

Analyzing the Leading Causes of Death in New York City From 2007-2014

Daneesh Shahar (ads766), Christopher Cai (cc4917), Maame Buckman (meb808)

Data Bootcamp Final Project Fall 2018
NYU Stern School of Business

I. INTRODUCTION

As college students in New York City, we are berated with advertisements and campaigns, both physically and digitally, that advocate a healthy lifestyle. With a sweetgreen and Poké bowl spot seemingly sprouting out of every corner, the ‘health craze’ that is spreading in the city prompted us to question the kinds of health trends going on around us and how they affected society as a whole. This led us to become interested in looking into data regarding the leading causes of death amongst New Yorkers over time to see whether this health craze was actually leading to changes. We also felt that looking into the leading causes of death amongst New Yorkers may help us lead longer and healthier lives if we were able to pinpoint things to avoid in our everyday lifestyles.

I. DATA UNDERSTANDING

Our main dataset was downloaded from the NYC OpenData website [1] and is 1094 rows by 7 columns. It describes data on the leading causes of death in New York City from 2007 to 2014. Each row contains the leading cause, its scientific code and the corresponding year, sex, race, ethnicity, number of deaths, death rate and age adjusted death rate (Fig 1).

	Year	Leading Cause	Sex	Race Ethnicity	Deaths	Death Rate	Age Adjusted Death Rate
0	2010	Assault (Homicide: Y87.1, X85-Y09)	M	Black Non-Hispanic	299	35.1	35.5
1	2011	Mental and Behavioral Disorders due to Acciden...	M	Not Stated/Unknown	5	.	.
2	2011	Diseases of Heart (I00-I09, I11, I13, I20-I51)	M	Black Non-Hispanic	1840	215.7	268.3
3	2008	Certain Conditions originating in the Perinata...	F	Other Race/ Ethnicity	.	.	.
4	2014	Accidents Except Drug Posioning (V01-X39, X43,...	F	Hispanic	64	5.1	5.4
5	2007	Intentional Self-Harm (Suicide: X60-X84, Y87.0)	M	Not Stated/Unknown	5	.	.
6	2012	Accidents Except Drug Posioning (V01-X39, X43,...	M	Black Non-Hispanic	152	17.8	18.6
7	2009	All Other Causes	M	Asian and Pacific Islander	220	43.1	56.1
8	2013	Diseases of Heart (I00-I09, I11, I13, I20-I51)	F	Asian and Pacific Islander	437	72.8	81.8
9	2014	Accidents Except Drug Posioning (V01-X39, X43,...	M	Other Race/ Ethnicity	12	.	.

Fig 1 - Head sample of dataset

Upon seeing the dataset for the first time, we were surprised to see assault, suicide and other causes present as we expected only well-known, health-related causes of death like heart disease and cancer to show up. These other causes revealed a blind spot in our awareness of the different socio-economic situations of the population of New York City.

II. DATA PREPARATION

The cleaning process for this dataset was fairly straightforward. First, we noticed that rows containing other ‘Race/Ethnicity’ and ‘Not Stated/Unknown’ were mostly unpopulated - some rows had their ‘Deaths’ column filled in while all of them had a ‘.’ Value in their ‘Death Rate Per 100,000’ column.

These rows were not useful to us, so we converted and replaced the ‘.’ with a ‘None’ value and had any rows containing a ‘None’ value dropped from the dataset. Next, we changed all the values in the ‘Deaths’ and ‘Death Rate Per 100,000’ from objects to floats. Finally, we removed the ‘Age Adjusted Death Rate’ column as it wasn’t relevant for our purposes. This reduced the size of the dataset down to 708 rows by 6 columns.

	Year	Leading Cause	Sex	Race Ethnicity	Deaths	Death Rate Per 100,000
0	2010	Assault (Homicide: Y87.1, X85-Y09)	M	Black Non-Hispanic	299.0	35.1
2	2011	Diseases of Heart (I00-I09, I11, I13, I20-I51)	M	Black Non-Hispanic	1840.0	215.7
4	2014	Accidents Except Drug Posioning (V01-X39, X43,...	F	Hispanic	64.0	5.1
6	2012	Accidents Except Drug Posioning (V01-X39, X43,...	M	Black Non-Hispanic	152.0	17.8
7	2009	All Other Causes	M	Asian and Pacific Islander	220.0	43.1
8	2013	Diseases of Heart (I00-I09, I11, I13, I20-I51)	F	Asian and Pacific Islander	437.0	72.8
12	2012	Essential Hypertension and Renal Diseases (I10...	F	White Non-Hispanic	199.0	14.0
15	2014	Cerebrovascular Disease (Stroke: I60-I69)	M	Hispanic	165.0	13.8
16	2011	Diseases of Heart (I00-I09, I11, I13, I20-I51)	M	White Non-Hispanic	4220.0	316.4
17	2014	Chronic Lower Respiratory Diseases (J40-J47)	F	Hispanic	193.0	15.2

Fig 2 - Head Sample of Dataset post-cleaning

III. HYPOTHESES

HYPOTHESIS I - All Leading Causes Affect All Races, Ethnicities and Genders.

