# QUANTITATIVE METHODS

Week One

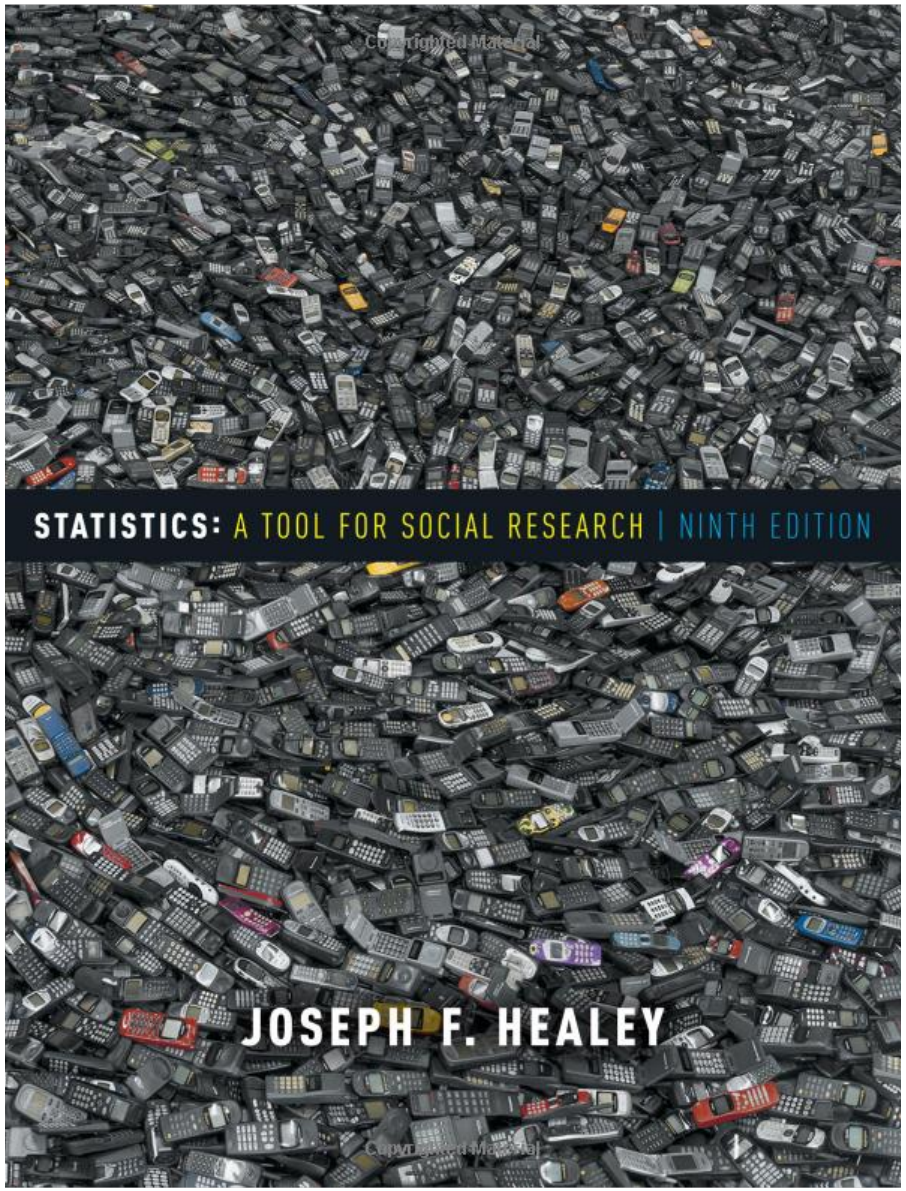
The New School

Aaron Hill

# Introductions

- What is your previous experience with statistics?
- What is your career trajectory?
- What is something interesting about you?

# Review Syllabus

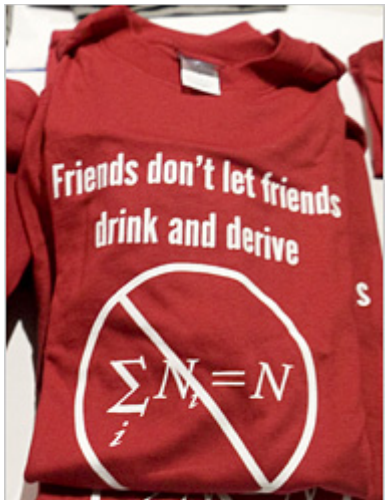

STATISTICS: A TOOL FOR SOCIAL RESEARCH | NINTH EDITION
JOSEPH F. HEALEY


DATA
ANALYSIS
FOR
POLITICS
AND
POLICY
EDWARD R. TUFTE

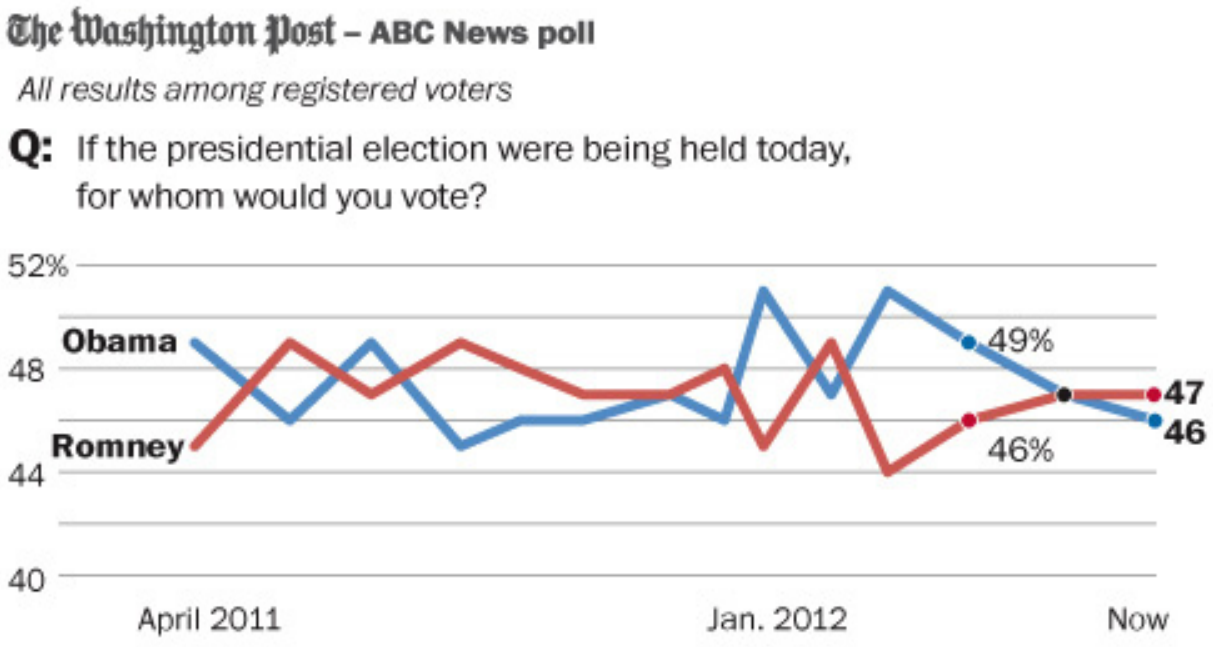
# Expectations

- Complete assigned reading of text, lecture notes, and additional readings
- Manual calculations, numeric and algebraic
- Application of calculated statistics
- Demonstrate proficiency in statistical computing
- Strong writing about your findings
- Attendance, participation, and punctuality

# STATISTICS IN NEWS AND LIFE




# Election Polls



# Romney With Longer Odds

By NATE SILVER

Mitt Romney has always had difficulty drawing a winning Electoral College hand. Even during his best period of polling, in the week or two after the first presidential debate in Denver, he never quite pulled ahead in the polling averages in Ohio and other states that would allow him to secure 270 electoral votes.

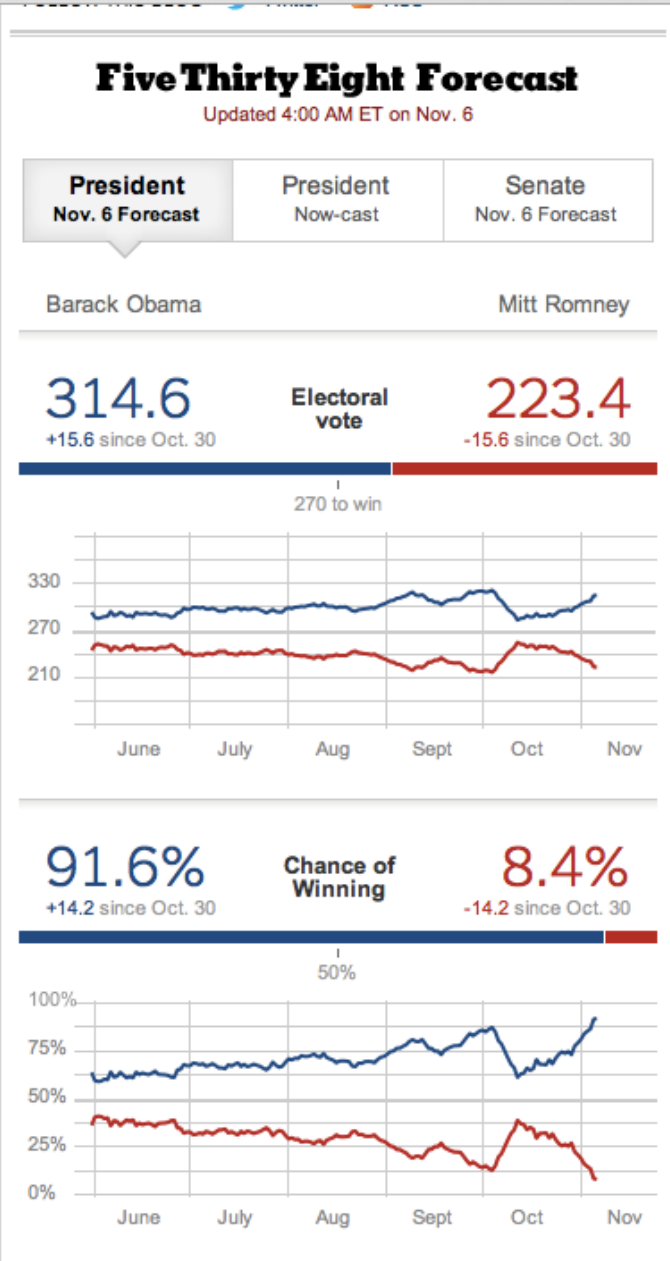
But the most recent set of polls suggest another problem for Mr. Romney, whose momentum in the polls stalled out in mid-October. Instead, it is President Obama who is making gains.

