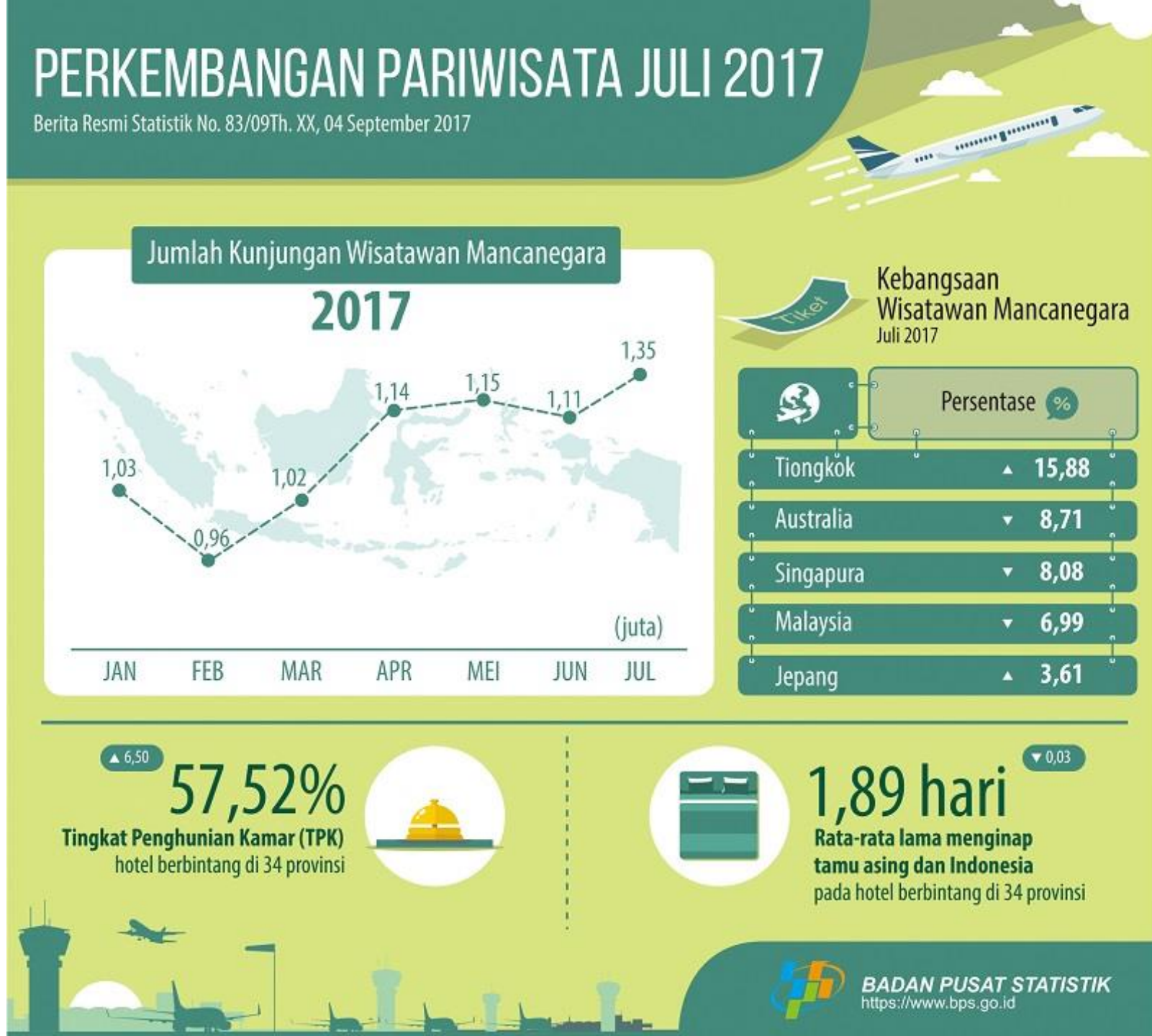


Short Quiz

1. Sebutkan minimum 5 informasi yg Anda peroleh dari gambar di samping?
2. Sebutkan peubah apa saja yg diamati pada kasus ini?
3. Sebutkan skala pengukurannya.
4. Berikan komentar Anda secara umum tentang infografis di samping.

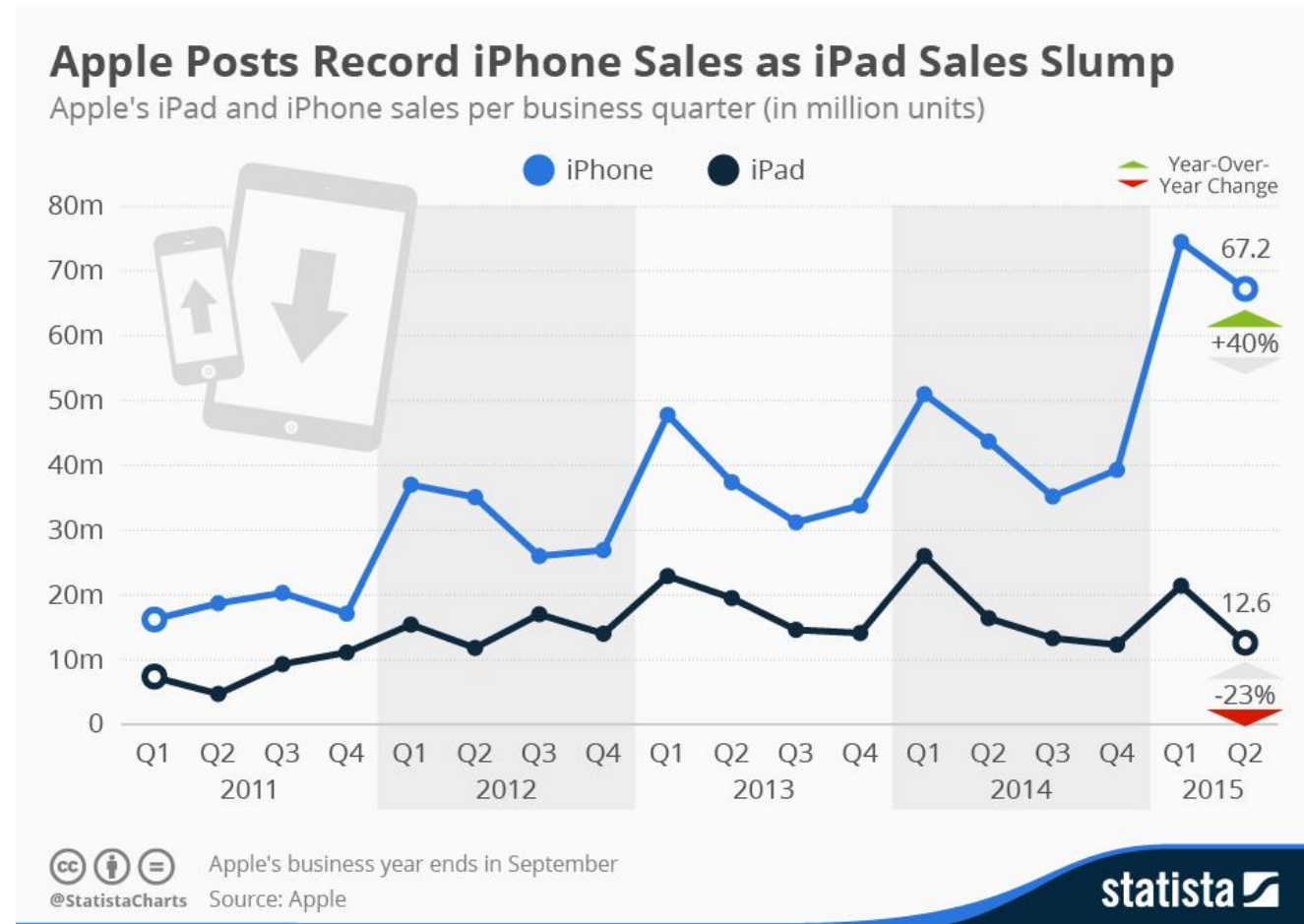
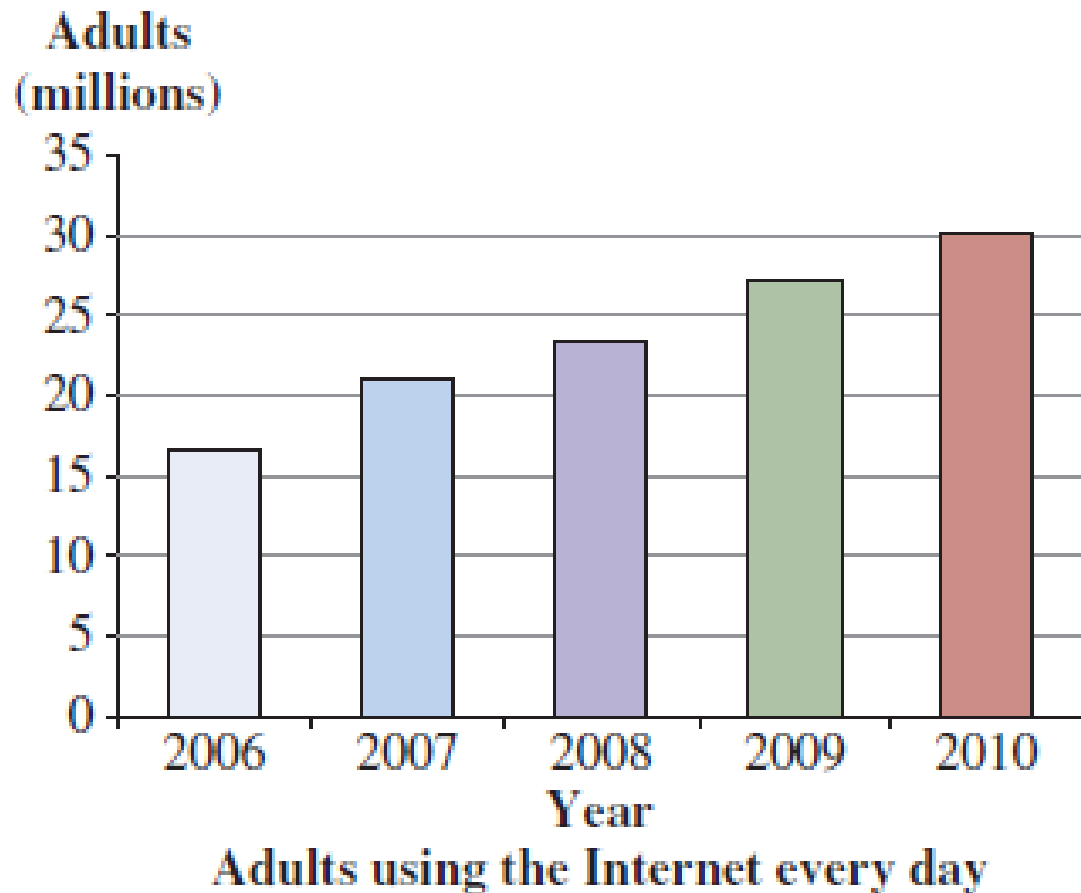


