GGPlot Class Reflections

**Lesson 7-Boxplots part II**

**What was your level of comfort with the lesson/application?**

I was very comfortable with this lesson and application as it mostly covered material that we had seen before (how to calculate quantiles, convert values to categories/levels, make boxplots). I appreciated this because it gave me the opportunity to reinforce what I had learned and highlighted the areas I had to go back and read. There was some new material (facet grid, using factor (levels=) in ggplot, the labeller function, and scale\_x\_discrete(limits=) which was great to learn because it is handy.

**What areas of the lesson/application confused or still confuses you?**

Why didn’t my for loop (see my code: method 1) and my non for loop (see my code: method 2) not give the same values when categorizing relative humidity into low, medium, and high?

RGB colors (red=1, green=0.2, blue=0)—it is hard to wrap my head around the exact quantities to get the color I want. I would much rather just specify the actual colors (i.e. “aquamarine” or the HEX code “#8dd3c7”).

Can you use labeller(as\_labeller) under geom\_boxplot as well, in addition to under facet\_grid?

**What is a way you can apply the material in this lesson towards your research or area of study?**

I am constantly battling ggplot because it treats variables that I want as factors not as factors, and because it does not order them and label them the way I want when plotting. Material from this application has solved all of those issues for me! Some plots require very specific coloring of boxplots and knowing that I can supply the vector of colors (in order that the boxplots appear) is helpful. For example, coloring one boxplot white and all of the rest green.

**What are some things you would like to learn related to, but not covered in, this lesson?**

Other ways boxplots could look

How to change the shading of the facet grid titles (default is grey shading)

**Lesson 6-Boxplots I**

**What was your level of comfort with the lesson/application?**

In the prior class reflection, I asked for more for loops, and I got that in this lesson! I was the most comfortable with the plotting (boxplots, violin plots) and plotting components (adding whiskers, annotating). I was also comfortable with subsetting and extracting values from a data frame. I am getting slightly more experienced with for loops, but they still intimidate me a little bit. I think I just need to do lots and lots of for loops.

**What areas of the lesson/application confused or still confuses you?**

When else will we use stat\_boxplot

Are annotate (geom=”text”) and geom\_text(aes()) equivalent? Reason for using one over the other?

In lesson 6, you had code to calculate the median values for North and South directions. In my own research, I conduct similar operations but the code is more condensed. For example, for the median value:

A=median(weatherData$changeMaxTemp[weatherData$windDir=="North"])

Is the code I have written I have bad practice? Should I be aiming to structure my code like yours:

northVals=which(weatherData$windDir == "North");

weatherData[northVals, "changeMaxTemp"]

northMed = median(weatherData[northVals,"changeMaxTemp"], na.rm=TRUE);

the extension: condensed for loop (many for loops in one). I followed the code and understand superficially.

**What is a way you can apply the material in this lesson towards your research or area of study?**

Similar as last week, the for loops, data indexing, and data binning/categorizing was useful and I am going to use this code frequently for my own research. I do plot lots of boxplots for my manuscripts, and this week, I learned how to modify the appearance of outlier values and how to annotate the plot, which I have wanted to learn how to do.

**What are some things you would like to learn related to, but not covered in, this lesson?**

for loops to conduct the same mathematical operation over multiple columns (e.g. converting 3 columns of data from F to C)

how to modify the appearance of certain outlier values, not all outlier values in a plot

**Lesson 5-Histograms**

**What was your level of comfort with the lesson/application?**

Of all of the lessons, I was the least comfortable with this one, but the content was still familiar enough that I knew I would be able to complete the lesson and application. I had worked with stacked plots before, and knew about moving around legends but I had not worked with for loops or the which function before, so that was new.

**What areas of the lesson/application confused or still confuses you?**

When to use & vs. &&

I added a script file to your ***scripts*** folder that talks about this

facet\_grid vs. facet\_wrap

Will need to look into this…

for loops: I understand them at their most basic, but need more practice constructing for loops in different ways to feel comfortable using them for my own research. Perhaps more examples and more applications of the loops.

We will get to an example in class

**What is a way you can apply the material in this lesson towards your research or area of study?**

The for loops will be very handy because often, I have meta data on my biological samples that I need to categorize/bin in different ways. Most of the tasks I do in R involve data indexing and sub-setting, so having learned the which function will also be useful. I do not really use histograms for my research but anticipate using stacked histograms to visualize microbial community data.

