Deaths in Police Custody & Officer Involved Shootings in Texas

•••

Capstone 2 Project for
Springboard's Data Science Career Track

The Problem

- From 2005-2016: 7700+ people died in police custody in Texas
- From 2010-2016: 640 reported incidents of officer involved shootings by major police departments in Texas, over 200 were fatal
- How could these have been prevented?
- What factors contributed to fatalities?
- Could discovering indicators improve officer training or reduce deaths?

Potential Clients

Non-Profits/Activists

- Texas Justice Initiative
- ACLU of Texas
- EB Wiki
- Other activist groups

May find the analysis and models useful in predicting future deaths and preventing them.

Journalists

News organizations, such as ProPublica, that do investigative reporting could use this analysis as a jumping off point for further investigation to expose biases within police departments or to report on trends.

Law Enforcement

Police Departments or the Texas government could use the analysis to improve officer training in order to reduce deaths and improve outcomes.

The Data

Deaths in Police Custody

- reported to the Attorney General of Texas
- Police departments are required by law to report deaths in custody
- data obtained by the Texas Justice Initiative and published on their website
- http://texasjusticeinitiative.org/
- downloaded as a CSV file

Officer Involved Shootings

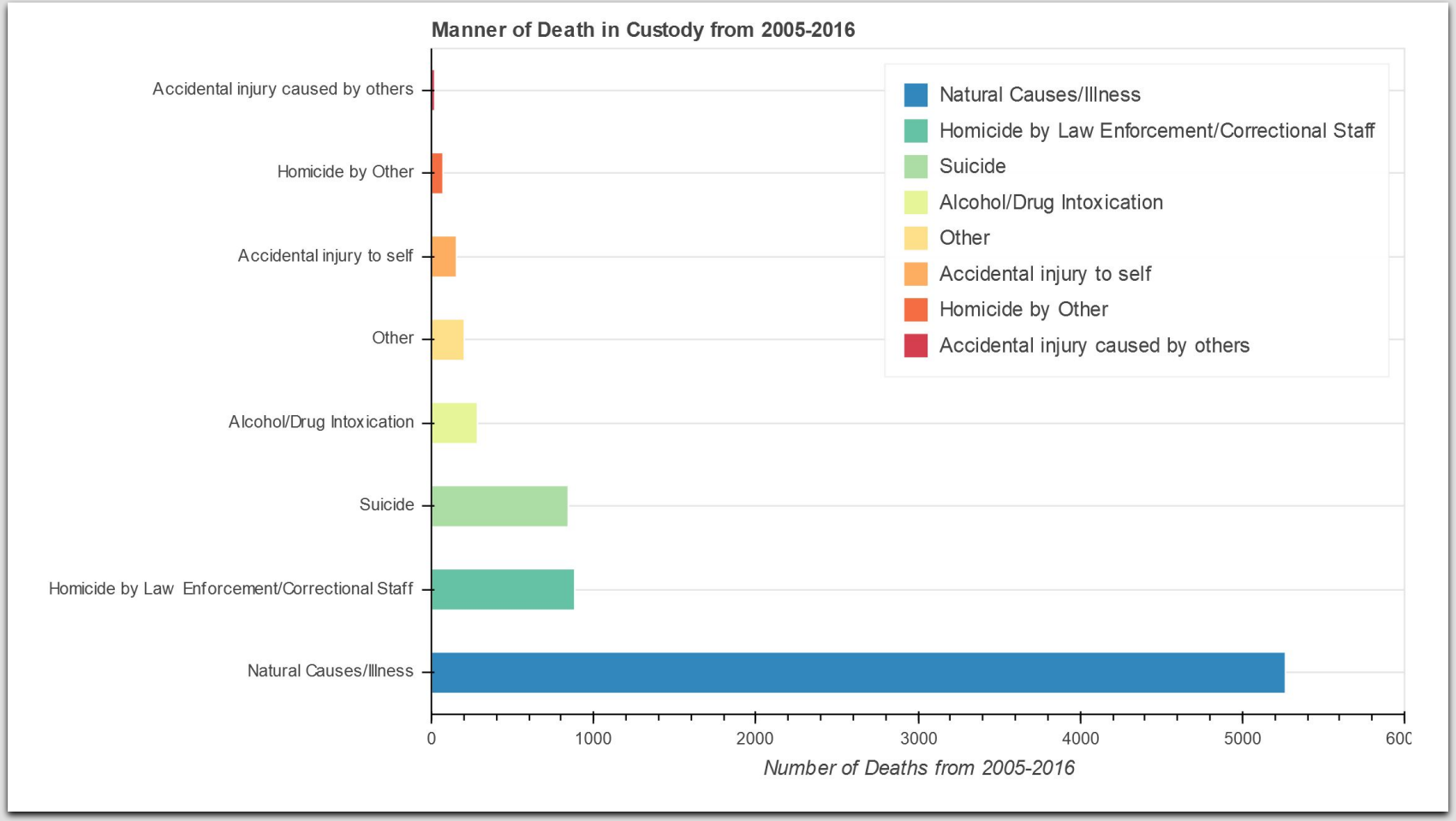
- collected by VICE news, by contacting the 50 largest police departments in the country.
- https://news.vice.com/en_us/article/a3jjpa/nonfatal-police-shootings-data
- downloaded as a CSV file

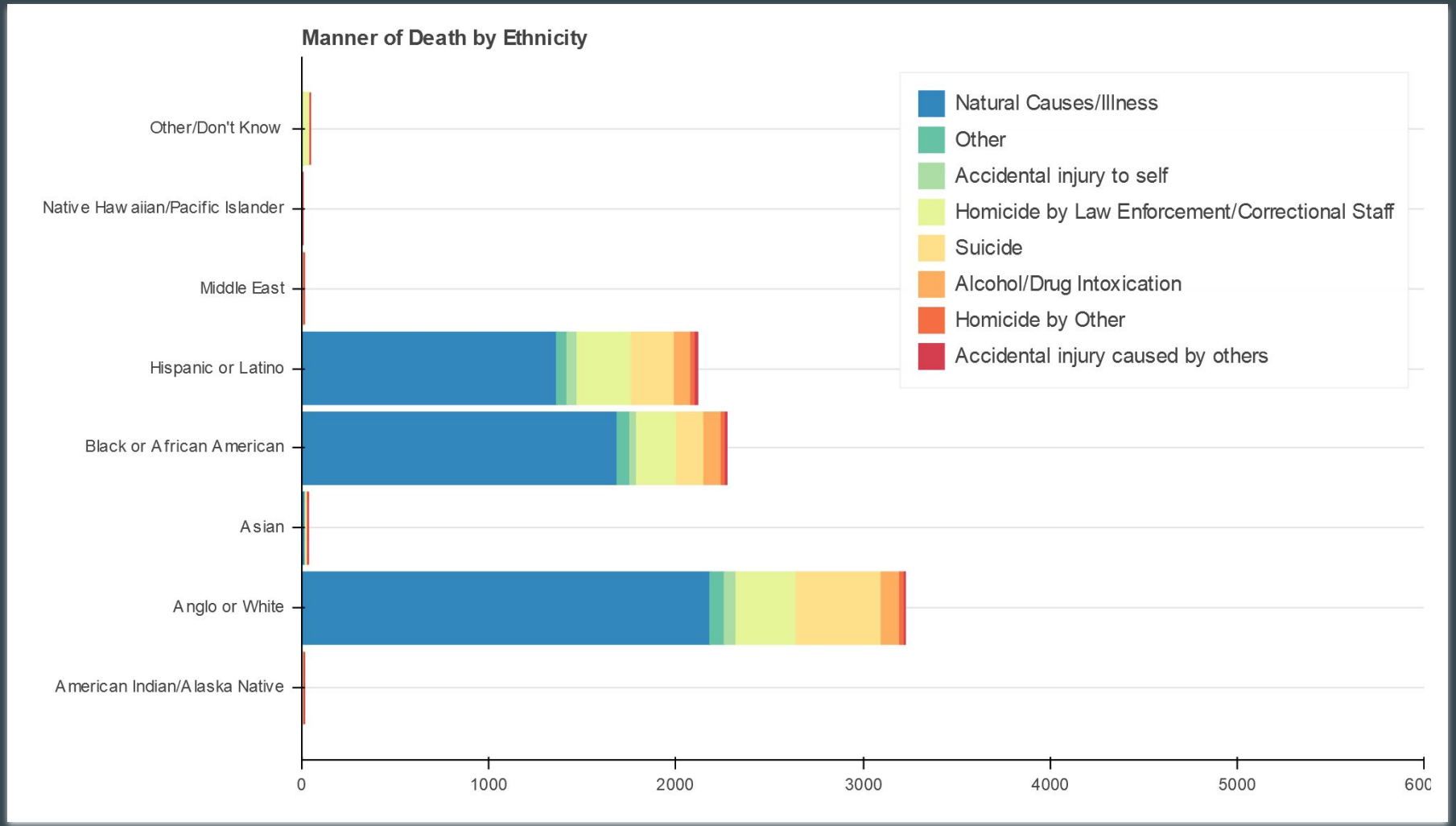
Other Potential Datasets:

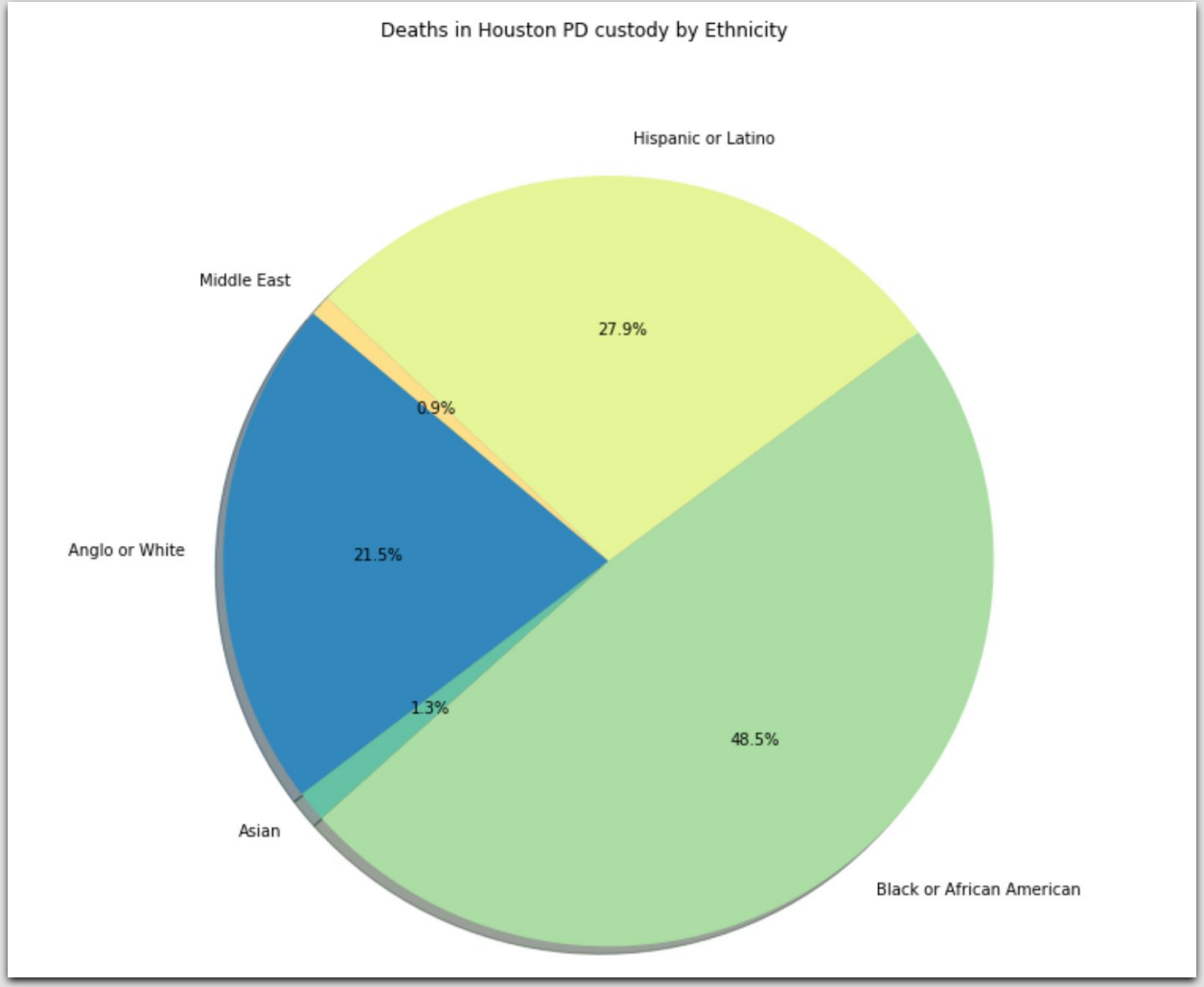
- General population demographics are not comparable for many reasons
    - FBI Uniform Crime Report - could be used to obtain statistics on the greater population in custody in Texas

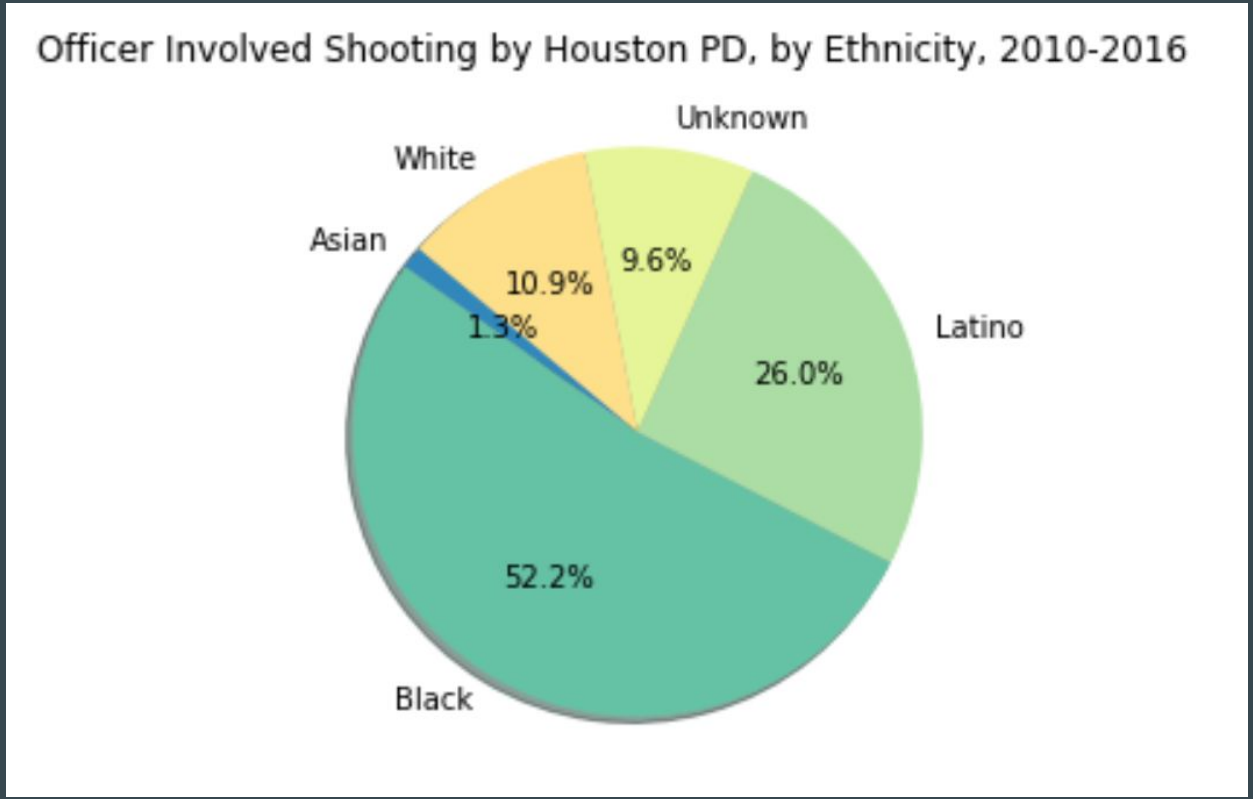
- Data on all people in custody could possibly be obtained from individual departments
    - would likely be arduous, time-consuming

