

Exploratory vs explanatory graphics & R graphics systems

Introduction to Data Visualization

Table of Contents

- [1. What's a snail?](#)
- [2. Infographics vs. data visualization](#)
- [3. Infographics](#)
- [4. Data visualization](#)
- [5. Exploratory vs explanatory data visualization](#)
- [6. Exploratory data visualization](#)
- [7. Explanatory data visualization](#)
- [8. Preparations to code along](#)
- [9. Extended example: US cereal data](#)
- [10. Exploratory plot array](#)
- [11. Explanatory scatterplot](#)
- [12. Base graphics plot functions](#)
- [13. Many many types of built-in diagrams](#)
- [14. grid graphics](#)
- [15. grid graphics example](#)
- [16. lattice graphics](#)
- [17. Grammar of graphics with ggplot2](#)
- [18. Concept summary](#)
- [19. Code summary](#)
- [20. Glossary](#)
- [21. References](#)



- What's a snail?
- Exploratory vs explanatory graphics
- Plot arrays and their use
- Base graphics plot functions
- R packages: `grid`, `lattice`, `ggplot`
- Which graphics package should you use?

Image: Matisse, The Snail/L'Escargot (1953). Gouache on paper, cut and pasted, mounted on canvas, 268.4 x 287 cm). Tate Gallery, London.

1 What's a snail?



- What is the purpose of "graphics". Give some examples!
- What does "The Snail" by Matisse achieve as graphics?
- See also: [The Swimming Pool](#), [Matisse and the Chapelle du Rosaire](#)
- Graphics: beauty/aesthetics, abstract representation of data
- Technique: color, line, medium, object, artist (subject)
- Origin: something written or drawn (also: basic unit of speech)
- John 8:6-8 "Jesus bent down and started writing on the ground with his finger." The Pharisees leave and Jesus is alone with the woman whom he saved from being stoned to death for adultery.
- Graphics examples: artistic, [architecture](#), [marketing](#), [product design](#)

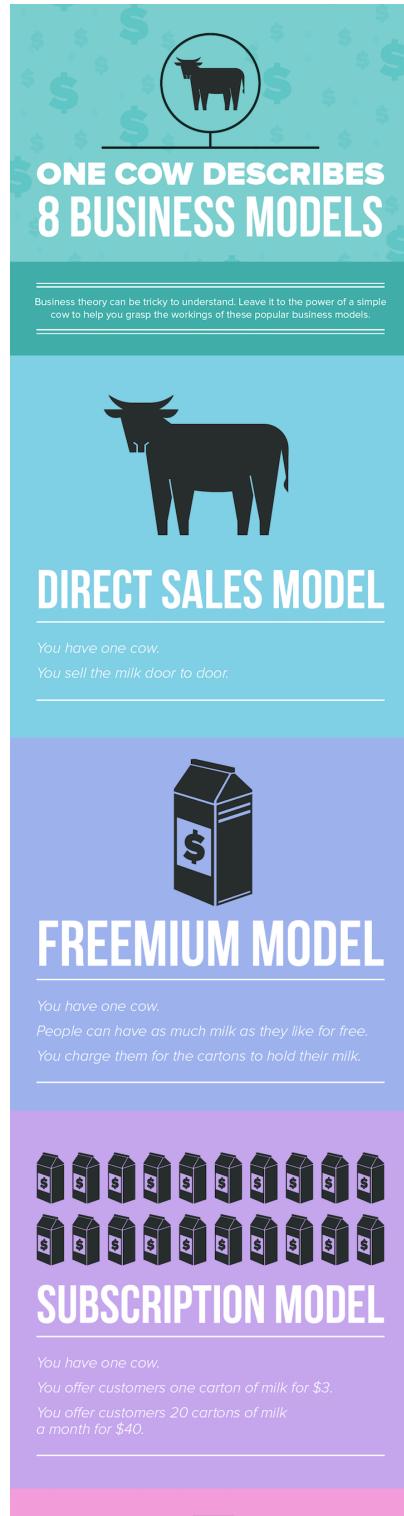
2 Infographics vs. data visualization

How would you describe the difference? Got an example?

