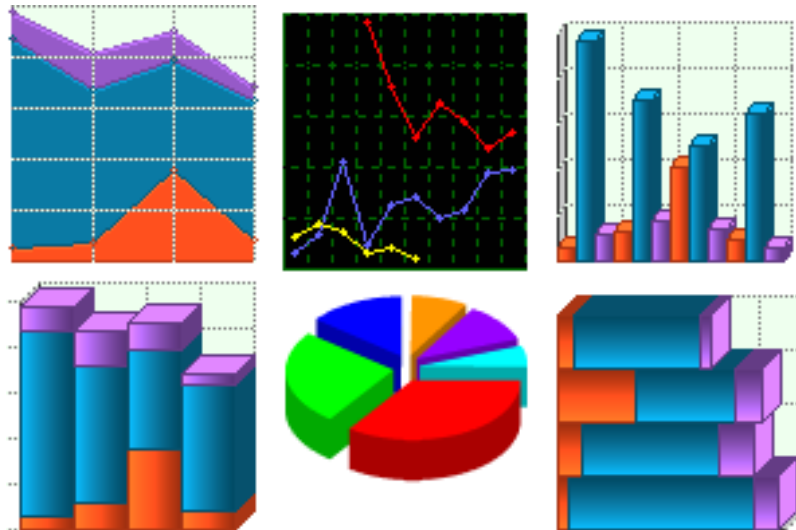


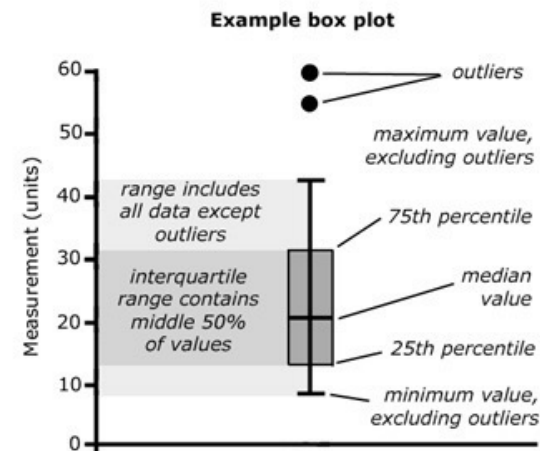
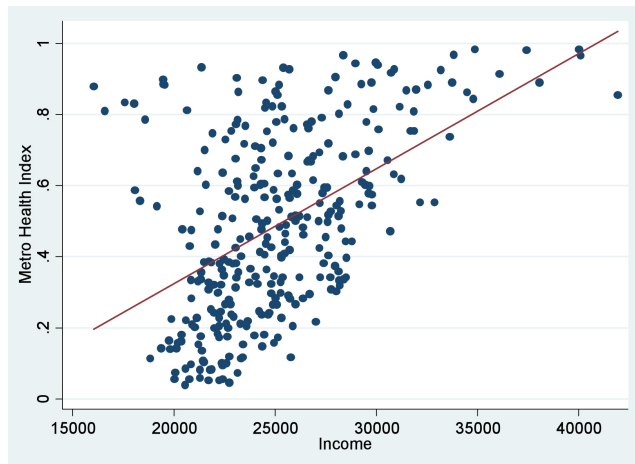
# Chapter 2

## Picturing Variation with Graphs



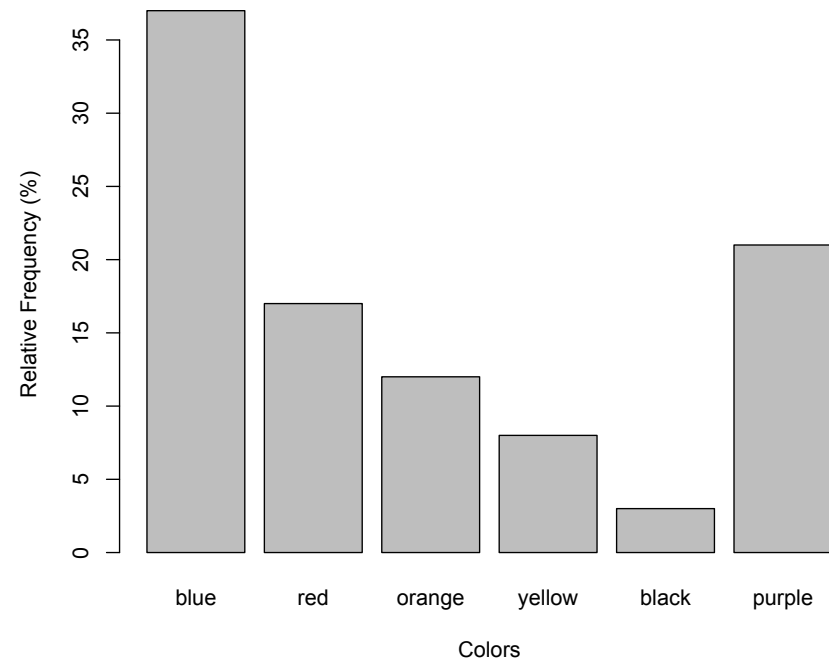
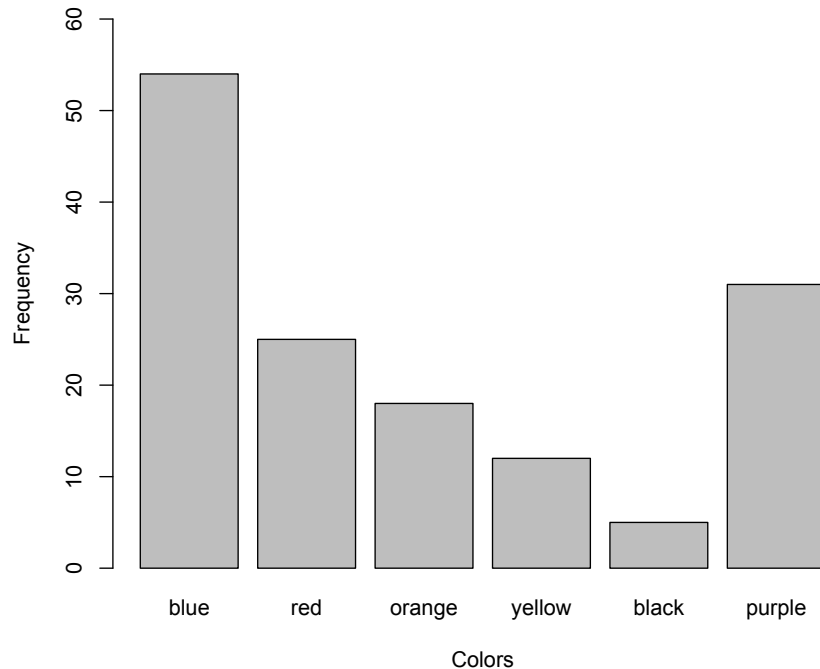
# How do we visualize data?

- By making a picture!
- Organizing the data in the form of a chart, graph or plot can be an effective way of looking at trends, patterns, outliers and other relationships.
- Visualizing data can also help us to make predictions and conclusions.



# Bar Charts

- **Bar charts** display the counts or frequencies of each category next to each other for easy comparison.
  - Bar charts are used for categorical variables.
  - Typically, the order of the bars does not matter.
- **Relative frequency bar charts** display relative proportions of each category.
  - Frequency is how many times a value occurred in the data.



145 people were asked to state their favorite color.

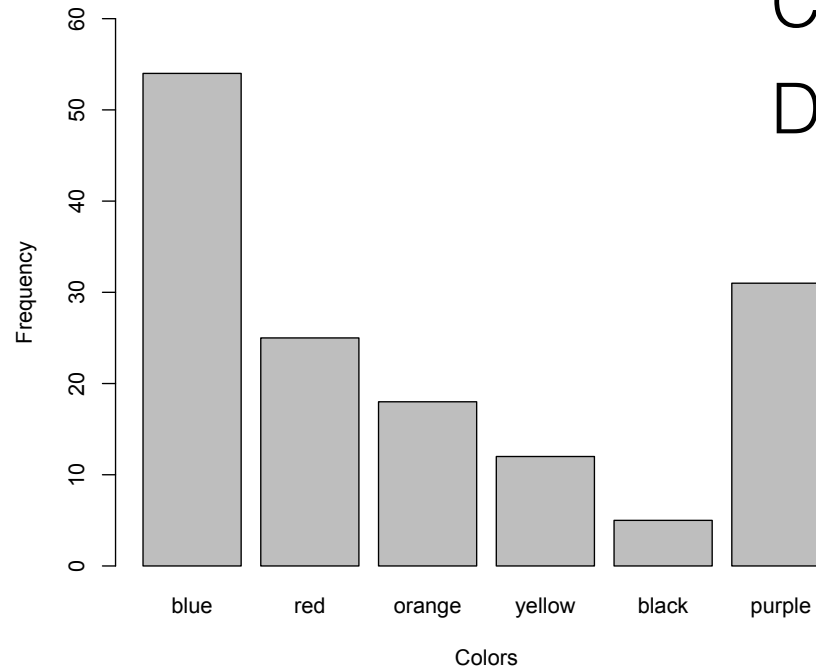
# Bar Charts

Clicker!

- Which color is the most popular favorite color for the surveyed group of people?

Answer:

- A. Red
- B. Orange
- C. Blue
- D. Purple



145 people were asked to state their favorite color.

# Bar Charts

- Which color is the most popular favorite color for the surveyed group of people?

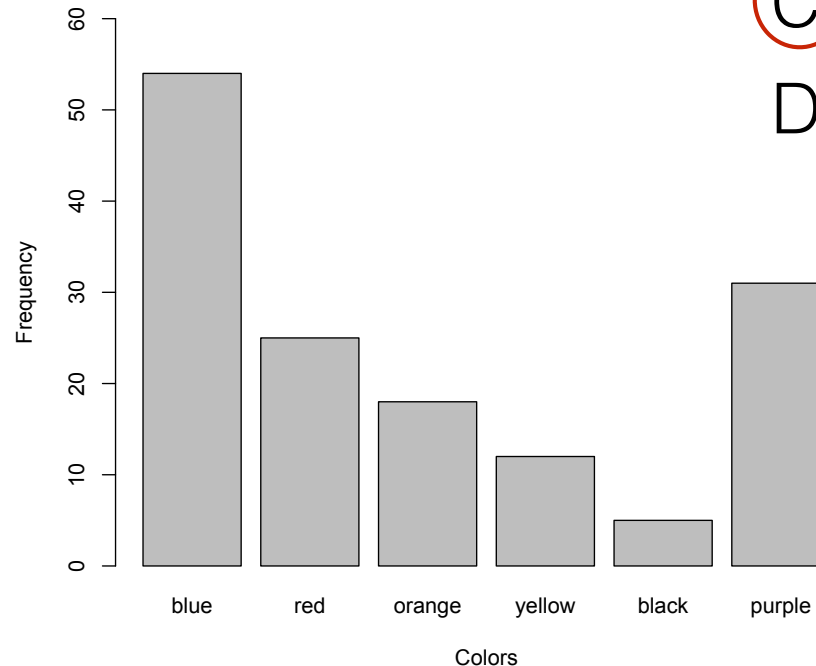
Answer:

A. Red

B. Orange

☒ C. Blue

D. Purple



145 people were asked to state their favorite color.

# Bar Charts

Clicker!

- Which color is the least popular favorite color for the surveyed group of people?

