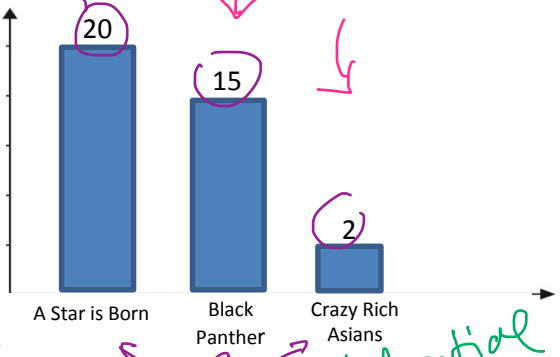
Jan 7

Chapter 2: Exploring Data with Tables and Graphs

2.1 Frequency Distributions for Organizing & Summarizing Data

BAR GRAPHS

Ex: We ask a sample of people who they believe will win the Oscar for Best Picture.



a) Is this quantitative or qualitative data?

b) What percentage of people said "A Star is Born"?

c) What percentage of people said "Black Panther" or "Crazy Rich Asians"?

d) Is this an example of descriptive statistics or inferential statistics?

e) Our above sample consists of Marina Del Rey residents. Let our population be Marina Del Rey residents and there are 10,000 residents in Marina Del Rey. How many would you expect to say "A Star is Born" will win the Oscar for Best Picture?

FREQUENCY DISTRIBUTIONS

Def A ***frequency distribution*** lists all of the categories along with their corresponding frequencies.

Note: A frequency table shows how a data set is distributed among several classes.

STANDARD TERMS OF A FREQUENCY DISTRIBUTION

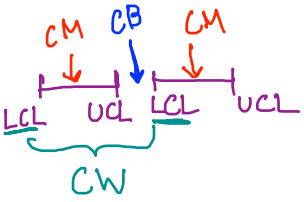
Def ***Lower class limits*** (LCL) are the smallest numbers that can belong to the different classes.

Def ***Upper class limits*** (UCL) are the largest numbers that can belong to the different classes.

Def ***Class Boundaries*** are the numbers used to separate the classes without gaps.

How to find?

1. Find the size of the gap between the upper class limit of one class and the lower class limit of the next class. Take half that amount.
2. Subtract that amount from the first LCL, and add that amount to the last UCL.
3. For the other boundaries, add the LCL from one class to the UCL from the next class, and divide by two.



Def ***Class midpoints*** are the values in the middle of the classes.

How to find? Class midpoint is the average of the lower class limit and the upper class limit.

$$\text{Class Midpoint} = \frac{LCL + UCL}{2}$$

Def ***Class width*** is the difference between two consecutive lower (or upper) class limits, or, if the classes are not given, is found by doing the steps on the next page.

Steps to Create Intervals/Classes:

1st: Decide on the number of classes *(let's say 6 classes)*

2nd: Find the class width using…

$$class\ width \approx \frac{largest\ data\ value - smallest\ data\ value}{number\ of\ classes}\ (round\ up!)$$

$CW = \frac{max - min}{6}$

3rd: Start with the smallest data value as the *lower class limit* and add the class width to get the next *lower class limit.* Use the class width to determine the *upper class limits*.

4th: Continue until you get to the last class and it should include all of your numbers.*The last class should include you

Ex: The following data set represents the age at which each president was first inaugurated.

57	61	57	57	58	57	61	54	68	51	49	64	50
48	65	52	56	46	54	49	51	47	55	55	54	42 (min)
51	56	55	51	54	51	60	62	43	55	56	61	52
69	64	46	54	47	70 (max)							

$CW = \frac{70 - 42}{6} = 4.7 \approx 5$

Use the data to complete the given frequency distribution and find the following. (+4)

Warning
Make sure there's always 2 extra CB; one at beginning & one at end.

Ages	*Frequency*
42 – 46 (pick age before next class)	𝍸 (4)
47 – 51 (+CW)	𝍸 𝍸 | (11)
52 – 56 (+CW)	𝍸 𝍸 𝍸 (14)
57 – 61 (+CW)	𝍸 𝍸 (9)
62 – 66 (+CW)	𝍸 (4)
67 – 71 (+CW)	||| (3)
Total	45

