

Capitalizing on the Power of Dynamic Data Visualization

Neil W. Polhemus

Statpoint Technologies, Inc., P.O. Box 134, The Plains, VA 20198

neil@statpoint.com

Abstract

Dynamic graphs are extremely effective at extracting information from data in ways that provide the viewer with insights that could not be obtained from static images. Dynamic data visualization is especially useful for illustrating changes over time, for changing the viewpoint at which 3D images are displayed, and for demonstrating the impact of changing model parameters. This paper demonstrates the power of dynamic data visualization by showing several Statlets, which are dynamic graphs controlled by user interactions. Examples include dynamic demographic maps for visualizing multiple time series, dynamic population pyramids and wind rose charts, multivariate bubble charts, bivariate density estimators, and response surfaces.

Key Words: data visualization, dynamic graphs, statistical graphics, Statlets

1. Dynamic Data Visualization

The addition of dynamic features to statistical graphs adds a level of visualization not obtainable by static displays. Dynamic features are particularly important in showing:

- Changes over time.
- Sensitivity of models to estimated parameters.
- Multidimensional data.

Given modern technology, dynamic graphs can be imbedded in PowerPoint slides and other media.

This paper shows various graphs that are greatly enhanced by the addition of animation and interactive control. While displayed in a static format here, the dynamic version is viewable at www.statgraphics.com/statistics-library.

1.1 Dynamic Demographic Maps

Maps are commonly used to display demographic data. For example, Figure 1 shows the crime rates in each of the 50 states and the District of Columbia in 1965. The amount of crime in each state is coded using a gradient scale that ranges from dark blue to dark red.

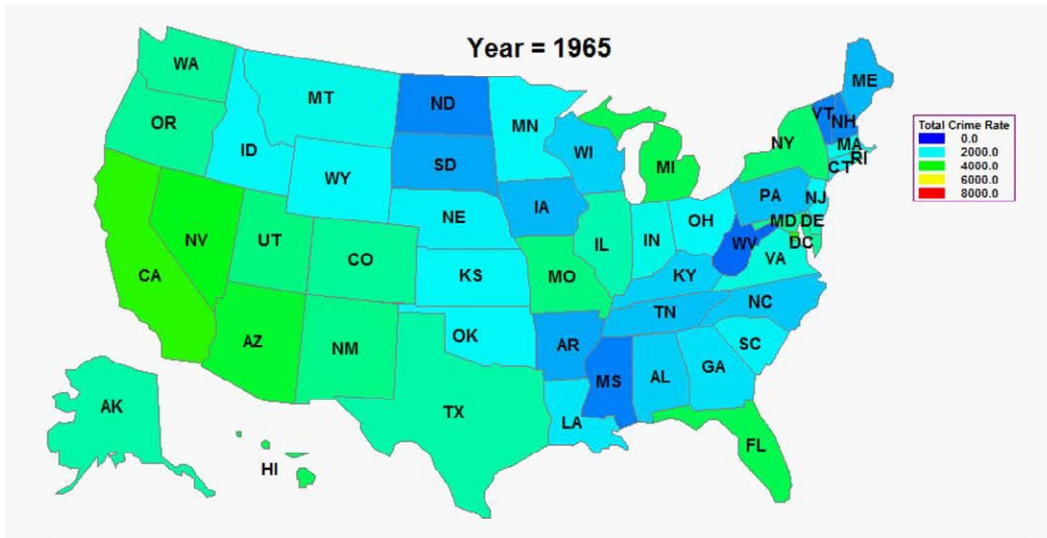


Figure 1: Crime rates by state in 1965.

In the dynamic version of this graph, a slider bar is added to change the year interactively. The increase in crime during the 1980's and 1990's and subsequent decrease after the turn of the century is readily apparent. Figure 2 shows the same data in 1990.

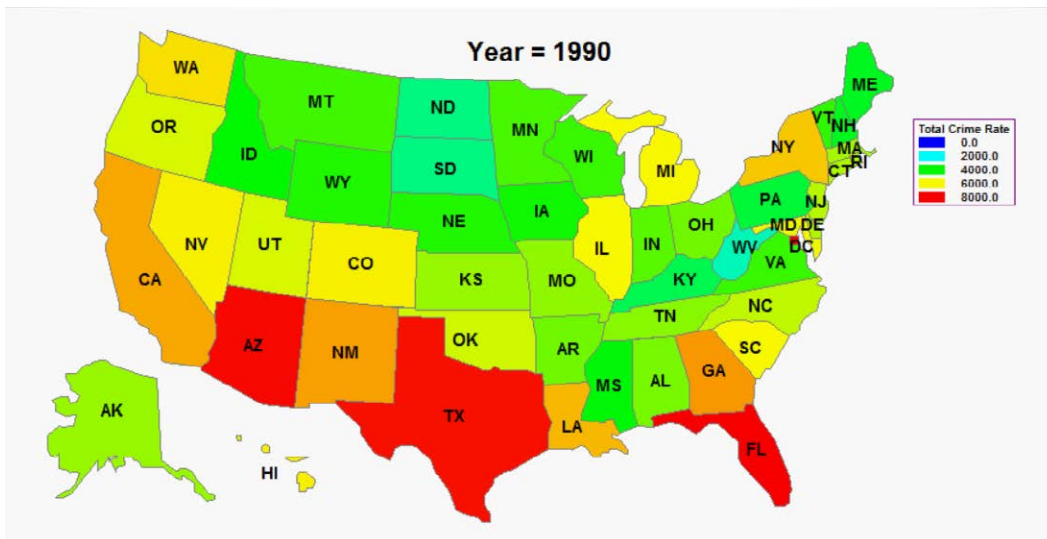


Figure 2: Crime rates by state in 1990.

