Lab 4

Math 241, Week 4

```
# Put all necessary libraries here
library(tidyverse)
#install.packages("ggrepel")
library(ggrepel)
```

Problem 1: COVID survey - interpretation

There are some statements that have general agreement across the board, most people seemed to agree that they would recommend the vaccine to others and are confident in the vetting process for new vaccines - except those who did not get the vaccine themselves. This feels intuitive. Indigenous people also appear more hesitant to trust the vaccines, which is also unsurprising given the historical context of the medical system's role in colonization. Something surprising from this data is that even though people tend to lean towards trusting the vaccines, there is a lot more concern regarding the safety and side effects of the vaccine than expected. I would be curious to learn more about the group that prefer not to say a gender identity as they also seemed particularly skeptical of the vaccine.

Problem 2: COVID survey - reconstruct

```
library(readr)
library(dplyr)
library(tidyverse)
library(knitr)
library(ggplot2)
library(moderndive)

#loading dataset
covid_survey <- read_csv("~/Rstudio/DataScience/math241repo/labs/lab04/data/covid-survey.csv", skip = 1)

#recoding
covid_survey <- covid_survey %>%
  filter(if_any(-response_id, ~ !is.na(.))) %>%
  mutate(exp_already_vax = recode(exp_already_vax,
                                  "1" = "Yes",
                                  "0" = "No"),
         exp_flu_vax = recode(exp_flu_vax,
                                  "1" = "Yes",
                                  "0" = "No"),
         exp_profession = recode(exp_profession,
                                  "1" = "Nursing",
                                  "0" = "Medical"),
         exp_gender = recode(exp_gender,
```

```
                              "0" = "Male",
                              "1" = "Female",
                              "3" = "Non-binary",
                              "4" = "Prefer not to say"),
         exp_race = recode(exp_race,
                              "2" = "Asian",
                              "1" = "American Indian / Alaskan Native",
                              "3" = "Black / African American",
                              "4" = "Native Hawaiian / Other Pacific Islander",
                              "5" = "White"),
          exp_ethnicity = recode(exp_ethnicity,
                              "1" = "Hispanic / Latino",
                              "2" = "Non-Hispanic/Non-Latino"),
         exp_age_bin = recode(exp_age_bin,
                              case_when(
                                exp_age_bin <20 ~ "<20",
                                exp_age_bin >= 20 & exp_age_bin <= 25 ~ "21-25",
                                exp_age_bin >= 26 & exp_age_bin <= 30 ~ "26-30",
                                exp_age_bin > 30 ~ ">30"
                              ))
  )


covid_survey_longer <- covid_survey %>%
  #this turns the explanatory columns into one column with each option as a variable (as opposed to a s
  pivot_longer(
    cols = starts_with("exp_"),
    names_to = "explanatory",
    values_to = "explanatory_value"
    ) %>%
  filter(!is.na(explanatory_value)) %>%
  #this does the same thing but with response columns, compiling them into one column called response
  pivot_longer(
    cols = starts_with("resp_"),
    names_to = "response",
    values_to = "response_value"
)
covid_survey_longer
```

```
## # A tibble: 37,524 x 5
##    response_id explanatory    explanatory_value response          response_value
##          <dbl> <chr>          <chr>             <chr>                     <dbl>
##  1           1 exp_profession Nursing           resp_safety                   5
##  2           1 exp_profession Nursing           resp_confidence_~             2
##  3           1 exp_profession Nursing           resp_concern_saf~             2
##  4           1 exp_profession Nursing           resp_feel_safe_a~             1
##  5           1 exp_profession Nursing           resp_will_recomm~             1
##  6           1 exp_profession Nursing           resp_trust_info               1
##  7           1 exp_flu_vax    Yes               resp_safety                   5
##  8           1 exp_flu_vax    Yes               resp_confidence_~             2
##  9           1 exp_flu_vax    Yes               resp_concern_saf~             2
## 10           1 exp_flu_vax    Yes               resp_feel_safe_a~             1
## # i 37,514 more rows
```

```r
#calculating mean, 10th percentile and 90th percentile
covid_survey_summary_stats_by_group <- covid_survey_longer %>%
  group_by(explanatory, explanatory_value, response) %>%
