IMT 573: Module 2 Lab

Data Visualization

Tommy Huynh

Due: July 2, 2021

Collaborators: List collaborators here.

Objectives

In this module, we have focused on exploring data. Visualization is a great way to do this. Let's play around with visualization in this lab. You objective in this assignment is to create and reflect on different ways to visualize data. Think about what visuals you like and which enable you to tell compelling stories with data. And think about which charts you create might be misleading!

Instructions

Before beginning this assignment, please ensure you have access to R and RStudio; this can be on your own personal computer or on the IMT 573 R Studio Cloud.

1. Open the `02_lab_viz.Rmd` and save a copy to your local directory. Supply your solutions to the assignment by editing `02_lab_viz.Rmd`.

2. First, replace the "YOUR NAME HERE" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.

3. Be sure to include well-documented (e.g. commented) code chucks, figures, and clearly written text chunk explanations as necessary. Any figures should be clearly labeled and appropriately referenced within the text. Be sure that each visualization adds value to your written explanation; avoid redundancy – you do no need four different visualizations of the same pattern.

4. Collaboration on problem sets is fun and useful, and I encourage it, but each student must turn in an individual write-up in their own words as well as code/work that is their own. Regardless of whether you work with others, what you turn in must be your own work; this includes code and interpretation of results. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.

5. All materials and resources that you use (with the exception of lecture slides) must be appropriately referenced within your assignment.

6. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit`. When the PDF report is generated rename the knitted PDF file to `lab2_YourLastName_YourFirstName.pdf`, and submit the PDF file on Canvas.

Setup

In this lab you will need, at minimum, the following R packages.

```
# Load standard libraries
library(tidyverse)
```

```
## Warning: package 'tidyr' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
library(ggplot2)
library(scales)
```

```
## Warning: package 'scales' was built under R version 4.1.1
```

In the demonstration for Module 2, we encountered data from the sinking of the RMS Titanic in the North Atlantic Ocean in the early morning of 15 April 1912. We will revisit this data.

```
# Load titanic data
  titanic <- read_csv("titanic.csv")
```

Recall, our two questions for exploration in the demonstration were:

- Who were the Titanic passengers and what characteristics did they have?
- What passenger characteristics or other factors seem to be associated with survival?

You job is to create a new visualization for each of these questions and comment on their ability to speak to these questions. Have fun and be creative!
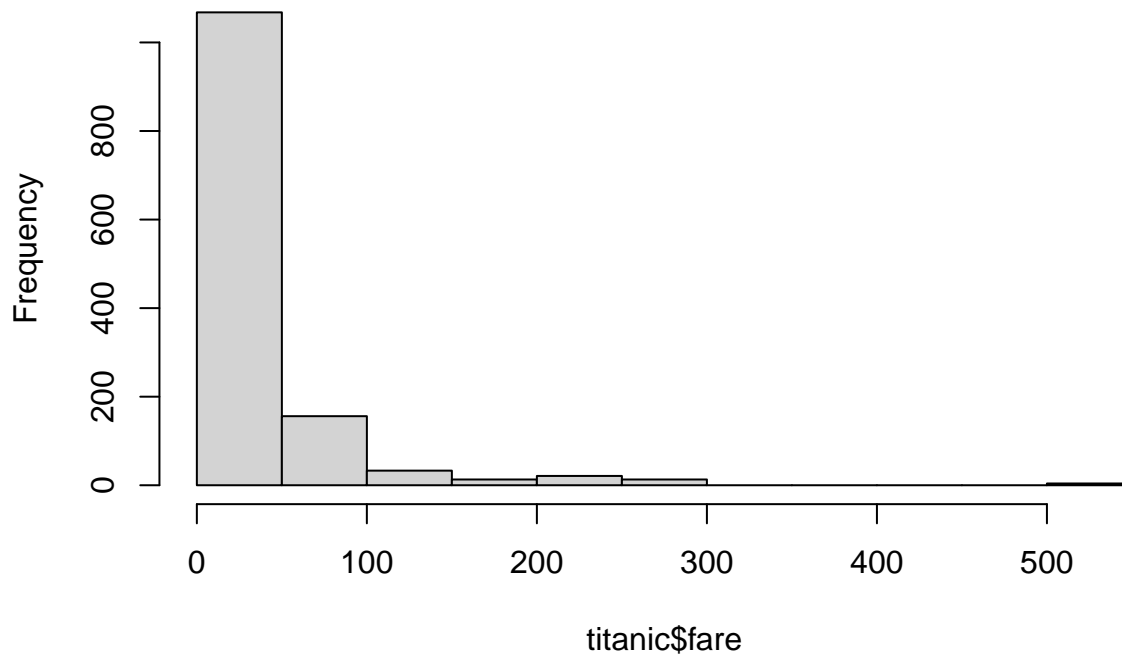
Problem 1: Who were the Titanic passengers and what characteristics did they have?

