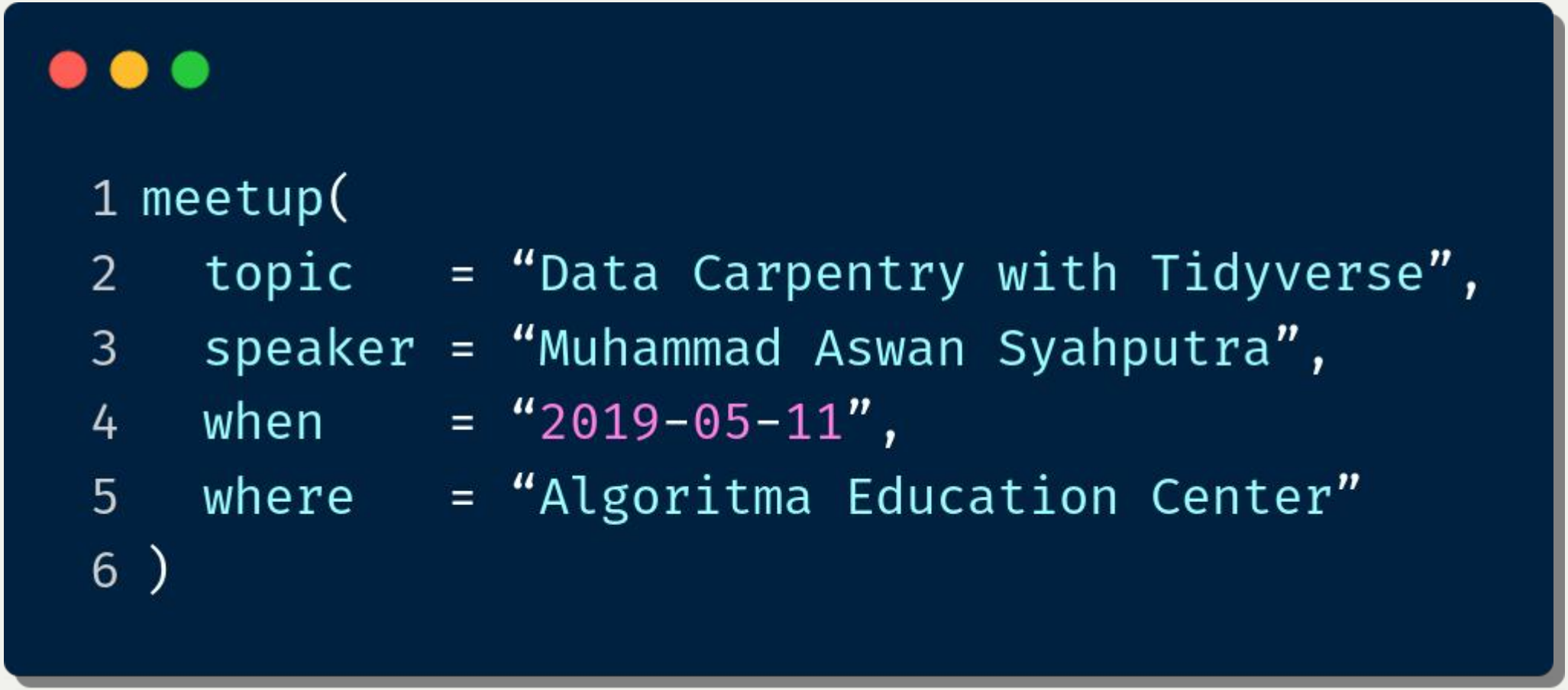

meetup(
topic = “Data Carpentry with Tidyverse”,
speaker = “Muhammad Aswan Syahputra”,
when = “2019-05-11”,
where = “Algoritma Education Center”
)

- Sensory Scientist @ Sensolution.ID
- Trainer @ R-Academy Telkom University and The Datanomics Institute (TDI)
- Initiator of Komunitas R Indonesia
- Pkgs: sensehubr, nusandata, bandungjuara, prakiraan, etc
- Shinyapps: sensehub, thermostats, aquastats, bcrp, bandungjuara, etc

