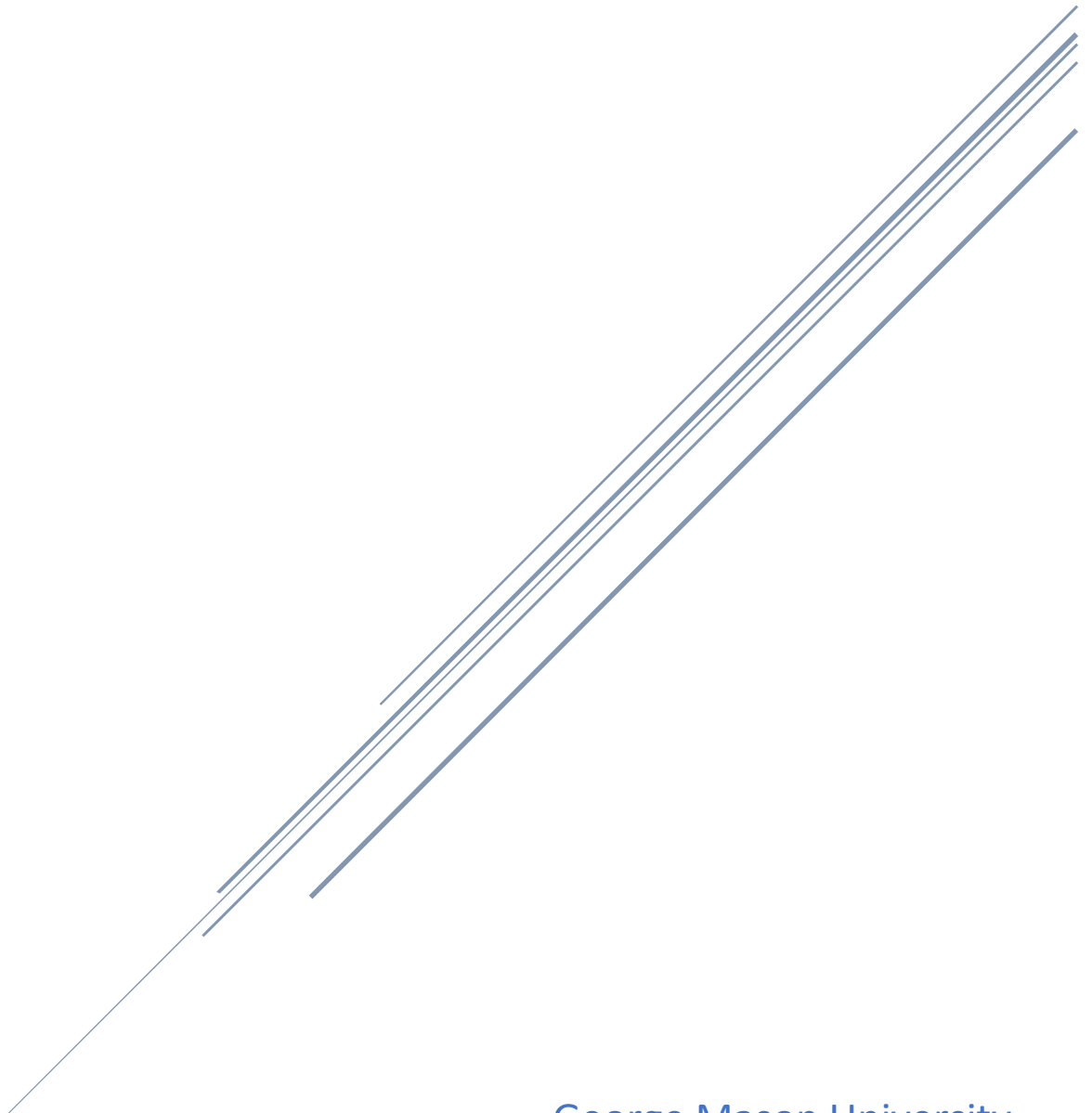# ANALYZING LEADING CAUSES OF DEATH IN NEW YORK CITY

By: Nafisa Ahmed & Ubaid Tariq

Analyzing Leading Causes of Death in New York City

1.0   Abstract

The targeted goal for this research is to get an overall picture of the main causes of death in a big metropolitan city like New York. New York City is also known as the human jungle due to its population and cultural diversity. People from almost every nation reside in New York. Population is getting bigger in numbers with passage of time as well as it is quite balanced with the mortality rate. According to this research leading causes of deaths in this city are multivarious, and it quite depends upon from which ethnicity a deceased person belongs. This research focuses on the 10 leading causes of deaths among all the ethnicities present in the city. As well as studying what the specific causes of death are among specific genders and ethnicities. Similar research has also been conducted in Australia as well. Different data has been analyzed on large scale by applying multiple filters of age, gender and what causes are consistent among the patients.

2.0   Introduction

This research was conducted to investigate the top 10 leading causes of death in New York City which greatly impact the mortality rate of the city. This study has been based on multiple ethnicities since the city residents are multicultural. For the past 3 decades diseases such as: Heart related, Cancer and accidents remained top on chart for prominent causes for mortality.

According to Kimberley Holland, who analyzed the causes of death collected by the Centers for Disease Control and Prevention (CDC), Heart related diseases are the most common among all age groups, especially more consistent among older individuals. The main cause of heart disease among people may be due to their lifestyle and daily diet choices. Cancer is the second leading cause that heavily effects mortality rate. This may be due to smoking being very common among people of all ages, and it is one of the leading causes for Cancer. Since cigarette ingredients mostly consist of carcinogenic elements that play a vital role in contributing to this disease.

