Songlines*

Sotiris Kampanelis† Cardiff University

Aldo Elizalde‡ Queen's University Belfast

Yannis M. Ioannides§ Tufts University

August 22, 2026

Abstract

Can colonisers' reliance on Indigenous knowledge leave a lasting mark on economic geography? This paper studies how knowledge transmission from Indigenous societies during colonisation shaped the long-run economic geography of settler states. We focus on Australia, where Aboriginal people encoded routes, water sources, and seasonal knowledge in an oral tradition known as *Songlines*. We construct the first georeferenced database of almost 700 *Songlines* spanning mainland Australia and combine it with 19th-century exploration routes, cities, towns, railway foundation dates, and modern satellite data. *Songlines* strongly predict the routes of European explorations, whereas their counterfactual optimal paths based on geographic conditions (HMI Paths) exhibit little predictive power. Applying transformer-based language models to explorers' diaries, we find that explorations reporting cooperative interactions with Aboriginal people travelled significantly closer to *Songlines*, unlike those associated with hostile interactions. The imprint is persistent. Cities and towns founded before 1900, the earliest railway stations, and present-day nighttime light density all point to the same origin: *Songlines*.

Keywords: Indigenous knowledge; colonial exploration; knowledge transmission; economic geography; Australia.
JEL Codes: N77; O10; R12; Z10; Z13.

*We thank Luis Angeles, Panle Barwick, Jeanet Bentzen, Marcelo Bianconi, Yuxian Chen, Arcangelo Dimico, Evangelos V. Dioikitopoulos, Andrew Foster, Oded Galor, Luís Guimarães, Edward Glaeser, Pauline Grosjean, Eduardo Hidalgo, Erik Hornung, Allan Hsiao, Stelios Michalopoulos, Federico Masera, Frederic Robert-Nicoud, Jordan Rosenthal-Kay, Nathan Nunn, Sascha O. Becker, Ömer Özak, Elias Papaioannou, Adam Storeygard, Alex Trew, Serena Trucchi, and David Weil for helpful comments and discussions on our work. We would also like to thank the audiences at the seminars, conferences, and workshops in CRETE 2022, ASREC-London, ASREC-Australasia, ASREC-California, IESG-AERIP, Queen's University Belfast, RES 2023, RIDGE-TSW Bogota Forum, RIDGE-TSW Chile Forum, NBER-Urban Economics 2023, Uppsala Workshop in Economic History, Brown University, University of Toronto, Urban Economics Association in Washington DC, University of California Irvine, Australasian Development Economics Workshop 2025, UNSW, and the Winter North American Meetings of the Econometric society, San Francisco, 2025. We also thank Maria Bableki and Charlie Joyez for excellent research assistance, and John (HyunSu) Lee for his valuable assistance with NLP tools and the BERT model. We are grateful for financial support from the British Academy/Leverhulme Small Research Grants and Neubauer Family Foundation. The paper has been presented under an earlier title: "How Local Knowledge Fuelled Colonial Settlement and Long-Run Economic Activity: Evidence from Australia".

† Cardiff Business School, Cardiff University. kampaneliss@cardiff.ac.uk
‡ Department of Economics, Queen's University Belfast. a.elizalde@qub.ac.uk
§ Department of Economics, Tufts University. Yannis.Ioannides@tufts.edu

1. Introduction

The determinants of colonial expansion and its long-run economic consequences have been largely attributed to what European settlers brought with them, including institutions, technology, and disease (Acemoglu et al. 2001, 2002; Dell 2010; Easterly and Levine 2016; Glaeser et al. 2004; Sokoloff and Engerman 2000). This paper focuses on a dimension that has received little attention: the knowledge that colonisers acquired from Indigenous societies upon arrival. To study this, we exploit the case of Australia, where European colonisers had first to learn from Aboriginal guides how to navigate and survive in unfamiliar terrain before colonial power could take root. This was made possible through the adoption of a network of paths that Aboriginal communities had transmitted orally via songs across generations, in the absence of a written language, known as *Songlines*. Therefore, this paper investigates whether European adoption of *Songlines* left a measurable imprint on Australia's economic geography, and whether that imprint endures to the present day.

Songlines have deep roots. Aboriginal people had traversed Australia for millennia prior to European contact, accumulating detailed knowledge of the environment and cultural life of Aboriginal communities that Europeans could not have independently acquired. Critically, *Songlines* were not simply the most direct or environmentally optimal paths across the terrain. They were shaped by Aboriginal cultural life, following directions dictated by stories, traditions, and social relationships between communities (Neale 2017). For instance, a *Songline* connecting the southern and northern parts of Australia, would not follow the geographically shortest path. It would pass through a series of waypoints determined by the social and cultural relationships between the Aboriginal communities along the way. In adopting *Songlines*, European explorers were therefore accessing a body of knowledge they could not have acquired independently.

The historical evidence leaves little doubt about this. Early colonial documentation attests to the centrality of this knowledge system for European exploration and survival. For instance, Reynolds (1980) records that "*the natives were the parties who first guided the White Man through*

the intricacies of their forests, led them to their rivers, their springs, and rich pastures...The knowledge of their Country was thus acquired, was turned to account."[1]

To examine this empirically, we construct a new georeferenced dataset of *Songlines*. We do so by digitising anthropological data on Aboriginal trading-cultural sites published by McCarthy (1939), which provides one of the most comprehensive accounts of known origins, waypoints, and destinations associated with Aboriginal movement prior to colonisation. Using these records, we compile a catalogue of 581 such trading-cultural sites spanning mainland Australia.[2]

Using this dataset, we construct a georeferenced map of codified *Songlines*.[3] To do so, we employ more than 12,000 Indigenous heritage sites (primarily rock art sites) compiled from multiple datasets.[4] We first aggregate nearby sites into cultural hubs and use them to create a continuous measure of cultural intensity across mainland Australia. This measure assigns greater importance to areas surrounded by a dense concentration of cultural activity, while allowing the influence of each hub to decline with distance. We then reconstruct each *Songline* as the least-cost path (LCP) between ethnographically documented origins and destinations recorded by McCarthy (1939), such that movement is more likely to occur through culturally significant areas than across regions with little cultural evidence. Because the cost surface is derived exclusively from cultural heritage places

[1] To give an example of how traditional *Songlines* were sung and how they described travel routes we present in the Appendix Section H a contemporary song describing the path of a train, composed in the 1960s by a native Aboriginal. The song provides a glimpse into how *Songlines* once depicted travel paths before colonisation. Moreover, it highlights the enduring tradition of creating and passing down *Songlines* as a vital cultural practice among Aboriginal communities.

[2] McCarthy (1939)'s trilogy contains 671 unique locations. These include origins, waypoints, and destinations forming connected segments (e.g., a route from A to B and then from B to C, where B serves as the destination of the first segment and the origin of the next). Locations that fall in inland Australia are retained directly. When only one endpoint of a segment falls in water, it is moved to the nearest inland grid cell, since the other endpoint is already on the mainland. Segments where both endpoints fall in the sea or on distant islands are removed. Finally, a small number of segments whose two endpoints are located very close to one another are also excluded. After these adjustments, 581 origins, waypoints, and destinations are retained.

[3] We are grateful to Pauline Grosjean for proposing the term "codified" *Songlines*. The term reflects that, following extensive contact with Europeans, part of the knowledge encoded in the *Songlines* had been lost in oral transmission or not conveyed to McCarthy (1939), who recorded only the knowledge that had survived and was subsequently codified in his account. For simplicity, we will just use the term *Songlines* throughout the paper.

[4] These are drawn from the Aboriginal Cultural Heritage Register (DPLH-099), Department of Planning, Lands and Heritage, Government of Western Australia; the Australian Heritage Database (AHDB) and the National Heritage List (NHL), both maintained by the Department of Climate Change, Energy, the Environment and Water, Australian Government; approximately 170 academic publications on Australian rock art; and web-based research using public sources including Wikipedia, national park databases, tourism registers, and the Trove digital library.

and is not based on environmental variables such as terrain and water availability, the resulting routes reflect the cultural logic of movement embedded in Aboriginal knowledge systems. This yields 697 distinct *Songlines*, comprising the entire network.

We divide mainland Australia into a grid of 10 km $\times$ 10 km cells and record, for each cell, whether a reconstructed *Songline* passes through it. We then bring together several additional datasets at the same grid level: digitised 19th-century exploration routes, the founding dates of 249 cities and 1,133 towns, the opening years of 4,953 railway stations, and modern satellite measures of economic activity.

The first empirical analysis reaffirms historical records: *Songlines* predict 19th-century European exploration routes. Since our reconstructed *Songlines* reflect culturally determined routes rather than environmentally optimal ones, we also connect the same 581 trading-cultural sites using a least-cost path algorithm based solely on the Human Mobility Index (HMI) (Özak 2018). This index estimates the time required to traverse a square kilometre of land in pre-industrial times, accounting for temperature, humidity, cloud cover, slope, terrain, and risk of heat exhaustion. The resulting HMI Paths represent the routes that travellers would have followed in the absence of Aboriginal guidance, relying only on physical geography. Controlling for climate, elevation, ruggedness, proximity to the coast and capital, and historical mining activity during the colonial gold rush, *Songlines* consistently outperform HMI Paths as predictors of European exploration routes.

Several concerns may arise regarding the relationship between *Songlines* and European exploration routes. We address them in turn. First, we address potential archaeological recording and preservation bias by combining multiple independent heritage sources, excluding regions with low archaeological discovery potential, restricting the analysis to environments where evidence is most likely to survive, and accounting for differences in recording practices across states. Second, we demonstrate that the results are not driven by coastal accessibility, localised spatial confounding, or any single region by progressively excluding coastal cells, comparing contiguous grid cells through a pair fixed-effects design in the spirit of Dube et al. (2010), and sequentially dropping individual

states. Third, we show that the results hold after controlling for pre-colonial settlement intensity using radiocarbon and luminescence dates from archaeological sites across Australia. Across all of these exercises the estimated effect of *Songlines* remains stable and statistically significant.

We then test our main hypothesis directly by combining spatial data on exploration routes with textual evidence drawn from explorers' diaries. Specifically, we digitise 29 explorer diaries, comprising several thousand pages of daily accounts of encounters with Aboriginal people, ranging from guidance and cooperation to hostility and conflict. We apply transformer-based language models (BERT and RoBERTa) to classify interactions between European explorers and Aboriginal people as positive, neutral, or negative based on the language used. Sentences describing that explorers were "*shown a safer crossing*," "*guided to water*," or "*helped to find food*" reflect positive interactions and the sharing of local knowledge, whereas sentences such as "*refused us water*," "*blocked the track*," or "*threatened the party*" indicate hostility and lack of cooperation. These classifications allow us to construct an exploration-level average sentiment score, which we link to the corresponding exploration route. The results are striking: explorations reporting more cooperative interactions with Aboriginal people travelled significantly closer to *Songlines*, while hostile explorations drifted away.

Lastly, our findings suggest that the imprint of *Songlines* extends beyond early explorations. Cities and towns founded before 1900, the earliest railway stations, and present-day patterns of nighttime light density all cluster along *Songlines*, whereas HMI Paths are at most weakly associated with these outcomes. The effect is concentrated in the early colonial period, consistent with an informational advantage that mattered most when the land was unfamiliar and diminished thereafter as technology, rail, and settlement density reduced the knowledge gap.

We contribute to the literature in various ways. First and foremost, we extend the literature on the economic effects of colonisation (Sokoloff and Engerman 2000; Acemoglu et al. 2001; Acemoglu et al. 2002; Dell 2010; Easterly and Levine 2016), which has emphasised what Europeans brought with them. We show that what they learned on arrival mattered too: orally transmitted Indigenous knowledge shaped the paths of exploration, the location of early settlements, and the

long-run geography of economic activity. Second, we speak to the deep-roots literature on pre-colonial factors (Michalopoulos and Papaioannou 2013; Angeles and Elizalde 2017), which has focused on centralised states and physical legacies; we show that stateless societies transmitting knowledge orally can leave equally persistent marks. It also extends work on ancient trade routes (Flückiger et al. 2022; Barjamovic et al. 2019) from path persistence to knowledge persistence — and, through our sentiment analysis, tests the knowledge-transfer mechanism directly. The result has implications for urban and spatial economics (Glaeser et al. 1992; Duranton and Puga 2014; Barsanetti 2021), offering a new source of first-mover advantage in the spatial distribution of cities. Methodologically, we provide a template for extracting relational sentiment from 19th-century archival text at scale, which can be applied to other colonial settings where similar diaries survive.

The remainder of the paper proceeds as follows. Section 2 introduces the *Songlines*. Sections 3 documents the role of Aboriginal guides in European exploration. Sections 4 through 9 present the data, empirical strategy, and main results. Section 10 concludes. An extensive set of supplementary materials backing up our analysis is given in online Appendices.

2. The *Songlines*

Aboriginal people traversed Australia for at least fifty thousand years prior to European contact (Flood 1983, p. 77). Without a written language, they developed a remarkable system for recording and transmitting navigational and ecological knowledge across generations: *Songlines*. A *Songline* is a path across the landscape encoded as a song, naming mountains, waterholes, landmarks, and territorial boundaries in sequence, such that knowledge of the song is equivalent to knowledge of the route (Chatwin 1987; Kerwin 2010; Norris and Harney 2014). A single *Songline* could span thousands of kilometres, and the same song was reportedly recognisable across multiple language groups along its path, adapting in style and rhythm as it traversed the continent while preserving the sequence of landmarks that made it navigable (Norris and Harney 2014).

Songlines served not only as navigational tools but also as the physical paths along which goods, people, and ideas moved across the continent, forming the foundation of a sophisticated

Aboriginal trading network. Goods exchanged along these routes included red ochre, stone axes, pearl shells, pituri, boomerangs, and ceremonial knowledge (Kerwin 2010). A journey along a *Songline* could take months and served several functions simultaneously, including attending ceremonies, exchanging material goods, and transmitting cultural goods, such as songs and dances (Norris and Harney 2014). As Kerwin (2010) argues, the trade routes maintained by Aboriginal societies were not inferior to other ancient trade networks, and cannot be understood without reference to the oral histories that governed them.

In sum, *Songlines* constituted an oral map of the Australian landscape, encoding navigational, ecological, and social knowledge within a single system of song and ceremony (Neale 2017; Norris and Harney 2014). Critically, this knowledge was not simply a record of the most direct or geographically optimal paths across the terrain. It was shaped by Aboriginal cultural life, following directions dictated by stories, traditions, and social relationships between communities, such that knowledge of the song was equivalent to knowledge of the route, the resources along it, and the social protocols governing its use (Neale 2017). For European explorers, adopting *Songlines* meant drawing on generations of accumulated Aboriginal knowledge that no outsider could have assembled independently.

3. Adopting Knowledge from Aboriginal Guides

The exploration of Australia in the 19th century relied heavily on the knowledge and skills of Aboriginal guides. European explorers, unfamiliar with the Australian landscape, relied on these guides to navigate, locate water and food, and communicate with local communities. For instance, Edward John Eyre was blunt about it in his diary: "*A European would perish in a district capable of supplying the necessaries of life, because he had not the experience to direct him where to search for food. It is equally the same with respect to procuring water — the native has many resources at command, where the white man would faint or perish from thirst.*" Explorers secured this help in two ways: by recruiting guides at the outset of an exploration, or by engaging with Aboriginal people encountered along the route (Clarke 2008; McLaren 1996). Patterns from four regions

illustrate what guides contributed; and what their absence cost.

Southeastern Australia. From the first colonies in southeastern Australia, Aboriginal guides were vital to European exploration efforts. Hamilton Hume drew on Aboriginal expertise from his first journey in 1814 and continued to do so throughout his career. His ability to communicate and negotiate with Aboriginal communities proved essential during his 1828 exploration with the British explorer Charles Sturt, securing safe passage and local knowledge that contributed directly to its success (Hume 1966). This pattern of dependence is further illustrated by Charles Throsby, whose explorations between 1817 and 1821 relied on Broughton, an Aboriginal man also known as Toodwick, to navigate unfamiliar country (Campbell 2005). George William Evans's 1812 exploration from Jervis Bay to the Shoalhaven River was guided in a similar fashion by Bundle, a Gundungurra man, along a traditional *Songline* used by the Dharug and Gandangara peoples to reach Wiradjuri territory (Bathurst Local Aboriginal Consultative Committee 2011). These explorations formed part of early efforts to cross the Blue Mountains, leading Evans to the identification of Bathurst, the first inland European settlement in Australia.

Central Australia. The harsh landscapes of Central Australia posed no less severe a challenge. Ernest Giles's survival during his explorations depended critically on his Aboriginal guide Tommy Oldham, whose ability to locate water sources, most notably Queen Victoria Spring, averted the collapse of the exploration (Giles 1889). John McDouall Stuart's explorations through the same region equally relied on Aboriginal trackers. Following the departure of his Aboriginal guide in 1858, Stuart's party struggled to survive the return journey, attesting to the importance of Indigenous knowledge for the success of these explorations (Stuart 1865).

Western Australia. Western Australia offers further evidence of this dependence. John and Alexander Forrest relied critically on their Aboriginal guide Windich, whose knowledge of water sources and forage for horses proved essential to their explorations of Lake Moore and Lake Barlee during the 1860s and 1870s (Forrest 1875). Peter Egerton Warburton's explorations through the same region depended in turn on Charley, whose tracking skills and knowledge of water sources were indispensable to the survival of the exploration (Warburton 1875).

Northern Australia. The same pattern of dependence characterised explorations in Northern Australia. Ludwig Leichhardt's 1844 exploration depended critically on his Aboriginal guide Harry Brown, whose corrections to Leichhardt's course proved essential to the survival of the party (Leichhardt 1847). Edmund Kennedy's exploration was no different, relying on his Aboriginal guide Jackey-Jackey. When Kennedy was fatally attacked, Jackey-Jackey stayed by his side until his death and subsequently navigated alone through unfamiliar terrain to reach the coast and report the exploration's fate (Kennedy 1849).

Overall, Aboriginal guides were not simply companions on European explorations; their knowledge and skills were what made those explorations possible. Notable examples, such as the partnerships between John Forrest and Windich, Ludwig Leichhardt and Harry Brown, and Peter Egerton Warburton and Charley, each contributed directly to opening the Australian interior to European settlement (Reynolds 1990; McLaren 1996; Clarke 2008). Appendix F provides a detailed account of Aboriginal guides and their respective European explorers across four regions of Australia, documenting their names and the nature of their contributions to each exploration.

4. Data and Construction of Variables

4.1. Songlines and Counterfactuals

The construction of the first georeferenced map of Australian *Songlines* proceeds in two steps. The first step establishes the network of trading-cultural sites (origins, waypoints, and destinations). The origin of this dataset is the trilogy of research articles by McCarthy (1939), an anthropologist at the Australian Museum who devoted years to compiling trading-cultural sites from early colonial accounts of Aboriginal exchange networks. His work remains the most comprehensive single source on pre-contact Aboriginal trade geography. From these publications, we manually extract 581 trading-cultural sites covering all of mainland Australia, each documented with the specific goods exchanged and the tribes involved. Two examples drawn from the source material are presented below.

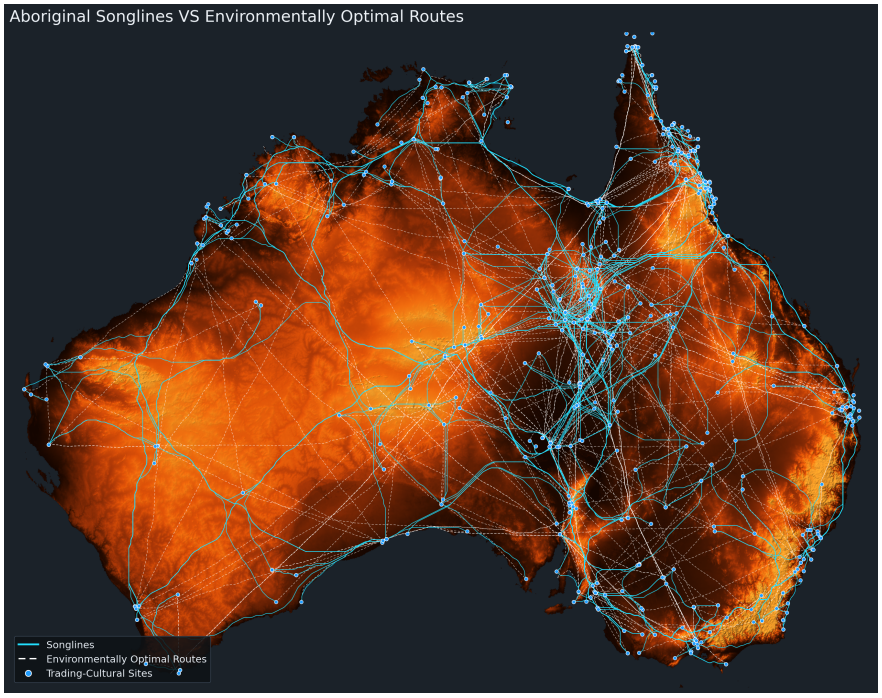


Figure 1. Aboriginal *Songlines* versus Environmentally Optimal Routes in Australia

Notes: This map contrasts the reconstructed *Songlines* (cyan solid lines) with environmentally optimal routes (white dashed lines) across mainland Australia. *Songlines* are least-cost paths connecting McCarthy (1939)'s 581 documented trading-cultural sites (blue dots), routed through a cultural cost surface built from more than 12,000 recorded Indigenous heritage places. Environmentally optimal routes connect the same endpoints using only the Human Mobility Index terrain cost of Özak (2018). The systematic divergence between the two networks reflects Aboriginal cultural knowledge embedded in the routes. See Appendix E for full construction details.

Example 1: "*Jessop...says that at 'these three places, Noarlunga, Augusta and Aroona, situated at distances of 150 miles in a direct line from south to north, where they interchanged their respective earths or clays, the natives drove also a good trade in skins with those who lived further inland'*".

Example 2: "*Johnston and Cleland. . . state that 'it is probable that the supply of pituri of the Ooldea area is secured from the Musgrave and Everard Ranges. . . along the recognised and known trade routes between these areas. . . there is a native trade route passing more or less north-east from Alice Springs via Aritunga, Ambalindum, the Plenty, MacDonald Downs, The Sandover, thence via the native wells to the Georgina at Lake Nash'*".

The second step connects the 581 trading-cultural sites into culturally oriented routes. Instead of using terrain or environmental features, the routing algorithm draws exclusively on more than 12,000 recorded Indigenous heritage places to guide the paths through culturally significant country.

