Homework 5

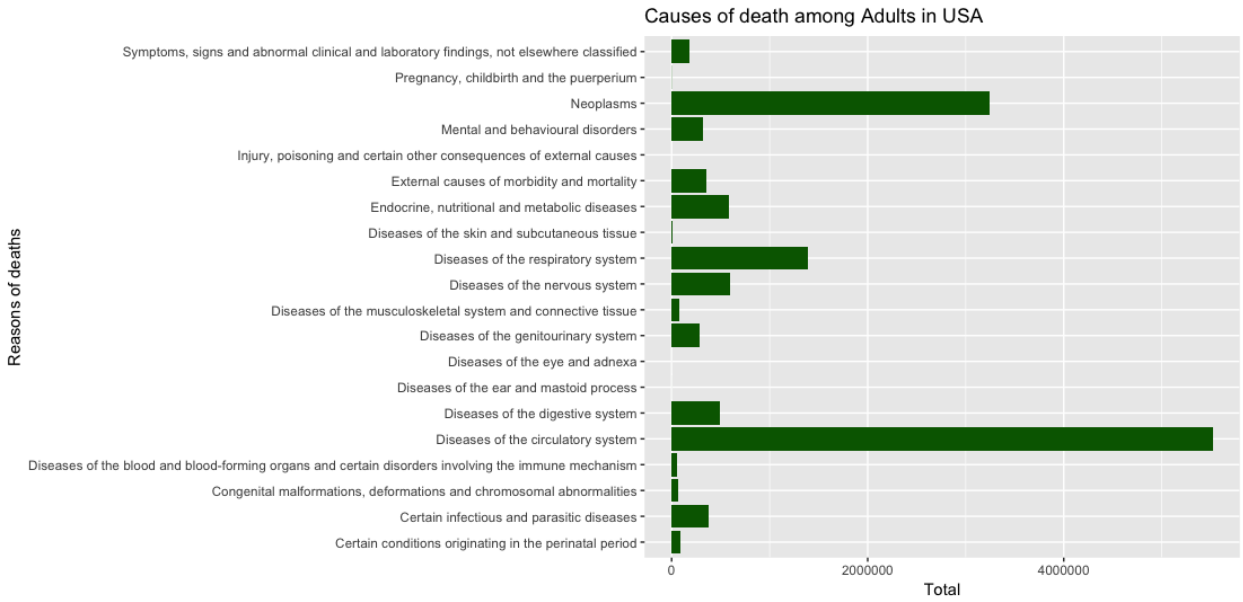
Visualizing U.S. Mortality Statistics for 2004, as classified by the World Health Organization's International Classification of Diseases 10 (ICD10)

Problem Statement: Choose the best information visualization technique(s) (e.g. statistical visualizations, tree maps, networks, etc.) that provides some insights about what causes death in the United States based on the WHO ICD10 classification statistics. You do not have to show the complete data set in your project, but you will need to provide some comparison between genders in the final product, where appropriate.

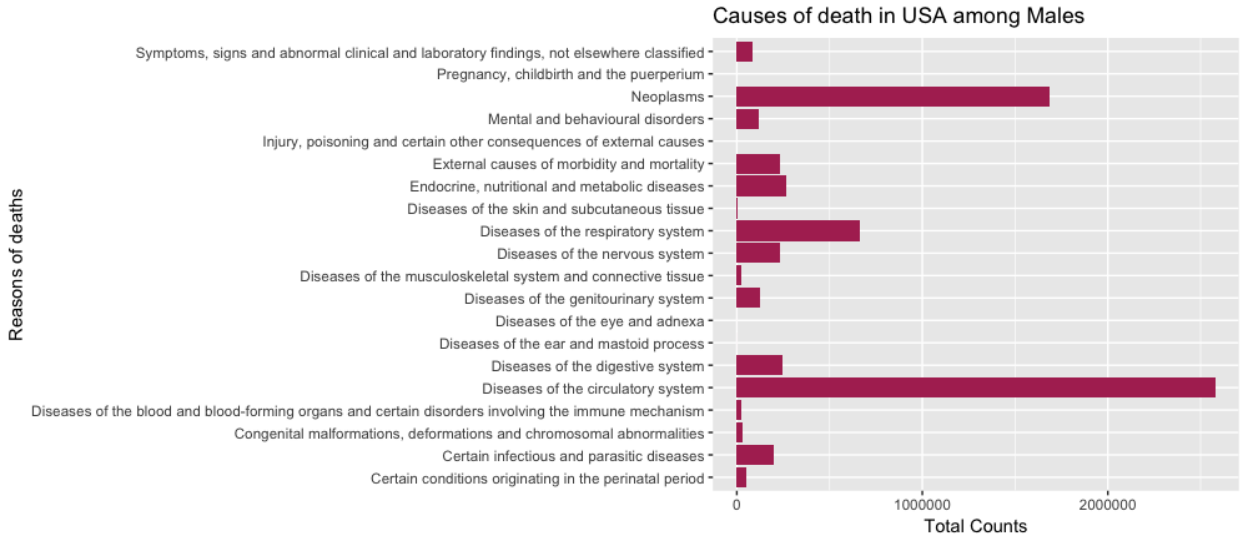
Data: The data had 16 attributes and 135878 observations. The data consists of first 20 chapters out of the 22 chapters which are defined as WHO ICD10 classification (More information. Click here). In each chapter, there are block codes and title which labels the reason of death. The data has total death counts by males and females. We were given data in 3 formats – csv, treeml, and graphml.

