Lab #2 - Gapminder Dataset

Econ 224

August 30th, 2018

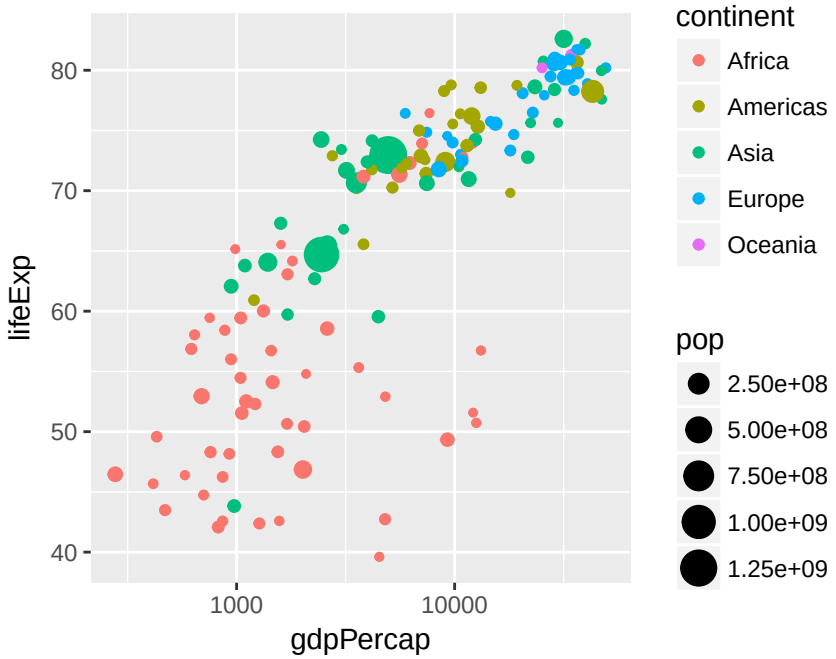
Introduction

Today we'll revisit the `gapminder` dataset and use it to introduce some more advanced features of `dplyr` and `ggplot2`, building on the material from our first lab. Before you begin, make sure that you have loaded the `tidyverse` and `gapminder` packages.

