This short essay tries to apply everything learned during the Data Visualization class on a visualizaiotn by the Wall Street Journal. There are multiple visualizations shown on the page, all of which cover the employment rate in the US in some form. The first visualization delivers a focus on the employment rate from 2007 to present day. The employment rate of twenty different sectors have been taken into account, ranging from Mining & Logging to Local Government. The first visualization can be seen as a stacked scatter plot, where every sector is placed based on its growth or shrinkage. The Y-scale shows the growth- or shrinkage percentage and the colour shows the difference in growth compared to the month before. Leaving your mouse on a certain point will result in all the dots of the sector associated with the dot your mouse is on to turn black and to be able to see the growth or shrinkage over time. This shows in-depth information about which sector showed what growth at a certain given time. The visualization can show which sector flourished at what moment and which sector suffered the most from the recession as well as which grew the most during the recession. The second visualization is a list with details about the twenty sectors at hand that can clarify the data from the first visualization. Clicking a month in the stacked scatter plot shows the data for that month per sector, making it easier to spot smaller changes and see some more numbers that can be important, such as the sector size and monthly change of jobs.

The second set of visualizations shows the unemployment rate from 1948 to present day for the entire economy without splitting it into sectors. A line graph is added to create a nice overview of the data. It shows when there is considered to be a recession and at what times this was reached. There is also a set of selectors that can change the data real-time, giving even more in depth information.

Problem domain, context and tasks summary

All the tools given in the visualization can give in-depth information about the unemployment rate in the USA over the years. This can show that “recession” is not something that is new to the world, but has been around through the 20th century. The stacked scatter plot shows information about the last eight years in twenty sectors and shows more detailed data about the growth of said sectors.

Tufte’s lie factor

The visualization itself is very clean, the data is presented to be clear and understandable. Tufte introduced multiple design principles which include the Data-Ink ratio, Chart Junk, Data Density and Layering. Tufte also introduced a concept known as the “Lie Factor”.

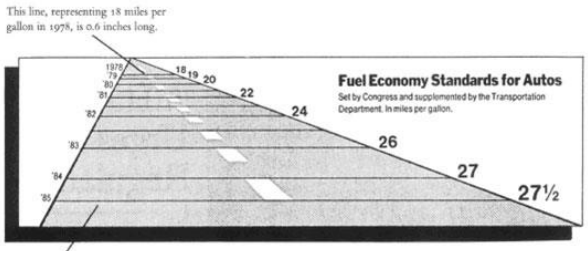

Figure 1 - An example of Tufte's Lie Factor - http://www.infovis-wiki.net/images/0/05/Lie_factor_example1_image.jpg

A famous example of this Lie Factor is shown in Figure 1 - An example of Tufte's Lie Factor The way in which data is shown can alter the readers perception of it, This Lie Factor is defined as the difference between the size of the effect in the graphic compared to the size of the effect in the data itself. In the example the data is said to have a Lie factor of 14.8, the numerical change of the data is 53%, while it is represented by a graphical change of 783%[i].

This Lie Factor is only partly present in the visualization by the Wall Street Journal and is mostly seen in the first set of visualizations in some form, be it small. The different dots that represent the sectors in the stacked scatter plot are stacked around the X-axis, where being under it represents shrinkage and being above it represents growth.

This does not mean being high up in the stack means large growth, it just means that there are a lot of sectors in between the selected sector and the
X-axis. This means that having dots higher up the stack might look like there is a bigger absolute growth, it only means that the growth is bigger relative to the different sector, fooling the reader as a result. Though is must be said that Tufte's Lie Factor is not completely fitting in this sense, since no absolute numbers are used on the Y-axis to show the growth.

Design Principles

Tufte's Design Principles can also be used to grade the quality of the visualization. The Data-Ink ratio defines the amount of ink used for data compared to the amount of ink used for the entire visualization. The ratio can show if too much emphasis is given to non-data related parts of the visualization. This visualization does not suffer from a small Data-Ink ratio since there is almost no clutter to be seen. All of the four visualizations are free of excess shapes, colours or other details.[ii]

This means that the amount of Chart Junk in this visualization is also kept to a minimum. Chart Junk is the second design principle by Tufte and it embodies everything in the chart that has nothing to do with the data itself of the clarification of said data. Things like added depth, colours without meaning and pictures or drawings that add to the "theme" of the graph. Usually, this Chart Junk only diminishes the readability of graphs, rendering it less useful than it would be without it.[i]

Since there is almost no Chart Junk to be seen, it makes it likely that the data density is high. This is also the case, since all of the visualizations are focussed around presenting their data, meaning that the amount of data per surface area of visualization.

