Leading Causes of Death

Ayodele, Cenk

2019-12-15

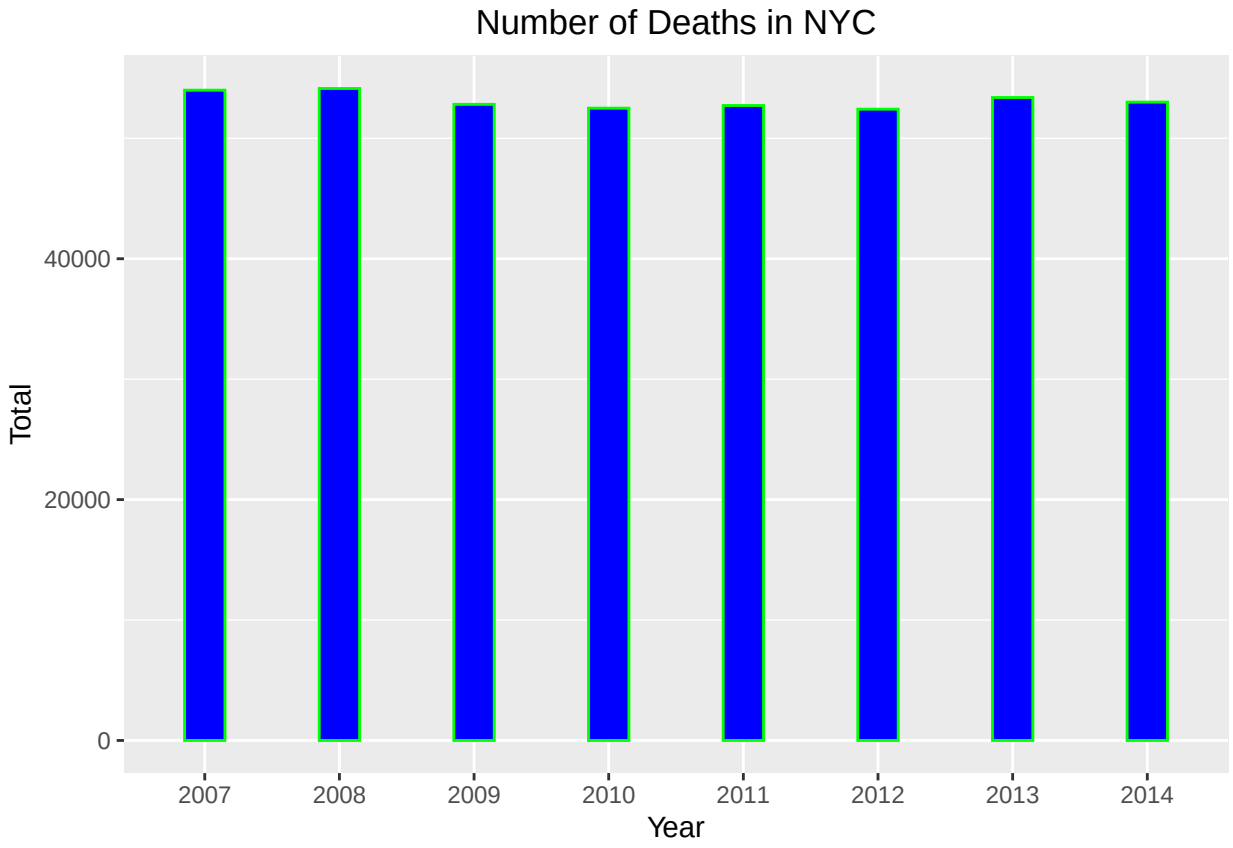
Background

Leading Causes of Death from NYC OpenData, gives a combination of observations collected between 2007 to 2014. The data collected features multiple biological diseases that are impacting the community and the number of deaths per year. In this report, we will examine the effect the leading causes of diseases has on the public.

Summary of the dataset for the Death

The graphs for the male and female show that for every year there are always more female deaths than male deaths. The graph shows there are some consistency with the deaths per year for each gender. As seen in the before graph where the death numbers don't drastically decrease or increase. The numbers are around the same, but looking at the summmary for the graphs of deaths, the average mean is 165 and the standard deviation is 186.

Question: What is the total number of deaths per year in New York City?