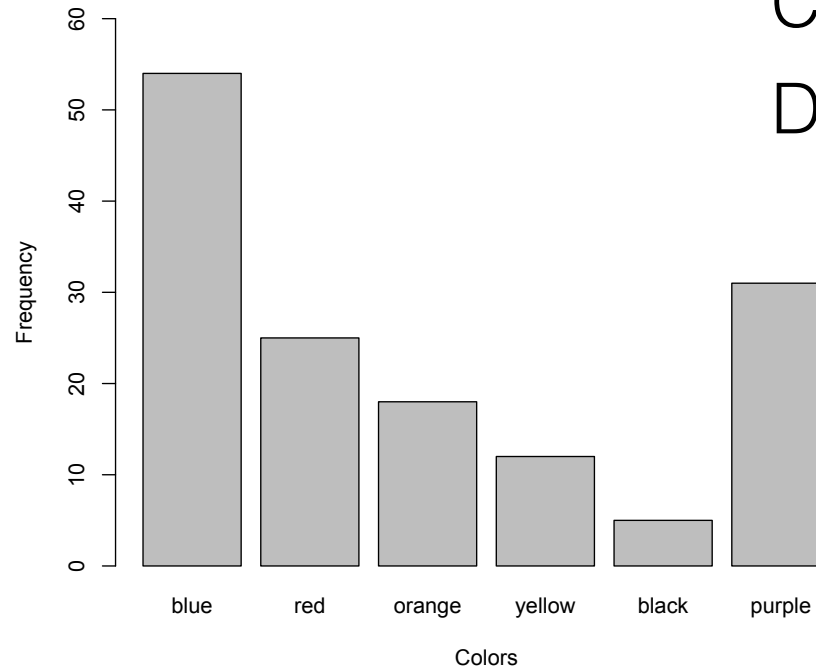
Answer:

A. Red

B. Black

C. Blue

D. Yellow



145 people were asked to state their favorite color.

# Bar Charts

- Which color is the least popular favorite color for the surveyed group of people?

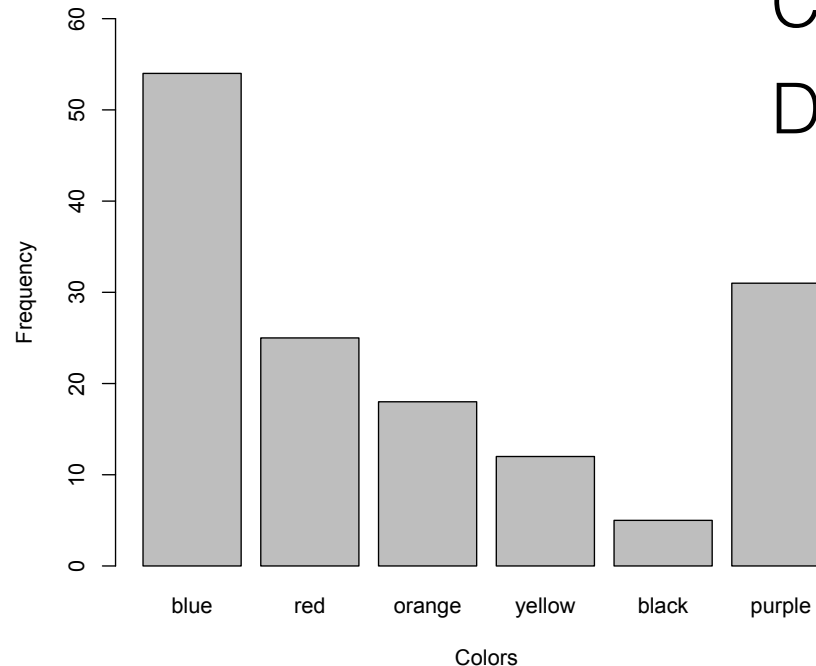
Answer:

A. Red

☒ B. Black

C. Blue

D. Yellow



145 people were asked to state their favorite color.

# Bar Charts

Clicker!

- About how many people said their favorite color was red?

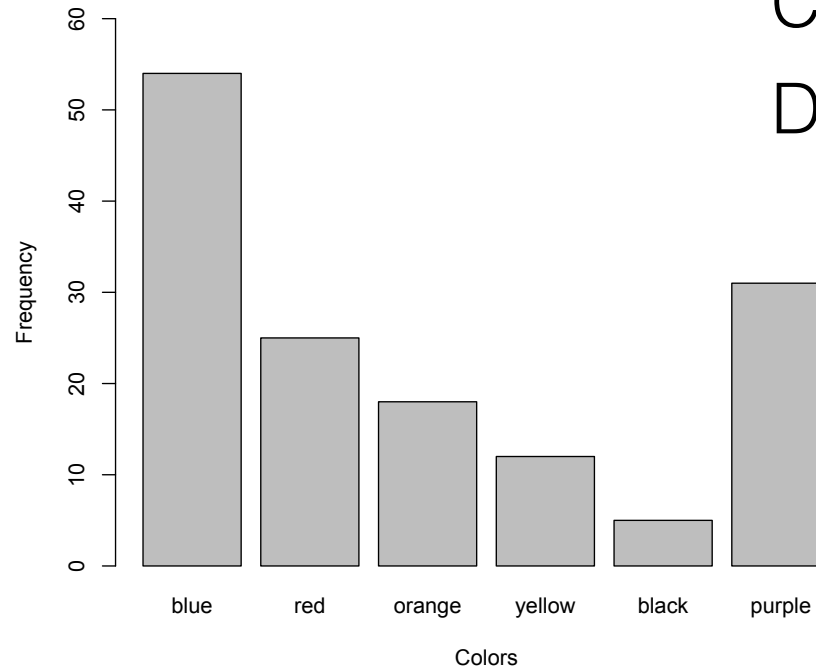
Answer:

A. 20

B. 25

C. 30

D. 40



145 people were asked to state their favorite color.



# Bar Charts

- About how many people said their favorite color was red?

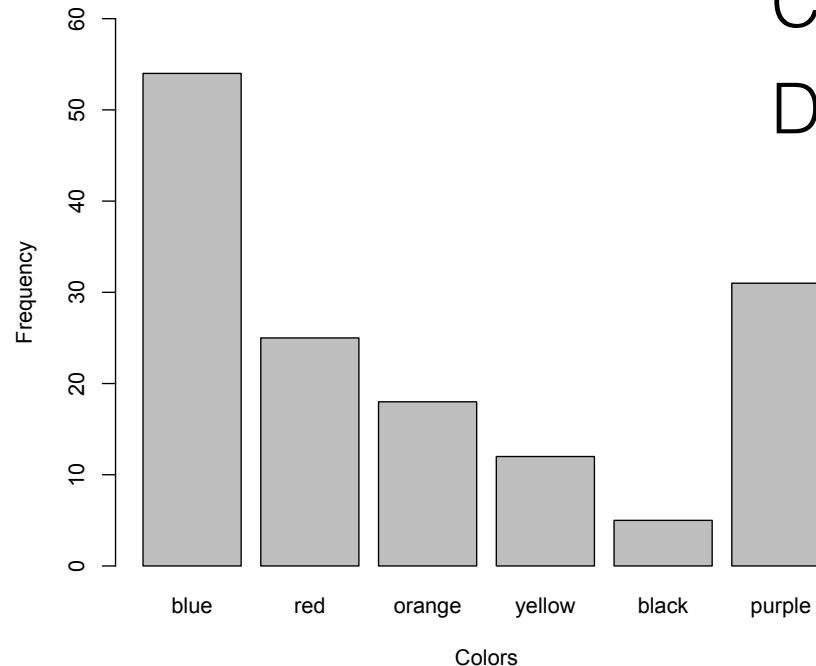
Answer:

A. 20

☒ B. 25

C. 30

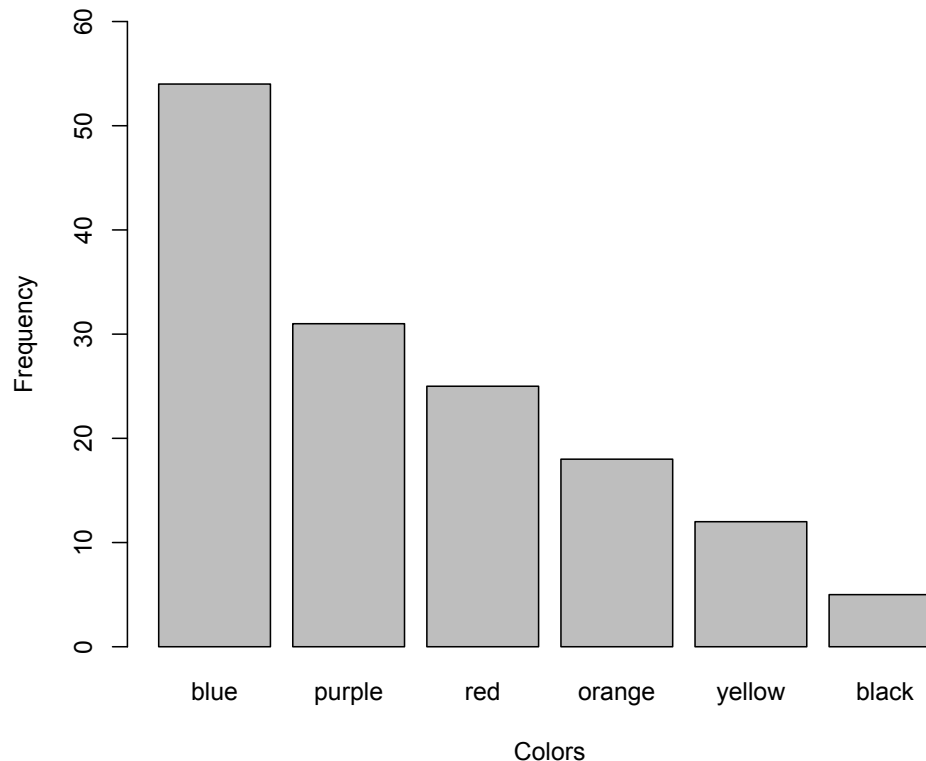
D. 40



145 people were asked to state their favorite color.

# Pareto Charts

- **Pareto chart** is a bar chart that orders the categories from largest to smallest frequency

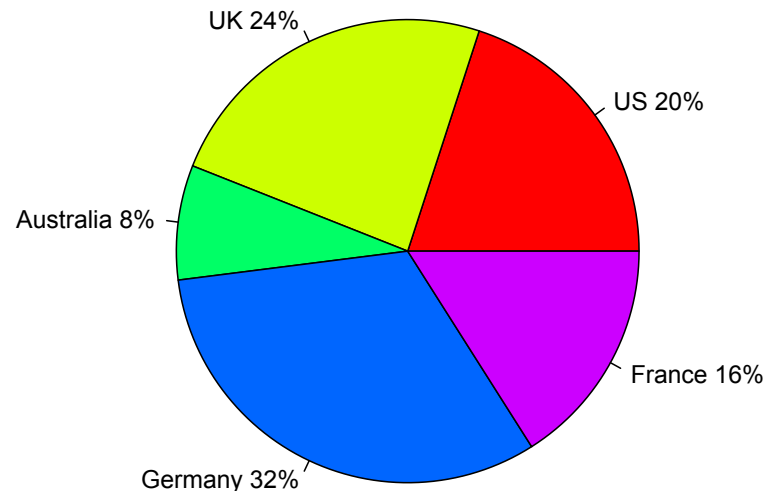


145 people were asked to state their favorite color.

