COLLEGE OF CHARLESTON

Week 2

Math 104: Elementary Statistics

Recall

- **Descriptive Statistics**
  - Summarize/organize data
  - Methods
    - Numerical descriptions of data
    - Graphical descriptions of data
- **Numerical descriptions of data**
  - Measures of center (3.1)
    - Mean
    - Median
    - Mode
  - Measures of dispersion (3.2)
    - Range
    - Standard Deviation
    - Variance
  - Measures of relative position (3.3)
    - Percentile/Quartile
    - Box plots (graphical display)

Recall

- **Qualitative** data
  - labels of samples
  - often count the number of samples with the same label, or **class**
- **Quantitative** data
  - Numerical measurements of samples
  - **discrete**, only takes on particular values
  - **continuous**, take on any value (at least in an interval)
- Levels of measurement
  - **nominal**, labels or names for each sample (e.g. Trump/Harris/Undecided)
  - **ordinal**, relative ordering of each sample (e.g. satisfied/indifferent/unsatisfied)
  - **interval**, numbers for each sample and zero is a placeholder (e.g. temperature)
  - **ratio**, numbers and zero indicates absence of something (e.g. weight)

3.1 Numerical Descriptions of Data

What is the “center” of data?

- The **center** of a dataset is a typical value in the dataset
  - For quantitative data, describes approximate location of data on number line
- In statistics, there are many measures of center of a dataset. We will explore
  - Mean (interval, ratio)
  - Median (ordinal*, interval, ratio)
  - Mode (nominal, ordinal, interval*, ratio*)
- The “**average**” is a popular and ambiguous term for a measure of center
  - The term “average” can be abused

Mean

- Suppose a population has size N
- Suppose a sample from the population has size n
- The **sample mean** is given by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

Uppercase Greek letter "sigma" Σ is a symbol of summation

- The **population mean** is given by

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum x_i}{N}$$

Greek letter μ "mu"

- The mean is typical NOT a value in the data set
- **Example**

Students were surveyed to find out the number of hours they sleep per night during the semester. Here is a sample of their self-reported responses. Calculate the mean.

5, 6, 8, 10, 4, 6, 9

Answer: $(5 + 6 + 8 + 10 + 4 + 6 + 9)/7 = 48/7 \approx$ **6.9**

Median

- The **median** is the middle value of an ordered array
  - List the dataset in ascending order
  - If there are an *odd* number of values, the median is the middle value
  - If there are an *even* number of values, the median is the mean of the two middle values
- The median may NOT be a value in the dataset
- **Example**

  Given the number of absences for two samples of students, find the median for each sample.

  a. 3, 4, 6, 7, 2, 8, 9

  b. 5, 7, 8, 1, 4, 9, 8, 9

Answer: a) 2, 3, 4, **6**, 7, 8, 9. b) 1, 4, 5, **7**, **8**, 8, 9, 9. Then, (7+8)/2=**7.5**

Mode

- A **mode** of a dataset is a value that occurs most frequently
  - If each value occurs an equal number of times, there is no mode.
- If only one value occurs most often, the data is **unimodal**.
- If two values occur most often, the data is **bimodal**.
- If more than two values occur most often, the data is **multimodal**.
- **Example**

  Given the number of phone calls received each hour during business hours for four different companies, find the mode of each, and state if the data set is unimodal, bimodal, multimodal, or no mode.

  a. 6, 4, 6, 1, **7**, 8, **7**, 2, 5, **7** Answer: 7; unimodal

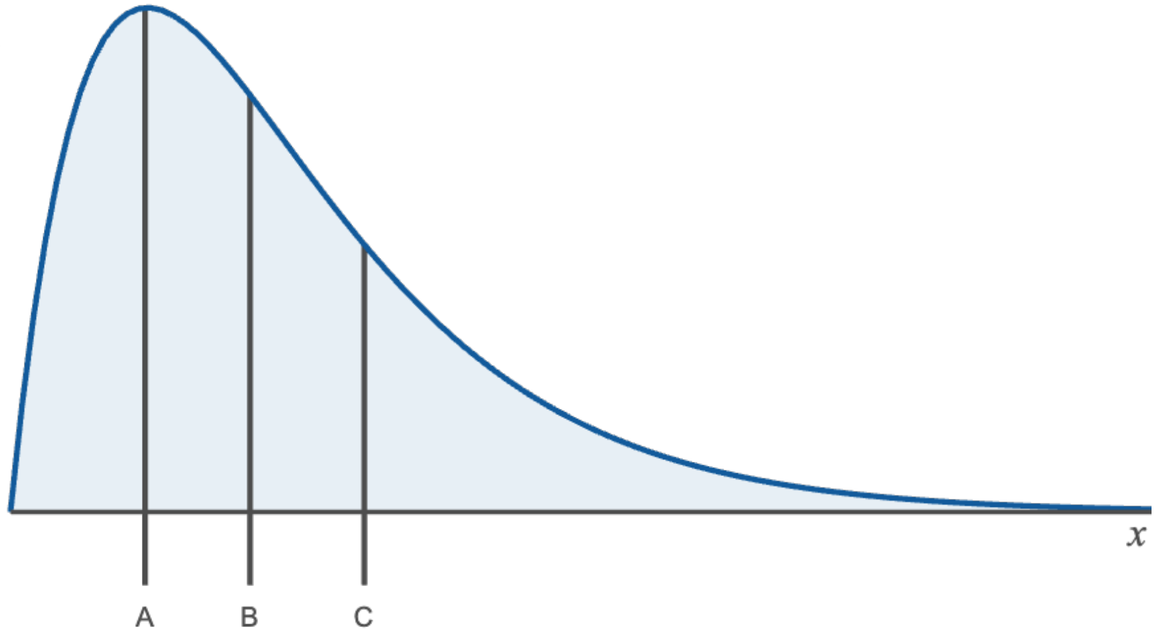
  b. 3, 4, 7, 8, 1, 6, 9 Answer: no mode

  c. **2**, 5, **7**, **2**, 8, **7**, 9, 3 Answer: 2, 7; bimodal

  d. **2, 2, 3, 3, 4, 4, 5, 5** Answer: 2, 3, 4, 5; multimodal

Comparing measures of center

- When to use?
  - **Mean**: Quantitative data with no outliers
  - **Median**: Quantitative data with outliers
  - **Mode**: Qualitative data
- Determining mean/median/mode from a graph?
  - Mode is data value with highest peak
  - Median divides the area of the distribution in half
  - Mean is pulled towards outliers



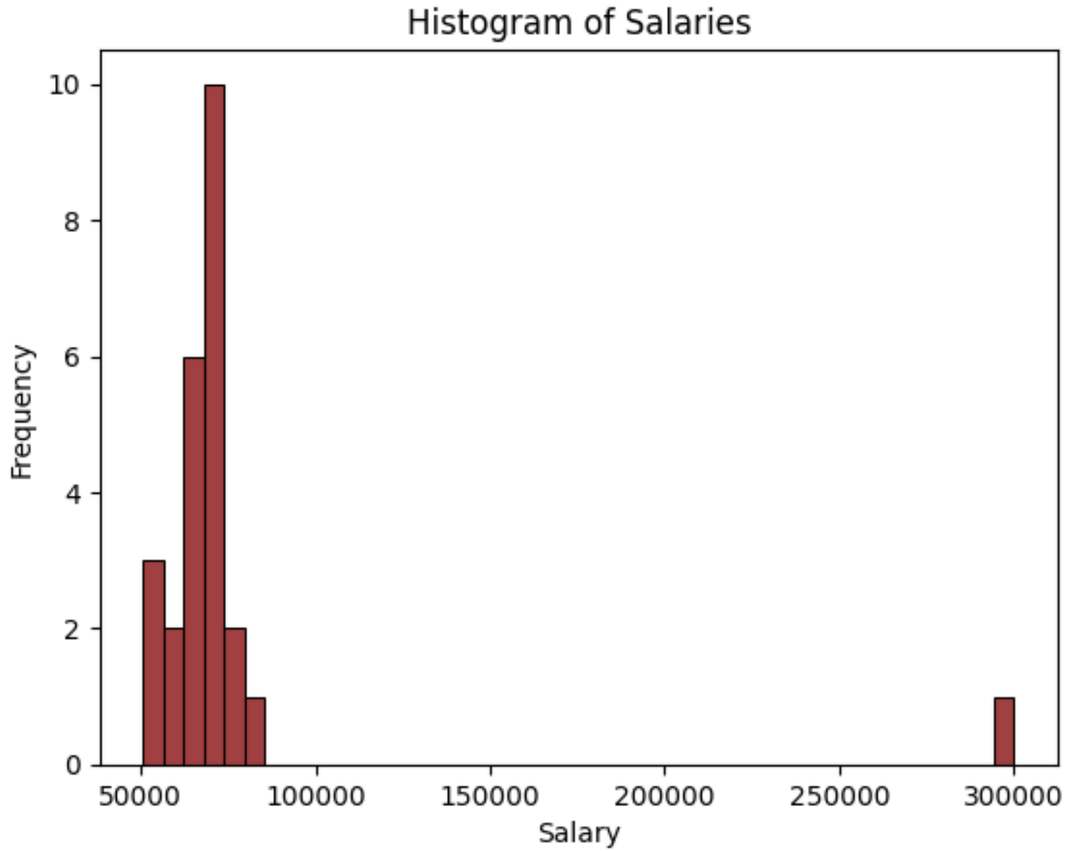
Which of A, B, C is the mean? median? mode?

