CLIMA-CAPPERS PROJECT REPORT

Q. Name of the group members and CNetIDs

Kaveri Chhikara - kaveri@uchicago.edu
Vishal Vincent Joseph- vjoseph@uchicago.edu
Dhruv Sinha- dhruvs@uchicago.edu

Q. A brief overview of the final project (200 words maximum)

In this project, we are mapping trends in economic and climate change indicators over the years for every country. Change in economic parameters and climate change indicators are very closely related. The application that we have created will allow the user to compare indicators at the following levels-

1. variation among countries for a given year (the user can select the year)
2. variation among countries aggregated for several years (user can select a range of years)
3. variation in difference between average growth rate of gdp and co2 emissions for several years (again, user can select a range of years)

The aim of this project is to allow users to use our application as a tool to probe into any climate or economic related research questions that they might have. We started with the following questions-

1. does difference between growth rate of climate change indicator and economic indicators grow with time
2. are developing countries polluting more as they grow economically

We have also provided a brief summary of each indicator (for each year as well as aggregated over several years) that lists the top performers (country-wise) for the indicator and the years selected.

Q. The overall structure of the software (1-page maximum). It would be nice to include a helpful diagram of how the modules are connected with each other but this is not required.

We have presented the visualizations in the Dash app. The dashboard is divided into four layers. The helper functions for each of these layers are placed in **/dash_app/helpers**

1. The first layers presents a comparison of climate and economic indicators via world maps for all the countries in a given year. For every indicator, top 5 countries are presented using a bar chart below the world map.

2. Layer 2 provides an aggregated view of Layer 1. Layer 2 allows the user to select a range of years.
3. The next layer presents a consolidated performance of each country in terms of climate and economic parameters. Each bubble in layer 3 represents the difference between average GDP growth rate and average growth rate of CO2 emissions. The size of the bubbles correspond to the magnitude of difference between these metrics and the color of the bubbles represents the sign of the difference. A positive sign indicates average GDP growth is more than average CO2 emissions per year for the selected range of years.
4. The final layers presents the regression results along with an animated scatter plot that visualizes the relationship between log_CO2_emissions_per_capita and log_GDP_per_capita. According to the environmental Kuznets curve, for developing nations at low levels of growth, the level of emissions should increase until the increase in emissions plateaus and eventually starts falling when the gdp per capita is high.

Our database, **indicators.sqlite3**, provides a scaffolding for the web project. We have three different data source- world bank API, IMF and a mapping of countries to regions and sub-regions. The climate indicators dataset is dynamically populated through calls to the world bank API for a selected list of parameters for all countries from 1995-2021. The economic indicators dataset was downloaded from the IMF databank, put into Pandas dataframes and cleaned using **/analysis/econ_clean.py** and **/analysis/import_export_energy_clean.py**. We also downloaded the mean surface temperature dataset as a csv, cleaned it using **mean_st.py** and added it to the climate_indicators table in the database. The country to region mapping database is directly imported to the database.

The visualizations for each layers are created in the app.py script under /dash_app using the appropriate helper scripts for each layer. Exploratory data analysis and model development was done using a Jupyter Notebook **/analysis/eda.ipynb.** We cleaned the data, dropped indicators with more than 35% missing values and countries with more than 35% of their data missing. Finally, we imputed data using KNNImputer from sklearn.

Q. A description on the code responsibilities for each group member (i.e., who was responsible for what module, files, tasks, etc.).

Vishal- Incharge of the overall architecture of the app. Wrote app.py that creates the DASH webpage. Files worked on- app.py, climate_parameters_api.py, dynamic_dropdown_list.py, layer_1.py

Kaveri- Incharge for implementing the API script that imports dataset from the World Bank. Also incharge of creating the functions that produce regressions on the main webpage. Files worked on- climate_parameters_api.py, eda.ipynb, regressions.py

Dhruv- Incharge for cleaning the Economic parameters dataset that was downloaded from World Economic Forum. Also incharge of creating the functions for layer 2 and layer 3 on the main webpage. Files worked on- econ_clean.py, mean_st.py, import_export_energy_clean.py, layer_2.py, layer_3.py.

Q. Short description on how to interact with the application and what it produces.

1. The first map of the web page allows the user to select a year, climate change indicator, and an economic parameter from the dropdown. The result of this query will be reflected in the two maps shown on the page. Below the two maps, there will be a list of top 4 performers for each indicator

2. As we scroll down, the next layer lets users aggregate the indicators over the years. The user will select the years, climate change indicator, and an economic parameter from the dropdown. Like the previous layer, the result of this query will be reflected in the two maps. This time, it's aggregated for several years (whatever user selects). Below the two maps, there will be a list of top 3 performers for each indicator.

3. In the next layer, the user only selects years. The outcome of this layer is the bubbles on each country on the world map. The size of the bubble reflects the magnitude of difference between average gdp growth and average growth of co2 emissions. The sign of the bubble reflects whether the difference is positive or negative.

4. In the last layer, we allow the user to do regression analysis. By default, the layer regresses log of Co2 emissions on log GDP/capita and log GDP/capita squared. The user can add imports and exports, but the model is vulnerable to the problem of multicollinearity on adding additional variables.

Q. What the project tried to accomplish and what it actually accomplished (200 words)

Given time and resource constraints, we had to modify the scope of the project. Our goal was to create a geospatial tool with two visualization components - 1) The trend of key climate indicators over time at the country level with special emphasis on how these trends vary around major climate change conventions, and 2) The trend of key economic indicators over the same time period. We have successfully created a dashboard using Dash to compare the trends in climate and economic indicators. However, we have been unable to find a relation between these trends and major climate change conventions. The current dashboard is divided into three layers-mentioned in the overview above. The final layer performs time fixed effects regression on the panel dataset at a regional level. We would have liked to incorporate structural breaks in this analysis during the years of major climate change conventions to find meaningful relationships between economic indicators and greenhouse gas emissions. This would require using statistical tests like the Chow test. We would also like to incorporate other socio-economic indicators like the human development index, level of corruption in a country and energy use.