Among 12 national polls published on Monday, Mr. Obama led by an average of 1.6 percentage points. Perhaps more important is the trend in the surveys. On average, Mr. Obama gained 1.5 percentage points from the prior edition of the same polls, improving his standing in nine of the surveys while losing ground in just one.

**Monday's National Polls**

| Pollster | Monday Poll | Change from Prior Poll |
|---|---|---|
| American Research Group | tie | unchanged |
| Angus Reid | Obama +3.0 | Obama +3.0 |
| CVOTER / U.P.I. | Obama +1.0 | unchanged |
| Democracy Corps | Obama +4.0 | Obama +1.0 |
| Gallup | Romney +1.0 | Obama +4.0 |
| Google Consumer Surveys | Obama +2.3 | Obama +2.0 |
| Ipsos / Reuters (online) | Obama +2.0 | Obama +1.0 |
| Monmouth U. / SurveyUSA | tie | Obama +3.0 |
| Public Policy Polling | Obama +2.0 | Obama +2.0 |
| RAND Corporation | Obama +4.4 | Obama +1.2 |
| Rasmussen Reports | Romney +1.0 | Romney +1.0 |
| Wash. Post / ABC News | Obama +3.0 | Obama +2.0 |
| **Average** | Obama +1.6 | Obama +1.5 |

Because these surveys had large sample sizes, the trend is both statistically and practically meaningful. Whether because of Hurricane Sandy, the relatively good economic news of late, or other factors, Mr. Obama appears to have gained ground in the closing days of the race.



## Monday's National Polls

| Pollster | Monday Poll |
|---|---|
| American Research Group | tie |
| Angus Reid | Obama +3.0 |
| CVOTER / U.P.I. | Obama +1.0 |
| Democracy Corps | Obama +4.0 |
| Gallup | Romney +1.0 |
| Google Consumer Surveys | Obama +2.3 |
| Ipsos / Reuters (online) | Obama +2.0 |
| Monmouth U. / SurveyUSA | tie |
| Public Policy Polling | Obama +2.0 |
| RAND Corporation | Obama +4.4 |
| Rasmussen Reports | Romney +1.0 |
| Wash. Post / ABC News | Obama +3.0 |

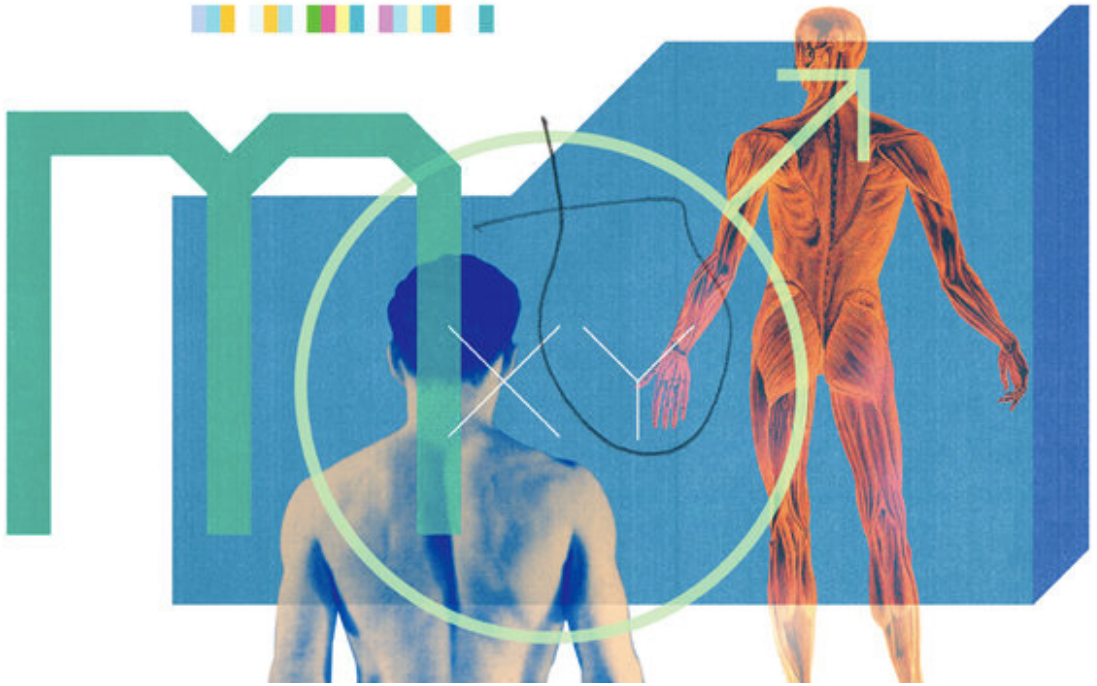
# Health

Well

BODY | JANUARY 29, 2014, 6:58 PM | 18 Comments

## New Concern About Testosterone and Heart Risks

By ANAHAD O'CONNOR

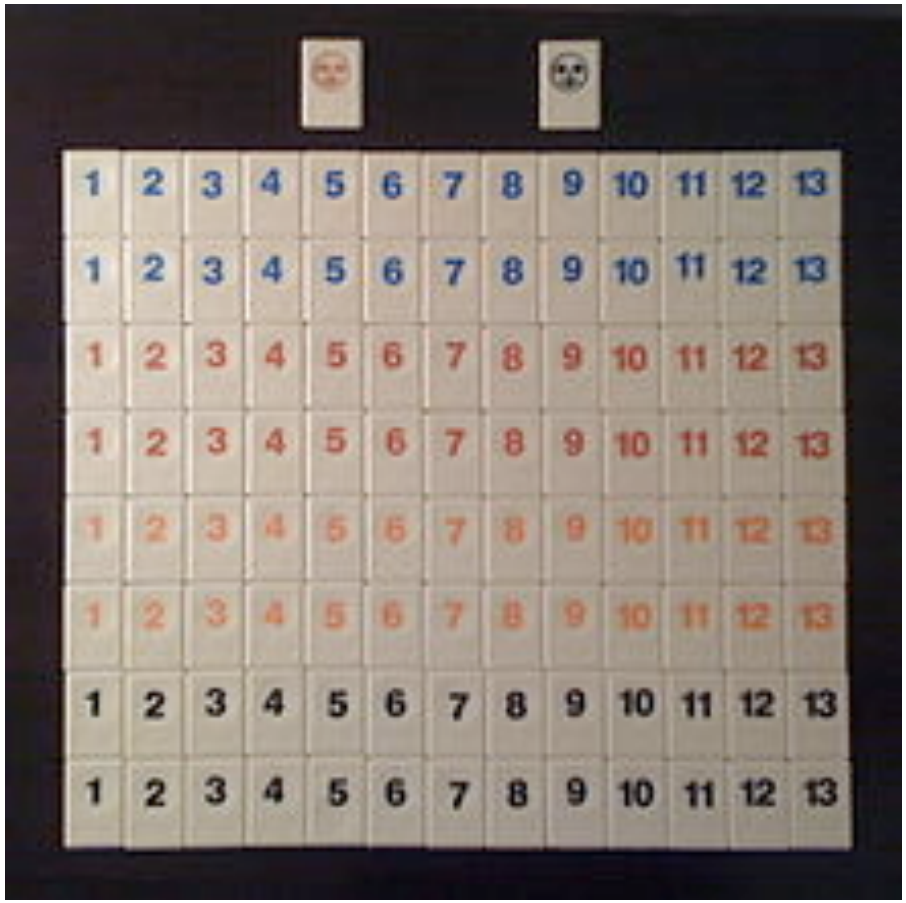


# Scientific Testing

- If schools reinstated physical education classes, a lot of fat children would lose weight.
- And they might never have gotten fat in the first place if their mothers had just breast fed them when they were babies.
- But be warned: obese people should definitely steer clear of crash diets.
- And they can lose more than 50 pounds in five years simply by walking a mile a day.
- *Myths of Weight Loss Are Plentiful, Researcher Says, New York Times January 30, 2013*

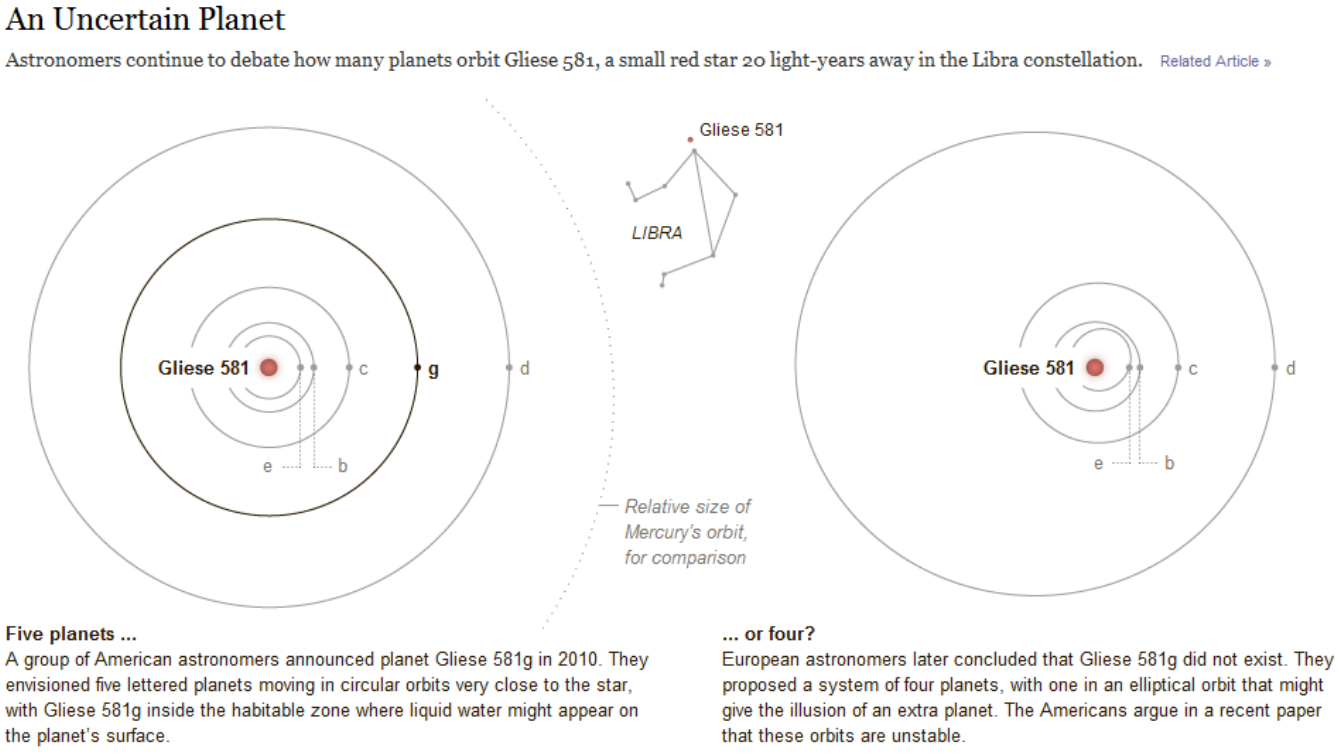
# Aaron's Holiday Dilemma

- 106 Tiles. I'm exhausted from all that tile flipping!



# Goldilocks Planet?

- “More to the point, he and other astronomers said, was that Dr. Vogt’s planet had a 4 percent chance of being a false alarm. That is far above the 1 percent chance normally considered a benchmark for planet detection.”

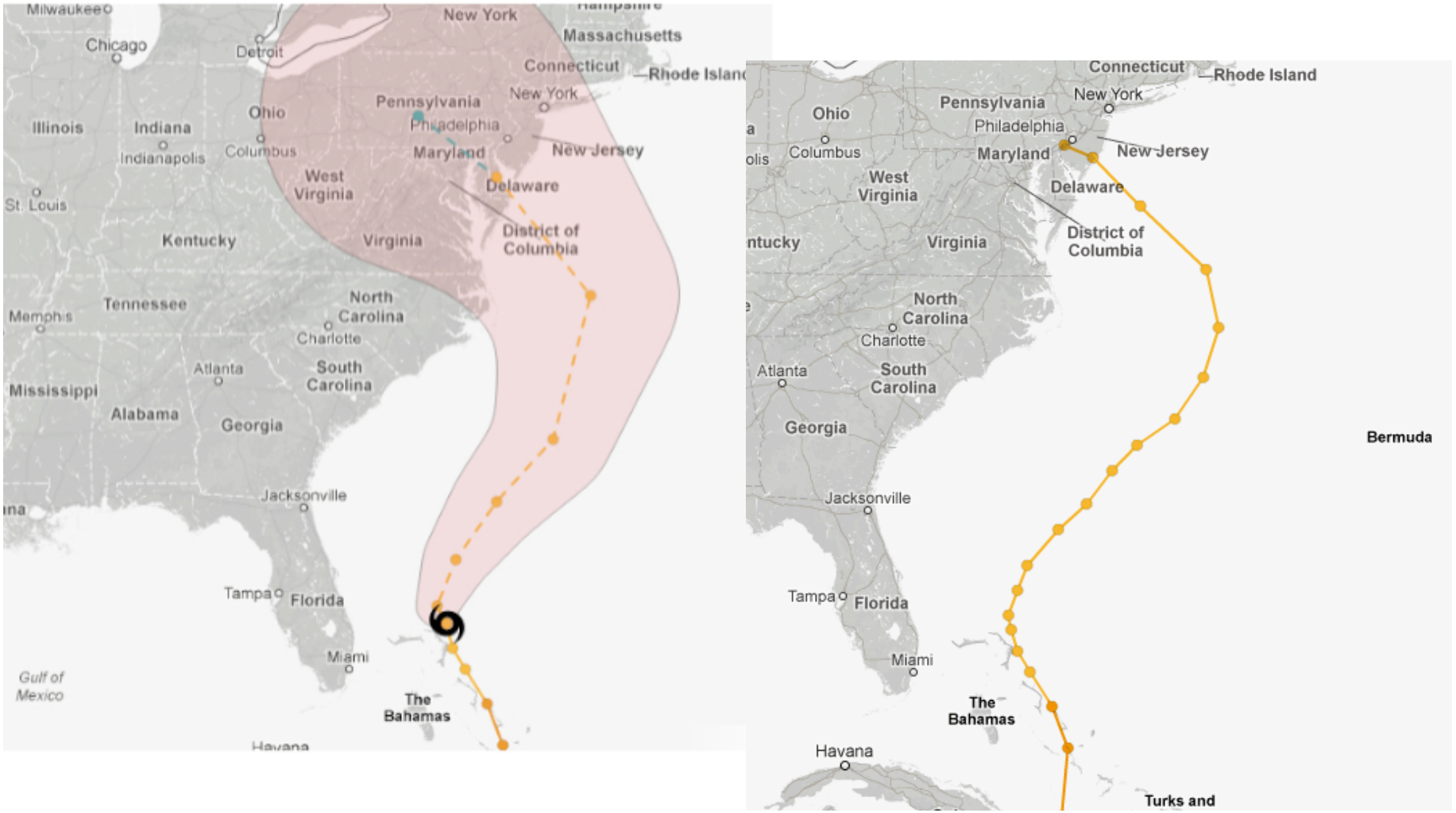


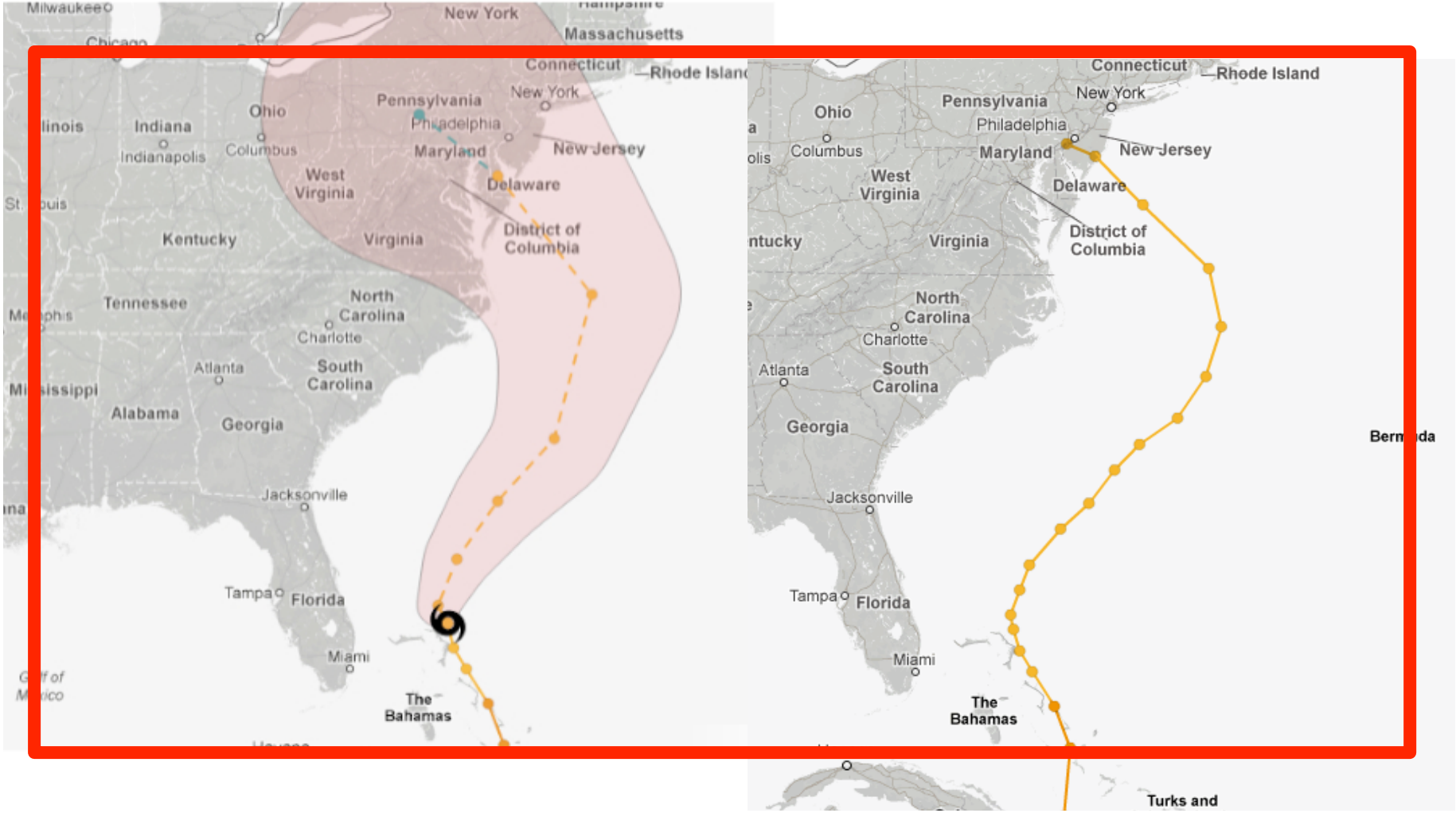
“I always knew I’d die, without knowing when. But now I knew it acutely.”

--Paul Kalanithi

# Teen Drivers

- “Teenage drivers’ risk of a crash increases 44 percent with one teenage passenger, and quadruples with three or more.” –New York Times


Milwaukee
Chicago
Detroit
New York
Massachusetts
Connecticut
Rhode Island
Pennsylvania
New York
Ohio
Illinois
Indiana
Indianapolis
Columbus
Philadelphia
Maryland
New Jersey
West Virginia
Delaware
St. Louis
Kentucky
Virginia
District of Columbia
Tennessee
Memphis
North Carolina
Charlotte
Atlanta
South Carolina
Mississippi
Alabama
Georgia
Jacksonville
Tampa
Florida
Miami
Gulf of Mexico
The Bahamas
Havana
Bermuda
Turks and


Milwaukee
Chicago
New York
Massachusetts
Connecticut
Rhode Island
Pennsylvania
New York
Ohio
Indiana
Philadelphia
Indianapolis
Columbus
Maryland
New Jersey
West Virginia
Delaware
Kentucky
Virginia
District of Columbia
Tennessee
North Carolina
Charlotte
Atlanta
South Carolina
Alabama
Georgia
Jacksonville
Tampa
Florida
Miami
The Bahamas
Turks and


Milwaukee
Chicago
Detroit
New York
Massachusetts
Connecticut
Rhode Island
New York
Pennsylvania
Ohio
Illinois
Indiana
Indianapolis
Columbus
Philadelphia
Maryland
New Jersey
West Virginia
Delaware
St. Louis
Kentucky
Virginia
District of Columbia
Tennessee
North Carolina
Memphis
Charlotte
Atlanta
South Carolina
Mississippi
Alabama
Georgia
Jacksonville
Tampa
Florida
Miami
Gulf of Mexico
The Bahamas
Havana
Turks and

# What Would You Do?

…If I said it were going to rain tomorrow?

- …If the chance of rain is 10%?
- …if the chance of rain is 90%?

# The Weather: a Metaphor for Inferential Statistics

- Making an estimate:

  *It's going to rain.*

- Assessing the estimate:

  *Probability of rain is 90%.*

- Drawing conclusions:

  *Bring umbrella.*

# Types of Data Analysis

- Quantitative Methods
  - Testing theories using numbers
- Qualitative Methods
  - Testing theories using language

# Statistics Defined

- A scientific framework for collecting quantitative data, making estimates, and drawing conclusions.

Humans are good, she knew, at discerning subtle patterns that are really there, but equally so at imagining them when they are altogether absent."

Carl Sagan (*Contact*, 1985)


N.Y.C.TAXI
2L75

# Why Do We Need Statistics?

*Objective analysis*

- Who will win the presidential race?
- What is the rate of poverty in the United States?
- Does Xanax decrease anxiety?
- Does education lead to increases in earnings?
- How many licks does it take to get the Tootsie Roll center of a Tootsie Roll Pop?

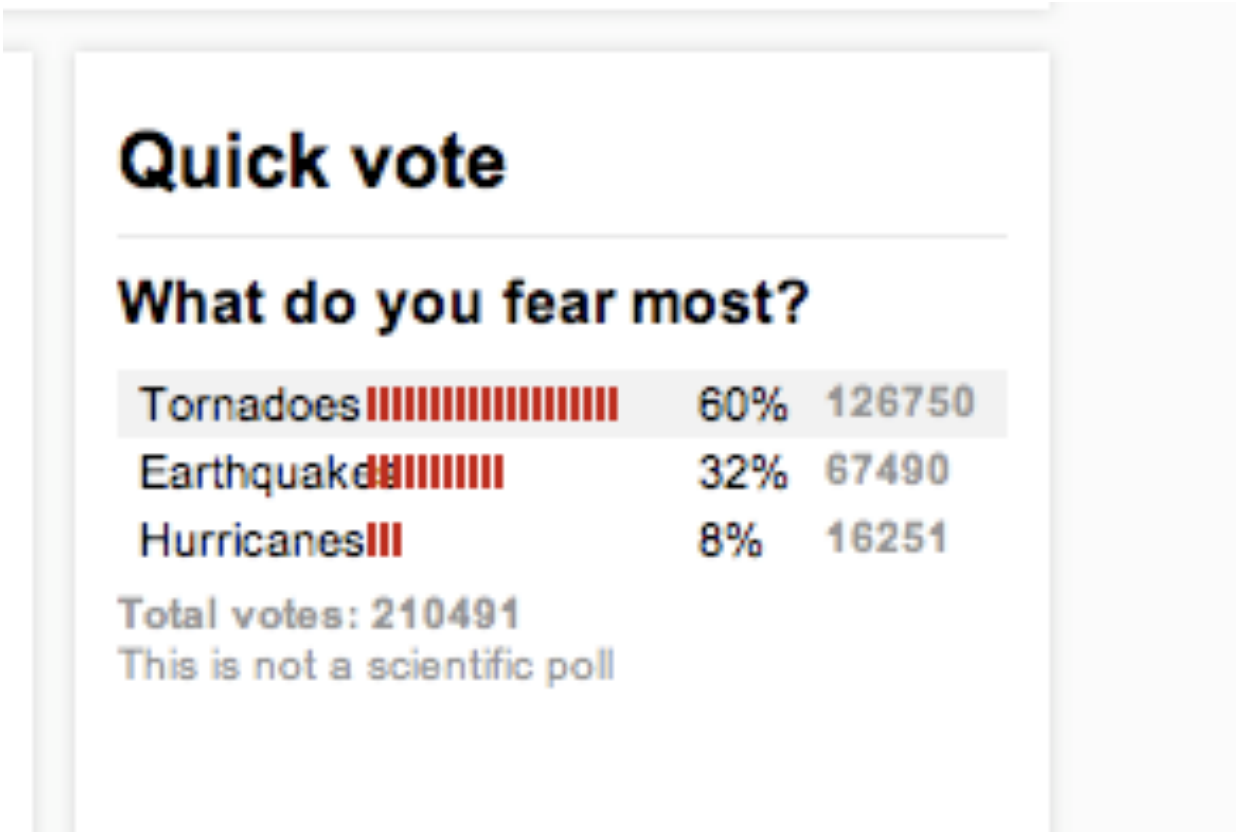
# 8% OF AMERICANS FEAR HURRICANES

# 8% of Americans fear hurricanes

- Who calculated the number?
- How did they calculate it?
- How many people did they ask?  How did they ask them?
- Do the people they asked reasonably represent the population of America?
- Who reported the statistic?

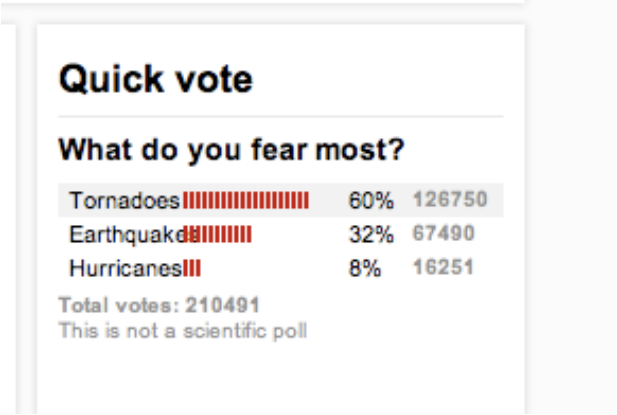
# 8% of Americans fear hurricanes?

- CNN.com, August 2011:



# Why Do We Need Statistics?

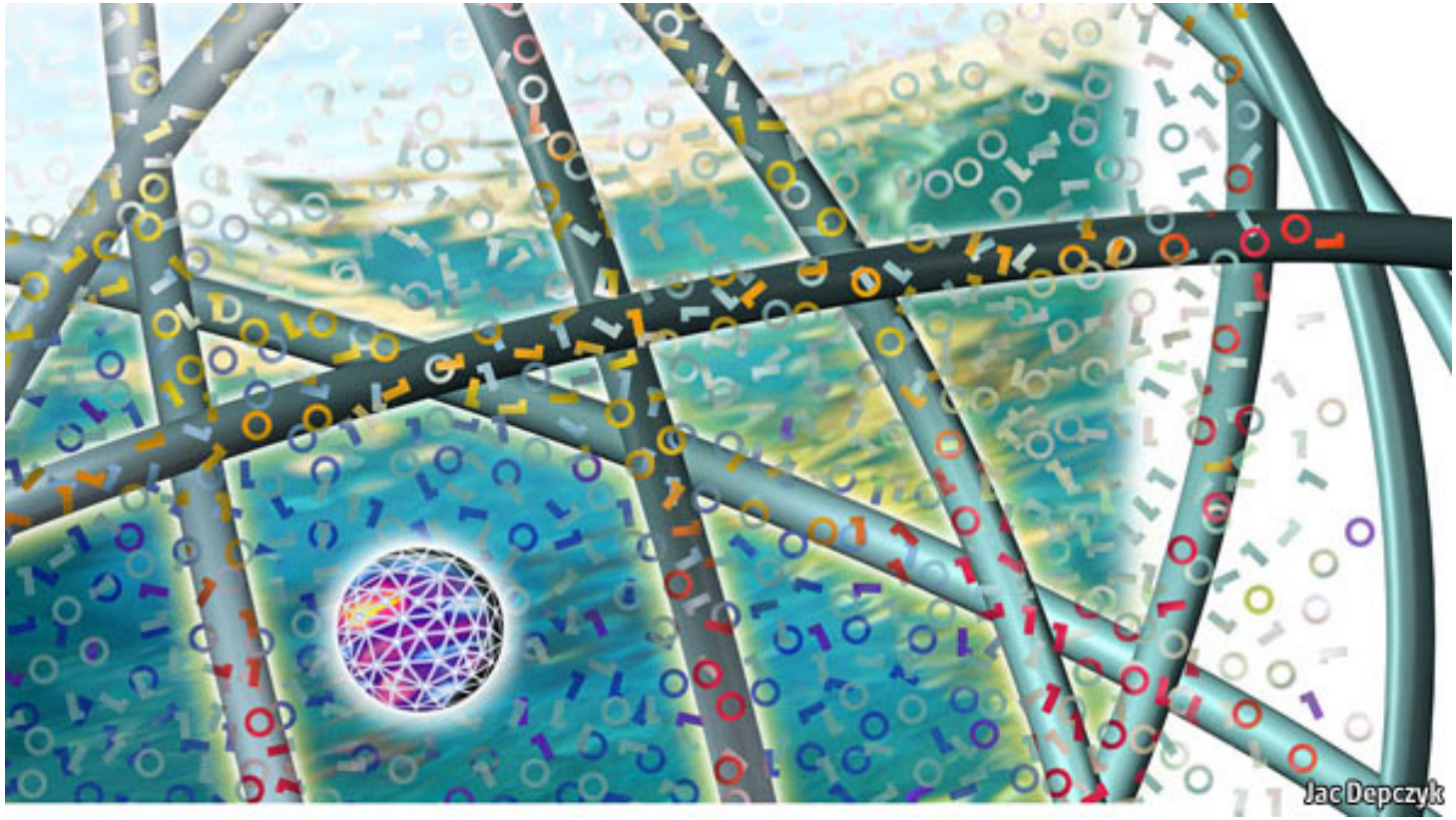
*Interpret analysis by others*



# Why do people hate statistics?

- “I keep saying that the sexy job in the next 10 years will be statisticians,” said Hal Varian, chief economist at Google. “And I’m not kidding.”

# Data, data everywhere



# It’s a numbers world!

- We’re able to present and/or estimate millions of occurrences, behaviors, or feelings in one or two numbers

# Where do we get these numbers?

- Data
  - information
- Where do we get the data?
  - Population: we have data for everyone
  - Sample: we don't have data for everyone, so we collect data on a smaller number of people within the population, but we have plans to *infer* meaning to the entire population

# Where do we store these numbers?

- Computers.  Databases.  Software.
- Organized into columns and rows
- Rows are observations
- Columns are variables

# Where do we store these numbers?

- Observation: the unit of analysis.
  - People, countries, schools, groups, individuals, organizations
- Variable: a trait, on which one can observe and record values from case to case
  - Can be quantitative or qualitative

# Where do we store these numbers?

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Hospt | Treat | Outcome | Time | AcuteT | Age | Gender | |
| 2 | 1 | Lithium | Recurrence | 36.143 | 211 | 33 | 1 | |
| 3 | 1 | Imipramine | No Recurrence | 105.143 | 176 | 49 | 1 | |
| 4 | 1 | Imipramine | No Recurrence | 74.571 | 191 | 50 | 1 | |
| 5 | 1 | Lithium | Recurrence | 49.714 | 206 | 29 | 2 | |
| 6 | 1 | Lithium | No Recurrence | 14.429 | 63 | 29 | 1 | |
| 7 | 1 | Placebo | Recurrence | 5 | 70 | 30 | 2 | |
| 8 | 1 | Lithium | No Recurrence | 104.857 | 55 | 56 | 1 | |
| 9 | 1 | Placebo | Recurrence | 2.857 | 512 | 48 | 1 | |
| 10 | 1 | Placebo | No Recurrence | 102.429 | 162 | 22 | 2 | |
| 11 | 1 | Placebo | Recurrence | 55.714 | 306 | 61 | 2 | |
| 12 | 1 | Imipramine | No Recurrence | 106.429 | 165 | 58 | 1 | |
| 13 | 1 | Imipramine | No Recurrence | 105.143 | 129 | 31 | 1 | |
| 14 | 1 | Imipramine | No Recurrence | 83 | 428 | 44 | 1 | |
| 15 | 1 | Imipramine | Recurrence | 27.286 | 256 | 55 | 2 | |
| 16 | 1 | Lithium | No Recurrence | 105.857 | 197 | 57 | 2 | |
| 17 | 1 | Lithium | Recurrence | 5.571 | 227 | 46 | 1 | |
| 18 | 1 | Imipramine | No Recurrence | 98 | 168 | 58 | 1 | |
| 19 | 1 | Lithium | No Recurrence | 16.286 | 194 | 57 | 1 | |
| 20 | 2 | Lithium | Recurrence | 1.286 | 173 | 54 | 1 | |
| 21 | 2 | Lithium | No Recurrence | 2.143 | 48 | 23 | 1 | |
| 22 | 2 | Imipramine | No Recurrence | 100 | 47 | 65 | 1 | |
| 23 | 2 | Imipramine | Recurrence | 27.143 | 95 | 27 | 1 | |
| 24 | 2 | Lithium | Recurrence | 4 | 148 | 50 | 1 | |
| 25 | 2 | Lithium | Recurrence | 74.143 | 127 | 41 | 2 | |
| 26 | 2 | Placebo | No Recurrence | 104.857 | 129 | 65 | 1 | |
| 27 | 2 | Placebo | Recurrence | 0.143 | 182 | 52 | 1 | |
| 28 | 2 | Placebo | Recurrence | 1.429 | 90 | 60 | 1 | |
| 29 | 2 | Placebo | Recurrence | 45.857 | 177 | 25 | 2 | |
| 30 | 2 | Imipramine | Recurrence | 17.429 | 234 | 27 | 2 | |

Sheet1 Sheet2 Sheet3

# Why we need statistics

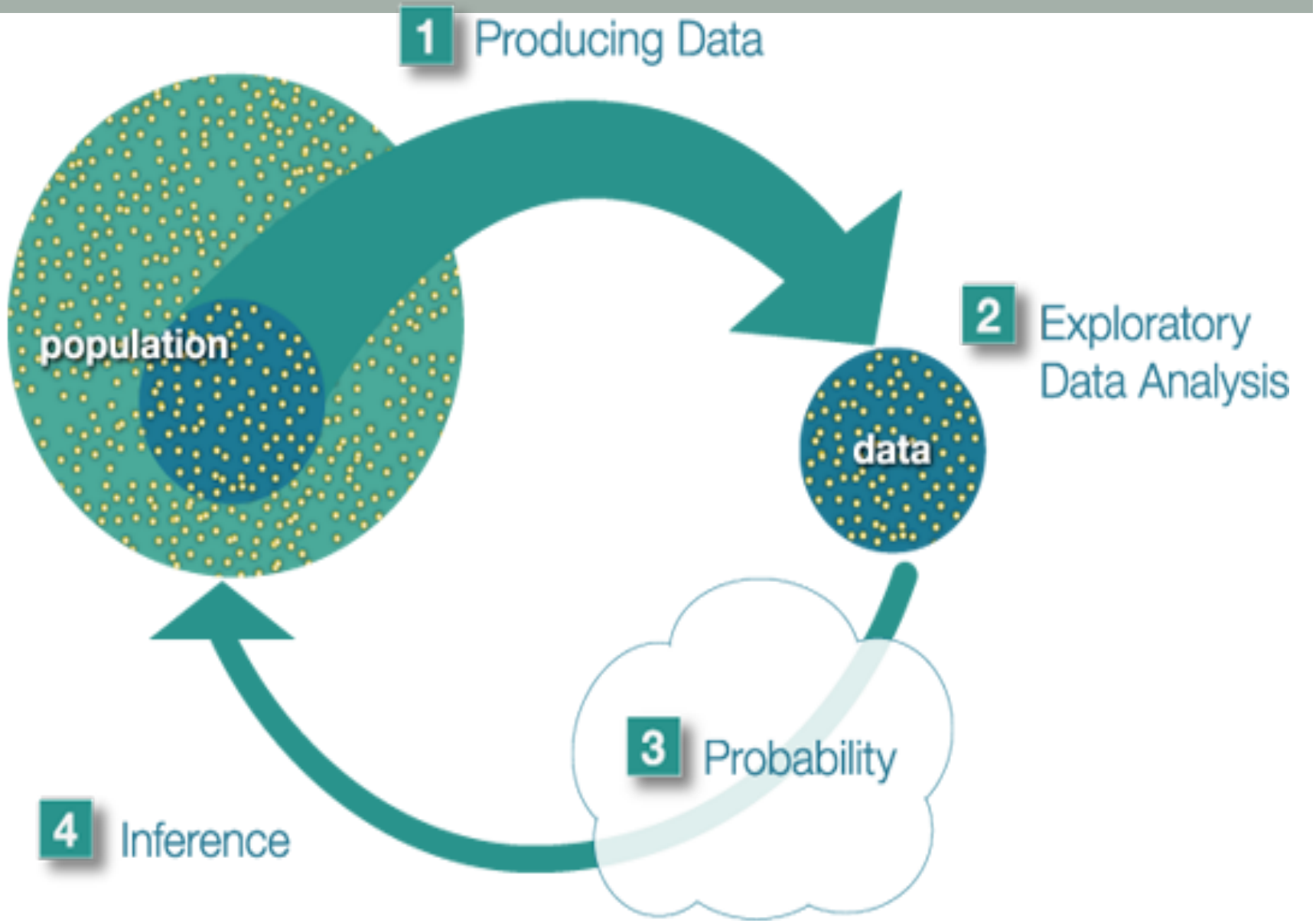
- Objective analysis
- Interpret analysis by others
- Summarize massive amounts of data

# What you get from this class

- Highly marketable research and analytic skills
- Proficiency in statistical computing
- Judge Judy-like abilities to determine which statistics are lies and which are believable
- Interpretation skills & statistical fluency: ability to make sense of complicated analyses
- Knowledge and use of the research process and the scientific method

# Two Part Course

- Part one: Descriptive statistics
- Part two: Inferential statistics
  - Using the scientific method
    - Systematic pursuit of knowledge; formulation of hypotheses, collection of data, and testing of hypotheses


1 Producing Data
population
2 Exploratory Data Analysis
data
3 Probability
4 Inference

# What are *Quantitative Methods*?

- Quantitative
  - Numeric, measurable with numbers, statistical
- Qualitative
  - Not numeric, measurable on some quality, not statistical
- Methods
  - Techniques

# Quantitative or Qualitative?

- Your annual salary
- What did you like about your last job?
- Did you like your last job?
- On a scale of 1 to 5, with 5 being ‘love’ and 1 being ‘hate’, how did you feel about your last job?
- *Which is subjective: quantitative or qualitative?*

# Quantitative Methods a.k.a. ‘Statistics’

- Not a Math class
- This is a Science & Art class
- Statistics: mathematical techniques and probability theory, used as one tool to describe, test, estimate, predict, and learn.
- You all already know how to calculate the most widely used statistic!

# Variables: Level of Measurement

- Variables can be quantitative or qualitative
- If quantitative, there are three possible levels of measurement:
  - Nominal
  - Ordinal
  - Interval-Ratio

# Level of measurement: nominal

- Data that is not numeric in any way, but strictly fits into categories
- Examples:
  - Type of tree
  - Type of pet
  - Zip code

# Level of measurement: ordinal

- Data that may or may not be numerical, but can be arranged in a universally logical order.
- Examples:
  - Belief scales
  - Categories (High, Medium, Low)

# Level of measurement: interval-ratio

- Data that are numerical and the distance between values can be exactly defined.  There is a true zero point, and the intervals are equal.
- Examples:
  - Number of children
  - Weight
  - Income

# Level of measurement

- Why do we care?
  - The method you can use to describe a variable or conduct a statistical test often depends on how it is measured.
  - Can you calculate the average (mean) for a nominal variable such as type of pet?
  - If a ‘population’ of interest is this class, what would a pie chart of age look like?

# Nominal, ordinal, or interval-ratio?

- Gallons of milk
- Name
- Homeland security terrorist level
- Height
- Social security number
- GPA
- Pay scale

# Level of measurement

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Hospt | Treat | Outcome | Time | AcuteT | Age | Gender | |
| 2 | 1 | Lithium | Recurrence | 36.143 | 211 | 33 | 1 | |
| 3 | 1 | Imipramine | No Recurrence | 105.143 | 176 | 49 | 1 | |
| 4 | 1 | Imipramine | No Recurrence | 74.571 | 191 | 50 | 1 | |
| 5 | 1 | Lithium | Recurrence | 49.714 | 206 | 29 | 2 | |
| 6 | 1 | Lithium | No Recurrence | 14.429 | 63 | 29 | 1 | |
| 7 | 1 | Placebo | Recurrence | 5 | 70 | 30 | 2 | |
| 8 | 1 | Lithium | No Recurrence | 104.857 | 55 | 56 | 1 | |
| 9 | 1 | Placebo | Recurrence | 2.857 | 512 | 48 | 1 | |
| 10 | 1 | Placebo | No Recurrence | 102.429 | 162 | 22 | 2 | |
| 11 | 1 | Placebo | Recurrence | 55.714 | 306 | 61 | 2 | |
| 12 | 1 | Imipramine | No Recurrence | 106.429 | 165 | 58 | 1 | |
| 13 | 1 | Imipramine | No Recurrence | 105.143 | 129 | 31 | 1 | |
| 14 | 1 | Imipramine | No Recurrence | 83 | 428 | 44 | 1 | |
| 15 | 1 | Imipramine | Recurrence | 27.286 | 256 | 55 | 2 | |
| 16 | 1 | Lithium | No Recurrence | 105.857 | 197 | 57 | 2 | |
| 17 | 1 | Lithium | Recurrence | 5.571 | 227 | 46 | 1 | |
| 18 | 1 | Imipramine | No Recurrence | 98 | 168 | 58 | 1 | |
| 19 | 1 | Lithium | No Recurrence | 16.286 | 194 | 57 | 1 | |
| 20 | 2 | Lithium | Recurrence | 1.286 | 173 | 54 | 1 | |
| 21 | 2 | Lithium | No Recurrence | 2.143 | 48 | 23 | 1 | |
| 22 | 2 | Imipramine | No Recurrence | 100 | 47 | 65 | 1 | |
| 23 | 2 | Imipramine | Recurrence | 27.143 | 95 | 27 | 1 | |
| 24 | 2 | Lithium | Recurrence | 4 | 148 | 50 | 1 | |
| 25 | 2 | Lithium | Recurrence | 74.143 | 127 | 41 | 2 | |
| 26 | 2 | Placebo | No Recurrence | 104.857 | 129 | 65 | 1 | |
| 27 | 2 | Placebo | Recurrence | 0.143 | 182 | 52 | 1 | |
| 28 | 2 | Placebo | Recurrence | 1.429 | 90 | 60 | 1 | |
| 29 | 2 | Placebo | Recurrence | 45.857 | 177 | 25 | 2 | |
| 30 | 2 | Imipramine | Recurrence | 17.429 | 234 | 27 | 2 | |

Sheet1 Sheet2 Sheet3

# DESCRIPTIVE STATISTICS AND CENTRAL TENDENCY

Week One, Part Two

# Basic Descriptive Statistics

- Percentages & Proportions
  - Standardizes distributions to the base of 100

# Basic Descriptive Statistics

- Percentages & Proportions
  - Proportion: $p = f / N$
  - Percentage: $\% = (f / N) \times 100$

$f$ = frequency, or number of cases in category

N = the number of cases in all categories

# Basic Descriptive Statistics

- Percentages & Proportions
  - Things to remember:
    - Not very good for very small samples
    - ***Always*** show N (number of observations)
    - Percentages & Proportions can be calculated for variables at ***ordinal*** and ***nominal*** levels
    - Ex: 29% of the sample was male

# Basic Descriptive Statistics

- Ratios
  - Shows relative frequency when comparing two categories
  - Tell us exactly how much one category outnumbers another
  - Often multiplied by a larger base (i.e., 10) to eliminate decimal points

# Basic Descriptive Statistics

- Ratios
  - Ratio = *f* 1 / *f* 2

*f* 1 = the number of cases in the first category

*f* 2 = the number of cases in the second category

# Basic Descriptive Statistics

- Ratios
  - What is the ratio of Protestant families to Catholic families?

Given:

1,370 Protestant families

930 Catholic families

# Basic Descriptive Statistics

- Rates
  - Number of actual occurrences divided by number of possible occurrences
  - Also often multiplied to higher power to eliminate decimal points
  - If Statisticsville has a population of 218,000 people and there are 120 auto thefts a year, what is the rate of auto thefts?

# Basic Descriptive Statistics

- Percentage Change
  - How much a variable has increased or decreased ***over time***
  - We need scores of variables over two points of time
  - Scores can be in the form of frequencies, rates or percentages

# Basic Descriptive Statistics

- Percentage Change
  - Percentage Change = (***f2*** - ***f1*** / ***f1***) X 100

***f1*** = first score, frequency or value

***f2*** = second score, frequency or value

# Basic Descriptive Statistics

- Percentage Change

In 2000 the death rate in Statisticsville was 20. In 2010 the death rate in Statisticsville was 17. What is the percentage change from 2000 to 2010?

# Basic Descriptive Statistics

- Frequency Tables
  - Summarize the distribution of a variable
  - Nominal, ordinal, or interval ratio but interval ratio may have to be collapsed to ordinal
  - Must be ***exhaustive*** and ***mutually exclusive***
  - There is a lot of discretion when creating frequency tables

# Basic Descriptive Statistics

- Frequency Tables
  - Nominal Level
    - For each category, the occurences are counted and subtotaled, along with the number of cases reported (N)
    - Make a frequency table with the data given below:

    M, F, M, F, M, M, M, F, M, F, M, F, F, F, F, F, M, M, M, F, F, F, M

    M = Male

    F = Female

# Basic Descriptive Statistics

- Frequency Tables
  - Ordinal Level
    - Percentages make it easier on the reader
    - Make a frequency table for the levels of satisfaction with the waste management services of Statisticsville with the data below and add a column for percentages

VS, S, D, VD, VS, D, S, VD, S, S, D, S,
S, VS, VD, S, S, S, VS

VS = very satisfied
VD = very dissatisfied

S = satisfied
D = dissatisfied

# Basic Descriptive Statistics

- Frequency Tables
  - Interval Ratio Level
    - To construct a frequency distribution table you need to:
      - Sort/list data (ex: ages, list in order from youngest to oldest)
      - Count the number each time the score (age) appears
      - Total the number of scores for each category

# Basic Descriptive Statistics

- Frequency Tables
  - Interval Ratio Level - Discreet
    - Ages of students at a language school:

      18, 34, 81, 29, 22, 33, 25
    - How should this data be sorted?
    - Is this exhaustive and mutually exclusive?

# Basic Descriptive Statistics

- Frequency Tables
  - Interval Ratio Level-Continuous
    - Recall: Continuous means it can be divided infinitely
    - What if we measured were making a frequency table of track times from a race at Statisticsville High School?

# Basic Descriptive Statistics

- Frequency Tables
  - Construct a frequency distribution w/ columns that show percentage distribution, cumulative frequency and cumulative percentages

  Following are reported the number of times 25 residents of a community for senior citizens left their homes for any reason during the past week.

