

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

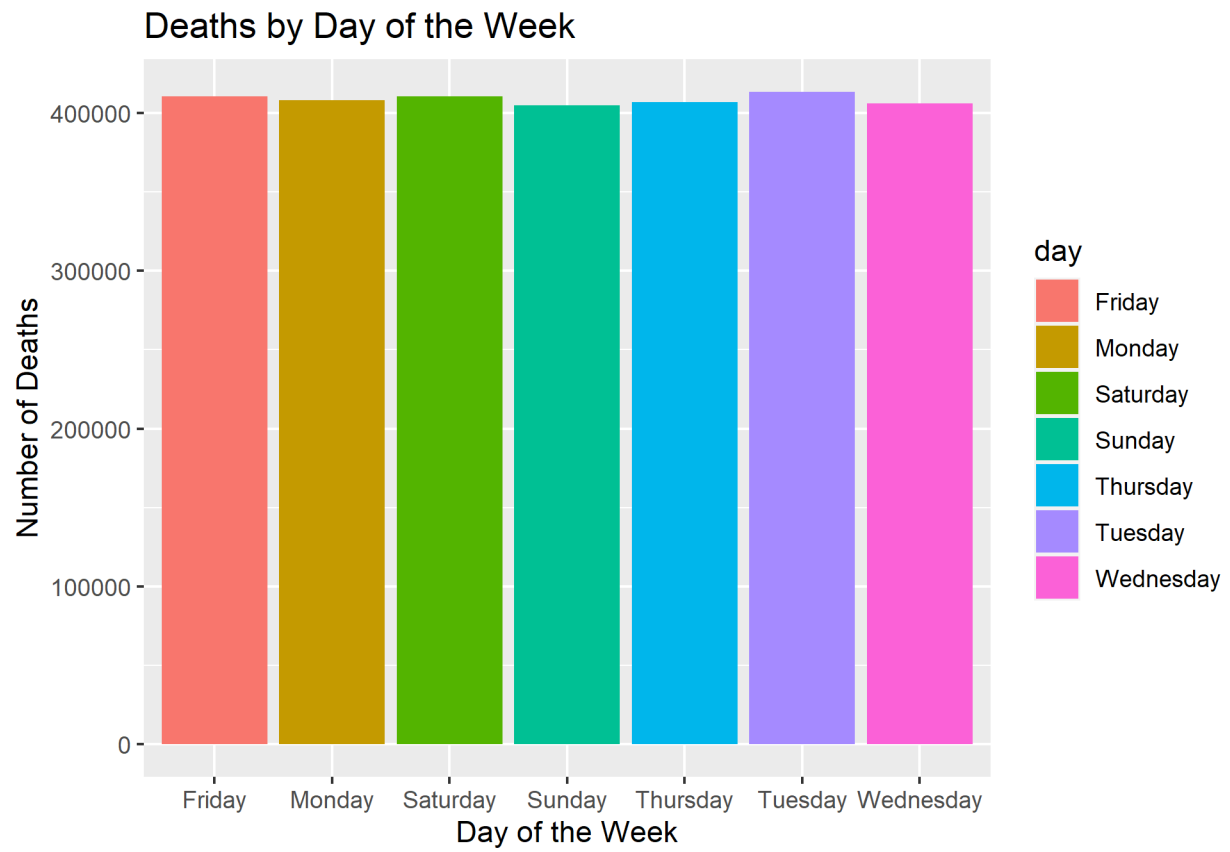
We are interested in what different factors correlated with higher death rates based on cause of death, age groups, and weekdays.

