**Questions**

**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.**

<http://xkcd.com/1127/large/>

1. 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.

**Color**

The xkcd visualization, in terms of color, without taking into account value, is limited to red (Republican Party) and blue (Democratic Party).

**Selective:** Color would be considered selective in this visualization as it is easy to determine the different parties based on their color.

**Associative:** The color in the Congress visualization is associative since it is possible to group the Republican and Democratic parties based on their color. This is possible even with the two visualizations of the House and the Senate.

**Quantitative:** There are no numerical values associated with the colors in this visualization.

**Order:** The color is not ordered in this visualization – red does not come before blue and vice versa. The rainbow scale discussed in the reading also does not apply to this visualization.

**Length:** Bertin discusses limiting the number of color values to six or seven to allow the colors to retain their selectiveness even though it is theoretically possible to have an infinite length of colors. The xkcd visualization is limited to only two colors.

**Value**

The xkcd visualization of the United States Congress uses color values to show the ideological makeup of the Congress at a given moment in time. The values range from dark, to medium, to light blue for the Democratic Party representing Far Left, Left and Center Left. On the Republican side the color values range from dark, to medium to light red representing the Far Right, Right and Center Right.

**Selective:** The values in this visualization would be considered selective as the colors are ‘selectable’ based on their value. It is easy to distinguish between the different values of blue and red and gain meaning from that change (i.e. a partisan shift right or left).

**Associative:** Again this visualization would be considered associative in terms of value. The different groupings of ideology can be seen by darkness or lightness of either the blue or red.

**Quantitative:** No numerical information can be gained from the different color values in this visualization.

**Order:** The values in this visualization would be considered ordered – going from the most liberal to the center to the most right ideology.

**Length:** Bertin says that the value length is theoretically infinite, but practically limited. In this visualization there are only a total of six different values; three for blue and three for red.

1. Munzner proposed a nested model for visualization design and validation. Discuss/validate your visualization with respect to domain problem characterization and data/operation abstraction design.

**Domain Problem Characterization**Munzner states that in the Domain Problem and Characterization phase that the author must learn about the users and the data they are representing. It is not really possible to tell whether the author did a thorough analysis of the of his users though one could argue that he does know the users of the xkcd site. This is a visualization meant for a mass audience so it would have been difficult to engage with target users before creating this visualization.

**Operation and Data Type Abstraction**The data source for this visualization is from DW-NOMINATE which is a statistical system used to calculate a politicians ideology. The DW-NOMINATE system was created by Keith Poole and Howard Rosenthal. There are two aspects to this phase, the first being to map the data from the vocabulary of the domain (DW-NOMINATE) to something more generic. A quick look at the DW-NOMINATE site will quickly show that the author must have done this to put the data in a more generic format his users could understand. The DW-NOMIATE data is highly technical in nature.

The second aspect to this phase is to manipulate the raw data into a format the visualization can use. It is not possible to know how the author abstracted this data into a format that was usable by his visualization, but it is highly likely that some form of transformation was done. A sample of the raw data is below and it is likely that the author put this into a form (tabular for instance) that was easier to read and manipulate.

Figure 1: DW-NOMINATE Data from the 1998 House

105 99909 0 99 0 USA 100 0 0 CLINTON -0.977 -0.211

105 15090 15208 41 1 ALABAMA 200 0 1 CALLAHAN 0.802 0.456

105 29300 0 41 2 ALABAMA 200 0 1 EVERETT 0.799 0.601

105 29700 0 41 3 ALABAMA 200 0 1 RILEY 0.795 0.607

105 29701 0 41 4 ALABAMA 200 0 1 ADERHOLT 0.704 0.710

105 29100 0 41 5 ALABAMA 100 0 1 CRAMER -0.041 0.476

Munzner also discusses the validity threats in her nested model. The threats in the Domain can be characterized as the author trying to solve the wrong problem. Either the author has misunderstood the problem or no problem exists that could be solved by a visualization. In the case of the Congress visualization the graphic was created for the users of the xkcd site – there is no problem to solve per se, but this concept could be translated to ‘would the visitors of the xkcd site find this visualization interesting.’ It is hard to have an answer for this, but one could infer that it was popular by the fact that there is a poster of this visualization for sale in the xkcd store.

The validity threat in the Abstraction phase is that the data types do not solve the problems of the users. In the case of the xkcd visualization the question would be whether the DW-NOMINATE data provides the ideological mappings correctly for the members of Congress. Munzner states that this threat can be eliminated by testing the data on the target audience, which in the case of xkcd is not realistic.

1. Based on Cleveland and McGill’s results, does your visualization embody good practices (i.e. can people accurately perform the tasks based on the encodings?)

The closest type of chart that was reviewed by Cleveland and McGill to the Congress visualization is the Statistical Map with Shading. Cleveland and McGill recommend replacing this type of chart with a Framed-Rectangle chart. However, since the Congress visualization represents data over time a grouped bar chart such as the one below would probably be a better choice.

Figure 2: Grouped bar chart with imaginary data

Cleveland and McGill also state that their results merely provide guidelines. I think that presenting the data in a grouped bar chart provides clarity for a given year, but it does not give the overall picture of how Congressional partisanship has changed over the years. The xkcd visualization does provide this and I think people can accurately perform the tasks based on the chosen encoding.

1. Do you agree that visualization is a functional art? Explain.
2. 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.)