*\*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 principles we learned in class are as follows:

* Show the data
* Focus on substance, not design of graphics
* Avoid distorting what the data have to say
* Make large data coherent in small space
* Encourage comparisons
* Reveal data as several levels (broad to dine details)
* Clear purpose: description, exploration, tabulation, decoration
* Integrated with descriptions of data

When considering these principles while examining this graph, I am particularly concerned about the principle “avoid distorting the data and what they have to say” because the y-axis starts at 94,000,000. This makes all of these yearly quartiles look like HUGE differences – when in actuality, they aren’t that large of differences when considering the roughly 97,000,000 people on welfare a timepoint 1 (2009 Q1). Additionally, since they are using this chart to state there are now more than 100,000,000 people on welfare, I would appreciate a line at 100,000,000 to focus on the point. Finally, the article using this figure was released in 2012, stating “over 100 million people in U.S. now receiving Federal Welfare.” However, this was already true in 2009 Q4, so it is not new news. This is another reason adding a line at y=100,000,000 would be nice to add to this graph, however, this would make it clearer that the news headline is making a claim that was true 3 years prior as well… Not breaking news, but definitely click bait!

|  |
| --- |
|  |

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

Showing temperatures in one state of the United States at 7 different months of presumably the same year just shows temperature fluctuations based on earth’s axis tilt at different times of the year (i.e., seasonal changes). This graph is showing us temperature changes going from winter (January – March) to Spring (March-May) and Summer (May-July). If the graph showed average temperatures per year for 100’s of years, then that would make a more convincing argument about rising temperatures due to climate change. Ultimately, the x-axis distorts what the data have to say and does show the data that we would need to see to be convinced of an argument about rising temperatures due to climate change.

|  |
| --- |
|  |

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 = Factor (cyl)  color=Factor (cyl) | **geom\_boxplot()**  **coord\_flip()** |
| How to Make Boxplot in R with ggplot2? - Python and R Tips | **x** (continent)  **y** (lifeexp)  fill = Continent  color= Continent | **geom\_boxplot()**  **geom\_jitter()** |
|  | **x** (weight)  **y** (count)  fill = Gender  color=Gender | **geom\_histogram()** |
|  | **x** (dose)  **y** (len)  fill = supp | **geom\_col(**  **position= “dodge”, width = .5)** |

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 am in biology education research, so I can see myself using facet\_wrap to create plots comparing class size (small, medium, large), group size (small, medium, and large), and other potential moderators of interest of student performance, student affect, etc. Since I work in a diversity, equity, and inclusion lab, I can also see the utility of using facet\_wrap to parse data by demographics.**

|  |
| --- |
|  |

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? Increased 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.