



KMMI 2021

Eksplorasi dan Visualisasi Data

Pertemuan 8:
Visualisasi Data Menggunakan RStudio

Sub CPMK

- Mahasiswa menampilkan informasi grafis dan hubungan kompleks dalam kumpulan data menggunakan Rstudio

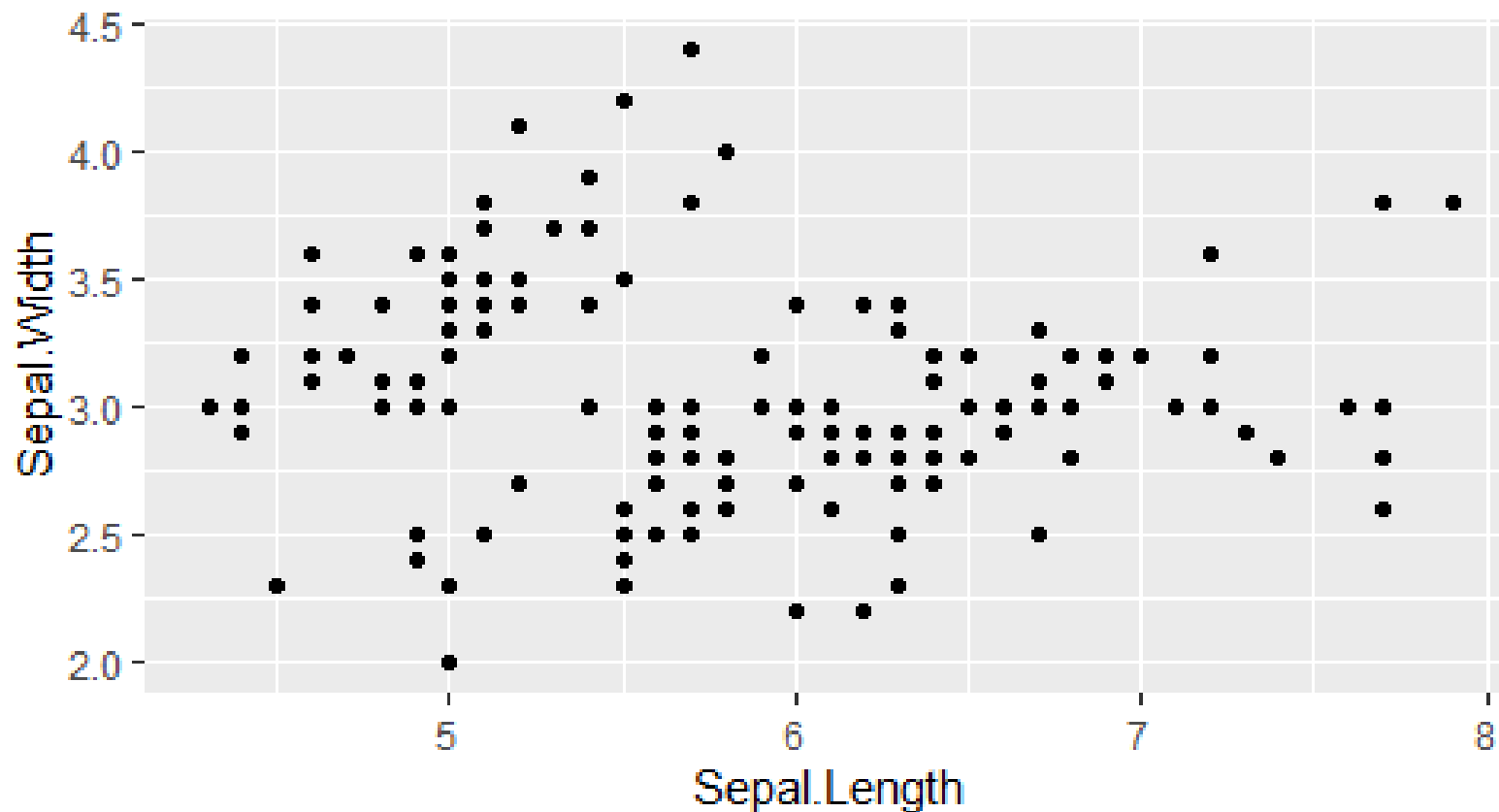
Pokok Bahasan

- Membuat visualisasi dasar: *scatterplot, bar plot, pie chart*
- *Multiple plots*
- *Pixel-oriented visualization techniques*
- *Geometric projection visualization techniques*
- *Icon-based visualization techniques*
- *Hierarchical visualization techniques*
- *Visualizing complex data*

