Homework 3

Bree, Cass, and Sergio

2025-10-23

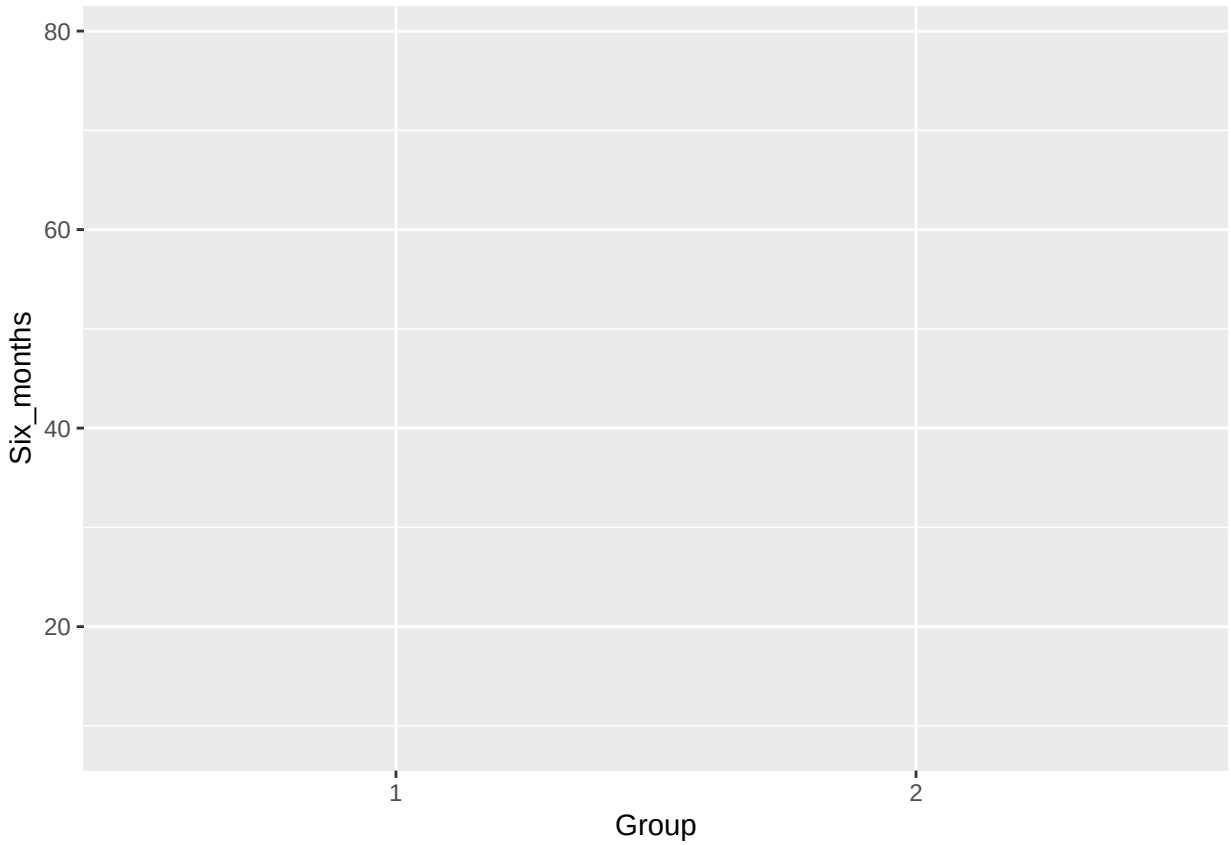
```
#Create stratified bar charts of text messages by Time

setwd("/cloud/project")

#installing the package needed to read an .xlsx file
#install.packages("readxl")

#Loading the packages
library(readxl)

TM <- read_excel("TextMessages.xlsx")
is.factor (TM$Group)
```

```
## [1] FALSE
```

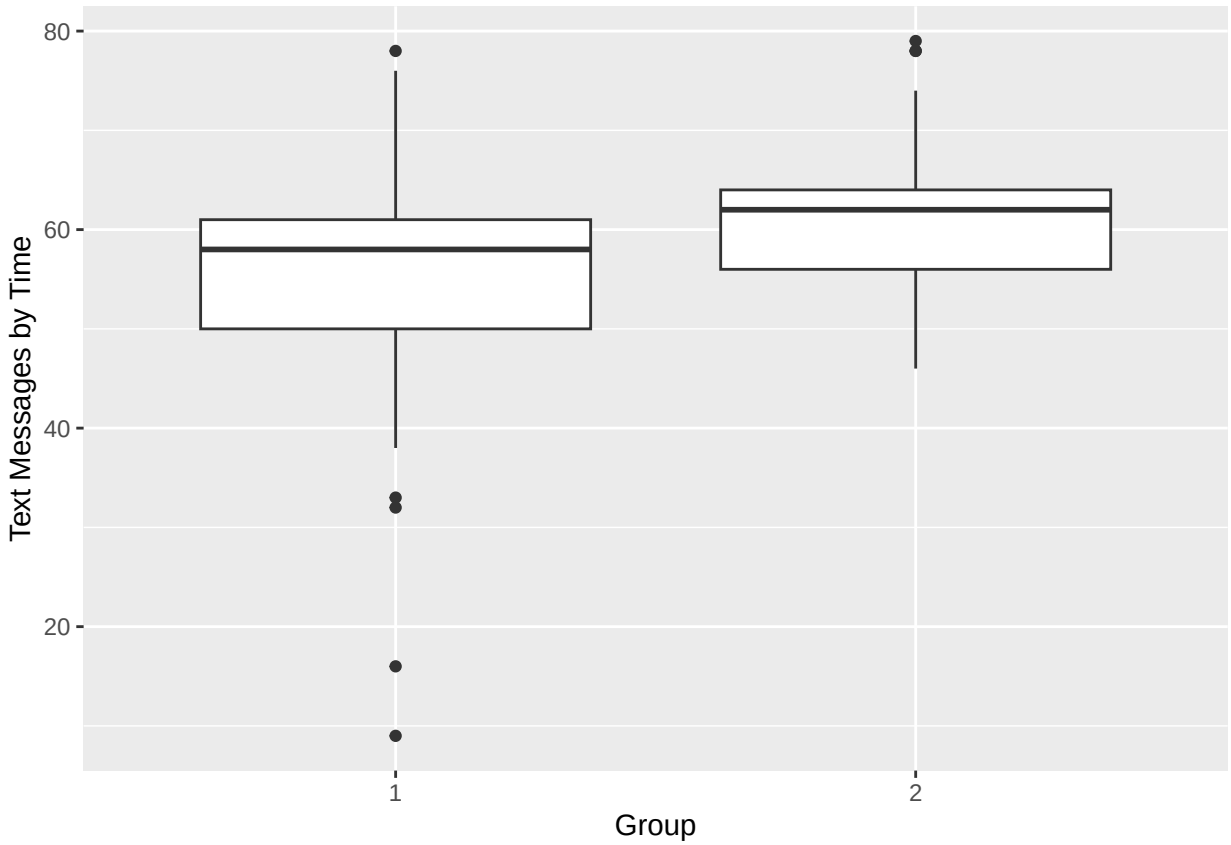
```
TM$Group <- as.factor(TM$Group)
is.factor (TM$Group)
```

```
## [1] TRUE
```

```
#install.packages("ggplot2")
library(ggplot2)
boxplot_TMbyTime <- ggplot(TM, aes(Group, Six_months))
boxplot_TMbyTime
```




```
boxplot_TMbyTime + geom_boxplot() + labs(x="Group", y="Text Messages by Time")
```




```
#importing the data needed
TextMessages <-read_excel("TextMessages.xlsx")

#to view and print on console window the first 5 observations
head(TextMessages)
