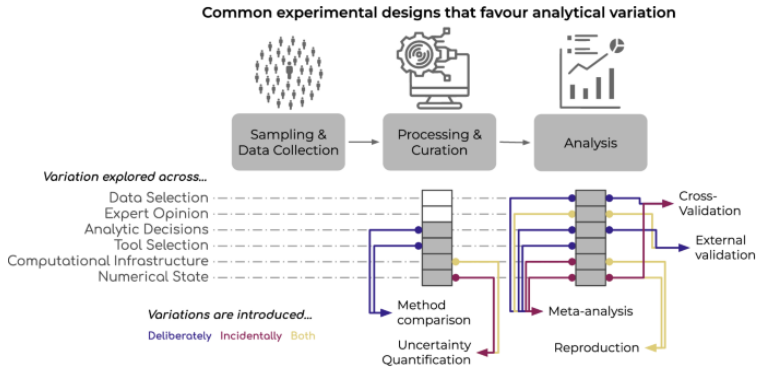
Processing and interpretation of neuroscience data:

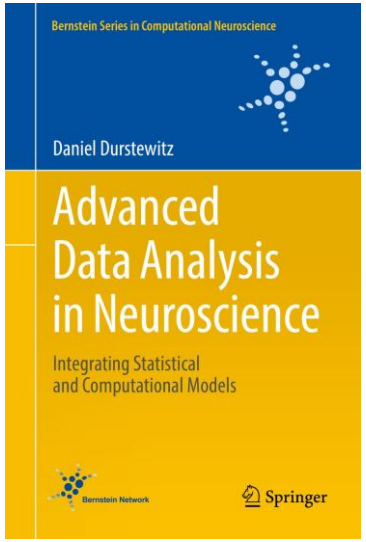
Module 1 – Data overview and single cell sequencing



Dr. Kaja Moczulska & Dr. Łukasz Piszczek

Univ.-Prof. Dr. Wulf Haubensak
Department of Neuronal Cell Biology

Center For Brain Research (Zentrum für Hirnforschung)



Module 2 – Data overview and single cell sequencing (18.11-09.12)
Prof. Vered Kellner and Dr. Hugo Malagon-Vina

Overview

Resources needed:

- Laptop
- Internet access
- Materials:
  - VM machine
  - https://github.com/HugoMalagon/NeuroData_860.053-MUW

21.10
- Introduction to R, basics
- Visual analytics

28.10
- dimensional reduction: PCA, UMAP
- Normalization/scaling
- clustering: k-means, knn

4.11
- Intro to Seurat
- Single cell RNA seq
- Dataset merging and preprocessing

11.11
- clustering in Seurat
- DEG interpretation

„Exam“: Assessment via email after each class

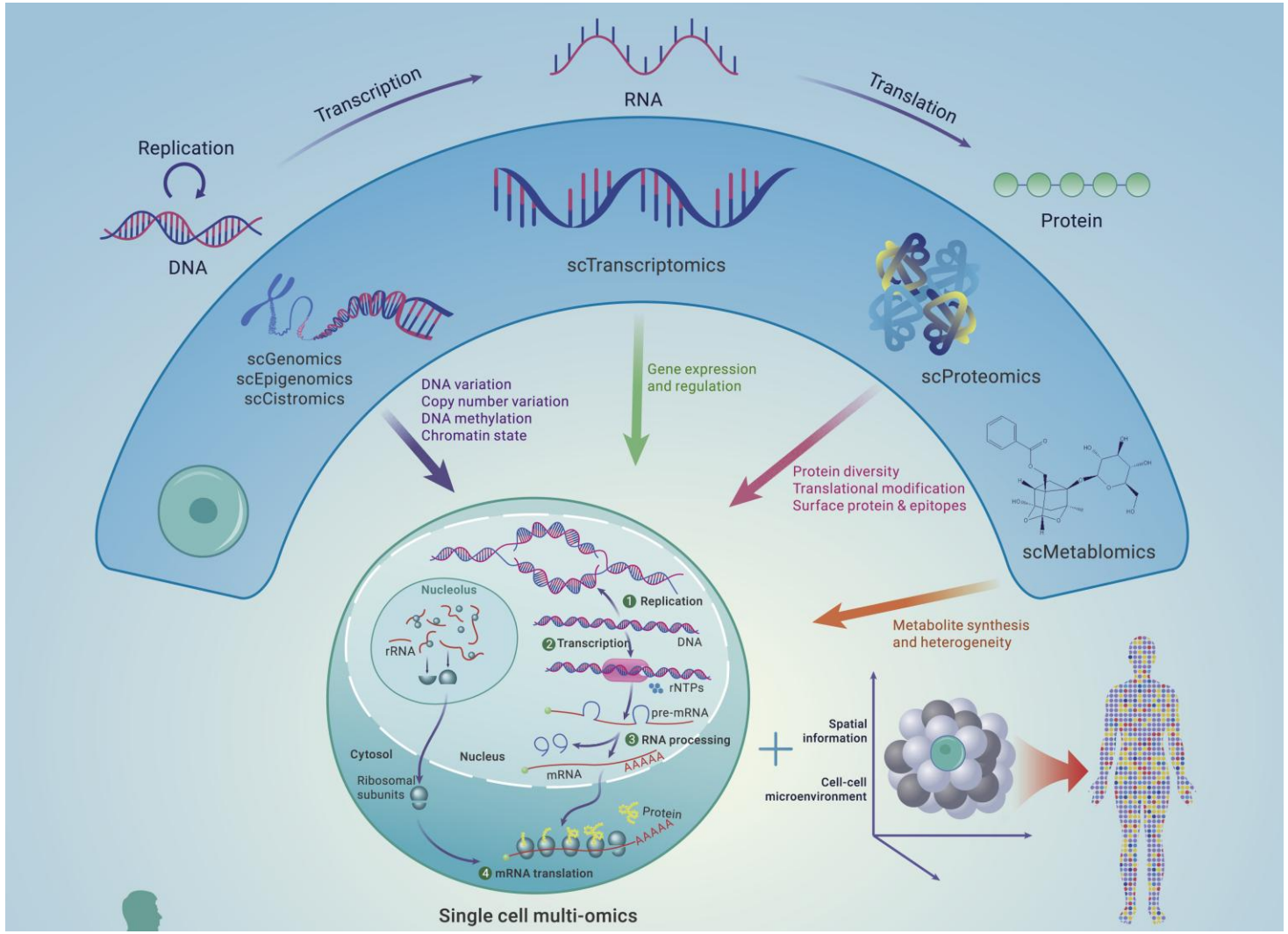
Day 1

Introduction to R

Basics

Why is single cell analysis interesting neuroscience?

- Neurogenomics and neurotranscriptomics are rapidly developing fields
- First sc transcriptome was published in 2009
- Single cell and spatial analysis techniques are constantly being developed and widely used in areas of biology
- Push to integrate multimodal data
- Valuable skillset: basics of data analysis tools and techniques are shared between different type of data
- Lot of data available: with the skillsets and ideas one can get new information out of already published available datasets



Wen et al, Innovation, 2022

Introduction into R

Visualization

Open source

Data science

Platform agnostic

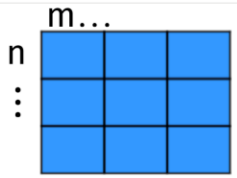
Computational statistics

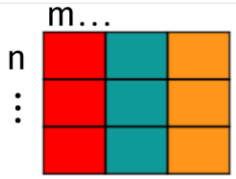
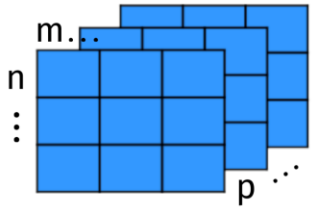
R:

- is a GNU Project and licensed under the GNU General Public License.
- It is written primarily in C, Fortran, and R itself.
- Precompiled executables are provided for various operating systems.
- As an interpreted language, R has a native command line interface.

