

Introduction to \mathcal{R}

Session 4. ggplot2 Graphics

Dag Tanneberg¹

Potsdam Center for Quantitative Research
University of Potsdam, Germany
November 8/9, 2019

¹Quantitative Social Scientist, www.dtanneberg.de

Introduction

Before we start...

Here is our in-class solution to yesterday's puzzle. We have built in a semantic error. Can you spot and fix the error? Hint: Try different parameter settings.

```
r <- 1; n <- 10000
hits <- vector("logical", length = n)
for (i in seq(n)) {
  hits[i] <- sqrt(sum(runif(2, -r, r) ^ 2)) <= 1
}
4 * mean(hits)
```

Before we start...

In 1987, the Indiana General Assembly attempted to establish the value of π by legislative fiat. The **Indiana Pi Bill** goes back Edward J. Goodwin's scientifically proven method to square the circle.

```
pi; pi <- 3.2; pi; base::pi
```

```
## [1] 3.141593
```

```
## [1] 3.2
```

```
## [1] 3.141593
```

Now, an object sits on your workspace that has name `pi` and value 3.2. You can still access the mathematical constant, though. See the manual: `?"::"`.

Today's Agenda

- **Place:** Campus Griebnitzsee, House 7, Room 2.41
- **Coffee break:** 10:45 - 11:00
- **Lunch break:** 12:30 - 13:30
- **Materials:** Go to <https://github.com/dagtann/pcqr/>

Day	Start	End	Official Topic
2	09:15	10:45	Graphics
	11:00	12:30	Basic Statistics
	13:30	15:00	(G)LMs

A Taste of ggplot2

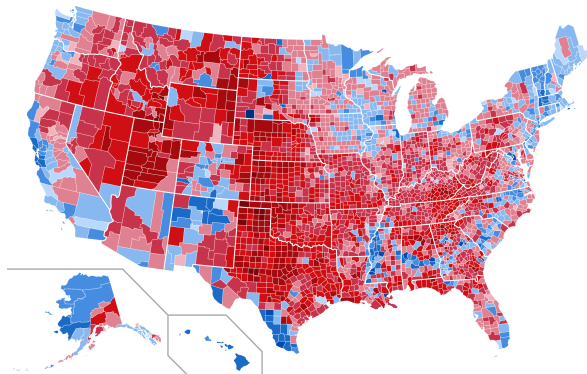


Figure 1: Vote Share in the 2012 Presidential Elections by County²

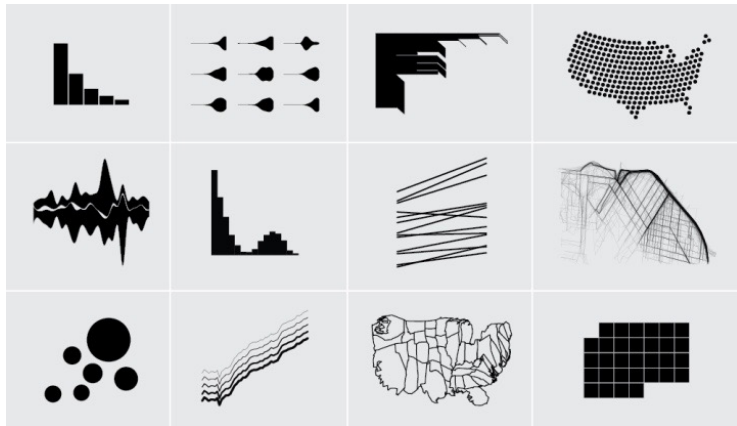
²Healan, K. 2019. Data Visualization. A Practical Introduction. Princeton: Princeton University Press, <https://socviz.co/maps.html> (last access: 11/05/2019).

Second Serving: Hierarchical Edge Bundling³



³Holtz, Y. 2017. The R Graph Gallery. <https://bit.ly/2QKdhLU> (last access: 11/05/2019).

An Embarrassing Riches of Options⁴



⁴Yau, N. 2016. 5 Tips for Learning to Code for Visualization.

<https://bit.ly/2yu0fe0> (last access: 10/11/2018).

Outline

- 1 Introduction
- 2 What is ggplot2?
- 3 How do you talk to ggplot2?
- 4 Summary

What is ggplot2?

Graphics Engines in \mathcal{R}

- Three major graphics systems in \mathcal{R}
 - base graphics: ships with \mathcal{R}^5
 - lattice developed by Deepayan Sarkar
 - ggplot2 developed by Hadley Wickham
- All build on Paul Murrell's grid Graphics
- All differ remarkably on usability & quality of output

⁵Check out Nathan Yau's tutorials at [Flowing Data](#).

