

DEATHS IN THE UNITED STATES

GROUP 2

Sibeso Mubonda, Pranita Premnath Shetty, Man Kuei Chen,
Zhipeng Zhao, Pawan Kumar

Summary of Project

Description

- ❑ We looked at the dataset on Mortality Multiple Cause of Deaths in the United States from years 2013-2015
- ❑ The dataset was obtained from Kaggle and the csv files can be found [here](#). The dataset contains data from the Centers for Disease Control and Prevention (CDC).
- ❑ Size: 1.2GB
- ❑ Dataset: 77 columns and a total of 7,590,821 records.
- ❑ Every record represents a deceased individual from ranging from 2013-2015.

Why this dataset?

- ❑ We found it interesting to analyze the patterns of mortality causes in the United States over the years.
- ❑ We found the dataset to be interesting because we can look to find if there is any correlation between certain variables like marital status, race and age and the causes of death.
- ❑ It has data that is annually updated by the CDC which is a very reliable public health agency.
- ❑ Dataset has clear variables and a data dictionary that explains the data. It also has a lot of data rows which makes the analysis more efficient as the sample is large.

Business Questions

- ❑ What is the distribution of the number of deaths across sex and race? In totality, what is the impact heart disease has over the three years on various age groups? Does an individual living alone decrease their probability of getting immediate help when one is suffering from a heart disease?
- ❑ Is there any difference in the top 3 death causes for single people versus married people at certain ages? If so, what is that death cause? What preventive measures could we do to increase those individual's lifespan?
- ❑ What manner of death and age groups have the highest count of autopsies performed after death? What is the relationship between manner of death and autopsy in relation to age?

Data Modeling

❑ Merging

- ❑ Generated a ZIP file link for loading data on AWS from Amazon S3
- ❑ Merged three Target CSV files (2013,2014,2015) by using “csv stack”

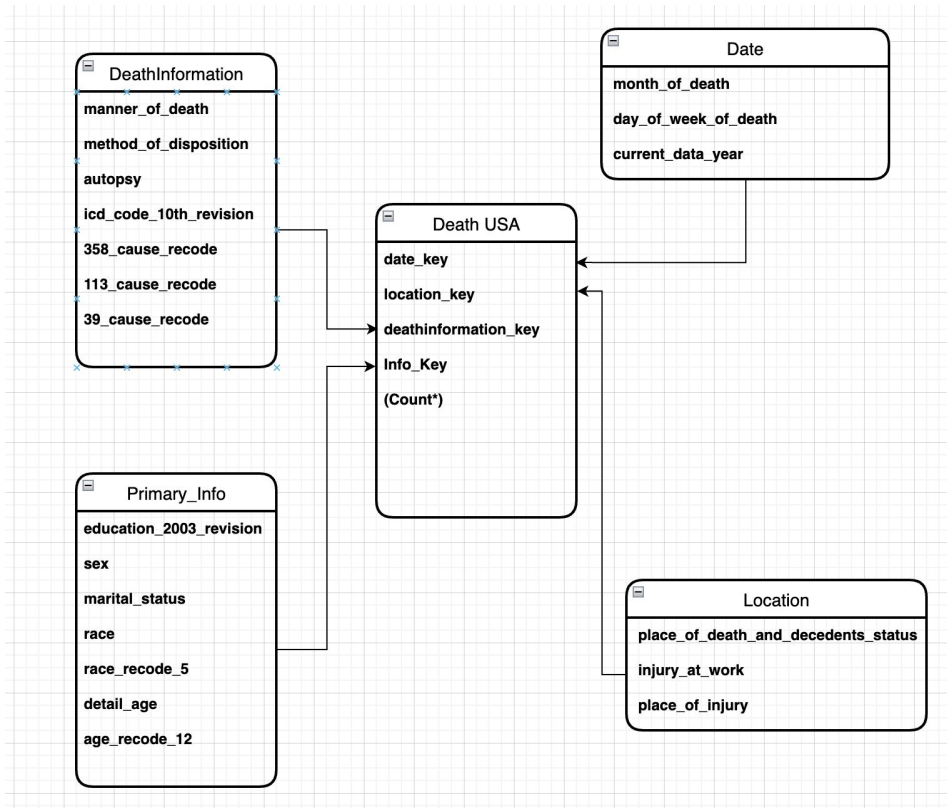
❑ Cleaning

- ❑ Dropped the column without analytical significance
- ❑ Replaced the missing value with suitable numeric value per data-dictionary

❑ Modeling steps

- ❑ Designed the Star Schema and made the Fact table
- ❑ Final dataset contains 7,950,821 rows

Star Schema



Business Question 1:

- ❏ What is the distribution of the number of deaths across sex and race? In totality, what is the impact heart disease has over the three years on various age groups? Does an individual living alone decrease their probability of getting immediate help when one is suffering from a heart disease?

