# Chapter 1 & Section 2.1

## Exploring Data

## Introduction

- **Statistics:**
  - the science of data. We begin our study of statistics by mastering the art of examining data. Any set of data contains information about some group of individuals. The information is organized in variables.
- **Individuals:**
  - The objects described by a set of data. Individuals may be people, but they may also be other things.
- **Variable:**
  - Any characteristic of an individual.
  - Can take different values for different individuals.

2

## Variable Types

- **Categorical variable:**
  - places an individual into one of several groups of categories.
- **Quantitative variable:**
  - takes numerical values for which arithmetic operations such as adding and averaging make sense.
- **Distribution:**
  - pattern of variation of a variable
  - tells what values the variable takes and how often it takes these values.

3

1.1 **FUEL-EFFICIENT CARS** Here is a small part of a data set that describes the fuel economy (in miles per gallon) of 1998 model motor vehicles:

| Make and Model | Vehicle type | Transmission type | Number of cylinders | City MPG | Highway MPG |
|---|---|---|---|---|---|
| ⋮ | | | | | |
| BMW 318I | Subcompact | Automatic | 4 | 22 | 31 |
| BMW 318I | Subcompact | Manual | 4 | 23 | 32 |
| Buick Century | Midsize | Automatic | 6 | 20 | 29 |
| Chevrolet Blazer | Four-wheel drive | Automatic | 6 | 16 | 20 |
| ⋮ | | | | | |

(a) What are the individuals in this data set?

(b) For each individual, what variables are given? Which of these variables are categorical and which are quantitative?

4

- A. The individuals are the BMW 318I, the Buick Century, and the Chevrolet Blazer.

- B. The variables given are
  - Vehicle type (categorical)
  - Transmission type (categorical)
  - Number of cylinders (quantitative)
  - City MPG (quantitative)
  - Highway MPG (quantitative)

5

## 1.1 & 1.2: Displaying Distributions with graphs

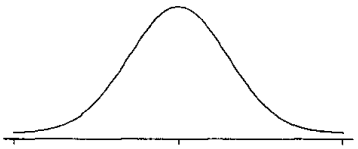
- **Graphs used to display data:**
  - bar graphs, pie charts, dot plots, stem plots, histograms, and time plots
- **Purpose of a graph:**
  - Helps to understand the data.
  - Allows overall patterns and striking deviations from that pattern to be seen.
- **Describing the overall pattern:**
  - Three biggest descriptors:
    - shape, center and spread.
  - Next look for outliers and clusters.

6

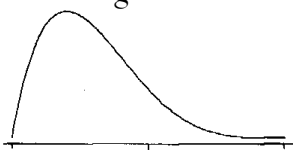
## Shape

- Concentrate on main features.
  - Major peaks, outliers (not just the smallest and largest observations), rough symmetry or clear skewness.
- Types of Shapes:

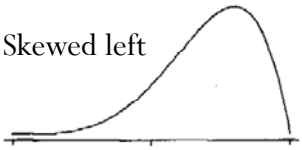
Symmetric



Skewed right



Skewed left



7

## How to make a bar graph.

**1.5 FEMALE DOCTORATES** Here are data on the percent of females among people earning doctorates in 1994 in several fields of study:[3]

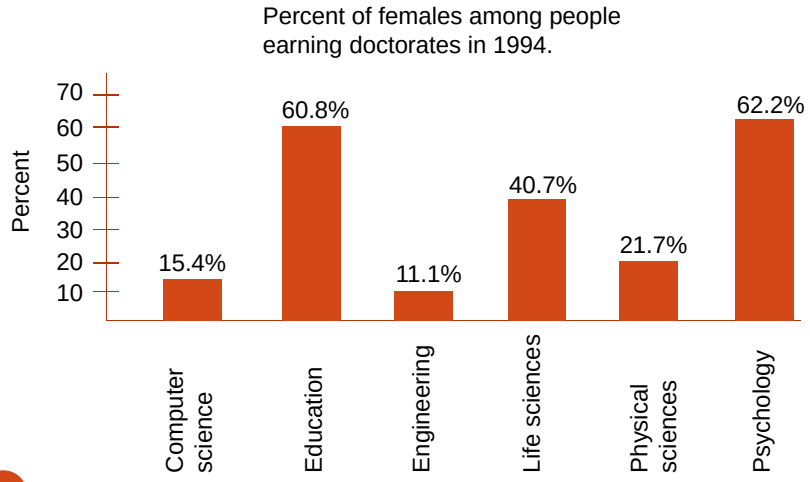
| | | | |
|---|---|---|---|
| Computer science | 15.4% | Life sciences | 40.7% |
| Education | 60.8% | Physical sciences | 21.7% |
| Engineering | 11.1% | Psychology | 62.2% |

(a) Present these data in a well-labeled bar graph.

(b) Would it also be correct to use a pie chart to display these data? If so, construct the pie chart. If not, explain why not.