Introduction into R: Data structures

	Dimensions	Mode (data "type")	Example
Vector		Identical	`c(10,0.2,34,48,53)`

Matrix		Identical	`matrix(c(1,2,3,`
			`11,12,13),`
			`nrow = 2,`
			`ncol = 3)`

Data frame		Can be different	`data.frame(x = 1:3,` `y = 5:7)`
Array		Identical	`array(data = 1:3,` `dim = c(2,4,2))`
List	Vector Matrix Data frame Array	Can be different	`list(x = cars[,1],` `y = cars[,2])`

See Rstudio IDE cheatsheat

https://r.qcbs.ca/workshop01/pres-en/workshop01-pres-en.html#1

Introduction into R: basic code structure

rescale-function.R

Source on Save | Run | Source

```
# create rescale function ----
rescale01 <- function(x) {
  if(!is.numeric(x)){
    stop("Value isn't numeric")
  }

  rng <- range(x, na.rm = TRUE, finite = TRUE)
  (x - rng[1]) / (rng[2] - rng[1])
}
```

Code Section

Function body

Conditional block

1:1 | # create rescale function | R Script

Conventional assignment types

Practical example

```
> mean(x=1:10)
[1] 5.5
> x
Error: object 'x' not found

> mean(x<-1:10)
[1] 5.5
> x
 [1]  1  2  3  4  5  6  7  8  9 10
```

https://docs.posit.co/ide/user/ide/guide/code/code-sections.html

Introduction into R

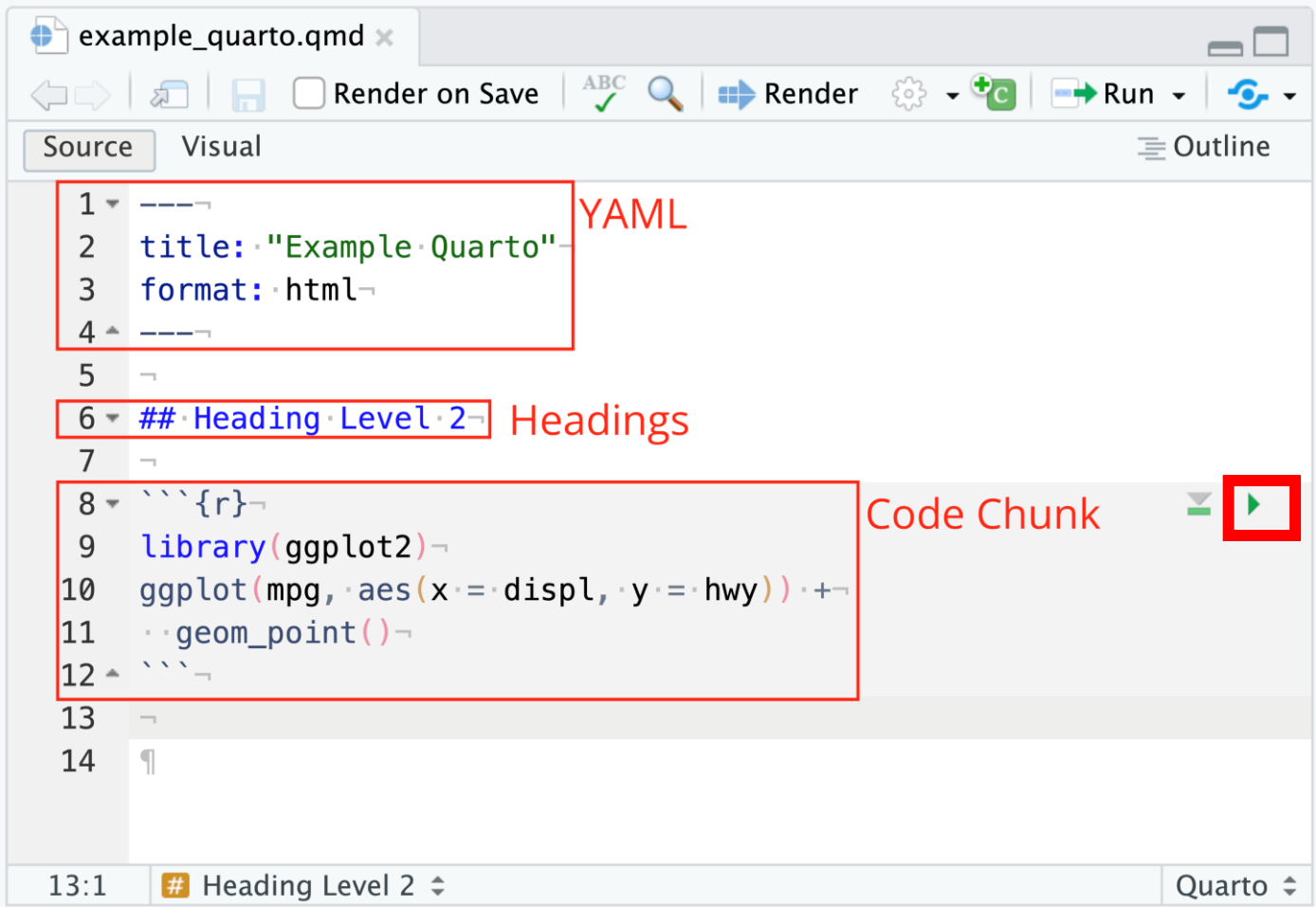
Operator	Primary Use	Scope/Behavior	Community Convention
<-	Assignment	Assigns in current environment	Preferred for assignment
=	Assignment, Arguments	Assigns or sets function argument	Preferred for arguments only

S and APL had its own keyboards

S language didn't have == for equality testing, so that was left to the single equal sign

