Graphs That Enlighten and Graphs That Deceive

Colby Community College

Definition

A **dotplot** consists of a graph of quantitative data in which each data value is plotted as a point above a horizontal scale of values. Dots representing equal values are stacked.

Definition

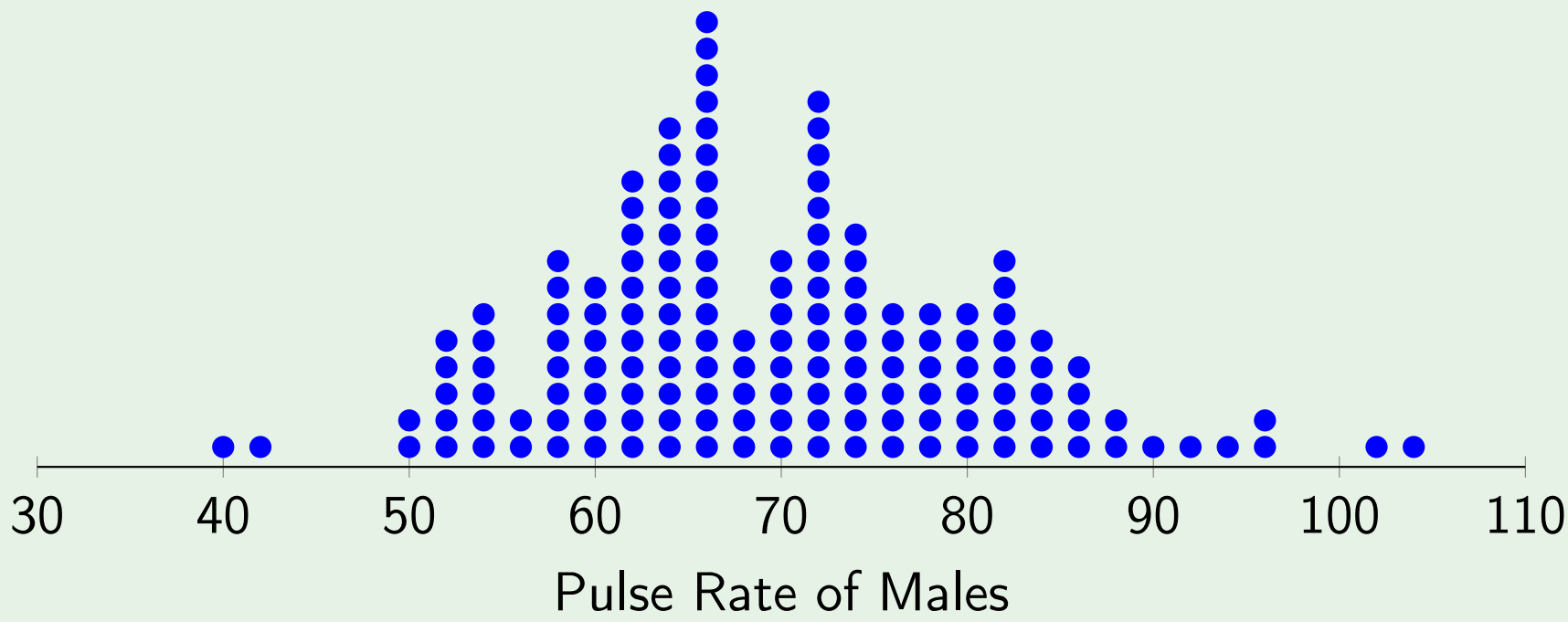
A **dotplot** consists of a graph of quantitative data in which each data value is plotted as a point above a horizontal scale of values. Dots representing equal values are stacked.

Features

- Displays the shape of the distribution of data.
- It is usually possible to recreate the original list of data values.