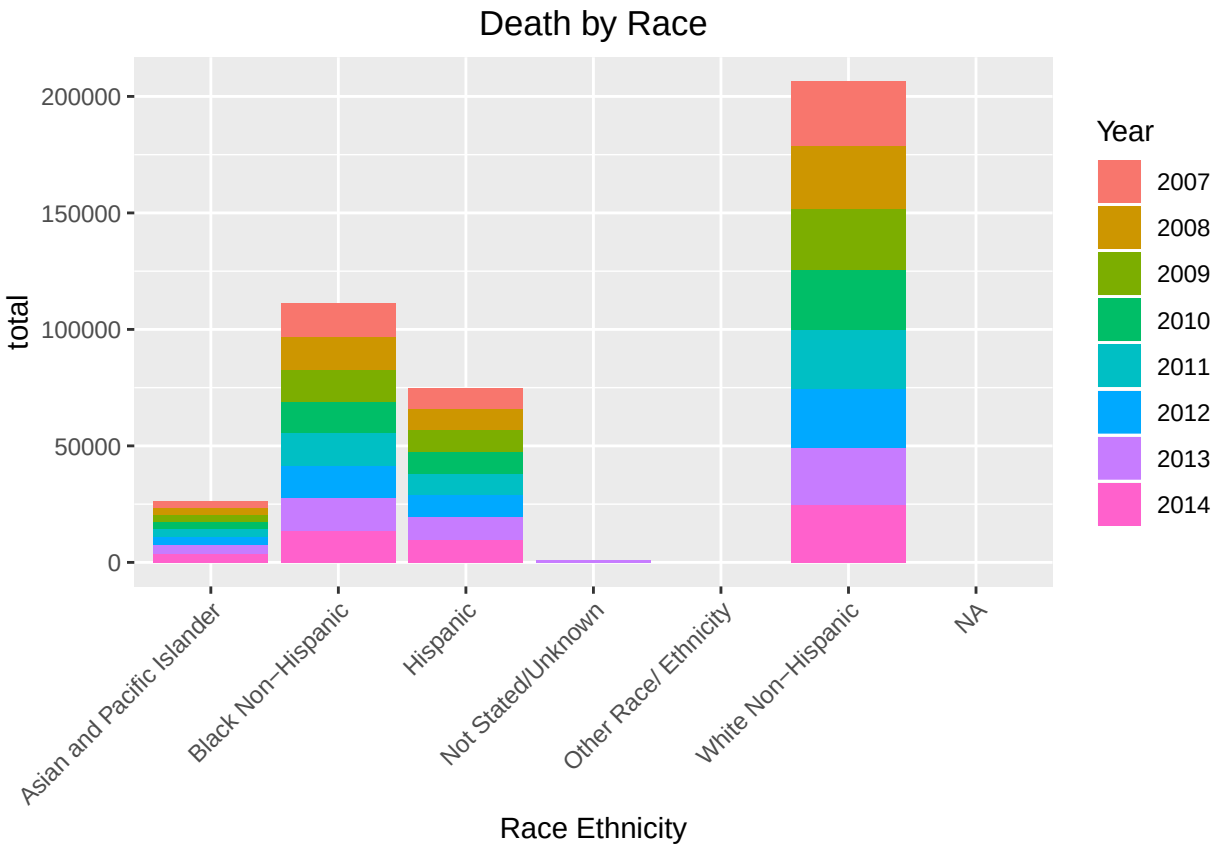

```
## # A tibble: 1 x 5
##   mean_death sd_death median_death min_death max_death
##        <dbl>    <dbl>        <dbl>     <dbl>     <dbl>
## 1       165.     186.          136         5      1557
```

Given the data: What are death counts of each sex in New York per year?

