

Fundamentals of R, Block 3 Practical

Block 3 - Practical Visualizations

Henrique Sposito and Livio Muller-Silva

R Markdown: another way to store code

Markdown is a simple formatting syntax for authoring HTML, PDF, and Word documents.

Creating an R Markdown document is just like an R script, you just have to click the new document button and select R Markdown from the options.

Markdown allows you to mix chunks of code (in light grey) with text and export a document with your code, text, and plots.

You can embed an R code chunk like this:

In the case above, we are just adjusting the setup for the document and loading some packages for our R Markdown document.

This is the best resource for information on R Markdown!

Tip: remember to always load the data at the beginning so your Rmarkdown will knit properly! Adding your data and Rmd files on the same folder also helps with path issues when knitting.

Basics:

Section headers work with #:

First-level header

Second-level header

Third-level header

For changing text styles use *:

Italics

Bold

Italics and bold

For inserting R code click on the **C** button above or use Cmd + Option + I on MAC (for Windows: Ctrl + Alt + I).

```
## [1] "R Markdown is awesome"
```

```
print("hello")
```

```
## [1] "hello"
```

Code chunks can be evaluated (should code be run?), included (should the code displayed in knitted document?), and much more. `rmarkdown`, as a tidyverse package, also has a cheat sheet!

When you click the **Knit** button a document in HTML or PDF can be generated that includes both content as well as any embedded R code chunks within the document.

Lastly, R Markdown can be further used to create presentations in R (as the ones we use in class, see the `xaringan` package) or even to write your Master's thesis (check out `iheidown`).

If we want to knit this to PDF, we need to download a Latex engine first:

Visualizations

Setting up the Gapminder data

```
##           country      continent      year      lifeExp
## Afghanistan: 12 Africa :624 Min. :1952 Min. :23.60
## Albania : 12 Americas:300 1st Qu.:1966 1st Qu.:48.20
## Algeria : 12 Asia :396 Median :1980 Median :60.71
## Angola : 12 Europe :360 Mean :1980 Mean :59.47
## Argentina : 12 Oceania : 24 3rd Qu.:1993 3rd Qu.:70.85
## Australia : 12 Max. :2007 Max. :82.60
## (Other) :1632
##           pop      gdpPercap
## Min. : 60011 Min. : 241.2
## 1st Qu.: 2793664 1st Qu.: 1202.1
## Median : 7023596 Median : 3531.8
## Mean : 29601212 Mean : 7215.3
## 3rd Qu.: 19585222 3rd Qu.: 9325.5
## Max. :1318683096 Max. :113523.1
##
```

Before we start, the `ggplot2` book is a great source for you to learn the details of visualizations in R (and the book was written using an R Markdown).

Line plots: The evolution of life expectancy

To create line plots in `ggplot2` we use the `geom_line()` function.

