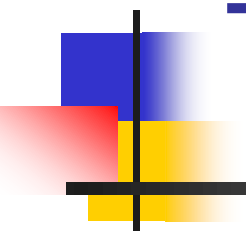
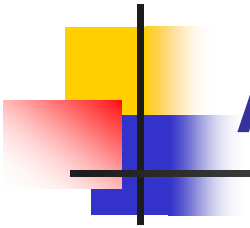
Business Statistics: Quantitative Methods and Techniques

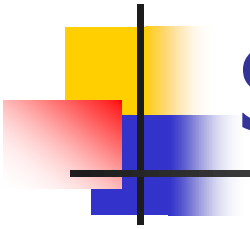


Lecture 3: Graphical Descriptive Techniques



Agenda

- Frequency tables and graphical descriptive techniques
  - measurement level of variables (types of data)
  - graphical techniques for quantitative data
  - graphical techniques for qualitative data



Some Basic Concepts

Values of the variable are the range of possible values for a variable.

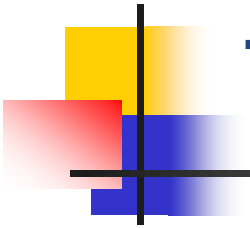
E.g. student marks (0..100)

Data are the ***observed values*** of a variable.

E.g. student marks: {67, 74, 71, 83, 93, 55, 48}

Types of Data

- An important first step for making decisions is to find the right data and prepare it.
  - Compilation of facts, figures, or other content
  - Numerical and non-numerical
  - All types and formats are generated from multiple sources
  - Often we have a large amount of data
  - Even small data can give insights
- Data that have been organized, analyzed, and processed in a meaningful and purposeful way become information.
- Use a blend of data, contextual information, experience, and intuition to derive knowledge.



Types of Data

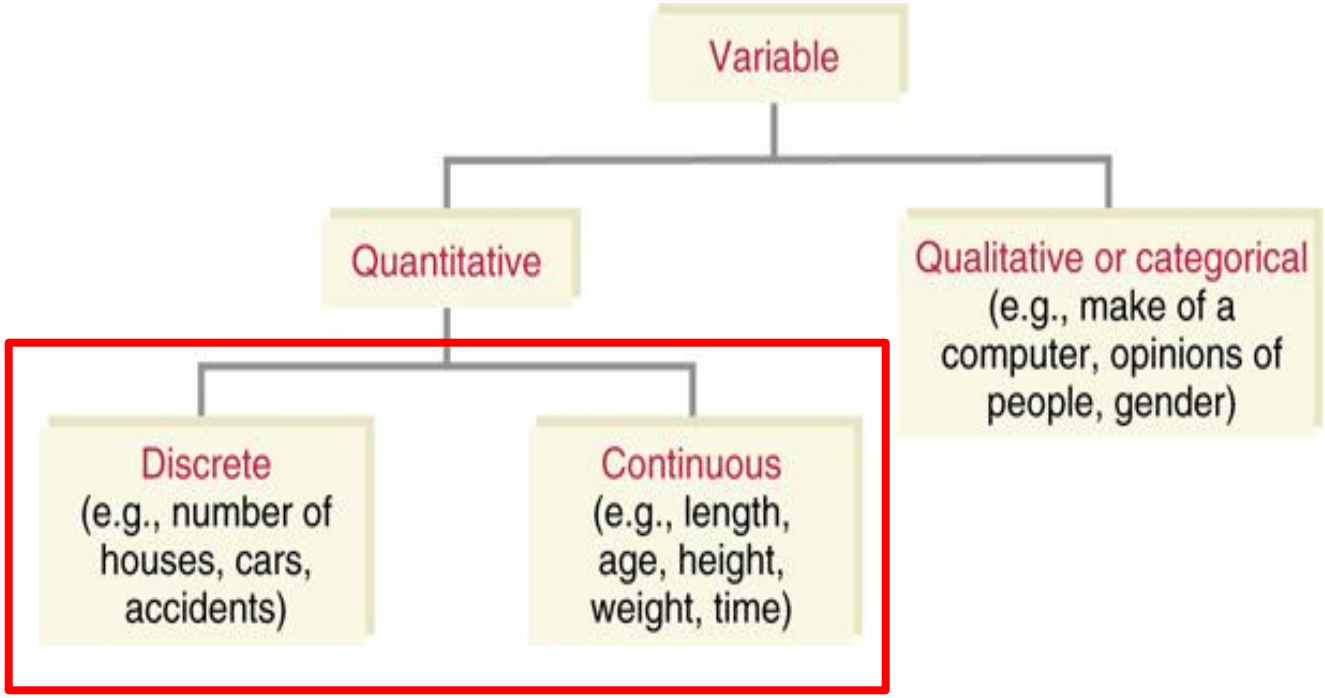
- It is not feasible to collect data that comprise a population of all elements of interest.
- A sample is a subset of the population and is used for analyses.
- Traditional statistical techniques use sample information to draw conclusions about the population.
- Cross-sectional data, Time series data

Descriptive statistics

Descriptive statistics involves arranging, summarizing, and presenting a *set of data* in such a way that useful *information* is produced.

