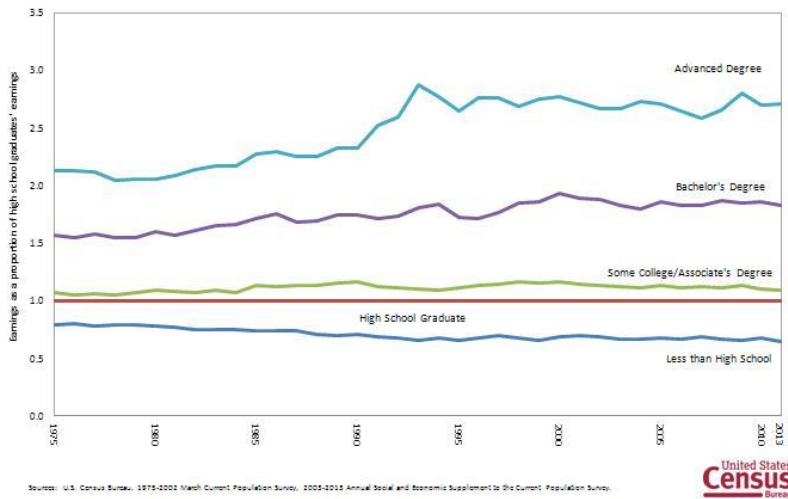MT 112, Homework #6, Spring 2020   Name _____

Instructions: Work problems on a separate sheet of paper and attach work to this page. You should show all work to receive full credit for problems. Questions with compact answers can be recorded directly on this page. Graphs and longer answers that won't fit here, indicate which page of the work the answer can be found on and be sure to clearly indicate it on the attached pages. You may use Excel to complete the problems, but then submit Excel files online.

1. Do a search for 'dubious statistics' on the web. Select one of the articles. Describe the dubious use of statistics in the cited research. What makes the use of stats dubious in the case you chose?

2. Create a pie chart for the following data. Create a summary table to start. A sample of 20 students who had recently taken elementary statistics yielded the following information on the brand of calculator owned (T=Texas Instrument, H=Hewlett Packard, C=Casio, S=Sharp). Be sure that your graph is properly labeled, displays the percents and has a title.

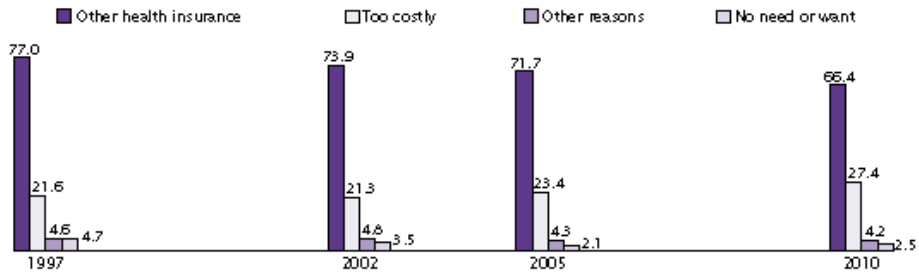| T	T	H	T	C	T	T	S	C	H
S	S	T	H	C	T	T	T	H	T

3. Below are a series of graphs. Interpret the graph. For time series graphs compare information from two points in time and say something about the information obtained from the graph. Is there a better way of displaying the data that might make the comparison you choose more salient? Explain what kind of graph you are looking at. If this qualifies as a 'bad graph' what is bad about it?

Figure 10: Average Earnings of Full-Time, Year-Round Workers as a Proportion of the Average Earnings of High School Graduates by Educational Attainment: 1975 - 2013

Sources: U.S. Census Bureau, 1975-2002 March Current Population Survey, 2003-2013 Annual Social and Economic Supplement to the Current Population Survey.

a)

