

IMT 573: Module 2 Lab

Data Visualization

Ali Qazi

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. Your 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 chunks, 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 not 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)
```

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_data <- read_csv("data/titanic.csv")
```

- What passenger characteristics or other factors seem to be associated with survival?

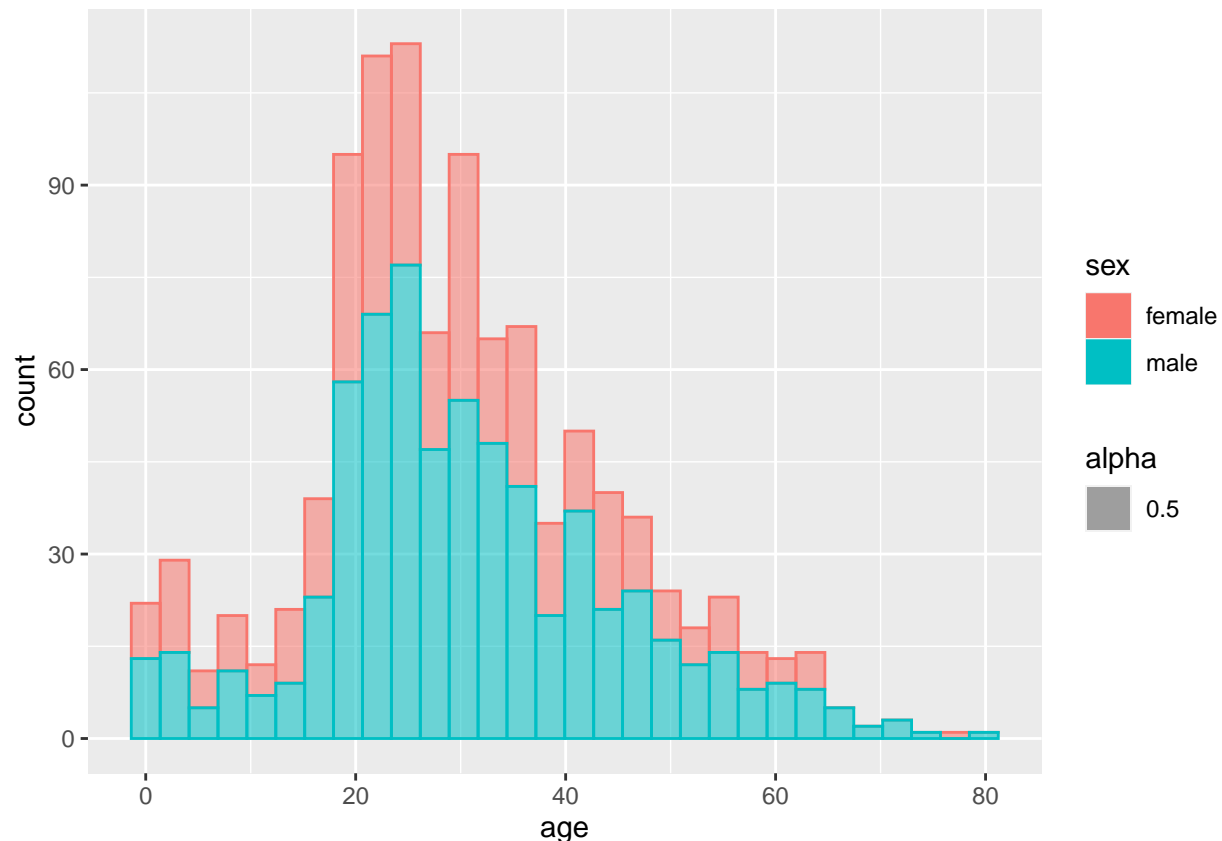
Your job is to create a new visualization for each of these questions and comment on their ability to speak to these questions. Have fun and be creative!

Problem 1: Who were the Titanic passengers and what characteristics did they have?

```
ggplot(data = titanic_data) +  
  geom_histogram(mapping = aes (x=age, color = sex, fill = sex, alpha = 0.5))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 263 rows containing non-finite values (stat_bin).
```



The graph shows the age and gender of the passengers of the Titanic. The graph clearly shows that the passengers were mostly between 20-40 years old and that there were more men on the Titanic.

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

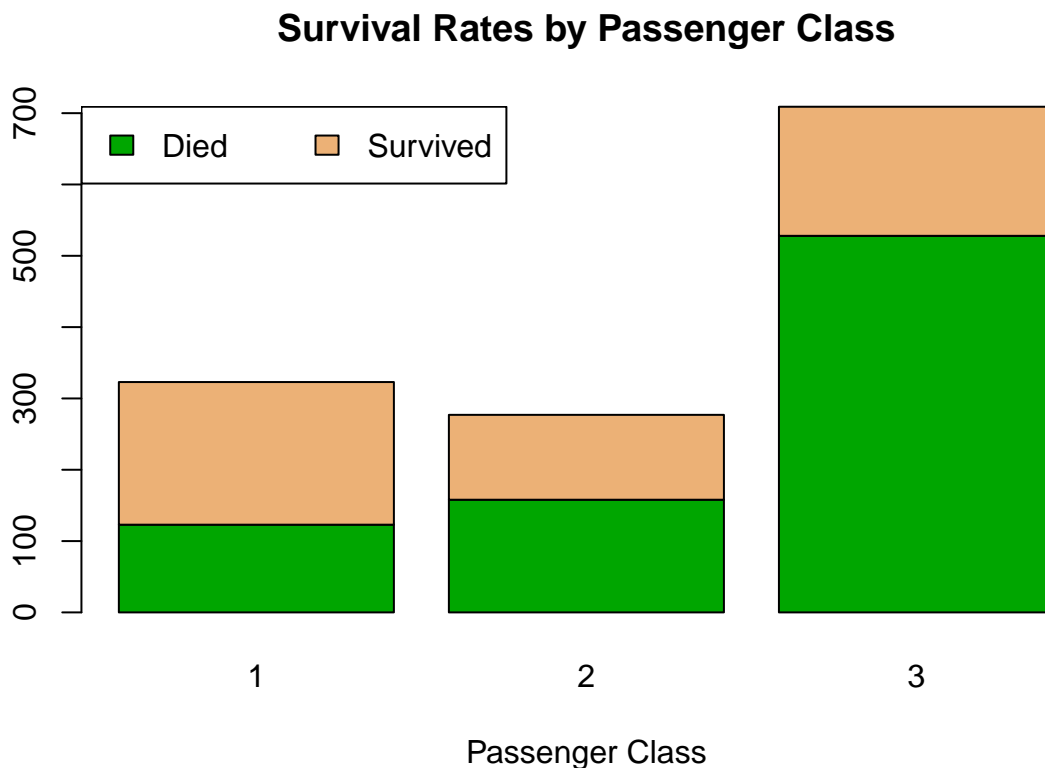
```

# Stacked barchart with color and legend
counts = table(titanic_data$survived, titanic_data$pclass)

# Builds barchart
barplot(counts,
main = "Survival Rates by Passenger Class",
xlab = "Passenger Class" ,
col = terrain.colors(3))

# Adding Legend
legend ("topleft",
legend = c("Died", "Survived") ,
fill = terrain.colors(3),
horiz = TRUE)

```



The stacked bar graph shows the survival rates of the passengers and characterizes it by the passengers class. The graph shows us that among class 1 more survived than died, in class 2 more died than survived and that in class 3 significantly more died than survived.