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: Kristen Karsburg Arguello Student number: 240697242

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.	The pre-processing transformed the dataset into a Pandas DataFrame with defined data types. Monetary columns with "$" and commas were converted from "object" to float by removing these symbols, and titles were updated to include "(in $)". Non-numeric "object" columns were replaced with spaces. The prefix "population" was removed from the labour force participation column, and other names were simplified. An additional continent column was created using the "pycountry_convert" library for easier grouping in the visualisation. Rows with missing values in the "Continent" column were dropped. In contrast, other missing values were handled separately for each graph by creating secondary data frames with rows containing missing values in relevant attributes removed. These changes were saved into a new CSV file used for visualisation creation. The visualisation analysed each country's economic power using mainly GDP and GDP per capita (calculated using GDP and population). Life expectancy and physicians per thousand were used to compare health impacts concerning economic power, showing how wealth can influence these values. Urban population percentage (calculated using urban population and total population), density, latitude, and longitude were used to map urbanisation, illustrating population distribution trends and allowing comparison with other metrics. Land area, CO2 emissions, and GDP were examined to assess the impact of country size on these attributes, highlighting economic power and pollution trends.
Fit to task: Describe how the visualization support exploration,	It integrates diverse metrics highlighting each country's economy and impacts on health,

comparison and understanding of differences between countries.	pollution, and urbanisation. It facilitates the exploration of urbanisation through urban population percentage, density, latitude, and longitude. By enabling direct comparisons, it shows how economically powerful countries impact CO2 emissions and highlights health disparities in less urbanised and wealthy countries, illustrating trends between wealth and these characteristics. The data-centred visualisation provides insights into the complex relationship between economic power, environmental impact, and health outcomes. With interactivity, users can highlight and filter data across the visualisation, offering a comprehensive view of country aspects globally and general information related to each continent.
Fit to user: Describe what steps have been taken to ensure that the visualization is appropriate for the intended user.	To ensure the visualization suits non-specialists interested in comparing countries and continents, several steps have been taken. The design utilizes visual variables from Bertin et al.'s semiology of graphs (2011) [1], such as positioning countries on a map using latitude and longitude, making it easier for users to identify each country and its location relative to others. It also employs size and colour to create contrast and effectively display information and data ranges. Data can be aggregated and filtered by continent to provide a broader understanding and highlight regional trends, or users can highlight specific continents among others to easily identify regions. The visualization offers an immediate understanding of its global focus through elements like maps, which indicate the international scope of the data. A title is included to clarify the visualization's goal and how selections work, aiding user comprehension. With multiple graphs displaying different attributes, users can compare data to identify economic trends across countries and continents. For example, Asia is noted as the continent with the second-highest average GDP, despite having many countries with low urban population percentages, high life expectancy, and significant CO2 emissions. Emphasis is placed on exploration, allowing users to interact with the data and explore different highlighting options to gain insights without needing technical expertise.

Analysis: Describe how statistical/analytical patterns of relevance are presented in the visualization.	A correlation matrix and a bar chart of averages are shown to present statistical and analytical patterns. The correlation matrix numerically illustrates the impact of one attribute on another, providing analytical insights into how various graphs function by highlighting relationships between selected attributes, and indicating suggestions on what attributes the user should compare to gain knowledge. Additionally, the bar chart focuses on the average GDP of each continent, allowing for a comparative economic statistical analysis of countries within the same continent. This chart is interactive, enabling users to highlight data by clicking on continent values, thus enhancing the analytical depth and user engagement with the economic data.
Visualization design: Describe and justify your use of visual channels in the visualization.	Maps used latitude and longitude to display population distribution and urbanisation patterns, employing choropleth and bubble techniques for equitable visual comparison, with bubble size representing magnitude. The map highlighted urban area concentrations geographically. Scatter plots, ideal for showing relationships between attributes, used axes for independent and dependent variables, with colour and size adding depth, such as illustrating GDP's influence on life expectancy (Y-axis) and the number of physicians (colour and size). Another scatter plot showed the relationship between land area and GDP, with CO2 emissions as an extra attribute, highlighting environmental impacts through bubble colour and size. A correlation matrix identified inter-variable relationships, while bar charts highlighted specific statistical patterns. These choices enhanced data clarity and interpretability, revealing key patterns and relationships.
Visualization design: Describe how you have made use of Gestalt theory and design principles.	Gestalt psychology suggests that human perception is influenced by both present stimuli and our motivations and expectations, introducing design principles that aid in building information visualisation, as noted by Kafka et al [3]. In scatter plots, the principle of proximity groups elements with similar values, revealing countries with similar economic, pollution, and health characteristics. Similarity is applied through colour and size, allowing quick comparison of countries with the same number of physicians per thousand, categorising by

