



STATISTICA Icon Plots

Last Update: 2009

U.S. Headquarters: StatSoft, Inc. • 2300 E. 14th St. • Tulsa, OK 74104 • USA • (918) 749-1119 • Fax: (918) 749-2217 • info@statsoft.com • www.statsoft.com

Australia: StatSoft Pacific Pty Ltd.
Brazil: StatSoft South America
Bulgaria: StatSoft Bulgaria Ltd.
Czech Rep.: StatSoft Czech Rep. s.r.o
China: StatSoft China

France: StatSoft France
Germany: StatSoft GmbH
Hungary: StatSoft Hungary Ltd.
India: StatSoft India Pvt. Ltd.
Israel: StatSoft Israel Ltd.

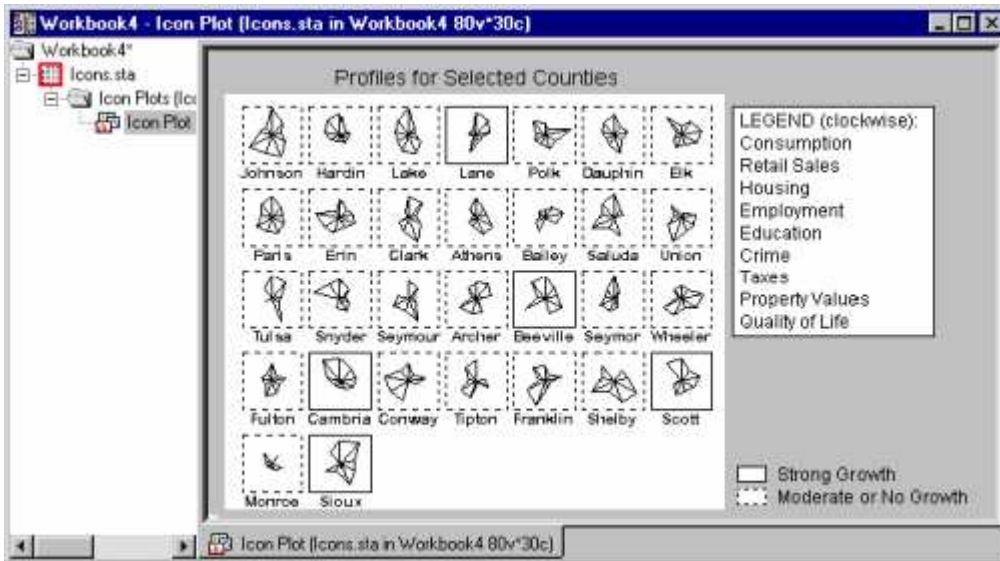
Italy: StatSoft Italia srl
Japan: StatSoft Japan Inc.
Korea: StatSoft Korea
Netherlands: StatSoft Benelux BV
Norway: StatSoft Norway AS

Poland: StatSoft Polska Sp. z.o.o.
Portugal: StatSoft Ibérica Lda
Russia: StatSoft Russia
Spain: StatSoft Ibérica Lda

S. Africa: StatSoft S. Africa (Pty) Ltd.
Sweden: StatSoft Scandinavia AB
Taiwan: StatSoft Taiwan
UK: StatSoft Ltd.

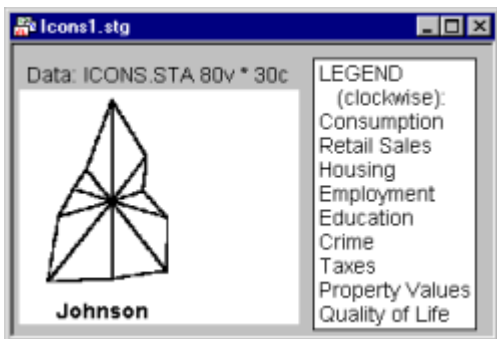
STATISTICA Icon Plots

Icon plots provide a selection of plots that represent cases or units of observation as multidimensional symbols.



The general idea behind this method capitalizes on the human ability to "automatically" spot complex (sometimes interactive) relations between multiple variables if those relations are consistent across a set of instances (in this case "icons"). Sometimes the observation (or a "feeling") that certain instances are "somehow similar" to each other comes before the observer (in this case an analyst) can articulate which specific variables are responsible for the observed consistency (Lewicki, Hill, & Czyzewska, 1992). However, further analysis that focuses on such intuitively spotted consistencies can reveal the specific nature of the relevant relations between variables.

The basic idea of icon plots is to represent individual units of observation as particular graphical objects where values of variables are assigned to specific features or dimensions of the objects (usually one case = one object). The assignment is such that the overall appearance of the object changes as a function of the configuration of values.



Thus, the objects are given visual "identities" that are unique for configurations of values and that can be identified by the observer. Examining such icons may help to discover specific clusters of both simple relations and interactions between variables.