Further deep analysis of the data proposed that accidents are the third leading cause of death in the United States. These are more common among the young people, that are related to motor vehicle related deaths like driving under the influence. According to the Centers for Disease Control and Prevention (CDC) there are 161,374 deaths alone each year by traffic accidents and there are around 28 million emergency room visits due to this cause every year (Holland, 2019).

Respiratory diseases are also another common leading cause of death in the United States. This is mainly due to secondhand smoke, poor life standards and common among people with preexisting conditions (Holland, 2019). According to analyzed data from the CDC chronic respiratory diseases is the fourth leading cause of death, followed by Stroke as the fifth cause of death, which is more common among men as compared to women. Alzheimer's disease is very common now a days among older individuals, but doctors believe that genetics plays vital role in

the cause of this disease ((Holland, 2019). Diabetes, influenza, kidney related diseases and suicide are major contributors among the causes which impacts mortality rate in the United States.

It's important now more than ever to know what the causes of death for individuals in specific cities/states are. Especially, since the covid pandemic has caused more deaths than anything else at the current moment. By studying the leading causes of death, researchers would be able to measure the population health of a specific city or state. This study focuses on what the leading causes of death in New York City were since 2007. By utilizing the data collected by the Department of Health and Mental Hygiene (DOHMH), this paper will explore the following research questions:

- What is the leading cause of death in New York City?
- What is the leading cause of death among males and females in New York City?
- Which Ethnicities had the most/least reported death counts in New York City?
- Were there specific causes of deaths that occurred only within specific ethnicities and not others?
- Have leading causes of death changed since 2007?

3.0   Literature Review

In the article, "Leading Causes of Death, New York, 2012", Mark Sharp along with the other authors have analyzed what the leading causes of death in New York were using the New York State Vital Statistics mortality data. In this article, the authors were able to analyze the top 10 leading causes of death in New York in the year 2012. The top 10 causes accounted for 75.7% of all the deaths in New York State. The top two causes were malignant neoplasms and diseases of the heart that accounted for 54% of the causes of death in New York State (Sharp et al., Leading Causes of Death, New York State, 2012). This article explored what the leading causes of death were by Age, Sex, and Race in 2012. The leading causes of death by Age were shown in a table for each Age Group. To summarize that table, accidents were the leading cause of death among the younger people. This accounted for around 20% to 29% of deaths among people ranging in Age 1 to 44 (Sharp et al., Leading Causes of Death, New York State, 2012). When analyzing the leading causes of death by Sex, heart disease and cancer were among the top causes of death in New York State among both men and women. Among men, the leading causes of death in 2012 were accidents and suicide, while hypertension and Alzheimer's are ranked the top causes of death among women. When analyzing the data by race, cancer, and heart diseases were the leading causes of death among Hispanics, Whites, and Blacks. This article confirms and duplicates the research questions that we are planning to analyze for our dataset. It relates to the research questions we are planning to analyze like the leading causes of death by sex, age, and ethnicities. In our research, we will be analyzing the leading causes of death from 2007 to the present for New York City, whereas this article analyzed New York State's leading causes of death specifically for the year 2012 and compared it to previous years.

The Australian Government's Institute of Health and Welfare conducted a similar study to analyze what the leading causes of death in Australia were to measure population health. In the article, "3.2 Leading causes of death", Australia had analyzed deaths from the year 2016,

where there was a total of 158, 500 deaths, 81, 900 of these deaths were males whereas 76, 700 of the deaths were females (3.2 leading causes of death, chapter 3 causes of ill health (Australia's Health 2018), 2018). When analyzing what the leading causes of death by sex were, among males it was coronary heart disease which accounted for 13% of the deaths in 2016. The leading causes of death for women were Alzheimer's and dementia, which accounted for 11% of the deaths in 2016. These results were similar to what Sharp et al., found in their study when analyzing the leading causes of death by sex in New York State for the year 2012. Other similar results Australia had gathered when analyzing leading causes of death by age were from the age of 1-44, among younger people, the leading causes of death were accidents and suicide. The leading causes of death among middle-aged individuals (ages 45-64) were predominantly, lung cancer and coronary heart diseases. For individuals older than 75, the leading cause of death was mainly Alzheimer's and dementia. This article, although not New York State specific, confirms and duplicates some of the research questions that we were planning to analyze for our dataset. Our research and analysis will be specific to New York City whereas this research and analysis were specific to the country of Australia. When analyzing our research questions, we would be able to see how our results, that are New York City specific compares to Australia and be able to see if leading causes of death by age and sex are similar.

According to Farida Ahmad and Richard Anderson's article "The Leading Causes of Death in the US for 2020", there was an estimated 17.7% increase in the number of deaths that occurred in the year 2020. Covid-19 was a strong factor in the increase of deaths since it was the third leading cause of death along with heart disease and cancer (Ahmad & Anderson, 2021). The data that was used for this article was collected from the National Center for Health Statistics (NCHS) National Vital Statistics System (NVSS). 2020 in the United States, also saw a 4.8% increase in deaths due to heart diseases, deaths due to suicide, and chronic lower respiratory diseases decreased by 5.6% and 3.4% respectively from 2019 to 2020 (Ahmad & Anderson, 2021). This article and the analysis somewhat differ from our research questions because we are trying to analyze specific trends between sex, age, and ethnicities whereas this article had analyzed the trends of what the leading causes of death for 2020 were in the United States.