# Pie Charts

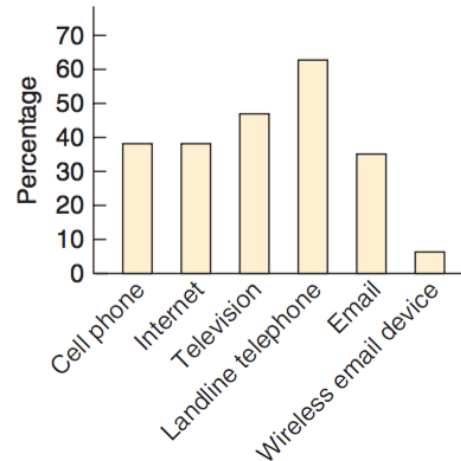
- **Pie charts** display data counts as percentages of subjects in each category.
- Each piece represents a category of the variable.
- The area of the piece is proportional to the relative frequency of that category.
- Even though pie charts are popular, they are not preferred amongst Statisticians. One reason is that it is difficult to visually compare the categories.

**Pie Chart of Countries**

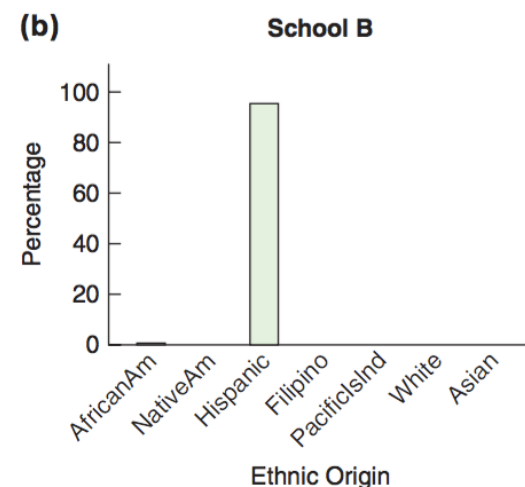
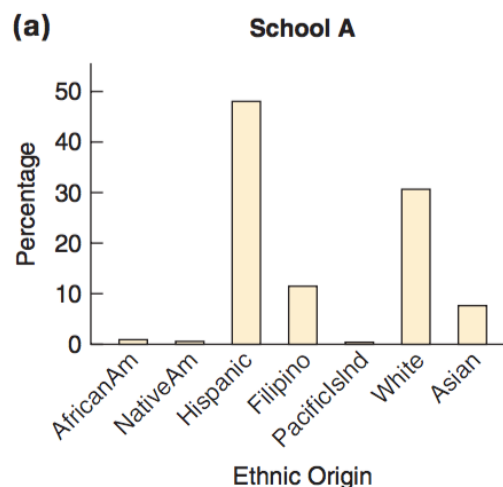


# Descriptions of Categorical Distributions

- **Mode** is the category that occurs the most or the category with the highest frequency.



- **Variability** refers to the distribution of counts amongst the different categories. The more categories there are the more variability there is.



# Stem-and-leaf Plots

- **Stem-and-leaf plots**, or **stemplots**, divide each observation into a “stem” and a “leaf”.
- For the number 50, the 5 is the stem and the 0 is the leaf.
- For the number 573, the 57 is the stem and the 3 is the leaf.

Example:

Exam scores (out of 50) from midterm 1 are:

34, 45, 48, 41, 32, 28, 19, 25, 39, 40, 34

1		9
2		58
3		2449
4		0158

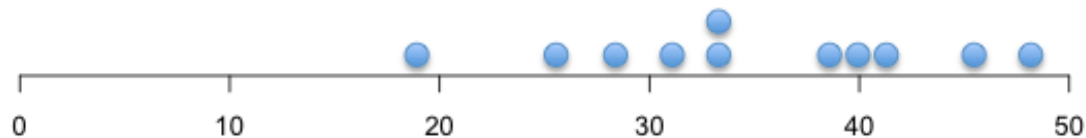
# Dot Plots

- In a **dot plot**, a dot is placed along the axis for every observation for that respective.

Examples:

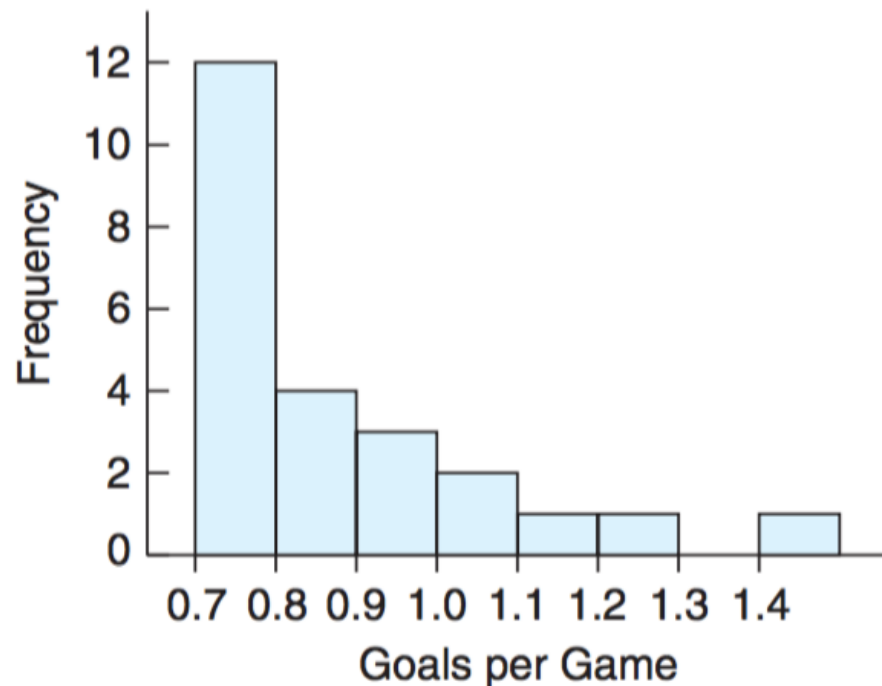
Exam scores (out of 50) from midterm 1 are:

34, 45, 48, 41, 32, 28, 19, 25, 39, 40, 34



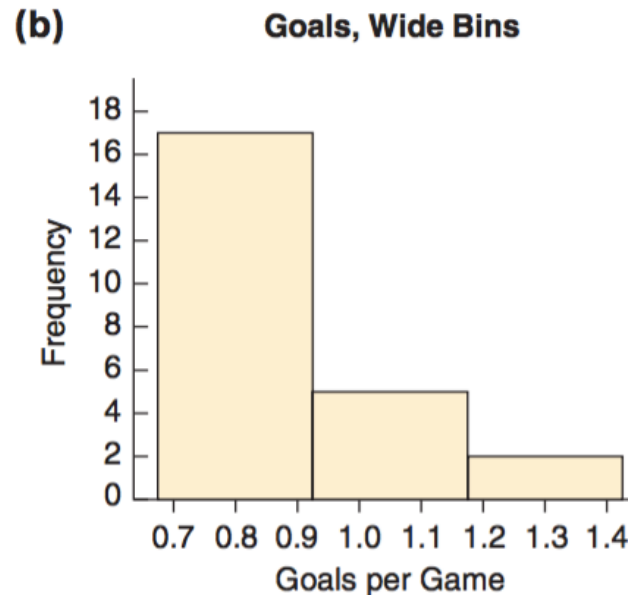
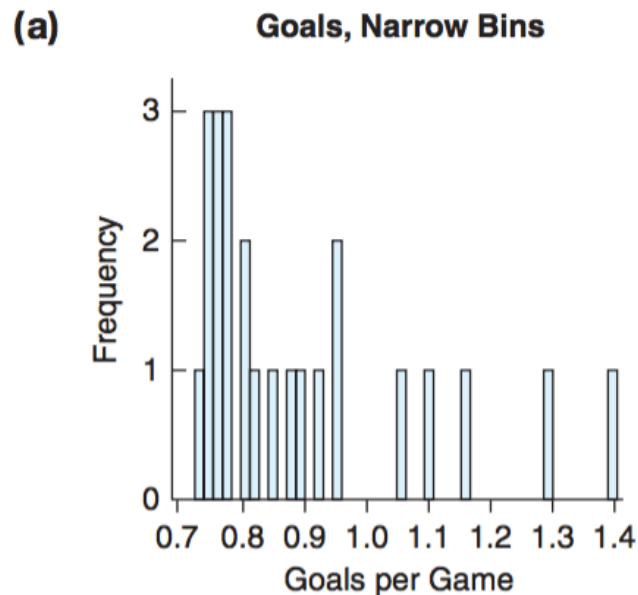
# Histograms

- Similar to dot plots and stem-and-leaf plots, a **histogram** is a way to view a numerical variable.
- The x-axis (horizontal) represents the numerical variable while the y-axis (vertical) represents the frequencies of observations.
- A histogram displays the number of cases in each bin.



# Changing Bin Widths of Histograms

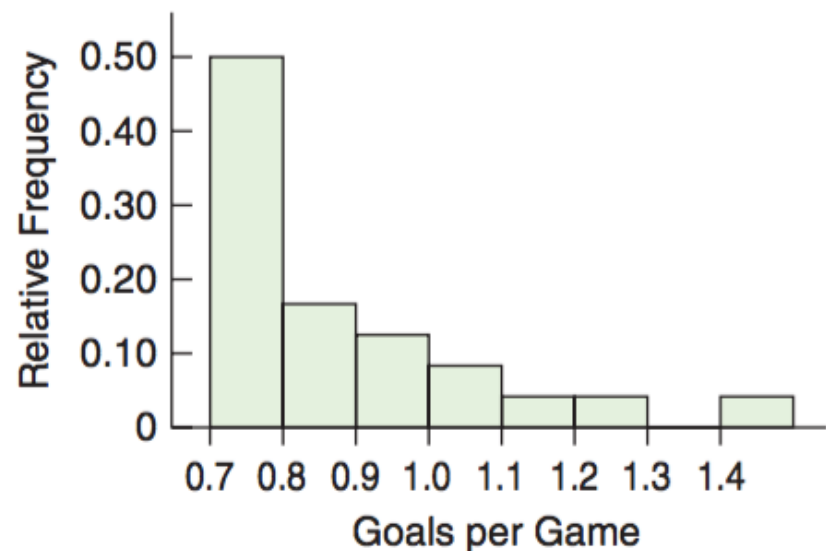
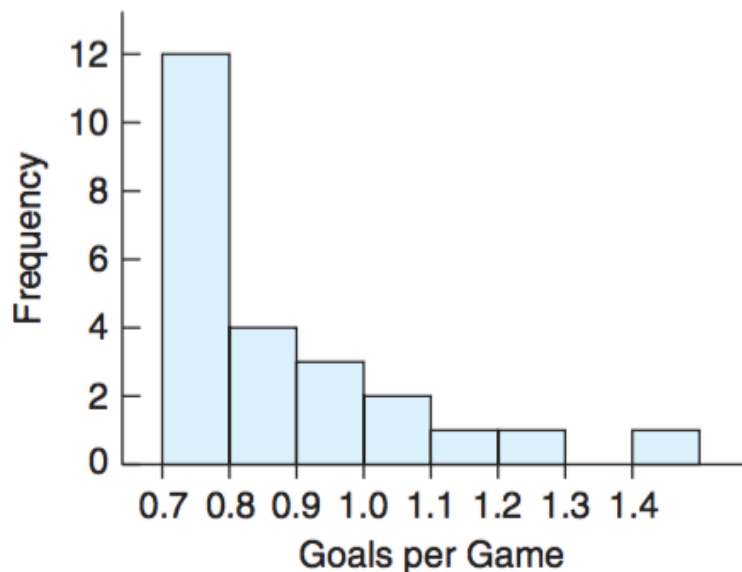
- Bin widths are a personal choice.
- Too small a width shows too much detail while too large a width shows too little.
- Histogram (a) has smaller bins whereas Histogram (b) has wider bins.