What's the buzz about ggplot2?

- Implements the Grammar of Graphics⁶

“In brief, the grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars).”⁷

- Very generic, schematic approach to data viz
- Seamless integration with **tidyverse**
- Numerous extensions: Animations, maps, daps, etc.
- Delivers high quality results – fast.

⁶Wilkinson, L. 1999/2005. The Grammar of Graphics. New York: Springer Science + Business.

⁷Wickham, H. 2016. ggplot2. 2⁸ ed. New York: Springer Science + Business.

How do you talk to ggplot2?

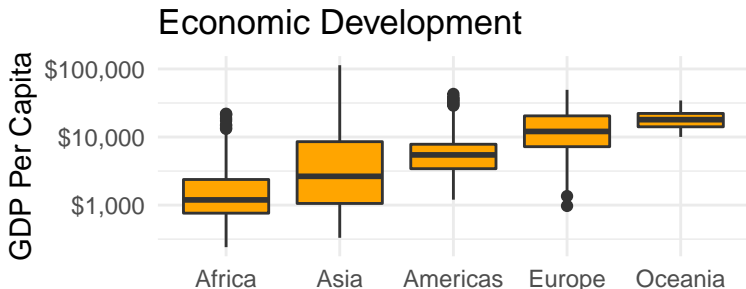
Intuition

- Idea: logical connection b/w data & plot elements
 - Example: Gender \mapsto Color
 - a.k.a. aesthetic mapping
- Answer the following questions for ggplot2:
 - 1 What is your data?
 - 2 What relationships do you want to see?
 - 3 How do you want to see it?
 - 4 What additional information do you want to see?
 - 5 What scales, axes, labels should be shown?
- Complete reference:
 - Hadley's book: **ggplot2**
 - ggplot2 Online Reference: **tidyverse**
 - Kieran Healy's book: **Data Visualization**

A Schematic for Making a Plot

```
p <- ggplot(  
  data = <data>, # 1. What is your ("tidy") data?  
  mapping = aes( # 2. What var's map unto the plot?  
    <aesthetic> = <variable>, # x  
    <aesthetic> = <variable>, # y  
    <...> = <...> # colour, fill, shape, size,  
    # alpha, ...  
  )  
) + # Add layers to your plot  
  geom_<type>(<...>) + # Define your plot type  
  scale_<mapping>_<type>(<...>) + # Adjust scales  
  coord_<type>(<...>) + # Adjust co-ordinates  
  labs(<...>) # Label plot elements  
ggsave(file = <...>, plot = <...>, ...) # Save ur work
```

Our Target Graph



Source: Gapminder.

- Open the notebook file notebook_04.Rmd.
- Using the notebook, answer questions 1-5.
- Use the generic schematic and **tidyverse** to generate the plot.

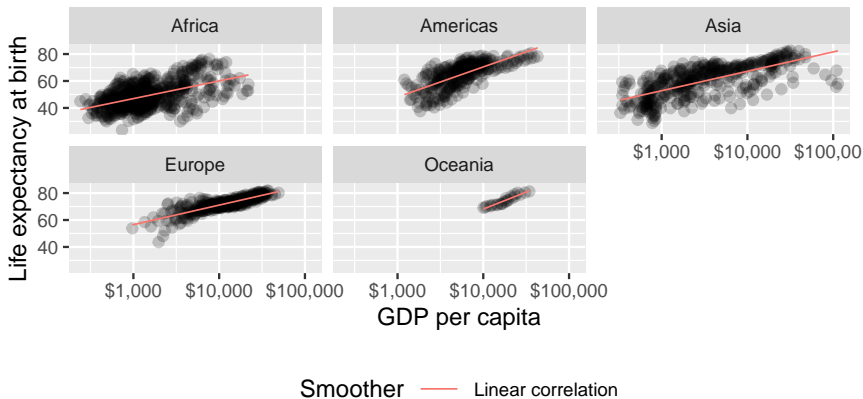
Summary

Summary

- There are three major graphics engines: base, ggplot, lattice.
- ggplot2's attraction
 - very generic, schematic approach to viz
 - fast, visually pleasing results
- Principle strategie:
 - a. Connect (map) data to plot elements
 - b. Layer plot elements
- Read Healan, K. 2019. Data Visualization. A Practical Introduction. Princeton: Princeton University Press. Available at <https://socviz.co/> (last access: 11/05/2019).

Self-practice: Replicate these figures.

Life Expectancy vs. Economic Development



Self-practice: Replicate this figure.

Life Expectancy vs. Economic Development

Graph shows 1997, 2003, and 2007.