Introduction into R: basic code structure

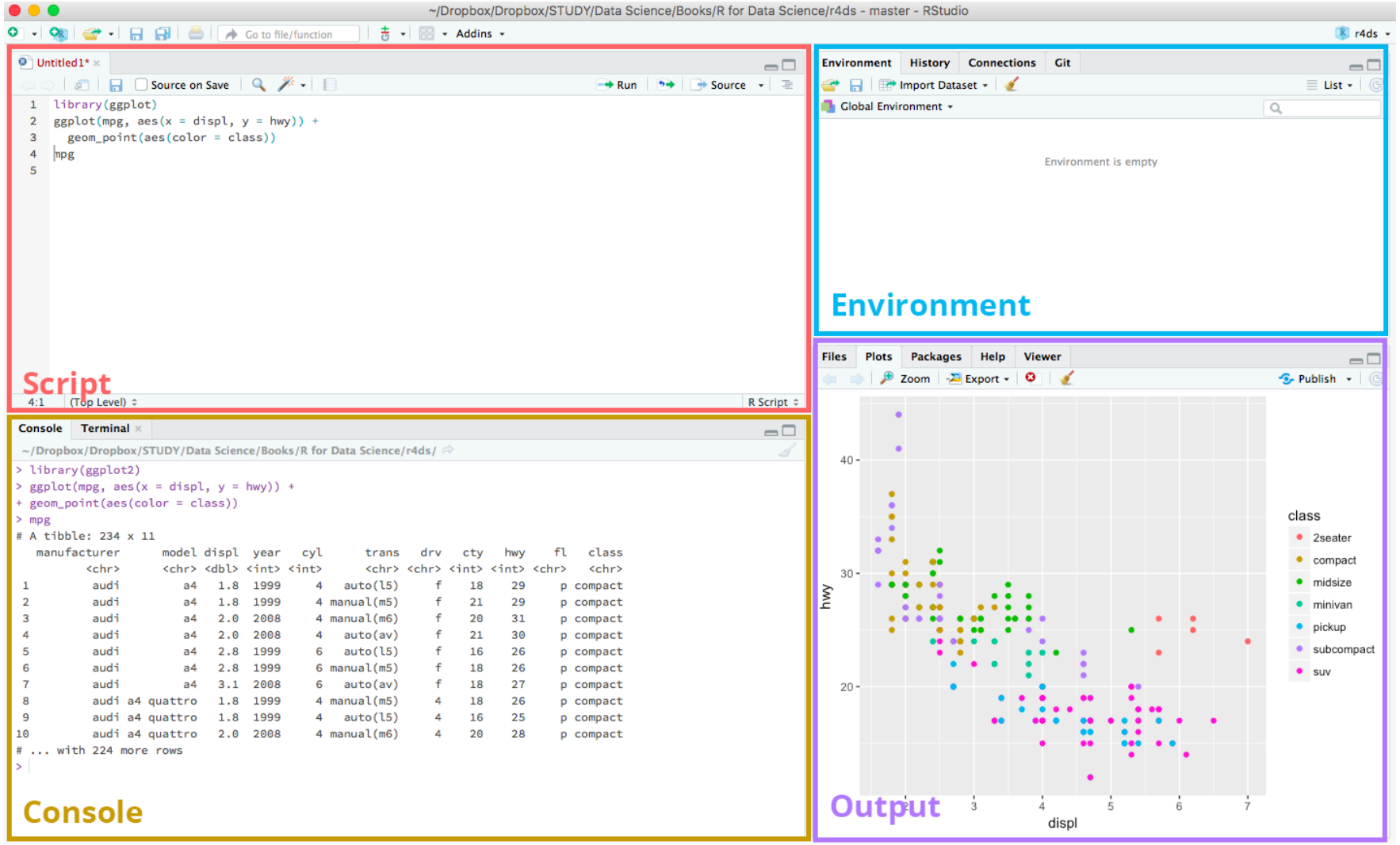
YAML is a human-readable and easy to write language to define data structures. Rmd makes it possible to use a YAML header to specify certain parameters right at the beginning of the document. Built-in YAML parameters make it easier to create more organized and informative reports.



Run Code Chunk

https://docs.posit.co/ide/user/ide/guide/code/code-sections.html

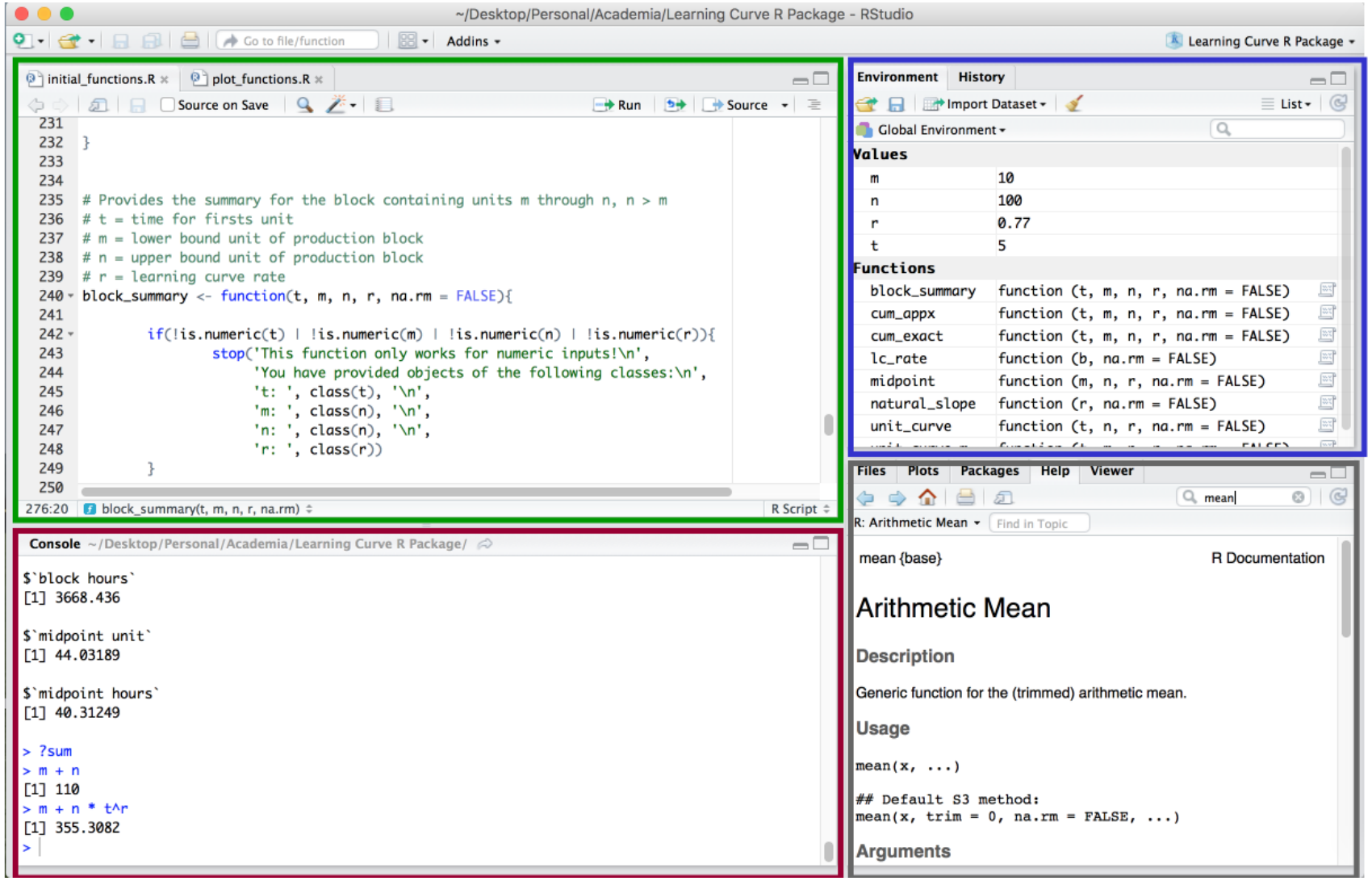
Introduction into R: RStudio layout



Introduction into R: RStudio layout

Script files
- Saves your script
- Allows code & comments
- Can have multiple files open at a time

