HW3_379

Vladyslav Faychuk

2025-10-23

```
read.csv("TextMessages.csv")
```

```
##    Group Baseline Six_months Participant
## 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
## 7      1       63         59           7
## 8      1       50         58           8
## 9      1       66         59           9
## 10     1       60         57          10
## 11     1       51         60          11
## 12     1       72         56          12
## 13     1       77         61          13
## 14     1       57         52          14
## 15     1       79          9          15
## 16     1       75         76          16
## 17     1       53         38          17
## 18     1       72         63          18
## 19     1       62         53          19
## 20     1       71         61          20
## 21     1       53         50          21
## 22     1       64         78          22
## 23     1       79         33          23
## 24     1       75         68          24
## 25     1       60         59          25
## 26     2       65         62          26
## 27     2       57         50          27
## 28     2       66         62          28
## 29     2       71         61          29
## 30     2       75         70          30
## 31     2       61         64          31
## 32     2       80         64          32
## 33     2       66         55          33
## 34     2       53         47          34
## 35     2       62         61          35
## 36     2       61         56          36
## 37     2       77         64          37
## 38     2       66         62          38
```

```
## 39     2       52         47          39
## 40     2       60         56          40
## 41     2       58         78          41
## 42     2       54         74          42
## 43     2       72         61          43
## 44     2       71         61          44
## 45     2       87         78          45
## 46     2       75         62          46
## 47     2       57         71          47
## 48     2       59         55          48
## 49     2       46         46          49
## 50     2       89         79          50
```

