

# Statistik Bisnis

Week 1-2

Collecting, Organizing and Visualizing  
Data

# Agenda

Time	Activity
<b>First Session</b>	
90 minutes	Collecting and Organizing Data
<b>Second Session</b>	
90 minutes	Visualizing Data

# Objectives

By the end of this class, students will:

- Understand how to collect data in statistic
- Be able to organize categorical and numerical data
- Understand how to read and interpret an organized data (table)
- Be able to visualize categorical and numerical data
- Understand how to make conclusion based on the data visualizations (charts and graphs)

# REVIEW

# 1.4

Untuk masing-masing variabel berikut, tentukan apakah jenisnya kategorikal atau numerikal. Jika variabel tersebut numerikal, tentukan apakah diskrit atau kontinyu. Selain itu, tentukan juga skala pengukurannya.

- a. Jumlah telepon per rumah tangga
- b. Lama waktu (dalam menit) menelepon terlama yang dibuat dalam sebulan
- c. Apakah seseorang didalam rumah memiliki HP yang memiliki fitur Wi-Fi (Wi-Fi-capable cell phone)
- d. Apakah terdapat koneksi internet cepat dirumah tangga

# 1.5

Pada tahun 2008, sebuah universitas di daerah midwestern United States melakukan survei pada mahasiswa tingkat satu yang telah menyelesaikan semester pertamanya. Survei dibagikan secara elektronik pada seluruh 3.727 mahasiswa, dan yang mengisi survei tersebut hanya 2.821 mahasiswa. Dari semua mahasiswa yang disurvei, 90,1% mengindikasikan bahwa mereka belajar dengan mahasiswa lainnya, dan 57,1% mengindikasikan bahwa mereka mengajar mahasiswa lainnya. Laporan tersebut juga mencatat bahwa 61,3% dari seluruh mahasiswa yang disurvei terlambat masuk kelas paling tidak satu kali, dan 45,8% mengakui bahwa mereka bosan di kelas paling tidak satu kali.

- a. Deskripsikan populasinya.
- b. Deskripsikan sampel yang terkumpul.

# Content

## Data Collection

## Organizing Data

- Categorical Data
- Numerical Data

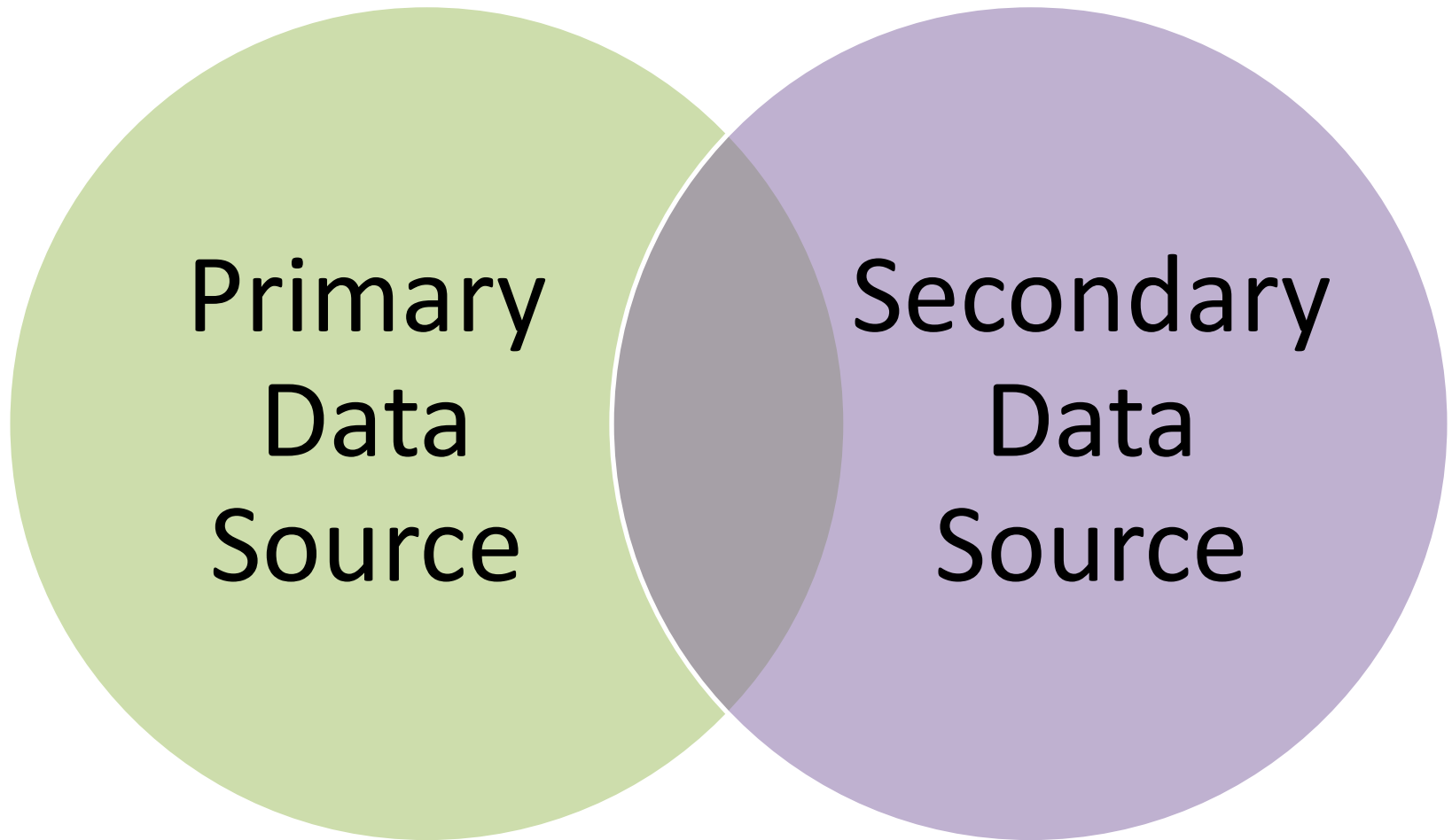
## Visualizing Data

- Categorical Data
- Numerical Data
- Two Numerical Data

# DATA COLLECTION



# Data Collection



# Data Source



# **ORGANIZING DATA**

# Organizing Data

## Categorical Data

The Summary Table (one categorical variable)

The Contingency Table (two categorical variable)

## Numerical Data

The Ordered Array

The Frequency Distribution

The Cumulative Distribution

# CATEGORICAL DATA

# Class Survey



What is your hand phone brand?



What is your phone carrier?

# The Summary Table

## Asal Provinsi Mahasiswa Statistik Bisnis 1 Tahun 2014

Province	Frequency	Percentage
Jawa Barat	13	46.43%
Sulawesi Selatan	5	17.86%
Jakarta	2	7.14%
Jawa Timur	2	7.14%
Sumatera Utara	1	3.57%
Sumatera Selatan	1	3.57%
Sulawesi Tengah	1	3.57%
Banten	1	3.57%
Bali	1	3.57%
Sumatera Barat	1	3.57%
<b>Total</b>	<b>28</b>	<b>100.00%</b>

# The Contingency Table

**Mahasiswa Statistik Bisnis 1 Tahun 2014 Dikelompokkan Berdasarkan Jenis Kelamin dan Status Saudara Kandung**

Jenis Kelamin	Saudara Kandung		Total
	Ada	Tidak ada	
Laki-laki	6	1	7
Perempuan	18	2	20
Total	24	3	27



# The Contingency Table

## Overall Percentage

**Mahasiswa Statistik Bisnis 1 Tahun 2014 Dikelompokkan Berdasarkan Jenis Kelamin dan Status Saudara Kandung**

Jenis Kelamin	Saudara Kandung		Total
	Ada	Tidak ada	
Laki-laki	22%	4%	26%
Perempuan	67%	7%	74%
Total	89%	11%	100%

# The Contingency Table

## Row Percentage

**Mahasiswa Statistik Bisnis 1 Tahun 2014 Dikelompokkan Berdasarkan Jenis Kelamin dan Status Saudara Kandung**

Jenis Kelamin	Saudara Kandung		Total
	Ada	Tidak ada	
Laki-laki	86%	14%	100%
Perempuan	90%	10%	100%
Total	89%	11%	100%

# The Contingency Table

## Column Percentage

**Mahasiswa Statistik Bisnis 1 Tahun 2014 Dikelompokkan Berdasarkan Jenis Kelamin dan Status Saudara Kandung**

Jenis Kelamin	Saudara Kandung		Total
	Ada	Tidak ada	
Laki-laki	25%	33%	26%
Perempuan	75%	67%	74%
Total	100%	100%	100%

# NUMERICAL DATA

# Class Survey

How tall are you?



What is your shoe size?



# The Ordered Array

150 155 155 155 155 156 156 156 156 157  
157 160 160 160 160 162 168 168 168 170  
170 171 173 173 174 174 175

# The Frequency Distribution

Sort raw data in ascending order:

150 155 155 155 155 156 156 156 156 157 157 160 160 160 160 162  
168 168 168 170 170 171 173 173 174 174 175

- Find range: **175 - 150 = 25**
- Select number of classes: **5 (usually between 5 and 15)**
- Compute class interval (width): **5 (25/5 then round up)**
- Determine class boundaries (limits):
  - Class 1: 150 to less than 155
  - Class 2: 155 to less than 160
  - Class 3: 160 to less than 165
  - Class 4: 165 to less than 170
  - Class 5: 170 to less than 175
  - Class 6: 175 to less than 180
- Compute class midpoints: **152.5, 157.5, 162.5, 167.5, 172.5, 177.5**
- Count observations & assign to classes

# The Frequency Distribution

## Tinggi Badan Mahasiswa Statistik Bisnis 1 Tahun 2014

Height	Frequency
150 but less than 155	1
155 but less than 160	10
160 but less than 165	5
165 but less than 170	3
170 but less than 175	7
175 but less than 180	1
<b>Total</b>	<b>27</b>



# The Relative Frequency Distribution and the Percentage Distribution

## Tinggi Badan Mahasiswa Statistik Bisnis 1 Tahun 2014

Height	Relative Frequency	Percentage
150 but less than 155	0.04	4%
155 but less than 160	0.37	37%
160 but less than 165	0.19	19%
165 but less than 170	0.11	11%
170 but less than 175	0.26	26%
175 but less than 180	0.04	4%
<b>Total</b>	<b>1</b>	<b>100.00%</b>

# Developing the Cumulative Percentage Distribution

## Tinggi Badan Mahasiswa Statistik Bisnis 1 Tahun 2014

Height	Percentage (%)	Percentage of Meals Less Than Lower Boundary of Class Interval (%)
150 but less than 155	4	0
155 but less than 160	37	4
160 but less than 165	19	41=4+37
165 but less than 170	11	50=4+37+19
170 but less than 175	26	70=4+37+19+11
175 but less than 180	4	96=4+37+19+11+26

# The Cumulative Distribution

## Tinggi Badan Mahasiswa Statistik Bisnis 1 Tahun 2014

Height	Cumulative Percentage less than indicated value
150	0
155	4%
160	41%
165	59%
170	70%
175	96%
<b>180</b>	<b>100%</b>

# **VISUALIZING DATA**

# Visualizing Data

## Categorical Variable

- Visualizing one variable
  - Bar chart, Pie chart an Pareto chart
- Visualizing two variables
  - Side-by-side bar chart

## Numerical Variable

- Visualizing one variable
  - Stem-and-leaf display
  - Histogram, polygon and ogive
- Visualizing two variables
  - Scatter plot and time-series plot

# Visualizing Data

## Categorical Variable

- Visualizing one variable
  - Bar chart, Pie chart and Pareto chart
- Visualizing two variables
  - Side-by-side bar

