



02

# AUTO-MPG

Informasi tentang otomotif

# Pemrosesan Awal

## Import packages

```
library(dplyr)
```

Untuk memanipulasi *data frame*

```
library(DataExplorer)
```

Untuk membuat *plot* dari persentase *missing value*

```
library(ggplot2)
```

Untuk membuat visualisasi data

## Import Data

```
auto <-  
read.table('auto-mpg.data')
```

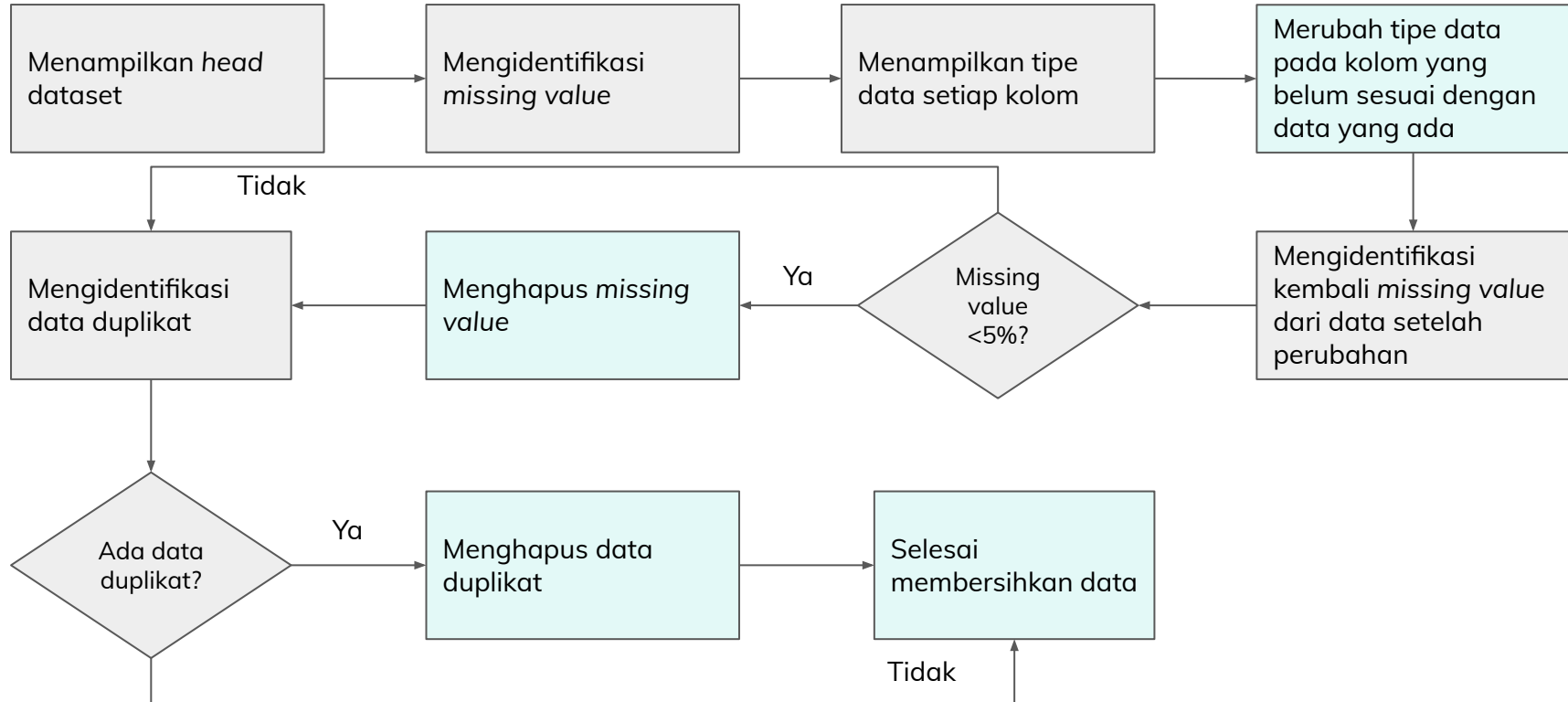
Membuat variabel 'auto' sebagai nama dataset yang akan digunakan. Data dapat dipilih secara manual setelah ada tampilan *window* baru ataupun secara langsung.

```
colnames(auto) <- c('mpg','cyl',  
'dis', 'hp', 'tor', 'wt', 'acc',  
'my', 'or', 'name')
```

