Final Report Exercise

Minh Tri Bui

Problem 1

Create a vector of five numbers of your choice between 0 and 10, save that vector to an object, and use the sum() function to calculate the sum of the numbers.

```
object = c(1:3,6,9)
sum(object)
```

```
## [1] 21
```

Problem 2

Create a data frame that includes two columns. One column must have the numbers 1 through 5, and the other column must have the numbers 6 through 10. The first column must be named "alpha" and the second column must be named "beta". Name the object "my_dat". Display the data.

Put your code and solution here:

```
alpha = c(1:5)
beta = c(6:10)
my_dat = data.frame(alpha,beta)
my_dat
```

```
##   alpha beta
## 1     1    6
## 2     2    7
## 3     3    8
## 4     4    9
## 5     5   10
```

Problem 3

Using the data frame created in Problem 2, use the summary() command a create a five-number summary for the column named "beta".

Put your code and solution here:

```
summary(my_dat$beta)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       6       7       8       8       9      10
```
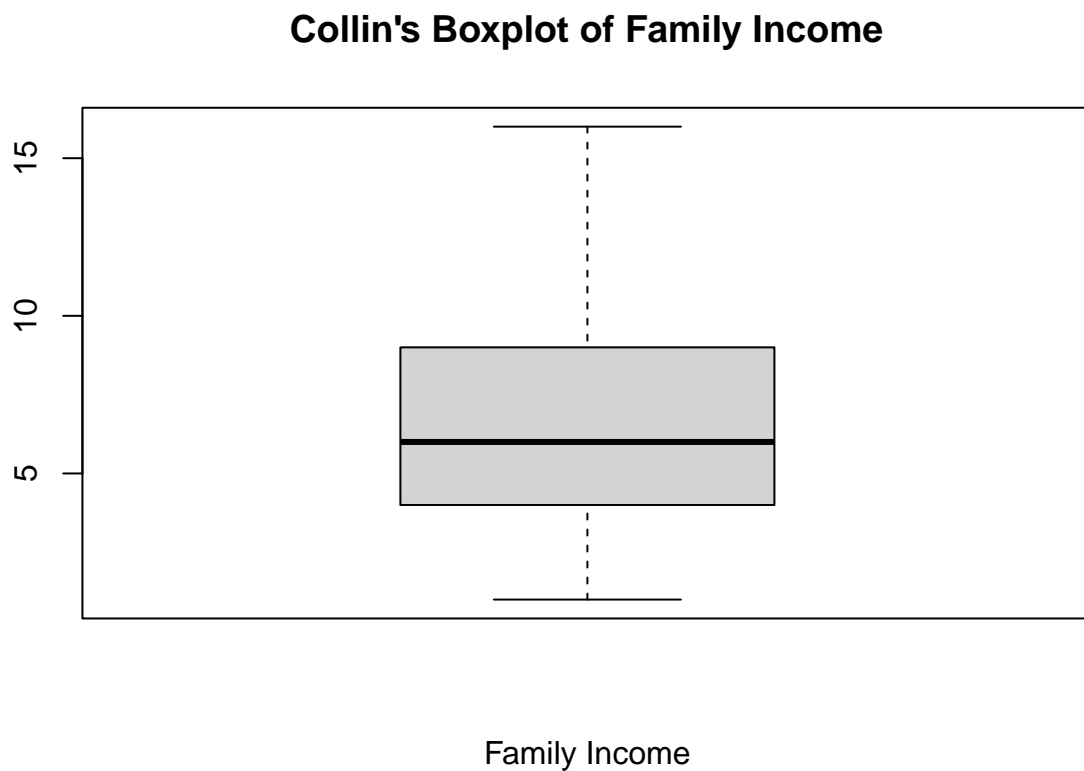
Problem 4

There is code for importing the example survey data that will run automatically in the setup chunk for this report (Line 13). Using that data, make a boxplot of the Family Income column using the Base R function (not a figure drawn using qplot). Include your name in the title for the plot. Your name should be in the title. Relabel that x-axis as "Family Income".

Hint: consult the codebook to identify the correct column name for the family income question.

Put your code and solution here:

```
boxplot(dat$faminc_new, data = dat, xlab = "Family Income", main = "Collin's Boxplot of Family Income")
```


Collin's Boxplot of Family Income

Family Income

Problem 5

Using the survey data, filter to subset the survey data so you only have male survey respondents who live in the northwest or midwest of the United States, are married, and identify as being interested in the news most of the time.

Use the str() function to provide information about the resulting dataset.

