WAL Data Analysis

```
usethis::use_git_config(user.name = "hanastepnick", user.email = "hana.stepnick@duke.edu")
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------
## v tibble  3.0.3     v purrr   0.3.4
## v tidyr   1.1.1     v dplyr   1.0.1
## v readr   1.3.1     v forcats 0.5.0
## -- Conflicts -----------------------------------------------------------------
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x readr::guess_encoding()  masks rvest::guess_encoding()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x purrr::pluck()           masks rvest::pluck()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
library(infer)
wal <- read_csv("wal.csv")
```

```
## Parsed with column specification:
## cols(
##   year = col_character(),
##   major = col_character(),
##   major_type = col_character(),
##   hispanic = col_character(),
##   race = col_character(),
##   ask_question = col_double(),
##   answer_question = col_double(),
##   reach_male = col_double(),
##   reach_female = col_double(),
##   lead_group = col_double(),
##   disagree = col_double(),
##   accountable = col_double(),
##   breakdown = col_double(),
##   gender = col_character(),
##   school = col_character()
## )
```
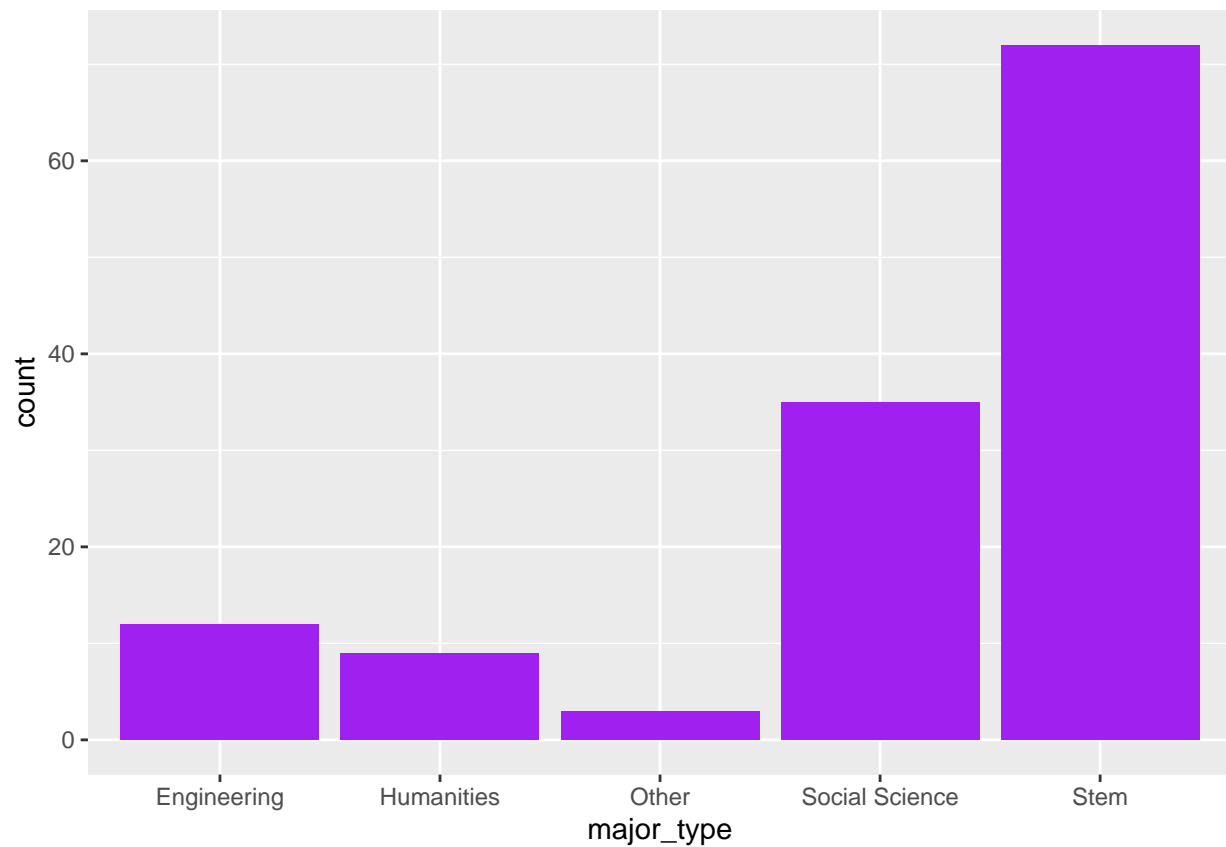
```
majors <- wal %>%
  distinct(major)
```

