**Handout #2: Descriptive Statistics**  
  
After the data has been collected, it needs to be organized, examined using various graphs, and summarized in order to learn from the data.   
  
**Note: Data always tell you some kind of story.**  
Here we learn various methods to accomplish these.  
  
- Construct frequency distribution (for organizing data)  
- Display data in graphs  
- Calculate numerical summaries: measures of central tendency, measures of variation & measures   
 of position

**Frequency Distribution and Graphs for Qualitative Data**

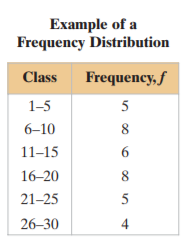
**- Frequency distribution:** Data are categorized into classes and then count the number of observations (data values) in each class, known as Class Frequency.  
  
**- Graphical display:** bar graph, pie chart

**Example:** Top Broadcast Shows. The networks for the top 20 television shows, as determined by the Nielsen Ratings for the week of April 18–24, 2005, are shown below. Construct a frequency distribution and a relative-frequency distribution for these network data. Using the constructed frequency distribution, what do you learn about the data? Construct bar graph and pie chart.  
  
CBS Fox ABC Fox CBS CBS ABC Fox CBS ABC CBS CBS NBC CBS NBC NBC CBS CBS NBC NBC

**Frequency Distribution and Graphs for Quantitative Data**

**Frequency Distribution:** construct a frequency distribution using class intervals.

**Graphical display:** Histogram, Frequency Polygon, Ogive, Stem-and-leaf plot, Dot plot, Time Series plot, Scatter plot etc.



**Important terms**  
**1. Lower class limit:** the least number that can belong to the class.

**2. Upper class limit:** the greatest number that can belong to the class.

**3. Class width:** the distance between lower (or upper) limits of consecutive classes.

**5. Midpoint**   
  
  
**6. Relative Frequency**

**7. Cumulative frequency (of a class):** is the sum of the frequencies of that class and all previous classes.

**8. Class boundaries:** are the numbers that separate classes without forming gaps between them.   
If data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries. To find the upper class boundaries, add 0.5 to each upper limit. The upper boundary of a class will equal the lower boundary of the next higher class.

**Guidelines to construct a Frequency Distribution:**  
1. Decide on the number of classes. (usually between 5 to 20 classes)   
  
2. Calculate the class width: , where Range = Max – Min.  
 Round up the class width to the nearest integer.   
  
3. Choose a number (usually minimum data value) for the lower limit of the first class to start with.  
4. Add class width to the lower limit of the first class to get the lower limit of second class.   
 Continue this for the other classes.  
  
5. Enter the upper class limits (1 or 0.1 or 0.01 unit less than the lower limit of next class depending   
 on the number of decimal in the data, e.g. for integer type, subtract 1).   
  
6. Count the observations within a class to obtain class frequency. You can add three more additional columns: class mid-point, relative frequency and cumulative frequency to construct a complete frequency distribution.

**Example:** Construct a frequency distribution table for the following data values using 5 classes. Construct histogram, frequency polygon and Ogive also.

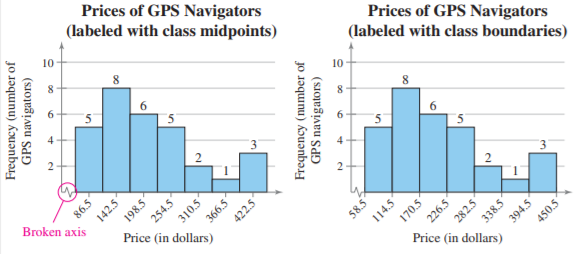
44, 32, 35, 38, 35, 39, 42, 36, 36, 40, 51, 58, 58, 62, 63, 72, 78, 81, 25, 84, 20.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classes | Tally Mark |  | Class mid -point | Relative Frequency | Cumulative  Frequency |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Total |  |  |  |  |  |

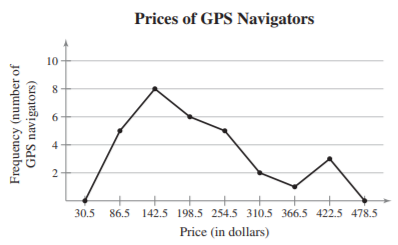
**Note:** This type of frequency distribution table is useful for large number of observations.

