

Numbers as pictures: Examples of data visualization from the Business Employment Dynamics program. October 2009

Charles M. Carson¹

¹U.S. Bureau of Labor Statistics, Washington, DC

Abstract

The Bureau of Labor Statistics presents a wealth of information to the general public. Much of this information is published in the form of tables and simple charts; however recent research has shown that humans are able to interpret information from visual cues better than they can interpret through tables of numbers. This paper applies principles from leading researchers in the field of data visualization, such as Dr Edward Tufte, to the datasets available from the Business Employment Dynamics program. New ways of thinking about the presentation of information, combined with tools from Microsoft Excel, SAS, and Google, can better inform the public about current issues in labor statistics.

Keywords: data visualization, graphs, graphics, charts, Business Employment Dynamics, Bureau of Labor Statistics, animation, sparklines, Tufte, Google docs, Trendalyzer,

Introduction

“Vision dominates our sensory landscape,” writes Stephen Few (2004) in his book, *Information Dashboard Design*. Indeed people get most of their information about the world through visual perception. A well-designed data graphic can help someone understand the underlying data much better than a simple table of numbers.

Modern computers have made it much easier to make graphics out of tabular data. However, there are good data graphics and bad data graphics, and a computer alone doesn't make a good graphic. In this paper, I explore some principles of good data graphics, describe the Business Employment Dynamics data series, and apply some simple principles to improve the Business Employment Dynamics graphics. Finally, I share some new tools that are available to create unique ways of visualizing data.

¹ Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

Principles of good data graphics

Edward Tufte (2001) writes, “Graphical elegance is often found in simplicity of design and complexity of data.” (p.177) He follows by laying out suggestions for making the usual graphs and charts better:

Attractive displays of statistical information

- have a properly chosen format and design
- use words, numbers, and drawing together
- reflect a balance, a proportion, a sense of relevant scale
- often have a narrative quality, a story to tell about the data
- are drawn in a professional manner, with the technical details of production done with care
- avoid content-free decoration, including chartjunk. (p.177)

The same principles that allow creative cartoonists to trick the mind’s eye with visual illusions can provide guidance for developing good graphical displays. Gestalt psychology suggests there are six principles of visual perception. Proximity: we tend to group things together if they are closer to each other. Similarity: we group objects together that have similar visual characteristics such as size, shape, or color. Common fate: we group things by the direction they appear to be moving. Good continuation: we see things as whole lines, even where they are interrupted. Closure: similar to continuation, we will finish a shape that appears broken. Area and symmetry: we usually perceive the larger object as background, and we notice asymmetry on a symmetric background.

Business Employment Dynamics

Business Employment Dynamics (BED) is a relatively new program at the Bureau of Labor Statistics. It is one of many products of the Quarterly Census of Employment and Wages (QCEW), which represents about 97% of all U.S. employment and serves as the BLS establishment sampling frame (Spletzer, J., Faberman J., Sadeghi, A., Talan, D., and Clayton, R., 2004, April).

For each of the 9 million establishments included in the QCEW, BED determines which establishments opened, which closed, which expanded, and which contracted. BED then sums the job gains from opening and expanding establishments as well as the job losses from closing and contracting establishments.

These measures add perspective to the economic situation by showing the job creation and job destruction underneath the net change in employment. Over an average quarter, 7.9 million jobs are created at opening and expanding establishments, and 7.5 million jobs are lost at closing and contracting establishments. Even though the net change in employment is very small by comparison, these measures of job churn lend important insight into the health of the economy.