# Relative Frequency Histograms

- A **relative frequency histogram** is a histogram where, similar to the relative frequency bar chart, the y-axis represents the relative proportions of each value.
- The value for the y-axis is chosen by dividing frequency by the total observations.
- Relative frequencies does not change the shape of the graph.
- This method is useful when comparing two samples with very different frequencies.



# Differences Between Bar Charts and Histograms

- A histogram displays numerical data. A bar chart displays categorical data.
- The bar widths of a histogram are meaningful and must all be the same size. The bar widths for a bar chart are meaningless.
- The bars of a histogram must touch each other. For a bar chart, there are gaps between bars.
- There is only one choice, ascending by  $x$ , for the order of the histogram, while there are many choices of order for a bar chart.

# Describing Numerical Distributions

- When describing distributions we talk about 3 characteristics of the distribution:
  - Shape
  - Center
  - Variability (Spread)

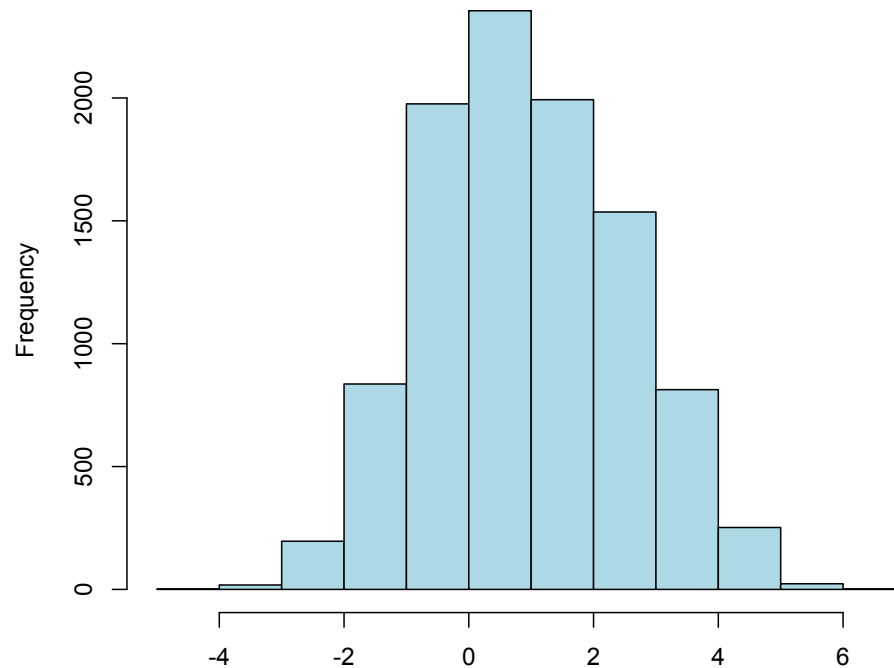
# Shape

- How many peaks does the histogram have?
  - A single peak is a unimodal distribution.
  - Two peaks is a bimodal distribution (more than two peaks is multimodal)
  - No apparent peaks is a uniform distribution

# Shape

Clicker!

- How many peaks does the histogram have?



Answer:

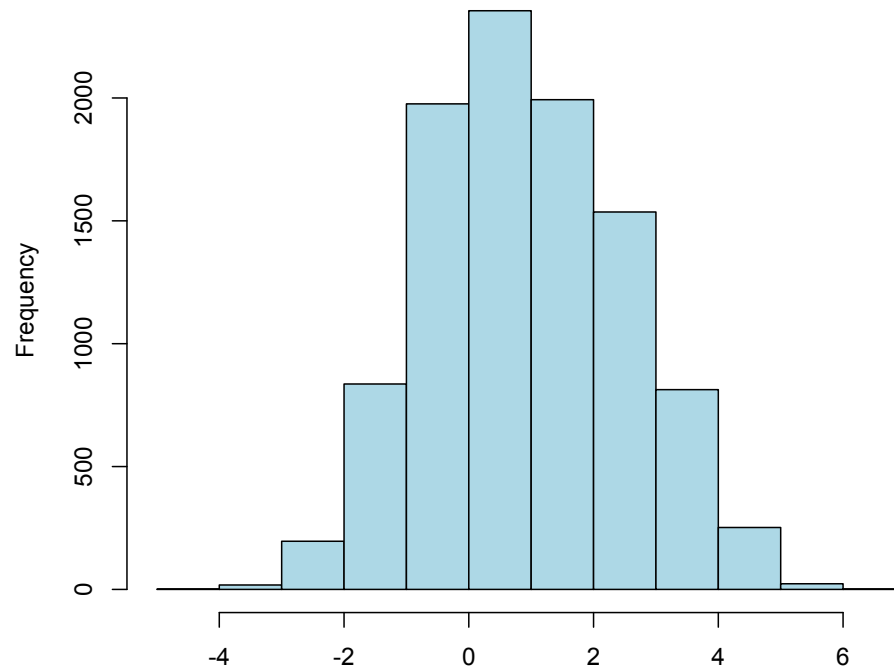
A. 0

B. 1

C. 2

# Shape

- How many peaks does the histogram have?



Answer:

A. 0

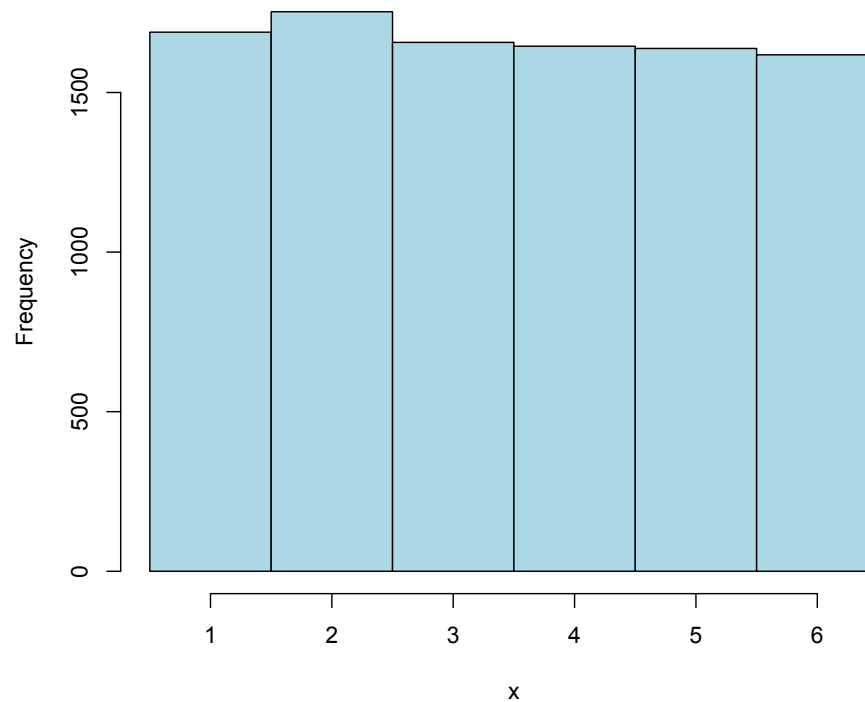
**B. 1**

C. 2

# Shape

Clicker!

- How many peaks does the histogram have?



Answer:

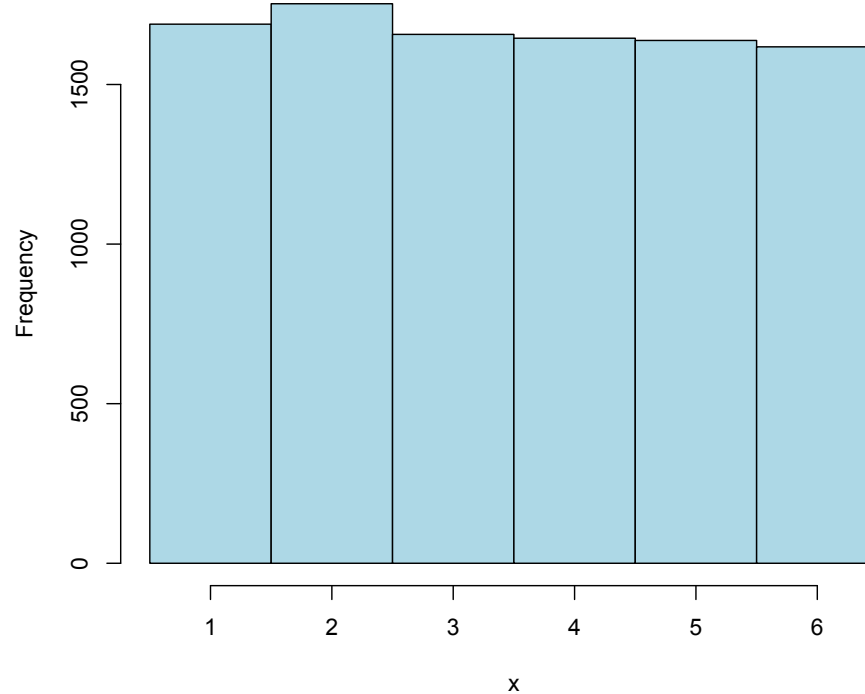
A. 0

B. 1

C. 2

# Shape

- How many peaks does the histogram have?



Answer:

A. 0

B. 1

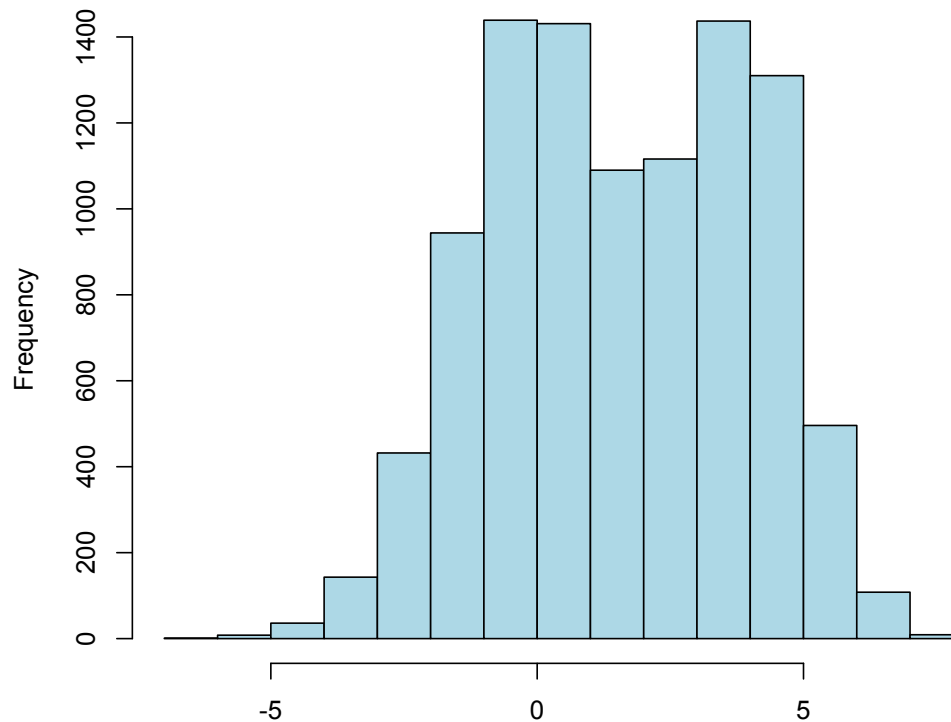
C. 2



# Shape

Clicker!

