Math 140
Introductory Statistics

Math 140 tutoring:
LIVE OAK 1319

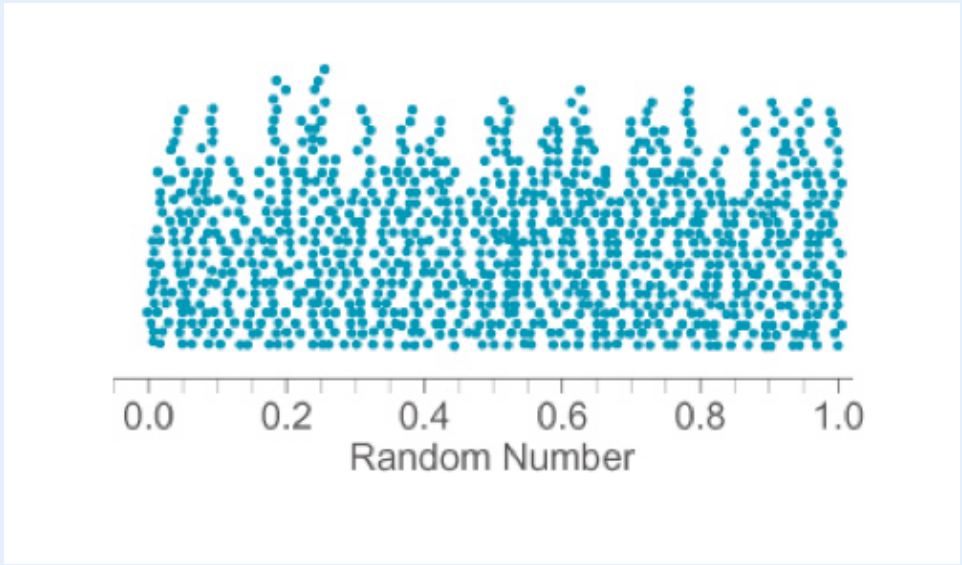
MW 11:30 - 4:30 TTh 3:30-5:30 F 10:30-12:30

General Hours:

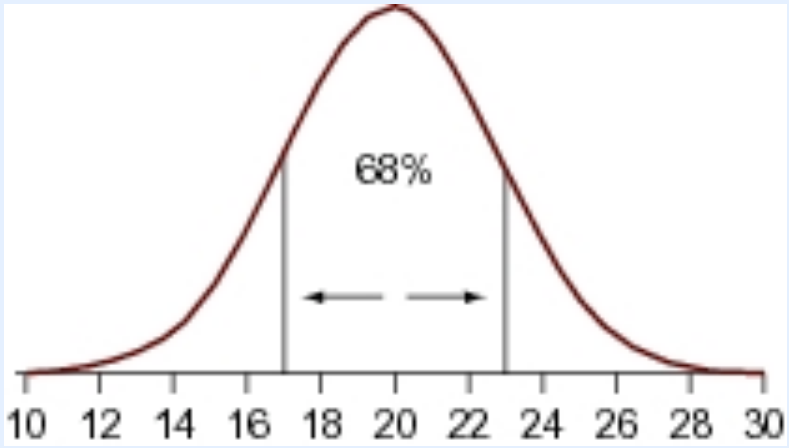
M - Th 10:00 - 5:30 F 10:00 - 3:00

Later: Saturday from 11 to 2.

Last time

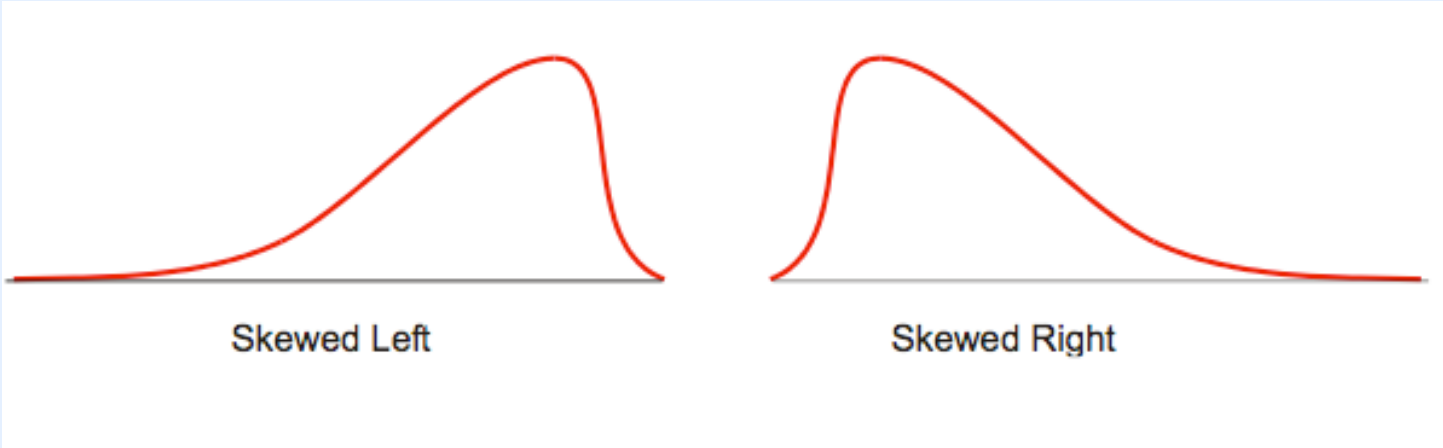


Uniform - rectangular distribution



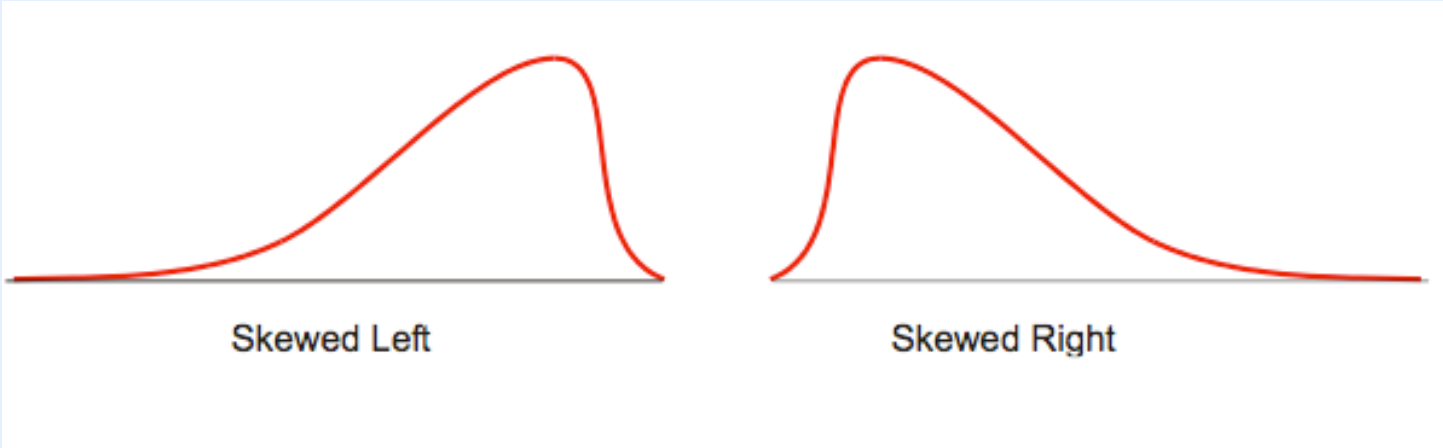
Normal distribution
mean
inflection points
standard deviation

Skewed distributions



Not symmetric curves
Data is bunched on one end
and a tail appears on the other side

