Week 2.1 - Data (Tidy Data)

* Set up polleverywhere; <http://bit.ly/fall18-pollev>
* Questions from Reading
* Head Sketch Survey

1. Do you think head lengths might vary in the class? What causes variation in head length? What are some important characteristics about a person that affects how tall their head is?
2. Did your partner come up with identical measurements? What causes variation in these measurements?
3. What are some things you could do if you wanted to make these measurements more similar? That is, how can you reduce the variability in your measurements?
4. Can we completely reduce the variability in measurement? Why or why not?
5. Now we’re going to have the 3 tall and 3 short students in the class share with us their head measurements.
   1. Let’s create a rough plot of our data.
   2. Notice something super obvious… these measurements vary! Now try to “explain” some of this variation. In the previous dotplot, color code the measurements that you think come from the tallest students vs shortest students. Were you pretty successful? Was the class pretty successful?
6. What do you think it means to “explain” variation?
7. Does explaining variation mean that the variation just goes away?
8. How would we put this data (from the survey) into “tidy” format?

* Three principles of tidy data
* <http://bit.ly/tidydata-fall18>

1. How do we make a histogram?

* <http://tinyurl.com/datacampsandbox>

<pre><code>csvfile &lt;- &quot;http://bit.ly/headsurvey-fall18&quot;  
headsurvey &lt;- read.csv(csvfile, header=TRUE)  
head(headsurvey)  
gf\_histogram(~ selfheadsize\_cm, data = headsurvey, fill = &quot;magenta&quot;)  
</code></pre>