STA404/504 HW4

Pei Wang

08/21/2022

Learning Objectives:

- Basic visualizations and aesthetic settings

The items below will be considered for grading:

- The plots are correct, with professionalism.
- Axis labels and titles are correct and complete.
- Units are clearly labeled.
- Proper grammar in the write-up.
- The discussion and the story told is interesting and appropriate.

Question 1

Load the shuffledplaylist.csv dataset into R by the code below. This data includes information on songs from a shuffled Spotify playlist.

#read the dataset to R-studio
music<-read.csv("shuffled_playlist.csv")

```
# Read data set
# music <- read.csv("shuffledplaylist.csv")
```

How do the tempo values differ by genre? Use a side-by-side style of plot that shows both the distribution, as well as summary statistics to help tell a story of how they differ. Describe in a few sentences what this display tells us about the relationship between the variables.

Question 2

The data "**Arthritis**" is a dataset from the package "**vcd**". It contains data in Koch & Edwards (1988), which is collected from a double-blind clinical trial investigating a new treatment for rheumatoid arthritis. To access the dataset, you may use the following code:

install.packages("vcd")
library(vcd)
#Load Arthritis dataset (data frame)
data(Arthritis)
?Arthritis

Please create a stacked/grouped bar graph that summarizes the number of patients who received Placebo/Treatment, and the outcome of the treatment (None, Some, or Marked improvement). On the plot,

please have the numbers displayed in the center of the corresponding bar. Make sure appropriate aesthetic settings such as color, labels, titles (make it centered), legends, etc, are used, if applicable.