Name: Date:

**Lab 1A: Data, Code & Google Colab *Response Sheet***

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

**So let's get started!**

• Describe the data that appeared after calling your variable “cdc”.

I see rows and columns, ages and genders.

• *Who* is the information about?

People who have traveled

• What sorts of information about them was collected?

Ages, genders, grade, race

**Data: Variables & Observations**

• How are our *observations* represented in our data?

They are represented

• What does the first column tell us about our observations?

Tells us the ages

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

Told us they always wore a seatbelt

**Uncovering our Data's Structure**

• How many students are in our cdc data set?

15,623

• How many variables were measured for each student?

*34*

**Type the following commands into google colab**

**cdc.shape**

**cdc.size**

**len(cdc.index)**

**len(cdc.column)**

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

cdc.shape

• Which of these functions tell us the number of variables?

**len(cdc.columns)**

**Syntax matters**

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

**List(cdc.columns)**

Shows the amount of variables

**list(cdc)**

**Shows the name of the variables**

**list(cdc.columns)**

Shows amount of variables

**LIST(CDC.COLUMNS)**

**Error**

**Syntax in action**

• 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?*

cdc.hist("height")

cdc.plot(kind="hist", y="height")

cdc['drive\_text'].value\_counts().plot.bar()

cdc.plot(kind="scatter", x="weight", y="height")

cdc.plot(kind="scatter", x="height", y="weight")this one

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

No because 5 feet is a normal height

**On your own:**

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

How healthy citizens are and we collect data to improve health

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

To help better health. No it did not

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

*They might use is to implement health changes. The public can benefit from 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?