Tool: I chose to use the .csv file and create intuitive statistical visualizations to provide insights in R. R is a statistical programming language. It has a library called Grammar of Graphics (ggplot2) which I used to create the graphs.

Causes of death among Adults in the United States-



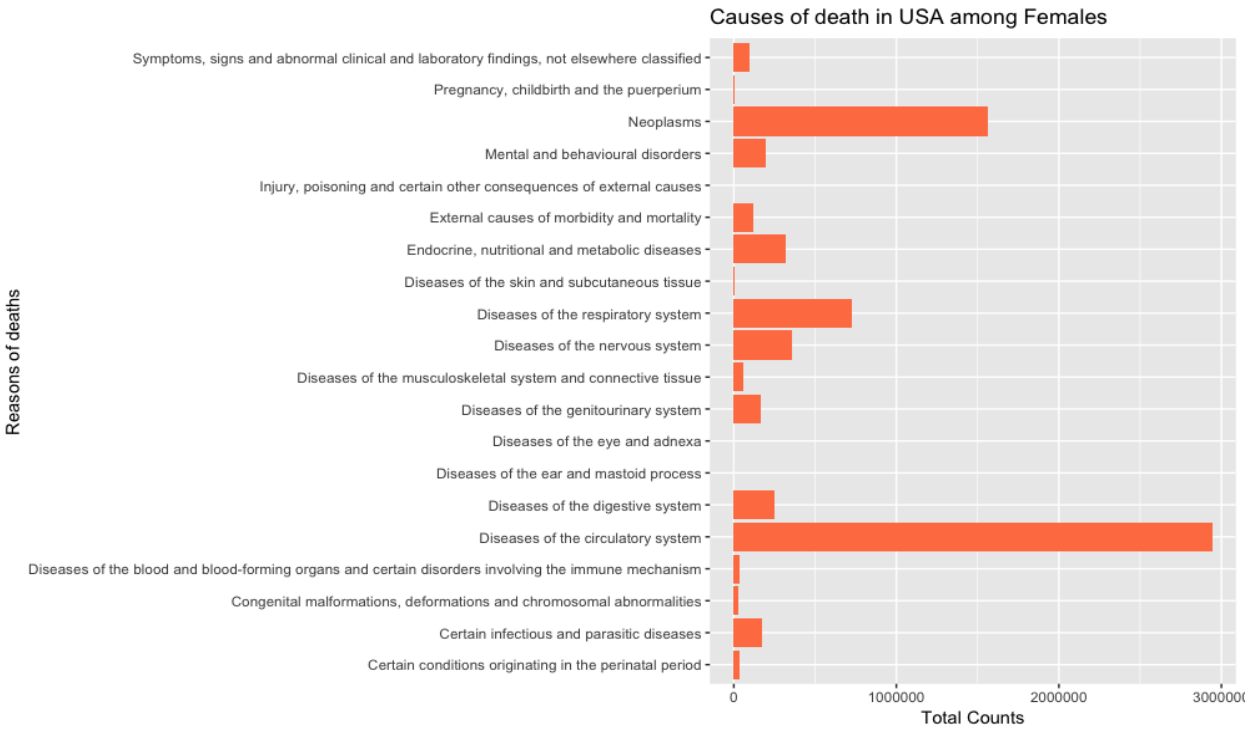
Causes of death among Males in the United States-