3 Infographics

- Aesthetically rich (i.e. beautiful/nice to look at)
- Manually drawn (i.e. hard to create/alter/check)
- Represents specific data source

Example: 8-business-model Cow ([Leadem, 2017](#))





Sources

Investopedia. (2015). Where are some examples of different types of business models in major industries. [investopedia.com](#)

Inc. Loss Leader Pricing. [inc.com](#)

Colby, C. (2016). The On-Demand Economy Is Growing, and Not Just for the Young and Wealthy. [fors.org](#)

Zifflestat. Story. [zifflestat.co.uk](#)

Crowdsourcing Week. What is Crowdsourcing. [crowdsourcing.com](#)

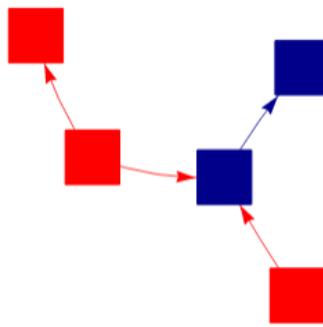
This image is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. More information: [creativecommons.org/licenses/by-nc-sa/4.0/](#)

© 2018 The Board of the University of Michigan. All rights reserved.

4 Data visualization

- Aesthetically barren (i.e. not decorated/beautiful)
- Algorithmically drawn (i.e. easy to create/alter/check)
- Rich in data details

Example: interactive network visualization with [visNetwork in R](#) (Junker, 2019)



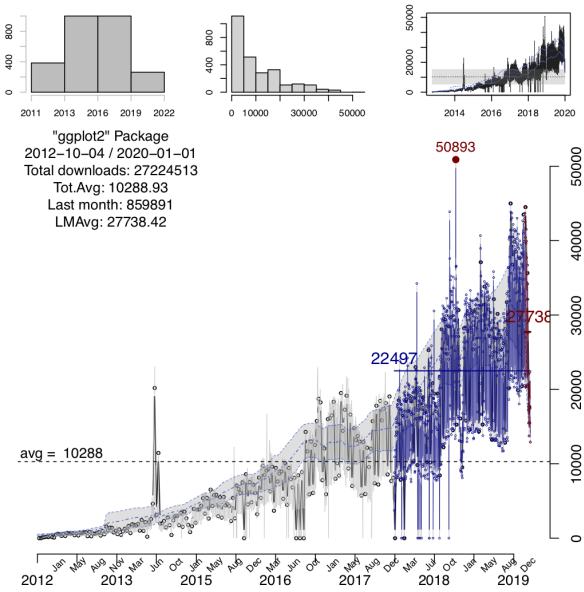
5 Exploratory vs explanatory data visualization

How would you describe the difference? Got an example?

6 Exploratory data visualization

- Helps us to understand what is in a **data set**
- Alternate name: Exploratory Data Analysis (EDA)
- Quickly **identify** features, curves, lines, trends, anomalies
- Best done at a high level of **granularity**
- Difficulty: separate **signals** from **noise**

Example: CRAN ggplot2 package downloads 2012-2019

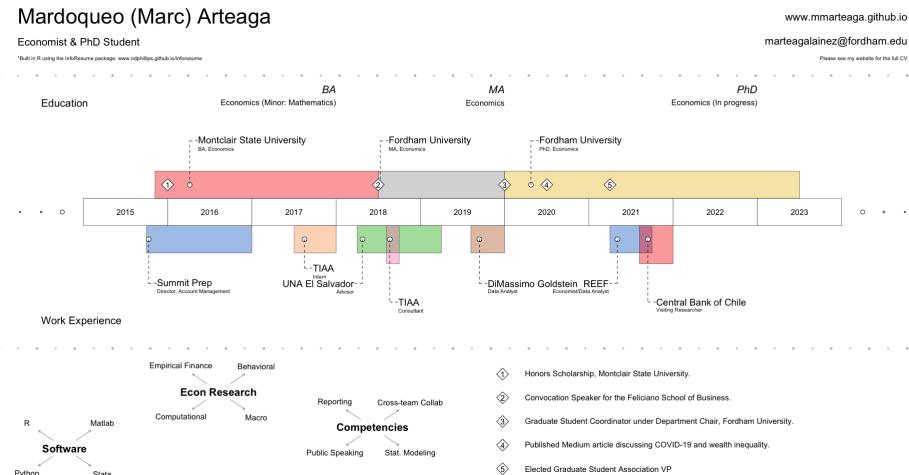


[Image source: Visualize.Cran.Downloads](#) (CRAN, 2021)

