The New York Times

Before an Arrest, Officers Tossed a (Virtual) Coin

July 14, 2018

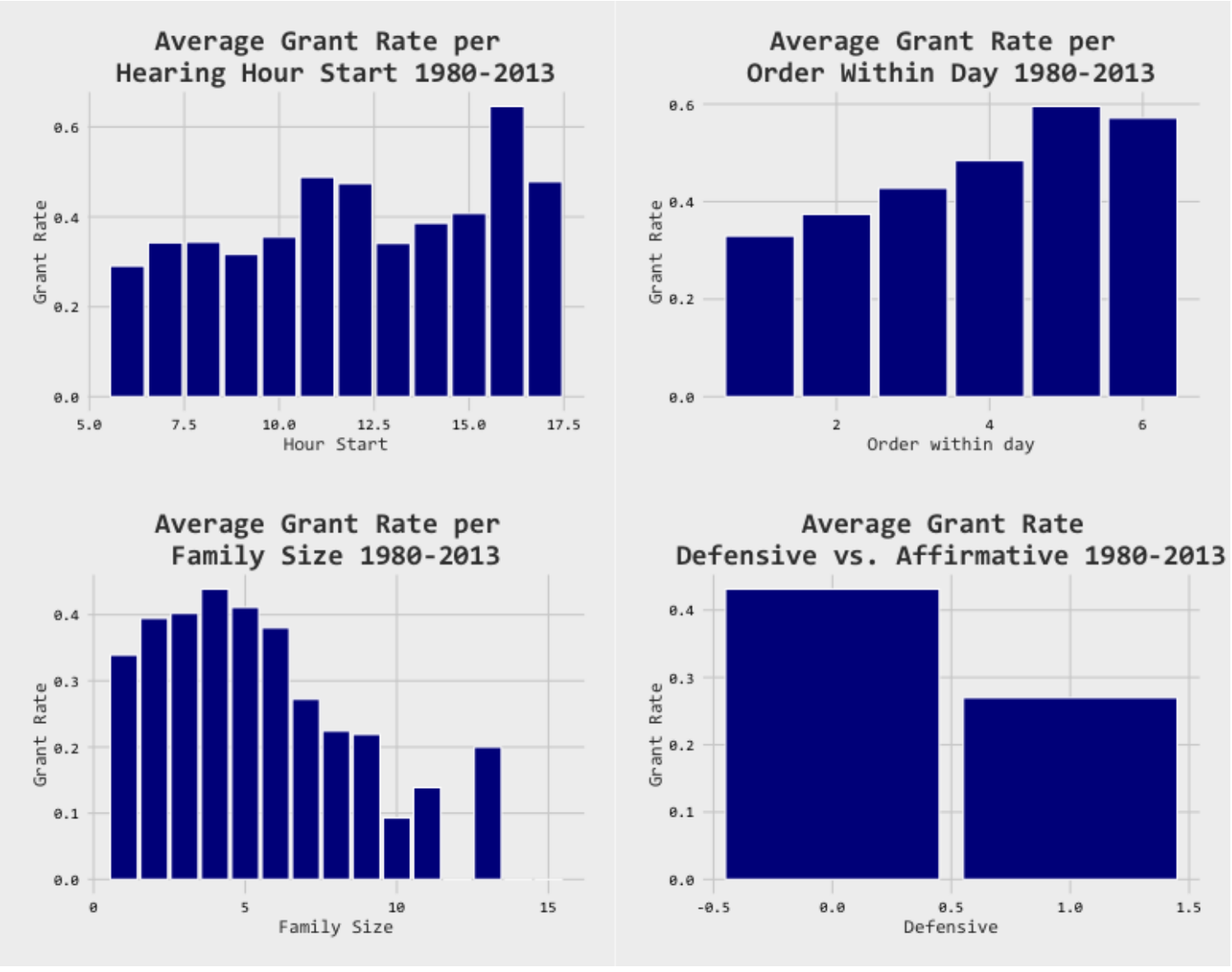
Three uses of judicial analytics

- Predictive analytics of judges
  - Score judicial performance
- Predictive analytics for causal inference
  - Law platform for automated prospective impact analysis
- Predictive analytics to increase recognition, dignity
  - Randomized control trials

Administrative universe on Asylum Courts 1981-2013

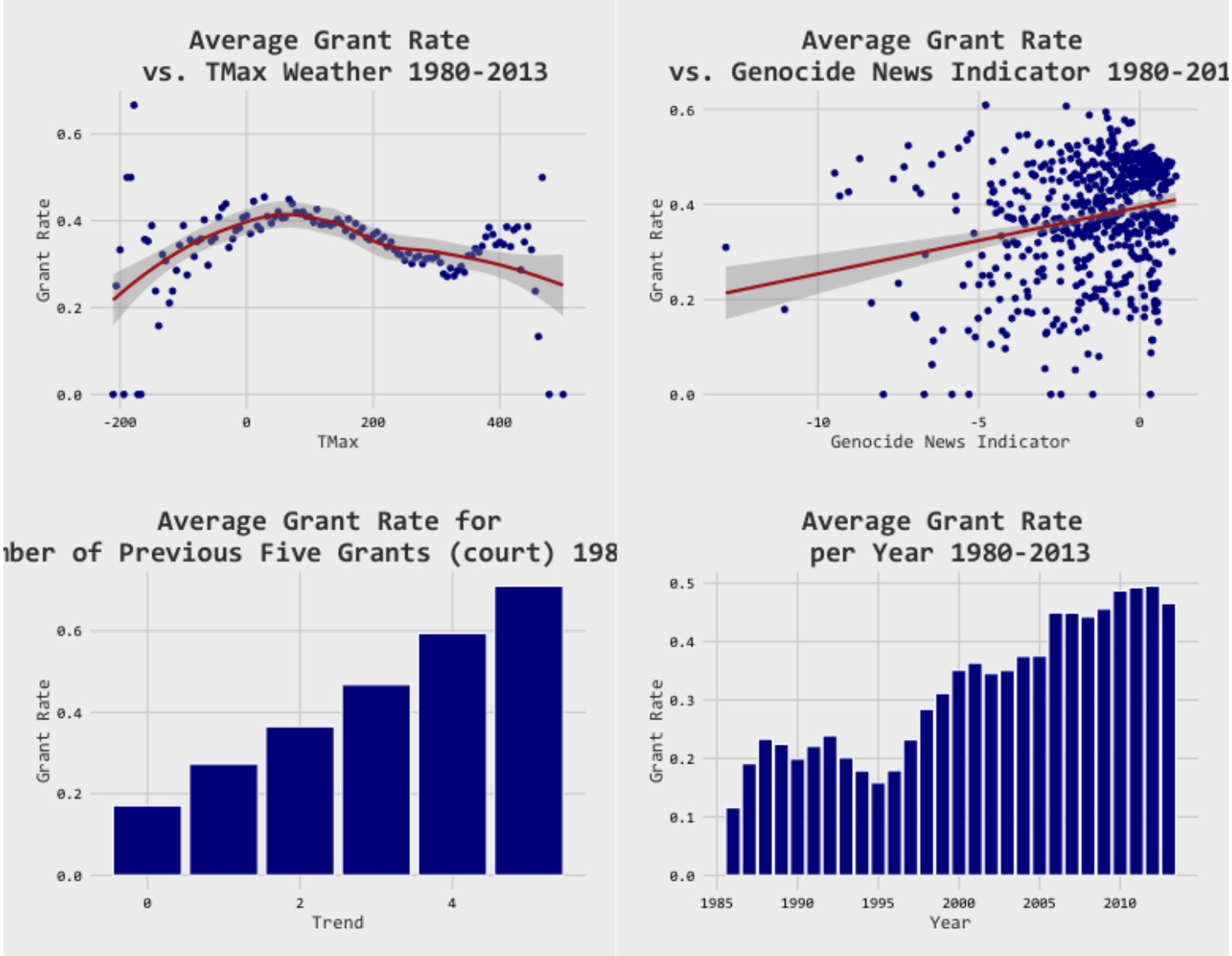
- 492,903 asylum decisions, 336 hearing locations, 441 judges
- High stakes: Denial of asylum usually results in deportation
  - "Applicant for asylum reasonably fears imprisonment, torture, or death if forced to return to her home country" (Stanford Law Review 2007)
  - Cases filed within each court are randomly assigned to judges
- Average grant rate is roughly 35%
- Using only data available up to the **decision date**, 82% accuracy
- Using only data available up to the **case opening**, 78% accuracy

Predictability of Asylum Decisions



More lenient before lunch & towards end of day & for affirmative asylum, U-shape with family size

