```
> workshop(
    topic   = “R Introduction for Data Science”,
    trainer = “Muhammad Aswan Syahputra”,
    when    = “2019-04-13”,
    where   = “Telkom University, Bandung”
  )
>  ...
```

- Sensory Scientist @ Sensolution.ID
- Using R for 4+ years, keen on Data Carpentry
- Initiator of Komunitas R Indonesia
- Pkgs: sensehubr, nusandata, bandungjuara, prakiraan, etc
- Shinyapps: sensehub, thermostats, aquastats, bcrp, bandungjuara, etc

aswansyahputra

# Know your neighbour!

- Who are you?
- What you do with data?
- How would you describe your experience with R?

# Our goal

Artwork by @allison_horst

HELLO

My name is

# ANDIKA

HELLO

My name is

# ALISYA

HELLO

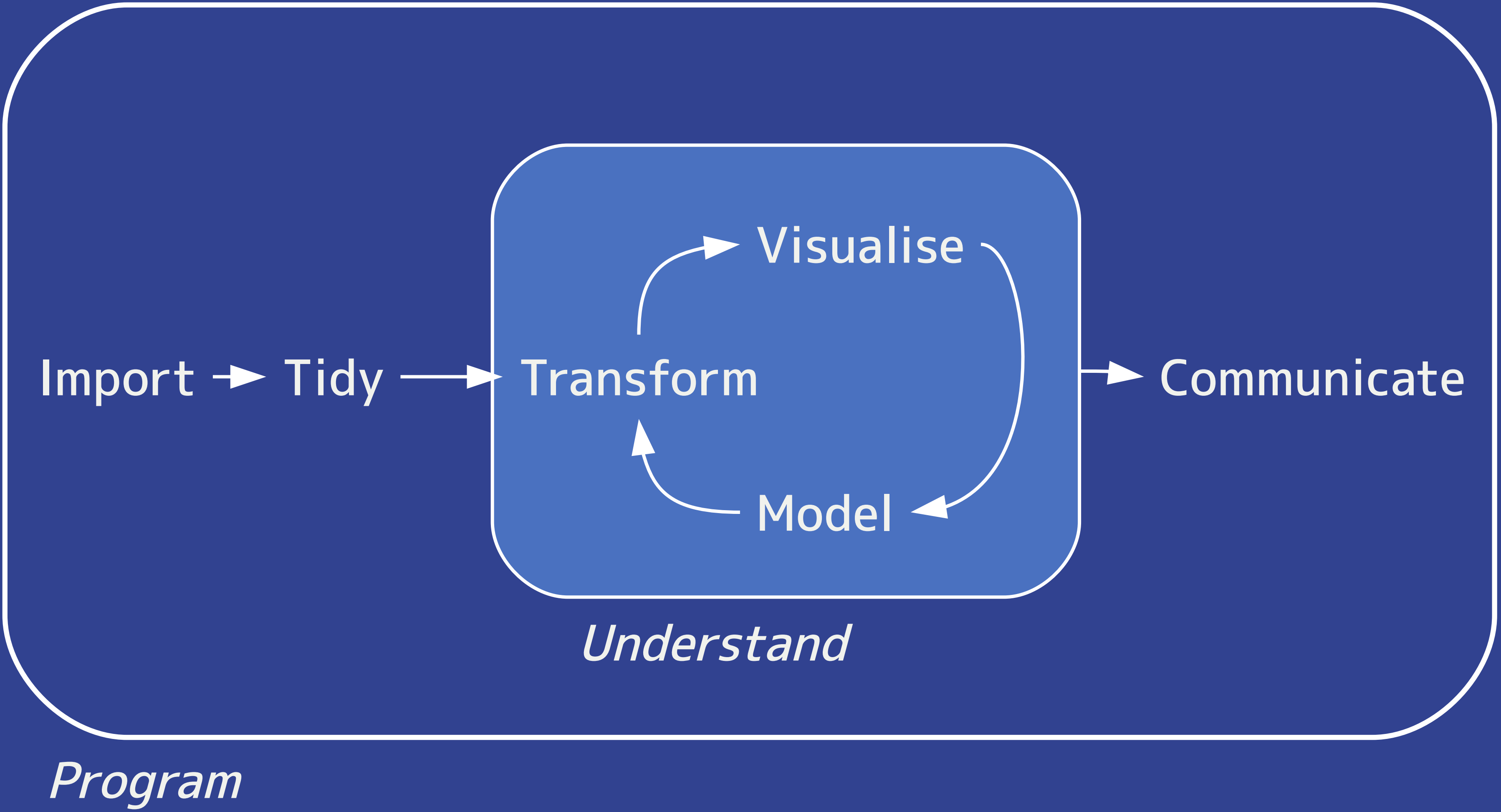
My name is

NAVIZ

HELLO

My name is

ISHLA



Import

Tidy

Transform

Visualise

Model

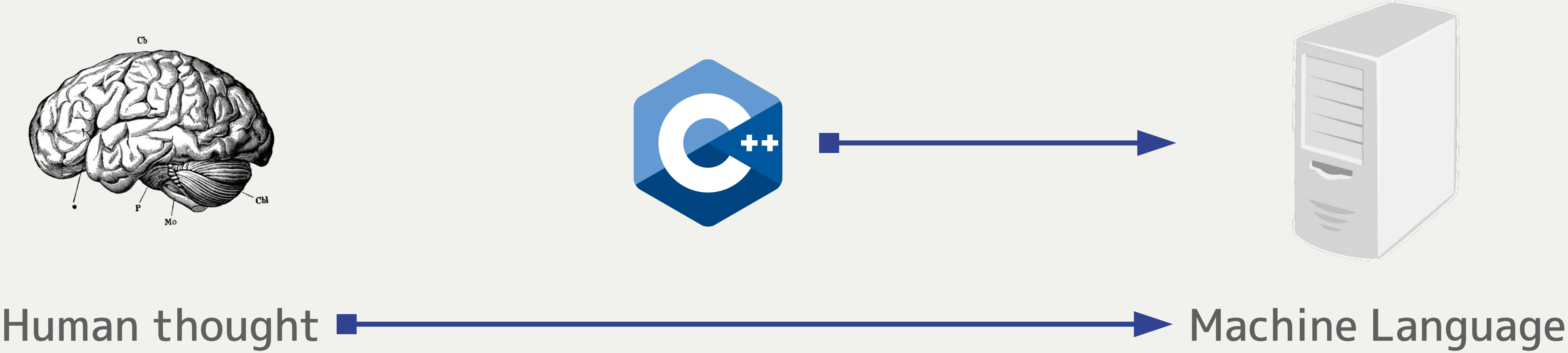
Communicate

*Understand*

*Program*

Human thought → Machine Language

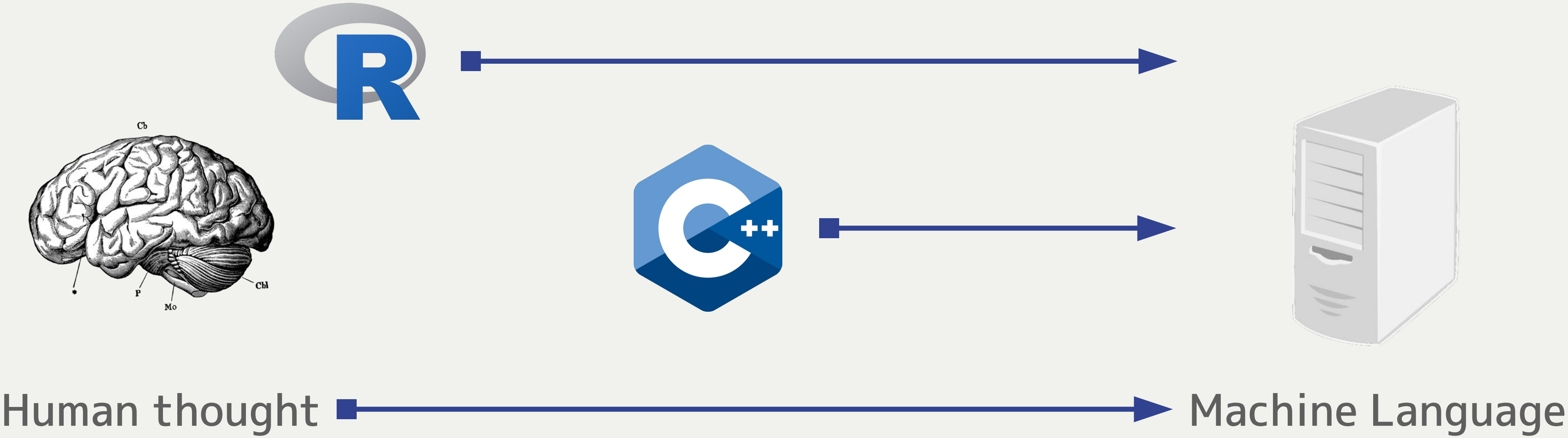
Source: https://github.com/rstudio-education/arm-workshop-rsc2019



Human thought

Machine Language

Source: https://github.com/rstudio-education/arm-workshop-rsc2019

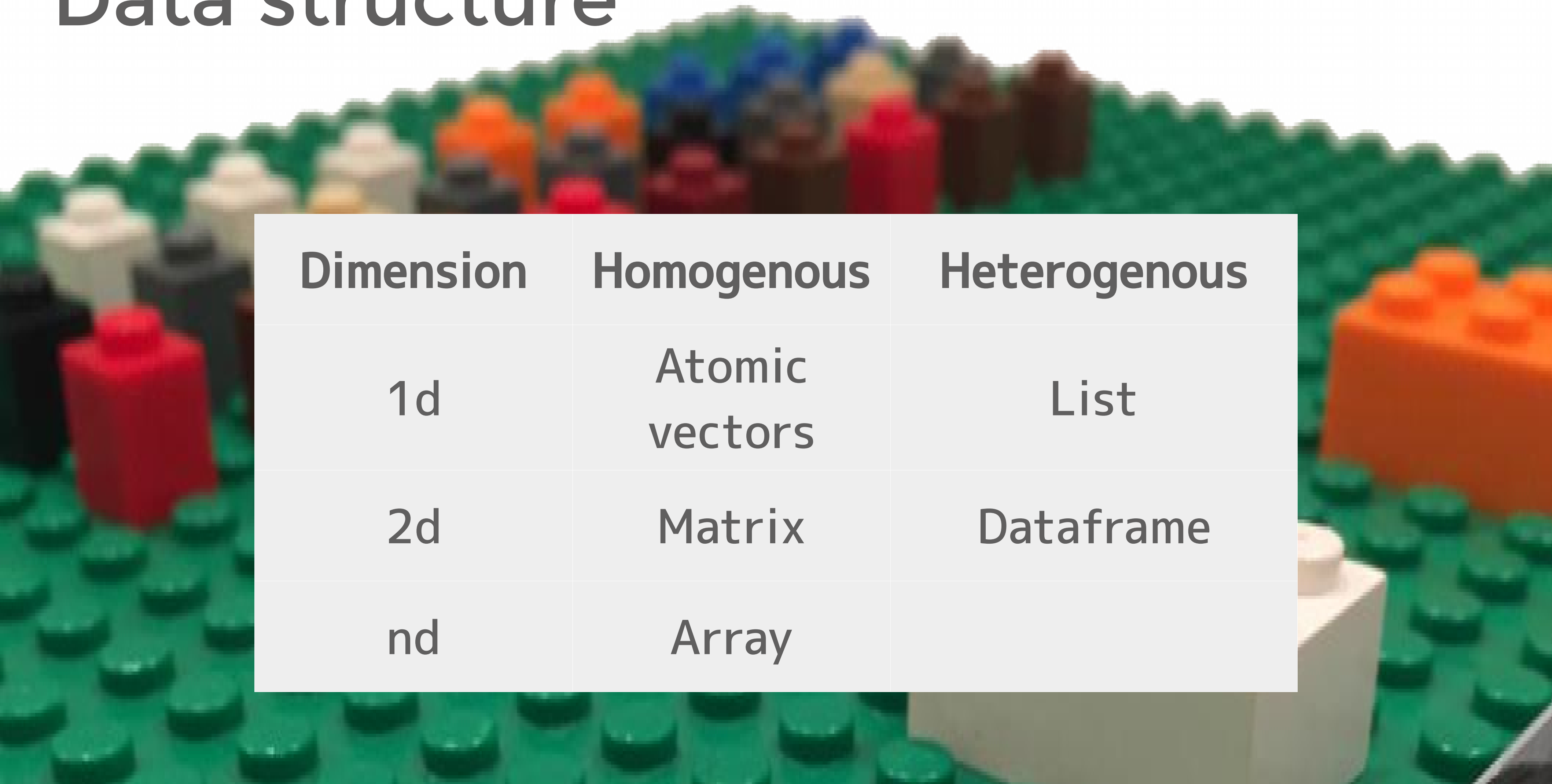


Human thought

Machine Language

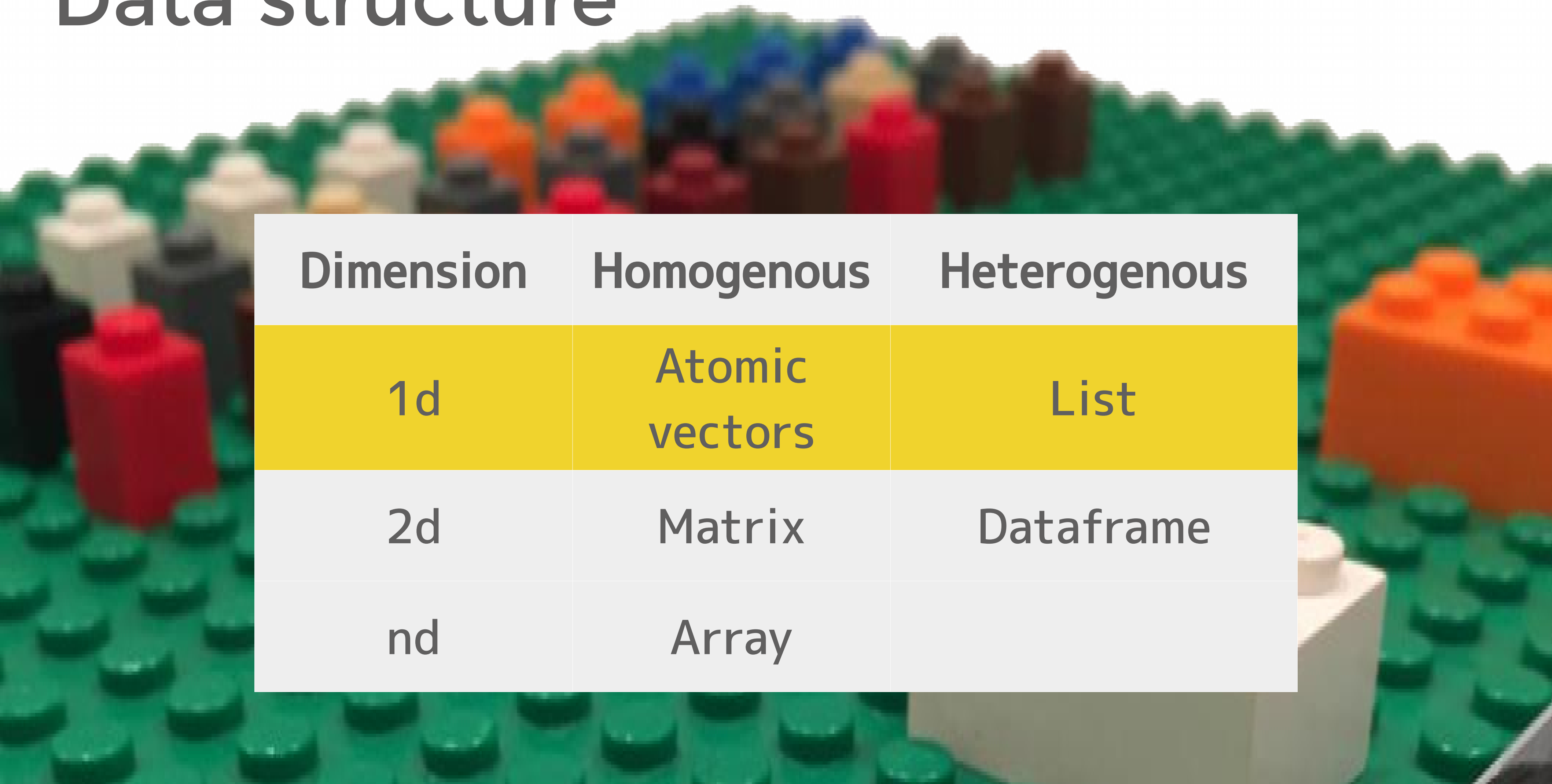
Source: https://github.com/rstudio-education/arm-workshop-rsc2019

# Data structure

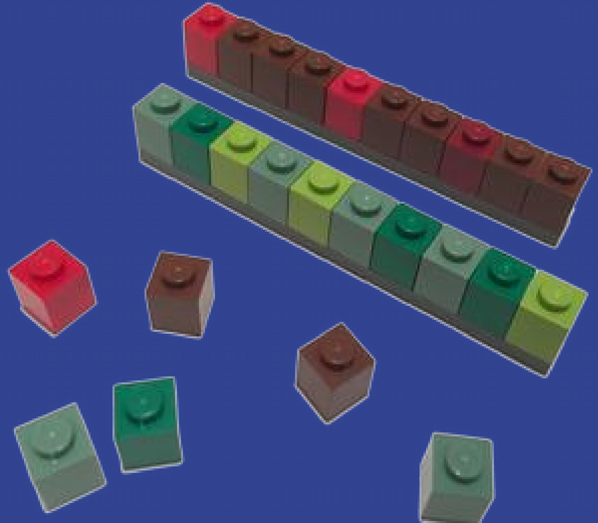


| Dimension | Homogenous | Heterogenous |
|---|---|---|
| 1d | Atomic vectors | List |
| 2d | Matrix | Dataframe |
| nd | Array | |

