DYSON SCHOOL
OF DESIGN ENGINEERING

MENG DESIGN ENGINEERING

Data Product to Visualize Environmental and Economic Data

An interactive visualisation that allows the exploration of various country data over time (namely, CO_2 emissions, GDP and renewables)

Authors:

Andrea Contri (01942886) & Aran Singh Mahal (01769144)

	Aran	Andrea
Ideation	50%	50%
Data Sourcing	40%	60%
Data Curation	45%	55%
Data Product Canvas	65%	35%
Conceptual and System Design	50%	50%
Programming	40%	60%
Testing and Validation	50%	50%
Archiving and Documentation	30%	70%
Report Writing	60%	40%

December 27^{th}, 2023

The Problem

Our world currently is being driven by data and the insights provided by it. With vast amounts of data revolving around environmental and economic factors, we are presented with a unique opportunity to not only understand but also communicate this data. This report outlines the development of a data visualisation tool that is intuitive and accessible to a wide range of stakeholders, from the general public to experts in policy-making and science. We have looked at a range of metrics under the umbrella of CO2 emissions, renewable energy and GDP for countries over the years. This application aims to allow users to identify trends in data, have more informed discussions on topics surrounding climate change and, eventually, be able to predict how countries will perform in the future.

We plan on making the code for this project open-source so that more tech-savvy users can adjust features as they see fit or add more useful functionality such as future predictions for best- and worst-case scenarios and impacts of this on the world.

Data sourcing and curation

After the decision on the type of problem we wanted to tackle, an initial evaluation of the reliability of the data was carried out. Four different datasets were chosen and isolated, and of these, three datasets include data over a year range. The datasets contained data on different timespans.

- "info.csv": contains general data for each country. This data is not tracked over the years as the majority does not vary. For data that did vary, we decided it was not relevant for this visualization. It includes parameters such as the area of the country, the population, and the currency. This dataset was sourced from "Our World in Data" [1], a project of the Global Change Data Lab that presents empirical research and data on a wide range of topics, such as public health, poverty, education, and the environment.
- "gdp.csv": contains the GDP (total and per capita) of the countries from 1960 to 2021. The data was adapted from a dataset sourced from the World Bank [2].
- "emissions.csv": sourced from the International Energy Agency [3]. It includes the data on CO2 emissions. The data is divided into the main sources of CO2 Emissions: coal, oil, gas, cement production, and flaring. It spans over two centuries of data, from 1750 to 2021.
- "sus.csv": also sourced from the International Energy Agency, it includes the main sources of renewable energy data per country and some other interesting statistics on the electricity usage and sustainability of the countries. It spans from 2000 to 2021 and contains parameters such as the percentage of the population with access to electricity, and the source of electricity (from fossil fuels, nuclear, and renewable sources).

Since the full names of the countries varied between datasets, it was decided to use the ISO 3166 alpha-3 codes. These are three-letter country codes defined in the ISO 3166 standard published by the International Organization for Standardization [4]. For datasets that were missing these codes, a column with all the ISO codes was added.

As expected, the data for some countries was missing. To create a uniform country list, countries that were missing data in the "info.csv" datasets were removed from all datasets. Countries missing in the other datasets were inserted with their ISO code, and all feature entries were set to NA. The list of the missing countries can be found in the Appendix.

To avoid excessive computation and the consequent slowing down of the visualization, we decided to build two different data structures. This slowed down the initial loading of the app but significantly reduced the lag while changing the parameters of the visualization. The first data structure "selectedData" is for the map visualization. Since accessing the parameters from the nested lists became too slow, we separated the parameters into different datasets. This way only one dataset had to be loaded for each visualization. The application has reactive variables that, based on the parameters selected by the user, help return the respective dataset and columns. The second data structure "data" was used for the information panel on the map, and the time graph plot of the selected parameters and country. The outline of the data structure can be found in the appendix.

The Solution

The final visualization is shown in Figure 1. It's formed by two columns. In the left column, the user can select the country, the sub-section and the final parameter through three dropdown menus. At the bottom of the column, there is a panel that plots the timeline data for the chosen parameter and country. The second column shows the map of the world, with three panels on top: the legend, the country information, and the year slider. By sliding the year, the user can visualize the data of that year on the map. The user can also view more specific information on a country by hovering and clicking directly on the map. All panels are reactive and automatically updated based on the different user inputs. The map panels can also be dragged around within the boundaries of the map. A "How to Use" button is also available on the top right of the page, toggling an instructions pop-up.