New tools

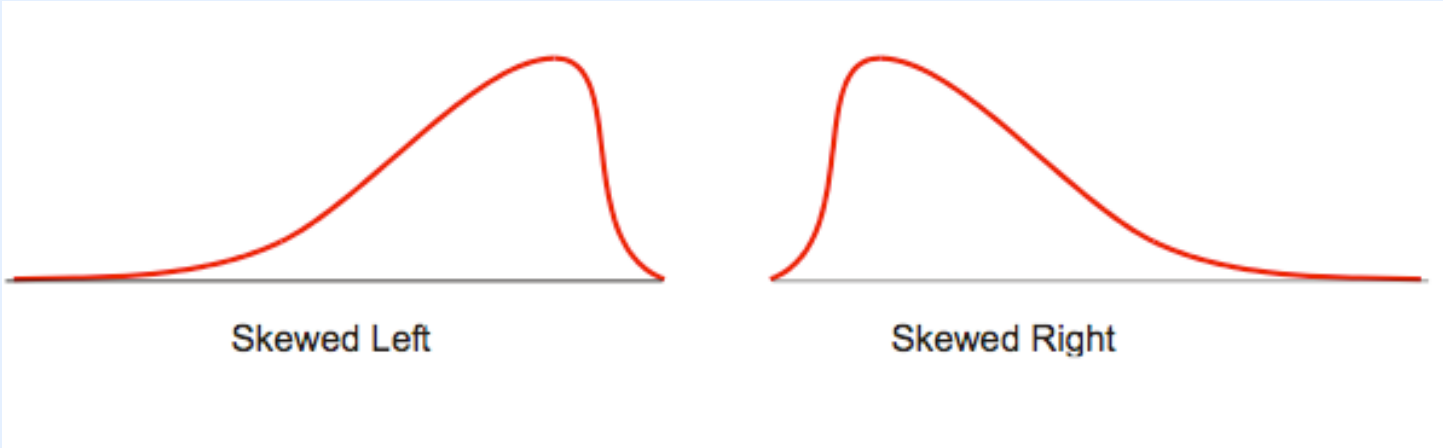


Median:

The value of the line dividing the number of values in equal halves

The area (or the number of points) to the left or to the right of the median are equal

New tools



Quartiles:

Once you have found the median, look at the left of the distribution and repeat the same procedure. This new value is called the lower quartile Q1

Repeat on the right, and find the upper quartile Q3

Median, lower and upper quartiles

They divide the distribution in quarters.

How much data is contained between Q1 and Q3?

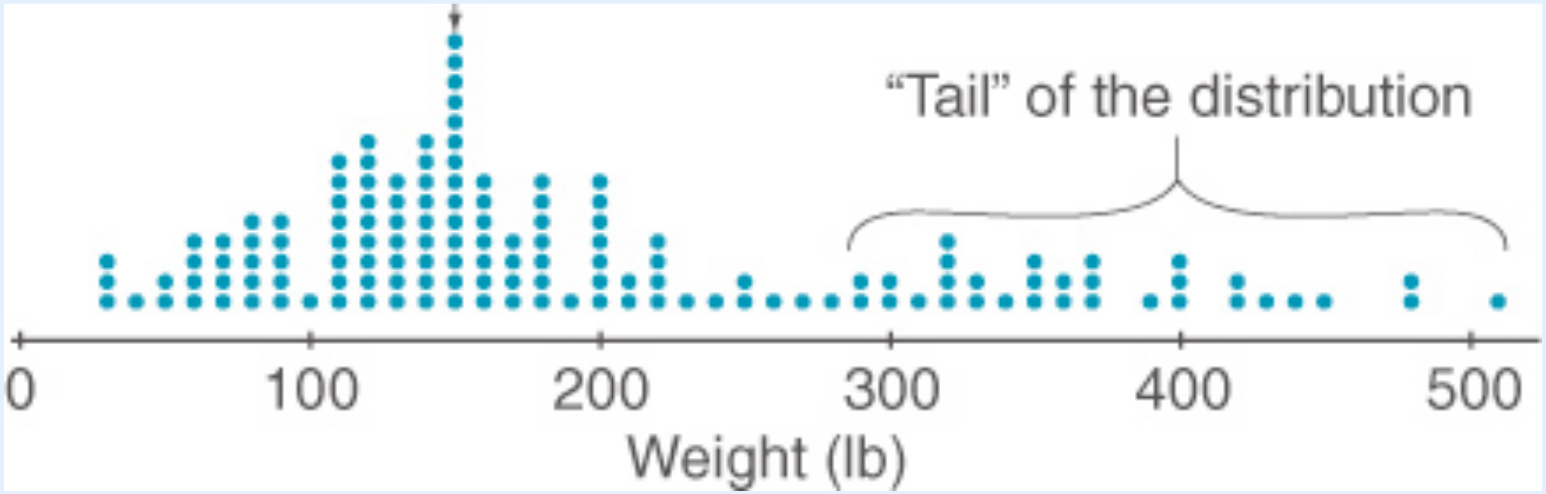
Median, lower and upper quartiles

They divide the distribution in quarters.

How much data is contained between Q1 and Q3?