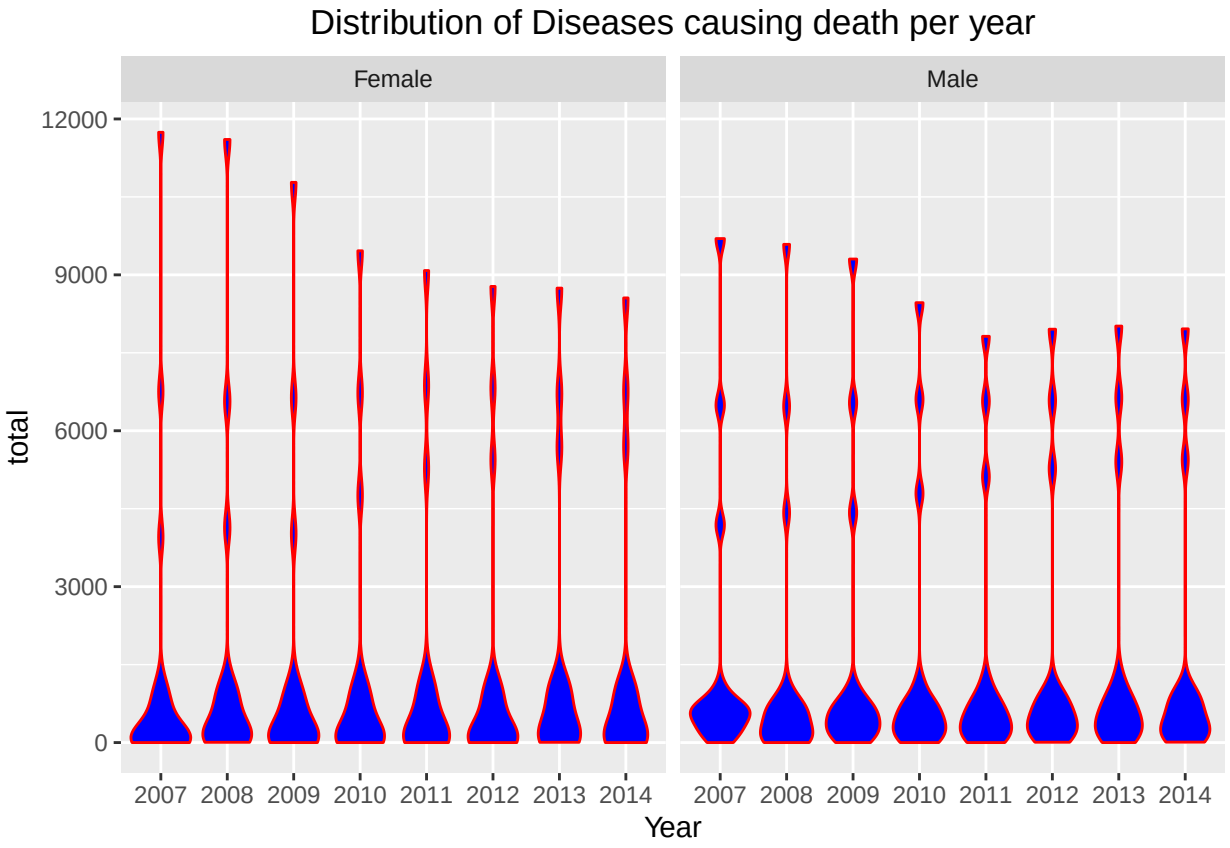
Summary of the Data

This graph shows the separtio of the data by gender. In this graph you can see what gender is being effected by biological diseases the most. Women tend to have more of a mortality than there male counter part. In the seven years of observation, each year the women's numbers of death have remained drastically higher than those of men. We thought that there might be a biological reasons that women die more than men. There are more graphs that later help us with knowing those causes.

