DC Crash Data

Melissa Cirtain
Lead Technologist
Booz Allen Hamilton
SIG Digital
June 25, 2018
cirtain_melissa@bah.com

40,327

US Traffic Fatalities in 2016

Costs

$416B in 2016

Principal cause of death in the US through age 23

Top-ten cause of death in the US across population

Can historical crash data predict major injuries or fatalities?

Data

SOURCE: DC Metropolitan Police Crash Data Management System

- 186K rows
- 49 columns
- Over 10 years of data

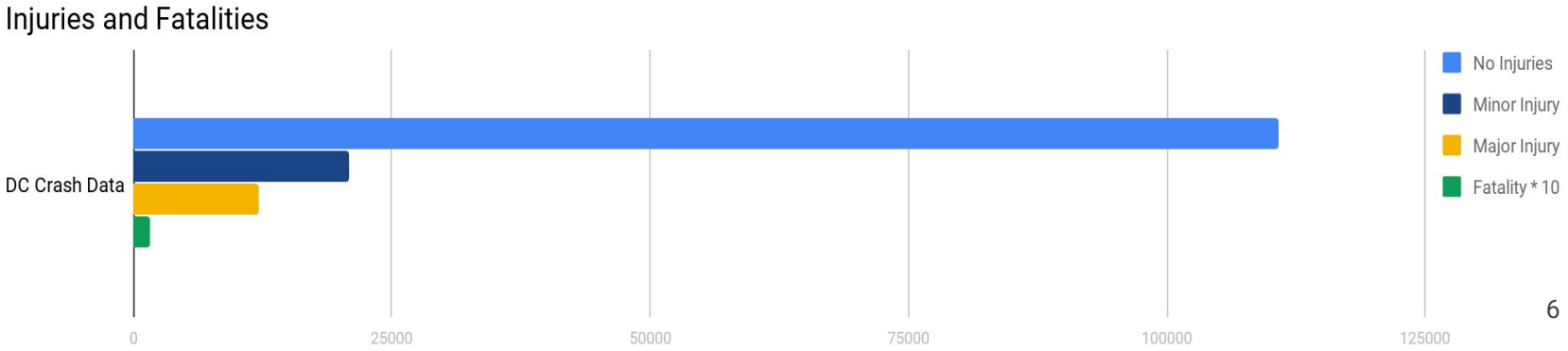
ROWS: Each row represents one crash and contains details for each record.

- date
- location
- vehicle types
- vehicle counts
- bicyclists or pedestrian involvement
- street ID
- speeding involvement

Data

<u>Unbalanced data</u> - The values we most need to predict, fatal accidents, are very rare in the data.

- **~ 10% crashes involve major injury**
- **Less than 0.1% crashes involve fatality**



