Exploratory Data Analysis With

Workshop R Bulan Juli

Pembicara

Mohammad Rosidi

- **Course Producer** Technaut Education
- **Asisten Peneliti** Teknik Lingkungan ITB bidang keahlian Teknologi Management Lingkungan
- **Penulis buku**
  - Metode Numerik Menggunakan R untuk Teknik Lingkungan
  - Analisis Statistika Menggunakan R Commander

rosidi.moh

@rosidi2610

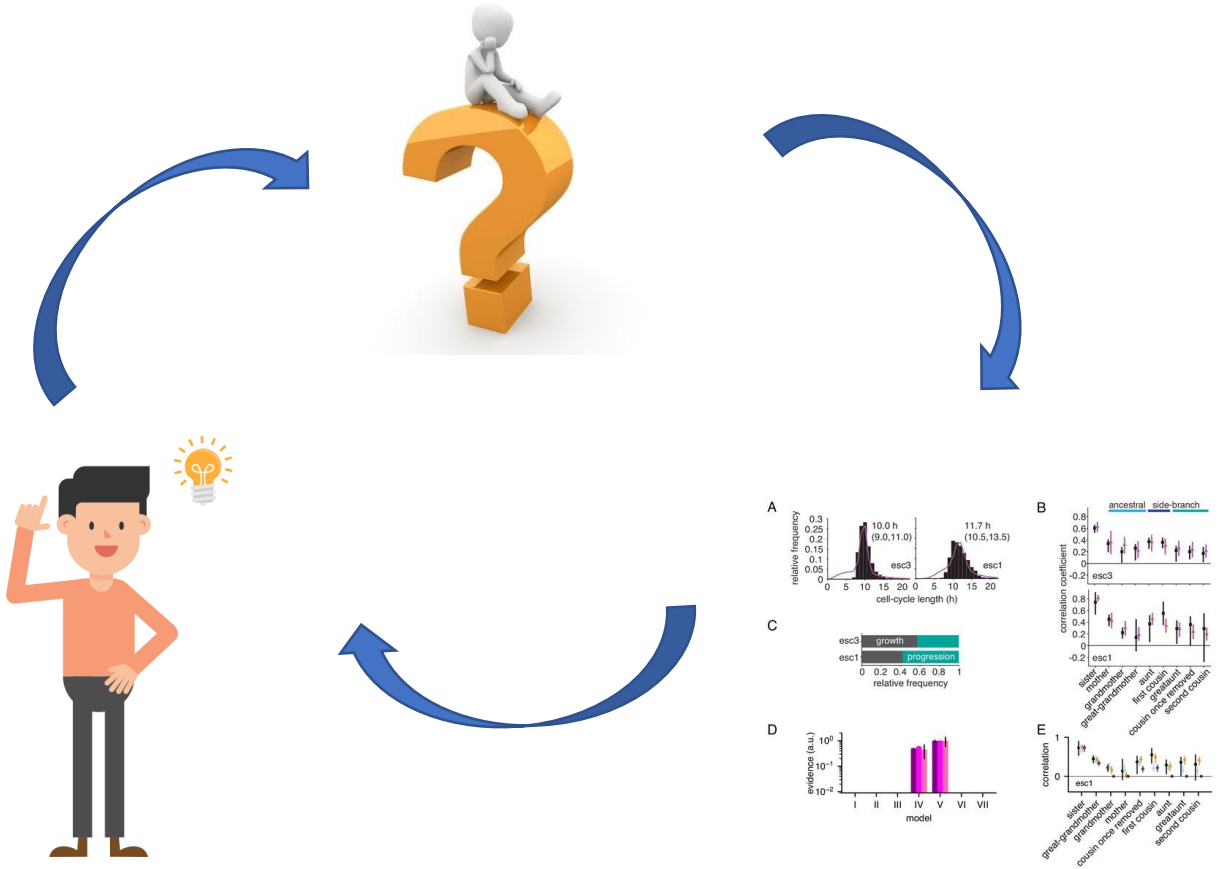
mohrosidi

Apa yang akan dipelajari?

1. Konsep dasar EDA
2. Paket yang digunakan
3. Studi Kasus

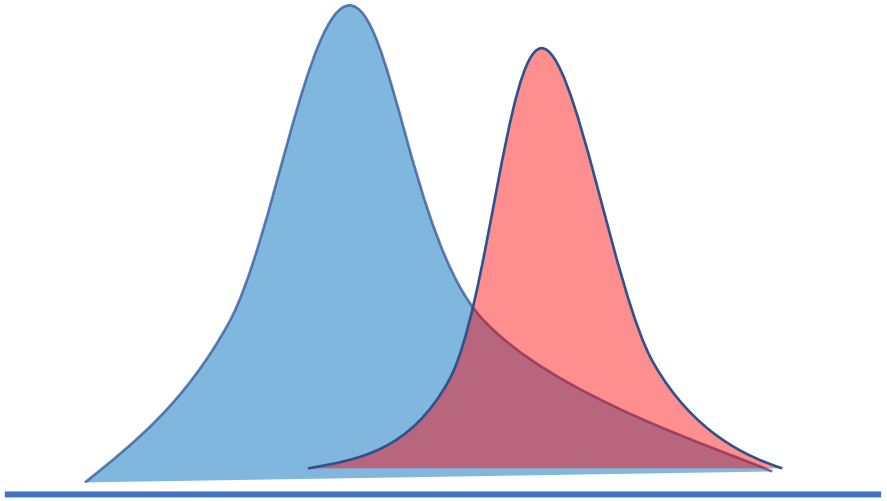
Konsep Dasar EDA

Apa itu EDA?