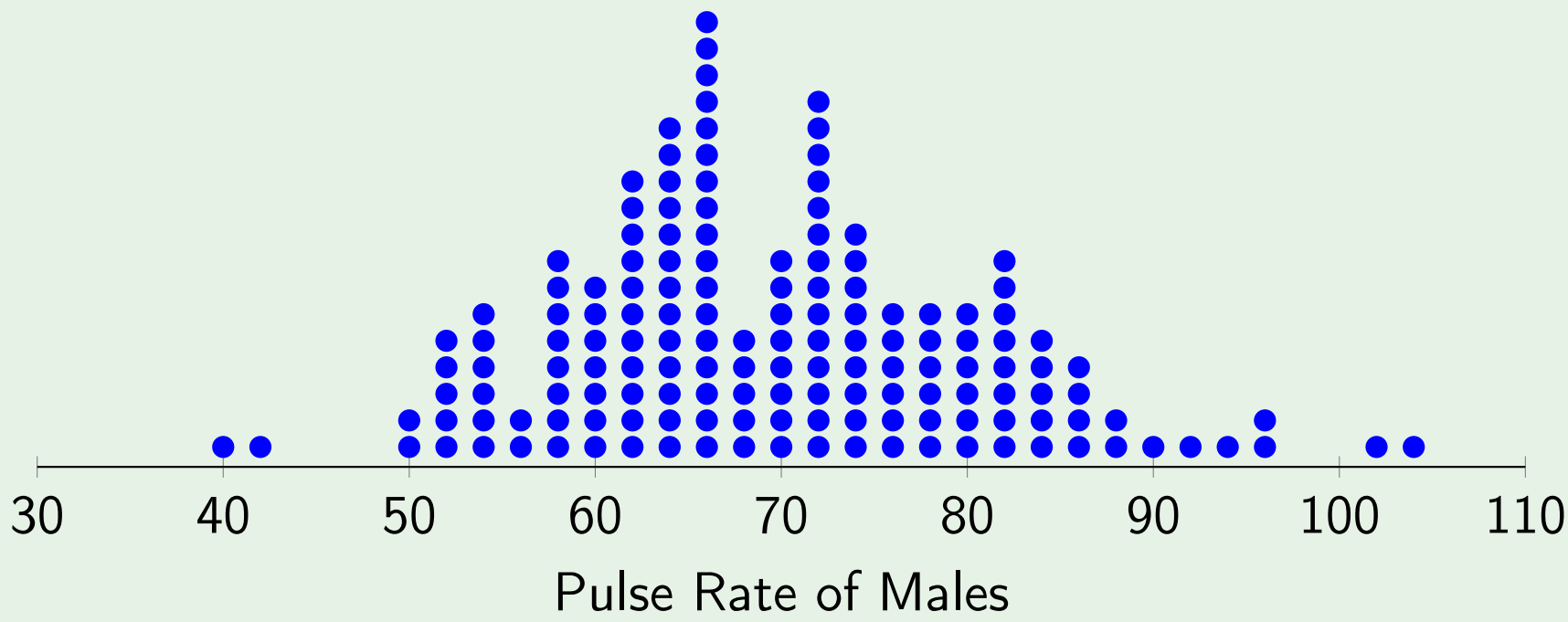
Example 1

The dotplot shows the pulse rates (beats per minute) of males.
(Data Set 1 in Appendix B.)



Example 1

The dotplot shows the pulse rates (beats per minute) of males.
(Data Set 1 in Appendix B.)



Note

A Histogram counts how many data values fall within an interval.
A Dotplot counts individual data points.

Definition

A **stemplot** (or **stem-and-leaf plot**) represents quantitative data by separating each value into two parts:

The Stem: Usually the leftmost digit.

The Leaf: Usually the rightmost digits.

Definition

A **stemplot** (or **stem-and-leaf plot**) represents quantitative data by separating each value into two parts:

The Stem: Usually the leftmost digit.

The Leaf: Usually the rightmost digits.

Note

You may have to round data values first.

Definition

A **stemplot** (or **stem-and-leaf plot**) represents quantitative data by separating each value into two parts:

The Stem: Usually the leftmost digit.

The Leaf: Usually the rightmost digits.

Note

You may have to round data values first.

Features

- Shows the shape and distribution of the data.
- Retains the original data values.
- The sample data are sorted.

Example 2

The stemplot shows the pulse rates (beats per minute) of males. (Data Set 1 in Appendix B.)

4	02
5	0022222444446688888888
6	000000022222222224444444444466666666666666688888
7	00000000222222222222444444444666666888888
8	000000222222224444666688
9	02466
10	24

The left is the rightmost digits, the stem is all the remaining leftmost digits. The stems are listed in increasing order, not the order in which they occur in the dataset.

Definition

A **time series graph** is a graph of quantitative data that have been collected at different points in time.

