

# MODUL 1: PENGENALAN R

## PRAKTIKUM KOMPUTASI STATISTIKA 1

### 1. Pengenalan R

#### R GUI

Bentuk R GUI (*General User Interface*) ada banyak, termasuk di dalamnya R CLI (*Command Line Interface*), R Commander, dan R Studio yang merupakan jenis R GUI paling sering dipakai. Di modul ini, akan dijelaskan sedikit mengenai masing-masing R GUI, cara installnya, fitur-fitur, serta kelebihan dan kekurangannya.

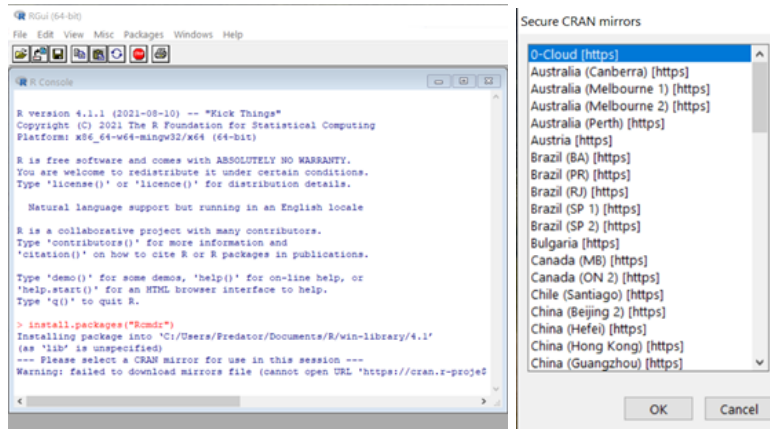
#### a. R Commander

Seperti kita telah pelajari dan lihat sebelumnya, R adalah perangkat lunak statistik berbasis perintah (*command driven*), yang sepertinya dapat memberi kesulitan bagi pengguna pemula untuk mengingat serangkaian perintah ataupun formula. Hal ini sudah diantisipasi oleh perangkat lunak tersebut dengan menyediakan penggunaan Graphical User Interface (GUI) yang memudahkan pengguna tanpa harus mengingat serangkaian perintah. GUI untuk R terbungel dalam paket yang bernama Rcmdr yang berisikan GUI untuk statistika dasar. Statistika dasar meliputi diantaranya pemodelan uji hipotesis, distribusi (normal, F, binomial, t, Chi-square, poisson), graph (histogram, scatterplot, boxplot, dan lain – lain), statistik deskriptif. Namun sebagai open source yang memiliki banyak kontributor, dalam pengembangannya juga mendukung untuk statistik lanjut, seperti pemodelan linier dan generalized-linear. Secara arsitektur menu, R terdiri dari menu utama seperti *File*, *Edit*, *Data*, *Statistics*, *Graphs*, *Models*, *Distributions*, *Tools*, dan *Help*.

#### Instalasi & Menu dalam Rcmdr

##### 1. Instalasi

Cara termudah melakukan instalasi R Commander pada sistem operasi Windows adalah jika anda terkoneksi dengan Internet, yakni melalui Menu RGui Packages > Install package(s). R secara otomatis akan meminta anda untuk memilih mirror tempat Rcmdr akan didownload dan diinstall.



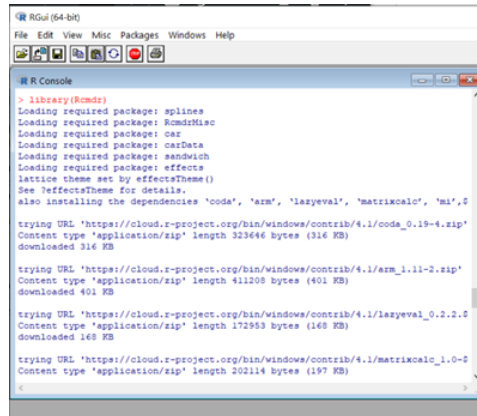
Gambar 1: Instalasi Rcmdr dengan koneksi Internet

Paket Rcmdr dan semua paket yang tidak terdapat (missing) yang sering terjadi jika instalasi tanpa koneksi internet, akan langsung di download dan diinstal secara otomatis. Sebagai contoh, jika hendak menginstall paket graf 3D graphics, anda memerlukan koneksi Internet untuk mendapatkan sumbernya (source). Pada Windows, Rcmdr hanya berfungsi pada SDI (Single-Document Rgui Interface). Untuk mengaktifkan SDI, buatlah copy ikon R desktop (klik-kanan pada ikon, seret dpada lokasi yang berbeda di desktop, pilih Copy Here dari menu context). Kemudian klik-kanan pada ikon yang baru dan pilih Properties. Tambahkan ---sdi (didahului oleh spasi) ke bagian/daerah Target pada tabulasi Shortcut dalam kotak dialog Properties. Bagian/daerah yang tadi dimaksud harus dapat membaca ekstensi alamat direktori seperti "C:\Program Files\R\rw2010\bin\Rgui.exe" - - sdi. Jika anda menginginkan , ubah nama ikon pada tabulasi General (misal ,R2.1.0. SDI). Kemudian klik OK.

Catatan: dalam instalasi Rcmdr untuk penulisan buku ini, kami menggunakan koneksi internet untuk mendownload dan menginstal file dan paket-paket yang diperlukan untuk instalasi lengkap.