We hypothesize that deaths in New York City from each leading cause affect all races, ethnicities and genders. Although we understand that a myriad of factors can affect one’s chance of death from certain causes (i.e. deaths from treatable diseases are less likely to happen to high income individuals due to access to better healthcare), we are purposefully naïve with this hypothesis and expected the leading causes of deaths to overall be agnostic to race, ethnicity and gender.

HYPOTHESIS II - Deaths from leading causes should decline each year due to of New York City’s increasing life expectancy

Everywhere we look nowadays, we see evidence of New York’s move towards more health-conscious lifestyles, from the rise of healthier food options to the increased importance of healthcare and general safety. The world as a whole has also seen increases in life expectancies due to these changes and trends, so we expected New York City to follow suit. With an increase in life expectancy over time should come a decline in death rates, especially for the leading causes which we believe will be targeted the most.

IV. THE EFFECTS OF LEADING CAUSES ON CATEGORIES OF PEOPLE (HYPOTHESIS 1)

a. Approach and General Understanding

We will first explore our hypothesis that each cause of death affects all races, ethnicities and genders. To approach this, we categorized the dataset by leading cause. We did this by using the groupby function on ‘leading cause’ and aggregating the sum of deaths and the average of the death rate per 100,000 for each cause (see Fig 3). The average death rate per 100,000 allows us to see, on average, how

many people in a sample size of 100,000 is affected by each leading cause. Sorting values by descending order for the now renamed ‘average death rate per 100,000’ we expected ‘total deaths’ to be descending as well.

However, looking at the dataset (Fig 3), we noticed that ‘Assault (Homicide)’, ‘Human Immunodeficiency Virus Disease’ (HIV) and ‘Mental and Behavioral Disorders due to Accidental Poisoning and Other Psychoactive Substance Use’ (MBD) do not follow the same descending order as the ‘total deaths’ column. This prompted us to conclude that a specific category of people (i.e. Male, Asian Pacific Islander; Female, Black Non-Hispanic) is significantly disproportionately affected by these three leading causes, and that potentially, the leading cause does not affect all races, ethnicities and genders. Additionally, due to the different population sizes of each category of people in New York City, the aligned trend of average death rate and total deaths is not indicative that each leading cause affects all categories of people. As such, we’ll also be comparing the dataset sizes of each category to the original.

Leading Cause	Total Deaths	Average Death Rate Per 100,000
Diseases of Heart (I00-I09, I11, I13, I20-I51)	145354.0	199.351563
Malignant Neoplasms (Cancer: C00-C97)	105051.0	149.796875
All Other Causes	76672.0	109.010937
Assault (Homicide: Y87.1, X85-Y09)	2082.0	26.788889
Influenza (Flu) and Pneumonia (J09-J18)	18433.0	25.889063
Diabetes Mellitus (E10-E14)	13564.0	20.287500
Cerebrovascular Disease (Stroke: I60-I69)	12808.0	18.843750
Human Immunodeficiency Virus Disease (HIV: B20-B24)	5354.0	18.803448
Chronic Lower Respiratory Diseases (J40-J47)	13082.0	18.320635
Mental and Behavioral Disorders due to Accidental Poisoning and Other Psychoactive Substance Use (F11-F16, F18-F19, X40-X42, X44)	4191.0	13.764000
Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)	7361.0	12.383929
Essential Hypertension and Renal Diseases (I10, I12)	6896.0	11.863636
Alzheimer's Disease (G30)	2972.0	10.286957
Intentional Self-Harm (Suicide: X60-X84, Y87.0)	2030.0	8.952174
Chronic Liver Disease and Cirrhosis (K70, K73)	1464.0	8.217647
Septicemia (A40-A41)	627.0	6.957143
Nephritis, Nephrotic Syndrome and Nephrosis (N00-N07, N17-N19, N25-N27)	753.0	5.961538
Certain Conditions originating in the Perinatal Period (P00-P96)	37.0	3.400000
Congenital Malformations, Deformations, and Chromosomal Abnormalities (Q00-Q99)	14.0	2.800000
Viral Hepatitis (B15-B19)	15.0	2.500000

Fig 3 - Categorizing by Leading Cause and Sorting Values for Average Death Rate Per 100,000

b. Categorizing by Race/Ethnicity

Our next step was to identify the category of people by race/ethnicity and gender that are disproportionately affected by assault, HIV and MBD as a cause of death. To categorize by race/ethnicity we used the loc function to create a new dataset for every race. We then used the groupby function to categorize by cause and aggregated the total deaths and average death rate per 100,000 as the next two columns. When grouping for the Hispanic category, it would include Black Non-Hispanic and White Non-Hispanic as well. We had to remove rows containing these strings for the dataset to only contain ‘Hispanic’.