Outline

- ✓ Plot Data Deret Waktu
- ✓ Menghitung ukuran lokasi (rataan, rataan terboboti, median, dan modus)
- ✓ Menghitung ukuran keragaman (kisaran, ragam dan simpangan baku)
- ✓ Menghitung persentil dan Kuartil
- ✓ Membuat diagram kotak garis



Menyajikan Data dari Waktu ke Waktu



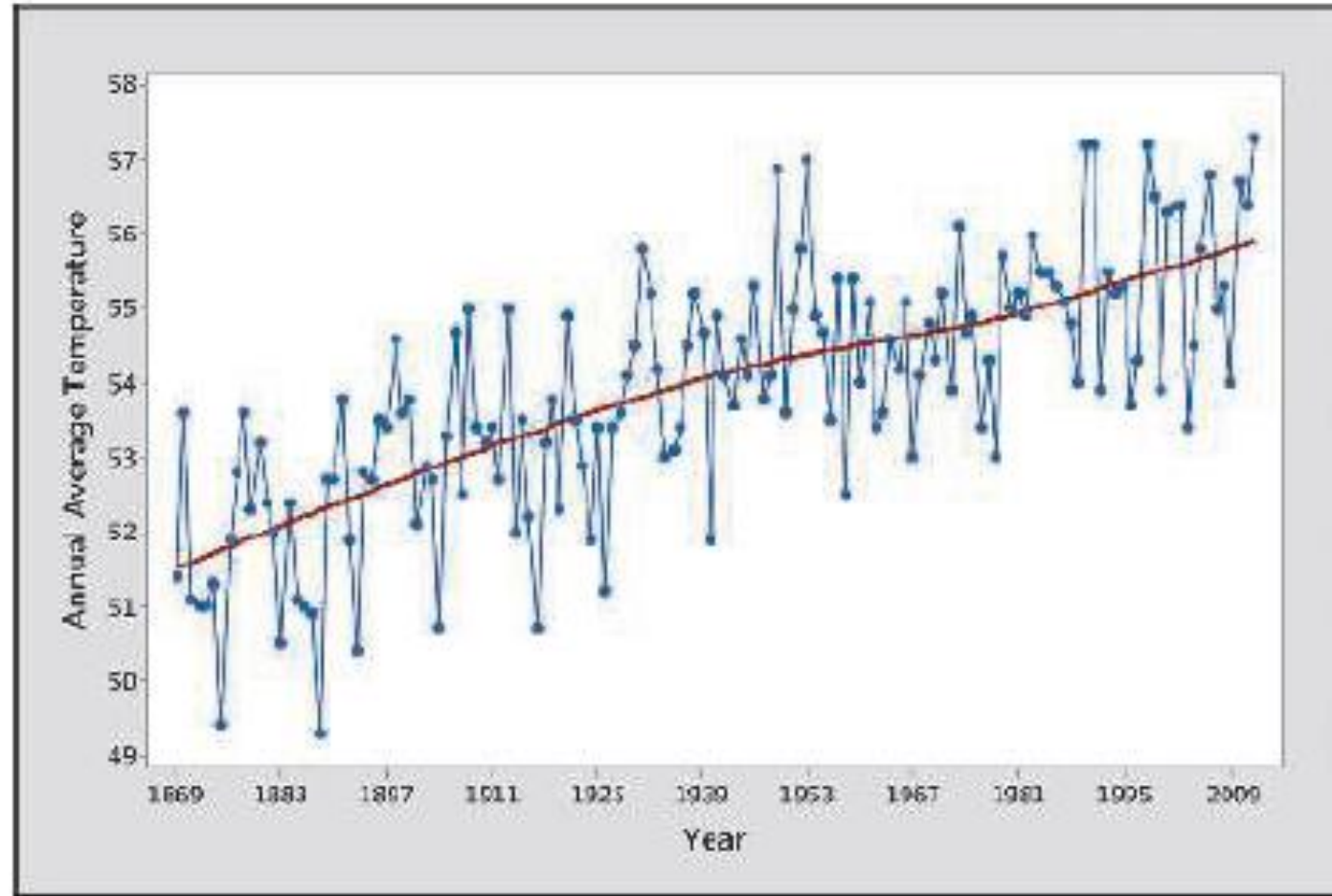
Ilustrasi



Warming Trend in New York City

- In a given year, the annual average temperature is the average of the daily average temperatures for that year.
- Let's analyze data on annual average Temperature (in degrees Fahrenheit) in Central Park, New York City, from 1869 to 2012. This is a continuous, quantitative variable.

Ilustrasi



▲ **Figure 2.8 MINITAB Output for a Time Plot of Central Park, New York City, Average Annual Temperatures.** The annual average temperatures are plotted against the year from 1869 to 2012. A smoothing curve is superimposed. **Question:** Are the annual average temperatures tending to increase, decrease, or stay the same over time?

Sedikit Cerita tentang Visualisasi Data



On the Shoulders of...Florence Nightingale

Graphical Displays Showing Deaths From Disease Versus Military Combat

Florence Nightingale

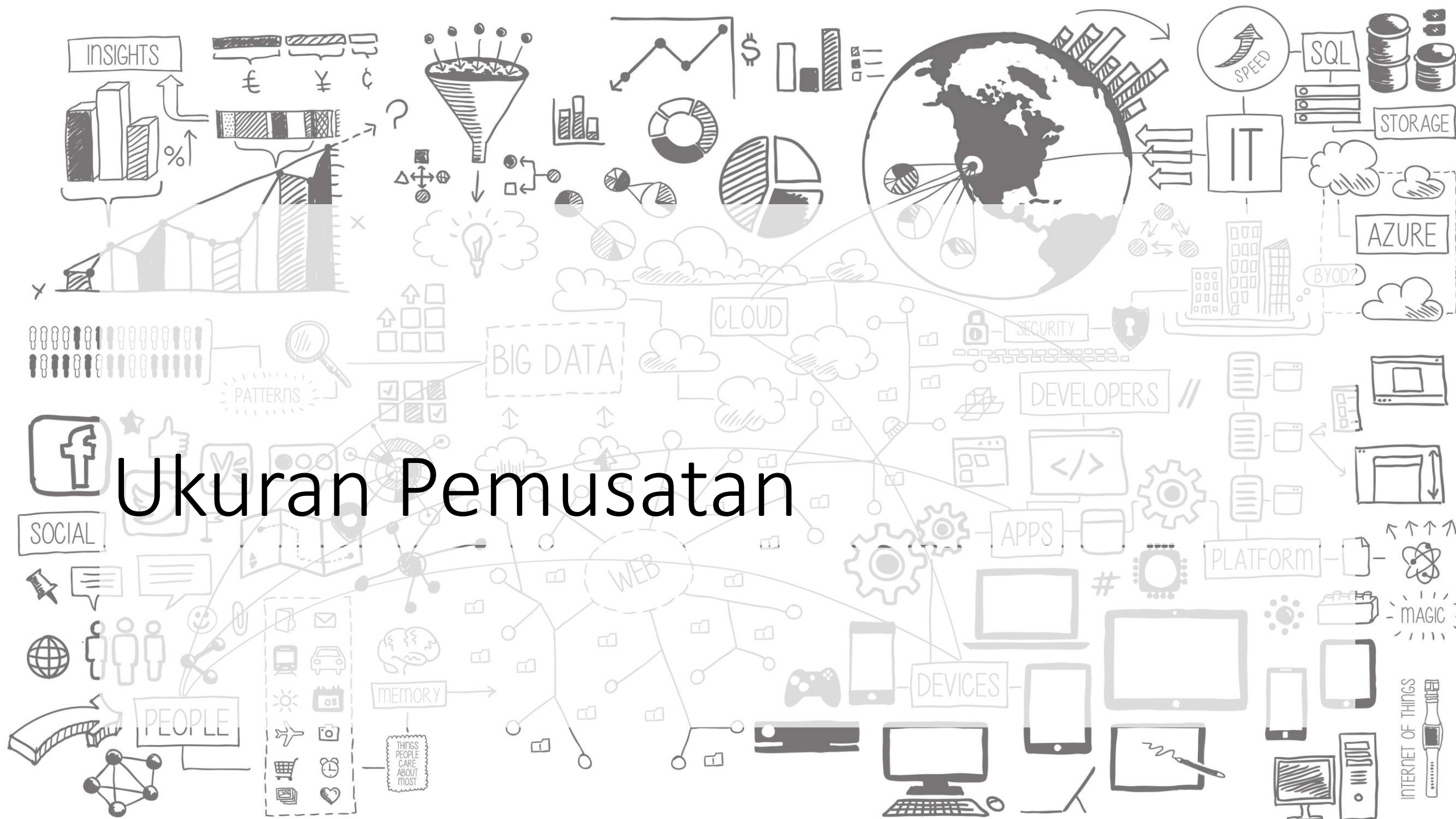
During the Crimean War in 1854, the British nurse Florence Nightingale (1820–1910) gathered data on the number of soldiers who died from various causes. She prepared graphical displays such as time plots and pie charts for policy makers. The graphs showed that more soldiers were dying from contagious diseases than from war-related wounds. The plots were

revolutionary for her time. They helped promote her national cause of improving hospital conditions.

After implementing sanitary methods, Nightingale showed with time plots that the relative frequency of soldiers' deaths from contagious disease decreased sharply and no longer exceeded that of deaths from wounds.

Throughout the rest of her life, Nightingale promoted the use of data for making informed decisions about public health policy. For example, she used statistical arguments to campaign for improved medical conditions in the United States during the Civil War in the 1860s (Franklin, 2002).

Ukuran Pemusatan



Rataan (*Mean*) dan Median

Center of the Cereal Sodium Data

Picture the Scenario

In Examples 4, 5, and 7 in Section 2.2, we investigated the sodium level in 20 breakfast cereals and saw various ways to graph the data. Let's return to those data and learn how to describe their center. The observations (in mg) are

0	340	70	140	200	180	210	150	100	130
140	180	190	160	290	50	220	180	200	210

Tentukan Rataan dan Median dari kandungan sodium pada sereal.

