WAL Data Analysis

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

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## 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 ------------------------------------------ tidyverse_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()
```

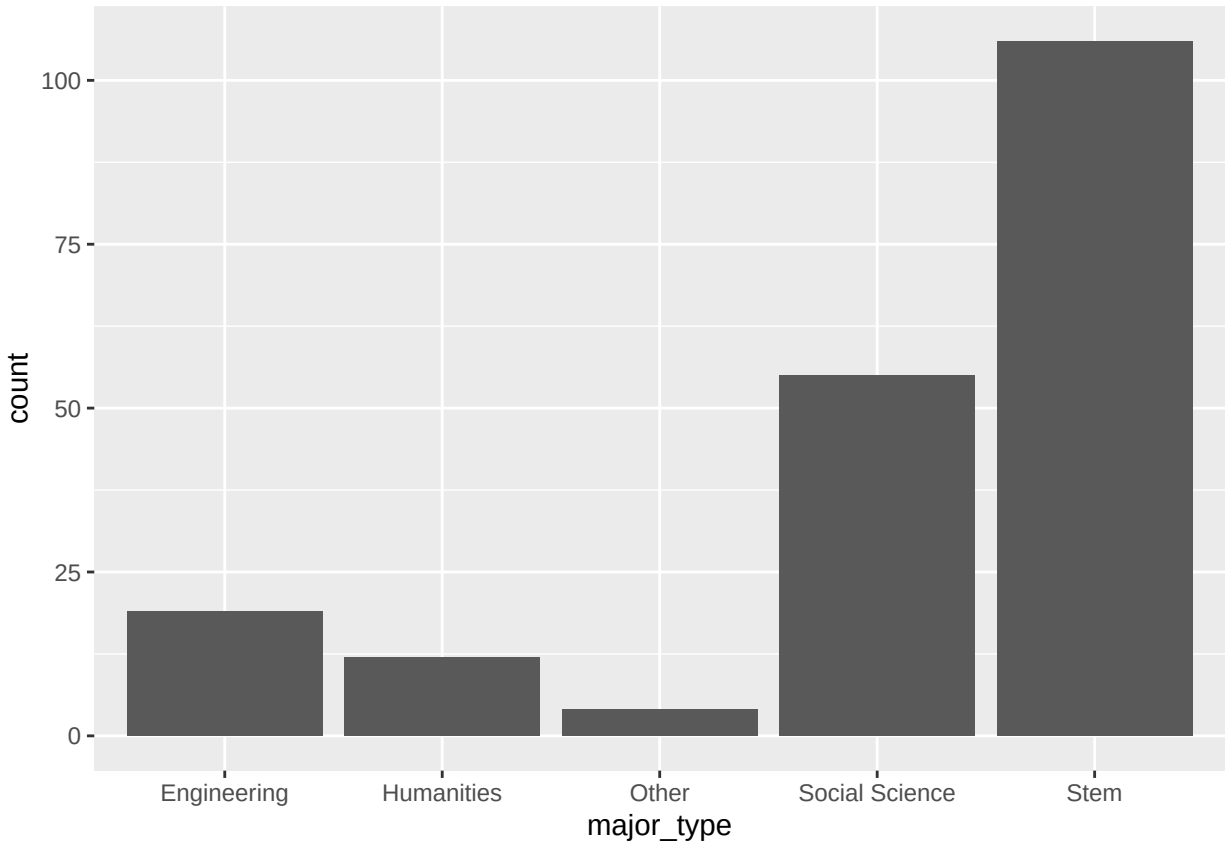
```
library(infer)
wal <- read_csv("wal.csv")
```