# Data structure

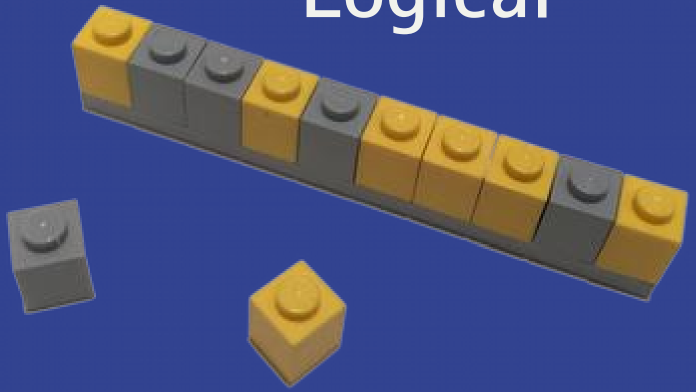


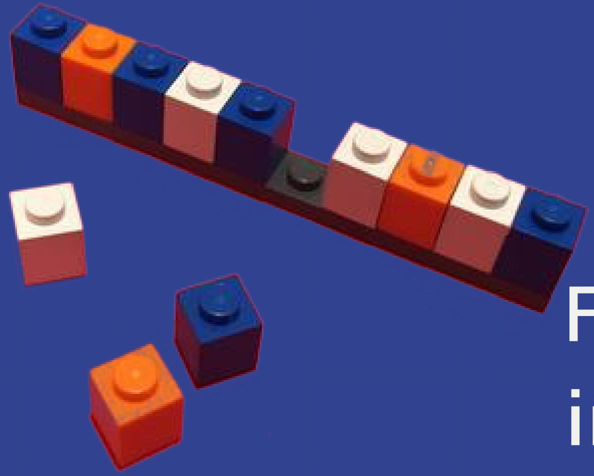
| Dimension | Homogenous | Heterogenous |
|---|---|---|
| 1d | Atomic vectors | List |
| 2d | Matrix | Dataframe |
| nd | Array | |



Integer, Double, Character

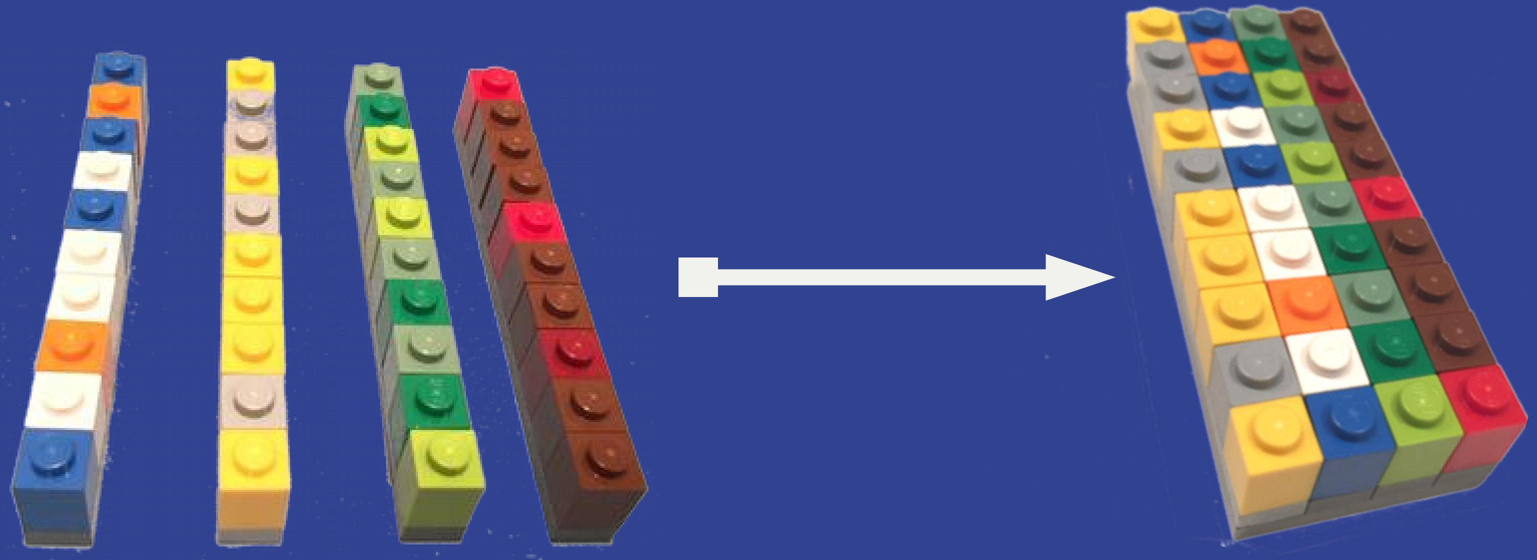
Logical





Factor (basically integer with class)

Artwork by @JennyBryan

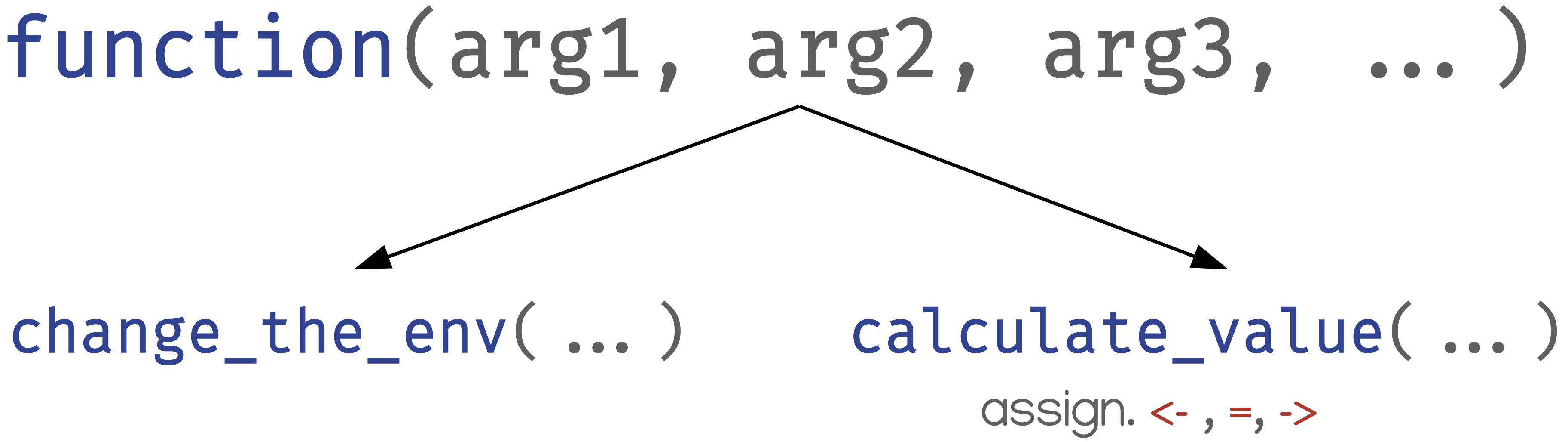


Vectors of same length

Dataframe

Artwork by @JennyBryan

# How do we process the data?



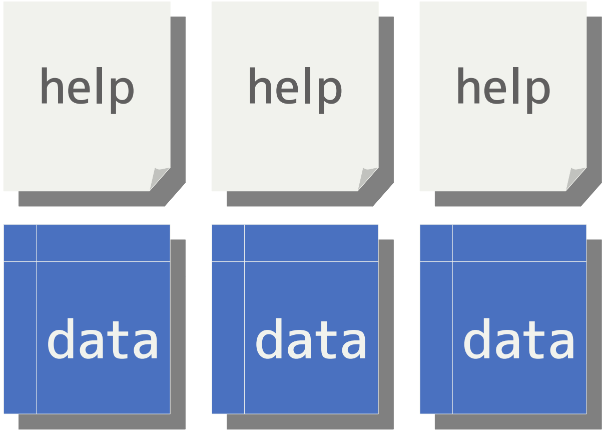
```
function(arg1, arg2, arg3, ... )
```

```
change_the_env( ... )
```

```
calculate_value( ... )
```

assign. <- , =, ->

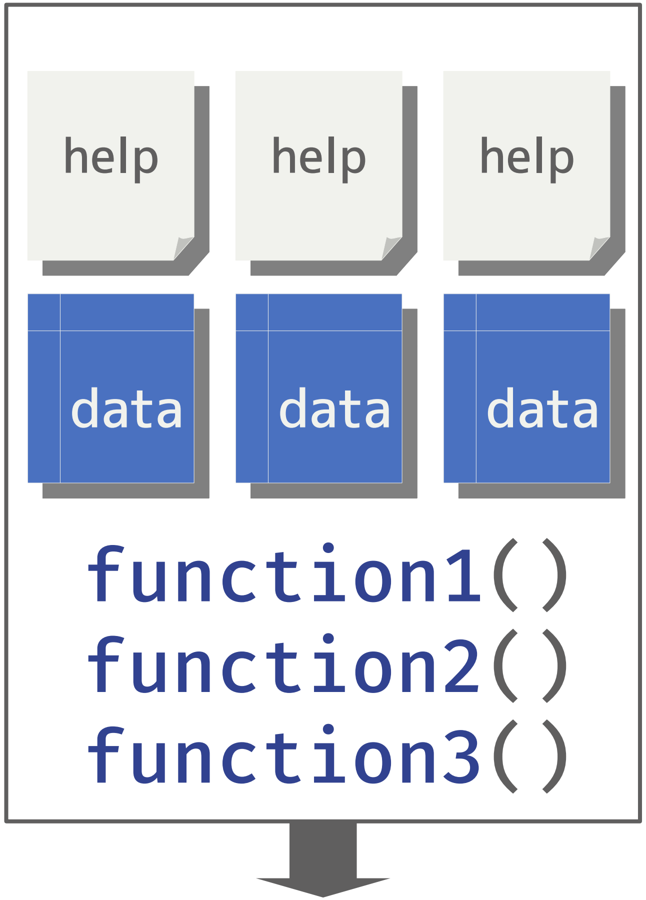
- arguments are contexts of a function
- arguments are matched by name, or
- arguments are matched by position, **be careful!**



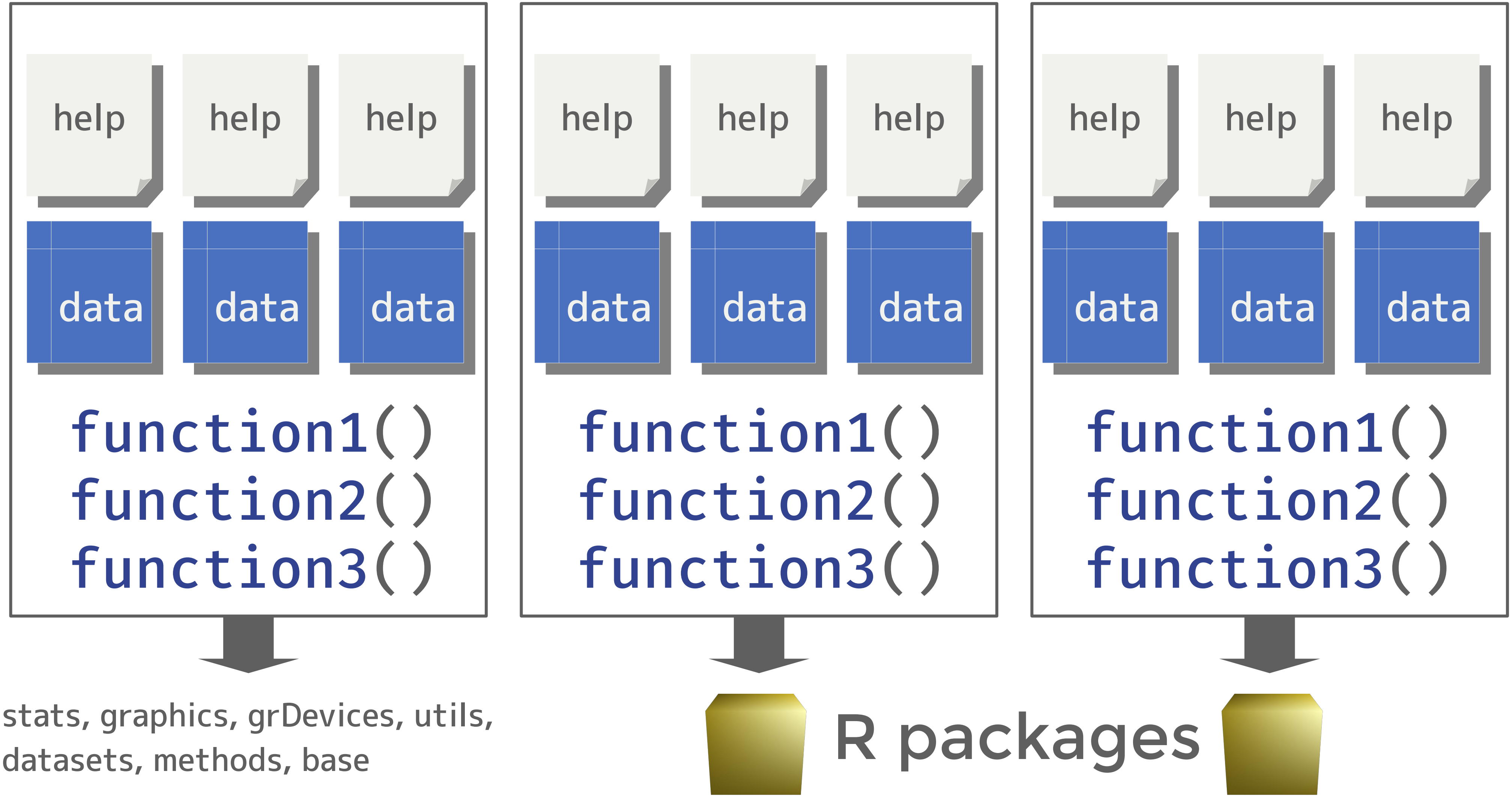
help help help

data data data

```
function1()
function2()
function3()
```



stats, graphics, grDevices, utils,
datasets, methods, base



function1()
function2()
function3()

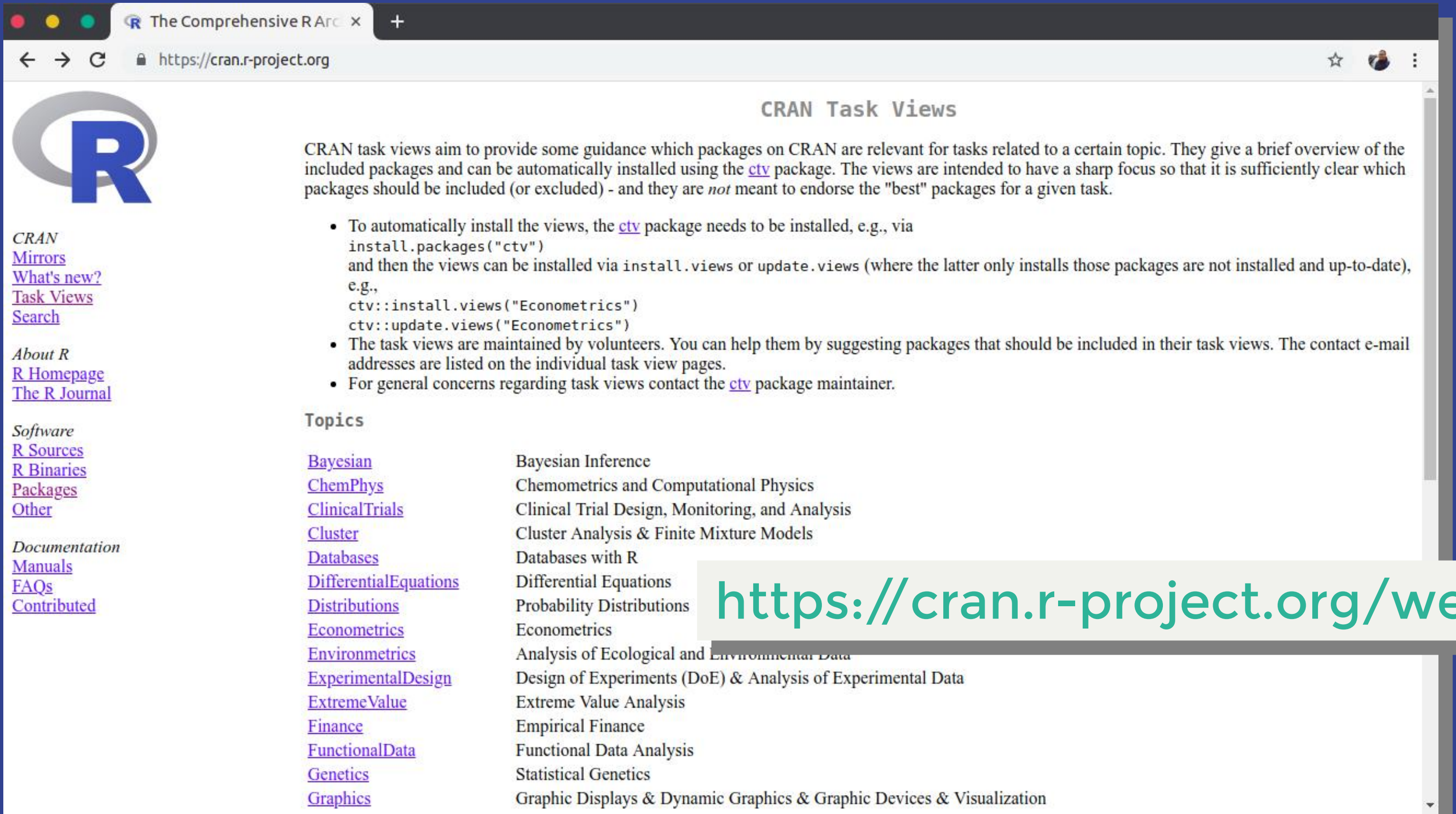
function1()
function2()
function3()

function1()
function2()
function3()

stats, graphics, grDevices, utils,
datasets, methods, base

R packages

# CRAN Task Views



## CRAN Task Views

CRAN task views aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They give a brief overview of the included packages and can be automatically installed using the ctv package. The views are intended to have a sharp focus so that it is sufficiently clear which packages should be included (or excluded) - and they are *not* meant to endorse the "best" packages for a given task.