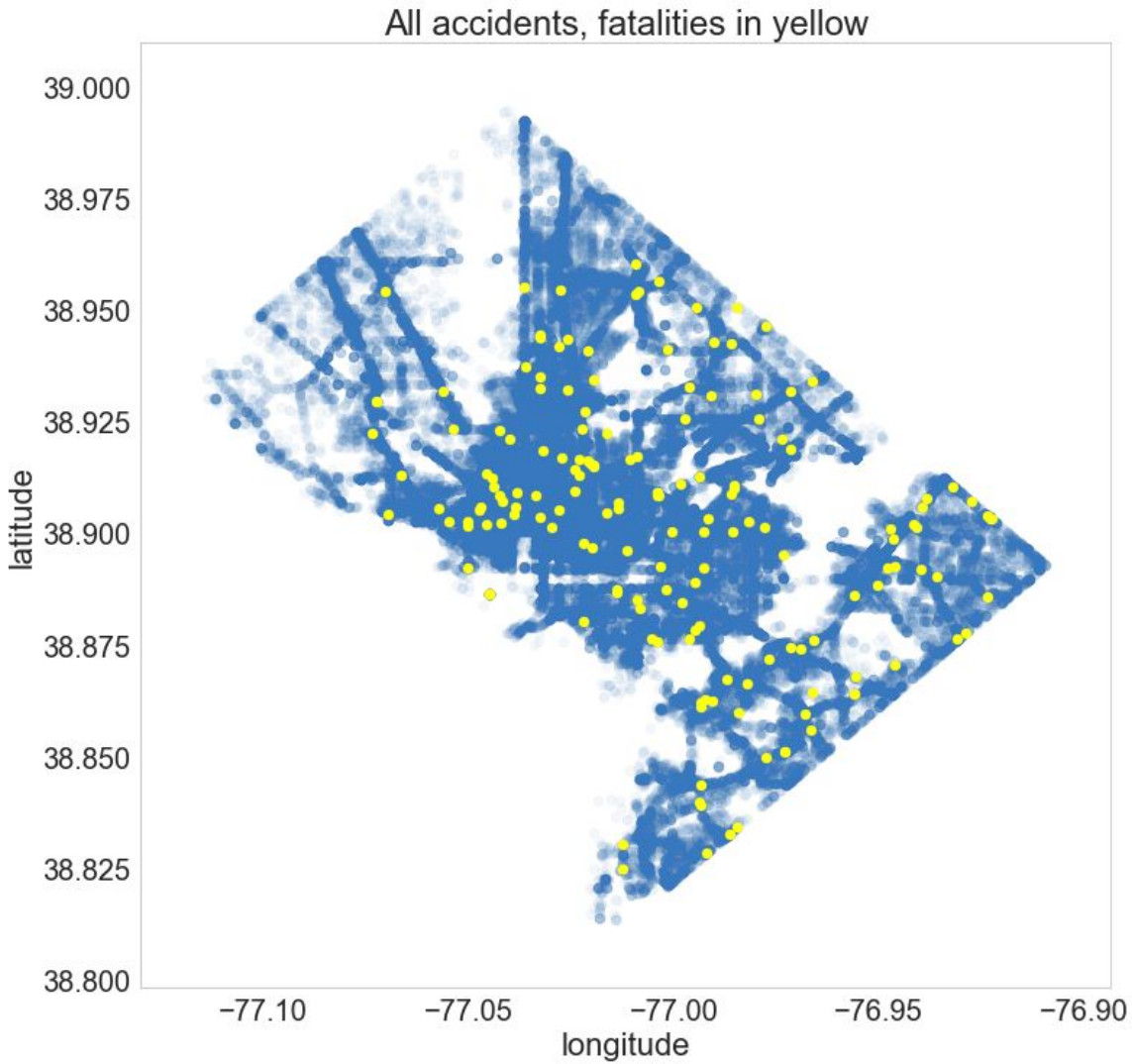
Exploratory Analysis

Removed rows missing fields, reducing ~20% of data

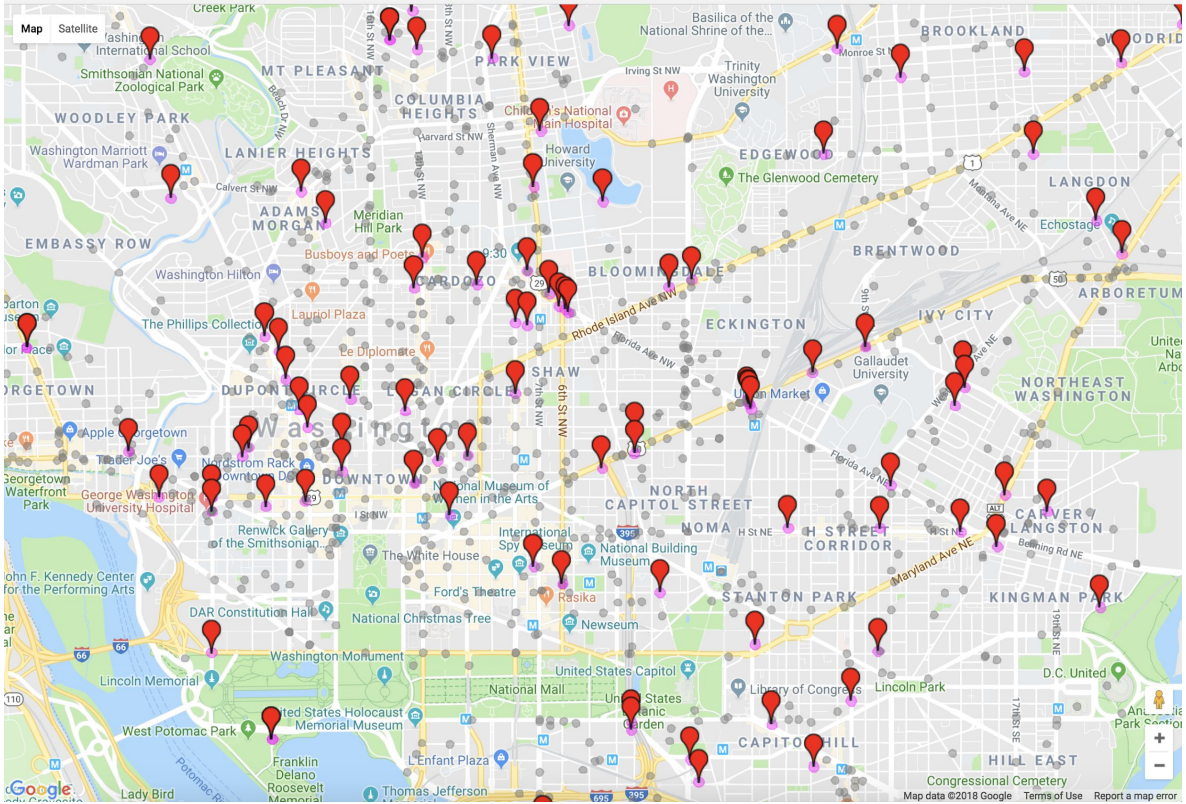
Weak correlations revealed through analysis

