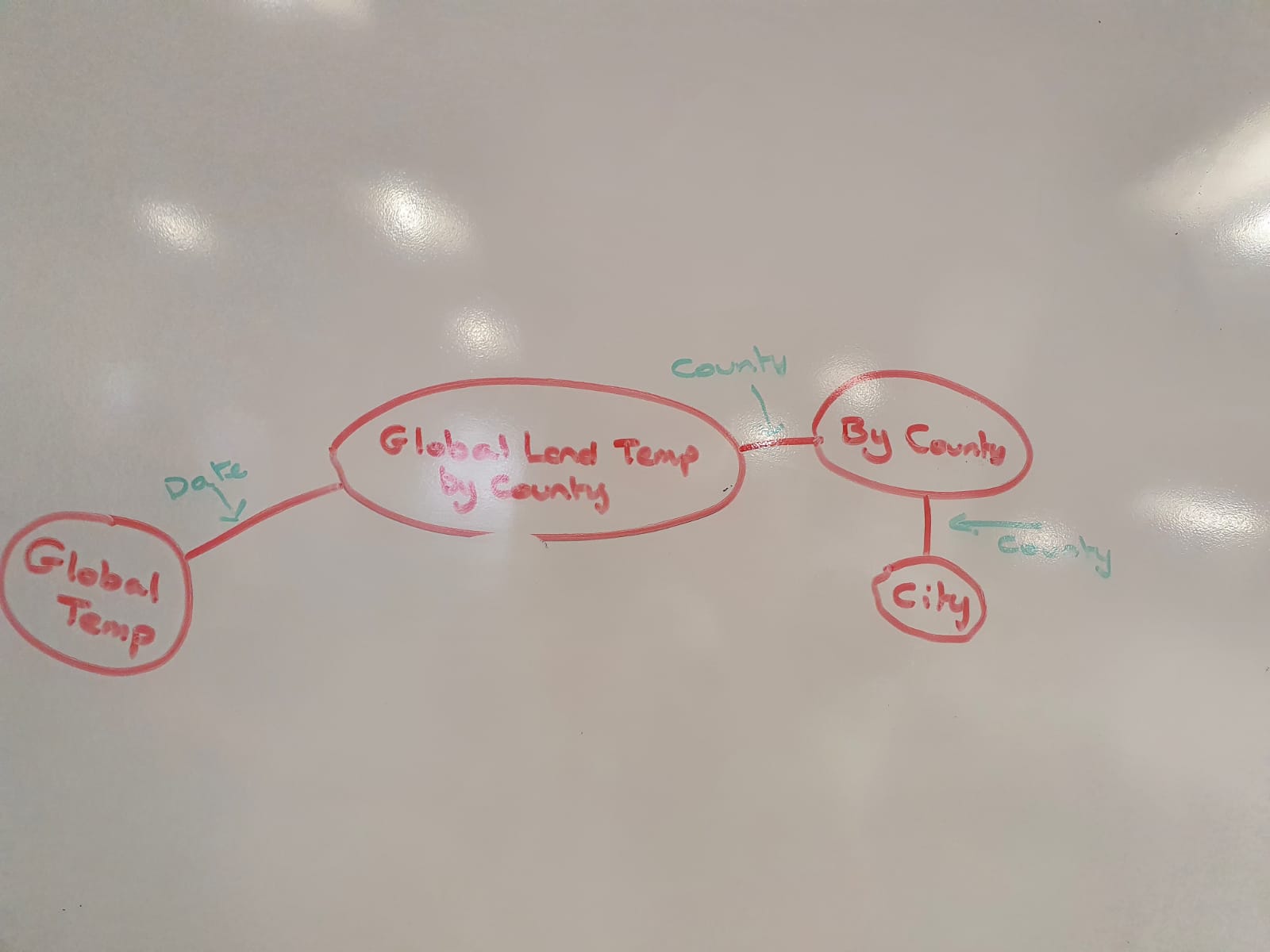
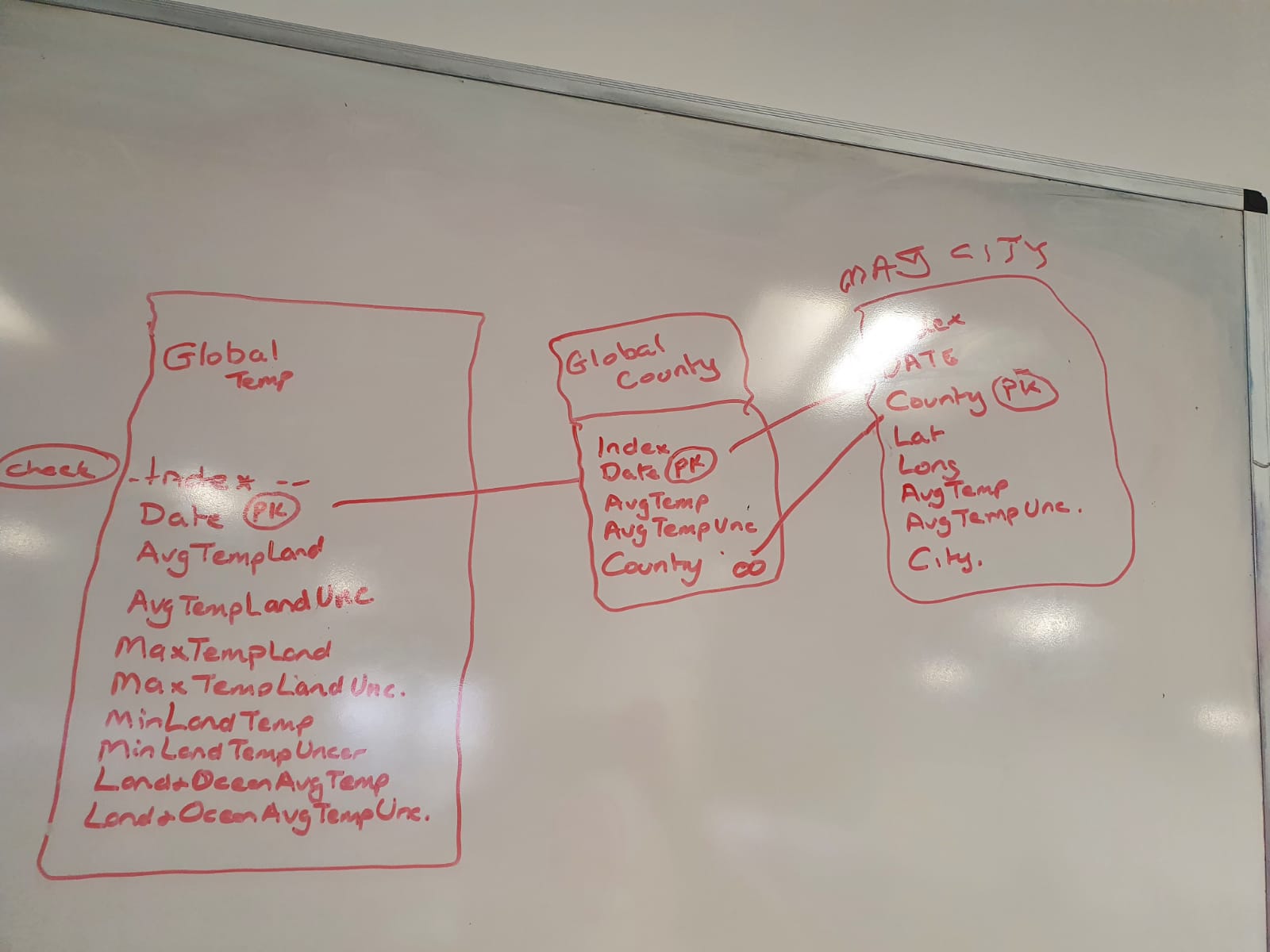
**Discussion of the open data considered by the group and the rational for the final selection.** Several data sources where considered by the team, firstly: a climate change data thingy, secondly: a criminal activity data both where found to be unsuitable, one for number of records being incomplete, the other for having such a small quantity of records overall, this has since been discounted, the third piece of data that we reviewed, we accepted as a team because of the large volume of data (13,000+ records) and all data entities where complete and useable. We fee; that the data set we chose had flexibly in 3 larger areas and provides meaningful comparison between those 3 areas.  
  
