Introduction to Statistics

Individuals – objects described by data

Variables – characteristic of individual

When we examine a data set we ask the following questions:

1. Who are the individuals described by the data and how many are there?
2. What are the variables and in what units is each variable recorded?
3. When was the data recorded?
4. Where was the data recorded?
5. How was the data recorded?

Categorical Variables – categories!

Quantitative Variables – numerical (measurable) data

Distribution – tells us what values the variable takes and how often

Inference – ← inferring things about the population based on our sample

Example 1: CensusAtSchool is an international project that collects data about primary and secondary school students using surveys. Hundreds of thousands of students from Australia, Canada, New Zealand, South Africa, and the United Kingdom have taken part in the project since 2000. We used the website's "Random Data Selector" to choose 10 Canadian students who completed the survey in a recent year. The table displays the data.

Province	Gender	Languages spoken	Handed	Height (cm)	Wrist circum. (mm)	Preferred communication	Travel to school (min)
Ontario	Male	1	Right	175	175	Internet chat or MSN	25
Alberta	Female	3	Right	147	140	MySpace/Facebook	20
Ontario	Male	1	Right	165	170	Internet chat	4
British Columbia	Female	1	Right	155	145	In person	10
New Brunswick	Male	9	Left	130.5	130	Other	40
Ontario	Male	2	Right	170	165	In person	7
Ontario	Male	3	Left	150	100	Internet chat	10
New Brunswick	Male	2	Both	167.5	220	Internet chat	30
Ontario	Female	1	Right	161	104	Text messaging	10
Ontario	Male	6	Right	190.5	180	Internet chat	10

a) Who are the individuals in this data set?

Provinces

b) What variables were measured? Identify each as categorical or quantitative.
In what units were the quantitative variables measured?

Gender
languages
Handed
prefered comm.
} categorical

height
wrist cir.
travel
} quantitative

Example 2: 7 of the 10 students sampled are right-handed. Can we conclude that 70% of the population of Canadian students who participated in the CensusAtSchool are also right-handed? Explain.

no.

1.1 Analyzing Categorical Data

Frequency Table – a table that displays Counts

Relative Frequency Distribution – a table that displays percentages

We can "pile" the data by counting the number of data values in each category of interest. We can organize these counts into a frequency table, which records the totals and the category names.

Class	Count
First	325
Second	285
Third	706
Crew	885
	2201

325/2201 = .1477
285/2201 = .1295
706/2201 = .3208
885/2201 = .4021
1.000

A ***frequency table*** *of the Titanic passengers*

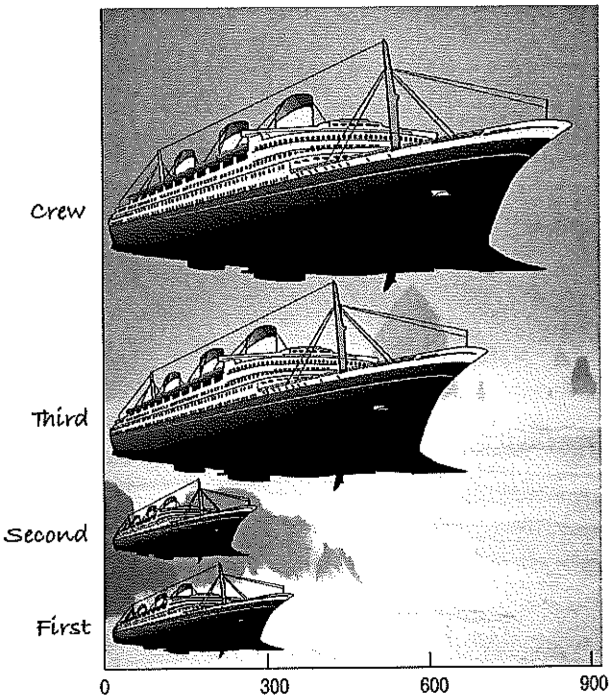
(counts)

A ***relative frequency*** *table of the Titanic passengers*

(percents)

What's Wrong With This Picture?

You might think that a good way to show the Titanic data is with this display:



This violates the area principle. The ship area proportions don't match. It looks like there's about 4x as many crew as 1st class but that's not right.

Bar Charts

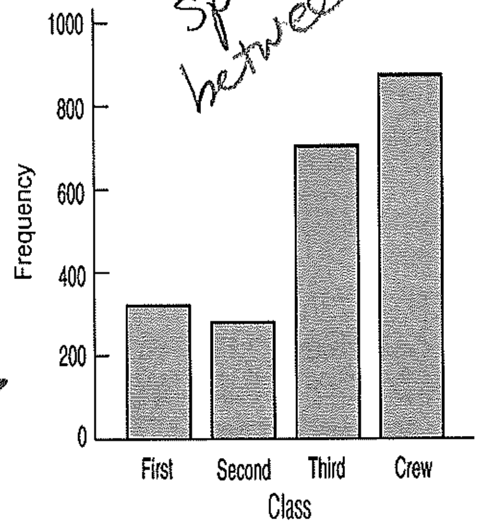
A bar chart is often used to display categorical data. The height of each bar represents the **COUNTS** for each category. Bars are displayed next to each other for easy comparison. When constructing a bar chart, note that the bars do not touch one another.

Categorical variables usually cannot be ordered in a meaningful way; therefore the order in which the bars are displayed is often meaningless.

This bar chart stays true to the area principle. Thus, a better display for the ship data is:

Counts

spaces in between!



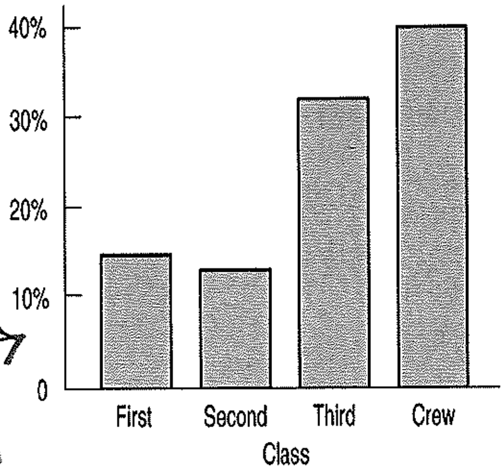
Relative Frequency Bar Chart

A relative frequency bar chart displays the relative frequency of counts for each category.

A relative frequency bar chart also stays true to the area principle.

Replacing counts with percentages in the ship data:
The sum of the relative frequencies is 100%

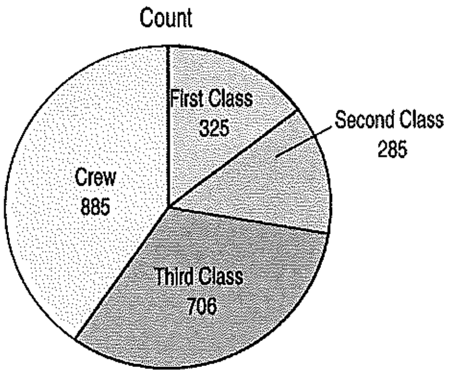
percentages



A pie chart is another type of display used to show categorical data. Pie charts show parts of a whole. Pie charts are often difficult to construct by hand.