We compiled this dataset from five independent sources: the Aboriginal Cultural Heritage Register published by the Department of Planning, Lands and Heritage, Government of Western Australia; the Australian Heritage Database (AHDB) and the National Heritage List (NHL), both maintained by the Department of Climate Change, Energy, the Environment and Water; approximately 170 academic publications on Australian rock art; and web-based research using Wikipedia, national park databases, tourism registers, and the Trove digital library. Together, these sources cover a wide range of site types, including rock art, grinding areas, creation and dreaming narratives, ritual and ceremonial locations, meeting places, and quarries. These sites are placed into roughly 4,000 cultural hubs by grouping sites within 5 km of one another, and each hub is weighted to correct for uneven archaeological survey effort across states.[5]

Using these weighted hubs, we construct a continuous cultural influence surface across the country. Each hub's influence decays exponentially with distance (half-life $\lambda = 50$ km), so that cells surrounded by many documented Indigenous heritage places score high while culturally empty land scores low. The raw influence is rank-scaled onto a 0–1 ladder, inverted into a travel-cost map. Culturally rich cells are assigned lower travel costs, while areas with little documented cultural activity are assigned higher travel costs. We further impose a remoteness penalty on regions located far from any recorded Indigenous heritage place. Major inland lakes ($\geq 1{,}000$ km^2) are treated as impassable barriers.[6]

For each observed *Songline*, a least-cost path algorithm traces the single lowest-cost route across this cultural cost surface, yielding 697 reconstructed *Songlines*. The resulting network consists of routes that follow culturally connected corridors rather than traversing areas with limited cultural evidence. Grid cell i receives a value of $Songlines_i = 1$ if at least one reconstructed *Songline* passes through it, and 0 otherwise. This yields 7,506 treatment cells out of 82,893; see

[5] Some parts of Australia have been recorded far more intensively than others. Without correction, well-surveyed regions would dominate the reconstruction simply because of survey effort rather than genuine cultural importance. Each hub is weighted by the inverse square root of the number of hubs recorded in its state.

[6] The void-friction component increases movement costs in locations far from any documented cultural hub. This does not assume that Aboriginal travellers moved exclusively between known sites; rather, it reflects the idea that areas containing documented cultural evidence are more likely to form part of broader cultural landscapes than areas where no evidence has been recorded.

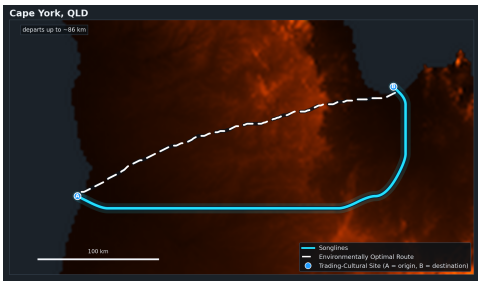
Figure 1. Full technical details of the construction procedure, including formulas and all parameter values, are reported in Appendix E.[7]

The cultural cost surface ensures that the reconstructed *Songlines* reflect Aboriginal cultural logic rather than terrain conditions. To illustrate this distinction, we construct counterfactuals to our *Songlines* named as HMI Paths. The construction proceeds as follows. We take the same set of origin–destination pairs used for the *Songlines* (the McCarthy (1939)'s trading-cultural sites) and run a least-cost path algorithm using only the Human Mobility Index (HMI) as the cost surface (Özak 2018). The HMI estimates the time required to traverse a square kilometre of land in pre-industrial conditions, accounting for temperature, relative humidity, cloud cover, slope, terrain type, and risk of heat exhaustion. It is derived entirely from physical geography and contains no cultural or archaeological information. The resulting paths represent the routes that would have connected Aboriginal trading-cultural sites had travellers relied solely on the physical landscape without any knowledge encoded in rock art sites or other cultural markers. Grid cell i receives a value of HMI Paths$_i = 1$ if at least one such path passes through it, and 0 otherwise.

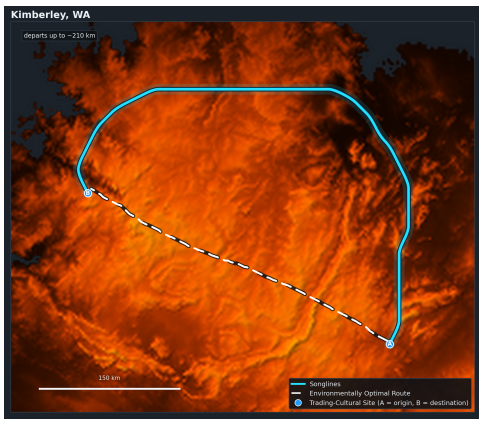
Because the HMI Paths use the same trading-cultural sites (origins, waypoints, and destinations) and the same cost algorithm as the *Songlines* but omit the cultural layer of the cost surface, they provide a sharp counterfactual: any difference in the predictive power of *Songlines* versus HMI Paths may be attributed to the Aboriginal practices embedded in the cultural cost surface. The HMI Paths are plausibly exogenous to the actions of both Aboriginal people and European explorers, as the travel cost parameter is derived entirely from physical geography. Figure 1 overlays the reconstructed *Songlines* against these environmentally optimal routes. The divergence between the two networks is striking. Although Aboriginal travellers could have followed the environmentally optimal path between any pair of trading-cultural sites, they frequently chose routes that departed substantially from the terrain-dictated shortest paths. Figure 2 presents three illustrative cases. In

[7] The structure of the routes displays substantial heterogeneity: some *Songlines* extend over very long distances with few or no waypoints, while others are shorter and characterised by a dense sequence of waypoints. This is consistent with the historical record. A notable example is the Dieri people of the Lake Eyre Basin, who are documented to have travelled directly to the Adnyamathanha some 500 kilometres to the south to obtain prized red ochre from the Flinders Ranges quarries, bypassing more accessible ochre quarries entirely (McBryde 1987).

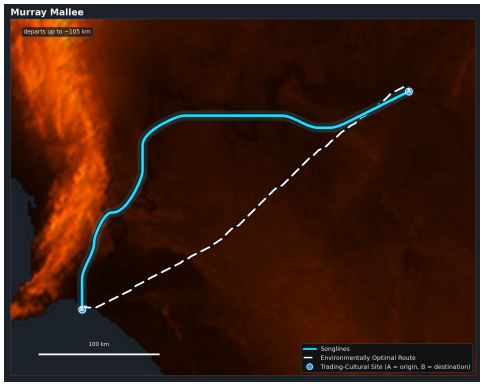
Cape York (Queensland), the reconstructed *Songline* departs by up to 86 km from the environmentally optimal route. In the Kimberley (Western Australia), the divergence reaches approximately 210 km, with the *Songline* sweeping around the plateau to pass through culturally significant country. In the Murray-Mallee region, the departure is roughly 105 km. These differences are consistent with the role of culture in Aboriginal travel: routes were shaped mainly by the network of inter-tribal relationships, ceremonial obligations, and access rights that governed movement across Australia.



(a) Cape York, QLD — departs up to ~86 km



(b) Kimberley, WA — departs up to ~210 km



(c) Murray-Mallee — departs up to ~105 km

Figure 2. Examples of *Songlines*–Optimal Route Divergence (HMI Paths)

Notes: Each panel shows a single origin–destination pair. Cyan solid lines are the reconstructed *Songlines*, routed through the cultural cost surface. White dashed lines are the environmentally optimal routes, computed using only the HMI terrain cost. Blue dots mark the McCarthy (1939)'s trading-cultural sites (A = origin, B = destination). In each case, the *Songline* deviates substantially from the terrain-optimal path, consistent with Aboriginal travellers following culturally defined routes shaped by inter-tribal relationships and ceremonial obligations rather than selecting the shortest or easiest geographic path.

4.2. Unit of Analysis and Outcomes

Our unit of observation is a regular grid cell covering the territory of Australia. Of the 82,893 grid cells in the original sample, we exclude the coastal layer of cells from the analysis. This restriction is motivated by the fact that Indigenous populations, *Songlines*, and European exploration activity were all disproportionately located along the coast. Retaining these cells would therefore create a mechanical correlation between *Songlines* and exploration routes, potentially biasing the results in favour of our hypothesis. By removing the first layer of coastal pixels, we focus on the interior of the country, where both Indigenous settlement and European exploration were substantially less intensive and where the relationship between *Songlines* and exploration can be examined more conservatively. This approach ensures that our findings are not driven by the historical concentration of human activity along Australia's coastline, reducing the sample from 82,893 to 79,722 grid cells. Each cell carries three types of outcome variable:

Exploration routes. We digitise Robinson (1927)'s map of 19th-century European inland exploration, covering 1813–1901 (Figure A.1).[8] A grid cell takes value one if crossed by at least one exploration route. This produces 12,135 cells with 1s out of 79,722.

Settlements and infrastructure. For cities, we use 249 foundation dates from Kampanelis (2019); for inland towns, 1,133 dates hand-collected from Elder (2024). For railways, we manually collect the opening year and location of 4,953 stations from local archives, historical pamphlets, and railway museum records, comprising nearly the full universe of Australian railway stations.

Modern economic activity. We use VIIRS nighttime light data (Elvidge et al. 2017) and GHSL population density (European Commission 2023) for 2015. Nighttime light enters the analysis as the mean VIIRS radiance within each grid cell. Appendix G shows that (the logarithm of) nighttime-light intensity at the Local Government Area level is positively associated with mean and median employee income. This relationship suggests that nighttime luminosity is a reliable proxy for local economic activity and long-run development in the Australian context, supporting

[8] Although this is a detailed map showing both nodes and sections of the early explorations, we should also note the limited technology of geographical tools in the early 20th century.

its use as an outcome measure in our analysis.

4.3. Controls

Although our *Songlines* are reconstructed exclusively using Indigenous heritage places, they may nevertheless pass through environmentally favourable locations that facilitated movement and settlement. To account for this, our empirical analysis controls for a comprehensive set of baseline geo-climatic, accessibility, and historical covariates, thereby ensuring that the estimated effects of *Songlines* are not simply capturing underlying environmental advantages. These include latitude and longitude, elevation, ruggedness, agricultural suitability, precipitation and temperature (and their variability), water cover, distance to the coast, distance to major urban centres (including the national and state capitals), an indicator for second-layer coastal cells, and proximity to historical mines.[9] The variable on historical mines is included to absorb the effects of the 19th-century gold rush on the spatial distribution of early economic activity.[10]

5. Empirical Strategy and Results

The empirical analysis examines the association between *Songlines* and European exploration routes, estimating the following model:

$$ExplorationRoute_i = \alpha + \beta\, Songlines_i + \gamma\, HMI\ Paths_i + Z_i'\theta + \varepsilon_i, \tag{1}$$

where i indexes grid cells covering mainland Australia. The dependent variable $ExplorationRoute_i$ takes the value 1 if grid cell i contains at least one European exploration route, and 0 otherwise. The main variable of interest, $Songlines_i$, takes the value 1 if grid cell i hosts at least one *Songline*, and 0 otherwise. $HMI\ Paths_i$ takes the value 1 if grid cell i hosts at least one *HMI Path*, and 0 otherwise.

[9] While the first layer of coastal grid cells has been excluded due to data limitations as explained in Section 4.2, we have created a coastal dummy using the second layer of coastal cells as these cells still offer valuable insight into the proximity of places to the coast.

[10] Historical mines are selected because the less advanced technology of the early 20th century suggests distances from those mines are irrelevant for the modern economy, given today's transport technology and infrastructure, and are therefore likely to be exogenous. Active mines are excluded from this measure, as their number and location may be correlated with contemporary economic activity, rendering them endogenous.

Z_i' is a vector of our baseline geo-climatic, accessibility, and historical covariates. All specifications include state fixed effects to absorb time-invariant differences across Australian states, including variation in archaeological recording practices, heritage management policies, historical settlement patterns, and the intensity of data collection efforts. Standard errors are clustered at the Local Government Area level throughout to account for potential spatial correlation and shared unobserved factors among grid cells within the same administrative unit.[11]

Table 1 reports the results of estimating equation 1. Column (1) establishes the association between *Songlines* and European exploration routes, conditioning only on state fixed effects and HMI Paths. The estimated coefficient on *Songlines* is positive and significant at the 1% level. Column (2) introduces geo-climatic baseline controls, and the association remains stable. The most conservative estimate, reported in column (3), further accounts for accessibility and historical controls.

Table 1. Songlines and European Explorations

	Dependent Variable: European Exploration Route (dummy)		
	(1)	(2)	(3)
Songlines	0.0139***	0.0130***	0.0121***
	(0.0033)	(0.0032)	(0.0032)
HMI Paths	0.0044	0.0025	0.0019
	(0.0032)	(0.0031)	(0.0031)
Geo-Climatic Controls		✓	✓
Accessibility Controls			✓
Historical Controls			✓
State FE	✓	✓	✓
N	79,722	79,722	79,722
R^2	0.008	0.013	0.015

Notes: The unit of observation is a grid cell of 10km × 10km. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable is a dummy variable indicating whether grid cell i contains a European exploration route. *Songlines* and HMI Paths are standardised. Column (1) includes only the treatment variables and state fixed effects. Column (2) additionally controls for geo-climatic characteristics. Column (3) further accounts for accessibility and historical controls. All specifications include state fixed effects. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

The estimated coefficient on *Songlines* is 0.0121. Since the *Songlines* variable is standardised, we rescale the coefficient by its standard deviation (0.283) to recover the effect of a discrete change

[11] *Songlines* and HMI Paths are standardised to have mean zero and unit standard deviation, allowing for direct comparison of coefficients across specifications.

from zero to one, yielding an effect of approximately $0.0121/0.283 \approx 0.043$. This implies that grid cells through which a *Songline* passes are about 4.3 percentage points more likely to contain a European exploration route. Given an unconditional probability of 15.22% across all grid cells, this corresponds to an increase to approximately 19.52% (an increase of about 28% relative to the sample mean).

HMI Paths are positively associated with exploration routes across all specifications, but their estimated magnitude is small and statistically insignificant, consistent with the interpretation that *Songlines* capture factors beyond purely geographic determinants of movement. The stability of the *Songlines* coefficient across specifications, and in the presence of HMI Paths, further supports this interpretation.

6. Robustness Checks

As a principal robustness check, we examine whether McCarthy (1939)'s documented trading-cultural sites predict European exploration routes after accounting for the terrain-only HMI Paths counterfactual. As reported in Table C.1, trading-cultural sites are positive and statistically significant across all specifications, while HMI Paths remain small and statistically insignificant. This pattern is consistent with the interpretation that culturally embedded Aboriginal route knowledge, rather than terrain alone, helps explain the geography of European exploration.

Moreover, the trading-cultural sites themselves may independently attract economic activity, much as historical portage sites do in Bleakley and Lin (2012). If the *Songlines* effect were driven by the endpoints rather than by the routes connecting them, it would disappear once those locations are accounted for. Table C.2 tests this by progressively controlling for proximity to the origin and destination sites identified by McCarthy (1939), then dropping all grid cells that contain such a site, and finally excluding all cells within a 50 km buffer around them. The coefficient on *Songlines* remains virtually unchanged in every case, confirming that the effect operates through the route knowledge encoded in the paths rather than through the specific trading-cultural locations at their endpoints.

In addition, several concerns may arise regarding the interpretation of the relationship between *Songlines* and European exploration routes. In particular, the results may reflect differences in the completeness of the archaeological record, variation in regional accessibility, or patterns of pre-colonial Aboriginal Settlement rather than the influence of *Songlines* themselves. We address these concerns through a broad set of exercises presented in Section C.

First, we address potential archaeological recording bias. By construction, the cultural influence surface combines information from multiple independent heritage datasets and adjusts for differences in recording intensity across states. Moreover, the results remain robust to excluding highly arid regions where archaeological discovery is less likely (Table C.3) and to sequentially omitting individual Australian states from the sample (Table C.5). All specifications also include state fixed effects to absorb systematic differences in archaeological recording, heritage management, and data collection practices.

Second, we address concerns related to accessibility and spatial confounding. The results remain robust to progressively excluding coastal areas and focusing on increasingly remote parts of the Australian interior (Table C.4), where successful exploration was more likely to depend on Aboriginal guides and Indigenous knowledge. They also remain robust when comparing *Songlines* cells only to immediately adjacent non-*Songlines* cells through a contiguous-neighbour fixed-effects design (Table C.10), ensuring that identification comes from highly local variation between geographically contiguous locations.

Third, concerns may arise that *Songlines* merely identify areas with historically larger Aboriginal populations. To address this, we control for *Pre-colonial Settlement* using radiocarbon and luminescence dates from the AustArch database, which provide direct evidence of long-term human presence and land use across Australia, and further restrict the sample to environments where archaeological evidence is most likely to be preserved and recovered. Table C.9 presents these results and additionally includes a specification that absorbs fixed effects for 436 Aboriginal language groups, accounting for unobserved pre-colonial institutional and societal heterogeneity. The *Songlines* coefficient remains stable and highly significant throughout.

Finally, the findings remain robust to alternative approaches for dealing with spatial dependence and omitted variables, including Conley spatial-HAC standard errors (Table C.6), Oster's coefficient-stability test (Table C.7), and spatial first-difference specifications (Table C.8). Across all exercises, the estimated effect of *Songlines* remains stable and statistically significant, suggesting that the relationship between *Songlines* and European exploration is unlikely to be driven by archaeological recording practices, regional accessibility, pre-colonial population density, or other forms of spatial confounding.

7. Adoption of Knowledge: Sentiment Analysis

The previous section shows that European exploration routes closely followed Aboriginal *Songlines*. The key question is whether this happened because Europeans relied on Aboriginal knowledge. The idea is that European explorers, who were unfamiliar with the landscape, relied on Aboriginal knowledge about routes, water, and terrain, either by travelling with Aboriginal guides or by receiving help along the way. This implies a clear pattern. Explorations that involved more cooperation with Aboriginal people should be closer to *Songlines*, while those with little or no cooperation should be further away.

To test this hypothesis, we draw on the diaries of 29 European explorers whose inland explorations are documented in Robinson (1927)'s map, comprising several thousand pages of daily accounts.[12] These diaries are drawn from *Project Gutenberg Australia* through eBooks, supplemented by additional online repositories and archival materials held at the *State Library of New South Wales*. They provide direct evidence on the nature of European-Aboriginal interactions during each exploration, describing practical decisions, movement through the landscape, and encounters with Aboriginal people. To classify the sentiment of each sentence from the explorers' diaries that

[12] Although we focus on 29 key explorers identified in Robinson (1927)'s map, for whom we can trace detailed routes and consult original diaries, we also collect an additional set of 48 inland and maritime explorers as a separate exercise. For these explorers, narrative and biographical sources document assistance from Aboriginal people, drawing on the Australian Dictionary of Biography, local council records, and material held at the State Library of New South Wales. For each explorer, at least one reliable source confirms the presence of one or more Aboriginal guides. The full dataset, including the underlying sources for each explorer, is available in the online appendix and can be accessed in the following online file here.

refers to Aboriginal people as positive, neutral, or negative, we use two transformer-based language models, BERT and RoBERTa. These models read the whole sentence and use the surrounding words to understand whether the text describes cooperation, reliance, or conflict, even when this is not stated directly. If explorations with more cooperative interactions are systematically closer to *Songlines*, this provides direct evidence that European explorers followed Aboriginal knowledge, rather than simply moving along routes determined by environmental features.[13]

The construction of the sentiment dataset proceeds in three steps. The first step processes the raw diary text into a clean corpus of daily exploration narratives. We apply HTML parsing to diaries sourced from *Project Gutenberg Australia* through eBooks to retain only the narrative text, excluding menus, headings, appendices, and metadata. For diaries obtained from additional online repositories and archival materials held at the *State Library of New South Wales*, we apply optical character recognition (OCR) to extract the text, followed by equivalent cleaning procedures. All diaries are converted into plain text files containing only day-by-day descriptions of events, without editorial comments or supplementary material. We normalise the text by removing unnecessary line breaks and correcting typographic inconsistencies, and handle punctuation to ensure that common abbreviations do not lead to incorrect sentence splitting. Sentences shorter than six words are discarded, as these typically correspond to headings, fragments, or contextual notes that do not convey meaningful narrative information.

The second step identifies which sentences in the cleaned corpus describe any form of relation-

[13] Other approaches to sentiment analysis include lexicon-based methodologies, where sentiment is inferred using predefined dictionaries of positive and negative words and measured by counting their number or frequency (relative share) within a text (Shapiro and Wilson 2022). However, we argue that such dictionary-driven techniques, and similar rule-based methods, are not well suited to interpreting historical diaries and comparable archival sources, where language is highly context-dependent, archaic, and often semantically ambiguous. In addition, modern sentiment lexicons may not align with the vocabulary and linguistic conventions of the historical period under study, further reducing the accuracy of sentiment classification in this context. In other research that uses contemporary text, such as news articles, political speeches, or social media posts, relational tone is often summarised with polarity measures that capture the balance between positive and negative expressions (Shapiro and Wilson 2022). These measures focus on clearly emotional language. In the case of nineteenth-century exploration diaries, however, this approach is less appropriate. Many interactions between explorers and Aboriginal groups are described in neutral terms. These neutral passages often reflect routine contact, the use of guides, or a non-hostile environment that allowed explorations to continue. Excluding such observations risks overlooking the non-hostile and cooperative conditions that enabled exploration to proceed through Indigenous territories. In this setting, neutrality does not mean the absence of relations. It often signals stable or non-conflictual interaction. For this reason, relying only on sentiment polarity would provide an incomplete picture of the relationships that shaped exploration routes.

ship between Aboriginal people and Europeans. To do so, we hand-label 1,599 sentences drawn from four exploration diaries chosen to span the continent and the century (John Oxley, Ludwig Leichhardt, Ernest Giles, and David Carnegie) as either RELATIONSHIP or NO RELATIONSHIP, and fine-tune a pretrained BERT model on these hand labels; no automated or external-model labels are used. A sentence is labelled RELATIONSHIP if it reveals any form of interaction or relational dimension between Europeans and Aboriginal people, including direct interactions such as conflict, assistance, guidance, or exchange of goods, expressions of perceptions or judgements, and references to Aboriginal people that can be identified from the surrounding text even when not explicitly named. Sentences are labelled NO RELATIONSHIP if they contain no explicit or implicit reference to Aboriginal people, such as descriptions of geography, flora, logistics, or administrative details. For instance, sentences from Ernest Giles's diary such as *"The natives threw stones and spears at the invaders, but nobody was killed"* and *"Here we met a considerable number of natives, and although the women would not come close, several of the men did, and made themselves useful by holding some of the horses' bridles and getting firewood,"* are classified as RELATIONSHIP, whereas *"I named these mounts Ayers Range"* or *"These trees are very handsome, with round umbrageous tops, the leaves are round and fringe-like"* are classified as NO RELATIONSHIP.

We assess whether the classifier generalises to diaries it has never seen using leave-one-diary-out cross-validation, which yields a mean out-of-sample F1-score of 0.725.[14][15] On a held-out sample of hand-labelled sentences the model classifies 92.5% correctly, recovering 71% of the sentences that genuinely describe a relationship while ensuring that 76% of those it flags are true

[14] Leave-one-diary-out cross-validation works as follows. Suppose we have four hand-labelled diaries, A, B, C, and D. We run four rounds. In round one we train the model on A, B, and C and test it on D, a diary the model has never seen; in round two we train on A, B, and D and test on C; in round three we train on A, C, and D and test on B; and in round four we train on B, C, and D and test on A. Every diary thus serves once as a completely unseen test set and three times for training. Averaging the four test scores gives the cross-validation F1-score of 0.725 reported in the text, an out-of-sample measure of how the model performs on a diary it did not learn from. Having confirmed that the model generalises, we then re-train it on all four diaries together, so that it learns from every hand-labelled sentence; this is the final classifier applied to the remaining, unlabelled diaries.

[15] The F1-score measures how well the model finds relationship sentences, on a scale from 0 (finds nothing correctly) to 1 (perfect). The model can make two mistakes: picking up sentences that are not really about the European–Aboriginal relationship, and overlooking sentences that are. The F1-score is high only when both mistakes are rare. We use it instead of the simple share of correct answers because relationship sentences are rare (about 14% of the text): a model that simply labelled every sentence as "no relationship" would be right 86% of the time yet find nothing, while its F1-score would be zero.

relationships (an F1-score of 0.735), indicating that it reliably detects relational sentences while maintaining a good balance between identifying relevant interactions and limiting false positives.