Its methods make use of graphical techniques and numerical descriptive measures to summarize and present the data.

Types of Variables

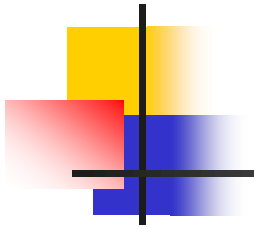


Variables and Scales of Measurement

- There are two types of variables: categorical and numerical
- Categorical
  - Also called qualitative
  - Represent categories
  - Labels or names to identify distinguishing characteristics
  - Arithmetic operations on the labels/values are not meaningful
  - Coded into numbers for data processing
    Example: marital status
- Numerical
  - Also called quantitative
  - Represent meaningful numbers
  - Arithmetic operations are meaningful
  - Discrete: assumes a countable number of values
    Example: number of children in a family
  - Continuous: assumes an uncountable number of values within an interval
    Example: investment returns

Measurement level of variables

- Choice of appropriate statistical technique depends on measurement level (or type) of variables analyzed
- Variables may either be qualitative or quantitative:
  - qualitative: outcomes are categorical
    - nominal: mutually exclusive categories, labeling (e.g. country of origin)
    - ordinal or ranked: natural ordering (e.g. preference for cola)
  - quantitative: outcomes are numerical
    - interval: equal distance (e.g. shoe size {5, 5.5, 6, 6.5, etc.})
    - ratio: absolute zero (e.g. number of vehicles owned in the last 10 years, market share)



A pain rating scale from 0 (no pain) to 10 (worst possible pain) is interval. It has a fixed measurement unit.

A pain rating scale that goes from no pain, mild pain, moderate pain, severe pain, to the worst pain possible is ordinal.

Measurement level of variables (cont.)

Information content	Type of variable	Operations that are allowed
absolute zero	ratio	=, ≠, <, >, +, –, ×, /
distance	interval	=, ≠, <, >, +, –
ordering	ordinal	=, ≠, <, >
labeling	nominal	=, ≠

qualitative

quantitative

Variables and Scales of Measurement

- Analysis techniques depend on the type of data.
- Nominal
  - Categorical
  - Least sophisticated
  - Values differ by label or name
  - Example: marital status
- Ordinal
  - Categorical
  - Reflect labels or name, but can be ranked
  - Cannot interpret the difference between the ranked values
  - Example: reviews from 1 star (poor) to 5 stars (outstanding)
- Interval
  - Numerical
  - Categorize and rank, differences are meaningful
  - Zero value is arbitrary and does not reflect absence of characteristic
  - Ratios are not meaningful
  - Example: temperature
- Ratio
  - Numerical
  - Most sophisticated
  - A true zero point, reflects absence of characteristic
  - Ratios are meaningful
  - Example: profits

Exc.1

Baseball fans are regularly asked to offer their opinions about various aspects of the sport. A survey asked the following questions. Identify the type of data.

Questions and Answers

a. **Q:** How many games do you attend annually?
A: Ratio

b. **Q:** How would you rate the quality of entertainment? (excellent, very good, good, fair, poor)
A: Ordinal

c. **Q:** Do you have season tickets?
A: Nominal

d. **Q:** How would you rate the quality of the food? (edible, barely edible, or horrible)
A: Ordinal

FIGURE.1

Assessing a Respondent's Liking of Soft Drinks with Nominal, Ordinal, Interval, and Ratio Scales

( )

Which of the soft drinks in the following list do you like? (Check ALL that apply):

___ Coke
___ Dr Pepper
___ Mountain Dew
___ Pepsi
___ 7UP
___ Sprite

( )

Rank the soft drinks according to how much you like each (most preferred drink = 1, and least preferred drink = 6):

___ Coke
___ Dr Pepper
___ Mountain Dew
___ Pepsi
___ 7UP
___ Sprite

( )

Please indicate how much you like each soft drink by checking the appropriate position on the scale:

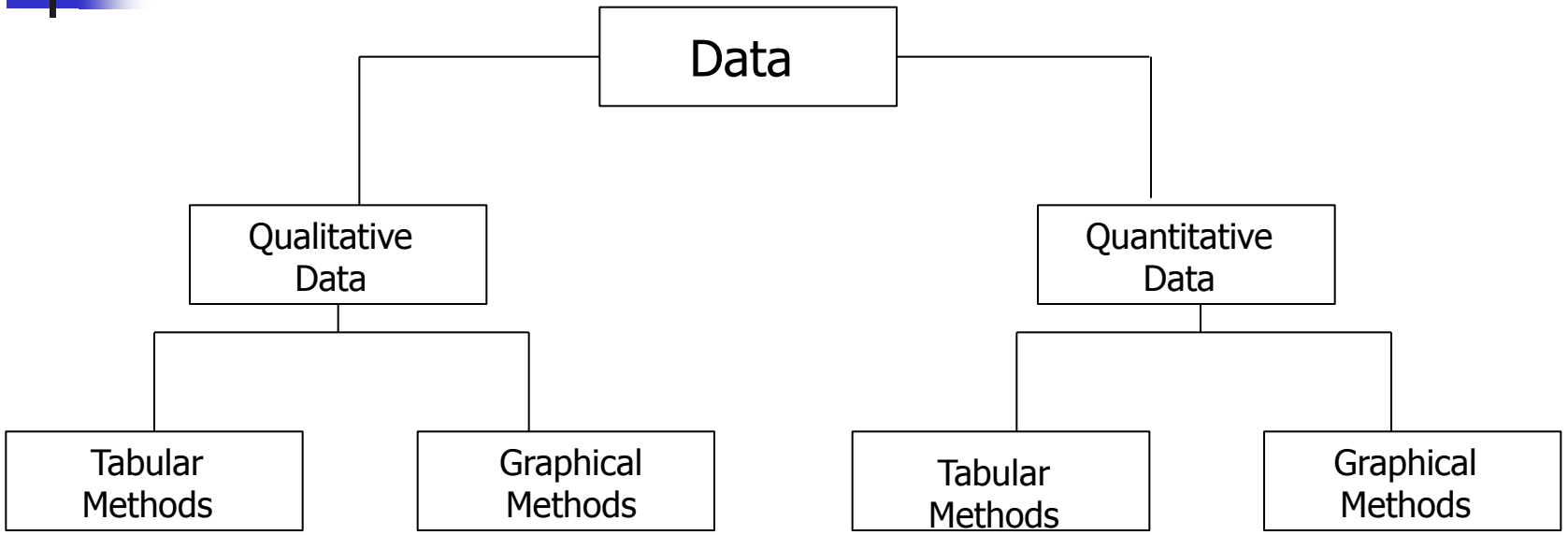
	Dislike a Lot	Dislike	like	Like a Lot
Coke	___	___	___	___
Dr Pepper	___	___	___	___
Mountain Dew	___	___	___	___
Pepsi	___	___	___	___
7UP	___	___	___	___
Sprite	___	___	___	___

( )

Please divide 100 points among these soft drinks to represent how much you like each:

___ Coke
___ Dr Pepper
___ Mountain Dew
___ Pepsi
___ 7UP
___ Sprite
100

Tabular and graphical methods for summarizing data



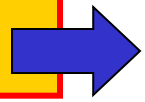
- **Frequency Distribution**
- **Bar Graph**
- **Pie Chart**
- **Frequency Distribution**
- **Histogram**
- **Stem-and-Leaf Display**

Frequency tables and histograms

Descriptive techniques for quantitative data

- Example 1: a telephone company has examined the telephone bills of new subscribers in first month after signing on
  - collect data
  - prepare a frequency distribution
  - draw a histogram
- Sample size: n = 200

See also next slides…



Frequency tables and histograms (cont.)

Collect data

Bills
42.19
38.45
29.23
89.35
118.04
110.46
0.00
72.88
83.05
.
.

Prepare a frequency distribution

How many classes to use?

# observations	# classes
Less then 50	5-7
50 - 200	7-9
200 - 500	9-10
500 - 1,000	10-11

Largest observation: 119.63

Smallest observation: 0

$$\text{Class width} = \frac{\text{Range}}{\text{\# Classes}} = \frac{119.63 - 0}{8} = 14.95 \rightarrow 15$$

