FINALNOW3

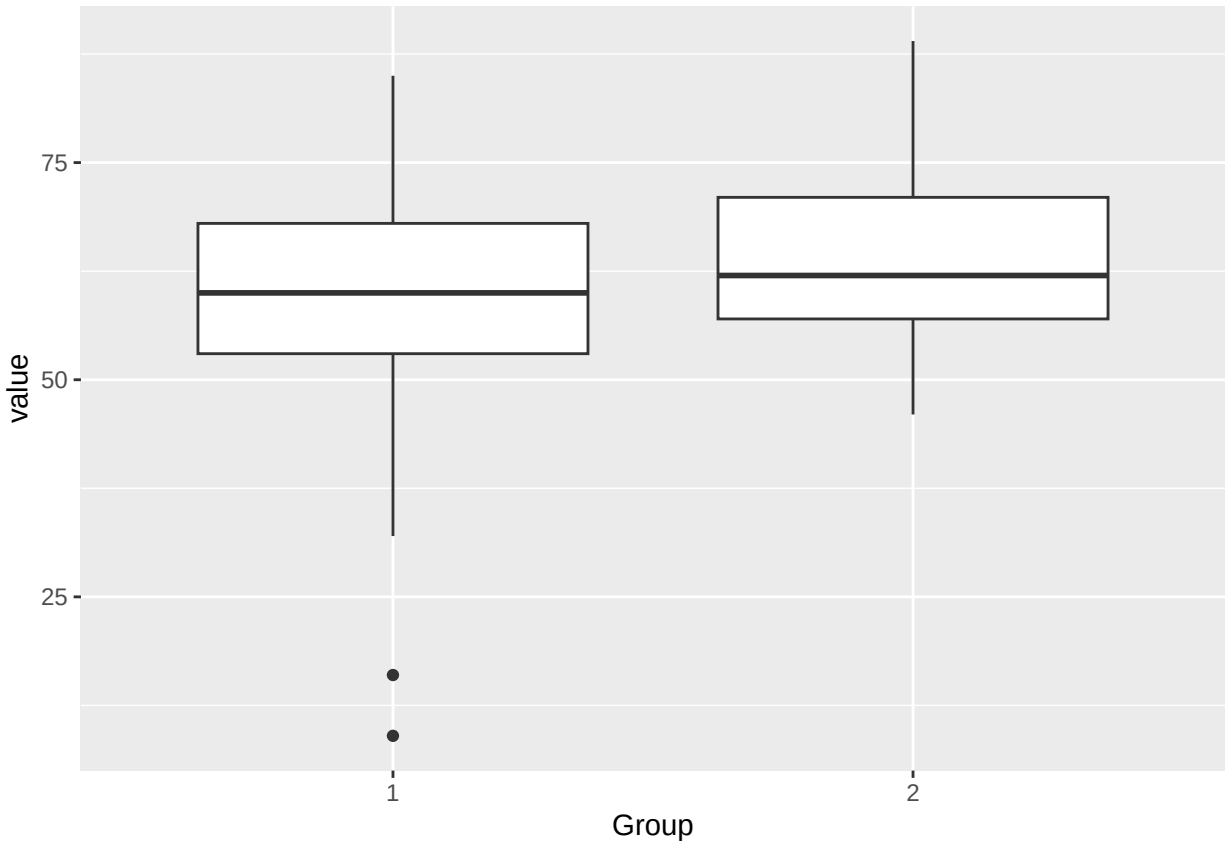
Lizbeth, Nicole, Anabel

2025-10-22

```r
setwd("/cloud/project")

TextMessages <- read.csv("TextMessages.csv", header=TRUE)

table(TextMessages$Group)
```

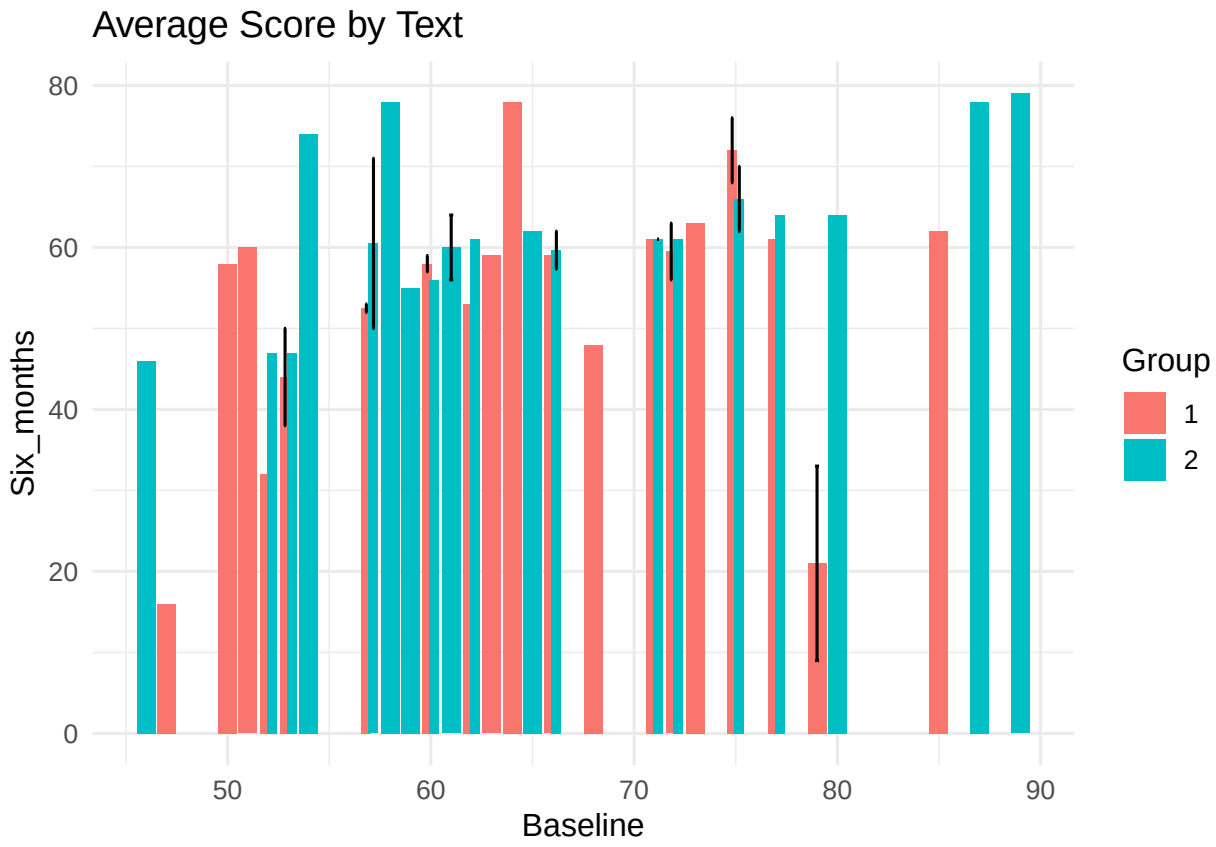
```
##
##  1  2
## 25 25
```

```r
install.packages("reshape")

library(reshape)

text <- melt(TextMessages, id=c("Participant","Group"),
             measured=c("Baseline","Six_months"))

text$Group <- as.factor(text$Group)
install.packages("ggplot2")

library(ggplot2)

ggplot(text,aes(x=Group,y=value)) + geom_boxplot()
```




```
#Visually, the graph for participants in group 1 shows that the median is about
#65. Also, we can see that there are two outliers below the minimum value. For
#group 2, the median seems to be about 68. Meanwhile, there are no outliers.
#Group 1 is left skewed, this means that participants fall mostly towards the
#values below the median. Thus, for group 2, it is right skewed, so the value
#for the paricipant mostly goes above the median.