We find the mean by adding all the observations and then dividing this sum by the number of observations, which is 20:

$$\text{Mean} = (0 + 340 + 70 + \dots + 210)/20 = 3340/20 = 167.$$

To find the median, we arrange the data from the smallest to the largest observation.

0	50	70	100	130	140	140	150	160	180
180	180	190	200	200	210	210	220	290	340

For the 20 observations, the smaller 10 (on the first line) range from 0 to 180, and the larger 10 (on the second line) range from 180 to 340. The median is 180, which is the average of the two middle values, the tenth and eleventh observations, $(180 + 180)/2$.

- Bagaimana jika data ke-10 dan 11 nilainya tidak sama, misal $X_{10} = 180$ sedangkan $X_{11} = 190$?
- Maka median adalah:

$$\text{Median} = \frac{180 + 190}{2} = 185$$

Bagaimana menentukan Median?

- Urutkan data dari berdasarkan nilainya, yang terkecil s.d terbesar.
- Jika banyak data berjumlah ganjil, maka median adalah nilai dari pengamatan yang berada di tengah-tengah data.
- Jika data berjumlah genap, maka median adalah rata-rata dari dua pengamatan yang berada di tengah-tengah data.



Suatu nilai yang membagi data menjadi dua bagian yang sama besar

Rataan(*Mean*)

Variables are symbolized by letters near the end of the alphabet, most commonly x and y . The sample size is denoted by n . For a sample of n observations on a variable x , the mean is denoted by \bar{x} (pronounced “x bar”). Using the mathematical symbol Σ for “sum,” the mean has the formula

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Rataan(*Mean*)

Center of the Cereal Sodium Data

0	340	70	140	200	180	210	150	100	130
140	180	190	160	290	50	220	180	200	210

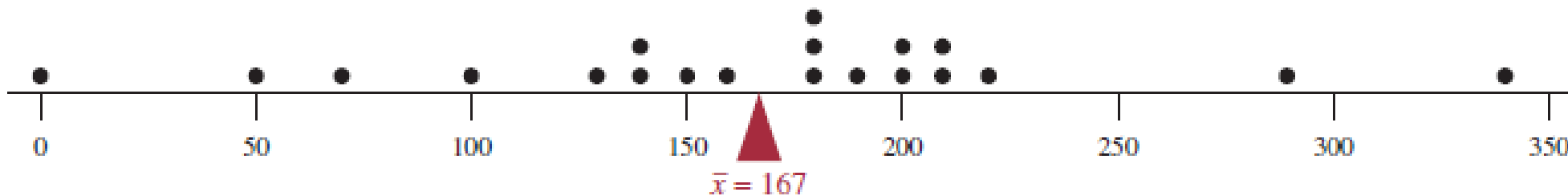
Rataan data:

$$\begin{aligned}\bar{x} &= (\Sigma x) / n \\ &= (0 + 340 + 70 + \dots + 210) / 20 \\ &= 3340 / 20 = 167\end{aligned}$$

Rataan(*Mean*)

- Merupakan pusat massa (centroid)

The Fulcrum Shows the Mean of the Cereal Sodium Data



- Notasi:
 - Populasi $\rightarrow \mu$
 - Contoh $\rightarrow \bar{x}$
- Digunakan untuk peubah numerik
- Sangat tidak resisten terhadap pencilan (*outlier*)

Pencilan (*Outlier*)

- Suatu pengamatan yang nilainya jauh lebih besar atau lebih kecil dibandingkan dengan sekumpulan pengamatan pada sebuah data.

Ilustrasi



CO₂ Pollution

Picture the Scenario

The Pew Center on Global Climate Change⁴ reports that global warming is largely a result of human activity that produces carbon dioxide (CO₂) emissions and other greenhouse gases. The CO₂ emissions from fossil fuel combustion are the result of electricity, heating, industrial processes, and gas consumption in automobiles. The International Energy Agency⁵ reported the per capita CO₂ emissions by country (that is, the total CO₂ emissions for the country divided by the population size of that country) for 2011. For the nine largest countries in population size (which make up more than half the world's population), the values were, in metric tons per person:

China 5.9	Indonesia 1.8	Nigeria 0.3
India 1.4	Brazil 2.1	Bangladesh 0.4
United States 16.9	Pakistan 0.8	Russian Federation 11.6

⁴Source: www.pewclimate.org/global-warming-basics/facts_and_figures/.

⁵Source: www.iea.org/.

Ilustrasi

Pertanyaan:

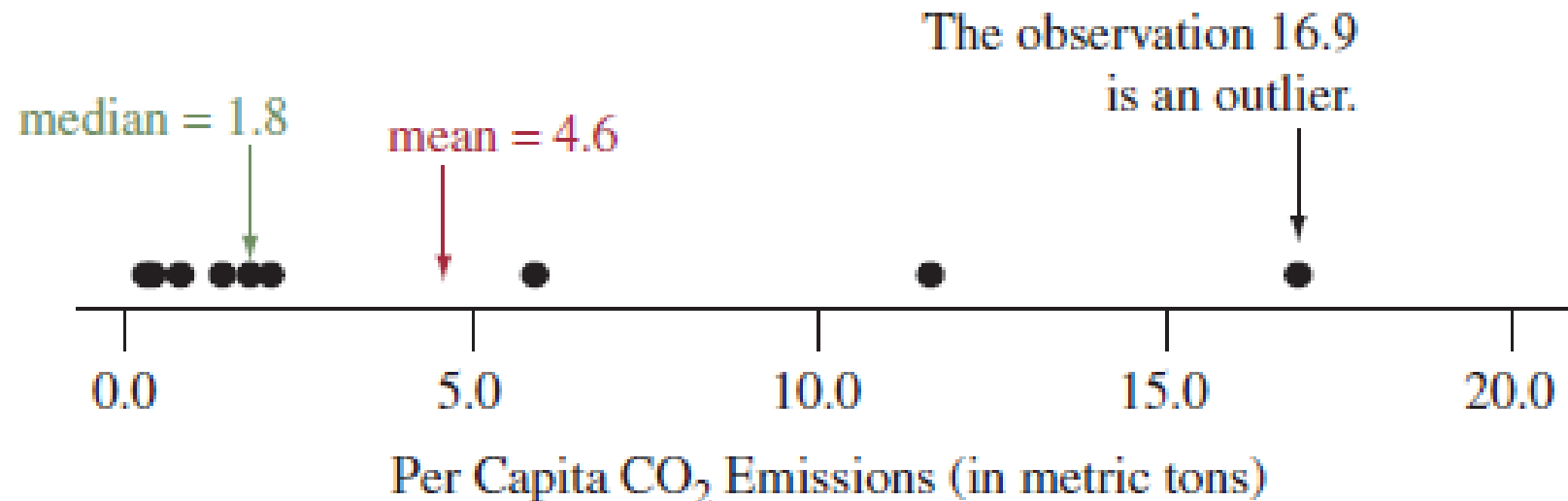
- a. Tentukan rata-rata dan median dari data polusi pada kasus ini!
- b. Menurut Anda, apakah terdapat pencilan?
- c. Berdasarkan nilai yang Anda peroleh pada butir (a), sebutkan pengaruh dari pencilan terhadap nilai rata-rata!

Ilustrasi

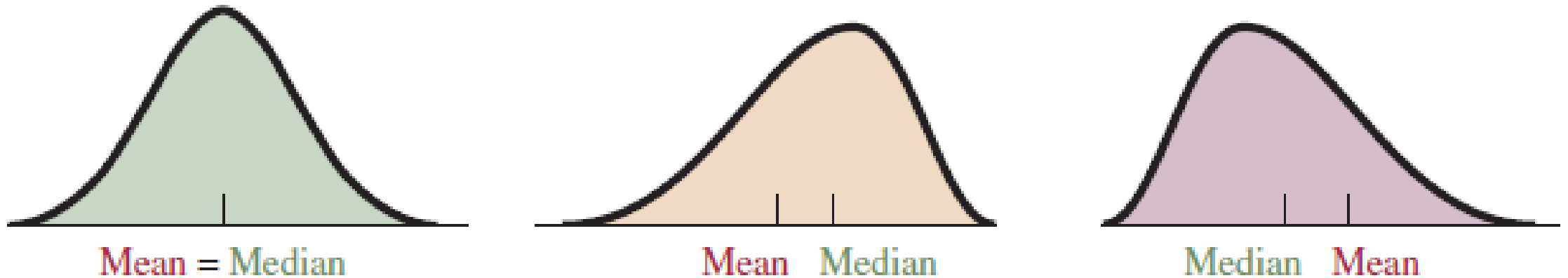
The CO₂ values have $n = 9$ observations. The ordered values are

0.3, 0.4, 0.8, 1.4, 1.8, 2.1, 5.9, 11.6, 16.9

Because n is odd, the median is the middle value, which is 1.8.



Perbandingan Rataan dan Median



▲ **Figure 2.9 Relationship Between the Mean and Median.** Question For skewed distributions, what causes the mean and median to differ?

Perbandingan Rataan dan Median

- Perhatikan data polusi pada ilustrasi sebelumnya.

0.3, 0.4, 0.8, 1.4, 1.8, 2.1, 5.9, 11.6, 16.9

- Dari data tsb diperoleh bahwa: Rataan = 4.6 dan Median=1.8

- Misalkan nilai dari pengamatan terakhir diubah, **16.9 diganti menjadi 90:**

0.3, 0.4, 0.8, 1.4, 1.8, 2.1, 5.9, 11.6, 90

- Berapakah rataan dan median-nya sekarang?