Answer: A = Mode, B = Median, C= Mean

What is the “average”?

discuss in small groups

- How do the median and mean compare?
  - $\bar{x}$ = 76560
  - median = 69900
- If you were a CEO of a company, and these values were the salaries of employees at the company, which measure of center would you report as the average in a press release?




```
50700,  53400,  54700,  60700,  60800,  62700,  63200,
63900,  64000,  65400,  67900,  68900,  69900,  70700,
70800,  71100,  71300,  71400,  71800,  71900,  72600,
74000,  79600,  82600,  300000
```

3.2 Measures of Dispersion

Range

- The **range** is the difference between the largest and smallest values in the dataset

Range = Maximum Data Value - Minimum Data Value

- **Example**

The following data were collected from samples of call lengths (in minutes) observed for two different mobile phone users. Calculate the range of each data set.

a. 2, 25, 31, 44, 29, 14, 22, 11, 40 Answer: 44 - 2 = 42

b. 2, 2, 44, 2, 2, 2, 2, 2 Answer: 44 - 2 = 42

- In this example, almost all of the data in (b) is 2, but the range is the same.

Variance

- The **variance** is a measure of how far the data values are spread out from the mean.
- The **population variance** is given by

Greek letter σ "sigma"

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

- The **sample variance** is given by

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

- **Example**: Jesmyn has just started driving for Uber. The number of rides she has given each day for a sample of 5 consecutive days is listed below. Find the variance for the number of rides Jesmyn has given: 3, 2, 5, 6, 4.
  - Answer:

$$\bar{x} = (3 + 2 + 5 + 6 + 4)/5 = 20/5 = 4$$

$$s^2 = \frac{1}{5 - 1}\left((3 - 4)^2 + (2 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + (4 - 4)^2\right)$$

$$= \frac{1^2 + 2^2 + 1^2 + 2^2 + 0^2}{4}$$

$$= \frac{10}{4} = 2.5$$

Standard deviation

- The **standard deviation** is a measure of how much we might expect a typical member of the data set to differ from he mean
- The **population standard deviation** is given by

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

- The **sample standard deviation** is given by

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

- **Example**: $s = \sqrt{s^2} = \sqrt{2.5} = 1.58$

Why standard deviation?

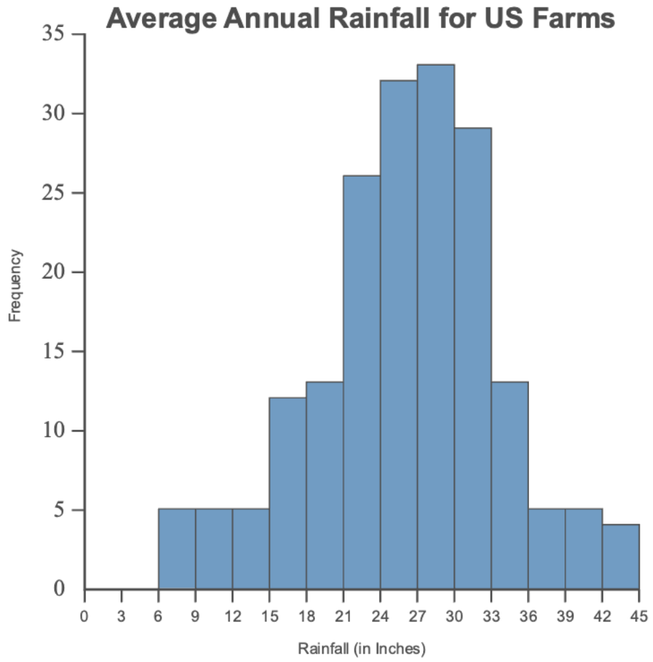
- Why not sum of deviations?

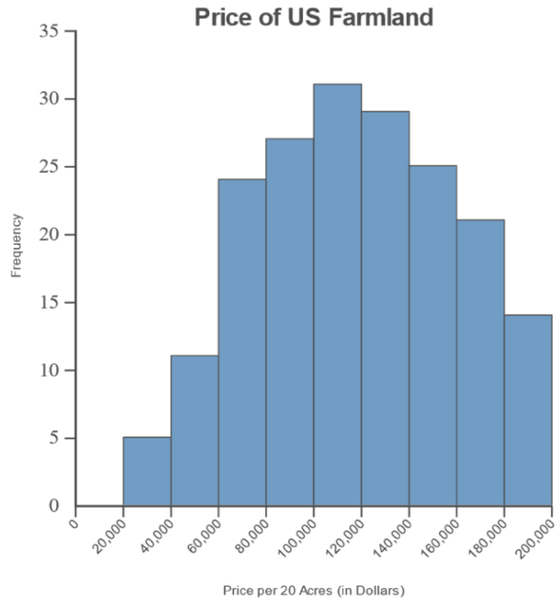
$$\sum (x_i - \bar{x}) = \left(\sum x_i\right) - n\bar{x} = \left(\sum x_i\right) - n\frac{1}{n}\sum x_i = \left(\sum x_i\right) - \left(\sum x_i\right) = 0$$

- What are the units of standard deviation v.s. variance?
  - Suppose x_i is in units. Then,
    - $\sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$ is in units
    - $\frac{1}{n-1}\sum (x_i - \bar{x})^2$ is in units squared
- Why $\frac{1}{n-1}$ instead of $\frac{1}{n}$?
  - The sample standard deviation is an **unbiased estimator** of the population standard deviation! Estimating with $\frac{1}{n}$ will consistently lead to underestimates.

Coefficient of Variation

- Without looking at the numerical labeling of a histogram, we cannot "eyeball" which graph has the larger standard deviation.
- The **coefficient of variation** is the ratio of the standard deviation to the mean as a percentage
  - $CV = \frac{\sigma}{\mu} \cdot 100\,\%$
  - $CV = \frac{s}{\bar{x}} \cdot 100\,\%$
- The graph with the wider-looking spread will, then, correspond to the graph with the larger coefficient of variation.





Case Study: Stocks v.s. Bonds

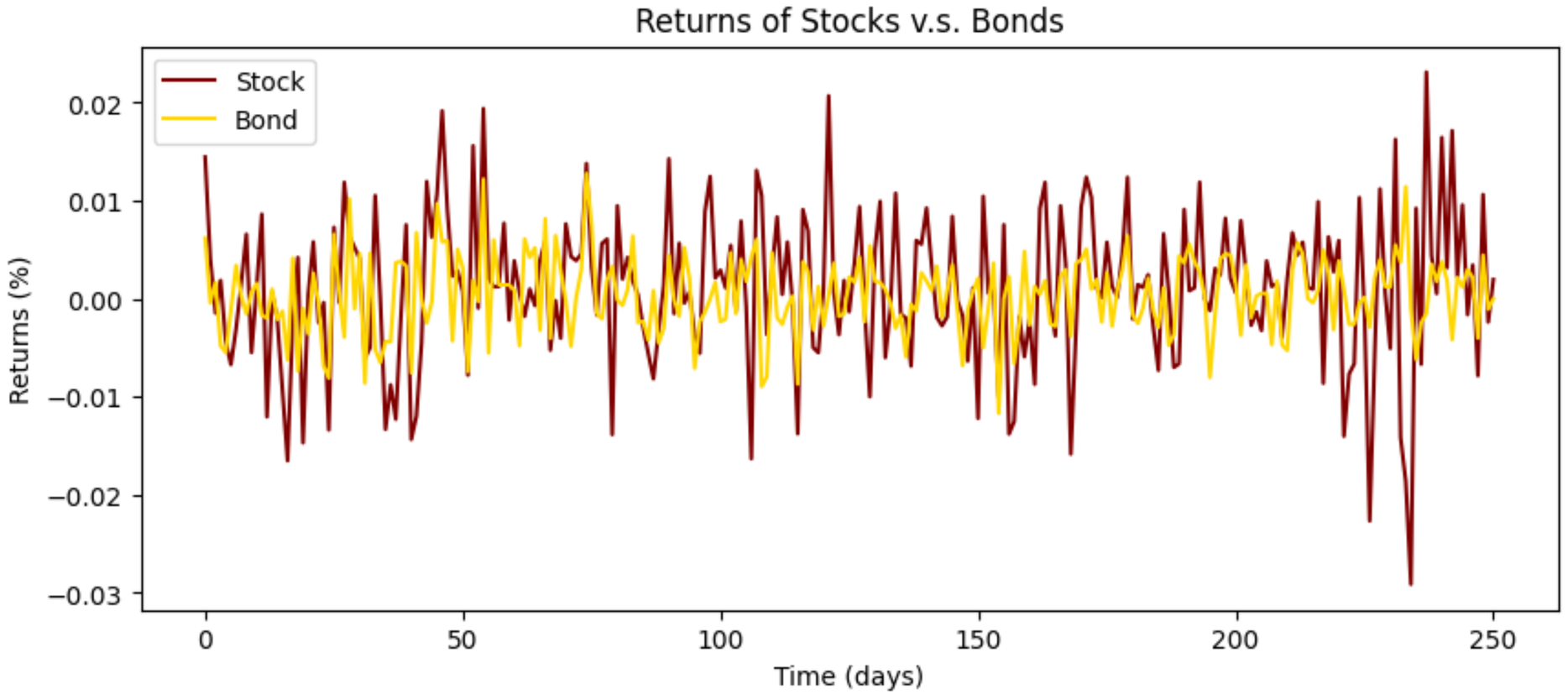
- The returns of a asset are given by $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$ (percent increase/decrease)
- The daily **volatility** is the standard deviation

$$s = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})^2}$$

- The annualized **volatility** is $V = s\sqrt{T}$ where $T = 252$ is the number of yearly trading days.

