EDUC112

Nahian Haque

2024-09-11

```
library(sf)
```

```
## Warning: package 'sf' was built under R version 4.4.1
```

```
## Linking to GEOS 3.12.1, GDAL 3.8.4, PROJ 9.3.1; sf_use_s2() is TRUE
```

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
survey <- read.csv("C:/Users/nahia/Downloads/sfusd_survey_4_9_24.csv")
```

```
# Calculate proportions for each unique value in multiple columns
cols <- c("race", "education", "gay", "child_gay", "user_lang", "spec_ed", "ell", "lang_path", "foster"

proportions_list <- lapply(survey[cols], function(x) prop.table(table(x)))

# Print the proportions for each column
for (i in 1:length(proportions_list)) {
  cat("Column:", cols[i], "\n")
  print(proportions_list[[i]])
}
```

```
## Column: race
## x
##                                                 African American
##                       0.422792304                     0.021312284
## American Indian or Alaskan Native                            Asian
##                       0.002269364                     0.150074001
```

```
##                   Decline to State                          Filipino
##                         0.049531327                       0.017661569
##                    Hispanic/Latino                    Pacific Islander
##                         0.102121362                       0.004736063
##                  Two or More Races                             White
##                         0.040157869                       0.189343858
## Column: education
## x
##
##                                                          0.32136162
##                                       Associates or technical degree
##                                                          0.02920572
##                                                  Bachelor's degree
##                                                          0.21243217
## Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)
##                                                          0.25939813
##                                          High school diploma or GED
##                                                          0.04666996
##                                                  Prefer not to say
##                                                          0.03650715
##                                        Some college, but no degree
##                                                          0.04548594
##                                           Some high school or less
##                                                          0.04893932
## Column: gay
## x
##                   Decline to state                No               Yes
##       0.32432166       0.07666502       0.54366058       0.05535274
## Column: child_gay
## x
##                   Decline to state                No   Not applicable
##       0.49659595       0.05476073       0.38283177       0.03127775
##               Yes
##       0.03453379
## Column: user_lang
## x
##           AR           EN           ES           SMO          TGL           VI
## 0.0009866798 0.8245683276 0.1003453379 0.0002960039 0.0001973360 0.0021706956
##         ZH-S
## 0.0714356191
## Column: spec_ed
## x
##          0          1
## 0.92816971 0.07183029
## Column: ell
## x
##          0          1
## 0.94070054 0.05929946
## Column: lang_path
## x
##          0          1
## 0.90823878 0.09176122
## Column: foster
## x
```

