Design 2

Group:
Houda Alberts, 10740287
Jasper Bakker, 10260250
Jelle de Boer, 10540075
Milou Bisseling, 10427538
David Lips, 6225950

Problem 1 Design Critique

Who is the intended audience?
The (young) adults who watch movies.

What tasks does the visualization enable?
It enables seeing the weekly box office revenue and the total domestic gross in dollars from movies between 1986 and 2008. The viewer can find out which movies are successful and when they are successful. You can also find out through the graph how long the movie lasted in the theatres. Another task that the visualization enables is the presentation of trends. For example, box office revenues tend to get bigger in holiday periods. The caption also mentions the task of the visualization to show that Oscar movies tend to attract smaller audiences that build over time, but it's not clear how to answer that question with the graph.

What data is represented in this visualization? Be specific.
The data that is represented in this visualisation is presented in areas, each area (and its color) represent a film's total domestic gross through in dollars adjusted for inflation between 1986 and 2008.

How is each data type visually encoded?
The legend shows the different colors which correspond to the film's total domestic gross in dollar adjusted for inflation. The height of the area shows the weekly box office revenue and the width of the area shows the longevity of the weekly box office revenue in dollars.

The data for each movie has its own area that is clickable for more information. This includes an overview of the movie. The areas show the title while hovering over these areas, they do not show the total box revenue. There's an x-axis with the months of the year starting at 1986 and ending at 2008. There's no y-axis in the graph, the y-axis is meaningless. The colors of the color map look very similar, they do not distinguish very well. The scale of the legend is arbitrary (0, 25, 100, 250, 862 million).

How do the visual elements and user interactions support the tasks?
By hovering over areas, you can select those, which makes you able to see certain areas better. Another addition is that details of the movies can only be seen when they are selected. The visual and user interactions support some of the tasks named before. The visual elements do show trends and how long the movie lasted in the theatres. But the visual elements don't show the answer to the question if Oscar movies tend to attract smaller audiences that build over time.

Why do you like / dislike this visualization?
We like the color scheme of the graph. It is also very unique and compact. The selecting makes the compactness less difficult. All information can also be found within the graph. What we dislike about the visualization is that the colors of the areas cover a high range. This does not say very much on detail and there is no explanation why this is chosen. So, the scale is chosen quite bad. The lack of a y-axis makes it more difficult to extract quantitative information. You can also not see which movies are Oscar movies. There is no origin in the graph, which can be solved by two positive y-axes and using layers. This is not done in the graph.

What is the problem domain or context of the visualization under consideration?
Movie box office revenue, and how movies - from 1986 to february 2004 - have generated box office revenue over the time of their appearance both weekly and longevity.

Which tasks can be achieved with this visualization?
Seeing that the Oscars attract smaller audience over a longer period of time. Another task is seeing the summaries of the shown movies. Their weekly and longevity box office revenue are also mentioned in the graph and can be read from it.

Tufte's principles of graphical integrity:

– Are the scales appropriately labeled?
Timewise the scales are clear, showing months and years. In terms of income, however, the only scale available is a depiction of how four different colors indicate a range of box office revenue from 1 to 862 million $$, supposedly in combination with the area of each movie's doodle-wave figure, while the height specifically translates to weekly income. Needless to say, this is not very clear.

– Is the Lie factor high?
The lie factor - as defined as the ratio of data variation and design variation - is hard to determine in this graphic. The movie's total income is depicted by the total area of its surface, which cannot be read accurately and are not described further upon clicking the movie's area. It is therefore hard to say whether the differences between each movie's exact gross income is exactly represented by the difference in the surface areas. Nevertheless, we suspect the lie factor is larger than 1 (i.e. the design variation is not

proportional to the data variation), as many of the area sizes seem arbitrarily chosen and to be less than proportional given the scale.

– Does the visualization show data variation and not design variation?
Given the answer above, probably both. And given the random shapes, however, it's probably mostly design variation too.

Tufte's visualization design principles, are they adhered to?

– Maximize the data-ink ratio.
Given the density of movies stacked upon one another, it seems the data-ink ratio is good. That said, one or multiple bar charts might be able to convey the same information with less space/ink, although this would surely obliterate the playful and flowing nature of the stream-graph as it is now.