Predictability of Asylum Decisions

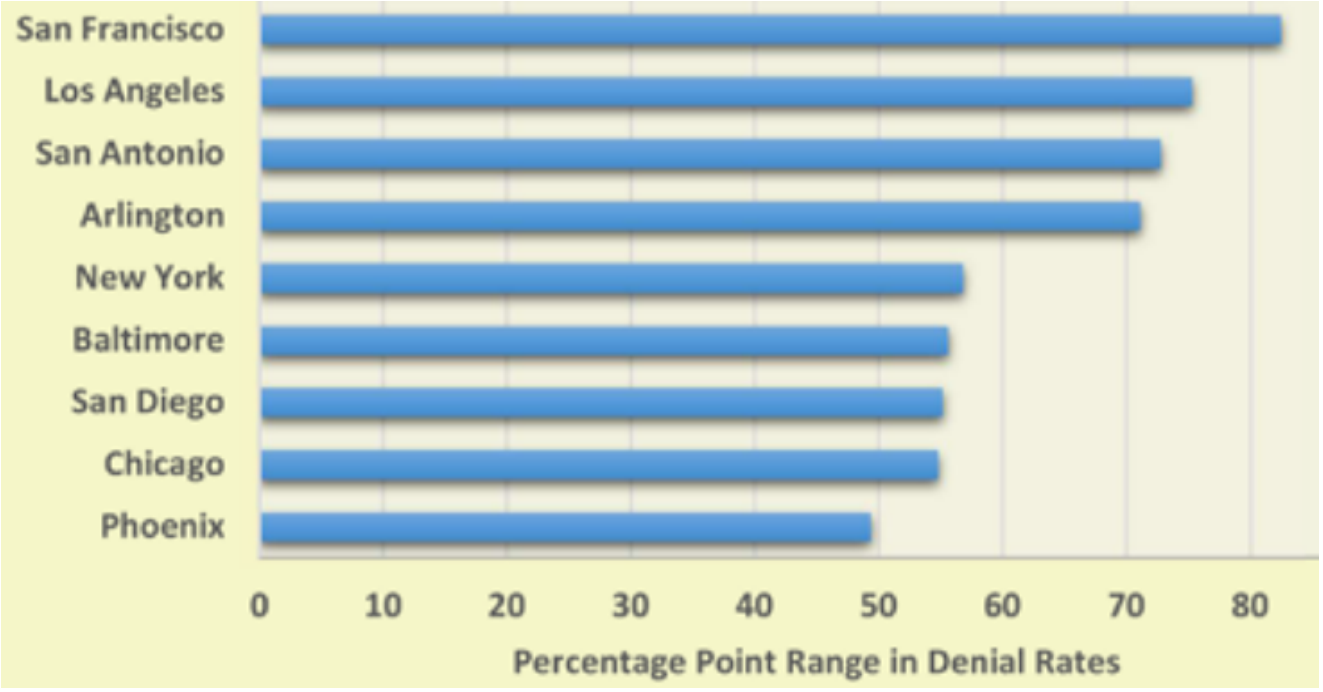


More lenient with good weather & genocide news indicator & over time & female judges. Strong trend factors within-court & within-judge.

Top Ten Countries by Applicants

Country	Count	Percentage	Grant Rate
China	107964	0.19	0.53
Haiti	42013	0.074	0.16
El Salvador	41626	0.074	0.087
Guatemala	34705	0.061	0.11
Colombia	27713	0.049	0.35
India	19161	0.034	0.37
Mexico	19031	0.034	0.073
Nicaragua	15987	0.028	0.2
Albania	12036	0.021	0.52
Indonesia	11399	0.02	0.32

Disparities in Grant Rates



- In San Francisco, one judge grants 90.6% of asylum requests, while another judge grants just 2.9%!

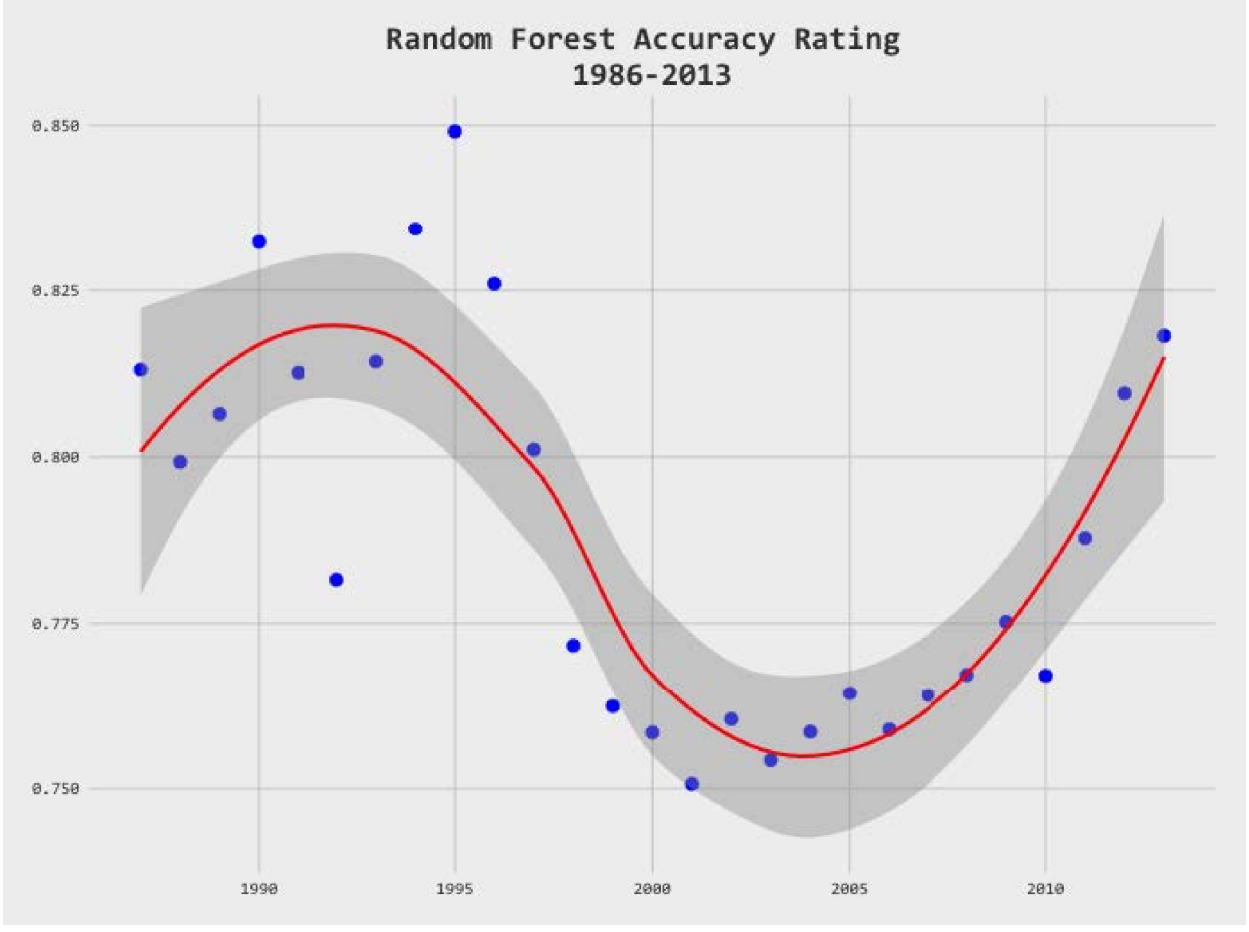
Machine Learning Approach

1. To predict an outcome
    1. asylum granted, appeal denied, charges brought
2. Train a classifer
    1. logistic regression, random forest, or gradient boosted machine
3. Using high-dimensional feature set for defendant and case

Classification Steps

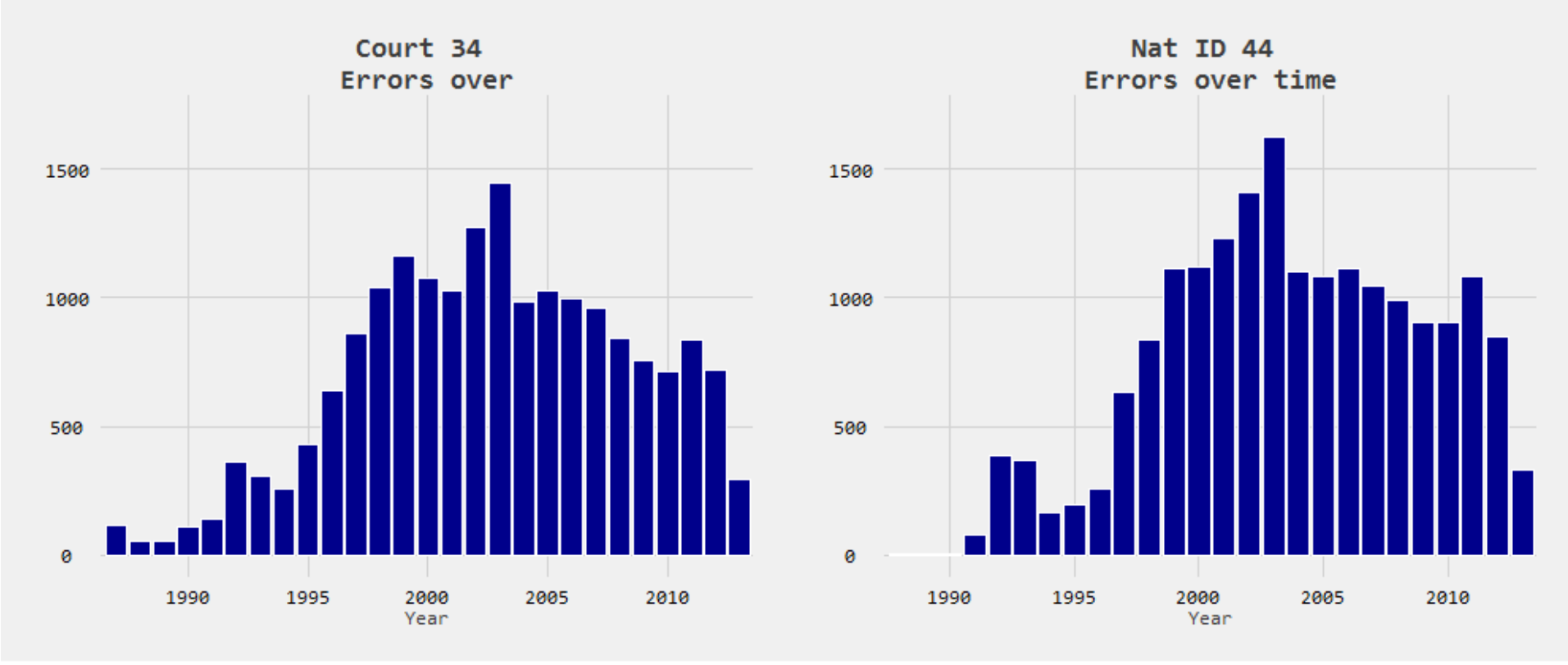
1. Do train/test split on data (e.g. 80% / 20%)
2. Cross-validation grid search on training set to select hyperparameters
3. Form predictions in test set and compare to true outcome
4. Sequential approach
    1. Trained parameter set on all cases up to preceding Dec 31 to make predictions for the following twelve months.