Definition

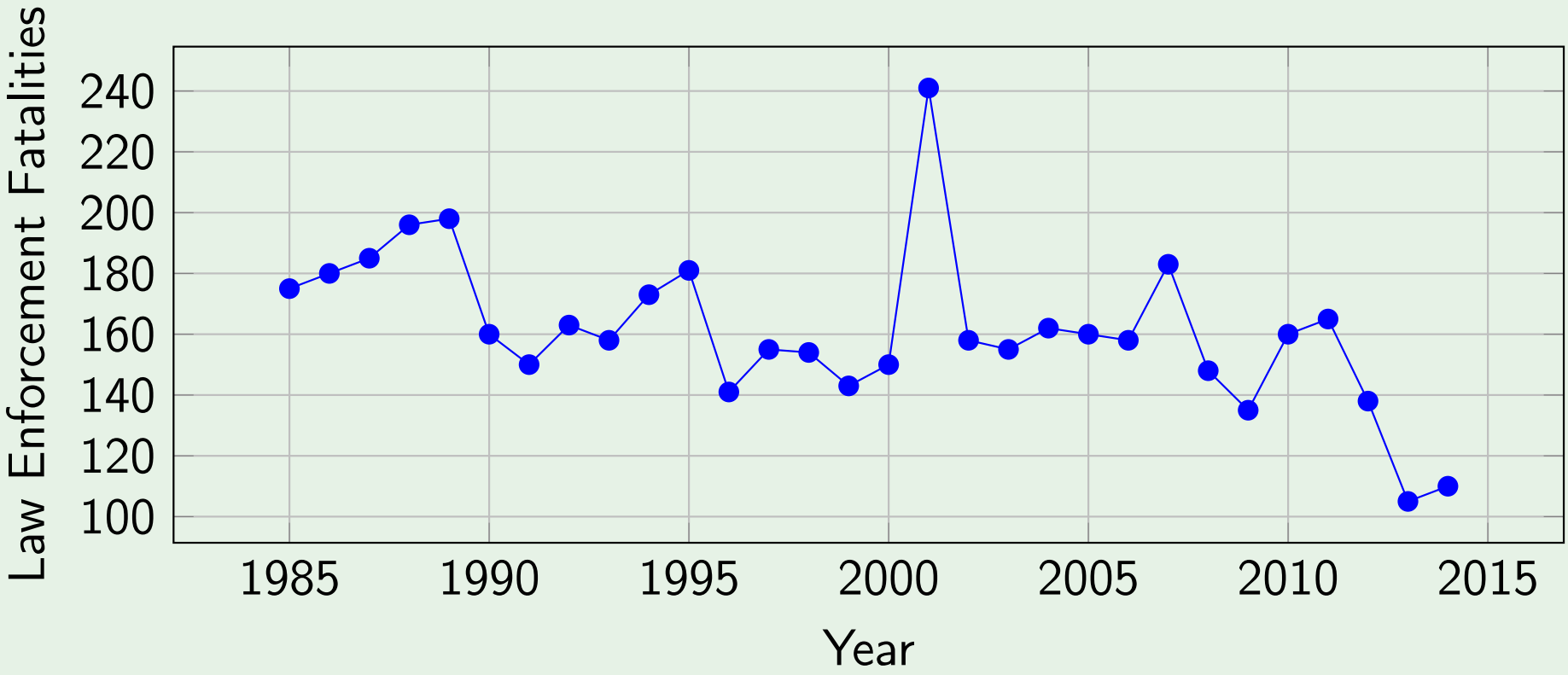
A **time series graph** is a graph of quantitative data that have been collected at different points in time.

Features

- Reveals information about trends over time.