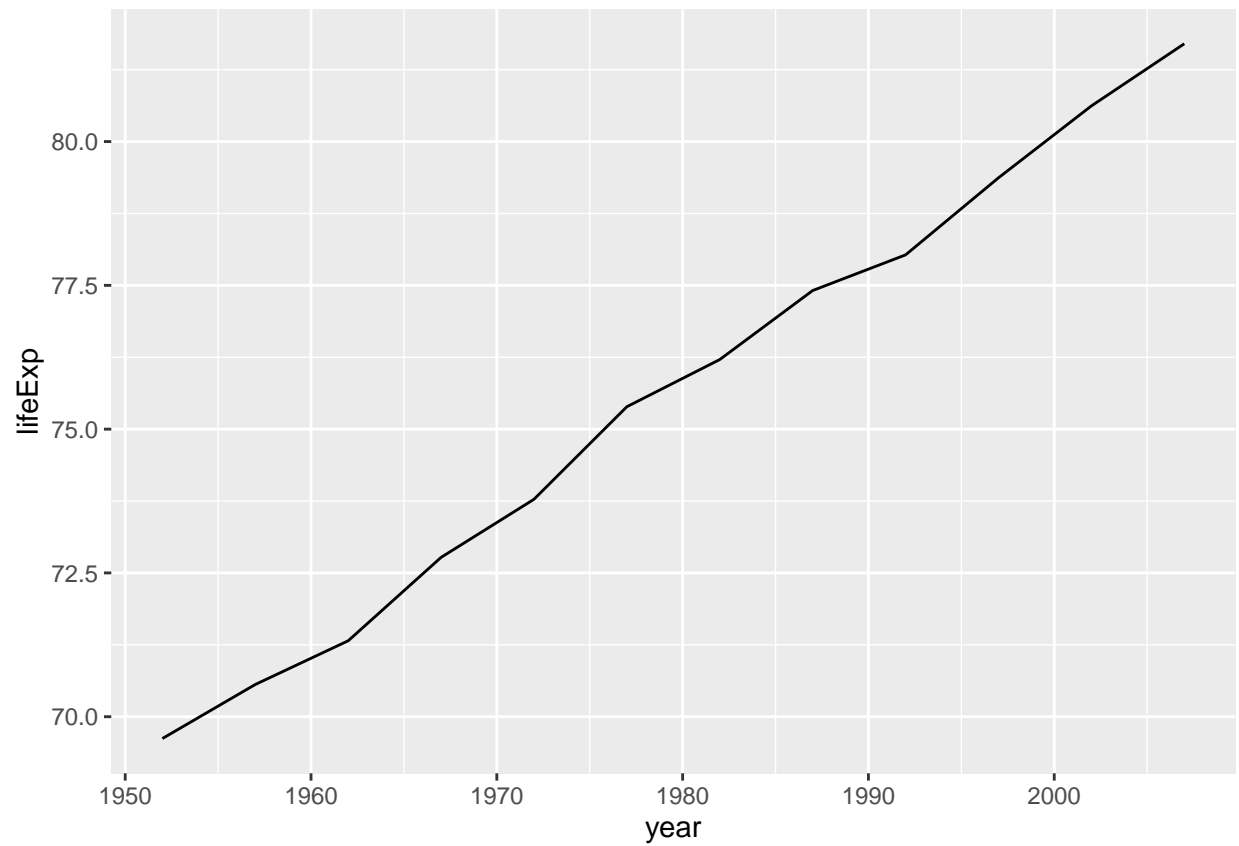
What are line plots good for?

