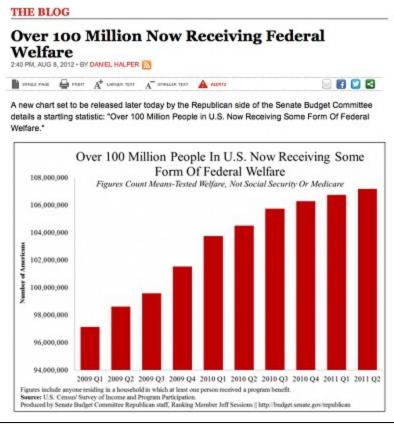*Please save this file as "LAST NAME_Assignment 3.docx"*
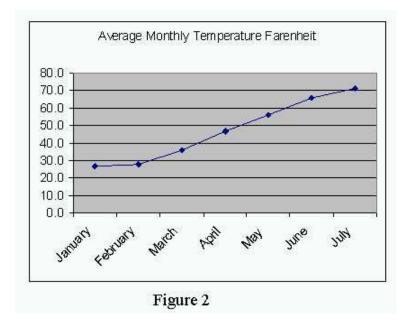
Open-Ended Responses

1. **III.1** This is a graph from the USA Today back in 2012; I would argue that it is misleading because of the y-axis. According to the principles of data (Tufte), how might the y-axis distort the data?
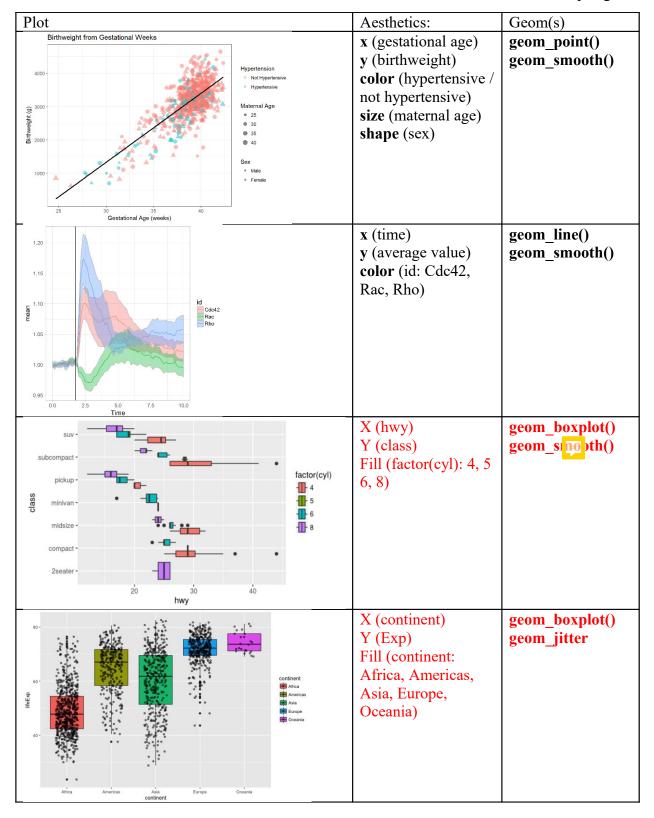

The Y axis is distorting the data by showing a very small variation between the X values to appear to be a larger variation. The X values are with single percent differences but since the Y value starts so high and covers a vary small variation the X values when comparing them end up looking distorted.
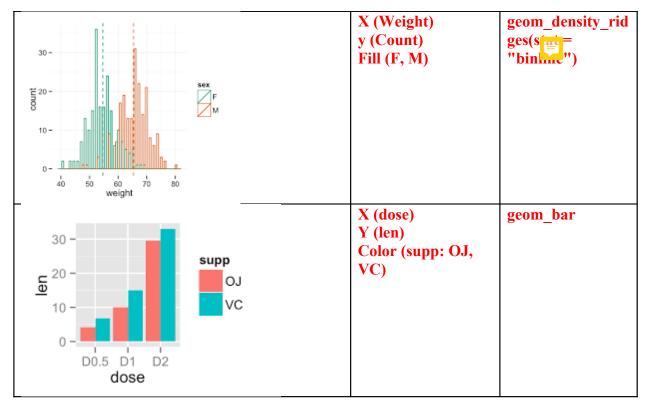
2. **III.1** Let's say I'm trying to convince someone that we should be very concerned about rising temperatures due to climate change so I show the average temperatures in Connecticut. I would argue that the below graph is misleading because of the x-axis. According to the principles of data (Tufte), how might the x-axis distort any claims I'm making about global climate change?

Figure 2

The data is inconclusive to compare temperature rates you need a larger sample size then a couple of months.

3. **III.5** For each plot, label the **aesthetics & geom(s)** that are present. The first two rows are filled out as examples of what I'm looking for. I have mapped the variable to the aesthetics for clarity, <u>but you do not need to do that</u>. Note: Color vs Fill can be had to tell and depends on the geometry; I'll be lenient with these.

