

Chapter 3: Problem Set

Numbers with an asterisk* have solutions in the back of the book.

Frequency Distributions and Histograms (3.1 & 3.2)

1.* **American Made Cars:**

To the right is a frequency distribution for the estimated miles per gallon (MPG) for 518 different models of American-made cars.

American-Made Cars	
MPG	Frequency
9 - 12	56
13- 16	177
17 - 20	167
21 - 24	72
25 - 28	34
29 - 32	12

- Identify the following: class midpoints, class boundaries, and class width.
- Create a histogram from the table.
- Create a relative frequency table from the frequency table.
- Create a cumulative and relative-cumulative frequency table.
- Estimate the mean MPG for the American-made cars included in this set.

2. **Japanese Made Cars:**

To the right is a frequency distribution for the estimated miles per gallon (MPG) for 281 different models of Japanese-made cars.

Japanese-Made Cars	
MPG	Frequency
9 - 12	8
13- 16	40
17 - 20	97
21 - 24	92
25 - 28	28
29 - 32	12
33 - 36	2
37 - 40	0
41 - 44	1
45 - 48	1

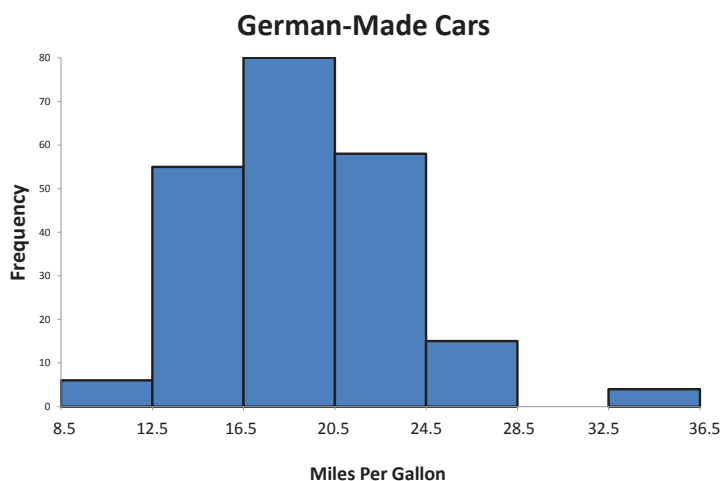
- Identify the following. Class midpoints, class boundaries, and class width.
- Create a histogram from the table. What makes this histogram difficult to create and read?
- Create a relative frequency table.
- Create a cumulative and relative cumulative frequency table.
- Estimate the mean MPG for Japanese-made cars included in this set.

3. **Comparing Distributions:** Use your answers from the previous two problems to answer the following questions regarding the difference in distributions for MPG between cars made in America and cars made in Japan.

- Do the distributions appear to be normal, skewed, uniform, or multi-modal?
- Which country appears to make the most fuel efficient cars?
- Which representation (frequency distribution, histogram, relative frequency distribution, cumulative frequency distribution, or relative-cumulative frequency) best allows you to compare the fuel-efficiency of the American-made and Japanese-made cars and why?

4.* **German-Made Cars:**

To the right is a histogram for the estimated miles per gallon (MPG) for 218 different models of German-made cars. Use this data (in graphical form) to compare the MPG distribution of German-made cars with those made in America and Japan (from problems 1 and 2).



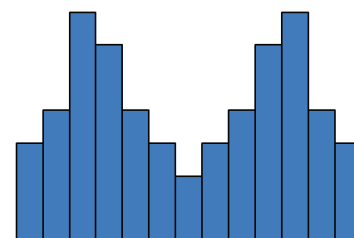
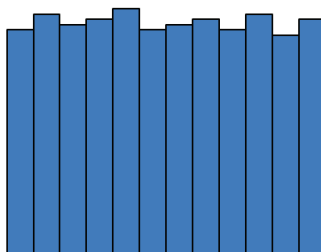
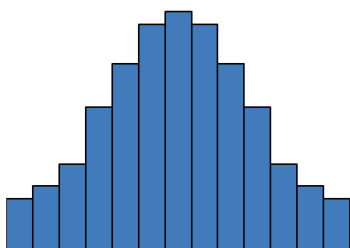
5.* **Estimating a Mean:** Estimate the mean score from the frequency distribution below.

