

# Many Eyes

Social Data Visualization Software

rcfpr1\_1\_1t\_5

rcfpr1\_1\_1t\_22 tymsgfpr1\_1\_1t\_

tymsgfpr1\_1\_3t\_3 tymsgfpr1\_tlfp

\_\_\_\_tymsgfpr1\_1\_1t\_

\_rcfpr1\_1\_1t\_100

#### Many Eyes is a public web site where

**About Many Eyes** 

users may upload data, create interactive visualizations, and carry on discussions. The goal of the site is to support collaboration around visualizations at a large scale by fostering a social style of data analysis in which visualizations not only serve as a discovery tool for individuals but also as a medium to spur discussion among users. To support this goal, the site includes novel mechanisms for end-user creation of visualizations and asynchronous collaboration around those visualizations. In addition to describing these technologies, this site provieds a preliminary report on the activity of the

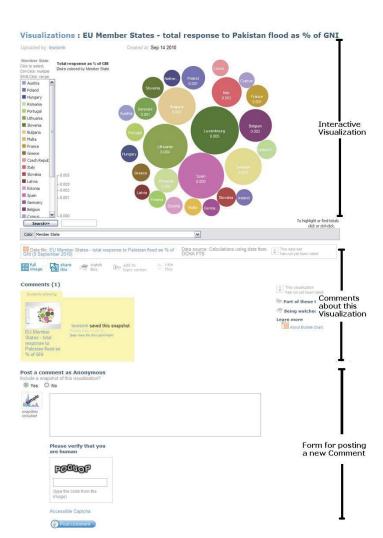
Many Eyes addresses the challenges of end user construction and the unique environment of open web-based collaboration. The site launched on January 23, 2007 and allows users to upload data, create visualizations of that data and leave comments on both visualizations and data sets. In addition to constituting a research platform, the site is an effort to "democratize" visualization technology; both providing the technology to the broadest possible audience and fostering a democratic deliberative style of data analysis.

#### On Many Eyes users can:

1. View and discuss visualizations 2. View and discuss data sets 3. Create visualizations from existing data

#### Registered users can also:

- 4. Rate data sets and visualizations 5. Upload data
- 6. Create and participate in topic centers 7. Select items to watch 8. Track contributions, watchlist, and
- topic centers 9. See comments that others have written



#### **Data Format and** Style

All the visualizations in Many Eyes accept the same data table format except <Word Tree and Phrase Net, where as, all the charts in the category (Analyze a text) understand free text also. A data table consists of rows where the values in each row are separated by tabs. The first row of the table should be «headers» that describe the columns. Here is an example table in the Many Eyes format:

Calories 100 Apple Yellow Banana 120

The table has a header row (Fruit, Color, Calories) and two data rows. The first two columns are freeform text and the third column will be automatically interpreted as numbers. In each row the values are separated by tab characters, not spaces. This format can be easily exported from databases or spreadsheets. The size limit is 5 MB. per file.

That is all there is to the basic format. Some visualizations add their own twist For example, in our map of the United States, text values such as «AZ» or «MA» will be interpreted as state abbreviations. If you are interested in a particular kind of visualization, read its explanation page before you upload your data.

**New Jersey &** 

**DAI 523:** 

**Instructor:** 

Pino Trogu

**Department** 

University

California, USA

Designed by:

BACK FRONT COVER

grid.

8.5x11 as shown on the **word** 

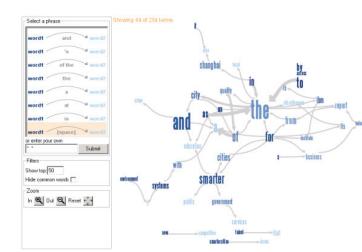
Elnaz Davoudi

December 2010

Information Design I

#### **Network Diagram**

Real-world information often comes in the form of relationships between entities or items, such as people who know each other (social networks), or Web pages that are connected to each other. In a network diagram, entities are connected to each other in the form of a node and link diagram.