Distribution of all major breakdowns:

```
ggplot(data = wal, mapping = aes(x = major_type)) +
  geom_bar()
```
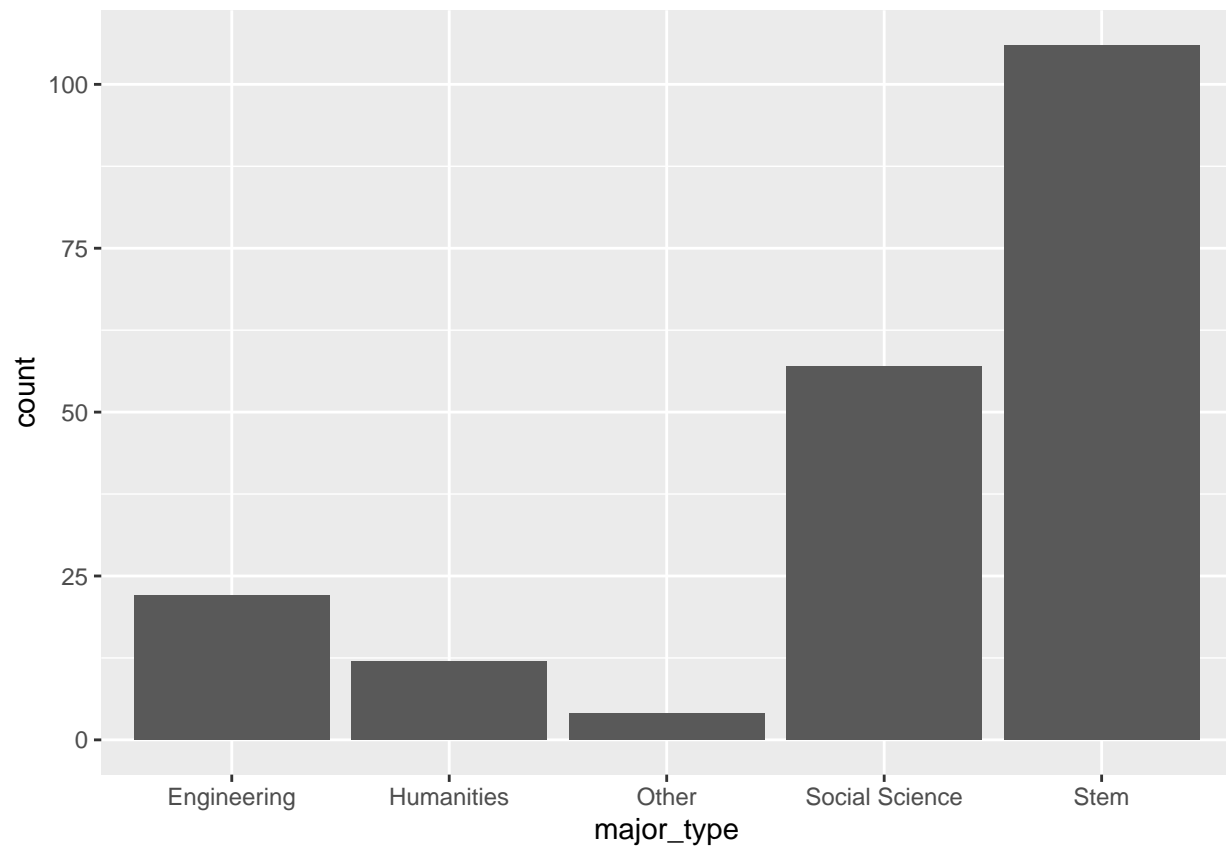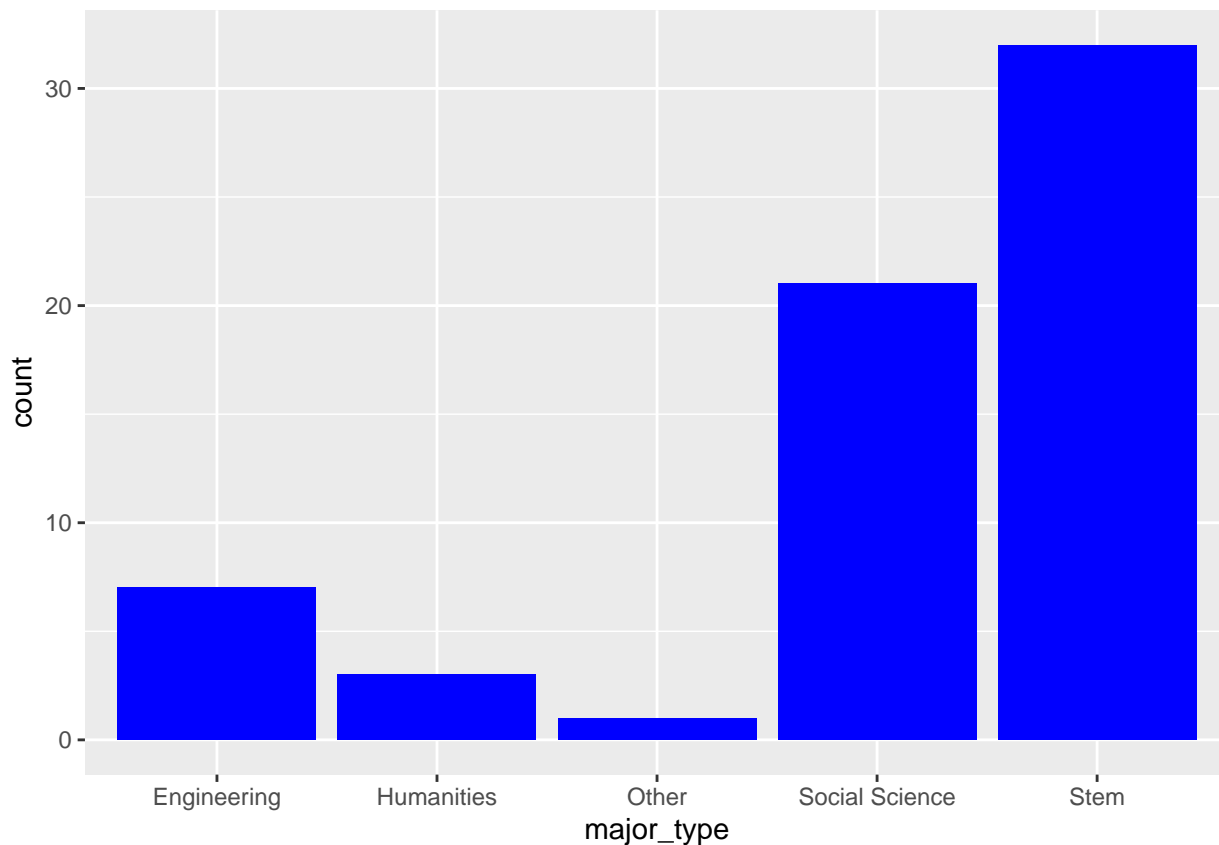
Distribution of major breakdowns for females:

```r
wal_female <- wal %>%
  filter(gender == "Female")

ggplot(data = wal_female, mapping = aes(x = major_type)) +
  geom_bar(fill = "purple")
```

