How can data help us heal communities at high risk for suicide?

Introduction

Suicide is a major public health concern. Over 41,000 people die by suicide each year in the United States; it is the 10th leading cause of death overall. Suicide is tragic. But it is often preventable. Knowing the risk factors for suicide and who is at risk can help reduce the suicide rate. There are approximately 300 deaths per year on railroads, usually a result of traffic accidents or suicides.

We have focused our efforts to study the various data sets available to derive insights that can help the authorities plan and implement policies to aid the high risk communities.
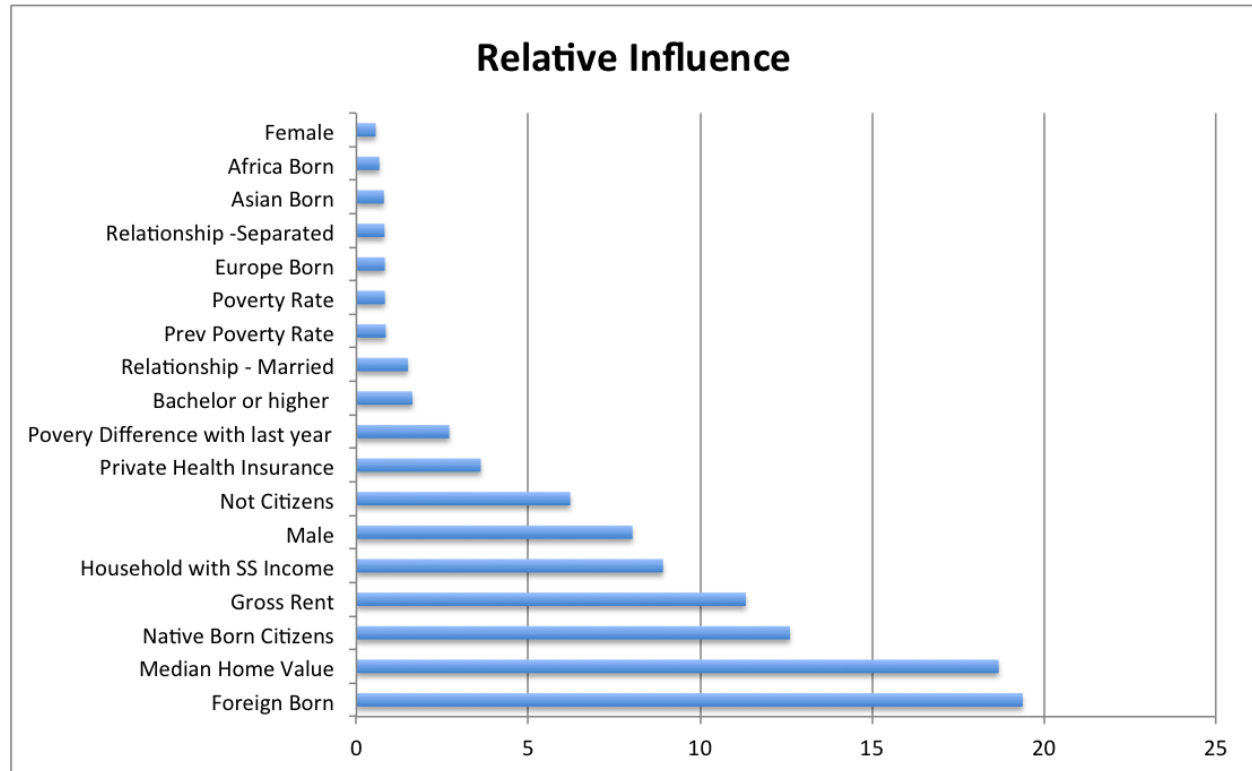
Methodology

