Motion Charts: Telling Stories with Statistics <small>October 2011</small>

Victoria Battista, Edmond Cheng

U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue, NE Washington, DC 20212

Abstract

In the field of statistical and computational graphics, dynamic charts bring rich data visualization to multidimensional information and make it possible to tell a story with statistics. As statistical and computational graphics have advanced, motion charts have made it possible to display multivariate data using two-dimensional bubble charts, scatter plots, and series relationships over temporal periods. This visualization of data enriches its exploration and analysis and more effectively conveys information to users. Incorporating examples from actual business and economic data series, our paper illustrates how motion charts can tell dynamic stories with data.

Key Words: Visualization, motion chart, bubble graph, exploratory data analysis, statistical graphics, multidimensional data

1. Introduction

A motion chart is a dynamic and interactive visualization tool for displaying complex quantitative data. Motion charts show multidimensional data on a single two dimensional plane. Dynamics refers to the animation of multiple data series over time. Interactive refers to the user-controlled features and actions when working with the visualization. Innovations in statistical and graphics computing made it possible for motion charts to become available to individuals. Motion charts gained popularity due to their use by visualization professionals, statisticians, web graphics developers, and media in presenting time-related statistical information in an interesting way. Motion charts help us to tell stories with statistics.

2. Data Visualization

Advancements in technology are filling up our world with information. The number of data sets and elements within these datasets are growing larger, the frequency of data collection is increasing, and the size of databases is expanding exponentially. Statisticians face new challenges in presenting the richness of the raw data with simplicity and clarity. Putting this vast amount of data into meaningful charts and graphs has become more troublesome. Most statistical graphics are presented in two-dimensional forms limited to the viewable size of a single sheet of paper or computer monitor. General audiences are not trained to analyze complicated graphics and cannot instantly understand the information in highly technical datasets.

Data visualization resolves this problem by mean of a process – acquire, parse, filter, mine, represent, refine, and interact with the data. It shifts the user limitation on filtering through the statistics to understanding the meaning behind all of the data. Data visualization illuminates the message behind the data.

3. Motion Charts

3.1 Data presentation
A motion chart is an animated, multidimensional, data presentation tool. It is a way of presenting multiple elements, or dimensions, of the data series on a single plane. As the name implies, a motion chart involves data series over temporal intervals using animation to show the changes in the series within the period.

One example is the number of jobs in the United States over the past 25 years with jobs on the y-axis and years on the x-axis. A traditional column chart or line chart shows this relationship in a simple and static way, while a motion chart can vary the descriptive characteristics of points and lines over time to illustrate other elements of how the U.S. labor market behaved during the ups/downs of the business cycle. Motion charts can be designed to explore more than one aspect of a series at once. For example, a color can represent a single series and the size or the shape of each point can represent a certain descriptor of that series at that point in time.

3.2 History
The pioneer of motion charts is Professor Hans Rosling at Sweden's Karolinska Institute. Dr. Rosling practices medicine in Sweden and collects and analyzes data about international health and disease. He researches global economic and social issues among countries in various developmental stages, and his research makes extensive use of empirical data. In February of 2005, Dr. Rosling co-founded the non-profit Gapminder Foundation, through which he and a team of graduate students developed the Trendalyzer software. This software is used to create dynamic charts and graphics based, as of September 2011, on nearly 500 highly verifiable statistical series about global health and economics. The team continues to expand the number of series in this data visualization tool, including series in their visualization tool only after checking for continuity, quality, and a reasonable assurance that the data will be produced in the future. In this way Gapminder can preserve the quality of the data incorporated in their software.

Dr. Rosling has used the Trendalyzer software in his lectures around the world, including his famous 2006 TED lecture presenting visually dynamic and easily understandable statistics about international development and health. Dr.Rosling's TED lecture captured the attention of both the statistical community and the business world. Google's interest was piqued, and they acquired the Trendalyzer software in 2007. By 2008, Google had made their version of this software, called Google Motion Charts, available to the public, allowing users to input any data they choose.