Distribution of major breakdowns for males:

```r
wal_male <- wal %>%
  filter(gender == "Male")

ggplot(data = wal_male, mapping = aes(x = major_type)) +
  geom_bar(fill = "blue")
```

Asking question:

```r
mean(wal_female$ask_question, na.rm = FALSE)
```

```
## [1] 3.633588
```

```r
mean(wal_male$ask_question, na.rm = FALSE)
```

```
## [1] 3.71875
```

```r
wal_male_female <- wal %>%
  filter(gender == "Female" | gender == "Male")


t_ask <- wal_male_female %>%
  t_test(ask_question ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)
t_ask
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.492  113.   0.312 less
```

Answering question:

```r
mean(wal_female$answer_question, na.rm = FALSE)
```

```
## [1] 3.335878
```

```
mean(wal_male$answer_question, na.rm = FALSE)
```

[1] 3.609375

```
t_answer <- wal_male_female %>%
  t_test(answer_question ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)
t_answer
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1     -1.56  114.  0.0612 less
```

Reach out to male prof:

```
mean(wal_female$reach_male, na.rm = FALSE)
```

[1] 3.473282

```
mean(wal_male$reach_male, na.rm = FALSE)
```

[1] 4.34375

```
t_reach_male <- wal_male_female %>%
  t_test(reach_male ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)
t_reach_male
```

```
## # A tibble: 1 x 4
##   statistic  t_df      p_value alternative
##       <dbl> <dbl>        <dbl> <chr>
## 1     -5.77  152. 0.0000000220 less
```

Reach out to female prof:

```
mean(wal_female$reach_female, na.rm = FALSE)
```

[1] 4.381679

```
mean(wal_male$reach_female, na.rm = FALSE)
```

[1] 4.359375

```
t_reach_female <- wal_male_female %>%
  t_test(reach_female ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)
t_reach_female
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.156  105.   0.438 less
```

