

Data Visualization

Statistical Charts

Components of Statistics

- A general process of investigation:
 - 1. Identify a question or problem.
 - 2. Collect relevant data on the topic.
 - 3. Analyze the data.
 - 4. Form a conclusion.
- **Statistics** is the study of how best to collect, analyze, and draw conclusions from data (stages 2-4).
 - How best can we collect data?
 - How should it be analyzed?
 - What can we infer from the analysis?

Data Matrix

Variable

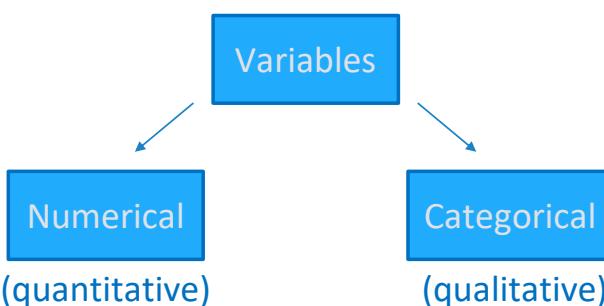
↓

email	spam	num_char	line_breaks	format	number
1	No	21705	551	html	small
2	No	7011	183	html	big
3	Yes	631	28	text	none
⋮	⋮	⋮	⋮	⋮	⋮
50	No	15829	242	html	small

← *Observation
(case)*

3

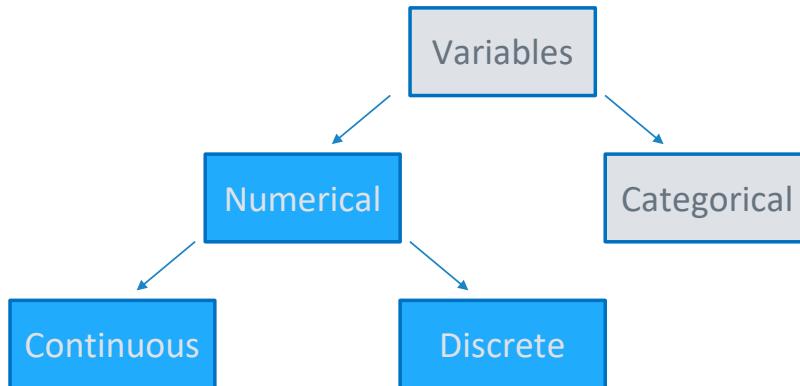
Types of Variables



- take on numerical values
 - sensible to add, subtract, take averages, etc. with these values
- take on a limited number of distinct categories.
 - categories can be identified with numbers, but not sensible to do arithmetic operations

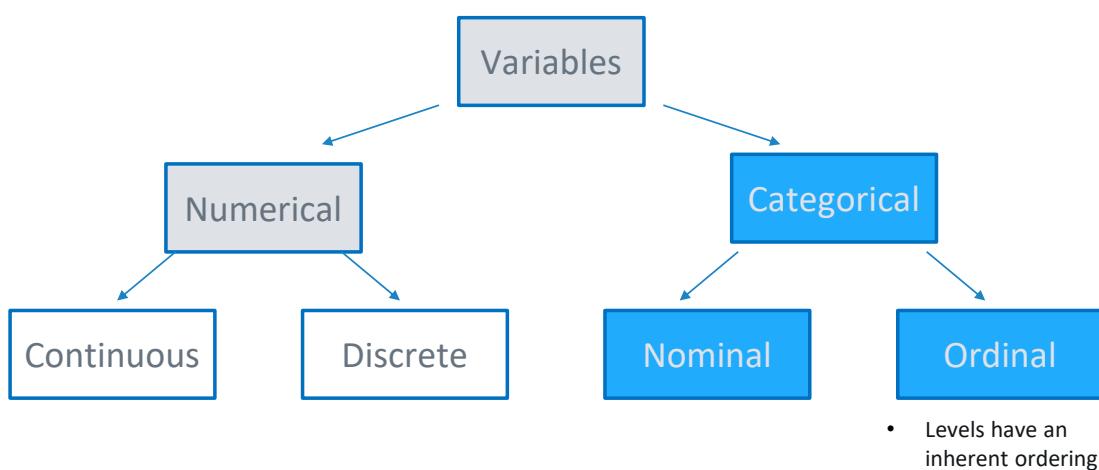
4

Numerical Variables



5

Categorical Variable



6

Example

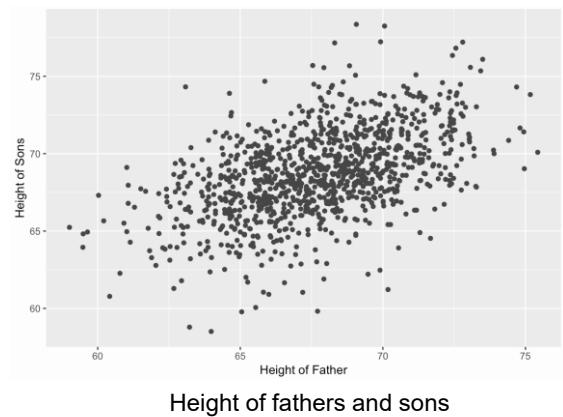
email	spam	num_char	line_breaks	format	number
1	No	21705	551	html	small
2	No	7011	183	html	big
3	Yes	631	28	text	none
:	:	:	:	:	:
50	No	15829	242	html	small

↓ ↓ ↓ ↓ ↓ ↓
 Identity Nominal Categorical Discrete Numerical Discrete Numerical Nominal Categorical Ordinal Categorical

7

Relationships between variables

- Two variables that show some connection with one another are called **associated**.
- Association can be further described as **positive** or **negative**.
- If two variables are not associated, they are said to be **independent**.



8

Population and Sample

9

Population

- Each research question refers to a target **population**.
- Example:
 - **Research question:** Can adult men become better, more efficient runners on their own, merely by running?
 - **Population of interest:** All men over 18
- Often it is too expensive to collect data for every case in a population.

10

Census

- Census: collect data from *everyone* in the population.

هزینه سرشماری سال 1395 پنج هزار میلیارد ریال است/ آغاز سرشماری نفوس از سوم مهر



11



12



13

Sampling

- A **sample** represents a subset of the cases and is often a small fraction of the population.
- Think about sampling something you are cooking: you taste a small part of what you're cooking to get an idea about the dish as a whole.
- If you generalize and conclude that your entire soup needs salt, that's an **inference**.



14

Anecdotal Evidence

- Consider the following statements:
 - My uncle smokes three packs a day and he's in perfectly good health, so smoking doesn't affect your health.
- The conclusion is based on data, but there are two problems:
 - First, the data only represent one or two cases.
 - Second, it is unclear whether these cases are actually representative of the population.
- Data collected in this haphazard fashion are called **anecdotal evidence**.

15

Sampling Bias



16

Some Sources of Sampling Bias

- *Non-response*: If only a *non-random* fraction of the randomly sampled people choose to respond to a survey, the sample may no longer be representative of the population.
- *Voluntary response*: Occurs when the sample consists of people who volunteer to respond because they have strong opinions on the issue.
- *Convenience sample*: Individuals who are easily accessible are more likely to be included in the sample.

17

Sampling Bias Example

- A historical example of a biased sample yielding misleading results:



Alf Landon

- In 1936, Landon sought the Republican presidential nomination opposing the re-election of FDR.

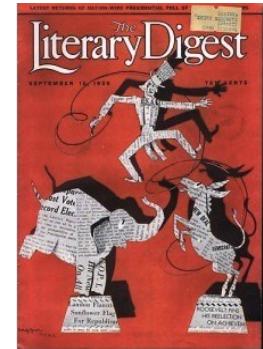


Franklin D. Roosevelt

18

The Literary Digest Poll

- The Literary Digest polled 10 million Americans, and got responses from about 2.4 million.
- The poll showed that Landon would likely be the winner and FDR would get 43% of the votes.
- Election result: FDR won, with 62% of the votes.
- The magazine was completely discredited because of the poll, and was soon discontinued.



19

What went wrong?

- The magazine had surveyed:
 - its own readers
 - registered automobile owners, and registered telephone users
- These groups had incomes well above the national average of the day which resulted in lists of voters far more likely to support Republicans than a truly *typical* voter of the time.
- The Literary Digest election poll was based on a sample size of 2.4 million, which is huge, but since the sample was *biased*, the sample did not yield an accurate prediction.

20

Type of Studies

21

Explanatory and Response Variables

- To identify the explanatory variable in a pair of variables, identify which of the two is suspected of affecting the other:

Explanatory variable $\xrightarrow{\text{might affect}}$ Response variable

- Labeling variables as explanatory and response does not guarantee the relationship between the two is actually causal, even if there is a high correlation between the two variables.
- We use these labels only to keep track of which variable we suspect affects the other.

22

Observational Studies & Experiments

-
- ```

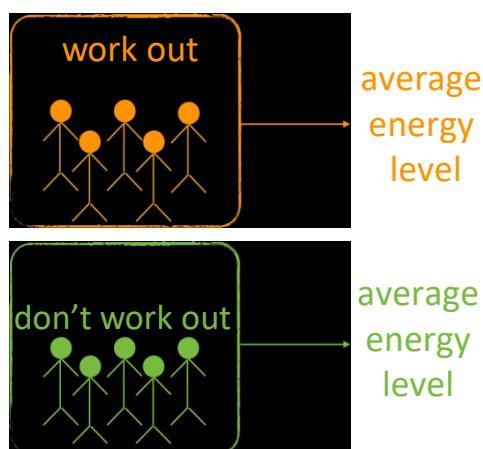
graph TD
 Studies[Studies] --> Observational[Observational]
 Studies --> Experimental[Experimental]

```
- collect data in a way that does not directly interfere with how the data arise ("observe")
  - only establish an association
  - **retrospective**: uses past data
  - **prospective**: data are collected throughout the study
  - randomly assign subjects to treatments
  - establish causal connections between explanatory and response variables.

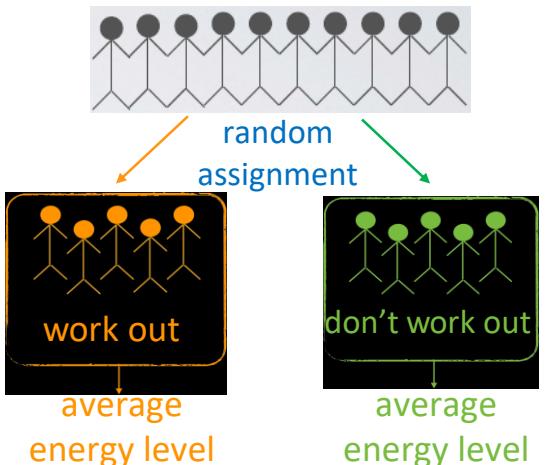
23

## Observational vs. Experimental Studies

### Observational Study



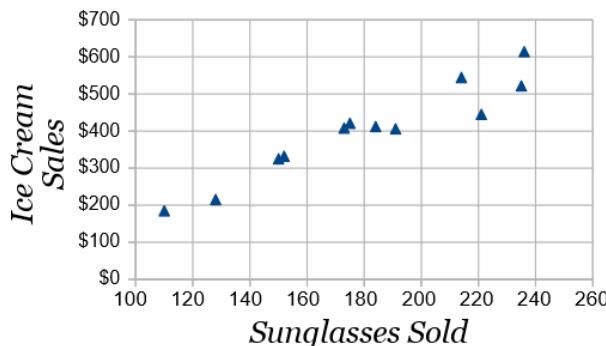
### Experiment



24

# Correlation does **not** imply causation

- The local ice cream shop keeps track of how much ice cream they sell.
- The ice cream shop finds how many sunglasses were sold by a big store for each day and compares them to their ice cream sales.



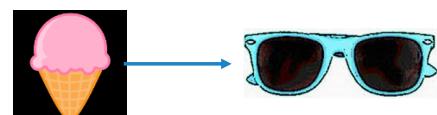
25

## Three possible explanations

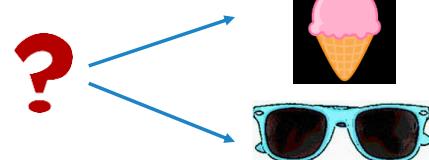
1. Sunglasses make people want ice cream!