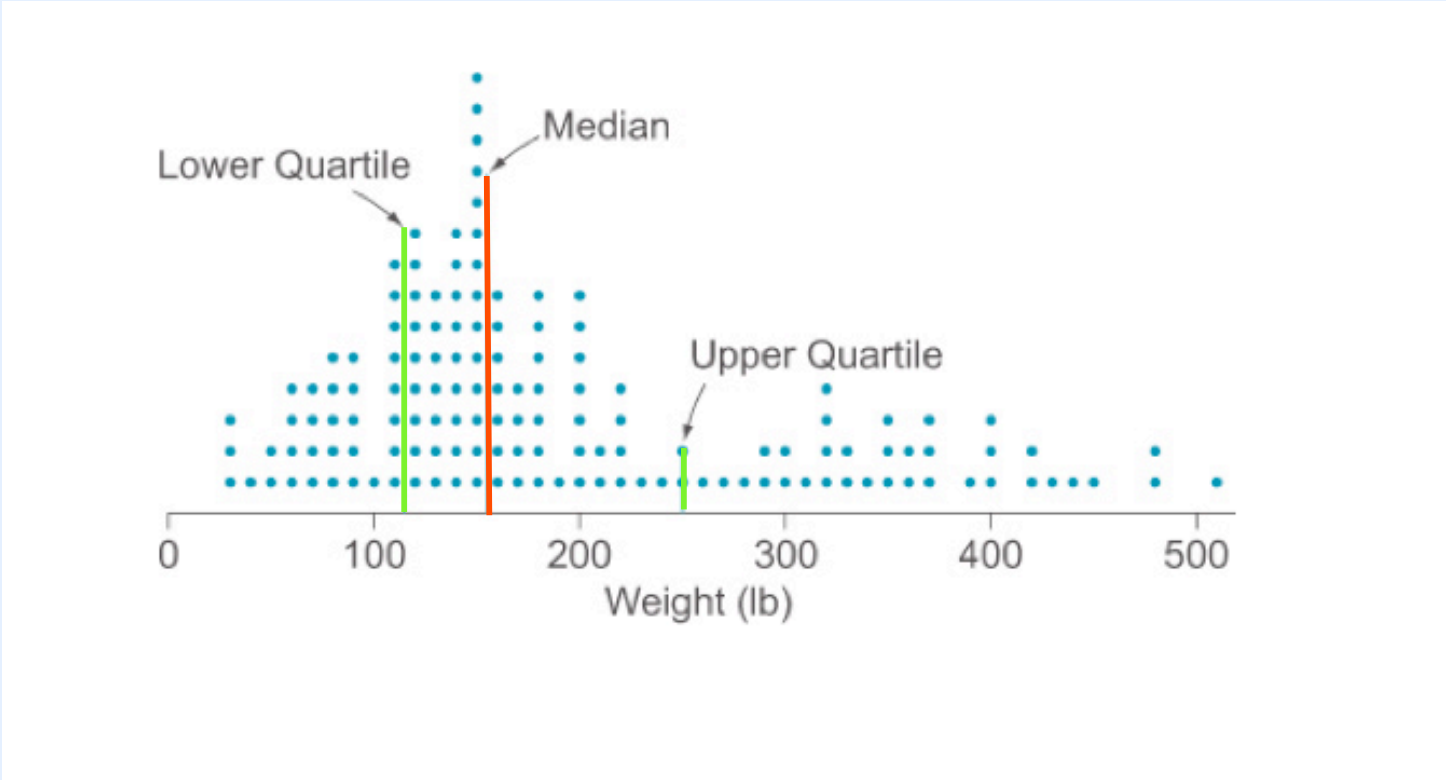
50%

Example - the weight of bears



Find median, Q1 and Q3

Example - the weight of bears



Median ~ 155 lb

Q1 ~ 115 lb

Q3 ~ 250 lb

Outliers, gaps and clusters

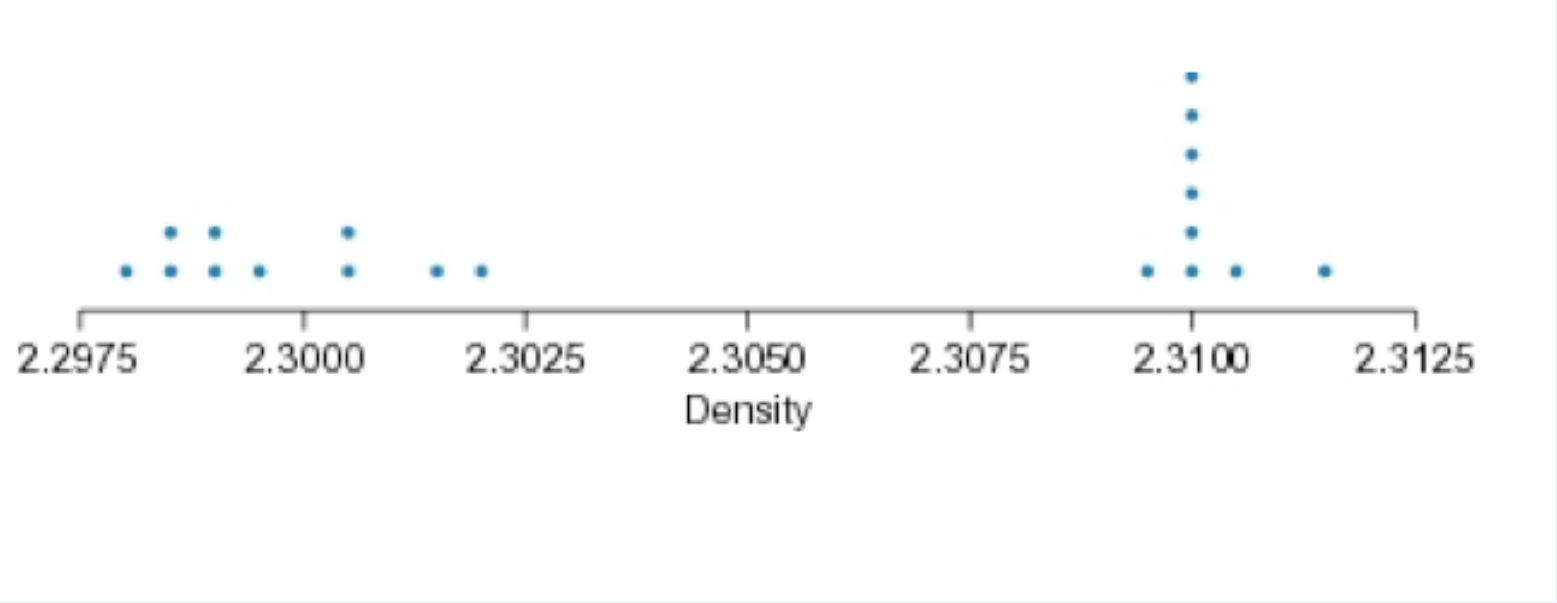
outliers are “special” values that stand out
when we look at the distribution

mistakes? Just flukes (a really really big bear!)
sometimes they can lead to interesting discoveries

gaps and clusters

“informal” definitions

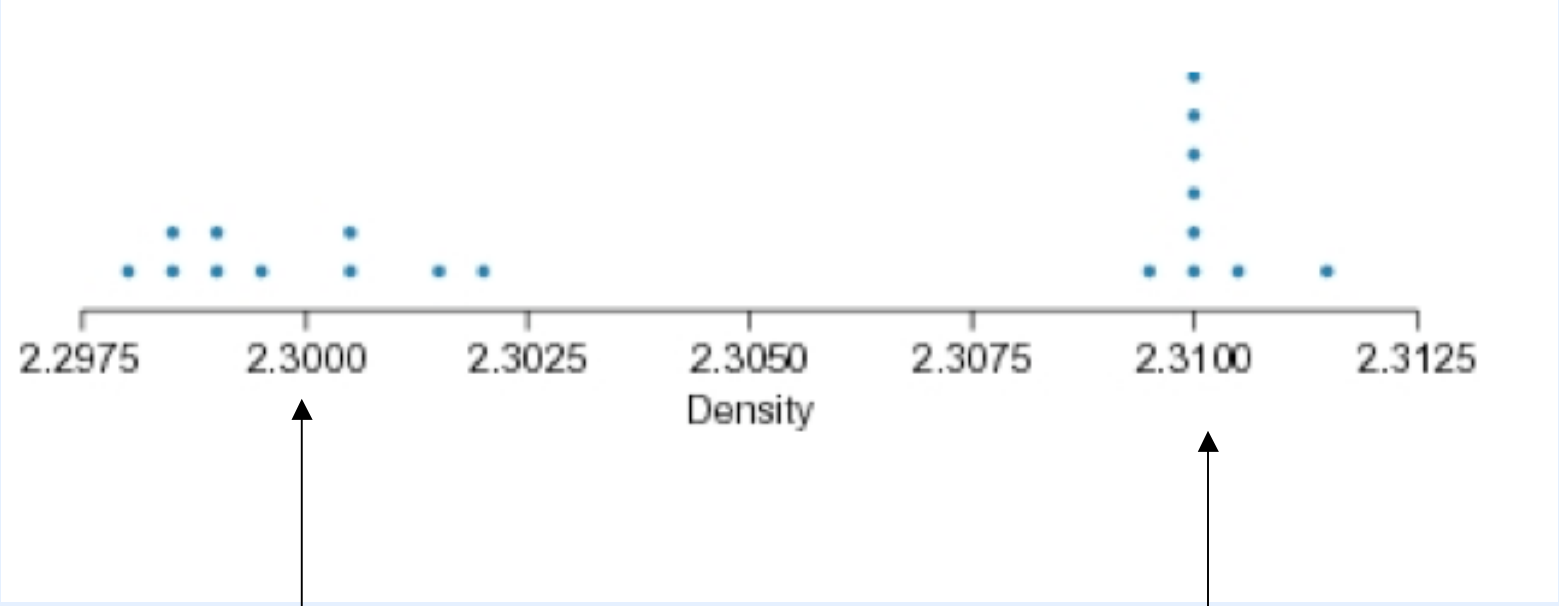
Outliers, gaps and clusters



Lord Rayleigh's densities of nitrogen

what is different between the two?
why two clusters?

