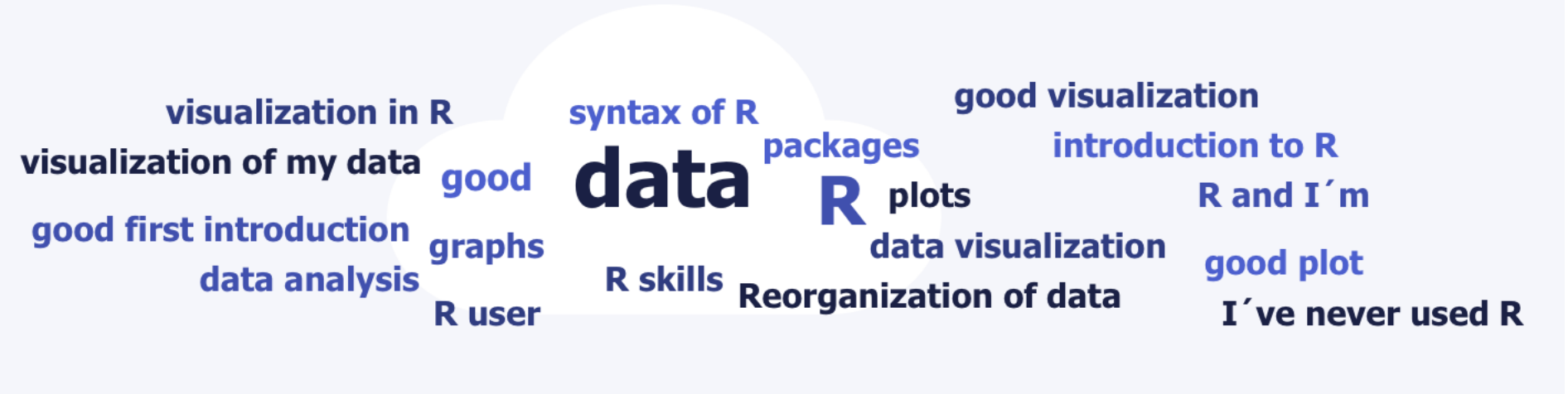
Introduction to Data Visualization: Theory and practice using R

Juliana Assis
Senior Data Scientist
jasge@dtu.dk

André Cunha
Guest PhD
afecun@dtu.dk

DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Welcome!



How to get started without previous experience

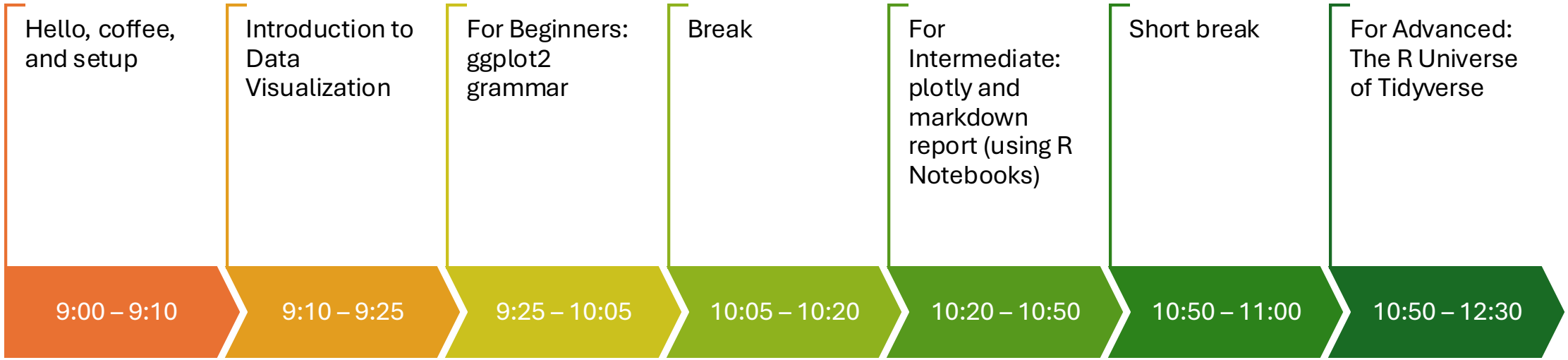
I´ve never used R and I´m interested in learning how to use it and how to represent data through it

best practices for data storytelling.

