1. Research Question(s):

The issue of climate change and its connection to GHG's has been widely studied and most notable by the IPCC. The Fifth Assessment Report for the IPCC [1] looks at the physical science basis for climate change and the different forcing mechanisms that affect it such as solar radiative forcing, greenhouse gas emissions such as carbon di oxide (CO^2) in addition to others such a biogenic methane from cattle farming among others. The report also has a chapter on the different regions and the chapter covering Europe [2] considers the impact on many sectors such as farming, water supply and biological conversation as a result of climate change. As one of the principal causes of climate change are GHG Emissions there is a large body of work that looks at the relationships between economic factors and emissions such as [3]. Here the authors note that there is an inverted U shaped relationship between some environmental degradation and income indicators. However, this does not hold GHG Emissions as their effect tend to be global rather than just local and there is an inter-generational nature of its negative effect. They find that capacity and responsibility of regulatory agencies to make intergenerational transfers in the presence of such long term effects were key determinants of the relationship between economic growth and GHG's. In [4], the authors note the lack of analysis of the aggregate emissions of GHG from land use changes such as agriculture, forestry which is termed non-industrial emissions. Based on the statistical analysis of 129 countries from the period of 1971-2010 they find that economic growth is a driver of industrial and non-industrial emissions. Also economic growth is found to have twice the effect on industrial emissions that non-industrial ones. Also the impact of industries such as farming on GHG Emissions are considered by [5]–[7]. The authors in [7] considered the dietary choices of people in the UK and its impact on GHG emissions and found that switching to more plant based diets as opposed to meat based diets could contribute to a lower GHG emissions and [5] notes that GHG savings of 22% and 26% can be made by changing from the current UK-average diet to a vegetarian or vegan diet. In [6] the authors conclude that with appropriate policies UK GHG emissions from farming could be cut at near zero to no cost.

With these readings from the literature, I decided to focus my analysis on the UK and look at Total GHG Emissions and their relationship to socioeconomic variables beyond income or economic growth such as Population as well as Total Energy Use. The data is from the time period of 1960 – 2012 to ensure completeness of the variables.

The research questions that I intend to explore are:

- Explore the link between socioeconomic variables such as Total CO2 Emissions, Total Energy Use, GDP and Population. Is there correlation among them?
- Explore the yearly trends between these variables. Are certain years better/worse than others?
- Confirm whether the trends noted in the literature can be observed in the UK data.

2. Data Source(s):

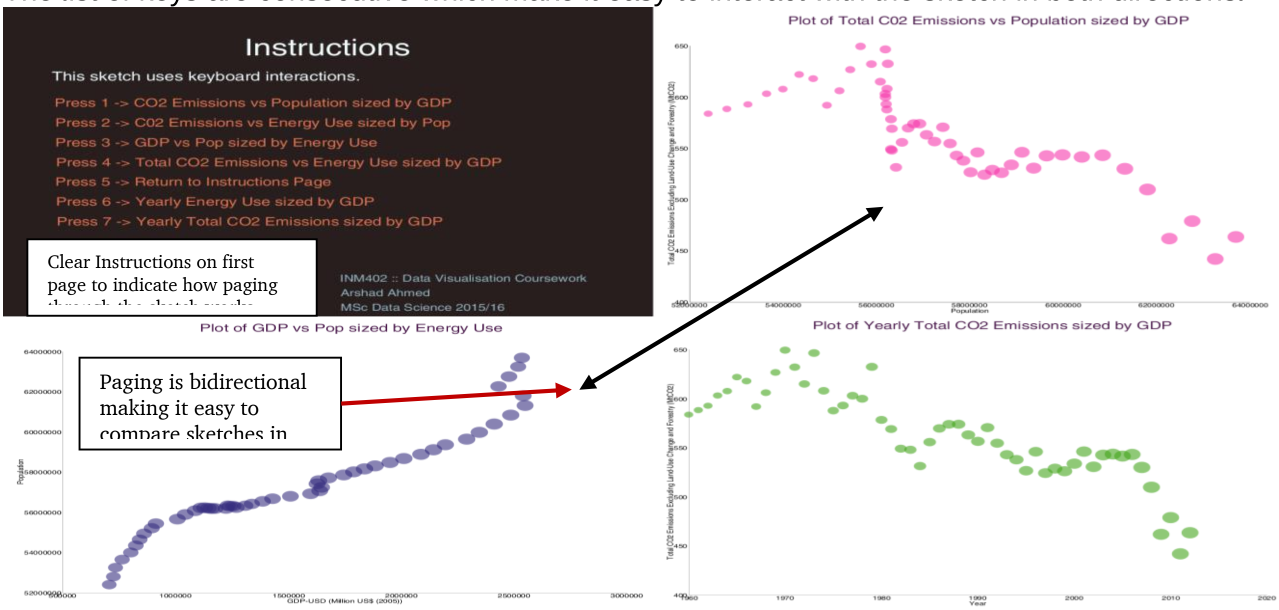
The data used for this sketch comes from the World Resources Institute's historical emissions data. Formally the dataset is cited as:

- CAIT Climate Data Explorer. 2015. Washington, DC: World Resources Institute. Available online at: http://cait.wri.org.

The original dataset comes in four different files but I only use three of them namely the Socioeconomic, Total C02 Emissions and Total GHG Emissions data for all countries. From this I extracted the dataset for the UK for the time period 1960 – 2012. This was done because the emissions data did not stretch far back beyond this point. Therefore to reduce the number of missing data this range is chosen for completeness. Also it was noticed that the Total GHG data did not extend beyond 1991 so when the scatter plot markers was sized using this data the points prior to 1991 become zero. To get around this problem I filled this missing values with a constant of 5 just to have a sensible size of the scatter plot circle markers. Also all the extracted data for the UK is merged into 1 combined file and this is used in the sketches. No additional scaling or aggregation of the data is performed as the data is already curated.

3. User Instructions:

Interacting with the sketch is very easy and is indicated on the first page. The instructions page explains clearly what type of interaction the sketch uses, which in this case is keyboard interaction. The list of keys are consecutive which make it easy to interact with the sketch in both directions.



4. Design Justification:

4.1 Design Framework

The guiding principles for the visualization design come from Tufte [8]. The key principles utilized here are that the visualization should aid the analyst/user to think about the important messages from the data rather than spend time worrying about methodology such as graphic design, technology of production. The visualization should avoid distorting the data, while not present many numbers in a small space but take care to present the most important ones. Also it should make coherent large and complex datasets so the end user is able to compare different data. It should provide multiple layers of detail starting with an overview and providing details on demand.

In this sketch I use paging to present subsequent data charts and provide clear instructions on how to navigate between them. This is to prevent data trends being skewed due to variable scales in the data. Also it is a practical manifestation of the principle of not presenting too many numbers in a small space but to maximize data to ink ratio. The paging with keyboard interaction help with comparison of different data as initially paging with mouse interaction was tested and found that it is more difficult for the end user to quickly switch between say the first and sixth chart quickly to compare trends with a mouse than it was using the keyboard. Multiple layers are shown in the scatterplots through the use of variable area marks.

4.2 Visual Encoding

The four levels of design identified by Munzner[9] are 1) domain situation, 2) task abstraction, 3) Visual Encoding / Interaction Idiom and 4) Algorithm. The algorithm cases is not relevant for this application as the visualization consists of drawing scatterplots and the dataset is not large so there no concerns over performance and memory issues. The domain situation in this case the GHG Emissions and their relationship to socioeconomic variables such as GDP, Population and Total Energy Use. The task abstraction is motivated by the data type which is tabular and numerical in this case. Therefore scatterplots are an appropriate for this dataset. The visual encoding controls what the users ultimately see. So for my sketch this means that when the sketch is run the first page the user encounters in the 'Instructions' page. This sets out clearly what data is presented and how to navigate through it. This translates to 7 options in total from which 6 keys are allocated to the data charts and 1 is allocated to returning to the homepage at any time.

The scatterplot is the visual encoding of choice for this sketch. The scatterplot encodes two quantitative variables using both horizontal and vertical spatial position channels where the mark type is a point. Munzner [9] also notes that area as a magnitude channel for marks are more effective than Volume, Saturation and Hue. Therefore, we avoid the latter in our design and only use area encoding for the marks and the color is kept constant within the scatterplots. Setting the mark type as a circle allows for additional augmentation as area can be encoded in to the mark and thus a third variable can be represented on the plot. This visual encoding supports the research questions of exploring relationships between variables by essentially performing the perceptual task of looking for patterns such as whether the points form a line along the diagonal. In addition this allows for easy identification of clusters, outliers, trends and distributions. [9]

4.3 Interaction

The interactions used in this sketch are keyboard interactions and paging to control the data narrative. Since I am presenting a series of charts in order to prevent over plotting and cognitive overload I opted to present each chart in a separate page with clear instructions on how to reach each chart. The keyboard interactions help the user to compare easily between charts that are not plotted in proximity. This allows for faster switching between views for example it is faster to switch between charts 1 and 6 using the keys on the keyboard than it is to switch by clicking on tabs using a mouse. This was noticed in the initial design and then changed for better usability.

