DATA 606 Data Project Proposal

Avery Davidowitz

Import Libraries

```
library(tidyverse)
```

Data Preparation

```
# load data
url <- "https://raw.githubusercontent.com/adavidowitz100/DATA606/main/FinalProject/email_data.csv"

raw_email_data <- readr::read_csv(url, col_types = list(
              Seniority = col_factor(),
              Function = col_factor(),
              "Subject Line" = col_character(),
              "Hook 1" = col_character(),
              "Z_Contact Status" = col_factor(),
              "Email Opened" = col_date(format = "%m/%d/%Y"),
              "Email Replied" = col_date(format = "%m/%d/%Y")
))

df <- raw_email_data |> dplyr::rename(seniority = Seniority,
                                      department = Function,
                                      subject = "Subject Line",
                                      hook = "Hook 1",
                                      status = "Z_Contact Status",
                                      opened_date = "Email Opened",
                                      replied_date = "Email Replied") |>
                        dplyr::mutate(opened = if_else(!is.na(opened_date), TRUE, FALSE)) |>
                        dplyr::mutate(replied = if_else(!is.na(replied_date), TRUE, FALSE)) |>
                        dplyr::mutate(days_diff = difftime(opened_date, replied_date, units = "days"))
head(df)
```

```
## # A tibble: 6 x 10
##   seniority department subject hook  status opened_d~1 replied_~2 opened replied
##   <fct>     <fct>      <chr>   <chr> <fct>  <date>     <date>     <lgl>  <lgl>
## 1 Director  Logistics  "Your ~ "Imp~ Unres~ NA         NA         FALSE  FALSE
## 2 Director  Distribut~ "Your ~ "Loo~ Attem~ NA         NA         FALSE  FALSE
## 3 VP        Supply Ch~ "Your ~ "Saw~ Attem~ NA         NA         FALSE  FALSE
## 4 Director  Logistics  "Your ~ "Enj~ Unres~ NA         NA         FALSE  FALSE
## 5 VP        Supply Ch~ "You +~ "Imp~ Unres~ NA         NA         FALSE  FALSE
## 6 VP        Distribut~ "Your ~ "Rea~ Unres~ 2022-07-20 NA         TRUE   FALSE
```

```
## # ... with 1 more variable: days_diff <drtn>, and abbreviated variable names
## #   1: opened_date, 2: replied_date
```

Research question

Does the recipient's seniority and department influence the effectiveness of this email marketing campaign? If the current campaign in not equally effective for all recipients the next campaign could be designed to be targeted more effectively by seniority and department.

Cases

There are 1449 cases/observations representing marketing emails sent to specific people.

Data collection

Real sanitized email market research data was used.

Type of study

The data is observational in nature.

Data Source

The data is from a business to business email market research campaign. The data can be viewed on Github at https://raw.githubusercontent.com/adavidowitz100/DATA606/main/FinalProject/email_data.csv

Dependent Variable

The dependent variables are "opened", "replied" and "days_diff" representing if an email was opened, responded to and the days between those events.

Independent Variable(s)

The independent variables are department and seniority.

Relevant summary statistics

```
seniority_open_xtab <- xtabs(~ seniority + opened, data=df)
seniority_open_xtab
```

```
##              opened
## seniority     FALSE TRUE
##   Director      372  428
##   VP             99  102
##   Sr. Manager    56   51
##   Head           16   18
##   Vp             72   88
```

```
##   Manager        54   78
##   C suite         2    7
##   Senior          1    2
##   Partner         0    3
```

```
prop.table(seniority_open_xtab, 1)
```

```
##              opened
## seniority         FALSE      TRUE
##   Director    0.4650000 0.5350000
##   VP          0.4925373 0.5074627
##   Sr. Manager 0.5233645 0.4766355
##   Head        0.4705882 0.5294118
##   Vp          0.4500000 0.5500000
##   Manager     0.4090909 0.5909091
##   C suite     0.2222222 0.7777778
##   Senior      0.3333333 0.6666667
##   Partner     0.0000000 1.0000000
```