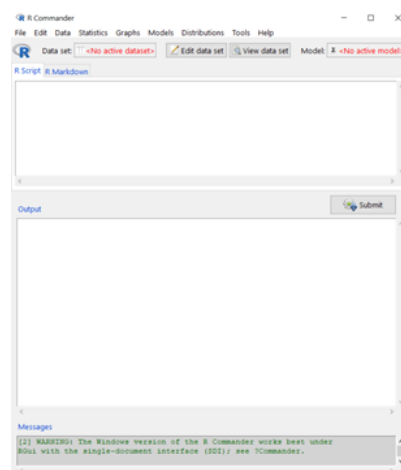
## 2. Load/menampilkan Rcmdr

Menuliskan sintaks "library(Rcmdr)" dan *run*, maka akan muncul tampilan :



Gambar 2: Loading paket R commander (Rcmdr)

### 3. Tampilan & Menu Rcmdr



Gambar 3: Tiga sub window dalam Rcmdr

Dalam GUI Rcmdr, terdiri dari tiga sub window (layar) utama, yakni:

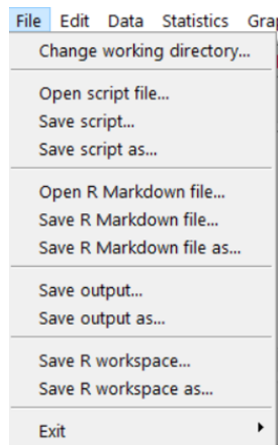
- Script Window: Layar yang menampilkan perintah (command) R. Perintah dalam bagian ini dapat di edit dan di eksekusi dan di eksekusi ulang (dengan menekan tombol submit)
- Output Window: Layar yang menampilkan hasil perhitungan yang di eksekusi dari Script Window

- Warning Window: Layar yang menampilkan peringatan seperti pesan error, data missing, atau adanya paket yang belum diinstall

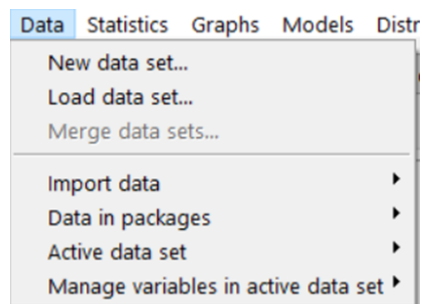
Berikut adalah menu utama dan submenu dalam Rcmdr:

- Menu File untuk loading dan menyimpan file script; untuk menyimpan output dan R workspace (ruang kerja R); dan untuk keluar aplikasi Rcmdr dan atau keluar aplikasi Rcmdr sekaligus R.
- Menu Edit (Cut, Copy, Paste, dll.) untuk penyuntingan (editing) konten dari script dan output windows. Klik kanan pada script atau output window akan menampilkan menu edit “konteks”.
- Submenu Data terdiri atas item untuk membaca dan memanipulasi data, seperti import data dari file ber-ekstensi Text, SPSS, minitab atau STATA.
- Submenu Statistics terdiri atas item-item berupa alat analisis statistika dasar.
- Menu Graphs untuk membuat grafik statistic dasar, seperti histogram, scatterplot, Boxplot, Pie chart, 3D dan lain lain.
- Menu Models item dan submenu-nya digunakan untuk menghitung summary numeric, confidence intervals, hypothesis tests, diagnostics, dan graphs untuk model statistic dan untuk menambahkan kuantitas diagnostic, misalkan,
- Residuals.
- Menu Distributions terdiri dari submenu distribusi probabilitas, quantiles, dan grafik distribusi statistika standar, yaitu distribusi Normal, Chi-Square, t, F, binomial dan Poisson.
- Menu Tools untuk loading paket R yang tidak berelasi dengan paket Rcmdr (misal, untuk mengakses data yang tersimpan dalam paket lain), dan untuk men-setting beberapa option.
- Menu Help yang berisikan informasi mengenai R dan R Commander termasuk manualnya. Setiap kotak dialog (dialog box) mempunyai tombol Help button.

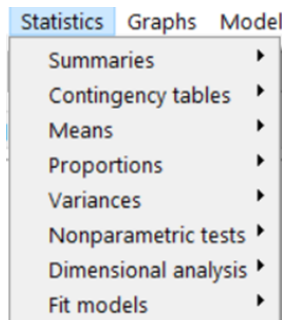
Cuplikan gambar (screenshot) menu dalam Rcmdr, disajikan pada gambar-gambar berikut ini.