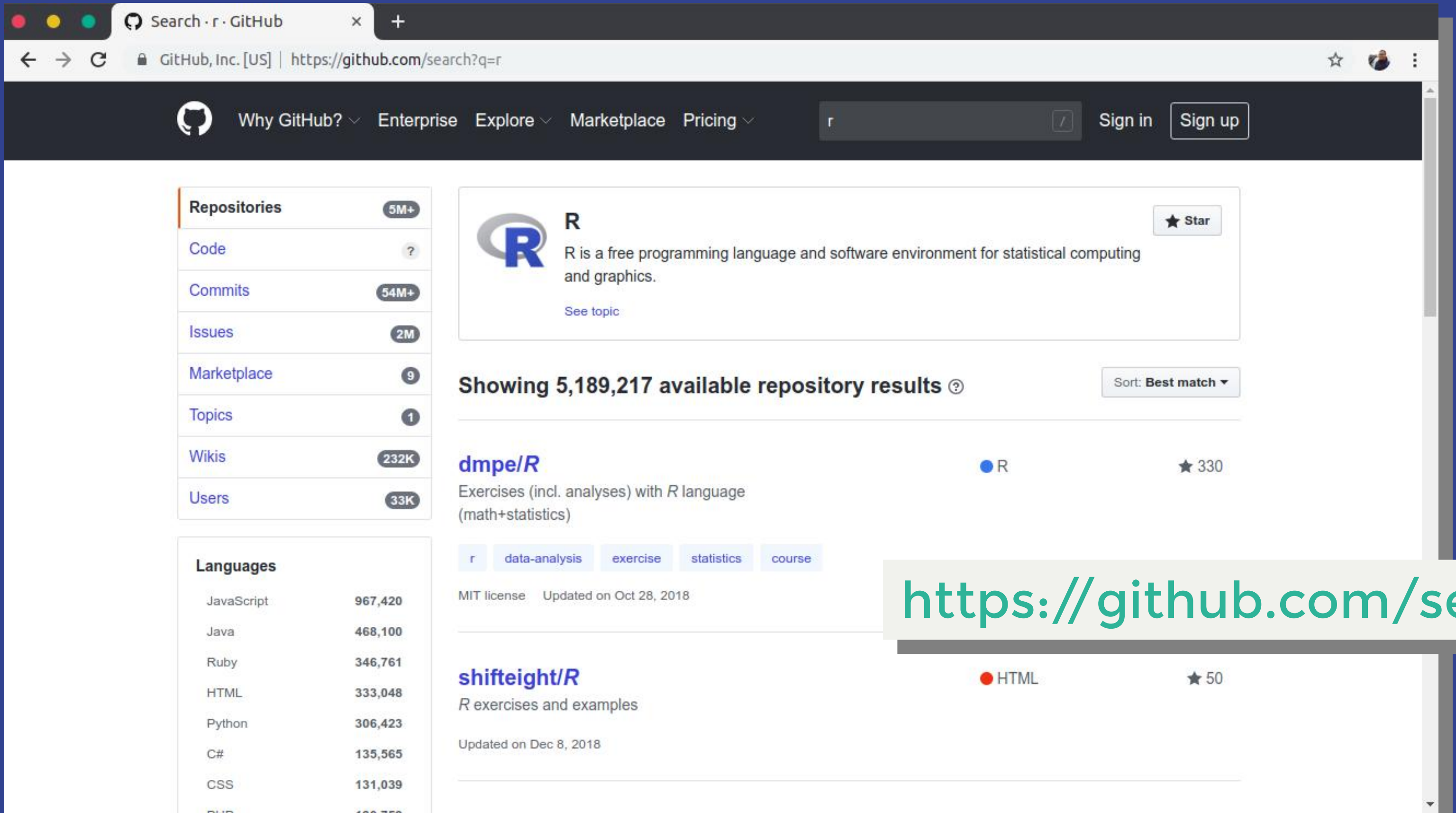
- To automatically install the views, the ctv package needs to be installed, e.g., via
  ```
  install.packages("ctv")
  ```
  and then the views can be installed via `install.views` or `update.views` (where the latter only installs those packages are not installed and up-to-date), e.g.,
  ```
  ctv::install.views("Econometrics")
  ctv::update.views("Econometrics")
  ```
- The task views are maintained by volunteers. You can help them by suggesting packages that should be included in their task views. The contact e-mail addresses are listed on the individual task view pages.
- For general concerns regarding task views contact the ctv package maintainer.

### Topics

| | |
|---|---|
| Bayesian | Bayesian Inference |
| ChemPhys | Chemometrics and Computational Physics |
| ClinicalTrials | Clinical Trial Design, Monitoring, and Analysis |
| Cluster | Cluster Analysis & Finite Mixture Models |
| Databases | Databases with R |
| DifferentialEquations | Differential Equations |
| Distributions | Probability Distributions |
| Econometrics | Econometrics |
| Environmetrics | Analysis of Ecological and Environmental Data |
| ExperimentalDesign | Design of Experiments (DoE) & Analysis of Experimental Data |
| ExtremeValue | Extreme Value Analysis |
| Finance | Empirical Finance |
| FunctionalData | Functional Data Analysis |
| Genetics | Statistical Genetics |
| Graphics | Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization |

https://cran.r-project.org/web/views/

# GitHub



| | |
|---|---|
| Repositories | 5M+ |
| Code | ? |
| Commits | 54M+ |
| Issues | 2M |
| Marketplace | 9 |
| Topics | 1 |
| Wikis | 232K |
| Users | 33K |

**Languages**

| | |
|---|---|
| JavaScript | 967,420 |
| Java | 468,100 |
| Ruby | 346,761 |
| HTML | 333,048 |
| Python | 306,423 |
| C# | 135,565 |
| CSS | 131,039 |

**R**

R is a free programming language and software environment for statistical computing and graphics.

See topic

**Showing 5,189,217 available repository results**

**dmpe/*R*** — ● R — ★ 330

Exercises (incl. analyses) with *R* language (math+statistics)

r · data-analysis · exercise · statistics · course

MIT license · Updated on Oct 28, 2018

**shifteight/*R*** — ● HTML — ★ 50

*R* exercises and examples

Updated on Dec 8, 2018

https://github.com/search?q=r

## Installing package, only once

```
install.packages(“pkg”) # from CRAN/local
remotes::install_github(“user/pkg”) # from GitHub
remotes::install_bioc(“repo”) # from Bioconductor
```

## Loading package, once in every session

```
library(pkg)
pacman::p_load(pkg) # load or install if not
available
```

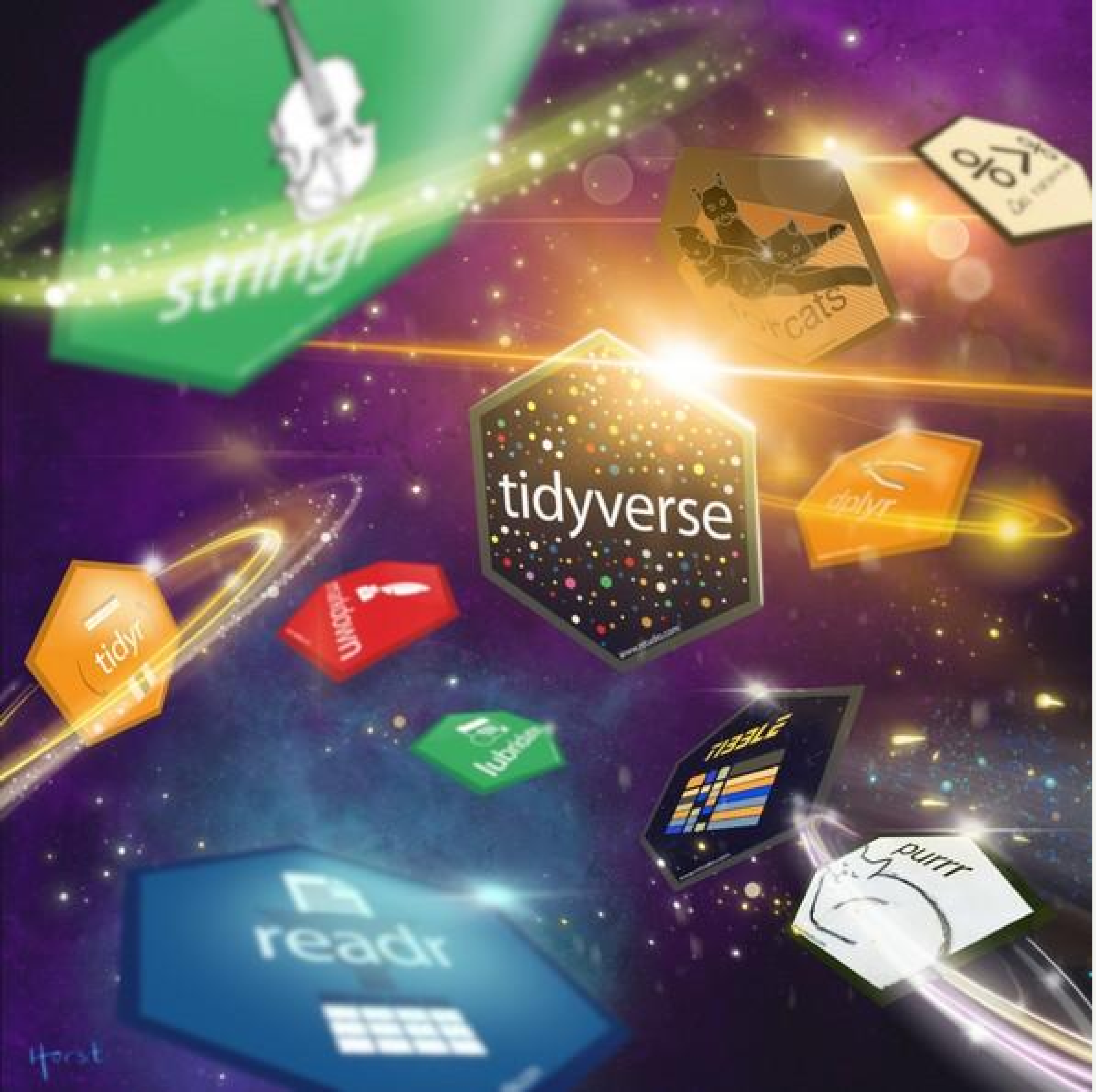
A lot of R packages

# to use! :D

A lot of R packages

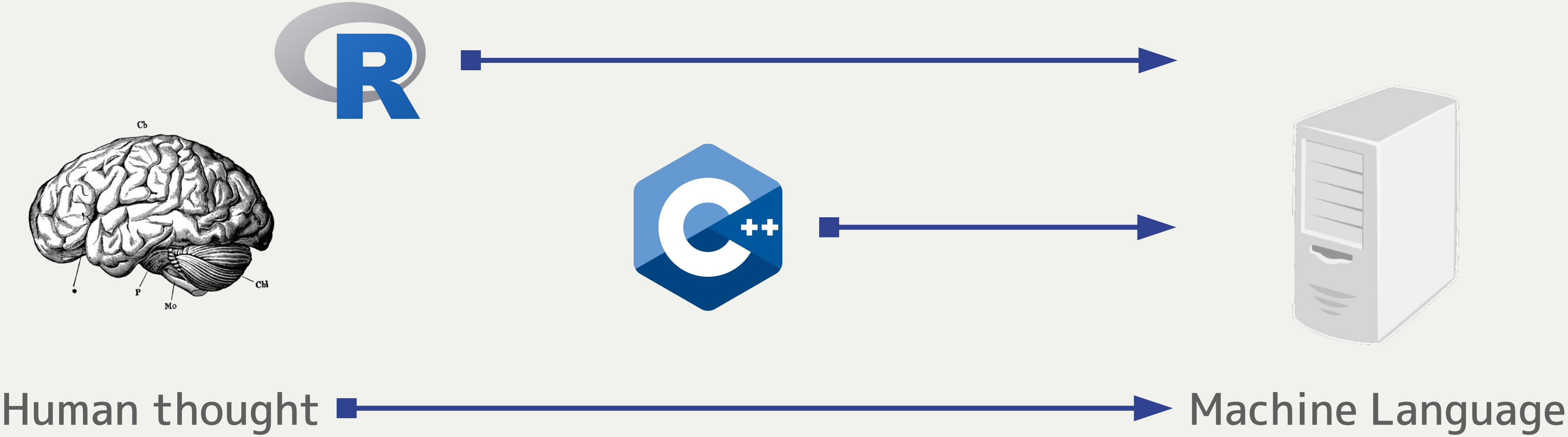
# to use! :(

tidyverse
www.rstudio.com



The tidyverse is an opinionated **collection of R packages** designed for **data science.** All packages **share** an underlying **design** philosophy, **grammar**, and **data structures**.

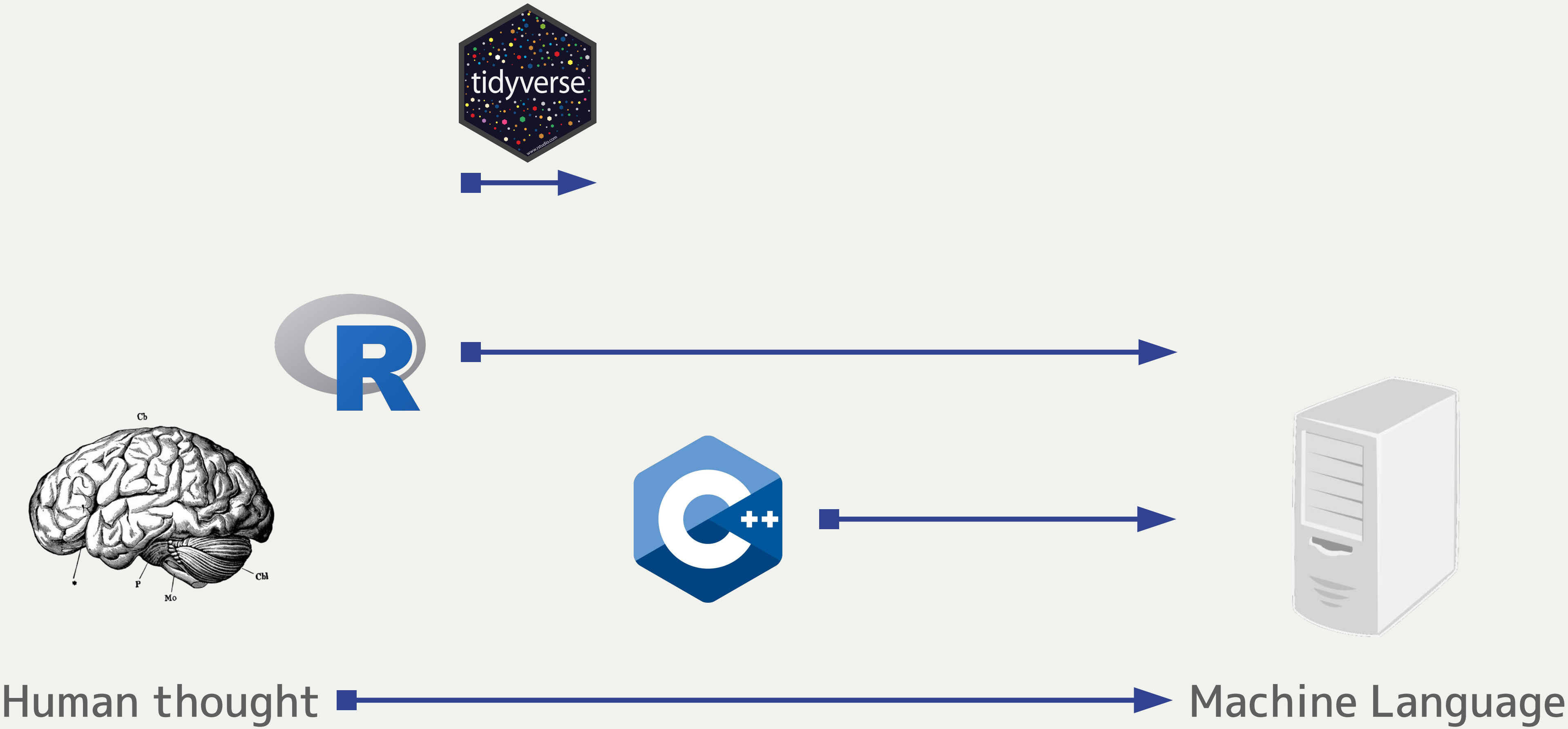
Artwork by @allison_horst



Human thought

Machine Language

Source: https://github.com/rstudio-education/arm-workshop-rsc2019



Human thought

Machine Language

Source: https://github.com/rstudio-education/arm-workshop-rsc2019

# R Syntax Comparison

## Dollar sign

```
goal(data$x, data$y)
```

- A.k.a base syntax
- Subsetting data by using '[ ]'

## Formula

```
goal(y~x, data=data)
```

Mostly used in modeling and statistical test

## Tidyverse

```
data %>% goal(x, y)
```

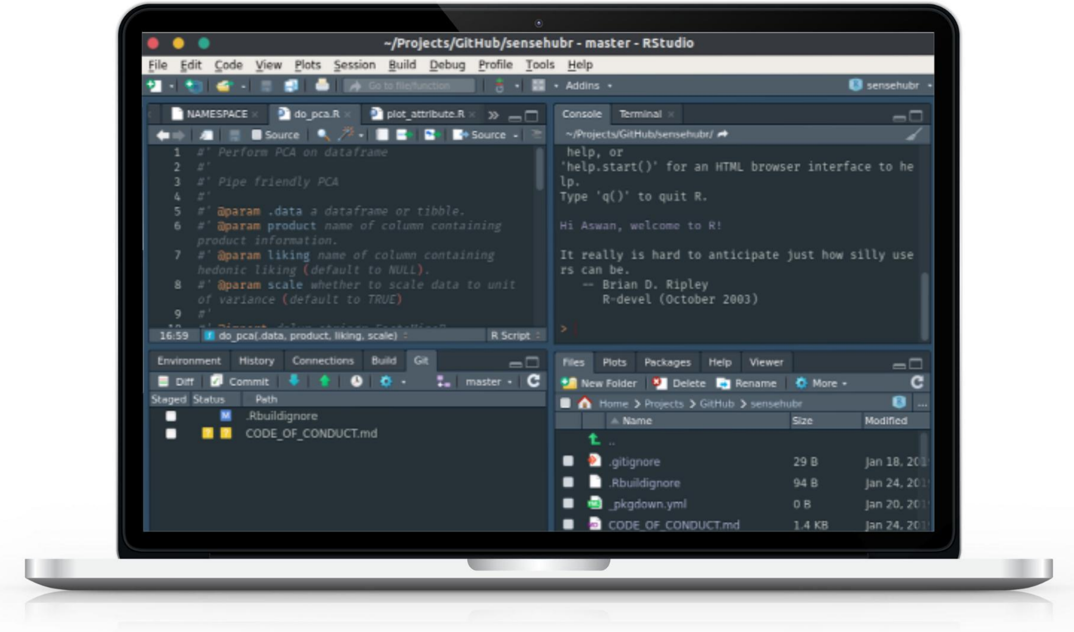
- Expecting data as the first argumen
- Plotting using '+' flavour

Cheatsheet: https://github.com/rstudio/cheatsheets/raw/master/syntax.pdf

# RStudio®

## Main features:

- Console
- Syntax-highlighting editor
- Tools for plotting, history, debugging and workspace management



