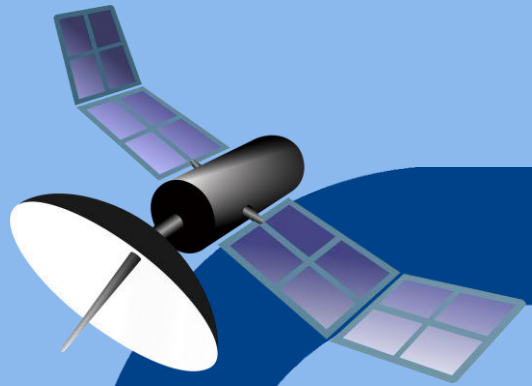
Research Institute for
Farm Animal Biology



Satellite Day @FBN

Making plots with `ggplot()`

Satellite Day @FBN – 05-March-2025

Learning Objectives

By the end of the workshop, you will be able to:

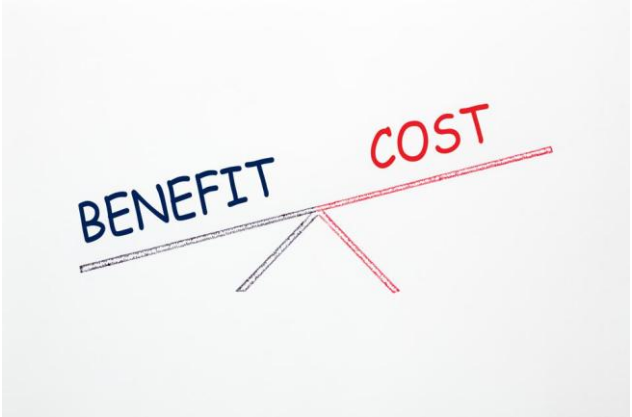
1. Understand the philosophy and structure of ggplot2
2. Create basic and advanced ggplots
3. Customize plots effectively (themes, labels, scales, etc.)
4. Work with different types of data and visualizations
5. Export high-quality plots for publications

The Grammar of Graphics: `ggplot()`

- The Grammar of Graphics
  … theoretical foundation

- The `ggplot()` function
  … the benefit of beeing general has costs

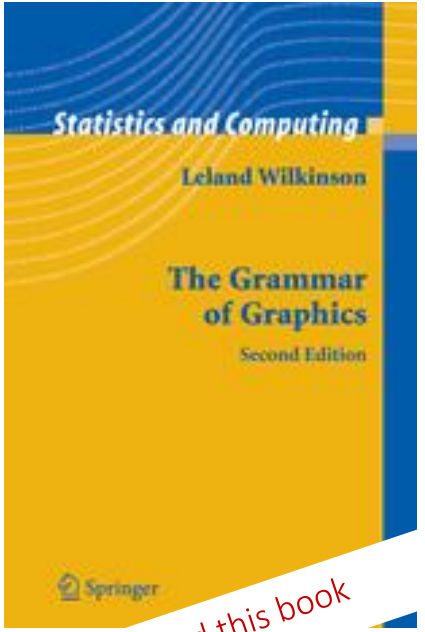
- Drawing anything
  … the benefits by far outweigh the costs



How it started

- `Grammar of Graphics` has been developed by Leland Wilkinson
  - 1st edition of his book in 1999
  - No focus on chart type, beauty of plots
  - Theoretical deconstruction of data graphics and how to design the system that allows to draw any plot

- `Grammar of Graphics` became foundation of different graphic applications, including the R package ggplot2

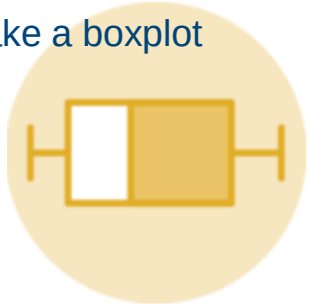


No need to read this book

One concept for all chart types !

- But there are many more than 4 chart types

one function to make a boxplot



Boxplot

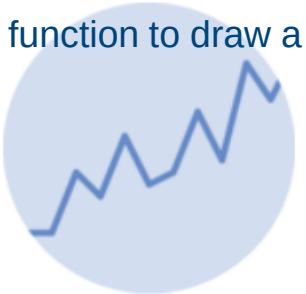
one function to make a bar plot

Barplot

one function to make a scatterplot

Scatter

one function to draw a line plot



Line plot

The idea of `Grammar of Graphics`

- Just like constructing **a meaningful sentence** in a language (with a subject, verb, and object), **creating a ggplot** involves combining different components in a structured way.

