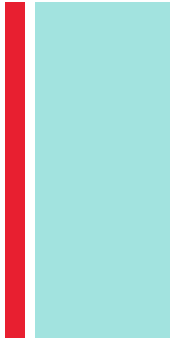




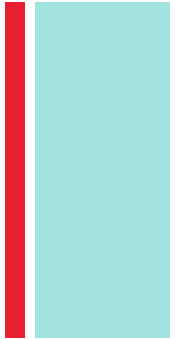
Today's AGENDA - Aug 2021



- TAKE out your **HW #1** & *Signed Syllabus* ready to check and to turn in
- Discuss Website and syllabus
- Review Expectations, & Norms
- Review August Calendar, Chap. 1 & HW
- Questions & Next Steps
- Aug 30th- **QUIZ #1** on intro concepts



General Class Reminders:



- Start on time, end on time
- Please **keep phones put away** unless we are using them for an activity. **Note:** Mr. L. will ask to take your phone if you're using it without permission.
- Warm-Ups: Whenever we have warm-ups, you are expected to write the problems (what are you trying to find?) and your solution, with work.
- Class Meetings \neq *Spectators sport*
- Questions, concerns?

+ Warm-UP

- 1) A _____ is a subset of the defined population. The characteristic or variable of a sample is called a _____.
- 2) What is 2.1 percent of 60?
- 3) Who won the free-throw battle for the first 20 games?

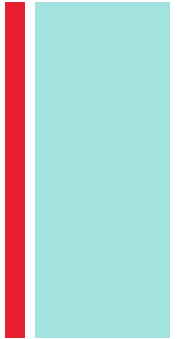
(from 2013)	First ten games	Next ten games
S. Curry (%)	0.90	0.80
K. Durant (%)	0.85	0.70

+ Warm-UP

1) A **sample** is a subset of the defined population (usually selected for study in some manner).

The characteristic or variable of a **sample** is called a **statistic**.

The characteristic or variable of a **population** is called a *parameter*.



+ What is 2.1 percent of 60?

■ What is 100 percent of 60?

60

■ What is 50 percent of 60?

30

■ What is 10 percent of 60?

6

$$60(10\%) =$$

$$60(0.1) = 6$$

■ What is 1 percent of 60?

$$60(1\%) =$$

$$60(0.01) = \mathbf{0.6}$$

$$\text{or } 6(10\%) =$$

+ What is 2.1 percent of 60?

■ What is 1 percent of 60?

$$60(0.01) = 0.6$$

■ What is 2 percent of 60?

$$60(0.02) = 1.2$$

■ What is 0.1 percent of 60?

$$60(0.001) = 0.06$$

■ What is 2.1 percent of 60?

$$60(0.021) = 60(0.02) + 60(0.001) = 1.26$$

+ Warm-UP

- Who won the free-throw battle for the first 20 games (from 2013 season)?

Statistics necessitates that we make decisions with incomplete information (**statistics** from ***samples!***)

S. Curry (%)	$\frac{0.90}{10}$	$\frac{0.80}{100}$	0.809 $= \frac{89}{110}$
K. Durant (%)	$\frac{0.85}{100}$	$\frac{0.70}{10}$	0.836 Winner! $= \frac{92}{110}$

Mistakes can occur when you try to “average” averages!



Chapter 1: Role of Statistics & the Data Analysis Process

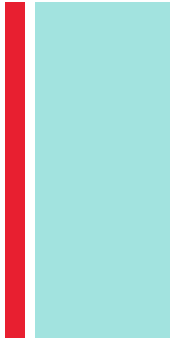
Introduction

Data Analysis: Making Sense of Data



Chapter 1

Role of Statistics & the Data Analysis Process

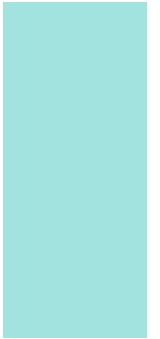


- **Introduction 1.1 – 1.3:** Statistics, Variability, and the Data Analysis Process
- **1.4** Types of Data & Graphical Displays of Data



Introduction

Data Analysis: Making Sense of Data



Learning Objectives

After this section, you should be able to...

- ✓ DEFINE “Individuals” and “Variables”
- ✓ DISTINGUISH between “Categorical” and “Quantitative” variables
- ✓ DEFINE “Distribution”
- ✓ DESCRIBE the idea behind “Inference”

- **Statistics** is the science of data.
 - **Data Analysis** is the process of *organizing, displaying, summarizing, and asking questions* about data.