- How many peaks does the histogram have?



Answer:

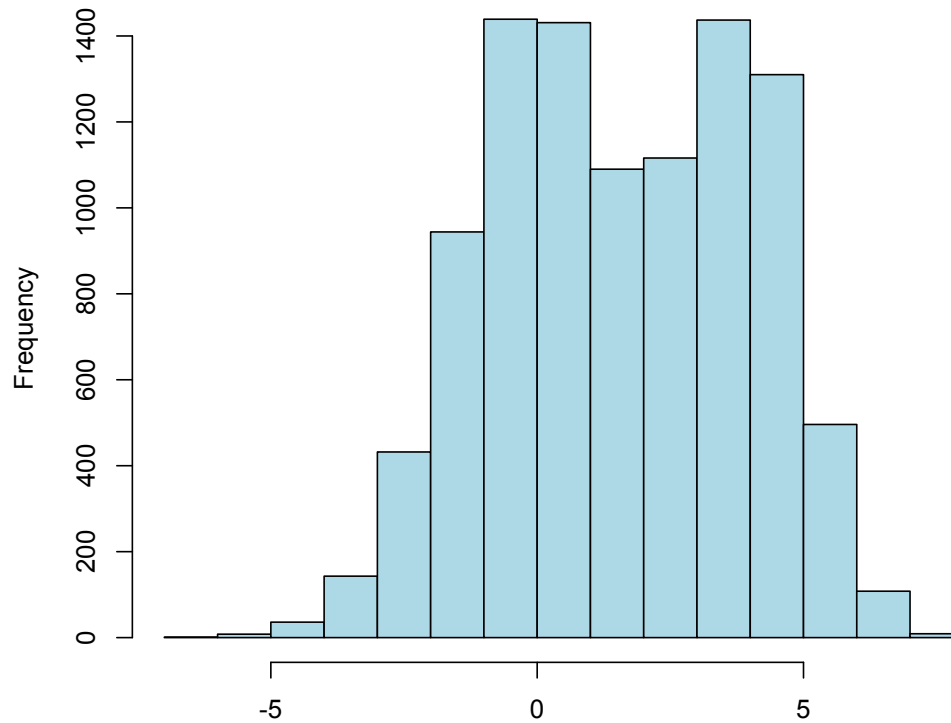
A. 0

B. 1

C. 2

# Shape

- How many peaks does the histogram have?



Answer:

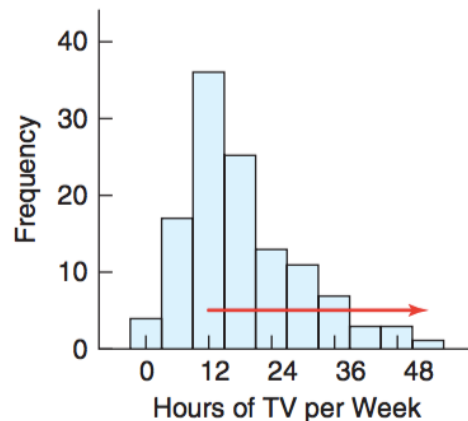
A. 0

B. 1

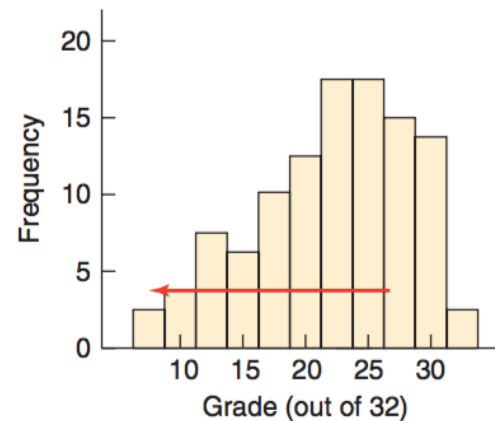
C. 2

# Shape

- Is the histogram symmetric or skewed?
  - If the histogram has a long tail it is skewed to whichever direction the tail occurs.



▲ **FIGURE 2.12** This data set on TV hours viewed per week is skewed to the right (Source: Minitab Program)



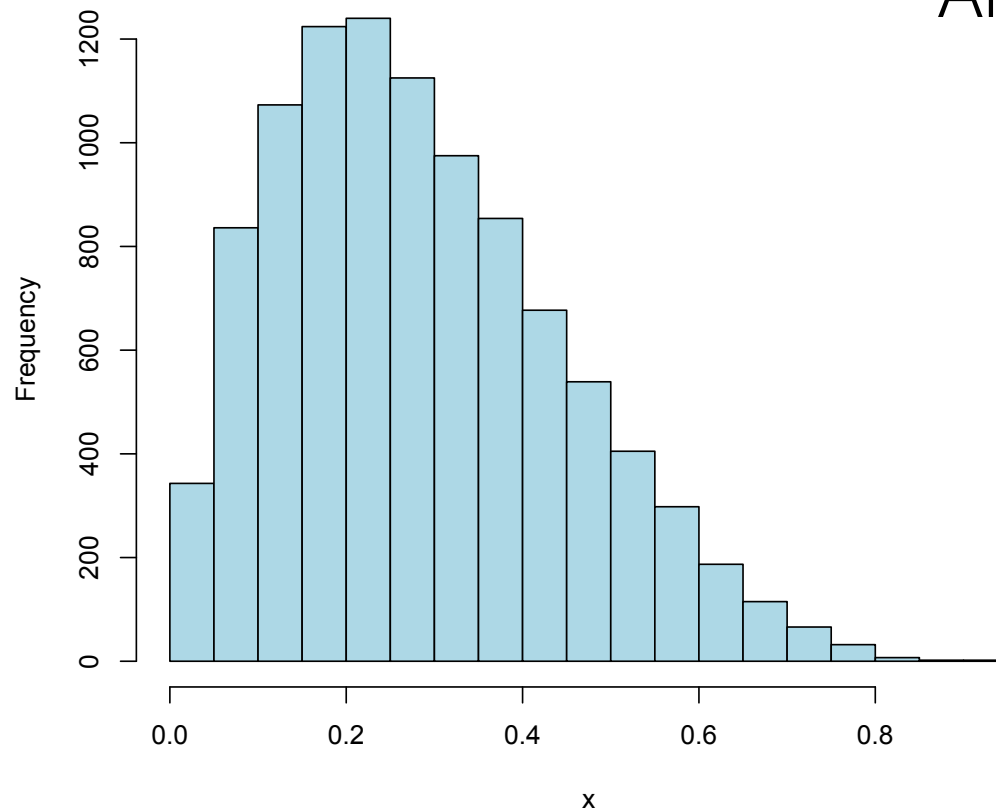
▲ **FIGURE 2.13** This data set on test scores is skewed to the left. (Source: Ryan 2006)

- The histogram is symmetric if the left hand side is roughly the mirror image of the right hand side.

# Shape

Clicker!

- Is the histogram symmetric or skewed?



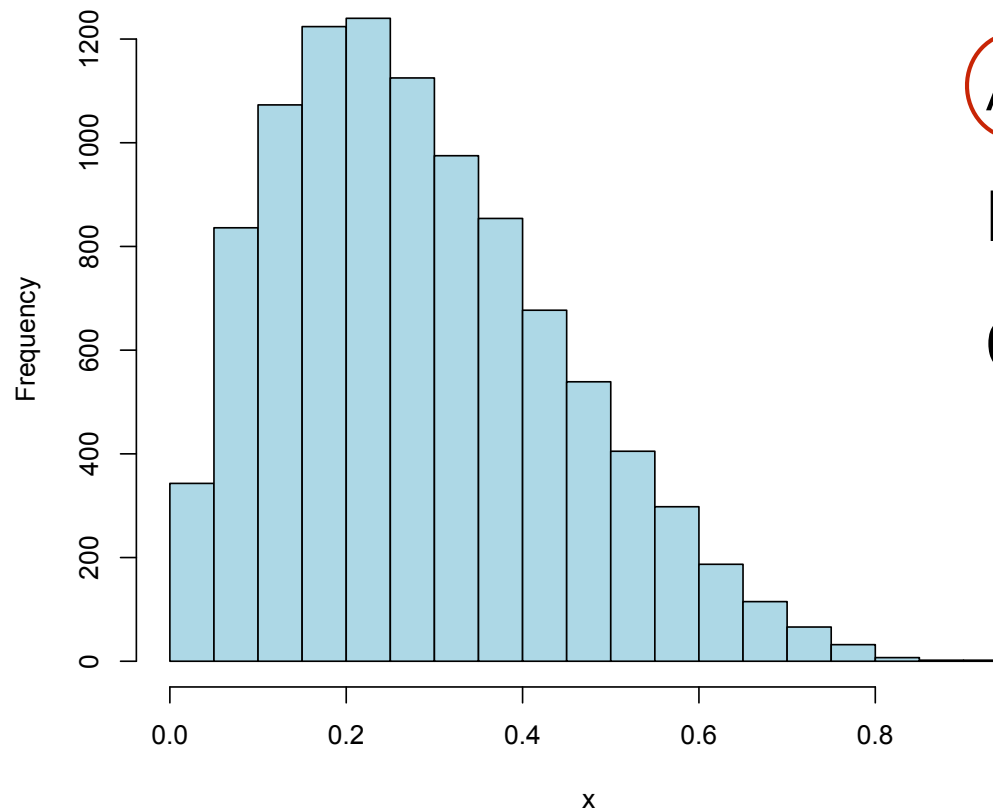
Answer:

- A. Right Skewed
- B. Left Skewed
- C. Symmetric

# Shape

- Is the histogram symmetric or skewed?

Answer:



- ☒ A. Right Skewed
- ☐ B. Left Skewed
- ☐ C. Symmetric

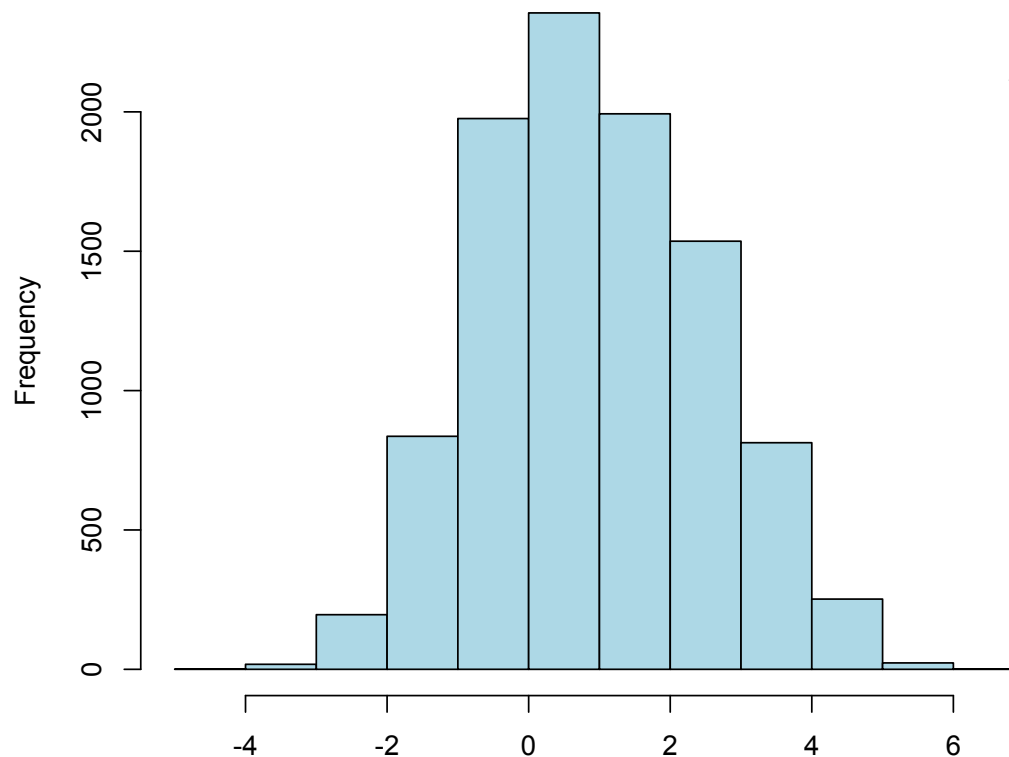
# Shape

Clicker!