The third step classifies the identified relationship sentences into positive, negative, or neutral, depending on the nature of the interaction between Europeans and Aboriginal people. Importantly, this is not a standard emotional polarity task but a form of relational classification that identifies the type of interaction, distinguishing cooperative, neutral, and hostile encounters, rather than simply measuring narrative tone. Following the supervised text classification framework of Gentzkow et al. (2019), we hand-label 701 relationship sentences drawn from the four training diaries, classifying each as cooperative (positive), neutral, or hostile (negative).[16] Because hostile interactions are described with varied vocabulary (attack, theft, spearing, intimidation) and are comparatively rare, the sample deliberately over-represents conflict-rich passages to ensure that the model has enough negative examples to learn from. The final training set comprises 492 neutral, 136 negative, and 73 positive sentences.[17] We then fine-tune a RoBERTa-large model on this labelled sample. Figure 3 summarises the procedure used to construct the relational sentiment dataset.

The final dataset consists of 12,319 sentences classified as positive, negative, or neutral across all diaries. Negative sentiment, assigned a value of 0, captures attack, threat, coercion, intimidation, or attempts to force departure. Neutral sentiment, assigned a value of 0.5, captures co-presence, silent withdrawal, avoidance of contact, or non-engagement. Positive sentiment, assigned a value of 1, captures guidance, assistance, provision of information, or other forms of cooperative interaction.

The following examples illustrate the classification. Carnegie writes that without using "rough treatment" toward captured natives, the exploration "could not by any possibility have succeeded in crossing the desert." This is classified as negative because it explicitly justifies coercion against Aboriginal people as necessary for survival. Hann writes that the camp was never to be left unattended because the natives "are never to be trusted." This is coded as negative due to the clear

[16] In this framework, a researcher hand-labels a subset of documents into predefined categories (in our case, cooperative, neutral, or hostile) and uses these labels to train a statistical model that classifies the remaining documents automatically.

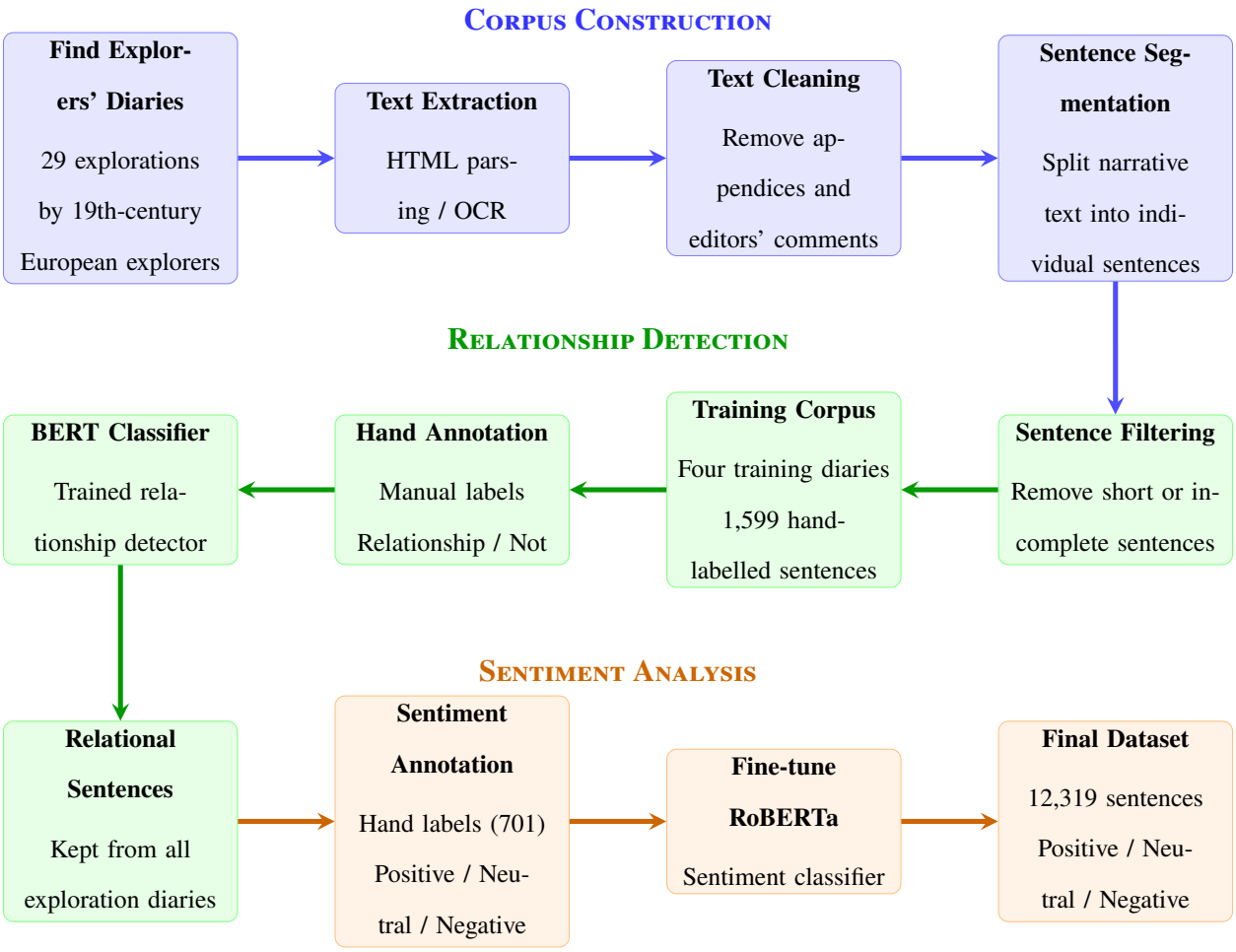
[17] The positive class is the smallest because cooperative interactions that materially aid the journey are genuinely rare in the diaries. Hostile sentences are over-sampled relative to their corpus frequency so that the classifier can distinguish them reliably.

expression of distrust and anticipated hostility. McDouall Stuart notes that if the camp were left with one man, "the natives might attack him, and destroy the lot." This is classified as negative because it reflects a perceived threat of violence. In contrast, Mitchell writes that he followed "the route proposed by my native friend" in order to approach the Darling. This is classified as positive because it shows reliance on Aboriginal guidance in determining the direction of travel. Likewise, Warburton reports that Charley followed native tracks and returned with news of a well. This is coded as positive because it reflects the use of Aboriginal knowledge to locate water and facilitate continued exploration.[18]

The sentence-level scores are then aggregated to the diary level by averaging across all sentences within each exploration route, as defined by Robinson (1927), yielding a route-level sentiment measure for each exploration. These scores per exploration are presented in Table 2. The sentiment variable ranges from 0.386 to 0.728, with a mean of 0.533. The relatively limited variation in this measure is consistent with the historical nature of exploration encounters, in which explorations typically interacted with multiple Aboriginal groups along their routes, experiencing both cooperative and tense moments depending on the specific tribe, circumstances, and stage of the journey. Sentiment values clustering around the midpoint of the scale are therefore precisely what one would expect from historical exploration narratives, and would raise concerns about the validity of the classification procedure only if they were systematically very low or very high, implying uniformly hostile or uniformly cooperative relationships throughout entire explorations, which is inconsistent with the historical record.

[18] A detailed description of the algorithm and the individual steps of the text-processing pipeline is provided in Appendix Section B.

Figure 3. Construction of the Relational Sentiment Dataset



Notes: The figure illustrates the pipeline used to construct the relational sentiment dataset from 19th-century exploration diaries. **Blue** boxes represent the corpus construction stage, where diary texts are collected, cleaned, and segmented into sentences. **Green** boxes represent the relationship detection stage, where sentences describing interactions between Europeans and Aboriginal people are identified using a BERT classifier trained on hand labels. **Orange** boxes represent the sentiment classification stage, where relational sentences are labelled as positive, neutral, or negative using a fine-tuned RoBERTa model. The final dataset contains 12,319 relational sentences used in the empirical analysis.

Table 2. Explorer Diaries and Mean Sentiment Toward Aboriginal People

Explorer	Diary Title	Sent.
A. C. Gregory	Journals of Australian Explorations by Augustus Charles Gregory and Francis Thomas Gregory	0.386
Allan Cunningham	Brief View of the Progress of Interior Discovery in New South Wales	0.545
Burke & Wills	Successful Exploration Through the Interior of Australia	0.611
Charles Sturt	His Life and Journeys of Exploration	0.494
Charles Winnecke	Mr. Winnecke's Explorations During 1883	0.621
David Carnegie	Spinifex and Sand	0.495
David Lindsay	Journal of the Elder Exploring Expedition	0.606
Edmund Kennedy	Expedition to Ascertain the Course of the River Victoria	0.577
Edward John Eyre	Journals of Expeditions of Discovery into Central Australia	0.470
Ernest Giles	Australia Twice Traversed	0.467
F. T. Gregory	Journals of Australian Explorations by Augustus Charles Gregory and Francis Thomas Gregory	0.421
Frank & Alexander Jardine	The Overland Expedition from Rockhampton to Cape York	0.429
Frank H. Hann	Do Not Yield to Despair: Frank Hugh Hann's Exploration Diaries in the Arid Interior of Australia, 1895–1908	0.648
Frederick Walker	Journal of Expedition in Search of Burke and Wills	0.539
George W. Evans	Two Journals of Early Exploration in New South Wales	0.597
Hume & Hovell	Journey of Discovery to Port Phillip, New South Wales	0.539
John Forrest	Explorations in Australia by John Forrest	0.579
John Lewis	Journal of Mr. Lewis's Lake Eyre Expedition, 1874–5	0.728
John McDouall Stuart	Explorations in Australia by John McDouall Stuart	0.499
John McKinlay	McKinlay's Journal of Exploration in the Interior of Australia	0.545
John Oxley	Journals of Two Expeditions into the Interior of New South Wales	0.494
Ludwig Leichhardt	Journal of an Overland Expedition in Australia	0.617
Peter Egerton Warburton	Journey Across the Western Interior of Australia	0.493
Thomas Mitchell	Three Expeditions into the Interior of Eastern Australia	0.525
Wentworth, Lawson & Blaxland	Journal of a Tour Across the Blue Mountains	0.519
William C. Gosse	W. C. Gosse's Explorations, 1873	0.643
William Hann	Diary of the Northern Expedition under the Leadership of William Hann	0.480
William Landsborough	Journal of Landsborough's Expedition	0.617
William O. Hodgkinson	The North-West Expedition (1876)	0.571

Using the sentiment data described above, we examine whether exploration routes associated with cooperation and positive interactions were located closer to *Songlines*. To do so, for each exploration route from Robinson (1927), shown in Figure A.1, we identify the grid cells it crosses. For each of these cells, we measure the distance to the closest *Songline*. This results in 12,041 cells out of the universe of 79,722. We then estimate the following model:

$$Y_{ie} = \alpha + \beta \bar{y}_e + \gamma X_{ie} + \mu_{g(i)} + \varepsilon_{ie}, \tag{2}$$

where i indexes grid cells and e indexes exploration routes. Y_{ie} denotes the geodesic distance from grid cell i, located on exploration route e, to the closest *Songline*. $\bar{y}_e$ is the exploration-level average relational sentiment expressed toward Aboriginal people in the diary of exploration e.[19] X_{ie} is a vector of control variables defined at the grid cell level, $\mu_{g(i)}$ are state fixed effects, where $g(i)$ denotes the state containing grid cell i. ε_{ie} is an error term. Moreover, standard errors are clustered at the exploration level. Clustering at this level accounts for potential correlation in the error terms across grid cells traversed by the same exploration, as observations within an exploration are likely to share common unobserved shocks, leadership characteristics, and strategic decisions.

Notably, Table 3, column (1) shows a negative and highly significant coefficient for the route-level sentiment when the dependent variable is distance to the nearest *Songline*. In simple terms, when the average sentiment along an exploration between European explorers and Aboriginal people is more positive, the route stays closer to *Songlines*. This fits our mechanism: cooperation and positive interaction with Aboriginal people made it easier for European explorers to learn from Aboriginal guidance, so European parties followed established routes instead of wandering off. We take this as direct evidence that Europeans adopted Aboriginal knowledge about how to travel, where to find water and food, and how to navigate safely, which influenced both how they explored and how they later set up economic activity.[20]

[19] If several sentiment-bearing routes cross the same pixel, the pixel takes the sentiment of the route whose line passes closest to the pixel centroid.

[20] In unreported results, we verify that the equal spacing of the three sentiment values (0, 0.5, 1) is not driving our findings. First, we replace the average sentiment score with the shares of positive and negative sentences entered separately, which imposes no assumption on the distance between categories. The positive share is negative and highly significant (-7.51, s.e. 2.18), while the negative share adds little explanatory power, indicating that the result operates through cooperation rather than the mere avoidance of conflict, consistent with our knowledge-transfer mechanism. Second, placing the neutral category off-centre, at 0.3 or 0.7 instead of 0.5, leaves the standardised coefficient negative and significant (-4.49 and -2.78, respectively). Third, we discard the sentiment scores altogether and keep only their ordering, assigning rank 1 to the explorer with the most hostile average sentiment and rank 29 to the one with the most cooperative, and use this rank in place of the score. The coefficient remains negative and significant (-3.49), even though the rank carries no information about the distance between sentiment categories. Because the sentiment measure is standardised before entering the regressions, the results are also invariant to any linear rescaling of the assigned values.

In column (2), after adding all baseline control variables, the coefficient remains negative and statistically significant, so the main result holds.[21] Sentiment scores (reported in Table 2) are standardised before entering the regressions (mean = 0, standard deviation = 1), so that the coefficient can be interpreted in terms of standard deviation changes and remains comparable across specifications. Since the dependent variable measures the distance to the nearest *Songline* in kilometres, the coefficient of -3.69 implies that a one standard deviation increase in the standardised sentiment measure is associated with a reduction of approximately 3.69 km in the distance to the nearest *Songline*. In the sample, the standardised sentiment measure ranges from -1.95 to 2.58, implying that moving from the lowest to the highest observed value corresponds to a decrease of roughly 16.7 km in the distance to the nearest *Songline*.

Table 3. Sentiment Analysis

	Dependent Variable: Distance to the nearest *Songline*	
	(1)	(2)
Sentiment	-5.173***	-3.688***
	(1.042)	(1.175)
Geo-Climatic Controls		✓
Accessibility Controls		✓
Historical Controls		✓
State FE	✓	✓
N	12,041	12,041
R^2	0.150	0.184

Notes: The unit of observation is a 10km × 10km grid cell along exploration routes. The dependent variable is the distance from each cell to the nearest *Songline*. Sentiment measures the standardised sentiment score of each exploration route, capturing the variation in the tone of interactions between European explorers and Aboriginal people recorded in explorers' diaries. All specifications include state fixed effects. The constant term is omitted for space. Robust standard errors clustered at the explorer level are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Table 4. Establishment of Cities and Towns

	All	est. before 1850	est. in 1850–1900	est. in 1900–1950	est. in 1950–2000
	(1)	(2)	(3)	(4)	(5)
Panel A: Cities					
Songlines	0.0017***	0.0008***	0.0007***	0.0001	0.0001
	(0.0003)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
HMI Paths	0.0003**	-0.0000	0.0004***	-0.0001**	0.0001
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
Geo-Climatic Controls	✓	✓	✓	✓	✓
Accessibility Controls	✓	✓	✓	✓	✓
Historical Controls	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓
N	79,722	79,722	79,722	79,722	79,722
R^2	0.007	0.007	0.002	0.000	0.001
Panel B: Towns					
Songlines	0.0032***	0.0014***	0.0017***	-0.0000	0.0001
	(0.0006)	(0.0004)	(0.0005)	(0.0001)	(0.0001)
HMI Paths	0.0015***	0.0009***	0.0005**	0.0000	0.0000
	(0.0004)	(0.0004)	(0.0003)	(0.0001)	(0.0000)
Geo-Climatic Controls	✓	✓	✓	✓	✓
Accessibility Controls	✓	✓	✓	✓	✓
Historical Controls	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓
N	79,563	79,563	79,563	79,563	79,563
R^2	0.040	0.035	0.011	0.001	0.001

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. Panel A examines the establishment of major cities, where the dependent variable in column (1) is a dummy equal to one if grid cell i contains a major city established by Europeans between 1788 and 2000. Panel B examines the establishment of towns, where grid cells containing a city are excluded. In both panels, columns (2) through (5) consider settlements established in the sub-periods before 1850, 1850–1900, 1900–1950, and 1950–2000, respectively. City establishment dates are drawn from Kampanelis (2019), covering 249 major cities in Australia. Town establishment dates are drawn from Elder (2024), covering 1,133 towns of which 866 are inland. *Songlines* and HMI Paths are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

8. Urban Evolution of Australia

8.1. Cities and Towns

The analysis shows that explorers followed routes close to *Songlines*. As Aboriginal knowledge shaped these journeys, early European settlements should also have appeared near these routes. Once paths inland were known, settlers likely moved along the same paths. We therefore examine whether early settlements are located near *Songlines*, focusing on the first settlement within each grid cell, since later settlements may reflect subsequent growth rather than the initial placement decision.

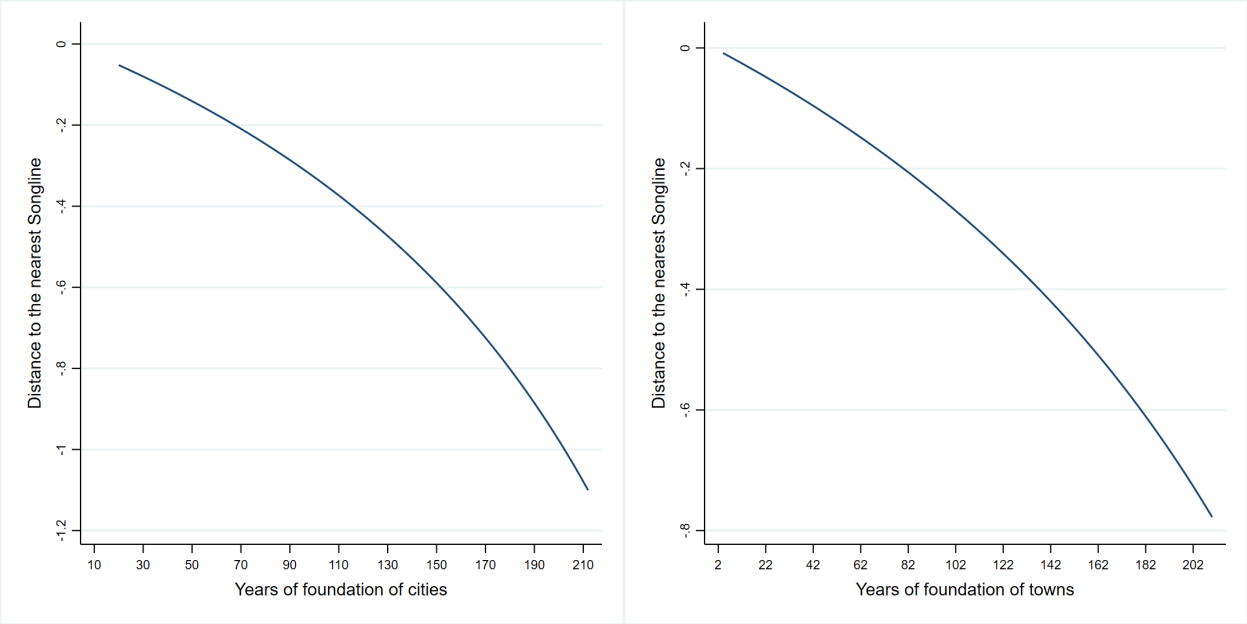
We begin with the timing and location of major cities, using data on 249 large Australian cities and their founding years. These cities account for a large share of the total Australian population and help trace early urban development. Certainly, given that Australia's largest cities are predominantly coastal, they possibly benefited more from early interactions between explorers, for whom coastal areas were more accessible, and Aboriginal communities. Therefore, we restrict the sample to inland cities, situated over 10 kilometres from the nearest coastline. We also ensure that each grid cell is assigned the earliest established city. This yields 159 grid cells, each containing an inland city and its foundation date. To see how the influence of *Songlines* may have shifted over time, we group cities into four periods: 1788–1850, 1850–1900, 1900–1950, and 1950–2000. For each grid cell, we create a dummy variable equal to one if a city was founded in that period, and use these as our dependent variables.

Table 4 Panel A reports results from regressing each period-specific city dummy on *Songlines* and HMI Paths. The coefficient on *Songlines* in column (1), which includes all cities, is positive and significant at the 1% level. Since only 159 out of 79,722 grid cells contain a major city, the unconditional probability of city presence is about 0.20 percent. The estimated coefficient implies that a one standard deviation increase in *Songlines* exposure raises this probability by 0.17

[21] In unreported results, we also control for the within-diary standard deviation of sentence sentiment to account for variation in the consistency of interactions described by each explorer. The coefficient on sentiment remains negative, highly significant, and becomes somewhat larger in magnitude (-5.52, s.e. 1.05), indicating that our findings are not driven by differences in the variability of reported interactions.

percentage points, an 85 percent increase relative to the mean. This suggests that *Songlines* strongly shaped where cities emerged. Columns (2) and (3) show that this relationship is concentrated among cities founded before 1900. For cities established after 1900, the estimates in columns (4) and (5) are near zero and statistically insignificant. Across all periods, the coefficient on HMI Paths remains small and largely insignificant, suggesting that geography alone does not explain the settlement pattern.

These results suggest that Aboriginal knowledge mattered most at the outset of colonisation. After 1900, forces such as rail expansion, population growth, and the gold rush opened up locations that had previously been difficult to reach, reducing the informational advantage that *Songlines* had previously provided.



(a) Emergence of cities and proximity to *Songlines*

(b) Emergence of towns and proximity to *Songlines*

Figure 4. Urban evolution and proximity to *Songlines*

Notes: These figures show the marginal association between the distance to the nearest *Songline* to major cities (Figure 4a) and towns (Figure 4b) and their corresponding year of establishment in Australia. Data on the foundation years of cities comes from Kampanelis (2019), and the data on the foundation years of towns comes from Elder (2024).

To further investigate the spatial expansion of the Australian urban system in relation to *Songlines*, we first calculate the distance of each city from the nearest *Songline*. We then examine the relationship between these distances and the founding year of each city by modelling how the

establishment of cities has evolved over time relative to their proximity to *Songlines*. The results in Figure 4a demonstrate a concave decreasing pattern. This finding implies that as European colonisation evolved, newer settlements are located further away from *Songlines*, with the marginal increase in the distance decreasing as time since foundation decreases. We interpret this result as evidence of *accelerating spatial agglomeration* around *Songlines* as the urban system thickens.

We extend our analysis to the evolution of economic activity into inland and typically arid regions, which may have developed under different patterns, not strictly linked to the *Songlines*. To do this, we hand-collect and digitise information from Elder (2024), which documents the years of establishment of 1,133 towns. From this dataset, we select 866 inland towns, each situated over 10 kilometres from the nearest coastline. After aggregating to the grid level and retaining only the earliest established town per grid cell, we obtain 831 unique grid-level observations. We further exclude 119 grid cells that already contain a city examined in the preceding analysis, yielding a final sample of 712 grid cells.

