Visualizing Country Data for the General Public – Technical Report

Please use this template table to describe and justify how you addressed the assignment and your application of visualization skills and techniques. Techniques that are correctly implemented, but not correctly described in the table will not receive full points. Please add any references that you use after the table; these should be cited where appropriate in the reporting. Your report (second column of the table) can be up to 2000 words, excluding references.

Name: Kevin Alexander **Student number:** 21014655

Visualization design	**Description of how you addressed the task**
Pre-processing: Describe and justify the data pre-processing steps carried out prior to visualization and the selection of attributes to visualize.	1. **Stripped extra spaces from column names** I used the str.strip() method to strip extra spaces from column names. I chose to do this as a precautionary measure as it ensures that there are no unexpected errors when attempting to access columns. For example, trying to access df['Unemployment'] and df['Unemployment '] could cause an error. 2. **Removed currency special characters and converted to decimal** I removed all of the currency special characters from data that had '$' prefixed at the beginning of the value. After removing the dollar sign, I converted the new values into the float data type. I chose to remove the prefixed symbols and convert the values into floats as by doing so it allows the new data to be numerically processed and plotted onto future graphs, due to the lack of special symbols. 3. **Filled missing values with the mean of the column** I filled in missing numeric values with the mean of the columns. I chose to do this in order to preserve the accuracy of the data while being able to show meaningful statistics. The absence of missing data allows graphs to be created and displayed in their entirety, while providing educated assumptions

	on what the data would have shown if it were not missing. If I were to use other methods to handle missing values such as deleting the row the missing value is associated with, it would reduce the overall dataset and would be less meaningful to the user. 4. **Filtered select columns after data cleaning** I filtered out the existing, now cleaned, dataset by the specific columns that I planned to use throughout my visualisation. This minimises the size of the final cleaned version of the dataset and increases the efficiency of the overall visualisation, as only the necessary data is being processed. 5. **Cleaned Column Names** I then cleaned the column names that I filtered out previously. I replaced all of the white spaces with underscores in between the filtered out column names (e.g. Birth Rate -> Birth_Rate). I chose to clean column names as spaces can increase complexity when referencing columns. For example instead of accessing a column using df.Birth_Rate, I would have to use bracket notation (df['Birth Rate']). I have chosen to visualise the following selection of attributes: 1. Population 2. GDP 3. Life Expectancy 4. Unemployment Rate 5. Fertility Rate 6. Infant Mortality 7. Population Density 8. + GDP per capita I have chosen to explore these attributes due to their interconnecting relevance toward one

	another, giving the user an insight into a country's social and economic state through these different attributes. The connections and correlations that these attributes have between each other allow the user to efficiently analyse the data and give them a broader understanding of the reasoning behind the data.
Fit to task: Describe how the visualization support exploration, comparison and understanding of differences between countries.	**Exploration** I have implemented the ability for the user to explore and interact with the scatterplots through a regional filter and search dropdown. This filter provides a legend that colour-codes the country by the continent they belong to and allows the user to filter the scatterplot data by continent. Additionally, the search dropdown allows the user to filter the data by a specific country, highlighting only their individual data plot on each scatter plot and on the choropleth map. I have also implemented tooltips, allowing the user to hover over plot points or countries to see additional details. **Comparison** The choropleth map allows the user to compare the population of each country through the colour scale. The darker the red, the higher the total population of the country. The colour-coded aspect allows users to compare countries visually at a glance, while also giving them more detailed values in the tooltip. 3 scatter plots: 1. GDP vs life expectancy 2. Fertility Rate vs Infant Mortality 3. Minimum Wage vs Unemployment Rate

	Understanding of differences between countries The use of the region legend helps the user understand differences in data between regions, giving the user a broader and more global perspective. Through the legend, users can view patterns and differences in how countries in different regions are doing in terms of GDP, unemployment rate…etc. The addition of the GDP per Capita attribute allows the user to have a more insightful understanding of how wealth is spread in comparison to its overall population size. This addition can highlight differences between countries that have similar total GDP, but a different population size.
Fit to user: Describe what steps have been taken to ensure that the visualization is appropriate for the intended user.	**Layout** The visualisation features one choropleth map on the left-hand side, with three scatter plots on the right. This gives the user a general overview of the world's population but also allows further exploration through the scatter plots. The visualisation title, labels and axis are appropriately sized and are clearly readable to ensure the user can understand the relationship between the attributes displayed in the graphs. **Colour** To help accommodate all types of users, I used a colour-blind safe palette for the region legend to ensure all types of users can use and interact with the visualisation. **Audience** To help make the visualisation easier to understand, I took into consideration that the user will have little to no mathematical knowledge. I converted large numeric values to words to make it as readable as possible. For

	example: 1000 to 1 thousand, or 1,000,000 million...etc.
Analysis: Describe how statistical/analytical patterns of relevance are presented in the visualization.	The scatter plots display the attributes as pairs, allowing the user to explore their relationships, differences, and patterns. For example: In my first scatter plot, I chose to pair GDP and Life Expectancy. In this scatter plot, the user can see the relationship a country has between its economic output and the life expectancy of its citizens. As GDP shows how much money the country generates, it directly influences how much they can allocate toward public resources, such as education, healthcare…etc. By analysing this scatter plot, the user can analyse the relations and draw conclusions. An example could be that having a higher overall GDP generally leads to higher life expectancy, due to having more budget to put towards public health services like hospitals, ambulances, and medicine.
Visualization design: Describe and justify your use of visual channels in the visualization.	The visualisation I have created uses a variety of different visual channels in order to effectively display data. **Colour** I have assigned specific colours that are colour-blind safe to group countries within the same continent to allow the user to easily distinguish between different regions. The colour scheme gives the user a visual cue, allowing them to highlight trends between regions. **Interaction** Through interaction, the tooltips, country dropdown and zoom allow the user to explore the data they are interested in, while potentially and unintentionally discovering extra data that they might not have initially known.

