

Assignment: ASSIGNMENT 3

Name: Campbell, Christian

Date: 7/1/2023

Load the ggplot2 package

```
library(ggplot2) theme_set(theme_minimal())
```

Install and load the pastec, plyr and dplyr package

```
library(plyr) library(dplyr) library(pastecs)
```

Set the working directory to the root of your DSC 520 directory

```
setwd("C:/Users/Owner/Documents/Important/Bellevue/Statistics for Data Science/dsc520-master")
```

Load the data/scores.csv to

```
course_df <- read.csv("data/scores.csv")  
print(course_df)
```

Create one variable to hold a subset of your data set that contains only the Regular

Section and one variable for the Sports Section.

```
reg_sec <- filter(course_df, Section == "Regular") print(reg_sec)  
sports_sec <- filter(course_df, Section == "Sports") print(sports_sec)
```

Use the Plot function to plot each Sections scores and the number of students achieving

that score. Use additional Plot Arguments to label the graph and give each axis an appropriate label.

```
ggplot(reg_sec, aes(x=Score, y=Count)) + geom_col() + ggtitle("Regular Section") + xlab("Score (Points)") + ylab("Count (Student)")  
ggplot(sports_sec, aes(x=Score, y=Count)) + geom_col() + ggtitle("Sports Section") + xlab("Score (Points)") + ylab("Count (Students)")
```

Load a data set I made up

```
mu_df <- read.csv("data/Made_Up.csv") ls(mu_df)
```

2.1 Use the apply function on a variable in your data set.

```
apply(mu_df[, c(1,2,3,4)], 2, mean)
```

2.2 Use the aggregate function on a variable in your data set.

```
aggregate(Age~Seniority, mu_df, mean)
```

2.3 Use the plyr function on a variable in your dataset – more specifically, I want to see you split some data,

perform a modification to the data, and then bring it back together.

I extracted the weekly_wrk_hrs from the df and made it its own df

```
weekly_wrk_hrs <- mu_df[2]
```

I created a new df that didn't contain weekly_wrk_hrs

```
new_mu_df <- mu_df[, c(1,3,4,5)]
```

I transformed the workweek df to a work day df.

```
daily_wrk_hrs <- weekly_wrk_hrs/5
```

I changed the name of the column.

```
colnames(daily_wrk_hrs) = "daily_work_hrs"
```

I combined daily work hours with new_mu_df thereby replacing weekly_wrk_hrs.

```
actl_mu_df <- bind_cols(new_mu_df, daily_wrk_hrs)
```

2.4 Check distributions of the data.

```
stat.desc(actl_mu_df)
```

2.6 Create at least 2 new variables.

```
Pay_per_hr <- c(40,49,71,32,21,28,26,51,34,39,21,23,29,41,52,45,43,29,35,41) vacation_hrs <- c(650,607,896,203,80,302,491,6
```

I'm combining the newly created variables with the established df.

```
actl_mu_df <- bind_cols(new_mu_df, Pay_per_hr, vacation_hrs)
```

I'm giving the respective columns a label.

```
colnames(actl_mu_df)[5] = "Pay_per_hr" colnames(actl_mu_df)[6] = "Vacation_hrs"  
print(actl_mu_df)
```