Data Visualization with ggplot2

R for Data Science workshop

2019-05-01 (updated: 2019-07-15)

Data Visualization with ggplot2

Outline

- Overview of ggplot2
- Prerequisites
    - Tidyverse
    - Data frames
- Grammar of Graphics
- Exploratory vs Expository Plots

Data science workflow

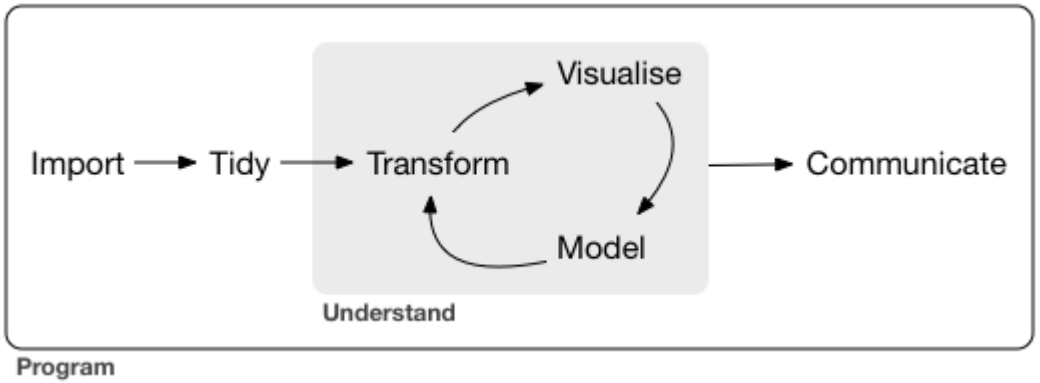


Image source: R for Data Science by Hadley Wickham & Garrett Grolemund.

Data science workflow

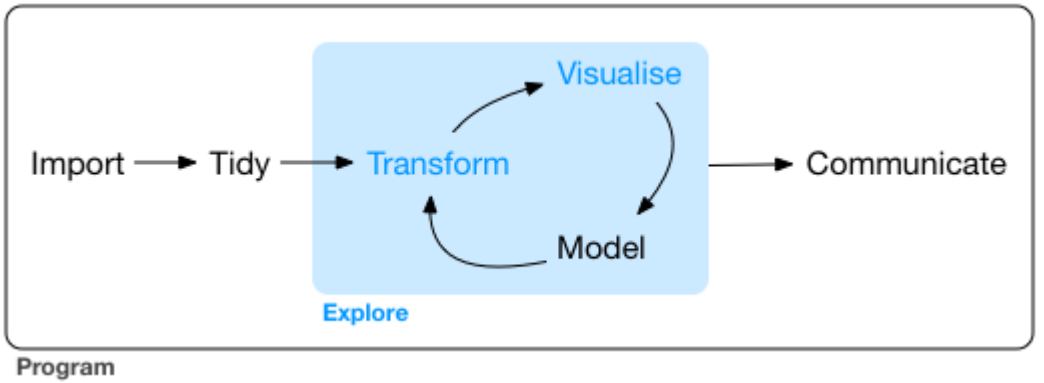


Image source: R for Data Science by Hadley Wickham & Garrett Grolemund.

ggplot2

From the official ggplot2 website:

> ggplot2 is 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.

- R package
- Developed by Hadley Wickham (PhD, Iowa State, 2008)
- Now part of the tidyverse
- Other R graphics systems: base, lattice, grid
- Extensible
- Worth the learning curve

Prerequisites

Tidyverse

- Opinionated collection of R packages
- Share underlying design philosophy, grammar, and data structures

Data frames

- Rectangular data structure
- Columns represent variables
- Rows represent observations
- Tidyverse data frames are called "tibbles"