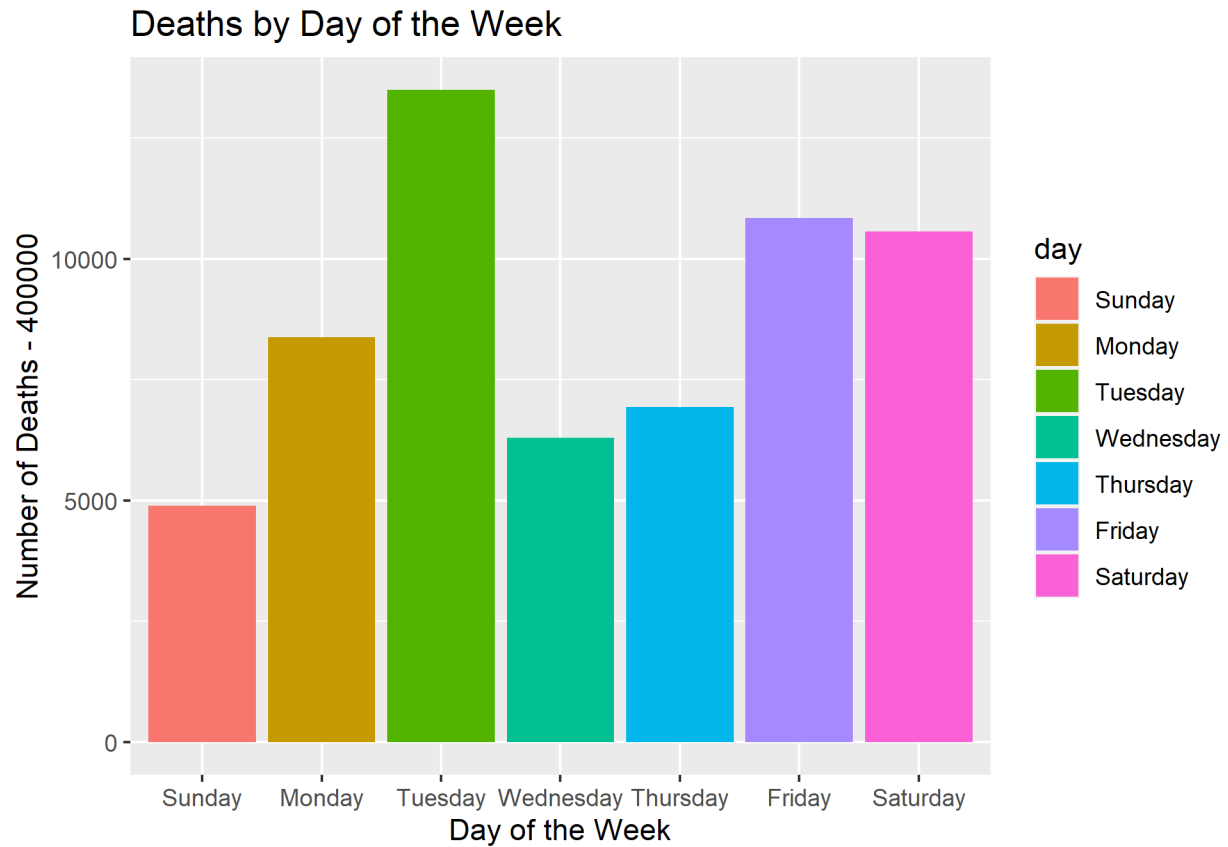
Exploration

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.

