R-intro - Session 4 (part 2 - ggplot2)

João Gonçalves

1 de Agosto de 2018

Contents

Objectives of session 4

- A brief tour to the `tidyverse`
- Data manipulation with the `dplyr` package
  - Main `dplyr` verbs
  - The pipe `%>%` operator
- Making pretty graphics with `ggplot2` - design principles and main plot types
  - Histograms
  - Boxplots
  - Bar-plots
  - Scatter-plots
  - Prediction, trend and smooth lines
  - Making plots by groups
  - Facets

The *ggplot2* package

---

ggplot2 is a powerful data visualization package for the statistical programming language R providing an alternative for making graphs in this software.

The creators of *ggplot2* define it as:

"a system for declaratively creating graphics, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details."

So what is the rationale behind the *grammar of graphics*?

Put simply, a statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

So, building on this, the basic idea behind the *grammar of gaphics* is to break up graphs into semantic components and thus giving you the ability to independently specify each building block of a plot.

Combining them allows to create just about any kind of graphical display you want.

These building blocks include:

- Data

- Aesthetic mapping (e.g., what to put in x or y, shape, size, colors)
- Geometric object (e.g., points, lines)
- Statistical transformations (e.g., scale, log)
- Scales
- Coordinate system (e.g., cartesian, polar)
- Position adjustments
- Faceting

Before proceeding you need to install the *ggplot2* package:

```
install.packages("ggplot2", dependencies = TRUE)
```

Then load the package using:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```