```

```
## # A tibble: 6 x 4
##   Group Baseline Six_months Participant
##   <dbl>    <dbl>      <dbl>       <dbl>
## 1     1       52         32           1
## 2     1       68         48           2
## 3     1       85         62           3
## 4     1       47         16           4
## 5     1       73         63           5
## 6     1       57         53           6
```

```
#to output the total number of rows of data
nrow(TextMessages)
```

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

```
#output on the console window displays 50, confirming that our data has 50 rows

#to view and print the variable names on the console window
names(TextMessages)
```

```
## [1] "Group"       "Baseline"    "Six_months"  "Participant"
```

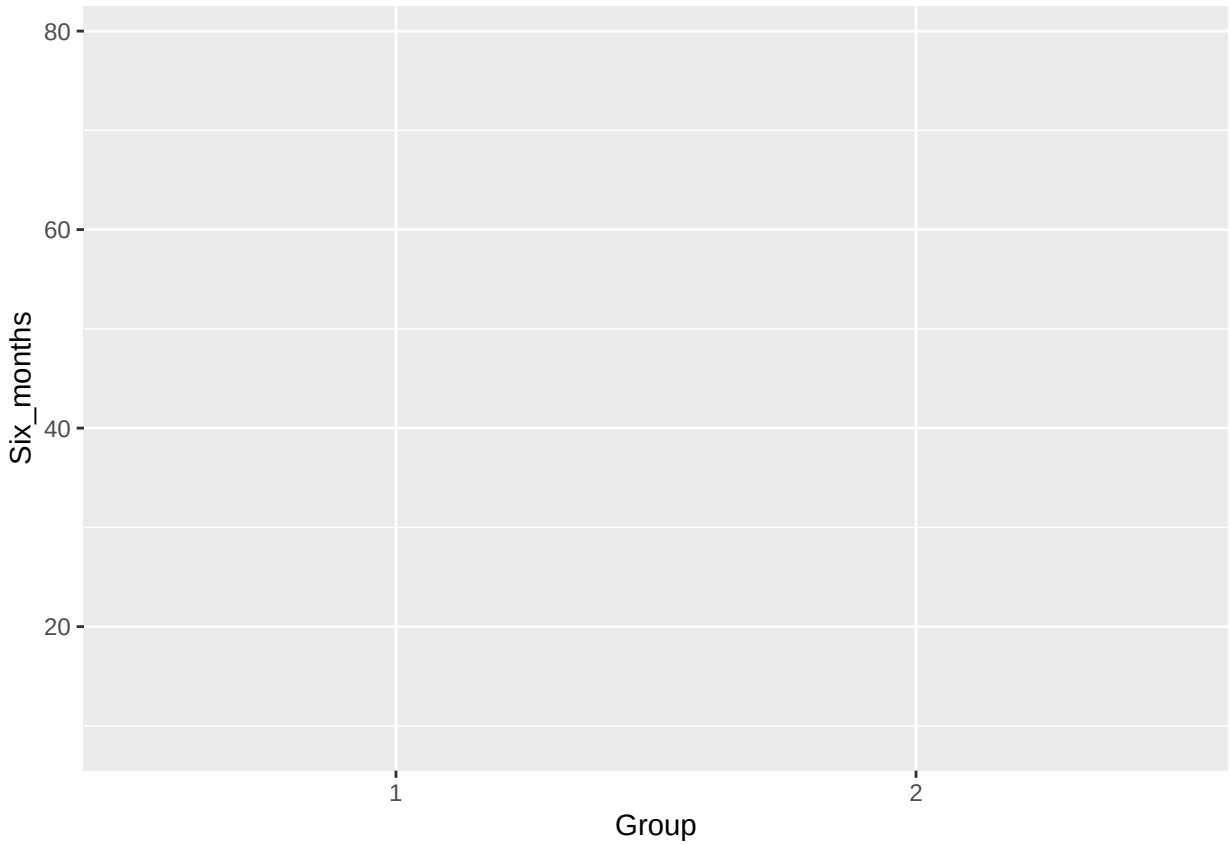
```r
#we confirm the 4 variable names from the data description:Group, Baseline, Six_months, and Participant