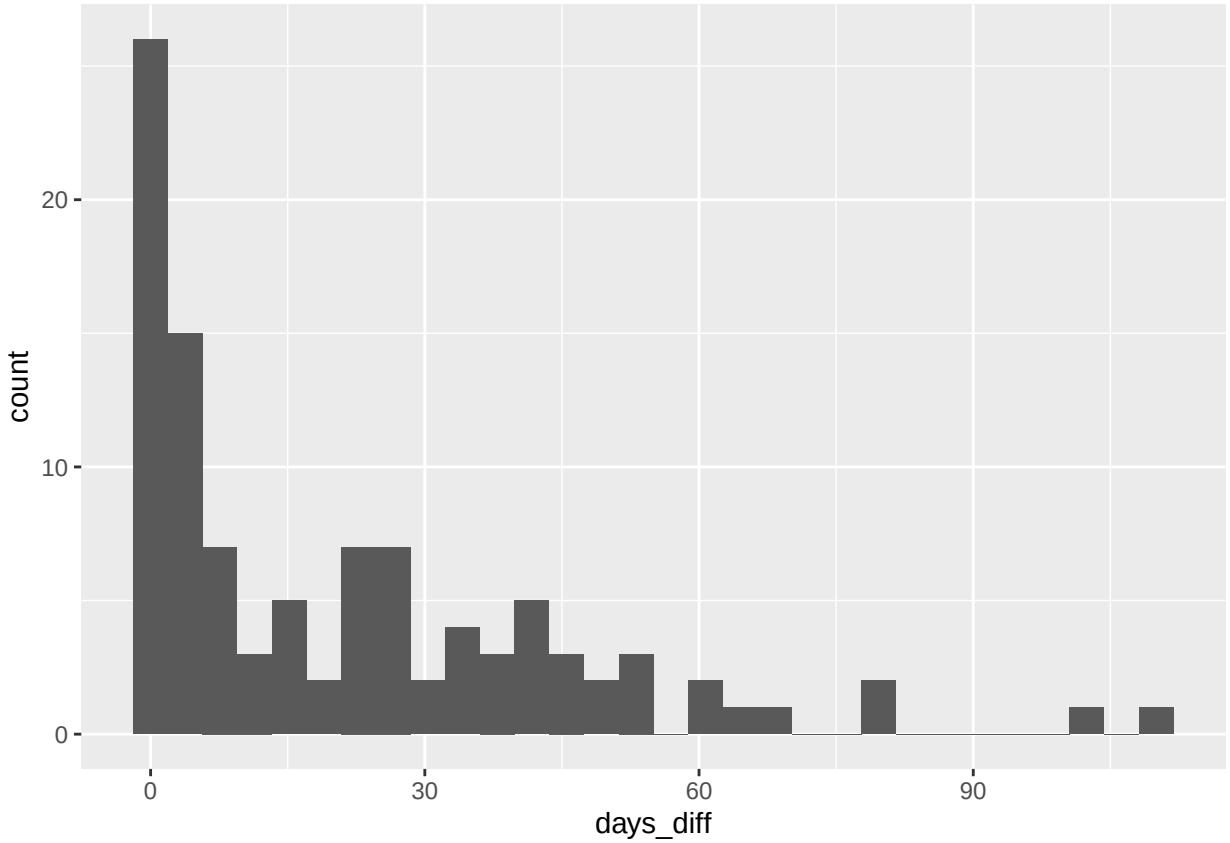
```
seniority_replied_xtab <- xtabs(~ seniority + replied, data=df)
seniority_replied_xtab
```

```
##              replied
## seniority     FALSE TRUE
##   Director      749   51
##   VP            188   13
##   Sr. Manager    95   12
##   Head           34    0
##   Vp            145   15
##   Manager       122   10
##   C suite         8    1
##   Senior          3    0
##   Partner         3    0
```

```
prop.table(seniority_replied_xtab, 1)
```

```
##              replied
## seniority          FALSE       TRUE
##   Director    0.93625000 0.06375000
##   VP          0.93532338 0.06467662
##   Sr. Manager 0.88785047 0.11214953
##   Head        1.00000000 0.00000000
##   Vp          0.90625000 0.09375000
##   Manager     0.92424242 0.07575758
##   C suite     0.88888889 0.11111111
##   Senior      1.00000000 0.00000000
##   Partner     1.00000000 0.00000000
```

```
ggplot(df, aes(x=days_diff)) + geom_histogram()
```


count
20
10
0
0
30
60
90
days_diff