3.3 Significance
Dynamic motion charts allow us to effectively describe several aspects of a series of data in a single plane. Motion charts easily show multiple cross-sectional data series with temporal animation. These charts minimize the presentation of „chartjunk', or non-data details, in quantitative information displays, making use of high data-ink ratio and high data density of the graphic, meaning that each detail of a motion chart is used to describe an additional aspect of the dataset. Motion charts plot thousand of data points with clarity and precision.

Most motion charts are designed to allow for users' interaction and control over display properties. The complexity of analyzing the underlying data is simplified by the user's ability to control how data is displayed in the visualization. Users can interact with the graphics to analyze the relationships among series. Built-in features allow switching variables between axes, series selection, scaling settings, and different types of graphical representations for the same data. Users can effectively assemble the interested elements and perform necessary analytics, without concerning themselves with the actual dataset.

3.4 Creation and development

Motion Charts can be created with visualization kits available for free or for purchase. Generally, software with the capability to produce motion charts is simple to use. Data can be imported and converted to a readable format by the software. Users can set the display attributes based on analyses of interest. Web-based tools can also easily create motion charts. Pre-formatted data are uploaded to the web-based tools which then create the motion charts for display.

Chart 1: List of tools to create Motion Charts

Name of Tool	Format	Availability	Data Input	Skill Level
Google Motion Charts Gadget	Software	Free	You can input your own data	Low to Mid, somewhat intuitive
Gapminder World	Software	Free	Nearly 500 publicly collected data series, unable to input your own data	Low, very intuitive
Trend Compass	Software	License $3,500	You can input your own data	Low, fairly intuitive
JMP by the SAS Institute	Software	License $1,000+	You can input your own data	Low, fairly intuitive
Flash and others	Web	Varies	You can input your own data	High, program in various languages
Google APIs	Web	Mostly free	You can input your own data	High, program in various languages
Tableau Public	Web	Free	You can input your own data	Low, very intuitive

Software-based visualization tools are generally simple to use. Some software packages capable of creating motion charts are Gapminder World, Google Docs Gadget, Trend Compass, and JMP. The data are prepared in a designated format readable by the software. The software acquires, parses, and represents the data depending on its functionalities. Most visualizations enable interaction and control from the user to customize the display of the information, though in some of the software packages, the display control is somewhat limited. For example, in Google Motion Charts Gadget, customization options beyond axis selection and scaling, color of series, and size of series are limited. Gapminder's motion charts are limited to selected international social and economic series though these series cover an extremely wide range of data that a user can choose.

Web-based visualization tools offer the flexibility for developing interactive controls and displaying customized graphics. Motion charts created thru web programming languages are often displayed directly on web pages. Creating web-based motion charts requires computer technical skills such as web page authorizing, computer language programming, and graphic design. Often times, the skills to develop programming based visualization might be far beyond a statistician's toolkit. Examples of web-based motion charts are Google API, Flash, and HTML5.

Whether choosing a software- or web-based development kit, the most important factor in planning is to know the intended audiences. Think about what an audience is supposed to perceive, and how they are accessing the graphics. Consider the level of control given to the audiences over the graphics.

4. Motion Charts with Labor Force Statistics

4.1 Bureau of Labor Statistics
The Bureau of Labor Statistics (BLS states as its mission:

> "The Bureau of Labor Statistics of the U.S. Department of Labor is the principal Federal agency responsible for measuring labor market activity, working conditions, and price changes in the economy. Its mission is to collect, analyze, and disseminate essential economic information to support public and private decision-making. As an independent statistical agency, the BLS serves its diverse user communities by providing products and services that are objective, timely, accurate, and relevant."

The BLS publishes data on inflation, prices, employment, unemployment, pay and benefits, occupations, productivity, and many other labor related subjects. Current data and analysis produced by the BLS are made available in the forms of news releases and websites. The BLS also publishes bulletins, periodicals, and reports on facts, analysis, and research based on its data. These economic and social statistics provide a wealth of information on the labor market for the public.

4.2 Current Employment Statistics
The Current Employment Statistics program surveys payroll records of more than 410,000 worksites on a monthly basis and provides detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls for the Nation, all States, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and more than 400 metropolitan areas.