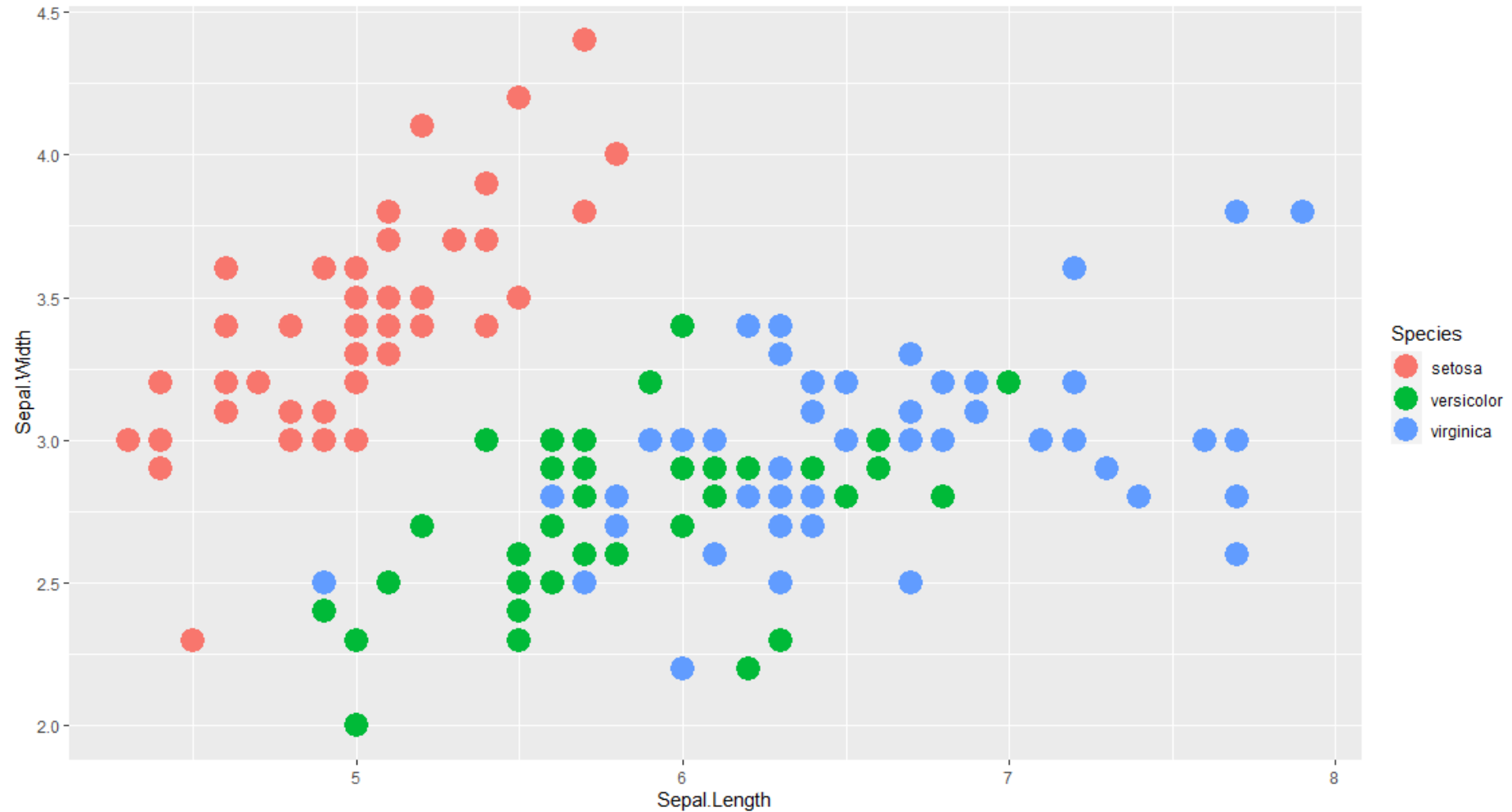
Membuat visualisasi dasar: **scatterplot**

- ***Scatterplot*** memvisualisasikan hubungan antara dua variabel numerik dengan menampilkan titik data pada koordinat x sebagai satu variabel dan koordinat y sebagai variabel lain.
- Secara konvensional, variabel pada sumbu x adalah variabel bebas dan variabel pada sumbu y adalah variabel terikat.
- **Scatterplot:**
 - menampilkan nilai dua variabel di sepanjang dua sumbu
 - menunjukkan hubungan di antara keduanya (mengungkapkan korelasi)

```
1 #basic scatterplot
2 library(ggplot2)
3 head(iris)
4 ggplot(iris, aes(x=Sepal.Length, y=Sepal.Width)) + geom_point()
```



```
6 #basic scatterplot with color depending on Species
7 ggplot(iris, aes(x=Sepal.Length, y=Sepal.Width, color=Species)) +
8   geom_point(size=6)
```