We can see that the **top three** reasons of death among Males are:

1. Diseases of the Circulatory system (more information)
2. Neoplasms (more information)
3. Diseases of the Respiratory system (more information)

These were also the top three reasons among Females which can be seen below:



Observations: There are more deaths among females than males. In the above graph, we can see that more females (n = 2,941,819) die due to circulatory diseases in the USA which is higher than males (n = 2,577,419). Almost, equal number of males (n = 1,682,771) die due to neoplasms as females (n = 1,561,220). Diseases of the eye and ear is uncommon among both genders. Injury, poisoning, and external causes data wasn't well recorded in the WHO data.

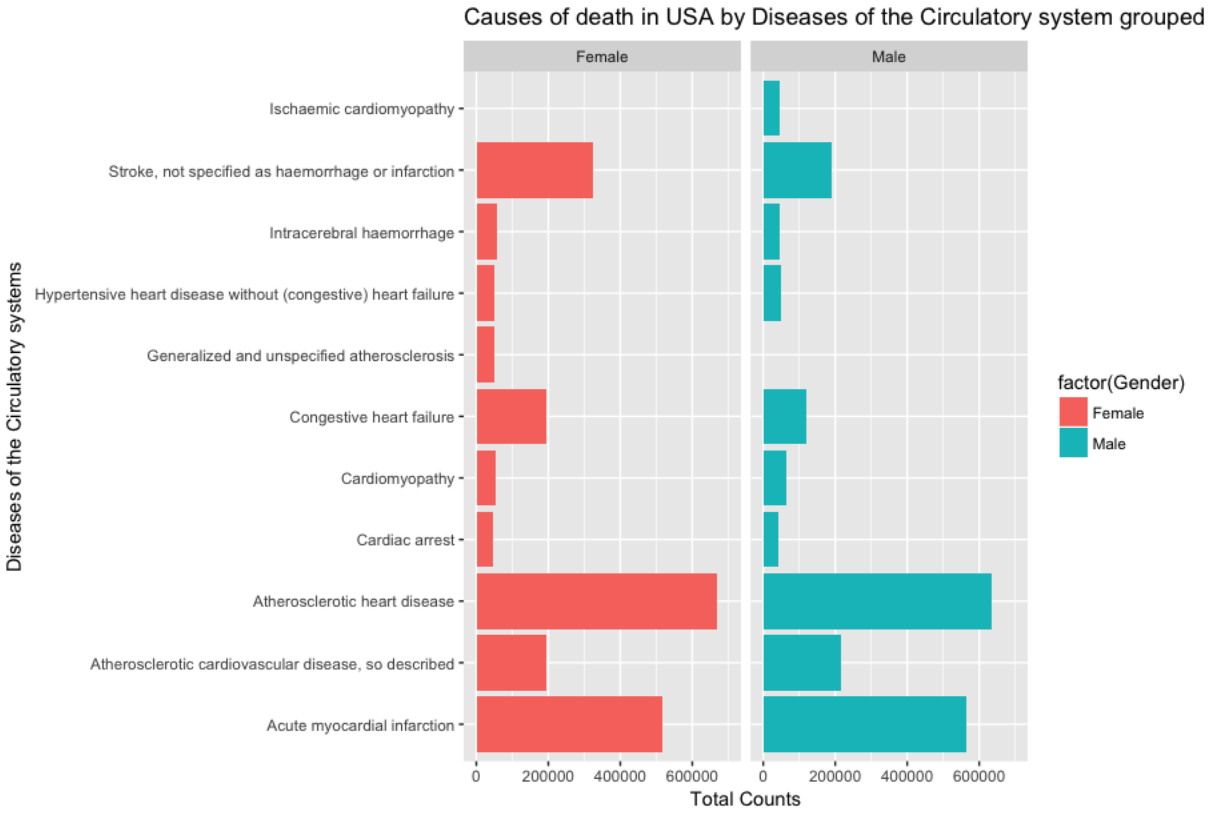
I decided to explore these top reasons further to answer some of the following questions:

Q1. Which disease that is common among Females is uncommon among Males and vice-versa?

Q2. Do reproductive organs contribute to any of these three categories?

