# ¶ Stat 405 Final

Meera Borle, Isis Burgos, Naomi Consiglio, Carson Foster, Aidan Gerber

December 10th, 2022

## Introduction

Our main dataset is a mortality dataset from the National Bureau of Economic Research. There are new versions of it each year and we are using the 2019 version. Each row is one death based on information filed in death certificates in the 50 states plus Washington DC. Columns are factors regarding the status of persons involved in deaths that can be gleaned from public death certificates. The dataset is focused on the different causes of death. Notable information the dataset contains is education, sex, age classification, day of month, place of death, weekday, manner of death, cause of death, and different risk factors that the deceased had.

## Questions

We are interested in what different factors correlated with higher death rates based on cause of death, age groups, and weekdays.
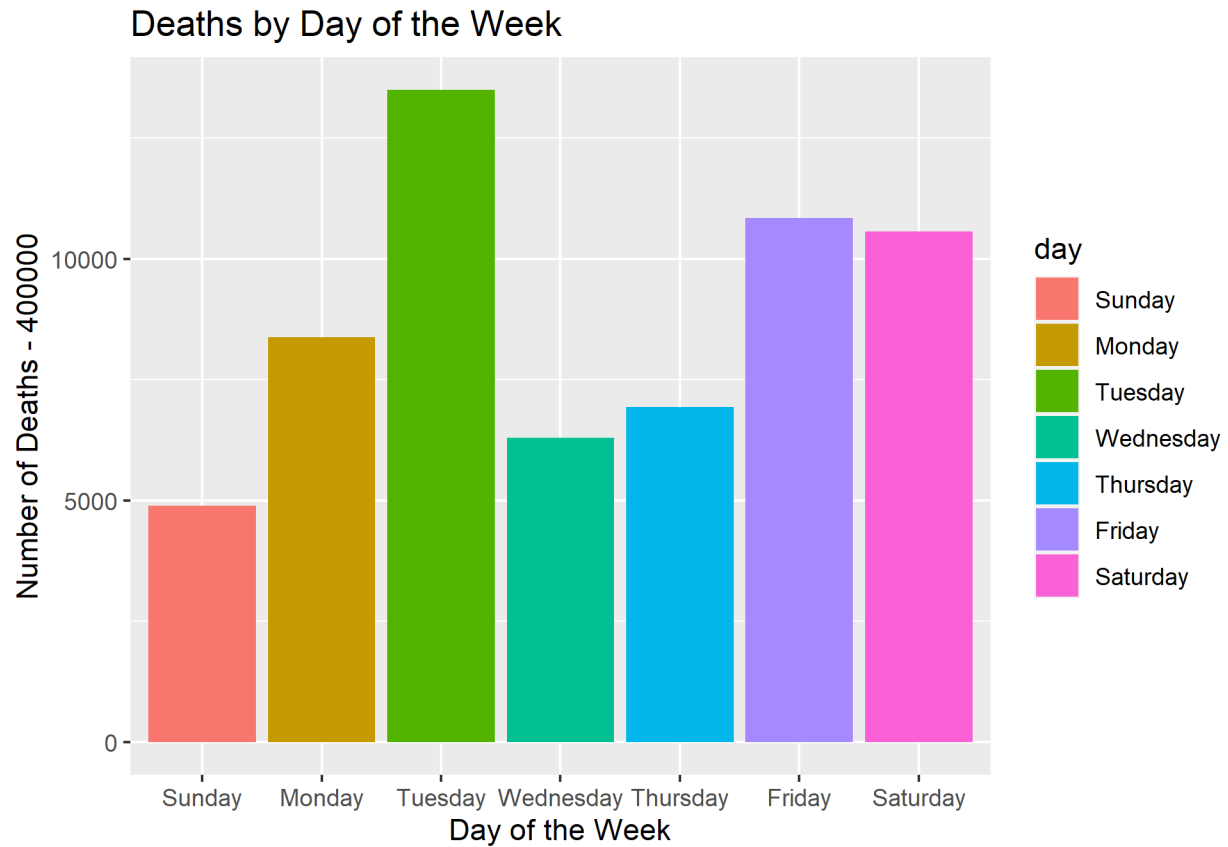
## Exploration

### Deaths by Weekday

First, we plotted weekday of death versus death counts. There were the most deaths on Tuesday. However, days have an average of 7839.79 deaths and 2019 had an extra Tuesday

so adjusting for that, the most deaths were on Fridays.

