

DATA VISUALIZATION

1st Assignment

VOULGARI ELENi

A.M. 17005



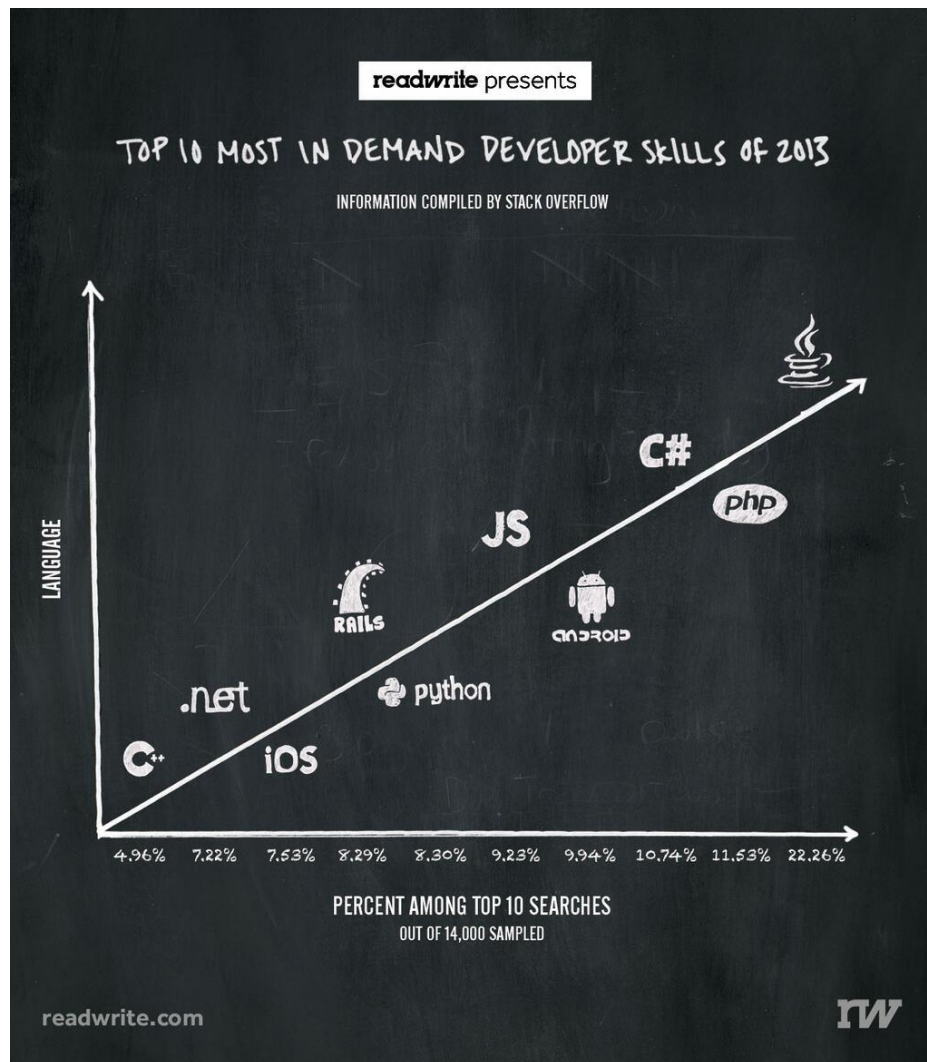
Introduction

For a visualization to be successful, it should have a clear purpose, include only the relevant context, use appropriate structure and has useful formatting, so as to be able to help the user extract useful information.

The aim of this assignment is to find examples of bad visualizations on the internet. For each one of these examples, there will be comments on the problems it has and an explanation of how it could be improved.

Examples of Bad Visualization and Improvement Suggestions

1. http://cs.colby.edu/courses/S14/cs251/LectureNotes/figures/developer_skills.jpg