Table 4 Panel B reports the same analysis for towns. The results are remarkably consistent with those for cities. The coefficient on *Songlines* in column (1), which includes all towns, is positive and significant at the 1% level. Since only 712 out of 79,563 grid cells contain a town, the unconditional probability of town presence is about 0.89 percent. The estimated coefficient implies that a one standard deviation increase in *Songlines* exposure raises this probability by 0.32 percentage points, a 36 percent increase relative to the mean. Columns (2) and (3) reveal that this effect is driven by earlier settlements, with *Songlines* being positive and significant at the 1% and 5% levels for towns established before 1850 and between 1850 and 1900, respectively. In contrast, columns (4) and (5) show no significant effect for towns established after 1900, mirroring the pattern found for cities. HMI Paths also enters positively and significantly in columns (1) through (3), though the magnitude of its coefficient is considerably smaller, further underscoring the dominant role of *Songlines* in shaping the geography of early town settlement.

The results from Figure 4b align with our earlier findings (t-statistic=-2.64), showing that towns closer to the *Songlines* emerged earlier. This supports the idea that the *Songlines* acted as

focal points for the development of economic activity for more than two centuries. Even though inland areas are not as thickly settled as coastal ones, and the marginal effect of year of settlement weakens over time, clustering near *Songlines* is still present.

8.2. Early Infrastructure

The settlement patterns documented in the previous section naturally lead to a further question: once Europeans established towns and cities, did the transport networks they built follow similar routes? If early settlements emerged near *Songlines*, the roads and railways constructed to connect these places may also have been constructed in ways that followed similar lines. This would be consistent with a form of path dependence, where the initial transfer of Aboriginal knowledge shaped not only exploration and settlement but also the emerging infrastructure network, which in turn, may have driven the agglomeration of economic activity in the long-run.

To explore this idea, we hand-collected data on 4,953 historical railway stations across Australia. Specifically, we compiled information from multiple online and archival sources, including Wikipedia entries, local newspaper archives, historical pamphlets, and railway museum websites. From these materials, we identified the opening year and location of almost the full universe of railway stations in all Australian states.

The earliest steam-powered railway stations opened in the mid-1850s. For example, stations at Port Elliot (1854) and Victor Harbor (1854) in South Australia were among the first to operate. In Victoria, Flinders Street Station opened in 1854 as part of the line connecting Melbourne to Sandridge (now Port Melbourne). In New South Wales, the first railway line from Sydney to Parramatta was inaugurated in 1855, with Redfern Station (often referred to as the "Sydney Terminal") serving as its original city terminus.

Figure 5 illustrates the spatial and temporal expansion of the Australian railway network. Lighter red dots denote earlier railway stations, while the blue labels indicate the opening year for a representative subset of stations across the country. Using these data, we create indicator variables for early transport infrastructure at the grid-cell level. A grid cell receives a value of 1

if it contains at least one historical railway station, and 0 otherwise. Consistent with the approach used for cities and towns, we assign each grid cell its earliest railway station, capturing the initial arrival of railway infrastructure. This allows us to examine whether the early placement of the railway network followed *Songlines*. We distinguish three periods: up to 1900, given that railway construction began after 1850, 1900 to 1950, and 1950 to 2000.

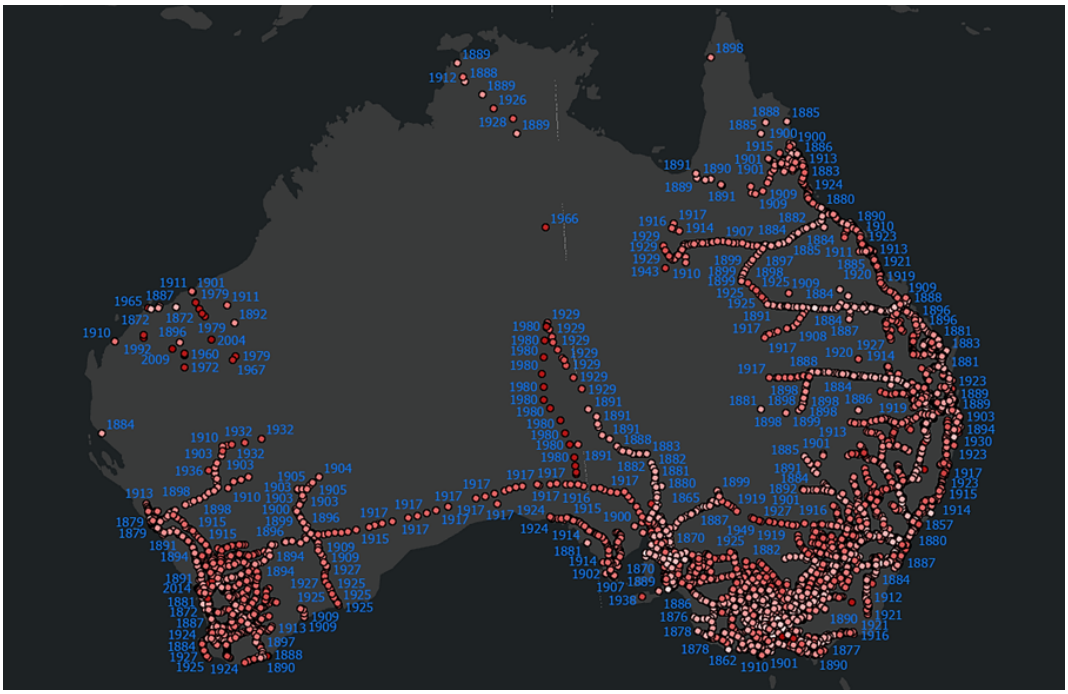


Figure 5. Construction of the Australian railway network

Notes: This figure maps the opening of 4,953 historical railway stations across Australia. Lighter red dots indicate earlier stations, and blue labels report the opening year for a representative subset of stations.

Table 5. Establishment of Railway Stations

	All Stations	Stations est. before 1900	Stations est. in 1900–1950	Stations est. in 1950–2000
	(1)	(2)	(3)	(4)
Songlines	0.0070***	0.0049***	0.0019**	0.0002
	(0.0013)	(0.0010)	(0.0008)	(0.0001)
HMI Paths	0.0007	0.0020***	-0.0014*	0.0001
	(0.0010)	(0.0008)	(0.0008)	(0.0001)
Geo-Climatic Controls	✓	✓	✓	✓
Accessibility Controls	✓	✓	✓	✓
Historical Controls	✓	✓	✓	✓
State FE	✓	✓	✓	✓
N	79,722	79,722	79,722	79,722
R^2	0.118	0.086	0.046	0.001

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable in column (1) is a dummy variable equal to one if grid cell i contains a railway station. Each grid cell is assigned its earliest established railway station. Columns (2) through (4) consider stations established in the sub-periods before 1900, 1900–1950, and 1950–2000, respectively. *Songlines* and HMI Paths are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

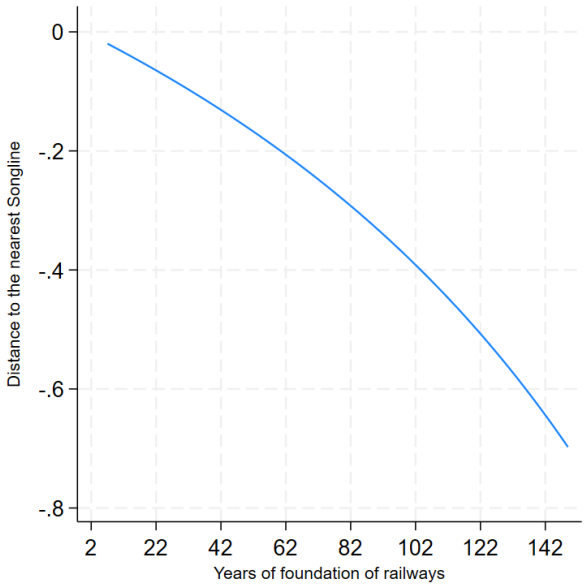


Figure 6. Railway expansion and proximity to *Songlines*

Notes: This figure shows the marginal association between the distance to the nearest *Songline* to railway stations and their corresponding year of establishment in Australia. Data on the foundation years of railway stations comes from historical records on railway expansion in Australia.

In Table 5, we regress a period-specific dummy variable for the construction of a railway station at the grid-cell level on *Songlines* and HMI Paths. The results closely mirror those reported

in Table 4. The coefficient capturing the overall presence of railway stations in column (1) is positive and statistically significant. The coefficient for stations built before 1900 in column (2) is positive and significant at the 1% level. For the period 1900–1950 in column (3), the coefficient remains significant at the 5% level. By contrast, the coefficient for stations constructed between 1950 and 2000 in column (4) is economically small and statistically insignificant.

Across all specifications, the HMI Paths variable has a negligible effect on railway station construction. These findings are consistent with those of the previous Section 8.1, where we show that *Songlines* shaped the early spatial expansion of European settlement, while over time other economic and institutional forces became more important in determining the geography of economic activity in Australia.

Moreover, following the previous Section 8.1, where we document the relationship between the founding of cities and towns and their proximity to *Songlines*, we replicate the same exercise for railway stations. Specifically, we compute the distance of each historical railway station to the nearest *Songline* and relate this distance to the station's opening year. This enables us to assess whether the temporal expansion of the railway network followed a similar spatial pattern to that observed for urban settlements.

Figure 6 reveals a comparable non-linear relationship. The earliest railway stations are located closer to *Songlines*, whereas stations constructed in later periods tend to be situated further away. Although the marginal increase in distance becomes smaller over time, the overall pattern suggests that *Songlines* continued to shape the geography of railway development throughout the 19th and 20th centuries, leaving a lasting imprint on Australia's transport network.

9. Long-Run Effects of *Songlines*

In this section, we examine the effect of *Songlines* on long-run economic development. Although Australia is classified as a developed economy, its interior regions exhibit substantially lower levels of economic activity. Following the established literature (Gibson et al. 2021; Henderson et al. 2012), we employ a binary indicator of nighttime light activity and population density as

proxies for local economic activity.

The estimating equation mirrors the exploration specification:

$$NightLight_i = \alpha + \beta\, Songlines_i + \gamma\, HMI\ Paths_i + Z_i'\theta + \varepsilon_i, \tag{3}$$

where $NightLight_i$ is mean nighttime luminosity (columns 1–3) or population density (column 4) in grid cell i; $Songlines_i$ and $HMI\ Paths_i$ are the standardised treatment and terrain-only control; Z_i is the vector of baseline geo-climatic, accessibility, and historical controls; and standard errors are clustered at the Local Government Area level.

Column (1) of Table 6 establishes a robust positive association between the presence of *Songlines* and contemporary economic activity, controlling for HMI Paths, significant at the 1% level. Column (2) introduces geo-climatic baseline controls, and the association remains stable. The most conservative estimate, reported in column (3), further accounts for accessibility and historical controls. The estimated coefficient on *Songlines* is 0.0202. Since the *Songlines* variable is standardised, we rescale the coefficient by its standard deviation (0.283) to recover the effect of a discrete change from zero to one, yielding an effect of approximately $0.0202/0.283 \approx 0.071$. This implies that grid cells through which a *Songline* passes exhibit nighttime luminosity about 0.071 units higher. HMI Paths are also positively and significantly associated with economic activity, though their estimated magnitude is roughly one-third that of *Songlines*, consistent with the interpretation that *Songlines* capture forces beyond purely geographic determinants of movement.

Column (4) of Table 6 reports results using population density as an alternative measure of economic activity, reinforcing these findings. The estimated coefficient on *Songlines* is 1.5230, indicating that a one standard deviation increase in *Songlines* is associated with a 1.52 unit increase in population density. HMI Paths are also positively and significantly associated with population density (coefficient 0.5651), though their estimated magnitude is again notably smaller than that of *Songlines*. Taken together, the evidence points to persistent magnet effects of *Songlines* on human settlement, consistent with a long-run cultural and economic legacy.

Table D.1 shows that the relationship between *Songlines* and long run economic activity is

Table 6. Long Term Effects

	Night Lights			Pop. Density
	(1)	(2)	(3)	(4)
Songlines	0.0208***	0.0207***	0.0202***	1.5230***
	(0.0049)	(0.0047)	(0.0047)	(0.3138)
HMI Paths	0.0038*	0.0068***	0.0076***	0.5651***
	(0.0021)	(0.0025)	(0.0025)	(0.1713)
Geo-Climatic Controls		✓	✓	✓
Accessibility Controls			✓	✓
Historical Controls			✓	✓
State FE	✓	✓	✓	✓
N	79,722	79,722	79,722	79,722
R^2	0.016	0.027	0.035	0.045

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable in columns (1) through (3) is mean nighttime luminosity (VIIRS). The dependent variable in column (4) is population density (GHSL, 2015). *Songlines* and HMI Paths are standardised. All specifications include state fixed effects. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

highly robust. Replicating the full set of robustness exercises from the exploration analysis, together with additional checks for luminosity and population density, leaves the estimated coefficient on *Songlines* positive and statistically significant throughout. The results are robust to alternative samples, spatial correlation, omitted variable bias, spatial first differences, and additional controls for pre-colonial settlement, Aboriginal language groups, and *Trading-Cultural Sites*. In several specifications, the estimated effect becomes even larger when the sample is restricted to the more remote interior.

10. Conclusion

This paper studies the long-run economic consequences of colonisers' reliance on Indigenous knowledge in Australia. Aboriginal people encoded routes, water sources, and seasonal knowledge in songs passed down across generations. These songs define a network of paths known as *Songlines*. European explorers, unfamiliar with the landscape, relied on Aboriginal guides to navigate, locate water and food, and communicate with local populations. Through such interactions, they adopted the knowledge embedded in *Songlines*. The paper establishes empirically that: (i) *Songlines* predict

the direction of 19th-century European exploration routes more strongly than environmentally optimal paths, reflecting the reliance of European explorers on Aboriginal navigational knowledge; (ii) explorations reporting cooperative interactions with Aboriginal guides travelled significantly closer to *Songlines*, while those associated with hostile interactions were further away, providing direct evidence for the knowledge-transmission mechanism; and (iii) the imprint of *Songlines* is persistent, with cities and towns founded before 1900, the earliest railway stations, and present-day economic activity all clustering along these routes. Routes based on geography alone show at most a weak association with the same outcomes.

The findings shed new light on the determinants of colonial expansion and its long-run consequences by showing that what colonisers learned upon arrival mattered alongside what they brought with them. They contribute to the deep-roots literature by demonstrating that stateless societies transmitting knowledge orally can leave marks on economic geography as persistent as those associated with centralised pre-colonial institutions. They further extend the literature on ancient trade routes by shifting the focus from path persistence to knowledge persistence. Methodologically, the application of language models to 19th-century exploration diaries provides a novel approach to extracting information and testing knowledge-transfer mechanisms in settings where comparable historical sources exist. More broadly, the findings highlight the role of Indigenous people not as passive subjects of colonisation, but as active contributors to the spatial foundations of modern economies.

References

Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). "The colonial origins of comparative development: An empirical investigation". In: *American Economic Review* 91.5, pp. 1369–1401.

— (2002). "Reversal of fortune: Geography and institutions in the making of the modern world income distribution". In: *The Quarterly Journal of Economics* 117.4, pp. 1231–1294.

Angeles, Luis and Aldo Elizalde (2017). "Pre-colonial institutions and socioeconomic development: The case of Latin America". In: *Journal of Development Economics* 124, pp. 22–40.

Barjamovic, Gojko, Thomas Chaney, Kerem Coşar, and Ali Hortaçsu (2019). "Trade, merchants, and the lost cities of the bronze age". In: *The Quarterly Journal of Economics* 134.3, pp. 1455–1503.

Barsanetti, Bruno (2021). "Cities on pre-Columbian paths". In: *Journal of Urban Economics* 122, p. 103317.

Bathurst Local Aboriginal Consultative Committee (2011). *The Wiradjuri people*. Bathurst, N.S.W.: Bathurst Local Aboriginal Consultative Committee.

Bleakley, Hoyt and Jeffrey Lin (2012). "Portage and path dependence". In: *The Quarterly Journal of Economics* 127.2, pp. 587–644.

Campbell, Keith (2005). "Broughton (c.1798–c.1850)". In: *Australian Dictionary of Biography*. URL: https://adb.anu.edu.au/biography/broughton-12820.

Chatwin, Bruce (1987). *The Songlines*. London: Jonathan Cape.

Clarke, Philip A (2008). *Aboriginal plant collectors*. Rosenberg Pub.

Dell, Melissa (2010). "The persistent effects of Peru's mining mita". In: *Econometrica* 78.6, pp. 1863–1903.

Dube, Arindrajit, T. William Lester, and Michael Reich (2010). "Minimum Wage Effects Across State Borders". In: *The Review of Economics and Statistics* 92.4, pp. 945–964. DOI: 10.1162/REST_a_00039. URL: https://doi.org/10.1162/REST_a_00039.

Duranton, Gilles and Diego Puga (2014). "The growth of cities". In: *Handbook of Economic Growth* 2, pp. 781–853.

Easterly, William and Ross Levine (2016). "The European origins of economic development". In: *Journal of Economic Growth* 21, pp. 225–257.

Elder, Bruce (2024). *Aussie Towns: A Definitive Travel Guide to Australian Towns and Destinations*. Accessed: 2024-09-10. URL: https://www.aussietowns.com.au/.

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh (2017). "VIIRS night-time lights". In: *International Journal of Remote Sensing* 38.21, pp. 5860–5879.

European Commission (2023). *GHSL Data Package 2023*. URL: https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2023.pdf?t=1683540422 (visited on 09/06/2023).

Flood, Josephine (1983). *Archaeology of the Dreamtime: The Story of Prehistoric Australia and its People*. Collins Publishers Australia.

Flückiger, Matthias, Erik Hornung, Mario Larch, Markus Ludwig, and Allard Mees (2022). "Roman transport network connectivity and economic integration". In: *The Review of Economic Studies* 89.2, pp. 774–810.

Forrest, John (1875). *Explorations in Australia*. London: Sampson Low, Marston, Low, & Searle.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as Data". In: *Journal of Economic Literature* 57.3, pp. 535–574.

Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li (2021). "Which night lights data should we use in economics, and where?" In: *Journal of Development Economics* 149, p. 102602.

Giles, Ernest (1889). *Australia Twice Traversed: The Romance of Exploration, Being a Narrative Compiled from the Journals of Five Exploring Expeditions into and through Central South Australia and Western Australia, from 1872 to 1876*. Vol. 1. London: Sampson Low, Marston, Searle, & Rivington. URL: https://books.google.gr/books/about/Australia_Twice_Traversed.html?id=gwsOAAAAYAAJ&redir_esc=y.

Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer (1992). "Growth in cities". In: *Journal of Political Economy* 100.6, pp. 1126–1152.

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004). "Do institutions cause growth?" In: *Journal of Economic Growth* 9.3, pp. 271–303.

Henderson, J Vernon, Adam Storeygard, and David N Weil (2012). "Measuring economic growth from outer space". In: *American Economic Review* 102.2, pp. 994–1028.

Hume, S. H. (1966). “Hamilton Hume (1797–1873)”. In: *Australian Dictionary of Biography*. URL: https://adb.anu.edu.au/biography/hume-hamilton-2211.

Kampanelis, Sotiris (2019). “It’s time for Westernization: the advantages of the early start for long-term economic development at the local level”. In: *Oxford Economic Papers* 71.4, pp. 996–1025.

Kennedy, Edmund (1849). “Expedition to Ascertain the Course of the River Victoria”. In: *Journal of the Royal Geographical Society of London* 19, pp. 193–199.

Kerwin, Dale (2010). *Aboriginal Dreaming Paths and Trading Routes: The colonisation of the Australian economic landscape*. Sussex Academic Press.

Leichhardt, Ludwig (1847). *Journal of an Overland Expedition in Australia, From Moreton Bay to Port Essington, A Distance of Upwards of 3000 Miles, During the Years 1844-1845*. London: T. & W. Boone.

McBryde, Isabel (1987). “Goods from another country: exchange networks and the people of the Lake Eyre Basin”. In: *Australians to 1788*. Ed. by D.J. Mulvaney and J.P. White. Sydney: Fairfax, Syme and Weldon, pp. 252–273.

McCarthy, Frederick David (1939). ““Trade in Aboriginal Australia, and ”Trade” Relationships with Torres Strait, New Guinea and Malaya””. In: *Oceania* 9.4, pp. 405–438.

McLaren, Glen (1996). *Beyond Leichhardt: Bushcraft and the Exploration of Australia*. Fremantle, W.A.: Fremantle Arts Centre Press.

Michalopoulos, Stelios and Elias Papaioannou (2013). “Pre-colonial ethnic institutions and contemporary African development”. In: *Econometrica* 81.1, pp. 113–152.

Neale, Margo, ed. (2017). *Songlines: The Power and Promise*. Canberra: National Museum of Australia Press.

Norris, Ray P. and Bill Yidumduma Harney (2014). “Songlines and Navigation in Wardaman and Other Aboriginal Cultures”. In: *Journal of Astronomical History and Heritage* 17.2, pp. 141–148.

Özak, Ömer (2018). "Distance to the pre-industrial technological frontier and economic development". In: *Journal of Economic Growth* 23.2, pp. 175–221.

Reynolds, Henry (1980). "The land, the explorers and the Aborigines". In: *Australian Historical Studies* 19.75, pp. 213–226.

— (1990). *With the white people: the crucial role of Aborigines in the exploration and development of Australia*. Penguin Books, Ringwood, Vic.

Robinson, H.E.C. (1927). *The discovery and exploration of Australia, 1519 to 1901 by sea and land*. Sydney: H.E.C. Robinson Ltd.

Shapiro, Adam H. and Daniel J. Wilson (2022). "Taking the Fed at Its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis". In: *The Review of Economic Studies* 89.5, pp. 2768–2805.

Sokoloff, Kenneth L and Stanley L Engerman (2000). "Institutions, factor endowments, and paths of development in the new world". In: *Journal of Economic Perspectives* 14.3, pp. 217–232.

Stuart, John McDouall (1865). *Explorations in Australia: The Journals of John McDouall Stuart during the Years 1858, 1859, 1860, 1861, & 1862, When He Fixed the Centre of the Continent and Successfully Crossed It from Sea to Sea*. Ed. by William Hardman. 2nd. London: Saunders, Otley, and Co.

Warburton, Peter E. (1875). *Journey Across the Western Interior of Australia: With an Introduction and Additions*. London: Sampson Low, Marston, Low, & Searle.

Online Appendices

A. Summary Statistics and Main Figures



Figure A.1. European exploration routes by Robinson (1927)

Notes: This map shows the European explorations routes as drawn by Robinson (1927).