2. Eating ice cream makes people buy sunglasses!



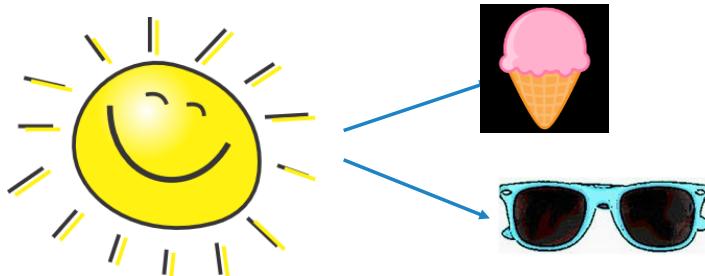
3. A third variable is responsible for both.



26

# Confounding Variable

- An extraneous variable that affects both the explanatory and the response variable and that make it seem like there is a relationship between the two are called **confounders** or **confounding variables**.



27

# MMR Vaccination and Autism

THE LANCET

EARLY REPORT | VOLUME 351, ISSUE 9103, P637-641, FEBRUARY 28, 1998

RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

Dr AJ Wakefield, FRCS & + SH Murch, MB + A Anthony, MB + J Linnell, PhD + DM Casson, MRCP - M Malik, MRCP - et al  
Show all authors

Published: February 28, 1998 DOI: [https://doi.org/10.1016/S0140-6736\(97\)11096-0](https://doi.org/10.1016/S0140-6736(97)11096-0)

PDF [942 KB] Figures Save Share Reprints Request

PlumX Metrics

Request Your Institutional Access

**Summary**

**Background**

We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorders.

**Methods**

12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent

**RETRACTED**

28

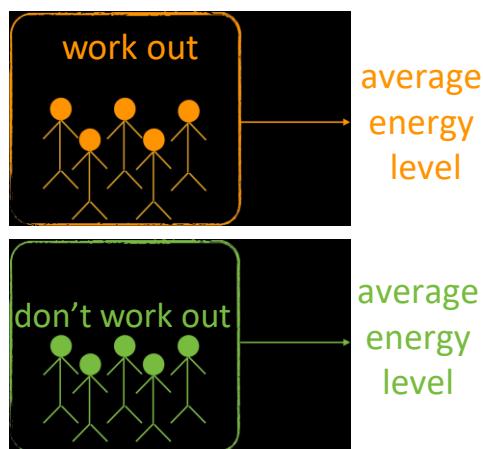
# Do popes live longer?

The screenshot shows the BBC More or Less website. At the top, there's a red header with the BBC logo, a search bar, and a 'Menu' button. Below the header, the title 'More or Less' is displayed, followed by a large image of Pope Francis smiling. A 'Listen now' button is overlaid on the image. To the right of the image, text reads 'Last on Mon 18 Apr 2016 22:20 Local time BBC WORLD SERVICE ANR'. Below this, there are 'More episodes' sections for 'PREVIOUS The story of average' and 'NEXT Most Expensive Building', each with a 'Listen now' button. At the bottom, it says 'See all episodes from More or Less'.

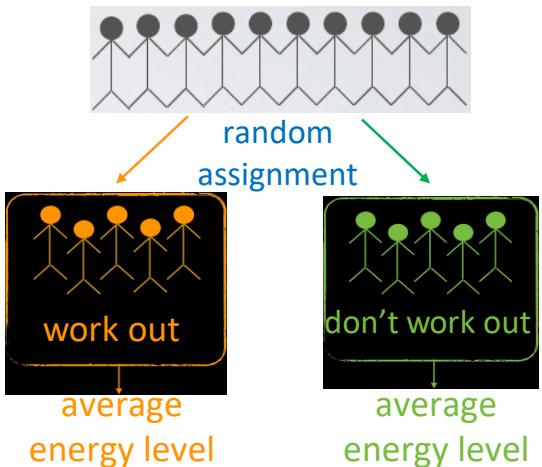
29

## Observational vs. Experimental Studies

### Observational Study



### Experiment



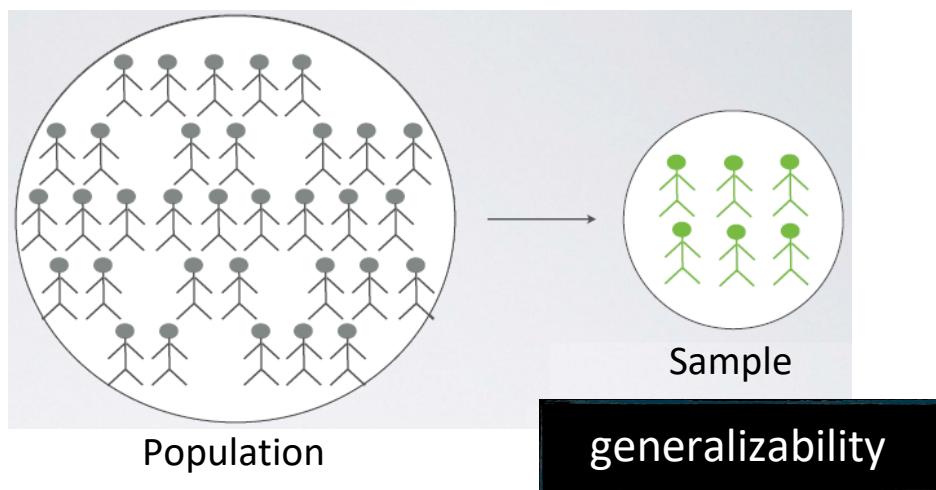
30

# Principles of Experimental Design

- *Control*: Compare treatment of interest to a control group.
- *Randomize*: Randomly assign subjects to treatments, and randomly sample from the population whenever possible.
- *Replicate*: Within a study, replicate by collecting a sufficiently large sample. Or replicate the entire study.
- *Block*: If there are variables that are known or suspected to affect the response variable, first group subjects into *blocks* based on these variables, and then randomize cases within each block to treatment groups.

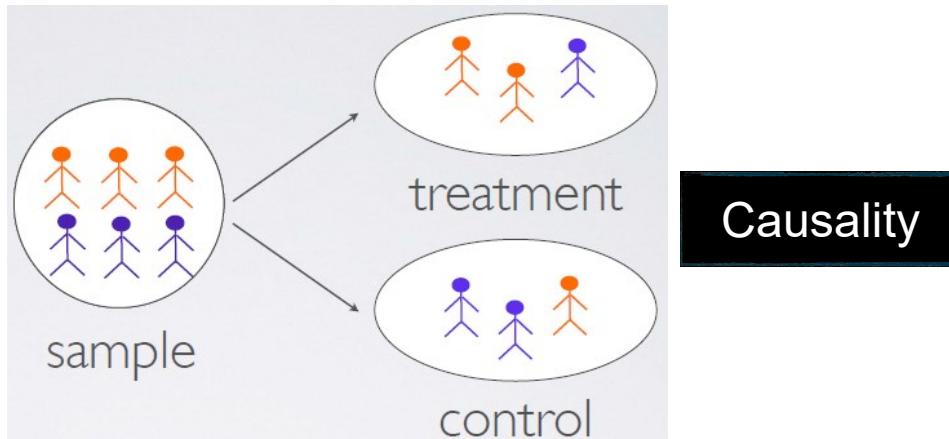
31

## Random Sampling



32

# Random Assignment



33

## Random Assignment vs. Random Sampling

| <i>ideal experiment</i> | Random assignment                                       | No random assignment                                                             | <i>most observational studies</i> |
|-------------------------|---------------------------------------------------------|----------------------------------------------------------------------------------|-----------------------------------|
| Random sampling         | Causal conclusion, generalized to the whole population. | No causal conclusion, correlation statement generalized to the whole population. | <i>Generalizability</i>           |
| No random sampling      | Causal conclusion, only for the sample.                 | No causal conclusion, correlation statement only for the sample.                 | <i>No generalizability</i>        |
| <i>most experiments</i> | Causation                                               | Correlation                                                                      | <i>bad observational studies</i>  |

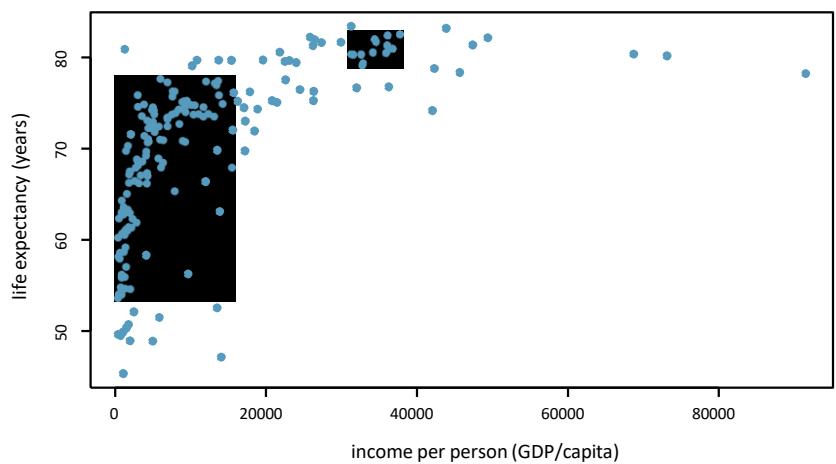
34

# Visualizing Numerical Data

35

## Scatterplot

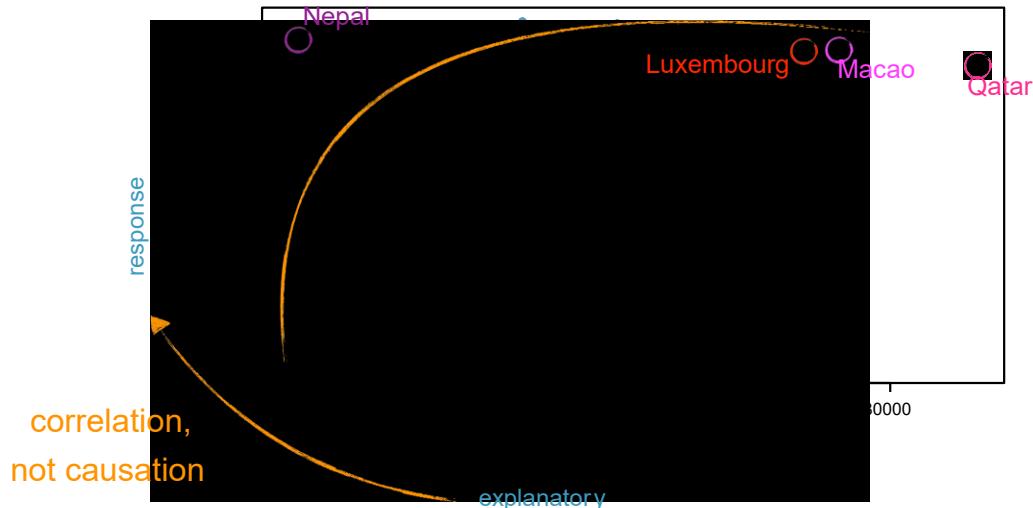
| data        | income /person | life expectancy |
|-------------|----------------|-----------------|
| Afghanistan | 1359.7         | 60.254          |
| Albania     | 6969.3         | 77.185          |
| Algeria     | 6419.1         | 70.874          |
| :           | :              | :               |
| Zimbabwe    | 545.3          | 58.142          |



- *Scatterplots* are useful for visualizing the relationship between two numerical variables.

