The results below are generated from an R script.

```
# Assignment: ASSIGNMENT 4_1
# Name: Reppeto, Brian
# Date: 2023-06-27

library(tidyverse)
library(readxl)
library(ggplot2)
library(dplyr)
library(conflicted)
#library(plyr)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory
setwd("~/DSC520/Week_4")

## Load the `data` to
scores_df <- read.csv("scores.csv")

head(scores_df)
```

```
##   Count Score Section
## 1    10   200  Sports
## 2    10   205  Sports
## 3    20   235  Sports
## 4    10   240  Sports
## 5    10   250  Sports
## 6    10   265 Regular
```

```
## 1. The observational units are the students in the two sections
## 2. Section Type (Categorical),Course Grades (Categorical), total
##    Points Earned (Quantitative)

regular_section <- subset(scores_df, Section == "Regular", select=Count:Section)
sports_section <- subset(scores_df, Section == "Sports",select=Count:Section)

#head (regular_section)
#head (sports_section)
#table (scores_df['Section'])

ttl_reg_score <- sum(regular_section$Score)
ttl_sport_score <- sum(sports_section$Score)

ttl_reg_count <- sum(regular_section$Count)
ttl_sport_count <- sum(sports_section$Count)

avg_reg_score <- mean(regular_section$Score)
avg_sport_score <- mean(sports_section$Score)

med_reg_score <- median(regular_section$Score)
med_sport_score <- median(sports_section$Score)

sd_reg_score <- sd(regular_section$Score)
sd_sport_score <- sd(sports_section$Score)
```

```r
iqr_reg_score <- IQR(regular_section$Score)
iqr_sport_score <- IQR(sports_section$Score)

avg_reg_count <- mean(regular_section$Count)
avg_sport_count <- mean(sports_section$Count)

ttl_reg_score
```

```
## [1] 6225
```

```r
ttl_sport_score
```

```
## [1] 5840
```

```r
ttl_reg_count
```

```
## [1] 290
```

```r
ttl_sport_count
```

```
## [1] 260
```

```r
avg_reg_score
```

```
## [1] 327.6316
```

```r
avg_sport_score
```

```
## [1] 307.3684
```

```r
avg_reg_count
```

```
## [1] 15.26316
```

```r
avg_sport_count
```

```
## [1] 13.68421
```

```r
med_reg_score
```

```
## [1] 325
```

```r
med_sport_score
```

```
## [1] 315
```

```r
sd_reg_score
```

```
## [1] 33.26528
```

```r
sd_sport_score
```

```
## [1] 58.0318
```

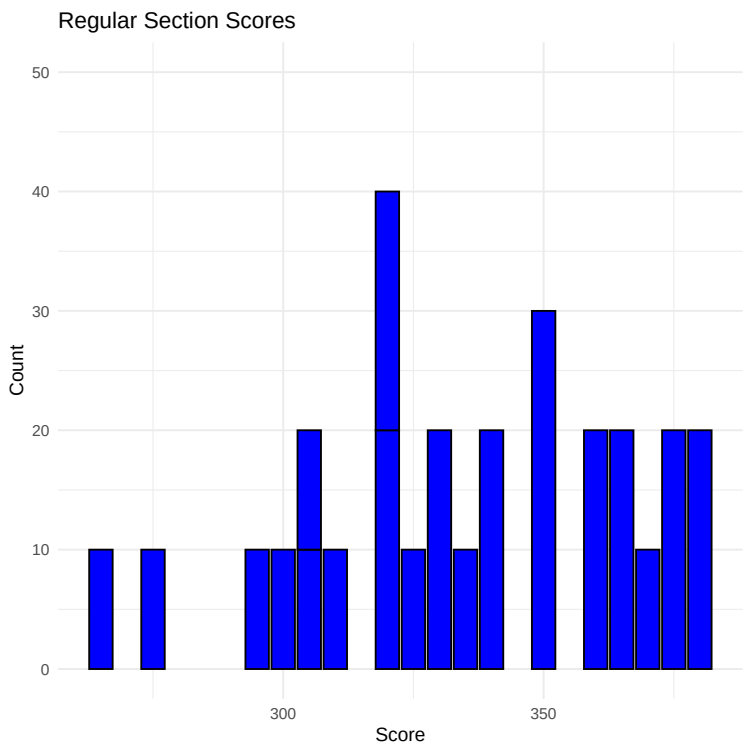
```r
iqr_reg_score
```

```
## [1] 50
```

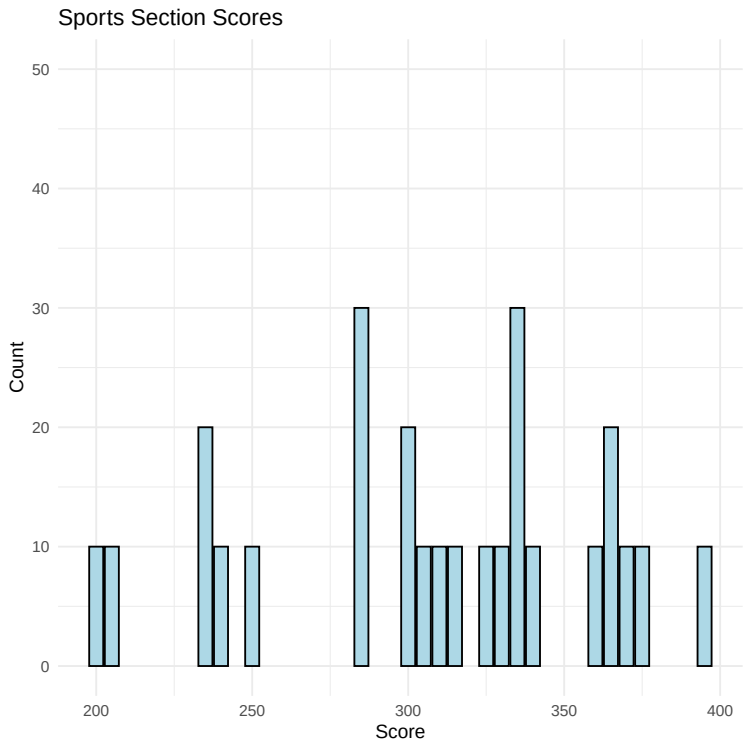
```r
iqr_sport_score
```

```
## [1] 82.5
```

```r
ggplot(regular_section, aes(x = Score, y = Count)) +
