

Pre-processing data di R

(Pelatihan data sains menggunakan R dan Gephi)

Ujang Fahmi

Pelajaran ke-1



Salam kenal dan selamat datang.

Semoga kita semua bisa saling berbagi pengalaman dan pengetahuan. Saya adalah Ujang Fahmi, Co-founder dan mentor Sadasa Academy.

Jika anda berada dan sedang membaca tutorial ini, maka kemungkinan anda adalah orang yang sedang ingin belajar data sains, atau mungkin ditugaskan untuk mempelajari R oleh institusi atau organisasi anda. Sama seperti saya dulu, dimana tanpa latar belakang engineering saya didiharuskan untuk belajar R, demi menyelesaikan tugas akhir dan akhirnya jadilah seperti saya sekarang ini.

Satu hal yang pasti, ini adalah langkah pertama dari banyak langkah yang harus dilalui, entah melalui lembaga resmi atau belajar secara mandiri. Jadi selamat belajar!!!

Ujang Fahmi, Yogyakarta, 2021-09-16

Error yang umum 1

Di worksapce kita menulis

```
a + 198
```

Di console menampilkan:

Error: object 'a' not found

Artinya objek a tidak bisa ditemukan sehingga R tidak bisa menyelesaikan perintah yang diberikan

Error yang umum 2

Di workspace kita mencoba untuk memberikan perintah impor data dengan menggunakan fungsi `read_csv`, kita juga udah yakin argumen yang dibutuhkan sudah sesuai.

```
df1 = read_csv("folder1/file.csv")
```

Tapi, di console kita masih tetap menerima error berikut:

```
Error in read_csv("folder1/file.csv") :  
could not find function "read_csv"
```

Error menunjukkan bahwa fungsi `read_csv` tidak ditemukan. Artinya kita belum memanggil library yang dibutuhkan.

Impor dan Ekspor Data

Di R kita hanya bisa mengolah data yang sudah kita impor atau berada di environment R. Maka ketika ada `error: object not found` di console, sebagian besar disebabkan oleh data/objek yang belum ada.

Impor Data

- Impor data merupakan salah satu langkah pertama dan utama dalam pengolahan data
- impor umumnya menggunakan sintak dengan awalan `read`. Misal `read.csv()` atau `read_csv`, sama-sama digunakan untuk impor data `.csv`

Impor data frame

Data frame atau data flate di R bisa diimpor dengan beberapa sintaks dari beberapa package.

Argumen yang dibutuhkan:

1. namafolder
2. namafile atau link data

```
# dasar
df1 = read.csv(file = "namafolder/namafile.csv", header = TRUE,
               sep = ",")

# readr
library(readr)
df2 = read_csv(file = "namafolder/namafile.csv", trim_ws = TRUE)

# xlsx
library(readxl)
df3 = read_excel(path = "namafolder/namafile.xlsx", sheet = 1)
```

Impor data json

Json atau `java script object notation` adalah format data yang umum digunakan untuk ditampilkan di sebuah laman/website.

Impor data lain

Ekspor data frame

Ekspor data lain

Nama bari menjadi kolom (Row names to column)

Membuat kolom baru

Memisah dan Menggabungkan nilai kolom

Merangkum variabel

Table of Contents

Mendapatkan Bantuan dan Mengatasi Error

- Error yang umum 1

- Error yang umum 2

Impor dan Ekspor Data

- Impor Data

- Ekspor Data

Pengecekan Data

- Jumlah observasi dan variabel

- Rangkuman data

Penataan Data (*Data Wrangling*)

- Memilih Kolom dan Menyusun nama kolom

- Memilih Baris

- Transformasi kolom

Mengatasi Nilai Hilang

- Menggunakan teknik manual

- Menggunakan `package`