Data Visualization for Political Scientists

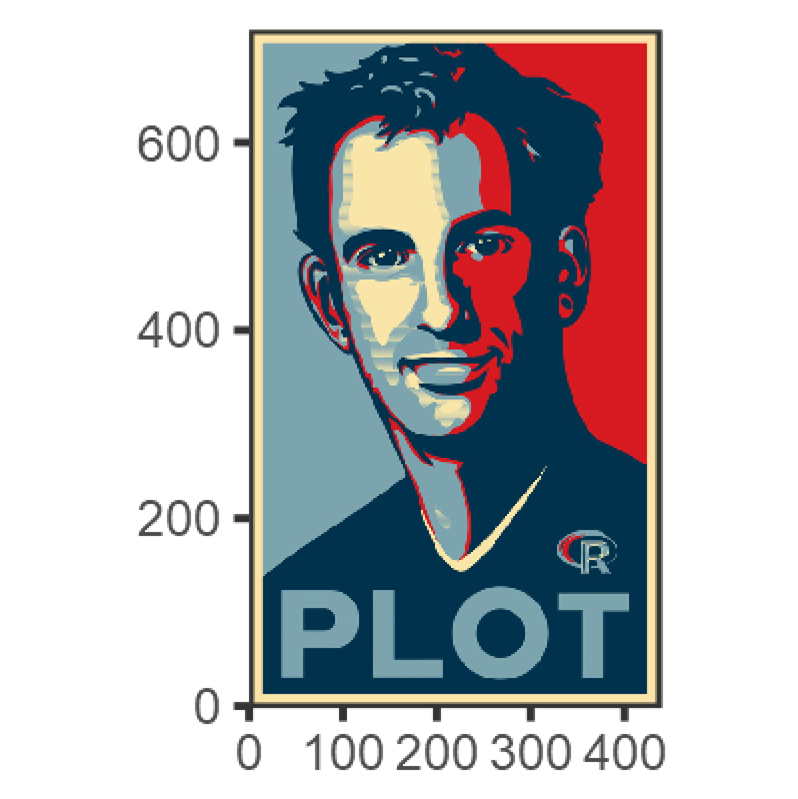
Session 2 - The Anatomy of a ggplot2 plot

Felix Haass

15 Januar 2018

ggplot2

The "gg" in stands for the "Grammar of Graphics." The grammar of graphics is a philosophy of data visualization which forces you to think about *what* you want to visualize *how*. Hadley Wickham followed this philosophy to implement the package.



The anatomy of a ggplot2 plot

The grammar of graphics specifies **building blocks** out of which an analyst builds a plot. These include, in the order of application:

1. Data (*what* do you want to plot?)
2. Aesthetic mapping (what comes on the x and y axes?)
3. Geometric object ( ) (How do we want to see our data? Points, lines, bars, …)
4. Add more (e.g. add regression lines to a scatterplot)
5. Polish labels, scales, legends, and appearance.

(see this link for more details)

Useful tips from the dataviz ninja

1. **Think hard about you want to visualize!**

> "Think of graphs as comparison" - Andrew Gelman

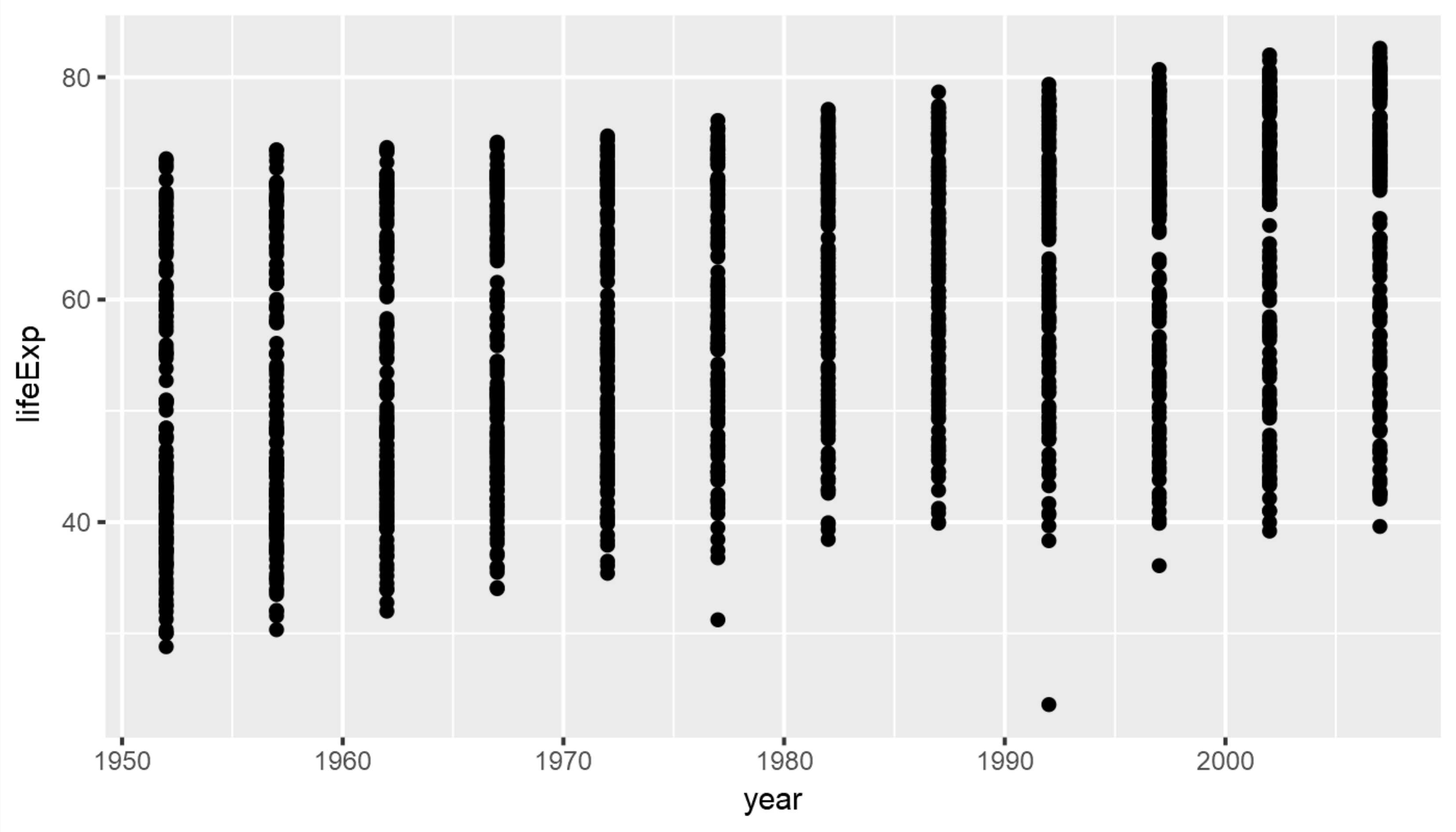
ggplot2 building blocks

Let's look at the ggplot2 building blocks in practice:

Note the that ties the building blocks together.

