# Math 140
# Introductory Statistics

First midterm

February 20 2013

# 3. Stemplots

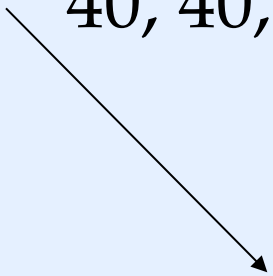
Speeds of mammals (mph)

11, 12, 20, 25, 30, 30, 30, 32, 35, 39,
40, 40, 40, 42, 45, 48, 50, 70

# 3. Stemplots

Speeds of mammals (mph)

11, 12, 20, 25, 30, 30, 30, 32, 35, 39,
40, 40, 40, 42, 45, 48, 50, 70

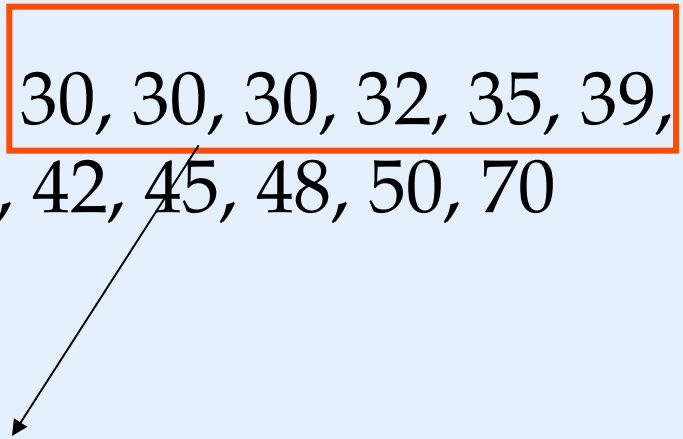


1 | 1 2

# 3. Stemplots

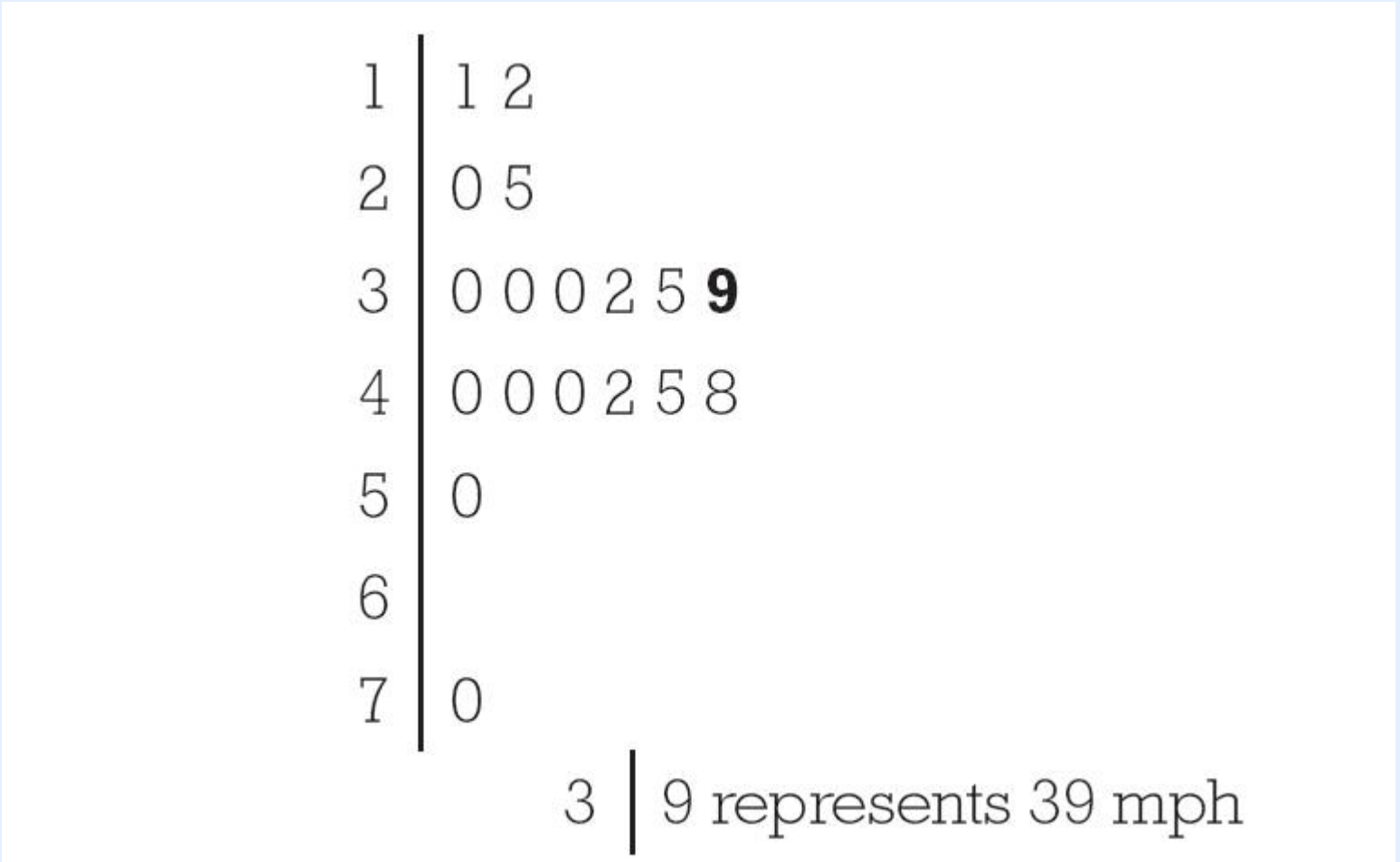
Speeds of mammals (mph)

