

Lab 4

Math 241, Week 4

```
# Put all necessary libraries here  
library(tidyverse)
```

Problem 1: COVID survey - interpretation

This graph seems to include a lot of data from a COVID survey, faceted across many different identities and separated by question responses. The way it is structured allows us to see some interesting situations where some specific explanatory variables (demographics) respond a specific way to some questions.

We can see a trend in regards to age, where the sub 20 group is often significantly different. We also notice specifically that there is not much of a difference between 21-25 and 26-30 groups. A specific example that was intuitive was that older demographics are more hesitant and less trusting than younger people. This might be intuitive due to lived experience and how recent it was. It's interesting how the 21-25 group thought the vaccine is unsafe, but would recommend it to family and friends and others as I would have not guessed that.

It's also pretty interesting how people who had the COVID vaccine and the flu vaccine had similar responses to the questions, which tells me that there is possibly some relationship between these thoughts. I remember people being worried about this specific COVID vaccine as people thought it was accelerated, which made it "unsafe" and it's interesting seeing that logic doesn't apply much according to this data.

Another thing I noticed was the lack of differences between some of the races listed, but not all of them. I wonder why people who marked down White, Black, and Asian had no drastic differences across all 6 questions, but if we compare it to Native Hawaiian or Pacific Islander or American Indian / Alaskan Native, we see a drastic change.

I also did not expect people under nursing to choose strongly disagree for the safety of the vaccine, but also expected that medicine would be the same or at most higher by one mark.

Problem 2: COVID survey - reconstruct

```
covid_survey <- read_csv("data/covid-survey.csv", skip = 1)  
dim(covid_survey)  
  
## [1] 1121 14  
  
covid_survey <- covid_survey %>% filter(if_any(-response_id, ~ !is.na(.)))  
dim(covid_survey)  
  
## [1] 1111 14  
  
covid_survey$exp_already_vax <- case_when(  
  covid_survey$exp_already_vax == 0 ~ "No",  
  covid_survey$exp_already_vax == 1 ~ "Yes"  
)  
  
covid_survey$exp_flu_vax <- case_when(  
  covid_survey$exp_flu_vax == 0 ~ "No",
```

```

  covid_survey$exp_flu_vax == 1 ~ "Yes"
)

covid_survey$exp_profession <- case_when(
  covid_survey$exp_profession == 0 ~ "Medical",
  covid_survey$exp_profession == 1 ~ "Nursing"
)

covid_survey$exp_gender <- case_when(
  covid_survey$exp_gender == 0 ~ "Male",
  covid_survey$exp_gender == 1 ~ "Female",
  covid_survey$exp_gender == 3 ~ "Non-binary third gender",
  covid_survey$exp_gender == 4 ~ "Prefer not to say"
)

covid_survey$exp_race <- case_when(
  covid_survey$exp_race == 1 ~ "American Indian / Alaskan Native",
  covid_survey$exp_race == 2 ~ "Asian",
  covid_survey$exp_race == 3 ~ "Black / African American",
  covid_survey$exp_race == 4 ~ "Native Hawaiian / Other Pacific Islander",
  covid_survey$exp_race == 5 ~ "White"
)

covid_survey$exp_ethnicity <- case_when(
  covid_survey$exp_ethnicity == 1 ~ "Hispanic / Latino",
  covid_survey$exp_ethnicity == 2 ~ "Non-Hispanic/Non-Latino"
)

covid_survey$exp_age_bin <- case_when(
  covid_survey$exp_age_bin == 0 ~ "<20",
  covid_survey$exp_age_bin == 20 ~ "21-25",
  covid_survey$exp_age_bin == 25 ~ "26-30",
  covid_survey$exp_age_bin == 30 ~ ">30",
)

```

In this section, the first pivot longer aims to convert the wide structure of the data to a long one that's mostly row based. It gets all the explanatory variables (the ones that start with exp_) and places them as rows with their value as a new column. In the middle, filter removes any NA values. The second one does something similar but with response variables, where it gathers them all and assigns them response values.

```

covid_survey_longer <- covid_survey %>%
  pivot_longer(
    cols = starts_with("exp_"),
    names_to = "explanatory",
    values_to = "explanatory_value") %>%
  filter(!is.na(explanatory_value)) %>%
  pivot_longer(
    cols = starts_with("resp_"),
    names_to = "response",
    values_to = "response_value")

covid_survey_longer