## Numerical Variable

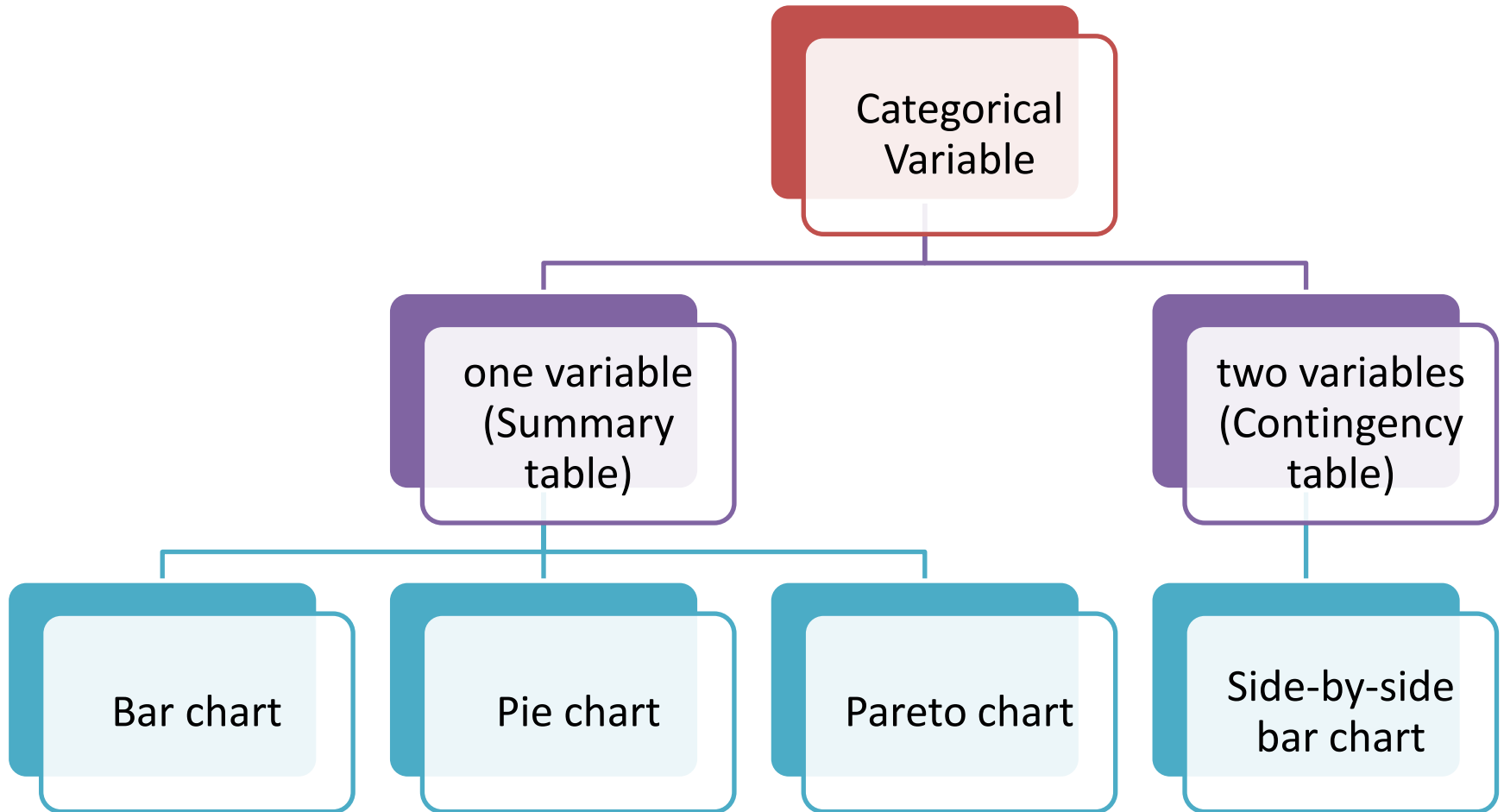
- Visualizing one variable
  - Stem-and-leaf plot
  - Histogram, polygon
- Visualizing two variables
  - Scatter plot and time-series plot