– Avoid chart junk.
Certainly, junk is minimal.

– Increase data density.
Yep, movies are stacked upon each other densely.

– Layer information.
Somewhat; the visualisation tries to show total gross income (area), weekly income (height), and cinema runtime (total width) in one graph, which is certainly a layered approach. Moreover, the interactivity enables you to obtain more information about a movie by clicking on it.

Graphic design principles:

– How is contrast used?
The different items are shown in different colors, movies with a total domestic gross between 1 and 25 million are shown in light yellow. Movies with a total domestic gross between 25 and 100 million are shown in dark yellow and so on. The contrast between the two sorts of yellow are very small, which makes it difficult to see the difference. It's also difficult to see the the difference between the white borders and the light yellow areas.

– How is repetition used?
The aspects of the design are used throughout the entire peace, each shape represents how one film did at the box office.

– How is alignment used?
Every movie has a visual connection with someone else, but it's difficult to compare the movies because the shapes are all different. It's also impossible to compare different movies from different years.

– How is proximity used?

Physical closeness implies a relationship between data. In this visualisation, the movies are shown on an interactive time scale but the group related items are not shown together. In the caption of the visualisation they speak about holiday hits and oscar movies, these different groups are not shown together.

Comment on the visual encodings that are used.

– Which visual encodings are used?

The visual encoding is the way in which data is mapped into visual structures, upon which we build the images on a screen. The visualization uses planar and retinal variables. The planar variables used are the months on a timescale on the x-axis. The retinal variables are the size and the color of the areas (movies). The size and the color of the areas are used to represent the total domestic gross in dollars and the shape is used to represent how the movie did at the box office.

– Are the visual encodings appropriate?

We think the use of the planar variables is partly inappropriate because the y-axis is now meaningless, and could have been used to add valuable encoding in terms of gross income, for example. Another reason why the visual encodings are inappropriate is that retinal variables as shapes, size and color are better used with qualitative data instead of quantitative data, as they are hard to compare and difficult to translate to specific numbers.

Comment on subjective dimensions such as aesthetics, style, playfulness and vividness.

The colors of the visualization look nice together, it's also interactive with playfulness and vividness as a result. The way the movies are organised in a 'stream-graph' shows the ebb and flow of box office revenue very well, it draws the reader in, and facilitates playful exploration of the data as you scroll back and forth through time.

What is the intended goal of the visualization and is that goal achieved?

The intended goal of the visualization was to show box office revenue in dollars adjusted by inflation per movie from 1986 and 2008. The second goal was to show the trend that box office revenues tend to be higher in holidays and the last goal was to show that Oscar movies tend to attract smaller audiences that build over time.

The first goal are partially achieved, since the viewer can see the longevity of the movie on the x-axis but it's difficult for the viewer to see how big the weekly box office revenue is and how big the film's total domestic gross out of the areas. The second goal is achieved because the viewer can see the trend that the total box office revenues are higher in holiday periods. The last goal is not achieved, because the visualization does not show anything about Oscar movies. If the view does not show which movies are Oscar movies, the goal is not achieved.

Are there any things you would do differently, and why?