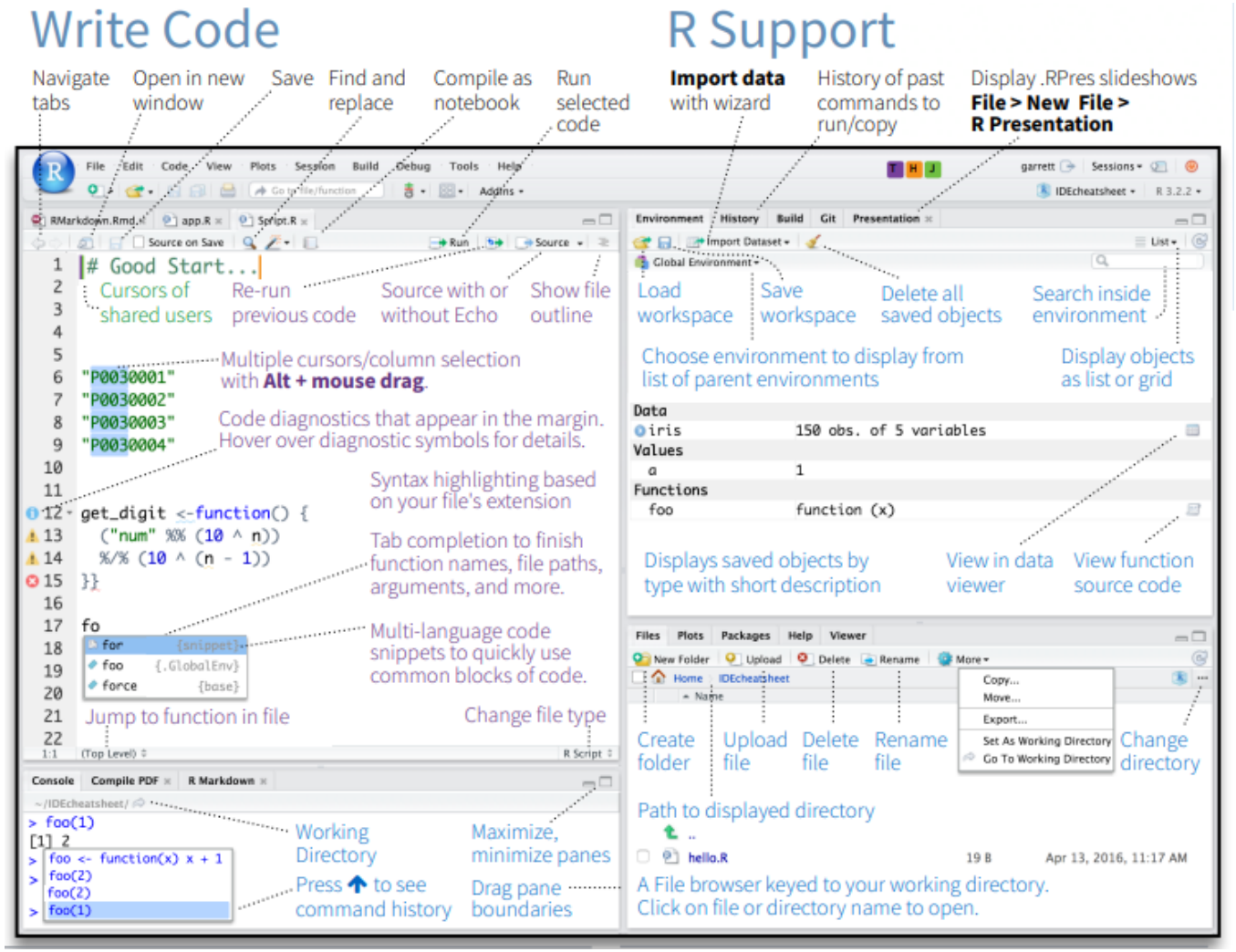
Console/Command line
- Can use as calculator
- Does not save code
- This is where your output is displayed



Workspace environment
- Holds your objects
- Can review history

Misc - Displays:
- files in working directory
- plots when produced
- help files/search

Introduction into R: RStudio layout



See Rstudio IDE cheatsheat

Introduction into R: help

Not sure what a function is doing or how to use it?

Use ?function_name or help(function_name)

```
> ?mean
>
```

Files | Plots | Packages | Help | Viewer | Presentation

R: Arithmetic Mean Find in Topic

mean {base} R Documentation

Arithmetic Mean

Description

Generic function for the (trimmed) arithmetic mean.