Case Study: Stocks v.s. Bonds

- Stock volatility: 12.3%
- Bond volatility: 6.3%



Poll Activity

Poll activity

Link: elementary-stats.com (go to the dropdown menu “Class Polls” on top right)

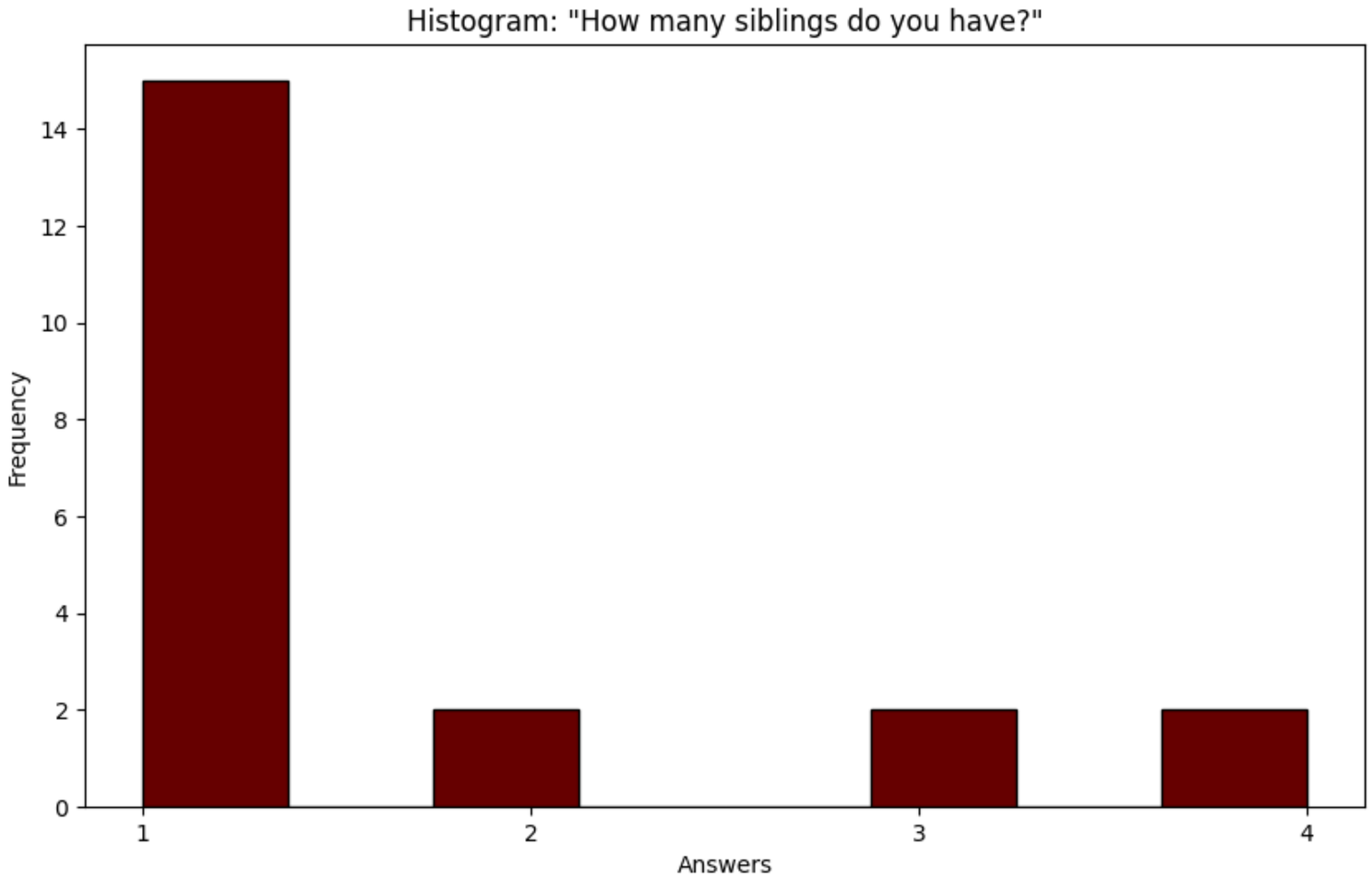
Instructions:

- Step 0: What is the **sample size**? **Population**?
- Step 1 (shape): What is the shape of the distribution? Is it **uniform** (frequency of each class is nearly the same) or **symmetric**? Is it **skewed**? To the right (majority on left side of distribution) or left (majority on the right side of the distribution)?
- Step 2 (outliers): What are the **outliers**?
- Step 3 (center): What is the **sample mean**? What is the **median**? What is the **mode**? Is the data set **unimodal**, **bimodal**, or **multimodal**?
- Step 4 (dispersion): what is the **range**? what is the **sample variance**? **standard deviation**?

Poll 1: How many siblings do you have?

- Step 0:
  - $N = 30$
  - $n = 21$
- Step 1 (shape): what is the shape of the distribution? Is it **uniform** (frequency of each class is nearly the same) or **symmetric**? Is it **skewed**? To the right (majority on left side of distribution) or left (majority on the right side of the distribution)?
  - Answer: right skewed
- Step 2 (outliers): What are the **outliers**?
  - Answer: HARD TO TELL
- Step 3 (measures of center): What is the **sample mean**? What is the **median**? What is the **mode**? Is the data set **unimodal**, **bimodal**, or **multimodal**?
  - $\bar{x} = 1.57$
  - median = 1
  - mode = 1
- Step 4 (measures of dispersion): what is the **range**? what is the **sample variance**? **standard deviation**?
  - range = 3
  - $s^2 = 1.01$
  - $s = 1.00$

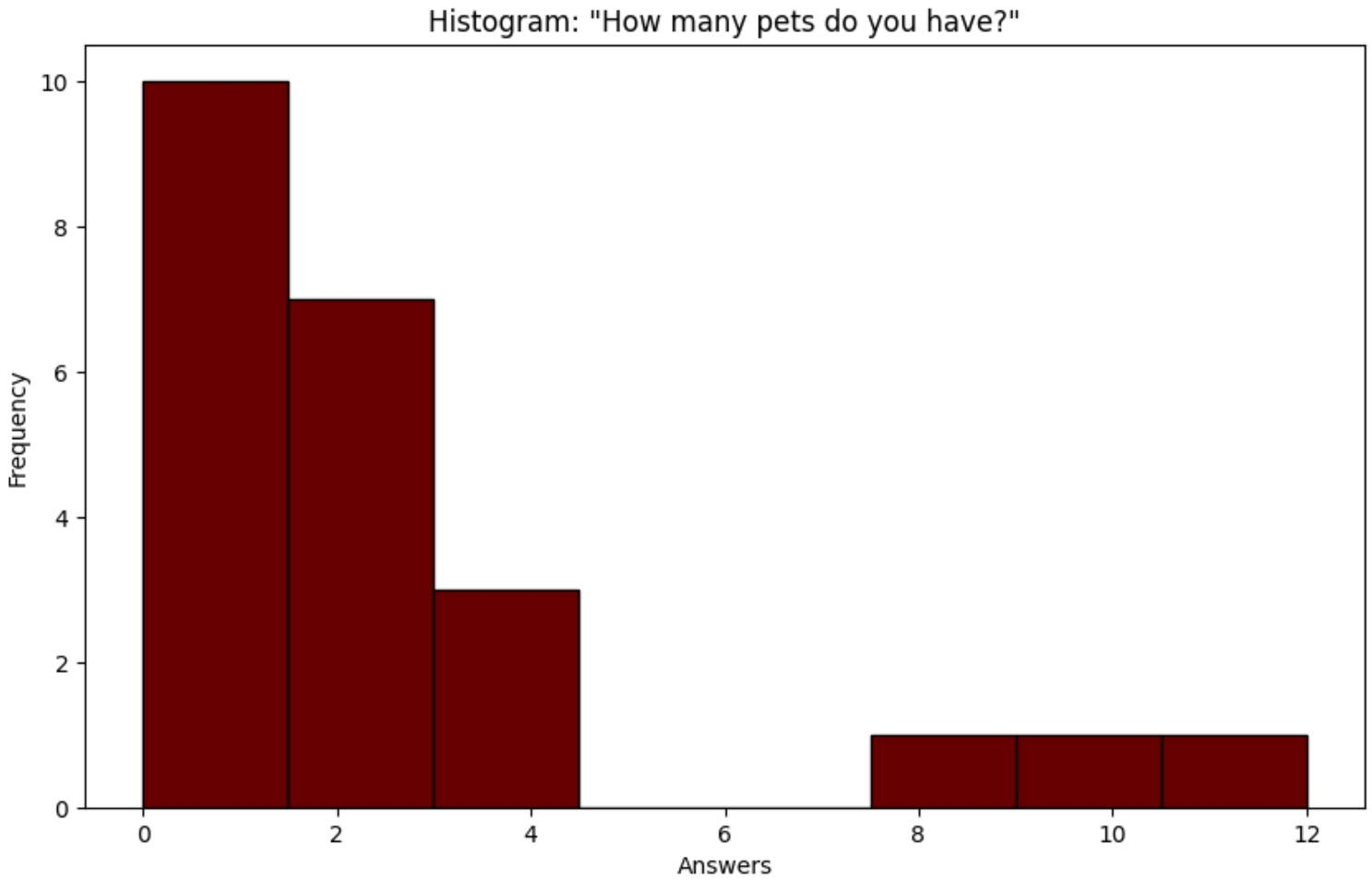
Poll 1: Histogram



Poll 2: How many pets do you have?

