Objectives:

1.1 Displaying Distributions with Graphs

1.2 Describing Distributions with Numbers

Summary of Commands

See Chapter 1 Appendix for detailed instructions

- Graph → Bar Chart (for Bar chart and Pareto chart)
- Graph → Pie Chart
- Graph → Histogram
- Graph → Stem-and-Leaf
- Stat → Basic Statistics → Display Descriptive Statistics
- Graph → Boxplot
- Graph → Time Series Plot

Problems 1 (**1.22 Garbage is big business. Data Set: GARBAGE**)
The formal name for garbage is "municipal solid waste." In the United States, approximately 254 million tons of garbage are generated in a year. Below is a breakdown of the materials that made up American municipal solid waste in 2007.[15]

Material	Weight (million tons)	Percent of total
Food scraps	31.7	12.5
Glass	13.6	5.3
Metals	20.8	8.2
Paper, paperboard	83.0	32.7
Plastics	30.7	12.1
Rubber, leather, textiles	19.4	7.6
Wood	14.2	5.6
Yard trimmings	32.6	12.8
Other	8.2	3.2
Total	254.1	100.0

(a) Make a bar graph of the percents. The graph gives a clearer picture of the main contributors to garbage if you order the bars from tallest to shortest.
(b) Also make a pie chart of the percents. Comparing the two graphs, notice that it is easier to see the small differences among "Food scraps," "Plastics," and "Yard trimmings" in the bar graph.

Problems 2 (1.27 U.S. unemployment rates.
Data Set: UNEMPLOYMENT) An unemployment rate is the number of people who are not working but who are available for work divided by the total number of people in the workforce, expressed as a percent. The table below gives the U.S. unemployment rates for each state as of August 2008.[21]

State	Rate	State	Rate	State	Rate
Alabama	4.9	Louisiana	4.7	Ohio	7.4
Alaska	6.9	Maine	5.5	Oklahoma	4.0
Arizona	5.6	Maryland	4.5	Oregon	6.5
Arkansas	4.8	Massachusetts	5.3	Pennsylvania	5.8
California	7.7	Michigan	8.9	Rhode Island	8.5
Colorado	5.4	Minnesota	6.2	South Carolina	7.6
Connecticut	6.5	Mississippi	7.7	South Dakota	3.3
Delaware	4.9	Missouri	6.6	Tennessee	6.6
Florida	6.5	Montana	4.4	Texas	5.0
Georgia	6.3	Nebraska	3.5	Utah	3.7
Hawaii	4.2	Nevada	7.1	Vermont	4.9
Idaho	4.6	New Hampshire	4.2	Virginia	4.6
Illinois	7.3	New Jersey	5.9	Washington	6.0
Indiana	6.4	New Mexico	4.6	West Virginia	4.1
Iowa	4.6	New York	5.8	Wisconsin	5.1
Kansas	4.7	North Carolina	6.9	Wyoming	3.9
Kentucky	6.8	North Dakota	3.6		

(a) Construct a histogram of these rates.
(b) Prepare a stemplot of the rates.

Problems 3 (1.35 Is the supply adequate? Data Set: OILWELLS)
How much oil the wells in a given field will ultimately produce is key information in deciding whether to drill more wells. Here are the estimated total amounts of oil recovered from 64 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousands of barrels:[27]

21.7	53.2	46.4	42.7	50.4	97.7	103.1	51.9	43.4	69.5
156.5	34.6	37.9	12.9	2.5	31.4	79.5	26.9	18.5	14.7
32.9	196.0	24.9	118.2	82.2	35.1	47.6	54.2	63.1	69.8
57.4	65.6	56.4	49.4	44.9	34.6	92.2	37.0	58.8	21.3
36.6	64.9	14.8	17.6	29.1	61.4	38.6	32.5	12.0	28.3
204.9	44.5	10.3	37.7	33.7	81.1	12.1	20.1	30.5	7.1
10.1	18.0	3.0	2.0						

Graph the distribution (histogram).

Problems 4 (**1.59 Recoverable oil. Data Set: OILWELLS**) The estimated amounts of recoverable oil from 64 oil wells in the Devonian Richmond Dolomite area of Michigan are given **Exercise 1.35**.
(a) Find the mean and the standard deviation.
(b) Find the five-number summary.
(c) Draw a boxplot.
(d) What is the value of IQR? What are the values of the three outliers, according to the 1.5*IQR rule?

Problems 5 (**1.38 A multimillion-dollar business is threatened. Data Set: BERINGSEAFISH**) Bristol Bay of Alaska has typically produced more wild-caught sock-eye salmon, *Oncorhynchus nerka,* than any other region in the world. In good years, the runs typically exceed 50 million fish. The sockeye salmon industry here provides thousands of jobs and generates millions of dollars per year.[28] Here are the numbers of sockeye salmon in runs at Bristol Bay between 1988 and 2007:[29]

Year	Runs (millions)	Year	Runs (millions)	Year	Runs (millions)	Year	Runs (millions)
1988	22.9	1993	52.7	1998	18.1	2003	26.5
1989	44.5	1994	50.3	1999	39.5	2004	43.5
1990	47.1	1995	60.8	2000	28.4	2005	39.3
1991	42.0	1996	37.0	2001	22.0	2006	43.1
1992	45.6	1997	18.8	2002	17.2	2007	44.3

(a) Make a graph to display the distribution of salmon run size, then describe the pattern and any striking deviations that you see.
(b) Make a time plot of run size and describe its pattern. As is often the case with data measured at specific time intervals, a time plot is needed to understand what is happening.

Problems 6 (**1.52 GDP growth in 120 countries. Data Set: COUNTRIES120**) The gross domestic product (GDP) of a country is the total value of all goods and services produced in the country. It is an important measure of the health of a country's economy. For this exercise, you will analyze the growth in GDP, expressed as a percent, for 120 countries.[33]

(a) Compute the mean and the standard deviation.
(b) Which two countries are extreme outliers for this variable?
(c) Recompute the mean and standard deviation without the outliers. Explain how the mean and standard deviation changed when you deleted the outliers.

Problems 7 (**1.53 Use the resistant measures for GDP. Data Set: COUNTRIES120**) Repeat parts (a) and (c) of the previous exercise using the median and the quartiles. Explain how the median and quartiles changed when you deleted the outliers.

Problems 8 (**1.71 $\bar{x}$ and *s* are not enough. Data Set: ABDATA**)
The mean $\bar{x}$ and standard deviation *s* measure center and spread but are not a complete description of a distribution. Data sets with different shapes can have the same mean and standard deviation. To demonstrate this fact, find $\bar{x}$ and *s* for these two small data sets. Make a stemplot of each. Do they have the same distributions?

Data A:	9.14	8.14	8.74	8.77	9.26	8.10
	6.13	3.10	9.13	7.26	4.74	
Data B:	6.58	5.76	7.71	8.84	8.47	7.04
	5.25	5.56	7.91	6.89	12.50	

Problem 9 (1.128 Evaluating the improvement in a product. Data Set: CORN) Corn is an important animal food. Normal corn lacks certain amino acids, which are building blocks for protein. Plant scientists have developed new corn varieties that contain these amino acids. To test a new corn as an animal food, a group of 20 one-day-old male chicks was fed a ration containing the new corn. A control group of another 20 chicks was fed a ration that was identical except that it contained normal corn. Here are the weight gains (in grams) after 21 days:
(a) Compute five-number summaries for the weight gains of the two groups of chicks. Then make boxplots to compare the two distributions. What do the data show about the effect of the new corn?
(b) The researchers actually reported means and standard deviations for the two groups of chicks. What are they? How much larger is the mean weight gain of chicks fed the new corn?

Normal corn				New corn			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	477	392
345	455	360	431	430	339	410	326