Outliers, gaps and clusters

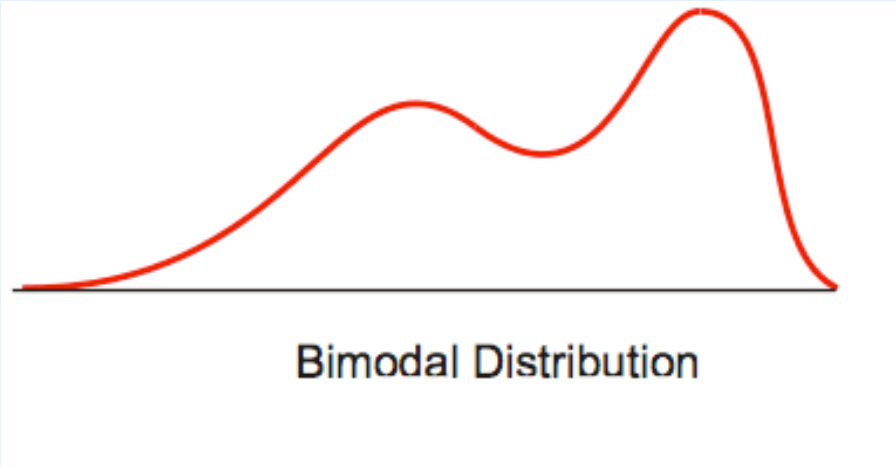


Chemically produced

Atmospheric

There might be something else in the atmosphere!

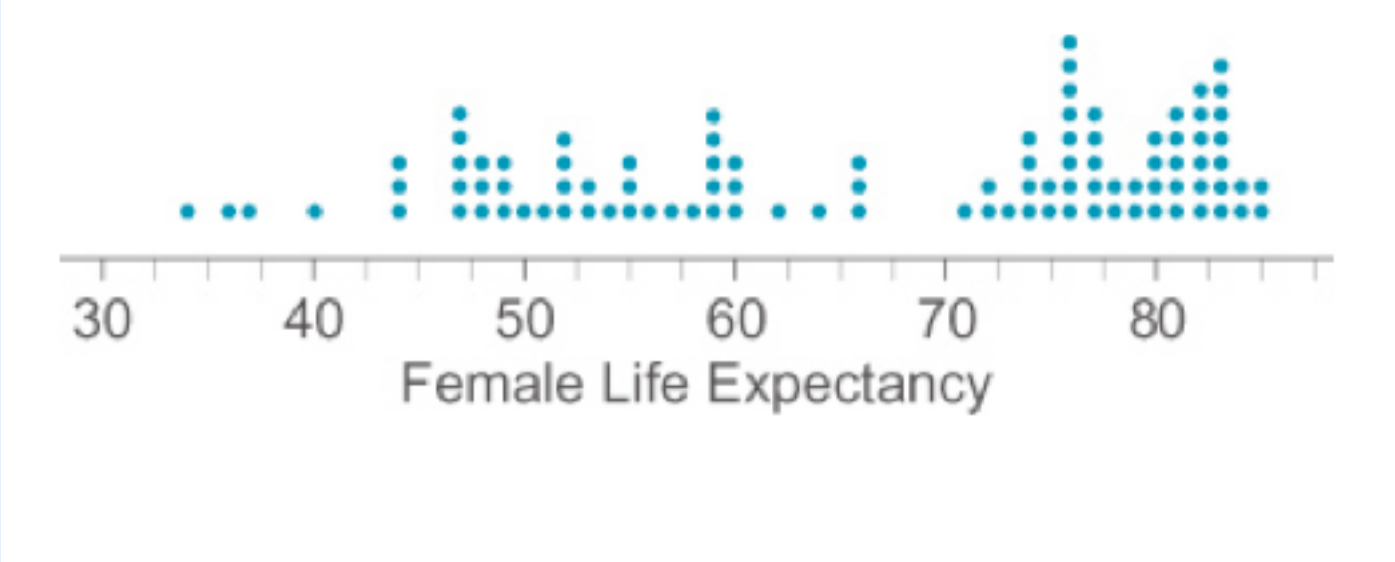
Bimodal distributions



Some distributions have two peaks instead of one

Unimodal (one peak)
Bimodal (two peaks)
Multimodal (many peaks)

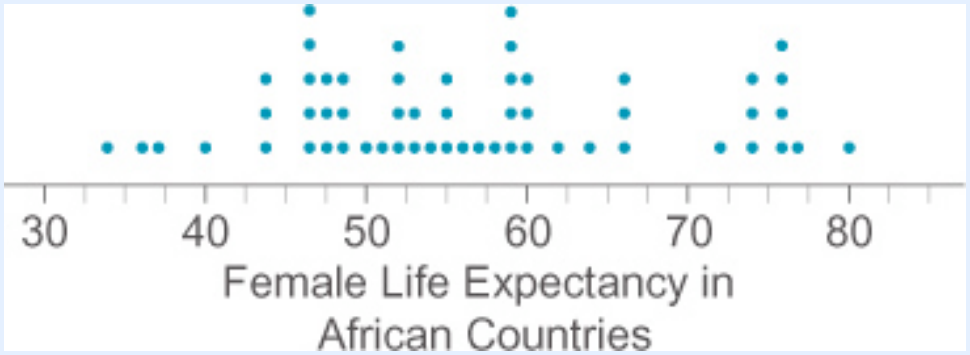
Example

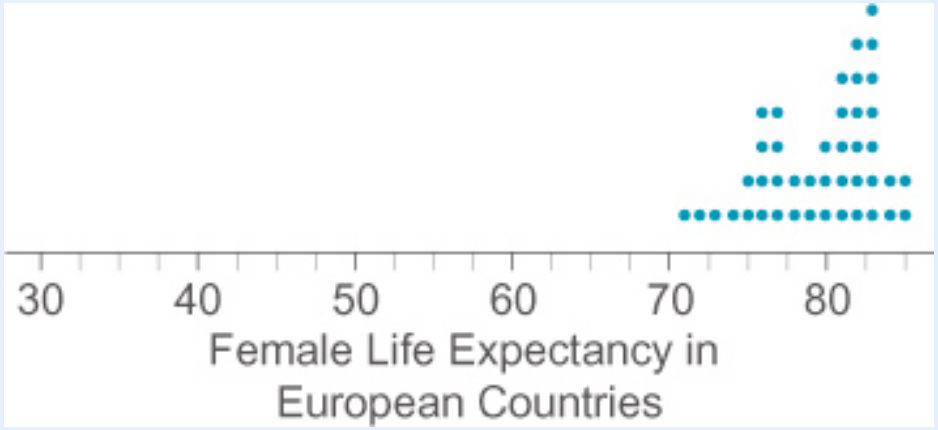


Bimodal - what to make of this?

is there other info we can use?

Splitting data





Africa - spread out

Europe - skewed to left

Quantitative vs. categorical data

Quantitative : data in form of numbers

that can be compared and

that can take a large range of values

Categorical : a case can belong to a category or not

How to look at quantitative data?