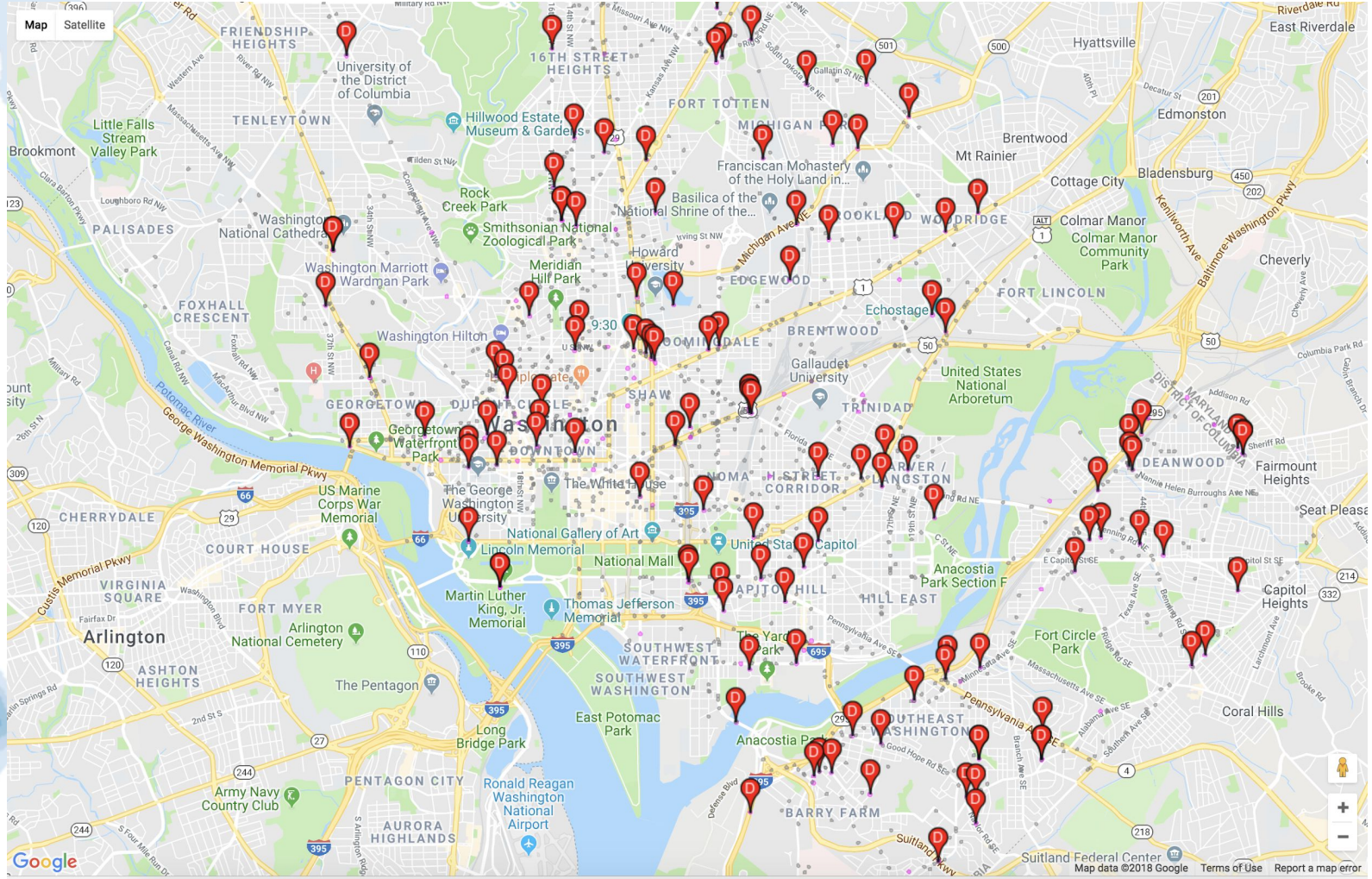
Leverage latitude and longitude in exploratory data analysis