```
##           0           1
## 0.997730636 0.002269364
## Column: homeless
## x
##           0           1
## 0.998223976 0.001776024
## Column: role
## x
##                       Central office staff     Community member
##          0.326689689          0.004045387          0.011445486
##    Community partner       Multiple roles     Parent/Caregiver
##          0.005722743          0.082683769          0.481499753
##    School site staff              Student
##          0.063640849          0.024272324
```

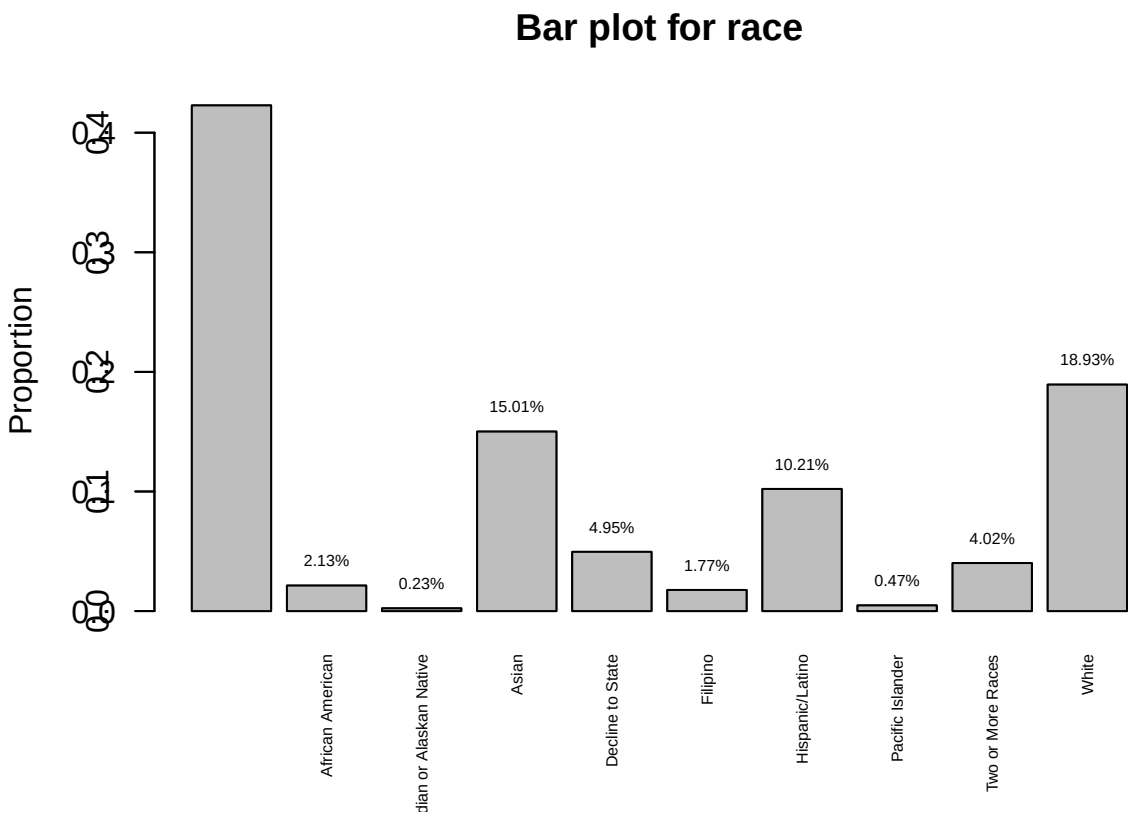
```r
dfs <- list()

for (i in 1:length(proportions_list)) {
  # Create a data frame for the current column
  df <- as.data.frame(proportions_list[[i]])
  # Set appropriate column names
  colnames(df) <- c("category", "proportion")
  # Add the column name as an identifier
  df$column <- cols[i]
  # Append the data frame to the list
  dfs[[i]] <- df
}
```

```r
dfs[[1]]
```

```
##                              category  proportion column
## 1                                     0.422792304   race
## 2                    African American 0.021312284   race
## 3  American Indian or Alaskan Native 0.002269364   race
## 4                               Asian 0.150074001   race
## 5                   Decline to State 0.049531327   race
## 6                           Filipino 0.017661569   race
## 7                    Hispanic/Latino 0.102121362   race
## 8                   Pacific Islander 0.004736063   race
## 9                  Two or More Races 0.040157869   race
## 10                              White 0.189343858   race
```