- Step 0:
  - $N = 30$
  - $n = 23$
- Step 1 (shape): what is the shape of the distribution? Is it **uniform** (frequency of each class is nearly the same) or **symmetric**? Is it **skewed**? To the right (majority on left side of distribution) or left (majority on the right side of the distribution)?
  - Answer: right skewed
- Step 2 (outliers): What are the **outliers**?
  - Answer: 8, 10, 12
- Step 3 (measures of center): What is the **sample mean**? What is the **median**? What is the **mode**? Is the data set **unimodal**, **bimodal**, or **multimodal**?
  - $\bar{x} = 2.61$
  - median = 2
  - mode = 2
- Step 4 (measures of dispersion): what is the **range**? what is the **sample variance**? **standard deviation**?
  - range = 12
  - $s^2 = 9.8$
  - $s = 3.13$

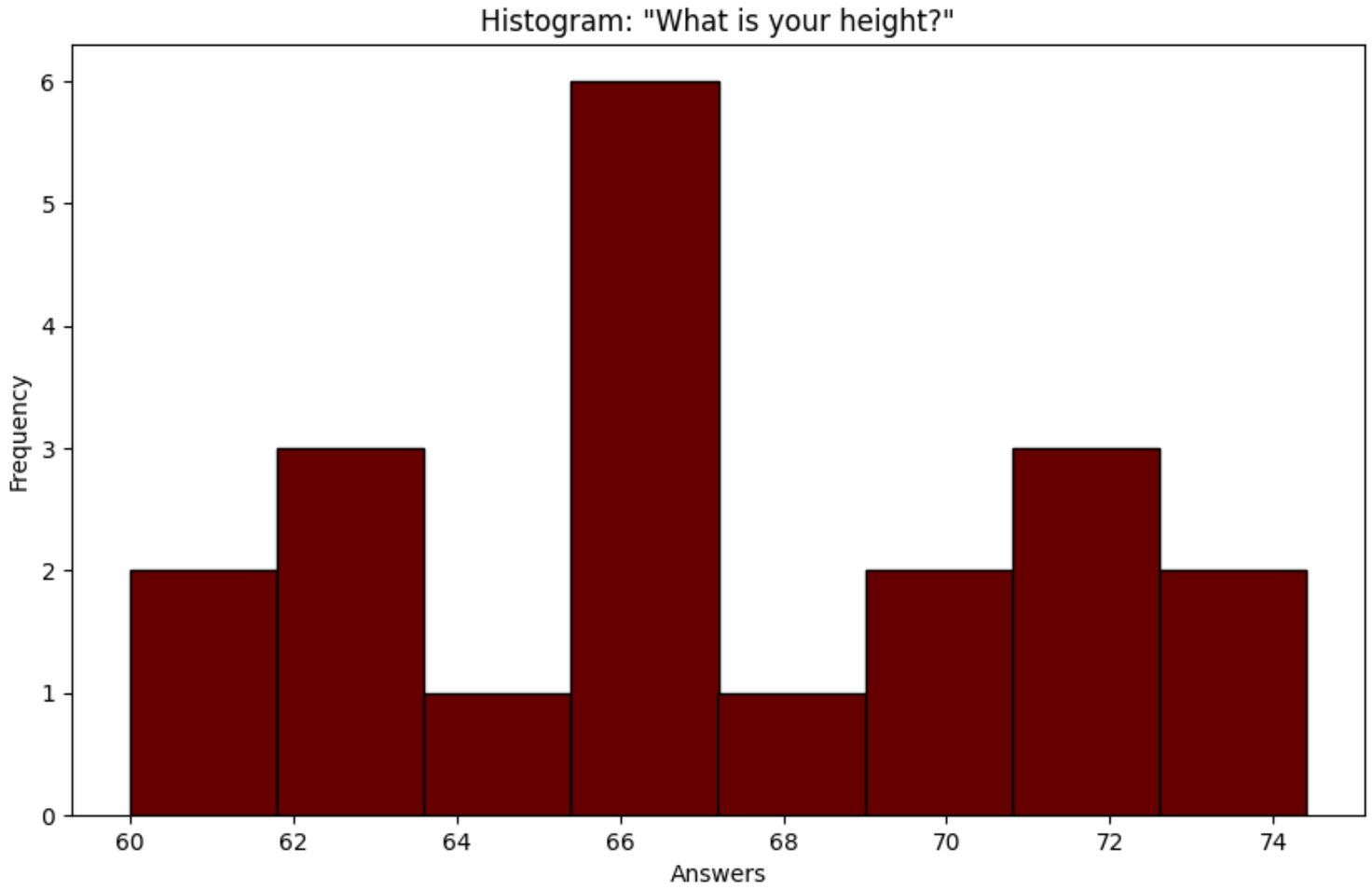
Poll 3: Histogram



Poll 3: What is your height?

- Step 0:
  - $N = 30$
  - $n = 21$
- Step 1 (shape): what is the shape of the distribution? Is it **uniform** (frequency of each class is nearly the same) or **symmetric**? Is it **skewed**? To the right (majority on left side of distribution) or left (majority on the right side of the distribution)?
  - Answer: symmetric
- Step 2 (outliers): What are the **outliers**?
  - Answer: none
- Step 3 (measures of center): What is the **sample mean**? What is the **median**? What is the **mode**? Is the data set **unimodal**, **bimodal**, or **multimodal**?
  - $\bar{x} = 67.07$
  - median = 67
  - mode = 66
- Step 4 (measures of dispersion): what is the **range**? what is the **sample variance**? **standard deviation**?
  - range = 14
  - $s^2 = 17.33$
  - $s = 4.16$

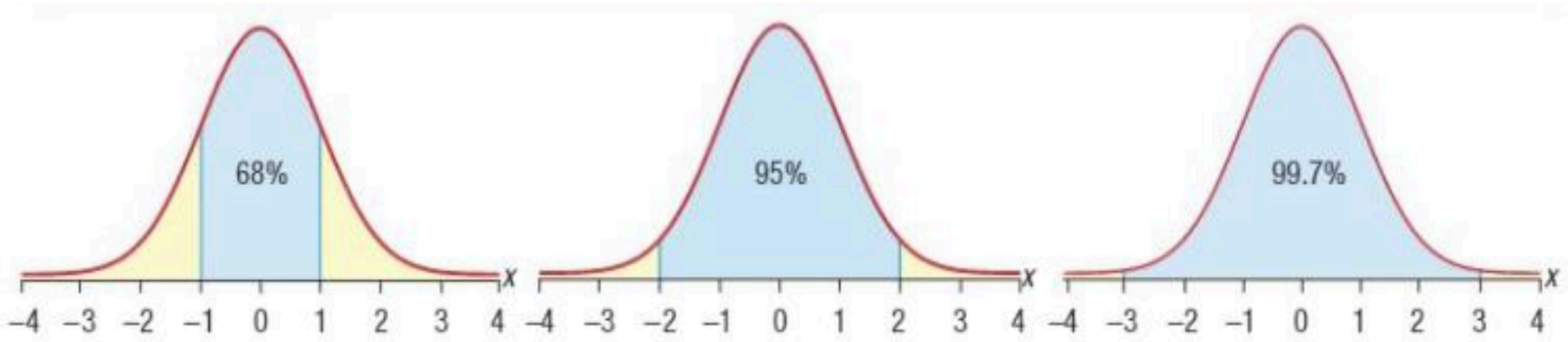
Poll 2: Histogram



3.2 Measures of Dispersion (continued)

Empirical rule

- When a distribution is, bell shaped the **Empirical Rule** can be used to estimate the percentage of values within a few standard deviations from he mean.
- **The Empirical Rule**
  - Approximately 68% of the data values lie within 1 standard deviation
  - Approximately 95% of the data values lie within 2 standard deviations
  - Approximately 99% of the data values lie within 3 standard deviations



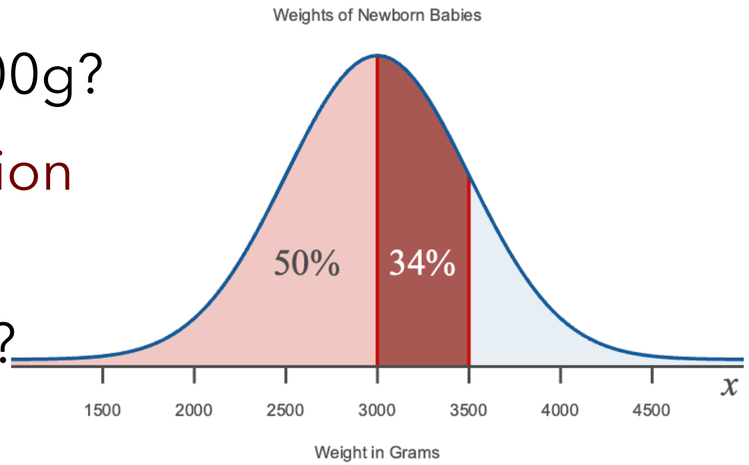
Empirical rule

- **Example**: weights of newborns is bell-shaped with μ = 3000g and σ = 500g
  - a) what percentage of babies weight between 2000-4000g?

    2000 = $\mu - 2\sigma$, 4000 = $\mu + 2\sigma$. Hence, 95% of population lies in this range.
  - b) what percentage of newborn weight less than 3500g?