The monthly employment series is one of the seven Principle Federal Economic Indicators published by the BLS. The change in employment is a key indicator of the state of the economy. The growth or decline in jobs, hours, and earnings are keenly watched measures of current economic conditions.

For policy making, Congress uses data from the Current Employment Statistics program to help make policy decisions on economics and social issues. Business uses earnings data for creating escalation clauses, decision making, negotiating contracts, and business planning. Local government uses the data to evaluate the economic health of States and areas. Academics use the rich data series for research and analysis.

4.3 Motion Chart with U.S. Employment Series
Economic data has traditionally been shown in static tables and charts. This presentation
style serves the purpose of relaying accurate and nonbiased quantitative data to the users.
Users assemble the necessary data from pre-defined displays and analyze them for
valuable information. Figures 1 and 2 illustrate a common display of economic data using
line graphs.

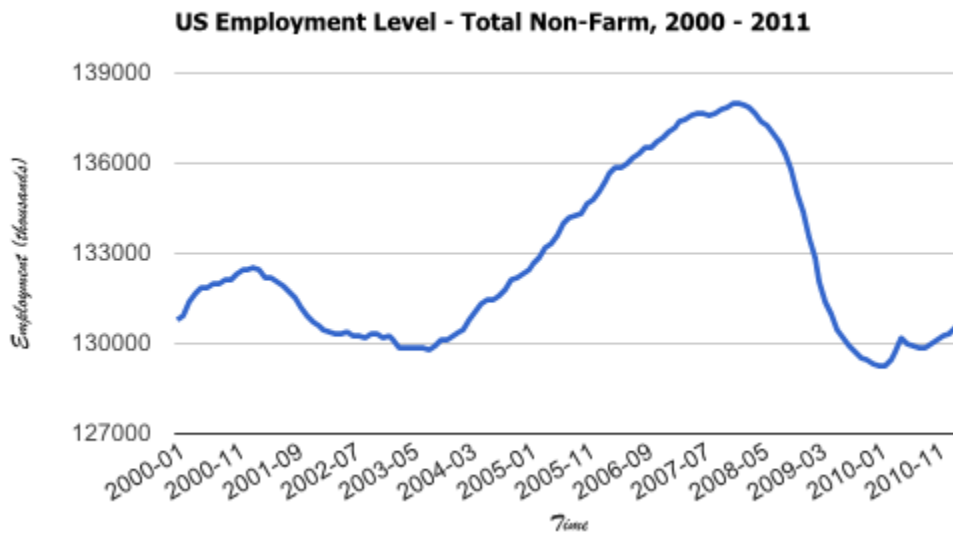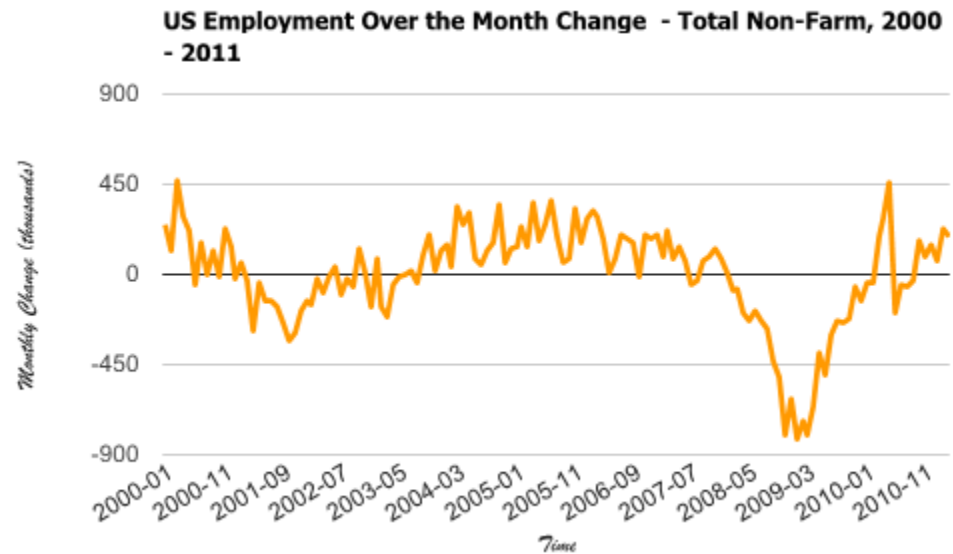Figure 1: U.S. Employment Level – Total Non-Farm, 2000-2011