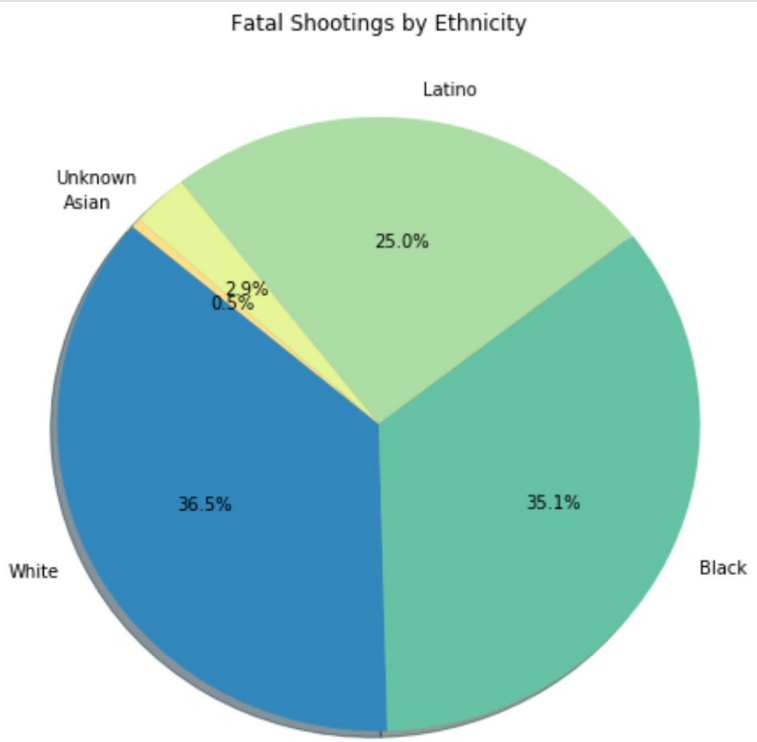
Exploratory Findings

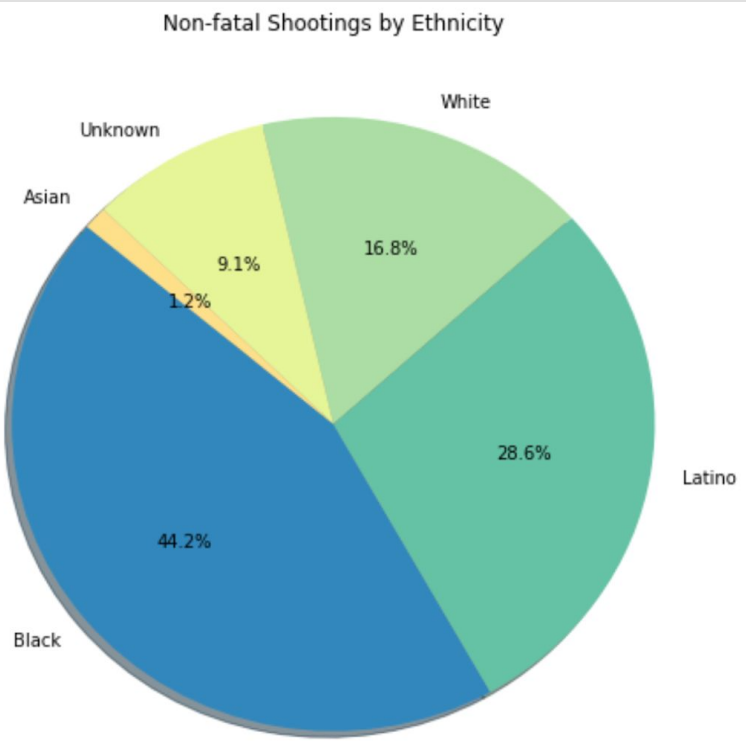

Manner of Death in Custody from 2005-2016
Accidental injury caused by others
Homicide by Other
Accidental injury to self
Other
Alcohol/Drug Intoxication
Suicide
Homicide by Law Enforcement/Correctional Staff
Natural Causes/Illness
Natural Causes/Illness
Homicide by Law Enforcement/Correctional Staff
Suicide
Alcohol/Drug Intoxication
Other
Accidental injury to self
Homicide by Other
Accidental injury caused by others
0
1000
2000
3000
4000
5000
600
Number of Deaths from 2005-2016

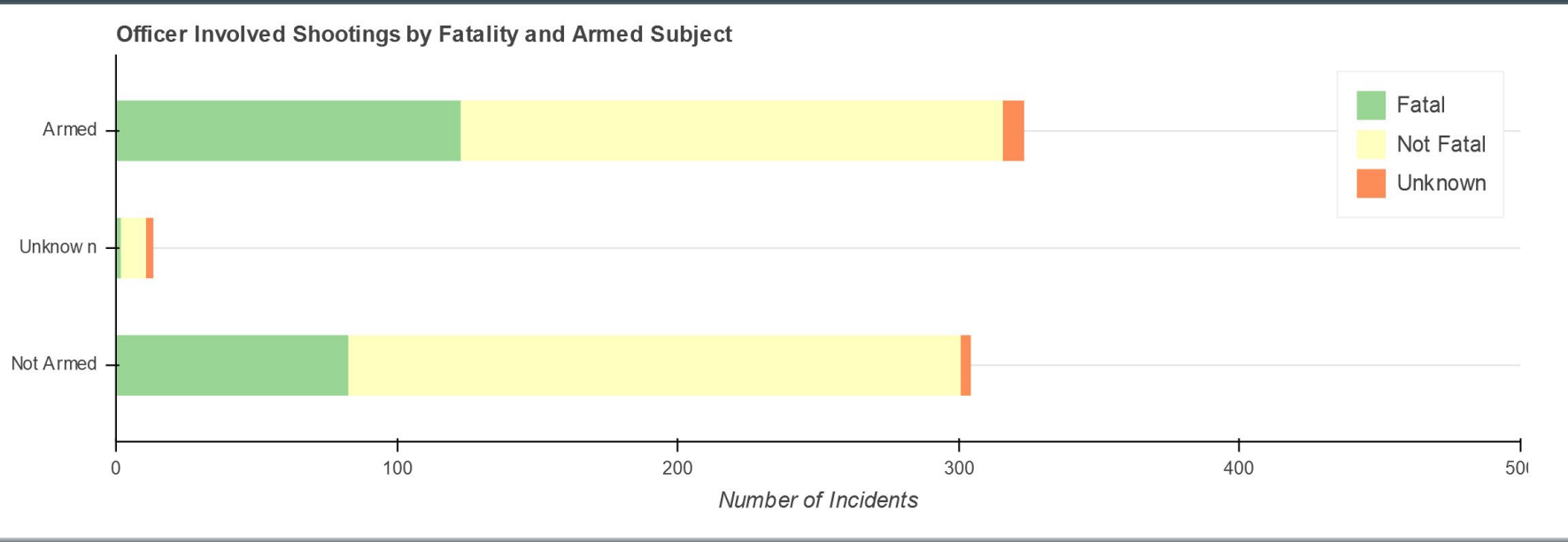

Manner of Death by Ethnicity
Other/Don't Know
Native Hawaiian/Pacific Islander
Middle East
Hispanic or Latino
Black or African American
Asian
Anglo or White
American Indian/Alaska Native
Natural Causes/Illness
Other
Accidental injury to self
Homicide by Law Enforcement/Correctional Staff
Suicide
Alcohol/Drug Intoxication
Homicide by Other
Accidental injury caused by others
0
1000
2000
3000
4000
5000
600

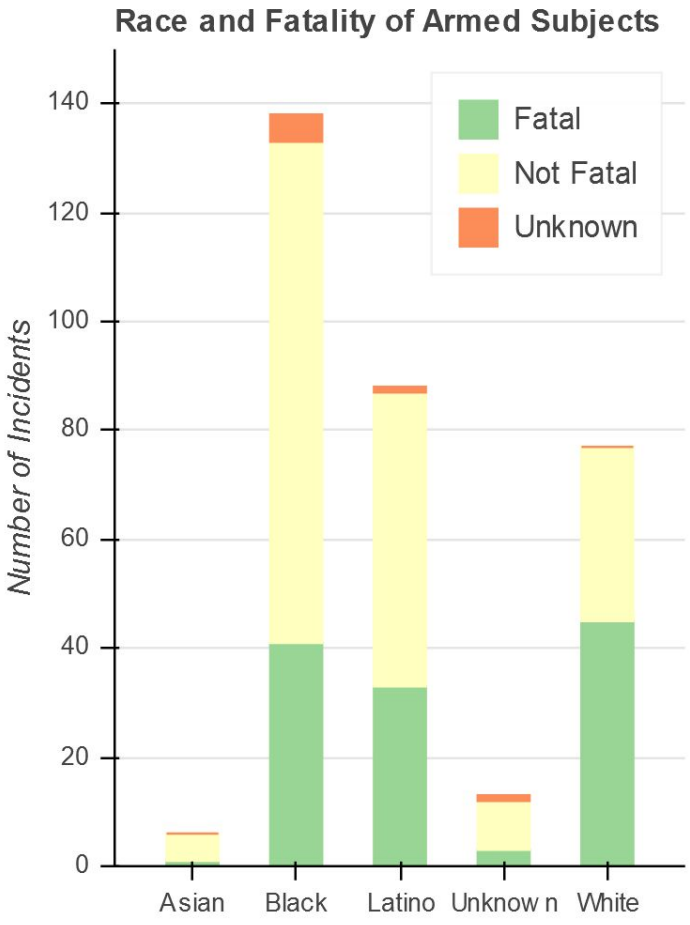

Deaths in Houston PD custody by Ethnicity
Hispanic or Latino
27.9%
Middle East
0.9%
Anglo or White
21.5%
1.3%
Asian
48.5%
Black or African American

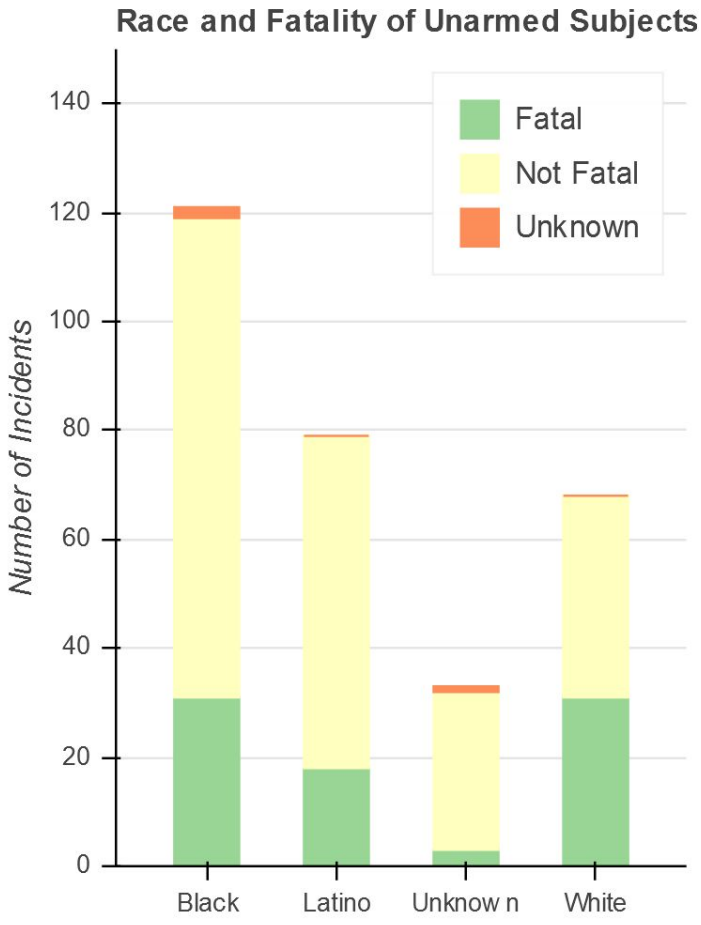

Officer Involved Shooting by Houston PD, by Ethnicity, 2010-2016
Unknown
White
Asian
9.6%
10.9%
1.3%
Latino
26.0%
52.2%
Black