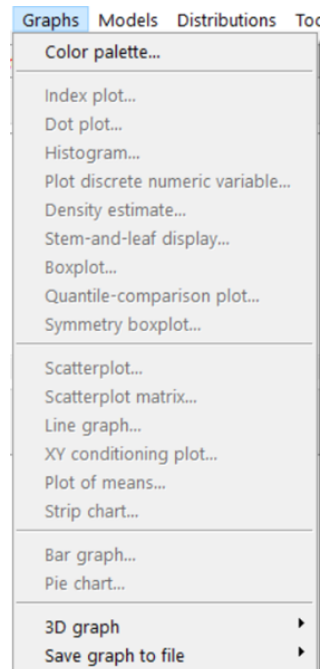
Gambar 4: Menu File



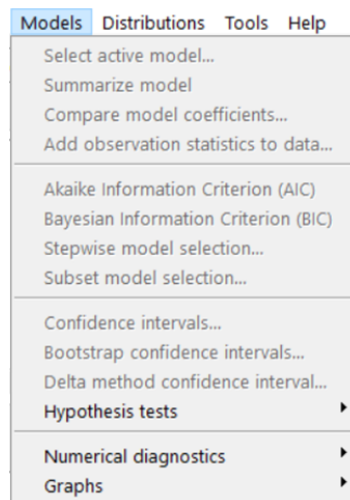
Gambar 5: Menu Data



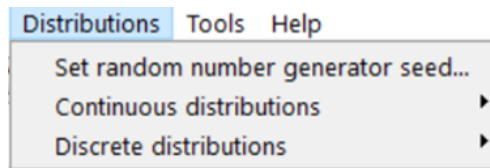
Gambar 6: Menu Statistics



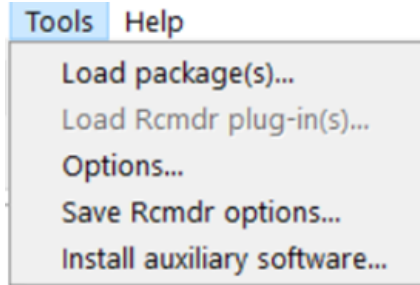
Gambar 7: Menu Graph



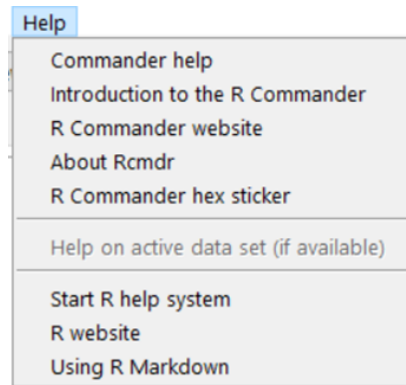
Gambar 8: Menu Models



Gambar 9: Menu Distributions



Gambar 10: Menu Tools



Gambar 11: Menu Help

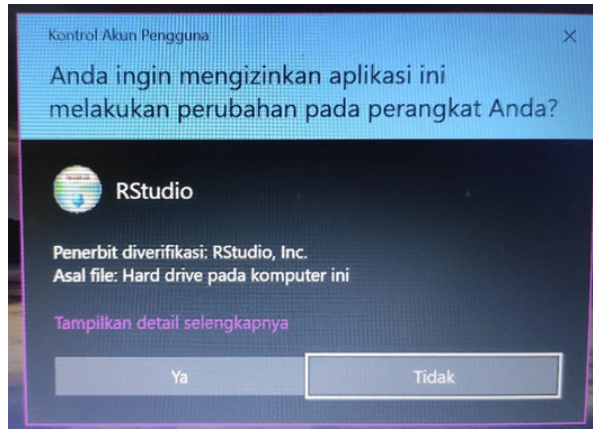
## b. R Studio

RStudio adalah Integrated Development Environment (IDE) untuk R yang banyak digunakan hingga saat ini. Dapat dikatakan bahwa hampir semua pengguna R yang sudah mengetahui RStudio akan lebih memilih menggunakan R melalui RStudio dibandingkan dengan menggunakan R GUI. R dan RStudio adalah dua program yang berbeda User tidak harus install RStudio untuk dapat menggunakan program R (melalui R GUI). Tapi Anda diwajibkan untuk install R terlebih dahulu sebelum install dan menggunakan RStudio karena RStudio membutuhkan program R yang sudah terinstal di PC atau server.

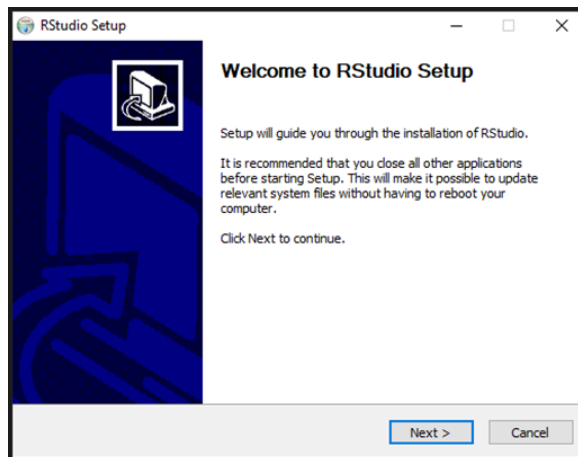
### 1. Tata Cara Install RStudio

Langkah langkah untuk install RStudio sebagai berikut :

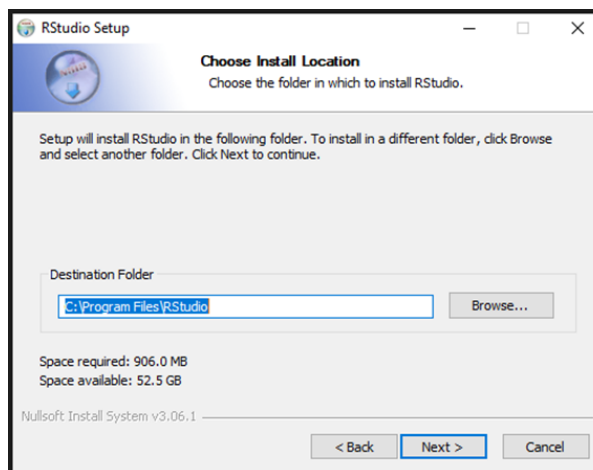
- Download RStudio Desktop  
<https://www.rstudio.com/products/rstudio/download/>
- Jalankan installer yang sudah didownload