1. Dot plots

Each dot represents a case

Dots may represent more than one case
(one dot may represent 1000 cases - USA births)

We can use different symbols for different
Categories of data

Dot plots work best when

Relatively small number of values to plot

Want to keep track of individuals

Want to see the shape of the distribution

Have one group or a small number of groups that we want to compare

Making plots by hand

2. Histograms

Similar to dot plots but where data is grouped

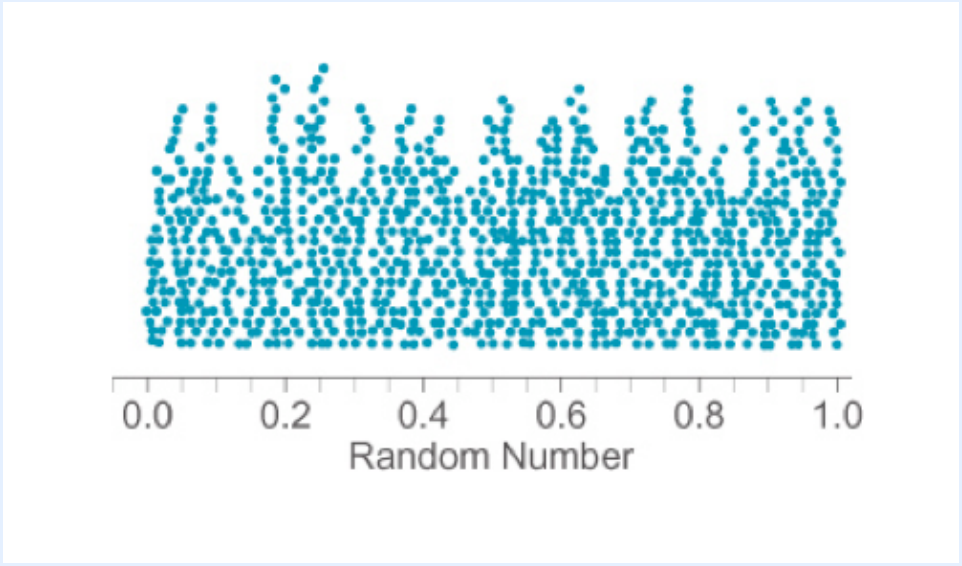
Groups of cases represented as rectangles or bars

The vertical axis gives the number of cases
(called frequency or count)

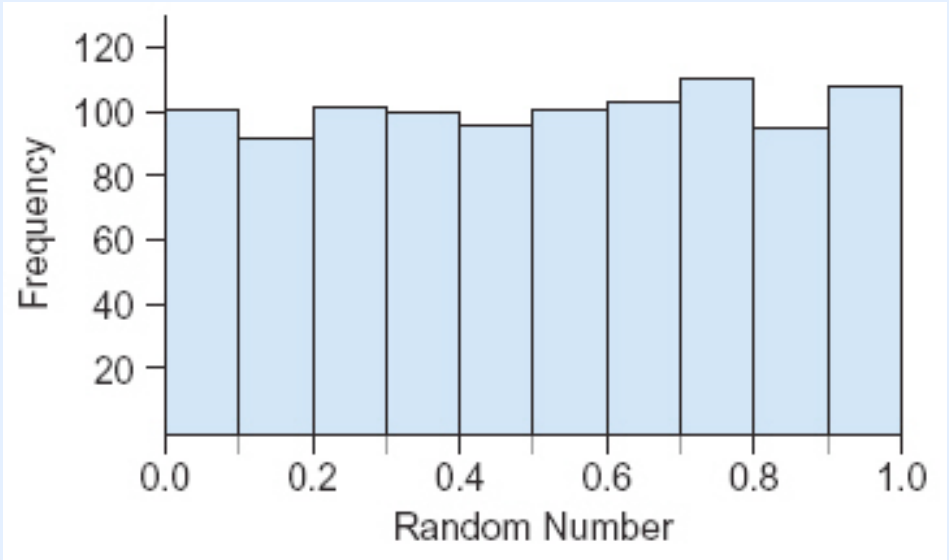
By convention borderline values
go to the bar on the right.

There is no prescribed number for the width of the bars.

Random numbers

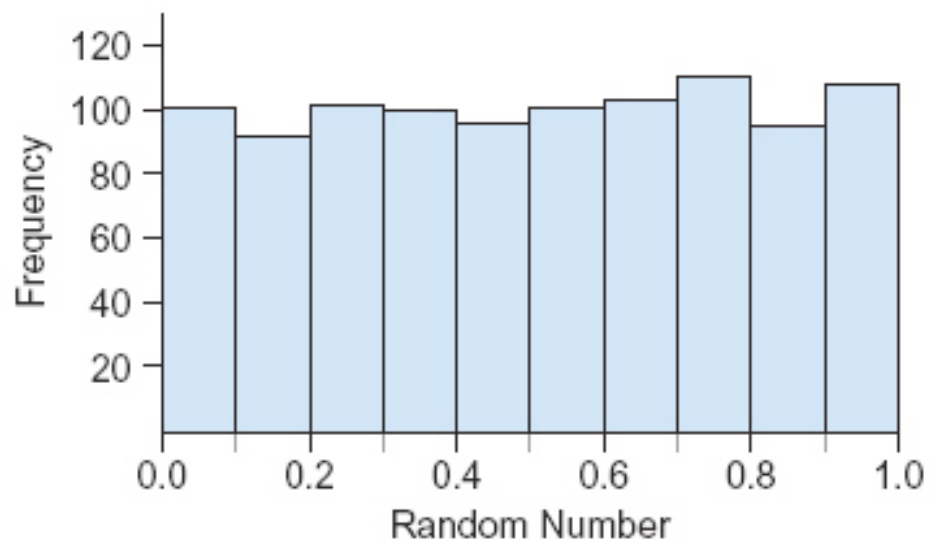


Dot plot



Histogram

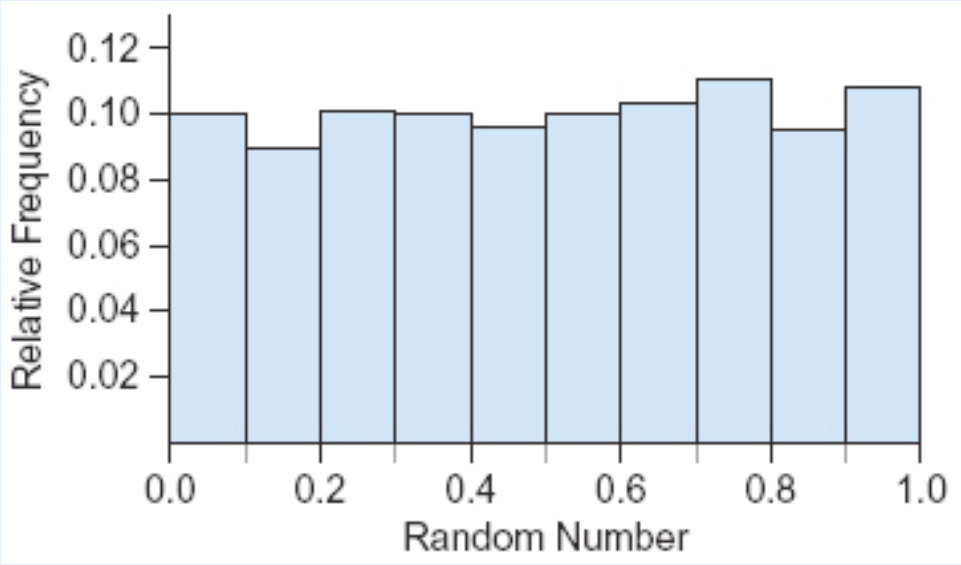
Histograms



A histogram is like a ‘coarse grained’ dot plot
‘bins’ on the x-axis
‘frequency’ on the y-axis

