Inferential Statistics and Probability
a Holistic Approach

Chapter 1

Displaying and Analyzing Data with Graphs

This Course Material by Maurice Geraghty is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Conditions for use are shown here: https://creativecommons.org/licenses/by-sa/4.0/

1

Introduction

- Syllabus– Homework 0
- Projects
- Computer Lab – S44
  - Minitab
- Website
  - http://nebula2.deanza.edu/~mo
- Tutor Lab - S43 (S41 for MPS)
  - Drop in or assigned tutors – get form from lab.
  - Group Tutoring
- Other Questions

2

Descriptive Statistics

- Organizing, summarizing and displaying data
  - Graphs
  - Charts
  - Measure of Center
  - Measures of Spread
  - Measures of Relative Standing

3

Problem Solving

- The Role of Probability
- Modeling
- Simulation
- Verification

4

Inferential Statistics

- Population – the set of all measurements of interest to the sample collector
- Sample – a subset of measurements selected from the population
- Inference – A conclusion about the population based on the sample
- Reliability – Measure the strength of the Inference

5

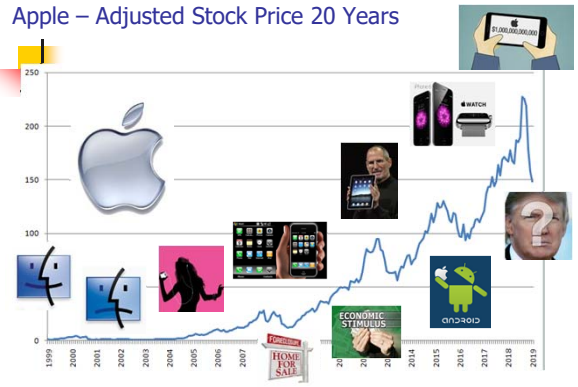
Raw Data – Apple

Monthly Adjusted Stock Price: 12/1998 to 12/2018

115.82	102.97	106.17	75.50	69.86	52.70	41.97	27.42	11.11	25.77	11.04	9.35	4.19	1.39	0.93	1.42	0.97
110.52	115.73	114.39	74.83	76.83	49.73	40.49	26.01	12.06	23.71	11.93	8.82	4.36	1.36	1.01	1.39	1.07
112.96	116.40	103.43	69.93	77.80	52.67	39.16	24.53	14.00	24.72	10.55	7.49	3.41	1.49	1.05	1.14	1.27
112.47	107.44	96.49	63.79	87.18	49.62	36.92	24.12	14.79	19.97	10.02	6.98	2.52	1.35	0.94	1.01	1.68
105.56	109.84	98.16	65.19	86.93	50.07	31.63	21.89	22.06	18.02	8.83	6.10	2.24	1.47	0.96	1.21	3.96
103.12	117.62	91.10	60.15	79.47	50.81	33.47	21.26	20.68	17.14	8.84	5.55	2.10	1.37	0.99	1.22	3.31
94.60	121.63	88.56	52.71	75.99	43.68	32.73	18.53	21.79	15.88	7.45	4.79	2.12	1.24	1.15	1.51	3.41
98.81	126.33	86.17	59.78	75.17	45.26	33.43	17.67	24.56	15.77	7.78	5.17	1.83	1.17	1.52	1.30	2.73
92.20	120.85	79.89	58.47	75.99	45.56	33.97	16.37	22.63	12.99	9.16	4.69	1.68	0.93	1.58	1.66	4.04
107.20	120.15	72.66	58.45	78.01	45.35	30.58	13.68	18.67	12.09	8.16	5.42	1.76	0.92	1.54	1.44	4.42
95.10	124.05	71.24	58.28	70.58	45.96	26.63	11.62	16.27	11.01	8.91	5.84	1.56	0.98	1.41	1.19	3.73
95.22	112.69	67.37	59.80	59.40	44.15	24.99	11.73	17.61	11.16	9.83	5.00	1.47	0.93	1.61	1.41	3.38

6

Apple – Adjusted Stock Price 20 Years



7

Crime Rate

- In the last 18 years, has violent crime:
  - Increased?
  - Stayed about the Same?
  - Decreased?

8

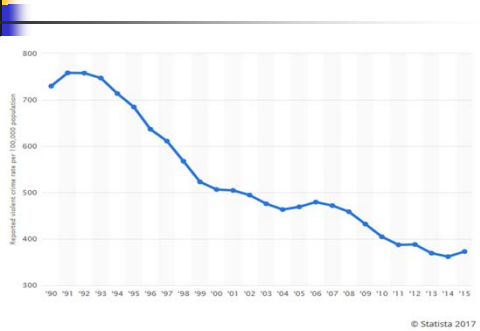
Perception – Gallup Poll

Is there more crime **in the U.S.** than there was a year ago, or less?