- Klik Next

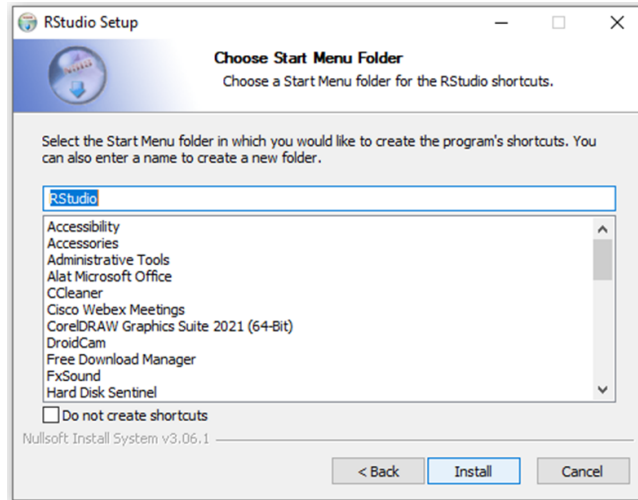


- Klik Next

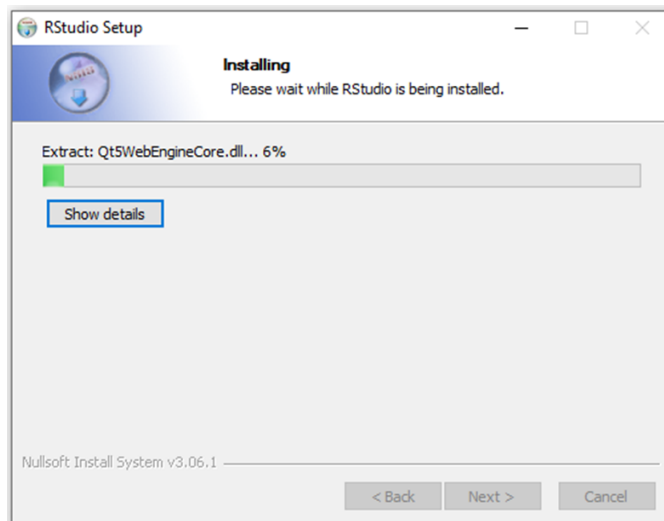


- Klik Install

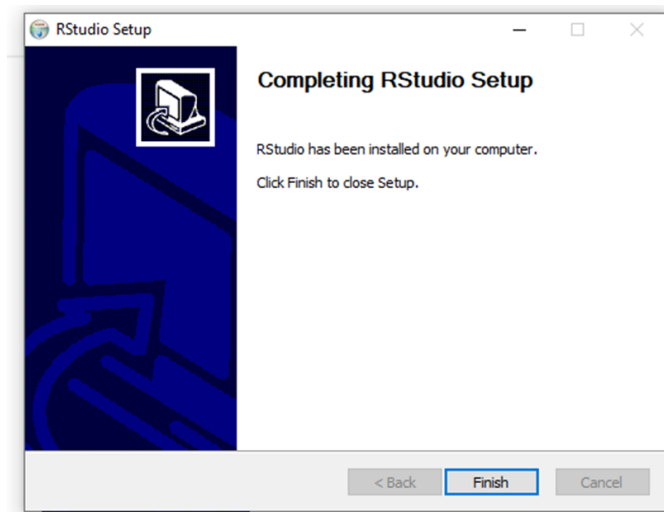




- Tunggu proses instalasi selesai

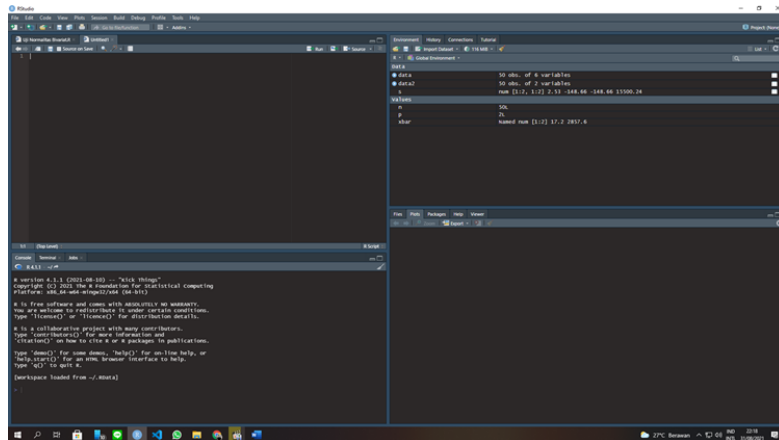


- Klik Finish

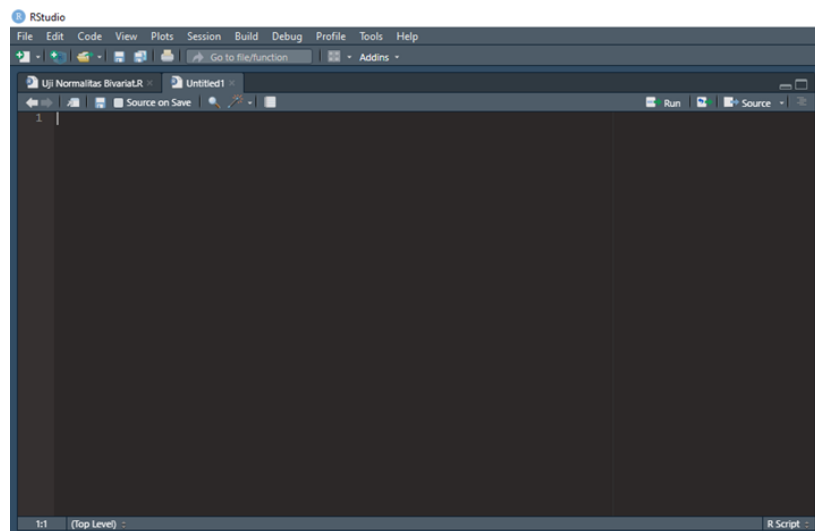


## 2. Tampilan Layout RStudio

Pada dasarnya, layout RStudio terbagi kedalam empat (4) bagian, yaitu: Source, Console, Environment/History/Connections, fitur lainnya. Gambaran umum untuk masing-masing panel dijelaskan sebagai berikut.



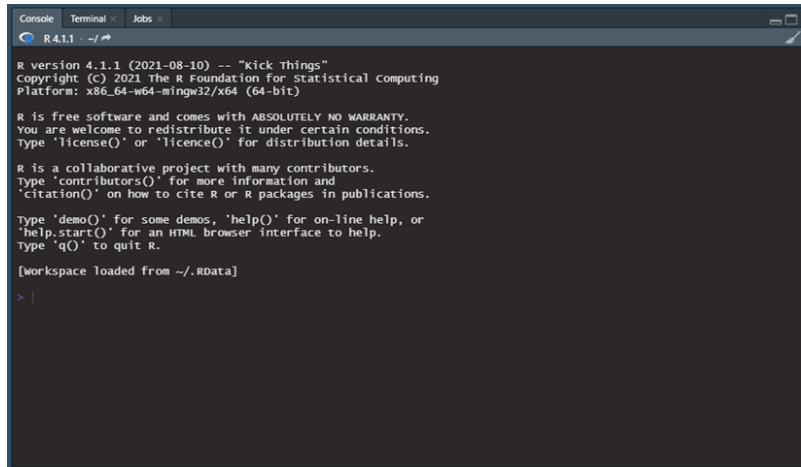
- Source/Editor



Source/Editor window adalah jendela yang dapat digunakan untuk membuat, mengedit, dan menyimpan script R. Pada jendela ini, tersedia fitur autocomplete yang akan memudahkan kita dalam membuat script. Apabila jendela tersebut tidak muncul saat pertama kali menginstall RStudio, langkah yang dapat dilakukan adalah sebagai berikut:

**Klik tab File -> New File -> R script.**

- Console



```
R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

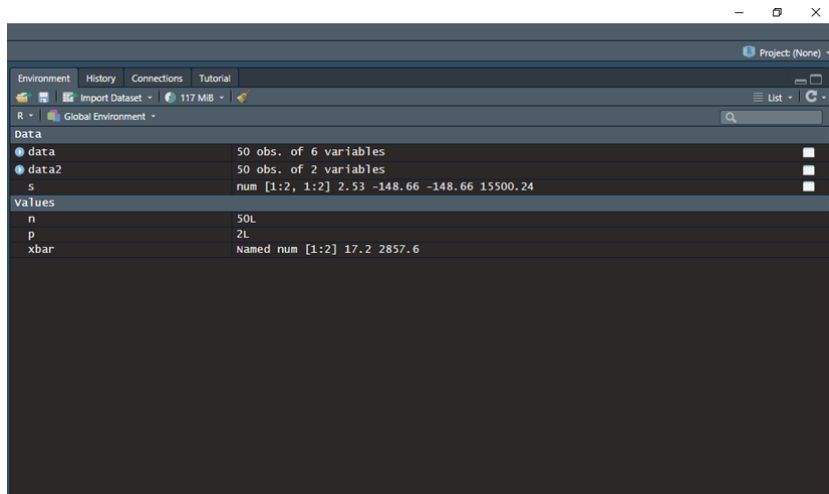
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[workspace loaded from ~/.RData]

> |
```

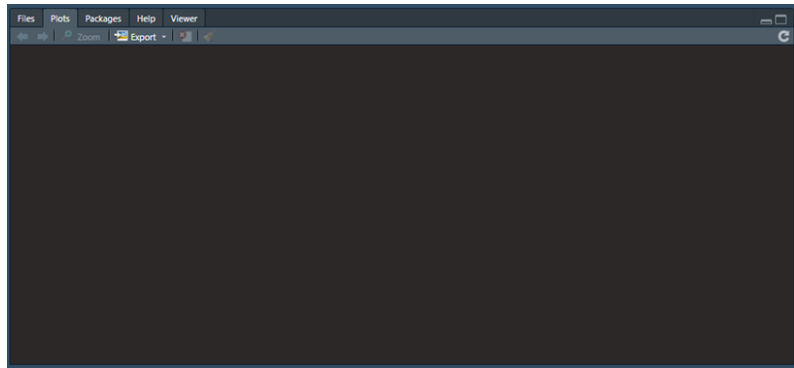
Console window adalah jendela tempat mengeksekusi command dari script R yang dibuat. Script tersebut dapat dijalankan dari Source/Editor window maupun dengan langsung mengetikkan script pada jendela ini. Jendela ini akan menampilkan output dari setiap satu proses/baris command R yang dijalankan. Jendela ini sebetulnya merupakan tampilan langsung dari setiap proses yang dikerjakan oleh R.

- Environment/History/Connections



Jendela ini terdiri dari beberapa tab, yaitu Environment, History, dan Connections. Tab environment akan menampilkan daftar data dan nilai yang sedang aktif tersimpan didalam memory (RAM). Kita dapat melihat data atau nilai tersebut dengan mengklik nama data tersebut. Tab History akan menampilkan daftar command yang telah dijalankan sebelumnya dalam satu session yang aktif. Adapun tab Connection merupakan tab khusus yang berkaitan dengan koneksi ke database seperti mySQL, postgreSQL, Spark, dll.

- Fitur Lainnya



Jendela ini merupakan fitur tambahan yang bermanfaat khususnya dalam manajemen file, menampilkan output command berupa plot, informasi dan bantuan dalam penulisan script, dan web viewer. Khusus pada tab Packages, kita dapat melihat daftar library R yang telah terinstall pada PC kita. Apabila kita ingin menambahkan library baru, cukup dengan menekan tombol install, kemudian mengetikkan nama dari library yang ingin diinstall.

**Files, Plot, Packages, Help**

Tab **files** menampilkan seluruh file dan folder.

Tab **plots** akan menampilkan seluruh grafik.

Tab **packages** akan menampilkan listing packages atau add-ons yang diperlukan agar program berjalan dengan baik

Tab **help** menampilkan penjelasan yang lengkap dan komprehensif apabila ada instruksi (code) yang tidak jelas, dilengkapi contohnya

c. **Contoh R GUI Lain**

Referensi R GUI yang lain dapat dibaca di link

<https://tinyurl.com/ReferensiRGUI>

2. **Import Data dengan R**

Jenis file yang akan dibahas adalah textfile (\*.txt), CSV (\*.csv) dan Excel file.

a. Textfile

Textfile adalah salah satu jenis file yang sering digunakan untuk menyimpan atau bertukar data. Fungsi yang sudah tersedia di R untuk import textfile adalah **read.table()** dan **read.delim()**.

b. CSV

Selain textfile, CSV juga merupakan format file yang sering digunakan untuk menyimpan atau bertukar data. Fungsi yang tersedia di R untuk dapat mengimport data dari file CSV adalah **read.csv()**.

c. Excel

Jenis file lain yang juga sering digunakan adalah file Excel. Kita dapat menggunakan package **{readxl}** dengan fungsi **read\_excel()** untuk import data

dari file Excel. Argumen **path** = adalah lokasi dan nama file Excel yang akan kita gunakan.

Ada juga cara lain yang dapat kita gunakan dalam mengimport data seperti perintah **read.delim()** yang cara kerjanya adalah pertama kita membuka file yang berisi data yang kita miliki, lalu copy bagian data yang kita ingin import dan selanjutnya tulis perintah yang sesuai, juga fungsi **file.choose()**.

### 3. Manipulasi *Data Frame* dengan R

Setelah berhasil mengimport data, kita akan mendapatkan suatu objek bertipe *data frame* yang berisikan data yang telah kita *import* di Bab 2. Di dalam praktikum atau untuk keperluan studi, data yang ada biasanya sudah bersih dan rapi agar memudahkan penggunaan studi. Namun, pada kasus nyata, data yang rapi jarang untuk ditemukan. Data tidak rapi dapat disebabkan oleh tahap pengumpulan data atau proses memasukkan data sehingga memiliki data yang hilang (*missing value*), duplikat, *redundant*, pencilan (*outlier*), dan lain-lain.

Data yang “berantakan” dapat mempersulit seorang analis atau peneliti dalam menginterpretasikan bentuk dan karakteristik dari data tersebut (Salk, 2020). Data yang berantakan bahkan juga dapat mengganggu kinerja model sehingga hasil dari model yang dibuat bisa menjadi kurang akurat dalam memprediksi karena bias-bias. Sehingga, diperlukan pembersihan data agar memudahkan analis dalam melanjutkan tahap-tahap analisis selanjutnya.

Untuk menganalisis kumpulan data, pertama-tama data harus dalam format yang rapi. Untuk dianggap sebagai kumpulan data yang rapi, data harus mengikuti prinsip *Tidy Data*, yaitu: (1) setiap pengamatan harus berada dalam barisnya sendiri, (2) setiap variabel memiliki kolomnya sendiri, dan (3) setiap nilai harus berada dalam selnya sendiri (Wickham, 2014).

“**TIDY DATA** is a standard way of mapping the meaning of a dataset to its structure.”

—HADLEY WICKHAM

### In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

Gambar 3.1. Bentuk data yang rapi (cfss.uchicago.edu)

Dari Gambar 3.1, kita dapat mengetahui bentuk data yang rapi, yaitu dengan kolom-kolomnya adalah variabel (fitur), dan baris-barisnya adalah observasi. Dengan bentuk data yang rapi, kita dapat dengan mudah mengolah data tersebut.

Menggunakan R, kita dapat dengan mudah merapikan data yang dikenal dengan istilah *Data Wrangling*. Apa itu *Data Wrangling*? *Data Wrangling* adalah metode pembersihan (*cleaning*), penataan (*structuring*), dan pengayaan (*enriching*) data mentah ke dalam format yang diinginkan untuk pengambilan keputusan yang lebih baik dalam waktu yang lebih singkat. Untuk melakukan *Data Wrangling* dengan menggunakan R, kita akan menggunakan *package tidyverse*. *Package* ini berisikan *package tidyr* dan *dplyr* yang masing-masing digunakan untuk merapikan dan memanipulasi serta mentransformasi data. Di dalam *package* ini juga berisikan *package* lain yang sering digunakan untuk keperluan *data science*.

### Operator Pipe

Untuk memudahkan dalam proses *wrangling*, operator pipe digunakan untuk memudahkan dalam menulis syntax. Operator ini memungkinkan kita untuk meneruskan hasil dari satu fungsi/argumen ke argumen lain secara berurutan. Operator pipe yang dimaksud adalah operator `%>%`. Operator ini juga dapat diakses dengan *shortcut key* Ctrl+Shift+M untuk Windows *user* dan Command+Shift+M untuk Mac *user*.

```
table5 %>%
  unite(col = "new", century, year, sep = "-")
```

Gambar 3.2. Contoh penggunaan Pipe Operator (Perhatikan %>%)

### Contoh *Tidying Data* dan *Manipulating Data* dengan R

Dengan menggunakan data billboard yang telah disediakan, kita akan melakukan *tidying data* dan *manipulating data* dengan *package tidyverse*.

```
> billboard
```

	artist	track	minutes	seconds	date.entered	wk1	wk2
1	2 Pac	Baby Don't Cry (Keep...	4	22	2000-02-26	87	82
2	2Ge+her	The Hardest Part Of ...	3	15	2000-09-02	91	87
3	3 Doors Down	Kryptonite	3	53	2000-04-08	81	70
4	2 Pac	Loser	4	24	2000-10-21	76	76
5	2Ge+her	wobble wobble	3	35	2000-04-15	57	34
6	3 Doors Down	Give Me Just One Nig...	3	24	2000-08-19	51	39

Gambar 3.3. *Head* dari Data Billboard

### *Tidying Data* (Merapikan Data)

Terdapat beberapa fungsi yang umum digunakan untuk merapikan data dengan R, yaitu:

- `gather()`

Fungsi ini digunakan untuk mengubah beberapa kolom menjadi dua kolom saja, atau secara mudahnya, membuat observasi dari variabel.

Argumen:

```
gather(data, 'name of key column', 'name of the new
value column', names or numeric index of columns to
collapse)
```

Contoh:

```
gather(billboard, "week", "rating", 6:7)
```

```
> gather(billboard, "week", "rating", 6:7)
```

	artist	track	minutes	seconds	date.entered	week	rating
1	2 Pac	Baby Don't Cry (Keep...	4	22	2000-02-26	wk1	87
2	2Ge+her	The Hardest Part Of ...	3	15	2000-09-02	wk1	91
3	3 Doors Down	Kryptonite	3	53	2000-04-08	wk1	81
4	2 Pac	Loser	4	24	2000-10-21	wk1	76
5	2Ge+her	wobble wobble	3	35	2000-04-15	wk1	57
6	3 Doors Down	Give Me Just One Nig...	3	24	2000-08-19	wk1	51
7	2 Pac	Baby Don't Cry (Keep...	4	22	2000-02-26	wk2	82
8	2Ge+her	The Hardest Part Of ...	3	15	2000-09-02	wk2	87
9	3 Doors Down	Kryptonite	3	53	2000-04-08	wk2	70
10	2 Pac	Loser	4	24	2000-10-21	wk2	76
11	2Ge+her	wobble wobble	3	35	2000-04-15	wk2	34
12	3 Doors Down	Give Me Just One Nig...	3	24	2000-08-19	wk2	39

Gambar 3.4. Data *Billboard* dengan fungsi `gather()` pada kolom *week1* dan *week2*

- `spread()`

Fungsi ini digunakan untuk mengubah dua kolom menjadi beberapa kolom, atau secara mudahnya, membuat variabel dari observasi. (kebalikan dari gather).