**Histogram:** Is a bar graph with the classes (or midpoints) on the x-axis and the class frequencies (or relative frequencies) on the y-axis. Consecutive bars must touch.

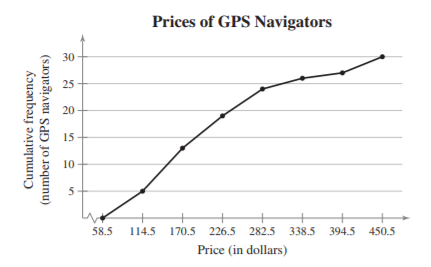
Because consecutive bars of a histogram must touch, bars must begin and end at class boundaries instead of class limits.



Use the frequency distribution of obtained above, construct a **histogram**.

**Frequency polygon**: line graph used to graph frequencies. Plot points that represent the midpoint and frequency of each class and connect the points in order from left to right.  
  
Use the frequency distribution of obtained above, construct a **frequency polygon**.

**Ogive (cumulative frequency graph):** is a line graph that displays the cumulative frequency of each class at its upper class boundary. The upper boundaries are marked on the horizontal axis, and the cumulative frequencies are marked on the vertical axis. An ogive tells how many data are less than the indicated value on the horizontal axis.

****

Use the frequency distribution of obtained above, construct an **ogive**.

**Stem-and-leaf plot (good for small data sets):** Separate each number in a data set into a stem and a leaf (usually the last digit).   
  
**Example:** The data represents the scores of a biology class on a midterm exam.  
75 85 90 80 87 67 82 88 95 91 73 80 83 92  
94 68 75 91 79 95 87 76 91 85

**Example:** Sam got his friends to do a long jump and got these results:  
2.3, 2.5, 2.5, 2.7, 2.8 3.2, 3.6, 3.6, 4.5, 5.0. Construct a stem-and-leaf plot and describe what do you learn about the data.

**Dot Plots:** Represent each data value with a dot along a numerical scale. When data values repeat, the dots pile up vertically at that value.

**Example:** Robert, a sixth grader at Roosevelt Middle School, usually goes to bed around 10:00 p.m. and gets up around 6:00 a.m. to get ready for school. That means he gets about 𝟖 hours of sleep on a school night. He decided to investigate the statistical question: How many hours per night do sixth graders usually sleep when they have school the next day? Robert took a survey of 𝟐𝟗 sixth graders and collected the following data to answer the question.   
  
𝟕 𝟖 𝟓 𝟗 𝟗 𝟗 𝟕 𝟕 𝟏𝟎 𝟏𝟎 𝟏𝟏 𝟗 𝟖 𝟖 𝟖 𝟏𝟐 𝟔 𝟏𝟏 𝟏𝟎 𝟖 𝟖 𝟗 𝟗 𝟗 𝟖 𝟏𝟎 𝟗 𝟗 𝟖  
   
 Robert decided to make a dot plot of the data to help him answer his statistical question.

a. What are the least and most hours of sleep reported in the survey of six graders?

b. What number of hours slept occurred most often in the data set?

**Time Series Graph**: A time series graph represents data that occur over a specific period of time.

**Example (Work together):** The percentages of voters voting in the last 5 Presidential elections are shown here. Construct a time series graph.

Year 1984 1988 1992 1996 2000

% of voters voting 74.63 72.48 78.01 65.97 67.50

a. Highest voters in: lowest voters in:

b. Data trend (overall)  
c. Sharp increase in? How much?  
d. Sharp decline in? How much?

