

The background is a deep blue gradient with a subtle pattern of white dots, resembling a starry sky. On the left side, there are several concentric circles and a large circular scale with degree markings from 140 to 260. Some of the circles have arrows indicating a clockwise direction. The overall aesthetic is technical and futuristic.

IN SEARCH OF THE UNINSURED

VINCE VELOCCI

OCTOBER 28, 2016

MOTIVATION



- There are organizations (Enroll America) whose goal is to reach out to those without health insurance to discuss available options
- Enroll America may work with data science consultancies for insight on how to reach out to those not likely to have health insurance

DATA SCIENCE TO THE RESCUE! HOPEFULLY...

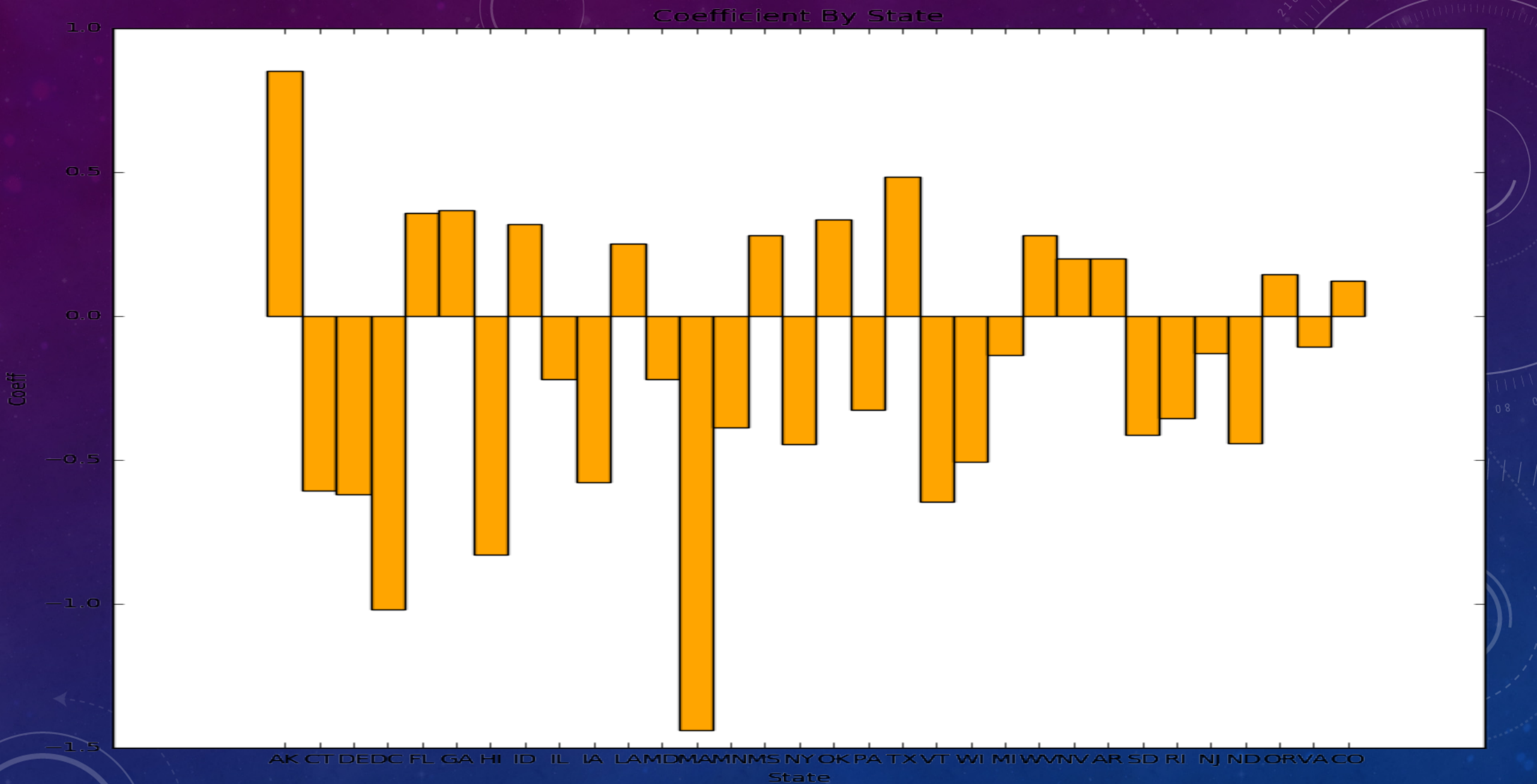
- Data Source: Data on individuals in the country (health ins + lots of demographic features) obtained from Civis Analytics
- You name it: income, age, education, military, state, marital, type of work, vehicles, citizenship
- Over 272,000 labeled data PLUS 188,000 of unlabeled data
- Lots of categories → over 1000 features with dummies
- Can we trust this data? (Self Reporting)

MODELING

- Need to cut down on features:
 - Logistic with penalty
 - PCA
- Tried Random Forest, KNN, Logistic Regression
- BEST MODEL: **Logistic Regression with L1 Reg**
 - Mean F1 Score = 0.48 ☹️ , (5-fold cross val) AUC: 0.84
 - Recall around 0.6 for threshold of 20% ☹️
 - Model then trained on full dataset and used to make predictions on unlabeled dataset. Result?