Fatal Shootings by Ethnicity
Latino
Unknown
Asian
25.0%
2.9%
0.5%
36.5%
35.1%
White
Black


Non-fatal Shootings by Ethnicity
White
Unknown
Asian
9.1%
16.8%
1.2%
28.6%
Latino
44.2%
Black

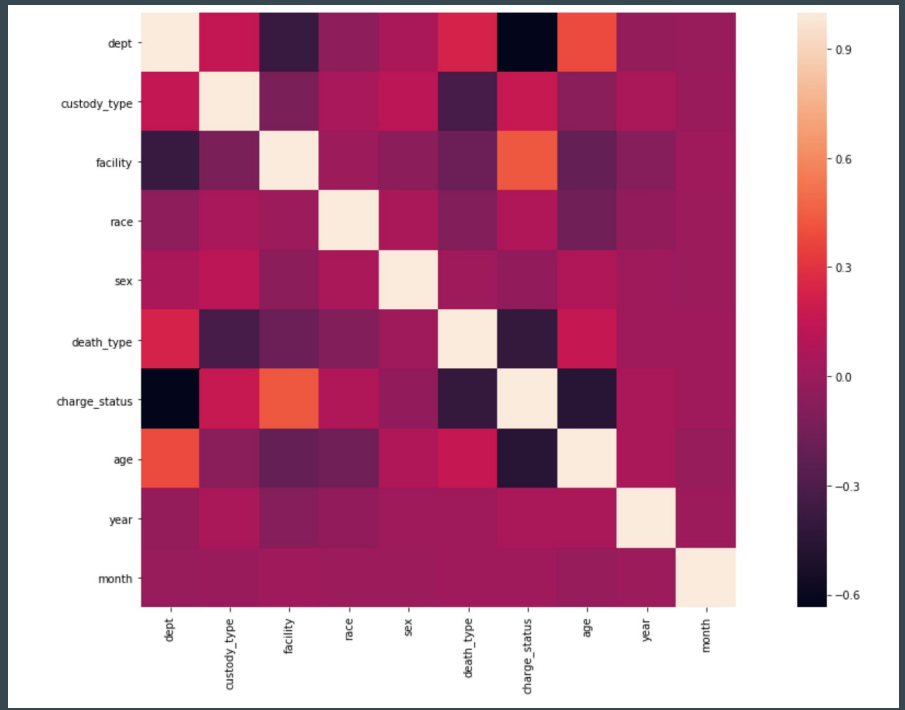

Officer Involved Shootings by Fatality and Armed Subject
Armed
Unknow n
Not Armed
Fatal
Not Fatal
Unknown
0
100
200
300
400
Number of Incidents


Race and Fatality of Armed Subjects
Number of Incidents
140
120
100
80
60
40
20
0
Fatal
Not Fatal
Unknown
Asian
Black
Latino
Unknow n
White

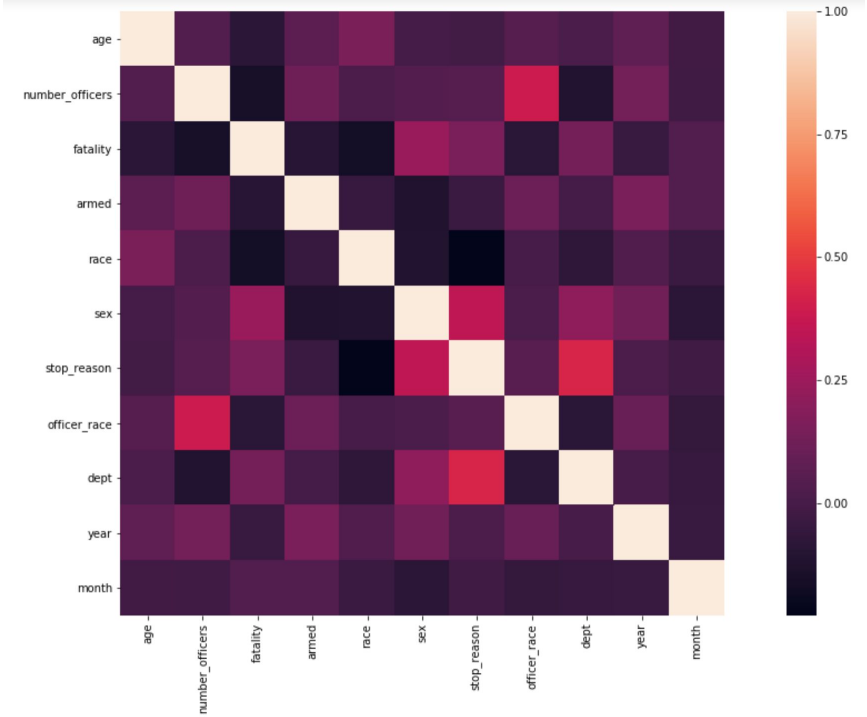

Race and Fatality of Unarmed Subjects
Number of Incidents
140
120
100
80
60
40
20
0
Fatal
Not Fatal
Unknown
Black
Latino
Unknow n
White

Heat Maps

Deaths in Custody



Officer Involved Shootings



Clustering - Model Building

Models Used

- KMeans
- Spectral Clustering
- Agglomerative Clustering
- DBSCAN

- Both datasets scaled before modeling
- Evaluated with Silhouette Score:
    - measures how similar a point is to its cluster compared to others

Deaths in Custody

Best Models

- Kmeans
  - Tuning : number of clusters set to 2-10
  - Best model: n_clusters = 7, Score: 0.44
- Spectral Clustering
  - Tuning: n_clusters = 2-10 , gamma = [0.001, 0.01, 0.1, 0.25, 0.5, 1.0]
  - Best model: 8 clusters, gamma = 0.01, Score: 0.45
- Agglomerative Clustering
  - Tuning: n_clusters = 2-10
  - Best model: 4 clusters, Score = 0.40
- DBScan
  - Tuning: epsilon = [0.25, 0.5, 0.75, 1, 1.5], “min_samples” = [3, 4, 5, 6]
  - Above these hyperparameters, DBScan only grouped data into one cluster
  - Highest scoring model: epsilon = 1.5, min_samples = 6, Score = 0.33