As seen from the image above, the data we focused on was of global appeal, this was a consideration of data that was available covering global temperatures, however, due to the size of this data source it would have been too cumbersome for us to use as this project, however, we were able too narrow down data for London UK, and this data set has underpinned the data used in this tool.  
  
  
  
The above image was used to demonstrate the complexity of the tables and led to the reason why we reduced the functionality to London alone.  
  
**Assessment of the potential insights that might be extracted from the data set and the interfaces needed to interact with your application (requirements)**  
  
The insight of the chosen data is the visualise the average temperature in London, within a 130-year period, this analysis can be further broken down to show patterns over years, seasons or months. **System that we're developing is designed to provide the end user with a functional capability for searching temperature data events between 1900 and 2013. This is underpinned by 1,300+ records captured during this time period by**: https://www.kaggle.com/yamqwe/global-climate-change-datae?select=GlobalLandTemperaturesByMajorCity.csv. Kaggle provided the temperature data that was used for this tool we developed.

User page, used for viewing all the data and you can click either month/year/season, with a drop-down choice, leading to another page, provides a data visualisation - allow the user to search with the queries, and with range, ect. It will become more flexible, and dynamic, as the user drills down on the data with the data changing, being dynamically changing graphs shown.

The interactive interfaces will provide insight by using the following:

**Seasons, Pie chart.**

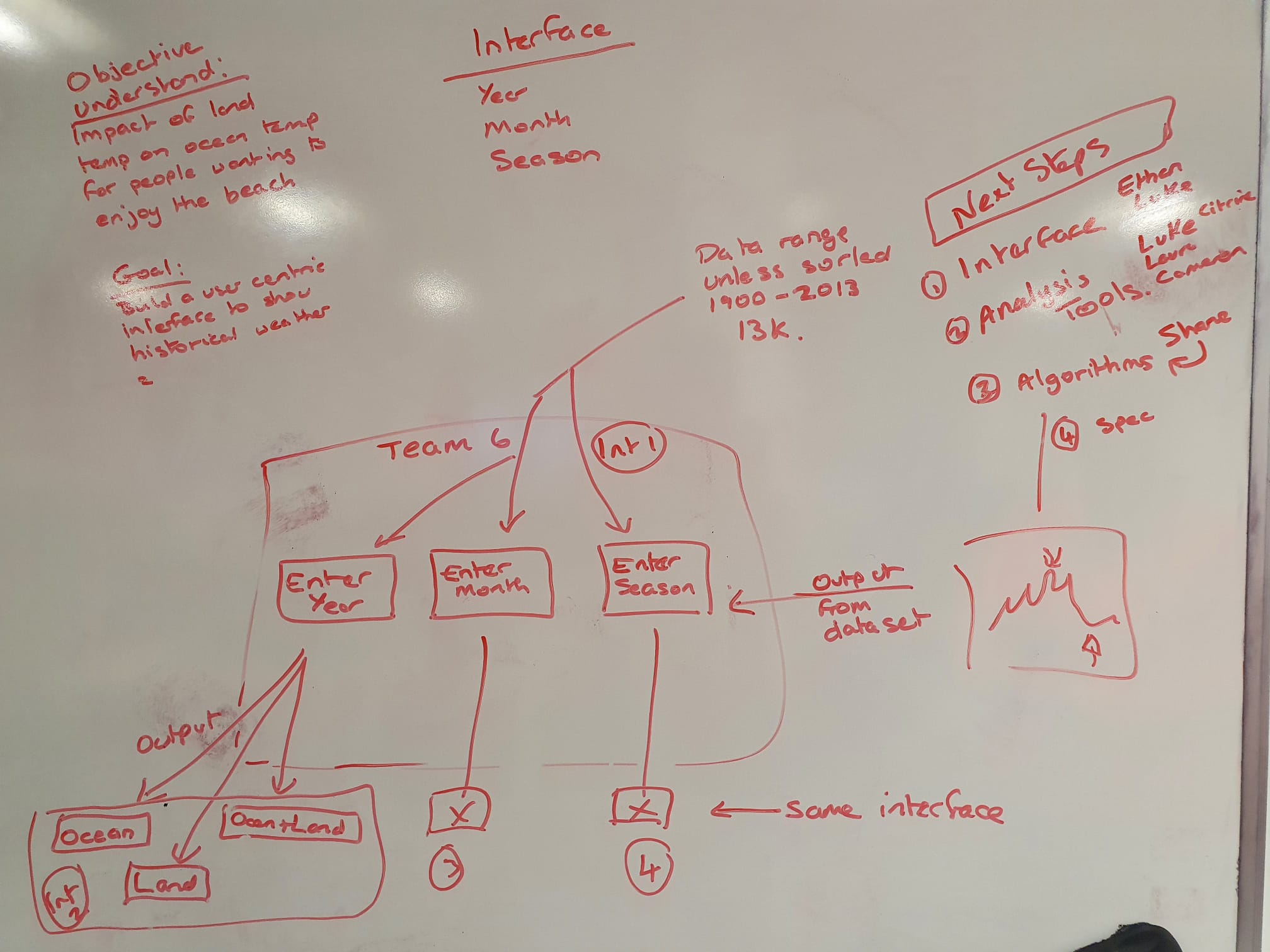
**Range of temperatures across the years, Bar chart. (Broken down).**

