Chapter 3: Graphs

In my last stats course I was amazed to hear my teacher
announce that If we did not like our results, all we needed
to do was change our levels of confidence. – *Author Unknown*

Learning Objectives

Upon completion of this chapter, students should know

- The purpose and guidelines for constructing graphs.
- How to construct and interpret frequency histograms, frequency polygons, and relative frequency polygons.
- How to construct and interpret cumulative frequency polygons, cumulative relative frequency polygons, and cumulative percent polygons.
- How to construct and interpret stem and leaf diagram.
- How the shape of graphs can be changed without altering the data.

Key Terms

Terms are listed in the order they appear in the chapter.

Axes are the horizontal and vertical lines of a graph.

***X*-axis**, also known as the **abscissa**, is the horizontal line of a graph. In a graph of a frequency distribution, it is customary to represent the dependent variable along the abscissa. However, the independent variable is represented along the abscissa in a graph depicting the results of an experiment.

***Y*-axis**, also known as the **ordinate**, is the vertical line of a graph. In a graph of a frequency distribution, it is customary to represent the frequency along the ordinate. When the graph depicts the results of an experiment, the dependent variable is represented along this axis.

Frequency histogram is a type of bar graph where the width of each bar indicates the class interval size and the height of each bar indicates the frequency of the class interval. Histograms are used to illustrate the number of participants in each class interval. Each bar of the histogram extends from the lower to the upper real limits of the class interval.

Frequency polygon is a type of line graph that portrays the same information found in a histogram except the midpoint of the class interval is the single point plotted to represent the entire class interval. Lines connect the points on the graph and connect to the abscissa on both the extreme right and left ends of the graph. Like the histogram, the frequency polygon is also used to illustrate the number of participants in each class interval.

Relative frequency polygon is plotted the same as a frequency polygon, except the relative frequencies are represented along the ordinate. Because relative frequencies are the proportion of scores that fall within the class interval, similar sets of information that have different numbers of participants can be compared graphically using the relative frequency polygon.

Cumulative frequency polygon is plotted using the upper real limit of the class interval as the most representative single point. The cumulative frequency is represented along the ordinate. Since cumulative frequency is the total number of scores that fall below the upper real limit of the class interval, the last point plotted at the highest class interval is equal to *n*. The cumulative frequency line is not brought down to the abscissa after the last point but is allowed to hang in the air.

Cumulative relative frequency polygon is plotted at the upper real limit of the class interval. The cumulative relative frequency is the proportion of scores that fall below the upper real limit of the class interval and the line connecting to point plotted at the highest class interval is left hanging.

Cumulative percent polygon, also called percentile, is plotted the same as the cumulative frequency polygon except cumulative percent is represented along the ordinate. Because the cumulative percent polygon is the percentage of scores below the upper real limit it is easier to interpret.

Stem-and-leaf diagram is a technique that displays data visually by dividing the raw score into two parts, a stem and a leaf. The leaf is normally the last digit of the score and the stem is the remaining digit(s) of the score.

Lecture and Demonstration Ideas

"A picture is worth a thousand words." *Barnard* (1927)

. Visual information can be very powerful and such information is sometimes used to draw conclusions or influence decisions. These days it seems a picture is worth a hundred thousand dollars. In addition to guidelines for proper graph construction (see Transparencies 3-1 to 3-7), the text also mentions several methods graphs can be modified to make data appear more or less favorable. Mistakenly, however, students tend to think that data are altered when graphs are drawn in distorted ways. Students might find it interesting to know that a comparison of annual reports of US. corporations from 1991-93 and from 1996-97 showed that 33 percent and 17 percent, respectively, modified graphs in some way to make them appear more favorable (Louwers, Pitman, & Radtke, 1999). Thus, it is possible data are altered in some cases of distortion, but it is also likely the data are factual.

While statisticians use graphs to display relationships and experimental effects, they are also used to visually examine the overall pattern of data in the distribution. Since careful visual examination of summary data

is an important step in data analyses, you may want to discuss some of the reasons for this procedure (identification of coding errors, outliers, etc.).