Let's plot life expectancy in time, for Switzerland!

```
##
## Attaching package: 'dplyr'

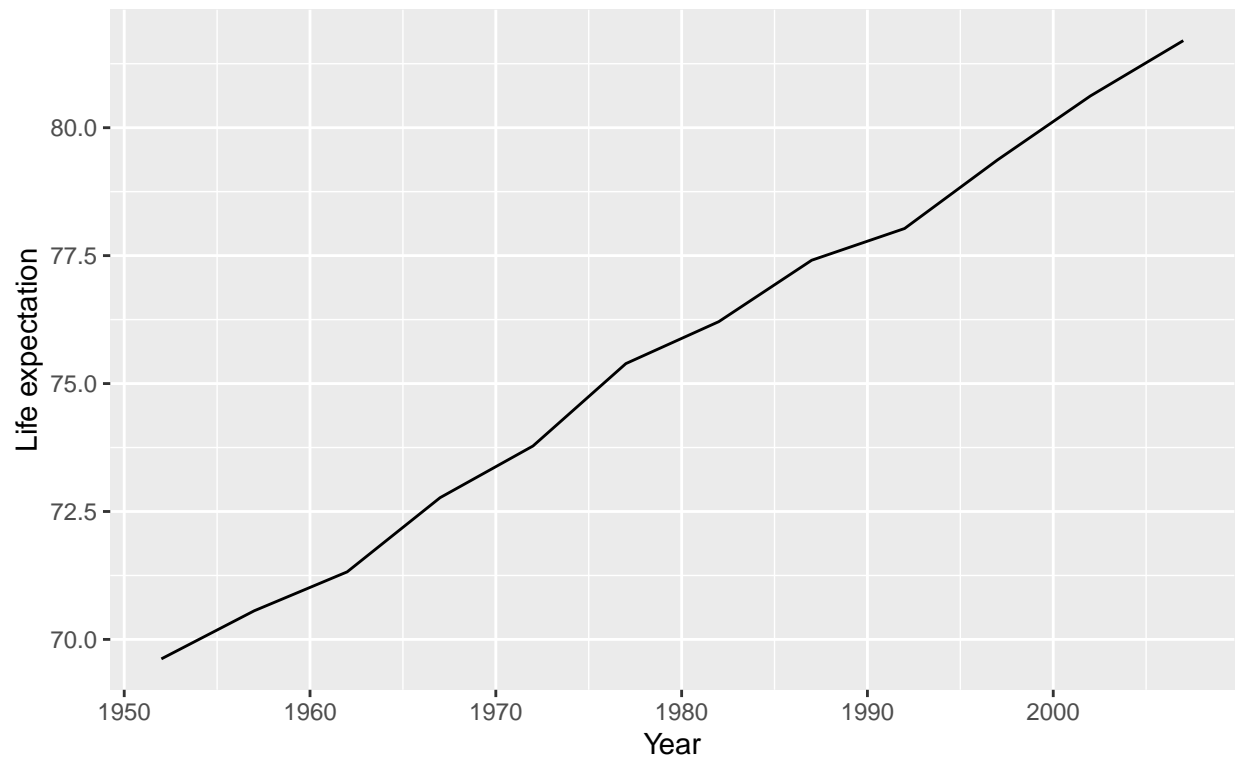
## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```



How can we improve this?

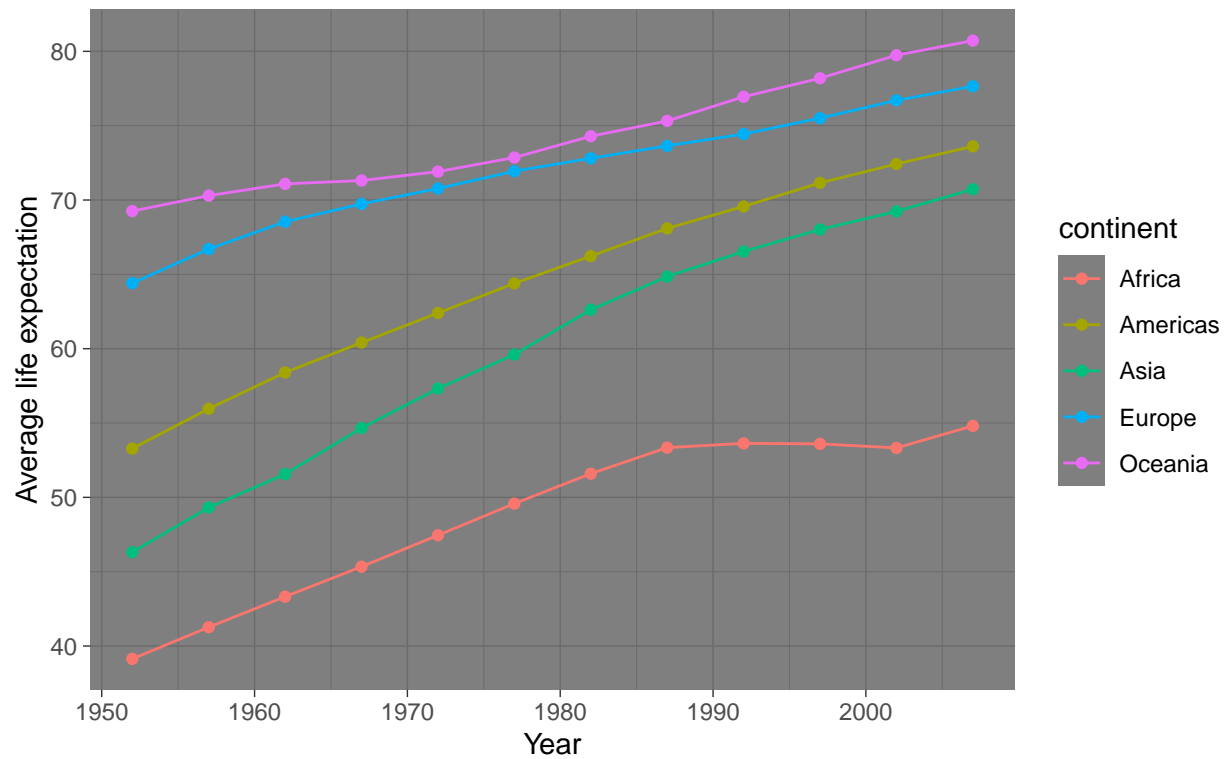
Life expectancy in Switzerland from 1957 to 2007



Source: Gapminder

Do you think that life expectancy increased for across all continents in time?

Life expectancy by continent from 1957 to 2007



Source: Gapminder

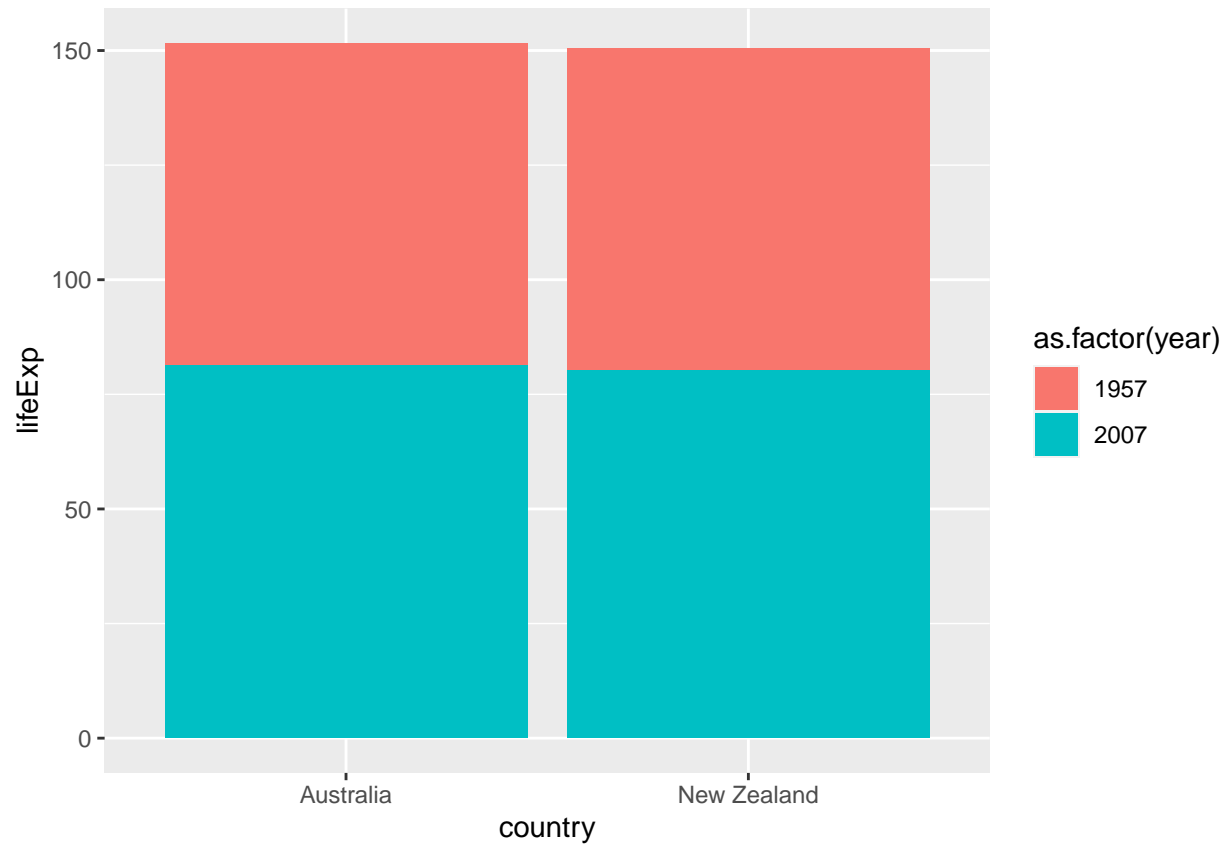
Could you make the same line plots for GDP per capita across continents in time?

Bar plots: Life expectancy from 1957 to 2007 across continents

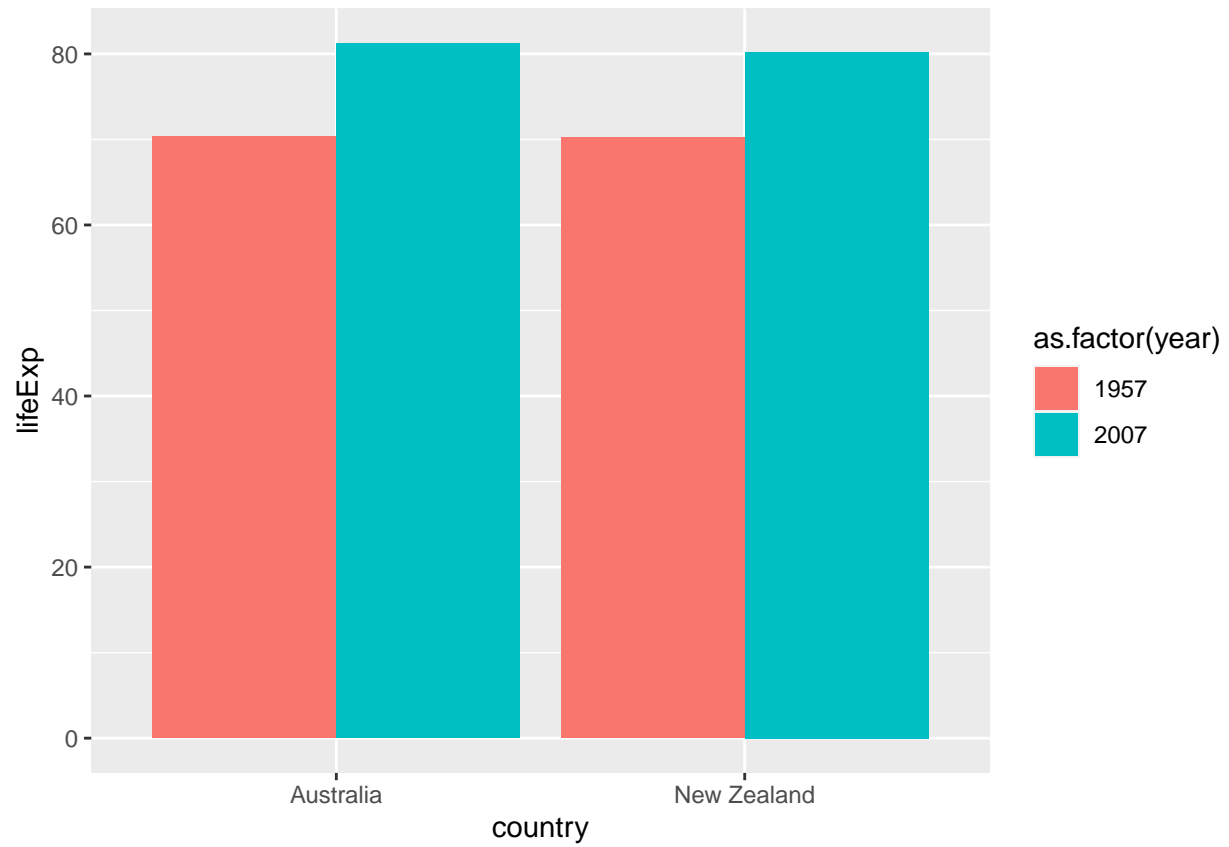