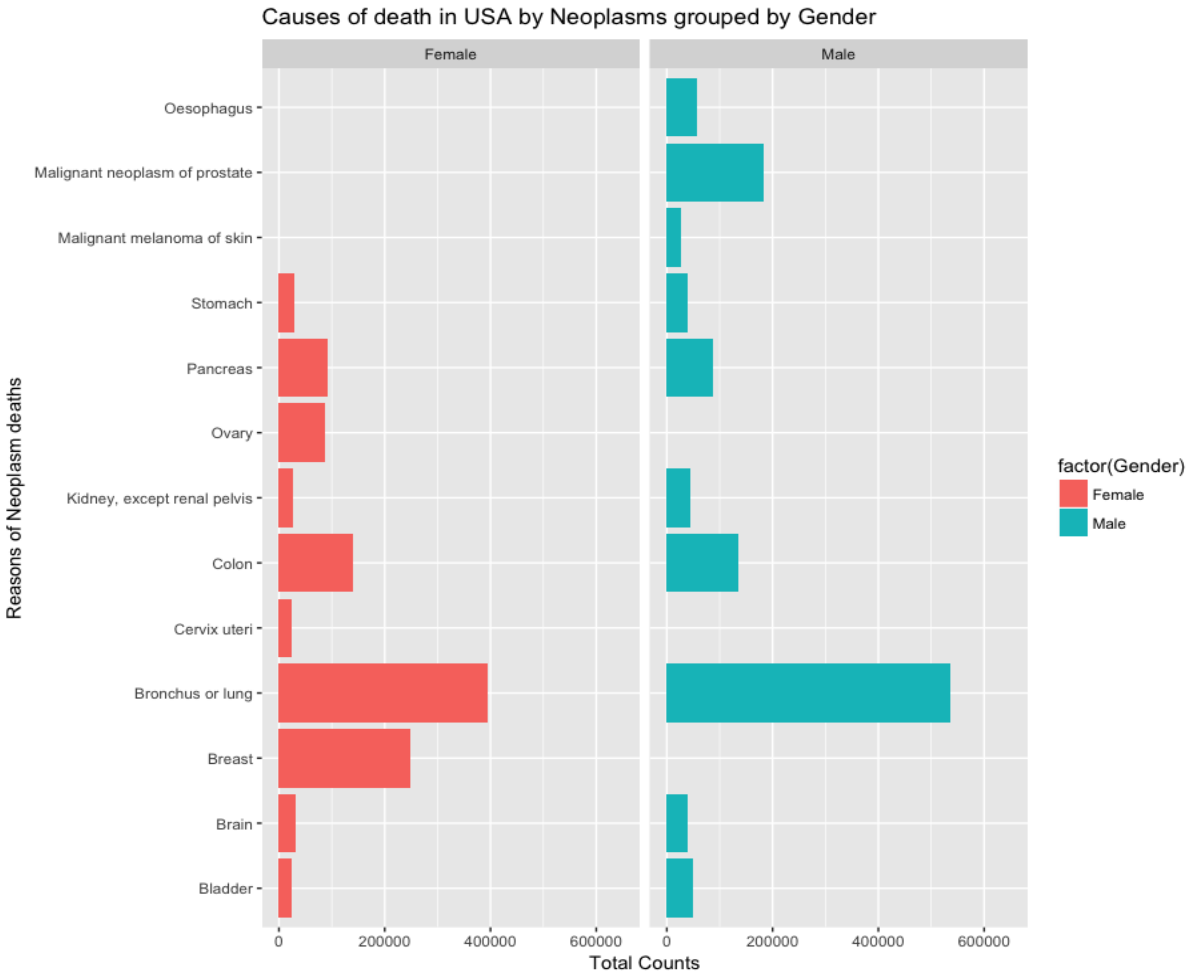
For convenience, I chose only **top ten reasons** among each category.