Use the concept map of descriptive statistics, shown on Transparency 3-15, to show students the relationship of the measurement scales to different graphs. Since graphs are used as visual summaries of a distribution, ask students to examine the histograms shown on Transparency 3-10 and discuss the differences in the three distribution patterns. Is there more variability in one distribution than the others? Is the center or average of the distribution easier to determine in one histogram than the others? This discussion ties into the magazine readability study mentioned below and the next two chapters (central tendency and variation) of the text.

Magazine Study. Shuptrine and McVicker (1981) examined readability of magazines relative to educational level. In this study, participants ranked 30 magazines according to educational level of their readers. Then, based on these rankings, magazines were grouped into high, medium, and low educational levels. From each of these three groups, three magazines were randomly selected. Next, six advertisements were randomly selected from each of the nine magazines. Among the data observed, as shown in Transparency 3-8, were the numbers of words with 3 or more syllables in the advertising copy. Use this dataset to demonstrate graph construction (Transparencies 3-1 to 3-7) for one condition of the study. Solutions are shown on Transparencies 3-9 to 3-11.

Critique Time. After covering the elements of graph construction, have students examine the graph shown on Transparency 13-14. This histogram was constructed using grouped frequency data on the average number of absences reported for full-time wage and salary workers by age and sex based on the 2000 census of annual household data (Bureau of Labor Statistics, 2001). Students should notice a problem with the histogram because interval sizes differ. (Note: The grouped frequency distribution from the census did not have equal interval sizes (see Transparency 3-13.)

Instructional Video. *Against All Odds: Inside Statistics.* Program Two, "Picturing Distributions." The first 14 minutes of this 30-minute video are specifically related to this chapter. Histograms are mentioned in examples of lightning prediction, etc. This video was produced by the Consortium for Mathematics and Its Applications and Chedd-Angier (1989). Available through Annenberg/CPB.

Active-Learning Activities

Magazine Readability Study. Ask your students to practice constructing histograms and polygons using data from the two conditions not used in your demonstration. In addition, have students compare the relative frequencies of the three conditions in a single relative frequency polygon. The solution is shown on Transparency 3-12.

Age and Stem-and-Leaf Diagrams. Generate class data by asking students to count the change in their pockets and/or purses. Write this data on the board and ask students prepare a stem-and-leaf diagram. Although our text does not cover back-to-back stem-and-leaf diagrams, you may wish to demonstrate this by separating the class data by gender.

Additional Assignments

Assignment 1. Ask students to use data from Shuptrine and McVicker (1981) study to construct either a histogram or polygon that makes the educational condition look very favorable. Along with the modified graph, have students include a brief description of the method they used to accomplish this "modification" without altering the data. Raw data and the grouped frequency distribution solutions are shown on Transparency 3-8 and Transparency 3-9.

Assignment 2. There are a number of websites with examples of problem graphs. You may want to print several to use as examples in class. Michael Friendly's (York University), *Gallery of Data Visualization* has a lot of examples of both good and bad graphs. The specific web page is called, *The Best and the Worst Statistical Graphs,"* located at the following URL: http://www.math.yorku.ca/SCS/Gallery/. Ask students to visit this website and write a critique on one of the problem graphs. Which of the major guidelines of graph construction was missing? Was the graph distorted in a way to influence a certain point? Explain how this can be determined.

References:

Bureau of Labor Statistics (2001). *Number of absences reported for full-time wage and salary workers by age and sex based on the 2000 census of annual household data*. Accessed 8/1/01. http://www.bls.census.gov/cps/cpsmain.htm

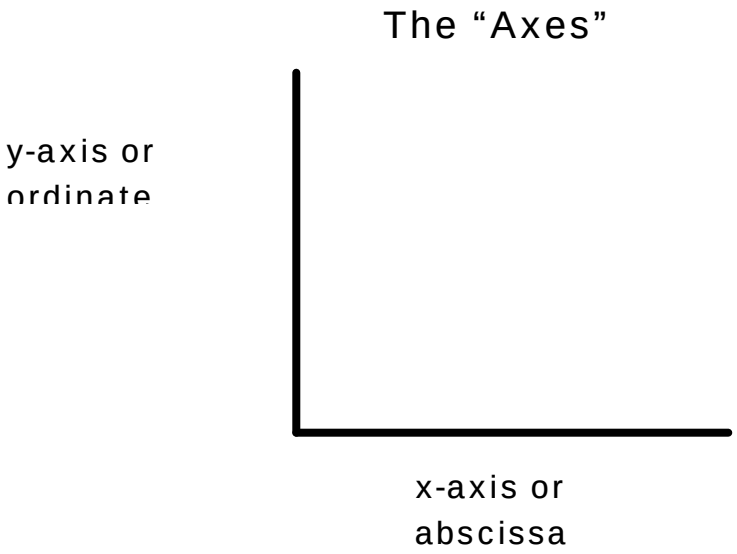
Clark, C. S. (1999. Editing Statistical and Illustrative Graphics. Accessed 6/12/01. http://www.peak.org/~cscottc/pdfs/graphics primer.pdf.

