

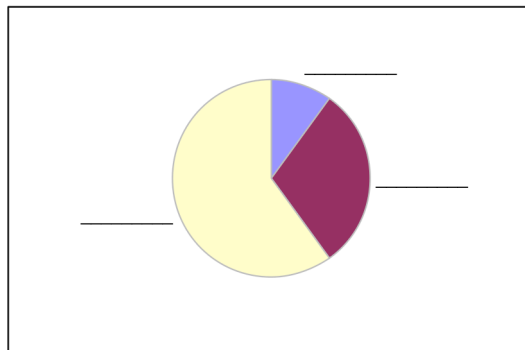
1. Below is some information about the tallest buildings in the world (completed by 2012).

Building	Country	Height, m	Floors	Use	Year built
Burj Khalifa	United Arab Emirates	828	163	Mixed	2010
Makkah Royal Clock Tower Hotel	Saudi Arabia	601	120	Hotel	2012
Taipei 101	Taiwan	508	101	Office	2004
Shanghai World Financial Center	China	492	101	Mixed	2008
International Commerce Center	China	484	108	Mixed	2010
Petronas Tower 1	Malaysia	452	88	Office	1998
Zifeng Tower	China	450	89	Mixed	2010
Willis (Sears) Tower	United States	442	108	Office	1974
Kingkey 100	China	442	100	Mixed	2012
Guangzhou International Finance Center	China	438	103	Mixed	2010

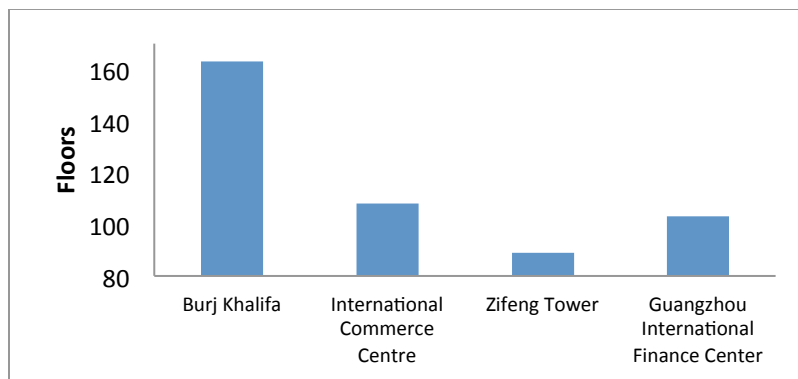
(a) What are the individuals in this data set?

(b) Identify the variables that were recorded, and indicate whether each one is categorical or quantitative.

(c) Here is a pie chart for the distribution of the variable “Use.” Fill in the blanks with the appropriate values of the variable.



(d) Below is a graph showing the total number of floors for the four buildings completed in 2010. What’s wrong with the way the information is presented in this graph?



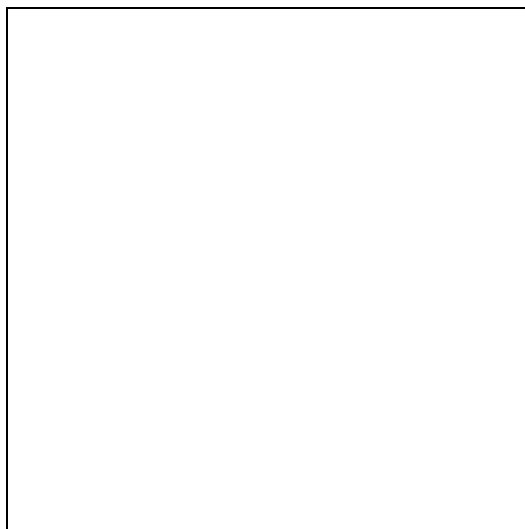
2. Researchers looking at the relationship between the type of college attended (public or private) and achievement gather the following data on 3265 people who graduated from college in the same year. The variable “management level” describes their job description 20 years after graduating from college.

		Type of College	
		Public	Private
Management level	High	75	107
	Medium	962	794
	Low	732	595

(a) Calculate the marginal distribution of management level in percents.

(b) Find the conditional distribution of management level for each college type, in percents.

(c) Sketch a segmented bar graph for the two conditional distributions in (b).



(d) Write a brief description of what the information in (b) and (c) tells you about the relationship between these variables.