	More	Less	Same (vol.)	No opinion
	%	%	%	%
2016 Oct 5-9	70	20	6	4
2015 Oct 7-11	70	18	8	4
2014 Oct 12-15	63	21	9	7
2013 Oct 3-6	64	19	9	7
2011 Oct 6-9	68	17	8	8
2010 Oct 7-10	66	17	8	9
2009 Oct 1-4	74	15	6	5
2008 Oct 3-5	67	15	9	9
2007 Oct 4-7	71	14	8	6
2006 Oct 9-12	68	16	8	8
2005 Oct 13-16	67	21	9	3
2004 Oct 11-14	53	28	14	5
2003 Oct 6-8	60	25	11	4
2002 Oct 14-17	62	21	11	6
2001 Oct 11-14	41	43	10	6
2000 Aug 29-Sep 5	47	41	7	5
1998 Oct 23-25	52	35	8	5
1997 Aug 22-25	64	25	6	5
1996 Jul 25-28	71	15	8	6
1993 Oct 13-18	87	4	5	4
1992 Feb 28-Mar 1	89	3	4	4
1990 Sep 10	84	3	7	6

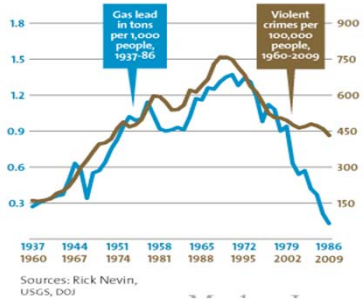
9

Reality - Reported Violent Crime Rate
(Source: US Justice Department)



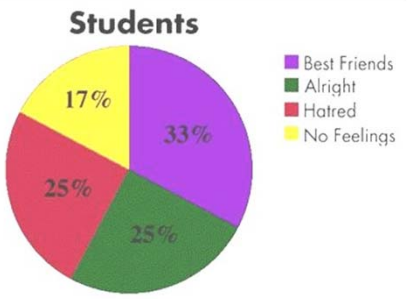
10

Line Graph - Crime and Lead



11

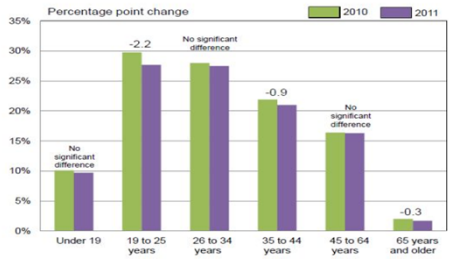
Pie Chart - What do you think of your College roommate?



12

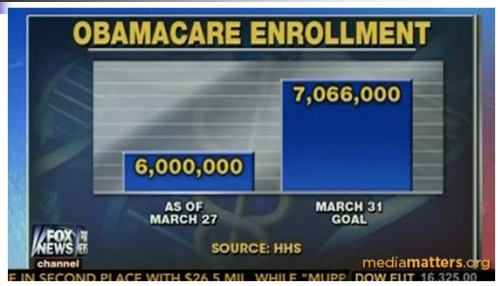
Bar Chart - Health Care

Uninsured Rates by Age: 2010 and 2011



13

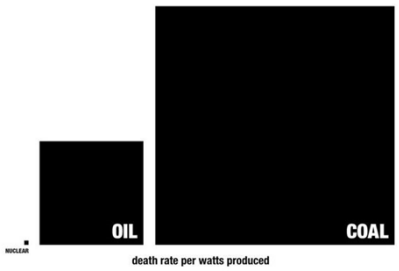
Distorting the truth with Statistics



14

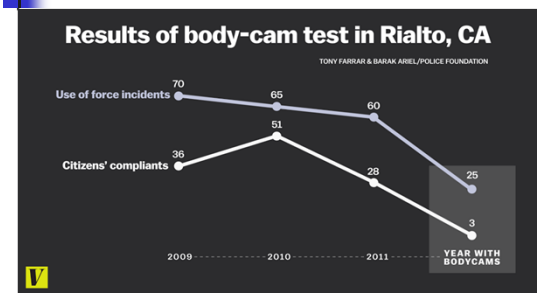
Nuclear, Oil and Coal Energy
Deaths per terawatt-hour produced

source: thebigfuture.com



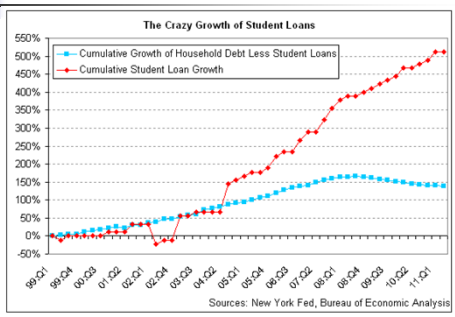
15

Should Police wear Body Cameras?



