Homicides in NYC: When are they domestic?

R-Rated - Holly Ansel, Jiayi Liang, Matt Jogodnik, Kehan Zhang

Section 1 - Introduction

For our final project, we (Team R-Rated) will be analyzing public data provided by the New York Police Department regarding homicides in the years 2016 through 2019. Our research question is the following:

Which factors make a homicide more likely to be domestic/are associated with domestic homicides?

The homicide data we are using comes from New York City's .gov website: https://www1.nyc.gov/site/nypd/stats/reports-analysis/homicide.page, and is collected by a department of the NYPD called CompStat. New York City's police department launched CompStat in 1994, successfully lowering crime rates through the use of professional management, statistical analysis, and implementation. All CompStat data is recorded by the NYPD in the Uniform Crime Reporting format. The data is collected by CompStat because every time a homicide is committed and reported it goes into their records.Each entry in the data set represents one recorded homicide incident in New York City. The variables include the date, the precinct, the victim and perpetrator's age, race, and sex, along with weapon used, arrest status, and other various notes. These variables serve to give important information on each of the homicides.

Section 2 - Data analysis plan

Since our research focuses on attributes of the victim and how they are/are not associated with domestic homicides, our main outcome variable will be domestic, a logical (Yes/No) variable associated with whether a homicide was committed by a family member of the victim. Predictor variables will include details surrounding the homicide such as in what precinct it was committed in, the time of year, and victim/perp age, sex, and race.

Statistical Methods

We plan on using the following statistical methods:

- Statistical Hypothesis Testing

  Through the use of statistical hypothesis testing, we can see if the proportion of homicides that are domestic in NYC is significantly different from other populations, such as NY state or the entire United States. In order to perform these tests, we would need domestic proportions for these areas, however, this data should not be difficult to obtain. Furthermore, since our data comes from multiple years, we could compare and see if one year has a statistically significantly higher proportion of domestic homicides than another and use this in our narrative on domestic homicides in NYC.

- Data Visualization

  Data visualization can help us break down and compare domestic homicides among different factors like race, sex, and age. Additionally, we can use map visualizations to show the proportion

of homicides that are domestic in each precinct to see if location has an effect on a homicide being domestic, directly answering our research question.

- (Linear) Regression

  Along with visualizations, regression could prove to be a very helpful tool in finding relationships between various predictor variables and homicides being domestic (our response variable). This will allow us to determine how several variables and the interactions between them contribute to domestic homicide vs. non-domestic homicide.

- Classification

  Similar to linear regression, we intend to use the classification regression methods such as using knn and logistic regression. These tools will be of use in answering our research question based on a hypothetical homicide victim we could predict the likelihood that they were a domestic homicide victim.

Sample Visualizations

Below are sample visualizations that highlight potential ways in which we might use the data to answer our research question. Underneath the code sample we explain the findings and how this will aid our answer of the research question.

1. Mean Age

```
# A tibble: 1 x 2
  mean_victim_age mean_perp_age
            <dbl>         <dbl>
1            34.7          32.9
```

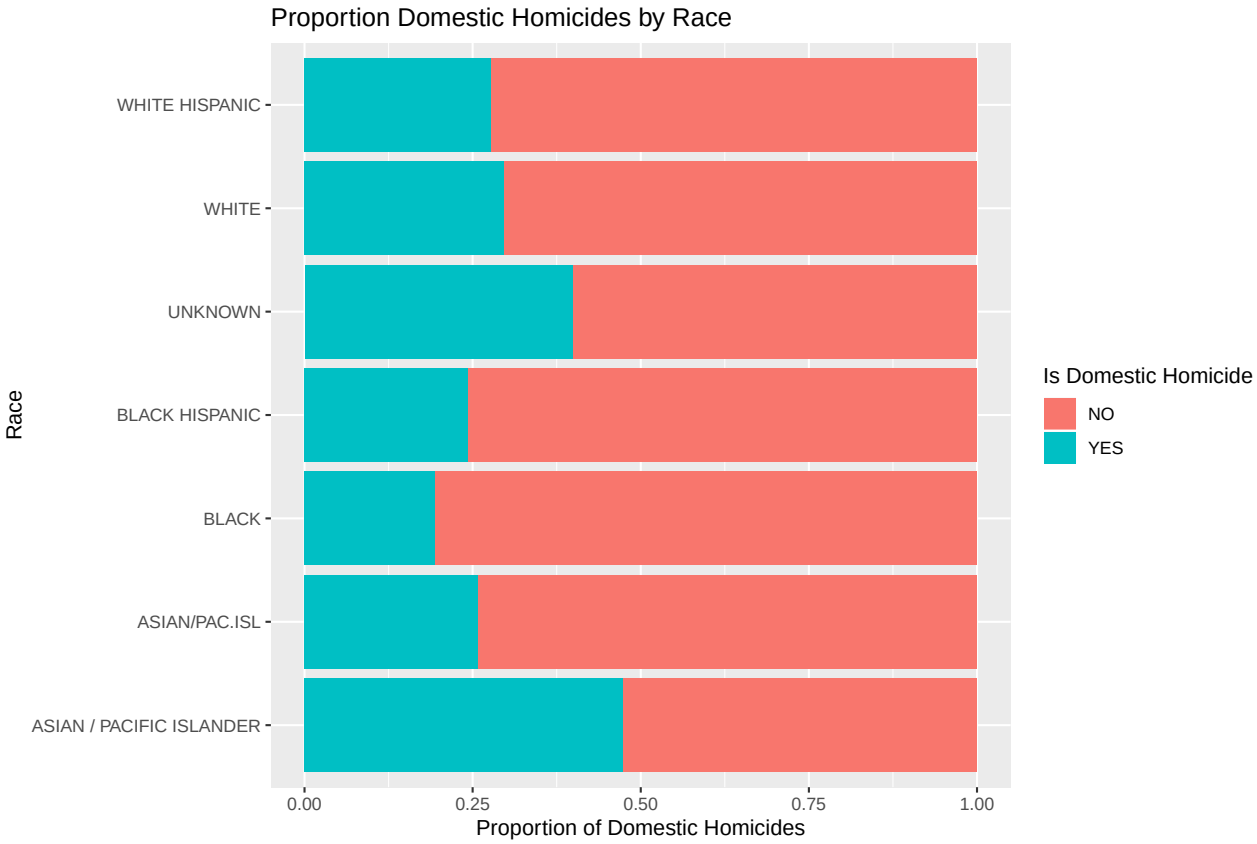
The mean age of victims of homicide in NYC between the years 2016-2019 was 34.70 while the mean age of perpetrators was 32.88. This will be used in general analysis of the effect age has on being a victim of domestic homicide as well as for general hypothesis testing of age in NYC as it compares to other areas.

