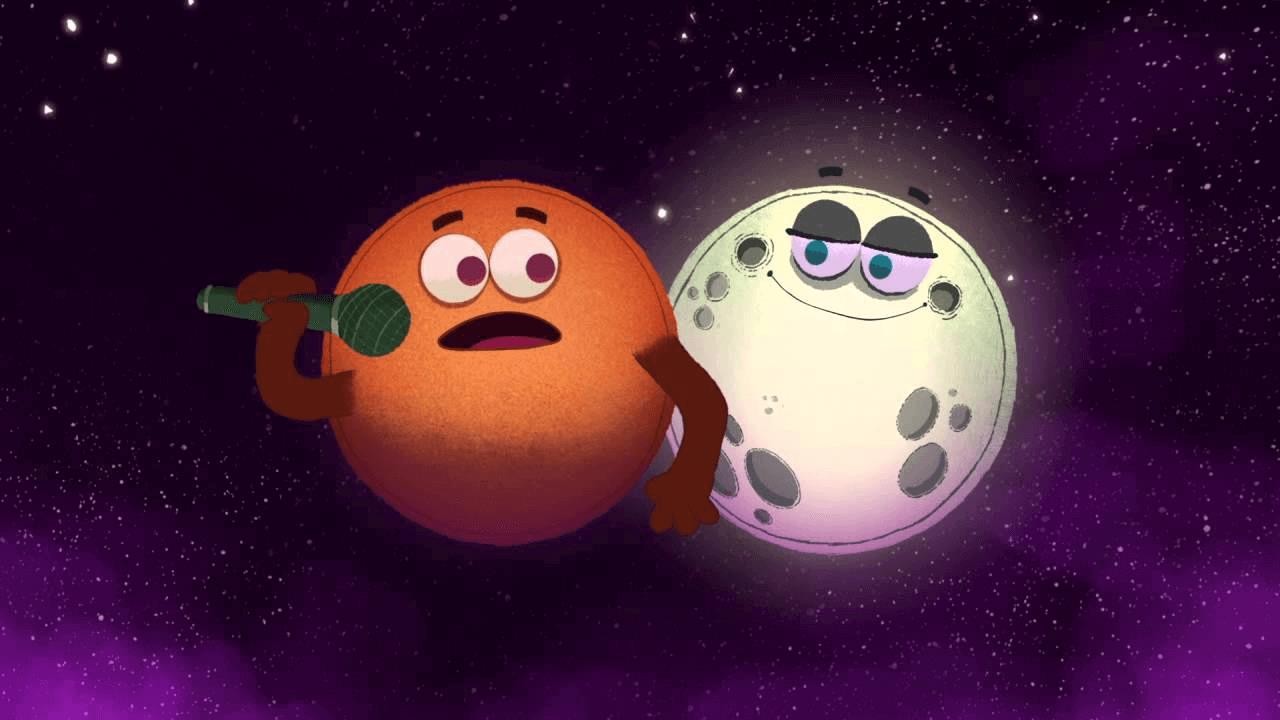
Moon & Mars



**BY-**

**VIDHU SHEKHAR TRIPATHI**

1

Contents

1. [Introduction 3](#_bookmark0)
2. [Design 4](#_bookmark1)
3. [Implementation 15](#_bookmark2)
4. [User guide 15](#_bookmark3)
5. [Conclusion 20](#_bookmark4)
6. [Bibliography: 20](#_bookmark5)
7. [Appendix: 21](#_bookmark6)

# Introduction

## Message-

Climate change is not a concept or theory anymore, it is happening in real time. Climate change results in the change of water cycle of the earth which in turn leads to extreme weathers and temperature changes across the globe. One of the leading causes for climate change is the extreme fossil fuel consumption across the globe.

Through the given narrative, I have tried to convey my message through a fictional conversation between planet Mars and our moon using factual dataset visualization.

## Intended Audience-

This narrative is aimed for the following-

* + Policy makers
  + Any one with a doubt about climate change
  + Corporate Head and Members
  + Students

#### Motivation-

The motivation was kindled by Prof. Jeremy Rifkin in his lecture on Industrial revolution 3.0 on vice.

(Source: https:/[/www.y](http://www.youtube.com/watch?v=QX3M8Ka9vUA&t=1267s))o[utube.com/watch?v=QX3M8Ka9vUA&t=1267s)](http://www.youtube.com/watch?v=QX3M8Ka9vUA&t=1267s))

# Design:

### Visualization Process-

The visualization process as explained in the appendix shows a detailed process of designing the entire application

The process followed steps as follows-

* + Brainstorming-

This part of the process involves bringing all the ideas on the table and simply focusing on everything that might work to support the narrative

* + Deciding what not to pick

Once we have all the ideas for the narrative its important to decide which one those works best for the given case scenario

* + Finalize the narrative

Decide the basic outline and details of the narrative around which the issue is focused

* + Brining it all together

Joining the above created template with the selected representation to get an idea of how the actual application may look and assess if it will suffice the requirement or not.

### Selected Plots vs Alternatives- 1)Map-

Selecting a map to represent the desired message was a tough decision. I considered following choices-

* + Leaflet Map
  + Plane Map with Heat Shade
  + Plane Map with Distinct Colours

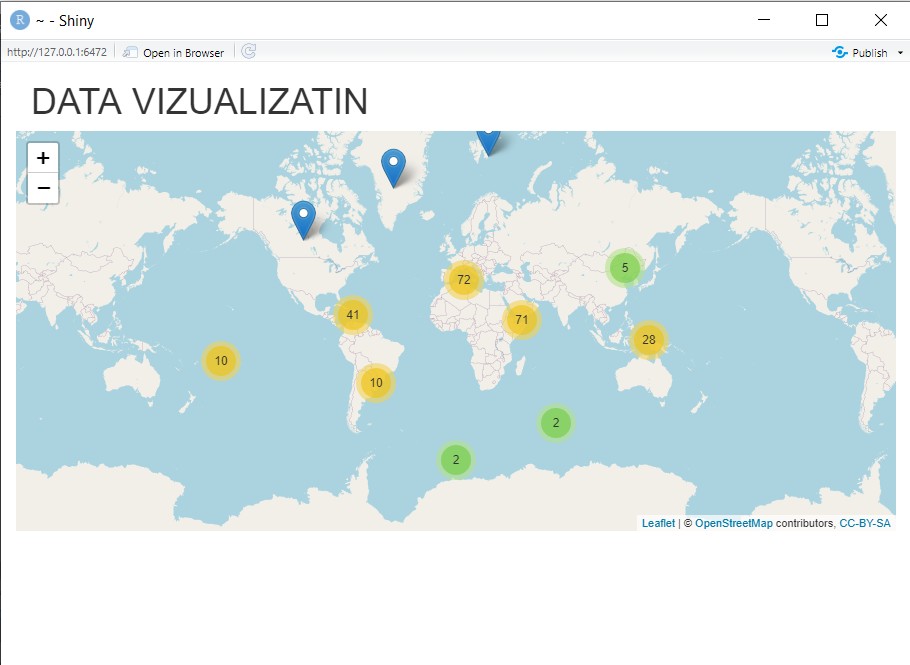
Leaflet Map-

Leaflet map was considered due to following reason-

* + Leaflet map enables user to zoom in to each particular country and check individual country score.
  + It also gives a description about number of records present in data in a particular region

But leaflet map was not considered due to following reasons-

* + Aim of the narrative is to describe a “conversation” and thus messages or graphs should be able to able convey a lot of information in a quick and abstract manner, where leaflet map is not the ideal choice
  + Its hard to study effect on values over time using a leaflet map



*Figure 1 Map using Leaflet*

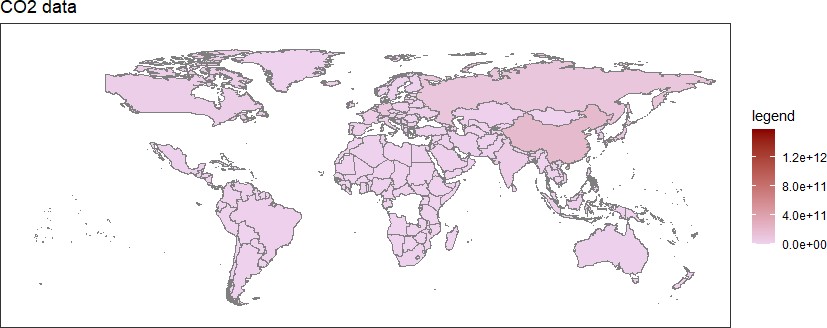
Plane map with Heat shade-

Plane map with Heat shade was considered due to following reasons –

* + Almost perfect choice to display change in temperature all over globe in a abstract and quick manner.
  + Change in values over time can be studies with ease.

