# AP Statistics

**Directions:**

1. This packet is designed to introduce you to one variable statistics and is to be handed in to your AP Statistics teacher on the first day of school.
2. Watch the video lessons and fill in the guided notes as you watch them. Links to the videos are provided at the top of each lesson (https://skewthescript.org/ap-stats-curriculum)
3. After completing the guided notes, complete the attached MCQs.
4. A test on the material included in this packet will be given during the first full week of the school year.

---------------------------------------------------------------------------------------------------------------------

**Name:** _____

# AP Statistics Handout: Lesson 1.1

*Topics:* quantitative and categorical data, misleading graphs

## Lesson 1.1 Guided Notes: Video Link - Misleading Data

*Quantitative vs. Categorical Data*

Quantitative data: Data that is _____ (think 'quantities'). The values have an inherent _____.

List several examples of quantitative data:

Categorical data: Data where values are categories or group labels, which often _____ _____.

List several examples of categorical data:

| Student | Height (in) | Dominant Hand | Final Exam Score | Home Zip Code |
|---------|-------------|---------------|------------------|---------------|
| Bill | 72 | Left | 77 | 68494 |
| Julius | 64 | Right | 83 | 68492 |
| Yesenia | 67 | Right | 91 | 68490 |

1. Label each of the variables (height, favorite color, final exam score, home zip code) as either quantitative or categorical. For each, explain your reasoning.

*Data Visualization & Misleading Graphs*

How to spot a misleading graphic:

1. It may not have axis labels or _____.
2. It may _____ the x or y axis, or start at a weird place.
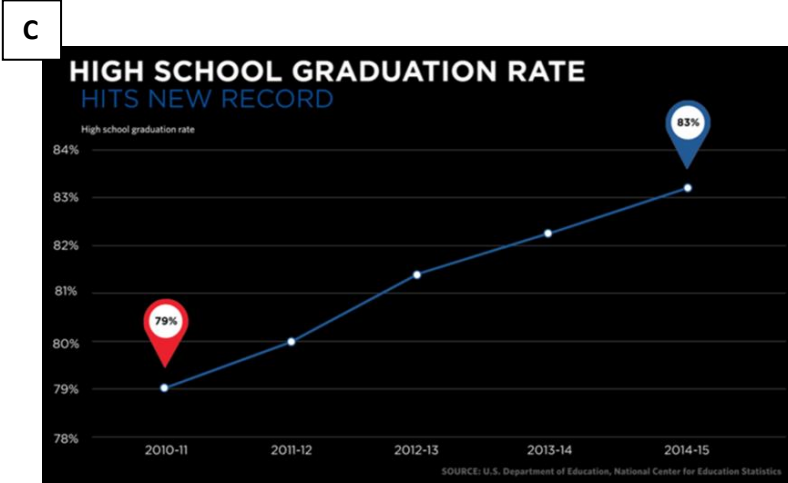3. It may use _____ for bar graphs (called a 'pictograph').

**Example 1:** "Kobe was a ball hog"

**Example 2:** "Chevy builds the most dependable trucks"
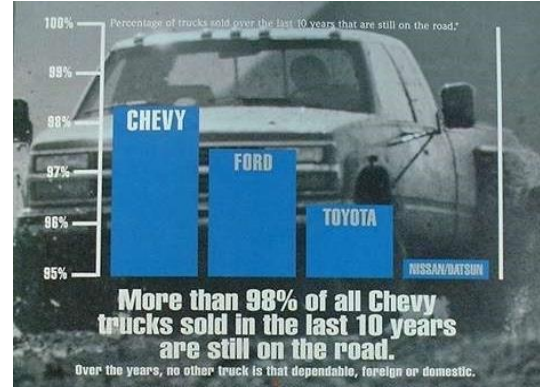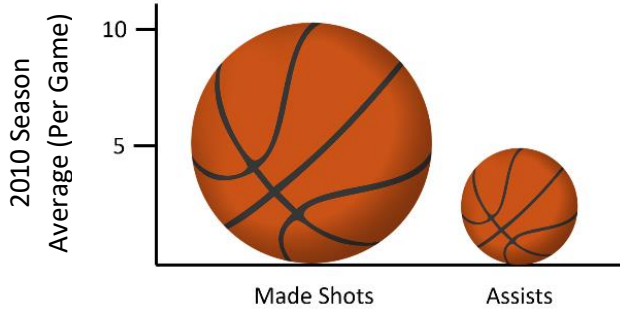




Example from TED-Ed: "How to spot a misleading graph"

For each example, describe why the data visual might be misleading. Then, sketch a more transparent graph of the same data.

Skew The
Script

# Lesson 1.1 Practice

*Please complete all exercises **before turning to the next page**!*

**C**



HIGH SCHOOL GRADUATION RATE
HITS NEW RECORD
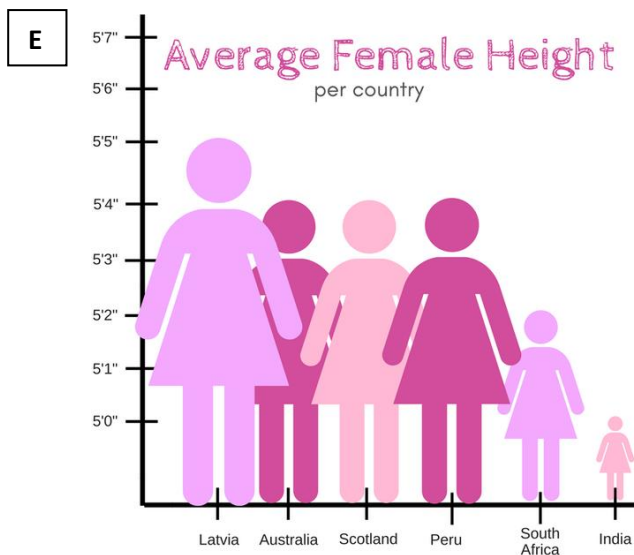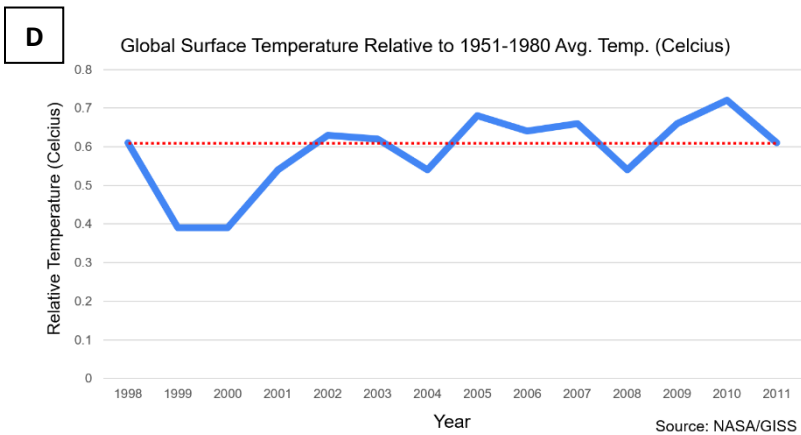
High school graduation rate

Source link: https://obamawhitehouse.archives.gov/blog/2016/10/17/graduation-rate-reaches-new-high-one-student-shares-his-story

**Graphic 'C'** was published on the White House's official blog during the Obama Administration. It uses national school data prepared by the Department of Education.

**Graphic 'D'** was presented at summit of climate change skeptics. It uses global land-ocean temperature data from NASA's Goddard Institute for Space Studies.

**Graphic 'E'** was shared on Twitter and has an unknown origin.

For each graph, answer the following questions: Is the visual misleading? Why or why not? If it is misleading, how would you change it?
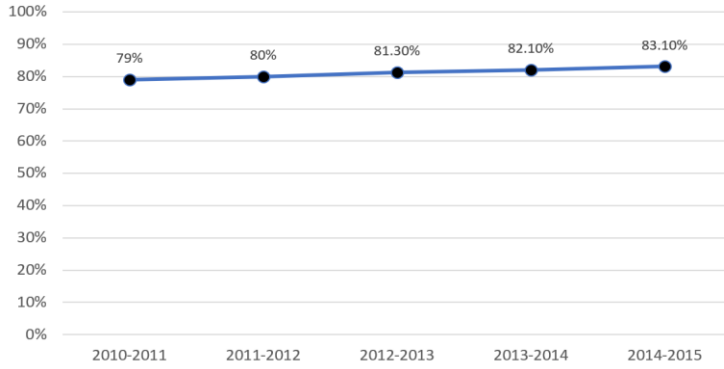
**D**



Global Surface Temperature Relative to 1951-1980 Avg. Temp. (Celcius)