11, 12, 20, 25, 30, 30, 30, 32, 35, 39,
40, 40, 40, 42, 45, 48, 50, 70



3 | 0 0 0 2 5 9

# 3. Stemplots



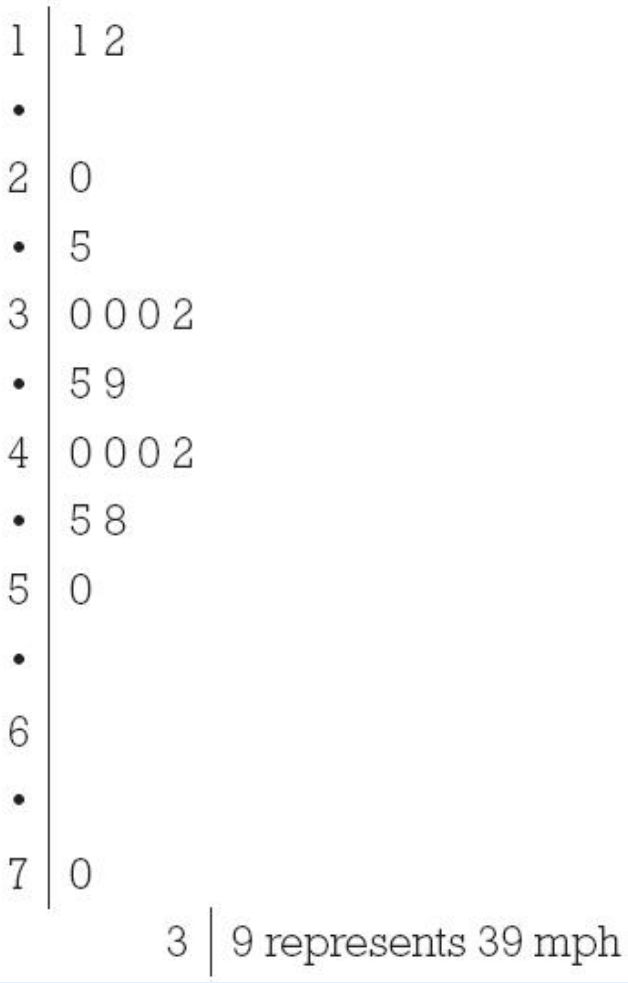
# 3. Stemplots

Or stem-and-leaf plots

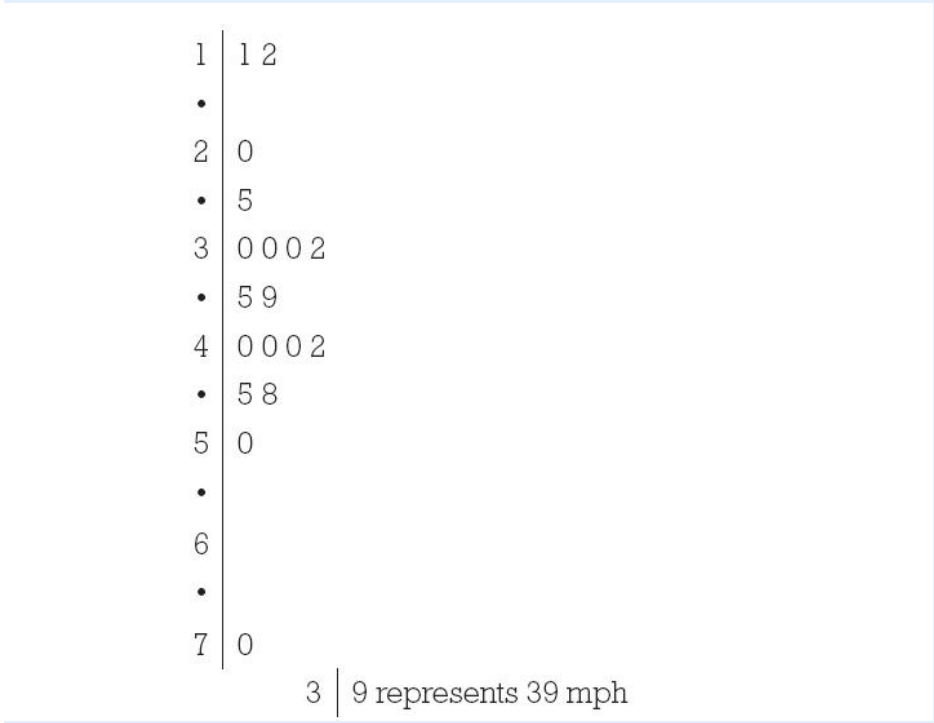
Numbers on the left are called stems
(the first digits of the data value)