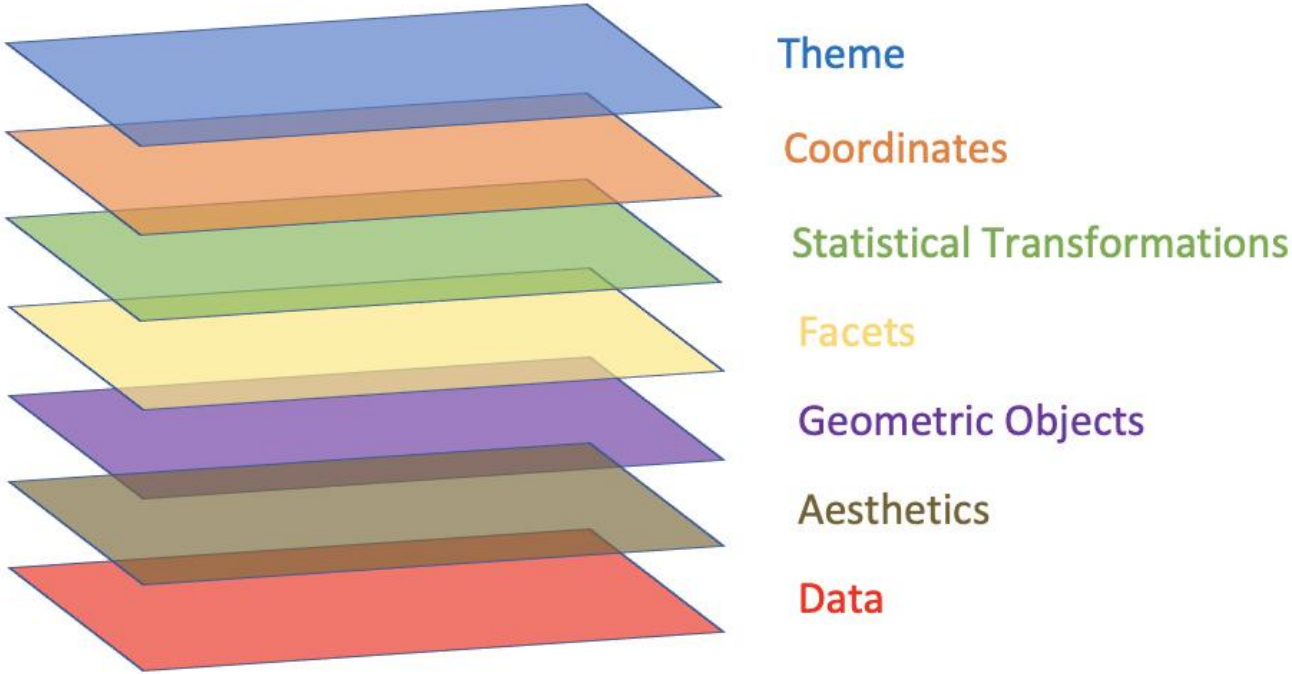
For example, in English:

- *"The cat (subject) sits (verb) on the table (object)."*

In ggplot2:

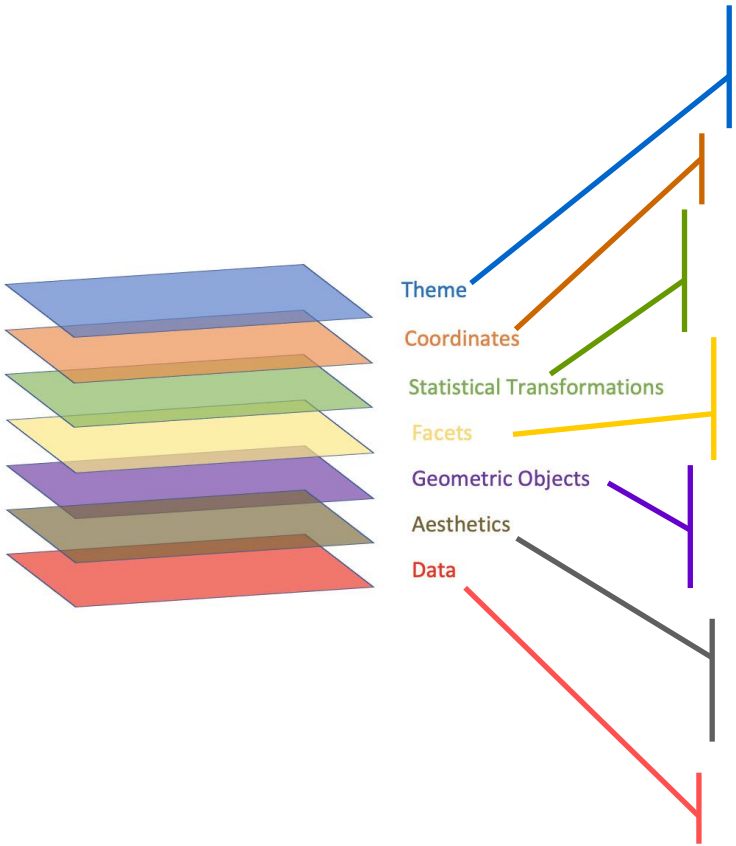
```
ggplot(data) (subject) +
aes(x, y) (verb: mapping variables to aesthetics) +
geom_*() (object: specifying the type of visualization).
```

7 main components of the `Grammar of Graphics`



https://towardsdatascience.com

7 main components of the `Grammar of Graphics`



Themes provide a variety of options to design all non-data elements (legend, background, or annotations)

Definition of the coordinate system

Including statistical models and summary statistics in the plot

Specifications of sub-plots (plotting several variables within your data next to one another in separate plots

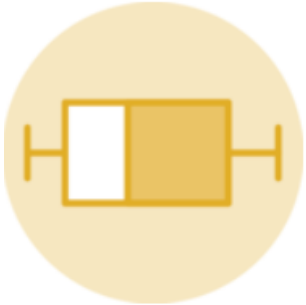
Shapes / geometric objects used to represent the data (plot type, added horizontal lines..)

The scales onto which selected variables are mapped (display one variable on the x-axis and another on the y-axis)

