Section 2.1

Organizing Qualitative Data

Objectives

- Organize Qualitative Data in Tables
- Construct Bar Graphs
- Construct Pie Charts

A **frequency distribution** lists each category of data and the number of occurrences for each category of data.

Color	Tally	Frequency
Brown	||||| ||||| ||	12
Yellow	||||| |||||	10
Red	||||| ||||	9
Orange	||||| |	6
Blue	|||	3
Green	|||||	5

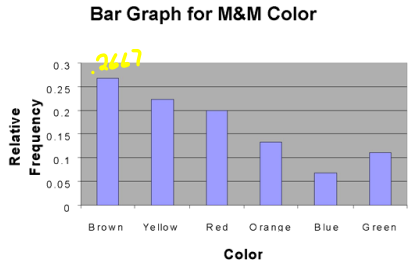
The **relative frequency** is the proportion (or percent) of observations within a category and is found using the formula:

$$\text{relative frequency} = \frac{\text{frequency}}{\text{sum of all frequencies}}$$

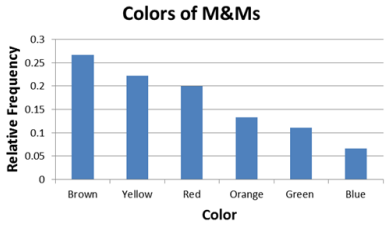
A **relative frequency distribution** lists each category of data with the relative frequency.

Color	Tally	Frequency	Relative Frequency
Brown	||||| ||||| ||	12	12/45 ≈ 0.2667
Yellow	||||| |||||	10	0.2222
Red	||||| ||||	9	0.2
Orange	||||| |	6	0.1333
Blue	|||	3	0.0667
Green	|||||	5	0.1111

45 1

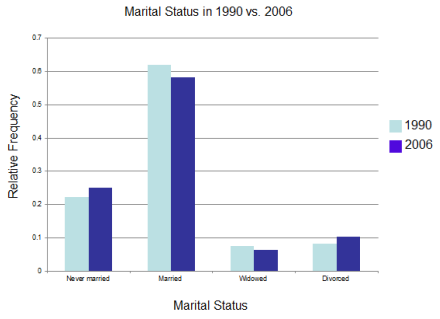


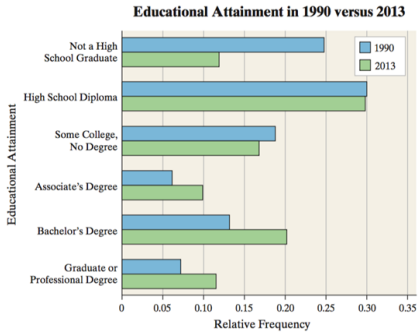
A **Pareto chart** is a bar graph where the bars are drawn in decreasing order of frequency or relative frequency.



The following data represent the marital status (in millions) of U.S. residents 18 years of age or older in 1990 and 2006. Draw a side-by-side relative frequency bar graph of the data.

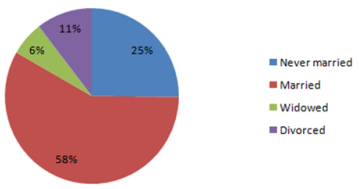
Marital Status	1990	2006
Never married	40.4	55.3
Married	112.6	127.7
Widowed	13.8	13.9
Divorced	15.1	22.8





EXAMPLE *Constructing a Pie Chart*

Marital Status, 2006



Section 2.2

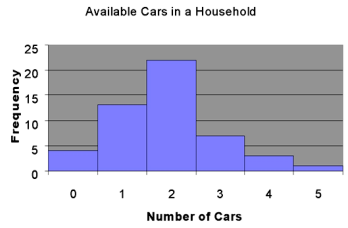
Organizing Quantitative Data: The Popular Displays

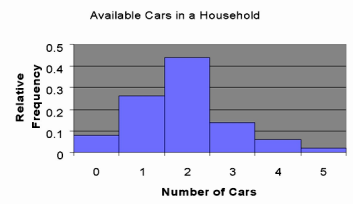
Objectives

- Organize discrete data in tables
- Construct histograms of discrete data
- Organize continuous data in tables
- Construct histograms of continuous data
- Draw stem-and-leaf plots
- Draw dot plots
- Identify the shape of a distribution

Draw a frequency and relative frequency histogram for the "number of cars per household" data.

# of Cars	Frequency	Relative Frequency
0	4	4/50 = 0.08
1	13	13/50 = 0.26
2	22	0.44
3	7	0.14
4	3	0.06
5	1	0.02





The following data represent the time between eruptions (in seconds) for a random sample of 45 eruptions at the Old Faithful Geyser in Wyoming. Construct a frequency and relative frequency distribution of the data.