The best way to characterize titanic passengers is by their age and how much they paid for their ticket.
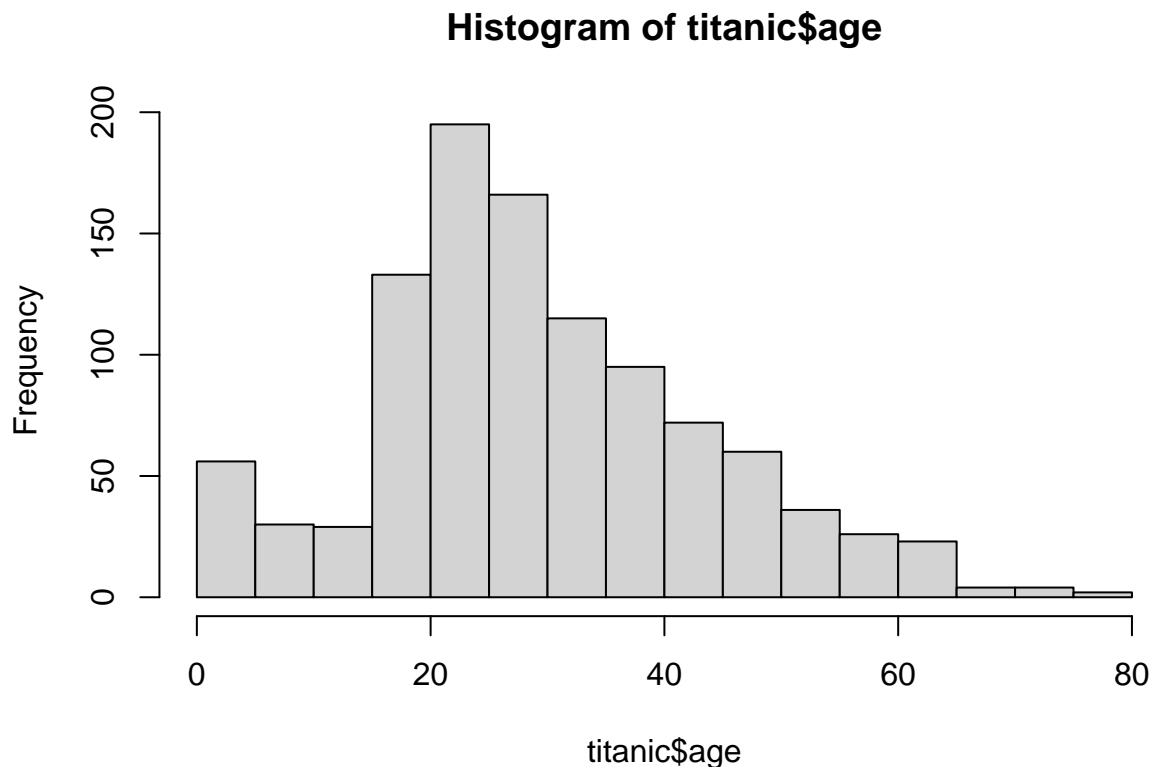
```
titanic <- as.data.frame(titanic)
```

```
hist(titanic$fare)
```

Histogram of titanic$fare


```
hist(titanic$age)
```

Histogram of titanic$age

Problem 2: What passenger characteristics or other factors seem to be associated with survival?

