

Table 3.4.13 – Percentage of Obese Adults by US County

FIPS Code	Percentage of Obese Adults 2016
1001	30.5
1003	26.6
1005	37.3
1007	34.3
1009	30.4
...	...

Using our new data, we generate the following map.

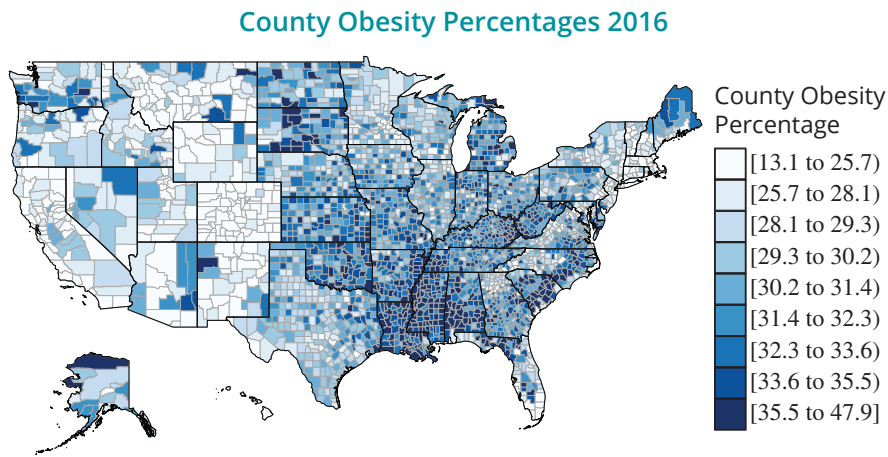


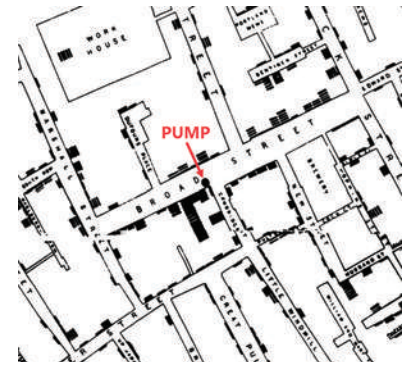
Figure 3.4.12

As you can see, Figure 3.4.12 is substantially different from Figure 3.4.11, and comes along with an entirely new set of conclusions. If we had used Figure 3.4.11 to come to a conclusion, we would have assumed that the Southwest and Northeast regions are the areas in the country that struggle with obesity the most. However, once we normalized our obesity variable by making it a ratio of the total county population, Figure 3.4.12 seems to suggest that it is actually the Southeast and the Midwest that struggle with obesity the most. The data shows that these regions have a higher percent of adults who are obese relative to their total population than the Northeast and Southwest do. This is a prime example of why it is necessary to normalize data when making comparisons.

3.4 Exercises

Basic Concepts

1. What is the main characteristic of data that a histogram reveals?
2. Describe the type of data that could be usefully described with a histogram.
3. True or false: A frequency distribution contains all of the information needed to construct a histogram.
4. List the important features to look for when studying a histogram.
5. Explain why the stem-and-leaf display is sometimes called a “hybrid graphical method.”
6. Identify the advantages of a stem-and-leaf display.
7. Consider the following data value: 39. What would be the stem and leaf for this value if we identified the stem as the tens digit? What would be the stem and leaf if we identified the stem as the hundreds digit?



Absence of Evidence is Not Evidence of Absence

During the London cholera outbreaks of the mid-1800s, thousands of people died within a relatively short period. At the time, the prevailing theory regarding how cholera was spread was called the miasma theory. It stated that the disease was spread through “bad air” that emanated from rotting organic matter. However, Dr. John Snow suspected that unsanitary water from the River Thames was the true culprit. Unfortunately, germ theory had not been developed yet, so Dr. Snow didn’t fully understand how the alternative transmission method worked. In 1854, Dr. Snow utilized sampling and data visualization to illustrate that most of the cholera outbreaks happening at the time were occurring in houses that were close to the water pump on Broad Street. Still, the skeptics endured. However, even though his examination of the water was absent of evidence for harmful microbes, that does not mean that the microbes themselves were absent. Over a decade later, Louis Pasteur would officially propose germ theory, vindicating the work of Dr. Snow.

