



MODUL MATA KULIAH

DATA MINING

(MIK 620 SESI 10)

DISUSUN OLEH

NOVIANDI, M.Kom
NIDN. 0318018202

PROGRAM STUDI MANAJEMEN INFORMASI KESEHATAN

FAKULTAS ILMU-ILMU KESEHATAN

UNIVERSITAS ESA UNGGUL

2018

BAB I

DOWNLOAD DAN INSTAL R PADA WINDOWS

Capaian pembelajaran

1. Mahasiswa mampu menjelaskan peranan bahasa pemrograman R dalam pengolahan data.
2. Mahasiswa mampu menjalankan program dasar R dari konsol dan mampu meng-install RCommander.

Pendahuluan

R adalah bahasa pemrograman *open source* yang biasa digunakan untuk komputasi dan pengolahan data statistik serta berhubungan dengan penampilan grafik menggunakan tools yang disediakan oleh paket-paketnya. R terdaftar dibawah GNU (*General Public License*). Versi awal R dibuat pada tahun 1992 di Universitas Auckland, New Zealand oleh Ross Ihaka dan Robert Gentleman.

R menyediakan beragam statistic, *machine learning* (pemodelan linier dan non linier, *classic statistic test*, *time series analyst*, klasifikasi, clustering). R memiliki berbagai fungsi *built-in* dan juga fungsi *extended* untuk tugas statistic, *machine learning* dan visualisasi, seperti:

1. Data extraction
2. Data cleaning
3. Data loading
4. Data transformation
5. Statistic analysis
6. Predictive modeling
7. Data visualization

Understanding features of R

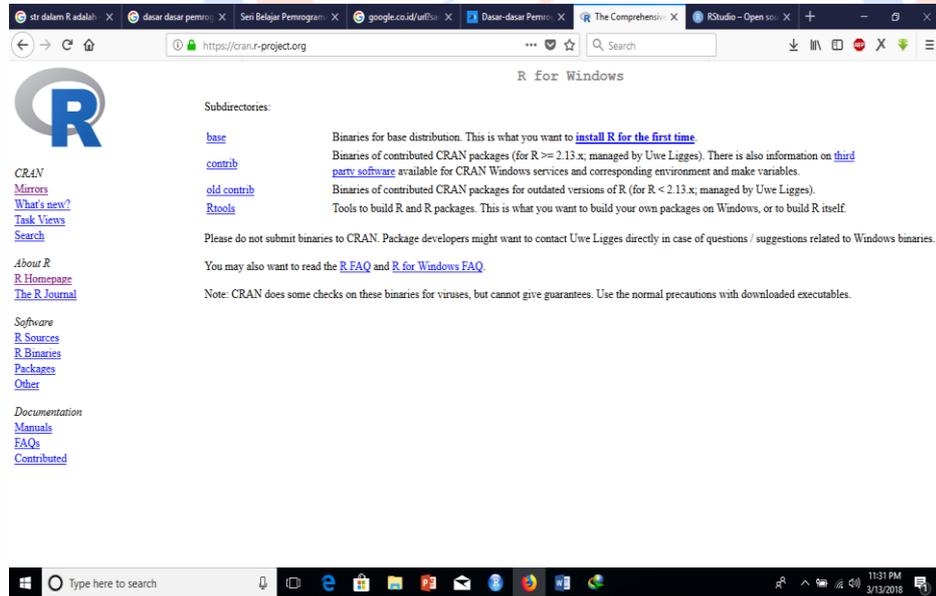
Beberapa fitur yang terdapat pada R,

1. Effective programming language
2. Relational database support

3. Data analytics
4. Data visualization
5. Package pada R sebagai penghubung dengan database yang besar.

Cara install R adalah:

1. Download paket R di <https://cran.r-project.org/>



Gambar 1 Situs resmi dan pengunduhan R

Setelah menginstal paket R dasar, disarankan untuk menginstal RStudio, yang merupakan Integrated Development Environment (IDE) yang hebat dan intuitif untuk R di <https://www.rstudio.com/>



Gambar 2 Situs resmi dan pengunduhan RStudio

2. Klik pada CRAN *section*, pilih CRAN mirror dan pilih sesuai dengan sistem operasi yang digunakan (windows).
3. Download R version dari mirror
4. Install R dengan ekstension .exe

Cara install RStudio

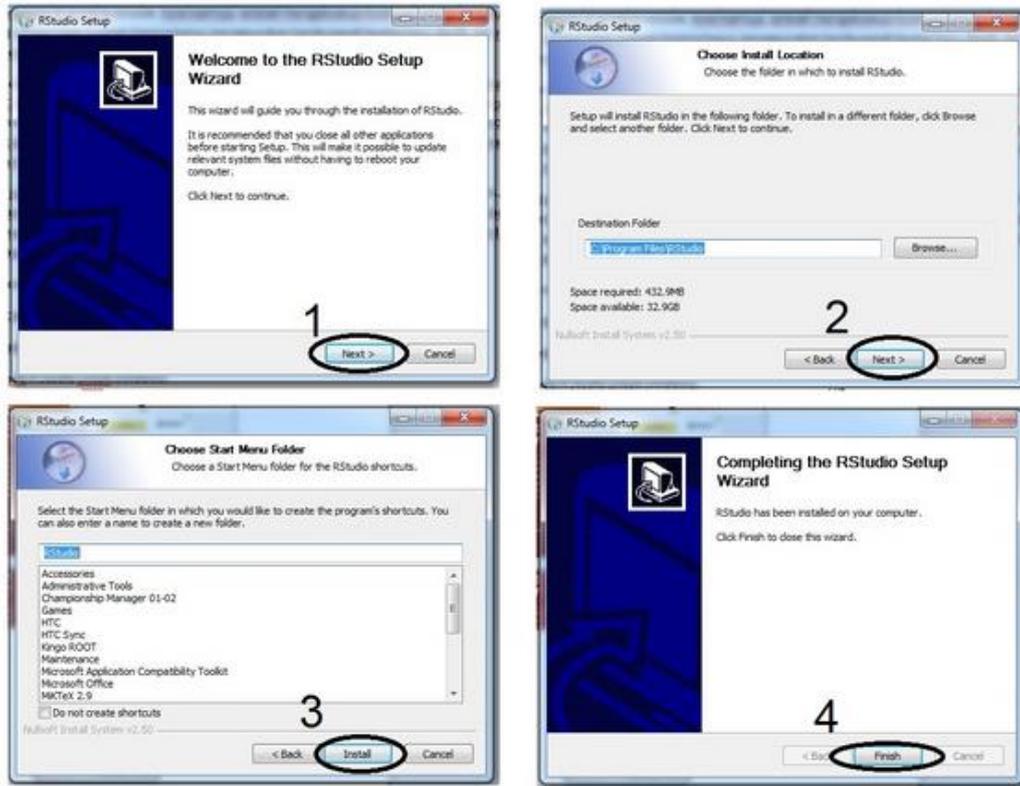
Rstudio diinstall setelah R diinstall. R dapat digunakan langsung, dengan cara mengetikkan kode-kode perintah pada jendela console. Kelemahannya adalah saat melakukan editing pada perintah yang sudah di-*create*. Hal ini dapat diatasi dengan menggunakan jendela R editor. R editor dapat dibuka dengan membuka file → new script, kemudian akan muncul jendela R editor.

Berikut cara mendownload Rstudio untuk windows:

1. Buka URL <https://www.rstudio.com/products/rstudio/download/>
2. Klik pada Rstudio untuk windows

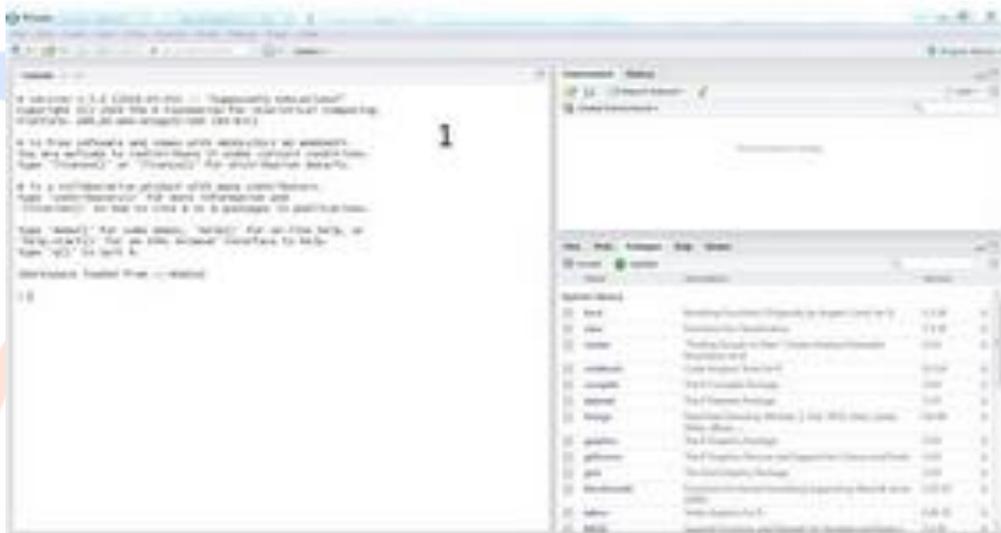
Cara menginstall

1. Klik kanan pada icon Rstudio dan pilih run as administrator.
2. Kemudian pilih next → next → install → finish



Gambar 3 proses install Rstudio pada windows

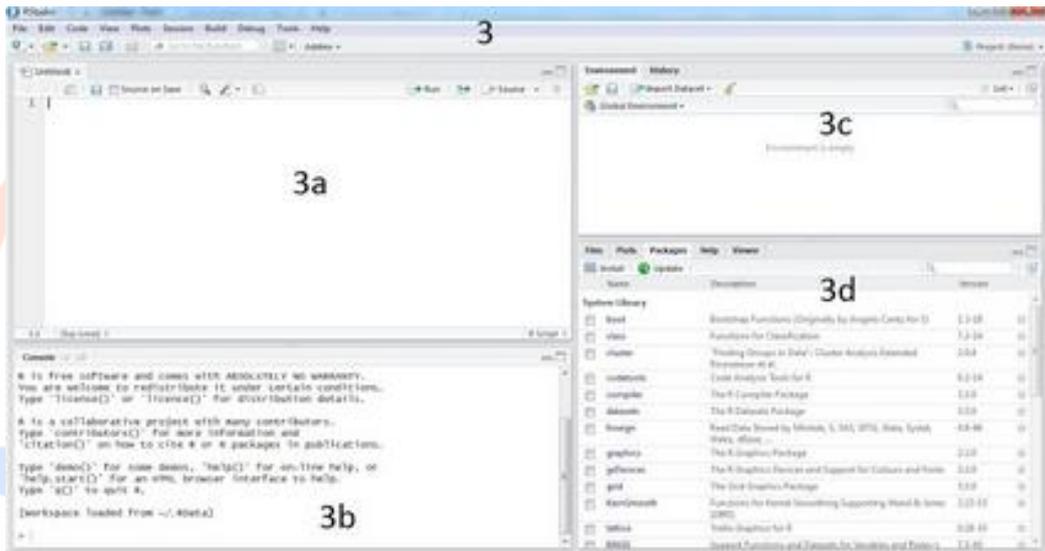
Jika proses install sudah selesai, maka Rstudio dapat dijalankan. Saat Rstudio pertama dijalankan, maka akan muncul jendela seperti Gambar 4.



Gambar 4 jendela awal Rstudio

Klik new file → new script, sehingga muncul jendela Rstudio yang baru.

Didalam Rstudio ada 4 jendela seperti yang terlihat pada Gambar 5, yaitu:



Gambar 5 Jendela editor pada Rstudio

Keterangan:

- 3a. = Untuk menuliskan kode perintah (jendela console)
- 3b. = Tempat kode perintah dieksekusi (jendela environment)
- 3c. = Tempat daftar objek-objek yang sedang aktif, jendela files, plots, packages, help dan viewer.
- 3d. = Tempat melihat root files, hasil plot, paket yang diinstal dan help (bantuan)

Untuk melihat kode perintah pada jendela editor dapat dieksekusi dengan menekan `alt+ enter` atau menggunakan tombol Run.

Fungsi-fungsi dasar pada R:

1. Visualisasi data:

Sintax yang digunakan antaranya: `barplot`, `pie`, `dotchart`, dan `histogram`.

2. Manipulasi data:

Sintax yang digunakan diantaranya: `sample`, `stack`, `unstack` dan `omit`

Beberapa *package* dan fungsi pada R yang digunakan, yaitu:

1. *Clustering*

Package yang digunakan untuk *clustering* diantaranya, **fpc**, **cluster**, **pvclust**, **mclust**

2. *Classification*

Package yang digunakan untuk *classification* diantaranya, **rpart**, **tree**, **marginTree**, **party**, **randomForest**, **maptree**

3. Asosiasi

Package yang digunakan untuk asosiasi diantaranya **arules** dan **drml**.

4. *Sequential pattern*

package yang digunakan untuk *sequential pattern* diantaranya **arulesSequences**

5. *Time series*

Package yang digunakan untuk *time series* diantaranya **ts**.

6. Statistik

package yang digunakan untuk statistik diantaranya **BaseR** dan **nlme**

Contoh kasus 1:

1. Buatlah file produksi.Rdata dan produksi.csv dari tabel berikut:

Bulan	Jawa	Sumatra	Sulawesi	Bali
Januari	12	6	4	2
Maret	9	5	3	5
Juni	8	4	9	7
September	12	5	2	5
Desember	10	8	6	6

2. Buatlah pie chart dari file produksi.Rdata

3. Buatlah histogram dari file produksi.Rdata.

Langkah-langkah menyelesaikan kasus 1:

1. Klik start → pilih Rstudio sehingga tampil layar editor seperti Gambar 5

2. Pada jendela console ketikkan perintah dibawah:

```
> Bulan <- c("Januari","Maret","Juni","September","Desember")
> Jawa <- c(12,9,8,12,10)
> Sumatra <- c(6,5,4,5,8)
> Sulawesi <- c(4,3,9,2,6)
> Bali <- c(2,5,7,5,6)
> a <- data.frame(Bulan,Jawa,Sumatra,Sulawesi,Bali)
> save(a,file = "produksi.Rdata")
> write.csv(a,"produksi.csv", row.names = FALSE)
> read.csv("produksi.csv")
```

Perintah `read.csv("produksi.csv")` berfungsi untuk menampilkan data `produksi.csv` sehingga tampil data tersebut pada jendela console, sebagai berikut.

```
> read.csv("produksi.csv")
  Bulan Jawa Sumatra Sulawesi Bali
1 Januari  12      6       4     2
2  Maret   9      5       3     5
3   Juni   8      4       9     7
4 September 12      5       2     5
5 Desember 10      8       6     6
```

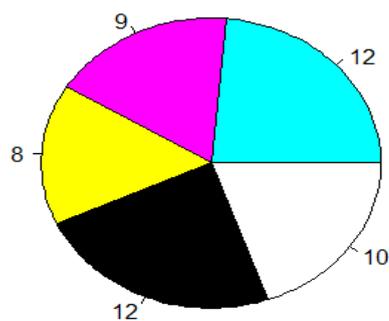
3. Tentukan warna setiap variabel dan jumlah produksi padi setiap wilayah dengan menggunakan perintah dibawah ini:

```
> color <- c("cyan","magenta","yellow","black","white")
> jawa_labels <- round(Jawa/sum(Jawa)*100,1)
> jawa_labels <- paste(jawa_labels,"%",sep="")
> sumatra_labels <- round(Sumatra/sum(Sumatra)*100,1)
> sumatra_labels <- paste(sumatra_labels,"%",sep="")
> sulawesi_labels <- round(Sulawesi/sum(Sulawesi)*100,1)
> sulawesi_labels <- paste(sulawesi_labels,"%",sep="")
> bali_labels <- round(Bali/sum(Bali)*100,1)
> bali_labels <- paste(bali_labels,"%",sep="")
```

4. Tampilkan bentuk pie chart dan histogram produksi padi pada jendela plot untuk masing-masing wilayah dengan menetikkan perintah dibawah ini:

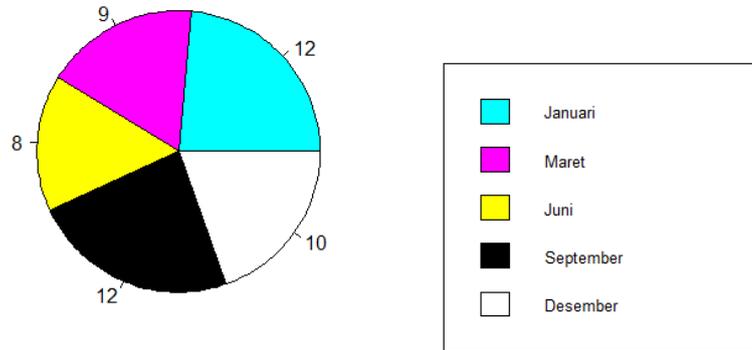
```
> pie(Jawa,main="Jumlah Prediksi Produksi Padi di Jawa
Tahun 2015\ndalam satuan (ribu ton)",col=color,label=
Jawa)
```

**Jumlah Prediksi Produksi Padi di Jawa Tahun 2014
dalam satuan (ribu ton)**



```
> legend(1.5, 0.5, c("Januari", "Maret", "Juni", "September", "Desember"), cex=0.8, fill=color)
```

Jumlah Prediksi Produksi Padi di Jawa Tahun 2014 dalam satuan (ribu ton)



```
> pie(Sumatra, main="Jumlah Prediksi Produksi Padi di Sumatra Tahun 2014\ndalam satuan (ribu ton)", col=color, label=Sumatra)
> legend(1.5, 0.5, c("Januari", "Maret", "Juni", "September", "Desember"), cex=0.8, fill=color)
> pie(Sulawesi, main="Jumlah Prediksi Produksi Padi di Sulawesi Tahun 2014\ndalam satuan (ribu ton)", col=color, label=Sulawesi)
> legend(1.5, 0.5, c("Januari", "Maret", "Juni", "September", "Desember"), cex=0.8, fill=color)
> pie(Bali, main="Jumlah Prediksi Produksi Padi di Bali Tahun 2014\ndalam satuan (ribu ton)", col=color, label=Bali)
> legend(1.5, 0.5, c("Januari", "Maret", "Juni", "September", "Desember"), cex=0., fill=color)
```

5. Menampilkan histogram untuk data produksi padi dengan mengetikkan perintah dibawah ini:

a. Panggil data produksi padi setiap wilayah

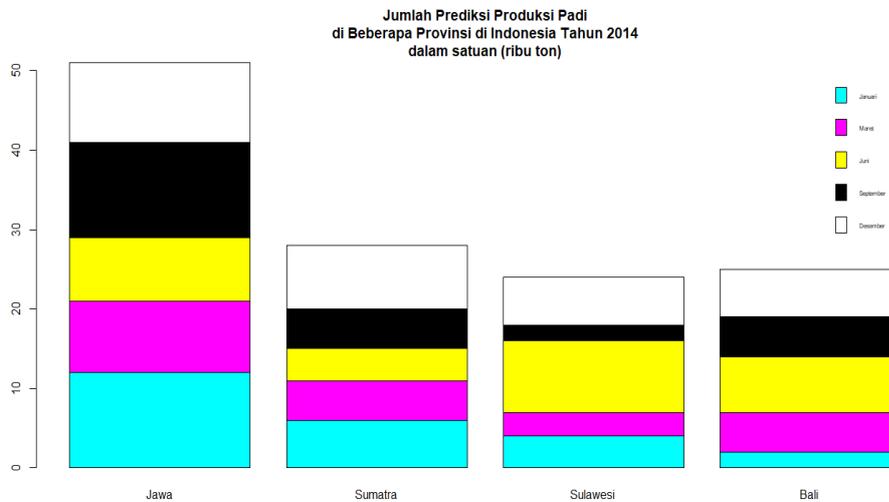
```
> b <- data.frame(Jawa, Sumatra, Sulawesi, Bali)
```

b. Tampilkan histogram untuk data produksi padi

```
> barplot(as.matrix(b), main="Jumlah Prediksi Produksi Padi\ndi Beberapa Provinsi di Indonesia Tahun 2014\ndalam satuan (ribu ton)", col=color)
```

c. Tampilkan legend histogram

```
> legend("topright",c("Januari","Maret","Juni","September",
"Desember"),cex=0.6,bty="n",fill=color)
```



Soal

1. Apa yang dimaksud dengan data mining?
2. Jelaskan, mengapa perlu dilakukan data mining?
3. Apa saja yang dilakukan pada data mining?
4. Jelaskan, apa yang anda ketahui tentang R?