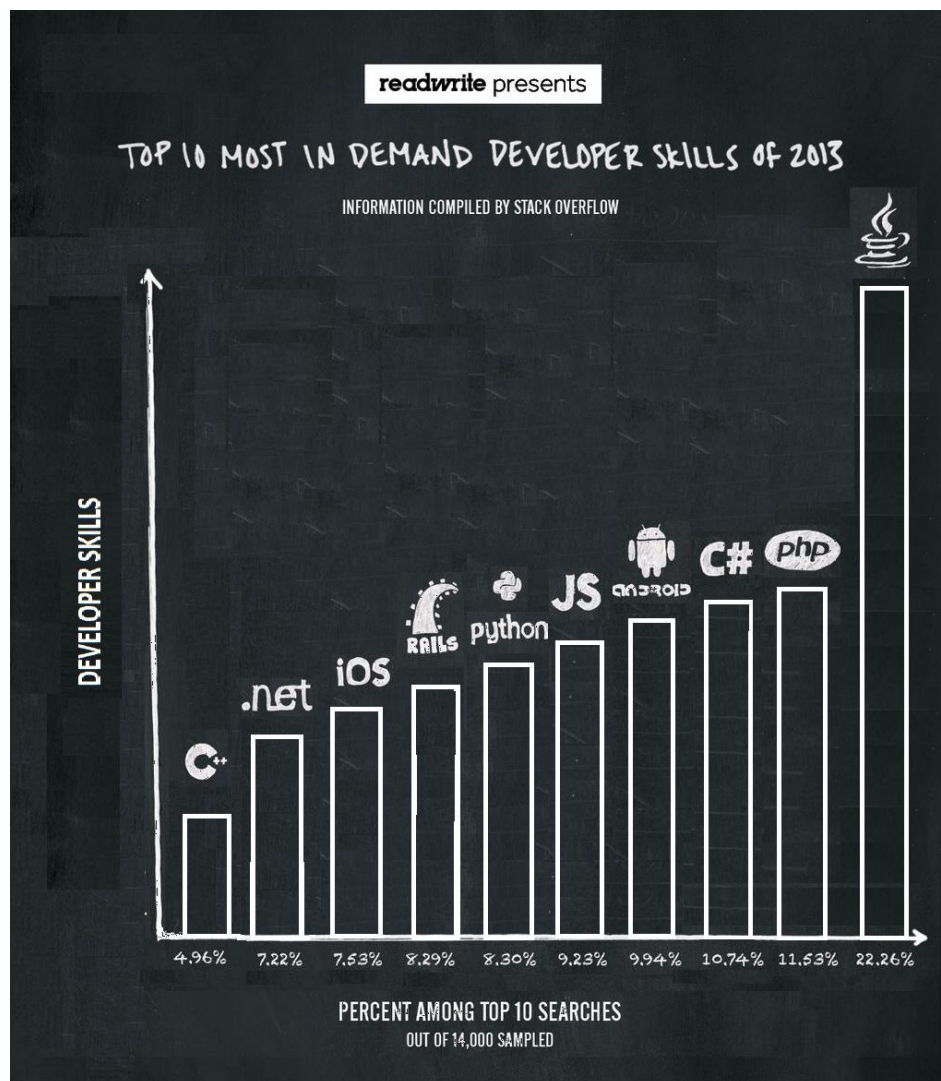
The above visualization shows the linear relationship between the percentage among top 10 searches and programming languages, where its aim is to show ‘The 10 Most in Demand Developer Skills of 2013’.

The logos mentioned in chart are not all languages, iOS and android are operating systems and .NET and Rails are frameworks. The Y-axis should have the label ‘Developer Skills’ instead of ‘Languages’.

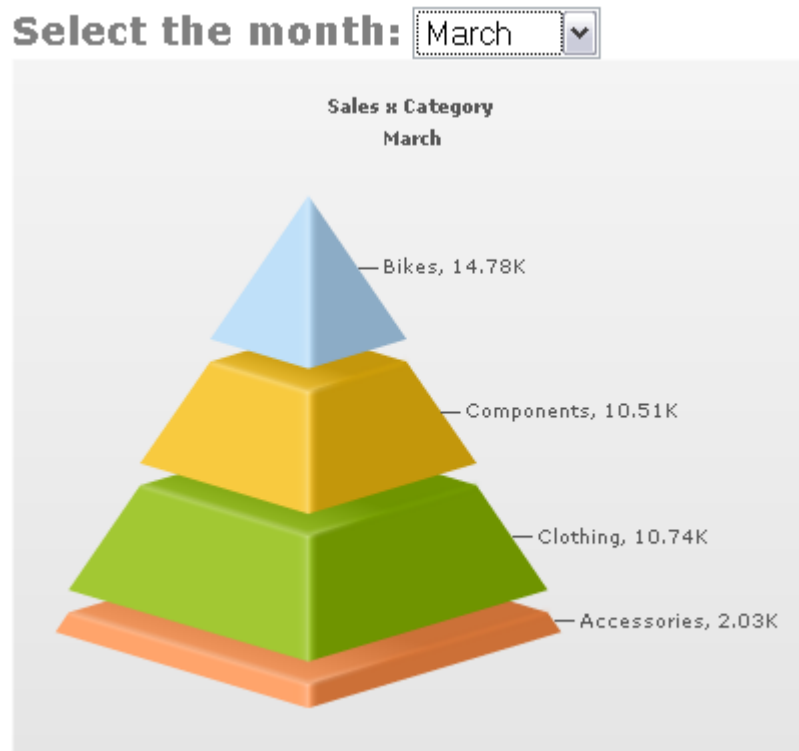
The Y-axis implies that one language is more of a language than another which is not something somebody could say, even though some of them are not even languages.