Plane map with Heat shade was not considered due to following reasons-

* + Map is used to describe threat of flooding and scarcity of water , which requires clear identification of each country which is harder when using heat shade.

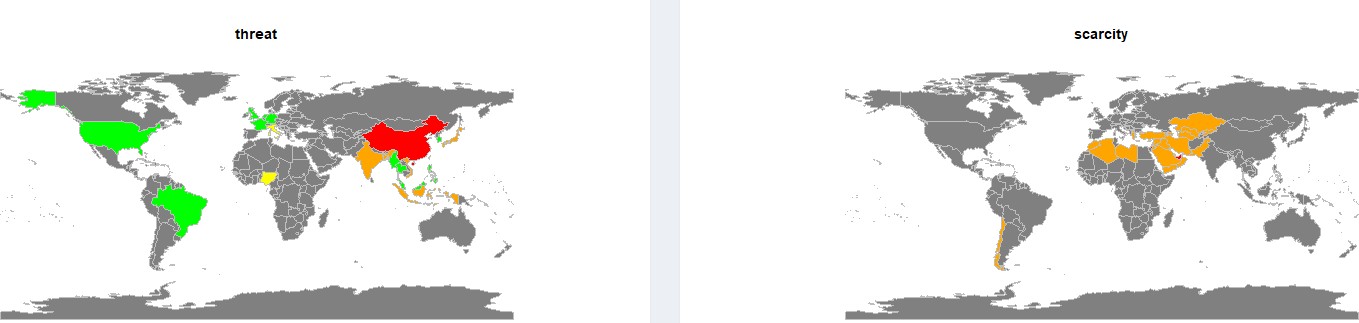


*Figure 2 Map using Heat Gradient*

Plane map with Distinct Colour-

Plane map with Distinct Colour was considered due to following reasons –

* + Almost perfect choice to identify exact county and also describe the categorical bracket it falls terms of values. all over globe in a abstract and quick manner.
  + Best choice for the narrative



*Figure 3 Map using Distinct colour*

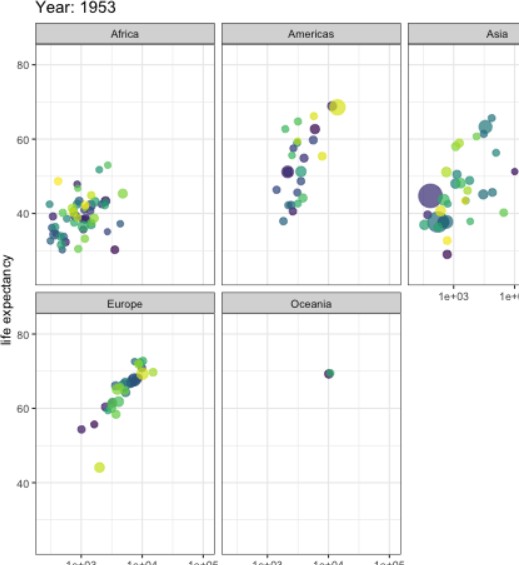
### Time Series Pollution data –

There are many ways in which we can plot values of Fossil fuel gasses in the atmosphere but in-order to suffice requirements of our narrative which here demands, that mars gets scared just by looking at one particular figure. It’s also important for the graph to be in form of animation , otherwise to plot growth of Fossil fuel emission (multiple gasses) for each year and for multiple countries will require a very large number of graphs which will destroy the narrative and will be very hard to for user to understand the message conveyed.

Following choices were considered-

* + Scatter plot with Circle Markers animation
  + Multiple line chart with animation

Scatter plot with Motion animation-

Scatter plot is a useful tool to represent large amount of data for multiple countries with circle markers making it convenient for the user to identify each country. It also has the ability to indicate multiple parameters such as change in size and coupled with facet wrap, it beautifully captures growth in major regions of globe without loosing details of individual countries over individual year, as shown in the figure below.

*Figure 4 Scatter Plot animation*

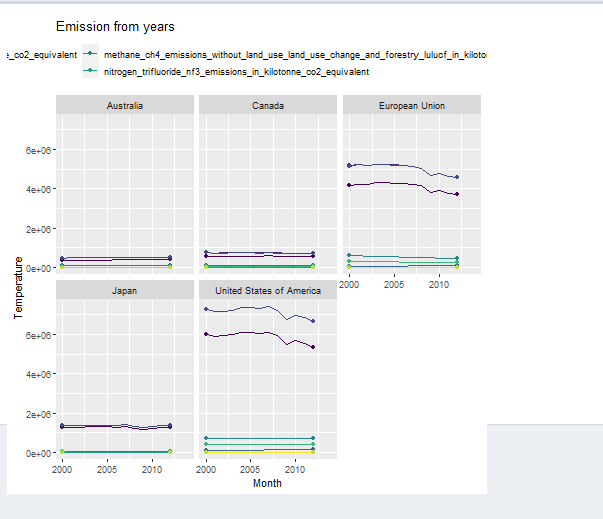
Say we want to represent increase in values of co2 with respect to sf6 over the years for given number of countries along with increase in level of any other gas, here our consideration fits well and can be considered a good choice but it has following limitations –

* + Any increase in number of parameters and the graph cannot suffice the requirement of either representation or usability.
  + Sometimes representation gets clustered and is hard to identify and extract exact inference or pattern.
  + If the message was about identifying correlation between data points, this was a better choice but not for identifying trends for multiple categories

### Line chart animation-

Line chart became the choice for the representation due to following reasons –

* + Each category of gas is clearly represented with clear trend identification.
  + Animation enables the plot to not only deliver the message but also captures the essence of the narrative.
  + Facet wrapping enables us to display trends of multiple parameters for multiple countries
  + Even if we want to increase number of gasses in our plot, this will still work perfectly, thus its highly robust.



### Temprature Variation-

*Figure 5 Line chart animation*

Our third part of the narrative focuses on how temperature is increasing all over the globe along with the uncertainty. Note that, it’s important that both the aspect of the global temperature is captured in-order to support the narrative. Also, since we are talking in global perspective its important that we represent data for multiple countries in an abstract manner.

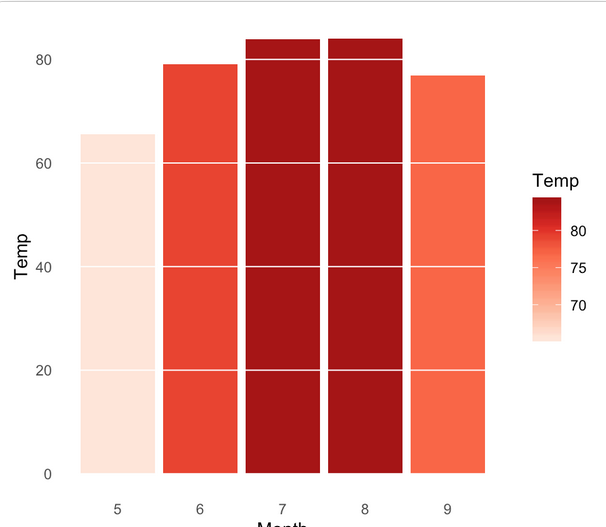
### Histogram-

Histogram is a useful tool and coupled with heat colour palette its one of the best choices when it comes to representing temperature data over a time series. We

could have gone with a facet wrap animation of the given histogram below as it would have helped to capture our narrative in a abstract and a manner that easily identifies trends and pattern.