Numbers on the right are called leaves
(the last digit of the data value)

# Split stemplots



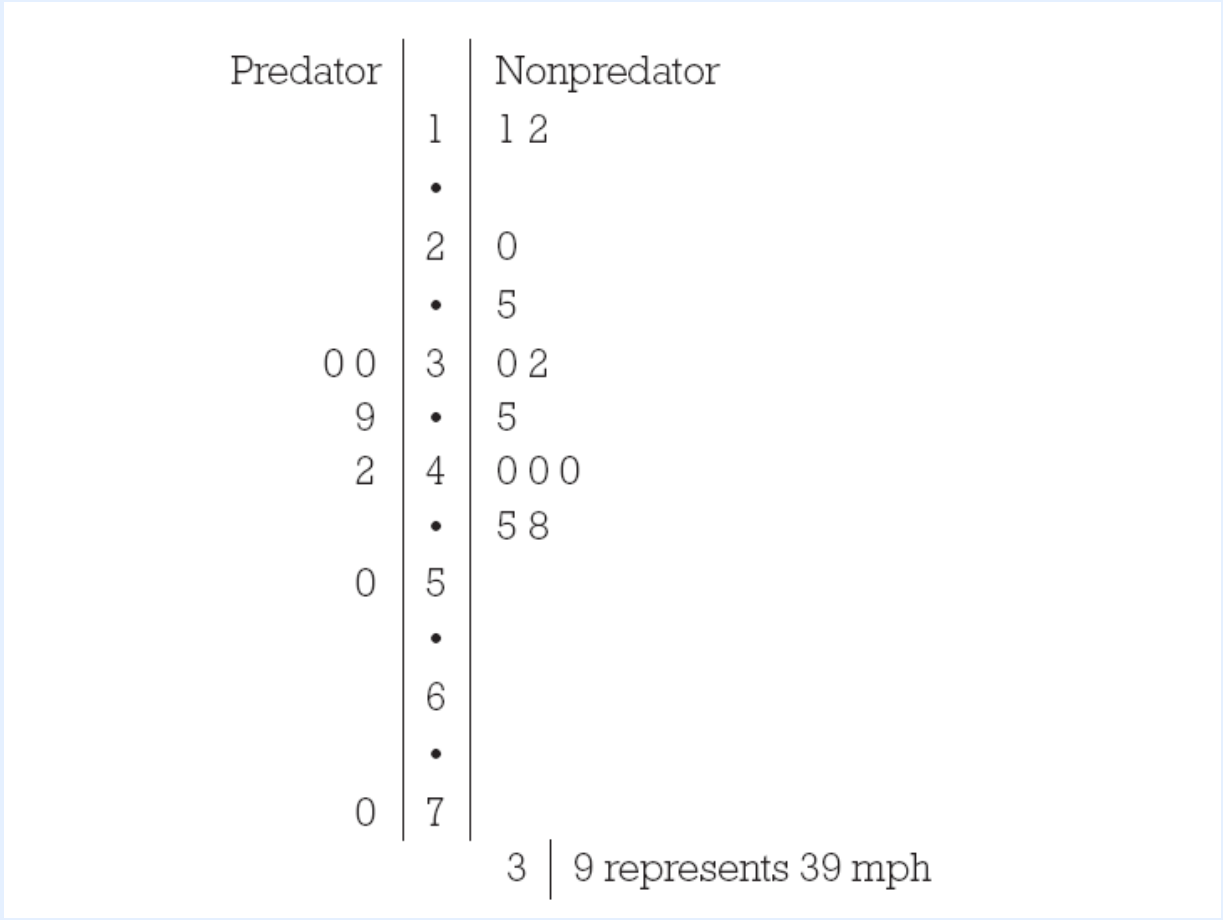
# Split stemplots




The unit digits 0,1,2,3,4 are associated with the first stem and they are placed on the first line.