Random Forest



Mid-2000's performance dip on test set.

Predictability of Asylum Decisions



- 40% of misclassifications from Congo applicants in one NYC court
- Second Congo War began in 1998 and ended in 2003

Predictability of Asylum Decisions

Table 1: Random Forest Final Importances

Category	Feature	Weight
Case Information	Attorney ID	0.01
	Court ID	0.01
	Defensive	0.01
	Hour Start	0.004
	Lawyer	0.02
	Nationality	0.024
	# in family	0.002
	Order in day	0.002
	Start time	0.004
	Other	0.11
	Total Case	**0.20**
Court Information	Hearing Location	0.01
	Other	0.06
	Total Court	**0.07**
Judge Information	College	0.007
	Judge ID	0.007
	Experience	0.006
	Male/Female	0.004
	Law School	0.007
	Graduation Year	0.006
	Military Years	0.001
	# of Cases	0.014
	President Appointed	0.002
	Year Appointed	0.005
	Other	0.051
	Total Judge	**0.10**
News Trends	Asylum	0.006
	Cleansing	0.005
	Crisis	0.006
	Genocide	0.006
	Refugee	0.006
	Aggregate	0.006
	Total News	**0.07**
Trend Features	Judge Avg. grant	0.179
	Avg. grant p. natn.	0.14
	Previous five	0.058
	Other	0.115
	Total Trend	**0.49**
Weather	Cloud Coverage	0.004
	Precipitation	0.002
	Snow	0.001
	Other	0.017
	Total Weather	**0.02**

- predominantly trend features and judicial characteristics - unfair?
  - one third-driven by case, news events, and court information

Early Predictability

- If case outcomes could be **completely predicted**
  - after a particular judge is assigned,
  - but **prior to judicial inquiry** into the case,
  - this would indicate that judges **did not take into account** any differences between cases.
- There may be cases for which country and date of application *should* completely determine outcomes (e.g., during violent conflict)
  - But significant inter-judge disparities in predictability suggest that this understanding of the country circumstances does not apply to all
- Some judges are highly predictable, always granting or rejecting
  - Snap judgments and predetermined judgments (Ambady and Rosenthal 1993)
  - Stereotypes pronounced with time pressure & distraction (Bless et al 1996)
    - ★ "In a crowded immigration court, 7 minutes to decide a family's future" (Wash Post 2/2/14)

Early Predictability of Asylum Decisions

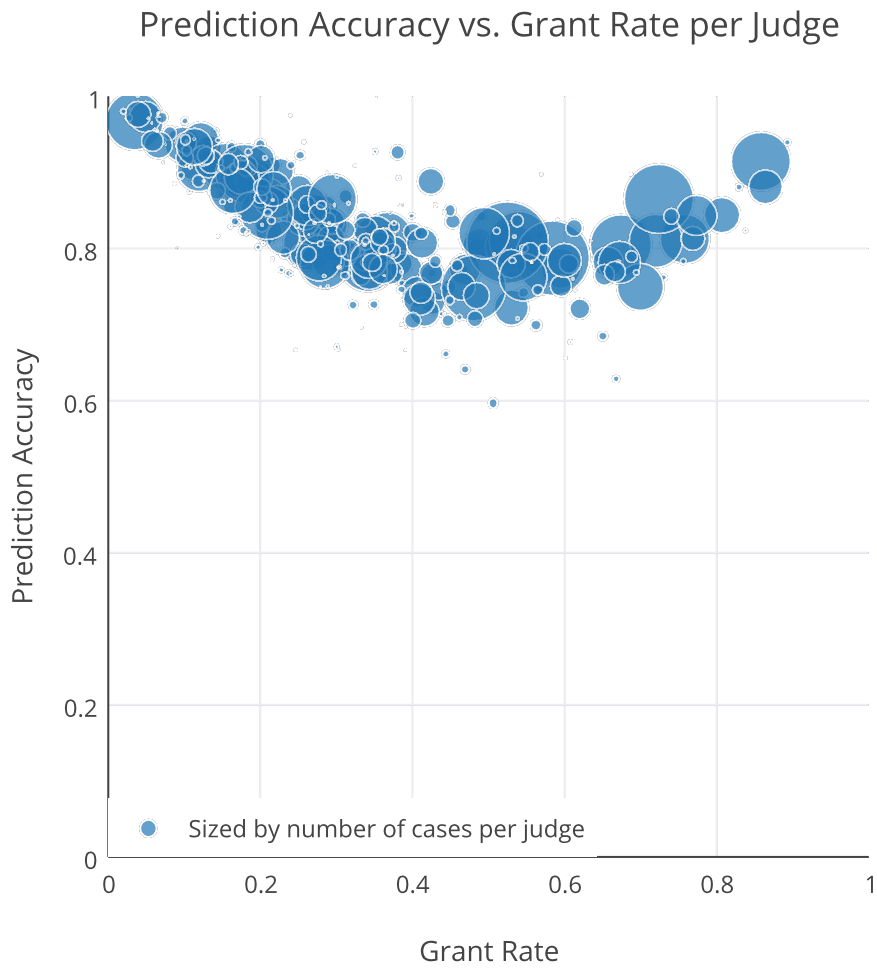
Model	Accuracy	ROC AUC
Judge ID	0.71	0.74
Judge ID & Nationality	0.76	0.82
Judge ID & Opening Date	0.73	0.77
Judge ID & Nationality & Opening Date	0.78	0.84
Full model at case completion	0.82	0.88

Variation over time has little additional impact on the outcome of adjuciations.

Dataset includes 70 additional features about the hearings

- Date, nationality, judge, and court motivate random forest

Inter-judge disparities in predictions vs. prediction accuracy

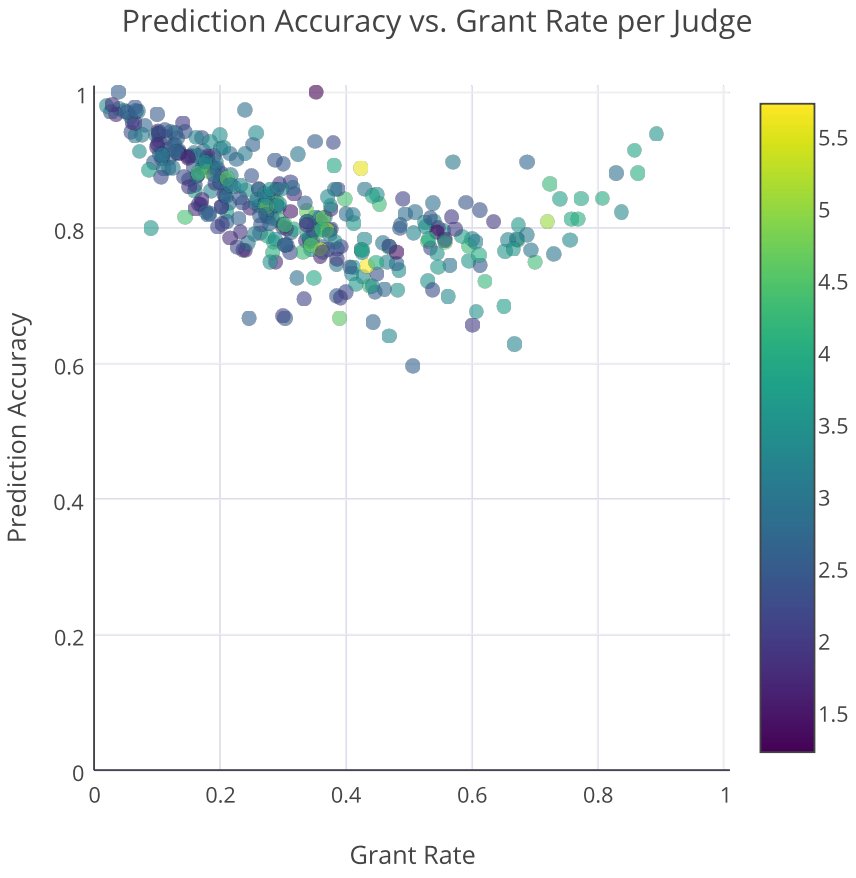


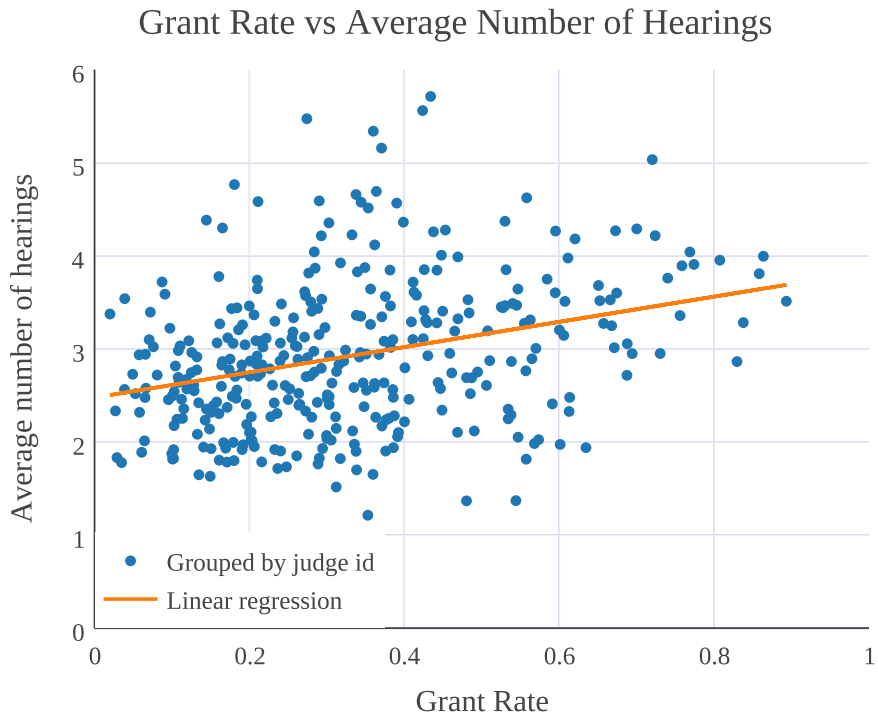
Judges with high and low grant rates are more predictable

Are less predictable judges simply flipping a coin?