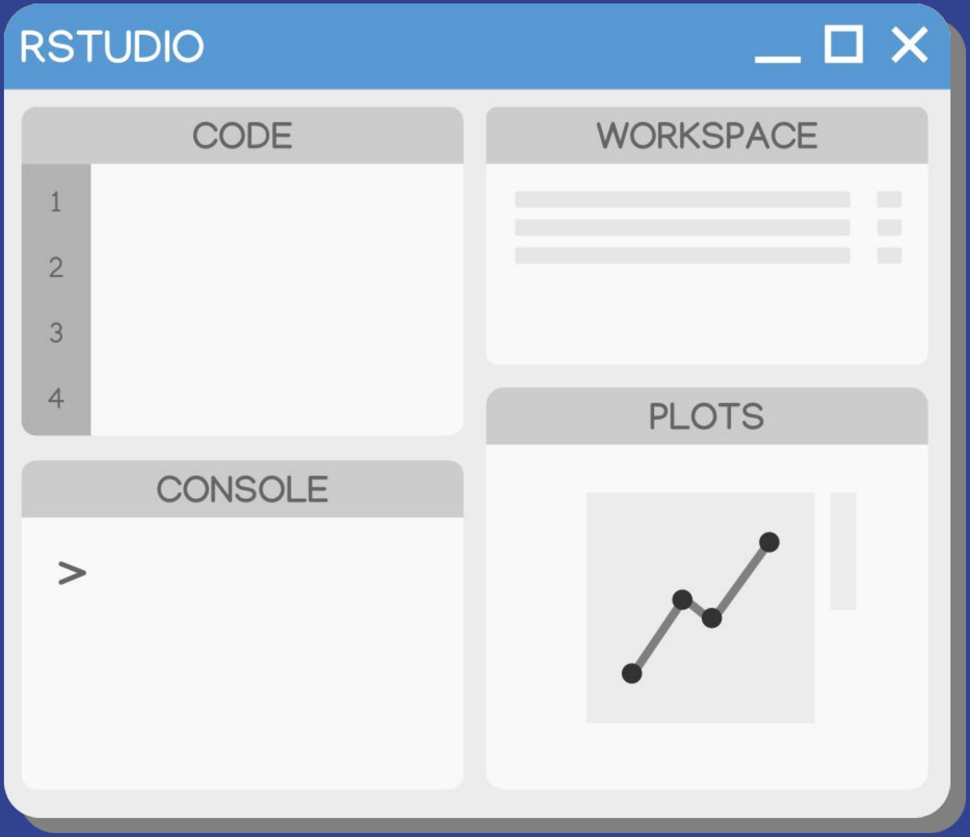
```r
#' Perform PCA on dataframe
#'
#' Pipe friendly PCA
#'
#' @param .data a dataframe or tibble.
#' @param product name of column containing product information.
#' @param liking name of column containing hedonic liking (default to NULL).
#' @param scale whether to scale data to unit of variance (default to TRUE)
#'
```

```
 help, or
'help.start()' for an HTML browser interface to he
lp.
Type 'q()' to quit R.

Hi Aswan, welcome to R!

It really is hard to anticipate just how silly use
rs can be.
   -- Brian D. Ripley
      R-devel (October 2003)

>
```

| Staged | Status | Path |
|---|---|---|
| □ | M | .Rbuildignore |
| □ | ? ? | CODE_OF_CONDUCT.md |

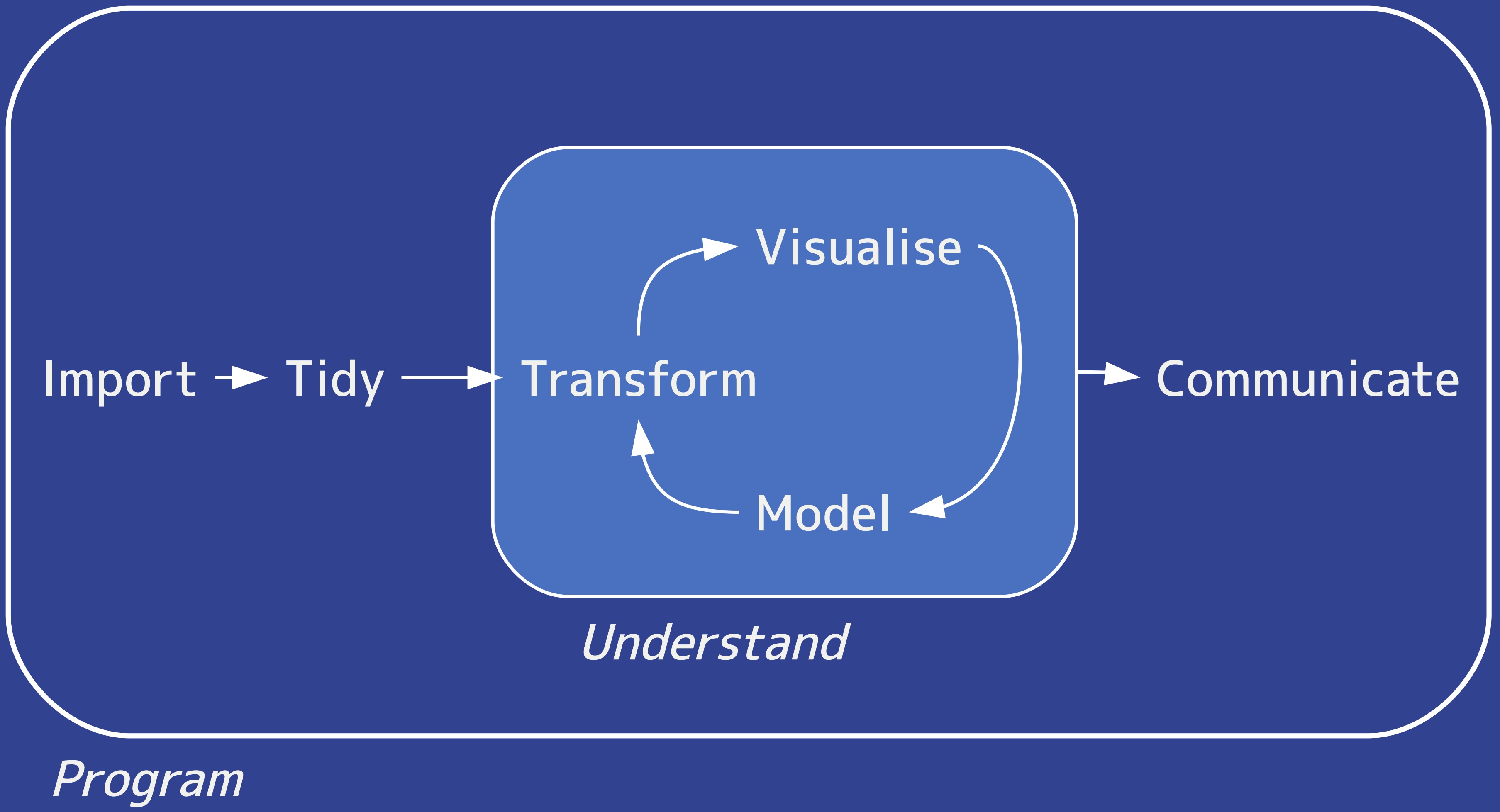
| Name | Size | Modified |
|---|---|---|
| .. | | |
| .gitignore | 29 B | Jan 18, 201 |
| .Rbuildignore | 94 B | Jan 24, 201 |
| _pkgdown.yml | 0 B | Jan 20, 201 |
| CODE_OF_CONDUCT.md | 1.4 KB | Jan 24, 201 |

rstudio.com/products/rstudio/download/



- **Tab**, autocompletion & path navigation
- **Alt + -**, for assignment operator <-
- **Ctrl + Shift + M**, for pipe operator %>%
- **Ctrl + Enter**, run current line code/example on help page
- **Ctrl + Up**, search for code history on console or editor pane
- **Alt + Up/Down**, move code to above or below
- **Alt + Shift + Up/Down**, copy code to above or below
- **Ctrl + D**, delete current line
- **Ctrl + Shift + F10**, restart R session
- **Ctrl + Alt + B**, run code up to current line

Cheatsheet: https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf



Import

Tidy

Transform

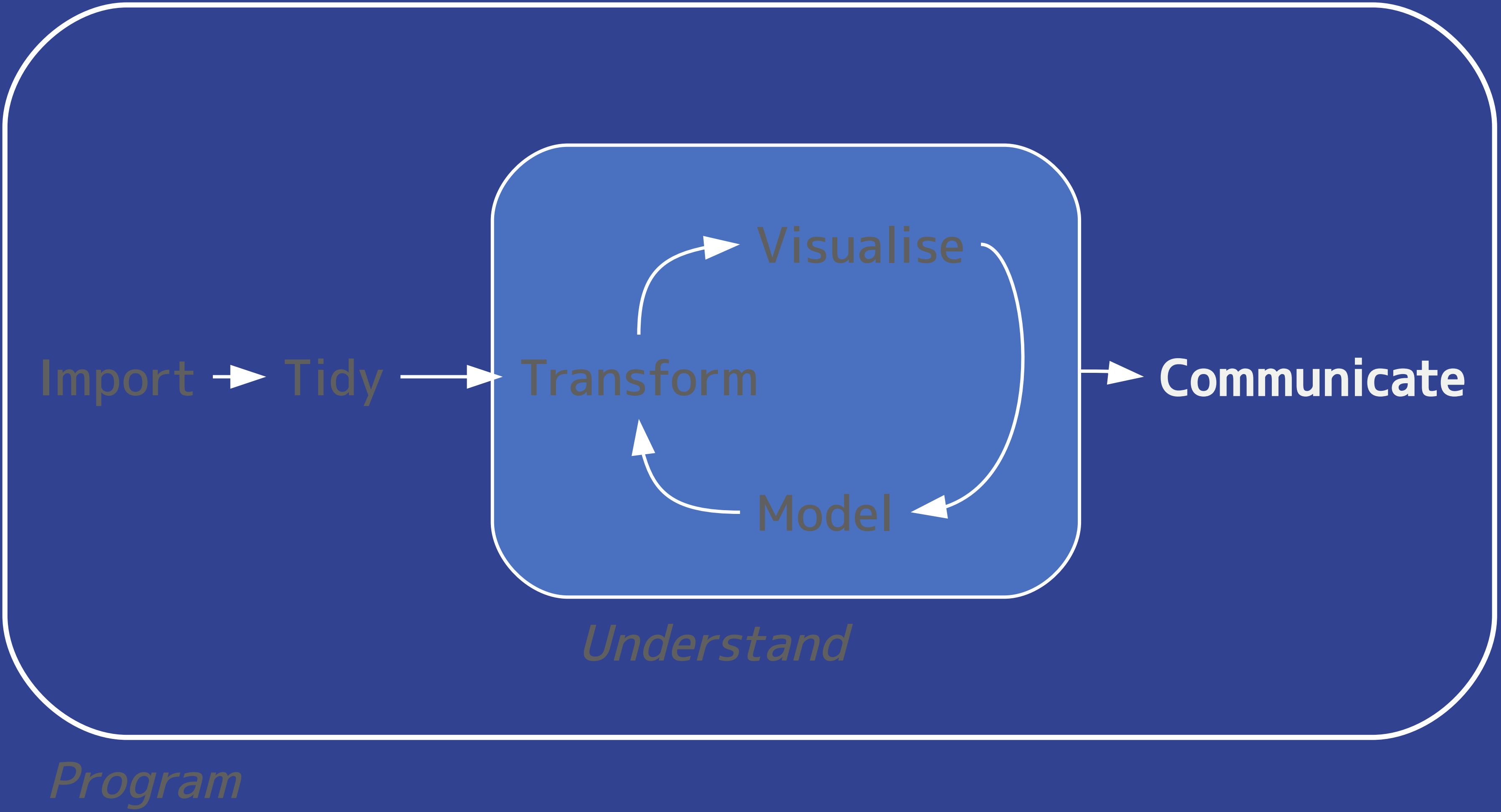
Visualise

Model

Communicate

*Understand*

*Program*



Import

Tidy

Transform

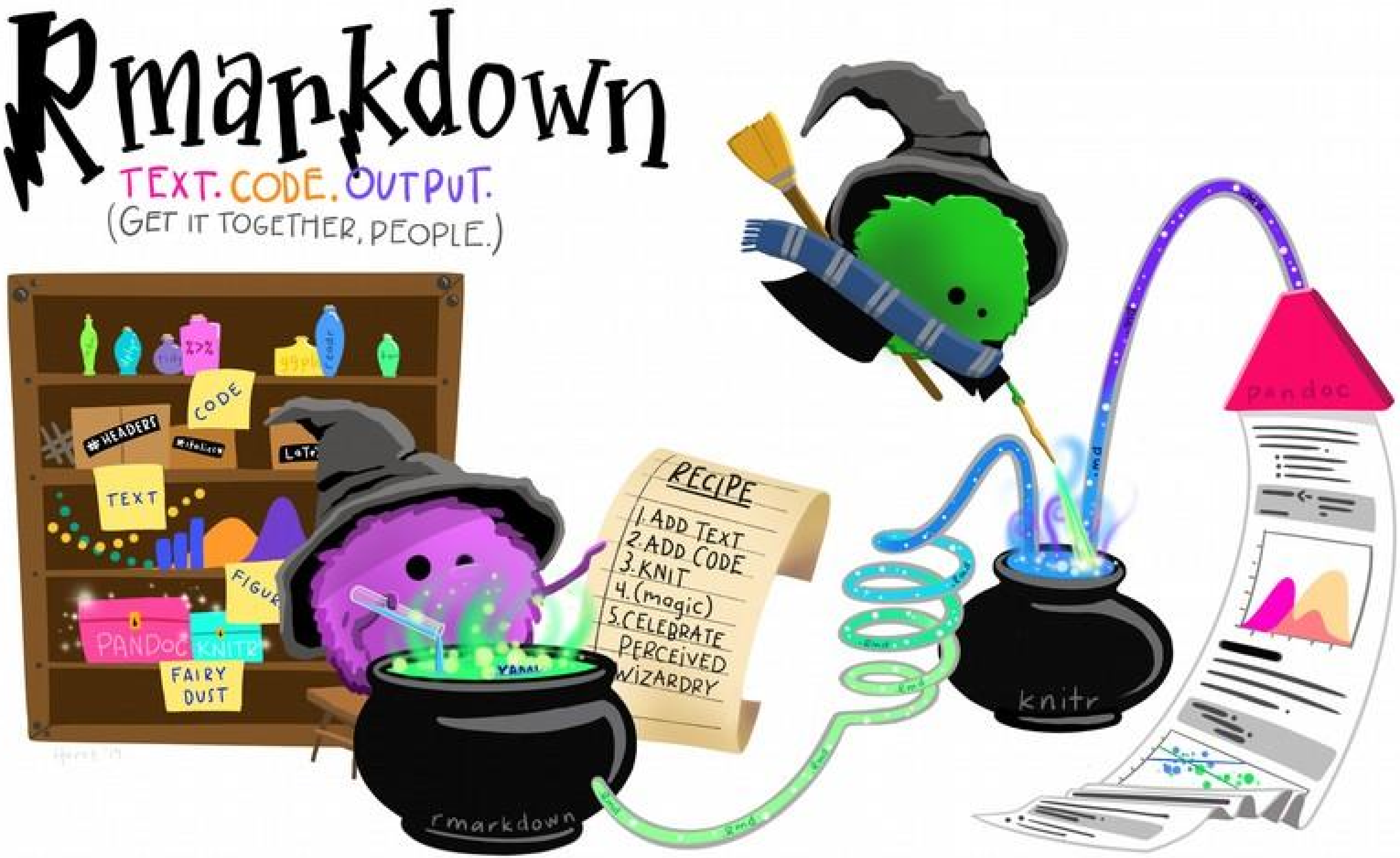
Visualise

Model

Communicate

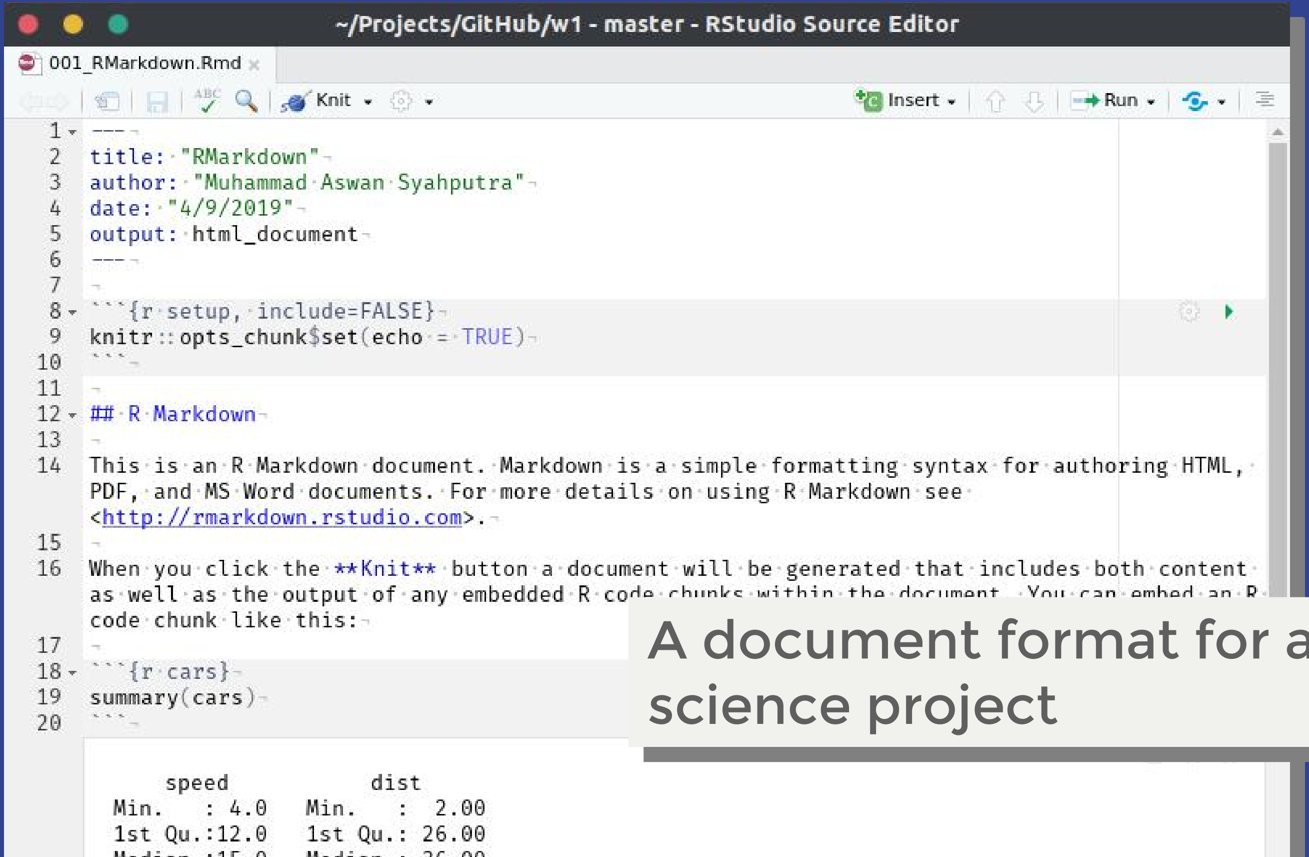
*Understand*

*Program*



Artwork by @allison_horst

# R Markdown



~~~
---
title: "RMarkdown"
author: "Muhammad Aswan Syahputra"
date: "4/9/2019"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
~~~

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML,
PDF, and MS Word documents. For more details on using R Markdown see
<http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content
as well as the output of any embedded R code chunks within the document. You can embed an R
code chunk like this:

```{r cars}
summary(cars)
```
```

```
     speed           dist
 Min.   : 4.0   Min.   :  2.00
 1st Qu.:12.0   1st Qu.: 26.00
 Median :15.0   Median : 36.00
```

A document format for authoring data science project
```

- Use script (R Markdown or R Script), try to avoid console
- Use Projects, not `setwd( ... )` in script
- Set `stringsAsFactor = FALSE`, but not in the .Rprofile
- Ctrl+Shift+F10 and Ctrl+Alt+B to clean up, not `rm(list=ls())`
- Learn the handy shortcuts
- Do not save and load .Rdata
- Use version control system: git!