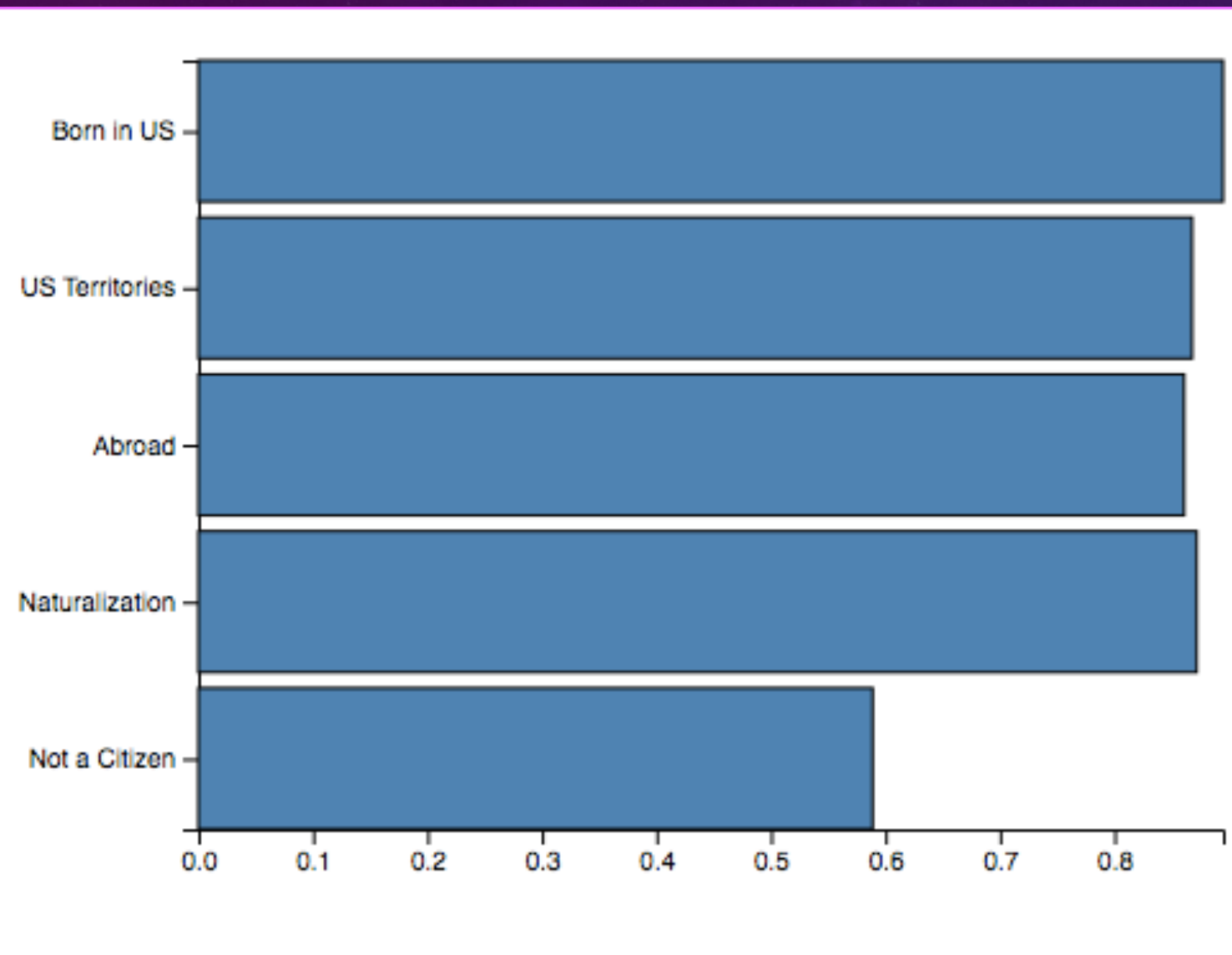
A FEW INSIGHTS?

- Many significant features according to this model
 - Age, # of children in household (negative -> chance of being uninsured down with age & # of children)
 - Not a US Citizen -> More likely uninsured than US Citizens (BUT see last slide)
 - Gov employees & those in their own practice more likely to have ins (negative coefficients)
 - Industries: Fishing, Hunting, Trapping most likely to be uninsured
 - Industries: Aerospace Parts Manuf and Rail Transportation LEAST likely to be uninsured
 - US States: Largest negative coefficient? MA Largest positive coefficient? AK
 - (Quick Google Search shows MA was the state with the largest percentage insured)



- From Left: AK, CT, DE, DC, FL, GA, HI, ID, IL, IA, LA, MD, MA, MN, MS, NY, OK, PA, TX, VT, WI, MI, WV, NV, AR, SD, RI, NJ, ND, OR, VA, CO

MEAN PROBABILITY OF INDIVIDUAL BEING UNINSURED BY CITIZENSHIP FOR UNLABELED SET



The background is a deep blue gradient with a subtle pattern of white dots, resembling a starry sky. Overlaid on this are several faint, white, concentric circles and arcs. A prominent circular scale is visible on the left side, with numerical markings ranging from 140 to 260 in increments of 10. Other smaller circular elements with arrows are scattered across the frame.

THANK YOU!

IN SEARCH OF THE UNINSURED

VINCE VELOCCI

OCTOBER 28, 2016