Usage

```
mean(x, ...)

## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments

`x`	an **R** object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects. Complex vectors are allowed for `trim = 0`, only.
`trim`	the fraction (0 to 0.5) of observations to be trimmed from each end of `x` before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
`na.rm`	a logical evaluating to `TRUE` or `FALSE` indicating whether `NA` values should be stripped before the computation proceeds.
`...`	further arguments passed to or from other methods.

Value

If `trim` is zero (the default), the arithmetic mean of the values in `x` is computed, as a numeric or complex vector of length one. If `x` is not logical (coerced to numeric), numeric (including integer) or complex, `NA_real_` is returned, with a warning.

If `trim` is non-zero, a symmetrically trimmed mean is computed with a fraction of `trim` observations deleted from each end before the mean is computed.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

See Also

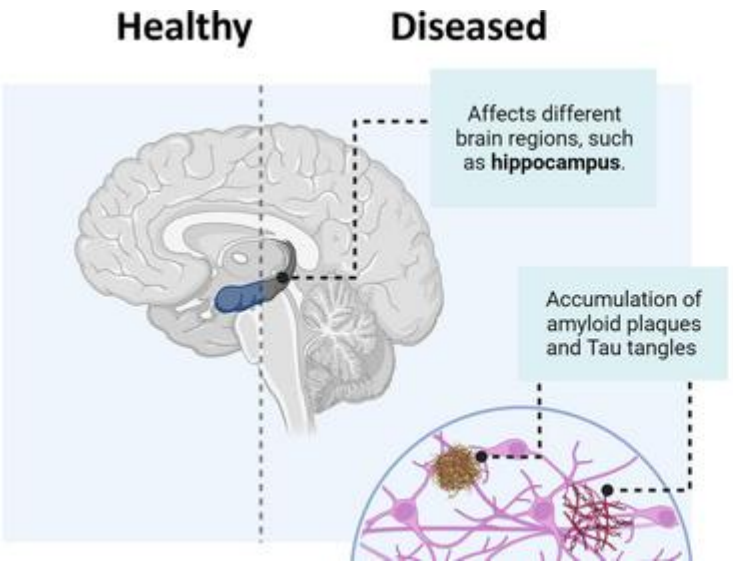
`weighted.mean`, `mean.POSIXct`, `colMeans` for row and column means.

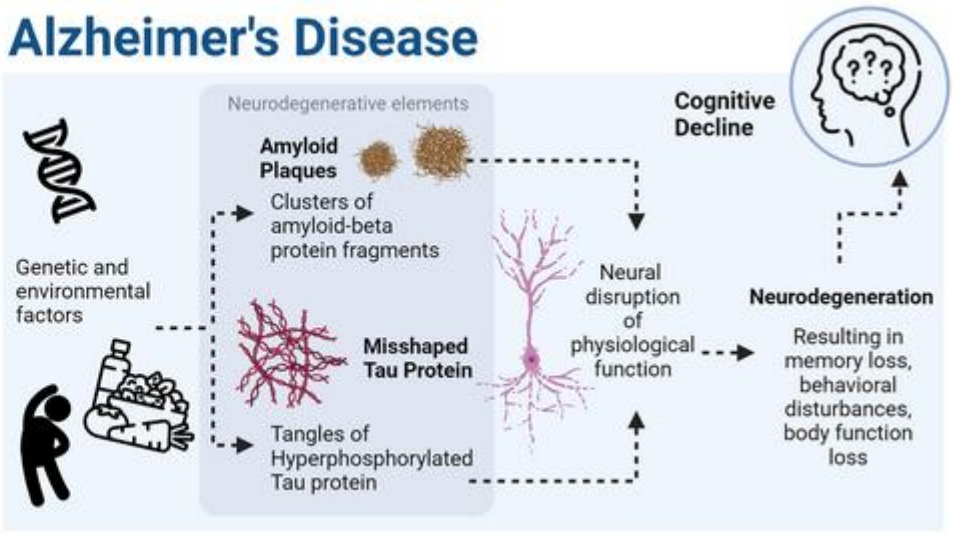
Examples

Run examples

See Rstudio IDE cheatsheat

Alzheimer's disease datasets for this module





Balestri et al., J Neuoinfl 2024

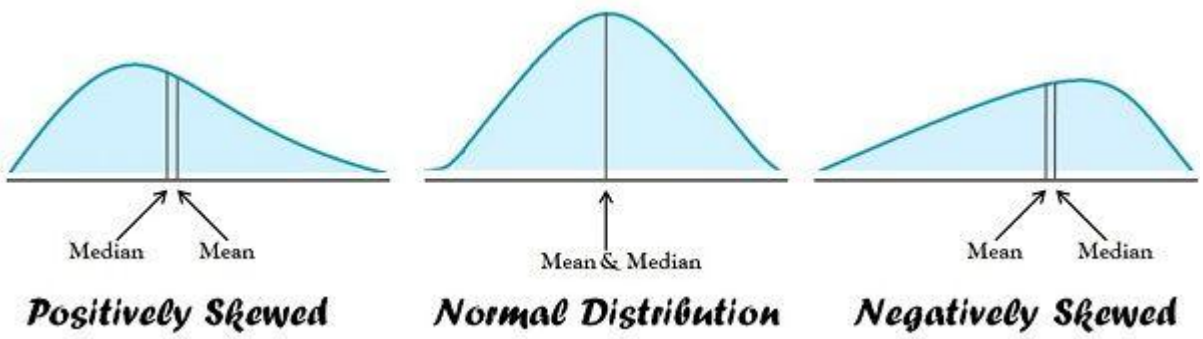
- Meeting 1 and 2: human patient data
  - Microarray
  - Selected genes expression, 3 brain regions, 2 sexes, control and AD patients
  - https://europepmc.org/article/MED/23595620
- Meeting 3 and 4: preclinical mouse model (ApoE KO) data
  - Single cell RNA seq
  - All detected genes, hippocampus, control and AD mice
  - https://www.sciencedirect.com/science/article/pii/S0168010221002182?via%3Dihub

Look at the data

- Get some info with what you work on. Useful 1st glimpse
  - dim(df) -> dimensions of data frame
  - nrows() -> number of rows
  - ncols() -> nuber of columns
  - head(df, 20) -> first 20 entries
  - tail(df, 20) -> last 20 entries