The difference between percent among the languages cannot be calculated, just by looking at the graph and the symbols that are arranged one below the other against the linear graph is pretty confusing.

The best way to represent the graph is by using a Bar Chart, like the one below:



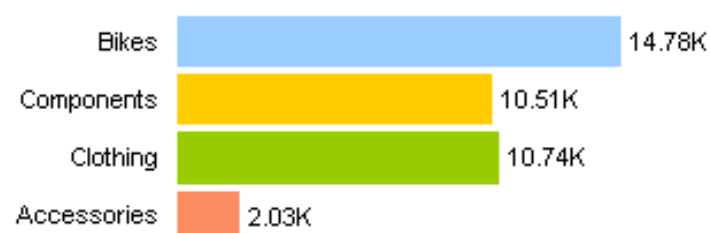
2. <https://peltiertech.com/bad-graphics-funnel-chart/>



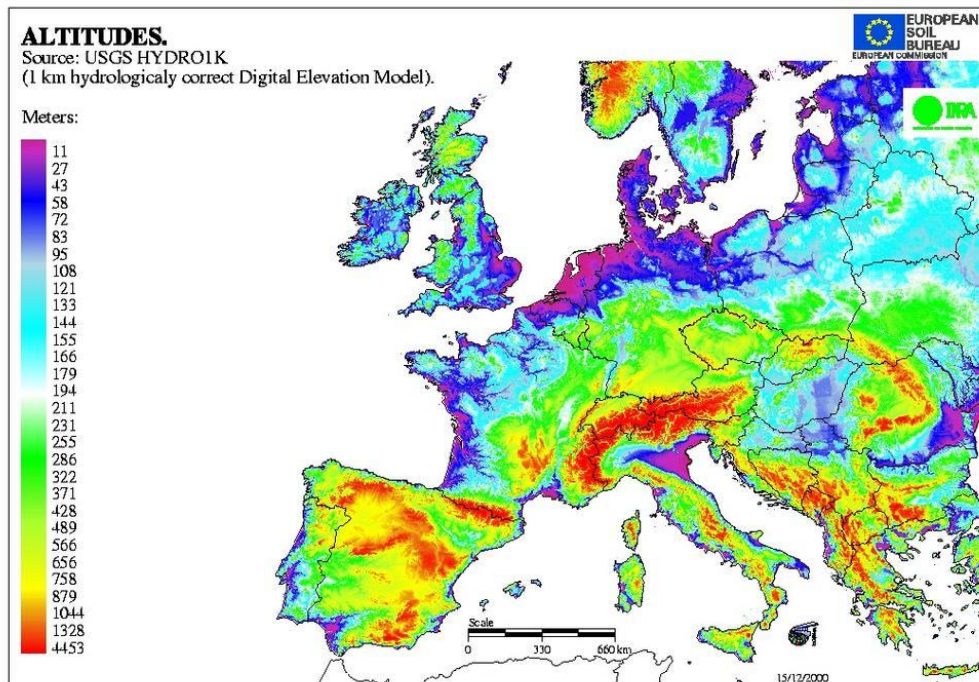
The above 3D pyramid chart shows the sales of a company, but it doesn't seem to be very efficient. The component of the base 'Accessories' is the smallest of all, which doesn't make sense for a pyramid, as the base should be the largest part to support the whole structure and thus the company.

