

Modul 02 - R for Data Analysis

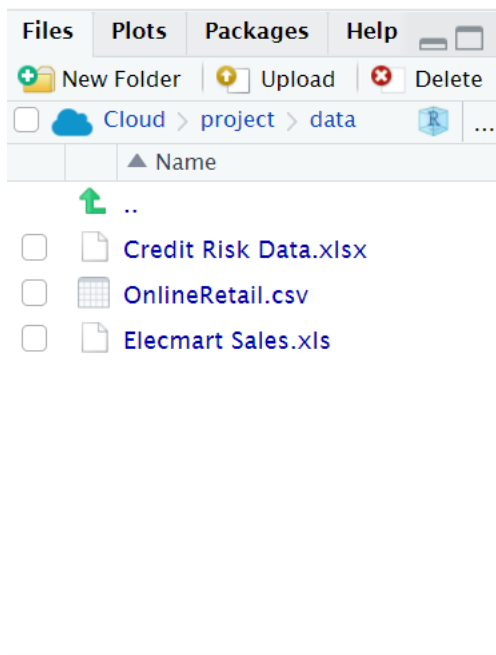
Roni Yunis

3/15/2022

Membaca dataset

Setelah memahami dasar dari pemrograman R seperti tipe dan struktur data, sekarang kita akan membahas bagaimana cara mengimport dan membaca dataset pada R. Import dataset pada R dapat dilakukan melalui 2 cara: 1. Melalui menu import dataset pada tab menu File. 2. Melalui upload file pada panel file.

Import dataset tergantung dari jenis format file dari dataset tersebut, ada yang berformat excel, csv, xlsx, dll. Dalam contoh ini kita sudah mengimport dataset pada folder data, dan bisa dilihat pada panel file yang menampung 3 buah dataset dengan format yang berbeda yaitu format: .csv, .xlsx dan .xls



data csv

Untuk membaca data dengan format .csv pada R bisa menggunakan fungsi `read.csv()`.

```
retail <- read.csv("data/OnlineRetail.csv")
```

setelah kita jalankan fungsi diatas maka nanti pada panel environment yang terbentuk sebuah data dengan nama *retail*

data excel(.xls)/(.xlsx)

Untuk membaca data dengan format excel atau .xls, dan .xlsx, kita harus menginstall sebuah packages yaitu “readxl”, untuk install packages nanti akan dibahas pada bagian khusus dibawah. Setelah packages berhasil di install, maka kita harus aktifkan packages tersebut dengan perintah seperti ini:

```
#panggil library readxl
library(readxl)
```

Setelah library “readxl” berhasil dipanggil, maka untuk membaca file .xls bisa menggunakan fungsi `read_excel()`

```
sales <- read_excel("data/Elecmart Sales.xls", sheet = "Data")
```

“sheet” menunjukkan nama sheet atau nomor sheet yang ada pada file excel tersebut.

Untuk membaca data dengan format .xlsx, caranya sama dengan yang sebelumnya

```
risk <- read_excel("data/Credit Risk Data.xlsx", sheet = "Base Data")
```

```
risk
```

```
## # A tibble: 425 x 12
##   LoanPurpose      Checking Savings MonthsCustomer MonthsEmployed Gender
##   <chr>          <dbl>    <dbl>          <dbl>          <dbl> <chr>
## 1 Small Appliance      0      739             13             12 M
## 2 Furniture            0     1230             25              0 M
## 3 New Car              0      389             19            119 M
## 4 Furniture          638      347             13             14 M
## 5 Education          963     4754             40             45 M
## 6 Furniture        2827         0             11             13 M
## 7 New Car              0      229             13             16 M
## 8 Business              0      533             14              2 M
## 9 Small Appliance    6509      493             37              9 M
## 10 Small Appliance   966         0             25              4 F
## # i 415 more rows
## # i 6 more variables: MaritalStatus <chr>, Age <dbl>, Housing <chr>,
## #   Years <dbl>, Job <chr>, CreditRisk <chr>
```

Memeriksa isi dataset

Setelah kita berhasil mengimport dan membaca dataset, maka langkah selanjutnya harus dilakukan adalah memeriksa isi dari dataset. Ada beberapa fungsi yang sering digunakan yaitu fungsi `str()`, `head()`, `tail()`, `summary()`, `Select()`, dan `filter()`.

str()

Fungsi `str()` digunakan untuk melihat struktur dan variabel dari data. Misalkan kita akan melihat struktur data dari objek *retail* yang sebelumnya kita gunakan untuk menyimpan data “OnlineRetail.csv”. Cara menggunakan fungsi `str()` adalah:

```
str(retail)
```

```
## 'data.frame':    541909 obs. of  8 variables:
## $ InvoiceNo : chr  "536365" "536365" "536365" "536365" ...
## $ StockCode : chr  "85123A" "71053" "84406B" "84029G" ...
## $ Description: chr  "white hanging heart t-light holder" "white metal lantern" "cream cupid hearts coat hanger" ...
## $ Quantity : int   6 6 8 6 6 2 6 6 6 32 ...
## $ InvoiceDate: chr  "12/1/2010 8:26" "12/1/2010 8:26" "12/1/2010 8:26" "12/1/2010 8:26" ...
## $ UnitPrice : num   2.55 3.39 2.75 3.39 3.39 7.65 4.25 1.85 1.85 1.69 ...
## $ CustomerID: int   17850 17850 17850 17850 17850 17850 17850 17850 17850 13047 ...
## $ Country : chr  "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...
```

kalau fungsi `str(retail)` tersebut kita jalankan, maka akan memperlihatkan struktur data (jumlah baris dan jumlah kolom), dan variabel-variabel dari data beserta dengan type datanya.