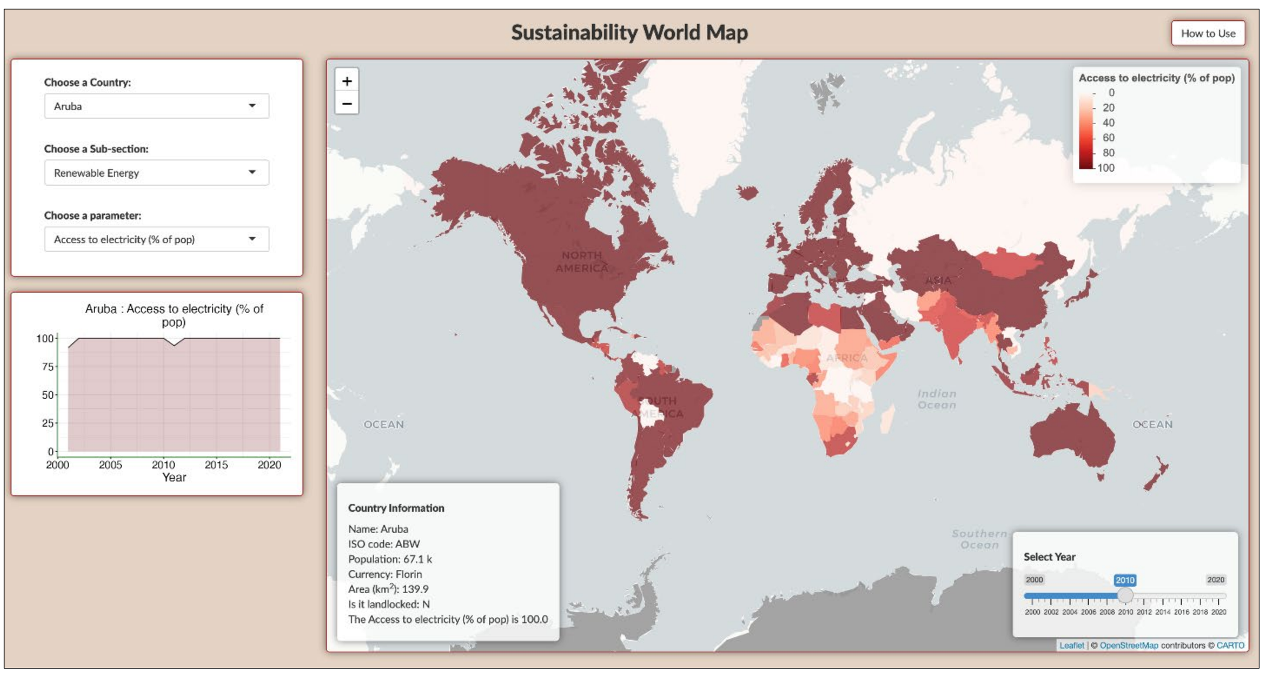


Figure 1: Final visualization

Development

File Structure and Version Control

Due to the large size of the code, this is divided into multiple files and folders. The structure is outlined below.

- "*app.R*": the main file containing the UI and server code for the app.
- "*global.R*": file that contains the libraries, the data cleaning, and the support functions.

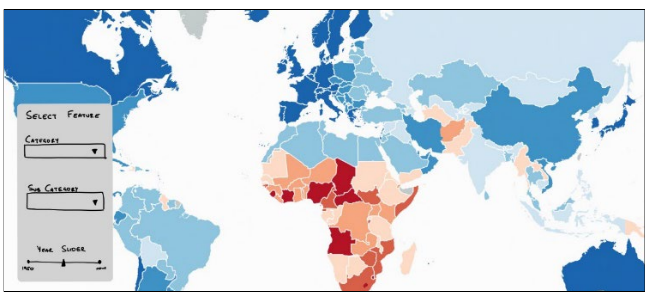
- "/datasets": contains the four .csv files.
- "/world": contains the files for the map rendering.
- "/www": contains the style.css file for the HTML formatting.

To back up the code and share it among us, the files were constantly pushed and updated to a GitHub directory and a OneDrive folder.

Map

For this sort of data product, utilizing choropleth plots with responsive inputs seemed to be the most suitable for an intuitive user experience. Early sketches of the design were used to focus on layout, functionality and understanding user interaction with the platform (Figure 2). Initially, we thought that having a whole screen map plot with draggable panels for user input and information would be best (this changed later). We also planned to include a feature where once a country is clicked, the interface zooms in on the country and then displays the information panel with the relevant category data (Figure 3).