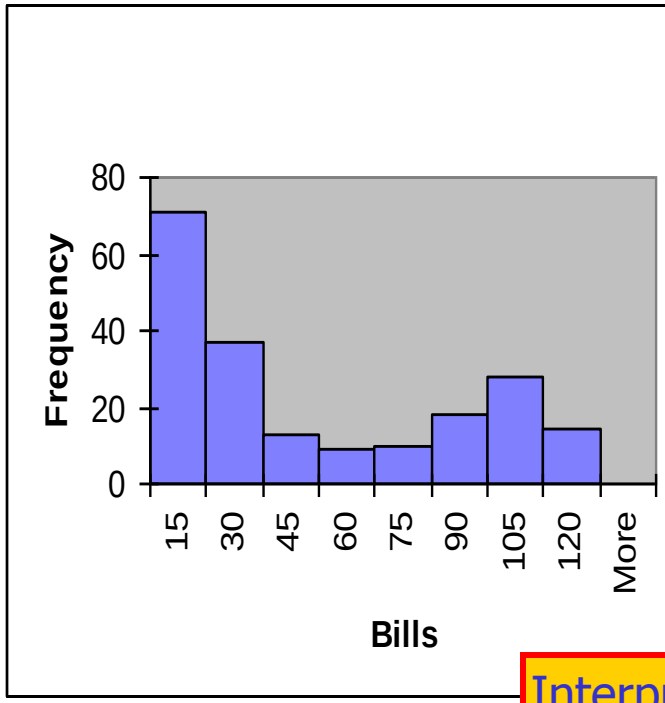
Frequency tables and histograms (cont.)

Excel output:

Frequency distribution

Bin	*Frequency*
15	71
30	37
45	13
60	9
75	10
90	18
105	28
120	14
More	0

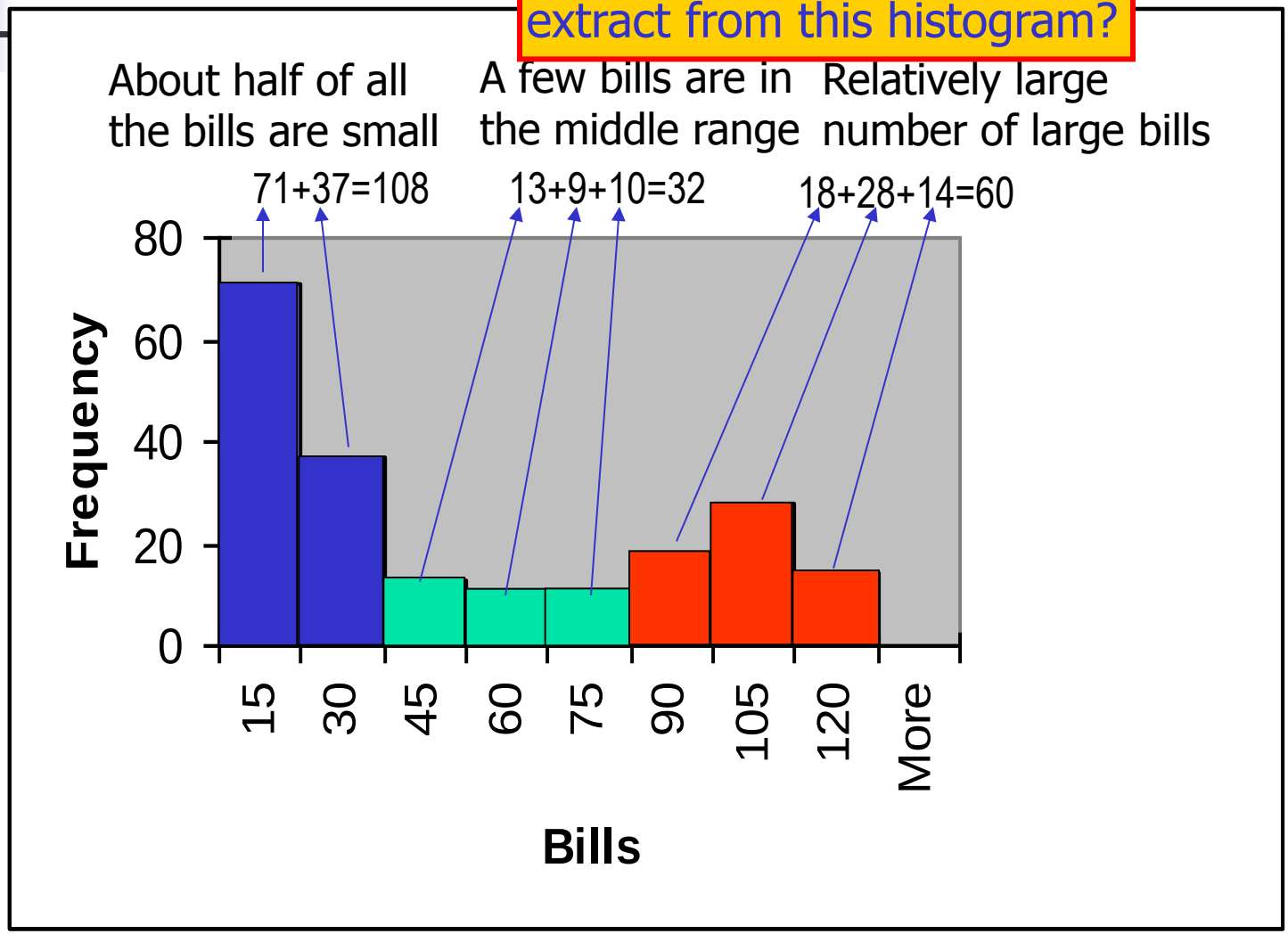
Histogram



Interpretation...

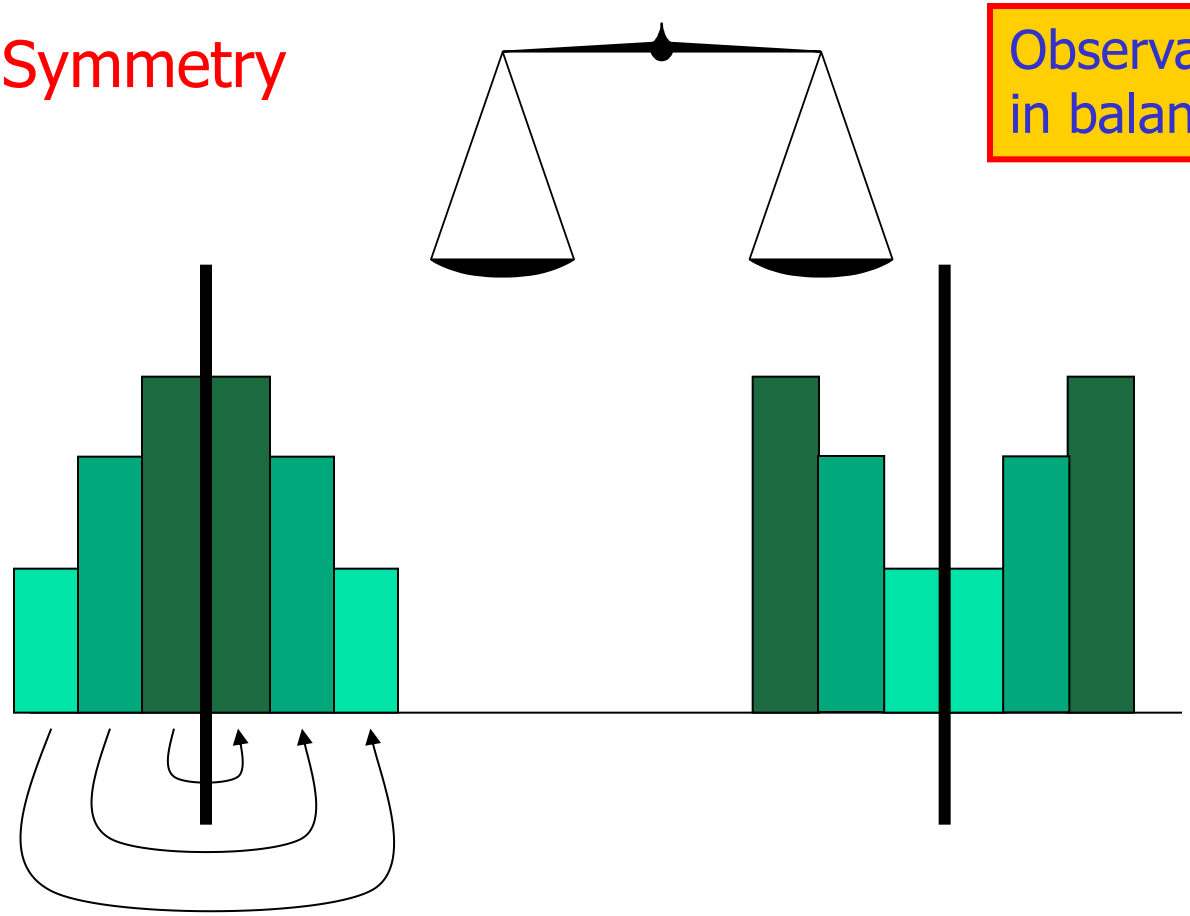
Interpretation of histograms

What information can we extract from this histogram?