4.0   Data Cleansing/Methods

The data was collected by the Department of Health and Mental Hygiene (DOHMH). We downloaded the data (CSV file) directly from the NYC Open Data website. The data set includes the causes of death by Ethnicity and Sex since 2007. The variables that were included in this data set included Year, Leading Cause, Sex, Race Ethnicity, Deaths, Death Rate, Age Adjusted Death Rate. The data has four different ethnicity groups: Asian and Pacific Islander, Non-Hispanic Black, Hispanic, and Non-Hispanic White. This data set also included options where the individual's ethnicity was not known, those options were; Not Stated/Unknown and Other Race/Ethnicity. Originally the data set included 1,516 observations/rows.

The first step in this study was to clean the data. Python was the main software used to clean the data. The first step was to make sure the variable "Sex" only had two options, which were "Female" and "Male". In the original CSV file, there were four options for the "Sex" variable, which were, Female, Male, F, and M. Having those four options were confusing, so we

had used python to replace any observation that had "M" or "F" replaced to either "Male" or "Female". Python was also used to update the "Leading Cause" column, to remove the code in the Leading Cause name for each observation, to just have the name of the cause of death in that column. The next step was to load the CSV file into R and remove any missing/null values. After removing the missing/null values and creating a new data frame we are left with 973 observations/rows with the following ethnicities: Asian and Pacific Islander, Non-Hispanic Black, Hispanic, and Non-Hispanic White. By removing the missing/null values, R also removed the Not Stated/Unknown and Other Race/Ethnicity for the "Race Ethnicity" variable. R was mainly used to remove the any missing/null values and data visualization for our research questions. In addition, mySQL workbench was used to pull some simple queries and create a table schema.

5.0   Results/Analysis

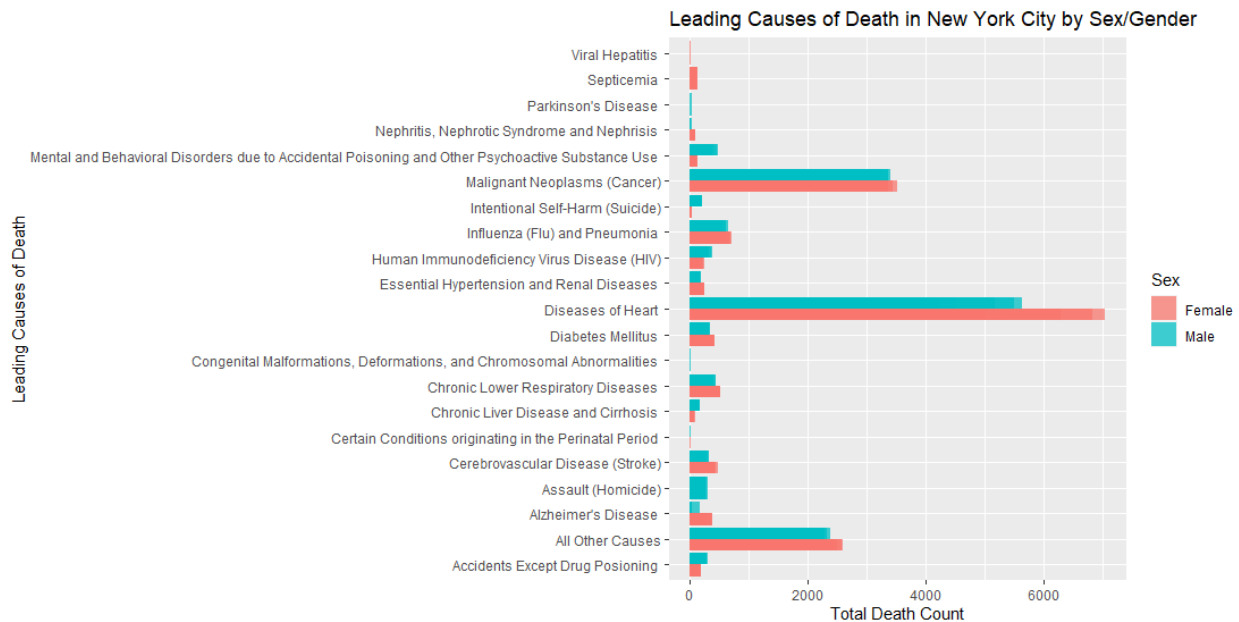### 5.1   *What is the leading cause of death in New York City?*
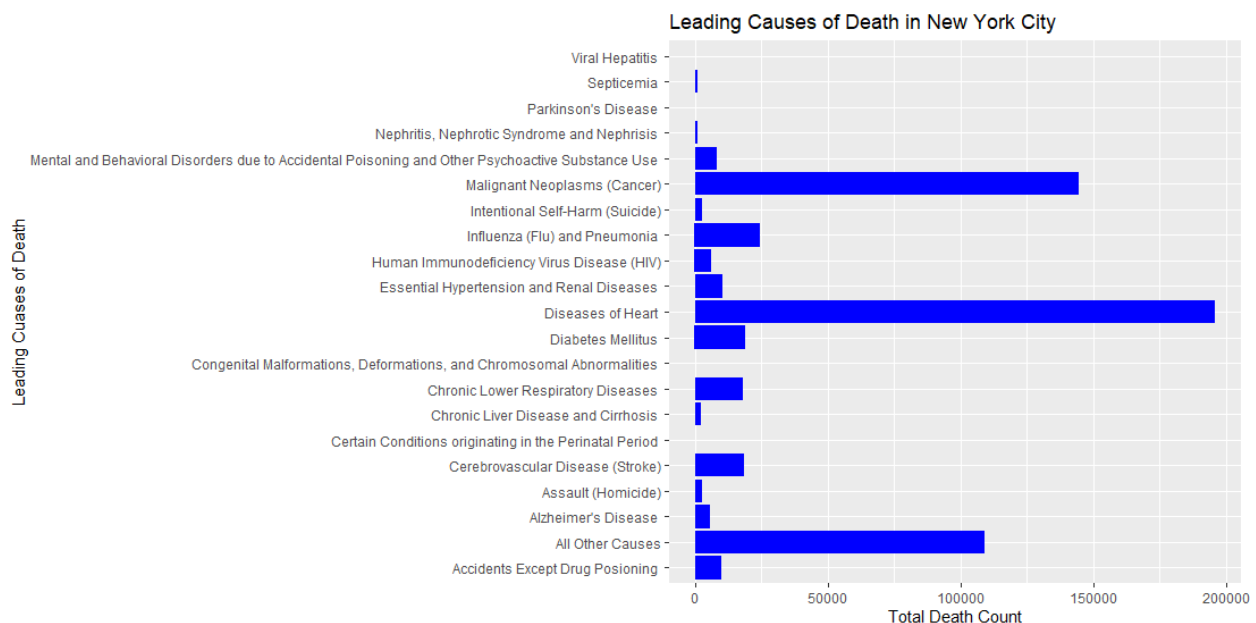
Figure 1: Leading Causes of Death in New York City