Pie charts show the whole group of cases as a circle.

They slice the circle into pieces whose size is proportional to the fraction of the whole in each category.



A 2-way-table ______ shows two categorical variables together. The margins give the frequency distributions for each of the variables, also called the marginal distributions

It shows how individuals are distributed along each variable, contingent on the value of the other variable.

Example: we can examine the class of ticket and whether a person survived the Titanic:

		Class				
		First	Second	Third	Crew	Total
Survival	Alive	203	118	178	212	711
	Dead	122	167	528	673	1490
	Total	325	285	706	885	2201

Marginal Distribution vs. Conditional Distributions:

The marginal distribution of Survival is... — look at margins. You can ignore class and only focus on survival status.

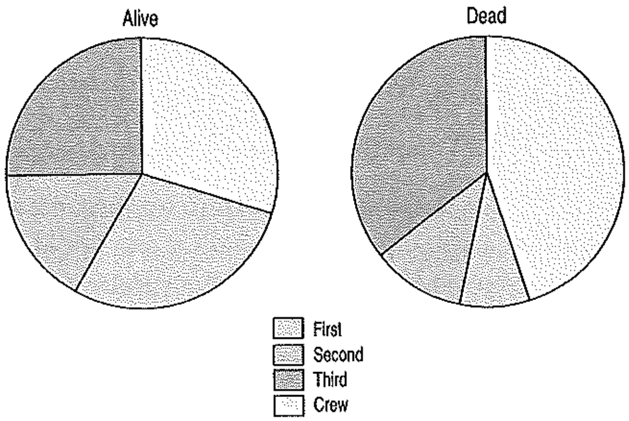
		Class				
		First	Second	Third	Crew	Total
Survival	Alive	~~203~~	118	~~178~~	212	711
	Dead	122	~~167~~	~~528~~	673	1490
	Total	~~325~~	285	706	885	2201

The conditional distribution of ticket Class, conditional on having perished... — focus on "those having perished" and ignore rest. conditional dist. always in middle of table.

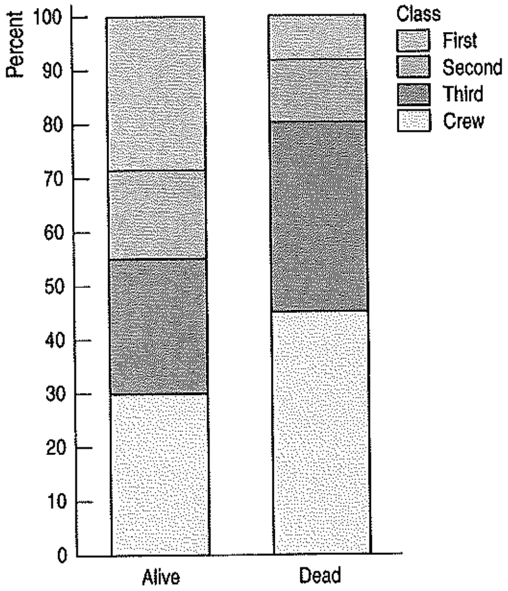
		Class				
		First	Second	Third	Crew	Total
Survival	Alive	203	~~118~~	~~178~~	212	711
	Dead	122	167	528	673	1490
	Total	325	~~285~~	~~706~~	~~885~~	2201

The conditional distributions tell us that there is a difference in class for those who survived and those who perished.

This is better shown with pie charts of the distributions:



Segmented Bar Charts



Is there an **association**??

do the marginal distributions match the conditional distributions?

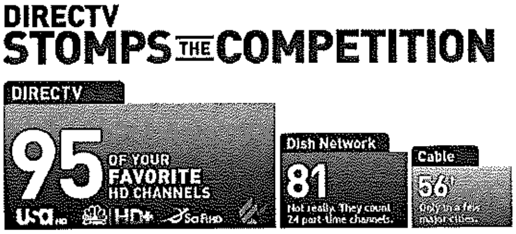
~~do~~ proportion survived = proportion of each class survived?

Are they **independent**??

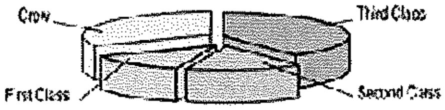
if 2 variables are independent, there's no association.

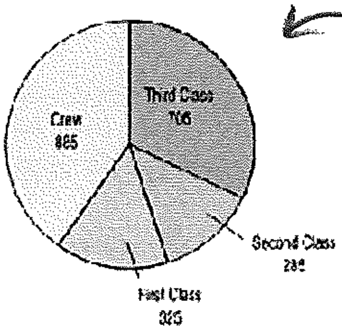
class and survival aren't independent (there is an association) because the proportion of 1st class passengers was .1477 but the proportion of 1st class who survived was 203/325 = .6246 which are very different.

What is wrong with the following graphs?

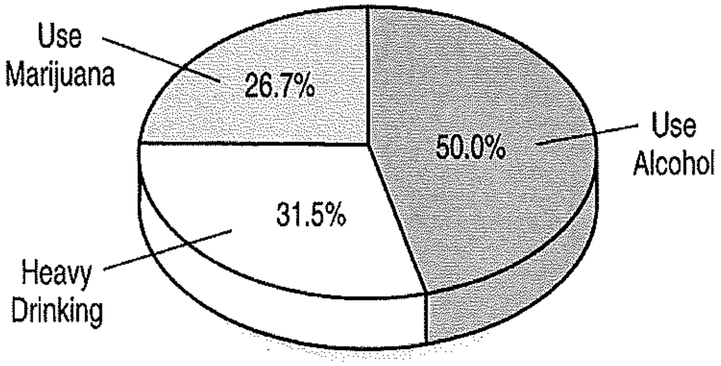


violates area principle!





ok



these don't add up to 100%

Example 3: A survey of 4826 randomly selected young adults (aged 19 to 25) asked, "What do you think are the chances you will have much more than a middle-class income at age 30?"

