

## Deskripsi 22 kategori situs

- *Games*: merupakan kategori yang berisi situs terkait dengan games
- *Blog*: merupakan kategori yang berisi situs blog dan personal web
- *News*: merupakan kategori yang berisi situs yang menampilkan berita dari berbagai macam bidang (sport, gosip, otomotif, politik, agama, lowongan kerja dan lain lain). Pada kategori ini juga terdapat situs tv online dan radio online.
- *Pornography*: merupakan kategori yang berisi situs erotis atau porno, mencakup fetish, animasi, kartun, cerita, dan pornografi
- *Online Shopping*: merupakan kategori yang berisi situs penjualan produk atau layanan secara online (jual beli online).
- *Services*: merupakan kategori yang berisi situs yang menyediakan layanan e-mail, search engine, online tools (contoh: speedtest.net), cloud, anonymous proxy, portal ke situs lain, dan hosting serta situs perusahaan penyedia layanan internet.
- *Finance/Banking*: merupakan kategori yang berisi yang menyediakan informasi keuangan atau layanan yang berhubungan pengelolaan uang, seperti situs perusahaan Bank, kartu kredit, credit unions, asuransi, dan keuangan lainnya.
- *Travel*: merupakan kategori yang berisi jasa perjalanan pribadi atau bisnis, seperti hotel, resort, airlines, penyewaan mobil, agen perjalanan, turis dan informasi perjalanan
- *Information and Communication Technology (ICT)*: merupakan kategori yang berisi tentang informasi umum mengenai komputer, Gadget dan internet serta situs perusahaan yang bergerak dibidang IT (*software/hardware* seperti: Samsung, Apple dll), situs *software developer*, update software seperti scripts, plugin atau driver. seperti situs framework programming (contoh: yiiframework.com), dan situs perancangan website.
- *File Sharing*: merupakan kategori yang berisi situs untuk *upload* dan *download* fail, *repository shareware* dan *freeware* yang dapat didownload, situs untuk sharing image, serta situs peer-to-peer
- *Hacking*: merupakan kategori yang berisi situs illegal *license key generators*, *software cracks*, dan *serial number*
- *Entertainment*: merupakan kategori yang berisi situs *video streaming*, situs manga, situs *anime*, situs gambar lucu, situs *movie*, situs drama, pertunjukan seni/budaya/hiburan dan situs musik.
- *Publication*: merupakan kategori yang berisi situs terkait publikasi jurnal ilmiah, situs untuk mengunduh *ebook*, dan situs penerbit buku.
- *Education/references*: merupakan kategori yang berisi situs tutorial, situs informatif (contoh: Wikipedia), situs akademis seperti mata pelajaran akademis (matematika, sejarah), halaman web sekolah atau universitas, dan halaman administrasi pendidikan (dewan sekolah, kurikulum guru), situs konferensi, seminar/symposium dan situs penyedia kamus online.

- *Religion*: merupakan kategori situs yang berisi tentang topik agama dan informasi yang berkaitan dengan agama-agama Besar. Kategori ini mencakup situs yang mencakup konten keagamaan seperti diskusi, kepercayaan, artikel, dan informasi untuk jemaat lokal seperti homepage gereja atau sinagog. Agama-agama dalam kategori ini adalah Baha'I, Buddhisme, Cina Tradisional, Kristen, Hinduisme, Islam, Jainisme, Yudaisme, Shinto, Sikhisme, Tenrikyo, Zoroastrianisme.
- *Organization and Company*: merupakan kategori yang berisi situs organisasi/asosiasi profesi, instansi/lembaga penelitian, situs badan dunia, serta situs perusahaan yang sudah memiliki badan hukum seperti PT, CV, UD, dan lainnya.
- *Government/Military*: merupakan kategori yang berisi situs pemerintahan atau militer
- *Business*: merupakan kategori yang berisi informasi terkait usaha dan bisnis. Kategori ini juga mencakup informasi, layanan, atau produk yang membantu rencana bisnis, pengelolaan, dan pasar perusahaan mereka.
- *Social Network/Forum/Bulletin Boards*: merupakan kategori yang berisi situs social networking, situs forum (*Technical/Business Forums*), situs komunitas, dan situs *chatting*.