Practical ways to construct nice visual plots. When sharing codes (for publications) does codes for visuals should also be there?

I would like to learn how to make good looking graphs for publication and how to combine multiple images into one with corresponding labels and legends. So far, I manage with the help of AI, but it would be nice to have more solid foundations.

Agenda


Hello, coffee, and setup
9:00 – 9:10
Introduction to Data Visualization
9:10 – 9:25
For Beginners: ggplot2 grammar
9:25 – 10:05
Break
10:05 – 10:20
For Intermediate: plotly and markdown report (using R Notebooks)
10:20 – 10:50
Short break
10:50 – 11:00
For Advanced: The R Universe of Tidyverse
10:50 – 12:30

What is Data Visualization?

- It is the graphical representation of data or information

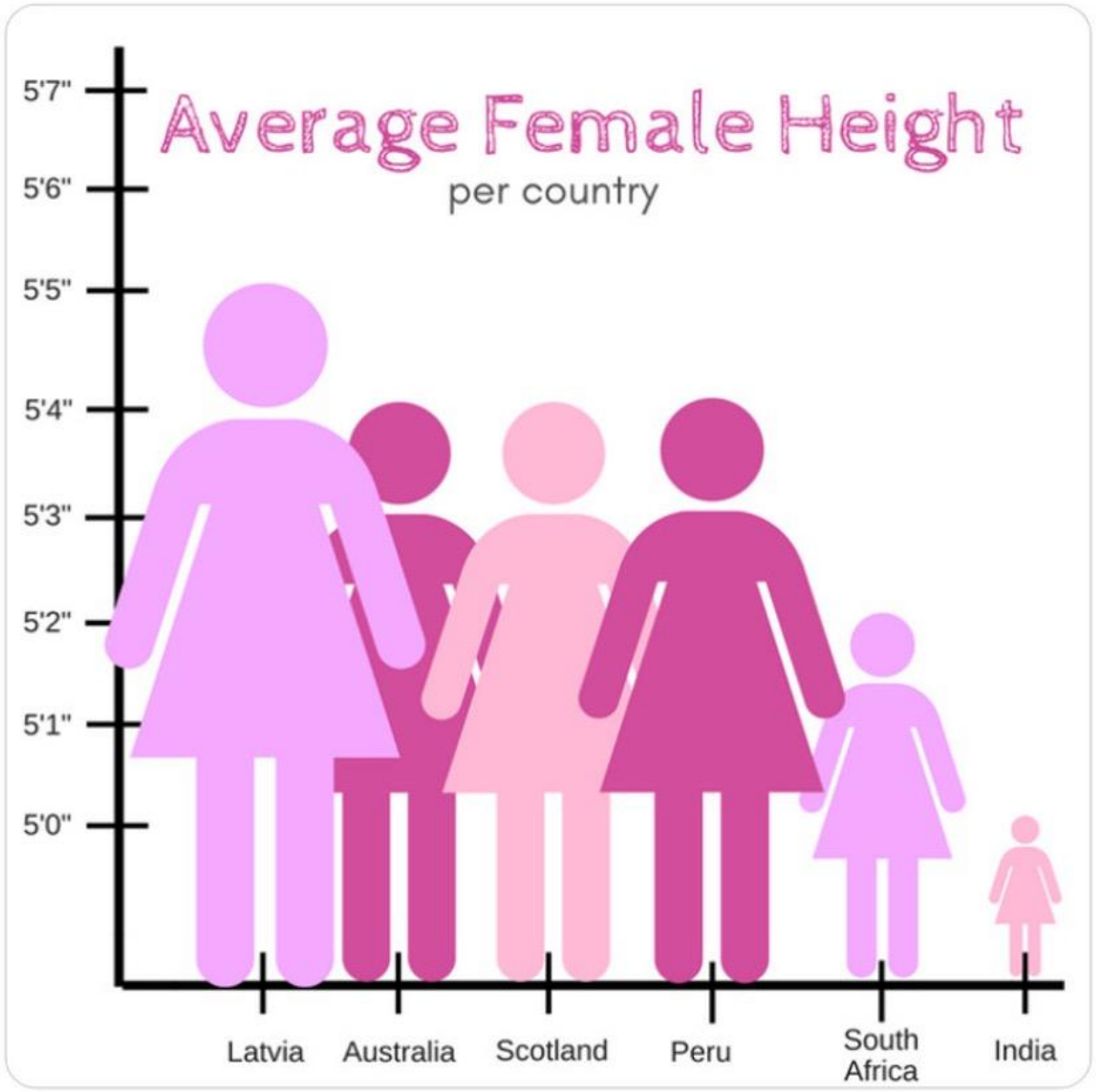
https://r-graph-gallery.com

Principles of Good Visualization

Principle	Description
Show the Data	Clearly present the data without unnecessary elements that distract from understanding.
Avoid Distorting Data	Represent data accurately; avoid misleading scales or transformations.
Enable Comparisons	Make it easy to compare data points, trends, or categories visually.
Show Variability	Highlight uncertainties or variations, such as with error bars or confidence intervals.
Integrate Evidence	Combine text, numbers, and graphics to strengthen the narrative and insights.
Use Visual Hierarchy	Organize elements so the most important information stands out clearly.
Consider the Audience	Tailor visuals to the audience's level of expertise and interests.
Aesthetic Integrity	Ensure the visualization is attractive but does not compromise accuracy or clarity.
Direct Attention	Use cues like color, labels, or annotations to focus the viewer on key aspects of the visualization.
Keep it Simple	Remove unnecessary elements and focus on the essential aspects of the data.