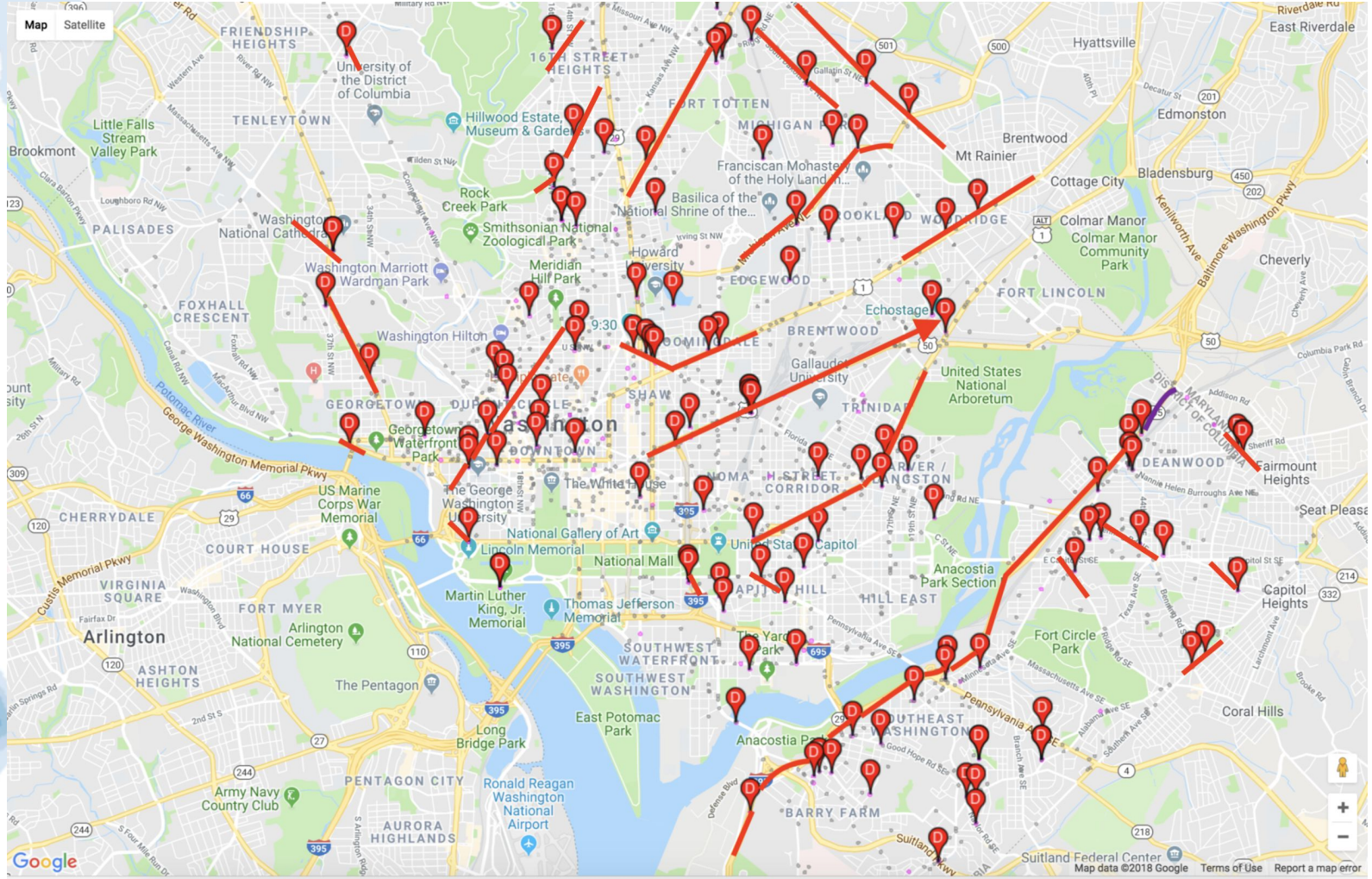
- Plot each crash on a grid
- Overlay fatalities to visualize geographic distribution


All accidents, fatalities in yellow
39.000
38.975
38.950
38.925
38.900
38.875
38.850
38.825
38.800
latitude
−77.10
−77.05
−77.00
−76.95
−76.90
longitude

Exploratory Analysis




Map
Satellite
FRIENDSHIP HEIGHTS
University of the District of Columbia
16TH STREET HEIGHTS
FORT TOTTEN
Hyattsville
East Riverdale
Little Falls Stream Valley Park
Brookmont
TENLEYTOWN
Hillwood Estate, Museum & Gardens
Franciscan Monastery of the Holy Land in...
Brentwood
Mt Rainier
Edmonston
Cottage City
Bladensburg
Rock Creek Park
Basilica of the National Shrine of the...
Smithsonian National Zoological Park
Washington National Cathedral
PALISADES
Washington Marriott Wardman Park
Meridian Hill Park
Howard University
EDGEWOOD
Colmar Manor
Colmar Manor Community Park
Cheverly
FOXHALL CRESCENT
Washington Hilton
BRENTWOOD
FORT LINCOLN
Echostage
Gallaudet University
United States National Arboretum
GEORGETOWN
SHAW
TRINIDAD
Georgetown Waterfront Park
Washington
DOWNTOWN
NOMA
H STREET CORRIDOR
DEANWOOD
Fairmount Heights
Seat Please
CHERRYDALE
US Marine Corps War Memorial
The George Washington University
The White House
National Gallery of Art
Lincoln Memorial
National Mall
United States Capitol
COURT HOUSE
Anacostia Park Section F
VIRGINIA SQUARE
FORT MYER
Arlington National Cemetery
Martin Luther King, Jr. Memorial
Thomas Jefferson Memorial
CAPITOL HILL
HILL EAST
Capitol Heights
Arlington
ASHTON HEIGHTS
The Pentagon
SOUTHWEST WATERFRONT
SOUTHWEST WASHINGTON
Fort Circle Park
Coral Hills
East Potomac Park
Long Bridge Park
Anacostia Park
SOUTHEAST WASHINGTON
PENTAGON CITY
Ronald Reagan Washington National Airport
Army Navy Country Club
AURORA HIGHLANDS
BARRY FARM
Suitland Federal Center
Potomac River
George Washington Memorial Pkwy
Baltimore-Washington Pkwy
Kenilworth Ave
Pennsylvania Ave
Cuetis Memorial Pkwy
Map data ©2018 Google
Terms of Use
Report a map error
Google