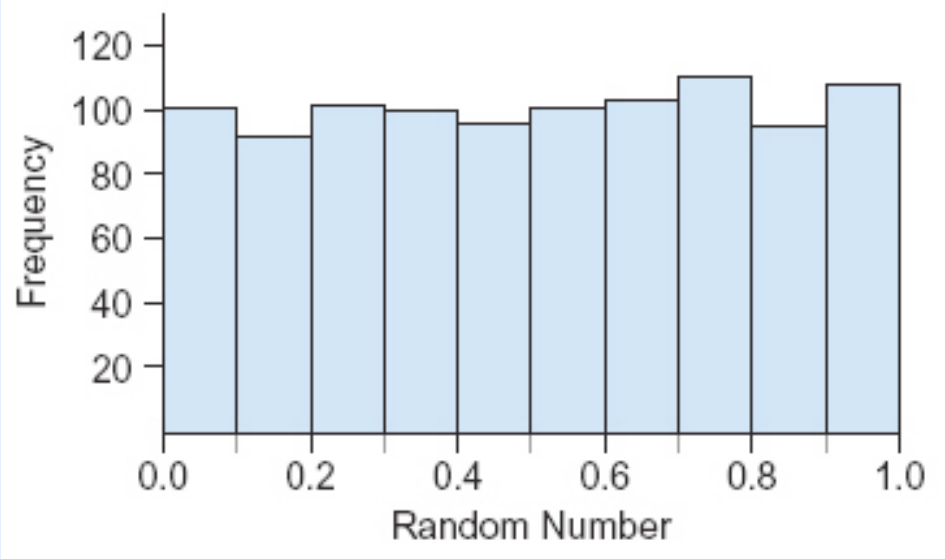
We can choose bin size any way we like

Relative Frequency

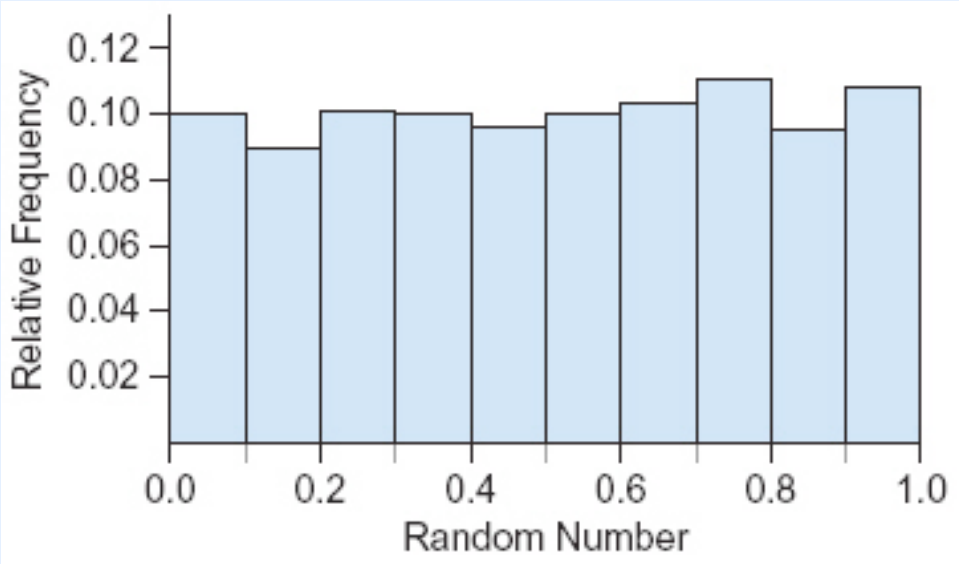


The sum of all heights is one

Frequency and Relative frequency

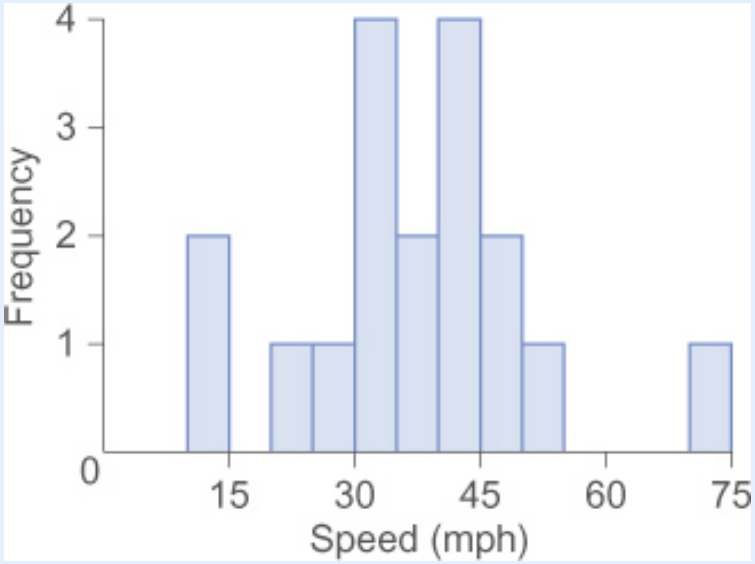


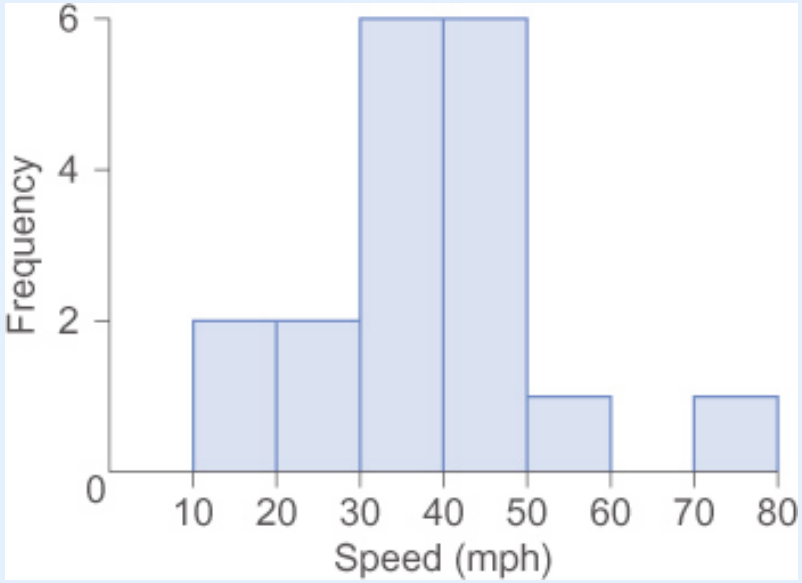
actual occurences



percent of total
(in this case
divide by 1000)

Different bin choices





Speed of mammal species
Using two bar widths

THERE IS NO RIGHT OR WRONG

Histograms work best when

Large number of values to plot

Don’t need to see individual values exactly

Don’t want to see exact shape of distribution

Have one distribution to look at

Use a calculator or computer

3. Stemplots

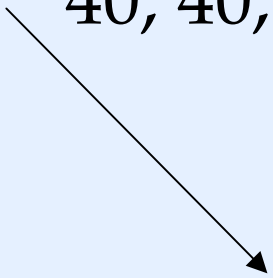
Speeds of mammals (mph)