Officer Involved Shootings

Best Models

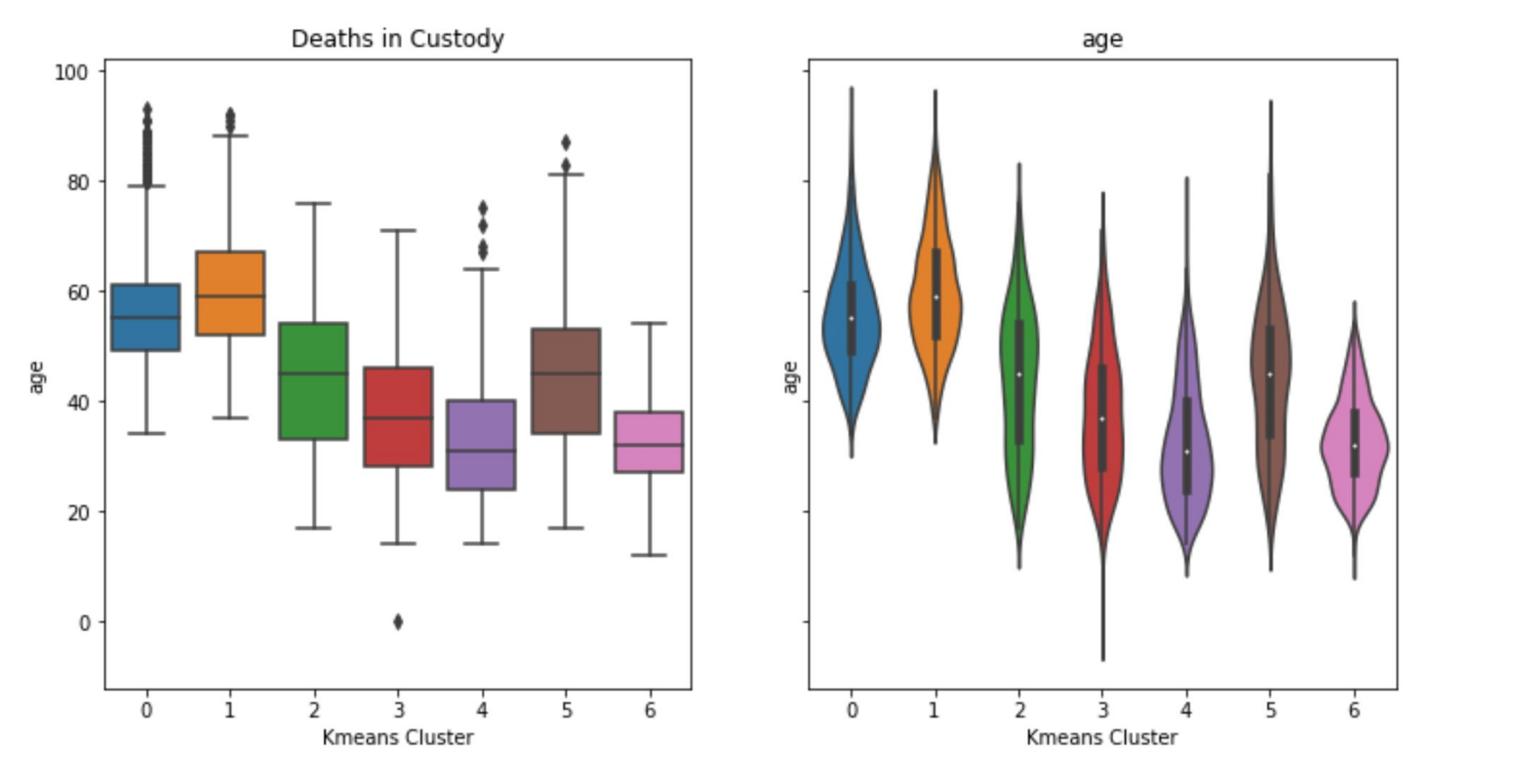
- Kmeans
  - Tuning : number of clusters set to 2-10
  - Best model: n_clusters = 9, Score: 0.38
- Spectral Clustering
  - Tuning: n_clusters = 2-20 , gamma = [0.01, 0.5, 1.0, 1.5]
  - Best Performing: 19 clusters, Score: 0.44 -- concern for overfitting
  - Selected Model: 11 clusters, gamma = 0.1, Score: 0.40
- Agglomerative Clustering
  - Tuning: n_clusters = 2-10
  - Best model: 7 clusters, Score = 0.31
- DBScan
  - Tuning: epsilon = [0.1, 0.25, 0.5, 0.75, 1, 2], “min_samples” = [1, 2, 3, 4, 5]
  - At epsilon = 3 DBScan only grouped data into one cluster
  - Highest scoring model: epsilon = 0.25, min_samples = 1, Score = 0.598
  - Clearly overfit: many cluster had only 1 point, total of 156 clusters

And the winner is...

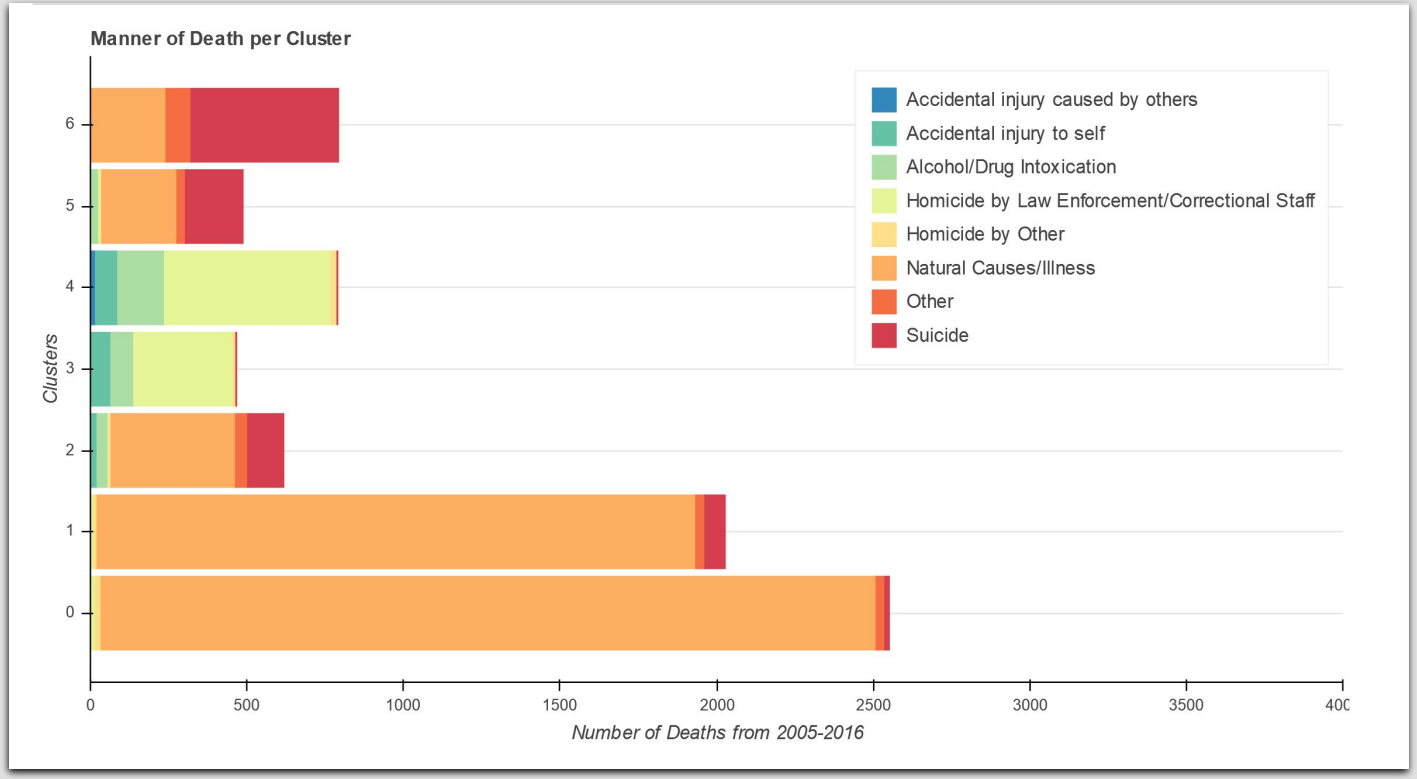
Kmeans Clustering !!!

