**DataViz**

**Lab: ggplot2: qplot and 1-variable ggplot**

Purpose:

* To explore some of the geoms, use of color, and other variations of qplot and ggplot. Investigate what types of “geoms” are available, and for which types of plots they might be appropriate.
* To explore the use of special variables (stat and e.g., ..density..)

**qplot:** Start with the basic scatter plot:

qplot(x=mpg, y=wt, data = mtcars, geom="point")

Now let's look at other geoms: <http://ggplot2.tidyverse.org/reference/>

1. Try using some of these geoms, and comment on them. (When you use the geom, the syntax is not geom\_jitter, but geom=”jitter”, at least in the context of our script.
* Jitter: What is it? Did it make any difference here? (I will explain jitter to you, but try it first and see if you can understand what it happening in the graph.)
* Tile: and add a color of red.
	+ From the documentation, it looks like “rect” and “tile” are pretty much the same thing. So try rect. It doesn’t work. Why not? Do you understand the error message?
* Hex. It works. But it doesn’t really inform much. Why is this not much better than a regular scatterplot? Hint: It’s not because of the cute little hex shapes, but related to the color. Why might a different dataset be better served with a geom\_hex.
* Try a 2d density plot. Isn’t that pretty? What on earth does it mean? (Can you interpret it?)
	+ Try a regular, 1-d density plot with just the mpg variable. Now can you see what the density plot means?
* Try crossbar.
	+ It doesn’t work.
	+ Add the params that might make it work. (What did you add?)
	+ What do you make of that chart?
	+ 🡪This geom is better [when layered with other geoms](http://ggplot2.tidyverse.org/reference/geom_linerange.html), and we can revisit it again when using ggplot2’s gpplot function.
* Try a bar chart. Whoa! Look at that error. What does that mean? We’ve run into something similar before. Recall what a bar chart is supposed to represent. How can you change this?
	+ 🡪BTW, we’ll see more of the “stat” relationship with “geom” when we study ggplot2. Each geom has a default “stat”. Sometimes, you can just specify the stat, or change the stat. Sometimes, you can get the same chart by specifying the geom or the stat.
	+ So make the bar chart work! (use only 1-d. The second dimension is already defaulted to “count”.)
1. In the statement below, what does “smooth” mean?

 qplot(mpg, wt, data=mtcars,

 geom =c ("point", "smooth"))

**ggplot:** Start with a basic density plot:

1. ggplot(wdata) + stat\_density(aes(x = weight))
* Give it a red outline and a blue fill. Is this dependent on any other variable? Or global to the entire graph?
* Will it work the same if you use geom density instead of stat?

Consider:

ggplot (mtcars, aes(cyl)) + geom\_bar()

* Try to give this a red outline and a blue fill, using a similar syntax as above for color and fill. It doesn’t work.
* Now try to “layer” it similarly to the previous example. That works…
1. Consider from our lab:

a<-ggplot(wdata,aes(x=weight))

a + geom\_density()

a + geom\_density(color = "black", fill = "gray")+

 geom\_vline(aes(xintercept=mean(weight)),

 color="#fc4e07", linetype = "dashed", size = 1)

Briefly explain each line of “layering” and what is being added with each line.

1. Consider from our lab:

a + geom\_density(aes(color=sex))

How is that different from: a + geom\_density(color="blue")

Explain the result of this (why??): a + geom\_density(aes(color="blue"))

1. Consider from our lab, below. Explain what each part of the stmt is doing. Be careful with ..density..

a+ geom\_histogram(aes(y = ..density..),color = "blue", fill = "yellow") +

 geom\_density(alpha = 0.2, fill = "grey")

1. What is the difference between a color (or fill) specification inside the aes parentheses, and a color (or fill) specification outside the aes parentheses? Give an example either from our lab, or one that you have explored.