aswansyahputra @aswansyahputra_

t.me/GNURIndonesia

indo-r

@r_indonesia_

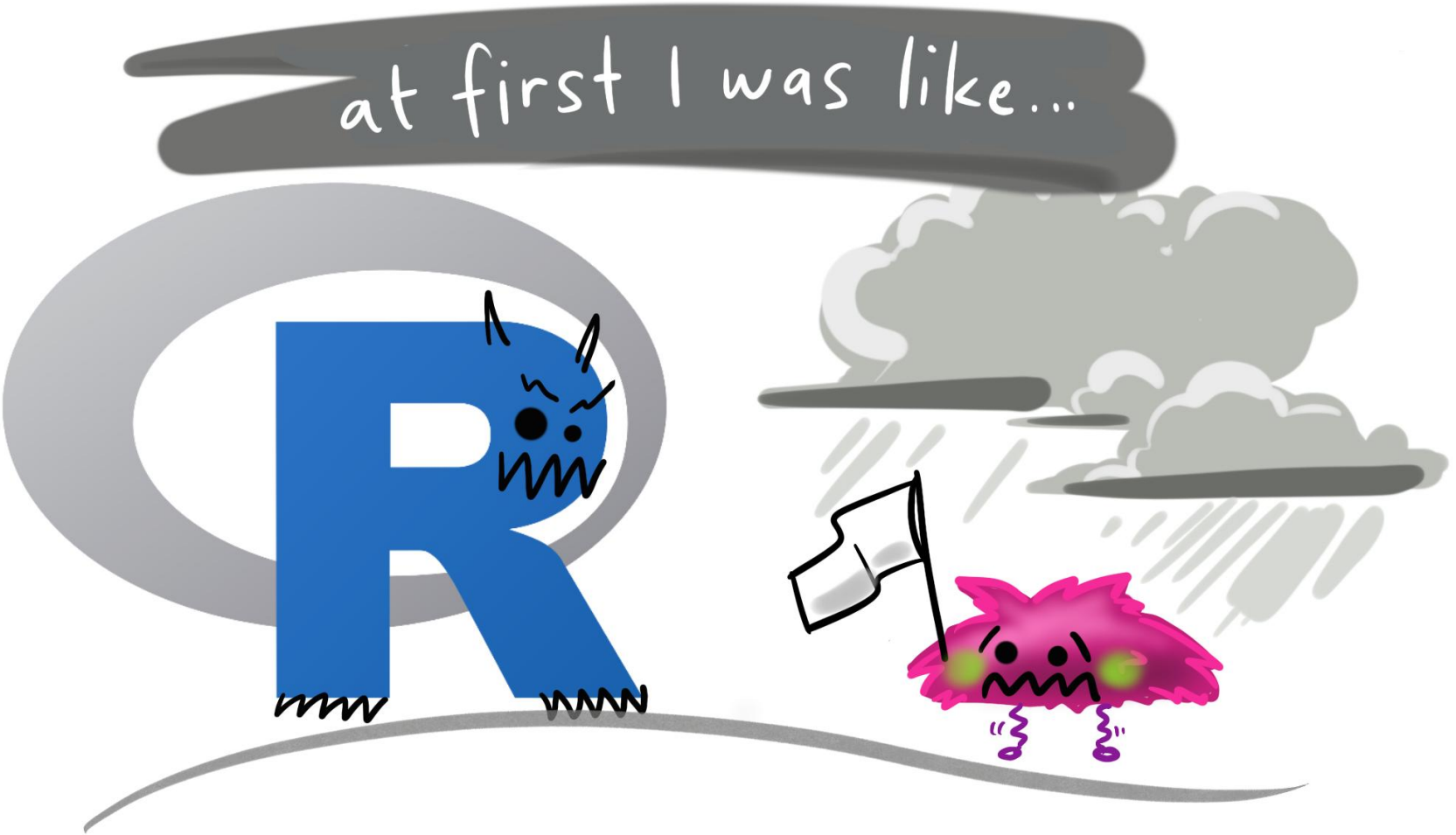
www.r-indonesia.id

Komunitas
R Indonesia


R Indones a
www.r- ndones a. d

Know your neighbour!

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



Artwork by @allison_horst

Artwork by @allison_horst

Data Carpentry?

It’s so relatable,

is it not?