**Graphical  
Errors**

# **CATEGORICAL VARIABLE**

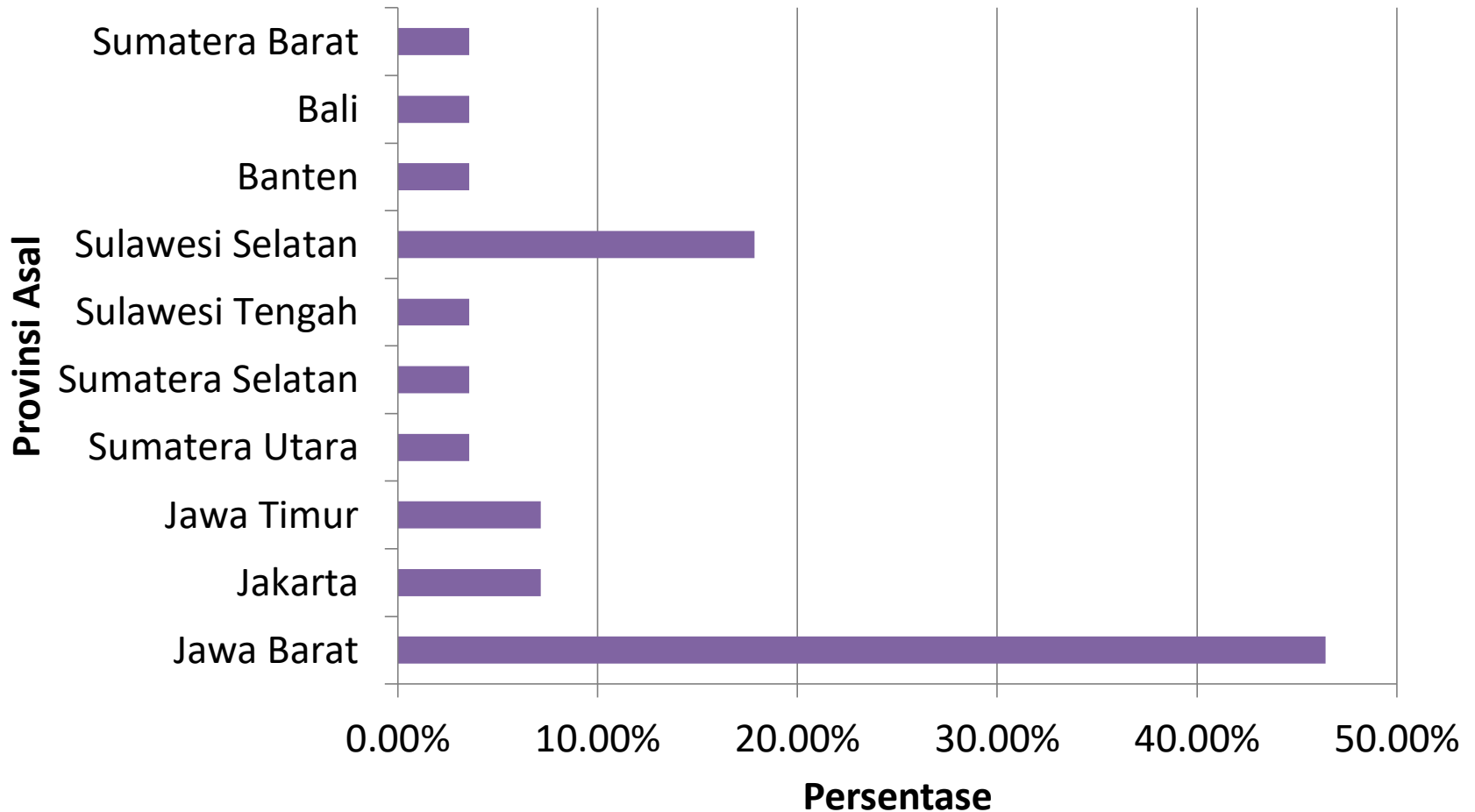
# Visualizing Data





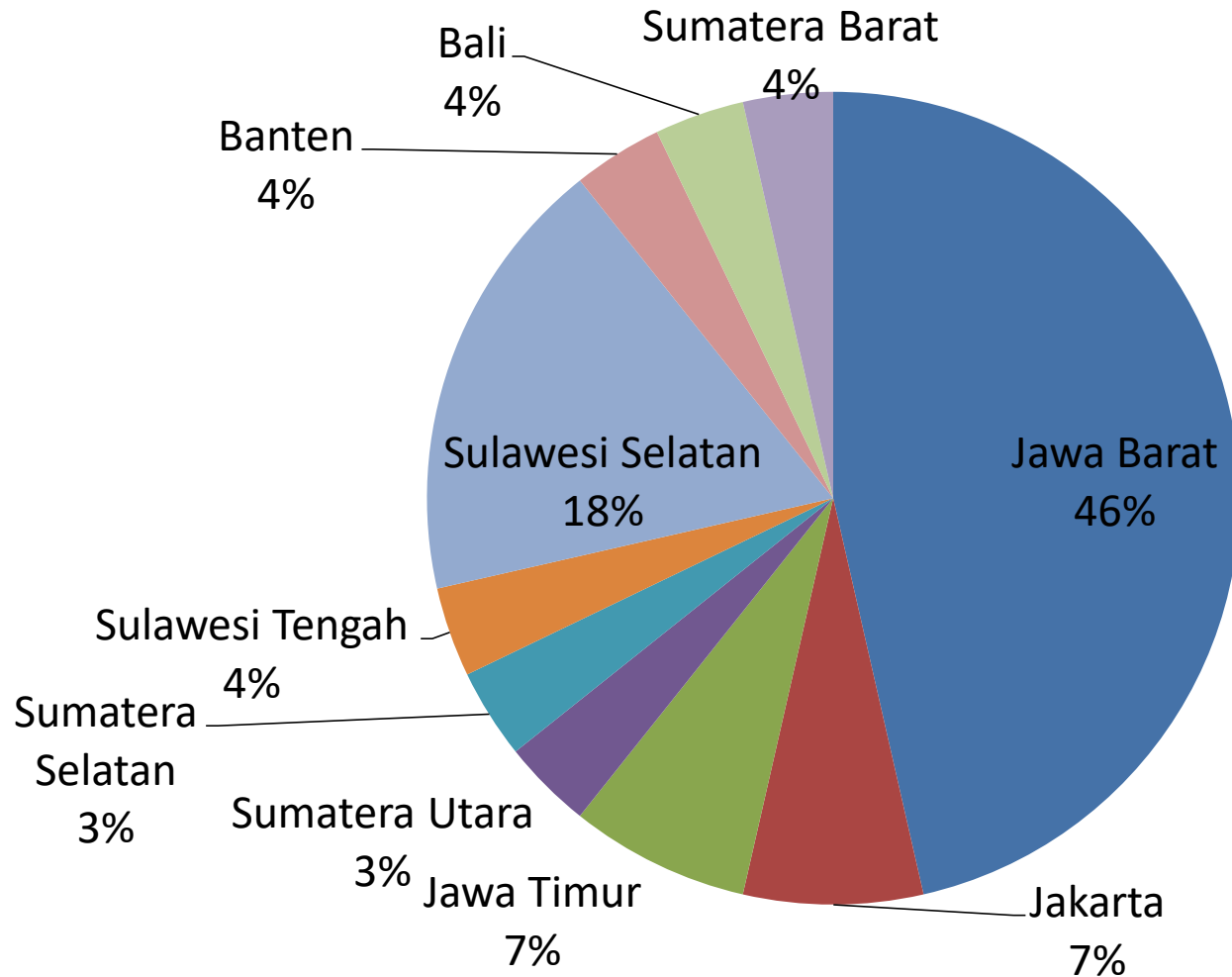
# Bar Chart

Provinsi Asal Mahasiswa Statistika Bisnis 1 tahun 2014



# Pie Chart

Provinsi Asal Mahasiswa Statistika Bisnis 1 tahun 2014

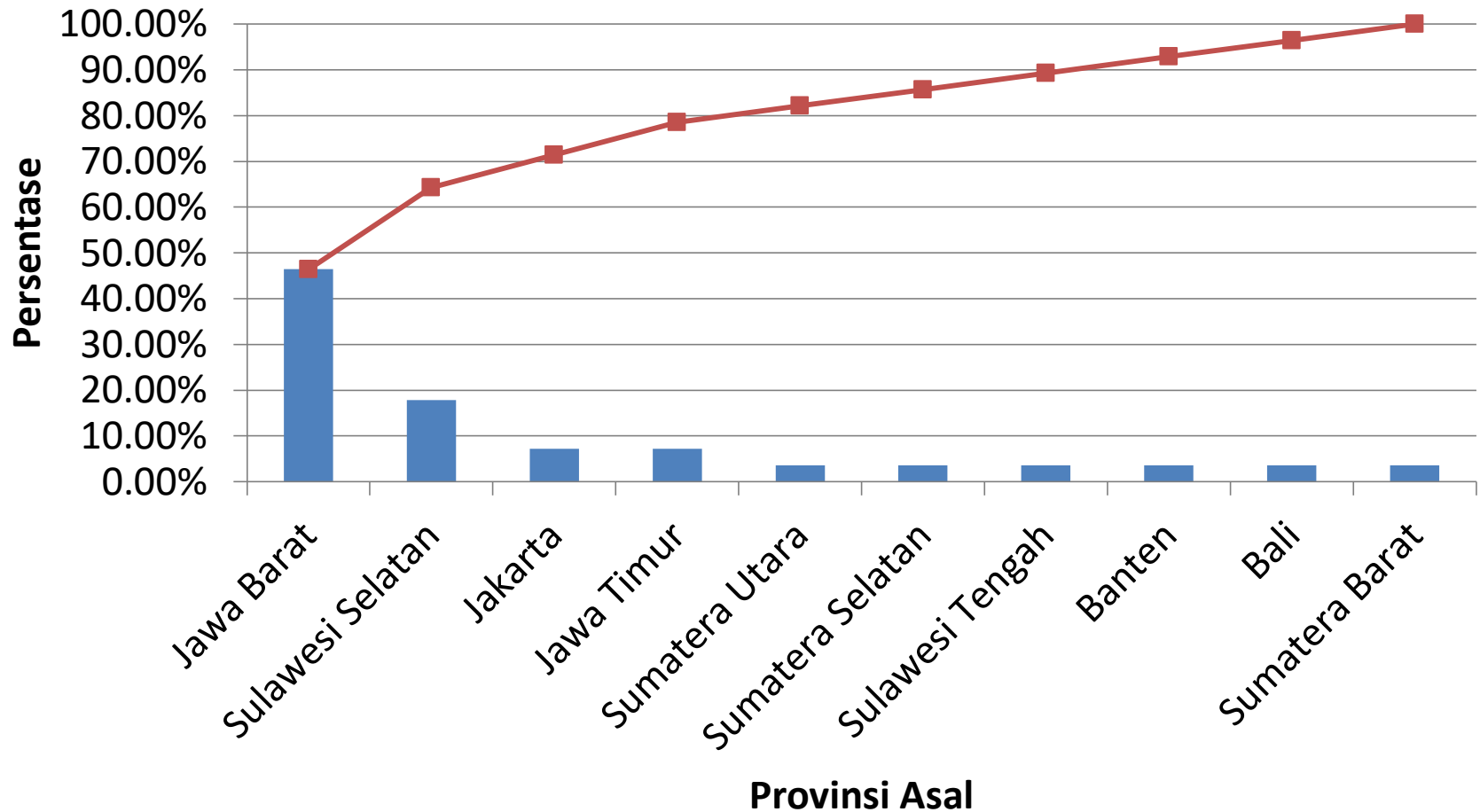


# Pareto Chart

- A Pareto chart has the capability to separate the “vital few” from the “trivial many,” enabling you to focus on the important categories.
- In situations in which the data involved consist of defective or nonconforming items, a Pareto chart is a powerful tool for prioritizing improvement efforts.

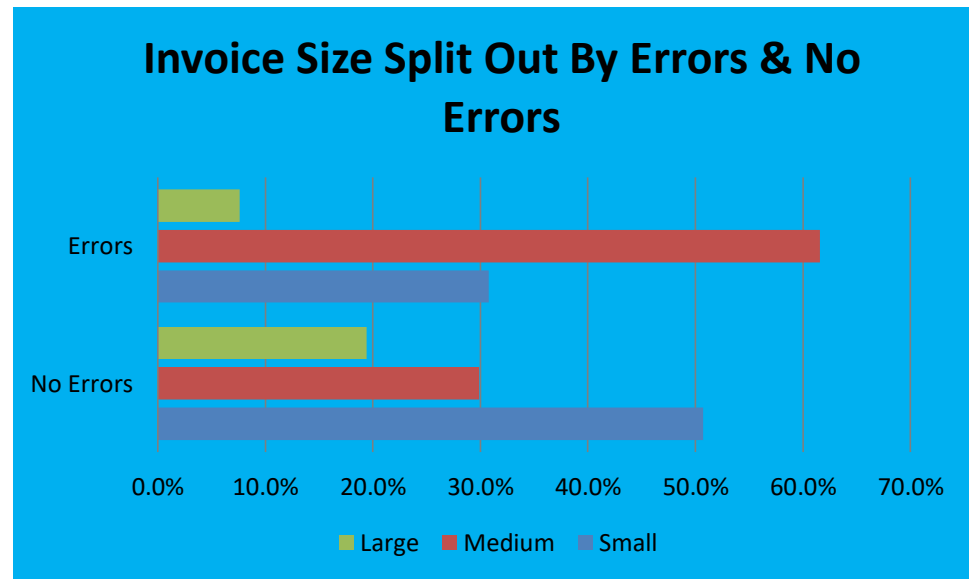
# Pareto Chart

Provinsi Asal Mahasiswa Statistika Bisnis 1 tahun 2014

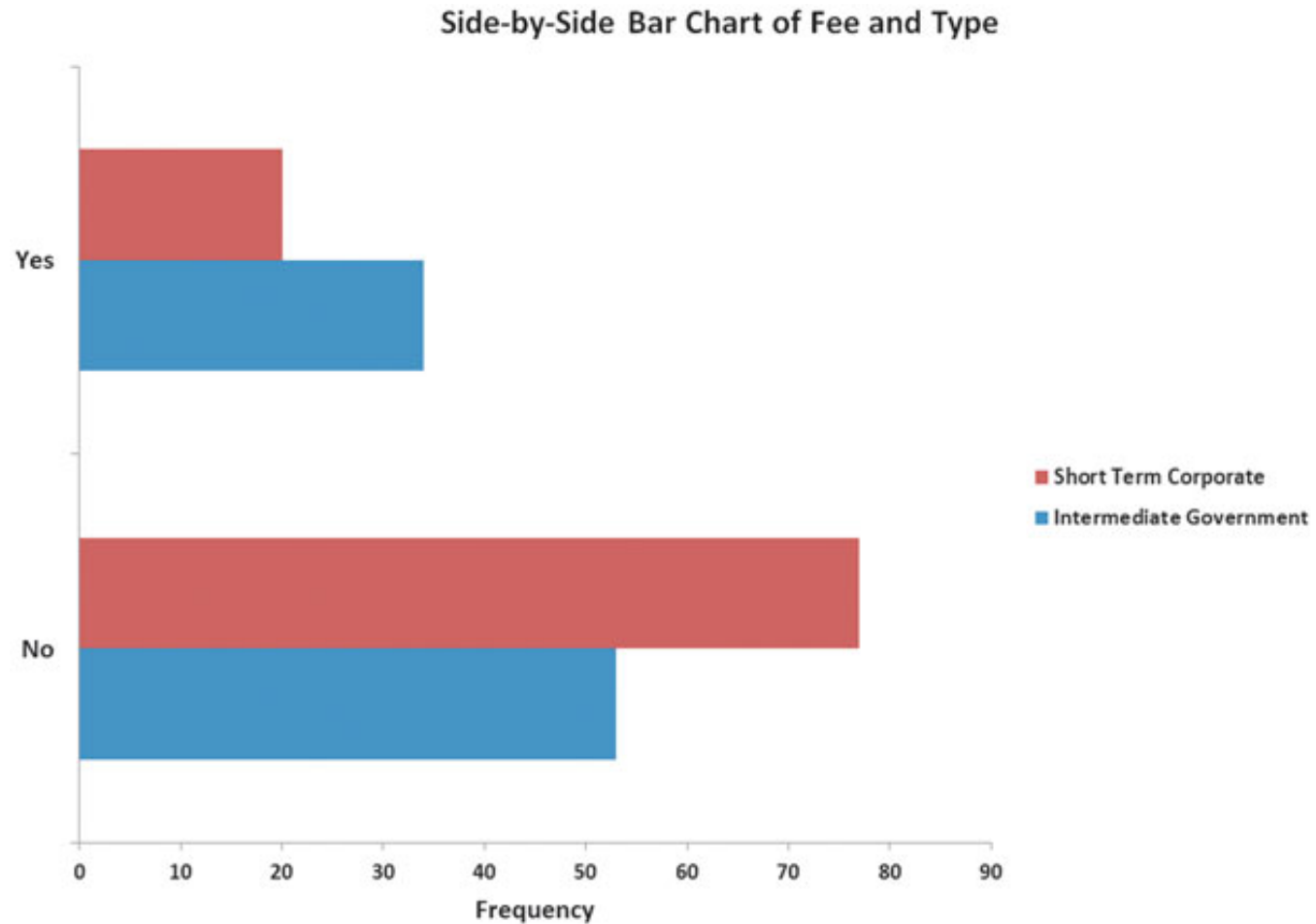


# Side-By-Side Bar Chart

	No Errors	Errors	Total
Small Amount	50.75%	30.77%	47.50%
Medium Amount	29.85%	61.54%	35.00%
Large Amount	19.40%	7.69%	17.50%
Total	100.0%	100.0%	100.0%