Figure 3.
Most Common Reasons for Nonparticipation by Workers Who Chose Not to Participate in Jobs With Employer-Offered Health Insurance: 1997, 2002, 2005, and 2010
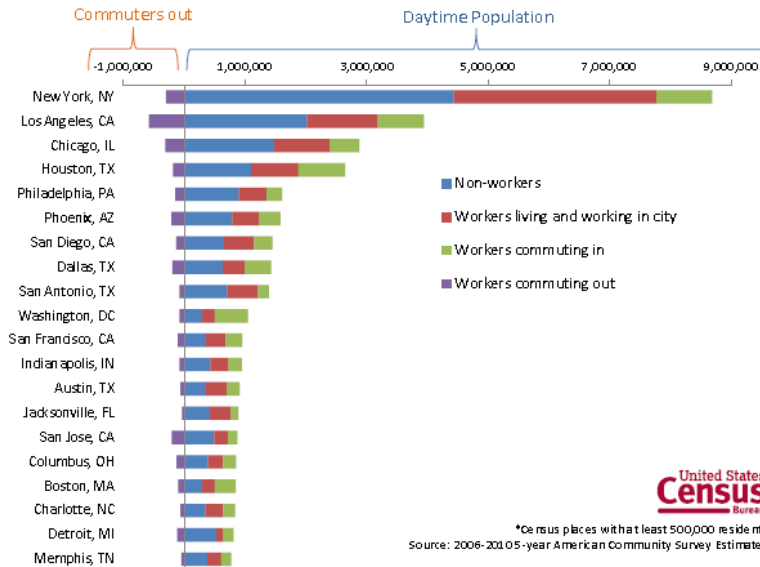
(In percent)

■ Other health insurance   □ Too costly   ▨ Other reasons   ▨ No need or want

1997
77.0
21.6
4.6
4.7

2002
73.9
21.3
4.8
3.5

2005
71.7
23.4
4.3
2.1

2010
66.4
27.4
4.2
2.5

Notes: "Other reasons" category is composed of write-ins. Categories do not sum to 100 percent since respondents may select more than one category.

b)

Source: U.S. Census Bureau, Survey of Income and Program Participation (1996, 2001, 2004, and 2008 Panels).

Cities among the Top 20 for Commuter-Adjusted Population*
2006-2010 5-year American Comunity Survey

Commuters out   Daytime Population

-1,000,000   1,000,000   3,000,000   5,000,000   7,000,000   9,000,000

New York, NY
Los Angeles, CA
Chicago, IL
Houston, TX
Philadelphia, PA
Phoenix, AZ
San Diego, CA
Dallas, TX
San Antonio, TX
Washington, DC
San Francisco, CA
Indianapolis, IN
Austin, TX
Jacksonville, FL
San Jose, CA
Columbus, OH
Boston, MA
Charlotte, NC
Detroit, MI
Memphis, TN

■ Non-workers
■ Workers living and working in city
■ Workers commuting in
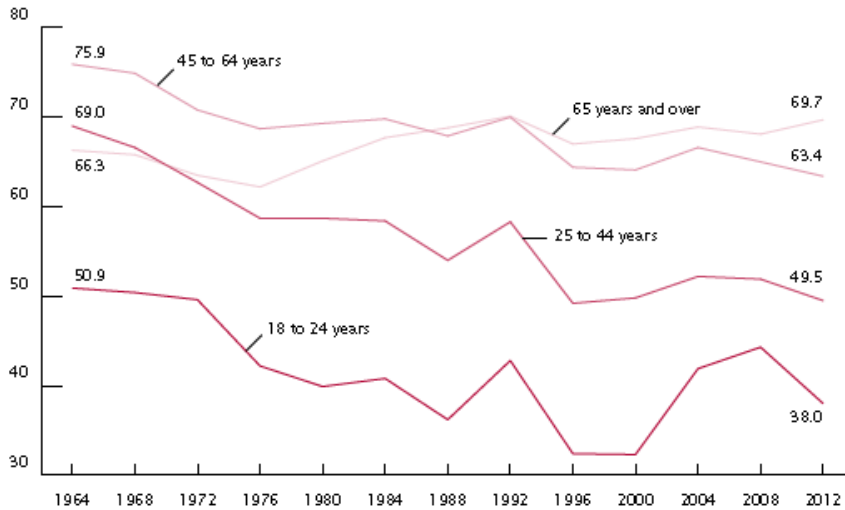■ Workers commuting out

United States Census Bureau

*Census places with at least 500,000 residents.
Source: 2006-2010 5-year American Community Survey Estimates.