The initial lectures on descriptive analytics and reviewing what makes data visualisations successful were very useful for this initial ideation stage. We researched and reviewed different types of visuals and decided that choropleths and time graphs would be most suitable for displaying the information. The techniques related to forecasting can also be used for future developments of the platform (outlined in Future Plans below). Also, the tutorials related to responsive elements were useful in determining which parts of the platform need different UI and server code.



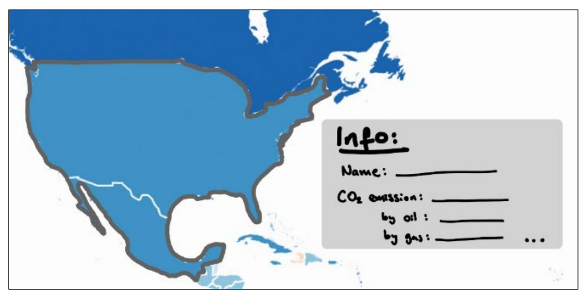


Figure 2 & 3: Initial Sketch of what interface and information panel could look like

The Leaflet library was used to create the maps and we used draggable panels on the page to display the inputs. Country names were able to be displayed in the top right depending on which country the user's mouse was hovering over (Figure 4). This initial iteration was developed when we created a cleaner user interface, with more in-depth information and slide panels. When we interacted with the initial version, we decided that the interface could be improved if the "Choose an Option" panel was not covering the map.

Figure 5 shows the updated user interface: a select box panel, a plot associated with the chosen feature and county, a slider panel, and an information panel. It is much cleaner than the previous iteration since less of the map is covered by the panels. The panel on the left contains a plot to show the chosen country's feature development and trends over time. The time graph on the left and the choropleth plots both needed the overall datasets to be processed in different ways. The feature-time graph required country-specific data for each feature across the different years, whereas the choropleth plot required annual data for each feature according to the country. Although it is quite computationally expensive to carry out the data extraction in this way, the results show a more comprehensive depiction of what the data can show. One issue we ran into when developing this data product is that due to the vast amounts of data, different data frames and processes, the application loads slower than would be ideal. However, once fully loaded, the application runs efficiently.

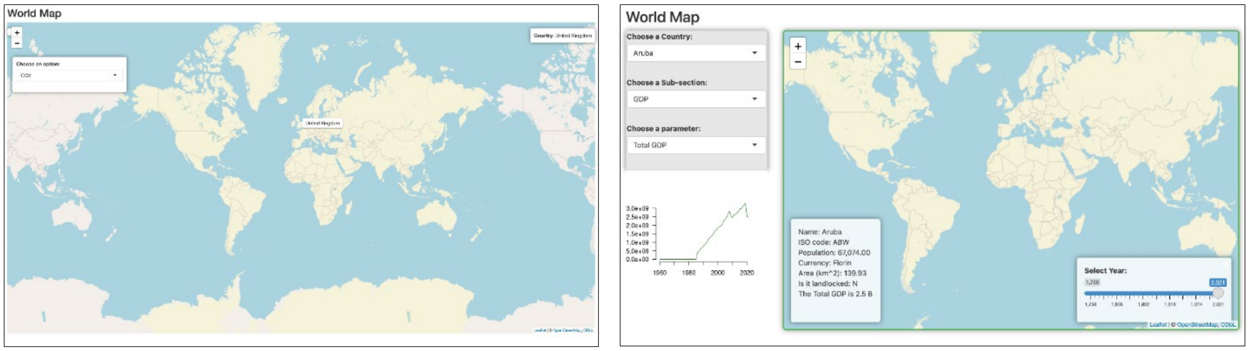


Figure 4 & 5 – First and second versions of the visualization

We ran into data reading and header referencing issues, which hindered the progress of implementing the choropleth map. In this iteration (Figure 6), we displayed the selected data directly on the page to verify data was loaded correctly and checked against the choropleth. To enhance user experience, we also revised the slider to be responsive; it now adjusts its start and end dates based on the available years for the chosen category.