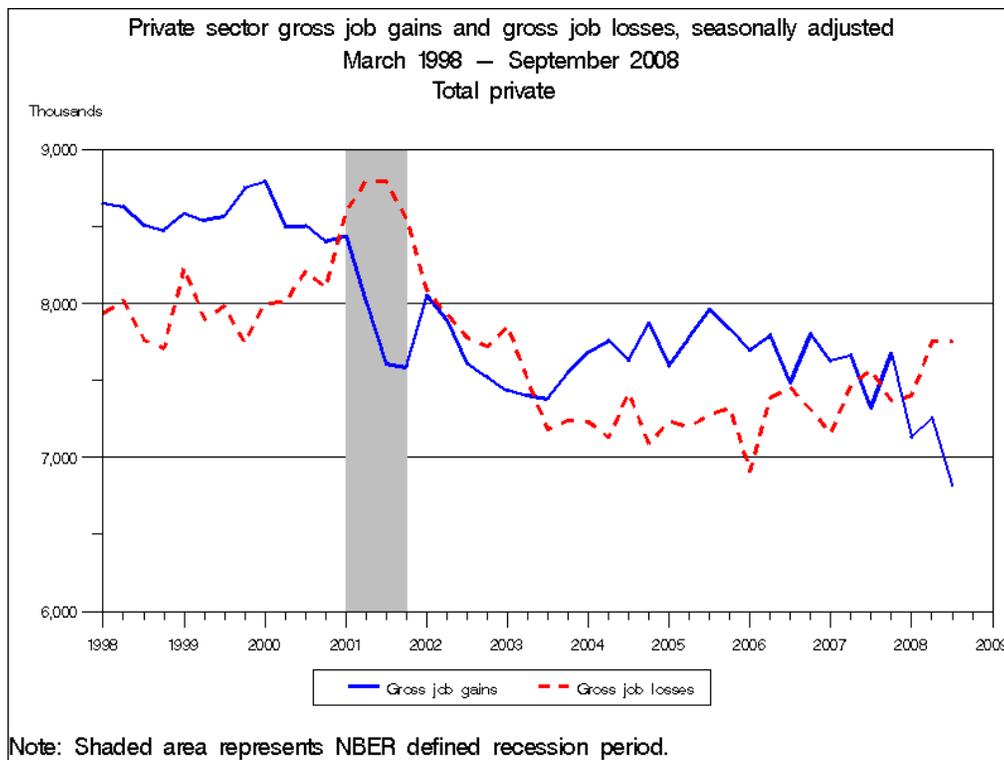
Moreover, since BED data are drawn from a virtual census of employment, there are many different types of detail available. BED publishes establishment data by major industry sectors, by state, and by the size of the change in employment. BED sums data

for firms (using IRS-issued Employer Identification Number, or EIN) and publishes by the size of the firm. BED even breaks out business births and deaths from openings and closings.

With such a rich dataset, it is all the more important to effectively and efficiently present this information. Charts can quickly become cluttered with too much data, but good graphical design can reveal new insights from these data.

Improving on current charts

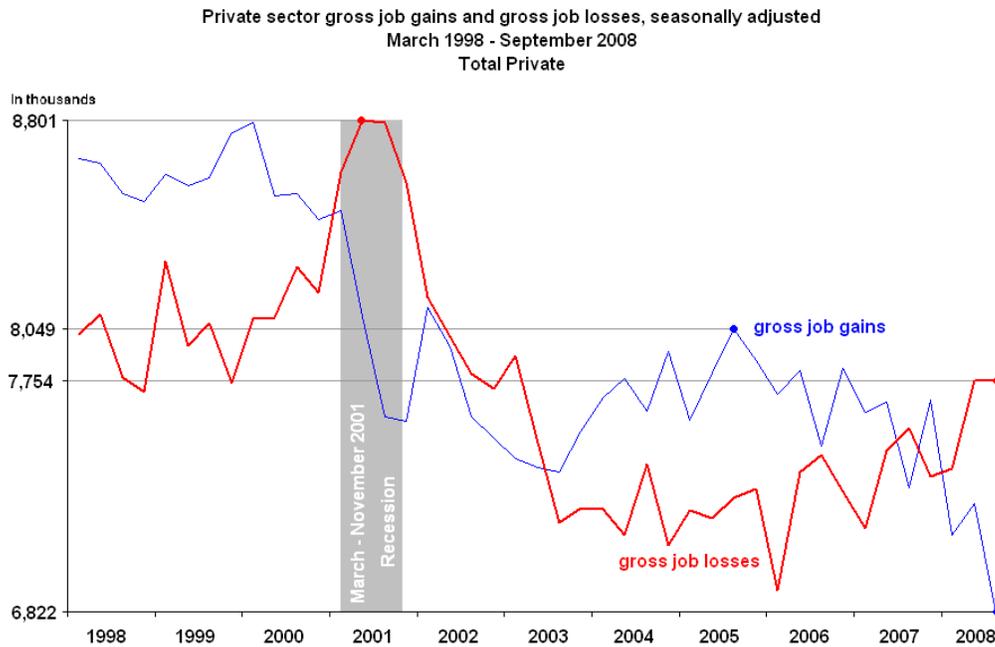
There is nothing particularly wrong about the chart that appears at the top of any Business Employment Dynamics news release (U.S. Bureau of Labor Statistics, 2009). It is a simple line chart, showing gross job gains and gross job losses. A shaded bar highlights the period of the recession of 2001.



