Visualisation

Theoretical and Empirical Research Methodology,
Implementation Lab 7

Prof. Dr. Claudius Gräbner-Radkowitsch
Europa-University Flensburg, Department of Pluralist Economics
www.claudius-graebner.com | @ClaudiusGraebner | claudius@claudius-graebner.com

Europa-Universität Flensburg

Europa-Universität Flensburg
International Institute of Management and Economic Education
Department of Pluralist Economics

Goals for today

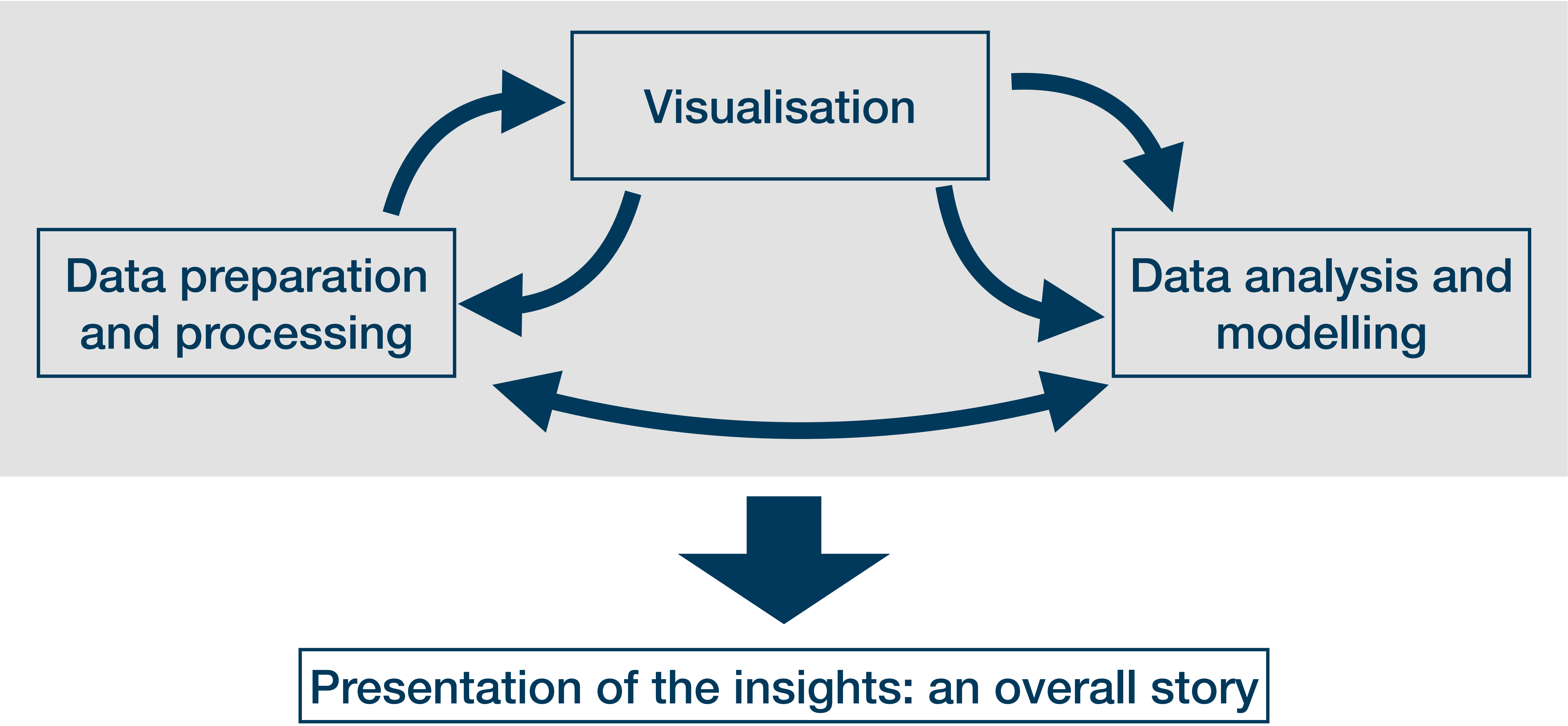
I. Understand how plots are created layer-wise via the `ggplot2` package

