# Homework 02 - Describing Data with Numbers

Prior to beginning this analysis, download the .qmd file from D2L Brightspace. This will allow you to work in a pre-formatted document this week. As you work through, take time to notice how the document is formatted - so you can format yours in a similar way in the future. Also download the uafootball.csv from D2L Brightspace and save it in the same location as your .qmd file.

## Problem 1

The University of Arizona Wildcat varsity football team plays 12 regular season games each year. The data set uafootball contains the difference in scores for the ten seasons from 2014 to 2024 with 2020 excluded due to Covid-19 (they played a shortened season that year).

### Part a)

First we will read in the data. The following code block is used to do so. To run the block, either click on the green play triangle in the upper right hand corner of the block or click on the line you wish to run and click run in the top right or cmd+enter or ctrl+enter depending on your device.

```r
football <- read.csv("uafootball.csv")
```

### Part b)

How many games did the Wildcats win each year?

```r
games_won <- c()

for (i in 1:10) {
  temp_data <- football[, i]
  temp_count <- sum(temp_data > 0)
  games_won <- c(games_won, temp_count)
}

print(games_won)
```
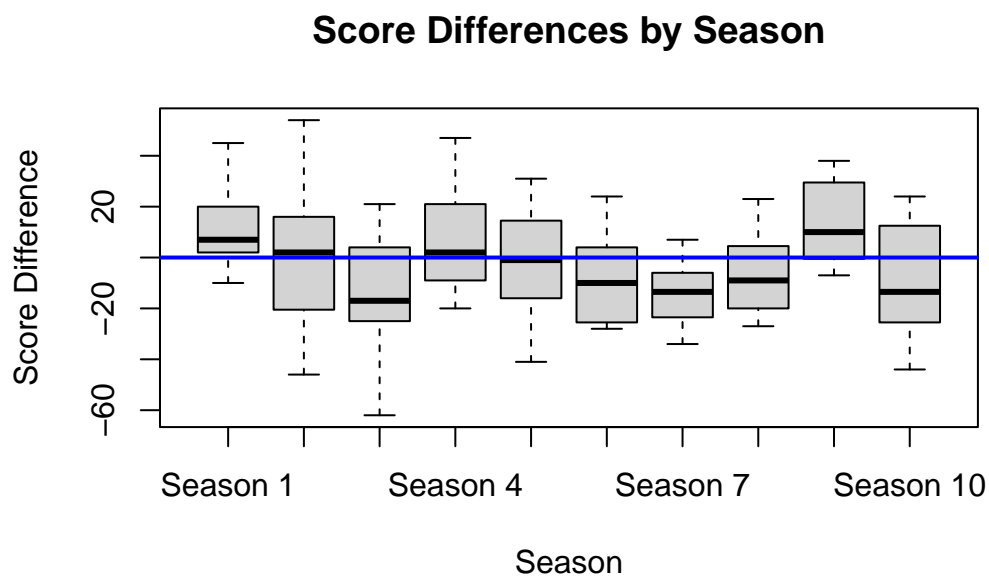
```
 [1] 10  6  3  7  5  4  1  5  9  4
```

**Part c)**

Give side by side box plots of the scores.

```r
boxplot(football[, 1:10],
        main = "Score Differences by Season",
        xlab = "Season",
        ylab = "Score Difference",
        names = paste("Season", 1:10))

abline(h = 0, col = "blue", lwd = 2)
```

**Score Differences by Season**



**Part d)**

What is the mean difference in scores for the seasons?

```r
avg_score_dif <- apply(football, 2, mean)
avg_score_dif
```

```
          f14          f15          f16          f17          f18          f19
   11.16666667   0.08333333 -13.41666667   7.75000000  -1.25000000  -8.83333333
```

```
        f21           f22           f23           f24
-14.25000000   -5.66666667   13.50000000  -10.00000000
```

**Part e)**

What is the standard deviation for the difference in scores for the seasons?

```
avg_score_dif_spread <- apply(football, 2, sd)
avg_score_dif_spread
```

```
     f14      f15      f16      f17      f18      f19      f21      f22
15.48509 29.38911 23.95245 21.10202 22.99061 17.36157 11.98579 16.42799
     f23      f24
17.06938 23.67200
```

**Part f)**

Write a short paragraph to describe how the Wildcat football team has performed over the past ten seasons. Use some of the summaries in your description.

The football team has performed poorly. They are frequently outscored, and do not have many wins. Their performances are also inconsistent with the high standard deviations, sometimes they win big and sometimes they lose big.

**Part g)**

Before submitting, clean up this document. Anywhere in the code where there are NOTES and unnecessary information, remove those. You can leave in comments that you find necessary but think about making this document as nice as possible for a classmate to read. Once you have done so, render the document and upload to Gradescope.

Make sure to tag your pages in Gradescope! This means that each problem should have the appropriate pages associated with it.