```r
barplot(height = dfs[[1]]$proportion, names.arg = dfs[[1]]$category,
        main = paste("Bar plot for", dfs[[1]]$column[1]),
        ylab = "Proportion",
        las = 2, cex.names = 0.5)
text(x = barplot(dfs[[1]]$proportion, add = TRUE),  # Add bars on existing plot and get x-coordinates
     y = dfs[[1]]$proportion,  # Use the same y-coordinates as bars
     labels = scales::percent(dfs[[1]]$proportion),  # Format labels as percentages
     pos = 3, cex = 0.5, col = "black")  # Adjust position, size, and color of text labels
```



Trying geospatial:

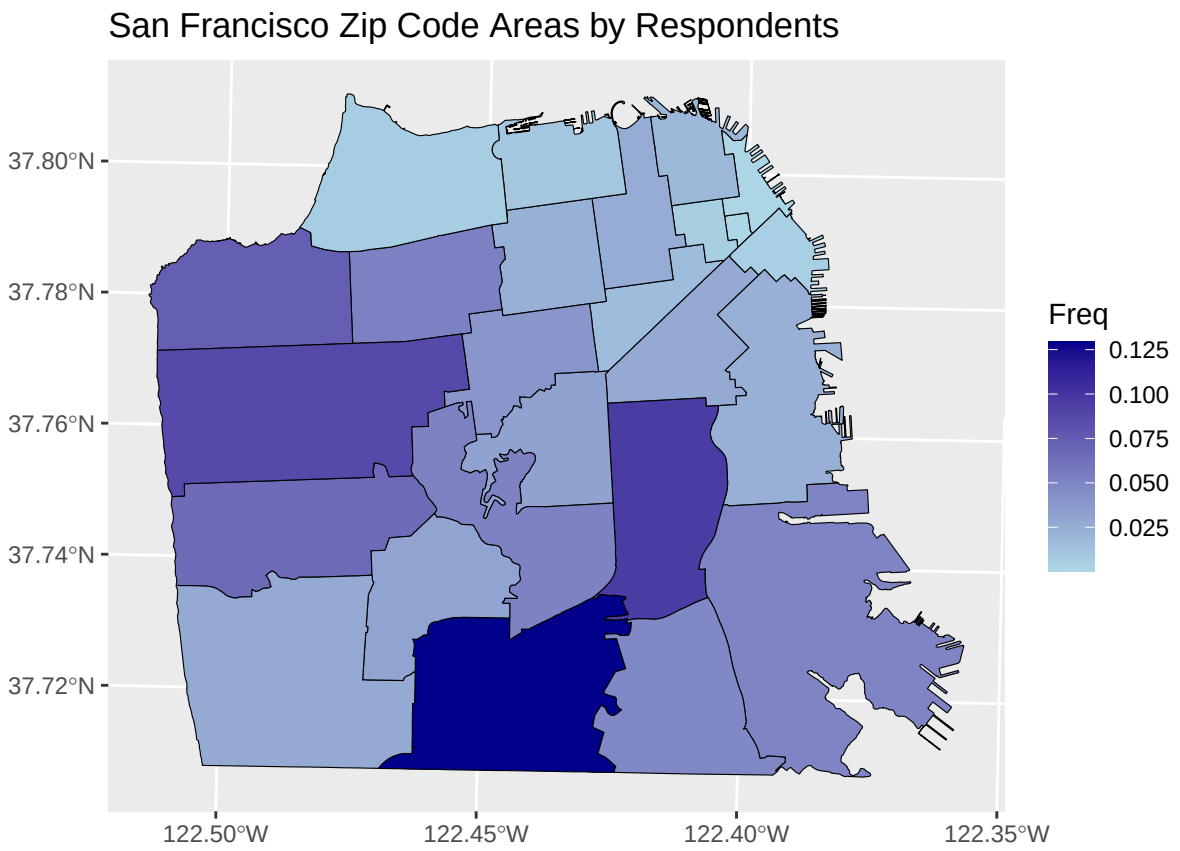
```
survey_zips <- as.data.frame(prop.table(table(survey$zip)))
```

```
sf_shp <- st_read("C:/Users/nahia/Downloads/data/sfzipcodes.shp")
```

```
## Reading layer ‘sfzipcodes’ from data source
##   ‘C:\Users\nahia\Downloads\data\sfzipcodes.shp’ using driver ‘ESRI Shapefile’
## Simple feature collection with 25 features and 3 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 5979385 ymin: 2085841 xmax: 6024665 ymax: 2123809
## Projected CRS: NAD83 / California zone 3 (ftUS)
```

```
sf_zip_merged <- merge(sf_shp, survey_zips,
                       by.x="ZIP_CODE", by.y="Var1", all.x=TRUE, all.y=TRUE)
```

```
ggplot() +
  geom_sf(data = sf_zip_merged,
          aes(fill = Freq), color = "black") +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  labs(title = "San Francisco Zip Code Areas by Respondents")
```



Note: There were 107 unique zip codes that did not correspond with a SF zip code (starting with 941)

Survey Analysis

1. Grouped by race, what are the average responses to each of the Equity, Excellence, Efficiency questions?
2. Same as 1, grouped by zip code
3. Same as 1, grouped by educ.

General Average Responses

```
colMeans(survey[, c(19:34, 40:42)], na.rm = TRUE)
```

```
##     eq_his     eq_hin      eq_as      eq_pa      ex_ap     ex_scc     ex_sel
##   3.346423   3.494602   4.182832   3.959218   4.161290   4.404356   4.191207
##      ex_sd     ex_att      er_se      er_fc      er_bu      er_bc     er_cpp
##   3.819392   3.772113   3.845383   4.058299   3.642577   3.854393   3.388194
##     er_ocs      er_tt     equity excellence efficiency
##   3.260597   3.998051   3.745805   4.070482   3.725140
```

Equity