II. Learn how to map variables in data frames to visual aspects of a plot

III. Figure out how you can re-use code across different visualisation tasks

Basics of visualization

The role of visualisation in data science



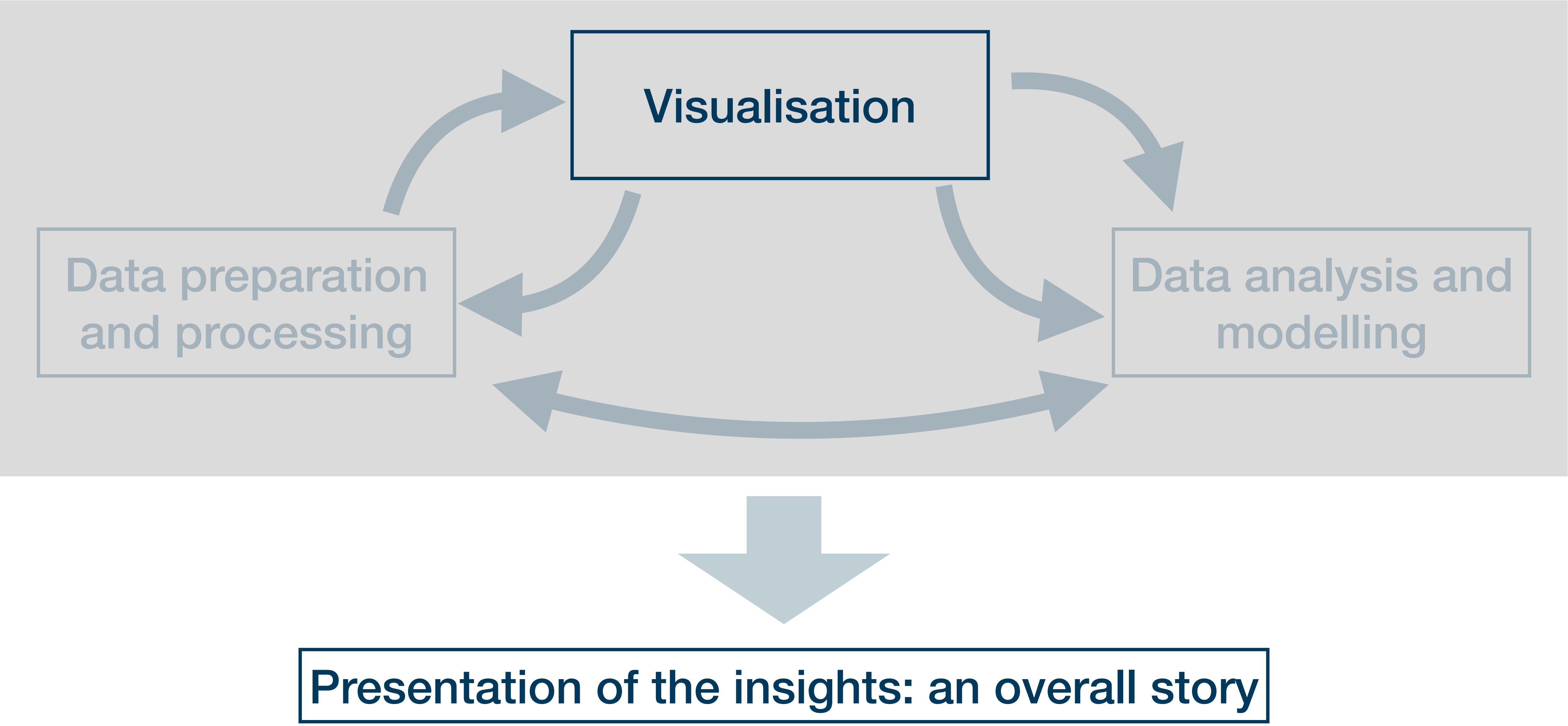
Visualisation

Data preparation and processing

Data analysis and modelling

Presentation of the insights: an overall story

The role of visualisation in data science



Visualisation

Data preparation and processing

Data analysis and modelling

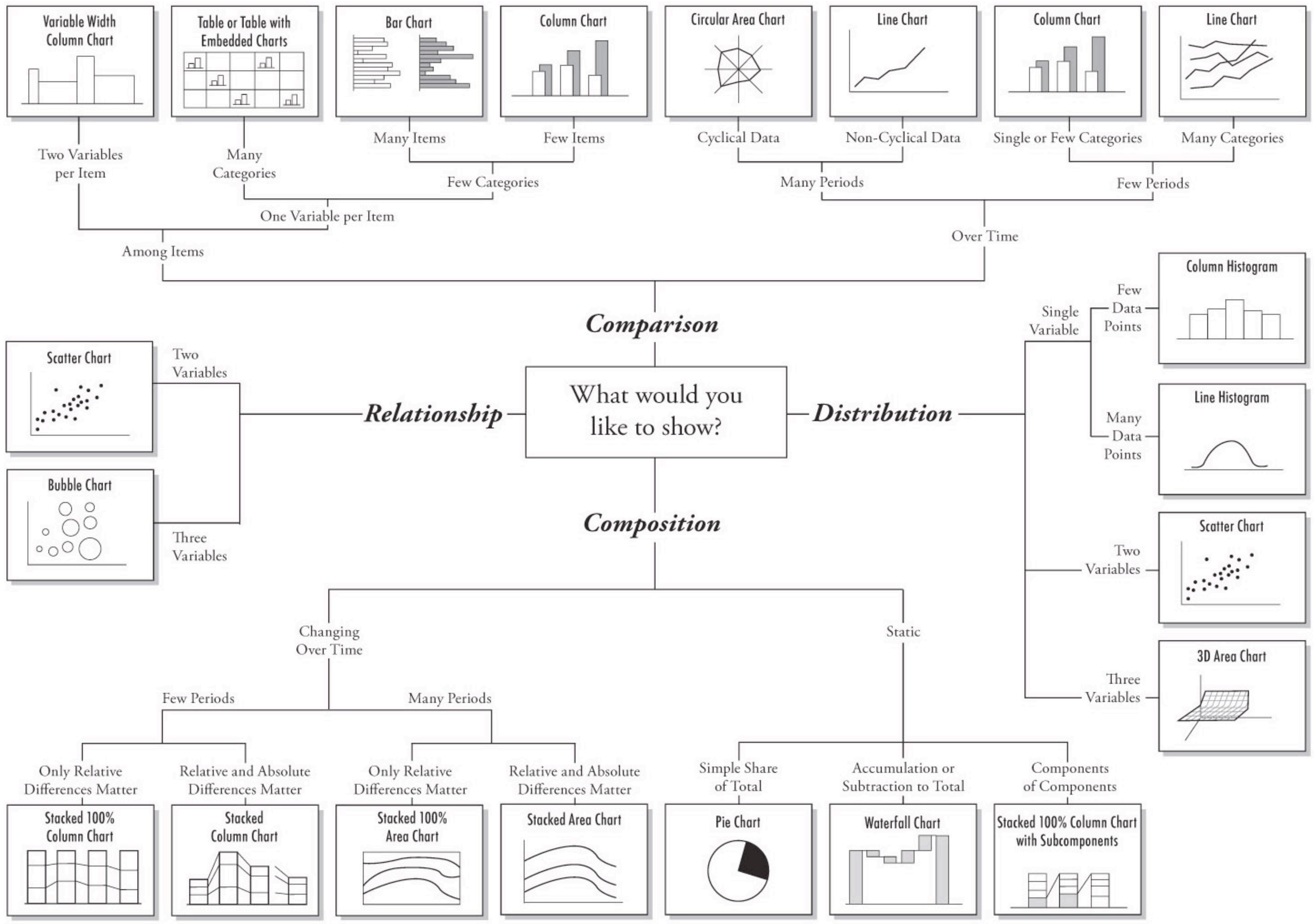
Presentation of the insights: an overall story

About visualisations

- Visualisations can be used for many purposes
  - **Exploratory data analysis** → understand your data → prepare/refine models
  - **Communication** → inform others about your results
  - **Manipulation** → convince others or recognise others trying to convince you
- Here we will learn about how to create visualisations using the package `ggplot2`
- An easy-to-read, widely-used and powerful visualisation engine
- Many great extensions, e.g. for animated GIFs, control charts, and many more…

What kind of plot do you want?