Perbandingan Rataan dan Median

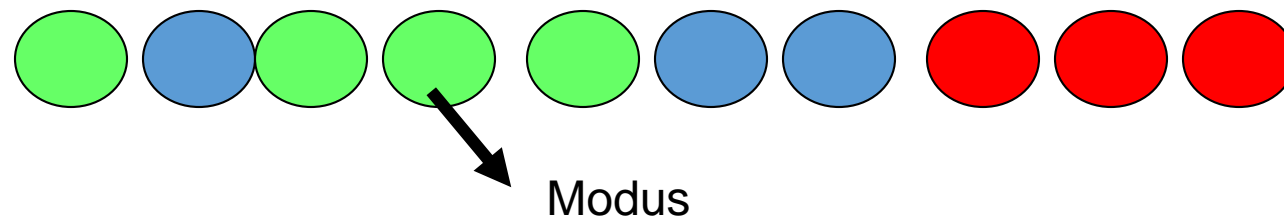
- Rataan menjadi 12.7 (sebelumnya 4.6)
- Median nilainya tetap 1.8



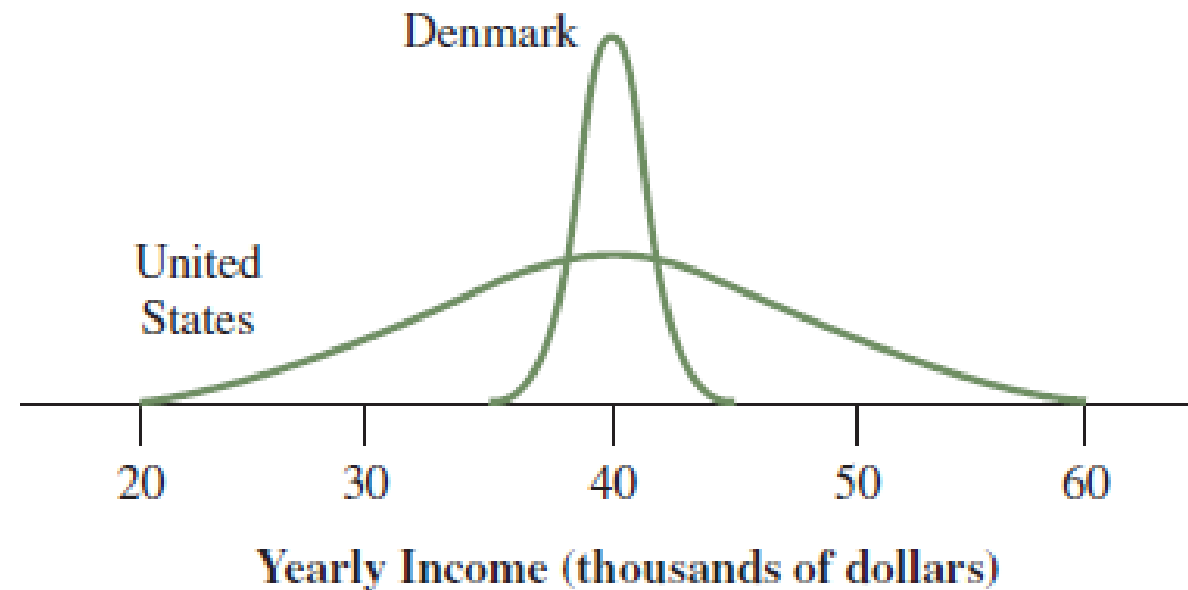
Median resisten terhadap pencilan, sedangkan rataaan tidak resisten terhadap pencilan.

Modus (Mode)

- Merupakan nilai pengamatan yang paling sering muncul
- Dalam satu gugus data dapat mengandung lebih dari satu modus
- Dapat digunakan untuk semua jenis data, tapi paling banyak digunakan untuk data kategorik atau data diskret dengan hanya sedikit nilai yang mungkin muncul



Mengapa diperlukan ukuran penyebaran?



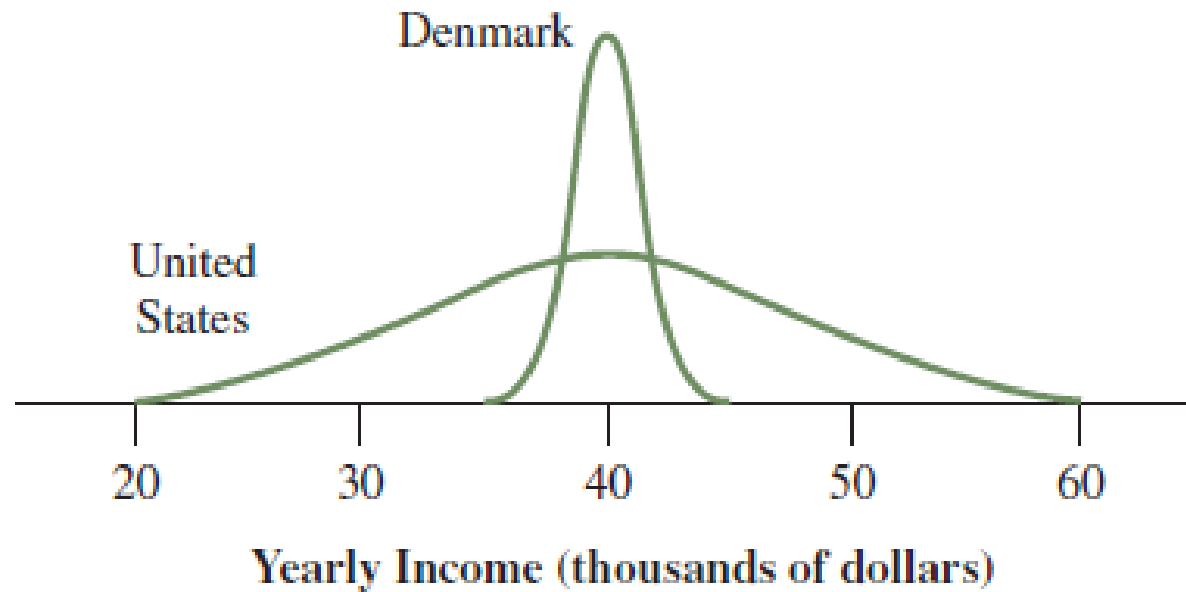
▲ **Figure 2.10** Income Distributions for Music Teachers in Denmark and in the United States. The distributions have the same mean, but the one for the United States shows more variability around the mean. **Question** How would the range for Denmark change if one teacher earned \$100,000?

Range (Wilayah)

- Merupakan selisih dari nilai terbesar dan nilai terkecil:

$$R = x_{max} - x_{min}$$

- Hanya memperhitungkan nilai terkecil dan terbesar, sedangkan sebaran nilai antara dua nilai tersebut tidak diperhitungkan
- Tidak resisten terhadap pencilan.

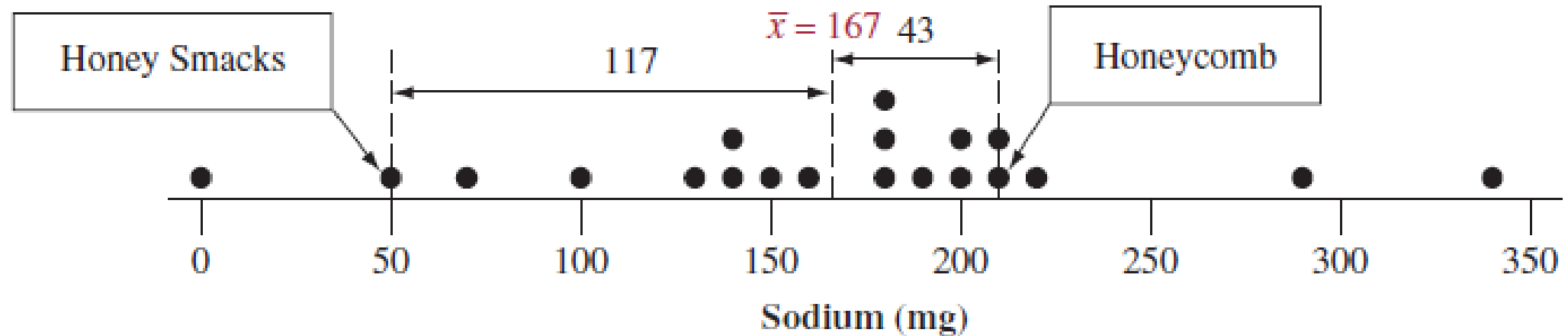


Pertanyaan:

- Berapakah range pendapatan tahunan di Denmark dan USA ? Manakah yang lebih besar?
- Jika ternyata diketahui bahwa terdapat seseorang di Denmark yang memiliki pendapatan 100 ribu dolar per tahun, apakah range bagi pendapatan tahunan di Denmark berubah?

Simpangan Baku (*Standard Deviation*)

- Deviasi = jarak antara suatu pengamatan terhadap rataannya ($x - \bar{x}$)



▲ **Figure 2.11** Dot Plot for Cereal Sodium Data, Showing Deviations for Two Observations. **Question** When is a deviation positive and when is it negative?

The Standard Deviation s

The **standard deviation** s of n observations is

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{\text{sum of squared deviations}}{\text{sample size} - 1}}.$$

This is the square root of the **variance** s^2 , which is an average of the squares of the deviations from their mean,

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}.$$

Notasi:

	Ragam	Simpangan Baku
Populasi	σ^2	s^2
Contoh	σ	s

Mengapa **pembagi** pada rumus
RAGAM maupun **SIMPANGAN BAKU**
contoh adalah **$n - 1$** ?

Perbandingan Simpangan Baku dan *Range*

Ilustrasi

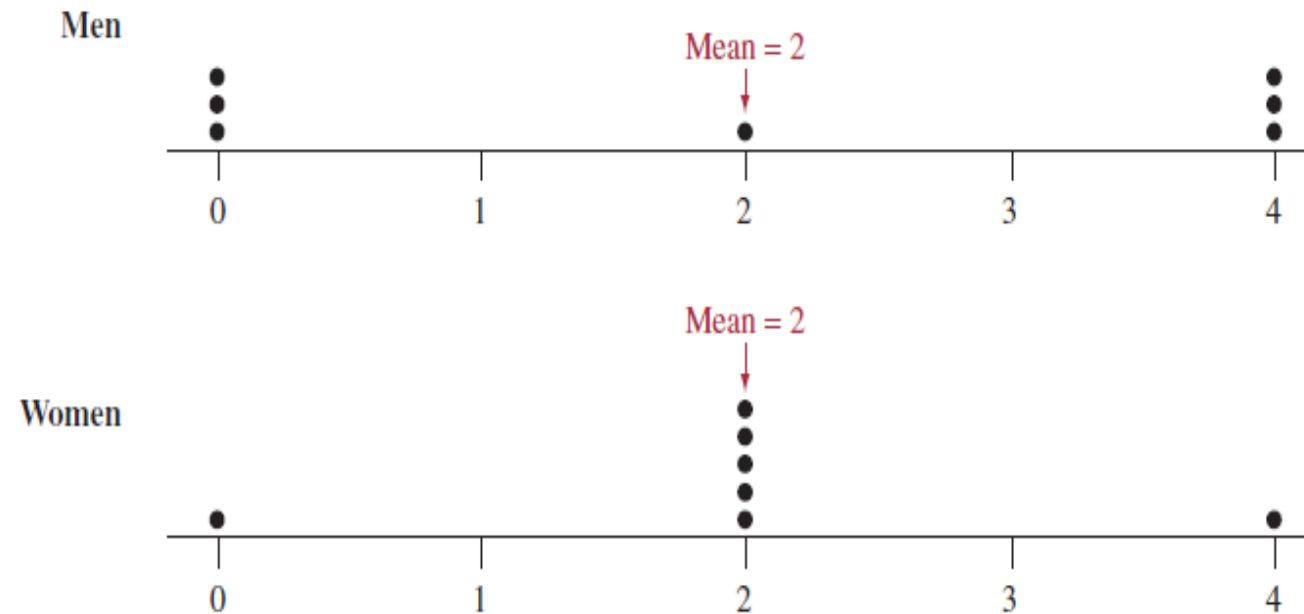
Sebanyak 7 pria dan 7 wanita ditanya, “Berapa banyak jumlah anak yg ideal dalam sebuah keluarga?”