There is, however, room for improvement with this graph. First, there is wasted space on the graph itself. While even numbers are pretty, there is no good data-based reason why the y-axis range starts at 6 million and ends at 9 million; neither is there a good reason for gridlines to be drawn at 7 and 8 million. The purpose of the scale of the y-axis is to help data-users know what numbers are represented by the lines of the graph. Instead of picking arbitrary numbers that happen to be even millions, it could use numbers that are directly useful for a data user to define the y-axis. The minimum and maximum of the series is more useful; specific gridlines could highlight other interesting data points, such as the current level of gross job losses, or the peak of gross job gains during the most recent expansion. Moreover, the labeling of the x-axis may confuse data-users, because the tick mark for each year actually shows data for March.

Second, the explanatory text of the graph could be integrated better onto the picture itself. Gestalt principles say we relate things that are close in proximity. By avoiding a legend and putting the text directly on the graph, we can reinforce what each line represents, and prevent people from having to refer back and forth to the legend.

Finally, the eye can focus on the more important parts of the graph by eliminating “chartjunk” such as borders and using grey instead of black for less important elements of the chart.

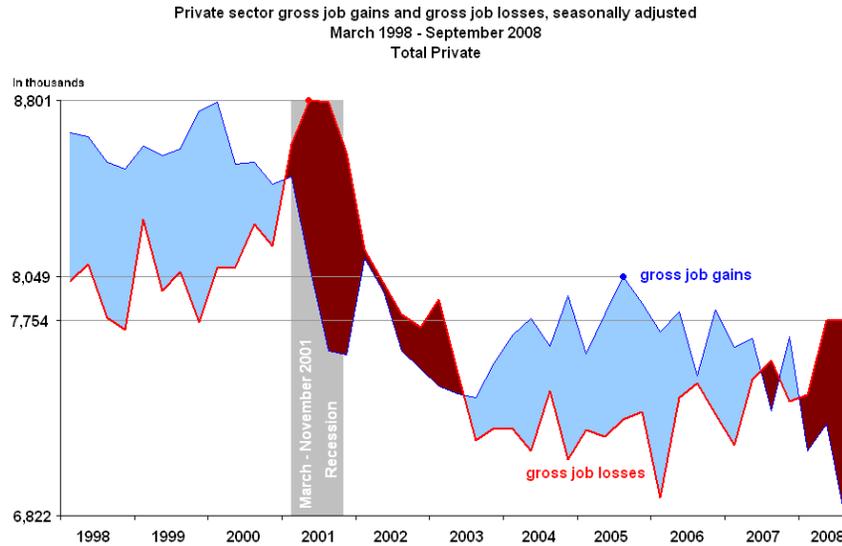


The redesigned chart is simpler. More of the ink (or pixels) used to draw it actually conveys information to the viewer. The scale lets a data user know what gross job gains are in the most recent quarter (6,822,000), as well as what gross job losses were during the peak of the recession of 2001 (8,801,000). The gridlines (reinforced with a dot) now also convey more useful data points by highlighting the current level of gross job losses (7,754,000) and the level of gross job gains at the peak of the last period of expansion (8,049,000). Text needed to describe the chart is integrated into the chart, eliminating the legend and note.

