02: Descriptive Analysis

2026-01-30

Contents

Setup

Load Libraries

```
library(ggplot2)
library(patchwork)
library(tidyr)
library(dplyr)
```

Set File Paths

```
root <- "/Users/pei-chin/Research/Behavioral Science and Marketing_data_task"

data_clean <- file.path(root, "data_clean")
output     <- file.path(root, "output")
figures    <- file.path(output, "figures")
```

Load Clean Data

```
load(file.path(data_clean, "clean_data.RData"))
```

Variable Check

Define Feeling Variables

```
feeling_vars <- c(
  "feelings_youalone",
  "feelings_bothyoufirst",
  "feelings_themalone",
  "feelings_boththemfirst",
  "feelings_neither",
  "feelings_youaloneforgiven"
)
```

Summary Statistics for Feeling Variables

```
summary(clean_data[, feeling_vars])
```

```
##  feelings_youalone feelings_bothyoufirst feelings_themalone
##  Min.   :-30.00    Min.   :-30.000       Min.   :-30.000
##  1st Qu.:-30.00    1st Qu.:  0.000       1st Qu.:-20.000
##  Median :-20.00    Median : 10.000       Median :-10.000
##  Mean   :-18.36    Mean   :  7.844       Mean   : -5.533
##  3rd Qu.:-10.00    3rd Qu.: 20.000       3rd Qu.: 10.000
##  Max.   : 10.00    Max.   : 30.000       Max.   : 30.000
##  feelings_boththemfirst feelings_neither feelings_youaloneforgiven
##  Min.   :-20.00         Min.   :-30.00   Min.   :-30.00
##  1st Qu.: 10.00         1st Qu.:-23.00   1st Qu.:-30.00
##  Median : 20.00         Median :-15.00   Median :-19.00
##  Mean   : 17.42         Mean   :-14.38   Mean   :-13.91
##  3rd Qu.: 26.00         3rd Qu.: -8.00   3rd Qu.: -5.00
##  Max.   : 30.00         Max.   : 20.00   Max.   : 22.00
```

Check Sample Sizes

Check how many people are in the high self-blame group (blame_1 > 50 = high-self group).

```
feeling_complete_n <- sapply(
  feeling_vars,
  function(var) sum(!is.na(clean_data[[var]]))
)

print(feeling_complete_n)
```

```
##         feelings_youalone     feelings_bothyoufirst        feelings_themalone
##                        45                        45                        45
##    feelings_boththemfirst          feelings_neither feelings_youaloneforgiven
##                        45                        45                        45
```

```
table(clean_data$high_blame)
```

```
##
##  0  1
## 35 10
```

Feeling Ratings by Self-Blame Group

```
by(
  clean_data[, feeling_vars],
  clean_data$high_blame,
  summary
)
```

```
## clean_data$high_blame: 0
##  feelings_youalone feelings_bothyoufirst feelings_themalone
##  Min.   :-30.00    Min.   :-30.000       Min.   :-30.0
##  1st Qu.:-30.00    1st Qu.:  0.000       1st Qu.:-14.0
##  Median :-22.00    Median : 10.000       Median : -8.0
##  Mean   :-20.89    Mean   :  6.629       Mean   : -4.2
##  3rd Qu.:-12.00    3rd Qu.: 19.000       3rd Qu.:  9.0
##  Max.   :  8.00    Max.   : 27.000       Max.   : 30.0
##  feelings_boththemfirst feelings_neither feelings_youaloneforgiven
##  Min.   :-20.00         Min.   :-30.0    Min.   :-30.00
##  1st Qu.: 11.00         1st Qu.:-23.5    1st Qu.:-30.00
##  Median : 20.00         Median :-19.0    Median :-20.00
##  Mean   : 17.54         Mean   :-15.4    Mean   :-15.66
##  3rd Qu.: 25.50         3rd Qu.: -8.5    3rd Qu.: -5.00
##  Max.   : 30.00         Max.   :  4.0    Max.   : 22.00
## ------------------------------------------------------------
## clean_data$high_blame: 1
##  feelings_youalone feelings_bothyoufirst feelings_themalone
##  Min.   :-30.00    Min.   :-10.00        Min.   :-30.00
##  1st Qu.:-20.00    1st Qu.:  6.25        1st Qu.:-22.50
##  Median :-11.50    Median : 15.50        Median :-20.00
##  Mean   : -9.50    Mean   : 12.10        Mean   :-10.20
##  3rd Qu.:  4.75    3rd Qu.: 20.00        3rd Qu.:  6.25
##  Max.   : 10.00    Max.   : 30.00        Max.   : 20.00
##  feelings_boththemfirst feelings_neither feelings_youaloneforgiven
##  Min.   :-10.00         Min.   :-30.00   Min.   :-30.00
##  1st Qu.:  8.50         1st Qu.:-18.25   1st Qu.:-20.00
##  Median : 20.00         Median :-10.00   Median : -7.00
##  Mean   : 17.00         Mean   :-10.80   Mean   : -7.80
##  3rd Qu.: 27.75         3rd Qu.: -5.75   3rd Qu.:  2.25
##  Max.   : 30.00         Max.   : 20.00   Max.   : 20.00
```

Binary Choice Preferences

First Set of Binary Choice Questions

```
# Participants who prefer apologizing first
sum(clean_data$prefer_I_apologize_first)
```

```
## [1] 35
```

```
# Participants who prefer neither option
sum(clean_data$prefer_Neither_I_nor)
```

```
## [1] 10
```

