

Name: _____

Date: _____

LAB 1A: Data, Code & RStudio Response Sheet

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

Welcome to the labs!

So let's get started!

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

(1) *Who* is the information about?

(2) What sorts of information about them was collected?

Data: Variables & Observations

(3) Based on the data, describe a few characteristics about the first observation.

(4) What does the first column tell us about our observations?

Uncovering our Data's Structure

(5) How many students are in our cdc data set?

(6) How many variables were measured for each student?

Name: _____

Date: _____

LAB 1A: Data, Code & RStudio Response Sheet

Some new functions

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

(8) Which of these functions tell us the number of variables?

First Steps

Syntax matters

(9) What happens after each command?

(10) Which does R understand?

R's most important syntax

Syntax in action

(11) Would a histogram, bargraph, or scatterplot be useful for answering the question: *Is it unusual for students in the CDC dataset to be taller than 1.8 meters?*

(12) Do you think it's unusual for students in the cdc dataset to be taller than 1.8 meters? Why or why not?

Name: _____

Date: _____

LAB 1A: Data, Code & RStudio
Response Sheet

On your own:

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

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

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

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

(17) Write and run the code to create a bargraph to visualize the distribution of how often students ate fruit. About how many students did not eat fruit over the previous 7 days?