Jawabannya:

Men: 0, 0, 0, 2, 4, 4, 4

Women: 0, 2, 2, 2, 2, 2, 4

Perbandingan Simpangan Baku dan *Range*



- Berapakah *range* dan simpangan baku dari kedua kelompok tsb?

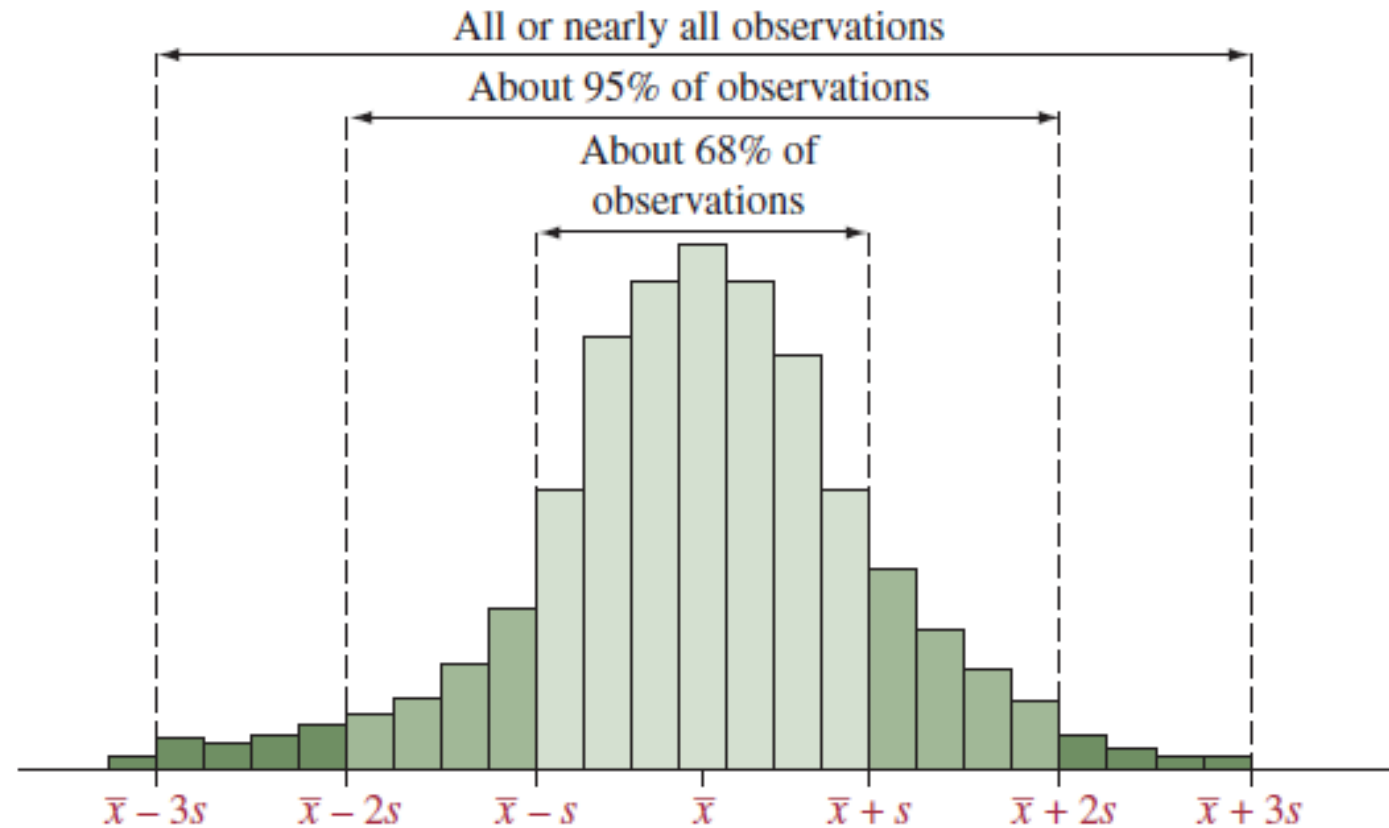
	Range	Simpangan Baku
Pria	4	2
Wanita	4	1.2

Bagaimana kesimpulan
Anda??

Sifat-sifat Simpangan Baku

- Semakin besar keragaman data, semakin besar nilai simpangan baku.
- Jika semua data bernilai sama, maka simpangan baku bernilai nol.
- Simpangan baku tidak resisten terhadap pencilan.

Empirical Rule

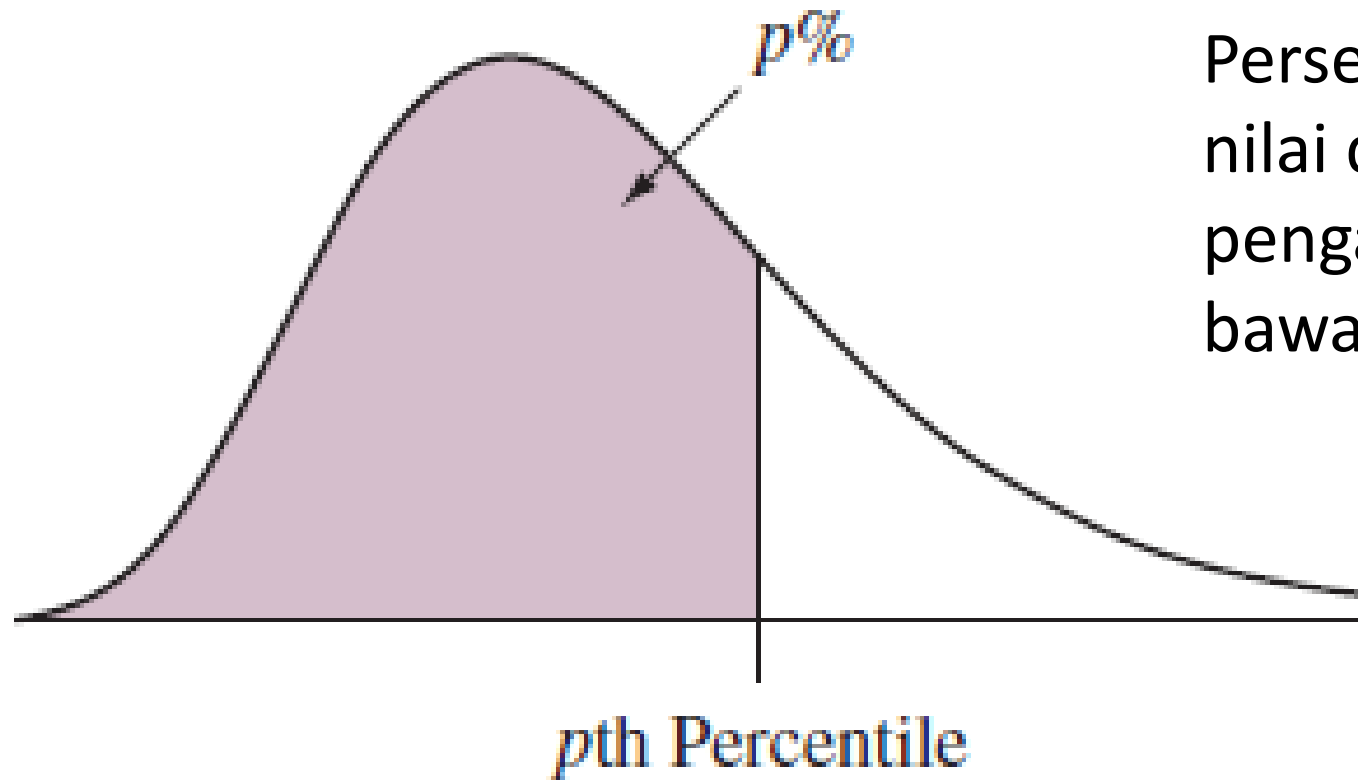


▲ **Figure 2.12 The Empirical Rule.** For bell-shaped distributions, this tells us approximately how much of the data fall within 1, 2, and 3 standard deviations of the mean. **Question** About what percentage would fall *more than* 2 standard deviations from the mean?

Ukuran Keragaman Berdasarkan Posisi Data

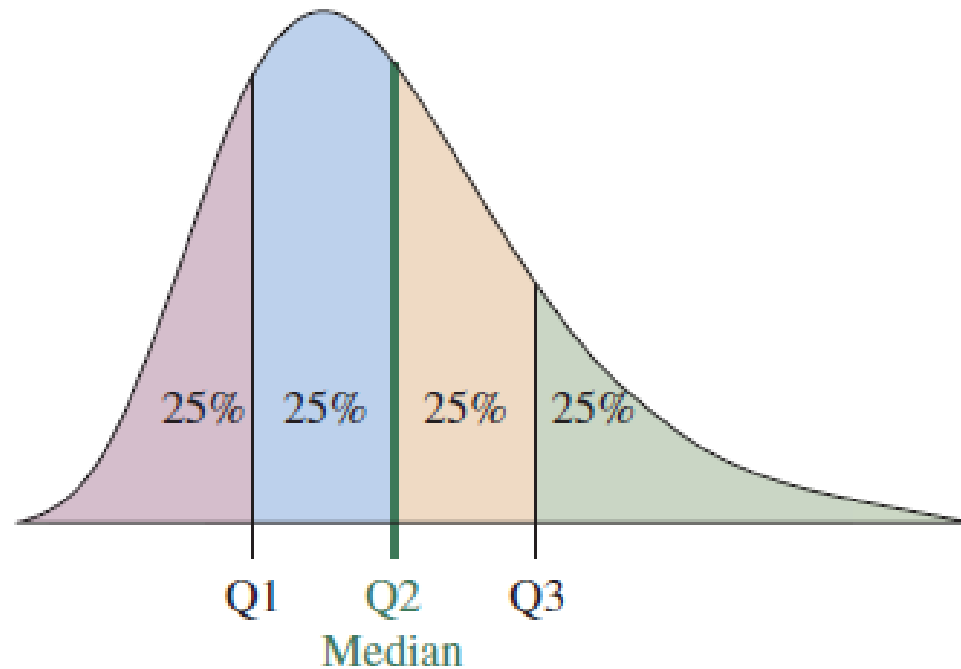
- Persentil
- Kuartil
- Jangkauan Antar Kuartil (*Interquartile Range/IQR*)

Persentil



Persentil ke- p adalah suatu nilai dimana sebanyak $p\%$ pengamatan terletak di bawahnya.