Argumen:

```
spread(data frame to reshape, column to use for keys,
column to use for value)
```

Contoh:

```
spread(billboard, week, rating)
```

```
> spread(billboard, week, rating)
```

	artist	track	minutes	seconds	date.entered	wk1	wk2
1	2 Pac	Baby Don't Cry (keep...	4	22	2000-02-26	87	82
2	2 Pac	Loser	4	24	2000-10-21	76	76
3	2Ge+her	The Hardest Part Of ...	3	15	2000-09-02	91	87
4	2Ge+her	wobble wobble	3	35	2000-04-15	57	34
5	3 Doors Down	Give Me Just One Nig...	3	24	2000-08-19	51	39
6	3 Doors Down	Kryptonite	3	53	2000-04-08	81	70

Gambar 3.5. Data *Billboard* dengan fungsi `spread()` pada kolom *week* dan *rating*

- `separate()`

Fungsi ini digunakan untuk memisahkan sebuah kolom berdasarkan karakter dari string dari value masing-masing observasi, atau secara mudahnya, memisahkan satu kolom menjadi beberapa kolom.

Argumen:

```
separate(data frame to reshape, column name to
separate, names of the new columns, "separator")
```

Contoh:

```
separate(billboard,
date.entered, c("year", "month", "day"), sep="-")
```

```
> separate(billboard, date.entered, c("year", "month", "day"), sep="-")
```

	artist	track	minutes	seconds	year	month	day	wk1	wk2
1	2 Pac	Baby Don't Cry (keep...	4	22	2000	02	26	87	82
2	2Ge+her	The Hardest Part Of ...	3	15	2000	09	02	91	87
3	3 Doors Down	Kryptonite	3	53	2000	04	08	81	70
4	2 Pac	Loser	4	24	2000	10	21	76	76
5	2Ge+her	wobble wobble	3	35	2000	04	15	57	34
6	3 Doors Down	Give Me Just One Nig...	3	24	2000	08	19	51	39

Gambar 3.6. Data *Billboard* dengan fungsi `separate()` pada kolom *date.entered* menjadi *year*, *month*, dan *day*.

- `unite()`

Fungsi ini digunakan untuk menggabungkan beberapa kolom menjadi sebuah kolom, atau secara mudahnya, menggabungkan beberapa kolom menjadi satu kolom.

Argumen:

```
unite(data frame to reshape, "column name to unite",
columns for unite, "separator")
```



Contoh:

```
unite(billboard, "time", minutes, seconds, sep=":")
> unite(billboard, "time", minutes, seconds, sep=":")
  artist track time date.entered wk1 wk2
1 2 Pac Baby Don't Cry (Keep... 4:22 2000-02-26 87 82
2 2Ge+her The Hardest Part Of ... 3:15 2000-09-02 91 87
3 3 Doors Down Kryptonite 3:53 2000-04-08 81 70
4 2 Pac Loser 4:24 2000-10-21 76 76
5 2Ge+her wobble wobble 3:35 2000-04-15 57 34
6 3 Doors Down Give Me Just One Nig... 3:24 2000-08-19 51 39
```

Gambar 3.7. Data *Billboard* dengan fungsi `unite()` pada kolom *minutes* dan *seconds* menjadi *time*.

### ***Manipulating Data (Memanipulasi Data)***

Dalam R, kita dapat memanipulasi data yang kita miliki untuk memudahkan dalam menginterpretasikan data yang kita punya dengan beberapa syntax, yaitu:

- `select()`

Fungsi ini digunakan untuk memilih kolom yang akan digunakan. Fungsi ini kurang lebih sama dengan fungsi `select()` yang ada pada MySQL. Untuk menyeleksi, kita dapat menggunakan argumen ini atau argumen bawaan dari *base R*, yaitu `df[, columns to be selected]`.

Argumen:

```
select(columns to be selected)
```

Contoh:

```
select(billboard, track, wk1, wk2)
> select(billboard, track, wk1, wk2)
  track wk1 wk2
1 Baby Don't Cry (Keep... 87 82
2 The Hardest Part Of ... 91 87
3 Kryptonite 81 70
4 Loser 76 76
5 wobble wobble 57 34
6 Give Me Just One Nig... 51 39
```

Gambar 3.8. Data *Billboard* dengan fungsi `select()` pada kolom *track*, *wk1*, dan *wk2*.

- `filter()`

Fungsi ini digunakan untuk menyaring data atau membuat subset data frame, dengan menampilkan semua baris yang memenuhi kondisi. Pada *base R*, syntax yang digunakan adalah `subset()`.

Argumen:

```
filter(data, condition)
```

Contoh:

```
filter(billboard, wk1>80)
```

```
> filter billboard, wk1>80)
  artist          track minutes seconds date.entered wk1 wk2
1 2 Pac Baby Don't Cry (keep... 4      22 2000-02-26 87 82
2 2Ge+her The Hardest Part Of ... 3      15 2000-09-02 91 87
3 3 Doors Down Kryptonite        3      53 2000-04-08 81 70
```

Gambar 3.9. Data *Billboard* dengan fungsi `filter()` pada kolom *wk1* dengan kondisi *week* > 80.

```
filter(billboard, minutes=="4")
> filter(billboard, minutes=="4")
  artist          track minutes seconds date.entered wk1 wk2
1 2 Pac Baby Don't Cry (keep... 4      22 2000-02-26 87 82
2 2 Pac Loser                4      24 2000-10-21 76 76
```

Gambar 3.10. Data *Billboard* dengan fungsi `filter()` pada kolom *minutes* dengan kondisi *minutes* = 4.

- `mutate()`

Fungsi ini digunakan untuk membuat variabel baru dengan memutasi variabel yang sudah ada.