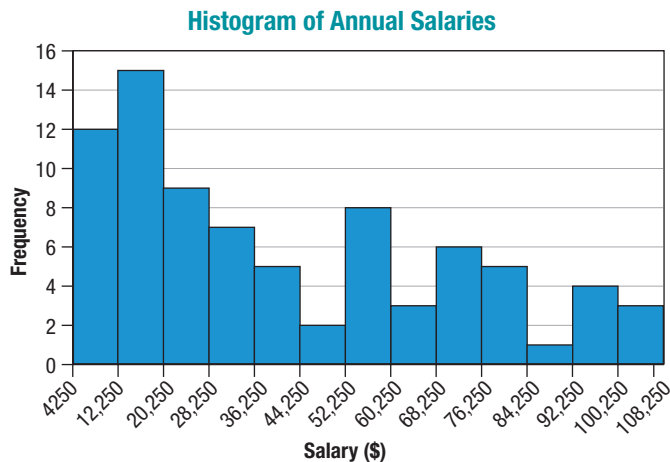
8. When constructing a stem-and-leaf display, how do you determine which part to make the stem and which part to make the leaf?
9. What is an ordered array?
10. What are some advantages of the ordered array?
11. What is a dot plot?
12. What are some advantages of using a dot plot?
13. How can the most frequently occurring value be identified by studying a dot plot?
14. Why is it important to plot time series data?
15. The time variable is always graphed on which axis?
16. Identify a variation on a time series plot that can make the data more visually interesting.

Exercises

17. The closing prices (in dollars) for selected stocks trading on the New York Stock Exchange and NASDAQ Exchange were as follows.

Closing Prices	
Stock	Closing Price (\$)
Citigroup (C)	34.70
Pfizer (PFE)	22.34
Herbalife (HLF)	69.72
JP Morgan Chase (JPM)	44.34
Intel (INTC)	28.07
WalMart (WMT)	60.67
Microsoft (MSFT)	31.52
PepsiCo (PEP)	66.17
General Motors (GM)	24.81
Verizon Communications (VZ)	37.66
Southwest Airlines (LUV)	8.31
Sprint Nextel (S)	2.76
Yahoo! Inc (YHOO)	15.07
International Business Machines (IBM)	205.47

- a. Construct a frequency distribution for the closing prices.
 - b. Construct a histogram for the closing prices.
18. A sample of 80 laborers is selected from a large city and their annual salaries are determined. The following histogram summarizes the data.



- a. What is the level of measurement of the variable?
 - b. How many of the laborers earn at least \$36,250?
 - c. What percent of the laborers earn at most \$28,249?
 - d. What percent of the laborers earn at least \$84,250?
19. A nutritionist is interested in knowing the percent of calories from fat which Americans consume on a daily basis. To study this, the nutritionist randomly selects 25 Americans and evaluates the percent of calories from fat consumed in a typical day. The results of the study are as follows.

Calories from Fat per Day (%)				
34	18	33	25	30
42	40	33	39	40
45	35	45	25	27
23	32	33	47	23
27	32	30	28	36

- a. Construct a frequency distribution for the percent of calories from fat.
 - b. Construct a relative frequency distribution for the percent of calories from fat.
 - c. Construct a histogram of the relative frequency distribution.
 - d. Comment on any information about the percent of calories from fat consumed by the participants in the study which you were able to ascertain by examining the distributions and the histogram.
20. Consider the assets (in billions of dollars) of the 10 largest commercial banks listed in the following table.

Assets (Billions of Dollars)				
216.9	138.9	115.5	110.3	103.5
98.2	76.4	64.0	53.5	49.0

- a. Construct a frequency distribution for the assets (in billions of dollars) of the 10 largest commercial banks.
- b. Construct a relative frequency distribution for the assets (in billions of dollars) of the 10 largest commercial banks.
- c. Construct a histogram of the relative frequency distribution.
- d. Comment on any information about the assets (in billions of dollars) of the 10 largest commercial banks which you were able to ascertain by examining the distributions and the histogram.

21. Fifty hospitals in a western state were polled as to their basic daily charges for a semi-private room. The results are listed in the following table, rounded to the nearest dollar.

Daily Charges for Semi-Private Rooms (Dollars)									
125	135	148	156	248	215	156	148	135	149
178	156	135	125	214	256	258	265	156	148
123	147	189	199	189	248	215	259	158	235
268	269	158	198	147	258	269	239	288	199
179	179	189	169	258	178	257	249	259	259

- What level of measurement do the data possess?
 - Construct a stem-and-leaf display for the data using the tens digits as the stems.
 - Comment on the shape of the distribution.
22. The data in the following table are the toxic emissions (in thousands of tons) for 10 states in the United States.

Toxic Emissions (Thousands of Tons)									
206	147	441	128	127	133	422	152	114	134

Source: Toxics in the Community, U.S. Environmental Protection Agency

- Construct a stem-and-leaf display for the data using the hundreds digits as the stems.
 - Comment on any information about the toxic emissions (in thousands of tons) of the 10 states that you were able to ascertain by examining the stem-and-leaf display.
23. Consider the following highway miles per gallon for 19 selected models of mini-compact, sub-compact, and compact cars.

Miles per Gallon									
26	46	36	31	28	28	27	38	42	36
37	33	23	29	37	34	29	40	28	

- Construct a stem-and-leaf display for the data.
 - Comment on any information about the highway mpg of the selected models which you were able to ascertain by examining the stem-and-leaf display.
24. An instructor is interested in comparing exam scores for fraternity and non-fraternity males in her class. Meaningful comparisons between two sets of data can be made using a side-by-side stem-and-leaf display. To illustrate this, note the following display summarizing the scores.

Leaf (Non-Fraternity)	Stem	Leaf (Fraternity)
	0	9
2	1	4 0 8
	2	5 7 9 4 5 5 1
3 9	3	2 6 6 9 7 7 3 2 1 6 0
	4	2 7 5
5 6 4 8 9 9 0 2	5	4 7 6 7
4 4 7 8 1 0 3 2 2 6 8 9	6	6 8 9 9 5
5 4 7 8 4 3 8 8 9 1	7	3 4 2 7 8 6 7 4 3
2 9 7 4	8	4 5 3 8 9 9 6 4 2 1 1 4 5
4 2	9	4 3 5 1 6 7 7 0 3
Key: Non-Fraternity 2 9 = 92		
Fraternity 9 4 = 94		

- a. What level of measurement do the data possess?
 - b. Based upon the stem-and-leaf display, compare the two groups. Think of the several ways in which this can be done.
 - c. Suppose that 60% is considered a passing score on the exam. What percent of the fraternity students passed the exam? Non-fraternity students?
 - d. If someone scores 90 or higher on the exam, they will be exempt from taking the next exam. What percent of the fraternity students will be exempt from taking the next exam? Non-fraternity students?
25. Microsoft's consumer PC sales growth for the last 16 quarters are listed in the following table. Examine the data (sales growth, in percentages) and answer the following questions.

PC Sales Growth			
Quarter	Sales Growth (%)	Quarter	Sales Growth (%)
1	20	9	20
2	24	10	19
3	22	11	33
4	19	12	37
5	23	13	24
6	27	14	10
7	16	15	0
8	10	16	-4

Source: Citi Investment Research and Analysis IDC, Company Reports, May 2011

- a. Construct an ordered array of the data in rank order.
 - b. What conclusions can you make based on the ordered array?
26. *Fortune* magazine publishes a list of the top 100 best companies to work for. For the top 10 companies on this list, the average annual employee salaries are given in the following table (in thousands of dollars).

Average Salaries (Thousands of Dollars)									
121	122	136	74	118	101	114	61	95	132

- a. Construct a stem-and-leaf display for the data using the tens digits as the stems.
 - b. Comment on any information about the average annual salaries (in thousands of dollars) of the top 10 companies which you were able to ascertain by examining the stem-and-leaf display.
 - c. Construct an ordered array of the average annual salaries in rank order.
 - d. Does the ordered array provide any additional insight into the nature of the data?
27. Construct a dot plot for the following set of data.

23	19	15	20	17
16	18	14	23	22
19	23	19	16	25
17	20	21	23	24

28. Listed in the following table is the number of passing attempts per game by Super Bowl champion Aaron Rodgers in the 2010 NFL season. Construct a dot plot for the data set.

Passing Attempts by Aaron Rodgers				
31	29	45	17	46
33	34	34	34	31
35	30	11	37	28

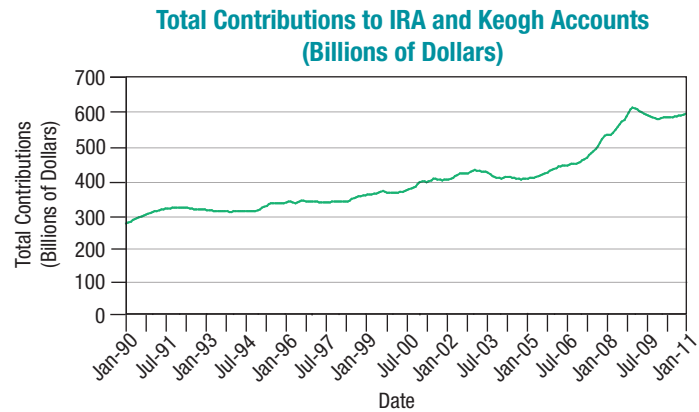
Source: ESPN

29. The following table contains the average monthly energy consumption (in kilowatt-hours) by household for nine South Atlantic states in 2007. Construct a dot plot for the data set.