Figure 2: U.S. Employment Monthly Change – Total Non-Farm, 2000-2011

The first figure shows the U.S. employment level movement from January 2001 to March
2011. Employment expanded and contracted through the business cycle, with the level
peaking in January 2008. The second figure reflects employment growth and loss over
the same period. The trend illustrates the massive job loss during the 2001 and 2008-2009

recession. With the line graph, time is on the x-axis and employment change is on the y-axis. The display is a static graph limited by the dimension of the X-Y plane.

Motion Charts offer dynamic display and interactive features to explore the data. The same U.S. employment data from the graphs above is used for the following examples. The data were formatted and imported to Google Docs. By loading the Motion Charts Gadget within the spreadsheet, the following visualization (Figure 3 is generated.

To view these graphs in their full dynamic format please visit:
http://tinyurl.com/jsm2011motioncharts

Figure 3: U.S. Employment– Total Non-Farm, 2000-2011
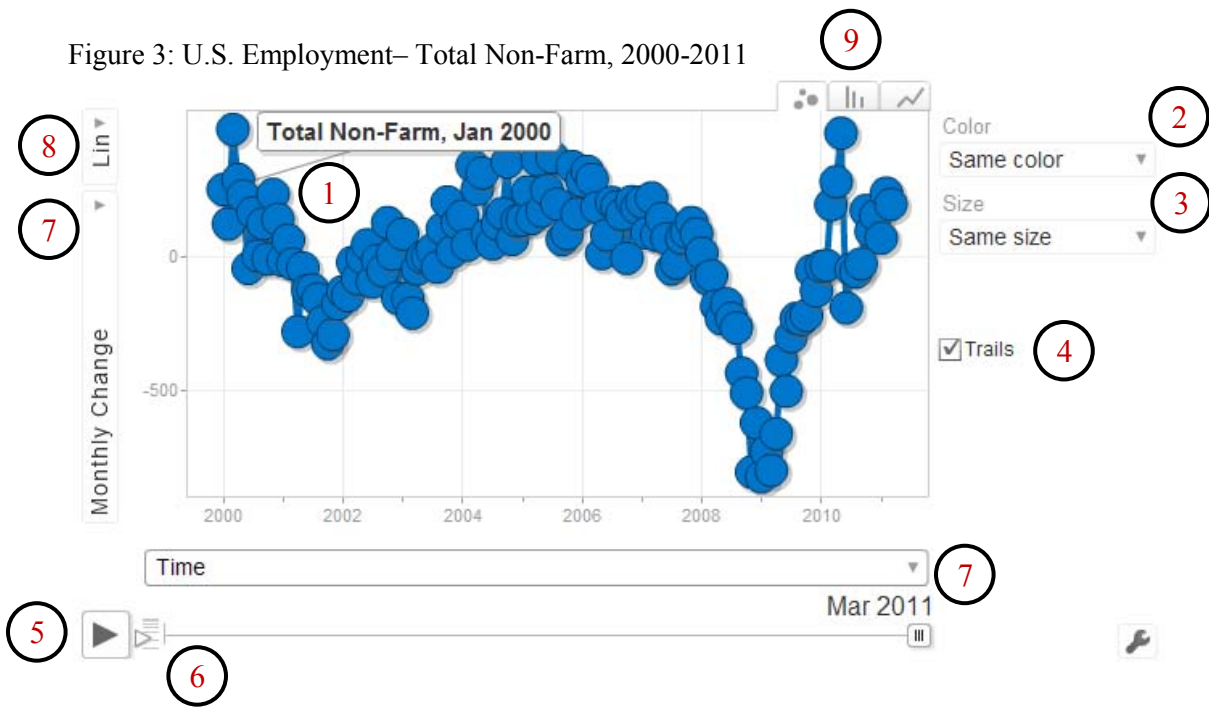
The built-in Motion Charts visualization from Google Docs offers an impressive selection of features which are not available in traditional charts. The handful of configurable interactive features encourages users to explore the data and discover its analytics. Features available in this visualization include:

1   Mouse-over data display
2   Color representation
3   Plot size representation
4   Plot trails
5   Animated time plot
6   Variable speed of animation
7   Changing of axis series and series selection
8   Changing of axis scaling
9   Three available visualizations (bubbles, bars, and lines)

Features and settings are configurable and allow users to customize the display according to their analytical needs. Users can arrange the information and interpret the statistics behind the data content. Motion charts help users to discover the story behind the data.

For example, by selecting the color feature to reflect the monthly employment level, the visualization displays the graph in multiple colors according to total employment. The color data points range from blue, or 128 million in employment, to red, or 138 million in employment. The darker blue the line is, the lower the employment level, the more red it is, the higher the employment level, as shown in Figure 4.

Figure 4: U.S. Employment– Total Non-Farm, 2000-2011
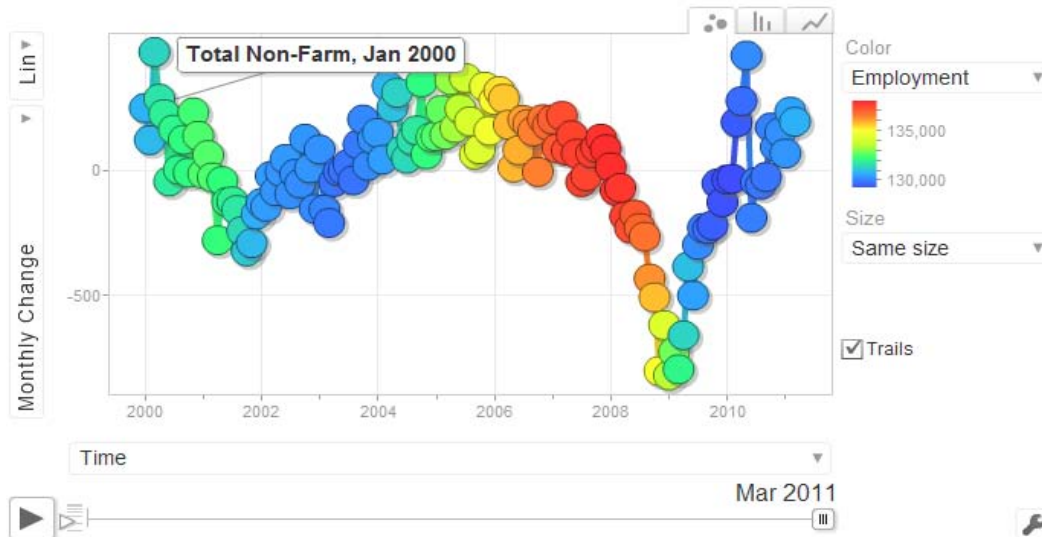
As job growth slowed and began declining thru 2006 and 2007, the U.S. economy held the highest levels of total employment. The downward trend signals the declining growth, while the data point color reflects the peak of employment level. Scrolling over the data points displays the months and employment levels reached in this time period. Nearly 138 million jobs in January 2008 fell to about 129 million by February 2010. At the highest amount of over-the-month job loss in January 2009, the economy still had employment of 134 million.

By disaggregating the U.S. total into major industry sectors, Figure 5 illustrates the shifting of employment composition within various industries.