Our first research question was "What is the leading Cause of Death in New York City?". The above graph/visualizations represents that the main leading cause of death in New York City was Diseases of the heart. It shows that Diseases of the Heart killed a little over 180,000 people in New York City since 2007. The second leading cause of death in New York City was Malignant Neoplasms (Cancer). Malignant Neoplasms (Cancer) caused the death of a over 140,000 New York City individuals. Diseases that were a major cause of death for New York City individuals included, Viral Hepatitis, Parkinson's Disease, Congenital Malformations, Deformations, and Chromosomal Abnormalities, Certain Conditions originating in the Perinatal Period. Although these diseases caused the deaths of a few New York City individuals, they did not have a death count as big as the other diseases. Our graph suggests that Diseases of the Heart,

Malignant Neoplasms (Cancer), All other Causes, were more common than any of the other diseases in New York City.

5.2 What is the leading cause of death among males and females in New York City?

Figure 2: Leading Causes of Death in New York City by Sex/Gender

Our Second research question was *"What is the leading cause of death among males and females in New York City?".* In the above graph, we have represented the data as a side-by-side bar graph, where for each disease we have the total death for males and females. According to our graph, the leading cause of death for both males and females in New York City since 2007 was Diseases of the Heart. Although more females died from the heart disease than males. About 7,000 females had died of Diseases of the Heart in New York City since 2007, whereas about 5,500 males had died of Heart Diseases since 2007. Our graph also suggests among the leading causes of death: Diseases of the Heart, Malignant Neoplasms (Cancer), and All Other Causes, more females had died from each of these diseases than males. The Diseases that had the least death count for females in New York City were, Viral Hepatitis, Intentional Self-Harm (Suicide), Chronic Liver Disease and Cirrhosis, and Certain Conditions originating in the Perinatal Period, whereas for males the diseases that had the least death count in New York City were Parkinson's Disease, Congenital Malformations, Deformations, and Chromosomal Abnormalities, and Certain Conditions originating in the Perinatal Period. After our data cleansing, diseases like Viral Hepatitis and Septicemia had no male death count.

5.3   Which Ethnicities had the most/least reported death counts in New York City?

Highest/Least reported Deaths in New York City by Race/Ethnicity
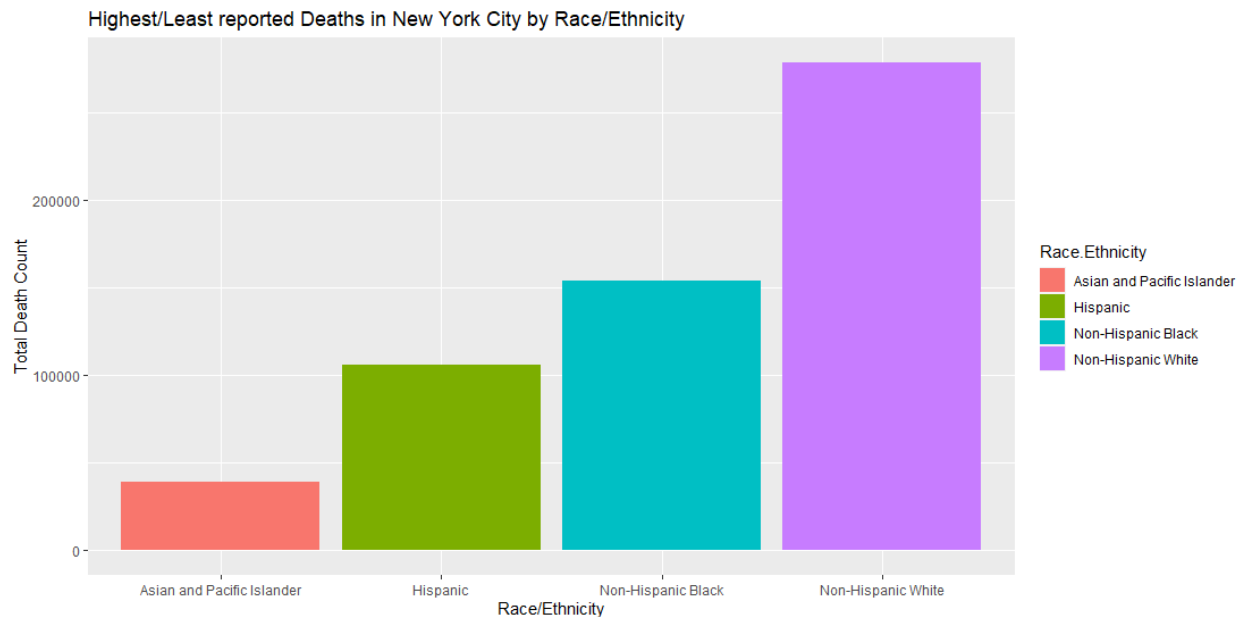
Figure 3: Highest/Least reported Deaths in New York City by Race/Ethnicity