728	678	723	735	703
730	722	708	714	713
726	716	736	719	672
698	702	738	725	711
721	703	735	699	695
722	718	695	702	731
700	703	706	733	726
720	723	711	696	695
729	699	714	700	718

The smallest data value is 672 and the largest data value is 738. We will create the classes so that the lower class limit of the first class is 670 and the class width is 10 and obtain the following classes:

670 - 679
680 - 689
690 - 699
700 - 709
710 - 719
720 - 729
730 - 739

Time between Eruptions (seconds)	Tally	Frequency	Relative Frequency
670 – 679	||	2	2/45 = 0.044
680 - 689		0	0
690 - 699	||||| ||	7	0.1556
700 - 709	||||| ||||	9	0.2
710 - 719	||||| ||||	9	0.2
720 - 729	||||| ||||| |	11	0.2444
730 - 739	||||| ||	7	0.1556

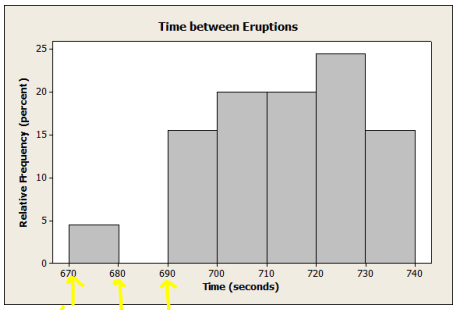
Guidelines for Determining the Lower Class Limit of the First Class and Class Width

Determining the Class Width

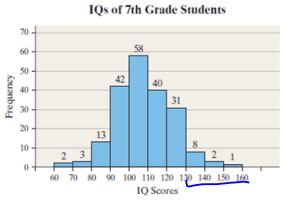
Decide on the number of classes. Generally, there should be between 5 and 20 classes. The smaller the data set, the fewer classes you should have. 6–12

Determine the class width by computing

$$\text{Class width} = \frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}$$



Example:



How many students sampled?
200

Class width?
10

Identify the class est Frequency.

Class	Freq.
60-69	2
70-79	3
80-89	13

and so on

Which class has the highest Frequency? 100-109

Lowest? 150-159

What % of students have an IQ of at least 130? $\frac{11}{200} = .055$

5.5%

A **stem-and-leaf plot** uses digits to the left of the rightmost digit to form the **stem.** Each rightmost digit forms a **leaf.**

For example, a data value of 147 would have 14 as the stem and 7 as the leaf.

147 — 1 Stem, 4 Leaf

1578 — 157 Stem, 8 Leaf

An individual is considered to be unemployed if they do not have a job, but are actively seeking employment. The following data represent the unemployment rate in each of the fifty United States plus the District of Columbia in June, 2008.

State13	Unemployment Rate	State	Unemployment Rate	State	Unemployment Rate
Alabama	4.7	Kentucky	6.3	North Dakota	3.2
Alaska	6.8	Louisiana	3.8	Ohio	6.6
Arizona	4.8	Maine	5.3	Oklahoma	3.9
Arkansas	5.0	Maryland	4.0	Oregon	5.5
California	6.9	Mass	5.2	Penn	5.2
Colorado	5.1	Michigan	8.5	Rhode Island	7.5
Conn	5.4	Minnesota	5.3	South Carolina	6.2
Delaware	4.2	Mississippi	6.9	South Dakota	2.8
Dist Col	6.4	Missouri	5.7	Tenn	6.5
Florida	5.5	Montana	4.1	Texas	4.4
Georgia	5.7	Nebraska	3.3	Utah	3.2
Hawaii	3.8	Nevada	6.4	Vermont	4.7
Idaho	3.8	New Hamp	4.0	Virginia	4.0
Illinois	6.8	New Jersey	5.3	Washington	5.5
Indiana	5.8	New Mexico	3.9	W. Virginia	5.3
Iowa	4.0	New York	5.3	Wisconsin	4.6
Kansas	4.3	North Carolina	6.0	Wyoming	3.2

5.1

3.9

3.2

We let the stem represent the integer portion of the number and the leaf will be the decimal portion. For example, the stem of Alabama (4.7) will be 4 and the leaf will be 7.

2	8
3	222388899
4	000012346778
5	0122333334555778
6	02344568899
7	5
8	5

4.0

A split stem-and-leaf plot:

This stem represents

This stem represents 3.5 – 3.9

2	8
3	2223
3	88899
4	00001234
4	6778
5	0122333334
5	555778
6	02344
6	568899
7	
7	5
8	
8	5

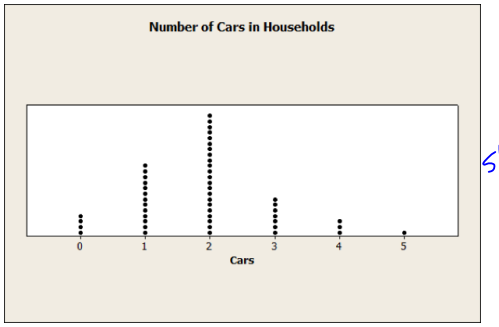
A **dot plot** is drawn by placing each observation horizontally in increasing order and placing a dot above the observation each time it is observed.

EXAMPLE *Drawing a Dot Plot*

The following data represent the number of available cars in a household based on a random sample of 50 households. Draw a dot plot of the data.

3	0	1	2	1	1	1	2	0	2
4	2	2	2	1	2	2	0	2	4
1	1	3	2	4	1	2	1	2	2
3	3	2	1	2	2	0	3	2	2
2	3	2	1	2	2	1	1	3	5

Data based on results reported by the United States Bureau of the Census.



Skewed Right

Uniform distribution the frequency of each value of the variable is evenly spread out across the values of the variable

Bell-shaped distribution the highest frequency occurs in the middle and frequencies tail off to the left and right of the middle

Skewed right the tail to the right of the peak is longer than the tail to the left of the peak

Skewed left the tail to the left of the peak is longer than the tail to the right of the peak.