Comfort with leading a group:

```
  mean(wal_female$lead_group, na.rm = FALSE)
```

[1] 3.877863

```
  mean(wal_male$lead_group, na.rm = FALSE)
```

[1] 3.75

```
t_lead <- wal_male_female %>%
  t_test(lead_group ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)
t_lead
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.814  108.   0.209 less
```

Comfort with disagreeing with group:

```
  mean(wal_female$disagree, na.rm = FALSE)
```

[1] 3.610687

```
  mean(wal_male$disagree, na.rm = FALSE)
```

[1] 3.546875

```
t_disagree <- wal_male_female %>%
  t_test(disagree ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)
t_disagree
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.373  116.   0.355 less
```

Comfort with keeping others accountable:

```
  mean(wal_female$accountable, na.rm = FALSE)
```

[1] 3.351145

```
  mean(wal_male$accountable, na.rm = FALSE)
```

[1] 3.375

```
t_accountable <- wal_male_female %>%
  t_test(accountable ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)
t_accountable
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
```

```
##        <dbl> <dbl>   <dbl> <chr>
## 1    -0.137  131.   0.445 less
```

Perceptions of major breakdowns:

```r
wal_stem <- wal %>%
  filter(major_type == "Stem")

mean(wal_stem$breakdown, na.rm = FALSE)
```

```
## [1] 5.028302
```

```r
wal_humanities <- wal %>%
  filter(major_type == "Humanities")

mean(wal_humanities$breakdown, na.rm = FALSE)
```

```
## [1] 5.083333
```

```r
wal_social_science <- wal %>%
  filter(major_type == "Social Science")

mean(wal_social_science$breakdown, na.rm = FALSE)
```

```
## [1] 4.929825
```

```r
wal_engineering <- wal %>%
  filter(major_type == "Engineering")

mean(wal_engineering$breakdown, na.rm = FALSE)
```

```
## [1] 4.227273
```

```r
wal_stem_engineering <- wal %>%
  filter(major_type == "Engineering" | major_type == "Stem")

mean(wal_stem_engineering$breakdown, na.rm = FALSE)
```

```
## [1] 4.890625
```

Engineering vs SS Gender Breakdown:

```r
wal_engineering_ss <- wal %>%
  filter(major_type == "Engineering" | major_type == "Social Science")

t_engineering_ss_breakdown <- wal_engineering_ss %>%
  t_test(breakdown ~ major_type,
         order = c("Engineering", "Social Science"),
         alternative = "less",
         conf_int = FALSE)

t_engineering_ss_breakdown
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1     -2.32  48.8  0.0122 less
```

Trinity vs. Pratt analyses:

Trinity:

```
trinity_female <- wal %>%
  filter(school == "Trinity") %>%
  filter(gender == "Female")
```

```
trinity_male <- wal %>%
  filter(school == "Trinity") %>%
  filter(gender == "Male")
```

```
mean(trinity_female$ask_question, na.rm = FALSE)
```

```
## [1] 3.655462
```

```
mean(trinity_male$ask_question, na.rm = FALSE)
```

```
## [1] 3.719298
```

```
wal_trinity <- wal %>%
  filter(school == "Trinity") %>%
  filter(gender == "Male" | gender == "Female")
```

```
t_ask_trinity <- wal_trinity %>%
  t_test(ask_question ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)
```

```
t_ask_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.353  100.   0.362 less
```

```
mean(trinity_female$answer_question, na.rm = FALSE)
```

```
## [1] 3.319328
```

```
mean(trinity_male$answer_question, na.rm = FALSE)
```

```
## [1] 3.614035
```

```
t_answer_trinity <- wal_trinity %>%
  t_test(answer_question ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)
```

```
t_answer_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -1.61   104.  0.0555 less
```

```
mean(trinity_female$reach_male, na.rm = FALSE)
```

```
## [1] 3.420168
```

```r
mean(trinity_male$reach_male, na.rm = FALSE)
```

```
## [1] 4.45614
```

```r
t_reachmale_trinity <- wal_trinity %>%
  t_test(reach_male ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_reachmale_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df  p_value alternative
##       <dbl> <dbl>    <dbl> <chr>
## 1     -7.20  155. 1.20e-11 less
```

```r
mean(trinity_female$reach_female, na.rm = FALSE)
```

```
## [1] 4.361345
```

```r
mean(trinity_male$reach_female, na.rm = FALSE)
```

```
## [1] 4.473684
```

```r
t_reachfemale_trinity <- wal_trinity %>%
  t_test(reach_female ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_reachfemale_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.827  102.   0.205 less
```

```r
mean(trinity_female$lead_group, na.rm = FALSE)
```

```
## [1] 3.957983
```

```r
mean(trinity_male$lead_group, na.rm = FALSE)
```

```
## [1] 3.701754
```

```r
t_leadgroup_trinity <- wal_trinity %>%
  t_test(lead_group ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)

t_leadgroup_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1     -1.63  98.0  0.0533 less
```

```
mean(trinity_female$disagree, na.rm = FALSE)
```

```
## [1] 3.655462
```

```
mean(trinity_male$disagree, na.rm = FALSE)
```

```
## [1] 3.526316
```

```
t_disagree_trinity <- wal_trinity %>%
  t_test(disagree ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)

t_disagree_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.736  107.   0.232 less
```

```
mean(trinity_female$accountable, na.rm = FALSE)
```

```
## [1] 3.378151
```

```
mean(trinity_male$accountable, na.rm = FALSE)
```

```
## [1] 3.421053
```

```
t_accountable_trinity <- wal_trinity %>%
  t_test(accountable ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_accountable_trinity
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.237  125.   0.406 less
```