Young adults by gender and chance of getting rich

	Gender		
Opinion	**Female**	**Male**	**Total**
Almost no chance	96	98	194
Some chance but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

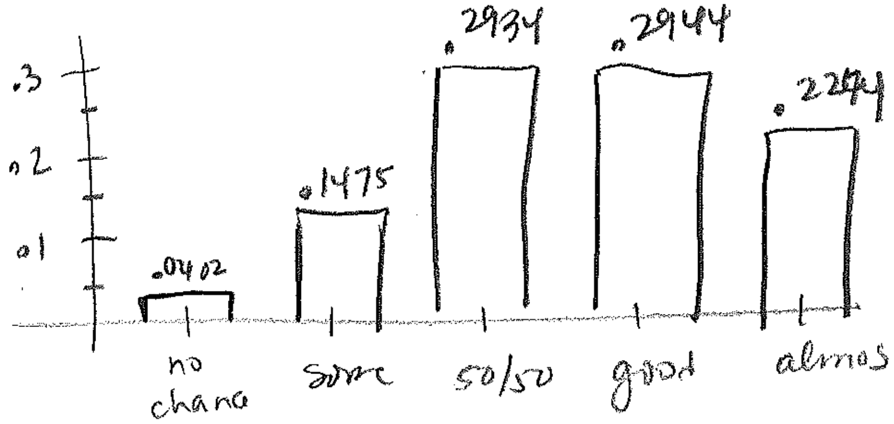
a) Calculate the marginal distribution (in percents) of opinions and make a table of the data.

ignore male/female

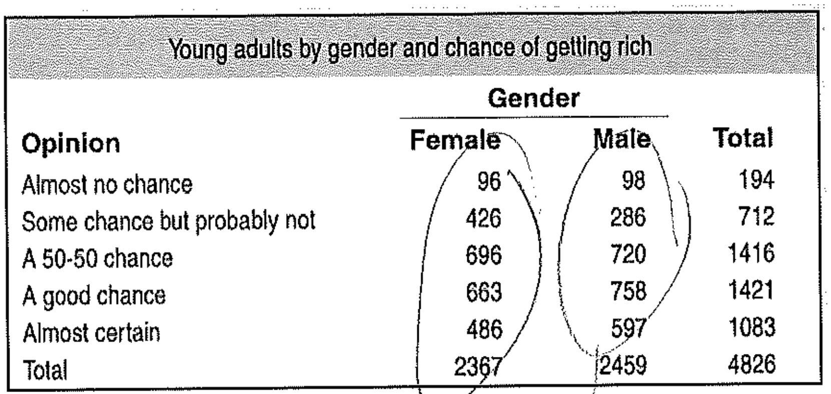
194/4826 = .0402 no chance
712/4826 = .1475 some but prob not
1416/4826 = .2934 50-50
1421/4826 = .2944 good
1083/4826 = .2244 almost certain

.9999

b) Create a graph of the distribution.



c) Calculate the conditional distribution of opinion among women.



Young adults by gender and chance of getting rich

	Gender		
Opinion	Female	Male	Total
Almost no chance	96	98	194
Some chance but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

no 96/ 2367 = .0406

some 426/ 2367 = .1800

50/50 696/2367 = .2940

good 663/2367 = .2801

certain 486/2367 = .2053

d) Calculate the conditional distribution of opinion among men.

no 98/ 2459 = .0399

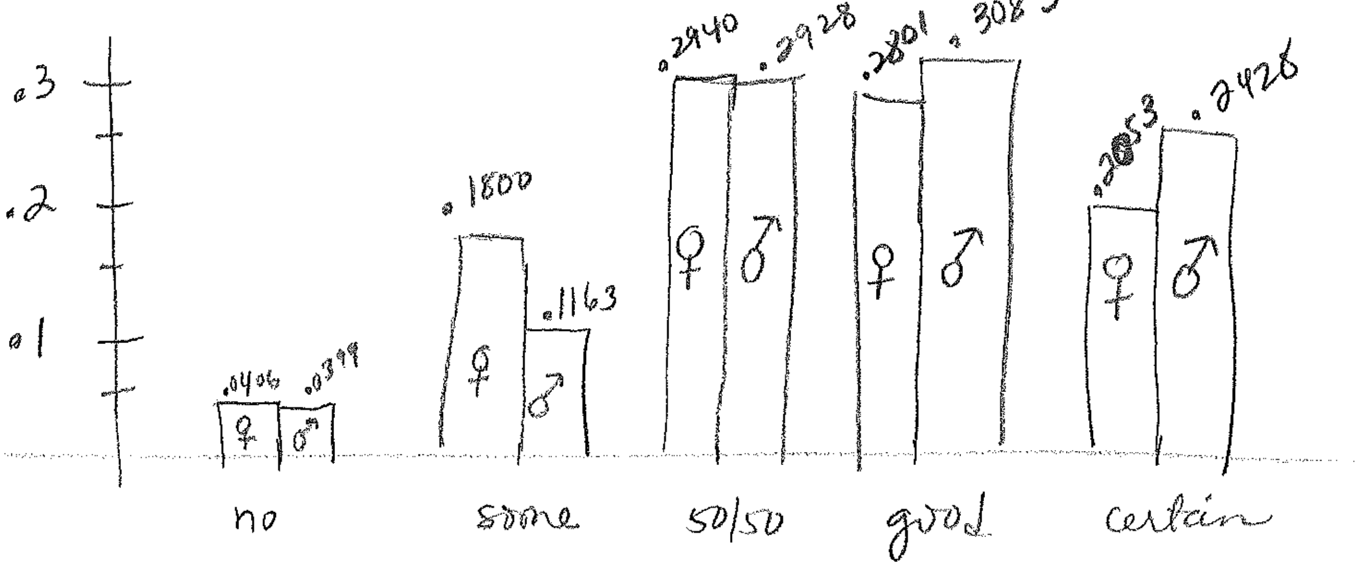
some 286 /2459 = .1163

50/50 720 /2459 = .2928

good 758 /2459 = .3083

certain 597/2459 = .2428

e) Compare the conditional distributions of men and women.



Simpson's Paradox – an association between two variables that holds for each individual value of a third variable can be changed or even reversed when the data for all values of the third variable are combined. ~~This reversal is called~~ ____________________.

Example 5: Do helicopters save lives? Accident victims are sometimes taken by helicopter from the accident scene to a hospital. Helicopter's save time. Do they also save lives? Let's compare the percent of accident victims who die with helicopter evacuation and with the usual transport to a hospital by road.

	Helicopter	Road
Victim died	64	260
Victim survived	136	840
Total	200	1100

a) What percent of helicopter patients die? What percent of road patients die?

H $\frac{64}{200} = .32$ R $\frac{260}{1100} = .2364$

higher % of Helicopter patients died

b) Here are the same data broken down by the seriousness of the accident. For both types of accidents, what percent of helicopter patients die? What percent of road patients die?

Serious Accidents

	Helicopter	Road
Died	48	60
Survived	52	40
Total	100	100

Less Serious Accidents

	Helicopter	Road
Died	16	200
Survived	84	800
Total	100	1000

H $\frac{48}{100} = .48$ R $\frac{60}{100} = .6$

higher % of road P died

H $\frac{16}{100} = .16$ R $\frac{200}{1000} = .2$

higher % of road P. died

Example (Simpson's Paradox continued): Two companies have labor and management classifications of employees. Company A's laborers have a higher averages salary than Company B's, as do Company As managers. But overall Company B pays a higher average salary. How can that be? And which is the better way to compare earning potential at the two companies?