36

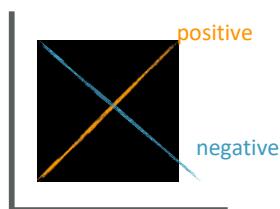
# Scatterplot



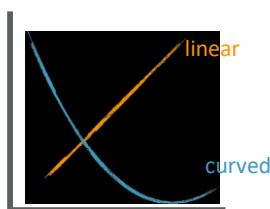
37

## Evaluating the relationship

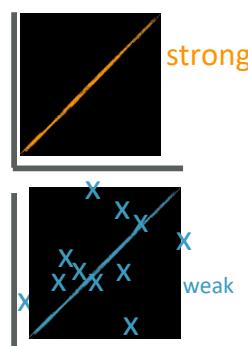
direction



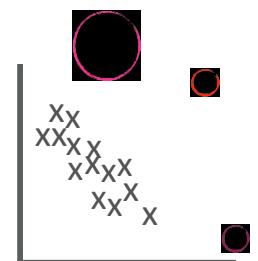
shape



strength



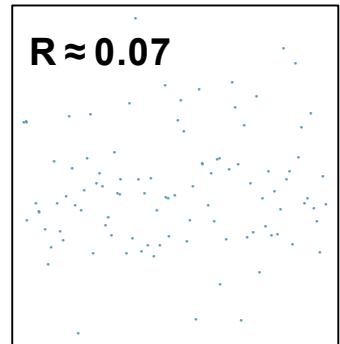
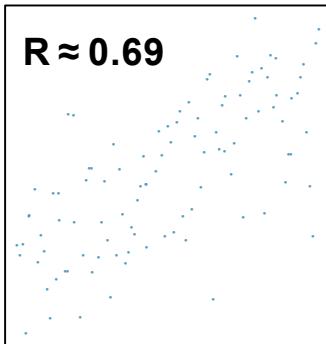
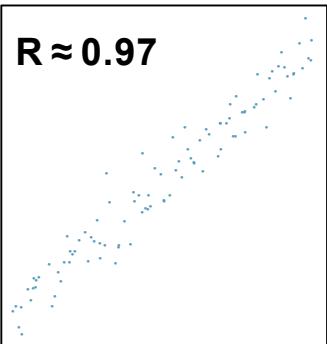
outliers



38

# Correlation Coefficient

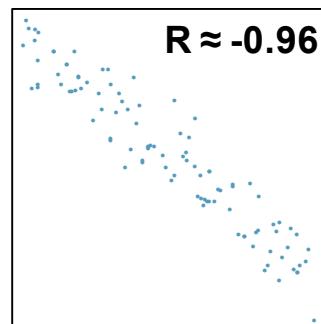
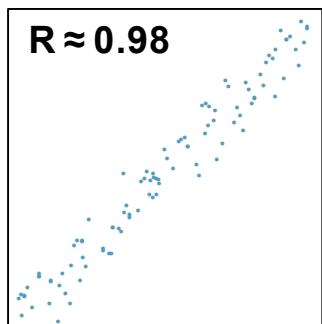
- Describes the strength of the **linear association** between two variables and is denoted as  $R$  or  $\rho$ .



39

# Correlation Coefficient

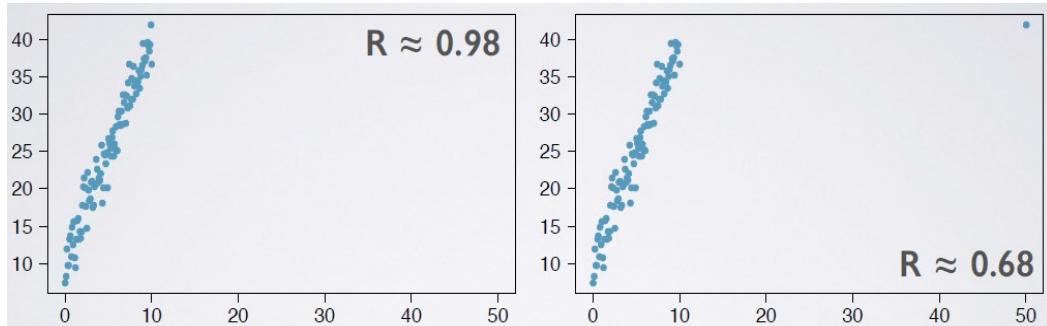
- The sign of the correlation coefficient indicates the direction of association.



40

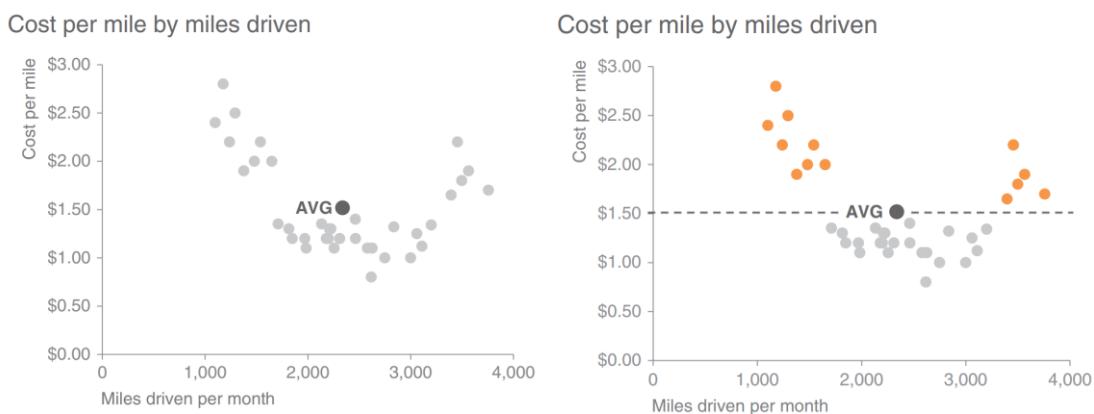
# Correlation Coefficient

- The correlation coefficient is sensitive to outliers.



41

# Modified Scatterplot

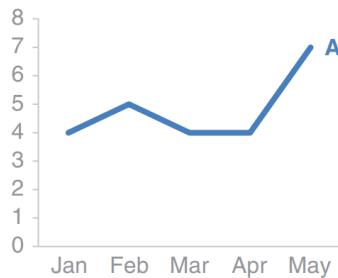


42

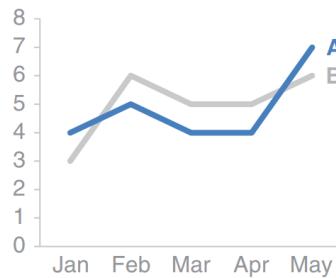
# Line Graph

- Line graphs are used to plot continuous data often in some unit of time.

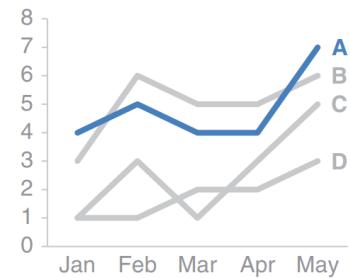
Single series



Two series



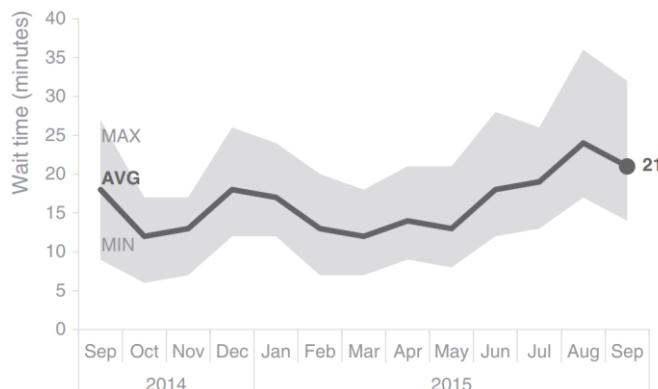
Multiple series



43

# Line Graph

Passport control wait time  
Past 13 months

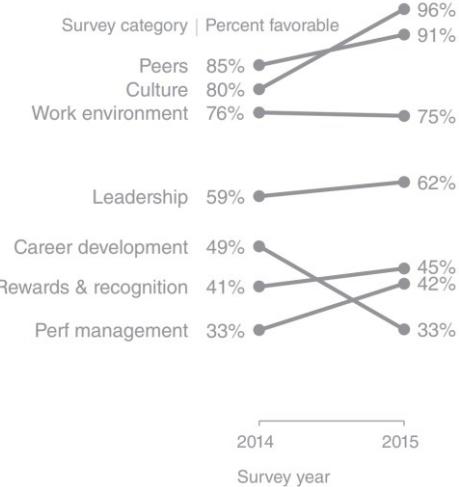


Showing average within a range in a line graph

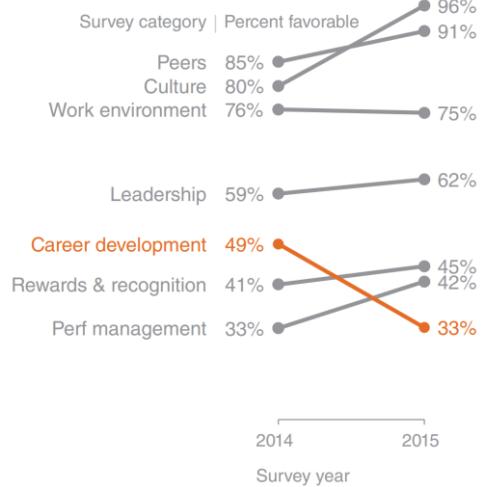
44

# Slopegraph

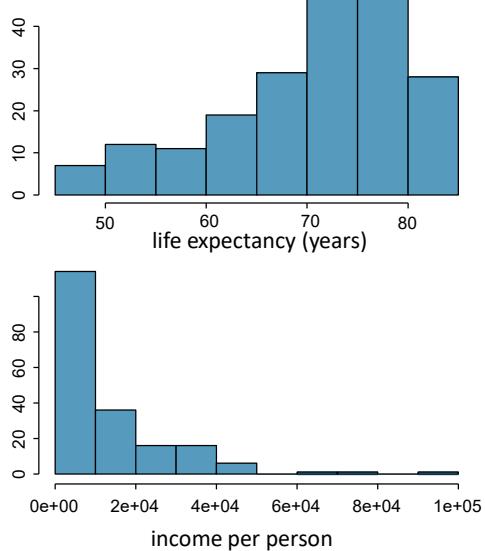
Employee feedback over time



Employee feedback over time



# Histogram

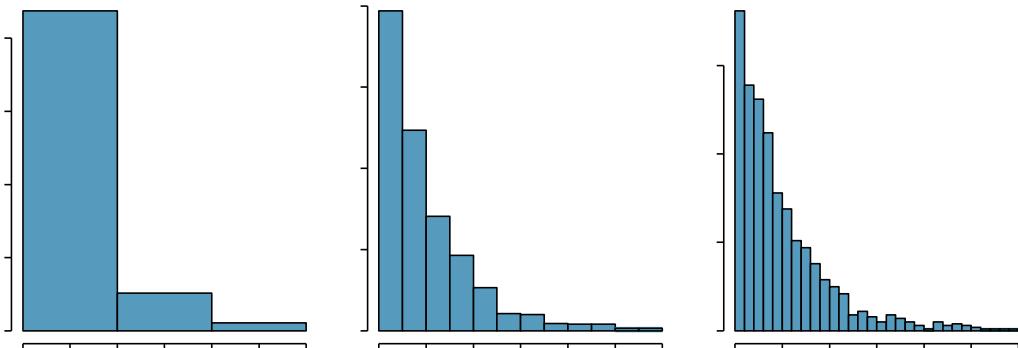


- Histograms provide a view of the **data density**.
- Histograms are especially convenient for describing the **shape** of the data distribution.
- The chosen **bin width** can alter the story the histogram is telling

46

# Bin Width

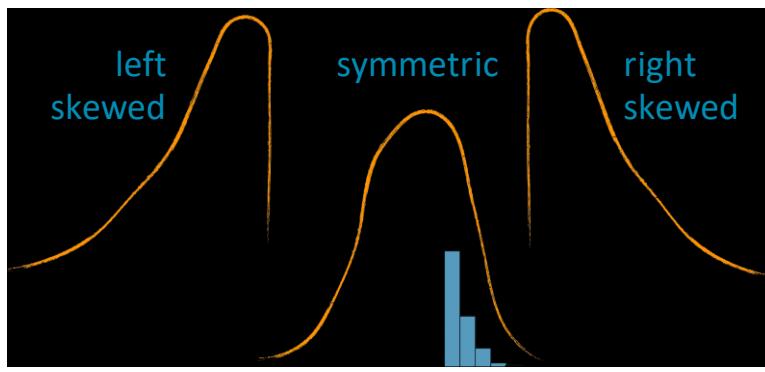
- When the bin width is too wide, we might lose interesting details.
- When the bin width is too narrow, it might be difficult to get an overall picture of the distribution.



