ggplot2 Introduction

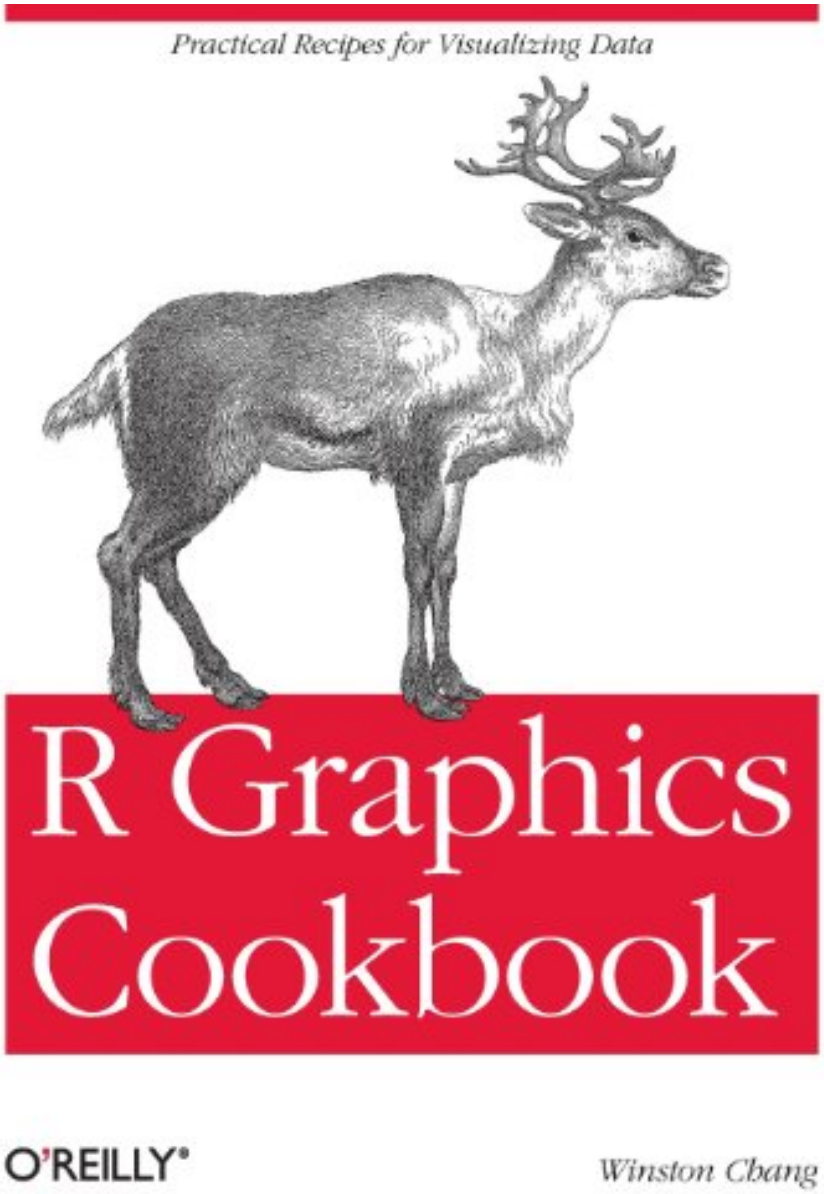
Jean-Baptiste Lecomte

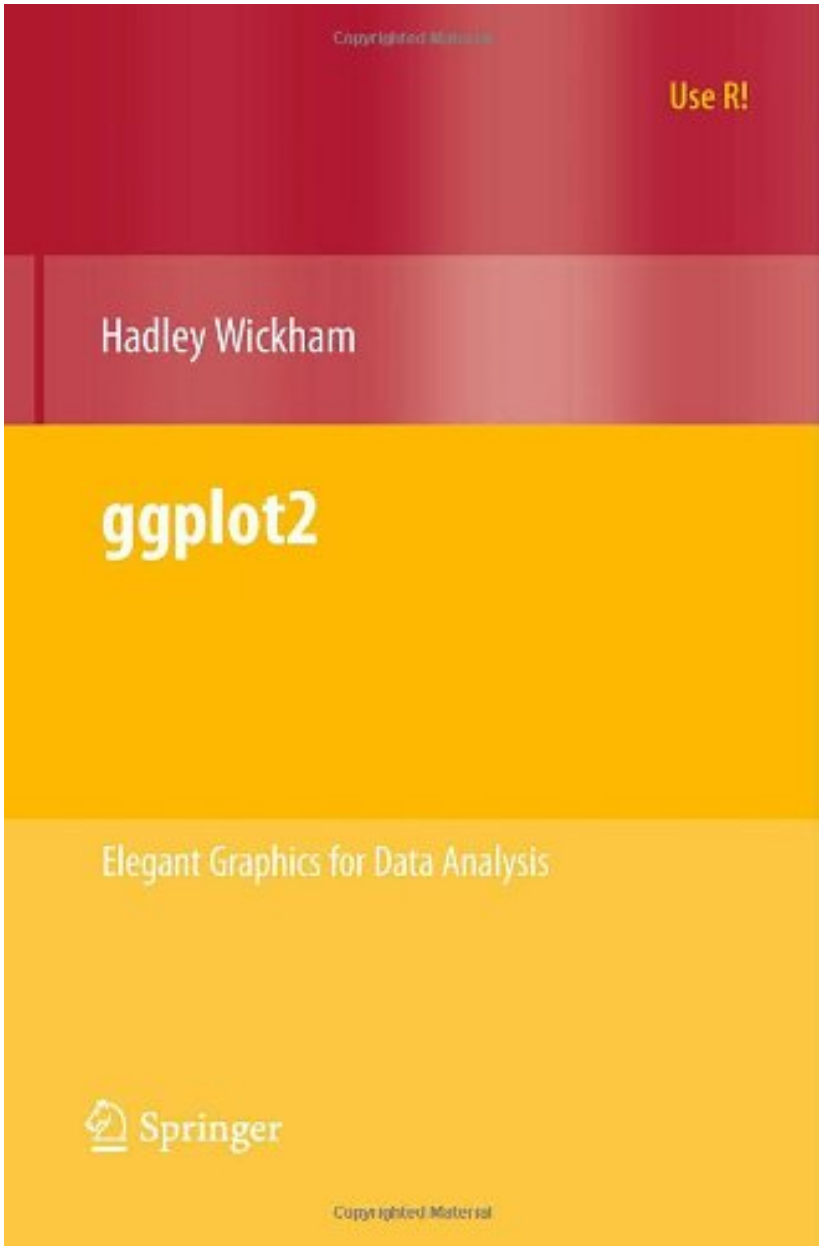
December 3, 2014

Introduction

- developped by Hadley Wickham (Rice University, Houston, USA)
- highly recoomaded R packages to work with ggplot2: reshape and plyr (also developped by H. Wickham)
- first version called in 2007

Useful books





Online ressources

- R code related to ggplot2 cookbook:
  http://www.cookbook-r.com/Graphs/
- R code related to useR! ggplot2 book:
  http://ggplot2.org/book/
- ggplot2 official documentation:
  http://docs.ggplot2.org/current/
- Google groups to ask questions:
  ggplot2@googlegroups.com
- Github repository:
  https://github.com/yhat/ggplot/

Introduction

- based on new aesthetic principles
- based on *The grammar of graphics* developed by Wilkinson in 2005
- efficient way to produce simple graphics with a length reduction of R code

Forget about R base graphics:

`plot()`, `hist()`, `par()`, `layout()`, `points()`, `lines()`,`legend()`

Principle

ggplot2 is based on a **layer** system which can be used as objects.

Main layers

- data → raw data
- mapping → graphic projection
- geom → geometric objects (points, lines, polygons, ...)
- stat → statistics transformation (histogram, model)
- scale → aesthetics customization (color, shape, size, axes, legend)
- coord → coordinate system (axes, grid)
- facet → subdivision (lattice, trellis)

Base functions

ggplot2 is based on two functions:

1. `qplot()` for **q**uick **plot**
   - easy and fast, but too simple in most cases
   - `qplot(x, y, data=data)`

2. `ggplot()`
   - more complex but more powerful and flexible by adding `layers`
   - `ggplot(data=data, aes(x, y)) + layers`

Getting Started

Data format

Always work with `data.frame`

Our data frame:

```
  str(df.data)

## 'data.frame': 0 obs. of  0 variables
```

Getting Started

Data format

Always work with `data.frame`

Our data frame:

```
  str(df_data)

## 'data.frame': 0 obs. of  0 variables
```