*\*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? It appears that y axis starting at non-zero point which make a significant difference in how a reader will understand the data visualization. Including the zero point, there should not be that much larger difference as that graph suggested.

Chart, bar chart

Description automatically generated

|  |
| --- |
|  |

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),
2. how might the x-axis distort any claims I’m making about global climate change? Because of omitting data as not including the remaining months since the warm season last for about 3 months which is from July-Sept and July is the hottest month with the average high ~ 80 Fahrenheit. Including all months help in given a clear image of the temp. I am not sure if the figure is trying to convince that person to be concerned or relaxed even though it shows a steady increase.

Chart, line chart

Description automatically generated

|  |
| --- |
|  |

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. My understanding the color is

|  |  |  |
| --- | --- | --- |
| 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(hwy)  y(class)  fill(factor(cyl): 4,5,6,8) | **geom\_boxplot()** |
| How to Make Boxplot in R with ggplot2? - Python and R Tips | x(continent)  y(lifeExp)  fill(continent: Africa, Americas, Aisa, Europe, Oceania) | **geom\_boxplot()**  **geom\_jitter()** |
|  | **x(weight)**  **y(count)**  **color(sex: F,M)** | **geom\_line()**  **geom\_histogram()** |
|  | **x(dose)**  **y(len)**  **fill(supp:OJ, VC)** | **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. Since the class example is related to COPUS and that is my project as well, I am going to choose your graduate elements paper as an example. This would be helpful if I am looking at the graduate element satisfaction from each subject of the chemistry dept like how satisfied the students from physical chemistry compared to biochemistry students

|  |
| --- |
|  |

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?

The plot will result in high (better) resolution.

|  |
| --- |
|  |

***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

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated

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

Chart, box and whisker chart

Description automatically generated

Calendar

Description automatically generated

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.