Understanding the effect of Victoria's energy consumption and emission on Victoria's Climate

Introduction

Domain: Community & Environment

The report analyses Victoria's energy consumption, emission and climate data across a 25-year period (1990 – 2014). It focuses on answering these questions:

1. Does Victoria's energy consumption have any effect on Victoria's increasingly hot climate[i]?
2. Do different parts of Victoria's energy consumption have variable effects on Victoria's climate?
3. Does Victoria's emission have any effect on Victoria's' climate and is it correlated to Victoria's energy consumption?

Datasets

- **VIC Climate Data (CSV)**
  - Adapted from the Climate Change Section under the official website of Bureau of Meteorology (BoM), Australian Government. It has various climate data including mean, maximum and minimum temperature data of Victoria from 1910 to present. Link: http://www.bom.gov.au/climate/change/#tabs=Tracker&tracker=timeseries
- **VIC Green House Gas Emission Data (CSV)**
  - Adapted from the National Greenhouse Gas Trend section under the official website of the Department of the Environment and Energy, Australian Government's official website. This dataset has detailed energy consumption and emission of various sectors of Victoria ranging from 1990 to 2014. Link: http://ageis.climatechange.gov.au/NGGITrend.aspx
- **VIC Energy Consumption Statistics (Excel)**
  - Adapted from the official website of the Office of Chief Economist, Department of Industry, Innovation and Science, Australian Government. It has detailed energy consumption data of Victoria since 1960. Link: https://industry.gov.au/Office-of-the-Chief-Economist/Publications/Pages/Australian-energy-statistics.aspx (Table B & G).

Pre-processing & Integration

Tools used include Pandas, Matplotlib and Seaborn Library. DataFrame in Pandas is extensively used due to its versatility in processing data. Jupyter Notebook is chosen as a python IDE because of its simplicity in parsing and executing python commands. Data pre-processing are divided into three parts, each pertaining to its corresponding datasets.

VIC Climate Data (CSV)

BoM offers separate Victoria climate data in form of .txt file with a custom format, hence it is necessary to convert these data into a form which is suitable for python data processing. The following procedures are applied on these data.

- Data with the date before 1990 or after 2014 are discarded to ensure consistency of time range later in the integration.
- Date of each data entry is trimmed to include year only, instead of its original date format (start yymmdd to end yymmdd). Since each year only contains one data entry after visual checking, it is safe to retain only the year information of each data.
- Separation between each year and its corresponding data is changed to a single comma strictly, instead of its original format (mixed space and tab). This is to facilitate the interpretation of csv file to ensure consistency in data integration.
- Column names are added to each file, with "year" on the left and data name on the right. This is to ensure correct merging of different climate data in data integration.
- A new file with "clean" added to its original name's suffix is created for each file to store cleansed data. These files will be used in data integration.

VIC Energy Consumption Statistics (Excel)

Data downloaded from the website provided above are in form of Microsoft Excel file. Only two tables are selected for their relevancy with this project.

- *Table B: Australian energy intensity and energy consumption per person, by state, energy units.*
  - This dataset is chosen to investigate the correlation of energy consumption, productivity and intensity as well as energy consumption per capita with climate.
- *Table G: Australian energy consumption, by state, by fuel, physical units.*
  - This dataset is chosen to investigate various fuel type of energy consumption.
  - The section of "Production of derived fuels" is removed since it is not relevant to the focus of this project.

Both CSVs have original well-phrased column names in the first row and corresponding unit on the second row. Their original indexes are reset and kept as a new column named "Year". The year of each data is changed to contain only one year (e.g. 2010), instead of its original format (e.g. 2010-2011) to ensure consistency. Both CSVs are trimmed by removing the first and second row to retain their data only. Then their columns are renamed using their original column names. Only data between 1990 to 2014 are kept for time range consistency. All numbers are converted to a unified format to avoid 'isnan' error later during integration.

