**Assessment Cover Sheet**

This Assessment Cover Sheet is only to be attached to

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ASSESSMENT DETAILS** | | | | | | |
| Unit title | | Data Visualisation | Tutorial /Lab Group | 1 | | Office use only |
| Unit code | | COS30045 | Due date | 30 Nov 2020 | |  |
| Name of lecturer/tutor | | Vong Wan Tze, Lee Sue Han | | | |  |
| Assignment title | | Visualisation Project | | | | Faculty or school date stamp |
| **STUDENT(S) DETAILS** | | | | | | |
| Student Name(s) | | | | | Student ID Number(s) | |
| (1) | Jonathan Lee Wei Han | | | | 101218655 | |

hard copy submission of assessments.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DECLARATION AND STATEMENT OF AUTHORSHIP** | | | | |
| 1. I/we have not impersonated or allowed myself/ourselves to be impersonated by any person for the purposes of this assessment. 2. This assessment is my/our original work and no part of it has been copied from any other source except where due acknowledgement is made. 3. No part of this assessment has been written for me/us by any other person except where such collaboration has been authorised by the lecturer/tutor concerned. 4. I/we have not previously submitted this work for this or any other course/unit. 5. I/we give permission for my/our assessment response to be reproduced, communicated, compared and archived for plagiarism detection, benchmarking or educational purposes.   I/we understand that:   1. Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to exclusion from the University. Plagiarised material can be drawn from, and presented in, written, graphic and visual form, including electronic data and oral presentations. Plagiarism occurs when the origin of the material used is not appropriately cited. | | | |
| **Student signature/s**  I/we declare that I/we have read and understood the declaration and statement of authorship. | | | |
| (1) |  |  |

Further information relating to the penalties for plagiarism, which range from a formal caution to expulsion from the University is contained on the Current Students website at **https://www.swinburne.edu.my/current-students/manage-course/exams-results-assessment**

Copies of this form can be downloaded from the Student Forms web page at **https://www.swinburne.edu.my/current-students/manage-course/exams-results-assessment/how-to-submit-work.php**

COS30045

Data Visualisation

Semester 2, 2020

Title

Overpopulation

|  |  |  |
| --- | --- | --- |
| Assessment | : | Visualisation Project |
| Name | : | Jonathan Lee Wei Han |
| Student ID | : | 101218655 |
|  |  |  |

**Table of Contents**

[1. Introduction 4](#_Toc50507401)

[1.1. Background & motivation 4](#_Toc50507402)

[1.2. Project objectives 4](#_Toc50507403)

[1.3. Project schedule 4](#_Toc50507404)

[2. Data 5](#_Toc50507405)

[2.1. Data source 5](#_Toc50507406)

[2.2. Data processing 5](#_Toc50507407)

[3. Requirements 6](#_Toc50507408)

[3.1. Must-have features 6](#_Toc50507409)

[3.2. Optional features 6](#_Toc50507410)

[4. Visualisation design 7](#_Toc50507411)

[5. Conclusion 8](#_Toc50507412)

[Reference 9](#_Toc50507413)

[Appendix 10](#_Toc50507414)

# Introduction

# Background & motivation

As we grow older, we tend to overlook things as they are called as a new norm in the 21st century. The topic on Environment in crisis has always been in the back of my mind especially the increased number of the human population. What came as a shock was in the year 1800, the world population was recorded to be populated with only 1 billion people. It may sound like a lot of people. However, it took around 2 centuries, that the population now is estimated to be around 6 billion people. China to date is the most populated country in the world with over 1 billion people living in the country, which is still increasing day by day not to mention other remaining countries.



Figure 1: World Population

According to (Mittal, R, Mittal, C, 2013), the environment has taken a massive impact such as emission of greenhouse gases, increase in soil erosion and the declining of animal population, which is corelates to the explosion in human population increase through the years. By selecting this topic for the data visualisation project, I intend to raise an awareness to all age groups on the crisis at hand and make a deeper change on the world as it is a quick and powerful tool to overview data. For the project, it will mainly display various kinds of data charts that will represent the topic on Environment crisis, Overpopulation in the world.



Figure 2: Greenhouse Gasses

Figure 3: Deforestation

# Project objectives

The primary objective of the project is to provide insight on the environmental crisis that is being faced by overpopulation. Accurate and up-to date data are used in projecting the project visualisation. To be more unique, the visualisation that is going to be produced will have emphasis on the years. In different years, the contribution of numbers of environmental crisis will vary between years. The number of animals killed, gasses emitted, and number of trees cut will be changed through the years.

The data visualisation will help the user and me to answer some primary questions including:

* What environmental issues might occur with an increase of population size?
* Does overpopulation cause air pollution?
* Does overpopulation lead to climate change?
* Does overpopulation lead to animal extinction?
* Does overpopulation cause deforestation?
* Does overpopulation lead to increase in greenhouse gas emission?

I believe that the answers to these questions beneficial as an awareness on the current environmental state. It will be an eye opener to the public as it is about the current issues regarding the environment situation, more importantly the state of the earth. For example, knowing about the amount of air pollution emitted every year, authorities would take actions in reducing the amount of air pollutants from being emitted such as reducing the number of cars built and factories made. For a better visualisation, the D3 tool is learnt as it is a tool to create interactive charts to illustrate the data more easily to be understood by users.

# Project schedule

|  |  |
| --- | --- |
| Week | Plan |
| 1-3 | Learn and get started with D3 |
| 3-4 | Research was conducted in finding relevant topics to the project theme “Environment in Crisis’.  In addition, data sets are discovered that would allow the user to do some interactivity. |
| 5 | Proposal for the project is due on Monday September 28th for further confirmation from the lecturer to proceed with the project. |
| 6-7 | Restructure data and proposal based on the feedback that was given.  Research was conducted to find more relevant topics regarding the project theme. |
| 8-10 | Begin the visualisation from creating the story  Creating D3 visualisations to answer the questions  Conduct a project consultation and fix some design problems. |
| 11-12 | Finishing up remaining visualisation |
| 13 | Finish up the final project report |

# Data

# Data source

The datasets for this visualisation will consists of six different datasets. The first dataset includes the data of global population from 1950 to 2019. The second dataset is the Air pollution dataset which has the data of the yearly global CO2 emission. The third dataset includes the animal extinction database of number of animals killed in the world every year. The fourth and fifth dataset regarding the global deforestation from the area of forest. The last database includes the global anomaly of temperature every year.