Kode Kategori	Kategori	Jumlah hit
k12	News	5,454,271
k9	Government/military	3,794,102
k5	File Sharing	2,974,468
k22	Uncategory	2,920,260
k19	Services	2,383,591
k4	Entertainment	2,176,330
k13	Online shopping	1,720,820
k1	Blogs/wiki	1,711,502
k3	Education/references	1,256,958
k15	pornography	1,240,496
k16	Publication	1,093,571
k2	Business	856,119
k11	ICT	726,264
k20	Social Network/Forum/Bulletin Boards	628,886
k14	Organization and company	521,244
k21	Travel	325,717
k18	Religion	265,629
k8	Games	184,309
k17	Real estate	98,982
k6	Finance/Banking	68,167
k10	Hacking	3,439
k7	Gambling	8

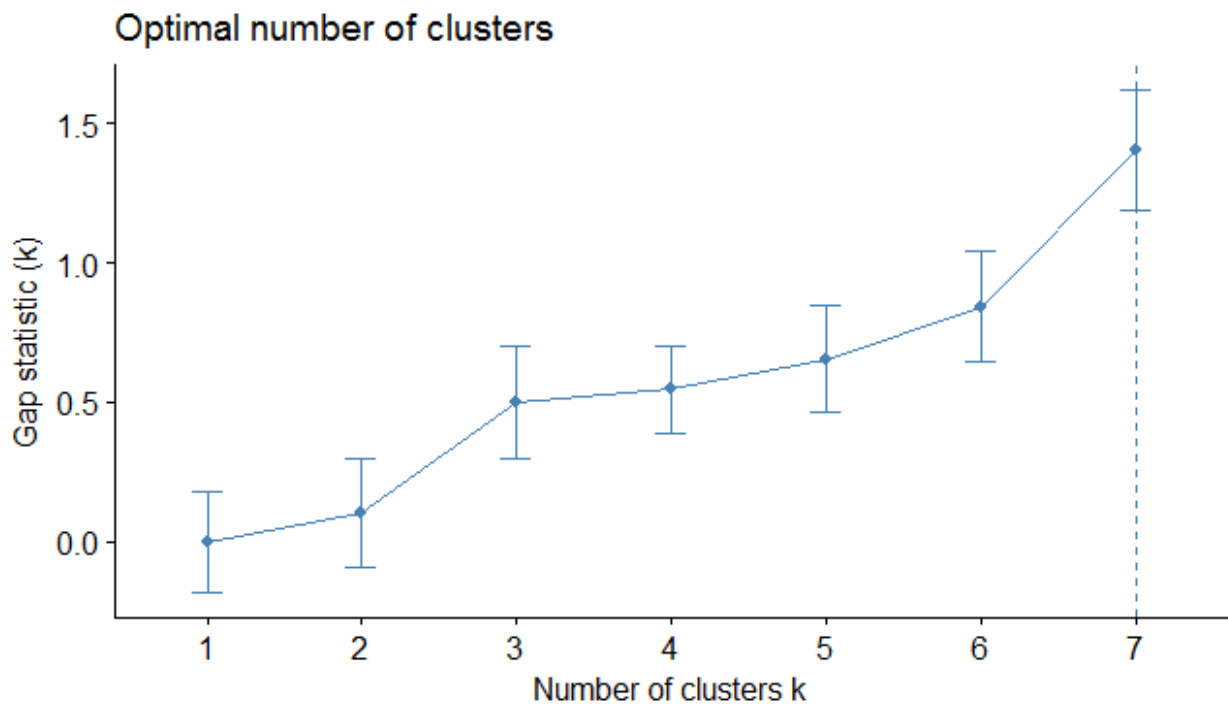
## Perintah RStudio

```
library(tidyverse)
library(cluster) # Algoritma clustering
library(factoextra) # Algoritma clustering dan visualisasi
```

Gap Statistic k optimal untuk menentukan jumlah cluster penggunaan internet di Pusbindiklat Peneliti LIPI berdasarkan kelompok Jabatan

1. Import Dataset from Excel (hit\_jabatan\_pusbin)
2. `df<-hit_jabatan_pusbin`
3. `set.seed(123)`
4. 

```
> gap_stat <- clusGap(df, FUN = kmeans, nstart = 25,
+                       K.max = 7, B = 9) # metode gap statistic
> fviz_gap_stat(gap_stat)
```



## Eksekusi K-Means

```
> final <- kmeans(df, 3, nstart = 25)
> print(final)
```

K-means clustering with 3 clusters of sizes 1, 7, 1

Cluster means:

	k1	k2	k3	k4	k5	k6	k8
1	4194.000	2440.0000	4177.000	52769.000	862.0000	130.00000	82.00000
2	1253.714	948.8571	1014.857	1589.857	692.5714	41.28571	33.71429
3	11605.000	6076.0000	4797.000	17703.000	13697.0000	296.00000	272.00000

	k9	k10	k11	k12	k13	k14	k15	k16
1	21292.000	0	1542.000	10186.000	6210.0000	622.0000	230.0000	2218.000
2	4705.286	15	1472.571	4170.286	844.2857	293.4286	215.2857	1147.143
3	41354.000	11	5227.000	45646.000	37987.0000	1673.0000	1638.0000	6482.000

	k17	k18	k19	k20	k21	k22
1	312.00000	98.0000	3843.000	226.0000	547	4593.000
2	17.85714	138.7143	3126.857	171.5714	313	2158.286
3	432.00000	1610.0000	17480.000	5087.0000	6768	43111.000

Clustering vector:

Analisis Kepegawaian	Fungsional Umum	Perencana	Pranata Humas
2	3	2	2
Pranata Komputer	widyaiswara	Honorar	Peneliti
2	1	2	2
Tamu			
2			

within cluster sum of squares by cluster:

```
[1] 0 578509157 0
(between_SS / total_SS = 93.6 %)
```

Available components:

```
[1] "cluster" "centers" "totss" "withinss" "tot.withinss"
"betweenss" "size" "iter" "ifault"
```

```
> table(hit_jabatan_pusbin$Jabatan, final$cluster)
```

	1	2	3
Analisis Kepegawaian	0	1	0
Fungsional Umum	0	0	1
Honorar	0	1	0
Peneliti	0	1	0
Perencana	0	1	0
Pranata Humas	0	1	0
Pranata Komputer	0	1	0
Tamu	0	1	0
Widyaiswara	1	0	0

```

> aggregate(df, by=list(cluster=final$cluster), mean)
  cluster      k1      k2      k3      k4      k5      k6
k8
1      1 4194.000 2440.0000 4177.000 52769.000  862.0000 130.00000  82.000
00
2      2 1253.714  948.8571 1014.857  1589.857  692.5714  41.28571  33.714
29
3      3 11605.000 6076.0000 4797.000 17703.000 13697.0000 296.00000 272.000
00
      k9 k10      k11      k12      k13      k14      k15      k16
1 21292.000  0 1542.000 10186.000  6210.0000  622.0000  230.0000 2218.000
2  4705.286 15 1472.571  4170.286   844.2857  293.4286  215.2857 1147.143
3 41354.000 11 5227.000 45646.000 37987.0000 1673.0000 1638.0000 6482.000
      k17      k18      k19      k20 k21      k22
1 312.00000  98.0000 3843.000  226.0000  547 4593.000
2  17.85714 138.7143 3126.857  171.5714  313 2158.286
3 432.00000 1610.0000 17480.000 5087.0000 6768 43111.000

```

karateristik cluster berdasarkan kelompok jabatan

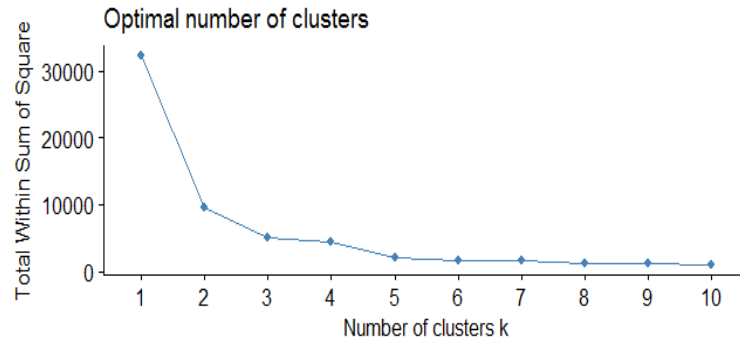
Cluster	Rendah	Sedang	Tinggi	Anggota
1	k10, k18	k1, k2, k3, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k17, k19, k20, k21, k22	k4	Widyaiswara
2	k1, k2, k3, k4, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k17, k19, k20, k21, k22	k18	k10	Peneliti Perencana Pranata Humas Pranata Komputer Honorar Tamud
3		k10	k1, k2, k3, k4, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k17, k18, k19, k20, k21, k22	Fungsional umum

3 Metode untuk menentukan k optimal jumlah cluster frekuensi akses internet oleh pegawai Pusbindiklat Peneliti LIPI

1. Import Dataset from Excel (log\_frekuensi\_akses2)
2. frek<- log\_frekuensi\_akses2

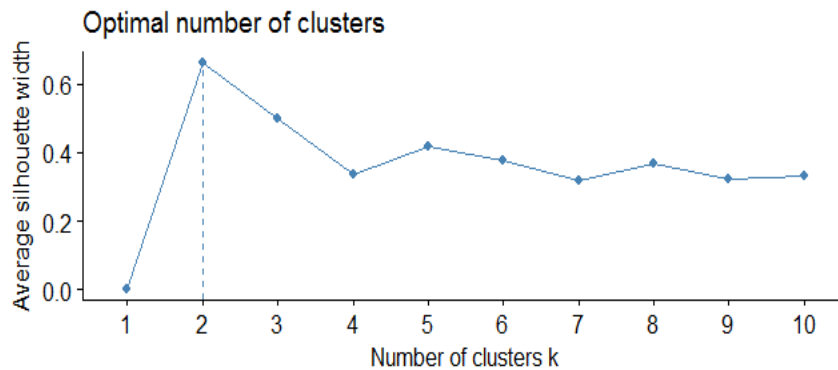
**Metode Elbow**

```
> fviz_nbclust(frek, kmeans, method = "wss")
```



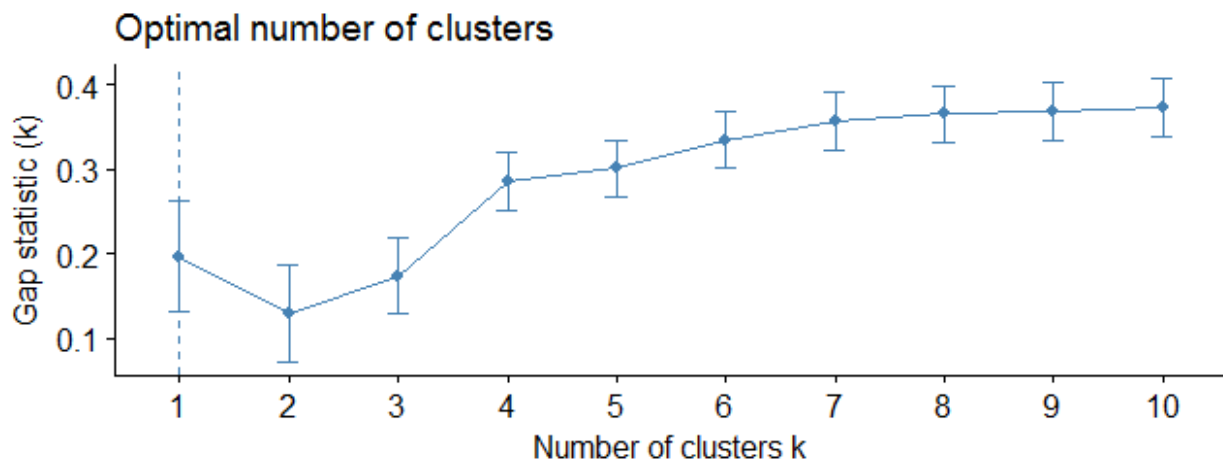
**Metode silhouette**

```
> fviz_nbclust(frek, kmeans, method = "silhouette")
```



**Metode Gap Statistik**

```
> set.seed(123)
> gap_stat <- clusGap(frek, FUN = kmeans, nstart = 25,
+                   K.max = 10, B = 49)
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 49) [one "." per sample]:
..... 49
> fviz_gap_stat(gap_stat)
```



## Eksekusi K-Means

```
> frekuen <- kmeans(frek, 3, nstart = 25)
> print(frekuen)
K-means clustering with 3 clusters of sizes 4, 17, 28

Cluster means:
  Agustus September  Oktober November Desember  jumlah
1 1.000000  0.75000  2.25000  0.50000  1.25000  5.75000
2 5.294118 10.94118 13.64706 13.41176 10.64706 53.94118
3 9.035714 15.25000 19.35714 19.50000 16.82143 79.96429

Clustering vector:
 [1] 3 3 2 3 3 3 1 3 3 2 2 2 2 3 3 3 2 2 3 3 2 3 3 1 3 2 3 1 1 3 2 2 3 3 3 3
 2 2
[39] 3 3 3 3 3 2 2 2 3 3 2

within cluster sum of squares by cluster:
 [1] 50.000 3169.294 1838.714
 (between_ss / total_ss = 84.3 %)

Available components:

 [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
 [6] "betweenss"    "size"         "iter"         "ifault"
```

```
> aggregate(frek, by=list(cluster=frekuen$cluster), mean)
 cluster Agustus September  Oktober November Desember  jumlah
1      1 1.000000  0.75000  2.25000  0.50000  1.25000  5.75000
2      2 5.294118 10.94118 13.64706 13.41176 10.64706 53.94118
3      3 9.035714 15.25000 19.35714 19.50000 16.82143 79.96429
```

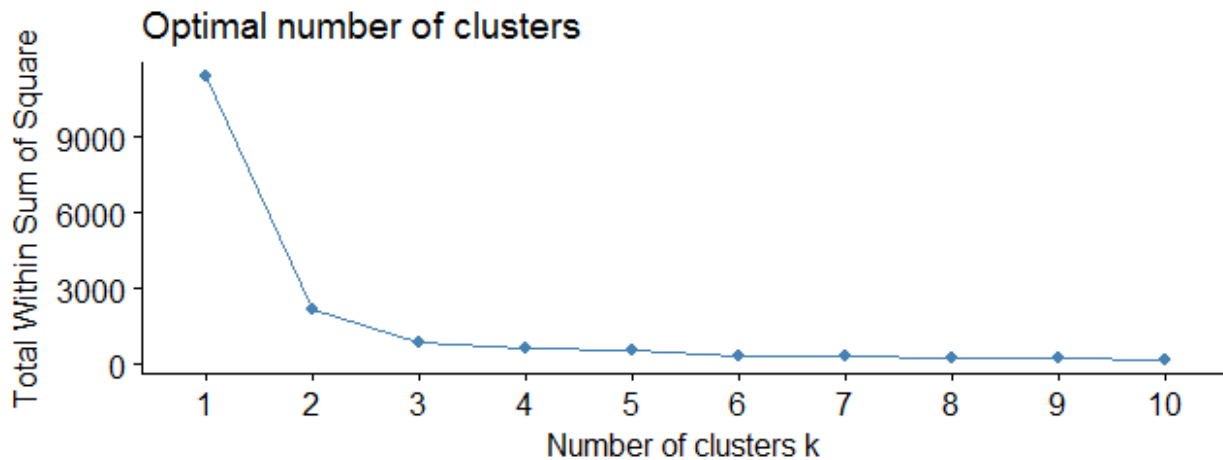
Frekuensi akses internet pegawai Pusbindiklat			
Cluster	Level	Jumlah	Persentase
1	Rendah (< 10 hari dalam 5 bulan)	4	8%
2	Sedang (10-66 hari dalam 5 bulan)	17	35%
3	Tinggi (> 66 hari dalam 5 bulan)	28	57%

3 Metode untuk menentukan k optimal jumlah cluster lama waktu akses internet terhadap situs yang tidak mendukung pekerjaan oleh pegawai Pusbindiklat Peneliti LIPI

1. Import Dataset from Excel (log\_waktu\_mingguan\_webpribadi)
2. waktu<- log\_waktu\_mingguan\_webpribadi

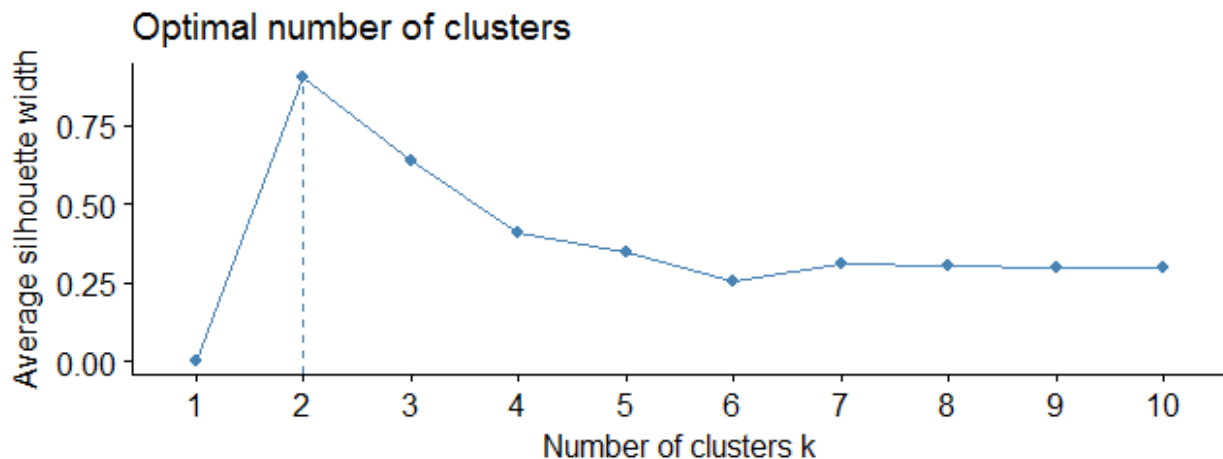
### Metode Elbow

```
> fviz_nbclust(frek, kmeans, method = "wss")
```