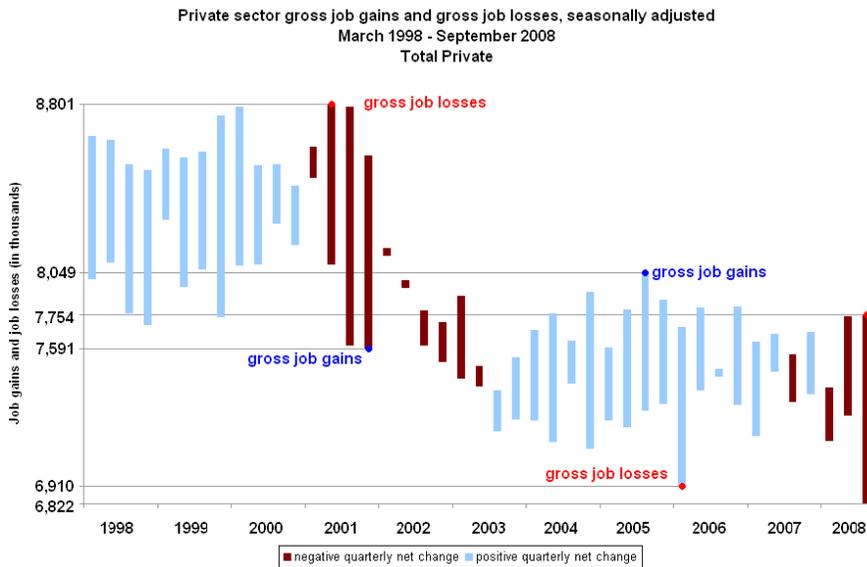
The redesigned chart allows a data-user to see more clearly the patterns in the data. The decrease in gross job gains and the increase in gross job losses since 2005 is more pronounced and observable. It is easier to observe how gross job gains and gross job losses interact during the “jobless recovery” in 2002 and 2003. Finally, the use of the labeling and gridlines makes it possible to see that the peak of job losses during the recession (8,801,000) was slightly higher than the peak of job gains during the previous expansion.

With any chart, it is important to ask, “What are data users trying to explain when they look at Business Employment Dynamics data?” For Business Employment Dynamics data, the simple answer is that data-users are trying to explain the variation in net change.

One of the things that this graph does not effectively highlight is a simple observation: the difference between gross job gains and gross job losses results in the net change in employment. The distance between the gross job gains line to the gross job losses line equals the net change; this relationship can be highlighted and reinforced by filling in that space.



The chart above is the same as the one before. The only difference is that the areas where gross job gains are larger than gross job losses are highlighted in light blue, and the areas where gross job losses are larger than gross job gains are highlighted in maroon. Now, the viewer's attention is focused on the size of these respective areas, and because of the Gestalt principles of area and closure, the viewer sees each of these as whole shapes. This is better, but it still does not really emphasize the vertical distance.



This last chart changes the focus of the viewer by taking the concept of a bar chart and expanding the quantity of information provided. The length of the bar represents the net change because it is the difference between gross job gains and gross job losses. The color of the bar indicates whether the net change is positive or negative (the colors are different shades, so this chart would still be legible if printed in black and white). When there is a negative net change, gross job losses are bigger than gross job gains; this is reinforced on the chart by labeling the top and bottom of these bars. Similarly, when gross job gains are bigger than gross job losses, there is a positive net change.

Because of the principle of similarity, the viewer will group the bars that are similar in color. The recession of 2001 and the following “jobless recovery,” stand out because of the group of maroon bars that begins in 2001 and continues through 2003. Because of the principle of common fate, the bars appear to move down in 2002 and 2003 and only move up slightly in 2004 and 2005, representing the decreased amount of “job churn” in the economy during the most recent period of recovery. Net change (the length of the bars) was smaller in 2007 (leading into the current recession) than in 2005; this is similar to the pattern that occurred in 2000 before the recession of 2001. Compared to simply filling in the area, this chart more effectively shows which quarters have a large net change and which quarters have a small net change, while still demonstrating how gross job gains and gross job losses interact.

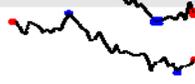
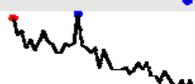
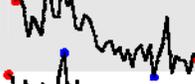
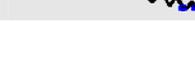
New ways of visualizing data

We continue to learn more about how humans process information. At the same time, computers continue to make it easier to change large tables of numbers into interesting and useful graphics.