    3500 = $\mu + \sigma$. Hence, by symmetry 34% lie between 3000-3500g. 50% of the data values is < μ = 3000. Hence, 84% weigh less than 3500g.
  - c) what is the range of weights containing 68% of the population?

    68% is the range $\mu - \sigma$ to $\mu + \sigma$ which is 2500-3500g



Chebyschev's Theorem

- Problem: cannot apply the Empirical Rule to distributions that are not bell-shaped.
- **Chebyschev's Theorem**
  - The proportion of data that lie within K standard deviations of the mean is AT LEAST

$$1 - \frac{1}{K^2} \text{ for } K > 0$$

  - For $K = 2$, at least 75% of the data lie within 2σ of the mean
  - For $K = 3$, at least 88.9% of the data lie within 3σ of the mean
  - For $K = 4$, at least 93.7% of the data lie within 4σ of the mean
- Can apply the theorem for K values that are not integers.
- **Example**: a dataset not bell-shaped has a mean of 50 and a standard deviation of 5. Determine the minimum percentage of data values that lie between 40 and 60.
  - Answer: $40 = \mu - 2\sigma = 50 - 2 \cdot 5.\ 60 = \mu + 2\sigma = 50 + 2 \cdot 5 = 60$. Hence, $1 - \frac{1}{2^2} = 0.75$ of the data, or **75%**, is between 40 and 60.

3.3 Measures of Relative Position

Relative Position: Percentiles

- Often useful to know
  - where in the data a particular value is located
  - how many values that are above or below a particular value
- Divide data into equal parts known as **percentiles**
- Location of data value in Pth percentile is $l = n \cdot \frac{P}{100}$
  - If *not a whole number,* the value is the value of the next location (i.e. round up)
  - If a *whole number*, the value is the mean of the value at the location and the value at the next location.
- The Pth percentile of a value is $P = \frac{l}{n} \cdot 100$ where l is the number of values in the data set *less than or equal* to the given value

Percentiles: Example

- **Example**: A car manufacturer is studying the highway miles per gallon (mpg) for a wide range of makes and models of vehicles. The following stem-and-leaf plot contains the average highway mpg for each of the 135 different vehicles the manufacturer tested.

  a. Find the value of the 10th percentile.

  b. Find the value of the 20th percentile.

Highway Gas Mileage for Various Vehicles	
Stem	**Leaves**
12	1
13	3
14	1
15	5 6
16	1 1 7 8
17	0 0 1 2 3 4 4 5 6 9
18	2 3 4 5
19	1 2 2 2 3 3 4 6 6 7 8 9
20	1 2 3 3 3 4 5 6 6 7 8
21	0 1 1 2 3 5 7 8 9
22	2 3 4 7 8 9
23	1 1 1 4 4 5 6 6 6 6 7 8 9 9
24	0 1 2 3 4 4 4 5 5 5 5 6 7 8 8 8 9 9
25	0 0 1 1 1 2 3 3 3 3 4 4 5 6 6 7 8 9
26	0 0 0 1 2 5 5 6 7 9
27	1 4 7
28	3 5
29	2 4 9
30	0 7
31	3
32	7
33	
34	5
35	9

Percentiles

- **Answer**: First, it is important to notice that the data values are presented in an ordered stem-and-leaf plot, as it is essential that the data values be in numerical order. This is an important first step since the location of the percentile refers to the location in the ordered array of values.
- Part a)
  - There are 135 values in this data set, thus n=135. We want the 10th percentile, so P=10. Substituting these values into the formula for the location of a percentile gives us the following.

$$l = n \cdot \frac{P}{100} = 135 \cdot \frac{10}{100} = 13.5$$

  - Since the formula resulted in a decimal value for l, we round the number 13.5 to the next larger whole number, 14, to determine the location. Thus, the 10th percentile is approximately the value in the 14th spot in the data set. Counting data values, we find that the 14th value is 17.3. Thus, the **value of the 10th percentile of this data set is 17.3 mpg**. This means that approximately 10% of the values in the data set are less than or equal to 17.3 mpg.

Highway Gas Mileage for Various Vehicles	
Stem	**Leaves**
12	1
13	3
14	1
15	5 6
16	1 1 7 8
17	0 0 1 2 3 4 4 5 6 9
18	2 3 4 5
19	1 2 2 2 3 3 4 6 6 7 8 9
20	1 2 3 3 3 4 5 6 6 7 8
21	0 1 1 2 3 5 7 8 9
22	2 3 4 7 8 9
23	1 1 1 4 4 5 6 6 6 6 7 8 9 9
24	0 1 2 3 4 4 4 5 5 5 5 6 7 8 8 8 9 9
25	0 0 1 1 1 2 3 3 3 3 4 4 5 6 6 7 8 9
26	0 0 0 1 2 5 5 6 7 9
27	1 4 7
28	3 5
29	2 4 9
30	0 7
31	3
32	7
33	
34	5
35	9

Percentiles

- Part b)
  - We still have n=135, but to find the value of the 20th percentile, P=20. Substituting these new values into the formula, we get the following.

$$l = n \cdot \frac{P}{100}$$

$$l = 135 \cdot \frac{20}{100}$$

$$l = 27$$

  - Since the value calculated for l is a whole number, we must find the mean of the data value in that location and the one in the next larger location. Thus, the 20th percentile is the arithmetic mean of the 27th and 28th values in the data set, which are 19.2 and 19.3, respectively. Hence the **value of the 20th percentile is 19.25 mpg.** This means that approximately 20% of the values in the data set are less than or equal to 19.25 mpg.

Highway Gas Mileage for Various Vehicles	
Stem	**Leaves**
12	1
13	3
14	1
15	5 6
16	1 1 7 8
17	0 0 1 2 3 4 4 5 6 9
18	2 3 4 5
19	1 2 2 2 3 3 4 6 6 7 8 9
20	1 2 3 3 3 4 5 6 6 7 8
21	0 1 1 2 3 5 7 8 9
22	2 3 4 7 8 9
23	1 1 1 4 4 5 6 6 6 6 7 8 9 9
24	0 1 2 3 4 4 4 5 5 5 5 6 7 8 8 8 9 9
25	0 0 1 1 1 2 3 3 3 3 4 4 5 6 6 7 8 9
26	0 0 0 1 2 5 5 6 7 9
27	1 4 7
28	3 5
29	2 4 9
30	0 7
31	3
32	7
33	
34	5
35	9

Percentiles: Example

- **Example**: The Nissan Xterra averaged 21.1 mpg. In what percentile is this value?

Highway Gas Mileage for Various Vehicles	
Stem	**Leaves**
12	1
13	3
14	1
15	5 6
16	1 1 7 8
17	0 0 1 2 3 4 4 5 6 9
18	2 3 4 5
19	1 2 2 2 3 3 4 6 6 7 8 9
20	1 2 3 3 3 4 5 6 6 7 8
21	0 1 1 2 3 5 7 8 9
22	2 3 4 7 8 9
23	1 1 1 4 4 5 6 6 6 6 7 8 9 9
24	0 1 2 3 4 4 4 5 5 5 5 6 7 8 8 8 9 9
25	0 0 1 1 1 2 3 3 3 3 4 4 5 6 6 7 8 9
26	0 0 0 1 2 5 5 6 7 9
27	1 4 7
28	3 5
29	2 4 9
30	0 7
31	3
32	7
33	
34	5
35	9

Percentiles: Example

- **Answer**: We begin by making sure that the data are in order from smallest to largest. We know from the previous example that they are, so we can proceed with the next step.
  - The Xterra's value of 21.1 mpg is repeated in the data set, in both the 48th and 49th positions, so we will pick the one with the largest location value, which is the 49th. Using a sample size of n=135 and a location of l=49, we can substitute these values into the formula for the percentile of a given data value, which gives us the following.

$$P = \frac{l}{n} \cdot 100$$

$$P = \frac{49}{135} \cdot 100$$

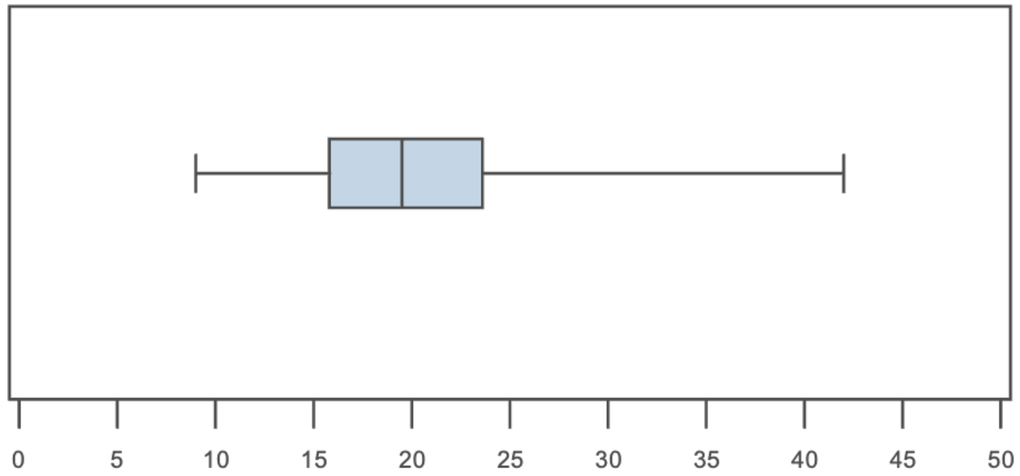
$$P \approx 36.3$$

  - Since we always need to round a percentile to a whole number, we round 36.3 to 36. Thus, approximately 36% of the data values are less than or equal to the Xterra's mpg rating. That is, 21.1 mpg is in the **36th percentile** of the data set.