Ref: https://ggplot2-book.org

Principles of Good Visualization

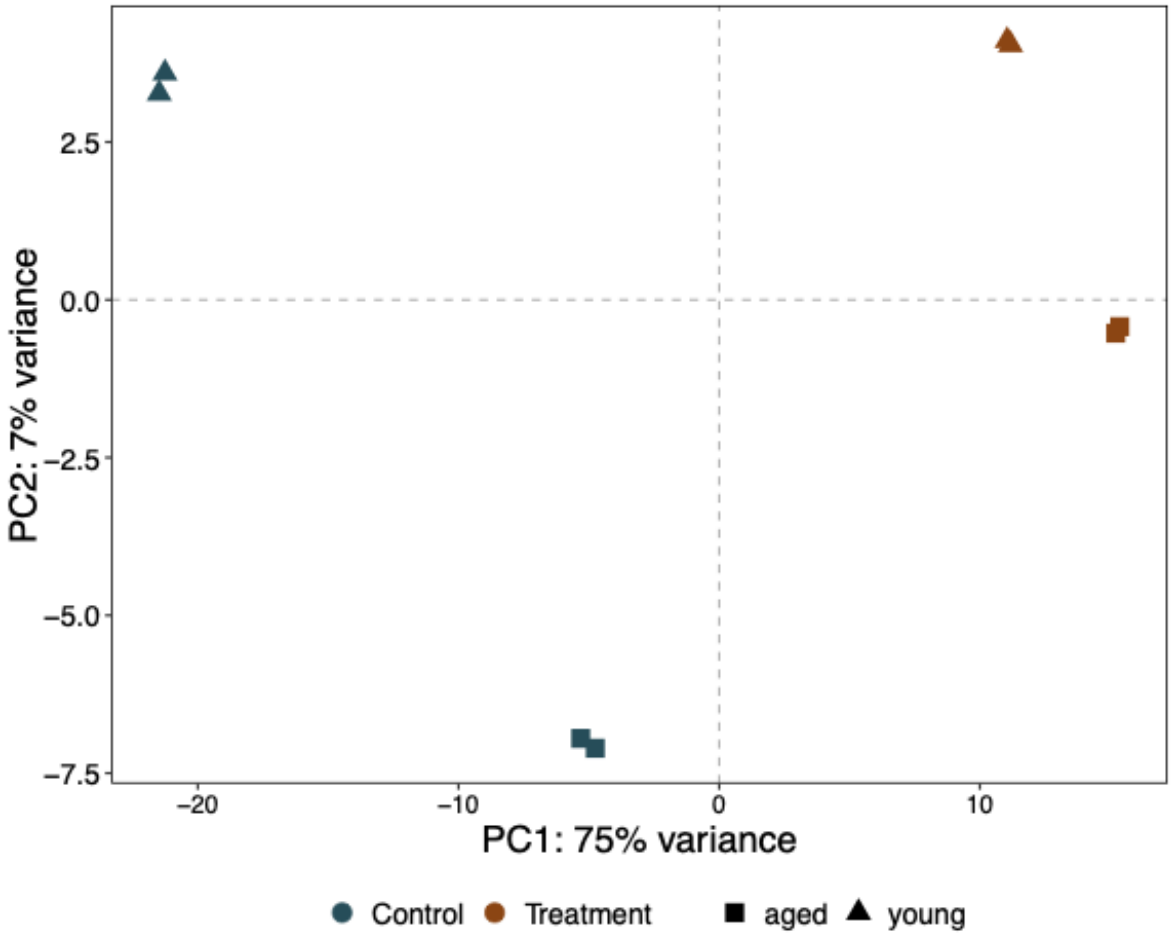


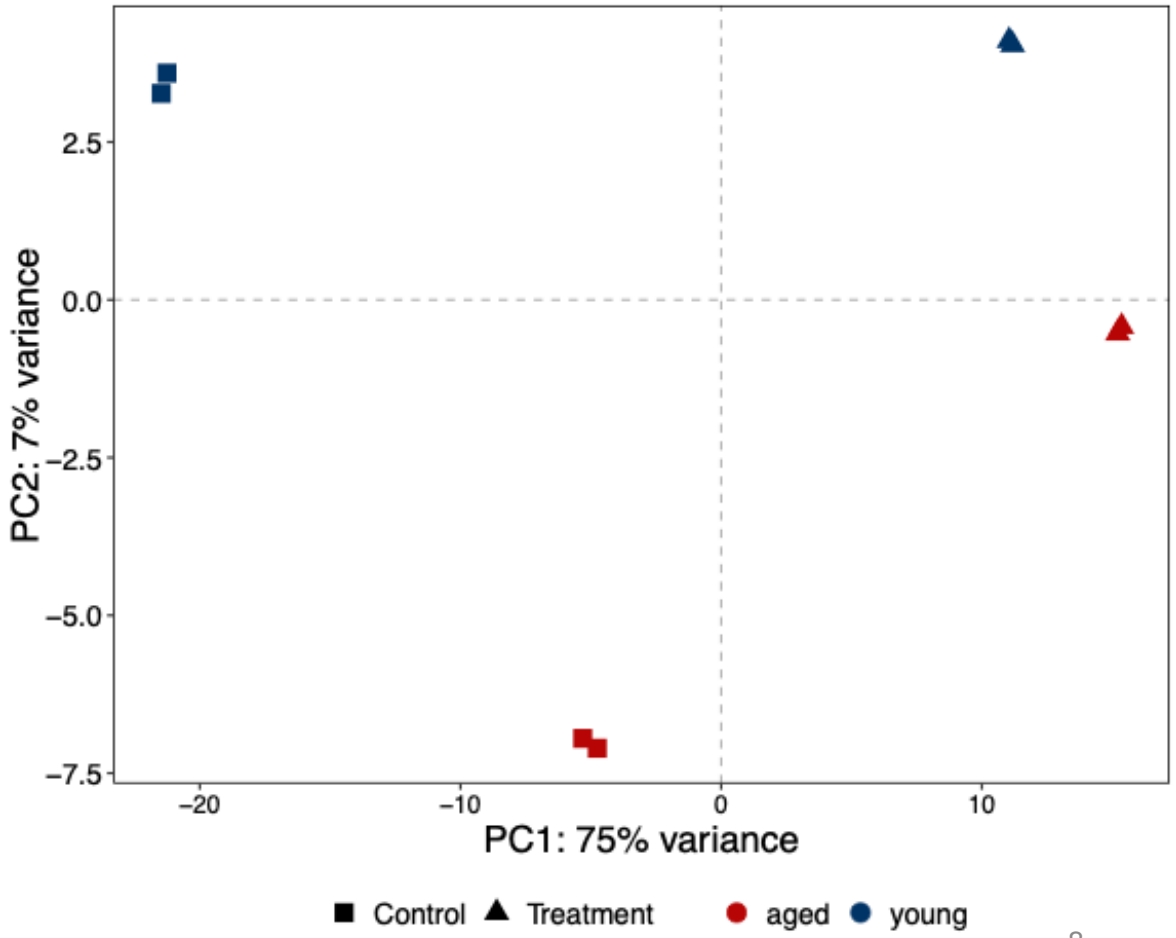
Principles of Good Visualization

- What can you say about the first plot?
- How about the second one?
- What would you make different?