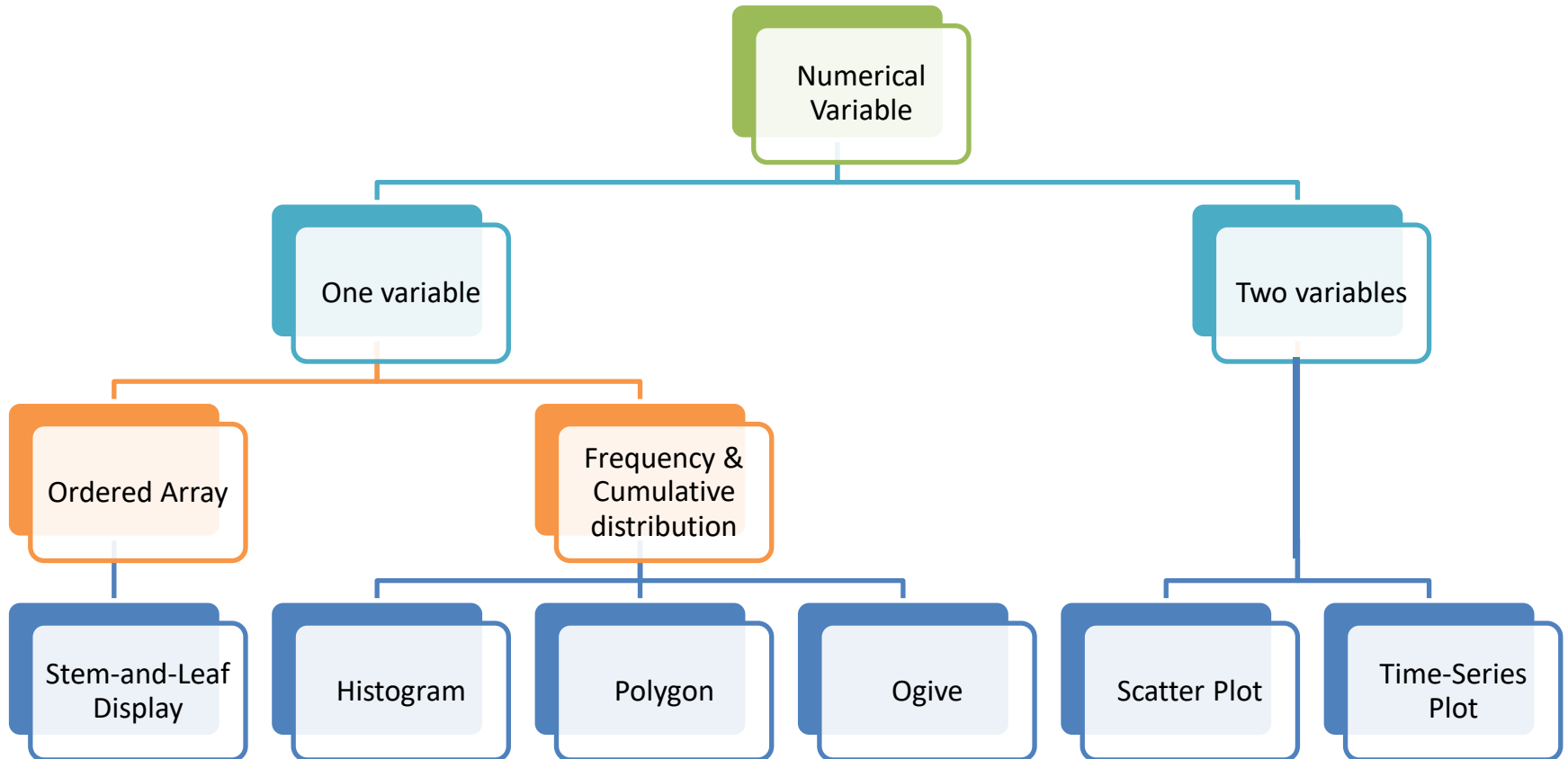


# Side-By-Side Bar Chart



# **NUMERICAL VARIABLE**

# Visualizing Data



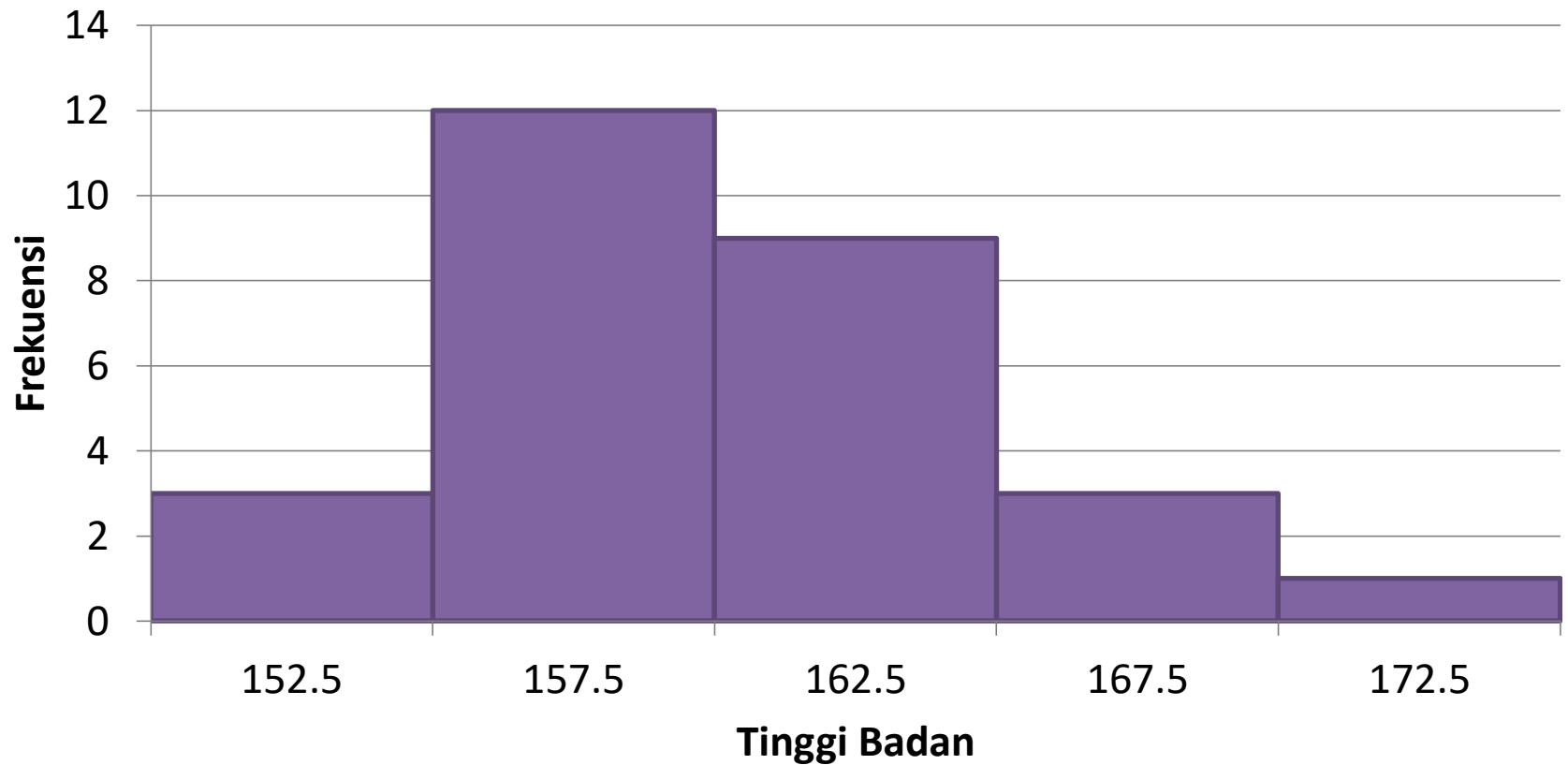


# Stem-and-Leaf Display

Stem (Batang)	Leaf (Daun)
15	0 2 4 5 5 5 5 5 5 7 8 8 8 9 9
16	0 0 0 0 0 0 1 2 3 5 5 5
17	0