For the Asian Pacific Islander population, we found that none of the three causes were in the dataset showing us that they are not a leading cause for this category of people (Fig 4). The dataset for this

category is 17 rows by 7 columns, 3 fewer rows than the base dataset. They were therefore not affected by 3 leading causes in the original dataset.

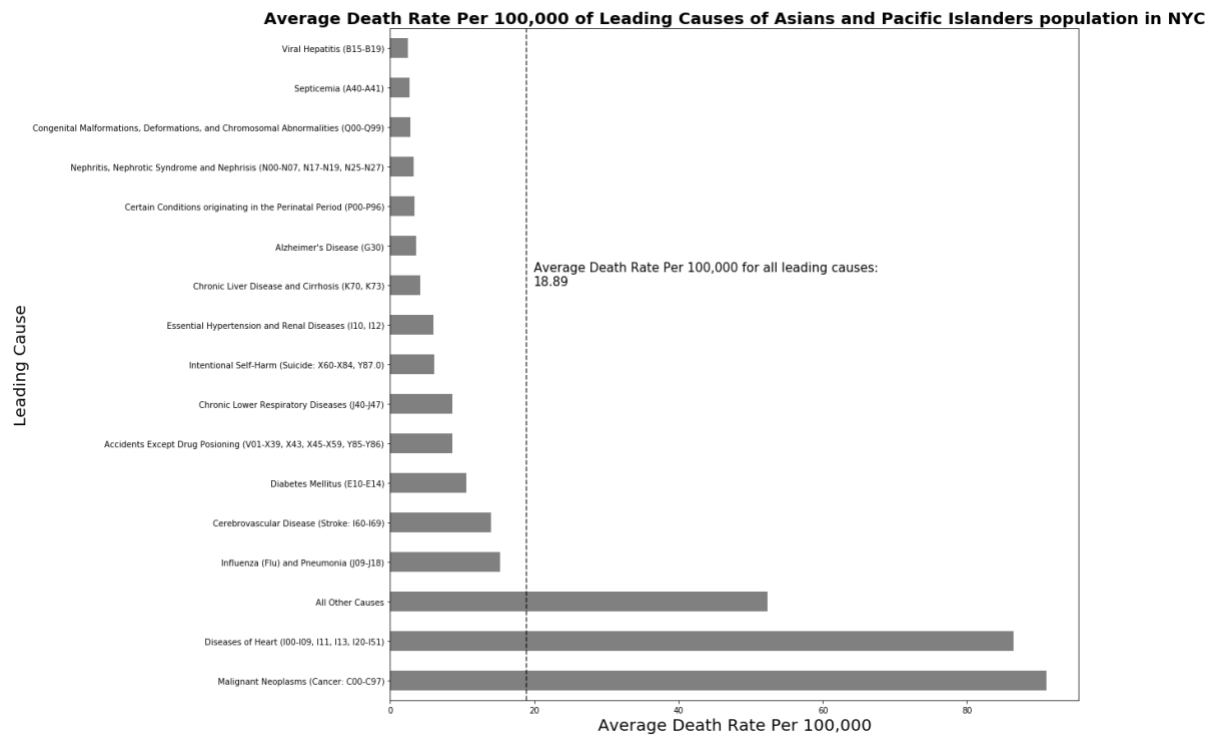


Fig 4 - Average Death Rate Per 100,000 of Leading Causes of Asians and Pacific Islanders in NYC

For the Black Non-Hispanic population, we found that all three causes affect this group. Notably, assault and HIV rank high in the Average Deaths Per 100,000. Upon closer analysis, we noticed that the total deaths caused by assault was out of the descending order when value sorting by Average Death Rate Per 100,000. This indicates that there is one gender is more heavily affected by these causes. This group was only affected by 15 of the 20 leading causes.