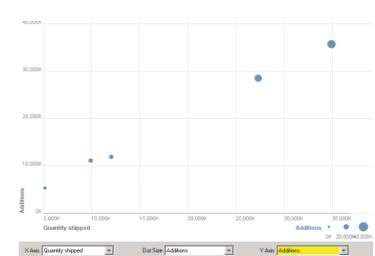
A network consists of a set of objects called vertices connected by edges. The visualization of the network is optimized to keep strongly related items in close proximity to each other. In this way, the overall arrangement of vertices in the network is very descriptive of the structure of the connections between vertices (vertices that are far away are weakly related to each other). In this visualization, the size of a vertex is proportional to the number of edges emanating from it. The network display in Many Eyes is fully interactive and can be zoomed and panned to obtain a detailed view of different sections of the graph. A network diagram takes a table consisting of at least two text columns of data, with each row representing a single relationship between two items. The column contents are interpreted as the display label for each item. Please keep in mind that in general, the layout optimization process can be very time consuming for larger networks (more than a thousand vertices) especially when running in a browser. In most cases the applet will eventually compute a correct layout in a few minutes, however. Also, this technique is not very

#### Scatterplot

A scatter plot is a classic statistical diagram that lets you visualize the relationship between numeric variables.

well suited for networks in which many

nodes have a large number of neighbors.



The Many Eyes scatter plot gives the conventional diagram an interactive twist. Each row in the data table is represented by a dot. The x-coordinate of a dot corresponds to one of the columns in the table. To select which columns are used for the x- and y-axes, and which for the dot size, use the three drop-down menus below the scatter plot. The y-coordinate corresponds to a different column. The size of a dot can also reflect another column. A scatter plot is one of the most flexible methods for visualizing numeric data. It works with any data table that has at least two numeric columns (for the x- and y-axes) and a text column (to label points as the mouse moves over them).

#### **Matrix Chart**

multidimensional data set in a grid. The matrix chart divides the screen into a grid. Rows represent the values in one text column (e.g., product line) and columns represent another text column (e.g., country). Each cell then shows a circle or bar that represents the value for its row/ column combination (e.g., returns for golfing equipment for Belgium).

#### **Bars and Bubbles**

The two modes of the matrix, circles and bars, are useful in different situations. Bars, whose height represents numeric values, are better for exact comparisons and allow space for more columns. Circles, which show values via area, are good for showing non-negative values that vary greatly, and allow space for more rows.

#### Colors

You can display a third dimension of data using color. When you select a column to be represented by color, circles will turn into miniature pie charts and bars will be broken into differently-colored pieces. This option should be used with caution, since it's possible to create a cluttered display, but it can often be effective.



#### Average, Total, and Count The matrix chart automatically performs

simple aggregation of the data set. For instance, if your data set has two rows that show returns for golfing equipment for Belgium, say \$1,000 and \$2,000 the grid cell for golfing egipment/Belgium will aggregate these values. If you choose «Total» in the options pane, the value will be \$3,000. Average would yield \$1,500, and Count would show 2, the number of rows that correspond to that cell. Note that if your data set has no numeric columns, the only option is to aggregate by count.

Size By Sometimes you may wish to compare data within columns or rows. Use the «Size By» buttons to compare show proportions in individual columns, rows, or (if you use color) even within cells. A matrix chart takes a table with at least two text columns, for the x- and

y-categories. If there is a third text column it can be used to determine colors. If there are no numeric columns, then the bubbles or bars simply show the count of each category combination. A matrix chart can be seen as the visual equivalent of a «cross-tab» or pivot table.

## Map

The maps on Many Eyes can overlay data values on geographic regions. Use a map if you want to see spatial patterns in the The map displays data associated with

geographic regions using either colors or **Colors versus bubbles** 

Colored maps are not always the best way to display values, since large regions are emphasized much more than small ones. To address this issue, the map provides an

#### alternate «bubble» mode in which values are represented by circles. The area of each circle corresponds to the absolute value of the data. Blue circles indicate

negative numbers. **Comparing maps** You can compare two data columns side-by-side, or show one small map for each column. This «small multiples» mode is often an effective way to show regional changes over time or to correlate

geographic trends. Zoom Align Map Scales | Colors or Bubbles?



