STOP GLOBAL WARMING: BECOME A PIRATE

Global Average Temperature Vs. Number of Pirates

WWW.VENGANZA.ORG

CORRELATION DOES NOT PROVE CAUSATION!

The Organization and Graphic Presentation of Data

Frequency Distributions for Qualitative Data

Definitions:

- The **frequency** of a category is the number of times it occurs in the data set.

- A **frequency distribution** is a table that presents the frequency for each category.

Example: Construct a frequency table for the numbers of newspapers sold at a local shop over the last 10 days:

$$22, 20, 18, 23, 20, 25, 22, 20, 18, 20$$

Solution:

Papers Sold	Frequency
18	2
19	0
20	4
21	0
22	2
23	1
24	0
25	1

Definitions:

- The **Percentage**, or **proportion** of a category is the frequency of the category divided by the sum of all the frequencies.
$$\text{Proportion} = \frac{\text{Frequency}}{\text{Sum of all frequencies}}$$
It is recommended that you use decimals rather than whole percents.

- A **percentage distribution** is a table that presents the relative percentage of each category.

Example: Construct a percentage table for the data in the previous example.

Solution:

1. Find the total number of observations by summing the frequencies:

$$\text{Sum of frequencies} \quad = \quad 2 + 0 + 4 + 0 + 2 + 1 + 0 + 1$$
$$= \quad 10$$

2. Find the relative frequencies for all of the viable categories:

$$\text{Percentage for 18 papers sold} \quad = \quad \frac{2}{10}$$
$$= \quad 0.20$$

$$\text{Percentage for 20 papers sold} \quad = \quad \frac{4}{10}$$
$$= \quad 0.40$$

$$\text{Percentage for 22 papers sold} \quad = \quad \frac{2}{10}$$
$$= \quad 0.20$$

$$\text{Percentage for 23 papers sold} \quad = \quad \frac{1}{10}$$
$$= \quad 0.10$$

$$\text{Percentage for 25 papers sold} \quad = \quad \frac{1}{10}$$
$$= \quad 0.10$$

3. Draw the table

Papers Sold	Frequency	Percentage (in decimal form)
18	2	0.20
19	0	0.00
20	4	0.40
21	0	0.00
22	2	0.20
23	1	0.10
24	0	0.00
25	1	0.10

On Your Own: Construct a percentage table for the data below that sampled students and asked the total number of children in their family (including themselves).

$$
\begin{array}{ccccccccc}
2 & 2 & 2 & 4 & 5 & 3 & 3 & 3 & 3 \\
2 & 1 & 2 & 3 & 5 & 3 & 4 & 3 & 1 \\
2 & 3 & 5 & 3 & 2 & 1 & 3 & 2
\end{array}
$$

Solution:

1. Find the total number of observations by summing the frequencies:

$$
\begin{aligned}
\text{Sum of frequencies} \;=\;& 2+2+2+4+5+3+3+3+3+2+1+2+3+5 \\
& +3+4+3+1+2+3+5+3+2+1+3+2 \\
=\;& 26
\end{aligned}
$$

2. Find the relative frequencies for all of the viable categories:

$$
\text{Percentage for 1 child} \;=\; \frac{3}{26}
$$
$$
\approx\; 0.12
$$

$$
\text{Percentage for 2 children} \;=\; \frac{8}{26}
$$
$$
\approx\; 0.31
$$

$$
\text{Percentage for 3 children} \;=\; \frac{10}{26}
$$
$$
\approx\; 0.38
$$

$$
\text{Percentage for 4 children} \;=\; \frac{2}{26}
$$
$$
\approx\; 0.08
$$

$$
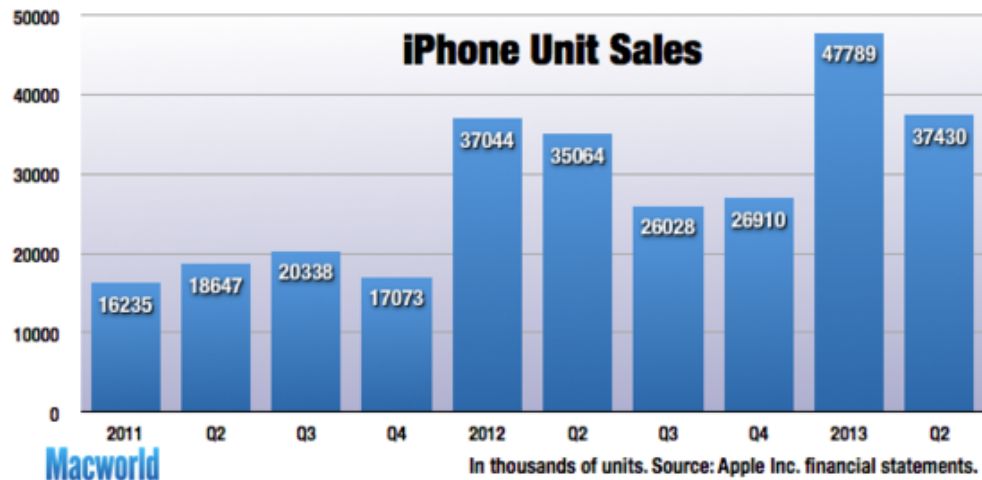\text{Percentage for 25 children} \;=\; \frac{3}{26}
$$
$$
\approx\; 0.12
$$

3. Draw the table

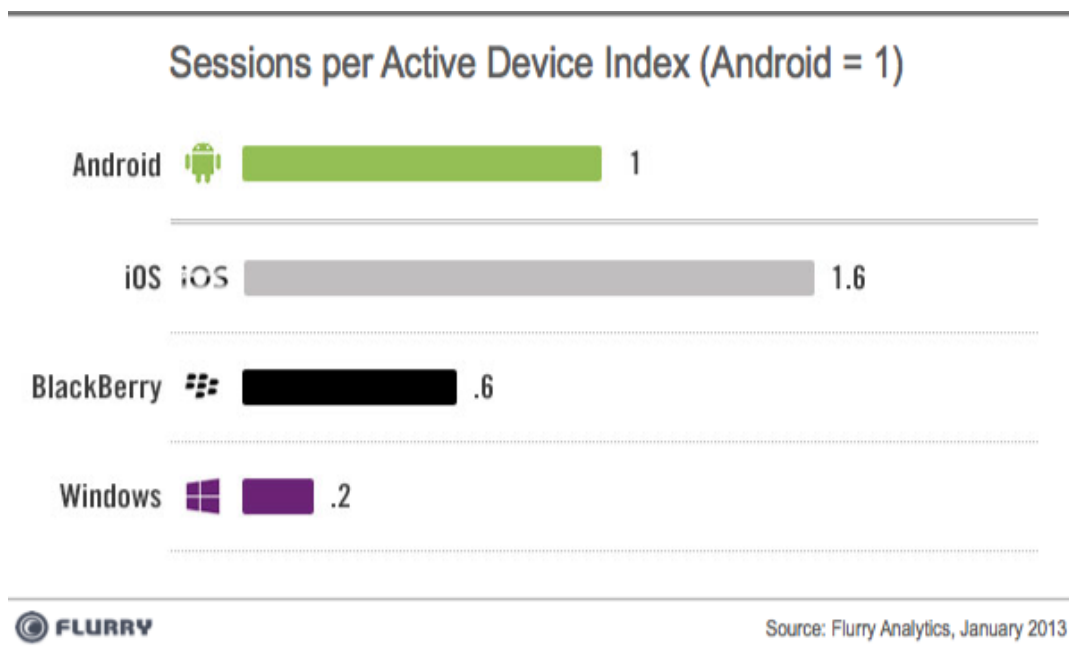| Children	Frequency	Percentage
1	3	0.12
2	8	0.31
3	10	0.38
4	2	0.08
5	3	0.12