# PREPARING THE VARIABLES #

#Check if "Group" is a factor
is.factor(TextMessages$Group)
```

```
## [1] FALSE
```

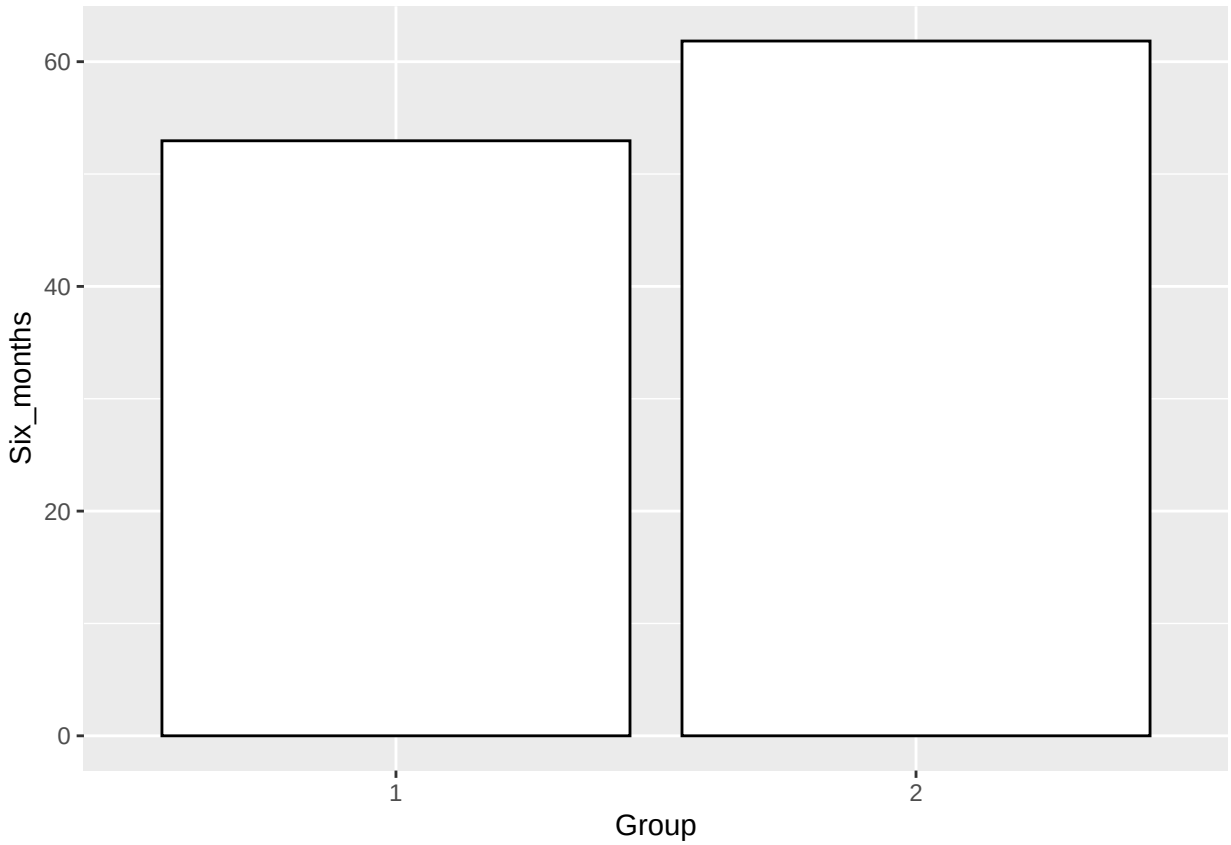
```r
#It is not a factor so we convert it into one
TextMessages$Group <- as.factor(TextMessages$Group)

#Confirm the conversation
is.factor(TextMessages$Group)
```

```
## [1] TRUE
```

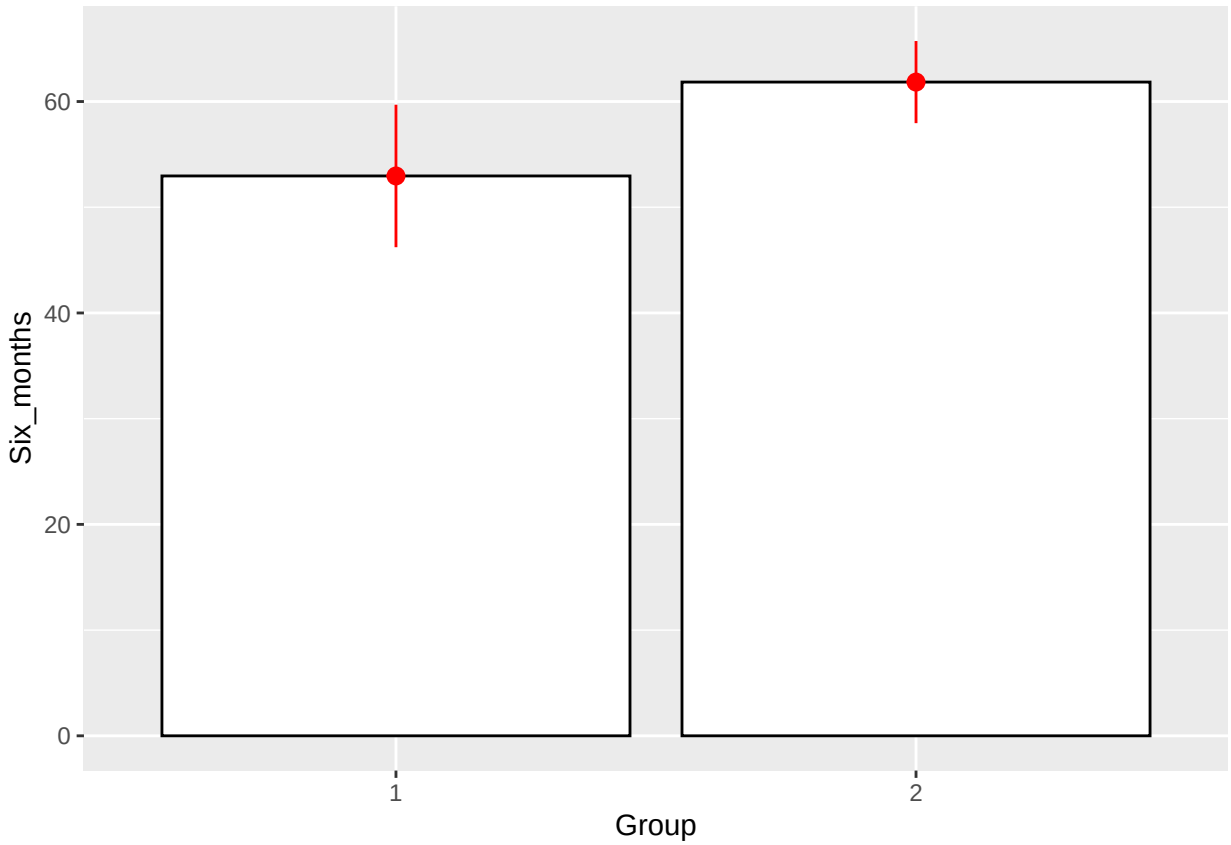
```r
#Install ggplot function
#install.packages("ggplot2")

#load the packages
library(ggplot2)


################
#ERROR BAR PLOT#
################

#Create a ggplot object using Group as x-axis and six_months as y-axis
barWithErrors_SixMonthsByGroup <- ggplot(TextMessages, aes(Group,Six_months))

#View the basic x-y axes setup
barWithErrors_SixMonthsByGroup
```



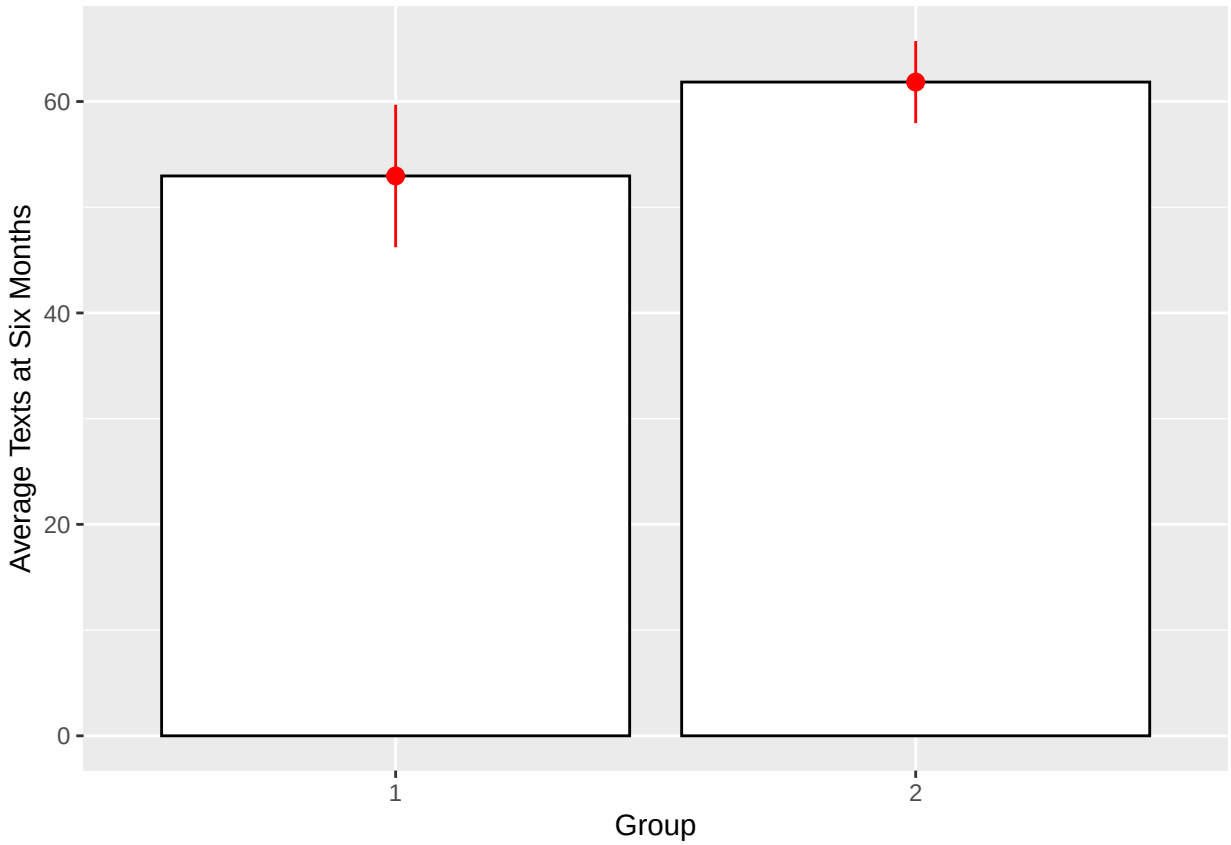

```
#Add the first layer showing bars representing the mean of each group