Chart Suggestions—A Thought-Starter

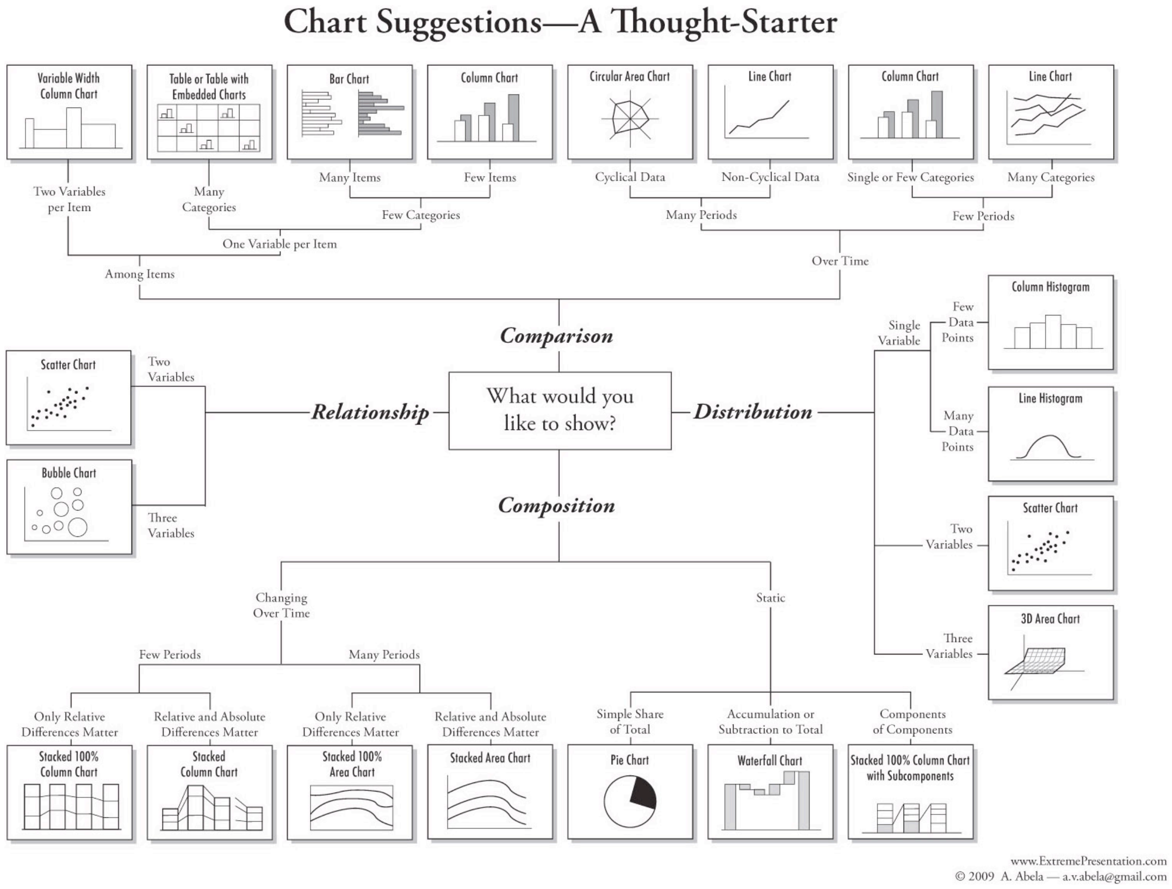


www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

What kind of plot do you want?

- Visualisation always involves prior thinking and theory
- The great thing about `ggplot2` is that the syntax is the same for all graphs