Early Predictability of Asylum Decisions

Hearing sessions are greater for less predictable judges





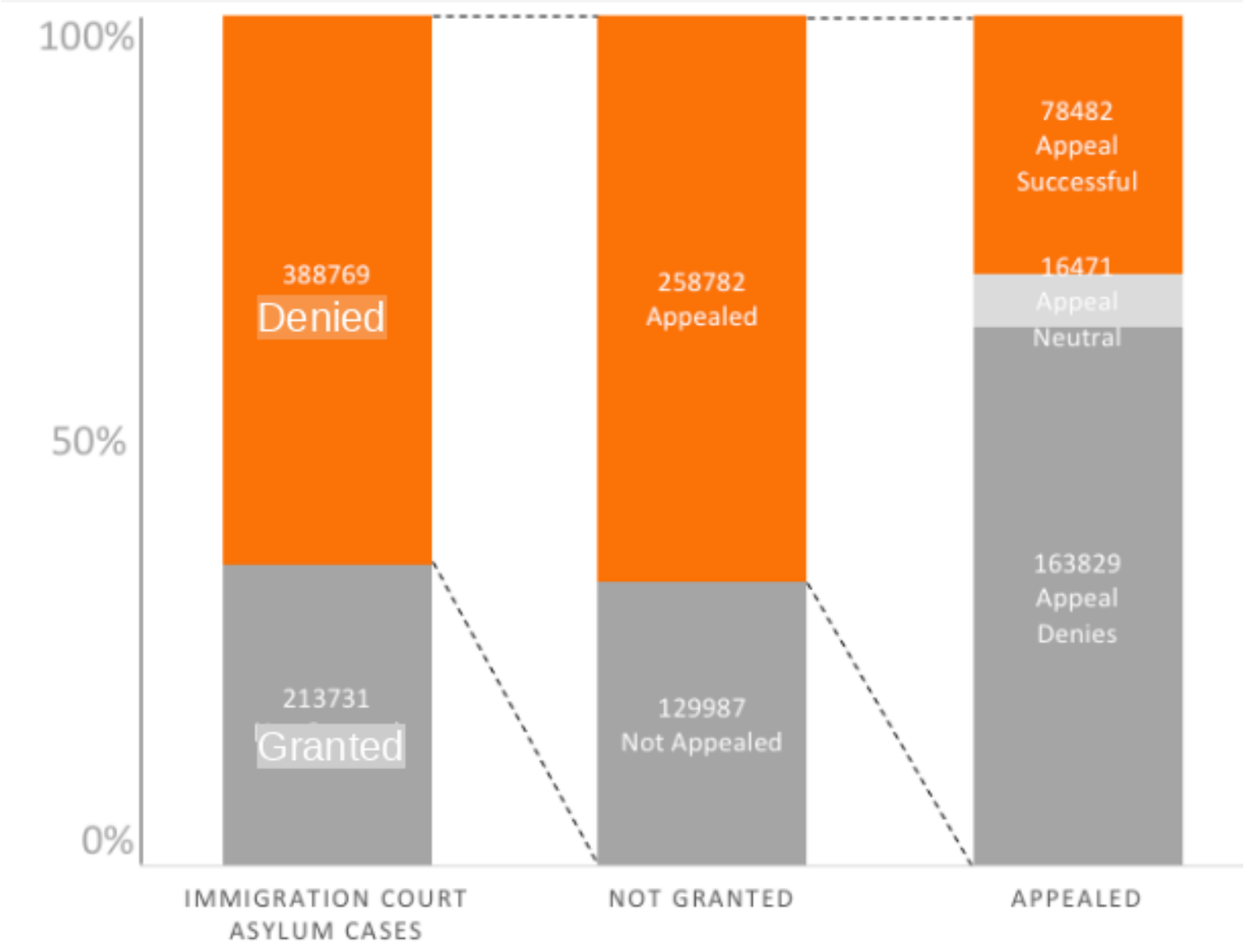
Are judges with low or high grant rates using snap judgments?

Hearing sessions are greater for judges with higher grant rates

Predictability of Asylum Appeals?

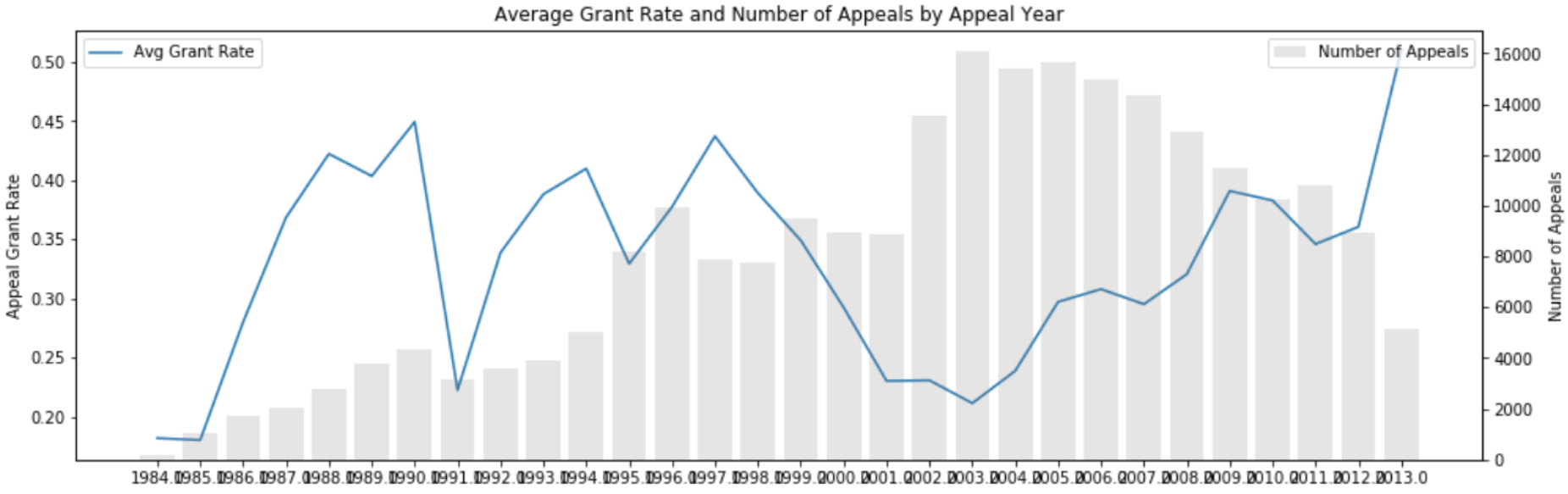
- We have shown evidence of early predictability that varies by judge
  - We see evidence of behavioral anomalies (more later)
- If making mistakes, we might expect judge identity to predict appeal
  - information that might be useful for an applicant considering an appeal
  - or a judge considering his/her decision