Maps are also used to display the value of a variable at different locations. For example, Figure 3 uses bubbles to display the size of the largest cities in the United States in 1790.

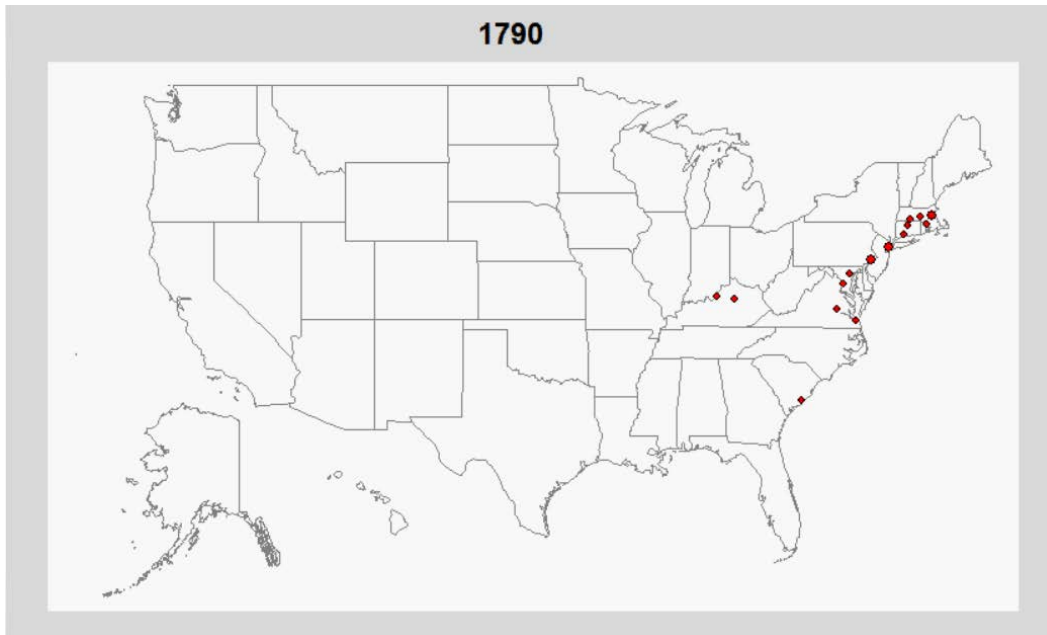


Figure 3: Bubble chart illustrating size of largest U.S. cities in 1790.

In the dynamic version of this graph, changing the year shows the westward expansion of the country and the recent migration from North to South.

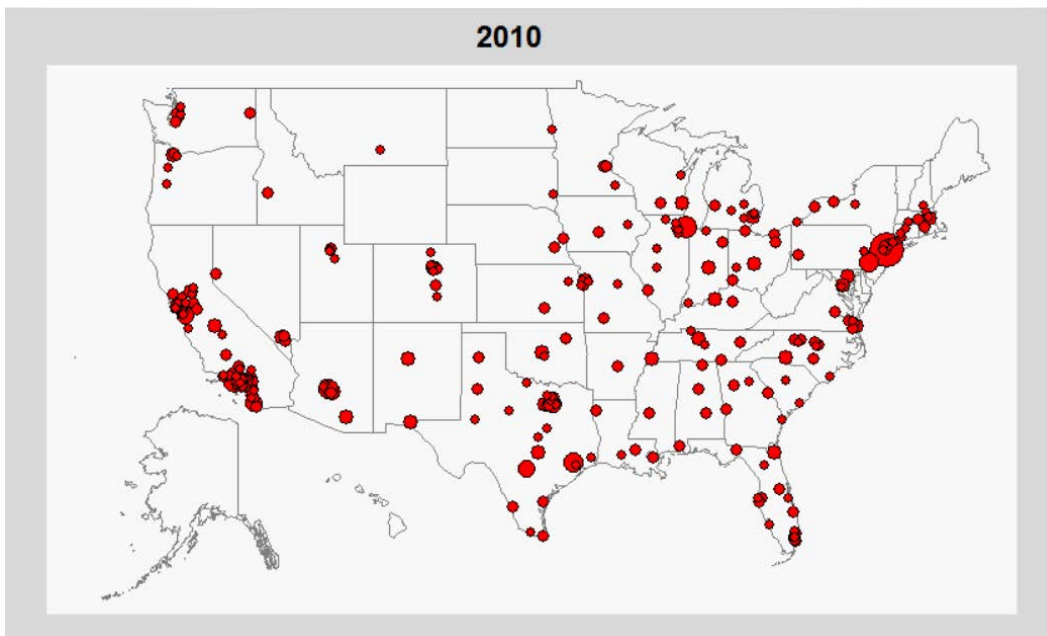


Figure 4: Bubble chart illustrating size of largest U.S. cities in 2010.