Source: NASA/GISS

**E**



Average Female Height
per country

Latvia   Australia   Scotland   Peru   South Africa   India

Graphic found by Sabah Ibrahim on twitter
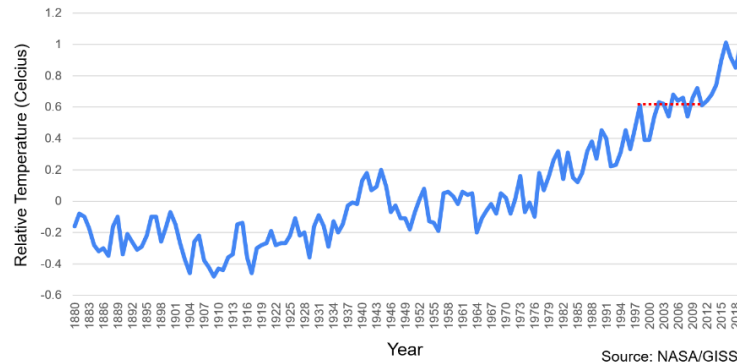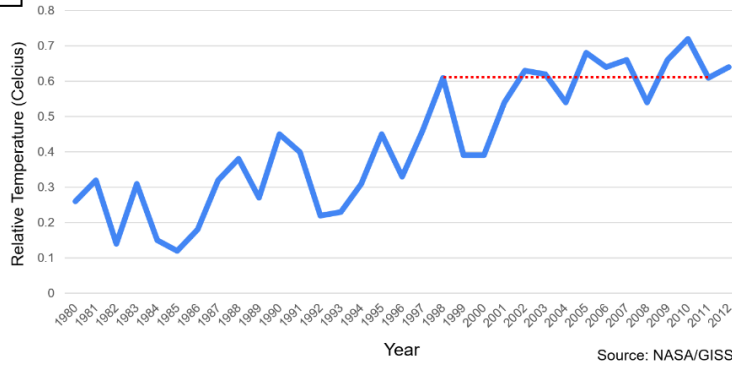
Skew The Script

**C**

### U.S. High School Graduation Rates



This page presents the **same data**, but with **adjusted graphs**. For each graph, answer the following question: What makes this version of the graph less misleading? Explain.
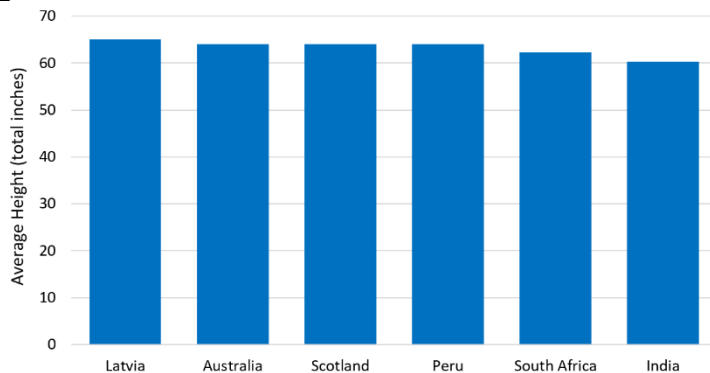
**D**

Global Surface Temperature Relative to 1951-1980 Avg. Temp. (Celcius)



Source: NASA/GISS



Source: NASA/GISS

**E**

### Average Height Among Women in Different Countries

Skew The Script

# AP Statistics Handout: Lesson 1.2
## [Guided Notes Video Link](#)
*Topics:* marginal & conditional distributions, bar graphs, association

**2021 - 2022**

**Dynamic Duo:** The Aces basketball team moved to Las Vegas in 2018, where they finished last place in the WNBA's western conference. By 2022, they won the league title, led by their two new stars: forward A'ja Wilson (6'4") and guard Kelsey Plum (5'8"). They had compatible skillsets: Wilson's height and skill allowed to her to score a lot from shorter distances (2-point shots), while Plum's range allowed her to score from longer distances (3-point shots). That season, Wilson and Plum became the first pair of teammates to each score 700 points in a single season.

**MVP & Shooting Accuracy:** In their championship season, both athletes were league MVP (most valuable player) contenders. One common criterion for evaluating an MVP candidate is their shooting percentage. Each player's 2021-2022 shooting percentages are given at left, where you'll notice a *paradoxical* result.

| Shooting Percentage | A'ja Wilson | Kelsey Plum |
|---|---|---|
| 2-Pointers | 48.3% | **48.8%** |
| 3-Pointers | 38.1% | **41.0%** |
| Overall | **47.4%** | 45.1% |

**Today's Key Analysis:** How can a player be better overall, but also worse in each category?

**Wilson**

| Shot Result | Type of Shot | | |
|---|---|---|---|
| | 2-Pointers | 3-Pointers | Total |
| Made | 435 | 32 | |
| Missed | 466 | 52 | |
| Total | | | |

**Plum**

| | Type of Shot | | |
|---|---|---|---|
| | 2-Pointers | 3-Pointers | Total |
| Made | 215 | 157 | 372 |
| Missed | 226 | 226 | 452 |
| Total | 441 | 383 | 824 |

**Data source:** Basketball Reference (*basketball-reference.com*). Data is shown from 2021 - 2022.
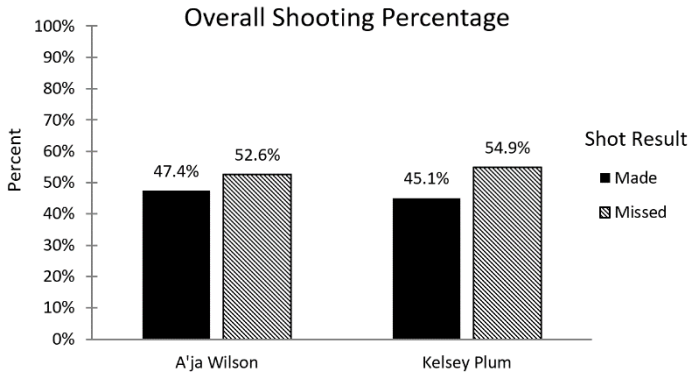
*Marginal distributions*

Two-Way Table: A table of counts describing two _____.

1) What are the two variables described in these tables?

2) Find the A'ja Wilson's **overall** shooting percentage (i.e. the marginal distribution of her shot results).

Skew The Script

## Side-by-side bar graphs

### Overall Shooting Percentage



3) What feature of this **side-by-side bar graph** is <u>not</u> needed to satisfy "title, tick, tick, label, label, label"?

4) Which player makes their shots more often? Which player misses their shots more often?

## Conditional distributions, segmented bar graphs, and associations

5) Find A'ja Wilson's shooting percentage for each shot type. Then, graph the distribution using a **segmented bar graph**. The distribution and graph for Kelsey Plum is given below.

**A'ja Wilson:** Shot accuracy by shot type (conditional distribution)

**Kelsey Plum:** Shot accuracy for each shot type (conditional distribution)

|          | 2-pointers | 3-pointers |
|----------|------------|------------|
| Made:    | 48.8%      | 41.0%      |
| Missed:  | 51.2%      | 59.0%      |

### Shooting Accuracy, Kelsey Plum
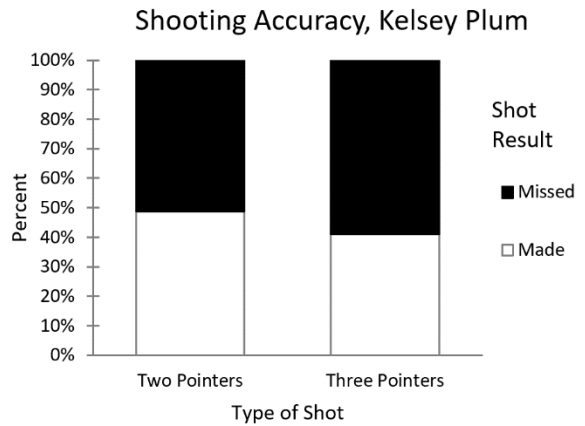


6) Do the data suggest an association between shooting percentage and shot type?

**Note:** What are the three key components you should mention when describing an association?

# Lesson 1.2 Discussion

**Discussion Question:** Why is Wilson more accurate overall, yet less accurate in every shot category? Explain.

_Hint:_ Below are two **mosaic plots** of the above data. Mosaic plots are the same thing as segmented bar graphs, but their bar widths are scaled by the sample size (number of shots) in each category.[1] Use these plots to inform your answer.

