Data Processing - Readings 2

Vincent Erich
10384081

November 3, 2015

Questions

1. For this assignment, I will review the 'U.S. age pyramid' visualization from 'The Next America'[1]. Figure 1 shows a screenshot of the visualization.

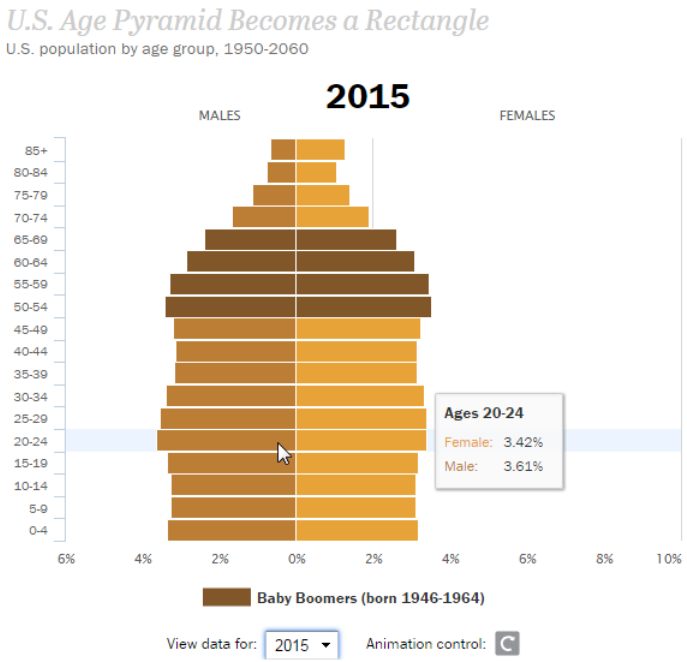

Figure 1: A screenshot of the 'U.S. age pyramid' visualization from the 'The Next America'.

[1]The visualization can be found here: `http://www.pewresearch.org/next-america/#Two-Dramas-in-Slow-Motion`

2. Bertin defines a number of visual variables that can be used to create varying marks: position, size, shape, value, color, orientation, and texture. Two of these variables particularly apply to the 'U.S. age pyramid' visualization: position and value.

   Of all the visual variables, position is the most versatile and most powerful one, and it possesses a number of characteristics that prove useful in the 'U.S. age pyramid' visualization[2]. First, position possesses the quantitative characteristic, meaning that the relationship between two marks differing in position can be seen as numerical. In the 'U.S. age pyramid' visualization, the position of a bar indicates a numerical value. The horizontal position of a bar represents the percentage of U.S. population in the corresponding age group. The wider a bar, the higher the percentage of U.S. population in the corresponding age group. The age group itself is represented by the vertical position of a bar. The higher the vertical position of a bar, the higher the age group it represents.

   Second, position possesses the order characteristic, meaning that changes in position support ordered readings. That is, a change in position will automatically be read as more or less. In the 'U.S. age pyramid' visualization, this holds true for left-right interpretations (i.e., the percentage of U.S. population in the corresponding age group) as well as for up-down interpretations (i.e., the age groups themselves).

   Third, position also possesses the selective characteristic, meaning that a mark changed in position makes it easier to select that changed mark from all the other marks. Although the distinction between the data for men and women in the 'U.S. age pyramid' visualization is primarily clarified using the visual variable value (see the next paragraph), the horizontal start and end position of a bar also aid in this distinction: the start position of a bar (the left end) represents the data for men (up to the vertical zero line), and the end position of a bar (the right end) represents the data for women (from the vertical zero line).

   Value is the relative lightness or darkness of a mark. Value is another visual variable that possesses a number of characteristics that prove useful in the 'U.S. age pyramid' visualization[3]. First, like position, value possesses the selective characteristic, meaning that marks that change in value can be easily distinguished and interpreted as different. In the 'U.S. age pyramid' visualization, two different values of the color orange are used to distinguish between the data for men and women. However, the two values of the color orange still allow the opposing bars to be read as a whole.