VIC Green House Gas Emission Data (CSV)

All thousand separators are removed to ensure correct interpretation of numbers by python. The separator is changed to commas instead of semicolons for consistent CSV format. This cleaned CSV file is changed to a transpose of itself to ensure the year column is vertical for time consistency. Then each column is renamed with its original descriptive name. The index is reset and its original index is kept as a new column named "Year". No trimming is applied since all data are perfectly in range.

It is not suitable to remove outliers within the time range of the three datasets above. These data are conclusive statistical data from official government websites, derived from years of individual data and should have done data cleansing already before computation.

After careful pre-processing of all datasets described above, the final integration is easily done using Pandas Library in Python. By loading all data files of each dataset using a for loop and `read_csv()` facility, data of each dataset are merged by their common column "Year". The resulting DataFrames consist a single energy consumption DataFrame, a single climate DataFrame and a single emission DataFrame. A new function for generating a correlation matrix between two DataFrames is created which uses to each column from one DataFrame to compute correlation against each column in the other DataFrame to form a correlation matrix. The resulting correlation matrix will be represented using a heat map later.

Limitations and Difficulties

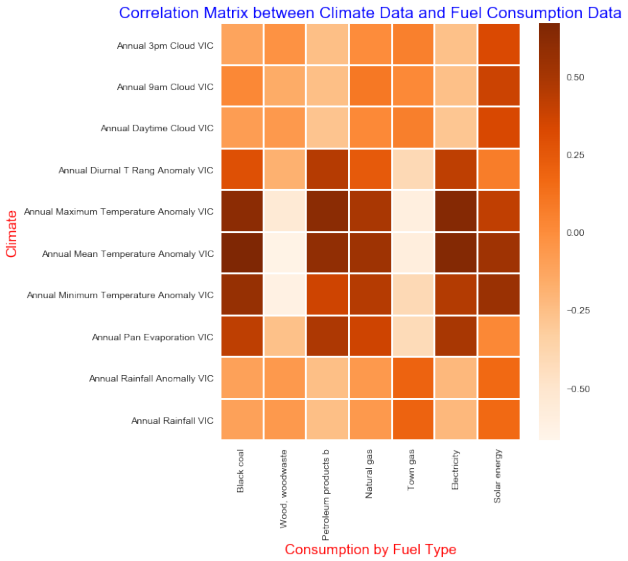
Energy consumption data are in original presented as an interval between two consecutive years instead of a single year. While the starting date and month is unknown, the assumption to categorise data into a single year may cause inconsistent data. It is also unknown how outliers are removed before computing the final data used in all three datasets. Different government agencies may collect and calculate their statistical data differently. It is unknown what kind of pre-processing they have done to these data. Due to these limitations, data inconsistency may arise.

Some DataFrames may still yield '`isnan`' error during plotting using Seaborn `heatmap()` or Pandas `corr()`. This is caused by python's misinterpretation of some numbers as strings. This can be fixed by applying a `astype(float)` action to all DataFrames. The resulting DataFrames is not changed, except the years are now in float rather than int. This may look aesthetically unappealing, but it does not affect the overall results since year is only used as a common column to join DataFrames or as an x-axis label reference during plotting. It does not affect any data consistency. Therefore, this problem could be put aside.

Results

the following results are generated by Integrating data of different domains.

By looking through every Pearson Correlation between each pair of columns from these two DataFrames, this correlation matrix is generated. This heat map suggests that while many data do not exhibit strong correlations, there are still some high correlations in the data. This is shown by the “deep orange” colours in the heatmap.