```
survey %>%
  group_by(race) %>%
  summarise_at(vars(eq_his, eq_hin, eq_as, eq_pa, equity),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 10 x 6
##    race            eq_his_Mean_Response eq_hin_Mean_Response eq_as_Mean_Response
##    <chr>                          <dbl>                <dbl>               <dbl>
##  1 ""                              3.27                 3.36                4.16
##  2 "African Ameri~                 3.96                 3.96                4.15
##  3 "American Indi~                 3.74                 3.84                3.95
##  4 "Asian"                         3.29                 3.38                4.15
##  5 "Decline to St~                 2.63                 2.81                3.95
##  6 "Filipino"                      3.71                 3.94                4.30
##  7 "Hispanic/Lati~                 3.85                 3.91                4.26
##  8 "Pacific Islan~                 3.73                 3.78                4.27
##  9 "Two or More R~                 3.29                 3.49                4.21
## 10 "White"                         3.26                 3.51                4.23
## # i 2 more variables: eq_pa_Mean_Response <dbl>, equity_Mean_Response <dbl>
```

```
survey %>%
  group_by(zip) %>%
  summarise_at(vars(eq_his, eq_hin, eq_as, eq_pa, equity),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 136 x 6
##      zip eq_his_Mean_Response eq_hin_Mean_Response eq_as_Mean_Response
##    <int>                <dbl>                <dbl>               <dbl>
##  1     0                  NaN                  NaN                 NaN
##  2  4102                    4                  NaN                 NaN
##  3  8327                  NaN                  NaN                 NaN
##  4 11221                    1                    2                   4
##  5 53219                  NaN                  NaN                 NaN
##  6 76100                    5                    5                   4
##  7 84107                    1                    1                   1
##  8 84108                    5                    4                   5
##  9 84114                  NaN                  NaN                 NaN
## 10 84115                    1                  NaN                   1
## # i 126 more rows
## # i 2 more variables: eq_pa_Mean_Response <dbl>, equity_Mean_Response <dbl>
```

```
survey %>%
  group_by(education) %>%
  summarise_at(vars(eq_his, eq_hin, eq_as, eq_pa, equity),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 8 x 6
##   education        eq_his_Mean_Response eq_hin_Mean_Response eq_as_Mean_Response
##   <chr>                           <dbl>                <dbl>               <dbl>
## 1 ""                               3                    3.17                4.06
## 2 "Associates or ~                 3.72                 3.79                4.30
## 3 "Bachelor's deg~                 3.20                 3.40                4.18
## 4 "Graduate or pr~                 3.26                 3.46                4.19
## 5 "High school di~                 3.83                 3.84                4.15
## 6 "Prefer not to ~                 3.21                 3.19                4.07
## 7 "Some college, ~                 3.70                 3.71                4.29
## 8 "Some high scho~                 3.67                 3.77                4.09
## # i 2 more variables: eq_pa_Mean_Response <dbl>, equity_Mean_Response <dbl>
```

```
survey %>%
  group_by(role) %>%
  summarise_at(vars(eq_his, eq_hin, eq_as, eq_pa, equity),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 8 x 6
##   role             eq_his_Mean_Response eq_hin_Mean_Response eq_as_Mean_Response
##   <chr>                           <dbl>                <dbl>               <dbl>
## 1 ""                               2.95                 3.18                3.81
## 2 "Central office~                 3.77                 4.03                4.17
## 3 "Community memb~                 3.35                 3.48                4.18
## 4 "Community part~                 4.18                 4.49                4.58
## 5 "Multiple roles"                 3.69                 3.84                4.29
## 6 "Parent/Caregiv~                 3.19                 3.33                4.15
## 7 "School site st~                 3.97                 4.10                4.32
## 8 "Student"                        3.45                 3.64                3.99
## # i 2 more variables: eq_pa_Mean_Response <dbl>, equity_Mean_Response <dbl>
```

Excellence