Plot	Aesthetics:	Geom(s)
	x (gestational age) **y** (birthweight) **color** (hypertensive / not hypertensive) **size** (maternal age) **shape** (sex)	**geom_point()** **geom_smooth()**
	x (time) **y** (average value) **color** (id: Cdc42, Rac, Rho)	**geom_line()** **geom_smooth()**
	X (hwy) Y (class) Fill (factor(cyl): 4, 5 6, 8)	**geom_boxplot()** **geom_smooth()**
	X (continent) Y (Exp) Fill (continent: Africa, Americas, Asia, Europe, Oceania)	**geom_boxplot()** **geom_jitter**

	X (Weight) y (Count) Fill (F, M)	geom_density_rid ges(stat = "binline")
	X (dose) Y (len) Color (supp: OJ, VC)	**geom_bar**

4. **III.6** Name an example in your own discipline where you would ever find the utility of using facet_wrap() or facet_grid() to produce multiple plots that are very similar, but change something each time.

When comparing multiple classes you can use facet wrap to put the classes plotted by grades next to each other for comparison

5. **III.8** What is the primary advantage to exporting your plot with ggsave() or png(), jpg(), pdf(), etc. versus just copying/pasting or grabbing a screen shot from the previewer pane in RStudio?
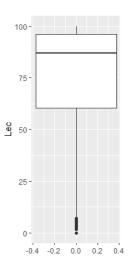
Exporting the plot gives you the ability to customize the resolution/size/file-size

Coding Section

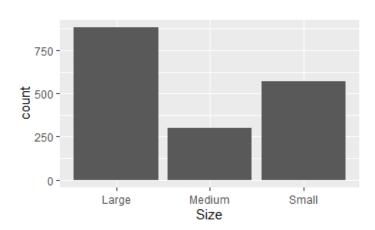To complete this section, start a new script file with the following layout:

```
# YOUR NAME
# Assignment 3 Data Visualizations

# #1 ---------------- (new section: CTRL + SHIFT + R)
here's my code # with adequate commenting

# #2 ---------------- (new section: CTRL + SHIFT + R)
here's my code # with adequate commenting
```

Using the copus data, make the following plots to the best of your ability (may not be exactly the same if using jittering, default colors, text, titles, exact theme, etc, but should tell the same story). Note: you might have to manipulate the data before you can make the plot. You just need to supply the code for this, no need to actually grab/submit a screen shot or save the plot.
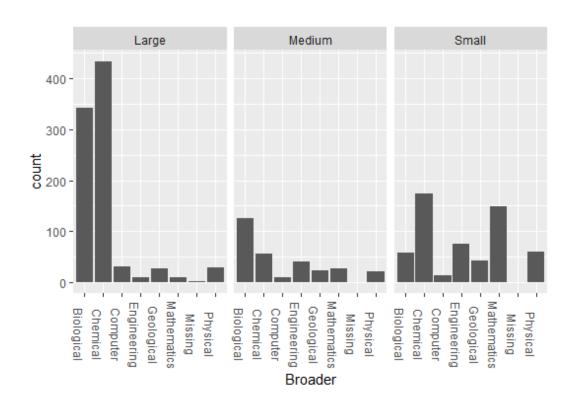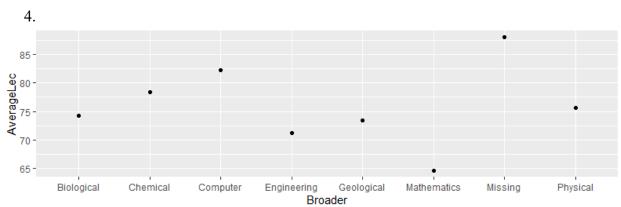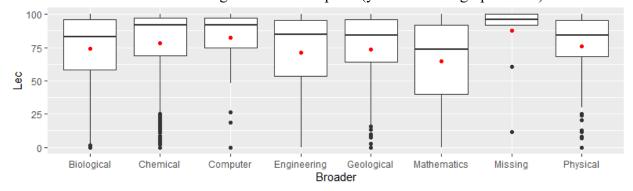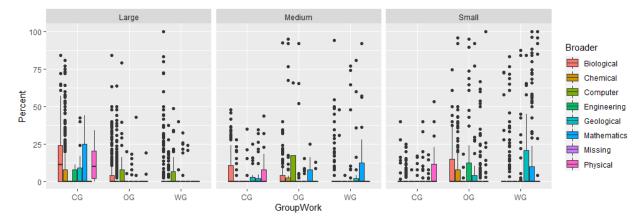
1.

2.

3.

4.
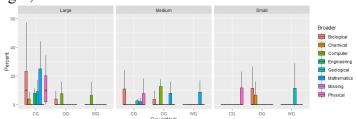
5. The red dots are the average for each discipline (you made this graph above).

6.

7. For the final plot:
   a. Take your code from the previous plot.
   b. Clean up the plot by:
      b.i.   Removing all the outlier points (all the individual points outside of the main boxplot; see ??geom_boxplot() and all of the arguments that start with "outlier.")
      b.ii. Change the y-axis limits for a maximum of 60.
      b.iii. After changes, it should look like this:

   c. Write code that will export it as a .pdf that is 2 inches tall, 6 inches wide.