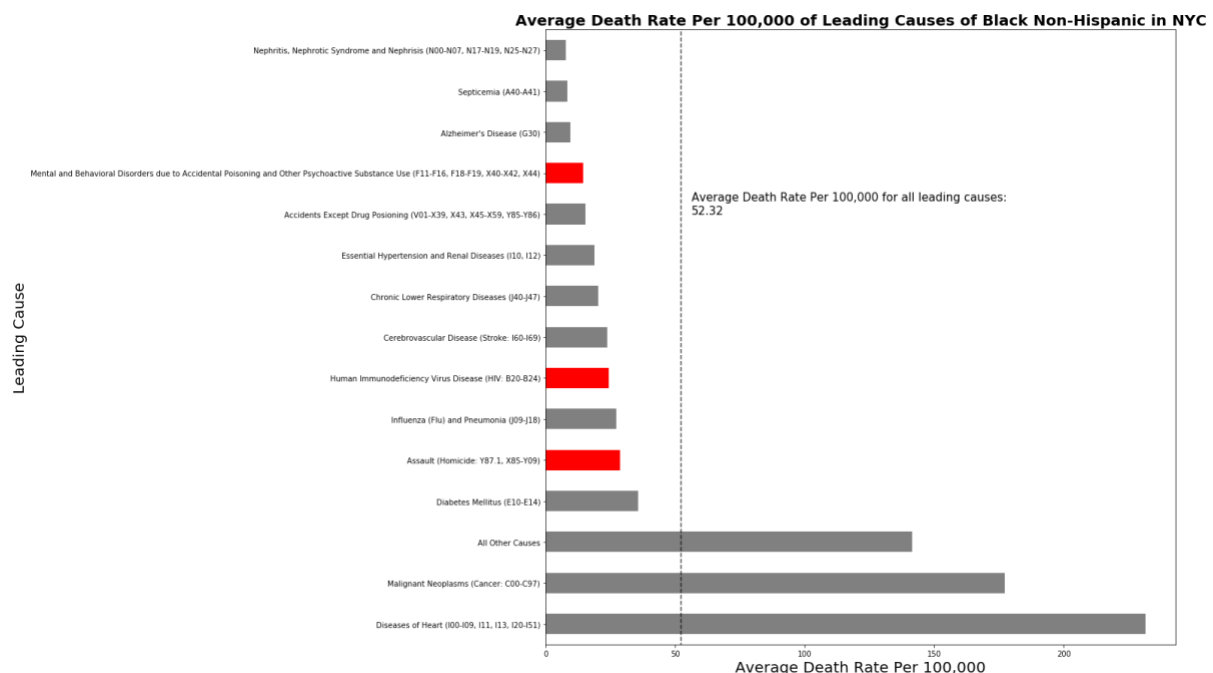


Fig 5 - Average Death Rate Per 100,000 of Leading Causes of Black Non-Hispanic in NYC

For the White Non-Hispanic population, we found that only MBD out of the three causes affects this group. The size of this dataset is also 14 rows by 3 columns which indicate that the group is not affected by 6 causes from the base dataset.

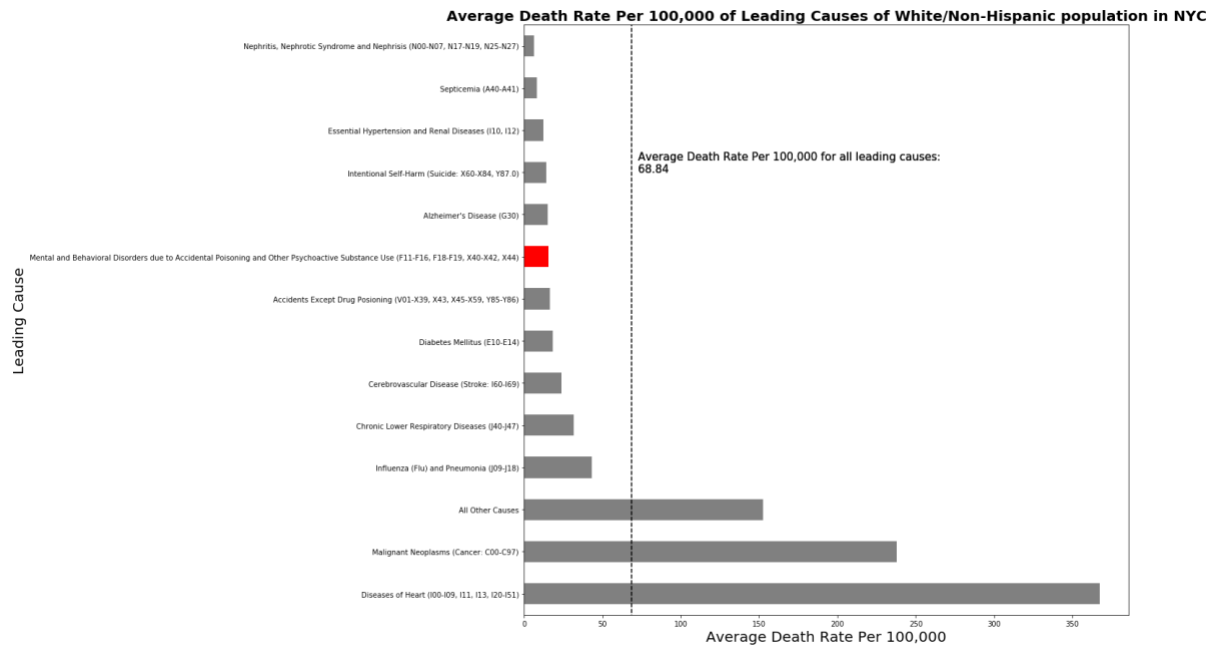


Fig 6 - Average Death Rate Per 100,000 of Leading Causes of White Non-Hispanic in NYC

For the Hispanic population, we found all three causes also affect this group. Looking at the dataset we also noticed that HIV and Assault are out of order when sorting the values of average death rate per 100,000 in descending order. Again, this signals that one gender is more heavily affected by these causes. This group was affected by 15 of the 20 leading causes.