Furthermore, the proportion of the other three components is not easy to understand as their levels seem to be of the same height. Last but not least, the major sales (bikes) seem to be the least, as the eye perceives the area it covers as smaller than the others.

A much better way to represent this kind of data is with a Bar Chart, like the one below:

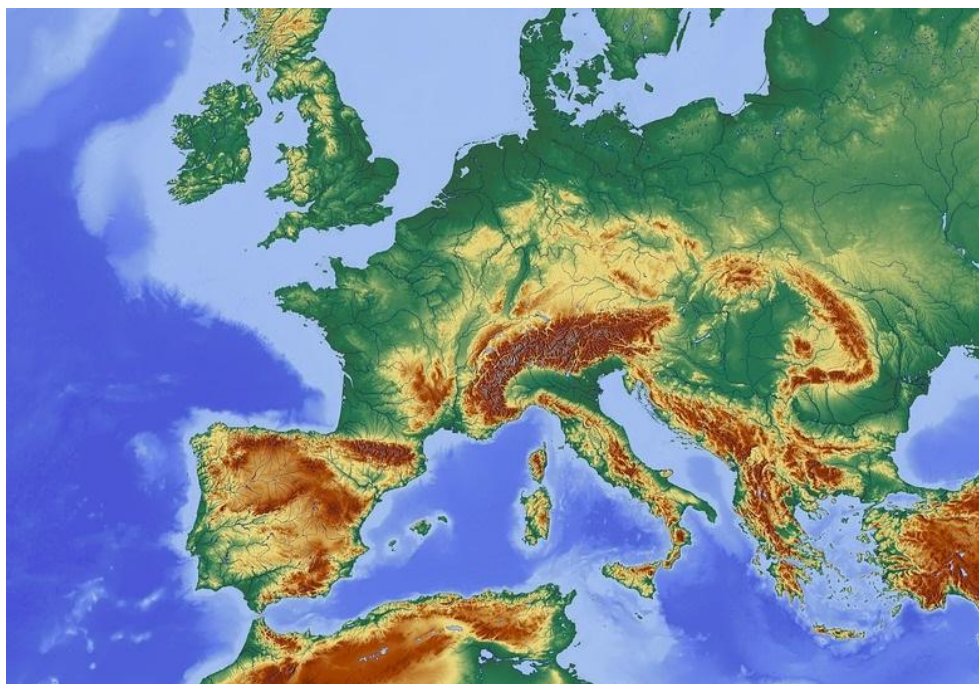


3. <https://www.safaribooksonline.com/library/view/designing-data-visualizations/9781449314774/ch04.html>

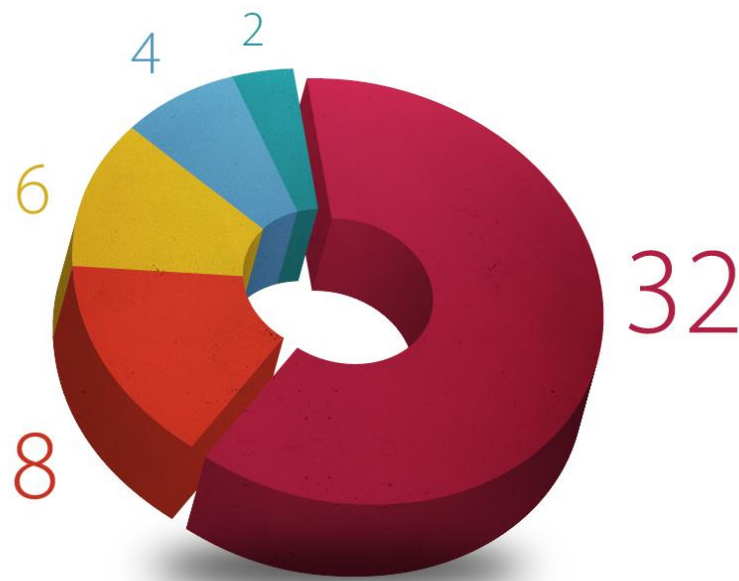


The above plot shows the altitudes in Europe, by using different colors on the map. The problem is that the colors that are used are not representative of what wants to be shown. These colors are mostly used to represent temperature and someone might believe that it is a map that shows how warm and cold are the areas of Europe, which would be totally wrong.