To create bar plot in ggplot2 we use the `geom_col()` function.

What are bar plots good for?

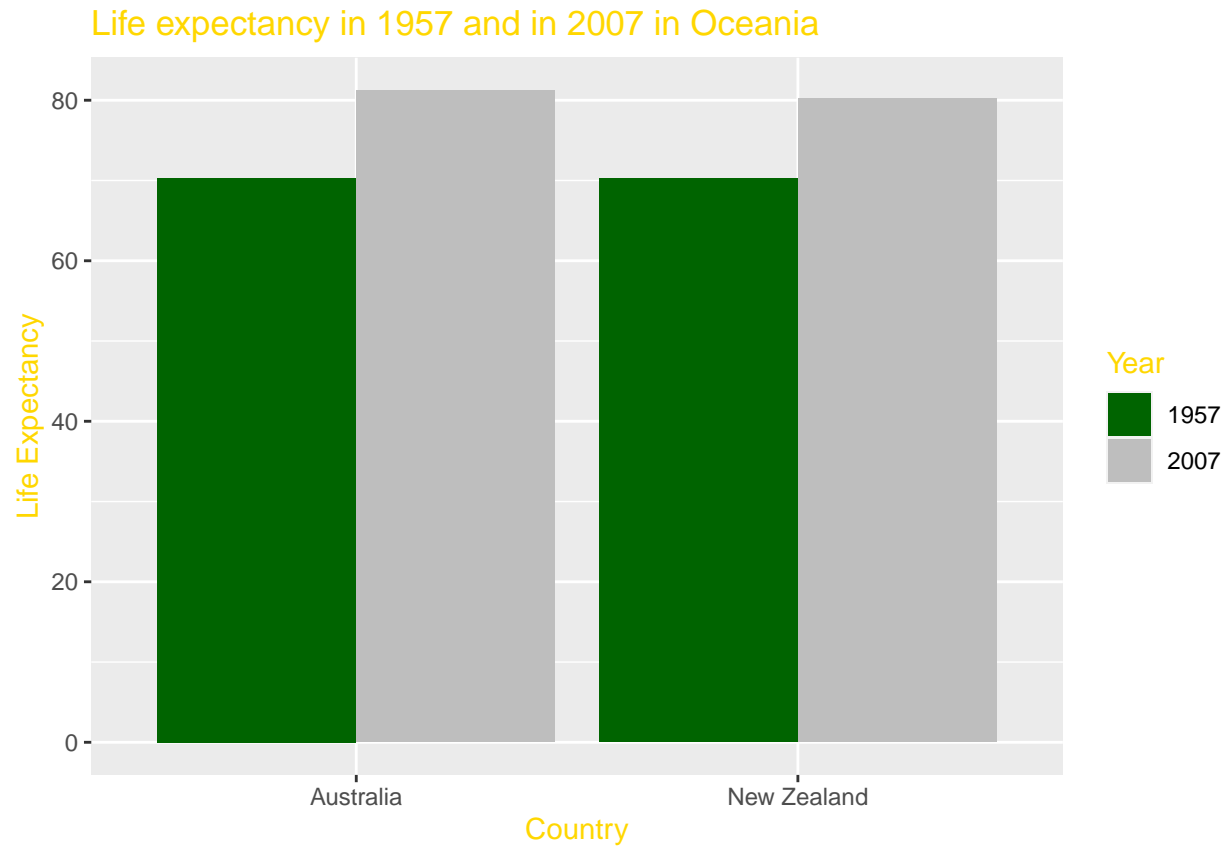
Let's start with a simple bar plot, one continent at two different points in time.



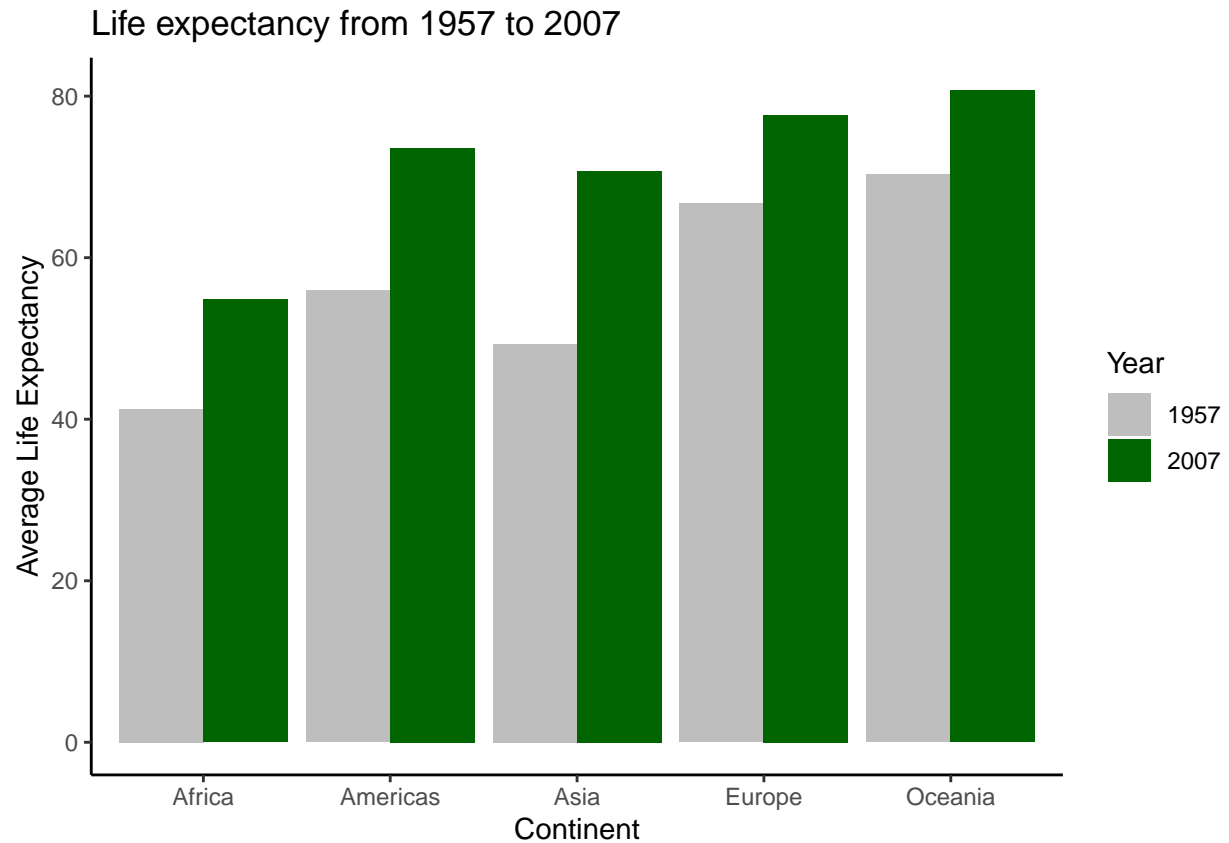
What is the issue with this plot?



This looks nicer, but there are several improvements we can still make!



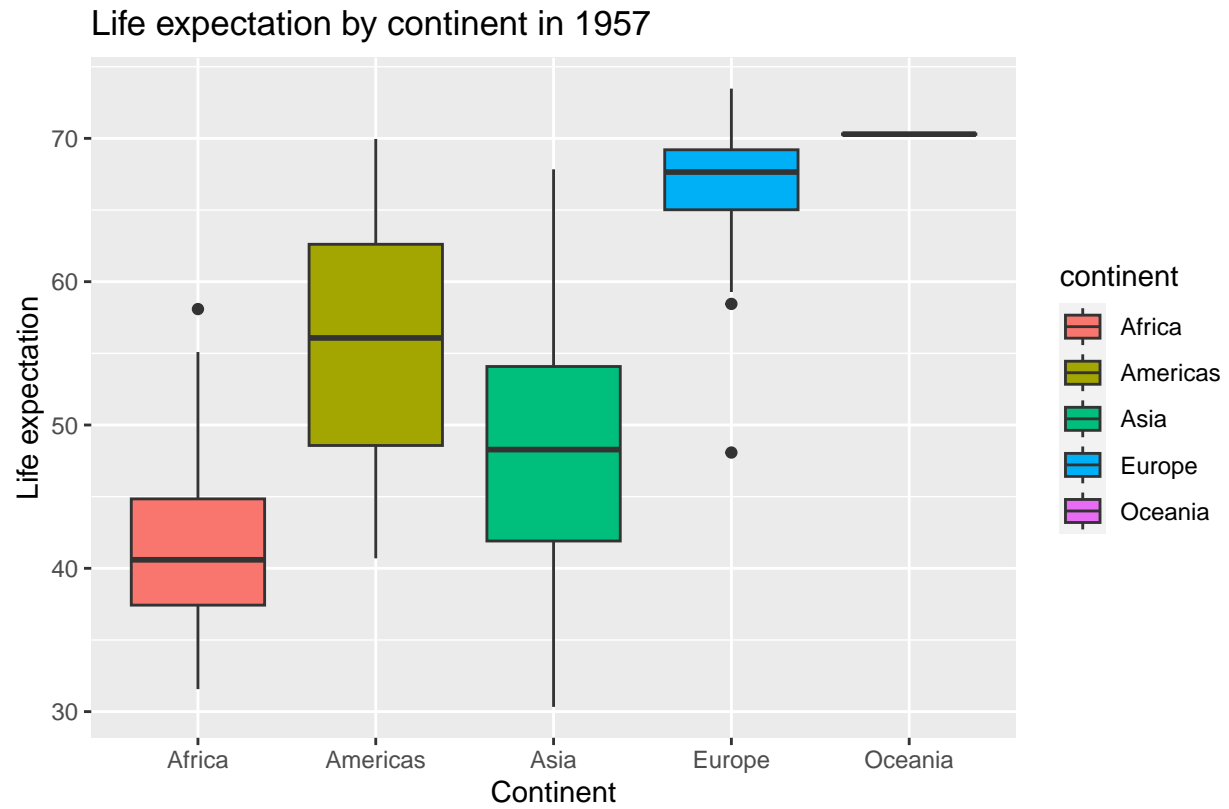
Lastly, let's use the same bar plots the average difference in life expectancy from 1957 to 2007 across continents.



Box plots: Distribution of life expectancies across continent

To create scatter plots in ggplot2 we use the `geom_boxplot()` function.

What are box plots good for?

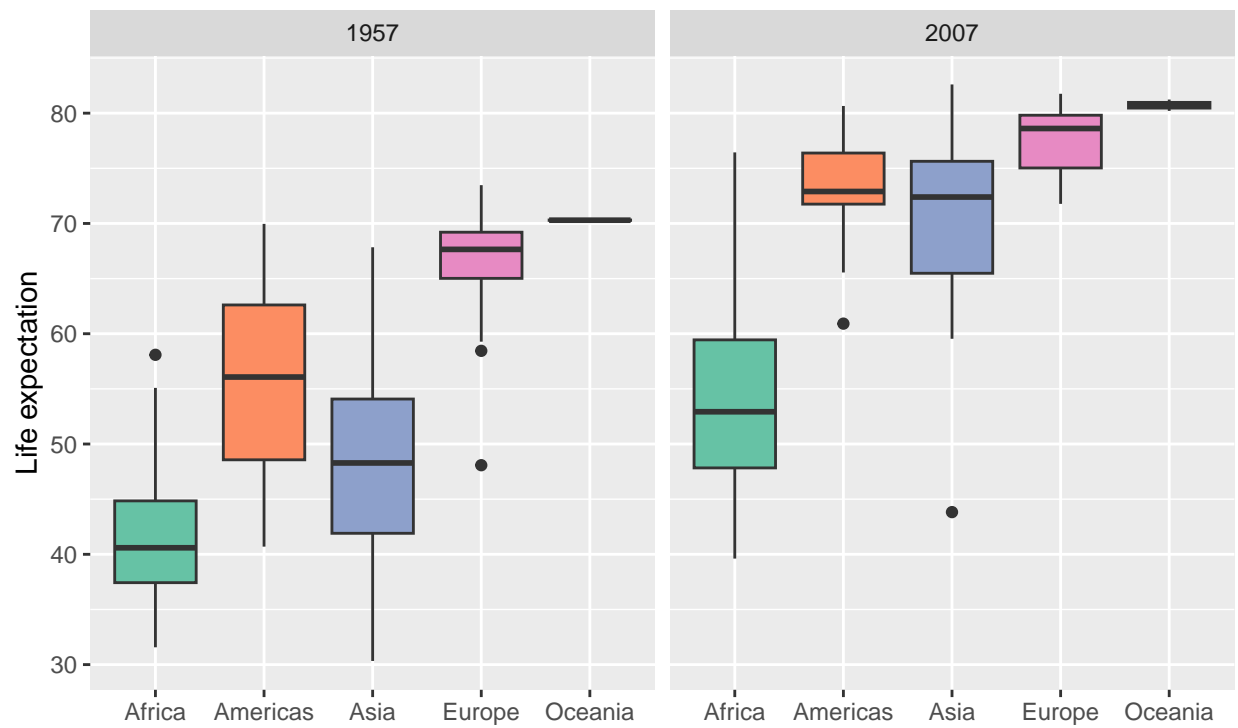


Source: Gapminder

There is a lot of redundant information in this plot, no?



Life expectation by continent in 1957 and 2007



Source: Gapminder

Could you make the same box plot for GDP?

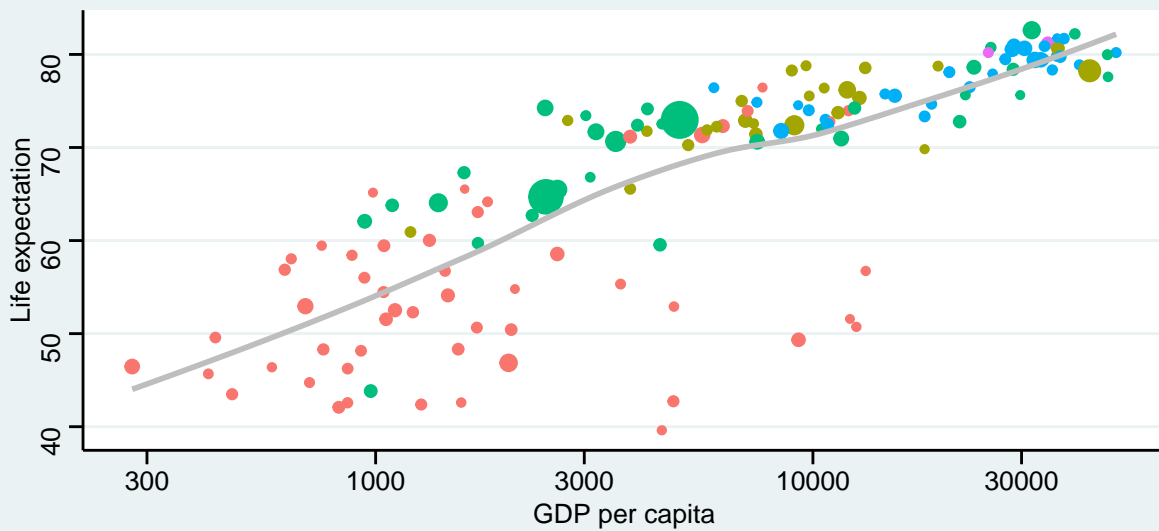
Scatter plots: Population, life expectancy and GDP

To create scatter plots in ggplot2 we use the `geom_point()` function.

What are scatter plots good for?

```
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

How much life can money buy?



pop

● 250000000 ● 500000000 ● 750000000 ● 1000000000 ● 1250000000

continent

● Africa ● Americas ● Asia ● Europe ● Oceania

Source: Gapminder