Climate Change Project Proposal

Adelynn Shirts, Spencer DenBleyker, and Zayne Maughan

1 Introduction

We will be doing the requirements for the Call of the Data Analysis competition. This year the focus of the competition will be on environmental data. For this reason, we have found four datasets regarding the global effects of climate change that we will analyze and create visualizations for.

2 Data

We will be working with four datasets regarding global climate change aggregated by the International Monetary Fund. Each dataset is comprised of Region names with ISO2 and ISO3 codes. Some regions do not have an associated ISO2 code and are left blank.

Annual Surface Temperature Change: This dataset is comprised of the annual mean surface temperature changes across various countries and regions from 1961-2023. The temperature change is given in Celsius.

Change in Sea Levels: This dataset contains mean sea levels at various regions from 1992-2025.

Forest and Carbon: This dataset contains time series information about forested areas and carbon stocks for different global regions and countries from 1992-2022. The measurement or index differs depending on the variable. Carbon stocks in forests are given in million tonnes and forest area is given in hectares.

Climate Related Disasters: This dataset contains the number of climate-related disasters in various global regions and countries from 1980-2024. There are several missing values.

3 Preliminary Data Analysis and Visualizations

Summary Statistics: For each dataset, we will calculate basic summary statistics, such as mean, median, and standard deviation, to get a better idea of their distributions.

Visualizations: As each dataset contains time-series data, we will create line plots, potentially with smoothing to show trends over time. We will also create scatterplots, histograms, bar charts, and boxplots, where appropriate for each dataset. Heatmaps may give us insight into correlations between data. We are also considering making maps to represent the country and region-specific data.

4 Methods

In the event we submit the project to the competition, we will analyze this data using a variety of techniques, varying depending on the composition of each dataset, including relevant visualizations.

Annual Surface Temperature Change: We will model temperature trends over time, in an attempt to understand global warming patterns. We may accomplish this using regression. We will check assumptions using visual indicators.

Forest and Carbon: To investigate the relationship between forest area and carbon stocks, we could fit a linear model. This may require some data transformations for non-normal

distributions. We may also conduct a Pearson or Spearman correlation analysis, depending on the distribution of the data and the number of outliers.

Climate Related Disasters: We would like to analyze if disaster frequencies differ significantly between different regions of the world. We plan to accomplish this using ANOVA if the frequencies are normally distributed, or using Kruskal-Wallis tests if they are not normally distributed.

5 Proposed Packages and Functions

We will use a variety of baseR and additional libraries for our project. We will use the reader, dplyr and tidyr packages to import and clean our data. From the reader package, we will use the read_csv() function to import our data from the csv files. From the dplyr package, we may use the mutate() function to modify any necessary columns, such as converting temperature values. From the same package, we may also implement the filter() function to extract specific rows of our data, and select() function to retain only columns in our dataset, which may be helpful for the columns that contain many missing values. Alternatvely, we may use the drop_na() and replace_na() functions from the tidyr package to remove or replace the na values in those columns, depending on which function is better suited to the data and how we intend to use it later.

Visualizations: We intend to use ggplot2 for our visualizations. Given its many functions, we can create histograms using the geom_histogram() function, bar charts using the geom_bar() function, boxplots using the geom_boxplot() function, scatterplots using the geom_point() function. We can further use the geom_line and geom_smooth functions to create line plots, and LOESS smoothing to trend lines. When plotting the spacial data, we can use sf's geom_sf() function to plot our map data, and add annotations using ggspatial's annotation_map() function. To create heatmaps, we can use the ggcorplot() function from the ggcorrplot package. Using ggplot2 as a base for the visualizations will give them a cohesive appearance in our write-up.

Analysis: We will use baseR's summary() and sd() functions to obtain basic summary statistics. Alternatively, we could use the describe() function from the psych library to get a more detailed output of the summary stats, including the number of non-missing observations, and the skew of the data. We may also consider using baseR's table() function to create frequency tables for our categorical data.

If we submit our project to the competition, we will complete the analysis section of this proposal using a combination of the baseR and stats libraries. We anticipate using the lm() function from baseR to fit linear regression models. From the stats package, we would use the cor.test(), aov(), and kruskal.test() functions to analyze the data. We may also need to normalize or transform our data using the scale(), log(), or sqrt() functions from baseR to model the data accordingly.

6 Planned Work Allocation:

Spencer and Adelynn found the data and chose the competition. Adelynn wrote the bulk of the proposal with suggestions from Spencer. Going forward, Spencer will delegate who implements the different aspects of the code and final report. Zayne will be responsible for creating the presentation slides. All members will be responsible for implementing the code and writing the sections of the report assigned to them, as well as verifying other member's results, proofreading the final write-up, and presenting their contributions in the presentation.

References

Data found at: https://climatedata.imf.org/pages/climate-and-weather#cc1