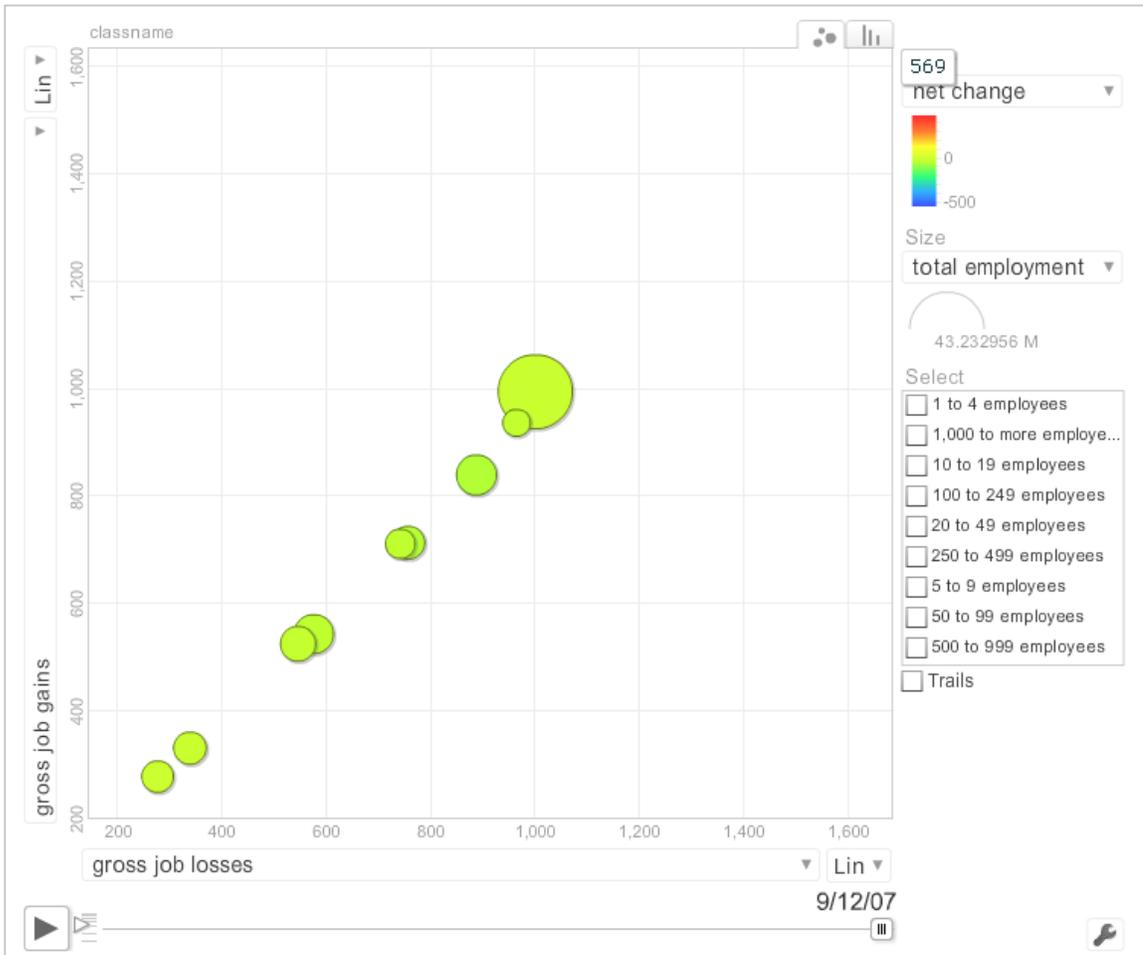
“Sparklines” are a way of visualizing data invented by Edward Tufte (2006). He describes them as “word-sized graphics.” (p. 47) In general, when we look at a chart, we’re only really interested in seeing the overall trend. Sparklines can efficiently convey this information in a very small amount of space. Pairing a sparkline with the number it represents provides context and direction for that number.