1.2 Dynamic Population Pyramids

At times, special dynamic displays are needed to properly demonstrate changes over time. For example, Figure 5 displays a population pyramid in which the length of each bar represents the number of individuals living in the United States in 1950 who fell into various age categories. The post-war baby boom is very evident. Interactively changing the year shows dramatically how the baby boomers have aged, how life expectancy has increased, and how women are living considerably longer than men.

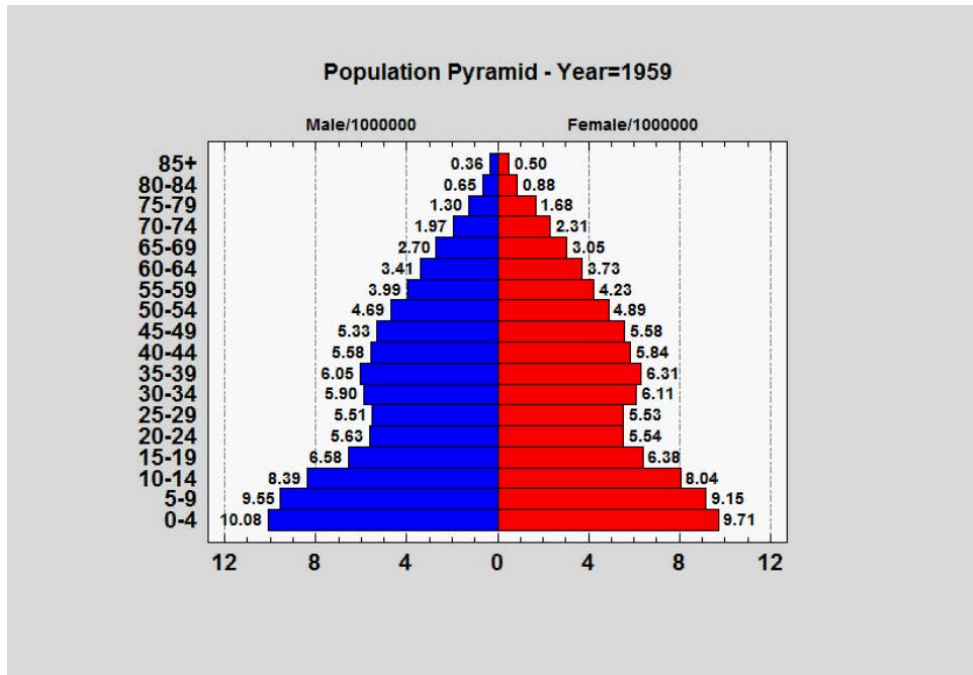


Figure 5: Population pyramid showing age distribution in the U.S. during 1959.

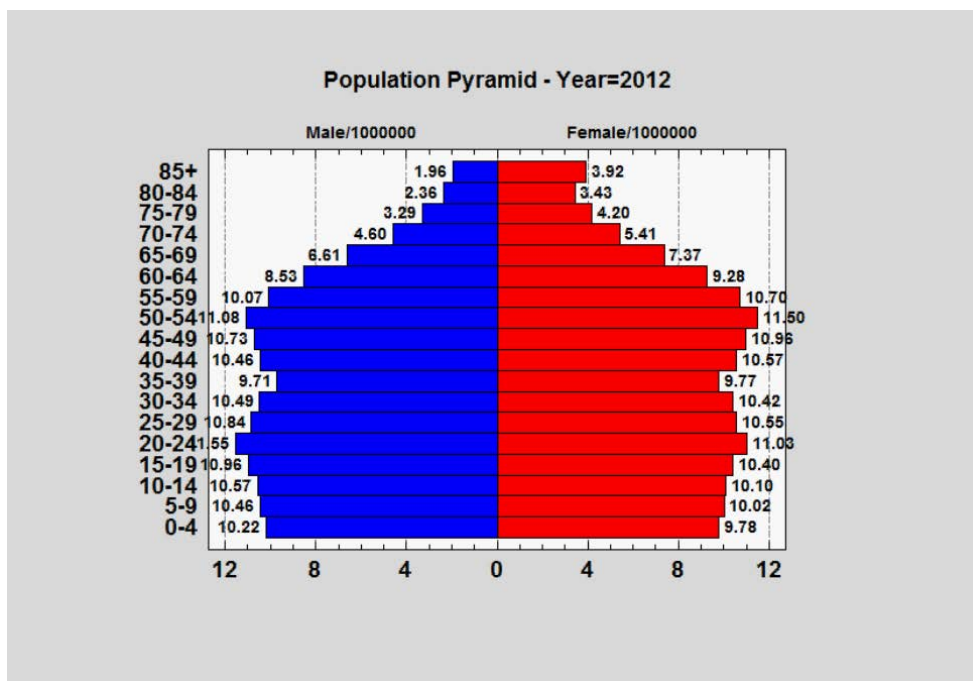


Figure 6: Population pyramid showing age distribution in the U.S. during 2012.

1.3 Dynamic Wind Roses

Another interesting graphic called a wind rose is used to visualize the speed and direction of wind at a specific location. Figure 7 is a wind rose generated at an airport showing the wind conditions at midnight, while Figure 8 shows the distribution at noon. Dynamically changing time shows how the wind speed and direction change throughout the day.