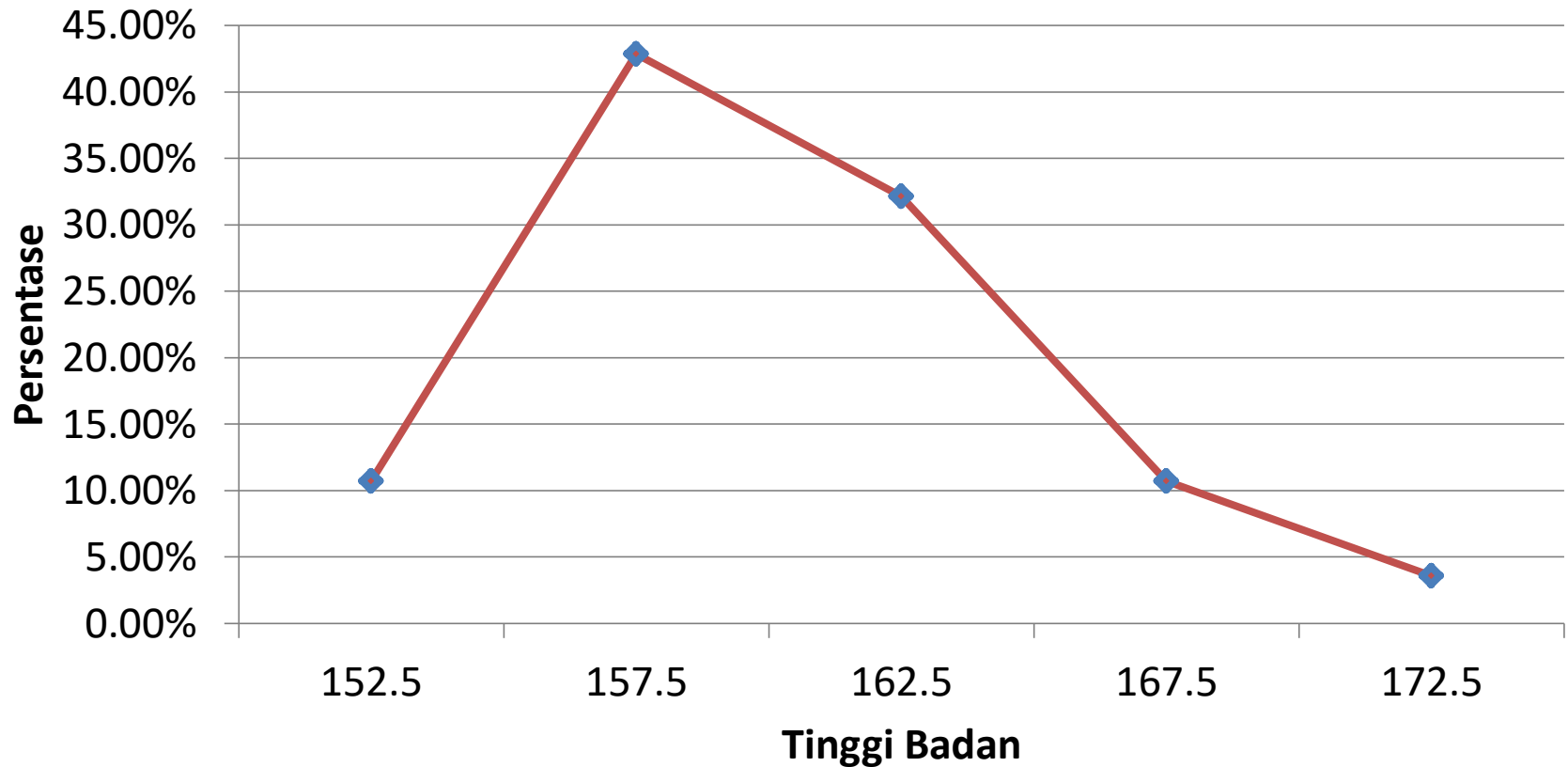
# Histogram

Tinggi Badan Mahasiswa Statistika Bisnis 1 Tahun 2014



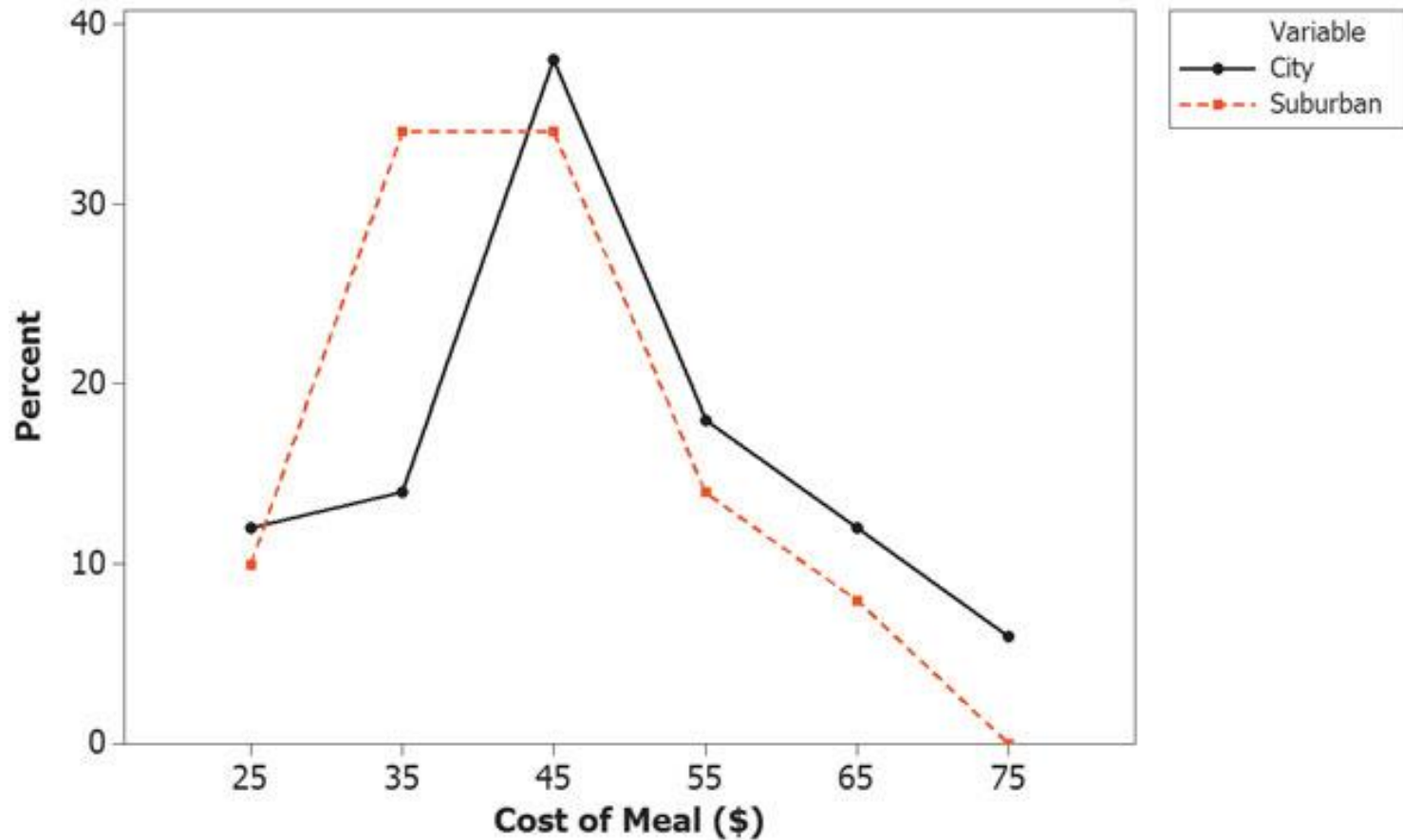
# Percentage Polygon

Tinggi Badan Mahasiswa Statistika Bisnis 1 Tahun 2014



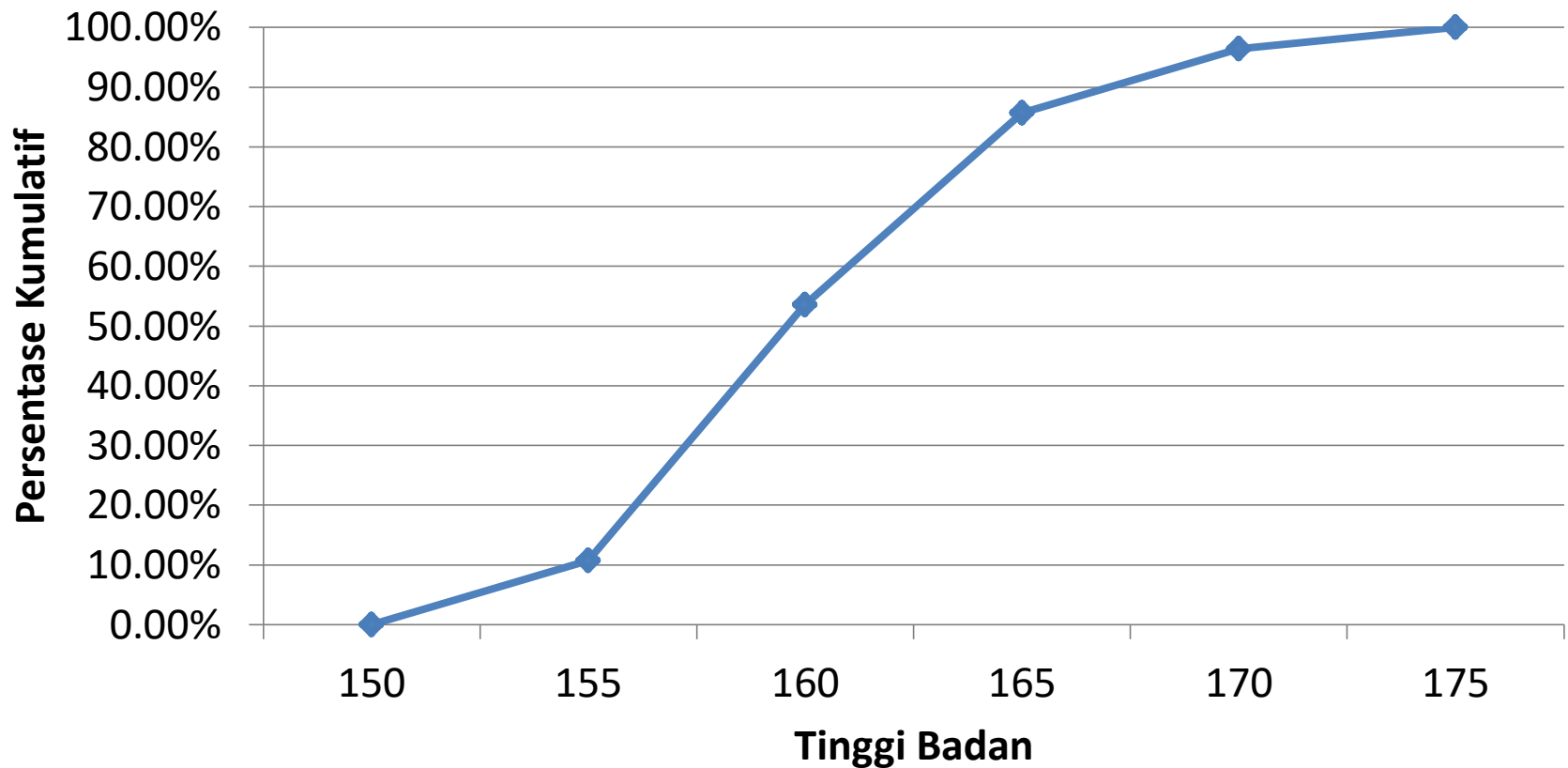
# Percentage Polygon

Percentage Polygons for Cost of Meals at City and Suburban Restaurants



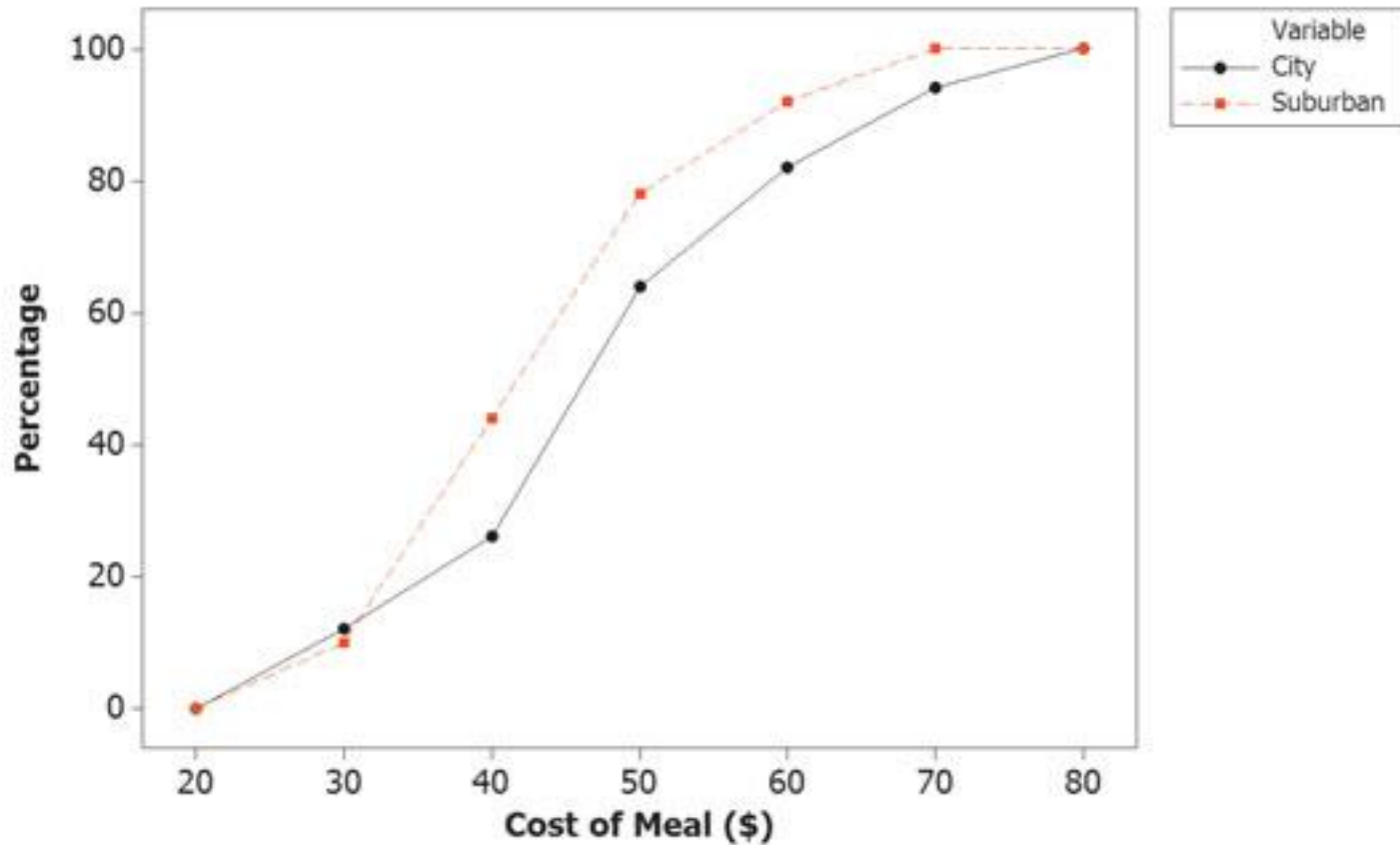
# Cumulative Percentage Polygon (Ogive)

Tinggi Badan Mahasiswa Statistika Bisnis 1 Tahun 2014

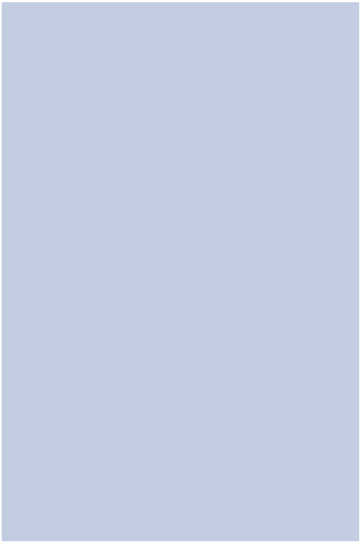


# Cumulative Percentage Polygon (Ogive)

Cumulative Percentage Polygons for Cost of Meals at City and Suburban Restaurants



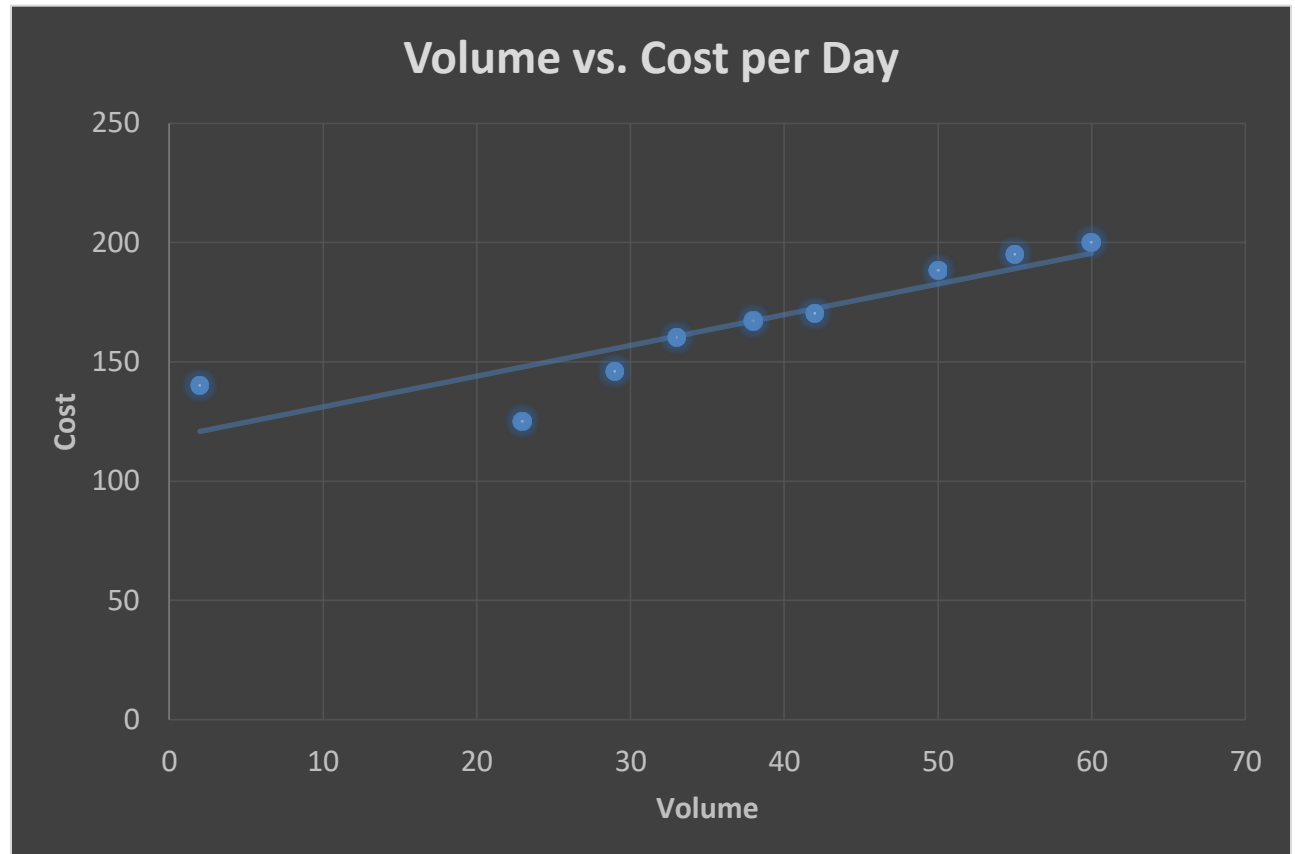
# Note!



When you construct polygons or histograms, the vertical ( $Y$ ) axis should show the true zero, or “origin,” so as not to distort the character of the data.