Bar Graphs

Definition:

- A **bar graph** is a graphical representation of a frequency distribution (may be vertical or horizontal).
  The bars in each category never touch each other.

Example:
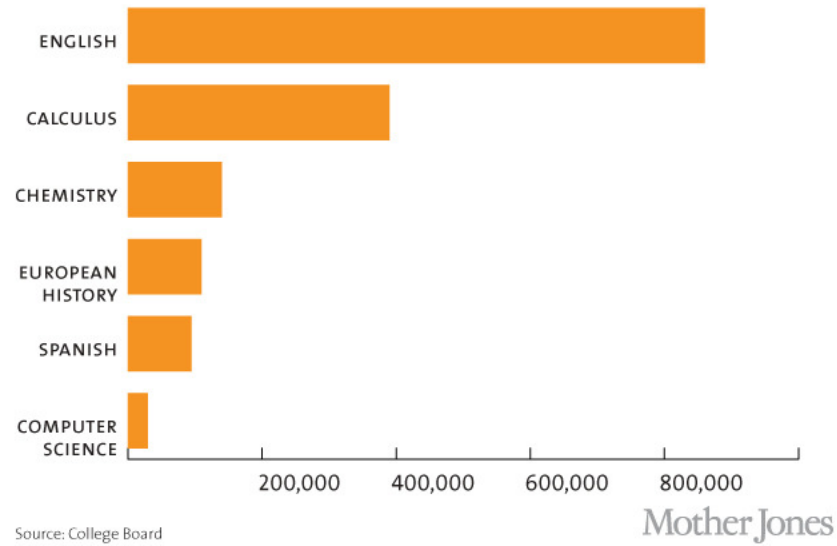

(Friedman, 2013)

Example:

(Gordon, 2013)

Definition:

- A **Pareto chart** is a bar graph in which categories are presented in order of frequency or Percentage.

Example:

Number of AP Exams in Selected Subjects, 2013
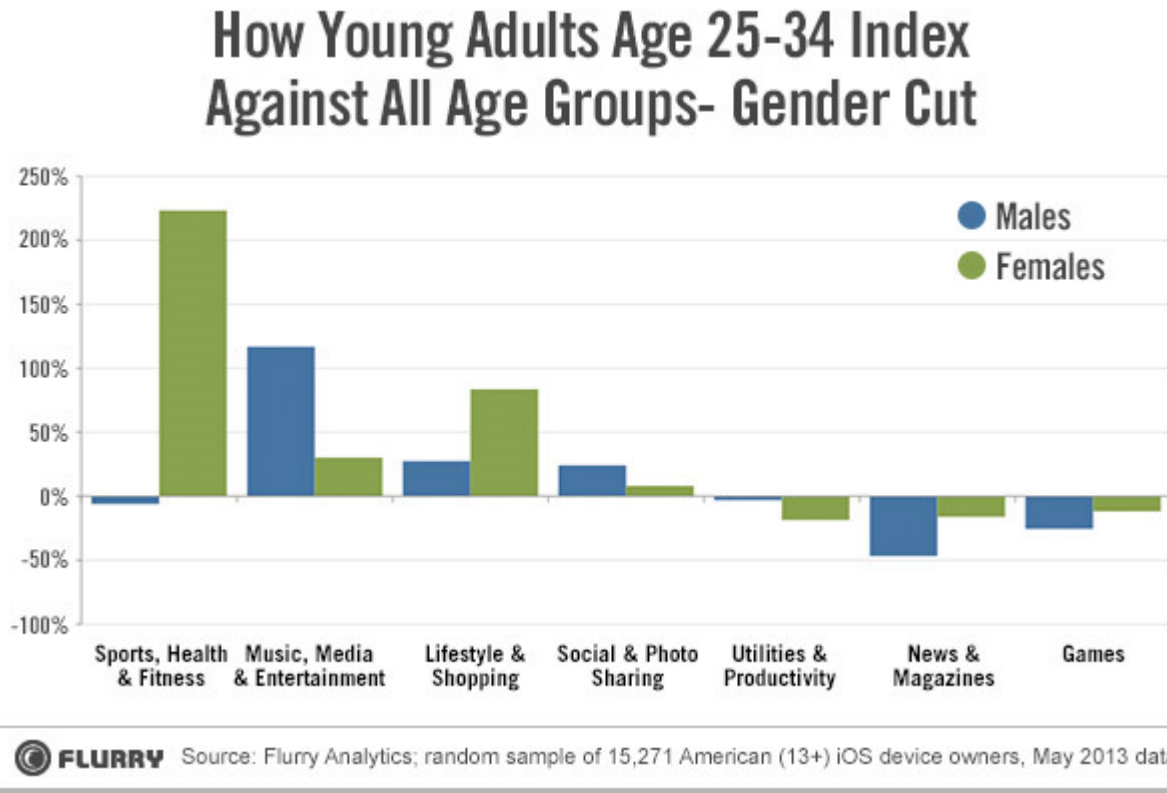
Source: College Board

Mother Jones

(Raja, 2014)

Definition:

- A **side-by-side bar graph** is a bar graph in which bars that correspond to the same category are set next to each other (typically for comparison).
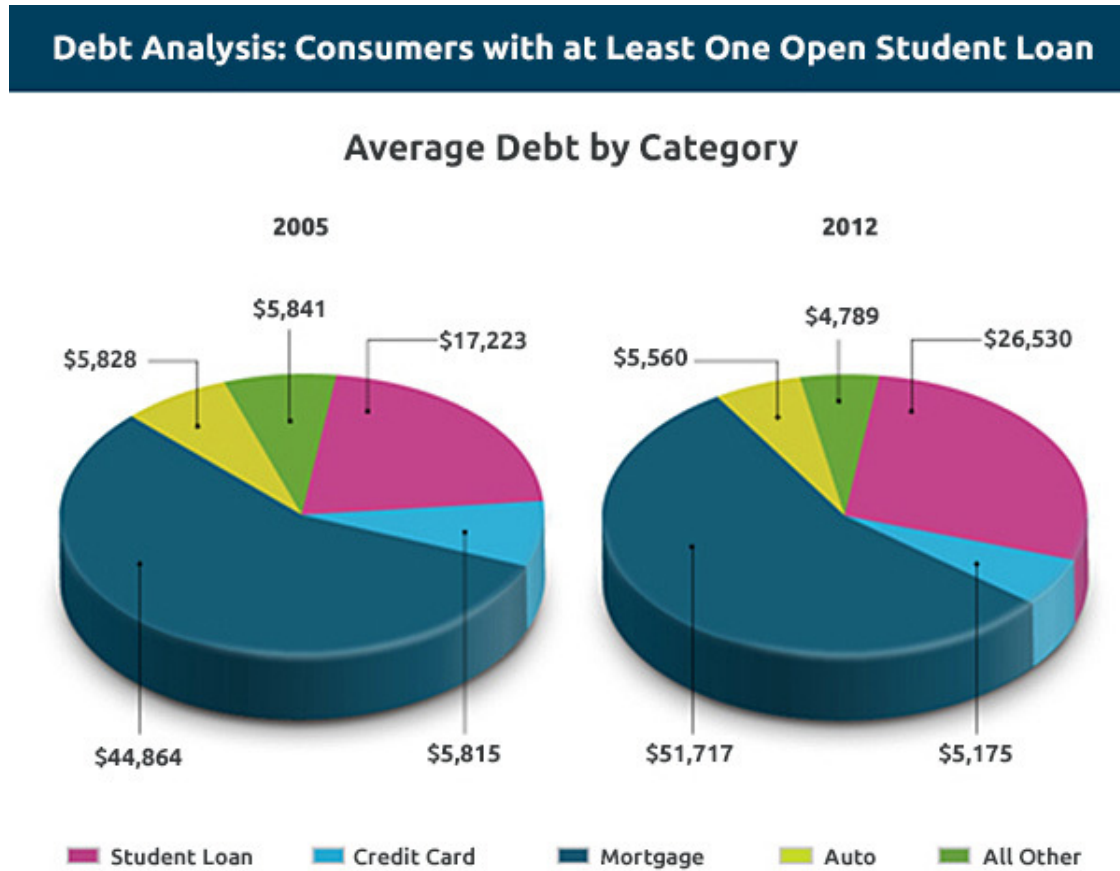
Example:

How Young Adults Age 25-34 Index Against All Age Groups- Gender Cut

(Khalaf, 2013)

Pie Charts

Definition:

- A **pie chart** is a circular chart divided into sectors, illustrating numerical proportion or frequencies.

Example:

Debt Analysis: Consumers with at Least One Open Student Loan

Average Debt by Category

2005

$5,841
$17,223
$5,828
$44,864
$5,815

2012

$4,789
$26,530
$5,560
$51,717
$5,175

Student Loan   Credit Card   Mortgage   Auto   All Other

Source: FICO™ Banking Analytics Blog. © 2012 Fair Isaac Corporation

(Quinn, 2013)

8

Frequency Distributions for Quantitative Data

Definition:

- A **class** is an interval in a frequency distribution. (e.g. 1.01-2.00, 2.01-3.00, 3.01-4.00, etc.)

- The **lower class limit** of a class is the smallest value that can appear in that class.

- The **upper class limit** of a class is largest value that can appear in that class.

- The **class width** is the difference between consecutive lower class limits.

- The **Percentage** of a category is the frequency of the category divided by the sum of all the frequencies.

$$\text{Percentage} = \frac{\text{Frequency}}{\text{Sum of all frequencies}} = \frac{\text{Class frequency}}{\text{Sample size}} = \frac{f}{n}$$

Note: Requirements for Choosing Classes:

- Every observation must fall into one of the defined classes.

- The classes must not overlap.

- The classes must be of equal width.

- There must be no gaps between classes.

Note: Choosing a Frequency Distribution:

1. Choose a class width. (typically there should be no less than 5 and no more than 20 classes so choose your width appropriately)

2. Choose a lower class limit. (typically slightly less than the minimum data value)

3. Compute the lower limit for the second class by adding the class with to the lower limit for the first class.

4. Continue this process until you have a class with the largest data value included.

5. Count the number of class and construct the frequency distribution.

Note: If you are given the number of classes, then the

$$\text{Class width} = \frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}$$

and then round up!

Example: Make a frequency distribution for the following data, using 5 classes:

$$5 \quad 10 \quad 7 \quad 19 \quad 25 \quad 12 \quad 15 \quad 7 \quad 6 \quad 8$$
$$17 \quad 17 \quad 22 \quad 21 \quad 7 \quad 7 \quad 24 \quad 5 \quad 6 \quad 5$$

Solution:

- Lower class limit: 5

- Upper class limit: 25

- Class width: $\dfrac{25 - 5}{5} = 4$

Classes will be in increments of 4: 5-9, 10-14, 15-19, 20-24, 25-29.

Class	Frequency
5-9	10
10-14	2
15-19	4
20-24	3
25-29	1

On Your Own: The following table presents the purchase totals (in dollars) of a random sample of gasoline purchases at a convenience store. Construct a frequency distribution using a class width of 10, and using 0 as the lower class limit for the first class.

76.59	48.55	93.66	60.17	39.10
93.28	65.43	34.12	80.41	77.16
80.07	93.46	39.19	43.84	44.70
68.74	89.98	6.97	52.86	68.93

Solution:

- Lower class limit: 0

- Upper class limit: 25

- Class width: 10

Class	Frequency
0.00-9.99	1
10.00-19.99	0
20.00-29.99	0
30.00-39.99	3
40.00-49.99	3
50.00-59.99	1
60.00-69.99	4
70.00-79.99	2
80.00-89.99	3
90.00-99.99	3

Histograms

Definition:

- A **histogram** is a graph of a frequency distribution in which rectangles with bases on the horizontal axis are given widths equal to the class intervals and heights equal to the corresponding frequencies. There are many types, but of enow, we have **frequency histograms** and **Percentage histograms**.
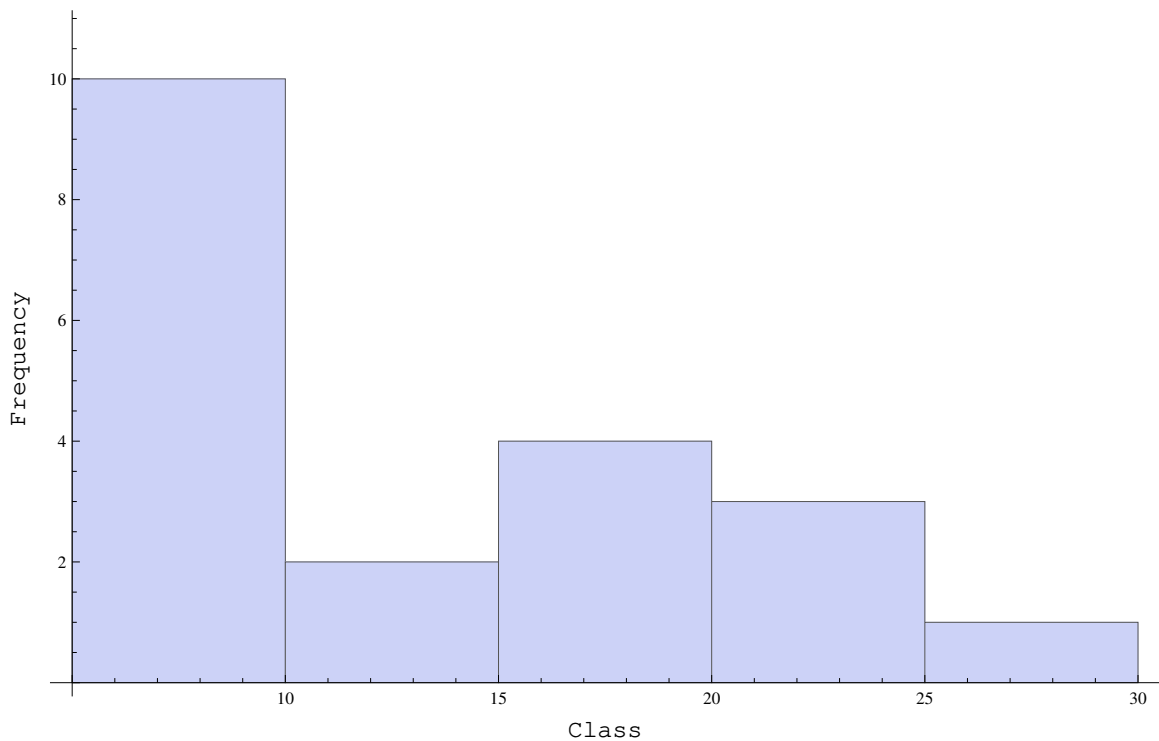
  Note: If any class is open ended, a histogram cannot be drawn.

Example: Construct a frequency histogram using the following data using only 5 classes:

5	10	7	19	25	12	15	7	6	8
17	17	22	21	7	7	24	5	6	5

Solution:

| Class	Frequency
5-9	10
10-14	2
15-19	4
20-24	3
25-29	1

On Your Own: The following table presents the purchase totals (in dollars) of a random sample of gasoline purchases at a convenience store. Construct a frequency histogram using a class width of 10, and using 0 as the lower class limit for the first class.

76.59	48.55	93.66	60.17	39.10
93.28	65.43	34.12	80.41	77.16
80.07	93.46	39.19	43.84	44.70
68.74	89.98	6.97	52.86	68.93

Solution:

Class	Frequency
0.00-9.99	1
10.00-19.99	0
20.00-29.99	0
30.00-39.99	3
40.00-49.99	3
50.00-59.99	1
60.00-69.99	4
70.00-79.99	2
80.00-89.99	3
90.00-99.99	3
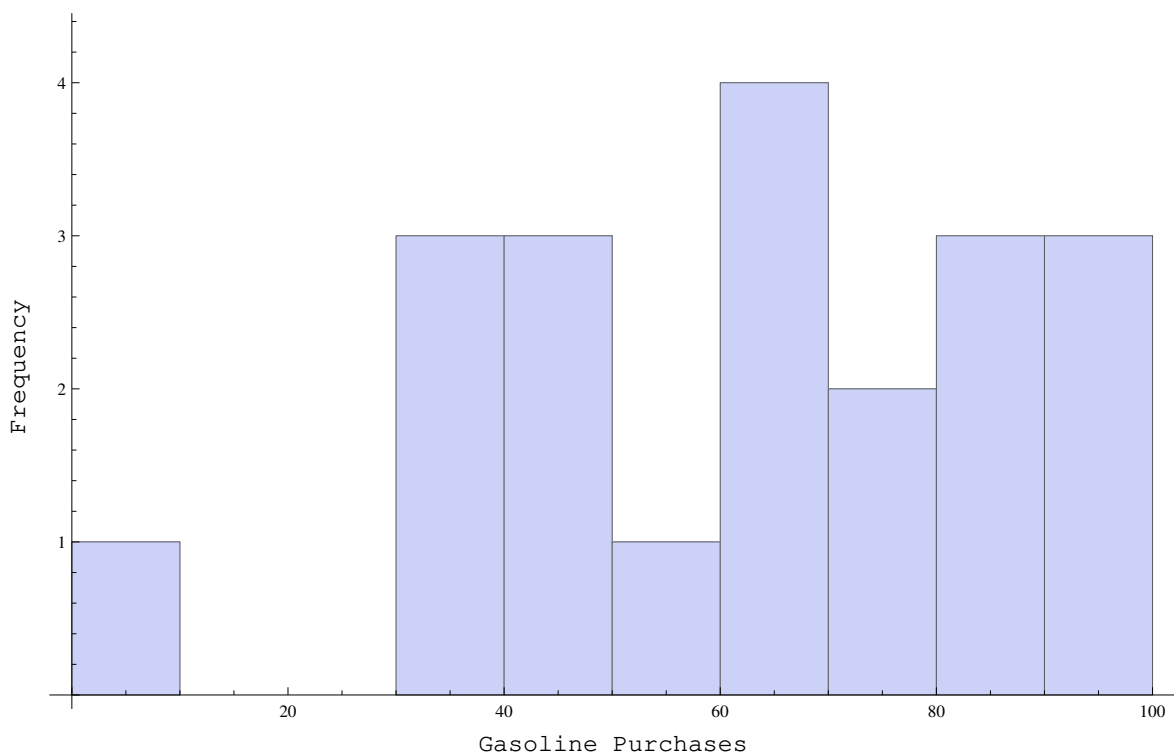
# Shapes of Histograms

Definitions:

- A histogram (or any distribution) is **skewed** if one side (called a tail) is longer than the other.