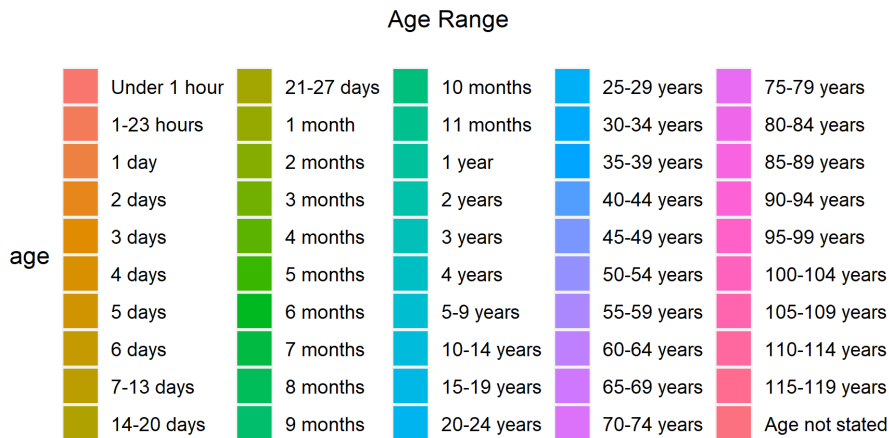## Deaths by Day of the Week
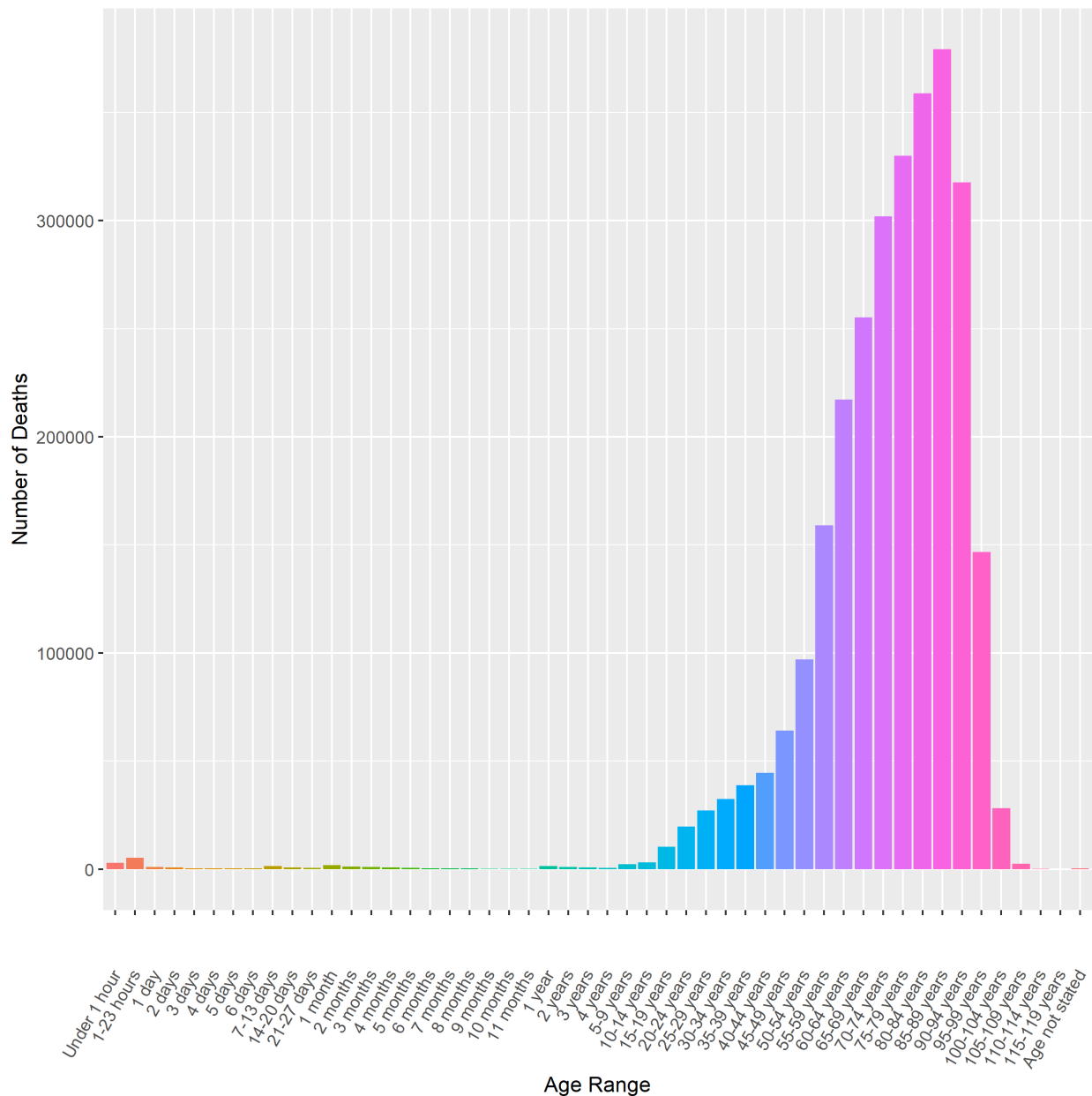
