Data Wrangling and Analysis of Experiment Results

Masoud Jasbi

7/8/2019

Data Wrangling

```
qualtrics_data <- read.csv("data/raw/Pilot-Quantifiers-data-trimmed.csv", na.strings = "")

N_dropout <-
  qualtrics_data %>%
  filter(Finished=="FALSE") %>%
  nrow()

N_participants <-
    qualtrics_data %>%
  filter(Finished=="TRUE") %>%
  nrow()
```

Total of 69 took this survey. 13 did not finish the survey and dropped out. The results for 56 are reported.

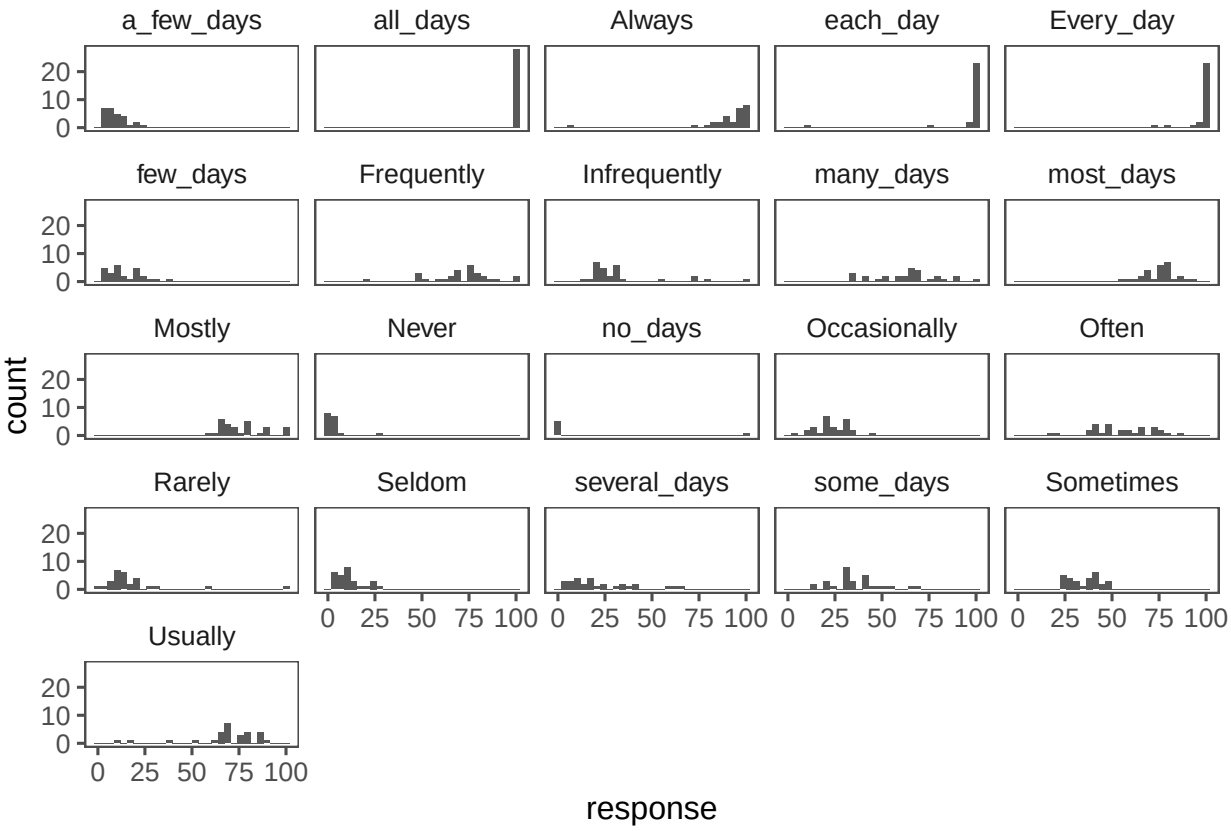
```
tidy_data <-
  qualtrics_data %>%
  filter(Finished=="TRUE") %>%
  mutate(sid=c(1:56)) %>%
  select(-Finished) %>%
  gather(quantifier, response, Every_day:Never) %>%
  drop_na(response) %>%
  mutate(condition = c(rep("nominal", 255), rep("temporal", 297))) %>%
  mutate(response = as.numeric(response))

write_csv(tidy_data, "data/processed/Pilot-Quantifiers-data-processed.csv")
```

