

Lab 4

Math 241, Week 4

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

Problem 1: COVID survey - interpretation

Variability Across Groups: There appears to be significant variability in attitudes towards vaccination across different explanatory variables (e.g., year of study, specialty interest). This suggests that there might be diverse opinions within the medical and nursing student community, possibly influenced by factors such as clinical exposure and personal beliefs.

General Agreement on Vaccination Importance: The points indicating mean Likert scores for statements related to the importance of vaccination and trust in vaccine efficacy seem to lean towards agreement (1 - Strongly Agree, 2 - Somewhat Agree), especially in the “All” category. This aligns with the expectation that future healthcare professionals recognize the value of vaccines in public health.

Concerns and Disagreements: For some statements, particularly those related to vaccine mandates or potential side effects, the wider spread of error bars and higher mean scores (possibly indicating disagreement or neutrality) suggest that there are concerns and reservations among students. This is an interesting find considering the overall pro-vaccine sentiment and could reflect the complex nature of decision-making in healthcare.

Problem 2: COVID survey - reconstruct

```
covid_survey <- read_csv("~/Desktop/DataScience/math241/labs/lab04/data/covid-survey.csv", skip = 1) #s  
  
covid_survey_clean <- covid_survey %>%  
  filter(rowSums(!is.na(.[1])) > 0)  
#removed rows where all values aside from response_id are NA  
  
covid_survey_relabeled <- covid_survey_clean %>%  
  mutate(  
    exp_already_vax = recode(exp_already_vax, '0' = 'No', '1' = 'Yes'),  
    exp_profession = recode(exp_profession, '0' = 'Medical', '1' = 'Nursing'),  
    exp_gender = recode(exp_gender, '0' = 'Male', '1' = 'Female', '3' = 'Non-binary third gender', '4' = 'Other'),  
    exp_race = recode(exp_race, '1' = 'American Indian / Alaskan Native', '2' = 'Asian', '3' = 'Black / African American', '4' = 'Hispanic / Latino', '5' = 'Other'),  
    exp_ethnicity = recode(exp_ethnicity, '1' = 'Hispanic / Latino', '2' = 'Non-Hispanic/Non-Latino'),  
    exp_age_bin = recode(exp_age_bin, '0' = '<20', '20' = '21-25', '25' = '26-30', '30' = '>30')  
  )  
#converts numerical codes to their corresponding categories based on guidelines
```

```
covid_survey_longer <- covid_survey_relabeled %>%
  mutate(across(starts_with("exp_"), as.character)) %>% # Convert all exp_ columns to character
  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"
  )
#pivots the data for analysis

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 1 resp_safety 5
## 8 1 exp_flu_vax 1 resp_confidence_~ 2
## 9 1 exp_flu_vax 1 resp_concern_saf~ 2
## 10 1 exp_flu_vax 1 resp_feel_safe_a~ 1
## # i 43,418 more rows
```

The first `pivot_longer()` statement in the code chunk above is taking multiple columns whose names start with “exp_” (which represent different explanatory variables) and converting them into a long format. In the resulting two-column layout, one column (`explanatory`) contains the names of the original explanatory variables, and the other column (`explanatory_value`) contains the corresponding values.

The second `pivot_longer()` statement does the same transformation for columns starting with “resp_” (which represent different response variables from the survey). These columns are also converted into a long format, resulting in two columns: `response`, containing the names of the original response variables, and `response_value`, containing the corresponding values.

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

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
```

```
library(ggplot2)
```

```
# Define the custom labels for the explanatory and response variables
```

```
explanatory_variable_labels <- c(
```

```
  "exp_age_bin" = "Age",
  "exp_already_vax" = "Had COVID vaccine",
  "exp_ethnicity" = "Ethnicity",
  "exp_flu_vax" = "Had flu vaccine this year",
  "exp_gender" = "Gender",
  "exp_profession" = "Profession",
  "exp_race" = "Race"
```

```
)
```

```
response_variable_labels <- c(
```

```
  "resp_safety" = "Based on my understanding, I\ndon't believe the vaccine is safe",
  "resp_feel_safe_at_work" = "Getting the vaccine will\nmake me feel safer at work",
  "resp_concern_safety" = "I am concerned about the safety\nand side effects of the vaccine",
  "resp_confidence_science" = "I am confident in the scientific\nvetting process for the new COVID vaccine",
  "resp_trust_info" = "I trust the information that I have\nreceived about the vaccines",
  "resp_will_recommend" = "I will recommend the vaccine to\nfamily, friends, and community members"
```

