ANALYZING THE LEADING CAUSE OF DEATH IN NYC FROM 2007- 2014

Abbas Khanzada

Abstract:

In this analysis, the NYC data is used from 2007-2014 and the leading cause of deaths is analyzed using both exploratory data analysis and regression analysis. Certain insights are generated which are useful for the health department to look into and make strategies accordingly. The results showed that stroke kills the most people in New York City, followed by diabetes and heart disease. Most deaths are heart-related. The leading cause of mortality from 2009 to 2014 was stroke, except in 2007 and 2008, when accidents (excluding drug poisoning) led. Accidental deaths are highest among white non-Hispanics. White non-Hispanic men had greater death rates than women. Death rates have declined, peaking in 2007 and falling in 2014.

Introduction:

As a matter of fact, health is a highly significant aspect of people's life, regardless of geographical location. It is possible for the causes to be quite diverse, and New York City is not an exception when it comes to people who are attempting to lead healthy lives. Due to the frenetic climate, living in New York City is a tremendously stressful experience [1]. In every part of the city and in every location that we visit, we hear a great number of ambulances. There is now a problem with noise pollution because it has reached a certain extent. The leading causes of death in New York City have changed over time; however, these causes may vary dramatically depending on factors such as gender, race, income level, or geographic location. We are better able to examine the living situations of citizens as well as the social and economic issues that are behind them when we have a better understanding of the primary causes of mortality. Furthermore, when it comes to future illness control and medical research, the data would be extremely important in order to enhance the health of the community. This analysis provides responses to a number of questions, including the following:

- What is the most common reason for death in New York City?
- What is the mortality rate between the sexes?
- In what direction is the death rate for each gender moving?
- In terms of gender, year, race, and ethnicity, what is the leading cause of death?

- If there is such a wide range of people living in the city, how would the death rate vary according to race?
- Develop a regression model to forecast the death rate by taking into account gender, race and ethnicity, and the year.

Background:

During the past few years, the New York City Department of Health and Mental Hygiene has increased its efforts to investigate and improve COD reporting. These efforts include the implementation of a hospital-based intervention to reduce the over-reporting of heart disease, the development of an e-learning COD training module, and the ongoing provision of on-site training for health-care providers who are responsible for completing death certificates [2]. However, quality difficulties continue to exist, which restrict the value of COD data. These challenges are caused by a number of variables, including administration practices and policies, concerns regarding risk management, and a lack of uniform training for physicians. Studies that have evaluated the accuracy of UCOD have discovered that heart disease is over reported, that deaths caused by pneumonia have underlying causes that are not stated on the death certificate, and that UCOD statistics may differ when vital records jurisdictions request more information for death certificates that are submitted with limited COD information [3].

Data & Methods:

The dataset that is used in this study is obtained from NYC website and is used for analysis. The data initially consists of 7 attributes and 1094 observations in total which provides us information. The exploratory data analysis as well as regression analysis techniques are applied. The tables are created to answer the research questions along with bar graphs and time series charts. Finally, the regression analysis is applied to predict the death rate based on sex, race & ethnicity and year.

Results & Discussion:

Starting with the leading cause of deaths overall in NYC, the table attached below shows the top 10 leading causes of deaths. The leading cause of death that tops the list in this case is cerebrovascular disease which is commonly referred to as Stroke. The leading cause of death that

stands at the second number in the list is Diabetes Mellitus followed by disease of heart. It is clear from the table attached that most of the deaths are caused by heart related diseases and issues.

	Cause_of_Death <fctr>	Count <int>
2	Cerebrovascular Disease (Stroke: I60-I69)	64
3	Diabetes Mellitus (E10-E14)	64
4	Diseases of Heart (I00-I09, I11, I13, I20-I51)	64
5	Influenza (Flu) and Pneumonia (J09-J18)	64
6	Malignant Neoplasms (Cancer: C00-C97)	64
7	Chronic Lower Respiratory Diseases (J40-J47)	63
8	Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)	56
9	Essential Hypertension and Renal Diseases (I10, I12)	55
10	Human Immunodeficiency Virus Disease (HIV: B20-B24)	29
11	Mental and Behavioral Disorders due to Accidental Poisoning and Other Psychoactive Substance Use (F11-F16, F18-F19, X40-X42, X44)	25

Similarly, when the leading cause of death is explored for each year it is found out that except 2007 and 2008, all other years from 2009-2014 have cerebrovascular disease as the leading cause of death. However, in 2007 and 2008, the leading cause of death was accidents except drug poisoning.