1. Identify the prompt data, create data dictionary
2. Research railway suicides (trends, causes, history) - see additional references
3. Identify additional sources of data
4. Retrieve data at the county level; merge with the prompt data
5. Build base models - use R to find coefficients
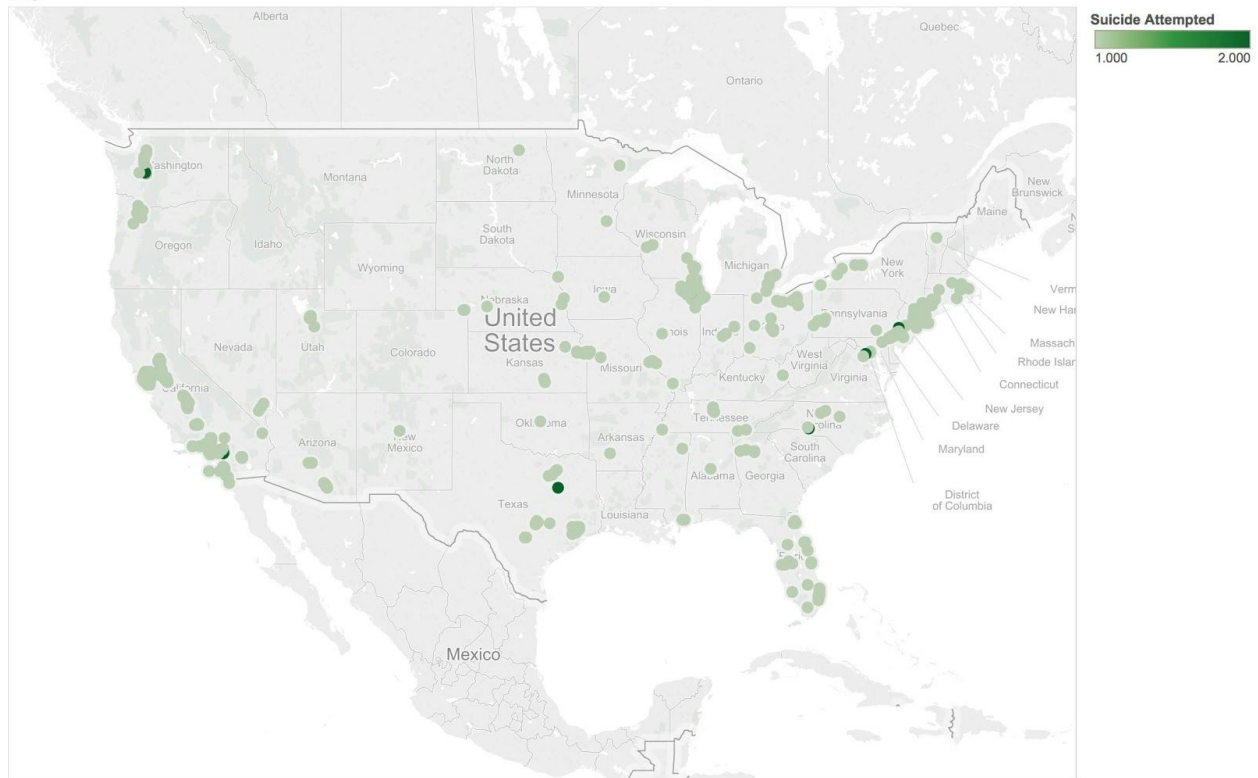
Data Processing

- The prompt data has railway incidents dating back 1975. However, the Federal Railway Authority (FRA) did not require records of suicides until 2011. Therefore, we concentrated all analysis on data from 2011 forward.

- We considered at the advice of the Judge that he would like to see data at the city/metropolitan level. All the data we had collected up to this point was county, but we thought to change paths anyway. In the end, we reverted back to county level analysis. There is very minimal data available to us on railway suicides; with only one or two suicides in any one city over 5 years of data, the results would not be significant or insightful.