## A'ja Wilson

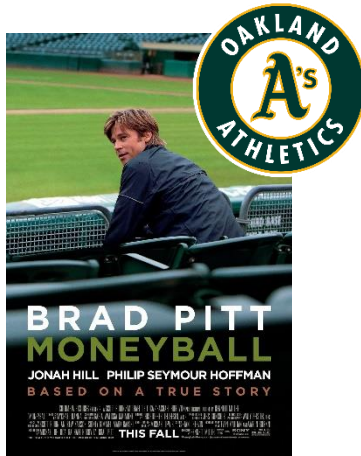## Kelsey Plum

**Shot Result**
- Miss
- Make
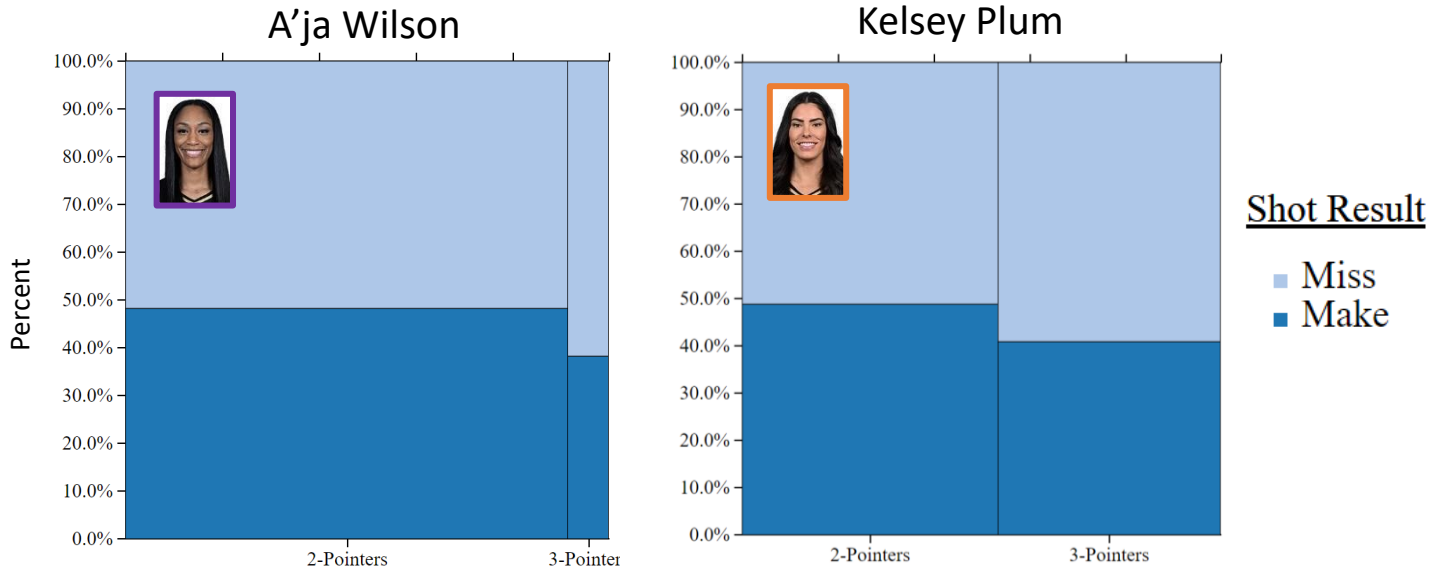
_Plots created using stapplet.com_

---

[1] On the AP Exam, you might be asked to interpret a mosaic plot, but you'll likely never have to draw one yourself.

Skew The Script

Skew The Script

# AP Statistics Handout: Lesson 1.3

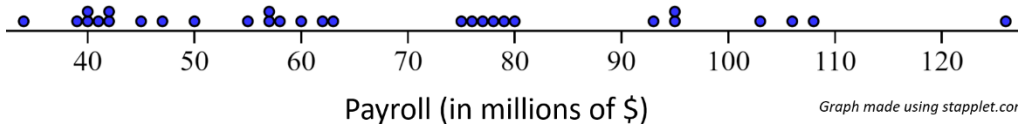*Topics:* dot plots, stemplots, histograms, CSOCS  **Guided Notes Video Link**

## Moneyball

"Moneyball" depicts the 2002 season of the Oakland Athletics (the A's) baseball team. When the movie was released, it was a hit – grossing $110 million. It was surprising that a move about the Oakland A's was successful. After all, in that 2002 season, the New York Yankees (not the A's) scored the most runs. The Atlanta Braves (not the A's) had the largest division lead. And the Anaheim Angels (not the A's) won the championship.

**Today's Key Analysis:** Why would anyone make (or watch) a movie about the Oakland A's?

*Dot Plots*

**"There are rich teams, and then there are poor teams, then there's 50 feet of crap, and then there's us."**

*– Brad Pitt, playing the GM of the Oakland A's in "Moneyball"*



Payroll (in millions of $)

*Graph made using stapplet.com*

| Team | Payroll (millions) |
|---|---|
| New York Yankees | $126 |
| Boston Red Sox | $108 |
| Texas Rangers | $106 |
| Arizona Diamondbacks | $103 |
| Los Angeles Dodgers | $95 |
| New York Mets | $95 |
| Atlanta Braves | $93 |
| Seattle Mariners | $80 |
| Cleveland Indians | $79 |
| San Francisco Giants | $78 |
| Toronto Blue Jays | $77 |
| Chicago Cubs | $76 |
| St. Louis Cardinals | $75 |
| Houston Astros | $63 |
| Anaheim Angels | $62 |
| Baltimore Orioles | $60 |
| Philadelphia Phillies | $58 |
| Chicago White Sox | $57 |
| Colorado Rockies | $57 |
| Detroit Tigers | $55 |
| Milwaukee Brewers | $50 |
| Kansas City Royals | $47 |
| Cincinnati Reds | $45 |
| Pittsburgh Pirates | $42 |
| Florida Marlins | $42 |
| San Diego Padres | $41 |
| Minnesota Twins | $40 |
| Oakland A's | $40 |
| Montreal Expos | $39 |
| Tampa Bay Devil Rays | $34 |

1. On the dotplot above, label the point the represents the New York Yankees. Then, label the Oakland A's. Based on their player payrolls, which team would you expect to win more games?

2. What are some advantages and disadvantages of using a dotplot to visualize data?

Skew The Script

## *Stemplots*

3. Below, create a stemplot of the payroll data. **Remember:** title, tick, tick, label, label, label.

5. What are some advantages and disadvantages of using a stemplot to visualize data?

## *Histograms*

6. Using the payroll data, complete the frequency table. Then, create a histogram of the data. **Remember:** title, tick, tick, label, label, label.

| Payroll (millions) | Frequency | Relative Frequency |
|---|---|---|
| 20 - 39 | | |
| 40 - 59 | | |
| 60 - 79 | | |
| 80 - 99 | | |
| 100 - 119 | | |
| 120 - 139 | | |
| Total | 30 | 100% |

7. What would change about your graph if you made a *relative frequency* histogram?

8. What are some advantages and disadvantages of using a histogram to visualize data?

Skew The Script

*CSOCS*

**Shape**

"Describe the distribution…"
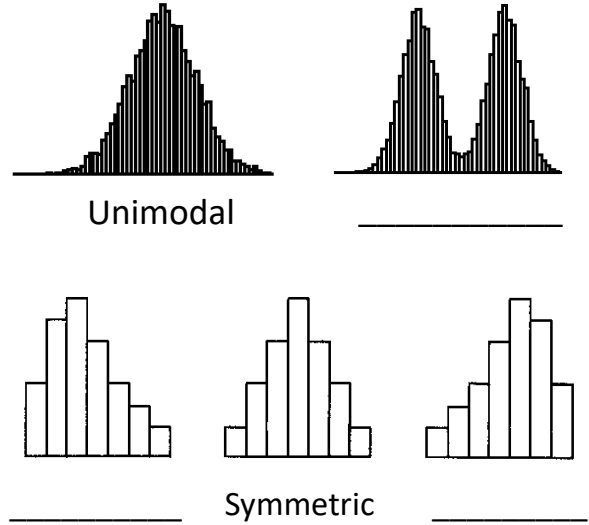
**C**ontext – What _____ is being measured?

**S**hape – _____ skew, symmetric, modes

**O**utliers – _____ points

**C**enter – Mean, median, _____

**S**pread – _____, IQR, standard deviation

Unimodal                     _____

_____        Symmetric        _____

9. Describe the distribution of payroll data, as depicted in your histogram.

**"Using stats the way we read them, we'll find value in players that nobody else can see…I believe that there is a championship team of 25 people that we can afford because everyone else in baseball undervalues them. "**
  – *Jonah Hill, playing the Oakland A's statistician in "Moneyball"*
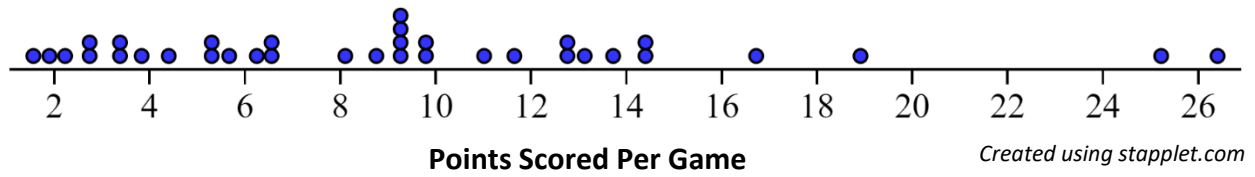
Number of Wins (per team)
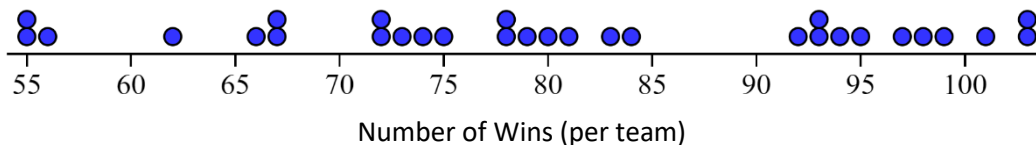
10. The dotplot above shows the number of wins each team had in the 2002 season. Below is a table with specific information for the Oakland A's and New York Yankees. Fill in the final column in the table, then label the teams on the dotplot above. Using the graph and the table, comment on why it makes sense to make a movie about the Oakland A's (even though they didn't win the championship).

| Team | Payroll | Wins | $ spent per win |
|------|---------|------|-----------------|
| NY Yankees | $125,928,583 | 103 | |
| Oakland A's | $40,004,167 | 103 | |

Skew The Script

# Lesson 1.3 Practice



**Points Scored Per Game**

*Created using stapplet.com*

1. The above dotplot displays the points scored per game by 35 NBA players during the 2017 basketball season.

    a) Describe the distribution of points scored per game.

    b) Among these players, there were two selected to be "All-Stars" (a recognition that they are among the best players in the league). These two players were LeBron James and Kyrie Irving. Circle the dots that likely represent these players. Then, comment on how their removal from the dataset would change your interpretation of the shape of the distribution.