Faceting - Plotting multiple subsets at once

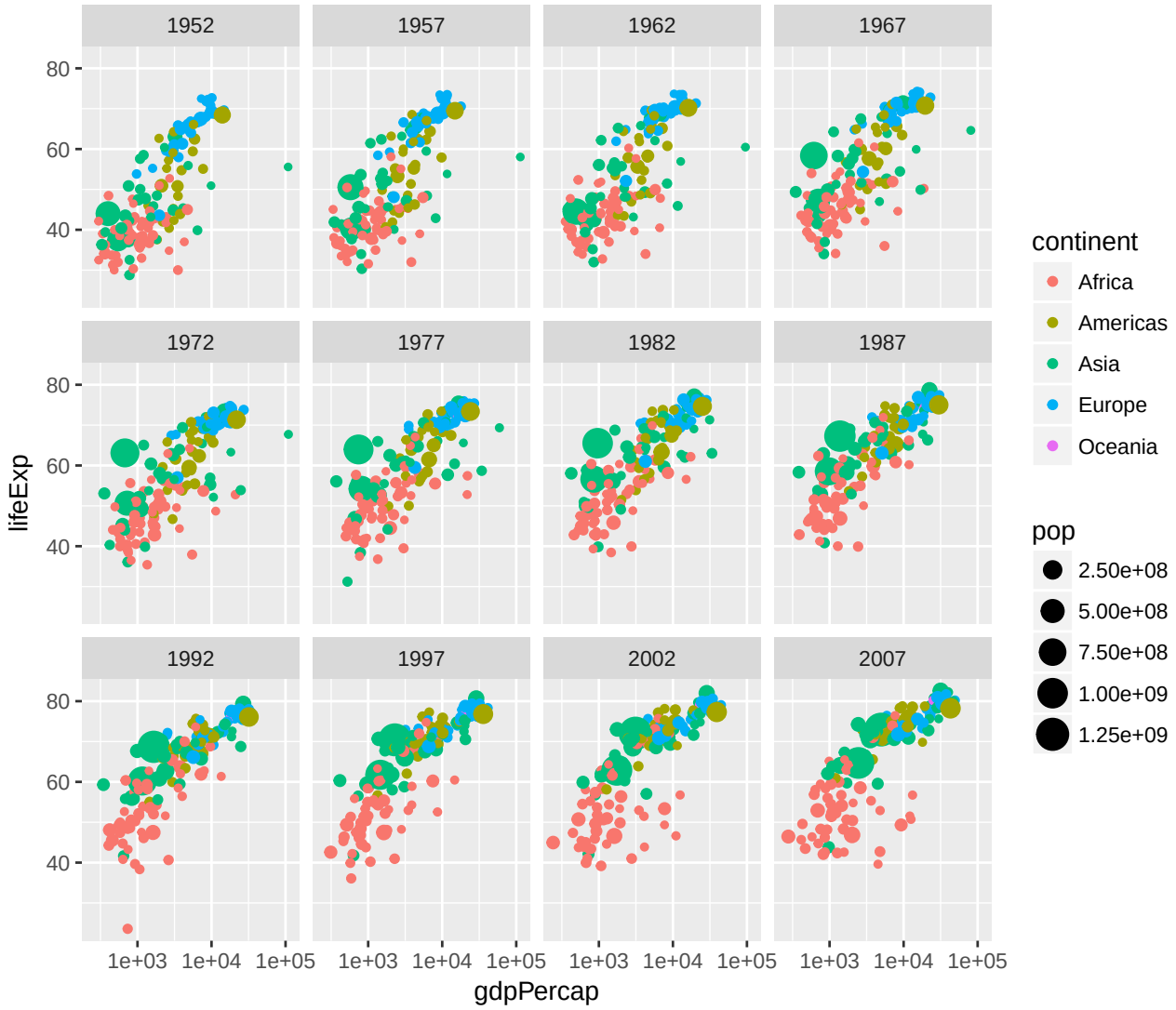
Let's pick up where we left off in lab #1, with a plot of GDP per capita and life expectancy in 2007:

```
gapminder_2007 <- gapminder %>%
  filter(year == 2007)
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color = continent, size = pop)) +
  scale_x_log10()
```



This is an easy way to make a plot for a single year. But what if you wanted to make the same plot for *every year* in the `gapminder` dataset? It would take a lot of copying-and-pasting of the preceding code chunk to accomplish this. Fortunately there's a much easier way: *faceting*. In `ggplot2` a *facet* is a subplot that corresponds to a subset of your dataset, for example the year 2007. We'll now use faceting to reproduce the plot from above for all the years in `gapminder` simultaneously:

```
ggplot(gapminder) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color = continent, size = pop)) +
  scale_x_log10() +
  facet_wrap(~ year)
```



Note the syntax here: in a similar way to how we added `scale_x_log10()` to plot on the log scale, we add `facet_wrap(~ year)` to facet by `year`. The tilde `~` is important: this has to precede the variable by which you want to facet.

Now that we understand how to produce it, let's take a closer look at this plot. Notice how this plot allows us to visualize five variables *simultaneously*. By looking at how the plots change over time, we see a pattern of increasing GDP per capita and life expectancy throughout the world between 1952 and 2007. Notice in particular the dramatic improvements in both variables in the Asian economies.

Exercise #1

1. What would happen if I were to run the following code? Explain briefly.

```
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color = continent, size = pop)) +
  scale_x_log10() +
  facet_wrap(~ year)
```