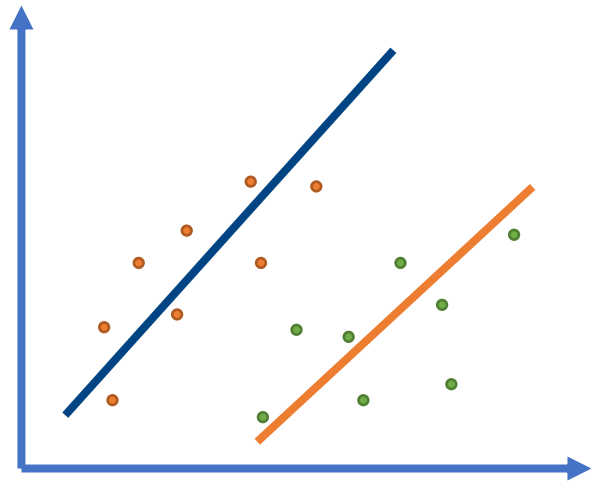


Proses iteratif mengenal data

Apa yang hendak di jawab?

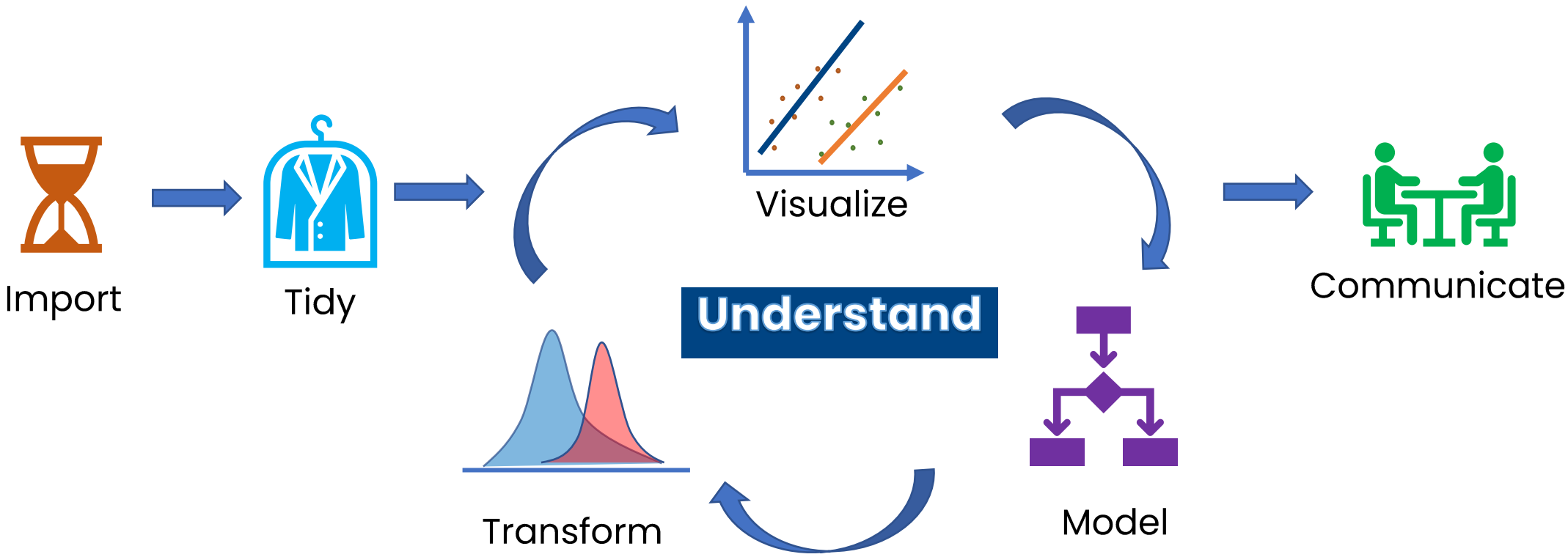


Variasi



Kovarian

Data Science Life Cycle



Tidy Up Your Data First

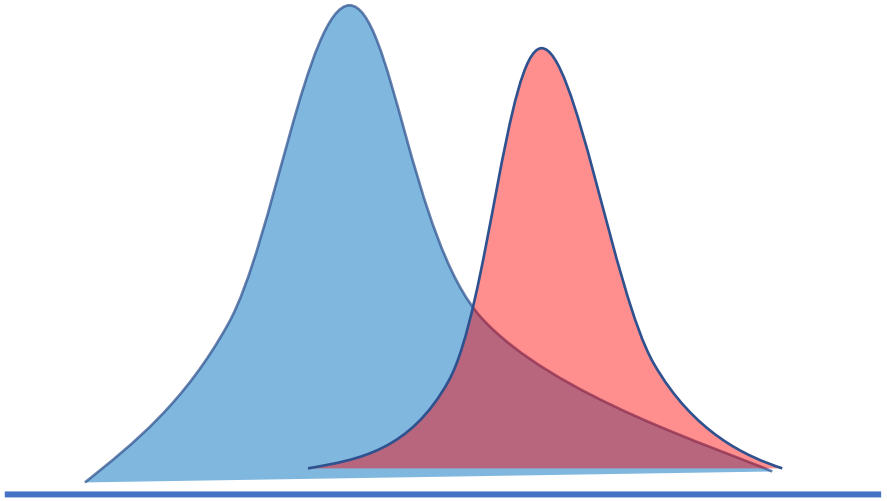
	Name	Gender	Age
1	Phil	Male	54
2	May	Female	46
3	Mack	NA	31

Setiap **variabel** memiliki kolomnya tersendiri

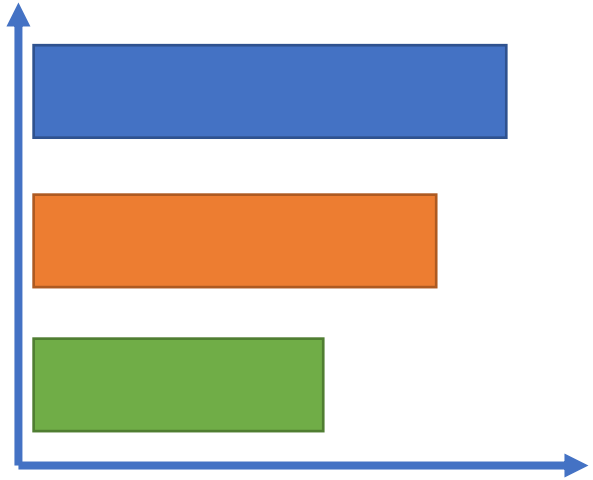
Setiap **observasi** data memiliki barisnya sendiri