The most prominent factor that is associated with passenger survival is their gender. As you can see, a much larger share of females survived the sinking than males.

```r
#Counts total number of females and total number of female survivors
female_count <- sum(titanic$sex == "female")
female_survived <- sum(titanic$sex == "female" & titanic$survived == "1")

#Counts total number of males and total number of male survivors
male_count <- sum(titanic$sex == "male")
male_survived <- sum(titanic$sex == "male" & titanic$survived == "1")

#Calculates percentage of males and females that survived
male_survival <- male_survived/male_count
female_survival <- female_survived/female_count

survival_percentages <- c(male_survival, female_survival)
survival_percentages <- label_percent()(survival_percentages)

#Subsets titanic class into 3 different cabins
first_class <- subset(titanic, pclass == "1")
second_class <- subset(titanic, pclass == "2")
```

```
third_class <- subset(titanic, pclass == "3")

#Calculates survival percentage for each class
first_survival <- sum(first_class$survived == "1") / sum(titanic$pclass == "1")

second_survival <- sum(second_class$survived == "1") / sum(titanic$pclass == "2")

third_survival <- sum(third_class$survived == "1") / sum(titanic$pclass == "3")

#Combines all data into a dataframe to create data visualization
row_name <- c("First Class", "Second Class", "Third Class")
percentage <- c(first_survival, second_survival, third_survival)
percentage <- label_percent()(percentage)

class_survival <- data.frame(row_name, percentage)


row_name1 <- c("Male", "Female")
gender_survival <- data.frame(row_name1, survival_percentages)

#Data visualization
ggplot(class_survival, aes(x=row_name, y=percentage)) + geom_bar(stat="identity") + labs(x="Passenger Cl
```
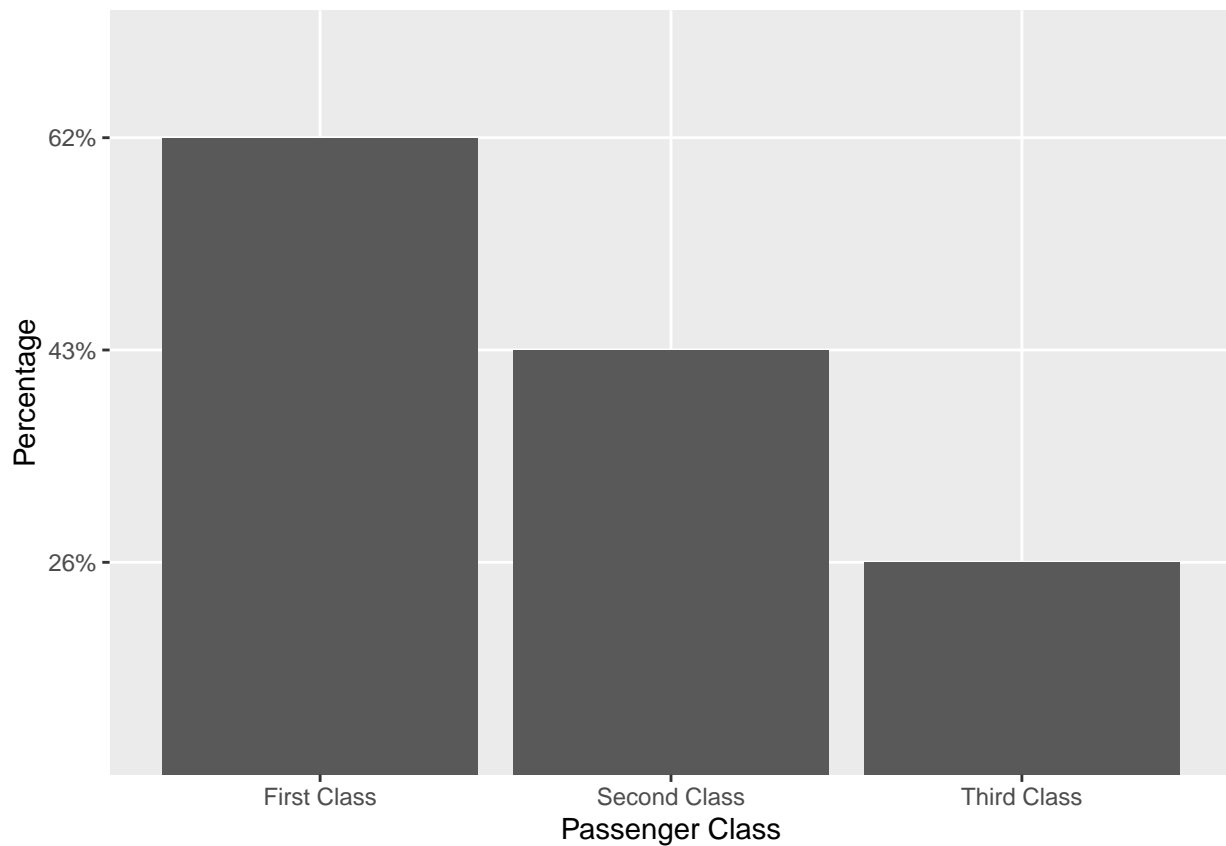
```
ggplot(gender_survival, aes(x=row_name1, y=survival_percentages)) + geom_bar(stat="identity") + labs(x=
```