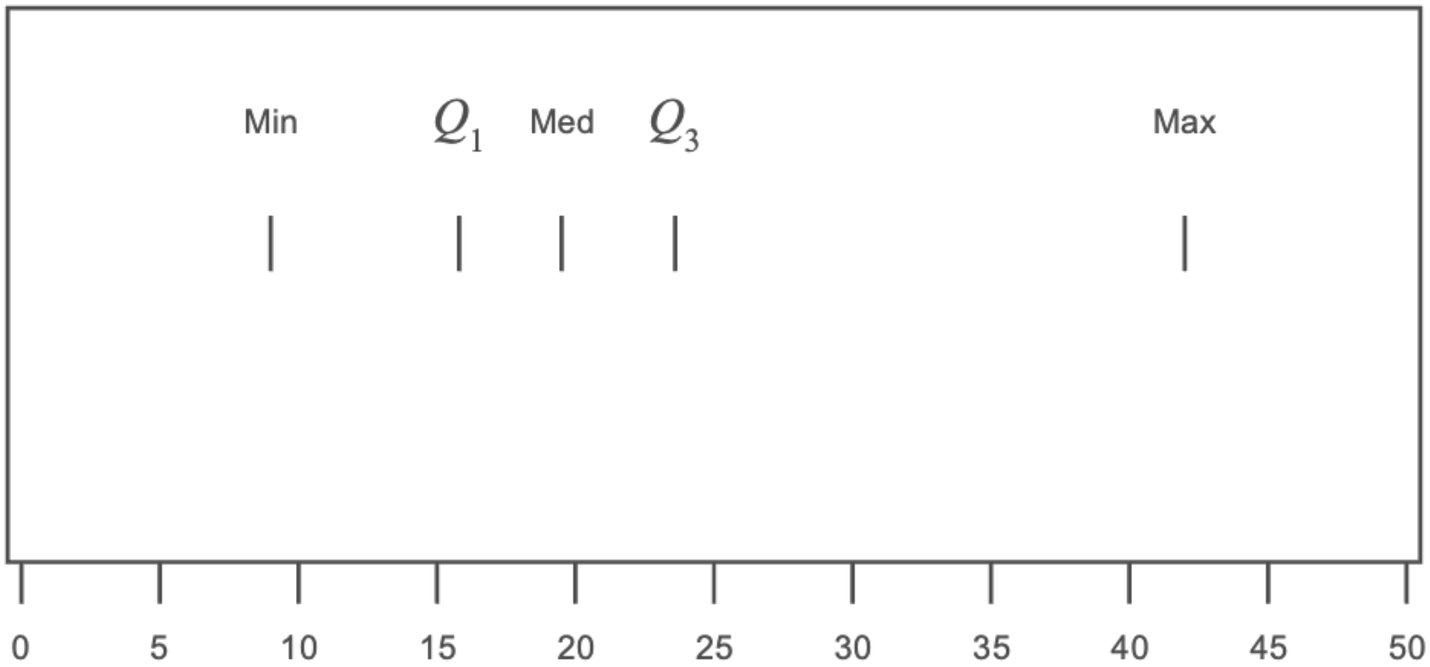
Highway Gas Mileage for Various Vehicles	
Stem	**Leaves**
12	1
13	3
14	1
15	5 6
16	1 1 7 8
17	0 0 1 2 3 4 4 5 6 9
18	2 3 4 5
19	1 2 2 2 3 3 4 6 6 7 8 9
20	1 2 3 3 3 4 5 6 6 7 8
21	0 1 1 2 3 5 7 8 9
22	2 3 4 7 8 9
23	1 1 1 4 4 5 6 6 6 6 7 8 9 9
24	0 1 2 3 4 4 4 5 5 5 5 6 7 8 8 8 9 9
25	0 0 1 1 1 2 3 3 3 3 4 4 5 6 6 7 8 9
26	0 0 0 1 2 5 5 6 7 9
27	1 4 7
28	3 5
29	2 4 9
30	0 7
31	3
32	7
33	
34	5
35	9

Box plots

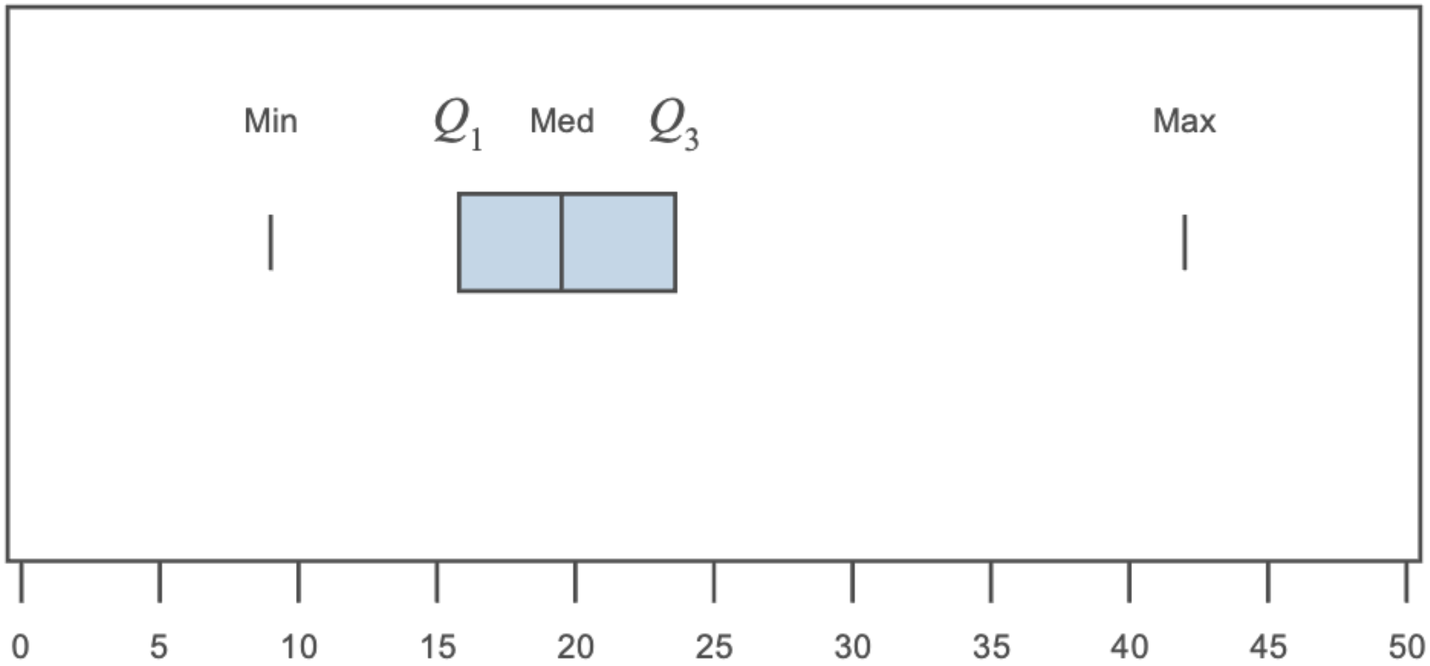
- 5-number summary
  - $\min$ (minimum)
  - Q_1 (1st quartile)
  - Q_2 (median)
  - Q_3 (3rd quartile)
  - $\max$ (maximum)
- Draw a box plot
  - $IQR = Q_3 - Q_2$ is the length of box
  - $\text{range} = \max - \min$ is the length of whiskers
- Sometimes…
  - length of whiskers is range with outliers removed
  - outliers marked with ∘



