ggplot2 Introduction

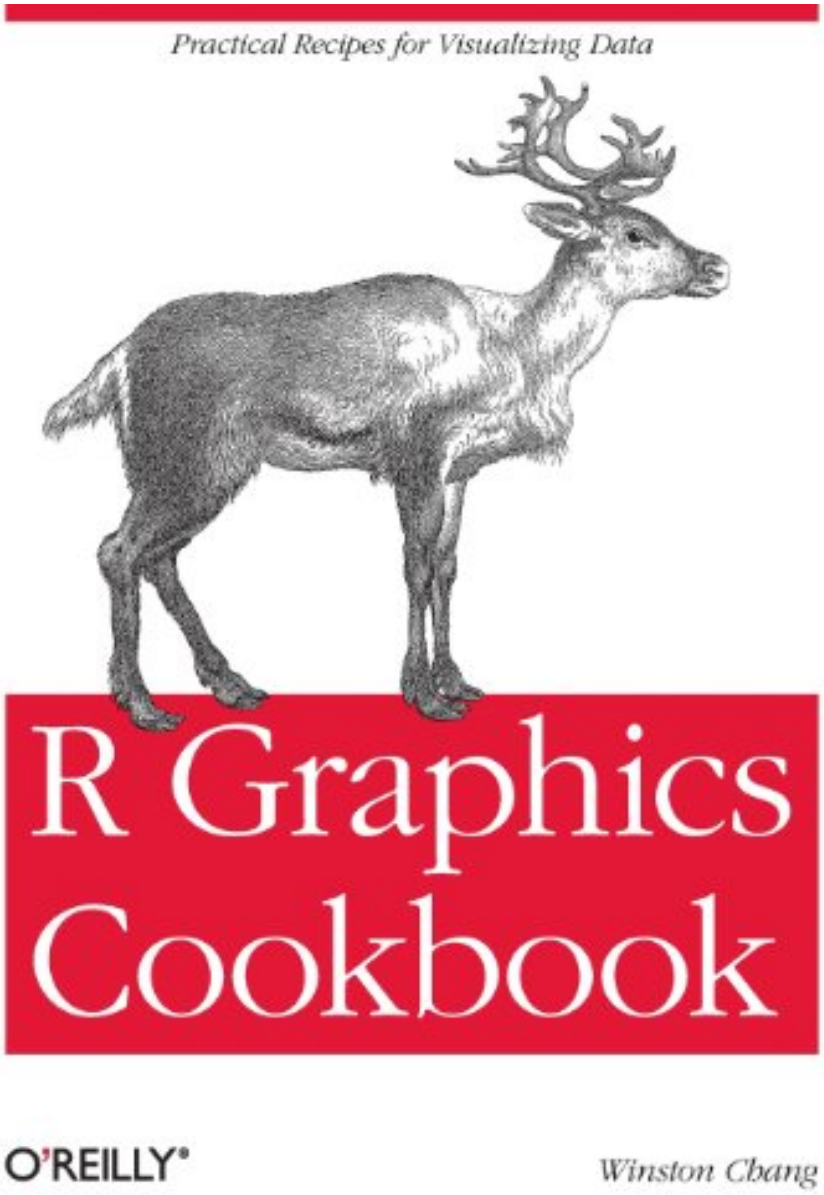
Jean-Baptiste Lecomte

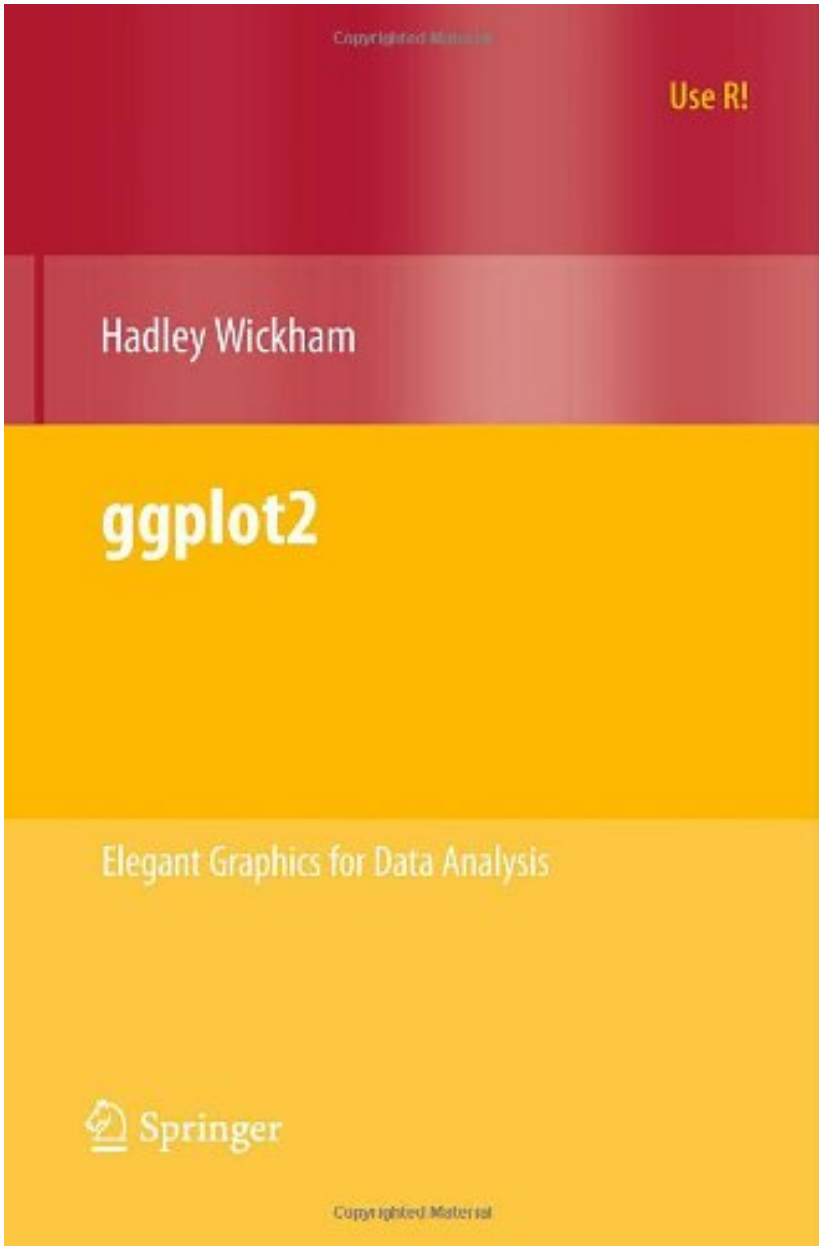
December 8, 2015

Introduction

- developped by Hadley Wickham (Rice University, Houston, USA)
- highly recommanded 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 a `data.frame`

Our data frame:

```
  head(df_data)

##     Year Month DURATION_MINUTES AREA Avg_net_depth Avg_net_
## 476 2005     7                21  5AB    -0.3162932     0.39
## 477 2005     7                20  5AB    -0.4353716     0.43
## 478 2005     7                21  5AB    -0.4418207     0.30
## 479 2005     7                21  5AB    -0.2340669     0.13
## 480 2005     7                20  5AB    -0.1713032    -0.02
## 481 2005     7                20  5AB    -0.1683089    -0.35
##           Lon    Lat         X       Y     X_km     Y_km Pres
## 476 -127.970 51.160 572025.0 5668122 572.0250 5668.122    1
## 477 -127.995 51.140 570307.2 5665874 570.3072 5665.874    1
```

Scatter plot: Depth and Biomass

```
scatter.plot <- ggplot(data=df_data, aes(x=Avg_net_depth,
                                         y=Biomass)) +
                geom_point()
print(scatter.plot)
```