Layering is only used in the line graph when selecting a certain gender, race or other options.

CRAP Framework

Tufte is not the only one that has created a framework for data visualisation, there are more. One of which is the CRAP model, this model handles Contrast, Repetition, Alignment and proximity to guide someone into making a high-quality visualization.

Contrast is used to actively show differences between data, this way you make it clear to the reader that the data is either different or the same. This will not only make things more clear to the reader, it can also add visual interest; contrasted things stand out more than bland ones. Adding contrast is not limited to the way you present your data, but also for colour and shape for instance. The Wall Street Journal uses colour to add contrast to their visualizations. The contrast catches the eye of the reader and makes it easy to spot a trend within the visualization. Not only that, the colour is also used to create cohesion between the different visualizations by the use of Repetition.

Repetition can be used to unify different parts of a design. It can even be used to control the readers eye across the page. In this case, colour is used for both contrast and repetition. The colours used in the overview per sector is the same as the colour scheme used in the stacked scatter plot and the cell-styled overview below.

In terms of alignment, the page as a whole makes a little less sense. The first three visualizations are centred around the middle of the page, making the fact that they are all a different size less of a deal. Alignment is all about making a page look smooth by aligning elements along "hard vertical edges". This is not the case on the Wall Street Journal page since it is centred.

There is both an upside and a downside to this. By not using a baseline when aligning elements a web designer can show, just like with contrast, that certain elements are different and should be looked at differently. On the other hand it makes the page look less fluent, it distorts the continuity.

Personally, I would have liked the visualizations to be of the same width to make the column they are in less jagged. Especially the last two visualizations are not aligned right with their title like the first two visualizations are, if they were stretched slightly it would have looked cleaner and stronger in my opinion. Especially since the proximity is applied nicely.

Grouping related items on a page together to show relation is what is called Proximity. Looking at the page show two groups of two, the first to being the stacked scatter plot and the list, the second being the cell-styled overview and the line graph. When looking at the data, this can only make sense; this first two visualizations are centred around different data then the second set. This is a result of the small line in between the second and the third visualization, splitting the page in half.[iii]

Subjective Dimensions

All in all the visualization as a whole is very pragmatic, everything is clear and easy to work with, conveying the message was without a doubt the top priority when design this page. A downside of this practicality is its result on the aesthetics and playfulness of the visualizations. It is only logical that practicality and pragmatism do not line up with playfulness. By adding more interactivity and gimmicks the playfulness be more apparent, but it will also blur the message, leaving the designer with a trade-off. Since the visualization is designed by a renowned journal, it only makes sense that pragmatism was chosen over playfulness, conveying the message is more important than showing the viewer a good time. Keeping the visualization as clean as possible does add to the aesthetic and style of the visualization, it is simple and scientific. This makes it look more believable than when more bells and whistles would be added.

Visual Encodings

Multiple visual encodings are used, starting with position and colour which are used to show the absolute growth or shrinkage and position to show growth relative to the other points. The points are selective, showing all the other points of the sector when hovering over a point with the mouse.[iv]

Conclusion

All in all the visualization does a great job at conveying its message, showing the growth or shrinkage of employment rates in the US from January 1948 to present day in general and in detail, and does so in a very effective and scientific way. The style is calm and scientific, leaving a lot of room for the data to be examined. The strangely shown rainbow coloured scale is the only real downside, not making sense at when first looking at it. The .75% should be deleted and swapped with a 1% one square to the right. Being able to click a dot in the first visualization in order to keep the selected sectot black would also have helped in giving insight in the data. When a viewer moves its mouse, the highlight disappears. Also, the rainbow colour scheme can create problems for colour-blind viewers, making the visualizations unnecessarily complicated. Personally I would try to make the titles and visualizations line up and try to make the selection screen combinable; only one of the options under one section can be selected, being able to combine options would be nice.

---

[i] Tufte's Principles. (2009, August 05). Retreived from http://thedoublethink.com/2009/08/tufte%E2%80%99s-principles-for-visualizing-quantitative-information/

[ii] Torsten Möller, 2015. Design Principles [PDF]. Retreived from http://www.cs.sfu.ca/~torsten/Teaching/Cmpt467/LectureNotes/02_design.pdf

[iii] Using Graphic Design Principles in Web Design. Retreived from (http://www.colorado.edu/AmStudies/lewis/Design/graprin.htm#summary

[iv] Carpendale M.S.T. (unknown). Considering Visual Variables as a Basis for Information Visualization. https://cdn.mprog.nl/dataviz/excerpts/w2/Carpendale_Considering_Visual_Variables.pdf