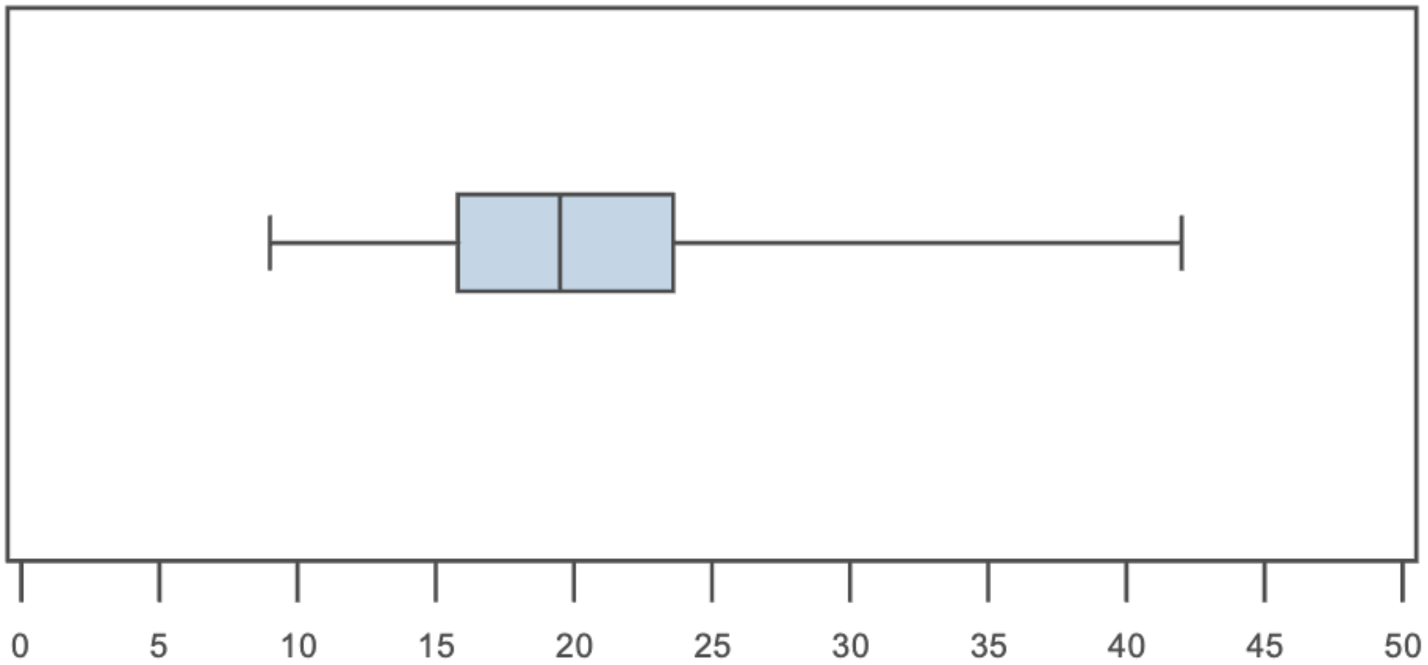
Box plots



Box plots

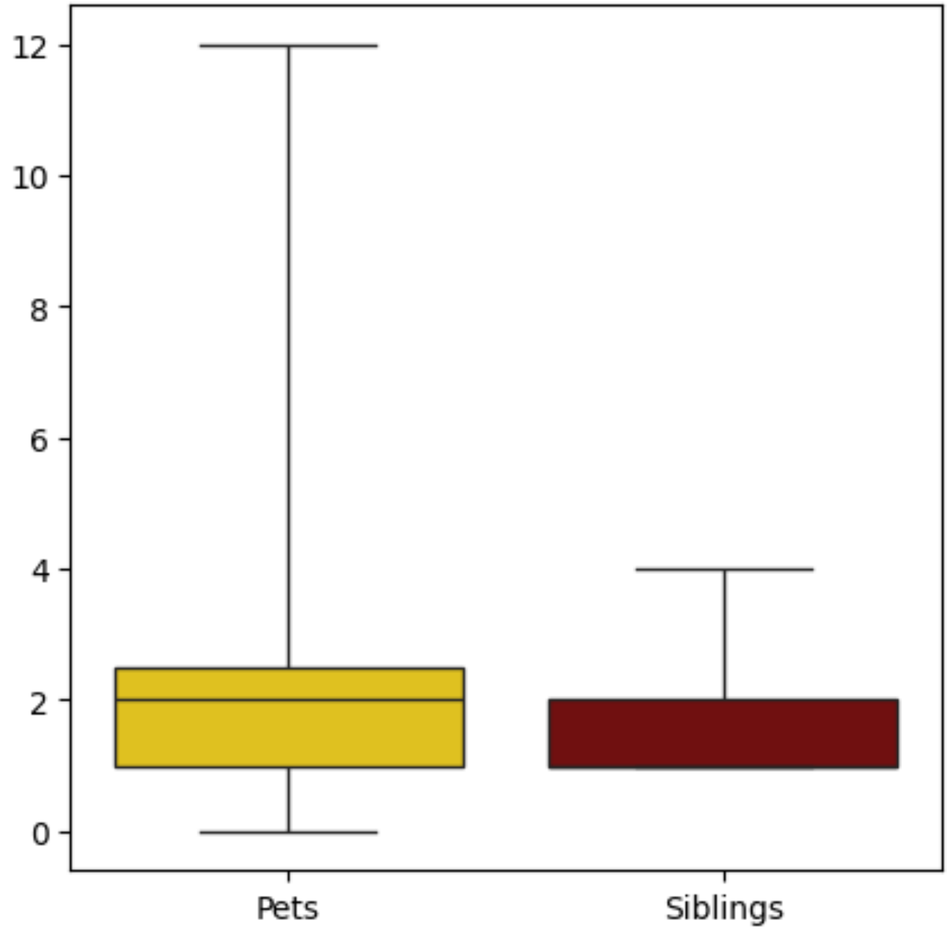


Box plots



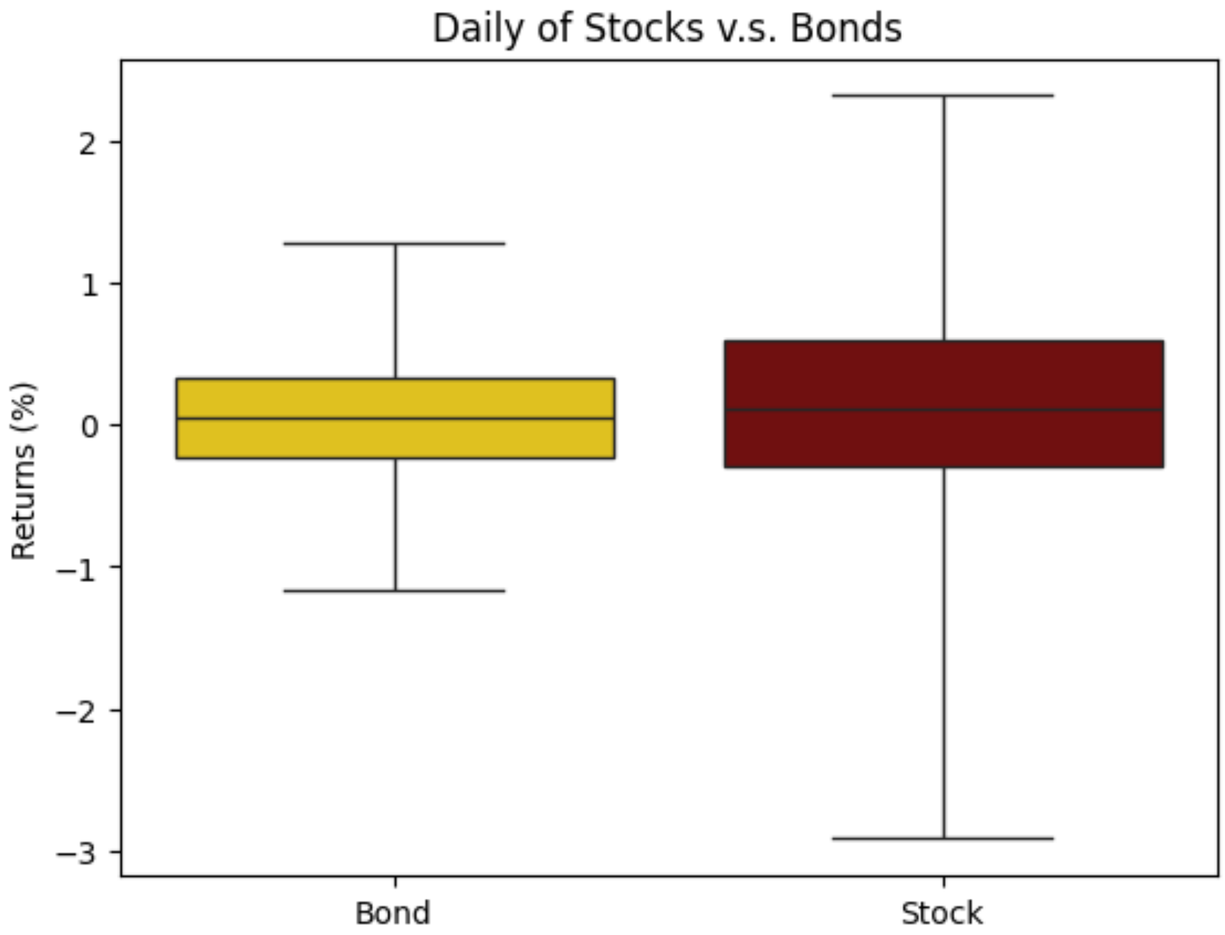
Example: Class Polls

- Pets
  - $\min = 0$, minimum
  - $Q_1 = 1$, 1st quartile
  - $Q_2 = 2$, median
  - $Q_3 = 2.5$, 3rd quartile
  - $\max = 12$, maximum
- Siblings
  - $\min = 1$, minimum
  - $Q_1 = 1$, 1st quartile
  - $Q_2 = 1$, median
  - $Q_3 = 2$, 2nd quartile
  - $\max = 4$, maximum