The actual data to be plotted

Costs and benefits

Line plot

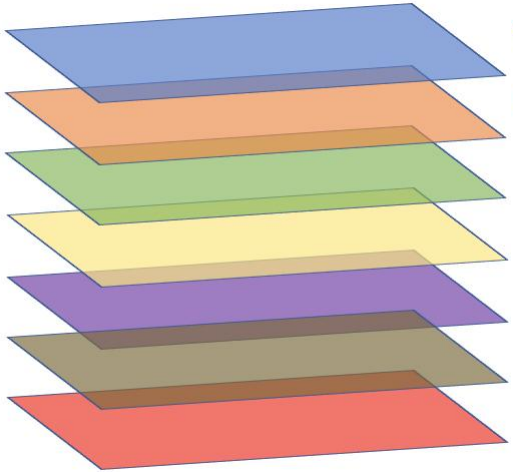


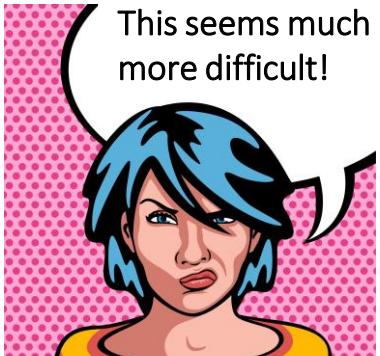
Boxplot

Scatter

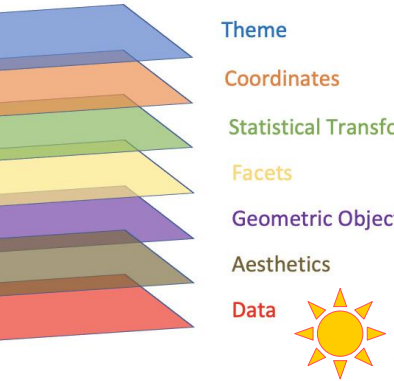
Barplot

versus



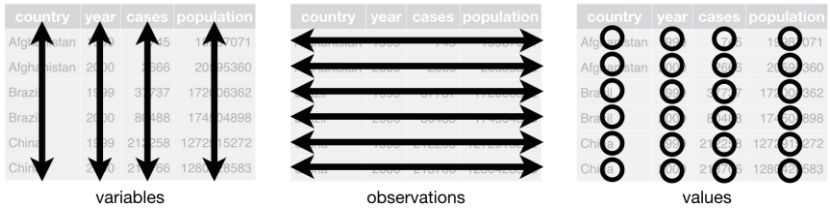


`Grammar of Graphics`: Data



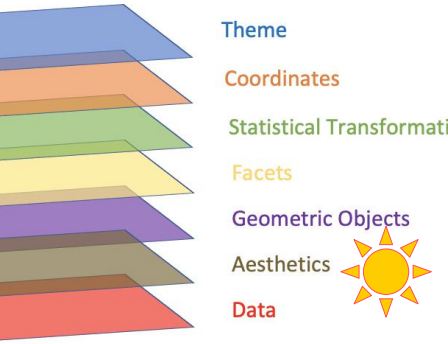
- Grammar requires a **tidy data format**
- *"Tidy datasets are all alike, but every messy dataset is messy in its own way."* — Hadley Wickham

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.



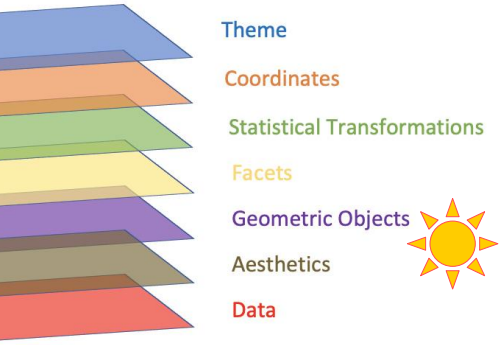
R for Data Science

`Grammar of Graphics`: Aesthetics

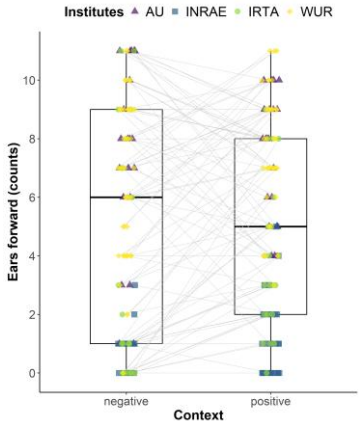


- Link variables in the data to graphical properties in the geometry
  - Each column in the tidy data set has a name and `ggplot()` needs to „know“ about the column names
  - The mapping defines which variables will be plotted on, *e.g.* the x-axis and the y-axis, which variable is the colour and which one the size of the symbols

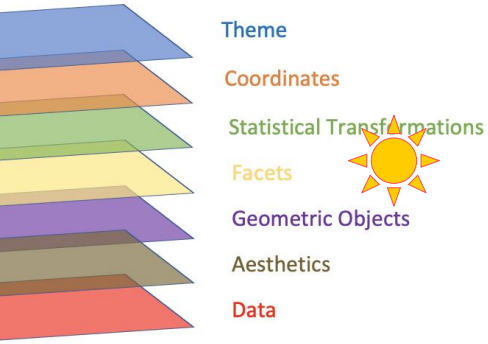
`Grammar of Graphics`: Geometries



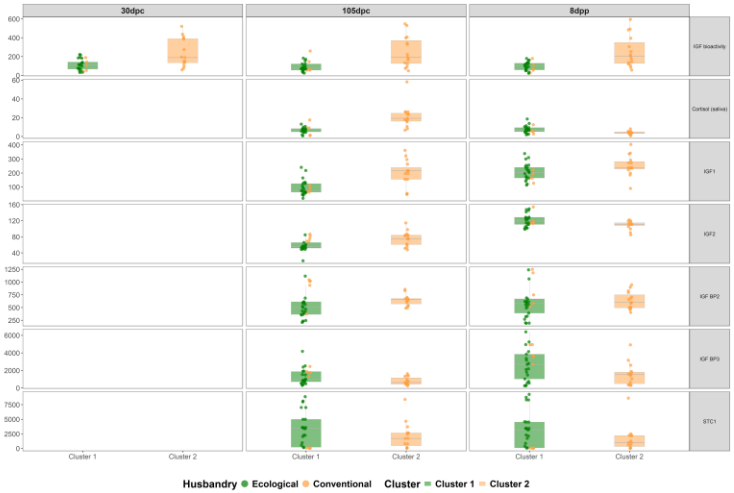
- Interpretation of the aesthetics to the graphic itself, it determines your plot type
  - the x- and y-coordinates can be plotted as points (`geom_point`) or lines (`geom_line`)
  - geometries can be more complex, *e.g.* box plots (`geom_boxplot`)
  - multiple layers of different geometries can be in one plot (+ + + +)