- Is the histogram symmetric or skewed?

Answer:

- A. Right Skewed
- B. Left Skewed
- C. Symmetric



# Shape

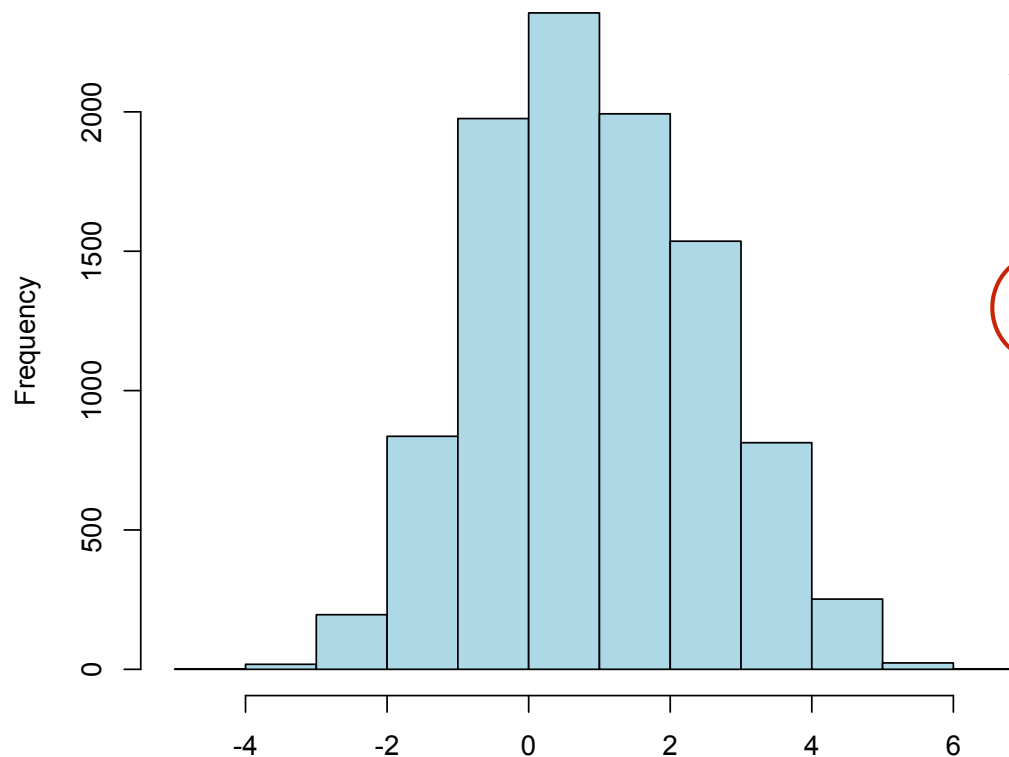
- Is the histogram symmetric or skewed?

Answer:

A. Right Skewed

B. Left Skewed

☒ C. Symmetric



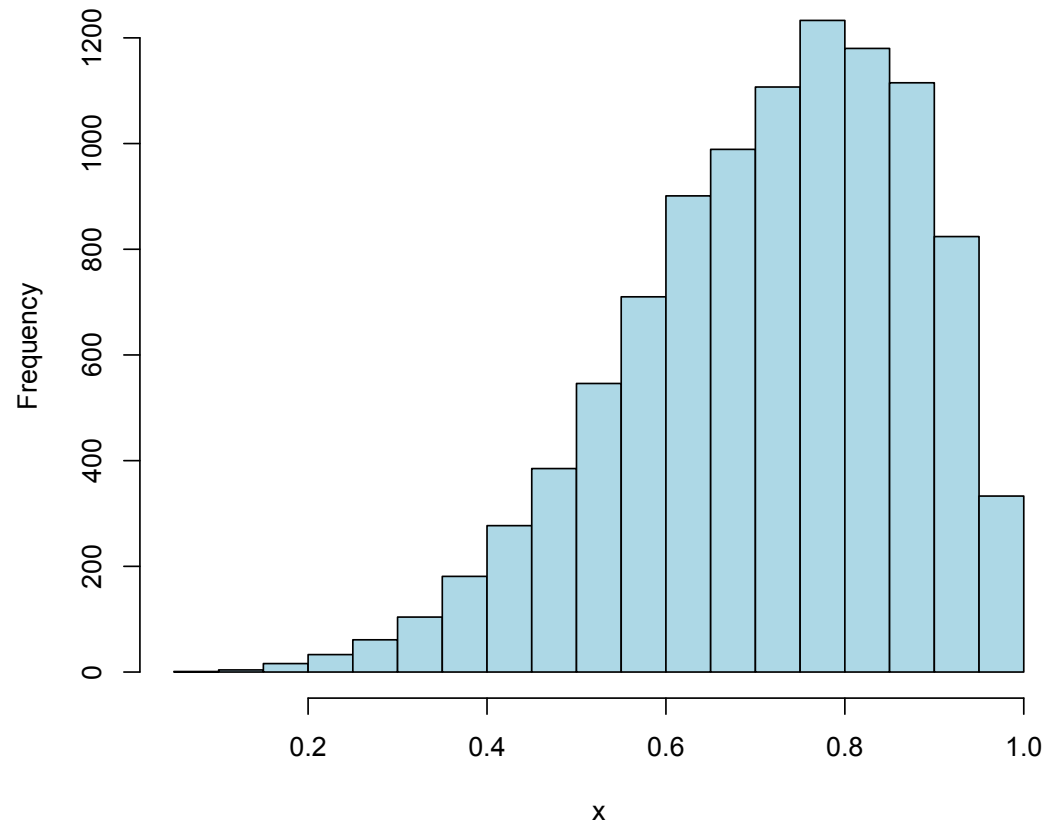
# Shape

Clicker!

- Is the histogram symmetric or skewed?

Answer:

- A. Right Skewed
- B. Left Skewed
- C. Symmetric

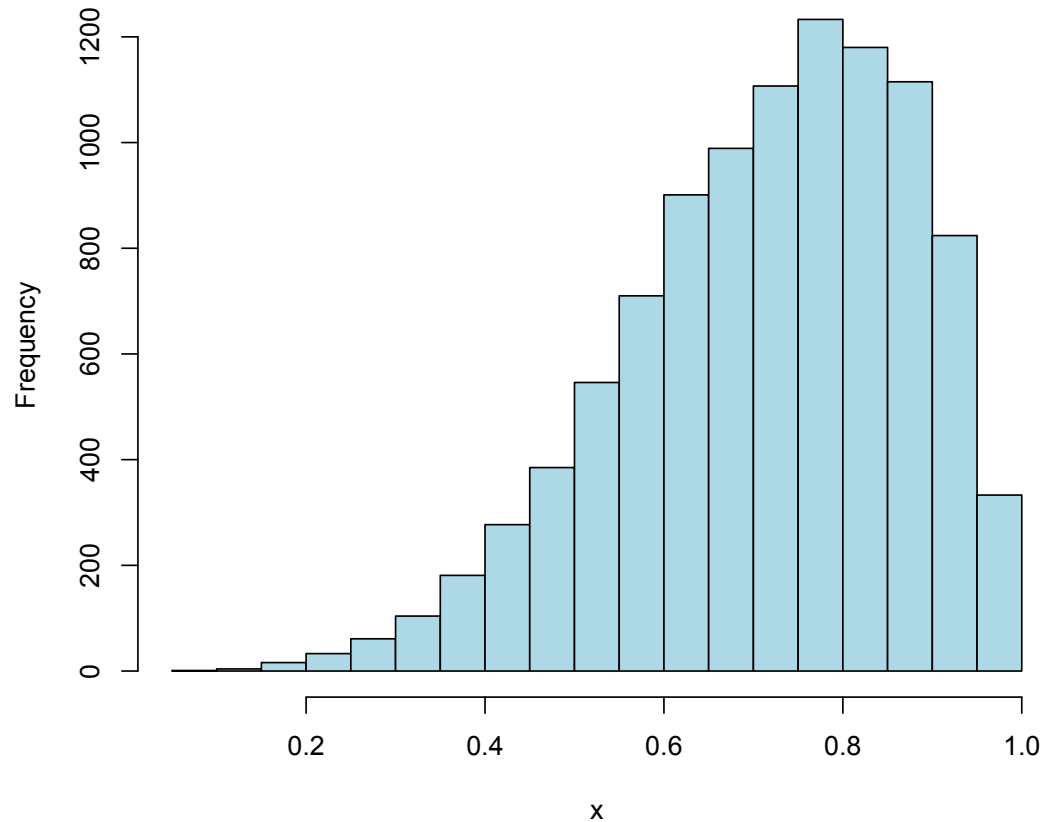




# Shape

- Is the histogram symmetric or skewed?

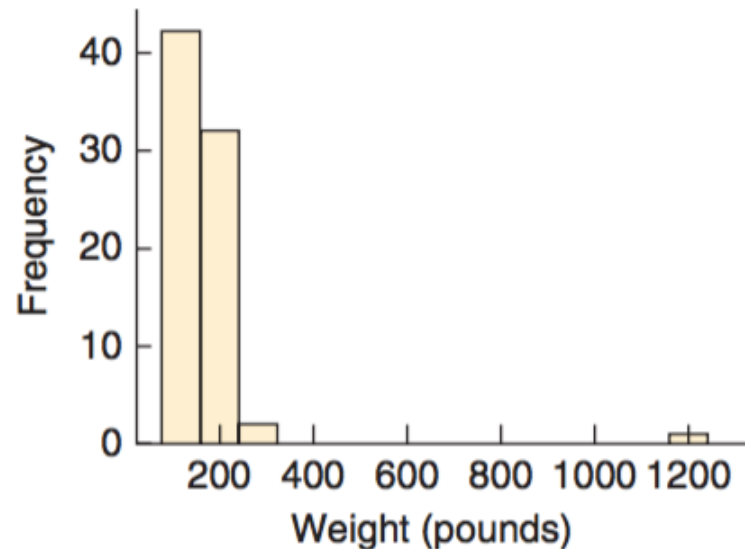
Answer:



- A. Right Skewed
- ☒ B. Left Skewed
- C. Symmetric

# Shape

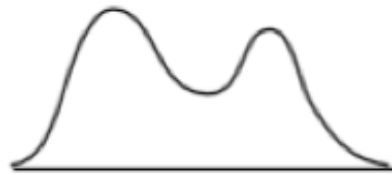
- Are there any unusually large or small values present in the histogram?
  - Extreme values are called **outliers**.
  - Sometimes outliers occur due to mistakes and sometimes they are just unusually interesting observation.