Asylum Cases and Appeals



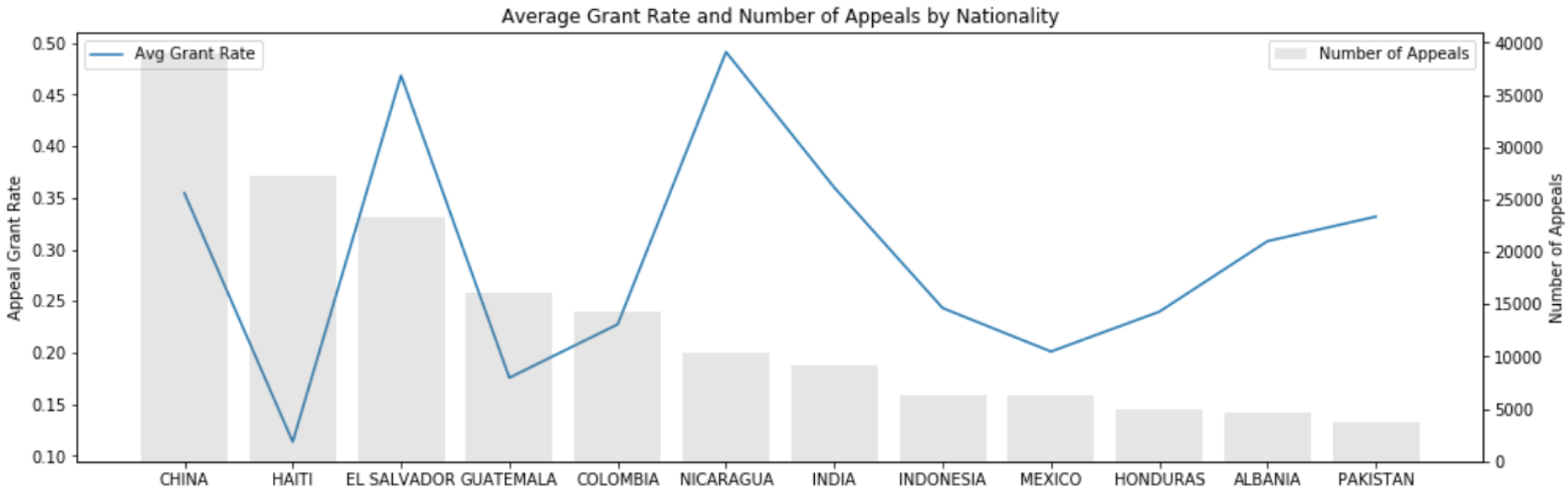
Predict final grant or deny

Appeal Grant Rate by Year



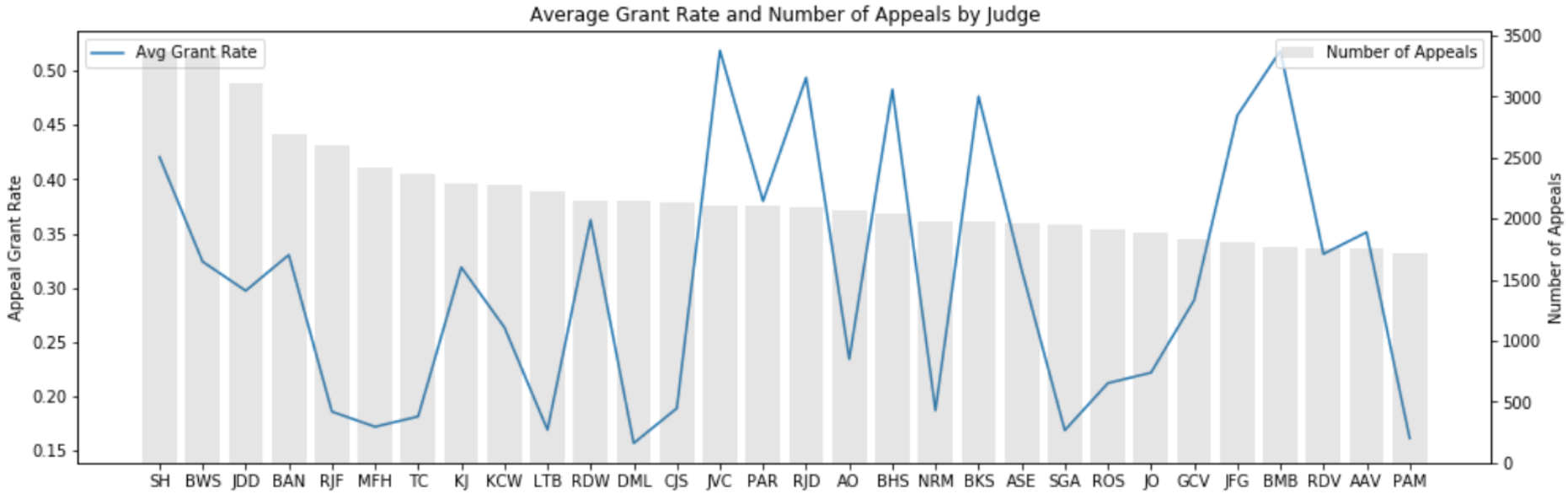
Grant rate varies over time

Appeal Grant Variability, by Nationality



Grant rate varies by nationality

Appeal Grant Variability, by Judge



Reversal rate varies by lower court judge

A successful appeal of asylum denial means original judge made a mistake.

Random Forest Performance

Model	Accuracy	ROC AUC
Full Model	**0.792**	**0.840**
Nat + Judge + Year	0.741	0.765
Nat + Judge	0.704	0.701
Nationality Only	0.683	0.665
Judge Only	0.675	0.625

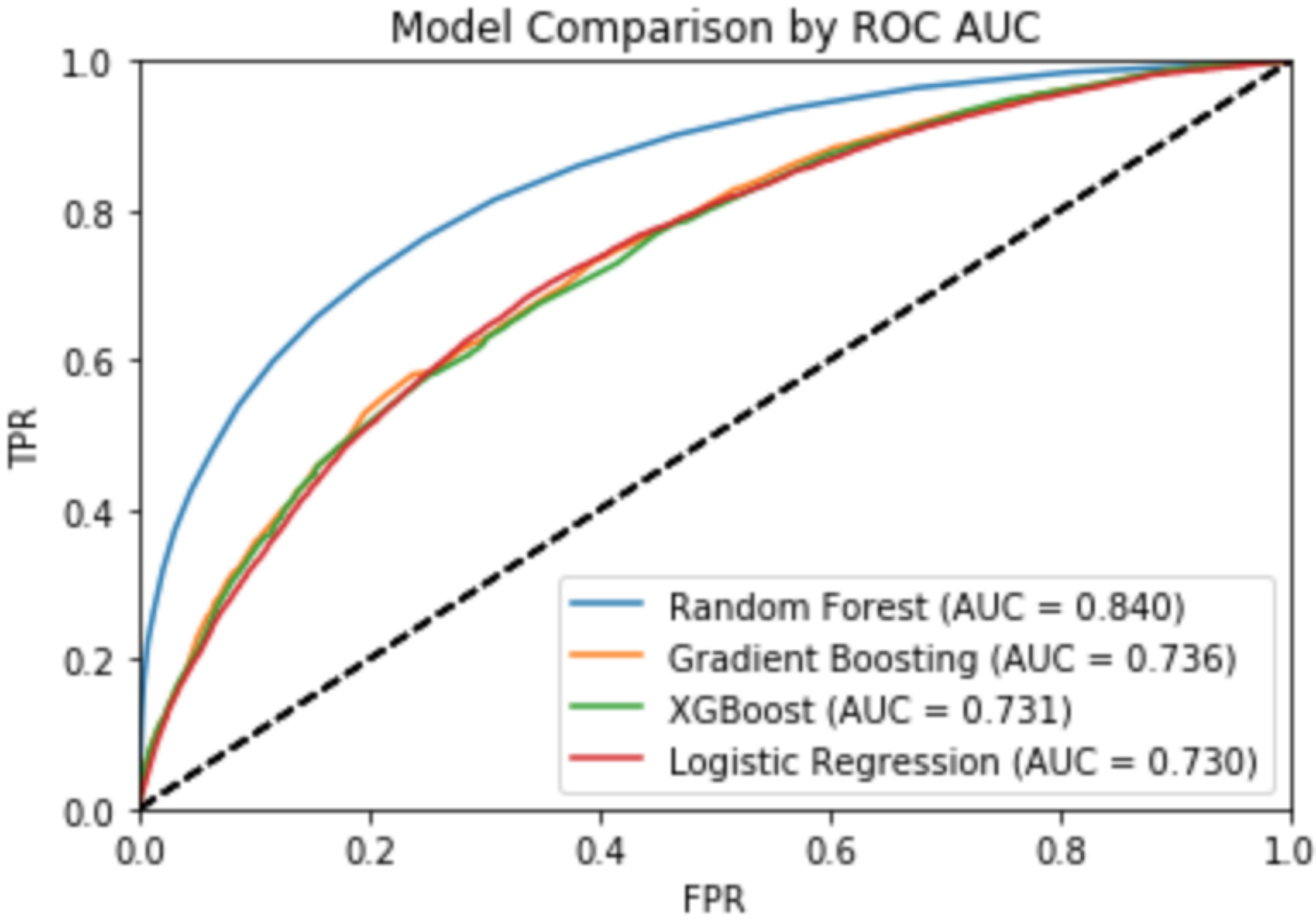
Prediction accuracy largely driven by identity of lower court judge