Definitions:

Individuals (*or Observations*) – objects (people, animals, things) described by a set of data

Variable - any characteristic of an individual

Categorical Variable

– places an individual into one of several groups or categories.

Quantitative Variable

– takes numerical values for which it makes sense to find an average.

- A variable generally takes on many different values. In data analysis, we are interested in how often a variable takes on each value.

Definition:

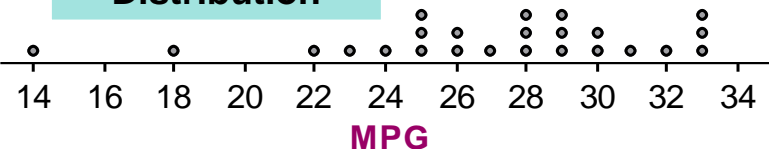
Distribution – tells us what values a variable takes and how often it takes those values

Example

MODEL	MPG	MODEL	MPG	MODEL	MPG
Acura RL	22	Dodge Avenger	30	Mercedes-Benz E350	24
Audi A6 Quattro	23	Hyundai Elantra	33	Mercury Milan	29
Bentley Arnage	14	Jaguar XF	25	Mitsubishi Galant	27
BMW 5281	28	Kia Optima	32	Nissan Maxima	26
Buick Lacrosse	28	Lexus GS 350	26	Rolls Royce Phantom	18
Cadillac CTS	25	Lincoln MKZ	28	Saturn Aura	33
Chevrolet Malibu	33	Mazda 6	29	Toyota Camry	31
Chrysler Sebring	30	Mercedes-Benz E350	24	Volkswagen Passat	29

Variable of Interest:
MPG

Dotplot of MPG Distribution

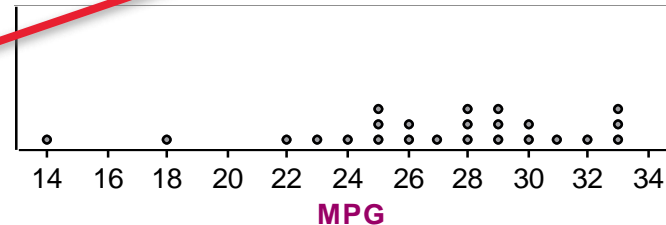


How to Explore Data

Examine each variable by itself.
Then study relationships among the variables.