Reading: happygitwithr.com

git

Download: git-scm.com

# With great codes,
# comes great bugs!

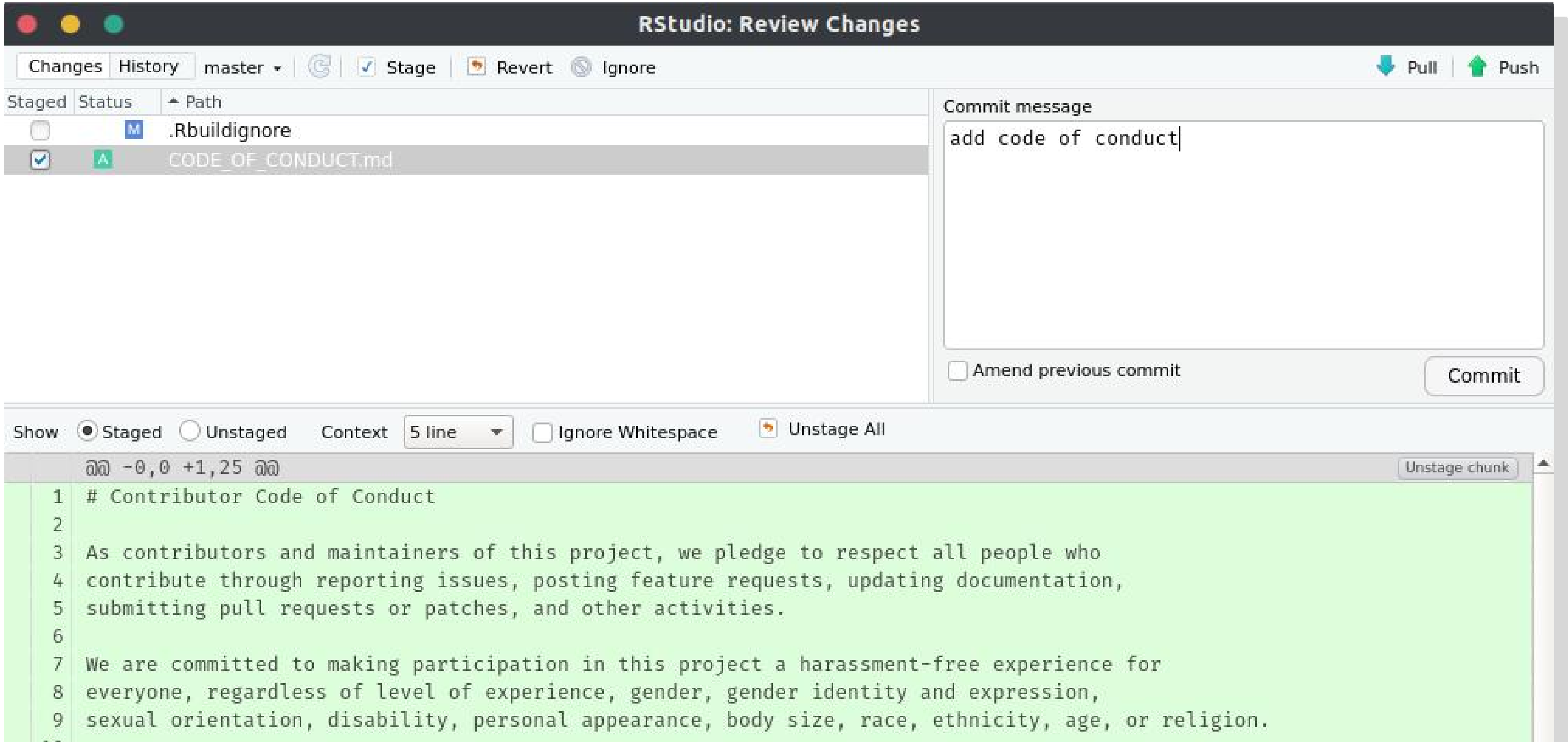
*- (not) Uncle Ben*

Store and share! Why sharing your work? Motivation here.

- `git clone https://github/user/repo`
- Do some works!
- `git add file.R` or `git add .`
- `git commit -m “what you have done”`
- `git push origin master`
- Repeat: work, git add, git commit, git push

- `git init`
- `git remote add origin https://github/user/repo`
- Do some works!
- `git add file.R` or `git add .`
- `git commit -m “what you have done”`
- `git push -u origin master` #use -u only once
- Repeat: work, git add, git commit, git push

# It is available in RStudio!



| Staged | Status | Path |
|---|---|---|
| □ | M | .Rbuildignore |
| ☑ | A | CODE_OF_CONDUCT.md |

Commit message

add code of conduct

□ Amend previous commit

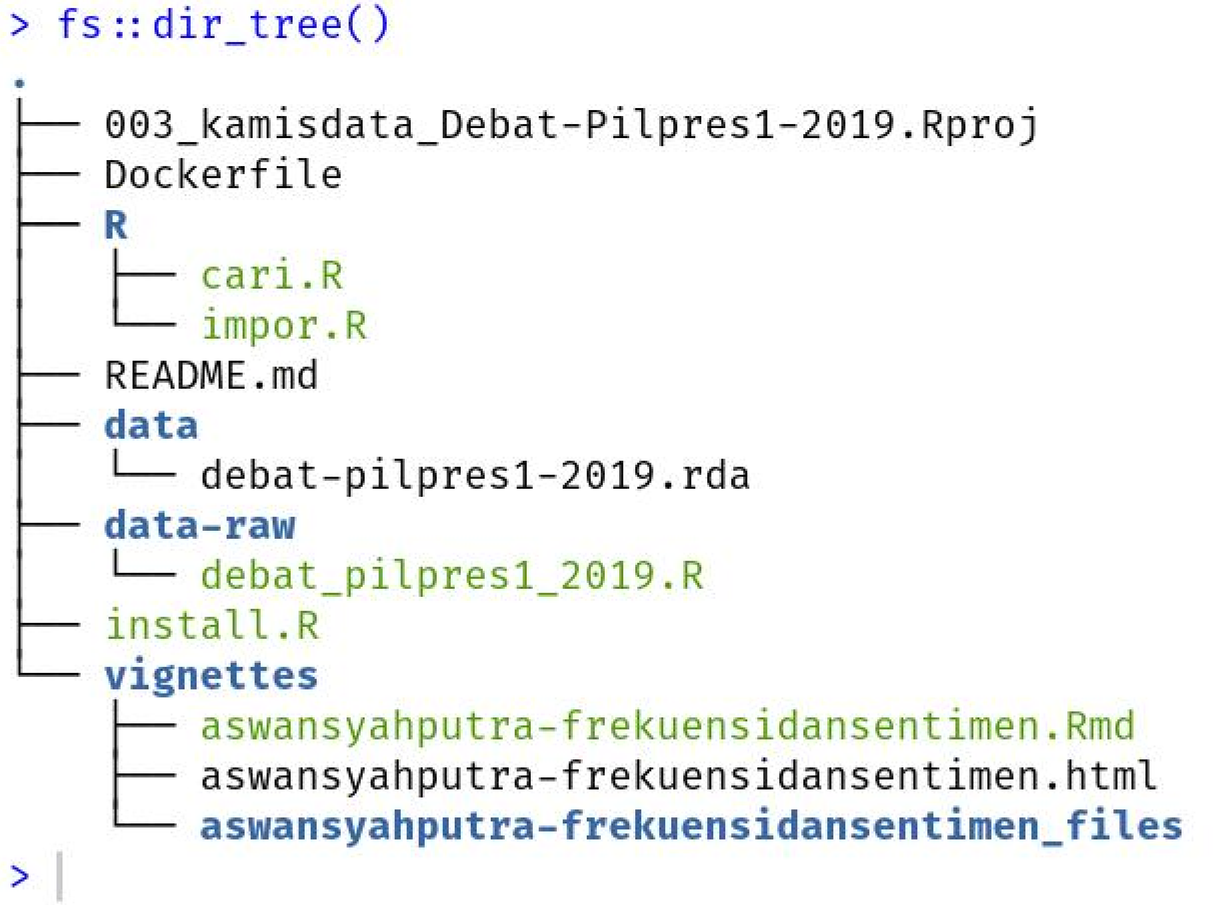
Show ● Staged ○ Unstaged Context 5 line □ Ignore Whitespace

```
@@ -0,0 +1,25 @@
# Contributor Code of Conduct

As contributors and maintainers of this project, we pledge to respect all people who
contribute through reporting issues, posting feature requests, updating documentation,
submitting pull requests or patches, and other activities.

We are committed to making participation in this project a harassment-free experience for
everyone, regardless of level of experience, gender, gender identity and expression,
sexual orientation, disability, personal appearance, body size, race, ethnicity, age, or religion.
```

# Let's get started!

- Go to github.com/r-academy/w1, click 'Fork' button
- Click 'Clone or Download', copy the URL
- In RStudio, File – New Project – Version Control – Git. Paste URL
- In File pane (bottom-right), click vignettes-'001_pendahuluan.Rmd' to open it
- You have **10 minutes** to play with it!

# Working directory



```
> fs::dir_tree()
.
├── 003_kamisdata_Debat-Pilpres1-2019.Rproj
├── Dockerfile
├── R
│   ├── cari.R
│   └── impor.R
├── README.md
├── data
│   └── debat-pilpres1-2019.rda
├── data-raw
│   └── debat_pilpres1_2019.R
├── install.R
└── vignettes
    ├── aswansyahputra-frekuensidansentimen.Rmd
    ├── aswansyahputra-frekuensidansentimen.html
    └── aswansyahputra-frekuensidansentimen_files
>
```

**3 principles for naming files:**

- Machine readable
- Human readable
- Default ordering

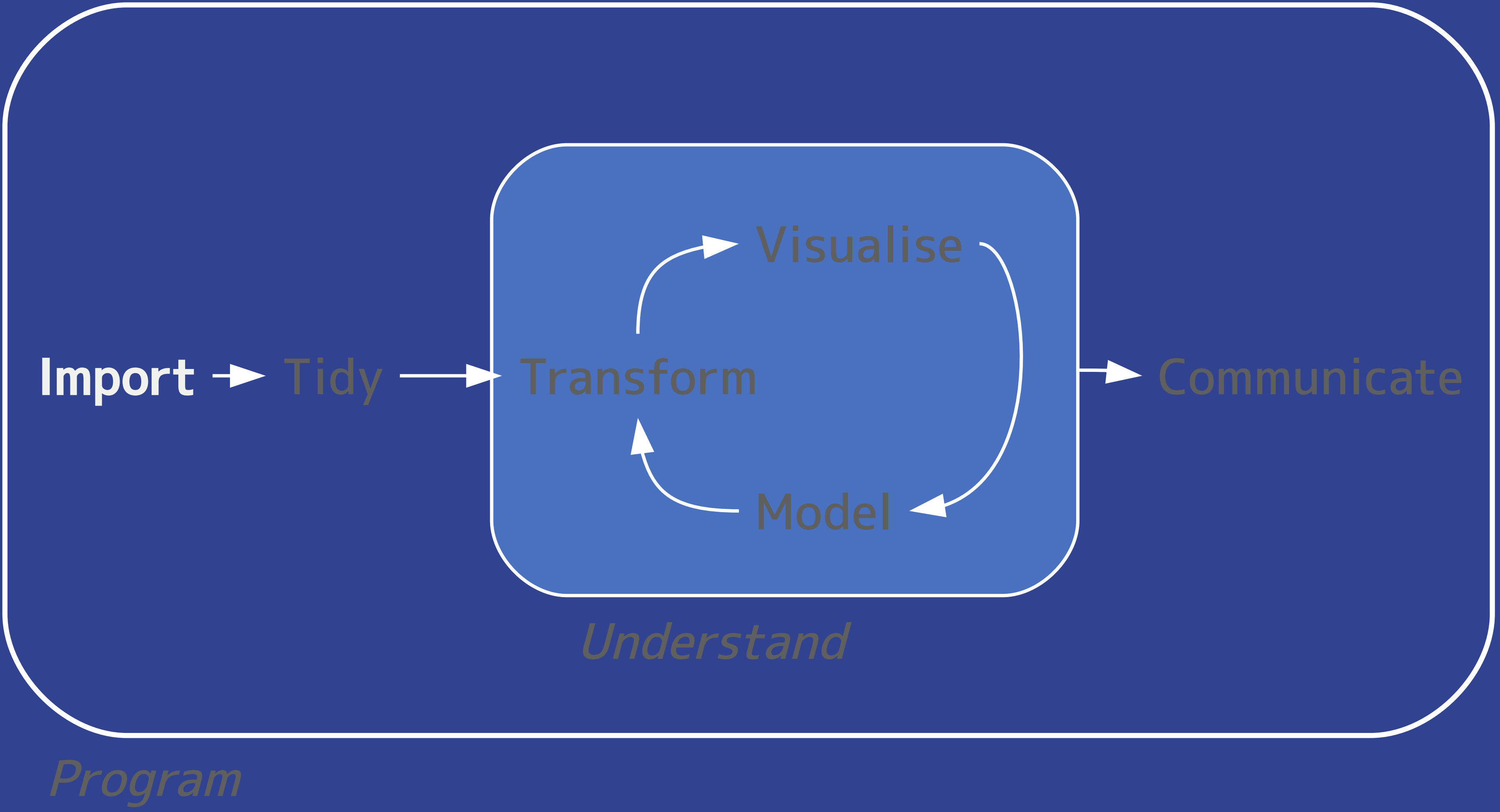
More info:
speakerdeck.com/jennybc/how-to-name-files

# How to set git?

```
git config --global
user.email
“email@domain.com”

git config --global
user.name “Your Name”
```

# How to use git?

- In Environment Pane, hit ‘Git’ tab
- Click commit, a window pane will appear
- Select all files (Ctrl + A), click ‘Stage’
- Fill commit message, the click ‘Commit’
- Hit ‘Push’ Button, done!
- You may check you GitHub now