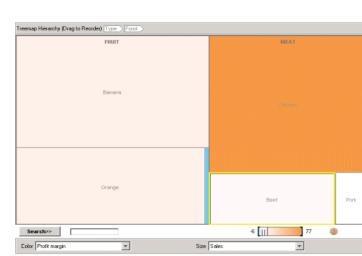
#### Treemap

A tree map is a visualization of hierarchical structures. It is very effective in showing attributes of leaf nodes using size and color coding. Tree maps enable users to compare nodes and sub-trees even at varying depth in the tree, and help them spot patterns and exceptions. Many data sets have a hierarchical structure: items are divided into categories, subcategories, and so on. Each data item, or row in the data set, is represented by a rectangle. The rectangles have different sizes, with the area being proportional to a user-defined attribute. The rectangles are colored to reflect other attributes of the

Many Eyes provides two types of treemap: a «plain» Treemap, and a special Change tree map. In the plain version, you can define a particular column of a data set to correspond to color. If your data set includes numbers such as yearly sales, this is a convenient way to visualize change

over time. A tree map requires a somewhat complex data table. To create the categories and subcategories, Many Eyes will look at all the text columns, in order from left to right. The leftmost text column will be used for the highest-level category, the next for the level below that, and so on.

Treemaps are a complex but powerful information visualization technique. They were introduced in Shneiderman, 1992. Our layout algorithm is based on Bruls, Huizing, and & van Wijk, 2000.



#### Phrase Net

A phrase net diagrams the relationships between different words used in a text. It uses a simple form of pattern matching to provide multiple views of the concepts contained in a book, speech, or poem. The result of this simple pattern matching scheme is a surprisingly coherent view of some of the concepts in the article. In a book, for example, you might find a large cluster of the main characters and their relationships is on one side and separate clusters related to emotion and attitude.

#### Phrase net analyzes a text by looking for pairs of words that fit particular patterns.

proportions. For example, the pie chart below shows the proportions of a budget for advertising: 50% television, 40% newspaper, and 10% yelling on a street corner. The percentages are encoded are «slices» of a pie, with the area corresponding to the percentage. The Many Eyes pie chart shows each slice in a different color. You don't need to put your numbers in percentage form: the visualization does this for you. (In the process it ignores any numbers that are negative or missing). You can move the mouse over slices for more information on

The pie chart needs one column of text, for labeling slices, and at least one configure the chart, you can specify more than one numeric column, in which case are looking at poll results from multiple

same phrase. The color of a word indicates whether it is more likely to be found in the first or second slot of a pattern. The darker the word, the more often it appears in the first position.

You can specify this pattern by using

asterisks as wildcard characters. For

instance, the pattern «\* and \*» will match

phrases like «play and sing» or «vexation

and regret.» Punctuation matters, so it will

not match «left, and then». You can choose

from some useful defaults or you can type

After you specify a pattern, the program

creates a network diagram of the words it

if they occur in the same phrase. The size

of times it occurs in a match; the thickness

of an arrow between words tells you how

many times those two words occur in the

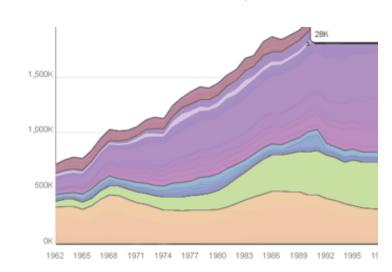
of a word is proportional to the number

finds as matches. Two words are connected

your own patterns in the field below the list.

#### Stack Graphs

A stack graph is a classic method for visualizing change in a set of items, where the sum of the values is as important as the individual items. For example, a stack graph is excellent for looking at revenue over time across several products. Because stack graphs use areas to convey numbers, they do not work for negative values. And in some situations it might not make sense to add up different data series (say, prices of different stocks over time). For these



cases you may want to use a line graph. Many Eyes includes two different stack graphs. The plain stack graph is appropriate for up to a few dozen items. If you have a larger number of items that are organized into categories and subcategories, you should consider the Stack Graph for Categories. A stack graph takes a table where each column corresponds one data series. One special column corresponds to the x-axis labels. For example, in this table the «Year» column would be used for the x-axis labels, and the two other columns would define two data series to be graphed. Note that this example is natural for a stack graph since the values are positive, and