```

```
## # A tibble: 43,428 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 43,418 more rows
```

```
covid_survey_summary_stats_by_group <- covid_survey_longer %>% group_by(explanatory, explanatory_value,
covid_survey_summary_stats_by_group
```

```
## # A tibble: 126 x 6
## # Groups:   explanatory, explanatory_value [21]
##      explanatory explanatory_value response      mean    low  high
##      <chr>          <chr>          <chr>    <dbl> <dbl> <dbl>
## 1 exp_age_bin 21-25      resp_concern_safety 3.32    2    5
## 2 exp_age_bin 21-25      resp_confidence_science 1.31    1    2
## 3 exp_age_bin 21-25      resp_feel_safe_at_work 1.20    1    2
## 4 exp_age_bin 21-25      resp_safety 1.95    1    5
## 5 exp_age_bin 21-25      resp_trust_info 1.29    1    2
## 6 exp_age_bin 21-25      resp_will_recommend 1.09    1    1
## 7 exp_age_bin 26-30      resp_concern_safety 3.35    1    5
## 8 exp_age_bin 26-30      resp_confidence_science 1.40    1    2
## 9 exp_age_bin 26-30      resp_feel_safe_at_work 1.29    1    2
## 10 exp_age_bin 26-30      resp_safety 2.16    1    5
## # i 116 more rows
```

```
covid_survey_summary_stats_all <- covid_survey_longer %>% group_by(response) %>% summarize(mean = mean(
covid_survey_summary_stats_all
```

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

```
covid_survey_summary_stats <- bind_rows(covid_survey_summary_stats_all, covid_survey_summary_stats_by_g
covid_survey_summary_stats
```

```
## # A tibble: 132 x 6
##      response      mean    low  high explanatory explanatory_value
##      <chr>    <dbl> <dbl> <dbl> <chr>          <chr>
## 1 resp_concern_safety 3.28    1    5 All          ""
## 2 resp_confidence_science 1.43    1    2 All          ""
```

```
## 3 resp_feel_safe_at_work 1.36 1 2 All ""
## 4 resp_safety 2.03 1 5 All ""
## 5 resp_trust_info 1.40 1 2 All ""
## 6 resp_will_recommend 1.21 1 2 All ""
## 7 resp_concern_safety 3.32 2 5 exp_age_bin "21-25"
## 8 resp_confidence_science 1.31 1 2 exp_age_bin "21-25"
## 9 resp_feel_safe_at_work 1.20 1 2 exp_age_bin "21-25"
## 10 resp_safety 1.95 1 5 exp_age_bin "21-25"
## # i 122 more rows
```

```
covid_survey_summary_stats$response[covid_survey_summary_stats$response == "resp_safety"] <- "Based on m
covid_survey_summary_stats$response[covid_survey_summary_stats$response == "resp_confidence_science"] <-
covid_survey_summary_stats$response[covid_survey_summary_stats$response == "resp_feel_safe_at_work"] <-
covid_survey_summary_stats$response[covid_survey_summary_stats$response == "resp_will_recommend"] <- "I
covid_survey_summary_stats$response[covid_survey_summary_stats$response == "resp_trust_info"] <- "I tru
covid_survey_summary_stats$response[covid_survey_summary_stats$response == "resp_concern_safety"] <- "I
```

```
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_age_bin"] <- "Age
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_gender"] <- "Gend
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_race"] <- "Race"
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_ethnicity"] <- "E
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_profession"] <- "I
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_already_vax"] <-
covid_survey_summary_stats$explanatory[covid_survey_summary_stats$explanatory == "exp_flu_vax"] <- "Had
```

```
ggplot(data = covid_survey_summary_stats, mapping = aes(mean, explanatory_value, xmin = low, xmax = high)) +
  geom_errorbarh(height = 0.1) +
  geom_point() +
  facet_grid(vars(factor(explanatory, levels = c("All", "Age", "Gender", "Race", "Ethnicity", "Professi
  theme_minimal() +
  theme(strip.background = element_rect(colour="black",
                                         fill="gray90"),
        axis.title.y=element_blank(),
        strip.text.y = element_text(angle = 0),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())) +
  labs(x = "Mean Likert score\n(Error bars range from 10th to 90th percentile)")
```