# Scatter Plot

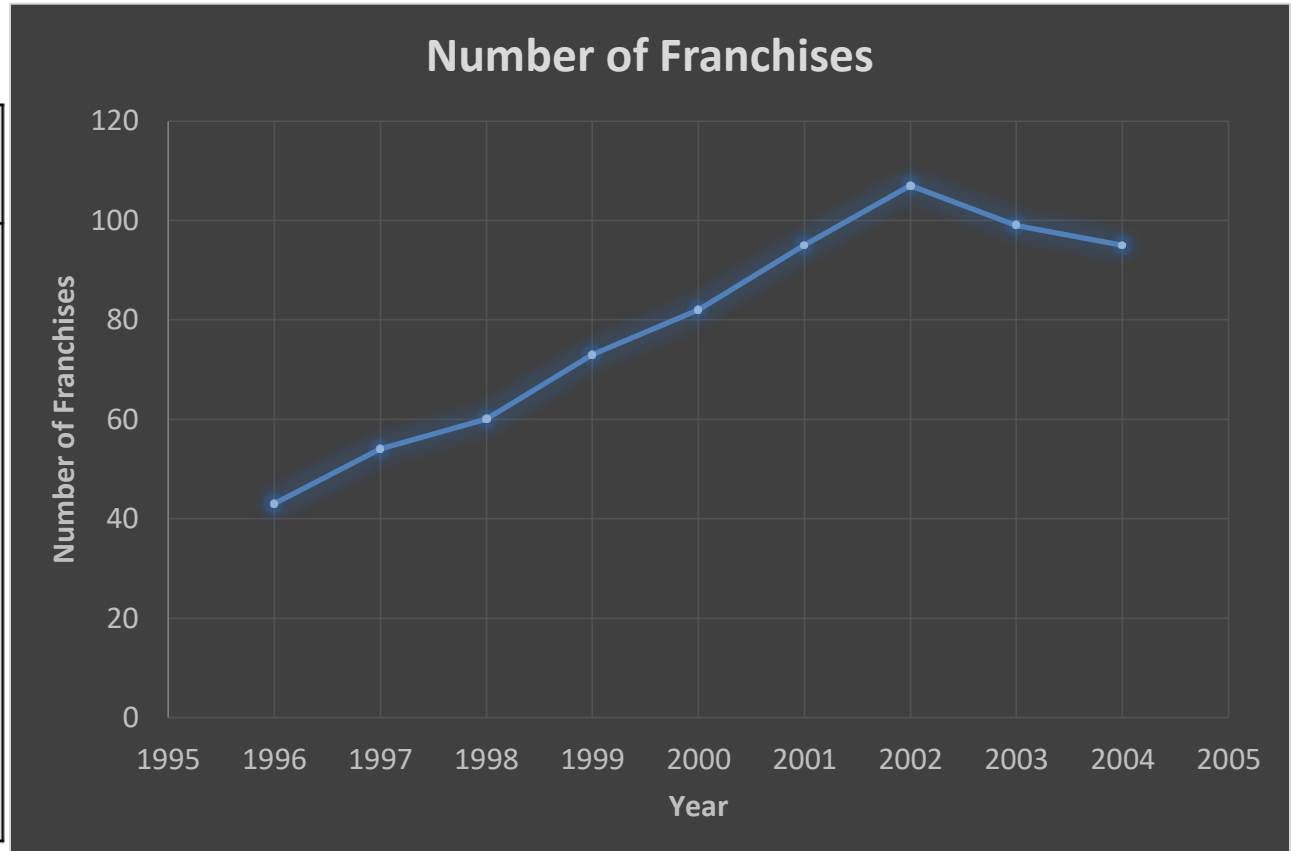
Volume per day	Cost per day
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200





# Time Series Plot

Year	Number of Franchises
1996	43
1997	54
1998	60
1999	73
2000	82
2001	95
2002	107
2003	99
2004	95



# Principles of Excellent Graphs

The graph should not distort the data.

The graph should not contain unnecessary adornments (sometimes referred to as chart junk).

The scale on the vertical axis should begin at zero.

All axes should be properly labeled.

The graph should contain a title.

The simplest possible graph should be used for a given set of data.

# Graphical Errors: Chart Junk

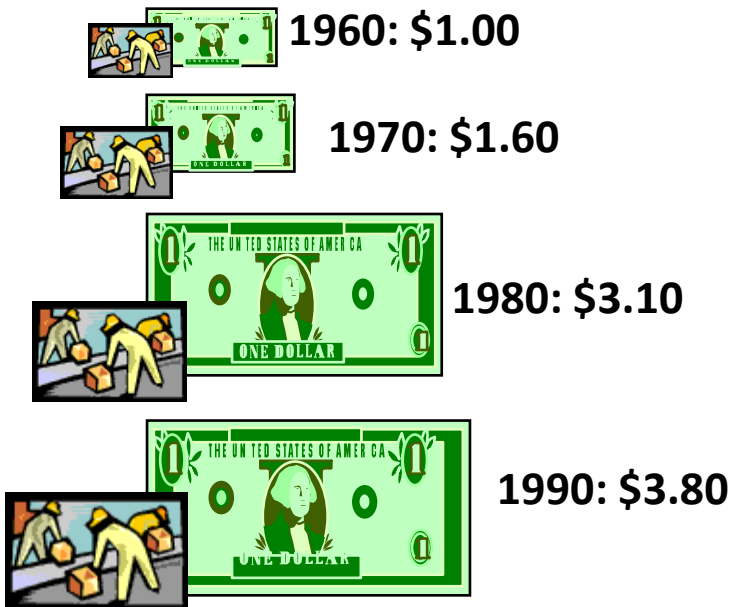


Bad Presentation

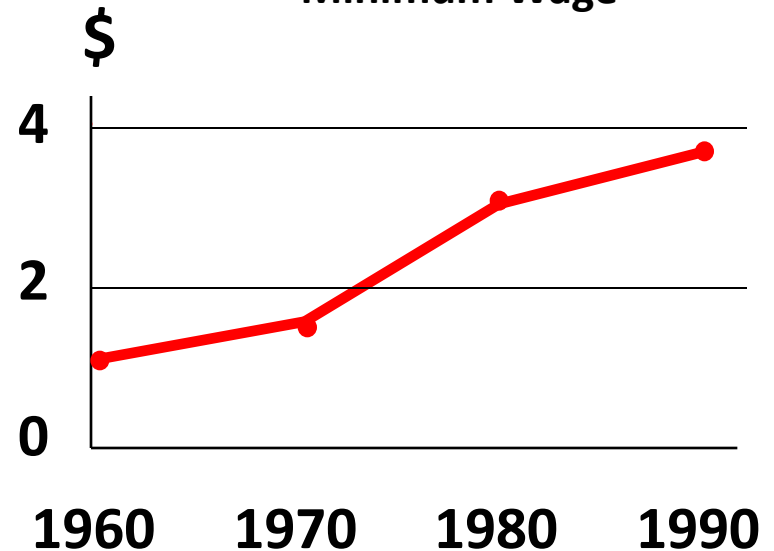


Good Presentation

Minimum Wage



Minimum Wage



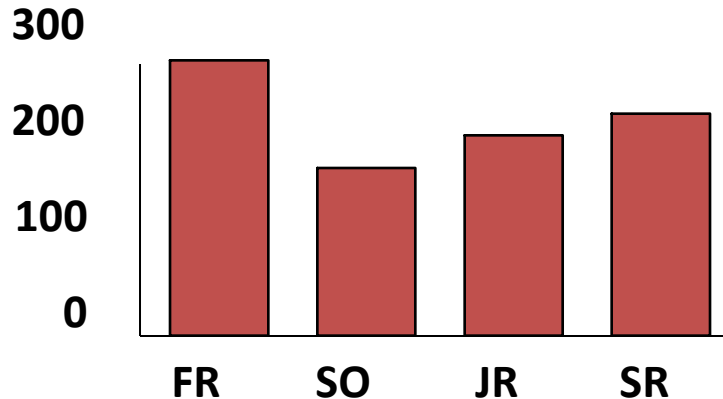
# Graphical Errors: No Relative Basis



## Bad Presentation

A's received by  
students.

Freq.



## Good Presentation

A's received by  
students.

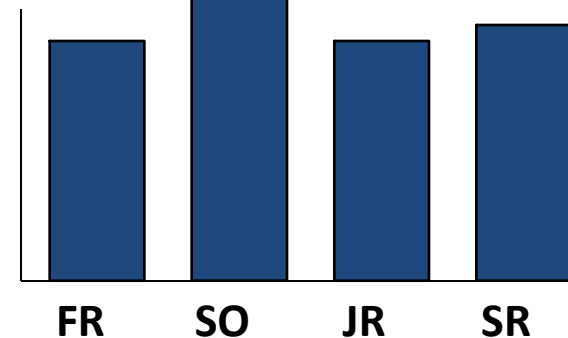
%

30%

20%

10%

0%

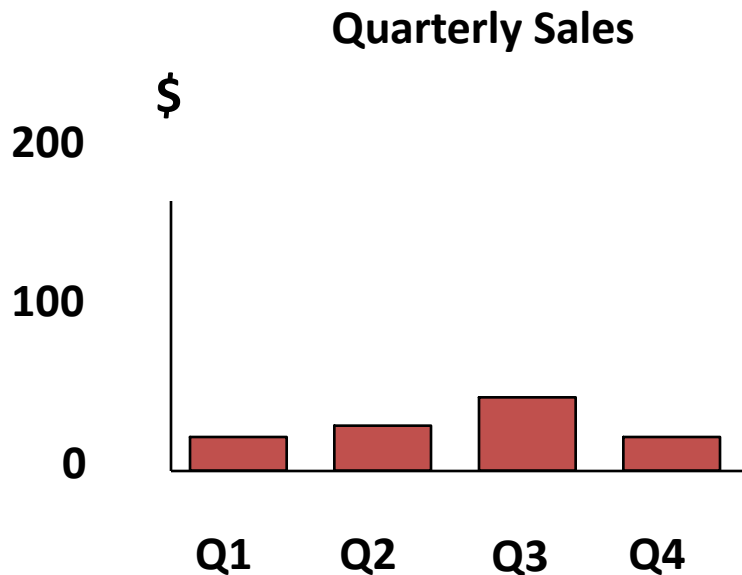


FR = Freshmen, SO = Sophomore, JR = Junior, SR = Senior

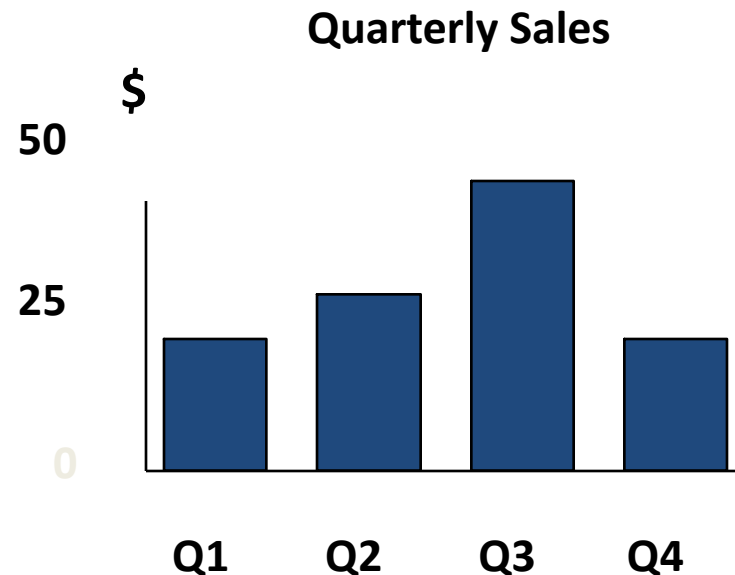
# Graphical Errors: Compressing the Vertical Axis



## Bad Presentation



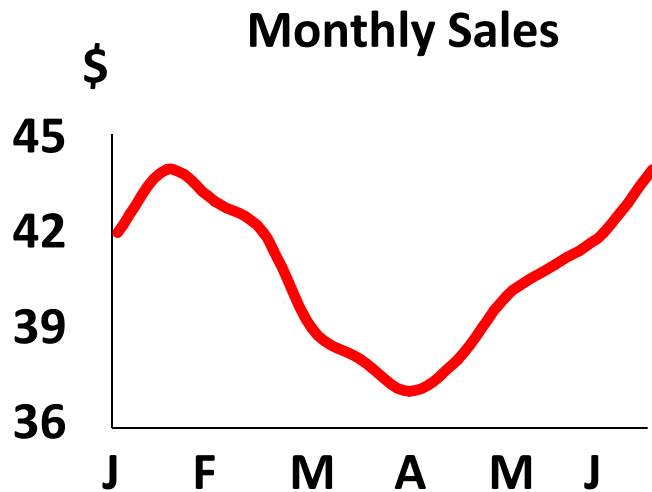
## Good Presentation



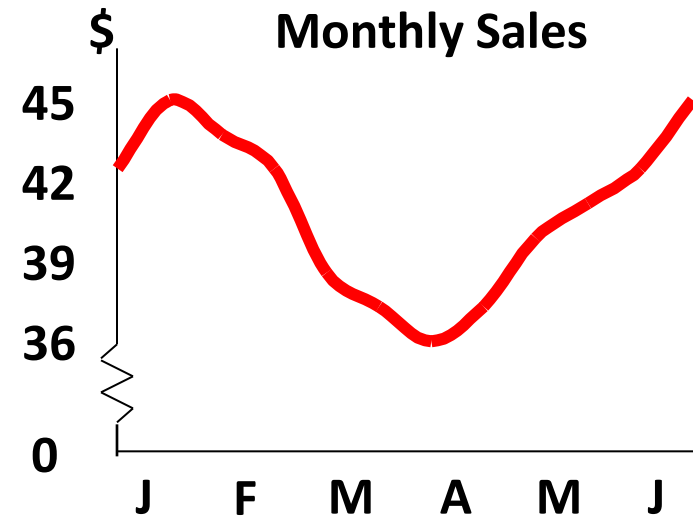
# Graphical Errors: No Zero Point on the Vertical Axis



**Bad Presentation**



**Good Presentations**



**Graphing the first six months of sales**

# EXERCISE

## 2.28

Tabel berikut menunjukkan persentase konsumsi listrik rumah tangga di Amerika Serikat yang disusun berdasarkan jenis alat elektronik pada tahun 2012:



## 2.28

Jenis Alat Elektronik	Persentase (%)
AC	18
Pengering pakaian	5
Mesin cuci	24
Komputer	1
Alat memasak	2
Pencuci Piring	2
Freezer	2
Penerangan	16
Kulkas	9
Penghangat ruangan	7
Pemanas Air	8
TV dan perangkatnya	6

## 2.28

- a. Gambarkan bar chart, pie chart, dan Pareto chart untuk data tersebut.
- b. Grafik manakah yang paling cocok menurut anda untuk menggambarkan data tersebut?

