

Lab 4

Cameron Adams

Math 241, Week 4

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

Problem 1: COVID survey - interpretation

Off the bat one of the biggest differences in opinions is correlated with if the medical and nursing students have had or not had the Covid vaccine. If the students had the Vaccine, they more likely to agree or strongly agree making their data more skewed right. However if a student hadn't had the vaccine, they would often answer with Neither Agree Nor Disagree. The exception to this was when asked if they were concerned about the side effects of the vaccine, where vaccinated students answered on average Neither Agree Nor Disagree and Non-vaccinated students agreed more. Another interesting question was when asked if the students would recommend the vaccine to family and friends. Vaccinated students overwhelmingly agreed with at least 80% of participants answering strongly agree. whereas the non-vaccinated students ranged from all answered averaging Neither Agree Nor Disagree. One thing that's important to take into account however is the All category. this suggests that overall there are much less non vaccinated students than vaccinated students since the data is still skewed right or much more in line with the vaccinated students.

There was also a similar correlation with if the students have had the Flu Vaccine. Students that took the Flue vaccine answered similar to students who had had the covid vaccine, showing similar or identical statistics. Students who didn't have the flu vaccine did seem to trust it a bit more than the Covid vaccine but did show similar trends as the question before.

Some other interesting shorter statistics were that binary gender had no effect on the answer with the differences only coming from non-binary or preferred not to answer responses.

Problem 2: COVID survey - reconstruct

prob 2.1

```
survey_data <- read.csv("data/covid-survey.csv", skip = 1)
head(survey_data)
```

##	response_id	exp_profession	exp_flu_vax	exp_gender	exp_race	exp_ethnicity
## 1	1	1	1	0	2	2
## 2	2	1	1	1	2	2
## 3	3	NA	NA	NA	NA	NA
## 4	4	1	1	0	5	2
## 5	5	1	1	0	5	2
## 6	6	1	1	1	5	2
##	exp_age_bin	exp_already_vax	resp_safety	resp_confidence_science		
## 1	25	1	5		2	
## 2	20	1	5		1	
## 3	NA	NA	NA		NA	
## 4	25	1	5		1	

```
## 5      25      1      5      1
## 6      25      1      5      1
##   resp_concern_safety resp_feel_safe_at_work resp_will_recommend
## 1              2              1              1
## 2              1              1              1
## 3              NA              NA              NA
## 4              1              1              1
## 5              1              1              1
## 6              1              1              1
##   resp_trust_info
## 1              1
## 2              1
## 3              NA
## 4              1
## 5              1
## 6              1
```

```
dim(survey_data)
```

```
## [1] 1121  14
```

```
#### Prob 2.2
```

```
# Filter out rows where all values except "response_id" are NA
```

```
cleaned_data <- survey_data %>%
  filter(!is.na(exp_profession) &
         !is.na(exp_flu_vax) &
         !is.na(exp_gender) &
         !is.na(exp_race) &
         !is.na(exp_ethnicity) &
         !is.na(exp_age_bin) &
         !is.na(exp_already_vax) &
         !is.na(resp_safety) &
         !is.na(resp_confidence_science) &
         !is.na(resp_concern_safety) &
         !is.na(resp_feel_safe_at_work) &
         !is.na(resp_will_recommend))
```

```
head(cleaned_data)
```

```