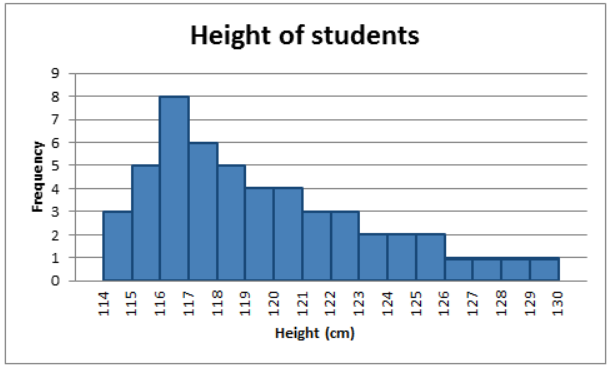
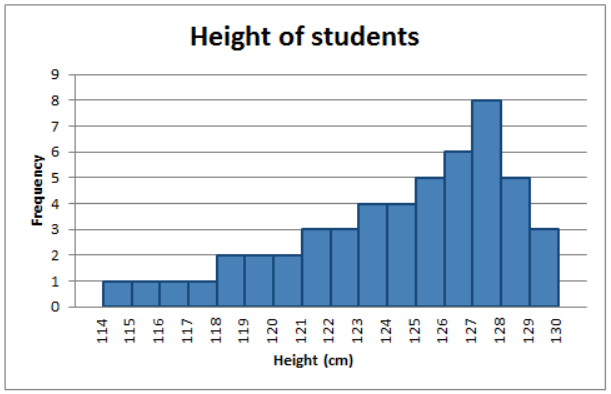
**Scatter Plot:** is a graph of order pairs of data values that is used to determine if a relationship exists between the two variables.

**Example (Work together):** A researcher wishes to determine if there is a relationship between the number of days an employee missed a year and the person’s age. Draw a scatter plot and comment on the nature of the relationship.

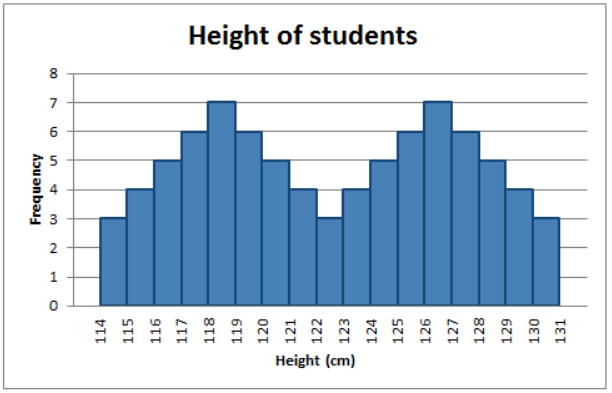
Age, x 22 30 25 35 65 50 27 53 42 58

Days missed, y 0 4 1 2 14 7 3 8 6 4

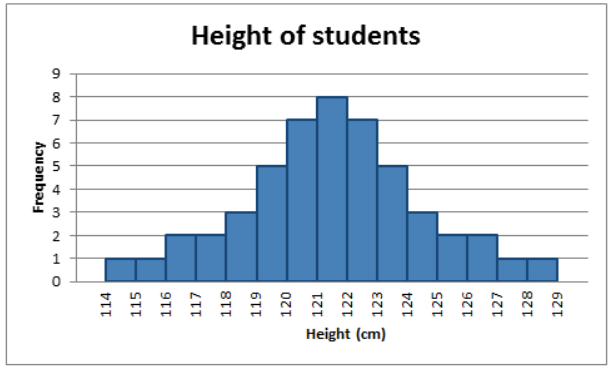
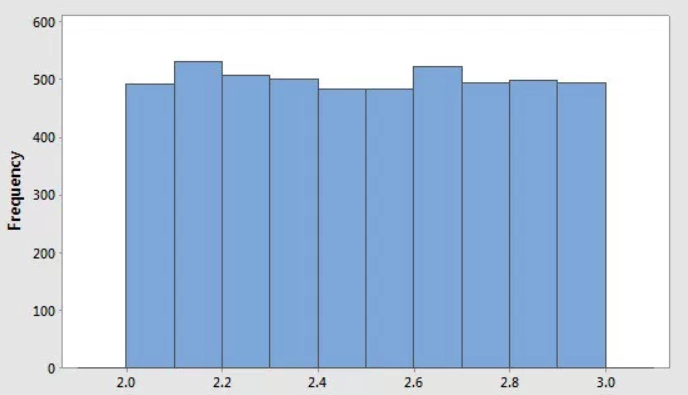
**Shapes of data distribution**:   
Describe the distribution (or pattern) of the data within a dataset.  
  
1. Symmetric distribution  
2. Asymmetrical distribution  
**Next Handout #3:** You will learn how to find numerical summaries of data

a. Positively skewed (right skewed) b. Negatively skewed (left skewed)



Bi-modal

a. Bell shaped symmetric (Normal) b. Uniform Distribution