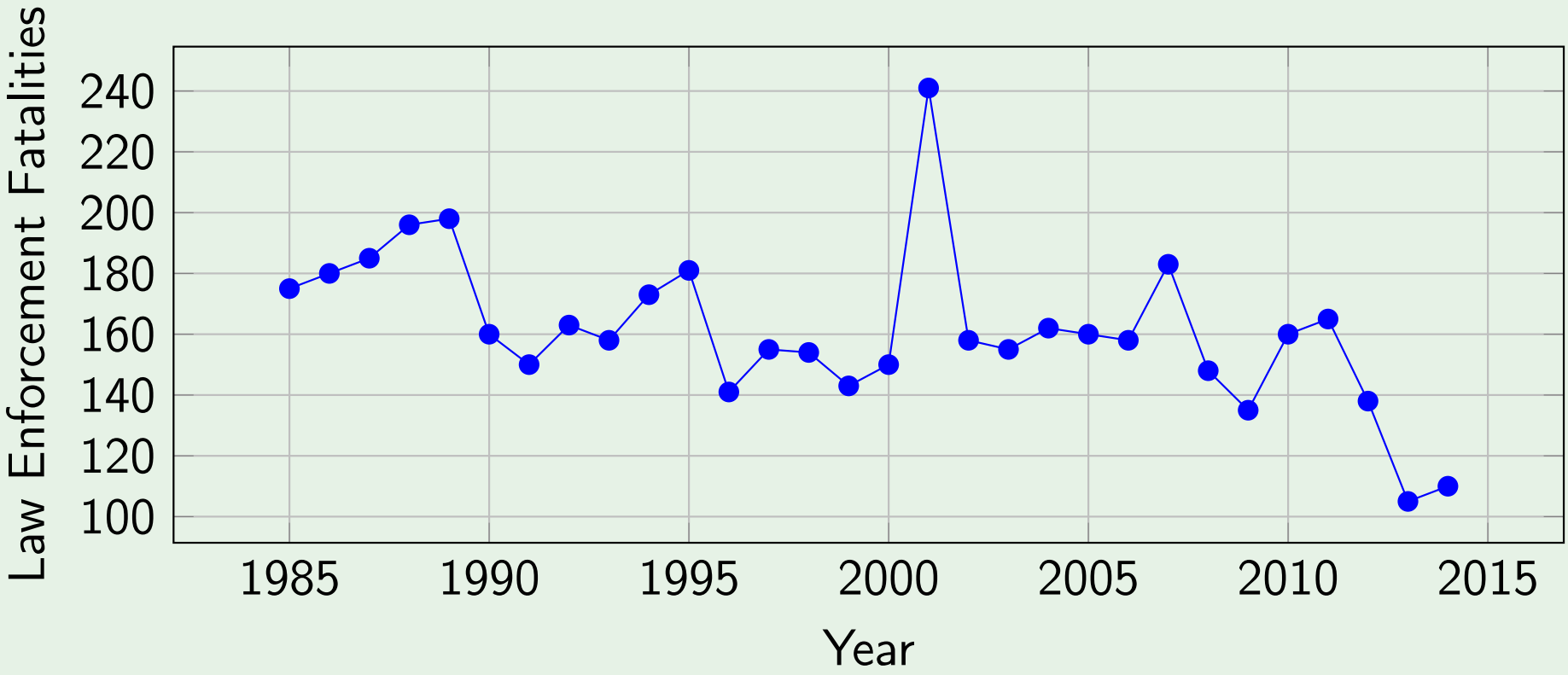
Example 3

The time-series graph depicts the yearly number of fatalities of law enforcement officers in the United States.



Example 3

The time-series graph depicts the yearly number of fatalities of law enforcement officers in the United States.



Notice that there is a large spike in 2001, most of these fatalities would have been during the terrorist attacks on September 11, 2001. If we exclude the spike, there appears to be a slight downward trend.

Definition

A **bar graph** uses bars of equal width to show frequencies of categories or categorical data. The bars may or may not be separated by small gaps.

Definition

A **bar graph** uses bars of equal width to show frequencies of categories or categorical data. The bars may or may not be separated by small gaps.

Features

- Shows the relative distribution of categorical data so that it is easier to compare the different categories.

Definition

A **bar graph** uses bars of equal width to show frequencies of categories or categorical data. The bars may or may not be separated by small gaps.

Features

- Shows the relative distribution of categorical data so that it is easier to compare the different categories.

Definition

A **Pareto chart** is a bar graph for categorical data, with the added stipulation that the bars are arranged in descending order according to the frequencies, so the bars decrease in height from left to right.

Definition

A **bar graph** uses bars of equal width to show frequencies of categories or categorical data. The bars may or may not be separated by small gaps.

Features

- Shows the relative distribution of categorical data so that it is easier to compare the different categories.