The unit digits 5,6,7,8,9 are associated with the second stem
and they are placed on the second line.

# Back to back stemplots



| Predator | | Nonpredator |
|---|---|---|
| | 1 | 1 2 |
| | • | |
| | 2 | 0 |
| | • | 5 |
| 0 0 | 3 | 0 2 |
| 9 | • | 5 |
| 2 | 4 | 0 0 0 |
| | • | 5 8 |
| 0 | 5 | |
| | • | |
| | 6 | |
| | • | |
| 0 | 7 | |

3 | 9 represents 39 mph

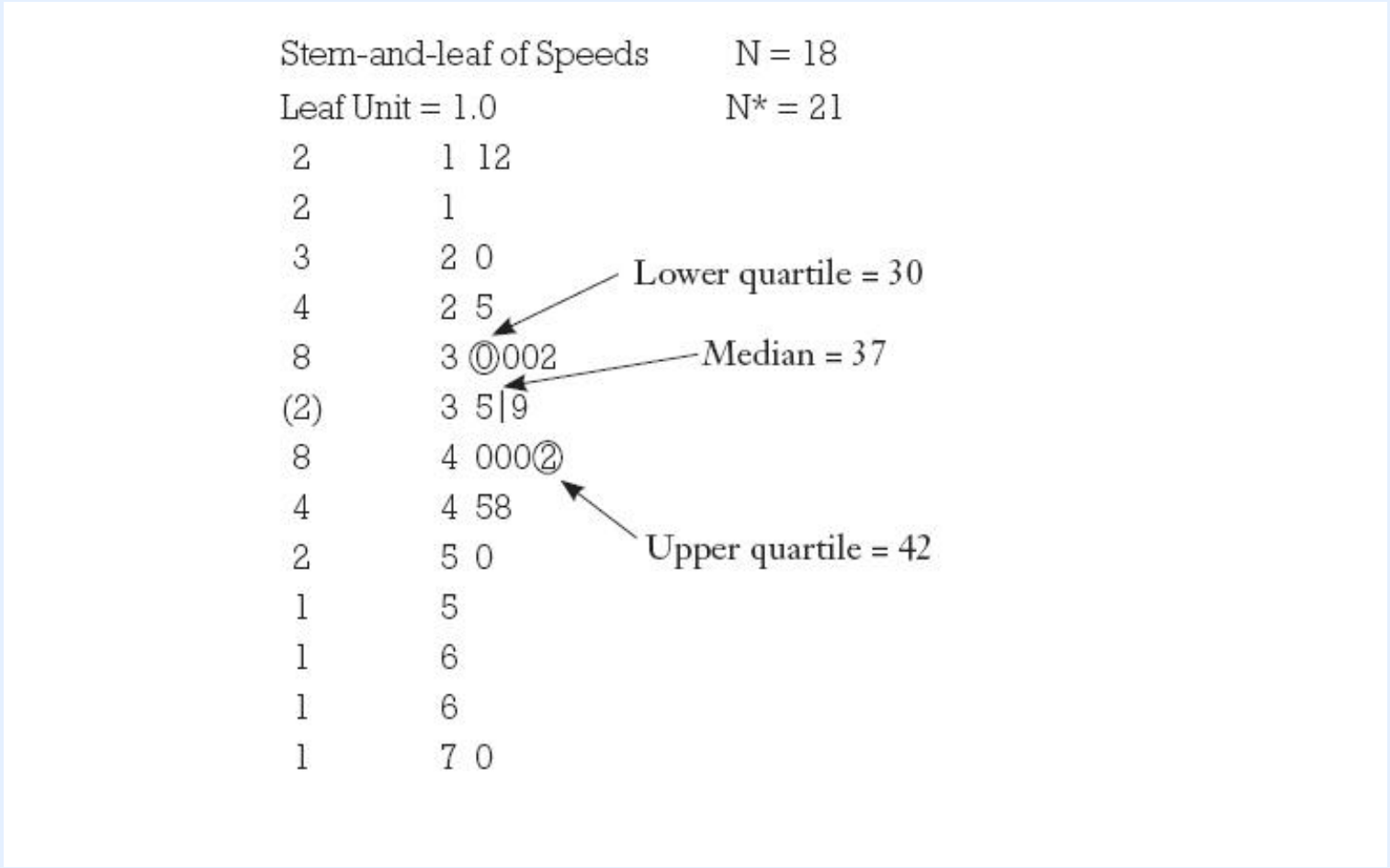
The data is differentiated on whether the mammals are predators or non-predators