barWithErrors_SixMonthsByGroup +
  stat_summary(fun = mean, geom = "bar", fill = "White", colour = "Black")
```




```
#Install and load Hmisc (For the creation of error bars)
#install.packages("Hmisc")
library(Hmisc)

#To add error bars showing 95% confidence intervals for group means
barWithErrors_SixMonthsByGroup +
  stat_summary(fun = mean, geom = "bar", fill = "White", colour = "Black") +
  stat_summary(fun.data = mean_cl_normal, geom = "pointrange", colour = "Red")
```



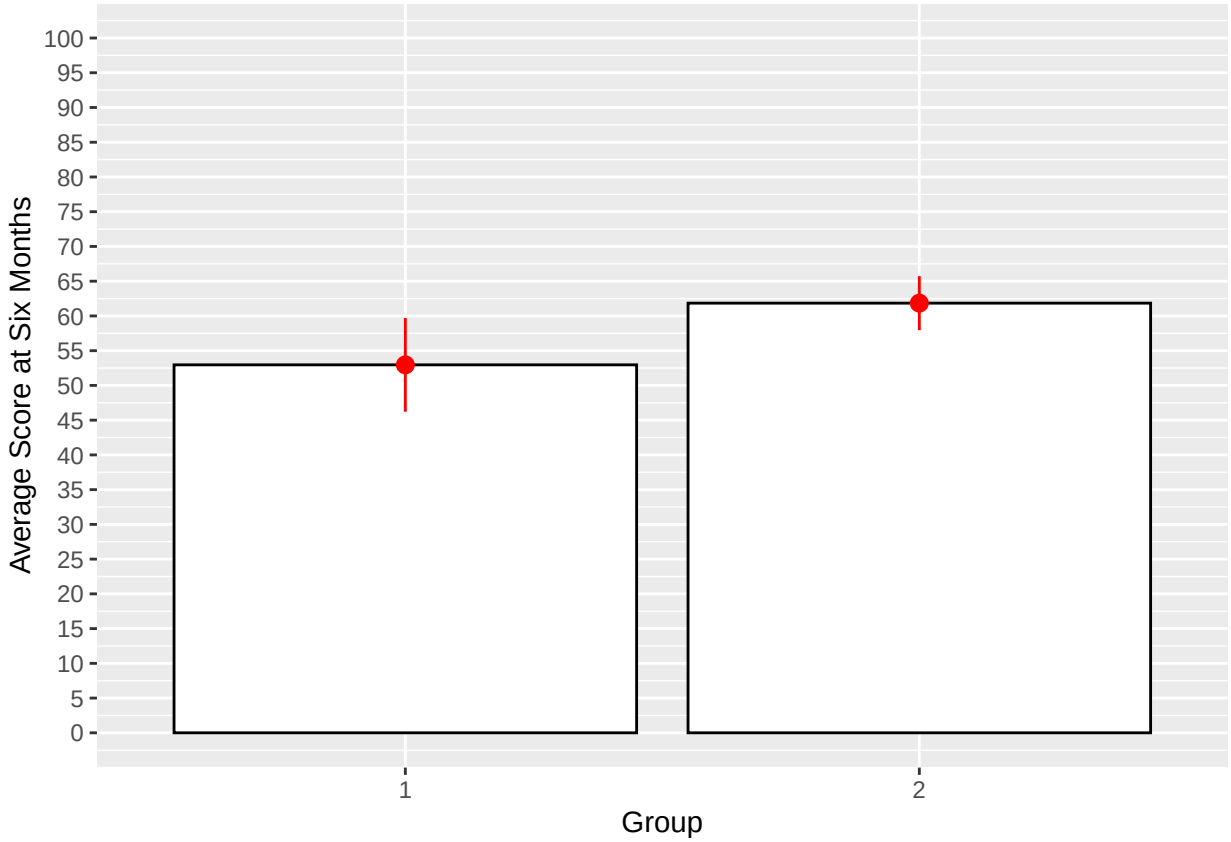

```
#Adding descriptive axis labels
barWithErrors_SixMonthsByGroup +
  stat_summary(fun = mean, geom = "bar", fill = "White", colour = "Black") +
  stat_summary(fun.data = mean_cl_normal, geom = "pointrange", colour = "Red") +
  labs(x = "Group", y = "Average Texts at Six Months")
```




```
#Check the mean values per group to help decice y-axis scale