Visualization design: Describe how you have made use of Gestalt theory and design principles.	**Proximity** In my choropleth map, I have placed the Population colour scale as well as the world map closely together. This creates a strong visual connection to the user as they will automatically associate the scale with the map due to their position within each other. **Similarity** The colour coding in the region legend groups the countries into regions, ensuring that countries from the same regions are visually grouped using colour. This makes it easier for the end user to identify patterns across entire regions. **Alignment** The graphs in my visualisation are arranged in an organised and structured layout. The choropleth map is positioned on the left, and the scatter plots on the right, with each scatter plot aligning with each other. This alignment makes the dashboard feel connected.
Visualization design: Describe and justify your use of colour in the visualization.	I have grouped the countries into regions and have allocated them a corresponding colour. This colour scheme is consistent throughout the entirety of the visualisation. This enables the user to quickly identify trends not just across countries, but across entire regions. Each colour has been selected to be clearly different to allow the user to easily distinguish between them. I selected a colour palette that aligns with the Color Universal Design (CUD). On top of this, I use the Viridis colour scheme, which is also colour-blind safe. This takes those who have different types of colour-blindness into account, ensuring that users of all types are able to effectively use the visualisation.
Visualization design: Describe the interactive features used in your	When selecting a country on the choropleth map, that country's plotted data on each scatter graph will be highlighted, making the surrounding plots of data more opaque. This

visualization, and how they facilitate exploration.	highlights the user's chosen country of interest, allowing them to explore attributes across multiple graphs. Additionally, the search dropdown feature allows the user to search for a specific country that they might not be able to manually find on the map, zooming in on that particular country and highlighting only their plots of the data on the scatter plots, disabling all others. This gives the user a more focused view on that specific country, allowing them to explore that country's attributes across all graphs. Finally, when viewing scatterplots, the user can zoom in and out to examine their position and data in relation to other countries in detail.
Visualization design: Describe the design of the multiple coordinated views (dashboard) visualization, and how it facilitates exploration.	In my visualisation, I implemented 4 graphs on the dashboard total. 1 choropleth map and 3 scatter plots. These views are interconnected through the integrated interactivity to allow the user to compare, analyse and explore the visualisation. This combination of graphs facilitates exploration due to the interconnected interactivity, depth, and visuals. The interactivity in the visualisation dynamically updates all graphs depending on what the user wants to view. This facilitates exploration as it ensures that users can explore data from multiple perspectives without tedious manual input.
Visualization design: Describe considerations made in the use of language and text in the visualization.	**Readability** The legend labels are simple and informative. Additionally, the axis labels are clear and include the appropriate units of measurement in brackets. This clarifies and defines what the values are measured in, preventing potential confusion for the user. **Audience considerations** I have simplified certain aspects of the graph with the target audience in mind in order to ensure the visualisation is readable and easy to

	understand for everyone. (e.g. 1,000,000 to 1 million)
Evaluation	
Reflect on and describe how your visualization meets or does not meet the following heuristics from Wall et al. (2019): ● The visualization facilitates answering questions about the data. ● The visualization provides opportunities for serendipitous discoveries. ● The visualization provides a big picture perspective of the data.	**Facilitates answering questions about the data:** The search bar allows the user to directly filter the data in order to search for information from a specific country. This feature paired with tooltips and regional filtering allows the user to gain a deeper insight into specific countries or regions in order to answer questions they may have about particular attributes. **Opportunities for serendipitous discoveries:** The graphs have connected interactivity. When the user selects a country from the dropdown menu, that specific country's plot points and part of the map are highlighted and shown across all graphs. Even though the user may have initially selected a country to observe one particular point on a plot, the integrated interactivity with the other graphs allows the user to unintentionally observe related data in the other charts, revealing connections that the user might not have initially considered. **Big-picture perspective of the data** The choropleth map allows the user to see insightful details about countries at a global scale. The use of the region legend helps categorise information and gives a high-level view of how the different continents of the world differ socially and economically.
Reflect on and describe how your visualization could be improved to better meet the following heuristics from Wall et al. (2019):	**Facilitates answering questions about the data:** To better meet this heuristic, I could add dynamic annotations to create significant insights which could help the user interpret and

• The visualization facilitates answering questions about the data. • The visualization provides opportunities for serendipitous discoveries. • The visualization provides a big picture perspective of the data.	explore the data more efficiently. An example of this could be “Top 5 countries by GDP per capita”, or “Regions with the highest unemployment rate” **Opportunities for serendipitous discoveries:** --- I could improve my visualisation to better meet this heuristic by adding a feature to display random facts based on the data displayed. For example “Did you know Country X has one of the highest GDP, but low fertility rates?”. By prompting the user with information they might not know, they are encouraged to explore the sections they might not have initially. **Big-picture perspective of the data** --- To improve this, I could implement regional comparisons to display regional statics in the dashboard in order to give the user a high-level overview of the data, allowing the user to further explore if they wish.

References:

- Color Universal Design (CUD) - How to make figures and presentations that are friendly to Colorblind people, https://jfly.uni-koeln.de/color/
- Visme, https://visme.co/blog/interactive-data-visualization/
- ThoughtSpot, https://www.thoughtspot.com/data-trends/data-visualization/interactive-data-visualization