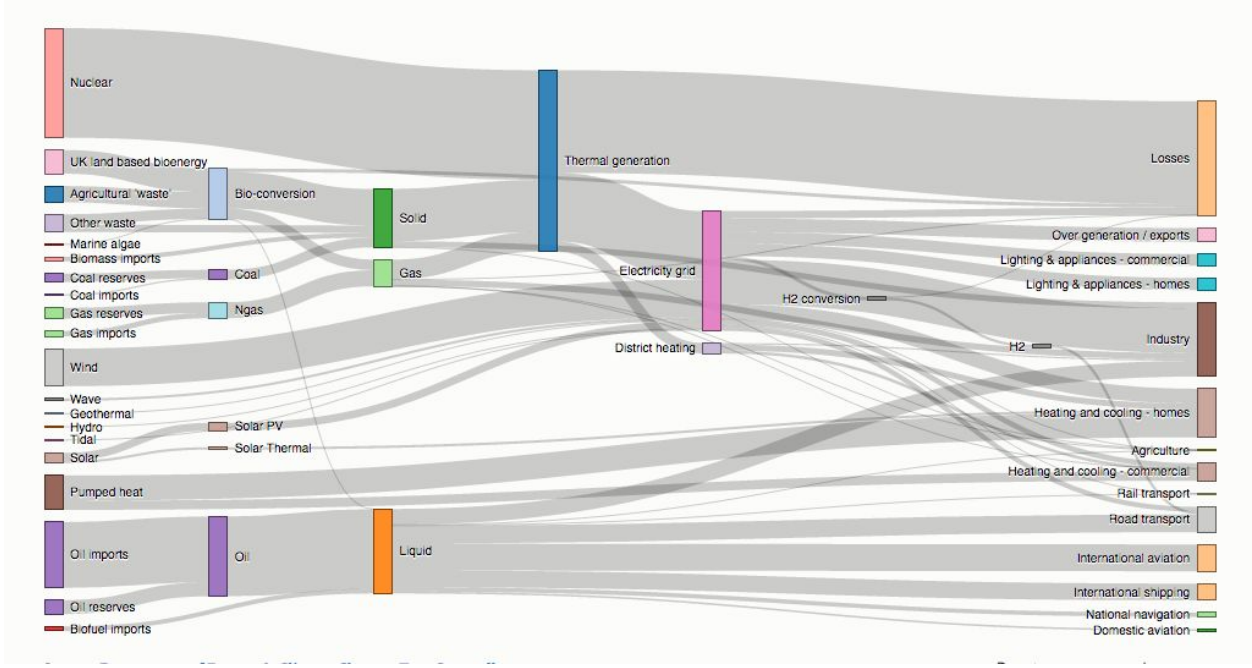
We would assign a numerical value to the y-axis to fully achieve the first goal of showing the total box revenue and the weekly box revenue. We would have also included more information in the interactive element that pops up when you click on a movie, like total box office revenue and oscar awards or nominations. Lastly, we would change the scales to be less arbitrary and depict more coherent income categories, perhaps using more colours (within the same scheme) to allow for more income categories and avoid lumping 250M-862M movies together.

Problem 2 Questions corresponding to the readings

Find a visualization not discussed in class or used in a homework and answer the following questions pertaining to that visualization. Attach the visualization as a screenshot in your submission.

https://bost.ocks.org/mike/sankey/

Sankey diagrams visualize the magnitude of flow between nodes in a network. This intricate diagram shows a possible scenario for UK energy production and consumption in 2050: energy **supplies** are on the left, and **demands** are on the right. Intermediate nodes group related forms of production and show how energy is converted and transmitted before it is consumed (or lost!). The thickness of each link encodes the amount of flow from source to target.



- **Consider Bertin's characterization of visual variables (position, size, shape, value, color, orientation, and texture). Pick 2 of Bertin's visual variables, and discuss them in relation to your visualization.**

a. Lines: The lines visualizes the flow from the origin of different forms of energy and it's transformation, which will end - on the right side - in the final use of the energy.
b. Size: The thickness of the lines are analogue to the amount of a specific form of energy. As you can see for example, relatively most of the demander forms of energy are derived from Thermal Generation.

- **Ask yourself what the designer is trying to convey and think of three to four possible tasks this visualization should help you with. Does the visualization achieve any of your tasks? (To view an example, see Albert Cairo, pages 26-28.)**

a. Our relative dependencies on different energy forms and it's use in the end. The thickness help's a lot. According to Steven Power's law (1961) we estimate length very well. The thickness of the bars can be seen as the length, since the width ('x axis') only indicates a direction, the total area of the bar doesn't indicate anything.
b. Furthermore the colors help to distinguish the different types of energy easily.
c. If you hover over a colored rectangle it shows you the corresponding tWh of either the energy source's production or the tWh used/lost. This could have been done better by letting it pop up quicker and more noticeable.

Following Albert Cairo's line of thought the following can be said about the graph:

a. The visualization does not show many different variables, but it does do a good job conveying the message it wants to get across.
b. It allows for comparison at a glance relatively easily due to the thickness of the bars
c. It helps organize and see relations between the different forms of energy (use)
d. It doesn't convey correlations very well, for instance: which form of energy is responsible for the size of the energy loss chunk?