Sex and Causes of death:

- ❑ Overall: No. of deaths observed in
 - **Males:** 4,017,696
 - **Female:** 3,933,125
- ❑ Top 5 causes of death in **Males:** *Ischemic heart diseases, other heart diseases, other malignant neoplasms, malignant neoplasms of trachea, bronchus and lung and chronic lower respiratory diseases.*
- ❑ Top 5 causes of death in **Females:** *Ischemic heart diseases, other heart diseases, chronic lower respiratory diseases, cerebrovascular diseases and malignant neoplasms of trachea, bronchus and lung.*

Race and Causes of death:

- ❑ Overall: No. of deaths observed in White, Black, American Indian and Asian/Pacific Islander
- ❑ Top 5 causes of death amongst **White population:** *Ischemic heart diseases, other diseases of heart, chronic lower respiratory diseases, other malignant neoplasms and malignant neoplasms of trachea, bronchus and lung.*
- ❑ Top 5 causes of death in **Black population:** *Ischemic heart diseases, other diseases of heart, other malignant neoplasms, cerebrovascular diseases and malignant neoplasms of trachea, bronchus and lung.*

Reference Table-Age

1 - Under 1 year (includes not stated infant ages)

2 - 1-4years

3 - 5-14years

4 - 15 - 24 years

5 - 25 - 34 years

6 - 35 - 44 years

7 - 45 - 54 years

8 - 55 - 64 years

9 - 65 - 74 years

10 - 75 - 84 years

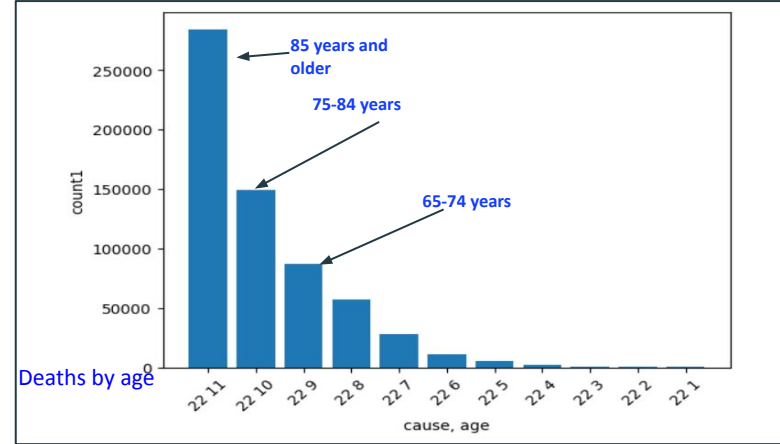
11 - 85 years and over

12 - Age not stated

Age and cause of death analysis:

In 2015 the following has been observed:

- a. Age groups under 1 year, 1-4 years, 5-14 years and 15-24 years represent 2.34% of the total deaths.
- b. Age groups 25-34 years, 35-44 years and 45-54 years represent 11.05% of total deaths.
- c. 86.61% of the deaths are represented by 55-64 years, 65-74 years, 75-84 years and above 85 years age group.



Relationship between place of death and marital status when cause of death is a heart disease:

- ❖ Place of death considered for the analysis is *decedent's home* and marital status is *married*.
- ❖ Our analysis is based on the assumption that single, widowed and divorced individuals live alone whereas married individuals live with their significant other.
- ❖ Those individuals who live alone have the highest percentage of deaths at their own homes in comparison to married individuals.
- ❖ Only 34% married individuals pass away at home due to heart diseases in comparison to the 64% individuals who live alone.

Specific analysis on cause of death - heart disease across age groups:

- The analysis is performed collectively over the three years 2013 - 2015.
- Heart disease occurs highest after the age of 55 years. This represents 92.90% of total deaths by heart disease.
- Highest heart disease deaths are observed in age group 85+ years.

Business Question 2:

- ❑ Is there any difference in the top 3 death causes for single people v.s. married people at certain age? If so, what is that death cause? What are preventive measures we could do to increase those individuals' lifespan?