**Years, Bar chart/Horizontal Chart. - The 113 years will be split into 16 data points, temperatures Y Axis.**

**Months, Bar chart/Horizontal Chart.**The potential analysis from this tool could provide correlation assessment between the land temperature, and potentially backed no Ocean temperature and vice versa, however due to to time constraints, this complexity may not be built into the tool at this time.  
  
The software that is used to develop this tool eg: Visio Studio, Python, Matplotlib; whilst this may seem complex, together they complement the objective of the tool by enabling us to develop code in python but provide the user with visualisation of data splits which is dynamically split, and generated by the user input – the data itself changing, providing a dynamic visualisation of the data -   
  
**Algorithm exploration with a description of any optimisations made.**The algorithm/process as such is that we're going too:

- Read from the data file (large data set)  
 - Search criteria (year, season, month) EG: Can look at seasons for all 113 years.

TOP LEVEL - Restrictive, extract data from the year.  
MID LEVEL - Seasons over 113 years.  
BOT LEVEL - Similar, but for month.

Will still show volumes, will still show months, but will be pre year.  
  
 The optimisation we spoke about already, and that was we looked at design needs and identified areas in which we could re-use the code with adaptions simply to with elements such as variables, and their names, ect. As a team, we followed the basic data model process of **concept, logical, physical**.   
  
As part of the concept process, we captured what we understood what was the requirement; once this was understood by the rest the team at a higher level, we where then able to challenge our thinking and the processes we understood at that point, we then looked at what areas we could simplify such as the volume of data was reduced to London, UK we simplified the capability to enables us to design a robust working system rather than as a project fail due to time. The functionality that we have developed; we are confident it could be replicated with comparative ease to provide further insight covering land temperatures and ocean temperatures, but as a team we felt we would lose sight of the objective of the task through overcomplicating the task at hand.  
  
**{PROGRAMMER INPUT HERE: When the programmers are looking at the pseudo code – pseudocode review here.}**

**Design that identifies how the application will process the data and interact with the user.  
  
System that we're developing is designed to provide the end user with a functional capability for searching temperature data events between 1900 and 2013. This is underpinned by 1,300+ records captured during this time period Kaggle.** (Text repeated twice, since it fits, should be similar but worded differently)  
  
Aiming at the general public, shows the data visualisation more clearly, chosen for specific audience use, customer centric. Originally considered NetworkX however it wasn't suitable for the kinds of graphs wanted to create, as such we decided to utilize Matplotlib, as it would allow us to generate the type of graphs needed.   
  
Although, we envisage that the data be used in a certain way, we are aware that the user may use it in a completely different way for their own purpose. **(Echo mention of what’s in the “Algorithm exploration” part of report, somewhat.)** The application allowing the user to be able to refine their data search, with the user being able to pick a year, and making the data pool smaller, allowing the user to have a more personal experience with the data, but also allowing them a lot of choice in the data they provide; allowing for dynamic data to be provided. **Description of the implimentations.**- What was implemented?  
- Why is it going to be implemented?  
- What was excluded?   
- Why was it excluded? **Testing.**- What was tested  
- Why was tested  
- How it was tested  
- What was the result of the tests?  
 **10 Project Meeting minutes.  
[Meeting minutes external ZIP file – all included].  
  
Evalulation.**- The build process  
- Teamwork/Improvement  
- Ease of development  
- Success of development  
- Development implementation/documentation  
- Algorithms  
- Ease of use of the final program  
- Summary (overall, teamwork, system design, system success, ect.)