Mendefinisikan nama kolom dengan nama baru

## Nomor 2.1

# Membersihkan data dan *output* dari `str()`



## Nomor 2.1

# Membersihkan data dan *output* dari `str()`

### 1. Melihat dan mengubah bentuk data

Syntax	Output
<p>Menampilkan variabel baru berisi nilai untuk 10 kolom dataset</p> <pre>head(auto)</pre>	<p>Tampilan <i>list</i> nilai data set</p> <pre>mpg cyl dis hp tor wt acc my or 1 18 8 307 130 74.0221 3504. 12.0 70 1 2 15 8 350 165 150.4898 3693. 11.5 70 1 3 18 8 318 150 143.9260 3436. 11.0 70 1 4 16 8 304 150 101.0950 3433. 12.0 70 1 5 17 8 302 140 113.3691 3449. 10.5 70 1 6 15 8 429 198 125.0070 4341. 10.0 70 1 name 1 chevrolet chevelle malibu 2 buick skylark 320 3 plymouth satellite 4 amc rebel sst 5 ford torino 6 ford galaxie 500</pre>
<p>Menentukan jumlah <i>missing value</i> yang terdapat pada setiap kolom</p> <pre>colSums(is.na(auto))</pre>	<p>Tampilan jumlah <i>missing value</i></p> <pre>mpg cyl dis tor hp wt acc my or name 8 0 1 6 8 0 0 0 0 0</pre>

## Nomor 2.1

# Membersihkan data dan *output* dari str()

### 1. Melihat dan mengubah bentuk data

Syntax	Output
<p>Menampilkan tipe data seperti apa yang ada pada setiap kolom</p> <pre>str(auto)</pre>	<p>Tampilan <i>list</i> tipe data setiap kolom variabel (auto)</p> <pre>'data.frame':   411 obs. of  10 variables:  \$ mpg : num  18 15 18 16 17 15 14 14 14 15 ...  \$ cyl : int   8  8  8  8  8  8  8  8  8 ...  \$ dis : num  307 350 318 304 302 429 454 440 455 390 ...  \$ hp  : chr   "130" "165" "150" "150" ...  \$ tor : num   74 150 144 101 113 ...  \$ wt  : chr   "3504." "3693." "3436." "3433." ...  \$ acc : num  12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...  \$ my  : int   70 70 70 70 70 70 70 70 70 ...  \$ or  : int    1  1  1  1  1  1  1  1  1 ...  \$ name: chr   "chevrolet chevelle malibu" "buick skylark 320" "plymo uth satellite" "amc rebel sst" ...</pre>
<p>Membuat variabel baru karena terdeteksi definisi pada kolom 'hp' tidak sesuai dengan tipe data karena ada <i>missing value</i></p> <pre>auto2 &lt;- auto %&gt;% replace(== '?', NA)</pre>	<p>Variabel baru: auto2</p>
<p>Menampilkan perubahan nilai <i>missing value</i> pada kolom 'hp' menjadi NA</p> <pre>unique(auto2\$hp)</pre>	<p>Tampilan isi data kolom 'hp' setelah perubahan nilai NA</p> <pre>[1] "130" "165" "150" "140" "198" "220" "215" "225" "190" "115" [11] "153" "175" "170" "160" "95" "97" "85" "88" "46" "87" [21] "90" "113" "200" "210" "193" NA "48" "100" "105" "180"</pre>