Skewed left	Approximately symmetric	Skewed right
Negatively skewed	Normal	Positively skewed
Long left tail	Basically symmetric	Long right tail
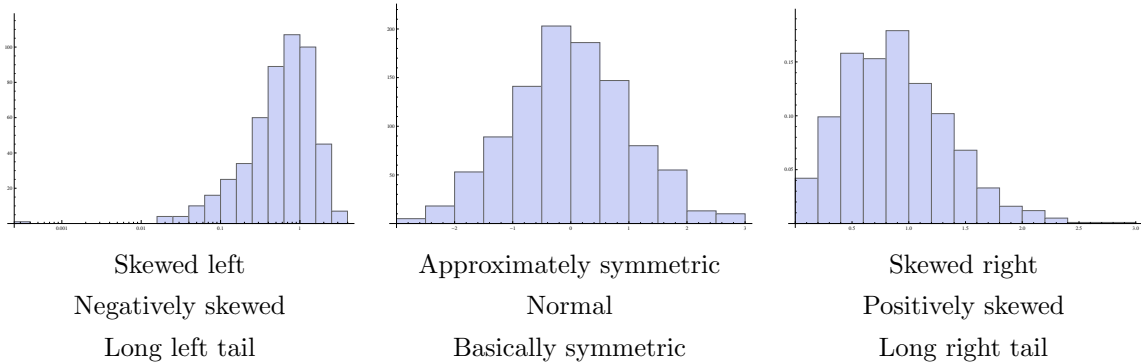
- A peak, or high point, of a histogram is referred to as a **mode**. A histogram is

  - **unimodal** if it has one mode.
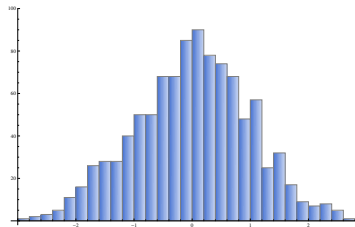
  - **bimodal** if it has two clearly distinct modes.
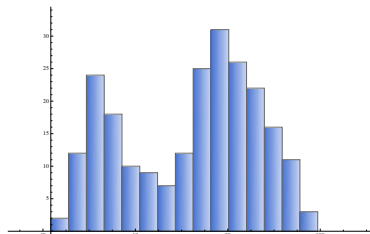
Frequency Polygons and Ogives

Definitions:

- The **midpoint** of a class is the average of its lower class limit and the lower class limit of the next class.

$$\text{Class midpoint} = \frac{\text{Lower limit} + \text{Higher limit}}{2}$$

- The **cumulative frequency** of a class is the sum of the frequencies of that class and all previous classes.

14

- The **cumulative Percentage** of a class is the cumulative frequency divided by the sum of all frequencies.

$$\text{Cumulative Percentage} = \frac{\text{Cumulative frequency}}{\text{Sum of all frequencies}}$$

- In a **frequency polygon**, a line graph is drawn by joining all the midpoints of the top of the bars of a histogram used for understanding the shapes of distributions (plotted midpoints vs. frequency or Percentage).

- An **ogive** is the distribution curve of a frequency distribution (plotted upper class limits v. cumulative frequency or cumulative Percentage).
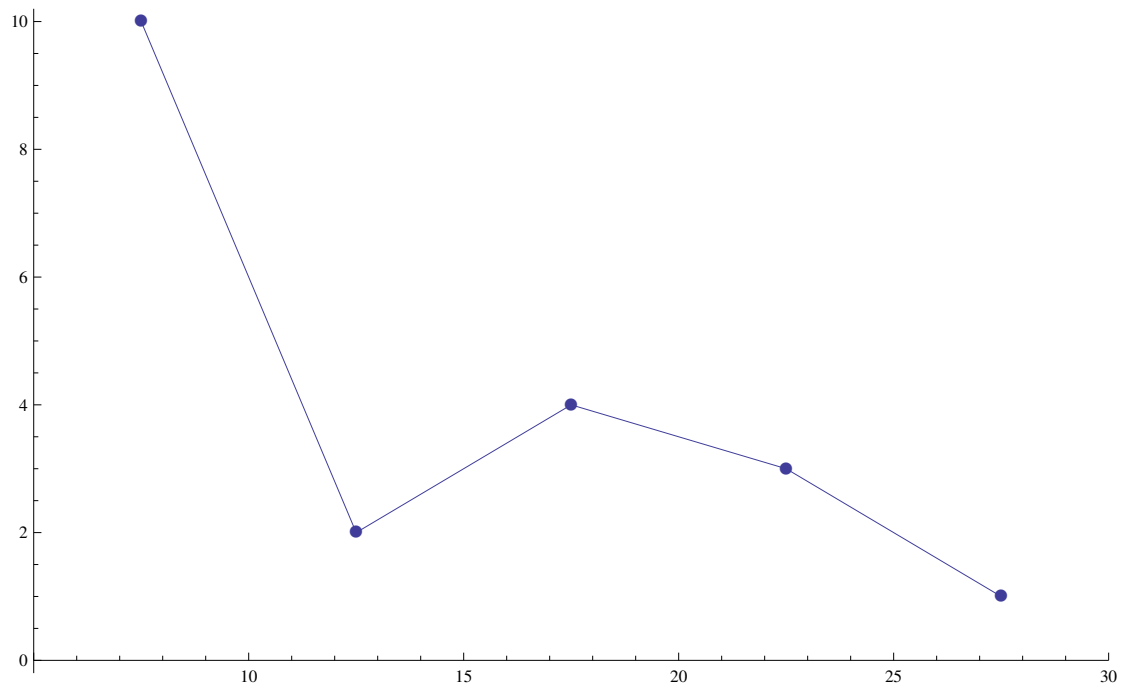

Example: Construct a frequency polygon and a Percentage ogive for the following data set:

$$5 \quad 10 \quad 7 \quad 19 \quad 25 \quad 12 \quad 15 \quad 7 \quad 6 \quad 8$$
$$17 \quad 17 \quad 22 \quad 21 \quad 7 \quad 7 \quad 24 \quad 5 \quad 6 \quad 5$$

Solution:

| Class	Midpoint	Frequency	Percentage	Cumulative Frequency	Cumulative Regular Frequency
5-9	$\frac{5+10}{2} = 7.5$	10	$\frac{10}{20} = 0.50$	10	$\frac{10}{20} = 0.50$
10-14	$\frac{10+15}{2} = 12.5$	2	$\frac{2}{20} = 0.10$	$10 + 2 = 12$	$\frac{12}{20} = 0.60$
15-19	$\frac{15+20}{2} = 17.5$	4	$\frac{4}{20} = 0.20$	$12 + 4 = 16$	$\frac{16}{20} = 0.80$
20-24	$\frac{20+25}{2} = 22.5$	3	$\frac{3}{20} = 0.15$	$16 + 3 = 19$	$\frac{19}{20} = 0.95$
25-29	$\frac{25+30}{2} = 27.5$	1	$\frac{1}{20} = 0.05$	$19 + 1 = 20$	$\frac{20}{20} = 1.00$

