DSC520 – Assignment 4.2.1

Supraja Rapuru

```
> library(ggplot2)
> theme_set(theme_minimal())
> library(pastecs)
> setwd("/Users/Supraja/dsc520")
> Scores_df <- read.csv("data/scores.csv")
> str(Scores_df)
'data.frame':	38 obs. of  3 variables:
 $ Count  : int  10 10 20 10 10 10 10 30 10 10 ...
 $ Score  : int  200 205 235 240 250 265 275 285 295 300 ...
 $ Section: chr  "Sports" "Sports" "Sports" "Sports" ...
> str(Scores_df)
'data.frame':	38 obs. of  3 variables:
 $ Count  : int  10 10 20 10 10 10 10 30 10 10 ...
 $ Score  : int  200 205 235 240 250 265 275 285 295 300 ...
 $ Section: chr  "Sports" "Sports" "Sports" "Sports" ...
> summary(Scores_df)
     Count          Score        Section
 Min.   :10.00   Min.   :200.0   Length:38
 1st Qu.:10.00   1st Qu.:300.0   Class :character
 Median :10.00   Median :322.5   Mode  :character
 Mean   :14.47   Mean   :317.5
 3rd Qu.:20.00   3rd Qu.:357.5
 Max.   :30.00   Max.   :395.0
> reg_df <-Scores_df[which(Scores_df$Section=='Regular'),]
> head(reg_df)
  Count Score Section
```

```
6     10   265 Regular
7     10   275 Regular
9     10   295 Regular
10    10   300 Regular
13    10   305 Regular
14    10   310 Regular
> sport_df<-Scores_df[which(Scores_df$Section=='Sports'),]
> head(sport_df)
  Count Score Section
1    10   200  Sports
2    10   205  Sports
3    20   235  Sports
4    10   240  Sports
5    10   250  Sports
8    30   285  Sports
> plot(reg_df$Score,reg_df$Count,type='h',xaxt="n",xlab="Score in Regular Section",ylab="Count of Students")
> axis(1, at = seq(200, 400, by = 10), las=2)
> plot(reg_df$Score,reg_df$Count,type='b',xaxt="n",xlab="Score in Regular Section",ylab="Count of Students")
> axis(1, at = seq(200, 400, by = 10), las=2)
> plot(reg_df$Score,reg_df$Count,type='h',xaxt="n",xlab="Score in Regular Section",ylab="Count of Students")
> plot(sport_df$Score,sport_df$Count,type='h',xaxt="n",xlab="Score in Sports Section",ylab="Count of Students")
> axis(1, at = seq(200, 400, by = 10), las=2)
> stat.desc(reg_df[,1:2], basic=TRUE, desc=TRUE, norm=FALSE, p=0.95)
               Count       Score
nbr.val     19.0000000  19.0000000
nbr.null     0.0000000   0.0000000
```

```
nbr.na        0.0000000    0.0000000
min          10.0000000  265.0000000
max          30.0000000  380.0000000
range        20.0000000  115.0000000
sum         290.0000000 6225.0000000
median       10.0000000  325.0000000
mean         15.2631579  327.6315789
SE.mean       1.4035088    7.6315789
CI.mean.0.95  2.9486625   16.0333524
var          37.4269006 1106.5789474
std.dev       6.1177529   33.2652814
coef.var      0.4008183    0.1015326
> stat.desc(sport_df[,1:2], basic=TRUE, desc=TRUE, norm=FALSE, p=0.95)
              Count       Score
nbr.val      19.0000000   19.0000000
nbr.null      0.0000000    0.0000000
nbr.na        0.0000000    0.0000000
min          10.0000000  200.0000000
max          30.0000000  395.0000000
range        20.0000000  195.0000000
sum         260.0000000 5840.0000000
median       10.0000000  315.0000000
mean         13.6842105  307.3684211
SE.mean       1.5691705   13.3134085
CI.mean.0.95  3.2967049   27.9704333
var          46.7836257 3367.6900585
std.dev       6.8398557   58.0318021
coef.var      0.4998356    0.1888021
> bar <- ggplot(Scores_df, aes(Score,Count, fill = Section))
```

```
> bar + stat_summary(fun = mean, geom = "bar", position="dodge",width = 8)+ facet_wrap( ~ Section)
Warning messages:
1: position_dodge requires non-overlapping x intervals
2: position_dodge requires non-overlapping x intervals
```