Our third research question was *"Which Ethnicities had the most/least reported death counts in New York City?".* To answer this question, we created a bar graph in R, that gave us the result above. Our graph represents that the Non-Hispanic White ethnicity had the Highest reported Death count in New York City since 2007. This was followed by the Non-Hispanic Black ethnicity as having the second highest reported death count in New York City since 2007, with just a little over 150,000 deaths. This was followed by the Hispanic Ethnicity with having a death count just over 100,000 deaths in New York City. Our graph represents that the Asian and Pacific Islander ethnicity had the least reported death counts in New York City since 2007, with a death count just a little under 50,000 deaths.

5.4   Were their specific causes of deaths that occurred only within specific ethnicities and not others?
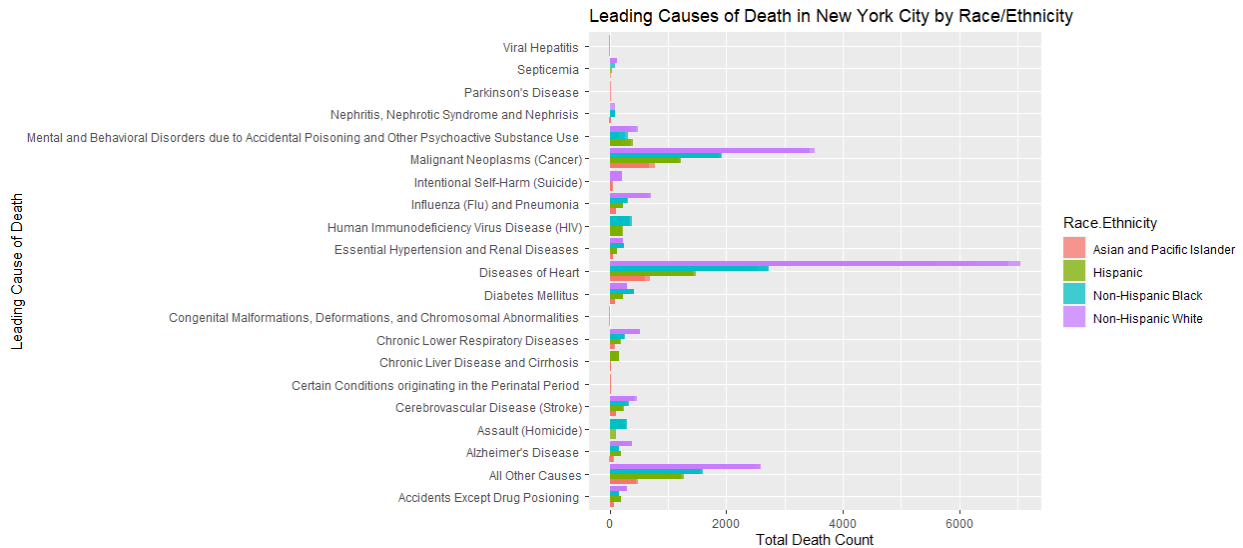
Figure 4: Leading Causes of Death in New York City by Race/Ethnicity

Our fourth research question was *"Were there specific causes of death that occurred only within specific ethnicities and not others".* To answer this question, we created a side-by-side bar graph that represents the leading causes of death and total death count for each race/ethnicity. The Leading Causes of Death in New York City by Race/Ethnicity chart suggests that more Non-Hispanic White individuals died from Diseases of the Heart than any other ethnicity, followed by the Non-Hispanic Black ethnicity and then the Hispanic Ethnicity. The Asian Pacific Islander Race/Ethnicity died more from the Malignant Neoplasms (Cancer), than any other disease. The top three leading causes of death for each ethnicity are as follows:

- Asian and Pacific Islander: Malignant Neoplasms (Cancer), Diseases of the Heart, and All other Causes.
- Hispanic: Diseases of the Heart, Malignant Neoplasms (Cancer), and All Other Causes.
- Non-Hispanic Black group were Diseases of the Heart, Malignant Neoplasms (Cancer), and All Other Causes.
- Non-Hispanic White: Diseases of the Heart, Malignant Neoplasms (Cancer), All Other Causes.