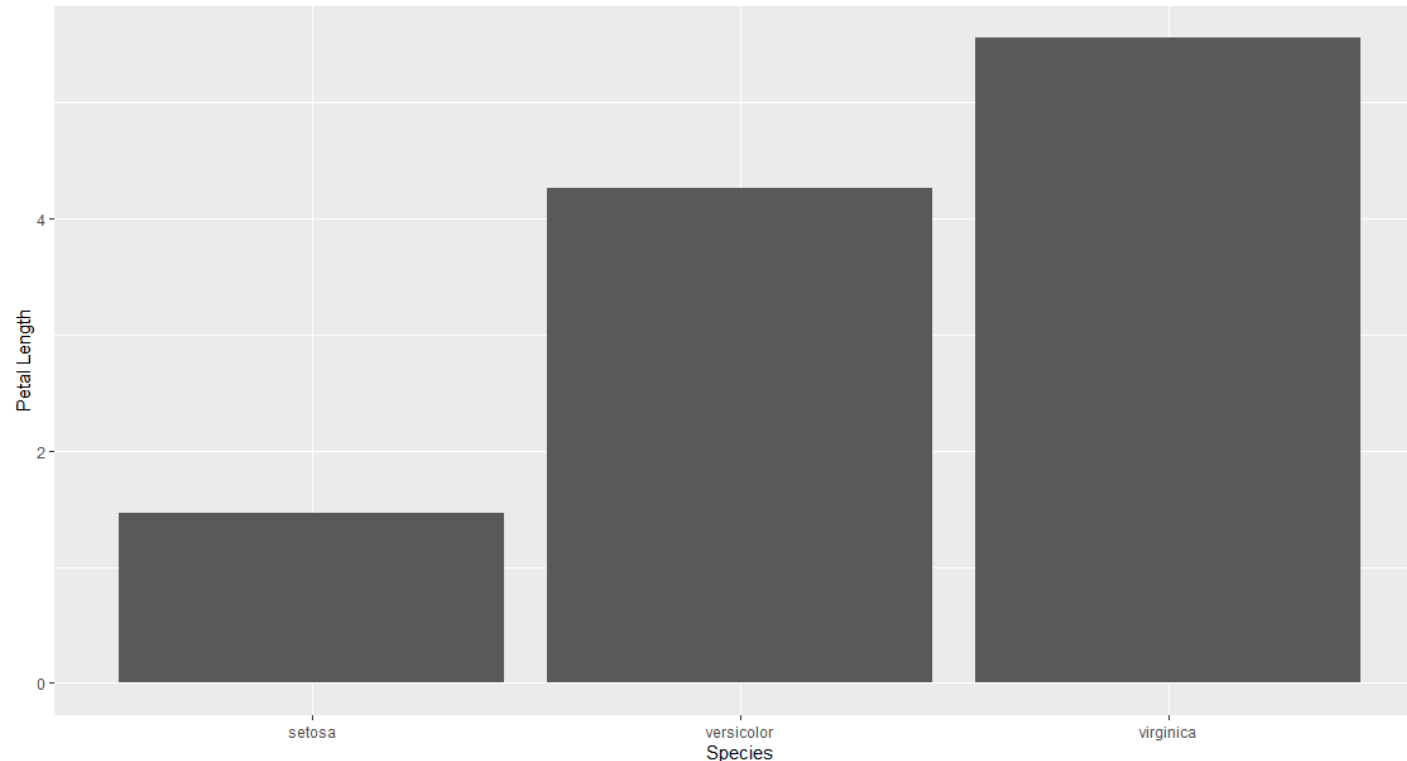
Membuat visualisasi dasar: **bar plot**

- **Diagram batang (*bar plot*)** dan **diagram lingkaran (*pie chart*)** adalah alat yang berguna untuk memvisualisasikan hubungan antara variabel numerik dan variabel kategorik.
- Diagram batang berisi sekumpulan batang (persegi Panjang) di mana tinggi batang sesuai dengan **nilai variabel**.
- **Diagram batang (*bar plot*)** memudahkan untuk melihat perbedaan antara nilai fitur dengan mengukur ukuran (panjang) batang.

```
10 # Create data aggregate (mean)
11 df = aggregate(iris[,1:4], by = list(iris$Species), FUN = mean)
12 df
```

	Group.1	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	setosa	5.006	3.428	1.462	0.246
2	versicolor	5.936	2.770	4.260	1.326
3	virginica	6.588	2.974	5.552	2.026

```
14 ggplot(df, aes(x = Group.1, y = Petal.Length)) +
15   geom_bar(stat="identity") +
16   xlab('Species') + ylab('Petal Length')
```



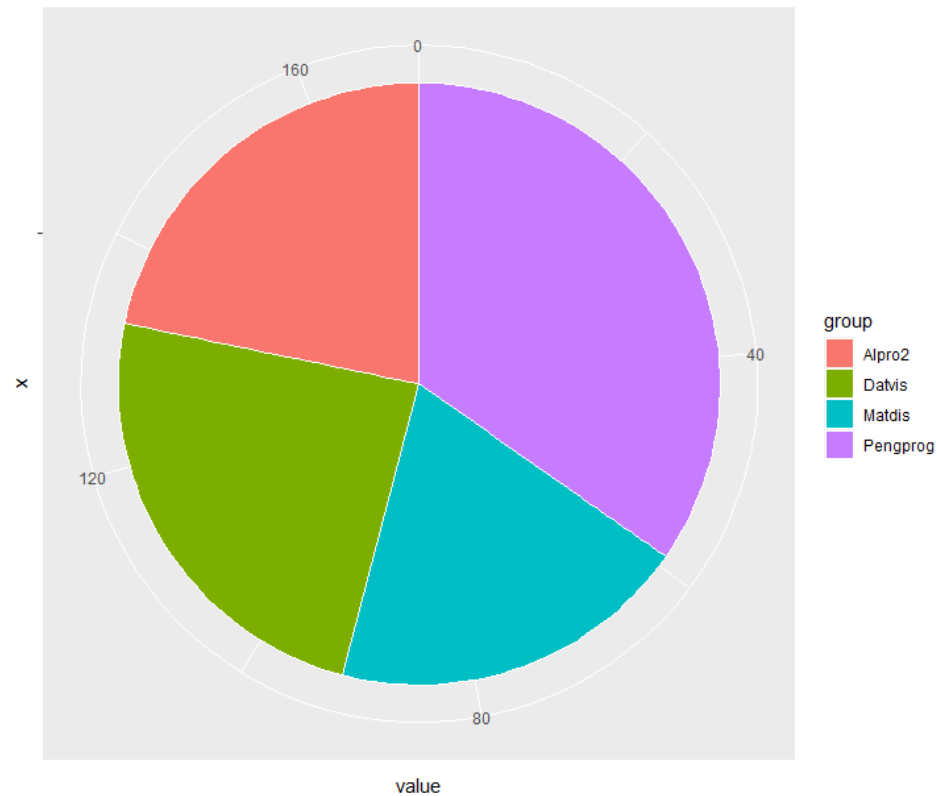
Membuat visualisasi dasar: **pie chart**

- **Diagram lingkaran** adalah visualisasi alternatif yang menunjukkan nilai gabungan dari variabel yang dihitung pada subkelompok data yang berbeda.
- Nilai agregat untuk subkelompok yang berbeda ditampilkan sebagai bagian dari keseluruhan.
- Lingkaran mewakili nilai agregat yang dihitung dari seluruh kumpulan data, dan irisan atau sektor yang mewakili, dihitung melalui subkelompok data yang berbeda.


```

19 # Create Data
20 data <- data.frame(
21   group=c('Matdis', 'Alpro2', 'Pengprog', 'Datvis'),
22   value=c(33,37,59,41)
23 )
24
25 ggplot(data, aes(x="", y=value, fill=group)) +
26   geom_bar(stat="identity", width=1, color="white") +
27   coord_polar("y", start=0)

```



Multiple plots

- Plot yang sama yang digambar pada subset yang berbeda dari dataset atau jenis plot yang berbeda yang digambar pada dataset yang sama.
- Parameter *mfrow* dari fungsi *par()* memungkinkan kita untuk menggambar banyak plot dalam satu tempat yang sama (*grid*).
- Parameter ini mengambil vektor dua elemen yang menunjukkan jumlah baris dan kolom plot

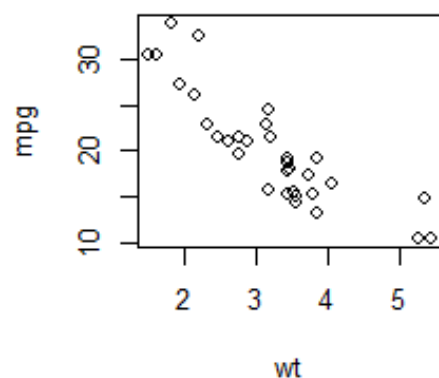
Mtcars dataset

A data frame with 32 observations on 11 (numeric) variables.