Year <int>	Leading.Cause <chr>
2007	Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)
2008	Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)
2009	Cerebrovascular Disease (Stroke: I60-I69)
2010	Cerebrovascular Disease (Stroke: I60-I69)
2011	Cerebrovascular Disease (Stroke: I60-I69)
2012	Cerebrovascular Disease (Stroke: I60-I69)
2013	Cerebrovascular Disease (Stroke: I60-I69)
2014	Cerebrovascular Disease (Stroke: I60-I69)

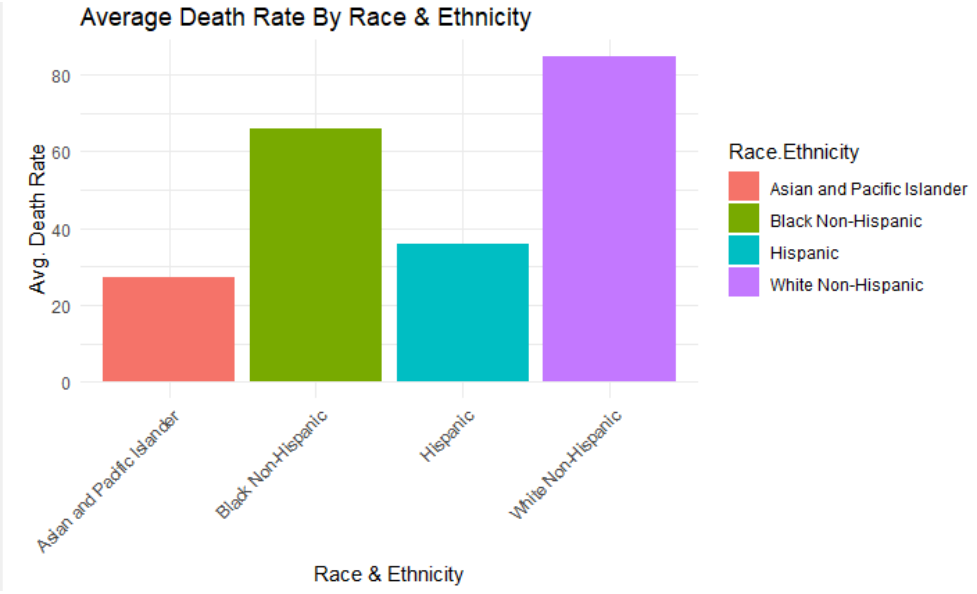
When the leading cause of death is explored by gender, both genders showed the same result and no difference was observed.

Sex <chr>	Leading.Cause <chr>
F	Cerebrovascular Disease (Stroke: I60-I69)
M	Cerebrovascular Disease (Stroke: I60-I69)

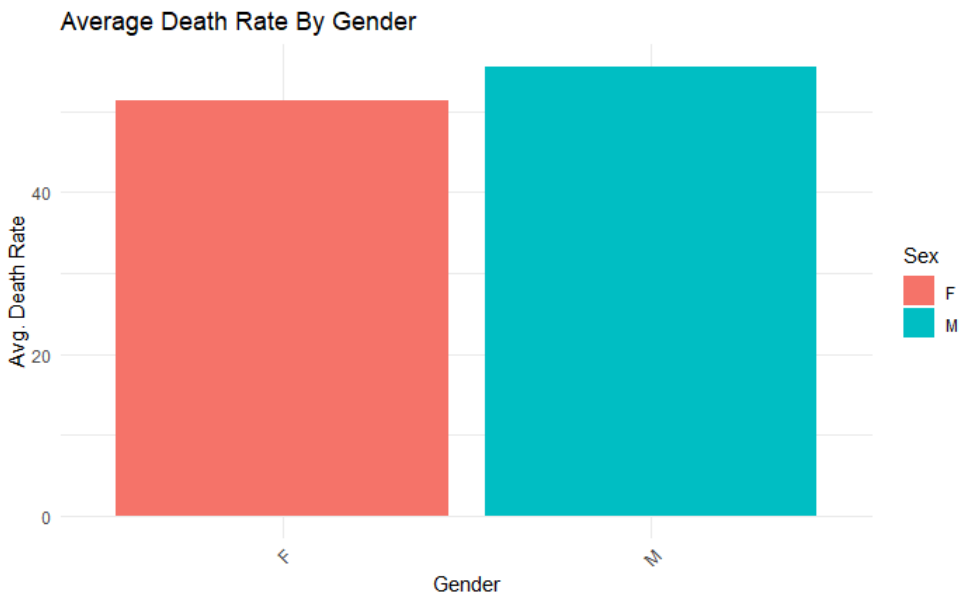
Based on the race and ethnicity, the Asian & Pacific Islander have accidents except drug poisoning as leading cause of death. The black non-Hispanic and Hispanic have cerebrovascular disease as the leading cause of death while for white non-Hispanic again it's accidents except drug poisoning.