Deaths by Day of the Week
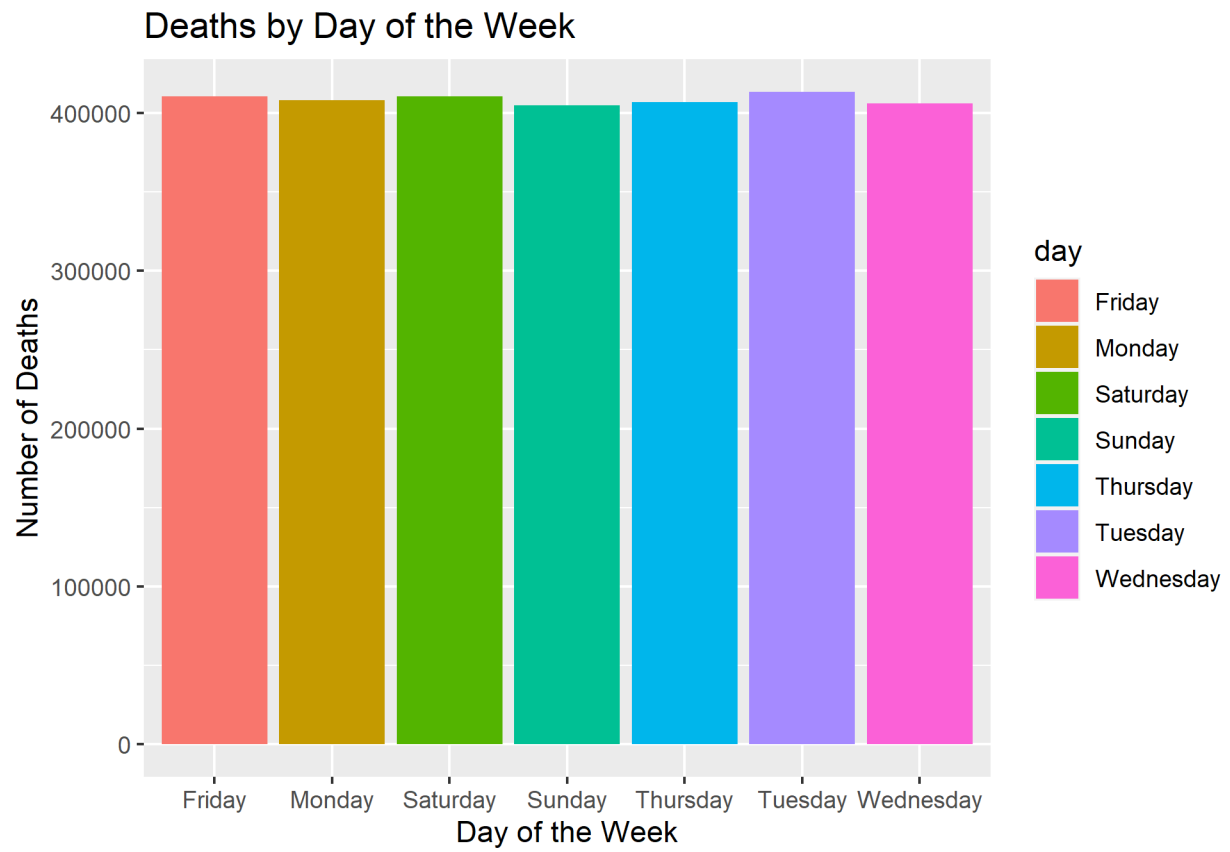
## Deaths by Age