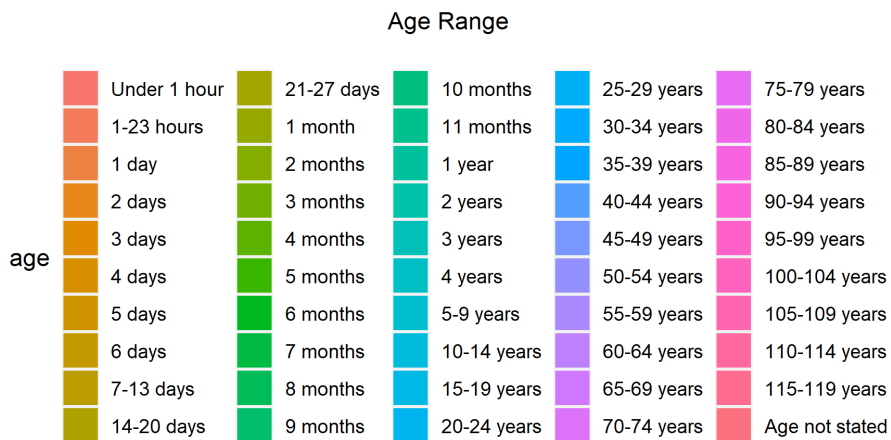
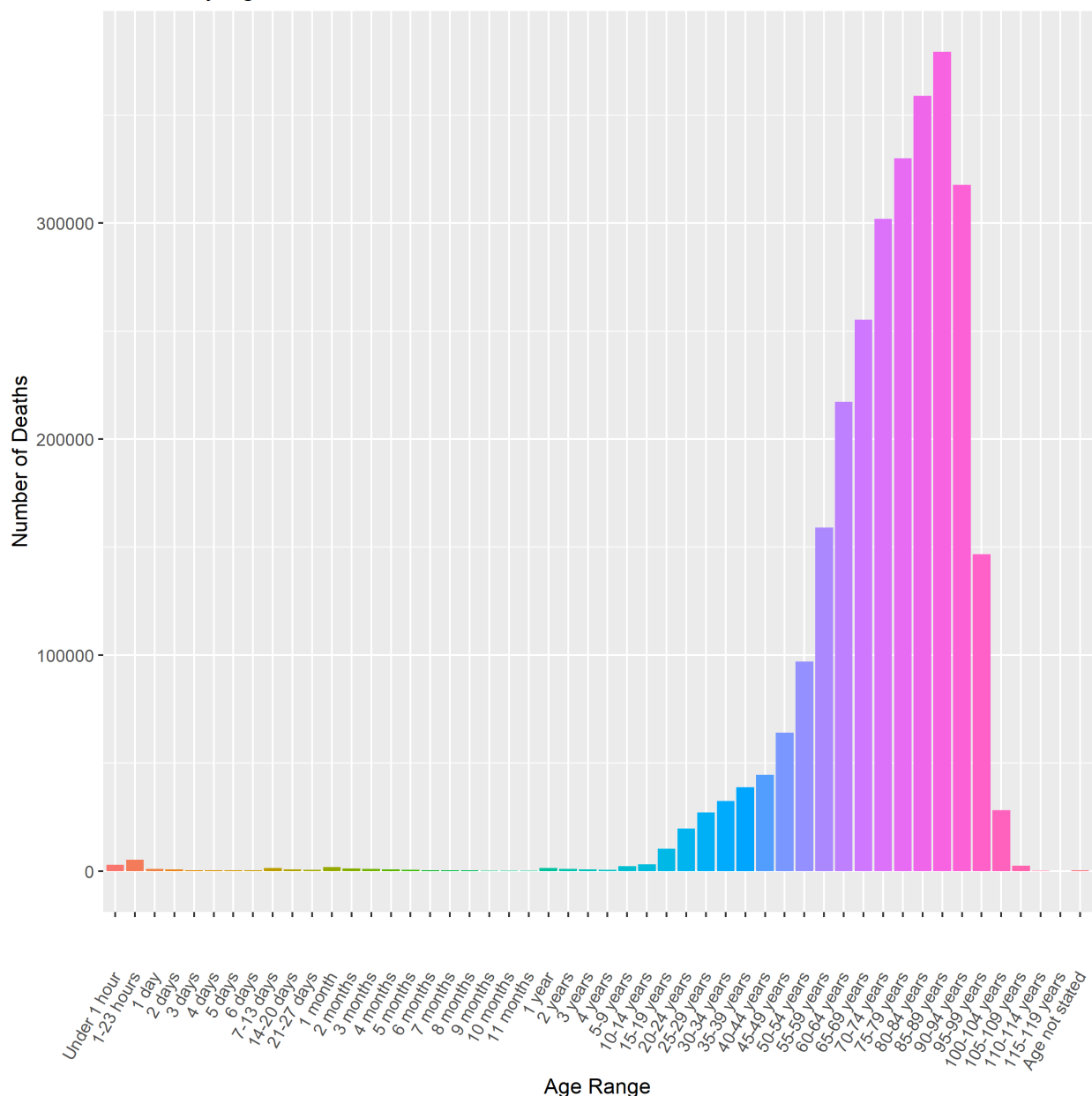