Table A.1. Summary Statistics

	N	Mean	St.Dev.	Min	Max
Songlines and Explorations					
Songlines	79,722	0.09	0.28	0	1
HMI Paths	79,722	0.28	0.45	0	1
McCarthy (1939)'s Trading-Cultural Sites	79,722	0.00	0.06	0	1
Exploration Routes	79,722	0.15	0.36	0	1
Sentiment Analysis					
Sentiment Score	12,041	0.53	0.08	0.39	0.73
Distance to Closest Songline	12,041	30.89	35.65	0	240.22
Cities and Towns					
Cities (all)	79,722	0.00	0.04	0	1
Cities est. before 1850	79,722	0.00	0.03	0	1
Cities est. in 1850–1900	79,722	0.00	0.03	0	1
Cities est. in 1900–1950	79,722	0.00	0.01	0	1
Cities est. in 1950–2000	79,722	0.00	0.01	0	1
Towns (all)	79,563	0.01	0.09	0	1
Towns est. before 1850	79,563	0.00	0.07	0	1
Towns est. in 1850–1900	79,563	0.00	0.06	0	1
Towns est. in 1900–1950	79,563	0.00	0.02	0	1
Towns est. in 1950–2000	79,563	0.00	0.01	0	1
Railways					
Railways (any)	79,722	0.04	0.19	0	1
Railways est. before 1900	79,722	0.02	0.13	0	1
Railways est. in 1900–1950	79,722	0.02	0.14	0	1
Railways est. in 1950–2000	79,722	0.00	0.02	0	1
Economic Development					
Night Light	79,722	0.02	0.39	0	31.11
Population Density	79,722	1.84	29.80	0	1866.42
Controls for Geo-Climatic, Accessibility, and Historical Factors					
Latitude	79,722	−25.30	5.82	−38.75	−11.09
Longitude	79,722	133.55	10.38	113.38	153.50
Agricultural Suitability	79,722	1.09	0.59	0	7
Elevation	79,722	281.23	184.07	−15.17	1819.27
Ruggedness	79,722	16.32	23.42	0.15	423.69
Precipitation	79,722	448.01	308.57	128.92	3751.36
StDev Precipitation	79,722	6.51	15.52	0	646.34
Temperature	79,722	217.48	37.95	49.59	293.02
StDev Temperature	79,722	1.09	1.46	0	24.09
Coastal Dummy	79,722	0.02	0.14	0	1
Distance to the Sea	79,722	343.75	236.34	4.60	915.17
Distance to Sydney	79,722	2083.83	936.82	18.51	3865.17
Distance to State Capital	79,722	907.88	480.62	2.99	2301.82
Distance to Historical Mine	79,722	174.41	123.05	0.50	650.23
Water Percentage	79,722	0.02	0.08	0	1

Notes: The unit of observation is a grid cell of $10km \times 10km$. Sentiment Score and Distance to Closest Songline are from the sentiment analysis dataset (N=12,041). All other variables use the main dataset after dropping grid cells with area below 100 km^2 (N=79,722). Railway variables reflect the earliest established station per grid cell. Town variables are summarised over the full sample prior to excluding city grid cells. Distances are measured in kilometres.

B. Sentiment Analysis Procedure

B.1. Step 1: Construction of the Sentence-Level Corpus

This step constructs a clean, uniform sentence-level corpus from all 29 exploration diaries listed in Table 2, which serves as the input for the subsequent relationship and sentiment classification tasks. Every diary is processed through the same pipeline so that the classifiers see consistent input across explorers.

The resulting corpus contains 85,324 sentences across the 29 diaries.

Step 1.1 Source text and scope

The source material consists of the digitised diaries obtained from Project Gutenberg Australia and, where necessary, from additional online repositories and archival materials. For diaries provided as HTML we parse the markup directly; for diaries available only as scanned images we apply optical character recognition (OCR). We focus only on the narrative portions of each diary, excluding all non-narrative material (such as navigation menus, copyright notices, and appendices).

Step 1.2 Extraction of narrative text

Using HTML parsing (or OCR for scanned sources), we extract only narrative content contained in paragraph and blockquote elements. These elements correspond to the diary's prose while excluding menus, headings, and metadata introduced during digitisation. The narrative text of each diary is then assembled into a single document for uniform preprocessing.

Step 1.3 Removal of non-narrative and boilerplate content

Digitised historical texts often contain material that is not part of the original narrative. To avoid contaminating the dataset, we explicitly remove:

- Project Gutenberg boilerplate and footer text
- Navigational lists and indices
- Poetic or reflective appendices that follow the diary narrative

For instance, in Ernest Giles's *Australia Twice Traversed*, all text appearing after the diary's concluding reflective section ("As a last remark... AUSTRALIA") is excluded. This ensures that the corpus contains only narrative prose produced during the expeditions.

Step 1.4 Text normalisation

The extracted text is normalised to ensure consistency and robustness:

- Line breaks and excess whitespace are removed
- Hyphenated words split across lines are rejoined
- Typographic quotation marks and apostrophes are standardised
- Em-dashes (—) are preserved, as they are frequently used in nineteenth-century prose to convey emphasis and narrative flow

Step 1.5 Sentence construction

Sentence construction is deliberately conservative, reflecting the complexity of nineteenth-century writing.

First, common abbreviations (e.g. Mr., Dr., a.m., p.m.) are protected so that periods within abbreviations do not trigger false sentence breaks.

Second, sentences are split using punctuation markers (periods, question marks, and exclamation marks). To correct digitisation artefacts, missing spaces after punctuation are inserted before splitting.

Third, fragments that begin with lowercase letters immediately after punctuation (e.g. "? an old discharged copper...") are treated as continuations of the preceding sentence rather than independent sentences.

Sentence splitting is performed paragraph-by-paragraph to ensure stable sentence construction.

Step 1.6 Filtering and de-duplication

To ensure that each observation corresponds to a meaningful narrative unit, we apply the following filters:

- Sentences shorter than six words are discarded
- Extremely long strings that reflect indices or digitisation artefacts are removed
- Duplicate sentences are dropped so that each sentence appears only once

Step 1.7 Final dataset

After preprocessing, the corpus contains 85,324 sentences across the 29 diaries. The final dataset consists of one row per sentence, with an explorer identifier, a sentence index, and the sentence text. This dataset serves as the input for the subsequent relationship classification and sentiment analysis.

B.2. Step 2: Identification of Relationship Sentences

In this step, we identify which sentences describe a relationship between Europeans and Aboriginal people. The aim is to distinguish sentences that express interaction, perception, judgement, or evaluative stance between the two groups from sentences that are purely descriptive or unrelated. This classification stage serves as a filtering step that separates relational content from background narrative and provides the foundation for the subsequent sentiment analysis.

Step 2.1 Input data

The procedure takes as input the sentence-level corpus constructed in Step 1, consisting of 85,324 sentences across the 29 diaries. Each sentence is associated with an explorer identifier and a unique index.

Step 2.2 Hand-labelled training data

Rather than relying on any automated or external-model annotation, we construct the training data entirely by hand. We manually label 1,599 sentences drawn from four diaries chosen to span the continent and the century—John Oxley (1817), Ludwig Leichhardt (1844–45), Ernest Giles (1872–76), and David Carnegie (1896)—of which 231 are labelled RELATIONSHIP.

Step 2.3 Classification task

Each sentence is classified into one of two categories:

- RELATIONSHIP
- NOT RELATIONSHIP

The definition of relationship is deliberately broad and conceptually aligned with historical analysis rather than requiring explicit interaction.

A sentence is labelled **RELATIONSHIP** if it reveals any relational dimension between Europeans and Aboriginal people, including:

1. Direct interactions or shared events, such as conflict, pursuit, violence, assistance, guidance, exchange of goods, communication, punishment, or arrest.
2. Perceptions, judgements, or evaluations, including European opinions, stereotypes, fears, hostility, admiration, moral reflections, or inferred Aboriginal attitudes toward Europeans.
3. Implicit references, where Aboriginal people are identifiable through pronouns or narrative context (e.g. “they,” “the tribe,” or similar expressions), even when not explicitly named.

Sentences are labelled **NOT RELATIONSHIP** only if they contain no explicit or implicit reference to Aboriginal people and no evaluative or relational content.

Step 2.4 Labelling guidelines

The hand labels follow a conceptual rather than mechanical definition of relationship, capturing attitudes, impressions, fear, prejudice, admiration, and tension. When uncertainty arises, the coder errs toward inclusion (RELATIONSHIP), reflecting the objective of preserving potentially meaningful social signals for the subsequent sentiment analysis.

Step 2.5 Training the classifier

Using the hand-labelled dataset, we fine-tune a supervised classification model based on BERT to distinguish between relational and non-relational sentences.

Step 2.6 Model validation and performance

We assess out-of-sample performance using leave-one-diary-out cross-validation, training on three of the four labelled diaries and testing on the unseen fourth, which yields a mean F1-score of 0.725. The final model, trained on all four diaries' hand labels, achieves an accuracy of 92.5%, an F1-score of 0.735, a precision of 0.758, and a recall of 0.714 on a held-out test set.

Step 2.7 Application to all diaries

The trained classifier is then applied to all 85,324 sentences in the corpus, identifying 12,319 relationship sentences across the 29 diaries. Only these relationship sentences are retained for the sentiment analysis.

B.3. Step 3: Sentiment Classification

After identifying sentences that contain relational content, the next step is to classify the sentiment expressed in these interactions.

Step 3.1 Filtering relational sentences

For each diary we retain only sentences classified as **RELATIONSHIP** by the fine-tuned BERT classifier. Sentences labelled **NOT RELATIONSHIP** are excluded from the subsequent sentiment analysis.

Step 3.2 Construction of the sentiment training sample

A sample of relational sentences from the four training diaries is annotated by hand into positive, neutral, or negative categories; no automated or external-model annotation is used. The sample deliberately over-represents conflict-rich passages to ensure that the negative class is adequately represented for training. The resulting training dataset contains 701 hand-labelled sentences: 492 neutral, 136 negative, and 73 positive.

Step 3.3 Training the sentiment classifier

We fine-tune a RoBERTa-large model on the hand-labelled sentiment dataset to perform sentence-level sentiment classification.

Step 3.4 Model validation and performance

Out-of-sample performance is assessed using leave-one-diary-out cross-validation, yielding a macro-F1 of 0.583; on a stratified held-out test set the model attains a macro-F1 of 0.706 and an accuracy of 0.802, with a negative-class recall of 0.67.

Step 3.5 Sentiment classification for all diaries

The trained RoBERTa classifier is applied to all sentences classified as **RELATIONSHIP**. For each sentence, the model assigns a sentiment label indicating whether the interaction described is positive, neutral, or negative. These sentence-level classifications form the basis for constructing the sentiment measures used in the empirical analysis.

C. Robustness for European Exploration Routes

C.1. Trading-Cultural Sites

Table C.1. Trading-Cultural Sites and European Explorations

Dependent Variable: European Exploration Route (dummy)							
	(1) All cells	(2) All cells	(3) All cells	(4) Excl. 100km	(5) Excl. 200km	(6) Excl. 300km	(7) Excl. 500km
Trading-Cultural Sites	0.0084***	0.0085***	0.0081***	0.0091***	0.0092***	0.0093***	0.0090**
	(0.0017)	(0.0017)	(0.0017)	(0.0020)	(0.0022)	(0.0024)	(0.0037)
HMI Paths	0.0051	0.0031	0.0023	0.0004	0.0032	0.0047	−0.0085**
	(0.0034)	(0.0032)	(0.0033)	(0.0033)	(0.0037)	(0.0041)	(0.0041)
Geo-Climatic Controls		✓	✓	✓	✓	✓	✓
Accessibility Controls			✓	✓	✓	✓	✓
Historical Controls			✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓	✓
N	79,722	79,722	79,722	64,815	51,845	40,638	22,298
R^2	0.007	0.012	0.015	0.015	0.014	0.017	0.030

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable is a dummy variable indicating whether a European exploration route passes through a grid cell. *Trading-Cultural Sites* is a standardised dummy equal to one if the grid cell contains at least one McCarthy (1939)'s trading-cultural site. HMI Paths is the standardised terrain-only counterfactual path variable. Column (1) includes only the treatment variables and state fixed effects. Column (2) additionally controls for geo-climatic characteristics. Column (3) further accounts for accessibility and historical controls. Columns (4) through (7) progressively exclude grid cells within 100km, 200km, 300km, and 500km of the coastline respectively. All specifications include state fixed effects. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Table C.2. Excluding Trading-Cultural Sites

Dependent Variable: European Exploration Routes (dummy)			
	(1) Control for Trading-Cultural Sites	(2) Exclude Trading-Cultural Sites	(3) Exclude 50km Buffer
Songlines	0.0110***	0.0109***	0.0110***
	(0.0033)	(0.0033)	(0.0035)
HMI Paths	0.0014	0.0013	−0.0034
	(0.0032)	(0.0032)	(0.0029)
Trading-Cultural Sites	0.0061***		
	(0.0017)		
Geo-Climatic Controls	✓	✓	✓
Accessibility Controls	✓	✓	✓
Historical Controls	✓	✓	✓
State FE	✓	✓	✓
N	79,722	79,408	64,141
R^2	0.016	0.015	0.013

Notes: The unit of observation is a grid cell of 10km × 10km. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable is whether the cell was crossed by a European exploration route. Column (1) controls for trading-cultural sites by including them as an additional regressor. Column (2) excludes grid cells that coincide with trading-cultural sites. Column (3) excludes all grid cells within a 50km buffer around each trading-cultural site. Trading-cultural sites are defined as origins, waypoints, and destinations of the *Songlines*, following McCarthy (1939). *Songlines*, HMI Paths, and trading-cultural sites are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.2. Excluding Arid Places

This robustness check addresses the heterogeneity that characterises the Australian landscape, examining whether the long-term effects of *Songlines* on exploration activity are driven by the inclusion of arid desert regions. We exclude grid cells below the 25th percentile of precipitation in column (1) and above the 75th percentile of temperature in column (2). Table C.3 shows that *Songlines* remains positive and statistically significant in both columns, confirming that the results are not sensitive to the inclusion of arid or climatically extreme regions.

C.3. Excluding Coastal Areas up to 500km from the Shore

This robustness check examines the role of geographic heterogeneity across the Australian landscape, particularly between coastal and inland regions. Table C.4 reports results for exploration routes while progressively excluding coastal areas up to 500km from the shoreline. Column (1)

Table C.3. Excluding Arid Places

	Dependent Variable: Exploration Routes	
	(1) Above 25th percentile of Precipitation	(2) Below 75th percentile of Temperature
Songlines	0.0116***	0.0138***
	(0.0031)	(0.0032)
HMI Paths	-0.0016	0.0002
	(0.0034)	(0.0033)
Geo-Climatic Controls	✓	✓
Accessibility Controls	✓	✓
Historical Controls	✓	✓
State FE	✓	✓
N	59,792	59,790
R^2	0.018	0.017

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable is a dummy variable indicating whether a European exploration route passes through a grid cell. Column (1) excludes grid cells below the 25th percentile of precipitation. Column (2) excludes grid cells above the 75th percentile of temperature. *Songlines* and HMI Paths are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

uses the full sample of grid cells. Columns (2) through (6) sequentially exclude cells within 50km, 100km, 200km, 300km, and 500km of the coast respectively. *Songlines* remains positive and statistically significant across all columns, confirming that the effects of *Songlines* on exploration activity persist well into the interior of Australia and are not driven by coastal areas.

C.4. Excluding States

Australia is a vast country with large environmental differences across its states. Western Australia is dominated by sparsely populated deserts, whereas New South Wales is far more urbanised and includes major cities such as Sydney. To examine whether our findings are driven by a particular region, we exclude each state individually from the analysis. Table C.5 shows that *Songlines* remains positive and statistically significant at the 1% level in all specifications, indicating that the results are not driven by any single state and remain robust across Australia's diverse environments.

Table C.4. Excluding Coastal Areas

Exploration Routes						
	(1) All cells	(2) Excl. 50km	(3) Excl. 100km	(4) Excl. 200km	(5) Excl. 300km	(6) Excl. 500km
Songlines	0.0121***	0.0101***	0.0114***	0.0130***	0.0157***	0.0120*
	(0.0032)	(0.0032)	(0.0035)	(0.0039)	(0.0044)	(0.0065)
HMI Paths	0.0019	0.0008	0.0000	0.0026	0.0040	-0.0088**
	(0.0031)	(0.0032)	(0.0031)	(0.0035)	(0.0039)	(0.0039)
Geo-Climatic Controls	✓	✓	✓	✓	✓	✓
Accessibility Controls	✓	✓	✓	✓	✓	✓
Historical Controls	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
N	79,722	72,152	64,815	51,845	40,638	22,298
R^2	0.015	0.014	0.015	0.015	0.018	0.031

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable is a dummy variable indicating whether a European exploration route passes through a grid cell. Column (1) uses the full sample. Columns (2) through (6) progressively exclude grid cells within 50km, 100km, 200km, 300km, and 500km of the coastline respectively. *Songlines* and HMI Paths are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Table C.5. Excluding Australian States

	Dependent Variable: Exploration Routes					
	(1) Excl. W. Australia	(2) Excl. N. Territory	(3) Excl. S. Australia	(4) Excl. Queensland	(5) Excl. Victoria	(6) Excl. New S. Wales
Songlines	0.0106***	0.0116***	0.0108***	0.0181***	0.0120***	0.0105***
	(0.0034)	(0.0035)	(0.0033)	(0.0038)	(0.0033)	(0.0035)
HMI Paths	0.0024	0.0020	-0.0007	0.0030	0.0016	0.0028
	(0.0038)	(0.0034)	(0.0032)	(0.0034)	(0.0032)	(0.0033)
Geo-Climatic Controls	✓	✓	✓	✓	✓	✓
Accessibility Controls	✓	✓	✓	✓	✓	✓
Historical Controls	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
N	50,769	65,954	69,813	62,752	77,542	71,780
R^2	0.014	0.021	0.012	0.019	0.015	0.017

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with robust standard errors clustered at the local government district level. The dependent variable is a dummy variable indicating whether a European exploration route passes through a grid cell. Each column excludes grid cells associated with one Australian state: Western Australia in column (1), Northern Territory in column (2), South Australia in column (3), Queensland in column (4), Victoria in column (5), and New South Wales and the Australian Capital Territory in column (6). *Songlines* and HMI Paths are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

C.5. Spatial Autocorrelation

We assess the potential impact of spatial autocorrelation on our results by computing Conley spatial-HAC standard errors (Conley 1999), implemented via the `acreg` command of Colella et al. (2019), with a distance cutoff of 50 km. This approach allows for arbitrary spatial correlation among residuals within the cutoff radius without imposing any grouping structure on the data. Table C.6 reports the results. The coefficient on *Songlines* remains positive and significant at the 1% level, with a Conley standard error of 0.0028, virtually identical to the baseline cluster-robust estimate. This confirms that our results are not driven by spatial autocorrelation.

Table C.6. Spatial Autocorrelation

	Dependent Variable: Exploration Routes
	(1)
Songlines	0.0121***
	(0.0028)
HMI Paths	0.0019
	(0.0028)
Geo-Climatic Controls	✓
Accessibility Controls	✓
Historical Controls	✓
State FE	✓
N	79,722
R^2	0.015

Notes: The unit of observation is a grid cell of $10km \times 10km$. OLS estimates are reported with Conley spatial-HAC standard errors (Conley 1999) computed via `acreg` (Colella et al. 2019) with a distance cutoff of 50 km. The dependent variable is a dummy variable indicating whether a European exploration route passes through a grid cell. *Songlines* and HMI Paths are standardised. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

C.6. Omitted Variables

In order to address omitted variable bias concerns, we apply the technique proposed by Oster (2019) to address the issue of omitted variable bias, which occurs when some relevant factors (unobserved controls) are left out of the model, potentially distorting the results. Oster (2019)'s approach improves our estimates by increasing the R^2 value, which is an indicator of how much

of the variation in the data is explained by the model, by 30% compared to our most conservative estimate ($R_{\max} = 1.3 \times R^2$). Increasing the R^2, Oster (2019)'s algorithm assumes that we employ all (observable and unobservable) variables as controls, allowing for the calculation of a new β coefficient and a parameter called δ. For the model to be free of omitted variable bias, the interval between our most conservative β (from our baseline estimate) and the new β coefficient should safely exclude zero, and the δ parameter must be greater than one. The parameter δ reflects how important the unobserved variables would need to be, compared to the observed ones, in order to significantly affect the model. If δ is greater than one, it suggests that the model is unlikely to be influenced by omitted variable bias. As shown in Table C.7, both these conditions are met: the new β coefficient is very close to the baseline estimate, and δ is greater than 5 ($\delta = 5.73$), which indicates that the model is highly unlikely to suffer from omitted variable bias.

Furthermore, we use the spatial first differences (SFD) design proposed by Druckenmiller and Hsiang (2018). This method accounts for the fact that adjacent units of analysis that are spatially closer to each other offer better counterfactuals, as unobservable confounders are not systematically correlated between them. To this end, we compare cell c to cell $c-1$ and assume that the expected potential outcome for each would be the same if they received the same treatment. We then distinguish each pair of adjacent cells in the west-east direction and regress the spatial first differences as follows:

$$\Delta ExplorationRoutes_i = \beta_{fsd} Songlines_i + \epsilon_i. \tag{4}$$

In equation (4), β_{fsd} captures the effect on exploration routes when a random grid cell c is treated while grid cell $c-1$ is not (or vice versa). Column (1) of Table C.8 calculates the differences in the west-east direction. We observe a positive and statistically significant effect of *Songlines* on exploration activity. This result lends confidence that spatially correlated unobserved heterogeneity and omitted variables are not biasing our findings.[22]

[22] Our results also remain highly significant when bootstrapped standard errors are used (Roodman et al. 2019).

Table C.7. Omitted Variable Bias (Oster's Test)

	Dependent Variable: Exploration Routes			
	(1) Controlled coefficient $\tilde{b}$	(2) Identified Set $(\dot{b}(R_{max}, \delta = 1), \tilde{b})$	(3) Excludes Zero	(4) Absolute Delta (δ)
Songlines	0.0121	[0.0105, 0.0121]	✓	5.73 > 1
Geo-Climatic Controls	✓	✓		
Accessibility Controls	✓	✓		
Historical Controls	✓	✓		
State FE	✓	✓		
N	79,722	79,722		
R^2	0.015	0.015		

Notes: The unit of observation is a grid cell of $10km \times 10km$. The dependent variable is a dummy variable indicating whether a European exploration route passes through a grid cell. Column (1) reports the controlled coefficient from the baseline estimate. Column (2) shows the identified set $(\dot{b}(R_{max}, \delta = 1), \tilde{b})$ obtained following Oster (2019), where R_{max} is set to 1.3 times the R^2 of the controlled regression. Column (3) indicates with ✓ that the identified set safely excludes zero, ruling out omitted variable bias. Column (4) reports the value of δ, which measures how much stronger selection on unobservables relative to observables would need to be to explain away the result. A value of $\delta > 1$ indicates that omitted variable bias is unlikely to drive the results. All specifications include state fixed effects, geo-climatic, accessibility, and historical controls. The constant term is omitted for space.

Table C.8. Omitted Variable Bias (Spatial First Differences)

	Dependent Variable: Exploration Routes
	(1) West–East direction
Songlines (Differences)	0.0054***
	(0.0011)
N	79,101
R^2	0.000

Notes: The unit of observation is a grid cell of 10km × 10km. The dependent variable is a dummy variable that takes the value of 1 if cell i has a European exploration route, and 0 otherwise. *Songlines* is standardised here, as in the main specifications, so the coefficient gives the change in exploration probability associated with a one standard deviation increase in the differenced *Songlines* measure between adjacent grid cells. The estimate calculates spatial first differences (SFD) in the west–east direction as proposed by Druckenmiller and Hsiang (2018). Robust standard errors are reported in parentheses and the constant term was omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

C.7. Pre-colonial Settlement Intensity

A natural concern is that our results simply reflect where more Aboriginal people lived rather than the *Songlines* network itself. If European explorers were attracted to densely settled areas because of water, food, local guides, or existing pathways, then the estimated *Songlines* effect could instead capture pre-colonial population density. We address this concern by controlling for an entirely independent measure of pre-colonial Aboriginal settlement.