  - …

- Libraries
  - dlookr
  - skimr
  - …

Basic statistics

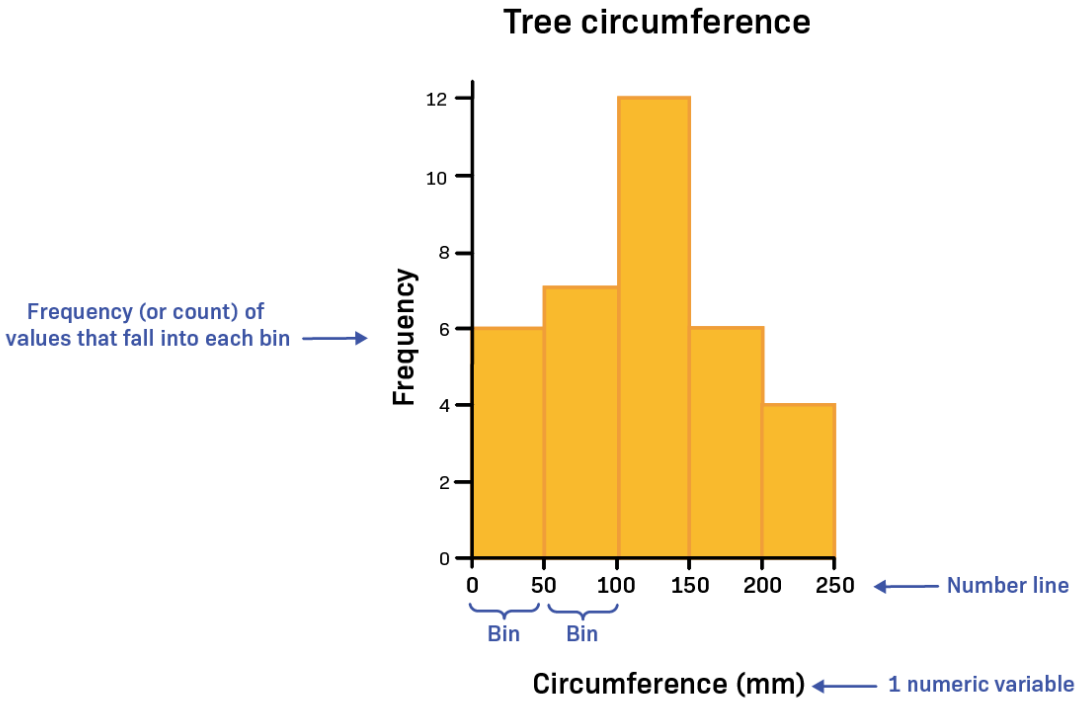


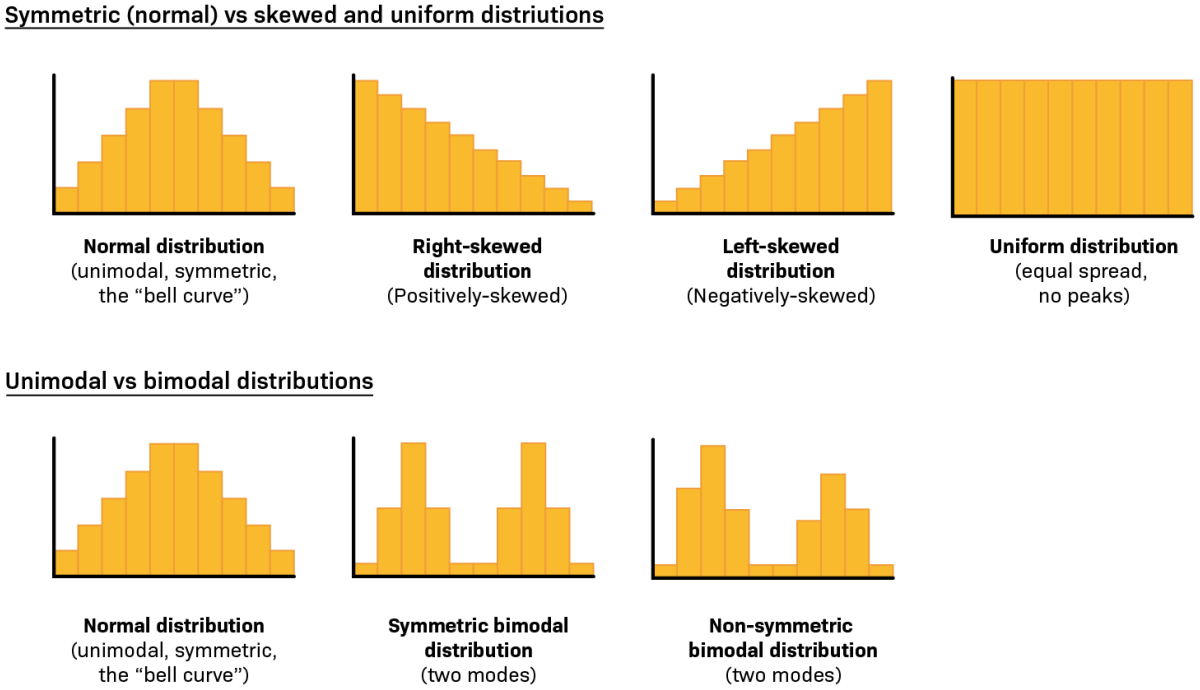
Normal distribution: test means
t-test
Skewed distribution: test medians
Wicoxon rank sum test

Useful functions in R:
mean(x) – calculates arithmetic mean
median(x) – calculates median
ttest(x, y) – calculates t-test (of the means)
wilcox.test() – calculates Wilcoxin test (of the medians)

How to check if your distribution is normal

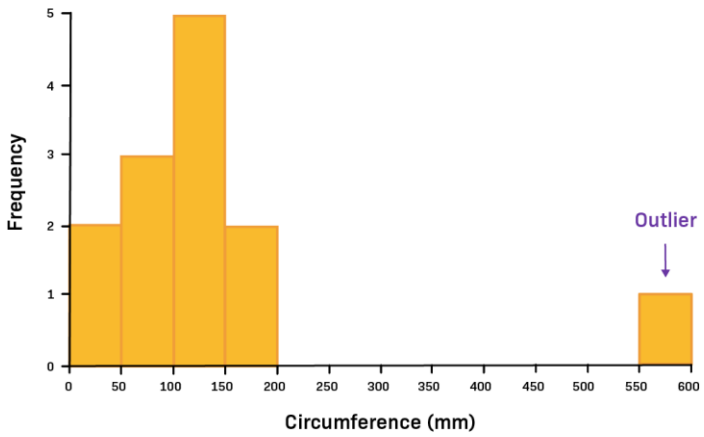
- Statistics (ex. shapiro.test())
- Visual (histogram)

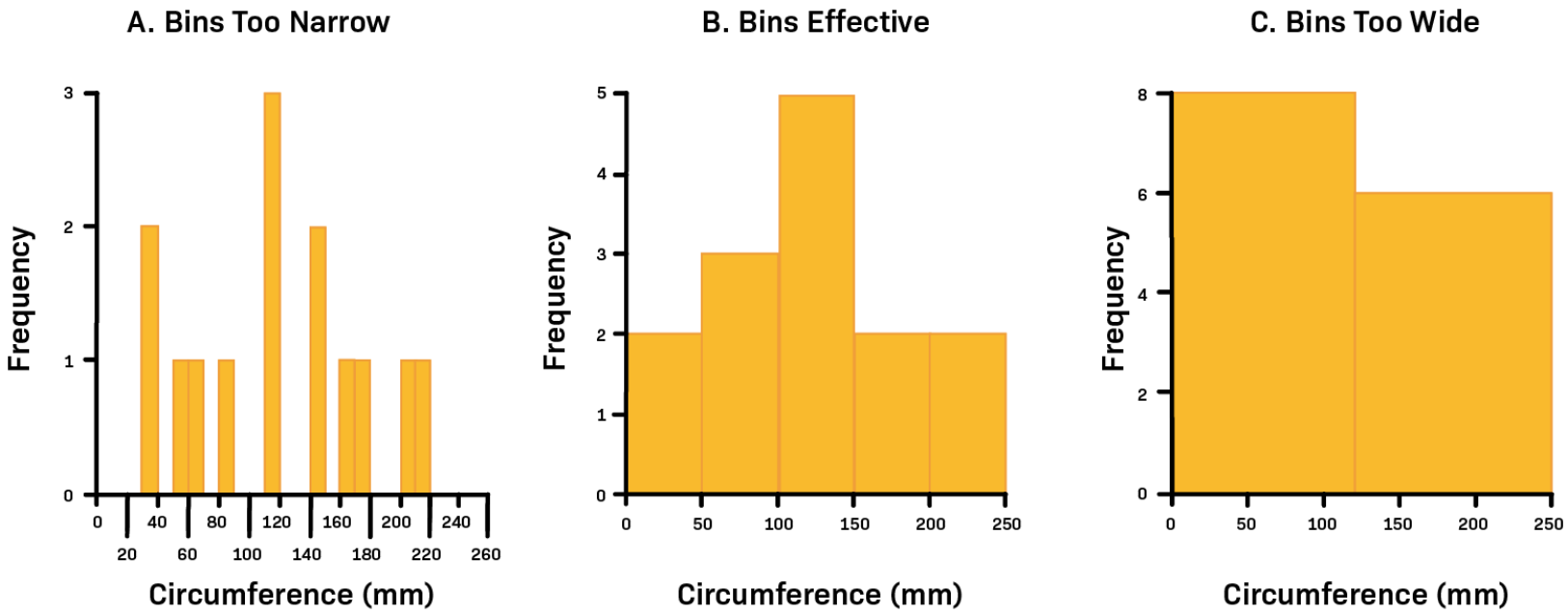




https://www.labxchange.org/library/items/lb:LabXchange:10d3270e:html:1

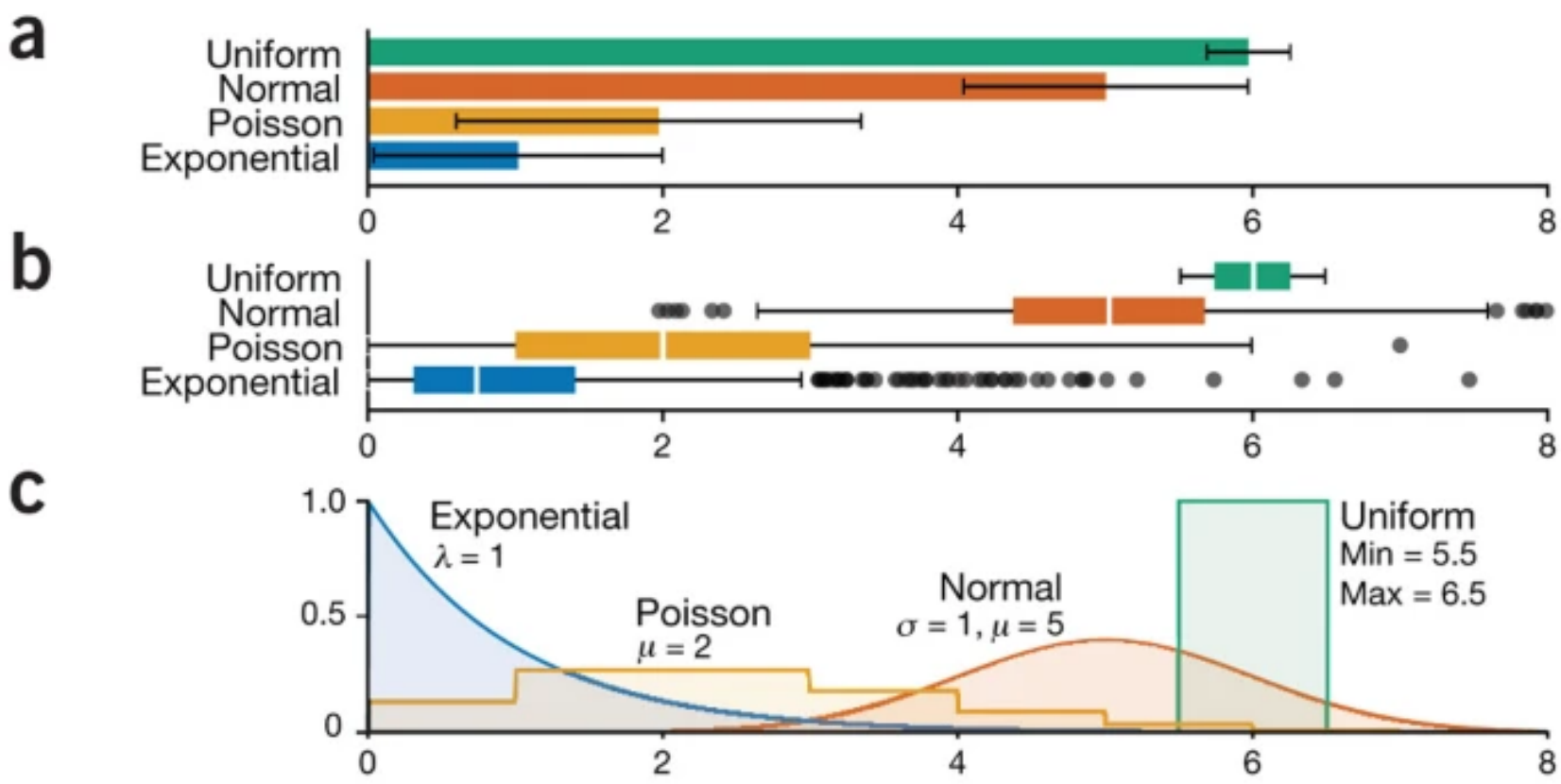
Histograms





https://www.labxchange.org/library/items/lb:LabXchange:10d3270e:html:1

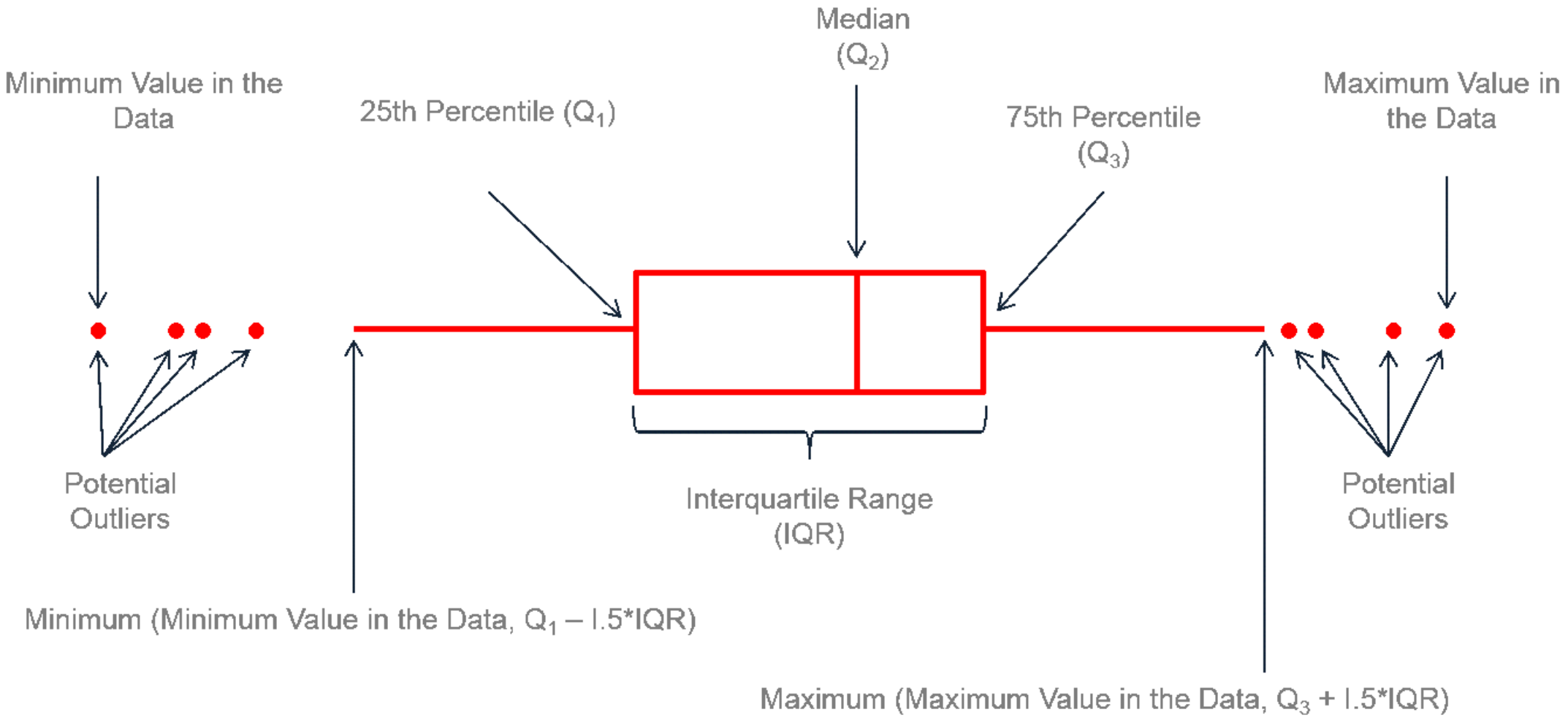
Bar plot vs box plot



(a) Bar chart showing sample means (n = 1,000) with standard-deviation error bars.