**What are some things you would like to learn related to, but not covered in, this lesson?**

How to add text to the inside of the plot (e.g. within the plot borders)

annotate(geom="text") – more here: <https://ggplot2.tidyverse.org/reference/annotate.html>

for loops to conduct the same mathematical operation over multiple columns (e.g. converting 3 columns of data from F to C)

I added a script file to your ***scripts*** folder that talks about this

**Lesson 4-Data objects and canvas styles**

**What was your level of comfort with the lesson/application?**

I was comfortable with both the lesson and application, although there were definitely things I had to read multiple times to make sure I understood, particularly the different ways to format Dates in R and what the different plot/panel parameters mean in ggplot.

**What areas of the lesson/application confused or still confuses you?**

What is a *tibble* exactly and when are they more appropriate to use over data frames

This is a philosophical discussion! They supposedly are modern versions of the data frame. I have not used them and people whom I work with have not found any convincing reason to switch to a tibble – unless you are already fully in the tidyverse.

**What is a way you can apply the material in this lesson towards your research or area of study?**

This is very useful for my research because I have to produce many visually appealing, colorful, publication-quality figures. Usually that means I have to make the axes and axis tick marks large, the plot border thick, and the plot elements colorful. My research deals with formatting Dates as well, and in the past, this has caused me many headaches but I believe I finally understand. I do a lot of data wrangling so knowing how to access different data frame values in different ways was helpful. I did not know that theDate = weatherData[["date"]] was the same as weatherData$theDate.

**What are some things you would like to learn related to, but not covered in, this lesson?**

How to add text to the inside of the plot (e.g. within the plot borders)

for loops to conduct a mathematical operation over multiple columns (e.g. converting 3 columns from F to C)

**Lesson 3-Plot Styles**

**What was your level of comfort with the lesson/application?**

I felt slightly more overwhelmed during this lesson compared to the previous, but it was very fun. The ggplot code is getting more busy, but the different components are making more sense now! I was able to follow along and complete the lesson and the application just fine.

**What areas of the lesson/application confused or still confuses you?**

In axis.title.x=element\_text(size=14, color="black",face="bold"); are there other parts instead of element\_text? Couldn’t ggplot have been designed so that we wouldn’t need to add element\_text to every single axis title/label component?

**What is a way you can apply the material in this lesson towards your research or area of study?**

Yes, for all of the plots I make, I have to set the shape, colors, and sizes of all of the text in the plot. I mostly always set the color of the x and y axis to black, but I do bold them and make the font large. I also have done scatter plots with regression lines but did not know how to change the color or shading of line. Now, I know and will use that code. Something important I learned is to set theme\_bw() before modifying the theme later; I have not done this in the past and it has caused much frustration.

**What are some things you would like to learn related to, but not covered in, this lesson?**

Which other ggplot components need to be in a certain order? For example, theme\_bw() needs to be set before theme(). Are there other ggplot components that need to be set in a certain order?

Center Plot title

Adding confidence intervals or standard errors to the regression line

Set dashed lines to denote upper and lower limits on y-axis

**Lesson 2-Components**

**What was your level of comfort with the lesson/application?**

I was very comfortable with the commands and skills covered in this lesson, from reading/saving as .CSV file, to plotting a scatterplot in ggplot. I knew how to set custom x and y tick marks and how to specify a plot title. I am sure as the course advances, plot components will get more challenging.

**What areas of the lesson/application confused or still confuses you?**

Not many in this lesson

**What is a way you can apply the material in this lesson towards your research or area of study?**

Absolutely. For my data, I create a lot of boxplots and a few scatterplots to showcase, for example, microbial community Shannon diversity vs. individual’s age, and using the code we created will come in handy. The code we created is more clean, structured, and easier to modify/navigate, which will save me many headaches down the road. Having the code written this way will also help me diagnose and solve R errors more quickly.

**What are some things you would like to learn related to, but not covered in, this lesson?**

What are the other components of mapping, besides aes (aesthetics)

Why certain mapping aesthetics go under geom\_point vs in ggplot

plotData = **ggplot**( data=deaths ) +

**geom\_point**(mapping=aes(x=time, y=accdeaths))

How to access a data file (.csv or .txt) directly from a URL