# Who has the faster speed?

| Predator | | Nonpredator |
|---|---|---|
| | 1 | 1 2 |
| | • | |
| | 2 | 0 |
| | • | 5 |
| 0 0 | 3 | 0 2 |
| 9 | • | 5 |
| 2 | 4 | 0 0 0 |
| | • | 5 8 |
| 0 | 5 | |
| | • | |
| | 6 | |
| | • | |
| 0 | 7 | |

3 | 9 represents 39 mph

# Calculating medians and quartiles



# Stemplots work best when

Small number of values to plot

Want to keep track of individual values
(at least approximately)

Want to see shape of distribution

Have two or more groups that we want to compare

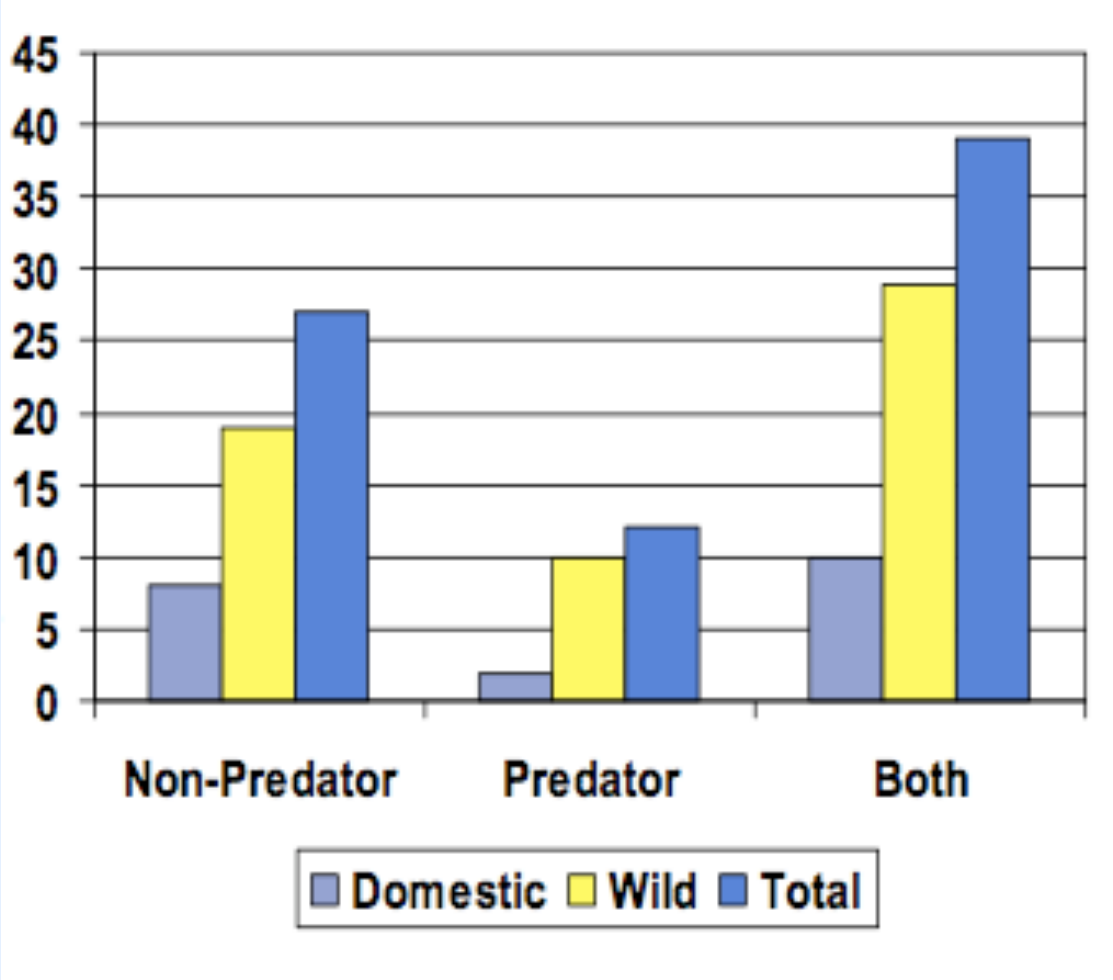
# 4. Bar graphs

One bar for each category

The height of the bar tells the frequency

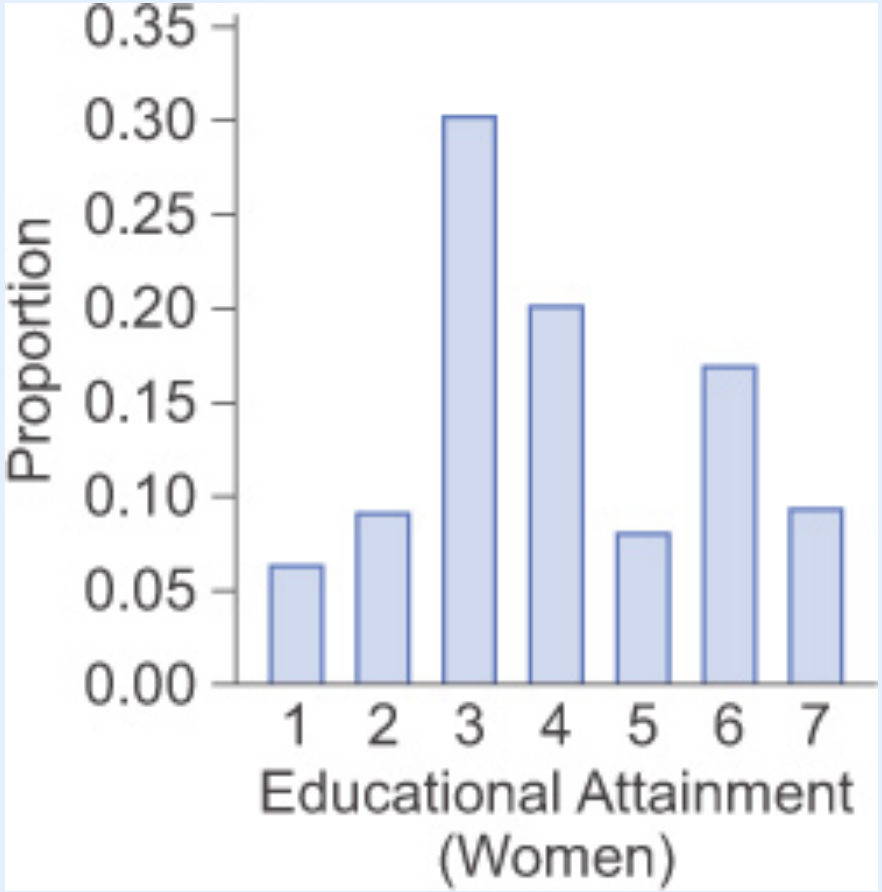
Bar graphs have categories in the horizontal axis, as opposed to histograms which have measurements.

# Bar graphs




Bars are separated so there is no confusion

# US working women age 25 or older




1. Less than 9th grade
2. 9th to 12th grade, no diploma
3. High school grad
4. Some college, no degree
5. Associate degree
6. Bachelor degree