Setiap **nilai** data memiliki selnya sendiri

EDA : Visualisasi Distribusi

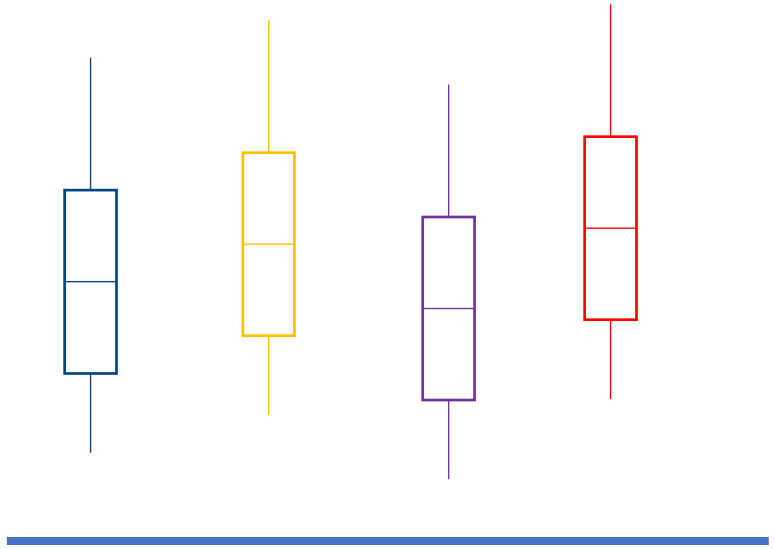


Numerik/kontinu

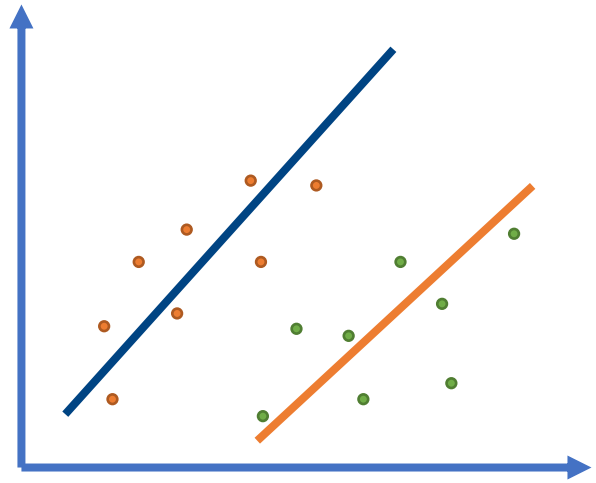


Kategori

EDA: Visualisasi Kovarian



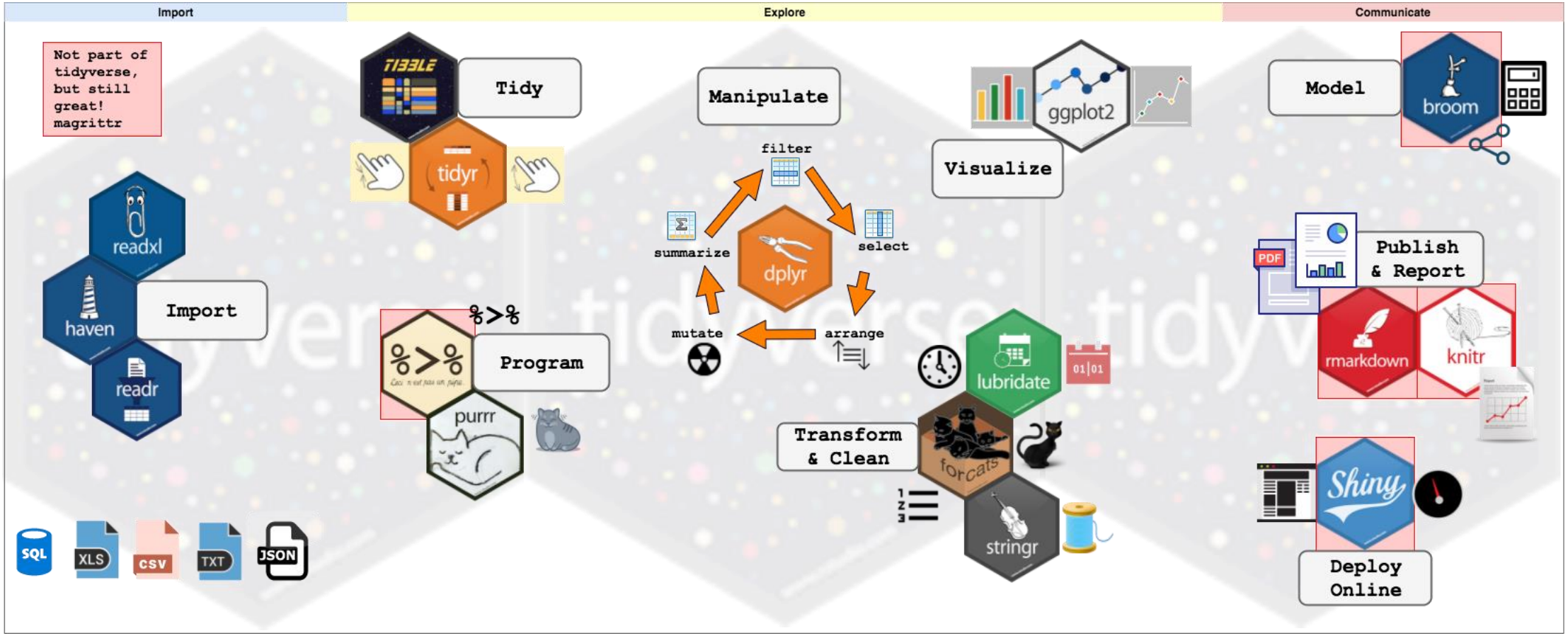
Kategori vs Numerik



Numerik vs Numerik

Paket EDA

Tidyverse


Import
Explore
Communicate
Not part of tidyverse, but still great! magrittr
tibble
Tidy
tidyr
readxl
haven
readr
Import
%>%
Program
