L9: Feb. 6, 2017.

Housekeeping.

- Writing assignment due 11:59 p.m. tonight on Canvas
- Homework due Wednesday in Class
- Quiz on Wednesday
- Choose your book — 1st week's reading & response will be due on Monday (assignment to be posted on Canvas by Wednesday).

Last time / last chapter:

- Definitions — parameter, statistic, variables, data, population, sample
- Data — qualitative / quantitative, cts. / discrete
- Sampling

... Questions?

Next chapter:

Descriptive Statistics

NEW CHAPTER... NEW GROUPS

In Chapter 1, we learned what constituted fair sampling and fair studies whose goal was collecting data.

In Chapter 2, the goal is to come up with ways of describing and displaying the data we collect.

(... after that, we'll learn how to use data to make inferences about populations — but that's later...)

Recall: Last Friday, we drew STEM-AND-LEAF PLOTS of heights. In general, stem & leaf plots work...

- For data sets that are relatively small (not an efficient/effective way of showing thousands of data points over a small range, for example)

- When data points can be meaningfully divided into STEMS and LEAVES — the leaf is the final significant digit, & the stem is everything else.

They're sometimes a good way of identifying OUTLIERS.

Example 2.2, p.69

Distances (km) from homes to nearest supermarkets =

STEM LEAF

{ 1 | 1, 1.5, 2.3, 2.5, 3.2, 3.3, 3.5, 4.0, 4.2, 4.5, 4.7, 4.8, 5.5, 5.6, 6.5, 6.7, 12.3 }

Draw the stem & leaf plot:

Stem	Leaf
1	1 5
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

An OUTLIER is a value in a data set that doesn't fit with the rest of the data — an EXTREME VALUE.

L9, Ct'd.

Sometimes, if ~~data~~ two data sets have the same (or close) stems, it can be useful to plot them SIDE-BY-SIDE.

For example...
2.3, p. 71

AGES AT INAUGURAT'N	STEM	AGES AT DEATH
9 9 8 7 7 7 3 2	4	6 9
~~7~~7766655544444111110	5	3 6 6 7 7 8
9 5 4 2 1 1 1 0	6	0 0 3 3 4 4 5 6 7 7 7 8
	7	0 0 1 1 1 4 7 8 8 9
	8	0 1 3 5 8
	9	0 0 3 3

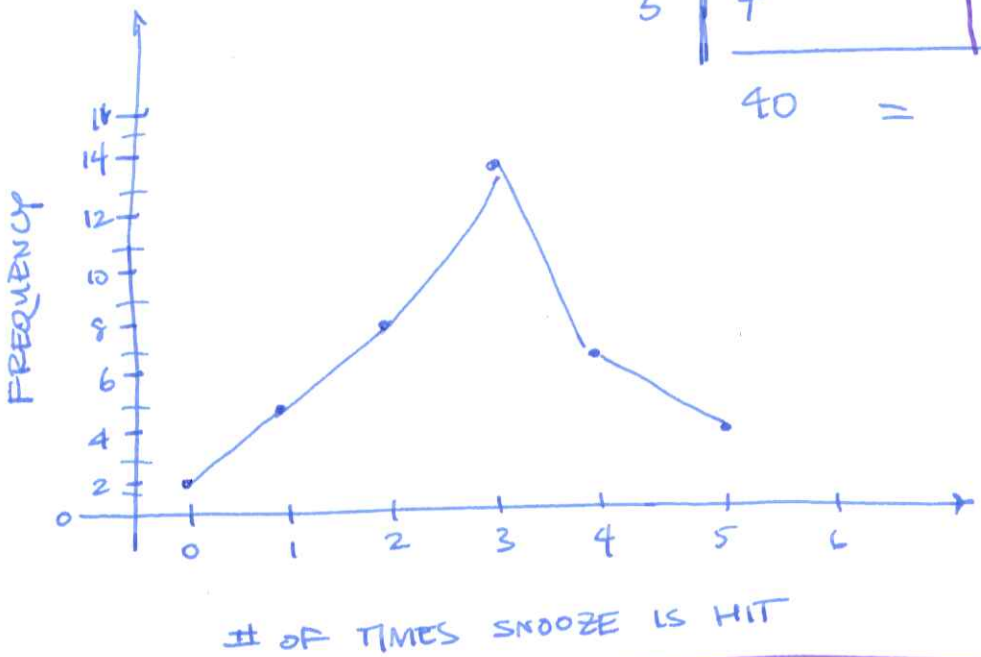
Leaves decrease toward stem (strictly speaking, which direction they increase/decrease is not really important, but it's nice to keep them in some kind of order).

Try It: Ex. 2.3, p. 72

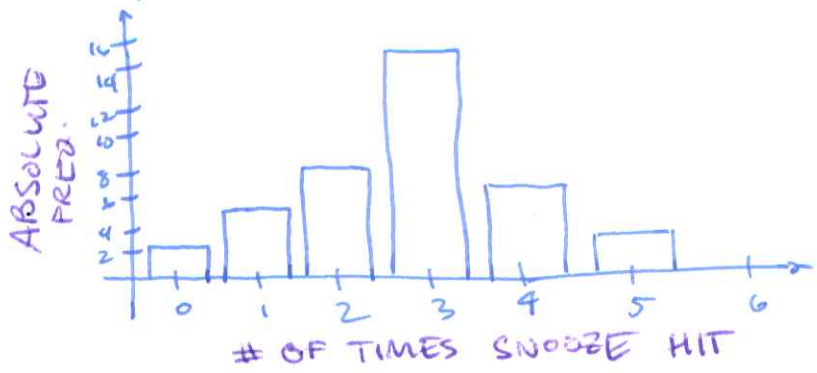
- Line charts are sometimes useful for displaying data that's already sorted into a frequency table:

Example: 2, 4, p. 73

# times student hits snooze button	Absolute Frequency	Relative Freq.
0	2	2/40 = 5%
1	5	5/40 = ____
2	8	8/40 = ____
3	14	14/40 = ____
4	7	7/40 = ____
5	4	4/40 = ____
	40 =	TOTAL # OF STUDENTS SAMPLED



... could also use a bar chart — but would have to separate the bars:



Bar charts also work well for frequency tables where the classes correspond to ranges of the data.

Example # of people in each age group & proportion of populatio

TRY IT 2.5
p. 74 ...

If your data is already in a frequency table, then a HISTOGRAM is the same thing as a bar chart.

If it isn't (maybe you just finished collecting it, or maybe someone was generous enough to share all of his or her data with you), then you need to come up with the frequency chart on your own...

① Decide how many classes you want
(usu. btwn. 5 & 15 ?) (sqrt. of # of data pts.?)

② Choose the starting pt. of the 1st class & interval to be less than the lowest data pt.
(maybe 1 more decimal place lower?)
e.g., 6.1 ⟹ start at 6.05 ?

TRY TO DO ① & ② SO THAT BOUNDARIES DON'T FALL ON (TOO MANY) DATA POINTS?