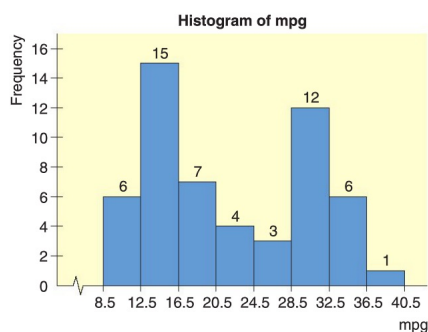
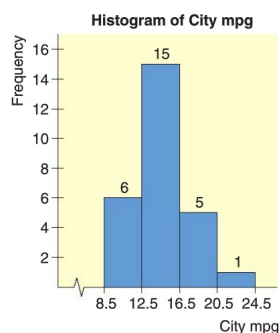


In-class Worksheet 2

1. Shown here is a histogram displaying the mileage, in miles per gallon, for a random selection of passenger cars.



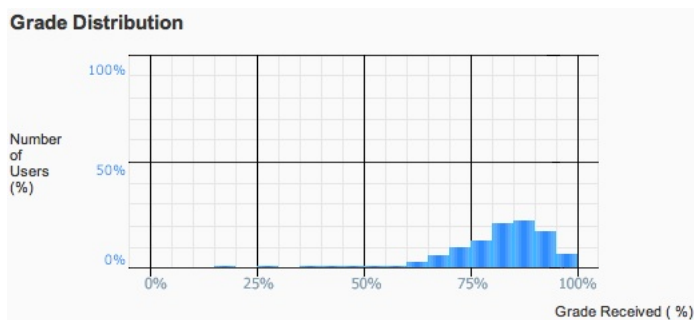
- (a) How many total data values were used to create this histogram?
 Summing the numbers sitting at the top of each bar in the histogram (which represent the frequencies of data in each category), $6 + 15 + 7 + 4 + 3 + 12 + 6 + 1 = 54$
- (b) What is the most common range of “miles per gallon” for a passenger cars in this sample?
 The interval from 12.5 to 16.5 contains 15 cars which is the highest frequency.
- (c) Describe the shape of this distribution. (Mound-shape symmetrical, Uniform, Skewed left, Skewed right, Bimodal).
 This is a bimodal distribution.
- (d) Curious about the bimodal shape of the distribution above (Is this consistent with your answer just above?), Jose did some research and found that this distribution included both city and highway mileage for each car in the sample. Looking at the raw data, Jose discovered the following histogram displays the the City Mileage for the cars in the sample.



Draw a histogram with the same axes as the original histogram and the following frequencies

Class	Frequency
8.5 – 12.5	0
12.5 – 16.5	0
16.5 – 20.5	2
20.5 – 24.5	3
24.5 – 28.5	3
28.5 – 32.5	12
32.5 – 36.5	6
36.5 – 40.5	1

2. Shown here is the relative frequency histogram (from D2L) of the overall course totals for MATH 1081 in Spring 2013 (600 students).



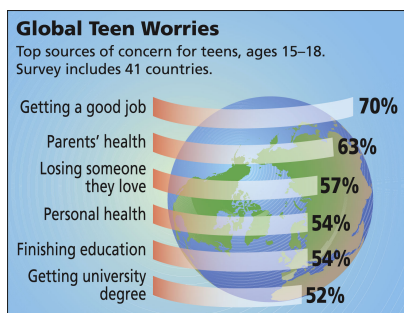
- (a) Describe the shape of this distribution. (Mound-shape symmetrical, Uniform, Skewed left, Skewed right, Bimodal).

This distribution is mound-shaped, but left-skewed.

- (b) Approximate the number of students that earned a 80% or higher in the class.

The students that scored 80% or higher are shown in the four rightmost bars. Those four bars represent about $21\% + 22\% + 19\% + 8\% = 70\%$ or 420 students.

3. Use the following figure to answer the questions below.



- (a) Is this a bar graph? Why or why not?

This is a bar graph, since it uses the lengths of the bars to represent the (relative) frequency of occurrence of each data value. It could be improved by using a scaled horizontal axis.

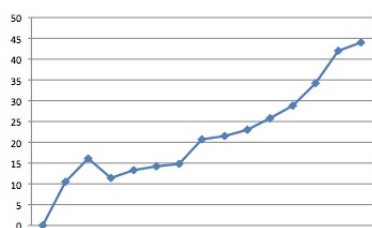
- (b) Could the same information be displayed in a Pareto chart? Explain.

The graph shown is essentially already a Pareto chart because the (relative) frequencies of occurrence of each data value are being displayed in decreasing order. Just rotate the figure 90-degrees counter-clockwise to make the presentation with vertical bars.

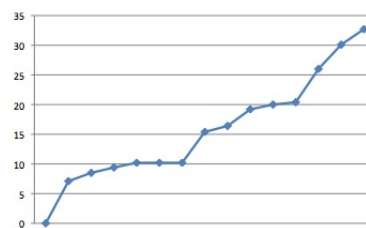
- (c) Could the same information be displayed in a circle graph? Explain.

This data could not be display in a circle graph, since the percentages do not represent the parts of a whole.

4. Without proper context and labels, ogives and time-series graphs can look very similar. Given here are two different graphs (both intentionally lacking labels and context). In *each* case determine if the graph can possibly represent an ogive and if it can possibly represent a time-series graph. Explain. (Note, I am not saying that one IS an ogive and one IS a time-series. For each graph decide if either, both, or neither option of “ogive” or “time-series” is possible and explain why or why not.)



In this case, there is a decrease in values from left to right, so this could not be an ogive, where the values of cumulative, hence always non-decreasing. It could be a time-series graph.



Since the data is always non-decreasing (from left to right), this could be both ogive and a time-series graph.

5. Using a *back-to-back* stem-and-leaf display, separate sets of data can more easily be compared. Consider the following display for Midterm scores for two different sections of the same class.

1 4 0 = 41 in Section 1 and 40 in Section 2	
SECTION 1	SECTION 2
7 4 3 1	4 0 0 1
9 8 8 7 7 5 5 3 1 1	5
6 5 4 4 1 0	6 4 4 5 5 5 7 8 8 9 9
0 0	7 0 1 1 3 5 5 5 6 6 7 8
8 8 6 5	8 0 0 2 5 5
9 9 7 7 5 5 5	9
0 0	10 0 0

- How many students took the Midterm in each section?
Section 1: 35 students; Section 2: 31 students
- What are the highest and lowest scores earned in Section 1? in Section 2?
Section 1: The highest score is 100. The lowest score is 41. Section 2: The highest score is 100. The lowest score is 40.
- What is the score with the highest frequency of occurrence in Section 1? in Section 2?
Section 1: The score with highest frequency is 95, with 3 occurrences. Section 2: The scores with highest frequency are 65 and 75, both with 3 occurrences.
- Which section displays a more *consistent* understanding of the material? Explain.
Section 2 shows 26 out the 31 students hovering in the interval of scores 64 to 85, where Section 1 is more spread out....not as notable of a *consistent* grouping.
- Which section displays a more *consistent* understanding of the material? Explain.
Without computing anything, it may look as though Section 1 may have a higher average, since there are more scores above 90. On the other hand, section 1 also as more scores below 60. So it may be difficult to decide which class is “better”. In fact, section 1 has a mean of ≈ 70.66 and section 2 has a mean of ≈ 71.42 . So if we define “better” to be the higher average, then the answer is section 2. But we don’t need to define better like this...