Among singles, there are more people that died at age **55-64** than at the other age groups while among married people, there are more people that died at **75-84** compared to the other age groups.

1 - Under 1 year (includes not stated infant ages)

2 - 1-4years

3 - 5-14years

4 - 15 - 24 years

5 - 25 - 34 years

6 - 35 - 44 years

7 - 45 - 54 years

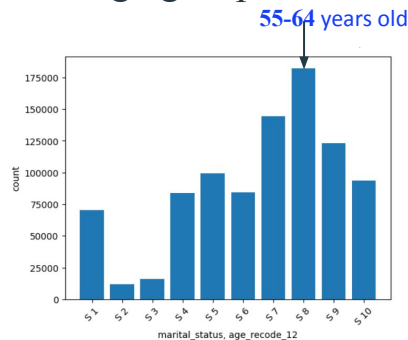
8 - 55 - 64 years

9 - 65 - 74 years

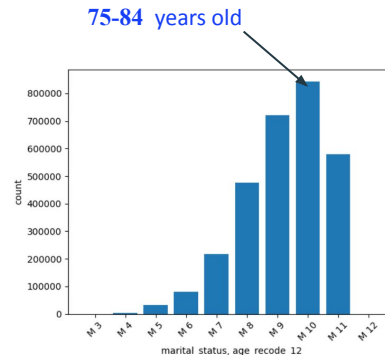
10 - 75 - 84 years

11 - 85 years and over

12 - Age not stated

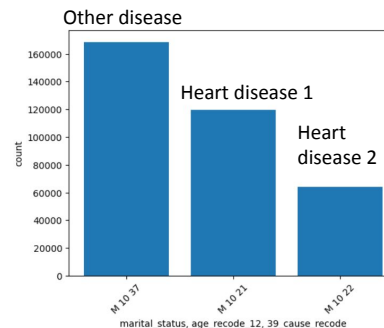
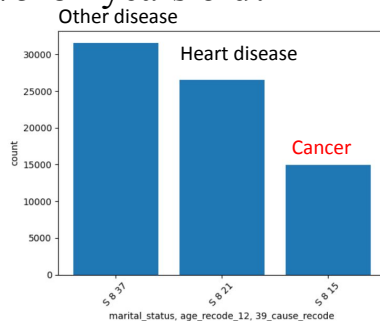


Deaths by age among single individuals

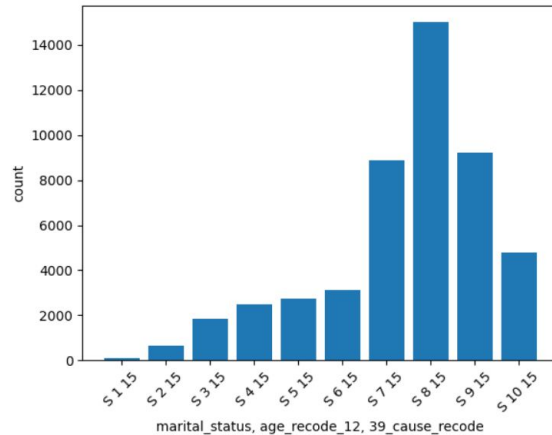


Deaths by age among married individuals

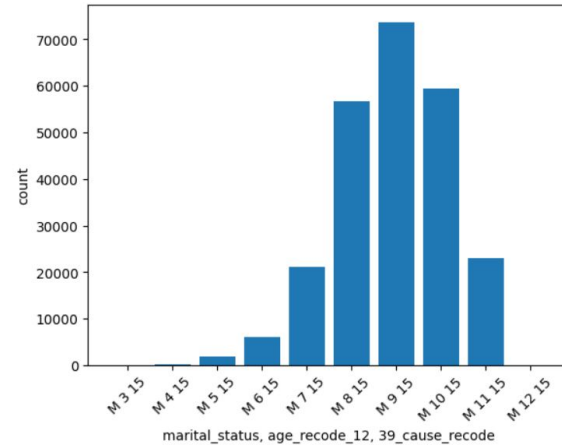
What are the top 3 death causes for single people at 55-64 years old v.s. Top 3 death causes for married people at 75-84 years old?



