*\*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?

Chart, bar chart

Description automatically generated

|  |
| --- |
| Graphical integrity is impacted by starting the Y axis at 94,000,000 and going up by 2,00,000 each time. If the true scale was shown the difference between 2009 and 2011 would not seem as stark. |

1. **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?

Chart, line chart

Description automatically generated

|  |
| --- |
| The X axis shows a monthly temperature from January to July of one year, rather than the temperature from the same time of year for several years. Due to natural fluctuations in temperature throughout the year, this graph is just showing a normal seasonal change in temperature. |

1. **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, but you do not need to do that. Note: Color vs Fill can be had to tell and depends on the geometry; I’ll be lenient with these.

|  |  |  |
| --- | --- | --- |
| Plot | Aesthetics: | Geom(s) |
| 3.3 A ggplot2 Tangent | R for Statistics in EPH | **x** (gestational age)  **y** (birthweight)  **color** (hypertensive / not hypertensive)  **size** (maternal age)  **shape** (sex) | **geom\_point()**  **geom\_smooth()** |
| Visualizing data with R/ggplot2 - One more time - the Node | **x** (time)  **y** (average value)  **color** (id: Cdc42, Rac, Rho) | **geom\_line()**  **geom\_smooth()** |
| ggplot2 extensions | X  Y  Color | **Geom\_boxplot()** |
| How to Make Boxplot in R with ggplot2? - Python and R Tips | X  Y  Color | **Geom\_boxplot()**  **Geom\_point()** |
|  | **X**  **Y**  **Color** | **Geom\_histogram()** |
|  | **X**  **Y**  **Color** | **Geom\_bar()** |

1. **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.

|  |
| --- |
| I would use facet\_wrap or facet grid to produce multiple plots showing the relationship between a scores on a test of cognitive functioning and scores for a test of cognitive persistence. By using the multiple plots I could show the difference in that relationship for females (ages 65 plus), males (ages 65 plus), females (18-35) and males (18-35). |

1. **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?

|  |
| --- |
| Resolution! When a plot is saved in a png, jpg, or pdf file it is a Raster file. In a Raster file, one color is associated with one pixel, which is very fine grain. |

***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.

Chart

Description automatically generated

Box plot of all Lec

Chart, bar chart

Description automatically generated

Bar chart of number of each class size

Chart

Description automatically generated

Faceted bar chart showing broader for each size

Chart

Description automatically generated

Scatter plot of the averagelec by broader

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

Chart, box and whisker chart

Description automatically generated

Old point plus box of lec by broader

Calendar

Description automatically generated

Box plot of group work in broader by percent faceted into size

1. For the final plot:
   1. Take your code from the previous plot.
   2. Clean up the plot by:
      1. 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.”)
      2. Change the y-axis limits for a maximum of 60.
      3. After changes, it should look like this:

Chart

Description automatically generated

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