Cluster Visualization

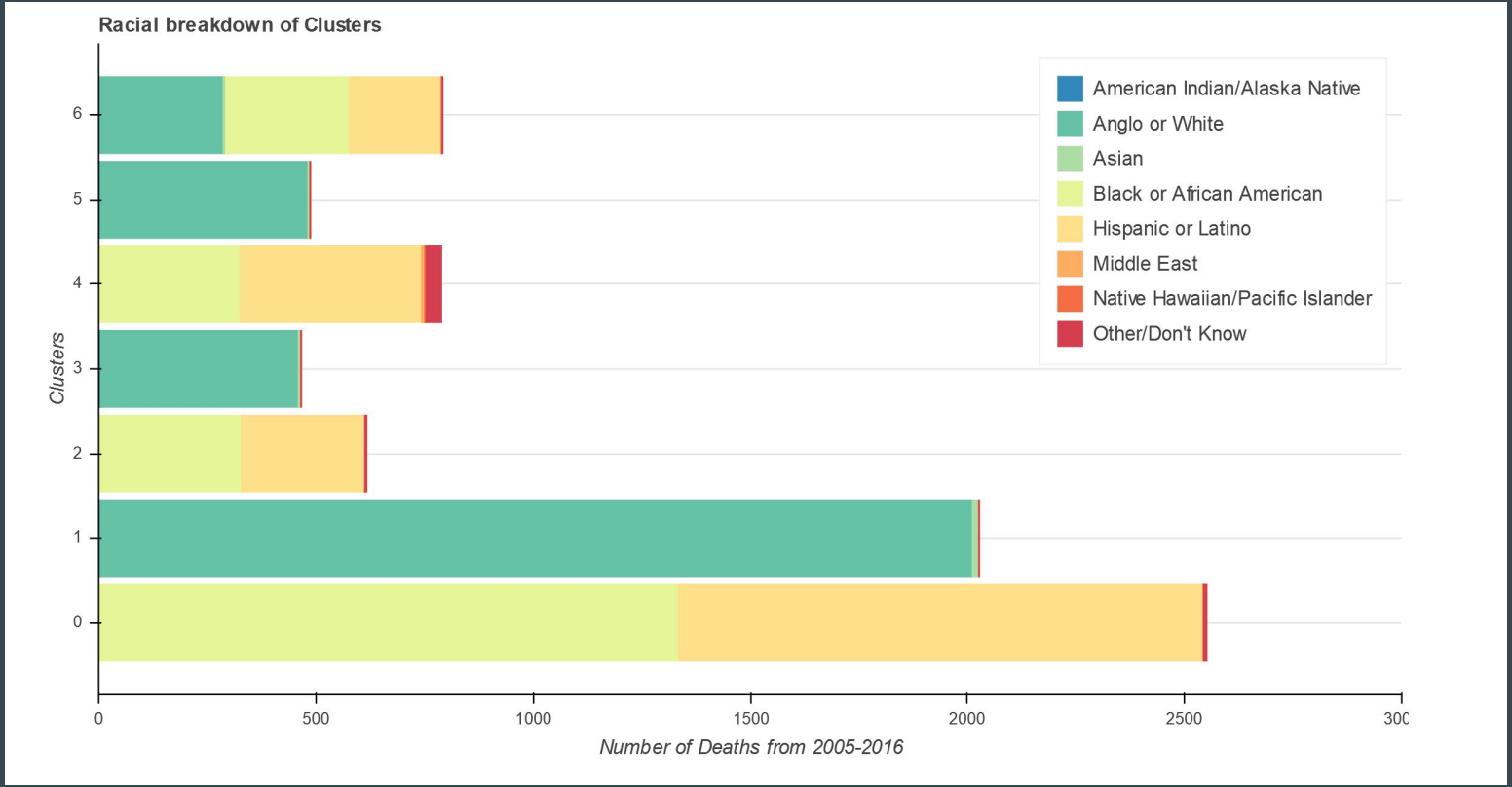
Deaths in Custody: Age



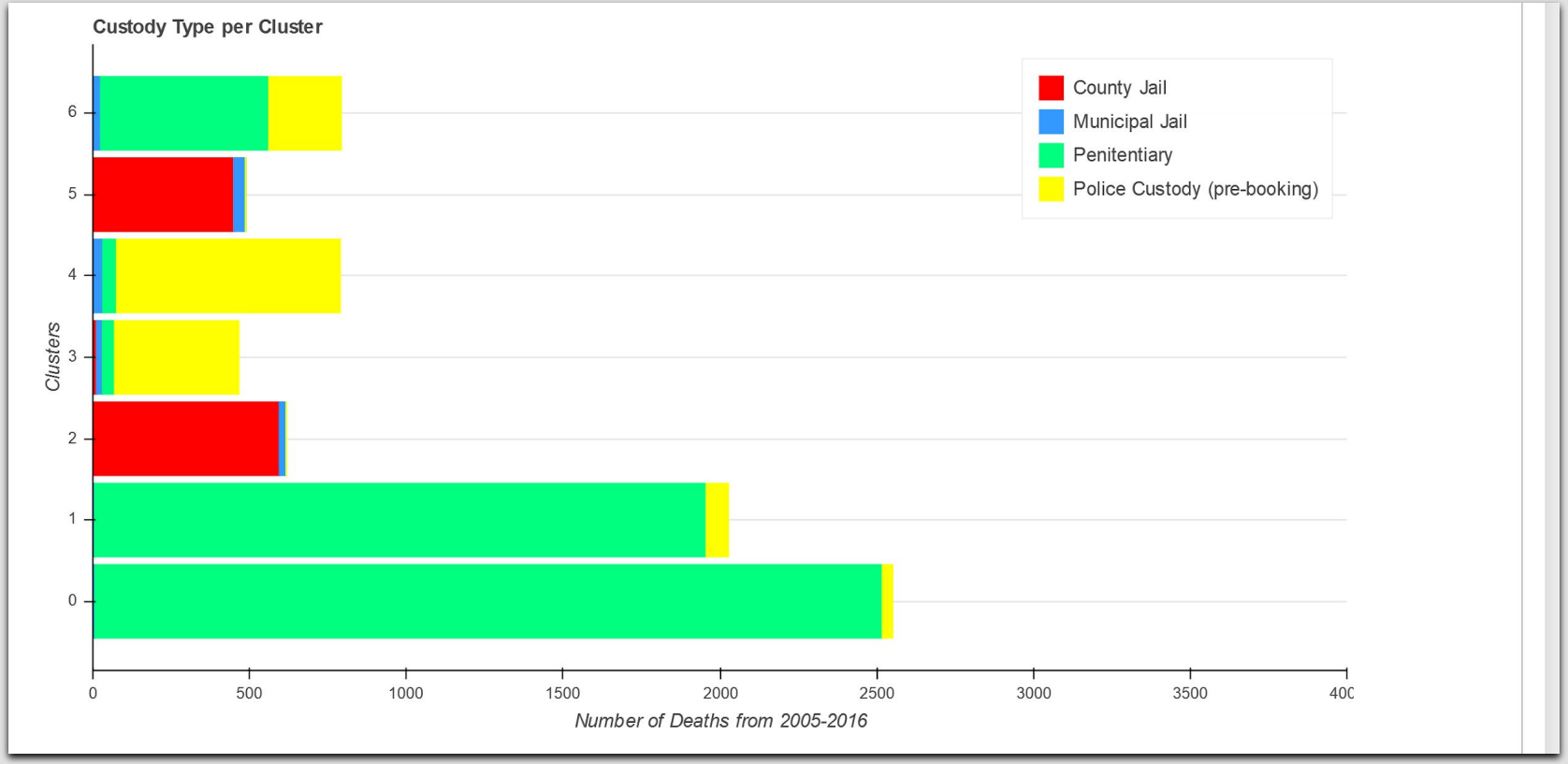
Deaths in Custody: Manner of Death



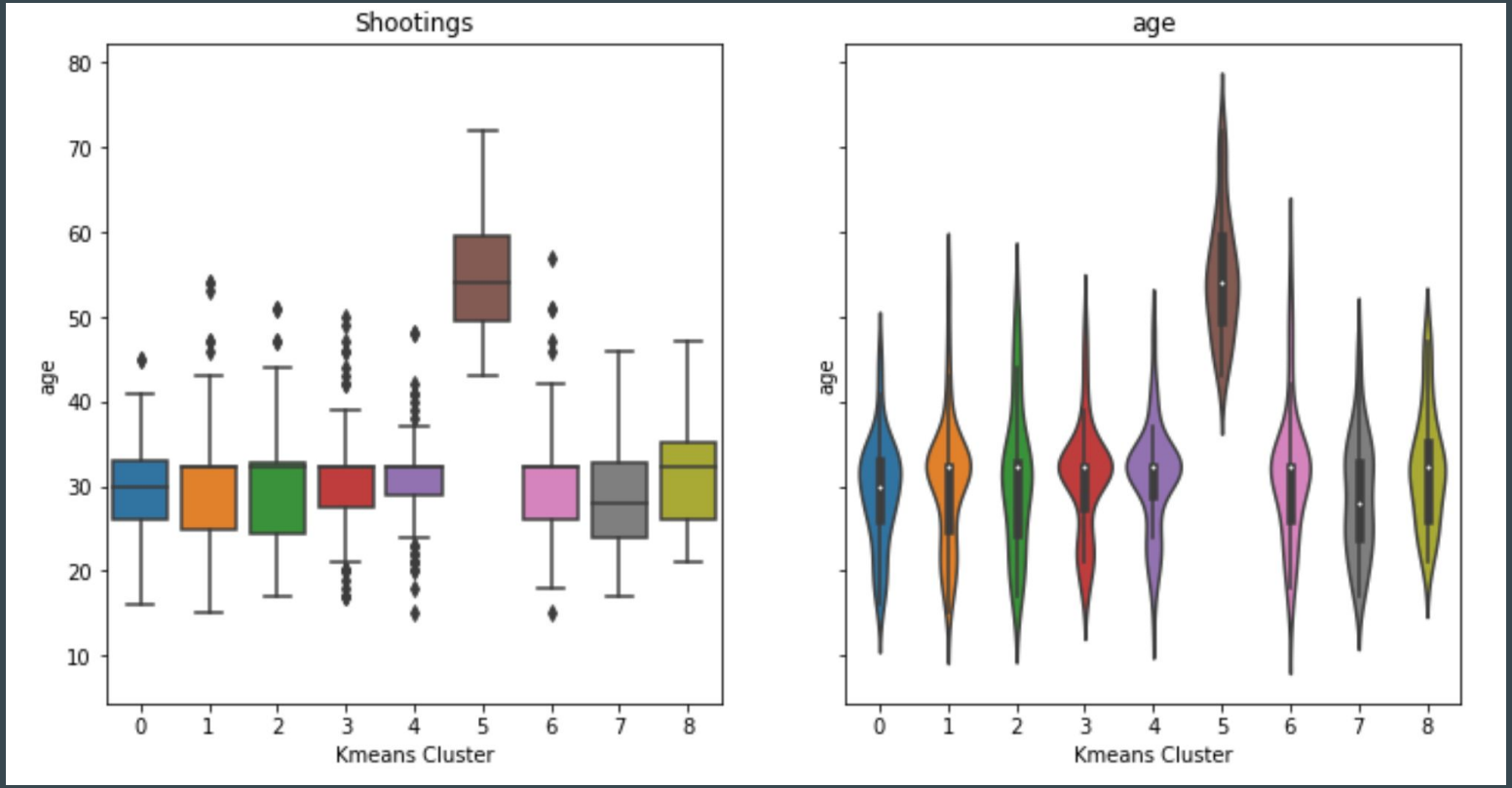
Deaths in Custody: Race