Race.Ethnicity <chr>	Leading.Cause <chr>
Asian and Pacific Islander	Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)
Black Non-Hispanic	Cerebrovascular Disease (Stroke: I60-I69)
Hispanic	Cerebrovascular Disease (Stroke: I60-I69)
White Non-Hispanic	Accidents Except Drug Posioning (V01-X39, X43, X45-X59, Y85-Y86)

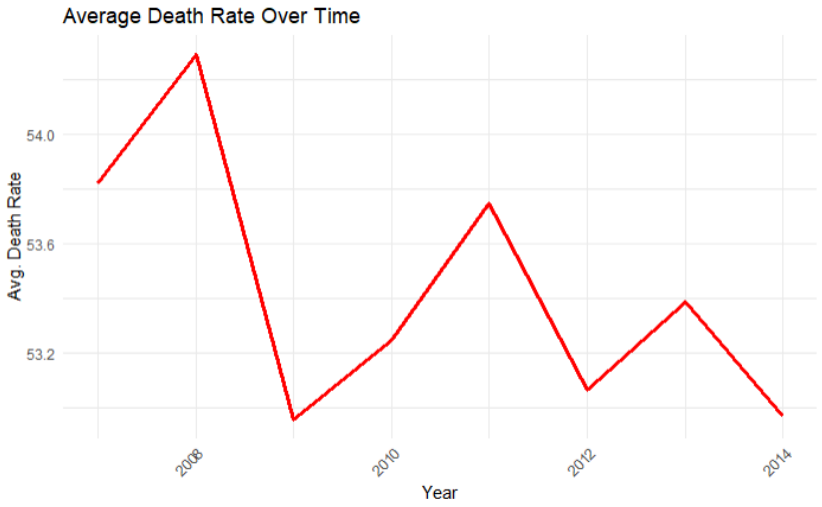
The average death rate by race and ethnicity, is plotted next and it can be seen that the highest death rate is observed for white non-Hispanic followed by black-non-Hispanic and then Hispanic. The least death rate is observed for Asian and Pacific Islander.



The average death rate for male is higher compared to females with average death rate for females being lower with values around 50 and higher for male around 55.



Finally, the average death rate over time shows a downward trend overall which means that as time passes by, the average death rate decreases. The highest death rate was reported in 2007 while the lowest is reported in 2014.



The regression model is fitted to the data and the summary of the regression model is attached below. The Sex and Year column is statistically insignificant which means that it has no effect on the death rate statistically. This is because the p value of the features is above significance level alpha = 0.05. The Race and Ethnicity feature has some categories being statistically significant with p values below alpha = 0.05. The R square of the model is 9.151%

which is low showing that the model is unable to explain a good amount of variance. However, the overall model is statistically significant with p value below alpha = 0.05.

```
Call:
lm(formula = Death.Rate ~ Sex + Year + Race.Ethnicity, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-78.00 -44.47 -22.28  34.08 408.30

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     258.3538  2410.0692   0.107    0.915
SexM                              4.0993     5.5020   0.745    0.456
Year                             -0.1159     1.1987  -0.097    0.923
Race.EthnicityBlack Non-Hispanic  38.6417     7.7701   4.973 8.29e-07 ***
Race.EthnicityHispanic             8.5690     7.7810   1.101    0.271
Race.EthnicityWhite Non-Hispanic  57.4423     7.7920   7.372 4.75e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.2 on 702 degrees of freedom
Multiple R-squared:  0.09151,   Adjusted R-squared:  0.08503
F-statistic: 14.14 on 5 and 702 DF,  p-value: 3.375e-13
```

Conclusion:

The leading cause of death in New York City is cerebrovascular illness (stroke), followed by diabetes and heart disease. Heart-related disorders account for the majority of deaths. Stroke was the leading cause of death from 2009 to 2014, with the exception of 2007 and 2008, which were dominated by accidents (excluding drug poisoning). Both genders have the same top causes of death. Accidents are the leading cause of death among Asians and Pacific Islanders, while stroke kills Black non-Hispanics and Hispanics. White non-Hispanics account for the majority of accidental deaths. The highest death rates are among White non-Hispanics, with men having higher rates than women. The average death rate has dropped over time, peaking in 2007 and falling to its lowest in 2014. Regression analysis reveals that race and ethnicity have a considerable impact on death rates, but the model explains little overall variance.

References:

1. Benn, E.K., Allen Hauser, W., Shih, T., Leary, L., Bagiella, E., Dayan, P., Green, R., Andrews, H., Thurman, D.J. and Hesdorffer, D.C., 2009. Underlying cause of death in incident unprovoked seizures in the urban community of Northern Manhattan, New York City. Epilepsia, 50(10), pp.2296-2300.

2. Johns, L.E., Madsen, A.M., Maduro, G., Zimmerman, R., Konty, K. and Begier, E., 2013. A case study of the impact of inaccurate cause-of-death reporting on health disparity tracking: New York City premature cardiovascular mortality. American journal of public health, 103(4), pp.733-739.

3. Gill, J.R., Ely, S.F., Toriello, A. and Hirsch, C.S., 2014. Adverse medical complications: an under-reported contributory cause of death in New York City. Public Health, 128(4), pp.325-331.