Our chart also suggests that the Viral Hepatitis, Parkinson's Disease, Congenital Malformations, Deformations, and Chromosomal Abnormalities Disease, and Certain Conditions Originating in the Prenatal Period were only common among people of Asian and Pacific Islander descent. Whereas diseases like Septicemia, Nephritis, Nephrotic Syndrome and Nephrisis, Mental and Behavioral Disorders due to Accidental Poisoning and other Psychoactive Substance User, Human Immunodeficiency Virus (HIV) and Assault were not that common among the Asian and Pacific Islander group.

Causes of Death like Viral Hepatitis, Parkinson's Disease, Nephritis, Nephrotic Syndrome and Nephrisis, Intentional Self-Harm (suicide), Congenital Malformations, Deformations, and Chromosomal Abnormalities Disease, Certain Conditions Originating in the Prenatal Period were not common among the New York City Individuals who were of Hispanic Ethnicity. Whereas causes of death like Mental and Behavioral Disorders due to Accidental Poisoning and other Psychoactive Substance User, Chronic Liver Disease and Cirrhosis, and Assault (Homicide) were common among people of Hispanic Ethnicity.

For people of Non-Hispanic Black descent, other common causes of death included: Septicemia, Mental and Behavioral Disorders due to Accidental Poisoning and other Psychoactive Substance User, Influenza (flu) and Pneumonia, Human Immunodeficiency Virus (HIV), Essential Hypertension and Renal Disease, Diabetes Mellitus, Chronic Lower Respiratory Diseases, Cerebrovascular Disease (Stroke), Assault (Homicide), Alzheimer's Disease, and Accidents except Drug Use. Causes of Death like Viral Hepatitis, Parkinson's Disease, Intentional Self0Harm (Suicide), Congenital Malformations, Deformations, and Chromosomal Abnormalities Disease, Chronic Liver Disease and Cirrhosis, Certain Conditions Originating in the Prenatal Period were not common among New York City individuals from Non-Hispanic Black ethnicity.

For people of Non-Hispanic White decent, other common causes of death are as follows: Septicemia, Nephrotic Syndrome and Nephrisis, Mental and Behavioral Disorders due to Accidental Poisoning and other Psychoactive Substance User, Intentional Self-Harm (Suicide), Influenza (flue) and Pneumonia, Essential Hypertension and Renal Disease, Diabetes Mellitus, Chronic Lower Respiratory Disease, and Cerebrovascular Disease (Stroke), Alzheimer's Disease, and Accidents Except Drug User. Some uncommon causes of death in New York City for individuals with Non-Hispanic White ethnicity were: Viral Hepatis, Parkinson's Disease, Human Immunodeficiency Virus (HIV), Congenital Malformations, Deformations, and Chromosomal Abnormalities Disease, Chronic Liver Disease and Cirrhosis, Certain Conditions originating in in the Perinatal Period, and Assault (Homicide).

5.5 *Have leading causes of death changed since 2007?*
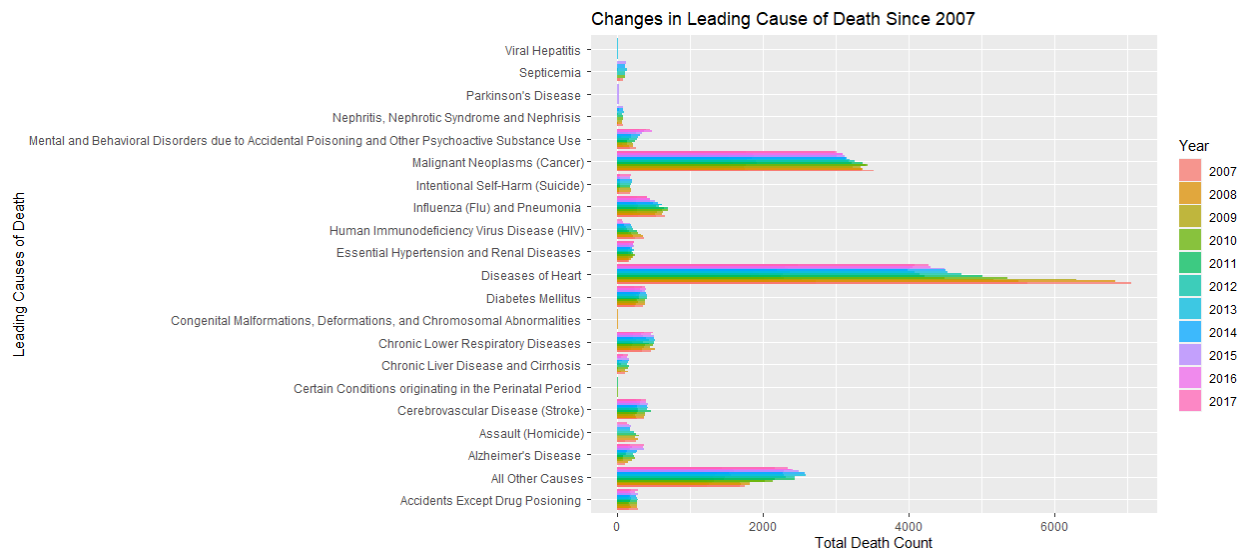
Figure 5: Leading Causes of Death Since 2007

Our fifth research question was, *"Have leading causes of death changed since 2007?"*. To answer this question, we created a side-by-side by graph in R, where we have the leading causes of death and total death count for each year in New York City since 2007. Our chart suggests that the number of deaths caused by Diseases of the Heart, has been exponentially decreasing since 2007 in New York City. Deaths from Heart Disease decreasing, could be due to individuals in New York City taking on a healthier lifestyle. The death count for "All Other Causes" has been steadily increasing since 2007. Causes of Death that have been steadily decreasing since 2007 in New York City include: Human Immunodeficiency Virus (Virus), Malignant Neoplasms (Cancer), and Assault (Homicide). Causes of Death that have been increasing since 2007 are Alzheimer's Disease. Some of the causes of death that have remained somewhat steady with no drastic increase/decrease in death include Intentional Self-Harm (Suicide), Essential Hypertension and Renal Disease, Diabetes Mellitus, Chronic Lower Respiratory Disease, Chronis Liver Disease and Cirrhosis, Cerebrovascular Disease (Stroke), and Accidents Except Drug User.