This white paper addresses the following topics:

- Analyzing Icon Plots
- Taxonomy of Icon Plots
- Standardization of Values
- Applications

Analyzing Icon Plots

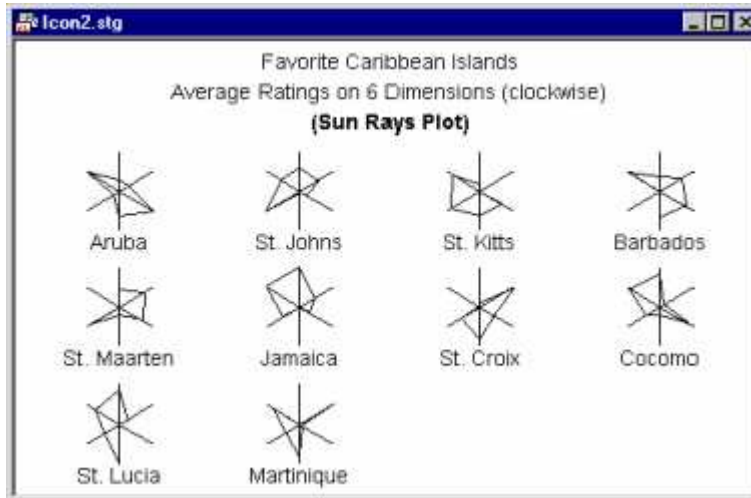
The "ideal" design of the analysis of icon plots consists of five phases:

1. Select the order of variables to be analyzed. In many cases, a random starting sequence is the best solution. You can also try to enter variables based on the order in a multiple regression equation, factor loadings on an interpretable factor, or a similar multivariate technique. That method can simplify and "homogenize" the general appearance of the icons, which would facilitate the identification of non-salient patterns. It can also, however, make some interactive patterns more difficult to find. No universal recommendations can be given at this point, other than to try the quicker (random order) method before getting involved in the more time-consuming method.
2. Look for any potential regularities, such as similarities between groups of icons, outliers, or specific relations between aspects of icons (e.g., "if the first two rays of the star icon are long, then one or two rays on the other side of the icon are usually short"). The circular type of icon plots (see *Taxonomy of Icon Plots*, below) is recommended for this phase.
3. If any regularities are found, try to identify them in terms of the specific variables involved.
4. Reassign variables to features of icons (or switch to one of the sequential icon plots, see *Taxonomy of Icon Plots*, below) to verify the identified structure of relations (e.g., try to move the related aspects of the icon closer together to facilitate further comparisons). In some cases, at the end of this phase it is recommended to drop the variables that appear not to contribute to the identified pattern.
5. Finally, use a quantitative method (such as a regression method, nonlinear estimation, discriminant function analysis, or cluster analysis) to test and quantify the identified pattern or at least some aspects of the pattern.

Taxonomy of Icon Plots

Most icon plots can be assigned to one of two categories: circular and sequential. *STATISTICA* icon plots include circular icons, sequential icons, pie icons, and Chernoff faces.

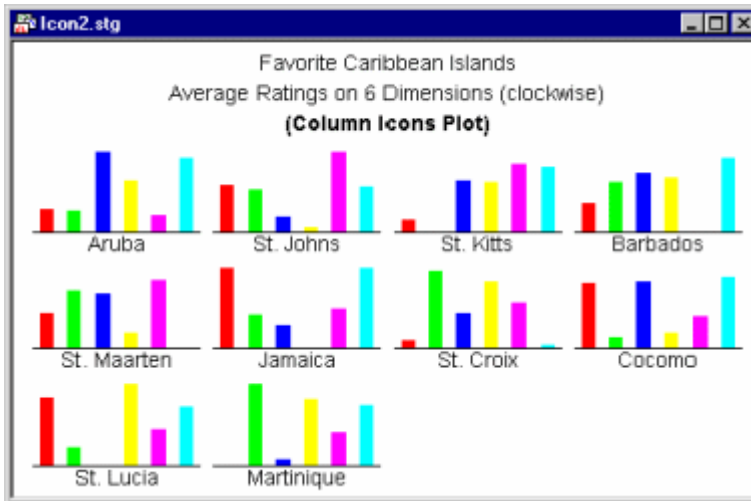
Circular icons. Circular icon plots (Star plots, Sun Ray plots, Polygon icons) follow a "spoked wheel" format where values of variables are represented by distances between the center ("hub") of the icon and its edges.



Those icons may help to identify interactive relations between variables because the overall shape of the icon may assume distinctive and identifiable overall patterns depending on multivariate configurations of values of input variables.