purrr
Manipulate
filter
select
dplyr
arrange
mutate
summarize
ggplot2
Visualize
Model
broom
Publish & Report
PDF
rmarkdown
knitr
lubridate
01|01
Transform & Clean
forcats
stringr
Shiny
Deploy Online
SQL
XLS
csv
TXT
JSON

Perbandingan Sintaks

Intermediate Step

```
one <- first(data)
two <- second(one)
result <- third(two)
```

Nesting Function

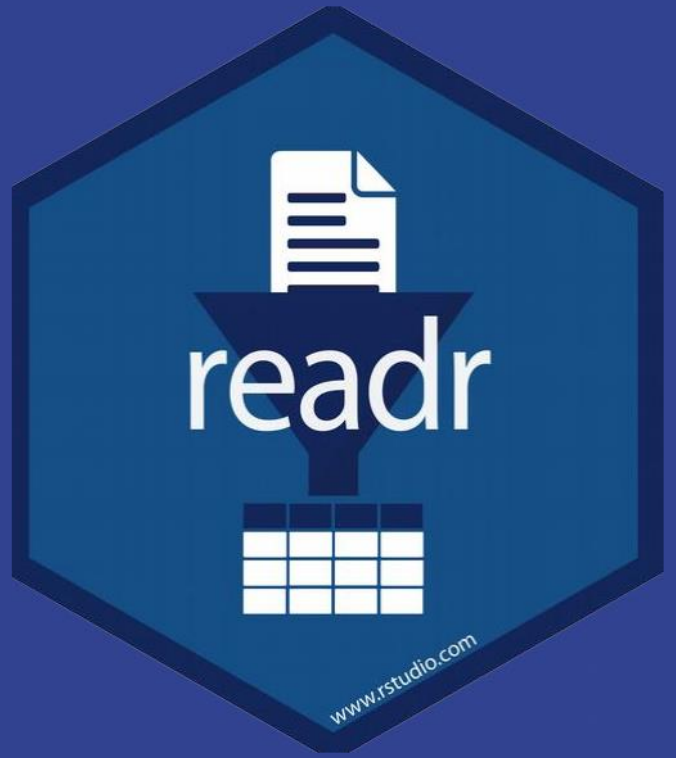
```
result <-third(
      second(
            first(data)
      )
)
```