Definition

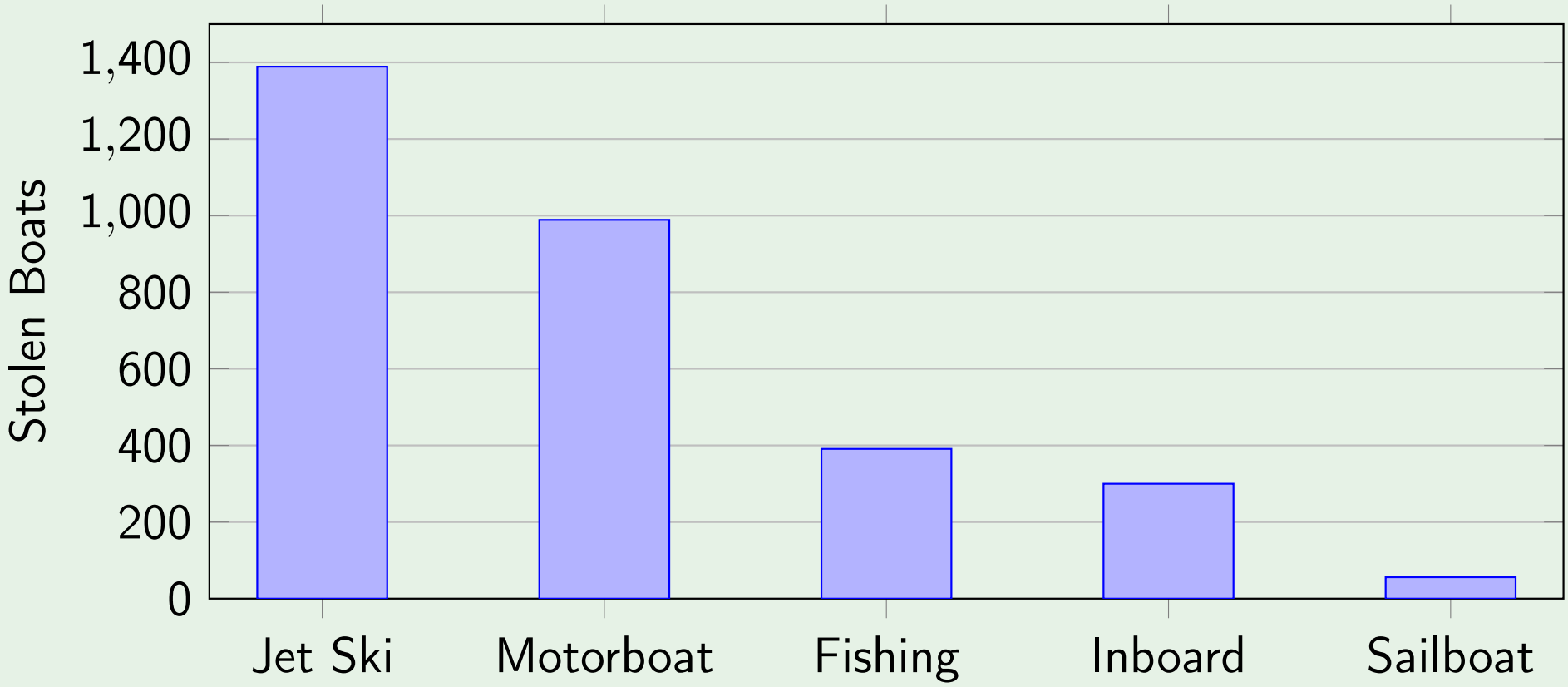
A **Pareto chart** is a bar graph for categorical data, with the added stipulation that the bars are arranged in descending order according to the frequencies, so the bars decrease in height from left to right.

Features

- Shows the relative distribution of categorical data so that it is easier to compare the different categories.
- Draws attention to the more important categories.

