Chp 1 Picturing Distributions w/ Graphs

Vocab

- **individuals**: objects described by a set of data. (in a spreadsheet, it's represented by a row of data)
- **variables**: any measurable quantity / characteristic of an individual (in a spreadsheet, it's represented by a column)
  - **categorical**: places an individual into one of several groups/categories (ex gender)
  - **quantitative**: takes numerical values & arithmetic ops can be done (like finding an average); requires a unit of measurement (ex weight)

Ex 1 Fuel Economy data

make/model	vehicle type	transmission type	# of cylinders	city mpg	hwy mpg
⋮					
honda civic	subcompact	auto.	4	25	36
toyota prius	midsize	auto.	4	48	45
Aston Martin vantage	2-seater	manual	8	12	19
⋮					

(a) what are individuals?

(b) what variables are given? Are they categorical or quantitative?

M1070

Chp 1 (Cont)

Ex 2] You are preparing to study tv-viewing habits of college students. Describe two categorical variables and two quantitative variables (w/ units of measure).

Vocab

- Distribution: distribution of a variable tells what values it takes on and how often.
- Graphs of distributions:
  1. Pie chart (or circle graph): useful for categorical vars; "slices" are sized by percents for categories; must include all the categories in the whole
  2. Bar graph: useful for categorical vars; horizontal axis = all categories and vertical axis = frequency (as a raw count or percent); more flexible than pie chart & easier to read; a bar graph can also compare quantities that are measured in same units

Chp 1 (cont)

③ • Histogram: useful for quantitative vars; horizontal axis = grouped variable values and vertical axis = frequency; visually it looks like bar graph except bars are next to each other w/ no space between them; all bars have same width.

☆ bar graph compares sizes of different quantities; histogram displays distribution of quantitative var.

Ex 3 In 2005, here were avg # of babies born on each day of the week.

Day	Births
Sun.	7,374
Mon.	11,704
Tues.	13,169
Wed.	13,038
Thurs.	13,013
Fri.	12,664
Sat	8,459

(a) make pie chart.

(b) make bar graph.

Chp 1 (cont)

Ex 4 If all categories of a variable add to only 61.2%, can we use a pie chart to display data? Why or why not?

Ex 5 Avg travel time to work (in minutes): make a histogram for this data.

STATE	TIME	STATE	TIME	STATE	TIME
Alabama	23.6	Louisiana	25.1	Ohio	22.1
Alaska	17.7	Maine	22.3	Oklahoma	20.0
Arizona	25.0	Maryland	30.6	Oregon	21.8
Arkansas	20.7	Massachusetts	26.6	Pennsylvania	25.0
California	26.8	Michigan	23.4	Rhode Island	22.3
Colorado	23.9	Minnesota	22.0	South Carolina	22.9
Connecticut	24.1	Mississippi	24.0	South Dakota	15.9
Delaware	23.6	Missouri	22.9	Tennessee	23.5
Florida	25.9	Montana	17.6	Texas	24.6
Georgia	27.3	Nebraska	17.7	Utah	20.8
Hawaii	25.5	Nevada	24.2	Vermont	21.2
Idaho	20.1	New Hampshire	24.6	Virginia	26.9
Illinois	27.9	New Jersey	29.1	Washington	25.2
Indiana	22.3	New Mexico	20.9	West Virginia	25.6
Iowa	18.2	New York	30.9	Wisconsin	20.8
Kansas	18.5	North Carolina	23.4	Wyoming	17.9
Kentucky	22.4	North Dakota	15.5	Dist. of Columbia	29.2

Ch 1 (cont)

More Graphs of Distributions

④. Stemplot: useful for quantitative data; can use instead of histogram, but only if it's a reasonably small data set; it presents more detailed info than histogram (☆ see table pg 19 of book on how to make one)

⑤. Time plot: useful for variable that is measured over time; horizontal axis = time & vertical axis = measurable variable; connect the dots (ex stock price or salary); presents time series data to show how it changes over time.

Ex 6 make stemplot for Ex 5 data.

Chp 1 (cont)

Ex 7 Average tuition table for in-state students at public 4yr colleges/universities.

Year	Tuition	Year	Tuition	Year	Tuition	Year	Tuition
1976	$2,197	1984	$2,426	1992	$3,444	2000	$4,221
1977	$2,225	1985	$2,532	1993	$3,623	2001	$4,411
1978	$1,986	1986	$2,656	1994	$3,758	2002	$4,715
1979	$1,986	1987	$2,699	1995	$3,802	2003	$5,231
1980	$1,939	1988	$2,721	1996	$3,913	2004	$5,624
1981	$2,018	1989	$2,792	1997	$4,022	2005	$5,814
1982	$2,194	1990	$2,977	1998	$4,131	2006	$5,918
1983	$2,358	1991	$3,187	1999	$4,183	2007	$6,185

make a time plot. Describe any patterns (trends) you see.

Chp1 (cont)

Interpreting Graphs (Vocab)

Look for overall patterns and notice any deviations.

- shape: is it skewed or symmetric?
  - skewed: a distribution is skewed to the right if the RHS of histogram extends much farther than the left; it's skewed to the left if LHS extends much farther than right.
  - symmetric: if RHS & LHS are basically mirror images of each other.
- spread: lowest and highest values; range of values.
- outlier: an individual value that is way outside overall pattern.

Ex 8 For histogram in ex 5, describe shape, spread, skewness, symmetry, overall pattern + outliers (if any)