# DISPLAYING DATA



## Bar Graphs and histograms:

Bar graphs is a graphical display of data using columns of different heights.

Histograms are similar to bar graphs but the columns are usually next to each other and show the frequency of quantitative variables.

DIFFERENCE BETWEEN BAR GRAPH AND HISTOGRAM

In a bar graph, each column represents a group defined by a categorical variable; labels on the x axis are categorical. In a histogram each column represents a group defined by a continuous quantitative variable; labels on the x axis are quantitative.

### BAR GRAPH

Data is collected on favorite types of books. Create a bar graph using the information below.

| Book Type | |
|---|---|
| Historical Fiction | 10 |
| Fiction | 13 |
| Fantasy | 16 |
| Nonfiction | 8 |

### HISTOGRAM

Data is collected on college tuition costs at a sample of colleges. Create a histogram using the information below.

COLLEGE TUITION: $,4661, $9,785, $10,699, $12,125, $12,551, $10,925, $16,402, $15, 350, $13, 530, $26,506, $22,300, $31,500, $46,200, $37,500

| College Tuition | # of colleges |
|---|---|
| $0-$5,000 | |
| $5,001-$15,000 | |
| $15,001-$25,000 | |
| $25,001+ | |

# DISPLAYING CATEGORICAL DATA



- Collect data. Keep track of gender
.
- Make a bar graph for the % of men and % of women saying "there are about the right number of women in high political offices in the country today."

*% saying there are ____ in high political offices in the country today*

| | All adults | Men | Women | Rep/Lean Rep | Dem/Lean Dem |
|---|---|---|---|---|---|
| **Too many women** | 6 | 8 | 4 | 9 | 3 |
| **Too few women** | 59 | 48 | 69 | 33 | 79 |
| Ideally there would be ... | | | | | |
| *More women than now, but still not as many women as men* | 5 | 3 | 6 | 5 | 4 |
| *Equal number of women and men* | 48 | 39 | 56 | 27 | 64 |
| *More women than men* | 6 | 5 | 7 | 2 | 11 |
| **About the right number of women** | 34 | 43 | 26 | 57 | 17 |

Note: Share of respondents who didn't offer an answer not shown. See findings for business at end of section.
Source: Survey of U.S. adults conducted June 19-July 2, 2018.
"Women and Leadership 2018"

- Make a pie chart of all respondents.

# DISPLAYING CATEGORICAL DATA



Is a pie chart appropriate for this data? Explain.

Think of an example where a pie chart would not be a good way to represent a set of data. Describe your example and explain why this would not be a good example.

.

# DISPLAYING DATA



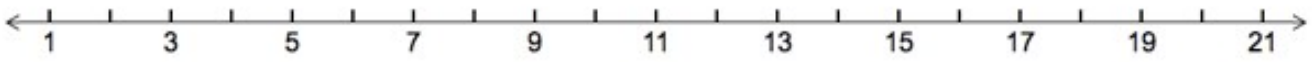
## DOT PLOTS

A dot plot is a simple way to graphically show the frequency of data within categories or groups. Dot plots are not good for large data sets.

Example. Quiz scores

- Label the axis and create a scale.
- Make a dot above each location for the value from the data

Data set: 12, 21, 18, 19, 20, 16, 17, 17, 19, 21, 20, 13,18, 15, 17, 17, 18, 9, 18, 18, 20




The purpose of a graph is to help us understand the data. *What do you see? What is the overall pattern and what data doesn't fit the pattern (outliers)?*

Describe the shape, spread, and center. Identify outliers.

- Shape:

- Spread: the spread shows how much variability in the data

- Center: can calculate the median mathematically or look at where the center seems to be.

- Outliers: outliers are data points that are distant from the other data points.

.

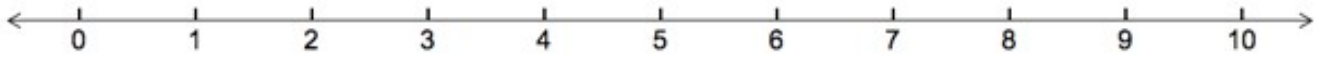
# DISPLAYING DATA



## NOW YOU TRY!

A group of students in a statistics class were surveyed about the number of siblings each student has. The data set shows the results of this survey. Create a dot plot for the data and respond to the analysis questions.

Data set: 0,1,2,0,3,0,1,4,0,5,0,1,8,2,0,1,2,2,1,3,0,4,3,6




Analysis

- Describe the shape:

- Describe the spread:

- Describe the center:

- Identify outliers.

.

# DISPLAYING DATA



Box Plots

**Box plots** (also called box and whisker diagram) is a way of displaying data based on **five key numbers**: minimum, maximum, median, first quartile and third quartile.

To create a box plot:

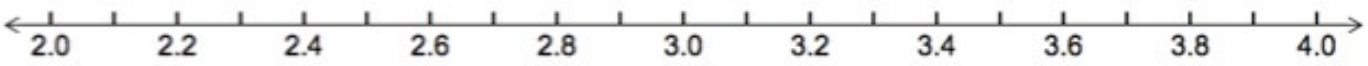
Komodo dragon lengths (in meters):
3.6, 2.7. 2.2, 3.1, 3.5, 3, 2.9, 2.8, 3.3, 3, 2.6, 3.1

- Put the list of numbers in order.