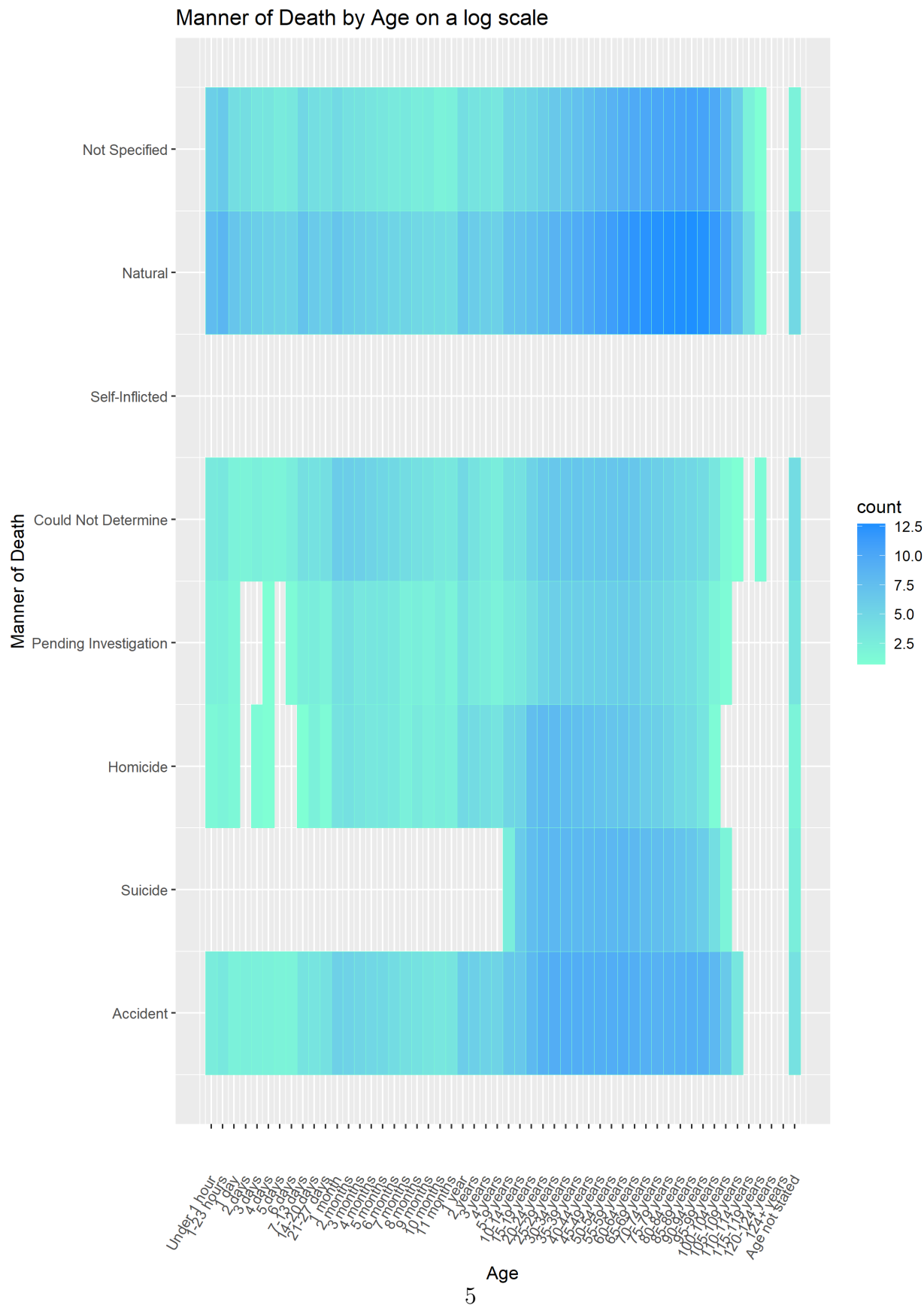
Deaths by Age

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.