Pratt:

```
pratt_female <- wal %>%
  filter(school == "Pratt") %>%
  filter(gender == "Female")
```

```
pratt_male <- wal %>%
  filter(school == "Pratt") %>%
  filter(gender == "Male")
```

```
mean(pratt_female$ask_question, na.rm = FALSE)
```

```
## [1] 3.416667
```

```
mean(pratt_male$ask_question, na.rm = FALSE)
```

```
## [1] 3.714286
```

```
wal_pratt <- wal %>%
  filter(school == "Pratt") %>%
  filter(gender == "Male" | gender == "Female")

t_ask_pratt <- wal_pratt %>%
  t_test(ask_question ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_ask_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.480  11.0   0.320 less
```

```
mean(pratt_female$answer_question, na.rm = FALSE)
```

```
## [1] 3.5
```

```
mean(pratt_male$answer_question, na.rm = FALSE)
```

```
## [1] 3.571429
```

```
t_answer_pratt <- wal_pratt %>%
  t_test(answer_question ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_answer_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.114  8.56   0.456 less
```

```
mean(pratt_female$reach_male, na.rm = FALSE)
```

```
## [1] 4
```

```
mean(pratt_male$reach_male, na.rm = FALSE)
```

```
## [1] 3.428571
```

```
t_reachmale_pratt <- wal_pratt %>%
  t_test(reach_male ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)

t_reachmale_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.869  9.96   0.203 less
```

```r
mean(pratt_female$reach_female, na.rm = FALSE)
```

```
## [1] 4.583333
```

```r
mean(pratt_male$reach_female, na.rm = FALSE)
```

```
## [1] 3.428571
```

```r
t_reachfemale_pratt <- wal_pratt %>%
  t_test(reach_female ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)

t_reachfemale_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1     -1.81  9.11  0.0521 less
```

```r
mean(pratt_female$lead_group, na.rm = FALSE)
```

```
## [1] 3.083333
```

```r
mean(pratt_male$lead_group, na.rm = FALSE)
```

```
## [1] 4.142857
```

```r
t_leadgroup_pratt <- wal_pratt %>%
  t_test(lead_group ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_leadgroup_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1     -1.66  7.82  0.0679 less
```

```r
mean(pratt_female$disagree, na.rm = FALSE)
```

```
## [1] 3.166667
```

```r
mean(pratt_male$disagree, na.rm = FALSE)
```

```
## [1] 3.714286
```

```r
t_disagree_pratt <- wal_pratt %>%
  t_test(disagree ~ gender,
         order = c("Female", "Male"),
         alternative = "less",
         conf_int = FALSE)

t_disagree_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
```

```
##       <dbl> <dbl>   <dbl> <chr>
## 1   -0.825  8.45   0.216 less
```

```
mean(pratt_female$accountable, na.rm = FALSE)
```

```
## [1] 3.083333
```

```
mean(pratt_male$accountable, na.rm = FALSE)
```

```
## [1] 3
```

```
t_accountable_pratt <- wal_pratt%>%
  t_test(accountable ~ gender,
         order = c("Male", "Female"),
         alternative = "less",
         conf_int = FALSE)

t_accountable_pratt
```

```
## # A tibble: 1 x 4
##   statistic  t_df p_value alternative
##       <dbl> <dbl>   <dbl> <chr>
## 1    -0.143  6.25   0.445 less
```