Though a good choice, we did not select this due to following reasons-

* + Its not a good tool to identify and represent temperature uncertainty growth over the years.
  + Any outlier is hard to identify in given histogram.
  + Trends are hard to identify when the values on y axis are large.
  + Can give cluttered look on the application

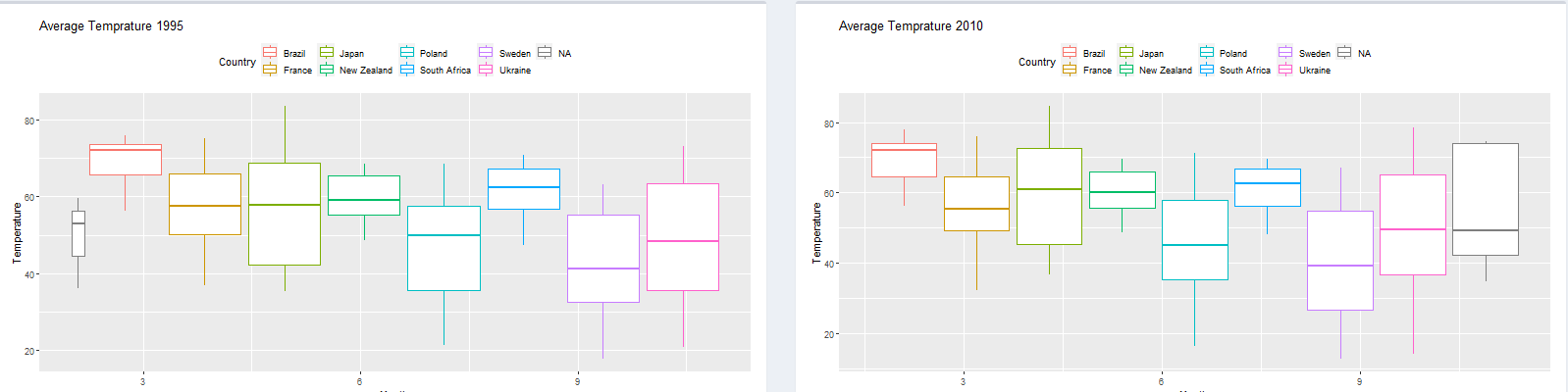


*Figure 6 Histogram*

### Box Plot-

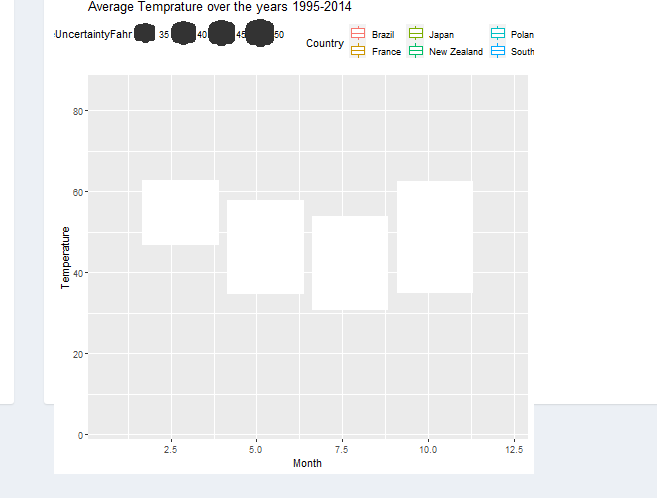
Boxplot comes as the perfect choice for our requirement. Boxplots are mainly used to identify outliers, but if analysed closely then its one of the graphical tools that can represent trends in average value and trend in variance, at the same time. Its clear representation of countries with different colours not only enables the user to have a

look at its own country and have a categorical analysis but also look at the big picture at the same time.



*Figure 7 Box Plot*

This is coupled with animation which describes this trend over time and also identifies outliers, which are marked in “red” because those are recorded data and thus represent a very extreme temperature for a country which should be seen as a sign of danger and conveys the feeling of unease and that is what the narrative is trying to convey at this stage



*Figure 8 Box Plot Animation*

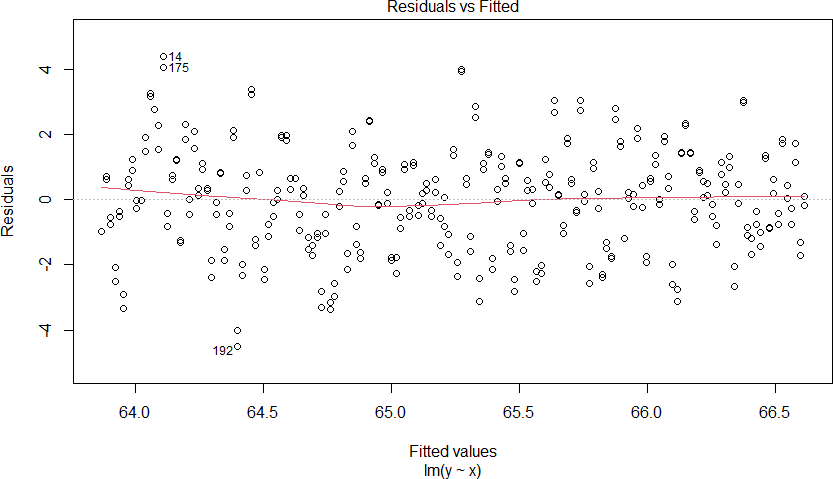
1. Detailed Analysis- This is probably the trickiest part of the design choices because as a data scientist we would like to convey as much information as possible but being in a narrative its important that only that information is used which
   * Conveys the message to each audience member
   * Does not give too much jargon that the narrative is lost
   * Important predications and statistical inferences are conveyed

For this we considered following graphs-

1. Residual vs Fitted –

This graph is very useful when we want to capture how well our trained model is with respect to the large amount of data that we have. Reasons for not selecting this graph are-

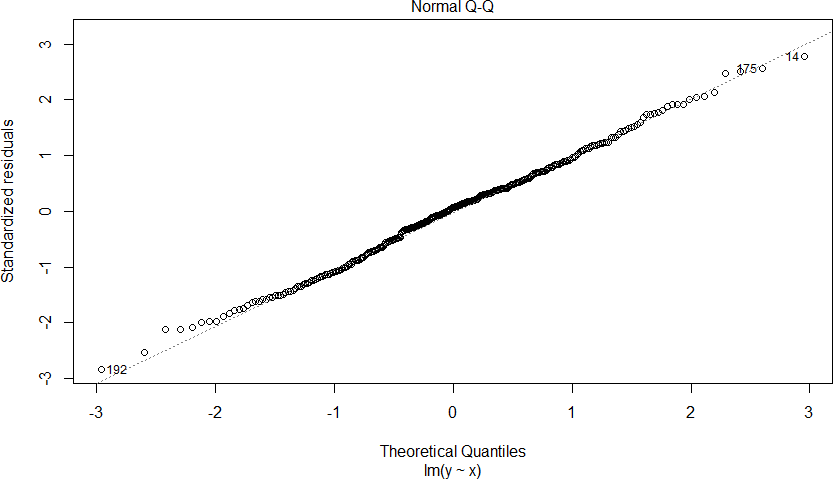
* 1. Does not cater to a wide audience
  2. Not really useful information for the narrative outlined



*Figure 9 Residual vs Fitted*

1. Normalized Graph for standard residual and standard deviation-

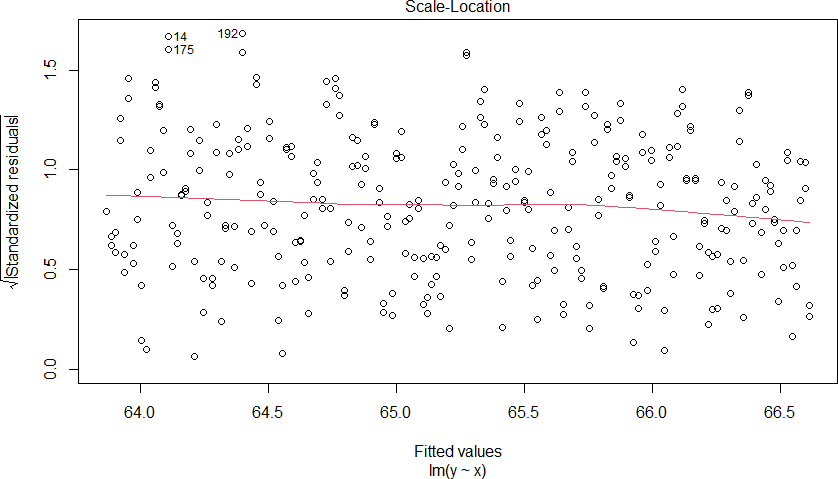
These graphs are very useful when you want to understand your model in detail. The normalized graph helps to scale down all the values and explain the curve and statistical coefficients of the trained model. When plotted against the standard deviation and standard residuals it helps the user understand what is the range of variance of the output ,one can expect out of the trained model.