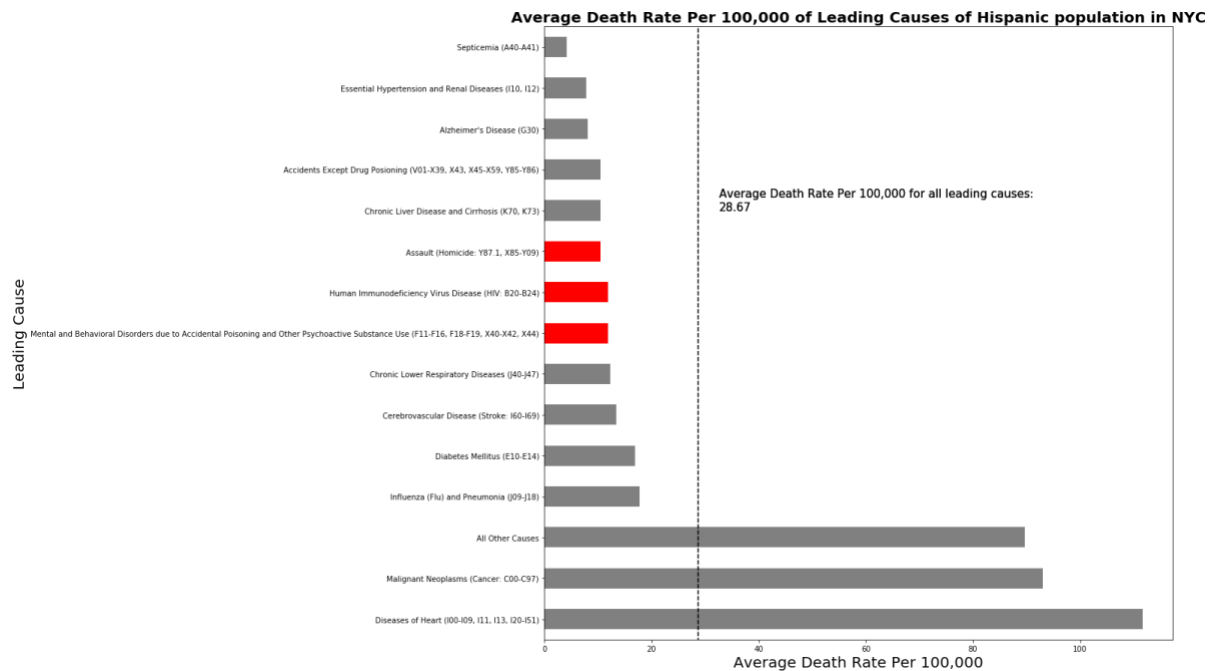


Fig 6 - Average Death Rate Per 100,000 of Leading Causes of Hispanic in NYC

c. Categorizing by Gender

To zero in on the specific group affected by these three causes, more specifically HIV and Assault, we categorized by gender to see how the leading causes affected them.

Females are affected by 18 of the 20 leading causes whereas males are affected only by 17 of them.

Comparing the graphs (Fig 7), we can clearly see that all the causes affect males significantly more than females, with Assault as a leading cause while at the same time completely absent from the female group.