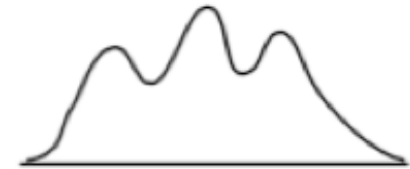
# Shapes of Distributions



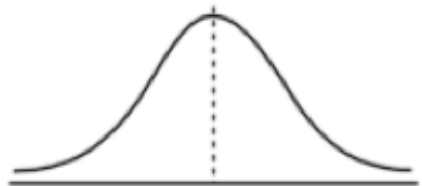
(a) Unimodal



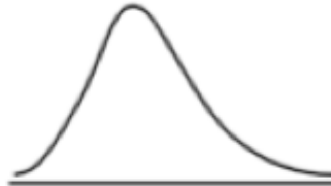
(b) Bimodal



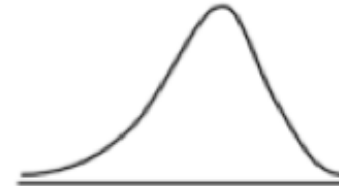
(c) Trimodal



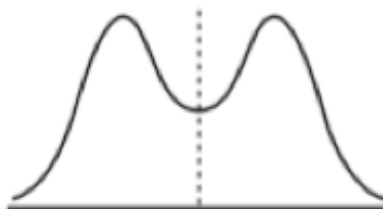
(d) Symmetric



(e) Positively skewed  
(long upper tail)



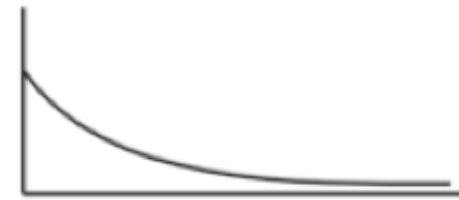
(f) Negatively skewed  
(long lower tail)



(g) Symmetric



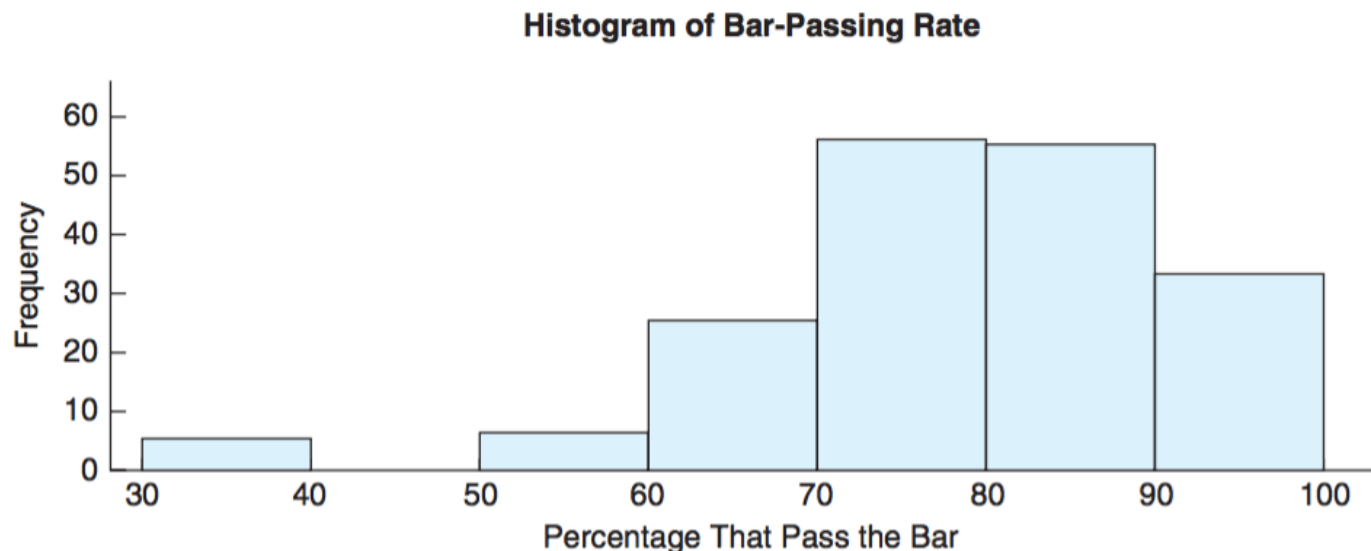
(h) Bimodal with gap



(i) Exponential shape

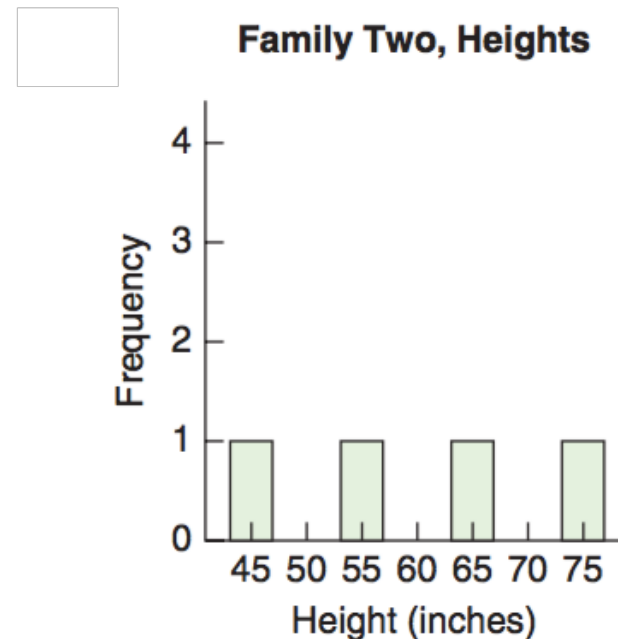
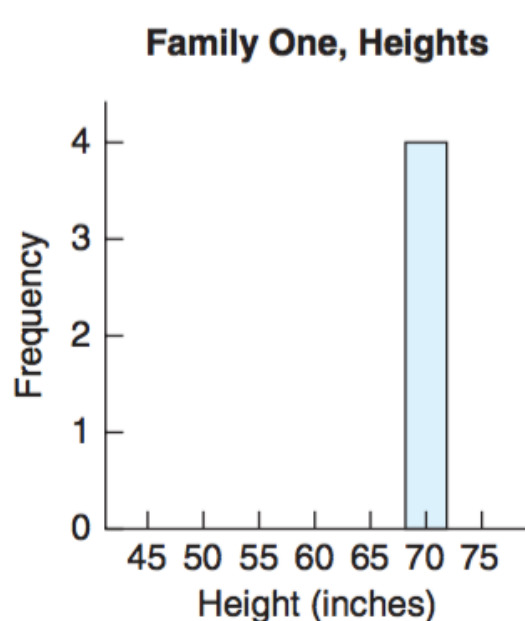
# Center

- What is the typical value?
  - The typical value is the one in the center.
  - It might not make sense to find the center for bimodal or multimodal distributions.
  - If there are two distinct groups in the data then it might make sense to find the typical value for each group.



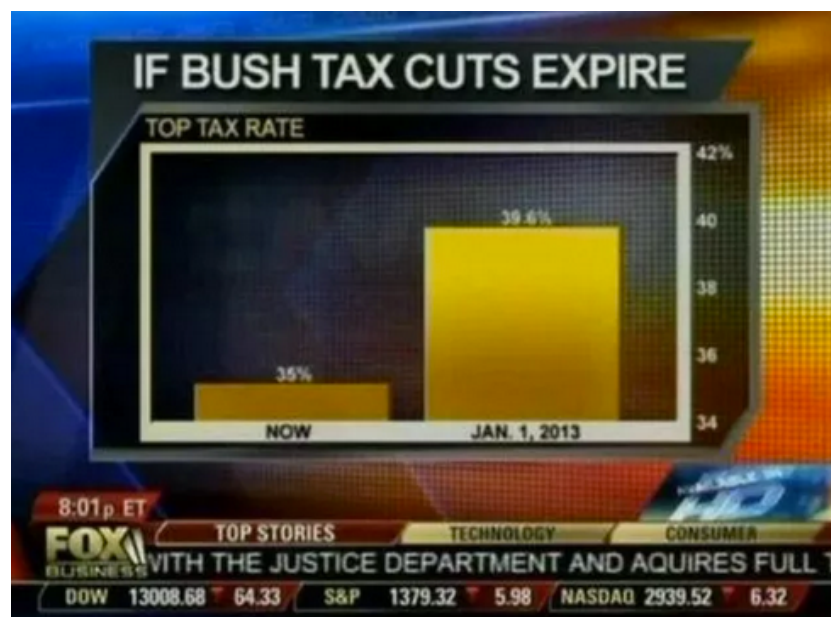
# Variability (Spread)

- Variability or spread describes how spread out the data is from the center.
- When the data values are tightly clustered around the center of the distribution, the spread is small, hence there is less variability.
- When the data values are scattered far from the center, the spread is large, hence there is more variability.



# Misleading Graphs

- Have the frequency scale not begin at 0 to create the illusion of greater differences
- Using symbols instead of bars of equal width can give an incorrect representation of the data
- Lack of labels can be misleading



# Example

The distribution of in-state annual tuition for all colleges and universities in the United States is bimodal. What is one possible reason for this bimodality?

# Example

The distribution of in-state annual tuition for all colleges and universities in the United States is bimodal. What is one possible reason for this bimodality?

Answer: Private colleges and public colleges tend to differ in amount of tuition.



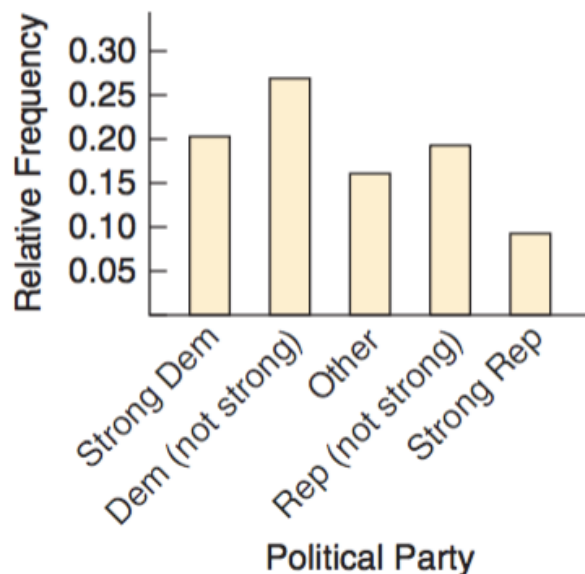
# Example

Clicker!

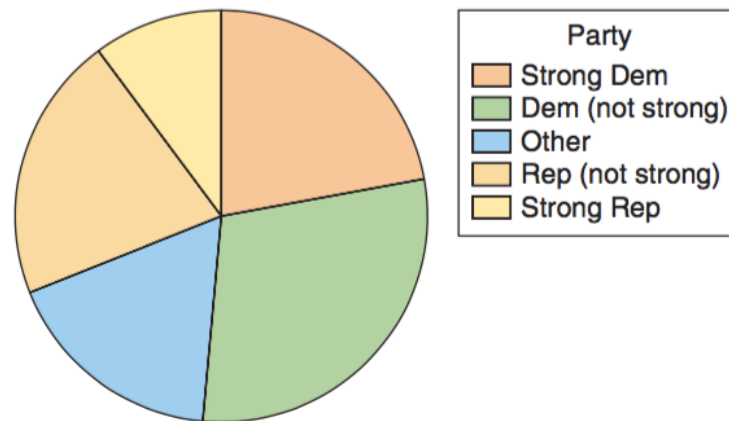
The 2008 GSS asked its participants to report their political party affiliation. The graphs show the distribution of political party affiliation for 1084 women.