Monthly Energy Consumption (Kilowatt-Hours)	
773	1143
960	1210
1163	1207
1171	1138
1086	

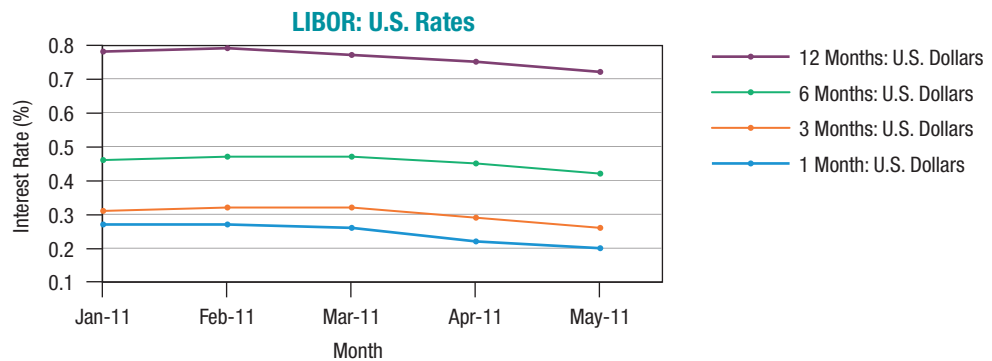
Source: U.S. Energy Information Administration

30. The following line graph displays the total IRA and Keogh accounts (in billions of dollars) in the U.S., charted from June 1990 to June 2011.



Source: www.economagic.com

- What conclusions can you make regarding the total contributed to the accounts?
 - Are the data time series data?
 - If the data are time series data, is the series stationary or nonstationary?
31. The following chart contains LIBOR (which stands for London Interbank Offered Rate) data for January 2011 through May 2011. LIBOR is the average interest rate that banks in London charge when lending funds to other banks. The line graphs in the figure represent 1-month, 3-month, 6-month, and 12-month interest rates.



Source: www.economagic.com

- Examine the chart and discuss the data. What conclusions can you make?
- If the data are time series data, is it a stationary or nonstationary time series? Explain your reasoning.

32. The Gallup Poll frequently obtains responses to the question, *At the present time, do you think religion as a whole is increasing its influence on American life or losing its influence?* The percentage of the respondents who answered “increasing” is given below for various polls.

Survey Responses										
Year	2001	1995	1992	1991	1990	1988	1986	1984	1982	1980
Percent	71	38	27	27	33	36	48	42	41	35
Year	1978	1977	1975	1974	1970	1969	1968	1965	1962	1957
Percent	37	37	39	31	14	14	19	33	45	69

- What level of measurement do the responses to the question possess?
 - Construct a time series plot for the data.
 - What conclusions can you make from the plot?
33. The following table gives the number of immigrants (in thousands) and the average annual immigration rate per 1000 people in the U.S. population for the decade ending in the year given.

Annual Immigration (per 1000 People)											
Year	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Number	3688	8795	5736	4107	528	1035	2515	3322	4493	7338	9095
Rate	5.3	10.4	5.7	3.5	0.4	0.7	1.5	1.7	2.1	2.9	3.2

- What levels of measurement do the three variables in this exercise possess?
- Construct a time series plot of the number of immigrants per decade.
- Find the percent change in the number of immigrants from the decade ending in 1900 to the decade ending in 2000.
- Find the percent change in the average annual immigration rate per 1000 people in the U.S. population from the decade ending in 1900 to the decade ending in 2000. Compare your answer to that which you obtained in part c. Can you explain why these answers are different?

3.5 Analyzing Graphs

Graphs that help us visualize data can either be enlightening, in the sense that they give us insight and understanding of a set of data, or misleading, either intentionally or unintentionally. When you see graphs in the media, you need to be cautious to ensure the data has been accurately represented by the graph. This section will help you analyze graphs for accuracy and appropriate presentation of the given information. Here are a few key ideas to consider when interpreting information displayed in graphical form.

Graph Labeling

Every graph should be properly labeled with an appropriate title that tells you what type of information is being displayed. Also, if the graph has a horizontal and vertical axis, these should be labeled and should include the unit of measurement when necessary for the understanding of the data. For example, in Figure 3.5.1 shown below, the title does not provide enough information about the data. Why were those countries chosen? Do they have relatively high or low prison populations compared to the rest of the world? Furthermore, we do not know whether this information is relevant to modern times. Is this data for a specific year? The countries are labeled along the horizontal axis, but note that the vertical axis is just labeled *Population*. We have no idea what the values along the vertical axis