4.4 Typography

Fonts can be broadly classified into 4 categories such as Serif, San Serif, Hand Writing and Decorative although other groupings are found in the literature. Of these font types Hand Writing and Decorative fonts are excluded for this application because they lack the tone and character necessary for this work. Hence, I consider Serif and San Serif font as they are considered to be appropriate for serious and technical work.[10]

The author in [11] uses a semantic differential methodology where participants are asked to rate the extent to which a certain adjective such as 'technical' or 'professional' characterized a certain typeface. They report that for their sample the participants rated the Helvetica font as being most appropriate for 'technical' writing. In [12], the author considers the how different fonts are perceived by the ageing human eye. They note that the key characteristics of fonts which are successfully read by older people or those with low vision have consistent stroke widths, open counter forms, pronounced ascenders and descenders and wider horizontal proportions. Sans Serif fonts are best suited for this and especially Helvetica with its larger x-height and wide proportions are found to improve legibility under such low vision conditions. These same characteristics make it easier for the font to be read on the screen and is perceived to be a modern and neutral font[13]. Neutrality is an important consideration as the presentation of this data should be not be associated with any bias. This is the basis of choosing this font for the visualization.

4.5 Color

The relevance of color theory in the data visualization context is discussed in [9]. They note that human visual system processes color into three opponent color channels which are 1) red to green 2) blue to yellow and 3) black to white this part encodes luminance information. The luminance conveys high resolution edge information while the other two convey lower resolution information. Color deficiency which is inability to perceive common color channels and a common version of this is deuteranopia and protanopia which are gender linked and forms of red green color blindness. This affects mainly men. The choice of colors will have to take this into account so the visualization can be utilized by a wide audience. Color Spaces are mathematical representations of our perception of color and is represented in 3D space. Therefore an ideal color scheme should be perceptually uniform and perform well under moderate color deficiency conditions. The inspiration for the colors used for this visualization come from the Viridis color map[14] based on the CAM02-UCS[15] color space which is freely available with Python. This color scheme is designed to be perceptually uniform with uniform perceptual deltas. So the change in value from 0.2 to 0.3 would give the same magnitude of change in color as if the value changed from 0.8 to 0.9. Also the color map performs well when tested with moderate color blindness. In addition color map also produces perceptually uniform profile when turned to grayscale. Thus the colors selected for the sketch make use of colors from the composed mainly from the yellow green part of the spectrum.

Also as a way of representing change the different charts are colored uniquely to indicate that different quantities are being compared and the colors used do not repeat so insights can easily be associated with the color of the chart. Light transparency is used so that overlapping regions do not form a concentrated mass but it is clear that the points overlap.

5. Further Work:

The key improvements that can be made to the sketch are in the area of increased interaction and reduction. In terms of interaction, the sketch could be improved to allow variables for plots to be chosen interactively or through drag and drop interaction. The encoding of the marks could also be made dynamic by allowing dynamic drag and drop interaction from all the variables available. Also tooltips that show additional details could be added with drag and drop interaction so the user is able to choose what details they would like to see. The sketch could be enhanced by reduction. In terms of reduction, item filtering could be added so that selecting a point or sets of points could give the option to be filtered and analyzed separately from the rest of the data. Attribute filtering could be added so we can eliminate attributes rather than items for example removing the encoding for the marks. Aggregation in this dataset would not be appropriate as the data is already at a national level and pre-aggregated. Also to aid in the assessment of correlations between variables the option to plot lines of fit through different functions such as linear, power or exponential functions could be added and the measure of fit score displayed. Higher scores would indicate better fit and a choice of functions would allow greater flexibility in analysis. The scale among the data variables vary so the option to plot on log scale or semi-log scale would be useful. This would allow for the easier investigation of non-linear correlations in the data.

