**Session 2 Group Activity**

Elena Leib & Willa Voorhies

**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 #7 and #11, then carry it out yourself.*

*Use the “questionnaire.docx” file from session 1 if you need a column guide.*

*If you need help let us know in the* [*help doc*](https://docs.google.com/document/d/1IesSkLxwvE3uUYDp6Qd3LZcYvGlwUweGFQLhM7_YIlY/edit?usp=sharing)*!*

1. Make sure your data and scripts from the bootcamp are organized in a coherent file structure.
2. Open a new R script from your bootcamp folder and check your working directory.
   * *Don’t forget to comment at the top and throughout to explain what you are doing*
3. Read in the data
   * *Make sure to give the data frame a meaningful variable name*
4. Look at the data and make sure it loaded correctly
   * Look at global environment
   * View()
   * str()
   * summary()
5. How many columns are in the data?
6. How many rows?
7. Choose two columns you think should be factors and make them into factors.
8. Which columns have the most NAs? List the top 3. Make sure you all agree!
   * *What do you notice about how the NAs are represented in different columns? In summary? When you View() it?*
9. On average, how likely do people think they are to catch covid-19?
10. How many types of living communities are there? And how many responses are there from each type of community?
11. Think of another question you can ask with your group and answer it.