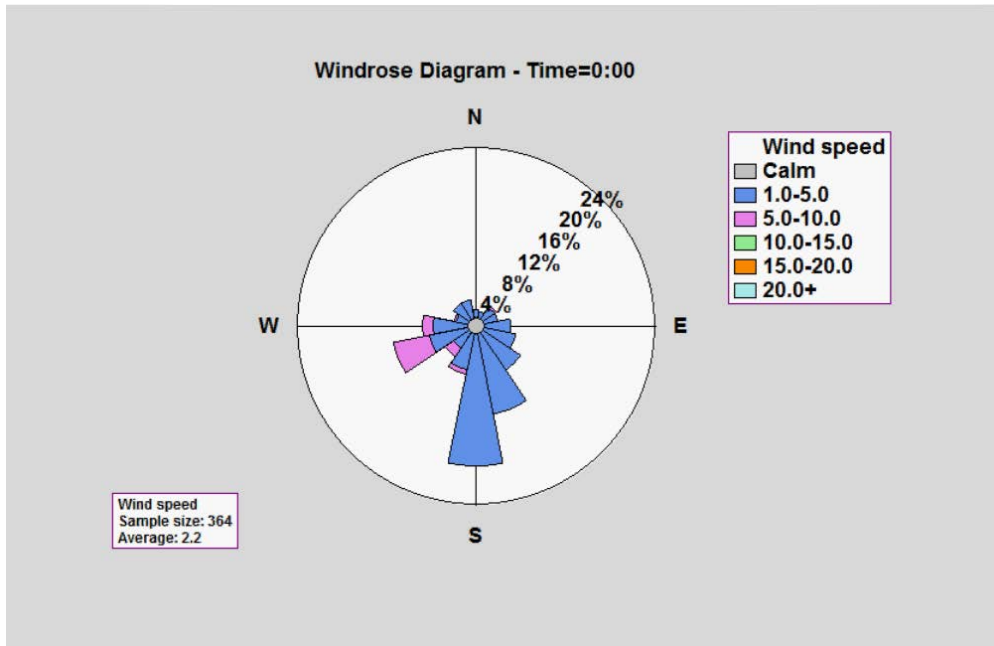


Figure 7: Wind rose showing wind direction and speed distribution at midnight.

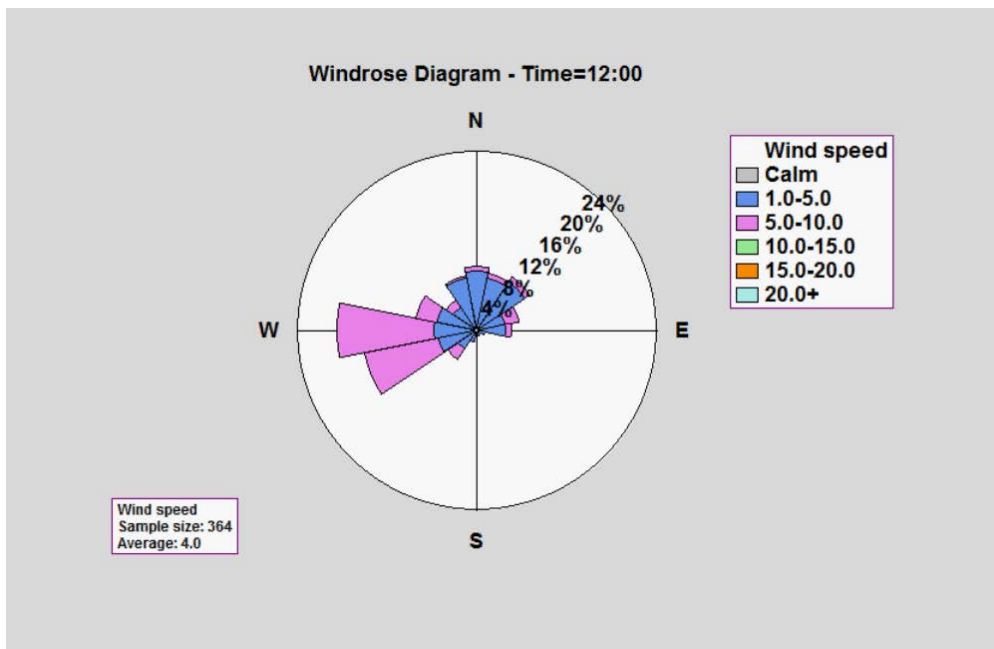


Figure 8: Wind rose showing wind direction and speed distribution at noon.

1.4 Dynamic Bubble Charts

Bubble charts of demographic data that change with time were first discussed in a Ted talk by Hans Rosling (2006). Figures 9 and 10 show data for many countries around the world in 1961 and again in 2009. In the dynamic version, you can see the general movement from lower right to upper left as well as significant reductions in infant mortality. Breadcrumbs may be used to track specific countries which show unusual patterns over time, such as Yemen and Sierra Leone.