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

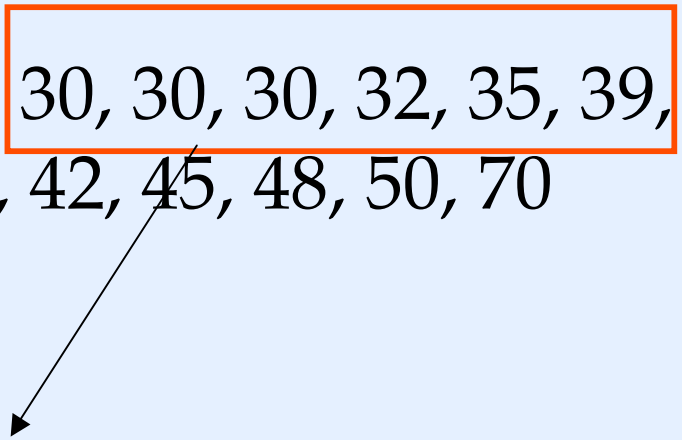


1 | 1 2

3. Stemplots

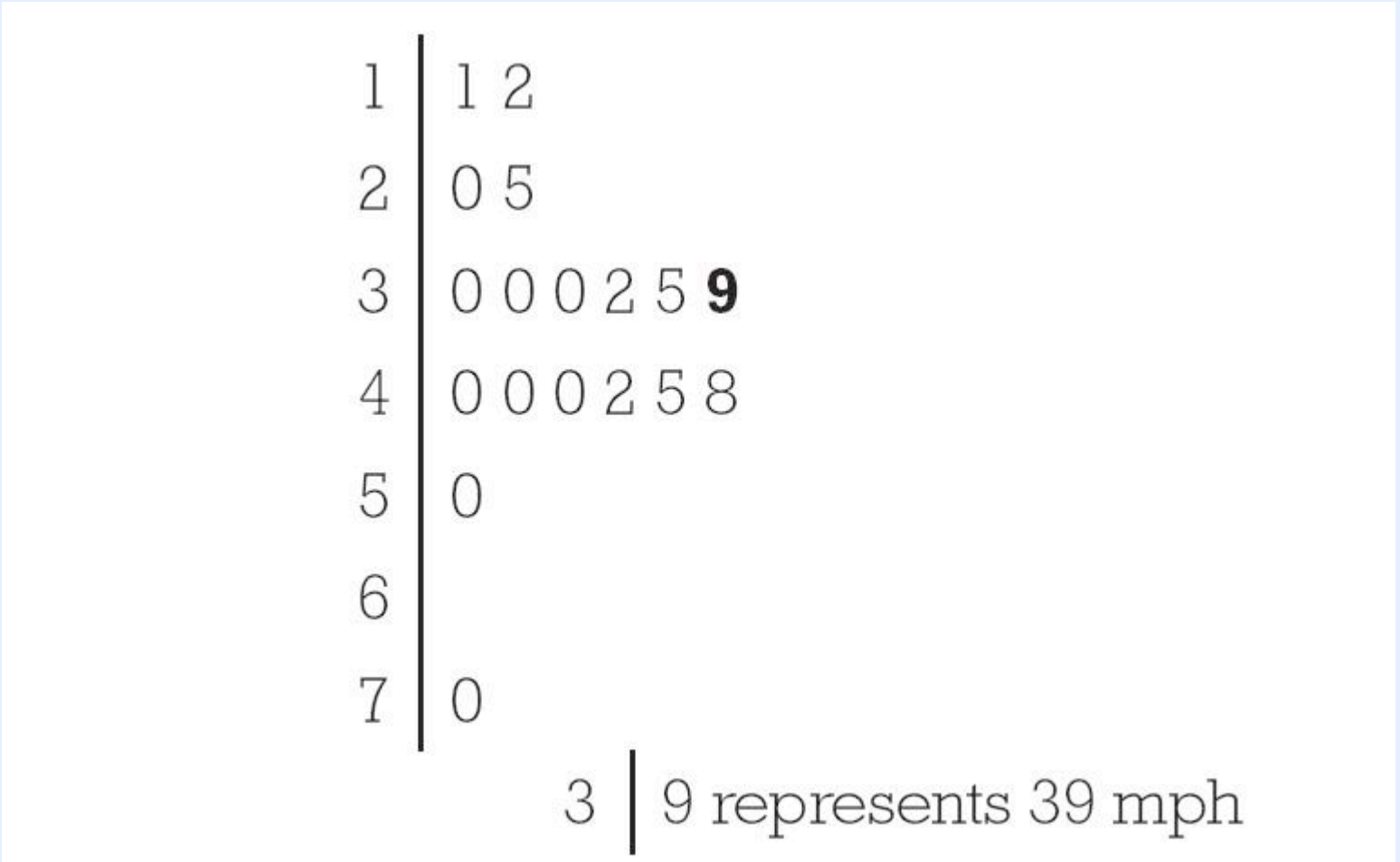
Speeds of mammals (mph)

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



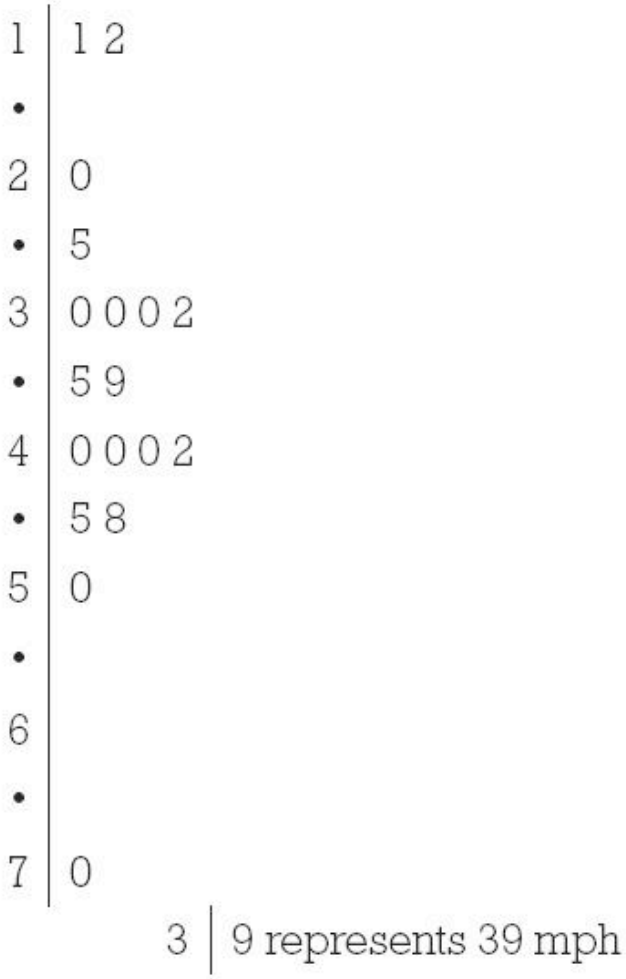
3. Stemplots

Or stem-and-leaf plots

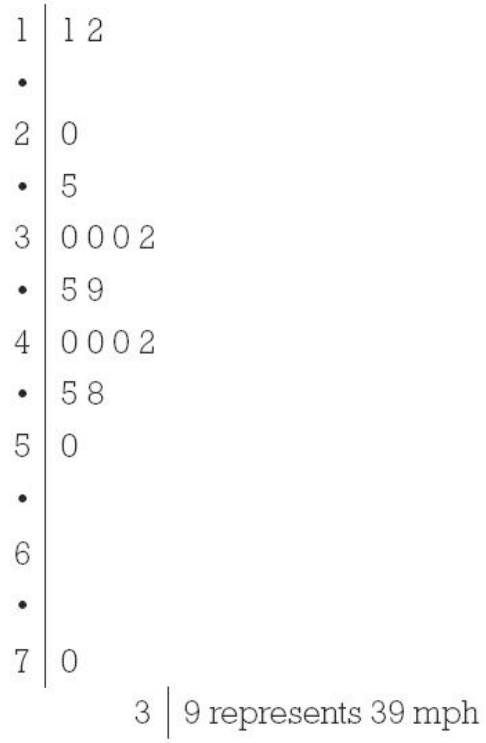
Numbers on the left are called stems
(the first digits of the data value)