(b) Box plot (n = 1,000) with whiskers extending to ±1.5 × IQR.
(c) Probability density functions of the distributions in **a** and **b**. λ, rate; μ, mean; σ, standard deviation.

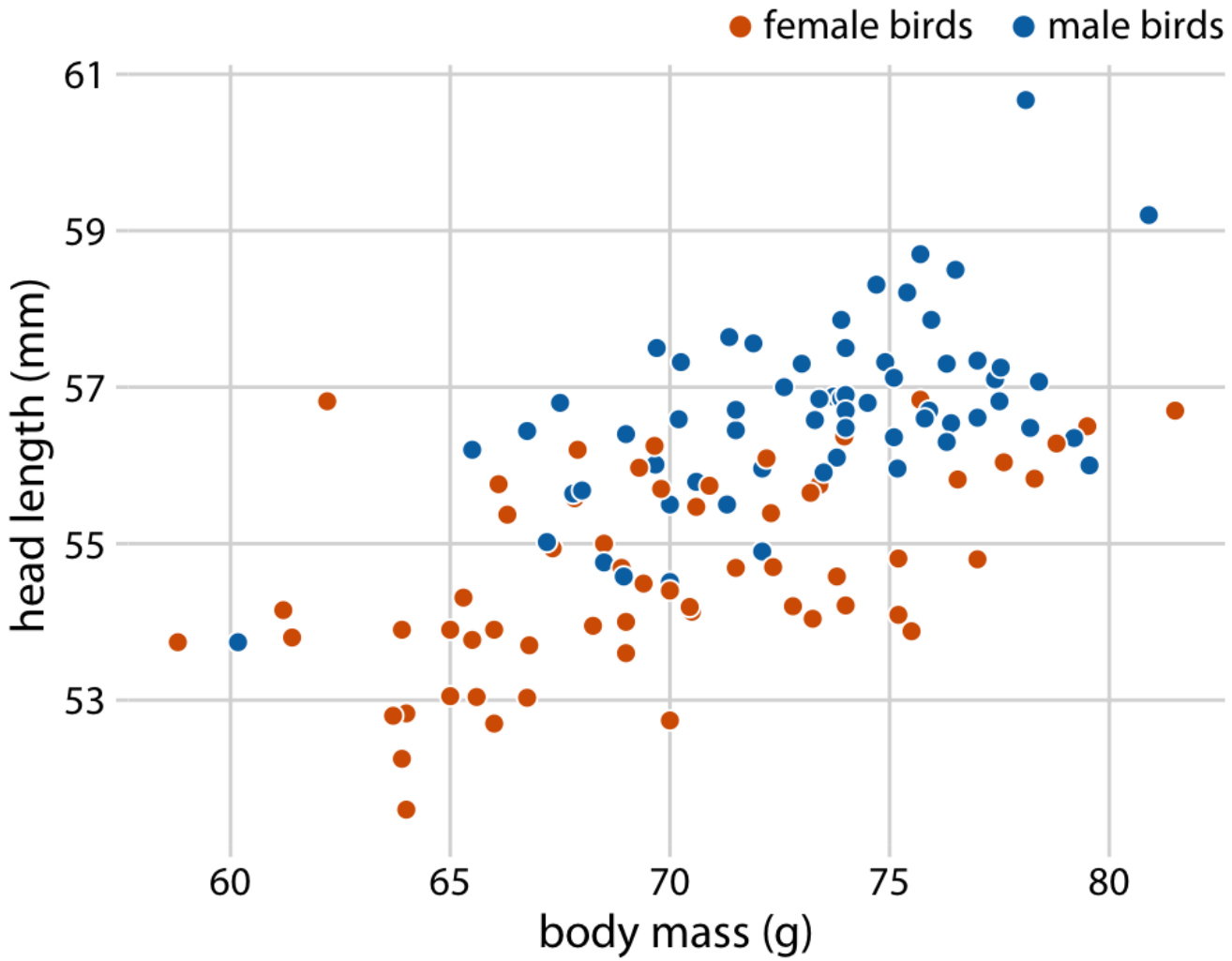
https://www.nature.com/articles/nmeth.2807

Box plot anatomy



https://www.data-to-viz.com/caveat/boxplot.html

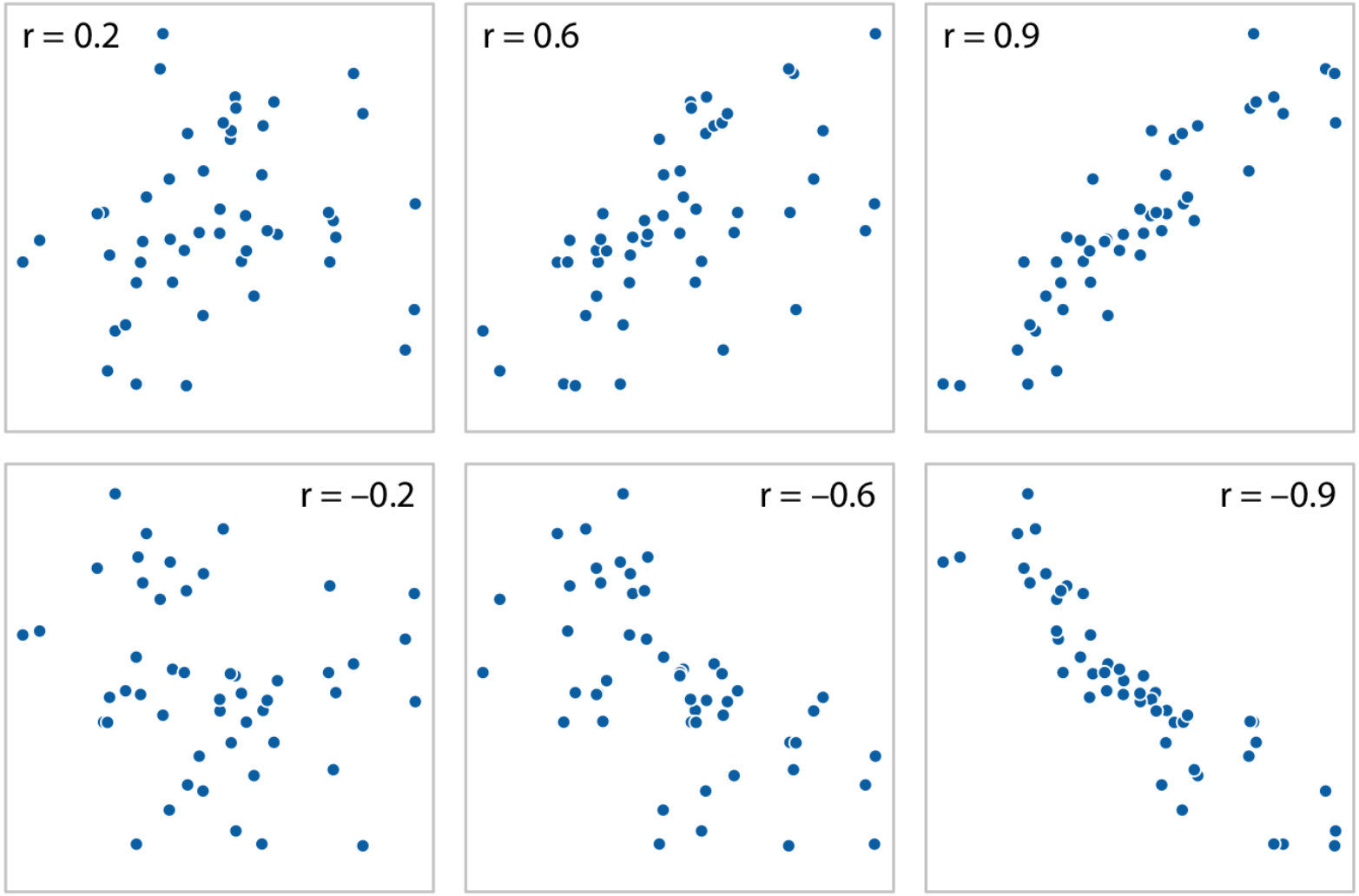
Scatter plot, correlation



Blue jays

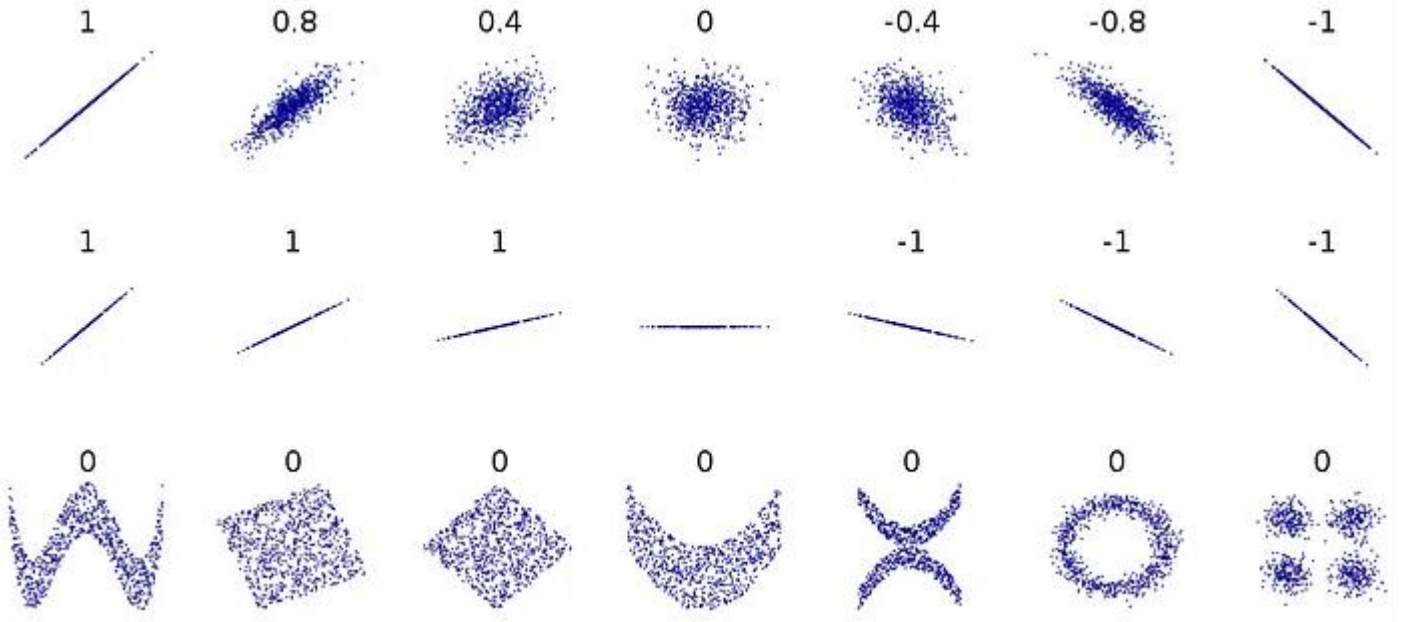
https://clauswilke.com/dataviz/visualizing-associations.html

Correlation coefficient (r)



https://clauswilke.com/dataviz/visualizing-associations.html

Why look at the data when we have correlation coefficient?



https://medium.com/@becaye-balde/visualizing-correlations-scatter-matrix-and-heat-map-d597436b7d23

Thank you for your attention

Now wake up and write some code!

For people working from home, go through the R-code lesson 1 intro, please.

There is a small assessment in the end of the R markdown file, where we ask you to send answers by email ☺