

3-Bangkitkan Data & Fungsi Built in

Ashari Ramadhan

11/20/2020

Membangkitkan Data

Seperti pada materi sebelumnya, kita bisa membuat vektor dengan mengisi nilainya secara manual seperti berikut:

```
vektor_1 <- c(1,4,7,9,1,3,3,6,5)
vektor_1
```

```
## [1] 1 4 7 9 1 3 3 6 5
```

Namun ada kalanya kita ingin membangkitkan data secara acak atau dengan pola tertentu.

Secara acak

Membangkitkan data secara acak dapat menggunakan fungsi `sample()`

```
vektor_2 <- sample(1:100, 50, replace = TRUE)
vektor_2
```

```
## [1] 45 89 15 92 13 30 12 91 25 3 34 99 79 56 32 61 31 87 9 42 56 42 59 49 38
## [26] 66 4 74 17 32 77 25 77 4 70 51 14 99 19 45 47 33 59 44 75 89 69 53 74 2
```

Catatan:

- 1:100 adalah parameter rentang nilai yang dibangkitkan
- 50 adalah banyaknya data yang ingin dibangkitkan
- `replace = TRUE`, artinya data yang sudah bangkitkan dapat kembali.

Berpola

Misalnya kita ingin membangkitkan data dengan selisih 4.

```
vektor_3 <- seq(1, 100, 4)
vektor_3
```

```
## [1] 1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97
```

Catatan:

- 1 adalah nilai terkecil
- 100 adalah terbesar
- 4 adalah selisih nilai

Mengikuti distribusi tertentu

Dalam statistika ada banyak distribusi, salah satu yang terkenal adalah distribusi normal. Jika kita membuat histogram data berdistribusi normal maka akan membentuk pola mirip lonceng dan puncak histogram berada di tengah.

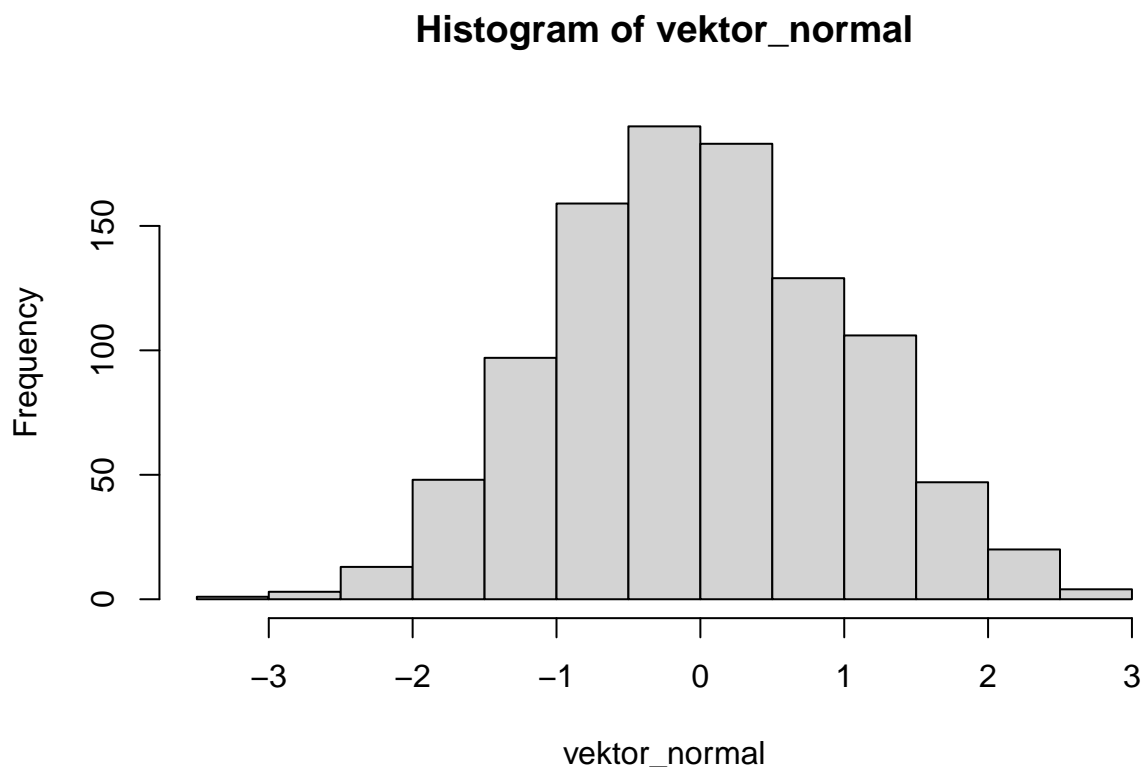
```
vektor_normal <- rnorm(1000, 0, 1)
# menampilkan 10 data
vektor_normal[1:10]
```

```
## [1] -0.03186792  0.92237752  0.05131862 -2.62486002  1.05016899 -1.29189851
## [7] -0.60378698  0.02611163 -0.77097910  0.66388574
```