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

Split stemplots



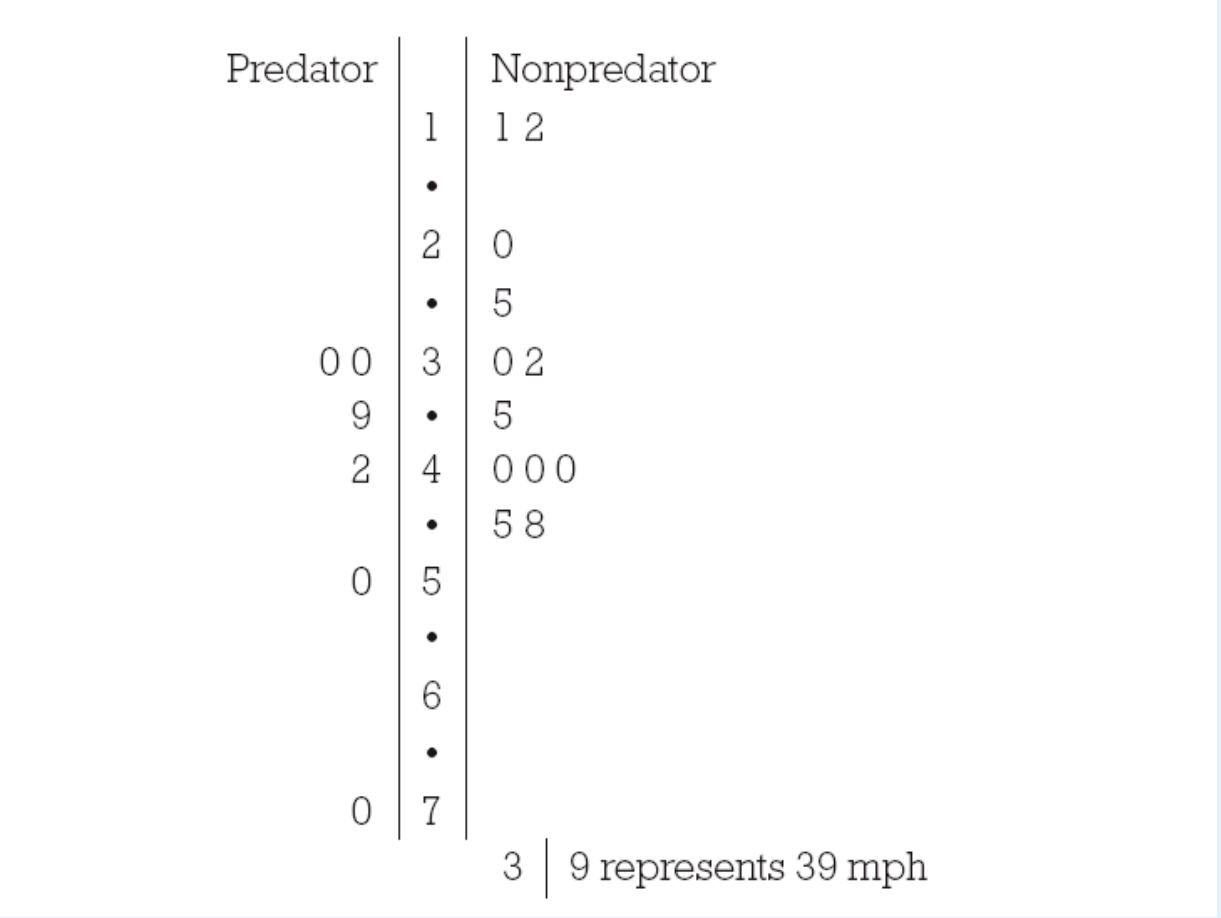
Split stemplots



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



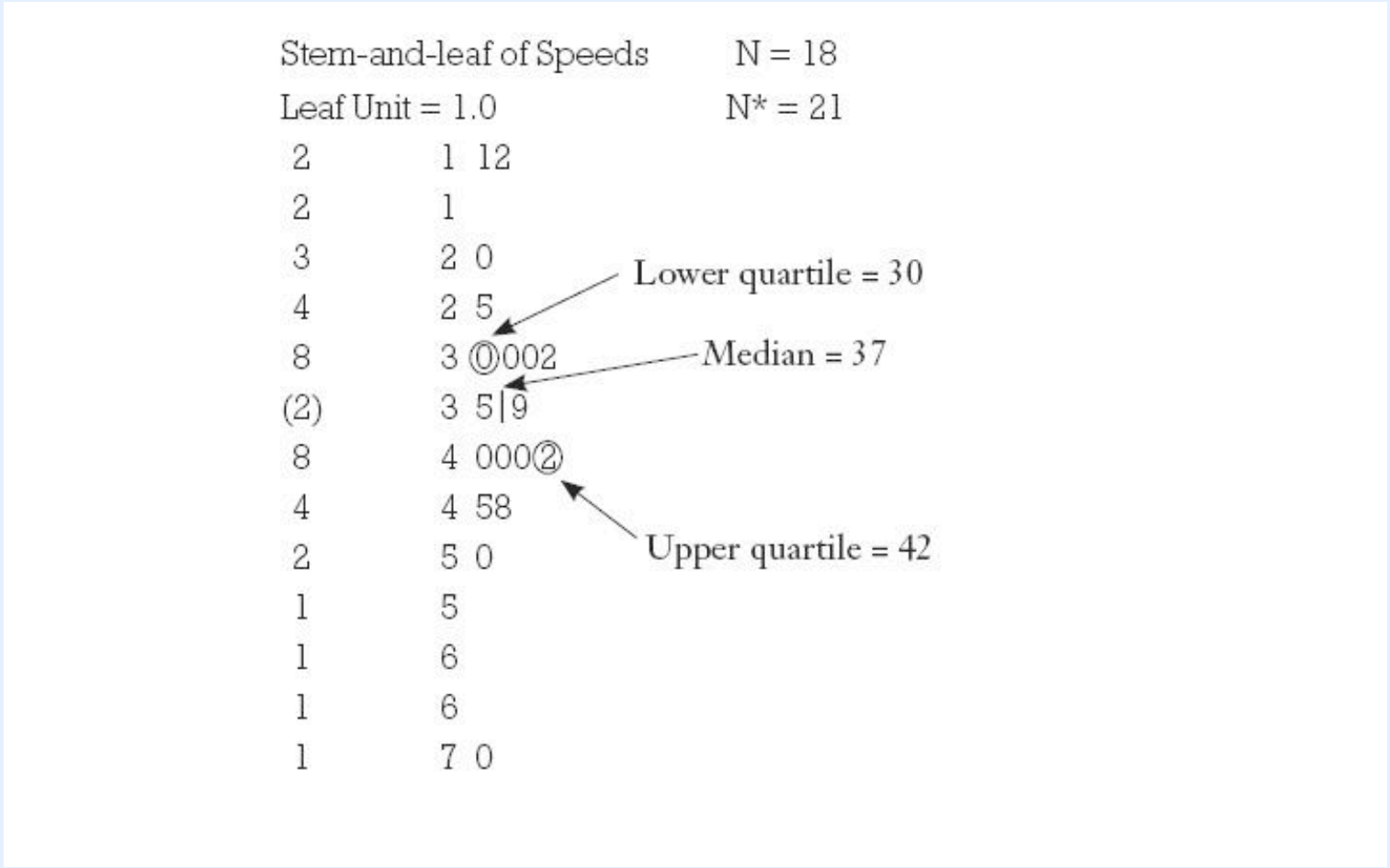
The data is differentiated on whether the mammals are predators or non-predators

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

Hk

Page 45 E16, E17 a/b, E13, E14