47

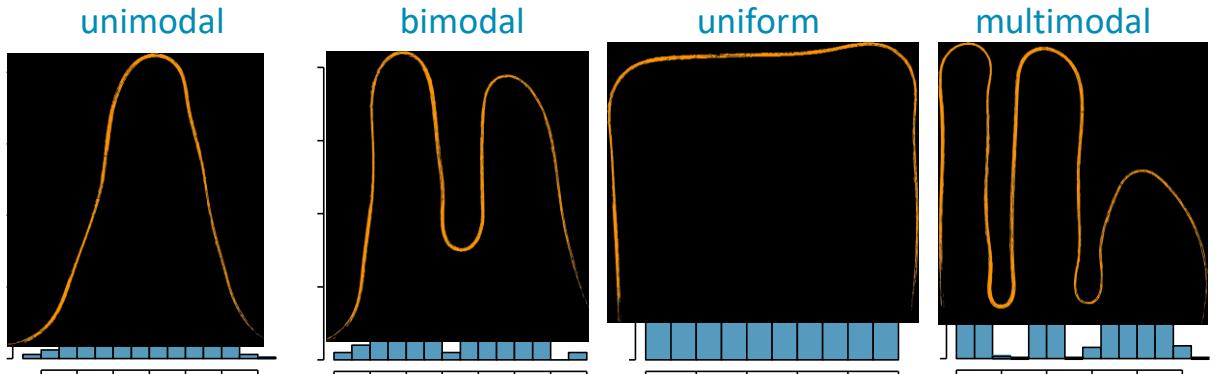
# Skewness

- Distributions are skewed to the side of the long tail



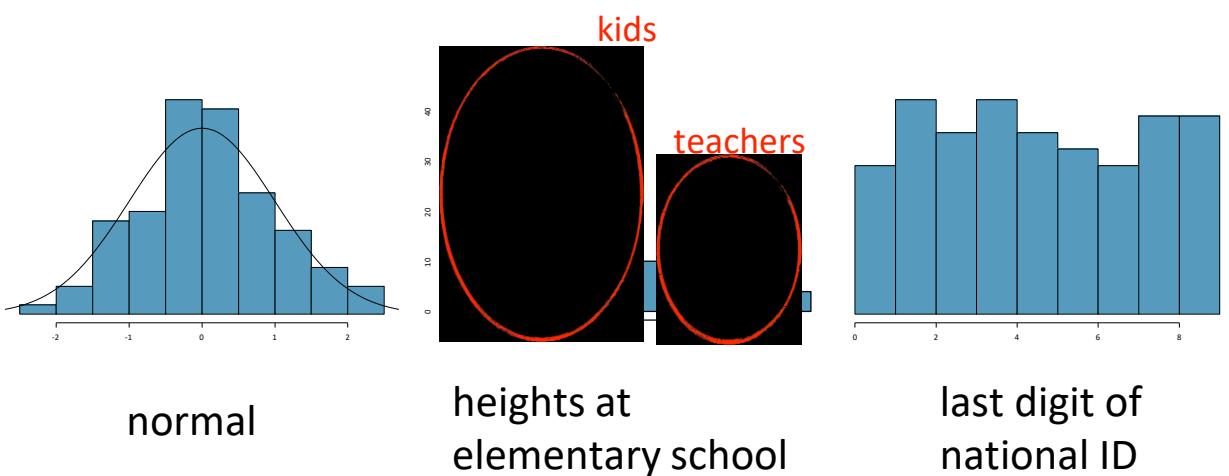
48

# Modality



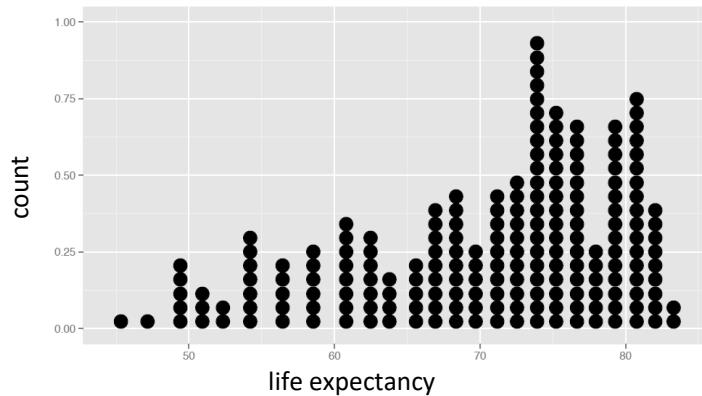
49

# Modality



50

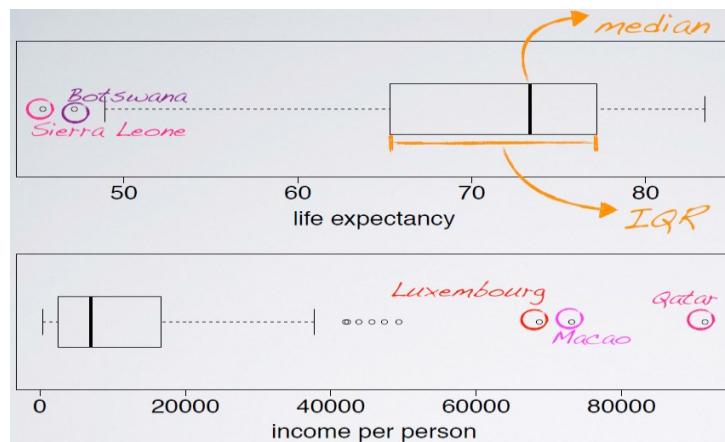
# Dotplot



- Useful when individual values are of interest
- Can get busy as the sample size increases

51

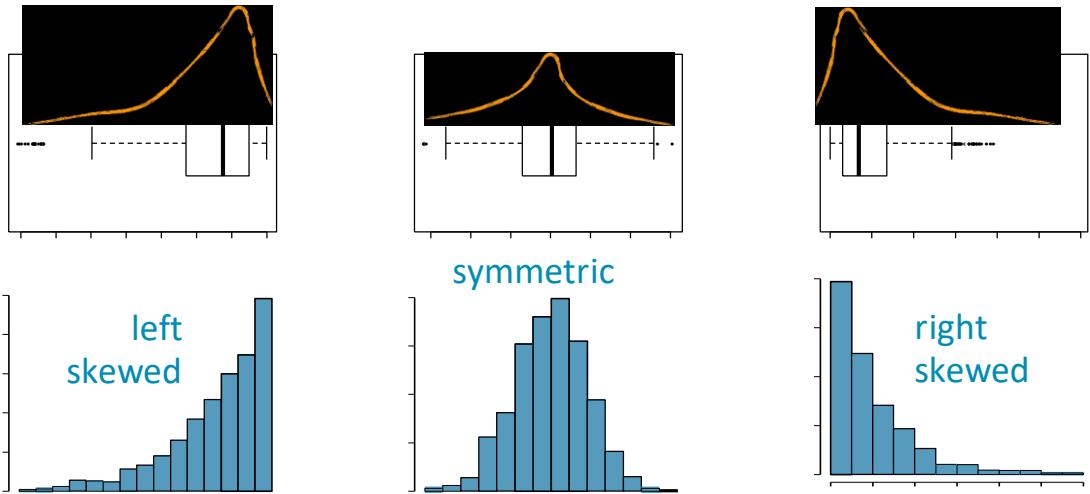
# Box plot



- Useful for highlighting outliers, median, IQR.

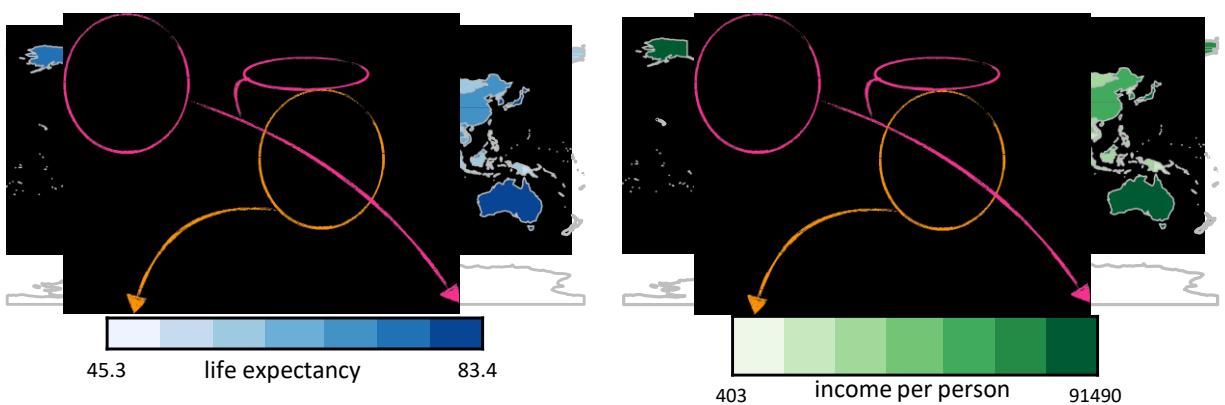
52

## Determining the skewness from a box plot



53

## Intensity Map

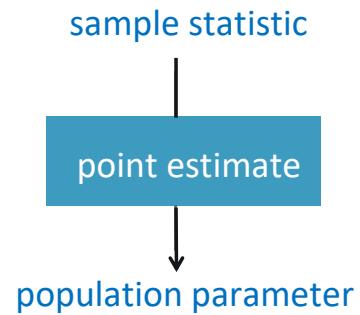


- Useful for highlighting the spatial distribution.

54

# Measures of Center

- **Mean:** arithmetic average
  - Sample mean:  $\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$
  - Population mean:  $\mu$
- **Median:** midpoint of the distribution
  - 50th percentile
- **Mode:** most frequent observation



## Example

- Nine students exam score:

75, 69, 88, 93, 95, 54, 87, 88, 27

$$\text{mean: } \frac{75+69+88+93+95+54+87+88+27}{9} = 75.11$$

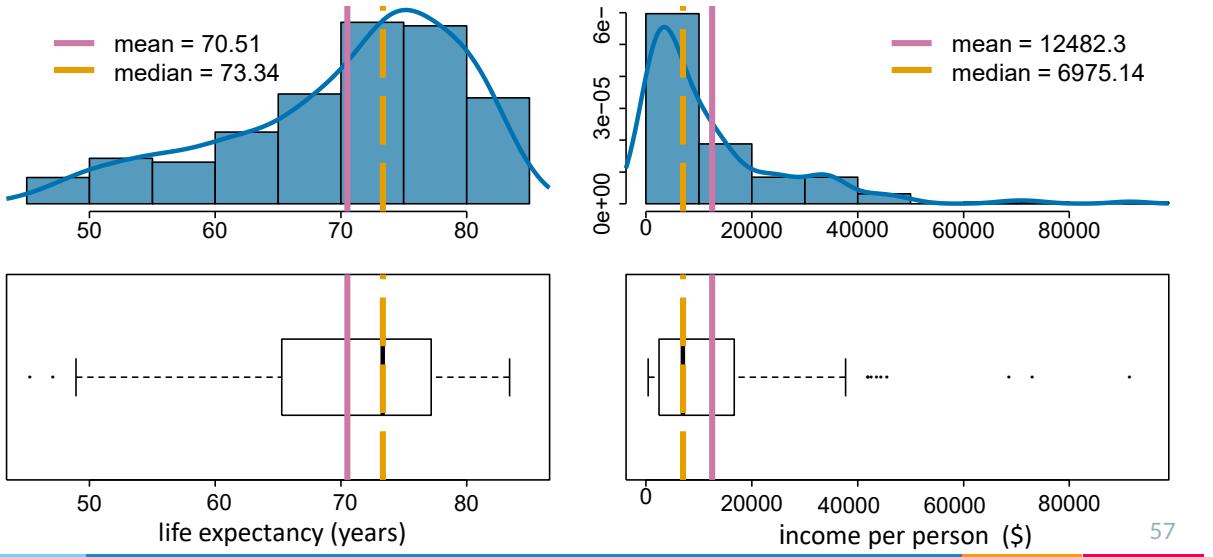
mode: 88

median: 27, 54, 69, 75, 88, 88, 93, 95

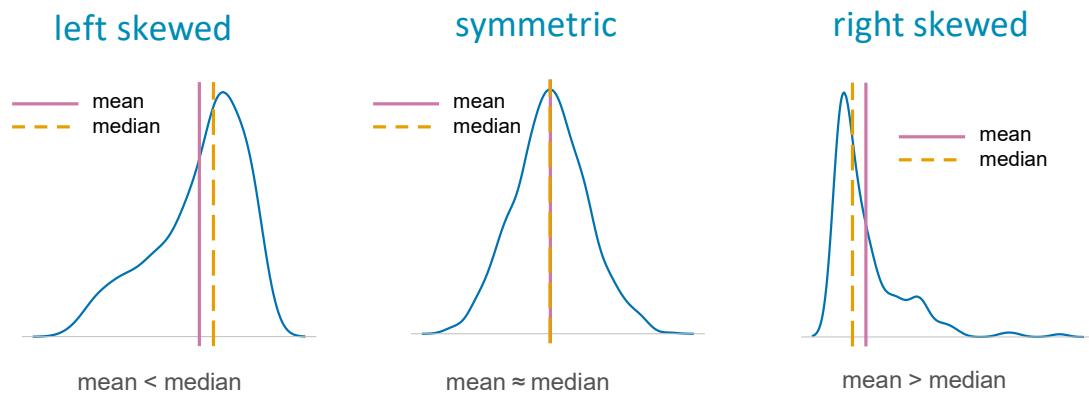
55

56

# Relation between Mean and Median



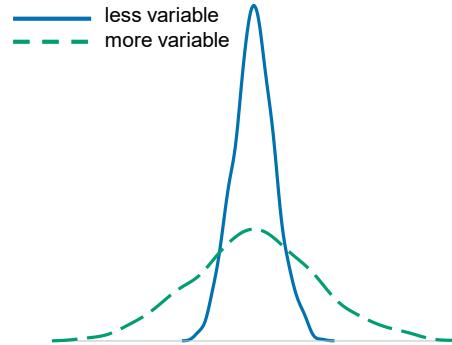
# Skewness vs. Measures of Center



# Measures of Spread

- In other words, statistics that tell us about the variability in the data:

- Range =  $(\max - \min)$
- Variance
- Standard deviation
- Inter-quartile range



59

## Variance

- Variance:** roughly the average squared deviation from the mean
  - Sample variance:  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
  - Population variance:  $\sigma^2$
- Example:** Given that the average life expectancy is 70.5, and there are 201 countries in the dataset:

$$s^2 = \frac{(60.3 - 70.5)^2 + (77.2 - 70.5)^2 + \dots + (58.1 - 70.5)^2}{201 - 1}$$

$$= 83.06 \text{ years}^2$$

|     | data        | life expectancy |
|-----|-------------|-----------------|
| 1   | Afghanistan | 60.254          |
| 2   | Albania     | 77.185          |
| 3   | Algeria     | 70.874          |
| :   | :           | :               |
| 201 | Zimbabwe    | 58.142          |

60

# Standard Deviation

- **Standard deviation:** roughly the average deviation from the mean that has the same units as the data

- Sample standard deviation:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

square root of  
the variance

- Population standard deviation:  $\sigma$

- **Example:** Given that the average life expectancy is 70.5, and there are 201 countries in the dataset:

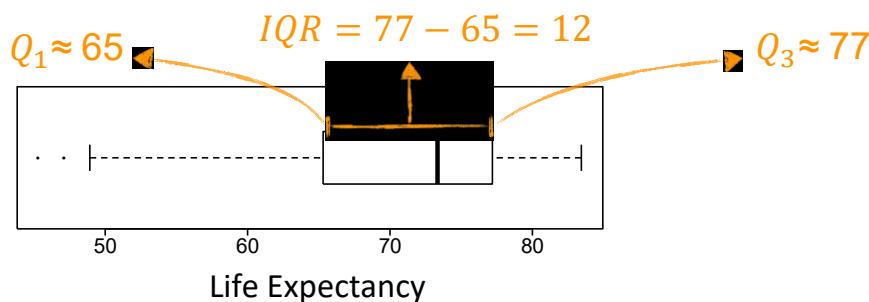
$$s = \sqrt{83.06} = 9.11 \text{ years}$$

61

# Interquartile Range

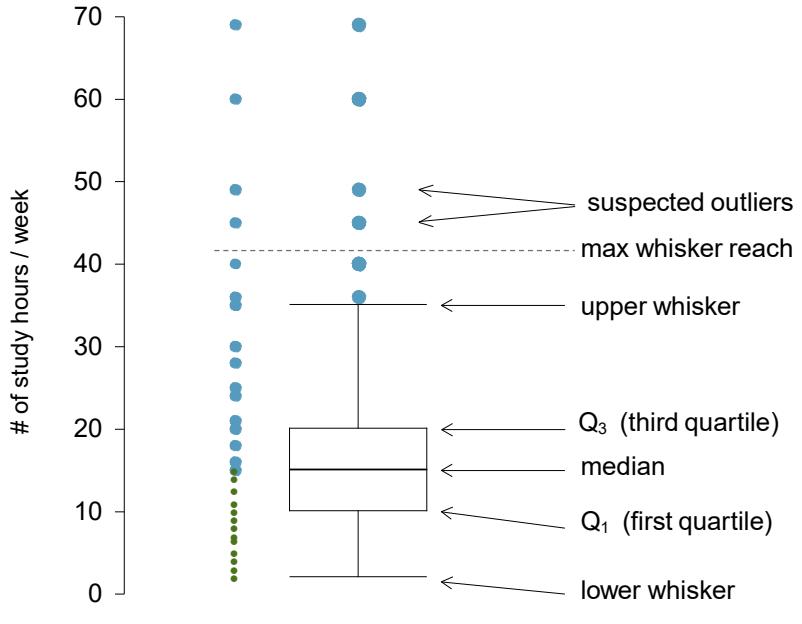
- Range of the middle 50% of the data, distance between the first quartile (25th percentile) and third quartile (75th percentile):

$$IQR = Q_3 - Q_1$$



62

# Boxplot



63

# Whiskers

- The **whiskers** attempt to capture the data outside of the box, however, their reach is never allowed to be more than  $1.5 \times IQR$ :
 
$$\text{max upper whisker reach} = Q_3 + 1.5 \times IQR$$

$$\text{max lower whisker reach} = Q_1 - 1.5 \times IQR$$
- Example:
 
$$\text{IQR : } 20 - 10 = 10$$

$$\text{max upper whisker reach} = 20 + 1.5 \times 10 = 35$$

$$\text{max lower whisker reach} = 10 - 1.5 \times 10 = -5$$
- A potential **outlier** is defined as an observation beyond the maximum reach of the whiskers.
  - An observation that appears extreme relative to the rest of the data.

64

# Outliers

- Why it is important to look for outliers?
- Examination of data for possible outliers serves many useful purposes, including:
  1. Identifying strong skew in the distribution.
  2. Identifying data collection or entry errors.
  3. Providing insight into interesting properties of the data.

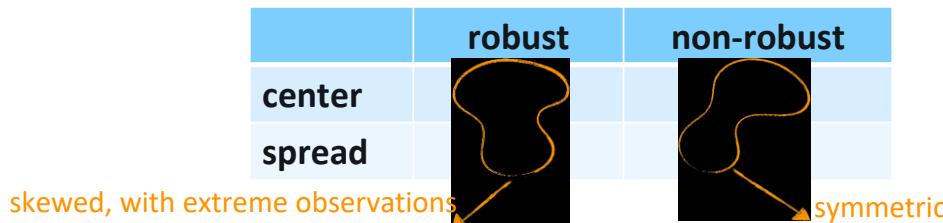
65

# Robust Statistics

- We define **robust statistics** as measures on which extreme observations have little effect.

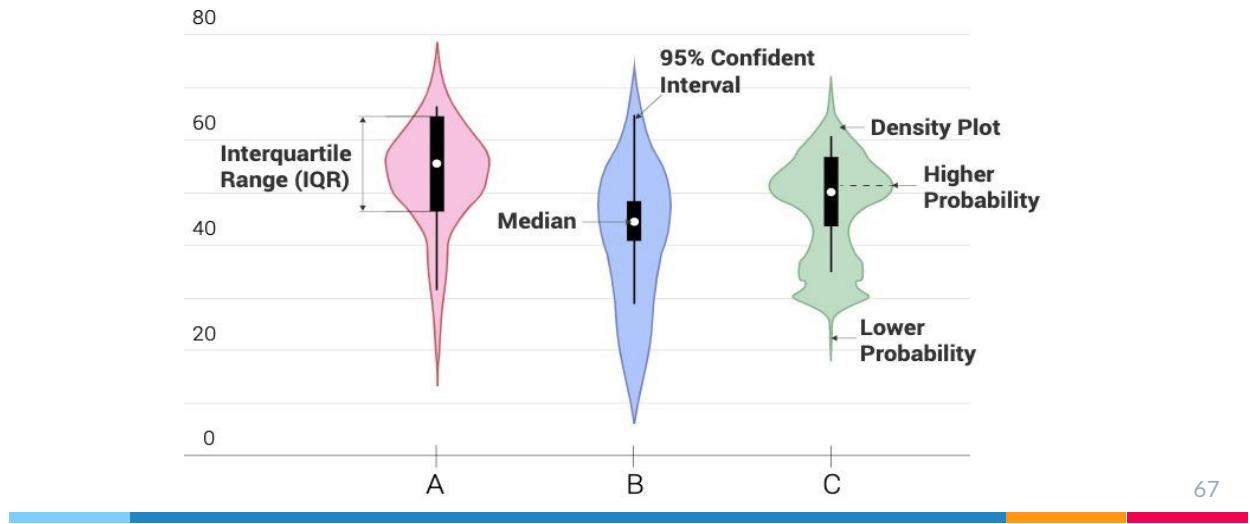
- Example:

|  | Data                | Mean | Median |
|--|---------------------|------|--------|
|  | 1, 2, 3, 4, 5, 6    | 3.5  | 3.5    |
|  | 1, 2, 3, 4, 5, 1000 | 169  | 3.5    |

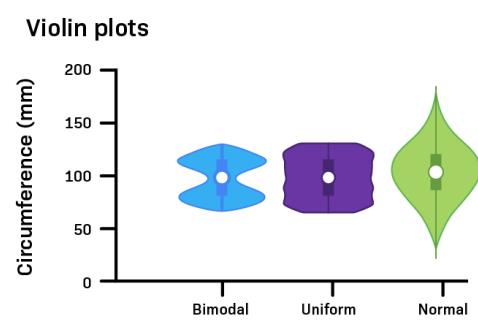
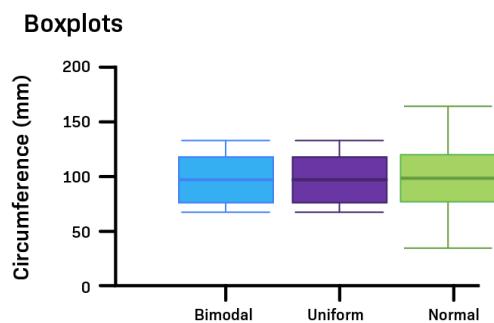


66

# Violin Plot



# Violin Plot vs. Box Plot



68

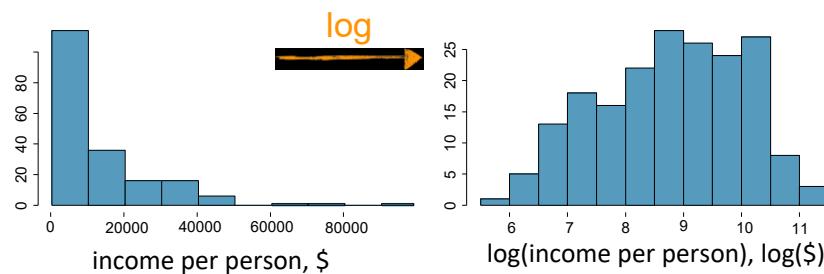
# Data Transformation

- A **transformation** is a rescaling of the data using a function.
  - Log transformation
  - Square root transformation
  - Inverse transformation
- When data are very strongly skewed, we sometimes transform them so they are easier to model.

69

## Log Transformation

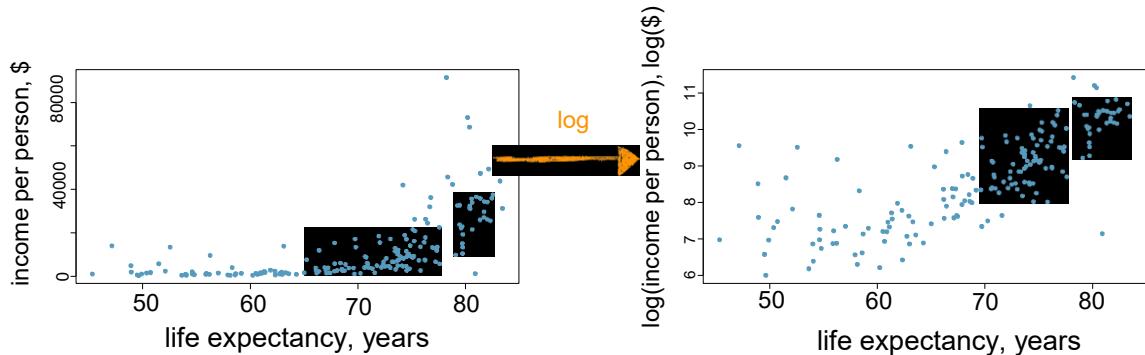
- Often applied when much of the data cluster near zero (relative to the larger values in the data set) and all observations are positive.