Catatan:

- 1000 adalah jumlah data yang ingin dibangkitkan
- 0 adalah nilai mean / rata-rata
- 1 adalah nilai standar deviasi

```
hist(vektor_normal, breaks = 20)
```



Masih banyak distribusi data yang lain seperti distribusi gamma, eksponensial, chi dll.

Mengunci bilangan acak

Seperti namanya “acak”, setiap kita membangkitkan data acak maka akan menghasilkan output yang berbeda, meskipun dengan kode yang sama.

```
contoh_1 <- sample(1:10, 5)
contoh_1
```

```
## [1] 3 8 1 10 7
```

```
contoh_2 <- sample(1:10, 5)
contoh_2
```

```
## [1] 3 6 10 7 8
```

```
contoh_3 <- sample(1:10, 5)
contoh_3
```

```
## [1] 1 2 3 10 9
```

Ada kalanya jika kita ingin membangkitkan bilangan acak secara konsisten. Untuk melakukan hal ini kita butuh fungsi `set.seed()`

```
set.seed(1000000000)
contoh_1 <- sample(1:10, 5)
contoh_1
```

```
## [1] 10 5 2 3 9
```

```
set.seed(10)
contoh_2 <- sample(1:10, 5)
contoh_2
```

```
## [1] 9 7 8 6 3
```

```
set.seed(10)
contoh_3 <- sample(1:10, 5)
contoh_3
```

```
## [1] 9 7 8 6 3
```

Dapat dilihat hasil yang dihasilkan sama. Nilai dalam `set.seed()` bebas dari rentang 1 hingga 1000000000

Fungsi Built In

Fungsi built in adalah fungsi yang secara default tersedia di R tanpa harus `install.packages` atau membangun fungsi sendiri. Ada begitu banyak fungsi built dan sudah kita gunakan sebelumnya seperti `c()`, `sample()`, `rnorm` dan lain-lain.

Fungsi built in yang akan praktekkan pada kesempatan kali ini, adalah fungsi untuk statistik deskriptif.

Kita kembali 2 vektor data numerik.

```
set.seed(100)
vektor_1 <- sample(1:10000, 100, replace = T)
vektor_2 <- sample(1:10000, 100, replace = T)
```