##   response_id exp_profession exp_flu_vax exp_gender exp_race exp_ethnicity
## 1           1             1           1           0           2           2
## 2           2             1           1           1           2           2
## 3           4             1           1           0           5           2
## 4           5             1           1           0           5           2
## 5           6             1           1           1           5           2
## 6           7             1           1           1           5           2
##   exp_age_bin exp_already_vax resp_safety resp_confidence_science
## 1          25              1           5              2
## 2          20              1           5              1
## 3          25              1           5              1
## 4          25              1           5              1
## 5          25              1           5              1
## 6          25              1           5              1
##   resp_concern_safety resp_feel_safe_at_work resp_will_recommend
## 1                   2                   1                   1
```

```
## 2          1          1          1
## 3          1          1          1
## 4          1          1          1
## 5          1          1          1
## 6          1          1          1
##   resp_trust_info
## 1          1
## 2          1
## 3          1
## 4          1
## 5          1
## 6          2
```

```
dim(cleaned_data)
```

```
## [1] 926 14
```

```
#2.3
```

```
# Relabel survey response values
```

```
cleaned_data2 <- cleaned_data %>%
```

```
  mutate(exp_already_vax = ifelse(exp_already_vax == 0, "No", "Yes"),
         exp_flu_vax = ifelse(exp_flu_vax == 0, "No", "Yes"),
         exp_profession = ifelse(exp_profession == 0, "Medical", "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 = ifelse(exp_ethnicity == 1, "Hispanic / Latino", "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",
```

```
))
```

```
head(cleaned_data2)
```

```
##   response_id exp_profession exp_flu_vax exp_gender exp_race
## 1          1      Nursing      Yes      Male      Asian
## 2          2      Nursing      Yes      Female      Asian
## 3          4      Nursing      Yes      Male      White
## 4          5      Nursing      Yes      Male      White
## 5          6      Nursing      Yes      Female      White
## 6          7      Nursing      Yes      Female      White
##           exp_ethnicity exp_age_bin exp_already_vax resp_safety
## 1 Non-Hispanic/Non-Latino      26-30          Yes          5
## 2 Non-Hispanic/Non-Latino      21-25          Yes          5
```

```
## 3 Non-Hispanic/Non-Latino      26-30      Yes      5
## 4 Non-Hispanic/Non-Latino      26-30      Yes      5
## 5 Non-Hispanic/Non-Latino      26-30      Yes      5
## 6 Non-Hispanic/Non-Latino      26-30      Yes      5
##   resp_confidence_science resp_concern_safety resp_feel_safe_at_work
## 1                2                2                1
## 2                1                1                1
## 3                1                1                1
## 4                1                1                1
## 5                1                1                1
## 6                1                1                1
##   resp_will_recommend resp_trust_info
## 1                1                1
## 2                1                1
## 3                1                1
## 4                1                1
## 5                1                1
## 6                1                2
```

```
dim(cleaned_data2)
```

```
## [1] 926  14
```

```
###2.4
```

```
covid_survey_longer <- cleaned_data2 %>%
  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: 38,892 x 5
##   response_id explanatory explanatory_value response response_value
##   <int> <chr> <chr> <chr> <int>
## 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 38,882 more rows
```

```
###2.5
```

```
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: 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 1.30     1     2
## 3 exp_age_bin 21-25      resp_feel_safe_at_work 1.18     1     2
## 4 exp_age_bin 21-25      resp_safety 1.97     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 3.32     1     5
## 8 exp_age_bin 26-30      resp_confidence_science 1.39     1     2
## 9 exp_age_bin 26-30      resp_feel_safe_at_work 1.27     1     2
## 10 exp_age_bin 26-30      resp_safety 2.17     1     5
## # i 116 more rows
```