Louwers, T.J., Pitman, M. K., & R.R. Radtke, R. R. (1999, May 1). Please pass the salt: A look at creative reporting in annual reports. *Today's CPA*, 20-23.

Shuptrine, F. K., & McVicker, D. D. (1981). Readability levels of magazine ads. *Journal of Advertising Research, 21*, 47-52.

Transparency 3-1

Basics in Constructing a Graph



1. Give the graph a clear, unambiguous title or figure caption.

2. Assign appropriate labels and meaningful numbers to each axis.

3. When Plotting Relationships:

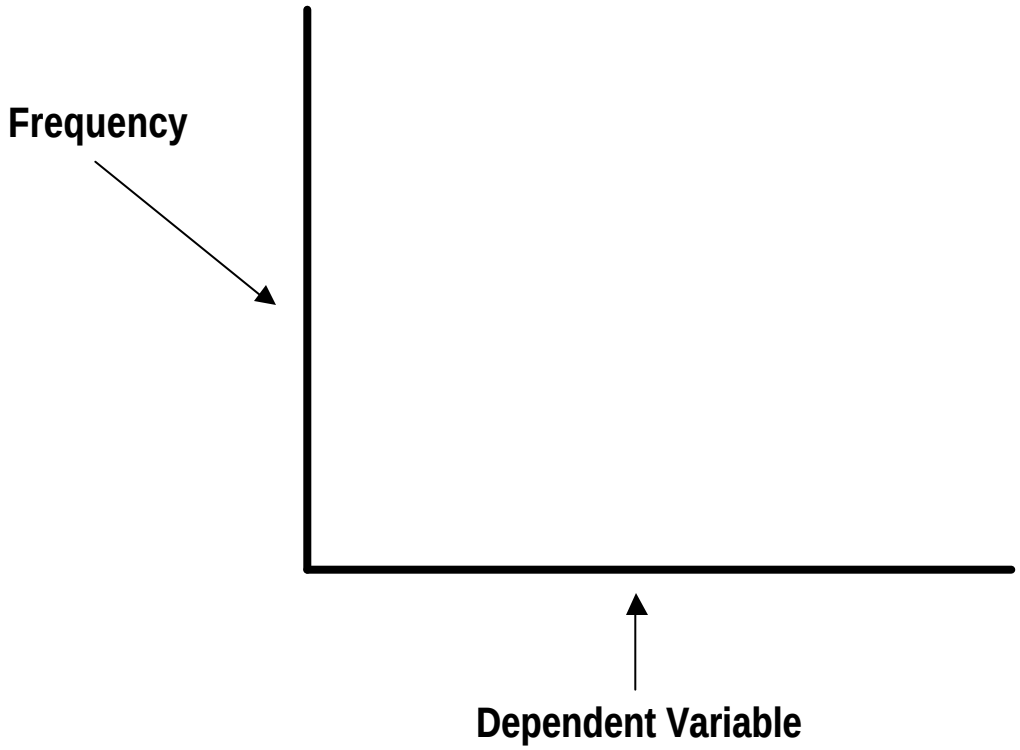
   Plot the independent variable along the abscissa
   Plot the dependent variable along the ordinate

4. When Plotting Data From a Frequency Distribution:

   Plot the dependent variable along the abscissa
   Plot the frequency along the ordinate

Transparency 3-2.

Graphing Frequency Distributions



Transparency 3-3.

Major Points to Setting Up Graphs

- ✓ Give the graph a clear, unambiguous title.
- ✓ Plot the independent variable along the abscissa and dependent variable along the ordinate.
- ✓ Plot the dependent variable along the abscissa and the frequency along the ordinate.

Transparency 3-4.

Histograms

- ✓ A frequency histogram consists of a number of bars placed side by side
- ✓ The width of each bar indicates the interval size.
- ✓ The height of each bar indicates the frequency of the interval.
- ✓ Plot the bars using the real limits.

Polygons

- ✓ Portrays the same information as a frequency histogram.
- ✓ Uses the midpoint, as the single point that represents the class interval.

Transparency 3-5.

Cumulative Frequency Polygon

- ✓ **The total number of scores that fall below the upper real limit of an interval.**
- ✓ **Plot the first point at the upper real limit of the interval below the lowest interval.**
- ✓ **The last point represents the cumulative frequency of the last interval, which is equal to *n*. It is left hanging in the air.**

Transparency 3-6.

Relative Frequency Polygon

- ✓ Plot at midpoint of class interval.
- ✓ Plot the relative frequencies.
- ✓ This type of graph can be used to compare similar sets of information that have different numbers of scores.

Cumulative Relative Polygon

- ✓ Plotted similar to cumulative frequency polygon except it is plotted using the cumulative relative frequencies.

Transparency 3-7.

Stem and Leaf Diagrams

- ✓ Allows raw data to be displayed visually.
- ✓ The leaf is normally the last digit (or in very large numbers, the last two or three digits) of the score.
- ✓ The stem is the remaining digits of the score.

Raw data
34
38
45
56
67
81

Stem	Leaf
3	4 8
4	5
5	6
6	7
8	1

Transparency 3-8.

Magazine Educational Level of Their Readers			
Highest Educational Level			
34.00	24.00	32.00	5.00
21.00	39.00	17.00	6.00
37.00	10.00	3.00	6.00
31.00	17.00	10.00	
10.00	18.00	6.00	
Medium Educational Level			
13.00	2.00	29.00	15.00
22.00	9.00	26.00	3.00
25.00	3.00	5.00	8.00
3.00	.00	5.00	
5.00	4.00	24.00	
Lowest Educational Level			
16.00	10.00	9.00	10.00
9.00	11.00	43.00	24.00
10.00	12.00	13.00	15.00
3.00	13.00	14.00	
12.00	1.00	32.00	

Highest Educational Level = Scientific American, Fortune, The New Yorker Group

Medium Educational Level = Sports Illustrated, Newsweek, People Group

Lowest Educational Level = National Enquirer, Grit, True Confessions

SOURCE:

Shuptrine, F. K., & McVicker, D.D. (1981). Readability Levels of Magazine Ads. *Journal of Advertising Research, 21,* 47-52.

Transparency 3-9.

Grouped Frequency Distribution of Words With 3 or More Syllables
Highest Educational Level Magazines

Real Limits	Apparent Limits	*f*	Midpoint	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
34.5 – 39.5	35 – 39	2	37	18	.111	.1.00
29.5 – 34.5	30 – 34	3	32	16	.167	.890
24.5 – 29.5	25 – 29	0	27	13	.000	.723
19.5 – 24.5	20 – 24	2	22	12	.111	.723
14.5 – 19.5	15 – 19	3	17	11	.167	.612
9.5 – 14.5	10 – 14	3	12	8	.167	.445
4.5 – 9.5	5 – 9	4	7	5	.222	.278
0 – 4.5	0 – 4	1	2	1	.056	..056

Grouped Frequency Distribution of Words With 3 or More Syllables
Medium Educational Level Magazines

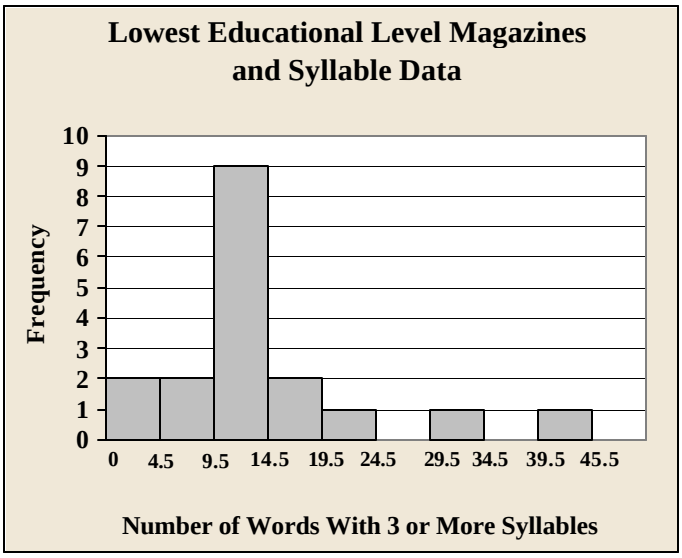
Real Limits	Apparent Limits	*f*	Midpoint	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
24.5 – 29.5	25 – 29	3	27	18	.167	1.00
19.5 – 24.5	20 – 24	2	22	15	.111	.833
14.5 – 19.5	15 – 19	1	17	13	.056	.722
9.5 – 14.5	10 – 14	2	12	12	.111	.666
4.5 – 9.5	5 – 9	4	7	10	.222	.555
0 – 4.5	0 – 4	6	2	6	.333	.333

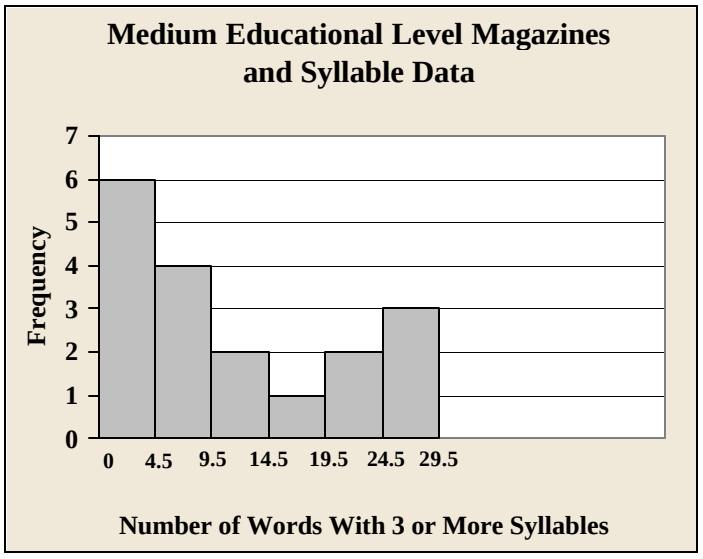
Grouped Frequency Distribution of Words With 3 or More Syllables
in Lowest Educational Level Magazines

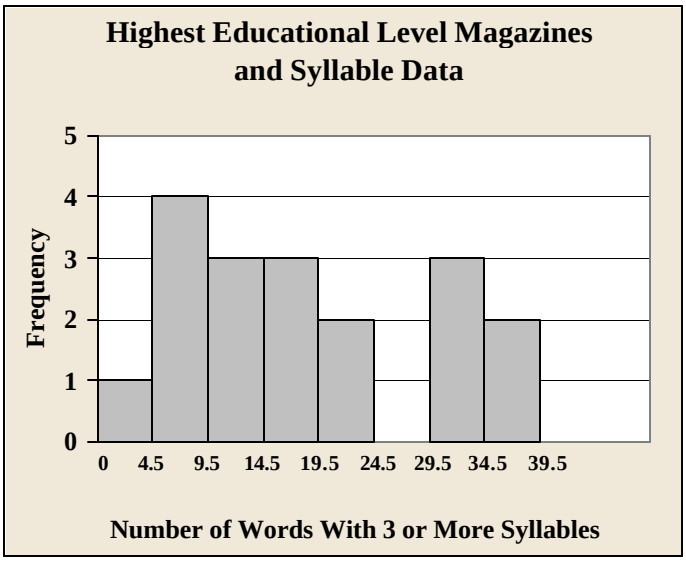
Real Limits	Apparent Limits	*f*	Midpoint	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
39.5 – 45.5	40 – 44	1	42	18	.056	1.00
34.5 – 39.5	35 – 39	0	37	17	000	.945
29.5 – 34.5	30 – 34	1	32	17	.056	.945
24.5 – 29.5	25 – 29	0	27	16	000	.889
19.5 – 24.5	20 – 24	1	22	16	.056	.889
14.5 – 19.5	15 – 19	2	17	15	.111	.833
9.5 – 14.5	10 – 14	9	12	13	.500	.722
4.5 – 9.5	5 – 9	2	7	4	.111	.222
0 – 4.5	0 – 4	2	2	2	.111	.111

Transparency 3-10

Histograms - Magazine Ad Readability Study

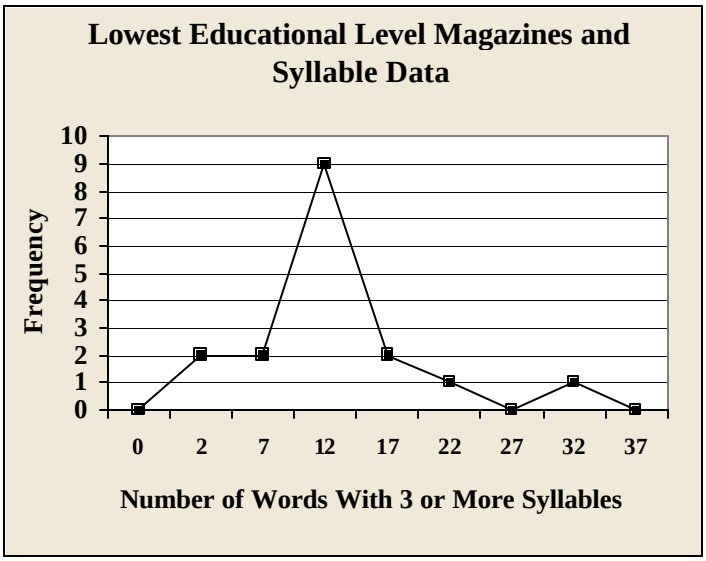


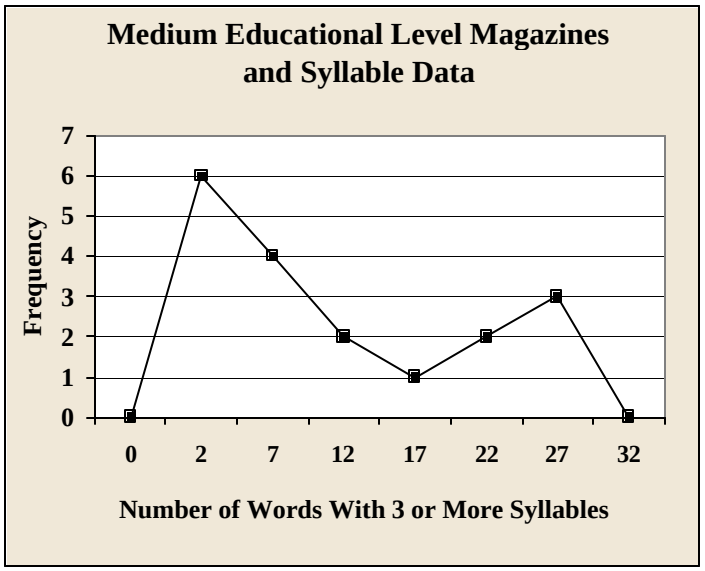


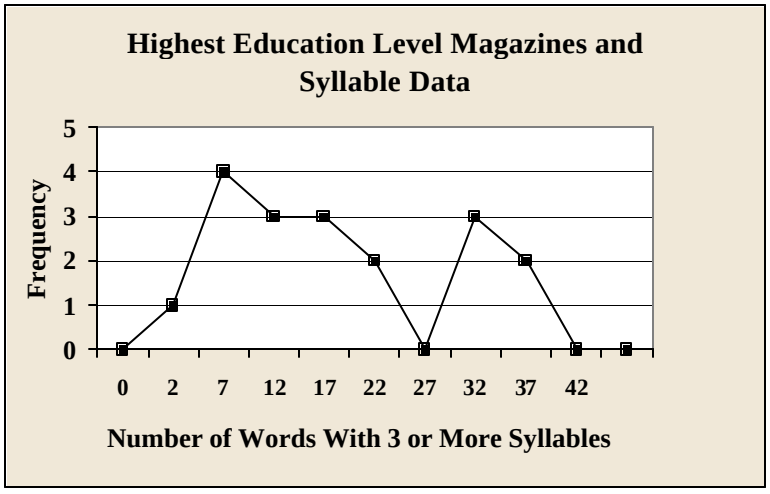


Transparency 3-11.

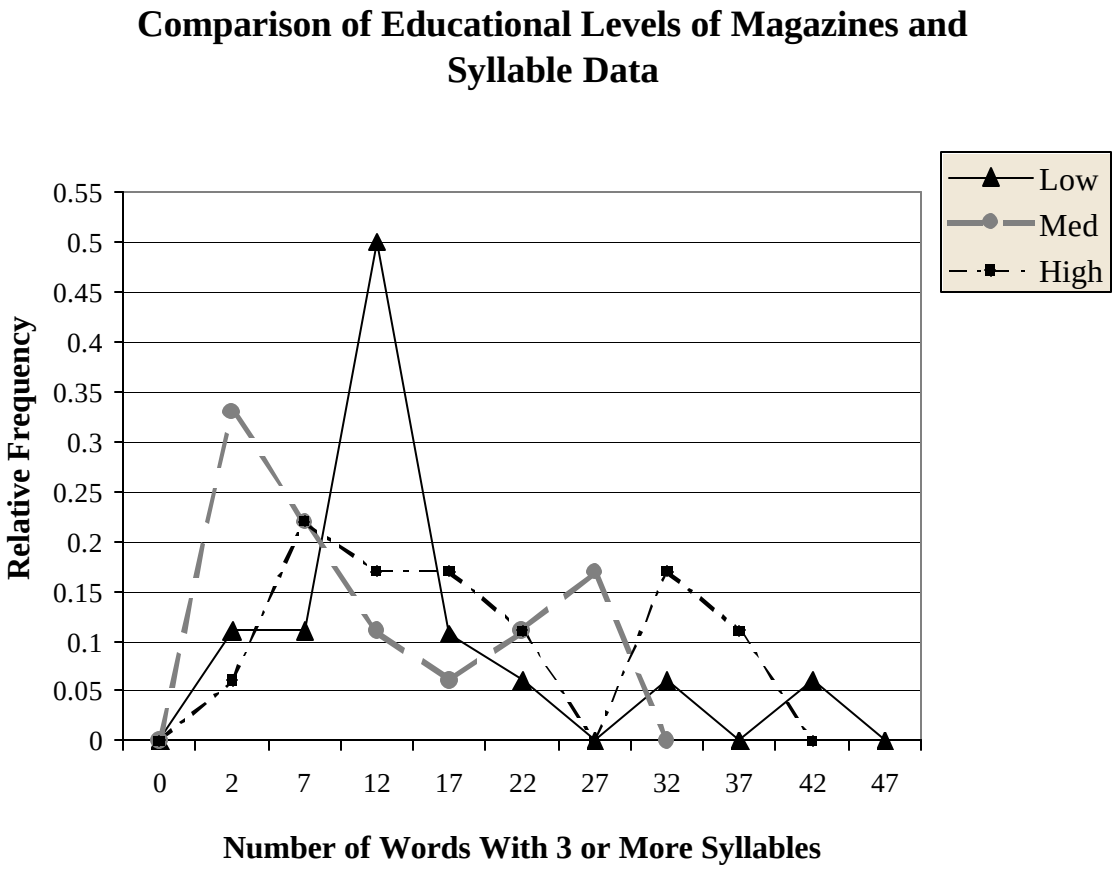
Polygons - Magazine Ad Readability Study







Transparency 3-12.



Transparency 3-13.

Age and Sex	Absences from work of employed full-time wage and salary workers by age and sex. Household Data: Annual Averages 2000						
	Total Employed	Absence Rate			Lost Worktime Rate		
		Total	Illness or Injury	Other Reasons	Total	Illness or Injury	Other Reasons
Total, 16 years and over	99,846	3.8	2.7	1	2	1.4	0.6
16 to 19 years	2,382	4.2	3.1	1.1	1.9	1.3	0.6
20 to 24 years	9,608	4	2.7	1.3	1.9	1.2	0.7
25 years and over	87,856	3.8	2.7	1	2	1.5	0.5
25 to 54 years	77,120	3.7	2.7	1.1	2	1.4	0.6
55 years and over	10,736	3.9	3.2	0.7	2.3	2	0.3
Men, 16 years and over	56,228	2.8	2.2	0.6	1.5	1.2	0.3
16 to 19 years	1,414	3.2	2.5	0.7	1.5	1.1	0.4
20 to 24 years..	5,420	2.8	2.1	0.7	1.3	1	0.3
25 years and over	49,394	2.7	2.2	0.6	1.5	1.2	0.3
25 to 54 years	43,353	2.7	2.1	0.6	1.4	1.2	0.3
55 years and over	6,041	3.3	2.7	0.6	2	1.7	0.3
Women, 16 years and over	43,618	5.1	3.5	1.6	2.7	1.8	1
16 to 19 years	968	5.6	3.9	1.7	2.6	1.5	1.1
20 to 24 years	4,188	5.5	3.5	2	2.7	1.5	1.2
25 years and over	38,461	5.1	3.5	1.6	2.7	1.8	0.9
25 to 54 years	33,767	5.1	3.4	1.7	2.7	1.7	1
55 years and over	4,694	4.8	3.9	0.9	2.6	2.3	0.4

Data Source:

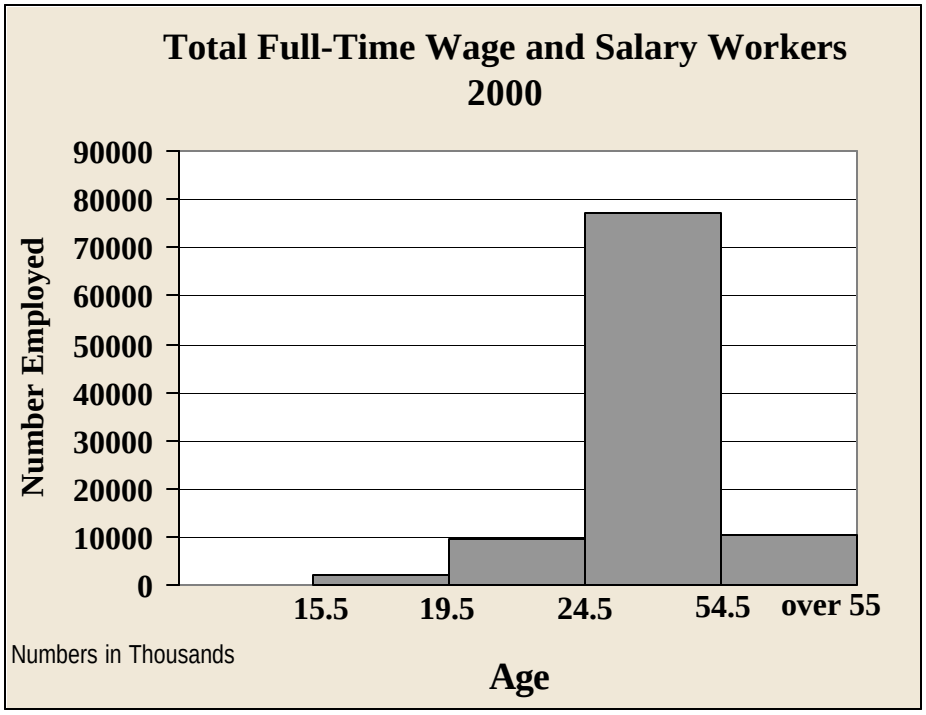
U.S. Bureau of Labor Statistics
Division of Labor Force Statistics
Suite 4675
2 Massachusetts Avenue, NE
Washington, DC 20212-0001

1 Absences are defined as instances when persons who usually work 35 or more hours a week worked less than 35 hours during the reference week for one of the following reasons: Own illness, injury, or medical problems; child-care problems; other family or personal obligations; civic or military duty; and maternity or paternity leave. Excluded are situations in which work was missed due to vacation or personal days, holiday, labor dispute, and other reasons. For multiple jobholders, absence data refer only to work missed at their main jobs. The absence rate is the ratio of workers with absences to total full-time wage and salary employment. The estimates of full-time wage and salary employment shown in this table do not match those in other tables because the estimates in this table are based on the full CPS sample and those in the other tables are based on a quarter of the sample only.

2 Hours absent as a percent of hours usually worked.

NOTE: Beginning in January 2000, data reflect revised population controls used in the household survey.

Transparency 3-14.



Transparency 3-15.

Concept Map