*Figure 10 Normal Q*

Reasons for not Selecting this graph-

* 1. Does not cater to a wide audience
  2. Not really useful information for the narrative outlined



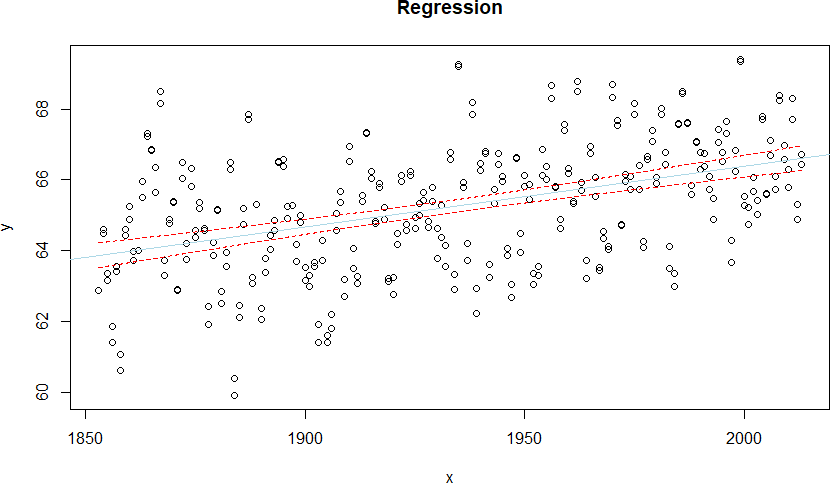
*Figure 11 location vs std deviation*

1. Regression using Plot vs ggplot-

The selected choice of graph to represent detailed analysis of temperature was to use regression analysis and derive confidence interval and prediction values from it.

This is the selected chart since

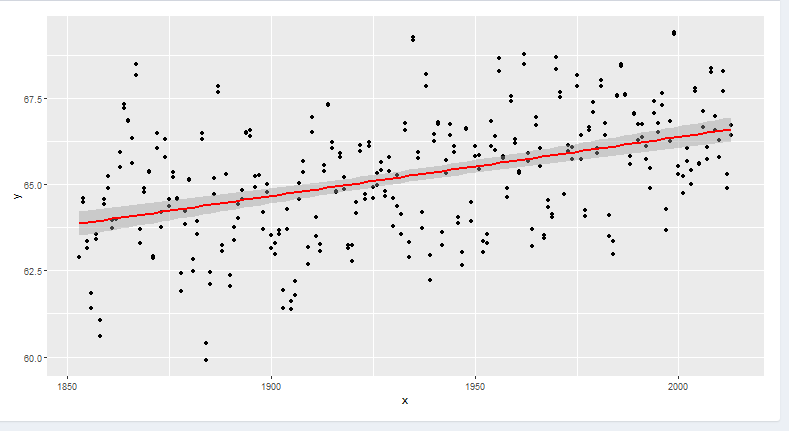
* 1. It is relatively easier to interpret and understand
  2. Highlights the confidence interval of 95% and thus also shows range over values
  3. Predicting future values of temperature is one of the key messages of the narrative



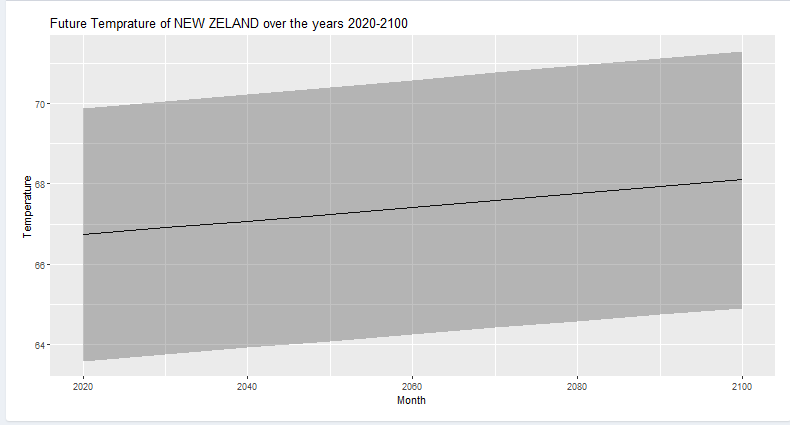
*Figure 12 Regression using Plot*

Another key decision to be taken was, how to plot it?

Use plot or ggplot , and as it is clearly evident form the plots above and below the graph drawn using ggplot is much more visually appealing than the other one.



*Figure 13 Regression using ggplot*



*Figure 14 Predicting future values*

The above graph is used to represent future temperature of New Zeeland with the black line representing the average temprature over the years where as the shaded area represents values between the upper and lower limits of the predicted values. Note we could have used facet wrap and animation here, but the narrative here uses this data as an example rather than explaination .

Thus this graph is the best choice.

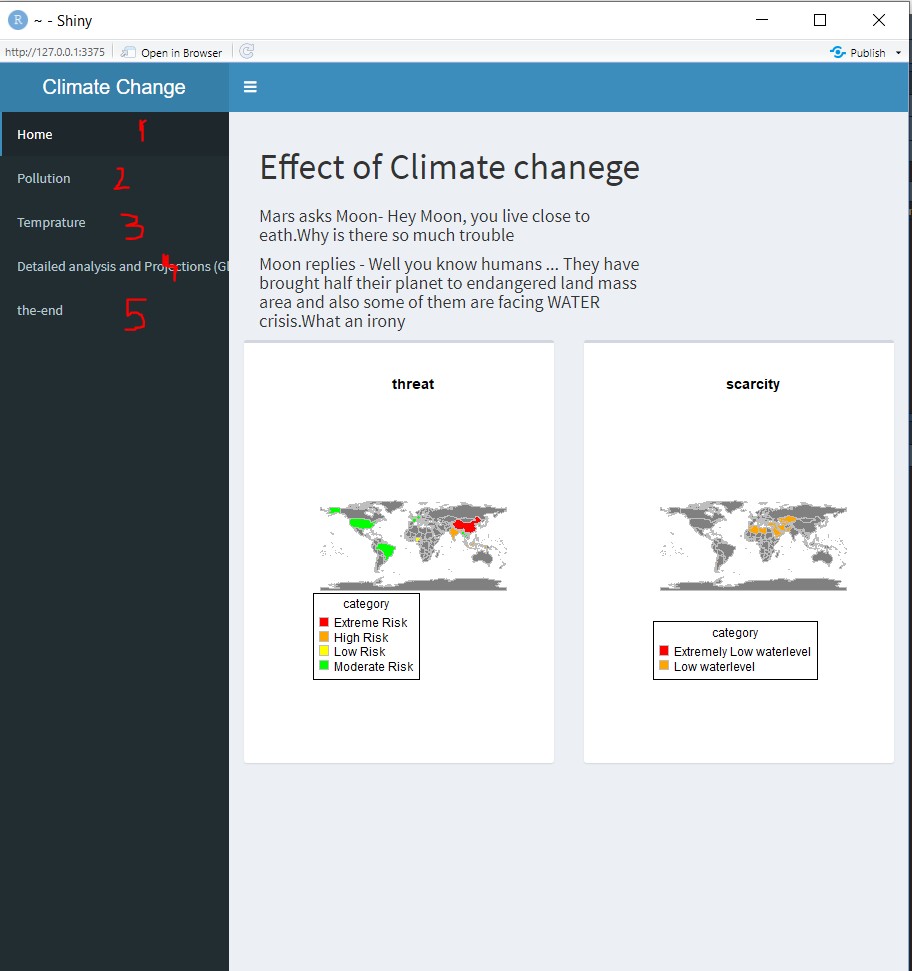
# Implementation

The entire project was implemented in Rstudio using multiple libraries namely Shiny, shinydashboard, ggplot2, devtools, dplyr etc. Steps employed to create the project are