8

# 1.5 How to make a bar graph.




9

**1.5 FEMALE DOCTORATES** Here are data on the percent of females among people earning doctorates in 1994 in several fields of study:[3]

| | | | |
|---|---|---|---|
| Computer science | 15.4% | Life sciences | 40.7% |
| Education | 60.8% | Physical sciences | 21.7% |
| Engineering | 11.1% | Psychology | 62.2% |

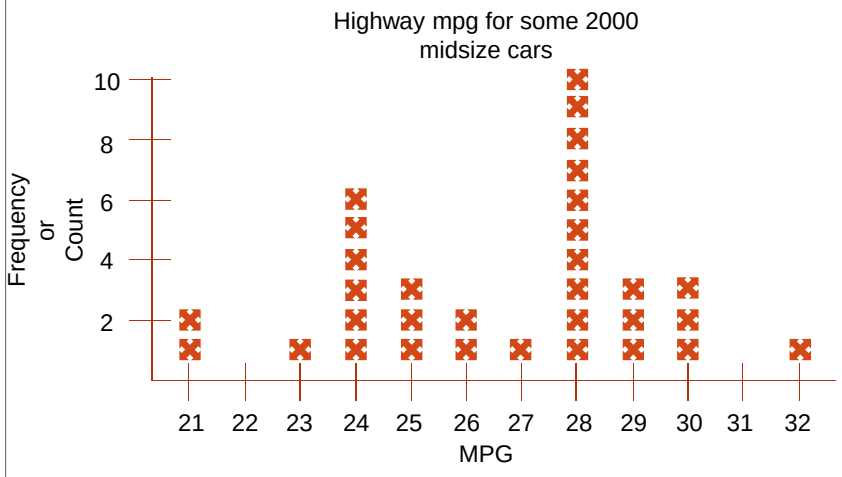
(a) Present these data in a well-labeled bar graph.

(b) Would it also be correct to use a pie chart to display these data? If so, construct the pie chart. If not, explain why not.

No, a pie chart is used to display one variable with all of its categories totaling 100%

10

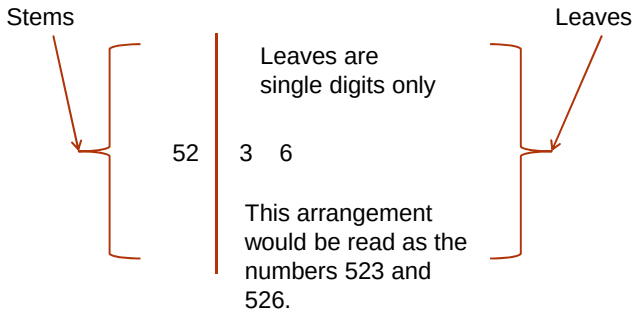
# How to make a dotplot




11

# How to make and read a stemplot

- A stemplot is similar to a dotplot but there are some format differences. Instead of dots actual numbers are used. Instead of a horizontal axis, a vertical one is used.

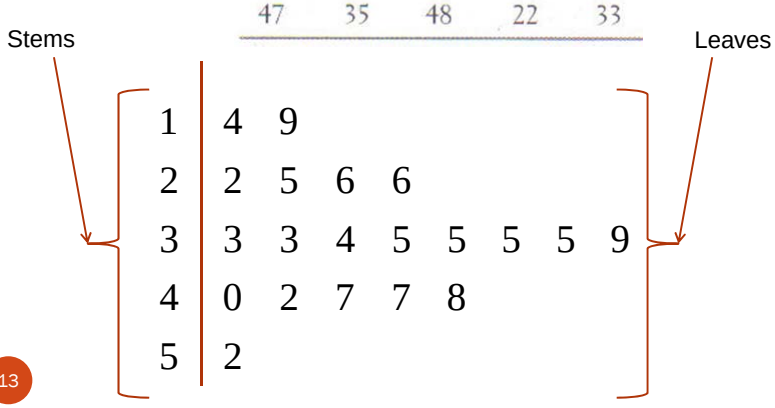



12

## How to make and read a stemplot

- With the following data, make a stemplot.

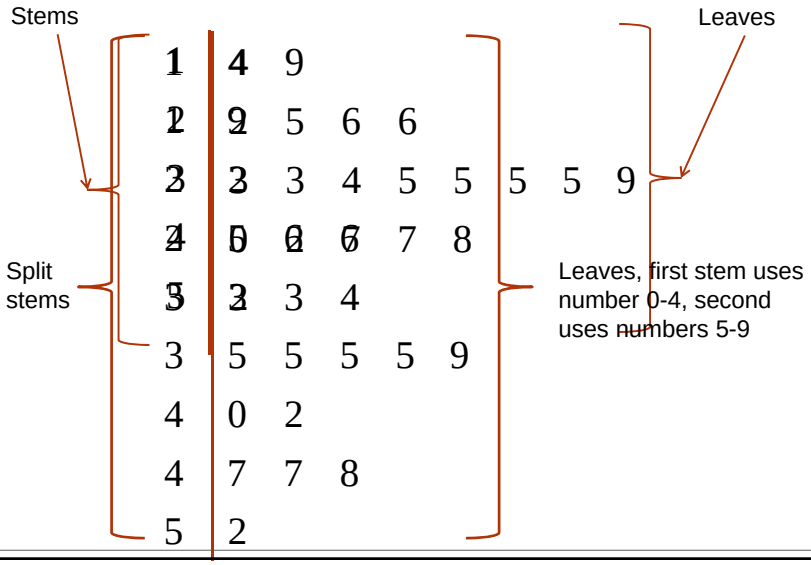
| | | | | |
|---|---|---|---|---|
| 40 | 26 | 39 | 14 | 42 |
| 47 | 19 | 26 | 35 | 34 |
| 52 | 25 | 35 | 35 | 33 |
| 47 | 35 | 48 | 22 | 33 |




13

## How to make and read a stemplot

- Lets use the same stemplot but now **split** the stems




14

## Section 1.1 & 1.2

- Homework: #'s 2, 4, 6, 9, 38a, 48a&b
- Any questions on pg. 1-4 in additional notes packet

15

## How to construct a histogram

- The most common graph of the distribution of one quantitative variable is a histogram.
- To make a histogram:
  1. Divide the range into equal widths. Then count the number of observations that fall in each group.
  2. Label and scale your axes and title your graph.
  3. Draw bars that represent each count, no space between bars.