Figure 5: U.S. Employment Level by Industries, 2000 – 2011
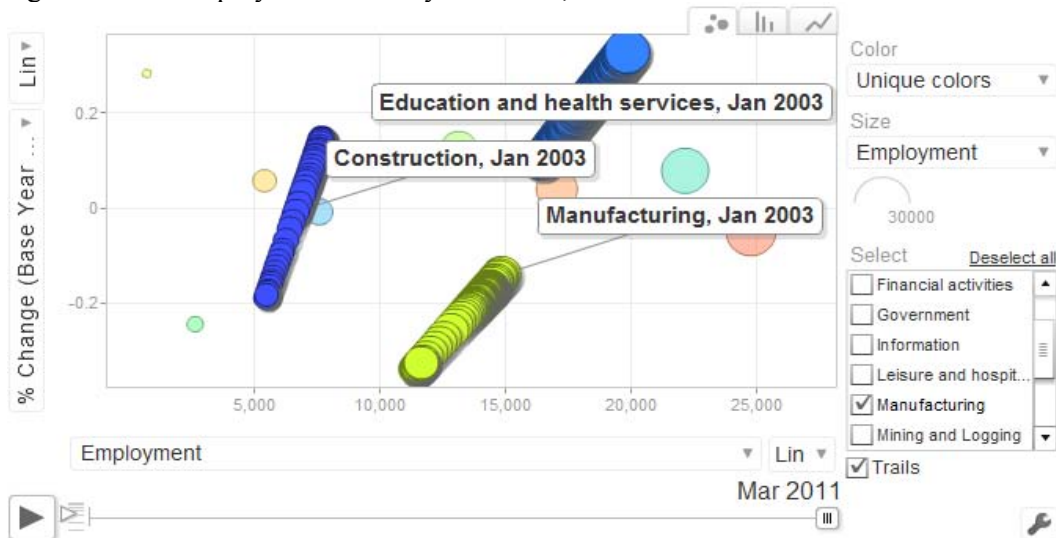
In this visualization, time has been completely removed from the x-axis, and instead employment level is located on the x-axis and employment change on the y-axis. For comparison purposes, the employment change is based on the growth or decline from year 2000 as the base, meaning that a positive 20 percent amounts to a 20 percent gain in employment since 2000. The unique color identifies the industry, and the size of the bubble represents the level of employment in that industry.

With animation, plotting the data points from January 2000 to March 2011, the bubbles representing industry and employment get larger as they move to the right, away from the zero line. Moving to the upper right hand corner indicates higher employment and greater percentage increase since 2000. Moving to the lower left shows less employment and larger percentage decrease since 2000. While not shown in Figure 5, the animation immediately displays the rapid employment contraction pulling almost every industry toward the left corner of the plane when the 2008-2009 recession hits. Nearly all industries decreased payrolls by laying off workers and hiring fewer employees.

Motion charts allow a user to draw trails behind selected series in order to better analyze series' movement over time. Figure 5 has the trails feature selected for manufacturing, education and health services, and construction. The trails show how construction started to climb out of the 2001 recession and then dropped drastically just ahead of the 2008-2009 recession. Manufacturing steadily declined over the decade, but at a less accelerated rate than construction. Education and Health Services is the only U.S. major industry sector which was growing in both the past two recessions. Also notice that the employment levels represented by the positioning of trails between Manufacturing and Education and Health Services appeared to completely switch places and then grow or decline further in opposite directions at nearly the same rate.

Figure 6: U.S. Employment Level by Industries, 2000 - 2011
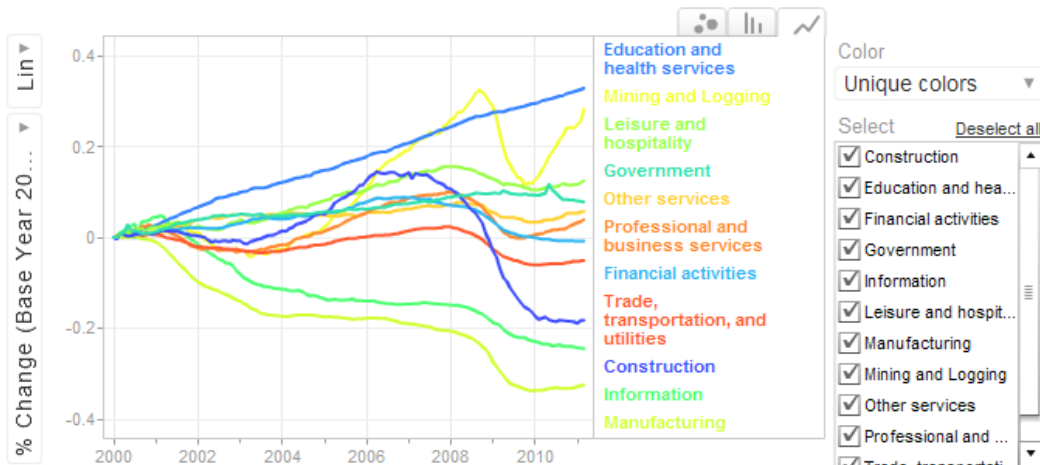
Another graphical representation of the same data is available using the tabs in the right top corner. The first is for the bubble graph used in Figures 3 to 5, the second is a bar graph, and the third is a line graph. Figure 6 is the linear display of the same data series. The visualization shows the light green line representing Manufacturing on a slow steady decline all the way at the bottom of the group of lines, the light blue line for Education and Health Services on a straight gradual trend upward at the top, and the dark blue line showing Construction increasing to a peak in 2006 and dropping off sharply thereafter.