(a) lower class limits

42, 47, 52, 57, 62, 67 years

(b) upper class limits

46, 51, 56, 61, 66, 71 years

(c) → class boundaries

41.5, 46.5, 51.5, 56.5, 61.5, 66.5, 71.5 ← use to label histogram

(d) class midpoints

$CM = \frac{42+46}{2} = 44$, $CM = \frac{47+51}{2} = 49$, $CM = \frac{52+56}{2} = 54$, $CM = \frac{57+61}{2} = 59$, $CM = \frac{62+66}{2} = 64$,

$CM = 69 = \frac{67+71}{2}$

(e) class width

5 years

Relative Frequency Distribution

Def A **relative frequency distribution** includes the same class limits as a frequency distribution, but relative frequencies are used instead of actual frequencies (or counts.)

$$\textit{relative frequency} = \frac{\textit{class frequency}}{\textit{sum of all frequencies}} = \frac{\textit{count}}{\textit{total}}$$

Note: Relative frequencies can also be re-expressed as percentage frequencies.

Ex: Use the previous data to complete the given relative frequency distribution. total = 45

(To 3 decimal places)

Ages	*Relative Frequency*	Freq	
42-46	4/45 ≈ 0.089	4	8.9%
47-51	11/45 ≈ 0.244	11	24.4%
52-56	14/45 ≈ 0.311	14	31.1%
57-61	9/45 ≈ 0.200	9	20.0%
62-66	4/45 ≈ 0.089	4	8.9%
67-71	3/45 ≈ 0.067	3	6.7%
	add these up = 1.00		add 100%

"sum"

Cumulative Frequency Distribution

Ex: Use the data to complete the given cumulative frequency distribution.

Ages	*Cumulative Frequency*	
Under 46	4	4
Under 51	15	4+11
Under 56	29	14+11+4
Under 61	38	9+29
Under 66	42	4+38
Under 71	45	3+42

2.2 Histograms

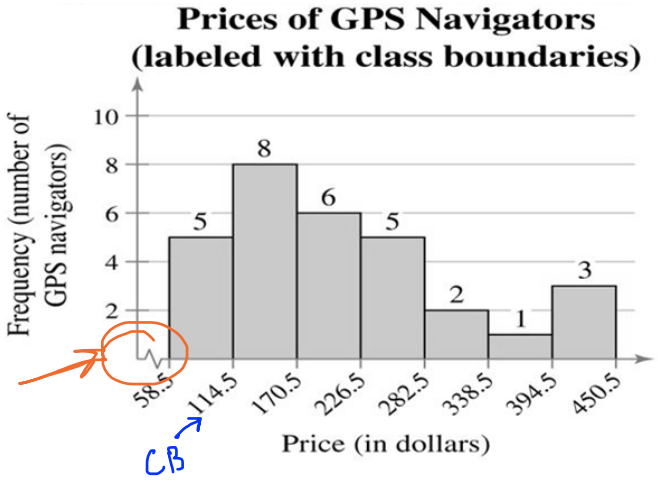
touch / no gaps

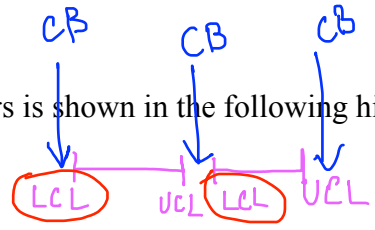
Def A ***histogram*** is a graph consisting of bars of equal width drawn adjacent to each other. The heights of the bars correspond to the frequency values.

Horizontal scale: class of quantitative data values (use class boundaries to partition)

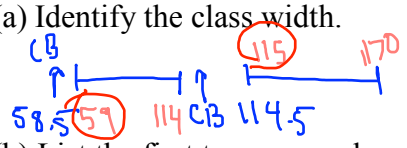
Vertical scale: frequencies

Ex: The prices of 30 randomly selected GPS navigators is shown in the following histogram.





(a) Identify the class width.



LCL LCL

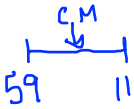
CW = 115 − 59

CW = 56

(b) List the first two upper class limits.

114 & 170

(c) Identify the first two class midpoint(s).