Analysis & Insights

Relative Influence

Feature	
Female	
Africa Born	
Asian Born	
Relationship -Separated	
Europe Born	
Poverty Rate	
Prev Poverty Rate	
Relationship - Married	
Bachelor or higher	
Povery Difference with last year	
Private Health Insurance	
Not Citizens	
Male	
Household with SS Income	
Gross Rent	
Native Born Citizens	
Median Home Value	
Foreign Born	

Features of Importance

April 24, 2016
Bayes Hackathon 2016
Data McDataface

City

Map based on Longitude and Latitude. Color shows sum of Suicide Attempted. The data is filtered on Suicide Attempted and Count Not Citizen as an attribute. The Suicide Attempted filter keeps 1. The Count Not Citizen as an attribute filter ranges from 10,446 to 1,843,693.

References

- http://transweb.sjsu.edu/PDFs/research/1129-2-preventing-suicide-on-US-rail-systems.pdf

- http://railwaysuicideprevention.com/railway-fatalities/overview.html#suicide

- https://drive.google.com/file/d/0B-velHZJGPpOVF9ueHJsNWl0ZUk/view

- http://safetydata.fra.dot.gov/OfficeofSafety/default.aspx

- http://oli.org/about-us/news/statistics

- http://safetydata.fra.dot.gov/OfficeofSafety/Default.aspx

- https://slack-files.com/files-pri-safe/T03V9PS2W-F0ZGP3WHX/fraguideforpreparingaccincre portspubmay2011__1_.pdf?c=1460357663-0a36f0ea3c2c6fd1b674cf33076f9b29681047ab

- http://www.lse.ac.uk/geographyAndEnvironment/whosWho/profiles/neumayer/pdf/Article%20i n%20Cross-Cultural%20Research.pdf

- http://bjp.rcpsych.org/content/187/1/49

- http://www.stat.columbia.edu/~madigan/DM08/descriptive.ppt.pdf

- https://safetydata.fra.dot.gov/officeofsafety/Documents/Railroad%20Safety%20Data%20Frequently%20Asked%20Questions.pdf?V=9
- http://www.ntsb.gov/news/events/Documents/2015_trespassing_FRM_Pres10GabreeVolpe1(2).pdf
- http://faculty.wcas.northwestern.edu/~ipsavage/211.pdf
- https://www.nimh.nih.gov/health/topics/suicide-prevention/index.shtml

Our data team

1. Angela Gunn
2. Hetal Chandaria
3. Karthik Chepudira
4. Prabhakar Gundugola
5. Deepesh Chaudri

GitHub Repository

Folder 'analysis'

1. BayesHack.R : R code for logistic regression and GBM
2. BayesHack2016_Team_Guardians-FeatureEngineering-Prabhakar.ipynb
3. BayesHack2016_Team_Guardians-Prabhakar.ipynb

Folder 'data'

1. ***rail_demo_pop_casaulties_2011.csv*** : final cleaned up dataset for analysis
2. county_pop_data_metrics_2011.csv
3. county_pop_data_metrics_2012.csv
4. county_pop_data_metrics_2013.csv
5. county_pop_data_metrics_2014.csv
6. population_race_2011.csv
7. Population_race_2014.csv
8. Poverty_median_income_2012.csv
9. Poverty_states.csv
10. unemployemnt_county_2012.csv
11. Population_race_2012.csv

12. Poverty_median_income_2010.csv
13. Poverty_median_income_2013.csv
14. Unemployemnt_county_2010.csv
15. unemployemnt_county_2013.csv
16. Population_race_2013.csv
17. Poverty_median_income_2011.csv
18. Poverty_median_income_2014.csv
19. Unemployemnt_county_2011.csv
20. unemployemnt_county_2014.csv

Future research topics

1. Proximity Analysis of the incidents to the following [HIFLD geospatial datasets](#).

- Railroad Bridges
- Road and Railroad Tunnels
- Railroads
- Intermodal Terminal Facilities
- Weigh in Motion Stations
- Fixed Guideway Transit Links
- Automatic Traffic Recorder (ATR) Stations
- Fixed Guideway Transit Stations
- Amtrak Stations
- Rail Junctions
- National Bridge Inventory (NBI) Bridges