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

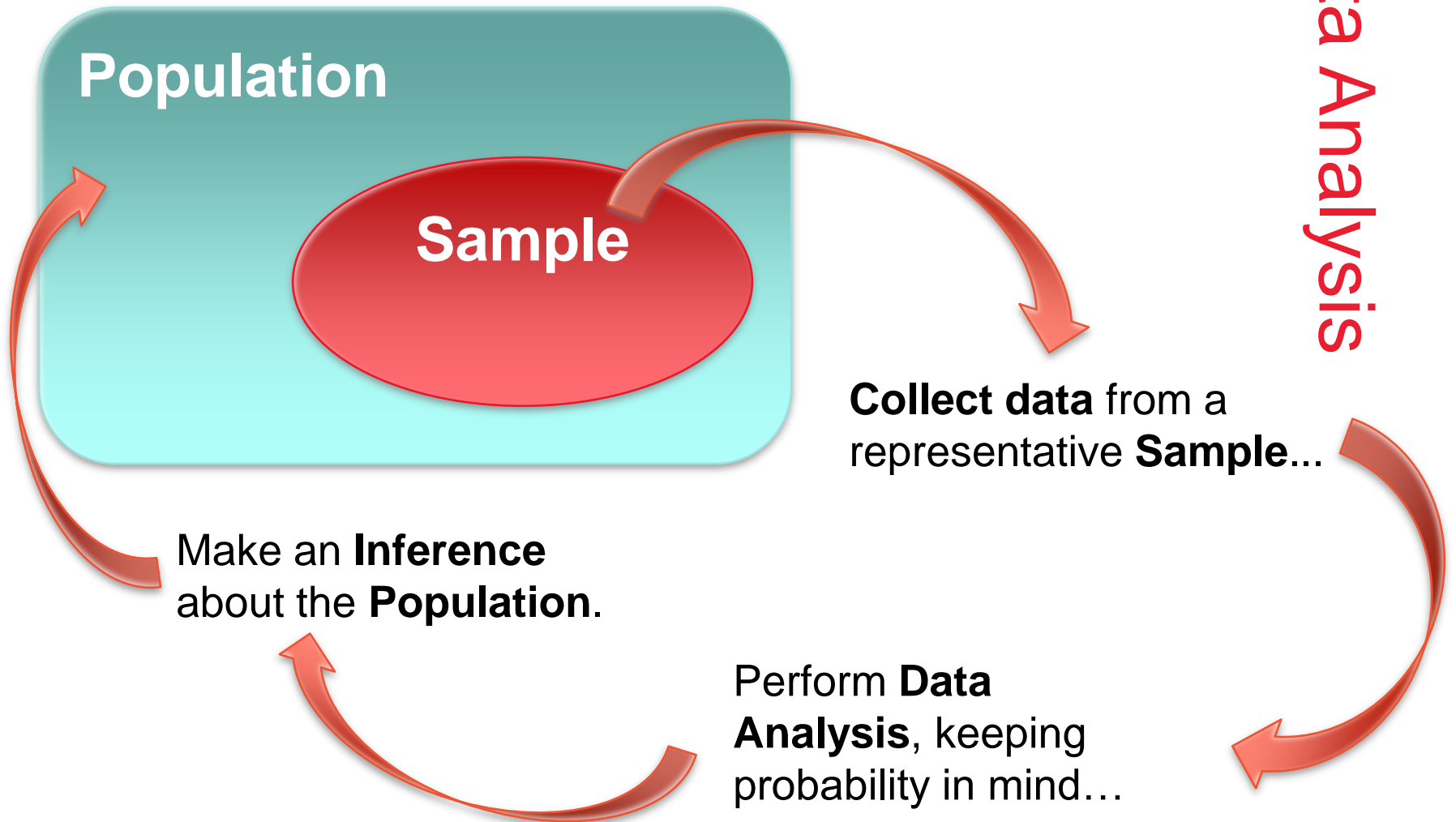
Start with a graph or graphs



Add numerical summaries

```
1-Var Stats
x̄=27
Σx=648
Σx²=17992
Sx=4.643836495
σx=4.546060566
↓n=24
```

From Data Analysis to Inference



+ The Data Analysis Process (p. 6 in your textbook)

- Understanding the nature of the problem
- Deciding what to measure and how to measure it
- Data Collection
- Data Summarization & preliminary analysis
- Formal data analysis
- Interpretation of results

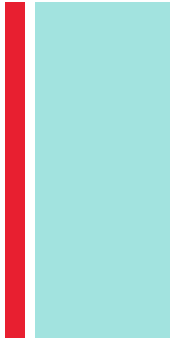
■ Let's talk
about stats
BABY!





Poll Everywhere (BYOD)

- wifi at: **PollEv.com/clowber280**
- **Text: clowber280** to **37607** enter my poll



📧 Respond at **PollEv.com/clowber280**

📧 Text **CLOWBER280** to **37607** once to join, then **A, B, C, D, or E**

What magnet are you in here at Manual?

- | | |
|------|---|
| HSU | A |
| J&C | B |
| MST | C |
| VA | D |
| YPAS | E |

+ Let 's Talk Stats...

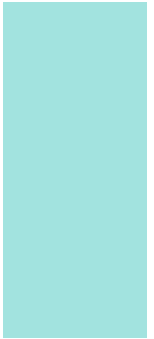
- 1.) How did deaths per year from natural disasters change in the last century?
- 2.) Worldwide, women aged 30 spent about how many (total) years in school?
(Note: Men of the same age spent 8 years)
- 3.) In the last 20 years, the percent of people living in extreme poverty has...





Introduction

Data Analysis: Making Sense of Data



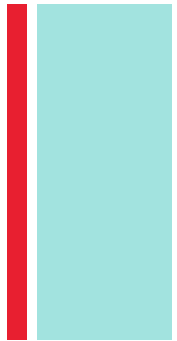
Summary

In this section, we learned that...

- ✓ A **dataset** contains information on **individuals**.
- ✓ For each individual, data give values for one or more **variables**.
- ✓ Variables can be **categorical** or **quantitative**.
- ✓ The **distribution** of a variable describes what values it takes and how often it takes them.
- ✓ **Inference** is the process of making a conclusion about a population based on a sample set of data.