Nilai minimum

```
min(vektor_1)
```

```
## [1] 223
```

Nilai maksimum

```
max(vektor_1)
```

```
## [1] 9732
```

Nilai rata-rata

```
mean(vektor_1)
```

```
## [1] 5029.56
```

Nilai median

```
median(vektor_1)
```

```
## [1] 5258
```

Nilai total

```
sum(vektor_1)
```

```
## [1] 502956
```

Panjang / banyak data

```
length(vektor_1)
```

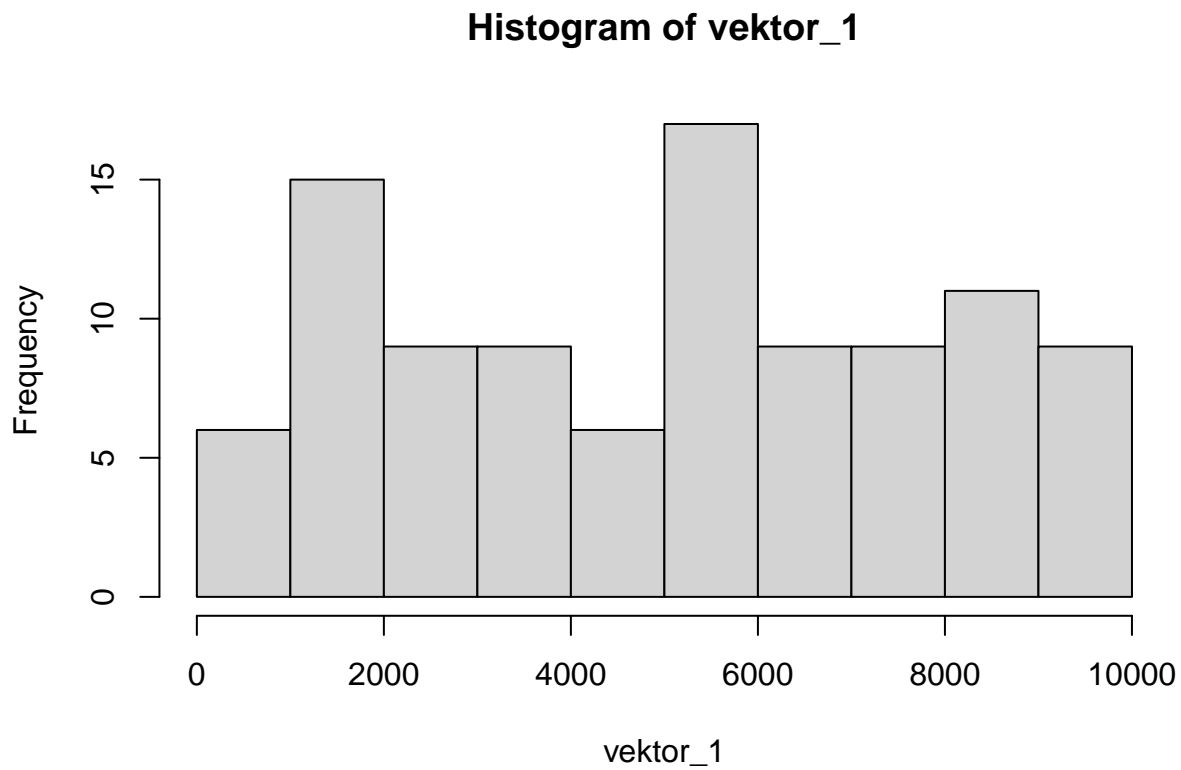
```
## [1] 100
```

Plot Base

Ada beberapa fungsi built in yang digunakan untuk membuat plot, Untuk data angka/numerik, kita bisa menggunakan `hist()`, `boxplot()` dan `plot`.

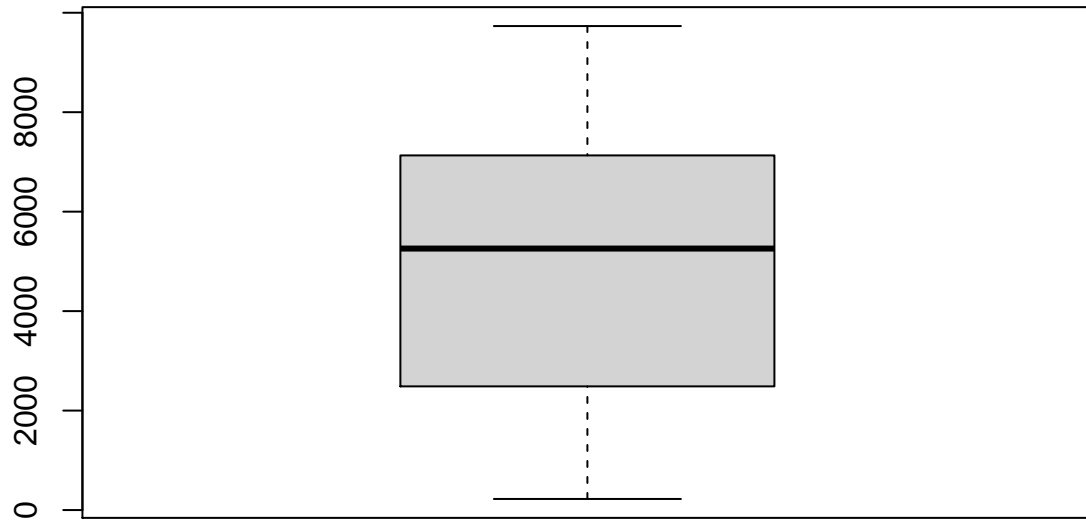
Histogram

```
hist(vektor_1)
```



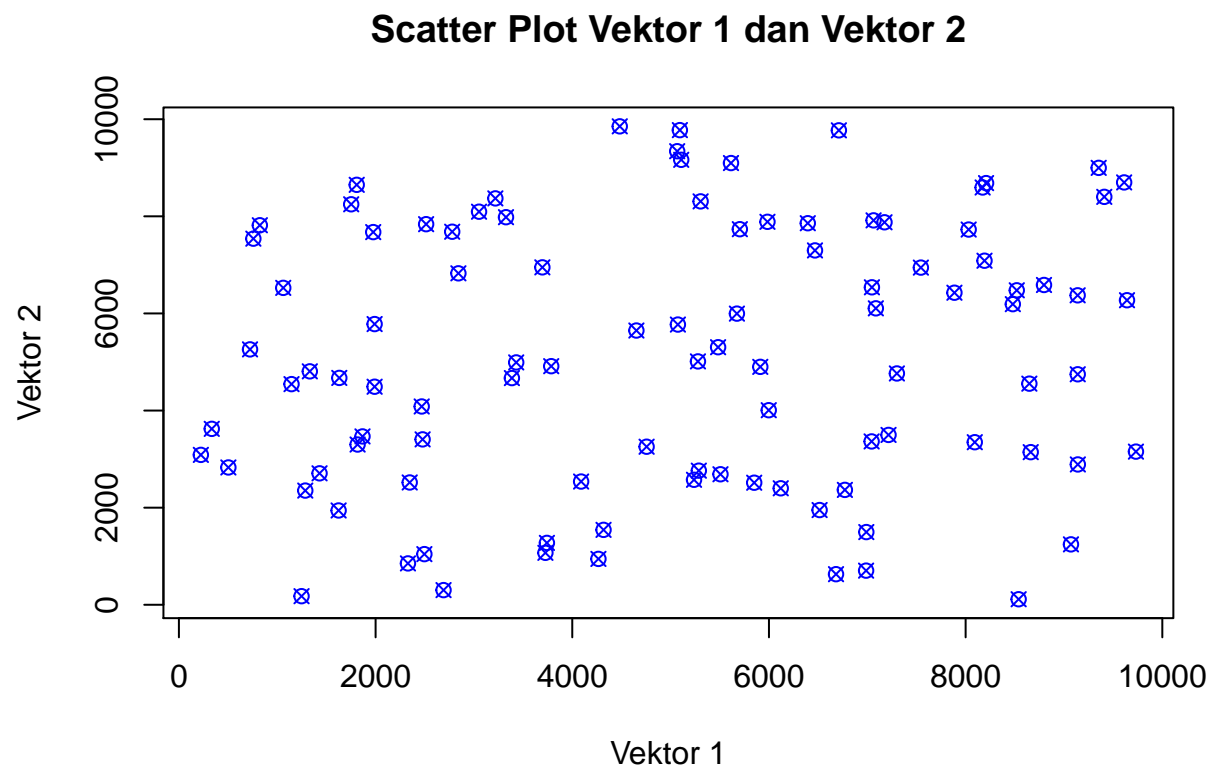
Boxplot

```
boxplot(vektor_1)
```



```
## Scatter plot
```

```
plot(vektor_1, vektor_2, main = "Scatter Plot Vektor 1 dan Vektor 2",  
     ylab = "Vektor 2", xlab = "Vektor 1", col="blue", pch = 13)
```



Catatan:

- main adalah judul plot
- ylab judul sumbu Y
- xlab judul sumbu x
- col adalah warna garis titik
- pch adalah bentuk titik