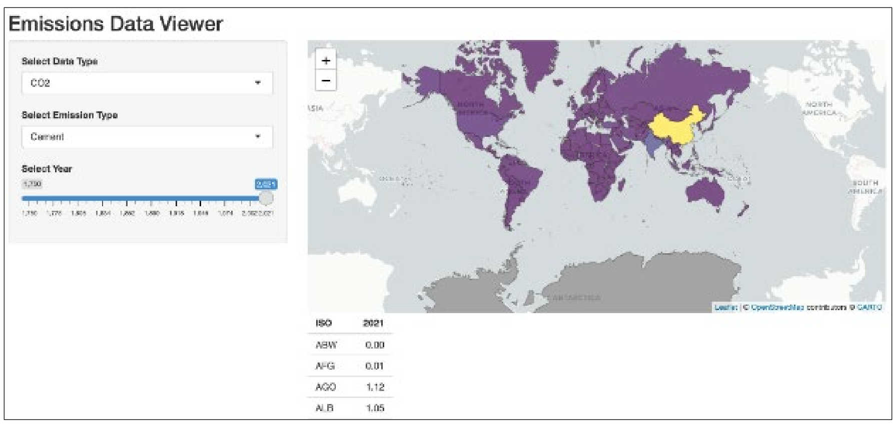


Figure 6: Final working prototype

Key Performance Indicators

Key Performance Indicators were used to evaluate the effectiveness of the data visualisation, each of which was focused on an aspect of the application's performance or user experience.

The data used has been sourced from primary or verified secondary sources. Representing accurate data is crucial for an application such as this since it could be used to inform of future scenarios. If we had more time, we could have implemented a feedback feature, where users can suggest changes or updates to current or missing data with cited references.

After we finished coding the project, we asked 8 people for some feedback on the platform. Here, we were primarily gauging how easy the application was to use, how effectively the data was communicated and gathering feedback on the visuals. All 8 of the people surveyed stated that the visuals for the platform were effective at delivering the information and they enjoyed the responsive elements (options and slider). However, they did say that the loading of the maps took some time. This is something we could look at making faster in the future. Out of all the people surveyed, 4 of them were university students, and they all stated that it was efficient for getting relevant information if needed. In future, we could carry out tests with expert researchers and policymakers to see how it could be improved and how effective it currently is. However, these are all qualitative measures for success. Once launched, we could look at user engagement metrics like users' session duration, how many new sign-ins and users' retention rate.

Audience and Interaction

As stated previously, the platform is targeted at a wide audience; from people who are just inquisitive on climate change matters to student researchers, policymakers and governments. The current features of the platform allow for basic exploration of the data, but future developments would allow for future scenario predictions, correlations between metrics and comparisons between countries (explored more in Future Steps). These features would open the door to more specialist researchers and policymakers.

On a more basic level, students of different experience levels can use the platform to help with coursework and can cite any of the data used in the platform. On the opposite side, we could keep our platform open-source so individual groups can build upon and improve or add features as they see fit. When it comes to the prediction of future scenarios, experts and scientists as well as ourselves can develop an algorithm for this.

Policymakers and governments can then make use of more advanced features of the platform to make informed decisions on future targets and plans to reduce CO_2 emissions or increase renewable energy output.

Expected Impact

This visualisation aims to improve the understanding of environmental and economic data for our wide audience. By using a tool like ours, we expect people to have more informed conversations about climate change and make more accurate predictions of how countries will perform in the future. By representing the data in a visual way such as this, we expect the public to understand the changes surrounding these topics and potentially lead to behavioural changes in terms of energy usage.

We expect that if countries can see the future trajectory of their CO_2 emissions or renewable energy output, they will put in place measures to ensure they hit their targets. With COP28 just gone by and countries taking more and more responsibility for their impact on the planet, we believe that this data product will not only help countries keep themselves but others accountable too. These future predictions for CO_2 emissions can be used to simulate how the world will look in X number of years and hopefully influence communities to act and reduce their carbon footprint. With companies pushing more towards being carbon neutral or positive, we can expect them to endorse or use these sorts of visuals to encourage others to also follow suit.

Future Plans:

Due to the time restrictions of this project, we were not able to implement all the features we wanted. During our user interviews, they did mention some features which we were also thinking of including if we had time. All the features mentioned below add a layer of visual information that can be useful for several different applications ranging from comparison to prediction.

Firstly, we would want to implement a forecasting feature, where users can see best and worst-case scenarios for a particular feature of a country. This could be implemented in both the time graph and the choropleth plot. These predictions can be particularly useful for governments and policymakers to identify if targets will be hit or not.