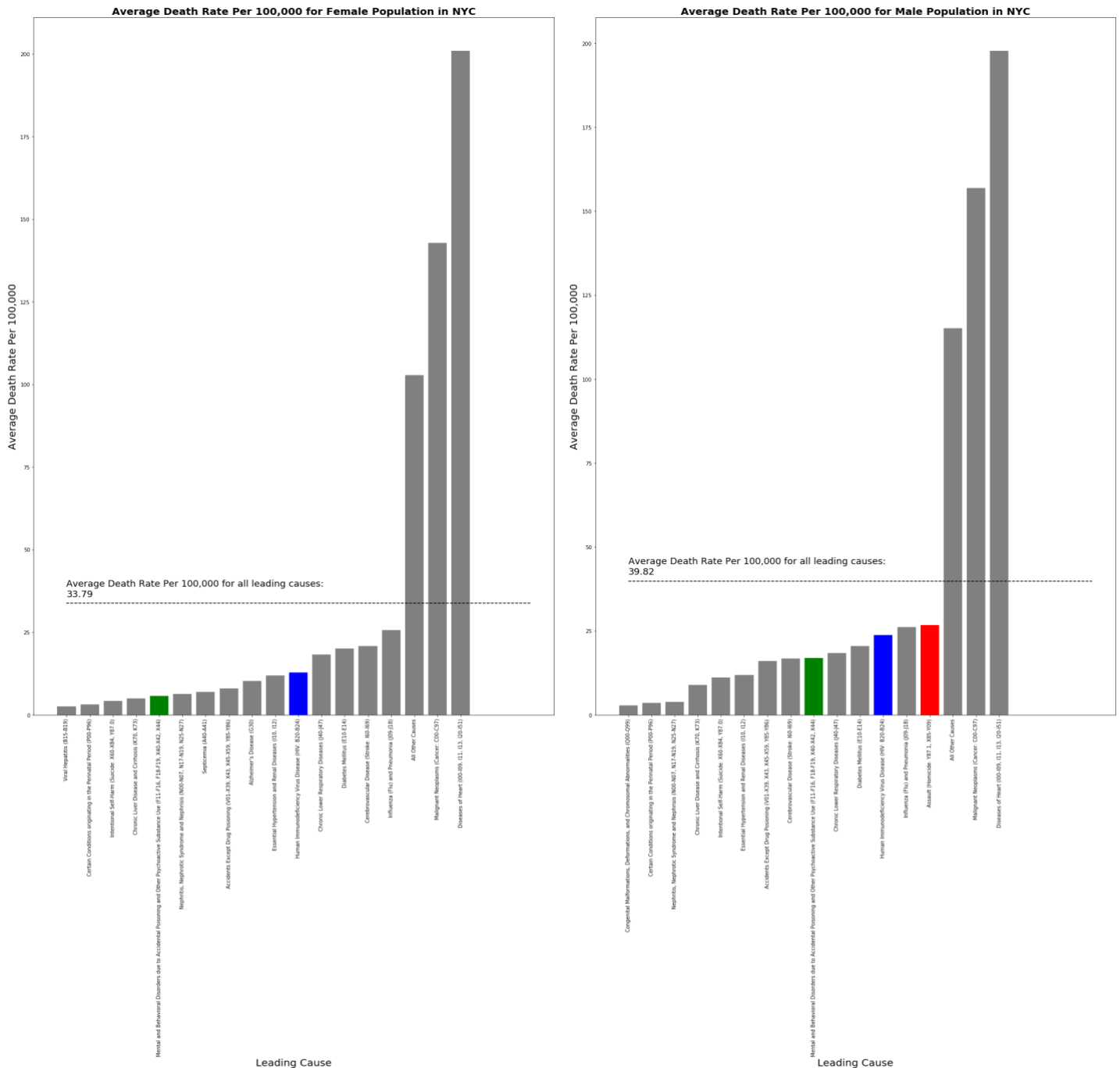


Fig 7 - Average Death Rate Per 100,000 of Leading Causes of Female (left) and Male (right) populations in NYC

d. Categorize by Race & Gender

Now that we have uncovered the two groups of people that are disproportionately affected by the three focus causes, we categorized them into separate datasets to formalize our findings.

Categorizing the dataset by Black Non-Hispanic males (see Fig 8). In relation to the three causes:

Leading Cause	Total Deaths	Average Death Rate Per 100,000
Diseases of Heart (I00-I09, I11, I13, I20-I51)	15887.0	232.587500
Malignant Neoplasms (Cancer: C00-C97)	12361.0	180.937500
All Other Causes	10228.0	149.687500
Diabetes Mellitus (E10-E14)	2308.0	33.787500
Human Immunodeficiency Virus Disease (HIV: B20-B24)	2259.0	33.100000
Assault (Homicide: Y87.1, X85-Y09)	1968.0	28.825000
Influenza (Flu) and Pneumonia (J09-J18)	1860.0	27.212500
Cerebrovascular Disease (Stroke: I60-I69)	1502.0	21.987500
Chronic Lower Respiratory Diseases (J40-J47)	1409.0	20.600000
Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)	1044.0	17.457143
Mental and Behavioral Disorders due to Accidental Poisoning and Other Psychoactive Substance Use (F11-F16, F18-F19, X40-X42, X44)	298.0	17.450000
Essential Hypertension and Renal Diseases (I10, I12)	1180.0	17.275000

Fig 8 - Leading Causes of Black Non-Hispanic Males in NYC from 2007-2014

Black Non-Hispanic males affected by assaults as a percentage of total Assault deaths
 $= (1968/2082) * 100 = 94.52\%$

Black Non-Hispanic males affected by HIV as a percentage of total HIV deaths
 $= (2259/5354) * 100 = 42.19\%$

Black Non-Hispanic males affected by MBD as a percentage of total MBD deaths
 $= (298/4191) * 100 = 7.11\%$

Categorizing the dataset by Hispanic males (see Fig 9). In relation to the three causes:

Leading Cause	Total Deaths	Average Death Rate Per 100,000
Diseases of Heart (I00-I09, I11, I13, I20-I51)	10433.0	114.112500
All Other Causes	9085.0	99.100000
Malignant Neoplasms (Cancer: C00-C97)	8892.0	97.112500
Influenza (Flu) and Pneumonia (J09-J18)	1640.0	17.887500
Diabetes Mellitus (E10-E14)	1499.0	16.375000
Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)	1378.0	15.025000
Mental and Behavioral Disorders due to Accidental Poisoning and Other Psychoactive Substance Use (F11-F16, F18-F19, X40-X42, X44)	1335.0	14.587500
Human Immunodeficiency Virus Disease (HIV: B20-B24)	1304.0	14.375000
Chronic Liver Disease and Cirrhosis (K70, K73)	1148.0	12.500000
Cerebrovascular Disease (Stroke: I60-I69)	1132.0	12.350000
Chronic Lower Respiratory Diseases (J40-J47)	934.0	11.557143
Assault (Homicide: Y87.1, X85-Y09)	114.0	10.500000

