*\*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 range of the y-axis is relatively small. Also, the y-axis starts roughly 3 million below the lowest bar graph value but ends approximately one million above the highest bar graph. This makes it appear that the increasing trend has increased to a “maximum” y-value, whereas the limitations of the y-axis have simply been constrained to make it appear this way. Either end of the y-axis should be the same distance from the lowest and highest data points so that the data does not look skewed. |

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 encourages you to compare monthly temperatures, but it only includes from January to July. During this time, the temperatures in most places will rise naturally, not simply because of climate change. If the x-axis went from January to December for multiple years and showed an increase or compared one month (January) across several years, this would more accurately display the trend. In this sense, the x-axis distorts the data because it does not show a wide enough range. |

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)  color(factor(cyl): 4, 5, 6, 8) | **Geom\_boxplot +**  **Coord\_flip** |
| How to Make Boxplot in R with ggplot2? - Python and R Tips | x (continent)  y (lifeExp)  color(continent: Africa, Americas, Asia, Europe, Oceania) | **Geom\_point()**  **Geom\_boxplot()** |
|  | x (weight)  y (count)  color (sex: F, M) | **Geom\_histogram(color = sex)** |
|  | x (dose)  y (len)  Fill (supp) | **Geom\_bar (position = “dodge”)** |

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 explore the sorption of Phosphorous (P) by Gypsum at various P conc and flow rates. So rather than having a giant graph with 5 different P concentrations at three different flow rates, I facet wrap based on flow rate and on P conc. This allows me to compare the effect P conc at each flow rate and to compare the effect of flow rate at each P conc. It is much easier to compare treatment effects this way rather than looking at a large plot containing everything. |

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 primary advantage is the resolution. ggsave() or other exporting methods allow you to increase the resolution or maintain a high resolution, whereas using a screen shot or copy/paste does not allow you to increase resolution and will save at the resolution and size that is has within R already. |

***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.