Map
Satellite
FRIENDSHIP HEIGHTS
University of the District of Columbia
16TH STREET HEIGHTS
FORT TOTTEN
MICHIGAN PARK
Hyattsville
East Riverdale
Edmonston
Brentwood
Mt Rainier
Cottage City
Bladensburg
Colmar Manor
Colmar Manor Community Park
Cheverly
Little Falls Stream Valley Park
Brookmont
TENLEYTOWN
Hillwood Estate, Museum & Gardens
Franciscan Monastery of the Holy Land...
Basilica of the National Shrine of the...
Rock Creek Park
Smithsonian National Zoological Park
Washington National Cathedral
PALISADES
Washington Marriott Wardman Park
Meridian Hill Park
Howard University
EDGEWOOD
FORT LINCOLN
Echostage
BRENTWOOD
FOXHALL CRESCENT
Washington Hilton
Gallaudet University
United States National Arboretum
GEORGETOWN
SHAW
Washington
Georgetown Waterfront Park
DOWNTOWN
The White House
NOMA
H STREET CORRIDOR
DEANWOOD
Fairmount Heights
Seat Pleasant
CHERRYDALE
US Marine Corps War Memorial
The George Washington University
National Gallery of Art
Lincoln Memorial
National Mall
United States Capitol
COURT HOUSE
VIRGINIA SQUARE
FORT MYER
Arlington National Cemetery
Arlington
Martin Luther King, Jr. Memorial
Thomas Jefferson Memorial
CAPITOL HILL
HILL EAST
Anacostia Park Section
Capitol Heights
Fort Circle Park
ASHTON HEIGHTS
The Pentagon
SOUTHWEST WATERFRONT
SOUTHWEST WASHINGTON
East Potomac Park
Long Bridge Park
Anacostia Park
SOUTHEAST WASHINGTON
Coral Hills
PENTAGON CITY
Ronald Reagan Washington National Airport
Army Navy Country Club
AURORA HIGHLANDS
BARRY FARM
Suitland Federal Center
Potomac River
George Washington Memorial Pkwy
Pennsylvania Ave
Baltimore-Washington Pkwy
Kenilworth Ave
Map data ©2018 Google
Terms of Use
Report a map error

Modeling

Logistic Classification	K Nearest Neighbors	Random Forest
10-fold cross validation score of 0.55	91% accuracy at k=17 **0.2%** minority recall	89% Accuracy **3.0%** minority recall
Best minority classification score of **76%** with clustering	With boosting, at k=19 **56%** minority recall	With boosting and clustering at k=8, minority recall , **0.28%** recall

Predicting major injury or fatality

Used Upsampling and K-Means Clustering

Most significant features: **distance from intersection** and **location**

Conclusions

- Scores lacked accuracy, both overall and minority identification.
- Shown areas where additional work is likely to improve models
- Fatalities are more likely to occur at or near intersections.
- Further feature engineering can improve modeling.