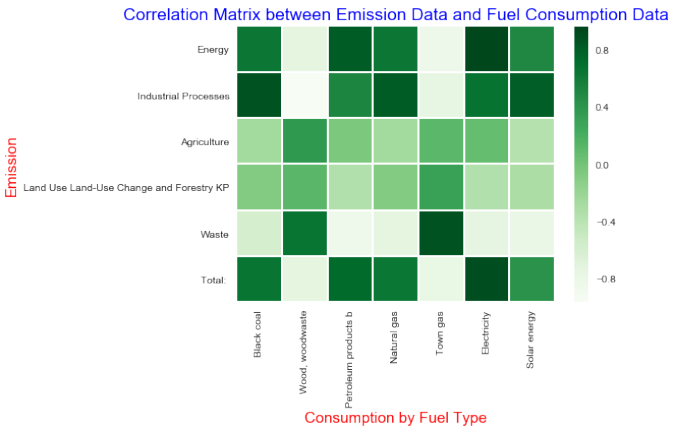


By taking a closer look at these data, here are some interesting findings:

- Temperature anomalies like maximum, minimum and mean temperature anomalies are likely to be affected by fuel consumption, regardless of fuel type. ($r \approx 0.7$)
- Wood, wood-waste and town gas have negative correlations with all temperature anomalies. ($r \approx -0.6$)
- Energy consumption, by fuel type generally have little or no correlation with all cloud data and rainfall. ($r \approx 0.1$)

A possible cause for this correlation stated above could be that fuel consumption caused too much emission like greenhouse gas emission. The use of wood-waste and town gas has plummeted these years due to technological advance but the temperature is rising. This could possibly be the cause of negative correlation. To verify if fuel consumption created too much greenhouse gas emission, we need to investigate the correlation between energy consumption and emission. Here are the results:

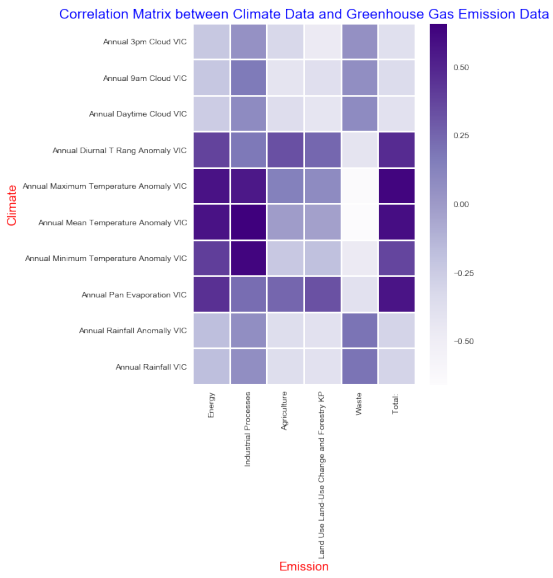
Investigating this result also yields some interesting findings:



- Energy consumption of all fuel types, except wood-waste and town gas has high correlations with the amount of energy emitted, total amount of emission created and the amount of emission produced by industrial process. ($r > 0.6$)
- Emission created by agriculture sector and land-use sector have little or no correlation with the type of fuel consumed. ($r \approx 0.1$)
- Waste have negative correlations ($r < -0.7$) with all fuel types except wood-waste and town gas, but it has a high correlation ($r \approx 0.7$) with these two.

This result suggests that consumption of fuels is very likely to create emission. More specifically, industrial process is the most “emissive” sector among all sectors. Again, wood-waste and town gas are used less often each year and most are emitted as waste rather than CO2 equivalent gas, therefore it shows a negative correlation with each sector.

By plotting the correlation matrix between emission and climate, a new heat map is generated:

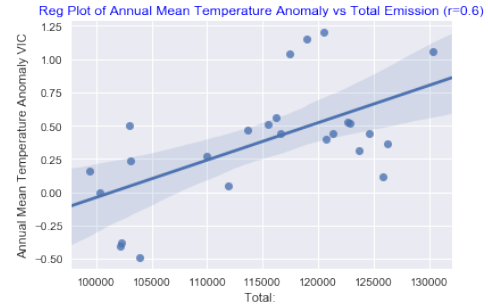


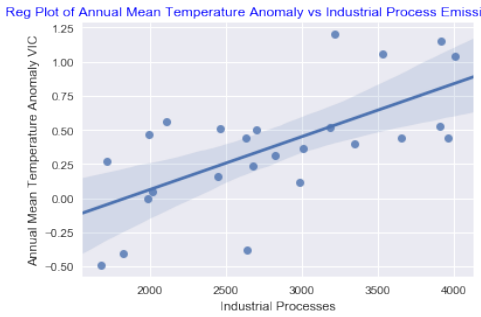
This heat map shows several notable features:

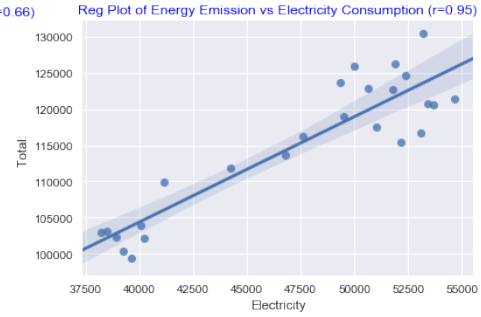
- Emission by the energy sector, industrial process and total emission has high correlations with annual mean, maximum and minimum temperature anomalies and pan evaporation ($r \approx 0.6$).
- Waste has negative correlations with all temperature anomalies ($r \approx -0.6$).

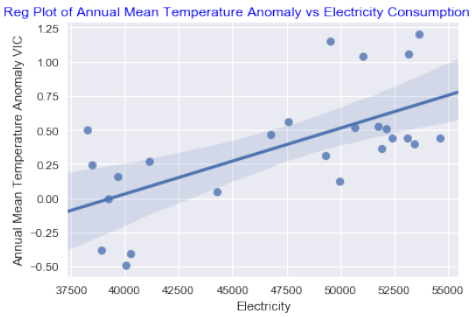
- Cloud and rainfall have little or no correlation with emission (r < 0.2).

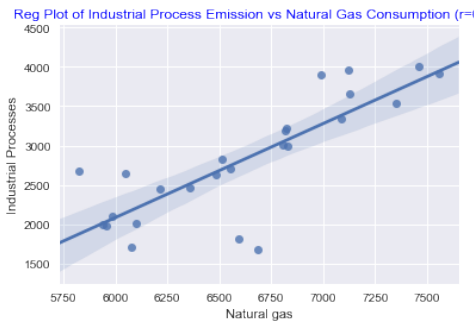
Since it is impractical to look at every correlation, several with high ‘r’ values are chosen and are individually plotted as a scatter plot with a linear regression line on the graph.

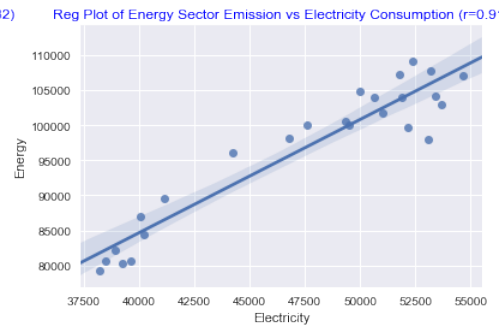












Regression plot between electricity consumption and energy emission shows a clear, straight linear regression line, suggesting a strong correlation between electricity consumption and emission from the energy sector. Similarly, regression plot between emission from energy sector with electricity consumption also shows a very strong correlation ‘r' = 0.95. These two graphs indicate that electricity consumption does cause more emission, at least in the energy sector.

Annual mean temperature anomaly has a positive correlation with total emission. The Pearson Correlation coefficient is not very high, an ‘r' = 0.6 is considered as ‘moderate'. This suggests that mean temperature anomaly is related to electricity consumption and the degree of correlation is not negligible. This is also true for annual mean temperature anomaly and industrial process emission, where ‘r’ = 0.66. These two graphs have demonstrated that temperature is affected by the amount of emission from some sector like industrial process.

Consumption of electricity also has a moderate correlation with annual mean temperature anomaly, as ‘r’ = 0.65 shown in the graph. Consumption of natural gas shows strong positive correlation with industrial process emission, with a few outliers at the bottom of the graph. These two graphs indicate that some fuel consumptions do have effects on climate temperature.