```
## 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()
## )
```

```
majors <- wal %>%
  distinct(major)
```

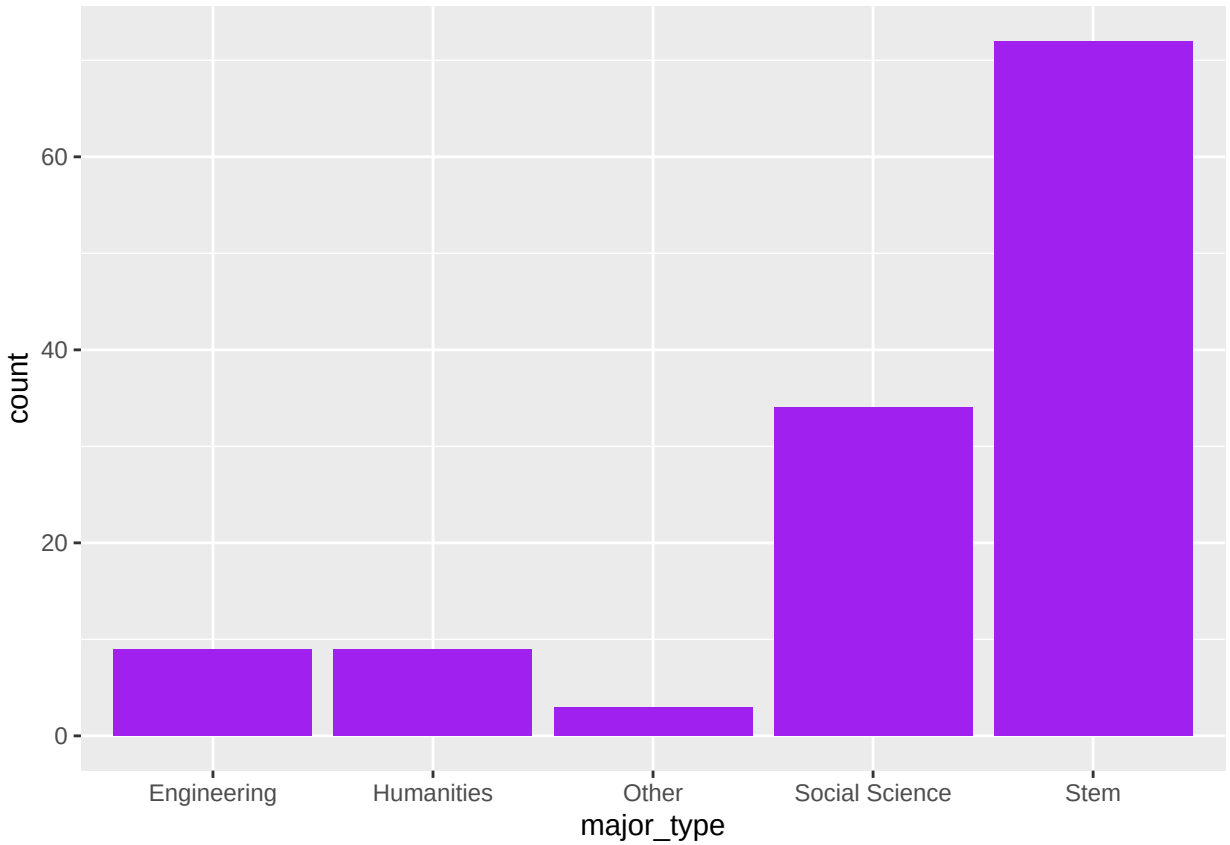
Distribution of all major breakdowns:

```
ggplot(data = wal, mapping = aes(x = major_type)) +
  geom_bar()
```



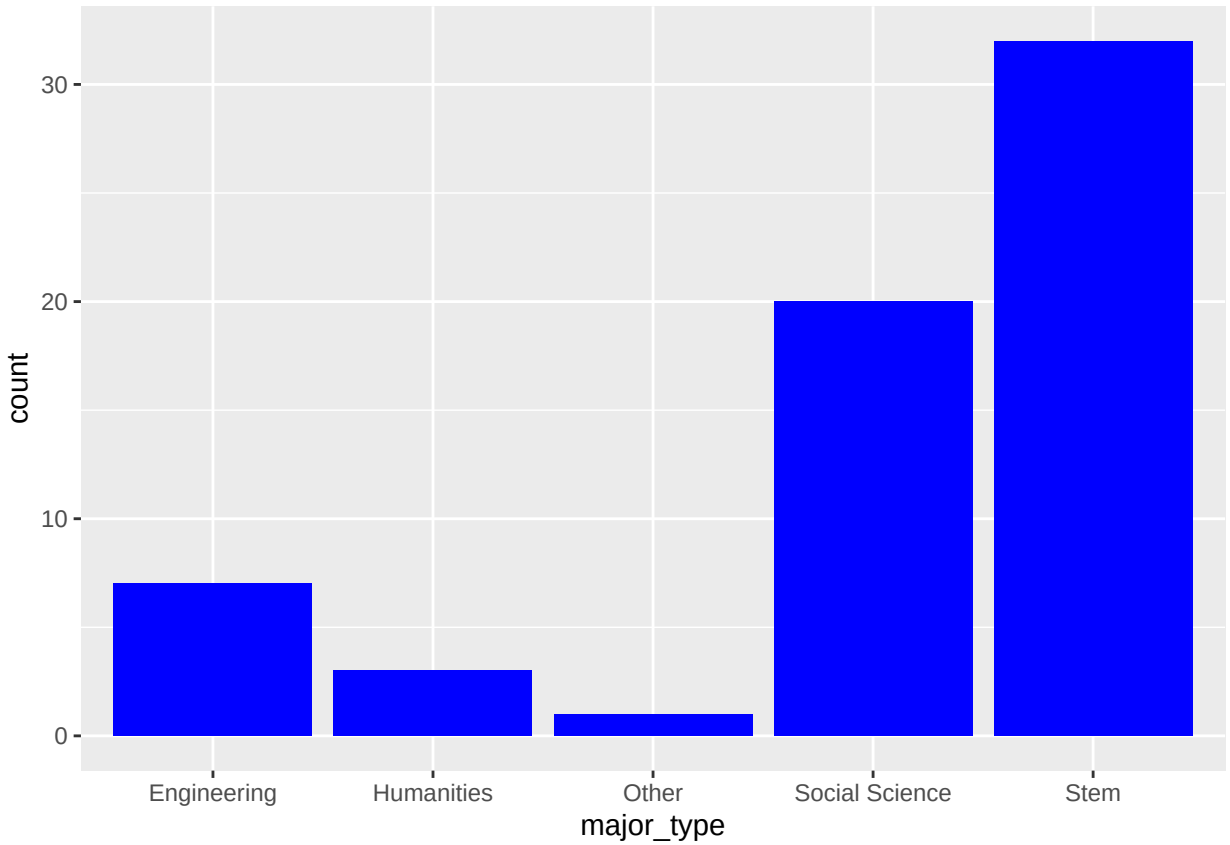
Distribution of major breakdowns for females:

```
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:

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

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



Asking question:

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

```
## [1] 3.661417
```

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

```
## [1] 3.698413
```

```
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    -0.212  112.   0.416 less
```

Answering question:

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

```
## [1] 3.338583
```

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

```
## [1] 3.587302
```

```
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     -1.40  114.  0.0823 less
```

Reach out to male prof:

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

```
## [1] 3.472441
```

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

```
## [1] 4.333333
```

```
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     -5.60  151. 0.0000000502 less
```

Reach out to female prof:

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

```
## [1] 4.362205
```

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

```
## [1] 4.349206
```

```
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   -0.0894  104.   0.464 less
```

Comfort with disagreeing with group:

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

```
## [1] 3.629921
```

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

```
## [1] 3.555556
```

```
t_disagree <- wal_male_female %>%
  t_test(lead_group ~ 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    -0.801  107.   0.212 less
```

Comfort with keeping others accountable:

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

```
## [1] 3.362205
```

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

```
## [1] 3.365079
```

```
t_accountable <- wal_male_female %>%
  t_test(lead_group ~ 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     0.801  107.   0.788 less
```

Perceptions of major breakdowns:

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

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

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

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

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

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

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

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

```
## [1] 4.909091
```

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

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

```
## [1] 4.105263
```

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

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

```
## [1] 4.888
```

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     -2.70  44.8 0.00486 less
```