

## Tornado and Butterfly Plots



Revised: 10/9/2017



Summary .....	1
Sample Data .....	2
Data Input.....	3
Analysis Summary .....	4
Tornado Plot.....	4
Butterfly Plot.....	5
Analysis Options.....	5

### Summary

The **Tornado and Butterfly Plots** procedure creates two similar plots that compare 2 samples of attribute data. Each plot consists of 2 sets of bars that show the frequency distribution of each sample over a set of categories. The only difference between the plots is where the labels are placed.

**Sample StatFolio:** *tornadoplot.sgp*

## Sample Data

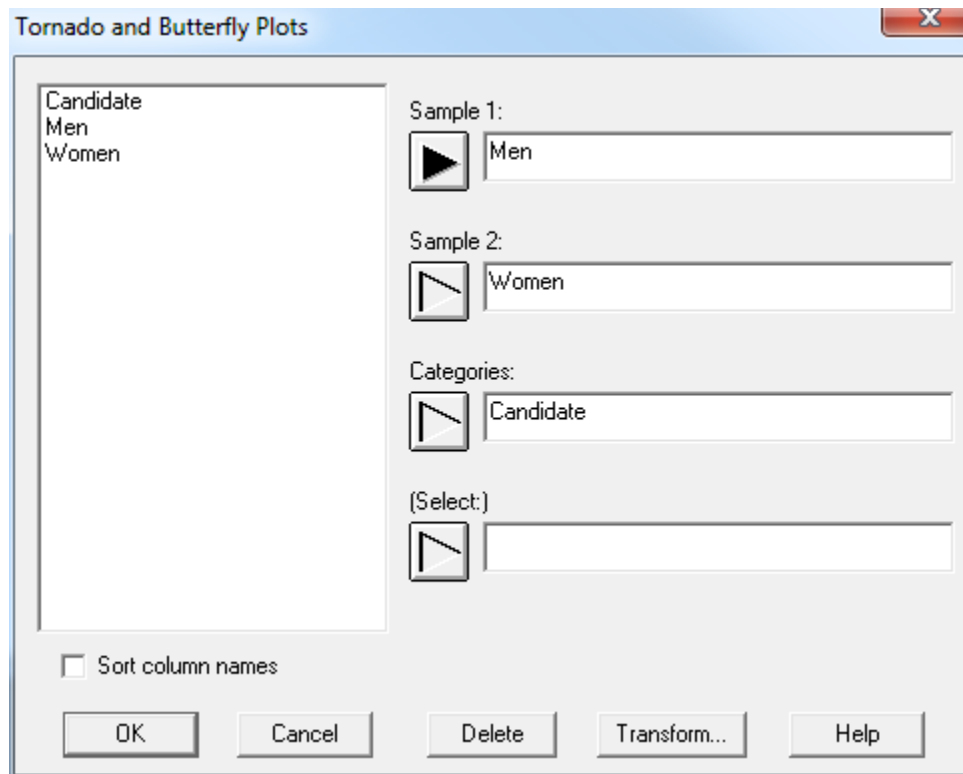
The file *Republican poll.sgd* contains the results of a poll taken by Quinnipiac University during the 2016 nomination process to determine the Republican Party's candidate for president of the United States. At the time, 6 candidates were vying for the nomination. The data show the percentage of likely voters who expressed a preference for each candidate, categorized by gender:

<i>Candidate</i>	<i>Men</i>	<i>Women</i>
Bush	2%	5%
Carson	2%	5%
Cruz	21%	16%
Kasich	6%	6%
Rubio	15%	22%
Trump	45%	34%

The goal of the plots is to illustrate the differences between the data in the 2 rightmost columns.

## Data Input

The data to be analyzed consist of 2 numeric columns and a categorical column identifying each row.



- **Sample 1:** numeric column containing the data for the first sample. The data normally consists of counts or percentages, although any set of non-negative values may be plotted.
- **Sample 2:** numeric column containing the data for the second sample.
- **Categories:** a numeric or non-numeric column identifying each row. The entries in this column are used as labels on the plots.
- **Select:** subset selection.

## Analysis Summary

The Analysis Summary displays the data for the 2 samples.