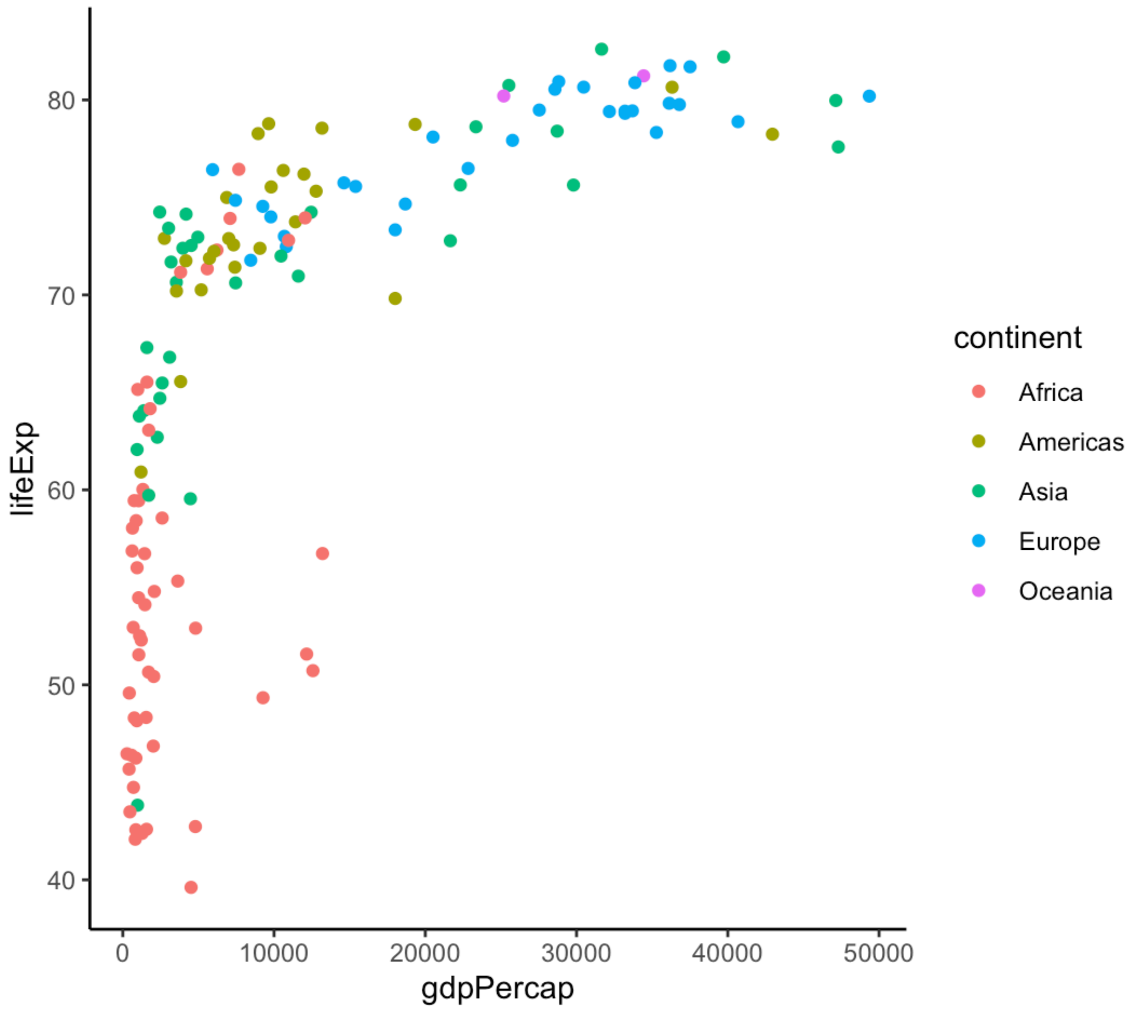
Chart Suggestions—A Thought-Starter



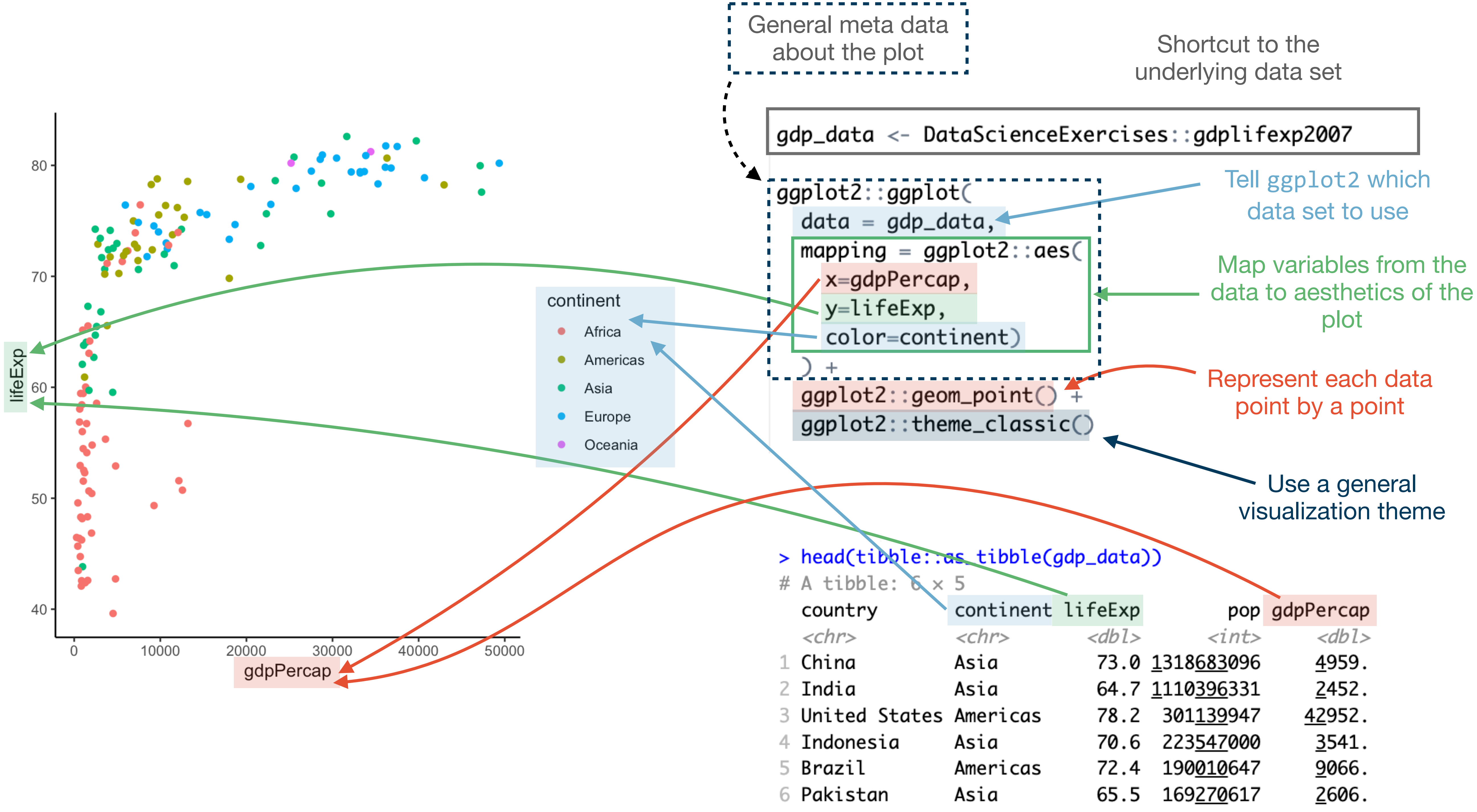
www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com

- During our lecture we focus on two examples:
  - The scatterplot/bubble chart from session 1
  - A line chart

A very simple example in the beginning



A very simple example in the beginning



General meta data about the plot

Shortcut to the underlying data set

```
gdp_data <- DataScienceExercises::gdplifexp2007
```

```
ggplot2::ggplot(
  data = gdp_data,
  mapping = ggplot2::aes(
    x=gdpPercap,
    y=lifeExp,
    color=continent)
  ) +
  ggplot2::geom_point() +
  ggplot2::theme_classic()
```