Stack graphs are a standard, useful chart. They can convey multiple levels of meaning at once, but have some drawbacks. It can be hard to judge the exact widths of a stripe in a stack graph, or compare the widths of two stripes. If exactitude or comparisons are of primary importance, consider a line graph instead. Another consideration is that some time series do not make sense to add. For example, adding the stock price of one

when you add them up you get something

meaningful: the price of a (not so healthful)

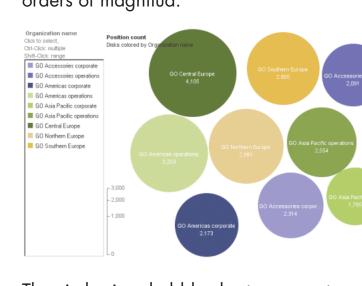
breakfast.

company to the stock price of another is pointless. In many data sets, items are arranged into categories and subcategories. The

tree outline provides a color key to these categories. It allows lets the user drill down into the data by expanding and collapsing categories. As the user drills down, the color scheme adjusts so that there is always a wide range of colors in the graph. The tree control itself contains tiny «sparkline» graphs that show the trend for each subcategory and category.

#### **Bubble Chart** A bubble chart displays a set of numeric

values as circles. It is especially useful for data sets with dozens to hundreds of values, or with values that differ by several orders of magnitud.

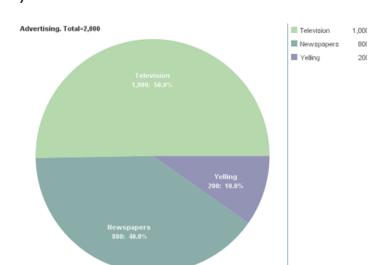


The circles in a bubble chart represent different data values, with the area of a circle corresponding to the value. Because a bubble chart uses area to represent numbers, it is best for positive values. If your data set includes negative values, they will be shown in a different color. If your data set has many negative numbers, consider using a block histogram instead.

#### Pie Chart

A pie chart is a familiar way of showing each one.

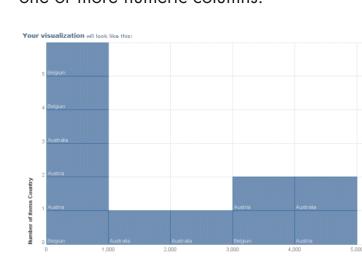
numeric column for slice values. When you viewers can use a menu to switch between different numeric values. For example, you might want multiple numeric columns if you



Pie charts have a mixed reputation. They are popular in business and the media but many information designers have criticized the technique. Some claim that the pie slice shape communicates numbers less exactly than other possibilities such as line length. But this remains unclear in the context of proportions: for example, we have seen no studies that looked at the task of judging whether an item is more or less than 50%. It is also unclear whether exact communication of numeric values is the only evaluation criterion; at least one study indicates that use of a pie chart for analyzing a problem as opposed to a bar chart changes the way people think about

#### **Block Histogram**

A block histogram lets you see the distribution of numeric values in a data set. The x-axis is divided into «bins» that correspond to value ranges. Each item in the data set is drawn as a rectangular block, and the blocks are piled into the bins to show how many values in each range. A block histogram is a bit like an interactive version of a stem-and-leaf diagram: it gives you an overall sense of the data, but you can get details if you want them. A block histogram takes a data table with one text column, for labels, and one or more numeric columns.



#### **Word Cloud** Generator

A «Word Cloud Generator» enables you to see how frequently words appear

in a given text, or see the relationship between a column of words and a column of numbers. You can tweak your word «clouds» with different fonts, layouts, and color schemes.

The size of a word is proportional to the quantity associated with that word, which, in the case of free text, is the word count. Word Cloud Generator draws each word at a size proportional to its frequency. By default, the visualization treats different letter cases («Hello» vs. «hello») as different words and filters out common words (such as «the»). The «Language» menu lets you adjust these settings.