```
)
```

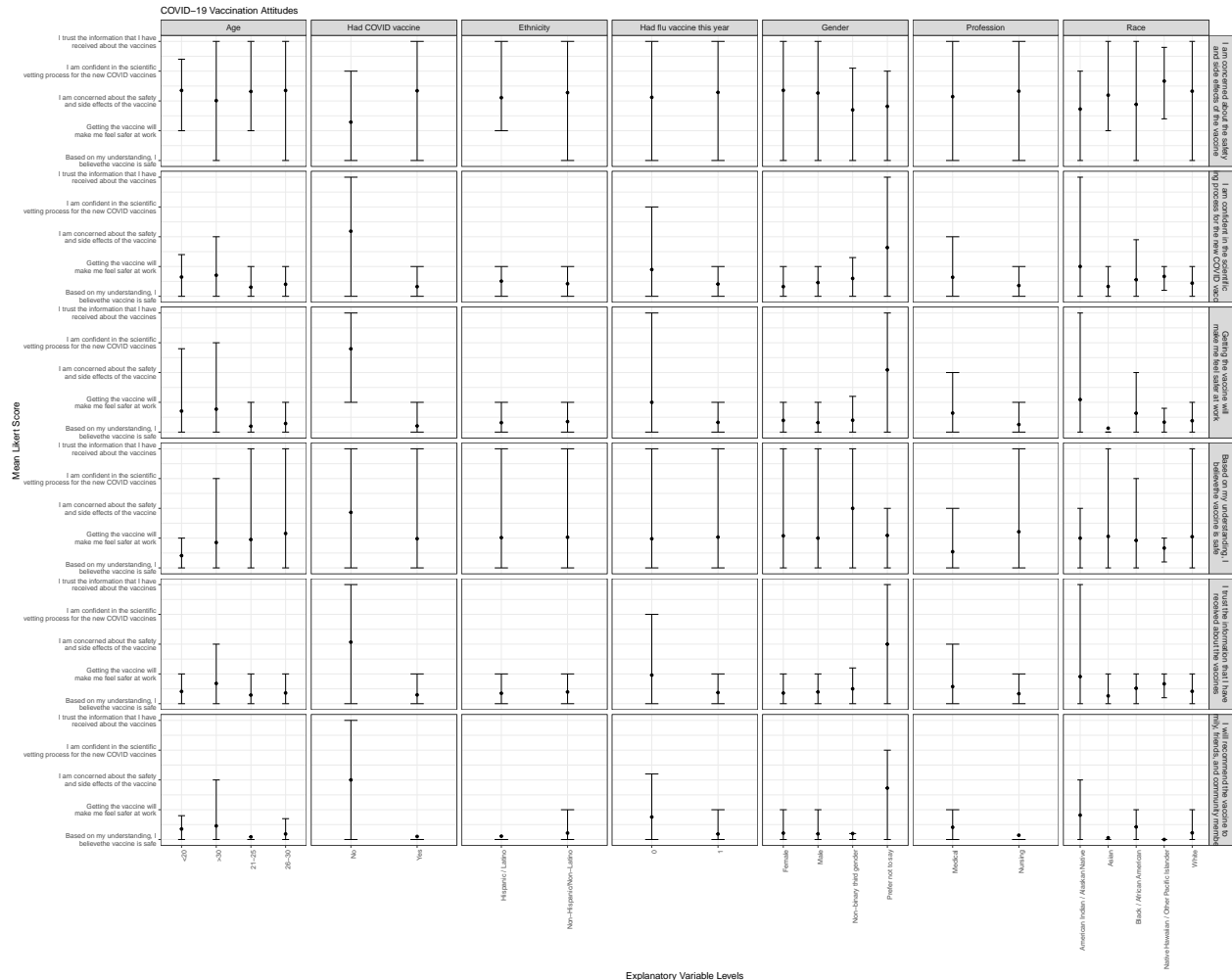
```
# Update the ggplot code
```

```
ggplot(covid_survey_summary_stats_by_group, aes(x = explanatory_value, y = mean, group = response)) +
  geom_errorbar(aes(ymin = low, ymax = high), width = 0.2) +
  geom_point() +
  facet_grid(response ~ explanatory, scales = "free_x", labeller = labeller(
    response = as_labeller(response_variable_labels),
    explanatory = as_labeller(explanatory_variable_labels)
  )) +
  theme_bw(base_size = 14) +
  theme(
    text = element_text(size = 12),
    axis.title = element_text(size = 14),
    axis.text.x = element_text(angle = 90, hjust = 1, size = 10),
    axis.text.y = element_text(size = 10),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
```

```

strip.text = element_text(size = 12)
) +
labs(
  x = "Explanatory Variable Levels",
  y = "Mean Likert Score",
  title = "COVID-19 Vaccination Attitudes"
) +
scale_x_discrete(labels = explanatory_variable_labels) +
scale_y_continuous(labels = function(x) response_variable_labels[x])

```



Notes: I had some trouble with the plot, on the left side we can see that the prompts from the questionnaire are repeated and I couldn't quite figure out how to get rid of those. Also, the axes are flipped since I had a hard time fitting all the text into the plot when the axes were on their correct sides.