Frequency Polygon:

Percentage Ogive:
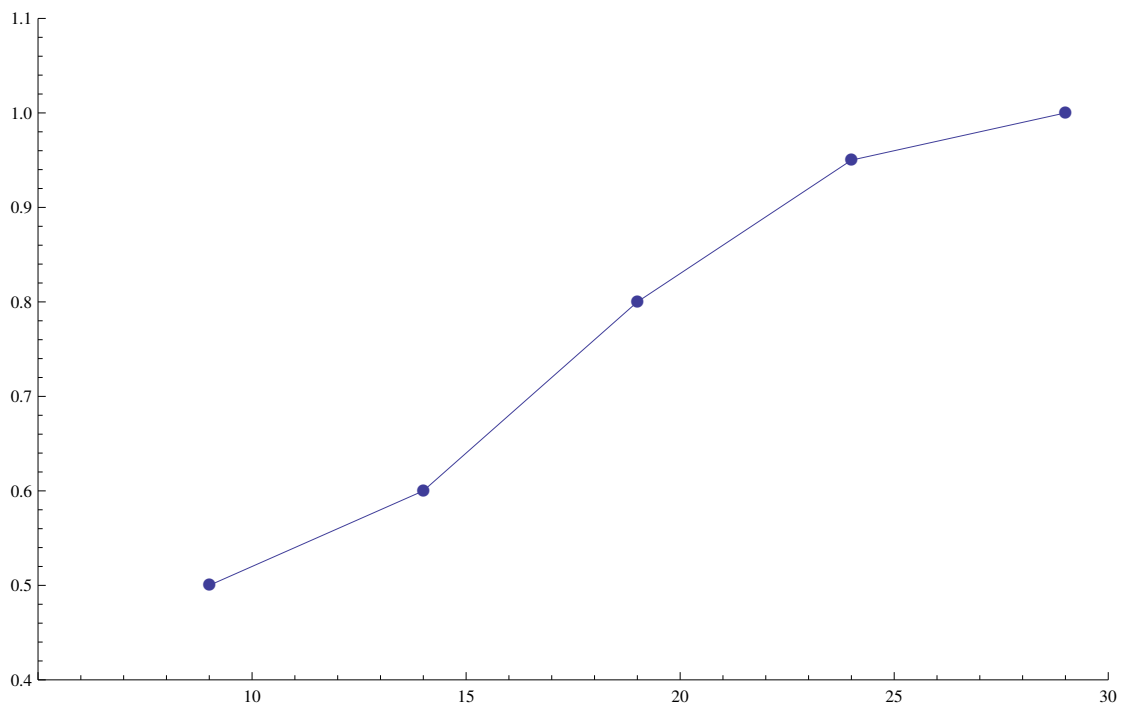
On Your Own: The following table presents the purchase totals (in dollars) of a random sample of gasoline purchases at a convenience store. Construct a Percentage polygon and a Percentage ogive for the following data set: using a class width of 10, and using 0 as the lower class limit for the first class.

76.59	48.55	93.66	60.17	39.10
93.28	65.43	34.12	80.41	77.16
80.07	93.46	39.19	43.84	44.70
68.74	89.98	6.97	52.86	68.93

Solution:

Class	Midpoint	Frequency	Percentage	Cumulative Frequency	Cumulative Regular Frequency
0.00-9.99	5.00	1	0.05	1	0.05
10.00-19.99	15.00	0	0.00	1	0.05
20.00-29.99	25.00	0	0.00	1	0.05
30.00-39.99	35.00	3	0.15	4	0.20
40.00-49.99	45.00	3	0.15	7	0.35
50.00-59.99	55.00	1	0.05	8	0.40
60.00-69.99	65.00	4	0.20	12	0.60
70.00-79.99	75.00	2	0.10	14	0.70
80.00-89.99	85.00	3	0.15	17	0.85
90.00-99.99	95.00	3	0.15	20	1.00