```
survey %>%
  group_by(race) %>%
  summarise_at(vars(ex_ap, ex_scc, ex_sel, ex_sd, ex_att, excellence),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 10 x 7
##    race             ex_ap_Mean_Response ex_scc_Mean_Response ex_sel_Mean_Response
##    <chr>                          <dbl>                <dbl>               <dbl>
##  1 ""                              4.20                 4.36                4.12
##  2 "African Ameri~                 3.88                 4.16                3.99
##  3 "American Indi~                 3.83                 3.89                4.11
##  4 "Asian"                         4.32                 4.50                4.34
##  5 "Decline to St~                 4.36                 4.42                4.01
##  6 "Filipino"                      4.31                 4.60                4.43
##  7 "Hispanic/Lati~                 4.17                 4.37                4.29
##  8 "Pacific Islan~                 4.16                 4.53                4.24
##  9 "Two or More R~                 4.19                 4.43                4.15
## 10 "White"                         3.99                 4.37                4.12
## # i 3 more variables: ex_sd_Mean_Response <dbl>, ex_att_Mean_Response <dbl>,
## #   excellence_Mean_Response <dbl>
```

```
survey %>%
  group_by(zip) %>%
  summarise_at(vars(ex_ap, ex_scc, ex_sel, ex_sd, ex_att, excellence),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 136 x 7
##      zip ex_ap_Mean_Response ex_scc_Mean_Response ex_sel_Mean_Response
##    <int>               <dbl>                <dbl>               <dbl>
##  1     0                 NaN                  NaN                 NaN
##  2  4102                   4                    3                   3
```

```
##  3  8327                NaN                NaN                NaN
##  4 11221                  5                  3                  3
##  5 53219                NaN                NaN                NaN
##  6 76100                  3                  5                  5
##  7 84107                  5                  5                  5
##  8 84108                  4                  5                  5
##  9 84114                NaN                NaN                NaN
## 10 84115                  5                  4                  4
## # i 126 more rows
## # i 3 more variables: ex_sd_Mean_Response <dbl>, ex_att_Mean_Response <dbl>,
## #   excellence_Mean_Response <dbl>
```

```
survey %>%
  group_by(education) %>%
  summarise_at(vars(ex_ap, ex_scc, ex_sel, ex_sd, ex_att, excellence),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 8 x 7
##   education        ex_ap_Mean_Response ex_scc_Mean_Response ex_sel_Mean_Response
##   <chr>                          <dbl>                <dbl>                <dbl>
## 1 ""                              4.53                 4.53                 4.31
## 2 "Associates or ~                4.34                 4.5                  4.35
## 3 "Bachelor's deg~                4.17                 4.42                 4.18
## 4 "Graduate or pr~                4.09                 4.39                 4.15
## 5 "High school di~                4.30                 4.36                 4.32
## 6 "Prefer not to ~                4.34                 4.41                 4.18
## 7 "Some college, ~                4.29                 4.55                 4.34
## 8 "Some high scho~                4.02                 4.26                 4.20
## # i 3 more variables: ex_sd_Mean_Response <dbl>, ex_att_Mean_Response <dbl>,
## #   excellence_Mean_Response <dbl>
```

```
survey %>%
  group_by(role) %>%
  summarise_at(vars(ex_ap, ex_scc, ex_sel, ex_sd, ex_att, excellence),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 8 x 7
##   role             ex_ap_Mean_Response ex_scc_Mean_Response ex_sel_Mean_Response
##   <chr>                          <dbl>                <dbl>                <dbl>
## 1 ""                              4                    4.25                 3.87
## 2 "Central office~                3.53                 4.34                 4.1
## 3 "Community memb~                4.10                 4.38                 3.84
## 4 "Community part~                3.53                 4.47                 4.28
## 5 "Multiple roles"                3.92                 4.34                 4.18
## 6 "Parent/Caregiv~                4.30                 4.46                 4.23
## 7 "School site st~                3.63                 4.18                 4.05
## 8 "Student"                       3.86                 4.15                 3.91
## # i 3 more variables: ex_sd_Mean_Response <dbl>, ex_att_Mean_Response <dbl>,
## #   excellence_Mean_Response <dbl>
```

Efficiency