## 2.1.1 World Population Dataset

The world population database includes the global population from different researchers. The dataset consists of yearly data from as far as 1 B.C.E. The dataset is given by Our World In Data organization and it is free to use and download [here](https://ourworldindata.org/grapher/world-population-since-10000-bce-ourworldindata-series?tab=table&time=1961..2019).

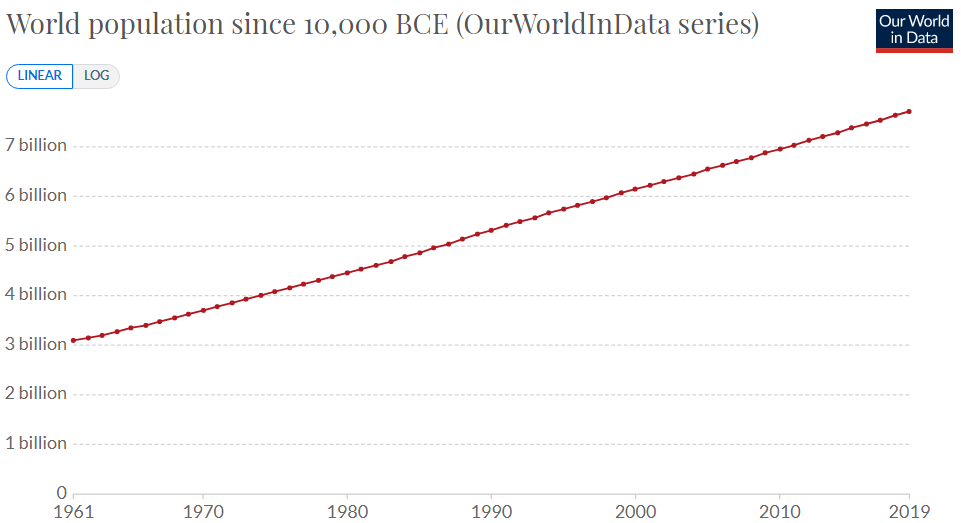


Figure 4: Example of World Population Chart

## 2.1.2 Air Pollution Dataset

The Air pollution database includes the atmospheric CO2 concentration of the world. The dataset consists of yearly data from as far as 803,719 B.C.E. The dataset is given by Our World In Data organization and it is free to use and download [Air pollution Database](https://ourworldindata.org/grapher/co2-concentration-long-term?time=earliest..2018).

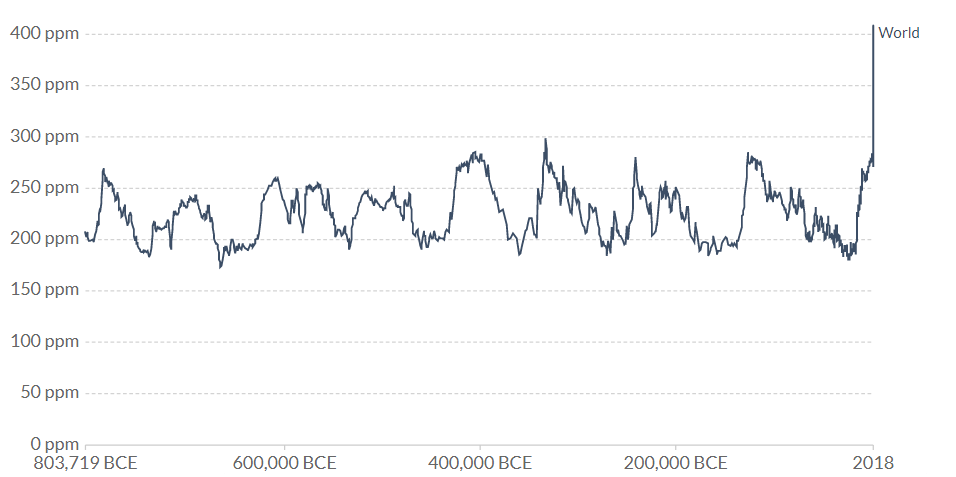


Figure 5: Example of Air pollution Chart

## 2.1.3 Animal Extinction Dataset

The Animal Extinction database includes all the animals killed every year by category such as chicken, pigs, turkey, sheep, goats and Cattle. The dataset consists of yearly data from 1961. The dataset is given by Our World In Data organization and it is free to use and download [Animal Extinction Database](https://ourworldindata.org/grapher/animals-slaughtered-for-meat?time=earliest..2018).

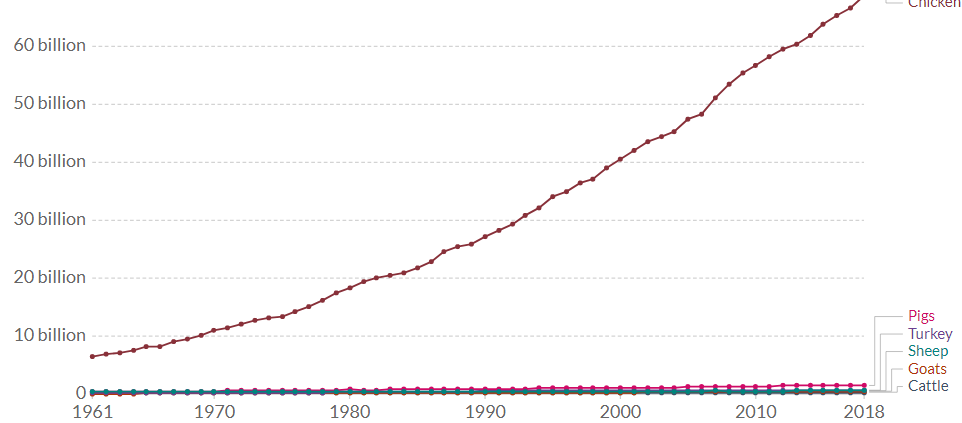


Figure 6: Example of Animal Extinction database

## 2.1.4 Deforestation Dataset

The Deforestation database includes the amount of area the forest is covered in the world or by countries. The dataset consists of yearly data from 1990. The dataset is given by Our World In Data organization and it is free to use and download [Deforestation Database](https://ourworldindata.org/grapher/forest-area-as-share-of-land-area?tab=table&time=1880..latest&region=World).