4.4 Motion Charts with Other Labor Statistics

The Phillips Curve is an inverse relationship between an economy's unemployment rate and inflation rate. The plot in Figure 7 is the U.S. Phillips Curve based on the monthly statistics published by the Bureau of Labor Statistics on unemployment rates and consumer price indexes. The period used in the graph covers years 2000 to 2011. The display puts the unemployment rate on the y-axis, inflation on the x-axis, and the net unemployment rate is the color of the bubble.

Figure 7: Phillips Curve – Rates of Inflation and Unemployment, U.S., 2000 - 2011
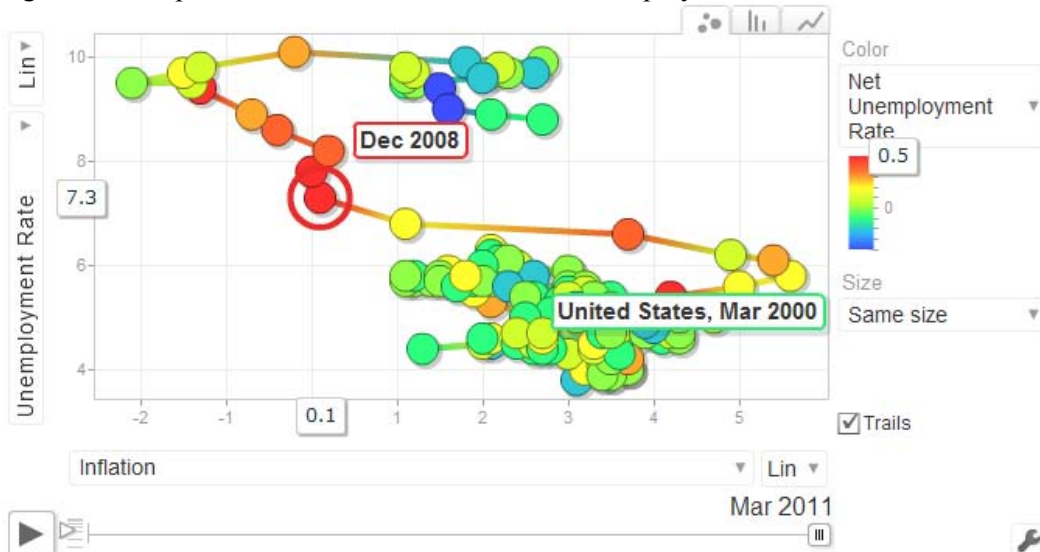
Prior to the 2008 to 2009 recession, the Phillips Curve tracked the U.S. economy with inflation ranging between 2 percent and 4 percent and unemployment staying at or below about 6 percent. A few months prior to the official declaration of the recession on December 2008 by the U.S. National Bureau of Economic Research (NBER, inflation spiked to nearly 6 percent. Then it dropped dramatically as unemployment increased to around 10 percent. Inflation remained negative for a few months in 2009, after which it rose slightly to between 1percent and 3 percent by early 2011. The unemployment rate, however, is still around 9 percent.

The last example is based on statistics published by the International Labor Comparisons (ILC program of the Bureau of Labor Statistics. The ILC adjusts international labor data to a common conceptual framework. Data from countries with differing concepts and methods can be compared in a common methodology. Figure 8 shows the distribution of employment in manufacturing and service sectors for selected countries from 1980 to the present.