Kuartil



▲ **Figure 2.14 The Quartiles Split the Distribution Into Four Parts.** Twenty-five percent is below the first quartile (Q1), 25% is between the first quartile and the second quartile (the median, Q2), 25% is between the second quartile and the third quartile (Q3), and 25% is above the third quartile. **Question** Why is the second quartile also the median?

Kuartil

- Nilai-nilai yang membagi data terurut menjadi 4 bagian yang sama
- Q0 (dibaca kuartil 0) merupakan nilai minimum dari data
- Q1(dibaca kuartil 1) merupakan nilai yang membagi data 25% data di kiri dan 75% data di kanan
- Q2 (dibaca kuartil 2) merupakan median, membagi data menjadi 50%
- Q3 (dibaca kuartil 3) merupakan nilai yang membagi data 75% data di kiri dan 25% data di sebelah kanan
- Q4 (dibaca kuartil 4) merupakan nilai maksimum dari data
- Nilai Q1, Q2, dan Q3 kekar terhadap pencilan

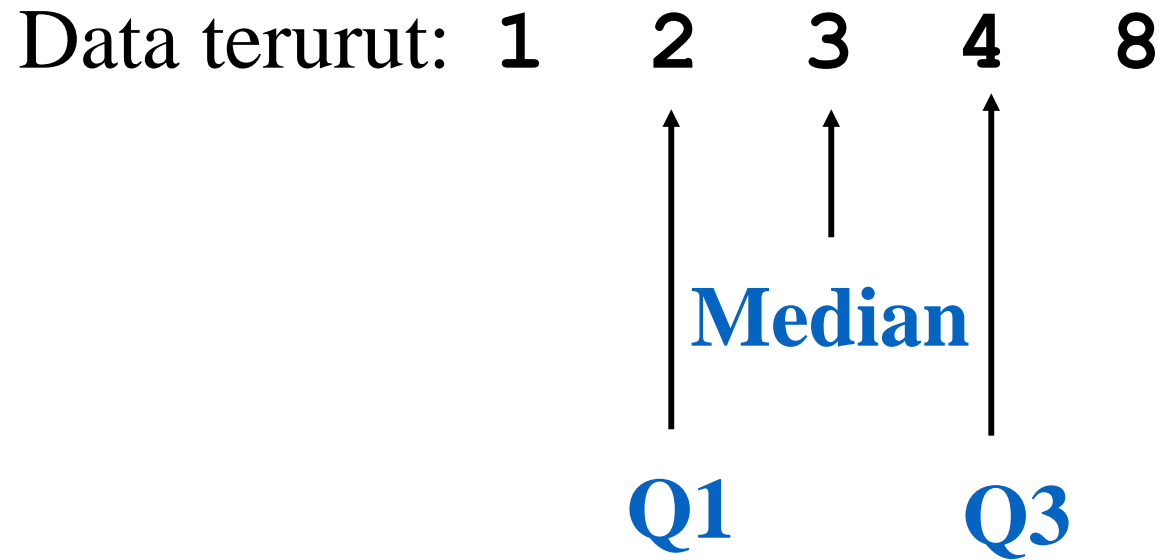
Langkah Teknis memperoleh Kuartil (Quartile)

Metode Belah dua

- Urutkan data dari kecil ke besar
- Cari posisi kuartil
 - $n_{Q2} = (n+1)/2$
 - $n_{Q1} = (n_{Q2}^* + 1)/2 = n_{Q3}$, n_{Q2}^* posisi kuartil dua terpangkas (pecahan dibuang)
- Nilai kuartil 2 ditentukan sama seperti mencari nilai median. Kuartil 1 dan 3 prinsipnya sama seperti median tapi kuartil 1 dihitung dari kiri, sedangkan kuartil 3 dihitung dari kanan.

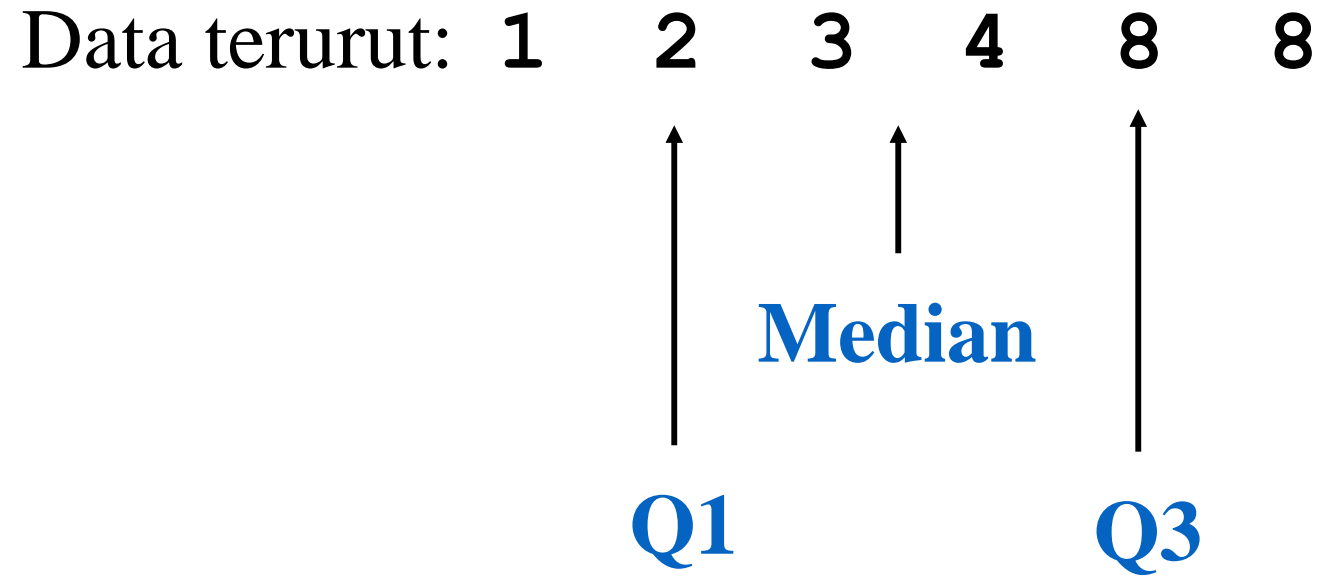
Ilustrasi

- Posisi $Q2 = n_{Q2} = (5+1) / 2 = 3$
- Posisi $Q1 = \text{Posisi } Q3 = (3+1)/2 = 2$



Ilustrasi

- Posisi $Q2 = n_{Q2} = (6+1) / 2 = 3.5$
- Posisi $Q1 = \text{Posisi } Q3 = (3+1)/2 = 2$



Langkah Teknis memperoleh Kuartil (Quartile)

Metode Interpolasi

- Urutkan data dari kecil ke besar
- Cari posisi kuartil
 - $n_{q1} = (1/4)(n+1)$
 - $n_{q2} = (2/4)(n+1)$
 - $n_{q3} = (3/4)(n+1)$
- Nilai kuartil dihitung sebagai berikut:
 - $X_{qi} = X_{a,i} + h_i (X_{b,i} - X_{a,i})$
 - $X_{a,i}$ = pengamatan sebelum posisi kuartil ke-i, $X_{b,i}$ = pengamatan setelah posisi kuartil ke-i dan h_i adalah nilai pecahan dari posisi kuartil

Ilustrasi

- Posisi $Q2 = n_{Q2} = (5+1) / 2 = 3$
- Posisi $Q1 = \frac{1}{4}(5+1) = 1.5$
- Posisi $Q3 = \frac{3}{4}(5+1) = 4.5$

Data terurut: **1 2 3 4 8**

Median



$$Q1 = 1 + 0.5(2-1) = 1.5$$

$$Q3 = 4 + 0.5(8-4) = 6$$

Ilustrasi

- Posisi $Q2 = n_{Q2} = (6+1) / 2 = 3.5$
- Posisi $Q1 = \frac{1}{4}(6+1) = 1.75$
- Posisi $Q3 = \frac{3}{4}(6+1) = 5.25$

Data terurut: **1 2 3 4 8 8**

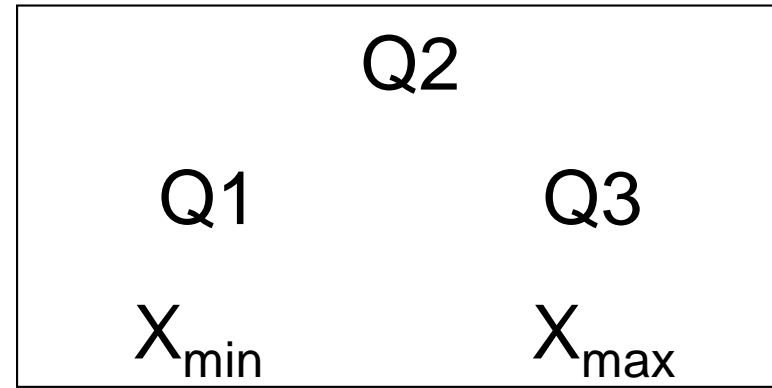
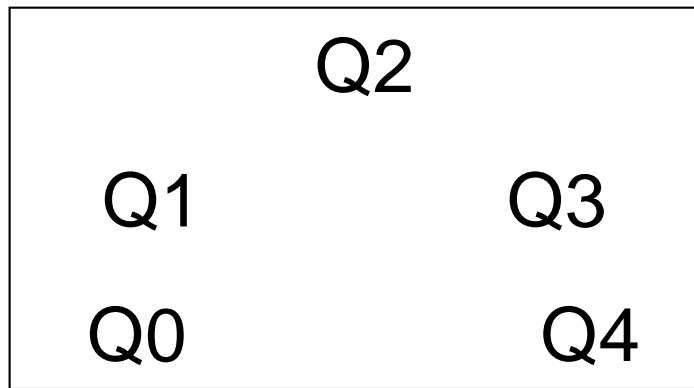
Median



$$Q1 = 1 + 0.75(2-1) = 1.75$$

$$Q3 = 8 + 0.25(8-8) = 8$$

Statistik 5 Serangkai



Jangkauan Antar Kuartil

- Merupakan selisih antara kuartil 3 dengan kuartil 1
$$IQR = Q_3 - Q_1$$
- Memperhitungkan sebaran antara nilai minimum dan nilai maksimum
- Kekar terhadap adanya nilai-nilai yang ekstrim (pencilan)

Latihan

Cereal Sodium Data

0	340	70	140	200	180	210	150	100	130
140	180	190	160	290	50	220	180	200	210

- Tentukan statistik 5 serangkai dari data tersebut!
- Tentukan pula jangkauan antar kuartil dari data tsb!