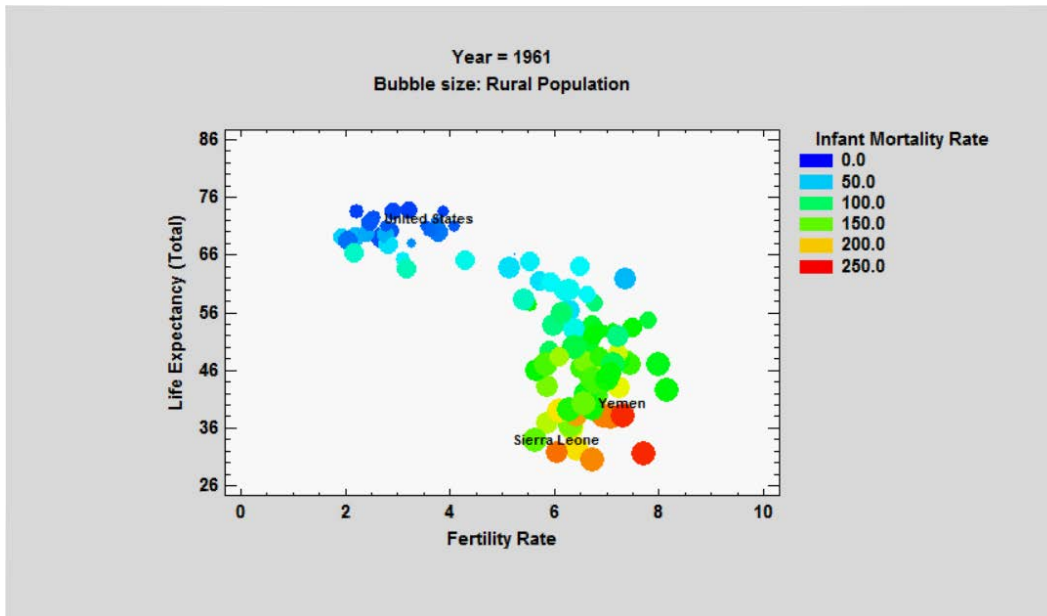


Figure 9: Bubble chart illustrating country statistics in 1961.

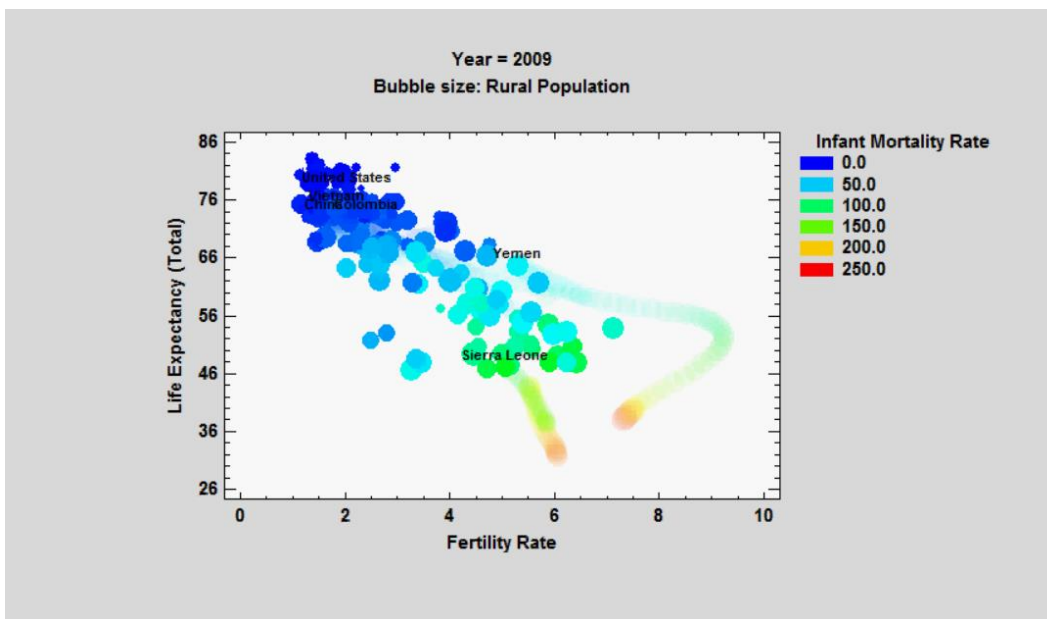


Figure 10: Bubble chart illustrating country statistics in 2009.

1.5 Deviation Dashboard

A useful way of visualizing changes in multivariate data over time is to create a deviation dashboard as shown in Figures 11 and 12. The dashboard shows fish counts taken at various locations in the Gulf of Maine during 1963 and 1979. Each count has been normalized by subtracting the sample mean at that location and dividing by the sample standard deviation.