* 1. Create a template of Shiny dash board representing basic architecture of the application
  2. Once the narrative outline is done , create multiple visualisation modulels for each section of the narrative
  3. Select the best design using 5 design sheet methodology and examining which module supports the narrative best
  4. Integrate the selected modules into the Shiny dashboard
  5. Test the complete application

# User guide:

As mentioned multiple times above this app focuses on telling a fictional conversation thus in-order to follow along with the converstion the graphical representaitons needed to be abstact and detaileld at the same time.

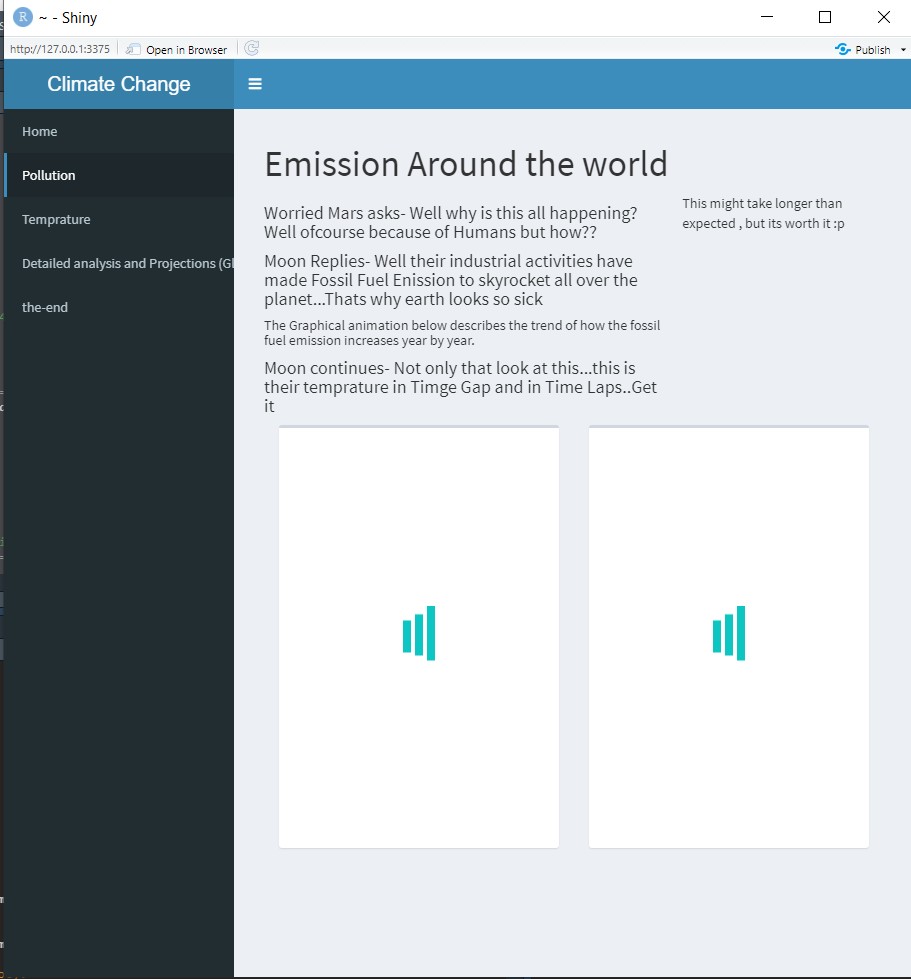


*Figure 15 App page1*

As shown in the above figure , the tabs of the application are marked in the numerical order in order to retain the flow of the coneresation.

It is important to note that since our application uses animation of latest version it can take some time to load the application, thus when you click a certain tab it will show a loading page.

(Recommendation- In order to have the best user experience click all the tabs once before starting from the first page, this starts the rendering process for all the animations and graphs as they are only invoked when tabs are clicked)