6.0 Limitations

Some limitations we ran into while doing this study were, we had no way of analyzing whether the causes of death related to any specific age group. In the future, it would be beneficial to collect the age of the individual in order to analyze, if specific leading causes of death were related to a specific age group. By having that variable in the data set researchers would be able to analyze if there are specific causes of death only affected by specific age groups and be able to reason as to why this might be. Another limitation that we ran into, is that after cleansing the data of any missing/null values we only had data specifically from 2007 to 2017, data from the 2017 and later was dropped since there was no way to accurately analyze that data. In the future, I think having accurate data for the later years in order to get an accurate representation of leading

causes death in New York City since 2007 would beneficial. Another limitation was the lack of study/research specifically in New York City. In our Literature review, we were not able to find any research on leading causes of death specific to New York city, we were able to find one research specific to the New York State. While there has been a lot of study in the past in the medical and health sector, it is extremely difficult to cover every cause of death, while the data set, we analyzed did not cover the causes of death holistically, this is a start and collecting more variables like the Age of the individual when they had passed away would be a step closer. Another limitation was that all causes related to the heart was grouped into one "Diseases of the Heart", it would have been helpful to know specifically what kind of heart disease had the highest death count.

7.0   Conclusion

In conclusion, this research/study was conducted in order gain insights into what the causes of death are in New York City, especially among specific genders and ethnicities. From our study and analysis, we were able to find that Disease of the Heart is the Leading Cause of Death In New York City and has been the leading cause since 2007. Diseases of the Heart is also the Leading Cause of Death among both Females and Males, whereas this specific disease has a higher death count among females. Diseases of the Heart was also the Leading Cause of Death among all four ethnicities in New York City since 2007. Based on our findings, even though Diseases of the heart has been the leading cause of death since 2007, it has also been steadily decreasing every year since 2007. This may be due to the people of New York City adopting a healthier lifestyle/choice. Heart disorders are the most frequent causes of death, they affect people of all ages and ethnicities. Blood pressure, which builds up when a person reaches a certain age, is one of the most prevalent causes of heart disease. This, in turn, produces further difficulties, which eventually leads to heart failure, also known as a heart attack.
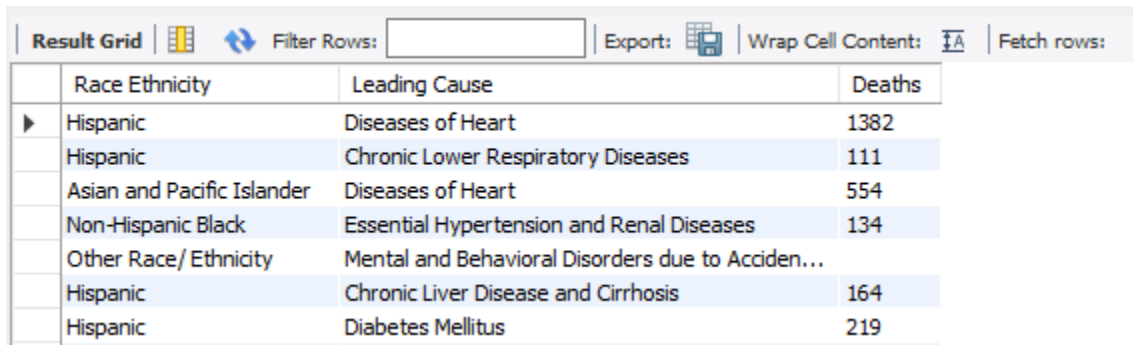
After Heart Diseases, Malignant Neoplasms (Cancer) was the second leading cause of death among New York City individuals since 2007. Cancer has been second leading cause of death among both males and females in New York City, but females had a higher death count because of this disease. Cancer was also the second leading cause of death among all four ethnicities. Cancer has been the leading cause of death since 2007, similar to Heart Disease, Cancer death counts have been dropping since 2007. To avoid the primary causes of mortality, people of New York City would need to adopt healthier lifestyles by eating healthier foods, and the government must impose a higher charge on tobacco-related items to discourage young people from smoking. The air-quality index must increase through maintaining strong carbon emission rules, limiting the manufacturing of non-renewable energy-based motor engines, and increasing tree planting.

8.0 Appendix

*Note: This section contains the queries that were ran int mySQL workbench. SQL queries were run on the CSV file that still contained null/missing values.