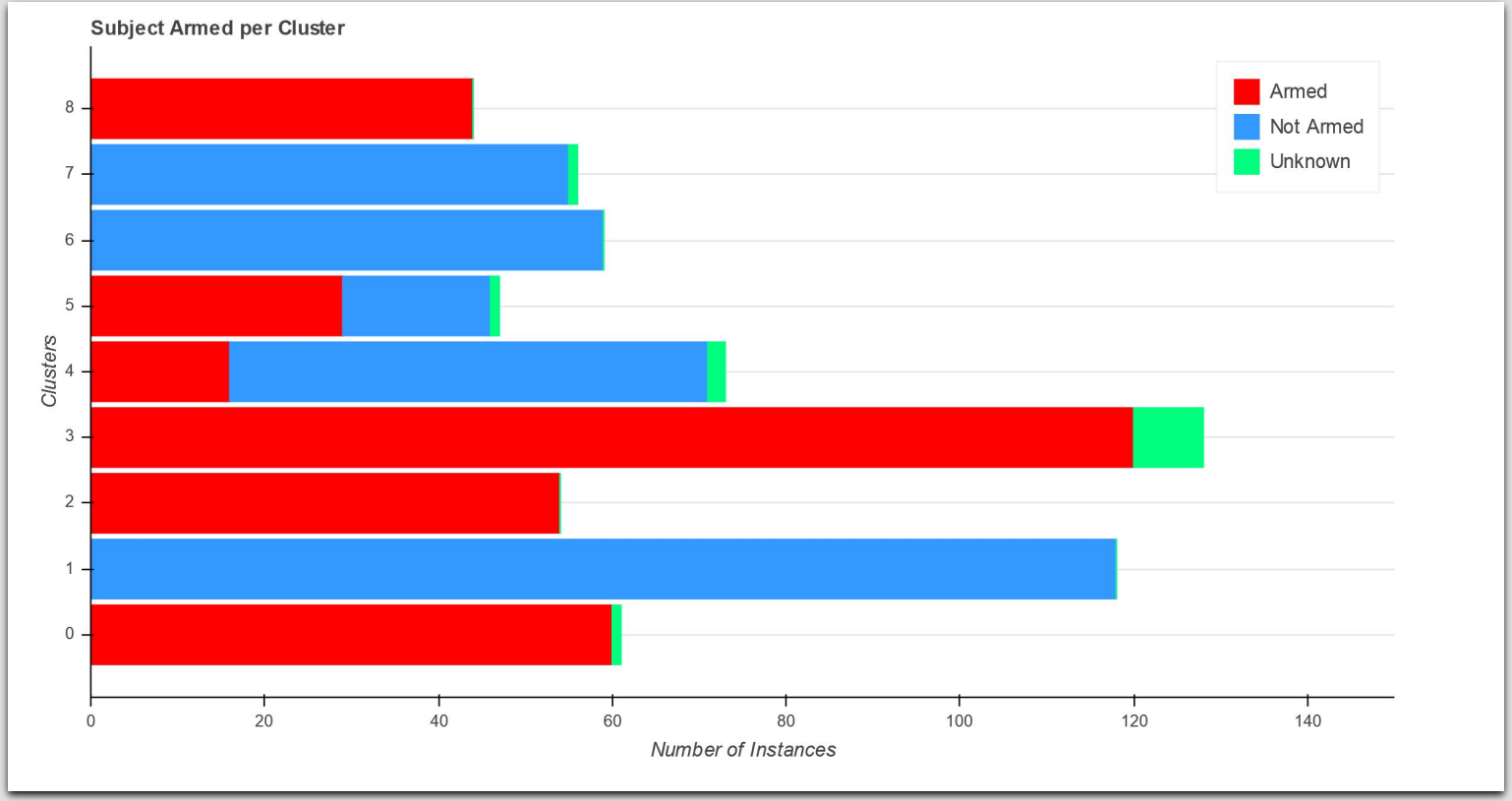
Deaths in Custody: Custody Type



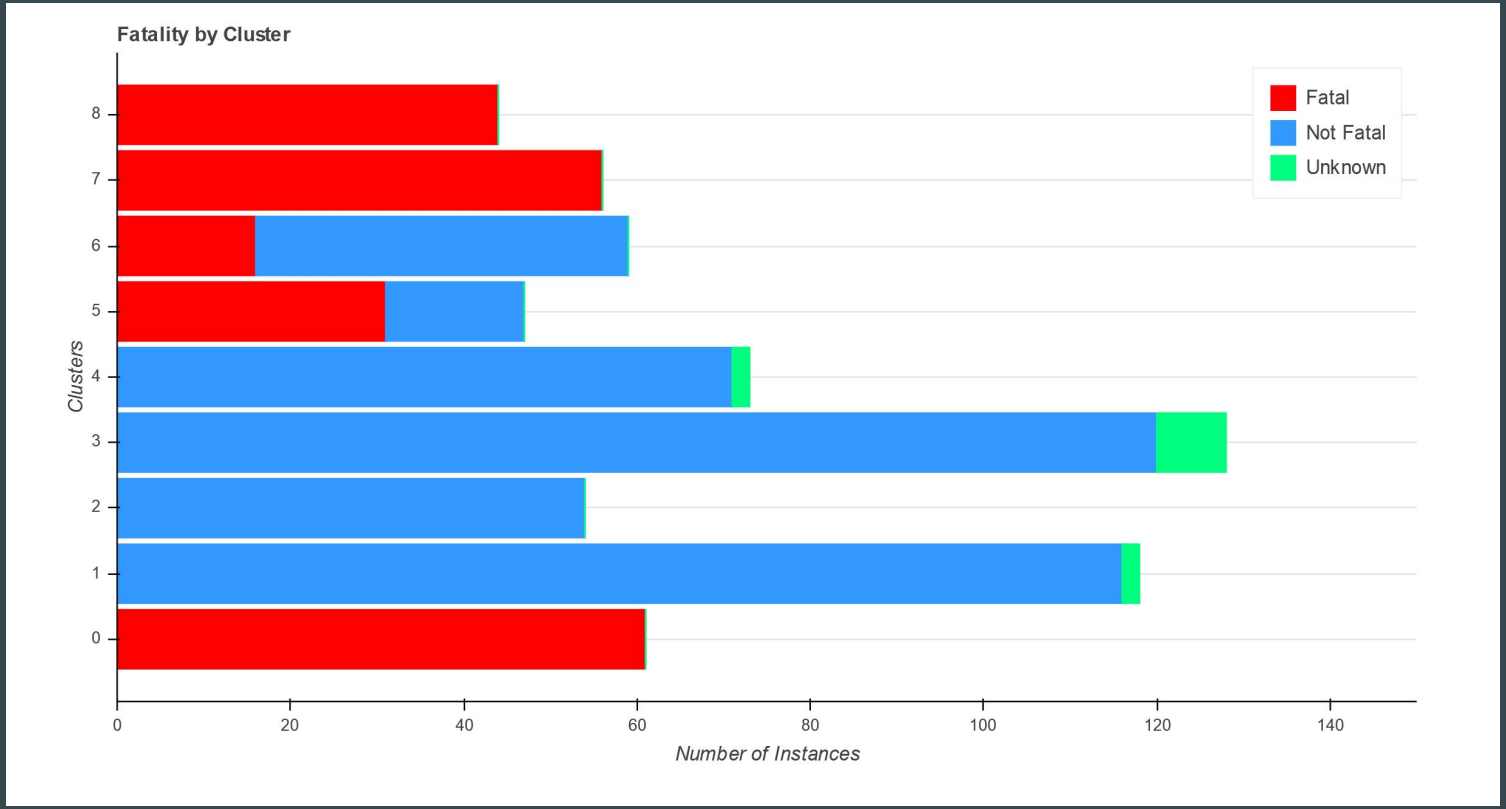
Officer Involved Shootings: Age



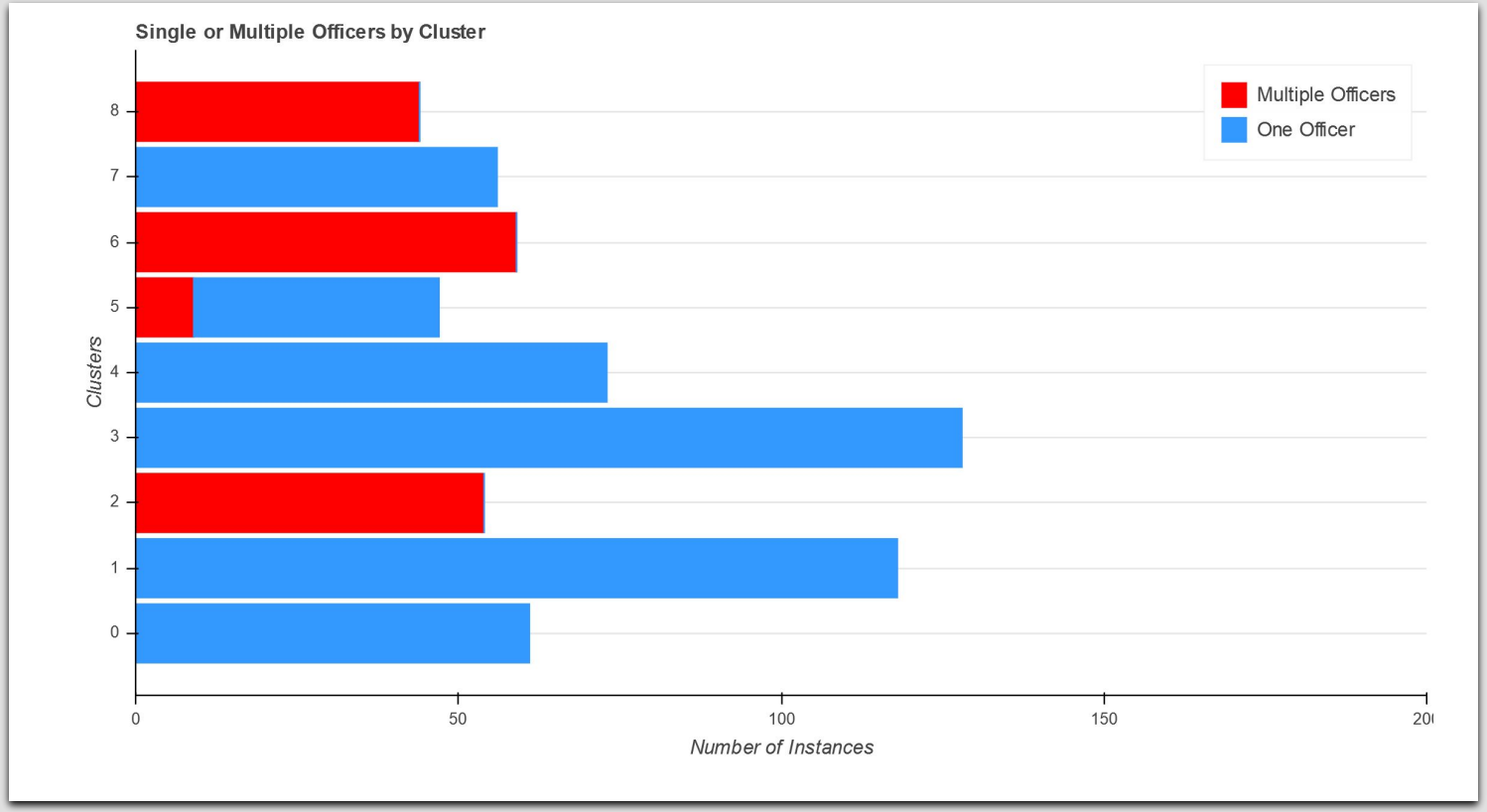
Officer Involved Shootings: Subject Armed