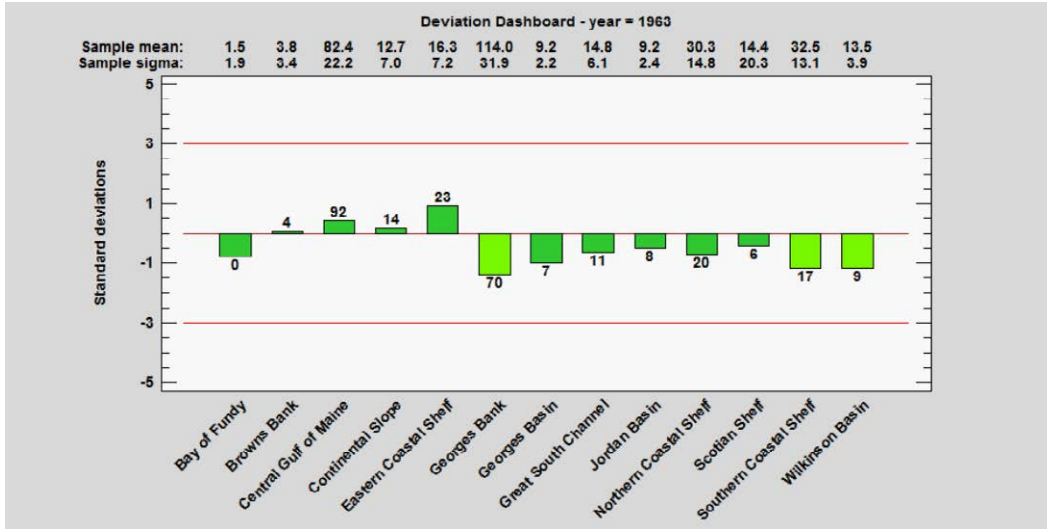


Figure 11: Normalized fish counts in the Gulf of Maine in 1963.

Values beyond ± 2 standard deviations are shown as yellow bars, while those beyond the 3-sigma limits are shown as red bars. In the dynamic version, you can detect correlations between the counts at various locations by noting that the bars change in sync with each other.

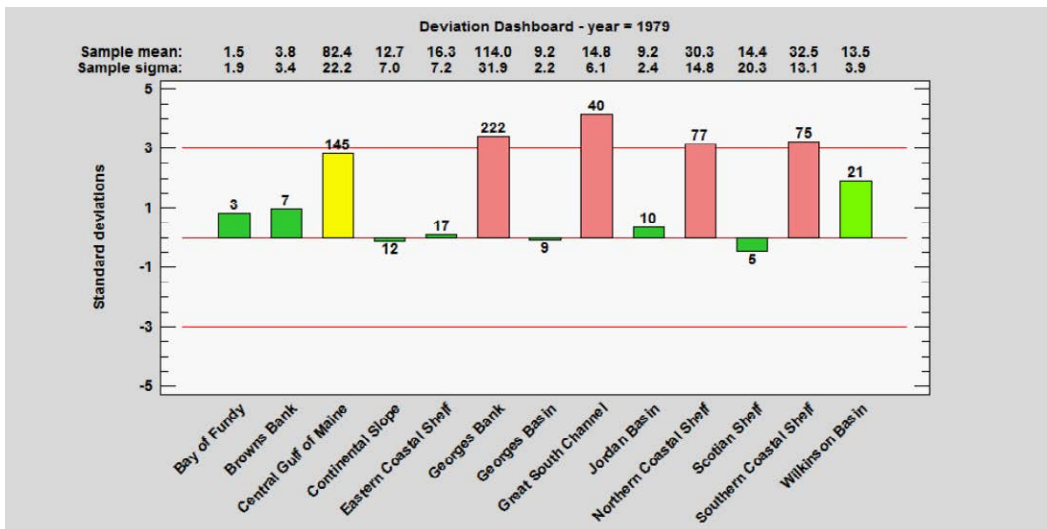


Figure 12: Normalized fish counts in the Gulf of Maine in 1979.

2. Visualizing the Effect of Analysis Parameters

The addition of dynamic features is also useful when illustrating the effect of statistical procedures applied to data. Letting viewers interactively change the value of analysis parameters provides important feedback about how well the analysis performs. It also assists analysts in rapidly finding optimal parameter values.

2.1 Bivariate Density Estimation

One procedure which benefits from interactive control is nonparametric estimation of a bivariate density function. Control over the bandwidth of the weighting function has a significant effect on how much detail is captured by the estimate. Figure 13 shows a typical estimate using a moderate bandwidth. The slider on the toolbar lets the viewer interactively change the bandwidth and immediately see the effect on the estimate.