Undeniably, there are some outliers in these graphs, while removing these outliers may yield more accurate results, the presence of these outliers is preferable here since the size of the datasets is not large and these outliers may yield potential significant results.

Hence it is clear that energy consumption does have effects on climate. These three are all correlated to each other as energy consumptions cause greenhouse gas emissions and subsequently affect climate, especially temperatures and pan evaporation. Some energy consumptions like electricity and petroleum products b do have variable effects on different parts of Victoria’s climate.

Value

With only the raw data from both datasets, one could hardly see any trend or correlations that lie between the datasets. The Government and various agencies would have major difficulties deciphering the value from the datasets and would unable to come up with potential corresponding policies and measures. The combination of these datasets will provide trend and degree of correlation which are valuable to government and environmental agencies for their policy-making. It also coloured degree of correlation,

providing a clear and intuitively way to look at these data instead of raw numbers. Correlations are plotted with data analysis techniques like linear regression and scatter plot provide detailed, in-depth graphs which clearly show the difference in the energy consumption and their effects on the climate regarding different industries and sectors. The results provide a broader picture of Victoria's climate for every Victorian to improve their awareness of their living environment, with valuable insights and references to environmental agencies in line with the current environmental policy.

Challenges and Reflections

The original project plan is to include some geographical information with the detailed emission data of different locations in Victoria. This project initially intends to find out the locations with high emissions in Victoria to see whether the climate at these locations are more affected or not. However, due to lack of geographical data, it is subsequently axed. There are also plans to download every climate data of every weather stations in Victoria and use these data to plot every climate variations of every place in Victoria. However, it is later realised that the amount of work to achieve this level of massiveness is enormous, this plan is abandoned halfway in the project due to time constraint. Another dataset about energy consumption per person is also included earlier. Due to the word limit and its relevance with the focus of this project, it is also cut off from the final report and included in the code section in the attachment.
The results of this project may be subjective since there is only limited data from 1990 to 2014. Climate problems should be investigated with a wide range of data support, this project only uses limited data, therefore the result may be more accurate if larger datasets are provided. Additionally, due to the lack of data, it is hard to remove outliers which may affect the result of the project. There may be other factors causing climate changes that have not been considered, these factors may affect the result as well.

Question Resolution

The questions are answered sufficiently, the goal to answer the correlations between climate, energy consumption and emission is achieved. However, it is not sufficient to conclude at this point since more structural and in-depth analysis are required to derive more insightful findings and resolutions. Future research could possibly include more data from different fields to refine the findings of this project.
With Victoria's increasing energy consumption each year, more CO2 equivalent gas is emitted by different sectors. This does have an impact on Victoria's climate. The government should limit CO2 emissions of sectors like Industrial and energy sectors as they have high emissions and have large effects on the climate. It should also treat fuel consumption of black coal, petroleum products b and electricity more seriously as these consumptions emit more and have more effects on the climate. The temperature of the climate is more volatile than other parts of the climate, therefore, to control global warming, the government should encourage the low carbon footprint energy and innovates in other energy sources, as most of the current energy sources do have some unneglectable effect on the climate.

Code

These Python Libraries were used in this project: `pandas`, `matplotlib`, `seaborn` and `os`

- Read .txt files, modify file and pre-processing were achieved using `os` library in python, this is necessary since different datasets are in different folders. To open, pre-process and save to a new clean CSV file, `os.listdir()` is necessary for file operations.
- Read data from .csv files and save them as DataFrame objects in python
- Clean and trim data to fitted in specified time range
- Inner join table on `Year` column
- Calculate Pearson coefficient for each pair of variables
- Generate regplot, correlation matrices, and scatter plots using `matplotlib` and `seaborn`

Bibliography

[i] ttp://www.climatechange.vic.gov.au, VIC Climate Change Act, 2017