70

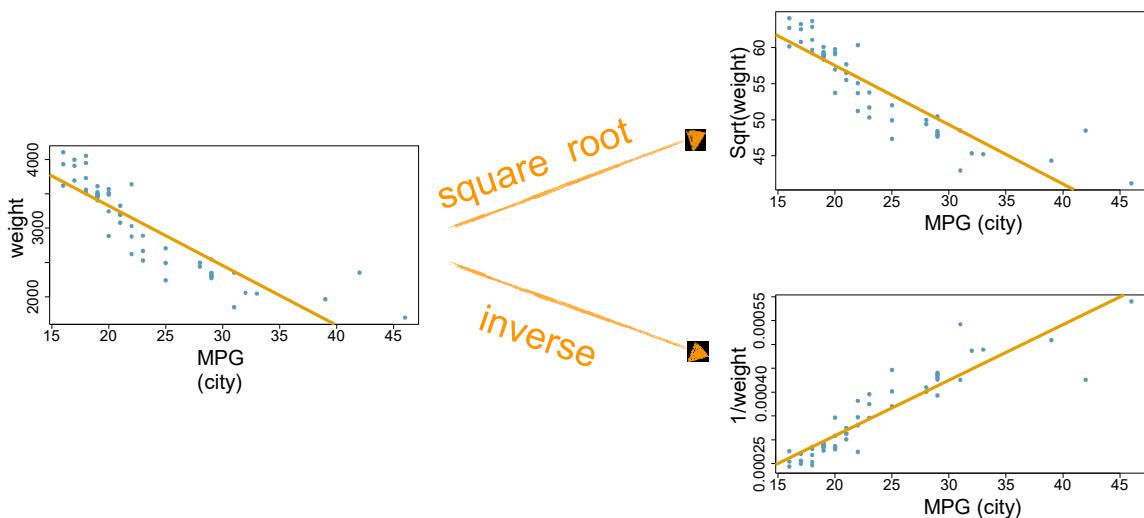
# Log Transformation

- To make the relationship between the variables more linear, and hence easier to model with simple methods



71

# Other Transformations



72

# Goals of Transformation

- To see the data structure differently.
- To reduce skew and assist in modeling.
- To straighten a nonlinear relationship in a scatterplot.
- To model the relationship with simpler methods.



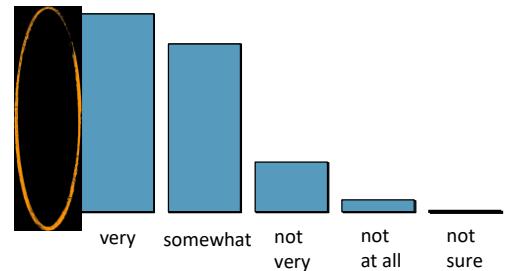
73



Describing Categorical  
Variables

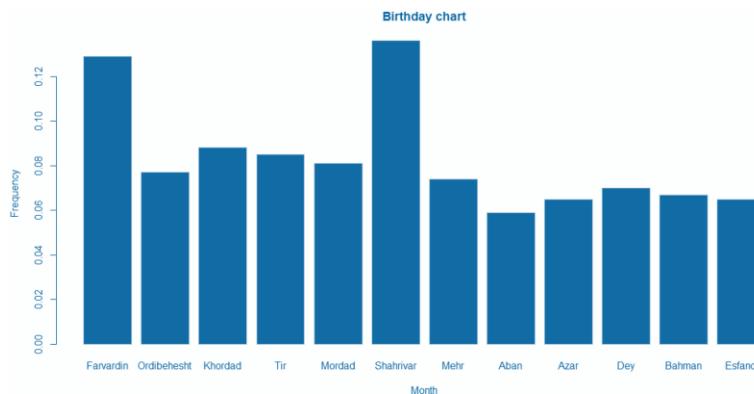
# Frequency Table & Bar Plot

| Difficulty saving money | Counts | Frequencies |
|-------------------------|--------|-------------|
| Very                    | 231    | 46%         |
| Somewhat                | 196    | 39%         |
| Not very                | 58     | 12%         |
| Not at all              | 14     | 3%          |
| Not sure                | 1      | ~0%         |
| Total                   | 500    | 100%        |



## Birthdays in Iran

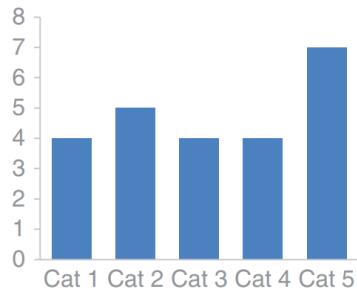
- Based on 1395 Census (A sample of 1,048,575 individuals)
  - Total number of valid data with Persian calendar: 1,000,222



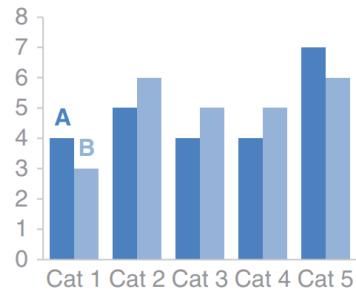
76

# Multiple Series in a Bar Chart

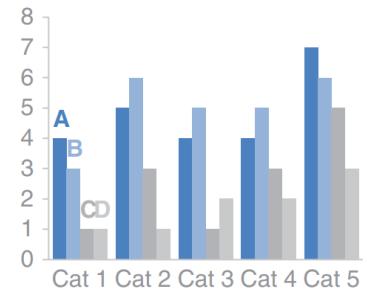
Single series



Two series

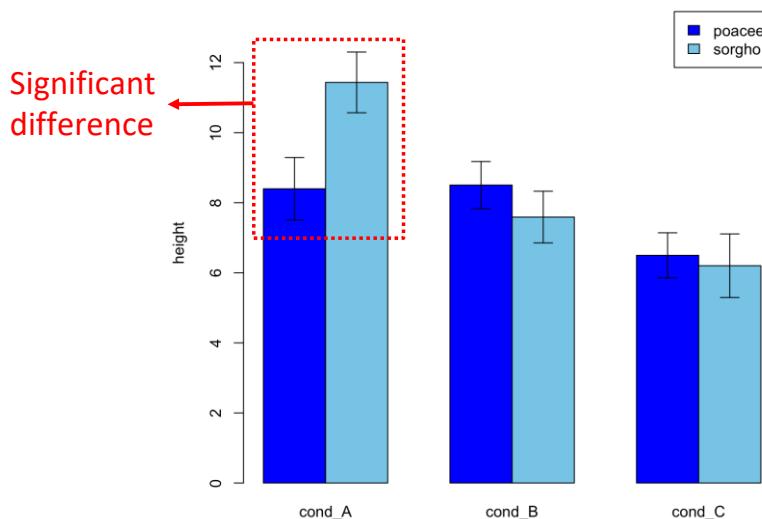


Multiple series



77

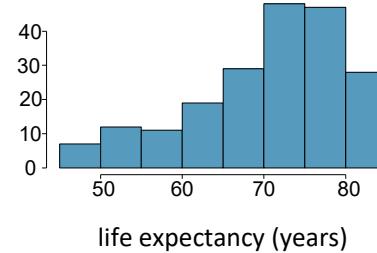
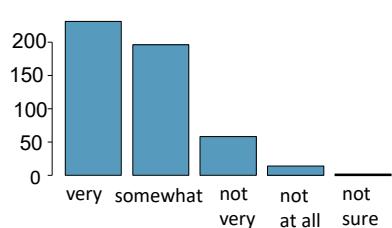
## Bar Plot + Error Bar



78

# Bar Plots vs. Histograms

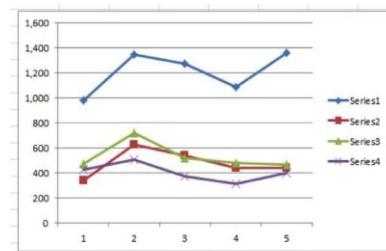
- Barplots for categorical variables, but histograms for numerical variables.
- x-axis on a histogram is a number line, and the ordering of the bars are not interchangeable.



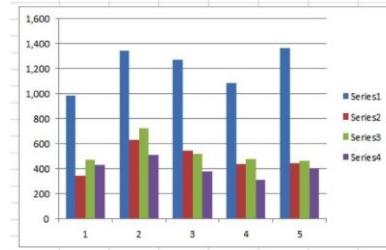
79

# Bar Plots vs. Line Charts

**Continuous values**  
e.g., time series

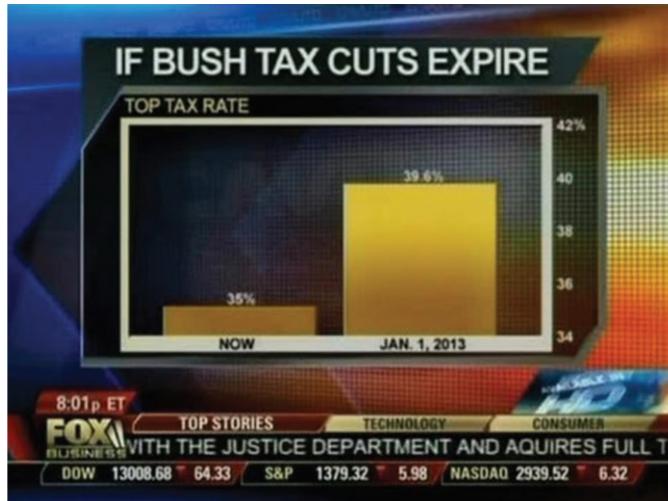


**Discrete values**  
e.g., countries



80

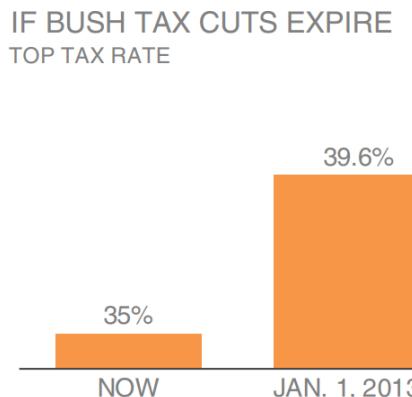
# Bar Plot Abuse



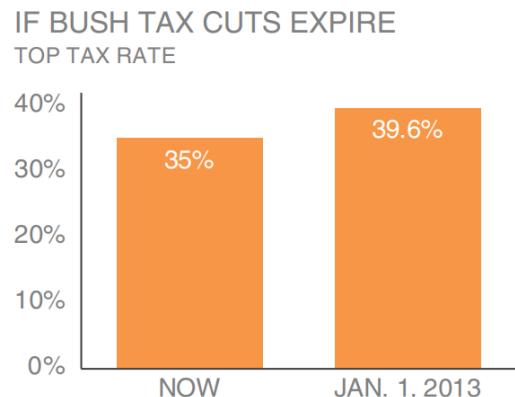
81

# Bar Plot Abuse

**Non-zero baseline:** as originally graphed



**Zero baseline:** as it should be graphed

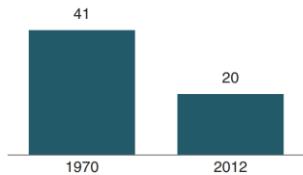


82

# Simple Text

## Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

# 20%

of children had a **traditional stay-at-home mom** in 2012, compared to 41% in 1970

83

# Contingency Table

|                   |            | Income  |          |         |         |       |
|-------------------|------------|---------|----------|---------|---------|-------|
|                   |            | < \$40K | \$40-80K | > \$80K | Refused | Total |
| Difficulty saving | Very       | 128     | 63       | 31      | 9       | 231   |
|                   | Somewhat   | 54      | 71       | 61      | 10      | 196   |
|                   | Not very   | 17      | 7        | 27      | 7       | 58    |
|                   | Not at all | 3       | 6        | 5       | 0       | 14    |
|                   | Not sure   | 0       | 1        | 0       | 0       | 1     |
|                   | Total      | 202     | 148      | 124     | 26      | 500   |

- A table that summarizes data for two categorical variables is called a **contingency table**.

