*\*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

|  |
| --- |
| The y-axis distorts because the scale does not show the total number of people receiving federal welfare, but only a portion of them. The scale starts at 94,000,000 and goes to 108,000,000 which would be 14 million of a possible 108 million or about 13%. That makes the rise over the 2.5 years look more dramatic with large jumps. If the entire 108,000,000 were included on the scale of the y-axis, the rise over 2.5 years would appear much smoother with less dramatic jumps. |

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 is showing average temperature for 6 months. In Connecticut, the average temperature in July will always be higher than the average temperature in January, with or without climate change. A graph to show climate change should show average yearly temperature over multiple years or if you want to show monthly average, then show each monthly average over multiple years, i.e. don’t compare months to other months but to that month in other years. |

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(hwy)  y(class)  fill(cyl: 4,5,6,8) | **geom\_boxplot()** |
| How to Make Boxplot in R with ggplot2? - Python and R Tips | x(continent: Africa, Americas, Asia, Europe, Oceania)  y(lifeExp)  fill(continent: Africa, Americas, Asia, Europe, Oceania) | **geom\_boxplot()**  **geom\_jitter** |
|  | **x(weight)**  **y(count)**  **color(sex: F, M)** | **geom\_histogram** |
|  | **x(dose: .5, 1, 2)**  **y(len)**  **fill(supp: OJ, VC)** | **geom\_col** |

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.

|  |
| --- |
| Academic librarians who work in public services usually have somewhat different roles than those who work in technical services. Researchers looking at academic librarians might be interested in plotting the relationship of the variables of interest but faceting to get 2 plots, one for public services and one for technical services to see if they differ. |

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?

|  |
| --- |
| Saving as a raster file(png, jpg, tiff) will let you makes changes to the size or resolution of an image, for example to meet the specific requirements of a journal. Saving as a vector file (pdf, eps, svg, ai) means the image is saved and will be drawn by a processor later which means it can be infinitely resolved. Copying/pasting or taking a screen shot does not allow you this same functionality meaning that images will often appear grainy or fuzzy, particularly if you try to enlarge them. |

***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.