Confusion matrix without judge identity

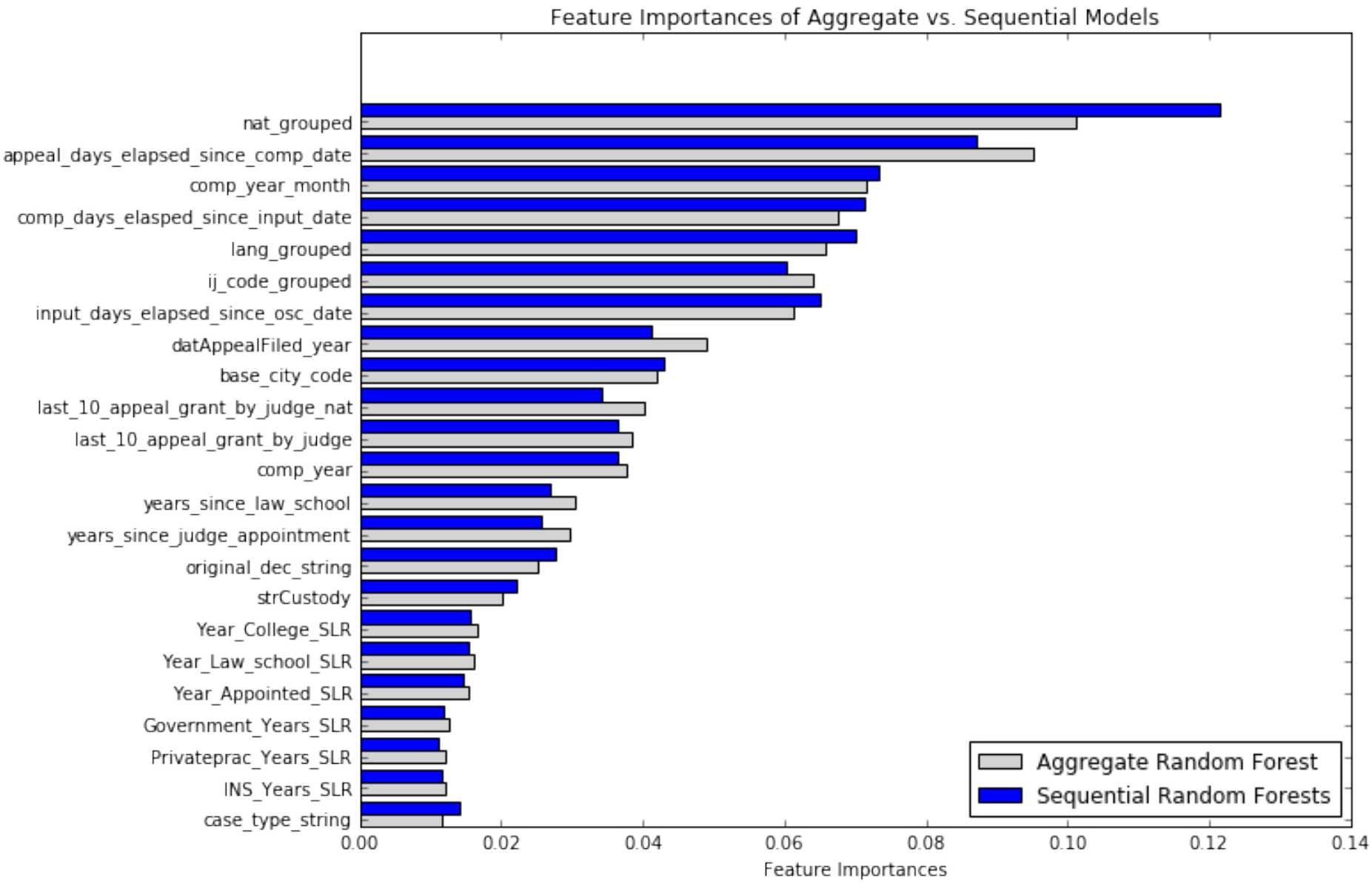
		Predicted	
		Denied	Granted
True	Denied	195,223	65,798
	Granted	73,269	104,406

Accuracy = 68.3%, F1 = 0.60

Asylum Appeal: RF vs. Other Models



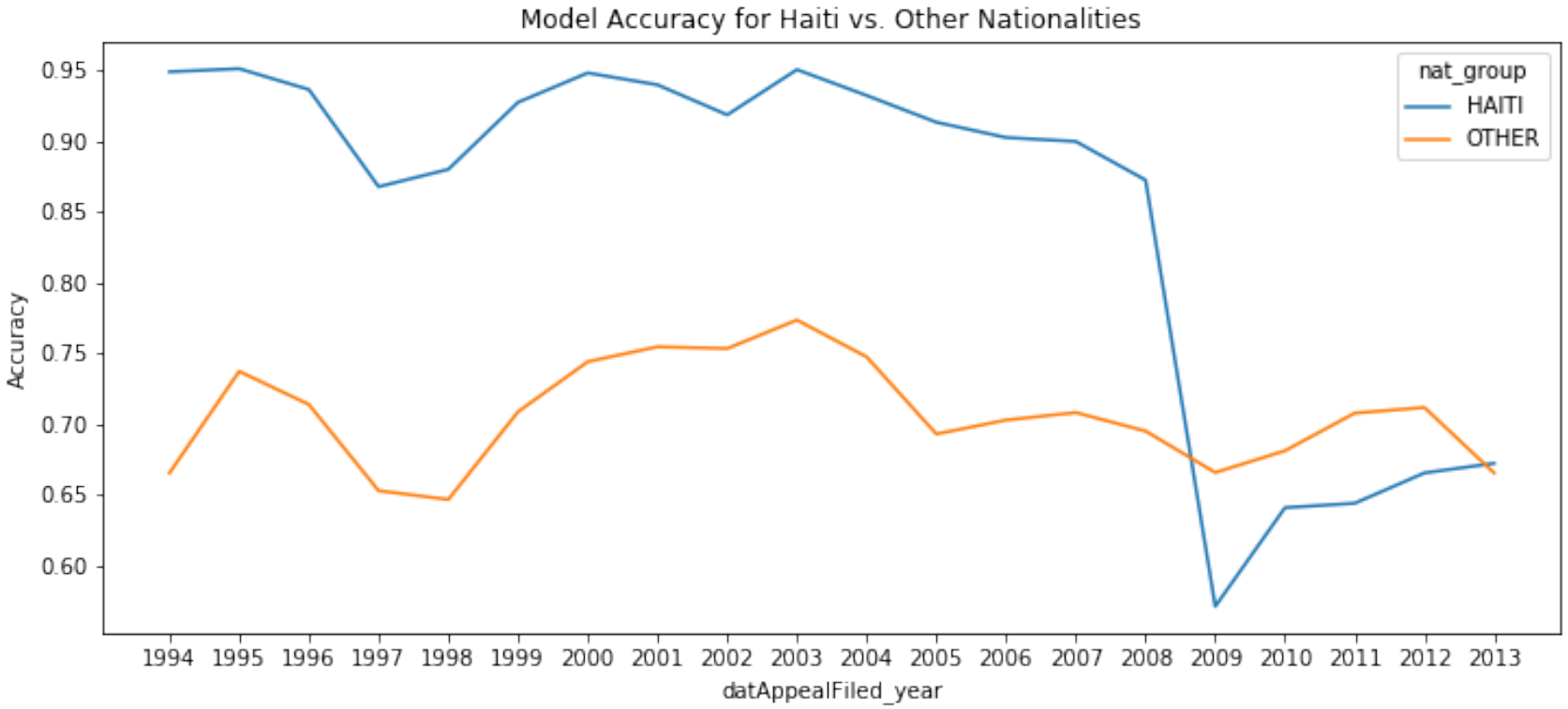
Asylum Appeal: Feature Importance



Asylum Appeal: Feature Importance

Feature Group	
Time Horizon Features	0.377804
Judge Features	0.277066
Respondent	0.177945
Trend Features	0.074494
Proceeding Features	0.060490
Location Features	0.042636

Limits of Model: Haiti Earthquake



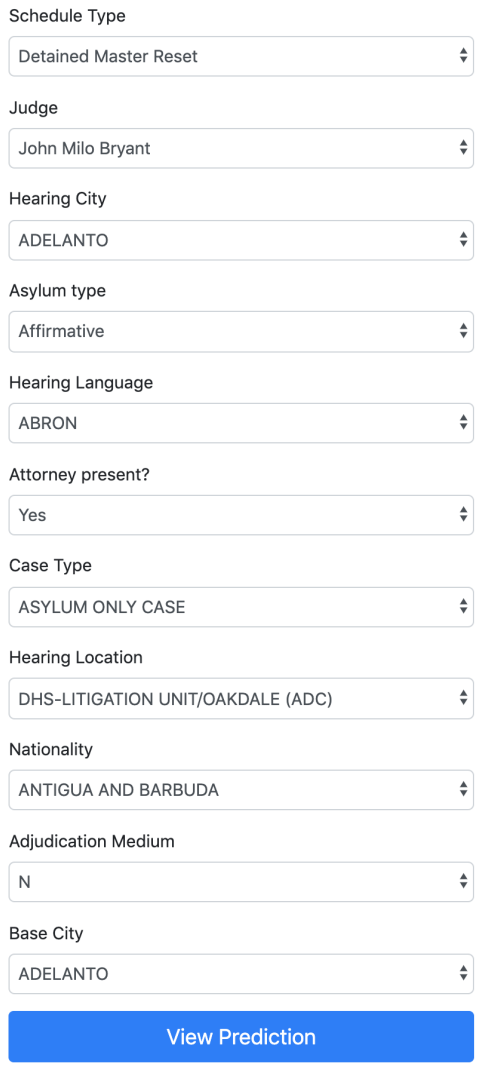
Many Haitian applicants denied asylum pre-2009, but granted after 2010.

Appellants have Rational Expectations

- Non-appeals likely have private information:
  - **6.4%** would have been successful in their appeal
  - vs. **32.4%** grant rate for the population that did appeal
- Among cases predicted to be successful in appeal, **84.3%** did appeal.
- Decision support tool for applicants may further reduce uncertainty

App (Screenshot)