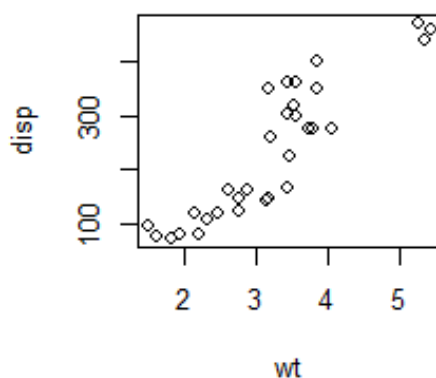
[, 1]	mpg	Miles/(US) gallon
[, 2]	cyl	Number of cylinders
[, 3]	disp	Displacement (cu.in.)
[, 4]	hp	Gross horsepower
[, 5]	drat	Rear axle ratio
[, 6]	wt	Weight (1000 lbs)
[, 7]	qsec	1/4 mile time
[, 8]	vs	Engine (0 = V-shaped, 1 = straight)
[, 9]	am	Transmission (0 = automatic, 1 = manual)
[,10]	gear	Number of forward gears

```
29 head(mtcars)
30 attach(mtcars)
31 par(mfrow=c(2,2))
32 plot(wt,mpg, main="Scatterplot of wt vs. mpg")
33 plot(wt,disp, main="Scatterplot of wt vs disp")
34 hist(wt, main="Histogram of wt")
35 boxplot(wt, main="Boxplot of wt")
```

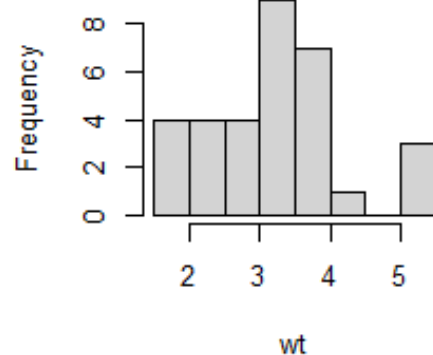
Scatterplot of wt vs. mpg



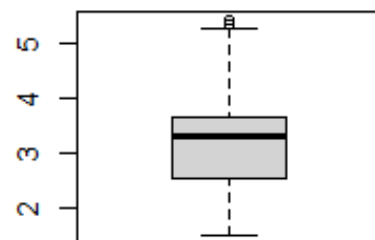
Scatterplot of wt vs disp



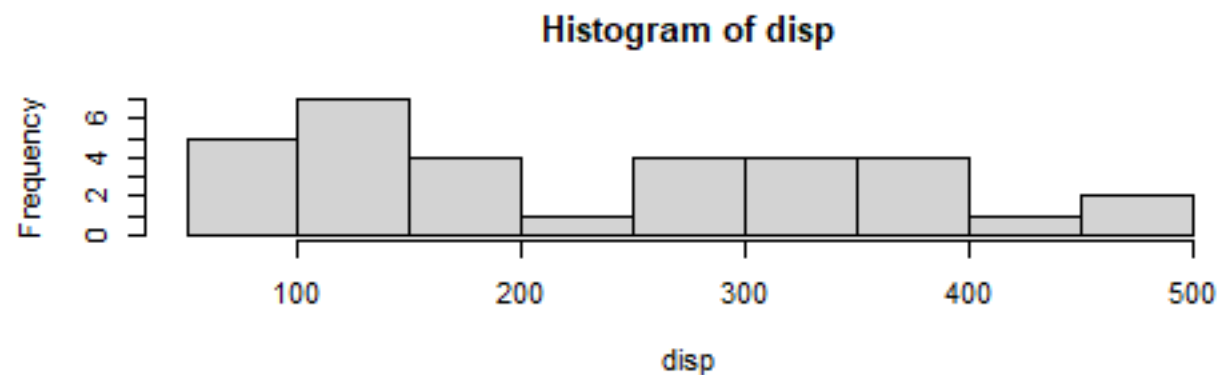
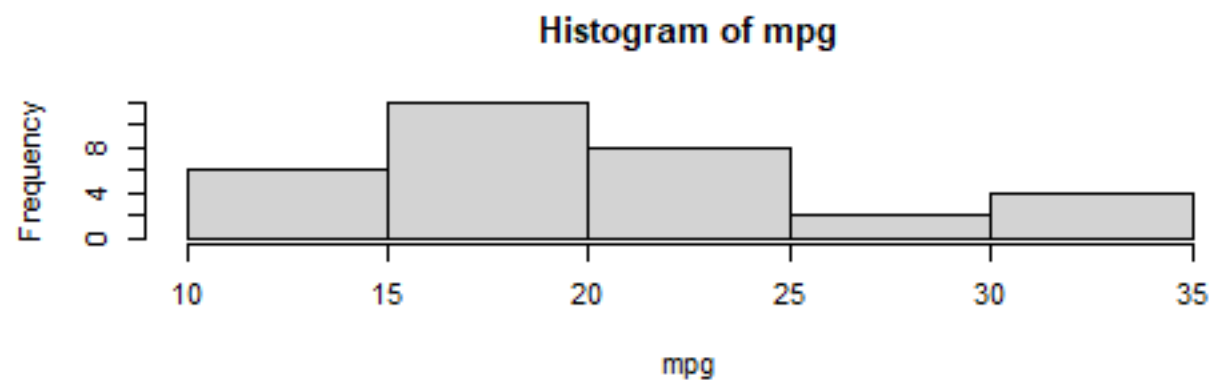
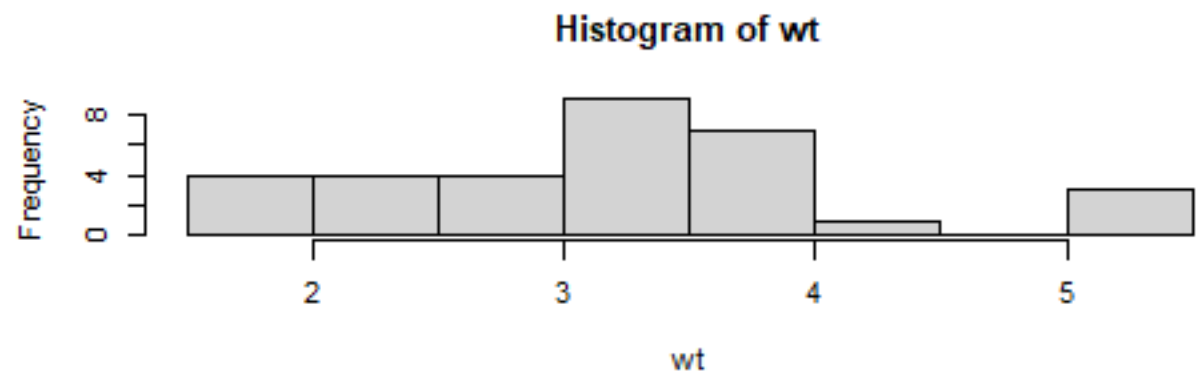
Histogram of wt