Head()

Fungsi `head()` digunakan untuk menampilkan 6 data teratas dari dataset. Misalnya kita akan melihat 6 data teratas dari objek *retail*, maka cara menggunakan fungsi `head()` adalah sebagai berikut:

```
head(retail)
```

```
## InvoiceNo StockCode Description Quantity
## 1 536365 85123A white hanging heart t-light holder 6
## 2 536365 71053 white metal lantern 6
## 3 536365 84406B cream cupid hearts coat hanger 8
## 4 536365 84029G knitted union flag hot water bottle 6
## 5 536365 84029E red woolly hottie white heart. 6
## 6 536365 22752 set 7 babushka nesting boxes 2
## InvoiceDate UnitPrice CustomerID Country
## 1 12/1/2010 8:26 2.55 17850 United Kingdom
## 2 12/1/2010 8:26 3.39 17850 United Kingdom
## 3 12/1/2010 8:26 2.75 17850 United Kingdom
## 4 12/1/2010 8:26 3.39 17850 United Kingdom
## 5 12/1/2010 8:26 3.39 17850 United Kingdom
## 6 12/1/2010 8:26 7.65 17850 United Kingdom
```

kalau kita ingin menampilkan 10 data teratas dari objek *retail*, maka bisa dituliskan seperti ini:

```
#menampilkan 10 data teratas
head(retail, 10)
```

```
## InvoiceNo StockCode Description Quantity
## 1 536365 85123A white hanging heart t-light holder 6
## 2 536365 71053 white metal lantern 6
## 3 536365 84406B cream cupid hearts coat hanger 8
## 4 536365 84029G knitted union flag hot water bottle 6
## 5 536365 84029E red woolly hottie white heart. 6
## 6 536365 22752 set 7 babushka nesting boxes 2
## 7 536365 21730 glass star frosted t-light holder 6
## 8 536366 22633 hand warmer union jack 6
```

```
## 9      536366      22632      hand warmer red polka dot      6
## 10     536367      84879      assorted colour bird ornament    32
##      InvoiceDate UnitPrice CustomerID      Country
## 1  12/1/2010 8:26      2.55      17850 United Kingdom
## 2  12/1/2010 8:26      3.39      17850 United Kingdom
## 3  12/1/2010 8:26      2.75      17850 United Kingdom
## 4  12/1/2010 8:26      3.39      17850 United Kingdom
## 5  12/1/2010 8:26      3.39      17850 United Kingdom
## 6  12/1/2010 8:26      7.65      17850 United Kingdom
## 7  12/1/2010 8:26      4.25      17850 United Kingdom
## 8  12/1/2010 8:28      1.85      17850 United Kingdom
## 9  12/1/2010 8:28      1.85      17850 United Kingdom
## 10 12/1/2010 8:34      1.69      13047 United Kingdom
```

Latihan Cobalah Anda tampilkan 15 data teratas dari objek *sales* dengan fungsi `head()`

```
#your code
```

Tail()

Fungsi `tail()` digunakan untuk menampilkan 6 data terbawah dari dataset. Misalnya kita akan melihat 6 data terbawah dari objek *retail*, maka cara menggunakan fungsi `tail()` adalah:

```
tail(retail)
```

```
##      InvoiceNo StockCode      Description Quantity
## 541904     581587     23256     childrens cutlery spaceboy      4
## 541905     581587     22613     pack of 20 spaceboy napkins     12
## 541906     581587     22899     children's apron dolly girl      6
## 541907     581587     23254     childrens cutlery dolly girl      4
## 541908     581587     23255     childrens cutlery circus parade      4
## 541909     581587     22138     baking set 9 piece retrospot      3
##      InvoiceDate UnitPrice CustomerID Country
## 541904 12/9/2011 12:50      4.15      12680 France
## 541905 12/9/2011 12:50      0.85      12680 France
## 541906 12/9/2011 12:50      2.10      12680 France
## 541907 12/9/2011 12:50      4.15      12680 France
## 541908 12/9/2011 12:50      4.15      12680 France
## 541909 12/9/2011 12:50      4.95      12680 France
```

Latihan Cobalah Anda tampilkan 10 data terbawah dari objek *sales* dengan fungsi `tail()`

```
#your code
```

Summary()