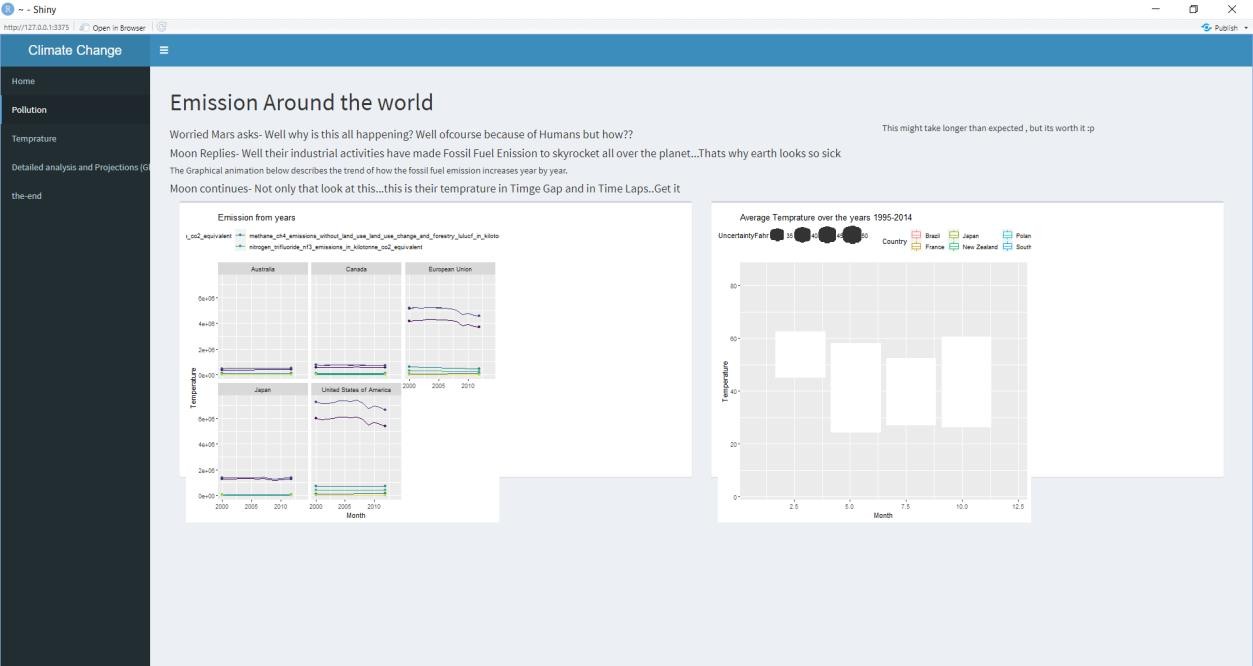


*Figure 16 App page2*

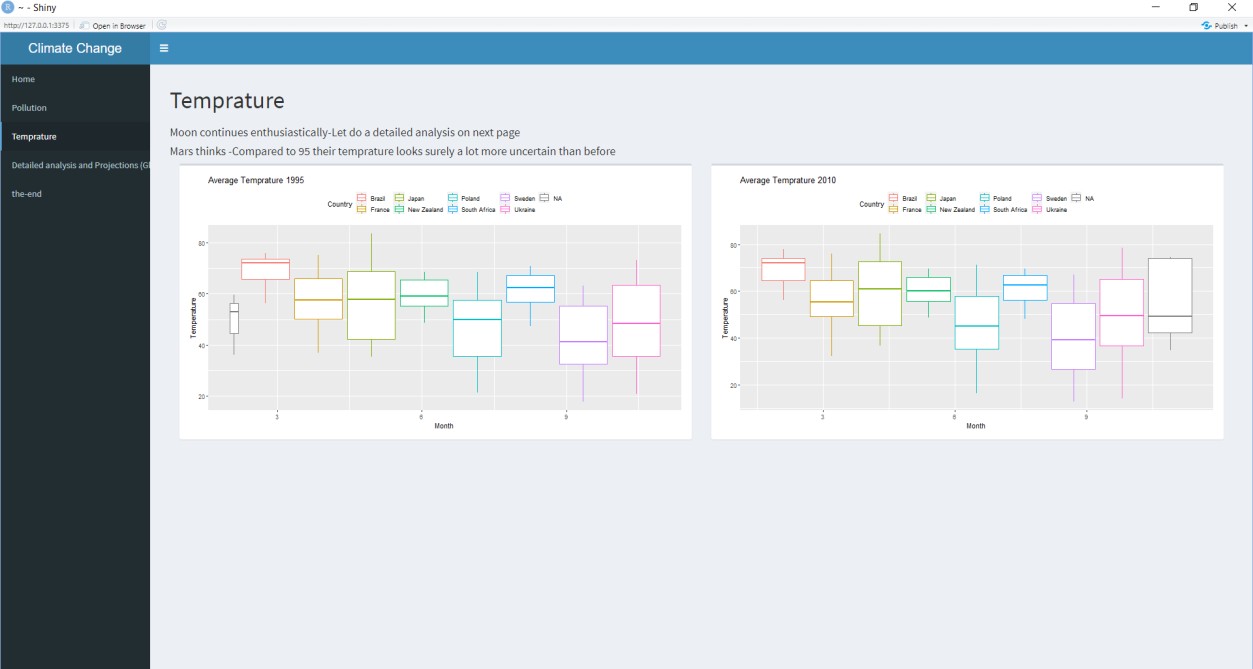
The application limits user interaction interms of choosing which plot or image to look at or which parameters should be chosen, since it’s a conversation and explanations and examples are given from moon to mars. So as a third person, we really don’t have a say.

First animation of second tab describes how fossil fuel emissions are increasing increasing all over the globe and second animation shows its direct consequence on different countries in terms of temprature rise and uncertainity.

The third tab explains the before shown animation by capturing the static states for two different years with large time gap

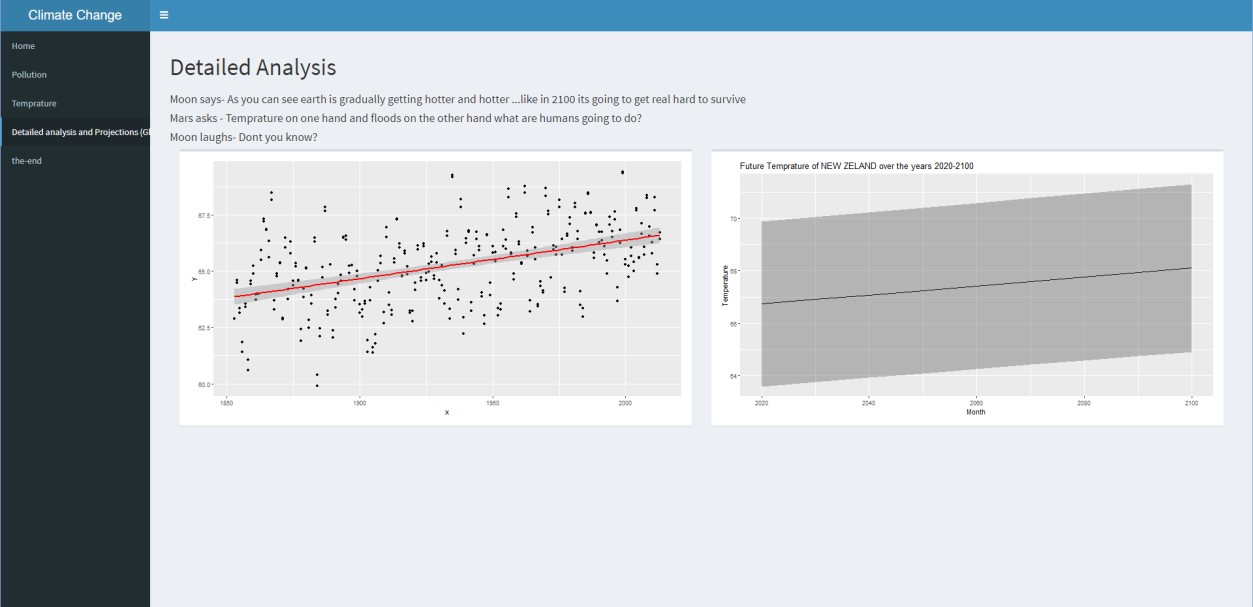


*Figure 17 App page 2*



*Figure 18 App Image 3*

Detailed analysis tab sums up the discussion by comparing how bad the present scenarios are and how bad the future can be if the current trends are not altered in a large magnitude and finally the last tab sums up the conversation at a lighter note (sarcastic joke).



*Figure 19 App image 4*

# Conclusion:

### Achievements-

* A Message which requires a global attention is conveyed with the right tone and right factual visualization
* Using shiny to create a web application improved my R skills two fold.
* Add animation to support a fictional narrative
* Creating and choosing correct visual representation that takes care of not only the technical aspect but also the creative aspect of the project

### Reflection-

* Visual representation for narrative could have been optimized
* Use of cartoon images could have made the application much more appealing
* Many other aspects of climate change could have been covered
* Map could have been more interactive

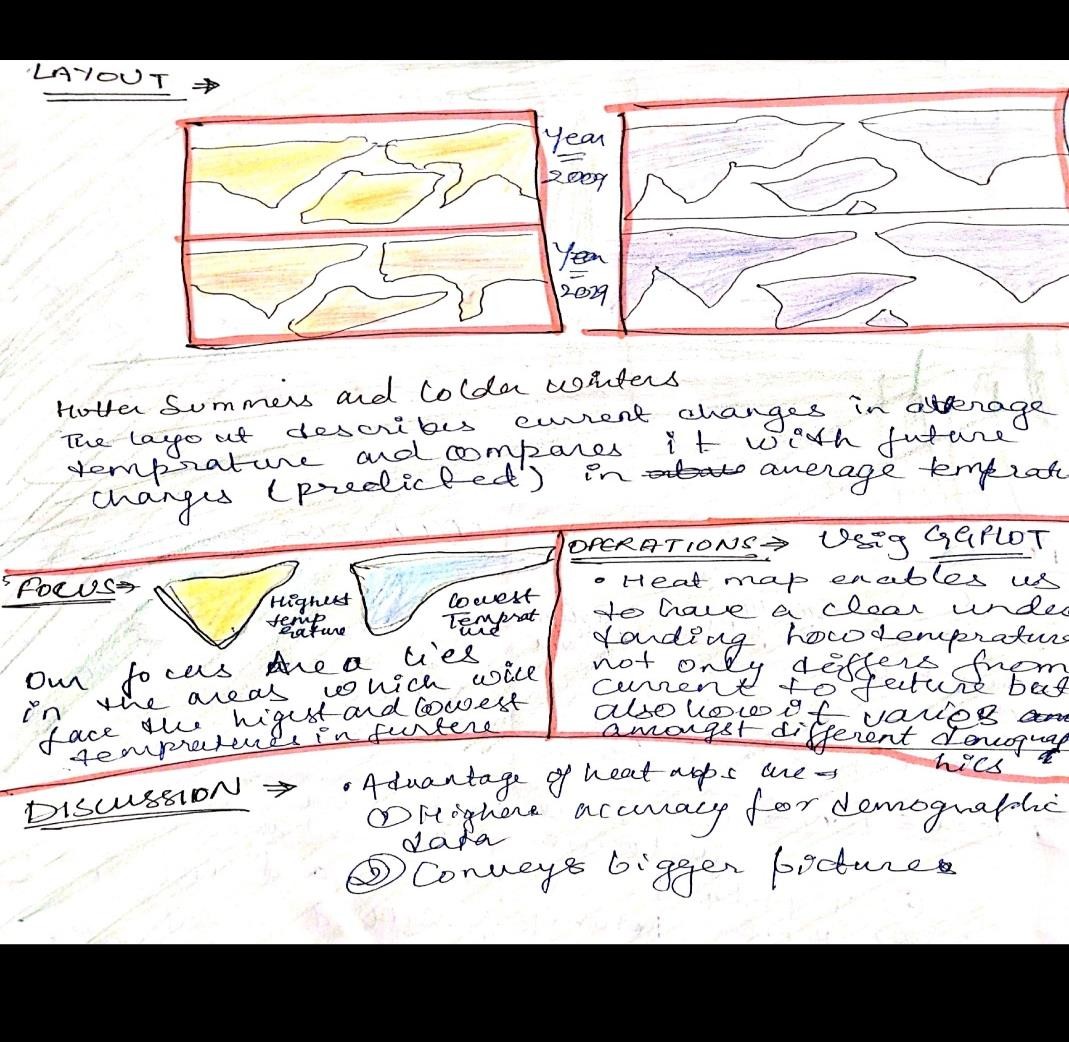
# Bibliography:

1. Generic way to plot a graph with establishing X and Y axis in R. Retrieved on 28 April 2020.From https://astrostatistics.psu.edu/su07/R/html/graphics/html/plot.html
2. Different type of line charts in R. Retrieved on 28 April 2020.From https://[www.statmethods.net/graphs/line.html](http://www.statmethods.net/graphs/line.html)
3. How to add column in a dataframe based on a condition. Retrieved on 28 April 2020.From https://stackoverflow.com/questions/15016723/how-to-add-column-into- a-dataframe-based-oncondition
4. Helsinki Social Statistics Course on Data Camp. Retrieved on 28 April 2020.From https://[www.datacamp.com/courses/helsinki-social-statistics](http://www.datacamp.com/courses/helsinki-social-statistics)
5. Accessing rows and columns of a dataframe in multiple ways. Retrieved on 28 April 2020.From https://pythonhow.com/accessing-dataframe-columns-rows-and- cells/
6. Countries under threat of flooding and water scarcity. Retrived 20 June 2020. From [https://www.wri.org/blog/2015/03/world-s-15-countries-most-people-exposed- river-floods](https://www.wri.org/blog/2015/03/world-s-15-countries-most-people-exposed-river-floods)
7. China under great threat of flooding. Retrieved 20 June 2020. From

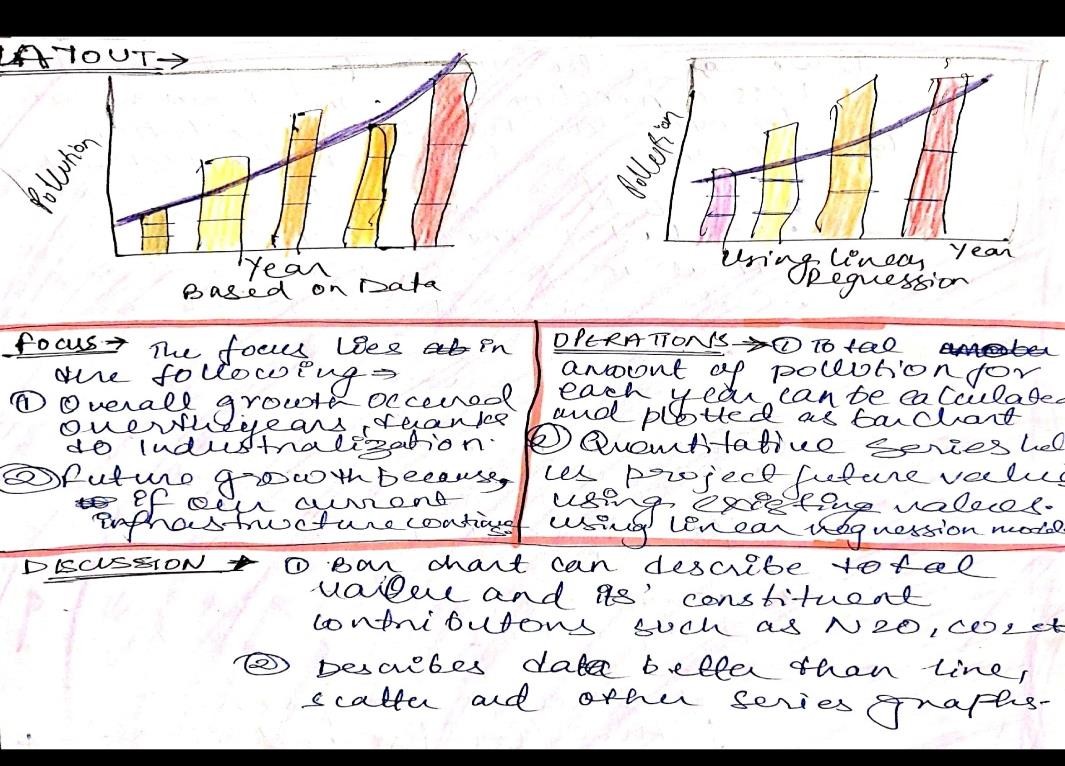
[https://www.carbonbrief.org/china-tops-new-list-of-countries-most-at-risk-from-](https://www.carbonbrief.org/china-tops-new-list-of-countries-most-at-risk-from-coastal-flooding) [coastal-flooding](https://www.carbonbrief.org/china-tops-new-list-of-countries-most-at-risk-from-coastal-flooding)