Pipe %>%

```
result <- data %>%
      first() %>%
      second() %>%
      third()
```

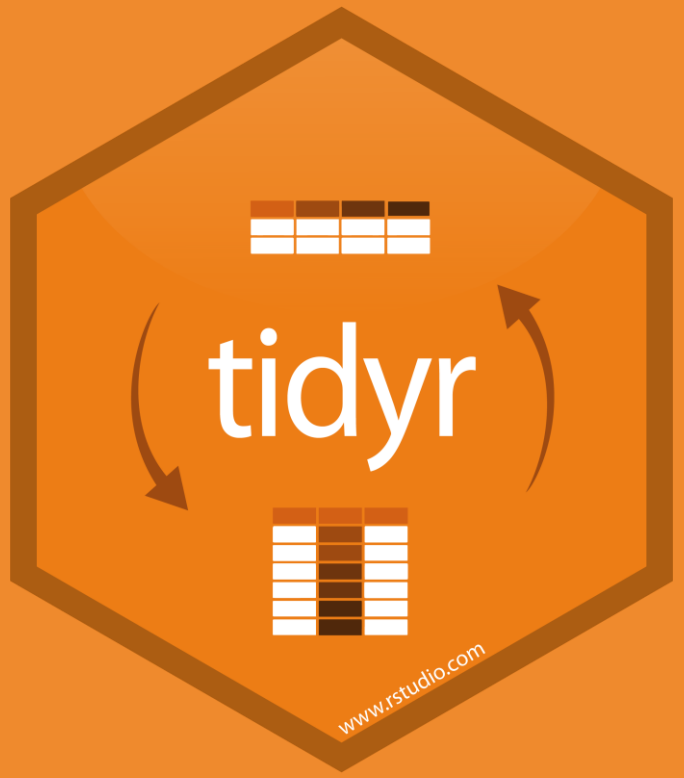
Readr

- **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



Tidyr

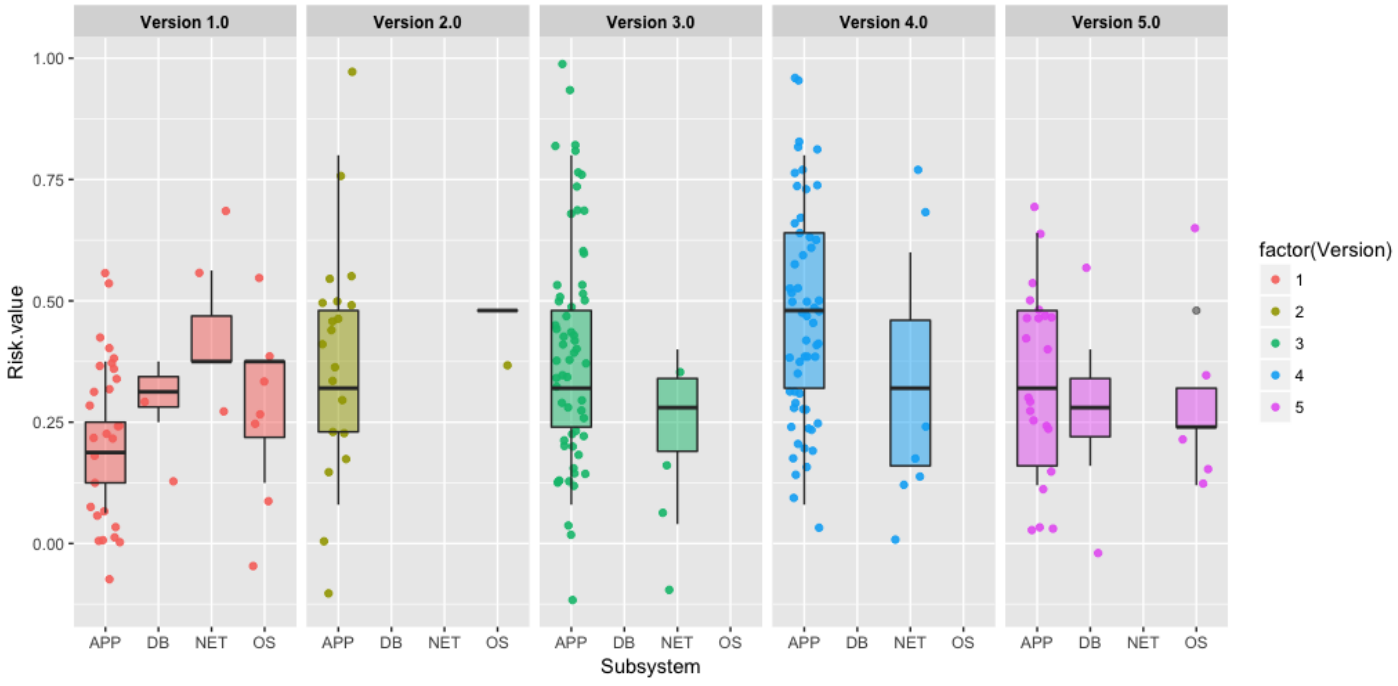
- **pivot_wide()**: long format → wide format
- **Pivot_long()**: wide format →long format
- **unite()**: menggabungkan beberapa kolom menjadi satu
- **separate()**: memisahkan sebuah kolom menjadi beberapa kolom