Argumen:

```
mutate(data, new variable with conditions or mutation
argument)
```

Contoh:

```
mutate(billboard, average=(wk1+wk2)/2)
```

```
> mutate(billboard, average=(wk1+wk2)/2)
  artist          track minutes seconds date.entered wk1 wk2 average
1 2 Pac Baby Don't Cry (keep... 4      22 2000-02-26 87 82 84.5
2 2Ge+her The Hardest Part Of ... 3      15 2000-09-02 91 87 89.0
3 3 Doors Down Kryptonite        3      53 2000-04-08 81 70 75.5
4 2 Pac Loser                4      24 2000-10-21 76 76 76.0
5 2Ge+her wobble wobble        3      35 2000-04-15 57 34 45.5
6 3 Doors Down Give Me Just One Nig... 3      24 2000-08-19 51 39 45.0
```

Gambar 3.11. Data *Billboard* dengan fungsi `mutate()` pada kolom *average* dengan argumen *average* = (wk1+wk2)/2.

Terdapat argumen yang dapat digunakan dalam fungsi `mutate()`, yaitu:

<code>pmin(), pmax()</code>	Element-wise min and max
<code>cummin(), cummax()</code>	Cumulative min and max
<code>cumsum(), cumprod()</code>	Cumulative sum and product
<code>between()</code>	Are values between a and b?
<code>cume_dist()</code>	Cumulative distribution of values
<code>cumall(), cumany()</code>	Cumulative all and any
<code>cummean()</code>	Cumulative mean
<code>lead(), lag()</code>	Copy with values one position
<code>ntile()</code>	Bin vector into n buckets
<code>dense_rank(), min_rank(), percent_rank(), row_number()</code>	Various ranking methods

- `summarise()`

Fungsi ini digunakan untuk membuat dataframe baru yang terdiri dari satu (atau lebih) baris untuk setiap kombinasi variabel pengelompokan. Jika tidak ada variabel pengelompokan, output akan memiliki satu baris yang meringkas semua observasi di input. Secara mudahnya, fungsi ini meringkas data sehingga informasi dari data tersebut dapat ditampilkan dengan cepat.

Argumen:

```
summarise(data, variable names with conditions or
summarise argument)
```

Contoh:

```
summarise(billboard, median=median(wk1),
variance=var(wk1))

> summarise(billboard, median=median(wk1), variance=var(wk1))
  median variance
1   78.5 265.7667
```

Gambar 3.12. Data *Billboard* dengan fungsi `summarise()`.

```
summarise(billboard, average=mean(wk1),
sum=sum(wk1), count=n())

> summarise(billboard, average=mean(wk1), sum=sum(wk1), count=n())
  average sum count
1 73.83333 443     6
```

Gambar 3.13. Data *Billboard* dengan fungsi `summarise()`.

Terdapat argumen yang dapat digunakan dalam fungsi `summary()`, yaitu:

<code>min(), max()</code>	Minimum and maximum values
<code>mean(), median()</code>	Mean and median values
<code>sum()</code>	Sum of values
<code>var(), sd()</code>	Variance and standard deviation of a vector
<code>first(), last()</code>	First and last value in a vector
<code>nth()</code>	N-th value in a vector
<code>n()</code>	The number of values in a vector
<code>n_distinct()</code>	The number of unique values in a vector

- `arrange()`

Fungsi ini secara mudahnya adalah fungsi yang digunakan untuk mengurutkan atau *sorting* data berdasarkan kolom atau variabel.

Argumen:

```
arrange(data, variable or column to be sorted)
```

Contoh:

```
arrange(billboard, wk1)
```

```
> arrange(billboard, wk1)
```

	artist	track	minutes	seconds	date.entered	wk1	wk2
1	3 Doors Down	Give Me Just One Nig...	3	24	2000-08-19	51	39
2	2Ge+her	wobble wobble	3	35	2000-04-15	57	34
3	2 Pac	Loser	4	24	2000-10-21	76	76
4	3 Doors Down	Kryptonite	3	53	2000-04-08	81	70
5	2 Pac	Baby Don't Cry (Keep...	4	22	2000-02-26	87	82
6	2Ge+her	The Hardest Part Of ...	3	15	2000-09-02	91	87

Gambar 3.13. Data *Billboard* dengan fungsi `arrange()` ke atas (kecil ke besar).

```
arrange(billboard, desc(wk1))
```

```
> arrange(billboard, desc(wk1))
```

	artist	track	minutes	seconds	date.entered	wk1	wk2
1	2Ge+her	The Hardest Part Of ...	3	15	2000-09-02	91	87
2	2 Pac	Baby Don't Cry (Keep...	4	22	2000-02-26	87	82
3	3 Doors Down	Kryptonite	3	53	2000-04-08	81	70
4	2 Pac	Loser	4	24	2000-10-21	76	76
5	2Ge+her	wobble wobble	3	35	2000-04-15	57	34
6	3 Doors Down	Give Me Just One Nig...	3	24	2000-08-19	51	39

Gambar 3.14. Data *Billboard* dengan fungsi `arrange()` ke bawah (besar ke kecil).

- `group_by()`

Fungsi ini secara mudahnya adalah fungsi yang digunakan untuk mengelompokkan satu atau lebih variabel agar dapat lebih spesifik dalam menganalisis suatu variabel. Penggunaan dari fungsi ini biasa dilanjutkan dengan fungsi `summarise()`.

Argumen:

`group_by(data, variable or column to be grouped)`

Contoh:

```
group_by(billboard, artist) %>%
  summarise(mean_wk1=mean(wk1), mean_wk2=mean(wk2))
```

```
> group_by(billboard, artist)%>%
+   summarise(mean_wk1=mean(wk1), mean_wk2=mean(wk2))
```

```
# A tibble: 3 x 3
```

	artist	mean_wk1	mean_wk2
1	2 Pac	81.5	79
2	2Ge+her	74	60.5
3	3 Doors Down	66	54.5

Gambar 3.15. Data *Billboard* dengan fungsi `group_by()` lalu di ringkas untuk mendapatkan *mean* dari *week1* dan *week2*.

Pada dasarnya, masih banyak fungsi-fungsi dan syntax lain yang dapat digunakan untuk memanipulasi dan merapikan data seperti `pivot_longer()`, `pivot_wider()`, `join()`, `rename()`, dan lain-lain yang dapat dipelajari lebih lanjut secara mandiri dengan melihat *cheatsheet* di pratinjau: [Data Wrangling with dplyr and tidyr](#)