Skew The Script

## AP Statistics Handout: Lesson 1.4
*Topics:* measures of center, measures of spread, using technology to find summary stats

## Guided Notes Video Link

*The table below shows the salaries of 12 employees at a company.*

### Measures of Center

| Salaries (thousands of $) |
|---|
| 39 |
| 34 |
| 34 |
| 35 |
| 34 |
| 32 |
| 43 |
| 34 |
| 185 |
| 35 |
| 29 |
| 67 |

1) Write the formula for **the mean** (in words and symbols). Then, find the mean salary (show your work).

2) Find the median salary (show your work). For convenience, the salaries (in thousands of $) are shown in order below:
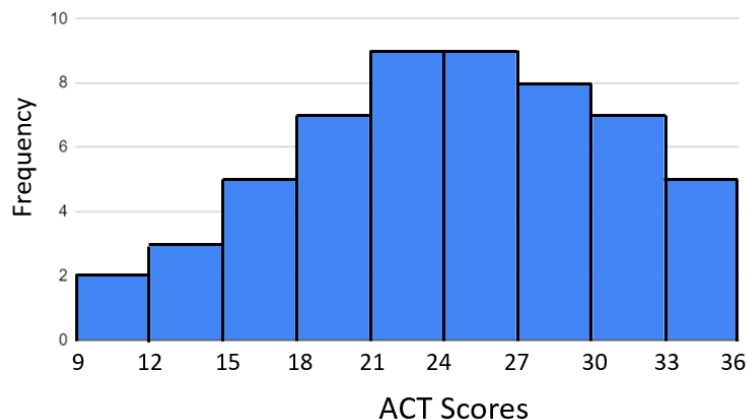
29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185

**Question to ponder (you'll answer it later in the lesson):** *The boss is trying to hire you to work at this company. She says, "Our typical salary is $50,100." Is this misleading? Why or why not?*

**Approximating a median in a histogram:** The graph below describes ACT scores among a group of students.

3) Approximate the median ACT score. Show your work.

*Measures of Spread*

4) Write the formula for the **range**. Then, find the range of the salaries (show your work).

5) The formula for the **standard deviation** ($s_x$) is displayed to the right:  $$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

Let's break down the formula piece-by-piece:

a) Why do you think we find the differences between each data point and the mean ($x_i - \bar{x}$)? Hint: the standard deviation is a measure of *spread.*

b) Why do you think we square the differences $(x_i - \bar{x})^2$ before we sum them?

c) We find the sum ($\sum$  ) and divide by n – 1. This is similar to what we did before for another statistic. Which statistic was that? Why do you think we are doing the (almost) same thing here?

d) At the end, we take the square root ($\sqrt{\phantom{xx}}$ ). Why do you think we do this?

6) Write down and interpret the standard deviation of the salaries.

7) Write down the formula for the **interquartile range (IQR)**. Then, find the IQR (show your work). For convenience, the salaries (in thousands of $) are shown in order below:

$$29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185$$

**Question to ponder (you'll answer it later in the lesson):** *Which measure of spread (range, standard dev., or IQR) best represents the "typical" distance between salaries? Why?*

Skew The Script

*Technology: Summary Statistics*
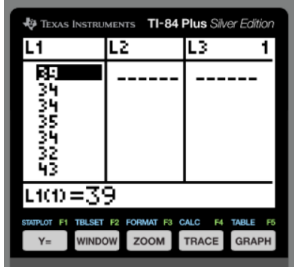


1. Put data into List 1
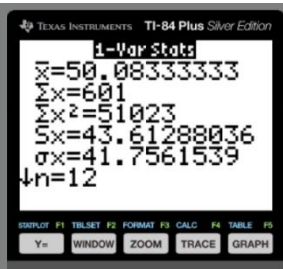(STAT → EDIT)



2. Find 1-Var Stats
(STAT → CALC → 2)



3. Select Data List
(→ Calculate)

4. Scroll through
the summary stats
(use Sx for stdev.)



## Lesson 1.4 Discussion

**1.** The boss is trying to hire you to work at this company. She says, "Our typical salary is $50,100." Is this misleading? Why or why not?

**Measures of Center**
Mean: 50.1 ($50,100)
Median: 34.5 ($34,500)

**2.** Which measure of spread (range, standard dev., or IQR) best represents the "typical" distance between salaries? Why?

**Measures of Spread**
Range: $156,000
$S_x$: $43,610
IQR: $7,000

| Salaries (thousands of $) |
|---|
| 39 |
| 34 |
| 34 |
| 35 |
| 34 |
| 32 |
| 43 |
| 34 |
| 185 |
| 35 |
| 29 |
| 67 |

**Post-Discussion Notes: "Resistance is futile"**

The median is _____ (not seriously affected by) skew and outliers. The mean _____ to skew and outliers. The mean follows skew/outliers.

The interquartile range (IQR) is resistant to skew and outliers. The range and _____ are not resistant to skew and outliers.

Why are the median and IQR resistant to outliers? *Let's explore with the salary data:*

29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185

**For the mean:** The outlier salary – $185,000 – drags up the mean because its high value is given _____ _____ in the calculation
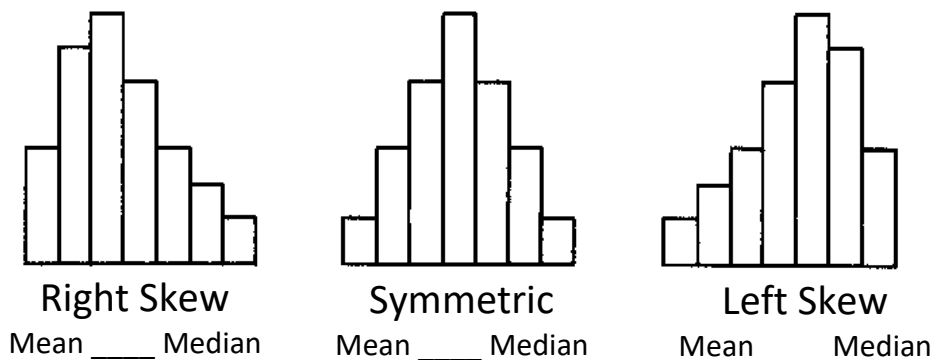
**For the median:** The _____ matters more than the _____. Because $185,000 is the highest data point, it's crossed off right away. The outlier is _____ given a large weight in the calculation.

**For the IQR:** Like the median, the position matters more than the value for the IQR.

**Applet:** Use the (very cool) simulation linked here to explore these properties of the mean and median.
     *Link:* http://digitalfirst.bfwpub.com/stats_applet/stats_applet_6_meanmed.html
     *Source: Digital First* project from Bedford, Freeman, & Worth publishers

Right Skew
Mean ____ Median

Symmetric
Mean ____ Median

Left Skew
Mean ____ Median

Skew The
Script

Name: _____

# AP Statistics Handout: Lesson 1.5 Guided Notes [Video Link](#)

*Topics:* five-number summary, determining outliers, boxplots, comparing distributions

**2017:** President Trump signs into law the largest tax reform in decades. Here are two different takes on the law:

"By eliminating tax breaks and loopholes, we will ensure that the benefits are focused on the middle class, the working men and women, not the highest income earners."

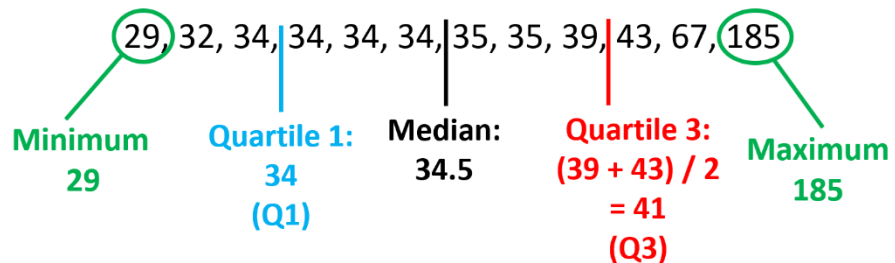– Donald Trump, 2017, during a speech in Indiana

"You remember just a few years ago when Trump and my Republican colleagues voted for almost $2 trillion in tax breaks for the wealthiest people in this country and the largest corporations."