7 Explanatory data visualization

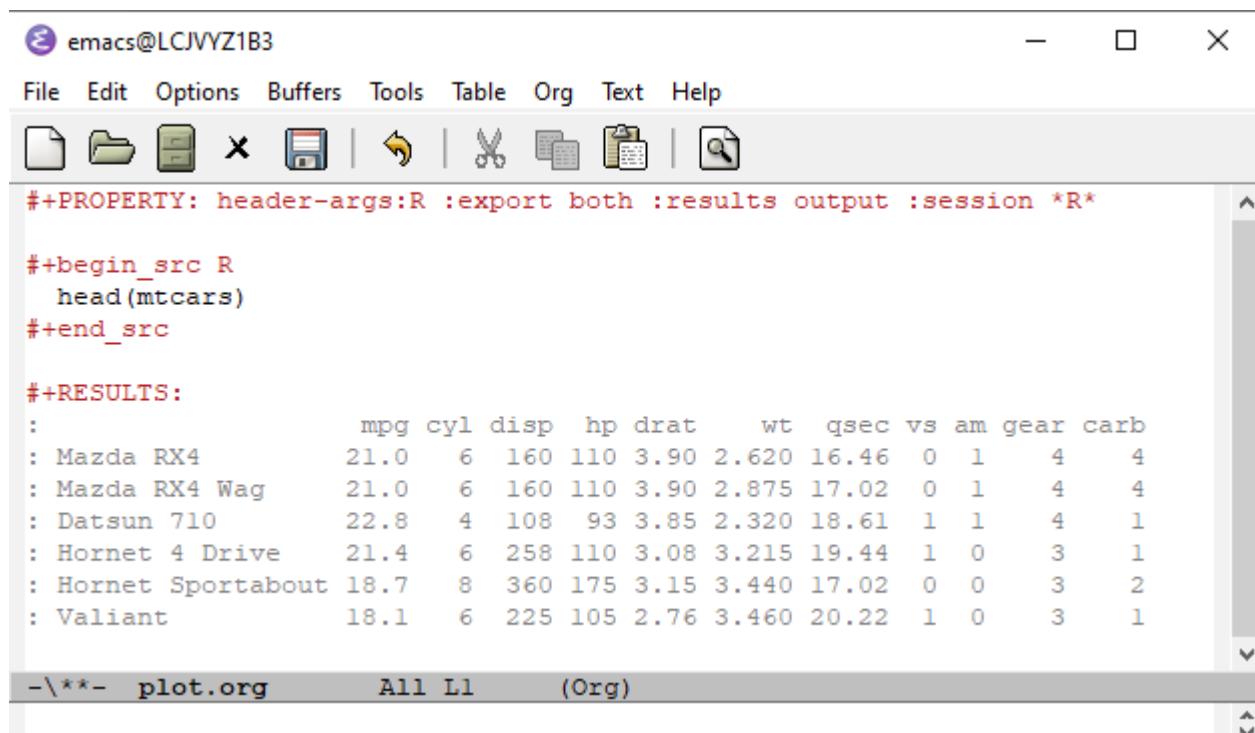
- Help us **convey** findings to others
- Alternate name: Data Storytelling
- Known to you at the outset (apart from **feedback**)
- **Design** to accommodate a particular **audience**
- Best done after **editorial decisions** what is relevant
- Difficulty: **selecting** focused data that support your **story**

Example: Creating a visual CV using R



Source: [How to create a visual CV using R!](#) (Arteaga, 2021)

8 Preparations to code along



```
#+PROPERTY: header-args:R :results output :session *R*

#+begin_src R
  head(mtcars)
#+end_src

#+RESULTS:
:          mpg cyl disp  hp drat    wt  qsec vs am gear carb
: Mazda RX4   21.0   6 160 110 3.90 2.620 16.46  0  1    4    4
: Mazda RX4 Wag 21.0   6 160 110 3.90 2.875 17.02  0  1    4    4
: Datsun 710   22.8   4 108  93 3.85 2.320 18.61  1  1    4    1
: Hornet 4 Drive 21.4   6 258 110 3.08 3.215 19.44  1  0    3    1
: Hornet Sportabout 18.7   8 360 175 3.15 3.440 17.02  0  0    3    2
: Valiant     18.1   6 225 105 2.76 3.460 20.22  1  0    3    1
```

-**- plot.org All L1 (Org)

- Open a new Org-mode file `plot.org` in Emacs
- Put this line at the top of the file `plot.org`:

```
#+PROPERTY: header-args:R :results output :session *R*
```

- Activate the code by putting your cursor on the line and entering `C-c C-c`. You should see the message `Local setup has been refreshed` in the minibuffer at the bottom of the editor.
- When you execute your first R code block, you'll be asked where you want the session named `*R*` to run: enter the path to `plot.org`
- For plots, use the header `:results graphics file :file plot.png`
- When you leave Emacs, you'll be warned that the session `*R*` is active: you can ignore this warning