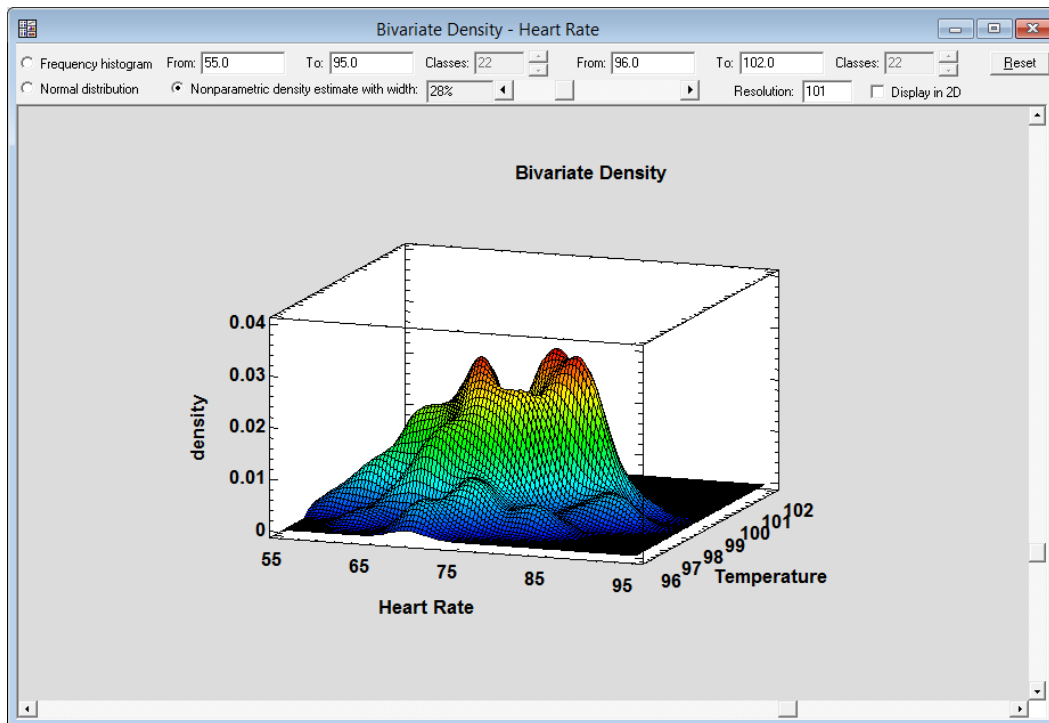


Figure 13: Bivariate density estimation Statlet with interactive bandwidth control.

2.2 Kriging

A second type of graph that benefits from interactive control over analysis parameters is Kriging. During Kriging, the analyst selects several parameters such as the increment at which the variogram is estimated, the maximum lag, the increment at which the response is plotted, the maximum radius for including nearby points in the estimate, the absence or presence of a nugget, etc. Letting the

analyst interactively change those options makes obtaining the desired result much quicker.

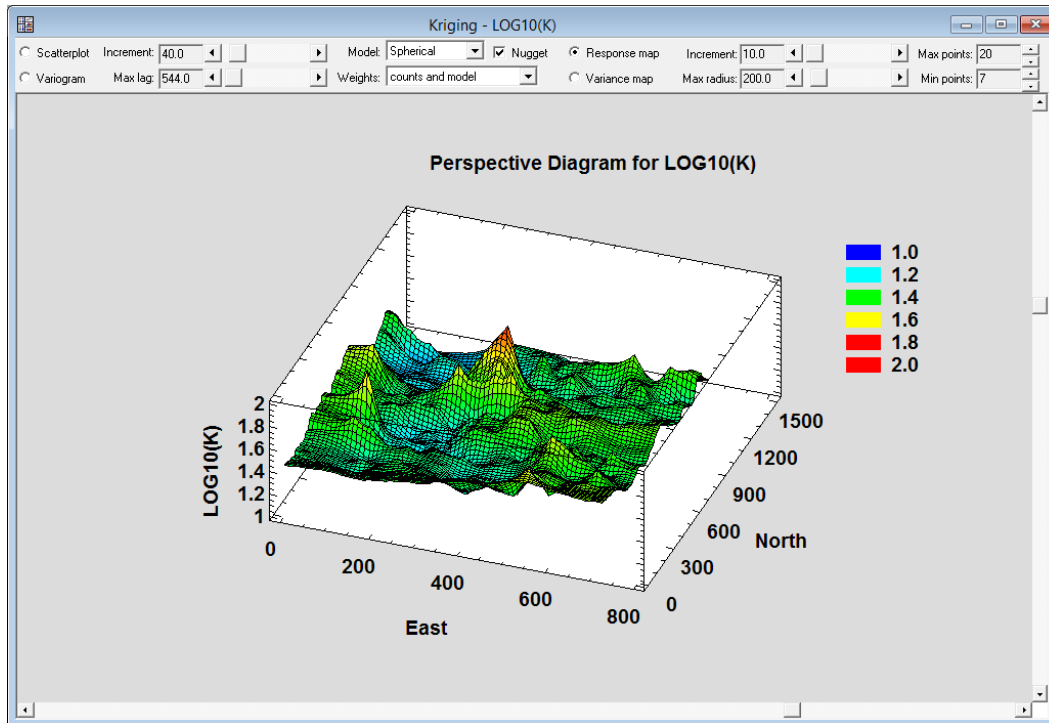


Figure 14: Perspective diagram from Kriging Statlet with interactive control over several analysis parameters.

2.3 Power Transformations

A third statistical procedure that benefits from interactive control over the analysis parameters is the application of a power transformation to normalize data that is skewed in its original metric. The Statlet illustrated in Figure 15 lets analysts interactively change the power to which the data are raised and examine the distribution of the transformed data using a quantile-quantile plot. Confidence limits, the RMSE value of the associated Box-Cox transformation, and the Shapiro-Wilk statistic all help the analyst select the best power for normalizing the data.