16

Increase in Debt since 1999

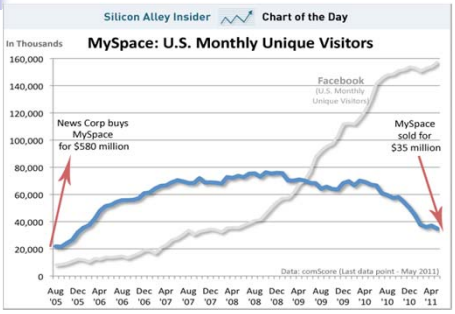


17

Most Popular Websites for College Students in 2007

18

Decline of MySpace



19

RATE MY PROFESSORS

Over 6,000 Schools, 1 million professors, 6 million opinions

De Anza College

Professor's Name	Department	Total Ratings	Overall Quality	Ease	Hot?
	Mandarin	3	4.3	2.0	
	Mandarin	8	1.6	1.6	
	Marketing	1	5.0	5.0	
	Mathematics	66	4.7	4.0	
	Mathematics	73	1.4	1.7	
	Mathematics	15	2.7	2.6	
	Mathematics	41	1.6	2.1	

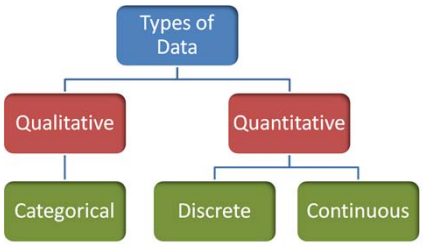
20

Types of Data

- Qualitative
  - Non-numeric
  - Always categorical
- Quantitative
  - Numeric
  - Categorical numbers are actually qualitative
  - Continuous or discrete

21

Types of Data



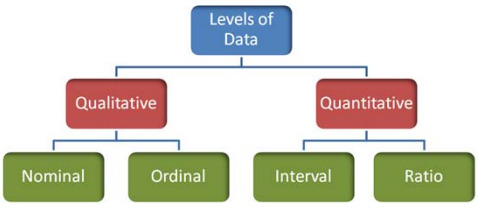
22

Levels of Data Measurement

- **Nominal:** Names or labels only
  - Example: What city do you live in?
- **Ordinal:** Data can be ranked, but no quantifiable difference.
  - Example: Ratings Excellent, Good, Fair, Poor
- **Interval:** Data can be ranked with quantifiable differences, but no true zero.
  - Example: Temperature
- **Ratio:** Data can be ranked with quantifiable differences and there is a true zero.
  - Example: Age

23

Levels of Data Measurement



24

Examples of Data

- Distance from De Anza College
- Number of Grandparents still alive
- Eye Color
- Amount you spend on food each week.
- Number of Facebook "Friends"
- Zip Code
- City you live in.
- Year of Birth
- How to prepare Steak? (rare, medium, well-done)
- Do you drive to De Anza?

25

Graphical Methods

- Qualitative Data
  - Pie Chart
  - Bar Chart
- Quantitative Data
  - Stem and Leaf Chart
  - Histogram
  - Ogive
  - Dot Plot

26

Graphing Categorical Data

A sample of 500 adults (age 18 and over) from Santa Clara County, California were taken from the year 2000 United States Census.

Marital Status	Frequency
Married	270
Widowed	22
Divorced - not remarried	42
Separated	10
Single - never married	156
Total	**500**

27

Graphing Categorical Data

- **n = sample size -** The number of observations in your sample size.
- **Frequency -** the number of times a particular value is observed.
- **Relative frequency -** The proportion or percentage of times a particular value is observed.
- **Relative Frequency = Frequency / n**

28

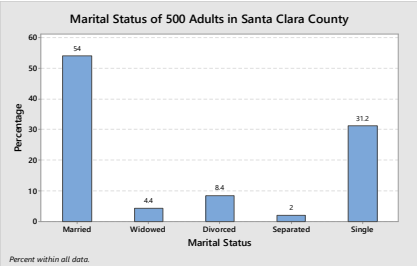
Graphing Categorical Data

A sample of 500 adults (age 18 and over) from Santa Clara County, California were taken from the year 2000 United States Census.

Marital Status	Frequency	Relative Frequency
Married	270	270/500 = 0.540 or 54.0%
Widowed	22	22/500 = 0.044 or 4.4%
Divorced - not remarried	42	42/500 = 0.084 or 8.4%
Separated	10	10/500 = 0.020 or 2.0%
Single - never married	156	156/500 = 0.312 or 31.2%
Total	**500**	**500/500 = 1.000 or 100.0%**

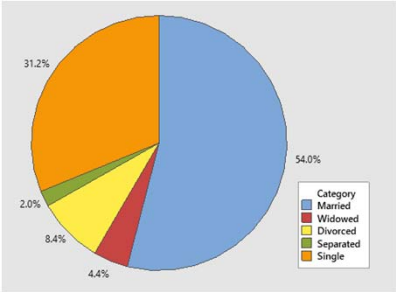
29

Bar Graph of Categorical Data



30

Pie Chart of Categorical Data



31

Daily Minutes spent on the Internet by 30 students

102	104	85	67	101
71	116	107	99	82
103	97	105	103	95
105	99	86	87	100
109	108	118	87	125
124	112	122	78	92

32

Describing Numeric Data

- Center?
  - Where is an "average" value
- Spread?
  - How far are data spread from the center
- Shape?
  - Symmetric or skewed?
- Anything Unusual?
  - Outliers, more than 1 peak?

33

Stem and Leaf Graph

```
 6  7
 7  18
 8  25677
 9  25799
10  01233455789
11  268
12  245
```

34

Back-to-back Example

- Passenger loading times for two airlines

11, 14, 16, 17, 19, 21, 22, 23, 24, 24, 24, 26, 31, 32, 38, 39	8, 11, 13, 14, 15, 16, 16, 18, 19, 19, 21, 21, 22, 24, 26, 31

35

Back to Back Example

	0	
	0	8
1 4	1	1 3 4
6 7 9	1	5 6 6 8 9 9
1 2 3 4 4 4	2	1 1 2 4
6	2	6
1 2	3	1
8 9	3	

36

Grouping Data

- Choose the number of groups
  - between 5 and 10 is best
- Interval Width = (Range+1)/(Number of Groups)
  - Round **up** to a convenient value
- Start with lowest value and create the groups.
- Example – for 5 categories
  Interval Width = (58+1)/5 = 12 (rounded up)

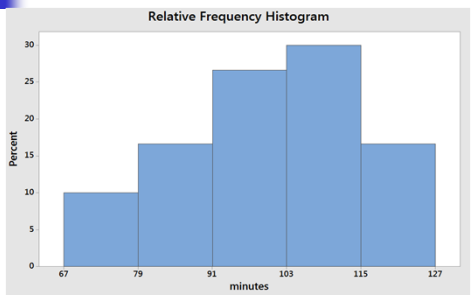
37

Grouping Data

Class Interval	Frequency	Relative Frequency
67 to 79	3	0.100 or 10.0%
79 to 91	5	0.167 or 16.7%
91 to 103	8	0.266 or 26.6%
103 to 115	9	0.300 or 30.0%
115 to 127	5	0.167 or 16.7%
Total	30	**1.000 or 100%**

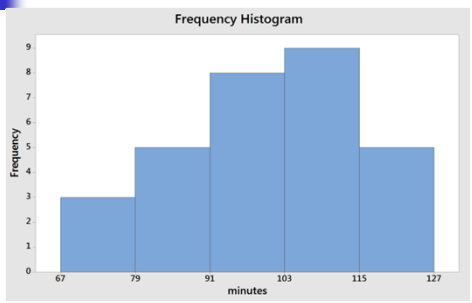
38

Histogram – Relative Frequency



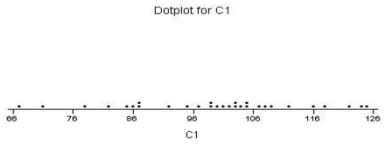
39

Histogram –Frequency



40

Dot Plot – Graph of Frequency



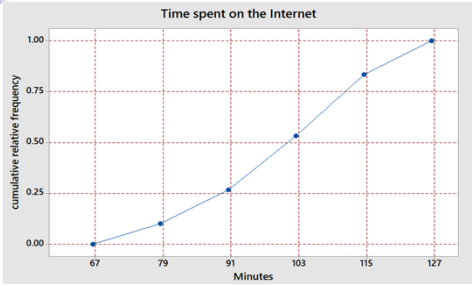
41

Cumulative Relative Frequency

Class Interval	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
67 to 79	3	0.100 or 10.0%	3	0.100 or 10.0%
79 to 91	5	0.167 or 16.7%	8	0.267 or 26.7%
91 to 103	8	0.266 or 26.6%	16	0.533 or 53.3%
103 to 115	9	0.300 or 30.0%	25	0.833 or 83.3%
115 to 127	5	0.167 or 16.7%	30	1.000 or 100%
Total	30	**1.000 or 100%**		

42

Ogive – Graph of Cumulative Relative Frequency



43