9 Extended example: US cereal data



- Using: UScereal data frame from the MASS package
- 11 characteristics of 65 breakfast cereals available for sale
- Information mostly based on the package label required by US FDA

```
library(MASS) # load MASS package
data(UScereal) # load UScereal data frame
str(UScereal) # display data frame structure
```

```
'data.frame': 65 obs. of 11 variables:
 $ mfr      : Factor w/ 6 levels "G","K","N","P",...: 3 2 2 1 2 1 6 4 5 1 ...
 $ calories  : num  212 212 100 147 110 ...
 $ protein   : num  12.12 12.12 8 2.67 2 ...
 $ fat       : num  3.03 3.03 0 2.67 0 ...
 $ sodium    : num  394 788 280 240 125 ...
 $ fibre     : num  30.3 27.3 28 2 1 ...
 $ carbo     : num  15.2 21.2 16 14 11 ...
 $ sugars    : num  18.2 15.2 0 13.3 14 ...
 $ shelf     : int  3 3 3 1 2 3 1 3 2 1 ...
 $ potassium: num  848.5 969.7 660 93.3 30 ...
 $ vitamins  : Factor w/ 3 levels "100%","enriched",...: 2 2 2 2 2 2 2 2 2 2 ...
```

10 Exploratory plot array

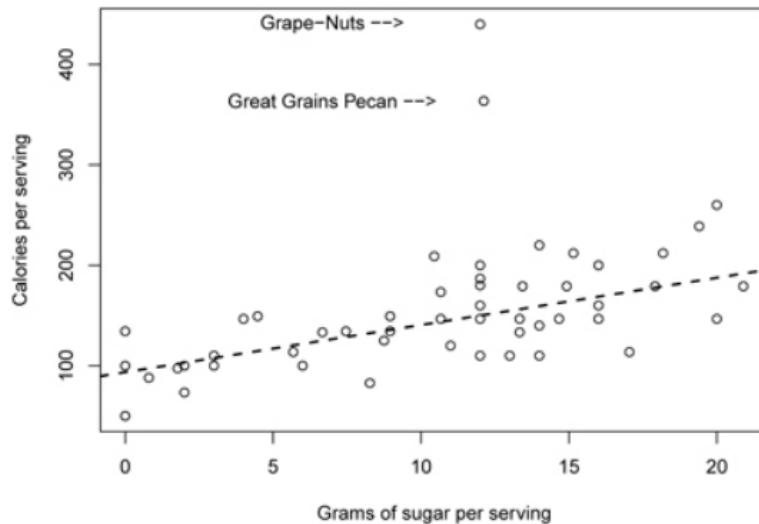
A useful, and common, exploratory plot is a panel of pairwise scatterplots to summarize the data frame.

```
plot(UScereal, las = 2)
```



- The `las` parameter sets the position of the tick labels - `las=2` means always perpendicular to the axis.
- Diagonal elements of the array list the name of the variable in the x-axis of all plots in that column, and the y-axis of all plots in that row.
- For 11 variables, 110 plots are shown. Some indicate strong relationships, e.g. `fat` and `calories`.
- Some variables, like `vitamins`, show only few values

11 Explanatory scatterplot



- Scatterplot of calories vs. sugar variables, augmented with a robust regression line ("robustness" refers to assumptions about the data)
- Dashed line highlights the trend our eye sees in the data if we ignore the two outlying points.
- The outliers correspond to cereals that have much higher calories than any of the others.
- The annotation of labels and text in the plot further aids the interpretation.

```
plot(data=UScereal, calories ~ sugars)
```



12 Base graphics plot functions

- Base graphics is the system originally built into the R language
- It's most common generic function is `plot`

- Base graphics are controlled by 72 graphics *parameters*
- Displays can be customized by *low-level* plotting functions
- Examples: abline, lines, points, text, legend etc.

FUNCTION	OBJECT TYPE	NATURE OF PLOT ¹
plot	Many	Depends on object type
barplot	Numeric	Bar plot
boxplot	Formula, numeric, list	Boxplot summary
hist	Numeric	Histogram
sunflowerplot	Numeric + Numeric	Sunflower plot
mosaicplot	Formula or table	Mosaic plot
symbols	Multiple numeric	Bubbleplots etc.