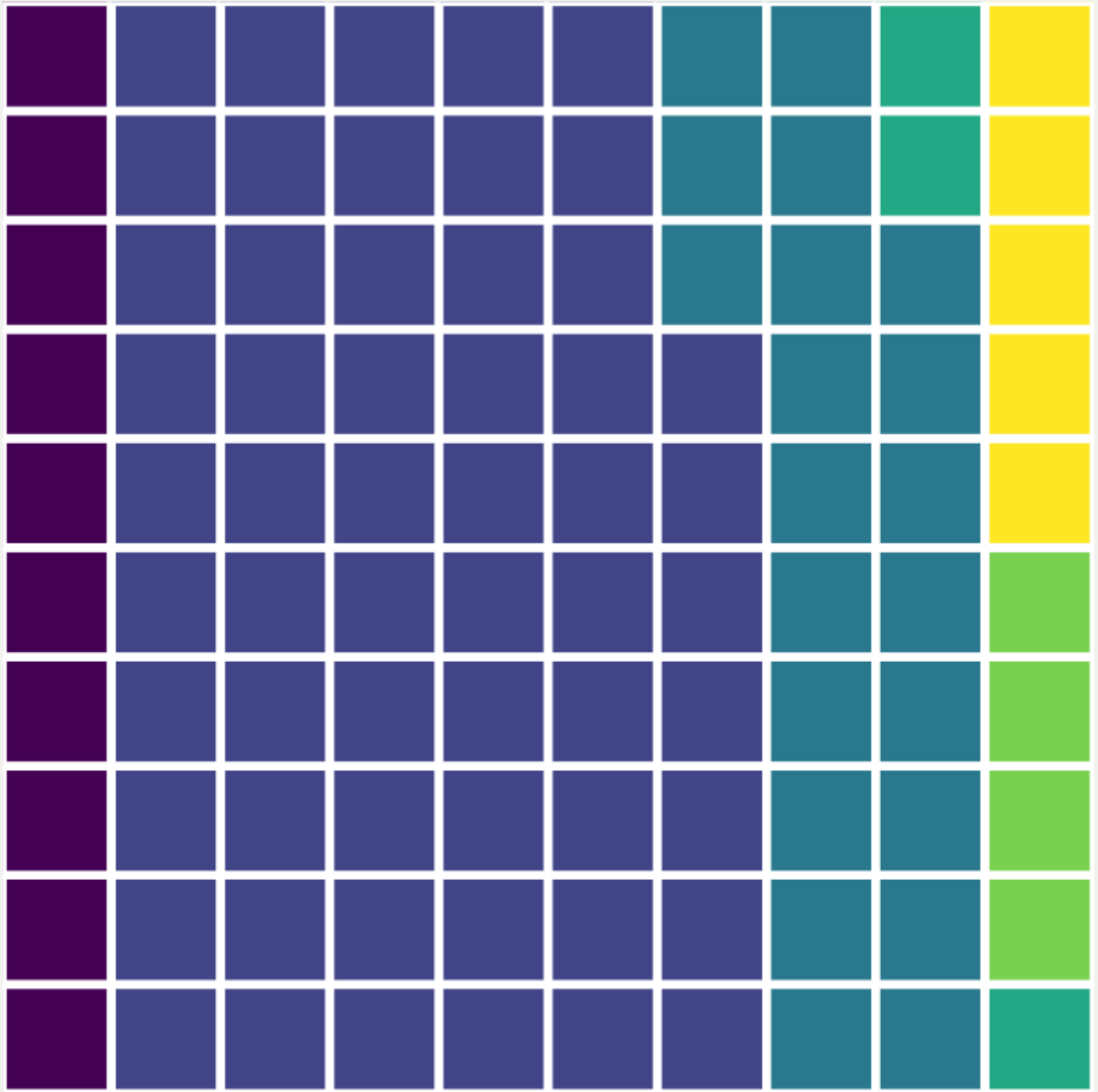
Source: Forbes, 2016

What data scientists spend the most time doing?

- Building training sets
- Cleaning and organizing data
- Collecting data sets
- Mining data for pattern
- Refining algorithms
- Other

What the least enjoyable part of data science?

- Building training sets
- Cleaning and organizing data
- Collecting data sets
- Mining data for pattern
- Refining algorithms
- Other