Prediction App (Beta): `https://floating-lake-11821.herokuapp.com/`

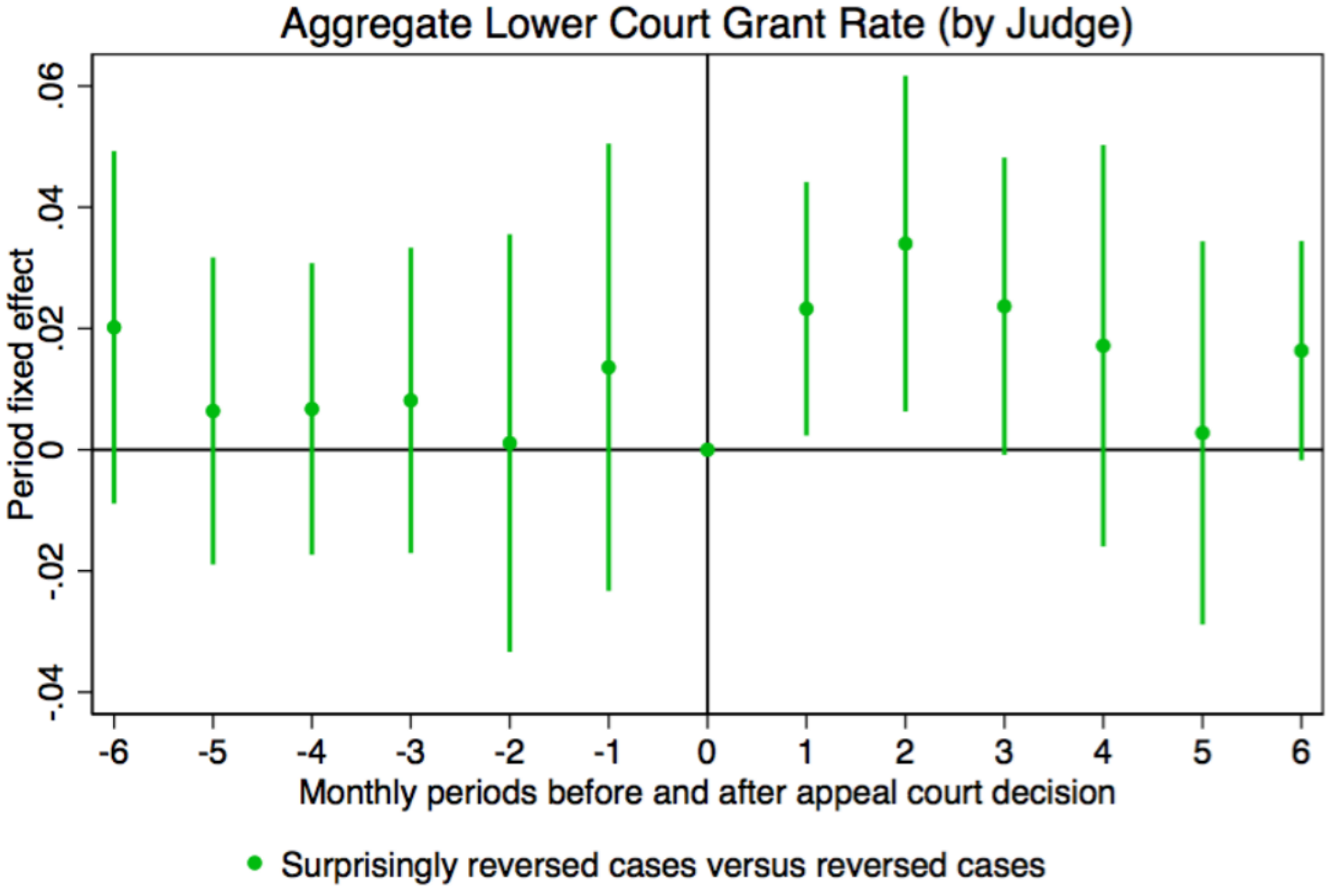


Signpost

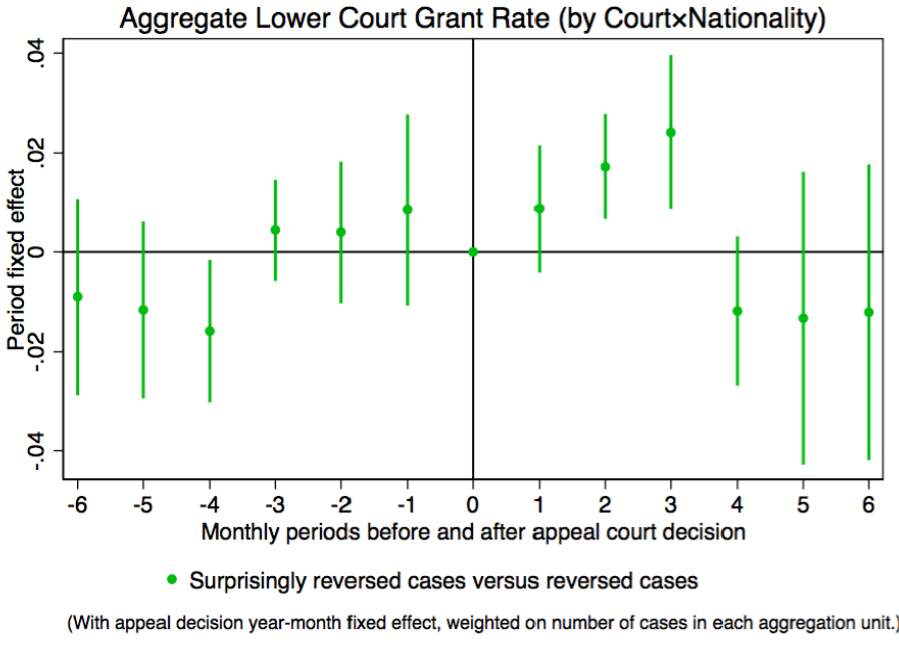
- We see judge identity playing a significant role in prediction of appeal
  - consistent with measurement of mistakes
  - early predictability
  - influence of behavioral factors
- Do we see judicial variation in responsiveness to reversal?
  - evaluate as event study (and ML)
- Do inattentive judges essentially randomize in implicit risk ranking?
  - evaluate with marginal treatment effects

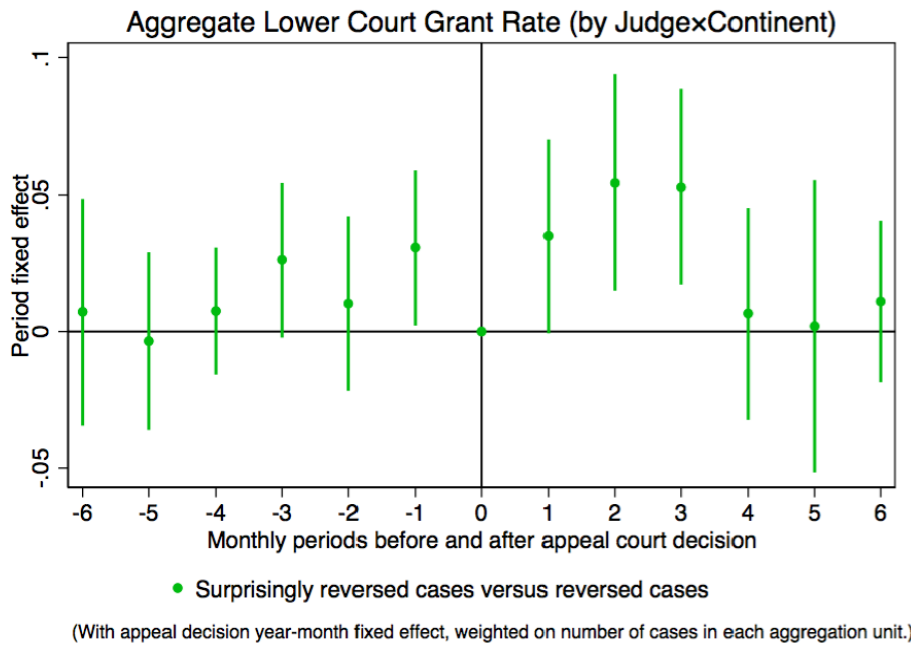
Effect of "Surprise" Appeal Rulings

Within-judge change in grant rates before/after "surprising" reversals (model predicts affirm), relative to unsurprising reversals (model predicts reverse):



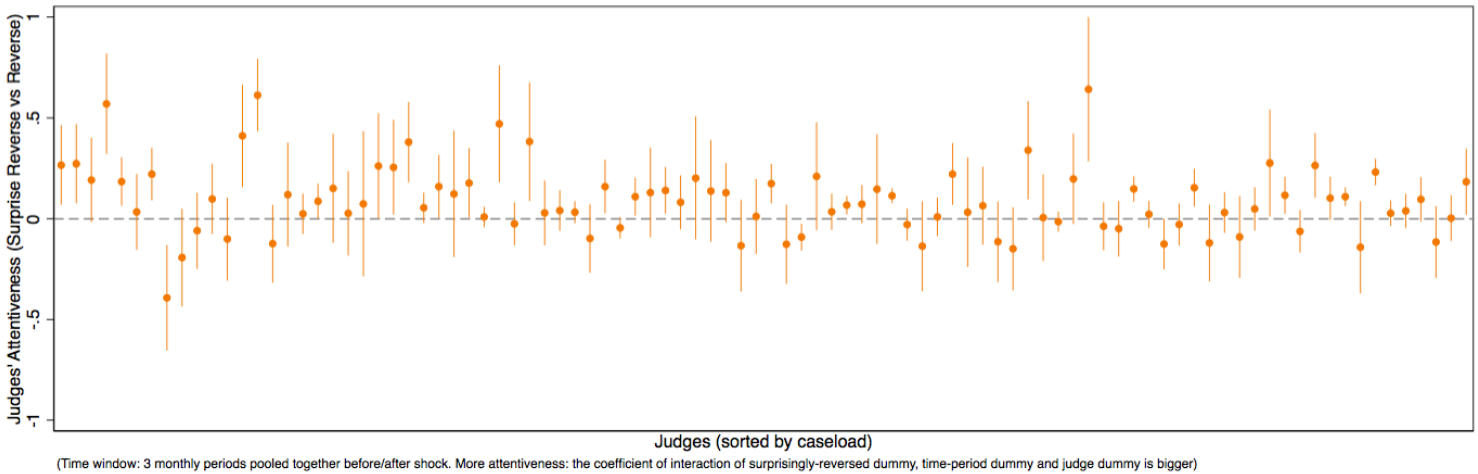
Surprise Rulings – By Nationality





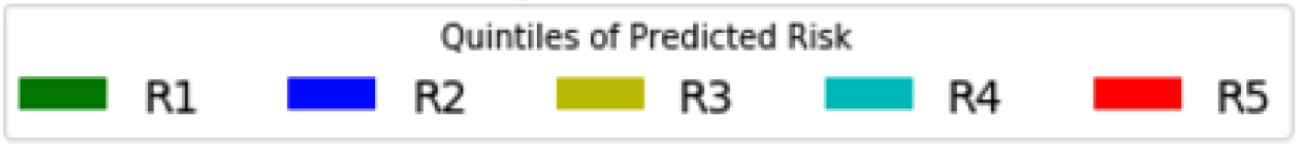
judge×nationality is too sparse for short time frame.