13 Many many types of built-in diagrams

- There are more than 40 types of useful diagrams in R

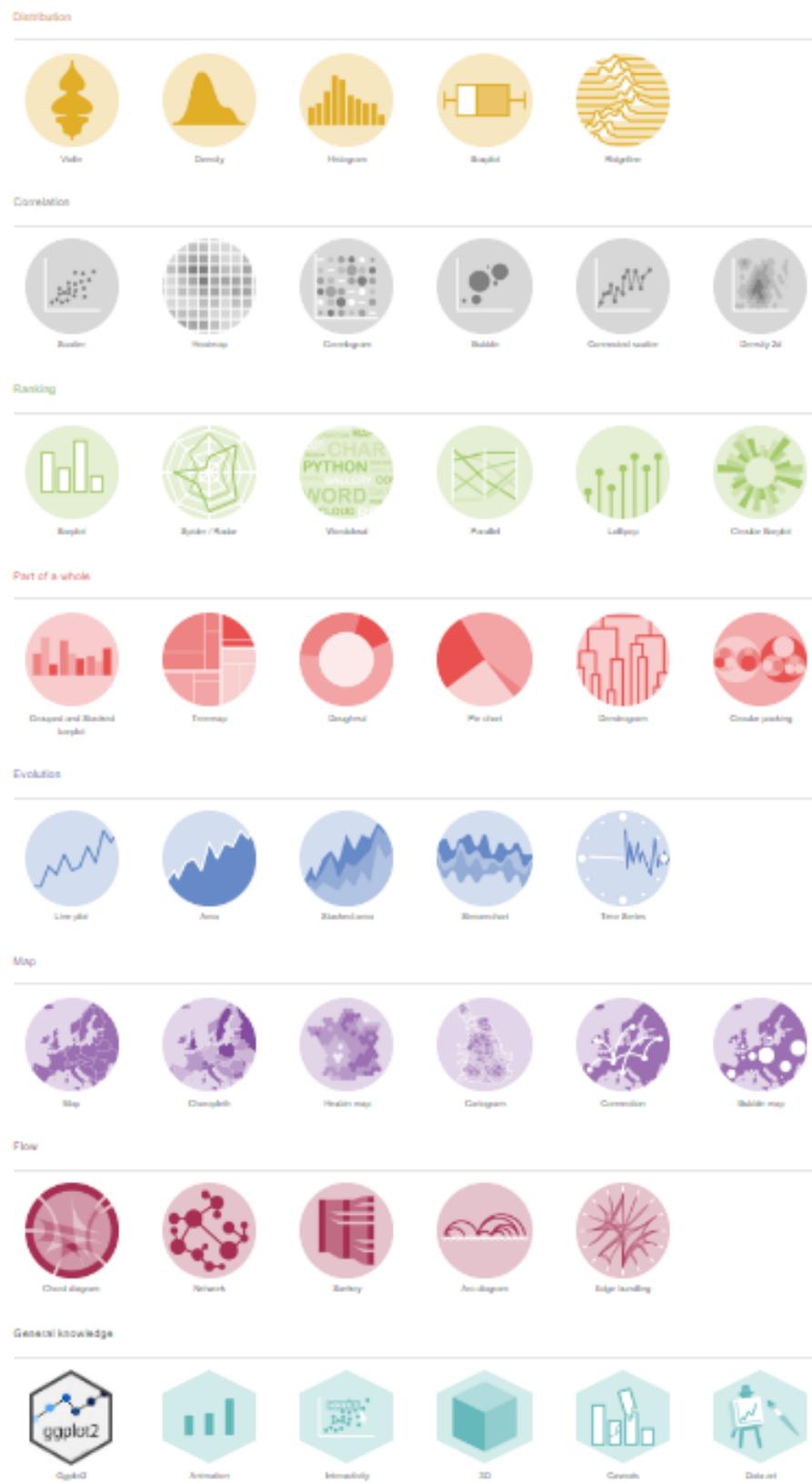
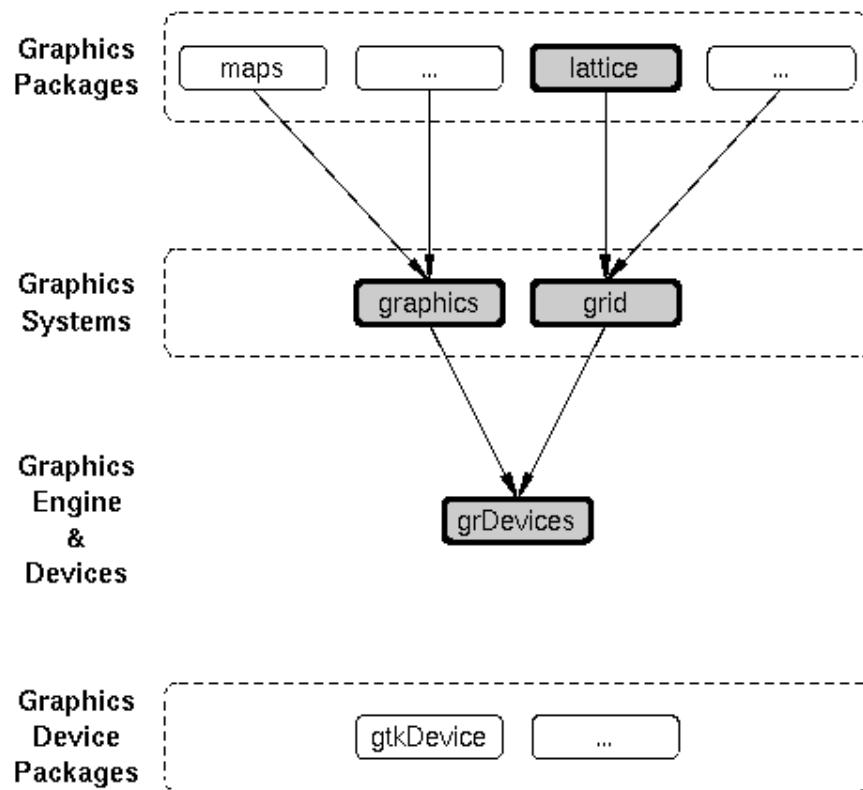