Source: Forbes, 2016

“

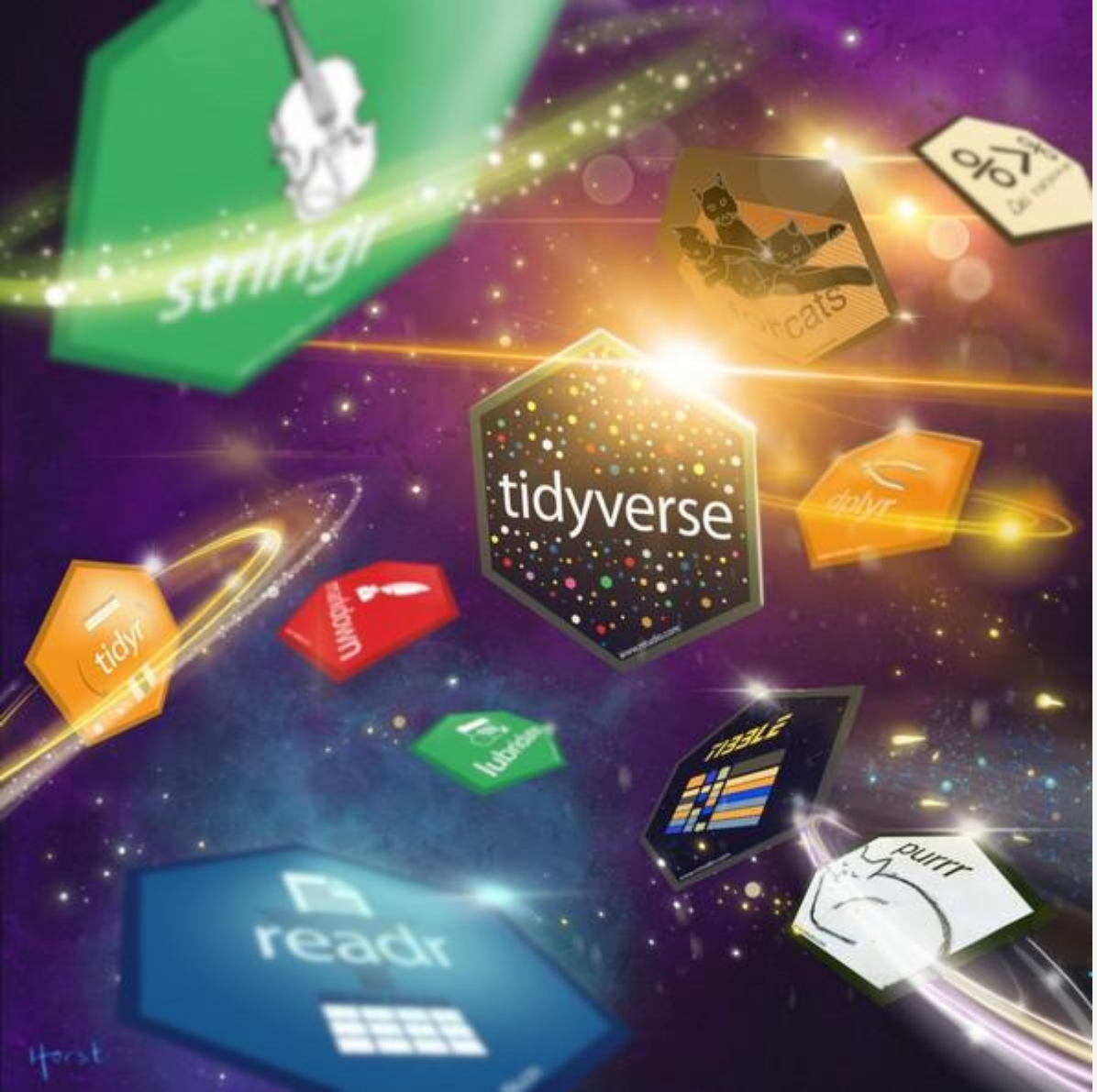
Do not underestimate

DATA PREPROCESSING

”

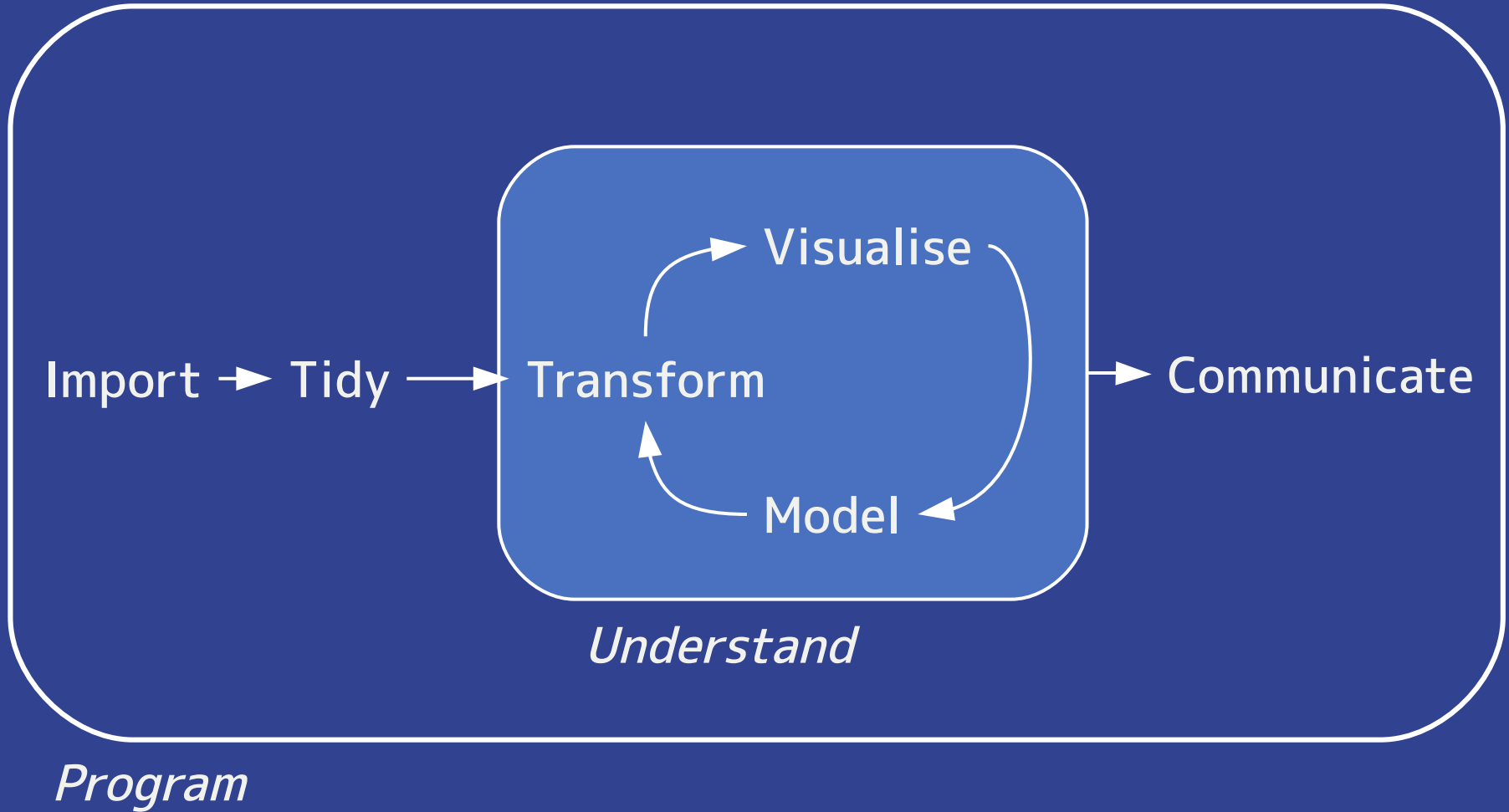
“is not a single process
but a thousand of
little skills and techniques”