- Divide the numbers into 4 sections (quartiles).

- Find the median (the number in the middle)

- Find the minimum and maximum:
- What is the value of Q1?
- What is the value of Q3?
- What is the Interquartile range (IQR)?
- Use the above to create the box plot:




- Calculate outliers: Q1-1.5(IQR);Q3+1.5(IQR)
  - Q1-1.5(IQR)=
  - Q3+1.5(IQR)=

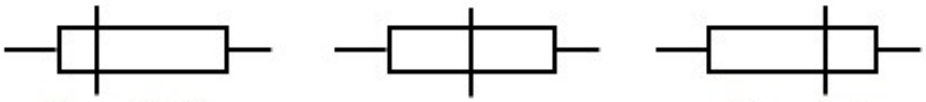
# DISPLAYING DATA



Box Plots: Interpreting

A box plot can provide key information about the distribution of data.

- The median is indicated by the vertical line running down the box.
- Range: the range indicates the distribution of data.
- The interquartile range (Q3-Q1) shows where the data is concentrated.
- Shape of the data set:



*Skewed Right* *Symmetrical* *Skewed Left*

The following are the scores of members of a golf team in tournament play: 89 90 87 95 80 93 99 86 81 111 108 83 88 91 79

- Construct a boxplot of the data.
- Find the minimum and maximum.
- Find the median.
- Find the range
- Find the interquartile range
- Find the outliers.
    - Q1-1.5
    - Q3+1.5
- Describe the distribution.

# DISPLAYING DATA



Box Plots: Now you try!

**Create a box plot for the data shown at the right:**

Ages of soccer players on an Olympic team:
24, 30, 30, 22, 25, 22, 18, 25, 28, 30, 25, 27, 19, 22, 25

What is the median_____

What is the minimum_____

What is the maximum_____

- What is the range of the data?

- What is Q1? What is Q3?

- Are there outliers? Explain.

- Is the data skewed right, skewed left, or not skewed? What does the distribution of data tell you about the data?

# DISPLAYING DATA



Stem-and Leaf-Plots

A stem and leaf plot shows the frequency of certain values occurring. Stem and leaf plots are used for quantitative data. Similarly, to dot plots or histograms, it is easy to see outliers and find clusters of data.

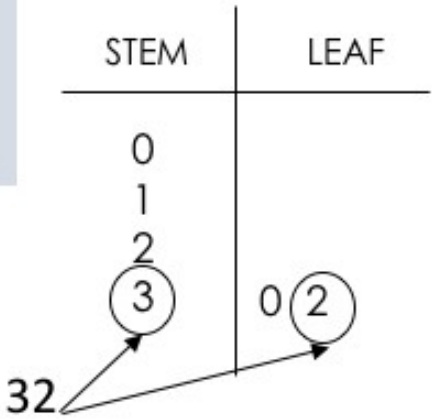
In a stem and leaf plot, the data is split into "stem" (the first digit) and leaf (the remaining digits).

For example: for 32 the stem is 3 and the leaf is 2.

**Step 1**: line up the data in order from least to greatest.

Data 24, 30, 32, 12, 7, 8, 25, 23, 23, 29, 20

**Step 2**: Place the data in to the stem and leaf plot. 32 and 30 have been completed for you.




Information about the data that we can access from stem and leaf plots:

The range (maximum – minimum):

How many entries are in the stem and leaf plot (i.e. what is *n)?*

What is the median (the center)?

What is the interquartile range?
- What is Q1?
- What is Q3?
- Interquartile Range (IQR)=

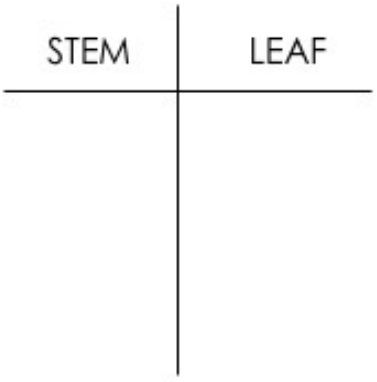
**Try it:** Create a stem and leaf plot given the data and complete the questions.

Data 10,22,31,22,12,14, 7,8,9,13,24,28

The range (maximum – minimum:

.

How many entries are in the stem and leaf plot (i.e. what is *n)?*




What is the median (the center)?

What is the interquartile range?
- What is Q1?
- What is Q3?
- Interquartile Range (IQR)=

# DISPLAYING DATA



## Stem-and Leaf-Plots One More Time...

Quiz scores

Convert the quiz scores in the table into percentages. Create a stem and leaf plot for the quiz score percentages.

| STEM | LEAF |
|---|---|
| | |

| Quiz Score/20 | Percentage |
|---|---|
| 12 | |
| 21 | |
| 18 | |
| 21 | |
| 12 | |
| 15 | |
| 17 | |
| 17 | |
| 18 | |
| 19 | |
| 14 | |
| 18 | |
| 15 | |
| 17 | |
| 17 | |
| 18 | |
| 9 | |
| 18 | |
| 18 | |
| 20 | |

The range (maximum – minimum):

How many entries are in the stem and leaf plot (i.e. what is *n)*?

What is the median (the center)?

What is the interquartile range?

What is Q1?

What is Q3?

Interquartile Range (IQR)=