	Co. A	Co. B
labor.	higher	lower
man.	higher	lower
overall	lower	higher

Picture This:

Co. A	Co. A
100 people making $ 10/hr	50 people making $9/hr
2 people making $50/hr	50 people making $45/hr

Alternate Example: Cell Phones

The Pew Research Center asked a random sample of 2024 adult cell phone owners from the United States which type of cell phone they own: iPhone, Android, or other (including non-smart phones). Here are the results, broken down by age category. Explain what it would mean if there was no association between age and cell phone type.

	18–34	35–54	55+	Total
iPhone	169	171	127	467
Android	214	189	100	503
Other	134	277	643	1054
Total	517	637	870	2024

If there were no association, then the % owning each phone brand in each age category would be close to the overall % owning each phone.

Alternate Example:

The following partially complete two-way table shows the marginal distribution of age and ice cream flavor preference. If there is no associate between age and flavor preference for the members of the sample, which of the following is the correct value of x?

	Chocolate	Vanilla	Total
Children	41	19	60
Adults		x	100
			160

~~since 60/160 = 37.5% of those are children, then~~ ~~$\frac{41}{41+x}$~~

~~since 100/160 = 62.5% of those are adults, then 62.5% of those who like vanilla should be adults~~

~~$\frac{x}{19+x}$~~

if 19/60 = 31.67% of kids like vanilla, 31.67% of adults should, so

$\boxed{x \approx 32 \text{ adults}}$

1.2 Displaying Quantitative Data with Graphs

How to Examine the Distribution of a Quantitative Variable

↑

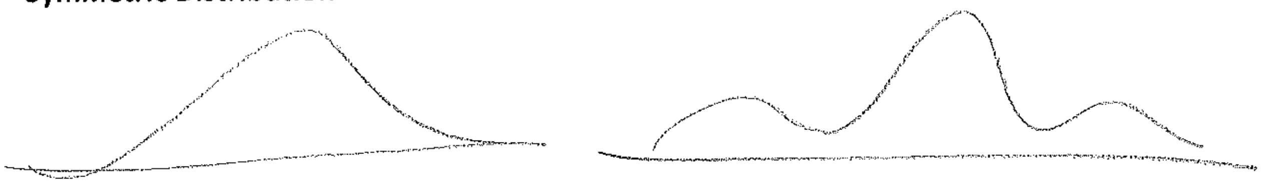
measurable

numerical values

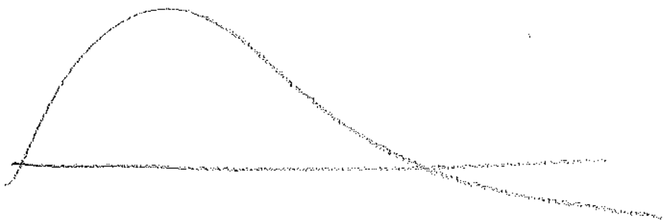
Shape

Example 1: Draw an example of each distribution type.

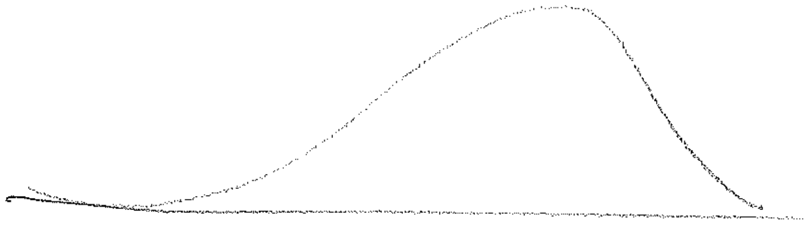
Symmetric Distribution –



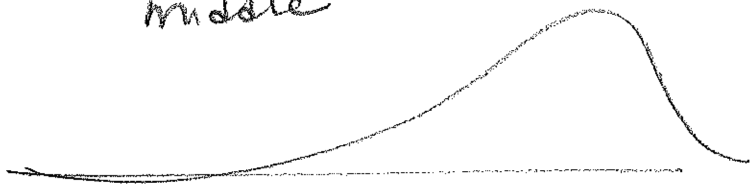
Skewed Right Distribution –



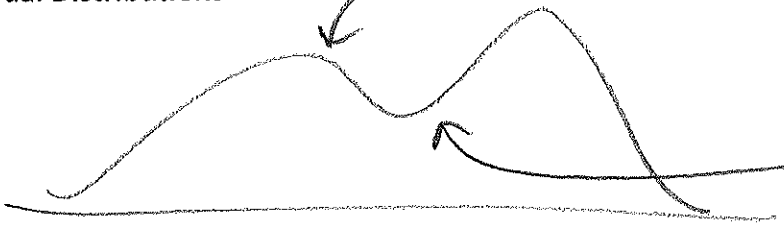
Skewed Left Distribution



Unimodal Distribution – don't assume one mode in middle

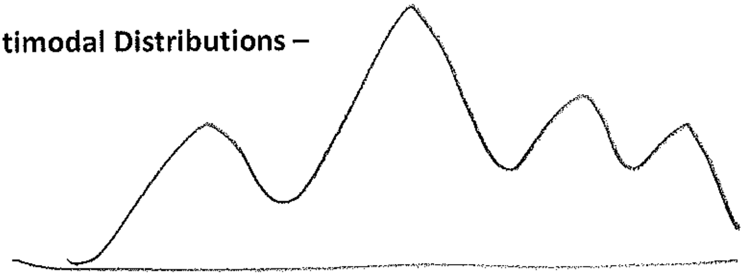

mode/peak

Bimodal Distributions –

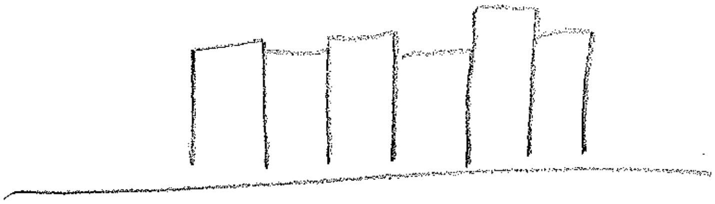

two modes/peaks

gap?

Multimodal Distributions –



Uniform Distributions –


roughly flat

"what day of the week were you born?"

To describe a univariate data distribution:

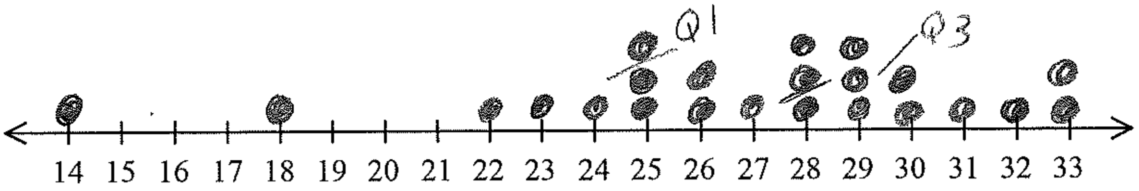
one variable

S – shape
O – outliers, if any
C – center (mean or median)
C – CONTEXT
S – spread/variability (IQR, range, standard deviation)

Example 2: The Environmental Protection Agency (EPA) is in charge of determining and reporting fuel economy ratings for cars. The table below displays the EPA estimates of highway gas mileage in miles per gallon (mpg) for a sample of 24 model year 2009 midsize cars.

