Lab 1A - Data, Code & RStudio

**Directions: Record your responses to the lab questions in the spaces provided.**

Describe the data that appeared after running View(cdc):

*Who* is the information about?

What sorts of information about them was collected?

How are our *observations* represented in our data?

What does the first column tell us about our observations?

How often did our first observation wear a seatbelt while riding in a car?

How many students are in our cdc data set?

How many variables were measured for each student?

Which of these functions tell us the number of observations in our data?

Which of these functions tell us the number of variables?

Run the following commands and write down what happens after each. Which does R understand?

Which one of these plots would be useful for answering the question: *Is it unusual for students in the CDC dataset to be taller than 1.8 meters?*

Do you think it's unusual for students in the data to be taller than 1.8 meters? Why or why not?

What is *public health* and do we collect data about it?

How do you think our data was collected? Does it include every high school aged student in the US?

How might the CDC use this data? Who else could benefit from using this data?

Write the code to visualize the distribution of weights of the students in the CDC data with a histogram. What is the *typical* weight?

Write the code to create a barplot to visualize the distribution of how often students wore a helmet while bike riding. About how many students never wore a helmet?