## Nomor 2.1

# Membersihkan data dan *output* dari str()

### 1. Melihat dan mengubah bentuk data

Syntax	Output
<p>Membuat variabel baru karena terdeteksi definisi pada kolom 'hp' tidak sesuai dengan tipe data karena ada <i>missing value</i></p> <pre>auto3 &lt;- auto2 %&gt;% replace(.== '?.', NA)</pre>	<p>Variabel baru: auto3</p>
<p>Menampilkan perubahan nilai <i>missing value</i> pada kolom 'wt' menjadi NA</p> <pre>unique(auto3\$wt)</pre>	<p>Tampilan isi data kolom 'hp' setelah perubahan nilai NA</p> <pre>[225] "4335." "1940." "2740." "2755." "2051." "2075." "1985." [232] "2190." "2815." "2600." "2720." "1800." "2070." "3365." [239] "3735." "3570." "3535." "3155." "2965." "3430." "3210." [246] "3380." "NA" "3620." "3410." "3445." "3205." "4080."</pre>
<p>Menampilkan tipe data seperti apa yang ada pada setiap kolom</p> <pre>str(auto3)</pre>	<p>Tampilan <i>list</i> tipe data setiap kolom variabel (auto)</p> <pre>'data.frame': 411 obs. of 10 variables:  \$ mpg : num 18 15 18 16 17 15 14 14 14 15 ...  \$ cyl : int 8 8 8 8 8 8 8 8 8 8 ...  \$ dis : num 307 350 318 304 302 429 454 440 455 390 ...  \$ hp : chr "130" "165" "150" "150" ...  \$ tor : num 74 150 144 101 113 ...  \$ wt : chr "3504." "3693." "3436." "3433." ...  \$ acc : num 12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...  \$ my : int 70 70 70 70 70 70 70 70 70 70 ...  \$ or : int 1 1 1 1 1 1 1 1 1 1 ...  \$ name: chr "chevrolet chevelle malibu" "buick skylark 320" "plymo uth satellite" "amc rebel sst" ...</pre>

## Nomor 2.1

# Membersihkan data dan *output* dari `str()`

### 1. Melihat dan mengubah bentuk data

Syntax	Output
<p>Membuat variabel baru berisi perubahan tipe data kolom 'hp' menjadi <i>numeric</i></p> <pre>auto4&lt;- transform(auto3, hp =   as.numeric(hp))</pre>	Variabel baru: auto5
<p>Membuat variabel baru berisi perubahan tipe data kolom 'wt' menjadi <i>numeric</i></p> <pre>auto5&lt;- transform(auto4, wt =   as.numeric(wt))</pre>	
<p>Membuat variabel baru berisi perubahan tipe data kolom 'or' menjadi <i>factor</i></p> <pre>auto6&lt;- transform(auto5, or =   as.factor(or))</pre>	

## Nomor 2.1

# Membersihkan data dan *output* dari `str()`

### 1. Melihat dan mengubah bentuk data

Syntax	Output
<p>Menampilkan tipe data seperti apa yang ada pada setiap kolom</p> <pre>str(auto6)</pre>	<p>Tampilan <i>list</i> tipe data setiap kolom variabel (auto)</p> <pre>'data.frame':  411 obs. of  10 variables:  \$ mpg : num  18 15 18 16 17 15 14 14 14 15 ...  \$ cyl : int   8  8  8  8  8  8  8  8  8  8 ...  \$ dis : num  307 350 318 304 302 429 454 440 455 390 ...  \$ hp  : num  130 165 150 150 140 198 220 215 225 190 ...  \$ tor : num   74 150 144 101 113 ...  \$ wt  : num  3504 3693 3436 3433 3449 ...  \$ acc : num   12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...  \$ my  : int   70 70 70 70 70 70 70 70 70 70 ...  \$ or  : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1  \$ name: chr  "chevrolet chevelle malibu" "buick skylark 320" el sst" ...</pre>



## Nomor 2.1

# Membersihkan data dan *output* dari `str()`

## 2. Menghapus *missing value* yang terdeteksi

Syntax	Output																						
<p>Menghitung jumlah data yang kosong pada setiap kolom</p> <pre>colSums(is.na(auto6))</pre>	<p>Tampilan jumlah <i>missing value</i> tiap kolom</p> <pre>mpg  cyl  dis  hp  tor  wt  acc  my  or  name 8    0    1    8    8    1    0    0    0    0</pre>																						
<p>Membuat grafik persentase <i>missing value</i> untuk menentukan proses perlakuan terhadap <i>missing value</i></p> <pre>plot_missing(auto6)</pre>	<p>Missing value pada profits kurang dari 5% → melakukan omit</p> <table border="1"><thead><tr><th>Feature</th><th>Missing Rows (%)</th></tr></thead><tbody><tr><td>name</td><td>0%</td></tr><tr><td>or</td><td>0%</td></tr><tr><td>my</td><td>0%</td></tr><tr><td>acc</td><td>0%</td></tr><tr><td>cyl</td><td>0%</td></tr><tr><td>wt</td><td>0.24%</td></tr><tr><td>dis</td><td>0.24%</td></tr><tr><td>tor</td><td>1.95%</td></tr><tr><td>hp</td><td>1.95%</td></tr><tr><td>mpg</td><td>1.95%</td></tr></tbody></table>	Feature	Missing Rows (%)	name	0%	or	0%	my	0%	acc	0%	cyl	0%	wt	0.24%	dis	0.24%	tor	1.95%	hp	1.95%	mpg	1.95%
Feature	Missing Rows (%)																						
name	0%																						
or	0%																						
my	0%																						
acc	0%																						
cyl	0%																						
wt	0.24%																						
dis	0.24%																						
tor	1.95%																						
hp	1.95%																						
mpg	1.95%																						