###2.6

```
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 3.26     1     5
## 2 resp_confidence_science 1.43     1     2
## 3 resp_feel_safe_at_work 1.34     1     2
## 4 resp_safety 2.04     1     5
## 5 resp_trust_info 1.38     1     2
## 6 resp_will_recommend 1.21     1     2
```

###2.7

```
covid_survey_summary_stats <- bind_rows(covid_survey_summary_stats_all, covid_survey_summary_stats_by_g

covid_survey_summary_stats
```

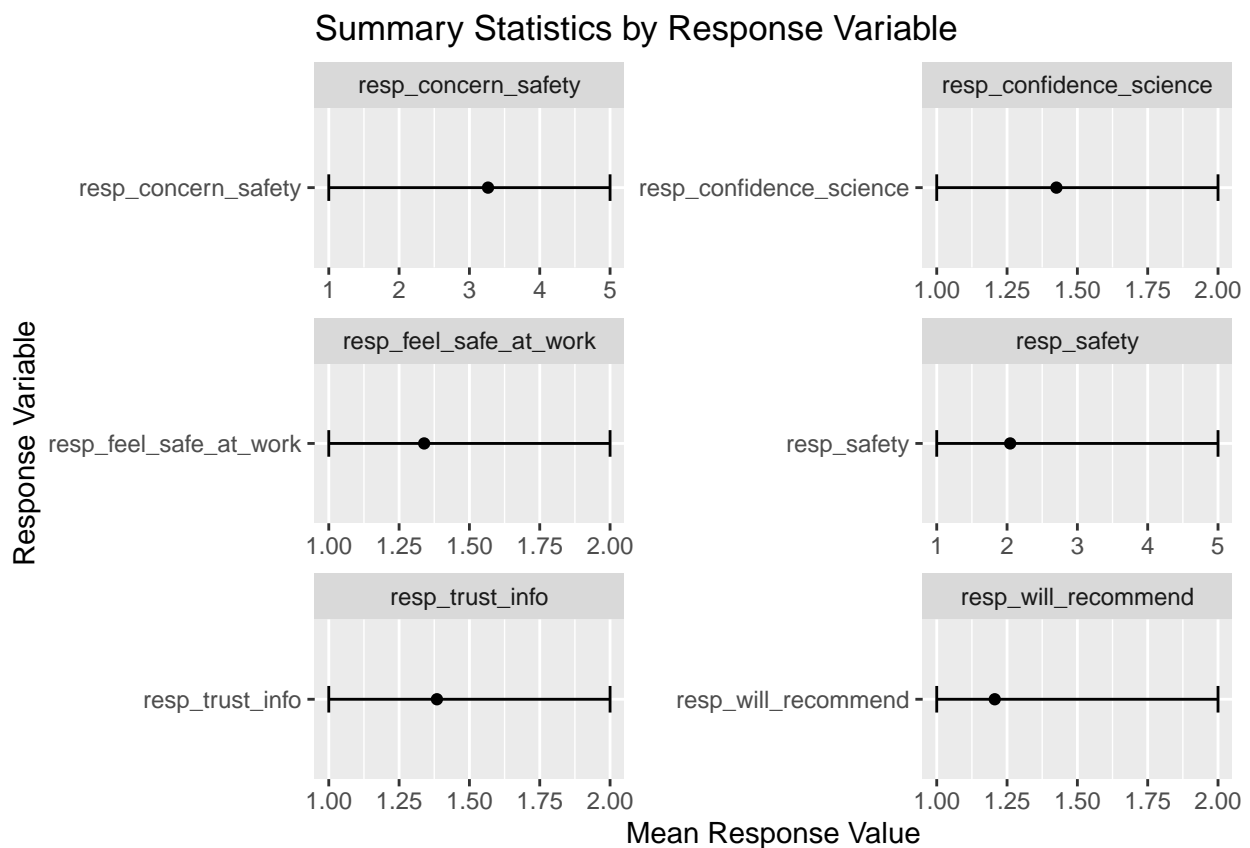
```
## # A tibble: 132 x 6
##   response      mean    low  high explanatory explanatory_value
```

```
##      <chr>                <dbl> <dbl> <dbl> <chr>          <chr>
## 1 resp_concern_safety    3.26    1    5 <NA>          <NA>
## 2 resp_confidence_science 1.43    1    2 <NA>          <NA>
## 3 resp_feel_safe_at_work  1.34    1    2 <NA>          <NA>
## 4 resp_safety            2.04    1    5 <NA>          <NA>
## 5 resp_trust_info        1.38    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 1.30    1    2 exp_age_bin 21-25
## 9 resp_feel_safe_at_work  1.18    1    2 exp_age_bin 21-25
## 10 resp_safety           1.97    1    5 exp_age_bin 21-25
## # i 122 more rows
```

```
###2.8
```

```
# Plot the faceted graph with rotated x-axis
```

```
ggplot(covid_survey_summary_stats_all, aes(x = response, y = mean)) +
  geom_point() +
  geom_errorbar(aes(ymin = low, ymax = high), width = 0.2) +
  facet_wrap(~ response, scales = "free", ncol = 2) +
  labs(x = "Response Variable", y = "Mean Response Value", title = "Summary Statistics by Response Vari")
  coord_flip()
```



```
covid_survey_summary_stats %>%
  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",
    is.na(explanatory) ~ "All"
  ),
  response = case_when(
    response == "resp_safety" ~ "Based on my understanding, I believe the vaccine is safe",
    response == "resp_feel_safe_at_work" ~ "Getting the vaccine will make me feel safer at work",
    response == "resp_concern_safety" ~ "I am concerned about the safety and side effects of the vaccine",
    response == "resp_confidence_science" ~ "I am confident in the scientific vetting process for the vaccine",
    response == "resp_trust_info" ~ "I trust the information that I have received about the vaccines",
    response == "resp_will_recommend" ~ "I will recommend the vaccine to family, friends, and community"
  ),
  explanatory_value = case_when(
    is.na(explanatory_value) ~ "All",
    TRUE ~ explanatory_value
  )
) %>%
ggplot(aes(x = mean, y = explanatory_value, xmin = low, xmax = high)) +
  geom_point() +
  geom_errorbar(aes(xmin = low, xmax = high)) +
  facet_grid(explanatory ~ response, scales = "free", labeller = labeller(explanatory = label_wrap_gen(
  labs(x = "Mean Likert Score", y = NULL) +
  theme_bw() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    strip.background = element_rect(fill = "gray90", color = "gray"),
    strip.text = element_text(face = "bold", size = 8),
    axis.text.x = element_text(size = 8),
    legend.position = "none",
    plot.background = element_rect(fill = "white"), # Set plot background to white
    plot.margin = margin(1, 1, 1, 1, "cm") # Set margin to create space around the plot
  )
)

```