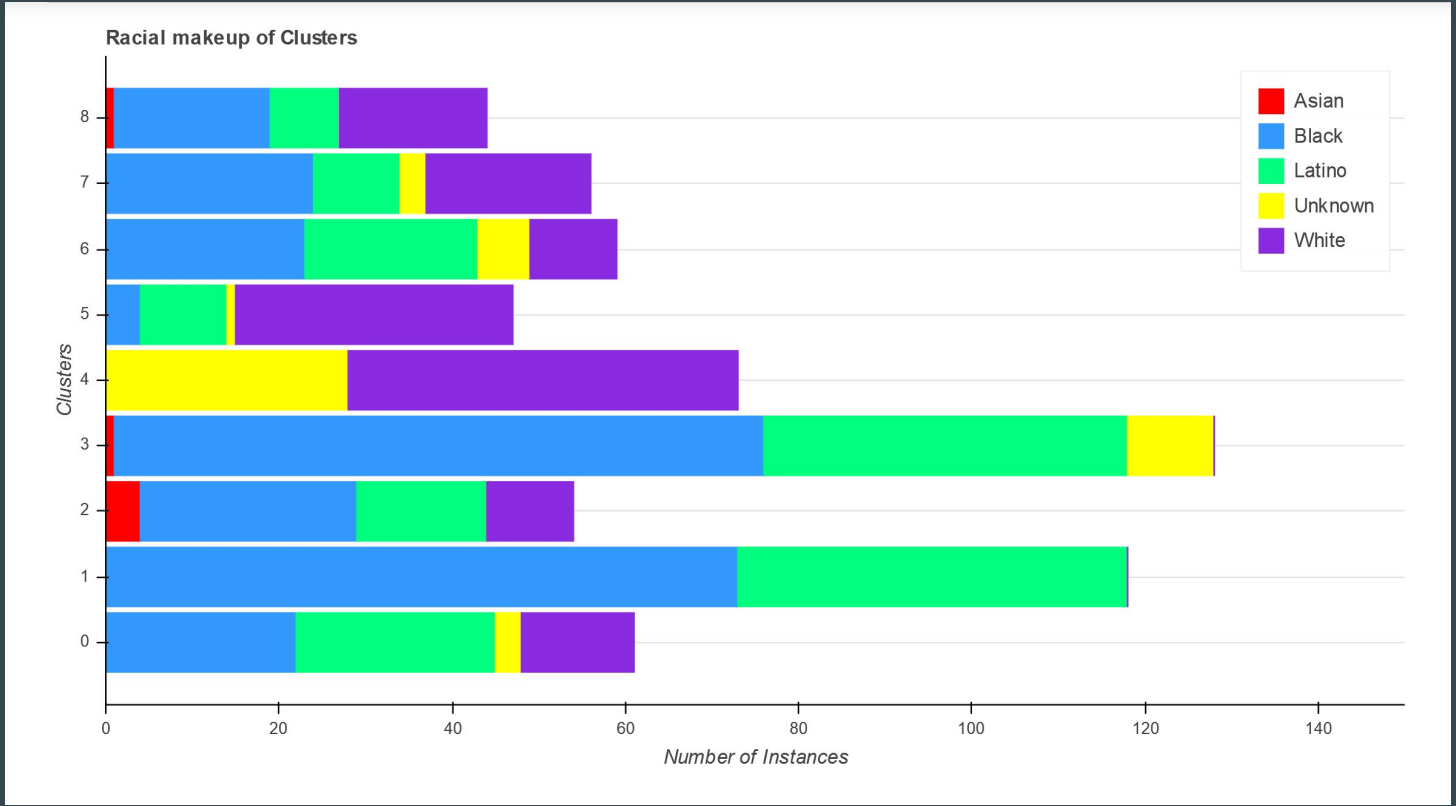
Officer Involved Shootings: Fatality



Officer Involved Shootings: Single/Multiple Officers?



Officer Involved Shootings: Race



Clusterings

Deaths in Custody		
Cluster 0 ● Older (but not much) ● Mostly natural causes ● Half black, Half Latino ● Penitentiary mostly	Cluster 1 ● Older (but not much) ● Mostly natural causes ● All White ● Penitentiary mostly	Cluster 2 ● Mostly natural causes, some suicide, bits of other manners of death ● Half black, half Latino ● County jail mostly
Cluster 3 ● Large % death by Law Enforcement ● All white ● Police custody mostly	Cluster 4 ● Largest % with death by Law Enforcement ● Mostly Latino, large Black % also ● Police custody mostly	Cluster 5 ● Half Natural cause, half Suicide, a few other manners of death ● White ● County jail mostly
Cluster 6 ● More than half Suicide ● Pretty evenly Black, White, and Latino ● Mostly penitentiary, but good portion in Police Custody		

Clusterings

Officer Involved Shootings		
Cluster 0 • Mostly armed • One Officer • Evenly Black and Latino, some white	Cluster 1 • All not armed • One officer • Mostly Black, good portion Latino	Cluster 2 • All armed • Multiple officers • About half Black, large portion Latino, some White, largest # of Asian of all clusters (but still low)
Cluster 3 • Almost all armed, rest unknown • One officer • More than half Black, good portion Latino, some unknown, one Asian	Cluster 4 • Mostly not armed, some armed • One officer • More than half white, rest unknown	Cluster 5 • Significantly older than other clusters • Majority armed, about 1/3 not armed • Most just one officer • Two-third white, some Latino, a few Black
Cluster 6 • Not armed • Multiple officers • Over 1/3 Black, about 1/3 Latino. Some White, a few unknown	Cluster 7 • Not armed, except 1 unknown • One officer • Over 1/3 Black, about 1/3 White, some Latino, a few unknown	Cluster 8 • Armed • Multiple officers • Equal of Black and White, some Latino

Conclusions

Exploratory Data Analysis

- 2nd leading cause of death in custody: “Homicide by Law Enforcement/Correctional Staff”
  - after clustering, became obvious most occurred in “Police Custody (pre-booking)”
  - likely happened during confrontation before/while being arrested
  - “Homicide by Law Enforcement/Correctional Staff” deaths that occurred in Penitentiary, County or Municipal Jail are concerning and much more overlooked
- Officer Involved Shootings: nearly 28% of unarmed subjects were killed vs 38% of armed subject
- More White people killed in armed & unarmed Officer Involved Shootings, despite Black and Latino people involved in more total shootings
  - Unclear why
  - When Houston PD excluded, White subjects become majority: makes sense White subjects have highest total fatalities
  - But still appears higher % of White subjects killed than others
  - No data on "Nature of Stop" for Houston PD: cannot be examined in relation to Race/Fatality
  - Same effect was found whether or not subjects were armed

Conclusions

Corrections and Further Analysis

- Many undiscovered insights
- Unanswered questions that can only be answered by additional data, which may not have been collected or is not currently in an accessible form.
    - ie: Deaths in Custody data: rates of occurrence in re race hard to determine as biased as rates of incarcerated etc not equal to population stats
    - Unclear whether there is racial bias in these instances w/out comparing rates to rates of all people in custody (who did not die)
    - Comparing data to population ratios just shows reflection of well-documented societal racial imbalance, not that being in custody is more deadly for people of particular races
- Many ways to expand and improve
    - Better feature engineering would help accuracy/usefulness of clustering.
    - Officer Involved Shootings-- “Officer Race” could re-coded as “White only” vs “at least one non-white officer”
    - Clustering data on “Department” level,
    - “Summary” data in Deaths in Custody: fields written about incident by a police officer
      Could be analyzed with NLP for insight into incidents & detect nefarious behavior