– Bernie Sanders, 2021, during an interview on CNN

## Today's Key Analysis: Did Trump's tax cut help the "average" American?

### Five-Number Summary

**Salaries at a company (in thousands of $)**

29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185

**Minimum** 29

**Quartile 1:** 34 (Q1)

**Median:** 34.5

**Quartile 3:** (39 + 43) / 2 = 41 (Q3)

**Maximum** 185

### Determining Outliers

Outliers are _____ high or low data values.

*Which of the following salaries are outliers? Show your work.*

29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185
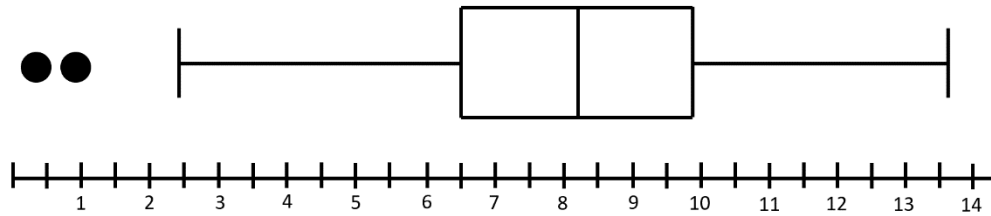
**Formulas for outlier boundaries:**

Upper Limit: Q3 + 1.5 x IQR
Lower Limit: Q1 − 1.5 x IQR
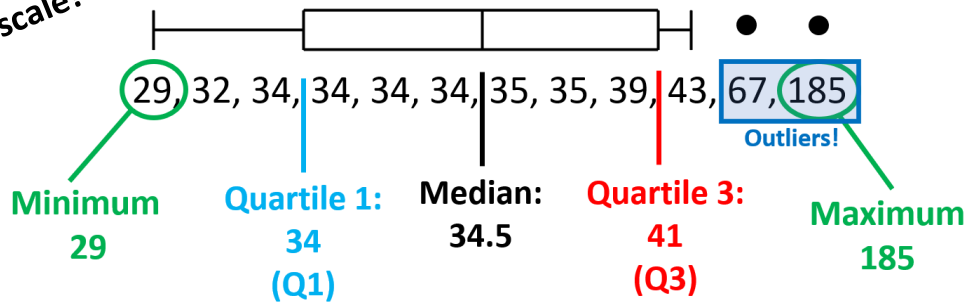
*Boxplots*

Label the components of the boxplot (Q1, Median, Q3, Min, Max, Outliers):



Label the type of skew for each of these boxplots:





Not to scale!



29, 32, 34, 34, 34, 34, 35, 35, 39, 43, 67, 185

Outliers!

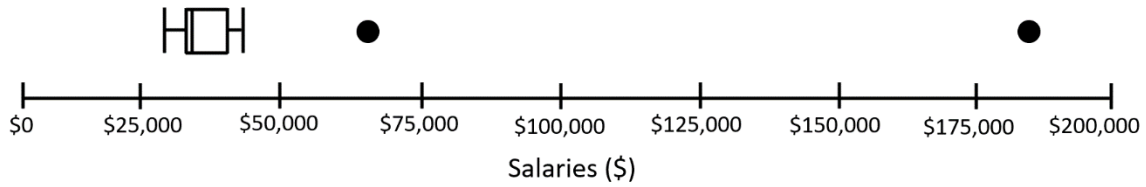| Minimum | Quartile 1: | Median: | Quartile 3: | Maximum |
|---|---|---|---|---|
| 29 | 34 | 34.5 | 41 | 185 |
| | (Q1) | | (Q3) | |
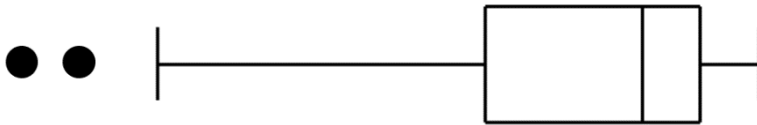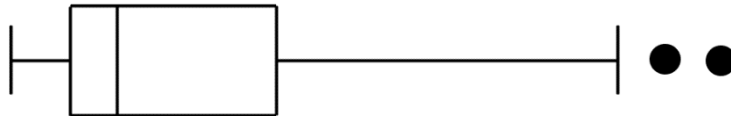
a) What percent of the data is below Q1?          b) What percent of the data is below Q3?

c) What percent of the data is above Q1?          d) What percent of the data is above the median?

e) What percent of the data is within the IQR?

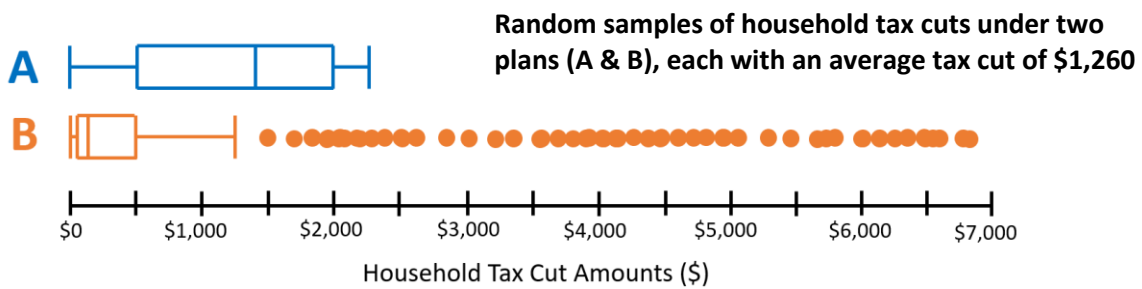Skew The Script

**Drawn to scale:**

# Salaries at a Company



Salaries ($)

**Describe the distribution:**

**C**ontext: Subject of data
**S**hape: Skew (_____)
**O**utlier: _____
**C**enter: _____
**S**pread: _____

## Comparing Distributions



**Random samples of household tax cuts under two plans (A & B), each with an average tax cut of $1,260**

Household Tax Cut Amounts ($)

**When asked to compare distributions…**
1. Use _____
2. Use _____ language for each feature
   - "less than", "greater than", "similar"

**Applies to dotplots, stemplots, histograms, and boxplots!**

*a) Compare the distributions:*

b) Which tax plan (A or B) shows tax cuts that benefit a broad swath of the American population? Which show tax cuts that mostly benefits a small minority of households? How can you tell?

c) These distributions (A & B) look very different. How is it possible that they have the same mean ($1,260)? Explain.

*Skew The Script*

# Lesson 1.5 Discussion

1. This table shows the actual tax cuts that Americans of different income levels experienced due to the 2017 tax law. Given the data presented here, explain why the following statements are misleading:

a) "Only the wealthy benefitted from the 2017 Tax Reform. The middle class **didn't benefit at all**."

| Income Group | Mean Tax Cut | Take-Home Income Growth Due to Tax Cut |
|---|---|---|
| Lowest 20% | $40 | 0.3% |
| Middle 20% | $780 | 1.4% |
| Top 20% | $5,790 | 2.2% |
| Top 1% | $32,650 | 2.2% |
| Top 0.1% | $89,060 | 1.3% |

b) "The **typical** American household paid **$1,260** less in taxes because of the 2017 Tax Reform." **Note:** $1,260 was the mean tax cut.
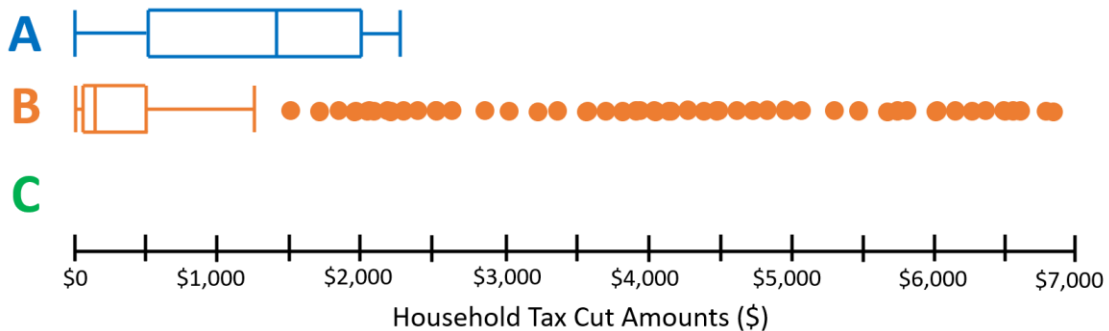


2. Above, sketch a boxplot (**C**) that could represent the actual distribution of tax cuts (based on the table above). Compare the medians and maximums of A, B, & C. In your view, is **C** a "fair" tax cut? Explain.

# Unit 1 - One Variable Data

**1.** A market researcher asked a group of men and women to choose their favorite color design from a sample of advertisements. The results are shown in the following table.

|  | Red with Black | Green with Black | Yellow with Black | Red with Blue | Green with Blue | Yellow with Blue |
|---|---|---|---|---|---|---|
| **Men** | 21 | 15 | 8 | 12 | 35 | 9 |
| **Women** | 15 | 3 | 11 | 31 | 24 | 16 |

Which of the following statements is <u>not</u> supported by the table?

(A) More men than women chose the color design red with black.

(B) More women than men chose the color design yellow with black.

(C) For men, the number who chose a design with black was greater than the number who chose a design with blue.

(D) The color design chosen by the most people was green with blue.