Judges Vary in Attention

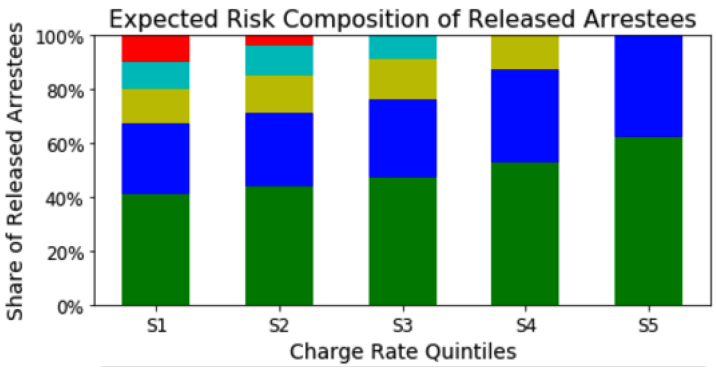


Do implicit rankings by judges differ by attentiveness?

How the judges rank the risk of asylees is unobserved. But, we can assess their implicit risk ranking by **comparing the distribution of outcomes of the asylees denied by the (randomly assigned) "strict" and the "lenient" judges.**

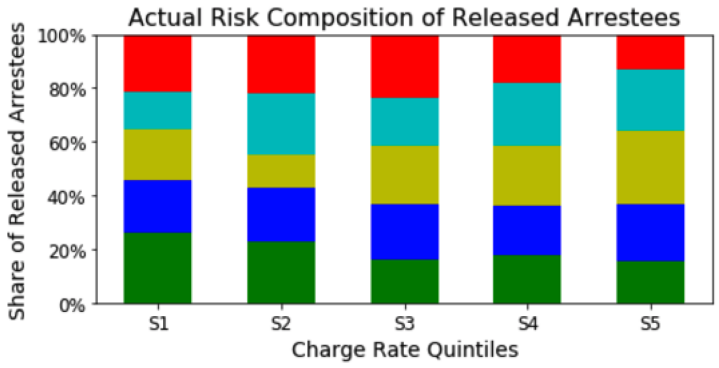


Robot Prosecutors



- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.

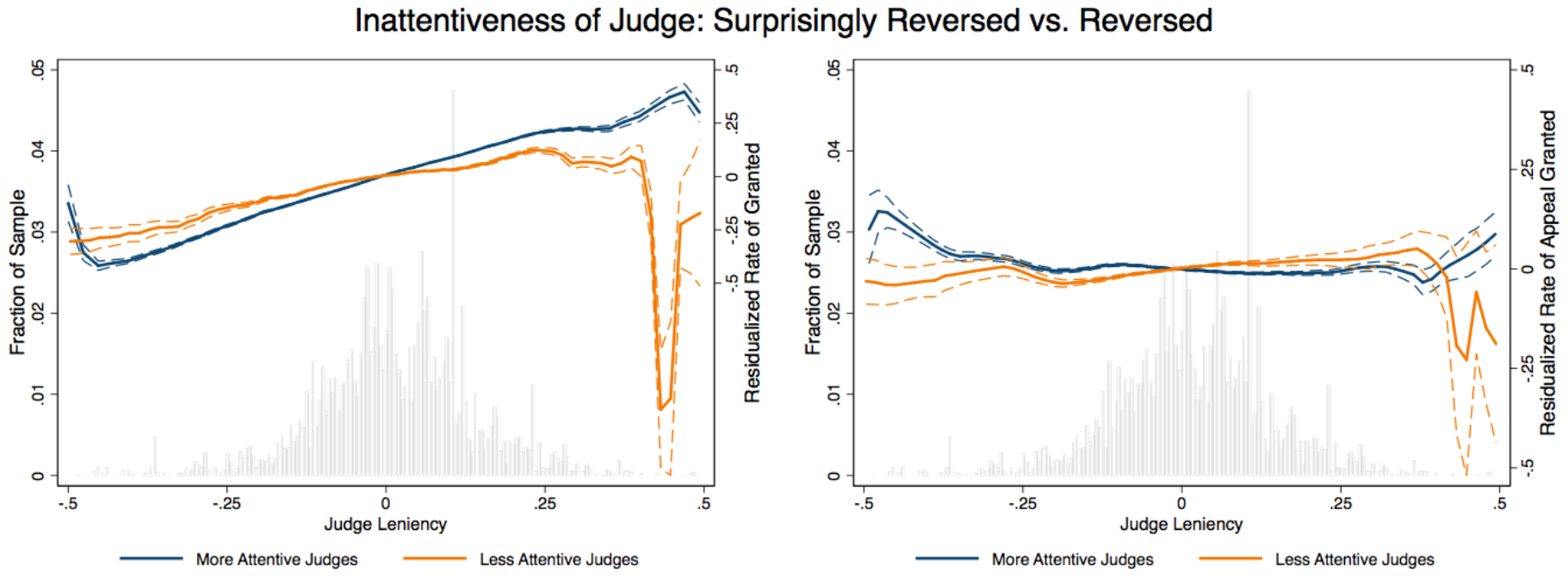
Human Prosecutors



- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.

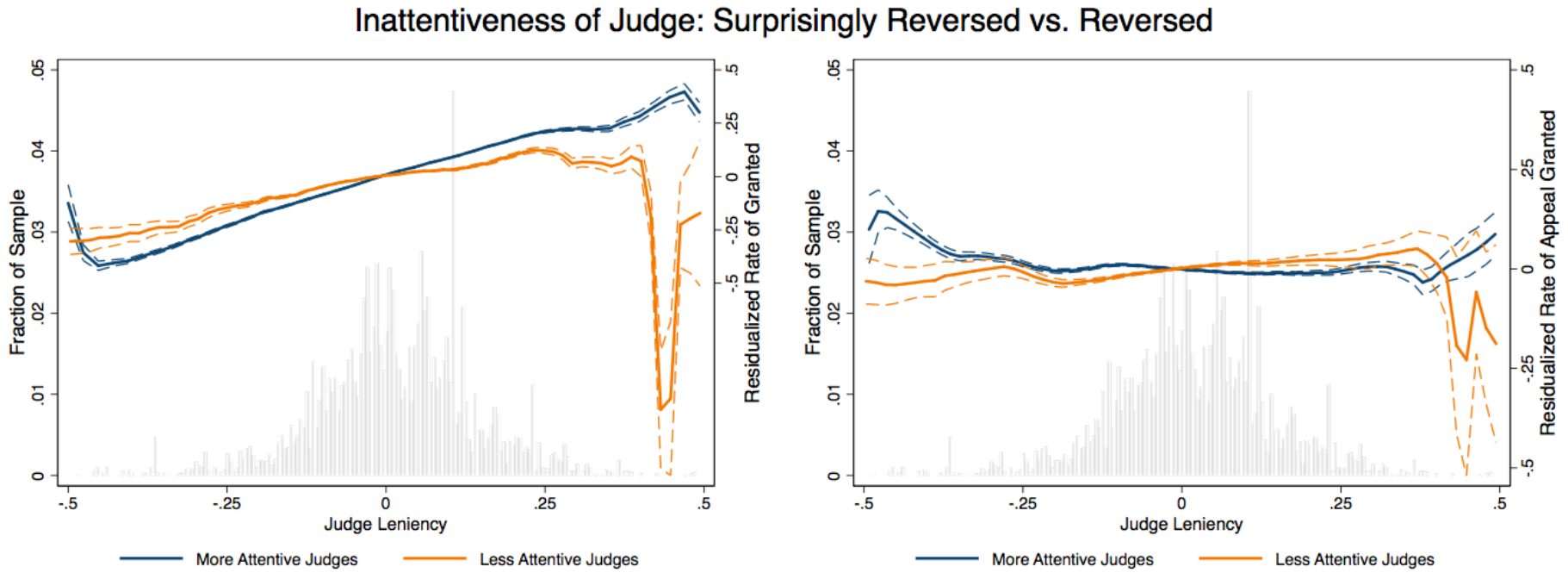
- Prosecutors:
  - charge/release decisions: 88% accuracy
  - defendant re-arrest: 77% accuracy

Distribution of appeal granted also similar..



(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly reversed dummy and time-period dummy is bigger)

But attentive judges rank asylees more like the appeal board



(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly reversed dummy and time-period dummy is bigger)

Policy

- Fragility of U.S. asylum courts
  - "7 minutes to decide a family's future" (Wash Post 2/2/14)

Judicial inattention

  - Early predictability
  - Behavioral bias
  - Response to appellate review
  - Implicit risk ranking

- Can we nudge judges to pay more attention?

Observational evidence suggests yes

- Experience, Incentives, Lawyers appear to nudge judges

Conclusion

Pathways to decrease judicial inattention

- lawyers, incentives, experience
- judicial education
- apps?
  - **RCT** (planning phases to evaluate app)
    - provide subset of court administrators, attorneys, or judges
    - link to point-in-time decisions in administrative data
- Assess effects on asylum applications, grant rates, appeals, disparities, etc.