6. Data Insights

From the charts presented we see some clear correlations emerge among some variables and no correlations among some. We can see that there is a negative correlation between Total CO2 Emissions and Population. This is surprising because my initial expectation was that there would be a positive correlation between these variables because other authors noted the positive correlation between economic growth and rising emissions. But in the case of the UK it is seen that GDP and Population are fairly positively correlated so as they increase the emissions decrease. But it must be noted that the Total CO_2 variable used here is excluding land use changes. But we noted from the literature that such industrial emissions are driven by economic growth. But even then we see that with increasing GDP and Population the Total CO_2 Emissions decrease. The Energy Use and Total CO_2 Emissions show no discernable correlations but clusters are obvious. We see that High GDP bubbles are clustered together and smaller GDP bubbles are clustered together. It could be argued that 3 clusters with small, medium and large GDP bubbles are visible. The high GDP bubbles correspond to periods of both high energy use and high emissions. The yearly trends surprisingly show that after 2000 the energy used reduced with a corresponding large fall in emissions.

It is interesting to note that for the time period considered from 1960 – 2012 the overall trend in CO_2 Emissions is on a downward trend while the energy use trend oscillates but overall there is a rise in the amount of energy used towards the end of the time period than at the beginning.. Also the plot of emissions and energy use sized by GDP is better than that sized by Population. This is because there is greater variance in the GDP variable than the Population variable.

7. References

[1] IPCC, "IPCC Fifth Assessment Report (AR5)," IPCC, 2013.

[2] IPCC, "IPCC Fifth report 2014 Chapter Europe," Ar5, no. October 2013, pp. 1–51, 2014.

[3] A. Ansuategi and M. Escapa, "Economic growth and greenhouse gas emissions," Ecol. Econ., vol. 40, no. 1, pp. 23–37, 2002.

[4] L. F. Sanchez and D. I. Stern, "Drivers of industrial and non-industrial greenhouse gas emissions," Ecol. Econ., vol. 124, pp. 17–24, Apr. 2016.

[5] M. Berners-Lee, C. Hoolohan, H. Cammack, and C. N. Hewitt, "The relative greenhouse gas impacts of realistic dietary choices," Energy Policy, vol. 43, pp. 184–190, 2012.

[6] D. Moran, M. Macleod, E. Wall, V. Eory, A. McVittie, A. Barnes, R. Rees, C. F. E. Topp, and A. Moxey, "Marginal Abatement Cost Curves for UK Agricultural Greenhouse Gas Emissions," J. Agric. Econ., vol. 62, no. 1, pp. 93–118, 2011.

[7] P. Scarborough, P. N. Appleby, A. Mizdrak, A. D. M. Briggs, R. C. Travis, K. E. Bradbury, and T. J. Key, "Dietary greenhouse gas emissions of meat-eaters, fish-eaters, vegetarians and vegans in the UK," Climatic Change, 2014.

[8] E. R. Tufte, "The Visual Display of Quantitative Information," Technometrics, vol. 2nd. p. 197, 2001.

[9] T. Munzer, "Visualization Analysis and Design," 2014. [Online]. Available: http://reader.eblib.com/(S(r2yhbeyw1xkv2hlwevlfmjec))/Reader.aspx?p=1664615&o=1821&u=ewcvJ4kR2ClEFiVMEf8F0LPq3bI%3d&t=1461357229&h=1D311413C30E3BD9AE8124D52966A04536B35EC1&s=44512535&ut=6083&pg=1&r=img&c=-1&pat=n&cms=-1&sd=2#. [Accessed: 22-Apr-2016].

[10] i. Hakki, "THE EFFECTS OF FONT TYPE CHOOSING ON VISUAL PERCEPTION AND VISUAL COMMUNICATION," vol. 1, no. 3, pp. 35–54, 2013.

[11] J. Mackiewicz, "What Technical Writing Students Do Know and Should Know about Typography."

[12] P. Nini, "Typography and the Aging Eye: Typeface Legibility for Older Viewers with Vision Problems," AIGA | The professional association for design, 2006. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0. [Accessed: 29-Apr-2016].

[13] "The Simplicity of Helvetica," 2010. [Online]. Available: http://www.webdesignerdepot.com/2010/01/the-simplicity-of-helvetica/. [Accessed: 02-May-2016].

[14] "A Better Default Colormap for Matplotlib | Berkeley Institute for Data Science." [Online]. Available: https://bids.berkeley.edu/resources/videos/better-default-colormap-matplotlib. [Accessed: 30-Apr-2016].

[15] A. Color, I. Processing, C. Fernandez-Maloigne, A. Color, and I. Processing, Advanced Color Image Processing and Analysis. 2013.