Name _____ Period _____

Visualizing Data

Your Tasks (Mark these off as you go)
☐ Tell a data story
☐ Obtain a dataset of objects
☐ Identify categorical variables
☐ Tidy the data into a computer readable format
☐ Create visualizations for your data
☐ Analyze your data
☐ Receive credit for this lab guide

☐ Tell a data story

With a partner, look through the graphs on this site: What's Going on in this Graph?

Scroll a bit and you will notice that the graphs are organized by topic and type,

Graphs Organized by Topic

Environment, Science and Technology

- Houston May Get 50 Inches of Rain. How Long Does it Take Your City to Get That Much?
- It's Not Your Imagination. Summers Are Getting Hotter.

Graphs Organized by Type

es Again and Again

Most?

Distribution (values and their frequency)

- Six Myths About Choosing a Major (boxplot)
- It's Not Your Imagination. Summers Are Getting Hotter. (histogram)

Paste a graph that you and your partner find interesting below.

What do you notice? Share what you are noticing and what this may imply.

What do you wonder? Where could you find the answers to what you wonder?
What's going on in this graph? Use what you notice. What can you infer from this graph beyond what it shows directly? What's the deeper story that comes from this graph?

□ Obtain a dataset of objects

Locate the datasets of objects you have been provided. Record the data set your group selected.

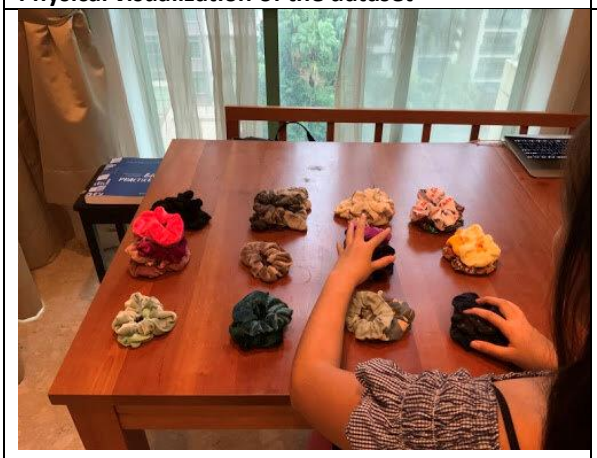
□ Identify categorical variables

Observe the dataset of objects you selected. What do you notice? Do the objects have different colors, shapes, size, design, etc.
What do you wonder? What you wonder about will depend on the dataset of objects you selected. For example, how many are there of each color? How many are there of each shape? Etc. Write what you wonder about as questions below. You must write two questions about your data set.
Question 1:
Question 2:

A **categorical variable** (also called **qualitative variable**) is a variable that can take on a fixed number of possible values based on some observed quality. List the categorical variables you identified below.
Categorical variable 1:
Categorical variable 2:

□ Prepare the data

It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data. Below is an example of scrunchies organized by color and frequency of wear.

Physical visualization of the dataset	**Numerical visualization of the data set**
	
The images above show the physical organization of the data and the total counts of each scrunchy by two categorical variables, Color and Frequency Worn.	

Consider the categorical variables you identified above. Organize your objects based on the categorical variables you identified. Take a picture of your physical visualization and insert this below.

Create a numerical visualization by counting and recording the number of items in each category. Do this on the whiteboard you have been provided. Take a picture of your numerical visualization below.

□ Tidy your data into a computer readable format

Tidy Data has become the standard format for science and business because it allows people to easily turn a data table into graphs, useful for analysis and insight.

There are three interrelated rules that make a dataset tidy,

1. Each variable is a column; each column is a variable.
2. Each observation is a row; each row is an observation.
3. Each value is a cell; each cell is a single value.

Below is an example of initial data collected in spreadsheet, and entered in a typical way that most novices would use (left), and the same data in Tidy Format (right),

Day	Plant A height (cm)	Plant B height (cm)
1	0.7	1.5
2	1.0	0.7
3	1.5	0.9
4	1.8	1.3
5	2.2	1.8

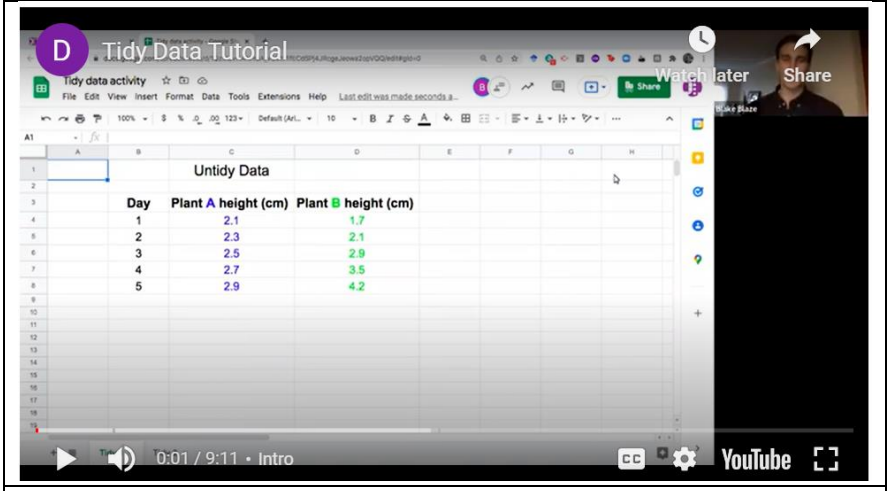
Day	Plant	Height (cm)
1	A	0.7
2	A	1.0
3	A	1.5
4	A	1.8
5	A	2.2
1	B	1.5
2	B	0.7
3	B	0.9
4	B	1.3
5	B	1.8

This is clearly the same data, but it now has:

- Three clearly identified **variables** (days, plant type, and height) each of which has its **own column.**
- Each **row** is an independent **observation** of the variables
- Each **cell** contains a single **value** of a single variable

Watch the video below, to learn more,

Tidy Data Tutorial



https://youtu.be/fFoK8Rdo4SA?si=GFELBp2hCL8X55zc

Below is an example of how the scrunchy dataset above was organized in a Tidy format using the categorial variables color and frequency worn.