Score	Frequency
60-64	8
65-69	4
70-74	2
75-79	1
80-84	1

6. **Estimating a Mean:** Estimate the mean score from the frequency distribution below.

Score	Frequency
61-70	2
71-80	4
81-90	5
91-100	1

- 7.* **Shapes of distributions:** Determine whether the following data, if put into a histogram would result in a distribution that is normal, skewed, bimodal, or uniform.
- The shoe sizes of 200 randomly selected adults.
 - The shoe sizes of 200 randomly selected adult men.
 - The shoe sizes of 200 randomly selected males (children included).
 - The last digit in the phone numbers of 1000 randomly selected people from across the country.
8. **Shapes of distributions:** Determine whether the following data, if put into a histogram would result in a distribution that is normal, skewed, bimodal, or uniform.
- The incomes of people living in an urban area.
 - The age of death for U.S. inhabitants (the mean is about 78 years).
 - The numbers produced by a random number generator.
 - The approval ratings (from a scale of 1 to 10) of a newly elected president.
9. **Shapes of Distributions - Variation:** Consider the three histograms below.



- Identify the distributions.
 - Assume all three histograms have the same classes and units. Which distribution would have the largest standard deviation and which one would have the smallest standard deviation?
- 10.* **Grading on a Curve:** Below are the results of 17 test scores from a 100 point exam. Based on the 90, 80, 70, 60, grading scale this results in 4 F's, 4 D's, 1 C, 3 B's, and 5 A's. Students are upset because nearly half of the scores (8) are D's and F's. The students ask if the instructor is going to **curve** the grading. There is no official definition of grading on the curve but the general idea is that when the process is complete, the grade distribution will be approximately bell shaped. Try to create a grading scale that curves these grades (the class widths do not have to be equal). Who are the winners? Who are the losers? Does your grading scale seem fair?

index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
score	42	48	48	55	60	64	64	66	75	82	85	88	92	95	96	98	100

Other Graphics (3.3)

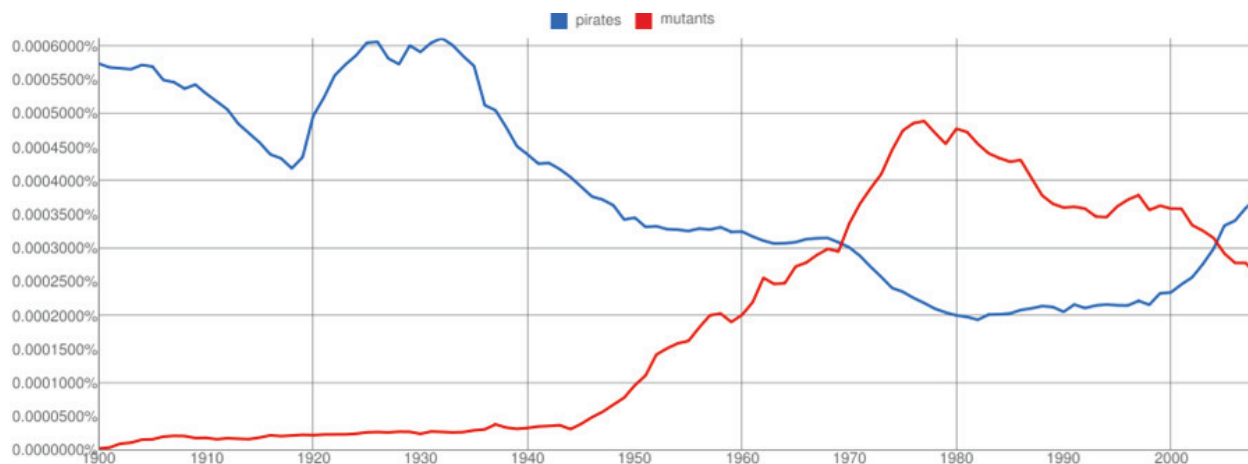
11. **Bar Graph - Distorting Data:** Consider the data in the table (from Gallup) representing the percentage of Americans who are in favor of legalizing marijuana divided by region. Make a bar graph that depicts these differences where the y -axis starts at zero. Make another where the y -axis starts at 25. How does this affect your interpretation of the data?

