

Data Processing - Design Critique 7

Vincent Erich
10384081

December 15, 2015

1 Introduction

In this paper I will critically evaluate the ‘Job Market Tracker’ visualization by Andrew Van Dam and Renee Lightner¹. I will evaluate the visualization using a number of concepts and ideas that have discussed during the course.

This critique will proceed in five parts. In Section 2 I will briefly explain the problem domain of the visualization, which tasks can be achieved, and what the intended goal is. In Sections 3, 4, and 5 I will talk about Tufte’s (1983) design principles of graphical integrity, Tufte’s (1983) visualization design principles (and whether they are adhered to), and William’s (2014) graphic design principles respectively. In Section 6 I will conclude with a short summary of my observations.

2 The visualization

The ‘Job Market Tracker’ visualization allows users to track national unemployment rates, and job gains and job losses in different sectors. The visualization consists of four graphs/plots: an interactive dot plot and an interactive table for the job gains and job losses, and an interactive heat map and interactive line graph for the national unemployment rates.

2.1 Job gains and job losses

As described above, the visualization allows users to track job gains and job losses in different sectors. This data is visualized using an interactive dot plot and an interactive table. In the interactive dot plot, time is plotted on the x-axis (in months, from April 2006 to November 2015), and the number of sectors gaining/losing jobs is plotted on the y-axis. Each month on the x-axis is represented by a column of color-coded dots. Each dot represents a sector (e.g., Mining and logging, Construction, Education, etc.), and the number of dots above the zero line is the number of sectors gaining jobs in that month,

¹Source: <http://graphics.wsj.com/job-market-tracker/>

and the number of dots below the zero line is the number of sectors losing jobs in that month. You can hover over a dot to see the sector it represents and the job gain/loss in percentages in that sector relative to the previous month (that is the color coded variable). Furthermore, if you hover over a dot (i.e., a sector) in a specific column (i.e., a month), that same dot is highlighted in all the other columns.

The interactive table gives a detailed overview of job gains and job losses for a specific month. The interactive table consists of five columns: ‘Sector name’, ‘Sector size’, ‘Monthly change (total jobs)’, ‘Monthly change (pct.)’, ‘Months rising’, and ‘Months falling’. If you click on a column in the interactive dot plot, the interactive table is updated to show the detailed information for that month. You can then sort the columns in the interactive table, and the interactive dot plot will update accordingly (i.e., you can alphabetically sort the ‘Sector name’ column, with all the sectors that gained jobs coming first, followed by those that lost jobs, and the dots in the dot plot will re-order accordingly). Furthermore, you can hover over a row (i.e., a sector) in the interactive table and the corresponding dots are highlighted in the interactive dot plot.

2.2 National unemployment rates

The visualization also allows users to track national unemployment rates. This data is visualized using an interactive heat map and an interactive line graph. In the interactive heat map, each row represent a month, and each column represents a year (from 1948 to 2015). Thus, each cell in the heat map represents a specific month in a specific year. The cells in the heat map are color-coded based on the unemployment rate. You can hover over a cell to see the exact unemployment rate in that month. Furthermore, you can select the option ‘Show Recessions’ and this will cause certain cells to be marked with a little black square (i.e., those cells represent months in which there was a recession).

The interactive line graph shows the unemployment rates over time. If you hover over the graph, a tooltip appears that shows the month, year, and unemployment rate in that month. Furthermore, if you hover over the graph, the corresponding cell and month in the interactive heat map are highlighted.

The interactive heat map and interactive line graph show the national unemployment rates by default. However, you can adjust both graphs/plots by selecting an unemployment rate based on gender, age (in years), race/ethnicity, or education (25+).

The goal of the visualization is to give those interested an insight in the history and development of the job market, and how the recession(s) has/have affected the job market.

In the subsequent sections I will go into detail about the visualization and evaluate the graphs/plots described above.

3 Graphical integrity

Communication should be the main goal of data visualization. Since a lack of graphical integrity can lead to miscommunication, graphical integrity is the most important aspect of graphics design. Tufte and Graves-Morris (1983) defined a number of principles of graphical integrity. In this section I will apply some of these principles to the ‘Job Market Tracker’ visualization.

In order to prevent graphical distortion and ambiguity, a graph should be clearly labelled, and scales should be appropriate (e.g., a bar chart should always start at zero). The graphs/plots in the ‘Job Market Tracker’ visualization is clearly labeled: all the graphs/plots have a title, and, where appropriate, a short explanation and a legend. Furthermore, important events in the data are labeled (e.g., recessions). The scales of the graphs/plots are also clearly labelled and appropriate. However, I do have a minus: the title of the interactive table is never updated (as opposed to the titles of the interactive heat map and interactive line graph). The title of the interactive table always says: ‘*Detailed View: Job Gains and Losses for November 2015*’. The interactive table gives a detailed overview of job gains and job losses for November 2015 by default. However, if you click on another column in the interactive dot plot (i.e., another month, see the previous Section), the title is not updated. This can lead to miscommunication.

Graphs sometimes under or over represent trends in the data. Tufte and Graves-Morris (1983) defined the *Lie Factor* as the size of the effect shown in the graph, divided by the size of the effect shown in the data. If the Lie Factor of a graph is smaller than one, then the graph under represents the trend in the data, and if the Lie Factor of a graph is greater than one, then the graph over represents the trend in the data. A distorted Lie Factor (i.e., a Lie Factor smaller or greater than one) is usually caused by using inappropriate scales and/or 3D effects in the graph. As described above, the scales of the graphs/plots in the ‘Job Market Tracker’ visualization are all appropriate and clearly labelled. Therefore, the Lie Factor of the visualization is not distorted.

Another important principle of graphical integrity is to show data variation and not design variation. This basically means that you have to show the data (i.e., the data is the most important aspect of a graph). The graphs/plots in the ‘Job Market Tracker’ visualization clearly show the data, and the data is not obscured. Furthermore, the data is shown in multiple ways (i.e., overview and detail). The ‘Job Market Tracker’ visualization clearly shows data variation and not design variation.

4 Visualization design principles

Besides the principles of graphical integrity, Tufte and Graves-Morris (1983) also defined a number of visualization design principles. In this section I will

apply some of these principles to the ‘Job Market Tracker’ visualization.

4.1 Data-Ink

Data-Ink is the ink used in a graph to actually communicate the data. According to Tufte and Graves-Morris (1983), good graphs maximize the so-called ‘Data-Ink Ratio’: the ink used in a graph to actually communicate the data, divided by the total ink used in the graph. Maximizing the Data-Ink Ratio means that you use the ink in a graph mostly to communicate the data.

The graphs/plots in the ‘Job Market Tracker’ visualization are all minimalist: the axes and legends show just enough information (i.e., ticks and labels) in order to understand the graphs/plots, and the focus really lies on the data (see also the last paragraph of the previous Section). The interactive table, for example, does not show horizontal and vertical lines that divide the table in rows and columns; this structure is implicit. This does, however, cause the focus to lie on the data. The creators of the ‘Job Market Tracker’ visualization succeeded in maximizing the Data-Ink Ratio.

4.2 Chartjunk

A concept that is related to Data-Ink is ‘Chartjunk’: the extraneous visual elements that distract from the message of a graph. An example of Chartjunk would have been the horizontal and vertical lines in the interactive table as discussed in the previous subsection.

Not surprisingly, lots of Chartjunk leads to a low Data-Ink Ratio, and no Chartjunk leads to a high Data-Ink Ratio. Although the creators of the ‘Job Market Tracker’ visualization succeeded in maximizing the Data-Ink Ratio, I believe that there are two types of Chartjunk in the visualization. First, in the interactive table, the sector size is represented by a bar with the exact size as a label (the bar is proportional to the sector size). Although this facilitates finding the largest or smallest sector, the bars do not necessarily contribute to the message of the table (and the accompanying interactive dot plot); only the labels would have been sufficient. Second, as described in Section 2, you can adjust the interactive heat map and interactive line graph by selecting an unemployment rate based on gender, age (in years), race/ethnicity, or education (25+). If you adjust both graphs/plots based on age by selecting the age group 16-24, for example, the interactive line graph shows a number of lines: the line for the overall unemployment rates (black, this line is always shown), the line for the unemployment rates under 16-24 year-olds (highlighted in blue), and the lines for the unemployment rates under the other age groups (light-grey). If you then hover over the interactive line graph, the tooltip not only shows the overall unemployment rate, but also the unemployment rate under 16-24 year-olds (obviously). Although the lines for the other age groups provide context (however, it is not clear which line belongs to which age group), they do not necessarily contribute to the message of the line graph (and the accompanying interactive heat map); only the line for the overall unemployment rates and

the line for the unemployment rates under 16-24 year-olds would have been sufficient².

4.3 Data Density

Tufte and Graves-Morris (1983) define ‘Data Density’ as the number of data items in a graph, divided by the area of the data in the graph. Tufte and Graves-Morris (1983) prefer high data density graphs. In the ‘Job Market Tracker’ visualization, the interactive dot plot and the interactive heat map are high data density graphs/plots: they show lots of data items in a relatively small area. The same goes for the interactive line graph.

5 Graphic design principles

Besides the visualization design principles of Tufte and Graves-Morris (1983), there are also a number of graphic design principles. Williams (2014) explains these graphic design principles, and I will apply some of them to the ‘Job Market Tracker’ visualization.

5.1 Contrast

Contrast is used to draw attention to important parts of a graph and to distinguish between (data) items that are different. There are a number of ways in which contrast is used in the ‘Job Market Tracker’ visualization. First, in the interactive dot plot, if you hover over a dot (i.e., a sector) in a specific column (i.e., a month), that same dot is highlighted in all the other columns. The dots are highlighted by making them black, and this provides contrast with respect to the other dots.

Second, in the interactive heat map, checking the ‘Show Recessions’ option will cause certain cells to be marked with a little black square (i.e., those cells represent months in which there was a recession), and this provides contrast with respect to the other cells.

Third, in the interactive line graph, the line that represents the unemployment rates for the category that was filtered on (e.g., age 16-24), is highlighted in blue. Furthermore, if you hover over the interactive line graph, the corresponding cell and month in the interactive heat map are highlighted (i.e., a black border appears around the cell and the month is made bold). These features also provide contrast.

²Although the visualization principles of Tufte and Graves-Morris (1983) are right, subjective dimensions should also be taken into account. As I said before, the bars in the ‘Sector size’ column of the interactive table do not necessarily contribute to the message of the table (and the accompanying interactive dot plot), but they do give *aesthetic* pleasure. Furthermore, the extra lines in the interactive line graph (e.g., the ones for the other age groups) are not necessary, but they do encourage exploration of those age groups (i.e., they are *playful*). The subjective dimensions sometimes justify certain design choices.

5.2 Repetition

The principle of repetition is to repeat some aspects of the design throughout the entire piece. In the ‘Job Market Tracker’ visualization, the color coding is repeated. Furthermore, the tooltip is also repeated throughout the visualization: the interactive dot plot, the interactive heat map, and the interactive line graph all show the same tooltip (in terms of design).

5.3 Alignment

Aligning items in a visual and readable arrangement is another important design principle; nothing should be placed on the page arbitrarily. The ‘Job Market Tracker’ visualization is clearly aligned: everything on the web page is center aligned, and the individual graphs/plots are vertically aligned. This alignment helps keeping the visualization balanced.

5.4 Proximity

The principle of proximity is to group related items together. In the ‘Job Market Tracker’ visualization, there are two groups of related items: the graphs/plots about the job gains and job losses in different sectors (i.e., the interactive dot plot and the interactive table), and the graphs/plots about the national unemployment rates (i.e., the interactive heat map and the interactive line graph). Related graphs/plots are positioned directly below each other (which implies a relationship), and the two groups are separated by white space and a small dividing line.

6 Conclusion

I have critically evaluated the ‘Job Market Tracker’ visualization by Andrew van Dam and Renee Lightner using a number of concepts and ideas that have been discussed during the course. I have used Tufte’s (1983) design principles of graphical integrity, Tufte’s (1983) visualization design principles, and William’s (2014) graphic design principles to evaluate the visualisation. I have shown that the visualization adheres to most of these principles, though some minor improvements should/could be made: the title of the interactive table should be updated based on the month selected in the interactive dot plot, and some chartjunk could be removed from the visualization. Nevertheless, it seems like the designers of the ‘Job Market Tracker’ visualization have put great thought into the design, thus providing a visualization that is both informative and engaging.

References

- Tufte, E. R. and Graves-Morris, P. (1983). *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT.
- Williams, R. (2014). *The non-designer's design book*. Pearson Education.