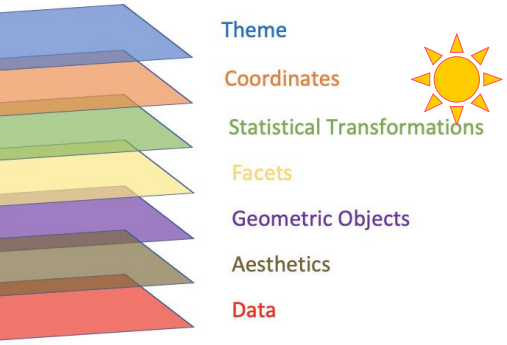
`Grammar of Graphics`: Facet mapping



- Link variables in the data to panels in the facet layout
- Variables of the data set are used to define the sub-plots, *i.e.* many plots with the same type of plot (`facet_grid`, `facet_wrap`)

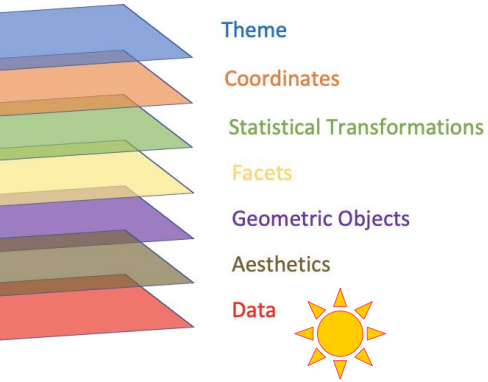


`Grammar of Graphics`: Statistical transf.

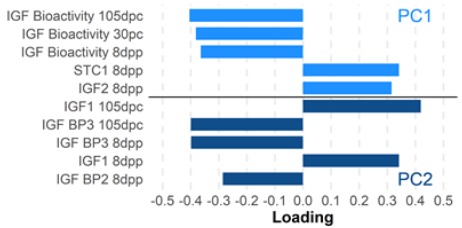


- You can even display data that are not contained in your data table
- Transform input variables to displayed values
  - count number of observations in each category for a bar chart
  - calculate summary statistics for a box plot

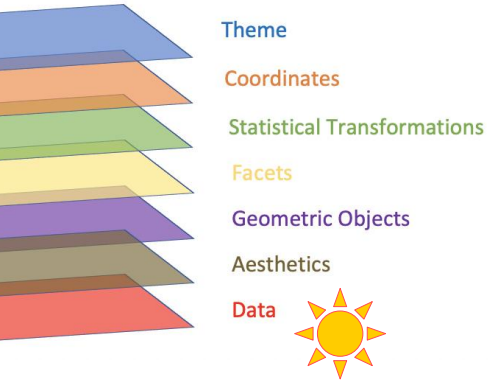
`Grammar of Graphics`: Coordinates



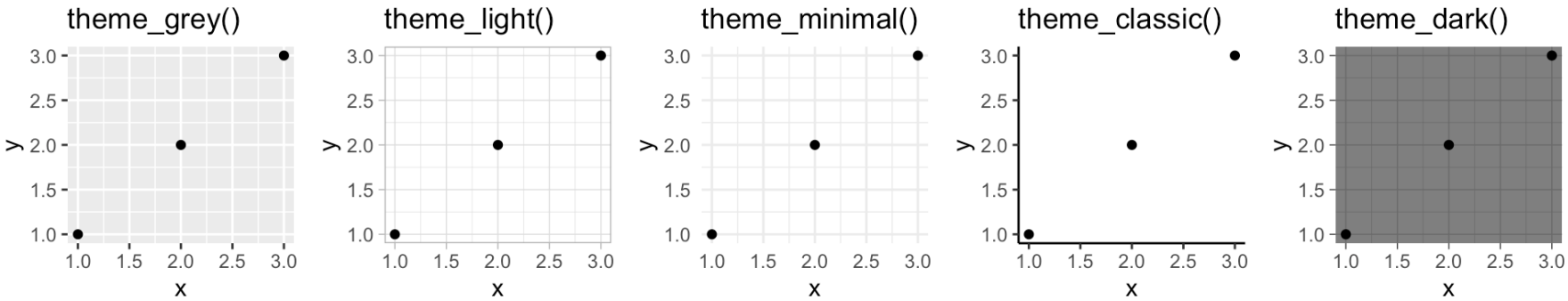
- Variables are mapped, scaled and applied to a geometry
- But in the end, the position values (i.e. not color, size) are interpreted by a coordinate system
- The coordinate system defines the physical mapping of the aesthetics to the paper
- Cartesian coordinate system is the most common one
- Flipped coordinate system is sometimes a good option
- But also: polar coordinates, spatial data (longitude, latitude) in cartography with many different projections