Figure 13: Source: R graph gallery

14 grid graphics

- The `grid` package uses the `grDevices` graphics engine
- Some packages use it, e.g. `vcd` (for graphing categorical variables)
- Nothing to do with the `grid` function of the base R package (which draws a grid over the plot, see `?grid`)
- More info: [Paul Murrell's documents](#)

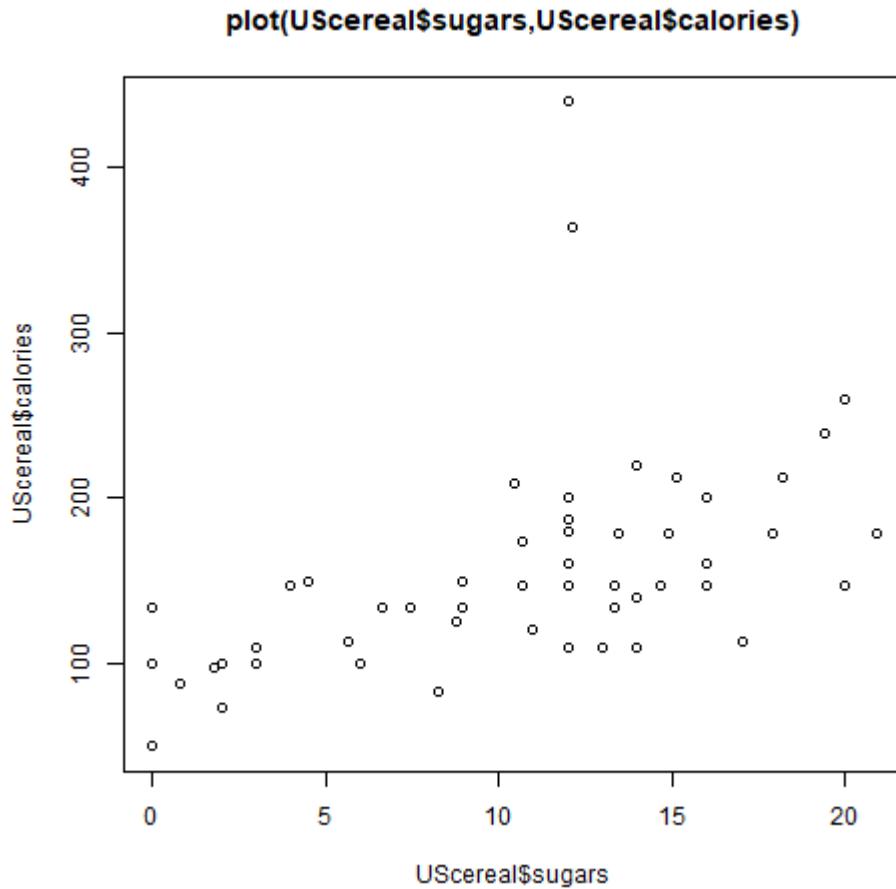


15 grid graphics example

- The following plot is generated with the base R (built-in) package. Below you find the code to create this plot using the `grid` package, demonstrating the greater flexibility but also steeper learning curve.
- Scatterplot with base R plot