Put your code and solution here:

```
a= filter(dat, gender == 1 & region <3 & marstat == 1 & newsint == 1)
str(a)
```

```
## 'data.frame':     75 obs. of  25 variables:
##  $ caseid      : int   420208067 412948037 411855595 414480371 416806180 414729651 412021973 41234883
##  $ region      : int   2 1 1 1 2 2 1 2 1 2 ...
##  $ gender      : int   1 1 1 1 1 1 1 1 1 1 1 ...
##  $ educ        : int   3 5 6 5 3 2 6 5 5 5 ...
##  $ edloan      : int   2 1 2 1 1 2 2 2 2 2 ...
##  $ race        : int   1 1 1 1 1 1 1 1 1 1 1 ...
##  $ hispanic    : int   2 2 2 2 2 2 2 2 2 2 2 ...
##  $ employ      : int   1 1 1 1 1 5 5 1 1 1 ...
##  $ marstat     : int   1 1 1 1 1 1 1 1 1 1 1 ...
##  $ pid7        : int   4 1 4 4 6 2 1 1 1 7 ...
##  $ ideo5       : int   5 1 3 3 3 1 2 3 1 5 ...
##  $ pew_religimp: int   4 4 4 1 2 4 4 2 4 1 ...
##  $ newsint     : int   1 1 1 1 1 1 1 1 1 1 1 ...
##  $ faminc_new  : int   10 11 10 9 10 2 13 8 7 11 ...
##  $ union       : int   1 2 3 3 1 3 2 2 2 3 ...
##  $ investor    : int   2 1 1 1 1 2 1 2 1 2 ...
##  $ CC18_308a   : int   4 4 4 1 2 3 4 4 4 1 ...
##  $ CC18_310a   : int   2 3 3 3 2 2 3 2 3 2 ...
##  $ CC18_310b   : int   2 3 3 3 2 3 3 5 3 3 ...
##  $ CC18_310c   : int   3 3 3 3 3 3 2 5 3 3 ...
##  $ CC18_310d   : int   2 5 3 3 3 2 2 2 3 3 ...
##  $ CC18_325a   : int   1 2 1 1 1 2 2 2 2 1 ...
##  $ CC18_325b   : int   2 2 1 1 1 1 2 1 2 2 ...
##  $ CC18_325c   : int   1 2 2 1 1 2 2 1 2 2 ...
##  $ CC18_325d   : int   1 1 1 1 1 1 1 1 2 1 ...
```

Problem 6

Filter the data the same as in Problem 5. Use a R function to create a frequency table for the responses for the question asking whether these survey respondents are invested in the stock market.

Put your code and solution here:

```
table(filter(dat, gender == 1 & region <3 & marstat == 1 & newsint == 1)$investor)
```

```
##
##  1  2
## 57 18
```

Problem 7

Going back to using all rows in the dataset, create a new column in the data using mutate that is equal to either 0, 1, or 2, to reflect whether the respondent supports increasing the standard deduction from 12,000 to 25,000, supports cutting the corporate income tax rate from 39 to 21 percent, or both (so, support for neither policy equals 0, one of the two policies equals 1, and both policies equals two). Name the column "tax_scale". Hint: you'll need to use recode() as well.

Display the first twenty elements of the new column you create.

Put your code and solution here:

```r
recode_325a <- recode(dat$CC18_325a, '1' = 1, '2' = 0)
recode_325d <- recode(dat$CC18_325d, '1' = 1 , '2' = 0)
tax_scale <- recode_325a + recode_325d
dat <- mutate(dat,tax_scale)
print(dat$tax_scale[1:20])
```

```
## [1] 2 1 1 1 2 2 2 1 0 1 0 2 0 1 0 1 1 1 2 1
```

Problem 8

Use a frequency table command to show how many 0s, 1s, and 2s are in the column you created in Problem 7.

Put your code and solution here:

```r
table(dat$tax_scale)
```

```
##
##   0   1   2
## 130 408 331
```

Problem 9

Again using all rows in the original dataset, use summarise and group_by to calculate the average (mean) job of approval for President Trump in each of the four regions listed in the "region" column.

Put your code and solution here:

```r
dat %>% group_by(region) %>% summarise(Trump_Approve_Mean = mean(CC18_308a))
```

```
## # A tibble: 4 x 2
##    region Trump_Approve_Mean
##     <int>              <dbl>
## ## 1     1               2.77
## ## 2     2               2.76
## ## 3     3               2.71
## ## 4     4               3.03
```

Problem 10

Again start with all rows in the original dataset, use summarise() to create a summary table for survey respondents who are not investors and who have an annual family income of between $40,000 and $119,999 per year. The table should have the mean, median and standard deviations for the importance of religion column.

Put your code and solution here:

```
q10 <- filter(dat , faminc_new == 5 | faminc_new == 6 | faminc_new == 7 | faminc_new == 8 | faminc_new =

q10 %>% summarise(Mean_imp_religion = mean(q10$pew_religimp), Median_imp_religion = median(q10$pew_relig
```

```
##   Mean_imp_religion Median_imp_religion Standard_Dev_Religion_Imp
## 1          2.325893                   2                  1.188906
```

Problem 11

Use kable() and the the summarise() function to create a table with one row and three columns that provides the mean, median, and standard deviation for the column named faminc_new in the survey data.

Put your code and solution here:

```
kable(dat %>% summarise(Mean_faminc = mean(faminc_new), Median_faminc = median(faminc_new), sd_faminc =
```

Mean_faminc	Median_faminc	sd_faminc
6.581128	6	3.247035

Problem 12

With the survey data, use qplot() to make a histogram of the column named pid7. Change the x-axis label to "Seven Point Party ID" and the y-axis label to "Count".

Note: you can ignore the "stat_bin()" message that R generates when you draw this. The setup for the code chunk will suppress the message.

Put your code and solution here:

```
qplot(dat$pid7, xlab = "Seven Point Party ID" , ylab = "Count")
```