## 2.37

Berikut data biaya per ons (\$) dari sampel 14 batang cokelat pekat:

0,68	0,72	0,92	1,14	1,42	0,94	0,77
0,57	1,51	0,57	0,55	0,86	1,41	0,90

- Urutkanlah data tersebut.
- Buatlah diagram batang-daun.
- Mana yang memberikan lebih banyak informasi, data yang telah diurutkan atau diagram batang-daun? Diskusikan.
- Disekitar nilai apakah, jika ada, biaya cokelat pekat batangan tersebut terkonsentrasi? Jelaskan.

## 2.38

Berikut data biaya listrik pada bulan juli 2010 dari sampel acak 50 apartemen dengan satu kamar tidur di kota besar:

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

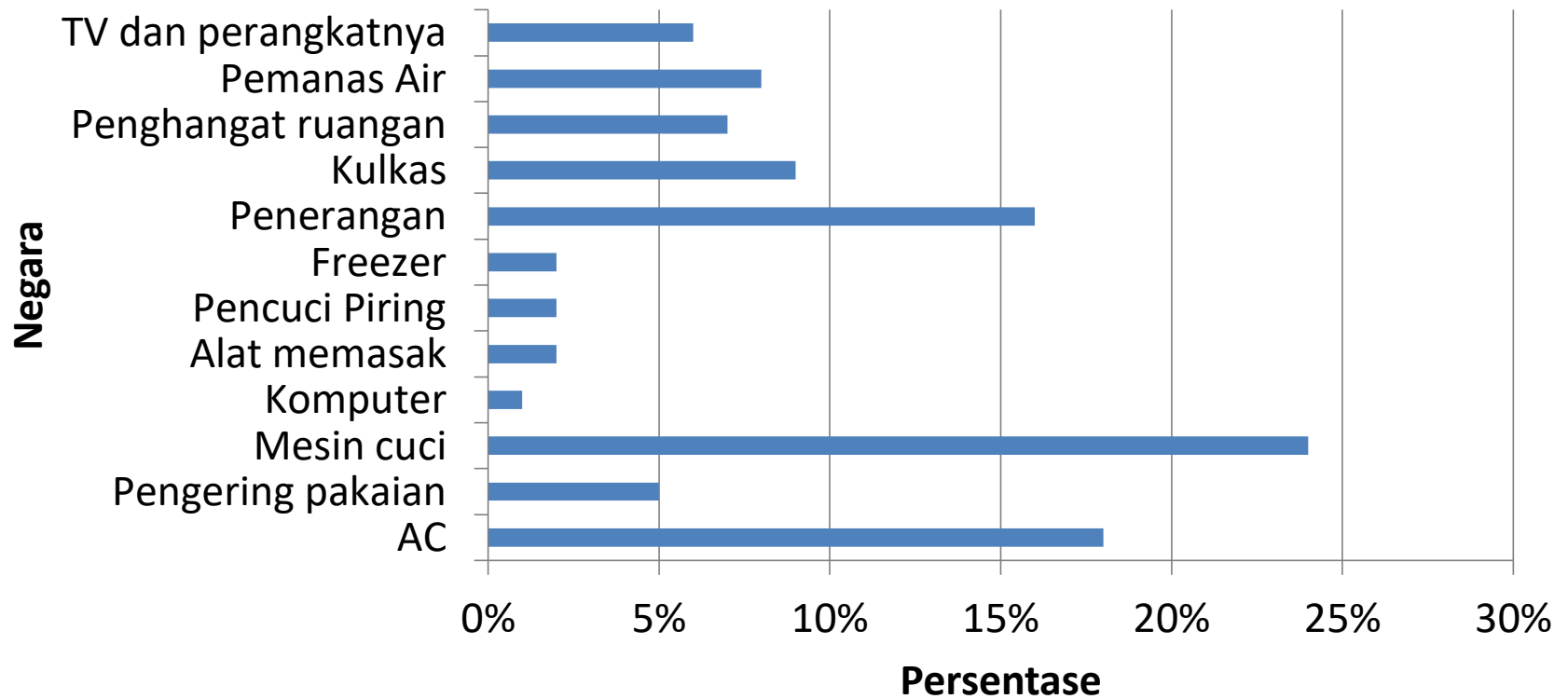
## 2.38

- a. Buatlah histogram dan percentage polygon.
- b. Buatlah cumulative percentage polygon (ogive).
- c. Pada nilai berapakah biaya listrik bulanan tersebut terkonsentrasi?

**ANSWER**

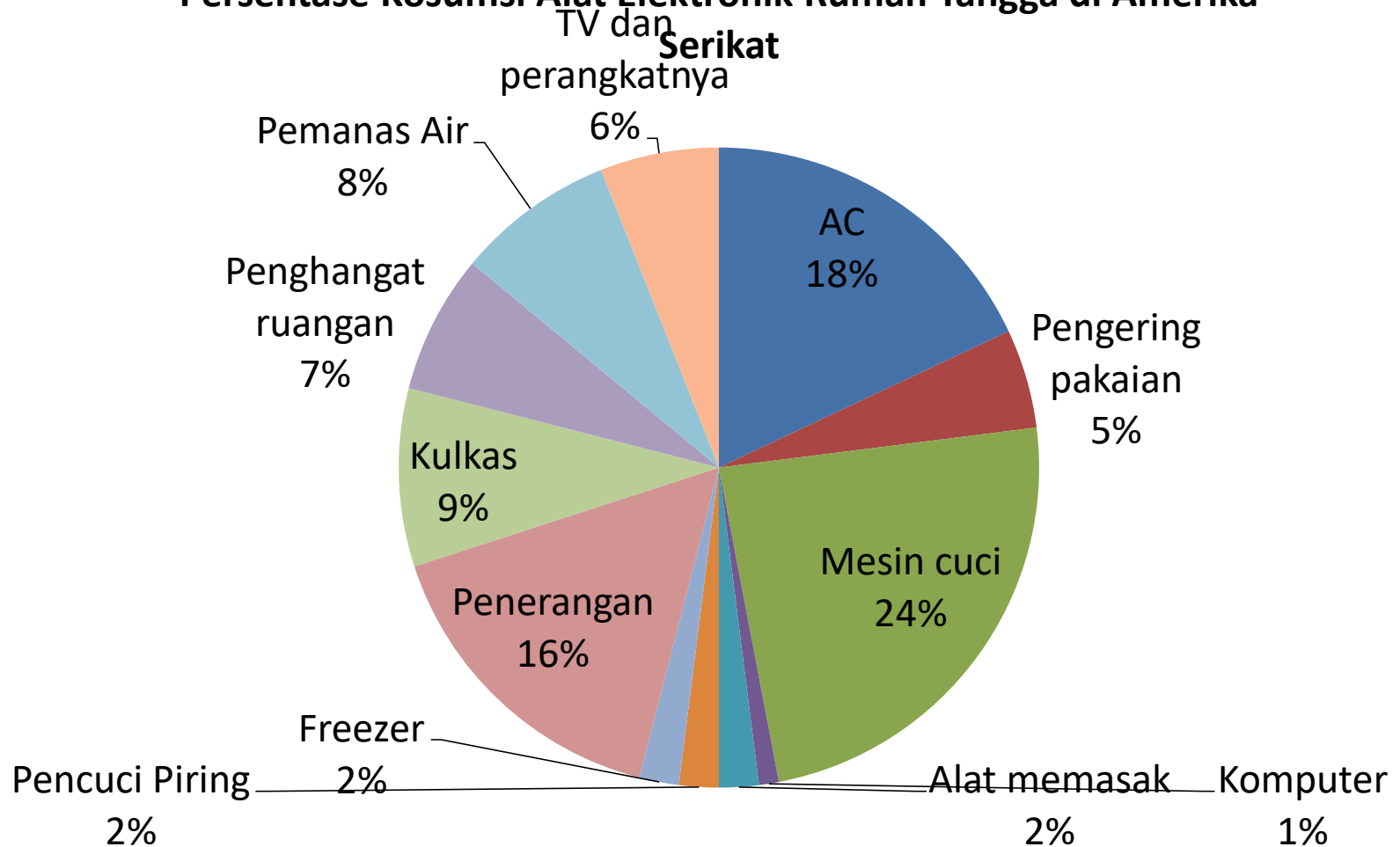
# 2.28

**Persentase Kosumsi Alat Elektronik Rumah Tangga di Amerika Serikat**



# 2.28

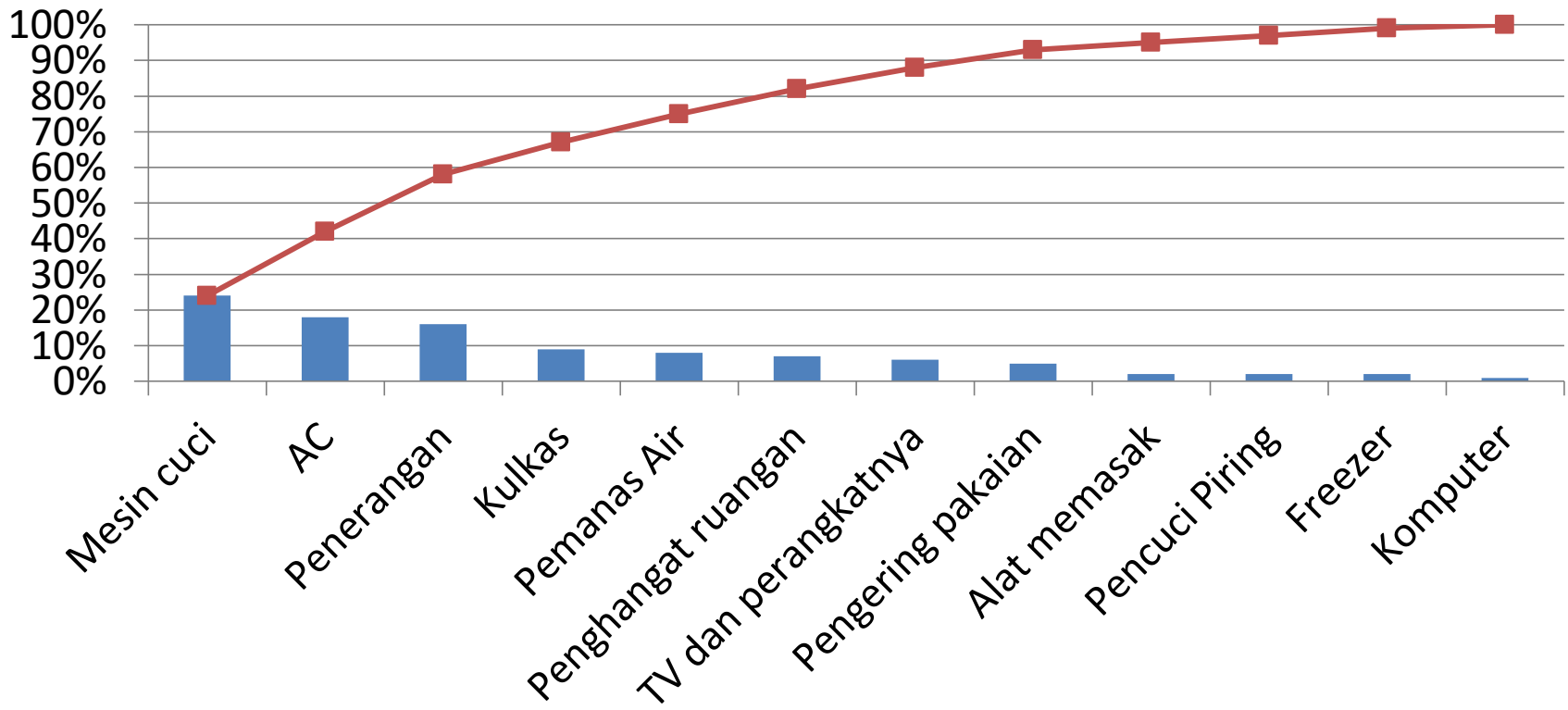
Persentase Kosumsi Alat Elektronik Rumah Tangga di Amerika





# 2.28

**Persentase Kosumsi Alat Elektronik Rumah Tangga di Amerika Serikat**



## 2.37

Data yang telah disusun:

0,55 0,57 0,57 0,68 0,72 0,77 0,86  
0,90 0,92 0,94 1,14 1,41 1,42 1,51

# 2.37

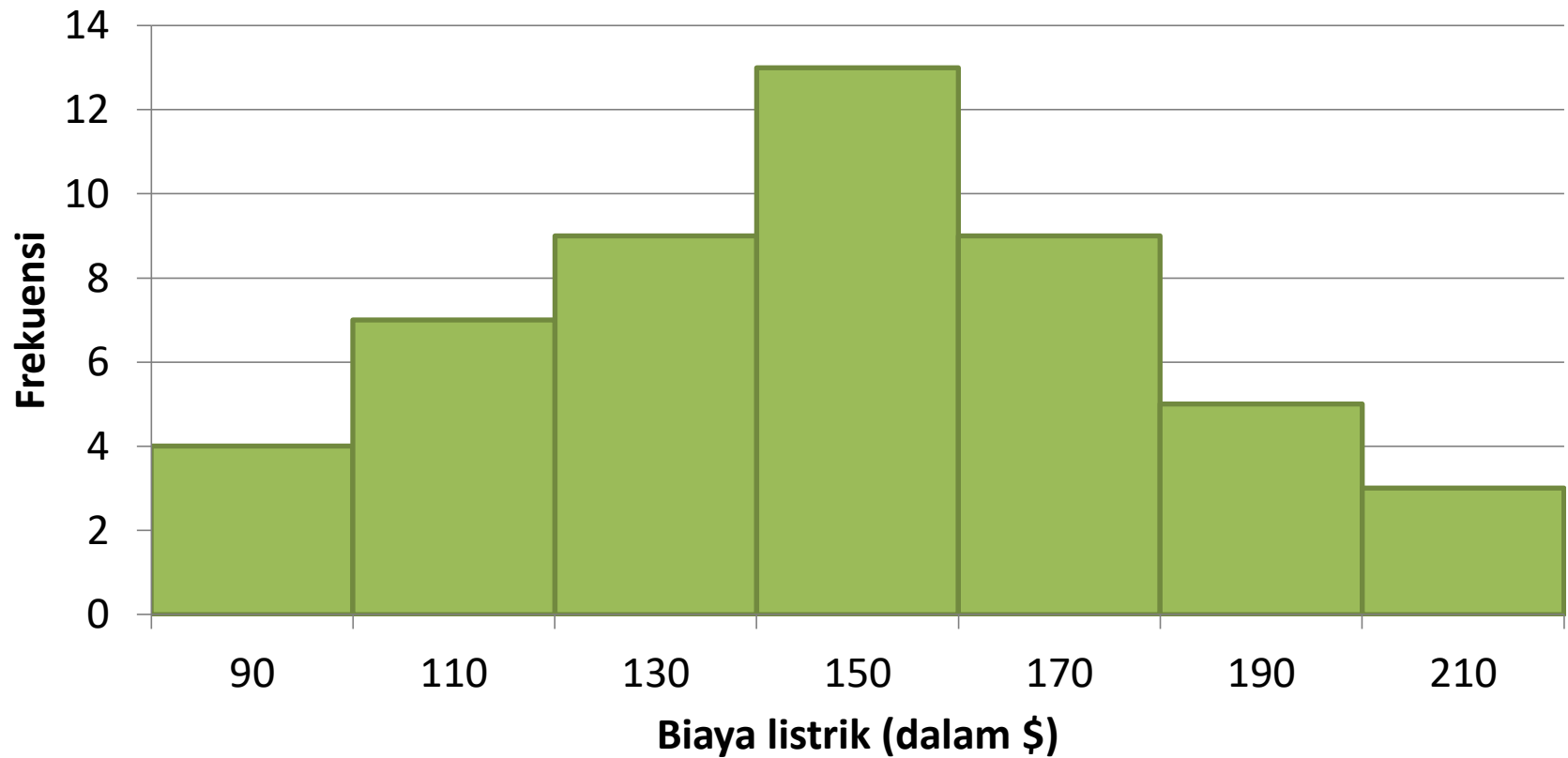
Diagram Batang-Daun:

5	5 7 7
6	8
7	2 7
8	6
9	0 2 4
1	
11	4
12	
13	
14	1 2
15	1

Catatan: 5 | 7 artinya: 0,57

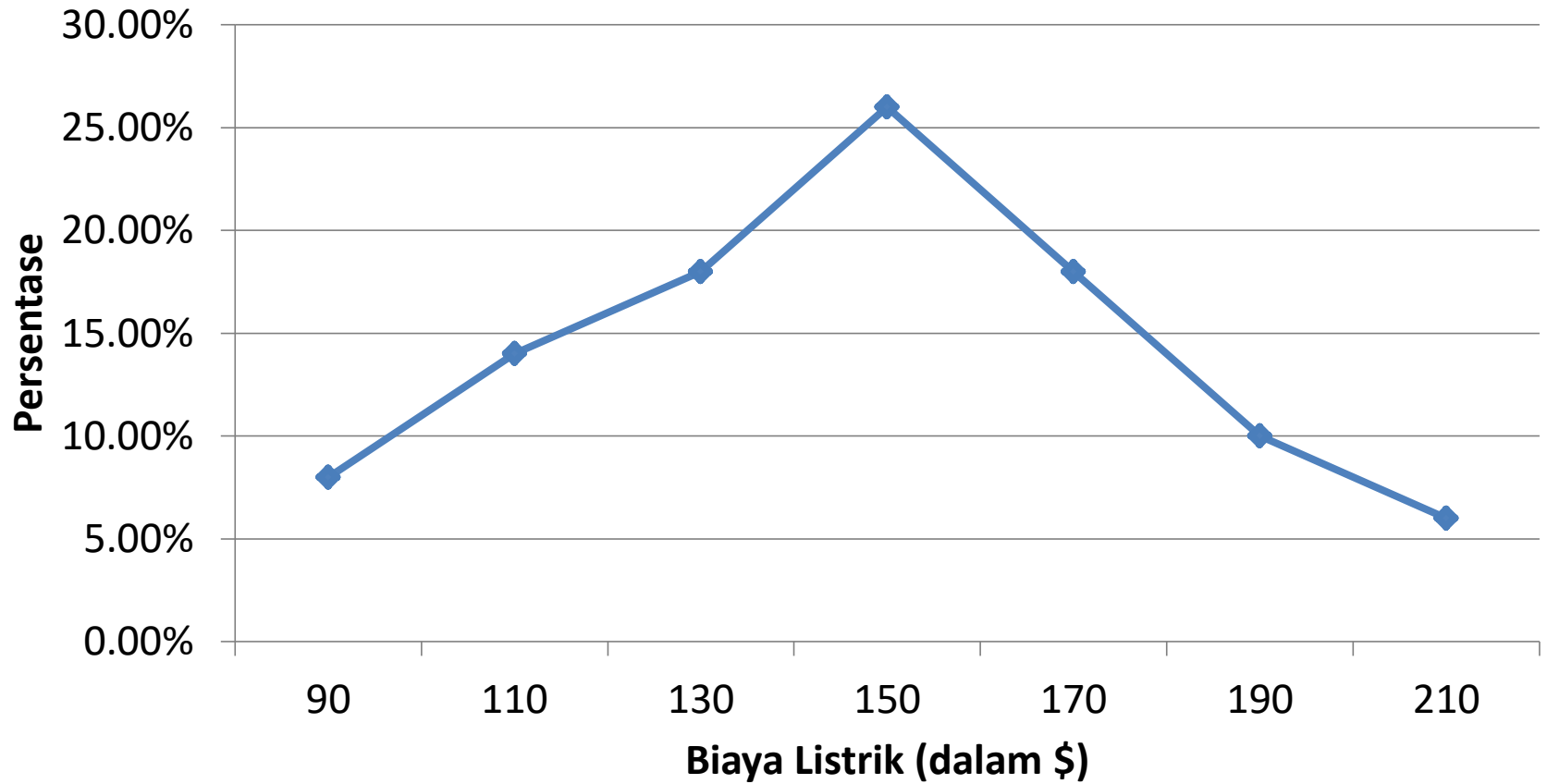
# 2.38

**Biaya Listrik pada bulan Juli 2010 untuk apartement satu kamar tidur di kota besar**



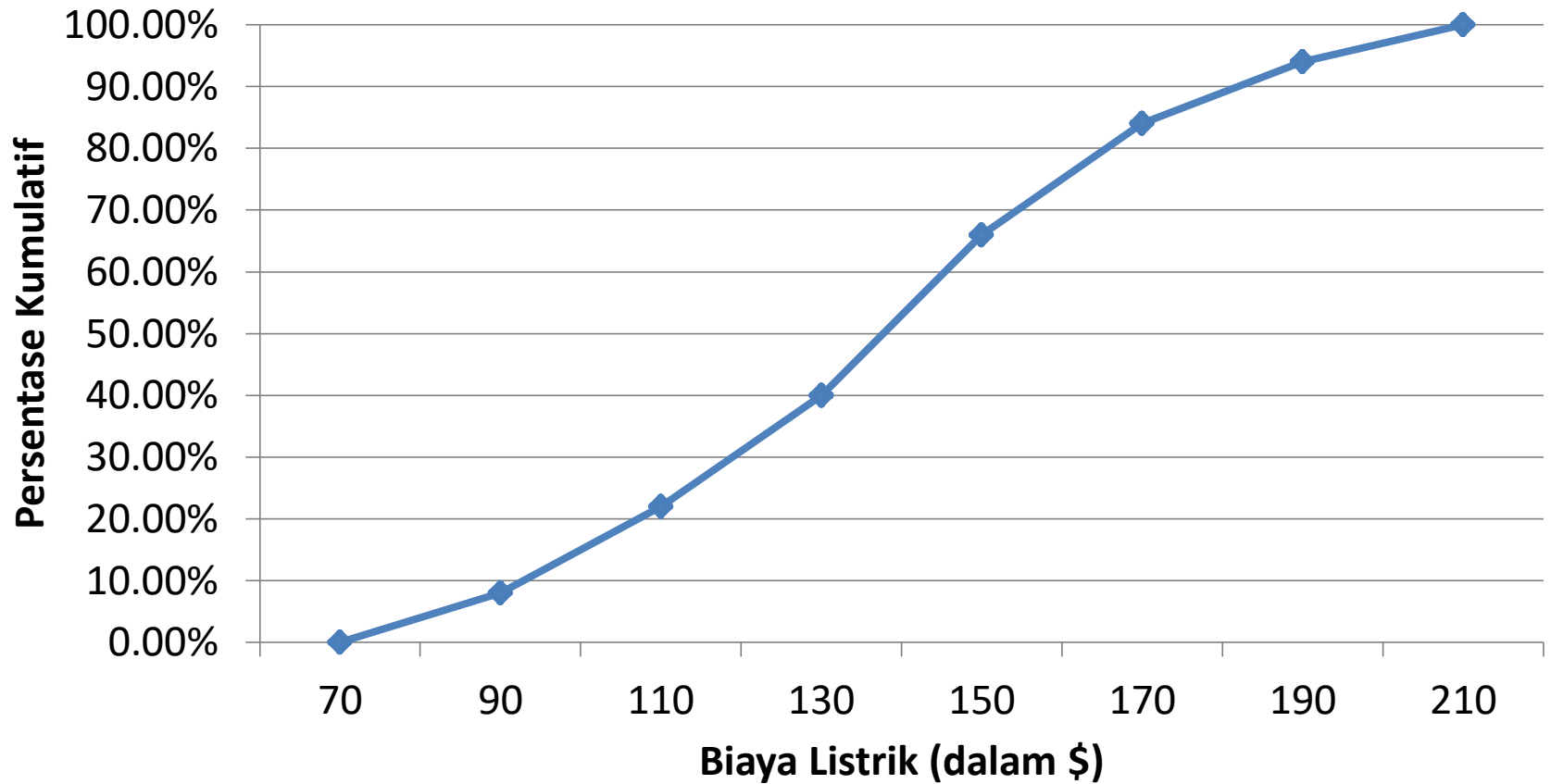
# 2.38

**Biaya Listrik pada bulan Juli 2010 untuk apartement satu kamar tidur di kota besar**



# 2.38

**Biaya Listrik pada bulan Juli 2010 untuk apartement satu kamar tidur di kota besar**



# **HOMEWORK**

# 1

- Kumpulkan data dari online shop/mall (amazon, lazada.com, etc.):
  - Satu Variabel Kategori
  - Satu Variabel Numerik

Masing-masing minimal 20 data

- Simpan data tersebut dalam tabel data mentah.
- Susun dan Visualisasikan data tersebut dalam tabel dan grafik yang sesuai.
- Perhatikan aturan pembuatan grafik yang baik (hal 50)



**THANK YOU**