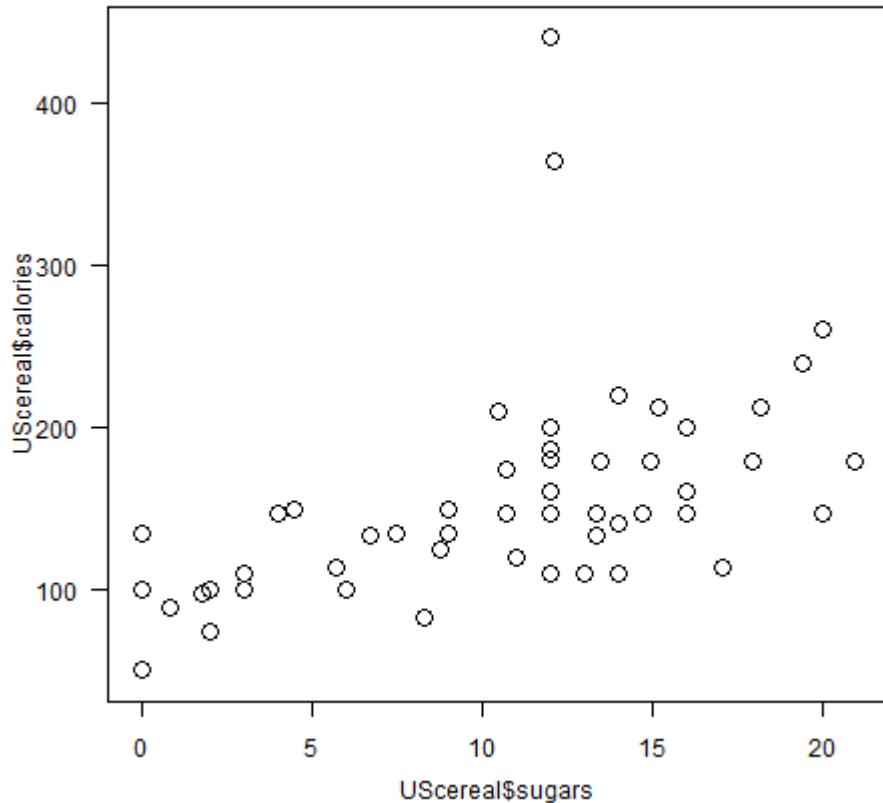
```

library(MASS)
plot(UScereal$sugars, UScereal$calories)
title("plot(UScereal$sugars, UScereal$calories)")
  
```



- Scatterplot with *grid* package:

```
library(MASS)
x <- UScereal$sugars
y <- UScereal$calories
library(grid)
pushViewport(plotViewport())
pushViewport(dataViewport(x,y))
grid.rect()
grid.xaxis()
grid.yaxis()
grid.points(x,y)
grid.text("UScereal$calories",x=unit(-3,"lines"),rot=90)
grid.text("UScereal$sugars",y=unit(-3,"lines"),rot=0)
popViewport(2)
```

