Homework 3

Practical R

## Supplemental Readings

1. [R4DS Chapter 3](https://r4ds.had.co.nz/data-visualisation.html)
2. [DV Chapter 3](https://socviz.co/makeplot.html#makeplot)

If you look in Resources, you’ll see links to the complete books: R for Data Science (R4DS) and Data Visualization: A Practical Introduction (DV)

#### Question 1

Read the data from the file gapminder\_w.csv. This file is not in tidy format

1. Identify how to make this data tidy. Sketch this out
2. Use pivot\_longer and pivot\_wider in the appropriate order to make this data tidy
3. Save the tidy version of this data set as gapm
4. Read the data from the file gapminder.csv. This should exactly match the data from part c. If you get stuck here, load the data from gapminder.csv as gapm.

## Visualizing the gapminder dataset

#### Question 2

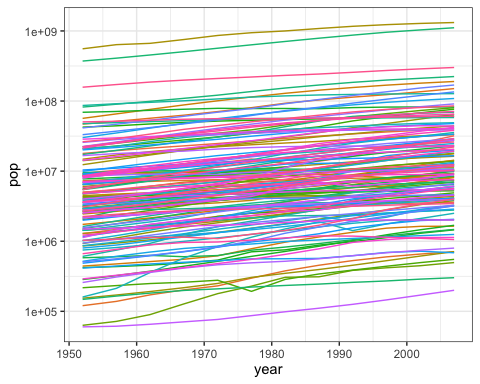
Let’s look at population growth over time.

ggplot(data = gapm,  
 mapping = aes(x = year, y = pop))+  
 geom\_line(color = country, show.legend = F)+  
 scale\_y\_log10()

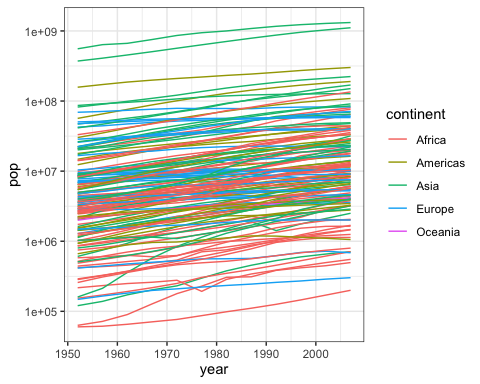
## Error in layer(data = data, mapping = mapping, stat = stat, geom = GeomLine, : object 'country' not found

Oops!! This doesn’t quite work.

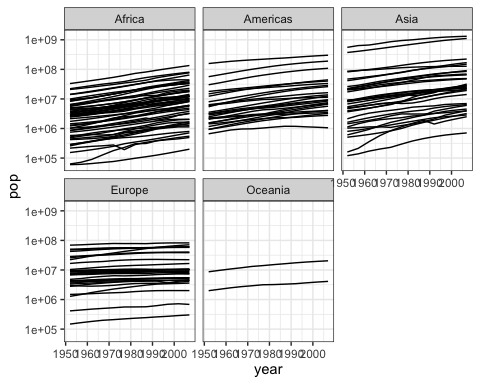
**a.** Update the code above so that the following plot is rendered.



**b.** This has too many colors, in my opinion. Lets just color the lines by continent rather than country, and allow the legend to also be drawn. The desired plot is as follows:



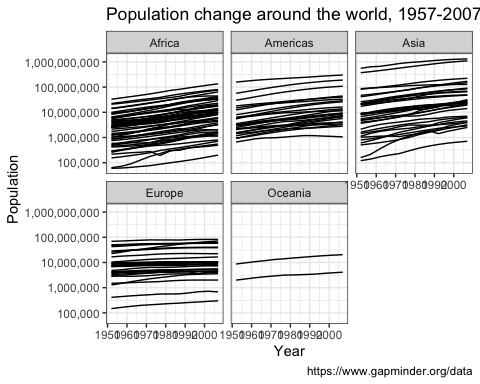
*Hint: you need to use both group and color, possibly with different variables in the data set*

**c.** This is still too messy. Let’s just separate out the continents and see the population growth by country. 

**d.** Okay. Let’s go with this as the final graph. Let’s clean it up with proper labels on the x- and y-axes, changing the numbers on the y-axis from scientific notation to usual numbers, and adding a title. The layers you’ll need to add involve

* labs(x = \_\_\_, y = \_\_\_) for the x and y labels
* adding an option to the scale\_y\_log10 function (see the lecture)
* ggtitle to add a title
* You can also change the theme to something else if you like.

The final picture should look something like this (you can decide on your own labels and title). Look at the documnetation for the layers and functions you use to see what options are available to make the plot look like mine.

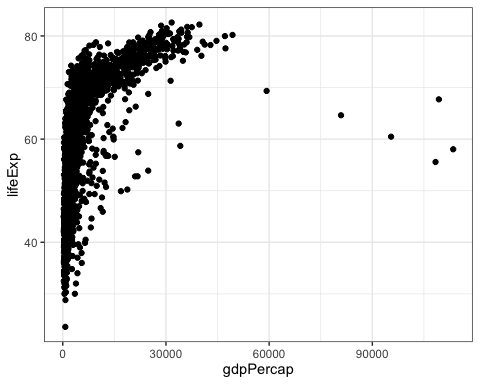


#### Question 3

Now, let’s look at the relationship between per capita GDP of a country and the life expectancy in the country in the gapminder dataset. The final visualization will mimic (with more limited data) a famous visualization that was shown in the Hans Rosling video [*The Joy of Stats*](https://www.youtube.com/watch?v=jbkSRLYSojo)

Let’s start out by plotting the life expectancy against per capita GDP in a scatter plot

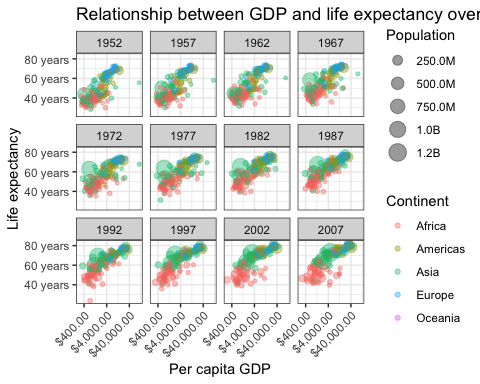
ggplot(gapm, aes(x = gdpPercap, y = lifeExp)) +   
 geom\_point()



We’re going to successively build up this visualization by adding (+) layers to the plot. Don’t create a legend for this plot

**a.** Change the x-axis to display on the log scale  
**b.** Change the size of the points to reflect the population  
**c.** Color the inside of each point by continent  
**d.** Separate the graphs out by year (use facet\_wrap)

The result should look something like this:



I will be impressed if you can get the annotation on the axis like mine. A good resource for this is [this page](http://www.sthda.com/english/wiki/ggplot2-essentials), scrolling down to the section titled “Graphical parameters”. You should also see the documentation of the functions in the scales package (In the lower right pane, there is a pane marked “Packages”. There you can search the name of a package and see all the available functions and then click to get the individual function documentations)