3. Literary scholars sometimes use the distribution of word lengths in a work as a test of authenticity. Here are the word lengths for the first 25 words on a randomly-selected page from Toni Morrison's *Song of Solomon*.

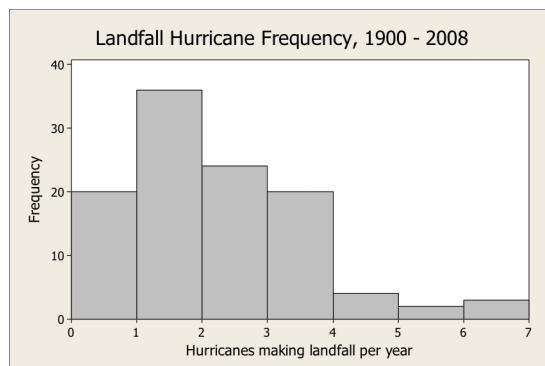
2 3 4 10 2 11 2 8 4 3 7 2 7  
5 3 6 4 4 2 5 8 2 3 4 4

- (a) Make a dotplot of these data.

\_\_\_\_\_

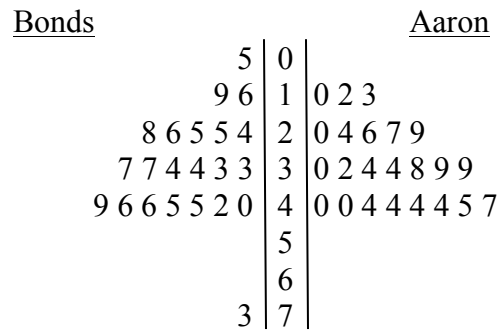
- (b) Describe the overall pattern of the distribution and any possible outliers.

4. The histogram below shows the number of hurricanes making landfall in the United States from 1900 to 2008. Describe the shape, center, and spread of the distribution.



5. On August 7, 2007 Barry Bonds hit his 756<sup>th</sup> home run, breaking the all-time career home run record, formerly held by Hank Aaron. Does that make Bonds a better home run hitter than Aaron? Let's compare their annual home run production over their entire careers. Below is a side-by-side stemplot. (Bonds played between 1986 and 2007. Aaron played between 1954 and 1978.)

Number of Home Runs per Year

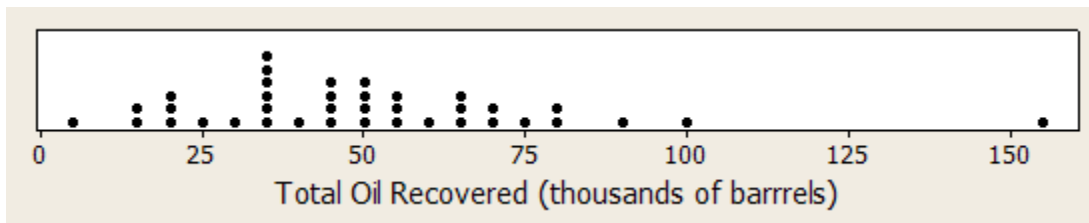


Key: 1|4 = 14 home runs

Use the plot to write a few sentences comparing Bonds and Aaron as home run hitters.

6. How much oil wells in a given field will ultimately produce is crucial information in deciding whether to drill more wells. Here are the estimated total amounts of oil recovered from 38 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousands of barrels. The data is provided in ascending order, along with a dotplot.

3	22	35	43	49	57	70	92
13	25	35	43	50	59	70	98
15	31	37	45	50	63	74	157
19	33	37	46	53	65	80	
21	35	38	48	56	66	82	



- (a) What measures would you use to describe the center and spread of these data? Justify your answer.
- (b) Find the five-number summary for these data.
- (c) Are there any outliers? Justify your answer.
- (d) Draw a boxplot of this distribution.

(e) For the oil well data on the previous page, how can you tell *without doing any calculations*, that the mean of these data is larger than the median?

7. Five students reported the amount of time (in minutes) they spent studying for an AP Statistics test the night before the test. Here are the results:

45    50    60    65    80

Calculate the mean and standard deviation of study time *using the formula* for each. Show your work!

8. Below are dotplots for three small datasets: A, B, and C. Without performing any calculations, rank the standard deviations of the datasets from lowest to highest. Justify your answer.