Model	Mpg	Model	Mpg	Model	Mpg
Acura RL	22	Dodge Avenger	30	Mercury Milan	29
Audi A6 Quattro	23	Hyundai Elantra	33	Mitsubishi Galant	27
Bentley Arnage	14	Jaguar XF	25	Nissan Maxima	26
BMW 528I	28	Kia Optima	32	Rolls Royce Phantom	18
Buick Lacrosse	28	Lexus GS 350	26	Saturn Aura	33
Cadillac CTS	25	Lincoln MKZ	28	Toyota Camry	31
Chevrolet Malibu	33	Mazda 6	29	Volkswagen Passat	29
Chrysler Sebring	30	Mercedes-Benz E350	24	Volvo S80	25

a) Construct a dotplot of the data.

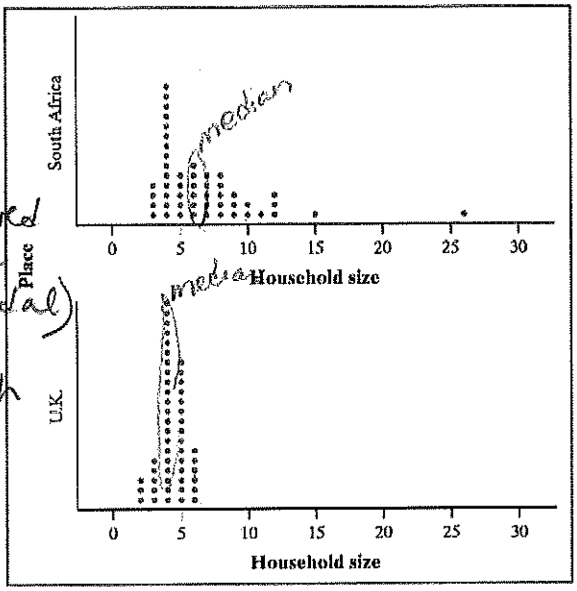


b) Describe the distribution. Are there any outliers?

- the fuel economy rating appears skewed left.
- the center appears to be at 28 (median).
- two cars have unusually low ratings at 14, 18.
- the IQR is 30-25 =5.

DIRECT COMPARE

Example 4: How do the number of people living in households in the United Kingdom (U.K.) and South Africa compare? We selected 50 random households from each country using the CensusAtSchool database. Use the dotplots to compare the distributions.



shape - South africa appears more skewed right than UK which is more normal-looking (symmetric/unimodal)

center - the center of South africa appears higher at 6 than UK which has a median at 4.

outlier - South africa has one or 2 possible unusually high points at 15, 26 whereas UK doesn't.

spread/variability - south africa has a range of 26-3=23 which is much higher than the range of UK household size which is only 4 people.

context ☺

Stemplots – stemplots are simple graphical displays for fairly small data sets.

- If a stemplot has too much data concentrated in one area you can also split stems.
- Stemplots do not work well with large data sets, but five is a good minimum.
- There is no magic number of stems to use.
- Rounding data and using the rounded digit as a leaf is acceptable.

Example 5: How many pairs of shoes does the typical teenager have? Let's sample this class and construct a stemplot. Describe the distribution once the stemplot is complete.

let's not.

Alternate Example: Which gender is taller, males or females? A sample of 14-year-olds from the United Kingdom was randomly selected using the CensusAtSchool website. Here are the heights of the students (in cm). Make a back-to-back stemplot and compare the distributions.

Male: 154, 157, 187, 163, 167, 159, 169, 162, 176, 177, 151, 175, 174, 165, 165, 183, 180 154 - 180
Female: 160, 169, 152, 167, 164, 163, 160, 163, 169, 157, 158, 153, 161, 165, 165, 159, 168, 152 -
153, 166, 158, 158, 166

♀		♂
3 3 2	15	4, 1
8 8 9 8 7	15	7, 9
1 3 0 3 4 0	16	3, 2
6 6 8 5 5 9 7 9	16	7, 9, 5, 5
	17	4,
	17	6, 7, 5
	18	3, 0
	18	7

Key

15|4 = 154 male

4|15| = 154 female

Shape - females seem unimodal, skewed left where males looks bimodal.

Center - median female height is around 161 which is lower than male height which is about 165.

spread - female height range is much smaller than males at 17 compared to 36.

Example 6: The table gives the percent of residents in each state born outside of the United States. Construct a histogram of the data.

i) Divide the range of the data into classes of equal width.

ii) Find the count (frequency) or percent (relative frequency) of individuals in each class.

iii) Label and scale your axes and draw the histogram.

on calculator

State	Percent	State	Percent	State	Percent
Alabama	2.8	Louisiana	2.9	Ohio	3.6
Alaska	7.0	Maine	3.2	Oklahoma	4.9
Arizona	15.1	Maryland	12.2	Oregon	9.7
Arkansas	3.8	Massachusetts	14.1	Pennsylvania	5.1
California	27.2	Michigan	5.9	Rhode Island	12.6
Colorado	10.3	Minnesota	6.6	South Carolina	4.1
Connecticut	12.9	Mississippi	1.8	South Dakota	2.2
Delaware	8.1	Missouri	3.3	Tennessee	3.9
Florida	18.9	Montana	1.9	Texas	15.9
Georgia	9.2	Nebraska	5.6	Utah	8.3
Hawaii	16.3	Nevada	19.1	Vermont	3.9
Idaho	5.6	New Hampshire	5.4	Virginia	10.1
Illinois	13.8	New Jersey	20.1	Washington	12.4
Indiana	4.2	New Mexico	10.1	West Virginia	1.2
Iowa	3.8	New York	21.6	Wisconsin	4.4
Kansas	6.3	North Carolina	6.9	Wyoming	2.7
Kentucky	2.7	North Dakota	2.1		

Frequency table

Class	Count
0 to < 5	
5 to < 10	
10 to < 15	
15 to < 20	
20 to < 25	
25 to < 30	
Total	

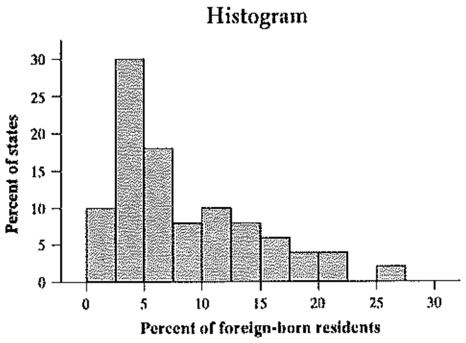
Relative frequency table

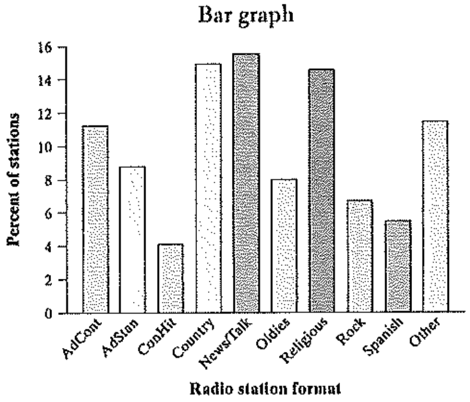
Class	Percent
0 to < 5	
5 to < 10	
10 to < 15	
15 to < 20	
20 to < 25	
25 to < 30	
Total	

Histograms vs. Bar Charts

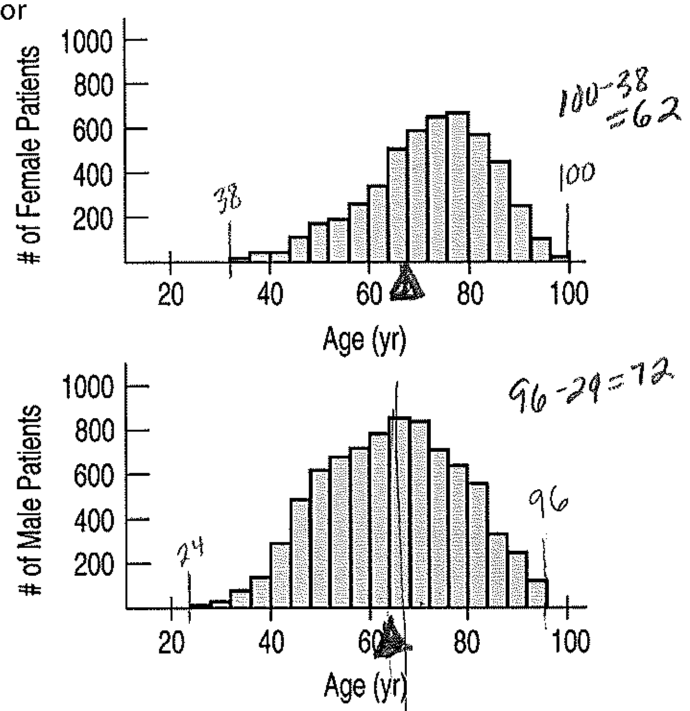
quantitative
no spaces

categorical
always spaces





Example 2: Compare the following distributions of ages for female and male heart attack patients.



@ shape - the ages of male patients appear more symmetric than female patients which appear skewed left.

center - the center of both appear to be similar, in the low 60's.

spread / variability - the range of male patients' ages was higher at 72 compared to 62 for female patients ages.

%'s counts

Why would we prefer a relative frequency histogram to a frequency histogram?

easier to compare w/ other data sets

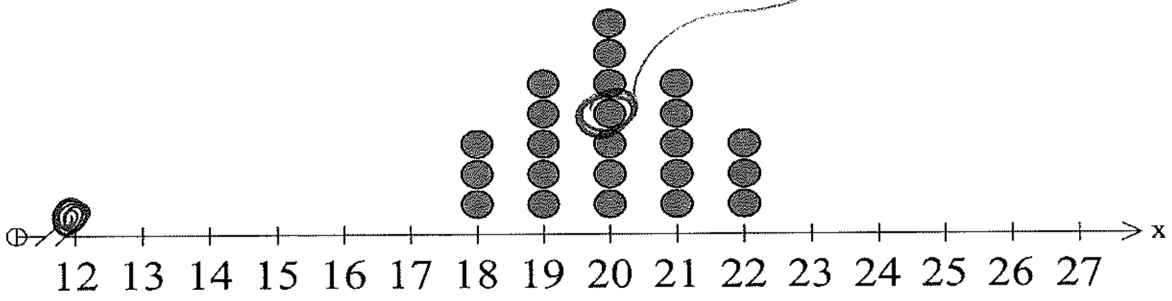
What will cause you to lose points on tests and projects (and turn the rest of my hair gray)?

- no key on ~~gap~~ stem/leaf
- no labels on axis

1.3 Describing Quantitative Data with Numbers

20 ↓ 20 ↓ 20 ↓

Here are the ages of students in a college course. Mark the median, mean, and mode.



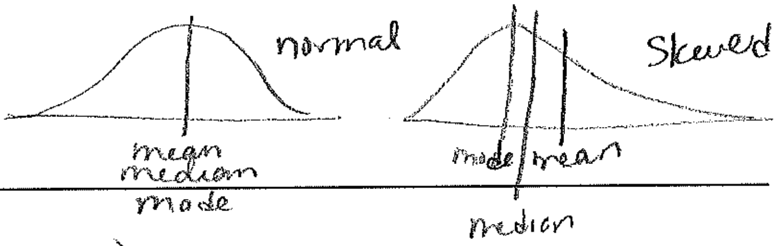
Let's say a 12-year old child genius is added to the course. Now mark the median, mean, and mode.

What has happened to these values?

median moved from here → to here → so still 20.

mean is now $[12 + 18(3) + 19(5) + 20(7) + 21(5) + 22(3)]/24 = 19.\overline{6}$

mode the same.



What is the difference between $\bar{x}$ and μ ?

$\bar{x}$ is sample (statistic) mean

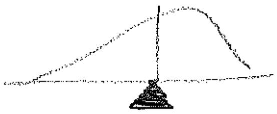
μ is population mean (parameter)

What is a resistant measure? Is the mean a resistant measure of center?

median is more Resistant to outliers than mean.

How can you estimate the mean of a histogram or dotplot?

balancing point



Is the median a resistant measure of center? Explain.

moreso than mean

How does the shape of a distribution affect the relationship between the mean and the median?

mean is skewed toward skew

median not so much

What is the range? Is it a resistant measure of spread? Explain.

no!! super susceptible to outliers
so we prefer IQR

What are quartiles? How do you find them?

divide data in ~~4th~~
quarters

What is the interquartile range (IQR)? Is the IQR a resistant measure of spread?

middle half of data

Example 2: Find the quartiles of each set of data. Then find the IQR.

a) The times it took 15 people in North Carolina to commute to work:

5 10 10 10 10 12 15 20 20 25 30 30 40 40 60

Q1=10 median =20 Q3=30

b) The times it took 20 people in New York to commute to work:

0	5
1	005555
2	0005
3	00
4	005
5	
6	005
7	
8	5

Key: 4|5 is a New York worker who reported a 45-minute travel time to work.

Q1 = 15

$$\text{median} = \frac{20+25}{2} = 22.5$$

$$Q3 = \frac{40+45}{2} = 42.5$$

Outliers – we will call an observation an outlier if it falls more than 1.5(IQR) above the third quartile or below the first quartile. The numbers that mark an outlier we will call unusual.?

Example 3: Examine the data from example 2 again. Are there any outliers?

a) 5 10 10 10 10 12 15 20 20 25 30 30 40 40 60

min=5 Q1=10 med =20 Q3=30 max=60

$Q1-1.5(Q3-Q1)$

left fence = $10-1.5(30-10) = -20$

right fence $= Q3+1.5(Q3-Q1)$

$= 30+1.5(30-10) = 60$

no outliers

b)

~~0~~	~~5~~
~~1~~	~~005555~~
~~2~~	~~0005~~
~~3~~	~~00~~
~~4~~	~~005~~
~~5~~	
~~6~~	~~005~~
~~7~~	
~~8~~	~~5~~

~~Key: 4|5 is a New York worker who reported a 45-minute travel time to work.~~

5-number summary

Create a 5-number summary from the data on the right, which represents the number of minutes people waited for their doctor when arriving on time for an appointment:

6	28
5	0468
4	0346
3	334446678
2	289
1	8

max=68

Q3=50

med=37

Q1=33

min=18

Min: 18
Q1:
Median:
Q3:
Max: 68

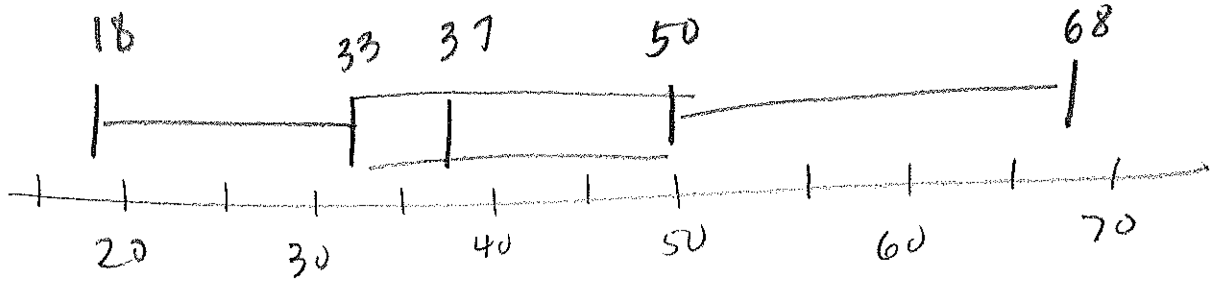
Find the range: $68-18=50$

Find the Interquartile Range: $50-33=17$

right fence $= Q3+1.5(Q3-Q1) = 50+1.5(17) = 75.5$

left fence $= Q1-1.5(Q3-Q1) = 33-1.5(17) = 7.5$

Now create a box plot from the 5 number-summary:



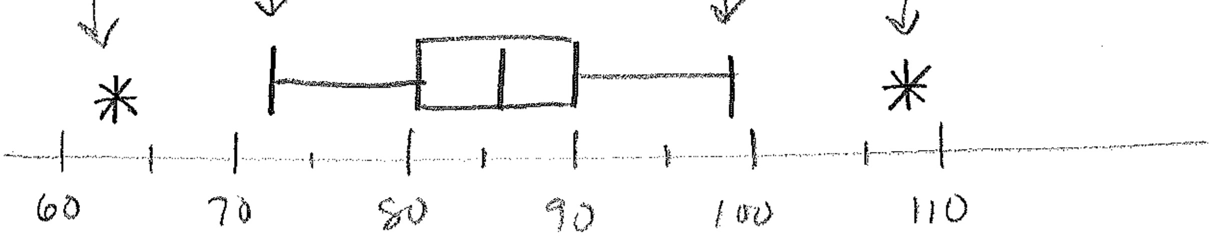
Example: Here are some Algebra 2 Honors test scores. Find the five-number summary, identify any outliers, and create a box plot. Denote the outliers with asterisks.

63, 72, 75, 78, 80, 81, 85, 86, 86, 86, 87, 87, 88, 90, 95, 96, 98, 108 max

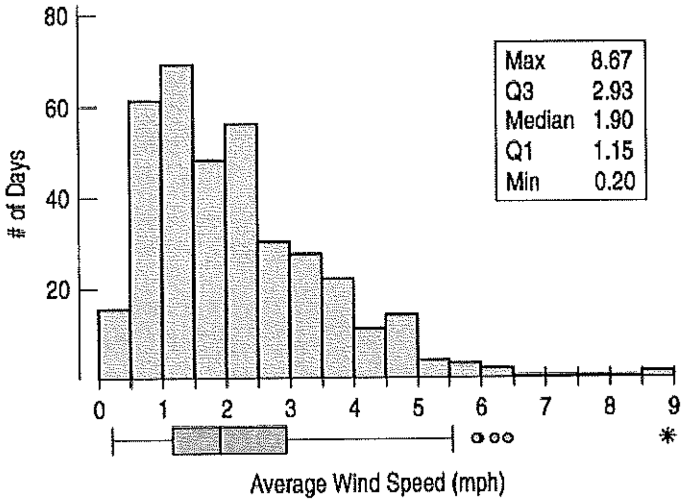
min Q1 Med Q3

left = 90 + 1.5(90-80) = 105

right = 80 - 1.5(90-80) = 65

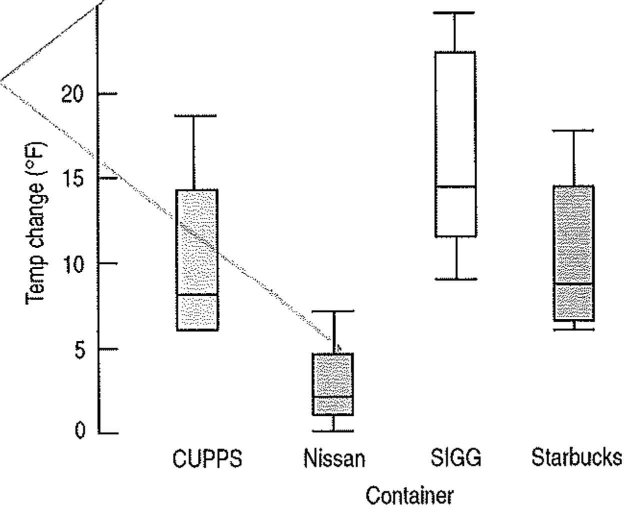


Example: A five-number summary of average wind speeds (mph) for the Hopkins Forest of western Massachusetts in 1989 is given. The data was taken every day for a full year. Compare the boxplot and histogram displaying the same data....



Example 3: For a class project, a student compared the efficiency of various coffee containers. For her study, she decided to try 4 different containers and to test each of them 8 different times. Each time, she heated water to 180°F, poured it into a container, and sealed it. After 30 minutes, she measured the temperature again and recorded the difference in temperature. Because these are temperature differences, smaller differences mean that the liquid stayed hot—just what we would want in a coffee mug. What can we say about the effectiveness of these four mugs?

	Min	Q1	Med	Q3	Max	IQR
CUPPS	6	6	8.25	14.25	18.50	8.25
Nissan	0	1	2	4.5	7	3.50
SIGG	9	11.50	14.25	21.75	24.50	10.25
Starbucks	6	6.5	8.50	14.25	17.50	7.75



Standard Deviation s_x - the standard deviation measures the average distance of the observations from their mean. It is calculated by finding an average of the squared distances and then taking the square root. The number you are taking the square root of is the variance.

Variance s_x^2 - the average squared distance is called the variance. It is the square of the standard deviation.

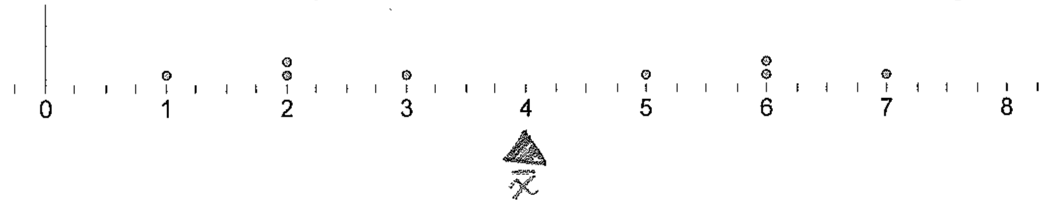
$$s_x^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1} = \frac{1}{n-1}\sum(x_i-\bar{x})^2$$

How to Find the Standard Deviation

1. Find the distance of each observation from the mean and square each of these distances.
2. Average distances by dividing their sum by n – 1.
3. The standard deviation s_x is the square root of this average squared distance:

In the distribution below, how far are the values from the mean, on average?

$\bar{x} = 4$



1−4		−3		9		
2−4	→	−2	→	4	→	
2−4		−2		4		
3−4		−1		1		
5−4	→	1	→	1	→	
6−4		2		4		
6−4		2		4		
7−4		3		9		

!! → 0

sum of $(\text{distances})^2$ is 1296

$$\to \frac{36}{8-1} = 5.142857$$

$$\downarrow$$

$$\sqrt{5.142857}$$

$$\downarrow$$

$$\boxed{2.268}$$

What does the standard deviation measure?

typical distance from $\bar{x}$.

What are some similarities and differences between the range, IQR, and standard deviation?

They all measure spread/variability.

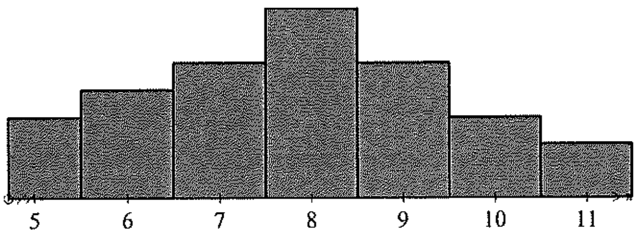
Example 5: A random sample of 9 children were asked how many pets they owned.
Here are the data: 1 3 4 4 4 5 7 8 9
Calculate the variance and standard deviation by hand.

Observations	Deviations	Squared deviations
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1		
3		
4		
4		
4		
5		
7		
8		
9		
	sum =	sum =

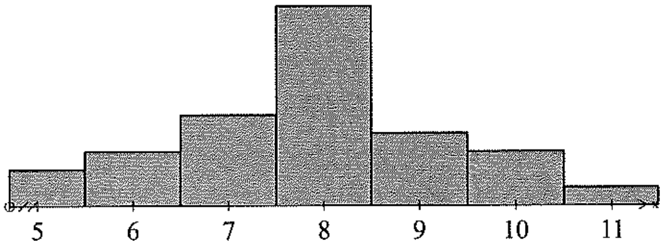
Good. Now we have done it by hand. Never again. From here on we will use technology (calculators).

Example: Compare the center and the range of the two classes:

Class A Quiz Scores



Class B Quiz Scores

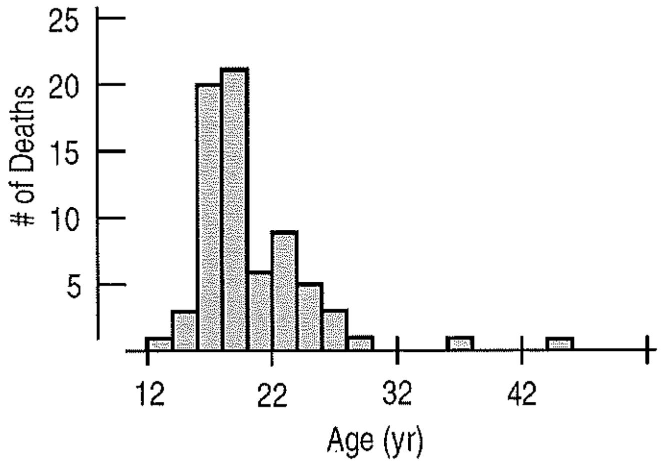


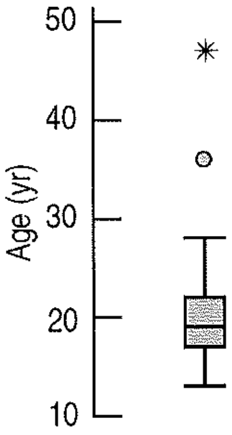
Which class would have a larger IQR? A - center more spread out

Which class shows less variability? B - typical distance from $\bar{x}$ smaller

Which class has a higher standard deviation? A

Example: Crowd Management Strategies monitors accidents at rock concerts. In their database, they list the names and other variables of victims whose deaths were attributed to "crowd crush" at rock concerts. Here are the histogram and boxplot of the victims' ages for data from 1999 to 2000:





a) What features of the distribution can you see in both the histogram and the boxplot? range "outliers" skew

b) What features of the distribution can you see in the histogram that you could not see in the boxplot? That it's bimodal

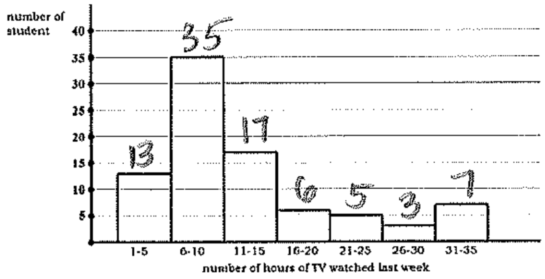
c) What summary statistic would you choose to summarize the center of this distribution? Why?

I would choose median because data isn't normal (symmetric/unimodal).

d) What summary statistic would you choose to summarize the spread of this distribution? Why?

I would choose IQR since there are outliers affecting range/SD

Example: In a survey, high school students were randomly selected and asked how many hours of television they had watched in the previous week. The histogram to the ~~right~~ left displays their answers.



a) Approximately how many students participated in the survey?

13+35+17+6+5+3+7 = 86

b) Describe the shape of the distribution.

skewed right

c) Approximately how many students watched 10 hours or less of TV last week? 35+13 = 48

d) Approximately how many students watched between 16 and 30 hours of TV last week?

14

e) In which category is the median of the data?

(look over to 43-44th student) $\boxed{6-10}$

f) Is it possible to calculate the mean of the data from a histogram?

no