#### **Word Tree**

A word tree is a visual search tool for unstructured text, such as a book, article, speech or poem. It lets you pick a word or phrase and shows you all the different contexts in which the word or phrase appears. The contexts are arranged in a tree-like branching structure to reveal recurrent themes and phrases. For instance, the word «word» occurs a number of times in the previous paragraphs. You will notice that in the words following

«word» there are many repeated phrases

times, and «or phrase» follows three times

Unlike most Many Eyes visualizations, the

word tree starts with a blank slate instead

of a full visualization of the data. You must

choose a search term to display a word

tree. After a word or phrase is typed, the

computer finds all the occurrences of that

term, along with the phrases that appear

\" word\" occurs a number of times in the previous paragraphs

You can manipulate the tree in several

clicking on a word in the tree. If you

ways. To zoom into a particular branch,

control-click on a word, the diagram will

use that new word as the main search

A line graph is a classic method for

numeric variables.

visualizing continuous change. The Many

Eyes line graph can chart one or more

If you are graphing more than one item,

values to make comparisons easier. Use the

«Relative» checkbox at lower right to index

you may want to standardize the data

A line graph takes a table where each

Line graphs are classic diagrams that

usually give a good picture of the data.

There are just a few pitfalls. The most

important thing is that line graphs should

only be used when the positions on the

x-axis have a natural ordering. If your

fine. If your labels are «US, England,

Germany» you should consider a bar

A bar chart is a classic method for

numerical comparisons. The Many Eyes

bar chart can show one or more sets of

A bar chart takes a table where each

column corresponds to one data series.

One special column corresponds to the

x-axis labels. For example, in the many

tables the «Year» column will be used for

define two data series to be graphed. In

certain cases you may want the x-axis

(many data spreadsheets have one

to flip rows and columns.

the x-axis labels, and the two other columns

labels to be taken from the column headers

column per year, for example). When you

configure the visualization, you can choose

Bar charts are classic diagrams that usually

give a good picture of the data. Their main

problem is that when there are many bars,

labeling becomes problematic. They also

is something that is plausibly continuously

imply that the data is discrete; if your data

changing over time, for instance, you might

graph instead.

variables.

**Bar Chart** 

labels are «2000, 2001, 2002» that>s

column represents one data series. One

special column corresponds to the x-axis

all the series to start at 100.

term. And if you wish to see the context

occurring before rather than after a phrase,

or phrase 

is a visual search tool for unstructured text, such as a book, art

made from an chapter of the installation guide, using the searc

works unlike most many eyes visualizations , the word tree starts

starts with a blank slate instead of a full visualization of the data

and shows you all the different contexts in which the wor

To create a word tree, the computer merges

For instance, «tree» follows «word» five

all the matching phrases.

after it.

select End.

**Line Graph** 

fostering a social style of data analysis in which visualizations not only serve as a discovery tool for individuals but also as a medium to spur discussion among users. However, upto this day, November 14th, 2010 the general rate of comments in this site is less than 1% of the whole activities that take place in this site. At the moment Many Eyes web site is providing a set of visualization creation and publishing tools to a large potential audience. The architecture of a site that aims to be useable by millions is nontrivial, and many choices and tradeoffs have been discussed by the IBM developers in the current design. In some cases these design choices involve simplifications to or reordering of the standard visualization pipeline—allowing data transformation in the context of creating a visualization, for instance. Also flexible collaborative capabilities are woven into all of the visualization

Many Eyes in the

Many Eyes is an experiment that was

mostly meant to support collaboration

around visualizations at a large scale by

**Future** 

components, as in a selection and bookmarking model. The site exists as part of a social ecosystem, with discussions on blogs providing a significant amount of attention and activity surrounding Many

As Martin wattenberg, Fernanda Viégas and Frank Van H, the developers of the software write in their paper, the future work will focus on three main areas. First, the site could benefit from a stronger set of community tools. In particular, as the collection of data sets grows, organization and quality control will become increasingly important. It would be beneficial to have mechanisms for the user community to do some of the work in organizing and filtering.