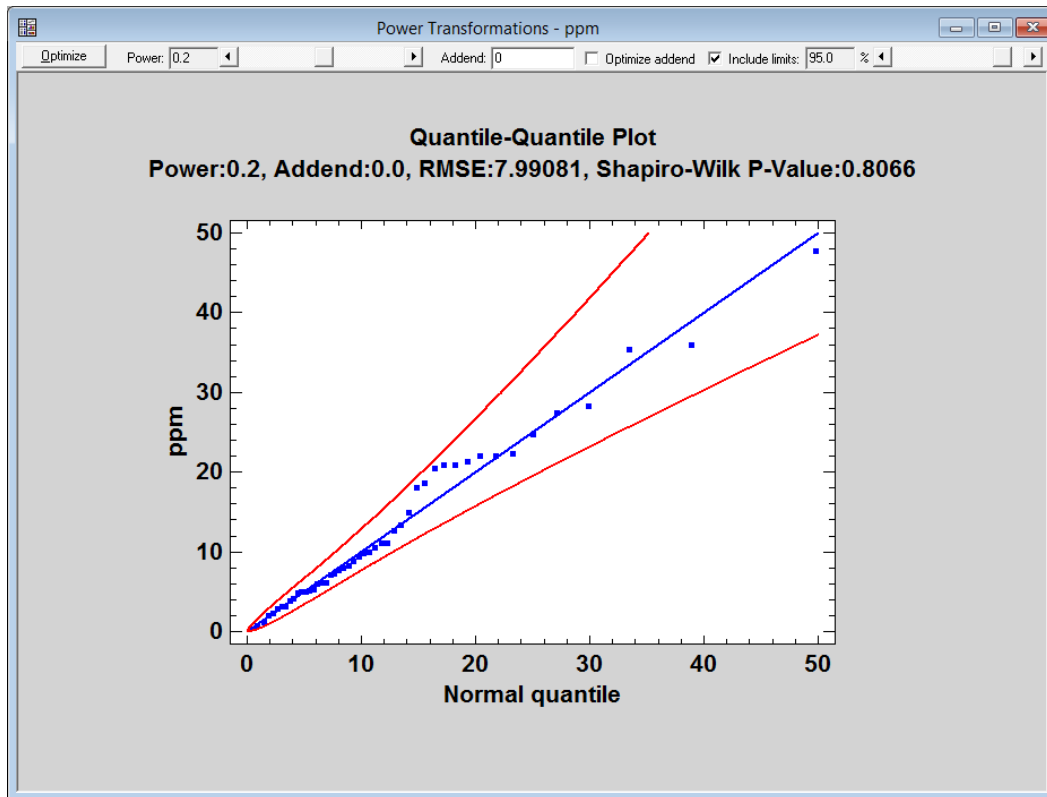


Figure 15: Quantile-quantile plot for examining effect of power transformation on normality of data.

2.4 Response Surfaces

A fourth procedure that benefits from dynamic interaction is the exploration of response surfaces. Having fit a surface to data from a designed experiment, analysts find it helpful to illustrate the fitted model so that optimal locations may be seen. Figure 15 shows a 3-dimensional contour plot illustrating how the estimated strength of a product varies as 3 experimental factors are changed. Contour plots are shown on the top, front and side of the plot. Analysts may interactively change the height at which the top surface is located as well as the values of other factors that are not included in the plot.

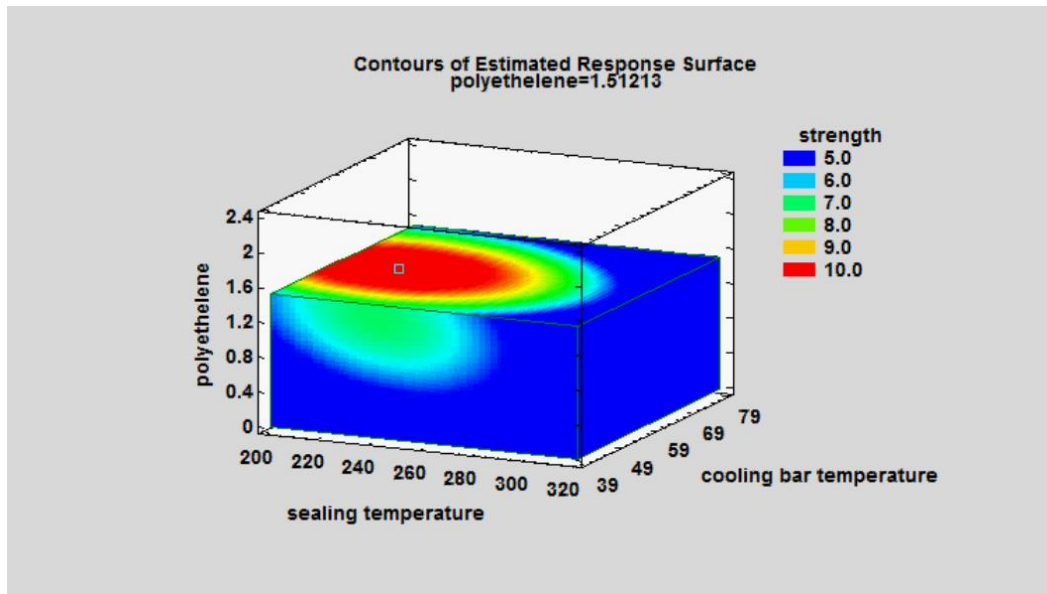


Figure 16: Dynamic response surface plot showing effect of 3 experimental factors on strength.

3. Conclusion

Adding dynamic elements to statistical graphs dramatically increases the amount of information they convey to the viewer. Two important areas which benefit from such additions are visualizing changes over time and examining the effect of analysis parameters on the final results. The speed of today's computers makes it possible to do a significant amount of computation between each redrawing of the graph so that the visual changes occur smoothly.

The original PowerPoint slides of this presentation contain the dynamic versions of each example described in this paper. They may be viewed at www.statgraphics.com/statistics-library.

References

- Rosling, Hans (2006). "The best stats you've ever seen."
www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen