

Lab 4

Math 241, Week 4

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

Problem 1: COVID survey - interpretation

This graphic is quite difficult to interpret as there are many variables with relatively large error bars and thus there is a lot of variability within the data.

Some interpretations: 1. Individuals age 30 and under tended to agree more with the statement “I trust the information that I have received about the vaccines” in comparison to those over the age of 30.

2. Regardless of gender, individuals on average agreed with the statement “Getting the vaccine will make me feel safer at work”
3. Individuals in the medical field on average agreed the most with the statement “Based on my understanding, I believe the vaccine is safe”

Problem 2: COVID survey - reconstruct

```
covid_survey <- read_csv("data/covid-survey.csv")
```

```
covid_survey <- read_csv("data/covid-survey.csv")
covid_survey_clean <- covid_survey %>%
  filter(rowSums(!is.na(.[-1])) > 0)
dim(covid_survey_clean)
```

```
## [1] 1112 14
```

```
colnames(covid_survey_clean) <- covid_survey_clean[1, ]
covid_survey_rmv <- covid_survey_clean[-1, ]
```

```
covid_survey_relab <- covid_survey_rmv %>%
  mutate(exp_already_vax = case_when(
    exp_already_vax == 0 ~ "No",
    exp_already_vax == 1 ~ "Yes"),
  exp_flu_vax = case_when(
    exp_flu_vax == 0 ~ "No",
    exp_flu_vax == 1 ~ "Yes"),
  exp_profession = case_when(
```

```

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

dim(covid_survey_relab)

```

```
## [1] 1111 14
```

```

covid_survey_longer <- covid_survey_relab %>%
  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
	<chr>	<chr>	<chr>	<chr>	<chr>
##	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
```

EXPLAINNNNNNNNNNNNNNNNNNNnnnnnnnsfower8r23tu[90t24309uo]

```
covid_survey_summary_stats_by_group <- covid_survey_longer %>%
group_by(explanatory, explanatory_value, response) %>%
summarise(mean = mean(as.numeric(response_value), na.rm = TRUE),
  low = quantile(as.numeric(response_value, na.rm = TRUE), 0.1, na.rm = TRUE),
  high = quantile(as.numeric(response_value, na.rm = TRUE), 0.9, na.rm = TRUE)
)
```

```
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) %>%
summarise(mean = mean(as.numeric(response_value), na.rm = TRUE),
  low = quantile(as.numeric(response_value, na.rm = TRUE), 0.1, na.rm = TRUE),
  high = quantile(as.numeric(response_value, na.rm = TRUE), 0.9, 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
```

```
covid_survey_summary_stats <- bind_rows(covid_survey_summary_stats_all, covid_survey_summary_stats_by_group)
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 <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 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 <- covid_survey_summary_stats %>%
  mutate(
    response = case_when(
      response == "resp_safety" ~ "Based on my understanding, I believe the vaccine is safe",
      response == "resp_confidence_science" ~ "I am confident in the scientific vetting process for the vaccine",
      response == "resp_feel_safe_at_work" ~ "Getting the vaccine will make me feel safer at work",
      response == "resp_will_recommend" ~ "I will recommend the vaccine to family, friends, and community",
      response == "resp_trust_info" ~ "I trust the information that I have received about the vaccines",
      response == "resp_concern_safety" ~ "I am concerned about the safety and side effects of the vaccine"
    )
  )
  mutate(
    explanatory = case_when(
      explanatory == "exp_age_bin" ~ "Age",
      explanatory == "exp_gender" ~ "Gender",
      explanatory == "exp_race" ~ "Race",
      explanatory == "exp_ethnicity" ~ "Ethnicity",
      explanatory == "exp_profession" ~ "Profession",
      explanatory == "exp_already_vax" ~ "Had COVID vaccine",
      explanatory == "exp_flu_vax" ~ "Had flu vaccine this year"
    )
  )
```

```
covid_survey_summary_stats %>%
  na.omit() %>%
  ggplot(aes(x = mean, y = explanatory_value)) +
  geom_point() +
  geom_errorbar(aes(xmin = low, xmax = high)) +
  facet_grid(rows = vars(explanatory), cols = vars(response), scales = 'free', labeller = label_wrap_gen(2))
  labs(
    title = NULL,
    x = "Mean Likert score \n (Error bars range from 10th to 90th percentile)",
    y = NULL
  )
```