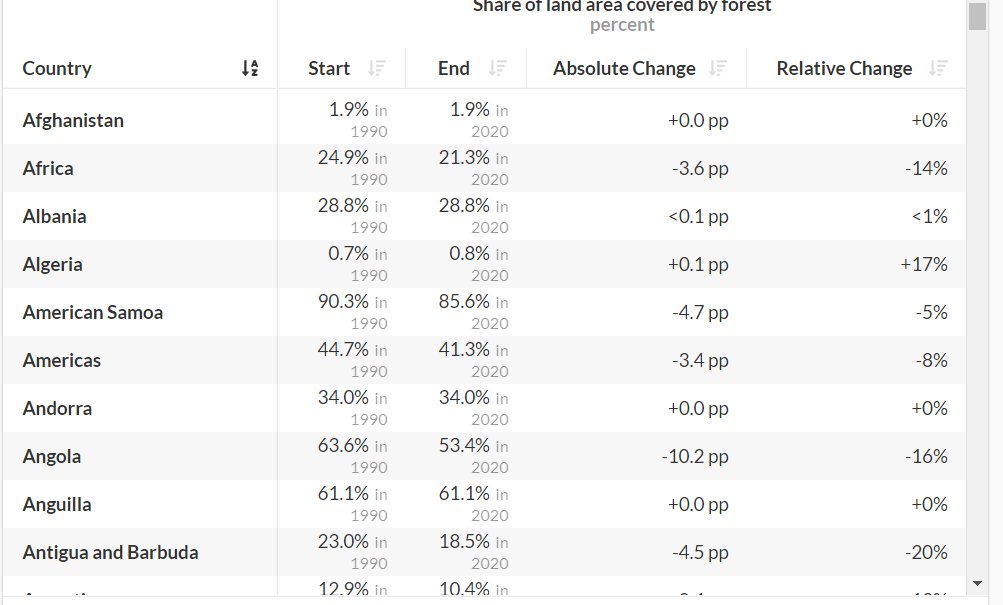


Figure 7: Example of Deforestation database

## 2.1.5 Anomaly temperature Dataset

The Greenhouse gasses database includes the temperature from upper, lower and median Celsius of Global, northern hemisphere, southern hemisphere and the tropics. The dataset consists of yearly data from 1850. The dataset is given by Our World In Data organization and it is free to use and download [Greenhouse gasses](https://ourworldindata.org/grapher/temperature-anomaly?time=1926..latest).

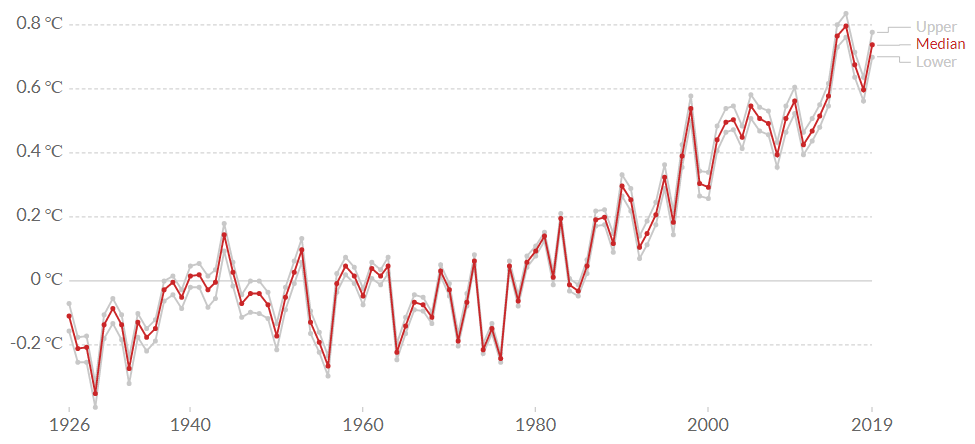


Figure 8: Example of Greenhouse gasses chart

# Data processing

## 2.2.1 Handling Missing Data

Data gathered from the dataset described, there may missing data values from the dataset. Databases that have missing data values will be given a 0 value as you can see from the figure below. This is due to the visualisation having to not show up on the graph/chart. Missing data is a null value.

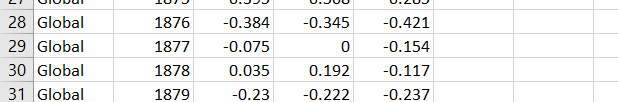


Figure 9: Example of missing value

## 2.2.2 Data splitting

The databases that was stated in 2.1 is collected and separated into different csv files to be combined with the world population database for a comparison of values. This would make the D3 to be easier to import by code. We must make sure that the values are corelated to the date of the world population for a better visualisation. After combining, there will be 6 csv files to be used in the d3 visualisation as shown below: -

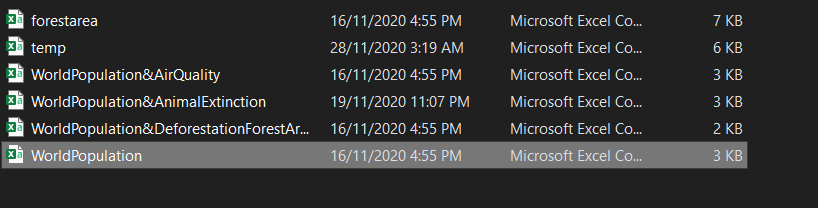


Figure 10: Csv files

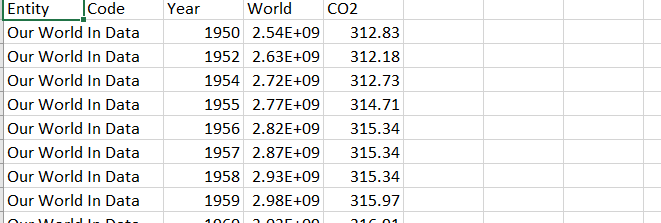


Figure 11: Csv content

# Requirements

# Must-have features

The final visualisation must follow the must have features in the project to be considered passed in the project.

## 3.1.1 Visual Element

* Majority of the screen must be occupied by the chart.
* Title should be accurate and follow the data described for user to understand easier.
* Axes, Scales and Labels are to be accurate and rescaled when the data ranged have been changed.
* Legends must be appropriate and accurate with the data.

## 3.1.2 Interactive Features

* Users should be able to start an animation of charts.
* Allow users to display 2 charts in 1 page
* Allow the user to have a tooltip to see the value of the chart.