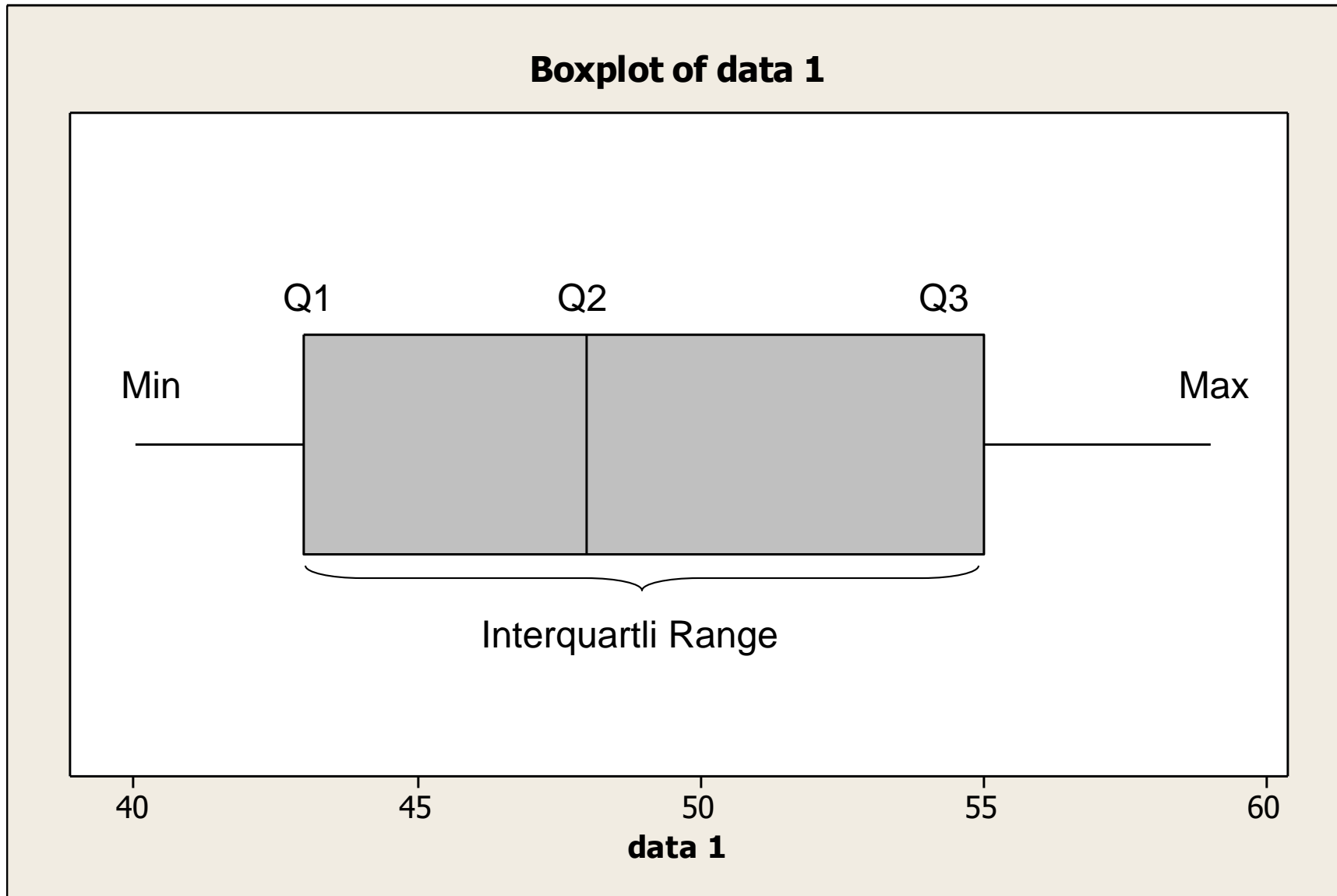


Diagram Kotak Garis (Boxplot)

Informasi yang diperoleh dari diagram kotak garis

- Melihat ukuran penyebaran dan ukuran pemusatan data
- Melihat adanya data pencilan
- Sebagai alat perbandingan sebaran dua kelompok data atau lebih

Penyajian Dengan Box-plot



Cara membuat box plot

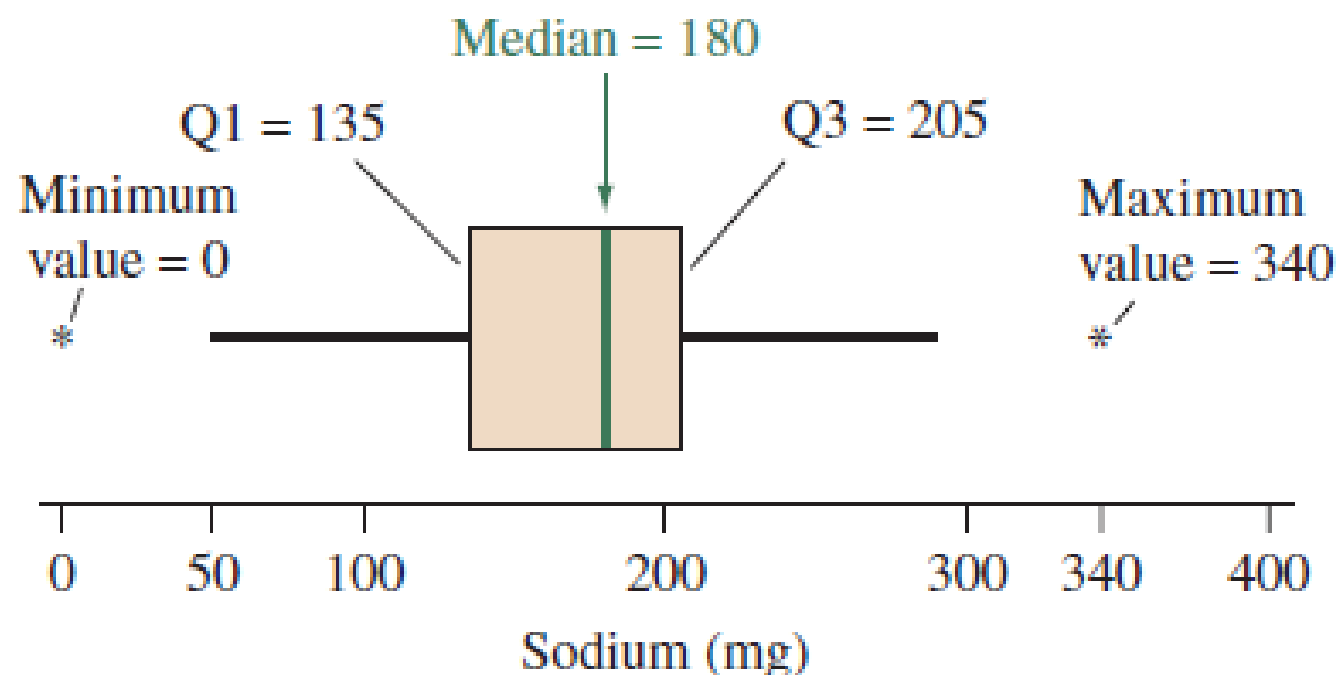
- 1) Hitung Statistik lima serangkai
- 2) Hitung Batas Atas (BA): $Q3 + 1.5(Q3 - Q1)$
- 3) Hitung Batas Bawah (BB): $Q1 - 1.5(Q3 - Q1)$
- 4) Identifikasi data. Jika data $< BB$ atau data $> BA$ maka data dikatakan outlier
- 5) Gambar kotak dengan batas $Q1$ dan $Q3$
- 6) Jika tidak ada pencilan : Tarik garis dari $Q1$ sampai data terkecil dan tarik garis dari $Q3$ sampai data terbesar
- 7) Jika ada pencilan : Tarik garis $Q1$ dan atau $Q3$ sampai data sebelum pencilan
- 8) Pencilan digambarkan dengan asterik

Latihan

Cereal Sodium Data

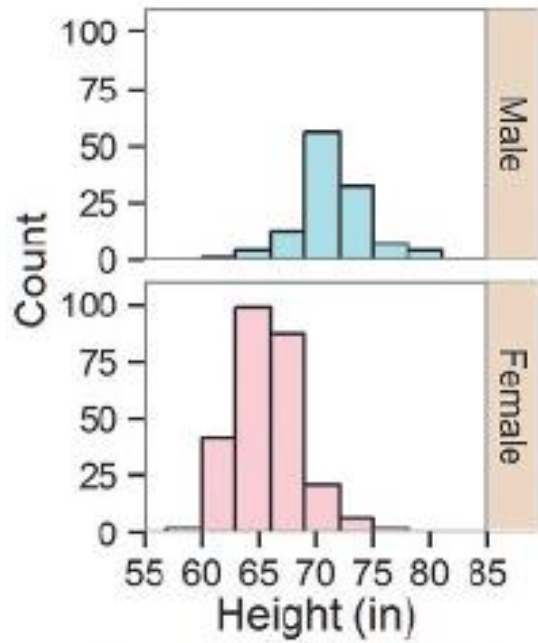
0	340	70	140	200	180	210	150	100	130
140	180	190	160	290	50	220	180	200	210

- Buatlah diagram kotak garis berdasarkan data tsb!
- Berikan komentar Anda tentang sebaran data kandungan sodium tsb!