```
TMData <- read.csv("TextMessages.csv")

is.factor(TMData$Group)
```

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

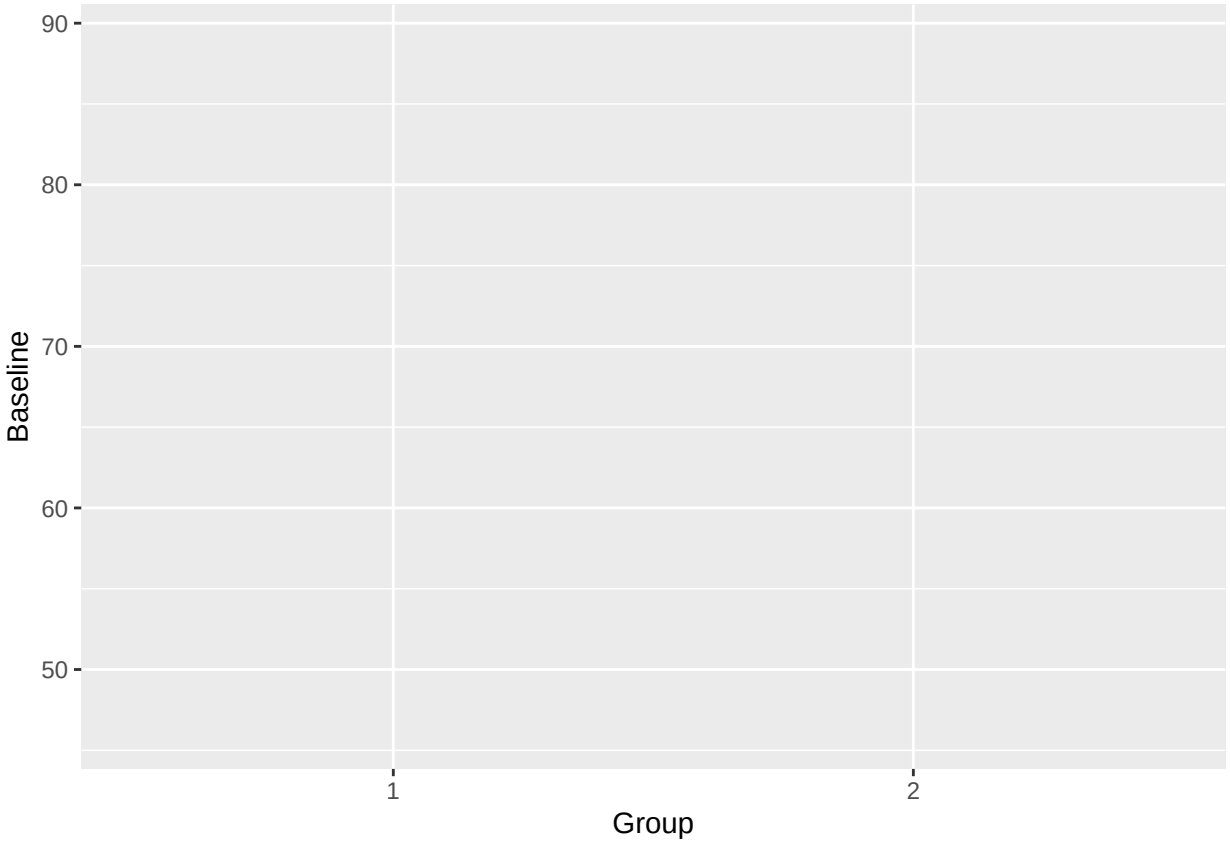
```
is.factor(TMData$Participant)
```

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

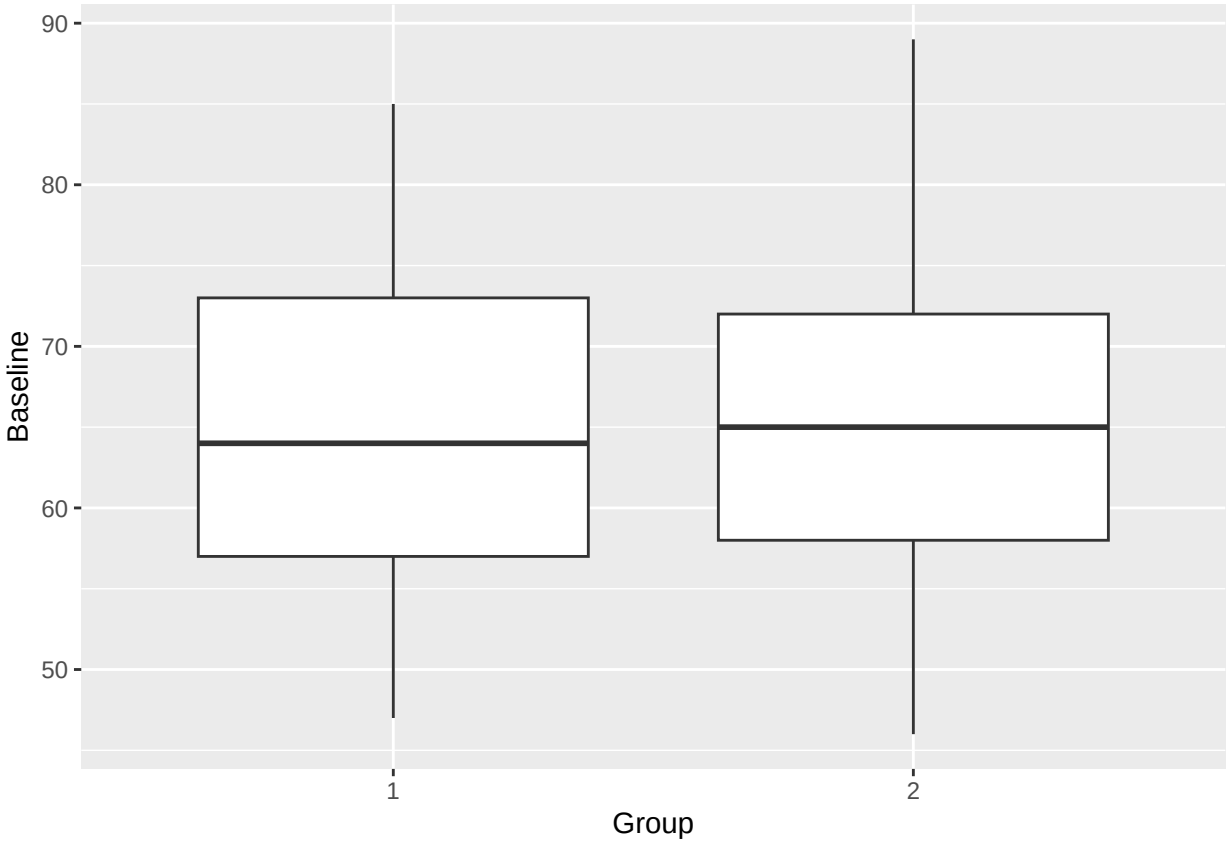
```
TMData$Group <- as.factor(TMData$Group)

TMData$Participant <- as.factor(TMData$Participant)

#install.packages("ggplot2")

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.1
```

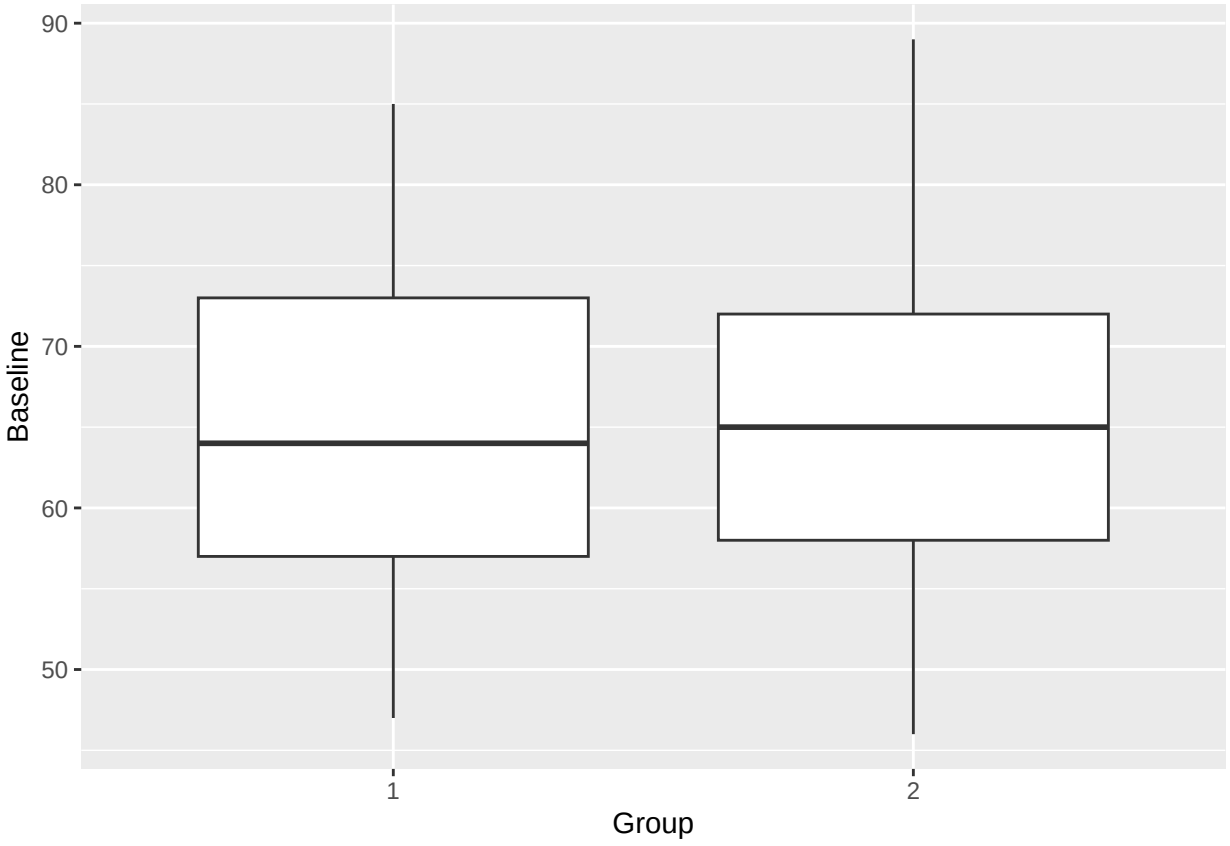
```
boxplot_BaselinebyGroup <- ggplot(TMData, aes(Group, Baseline))

boxplot_BaselinebyGroup
```




```
boxplot_BaselinebyGroup + geom_boxplot()
```




```
boxplot_BaselinebyGroup + geom_boxplot() + labs(x="Group", y="Baseline")
```



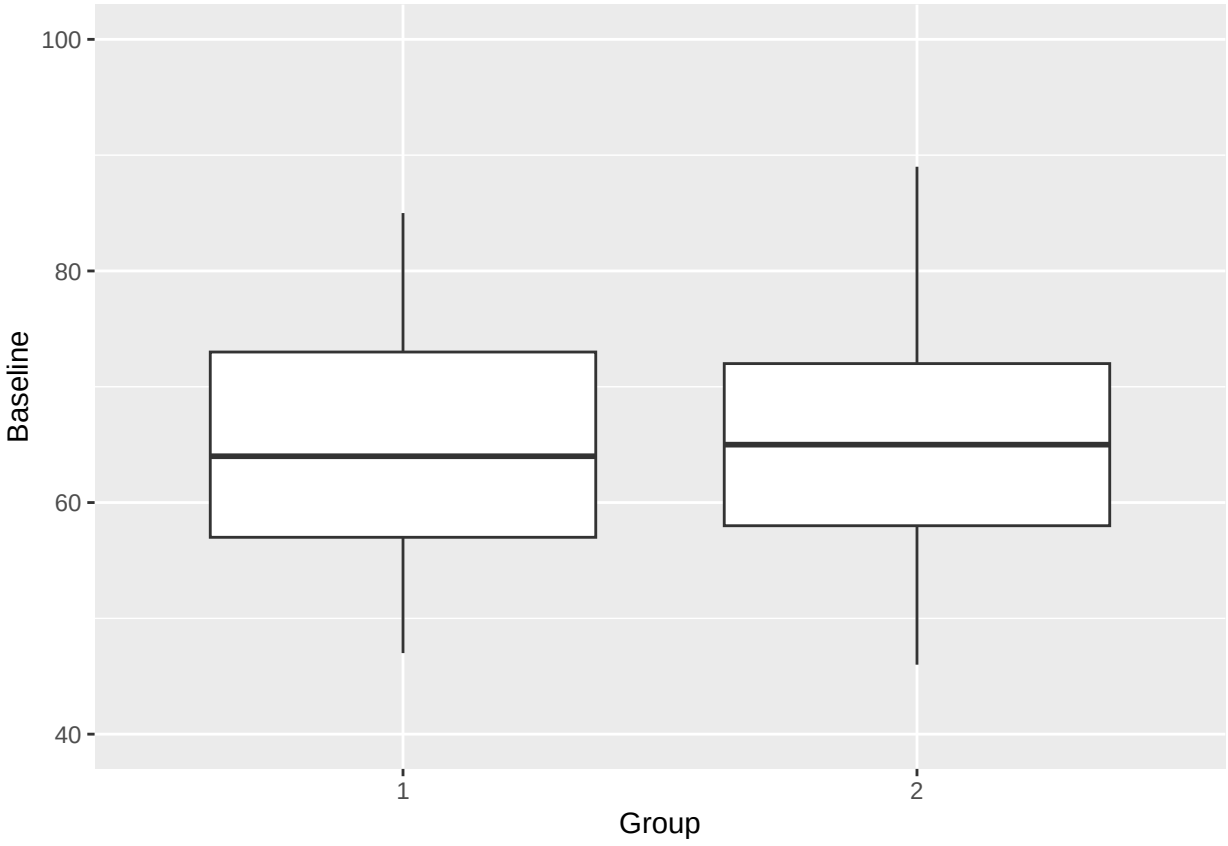

```
min(TMData$Baseline)
```

```
## [1] 46
```

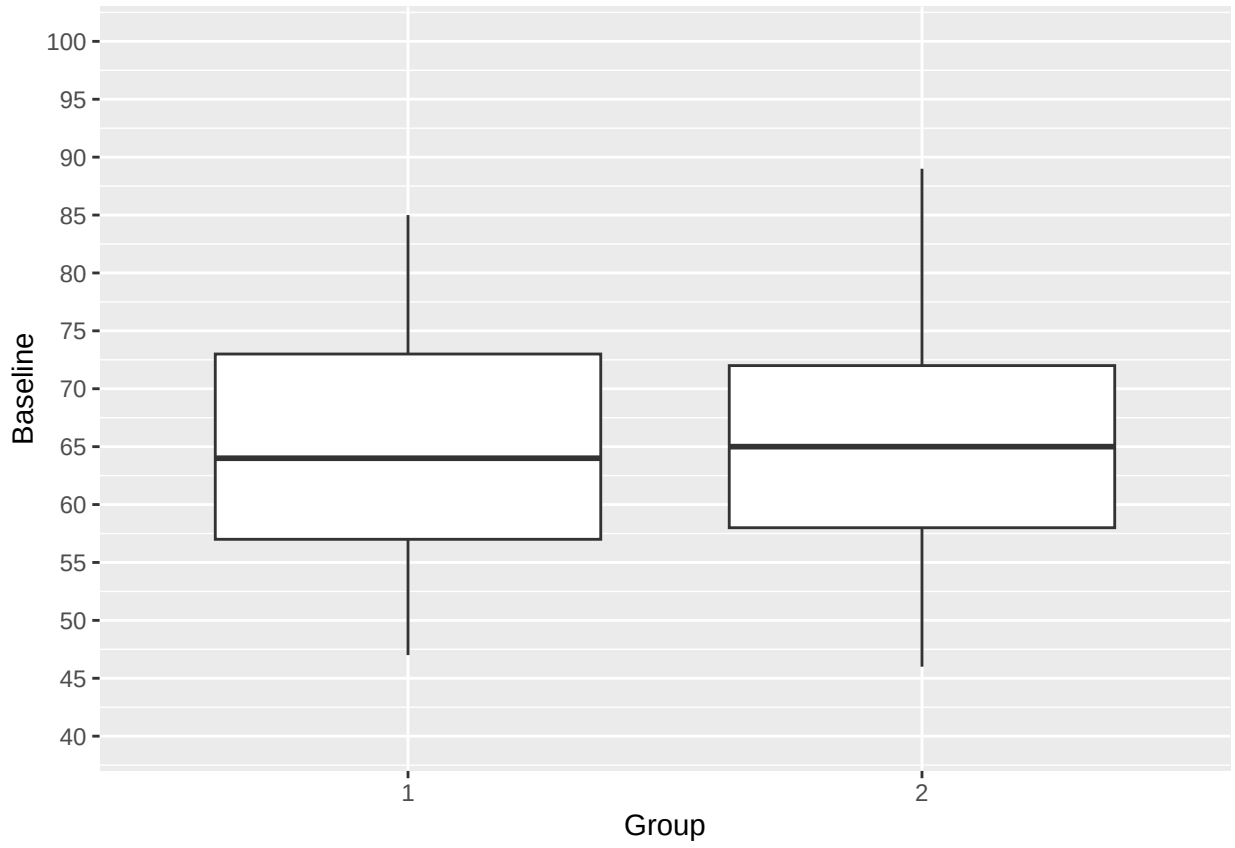
```