## 3.1.3 Technical Requirement

* Cross platform is to be possible as it can be viewed in different web-browsers and mobile.

In the final visualisation, all of the interactive features is able to be completed, visual elements and it meets the technical requirements.

# Optional features

* User will have an option to categorise the chart more in-depth by choosing exclusively to a part of country.
* Charts can be zoomed to be more detailed.

In the final visualisation, the optional features were unable to be developed, due to the time constraint it was not possible to complete the optional features in time.

# Design Alternatives

## 4.1 Alternative 1: Bar charts

### 4.1.1 Visual encoding

For the first design sketch, a linear bar chart data is presented in a rectilinear layout. The marks that was used are one qualitative data and one quantitative data. The qualitative data used in the visualisation is the period the data was recorded. As for the quantitative data, the number of % the recorded population is represented in a vertical spatial position channel. The qualitative data for the x-axis is an Identity type channel. Number of populations is suitable for a magnitude type channel.

From the figure below, the data used in indicating the period is using a Hue channel to interpret solid colours on different periods. Based on the colour guidelines, saturation channels or luminance channels are more suitable to be used for magnitude channel and hue is more used to describe categories. The data follows the correct colour guidelines as it uses Hue channel which is more used in categories.

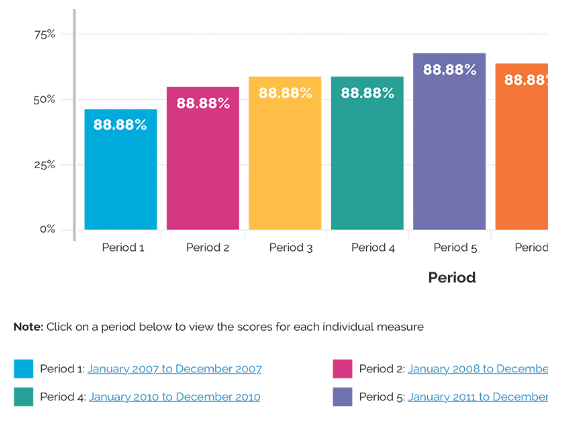


Figure 12: Alternative bar chart example

The chart would have an animation of each bar increasing from period 1 to the next when the page is being loaded. When the animation is finished, the user can add more periods of time by 1 period. The users are also able to hover on the bar chart rectangle by the use mouseover tooltip to display the value of the specific bar.

### 4.1.2 Design Critique

The main advantage of implementing the graph bar chart is that it ‘s ability to summarize a large data set into visual form. No matter how large the dataset, graph bar can create a visualisation that is simple for users to understand by using the quantitative data of % population. Besides that, the graph does not utilize a lot of ink, when more data is added, the bar chart uses hue channel to visualise the categorical data.

The chart has a big disadvantage which is the inability to be visible when there are a lot of bar graphs in a single chart. When the user keeps adding periods to the chart, the visibility of the bar graph is too thin and could not be seen. Other than that, the chart is unable to make a comparison between two categories. This means, the graph is only effective in one dimension not in two dimensions.

## 4.2 Alternative 2: Linear Line Chart

### 4.2.1 Visual Encoding

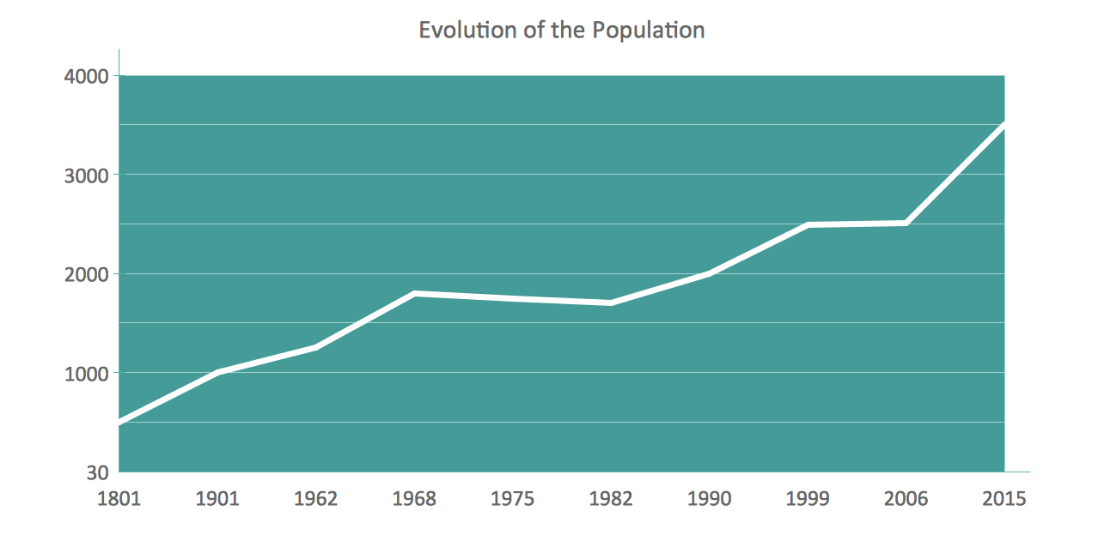
For the first design sketch, a linear line chart data is presented in a rectilinear layout. The x-axis of the line chart is encoded using the horizontal spatial position with linear scale. The y-axis of the line chart encodes the number of the world’s population. The higher the position of the line, it indicates the increase of the world population, where the lower the line indicates the decline of human population. The points are then connected to each other to form a line which will show the trend throughout the year (Increasing or Decreasing)

Figure 13: Alternative line chart example

### 4.2.2 Design Critique

The line chart is great in indicating the increasing and decreasing values of the data. The users are able to understand from the graph that whether the world population is increasing or declining. The data is readable due to the points of the data being connected to each other and forming into a line. It is a perfect chart to be used to indicate the amount of increase or decline in the human population from 100 years ago.

There are downsides to implementing this graph, when there are multiple categories plotted into the graph, it will be hard to interpret the chart as there are multiple lines in the graph which will cause clustering and would confuse the user in understanding the graph.

## 4.3 Final Designs

After explaining some of the different designs, the final visualisation will consist of multiple charts being some mentioned before and some not mentioned. The charts that would be used are Multi Line Charts, Linear Line Chart, Choropleth Map and Dual Axes Line Charts.

