Exercise 4 Report: Co2 Emissions Analysis for the US

Paul de Fusco - DSE241 - Cohort 3 - 2/1/2018

Outline:

The submission for exercise 4 consists in a Tableau Dashboard comprising four worksheets:

1) "CO2 Emissions for 50 States"
2) "CO2 Emissions and Population for Each State"
3) "Total CO2 Emissions by Sector"
4) "CO2 Emissions per State Resident"

Task Accomplished:

The dashboard provides insights for data visualization with respect to different combinations of variables. The main question that it addresses is: "How have emissions changed through the years?" - Each workbook accomplishes a different subtask

Data Model:

The exercise started by importing both the 'population' and 'emissions' files provided. The dataset was built by performing an inner join by columns 'state' and 'year'. As rows, only the 'year' dimension was added. In the columns section, the 'state' column was used as dimension while the sum of all emission values and the sum of all populations were added as attributes. As a result, the emissions and population data was grouped by year and state.

Workbook Discussions:

Workbook 1: "CO2 Emissions for 50 States"

The visualization idiom is of type tree map. The task it accomplishes is to display emissions for each of the 50 US states, for each year between 1990 and 2010, in a way that emphasizes emission quantities differences between states. Each large box area represents all years for a specific state depicted with the same color as shown on the legend to the right. Each of these areas is further divided into 20 rectangles, one for each year of measurement. The size of each reflects the amount of emissions for that particular year.

Expressiveness - Effectiveness Principles and Color Choices:

In terms of expressiveness, the idiom showcases each data point within the dataset and does not add data points not required. With respect to effectiveness, the channels' salience reflects the importance of the emissions with rectangle size - the larger the rectangle, the higher the emissions. The color is associated with each individual state irrespective of any color spectrum association as each state is different and unique from others. Tickmarks and grids are also used, as well as the ability to hover over any data point to visualize its values.

Channel Effectiveness Analysis

1. Accuracy: the size of each rectangle reflects the amount of emissions for that year. Although years with higher emissions are larger than others, the scale is linear and as emission levels are in many cases close to one another, this often results in rectangles of very similar size. The alternative would have been to apply a power function to the size of the rectangles, but that would have distorted the result.
2. Discriminability & Grouping: rectangles are grouped into larger boxes based on state belonging. Each state is uniquely identified by color. Thus, each state is easily distinguishable from others.
3. Separability: the choices of rectangles and boxes as physical elements, in addition to color associated with each state, makes the difference between rectangles easy to make.
4. Popout: rectangles with higher emissions appear larger than others. In addition, states are sorted from left to right in terms of emissions quantity, thus placing Texas and other states with higher emissions in a context where it is easy to spot them.

Workbook 2: "CO2 Emissions and Population for Each State"

The visualization is a series of plots for each state, arranged in sorted alphabetic order vertically from top to bottom. The task accomplished is to provide further insight into quantities of emissions in relation to population size, for each year and state. Unlike workbook 1, this idiom does not put all the emphasis on emissions because it does not just convey quantities of emissions by box sizes; instead, population quantity is also shown with the same level of emphasis.

Expressiveness - Effectiveness Principles and Color Choices:

In terms of expressiveness, the idiom showcases each data point within the dataset and does not add data points not required. With respect to effectiveness, the channels' salience reflects the importance of the emissions with plot lines - the higher the data point, the higher the quantity. The two variables are displayed on the same plot but represent different data quantities (through use of dual axis). This allows to compare quantities and patterns while still providing a compromises in terms of visual impact. For consistency, the color yellow is used for emissions while a light form of blue is used for population. To aid the viewer, the corresponding axis are highlighted in a color that is a more opaque version of the plot line it relates to.

The same color is associated with each individual curve irrespective of any color spectrum association as each state is different and unique from others.

Tickmarks and grids are also used, as well as the ability to hover over any data point to visualize its values.

Channel Effectiveness Analysis

1. Accuracy: to emphasize changes in quantity for each axis, it was important to use dual axis. If otherwise population and emissions had been on the same scale, the differences in emissions would have been unnoticeable.
2. Discriminability & Grouping: each state is clearly labeled; also, the emissions and population curves are easily discernible because of the highly contrasting colors chosen.
3. Separability: plot lines are easy to notice within the idiom. Visually, each plot line is associated with its reference axis to the left by means of background color choice for the axis title.

Workbook 3: "Total C02 Emissions by Sector"

This visualization includes a separate plot for each sector causing emissions, for each year in the dataset. The task accomplished is to show how some sectors are increasing the pace of emissions while other sectors have emitted very similar quantities for many years now. Unlike in idiom 2, this visualization aggregates all state emissions and is therefore much simpler as it only includes 5 subplots (one for each state).

Expressiveness - Effectiveness Principles and Color Choices:

Expressiveness: all the sectors and only the sectors from the dataset are shown in order to fulfill the visualization task. In terms of effectiveness, the choice of the plot/curve as a channel (with identical orders of magnitude for each subplot) allows to objectively compare emissions between different sectors. Tickmarks and grids are also used, as well as the ability to hover over any data point to visualize its values. Just as above, different shades of the same color (in this case green) are used by the title, the plot lines, and the axis title background, to highlight the fact that they refer to the same data points.

Workbook 4: "CO2 Emissions per State Resident"

Task accomplished: show the ratio of emissions to population through the years for each state. This workbook was designed including a combination features from idioms 2 and 3: the same color is used in different shades for axis, title, plot curves axis background to improve discriminability and separability. The subplots are small but large enough to showcase differences in quantity on an identical and objective scale. Because the ratio computes to very small numbers, the addition of tick marks and gridlines allows the viewer to distinguish among curves with ease.