Boxplot of wt



```
37 # 3 figures arranged in 3 rows and 1 column
38 attach(mtcars)
39 par(mfrow=c(3,1))
40 hist(wt)
41 hist(mpg)
42 hist(dis)
```



Pixel-oriented visualization techniques

Cara sederhana untuk memvisualisasikan nilai dimensi adalah dengan menggunakan piksel di mana warna piksel mencerminkan nilai dimensi



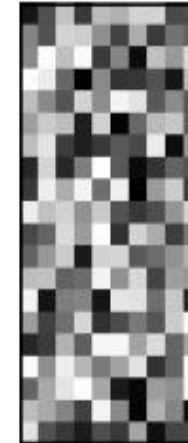
(a) *income*



(b) *credit_limit*



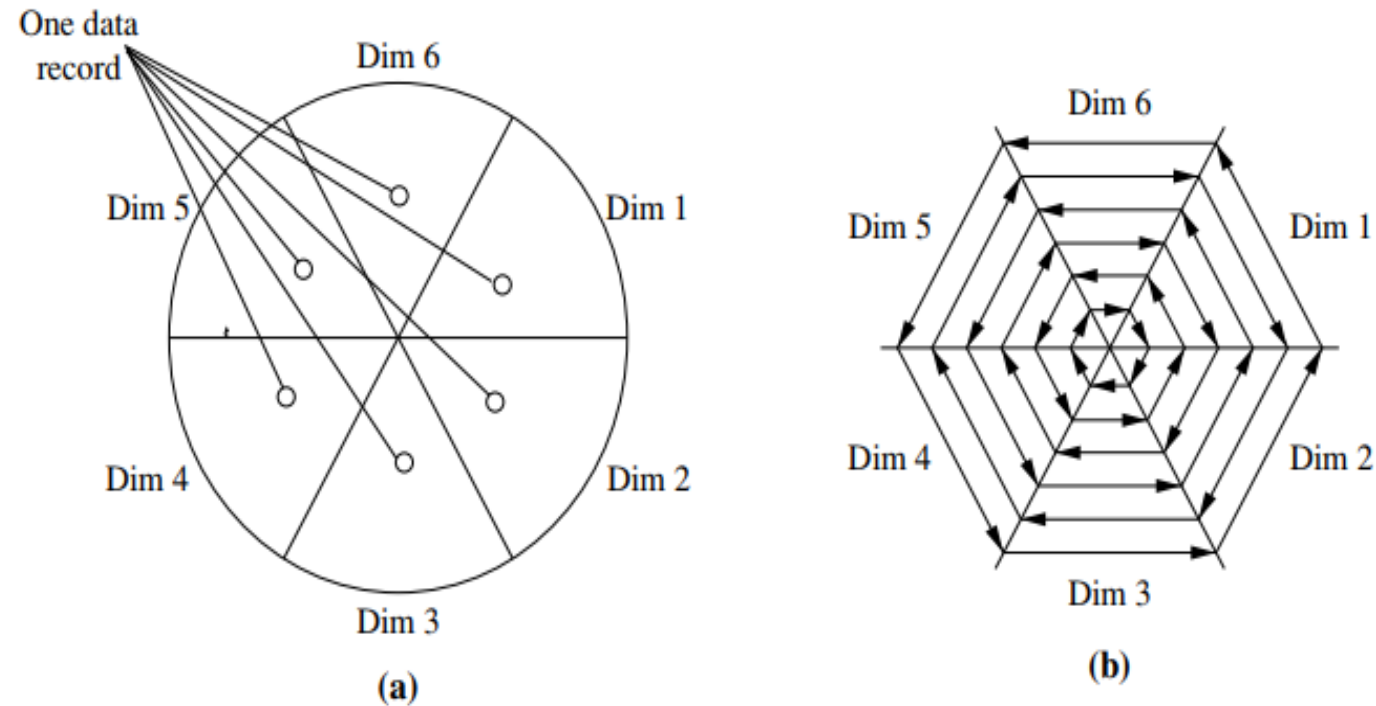
(c) *transaction_volume*



(d) *age*

Visualisasi berorientasi piksel dari empat atribut dengan mengurutkan semua pelanggan dalam pendapatan urutan menaik

Pixel-oriented visualization techniques



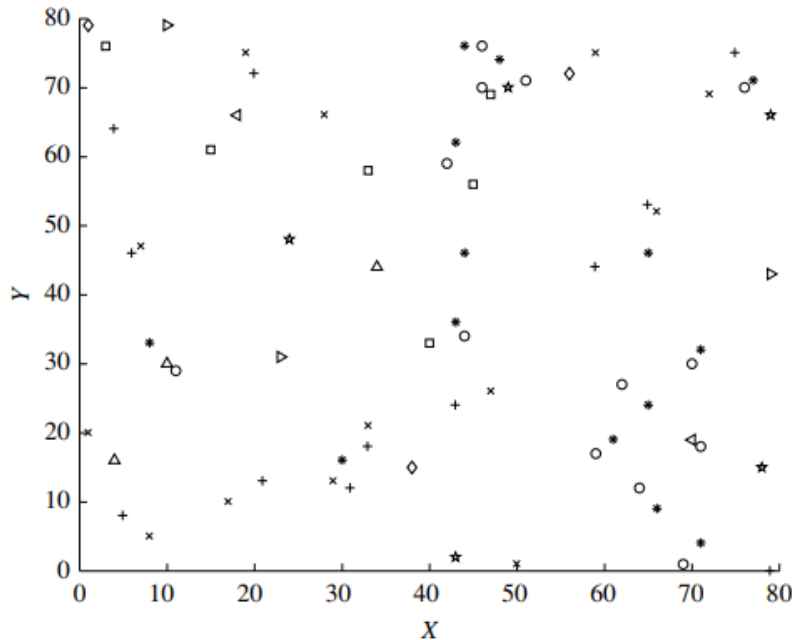
Circle segment technique.