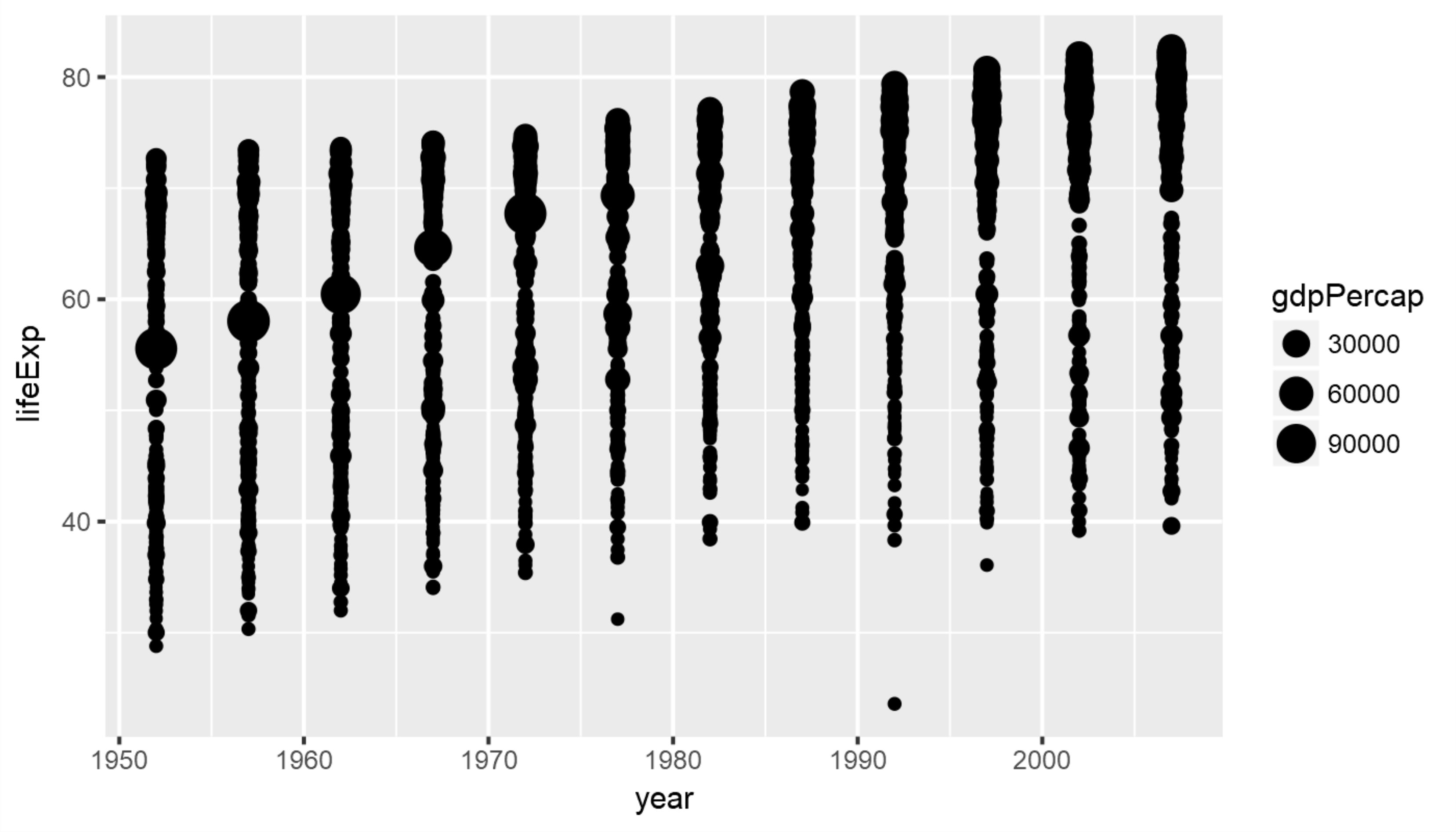
ggplot2 building blocks

[redacted]