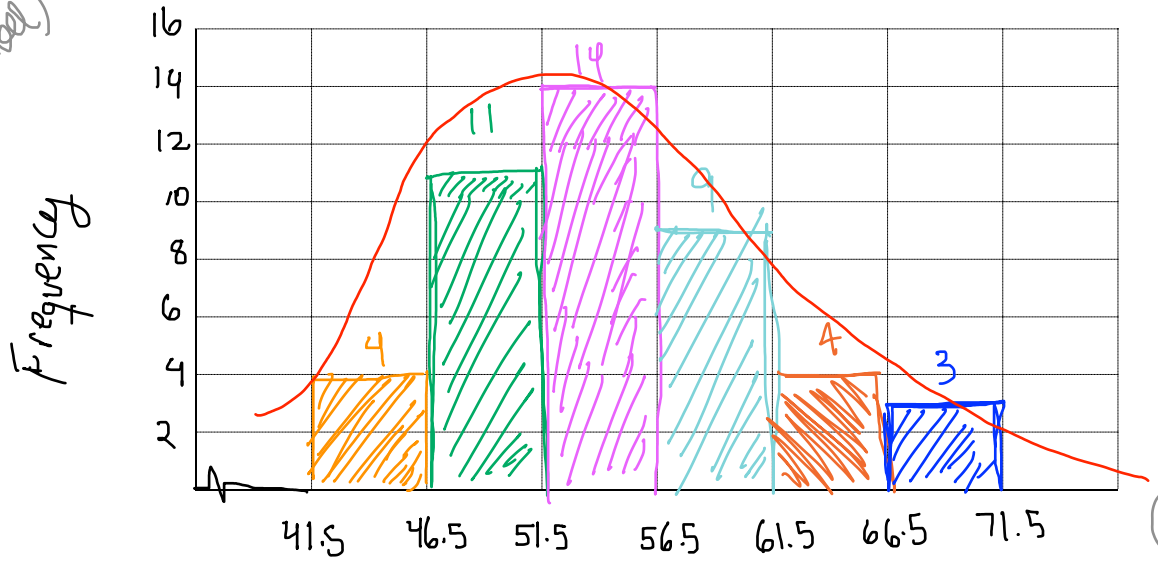
CM = $\frac{59+114}{2}$ = 86.5

CM2 = 142.5

Ex: Construct a histogram for the data set regarding inaugural age. Be sure to label your axes and give the histogram a title.

President's Inaugural Ages

(scale axes) (label)



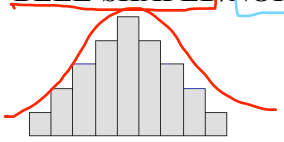
Shape of Distribution

→ approx. Bell Shaped

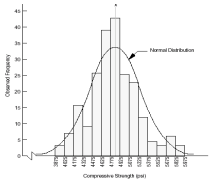
→ very slightly Skewed Right

COMMON DISTRIBUTION SHAPES

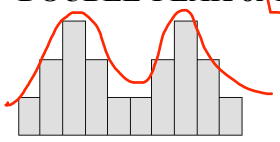
BELL-SHAPED/NORMAL or UNIMODAL



Examples: weight, height, show size, state standardized tests

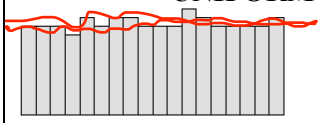


DOUBLE-PEAK or BIMODAL



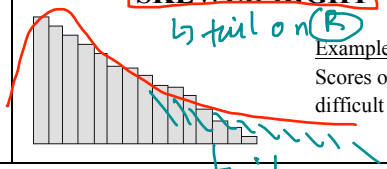
Example: Shoe size of males and females

UNIFORM



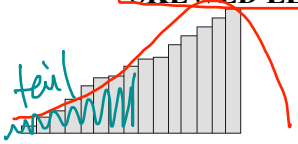
Example: rolling a die, and landing on 1

SKEWED RIGHT



Example: Scores of a difficult exam

SKEWED LEFT



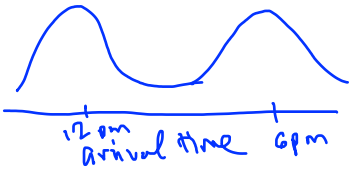
Example: Scores of an easy exam

Ex: For the following data, identify which of the above distributions would be most likely to occur.

(a) A sample of 1000 home prices in LA county.

Skewed Right

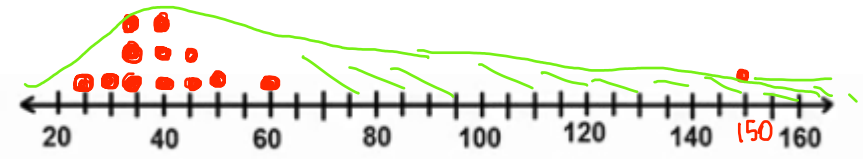
(b) A sample 500 of arrival times at a restaurant that is open for lunch and dinner.