Plots

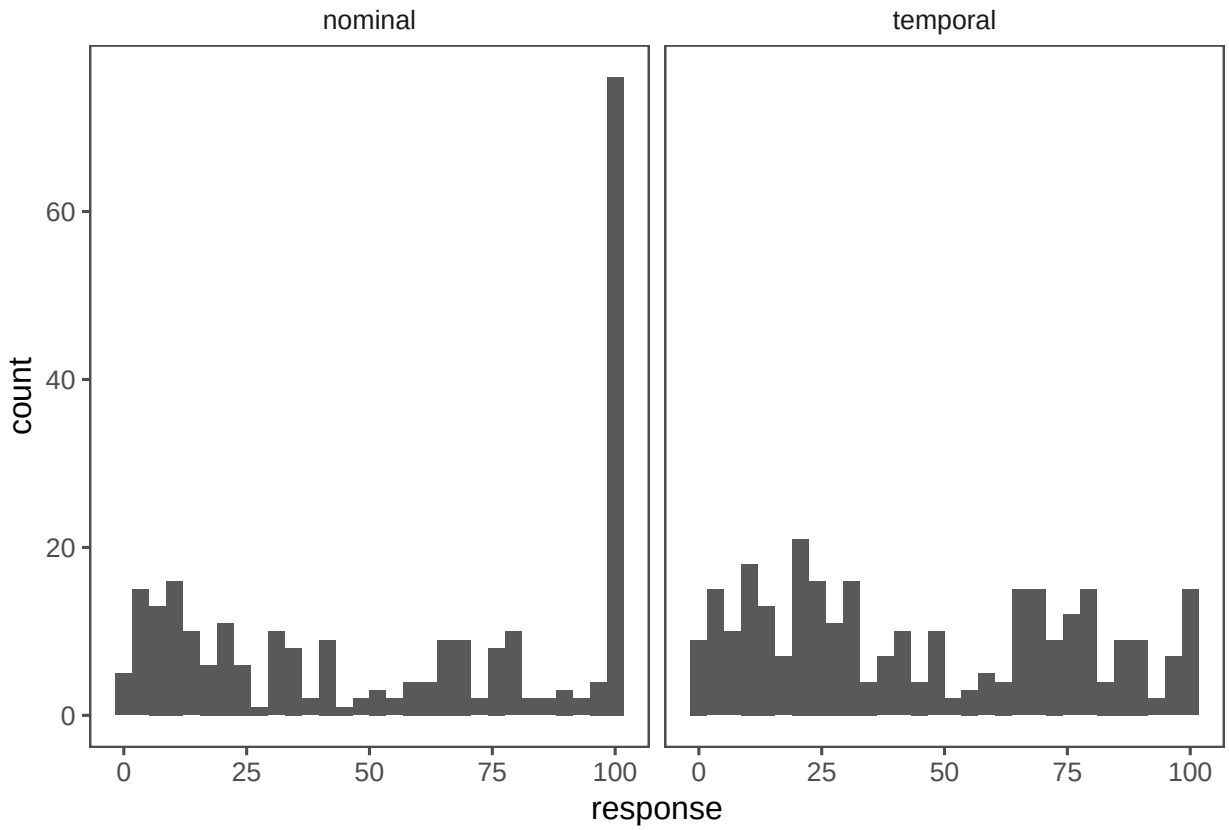
```
tidy_data %>%
  ggplot(aes(response)) +
  geom_histogram() +
  facet_wrap(.~quantifier) +
  theme_few()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



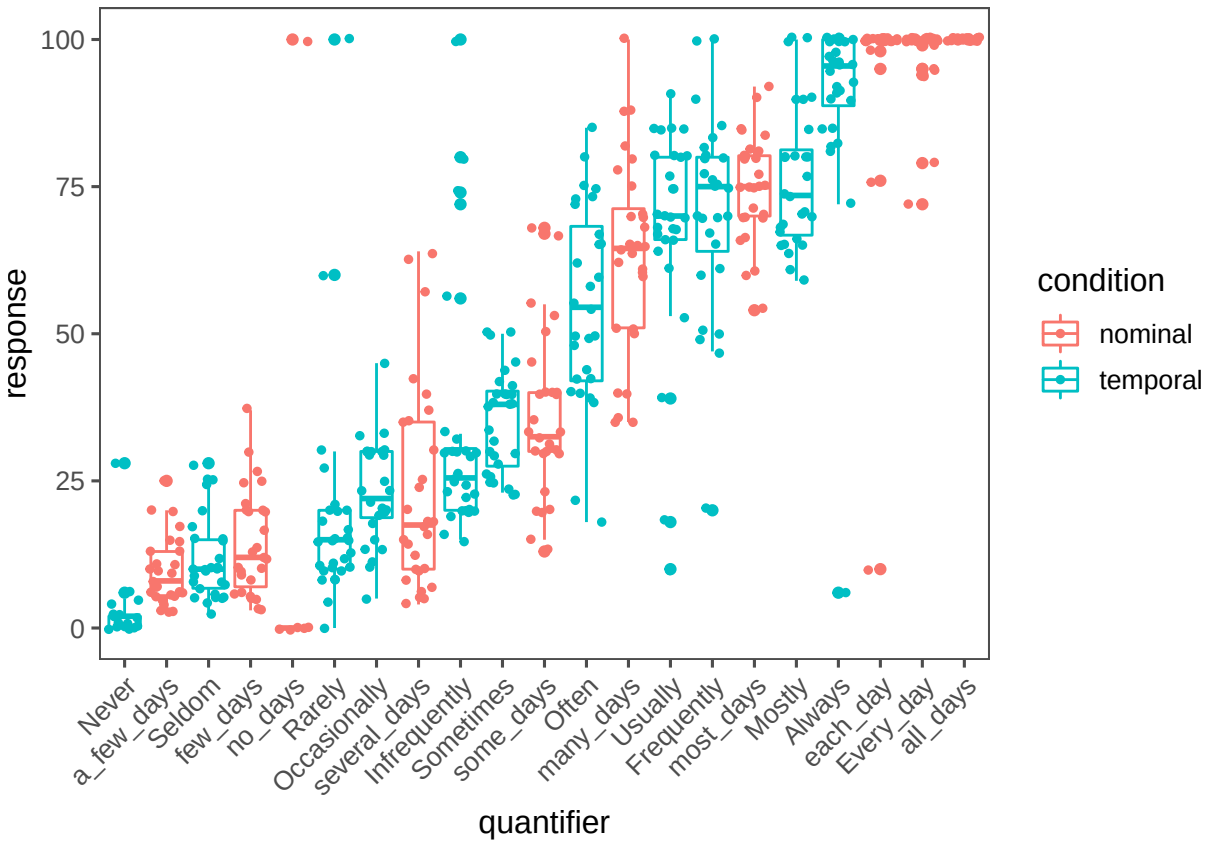