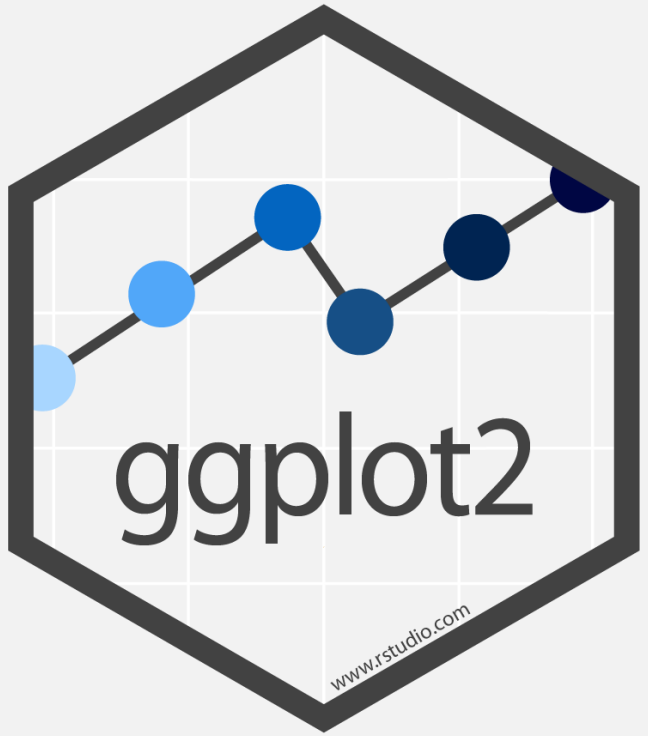


dplyr

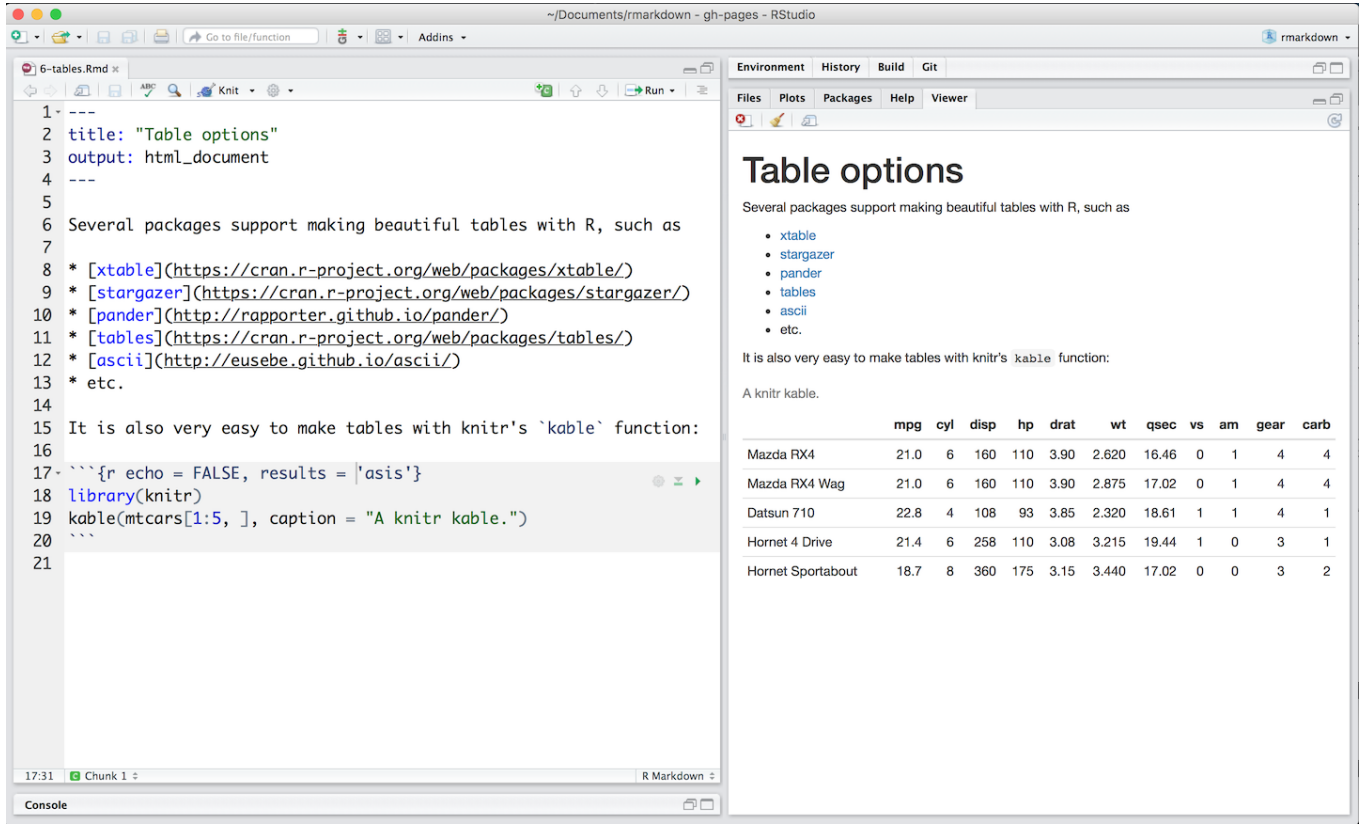
- **select()**: memilih kolom berdasarkan nama
- **filter()**: seleksi baris berdasarkan nilai pada sebuah kolom
- **mutate()**: membuat variabel baru
- **group_by()**: mengelompokkan data berdasarkan nilai variabel
- **summarise()**: menghitung ringkasan data
- **arrange()**: mengurutkan baris

ggplot


Version 1.0
Version 2.0
Version 3.0
Version 4.0
Version 5.0
Risk.value
1.00
0.75
0.50
0.25
0.00
APP
DB
NET
OS
Subsystem
factor(Version)
1
2
3
4
5


ggplot2
www.rstudio.com

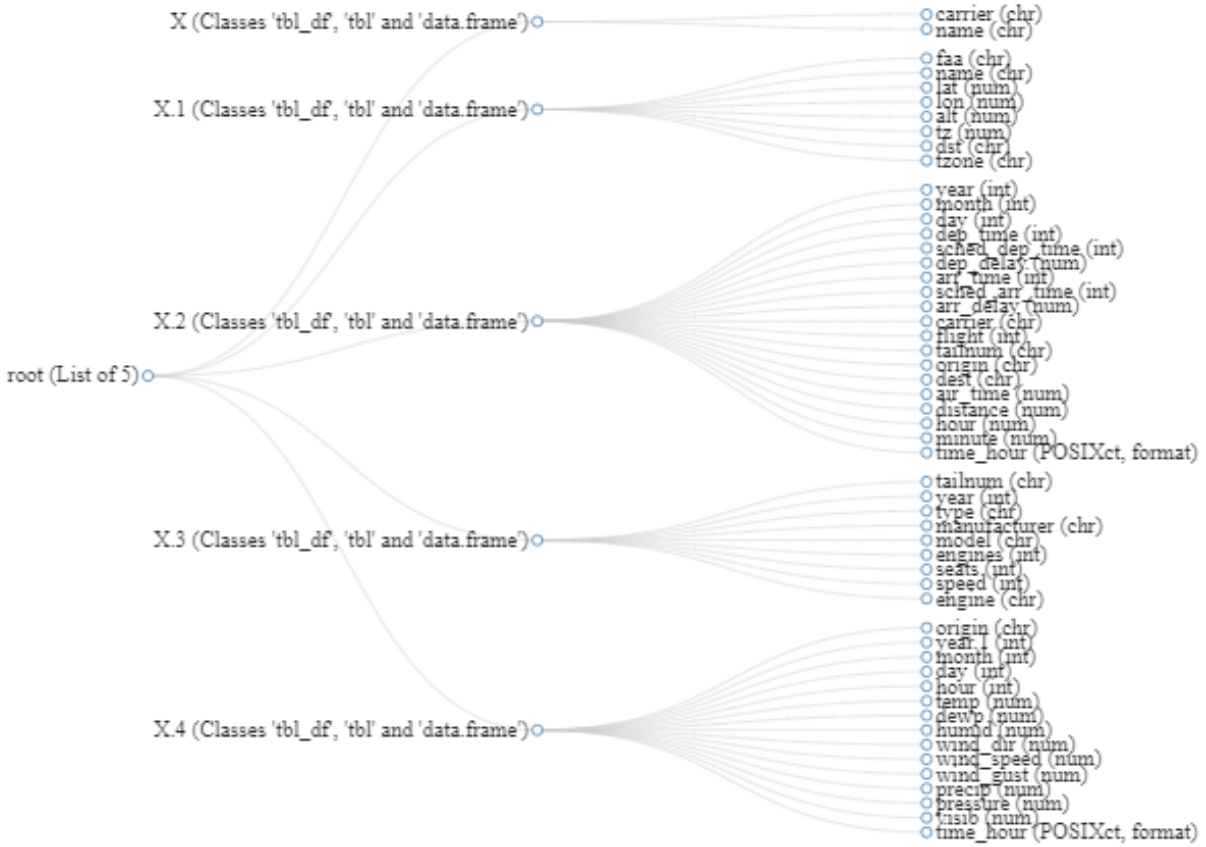
Rmarkdown


~/Documents/rmarkdown - gh-pages - RStudio
6-tables.Rmd
Knit
Run
---
title: "Table options"
output: html_document

Several packages support making beautiful tables with R, such as
* [xtable](https://cran.r-project.org/web/packages/xtable/)
* [stargazer](https://cran.r-project.org/web/packages/stargazer/)
* [pander](http://rapporter.github.io/pander/)
* [tables](https://cran.r-project.org/web/packages/tables/)
* [ascii](http://eusebe.github.io/ascii/)
* etc.
It is also very easy to make tables with knitr's `kable` function:
```{r echo = FALSE, results = 'asis'}
library(knitr)
kable(mtcars[1:5, ], caption = "A knitr kable.")
```
17:31
Chunk 1
R Markdown
Console
Environment
History
Build
Git
Files
Plots
Packages
Help
Viewer
Table options
Several packages support making beautiful tables with R, such as
xtable
stargazer
pander
tables
ascii
etc.
It is also very easy to make tables with knitr's kable function:
A knitr kable.
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

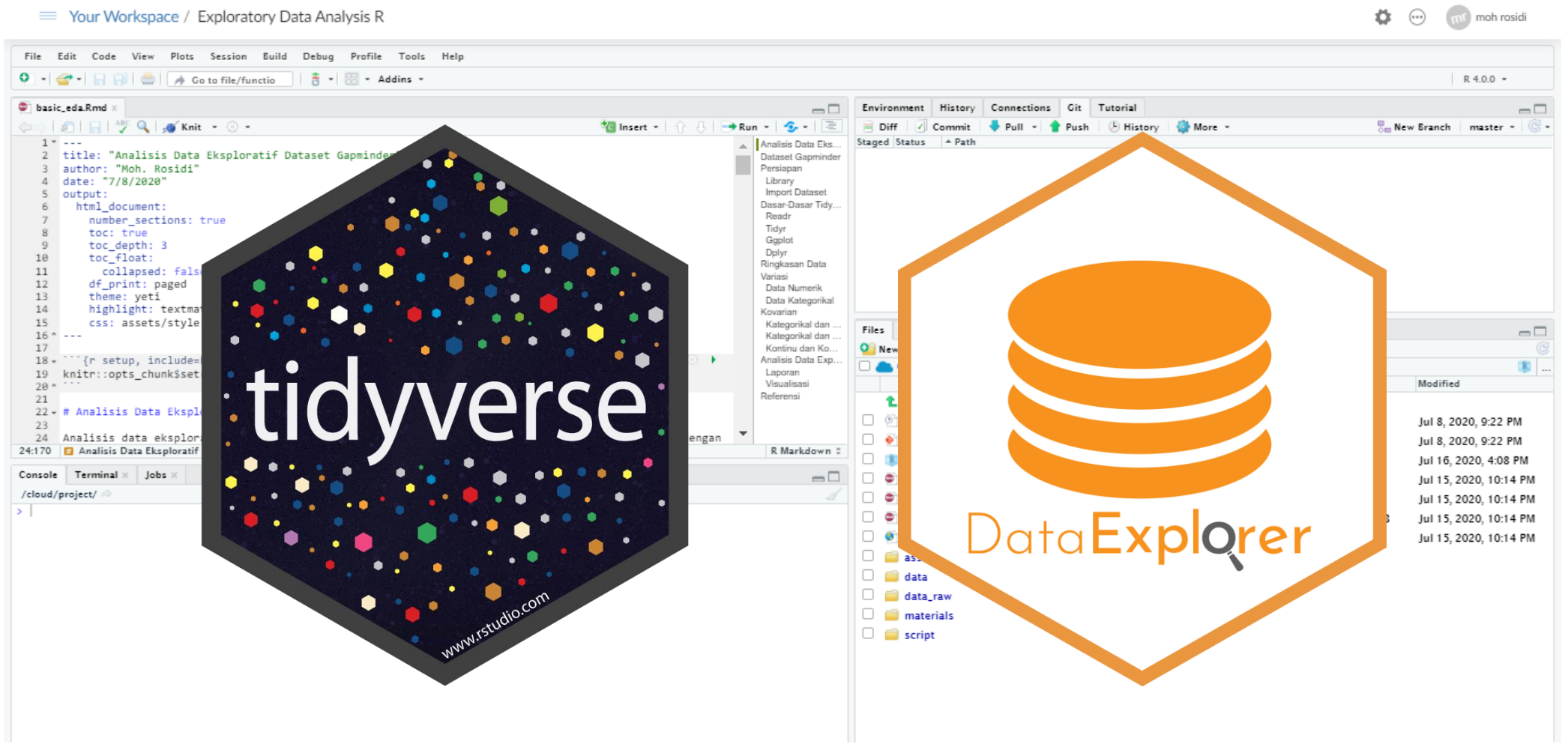

rmarkdown
www.rstudio.com

DataExplorer



DataExplorer

Praktek


Your Workspace / Exploratory Data Analysis R
moh rosidi
File Edit Code View Plots Session Build Debug Profile Tools Help
basic_eda.Rmd
Environment History Connections Git Tutorial
Console Terminal Jobs
/cloud/project/
tidyverse
www.rstudio.com
DataExplorer
data
data_raw
materials
script