Future Work

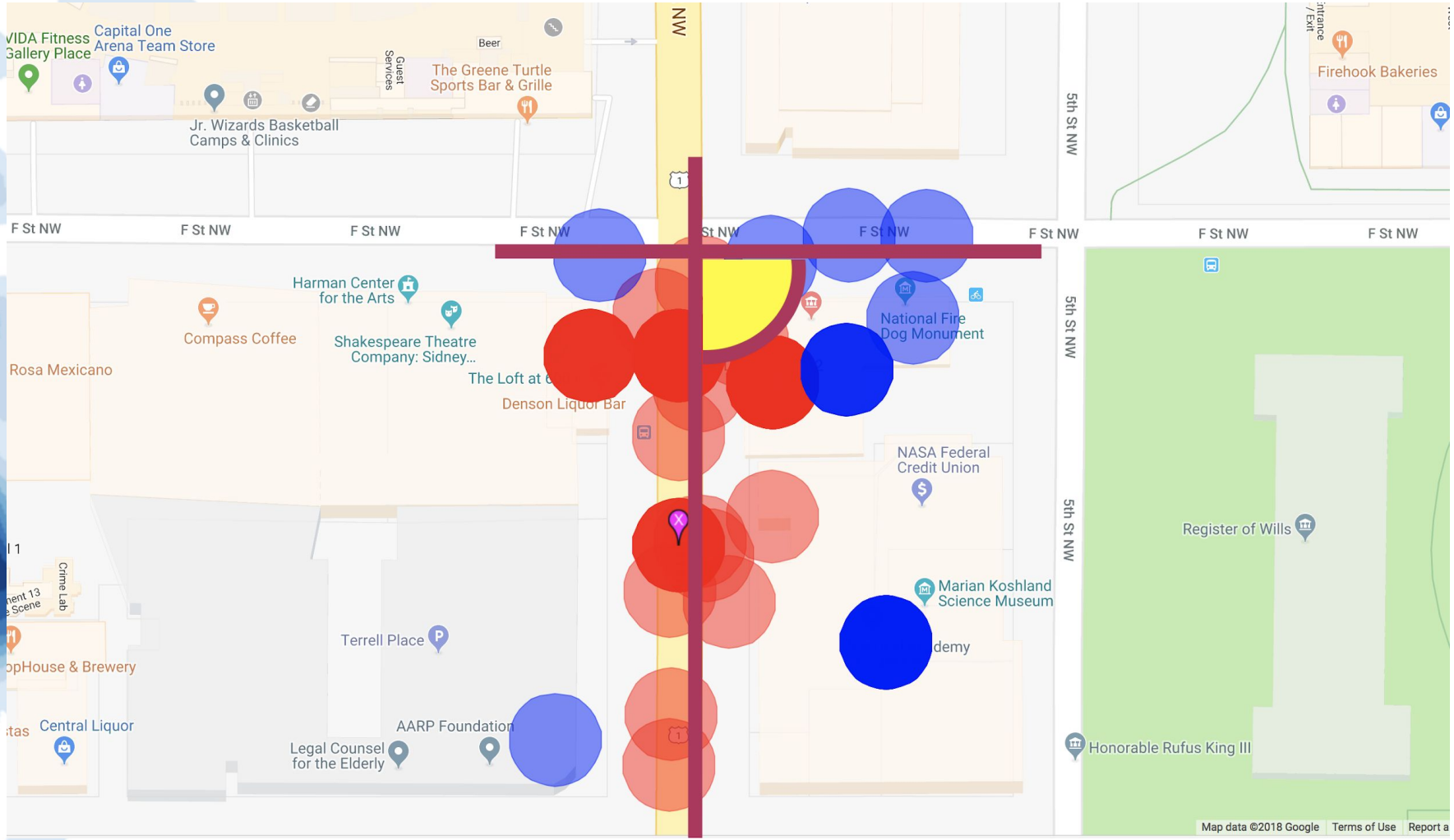
Acquire additional data

- Speed Limit
- Street Parking
- Number of Lanes
- Time of accident
- Compare DC data with other cities' data

Future Work

Engineer additional information

- Given importance of proximity to intersection, pursue adding a field to indicate the angle or acuteness of the nearest intersection.


VIDA Fitness
Gallery Place
Capital One
Arena Team Store
Beer
Guest Services
The Greene Turtle
Sports Bar & Grille
Jr. Wizards Basketball
Camps & Clinics
NW
5th St NW
Firehook Bakeries
F St NW
F St NW
F St NW
F St NW
F St NW
F St NW
F St NW
F St NW
Harman Center
for the Arts
Compass Coffee
Shakespeare Theatre
Company: Sidney...
Rosa Mexicano
The Loft at
Denson Liquor Bar
National Fire
Dog Monument
NASA Federal
Credit Union
Register of Wills
Marian Koshland
Science Museum
Crime Lab
Terrell Place
opHouse & Brewery
Central Liquor
AARP Foundation
Legal Counsel
for the Elderly
Honorable Rufus King III
Map data ©2018 Google
Terms of Use
Report a

References and Resources

Data:

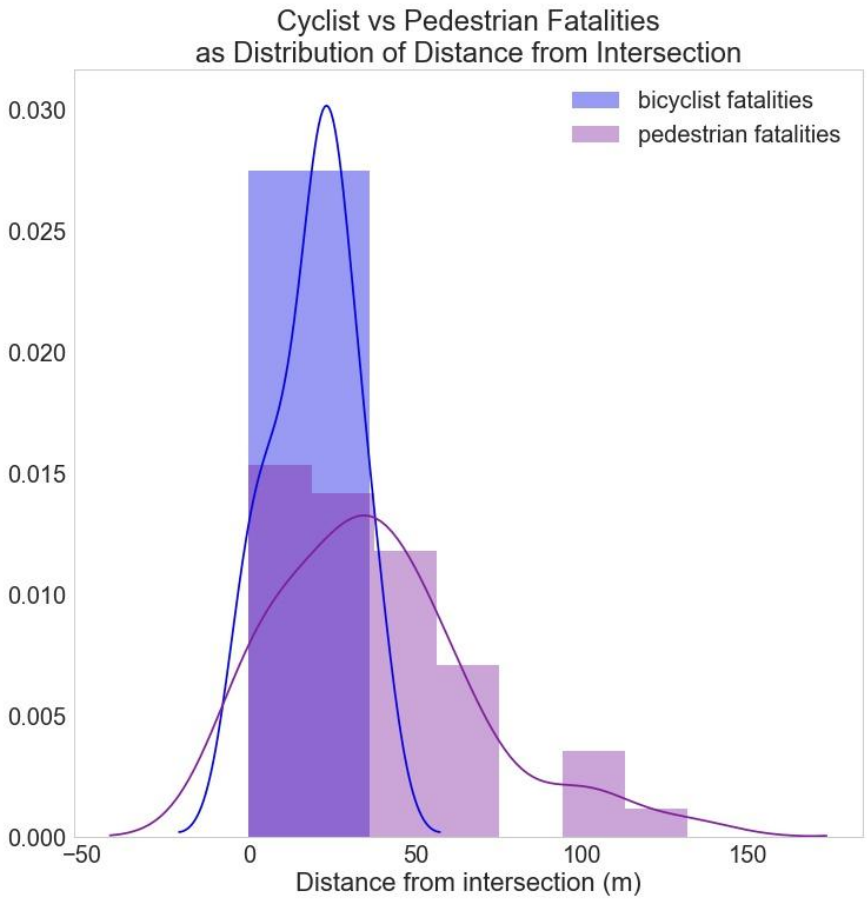
https://www.arcgis.com/home/item.html?id=70392a096a8e431381f1f692aaa06afd