Tell ggplot2 which data set to use

Map variables from the data to aesthetics of the plot

Represent each data point by a point

Use a general visualization theme

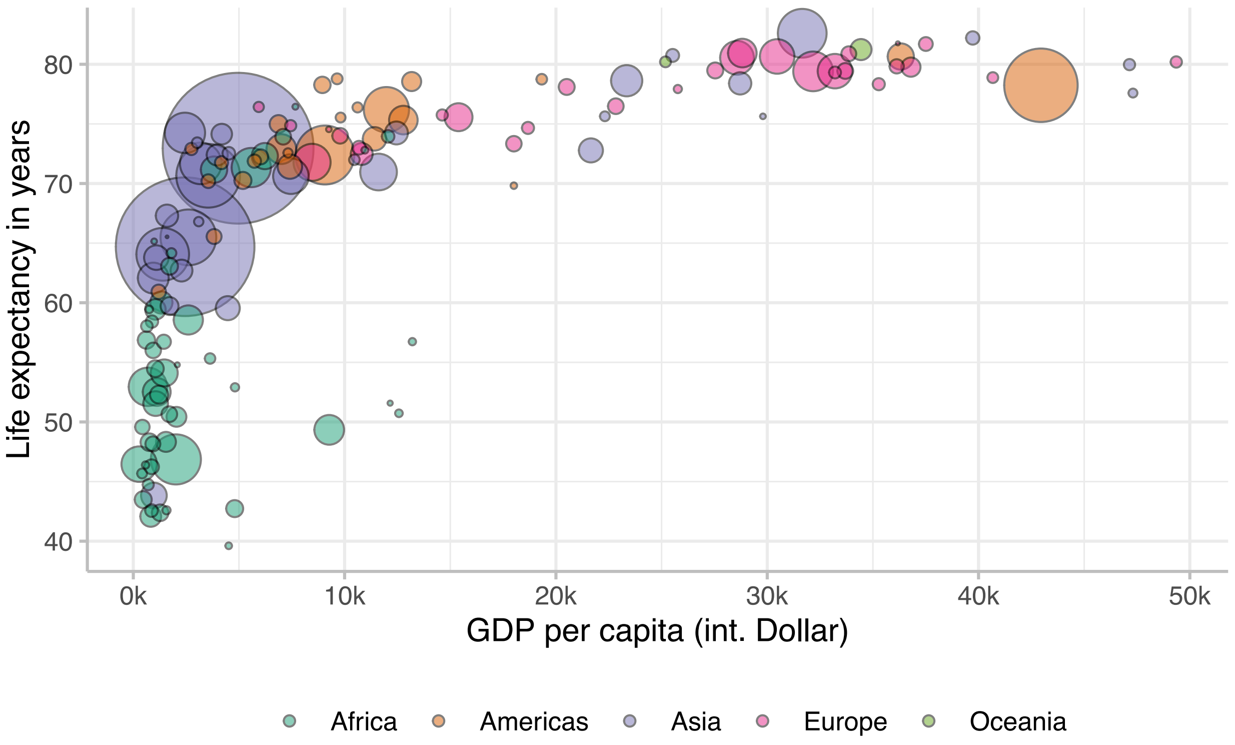
```
> head(tibble::as_tibble(gdp_data))
# A tibble: 6 x 5
  country       continent lifeExp        pop gdpPercap
  <chr>         <chr>       <dbl>      <int>     <dbl>
1 China         Asia         73.0 1318683096     4959.
2 India         Asia         64.7 1110396331     2452.
3 United States Americas     78.2  301139947    42952.
4 Indonesia     Asia         70.6  223547000     3541.
5 Brazil        Americas     72.4  190010647     9066.
6 Pakistan      Asia         65.5  169270617     2606.
```

The practical workflow

Where we want to go:

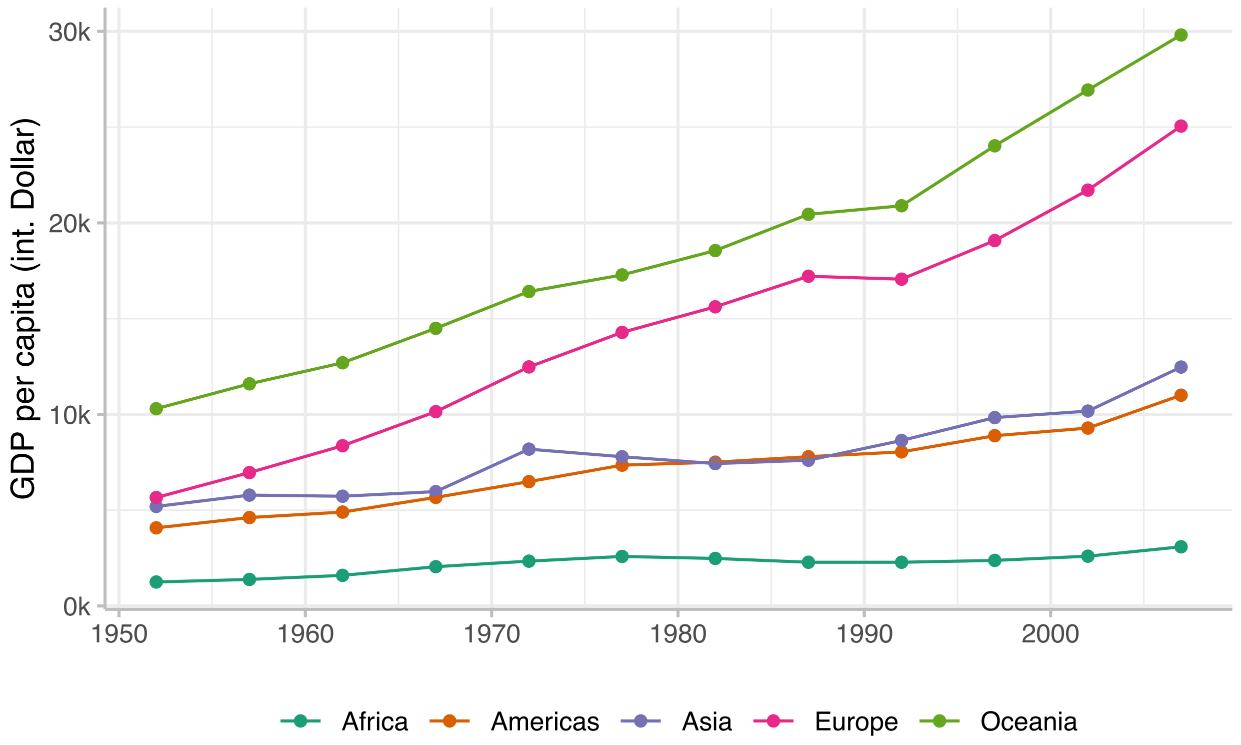
- Today we want to get started with visualisations and produce the following two plots:

Life expectancy and income per capita



Note: size of bubbles represents population. Data: Gapminder

The divergence of income per capita



Note: country data averaged over continants. Data: Gapminder

- We will see that the mechanics are very similar for different plots
    - Based on the readings you will be able to make even much more plots already now!

The general idea

- Every plot in `ggplot2` is generated in two major steps
  - You describe the plot in all its details via a list
  - You call the list and R renders the plot for you

This is where all the work gets done 😅

This is where errors become apparent 🤬

- To create the list-like description, `ggplot2` offers you a ton of helper functions

- You always start with an empty plot, then add layers above this empty plot, adjust details and that's it!

- Lets illustrate this using a subset of the gapminder data set only containing data for the year 2017
  - Readymade available to you via the DataScienceExercises package as `DataScienceExercises::gdplifexp2007`

Developing a ggplot - the general workflow

- Since we are working on the graph development interactively, see my **lecture notes** for documentation purposes

Summary & outlook

Summary

- Visualisations serve many purposes, including the exploration of your data and the communication of your results
- We learned how to visualise data stored in data frames via `ggplot2`
- While there are many different plot variants, their syntax is very similar

```
ggplot() +
     <GEOM_FUNCTION>(
       data = <DATA>
       mapping =aes(<MAPPINGS>),
          stat = <STAT>,
          position = <POSITION>
          ) +
     <COORDINATE_FUNCTION> +
     <FACET_FUNCTION> +
     <THEME_ADJUSTMENTS>
```

The geometric forms used to represent the data (points, lines, shades,…)

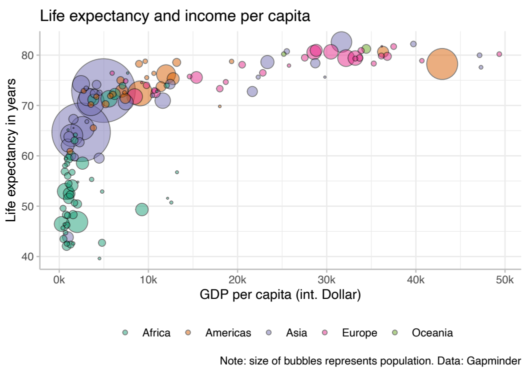
The data to be visualized

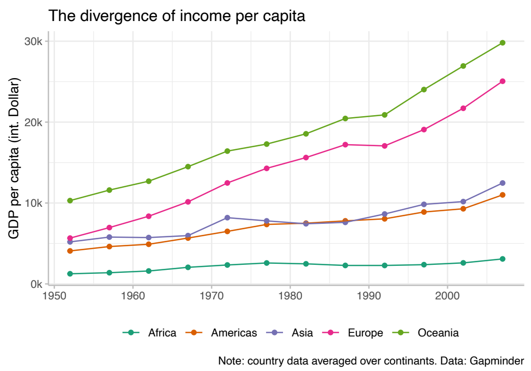
The mapping of the variables in data to the plot aesthetics (x/y-axis, size, form,…)