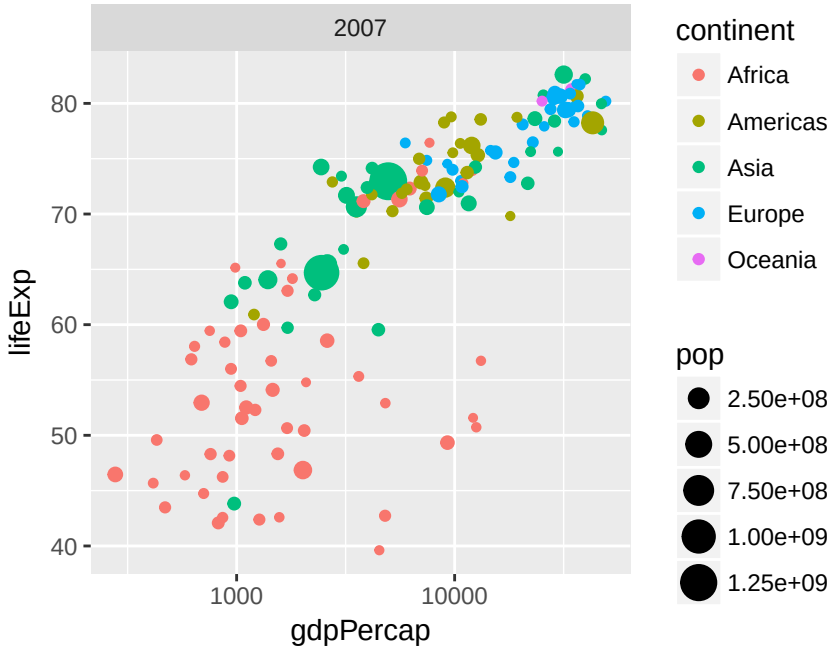
2. Make a scatterplot with data from `gapminder` for the year 1977. Your plot should be faceted by continent with GDP per capita on the log scale on the x-axis, life expectancy on the y-axis, and population indicated by the size of each point.
3. What would happen if you tried to facet by `pop`? Explain briefly.

Solution to Exercise #1

Write your code and solutions here

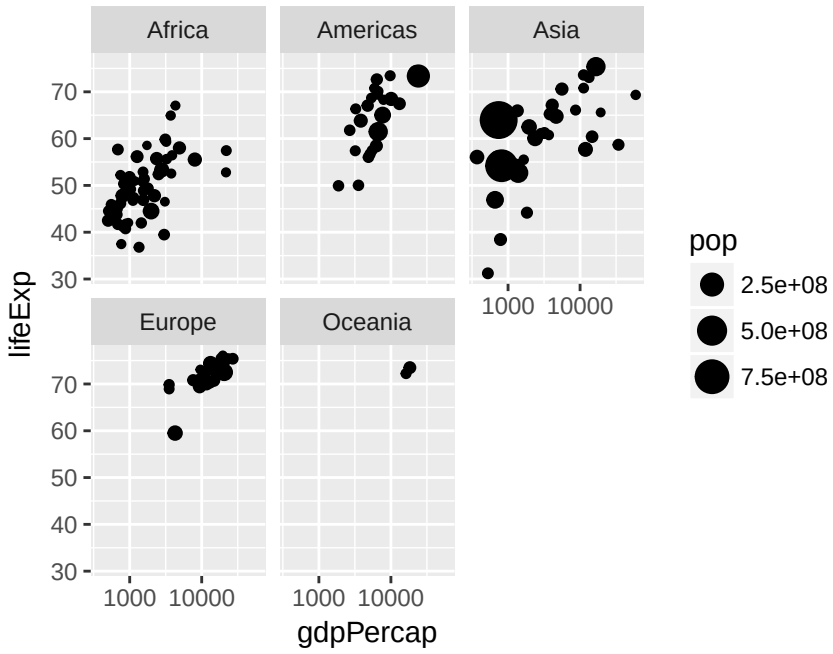
1. We'll only get one facet since the tibble `gapminder_2007` only has data for 2007:

```
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color = continent, size = pop)) +
  scale_x_log10() +
  facet_wrap(~ year)
```



2. Use the following code:

```
gapminder_1977 <- gapminder %>%
  filter(year == 1977)
ggplot(gapminder_1977) +
  geom_point(aes(x = gdpPercap, y = lifeExp, size = pop)) +
  scale_x_log10() +
  facet_wrap(~ continent)
```



3. You'll get something crazy if you try this. Population is continuous rather than categorical so every country has a different value for this variable. You'll end up with one plot for every country, containing a single point:

```
gapminder_1977 <- gapminder %>%
  filter(year == 1977)
ggplot(gapminder_1977) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color = continent)) +
  scale_x_log10() +
  facet_wrap(~ pop)
```