84

# Relative Frequency

|                   |            | Income         |                     |               |              | Total |
|-------------------|------------|----------------|---------------------|---------------|--------------|-------|
| Difficulty saving | Very       | < \$40K<br>128 | \$40K - \$80K<br>63 | > \$80K<br>31 | Refused<br>9 | 231   |
|                   | Somewhat   | 54             | 71                  | 61            | 10           | 196   |
|                   | Not very   | 17             | 7                   | 27            | 7            | 58    |
|                   | Not at all | 3              | 6                   | 5             | 0            | 14    |
|                   | Not sure   | 0              | 1                   | 0             | 0            | 1     |
|                   | Total      | 202            | 148                 | 124           | 26           | 500   |

< \$40K:  $128/202 = 63\%$  find it very difficult to save

\$40K-\$80K:  $63/148 = 43\%$

\$80K:  $31/124 = 25\%$

Refused:  $9/26 = 35\%$

feelings about difficulty of saving money and income are associated (dependent)

85

## Borders in Tables

Heavy borders

| Group   | Metric A | Metric B | Metric C |
|---------|----------|----------|----------|
| Group 1 | \$X.X    | Y%       | Z,ZZZ    |
| Group 2 | \$X.X    | Y%       | Z,ZZZ    |
| Group 3 | \$X.X    | Y%       | Z,ZZZ    |
| Group 4 | \$X.X    | Y%       | Z,ZZZ    |
| Group 5 | \$X.X    | Y%       | Z,ZZZ    |

Light borders

| Group   | Metric A | Metric B | Metric C |
|---------|----------|----------|----------|
| Group 1 | \$X.X    | Y%       | Z,ZZZ    |
| Group 2 | \$X.X    | Y%       | Z,ZZZ    |
| Group 3 | \$X.X    | Y%       | Z,ZZZ    |
| Group 4 | \$X.X    | Y%       | Z,ZZZ    |
| Group 5 | \$X.X    | Y%       | Z,ZZZ    |

Minimal borders

| Group   | Metric A | Metric B | Metric C |
|---------|----------|----------|----------|
| Group 1 | \$X.X    | Y%       | Z,ZZZ    |
| Group 2 | \$X.X    | Y%       | Z,ZZZ    |
| Group 3 | \$X.X    | Y%       | Z,ZZZ    |
| Group 4 | \$X.X    | Y%       | Z,ZZZ    |
| Group 5 | \$X.X    | Y%       | Z,ZZZ    |

86

# Heatmap

- A heatmap is a way to visualize data in tabular format, where instead of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

Table

|            | A   | B   | C   |
|------------|-----|-----|-----|
| Category 1 | 15% | 22% | 42% |
| Category 2 | 40% | 36% | 20% |
| Category 3 | 35% | 17% | 34% |
| Category 4 | 30% | 29% | 26% |
| Category 5 | 55% | 30% | 58% |
| Category 6 | 11% | 25% | 49% |

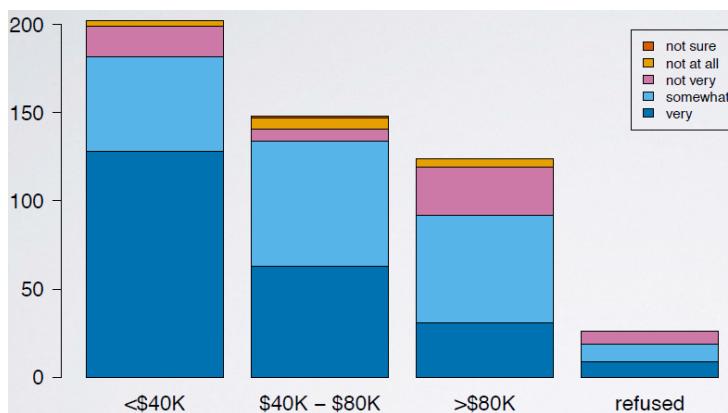
Heatmap

LOW-HIGH

|            | A   | B   | C   |
|------------|-----|-----|-----|
| Category 1 | 15% | 22% | 42% |
| Category 2 | 40% | 36% | 20% |
| Category 3 | 35% | 17% | 34% |
| Category 4 | 30% | 29% | 26% |
| Category 5 | 55% | 30% | 58% |
| Category 6 | 11% | 25% | 49% |

87

# Segmented (Stacked) Bar Plot

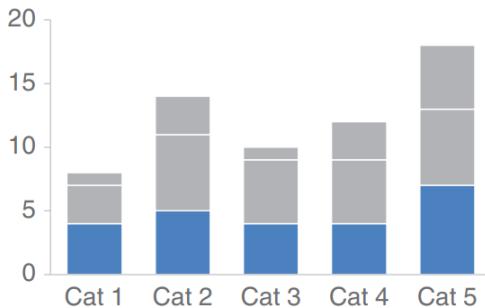


- Useful for visualizing conditional frequency distributions
- Compare relative frequencies to explore the relationship between the variables

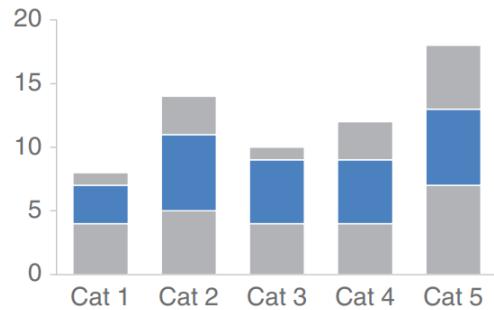
88

# Stacked Bar Plot

Comparing **these** is easy

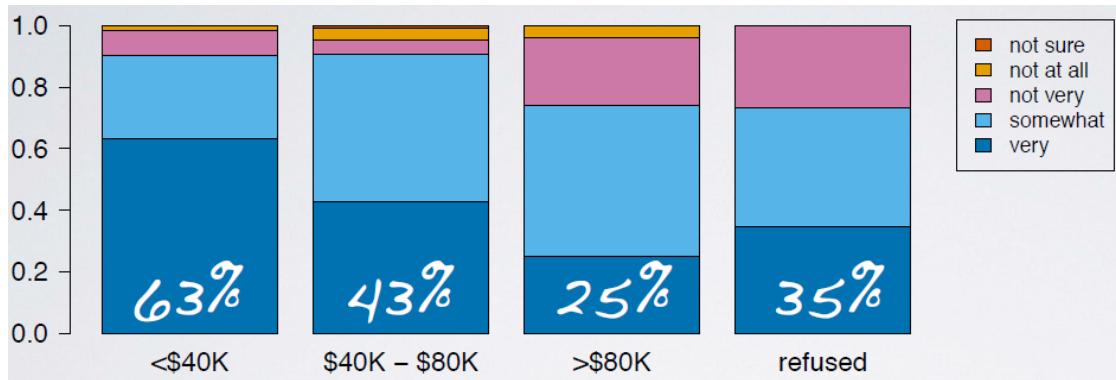


Comparing **these** is hard



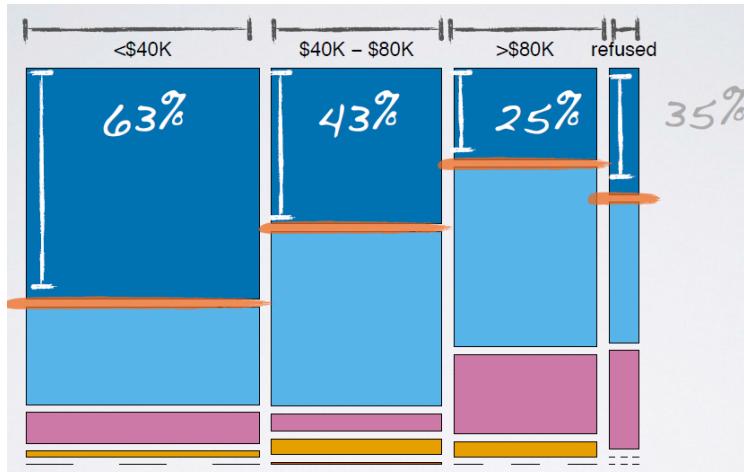
89

# Relative Frequency Segmented Bar Plot



90

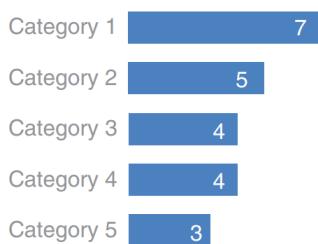
# Mosaic Plot



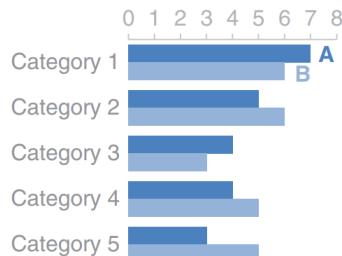
91

# Horizontal Bar Plot

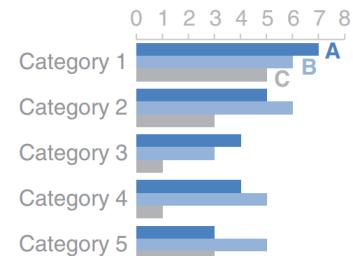
Single series



Two series



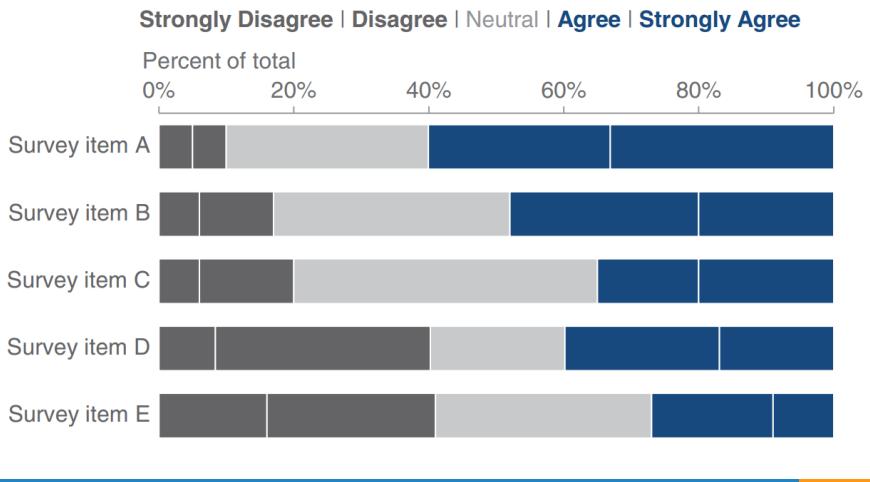
Multiple series



92

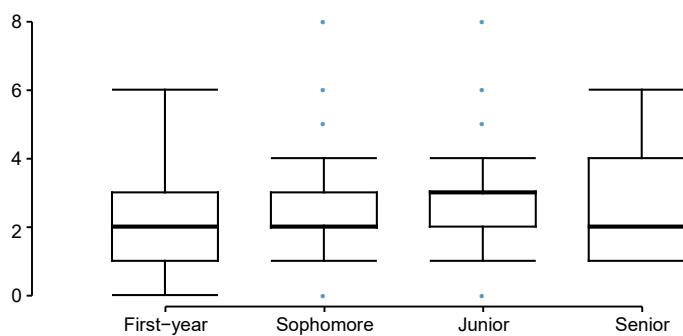
# Stacked Horizontal Bar Plot

Survey results



93

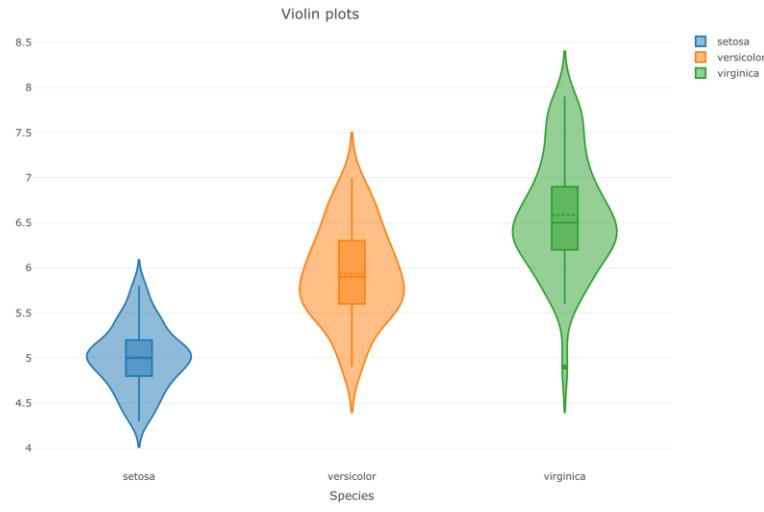
# Side-by-side box plots



- Does there appear to be a relationship between class year and number of societies students are in?