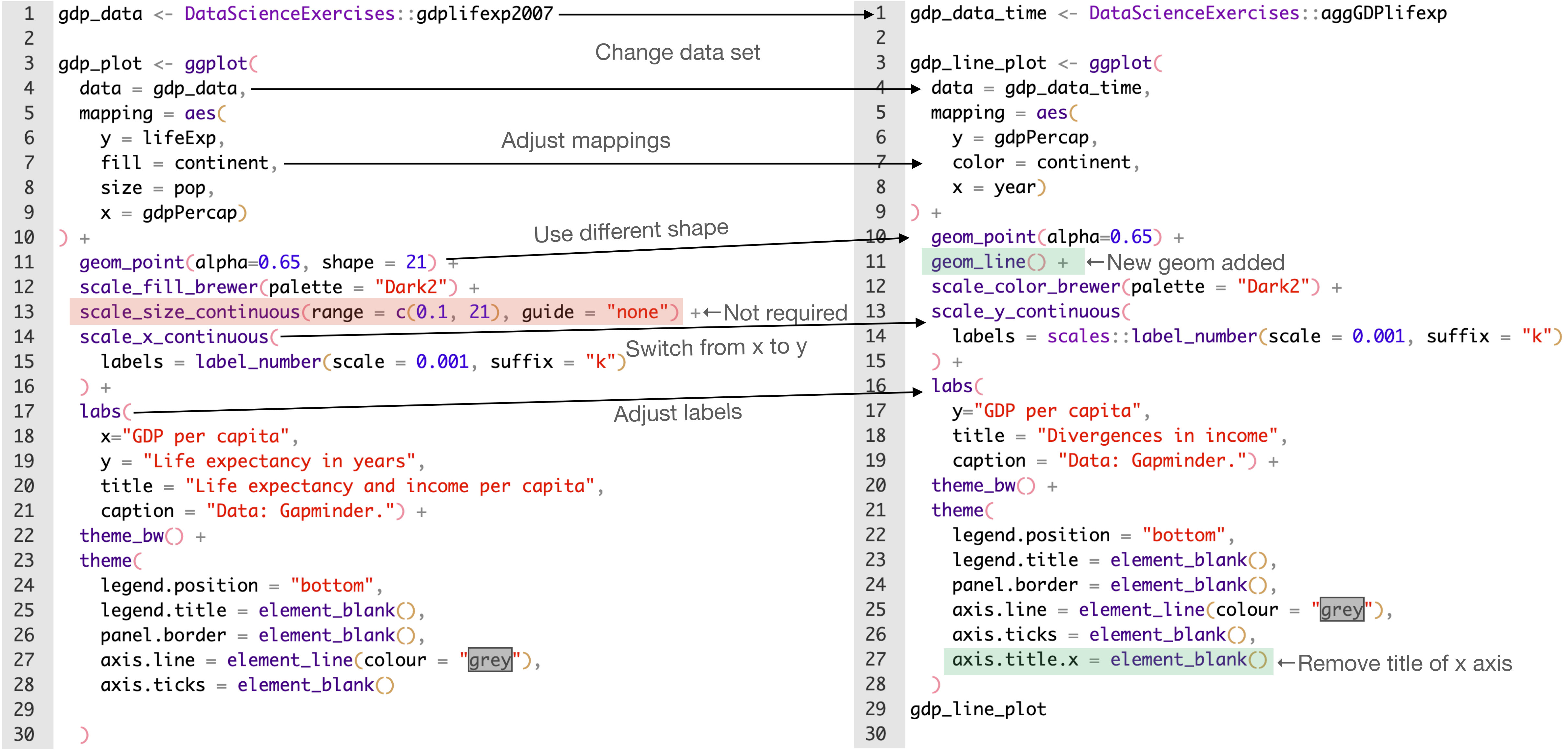
May be set as defaults within `ggplot()`, or separately for each geom

Adjustment to look, labels, etc.

From the bubble to the line chart







```
gdp_data <- DataScienceExercises::gdplifexp2007

gdp_plot <- ggplot(
  data = gdp_data,
  mapping = aes(
    y = lifeExp,
    fill = continent,
    size = pop,
    x = gdpPercap)
) +
  geom_point(alpha=0.65, shape = 21) +
  scale_fill_brewer(palette = "Dark2") +
  scale_size_continuous(range = c(0.1, 21), guide = "none") +
  scale_x_continuous(
    labels = label_number(scale = 0.001, suffix = "k")
  ) +
  labs(
    x="GDP per capita",
    y = "Life expectancy in years",
    title = "Life expectancy and income per capita",
    caption = "Data: Gapminder.") +
  theme_bw() +
  theme(
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.border = element_blank(),
    axis.line = element_line(colour = "grey"),
    axis.ticks = element_blank()

  )
```

Change data set

Adjust mappings

Use different shape

←Not required

Switch from x to y

Adjust labels

```
gdp_data_time <- DataScienceExercises::aggGDPlifexp

gdp_line_plot <- ggplot(
  data = gdp_data_time,
  mapping = aes(
    y = gdpPercap,
    color = continent,
    x = year)
) +
  geom_point(alpha=0.65) +
  geom_line() +
  scale_color_brewer(palette = "Dark2") +
  scale_y_continuous(
    labels = scales::label_number(scale = 0.001, suffix = "k")
  ) +
  labs(
    y="GDP per capita",
    title = "Divergences in income",
    caption = "Data: Gapminder.") +
  theme_bw() +
  theme(
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.border = element_blank(),
    axis.line = element_line(colour = "grey"),
    axis.ticks = element_blank(),
    axis.title.x = element_blank()
  )
gdp_line_plot
```

←New geom added

←Remove title of x axis

Summary

- Code for different plots differs mainly by the **aesthetic mappings** and the **geoms used** → allows you to **re-use a lot of code**
  - We produced two beautiful plots: a bubble plot and a line graph
- A great way to learn how to plot is to replicate examples from the internet, and adjust them to your own data:

from Data to Viz

The R Graph Gallery

- Once you understand the basic functioning, **AI** can be a very effective **assistant** for creating visualizations