Shapes of histograms

- Symmetry



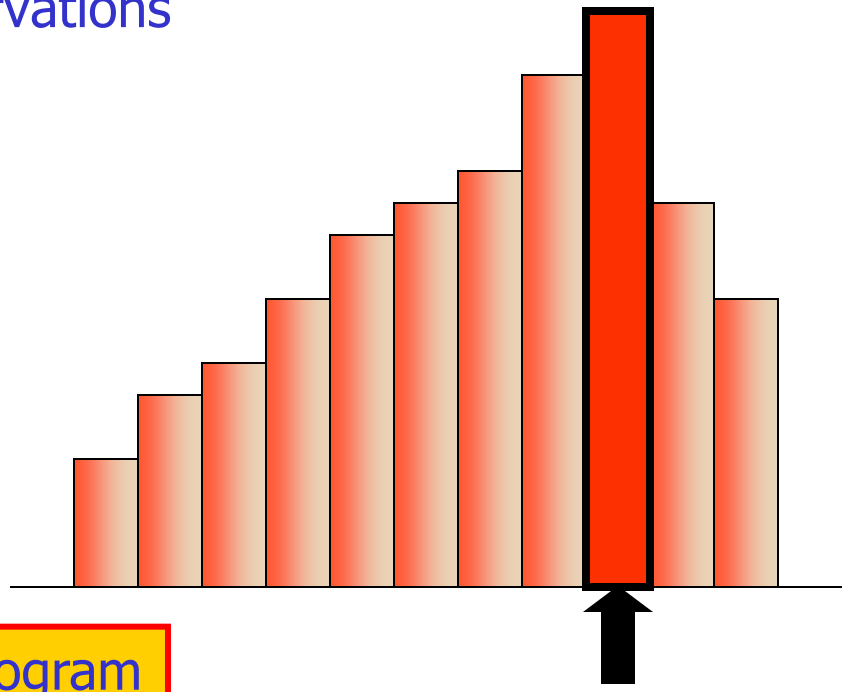
Observations are in balance

Shapes of histograms (cont.)

- Skewness

Shapes of histograms (cont.)

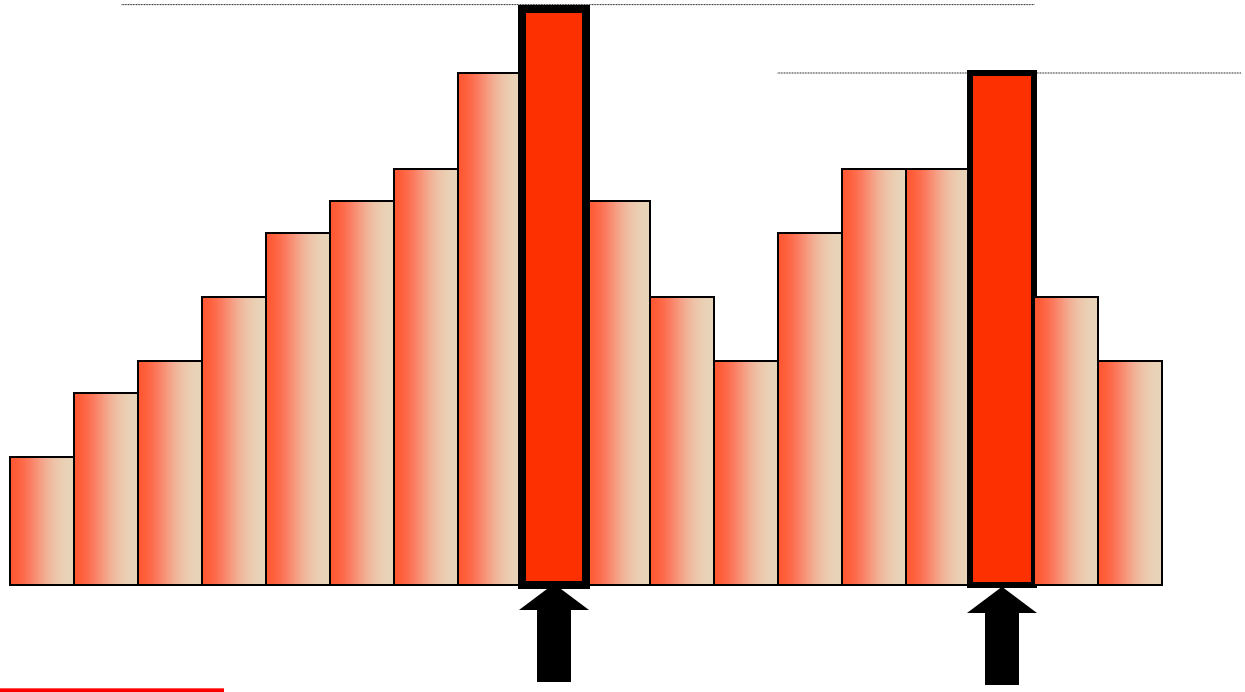
- Number of modal classes
  - a modal class is the one with the largest number of observations



A unimodal histogram

Shapes of histograms (cont.)