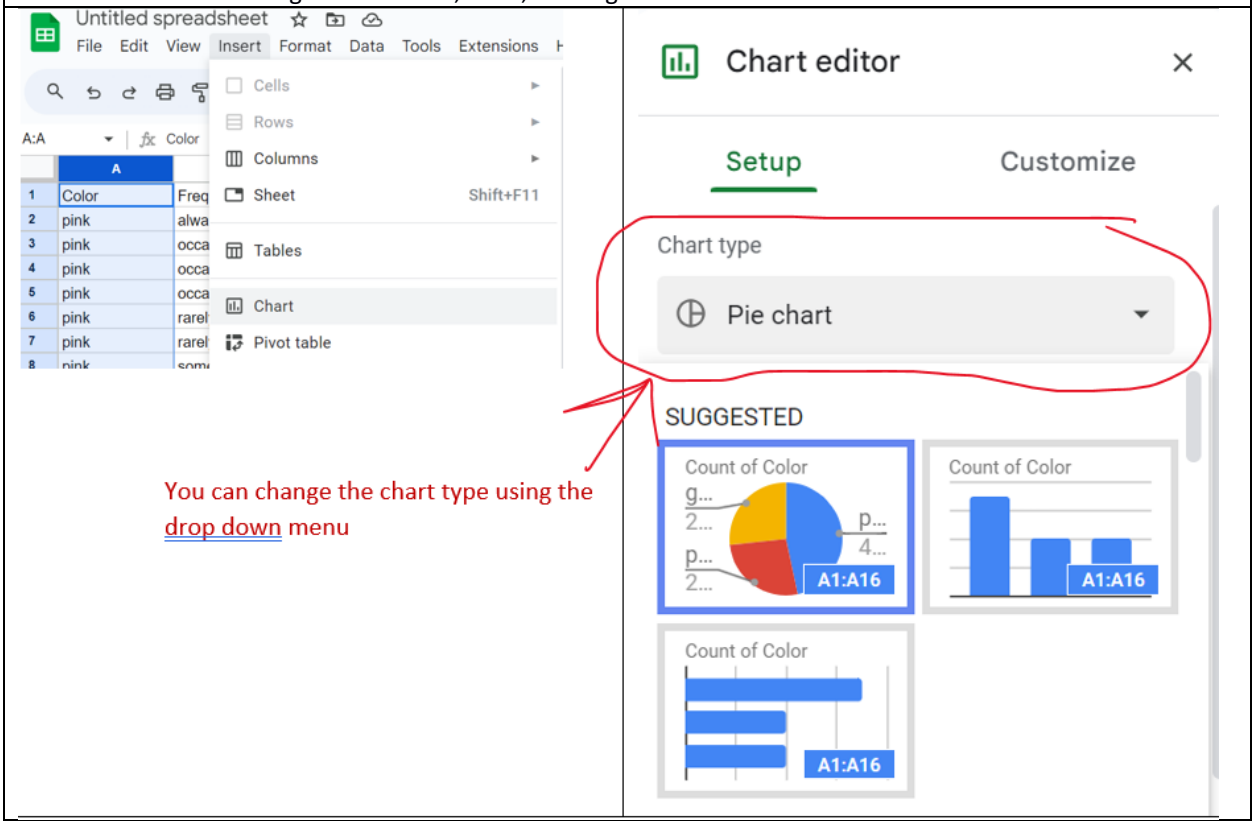
Scrunchy data set in Tidy format

	A	B
1	Color	Frequency Worn
2	pink	always
3	pink	occasionally
4	pink	occasionally
5	pink	occasionally
6	pink	rarely
7	pink	rarely
8	pink	rarely
9	purple	sometimes
10	purple	sometimes
11	purple	sometimes
12	purple	sometimes
13	purple	sometimes
14	purple	sometimes
15	gray	occasionally
16	gray	occasionally
17	gray	occasionally
18	gray	sometimes
19	gray	sometimes
20	gray	sometimes

Create a new Google Sheet. Enter your data in a Tidy format as explained in the video and illustrated above. Share your sheet and include a link to it below,

☐ Create visualizations for your data

Once you have your data in a Tidy format, it is easy to visualize your data using Google Sheets. Below is an example of how we can visualize the number each color in the scrunchy dataset,

How to visualize the categorical variable, color, in Google sheets


Create two visualizations using the categorical variables you identified above. Take a screen shot of your visualizations and paste them below.

☐ Analyze your data

Now that you have created your visualizations, you should be able to answer your original questions. Revisit your original questions, for each question use your visualizations to answer them	
Question	**Answer**

☐ Receive Credit for this lab guide

Submit this portion of the lab to Pluska to receive credit for the lab guide.