## Nomor 2.1

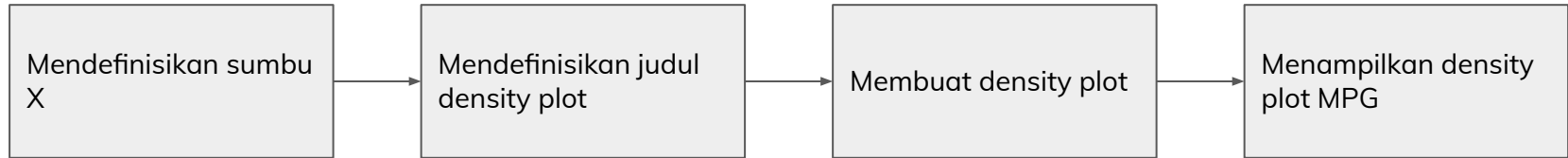
# Membersihkan data dan *output* dari `str()`

## 2. Menghapus *missing value* yang terdeteksi

Syntax	Output
<p>Membuat variabel baru yang sudah bersih dari data yang hilang</p> <pre>autompg&lt;-na.omit(auto6)</pre> <pre>colSums(is.na(autompg))</pre>	<p>Tampilan jumlah <i>missing value</i> tiap kolom</p> <pre>mpg  cyl  dis  hp  tor  wt  acc  my  or  name 0    0    0    0    0    0    0    0    0    0</pre>
<p>Mengecek apakah dalam dataset terdapat duplikasi data</p> <pre>sum(duplicated(autompg))</pre>	<p>Tampilan jumlah data duplikat</p> <pre>[1] 0</pre>

## Nomor 2.2

# Membuat density plot untuk variabel respon MPG



## Nomor 2.2

# Membuat density plot untuk variabel respon MPG

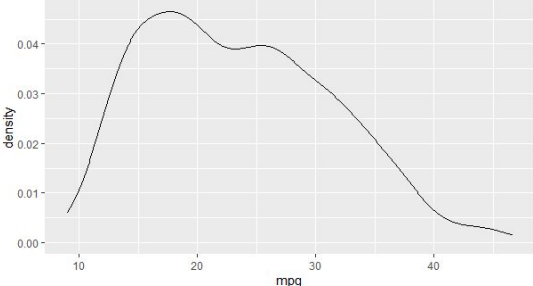
1. Melakukan pendefinisian sumbu x

Syntax	Output
<p>Membuat variabel dan mendefinisikan sumbu x menjadi mpg</p> <pre>vis_mpg &lt;- ggplot(data=autompg, aes(x=mpg))+ geom_density()</pre>	<p>Variabel: vis_mpg Density plot untuk sumbu x menjadi mpg</p>