2. Proportion of Homicides that are Domestic

```
# A tibble: 1 x 2
  domestic prop_domestic
  <chr>            <dbl>
1 YES              0.235
```

The proportion of homicides in NYC that were committed by relatives of the victim between the years 2016-2019 was 23.45%. This will be used in our analysis of domestic homicides and comparison to other areas - it's clear that a large proportion of homicide victims in NYC thus know their killers.

3. Domestic Homicides by Race Visualization



When looking at the proportion of domestic homicides by race it is evident that the proportion is highest among Asian/Pacific Islanders. This visualization serves our research question to show how race can potentially contribute as a factor of what makes someone more likely to be a domestic homicide victim. It also demonstrates visually that a lot of homicides are domestic.

Section 3 - Data

See data dimensions and codebook in README.

```
Observations: 1,241
Variables: 28
$ shooting_homicide_incident_id_anony <dbl> 25126, 31093, 31236, 31245, 312...
$ date                                <date> 2019-04-27, 2019-07-03, 2019-0...
$ month                               <ord> Apr, Jul, Feb, Jan, Jan, Jan, J...
$ precinct                            <chr> "041", "047", "075", "075", "03...
$ patrol_borough                      <chr> "PBBX", "PBBX", "PBBN", "PBBN",...
$ borough                             <chr> "BRONX", "BRONX", "BROOKLYN", "...
$ victim_age                          <dbl> 20, 25, 34, 29, 50, 44, 29, 38,...
$ victim_1                            <chr> NA, NA, NA, NA, NA, NA, NA, NA,...
$ victim_sex                          <chr> "M", "M", "M", "M", "F", "M", "...
$ victim_race                         <chr> "WHITE HISPANIC", "BLACK", "BLA...
$ victim_ethnic                       <chr> "HISPANIC", NA, NA, NA, "HISPAN...
$ perp_status                         <chr> NA, "ARRESTED", NA, NA, "DOA", ...
$ perp_age                            <dbl> NA, 34, NA, NA, 46, NA, 28, 38,...
$ perp_sex                            <chr> NA, "M", NA, NA, "M", NA, "M", ...
$ perp_race                           <chr> NA, "BLACK", NA, NA, "WHITE HIS...
$ perp_ethnic                         <chr> NA, NA, NA, NA, "HISPANIC", NA,...
```

```
$ relationship                  <chr> NA, NA, NA, NA, "INTIMATE PARTN...
$ weapon                        <chr> "HANDGUN", "HANDGUN", "HANDGUN"...
$ circumstance                  <chr> "UNKNOWN", "OTHER ARGUMENT", "G...
$ other_circumstance            <chr> NA, NA, NA, NA, "DOMESTIC", NA,...
$ in_out                        <chr> "O", "O", "O", "I", "I", "I", "...
$ case_n                        <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ...
$ record_n                      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ victim_n                      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ...
$ domestic                      <chr> NA, NA, NA, NA, "YES", NA, NA, ...
$ year                          <dbl> 2019, 2019, 2019, 2019, 2019, 2...
$ day                           <int> 27, 3, 14, 1, 1, 4, 6, 7, 9, 10...
$ wday                          <ord> Sat, Wed, Thu, Tue, Tue, Fri, S...
```