- (a) Mewakili catatan data dalam segmen lingkaran.
- (b) Meletakkan piksel dalam segmen lingkaran.

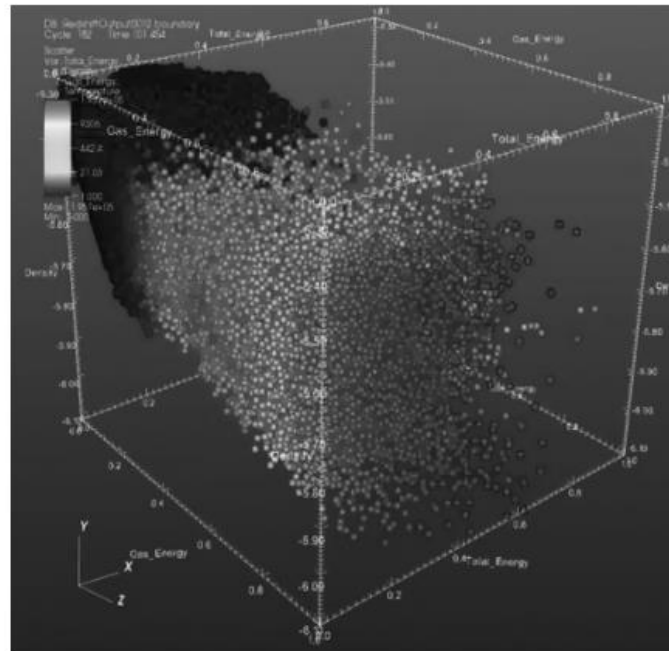
Geometric projection visualization techniques

- Kelemahan teknik visualisasi berorientasi piksel adalah bahwa teknik tersebut **tidak dapat banyak membantu kita dalam memahami distribusi data dalam ruang multidimensi.**
- Teknik proyeksi geometris membantu pengguna menemukan proyeksi menarik dari kumpulan data multidimensi.
- Tantangan utama yang coba diatasi oleh teknik proyeksi geometris adalah bagaimana memvisualisasikan ruang dimensi tinggi pada tampilan 3-D.
- *Scatterplot* menampilkan titik data 2-D menggunakan koordinat Cartesian

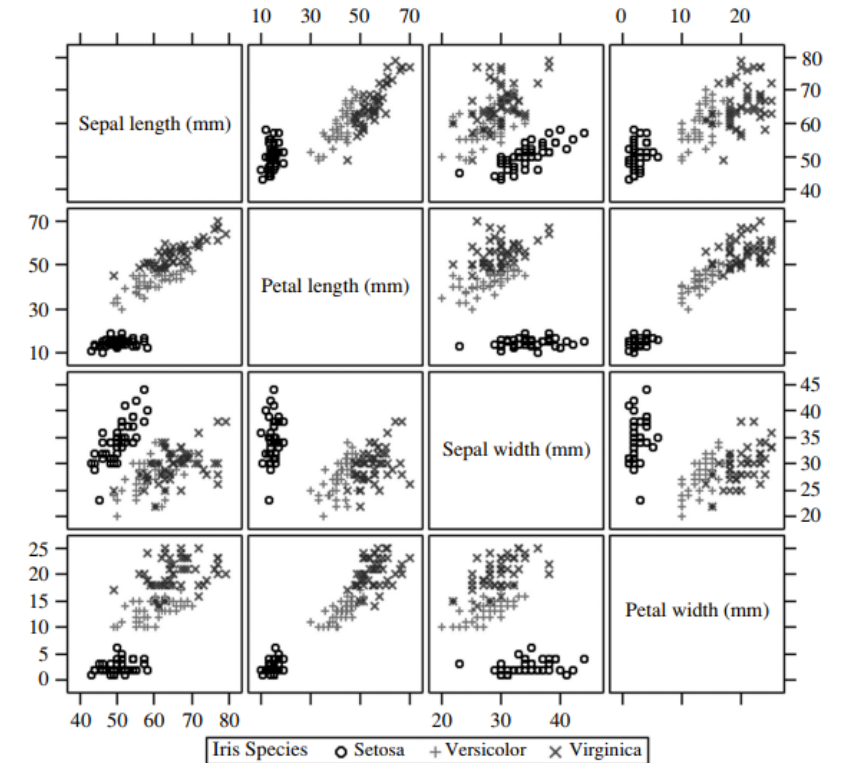
Geometric projection visualization techniques



Visualisasi kumpulan data 2-D menggunakan scatterplot



Visualisasi kumpulan data 3-D menggunakan scatterplot



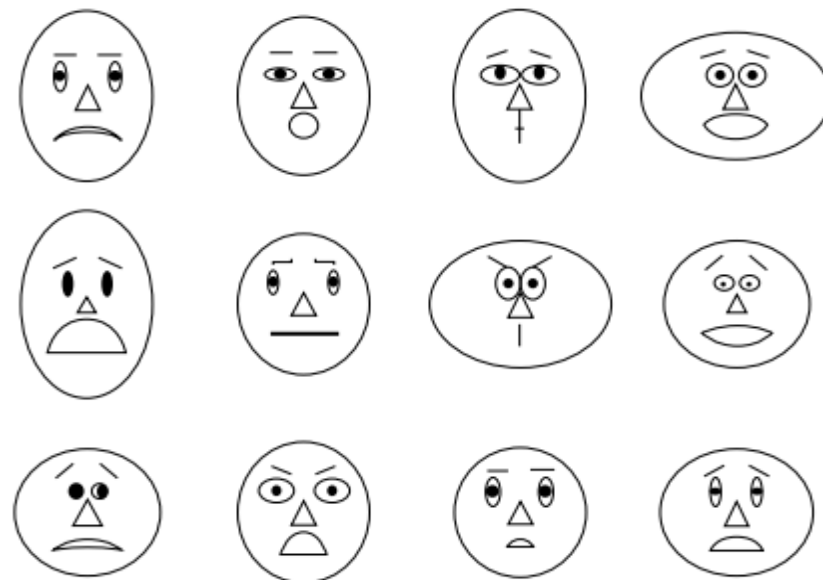
Visualisasi kumpulan data Iris menggunakan scatter-plot matrix

Icon-based visualization techniques

- Teknik visualisasi berbasis ikon menggunakan ikon kecil untuk mewakili nilai data multidimensi.
- Dua teknik berbasis ikon yang populer: **Chernoff faces** dan **stick figures**.

Icon-based visualization techniques

- Chernoff faces membantu mengungkap tren dalam data.
- Komponen wajah, seperti mata, telinga, mulut, dan hidung, mewakili nilai-nilai dimensi berdasarkan bentuk, ukuran, penempatan, dan orientasinya. Misalnya, dimensi dapat dipetakan ke karakteristik wajah berikut: ukuran mata, jarak mata, panjang hidung, lebar hidung, lengkungan mulut, lebar mulut, keterbukaan mulut, ukuran pupil, kemiringan alis, eksentrisitas mata, dan eksentrisitas kepala.



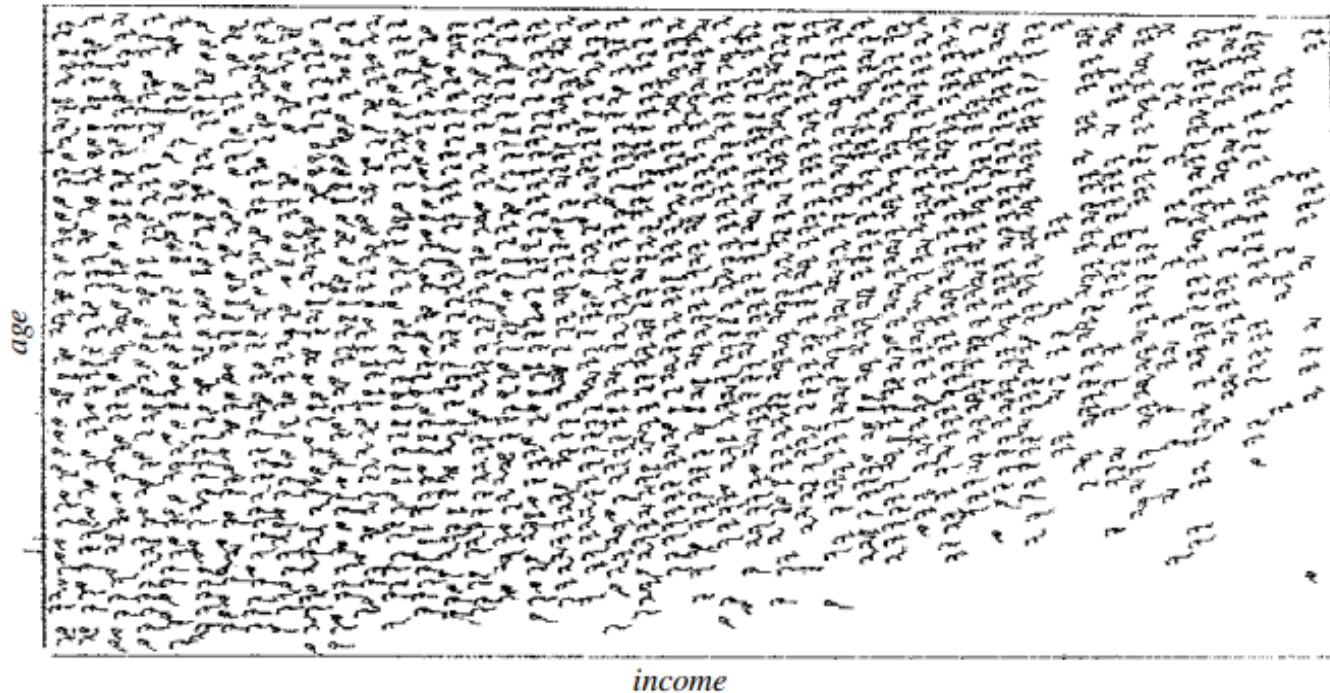
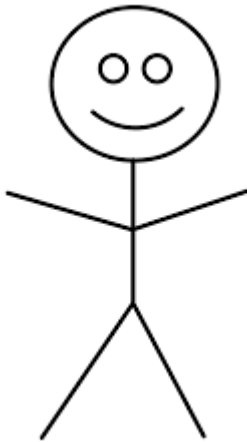
Icon-based visualization techniques

Iris dataset



Icon-based visualization techniques

- Teknik visualisasi ***stick figure*** memetakan data multidimensi ke figur stik lima potong, di mana setiap figur memiliki empat anggota badan dan satu tubuh.
- Dua dimensi dipetakan ke sumbu tampilan (x dan y) dan dimensi lainnya dipetakan ke sudut dan/atau panjang anggota badan.