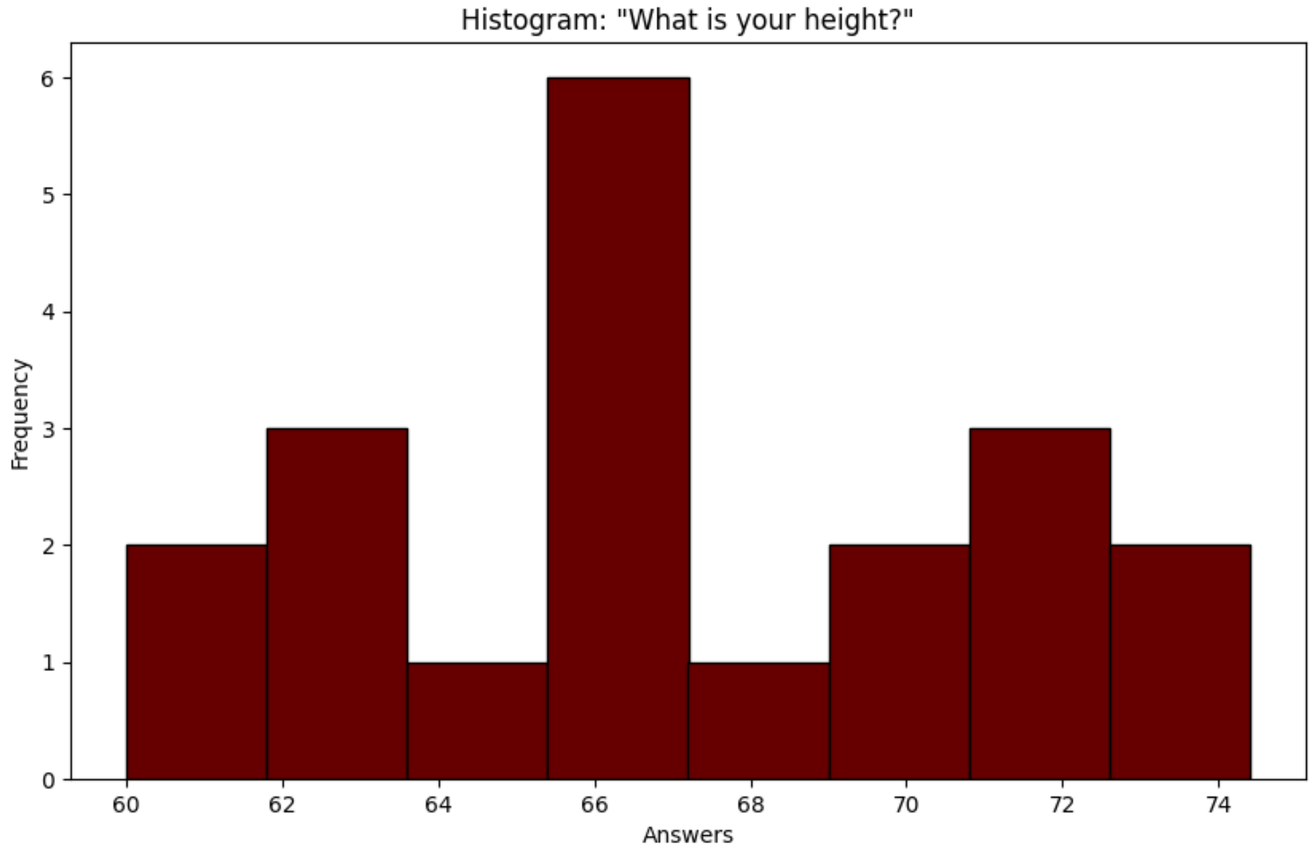


Example: Finance

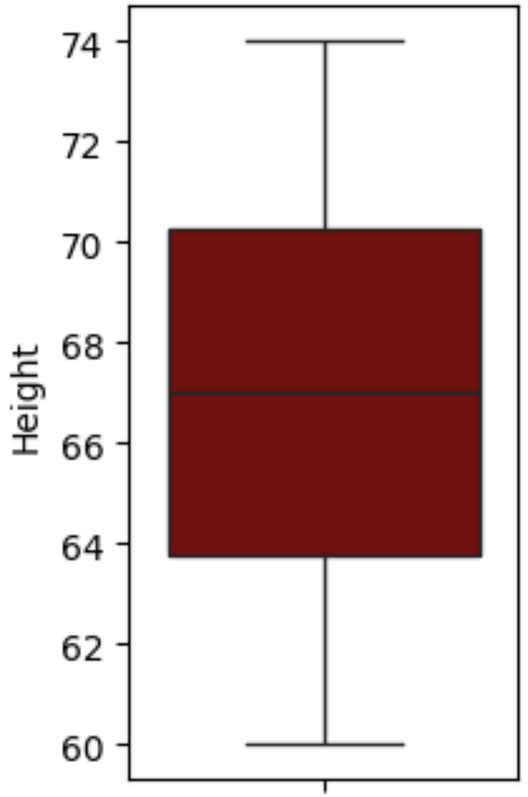
- Stocks
  - $\min$ = -2.9%, minimum
  - Q_1 = -0.28%, 1st quartile
  - Q_2 = 0.11%, median
  - Q_3 = 0.59%, 3nd quartile
  - $\max$ = 2.3%, maximum
- Bonds
  - $\min$ = -1.1%, minimum
  - Q_1 = -0.23%, 1st quartile
  - Q_2 = 0.05%, median
  - Q_3 = 0.32%, 3nd quartile
  - $\max$ = 1.2%, maximum



Discuss: Histograms v.s. Box Plots



v.s.

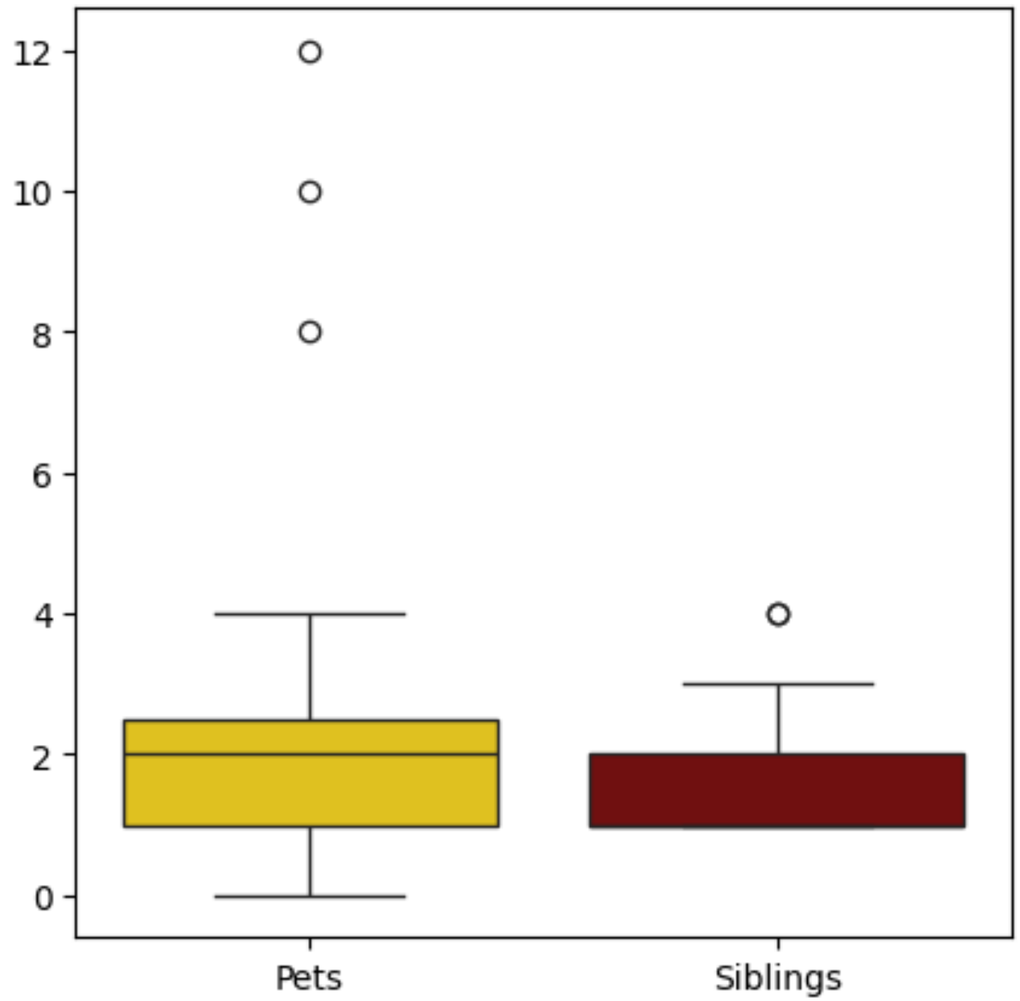


Outliers

- We can identify quartiles with percentiles
- Procedure:
  - The inter-quartile range (IQR) is the range $IQR = Q_3 - Q_2$
  - Multiply $1.5 \cdot IQR$
  - The lower threshold is $Q_1 - 1.5 \cdot IQR$
  - The upper threshold is $Q_3 + 1.5 \cdot IQR$
  - Any value greater than the upper threshold or lower than the lower threshold is an outlier!

Removing Outliers in Box Plot

- Pets
  - Q_1 = 1, 1st quartile
  - Q_2 = 2, median
  - Q_3 = 2.5, 3rd quartile
  - outliers:
    - 8, 10, 12
  - $\min$ = 0, minimum (no lower outliers)
  - $\max$ = 4, maximum (after removing outliers)
- Siblings
  - Q_1 = 1, 1st quartile
  - Q_2 = 1, median
  - Q_3 = 2, 3rd quartile
  - outliers:
    - 4
  - $\min$ = 1, minimum (no lower outliers)
  - $\max$ = 3, maximum (after removing outliers)



z-scores (very important)

- We can compare values from different populations by comparing their percentiles.
- We can also compare how the values relate to the respective means of datasets
  - leads to standard scores or **z-scores**!
- A **standard score for a population** value is given by

$$z = \frac{x - \mu}{\sigma}$$

- A **standard score for a sample** value is given by

$$z = \frac{x - \bar{x}}{s}$$

- Example: if the mean score of the math section of the SAT is 500 with a standard deviation of 100 points, what is the standard score for a student who scored 630?

Answer: $\mu = 500$ and $\sigma = 100$. If $x = 630$, then $z = \frac{x - \mu}{\sigma} = \frac{630 - 500}{100} = \frac{130}{100} = 1.3$