max(TMData$Baseline)
```

```
## [1] 89
```

```
boxplot_BaselinebyGroup + geom_boxplot() + labs(x="Group", y="Baseline") +
  scale_y_continuous(limits = c(40, 100))
```




```
boxplot_BaselinebyGroup + geom_boxplot() + labs(x="Group", y="Baseline") +
  scale_y_continuous(limits = c(40, 100), breaks = seq(from =40, to = 100, by = 5))
```




```
library(tidyverse)
```

```
## Warning: package 'tidyr' was built under R version 4.4.1
```

```
## Warning: package 'purrr' was built under R version 4.4.1
```

```
## Warning: package 'stringr' was built under R version 4.4.1
```

```
## Warning: package 'lubridate' was built under R version 4.4.1
```

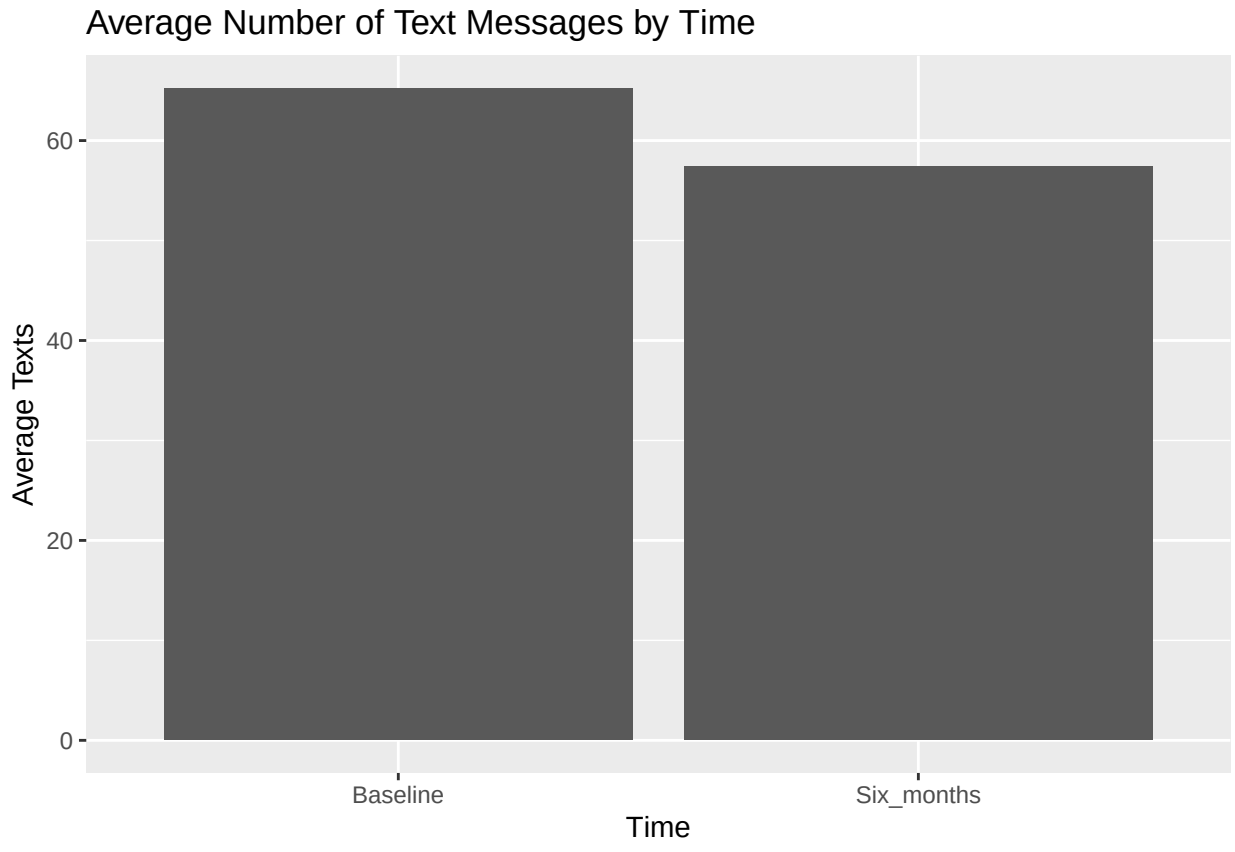
```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.2
## v lubridate 1.9.4     v tibble    3.2.1
## v purrr     1.0.4     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
textM <- read.csv("TextMessages.csv")
#Create stratified bar charts of text messages by Time
#Transform to long data to combine times (Baseline and six_months)