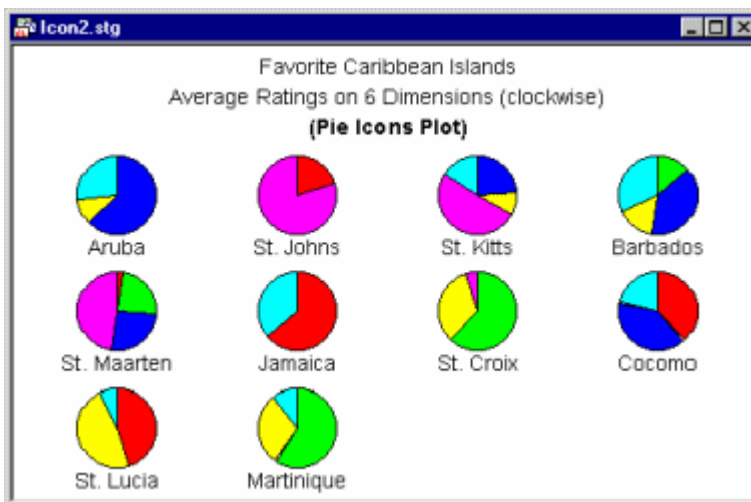
In order to translate such "overall patterns" into specific models (in terms of relations between variables) or verify specific observations about the pattern, it is helpful to switch to one of the sequential icon plots (see the next paragraph), which may prove more efficient when you already know what to look for.

Sequential icons. Sequential icon plots (Column icons, Profile icons, Line icons) follow a simpler format where individual symbols are represented by small sequence plots (of different types).



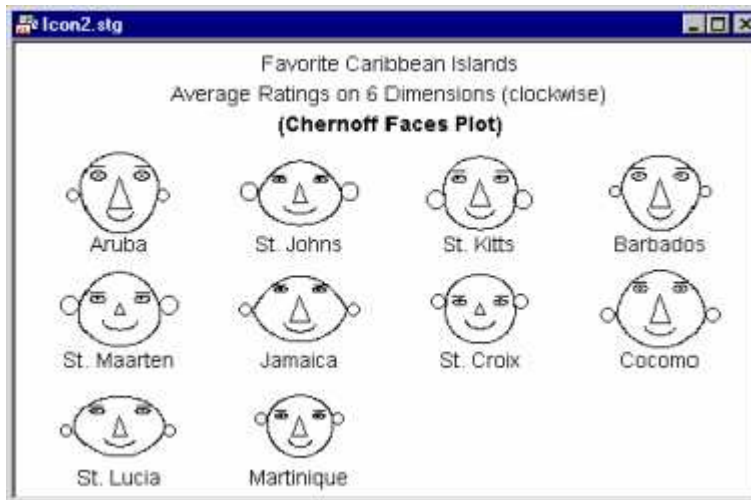
The values of consecutive variables are represented in those plots by distances between the base of the icon and the consecutive break points of the sequence (e.g., the height of the columns shown above). Those plots may be less efficient as a tool for the initial exploratory phase of icon analysis because the icons may look alike. However, as mentioned before, they may be helpful in the phase when some hypothetical pattern has already been revealed and you need to verify it or articulate it in terms of relations between individual variables.

Pie icons. Pie icon plots fall somewhere in between the previous two categories; all icons have the same shape (pie) but are sequentially divided in a different way according to the values of consecutive variables.



From a functional point of view, they belong rather to the sequential than circular category, although they can be used for both types of applications.

Chernoff faces. This type of icon is a category by itself. Cases are visualized by schematic faces such that relative values of variables selected for the graph are represented by variations of specific facial features.¹



Due to its unique features, it is considered by some researchers as an ultimate exploratory multivariate technique that is capable of revealing hidden patterns of interrelations between variables that cannot be uncovered by any other technique. This statement may be an exaggeration, however. Also, it must be admitted that *Chernoff Faces* is a method that is difficult to use, and it requires a great deal of experimentation with the assignment of variables to facial features.

Standardization of Values

Except for unusual cases when you intend for the icons to reflect the global differences in ranges of values between the selected variables, the values of the variables should be standardized once to assure within-icon compatibility of value ranges. For example, because the largest value sets the global scaling reference point for the icons, then if there are variables that are in a range of much smaller order, they may not appear in the icon at all (e.g., in a star plot, the rays that represent them will be too short to be visible).

¹ Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of American Statistical Association*, 68, 361-368.

Applications

Icon plots are generally applicable (1) to situations where you want to find systematic patterns or clusters of observations, and (2) when you want to explore possible complex relationships between several variables.

The first type of application is similar to cluster analysis; that is, it can be used to classify observations.

For example, suppose you studied the personalities of artists, and you recorded the scores for several artists on a number of personality questionnaires. The icon plot may help you determine whether there are natural clusters of artists distinguished by particular patterns of scores on different questionnaires (e.g., you may find that some artists are very creative, undisciplined, and independent, while a second group is particularly intelligent, disciplined, and concerned with publicly acknowledged success).

The second type of application (the exploration of relationships between several variables) is more similar to factor analysis; that is, it can be used to detect which variables tend to "go together." For example, suppose you were studying the structure of people's perception of cars. Several subjects completed detailed questionnaires rating different cars on numerous dimensions. In the data file, the average ratings on each dimension (entered as the variables) for each car (entered as cases or observations) are recorded.

When you now study the Chernoff faces (each face representing the perceptions for one car), it may occur to you that smiling faces tend to have big ears; if price was assigned to the amount of smile and acceleration to the size of ears, then this "discovery" means that fast cars are more expensive. This, of course, is only a simple example; in real-life exploratory data analyses, non-obvious complex relationships between variables may become apparent.