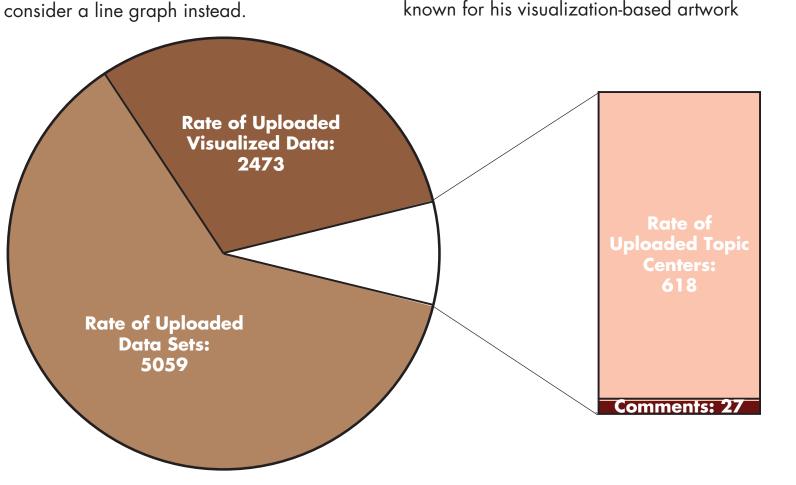
There are also natural extensions to the data model of Many Eyes. Some elements are logistical: the site would clearly gain from tools that allowed easy export from other site. More broadly, the capability of reading data from external sources would open up the possibility of "live" visualizations and the construction of composite visualization applications. All of these possibilities raise interesting questions around versioning and collaboration. A third natural future direction is to

augment the visualizations themselves. The current annotation/pointing scheme is extremely simple, and could be extended in a number of directions. For example, the site could exploit the fact that the annotations are tied to the data to allow inter-visualization brushing and selection. More generally, it would be natural to experiment with other types of visual metadata. For instance, as in one could provide visual indications to show which elements of a data set of have been examined closely.

Finally, it may make sense to use Many Eyes as a platform for rapid user testing of new visualization techniques. Conceivably the site could offer an API so that third-party developers and researchers could test their own offerings. Tests could consist of simply putting new components up and watching whether they are used, or could use more sophisticated methods. For instance, in the graph visualization component users may rearrange the graph layout—and they frequently do. Would it make sense to look at these humancreated layouts to deduce implicit aesthetic criteria? Such approaches seem promising, and point to the research benefits of a broadly available visualization site with a large user base.

#### The Developers

Fernanda B. Viégas is a computational designer whose work focuses on the social, collaborative, and artistic aspects of information visualization. She is a coleader; with Martin Wattenberg, the other developer of Many Eyes; of Google's «Big Picture» data visualization group in Cambridge, MA. Viégas holds a Ph.D. and M.S. from the Media Lab at MIT. Martin Wattenberg is a computer scientist and artist. He holds a Ph.D. in mathematics from U.C. Berkeley. Prior to joining IBM, Wattenberg was Director of Research and Development at SmartMoney.com, a joint venture of Dow Jones and Hearst. His work at SmartMoney included the groundbreaking Map of the Market. From 2005 to 2010, Wattenberg founded and managed IBM's Visual Communication Lab, exploring new forms of visualization and how they can enable better collaboration. A key project was Many Eyes an experiment in open, public data visualization and analysis. Wattenberg is



Many Eyes is an experiment that was mostly meant to support collaboration around visualizations at a large scale by fostering a social style of data analysis in which visualizations not only serve as a discovery tool for individuals but also as a medium to spur discussion among users. However, upto this day, November 14th, 2010 the general rate of comments in this site is less than 1% of the whole activities that take place in this site.