- Which political affiliation has the most women?
  - A. Dem (not strong)
  - B. Strong Dem
  - C. Rep (not strong)
  - D. Other

**Bar Chart of Political Party for Women**



**Pie Chart of Political Party for Women**



# Example

The 2008 GSS asked its participants to report their political party affiliation. The graphs show the distribution of political party affiliation for 1084 women.

- Which political affiliation has the most women?

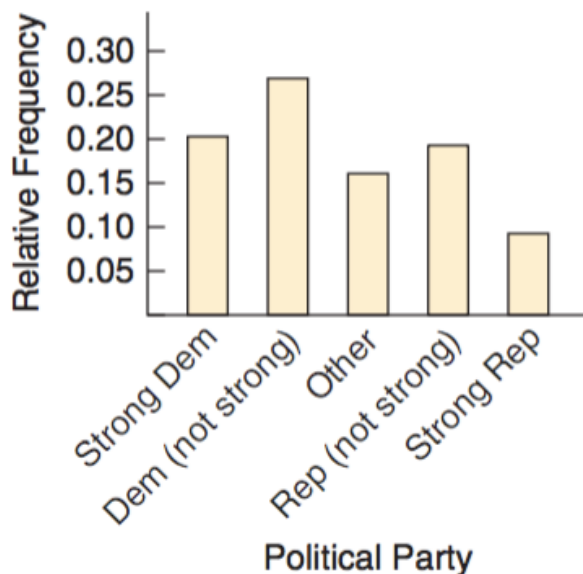
A. Dem (not strong)

B. Strong Dem

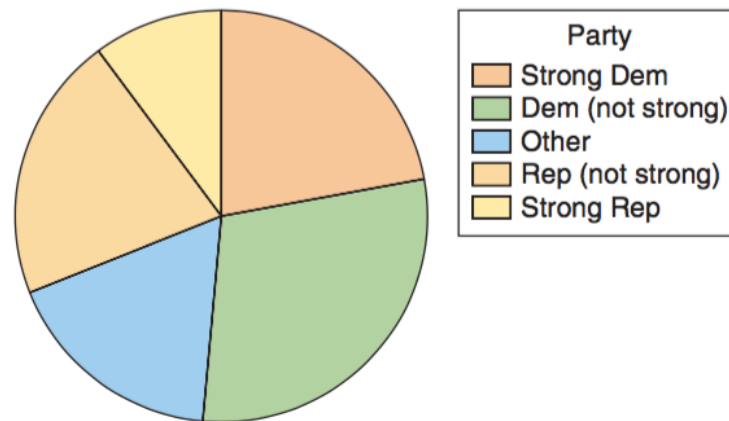
C. Rep (not strong)

D. Other

**Bar Chart of Political Party for Women**



**Pie Chart of Political Party for Women**



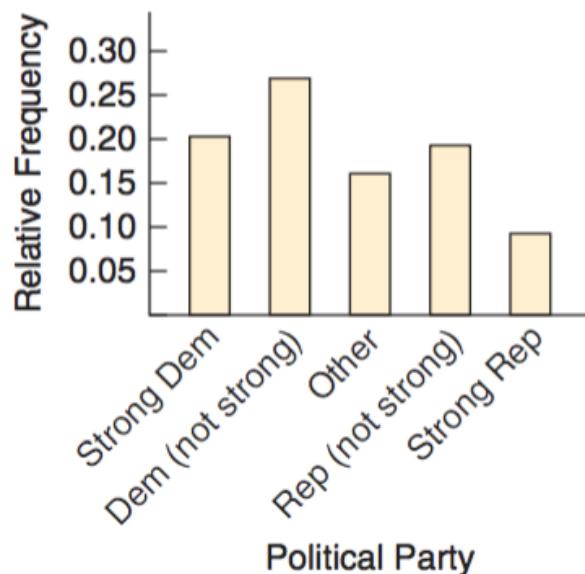
# Example

Clicker!

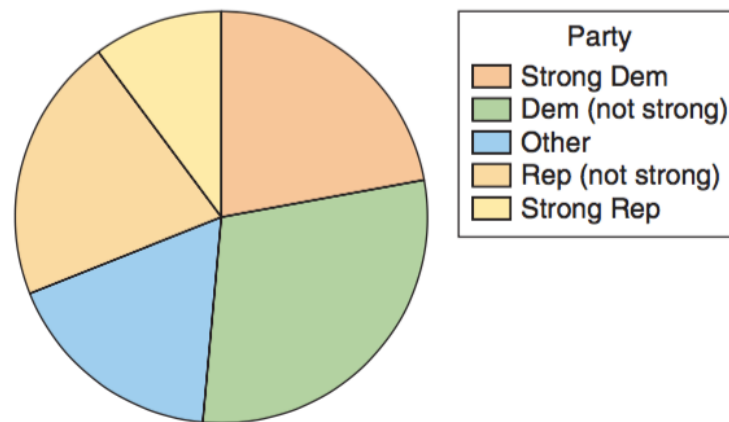
The 2008 GSS asked its participants to report their political party affiliation. The graphs show the distribution of political party affiliation for 1084 women.

- Which political affiliation has the 2nd largest number of women?
  - A. Dem (not strong)
  - B. Strong Dem
  - C. Rep (not strong)
  - D. Strong Rep

**Bar Chart of Political Party for Women**



**Pie Chart of Political Party for Women**



# Example

The 2008 GSS asked its participants to report their political party affiliation. The graphs show the distribution of political party affiliation for 1084 women.

- Which political affiliation has the 2nd largest number of women?

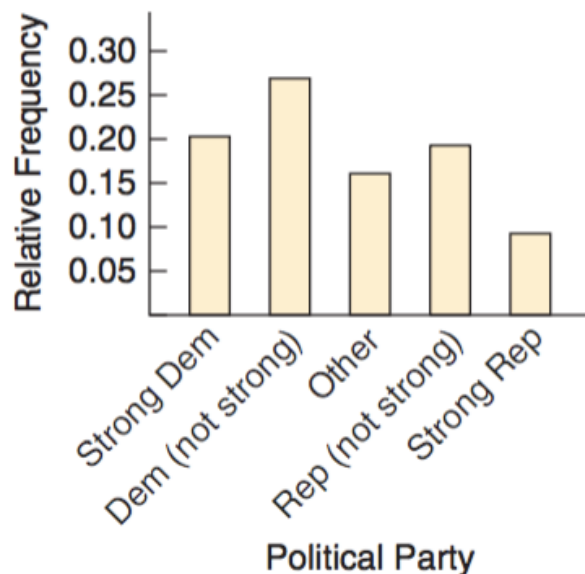
A. Dem (not strong)

☒ B. Strong Dem

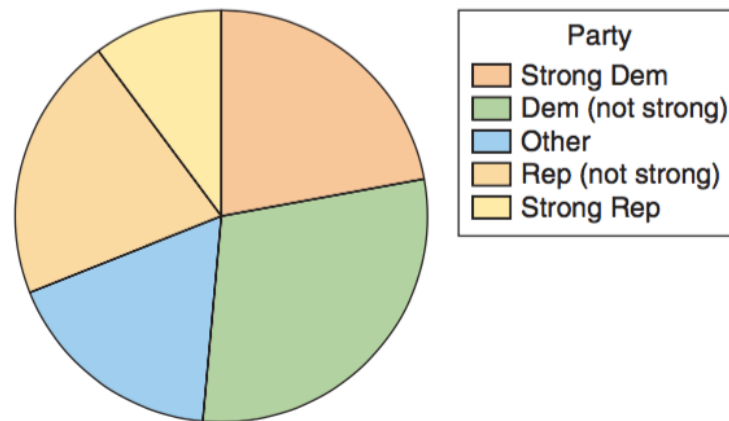
C. Rep (not strong) ← pretty close!

D. Strong Rep

**Bar Chart of Political Party for Women**



**Pie Chart of Political Party for Women**



# Example

Clicker!

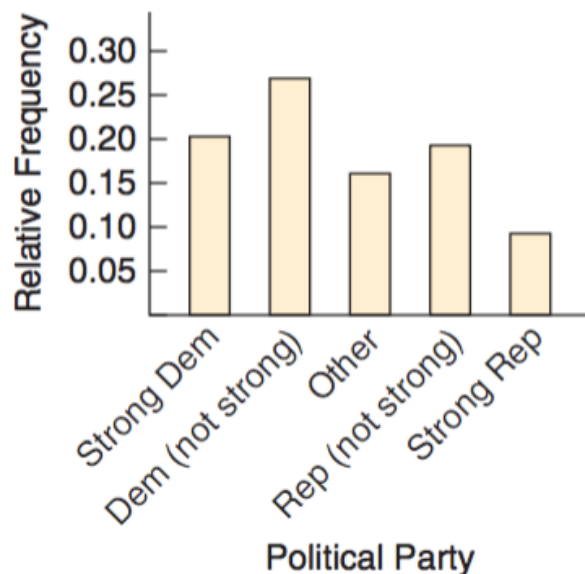
The 2008 GSS asked its participants to report their political party affiliation. The graphs show the distribution of political party affiliation for 1084 women.

- Is this easier to see from the bar chart or pie chart?

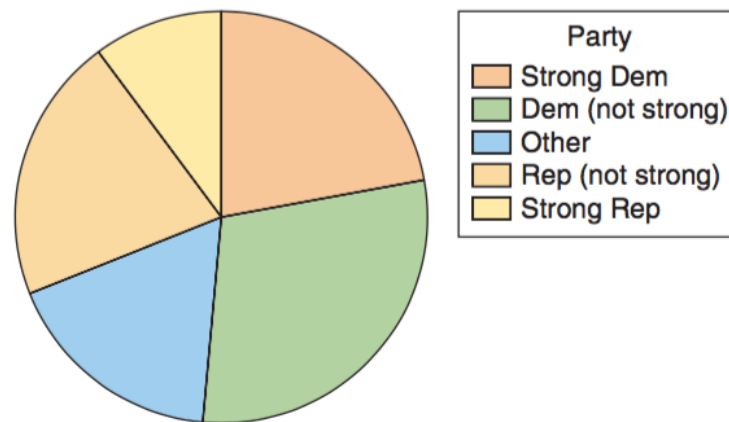
A. Bar chart

B. Pie chart

**Bar Chart of Political Party for Women**



**Pie Chart of Political Party for Women**



# Example

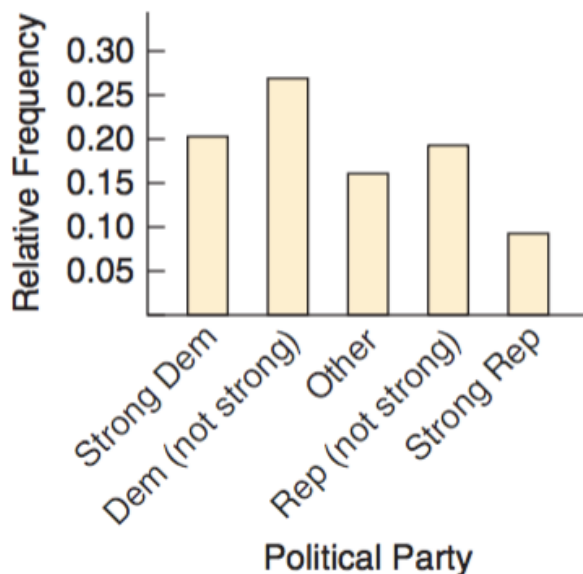
The 2008 GSS asked its participants to report their political party affiliation. The graphs show the distribution of political party affiliation for 1084 women.

- Is this easier to see from the bar chart or pie chart?

A. Bar chart

B. Pie chart

**Bar Chart of Political Party for Women**



**Pie Chart of Political Party for Women**