A comparison would also be a very useful piece of information to convey. Here, we would allow users to pick counties they want to compare, and it would print suitable data and create relevant plots to effectively demonstrate changes. Additionally, correlations between different features would also be beneficial to implement into the platform. This feature would be particularly useful for students or experts looking to carry out comparisons between certain countries of interest. This feature will not be as complex as the future predictions, and would merge with further plotting improvements, such as real time data labeling when the mouse hovers the plot.

Another thing to consider is the missing data. It would be beneficial to all if there was a way to add to the data so there are fewer missing entries. These entries will have to be sourced and then verified to ensure no inaccurate predictions are made.

Finally, a top 10 ranking could be added as a further panel on the map: this would allow the quick understanding of the best/worse performing countries per parameter

Figure 7 below shows an illustration of how some of these features could look in a future iteration of the data product.

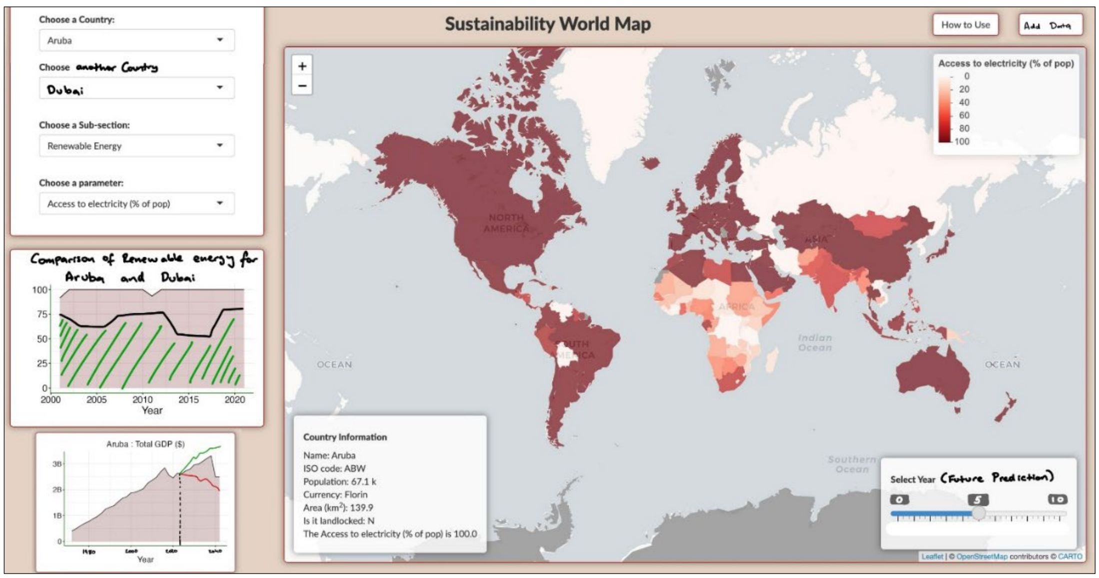


Figure 7 - Example of future iteration

Conclusion:

This report outlines an overview of how this data product was developed and highlights some of the potential impacts it could have. The tool has been created to make vast amounts of environmental and economic data easily understood by a wide range of stakeholders. The whole process made us understand more about what it takes to build a data product and the development processes that are required. We made a conscious effort to display the data in a very clear and efficient way using suitable plots. Another major component of the project was data sourcing, collation and formatting. During the initial stages, we had difficulties deciding on which data to use since some had many missing entries and others didn't have enough data. Once the data was all prepared, we were tasked with creating the platform. Due to each of us working across different files, we had issues merging code due to variable mis-referencing. With the large amounts of data and processing required, we were presented with some runtime issues. Although this doesn't affect the project too drastically at this stage, with future features and computation, the application may run slowly. We would have to look at more efficient ways of carrying out this data processing. But, overall, we believe this application can be used as a basis to form conclusions and eventually predictions of how countries will perform depending on certain features surrounding CO2 emissions, renewable energy data and GDP.

Appendix:

References

[1] - Our world in data (2023). Our World in Data. [online] Our World in Data.

[2] - World Bank (2023). World Bank open data. [online] World Bank.

[3] - IEA (2019). Data & Statistics - IEA. [online] IEA.

[4] - ISO (2019). ISO - International Organization for Standardization. [online] ISO.