Import

Tidy

Transform

Visualise

Model

Communicate

*Understand*

*Program*

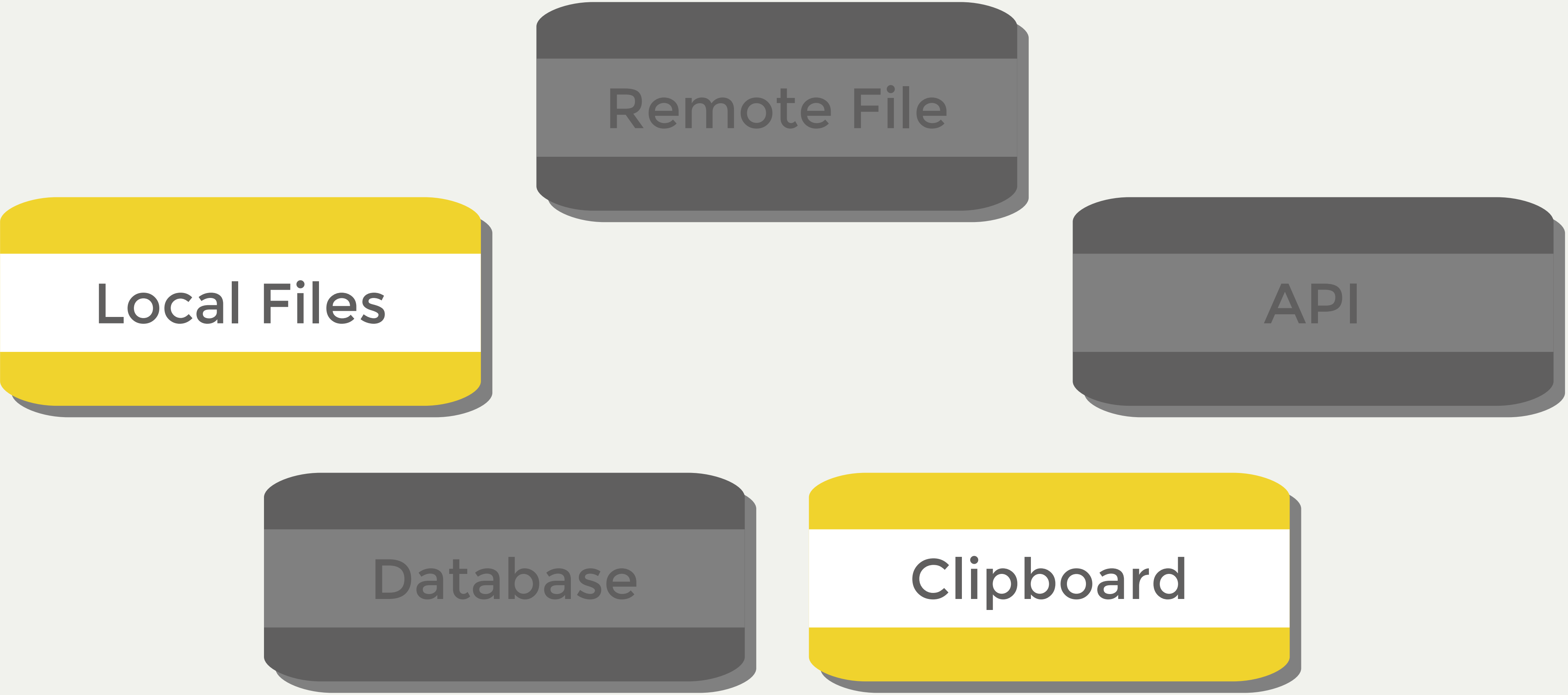
Remote File

Local Files

API

Database

Clipboard



Remote File

Local Files

API

Database

Clipboard

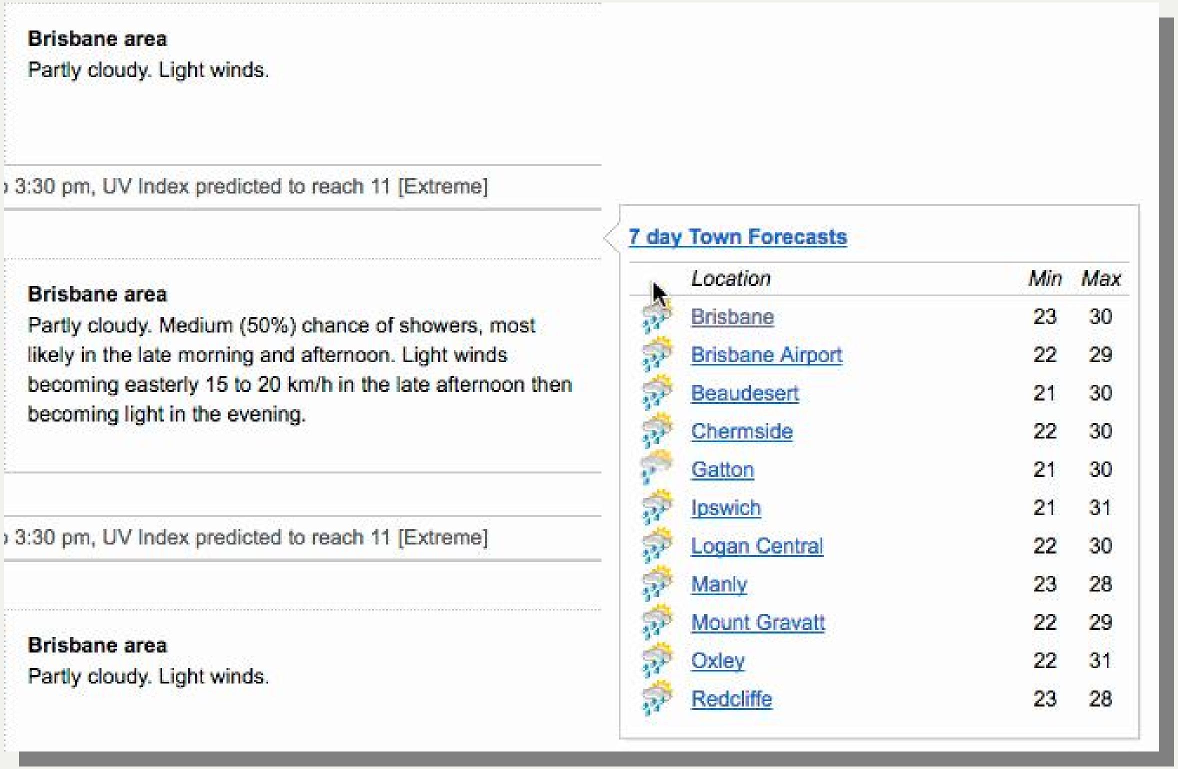
readr
www.rstudio.com

DATA
PASTA

# readr

- `read_csv()`: comma separated (CSV) files
- `read_tsv()`: tab separated files
- `read_delim()`: general delimited files
- `read_fwf()`: fixed width files
- `read_table()`: tabular files where columns are separated by white-space.
- `read_log()`: web log files

# datapasta



**Brisbane area**
Partly cloudy. Light winds.

› 3:30 pm, UV Index predicted to reach 11 [Extreme]

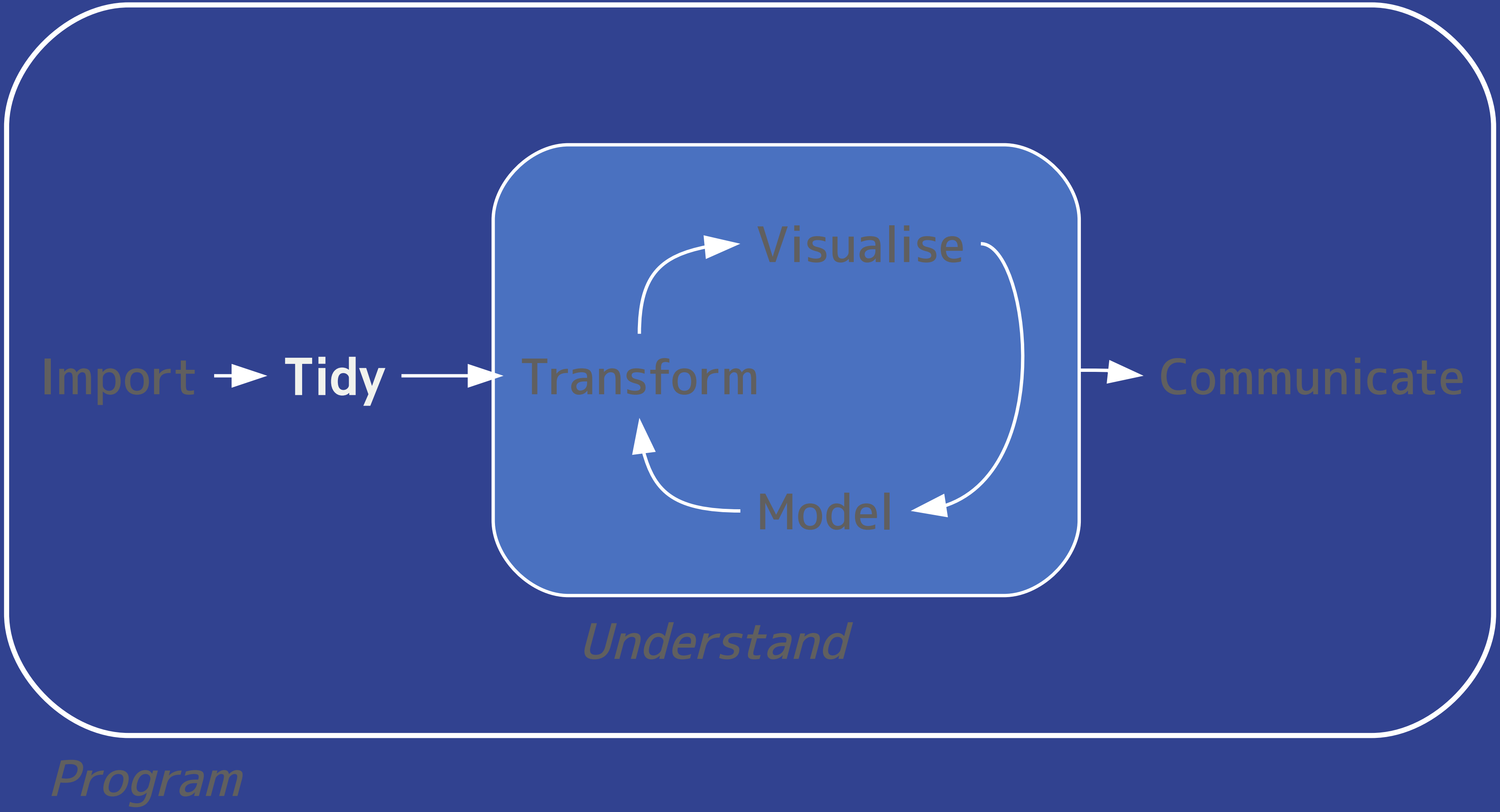
**Brisbane area**
Partly cloudy. Medium (50%) chance of showers, most likely in the late morning and afternoon. Light winds becoming easterly 15 to 20 km/h in the late afternoon then becoming light in the evening.

› 3:30 pm, UV Index predicted to reach 11 [Extreme]

**Brisbane area**
Partly cloudy. Light winds.

7 day Town Forecasts

| *Location* | *Min* | *Max* |
|---|---|---|
| Brisbane | 23 | 30 |
| Brisbane Airport | 22 | 29 |
| Beaudesert | 21 | 30 |
| Chermside | 22 | 30 |
| Gatton | 21 | 30 |
| Ipswich | 21 | 31 |
| Logan Central | 22 | 30 |
| Manly | 23 | 28 |
| Mount Gravatt | 22 | 29 |
| Oxley | 22 | 31 |
| Redcliffe | 23 | 28 |


Import
Tidy
Transform
Visualise
Model
Communicate
Understand
Program

Tidy datasets are all alike, but every messy dataset is messy in its own way!

*- Hadley Wickham*

# A Tidy dataset

| | Name | Gender | Age |
|---|---|---|---|
| 1 | Phil | Male | 54 |
| 2 | May | Female | 46 |
| 3 | Mack | NA | 31 |

# A variable has its own column

| | Var. 1 | Var. 2 | Var. 3 |
|---|---|---|---|
| Obs. 1 | A | B | C |
| Obs. 2 | D | E | F |
| Obs. 3 | G | H | I |

# An observation has its own row

| | Var. 1 | Var. 2 | Var. 3 |
|---|---|---|---|
| Obs. 1 | A | B | C |
| Obs. 2 | D | E | F |
| Obs. 3 | G | H | I |

# An value has its own cell

| | Var. 1 | Var. 2 | Var. 3 |
|---|---|---|---|
| Obs. 1 | A | B | C |
| Obs. 2 | D | E | F |
| Obs. 3 | G | H | I |

tidyr
GATHER
SPREAD

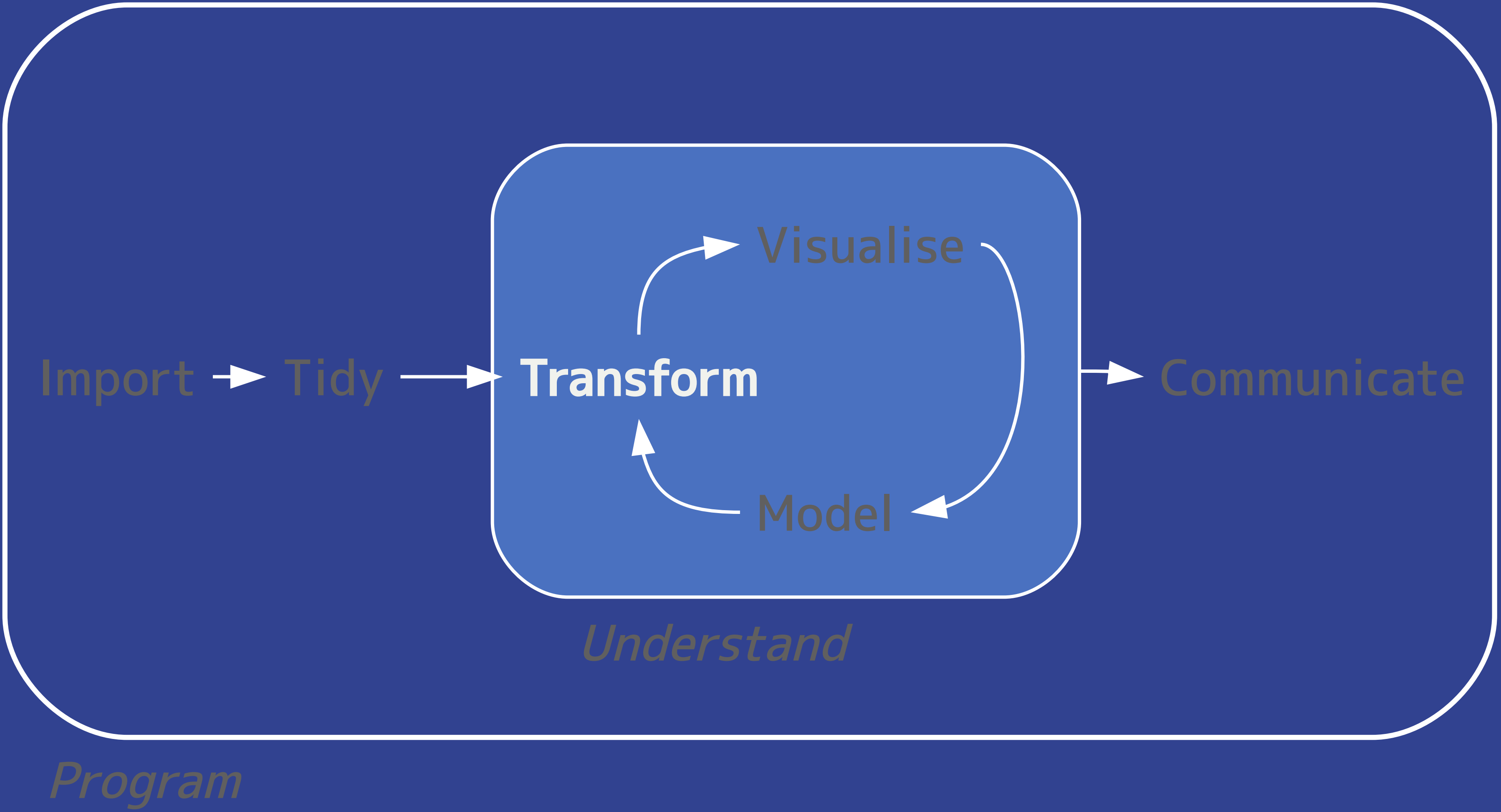
work by @allison_horst

# Not only tidy, but also tame

- Use synthetical names for column names
- Use consistent case: snake_case, camelCase, caterpillar.case
- Cast column type accordingly: <chr>, <dbl>, <lgl>, <date>, etc
- Treat <fct> carefully!
- Preferably turn implicit missing observation into explicit NA value

# Let’s do practice!

- Open ‘002_impor-tidy-data.Rmd’
- You have **15 minutes** to play with it
- Do not forget to push your works into GitHub!


Import
Tidy
Transform
Visualise
Model
Communicate
Understand
Program

Artwork by @allison_horst

## dplyr basic functions:

- filter() selects rows based on their values
- mutate() creates new variables
- select() picks columns by name
- summarise() calculates summary statistics
- arrange() sorts the rows

## tidyr basic functions:

- gather() wide-format >> long-format
- spread() long-format >> wide-format
- fill() fills value based on previous entry
- complete() turns implicit missing values into explicit

## Operators:

- ! (not)
- | (or)
- & (and)
- ==, !=
- <, <=, >, >=
- %in%
- is.na()

How can I

# chain?

1. diputar
2. dijilat
3. dicelupin
4. dimakan :D

1. putar(apa)
2. jilat(apa, berapa_kali)
3. celup(apa, ke)
4. makan(apa, output)

```
> oreo_putar ← putar(apa = “oreo”)
> oreo_jilat ← jilat(apa = oreo_putar,
                    berapa_kali = 2)
> oreo_celup ← celup(apa = oreo_jilat,
                    ke = “susu”)
> makan(apa = oreo_celup,
       output = “kenyang.perut”)
```

```
> oreo_putar ← putar(apa = “oreo”)
> oreo_jilat ← jilat(apa = oreo_putar,
                     berapa_kali = 2)
> oreo_celup ← celup(apa = oreo_jilat,
                     ke = “susu”)
> makan(apa = oreo_celup,
        output = “kenyang.perut”)
```

```
> makan(
    celup(
      jilat(
        putar(apa = “oreo”),
        berapa_kali = 2
        ),
      ke = “susu”
    ),
    output = “kenyang.perut”
  )
```

# magrittr

```
function(arg1, arg2, arg3, ... )

arg1 %>%
   function(arg2, arg3, ... )
```

---

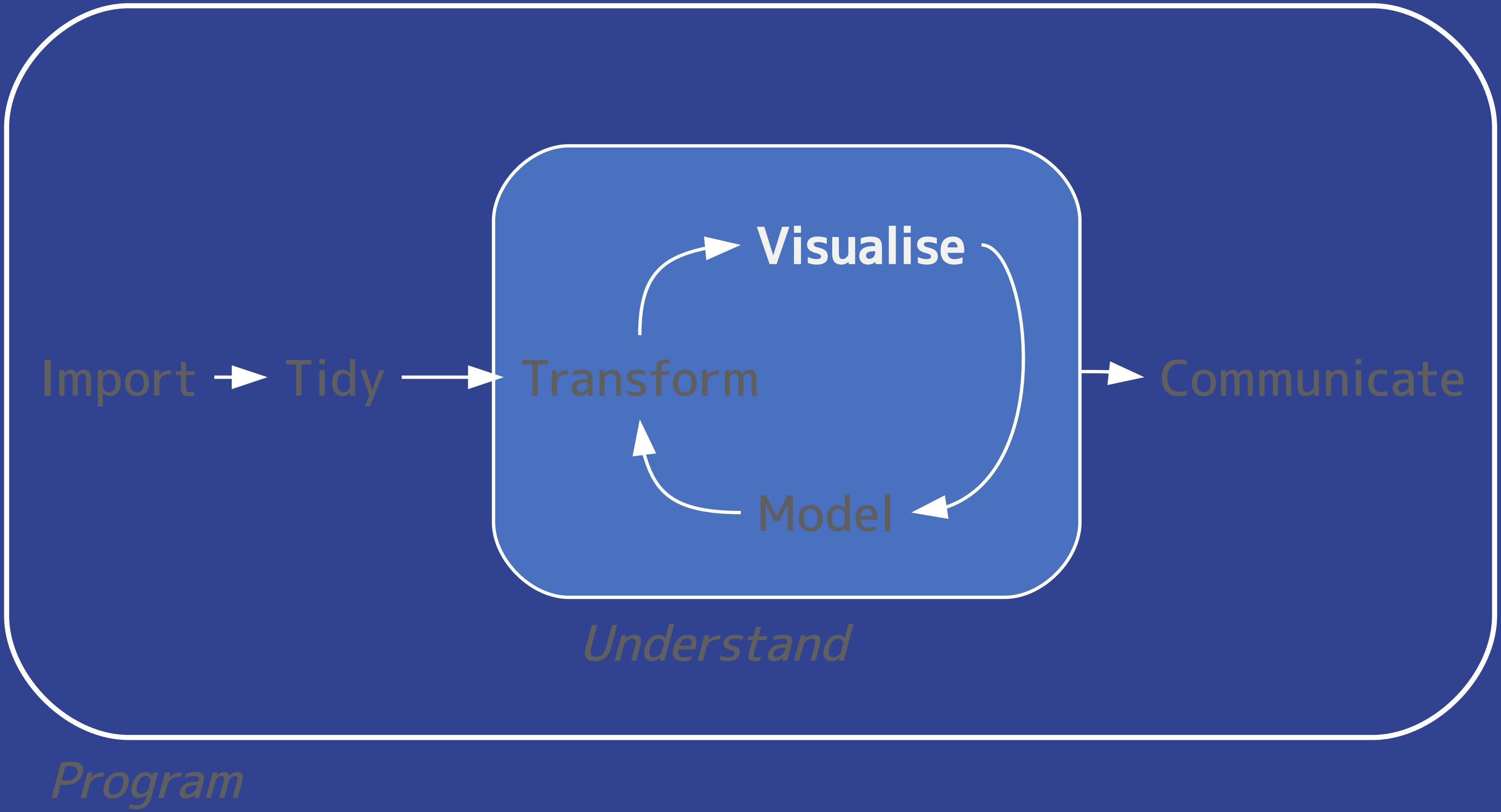
```
function(arg1, arg2, arg3, ... )

arg2 %>%
   function(arg1, arg2=.,arg3, ... )
```

C

```
> putar(apa = “oreo”) %>%
    jilat(berapa_kali = 2) %>%
    celup(ke = “susu”) %>%
    makan(output = “kenyang.perut”)
```

# Let’s do practice!

- Open ‘003_transformasi.Rmd’
- You have **30 minutes** to play with it
- Do not forget to push your works into GitHub!