Tidyverse packages work with tidy tibbles

Demo

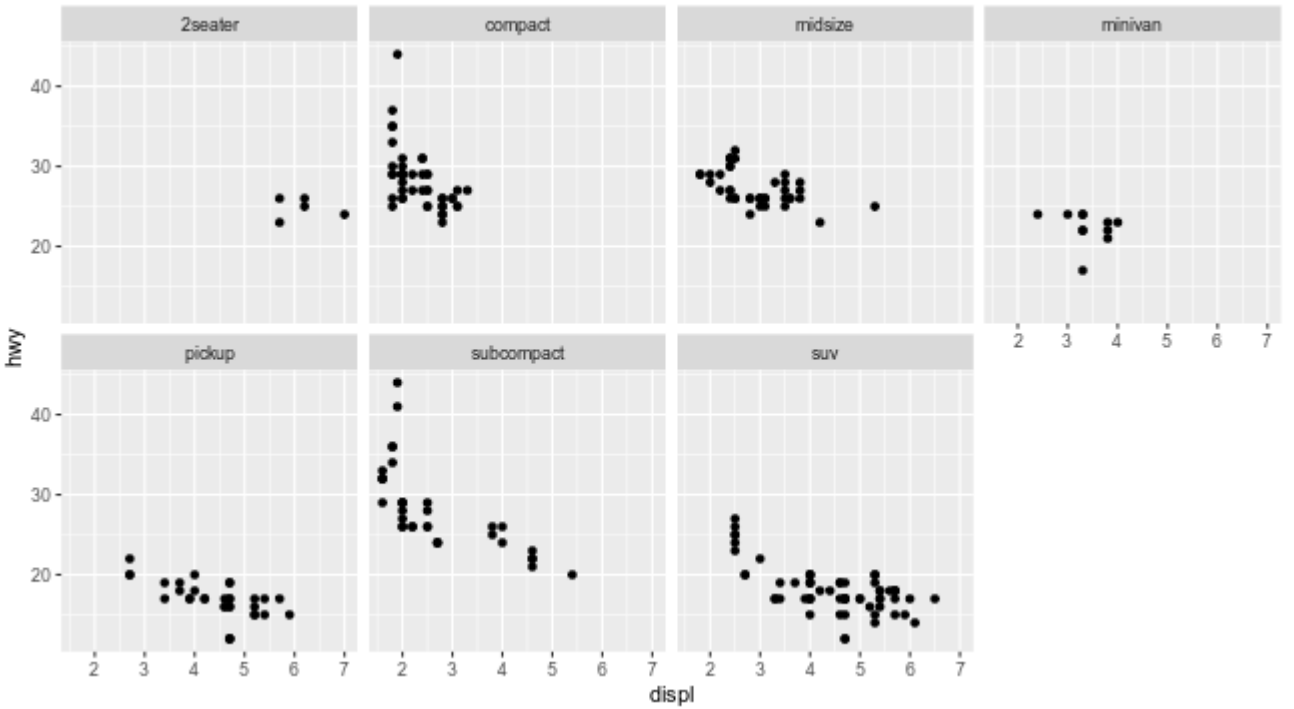
Grammar of Graphics

- **Data** (data frame in tidy format)
- Layers of **geom**etric objects and **stat**istical transformations
- **Scale**s that map data values to (and from) **aes**thetic space values (location, color, size, etc.)
- **Coord**inate system (Cartesion, map projections, polar)
- **Facet**ing (breaking data into subsets)
- **Theme**s

Facets are subplots

For example,

```
ggplot(data = mpg, mapping = aes(displ, hwy)) +
  geom_point() +
  facet_wrap(vars(class), nrow = 2)
```



Grammar of Graphics

We will focus on 3 main elements of the grammar:

- **Data** (data frame in tidy format)
- Layers of **geom**etric objects
- Mappings of data values to visual properties (**aes**thetics) of geoms

ggplot2 template

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>)) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>)) +
  ...
```

If the geoms use the same aesthetics:

```
ggplot(data = <DATA>, mapping = aes(<MAPPINGS>)) +
  <GEOM_FUNCTION>() +
  <GEOM_FUNCTION>() +
  ...
```

Demo

Full ggplot2 template

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(
     mapping = aes(<MAPPINGS>),
     stat = <STAT>,
     position = <POSITION>
  ) +
  <COORDINATE_FUNCTION> +
  <FACET_FUNCTION> +
  <SCALE_FUNCTION> +
  <THEME_FUNCTION>
```

Required elements: <DATA>, <GEOM_FUNCTION>, <MAPPINGS>

Other elements have sensible defaults.

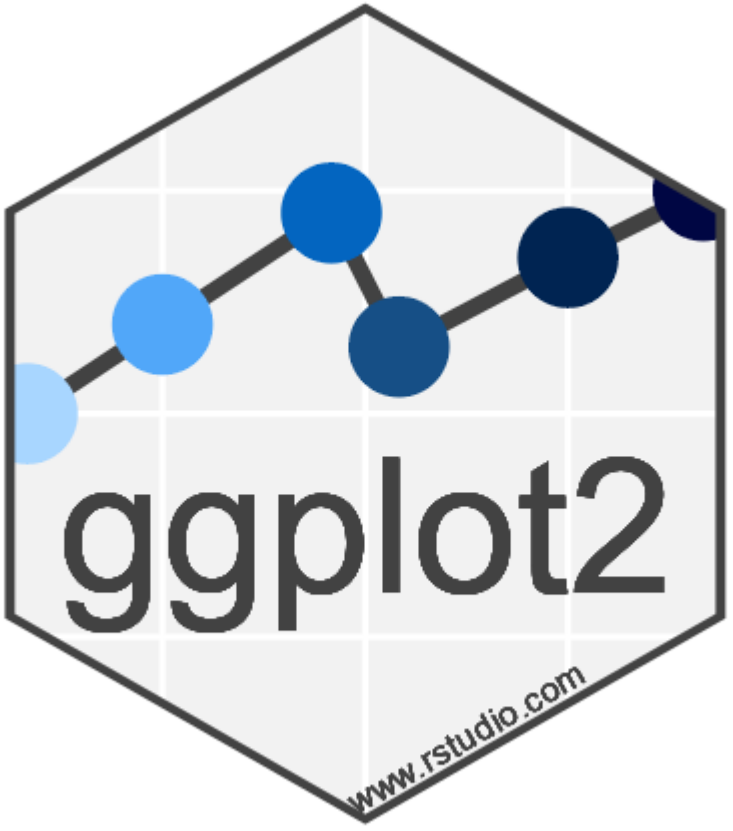
Some named graphs and their geoms

Name	Geom
Scatterplot	geom_point()
Linegraph	geom_line()
Histogram	geom_histogram()
Boxplot	geom_boxplot()
Barplot	geom_barplot()

Demo

ggplot2 is extensible

You can create almost any 2-dimensional plot with ggplot2



Source code is in demos/data-visualization-with-ggplot2/ggplot2-hex-logo.R

15:00

Your turn

Data visualization with ggplot2

your-turn/01-data-visualization-with-ggplot2.Rmd