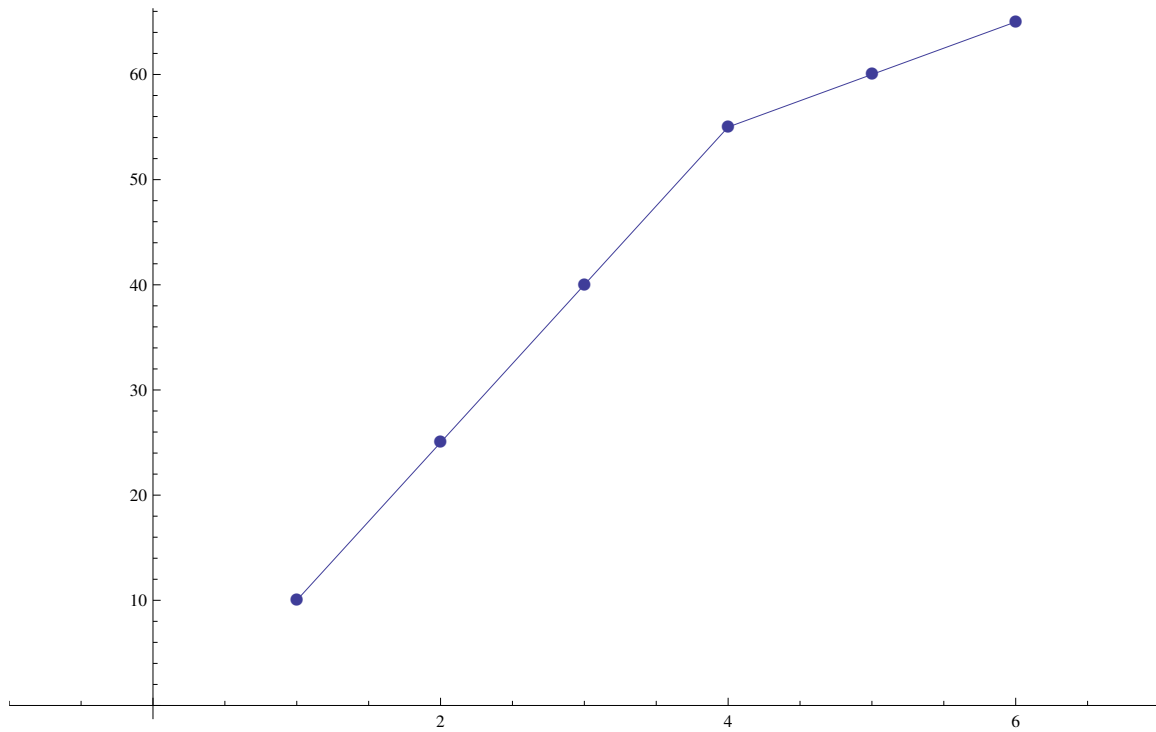Percentage Polygon:
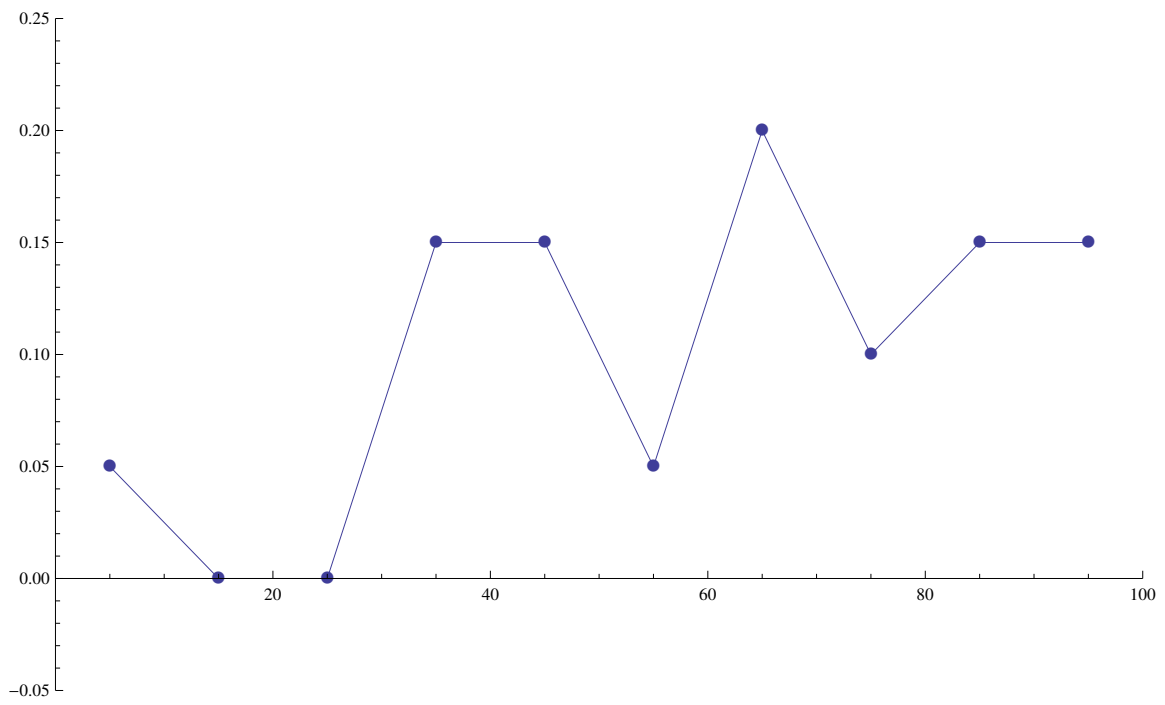
Frequency Ogive:
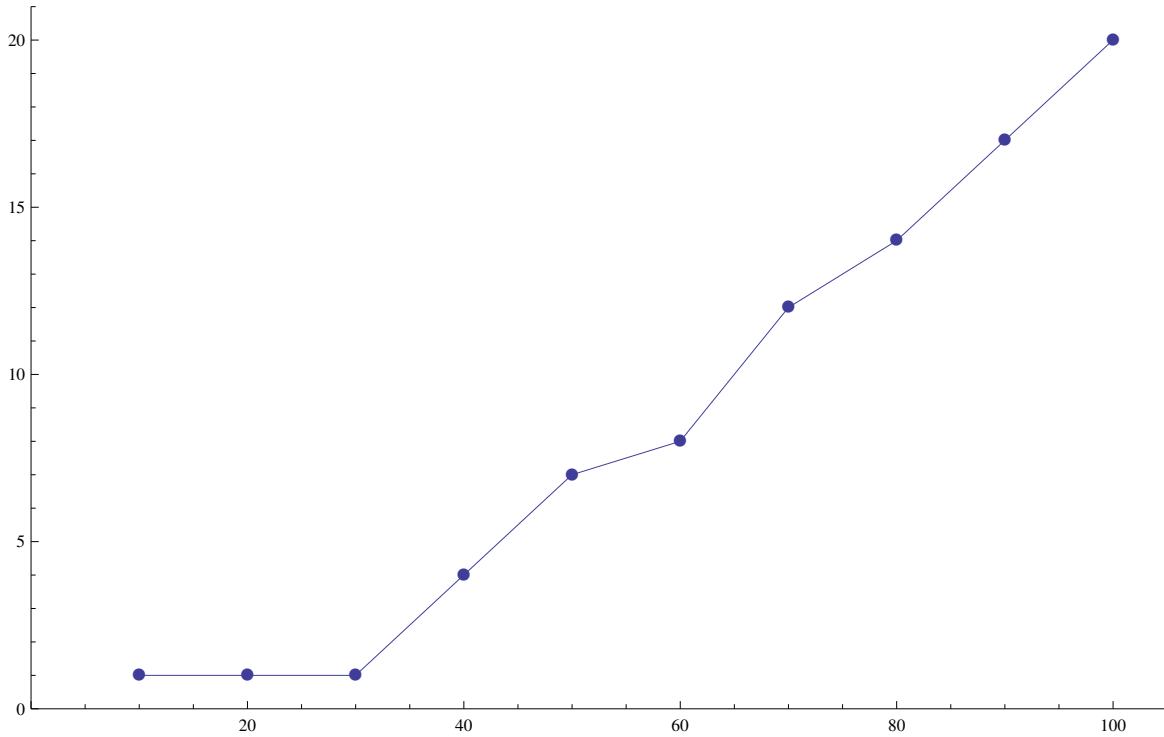
Time-Series Plots

Definition:

- A **time-series plot** is a display setting for a variable measured at different points in time.

Example: This data set gives number of staff in an organization. Draw a time-series plot of the data.