Other Resources:

http://injuryfacts.nsc.org/all-injuries/deaths-by-demographics/deaths-by-age/data-details/

http://injuryfacts.nsc.org/motor-vehicle/overview/introduction/

https://www.cdc.gov/motorvehiclesafety/

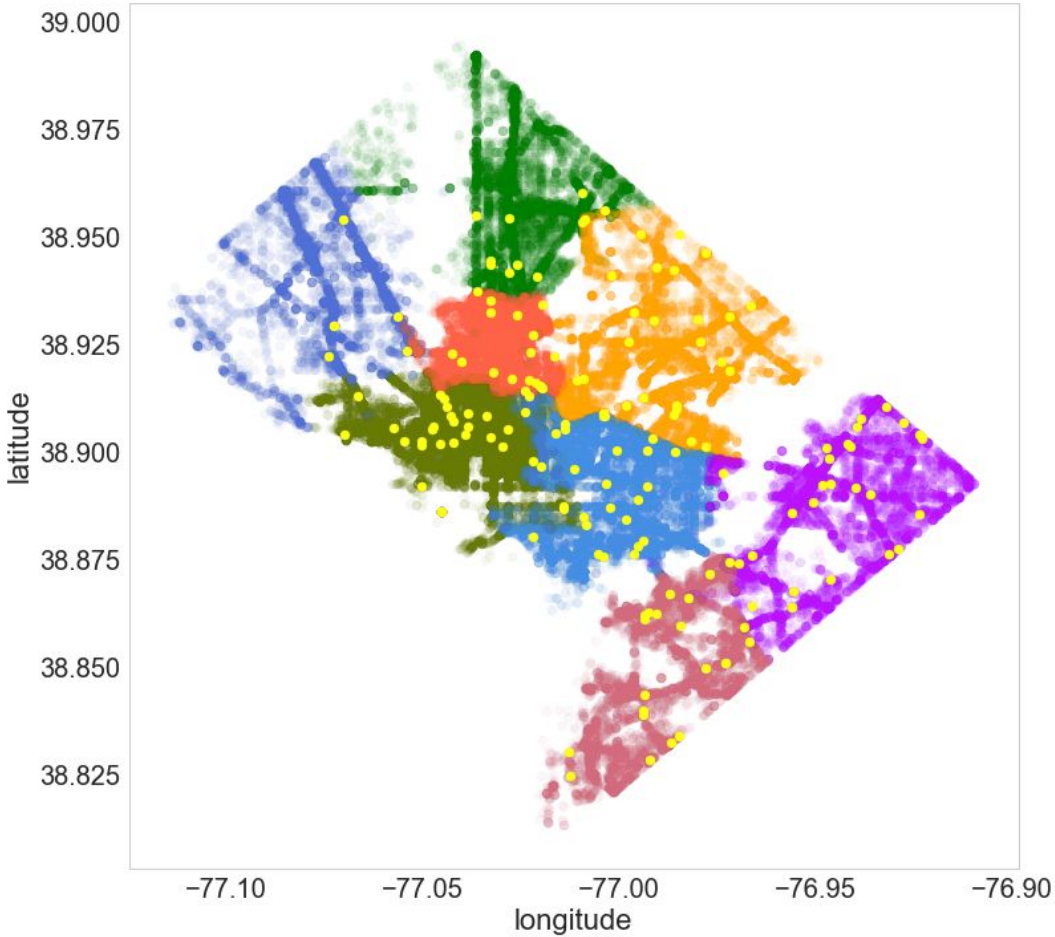

Cyclist vs Pedestrian Fatalities
as Distribution of Distance from Intersection
bicyclist fatalities
pedestrian fatalities
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−50
0
50
100
150
Distance from intersection (m)