I. Diseases of the Circulatory system



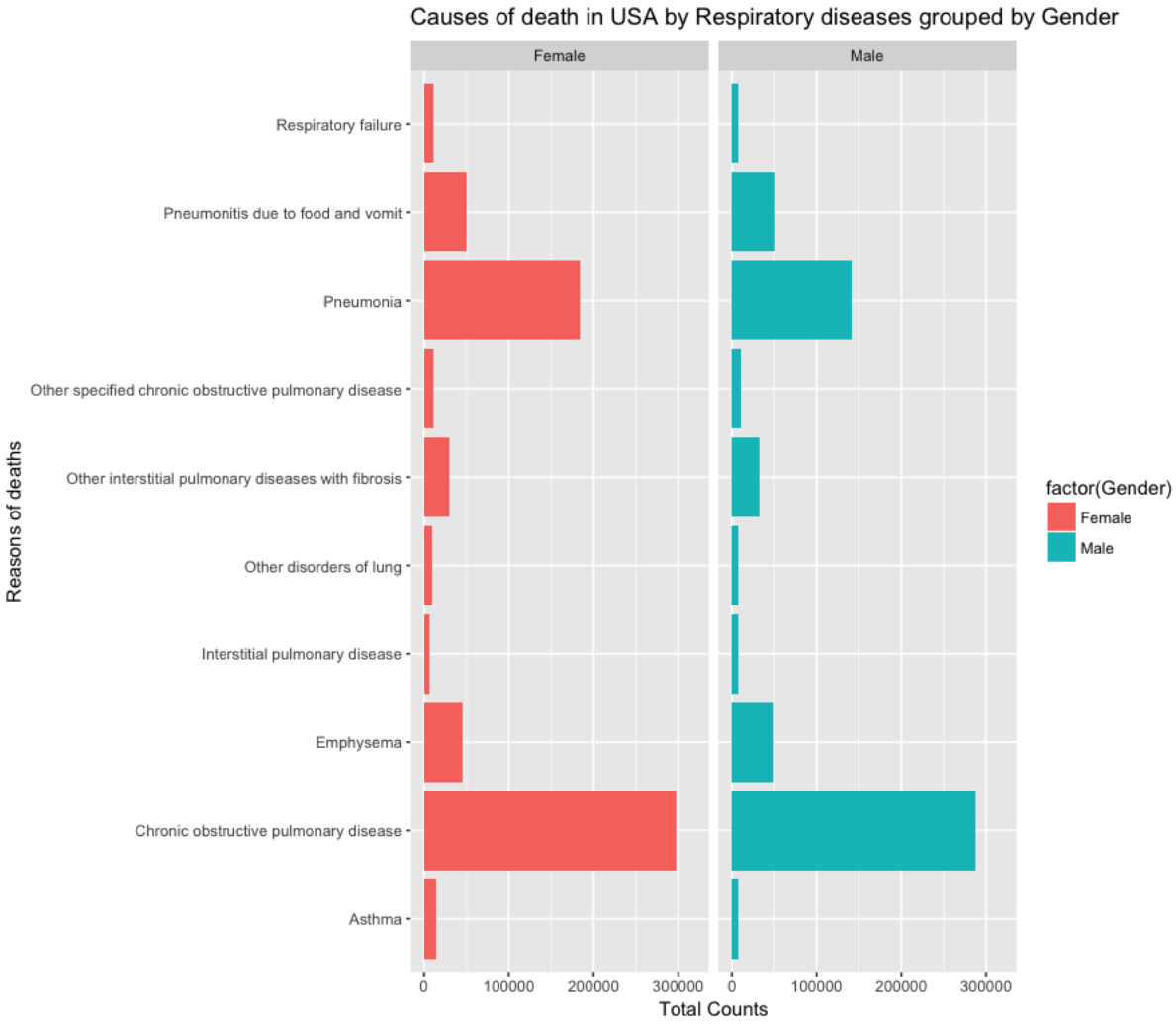
Observations: Death by **atherosclerotic heart disease** and **acute myocardial infarction** are the first two reasons irrespective of gender. Death by **stroke, hemorrhage or infarction** and **hypertensive heart disease with and without congestive heart failure** is more common among females than males. Further, death by **ischemic cardiomyopathy** is only caused among males.

II. Neoplasms



Observations: Although, death by neoplasms in organs like bronchus or lung is the top reason among both, males and females, the deaths by neoplasms in reproductive organs distinguishes across gender to a great extent. Females are more prone to deaths by neoplasms caused in **breasts** and **ovary** than males, who are prone to neoplasms by diseases caused in **colon** and **prostate**.

III. Diseases of the Respiratory system



Observations: For deaths by diseases of the circulatory system we see there is not much difference by gender. Nearly equal number of males and females die due to the same reasons. We can investigate further about deaths by diseases of the circulatory system by including various environment factors into the picture.