Our settlement measure comes from Williams et al. (2014) (AustArch), a continent-wide compilation of radiocarbon- and luminescence-dated archaeological sites assembled independently of our study. Unlike our Indigenous heritage dataset, it records excavated settlement sites such as shell middens, hearths, burials, open camps, and rockshelters rather than Indigenous heritage places. For each 10km $\times$ 10km cell, we count the number of distinct dated sites and standardise the measure, which we label *Pre-colonial Settlement*.

Panel A of Table C.9 validates our Indigenous heritage intensity surface against this independent archaeological record.[23] A one standard deviation increase in archaeological site density is associated with approximately a 0.05 standard deviation increase in the Indigenous heritage intensity surface, both with only state fixed effects (column 1) and after adding the full set of geographic, climatic, accessibility, and historical controls (column 2). The close correspondence between these two independently constructed datasets provides external validation that our Indigenous heritage intensity surface reflects genuine features of the pre-colonial Aboriginal landscape rather than artifacts of any single data source.

Panel B then asks whether the predictive power of *Songlines* survives once we control for *Pre-colonial Settlement*. Column (1) reproduces our benchmark specification, while column (2) adds *Pre-colonial Settlement*. Although settlement enters positively and significantly, the coefficient on *Songlines* remains virtually unchanged, declining only from 0.0121 to 0.0120. This suggests that

[23] The sample in Panel A is 79,372 rather than 79,722 because the cultural intensity surface is not defined on the 350 cells covered by the largest inland salt lakes, such as Lake Eyre and Lake Torrens, which were treated as impassable when the songlines were reconstructed. These cells therefore drop out of regressions using the surface, whereas Panel B retains the full sample.

Songlines capture dimensions of movement, connectivity, and landscape knowledge that cannot be explained by pre-colonial settlement alone.

A remaining concern is that archaeological sites may be more readily preserved and documented in some environments than others, making *Pre-colonial Settlement* a noisier proxy in those settings. Columns (3) to (5) therefore progressively exclude cells in the driest quartile of the precipitation distribution, the hottest quartile of the temperature distribution, and finally both simultaneously. Across all specifications, the coefficient on *Songlines* remains positive and statistically significant at the one percent level and, if anything, becomes slightly larger despite the substantial reduction in sample size. Together, the results in Table C.9 show that the predictive power of *Songlines* cannot be explained by pre-colonial settlement alone and remains robust where archaeological evidence is most reliably preserved and documented.

C.7.1. Pre-colonial Aboriginal Institutions

Although our specifications already control for pre-colonial settlement intensity, we take a further step to account for unobserved institutional and societal heterogeneity across Aboriginal groups along the lines of Michalopoulos and Papaioannou (2013) and Angeles and Elizalde (2017). To do so, we draw on the linguistic map compiled by O'Grady et al. (1966), who documented the boundaries of approximately 450 Aboriginal spoken languages across Australia prior to European colonisation. Despite inevitable imprecision in the delineated boundaries, this map remains the most comprehensive available source on pre-colonial societal structures in Australia. We assign each grid cell to the language group whose territory contains its centroid and construct a full set of language group fixed effects, absorbing all time-invariant differences in social organisation, cultural practices, and institutional arrangements across Aboriginal groups. Column (6) of Panel B in Table C.9 present the results.[24] The coefficient on *Songlines* remains positive and highly significant, closely matching the estimates in the preceding columns and indicating that the relationship between *Songlines* and

[24] Three language groups (Gambre, Raminjeri, and Riraidjango) contain a single grid cell in our sample. These singleton observations are automatically dropped by the fixed-effects estimator, as the group fixed effect perfectly absorbs the outcome and leaves no residual variation for estimation. This accounts for the minor difference in sample size relative to the preceding columns.

European exploration is not driven by unobserved heterogeneity across pre-colonial Aboriginal societies.

C.8. Contiguous Analysis

In this section, we turn to a contiguous-neighbour design that compares *Songlines* cells directly against the cells sitting right next to them. For each cell on the 10 km × 10 km grid, we define its neighbours as the up-to-eight surrounding cells, those sharing an edge or a corner. For each *Songlines* cell, we identify all neighbouring cells that are not themselves crossed by a *Songline*, forming treated and control pairs that share virtually identical geographic and climatic conditions. We implement two variants of this design. The first follows the contiguous-pair approach of Dube et al. (2010), in which each *Songlines* cell is paired separately with each of its non-*Songlines* neighbours, yielding 36,556 unique pairs. Each pair receives its own fixed effect, so identification comes exclusively from the outcome difference between two cells that sit side by side. Because a *Songlines* cell with multiple valid neighbours enters multiple pairs, the same treated observation is intentionally replicated, giving equal weight to every border segment, and the dataset contains $36{,}556 \times 2 = 73{,}112$ observations. We two-way cluster standard errors on the *Songlines* cell and the control cell to account for the dependence this replication introduces.

The second variant groups each *Songlines* cell together with all of its non-*Songlines* neighbours into a single neighbourhood, producing 6,881 groups. Here the treated cell appears only once within its group, alongside a variable number of control neighbours, so the dataset comprises 43,437 observations, fewer than in the pair design because the treated cell is not replicated across separate pairings. A neighbourhood fixed effect absorbs any feature common to the group, and the *Songlines* coefficient is identified from within-group variation between the treated cell and its surrounding controls. Standard errors are clustered at the neighbourhood level. Both specifications include the full set of geo-climatic and accessibility controls as well as the HMI least-cost-path indicator. The results, reported in Table C.10, confirm that the *Songlines* effect survives this demanding local comparison. The coefficient on the *Songlines* indicator remains positive and statistically significant

Table C.9. Pre-colonial Settlement, Cultural Sites, and European Exploration

Panel A. Dependent Variable: Indigenous Heritage Intensity Surface

	(1)	(2)
Pre-colonial Settlement	0.0475***	0.0263***
	(0.0059)	(0.0044)
Geo-Climatic Controls		✓
Accessibility Controls		✓
Historical Controls		✓
State FE	✓	✓
N	79,372	79,372
R^2	0.189	0.571

Panel B. Dependent Variable: European Exploration Routes (dummy)

	(1)	(2)	(3)	(4)	(5)	(6)
Songlines	0.0121***	0.0120***	0.0114***	0.0136***	0.0133***	0.0128***
	(0.0032)	(0.0032)	(0.0031)	(0.0032)	(0.0030)	(0.0026)
HMI Paths	0.0019	0.0018	-0.0017	0.0001	-0.0056	0.0010
	(0.0031)	(0.0031)	(0.0034)	(0.0033)	(0.0034)	(0.0023)
Pre-colonial Settlement		0.0053***	0.0052***	0.0065***	0.0064***	0.0047***
		(0.0018)	(0.0019)	(0.0019)	(0.0020)	(0.0016)
Geo-Climatic Controls	✓	✓	✓	✓	✓	✓
Accessibility Controls	✓	✓	✓	✓	✓	✓
Historical Controls	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
Language Group FE						✓
Sample	All	All	Excl. arid	Excl. hot	Excl. hot & arid	All
N	79,722	79,722	59,792	59,790	40,218	79,719
R^2	0.015	0.016	0.018	0.017	0.022	0.092

Notes: The unit of observation is a grid cell of 10km × 10km. OLS estimates are reported with robust standard errors clustered at the local government district level (in parentheses). *Panel A* asks whether our cultural record coincides with an independent archaeological measure of pre-colonial Aboriginal presence: the dependent variable is the *Indigenous Heritage Intensity Surface* (ranging from 0 to 1) (hub-clustered and state-weighted to mitigate recording-effort bias), and *Pre-colonial Settlement* is the standardised count of radiocarbon-dated archaeological sites (Williams et al. 2014) in the cell. *Panel B* asks whether the effect of *Songlines* on European exploration survives controlling for this settlement intensity: the dependent variable is a dummy for whether the cell contains a European exploration route. *Songlines*, HMI Paths and *Pre-colonial Settlement* are standardised. Geo-climatic controls are latitude, longitude, agricultural suitability, elevation, ruggedness, mean and standard deviation of precipitation and temperature, and water cover; accessibility controls are distance to the coast, to Sydney and to the nearest capital; historical controls are distance to the nearest historical mine and a secondary-coast indicator. Columns (3) to (5) of Panel B restrict the sample to easier-to-survey areas by excluding the driest quartile of precipitation ("arid"), the hottest quartile of temperature ("hot"), or both, so the result is not an artefact of where archaeology can be recovered. Column (6) of Panel B adds fixed effects for 436 Aboriginal language group homelands, absorbing all time-invariant unobserved heterogeneity within each ethnolinguistic territory. The constant term is omitted for space. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

at the 1% across all six outcomes, namely European exploration routes, cities, towns, railways, night-light luminosity, and population density.

Table C.10. Contiguous-Neighbour Fixed-Effects Estimates

	Exploration (1)	Cities (2)	Towns (3)	Railways (4)	Luminosity (5)	Population (6)
Panel A: Contiguous-Pair Fixed Effects						
Songlines	0.0216***	0.0045***	0.0111***	0.0118***	0.0276***	2.1374***
	(0.0043)	(0.0011)	(0.0021)	(0.0029)	(0.0053)	(0.3850)
HMI Paths	0.0080	0.0024*	0.0064**	0.0018	0.0217**	1.1659*
	(0.0068)	(0.0014)	(0.0032)	(0.0040)	(0.0100)	(0.6943)
Controls	✓	✓	✓	✓	✓	✓
Pair FE	✓	✓	✓	✓	✓	✓
N	73,112	73,112	73,112	73,112	73,112	73,112
R^2	0.715	0.506	0.547	0.652	0.812	0.835
Panel B: Neighbourhood Fixed Effects						
Songlines	0.0221***	0.0052***	0.0116***	0.0122***	0.0345***	2.4611***
	(0.0038)	(0.0010)	(0.0019)	(0.0026)	(0.0059)	(0.4029)
HMI Paths	0.0010	0.0006	0.0030*	0.0055**	0.0076	0.2024
	(0.0049)	(0.0007)	(0.0017)	(0.0026)	(0.0072)	(0.5104)
Controls	✓	✓	✓	✓	✓	✓
Neighbourhood FE	✓	✓	✓	✓	✓	✓
N	43,437	43,437	43,437	43,437	43,437	43,437
R^2	0.461	0.169	0.229	0.384	0.655	0.669

Notes: The unit of observation is a 10km × 10km grid cell. Queen contiguity defines the up-to-8 Moore neighbours (4 edge + 4 corner adjacent). *Panel A* implements a contiguous-pair design: each songline cell is paired with each non-songline queen neighbour (36,556 pairs, 73,112 rows). Standard errors are two-way clustered on the songline pixel and the control pixel. *Panel B* groups each songline cell with all its non-songline queen neighbours (6,881 groups, 43,437 rows); standard errors are clustered at the neighbourhood level. *Songlines* equals one if the cell is crossed by a reconstructed songline. HMI Paths equals one for cells on an HMI least-cost path. Controls include geo-climatic characteristics (agricultural suitability, elevation, ruggedness, precipitation, temperature, water), accessibility (distance to coast, coast layer, distance to Sydney, nearest capital, nearest historical mine), and coordinates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D. Robustness for long term effects

D.1. Robustness Checks

In this section, we replicate all robustness checks presented for exploration routes as the outcome variable and additionally introduce further robustness exercises related to long-run economic activity, measured through luminosity and population density. Table D.1 reports results across a range of sample restrictions and specification checks. Panel A excludes arid regions by dropping grid cells below the 25th percentile of precipitation, above the 75th percentile of temperature, or

both. Panel B progressively excludes grid cells within 50, 100, 200, 300, and 500 km of the coastline, while Panel C repeats this exercise using population density as the dependent variable. Panel D drops one state at a time. In all cases, the *Songlines* coefficient remains positive and statistically significant, with the point estimate often increasing as the sample is restricted to the more remote interior. Panel E reports Conley spatial-HAC standard errors with a 50 km distance cutoff; the *Songlines* coefficient remains significant at the 5% level despite the substantially larger standard errors. Panel F applies the Oster (2019) test for omitted variable bias: the bias-adjusted coefficient β^* is virtually identical to the baseline estimate, and the proportional selection ratio $\delta = 23.3$ indicates that unobservables would need to be over 23 times more important than observables to explain away the result. Panel G reports spatial first differences using east–west neighbours within a row; the differenced *Songlines* coefficient is positive and highly significant. Panel H controls for pre-colonial settlement intensity using the standardised AustArch archaeological site count and progressively restricts the sample to easier-to-survey environments; column (6) further absorbs fixed effects for 436 Aboriginal language groups from the O'Grady et al. (1966) map, and the coefficient remains significant at the 1% level. Panel I replaces *Songlines* with *Trading-Cultural Sites* and progressively excludes coastal areas. Finally, Panel J tests whether the *Songlines* effect is driven by trading-cultural sites as origin and destination points McCarthy (1939): the coefficient remains virtually unchanged when controlling for trading-cultural sites, excluding trading-site cells, or excluding all cells within a 50km buffer.

E. Construction of the Songline Variable

This appendix documents the precise procedure, formulas, inputs, and parameters used to construct the songline variable. All geometry is in EPSG:7855 (GDA2020 / MGA Zone 55), in metres.

Table D.1. Robustness Checks: Luminosity

Dependent Variable: Night-Light Luminosity

Panel A. Excluding Arid and Hot Areas

	(1) All cells	(2) Excl. arid	(3) Excl. hot	(4) Excl. hot & arid
Songlines	0.0202***	0.0259***	0.0257***	0.0364***
	(0.0047)	(0.0056)	(0.0064)	(0.0080)
HMI Paths	0.0076***	0.0104***	0.0118***	0.0189***
	(0.0025)	(0.0033)	(0.0036)	(0.0051)
N	79,722	59,792	59,790	40,218
R^2	0.035	0.038	0.043	0.047

Panel B. Excluding Coastal Areas

	(1) All cells	(2) >50km	(3) >100km	(4) >200km	(5) >300km	(6) >500km
Songlines	0.0202***	0.0047***	0.0040***	0.0024**	0.0017*	0.0008
	(0.0047)	(0.0014)	(0.0013)	(0.0010)	(0.0009)	(0.0007)
HMI Paths	0.0076***	0.0011	0.0007	0.0013**	0.0013**	0.0009**
	(0.0025)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0004)
N	79,722	72,152	64,815	51,845	40,638	22,298
R^2	0.035	0.089	0.104	0.007	0.004	0.005

Panel C. Excluding Coastal Areas: Population Density

	(1) All cells	(2) >50km	(3) >100km	(4) >200km	(5) >300km	(6) >500km
Songlines	1.5230***	0.3461***	0.2975***	0.2057***	0.1887***	0.1960**
	(0.3138)	(0.0731)	(0.0676)	(0.0543)	(0.0593)	(0.0775)
HMI Paths	0.5651***	0.0894**	0.0532	0.0596*	0.0755*	0.0517
	(0.1713)	(0.0426)	(0.0382)	(0.0334)	(0.0396)	(0.0348)
N	79,722	72,152	64,815	51,845	40,638	22,298
R^2	0.045	0.100	0.125	0.013	0.007	0.008

Panel D. Excluding States

	(1) Excl. WA	(2) Excl. NT	(3) Excl. SA	(4) Excl. QLD	(5) Excl. VIC	(6) Excl. NSW/ACT
Songlines	0.0164***	0.0238***	0.0239***	0.0256***	0.0165***	0.0170***
	(0.0041)	(0.0057)	(0.0053)	(0.0071)	(0.0042)	(0.0046)
HMI Paths	0.0054***	0.0098***	0.0105***	0.0086**	0.0050**	0.0074***
	(0.0021)	(0.0032)	(0.0031)	(0.0034)	(0.0023)	(0.0027)
N	50,769	65,954	69,813	62,752	77,542	71,780
R^2	0.046	0.038	0.039	0.037	0.034	0.026

Notes: The dependent variable is VIIRS night-light luminosity (Panels A, B, D) or GHSL population density (Panel C). All specifications include the full set of geo-climatic, accessibility, and historical controls plus state fixed effects, with standard errors clustered at the local government district level (in parentheses). *Panel A* restricts the sample by excluding the driest precipitation quartile ("arid"), the hottest temperature quartile ("hot"), or both. *Panels B–C* progressively exclude grid cells within 50, 100, 200, 300, and 500 km of the coastline. *Panel D* drops one state at a time. All variables are standardised. Continued on next page.

Table D.1. Robustness Checks: Luminosity (continued)

Dependent Variable: Night-Light Luminosity

Panel E. Conley Spatial-HAC Standard Errors (50 km cutoff)

	(1)
Songlines	0.0202**
	(0.0081)
HMI Paths	0.0076*
	(0.0045)
N	79,722
R^2	0.035

Panel F. Omitted Variable Bias (Oster Test)

$R_{\max} = 1.3 \times R^2$

β^* (bias-adjusted)	0.0198
δ	23.264

Panel G. Spatial First Differences

	(1)
Δ*Songlines*	0.0068***
	(0.0009)
ΔHMI Paths	0.0007
	(0.0011)
N	79,101
R^2	0.001

Panel H. Controlling for Pre-colonial Settlement Intensity

	(1) Baseline	(2) + Settl.	(3) Excl. arid	(4) Excl. hot	(5) Excl. hot & arid	(6) + Language FE
Songlines	0.0202***	0.0196***	0.0253***	0.0250***	0.0355***	0.0155***
	(0.0047)	(0.0046)	(0.0055)	(0.0063)	(0.0078)	(0.0035)
HMI Paths	0.0076***	0.0073***	0.0100***	0.0114***	0.0183***	0.0043**
	(0.0025)	(0.0025)	(0.0033)	(0.0036)	(0.0051)	(0.0019)
Pre-colonial Settlement		0.0187***	0.0193***	0.0226***	0.0233**	0.0135**
		(0.0069)	(0.0073)	(0.0086)	(0.0090)	(0.0057)
Language Group FE						✓
N	79,722	79,722	59,792	59,790	40,218	79,719
R^2	0.035	0.037	0.040	0.046	0.049	0.230

Notes (continued): Panel E reports Conley spatial-HAC standard errors with a 50 km distance cutoff. *Panel F* reports the Oster (2019) bias-adjusted treatment effect β^* and the proportional selection ratio δ under $R_{\max} = 1.3 \times R^2$; $\delta \gg 1$ indicates that unobservables would need to be over 23 times more important than observables to explain away the result. *Panel G* reports spatial first differences (east–west neighbours within a row). *Panel H* controls for pre-colonial settlement intensity (standardised AustArch site count); columns (3)–(5) further restrict to easier-to-survey environments; column (6) adds fixed effects for 436 Aboriginal language groups. Continued on next page.

Table D.1. Robustness Checks: Luminosity (continued)

Dependent Variable: Night-Light Luminosity

Panel I. Trading-Cultural Sites

	(1) State FE only	(2) All controls	(3) >100km	(4) >200km	(5) >300km	(6) >500km
Trading-Cultural Sites	0.0388***	0.0371***	0.0042**	0.0047**	0.0044*	0.0030
	(0.0119)	(0.0115)	(0.0019)	(0.0022)	(0.0024)	(0.0028)
HMI Paths	0.0024	0.0060***	0.0008	0.0012**	0.0012*	0.0007**
	(0.0019)	(0.0023)	(0.0006)	(0.0005)	(0.0006)	(0.0003)
N	79,722	79,722	64,815	51,845	40,638	22,298
R^2	0.023	0.042	0.104	0.009	0.006	0.008

Panel J. Excluding Origins and Destinations (Trading-Cultural Sites)

	(1) Control for Trading-Cultural Sites	(2) Exclude Trading-Cultural Sites	(3) Exclude 50km Buffer
Songlines	0.0137***	0.0134***	0.0041***
	(0.0035)	(0.0036)	(0.0013)
HMI Paths	0.0049**	0.0037*	−0.0006
	(0.0021)	(0.0019)	(0.0007)
Trading-Cultural Sites	0.0346***		
	(0.0113)		
N	79,722	79,408	64,141
R^2	0.043	0.033	0.272

Notes (continued): Panel I replaces *Songlines* with *Trading-Cultural Sites*, a standardised dummy equal to one if the cell contains at least one McCarthy (1939)'s trading-cultural site, and progressively excludes coastal areas. *Panel J* tests whether the *Songlines* effect survives after accounting for trading-cultural sites as origin and destination points: column (1) adds trading-cultural sites as a control, column (2) excludes trading-site cells, and column (3) excludes all cells within a 50km buffer around trading-cultural sites. All variables are standardised. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A. Inputs

- **Indigenous heritage sites.** The Database provides 12,047 recorded cultural locations across Australia: rock art sites (7,111), Creation/Dreaming narrative sites (1,967), quarries (1,020), ritual/ceremonial sites (990), grinding grooves (697), meeting sites/birthplaces (88), and other Aboriginal sites (174).
- ***Songline* endpoints.** McCarthy (1939)'s 581 documented trading-cultural sites serve as origins, waypoints, and destinations. Each connection between two consecutive trading-cultural sites defines one *Songline* segment (an origin–destination pair). For example, if three sites A, B, and C lie along a documented route, the segments A→B and B→C are treated as two separate *Songlines*, where B is the destination of the first and the starting point of the second. After dropping segments whose endpoints fall in the sea, snap to the same grid cell, or are otherwise unreachable, 697 *Songlines* are retained.
- **State borders.** GADM-derived state boundaries for hub-to-state assignment.
- **Inland water.** Large inland water bodies used as impassable barriers.
- **Grid.** A 10 km × 10 km fishnet covering mainland Australia, keyed by pixel identifier.

B. Grouping Indigenous heritage sites into hubs

Sites are grouped into hubs by single-linkage clustering at radius $r = 5$ km. The procedure works as follows: any two sites located within 5 km of each other are linked. If site A is within 5 km of site B, and site B is within 5 km of site C, then A, B, and C all belong to the same hub, even if A and C are more than 5 km apart. Formally, this is equivalent to identifying the connected components of a graph in which an edge connects any pair of sites separated by at most 5 km. Each hub's geographic location is then set to the average latitude and longitude of all its member sites.

This procedure groups the 12,047 recorded sites into 4,044 hubs. The majority of hubs (2,715, or 67%) consist of a single isolated site; 1,086 hubs contain 2–5 sites; 189 contain 6–20 sites; and 54 hubs contain more than 20 sites, with the largest hub comprising 1,522 sites (a densely recorded region in Western Australia). The median hub contains 1 site and the mean is approximately 3.

The rationale for collapsing sites into hubs and treating each hub as equally important is that the number of recorded sites in a given location reflects archaeological survey effort and government expenditure on heritage documentation at least as much as it reflects genuine cultural intensity. A cluster of 1,500 recorded sites in one region may indicate sustained government-funded surveying rather than a location that was 1,500 times more culturally significant than a single isolated site elsewhere. By merging nearby sites into a single hub and assigning equal initial weight to each hub (before the state-level correction in the next step), the construction avoids letting variation in recording intensity drive the reconstructed *Songline* network.