Sample Student responses from last year



When poll is active, respond at PollEv.com/clowber280

Text **CLOWBER280** to **37607** once to join

How did deaths per year from natural disasters change in the

Women aged 30 spent about how many years in school? (Men of the same age spent 8 years)

R
De



Powered by Poll Everywhere

+ HW 1 – Counting on Dyscalculia

Article and Q's (SAMPLE answers)

4B. The two phrases given are not the same, and therefore not equally likely. Given that there are many people around the world that speak English, but are not U.S. citizens, the conditional probability of

$$P(\text{Speak Eng} | \text{U.S. Cit}) > P(\text{U.S. Cit} | \text{Speak Eng})$$

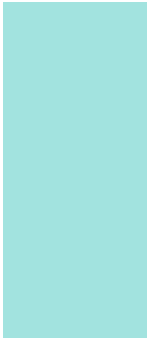
No they are not the same thing. One is asking the probability that an english speaker is American and the other is the probability that an American is an english speaker.

5. No because the chance of getting breast cancer in a 40 year old woman is 1.5 percent and then as they continue to age their chances of getting cancer increase.





Frequency Distributions & Bar Charts for Categorical Data



- **Frequency Distribution:** A table that displays the possible categories along with the associated *frequencies* (the count or number of times it occurs)
- **Relative Frequency Distribution:** A table that displays the possible categories along with the proportion of observations for each category.
- $relative\ frequency = \frac{frequency}{total\ observations\ in\ data\ set}$

- **Categorical Variables** place individuals into one of several groups or categories
 - The values of a categorical variable are labels for the different categories
 - The distribution of a categorical variable lists the count or percent of individuals who fall into each category.

Example

Frequency Table	
Format	Count of Stations
Adult Contemporary	1556
Adult Standards	1196
Contemporary Hit	569
Country	2066
News/Talk	2179
Oldies	1060
Religious	2014
Rock	869
Spanish Language	750
Other Formats	1579
Total	13838

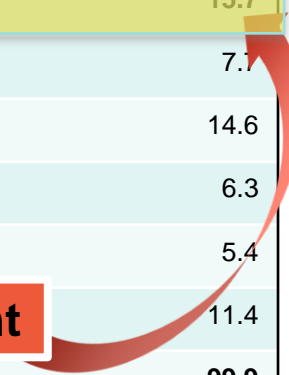
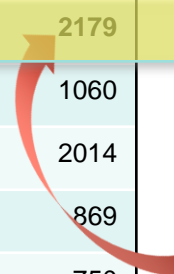
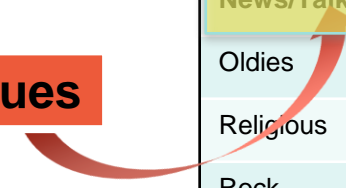
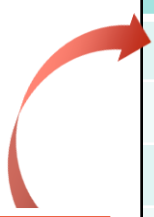
Relative Frequency Table	
Format	Percent of Stations
Adult Contemporary	11.2
Adult Standards	8.6
Contemporary Hit	4.1
Country	14.9
News/Talk	15.7
Oldies	7.7
Religious	14.6
Rock	6.3
Spanish Language	5.4
Other Form	11.4
Total	99.9

Variable

Values

Count

Percent



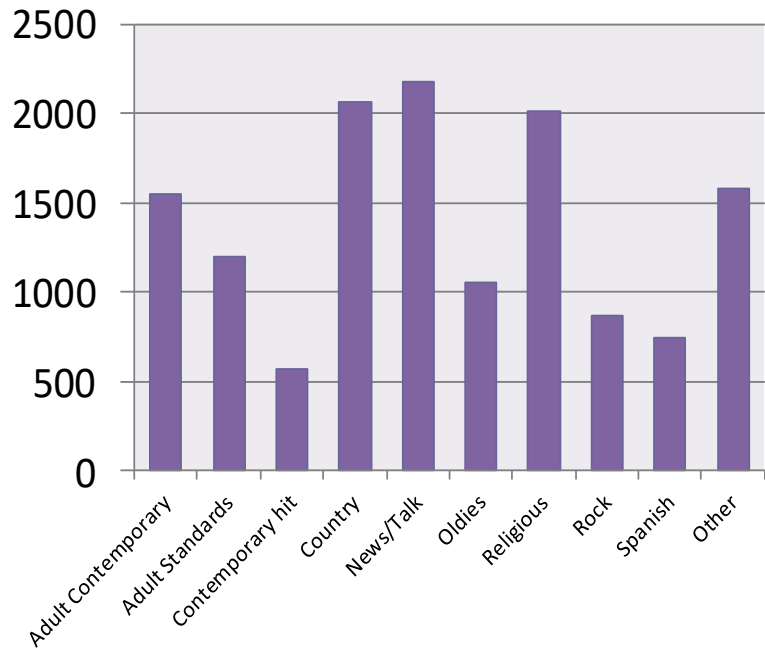
Displaying categorical data

Frequency tables can be difficult to read.