### 4.3.1 Line Chart Iterations

The implementation code of the line chart is a d3 line chart that could be found online. A line chart is used for the final visualisation because I’ve found that it is easier to express the change of population through the years. The number of populations in the world is a type of quantitative data, which is appropriate for the use of Line charts. The line of the chart uses a colour channel with red line. For interactivity, the line chart will have a start time animation button which allows the user to view the line chart from each year until the current year. After the animation is completed, a reset button can reset back the transition. A tooltip using a mouseover d3 function is also available to see the value of the line chart on a certain year.

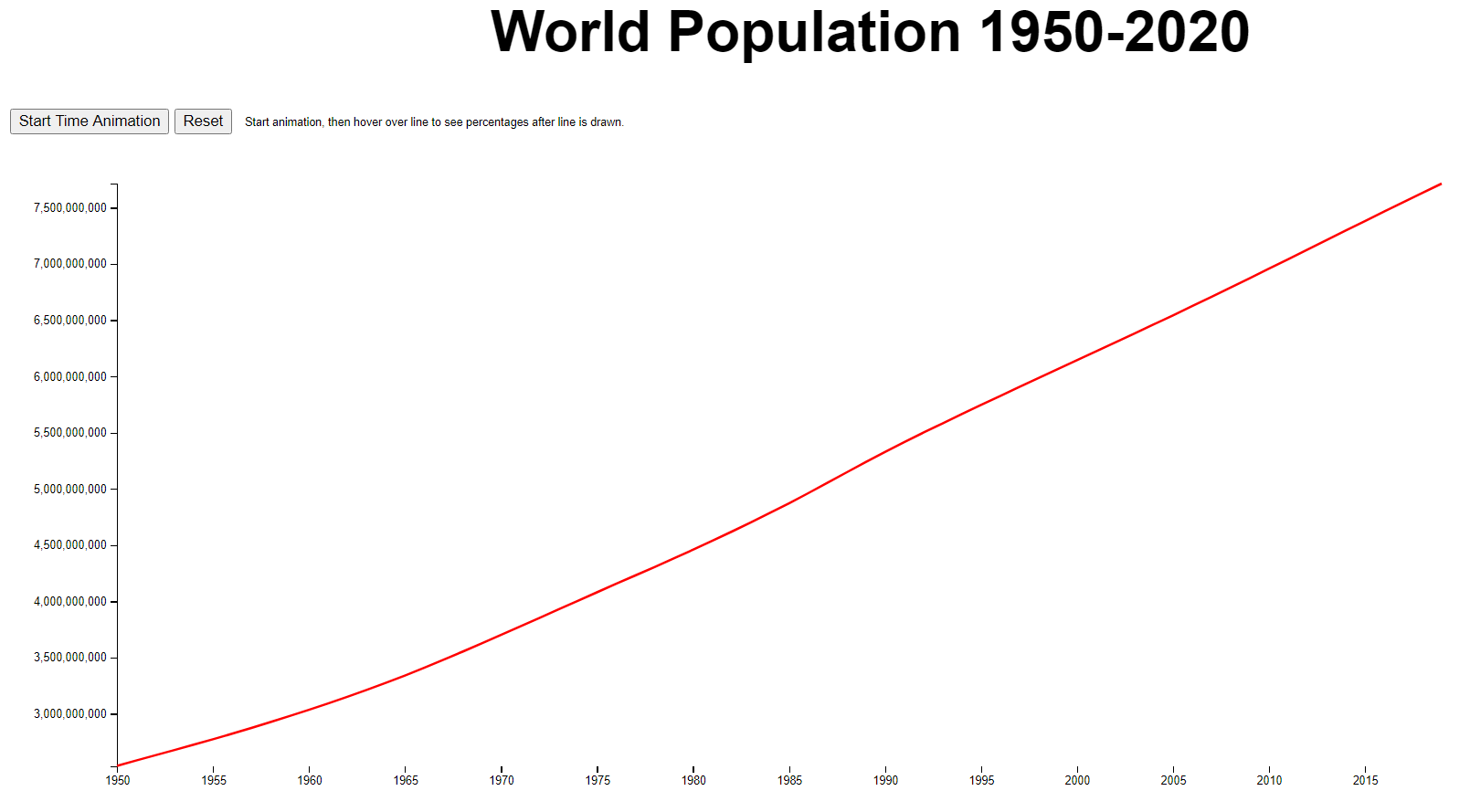


Figure 14: Final Line Chart Visualisation

### 4.3.2 Double Axes Multi Line Chart Iterations

For the second visualisation, a multi-line chart is used to indicate the relationship between the world population and the amount of Carbon dioxide released in the air. The purpose of this chart is to indicate whether the increase of the CO2 emitted is due to the increase amount of the human population. A multi-line chart is appropriately used in the project is due to the data being both world population and CO2 emitted is a quantitative data. Both lines are using a colour hue which are red indicating the world population and blue indicating the CO2 emission. A legend is created to avoid any confusion of the user in understanding which line represent what data. The left y-axis is coloured in blue to indicate it belongs to the world population and the right y-axis is indicating that it belongs to the CO2 emission data. A similar graph was plotted for the deforestation data chart which indicates the number of forest areas in %.