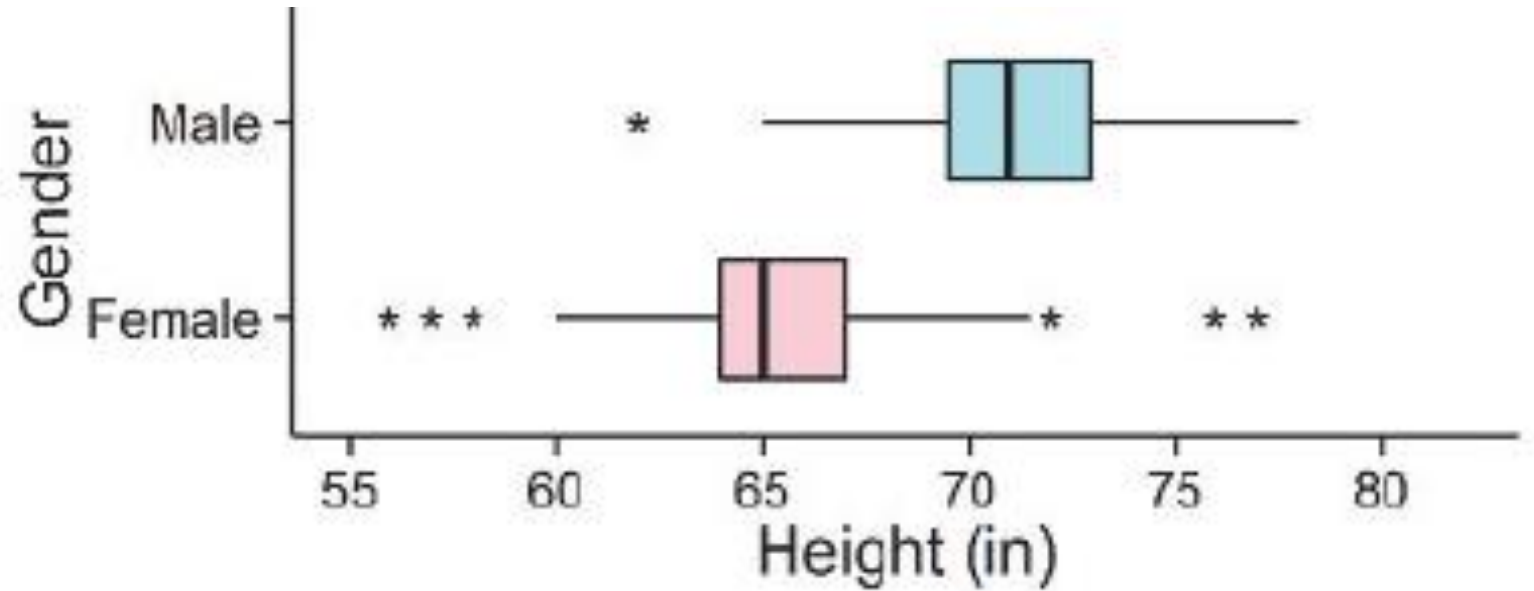


▲ **Figure 2.15 Box Plot and Five-Number Summary for 20 Breakfast Cereal Sodium Values.** The central box contains the middle 50% of the data. The line in the box marks the median. Whiskers extend from the box to the smallest and largest observations, which are not identified as potential outliers. Potential outliers are marked separately. The box plot can also be drawn vertically (see margin on the next page). **Question** Why is the left whisker drawn down only to 50 rather than to 0?

Ilustrasi



Histogram of Height
for Males and Females

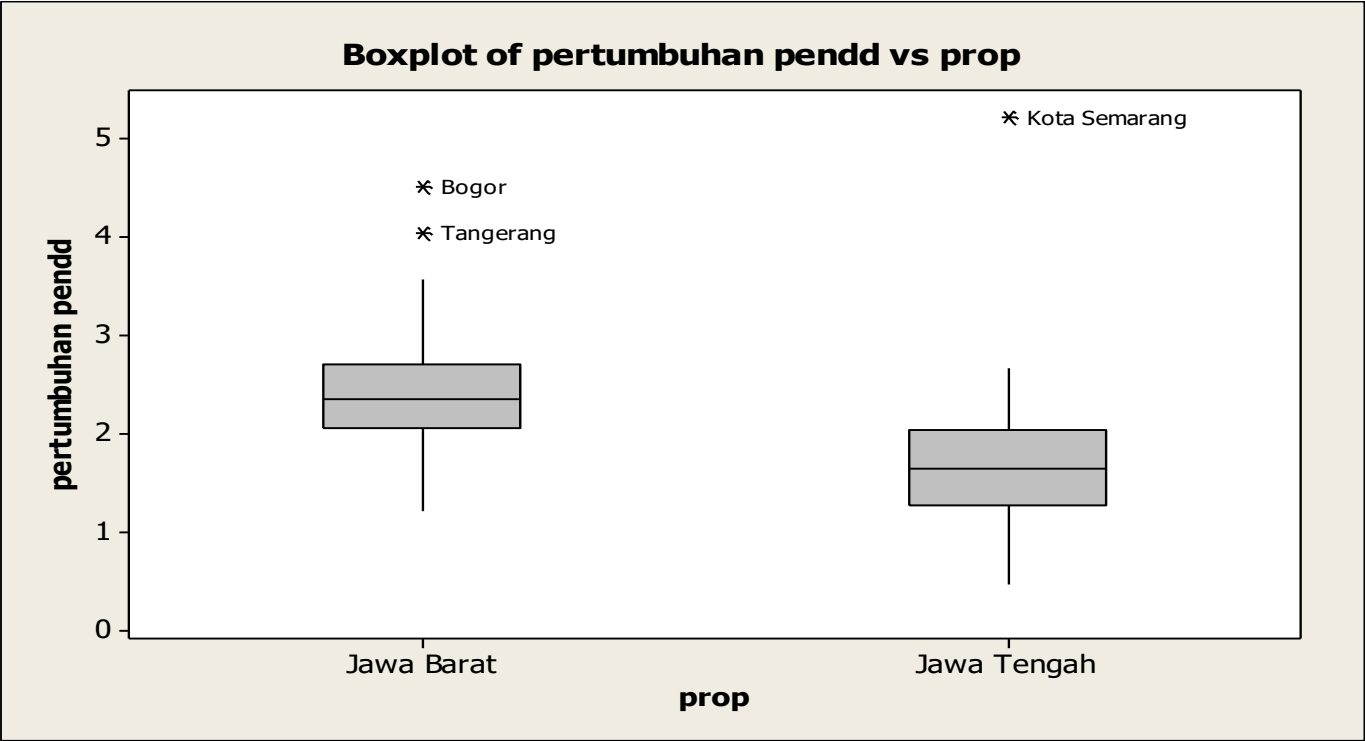


Jawa Barat		
No.	Kota/Kab	Pert. Pend.
1	Pandenglang	2.15
2	Lebak	2.48
3	Bogor	4.52
4	Sukabumi	2.51
5	Cianjur	2.33
6	Bandung	3.31
7	Garut	2.35
8	Tasikmalaya	2.15
9	Ciamis	1.21
10	Kuningan	1.97
11	Cirebon	2.73
12	Majalengka	2.01
13	Sumedang	1.41
14	Indramayu	2.53
15	Subang	1.89
16	Purwakarta	2.32
17	Karawang	2.31
18	Bekasi	3.57
19	Tangerang	4.04
20	Serang	2.85
21	Kota Bogor	2.60
22	Kota Sukabumi	1.48
23	Kota Bandung	2.20
24	Kota Cirebon	2.51

Rata-Rata:	
Jabar	2.48
Jateng	1.68
Minimum :	
Jabar	1.00
Jateng	1.00
Maksimum:	
Jabar	23.00
Jateng	34.00

Jawa Tengah		
No.	Kota/Kab	Pert. Pend.
1	Cilacap	1.28
2	Banyumas	1.78
3	Prubalingga	1.42
4	Banjarnegara	1.49
5	Kebumen	1.09
6	Purworejo	0.62
7	Wonosobo	1.64
8	Magelang	1.31
9	Boyolali	1.08
10	Klaten	1.19
11	Sukoharjo	2.10
12	Wonogiri	0.51
13	Karanganyar	2.07
14	Sragen	1.85
15	Grobogan	1.52
16	Blora	1.27
17	Rembang	2.08
18	Pati	1.62
19	Kudus	2.03
20	Jepara	1.87
21	Demak	1.38
22	Semarang	0.46
23	Temanggung	1.83
24	Kendal	0.83
25	Batang	1.70
26	Pekalongan	1.80
27	Pemalang	1.79
28	Tegal	2.67
29	Brebes	2.09
30	Kota Magelang	1.25
31	Kota Surakarta	1.39
32	Kota Slatiga	2.30
33	Kota Semarang	5.21
34	Kota Pekalongan	1.95
35	Kota Tegal	2.44

Ilustrasi



Pertumbuhan penduduk di Jawa Barat relatif lebih tinggi dibandingkan dengan pertumbuhan penduduk di Jawa Tengah. Secara umum, tingkat keragaman pertumbuhan penduduk antar kabupaten, di Jawa Tengah sedikit lebih besar dibanding dengan Jawa Barat. Kab Bogor dan Tangerang merupakan daerah yang tingkat pertumbuhan penduduknya cukup tinggi. Di Jawa Tengah Kota Semarang yang pertumbuhan penduduknya paling tinggi.



Terimakasih.

TUGAS

Bentuk kelompok yang masing2 terdiri dari 5 orang, lalu kerjakan tugas berikut ini:

1. Sebutkan, peubah apa saja yang dapat diamati pada database kelas kalian? Sebutkan pula skala pengukurannya.
2. Tuliskan minimal 3 informasi **menarik** yang dapat Anda peroleh dari data tersebut.
3. Buatlah tabel atau grafik yang menyajikan informasi yang telah Anda sebutkan pada nomor (2) (boleh menggunakan software).

Tuliskan hasil pekerjaan Anda dalam bentuk power point dan kirimkan ke r.rahma.anisa@gmail.com dengan subjek “**TUGAS STK211 Kelompok (nomor kelompok)**” paling lambat tgl **8 Oktober 2017 (Minggu) pk. 10.00 WIB**.