*\*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

|  |
| --- |
| This graphic is illustrating a point that may be presenting a forced bias, lacking in graphical integrity by forcing the scale of the y-axis to represent a population that may be gradually increasing regardless of the quarter or year and the scale does not range from zero to 100,000,000, inflating the effect in favor of the party |

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

|  |
| --- |
| These would be typical rising temperatures during the summer months in one state regardless of climatic changes. Having yearly temperature changes as the x-axis (potentially from the northern states vs. southern states) would better represent the claim and provide insight into the actual issue of warming events |

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 (id: FF6666, 66CC66, 33CCFF, CC66FF) | **geom\_boxplot() +**  **coord\_flip()** |
| How to Make Boxplot in R with ggplot2? - Python and R Tips | X (contintent)  Y (LifeExp)  Color (id: default boxplot color by continent) | **geom\_point()**  **geom\_boxplot ()** |
|  | **X (weight)**  **Y (count)**  **Color (sex)** | **geom\_histogram()** |
|  | **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 would use this code to incorporate different effects of plant-based algacidal treatments on; nontarget organisms (zooplankton), phytoplankton, toxin reduction, off-flavor removal, and nutrient concentrations, all in a consolidated form, all related to the increased concentration of the treatment. |

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?

|  |
| --- |
| This allows you to maintain the quality of the plot without having problems with pixilation. |

***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

ggplot(copus, aes(y = Lec)) +

geom\_boxplot()

Chart, bar chart

Description automatically generated

copus <- copus %>%

filter(!is.na(Size))

ggplot(copus, aes(x = Size)) +

geom\_bar() #creating bar chart

Chart

Description automatically generated

copus <- copus %>%

filter(!is.na(Size)) #filtering out NA’s

ggplot(copus, aes(x = Broader)) +

geom\_bar() +

facet\_wrap(~Size) +

theme(axis.text.x = element\_text(angle = 90, hjust = 1))

Chart

Description automatically generated

copus <- copus %>%

ungroup()

copus.a <- copus %>%

group\_by(Broader) %>%

summarize(AvgLec = mean(Lec))

ggplot(copus.a, aes(x=Broader, y = AvgLec)) + geom\_point()

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

Chart, box and whisker chart

Description automatically generated

copus <- copus %>%

group\_by(Broader) %>%

summarize(c(AvgLec = mean(Lec)), Lec) %>%

rename(“AvgLec = mean(Lec))”)

ggplot(copus, aes(x = Broader, y = Lec)) +

geom\_boxplot() +

geom\_point(aes(x = Broader, y = AvgLec, colour = “red”)) #creating red mean point

Calendar

Description automatically generated

Copus <- copus %>%

Group\_by(Broader) %>%

sapply()(FUN\_simplify = TRUE, USE.NAMES = TRUE)

Sapply(copus, class)

copus[3] <- sapply(copus[3],as.character) #need to have character, column 3

copus[4] <- sapply(copus[4],as.character) #need to have character, column 4

copus[5] <- sapply(copus[5],as.character) #need to have character, column 5

copus <- copus %>%

pivot\_longer(cols = CG:OG, names\_to = “GroupWork”) %>% #create GroupWork

#pivot around GroupWork

group\_by(“GroupWork”)

copus <- copus %>%

group\_by(Broader)

sapply(copus, class) #change values to numeric to plot

copus[4] <- sapply(copus[4], as.numeric)

ggplot(copus, aes( x = “GroupWork”, y = value, fill = Broader)) +

geom\_boxplot () +

geom\_point () +

facet\_wrap(~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.

ggplot(copus, aes( x = `GroupWork`, y = value, fill = Broader)) +

geom\_boxplot(outlier.shape = NA) + #remove outliers

ylim(0,60) + #change scale from 0 to 60

facet\_wrap(~Size) #split up data

ggsave("df\_dbl\_res\_in.pdf", #export dataset as pdf in good quality

height = 2,

width = 6,

units = "in",

dpi = 200)