Second Set of Binary Choice Questions

```
# Participants who prefer apologizing alone
sum(clean_data$prefer_feelings_youalone)
```

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

```
# Participants who prefer neither option
sum(clean_data$prefer__Neither_I_nor_2)
```

```
## [1] 24
```

Preferences by Self-Blame Group

First Preference Set

```
clean_data %>%
  group_by(high_blame) %>%
  summarise(
    I_apologize_first_1 = sum(prefer_I_apologize_first, na.rm = TRUE),
    Neither_1           = sum(prefer_Neither_I_nor, na.rm = TRUE),
    n                   = n()
  )
```

```
## # A tibble: 2 x 4
##   high_blame I_apologize_first_1 Neither_1     n
##        <dbl>               <dbl>     <dbl> <int>
## 1          0                  26         9    35
## 2          1                   9         1    10
```

Second Preference Set

Used to check whether patterns are consistent.

```
clean_data %>%
  group_by(high_blame) %>%
  summarise(
    I_apologize_alone_2 = sum(prefer_feelings_youalone, na.rm = TRUE),
    Neither_2           = sum(prefer__Neither_I_nor_2, na.rm = TRUE),
    n                   = n()
  )
```

```
## # A tibble: 2 x 4
##   high_blame I_apologize_alone_2 Neither_2     n
##        <dbl>               <dbl>     <dbl> <int>
## 1          0                  14        21    35
## 2          1                   7         3    10
```

Visualization

Prepare Data for Plotting

```
# Overall data
plot_data_all <- clean_data %>%
```

```
  select(all_of(feeling_vars)) %>%
  pivot_longer(
    cols      = all_of(feeling_vars),
    names_to  = "scenario",
    values_to = "feeling"
  )

# Grouped by blame level
plot_data_group <- clean_data %>%
  select(high_blame, all_of(feeling_vars)) %>%
  pivot_longer(
    cols      = all_of(feeling_vars),
    names_to  = "scenario",
    values_to = "feeling"
  )
```

Create Plots

Overall Distribution

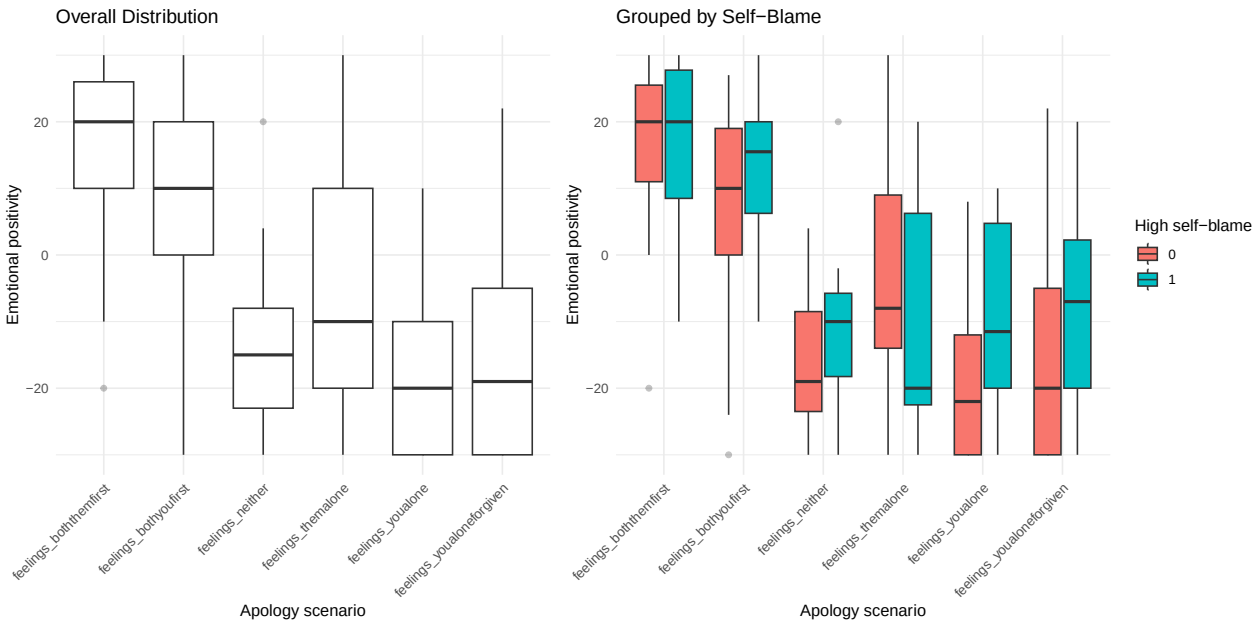
```
p1 <- ggplot(plot_data_all, aes(x = scenario, y = feeling)) +
  geom_boxplot(outlier.alpha = 0.3) +
  labs(
    title = "Overall Distribution",
    x     = "Apology scenario",
    y     = "Emotional positivity"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```

Distribution by Self-Blame Group

```
p2 <- ggplot(plot_data_group, aes(
  x    = scenario,
  y    = feeling,
  fill = factor(high_blame)
)) +
  geom_boxplot(outlier.alpha = 0.3) +
  labs(
    title = "Grouped by Self-Blame",
    x     = "Apology scenario",
    y     = "Emotional positivity",
    fill  = "High self-blame"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```

Combined Plot

```
p_all <- p1 + p2 + plot_layout(guides = "collect")
p_all
```



Save Figure

```
ggsave(
  filename = file.path(figures, "feeling_distribution.png"),
  plot     = p_all
)
```