HW5 - Make Data Move

Marie Mortensen

05/10/2020

# Explore the recent global developments with R

Today, you will load a filtered gapminder dataset - with a subset of data on global development from 1952 - 2007 in increments of 5 years - to capture the period between the Second World War and the Global Financial Crisis.

**Your task: Explore the data and visualise it in both static and animated ways, providing answers and solutions to 7 questions/tasks below.**

## Get the necessary packages

First, start with installing the relevant packages ‘tidyverse’, ‘gganimate’, and ‘gapminder’.

## Parsed with column specification:  
## cols(  
## key\_ID = col\_double(),  
## village = col\_character(),  
## interview\_date = col\_datetime(format = ""),  
## no\_membrs = col\_double(),  
## years\_liv = col\_double(),  
## respondent\_wall\_type = col\_character(),  
## rooms = col\_double(),  
## memb\_assoc = col\_character(),  
## affect\_conflicts = col\_character(),  
## liv\_count = col\_double(),  
## items\_owned = col\_character(),  
## no\_meals = col\_double(),  
## months\_lack\_food = col\_character(),  
## instanceID = col\_character()  
## )

## Look at the data

First, see which specific years are actually represented in the dataset and what variables are being recorded for each country. Note that when you run the cell below, Rmarkdown will give you two results - one for each line - that you can flip between.

unique(gapminder$year)

## [1] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007

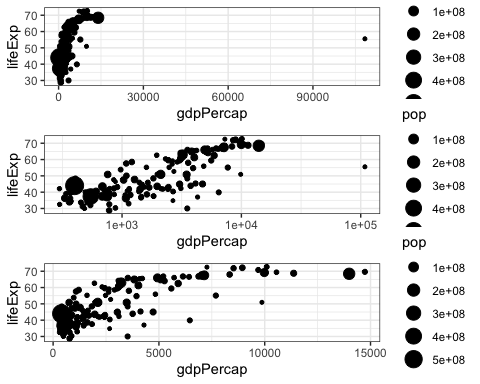
head(gapminder)

## # A tibble: 6 x 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Afghanistan Asia 1952 28.8 8425333 779.  
## 2 Afghanistan Asia 1957 30.3 9240934 821.  
## 3 Afghanistan Asia 1962 32.0 10267083 853.  
## 4 Afghanistan Asia 1967 34.0 11537966 836.  
## 5 Afghanistan Asia 1972 36.1 13079460 740.  
## 6 Afghanistan Asia 1977 38.4 14880372 786.

The dataset contains information on each country in the sampled year, its continent, life expectancy, population, and GDP per capita.

Let’s plot all the countries in 1952.

theme\_set(theme\_bw()) # set theme to white background for better visibility  
  
gap\_log <- gapminder %>% filter(year == 1952) %>% ggplot(aes(gdpPercap, lifeExp, size = pop)) +  
 geom\_point() +  
 scale\_x\_log10()   
  
gap\_normal <- gapminder %>% filter(year == 1952) %>% ggplot(aes(gdpPercap, lifeExp, size = pop)) +  
 geom\_point()  
  
gap\_no\_out <- gapminder %>% filter(year == 1952, gdpPercap<90000) %>% ggplot(aes(gdpPercap, lifeExp, size = pop)) + geom\_point()  
  
gridExtra::grid.arrange(gap\_normal, gap\_log, gap\_no\_out)



#filtering richest country  
gapminder %>% filter(year == 1952, gdpPercap>80000) #Kuwait

## # A tibble: 1 x 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Kuwait Asia 1952 55.6 160000 108382.

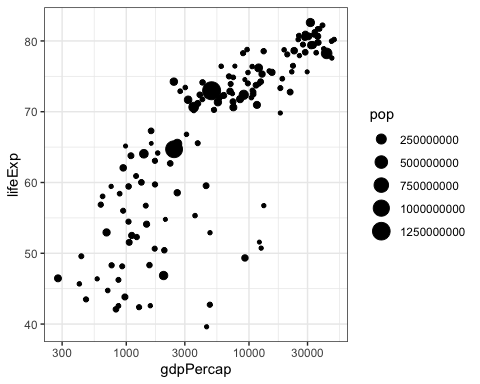
We see an interesting spread with an outlier to the right. Answer the following questions, please:

Q1. Why does it make sense to have a log10 scale on x axis? Changing the x-axis to a log scale helps in situations where there is a large range of values on either the x- or y-axis. In this case there is an outlier that extends the range on the x-axis to a gdp of 90000. All other datapoints are clustered between a gdp of 0 and 15000 and it is difficult interpreting the correlation of gdp and life expectancy. the logarithm of gdp reduces the range and makes an interpretation possible. When we try to remove the outlier by filtering out gdp above 90000 we also see that the plot looks like the semi-log graph but with actual gdp values.

Q2. What country is the richest in 1952 (far right on x axis)? Filtering out countries with a gdp over 80000 shows tha Kuwait is the richest country in 1952.

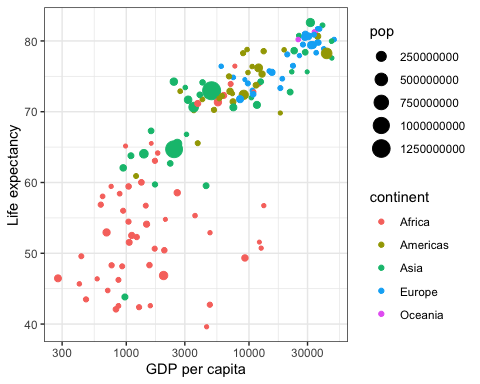
You can generate a similar plot for 2007 and compare the differences

options(scipen = 1000)  
gapminder %>% subset(year == 2007) %>% ggplot(aes(gdpPercap, lifeExp, size = pop)) +  
 geom\_point() +  
 scale\_x\_log10()

 The black bubbles are a bit hard to read, the comparison would be easier with a bit more visual differentiation.

Q3. Can you differentiate the continents by color and fix the axis labels?

gapminder %>% subset(year == 2007) %>% ggplot(aes(gdpPercap, lifeExp, size = pop, color = continent)) + #adding continent to a color argument  
 geom\_point() +  
 scale\_x\_log10() +  
 labs(x ="GDP per capita", y = "Life expectancy") #adding labels



Q4. What are the five richest countries in the world in 2007? To see the five richest countries in 2007 I select the columns and gdp per capita and afterwards arrange them in a descending order. This shows that Norway, Kuwait, Singapore, US and Ireland are the five richest countries in 2007.

gapminder %>%   
 subset(year == 2007) %>%   
 select(country, gdpPercap) %>%   
 arrange(desc(gdpPercap))

## # A tibble: 142 x 2  
## country gdpPercap  
## <fct> <dbl>  
## 1 Norway 49357.  
## 2 Kuwait 47307.  
## 3 Singapore 47143.  
## 4 United States 42952.  
## 5 Ireland 40676.  
## 6 Hong Kong, China 39725.  
## 7 Switzerland 37506.  
## 8 Netherlands 36798.  
## 9 Canada 36319.  
## 10 Iceland 36181.  
## # … with 132 more rows

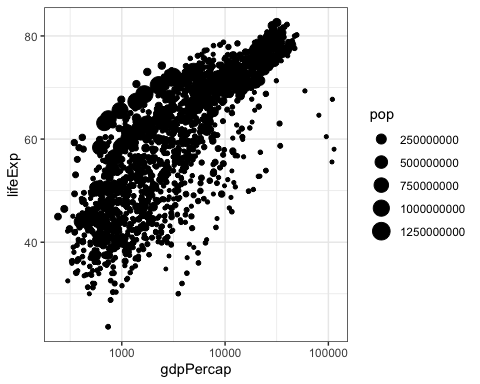
## Make it move!

The comparison would be easier if we had the two graphs together, animated. We have a lovely tool in R to do this: the gganimate package. And there are two ways of animating the gapminder ggplot.