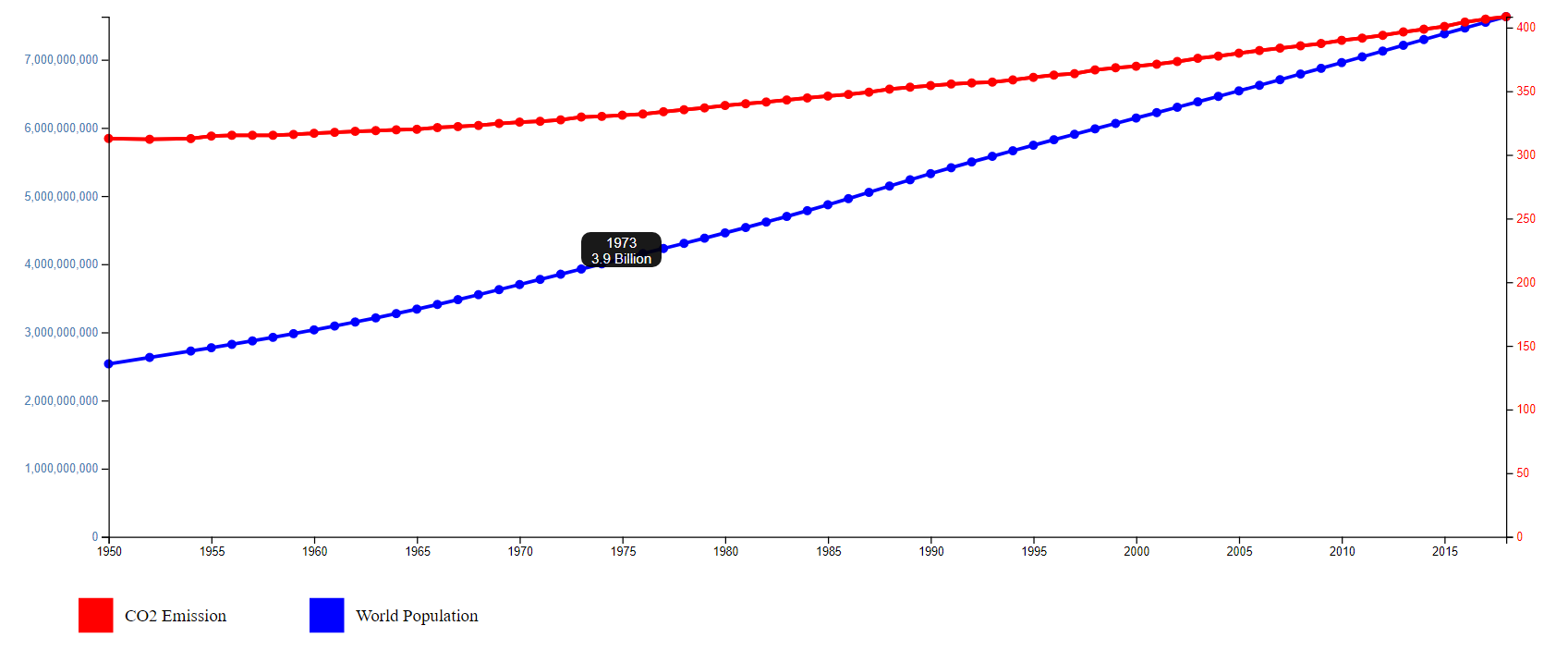


Figure 15: Final Double Axes Multi Line Chart

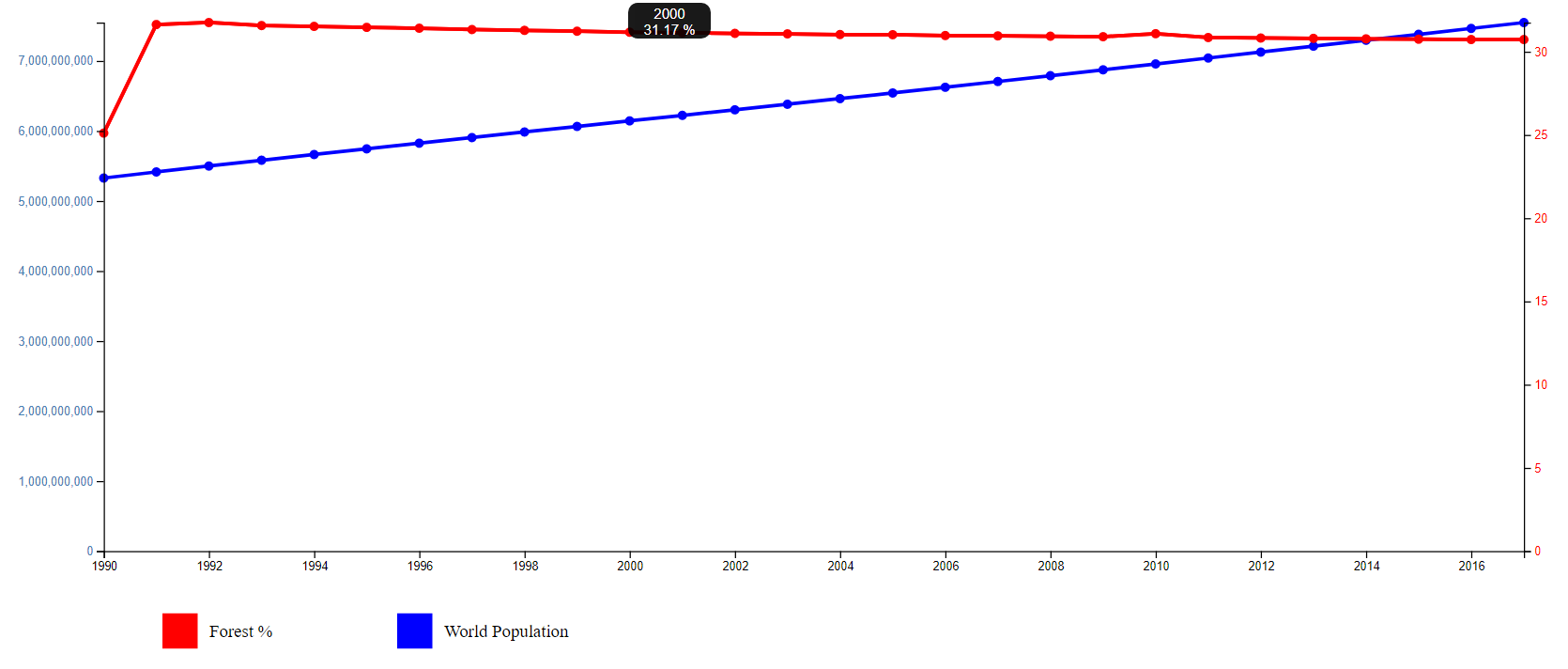
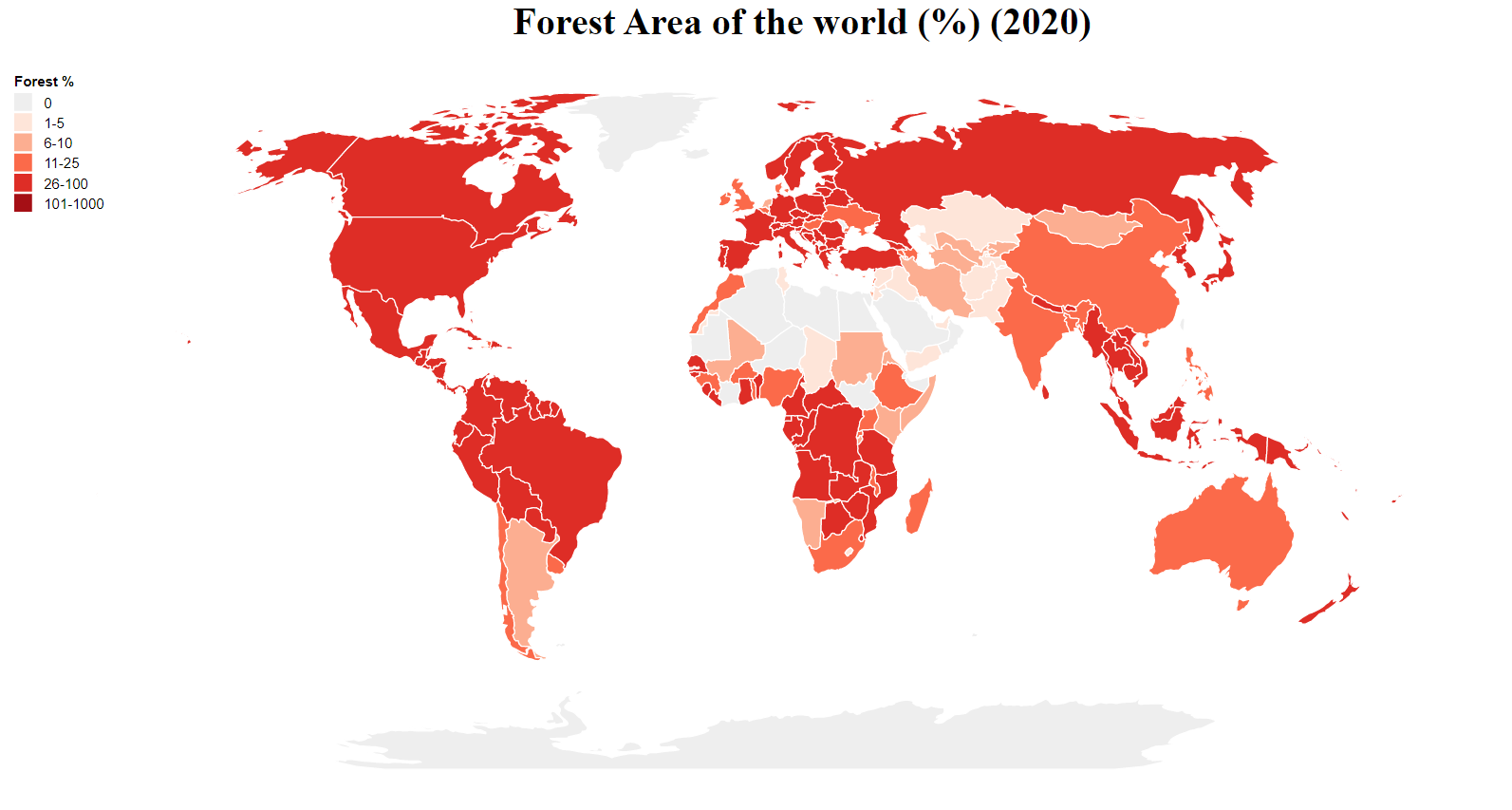


Figure 16: Final Double Axes Multi-line chart

### 4.3.3 Choropleth Map Chart Iterations

The third visualisation is the choropleth map using d3 encoding that can be found online. It is the most suitable chart for this kind of data as it represents the whole world. In terms of colour channel, the chart uses a saturation colour which is red from light to dark shades. The shades of red indicate the % of forest in a certain part of the world. The brighter the colour the more % of forest is in the area. While the lighter the colour of red indicates the lesser % of forest in the area. A label is presented to indicate the level of % forest of the area from lightest shade of red to the brightest red colour.



### 4.3.4 Multi Line Chart Iterations

The fourth visualisation is a multi-line chart, it is different from a dual axes multi line, a multi-chart line data of both compared databases are almost the same which is not necessary to use a double axes chart. From the chart below, we can see that there are dots on the line chart. The dot mark represents the data being plotted every year and then connected to form the line chart. An interactive feature is when a user hovers around the dots, it will indicate the value of the dot on that specific year. The colour channel used are like the other line charts which are hue colour of red and blue. A legend is created to make the user to be more aware of which line is corresponds to which data.

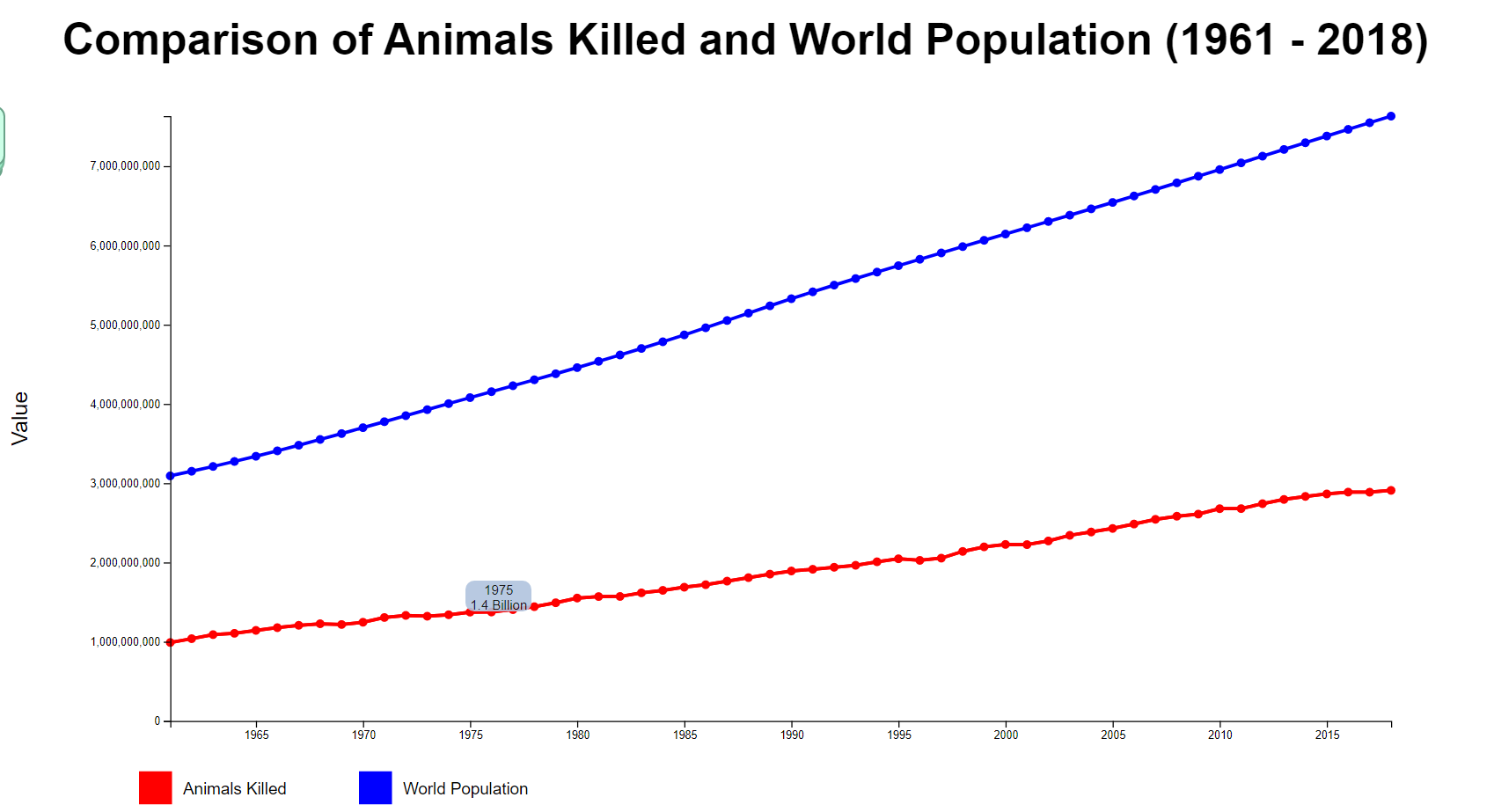


Figure 17: Final Multi Line Chart

### 4.3.5 Multi Line Chart Iterations

The last visualisation is another multi line chart. This chart however has 3 lines which are in the category of global, Northern and southern hemisphere. The chart indicates the temperature (Celsius) in different places of the world throughout the years. Instead of using legends, this chart labels the lines at the end to indicate that the line belongs to which data. When the page is loaded, an animation is played. The animation is where the lines of the chart appears from the bottom to its position. When we hover on the line, a tooltip will show the name of the line it belongs to and the line is brightly highlighted. The colour channel used is a hue colour which are red black and blue.

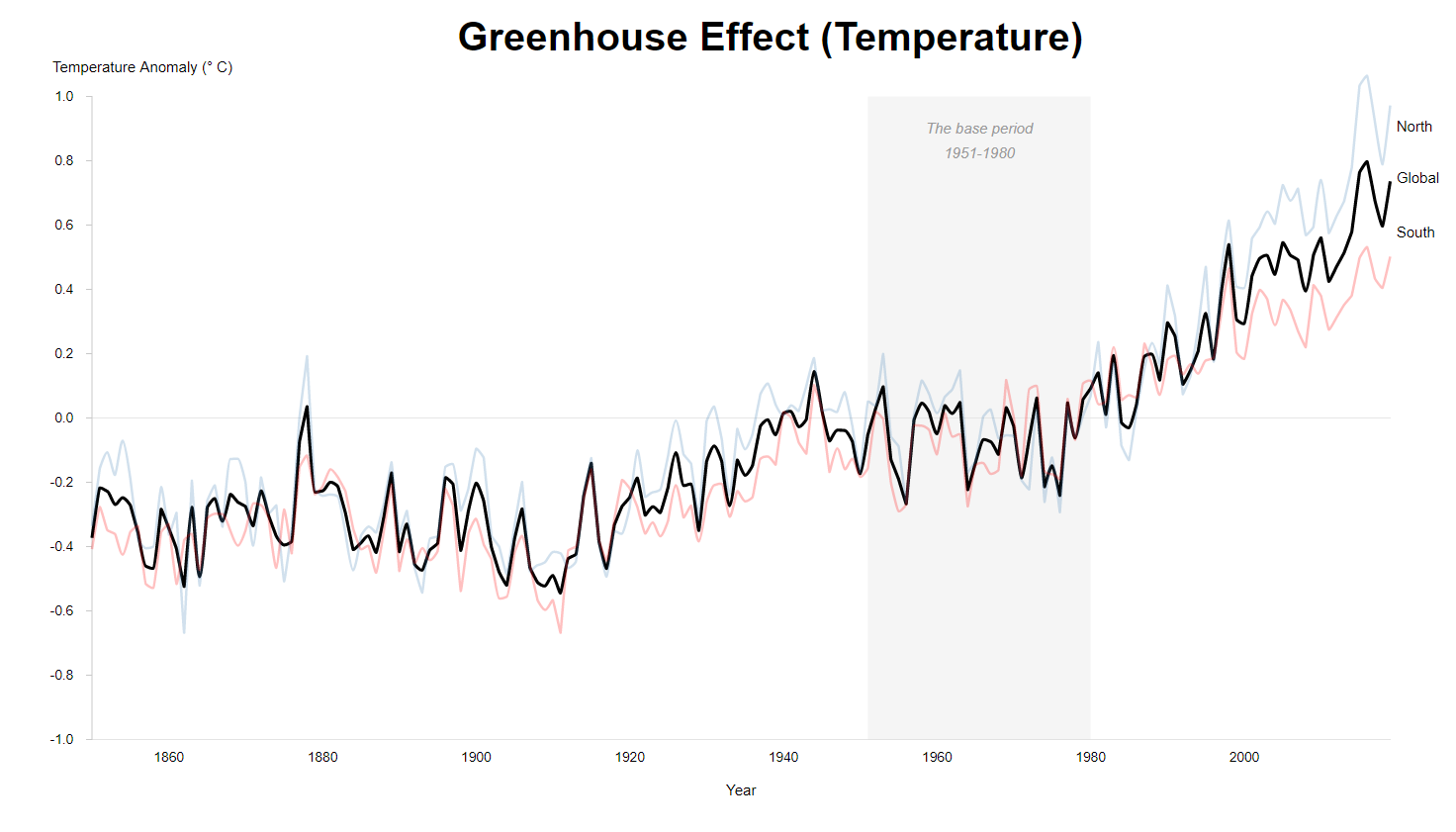


Figure 18: Final Greenhouse Line Chart Visualisation

# Conclusion

Lorem ipsum … …

# Reference

# Appendix

**Progress Report:**

Lorem ipsum … …

**Final Report:**

Lorem ipsum … …