(E) The total number of men surveyed by the market researcher was equal to the total number of women surveyed by the market researcher.

**2.** The distribution of the number of transactions per day at a certain automated teller machine $(\text{ATM})$ is approximately normal with a mean of 80 transactions and a standard deviation of 10 transactions. Which of the following represents the parameters of the distribution?

(A) $\bar{x} = 80; \quad s = 10$

(B) $\bar{x} = 80; \quad s^2 = 10$

(C) $\bar{x} = 80; \quad \sigma = 10$

(D) $\mu = 80; \quad \sigma = 10$

(E) $\mu = 80; \quad s = 10$

**3.** At a small coffee shop, the distribution of the number of seconds it takes for a cashier to process an order is approximately normal with mean 276 seconds and standard deviation 38 seconds. Which of the following is closest to the proportion of orders that are processed in less than 240 seconds?

(A) 0.17

(B) 0.25

(C) 0.36

(D) 0.83

(E) 0.95

## Unit 1 - One Variable Data

**4.** In a certain school district, students from grade 6 through grade 12 can participate in a school-sponsored community service activity. The following bar chart shows the relative frequencies of students from each grade who participate in the community service activity.
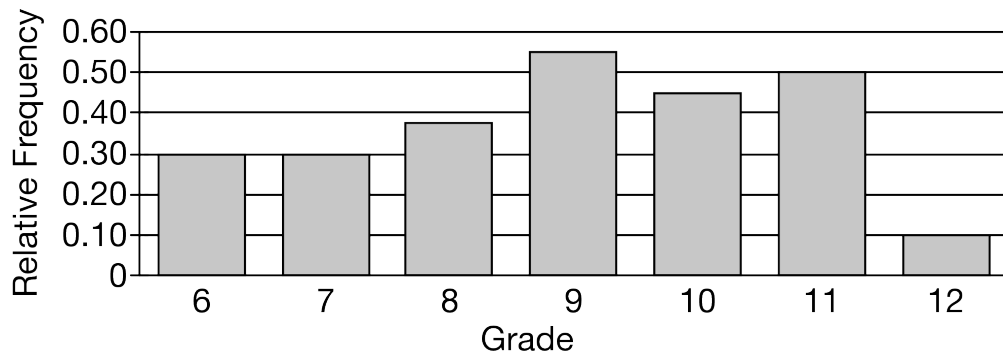


Which of the following statements is supported by the bar chart?

(A) The greatest number of participating students was in grade 9.

(B) The number of participating students in grade 6 was equal to the number of participating students in grade 7.

(C) The relative frequency of all participating students in grades 6 and 7 combined was 0.60.

(D) Grade 12 had the least relative frequency of participating students.

(E) Grade 11 had the greatest relative frequency of participating students.

**5.** Which of the following describes a continuous variable?

(A) The number of items sold at a craft booth for one day

(B) The number of apps downloaded from a website one day

(C) The diameters of the tree trunks at an evergreen farm

(D) The number of baskets made by a basketball player

(E) The shoe sizes of all shoes on sale at a department store

**6.** Data will be collected on the following variables. Which variable can be considered discrete?

(A) The height of a person

(B) The weight of a person

(C) The length of a person's arm span

(D) The time it takes for a person to solve a puzzle

(E) The number of books a person finished reading last month

## Unit 1 - One Variable Data

**7.** One statistic calculated for pitchers in baseball is called the earned run average, or $\text{ERA}$. The following boxplots summarize the $\text{ERA}$ for pitchers in two leagues, A and B.
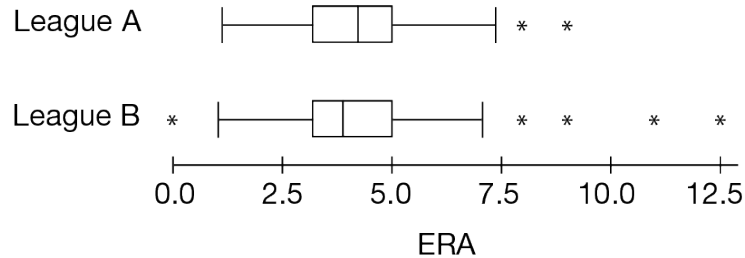


Based on the boxplots, which of the following statistics is the same for both leagues?

(A) The range

(B) The interquartile range

(C) The median

(D) The minimum

(E) The maximum

**8.** Roger claims that the two statistics most likely to change greatly when an outlier is added to a small data set are the mean and the median. Is Roger's claim correct?

(A) Yes, both the mean and median are likely to change greatly.

(B) No, only the mean is likely to change greatly.

(C) No, only the median is likely to change greatly.

(D) No, neither the mean nor the median are likely to change greatly.

(E) There is not enough information to determine if the mean or the median is likely to change greatly.

**9.** A child psychologist asked 100 five year olds and 50 ten year olds to name their favorite color. Their results are shown in the following table.

| | Red | Orange | Yellow | Green | Blue | Purple |
|---|---|---|---|---|---|---|
| **Five year olds** | 21 | 15 | 8 | 10 | 35 | 11 |
| **Ten year olds** | 7 | 2 | 6 | 15 | 12 | 8 |

Which of the following statements is supported by the table?

(A) The percentage of five year olds who selected red or blue as their favorite color is greater than the percentage of ten year olds who selected red or blue as their favorite color.

(B) The percentage of five year olds who selected yellow as their favorite color is greater than the percentage of ten year olds who selected yellow as their favorite color.

(C) The percentage of children who selected red, yellow, or blue as their favorite color was equal for both ages.

(D) Less than half of the five year olds selected red, yellow, or blue as their favorite color.

(E) Less than half of the ten year olds selected red, yellow, or blue as their favorite color.
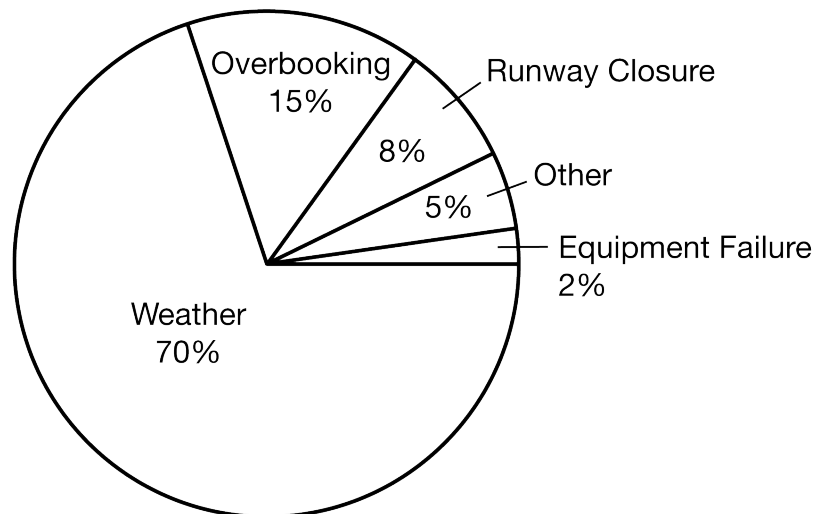
## Unit 1 - One Variable Data

10. The following frequency table shows the responses from a group of college students who were asked to choose their favorite flavor of ice cream.

| Flavor | Frequency |
|--------|-----------|
| Vanilla | 100 |
| Chocolate | 65 |
| Strawberry | 35 |
| Mint chip | 45 |
| Coffee | 30 |
| Butter pecan | 25 |

Which of the following statements is <u>not</u> supported by the table?

(A) The number of student responses is 300.

(B) One-third of the students chose vanilla.

(C) One-third of the students chose chocolate or strawberry.

(D) One-fourth of the students chose mint chip or coffee.

(E) One-half of the students chose vanilla or chocolate.

11. The following pie chart summarizes the results of a survey given to airlines about the primary reason for flight delays.



Which of the following statements is supported by the pie chart?

## Unit 1 - One Variable Data

    (A)   The reason given most frequently was runway closure.

    (B)   More delays were caused by weather than by all other reasons combined.

    (C)   More delays were caused by runway closure than were caused by overbooking.

    (D)   Overbooking and runway closure accounted for greater than one-fourth of the reasons given for flight delays.

    (E)   The combined percentage for other and runway closure was equal to the percentage for overbooking.

12. Which of the following statements is true about a distribution that appears to have a gap when displayed as a histogram?

    (A)   The distribution must have an outlier.

    (B)   The distribution has a region between two data values where no data were observed.

    (C)   The distribution is approximately normal.

    (D)   The distribution cannot be symmetric.

    (E)   The distribution must be bimodal.

13. The following boxplot shows the typical gas mileage, in miles per gallon, for 20 different car models.



Typical Gas Mileage
(Miles Per Gallon)
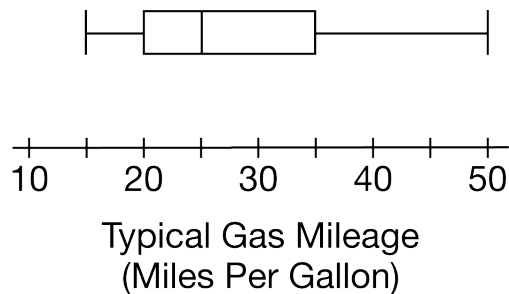
Based on the boxplot, the top 25 percent of the cars have a typical gas mileage of at least how many miles per gallon?