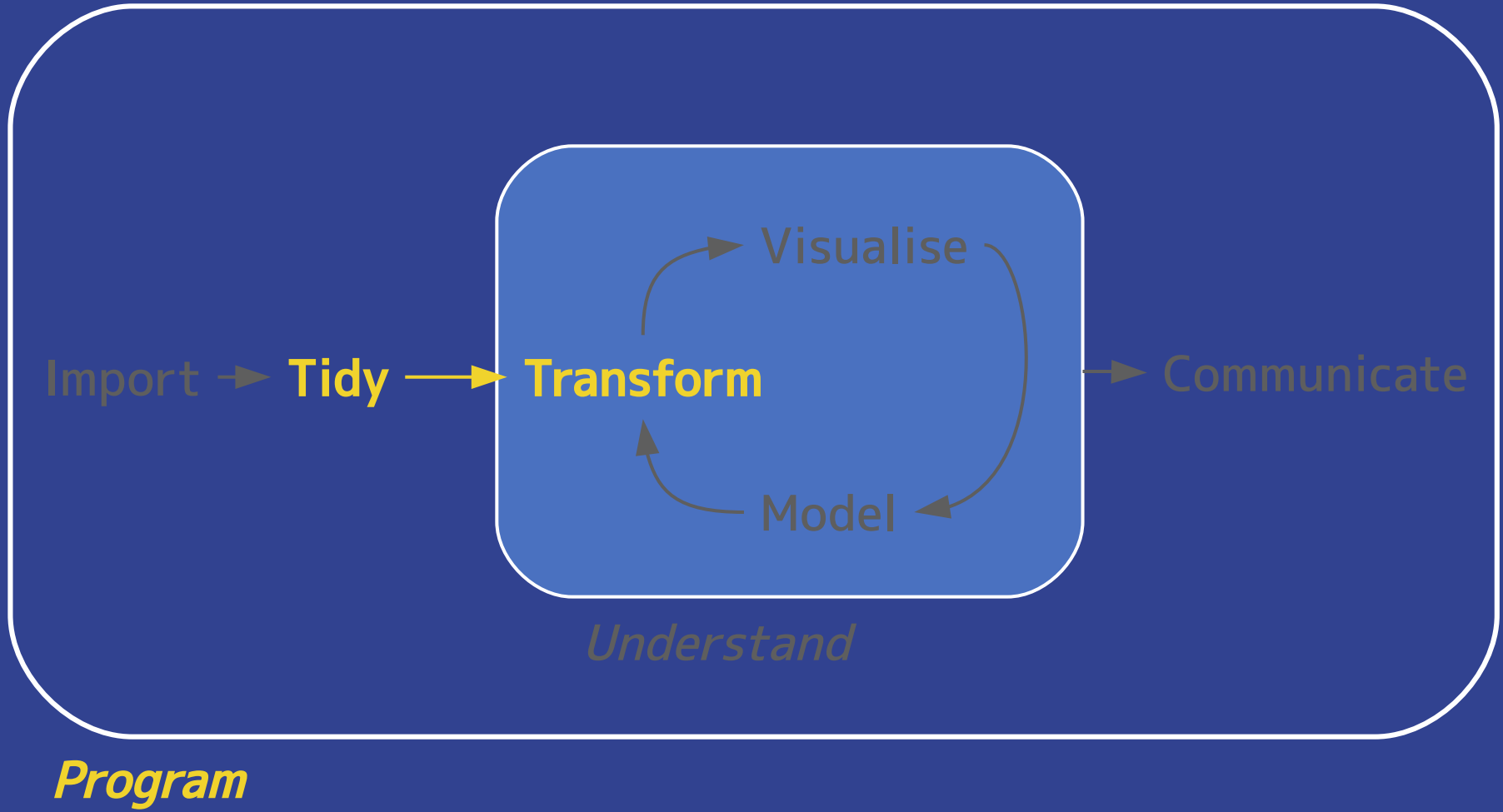
- David Minmo



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

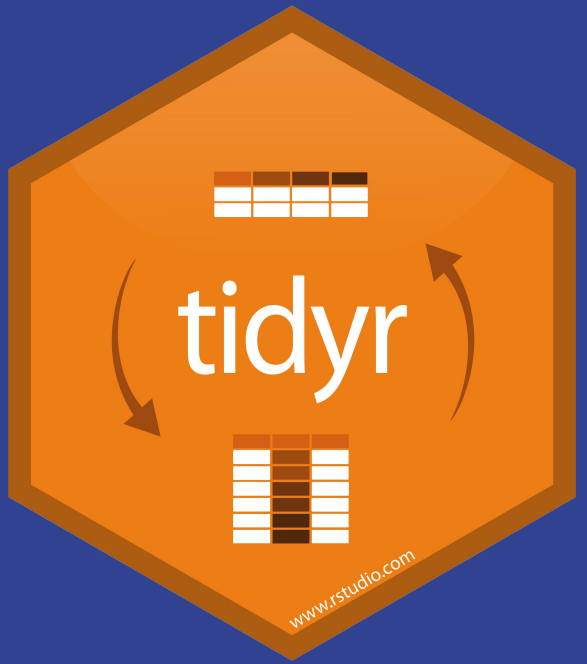
Artwork by @allison_horst


Import
Tidy
Transform
Visualise
Model
Communicate
Understand
Program


Import
Tidy
Transform
Visualise
Model
Communicate
Understand
Program

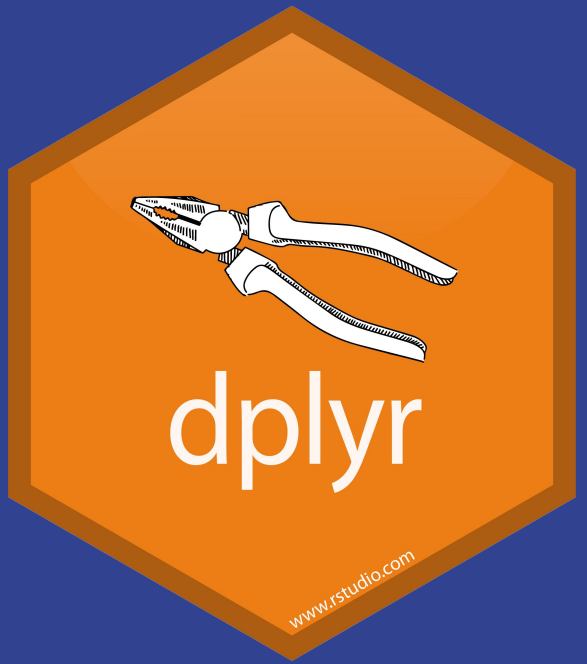
Artwork by @allison_horst

tidyr



Artwork by @allison_horst

dplyr



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