**Tornado and Butterfly Plots - Men & Women**  
 Sample 1: Men  
 Sample 2: Women

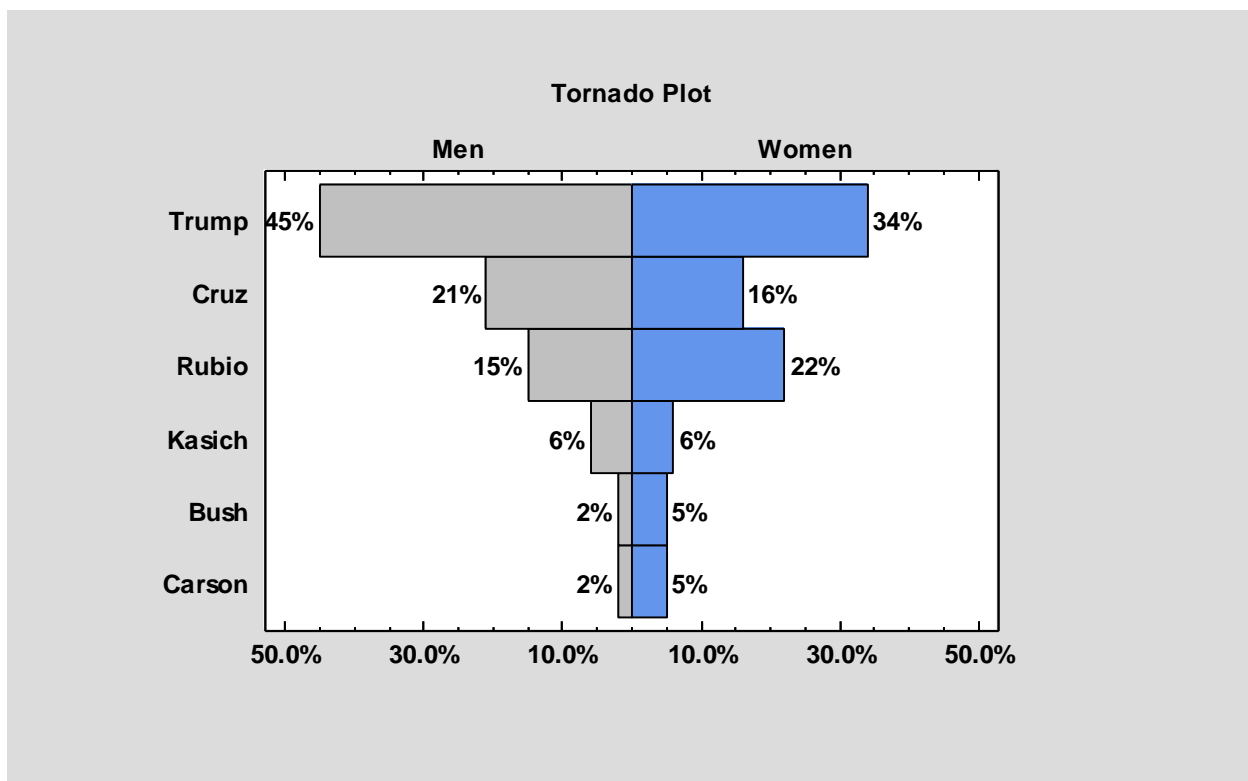
**Analysis Summary**

Candidate	Men	Women
Trump	45.0%	34.0%
Cruz	21.0%	16.0%
Rubio	15.0%	22.0%
Kasich	6.0%	6.0%
Bush	2.0%	5.0%
Carson	2.0%	5.0%

The order of the rows is controlled by the *Analysis Options* dialog box.