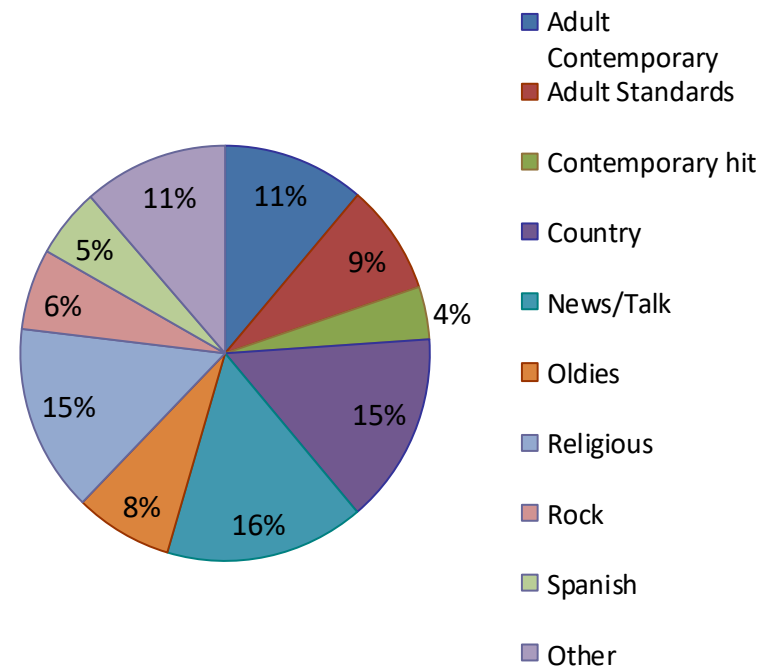
Sometimes it is easier to analyze a distribution by displaying it with a **bar graph** or **pie chart**.

Relative Frequency Table

Count of Stations



Percent of Stations

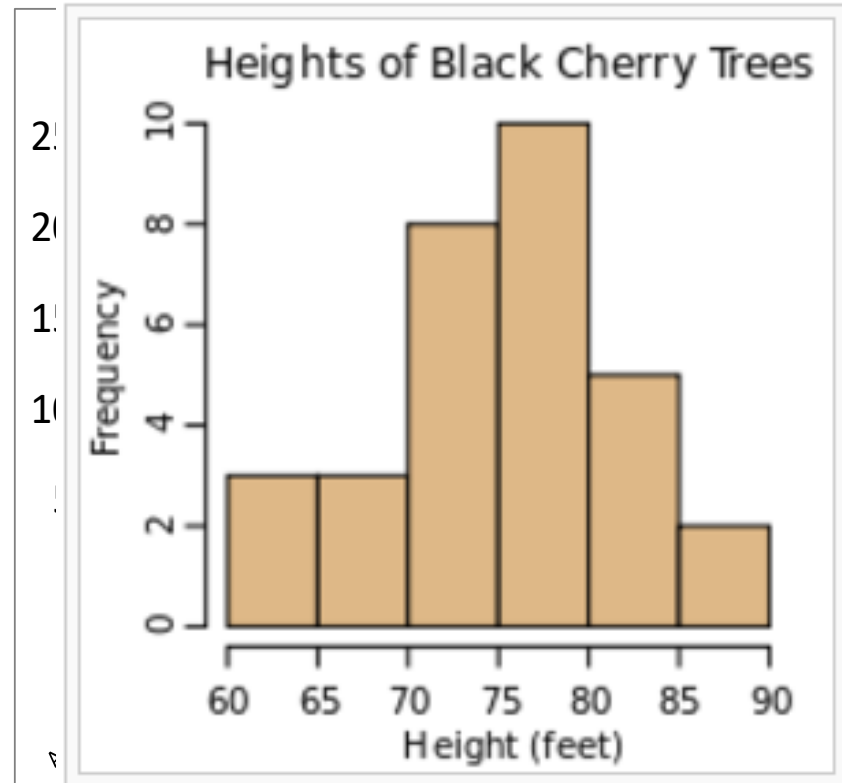


■ Displaying categorical data

Frequency tables can be easier to analyze by displaying the distribution with a **bar graph**.