```
tidy_data %>%
  ggplot(aes(response)) +
  geom_histogram() +
  facet_wrap(.~condition) +
  theme_few()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
tidy_data$quantifier <- reorder(tidy_data$quantifier, tidy_data$response, mean)

tidy_data %>%
  ggplot(aes(quantifier, response, color=condition), fill=condition) +
  geom_boxplot() +
  geom_jitter(size=1) +
  theme_few() +
  theme(axis.text.x = element_text(angle=45, hjust = 1, vjust = 1))
```



Analysis

K-means Clustering

```
kmeans_data1 <-
  tidy_data %>%
  filter(condition=="nominal") %>%
  spread(quantifier, response) %>%
  select(-sid, -condition)

kmeans_data2 <-
  tidy_data %>%
  filter(condition=="temporal") %>%
  spread(quantifier, response) %>%
  select(-sid, -condition)

kmeans_data <-
  bind_cols(kmeans_data1, kmeans_data2) %>%
  select(-no_days, Never) %>%
  drop_na()

df2 <- data.frame(t(kmeans_data[-1]))
colnames(df2) <- kmeans_data[, 1]

quantifier_clusters <- kmeans(df2, centers=4, nstart = 25)
quantifier_clusters$cluster
```

```
##     few_days several_days    some_days    many_days    most_days
##            2            2            4            1            1
##    each_day    Every_day     all_days        Never       Seldom
##            3            3            3            2            2
##       Rarely Occasionally Infrequently    Sometimes        Often
##            2            2            4            4            1
##     Usually   Frequently       Mostly       Always
##            1            1            1            3
```

4 clusters:

- each-day, every-day, all-days, always
- often, usually, frequently, mostly, many-days, most-days
- infrequently, sometimes, some-days,
- few-days, several-days, seldom, rarely, occasionally, (never)

5 clusters:

- every-day, each-day, always
- often, usually
- most-days, many-days, frequently, mostly
- few-days, several-days, rarely, occasionally, never
- some-days, some-times