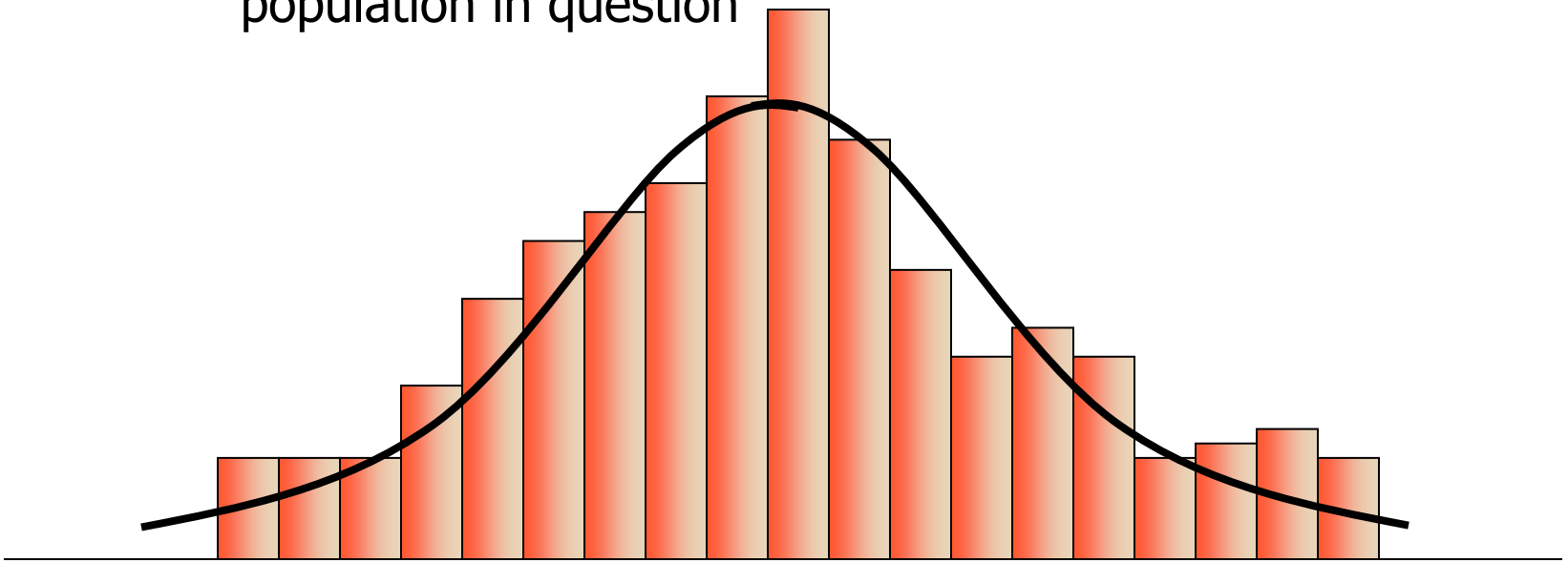
- Number of modal classes



A bimodal histogram

Shapes of histograms (cont.)

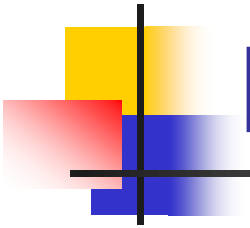
- Bell shaped histogram
  - many statistical techniques require that the population be bell shaped (or normally distributed)
  - drawing the histogram helps verify the shape of the population in question



Pie charts, bar charts

Descriptive techniques for qualitative data

- Qualitative data like country of origin, color of hair, etc., are represented by pie charts or bar charts (and not by histograms)
- When the raw data can be naturally categorized in a meaningful manner, we can display frequencies by
  - Bar charts – emphasize frequency of occurrence of the different categories
  - Pie charts – emphasize the proportion of occurrences of each category