### Option 1: Animate using transition\_states()

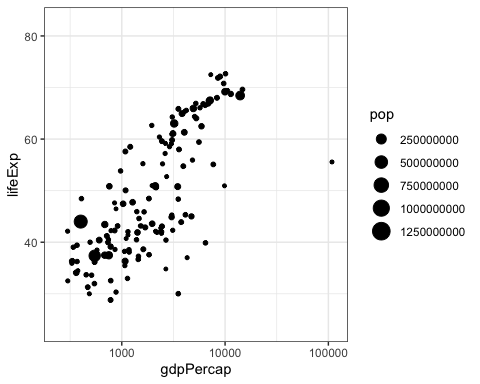
The first step is to create the object-to-be-animated

anim <- gapminder %>% ggplot(aes(gdpPercap, lifeExp, size = pop)) +  
 geom\_point() +  
 scale\_x\_log10() # convert x to log scale  
anim



This plot collates all the points across time. The next step is to split it into years and animate it. This may take some time, depending on the processing power of your computer (and other things you are asking it to do). Beware that the animation might appear in the ‘Viewer’ pane, not in this rmd preview. You need to knit the document to get the viz inside an html file.

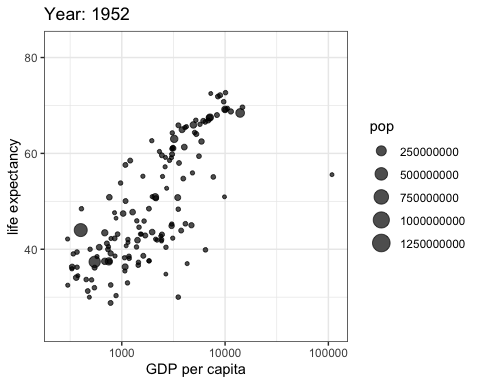
anim + transition\_states(year,   
 transition\_length = 1,  
 state\_length = 1)

 Notice how the animation moves jerkily, ‘jumping’ from one year to the next 12 times in total. This is a bit clunky, which is why it’s good we have another option.

### Option 2 Animate using transition\_time()

This option smoothes the transition between different ‘frames’, because it interpolates and adds transitional years where there are gaps in the timeseries data.

anim2 <- ggplot(gapminder, aes(gdpPercap, lifeExp, size = pop)) +  
 geom\_point(alpha = 0.7) +  
 # scale\_colour\_manual(values = country\_colors) +  
 scale\_x\_log10()+   
 labs(title = 'Year: {frame\_time}', x = 'GDP per capita', y = 'life expectancy') +  
 transition\_time(year)  
anim2



The much smoother movement in Option 2 will be much more noticeable if you add a title to the chart, that will page through the years corresponding to each frame.

Q5 Can you add a title to one or both of the animations above that will change in sync with the animation? [hint: search labeling for transition\_states() and transition\_time() functions respectively] Yes, by adding “labs” argument and placing frame\_time in curly brackets the animation recognizes when to change year in the title. Here it is also possible adding other labels to the x axis and y axis.

Q6 Can you made the axes’ labels and units more readable? Consider expanding the abreviated lables as well as the scientific notation in the legend and x axis to whole numbers.[hint:search disabling scientific notation] adding options(scipen = 100) it is possible to change values from having scientific notation. x and y labels have been changed by adding them to the title argument as described above.

Q7 Come up with a question you want to answer using the gapminder data and write it down. Then, create a data visualisation that answers the question and explain how your visualization answers the question. (Example: you wish to see what was mean life expectancy across the continents in the year you were born versus your parents’ birth years). [hint: if you wish to have more data than is in the filtered gapminder, you can load either the gapminder\_unfiltered dataset and download more at <https://www.gapminder.org/data/> ]

How has life expectancy and gdp changed in Spain (where I have lived) compared to Denmark (where i live :D )?

anim3 <- gapminder %>% filter(country == "Spain" | country == "Denmark") %>% ggplot(aes(gdpPercap, lifeExp,color = country, size = pop)) +  
 geom\_point(alpha = 0.7) +  
 scale\_colour\_manual(values = country\_colors) +  
 labs(title = 'Year: {frame\_time}', x = 'GDP per capita', y = 'life expectancy') +  
 transition\_time(year)  
anim3