c)
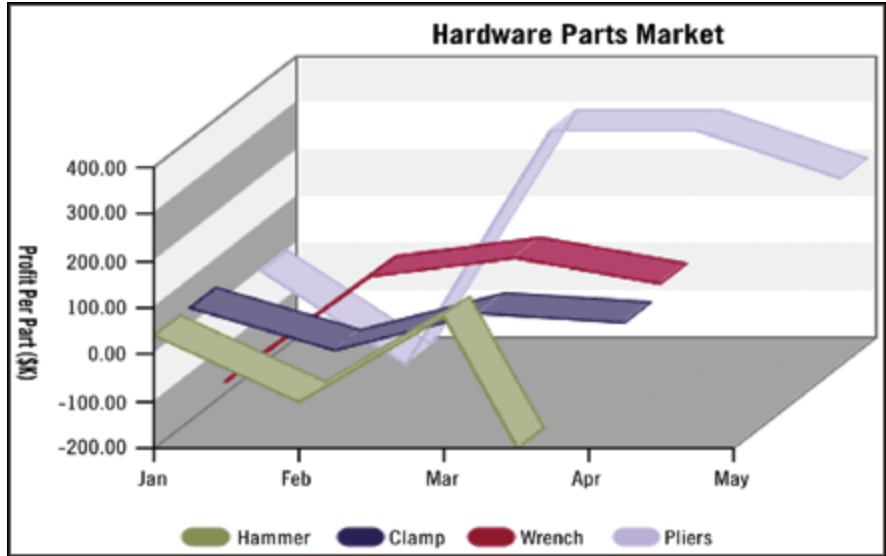
Figure 1.
Voting Rates Over Time for the Voting-Age Population: 1964–2012

(In percent)

80

75.9
45 to 64 years

69.0
65 years and over   69.7

66.3
63.4

70

60

50.9
25 to 44 years   49.5

18 to 24 years

38.0

50

40

30

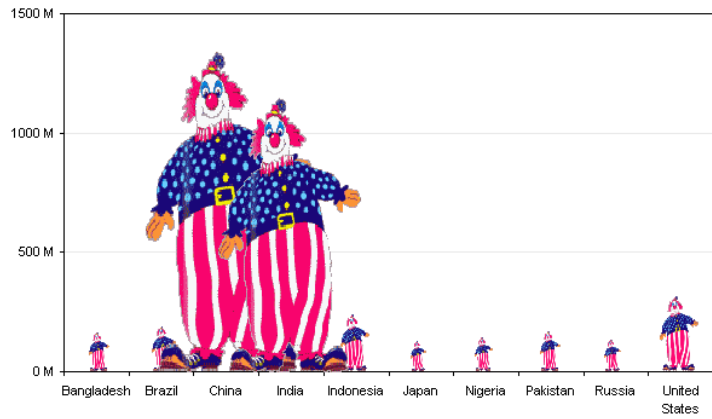1964 1968 1972 1976 1980 1984 1988 1992 1996 2000 2004 2008 2012

d)

Source: U.S. Census Bureau, Current Population Survey, Select Years.

e)

f)

g)