## Tornado Plot

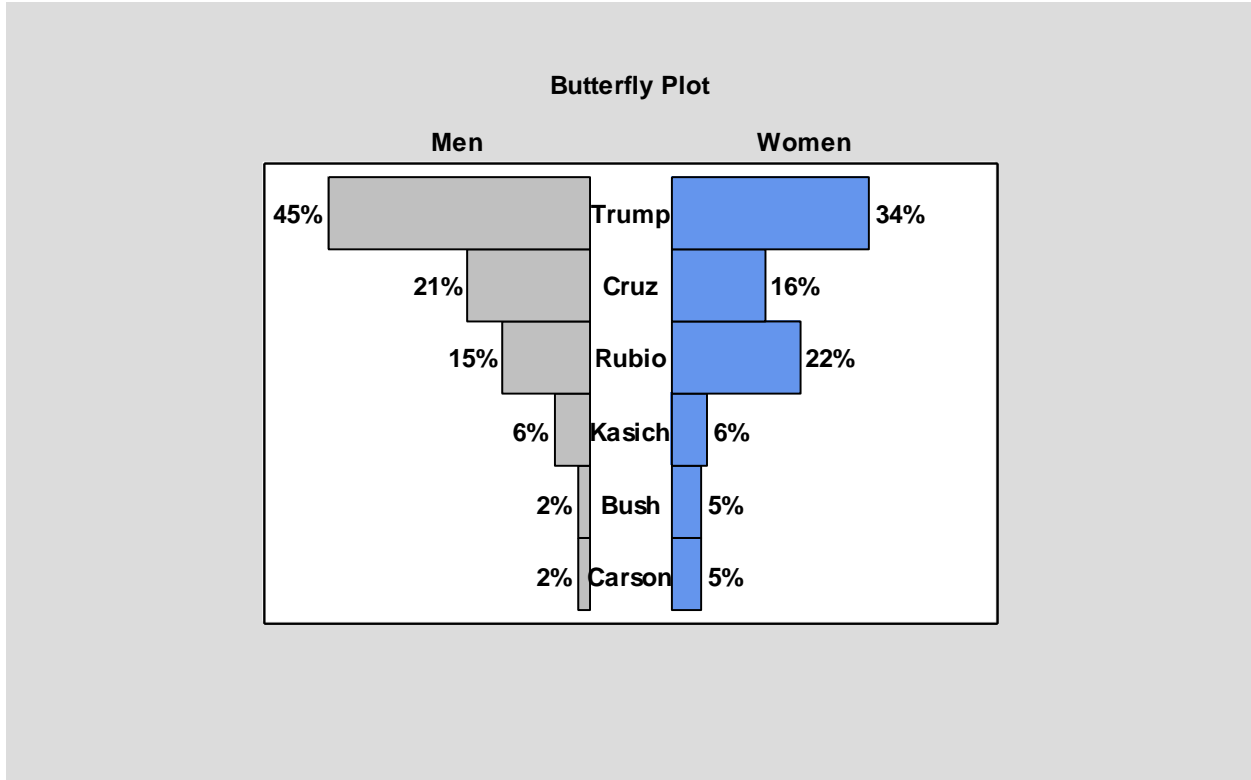
The tornado plot displays the data with bars that extend left and right of a centerline:



Category labels are displayed in the left margin of the plot.

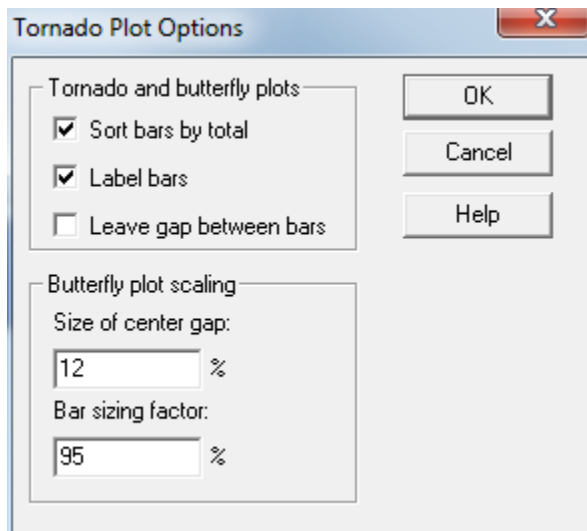
## Butterfly Plot

The butterfly plot displays the data with bars that extend left and right, leaving room for category labels in the middle:



## Analysis Options

The Analysis Options dialog box lets users control various aspects of the output:



- **Sort bars by total:** if selected, the categories are sorted in decreasing order of the sum of the 2 data columns.
- **Label bars:** if selected, labels are added at the end of each bar showing the data values.
- **Leave gap between bars:** if selected, a small gap is left between bars representing each row.
- **Size of center gap:** for the butterfly plot only, this field controls the amount of space left in the center of the plot for the category labels.
- **Bar sizing factor:** for the butterfly plot only, this field controls the relative length of the bars. Smaller values may be used to prevent bar labels from falling outside of the main plotting area.