For example, in third quarter 2008, the number of establishment births nationwide fell to  187,000. The sparkline helps put this number in context. It is easy to see that this is part of a trend of declining business births, following a period of higher births. We could add a few more data points to the sparkline to expand its usefulness. In third quarter 2008, the number of establishment births nationwide fell to **187,000**  from a peak of **223,000** in December of 2005 and lower than the **192,000** business births in December of 1998. One sentence now identifies three specific numbers that provide context for the most recent data, and the sparkline demonstrates the overall shape and trend of the data.

Gross job reallocation describes the total amount of “job churn” in the economy. It is created by summing gross job gains and gross job losses. For example, gross job reallocation nationwide has been declining since March 1998 when it was 16.1 percent of total employment . Since Business Employment Dynamics produces data for fifteen major industry sectors, a sparkline for each sector could be combined with a table of data. This uses the principles of similarity and proximity, allowing the data user to compare in one space both the numbers and the general shapes and trends.

	March 1998		Sept. 2008	Maximum	Minimum
Total Private	16.1		13.0	16.1 Mar. 1998	12.7 Mar. 2008
Goods-producing	15.7		14.1	15.8 June 2001	13.9 Sept. 2007
Natural resources and mining	39.3		28.2	39.3 Mar. 1998	28.2 Sept. 2008
Construction	27.1		22.2	27.1 Mar. 1998	21.8 Sept. 2006
Manufacturing	9.8		8.0	10.2 June 2001	7.4 Sept. 2007
Service-providing	16.1		12.6	16.1 Mar. 1998	12.4 Mar. 2008
Wholesale trade	13.8		9.7	13.8 Mar. 1998	9.7 Sept. 2008
Retail trade	15.8		12.7	16.0 June 1998	12.5 Mar. 2008
Transportation and warehousing	14.0		10.7	14.2 Dec. 2001	10.6 Mar. 2008
Utilities	7.5		4.8	8.7 Dec. 2000	4.0 Mar. 2006
Information	14.6		10.1	15.6 June 2001	9.1 Mar. 2007
Financial activities	13.5		10.4	14.3 June 1998	10.4 Mar. 2007
Professional and business services	19.4		14.8	19.4 Mar. 1998	14.4 Mar. 2008
Education and health services	11.1		8.5	11.1 Mar. 1999	8.2 Mar. 2008
Leisure and hospitality	21.3		17.7	21.3 Mar. 1999	17.2 Mar. 2008
Other services (except public admin)	18.3		14.9	18.3 Mar. 1998	14.9 Mar. 2008

As another tool available to designers of data graphics, Google has made its Trendalyzer software available for free on the internet. This software makes it possible to design animated graphs that change through time. What makes these graphs truly interesting and data-driven is their ability to let the user chose what comparisons to make, and what elements to study. Clickable options allow the user to pick what element is on each axis, and the user can play, pause, or drag to view a specific moment in time.



Conclusion

Many of the figures presented in this paper were produced using Microsoft Excel, which is commonly available to many data-users. The sparklines were created using SAS, but could just as easily have been made by any graphing program. The motion chart is available for free through Google Docs docs.google.com, and it is possible to upload spreadsheets from other programs such as Microsoft Excel. Chart making programs are useful tools, but they are only useful when graphics designers think critically about how best to display the data.

Well-designed data graphics can provide useful insight by helping data-users find the information they want. Graphics should be designed with care to be data rich and avoid distractions and “chart junk.” Designers don’t need to reinvent the wheel, but they can use the Gestalt principles, as well as other research and other graphs to apply to their own data.

References

- Few, S. (2004). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press.
- Google Docs (n.d.) Retrieved July 31, 2009 from Google: docs.google.com
- Spletzer, J., Faberman J., Sadeghi, A., Talan, D., and Clayton, R. (2004, April). Business Employment Dynamics: New Data on Gross Job Gains and Losses. *Monthly Labor Review*, 29-42 <http://www.bls.gov/opub/mlr/2004/04/art3full.pdf>
- 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.
- Tufte, E. (2006). *Beautiful Evidence*. Cheshire, CT: Graphics Press.
- U.S. Bureau of Labor Statistics. (2009). *Business Employment Dynamics: Third Quarter, 2008*. Washington, DC: Government Printing Office. <http://www.bls.gov/bdm/>
- Ware, C. (2000). *Information Visualization: Design for Perception*. San Diego, CA: Academic Press.