Name(s): _____

DATA 101 Assignment 2: Statistics and Data

Work in teams of 2

Instructions. Answer the following questions using the linked datasets and activities. Always note the source of your information when applicable. Write your responses in the spaces provided and show your work.

---

1. Complete this brief survey about your interests: https://forms.gle/ZG3U1yeCWdbbbNCe6.

2. Complete the activity about Johann Sebastian Bach's works from Week 2's course slides on Github, page 6. Write your answers in the space provided below.

3. Go to tidy tuesday's repository. Find a dataset you are interested in. Write or paste below the dataset in a spreadsheet format, including the first 5 rows. Why did you pick this dataset, and what makes it interesting to you? What are the units of the dataset? (Recall: the unit of a dataset is the object or subject we took measurements from.)

4. Go to https://enrollment.streamlit.app. Find the information below.

(a) What was the course with the highest enrollment in Althouse 201 and when?

(b) What was the course with the highest enrollment in the Althouse building and when?

(c) What was the course with the highest enrollment that I taught and when?

5. In each case determine whether the variable is categorical or numerical. If the variable is numerical, state whether it is discrete or continuous.

(a) The number of people entering the post office between 1pm and 3 pm on Monday

(b) The numbers on the uniforms of a soccer team.

(c) The length of time to swim 50 yards.

(d) The brand of phone purchased by a customer.

(e) Credit card numbers.

6. Below are the numbers of billionaires from five different states (consider this a sample of states) in the Northeast (according to *Forbes.com*).

State	**Billionaires**
New York	67
Connecticut	11
Pennsylvania	7
Massachusetts	7
New Jersey	5

(a) Compute the mean number of billionaires. Show your work.

(b) Compute the standard deviation (by hand and show your work) of the number of billionaires (round to the nearest tenth).

(c) Which of the given numbers contributes the most to the standard deviation? Why?