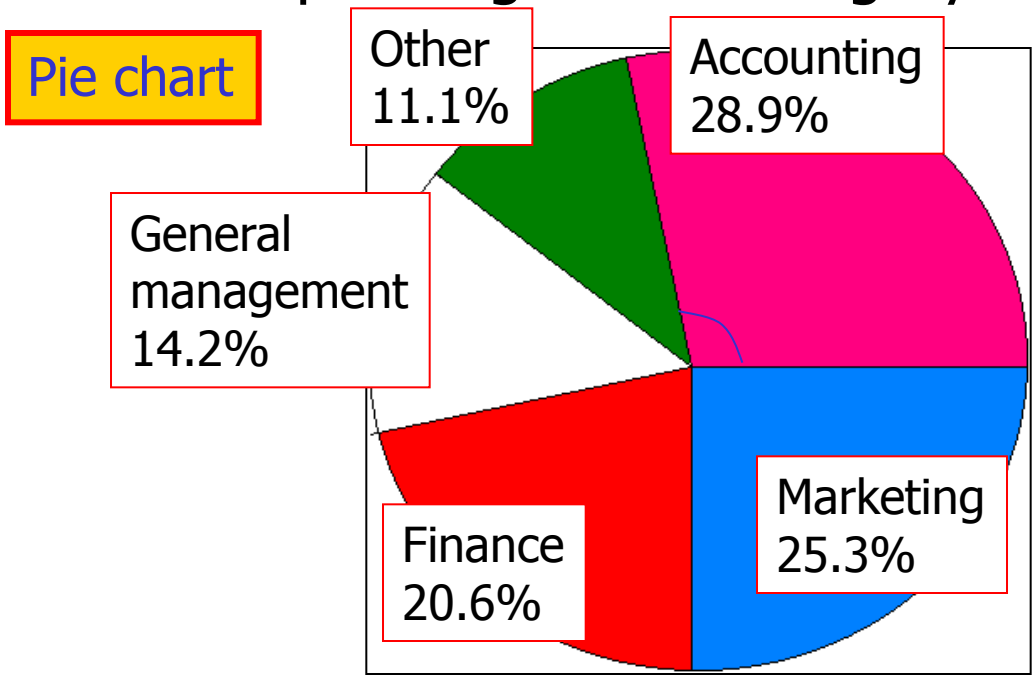


Pie charts

- Pie chart is very popular tool to represent the proportions of appearance for nominal data
- Example:
  - The student placement office at a university wanted to determine the general areas of employment last year
  - Data were collected, and the count of the occurrences was recorded for each area
  - These counts were converted to proportions and the results were presented in a pie chart

Pie charts

- Pie chart is a circle, subdivided into a number of slices that represent the various categories.
- Size of each slice is proportional to the percentage corresponding to the category it represents.

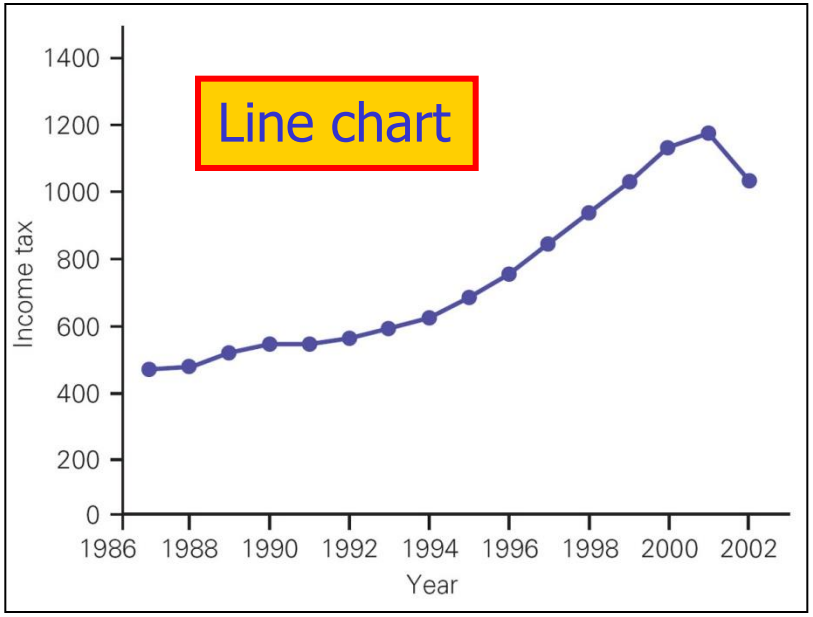


Describing time-series data

- Data can be classified according to the time they are measured:
  - Cross-sectional data are all collected at the same time
  - Time-series data are collected at successive points in time
- Time-series data are often depicted on a line chart (which is a plot of the variable over time)

Line charts

- Example :
  - The total amounts of income tax paid by individuals in 1987 to 2002 are given
  - Draw a graph of these data and describe the information produced
- Conclusion:
  - For the first five years – total tax was relatively flat
  - From 1993 there was a rapid increase in tax revenues
  - Finally, there was a downturn in 2002



Panel data

Cross-sectional data

Time-series data

Panel data

i = product

t = time