Thomas van der Veen, 10346481

This critique is about the rise and fall of unemployment in the US. It shows the data in multiple facets, using several visualizations. Each one is shows the variety in unemployment over time. In the visualizations, then, difference is shown between multiple demographical categories and per sector of production in the US labour market. This critique is meant to give the reader an insight in the different aspects that come to play in the making of a visualization, by analyzing this web page thoroughly. The way in which this analysis is structured and performed, gives an insight in what I have learned this course about the theoretical aspects of data visualization.

As the figures used on the web page are connected by, for instance colour coding, or even linked through the sharing of data, the analysis gives better insight when I show the pluriform theoretical aspects of data visualization and demonstrate in what manner they are applied to the example we see in front of us. Authors of theoretical studies, showing best and worst practices, function as framework through which the visualization is analyzed. They present us with clear, distinctive features, that can be either positive, or negative. I will focus on how the data is represented and whether that is in an honest way. But also on the story that the data is trying to tell us and whether that becomes clear and visible. Finally I shall elaborate on the more personal aspects of the analysis and eport my own opinion on aesthetic aspects, as well as language, subject and style.

The main goal of the figures is to show how the rise and fall of unemployment is distributed over different demographic and economic categories. He starts showing the economic aspects by visualizing the unemployment per sector. At first glance you just see a whole lot of dots in different colours on the screen. As I have come to experience in the working groups of the data processing course, when seeing a visualization, people tend to look at the visual stuff first. When they don’t understand things or understood the basics of the graph and would like to know what more it beholds, they search for extra explanation in text and legendas. Knowing that, it is very important to make the visualization intuitive. The first graph is not that intuitive, although the text and legendas explain the goal and way to interpret it, perfectly. When exploring the options the first graph has, by hovering over it, you’re all the more surprised seeing a lot of grey dots appear on the screen. They turn out to be connected data points. Every dot in the first graph is the growth (either positive or negative) per sector of the US’ economy in a certain month. When pointing with your cursor at a dot, the growth of that sector for the other months appear by marking the coloured dots. Thus, you can distinguish a path in the rise and fall of unemployment per sector. This is a task the visualization can perform. I will list the tasks it performs below. This way we can fully understand the visualization, before critiquing it further.

First visualization

1. Showing a path per sector, on how the sector gains or loses jobs.
2. Showing for the entire economy in which times jobs were created or lost.

Second visualization (linked to first)

1. Show absolute information on growth, sector size and consecutivity of growth.

Third visualization

1. Show the overall unemployment rate in the US over time.
2. Mark months in which there was a recession (decrease of production)
3. Show the unemployment rate for different demographic classifications. There can be selected on gender, age, ethnicity and education. When a category in a classification has been selected, the other categories are shown as well, but less distinctive.
4. Explore the impact of a recession on the unemployment of the demographic groups.

The fourth visualization has the same tasks as the third, as all their datapoints are shared. It shows the same data, but in a different way.

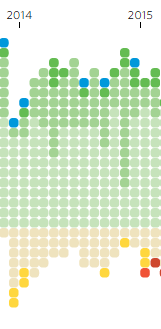
Now that the tasks and goals of the visualization have been stated, we can come to a deeper analysis. Tufte (1983) Uses six thumb rules to measure the integrity of a visualization. The main purpose is to make sure it represents the data as it is, without distorting it in any way. When that is done perfectly, the lie factor, as he calls it, is equal to 1. A higher lie factor means that the value of the data and the point it is trying to make, are overstated. In this case, The data is not distorted, but the way het visualizes data can give a wrong impression.

Figure , Selection of first visualization

Figure 1 shows a selection of the first visualization. Intuitively, the height of the graph shows the growth in that period of time. But in this case, it only shows how many sectors are above zero. It’s the colour value that shows the scale of the growth. It’s not possible to compare the overall growth in jobs per unit of time. In that sense, it can seem that the job demand is very high in a certain period, but every sector is barely above zero. Figure two shows the legenda, which shows a distortion: the extreme value below zero starts at minus 1% and represents everything below. The upper extreme value starts at 0.75% and represents everything higher. This is unbalanced.

Tufte would, however, be proud on the data ink ratio in the entirety of the page. Everything you see is represented data. There are no extra features that can draw your attention from the data; every dot or coloured rectangle represents a datapoint. Then, there are the blocks indicating recessions, but he rest is labels. There is no chart junk in any of the figures. In that sense, this visualization has good properties.

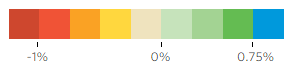
The difficulty in interpreting the different charts is in the variables. In the first graph, shades of green and yellow are used to indicate respectively growth and shrinkage. Their extremes are, then, colour coded as respectively blue and red. As the graph has a very strict distinction between the two sides, this works out pretty well. If the data would however be placed among each other - as is the case in figure three – this colour scheme would be hard to interpret. That’s the reason he used another colour set for coding the data in figure three. As the third visualization encodes different data than the first, it is a good thing that another coding scheme is used for this one. The C.R.A.P. acronyms teaches us that. But in this case the colour sets do not differ that much from each other. Both use red and green to indicate a good or a bad value. It is confusing to see two similar, but different, data sets, with two similar, but different colour sets. One shows growth in percentages using green and yellow with red and blue extremes, the other shows the percentage of people working, from witch growth can be deducted, using red and green. So, I cannot say that he did not abide the rules that Tufte gives, but when you look at the idea from which the rules stem, he does not abide to them.

Figure , legenda of first figure

So repetition is a thing that could be improved. Contrast is, in contrast, well executed as are alignment and proximity. The figures are well structured. This can be seen in the first graph where the main line is zero and the growth and shrinkage are respectively above and below that line.

In some sense, one could argue that the third visualization is one big piece of chart junk. The third and fourth figure are not only linked, they represent the same data in different ways. As Andrew Van Dam and Renee Lightner want to show two variables: time and the percentage of unemployment, it is far more conventional and clear, when they’d use a graph. They do so in figure four, so it seems to me they know the third figure is not clear enough, but they wanted to show it because of its aesthetics. The different colour codings are arbitrarily chosen and not comparable without coding scheme. The graph in figure four is perfectly clear.

Although they had better not used the third figure at all, a very good thing in the visualization as a whole is how they have managed to link the different figures. It is easy to understand how they are linked and it enables the user to zoom in on the data. Zooming in, and thus giving more detailed and specific information, without losing the overview and context is what I have come to consider the most valuable interactive feature. They start with overview and give the user the possibility to explore the data themselves.

They also highlighted the periods of time in which a recession took place. This explains some of the variation in the data set. In doing so the visualization manages to tell a story. A story that goes beyond the number of people, but one that shows how unevenly different demographic categories are affected by these recessions. Although the third graph is redundant, is can emphasize (and meanwhile distort) how black people are losing their jobs much faster than white people. In the first case, the figure shows a lot of red blocks, even without recession, and in the second there is only a little red in severe economic depressions.

I commented a lot on different aspects in this visualization. I do think their implementation would improve the quality of it and this paper is a critique so I should be critical. I do however enjoy this visualization very much. It tells a story, has well implemented interactive functionalities and yet is very easy to interpret. The data seems difficult at first, but is then shown in an interesting manner. As I am going to do a visualization project the coming month, the work of Andrew Van Dam and Renee Lightner give me something to reach and strive for.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.

Morris, M.E. S., & Hinrichs, R. J. (1996). Web Page Design: A Different Multimedia. Upper Saddle River. NJ: Prentice Hall.