```
survey %>%
  group_by(race) %>%
  summarise_at(vars(er_se, er_fc, er_bu, er_bc, er_cpp, er_ocs, er_tt, efficiency),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 10 x 9
##    race              er_se_Mean_Response er_fc_Mean_Response er_bu_Mean_Response
##    <chr>                           <dbl>               <dbl>               <dbl>
##  1 ""                               3.82                4.09                3.66
##  2 "African America~                3.65                3.82                3.60
##  3 "American Indian~                3.38                4                   3.62
##  4 "Asian"                          3.94                4.24                3.80
##  5 "Decline to Stat~                3.90                4.19                3.61
##  6 "Filipino"                       3.99                4.26                3.92
##  7 "Hispanic/Latino"                3.96                4.00                3.90
##  8 "Pacific Islande~                3.86                4.16                3.89
##  9 "Two or More Rac~                3.80                4.02                3.52
## 10 "White"                          3.75                3.92                3.41
## # i 5 more variables: er_bc_Mean_Response <dbl>, er_cpp_Mean_Response <dbl>,
## #   er_ocs_Mean_Response <dbl>, er_tt_Mean_Response <dbl>,
## #   efficiency_Mean_Response <dbl>
```

```
survey %>%
  group_by(zip) %>%
  summarise_at(vars(er_se, er_fc, er_bu, er_bc, er_cpp, er_ocs, er_tt, efficiency),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 136 x 9
##      zip er_se_Mean_Response er_fc_Mean_Response er_bu_Mean_Response
##    <int>               <dbl>               <dbl>               <dbl>
##  1     0                 NaN                 NaN                 NaN
##  2  4102                   3                   4                   3
##  3  8327                 NaN                 NaN                 NaN
##  4 11221                 NaN                 NaN                 NaN
##  5 53219                 NaN                 NaN                 NaN
##  6 76100                   5                   3                   4
##  7 84107                   5                   5                   5
##  8 84108                   5                   5                   5
##  9 84114                 NaN                 NaN                 NaN
## 10 84115                   4                   3                   3
## # i 126 more rows
## # i 5 more variables: er_bc_Mean_Response <dbl>, er_cpp_Mean_Response <dbl>,
## #   er_ocs_Mean_Response <dbl>, er_tt_Mean_Response <dbl>,
## #   efficiency_Mean_Response <dbl>
```

```
survey %>%
  group_by(education) %>%
  summarise_at(vars(er_se, er_fc, er_bu, er_bc, er_cpp, er_ocs, er_tt, efficiency),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 8 x 9
##   education          er_se_Mean_Response er_fc_Mean_Response er_bu_Mean_Response
```

```
##   <chr>                            <dbl>               <dbl>               <dbl>
## 1 ""                                3.95                3.89                3.82
## 2 "Associates or te~                3.92                4.23                3.86
## 3 "Bachelor's degre~                3.80                4.06                3.59
## 4 "Graduate or prof~                3.78                3.97                3.48
## 5 "High school dipl~                4.19                4.29                4.09
## 6 "Prefer not to sa~                4                   4.32                3.88
## 7 "Some college, bu~                3.96                4.23                4.03
## 8 "Some high school~                3.96                3.91                3.89
## # i 5 more variables: er_bc_Mean_Response <dbl>, er_cpp_Mean_Response <dbl>,
## #   er_ocs_Mean_Response <dbl>, er_tt_Mean_Response <dbl>,
## #   efficiency_Mean_Response <dbl>
```

```
survey %>%
  group_by(role) %>%
  summarise_at(vars(er_se, er_fc, er_bu, er_bc, er_cpp, er_ocs, er_tt, efficiency),
  list(Mean_Response = mean), na.rm = TRUE)
```

```
## # A tibble: 8 x 9
##   role               er_se_Mean_Response er_fc_Mean_Response er_bu_Mean_Response
##   <chr>                            <dbl>               <dbl>               <dbl>
## 1 ""                                3.79                4.06                3.76
## 2 "Central office s~                3.53                3.23                3.83
## 3 "Community member"                3.73                3.71                3.40
## 4 "Community partne~                3.37                3.72                3.72
## 5 "Multiple roles"                  3.72                3.92                3.54
## 6 "Parent/Caregiver"                3.91                4.17                3.67
## 7 "School site staf~                3.71                3.68                3.59
## 8 "Student"                         3.55                3.59                3.54
## # i 5 more variables: er_bc_Mean_Response <dbl>, er_cpp_Mean_Response <dbl>,
## #   er_ocs_Mean_Response <dbl>, er_tt_Mean_Response <dbl>,
## #   efficiency_Mean_Response <dbl>
```

#Format the written answers

```
unique_equity_sugg <- distinct(survey, equity_sugg)
```