Next, we plotted age versus death counts. Deaths were most prevalent among larger age groups such as those between 70 and 84, although deaths start increasing more quickly at age 60. There is also a spike in those less than 1 day old. However, those greater than 1 day old do not frequently die.

Deaths by Age

## Deaths by Manner

Here, we plotted the manner of death versus age and counted how many people of a certain aged died based on a certain manner of death. A few key finding of this analysis shows that the majority of people die from natural causes, especially those aged 60+ and less than 1 day old, and accidental causes, spanning across all age groups. What this plot may help to inform us about is the behavior and activities that people in a general age group may commonly engage in that may have lead to their manner of passing. By being observant of the manners of death based on age group, preventative methods can be used to decrease the number of accidental related deaths if we are able to determine commonly engaged activities for age groups. Using this plot will help us answer the cause of death amongst the different age groups, and further promote research in what actual activites people are participating in that lead to their manner of death.

Manner of Death by Age on a log scale

## Deaths by Age and Gender

In this plot, we plotted the number of deaths versus age ranges while demonstrating how many men compared to women passed away in each age category. In each of the bars, the red fill represents the amount of women who passed away in that particular age range while the blue accounts for the amount of men. The percentage seen in each bar represents the proportion of men in a given age range that passed away. This analysis shows that the majority of people under the age of 80 who pass away tend to be men, as nearly every bar from ages 0-80 shows the proportion of male deaths to be above 50%. This proportion of male deaths goes down after 80 years of age, and is likely because women who are of an older age tend to live to a complete life expectancy. What this plot may help to inform us about is the differences in male and females lives and life expectancies. Further research into differences in lifestyle choices for men versus women as a whole may help better explain why women tend to live longer than men. Furthermore, this plot accompanied with a plot on cause of death by gender, may assist in determining what kind of, potentially more risky, behaviors men may partake in during their lifetimes that lead to an earlier death than women.