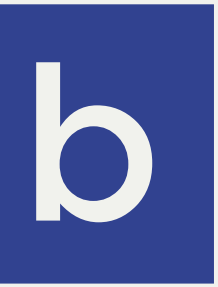

1. putar(apa)
2. jilat(apa, berapa_kali)
3. celup(apa, ke)
4. makan(apa, output)

a

```
> 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”)
```

a

```
> 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”
  )
```

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

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

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

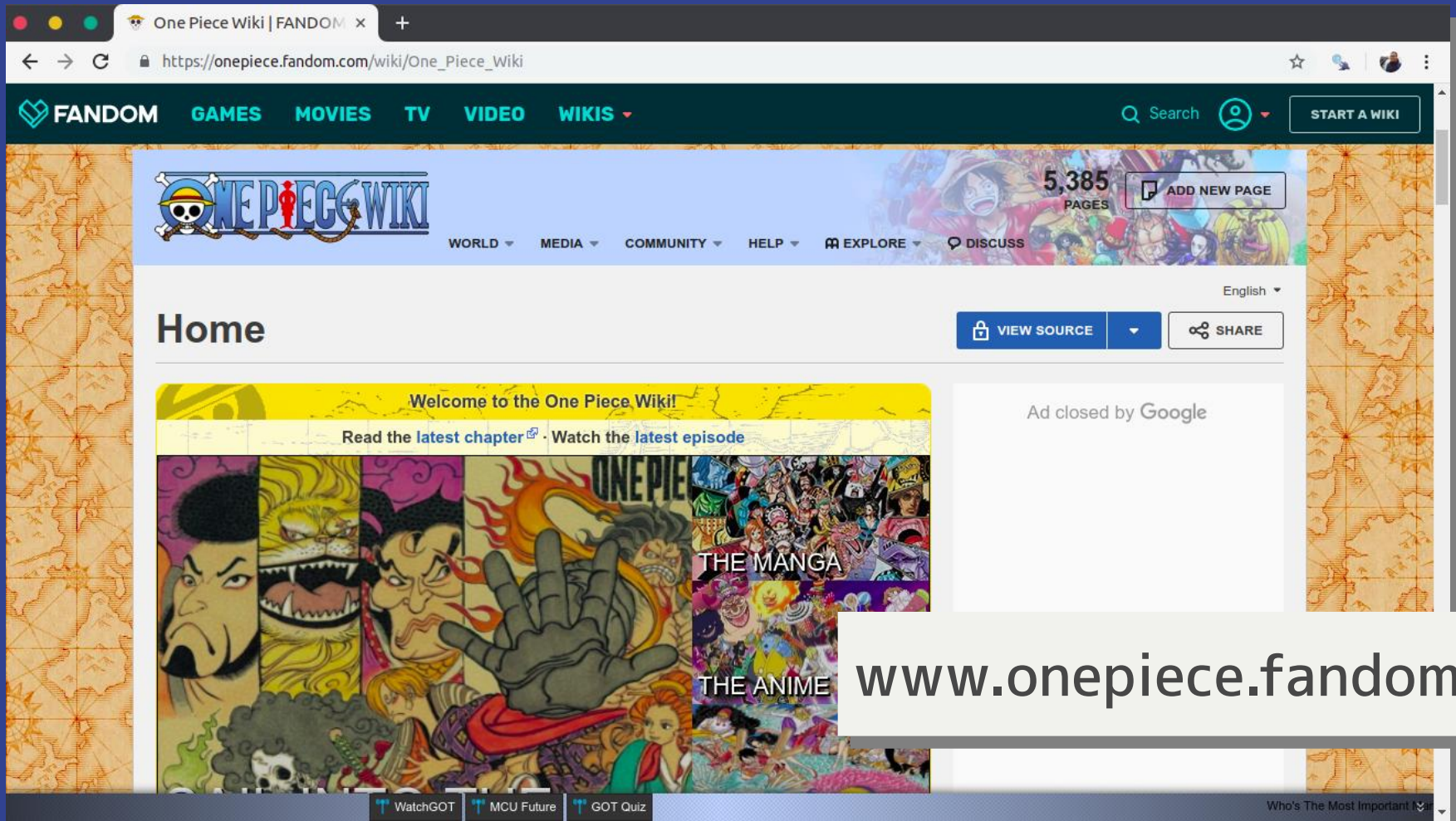
```r
arg2 %>%
  function(arg1, arg2=.,arg3, ... )
```

magrittr

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

What to do today?


ワンピース
ONE PIECE


One Piece Wiki | FANDOM
https://onepiece.fandom.com/wiki/One_Piece_Wiki
FANDOM
GAMES
MOVIES
TV
VIDEO
WIKIS
Search
START A WIKI
ONE PIECE WIKI
5,385
PAGES
ADD NEW PAGE
WORLD
MEDIA
COMMUNITY
HELP
EXPLORE
DISCUSS
English
Home
VIEW SOURCE
SHARE
Welcome to the One Piece Wiki!
Read the latest chapter · Watch the latest episode
THE MANGA
THE ANIME
Ad closed by Google
www.onepiece.fandom.com
WatchGOT
MCU Future
GOT Quiz
Who's The Most Important

Let's get started!

- Let's write R scripts together!
- I will demonstrate and explain the use of each code
- Access this presesentation at:

s.id/data-carpentry-with-tidyverse

```r
# Load packages ----------------------------------------------------------

library(tidyverse)
library(rvest)
library(janitor)
```

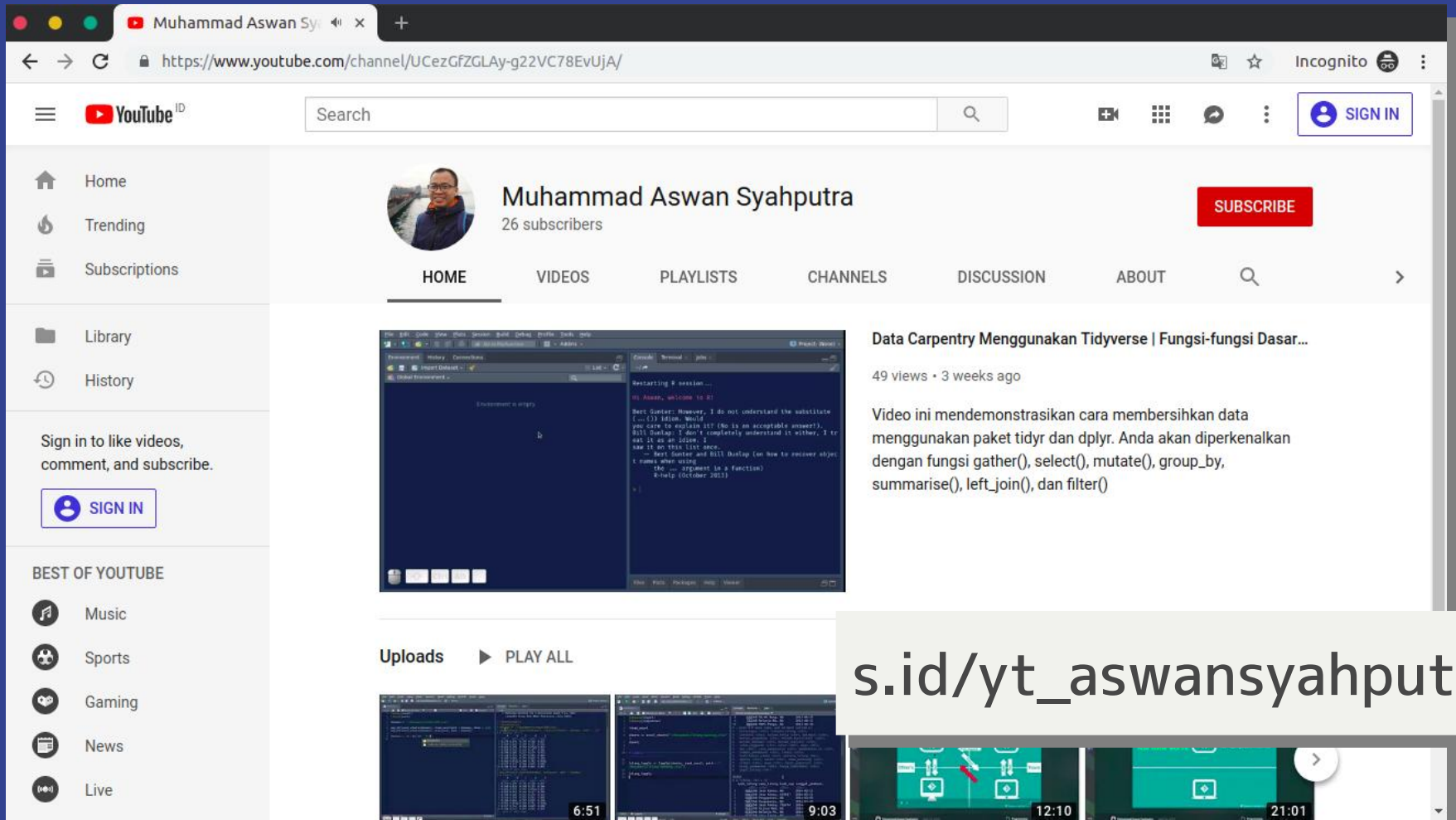
```r
# Define function to scrape OP character ----------------------------------

scrape_op_char ← function(char_url) {
  char_url %>%
    read_html() %>%
    {
      tibble(
        label = html_nodes(., ".pi-data-label") %>%
          html_text(),
        value = html_nodes(., ".pi-data-value") %>%
          html_text()
      )
    } %>%
    mutate(
      label = make_clean_names(label),
      label = recode(label,
                     japanese_name_2 = "devilfruit_jp_1",
                     english_name   = "devilfruit_en_1",
                     meaning = "devilfruit_meaning_1",
                     type = "devilfruit_type_1",
                     japanese_name_3 = "devilfruit_jp_2",
                     english_name_2 = "devilfruit_en_2",
                     meaning_2 = "devilfruit_meaning_2",
                     type_2 = "devilfruit_type_2"),
      value = str_remove_all(value, "\\[\\d+\\]")
    )
}
```

```r
# List all OP characters ----------------------------------------------------

op_chars_list <-
  "https://onepiece.fandom.com/wiki/List_of_Canon_Characters" %>%
  read_html() %>%
  html_node(".wikitable:nth-child(6)") %>%
  html_table(fill = TRUE, trim = TRUE) %>%
  as_tibble(.name_repair = make_clean_names) %>%
  select(-x, -x_2)

op_chars_list
```

```r
# Get url for all of OP characters ----------------------------------------

op_chars_urls_raw <-
  "https://onepiece.fandom.com/wiki/List_of_Canon_Characters" %>%
  read_html() %>%
  html_nodes(".wikitable:nth-child(6) > tr > td > a") %>%
  {
    tibble(
      name = html_attr(., "title"),
      url = str_c("https://onepiece.fandom.com", html_attr(., "href"))
    )
  }

op_chars_urls <-
  op_chars_urls_raw %>%
  semi_join(op_chars_list) %>%
  distinct(name, .keep_all = TRUE) %>%
  deframe()

op_chars_urls
```

```r
# Scrape all OP characters ---------------------------------------------

op_chars_raw <-
  map_dfr(op_chars_urls, scrape_op_char, .id = "name")

op_chars_raw
```

```r
# Prepare tidy dataset ----------------------------------------------------

op_characters <-
  op_chars_raw %>%
  mutate(label = factor(label, levels = unique(label))) %>%
  spread(label, value) %>%
  mutate_all(~na_if(.x, "N/A")) %>%
  unite("affiliations", matches("aff"), sep = ";") %>%
  mutate(affiliations = str_remove_all(affiliations, "(?:NA;NA|;NA)")) %>%
  mutate_all(~na_if(.x, "")) %>%
  inner_join(op_chars_list) %>%
  distinct(name, .keep_all = TRUE) %>%
  mutate(affiliations = str_remove_all(affiliations, "\\(.+\\)")) %>%
  separate_rows(affiliations, sep = "(?:;|,)") %>%
  mutate_if(is.character, ~str_trim(.x)) %>%
  select(name, debut_manga = chapter, debut_anime = episode, year, everything(),
-debut) %>%
  select_if(~mean(is.na(.x)) < 0.7)

op_characters
```


Muhammad Aswan Sy
https://www.youtube.com/channel/UCezGfZGLAy-g22VC78EvUjA/
Incognito
YouTube ID
Search
SIGN IN
Home
Trending
Subscriptions
Library
History
Sign in to like videos, comment, and subscribe.
SIGN IN
BEST OF YOUTUBE
Music
Sports
Gaming
News
Live
Muhammad Aswan Syahputra
26 subscribers
SUBSCRIBE
HOME
VIDEOS
PLAYLISTS
CHANNELS
DISCUSSION
ABOUT
Data Carpentry Menggunakan Tidyverse | Fungsi-fungsi Dasar...
49 views • 3 weeks ago
Video ini mendemonstrasikan cara membersihkan data menggunakan paket tidyr dan dplyr. Anda akan diperkenalkan dengan fungsi gather(), select(), mutate(), group_by, summarise(), left_join(), dan filter()
Uploads
PLAY ALL
6:51
9:03
12:10
21:01
s.id/yt_aswansyahputra

Thanks!

aswansyahputra@sensolution.id
www.aswansyahputra.com
speakerdeck.com/
aswansyahputra