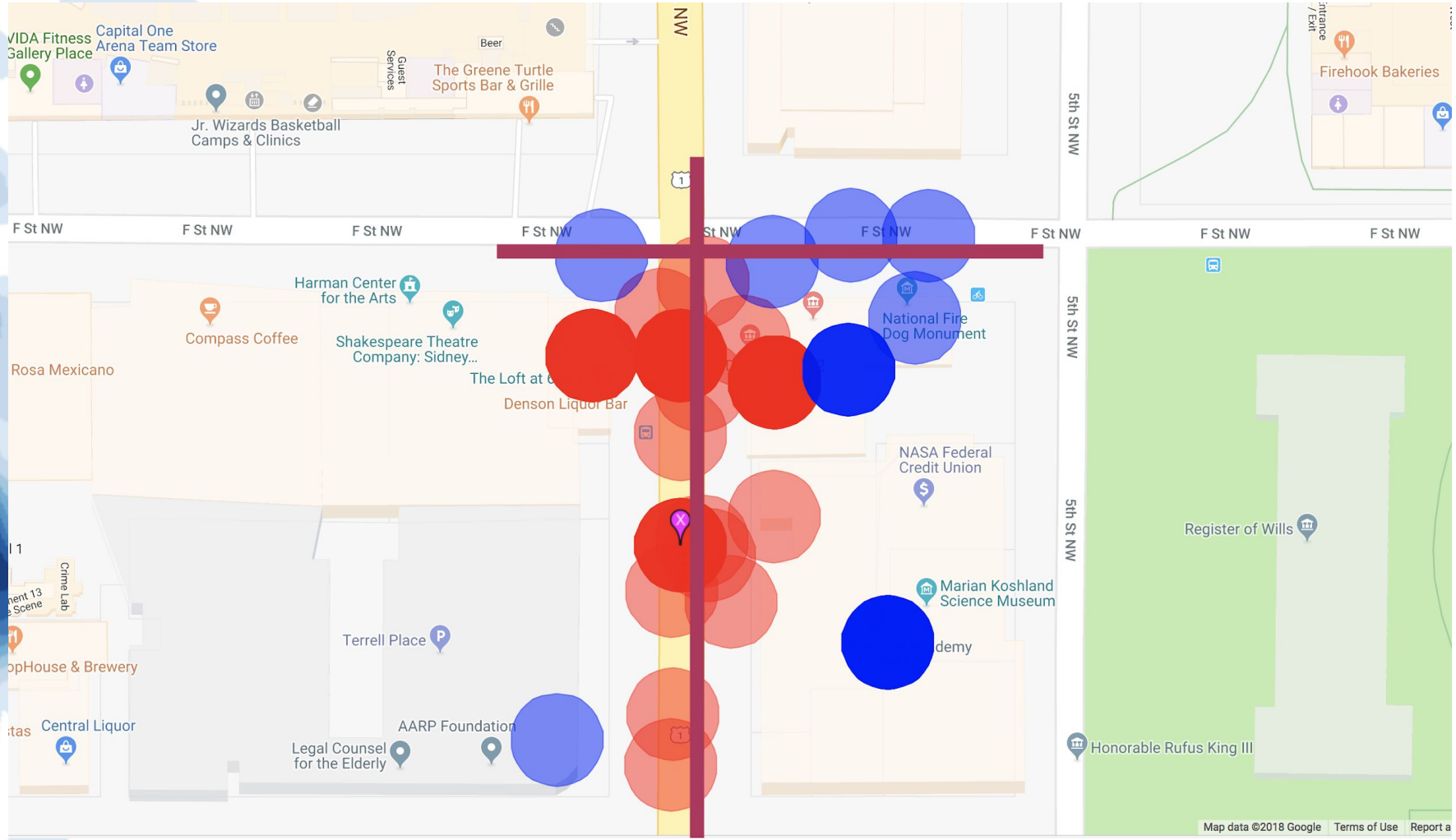
Weak Correlations

```
(df.drop(columns=['fatal_driver','fatal_pedestrian',
                  'fatal_bicyclist'])).corr()['fatal'].sort_values(ascending=False)
```

```
fatal                        1.000000
injuries_any                 0.061985
ped_inj_or_fatal             0.052655
driver_inj_or_fatal          0.045150
speeding_involved            0.033718
bike_inj_or_fatal            0.015380
total_pedestrians            0.015328
total_bicycles               0.013193
x                            0.009739
ward_number                  0.007511
pedestriansimpaired          0.005416
driversimpaired              0.003096
streetsegid                  0.001150
day_of_week                  0.001121
total_government             0.000408
minorinjuries_driver         0.000182
bicyclistsimpaired          -0.000301
roadwaysegid                -0.000618
majorinjuries_bicyclist     -0.000889
majorinjuries_pedestrian    -0.001277
minorinjuries_pedestrian    -0.002068
minorinjuries_bicyclist     -0.002563
total_taxis                 -0.002999
total_vehicles              -0.003412
majorinjuries_driver        -0.004386
offintersection             -0.005207
y                           -0.007379
Name: fatal, dtype: float64
```

All accidents, colored by ward,
with fatalities in yellow




VIDA Fitness
Gallery Place
Capital One
Arena Team Store
Beer
Guest Services
The Greene Turtle
Sports Bar & Grille
Jr. Wizards Basketball
Camps & Clinics
NW
5th St NW
Firehook Bakeries
F St NW
F St NW
F St NW
F St NW
F St NW
F St NW
F St NW
F St NW
Harman Center
for the Arts
Compass Coffee
Shakespeare Theatre
Company: Sidney...
Rosa Mexicano
The Loft at
Denson Liquor Bar
National Fire
Dog Monument
NASA Federal
Credit Union
Register of Wills
Marian Koshland
Science Museum
Terrell Place
Crime Lab
Central Liquor
AARP Foundation
Legal Counsel
for the Elderly
Honorable Rufus King III
Map data ©2018 Google
Terms of Use