  0, 2, 1, 7, 3, 7, 0, 2, 3, 17, 14, 15, 5, 0, 7, 5, 21, 5, 0, 7, 5, 21, 4, 7, 6, 2, 0, 10, 5, 7

# Basic Descriptive Statistics

- Bivariate Frequency Tables: Crosstabs

| | Right-handed | Left-handed | Total |
|---|---|---|---|
| Males | 43 | 9 | 52 |
| Females | 44 | 4 | 48 |
| Totals | 87 | 13 | 100 |

# Basic Descriptive Statistics

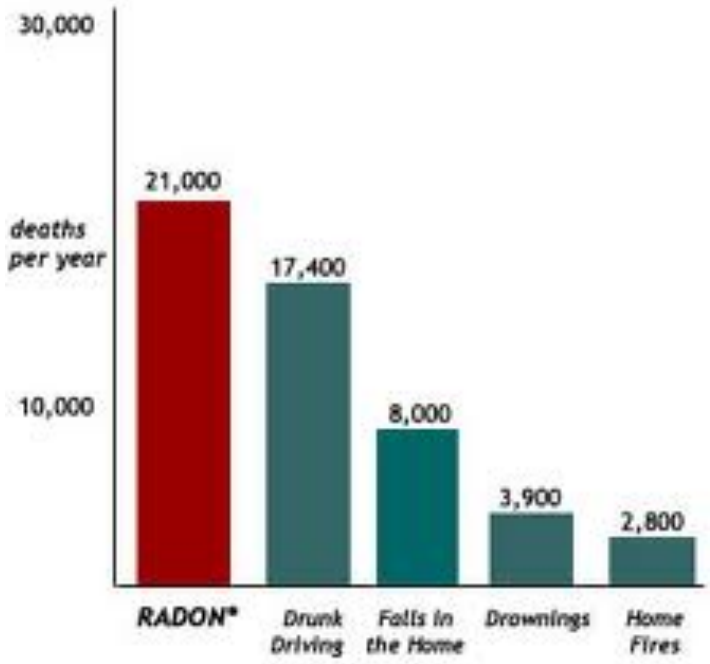
- Charts & Graphs
  - Pie Charts
    - Only look at percentages and always add to 100%
    - Ideal for nominal and ordinal
  - Bar Charts
    - Also communicates frequencies with percentages
    - Nominal and ordinal

# Basic Descriptive Statistics

- Charts & Graphs
  - Histograms
    - Looks like a bar chart but uses I/R data
    - The bars border each other and represent a range of collapsed values
  - Line Charts
    - Similar to histogram but uses points instead of bars and the points are connected
    - Looks at data over time, very flexible and good for showing trends

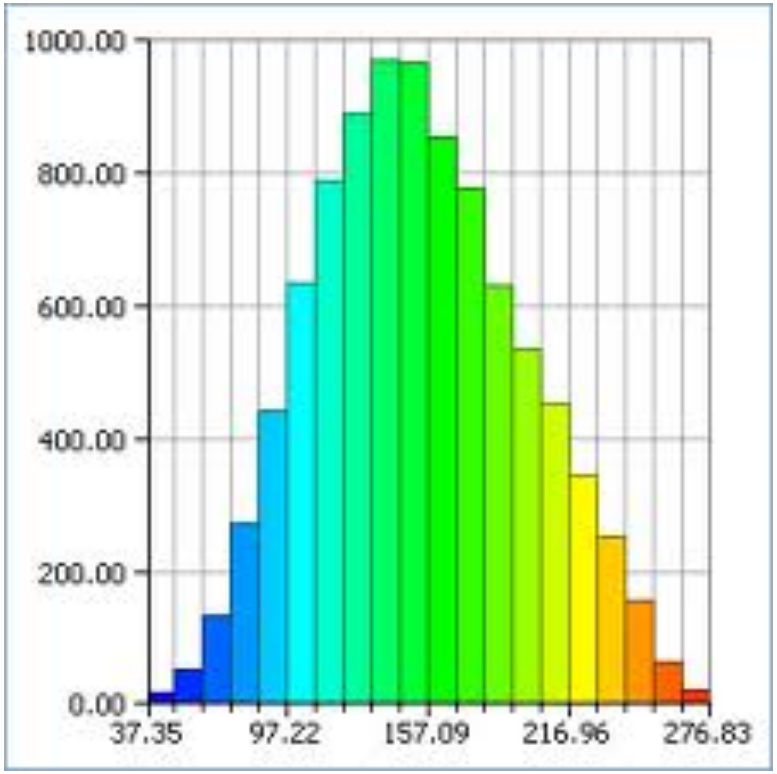
# Basic Descriptive Statistics

- Charts & Graphs - NAME THAT GRAPH!



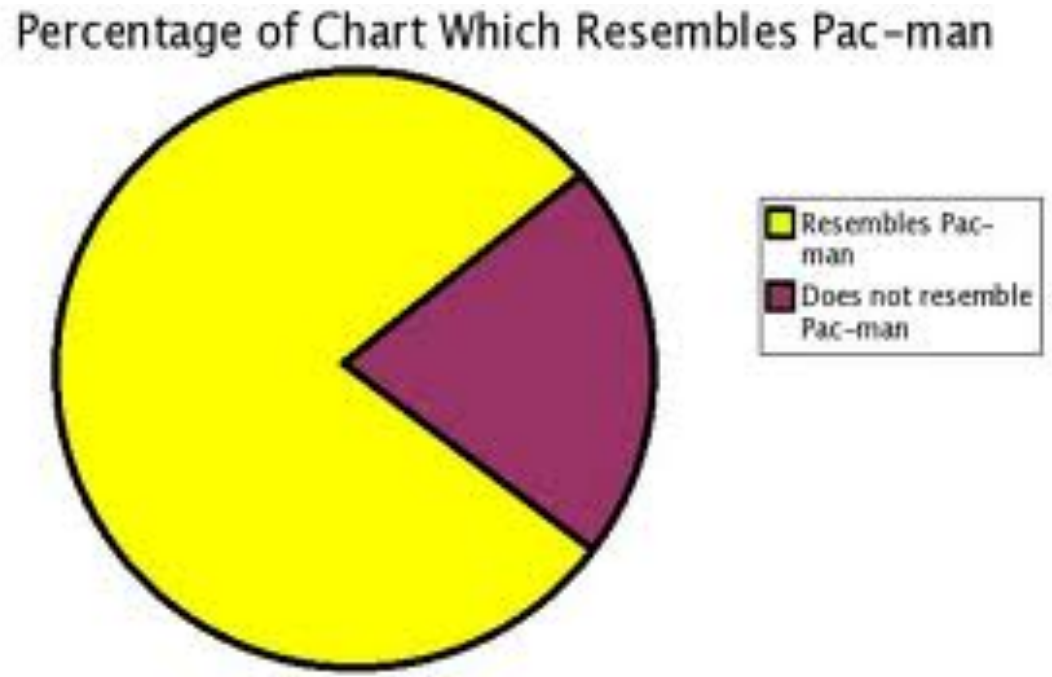
# Basic Descriptive Statistics

- Charts & Graphs - NAME THAT GRAPH!



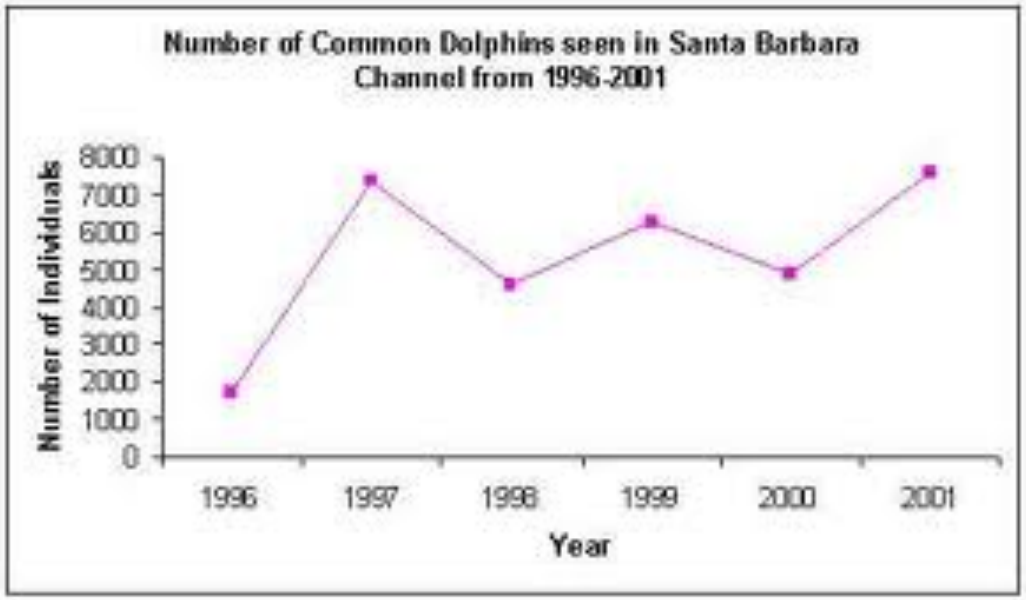
# Basic Descriptive Statistics

- Charts & Graphs - NAME THAT GRAPH!



# Basic Descriptive Statistics

- Charts & Graphs - NAME THAT GRAPH!



# How to describe?

- There are 7.9 million unemployed in Spain and 30.4 million unemployed in the U.S.  Which country has a bigger problem with unemployment?
- Population of Spain: 45.6 million
- Population of US: 304 million

# Mountains out of molehills

- I have two friends.  One is vegetarian, the other isn't.
  - 50% of people are vegetarian!
  - N=2

# Finding N from the percentage

- 15.2% in the U.S. don’t have health insurance

# What is the poverty rate?

- In 2008, 40.1 million lived in poverty in the US (as measured by the government)

# Ratios love to be confusing

- Formula: f1/f2
- Easier to interpret when the larger number is on top
- If I have two dogs and four horses, what is the ratio of *horses to dogs*?
- Expressed as horses:dogs, 4:2

$$\frac{horses}{dogs} = \frac{4}{2} = 2 = \frac{2}{1} = \frac{horses}{dogs}$$

# What happened?

- Before the holidays I weighed 150.  After, I weighed 160.

# Tables, charts, and graphs…

- Are easier to create with software.
- Summarize the ***distribution*** of a variable.
- Level of measurement matters!

# Level of Measurement

- Nominal
- Ordinal
- Interval-Ratio
  - Discreet – A unit of measurement cannot be subdivided
  - Continuous – Can be subdivided infinitely

# Level of Measurement

- Can be changed
- Interval-Ratio can be collapsed to ordinal or nominal
- Categories should be exhaustive and mutually exclusive. What's wrong here?:
  - Age: 1-18
  - Age: 18-65
  - Age: Greater than 65
- *Is age discreet or continuous?*

# Next Steps

- By next class, get textbook and read through Healey Chapter 3 (an old version of the textbook may be accessible on Google books while you wait for shipments)

# MEASURES OF CENTRAL TENDENCY

# Measures of Central Tendency

- Summarize the entire ***distribution*** of values
- Powerful because they can reduce an enormous amount of data to a simple summary statistic
- Can be misleading

# Mode

- Can be calculated for variables of any level of measurement: nominal, ordinal, or interval-ratio
- Usually used only for nominal or ordinal
- Definition: the value that occurs most frequently
- You can have more than one mode

# Mode

- What's the mode?
  - 58, 82, 82, 90, 98
- What's the mode?
  - 2, 9, 8, 7, 1, 7, 8
- What's the mode?
  - Chicken, chicken, dove, dove, alligator, alligator, alligator
- ***What does the mode tell us?***

# Median

- Can be calculated for ordinal or interval-ratio variables only.
- Definition: The exact center of a distribution of values.
- To find the exact center, the values must be **SORTED**
- If there is no exact center (because N is even), take the two center values and average them

# Median

- Finding the middle score when you have lots of values:
  - If N is odd: Add 1 to N and divide by 2.  This is the number of the value that is the median.
  - If N is even: Divide N by 2.  This is the number of the value that is the first middle value; take that value and the next one and average them.

