Track National Unemployment, Job Gains and Job Losses

Analysis and Critique of a Wall Street Journal Data Visualization

by Kim de Bie

11 December 2015

minor Programmeren: Data Processing

dr. Gosia Migut

University of Amsterdam, 2015-2016

Abstract: *In this paper, the quality of the Job Market Checker visualization as published by the Wall Street Journal is assessed. The different elements of the visualization are analysed for their ability to display and communicate data in a clear, unbiased, intelligible and attractive format. Although overall, the visualization seems to meet these goals, there are a number of points where the visualization is lacking quality. Most notably, Tufte's (2001) laws of graphical integrity and data density seem to have been violated. Lastly, some suggestions for improvements are made.*

Introduction

Data visualizations, in particular interactive visualizations published on the web, are increasingly becoming a key tool for journalists. In his book, Alberto Cairo distinguishes three levels in the process of understanding data: raw data, structured information and knowledge (Cairo 2012). Data visualizations are part of the second stage, namely that of structured information, and will help the viewer move to the next stage, namely that of knowledge. As such, whereas the role of the traditional journalist was to provide the public with processed 'knowledge' in the form of test articles, the modern journalist has moved one rank up the ladder as she also operates in the domain of providing structured information. This fundamentally changes the role of the journalist: interactive data visualizations make large datasets comprehensible for every viewer, whereas the public was traditionally limited to the interpretations and insights that a single journalist gathered from the data.

To facilitate this independent knowledge creation by the public, it is of crucial importance that data is displayed in a clear, unbiased, intelligible and attractive format. This paper will examine whether these goals have been met in the "National Unemployment, Job Gains and Job Losses" visualization as published by the Wall Street Journal (van Dam and Lightner 2015). Firstly, the visualization and its goals will be described. After this, the visualization will be analyzed vis-à-vis criteria established in the literature; lastly, some suggestions for improvement will be made.

Description

As its title indicates, the purpose of the "National Unemployment, Job Gains and Job Losses" visualization is to display the changes in employment rates within sectors and across the country, so that changes as well as abnormally high or low values can be identified. Translating this to abstract terms (Munzner 2009)[1], the visualization is intended to compare and to find anomalies. The visualization, as it is published by the Wall Street Journal, is intended for the general public and does not require any specialist knowledge. The data used is from government sources.

The visualization consists of two main sections, where the first describes changes in employment for different sectors and the second focuses on changes in national employment rates. In the first section, a dot graph contains data on relative job gains and losses per sector. By hovering over dots, the sector that the dot relates to, the corresponding month and the change in employment are displayed. Moreover, the same sector is highlighted across all months. Below this, the exact changes in employment per sector for the most recent month are displayed. The second section covers changes in national employment. A colored block grid displays changes in employment rates per month, and a line graph shows the same data.

[1] Munzner (2009) distinguishes between describing a task in *domain* and *abstract* vocabulary. An important step of a visualization design process is translating the task from the topic-specific or domain terms to general or abstract tasks, such as 'compare', 'expose uncertainty'.

Analysis and critique

In this section, the visualization will be assessed element by element, based on concepts and theories as found in the academic literature on data visualization. In the first object, the *dot graph*, the dots or circles are the most prominent visual elements. The dots are encoded with color, position and text labels (Carpendale 2003). Color and position are used to indicate order, whereas the text labels are meant to display more detailed information. The color usage in the dot graph is theoretically problematic as the colors are part of a rainbow scale (Borland and Taylor 2007), but because a clear legend with the colors is included and the colors are always in the same order, little harm seems to be done in practice. Color is also used to highlight the sector that is being hovered over in other months. Here, color is useful in the sense that all other dots highlighted are clearly visible. However, with regards its encoding through position, this graph is somewhat problematic. Implicitly, the overall graph seems to give the impression that the overall number of jobs decreased during the recession, and increased in other time periods. However, instead the graph actually visualizes the number of sectors in which jobs increased or decreased, which tells us nothing about the absolute increase or decrease: theoretically, more jobs may have been created during the recession than e.g. in 2013. In this sense, the visualization may – at first sight – give a wrong impression of the actual patterns in the data, and Tufte's Lie Factor therefore seems quite high (Tufte 2001). Besides, the dots highlighted on hover seem to display a pattern across months, which is quite difficult to distinguish in the first place, but which is potentially misleading in as well. The position of the selected sector in a month only indicates how many sectors had larger and smaller employment changes. Theoretically, a sector with exactly the same growth rates for each month could be at the top of the bar for one month, and at the bottom for the next. However, the visualization would in this case seem to display large variation within a sector, despite this being only partially true (or not at all): again, the Lie Factor seems rather high, and graphical integrity seems harmed (Tufte 2001).