### Metode Shilluete

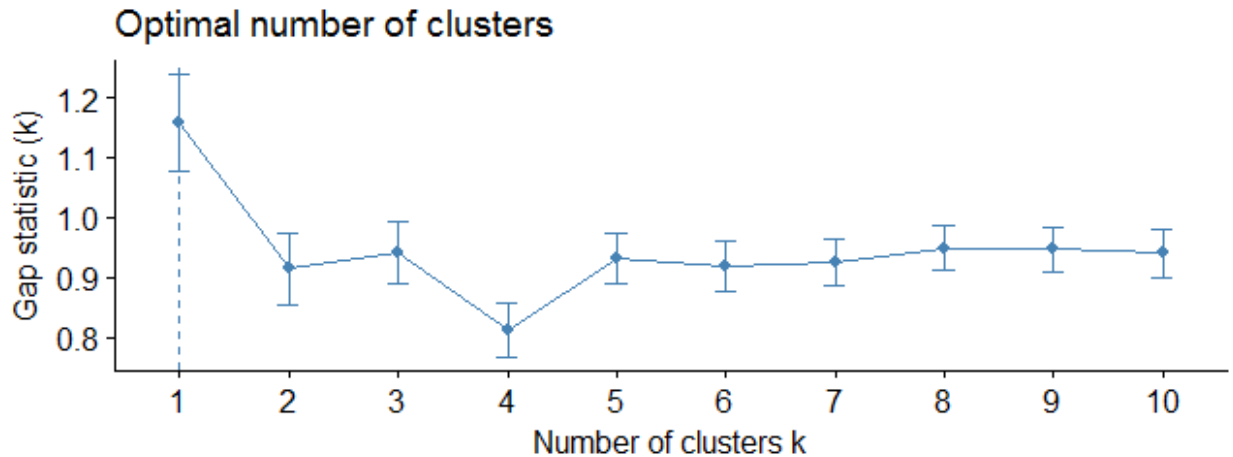
```
> fviz_nbclust(waktu, kmeans, method = "silhouette")
```



### Metode Gap Statistik

```
> set.seed(123)
> gap_stat <- clusGap(waktu, FUN = kmeans, nstart = 25,
+                     K.max = 10, B = 48)
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 48) [one "." per sample]:
..... 48
> fviz_gap_stat(gap_stat)
```





## Eksekusi K-Means

```
> akses <- kmeans(waktu, 3, nstart = 25)
> print(akses)
```

K-means clustering with 3 clusters of sizes 40, 7, 1

Cluster means:

	Agustus	September	Oktober	November	Desember	rataan
1	0.784250	1.360250	1.935750	1.895000	1.221750	1.439400
2	4.385714	7.594286	9.962857	8.495714	6.337143	7.355143
3	19.440000	40.770000	45.580000	58.000000	38.160000	40.390000

Clustering vector:

```
[1] 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
1 1
[39] 1 1 2 1 1 1 1 2 3 1
```

Within cluster sum of squares by cluster:

```
[1] 436.2043 436.1124 0.0000
(between_SS / total_SS = 92.3 %)
```

Available components:

```
[1] "cluster"      "centers"      "totss"       "withinss"    "tot.withinss"
"betweenss"    "size"        "iter"        "ifault"
```

```
> aggregate(waktu, by=list(cluster=akses$cluster), mean)
```

cluster	Agustus	September	Oktober	November	Desember	rataan
1	1 0.784250	1.360250	1.935750	1.895000	1.221750	1.439400
2	2 4.385714	7.594286	9.962857	8.495714	6.337143	7.355143
3	3 19.440000	40.770000	45.580000	58.000000	38.160000	40.390000

Waktu akses situs yang tidak mendukung pekerjaan selama jam kerja

<i>Cluster</i>	Kategori	Jumlah	Persentase
1	Rendah, akses situs yang tidak mendukung di tempat kerja (0-4 jam perminggu)	41	84%
2	Sedang, akses situs yang tidak mendukung di tempat kerja (>4-10 jam perminggu)	7	14%
3	Tinggi, akses situs yang tidak mendukung di tempat kerja (>10 jam perminggu)	1	2%

(Sumber level kategori akses: Fathonah dan Hartijasti, 2014)