2.3 Graphs that Enlighten and Ones that Deceive

Def **Dotplots** graph each data value as a dot above a horizontal scale of values. Dots representing equal values that are stacked.

Ex: Draw a dotplot for the following amount of cash that thirteen people have in their wallet.
25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, and 150



* *What is the shape of the distribution?*

Skewed Right

**What value appears to be an outlier?*

150

outlier

150 data point that is very different than the rest of data

Def A ***stemplot*** (or ***stem-and-leaf plot***) represents quantitative data by separating each into two parts: the stem (such as the leftmost digit) and the leaf (such as the rightmost digit).

Ex: The given data represents hourly wages (in dollars) of 30 randomly selected college students. Construct a stemplot for the given data.

11.5	13.4	12.8	15.5	10.9	15.0	12.5	17.2	11.0	11.2
12.7	14.3	12.5	18.0	11.8	10.9	11.6	13.5	12.5	16.5
18.5	12.4	12.9	10.8	14.8	12.9	12.3	15.5	17.6	15.0

Stems (units)	Leaves (tenths)
10	9 9 6
11	5 0 2 8 6
12	8 5 7 5 5 4 9 9 3
13	4 5
14	3 8
15	5 0 5 0
16	5
17	2 6
18	0 5

What is the shape of this distribution?

Skewed Right

Def A **pareto chart** is a bar graph with the condition of arranging the bars in descending order of frequency.

Def A **pie graph** depicts categories where each slice of the pie is proportional to the frequency count for the category.

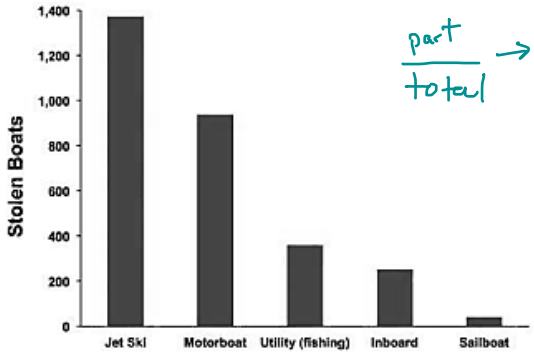


FIGURE 2-8 Pareto Chart of Stolen Boats

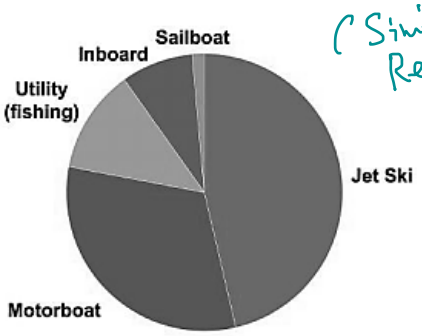
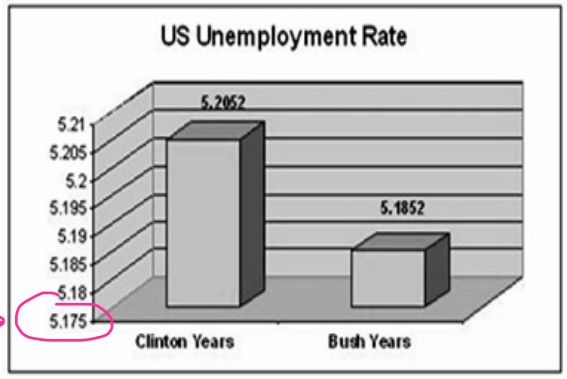


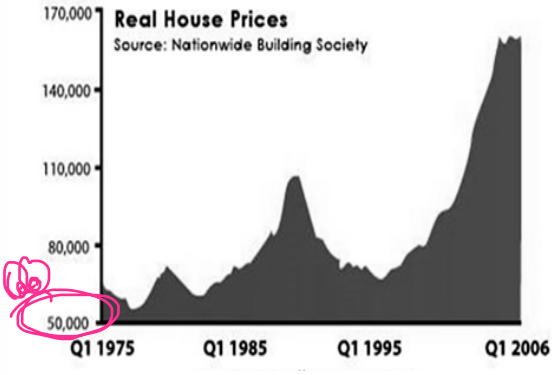
FIGURE 2-9 Pie Chart of Stolen Boats

DECEPTIVE GRAPHS

1. *Nonzero Vertical Axis*

Always examine a graph to see whether an axis begins at some point other than zero, which exaggerates differences.





2. *Pictographs* The given graphs are misleading because objects of area or volume are used to depict amounts that are actually one-dimensional.