100 Most Active Tweeters
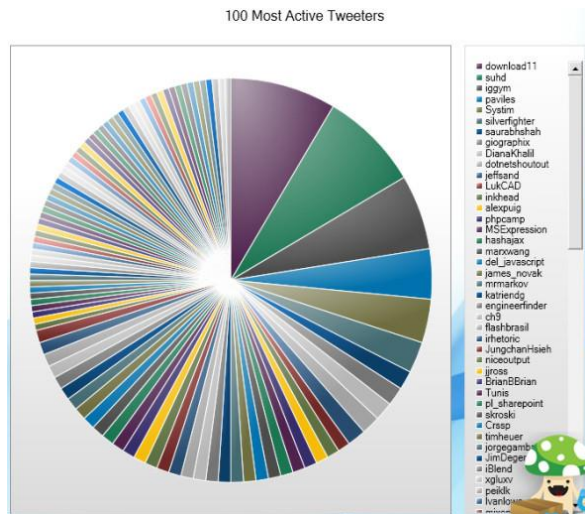

h)
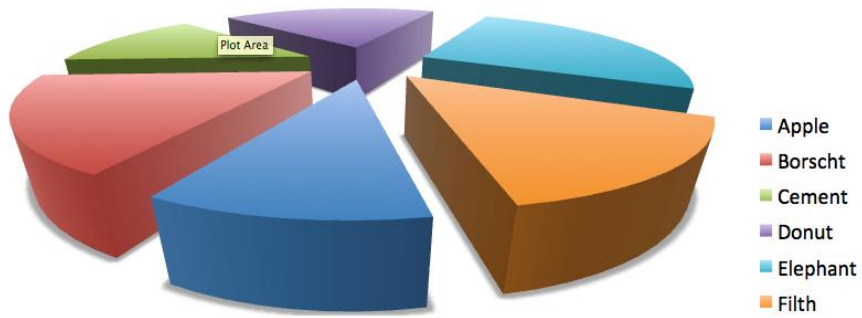
4. Suppose that you and your friends emptied your pockets of coins and recorded the year marked on each coin. The distribution of dates would be skewed to the left. Explain why?

5. Do you expect the distribution of the total player payroll for each of the 30 teams in Major League Baseball to be roughly symmetric, clearly skewed to the right, or clearly skewed to the left? Why?

6. For each of the histograms shown below, interpret the graphs as best as possible. Describe the distribution as symmetric or skewed (left or right), and if there are any outliers. Describe any other salient features of the graph.
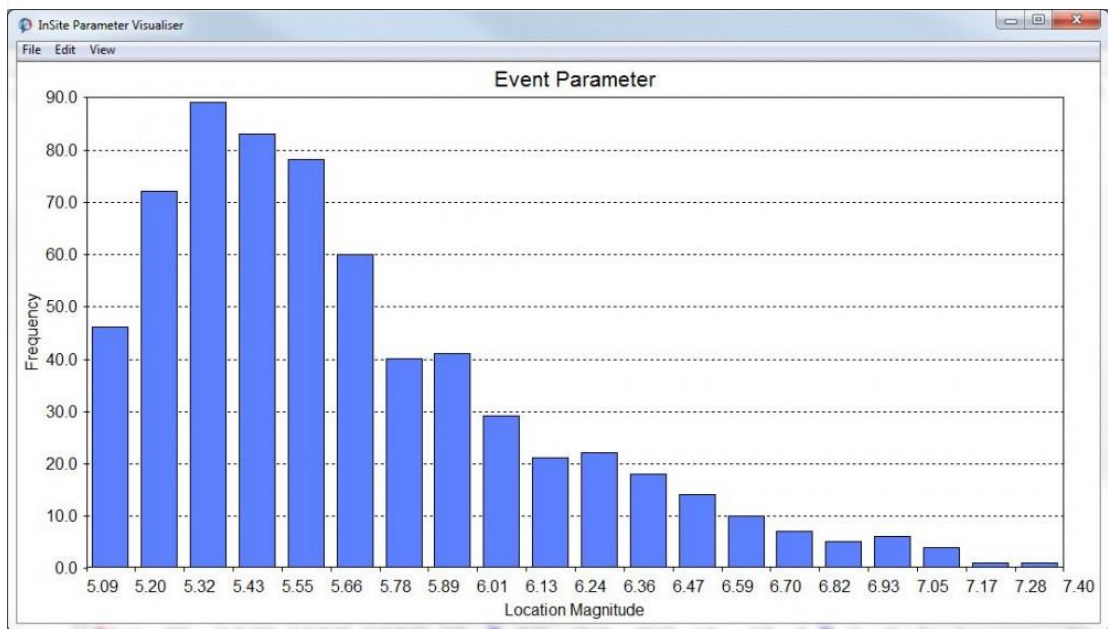