Subsetting Dataset by Grouping Year and Ethnicity



Showing the relationship of leading causes of death to the Year and the gender

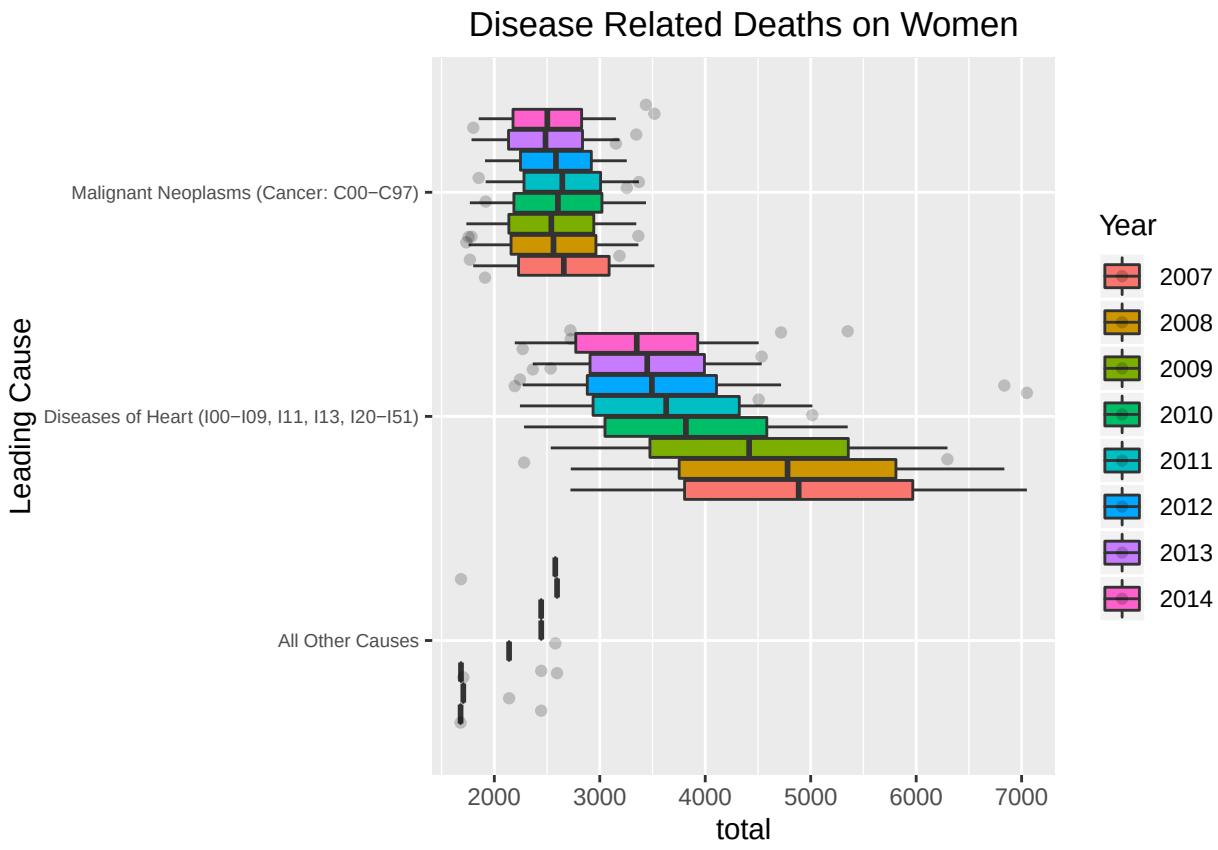


Summary of the Data

The reason for this graph is to show the data given the leading causes and the year and the number of deaths. The blue spots are the leading causes and there number of deaths. If you can see on the graph there are blue dots on the graph that stretch beyond the general population of leading causes that stay the around the same numbers. As we show the graph for each gender side by side, they show the different shapes each year makes based upon the year. With the women's graph mostly staying the same, the men's graph shapes have changed pver the years. Those maybe due to the outliers but from year 2007 to year 2011 the shapes changed in the graphs. We want to now figure out why women are dying so much from those leading causes who are

outliers on the graph.

Question: What diseases is causing women to die the most?




```
## # A tibble: 6 x 4
##   Sex    `Leading Cause`                                mean_death sd_death
##   <chr>  <chr>                                               <dbl>    <dbl>
## 1 Female All Other Causes                                    811.     813.
## 2 Female Diseases of Heart (I00-I09, I11, I13, I20-I51)     1641.    1993.
## 3 Female Malignant Neoplasms (Cancer: C00-C97)              1122.    1182.
## 4 Male   All Other Causes                                    814.     756.
## 5 Male   Diseases of Heart (I00-I09, I11, I13, I20-I51)     1433.    1634.
## 6 Male   Malignant Neoplasms (Cancer: C00-C97)              1094.    1121.
```

Summary of the data:

This graph shows the top 3 diseases that impacted the women population the most. The leading causes are "Malignant Neoplasms", "Diseases of Heart", and "All Other Causes". This data shows discrete data that has specific observations recorded.

Summary of All other Causes for Female Mortality

Results show: Mean, Standard Deviation, Median, Minimum, and Maximum

```
## # A tibble: 1 x 5
##   mean_total sd_total median_total min_total max_total
##        <dbl>    <dbl>        <dbl>     <dbl>     <dbl>
## 1       811.     813.          557        11      2595
```

Summary of Diseases of the Heart in the Female Group

```
## # A tibble: 1 x 5
##   mean_total sd_total median_total min_total max_total
##        <dbl>    <dbl>        <dbl>     <dbl>     <dbl>
## 1      1641.    1993.          846        31      7050
```

Summary for Cancer in Female

```
## # A tibble: 1 x 5
##   mean_total sd_total median_total min_total max_total
##        <dbl>    <dbl>        <dbl>     <dbl>     <dbl>
## 1      1122.    1182.         736.        11      3518
```

Therefore, we were wrong in our assumption to think that the data was normally distributed. Since the median and the mean are not the same, due to many ouliers in the data that stretched the data.