Fig 8 - Leading Causes of Hispanic Males in NYC from 2007-2014

Hispanic males affected by assaults as a percentage of total assault deaths
= $(114/2082) * 100 = 5.48\%$

Hispanic males affected by HIV as a percentage of total HIV deaths
= $(1304/5354) * 100 = 24.36\%$

Hispanic males affected by MBD as a percentage of total MBD deaths
= $(1335/4191) * 100 = 31.85\%$

V. CONCLUSION (HYPOTHESIS 1)

Our hypothesis that the leading causes would affect all races, ethnicities and gender similarly was incorrect, which we can see from not only the size of the datasets but also the detailed analysis of the three targeted causes we conducted. No category of people was affected by all 20 leading causes from the base dataset. Females were affected by the greatest number of leading causes at 18 causes while White Non-Hispanic people were affected by the smallest number of leading causes at 14.

For Assault, Black Non-Hispanic males are significantly disproportionately affected, making up a staggering 94.52% of all Assault deaths in New York City from 2007-2014.

For HIV deaths, Black Non-Hispanic males make up 42.19% of all deaths caused by HIV in New York City from 2007-2014 with Hispanic males coming in second at 24.36%.

MBD significantly disproportionately affects Hispanic males, with them making up 31.85% of all deaths caused by MBD in New York City from 2007-2014.

We found that as a leading cause, assault exclusively affected these groups of people.

VI. LEADING CAUSES OF DEATH ACROSS TIME ANALYSIS (HYPOTHESIS 2)

a. Approach

To organize the causes of death by Leading Cause, we created new data frames using the `loc` function to isolate the individual leading causes of Cancer, Assault, and Heart Disease using a `str.contains` function on the Leading Cause column. We then used the `groupby` and `agg` functions to sum up the Total Deaths column and get the mean for the Average Death Rate Per 100,000 by year to see how the total deaths and death rates changed for each of the causes over time. We were also curious as to the average deaths and death rates per year over the 8-year period so that we would have a frame of reference when analyzing our data.

b. Observations and Explanations

i. Assault (Homicide) Trend

In 2008, after a series of deadly assaults on traffic officers, New York City enacted a law that made the assault of traffic agents a felony [2]. This coincides with a sharp decrease in deaths by assault in 2008 and our data shows that death rates have not gone up to pre-enactment levels since.

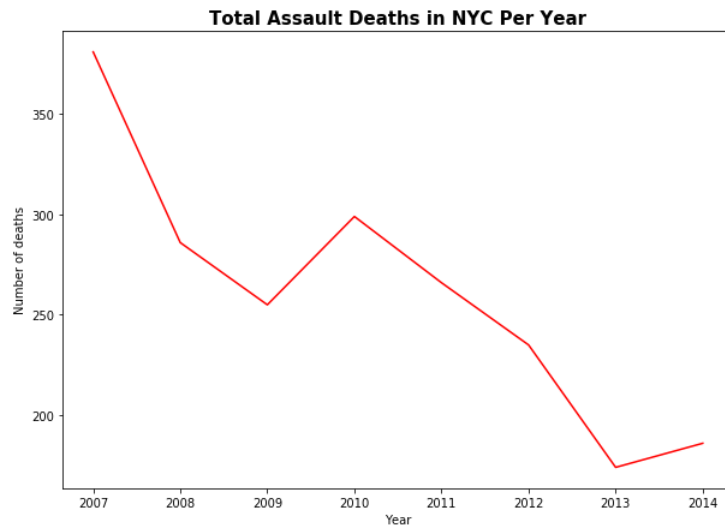


Fig 9 - Total Assault Deaths in NYC Per Year

We see another notable drop in death rates from assault in 2013. Following the Sandy Hook Elementary School shooting earlier that year in Connecticut and the Webster, New York shooting the year before, the New York Secure Ammunition and Firearms Enforcement Act (NY Safe Act) was enacted [3]. It is a very tight gun regulation law that broadened the legal definition of assault weapons to include semi-automatic firearms and had a high-capacity magazine provision to limit the number of rounds of ammunition a gun could legally hold. The Act also required ammunition dealers to conduct more stringent background checks. We believe this law to be one of the leading causes for the sharp decrease in 2013.

Generally, we have seen a decrease in death rates across the leading causes since 2007, partially due to increased regulation on assault and tightened gun laws. In terms of the health-related deaths, we attribute the decrease in death rates to an increase in health consciousness among society and more effective ways in treating medical illnesses.

The cause with the largest decline was death from cardiovascular diseases. We attribute this to NYC's regulations regarding the taxation, regulation, marketing of high-risk foods, drinks, and habits. For example, New York City has enacted a comprehensive tobacco control program that requires smoke-free workplaces, an increase in both the tax on cigarettes and the presence of anti-tobacco advertising [4]. It has also promoted healthy eating behaviors and diets through policies aiming to discourage fast food purchases and increase the accessibility of nutritious foods.