by(TextMessages$Six_months, TextMessages$Group, mean)
```

```
## TextMessages$Group: 1
## [1] 52.96
## ------------------------------------------------------------
## TextMessages$Group: 2
## [1] 61.84
```

```
# Set custom y-axis limits and tick marks for better readability
barWithErrors_SixMonthsByGroup +
  stat_summary(fun = mean, geom = "bar", fill = "White", colour = "Black") +
  stat_summary(fun.data = mean_cl_normal, geom = "pointrange", colour = "Red") +
  labs(x = "Group", y = "Average Score at Six Months") +
  scale_y_continuous(limits = c(0, 100), breaks = seq(from = 0, to = 100, by = 5))
```




```
#This bar plot by shows the mean Six_months score by Grouo with red error bars
#Representing 95% confidence intervals. Non-overlapping error bars suggest the
#significant differences between groups.

#install.packages("readr")
library(readr)

TextMessages <- read_excel("TextMessages.xlsx")

#View 5 observations
head(TextMessages, n=5)
```

```
## # A tibble: 5 x 4
##   Group Baseline Six_months Participant
##   <dbl>    <dbl>      <dbl>       <dbl>
## 1     1       52         32           1
## 2     1       68         48           2
## 3     1       85         62           3
## 4     1       47         16           4
## 5     1       73         63           5
```

```
#View variable names
names(TextMessages)
```

```
## [1] "Group"       "Baseline"    "Six_months"  "Participant"
```

```
#get Descriptive stats

#install.packages("dplyr")
```

```r
library(dplyr)

#summarize messages by group for both time points
summary_stats <- TextMessages %>%
  group_by(Group) %>%
  summarise(
    Mean_Baseline = mean(Baseline, na.rm = TRUE),
    SD_Baseline = sd(Baseline, na.rm = TRUE),
    Median_Baseline = median(Baseline, na.rm = TRUE),
    Mean_SixMonths = mean(Six_months, na.rm = TRUE),
    SD_SixMonths = sd(Six_months, na.rm = TRUE),
    Median_SixMonths = median(Six_months, na.rm = TRUE),
    Count = n()
  )

#print summary stats
summary_stats
```

```
## # A tibble: 2 x 8
##   Group Mean_Baseline SD_Baseline Median_Baseline Mean_SixMonths SD_SixMonths
##   <dbl>         <dbl>       <dbl>           <dbl>          <dbl>        <dbl>
## 1     1          64.8        10.7              64           53.0        16.3
## 2     2          65.6        10.8              65           61.8         9.41
## # i 2 more variables: Median_SixMonths <dbl>, Count <int>
```

```r
# Summary stats to Output File

write.csv(summary_stats, "SummaryStats_TextMessages.csv", row.names = FALSE)


# The summary table outputs summary stats for the text message data
#Standard deviation measures variability, with group 2 having a higher SD
```