The best practice would be to use different shades of green and brown to express altitude and show where the mountains and the valleys are, like the visualization shown below:



4. <http://f2017cpsc683.ucalgaryblogs.ca/2017/09/14/3d-pie-charts/>

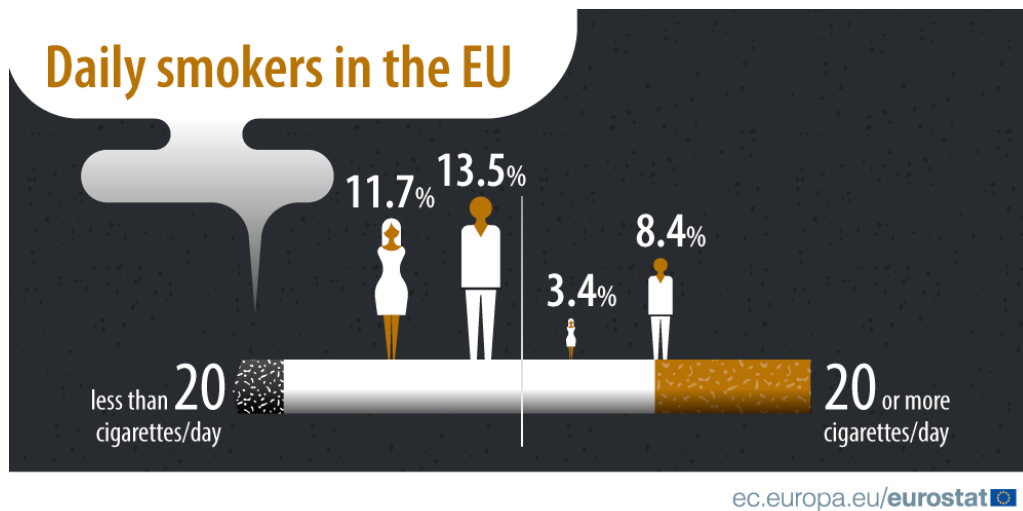


The chart above is a 3D pie chart, that we don't have knowledge of what it represents. Its problem is that with this kind of a diagram someone would expect the summation of the numbers to equal to 100, something that this visualization doesn't follow. Furthermore, the fact that it is 3D only adds to the complexity of the visualization and doesn't provide with any extra information.

We don't know what this visualization is about, but a better graph would be a Bubble Chart, like the one below:



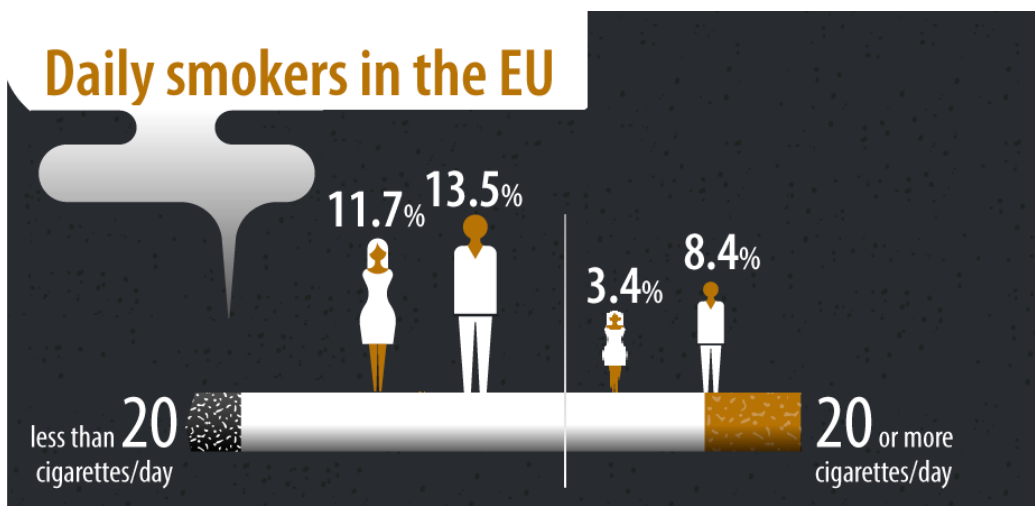
5. <https://kaplanviz.wordpress.com/2017/02/09/bad-viz-example/>