#####VISUAL1############
#barchart

# Install if not already installed
install.packages("tidyverse")

library(tidyverse)

text<- read.csv("TextMessages.csv", header=TRUE)
install.packages("reshape")

library(reshape)

# Reshape and assign to 'text'

text_long <- pivot_longer(data=text, cols = c(Baseline, Six_months),names_to = "time",
                          values_to = "Score" )

# Mean score by time & group
ggplot(text, aes(x = Baseline, y = Six_months, fill = as.factor(Group))) +
```

```
  stat_summary(fun = "mean", geom = "bar", position = position_dodge(width = 0.7)) +
  stat_summary(fun.data = mean_se, geom = "errorbar",
               position = position_dodge(width = 0.7), width = 0.2) +
  labs(title = "Average Score by Text ",
       x = "Baseline",
       y = "Six_months",
       fill = "Group") +
  theme_minimal(12)
```




```
######VISUAL2#########



#########SUMMARY(Descriptive Statistics with Raw Data)#########
install.packages("pastecs")
library(pastecs)

#In this process, the Original data in "TextMessages" was used to obtain summary
#statistics on time (Baseline) by the Group
by(TextMessages$Baseline, TextMessages$Group,stat.desc)
```

```
## TextMessages$Group: 1
##      nbr.val     nbr.null       nbr.na          min          max        range
##    25.000000     0.000000     0.000000    47.000000    85.000000    38.000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
##  1621.000000    64.000000    64.840000     2.135946     4.408377   114.056667
##      std.dev     coef.var
```

```
##    10.679732     0.164709
## ------------------------------------------------------------
## TextMessages$Group: 2
##      nbr.val     nbr.null       nbr.na          min          max        range
##   25.0000000    0.0000000    0.0000000   46.0000000   89.0000000   43.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 1640.0000000   65.0000000   65.6000000    2.1671794    4.4728385  117.4166667
##      std.dev     coef.var
##   10.8358971    0.1651814
```

```
#When looking at the data, the number of text messages ranges from 47 (min)-85
#(max)for Group 1 and 46(min)-89(max) for Group 2. This corresponds to the mean
#being 65.60 for Group 2 which is 0.76 greater than Group 1, at 64.80. Thus, the
#standard deviation for Group 1 and Group 2 are about 10, this means that it is
#about 15% of mean (10/65), so there is moderate variability on the spread of
#the data.

#In this process, the Original Data in "TextMessages" was used to obtain summary
#statistics on time (Six_Months) by the Group
by(TextMessages$Six_months, TextMessages$Group, stat.desc)
```

```
## TextMessages$Group: 1
##      nbr.val     nbr.null       nbr.na          min          max        range
##   25.0000000    0.0000000    0.0000000    9.0000000   78.0000000   69.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 1324.0000000   58.0000000   52.9600000    3.2662313    6.7411700  266.7066667
##      std.dev     coef.var
##   16.3311563    0.3083678
## ------------------------------------------------------------
## TextMessages$Group: 2
##      nbr.val     nbr.null       nbr.na          min          max        range
##   25.0000000    0.0000000    0.0000000   46.0000000   79.0000000   33.0000000
##          sum       median         mean      SE.mean CI.mean.0.95          var
## 1546.0000000   62.0000000   61.8400000    1.8820910    3.8844450   88.5566667
##      std.dev     coef.var
##    9.4104552    0.1521742
```

```
#For the data in Six_Months, the number of text messages ranges from 9(min)-78
#(max) for Group 1 and 46(min)-79 (max) for Group 2. The mean in this data set
#is higher in Group  2 (61.84) in comparison the number of Text messages in
#Group 1, at 52.96. The difference is 8.88. Hence, the standard deviation for
#Group 1 is about 16, this means that the data for text messages
#is about 30% (16/52.96) of the mean. 30% is a high standard deviation, so it
#means there is a great variance of text for Six_Months. Furthermore, in Group
#2, the standard deviation is about 9, this means that the data for text
#messages is about 15% (9/61.84) of the mean. The standard deviation is average
#for Group 2, so moderate variability of data for the data set.


#Overall, the Baseline for Text messages in Group 1 and Group 2, had text
#messages that mostly did fall between the number of 65 (mean). Most participants fell
#in this range. For the Six_Months, there was a great number of variance for the
#data in Group 1. Thus, for Group 2, there was moderate variance because they
#fell in the range of the mean at 61.
```