16

**1.14 CEO SALARIES** In 1993, *Forbes* magazine reported the age and salary of the chief executive officer (CEO) of each of the top 59 small businesses.[8] Here are the salary data, rounded to the nearest thousand dollars:

| | | | | | | |
|---|---|---|---|---|---|---|
| 145 | 621 | 262 | 208 | 362 | 424 | 339 |
| 750 | 368 | 659 | 234 | 396 | 300 | 343 |
| 862 | 204 | 206 | 250 | 21 | 298 | 350 |
| 149 | 350 | 242 | 198 | 213 | 296 | 317 |
| 250 | 396 | 572 | | | | |

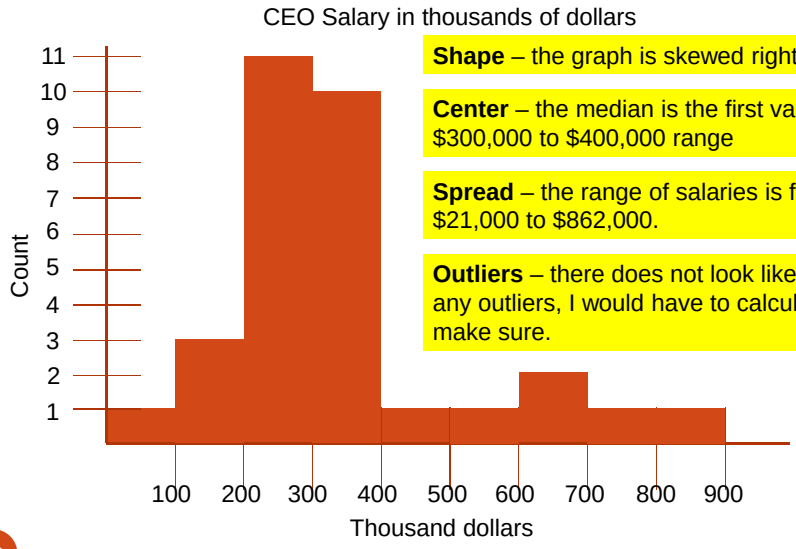
Construct a histogram for these data. Describe the shape, center, and spread of the distribution of CEO salaries. Are there any apparent outliers?

## Divide range into equal widths and count

| Scale | Counts |
|---|---|
| $0 \le$ CEO Salary $< 100$ | 1 |
| $100 \le$ CEO Salary $< 200$ | 3 |
| $200 \le$ CEO Salary $< 300$ | 11 |
| $300 \le$ CEO Salary $< 400$ | 10 |
| $400 \le$ CEO Salary $< 500$ | 1 |
| $500 \le$ CEO Salary $< 600$ | 1 |
| $600 \le$ CEO Salary $< 700$ | 2 |
| $700 \le$ CEO Salary $< 800$ | 1 |
| $800 \le$ CEO Salary $< 900$ | 1 |

18

# Draw and label axis, then make bars




**Shape** – the graph is skewed right

**Center** – the median is the first value in the $300,000 to $400,000 range

**Spread** – the range of salaries is from $21,000 to $862,000.

**Outliers** – there does not look like there are any outliers, I would have to calculate to make sure.

19

# New terms used when graphing data.

- Relative frequency:
  - Category count divided by the total count
  - Gives a percentage
- Cumulative frequency:
  - Sum of category counts up to an including the current category
- Ogives (pronounced O-Jive)
  - Cumulative frequencies divided by the total count
  - Relative cumulative frequency graph
- Percentile:
  - The $p^{th}$ percentile of a distribution is the value such that p percent of the observations fall at or below it.

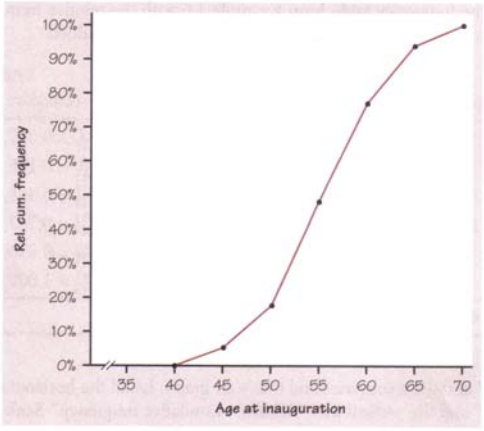
20

## Lets look at a table to see what an ogive would refer to.

| Class | Frequency | Relative frequency | Cumulative frequency | Relative cumulative frequency |
|---|---|---|---|---|
| 40–44 | 2 | $\frac{2}{43} = 0.047$, or 4.7% | 2 | $\frac{2}{43} = 0.047$, or 4.7% |
| 45–49 | 6 | $\frac{6}{43} = 0.140$, or 14.0% | 8 | $\frac{8}{43} = 0.186$, or 18.6% |
| 50–54 | 13 | $\frac{13}{43} = 0.302$, or 30.2% | 21 | $\frac{21}{43} = 0.488$, or 48.8% |
| 55–59 | 12 | $\frac{12}{43} = 0.279$, or 27.9% | 33 | $\frac{33}{43} = 0.767$, or 76.7% |
| 60–64 | 7 | $\frac{7}{43} = 0.163$, or 16.3% | 40 | $\frac{40}{43} = 0.930$, or 93.0% |
| 65–69 | 3 | $\frac{3}{43} = 0.070$, or 7.0% | 43 | $\frac{43}{43} = 1.000$, or 100% |
| TOTAL | 43 | | | |

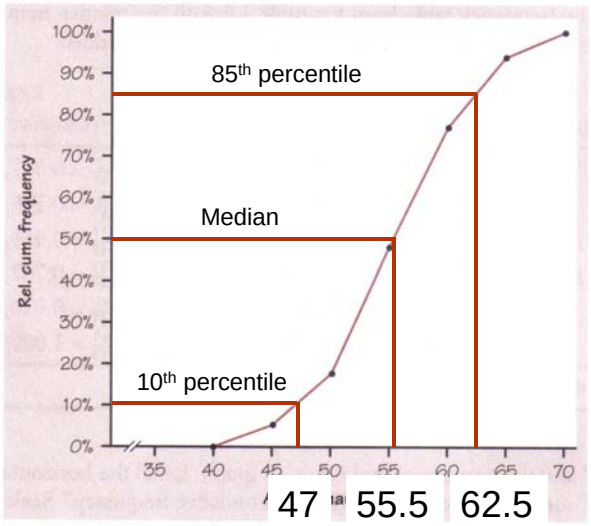
21

## The graph of an ogive for this data would look like this.




22

| Relative cumulative frequency |
|---|
| $\frac{2}{43} = 0.047$, or 4.7% |
| $\frac{8}{43} = 0.186$, or 18.6% |
| $\frac{21}{43} = 0.488$, or 48.8% |
| $\frac{33}{43} = 0.767$, or 76.7% |
| $\frac{40}{43} = 0.930$, or 93.0% |
| $\frac{43}{43} = 1.000$, or 100% |




Find the age of the 10th percentile, the median, and the 85th percentile?

47 55.5 62.5

23

## Last graph of this section

- **Time plots :**
  - Graph of each observation against the time at which it was measured.
  - Time is always on the x-axis.
  - Use time plots to analyze what is occurring over time.

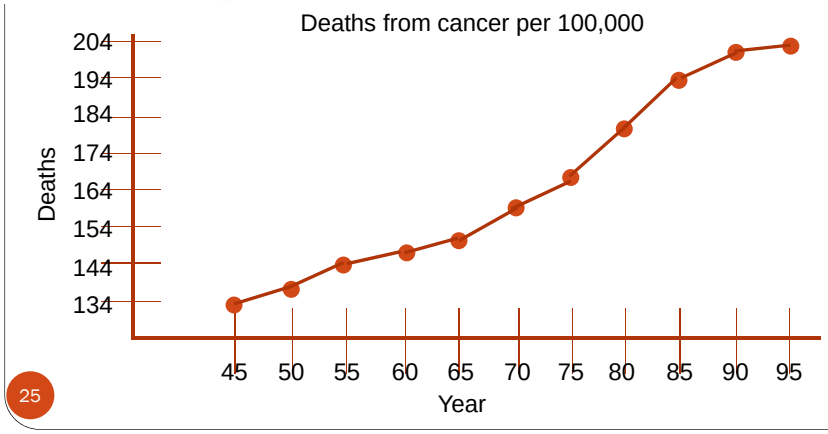
24

**1.21 CANCER DEATHS** Here are data on the rate of deaths from cancer (deaths per 100,000 people) in the United States over the 50-year period from 1945 to 1995:

| Year: | 1945 | 1950 | 1955 | 1960 | 1965 | 1970 | 1975 | 1980 | 1985 | 1990 | 1995 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Deaths: | 134.0 | 139.8 | 146.5 | 149.2 | 153.5 | 162.8 | 169.7 | 183.9 | 193.3 | 203.2 | 204.7 |

(a) Construct a time plot for these data. Describe what you see in a few sentences.

(b) Do these data suggest that we have made no progress in treating cancer? Explain.




25

## Section 1.2 & 2.1 Day 2

- Homework: #'s Section 1-2: 52, 56 (use scale starting at 7 with width of .5, make an ogive and use it to estimate the value of the center and also the 90th percentile) 58, Section 2-1: 9, R1.1 & 6, R2.2
- Any questions on pg. 5-8 in additional notes packet

26

## Section 1.3: Describing Quantitative Data with Numbers.

- **Center**:
  - Mean
  - Median
  - Mode – (only a measure of center for categorical data)
- **Spread**:
  - Range
  - Interquartile Range (IQR)
  - Variance
  - Standard Deviation

27

## Measuring center:

- **Mean**:
  - Most common measure of center.
  - Is the arithmetic average.
  - Formula:
    - $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$ or $\bar{x} = \frac{1}{n}\sum x_i$
  - Not **resistant** to the influence of extreme observations.

28

## Measuring center:

- **Median**
  - The midpoint of a distribution
  - The number such that half the observations are smaller and the other half are larger.
  - If the number of observations n is odd, the median is the center of the ordered list.
  - If the number of observations n is even, the median M is the mean of the two center observations in the ordered list.
  - Is **resistant** to the influence of extreme observations.

29

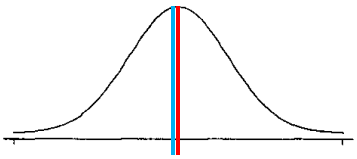
## Quick summary of measures of center.

| Measure | Definition | Example using 1,2,3,3,4,5,5,9 |
|---|---|---|
| **Mean** | $\frac{\text{sum of the data values}}{\text{number of data values}}$ | $\frac{1+2+3+3+4+5+5+9}{8}=4$ |
| **Median** | Middle value for an **odd** # of data values<br>Mean of the 2 middle values for an **even** # of data values | For 1,2,3,3,4,5,5,9, the middle values are 3 and 4. The median is: $\frac{3+4}{2}=3.5$ |
| **Mode** | The most frequently occurring value (Categorical data only) | Two modes: 3 and 5<br>Set is ***bimodal***. |

## Comparing the Mean and Median.

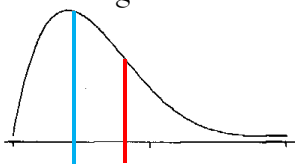
- The location of the mean and median for a distribution are effected by the distribution's shape.

Symmetric



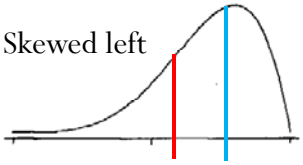
Median and Mean

Skewed right



Median and Mean

Skewed left



Mean and Median

31

**1.31** Joey's first 14 quiz grades in a marking period were

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | 84 | 91 | 75 | 78 | 80 | 74 | 87 | 76 | 96 | 82 | 90 | 98 | 93 |

(a) Use the formula to calculate the mean. Check using "one-variable statistics" on your calculator.

(b) Suppose Joey has an unexcused absence for the fifteenth quiz and he receives a score of zero. Determine his final quiz average. What property of the mean does this situation illustrate? Write a sentence about the effect of the zero on Joey's quiz average that mentions this property.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad \bar{x} = \frac{1190}{14}$$

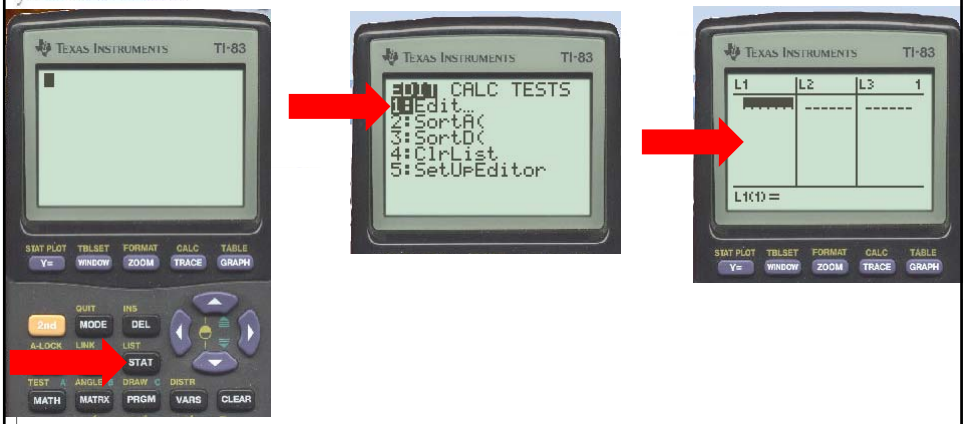
$$\bar{x} = \frac{86 + 84 + \ldots + 93}{14} \qquad \bar{x} = 85$$

32

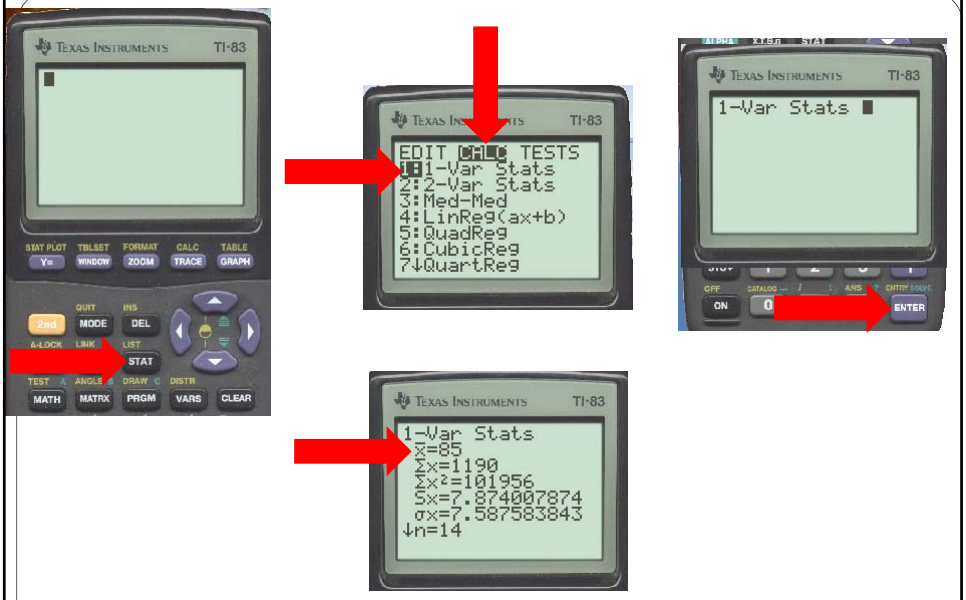
**1.31** Joey's first 14 quiz grades in a marking period were

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | 84 | 91 | 75 | 78 | 80 | 74 | 87 | 76 | 96 | 82 | 90 | 98 | 93 |

(a) Use the formula to calculate the mean. Check using "one-variable statistics" on your calculator.




33




34

**1.31** Joey's first 14 quiz grades in a marking period were

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 86 | 84 | 91 | 75 | 78 | 80 | 74 | 87 | 76 | 96 | 82 | 90 | 98 | 93 |

(b) Suppose Joey has an unexcused absence for the fifteenth quiz and he receives a score of zero. Determine his final quiz average. What property of the mean does this situation illustrate? Write a sentence about the effect of the zero on Joey's quiz average that mentions this property.

$\bar{x}_{new} = 79.\overline{3}$ $\bar{x}_{old} = 85$

Since zero is an outlier it effects the mean, since the mean is not a resistant measurement of the center of data.

35

**1.33** Suppose a major league baseball team's mean yearly salary for a player is \$1.2 million, and that the team has 25 players on its active roster. What is the team's annual payroll for players? If you knew only the median salary, would you be able to answer the question? Why or why not?

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$\$1,200,000 = \frac{1}{25} SUM$$

$$\$1,200,000 \times 25 = SUM$$

$$\$30 million = SUM$$

36

## Measuring spread or variability:

- Range
  - Difference between largest and smallest points.
  - Not **resistant** to the influence of extreme observations.
- Interquartile Range (IQR)
  - Measures the spread of the middle half of the data.
  - Is **resistant** to the influence of extreme observations.
  - Quartile 3 minus Quartile 1.

37

## To calculate quartiles:

1. Arrange the observations in increasing order and locate the median M.
2. The first quartile $Q_1$ is the median of the observations whose position in the ordered list is to the left of the overall median.
3. The third quartile $Q_3$ is the median of the observations whose position in the ordered list is to the right of the overall median.

38

## The five number summary and box plots.

- The five number summary
  - Consists of the
    - min, $Q_1$, median, $Q_3$, max
  - Offers a reasonably complete description of center and spread.
  - Used to create a boxplot.
- Boxplot
  - Shows less detail than histograms or stemplots.
  - Best used for side-by-side comparison of more than one distribution.
  - Gives a good indication of symmetry or skewness of a distribution.
  - Regular boxplots conceal outliers.
  - Modified boxplots put outliers as isolated points.

39

**1.36 SSHA SCORES** Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154 109 137 115 152 140 154 178 101 103 126 126 137 165 165 129 200 148

and for 20 first-year college men:

108 140 114 91 180 115 126 92 169 146 109 132 75 88 113 151 70 115 187 104

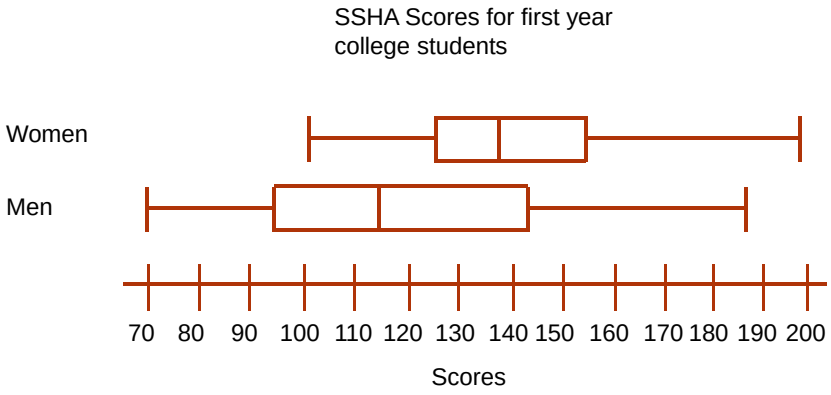
(a) Make side-by-side boxplots to compare the distributions.

•Start by finding the 5 number summary for each of the groups.
•Use your calculator and put the two lists into their own column, then use the 1-var Stats function.

| | Min | $Q_1$ | M | $Q_3$ | Max |
|---|---|---|---|---|---|
| Women: | 101 | 126 | 138.5 | 154 | 200 |
| Men: | 70 | 98 | 114.5 | 143 | 187 |

40

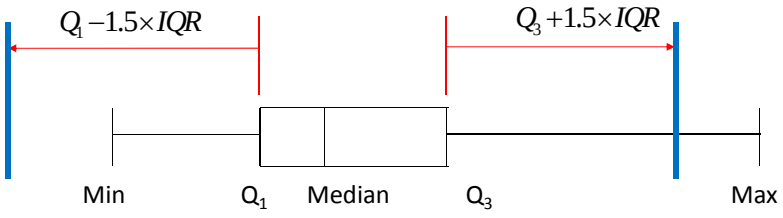
# How to construct a side-by-side boxplot




41

# Calculating outliers

- Outlier
  - An observation that falls outside the overall pattern of the data.
  - Calculated by using the IQR
    - Anything smaller than $Q_1 - 1.5 \times IQR$ or larger than $Q_3 + 1.5 \times IQR$ is an outlier




42

## Constructing a modified boxplot

1.36 **SSHA SCORES** Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154 109 137 115 152 140 154 178 101 103 126 126 137 165 165 129 200 148

| | Min | $Q_1$ | M | $Q_3$ | Max |
|---|---|---|---|---|---|
| Women: | 101 | 126 | 138.5 | 154 | 200 |

$IQR = 28$

$Q_1 - 1.5 \times IQR = 126 - 1.5 \times 28 = 84$

$Q_3 + 1.5 \times IQR = 154 + 1.5 \times 28 = 196$

43

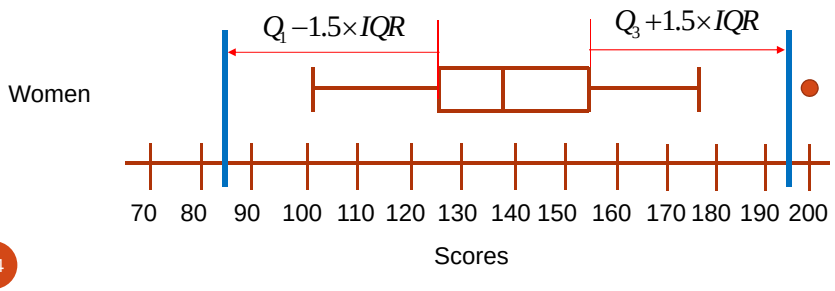
## Constructing a modified boxplot

154 109 137 115 152 140 154 178 101 103 126 126 137 165 165 129 200 148

*Lower bound for outlier* $= 84$ *Upper bound for outlier* $= 196$

| | Min | $Q_1$ | M | $Q_3$ | Max |
|---|---|---|---|---|---|
| Women: | 101 | 126 | 138.5 | 154 | 200 |

SSHA Scores for first year college students




44

## Section 1.3 Day 1

- Homework: #'s 84, 86, 88, 91, 92
- Any questions on pg. 9-12 in additional notes packet.

45

## Measuring Spread:

- **Variance** ($s^2$)
  - The average of the squares of the deviations of the observations from their mean.
  - In symbols, the variance of n observations $x_1, x_2, \ldots, x_n$ is

$$s^2 = \frac{\left(x_1 - \bar{x}\right)^2 + \left(x_2 - \bar{x}\right)^2 + \ldots + \left(x_n - \bar{x}\right)^2}{n-1} \quad \text{or} \quad s^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2$$

- **Standard deviation** (s)
  - The square root of variance.

$$s = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

46

## How to find the mean and standard deviation from their definitions.

- With the list of numbers below, calculate the standard deviation.
  - 5, 6, 7, 8, 10, 12

$$\bar{x} = \frac{5+6+7+8+10+12}{6}$$

$$\bar{x} = 8$$

$$s = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$s = \sqrt{\frac{(5-8)^2+(6-8)^2+(7-8)^2+(8-8)^2+(10-8)^2+(12-8)^2}{6-1}}$$

47

$$s = \sqrt{\frac{(5-8)^2+(6-8)^2+(7-8)^2+(8-8)^2+(10-8)^2+(12-8)^2}{6-1}}$$

$$s = \sqrt{\frac{(-3)^2+(-2)^2+(-1)^2+(0)^2+(2)^2+(4)^2}{5}}$$

$$s = \sqrt{\frac{9+4+1+0+4+16}{5}}$$

$$s = \sqrt{\frac{34}{5}}$$

$$s = \sqrt{6.8}$$

$$s = 2.61$$

48

## Properties of Variance:

- Uses squared deviations from the mean because the sum of all the deviations not squared is always zero.
  - Has square units.
- Found by taking an average but dividing by n-1.
  - The sum of the deviations is always zero, so the last deviation can be found once the other n-1 deviations are known.
    - Means only n-1 of the squared deviations can vary freely, so the average is found by dividing by n-1.
  - n-1 is called the **degrees of freedom**.

49

## Properties of Standard Deviation

- Measures the spread about the mean and should be used only when the mean is chosen as the measure of center.
- Equals zero when there is no spread, happens when all observations are the same value. Otherwise it is always positive.
- Not **resistant** to the influence of extreme observations or strong skewness.

50

## Mean & Standard Deviation Vs. Median & the 5-Number Summary

- Mean & Standard Deviation
  - Most common numerical description of a distribution.
  - Used for reasonably symmetric distributions that are free from outliers.
- Five-Number Summary
  - Offer a reasonably complete description of center and spread.
  - Used for describing skewed distributions or a distribution with strong outliers.

51

## Always plot your data.

- Graphs
  - Give the best overall picture of a distribution.
- Numerical measures of center and spread
  - Only give specific facts about a distribution.
  - Do not describe its entire shape.
  - Can give a misleading picture of a distribution or the comparison of two or more distributions.

52

## Changing the unit of measurement.

- Linear Transformations
  - Changes the original variable x into the new variable $x_{new.}$
  - $x_{new} = a + bx$
  - Do not change the shape of a distribution.
  - Can change one or both the center and spread.
  - The effects of the changes follow a simple pattern.
    - Adding the constant (a) shifts all values of x upward or downward by the same amount.
      - Adds (a) to the measures of center and to the quartiles but does not change measures of spread.
    - Multiplying by the positive constant (b) changes the size of the unit of measurement.
      - Multiplies both the measures of center (mean and median) and the measures of spread (standard deviation and IQR) by (b).

53

The table shows an original data set and two different linear transformations for that set.

| Original (x) | x + 12 | 3(x) - 7 |
|---|---|---|
| 5 | 17 | 8 |
| 6 | 18 | 11 |
| 7 | 19 | 14 |
| 8 | 20 | 17 |
| 10 | 22 | 23 |
| 12 | 24 | 29 |

What are the original and transformed mean, median, range, quartiles, IQR, variance and standard deviation?

54

- Original Data
  - Mean: $\bar{X} = 8$
  - Median: 7.5
  - $Q_1$: 6
  - $Q_3$: 10
  - IQR: 4
  - Range: 7
  - Variance: 6.8
  - St Dev: 2.61
- $x + 12$
  - Mean: $\bar{X} = 20$
  - Median: 19.5
  - $Q_1$: 18
  - $Q_3$: 22
  - IQR: 4
  - Range: 7
  - Variance: 6.8
  - St Dev: 2.61
- $3(x) - 7$
  - Mean: $\bar{X} = 17$
  - Median: 15.5
  - $Q_1$: 11
  - $Q_3$: 23
  - IQR: 12
  - Range: 21
  - Variance: 61.2
  - St Dev: 7.82

55

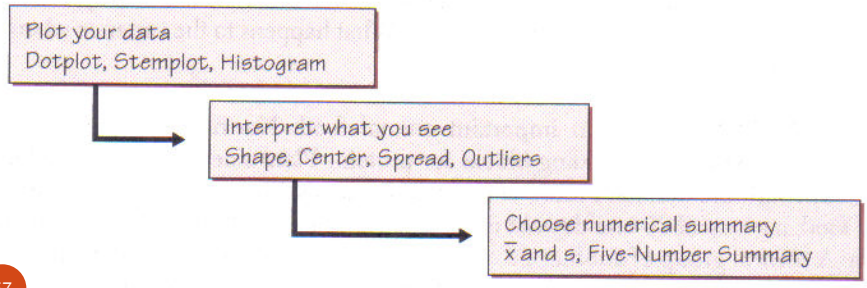
## Section 1.3 & 2.1 Day 2

- Homework: #'s Section 1-3: 97, 98, 103; Section 2-1: 19, 20, 22, R.2.3
- Any questions on pg. 13-16 in additional notes packet.

56

## Chapter review

Data analysis is the art of describing data using graphs and numerical summaries. The purpose of data analysis is to describe the most important features of a set of data. This chapter introduces data analysis by presenting statistical ideas and tools for describing the distribution of a single variable. The figure below will help you organize the big ideas.




57

### A. DATA

1. Identify the individuals and variables in a set of data.
2. Identify each variable as categorical or quantitative. Identify the units in which each quantitative variable is measured.

### B. DISPLAYING DISTRIBUTIONS

1. Make a bar graph and a pie chart of the distribution of a categorical variable. Interpret bar graphs and pie charts.
2. Make a dotplot of the distribution of a small set of observations.
3. Make a stemplot of the distribution of a quantitative variable. Round leaves or split stems as needed to make an effective stemplot.
4. Make a histogram of the distribution of a quantitative variable.
5. Construct and interpret an ogive of a set of quantitative data.

58

## C. INSPECTING DISTRIBUTIONS (QUANTITATIVE VARIABLES)

1. Look for the overall pattern and for major deviations from the pattern.

2. Assess from a dotplot, stemplot, or histogram whether the shape of a distribution is roughly symmetric, distinctly skewed, or neither. Assess whether the distribution has one or more major peaks.

3. Describe the overall pattern by giving numerical measures of center and spread in addition to a verbal description of shape.

4. Decide which measures of center and spread are more appropriate: the mean and standard deviation (especially for symmetric distributions) or the five-number summary (especially for skewed distributions).

5. Recognize outliers.

59

## D. TIME PLOTS

1. Make a time plot of data, with the time of each observation on the horizontal axis and the value of the observed variable on the vertical axis.

2. Recognize strong trends or other patterns in a time plot.

## E. MEASURING CENTER

1. Find the mean $\bar{x}$ of a set of observations.

2. Find the median $M$ of a set of observations.

3. Understand that the median is more resistant (less affected by extreme observations) than the mean. Recognize that skewness in a distribution moves the mean away from the median toward the long tail.

60

## F. MEASURING SPREAD

1. Find the quartiles $Q_1$ and $Q_3$ for a set of observations.

2. Give the five-number summary and draw a boxplot; assess center, spread, symmetry, and skewness from a boxplot. Determine outliers.

3. Using a calculator, find the standard deviation $s$ for a set of observations.

4. Know the basic properties of $s$: $s \geq 0$ always; $s = 0$ only when all observations are identical; $s$ increases as the spread increases; $s$ has the same units as the original measurements; $s$ is increased by outliers or skewness.

61

## G. CHANGING UNITS OF MEASUREMENT (LINEAR TRANSFORMATIONS)

1. Determine the effect of a linear transformation on measures of center and spread.

2. Describe a change in units of measurement in terms of a linear transformation of the form $x_{new} = a + bx$.

## H. COMPARING DISTRIBUTIONS

1. Use side-by-side bar graphs to compare distributions of categorical data.

2. Make back-to-back stemplots and side-by-side boxplots to compare distributions of quantitative variables.

3. Write narrative comparisons of the shape, center, spread, and outliers for two or more quantitative distributions.

62

## Chapter 1 Complete

- Homework: #'s
- Any questions on pg. 17-20 in additional notes packet.

63