

# Github Collaboration

Tingwei Adeck, Kaity Trinidad, Noor Aayla

March 21, 2025

## Import Data

```
library(readr)
TextMessages <- read_csv("TextMessages.csv") #1-use read_csv (parent df)
#View(TextMessages)
#head(TextMessages)

kable(head(TextMessages,5), format = "markdown", #
      caption = "Text Message Data Preview",
      table.envir = "table", align = "c")
```

Table 1: Text Message Data Preview

Group	Baseline	Six_months	Participant
1	52	32	1
1	68	48	2
1	85	62	3
1	47	16	4
1	73	63	5

## Factorization (Data Wrangling)

```
#factor a variable en place - risky
make_factor <- function(data, var){
  refactored <- as.factor(data[[var]]) #make factor
  data[[var]] <- refactored #assign factor back to var
  data #return
}

#factor with labels scalably
#### to unpack the vectors and setnames to map values to labels
make_factor_scale <- function(data, var, newvarName, fact_num_vect,
                              fact_char_vect) {
  data[[newvarName]] <- dplyr::recode(data[[var]],
                                     !!!setNames(fact_char_vect,
                                                  fact_num_vect))

  data
}

#check the groups
unique(TextMessages$Group)

## [1] 1 2

TextMessages <- make_factor_scale(TextMessages,
                                  "Group",
                                  "Group_factor", c(1,2),
                                  c("Group 1",
                                    "Group 2"))

TextMessages <- TextMessages %>%
  dplyr::relocate("Group_factor", .after = "Group")
```

## Summary Table Setup

```
Value_bs <- stat.desc(TextMessages$Baseline)
Value_sm <- stat.desc(TextMessages$Six_months)
em_summ_df_bs <- as.data.frame(Value_bs)
em_summ_df_sm <- as.data.frame(Value_sm)
row_names_bs <- rownames(em_summ_df_bs)
row_names_sm <- rownames(em_summ_df_sm)

em_summ_df_bs <- em_summ_df_bs %>%
  tibble::rownames_to_column(var = "Statistic")

em_summ_df_sm <- em_summ_df_sm %>%
  tibble::rownames_to_column(var = "Statistic")

# Pivot to wide format
wide_data_bs <- em_summ_df_bs %>%
  pivot_wider(names_from = "Statistic", values_from = colnames(em_summ_df_bs)[2])

wide_data <- em_summ_df_sm %>%
  pivot_wider(names_from = "Statistic", values_from = colnames(em_summ_df_sm)[2])

kable(em_summ_df_bs, format = "markdown",
      caption = "Statistical Summary Long Table for Baseline",
      table.envir = "table", align = "c")
```

Table 2: Statistical Summary Long Table for Baseline

Statistic	Value_bs
nbr.val	50.0000000
nbr.null	0.0000000
nbr.na	0.0000000
min	46.0000000
max	89.0000000
range	43.0000000
sum	3261.0000000
median	64.5000000
mean	65.2200000
SE.mean	1.5067982
CI.mean.0.95	3.0280244
var	113.5220408
std.dev	10.6546723
coef.var	0.1633651

```
kable(em_summ_df_sm, format = "markdown",
      caption = "Statistical Summary Long Table for Six Months",
      table.envir = "table", align = "c")
```

Table 3: Statistical Summary Long Table for Six Months

Statistic	Value_sm
nbr.val	50.0000000
nbr.null	0.0000000
nbr.na	0.0000000

Statistic	Value_sm
min	9.0000000
max	79.0000000
range	70.0000000
sum	2870.0000000
median	60.5000000
mean	57.4000000
SE.mean	1.9703931
CI.mean.0.95	3.9596532
var	194.1224490
std.dev	13.9327832
coef.var	0.2427314

## Breakdown

At the Baseline (Initial Time Point), participants in Group 1 sent an average of 64.8 text messages, with a standard deviation of 10.7, indicating moderate variability in texting behavior. The minimum number of messages sent was 47, while the maximum was 85. Similarly, participants in Group 2 had a slightly higher mean of 65.6 messages, with a standard deviation of 10.8, suggesting a similar spread of values. The minimum number of messages in this group was 46, and the maximum was 89.

After six months, both groups experienced a decline in the number of messages sent. Group 1 saw a more substantial reduction, with a new mean of 53.0 messages (a decrease of approximately 11.8 messages from baseline). Additionally, the standard deviation increased to 16.3, suggesting that individuals in this group displayed more varied texting behavior over time. The minimum and maximum values were not explicitly provided but are likely within the range of 9 to 78 messages. In contrast, Group 2 maintained a relatively higher number of messages, with a mean of 61.8 messages (a smaller decrease of about 3.8 messages from baseline). The standard deviation was 9.41, indicating that texting behavior remained more consistent within this group. The minimum and maximum values were at least 46 and 79 messages, respectively.

## Overall Interpretation

Both groups showed a decline in the number of text messages sent over time. However, Group 1 experienced a greater decrease, suggesting that participants in this group reduced their texting behavior more substantially. Additionally, the increase in standard deviation for Group 1 at six months indicates greater variability in how much texting behavior changed within this group—some participants may have significantly reduced their texting, while others maintained a more consistent level. On the other hand, Group 2 exhibited more stable texting behavior, with a smaller overall decline and less variability over time.

## References

ggplot2 facet : split a plot into a matrix of panels