Compare these 2 graphical displays:

Frequency Table	
Format	Count of Stations
Adult Contemporary	1556
Adult Standards	1196
Contemporary Hit	569
Country	2066
News/Talk	2179
Oldies	1060
Religious	2014
Rock	869
Spanish Language	750
Other Formats	1579
Total	13838



Bar Graphs vs. Histograms (or Bar Charts)

Bar charts and histograms compare sizes of different groups.

Bar charts

- Qualitative groups
- Symmetry and skewness not used
- Space between columns
- Columns can be vertical or horizontal

Histograms

- Quantitative groups
- Symmetry and skewness are used
- No space between columns
- Columns are always vertical

Bar Graphs \neq Histograms

Bar Graphs (or Charts)

Video Link: <http://stattrek.com/statistics/charts/histogram.aspx?Tutorial=AP>



Looking Ahead...



In the next Section...

We'll learn how to analyze categorical data.

- ✓ **Two-Way Tables**
- ✓ **Conditional Distributions**
- ✓ **Experimental Design**
- ✓ **Sampling Techniques**

We'll also learn how to organize a statistical problem.



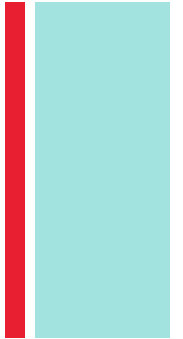
END of slides

See you next time!



Review HW #1

Counting on Dyscalculia



1) What are the three common causes of the problem for misinterpreting statistics that are cited by the author?

- **(1) psychological inability to objectively confront numbers or health hazards; (2) ignorance/confusion about the mathematics of statistics; (3) factual errors caused by how statistics were obtained/generated**

2) The author states that we have a preference for remembering statistics that are nice round numbers, typically ones that are multiples of what?

- **Multiples of 10** (base-10 number system)

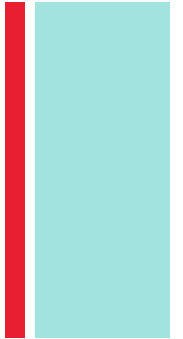
3) What aspect is most critical about a random sample?

- **Absolute size of a sample is most important**, not its percentage of the population of interest.



Review HW #1

Counting on Dyscalculia



4) Are these two phrases the same (meaning are they equally likely): (1) probability that someone is a U.S. citizen given that he or she speaks English, vs (2) probability that someone speaks English given that he or she is a U.S. citizen? Why or why not?

■ **Conditional probability issue: $P(a|b)$**

■ **Read as “What is the probability of a , given that b is known”**

$$(1) P(\text{U.S Citizen} \mid \text{speak English}) = \frac{20}{100}$$

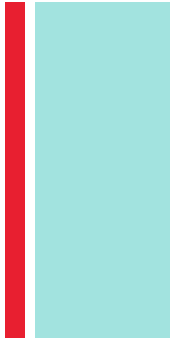
$$(2) P(\text{speak English} \mid \text{U.S Citizen}) = \frac{95}{100}$$

Conditional probability issue, especially w/ medical tests and false positives (*Bayes' Theorem*)



Review HW #1

Counting on Dyscalculia



- 5) Considering that hear that 1 in 8 women will develop breast cancer, should all women be equally fearful of developing the disease within the next few years? Give an example to help explain your answer.
- **No, a woman's age changes her risk factor.**
- Population of women in their early 20's, only have a **0.5% chance (5 out of 1000)** of developing breast cancer over the next 20 years
- Population of women in their early 40's, only have a **3.8% chance (38 out of 1000)** of developing breast cancer over the next 20 years



Review HW #1

Counting on Dyscalculia

- 6) I THOUGHT this article was a ____ (1 = easy to 5 = extremely difficult), in terms of my ability to understand the ideas presented. This is known as a **Likert scale**.
- 7) Based on the author's figure of "452,888,988,750 cases of *dyscalculia* recorded in this country annually", what was the population of the U.S. at the time this article was written?

$x = U.S. population in 1994$

88.47 % of x have 5.61 per day, $\times 365$ days = Number of annual cases

$$0.8847x(5.61)(365) = 452,888,988,750$$

$$x = \mathbf{250,000,000}$$