# Deaths by Age on a Log Scale with Percent Male
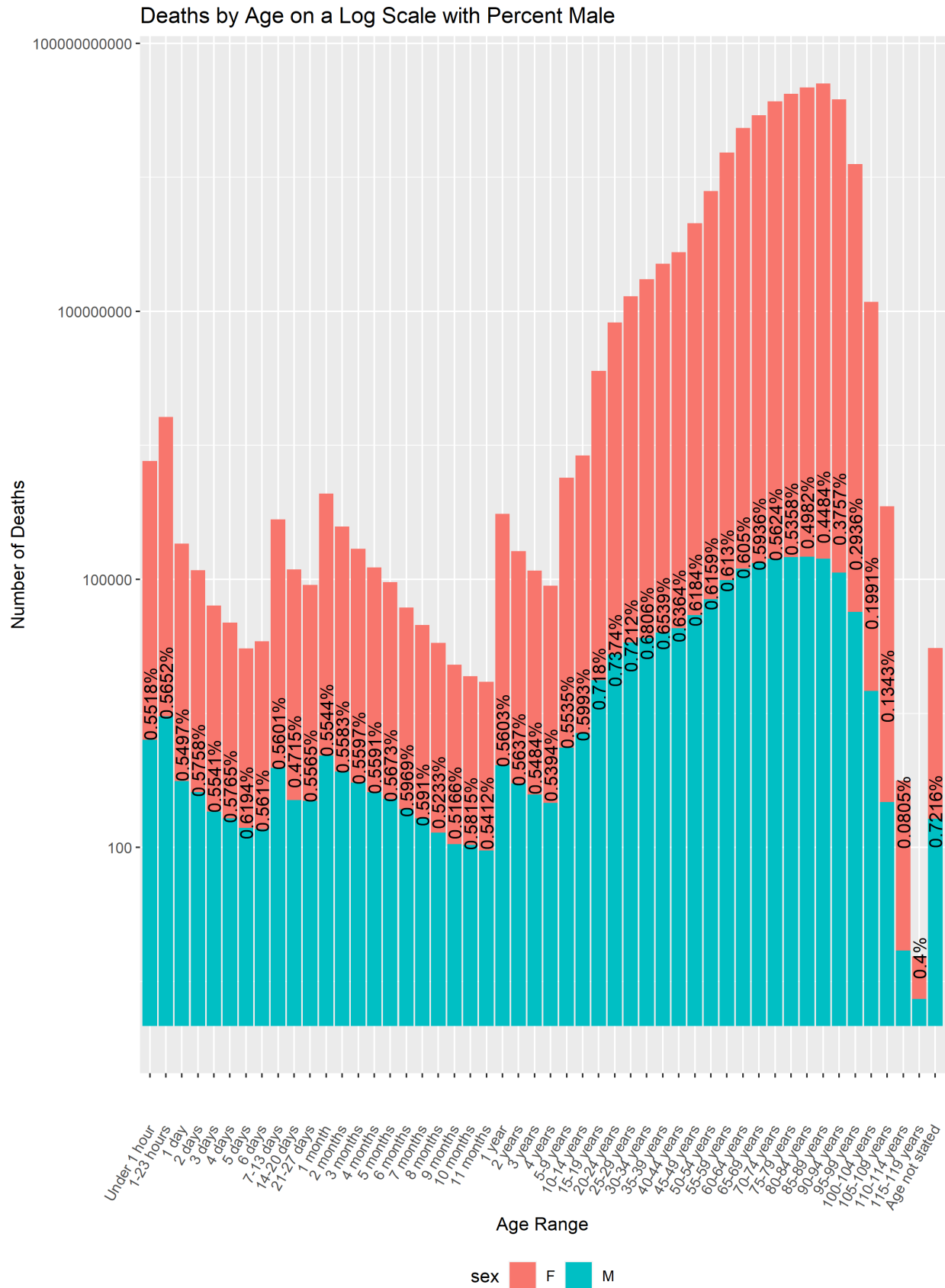
## Cause of Death by Education

For most causes of death, level of educuation does not have an impact on what proportion of people have that cause of death. The largest difference belongs to "Certain conditions originating in the perinatal period" with high occurences in those with 8th grade or less education and those with unknown education and nearly no occurences in all others. Another large proportion difference is in "Congenital malformations" where 8th grade or less has a much higher mortality proportion than other education levels. For causes of death that are not highly tied to conditions at birth, "Syphilis" and "Assault (homicide)" have the highest differing propotions. "Syphilis" has a quite small sample size but unknown education has the highest mortality propotion. For "Assault (homicide)", 9 - 12th grade, no diploma has thhe highest mortality proportion.

Cause of Death by Education Level