    (A)   15

    (B)   20

    (C)   25

    (D)   35

    (E)   50

14. A golfer recorded the following scores for each of four rounds of golf: 86, 81, 87, 82. The mean of the scores is 84. What is the sum of the squared deviations of the scores from the mean?

    (A)  $\sum(x - \bar{x}) = (86 - 84) + (81 - 84) + (87 - 84) + (82 - 84)$

    (B)  $\sum|x - \bar{x}| = |86 - 84| + |81 - 84| + |87 - 84| + |82 - 84|$

    (C)  $2\sum|x - \bar{x}| = 2[|86 - 84| + |81 - 84| + |87 - 84| + |82 - 84|]$

    (D)  $\sum(x - \bar{x})^2 = (86 - 84)^2 + (81 - 84)^2 + (87 - 84)^2 + (82 - 84)^2$

    (E)  $[\sum|x - \bar{x}|]^2 = [|86 - 84| + |81 - 84| + |87 - 84| + |82 - 84|]^2$
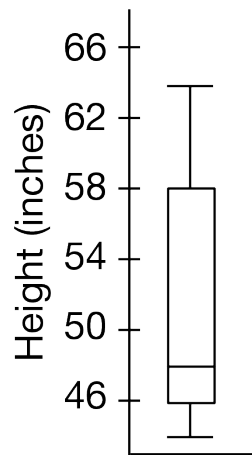
## Unit 1 - One Variable Data

**15.** One way to measure the duration of subterranean disturbances such as earthquakes and mining is to calculate the root-mean-square time. The following histograms summarize the distributions of the root-mean-square times for two sources of disturbances.
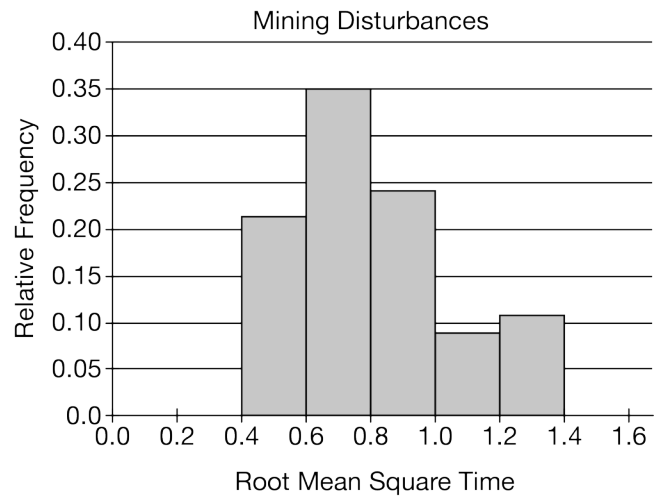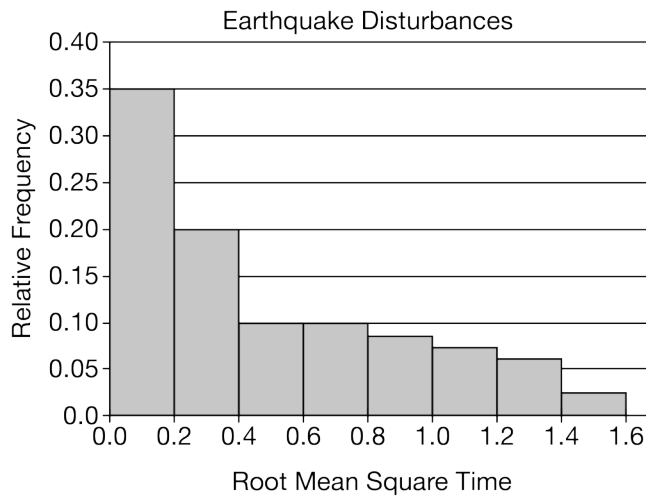


Based on the histograms, which of the following correctly compares the two distributions?

  (A)   The median of the earthquake disturbances is equal to the median of the mining disturbances.

  (B)   The median of the earthquake disturbances is less than the median of the mining disturbances.

  (C)   The range of the earthquake disturbances is equal to the range of the mining disturbances.

  (D)   The range of the earthquake disturbances is less than the range of the mining disturbances.

  (E)   The mode of the earthquake disturbances is equal to the mode of the mining disturbances.

**16.** An amusement park attraction has a sign that indicates that a person must be at least 48 inches tall to ride the attraction. The following boxplot shows the heights of a sample of people who entered the amusement park on one day.
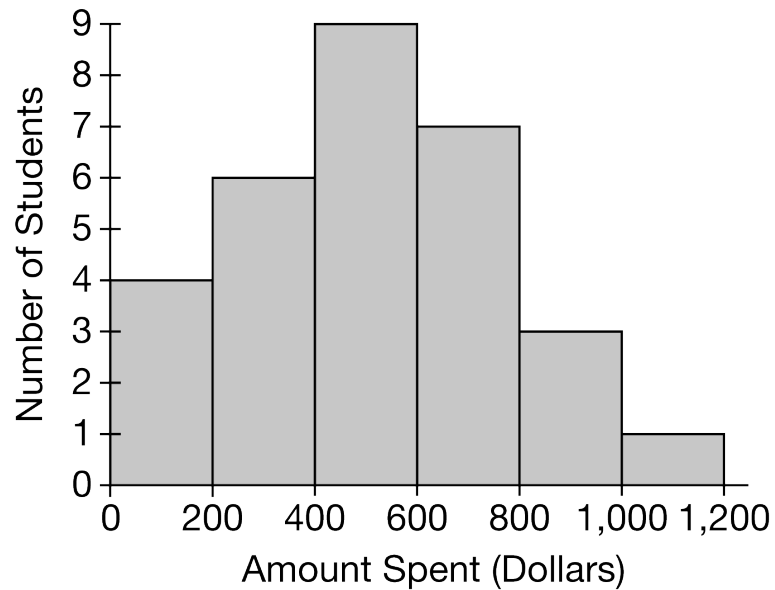


Based on the boxplot, approximately what percent of the people who entered the amusement park met the height requirement for the attraction?

## Unit 1 - One Variable Data

   (A)   25%

   (B)   48%

   (C)   50%

   (D)   75%

   (E)   100%

**17.** The following histogram summarizes the amount spent on plane tickets to travel home, in dollars, for a group of 30 college students.



If the interval size is decreased from $200 to $100, which of the following must remain the same on the new histogram?

   (A)   The heights of the bars

   (B)   The widths of the bars

   (C)   The number of bars

   (D)   The sum of the frequencies

   (E)   The shape of the distribution

# Unit 1 - One Variable Data

**18.** The following histogram shows the ages, in years, of the people who attended a documentary at a movie theater.
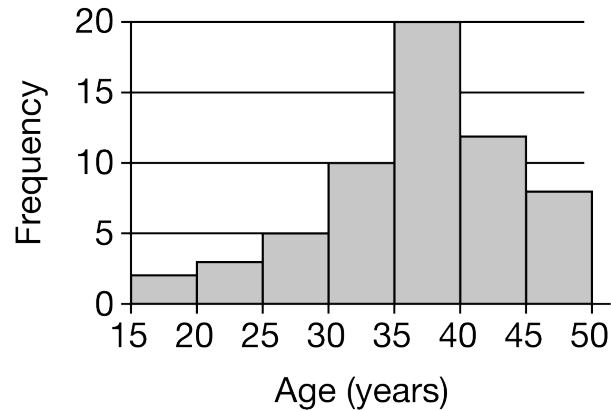


Based on the histogram, which of the following statements best describes the relationship between the mean and the median of the distribution of ages?

(A) The mean and the median are equal in value because the distribution is symmetric.

(B) The mean is most likely less than the median because the distribution is skewed to the right.

(C) The mean is most likely less than the median because the distribution is skewed to the left.

(D) The mean is most likely greater than the median because the distribution is skewed to the right.

(E) The mean is most likely greater than the median because the distribution is skewed to the left.

**19.** The following list shows the selling prices of 8 houses in a certain town.

| House | Price | House | Price |
|-------|-------|-------|-------|
| A | $302,100 | E | $275,800 |
| B | $275,800 | F | $295,000 |
| C | $305,400 | G | $281,900 |
| D | $250,600 | H | $284,700 |

What is the median selling price of the houses in the list?

(A) $263,200

(B) $283,300

(C) $288,450

(D) $290,600

(E) $293,400

## Unit 1 - One Variable Data

20. The following relative frequency table shows the political party affiliation for a sample of 500 people in a certain town.

| Political Party | Relative Frequency |
|---|---|
| Democrat | 0.35 |
| Republican | 0.30 |
| Independent | 0.20 |
| Green Party | 0.11 |
| Libertarian | 0.03 |
| Other | 0.01 |

Which of the following statements is supported by the table?

(A) The number of people affiliated with the Republicans is 30.

(B) The number of people affiliated with the Independents is 100.

(C) Less than half of the people are affiliated with the Democrats or the Republicans.