Among singles, there are more people that died from cancer at age **45-74** than at the other age groups while among married people, there are more people that died from cancer at **55-84** compared to the other age groups. The overall distributions show that single people die from cancer at relatively younger ages



Deaths due to cancer by age among single individuals



Deaths due to cancer by age among married individuals

The result highlights the potential impact marital status and likely the support shared by the married people can have on cancer mortality. For example, lacking support from spouses can lead to lower cancer screening rates and poorer lifestyle choices. We suggest companies , such as pharmaceutical and biomedical, designing cancer screening and prevention products targeting the younger single population.

Business Question 3:

- ❑ What manner of death and age groups have the highest count of autopsies performed after death? What is the relationship between manner of death and autopsy in relation to age?

Reference Table-Death Code

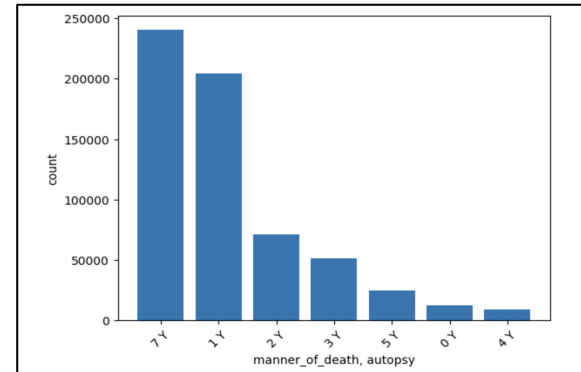
The data dictionary for manner of death is as follows:

- 1 - Accident
- 2 - Suicide
- 3 - Homicide
- 4 - Pending Investigation
- 5 - Could not determine
- 6 - Self-Inflicted
- 7 - Natural
- 0 - Not Specified

The most common manner of code for death's manner is 7. Autopsy is mainly performed when the cause of death is unknown. Counting the manner of death based on autopsy may bring various insights in relation to other aspects.

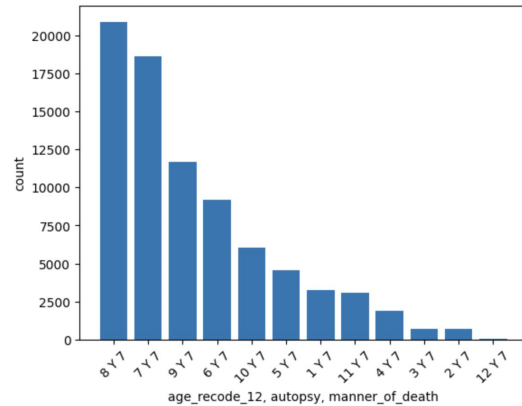
manner_of_death	autopsy	count
7	Y	240167
1	Y	204492
2	Y	71024
3	Y	51152
5	Y	24743
0	Y	12595
4	Y	9120

Manner of death

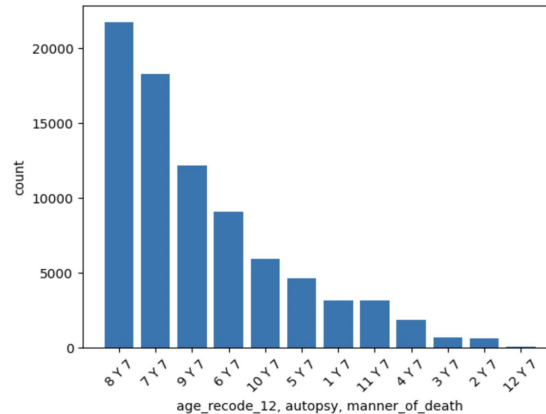


Count of Death

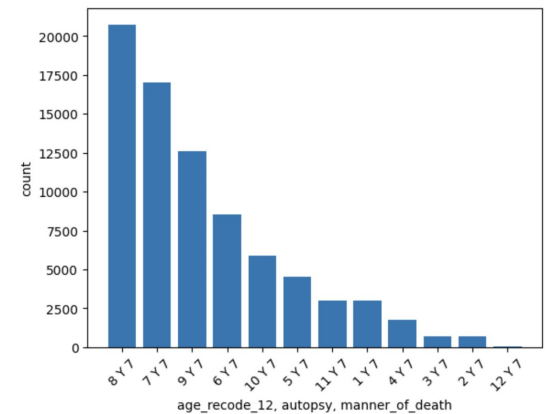
Age in relation to autopsy performed and natural cause of death



2013



2014



2015

The age group with the highest count of the deceased who had autopsies performed and died from natural causes in 2013, 2014 and 2015 is group 8 which refers to people around the age of 55-64 years old. An intuitive outcome .