Import

Tidy

Transform

**Visualise**

Model

Communicate

*Understand*

*Program*

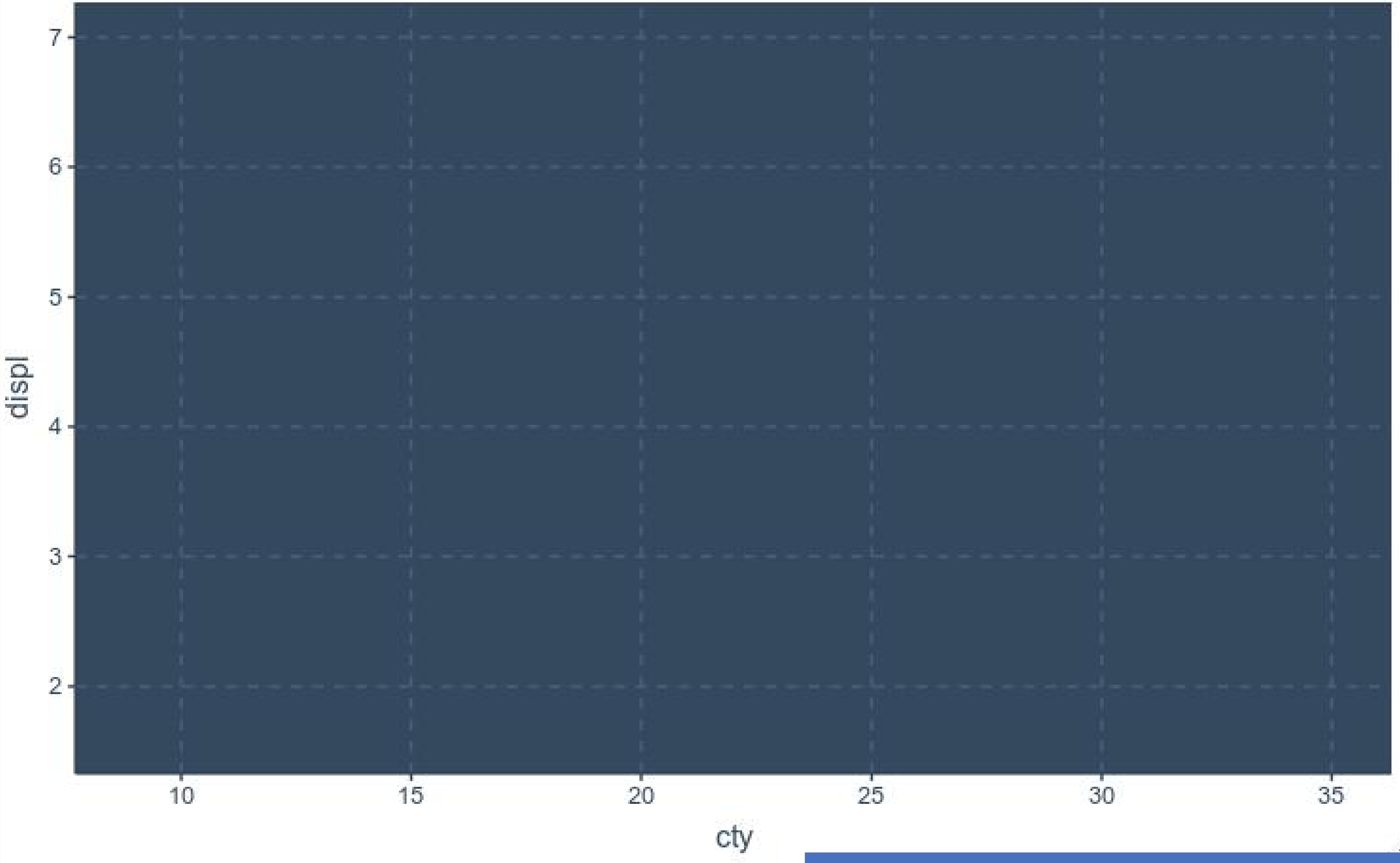
Artwork by @allison_horst

# ggplot2

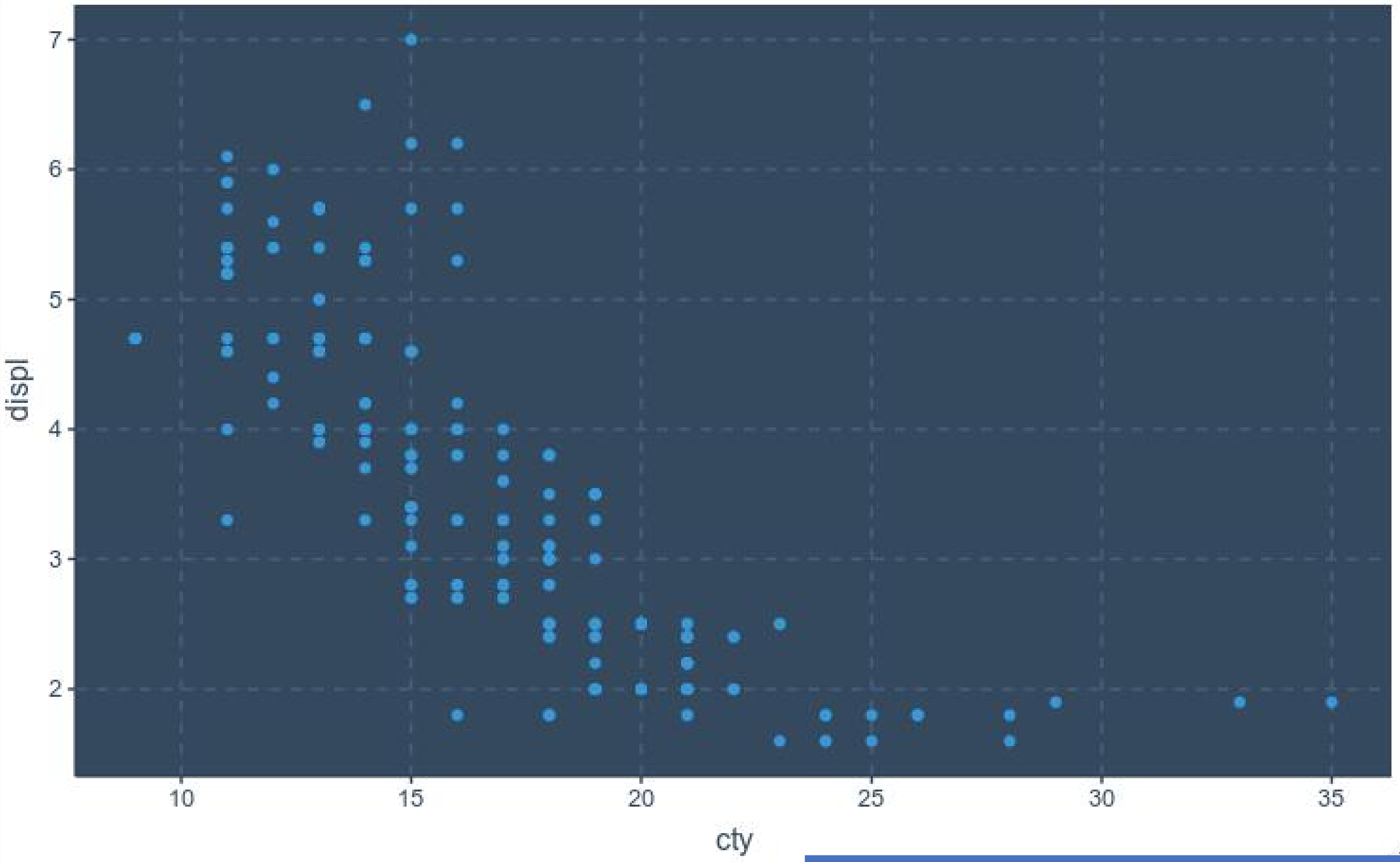
```
ggplot(data) +
   geom_X(mapping=aes(…)) +
   ...

ggplot(data, mapping=aes( ... )) +
   geom_X() +
   ...
```

---

```
data %>%
   ggplot(mapping=aes( ... )) +
   geom_X() +
   ...
```



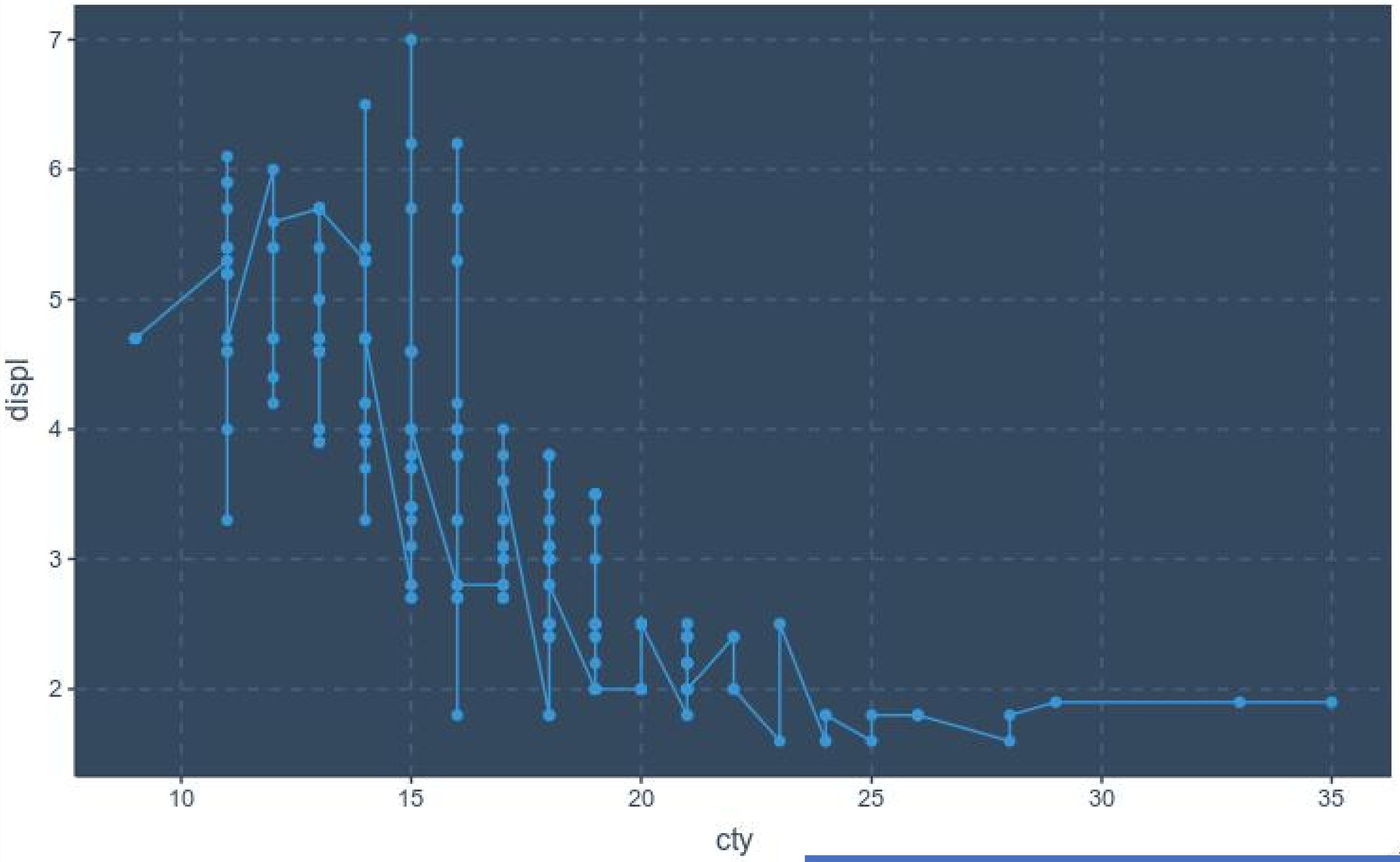
```
ggplot(mpg, aes(x = cty, y = displ))
```