C. Hub weighting

The clustering step above prevents dense clusters of recorded sites at a single location from dominating the reconstruction. However, differences in documentation effort also exist across states: some states have recorded far more sites than others. To prevent these state-level differences from influencing the reconstruction, we introduce a second correction. Each hub h receives a weight that down-weights hubs in heavily recorded states:

$$w_h = \frac{1}{\sqrt{N_s}},$$

where N_s is the number of hubs in hub h's state. This corrects for uneven archaeological survey effort across jurisdictions. The correction is substantial: Western Australia accounts for 2,526 of the 4,044 hubs (62.5%), while New South Wales contributes 355 (8.8%), Queensland 349 (8.6%), Victoria 260 (6.4%), the Northern Territory 232 (5.7%), South Australia 184 (4.5%), and the ACT 17 (0.4%). The over-representation of Western Australia arises because the Government of Western Australia maintains and publicly releases a comprehensive Aboriginal Cultural Heritage Register, whereas other states do not provide equivalent open-access databases, requiring us to rely on federal heritage lists, academic publications, and web-based research for the remaining jurisdictions. However, this imbalance does not necessarily indicate that Western Australia is

culturally over-represented in absolute terms. Western Australia constitutes approximately one-third of Australia's land area and encompasses the vast majority of the continent's arid interior, where site discovery and documentation is considerably more difficult than in the more accessible eastern states. Thus, having more complete government data for a region that is inherently harder to survey may simply bring Western Australian coverage closer to the level achieved through scattered independent sources elsewhere. Regardless, the $1/\sqrt{N_s}$ weighting corrects for remaining density differences: a hub in Western Australia receives a weight of 0.020, whereas a hub in South Australia receives 0.074, ensuring that no single state's recording patterns dominate the reconstructed *Songline* network.

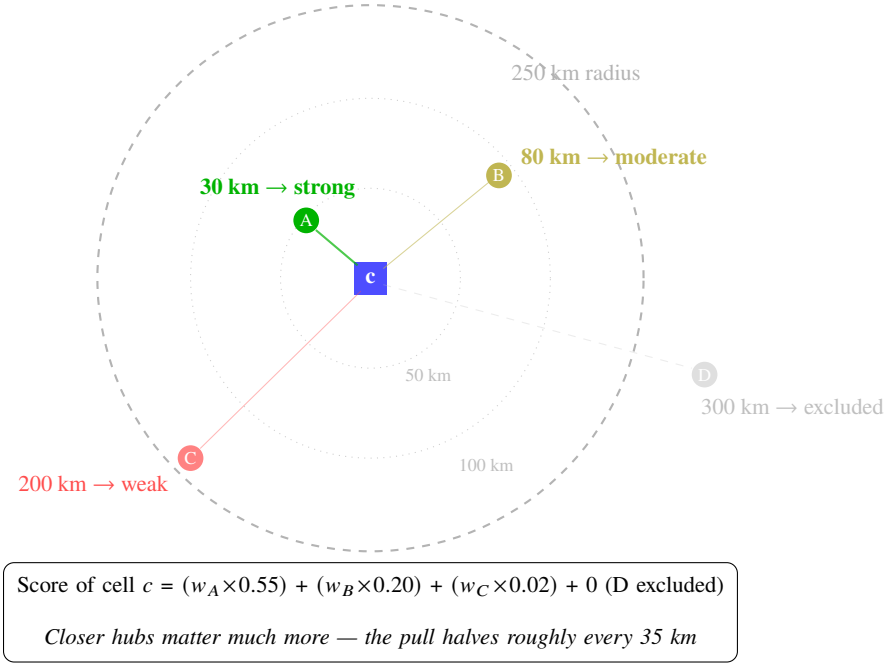
D. Cultural-influence surface

For each land-cell centroid c, cultural influence is the weighted sum of nearby hubs whose pull decays exponentially with distance:

$$\text{raw}(c) = \sum_{h:\, d(c,h) \le 5\lambda} w_h \cdot \exp\!\left(-\frac{d(c,h)}{\lambda}\right), \qquad \lambda = 50 \text{ km}, \text{ cutoff} = 5\lambda = 250 \text{ km},$$

where $d(c, h)$ is the Euclidean distance between cell c and hub h, and the sum runs over all hubs located within $5\lambda = 250$ km of c. Hubs beyond this radius contribute negligibly and are excluded. In simple terms, each hub acts like a magnet that pulls nearby cells toward it. The closer a hub is to a given cell, the stronger its pull. A hub 30 km away exerts a strong influence (the exponential equals approximately 0.55); a hub 80 km away contributes moderately (approximately 0.20); and a hub 200 km away barely registers (approximately 0.02). Beyond 250 km, the pull is effectively zero and the hub is ignored. For each cell, all contributions from hubs within range are added up: cells surrounded by many nearby hubs receive a high cultural score (they sit in culturally connected country), while cells far from any hub receive a score close to zero (they sit in culturally empty territory). The result is a smooth surface across Australia that is high near clusters of Indigenous heritage places and low in areas with no recorded cultural presence. Figure D.1 illustrates the logic.

Figure D.1. How hubs contribute to a cell's cultural influence score



E. Scaling to a 0–1 surface

The raw cultural-influence scores vary enormously: cells near dense hub clusters can score in the hundreds while remote cells score near zero. Without rescaling, these extreme differences would force the routing algorithm to use only the very highest-scoring cells, ignoring corridors with moderate cultural presence that still represent plausible *Songline* country. Rank-based normalisation solves this by sorting all cells from lowest to highest influence and assigning each cell a position on a uniform 0 to 1 scale. This preserves the ordering (a cell with more influence always ranks higher) while ensuring that the full range of culturally relevant territory can appear in reconstructed paths. The resulting surface does not depend on the specific value of the decay distance λ or on how hub weights are calibrated, so the reconstructed routes remain stable if those parameters are adjusted.

Raw influence is log-compressed and placed on a 0–1 ladder using rank scaling:

$$s(c) = \frac{\operatorname{rank}\big(\ln(1 + \operatorname{raw}(c))\big)}{n - 1}, \qquad s \in [0, 1],$$

where n is the number of land cells. $s = 1$ is the most culturally dense cell; $s = 0$ the least.

F. Void-friction penalty

A continuous penalty is added for country far from any hub:

$$V(c) = \min\Big(\frac{D_1(c)}{P_{95}(D_1)},\ 1\Big),$$

where $D_1(c)$ is the distance from cell c to the nearest hub, divided by the 95th percentile of D_1 across all land cells and capped at 1.

The cultural-influence surface in Section D already assigns low scores to cells far from hubs, and Section E preserves that ranking. However, two cells can both rank low while differing greatly in how isolated they are: one may sit 20 km from the nearest hub while another sits 400 km from anything. Without an additional penalty the routing algorithm would treat them almost identically. The void-friction term corrects this by measuring each cell's distance to its nearest hub and converting it to a 0–1 penalty. A cell close to at least one hub receives near-zero friction; a cell in deeply empty country receives high friction. This acts as a safety net that prevents reconstructed paths from crossing vast stretches where no archaeological record exists at all.

We normalise by the 95th percentile rather than a lower threshold (such as the 70th) so that the penalty reaches its maximum only for the most extreme 5% of cells. A lower percentile would penalise cells that are only moderately distant from hubs, effectively overriding the more detailed information already captured by the cultural-influence surface. The 95th percentile keeps the void penalty surgical: it intervenes strongly only where the evidence for cultural emptiness is overwhelming, and remains gentle everywhere else.

G. Travel-cost surface

Cultural influence is converted into a per-cell travel cost:

$$\text{cost}(c) = \varepsilon + (1 - s(c))^p + \theta \cdot V(c), \qquad \varepsilon = 0.01,\ p = 1,\ \theta = 0.5.$$

Culturally dense cells ($s \to 1$) cost ≈ 0.01 (cheap, preferred); empty cells ($s \to 0$) cost $\approx 1.01 + 0.5 \cdot V$. Lakes $\geq$ 1,000 km^2 are set to infinite cost—impassable barriers that routes must travel around.

H. Movement graph and routing

A lattice graph is constructed over land cells with 16-neighbour connectivity (8 king moves + 8 knight moves). The edge weight between adjacent cells a and b is:

$$\text{edge}(a, b) = \|\text{centre}_a - \text{centre}_b\| \cdot \tfrac{1}{2}\big(\text{cost}_a + \text{cost}_b\big),$$

The 16-neighbour scheme combines two types of grid connections, borrowed from chess terminology (Figure D.2). The 8 *king moves* connect each cell to its immediate neighbours in all directions (up, down, left, right, and the four diagonals), allowing travel at angles of 0°, 45°, 90°, etc. The 8 *knight moves* connect each cell to cells that are one step away in one direction and two steps away in the perpendicular direction (the L-shaped jump a knight makes in chess), adding intermediate angles of approximately 26.6° and 63.4° and their mirrors. Without these additional connections, a route travelling at, say, 30° would be forced into visible staircase-like zigzags between 0° and 45° steps. The knight-move connections allow the algorithm to approximate any travel direction smoothly, producing paths that resemble natural travel corridors rather than pixelated grid artefacts.

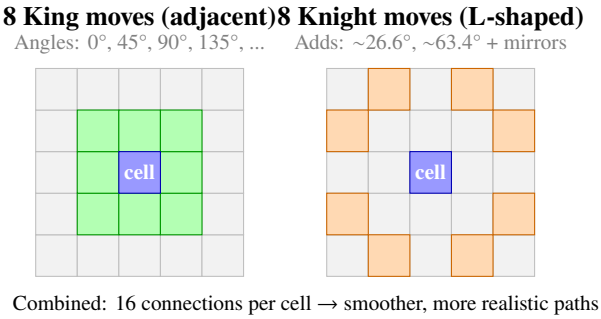


Figure D.2. Grid connectivity scheme. Left: 8 king-move neighbours (green). Right: 8 knight-move neighbours (orange). Together they give each cell 16 outgoing connections, allowing the routing algorithm to approximate any travel direction without staircase artefacts.

Each McCarthy (1939)'s segment is an origin–destination pair. Endpoints are snapped to the nearest land cell that is not a lake or ocean barrier. The cheapest path is found with Dijkstra's algorithm. A segment is dropped if its endpoints snap to the same cell or no path exists.

I. The *Songline* variable

The treatment set is the union of all 10 km grid cells crossed by the 697 kept paths:

$$Songline_g_i = \begin{cases} 1, & \text{if a reconstructed } \textit{Songline} \text{ passes through cell } i; \\ 0, & \text{otherwise.} \end{cases}$$

This yields 7,506 cells coded as 1 out of 82,893 total land cells (approximately 9%).

The reconstruction is anchored in real evidence: the routes are built from thousands of genuine Indigenous heritage sites and are required to connect independently documented *Songline* endpoints. Second, it follows cultural logic rather than modern convenience: paths prefer culturally rich, socially connected country and avoid empty land and major water barriers.

J. Alternative constructions

The reconstruction depends on two key parameters. The first is the decay distance λ, which controls how far each hub's cultural influence reaches across the landscape. A larger λ means that hubs cast influence over a wider area, allowing paths to travel further from recorded sites. The second is the clustering radius r, which determines how close two Indigenous heritage (such as rock-art) sites must be to each other in order to be grouped into the same hub. A smaller r produces more numerous but smaller hubs, while a larger r merges nearby sites into fewer, bigger clusters. These two parameters together define the fundamental cultural geography on which the routing operates: λ shapes the corridors available to paths, and r shapes the hub network itself.

Specification	b	s.e.	N	R^2
Baseline ($\lambda = 50, r = 5$)	0.0121^{***}	(0.0032)	79,722	0.015
$\lambda = 75, r = 5$	0.0069^{**}	(0.0030)	79,722	0.015
$\lambda = 50, r = 2$	0.0080^{***}	(0.0030)	79,722	0.015

The coefficient remains positive and statistically significant across all specifications, ranging from 0.007 to 0.012. The wider decay distance ($\lambda = 75$) produces a somewhat smaller point estimate because paths can use broader corridors and the treatment indicator becomes less selective. Similarly, the tighter clustering radius ($r = 2$) generates more but smaller hubs, slightly altering which cells are classified as treated. In both cases the direction and significance of the result are preserved, confirming that the positive association between reconstructed *Songlines* and European exploration routes is not driven by any particular parameter choice.

Table D.2. Variable definitions and sources (Part I)

Variable	Description	Source
A. *Songlines* and geographic counterfactual		
Songlines	Dummy equal to 1 if the grid cell contains at least one Aboriginal Songline reconstructed using least-cost path analysis.	Authors' elaboration from McCarthy (1939) and collected Indigenous heritage sites.
HMI Paths	Dummy equal to 1 if a least-cost path based on the Human Mobility Index crosses the grid cell.	Authors' computations using the HMI by Özak (2018) and McCarthy (1939).
B. Exploration and knowledge adoption		
Exploration Routes	Dummy equal to 1 if a European inland exploration route crosses the grid cell.	Authors' elaboration digitising Robinson (1927).
Sentiment Score	Exploration-level average relational sentiment toward Aboriginal people, classified as positive, neutral, or negative.	Authors' computations using BERT and RoBERTa models applied to 19th-century explorers' diaries.
Distance to Closest Songline	Distance in kilometres from each grid cell crossed by an exploration route to the nearest Songline.	Authors' elaboration using ArcGIS.
C. Settlement, infrastructure, and outcomes		
Cities	Dummy equal to 1 if the grid cell contains a major inland city established between 1788 and 2000. Cities are also grouped by foundation period.	Data from Kampanelis (2019).
Towns	Dummy equal to 1 if the grid cell contains an inland town established between 1788 and 2000. Towns are also grouped by foundation period.	Data from Elder (2024).
Railway Stations	Dummy equal to 1 if the grid cell contains at least one historical railway station. Stations are also grouped by opening period.	Authors' elaboration using historical railway records, local archives, and railway museum records.
Night Light	Average nighttime luminosity intensity within the grid cell.	Authors' elaboration using VIIRS data from Elvidge et al. (2017).
Population Density	Mean population density within the grid cell in 2015 (persons per km^2).	Authors' elaboration using GHSL data from European Commission (2023).
D. Geographic and historical controls		
Latitude / Longitude	Geographic coordinates of the grid cell centroid.	Authors' elaboration using ArcGIS.
Agricultural Suitability	Average suitability of soil conditions for crop production.	Authors' elaboration using Fischer et al. (2008).
Elevation	Average altitude in metres above sea level.	Authors' elaboration using DIVA-GIS.
Ruggedness	Standard deviation of altitude within the grid cell.	Authors' elaboration using DIVA-GIS.

Table D.3. Variable definitions and sources (Part II)

Variable	Description	Source
D. Geographic and historical controls (continued)		
Precipitation	Average annual precipitation and its standard deviation.	Authors' elaboration using WorldClim data from Hijmans et al. (2005).
Temperature	Average annual temperature and its standard deviation.	Authors' elaboration using WorldClim data from Hijmans et al. (2005).
Distance to the Sea	Geodesic distance from the grid cell centroid to the nearest coastline, in kilometres.	Authors' elaboration using ArcGIS.
Coastal Dummy	Dummy identifying second-layer coastal cells.	Authors' elaboration using ArcGIS.
Distance to Sydney	Geodesic distance from the grid cell centroid to Sydney, in kilometres.	Authors' elaboration using ArcGIS.
Distance to State Capital	Geodesic distance from the grid cell centroid to the corresponding state capital, in kilometres.	Authors' elaboration using ArcGIS.
Distance to Historical Mine	Geodesic distance from the grid cell centroid to the nearest historical mine, in kilometres.	Authors' elaboration using the Australian Mining History Association.
Water Percentage	Percentage of rivers, lakes, and canals within the grid cell.	Authors' elaboration using DIVA-GIS.
E. Appendix and robustness variables		
Trading-Cultural Sites	Dummy equal to 1 if at least one Aboriginal trading-cultural site documented by McCarthy (1939) is located in the grid cell.	Authors' elaboration from McCarthy (1939).
Pre-colonial Settlement	Count of radiocarbon-dated archaeological sites within the grid cell, used as an independent measure of pre-colonial settlement intensity.	Authors' elaboration using the AustArch database Williams et al. (2014).
Cultural Intensity Surface	Continuous measure (0–1) of Aboriginal cultural intensity, constructed from hub-clustered rock art sites and state-weighted to mitigate recording bias.	Authors' elaboration from collected Indigenous heritage sites.

F. Aboriginal Guides

The exploration of Australia during the 19th century was heavily dependent on the knowledge and skills of Aboriginal guides. These guides were instrumental in helping European explorers navigate the challenging and unfamiliar landscapes of Australia. Their expertise was vital in locating water, finding food, and facilitating communication with Indigenous communities. Explorers who employed Aboriginal guides had two main strategies: they could either recruit Aboriginal men from settled areas to join the expedition from the start or engage with local Aboriginal people when they encountered them during their journey to gain valuable knowledge directly from them (Clarke 2008; McLaren 1996). In this Appendix, we present comprehensive evidence on the Aboriginal companions of European explorers. The following online file here includes an extensive list of 77 documented inland explorers, primarily from the 19th century, along with compelling evidence demonstrating that they were indeed accompanied by Aboriginal guides. For each explorer, we also cite the source(s) that substantiate this evidence. These source(s), among others, include published explorers' diaries, the Australian Dictionary of Biography, websites of local councils, as well as documents from the National Library of Australia. We acknowledge that it is impossible to document every explorer who ventured across Australia, as some expeditions may have gone entirely unrecorded. Nevertheless, to the best of our knowledge, this is the first systematic compilation of consistent data on inland explorers and their Aboriginal guides. Based on this evidence, in the following subsections, we present key examples highlighting the importance of Aboriginal companions for a selected group of explorers from among all those presented in the Appendix. They are organised by specific regions across Australia.

F.1. Southeastern Australia: The Vital Role of Indigenous Knowledge

In Southeastern Australia, where two of the most significant urban centres, Sydney and Melbourne, emerged, Aboriginal guides played a crucial role in supporting European explorations. Hamilton Hume, a renowned Australian explorer, often relied on the invaluable guidance of Abo-

riginal people during his numerous expeditions. From his earliest explorations, Hume recognised the importance of Aboriginal knowledge and skills in navigating the challenging Australian terrain. On his first journey at the age of 17 in 1814, he was accompanied by an Aboriginal man who helped him traverse the Berrima-Bong Bong district, part of the Southern Highlands of New South Wales (Hume 1966). Hume's respect for Aboriginal expertise continued throughout his career, most notably during his collaboration with Charles Sturt on the 1828 expedition to the interior. Hume's ability to communicate and negotiate with Aboriginal communities was crucial in securing safe passage and gaining local insights, which significantly contributed to the success of his explorations. In addition to Hume, other explorers such as Charles Throsby who explored south-eastern Australia from 1817 to 1821, similarly relied on Aboriginal guidance. Broughton, whose Aboriginal name was recorded as Toodwick, Toodood, or Toodwit, became a trusted figure, serving as a guide and translator on several of Throsby's expeditions (Campbell 2005). In the same vein, George William Evans was accompanied by Aboriginal guides during his explorations from 1812 up to 1818. One notable instance is his 1812 expedition from Jervis Bay to the Shoalhaven River, where he was guided by an Aboriginal man named Bundle (also known as Bon-del or Bundel) from the Gundungurra area. In November 1813, he led a team of five men to survey a route through the mountains, aiming to reach the fertile plains beyond for agricultural expansion to support the Port Jackson colony. They traced the course of the Wambool (Macquarie River) westward into Wiradjuri territory, extending their journey to Killongbutta, approximately 40 kilometres from Bathurst. The path they took was along a well-established route (i.e., a songline) historically used by the Dharug and Gandangara peoples for trade with the Wiradjuri (Bathurst Local Aboriginal Consultative Committee 2011).

F.2. Central Australia: The Harsh Realities and Aboriginal Guidance

The stark and unforgiving landscapes of central Australia presented formidable challenges to early explorers. Ernest Giles, whose expeditions in the 1870s ventured into the arid interior, often relied on Aboriginal guides to survive. The presence of guides like Tommy Oldham, a Wirangu

man, was crucial in navigating the desolate deserts, finding scarce water sources, and avoiding natural dangers. On one significant occasion, Oldham discovered Queen Victoria Spring, a vital water source that saved Giles' expedition from dying of thirst. In his personal diary, Australia Twice Traversed (Giles 1889), Giles notes:

"It was Mr. Tietkens's turn to steer, and he started on foot ahead of the string of camels for that purpose. He gave Tommy his little riding-bull, the best leading camel we have, and told him to go on top of a white sandhill to our left, a little south of us, and try if he could find any fresh blacks' tracks, or other indications of water. I did not know that Tommy had gone, nor could I see that Tietkens was walking—it was an extraordinary event when the whole string of camels could be seen at once in a line in this country—and we had been traveling some two miles and a half when Alec Ross and Peter Nicholls declared that they heard Tommy calling out "water!" I never will believe these things until they are proved, so I kept the party still going on. However, even I, soon ceased to doubt, for Tommy came rushing through the scrubs full gallop, and, between a scream and a howl, yelled out quite loud enough now even for me to hear, "Water! water! plenty water here! come on! come on! this way! this way! come on, Mr. Giles! mine been find 'em plenty water!" I checked his excitement a moment and asked whether it was a native well he had found, and should we have to work at it with the shovel? Tommy said, "No fear shovel, that fellow water sit down meself (i.e. itself) along a ground, camel he drink 'em meself. Of course we turned the long string after him." [25]

Giles' explorations clearly demonstrate the critical impact of Aboriginal expertise; when skilled guides were involved, the survival and success of his party increased significantly. In the same vein, another significant explorer, John McDouall Stuart, also found that Aboriginal guidance was essential during his expeditions through central Australia. Stuart's journeys often involved traversing extremely harsh environments, and the Aboriginal guides' knowledge of the land proved to be the difference between success and failure. Stuart embarked on an expedition on 14 May 1858 with an assistant and an Aboriginal tracker, carrying provisions for four weeks, to explore beyond

[25] Mr. William Henry Tietkens (1844–1933) served as second-in-command on several of Giles's expeditions.

Lake Torrens and Lake Gairdner in search of grazing land. The group reached as far as Coober Pedy before turning south and then west. However, on 3 August, the Aboriginal tracker departed, leaving Stuart and his team to face significant challenges. With their supplies and water nearly depleted and their horses lame, they struggled to reach T. M. Gibson's outstation at Streaky Bay by the 22nd of August.[26] After resting for ten days, Stuart returned to Adelaide (Stuart 1865). The departure of the Aboriginal tracker on the 3rd of August clearly marked a turning point in Stuart's expedition. Without the essential guidance and knowledge provided by the Aboriginal guide, the European party encountered severe difficulties, including the depletion of their supplies and water, as well as the deteriorating condition of their horses.