Figure 8: Manufacturing to Service Based Jobs, Selected Countries, 1980 – 2010
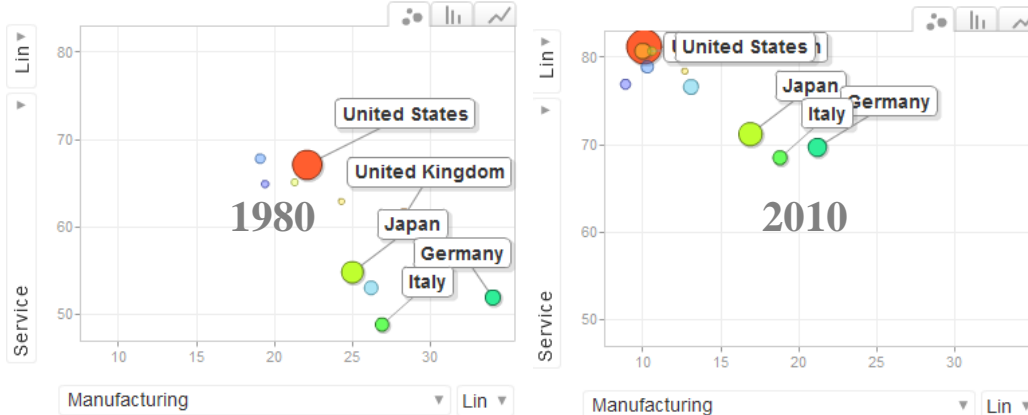
Employment in service industries is on the y-axis and employment in manufacturing industries is on the x-axis. Each country is represented with a unique color and the size of the bubble is based on the total employment level in that country. Playing the animation from 1980 to 2010, the visualization reveals how much these industrialized countries moved from manufacturing economies 30 years ago to service based economies now. The focus of employment in these countries has shifted from manufacturing and production based jobs to skills and service based jobs.

5. Conclusion

Motion charts have a significant role to play in data visualization. Applications are used in science, modeling, media, business intelligence, and analytical software. Motion charts display rich multidimensional data over time on a single plane. The dynamic interactive features allow data users to explore, interpret, and analyze information behind the data series. Motion charts are an excellent and interesting way of presenting valuable information that may otherwise be buried or lost in the data.

6. References

Carlson, Charles (2009). *Numbers as pictures: Examples of data visualization from the Business Employment Dynamics program.* JSM.

Cheng, Edmond (2010). *Explore, Analyze, and Present Statistics Data with Google Visualization Tools.* JSM.

Fry, Ben (2007). *Visualizing Data Exploring and Explaining Data with the Processing Environment.* O'Reilly Media.

Google Docs (n.d.) Retrieved July 31, 2009 from Google: docs.google.com

 "Google Visualization API Developer's Guide". Google.com. 2010. http://code.google.com/apis/visualization/documentation/index.html

Motion charts examples. *JSM2011 Motion Charts Telling Stores with Statistics.* https://docs.google.com/spreadsheet/pub?key=0AqwuDElaZtMIdDhralg5LWY4 LXR5Nk5wVDA1R0MyUGc&output=html

Tufte, E. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.

Tufte, E. (1997). *Visual Explanations*. Cheshire, CT: Graphics Press.

Tufte, E. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

U.S. Bureau of Labor Statistics. (2011). *Consumer Price Indexes: March, 2011.* Washington, DC. http://www.bls.gov/cpi/

U.S. Bureau of Labor Statistics. (2011). *Current Employment Statistics: March, 2011.* Washington, DC. http://www.bls.gov/ces/

U.S. Bureau of Labor Statistics. (2011). *Current Population Survey: March, 2011.* Washington, DC. http://www.bls.gov/cps/

U.S. Bureau of Labor Statistics. (2011). *International Labor Comparisons: International Comparisons of Annual Labor Force Statistics, Adjusted to U.S. Concepts, 10 Countries, 1970-2010.* Washington, DC. http://www.bls.gov/ilc/