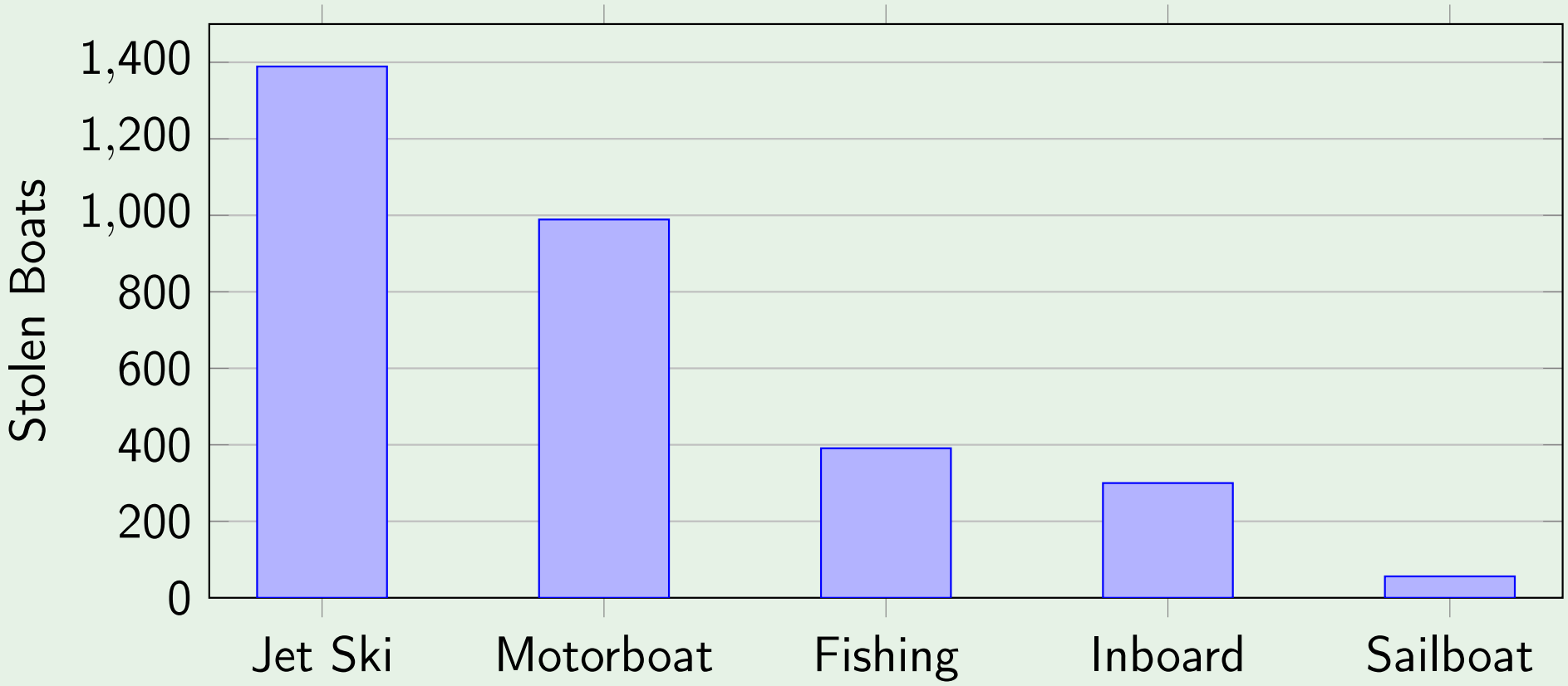
Example 4

For the boats stolen in a recent year, the bar graph shows the types most often stolen.



Example 4

For the boats stolen in a recent year, the bar graph shows the types most often stolen.



Notice that for boat thefts, jet skis are the worst problem.

Definition

A **frequency polygon** uses line segments connected to points located directly above class midpoint values.

Definition

A **frequency polygon** uses line segments connected to points located directly above class midpoint values.

Definition

A **relative frequency polygon** uses relative frequencies for the vertical scale.

Definition

A **frequency polygon** uses line segments connected to points located directly above class midpoint values.