(D) At least 200 people are affiliated with the Democrats.

(E) At least 80 people are affiliated with the Green Party or the Libertarians.

21. The following table summarizes the number of pies sold at a booth one day at a local farmers market.

| Type of Pie | Frequency |
|---|---|
| Apple | 18 |
| Blueberry | 14 |
| Cherry | 16 |
| Key Lime | 12 |
| Peach | 12 |
| Pumpkin | 18 |

Which of the following statements is supported by the table?

(A) More cherry pies were sold than any other type of pie.

(B) Twice as many apple pies as key lime pies were sold.

(C) More than half the pies sold were apple.

(D) Fewer than 50 pies were sold at the booth that day.

(E) The combined percentage of key lime pies sold and pumpkin pies sold was less than 50%.

## Unit 1 - One Variable Data

22. A sample of 100 students from Liberty High School and a sample of 60 students from Central High School were asked what they planned to do after graduation. Responses fell into five categories: four-year university $(4Y)$, community college $(CC)$, join the workforce $(W)$, join the military $(M)$, or undecided $(UD)$. The results are shown in the following bar chart.
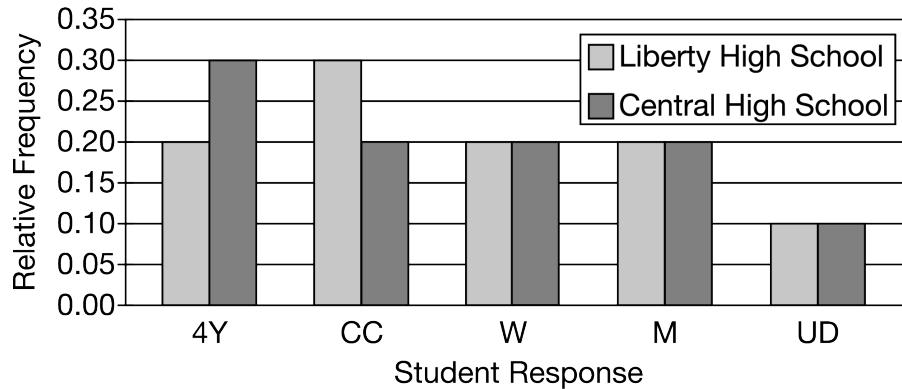


Which of the following statements is supported by the bar chart?

(A) For the category four-year university, the number of students from Central High School was 10 greater than the number of students from Liberty High School.

(B) At Liberty High School, more students selected a four-year university than any other activity.

(C) For the category join the workforce, the number of students from each school was equal.

(D) At Central High School, the same number of students selected four-year university and military.

(E) For the category undecided, the number of students from Liberty High School was 4 greater than the number of students from Central High School.

23. The following table shows summary statistics for the number of hours a group of students spent playing video games last Monday and last Saturday.

| Day | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|
| Monday | 0 | 1 | 2 | 4 | 12 |
| Saturday | 1 | 4 | 6 | 8 | 18 |

Based on the summary statistics, which of the following gives the best comparison of the range and the interquartile range $(IQR)$ of the two days?

## Unit 1 - One Variable Data

(A)  The range and $IQR$ of hours played on Monday are both greater than the range and $IQR$ of hours played on Saturday.

(B)  The range and $IQR$ of hours played on Monday are both less than the range and $IQR$ of hours played on Saturday.

(C)  The range and $IQR$ of hours played on Monday are both equal to the range and $IQR$ of hours played on Saturday.

(D)  The range of hours played on Monday is greater than the range of hours played on Saturday, and the $IQR$ of hours played on Monday is less than the $IQR$ of hours played on Saturday.

(E)  The range of hours played on Monday is less than the range of hours played on Saturday, and the $IQR$ of hours played on Monday is greater than the $IQR$ of hours played on Saturday.

24. The following bar chart shows the relative frequency of days of rain for 30 days in four regions of a certain state.
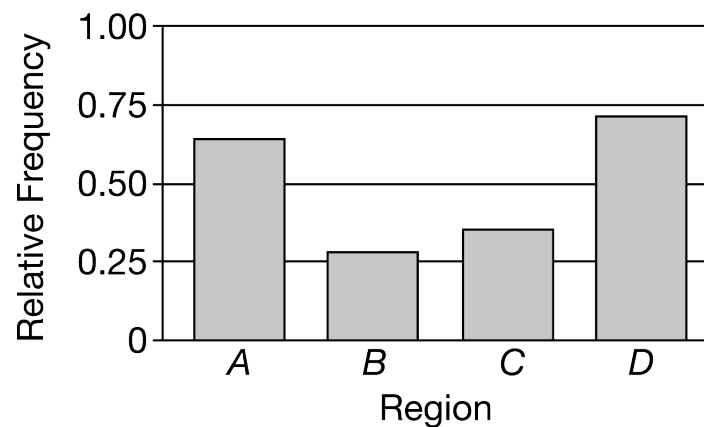


Which of the following statements is <u>not</u> supported by the bar chart?

(A)  Region $D$ had the greatest percentage of days of rain.

(B)  Region $B$ had the least percentage of days of rain.

(C)  Region $A$ had more than 15 days of rain.

(D)  Region $C$ had more than 25 days of rain.

(E)  Region $D$ had less than 23 days of rain.

## Unit 1 - One Variable Data

**25.** The following table shows data for the 8 longest roller coasters in the world as of 2015.

| Length (feet) | Type | Speed (miles per hour) | Height (feet) | Drop (feet) | Continent |
|---|---|---|---|---|---|
| 8,133 | Steel | 95 | 318 | 310 | Asia |
| 7,442 | Steel | 50 | 107 | 102 | Europe |
| 7,359 | Wood | 65 | 110 | 135 | North America |
| 6,709 | Steel | 81 | 259 | 230 | Asia |
| 6,602 | Steel | 95 | 325 | 320 | North America |
| 6,595 | Steel | 93 | 310 | 300 | North America |
| 6,562 | Steel | 149 | 171 | 168 | Asia |
| 6,442 | Wood | 67 | 163 | 154 | North America |

Which of the following variables is categorical?

(A) Length

(B) Type

(C) Speed

(D) Height

(E) Drop

**26.** Which of the following questions about cars in a school parking lot will allow for the collection of a set of categorical data?

(A) How many blue cars are in the lot?

(B) What are the gas mileages, in miles per gallon, of the cars in the lot?

(C) What are the weights, in pounds, of the cars in the lot?

(D) What is the number of cars in the lot with out-of-state license plates?

(E) What are the colors of the cars in the lot?

**27.** The following boxplot summarizes the heights of a sample of 100 trees growing on a tree farm.

28 30 32 34 36 38 40 42 44

Tree Height (inches)

Emily claims that a tree height of 43 inches is an outlier for the distribution. Based on the $1.5 \times \text{IQR}$ rule for outliers, is there evidence to support the claim?

## Unit 1 - One Variable Data

    (A)   Yes, because $(\max - Q3)$ is greater than $(Q1 - \min)$.

    (B)   Yes, because 43 is greater than $(Q3 + IQR)$.

    (C)   Yes, because 43 is greater than $(Q1 - 1.5 \times IQR)$.

    (D)   No, because 43 is not greater than $(Q3 + 1.5 \times IQR)$.

    (E)   No, because 43 is greater than $(Q1 - 1.5 \times IQR)$.

**28.** The following dotplot shows the scores of 25 people who played an online trivia game.
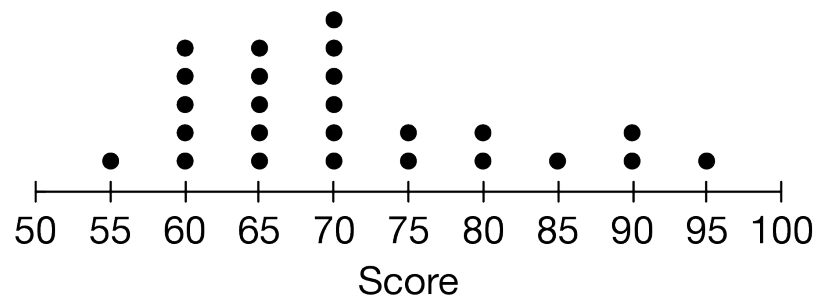


Which of the following statements is the best description of the distribution of scores?

    (A)   The distribution is roughly symmetric.

    (B)   The distribution is roughly uniform.

    (C)   The distribution is skewed left.

    (D)   The distribution is skewed right.

    (E)   The distribution is bimodal.

**29.** Data were collected on 100 United States coins minted in 2018. Which of the following represents a quantitative variable for the data collected?

    (A)   The type of metal used in the coin

    (B)   The value of the coin

    (C)   The color of the coin

    (D)   The person depicted on the face of the coin

    (E)   The location where the coin was minted

**30.** The following list shows the number of video games sold at a game store each day for one week.

15, 43, 50, 39, 22, 16, 20

Which of the following is the best classification of the data in the list?

    (A)   Categorical and continuous

    (B)   Quantitative and continuous

    (C)   Categorical and discrete

    (D)   Quantitative and discrete

    (E)   Neither categorical nor quantitative, and neither discrete nor continuous