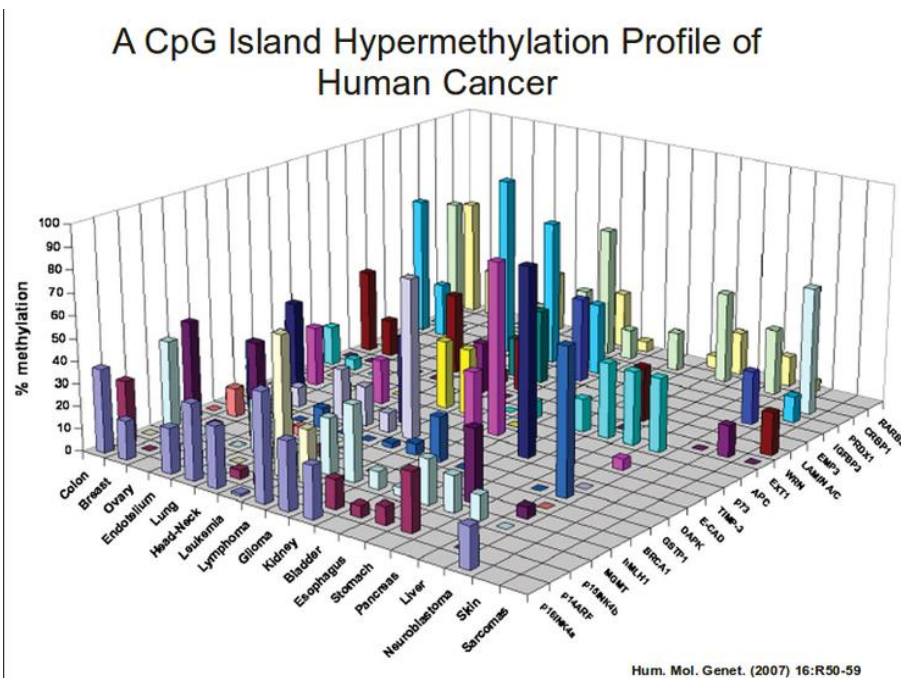
The first problem with the infographic above, seems to be that it compares two groups (less than 20 cigarettes and more than 20 cigarettes) but it looks more like a timeline to the human eye. The cigarette seems to be split in the middle but the foot of it and the burnt tobacco part don't participate in the visualization.

Furthermore, the proportion of the human that represents the 3.4% is too small in relation to the others and the visualization doesn't explain what these percentages represent, even though we can understand that it refers to the population of men and women in Europe.

The visualization would be somehow better with some changes as the graph shown below:

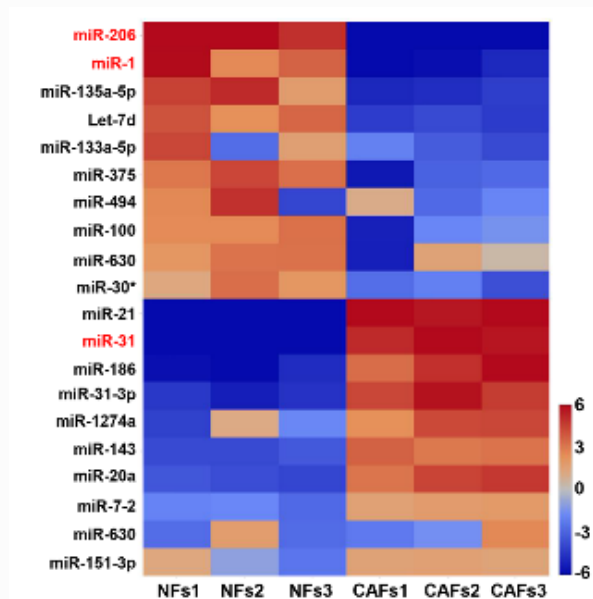


6. <https://www.livestories.com/blog/five-ways-to-fail-data-visualization>

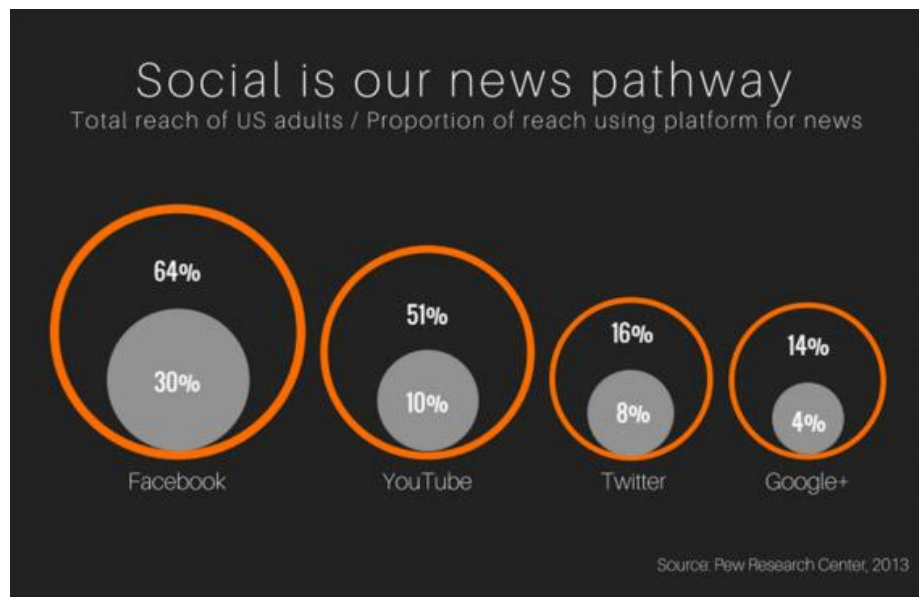


The 3D chart above is showing how patients with different cancers have different amounts chemical additions to their DNA. This visualization lacks simplicity as it is not possible for the human eye to understand which bar goes to which percentage and some bars are overlapping others.

The third dimension here is not helping in any way interpreting the data the right way. A much better way to represent the information would be a 2D heat map showing the percentage of the methylation with color gradation, as the one shown below (it is an example chart, not representing the same data):

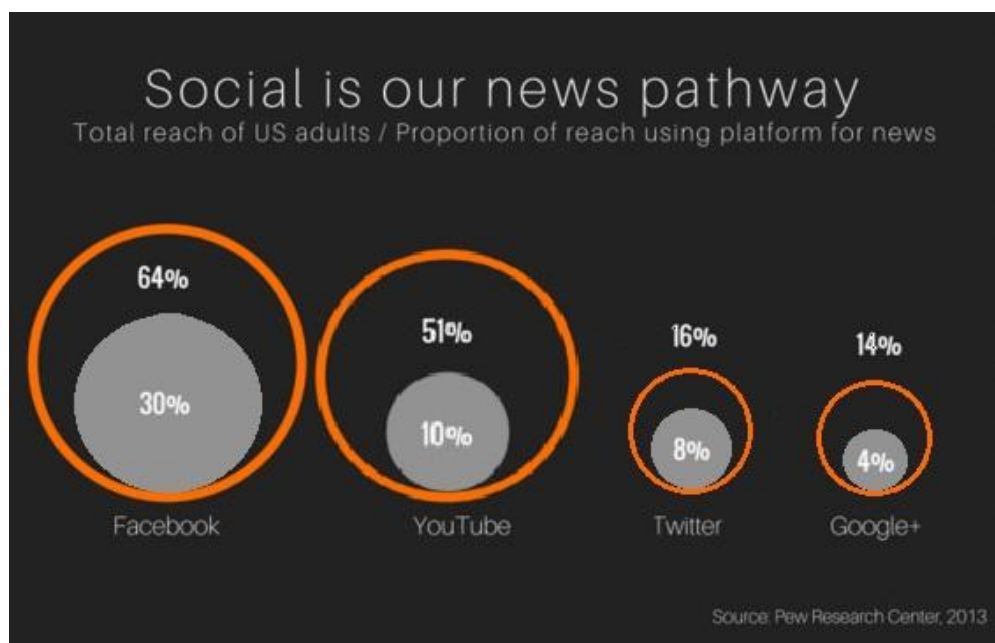


7. https://www.huffingtonpost.com/zach-kitschke/post_9445_b_7456664.html



The problem with this Bubble Chart is that the proportions of the bubbles are not right, meaning that, for example, the bubble which represents the 30% is almost twice as big as the bubble representing the 8% whereas it should have been bigger. The same applies to the other proportions of the visualization.

The chart with the rights proportions would be the following:



Conclusion

Data visualization is the process of converting data into easily understood graphs that enable effective decisions or conclusions. Someone should not miss a significant correlation or draw an incorrect conclusion that could affect these decisions or conclusions. So, the choice of the type of visualization is very important and all things must be considered and most of all the data itself, before picking one.