Untuk melihat ringkasan dari data, maka bisa digunakan fungsi `summary()`, fungsi ini akan menyajikan ringkasan dari variabel-variabel yang ada dalam dataset. Seperti nilai min, max, mean, Q1, dan Q3 serta juga melihat data kosong (NA's). Penggunaan fungsi `summary()` bisa seperti ini.

```
summary(retail)
```

```
## InvoiceNo      StockCode      Description      Quantity
## Length:541909 Length:541909 Length:541909 Min.    :-80995.00
## Class :character Class :character Class :character 1st Qu.:    1.00
## Mode  :character Mode  :character Mode  :character Median :    3.00
##                                     Mean  :    9.55
##                                     3rd Qu.:   10.00
##                                     Max.   : 80995.00
##
## InvoiceDate      UnitPrice      CustomerID      Country
## Length:541909 Min.    :-11062.06 Min.    :12346 Length:541909
## Class :character 1st Qu.:    1.25 1st Qu.:13953 Class :character
## Mode  :character Median :    2.08 Median :15152 Mode  :character
##                                     Mean  :    4.61 Mean  :15288
##                                     3rd Qu.:    4.13 3rd Qu.:16791
##                                     Max.   : 38970.00 Max.   :18287
##                                     NA's   :135080
```

Bekerja dengan Packages/Library

Dalam Bahasa Pemrograman R, banyak sekali tersedia packages/library yang mendukung analisis data. Setiap library diperuntukkan sesuai dengan kebutuhannya masing-masing. Dalam RStudio ada 2 cara instalasi library yang dapat dilakukan. Bisa melalui panel Console/Script atau melalui panel packages. Berikut ini kita akan bahas secara ringkas bagaimana instalasi packages/library tersebut dilakukan

Install packages melalui console

Untuk menginstall packages melalui console, dapat secara langsung dengan mengetik script `'install.packages("nama library")'`. Sebagai contoh kita akan menginstall packages *ggplot2*, packages ini digunakan untuk memvisualisasikan hasil analisis bisa dalam bentuk grafik (garis,batang,dll)



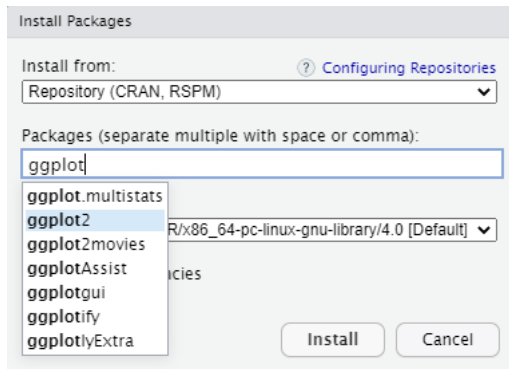
```
Console Terminal Jobs
/cloud/project/
> install.packages("ggplot2")
Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
(as 'lib' is unspecified)
trying URL 'http://package-proxy/focal/src/contrib/ggplot2_3.3.3.tar.gz'
Content type 'application/x-tar' length 4072450 bytes (3.9 MB)
=====
downloaded 3.9 MB

* installing *binary* package 'ggplot2' ...
* DONE (ggplot2)

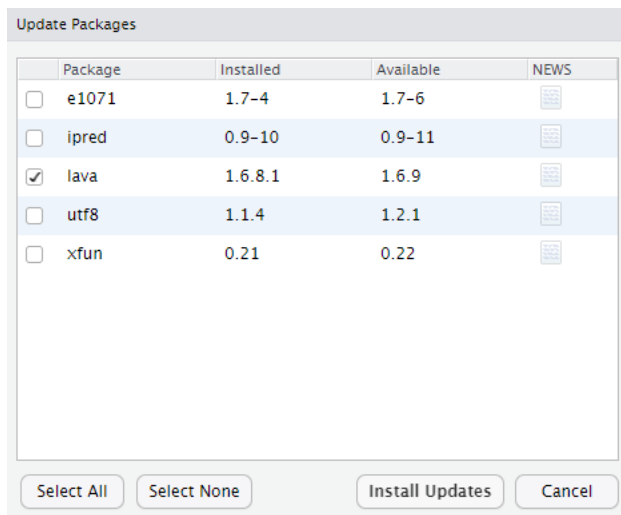
The downloaded source packages are in
  '/tmp/RtmpzqswSv/downloaded_packages'
> |
```

Install packages melalui panel packages

Untuk cara yang kedua ini adalah menginstall packages melalui panel *packages*. Ada 2 tab pilihan yaitu *Install* dan *Update*. Tab *Install* digunakan untuk menginstall packages yang diinginkan, dengan cara memasukkan nama packages yang ingin diinstall kemudian mengklik tombol *Install*.



Tab *Update* digunakan untuk memperbarui versi dari packages, dengan cara memilih nama packages dan mengklik tombol *Install Update*



Penutup

Demikian pembahasan kita tentang R untuk Analisis Data, semoga materi ini bisa Anda jadikan sebagai dasar untuk mengikuti materi-materi selanjutnya. *Selamat belajar dan salam sehat selalu*