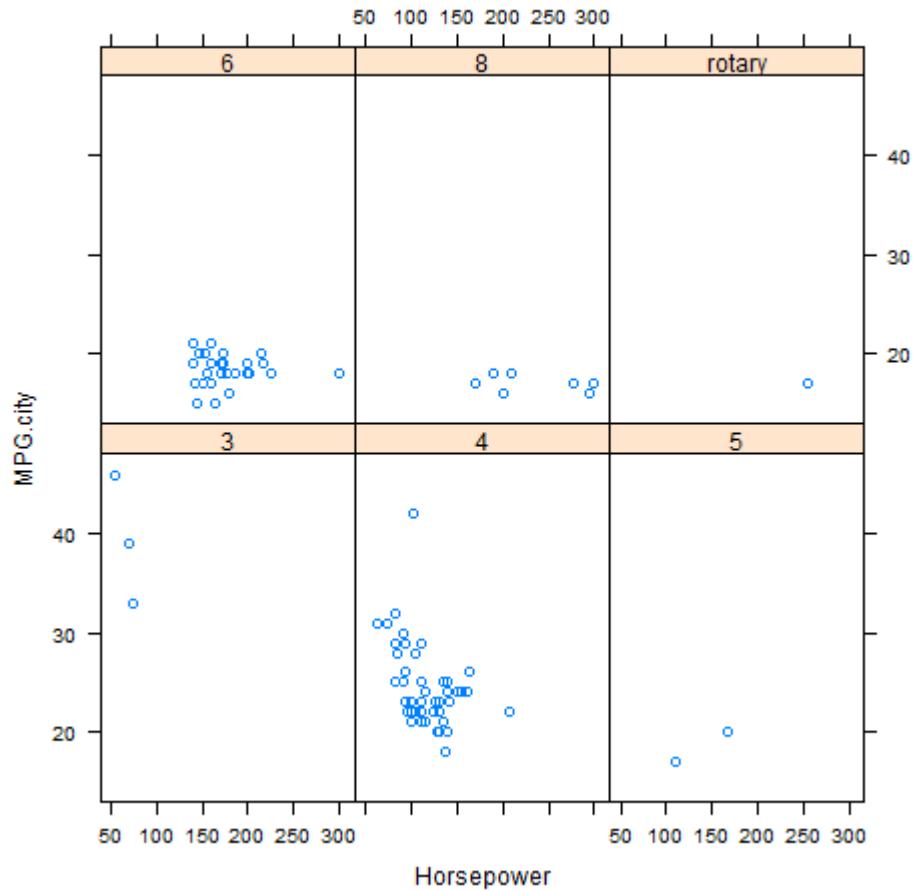


16 lattice graphics

- Based on grid graphics, shipped with base R (needs to be loaded)
- Alternative implementation to many standard plotting functions, including scatterplots, bar charts, boxplots, histograms, QQ-plots
- [lattice](#) has different default options for plot customization and some additional features, like the *multipanel conditioning plot*

```
library(MASS) # load MASS package for Cars93 data set
library(lattice) # load lattice package

## plot MPG.city vs. Horsepower, conditioned by Cylinders
xyplot(MPG.city ~ Horsepower | Cylinders, data = Cars93)
```



- You can group variables and get an automatic legend per group
- Price to be paid: simple annotations are harder to do than base R

17 Grammar of graphics with ggplot2

- Grammar of graphics construction based on human perception
- Better support for multipanel conditioning plots
- Highly extensible, complex, steep learning curve ([see here](#))

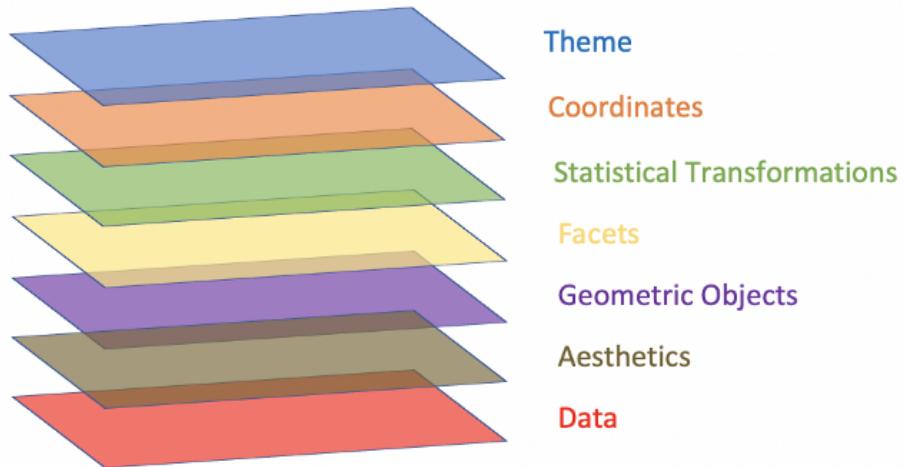


Figure 18: Grammar of Graphics (gg) philosophy

18 Concept summary

- Infographics are design-rich and built to inform, data visualizations (and dashboards) are data-rich and built to be flexible and alterable
- Exploratory/explanatory graphics have different challenges. EDA: separate signal from noise; storytelling: tell a good story!

19 Code summary

COMMAND	MEANING
method	Available methods for generic functions

20 Glossary

TERM	MEANING
Infographics	
Data visualization	
Exploratory graphics	
Explanatory graphics	

21 References

- Arteaga M (20 January, 2021). How to create a visual CV using R!. [Online: mmarteaga.github.io.](https://mmarteaga.github.io/)
- CRAN (27 April 2021). Visualize downloads from CRAN Packages. [Online: cran.r-project.org.](https://cran.r-project.org/)

Footnotes:

¹ Plot types not seen in this lecture yet: sunflower plots (scatterplots that reduce overplotting by turning multiple points into petals); mosaic plots (mosaic of rectangles whose height represents the proportional value); bubbleplots (scatterplot with a third dimension represented with the size of the dots).

Author: Marcus Birkenkrahe

Created: 2022-09-22 Thu 18:10