F.3. Western Australia: Success Through Collaboration

In Western Australia, the expeditions of John and Alexander Forrest are prime examples of successful collaboration between European explorers and Aboriginal guides. Windich, an Aboriginal guide who accompanied the Forrest brothers, played a pivotal role in their exploration of Lake Moore and Lake Barlee in the late 1860s and early 1870s. His deep understanding of the region's geography, including the intricate web of waterholes and their seasonal variations, was instrumental in the success of these expeditions. Windich was on particularly good terms with the Forrest family, and both John and Alexander Forrest heavily relied on him in their daily search for drinking water and horse feed. Windich usually acted as the scout, adept at finding native wells or waterholes in the rocky outcrops. While the Forrest brothers, being skilled surveyors, were capable of navigating using astronomical observations and were not at risk of getting lost in the inland deserts, they depended greatly on their Aboriginal tracker for finding drinking water and forage for their horses. The mutual respect and reliance between the Forrest brothers and Windich highlight how critical Indigenous knowledge was in unlocking the mysteries of Western Australia's remote

[26] T. M. Gibson's outstation refers to a remote or secondary station or homestead that was part of a larger pastoral property owned or managed by someone named T. M. Gibson. In the context of John McDouall Stuart's exploration, it would have been a place where the explorers could find shelter, supplies, and possibly assistance during their journey. Outstations were common in the Australian outback during the 19th century and served as critical support points for explorers, stockmen, and travellers, especially in isolated regions far from main settlements.

landscapes (Forrest 1875).[27]

Similarly, the explorations of Peter Egerton Warburton also exemplify the importance of Aboriginal guides in Western Australia. His mid-1850s expeditions benefited immensely from the guidance of an Aboriginal companion named Charley, whose expertise in navigating the vast, often hostile terrains was indispensable. Warburton credited Charley with saving the expedition multiple times by seeking out water and being sent ahead to scout paths and resources. In his book, called Journey Across the Western Interior of Australia (Warburton 1875), which is a published account of his diaries from his explorations, he notes:

"At 1 p.m. Charley returned with news that they had tracked the natives round from north-west to south-west, and found their well. Lewis remained there to clear it out, and I shall move on towards it this afternoon, in time to water the camels."

In another section of the same book he mentions:

"Sent Lewis and Charley at daylight to follow up the native tracks; Charley returned at 10 a.m. with news of a good well about five miles to the north-west. Lewis stayed at the well to save the camel."[28]

Warburton's diary notes suggest that his journeys were deeply intertwined with Charley's knowledge, underscoring his essential role as an Aboriginal guide. Charley's contribution to Warburton's expedition was twofold. Firstly, his ability to expertly track Aboriginal paths was crucial in guiding the entire Warburton party through the challenging terrain. Secondly, his deep knowledge of the natural environment and local geography was vital to the survival of the expedition, ensuring that the party could find essential resources such as water in the harsh and unforgiving landscape.

F.4. Northern Australia: The Mixed Fortunes of Exploration

Northern Australia, with its remote and often hostile environment, posed unique challenges to explorers like Ludwig Leichhardt. His expeditions from the Darling Downs to Port Essington in

[27] Other Aboriginal tracker(s) such as Tommy Pierre were also Aboriginal guides with Forrest's party.
[28] William Lewis was a European member of Warburton's expedition across the western interior of Australia.

the 1840s were marked by a reliance on Aboriginal guides, particularly Harry Brown and Charley Fisher, whose contributions were crucial for the survival and success of the exploration party. Brown, known for his exceptional navigation skills, frequently scouted ahead to find the best routes and locate vital water sources, such as the freshwater lagoon that became known as "Brown's Lagoons". His role was so essential that even when Leichhardt realised they had insufficient supplies during his first major expedition in 1844, he chose to send two European members (Caleb and Hodgson) back while retaining both Brown and Fisher (Leichhardt 1847). Brown's importance was further highlighted when he corrected Leichhardt's course during a critical moment, ensuring the party remained on the right path. More specifically, during the second expedition in 1846, Leichhardt mistakenly led the party in the wrong direction. Brown, with his deep knowledge of the terrain, realised the error and informed Mann, another member of the expedition, that they were off course.[29] Despite Leichhardt's initial confidence in his navigational skills, Brown's insight proved correct, and he was called upon to lead the party through the difficult terrain, ultimately ensuring their survival and continued progress (Blyton 2015). Among the exceptions to expeditions accompanied by Aboriginal guides that ultimately were not successful is Ludwig Leichhardt's third and final journey in 1848. Despite the presence of two Aboriginal guides, Wommai and Billy, the expedition ended in tragedy, with the entire party disappearing without a trace. This outcome contrasts sharply with Leichhardt's earlier expeditions, where the presence of skilled guides like Harry Brown had been critical to their success. Similarly, Edmund Kennedy and his guide Jackey-Jackey formed a partnership that was crucial in their exploration of northern Queensland. Despite the many dangers they faced, Jackey-Jackey's intimate knowledge of the land and his ability to navigate through dense rainforests and swamps were invaluable to Kennedy's expedition. Jackey-Jackey's loyalty and bravery were particularly evident towards the end of the expedition when he remained with Kennedy until his death and ultimately reached the supply ship alone.

[29] John F. Mann was one of the European members of Ludwig Leichhardt's second expedition in 1846.

F.5. Conclusion on the Role of Aboriginal Guides

The role of Aboriginal guides in the exploration of Australia extends far beyond the immediate survival of the explorers they accompanied. These guides were often recruited from survivors of the early waves of British colonisation and were present at pivotal moments in Australian exploration history. The partnerships formed between explorers and Aboriginal companions—such as those between Matthew Flinders and Boongaree, George Grey and Kaiber, John Mitchell and Barney, Throsby and Toodwick, Edward Eyre and Wylie, Ludwig Leichhardt and Charley Fisher, Peter Warburton and Charley, Ernest Giles and Jimmy, and John Forrest and Windich—were instrumental in opening up the Australian interior to European settlement (Reynolds 1990; McLaren 1996; Clarke 2008). Apart from the most well-known explorers, lesser-known figures such as James Blackman were also accompanied by an Aboriginal servant named Aaron during his exploration from Bathurst to the Cudgegong River in 1821, while Nathaniel ”Nat” Buchanan relied on a Warumungu man named Jack during his explorations in the Tennant Creek area in 1896. Similarly, William Hann, a less known explorer, was guided by an Aboriginal man named Jerry during his 1872 Northern Expedition of Cape York Peninsula. Many of these Aboriginal guides were later recognised for their contributions through the awarding of brass or copper breastplates, acknowledging their indispensable role in these exploratory missions (Cleary 1993; Troy 1993). Their deep understanding of the land, encoded in their ecological knowledge and cultural practices, provided European explorers with the tools they needed to survive and thrive in an unfamiliar and often unforgiving environment. The tragic fate of explorers like Burke and Wills, who failed to fully integrate Aboriginal guides into their team, highlights the dangers of neglecting this vital resource. Their experience stands as a stark reminder of the critical importance of Aboriginal expertise in the exploration of Australia (Clarke 2008; Cleary 1993).

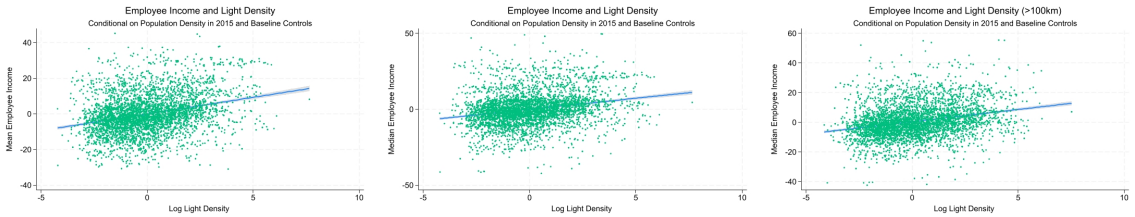
G. Luminosity as a Measure of Local Income

In our analysis, we utilise luminosity as an indicator of economic activity. While previous studies have demonstrated the efficacy of using light as a proxy for regional development in a few developed countries (Ebener et al. (2005), Ghosh et al. (2010), and Sutton et al. (2007)), the applicability of this approach to Australia warrants elaboration. Despite being considered a developed economy, Australia's unique historical trajectory, particularly stemming from European colonisation, alongside its varied geomorphology, has resulted in a predominantly coastal urban distribution with vast, sparsely inhabited interior regions. Although our primary analysis excludes the coastal areas hosting major Australian cities, questions may be raised regarding the ability of nightlights to accurately capture economic activity in both the interior and the coastal part.

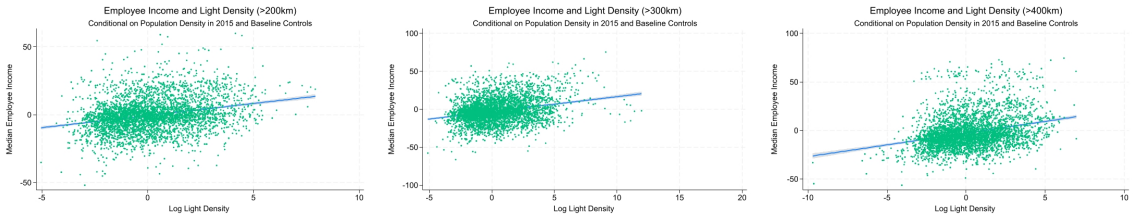
Hence, we explore the correlation between our measure of light density (the continuous measure in logs) and two conventional indicators of local economic growth: mean income and median income. Specifically, we draw upon data from Australian Bureau of Statistics (2019), which furnishes mean and median employee income values at the Local Government Area (LGA) level spanning 2013 to 2018. Our preference for 2015 data aligns with the temporal consistency of our light and population density metrics, both also sourced from 2015.

Each grid cell in our analysis is assigned to its corresponding LGA. In instances where a grid cell is intersected by more than one LGA, we join it with the LGA that covers its larger part. Subsequently, these grid cells inherit the mean and median income values of their respective LGAs. Leveraging both mean and median income offers a comprehensive perspective: while mean income provides a central tendency measure, the median is more robust to outliers, rendering it a robust measure in the presence of skewed distributions or income inequality.

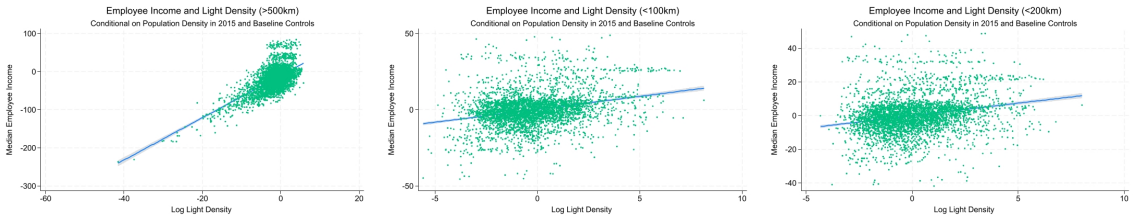
Given that our income data pertains to larger geographic areas than our grid cell analysis, we restrict our correlation analysis to 4,379 grid cells with non-zero values for both population density and nightlights. This is approximately 5.5% of our total sample. This approach excludes grid cells that lack income, light, or population data, thus enhancing the robustness of our correlation



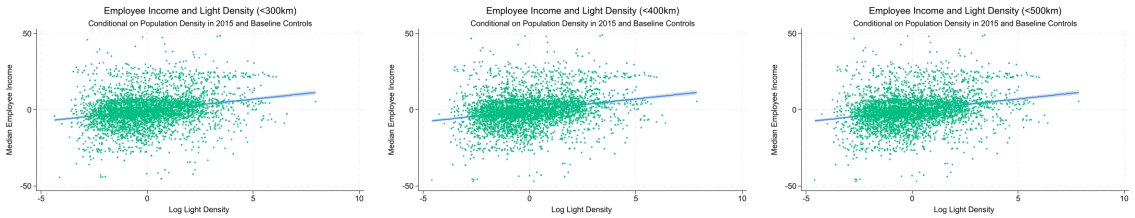
(a) Mean Income (b) Median Income (c) > 100 km



(d) > 200 km (e) > 300 km (f) > 400 km



(g) > 500 km (h) < 100 km (i) < 200 km



(j) < 300 km (k) < 400 km (l) < 500 km

Figure D.3. Relationship between luminosity and income conditional on population density and baseline controls.

assessments.

Figure D.3 demonstrates the robust positive correlation between our luminosity index and both income metrics, signalling significant associations. This positive relationship persists across both measures, affirming the strength of the correlation.

Figure D.3 also delves deeper into these associations by excluding coastal regions within the 100km to 500km range. Remarkably, even with this exclusion, the correlations remain high and the slopes of the relationships become steeper. This finding is particularly pronounced when focusing on the remote interior regions of Australia, traditionally characterised by lower development levels.

Conversely, we conduct a complementary analysis by retaining the coastal areas and excluding the interior regions extending beyond 100km to 500km from the coast. Despite this spatial shift, the positive correlations persist, reaffirming the reliability of our luminosity index as an indicator of economic activity at the local level across diverse geographic contexts.

In sum, our comprehensive analysis underscores the efficacy of luminosity as a reliable proxy for economic activity in Australia, even at the granular local level, thereby bolstering its utility for our empirical research.

H. A Contemporary *Songline*: *Spinifex Express Train Song*

Aboriginal Song	English Translation
Junirri karririlarnani	**Through the hills**
Parni parnta jirraparpa	**Driving along carrying its load**
Katinya Wungkuwungkunyanguru	**Past the Wungkuwungku Hills**
Wungkuwungkunyanguru munmurlkurtu	**Through Wungkuwungku Hills**
Pamti mirtamirta	**The white smoke billowing**
Yankakemanu yukurru minta	**In a chain from the 'dog's nose'**
Munmurlkurtu pamti mirtamirta	**White smoke billowing**
Yukurru mimta Tapata tapa	**The 'dog's nose' leaving behind**
Nyarukarra yukurru mimta yurarirrin wanyjamu	**Lots of smoke left behind**
Parni parnta jirraparpa katinya	**Driving the engine**
Yinyji tapa Tapata tapa nyarukarra	**Over Tahbatabba river**
Warrampa pilayi parrakurrarrakuman	**Country coming past and left behind**
Nyarukarra jurituraka wipirnlu	**Passing Red Rock Creek**
Wipirnlu janpirrirpa nguntuntu	**The big trucks shaking in the middle**
Karnirra kurturtula kamanyjanu	**(Train) calls out at**
Nyarukarrata marnkurrkamia mungkarta	**Big West Strelley Creek, going, going**
Yampangarri janpirrirpa nguntuntu karnirra	**Shaking in the middle**
Kurturtu kamanyjanu nyarukarrala	**Going and going**
Marnkurrkamia mungkarta yampangarri	**At Big West Strelley Creek**
Yampangarri warrampa	**Going along, the country going fast**
Jarangkunu murti pilurru	**Still going along to the destination**
Pajanpinpiyjarra Pimpinpala juraka	**It's Pinpin Creek coming up fast**
Wungkurrura purinju warrampa	**Truck shadow towing through the country**
Pinkarta jinyjipajarnu	**On the pedal**
Jinyjipajarnu warrampa	**Foot stamping on the pedal**
Jarangkunu murti pilurru	**The country going along, still running past**
Pajanpinpiyjarra Pimpimpala juraka	**Coming up fast, it's Pinpin Creek**

Notes: The story of the *Spinifex Express Train Song* reflects the composer's journey, using the train's route to recount and honour the connection to the land. It narrates the passage from waterhole to waterhole, blending the modern experience of train travel with the traditional practice of mapping country through song. This creates a new kind of songline or storyline, rooted in contemporary experience but resonating with cultural traditions. First recorded in 1964, the song was preserved by the *Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS)* for 50 years. In 2014, it was reintroduced to the community through an exhibition at the *Transport Museum* in Perth and shared with the public as part of efforts to preserve and celebrate Aboriginal cultural heritage. The song was displayed at the *Revolutions Train Museum* in Whiteman Park, Perth, as part of a collaboration with the *Wangka Maya Pilbara Aboriginal Language Centre* and AIATSIS. Its inclusion in the *ABC Open Mother Tongue Project* highlighted the significance of storytelling in Aboriginal languages and their connection to place and identity.

Appendix References

Angeles, Luis and Aldo Elizalde (2017). "Pre-colonial institutions and socioeconomic development: The case of Latin America". In: *Journal of Development Economics* 124, pp. 22–40.

Australian Bureau of Statistics (2019). *Personal Income in Australia, 2011–12 to 2017–18*. Accessed: 2026-05-08. URL: https://www.abs.gov.au/statistics/labour/earnings-and-working-conditions/personal-income-australia/2011-12-2017-18.

Bathurst Local Aboriginal Consultative Committee (2011). *The Wiradjuri people*. Bathurst, N.S.W.: Bathurst Local Aboriginal Consultative Committee.

Blyton, Greg (2015). "Harry Brown (c. 1819-1854): Contribution of an Aboriginal guide in Australian exploration". In: *Aboriginal History* 39, pp. 63–82.

Campbell, Keith (2005). "Broughton (c.1798–c.1850)". In: *Australian Dictionary of Biography*. URL: https://adb.anu.edu.au/biography/broughton-12820.

Clarke, Philip A (2008). *Aboriginal plant collectors*. Rosenberg Pub.

Cleary, Tania (1993). *Poignant Regalia: 19th Century Aboriginal Breastplates & Images: A Catalogue of Aboriginal Breastplates Held in Public, Regional and Private Collections in New South Wales, Victoria, Tasmania, Queensland, South Australia, Western Australia and the Australian Capital Territory*. Exhibition venues: Greenway Gallery, Hyde Park Barracks, 26 May-4 July 1993. Glebe, NSW: Historic Houses Trust of New South Wales.

Colella, F., R. Lalive, S.O. Sakalli, and M. Thoenig (2019). "Inference with Arbitrary Clustering". In: *IZA Discussion Paper* 12584.

Conley, T.G. (1999). "GMM Estimation with Cross Sectional Dependence". In: *Journal of Econometrics* 92.1, pp. 1–45.

Druckenmiller, Hannah and Solomon Hsiang (2018). *Accounting for unobservable heterogeneity in cross section using spatial first differences*. Tech. rep. National Bureau of Economic Research.

Dube, Arindrajit, T. William Lester, and Michael Reich (2010). "Minimum Wage Effects Across State Borders". In: *The Review of Economics and Statistics* 92.4, pp. 945–964. DOI: 10.1162/REST_a_00039. URL: https://doi.org/10.1162/REST_a_00039.

Ebener, S., C. Murray, A. Tandon, and C.C. Elvidge (2005). "From Wealth to Health: Modelling the Distribution of Income per Capita at the Sub-National Level Using Night-Time Light Imagery". In: *International Journal of Health Geographics* 4.1, p. 5.

Elder, Bruce (2024). *Aussie Towns: A Definitive Travel Guide to Australian Towns and Destinations*. Accessed: 2024-09-10. URL: https://www.aussietowns.com.au/.

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh (2017). "VIIRS night-time lights". In: *International Journal of Remote Sensing* 38.21, pp. 5860–5879.

European Commission (2023). *GHSL Data Package 2023*. URL: https://ghsl.jrc.ec.europa.eu/documents/GHSL_Data_Package_2023.pdf?t=1683540422 (visited on 09/06/2023).

Fischer, Guenther, Freddy Nachtergaele, Sylvia Prieler, HT Van Velthuizen, Luc Verelst, and David Wiberg (2008). *Global agro-ecological zones assessment for agriculture (GAEZ 2008)*. Tech. rep. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Forrest, John (1875). *Explorations in Australia*. London: Sampson Low, Marston, Low, & Searle.

Ghosh, T., R.L. Powell, C.D. Elvidge, K.E. Baugh, P.C. Sutton, and S. Anderson (2010). "Shedding Light on the Global Distribution of Economic Activity". In: *The Open Geography Journal* 3.1, pp. 148–161.

Giles, Ernest (1889). *Australia Twice Traversed: The Romance of Exploration, Being a Narrative Compiled from the Journals of Five Exploring Expeditions into and through Central South Australia and Western Australia, from 1872 to 1876*. Vol. 1. London: Sampson Low, Marston, Searle, & Rivington. URL: https://books.google.gr/books/about/Australia_Twice_Traversed.html?id=gwsOAAAAYAAJ&redir_esc=y.

Hijmans, Robert J, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis (2005). "Very high resolution interpolated climate surfaces for global land areas". In: *International Journal of Climatology: A Journal of the Royal Meteorological Society* 25.15, pp. 1965–1978.

Hume, S. H. (1966). "Hamilton Hume (1797–1873)". In: *Australian Dictionary of Biography*. URL: https://adb.anu.edu.au/biography/hume-hamilton-2211.

Kampanelis, Sotiris (2019). "It's time for Westernization: the advantages of the early start for long-term economic development at the local level". In: *Oxford Economic Papers* 71.4, pp. 996–1025.

Leichhardt, Ludwig (1847). *Journal of an Overland Expedition in Australia, From Moreton Bay to Port Essington, A Distance of Upwards of 3000 Miles, During the Years 1844-1845*. London: T. & W. Boone.

McCarthy, Frederick David (1939). ""Trade in Aboriginal Australia, and "Trade" Relationships with Torres Strait, New Guinea and Malaya"". In: *Oceania* 9.4, pp. 405–438.

McLaren, Glen (1996). *Beyond Leichhardt: Bushcraft and the Exploration of Australia*. Fremantle, W.A.: Fremantle Arts Centre Press.

Michalopoulos, Stelios and Elias Papaioannou (2013). "Pre-colonial ethnic institutions and contemporary African development". In: *Econometrica* 81.1, pp. 113–152.

O'Grady, Geoffrey N, Carl F Voegelin, and Florence M Voegelin (1966). "Languages of the world: Indo-Pacific fascicle six". In: *Anthropological Linguistics* 8.2, pp. 1–197.

Oster, Emily (2019). "Unobservable selection and coefficient stability: Theory and evidence". In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Özak, Ömer (2018). "Distance to the pre-industrial technological frontier and economic development". In: *Journal of Economic Growth* 23.2, pp. 175–221.

Reynolds, Henry (1990). *With the white people: the crucial role of Aborigines in the exploration and development of Australia*. Penguin Books, Ringwood, Vic.

Robinson, H.E.C. (1927). *The discovery and exploration of Australia, 1519 to 1901 by sea and land*. Sydney: H.E.C. Robinson Ltd.

Roodman, David, Morten Ørregaard Nielsen, James G MacKinnon, and Matthew D Webb (2019). "Fast and wild: Bootstrap inference in Stata using boottest". In: *The Stata Journal* 19.1, pp. 4–60.

Stuart, John McDouall (1865). *Explorations in Australia: The Journals of John McDouall Stuart during the Years 1858, 1859, 1860, 1861, & 1862, When He Fixed the Centre of the Continent and Successfully Crossed It from Sea to Sea*. Ed. by William Hardman. 2nd. London: Saunders, Otley, and Co.

Sutton, P.C., C.D. Elvidge, and T. Ghosh (2007). "Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery". In: *International Journal of Ecological Economics & Statistics* 8.S07, pp. 5–21.

Troy, Jakelin (1993). *King Plates: A History of Aboriginal Gorgets*. Aboriginal Studies Press.

Warburton, Peter E. (1875). *Journey Across the Western Interior of Australia: With an Introduction and Additions*. London: Sampson Low, Marston, Low, & Searle.

Williams, Alan N., Sean Ulm, Mike Smith, and Jacqueline Reid (2014). "AustArch: A Database of 14C and Non-14C Ages from Archaeological Sites in Australia – Composition, Compilation and Review (Data Paper)". In: *Internet Archaeology* 36. DOI: 10.11141/ia.36.6. URL: https://doi.org/10.11141/ia.36.6.