long_data <- textM %>% pivot_longer(cols = c(Baseline, Six_months),
```

```
               names_to = "Time",
               values_to = "Texts")

head(long_data)
```

```
## # A tibble: 6 x 4
##   Group Participant Time       Texts
##   <int>       <int> <chr>      <int>
## 1     1           1 Baseline      52
## 2     1           1 Six_months    32
## 3     1           2 Baseline      68
## 4     1           2 Six_months    48
## 5     1           3 Baseline      85
## 6     1           3 Six_months    62
```

```
# Create stratified bar chart
ggplot(long_data, aes(x = Time, y = Texts)) +
  stat_summary(fun = mean, geom = "bar", position = "dodge") +
  labs(
    title = "Average Number of Text Messages by Time",
    x = "Time",
    y = "Average Texts",
  )
```



```r
#Using Vlads long_data
textM <- read.csv("TextMessages.csv")
long_data <- textM %>% pivot_longer(cols = c(Baseline, Six_months),
                                    names_to = "Time",
                                    values_to = "Texts")

#Creating summary Tables

# Summary table by Group
group_summary <- long_data |>
  group_by(Group) |>
  summarise(
    mean_texts = mean(Texts, na.rm = TRUE),
    median_texts = median(Texts, na.rm = TRUE),
    sd_texts = sd(Texts, na.rm = TRUE),
    min_texts = min(Texts, na.rm = TRUE),
    max_texts = max(Texts, na.rm = TRUE),
    n_participants = n_distinct(Participant),
    n_rows = n(),
    .groups = "drop"
  )

print(group_summary)
```

```
## # A tibble: 2 x 8
##   Group mean_texts median_texts sd_texts min_texts max_texts n_participants
##   <int>      <dbl>        <dbl>    <dbl>     <int>     <int>          <int>
## 1     1       58.9           60     14.9         9        85             25
## 2     2       63.7           62     10.2        46        89             25
## # i 1 more variable: n_rows <int>
```

```r
#summary table by Time
time_summary <- long_data |>
  group_by(Time) |>
  summarise(
    mean_texts = mean(Texts, na.rm = TRUE),
    sd_texts = sd(Texts, na.rm = TRUE),
    n = n(),
    se_texts = sd_texts / sqrt(n),
    .groups = "drop"
  )

print(time_summary)
```

```
## # A tibble: 2 x 5
##   Time       mean_texts sd_texts     n se_texts
##   <chr>           <dbl>    <dbl> <int>    <dbl>
## 1 Baseline         65.2     10.7    50     1.51
## 2 Six_months       57.4     13.9    50     1.97
```

```r
# 3) By Group and Time
group_time_summary <- long_data |>
```

```r
  group_by(Group, Time) |>
  summarise(
    mean_texts = mean(Texts, na.rm = TRUE),
    sd_texts = sd(Texts, na.rm = TRUE),
    n = n(),
    se_texts = sd_texts / sqrt(n),
    .groups = "drop"
  )

print(group_time_summary)
```

```
## # A tibble: 4 x 6
##   Group Time       mean_texts sd_texts     n se_texts
##   <int> <chr>           <dbl>    <dbl> <int>    <dbl>
## 1     1 Baseline         64.8    10.7     25     2.14
## 2     1 Six_months       53.0    16.3     25     3.27
## 3     2 Baseline         65.6    10.8     25     2.17
## 4     2 Six_months       61.8     9.41    25     1.88
```