  geom_bar(stat ="identity", fill ="blue", color = "black") +
  ylim(0,50)+
  labs(title = "Regular Section Scores",
       x = "Score",
       y = "Count")
```




```
# Plot for Sports Section
ggplot(sports_section, aes(x = Score, y = Count)) +
  geom_bar(stat ="identity", fill ="lightblue", color = "black") +
  ylim(0,50)+
  labs(title = "Sports Section Scores",
       x = "Score",
       y = "Count")
```




```
## We cannot definitively say that one section tended to score more points than
## the other just by looking at the plots.  The avg scores for each are below.
## The avg_reg_score is 327.6316 which is just slightly larger than the
## avg_sport_score which is 307.3684.  The difference in avg scores is
## insignificant and does not determine if one is better than the other.


## No, based on the plots, we can see that there is some overlap in scores
## between the two sections.
## Statistical tendency means that on average, one section might have higher
## scores than the other, but individual variations exist.


## The students' prior knowledge or interest in sports could be an additional
## variable influencing the scores.
## If one section had more sports enthusiasts or students with prior
## knowledge in sports-related applications, they might perform better in the
## sports-themed section.
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.4.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
```

```
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] conflicted_1.2.0 readxl_1.4.2     lubridate_1.9.2  forcats_1.0.0    stringr_1.5.0
##  [6] dplyr_1.1.2      purrr_1.0.1      readr_2.1.4      tidyr_1.3.0      tibble_3.2.1
## [11] ggplot2_3.4.2    tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] gtable_0.3.3     highr_0.10       compiler_4.3.0   tidyselect_1.2.0 scales_1.2.1
##  [6] fastmap_1.1.1    R6_2.5.1         labeling_0.4.2   generics_0.1.3   knitr_1.43
## [11] munsell_0.5.0    pillar_1.9.0     tzdb_0.4.0       rlang_1.1.1      utf8_1.2.3
## [16] cachem_1.0.8     stringi_1.7.12   xfun_0.39        timechange_0.2.0 memoise_2.0.1
## [21] cli_3.6.1        withr_2.5.0      magrittr_2.0.3   grid_4.3.0       rstudioapi_0.14
## [26] hms_1.1.3        lifecycle_1.0.3  vctrs_0.6.3      evaluate_0.21    glue_1.6.2
## [31] farver_2.1.1     cellranger_1.1.0 fansi_1.0.4      colorspace_2.1-0 tools_4.3.0
## [36] pkgconfig_2.0.3
```

```
Sys.time()
```

```
## [1] "2023-07-02 18:39:39 EDT"
```