7. Phd or professional degree

Modal category: category with highest frequency

# Measures of center: mean and median

Earlier we used visual estimates
to find out center and spread

Now we will learn how to calculate them exactly

Measures of Center

Mean  Median

Measures of Spread

Standard Deviation  Inter Quartile Range

# Center: Mean (average)

Denoted as $\bar{x}$

$$\bar{x} = \frac{\text{sum of values}}{\text{number of values}} = \frac{\sum x}{n}$$

Example: 5, 12, 34, 18, 37, 11, 9, 21, 30, 6

$$\bar{x} = \frac{5+12+34+18+37+11+9+21+30+6}{10} = 18.3$$

# Center: Median

Denoted as Q2

Divides data into equal halves.

List all n values in increasing order
and find the middle one.

If n is odd the middle one is (n+1)/2
Say n=17 median is at (17+1)/2 = 9
And there 8 to the left, 8 to the right

If n is even the median is the average of the two
Values on and after n/2 positions

# Center: Median

Example: 5, 6, 9, 11, 12, 18, 21, 30, 34, 37, 41

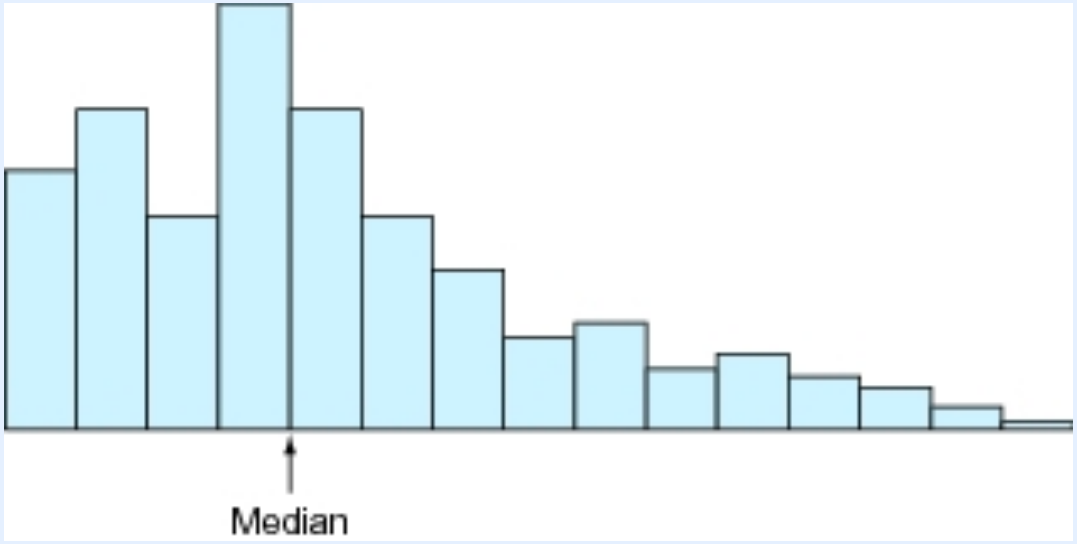
n=11 median is (n+1)/2 = $6^{th}$ position
18