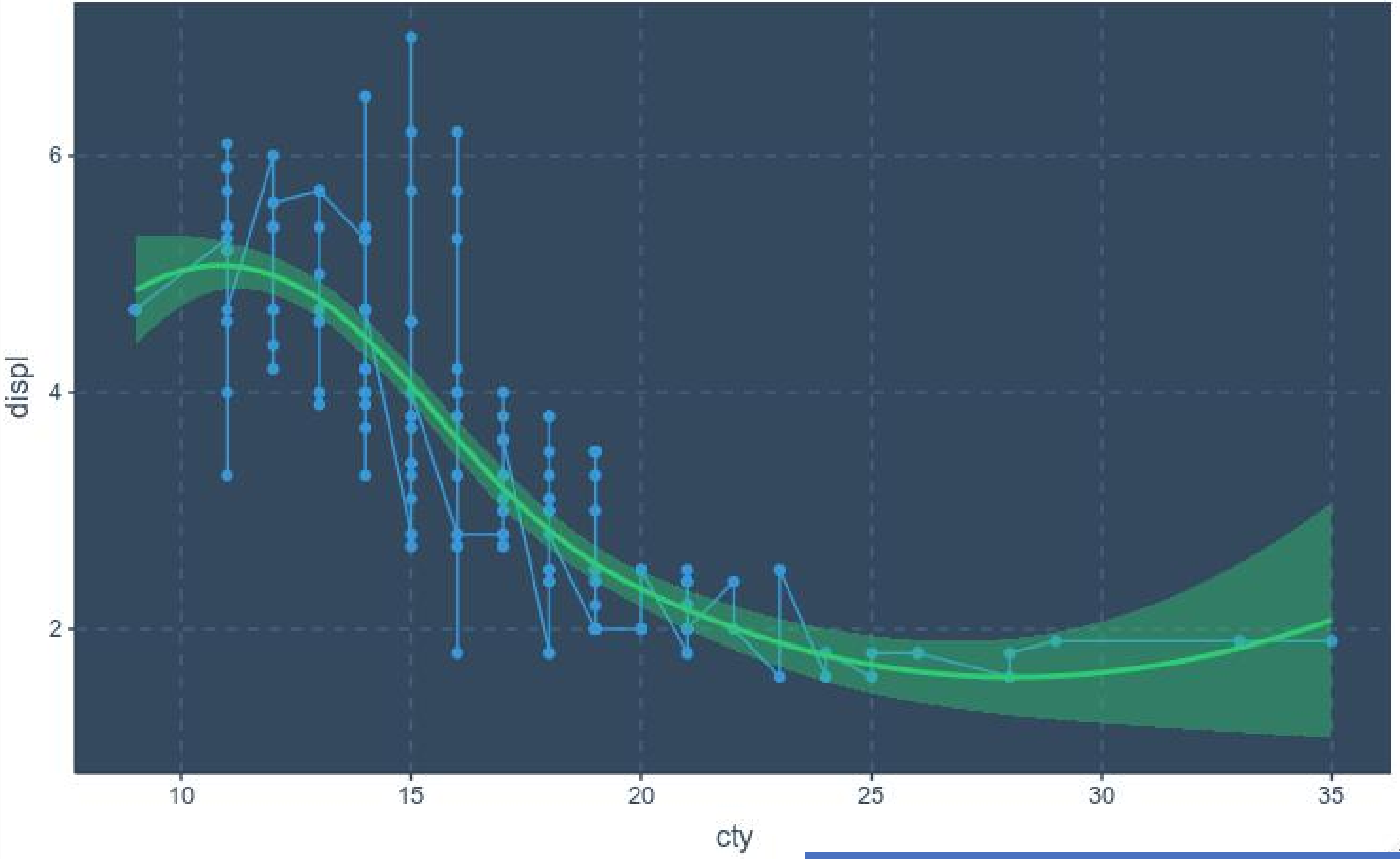
ggplot2

www.rstudio.com

```
ggplot(mpg, aes(x = cty, y = displ)) +
  geom_point()
```



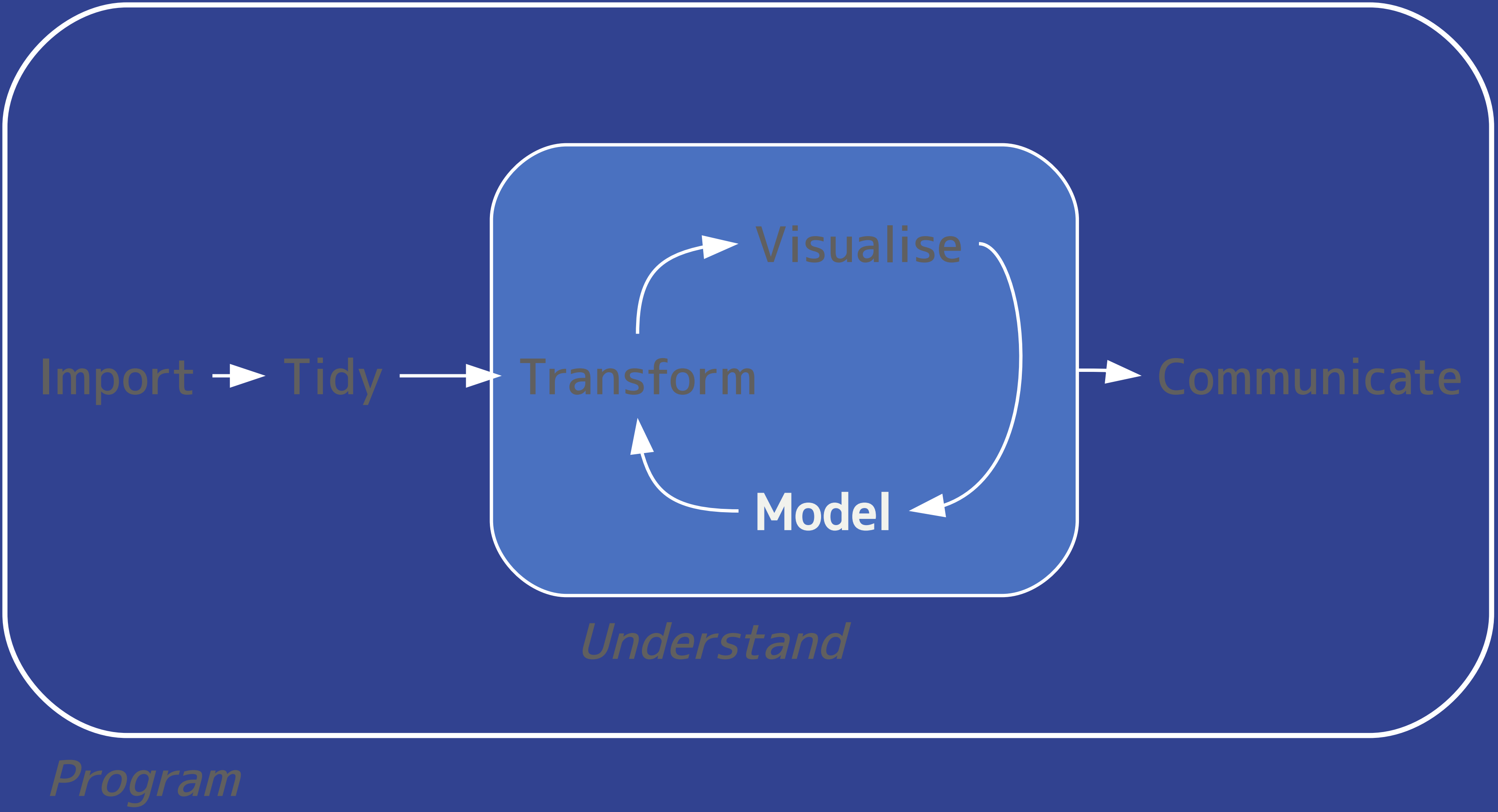
```
ggplot(mpg, aes(x = cty, y = displ)) +
  geom_point() +
  geom_line()
```



```
ggplot(mpg, aes(x = cty, y = displ)) +
  geom_point() +
  geom_line() +
  geom_smooth()
```

# Let’s do practice!

- Open ‘004_visualisasi.Rmd’
- You have **15 minutes** to play with it
- Do not forget to push your works into GitHub!



Import

Tidy

Transform

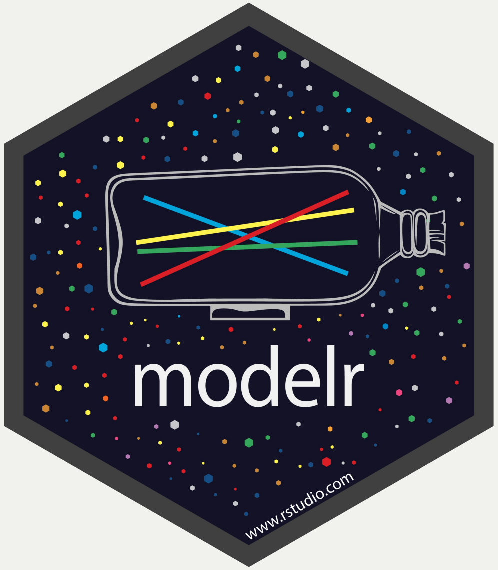
Visualise

**Model**

Communicate

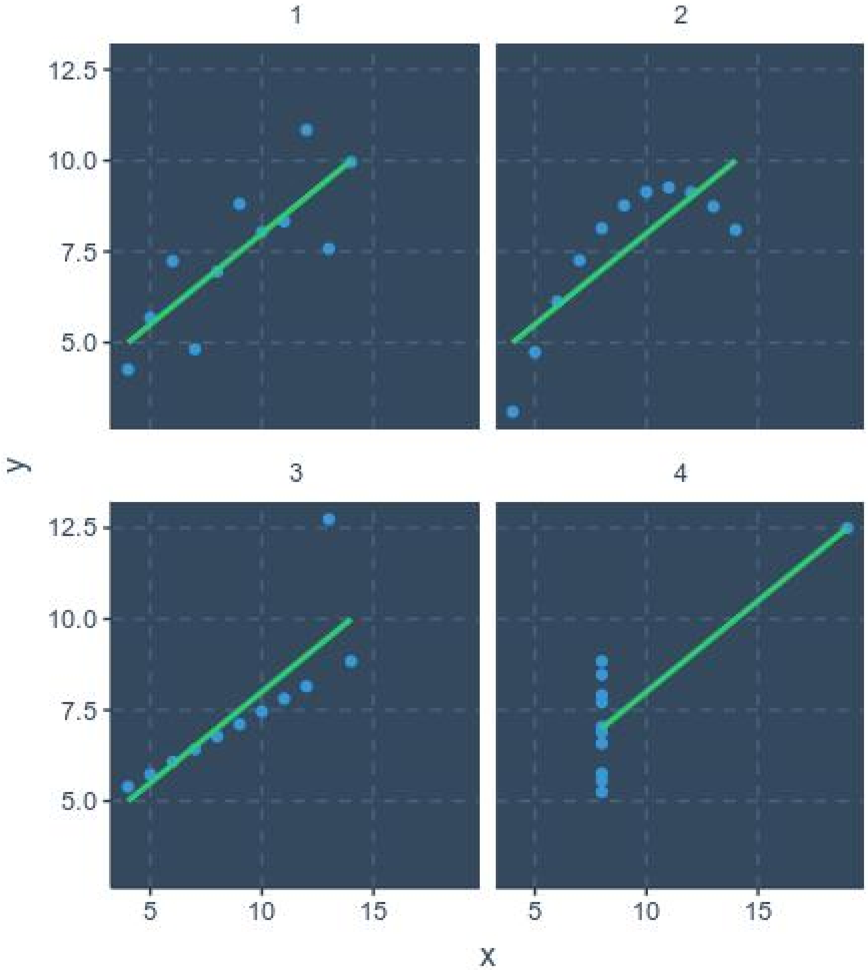
*Understand*

*Program*



A low dimensional description of a higher dimensional data set

Source: https://github.com/rstudio-education/arm-workshop-rsc2019



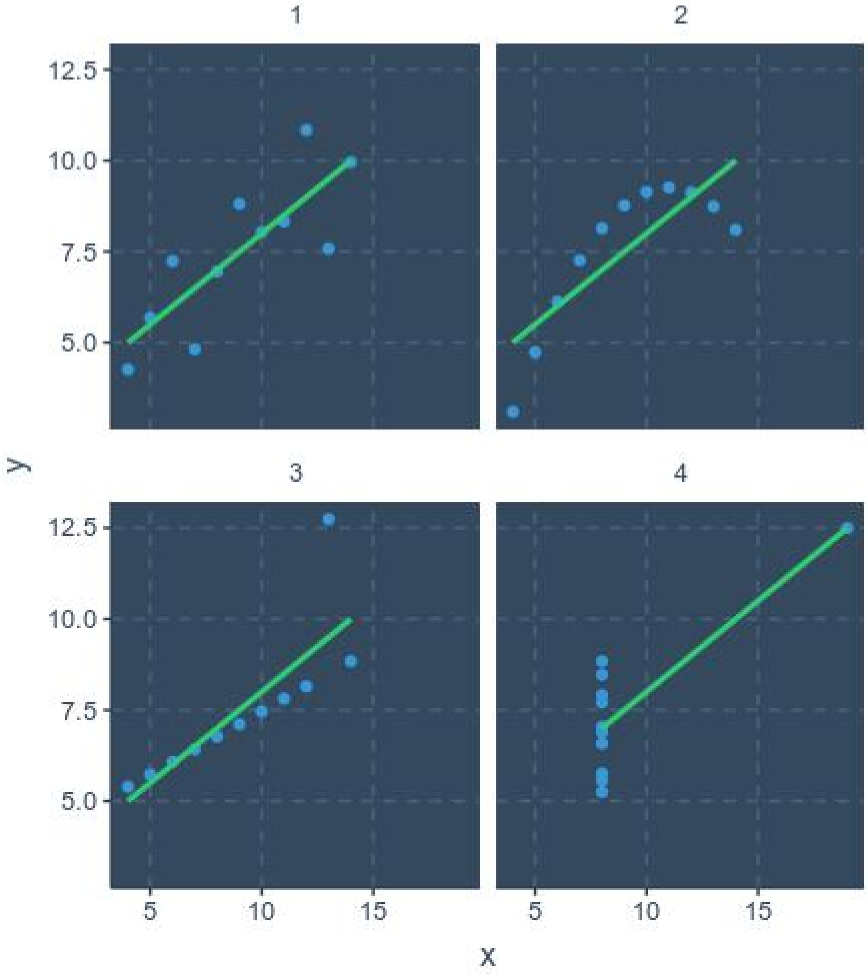
Outcome ~ Predictor/Explanatory

To predict

To explain

All models are wrong, but some are useful – George Box

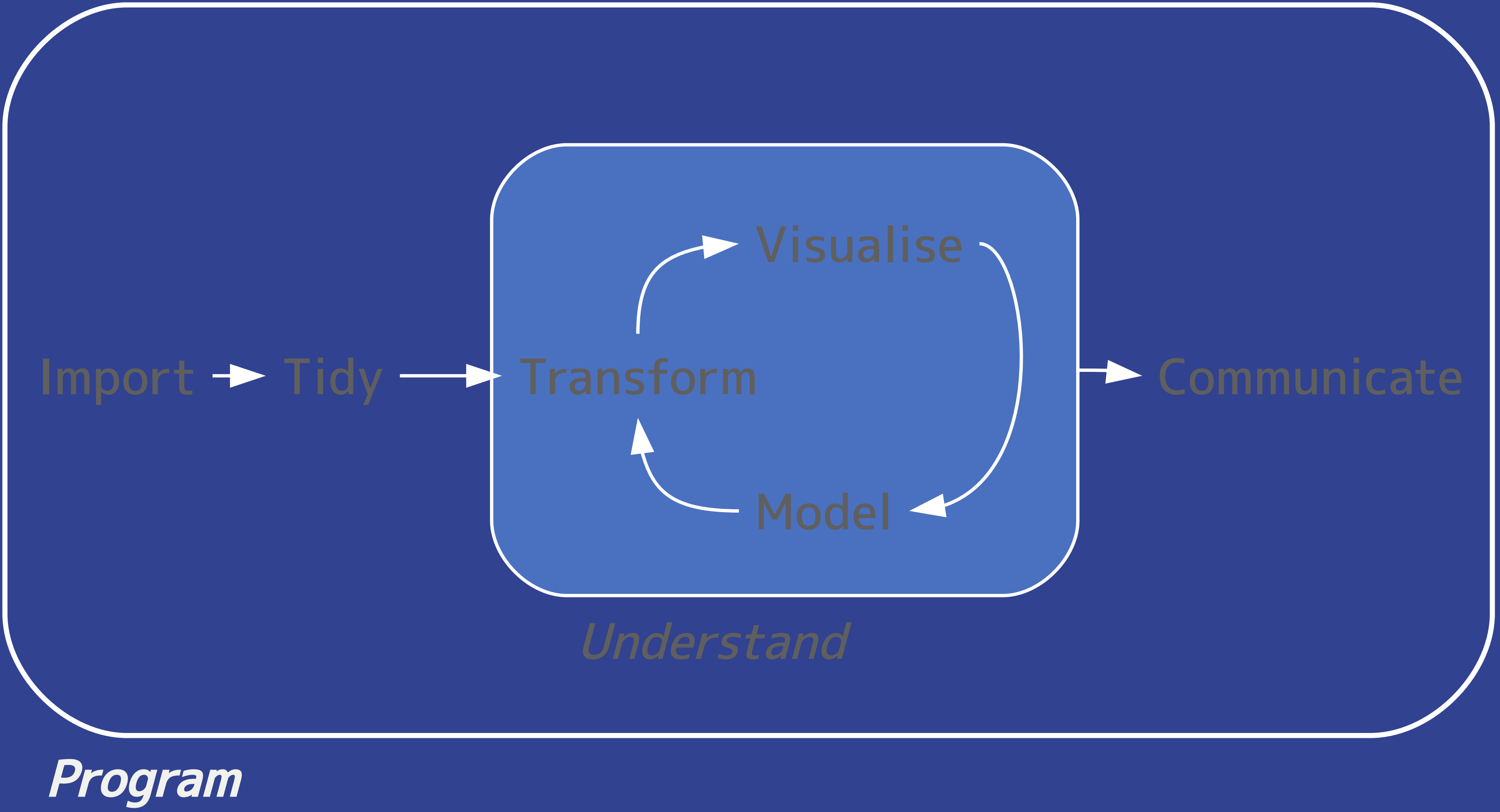
# Anscombe's Quartet



| | | |
|---|---|---|
| Mean of x | 9 | exact |
| Sample variance of x | 11 | exact |
| Mean of y | 7.50 | to 2 decimal places |
| Sample variance of y | 4.125 | ±0.003 |
| Correlation between x and y | 0.816 | to 3 decimal places |
| Linear regression line | y = 3.00 + 0.500x | to 2 and 3 decimal places, respectively |
| $R^2$ | 0.67 | to 3 decimal places |

# Let’s do practice!

- Open ‘005_model.Rmd’
- You have **40 minutes** to play with it
- Do not forget to push your works into GitHub!



Import

Tidy

Transform

Visualise

Model

Communicate

*Understand*

***Program***

```
x ← something
for (i in seq_along(x) {
  function(x[[i]])
}
```

```
x ← something
*apply(x, function)
```

```
x ← something
map_*(x, function)
```

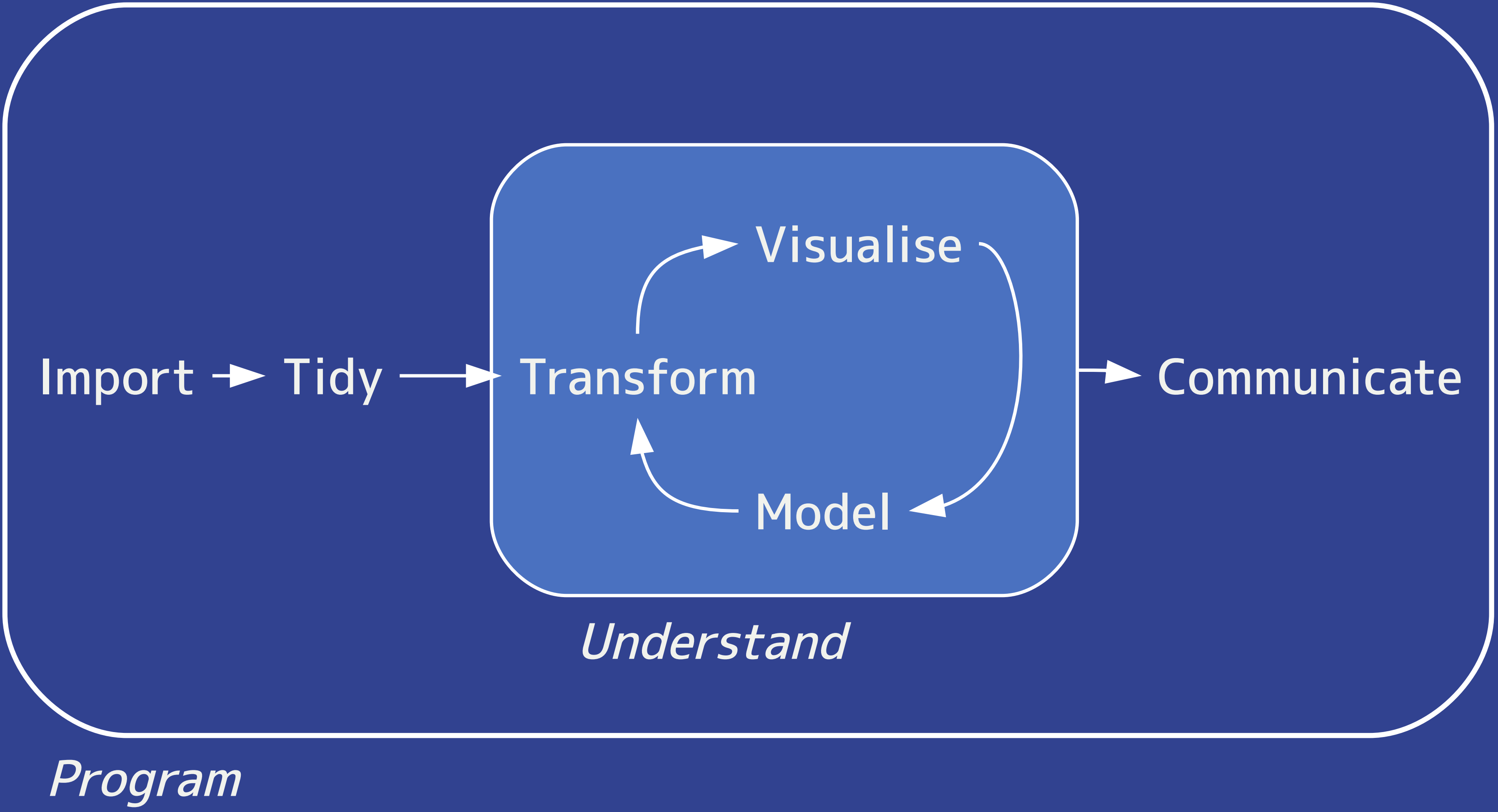
# purrr

# Let’s do practice!

- Open ‘006_iterasi.Rmd’
- You have **20 minutes** to play with it
- Do not forget to push your works into GitHub!

# Congrats!

Artwork by @allison_horst


Import
Tidy
Transform
Visualise
Model
Communicate
Understand
Program

```
> contact_me(
    name      = “Muhammad Aswan Syahputra”,
    email     = aswansyahputra@sensolution.id,
    Phone     = +62 822 3465 3816,
    twitter   = @aswansyahputra_
  )
> ...
```