# Median

- What is the median?
  - 1, 2, 3, 4, 5, 6, 7, 8, 9
- What is the median?
  - 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- What is the median?
  - 4, 7, 1, 8, 2
- What is the median?
  - 99, 1, 40, 18, 120, 11
- What is the median?
  - Almond, almond, pecan, cashew

# Median

- What is the median?

**QUIZ1**

| | | Frequency | Percent |
|---|---|---|---|
| Valid | 0 | 2 | 1.9 |
| | 2 | 1 | 1.0 |
| | 3 | 6 | 5.7 |
| | 4 | 6 | 5.7 |
| | 5 | 6 | 5.7 |
| | 6 | 12 | 11.4 |
| | 7 | 14 | 13.3 |
| | 8 | 16 | 15.2 |
| | 9 | 8 | 7.6 |
| | 10 | 34 | 32.4 |
| | Total | 105 | 100.0 |

# Median

- *What does the median tell us?*

# Percentiles, deciles, and quartiles

- The median is also the $50^{th}$ percentile, the $5^{th}$ decile, and the second quartile.
- Other ‘iles’ are calculated in the same way as the median, but rather than the middle (50%) value, you find the value at the point of interest (x%)
- N * ‘ile’ expressed as a proportion = the number sorted set that contains the value of your ‘ile’.  If you get a fraction, round up.

# Percentiles, deciles, and quartiles

- There are three quartiles:
  - 1st: 25%, 2nd: 50%, 3rd: 75%
- There are nine deciles:
  - 1st: 10%, 2nd: 20% …

# Percentiles, deciles, and quartiles

- N = 100; what is the number of the value with the 86$^{th}$ percentile?
- N = 259; what is the number of the value with the 3$^{rd}$ quartile?
- N = 450; what is the number of the value with the 6$^{th}$ decile?
- If you scored above the 75$^{th}$ percentile on a standardized test, how did you do?  What was your score?

# Mean

- Can be calculated for interval-ratio variables only.
- Definition: arithmetic average.
- To calculate: sum the values and divide by N

# Mean

- The mean is the point of the distribution around which the variation of the scores is minimized.  The mean is closer to all of the values than the other measures of central tendency.
- Least squares principle

# Mean

- Exam grades of 15 students:
  - 88, 48, 60, 51, 57, 85, 69, 75, 97, 72, 71, 79, 65, 63, 73

# Characteristics of the mean

- Very flexible
- Most commonly used and reported statistic
- Is very sensitive to outliers
- Can be misleading

# What is an outlier?

- an atypical and/or erroneous value that deviates markedly from the general behavior of the other values
- Salary:
  - 75,000
  - 55,000
  - 35,000
  - 62,000
  - 10,000,000

# The problem with outliers

- Skew
  - Heavy influences on the perimeters of your distribution
- Doesn’t always mean it’s wrong
- But it can wreak havoc on measures you might’ve liked to report

# Skew

- If the median is greater than the mean, you have negative skew
- If the median is less than the mean, you have positive skew
- If the median is equal to (or close to) the mean, you have no skew

# Mean

- Don’t be discouraged by these ‘cautions’ about the mean
- It’s a Superstar statistic that will be making frequent appearances throughout the course
- Knowing its downfalls is knowing when to use it, when not to use it, and when to ‘fix’ it

# Next up:

- Measures of dispersion
- These measures will help us determine properties of the ***distribution***

# MATH REVIEW

Or, remind me what I have forgotten since 7th Grade.

# Part 1: Why do we need a math review?

- A solid foundation of basic math skills goes a long way in this course
- Complicated math equations look easy once they are broken down
- Mastering the math now will free up your time for in depth analysis later
- We're probably all a little rusty

# Math Review-Operations

- 3 X 2
- 3 (2)
- (3)(2)
- 3*2
- 3 / 2
- 3 ÷ 2

- What does it all mean?
- Are there more operators?

# Math Review-Operations

- The first set is a variation on ways to represent multiplication and the second set is a variation of division
- We also use square root and summation

# Math Review-Operations

- Squares and Square Roots
  - Squaring a number / exponents
    - Multiplying the number by itself
    - 4 ^ 2 = 16
    - If the exponent is not a “2”, multiply by itself the number of times dictated by the exponent
    - 4 ^ 3 = 64 , or 4*4*4 = 64

# Math Review-Operations

- Squares and Square Roots
  - Square Roots
    - The value that when multiplied by itself, results in the original number.
    - Ex: $\sqrt{16} = 4$, or 4 squared is 16

# Math Review-Operations

- Summation / Sum operator
- ∑ is an operator like addition, subtraction, multiplication or division
- Directs us to add all of the scores on the variable indicated by the X symbol, the summation of the scores

# Math Review-Operations

- Ex: There are 4 families, all different sizes.  They are 2, 4, 5, & 7.
- $\sum X = 2 + 4 + 5 + 7 = 18$
- $\sum X$ means the sum of - It directs us to add all of the scores on the variable indicated by the *X* symbol

# Math Review-Operations

- $\sum X$
- $\sum X \text{ ^ } 2$
- $(\sum X) \text{ ^ } 2$
- $X$ = 10, 12, 13
- What's the difference?

# Math Review-Operations

- $\sum X$
  - Sum of scores
  - $\sum X = 10 + 12 + 13 = 35$
- $\sum X \text{ ^ } 2$
  - Sum of the squared scores
  - $\sum X \text{ ^ } 2 = (10) \text{ ^ } 2 + (12) \text{ ^ } 2 + (13) \text{ ^ } 2 =$
    $100 + 144 + 169 = 413$
- $(\sum X) \text{ ^ } 2$
  - Sum of the scores squared
  - $(\sum X) \text{ ^ } 2 = (10+12+13) \text{ ^ } 2 = (35) \text{ ^ } 2 = 1{,}225$

# Math Review-Operations

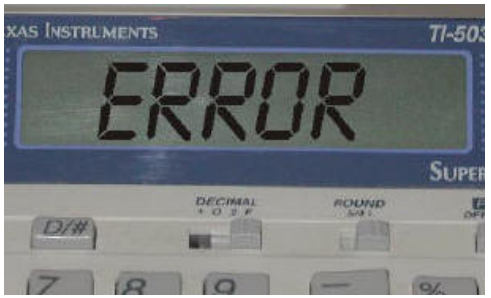
- Negative numbers
  - Adding a negative number is the same as subtracting
    - 3 + (-1) + 4 = 3 - 1 + 4 = 6
  - Subtraction changes the sign of a negative number
    - 3 - (-1) - 4 = 3 + 1 - 4 = 0

# Math Review-Operations

- Negative Numbers - Multiplying & Dividing
  - A negative number times or divided by a positive number will always result in a negative number
    - (3) (-4) = -12 or (-3)(4) = -12
    - (12)/(-4)= -3 or (-12)/(4) = -3
  - A negative number times or divided by a negative number will always result in a positive number
    - -3(-4) = 12
    - (-12)/(-4) = 3

# Math Review-Operations

- Square roots and negative numbers
  - Negative numbers do not have square roots
  - Multiplying a number by itself cannot result in a negative value
  - Squaring a negative number always results in a positive value
  - Calculator error



# Math Review-Scientific Notation

- Speaking of calculator errors…
- What happens when you see a readout of:
  - 8.9 X 10 ^ 3
  - 5.74 X 10 ^  5
  - 0.25 X 10 ^ -1
  - 7.925 X 10 ^ -4

# Please excuse my dear Aunt Sally!

**P**arentheses
**E**xponents
**M**ultiplication
**D**ivision
**A**ddition
**S**ubtraction

# Math Review-Rounding

- Rounding
  - When there are numbers with many values after the decimal point, we have to round
  - Rounding slightly decreases the accuracy of the number but is necessary
  - 15.87659374937493857 + 2.85904829052 is way too long to continue writing in a long equation

# Math Review-Rounding

- Rules for rounding in this class
  - Always round to the ***nearest hundredth***
  - Ex: 15.87659374937493857 = 15.88

# Math Review-Rounding

- Exceptions!!
  - If the number is close to 0, we must treat it differently
  - If you get an answer that is 0.00009 and our rounding rules are followed, the answer will be 0.00
  - When numbers are this small, extra detail is important
  - ***Always*** place a “0” before the decimal point to show the number is less than one
    - Ex: .123 should be written as 0.123

# Math Review - Practice

- Round according to our class rules:
  - 17.17532
  - 43.119
  - 1076.77337
  - 32.4651152301
  - 32.475112301
  - 0.009434534897
  - 1,999,425.0048123876

# Math Review - Practice

- 17 X 3
- 17 (3)
- (17)*(3)
- 17/3
- (42)^2
- √113

# Math Review - Practice

- 17 + (-3) + 4 + (-2) =
- 15 -3 -3 (-5) + 2 =
- (-27)(54) =
- (113)(-2) =
- (-14)(-100) =
- -34 / -2 =
- 322 / -11
- √-2 =
- (-17)^2 =

# Math Review - Practice

- (3 + 7) / 10 =
- 3 + 7 / 10 =
- ((4 - 3) + (7 + 2)) / ((4 + 5)(10)) =
- √(7(5 - 3)^2) / ((17 / 3)(4)) =
- 22 + 44 / 15 =

# Math Review - Practice

- Evaluate each of the expressions below when given (X) of 50, 55, 60, 65, 70:
  - $\sum X$
  - $\sum X \text{ ^ } 2$
  - $(\sum X) \text{ ^ } 2$