`Grammar of Graphics`: Themes

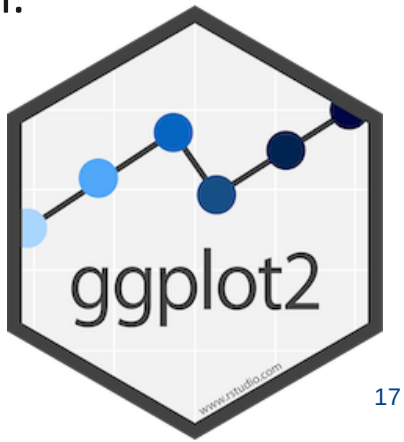


- Visual look of the plot, not related to the data
  - grids, ticks
  - labels
  - colour of background
  - *etc.*

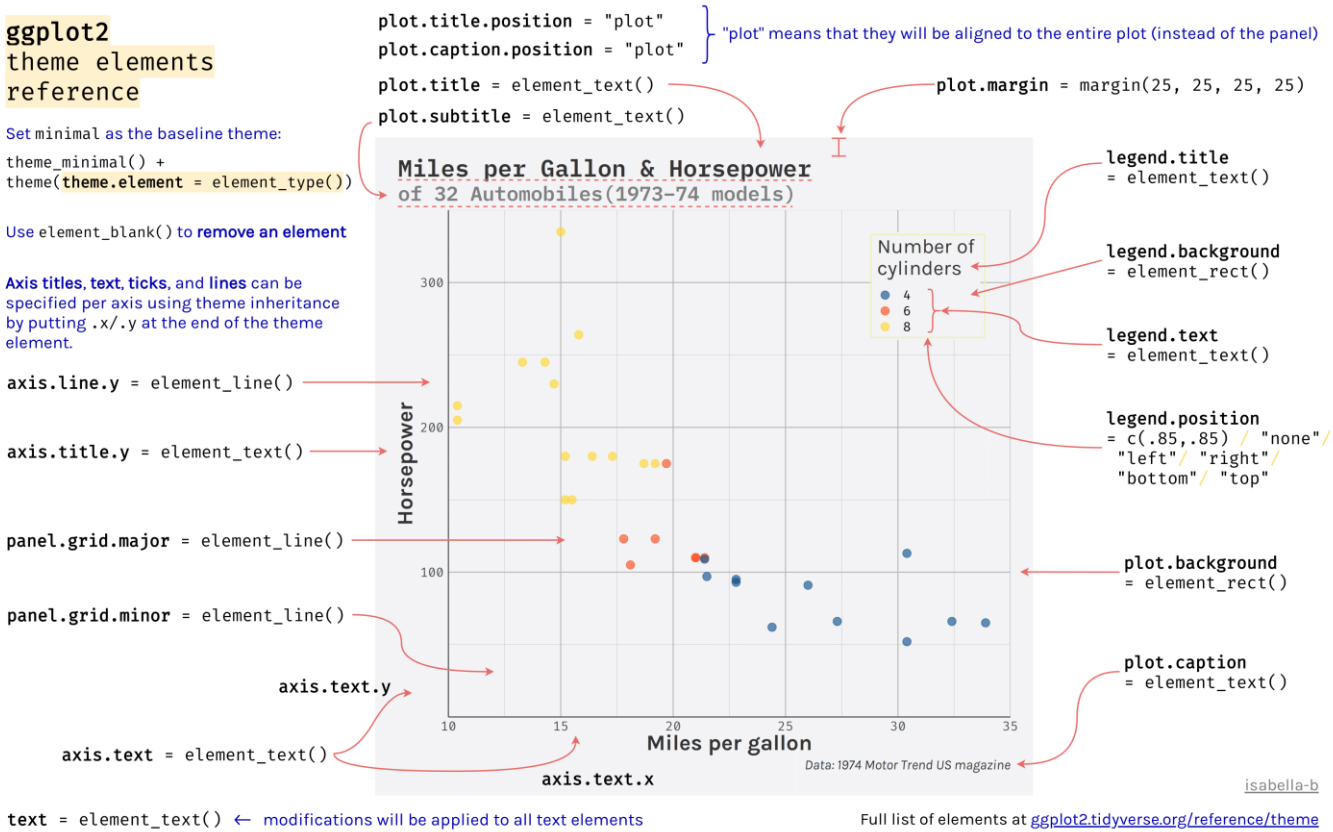


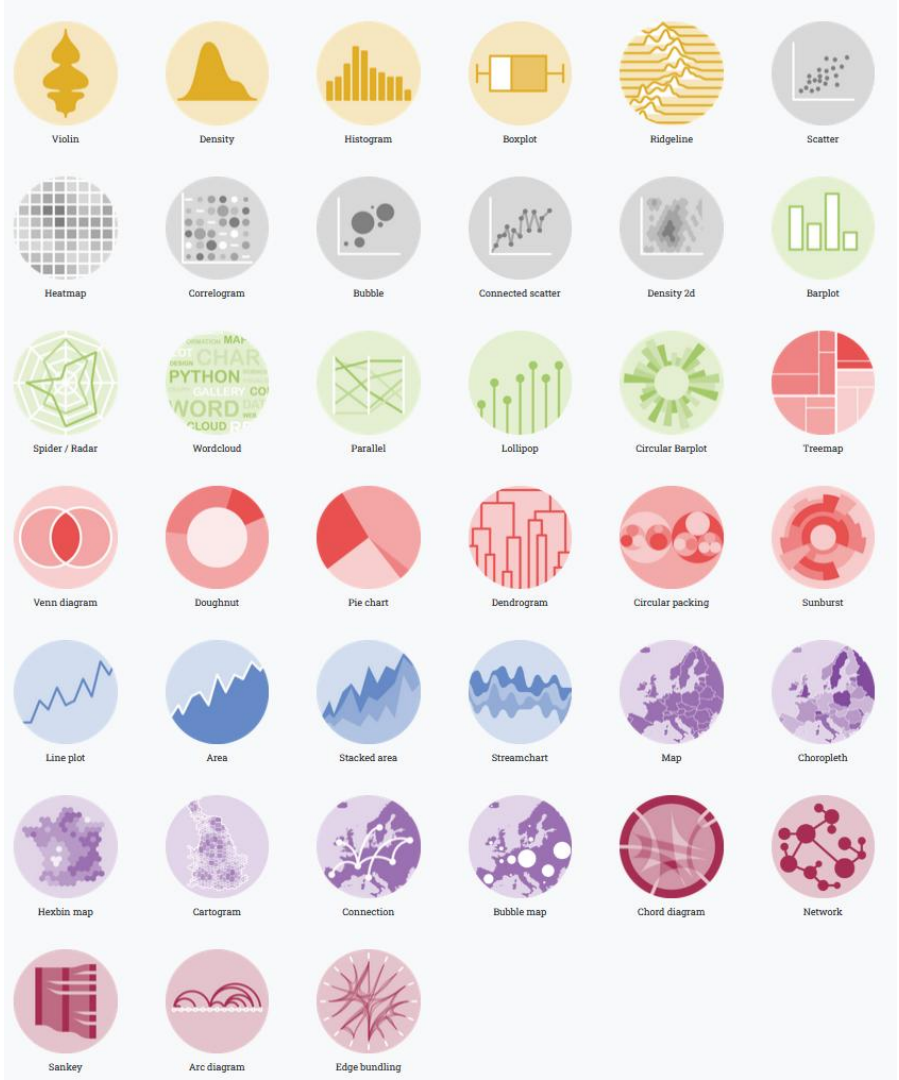
The R package ggplot2

- `ggplot2` is a data visualization package for **R** that is based on Leland Wilkinson's `` `Grammar of Graphics` ``, further developed by Hadley Wickham.
- `ggplot2` is a really powerful and flexible system for creating data graphics.
- `ggplot2` uses the layered grammar of graphics approach.



Anatomy of ggplot()





- Many very good online sources
- https://www.data-to-viz.com/

Research Institute for Farm Animal Biology
Wilhelm-Stahl-Allee 2
18196 Dummerstorf, Germany

www.fbn-dummerstorf.de