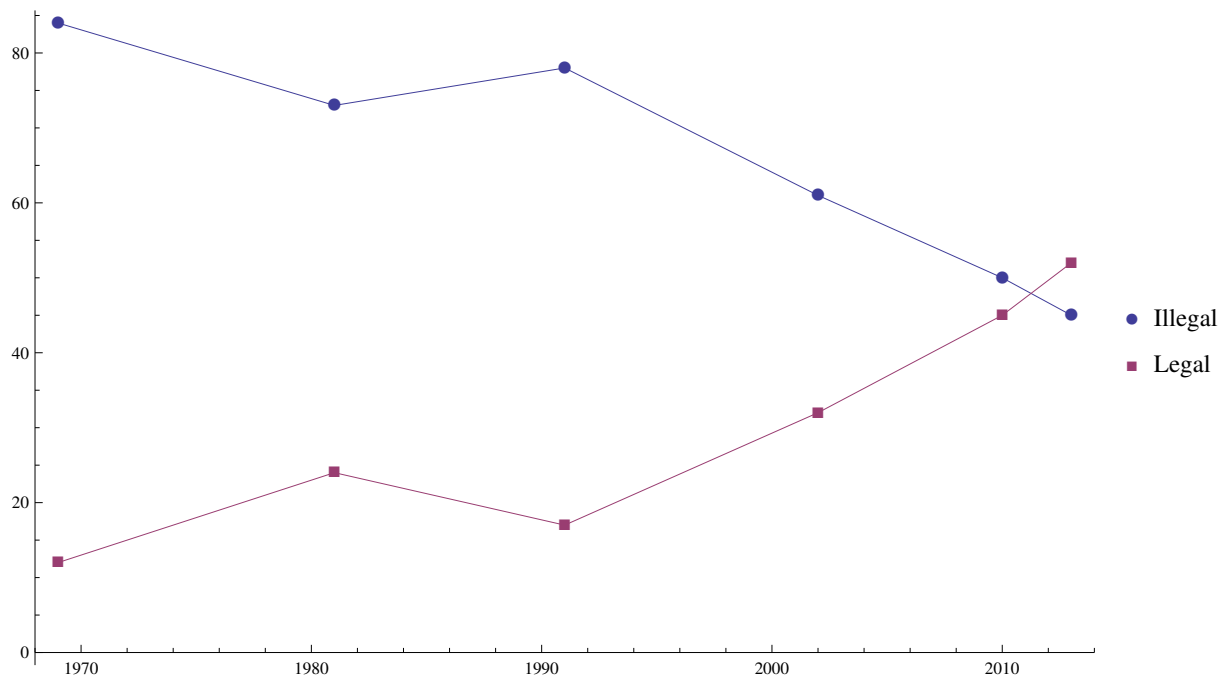| Year	1	2	3	4	5	6
Total	10	25	40	55	60	65

Solution:

On Your Own: This data set provides percentage of the US populous and their views on the legalization of marijuana between the years of 1969-2013. Draw a time-series plot of the data.

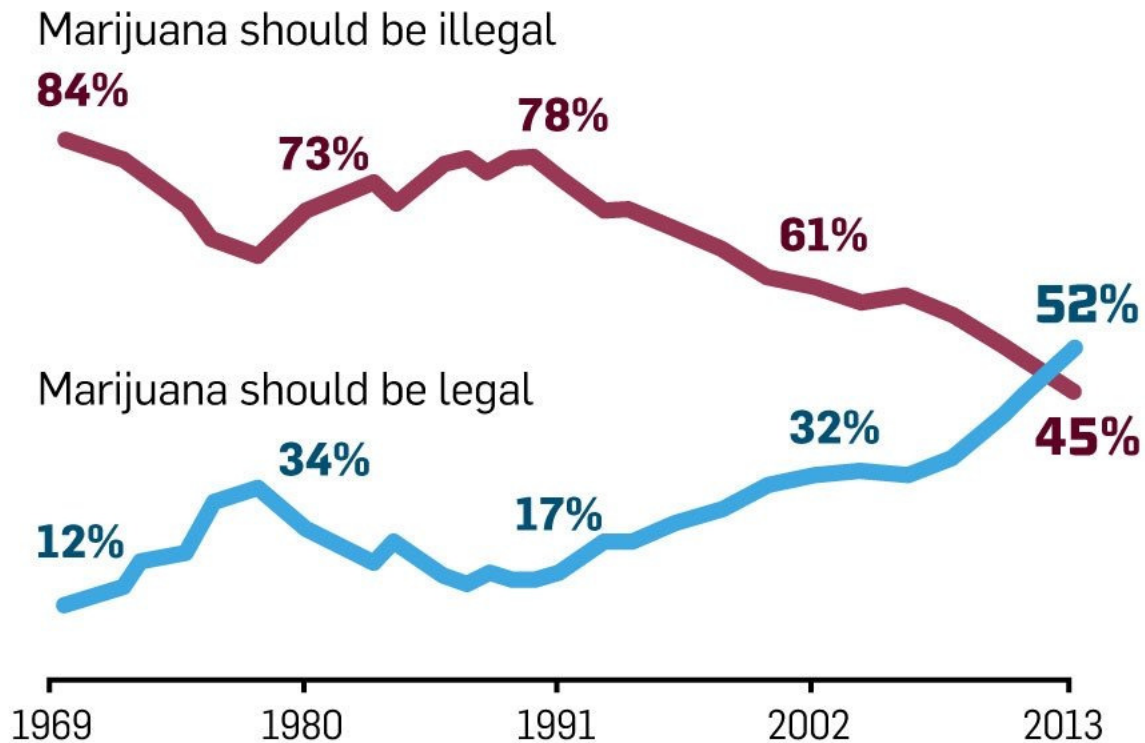| Year	1969	1981	1991	2002	2010	2013
Illegal	84	73	78	61	50	45
Legal	12	24	17	32	45	52

Solution:

Actual time series plot:

Pot's pirouette

For the first time, a majority of Americans say marijuana should be legal — and not just for medical purposes.

Marijuana should be illegal

84%

73%

78%

61%

52%

Marijuana should be legal

34%

32%

45%

12%

17%

1969   1980   1991   2002   2013

Source: Pew Research Center poll of 1,501 adults, conducted March 13 to 17, with a margin of error of plus or minus 2.9 percentage points.

THE STAR-LEDGER

2.4: Graphs Can Be Misleading

In Class Collaborative Learning Activity

References

Friedman, L. (2013, April 23). *iPad sales up, Macs flat, as Apple reports on q2 earnings.* Retrieved from http://www.macworld.com/article/2036232/ipad-sales-up-macs-flat-as-apple-reports-on-q2-earnings.html

Gordon, M. E. (2013, March 5). *Are indie app developers becoming an endangered species?.* Retrieved from http://blog.flurry.com/bid/94811/Are-Indie-App-Developers-Becoming-an-Endangered-Species

Khalaf, R. (2013, June 13). *A day in the life of a mobile consumer.* Retrieved from http://blog.flurry.com/bid/98080/A-Day-in-the-Life-of-a-Mobile-Consumer

Quinn, T. (2013, July 13). *How student loan debt factors into your FICO score.* Retrieved from http://blog.myfico.com/2013/07/how-student-loan-debt-factors-into-your-fico-score/

Raja, T. Y. (2014, January 17). *America's coming geek gap.* Retrieved from http://www.motherjones.com/media/2014/01/charts-ap-computer-science-2013

Stop Global Warming: Become a Pirate [Graphic]. Retrieved January 19, 2014, from: http://www.venganza.org/images/spreadword/pchart1.jpg