Definition

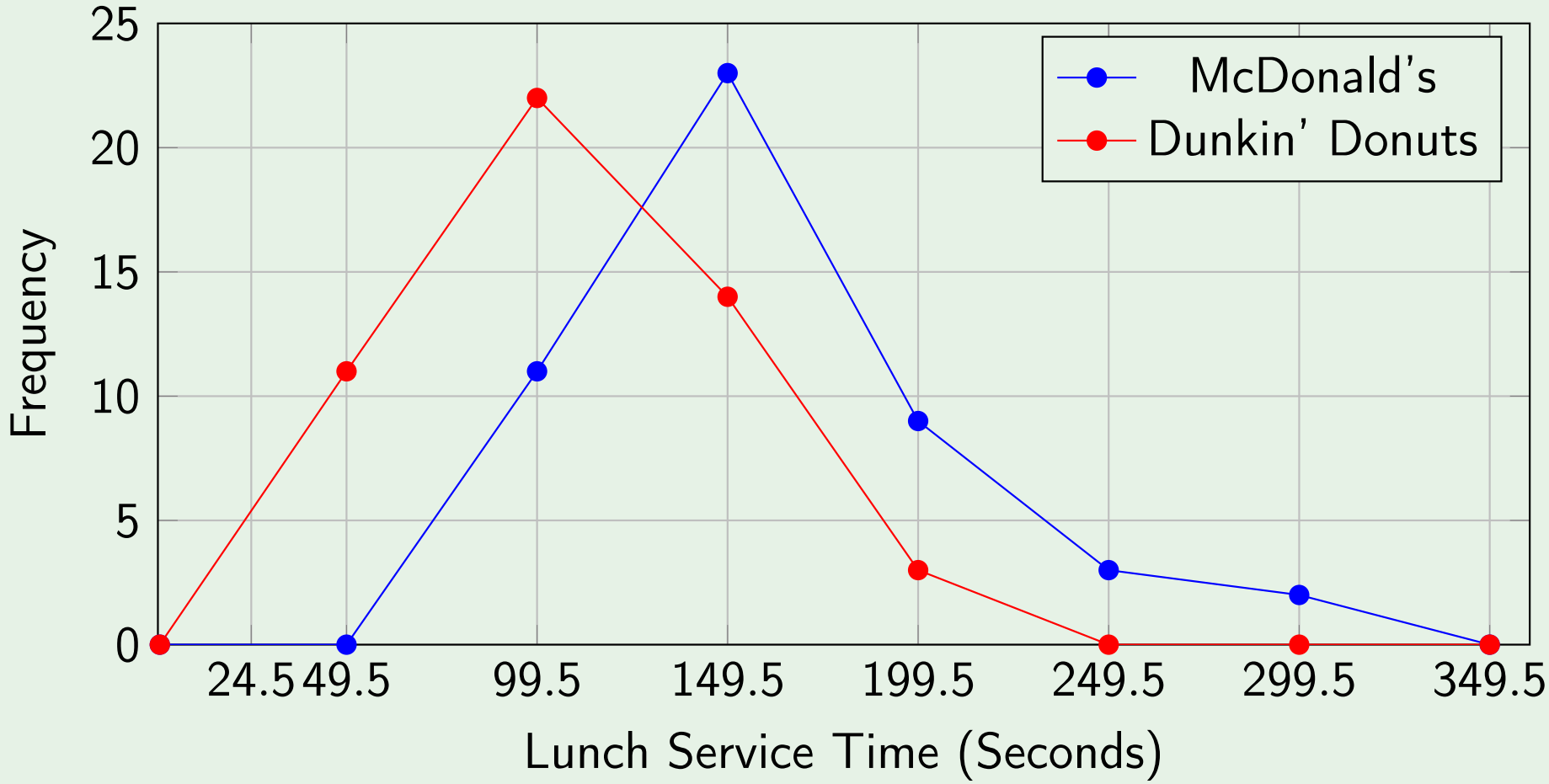
A **relative frequency polygon** uses relative frequencies for the vertical scale.

Features

- Frequency polygons make it easy to compare two data sets.

Example 5

The plot shows the wait times for both McDonald's and Dunkin' Donuts.



Pie Charts

Never use pie charts.

Not only are they a waste of ink, they lack an appropriate scale.

Pie Charts

Never use pie charts.

Not only are they a waste of ink, they lack an appropriate scale.

Note

A Pareto chart will depict the same data, but better.

