**Session 3 Group Activity**

Willa Voorhies & Elena Leib

**In your groups..**

*Try doing the following questions first on your own, and then check your answers with your group. Make decisions together as a group for #4, then carry it out yourself.*

*Remember you can get help on any function by typing ? followed by the function into the console. e.g., ?filter. You can also find help by searching on Google e.g., “tidyverse filter”*

*If you need help let us know in the* [*help doc*](https://docs.google.com/document/d/1cx-Ilgynv4U7PsobJPuVtDPZNXUH-vUr7-KejTYZ6ms/edit?usp=sharing)*!*

*Some questions have additional challenge problems. Do the challenge problems last if you have extra time or come back to them after the session!*

**Open an R script and load the covid\_attitudes data (and don’t forget to set options(stringsAsFactors = FALSE) before loading the data)! Then perform the following data processing steps. After you’ve performed each step, use the functions you learned last week to check if it worked.**

**1)** Remove the variable Q6.consent from your dataframe (we only have data for consented participants).

**2)** Only keep observations from large cities.

**3)** Remove participants from the dataframe that have NAs for any cells.

**A.** Use what you learned last week to see if all the NAs are removed.

**B**. If not, try replacing the #N/As for one column using case\_when(). Hint: must use NA\_character\_ in the case\_when, not just NA!

**C**. Now try to remove all NAs again and see what happens

**4)** Add a new variable to your dataframe called “apprehension\_score” that is a composite score of Q18, Q20, and Q21.

**5)** Recode the levels of Q13\_1.trust\_doctor\_news, which has TRUE and FALSE, to 1 for TRUE and 0 for FALSE, and make this variable into a factor

**6)** Make one of the likert-scale columns into a factor

**7)** Save your data frame as a csv file using write.csv().

*hint:*Your code should look something like this:

write.csv(your\_df, file =“add-your-file-name-here.csv”, row.names=FALSE)

What happens if you don’t include row.names=FALSE? Check the help file (?write.csv) to help you make your hypothesis and then check the output file to confirm.

***If you have extra time*… complete the challenge problems and/or explore some additional tidyverse functionality with the bonus tasks below.**

**Challenge Problems:**

Question 2 Challenge: Now keep observations for partcipants who live in large cities OR suburbs (*hint:* see the demo code for examples of logical operators you can use with filter or check out [this link with R operators](https://www.tutorialspoint.com/r/r_operators.htm)).

Question 3 Challenges:

1. You could use case\_when()to change #N/As for each column individually. OR here is some code that will help you change #N/As across the whole dataset.

mutate(across(where(is.character), ~ str\_replace(., "#N/A", NA\_character\_)))

Run this code and then try to write out this code in words. Is there any syntax (symbols or characters) you don’t understand?

1. Using the drop\_na() function only drop participants with missing values in Q17 and Q23. (hint: type ?drop\_na into the R console to get help specifying certain columns)

**Bonus:**

We can use summarise() to get summary statistics for our variables.

Run the following code to see how it works:

attitudes\_summary <- covid\_attitudes\_cleaned %>%

summarise(mean(attitudes\_score))

Try this yourself but now get the standard deviation instead of the mean (*hint:* you can use ?summarise if you’re stuck, and/or look up how to do standard deviation in R).

Now let’s look at the average apprehension score for each age group using another powerful tidyverse command: group\_by(). Run the following code:

attitudes\_summary <- covid\_attitudes\_cleaned %>%

# This is cleaning up all those wonky #N/As for us!

mutate(across(where(is.character), ~ str\_replace(., "#N/A", NA\_character\_))) %>%

# group by age

group\_by(Q40.age) %>%

# get mean apprehension score for each age group

summarise(apprehension\_mean=mean(apprehension\_score))%>%

# it’s important to ungroup!

ungroup()

**i)** View the new attitudes\_summary variable. What did it give you?

**ii)** Try grouping by another variable of interest (or more than one). Don’t forget to ungroup() after you’re done !

*Grouping doesn’t alter your data frame, it just changes how it’s listed and how it interacts with the other commands.*

Check out the tidy cheat sheet for more tidyverse and data wrangling functionality! <https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>