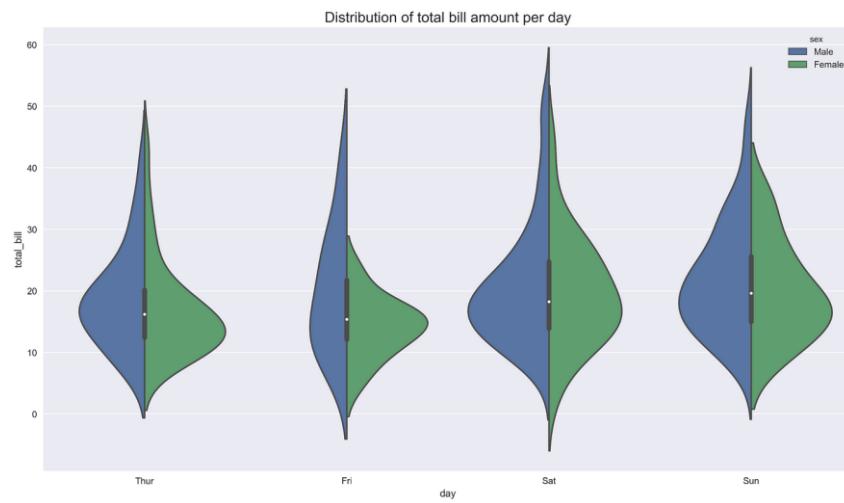
94

# Side-by-side violin plot



95

# Violin Plots for Comparison

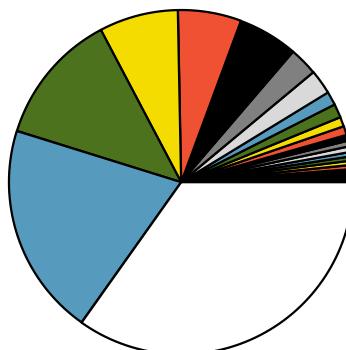
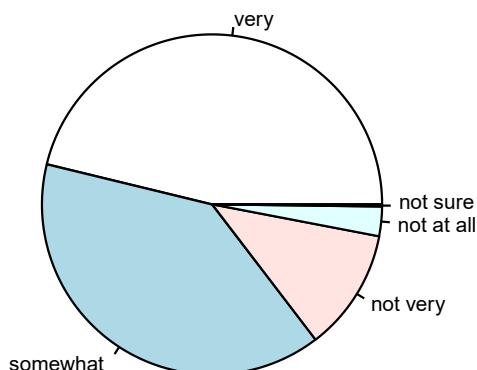


96

# To Be Avoided

97

~~Pie Chart?~~ NO!



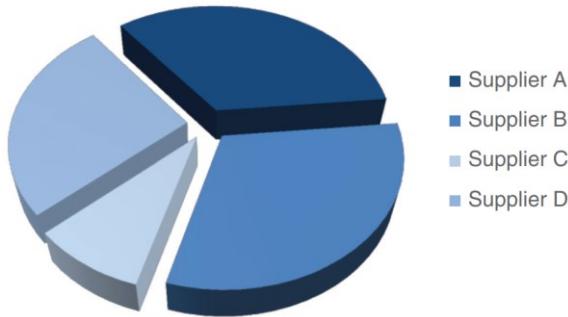
|                  |
|------------------|
| RODENTIA         |
| CHIROPTERA       |
| CARNIVORA        |
| ARTIODACTYLA     |
| PRIMATES         |
| SORICOMORPHA     |
| LAGOMORPHA       |
| DIPROTODONTIA    |
| DIDELPHIMORPHIA  |
| CETACEA          |
| DASYUROMORPHIA   |
| AFROSORICIDA     |
| ERINACEOMORPHYA  |
| SCANDENTIA       |
| PERISSODACTYLA   |
| HYRACOIDEA       |
| PERAMELEMORPHI   |
| A CINGULATA      |
| PILOSA           |
| MACROSCELIDEA    |
| TUBULIDENTATA    |
| PHOLIDOTA        |
| MONOTREMATA      |
| PAUCITUBERCULATA |
| SIRENIA          |
| PROBOSCIDEA      |
| DERMOPTERA       |
| NOTORYCTEMORPHI  |
| A MICROBIOTHERIA |

- A pie chart is actually much less informative than a bar plot.

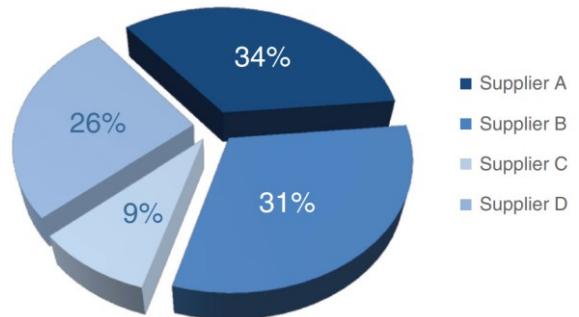
98

# 3D Pie Charts

Supplier Market Share



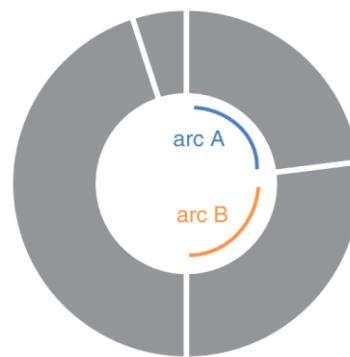
Supplier Market Share



99

# Donut Chart

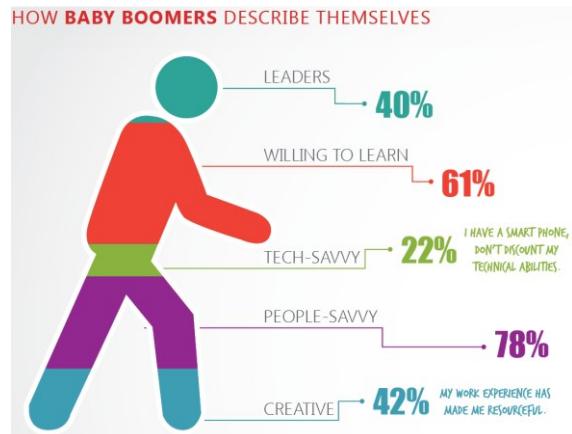
The donut chart



100

# Area Graphs

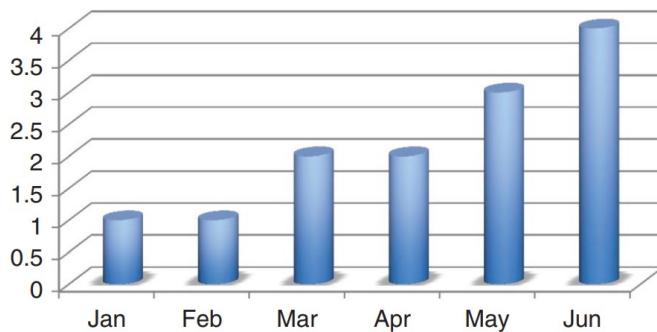
- Humans' eyes don't do a great job of attributing quantitative value to two-dimensional space.



101

## Never use 3D

Number of issues



102

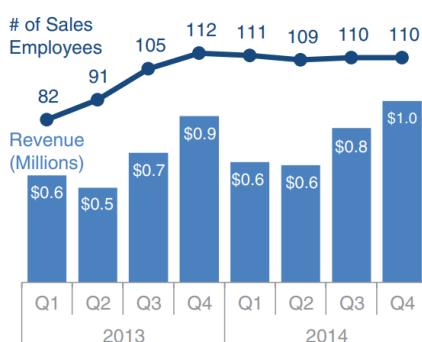
# Secondary y-axis



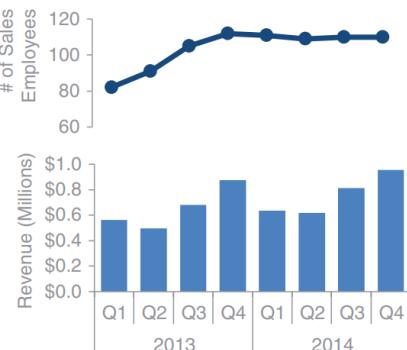
103

# Alternatives for Secondary y-axis

Alternative 1: label directly

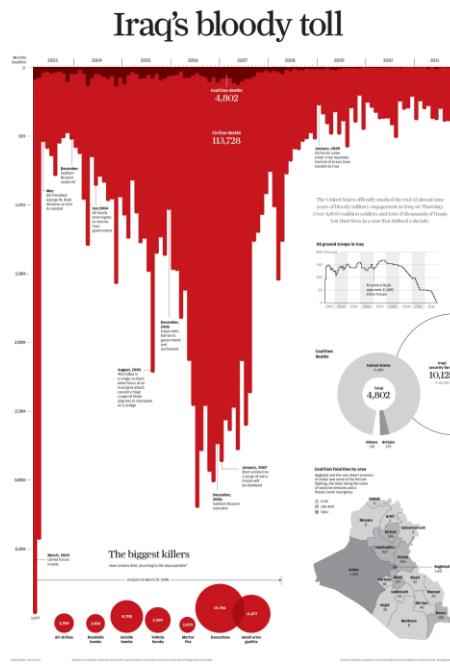
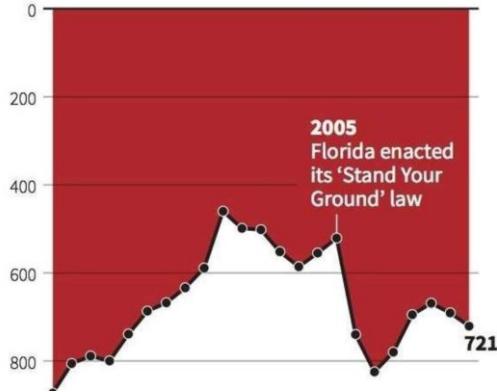


Alternative 2: pull apart vertically

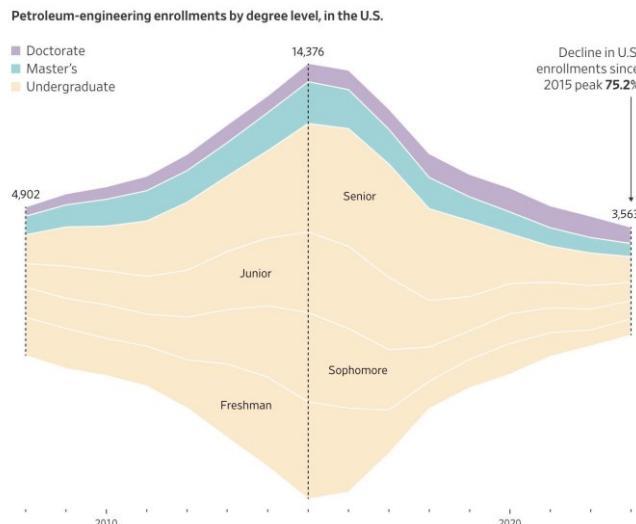


104

# Inverse Charts



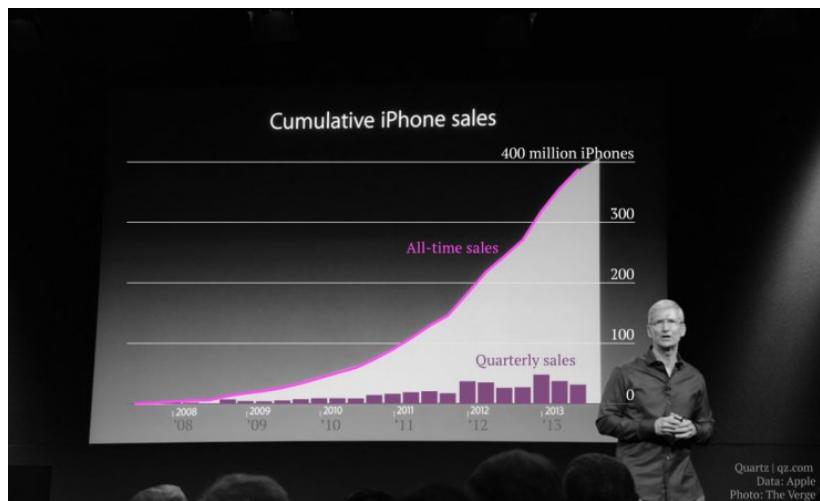
105



106



107



108

The image shows a segment from the CNBC program "Squawk on the Street". On the left, a man in a suit and glasses is speaking. To his right is a large graphic displaying financial information. At the top of the graphic, it says "GRayscale Bitcoin Trust (BTC) (GBTC)". Below that, the current price is listed as "19.88 -0.68 [-3.31%]" in red text. A line chart shows the price movement over the previous week, starting around \$17.50 on Thursday and rising to \$19.88 by Wednesday. The chart includes a blue horizontal line at the \$19.88 level. The days of the week are labeled as THU, FRI, MON, TUE, and WED along the x-axis. The y-axis has tick marks at 17, 18, 19, and 21. In the bottom left corner of the screen, there is a logo for "SQUAWK ON THE STREET". In the bottom right corner, the NBC/CNBC peacock logo is visible.

SQUAWK ON THE STREET

GRayscale Bitcoin Trust (BTC) (GBTC)

19.88 -0.68 [-3.31%]

1-WK [9.29%]

19.88

21  
20  
19  
18  
17

THU FRI MON TUE WED

GRAYSCALE'S BITCOIN BATTLE  
PAVING THE WAY FOR FIRST CRYPTO ETF?

CNBC