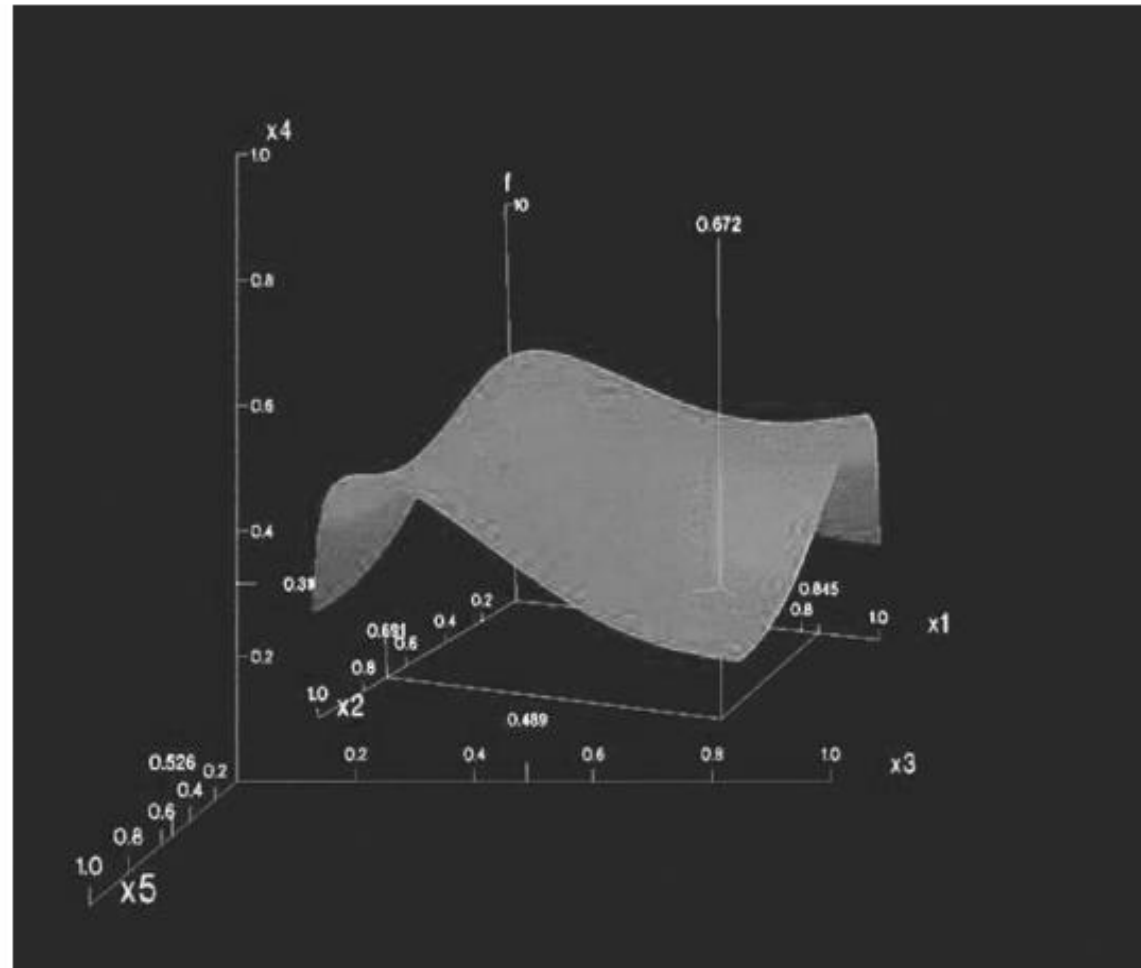
Hierarchical visualization techniques

- Teknik visualisasi yang dibahas sejauh ini berfokus pada memvisualisasikan beberapa dimensi secara bersamaan.
- Namun, untuk kumpulan data besar berdimensi tinggi, akan sulit untuk memvisualisasikan semua dimensi pada waktu yang sama.
- Teknik visualisasi hierarki mempartisi semua dimensi ke dalam himpunan bagian (yaitu, subruang). Subruang divisualisasikan secara hierarki.

Hierarchical visualization techniques

- **Worlds-within-Worlds** juga dikenal sebagai n-Vision, adalah representasi dari metode visualisasi hierarkis.
- Misalkan kita ingin memvisualisasikan kumpulan data 6-D, di mana dimensinya adalah F, X_1, \dots, X_5 .
- Kita ingin mengamati bagaimana dimensi F berubah sehubungan dengan dimensi lain.
- Pertama-tama kita dapat memperbaiki nilai dimensi X_3, X_4, X_5 ke beberapa nilai yang dipilih, misalnya, c_3, c_4, c_5 .
- Kita kemudian dapat memvisualisasikan F, X_1, X_2 menggunakan plot 3-D, yang disebut *world*, seperti yang ditunjukkan pada Gambar di slide berikutnya.

Hierarchical visualization techniques



Worlds-within-Worlds

Hierarchical visualization techniques

- Sebagai contoh lain dari metode visualisasi hierarkis adalah ***tree-maps*** yang menampilkan data hierarkis sebagai kumpulan persegi panjang bersarang.
- Contoh ditunjukkan gambar pada slide berikutnya, menunjukkan *tree-maps* yang memvisualisasikan berita Google.
- Semua berita diatur ke dalam tujuh kategori, masing-masing ditampilkan dalam persegi panjang besar dengan warna yang unik.
- Dalam setiap kategori (yaitu, setiap persegi panjang di tingkat atas), berita-berita tersebut selanjutnya dipartisi menjadi subkategori yang lebih kecil.

Hierarchical visualization techniques



Newsmap:
Penggunaan peta
pohon untuk
memvisualisasikan
berita utama berita
Google

Visualizing complex data

- Memvisualisasikan dan menganalisis data tidak terstruktur seperti teks dan social networks menarik banyak minat.
- Ada banyak teknik visualisasi baru yang didedikasikan untuk jenis data ini.
- ***Tag cloud*** adalah visualisasi statistik dari *tag* yang dibuat pengguna.
- Seringkali, dalam *tag cloud*, *tag* terdaftar menurut abjad atau dalam urutan yang disukai pengguna. Pentingnya *tag* ditunjukkan oleh ukuran font atau warna. Gambar di slide berikutnya menunjukkan *tag cloud* untuk memvisualisasikan *tag* populer yang digunakan di situs Web.

Visualizing complex data

animals architecture **art** asia australia autumn baby band barcelona **beach** berlin bike bird
birds birthday black blackandwhite blue bw california canada **canon** car cat
chicago china christmas church city clouds color concert cute dance day de dog
england europe fall **family** fashion festival film florida flower flowers food
football france friends fun garden geotagged germany girl girls graffiti green
halloween hawaii holiday home house india iphone ireland island italia **italy** japan july kids la
lake landscape light live london love macro me mexico model mountain mountains museum
music nature new newyork newyorkcity night **nikon** nyc ocean old paris
park party people photo photography photos portrait red river rock san
sanfrancisco scotland sea seattle show sky snow spain spring street summer
sun sunset taiwan texas thailand tokyo toronto tour **travel** tree trees trip uk urban
usa vacation washington water **wedding** white winter yellow york zoo

Menggunakan
tag cloud untuk
memvisualisasika
n *Web site tags*

Tugas

Gunakan Bahasa pemrograman R untuk membuat ***plot (scatterplot, barchart, piechart)*** pada atribut dataset tugas kelompok sebelumnya. Tambahkan plot untuk data berdimensi tinggi (teknik visualisasi bebas). Kemudian, artikan apa maksud dari plot tersebut.

- Tugas dikerjakan berkelompok.
- Tugas dikumpulkan paling lambat 13 November 2021 pukul 19.00 WIB di LMS.
- Beri nama file tugas: Tugas 08_Kelompok XX. (Contoh: Tugas 08_Kelompok 01)
- Tugas dikumpulkan dalam format pdf.



Terima Kasih