  summarize(
    mean = mean(response_value, na.rm = TRUE),
    low = quantile(response_value, 0.10, na.rm = TRUE),
    high = quantile(response_value, 0.90, na.rm = TRUE)
  )
covid_survey_summary_stats_by_group
```

```
## # A tibble: 102 x 6
## # Groups:   explanatory, explanatory_value [17]
##    explanatory     explanatory_value response                mean   low  high
##    <chr>           <chr>             <chr>                  <dbl> <dbl> <dbl>
##  1 exp_already_vax No                resp_concern_safety     2.29     1     4
##  2 exp_already_vax No                resp_confidence_science 3.19     1     5
##  3 exp_already_vax No                resp_feel_safe_at_work  3.80     2     5
##  4 exp_already_vax No                resp_safety             2.86     1     5
##  5 exp_already_vax No                resp_trust_info         3.07     1     5
##  6 exp_already_vax No                resp_will_recommend     3        1     5
##  7 exp_already_vax Yes               resp_concern_safety     3.34     1     5
##  8 exp_already_vax Yes               resp_confidence_science 1.33     1     2
##  9 exp_already_vax Yes               resp_feel_safe_at_work  1.21     1     2
## 10 exp_already_vax Yes               resp_safety             1.98     1     5
## # i 92 more rows
```

```r
covid_survey_summary_stats_all <- covid_survey_longer %>%
  group_by(response) %>%
  summarize(
    mean = mean(response_value, na.rm = TRUE),
    low = quantile(response_value, 0.10, na.rm = TRUE),
    high = quantile(response_value, 0.90, na.rm = TRUE)
  )
covid_survey_summary_stats_all
```

```
## # A tibble: 6 x 4
##   response                mean   low  high
##   <chr>                  <dbl> <dbl> <dbl>
## 1 resp_concern_safety     3.28     1     5
## 2 resp_confidence_science 1.43     1     2
## 3 resp_feel_safe_at_work  1.36     1     2
## 4 resp_safety             2.03     1     5
## 5 resp_trust_info         1.40     1     2
## 6 resp_will_recommend     1.21     1     2
```

```r
covid_survey_summary_stats <- bind_rows(
  covid_survey_summary_stats_all,
  covid_survey_summary_stats_by_group
)
covid_survey_summary_stats
```

```
## # A tibble: 108 x 6
```

```
##    response                 mean   low  high explanatory     explanatory_value
##    <chr>                   <dbl> <dbl> <dbl> <chr>           <chr>
##  1 resp_concern_safety      3.28     1     5 <NA>            <NA>
##  2 resp_confidence_science  1.43     1     2 <NA>            <NA>
##  3 resp_feel_safe_at_work   1.36     1     2 <NA>            <NA>
##  4 resp_safety              2.03     1     5 <NA>            <NA>
##  5 resp_trust_info          1.40     1     2 <NA>            <NA>
##  6 resp_will_recommend      1.21     1     2 <NA>            <NA>
##  7 resp_concern_safety      2.29     1     4 exp_already_vax No
##  8 resp_confidence_science  3.19     1     5 exp_already_vax No
##  9 resp_feel_safe_at_work   3.80     2     5 exp_already_vax No
## 10 resp_safety              2.86     1     5 exp_already_vax No
## # i 98 more rows
```

```
ggplot(data = covid_survey_summary_stats %>%
         mutate(explanatory = factor(explanatory,
                                     levels = c("NA", "exp_gender", "exp_race", "exp_ethnicity", "exp_pr
                                     labels = c("All", "Gender", "Race", "Ethnicity", "Profession", "Hac
         aes(x=mean, y = explanatory_value, xmin= low, xmax = high, color = explanatory_value)) +
  geom_errorbarh(position = position_dodge(width = 1)) +
  geom_point() +
  facet_grid(explanatory~response,
             labeller = labeller(response = c("resp_concern_safety" = "I am concerned
about the safety
and side effects of
the vaccine",
                              "resp_confidence_science" = "I am confident in
the scientific
vetting process for
the new COVID
vaccines",
                              "resp_feel_safe_at_work" = "Getting the vaccine
will make me feel
safer at work",
                              "resp_safety" = "Based on my
understanding, I
believe the vaccine
is safe",
                              "resp_trust_info" = "I trust the
information that I
have recieved about
the vaccines",
                              "resp_will_recommend" = "I will recommend
the vaccine to
family, friends, and
community members"
             )),
             scales = "free") +
  theme_minimal() +
  theme(legend.position = "none",
        strip.background = element_rect(color = "black", fill = "gray90"),
        strip.text.x.top = element_text(size = 12, hjust = 0.5, face = "bold")
        ) +
  labs(x = "Mean Likert Score
```

```
      (Error bars range from 10th to 90th percentile)",
      y = "")
```