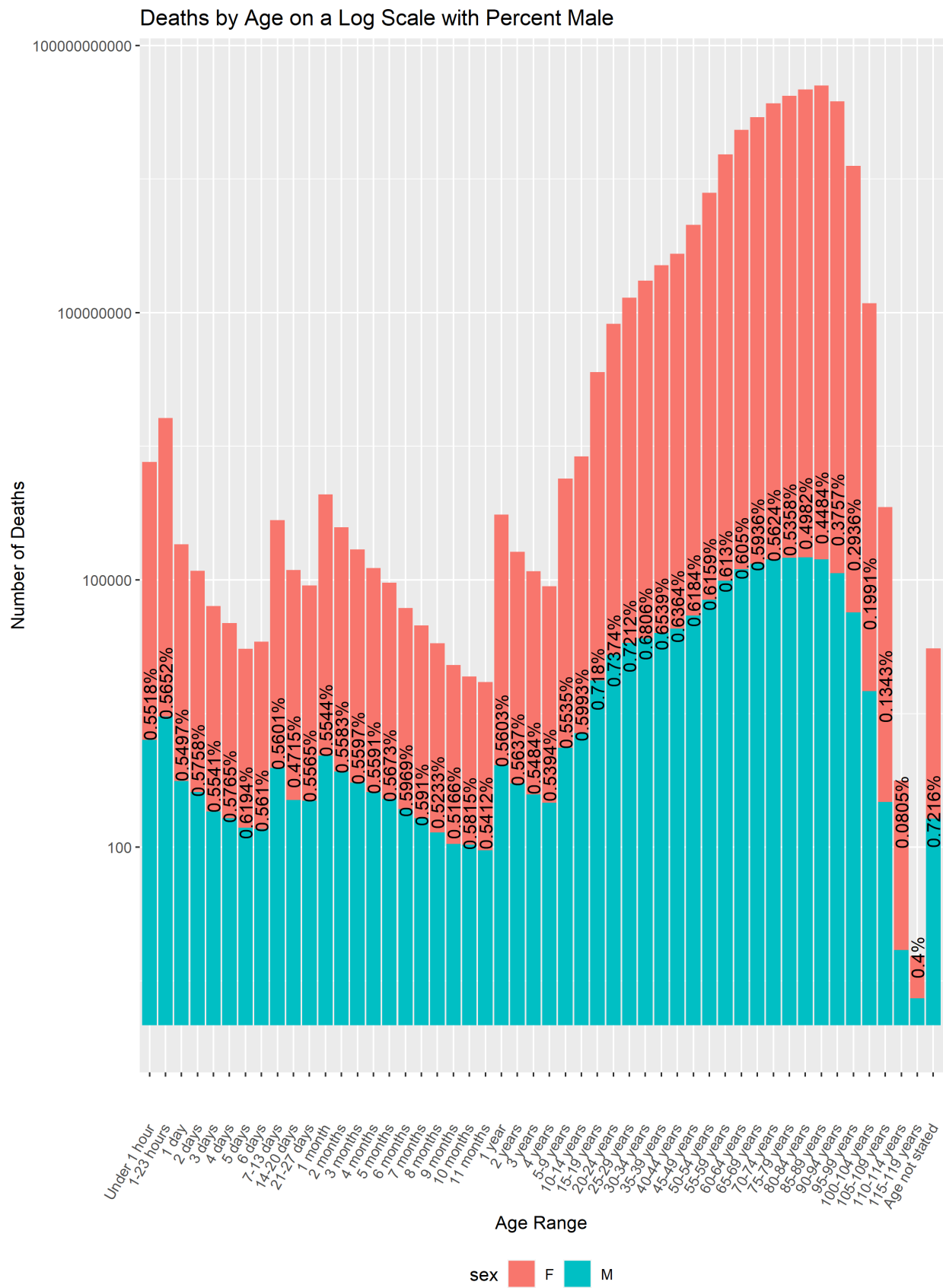
Deaths by Age



Deaths by Manner



Deaths by Age and Gender



Cause of Death by Education