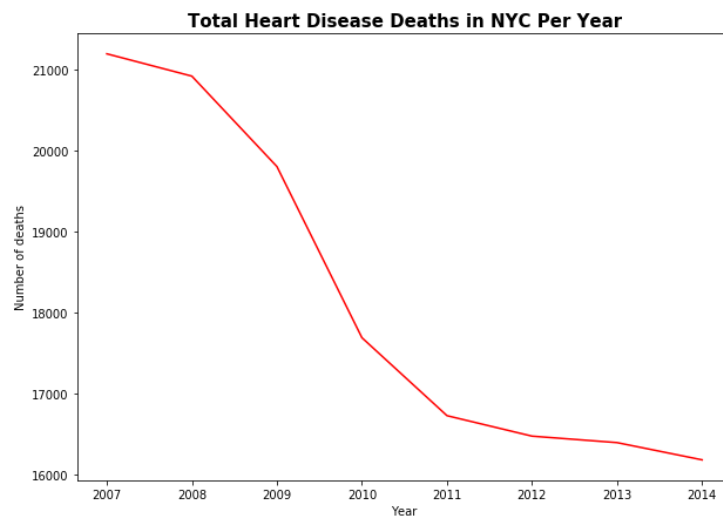


Fig 10 - Total Heart Disease Deaths in NYC Per Year

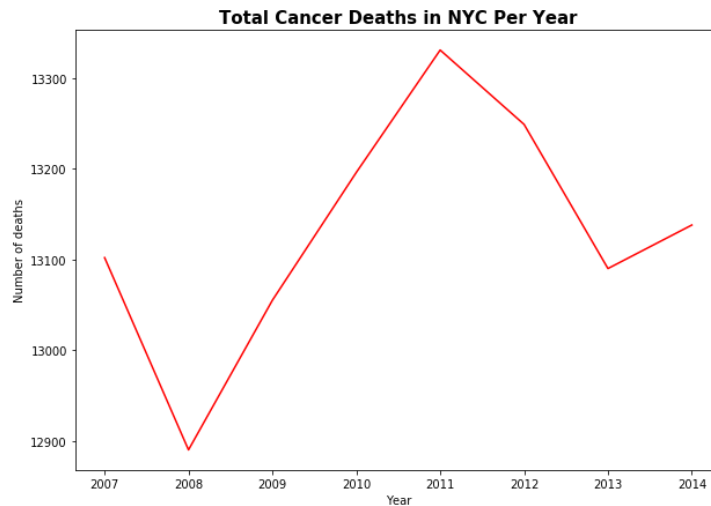


Fig 11 - Total Cancer Deaths in NYC Per Year

VII. CONCLUSION (HYPOTHESIS 2)

We found some startling results in our time series analysis. For one, we found that the death rates for Cancer remained basically the same over the 8-year period, signaling that despite the push for health-consciousness in New York City, death rates relating to cancer stayed relatively the same. However, we found that Total Deaths due to both Assault and Heart Disease were on a decline. The average number of deaths per year due to Assault was 231, but we found that the number of deaths declined from 381 to 186 in the 8-year period, equating to a $(381-186)/381 * 100 = 51.18\%$ decrease in assault-related deaths. The average Death Rate Per 100,000 due to Heart Disease was 199, but it declined from 231 in 2007 to 178 in 2014, equating to a $(231-178)/231 = 22.94\%$ decrease in the death rate due to Heart Disease. All in all, we have evidence to support that increased regulation over the time period has led to a much safer New York City due to the significant decline in Assault-related deaths. An increase in health-consciousness has also led to a lower death rate for Heart Disease-related deaths, a testament to the effectiveness of New York City's promotion of healthier eating habits.

LINK TO CODE:

<https://github.com/ChrisCai8/Data-Bootcamp-Final-Project>

BIBLIOGRAPHY

1. Calgary, Open. "New York City Leading Causes of Death." *NYC Open Data*, data.cityofnewyork.us/Health/New-York-City-Leading-Causes-of-Death/jb7j-dtam.
2. "NYC Makes Assaulting a Traffic Officer a Felony." *NBCNews.com*, NBCUniversal News Group, 25 July 2008, www.nbcnews.com/id/25851929/ns/us_news-crime_and_courts/t/nyc-makes-assaulting-traffic-officer-felony/#.XBfj-s9KjUo.
3. "Guide to The New York Safe Act | S.C.O.P.E. - Shooters Committee On Political Education - Dedicated Dedicated to Preserving the 2nd Amendment Rights for the Residents of New York State." *SCOPE NY*, www.scopeny.org/Guide-to-The-New-York-Safe-Act.
4. "Reducing Cardiovascular Disease Death Risk in New York City - School of Public Health - University of Minnesota." *School of Public Health*, 8 Jan. 2018, www.sph.umn.edu/news/new-york-city-cardiovascular-disease-death-risk-declines-following-aggressive-public-health-interventions/.