   Second, value also possesses the associative characteristic, meaning that

[2] Not all the characteristics of the visual variable position are discussed here.
[3] Not all the characteristics of the visual variable value are discussed here.

marks that are like in other ways can be grouped according to a change in value. In the 'U.S. age pyramid' visualization, this characteristic (obviously) groups the data for men and women respectively.

3. Munzner (2009) proposed a nested model for visualization design and validation. This nested model splits visualization design and validation into four levels. The first two levels of this model are: characterize the tasks and data in the vocabulary of the problem domain, and abstract into operations and data types.

   At the first level, the visualization designer must gain an understanding of the tasks and data of target users in some particular target domain, and must come up with a detailed description/characterization of the problems (and data) within this domain. The output of this level can be validated downstream by reporting the rate at which the visualization has been adopted by the target audience.

   I believe that the designer(s) of the 'U.S. age pyramid' visualization had a good understanding of the tasks and data of target users in the particular target domain (i.e., demographic transformations). I find that the visualization can be easily adopted, showing that the designer(s) successfully characterized the tasks (problems) and data in the vocabulary of the problem domain.

   At the second level, the visualization designer must map problems and data from the vocabulary of the particular target domain (i.e., the output of the first level) into a more abstract and generic description of (generic) operations and data types. The data types should by such that they can be addressed by visualization techniques (i.e., a table of numbers; a node-link graph or tree) and such that a visual representation of it will address the problem. The output of this level can be validated downstream by having a member of the target user community try the tool, in hopes of collecting anecdotal evidence (i.e., hypotheses confirmed) that the tool is in fact useful.

   I believe that the designer(s) of the 'U.S. age pyramid' visualization has/have successfully managed to map the problems and data from the vocabulary of the specific problem domain into a more abstract and generic description of (generic) operations and data types. The visualization confirms that in the future, the 'age pyramid' of the U.S. population will turn into a rectangle.

4. Based on the results of Cleveland and McGill (1984), I think that the 'U.S. age pyramid' visualization embodies good practices (i.e. people can accurately perform the tasks based on the encoding). First, people can accurately decode position along a common scale. In the 'U.S. age pyramid visualization', both the U.S. population percentage and the age group are

along a common scale (though the population percentage is not along one common scale; it is along a common scale in two directions).

Furthermore, people can accurately decode length, and in the 'U.S. age pyramid' visualization, the (horizontal) length of a bar represents the percentage of U.S. population in the corresponding age group.

5. I agree that visualization is a functional art. The main goal of a visualization is to be understandable and to provide a function. A visualization should be beautiful after that, or, as Alberto Cairo describes it, '*be beautiful thanks to its exquisite functionality*' (Cairo, 2012, p. xx). A visualization that is not understandable and that does not provide a function, but is appealing to the eye (i.e., a visualization that is merely art), has no added value. However, a visualization that is understandable and that does provide a function, but is not appealing to the eye (i.e., a visualization that is merely functional), probably fails to tell a story. A good visualization should be both functional and appealing to the eye; a good visualization should be functional art.

6. The designer(s) of the 'U.S. age pyramid' visualization is/are trying to convey that in the future, the 'age pyramid' of the U.S. population will turn into a rectangle, i.e., that there will be as many U.S. citizens over age 85 as under age 5. There are a number of tasks the visualization should help the reader with:

   (a) The visualization should **present** several variables: the *percentage* of U.S. population per *age group* for a number of age groups over a large period of *time*.

   (b) The visualization should allow **comparisons**. As an example, the visualization should allow the reader to compare the percentage of U.S. population in age group 0-4 in 2015 with the percentage of U.S. population in the same age group in 2030.

   (c) The visualization should **(actively) show** how the 'age pyramid' of the U.S. population will turn into a rectangle over a large period of time.

   The visualization successfully achieves the tasks described above.

References

Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.

Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554.

Munzner, T. (2009). A nested model for visualization design and validation. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):921–928.