Find the total number of deaths and their cause for each ethnicity in New York City:

Select `Race Ethnicity`,`Leading Cause`,Deaths
FROM nyc.lcod
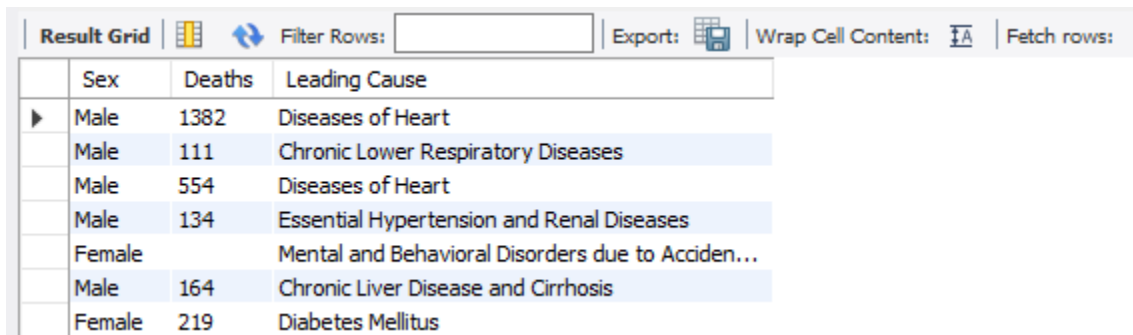GROUP BY `Race Ethnicity`,`Leading Cause`,Deaths;

Race Ethnicity	Leading Cause	Deaths
Hispanic	Diseases of Heart	1382
Hispanic	Chronic Lower Respiratory Diseases	111
Asian and Pacific Islander	Diseases of Heart	554
Non-Hispanic Black	Essential Hypertension and Renal Diseases	134
Other Race/Ethnicity	Mental and Behavioral Disorders due to Acciden...	
Hispanic	Chronic Liver Disease and Cirrhosis	164
Hispanic	Diabetes Mellitus	219

Find the Total Death count and the Leading Cause of Death:

Select Sex,Deaths,`Leading Cause`
FROM nyc.lcod;
GROUP BY Sex,Deaths

Sex	Deaths	Leading Cause
Male	1382	Diseases of Heart
Male	111	Chronic Lower Respiratory Diseases
Male	554	Diseases of Heart
Male	134	Essential Hypertension and Renal Diseases
Female		Mental and Behavioral Disorders due to Acciden...
Male	164	Chronic Liver Disease and Cirrhosis
Female	219	Diabetes Mellitus

What is the leading cause of death for Hispanic individuals, where the total number of deaths are more than 1000:

Select `Race Ethnicity`,Deaths,`Leading Cause`
FROM nyc.lcod
GROUP BY `Race Ethnicity`,Deaths
Having Deaths > 1000 And `Race Ethnicity` = 'Hispanic'

Leading Cause of Death and death count among Asian and Pacific-Islanders in 2009, where deaths are at least 100:

Select `Race Ethnicity`,Deaths,`Leading Cause`,Year
FROM nyc.lcod
WHERE Year = '2009'
GROUP BY `Race Ethnicity`,Deaths
Having Deaths > 100 And `Race Ethnicity` = 'Asian and Pacific Islander'

Find the top 5 Mortality Rate causes in NYC for Non-Hispanic White Individuals:

Select `Race Ethnicity`,`Death Rate`,`Leading Cause`
FROM nyc.lcod
GROUP BY `Race Ethnicity`,`Death Rate`
Having `Race Ethnicity` = 'Non-Hispanic White'
ORDER BY `Leading Cause`
ASC LIMIT 5;

Find the number of Deaths and their leading cause for Asian and Pacific-Islanders in New York City:

Select `Race Ethnicity`,Deaths,`Leading Cause`
FROM nyc.lcod
WHERE `Race Ethnicity` = 'Asian and Pacific Islander';

Race Ethnicity	Deaths	Leading Cause
Asian and Pacific Islander	554	Diseases of Heart
Asian and Pacific Islander	76	Diabetes Mellitus
Asian and Pacific Islander	63	Accidents Except Drug Posioning
Asian and Pacific Islander	452	Diseases of Heart
Asian and Pacific Islander	29	Essential Hypertension and Renal Diseases
Asian and Pacific Islander	39	Essential Hypertension and Renal Diseases
Asian and Pacific Islander	19	Certain Conditions originating in the Perinatal P...

References

Ahmad, F. B., & Anderson, R. N. (2021, May 11). *The Leading Causes of Death in the US for 2020*. JAMA. Retrieved November 20, 2021, from https://jamanetwork.com/journals/jama/fullarticle/2778234.

Australian Institute of Health and Welfare. (2018). *3.2 leading causes of death, chapter 3 causes of ill health (Australia's Health 2018).* Australian Government Australian Institute of Health and Welfare. Retrieved November 20, 2021, from https://www.aihw.gov.au/getmedia/96909f12-5e9d-4531-a32a-34e414c6b862/aihw-aus-221-chapter-3-2.pdf.aspx.

Department of Health and Mental Hygiene (DOHMH). (2020, December 16). New York City Leading Causes of Death. Retrieved from https://data.cityofnewyork.us/Health/New-York-City-Leading-Causes-of-Death/jb7j-dtam

Holland, K. (2019, March 29). *12 leading causes of death in the United States*. Healthline. Retrieved November 28, 2021, from https://www.healthline.com/health/leading-causes-of-death.

Sharp, M. J., Schoen, L. D., Wang, T., & Melnik, T. A. (n.d.). *Leading Causes of Death, New York State, 2012*. New York State Department of Health. Retrieved November 20, 2021, from https://www.health.ny.gov/statistics/vital_statistics/docs/leading_causes_of_death_nys_2012.pdf.