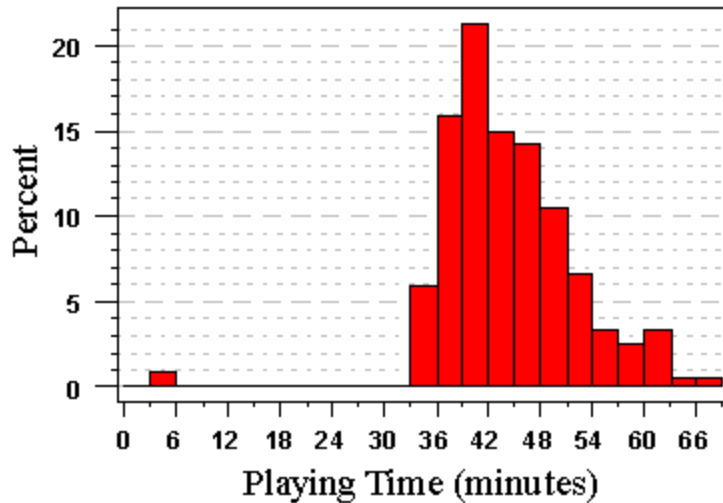Figure 3.4 Histogram (Ages of PSU Tenured Faculty)
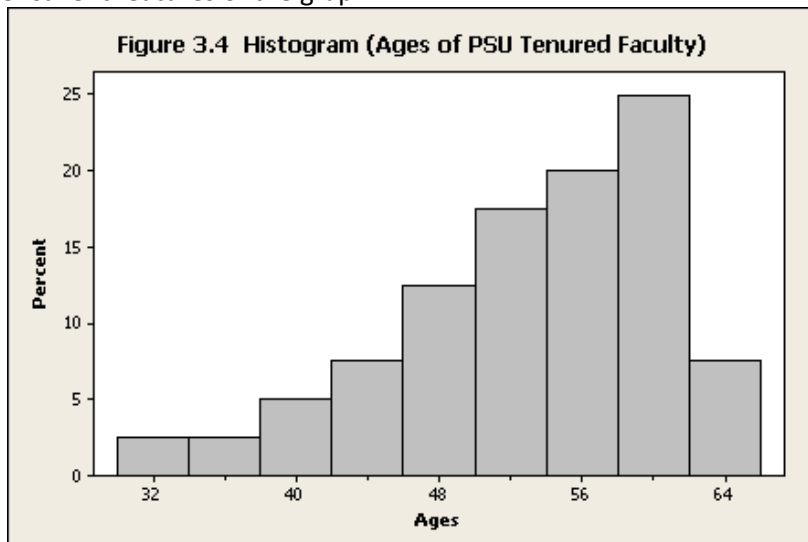
a)
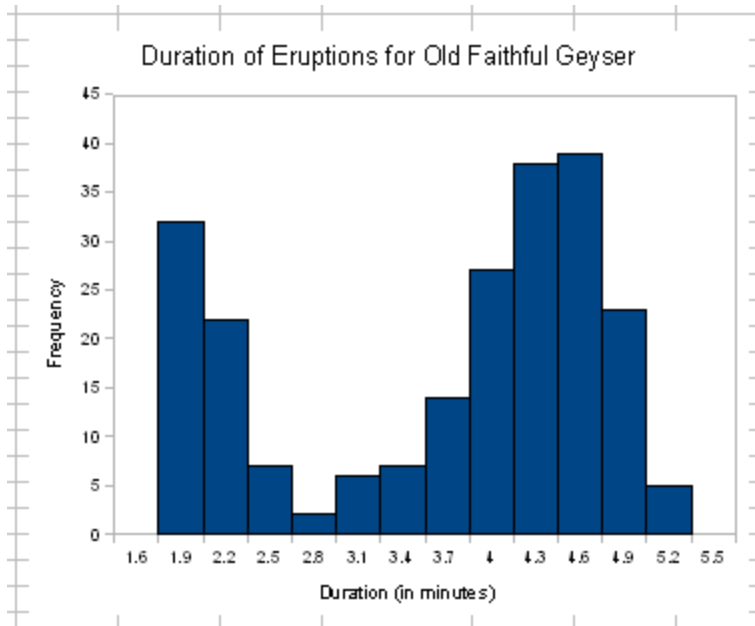
b)
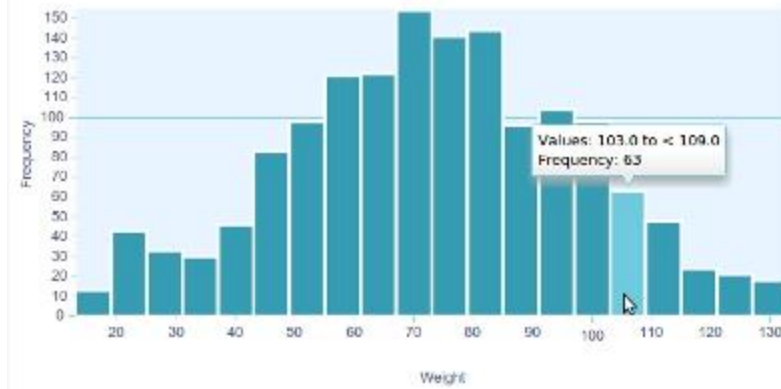
c)

d)

e)