Should marijuana be legal?	
Region	Percent who said yes
South	29
Midwest	33
East	34
West	47

12. Time-Series Graph: Google's Ngram Viewer

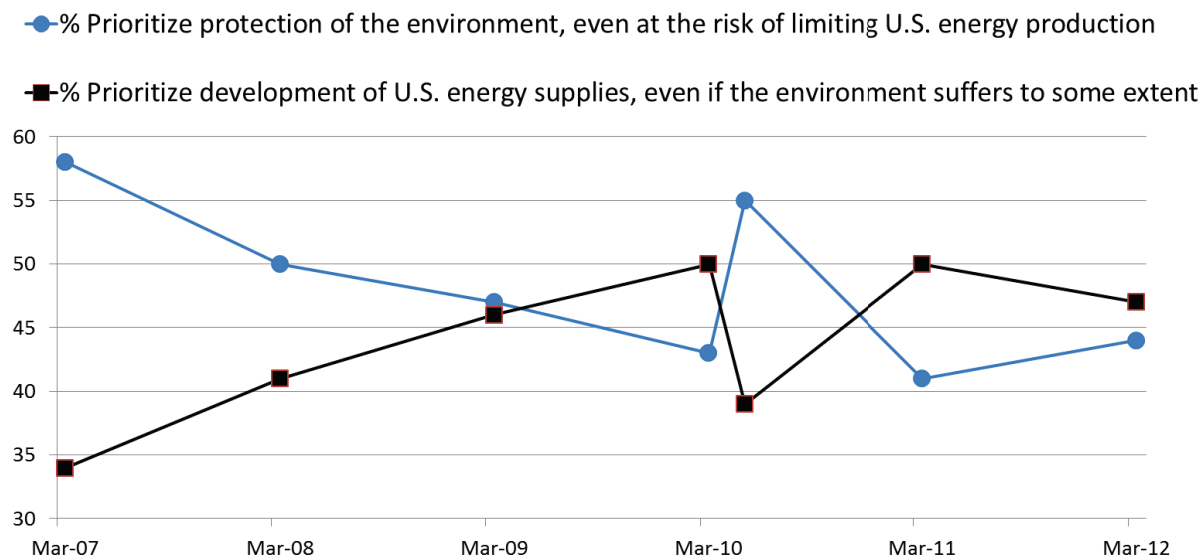
Interesting time-series graphs can be generated using Google's Ngram Viewer (<http://ngrams.googlelabs.com>). This site automatically generates a time-series graph of the relative frequencies for any word or words (ngrams) which appear in books from their digital library. You get to choose the words, the years, and the specifics of the library to search and the site creates the graph.

- (a)* Below is a time-series graph for relative frequency of the words *pirates*, and *mutants* during the years 1900-2008. Tell the story of *pirates* -vs- *mutants*.



- (b) Go to <http://ngrams.googlelabs.com> and create a time-series graph for the use of any two words or Ngrams and tell the story of the usage of those two words.

- 13.* **Time-Series, Gallup Opinion Poll:** Below is a time-series graph generated by a survey conducted by the Gallup Organization. ‡ It depicts whether the U.S. population favors environmental issues or energy production over the course of 5 years. What is the story that this graph tells? What *trick* did I use to make the differences appear more extreme?



14. Time-Series, Unemployment:

The table to the right gives the national unemployment rates from the Bureau of Labor and Statistics (www.bls.gov) for the 13 months starting in July 2008 (right at the beginning of the U.S. financial crisis) when unemployment rates started to drastically increase.

Make a time-series graph for this data where the y-axis starts at zero and one where the y-axis starts at 5.0. How does this affect the interpretation of the data?

National Unemployment	
Month	Rate (%)
July 2008	5.8
August 2008	6.1
September 2008	6.2
October 2008	6.6
November 2008	6.8
December 2008	7.3
January 2009	7.8
February 2009	8.2
March 2009	8.6
April 2009	8.9
May 2009	9.4
June 2009	9.4
July 2009	9.4

‡<http://www.gallup.com/poll/153404/Americans-Split-Energy-Environment-Trade-Off.aspx>