What can we say about both plots?

Direct Attention

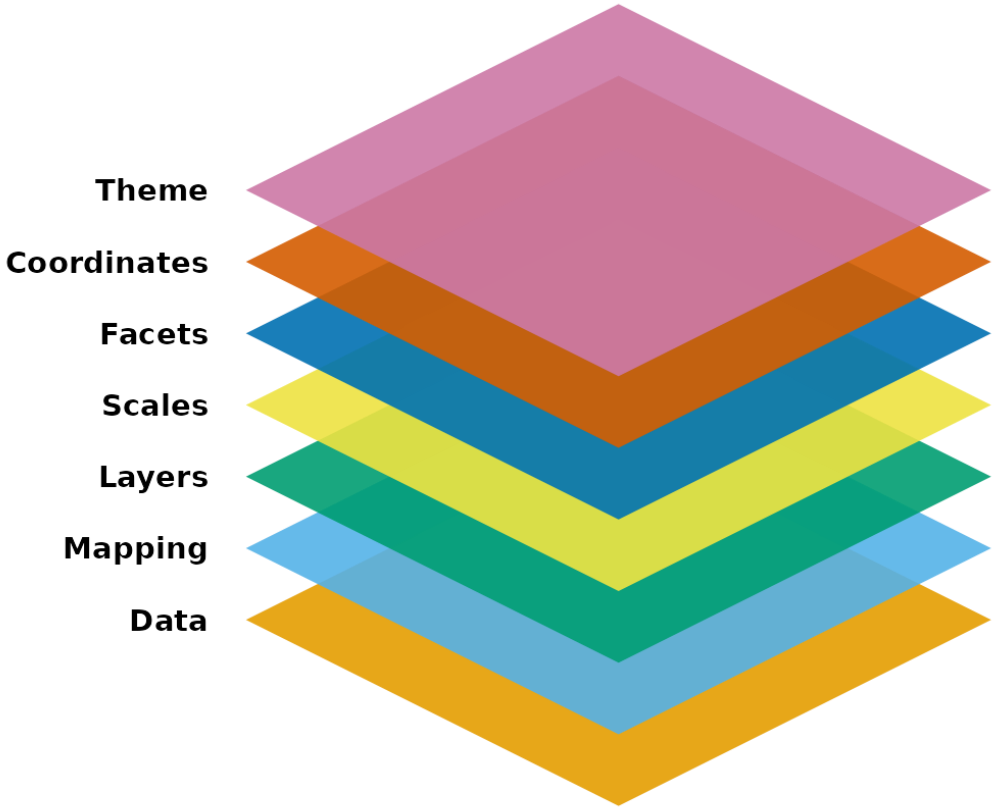




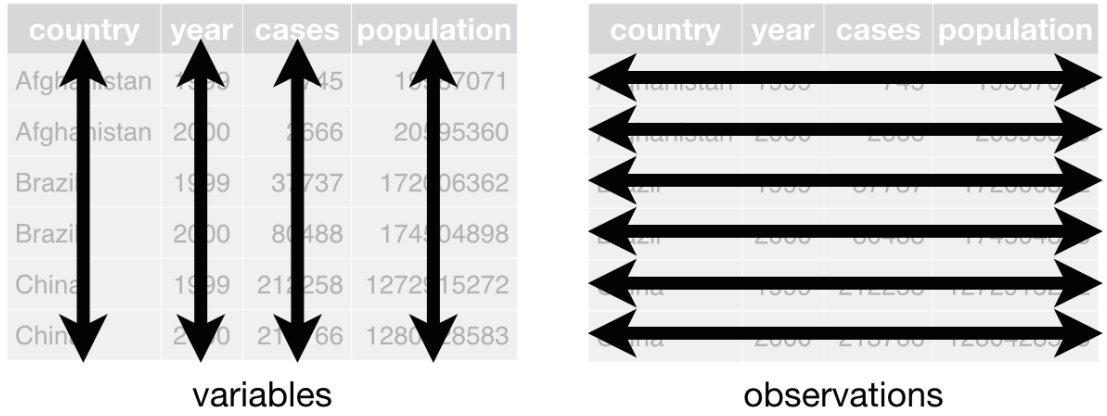
Introduction to ggplot2

ggplot2 uses a conceptual framework based on the grammar of graphics

7 composable parts



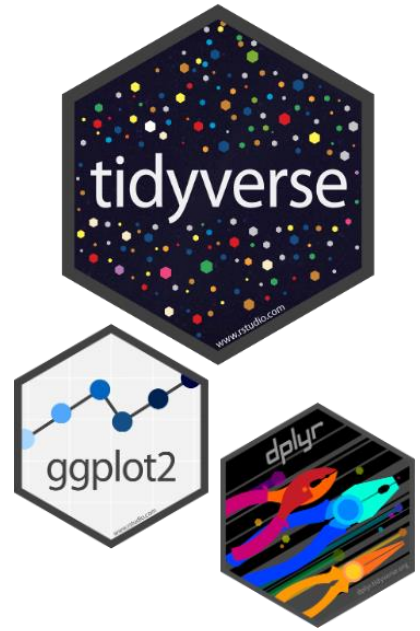
Mandatory elements: **data**, **mapping** and **layer**



Tidy data:
rows are observations and columns are variables

Pratical

Real life in DataViz



Communicate

Making Coding Easier

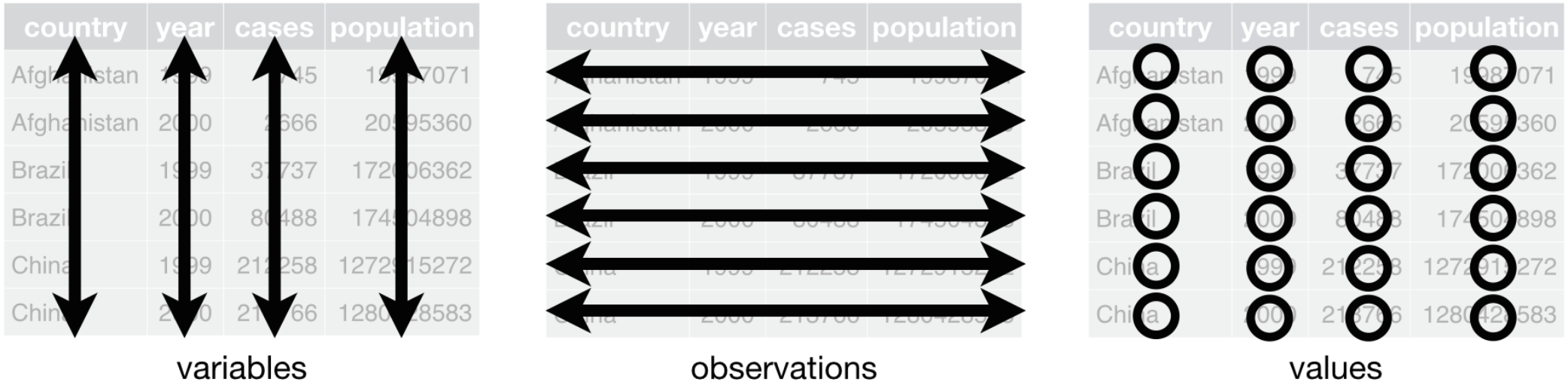
Pipe: pronounced "then"

Reverse assignment operator: "creates"

```
alpha_info_tab %>%
    filter(Study == "Exp1") ->
    Exp1
```

Variable

tidy format (we're going to explorer it in the next section), which briefly means a rectangular data frame structure where rows are observations and columns are variables



A good Plot, you cannot do, without first a bad Plot, doing.

"Content inspired by responses from ChatGPT, OpenAI's conversational AI model (accessed November 2024)."

Links for study

- R graph Gallery: https://r-graph-gallery.com
- Hadley Wickham: https://www.youtube.com/watch?v=9YTNYT1maa4
  - Books: https://r4ds.hadley.nz/data-visualize