Below the dot graph, *a table with bar chart* gives a specific overview of employment changes across sectors for the most recent month. Visual elements include bars and text, where the bars are encoded using color and length. Comparing length on a common axis is a task that can be easily completed by humans (Cleveland and McGill 1984), and indeed this part of the table seems quite clear. However, this element of the visualization builds on the same data as the graph above: the data-to-ink ratio (Tufte 2001) is therefore quite low. However, it could also be argued (and this seems a matter of taste) that displaying the same data (plus some additional variables) in a different format actually increases understanding as it highlights different parts of the dataset. Lastly, because the table is mainly built on text, it is not very visually attractive.

The third element of the visualization is the first that relates to variation in employment rates across the nation. This *block grid* is built up of one block per month over around fifty years. The blocks are encoded using color and text on hover. The usage of color is quite clear, at least, for the average viewer. For individuals with color blindness, the shades of green and red that are used will be much

more difficult to distinguish (Ware 2008). An interesting addition is the possibility to highlight recessions, which makes clear that employment rates did not necessarily decrease during these periods. However, it is unclear what displaying the data in this way adds to, for example, the line graph below. For every month in a given year, the colors are nearly always exactly the same, which shows that there are no clear patterns within years, and thus there seems little point in separating the months of the year so explicitly.

The last element of the visualization is a simple *line graph*. This line graph is linked to the block grid above: on hovering over the blocks, a pointer indicates the position of that block in the line graph and vice versa. Again, the two bottom graphs contain exactly the same data. Whereas in the upper two elements, on employment changes within sector, different variables of the same dataset were included, no additional variables are included in the line graph compared to the block grid. Thus, the two graphs merely express design variation and not data variation, and the data density is again very low (Tufte 2001).

Improvements

This section will describe two main ways in which this visualization could have been improved. Firstly, for both topics, **data density** could have been increased. Both the data on changes within sectors and that on changes across the country has been displayed twice. For the first topic, sector changes, the additional variables that the table provides could for example have been included in the text boxes already present in the dot graph. For the second two graphs, the line graph only would probably have been sufficient; the block grid does not provide extra information at all. As such, the visualization could have consisted out of two elements only, which would have made the web page less overwhelming: in its current format, it is quite long, containing four elements that have to be interpreted and understood separately. Secondly, for the first graph, the potential **Lie Factor** issue as described above could have been resolved by displaying the sector data in a different way. For example, grouped line graphs with one line for each sector could have been created instead of the stacked dots that are currently used. In this way, the changes in employment within different sectors would have been much easier to perceive.

Conclusion

The different elements of the visualization are quite helpful in getting a better picture of changes in employment within and across sectors in the United States. Nonetheless, there is quite some room for improvement. Firstly, the visualization shows that more is not always better: the visualization displays the same data in multiple occasions, which makes the web page look less attractive, but it is unclear what is gained. Secondly, this visualization once again highlights the importance of displaying data in a manner that leaves no room for bias. Although objectively, the authors are not 'lying with the data', it seems likely that the average viewer draws different conclusions from elements of the visualization because of issues of clarity. With the suggested improvements implemented, it seems that the Wall Street Journal has created a valuable tool for understanding employment issues in the United States.

Bibliography

Borland, D., and R.M. Taylor. 2007. "Rainbow Color Map (Still) Considered Harmful." *IEEE Computer Graphics and Applications* 27 (2): 14–17. doi:10.1109/MCG.2007.323435.

Cairo, Alberto. 2012. *The Functional Art: An Introduction to Information Graphics and Visualization*. Berkeley, California: New Riders.

Carpendale, M. S. T. 2003. "Considering Visual Variables as a Basis for Information Visualisation." http://prism.ucalgary.ca/handle/1880/45758.

Cleveland, William S., and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79 (387): 531–54.

Munzner, Tamara. 2009. "A Nested Process Model for Visualization Design and Validation." *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 921–28. doi:10.1109/TVCG.2009.111.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd edition. Cheshire, Conn: Graphics Pr.

Van Dam, Andrew, and Renee Lightner. 2015. "Track National Unemployment, Job Gains and Job Losses." *Wall Street Journal*. November 6. http://graphics.wsj.com/job-market-tracker.

Ware, Colin. 2008. *Visual Thinking: For Design*. Burlington, MA: Morgan Kaufmann.