Aesthetics - Size

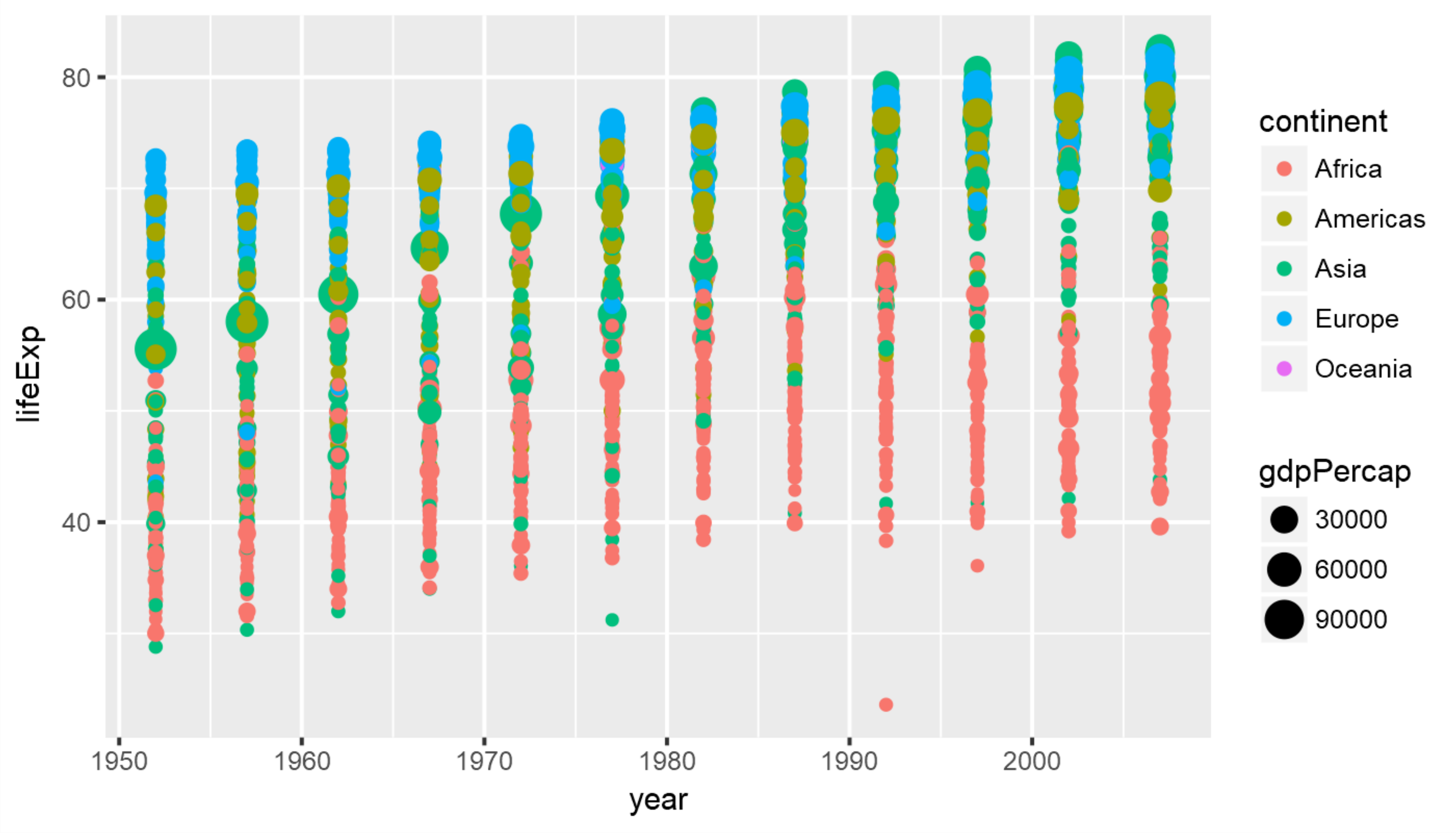
Aesthetics - Size



Aesthetics II - Color

Aesthetics II - Color

[redacted]

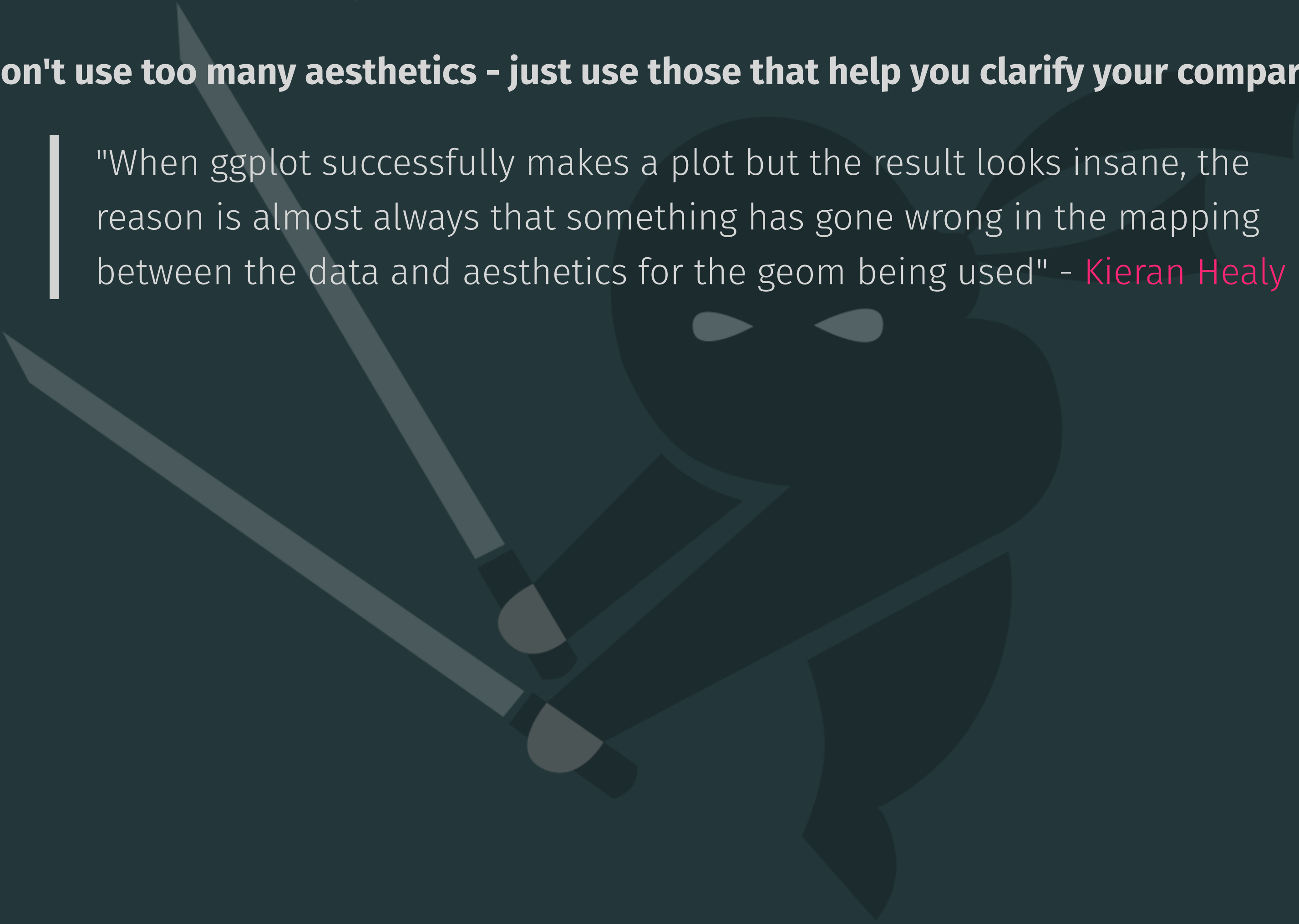


Useful tips from the dataviz ninja

1. Think hard about *what* you want to visualize!

2. **Don't use too many aesthetics - just use those that help you clarify your comparison!**

> "When ggplot successfully makes a plot but the result looks insane, the reason is almost always that something has gone wrong in the mapping between the data and aesthetics for the geom being used" - Kieran Healy



