DSC520 McKibben Ex 4.2

Makayla McKibben

2024-07-02

```
knitr::opts_chunk$set(echo = TRUE)
```

```
# Class: DSC520
# Assignment: Exercise 4.2
# Author: Makayla McKibben
# Date: 07.02.2024

# Installing the necessary packages
install.packages("ggplot2")
library(ggplot2)
install.packages("pastecs")
library(pastecs)
install.packages("dplyr")
library(dplyr)

scores <- read.csv(file = 'scores.csv', header = TRUE, sep =",", stringsAsFactors = FALSE)
str(scores)

# Observational units are the number of students achieving a certain score and
# which section they were in.
# (Count) Number of students is quantitative
# Scores is quantitative
# Section is categorical

# Separating the data by section
scores.sports <- scores[scores$Section == "Sports",]
head(scores.sports)
scores.regular <- scores[scores$Section == "Regular",]
head(scores.regular)

# Plotting the scores for each section individually
scores.sports.bar.chart <- ggplot(scores.sports, aes(x = Score, y = Count))
scores.sports.bar.chart + geom_bar(stat="identity", color="blue", fill="white") +
  labs(title = "Sports Lecture Scores", x ="Score",
       y = "Count") + xlim(180, 420)
scores.regular.bar.chart <- ggplot(scores.regular, aes(x = Score, y = Count))
scores.regular.bar.chart + geom_bar(stat="identity", color="darkgreen", fill="white") +
  labs(title = "Regular Lecture Scores", x ="Score",
       y = "Count") + xlim(180, 420)

# Gathering Data about each set of students
```

```
min(scores.sports$Score)
max(scores.sports$Score)
mean(scores.sports$Score)
median(scores.sports$Score)
min(scores.regular$Score)
max(scores.regular$Score)
mean(scores.regular$Score)
median(scores.regular$Score)

# From looking at the graphs we are able to see that the Sports section had a wide distribution
# encompassing both the lowest and highest scores of either section. The Sports section also had
# a lower mean than the regular section indicating that the regular section scored higher on average.

# No each group of students were dispersed, with no section scoring absolutely higher than the other.
# The statistical tendency in this case is meaning that students in the regular lecture scored, on
# average, higher than their sports counterparts. Not every student scored higher though, and the
# sports category had the highest score overall, but comparing the means of each section we see that
# the regular lecture was ~20 points higher on average.

# Could be the time of day of the lecture, the gender make-up of the section, or the level of investmen
# from each course section. If a bunch of jocks (I was one so allow me to trash them slightly here) are
# only enrolled because it's a requirement and they just need to get a C to stay on the team they will
# not be as invested as the other section is in their performance. The sports section
# scored lower but there could be an indicator that that section appealed more to the sporty people.
```