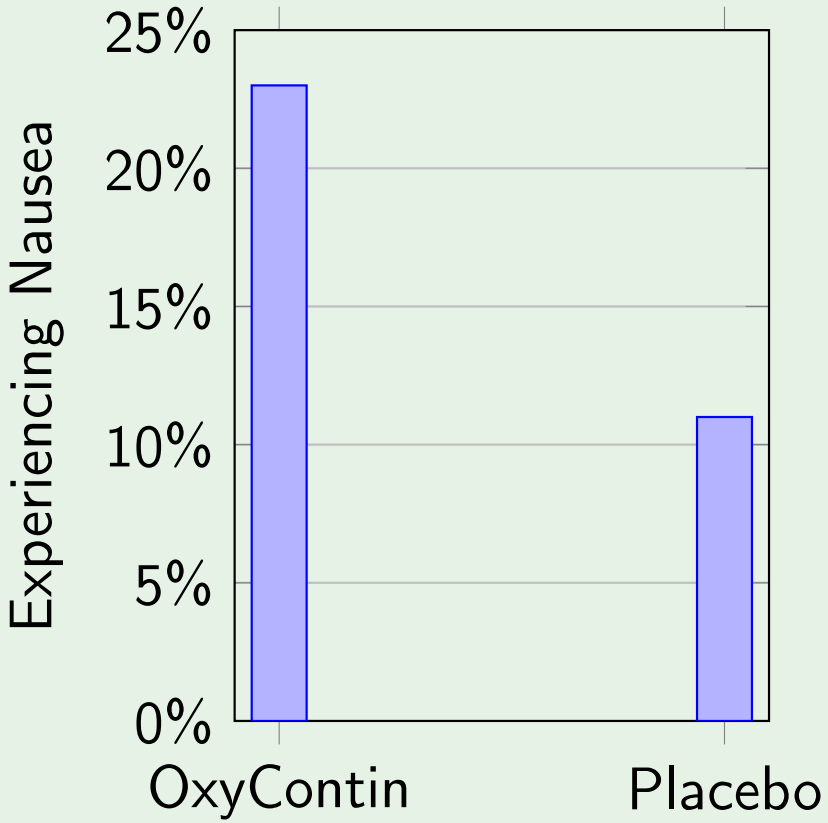
Nonzero Vertical Axis

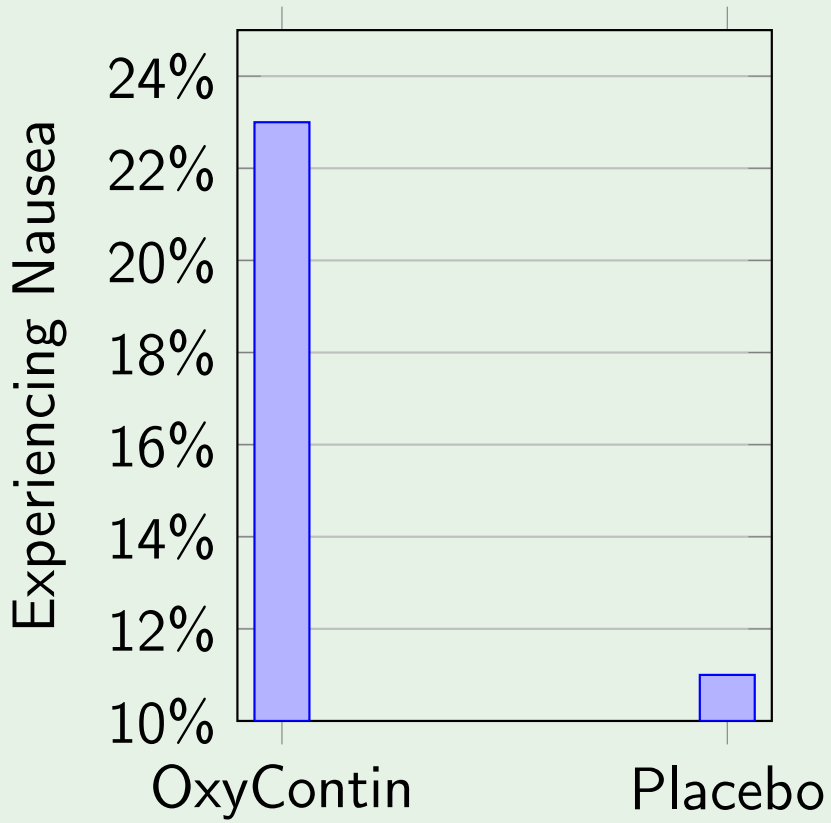
Always examine a graph carefully to see whether a vertical axis begins at some point other than zero so that differences are exaggerated.

Nonzero Vertical Axis

Always examine a graph carefully to see whether a vertical axis begins at some point other than zero so that differences are exaggerated.

Example 6





Pictographs

When examining data depicted with a pictograph, determine whether the graph is misleading because objects of area or volume are used to depict amounts that are actually one-dimensional.

Pictographs

When examining data depicted with a pictograph, determine whether the graph is misleading because objects of area or volume are used to depict amounts that are actually one-dimensional.

Example 7

The pictographs show data from the CDC.

(a) 1970: 37% of U.S. adults smoked.

(b) 2013: 18% of U.S. adults smoked.

The larger cigarette is about twice as long as the smaller, which means it has four times the area of the smaller cigarette. While the data shows that 37% is about twice of 18%.

Final Notes

- For small data sets of 20 values or fewer, use a table instead of a graph.

Final Notes

- For small data sets of 20 values or fewer, use a table instead of a graph.
- A graph of data should make us focus on the true nature of the data, not on other elements, such as eye-catching but distracting design features.

Final Notes

- For small data sets of 20 values or fewer, use a table instead of a graph.
- A graph of data should make us focus on the true nature of the data, not on other elements, such as eye-catching but distracting design features.
- Do not distort data. Construct a graph to reveal the true nature of the data.

Final Notes

- For small data sets of 20 values or fewer, use a table instead of a graph.
- A graph of data should make us focus on the true nature of the data, not on other elements, such as eye-catching but distracting design features.
- Do not distort data. Construct a graph to reveal the true nature of the data.
- Almost all of the ink in a graph should be used for data, not for other design elements.