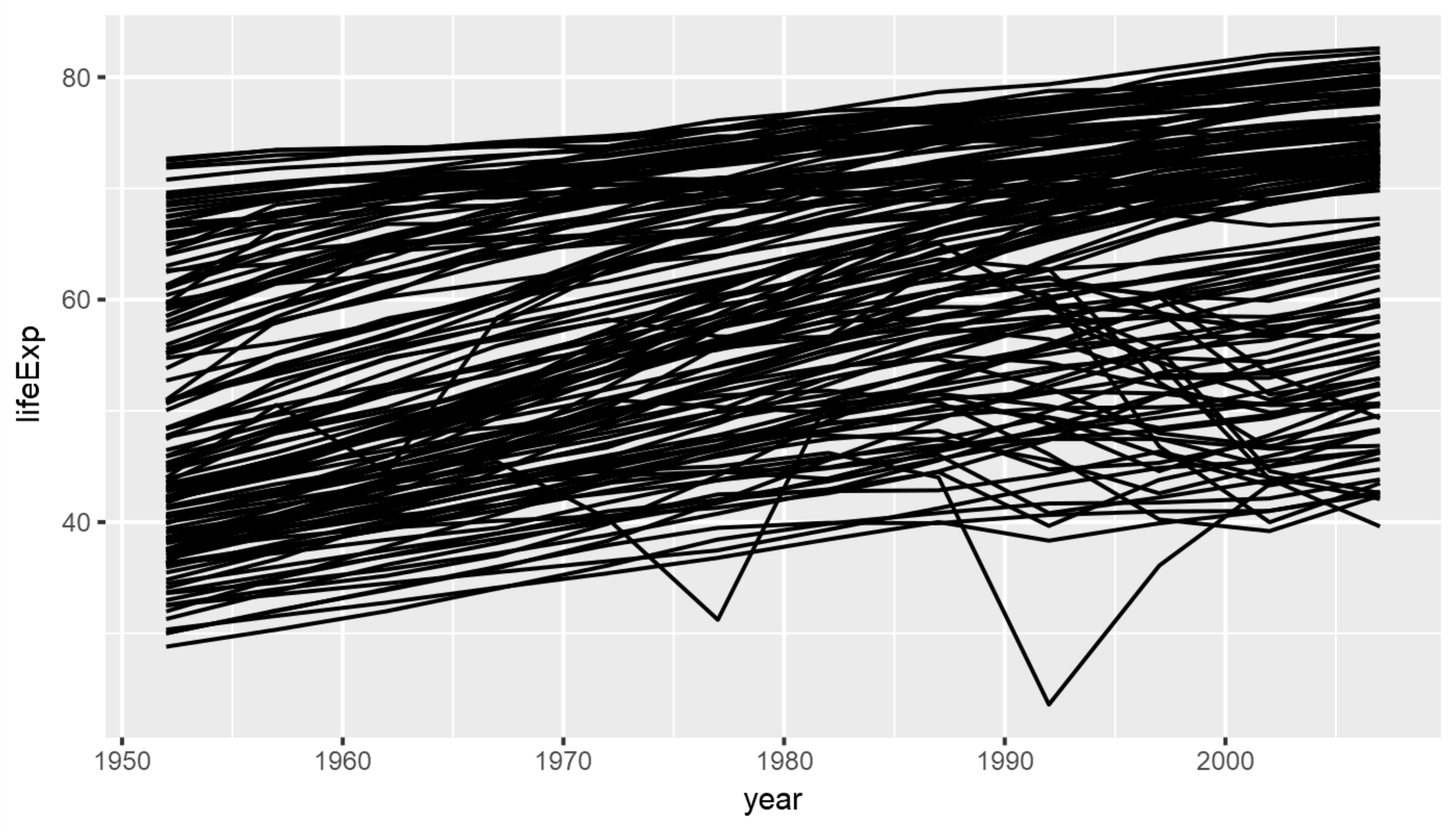
geoms

geoms

Whoops! What happened here?

geoms

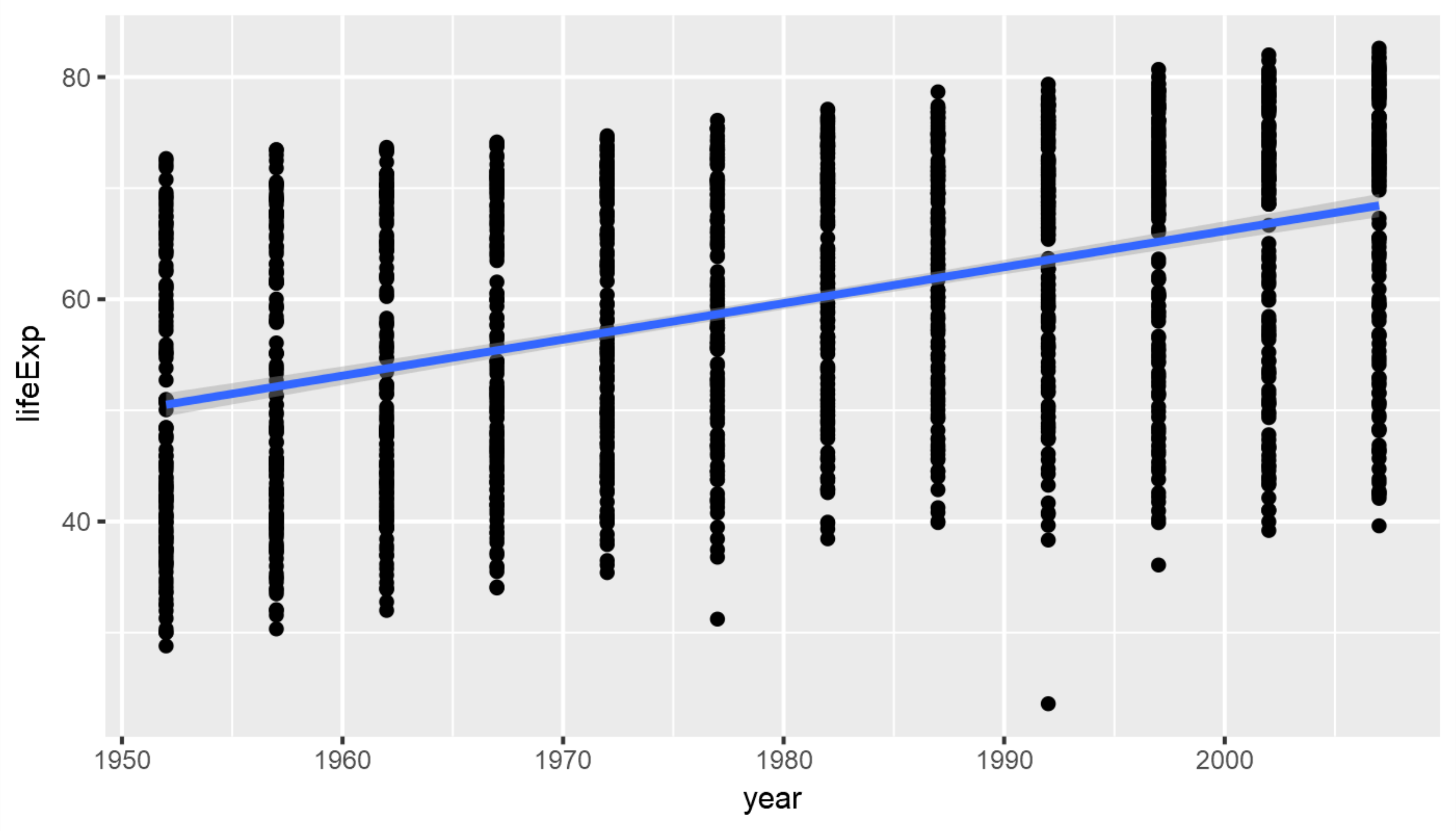
geoms



Combining geoms

Combining geoms

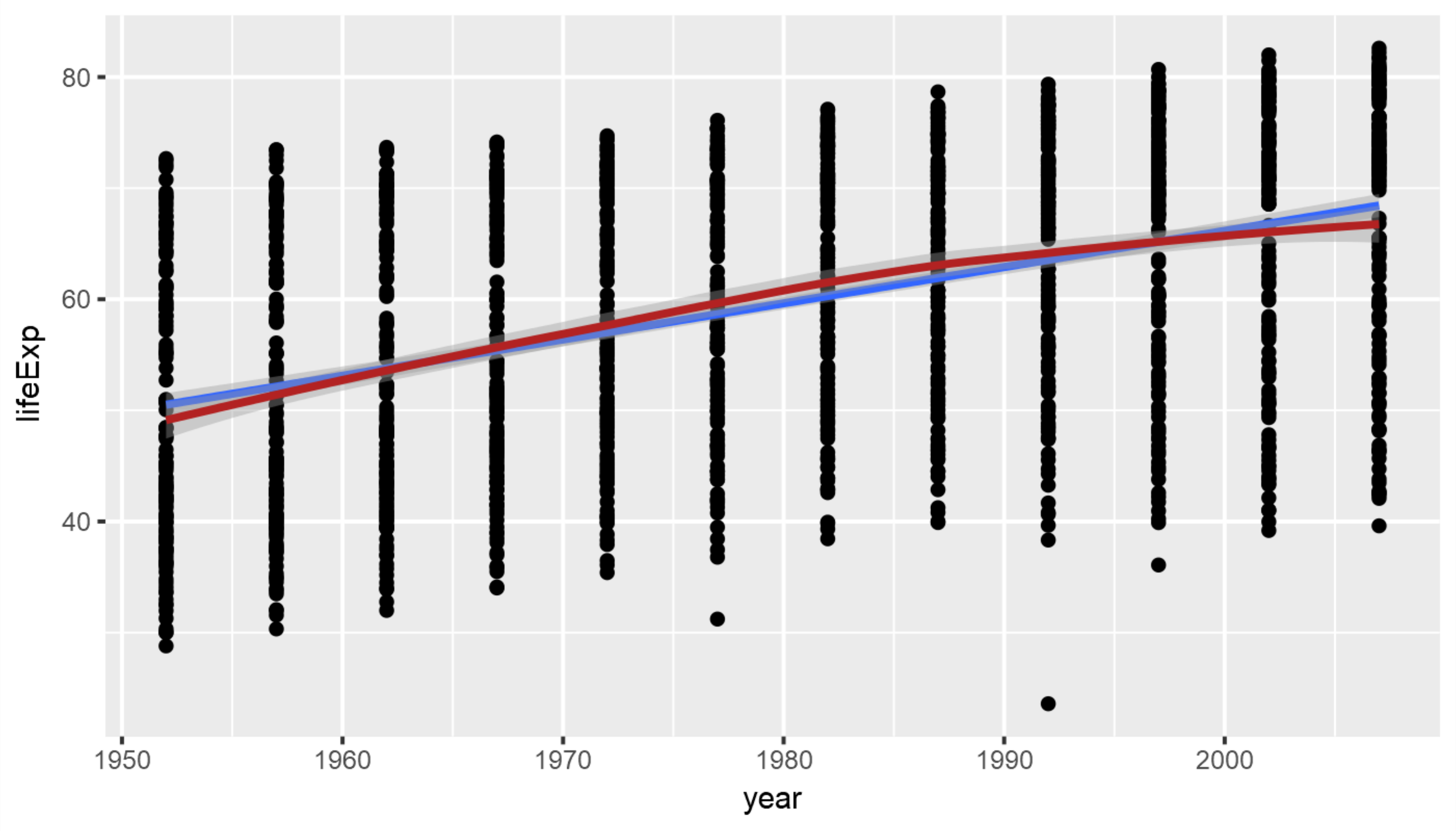
[redacted]



Combining geoms II

Bonus question: in this example we fix the color, i.e. we map it to a fixed value (which is red). What happens if we would map to a variable in the gapminder dataset, such as ?

Combining geoms II



Manipulate and Preprocess Data

Subsetting/filtering data helps to reduce complexity & get at the comparison that we want. To do that, we use the package which is part of the .

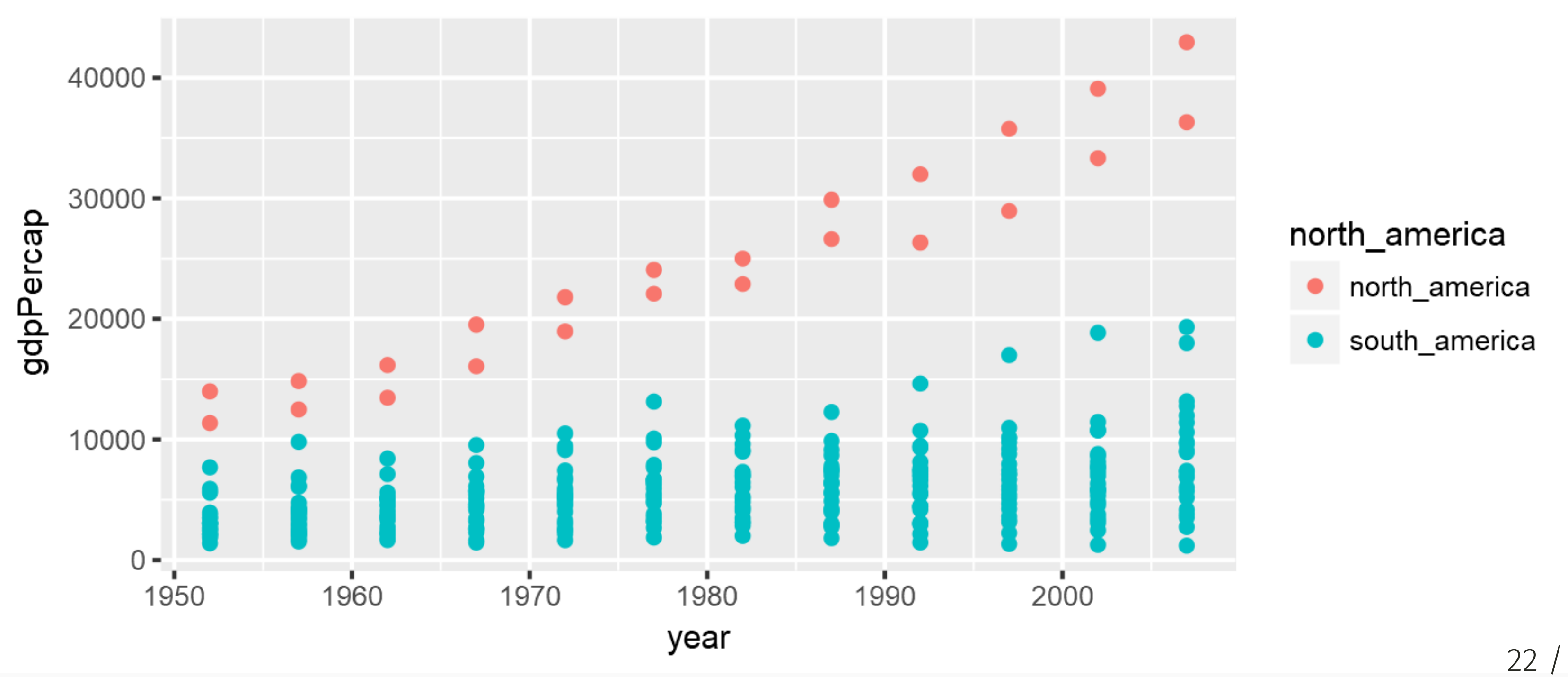
To filter data, we use the function.

Manipulate and Preprocess Data

Modify/add variables to existing data frame. We modify data with the function and chain them together using the pipe operator .

Manipulate and Preprocess Data

Use filtered and preprocessed data to highlight comparisons in ggplot:

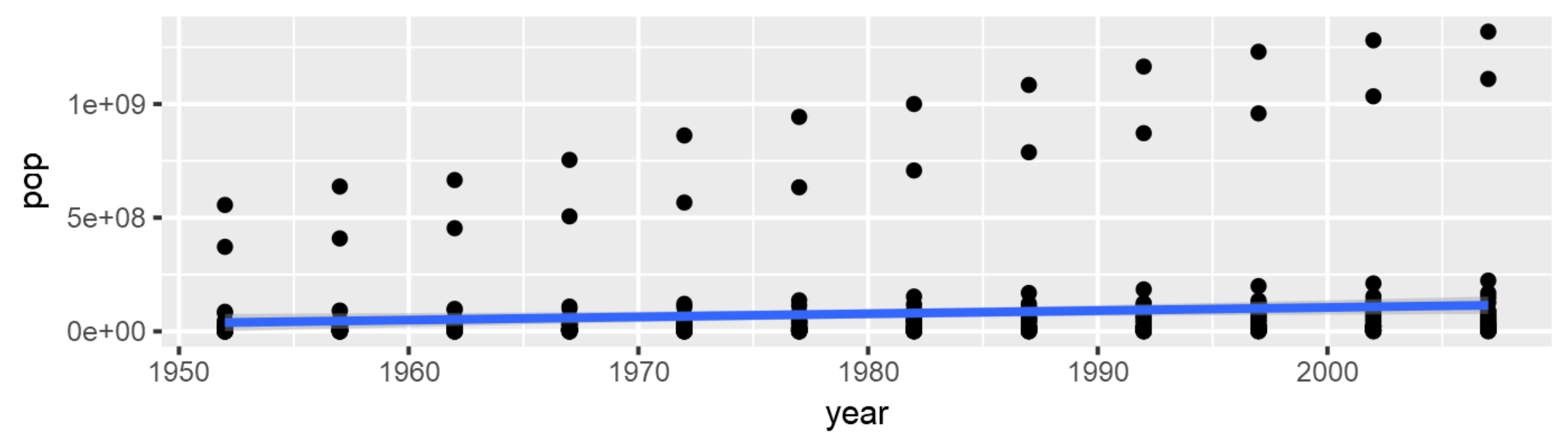


Exercise

Plot the development of population size (variable in the gapminder data) over time (variable in the gapminder data) in Asia (hint:). Add a trend line and/or smooth line.

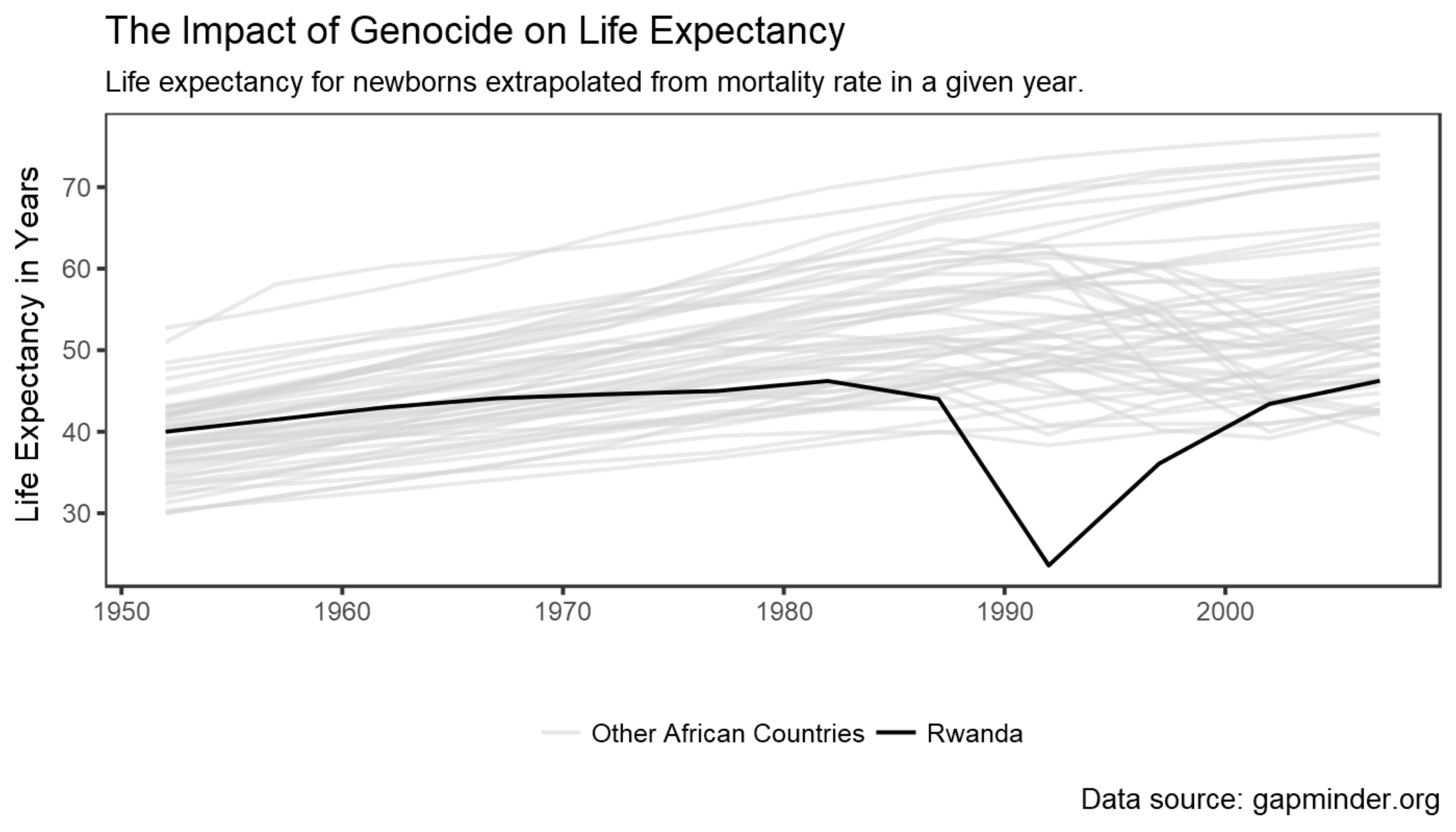
Bonus exercise: Plot the relationship between population size and ! (hint: might make sense to wrap and in).

Solution



Walkthrough Exercise

Goal:

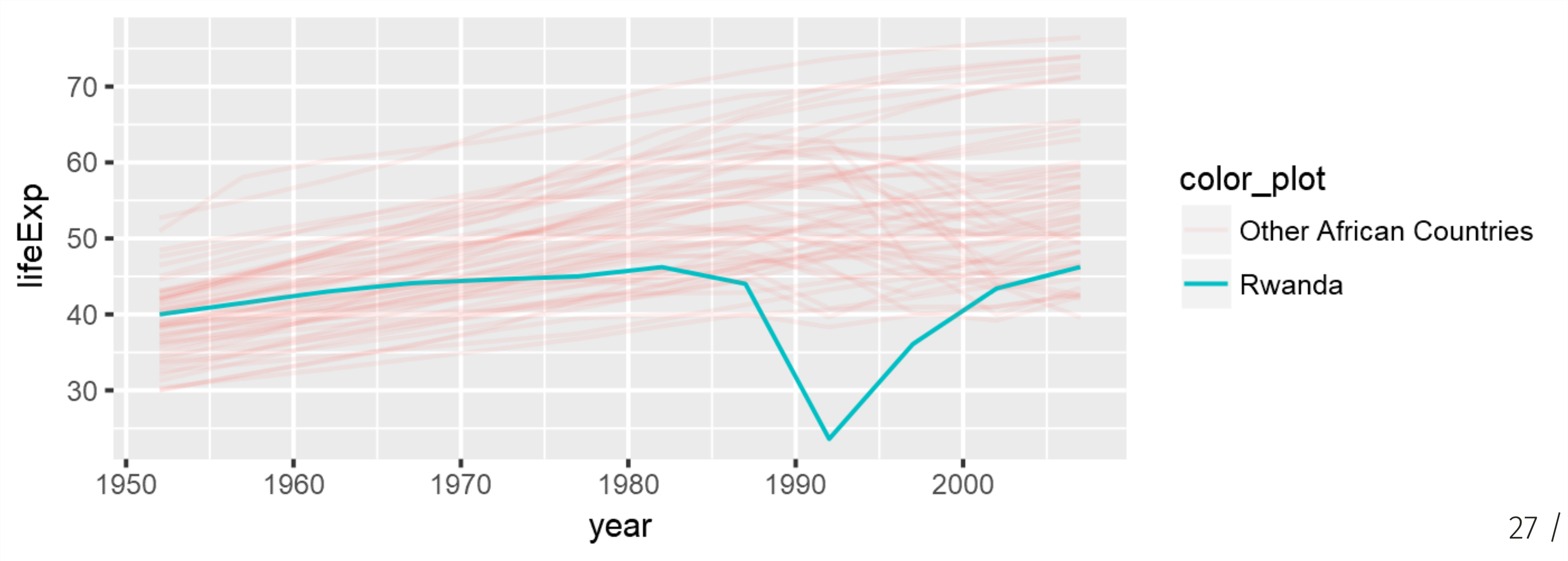


What do we want to visualize?

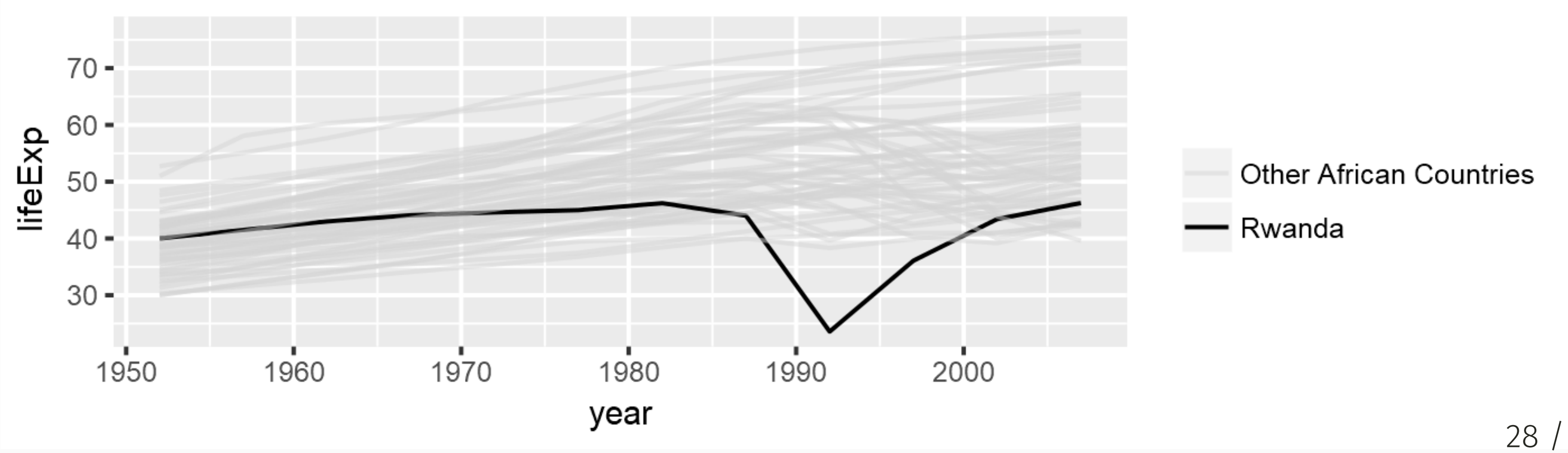
Think about the data! What is the comparison?

Genocide vs. non-genocide countries => Rwanda vs. rest of Africa

Add geom_line() + map color/alpha

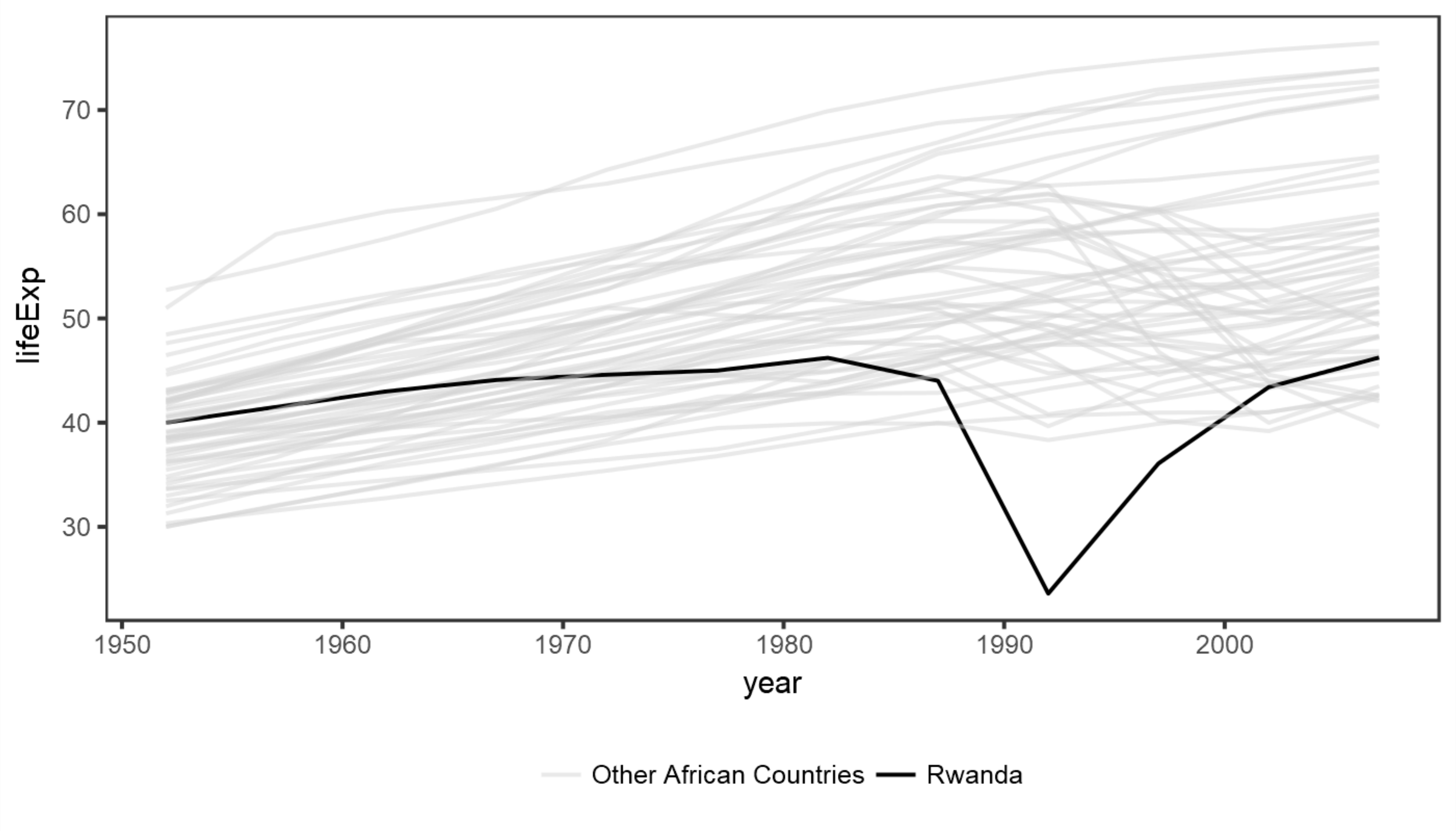


Add color/alpha scales



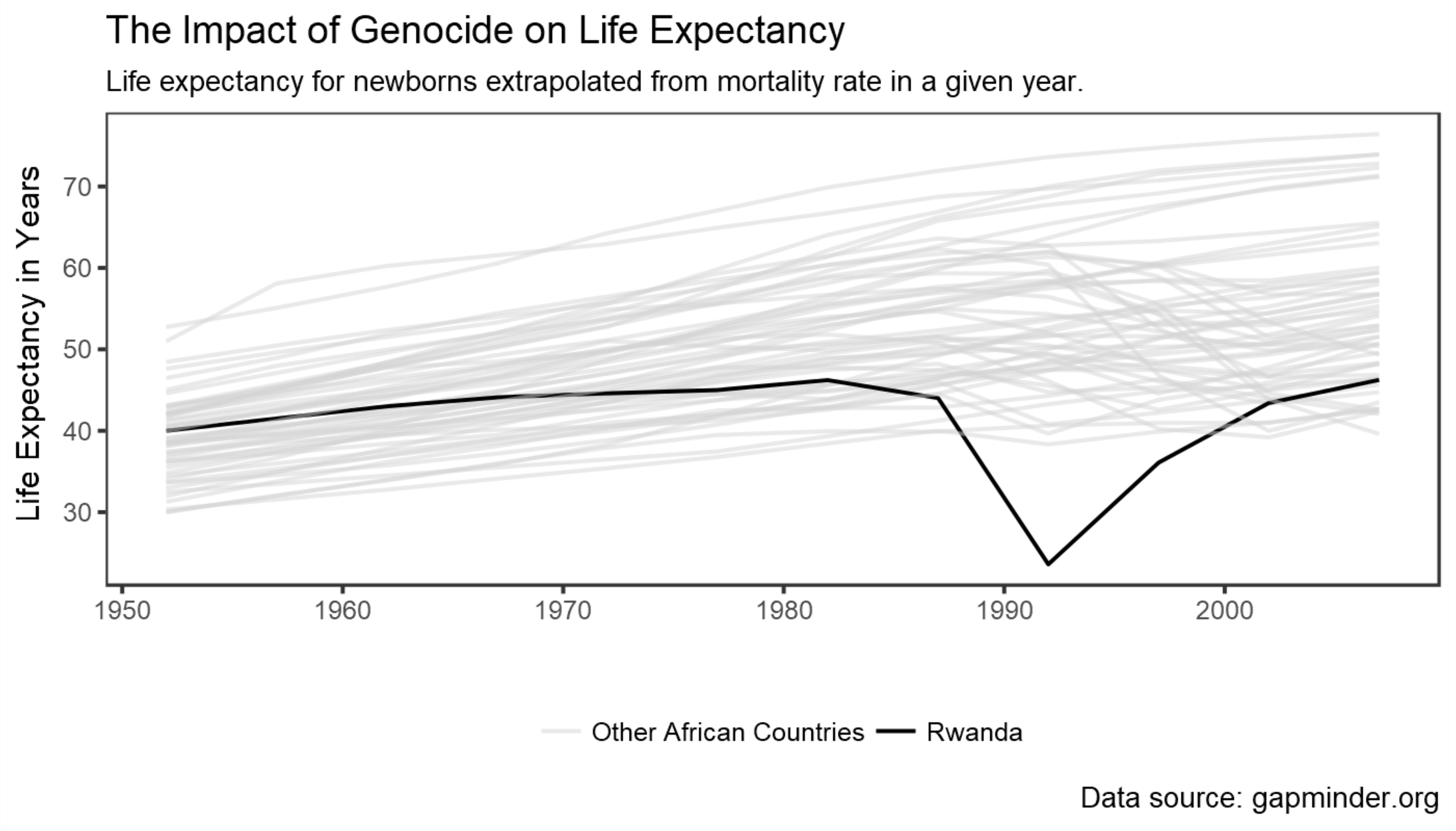
Manipulate appearance: add theme

Manipulate appearance: add theme



Manipulate appearance: change labels

Manipulate appearance: change labels



The Impact of Genocide on Life Expectancy

Life expectancy for newborns extrapolated from mortality rate in a given year.

Data source: gapminder.org

Useful tips from the dataviz ninja

1. Think hard about *what* you want to visualize!

2. Don't use too many aesthetics - just use those that help you clarify your comparison!

3. **Trial and error is your friend!**

   > "If you are unsure of what each piece of code does, take advantage of ggplot's additive character. Working backwards from the bottom up, remove each + some_function(...) statement one at a time to see how the plot changes." - Kieran Healy