7. The median income of US households in 2010 was about $49,455. Explain in plain language what the 'median income' is.

8. Echo Media reports that the average income for readers of the business magazine Forbes is $217,000. Is the median wealth of these readers greater or less than $217,000. Why?

9. Consumer Reports magazine presented the following data on the number of calories in a hotdog for each of 17 brands of meat hotdogs.

173	191	182	190	172	147	146	139	175
136	179	153	107	195	135	140	138	

Make a histogram and find the five-number summary. Use that data to create a box plot. How do the graphs compare? What data does the histogram represent better than the boxplot? What data does the boxplot represent better than the histogram? Then find the standard deviation.

10. A news article reported that of the 411 players on NBA rosters in February 1998, only 139 'made more than the league average salary' of $2.36 million. Was the figure reported the mean or the median? How can you tell?

11. Sketch density curves with the following properties: a) symmetric but not normal, b) skewed to the right.

12. Draw a histogram for the length of French movies shown below. With 8 bins.

81	90	90	92	93	94	95	96	96	102
103	105	106	111	113	116	119	122	122	123
125	125	128	137	158					

13. Convert the time data in the table above from minutes to hours (use 2 decimal places). Recreate your histogram. Then calculate the five-number summary, mean and standard deviation of the data. How does the unit conversion appear to affect the statistics? Do we get the same results if we convert the statistical summaries before or after calculating the conversions?

14. Use the data below in the table to calculate the weighted average.

Value	75	80	60	32	84	91	78
Weight	10%	20%	15%	5%	40%	3%	7%

15. Scores on the ACT college entrance exam vary normally with mean $\mu = 18$ and $\sigma = 6$. The range of reported scores is 1-36. a) What range contains the middle 95% of individual scores? b) If the ACT scores of 25 randomly selected students are averaged, what range contains the middle 95% of the averages?

16. The length of pregnancies from conception to birth varies according to a distribution which is approximately normal with mean 266 days and a standard deviation of 16 days. Use the 68-95-99.7 rule to answer the following: a) almost all (99.7%) pregnancies fall into what range? b) how long are the longest 2.5% of all pregnancies?

17. The Wechsler Adult Intelligence Scale (WAIS) is an IQ test. Scores on the WAIS for the 20-34 age group are approximately normal with a mean of 110 and standard deviation of 15. Scores for the 60-64 age group are approximately normal with a mean of 90 and standard deviation of 15. Sarah, who is 30, scores 130 on the WAIS. Her mother, who is 60, scores 110. a) Express both scores in terms of the number of standard deviations from the mean within their own age group; b) who scored higher relative to their age cohorts? Sarah or her mother?

18. It is possible to score higher than 800 on the SAT, but scores above 800 are reported as 800 (which is to say, a student doesn't need a perfect score to receive an 800). In 2011, the scores of college-bound senior men one the SAT Math test followed a normal distribution with mean of 531 and standard deviation of 119. What percentage of scores were above 800 (and so reported as 800)?

19. The quartiles of any distribution are the 25th and 75th percentiles respectively. About how many standard deviations from the mean are the quartiles for the normal distribution?

20. A variety of clubs and social organizations exist for people who are statistically exceptional in some way. High IQ societies are one such example (you can find a list and their standards for admission here: http://www.iqsocieties.com/), but clubs based on other exceptional characteristics exist as well, such a tall clubs (for example, http://tall.org/join-tci-main/how-to-join/). Why might the perception of high IQ societies and tall clubs differ in the general public? Can you find any other examples of statistically unusual traits that have formed social clubs or other organizations? Break down their admissions standards into standard deviations (for comparison, both Mensa and Tall Clubs International have a cut-off around 2 standard deviations above the mean). Why do you think this cut-off is commonly chosen?