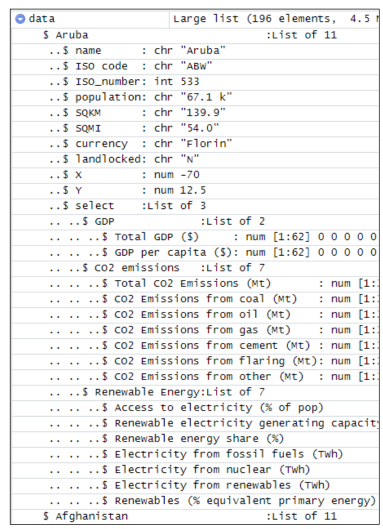
Contributions

We are both very happy with the contributions of our partner. We had no disagreement and believe there is no imbalance in the amount of work carried out by either of us.

Outline of the Data Structure

Below is the structure of “data”, one of the two main data structures used in the visualization.

- [Name of a country]
  - ISO code: [“ ”]
  - Population: [“ ”]
  - [Other Parameters] : [“ ”]
  - Select:
    - GDP:
      - Total GDP: [“ ”]
      - GDP per capita: [“ ”]
    - CO2 Emissions:
      - CO2 Emissions from coal: [“ ”]
      - [Other Parameters] : [“ ”]
    - Renewable Energy:
      - [Other Parameters] : [“ ”]
- [Name of another country]



```
data                              Large list (196 elements,  4.5 M
  $ Aruba                                :List of 11
   ..$ name      : chr "Aruba"
   ..$ ISO code  : chr "ABW"
   ..$ ISO_number: int 533
   ..$ population: chr "67.1 k"
   ..$ SQKM      : chr "139.9"
   ..$ SQMI      : chr "54.0"
   ..$ currency  : chr "Florin"
   ..$ landlocked: chr "N"
   ..$ X         : num -70
   ..$ Y         : num 12.5
   ..$ select    :List of 3
   .. ..$ GDP              :List of 2
   .. .. ..$ Total GDP ($)     : num [1:62] 0 0 0 0 0
   .. .. ..$ GDP per capita ($): num [1:62] 0 0 0 0 0
   .. ..$ CO2 emissions    :List of 7
   .. .. ..$ Total CO2 Emissions (Mt)       : num [1:
   .. .. ..$ CO2 Emissions from coal (Mt)   : num [1:
   .. .. ..$ CO2 Emissions from oil (Mt)    : num [1:
   .. .. ..$ CO2 Emissions from gas (Mt)    : num [1:
   .. .. ..$ CO2 Emissions from cement (Mt) : num [1:
   .. .. ..$ CO2 Emissions from flaring (Mt): num [1:
   .. .. ..$ CO2 Emissions from other (Mt)  : num [1:
   .. ..$ Renewable Energy:List of 7
   .. .. ..$ Access to electricity (% of pop)
   .. .. ..$ Renewable electricity generating capacity
   .. .. ..$ Renewable energy share (%)
   .. .. ..$ Electricity from fossil fuels (TWh)
   .. .. ..$ Electricity from nuclear (TWh)
   .. .. ..$ Electricity from renewables (TWh)
   .. .. ..$ Renewables (% equivalent primary energy)
  $ Afghanistan                          :List of 11
```

List of missing countries

These are the countries that did not have data in the “info.csv” database. Other “major” countries were missing data, but in that case the data was gathered manually.

American Samoa , Burma , Cayman Islands , Cook Islands , French Guiana , Falkland Islands (Malvinas) , Micronesia, Federated States of , Guam , Korea, Democratic People's Republic of , Lao People's Democratic Republic , Martinique , Montserrat , The former Yugoslav Republic of Macedonia, Niue , Anguilla , Hong Kong , Northern Mariana Islands , Faroe Islands , Gibraltar , Isle of Man , Macau , Monaco , Palestine , Montenegro , Mayotte , Aland Islands , Norfolk Island , Cocos (Keeling) Islands , Antarctica , Bouvet Island , French Southern and Antarctic Lands , Heard Island and McDonald Islands , British Indian Ocean Territory , Christmas Island , United States Minor Outlying Islands , Reunion , Republic of Moldova , Syrian Arab Republic , Tokelau , Sao Tome and , United States Virgin Islands , Wallis and Futuna Islands , Guadeloupe , Netherlands Antilles , Pitcairn Islands , Saint Pierre and Miquelon , Saint Helena , San Marino , Serbia , Holy See (Vatican City , Svalbard , Saint Martin , Saint Barthelemy , Guernsey , Jersey , South Georgia, South Sandwich Islands , Taiwan,