### Refrences:

• http://www-958.ibm.com/software/data/cognos/manyeyes

http://fernandaviegas.com/democratizing\_viz.html

http://www.bewitched.com/ research.html

• Viégas, F. B., Wattenberg, M., & van H, F. (2010, April 14). Many Eyes: A Site for Visualization at Internet Scale. Retrieved December 2, 2010, from IBM Reseach: http:// www.research.ibm.com/visual/papers/viegasinfovis07.pdf

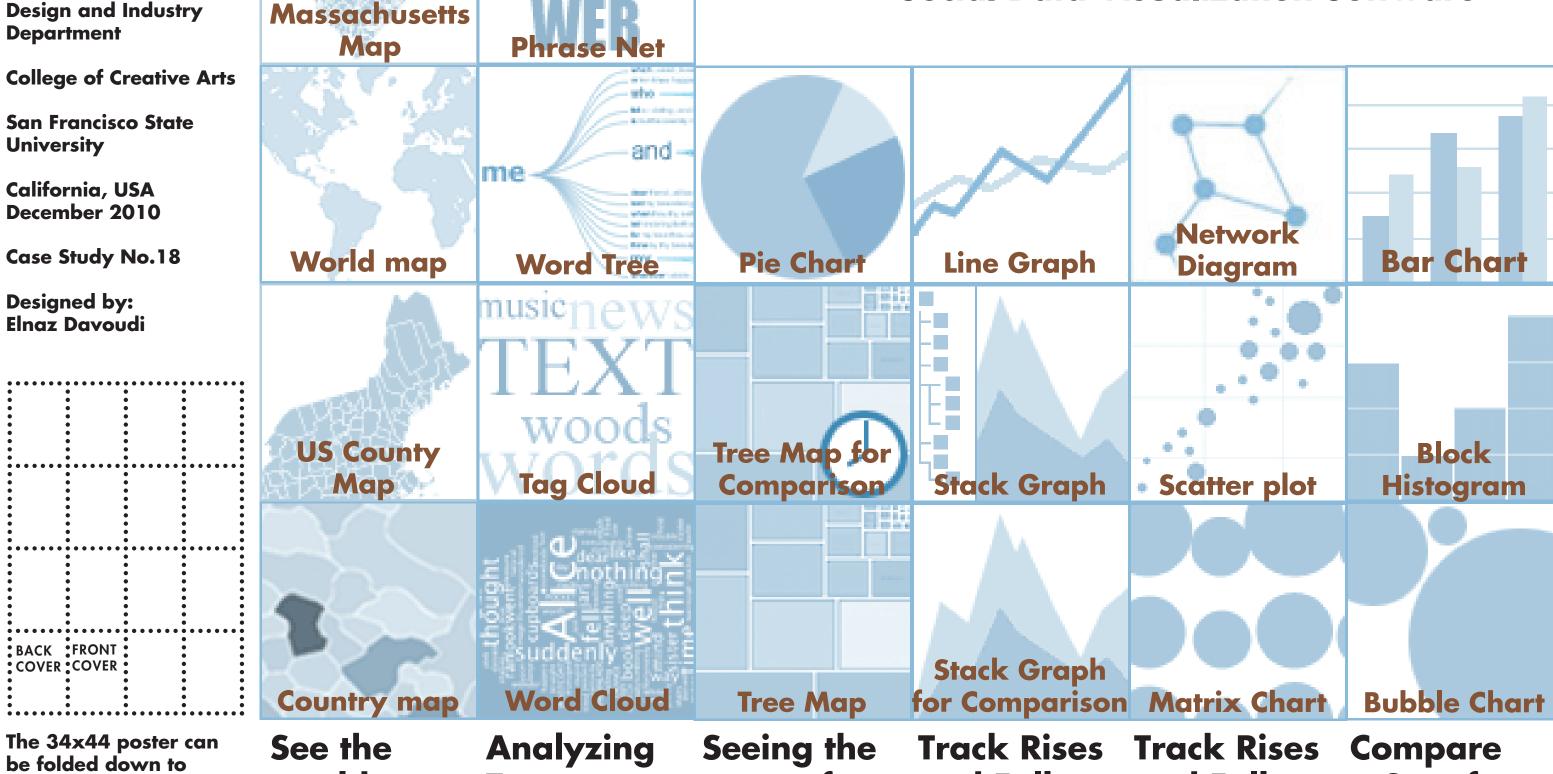
# Many Eyes **Social Data Visualization Software**

and Falls

**over Time** 

a Set of

**Values** 



parts of a

whole

**Text** 

and Falls

over Time