Example: 5, 6, 9, 11, 12, 18, 21, 30, 34, 37



n=10 median is between the two
values at n/2 = $5^{th}$ position  (12 + 18)/2 = 15

# Center: Median



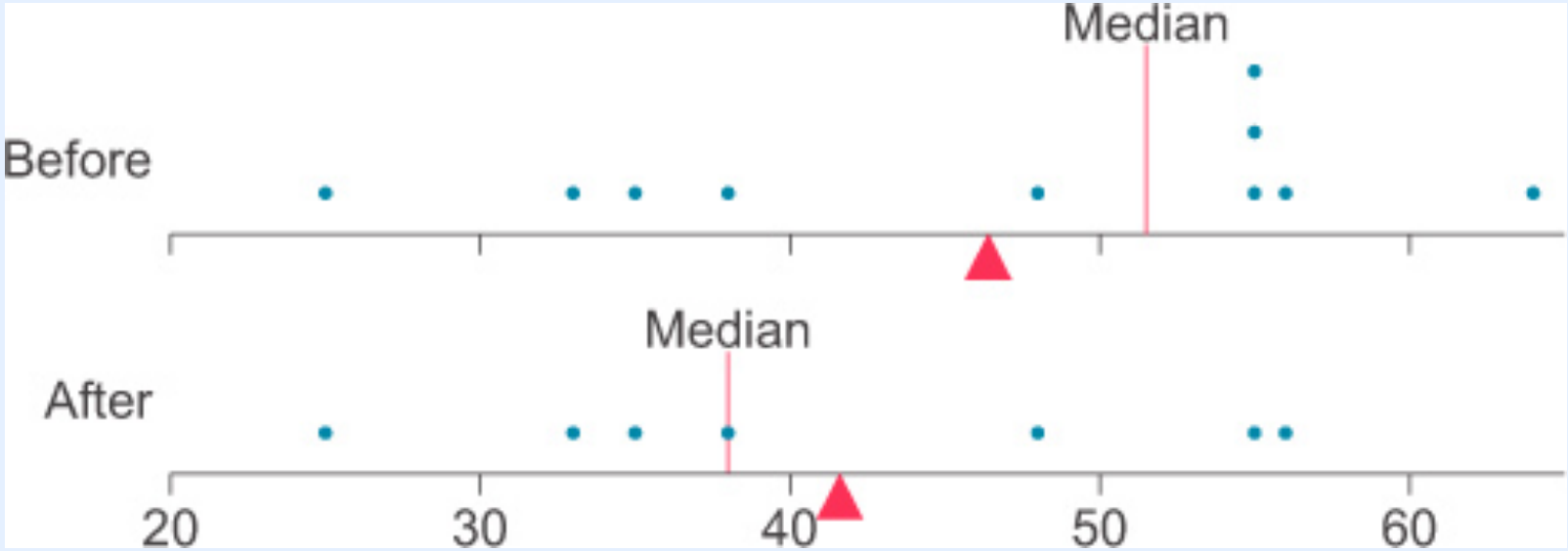

If placed in a histogram the median will divide the total area in two equal parts

# Median

Calculate means and
medians before and after Westvaco layoffs

25, 33, 35, 38, 48, 55, 56, 55, 55, 64

# Median




Calculate means and
medians before and after Westvaco layoffs

25, 33, 35, 38, 48, 55, 56, 55, 55, 64

# Spread - IQR

First Quartile or Lower Quartile Q1
Third Quartile or Upper Quartile Q3

Medians of left hand side of data and right hand side of Data with respect to the median

Inter Quartile Range
IQR = Q3 - Q1

# Five number summary
# Q1, Q3, median, min, max

11, 12, 20, 25, 30, 30, 30, 32, 35, 39,
40, 40, 40, 42, 45, 48, 50, 70

These give the five number summary
From which to calculate

IQR = Q3 - Q1
range = max-min

# Five number summary

11, 12, 20, 25, 30, 30, 30, 32, 35, 39,
40, 40, 40, 42, 45, 48, 50, 70

Min = 11
Max = 70
Q1 = 30
Median = Q2 = 37
Q3 = 42

Range = max-min = 70 - 11 = 59
IQR= Q3 - Q1= 42- 30 = 12

# Outliers

If a value is more than 1.5 times the IQR
from the nearest quartile it may be an outlier

Is the cheetah an outlier?
Is the pig an outlier?
Is the gazelle an outlier?
Is the lion an outlier?

Which animal is the largers non-outlier?

# Outliers - definitions

11, 12, 20, 25, 30, 30, 30, 32, 35, 39,
40, 40, 40, 42, 45, 48, 50, 70

A value is an outlier if it is more than 1.5 times the IQR from the nearest quartile

IQR= 12

1.5*IQR = 1.5*12 = 18

Q1=30 --- outliers are all data less than 30-18 = 12
Q3=42 --- outliers are all data more than 42+18 = 60

# Spread - Deviation

Deviation of a value x is how far it is from the mean

$$x - \bar{x}$$

This value is different for every data point x
and can be negative or positive

# Standard deviation

$$\sigma_n = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

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

The custom is to use $\sigma_n$

# Standard deviation

Data 2, 7, 8, 12, 12, 19  n=?  average $\bar{x}$ = ?

| x | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 2 | | |
| 7 | | |
| 8 | | |
| 12 | | |
| 12 | | |
| 19 | | |
| | | |
| total sum = 60 | | |

# Standard deviation

Example. Data: 2,7,8,12,12,19

$n = 6,\ \bar{x} = (2+7+8+12+12+19)/6 = 10$

| $x$ | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 2 | -8 | 64 |
| 7 | -3 | 9 |
| 8 | -2 | 4 |
| 12 | 2 | 4 |
| 12 | 2 | 4 |
| 19 | 9 | 81 |
| 60 | 0 | 166 |

Find $\sigma_n$ and $\sigma_{n-1}$

# Standard deviation

Example. Data: 2,7,8,12,12,19

$n = 6,\ \bar{x} = (2+7+8+12+12+19)/6 = 10$

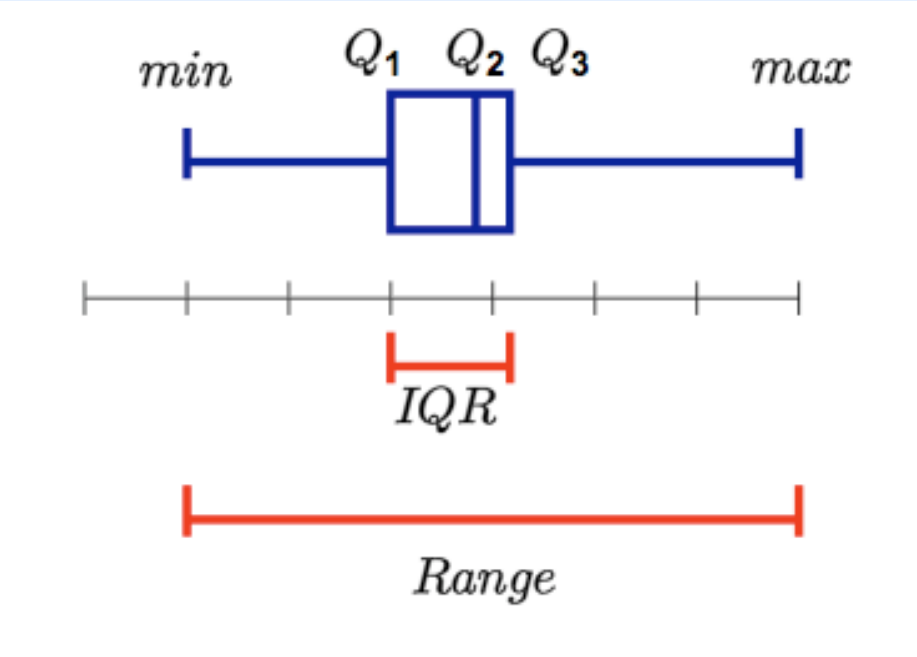
| $x$ | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 2 | -8 | 64 |
| 7 | -3 | 9 |
| 8 | -2 | 4 |
| 12 | 2 | 4 |
| 12 | 2 | 4 |
| 19 | 9 | 81 |
| 60 | 0 | 166 |

$$\sigma_n = \sqrt{\frac{166}{6}} \approx 5.2599$$

$$\sigma_{n-1} = \sqrt{\frac{166}{5}} \approx 5.7619$$

# Box Plots

Graphical display of 5 number summary
Q1, Q2, Q3, max, min



# Hk

Page 63, E25, E26, E28, E29, E32, E33, E36, E41,