# 7.Appendix:

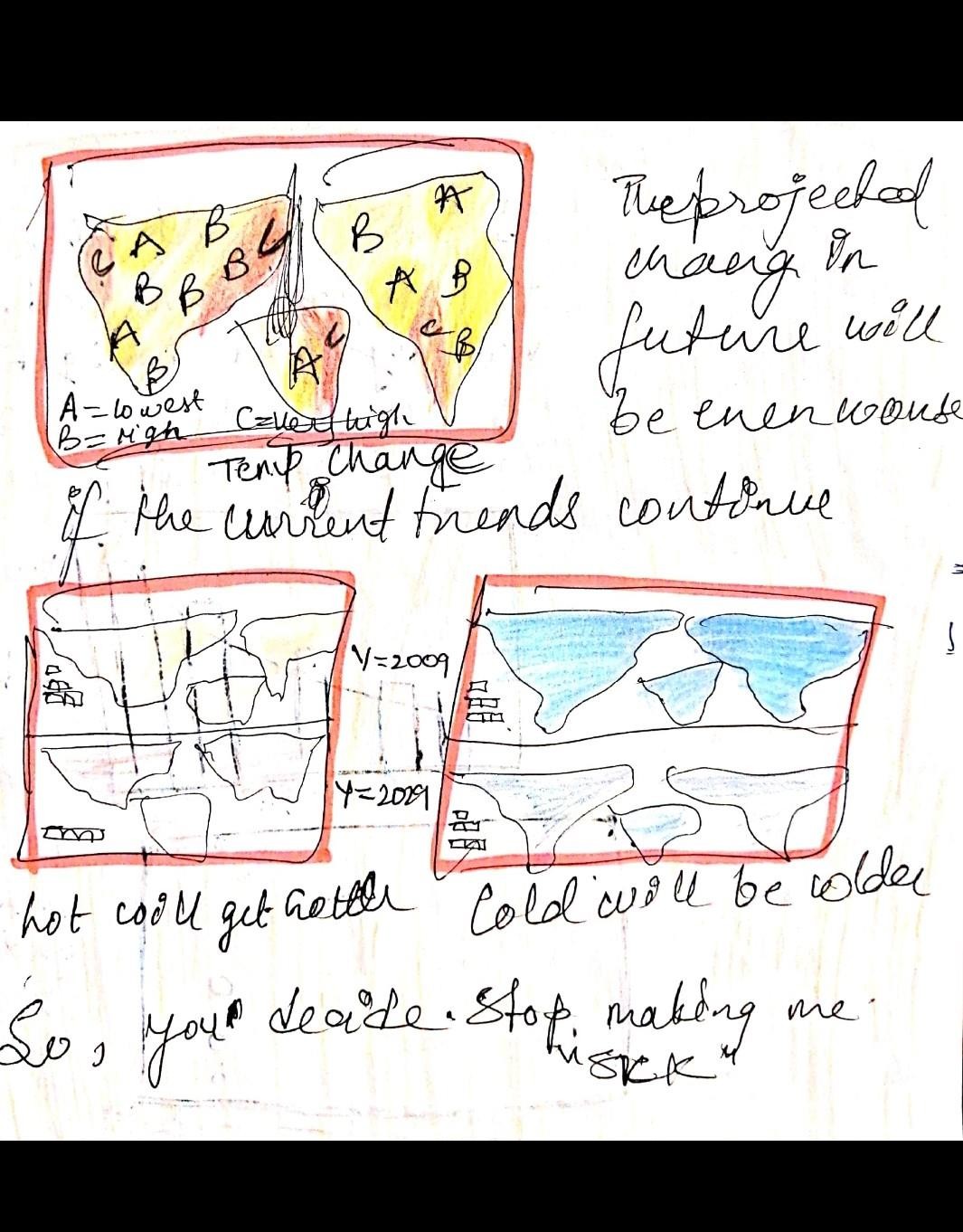
*Figure 20 Appendix 1*



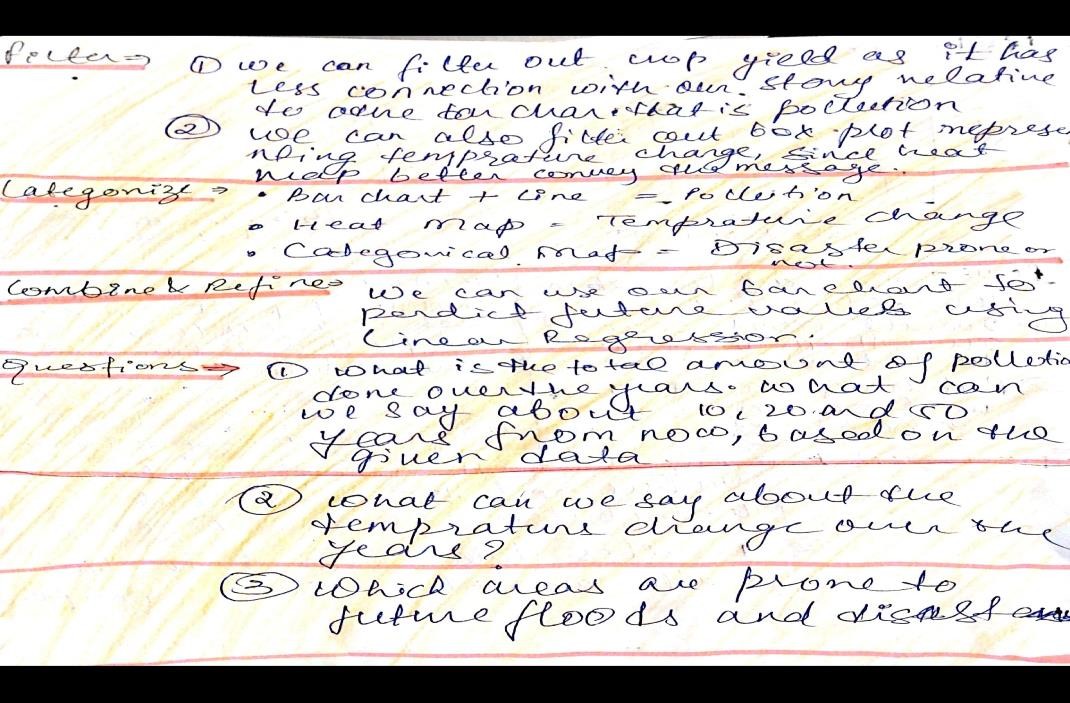
*Figure 21 Appendix 2*



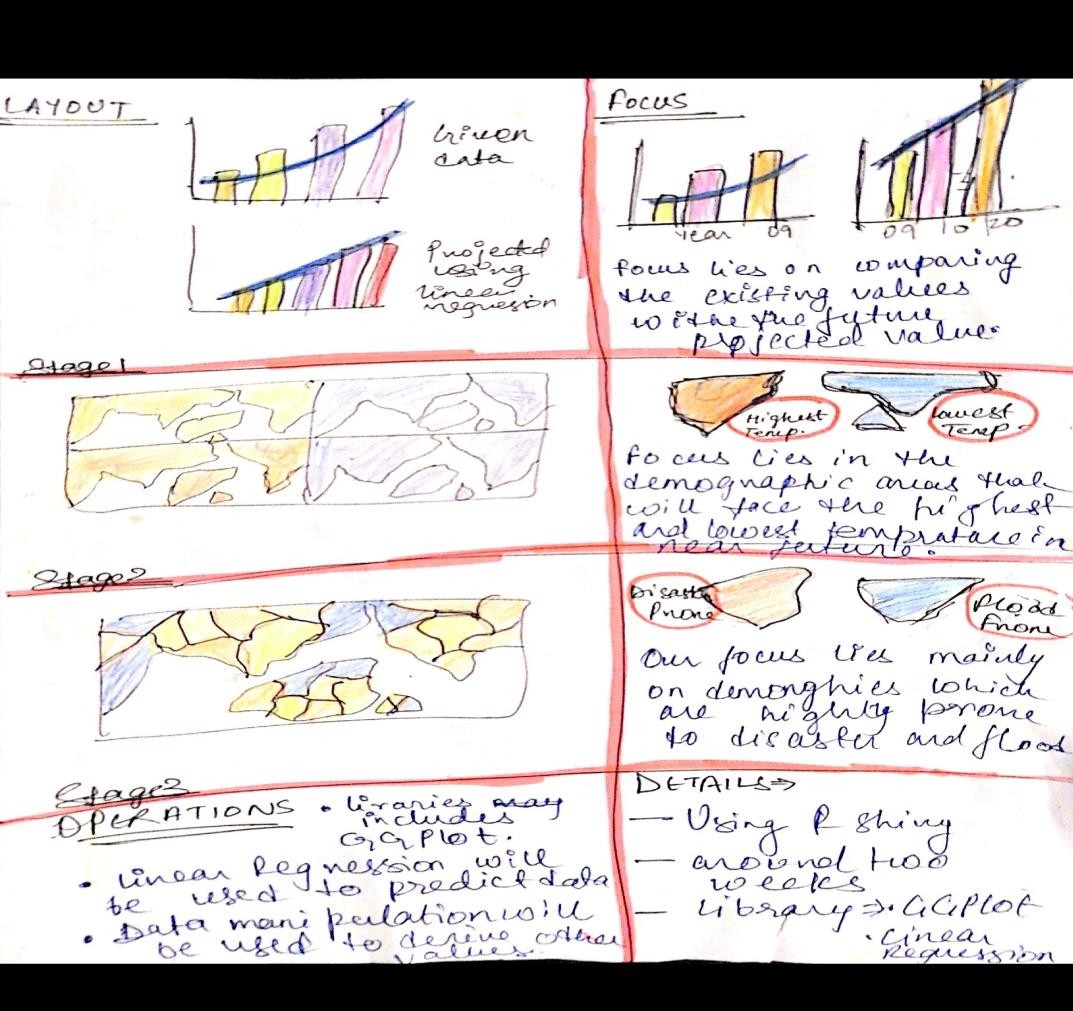
*Figure 22 Appendix 3*



*Figure 23 Appendix 4*



*Figure 24 Appendix 5*



*Figure 25 Appendix 6*