## Nomor 2.2

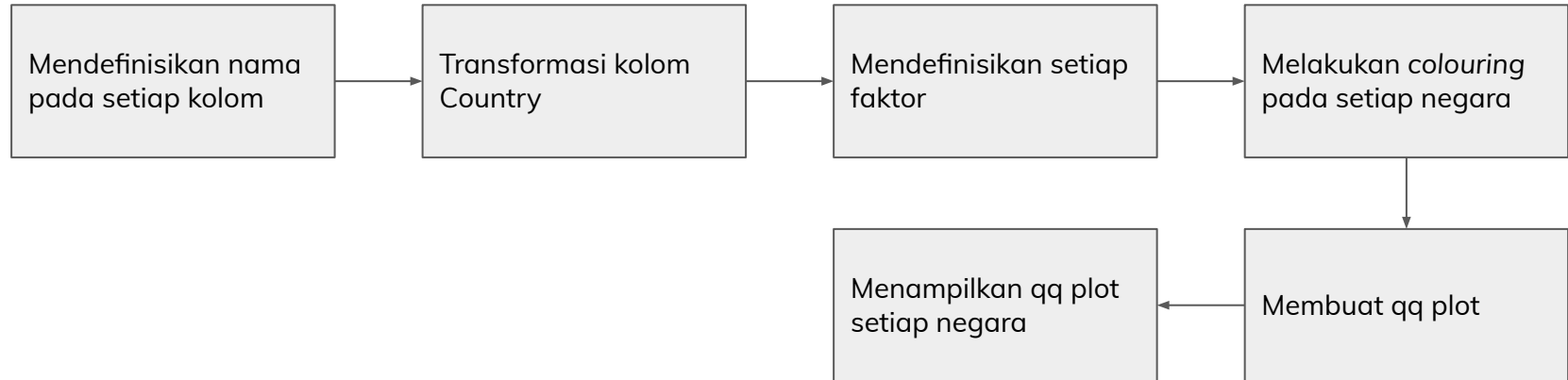
# Membuat density plot untuk variabel respon MPG

## 2. Menampilkan density plot MPG

Syntax	Output
<p>Membuat variabel baru yang berisi judul density plot</p> <pre>vis_mpg2&lt;- vis_mpg + labs(title = 'MPG Density Plot')</pre>	<p>Variabel baru: vis_mpg2 Judul density plot menjadi "MPG Density Plot"</p>
<p>Membuat density plot MPG</p> <pre>vis_mpg2</pre>	<p>Variabel respons tidak berdistribusi normal</p>  <p>The figure is a density plot titled "MPG Density Plot". The x-axis is labeled "mpg" and ranges from approximately 10 to 45, with major ticks at 10, 20, 30, and 40. The y-axis is labeled "density" and ranges from 0.00 to 0.04, with major ticks at 0.00, 0.01, 0.02, 0.03, and 0.04. The plot shows a unimodal distribution that is slightly right-skewed, with a peak density of approximately 0.045 at an mpg value of about 18. The distribution is non-normal, as indicated by the text above the plot.</p>

## Nomor 2.3

# Membuat qq plot untuk setiap negara MPG



## Nomor 2.3

# Membuat qq plot untuk setiap negara MPG

1. Melakukan pendefinisian nama pada setiap kolom

Syntax	Output
<p>Membuat variabel baru</p> <pre>no &lt;- autompg</pre>	<p>Variabel baru: no</p>
<p>Mendefinisikan nama pada setiap kolom</p> <pre>colnames(no) &lt;- c('mpg', 'cyl',   'dis', 'hp', 'tor', 'wt', 'acc',   'my', '<b>Country</b>', 'name')</pre>	<p>Setiap kolom memiliki nama sesuai dengan yang telah didefinisikan</p>

## Nomor 2.3

# Membuat qq plot untuk setiap negara MPG

2. Melakukan pendefinisian setiap faktor dan warnanya

Syntax	Output
<p>Mendefinisikan setiap faktor menjadi nama negara</p> <pre>levels(no\$Country) &lt;- c('USA',                         'Europe', 'Japan')</pre>	<p>Faktor menjadi 1 untuk USA 2 untuk Europe 3 untuk Japan</p>
<p>Membuat variabel baru yang berisi data mpg dan warna pada setiap negara</p> <pre>plot_mpg &lt;- ggplot(no, aes(sample = mpg, colour = Country)) +   stat_qq()</pre>	<p>Variabel baru: plot_mpg</p>



## Nomor 2.3

# Membuat qq plot untuk setiap negara MPG

### 3. Menampilkan qq plot untuk setiap negara

Syntax	Output
<p>Membuat variabel baru yang berisi tampilan qq plot dari setiap negara</p> <pre>plot_mpg1 &lt;- plot_mpg + labs(title = 'QQ Plot for MPG', fill = 'Country')</pre>	<p>QQ plot dengan faktor setiap negara</p> 