	urbanisation percentage and CO2 emission rates, or creating a hierarchy of GDP averages using hue and value for contrast. Symmetry and alignment in plot layout provide a balanced structure, making the visualisation approachable and easy to navigate. Different colours were chosen for various categories to avoid asymmetry and provide contrast, legends were placed on the right side, and size and colour ranges were standardised. Attributes in different graphs share the same formatting style, and other symmetric design choices were made. When filtering data, domain of scales was kept consistent to ensure repetition with different data manipulations, but scaling values and comparison change according to the filtered dataset; scales, domains, opacity, and stroke widths were repeated for similar plots. These principles create a cohesive and intuitive visualisation, making complex data more accessible and readable for the end user.
Visualization design: Describe and justify your use of colour in the visualization.	Distinct colour schemes were chosen to enhance clarity, ensuring consistent colours for the same categories and distinct ones for different categories. Variations in hue and value keep colours distinguishable in greyscale and for those with colour vision deficiencies, tested with filters. Contrasting colours like blue and yellow were selected over red and green to accommodate common colour blindness, avoiding colours with implicit meanings to prevent misinterpretation. Popout elements were highlighted with colours, while others were greyed out to emphasise a single feature, such as a country or continent. Scaling size helped interpret values when colour differentiation was subtle, while size was used for different attributes when colour variation was clear, such as on maps. Quantitative data used sequential or diverging schemes, designed for colour blindness and black-and-white printability.
Visualization design: Describe the interactive features used in your visualization, and how they facilitate exploration.	A continent selection feature is activated when the user clicks on the average GDP bar for a continent, highlighting all countries from that continent in other graphs using different colours, stroke widths, and opacities. Similarly, a country selection feature is activated by clicking on any point representing a country, highlighting its data across all graphs and indicating its continent, streamlining the

	exploration process. Tooltips reduce label clutter, keeping the visualisation clean and informative. A dropdown menu lists continents for filtering, allowing users to view data specific to a selected continent. The y-axes of scatter plots are interactive and can be zoomed in for detailed data inspection. In summary, highlighting enables analysis of data related to specific continents while comparing it to the rest of the dataset; zooming in allows detailed viewing of scatter plot values; and filtering facilitates the analysis of regional trends without unnecessary extra data.
Visualization design: Describe the design of the multiple coordinated views (dashboard) visualization, and how it facilitates exploration.	The dashboard features multiple coordinated views to explore how economic power influences country structures and their consequences. It integrates key visual analysis tasks, such as providing an overview, filtering, highlighting, and offering details on demand. The interconnected views ensure that interactions in one view, like country or continent selection, automatically update and highlight corresponding data in other views through dynamic querying. A dropdown filters data across all graphs by continent. Consistency is maintained with repeated visual elements, like circles of varying sizes, colours, and stroke widths, helping interpretation. This coordinated design facilitates seamless exploration of complex data relationships, enabling users to understand the interplay between economic power and its impacts effectively. Complementary views, including maps, scatter plots, and bar charts, provide a comprehensive analysis of each country's economic power, showing different attributes in each plot that individually convey important information but collectively offer a big picture. Averages and correlations provide additional support for interpretation with statistical and analytical analysis. When viewed together, all designed plots can enhance insights into economic power and its impacts on each country, showing the amount of money produced, health data, pollution contribution, and urbanisation percentage in one coordinated place, all crucial for defining a country's development level.
Visualization design: Describe considerations made in the use of	Labels with straightforward names were used to convey the meaning of each value, ensuring easy data interpretation. A title for the dropdown selection clarifies its function. Each

language and text in the visualization.	graph is titled to indicate the focus of the analysis and the variables included, providing context and guiding the viewer's attention. Additionally, values were formatted to prevent numbers from appearing too large, making the data more accessible and easier to read. A main title and subtitle were added to clarify the dashboard's main goal and how to use the multiple interaction features. This approach ensures that the visualisation communicates information effectively and intuitively.
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.	The tool facilitates answering questions through interactive elements like continent and country selection, tooltips, and filtering, enabling effective data navigation. Zooming in on scatter plot axes and using tooltips allows users to view values clearly. Texts, colours, sizes and other design principles were all applied to facilitate answers. Since the dataset contains data about each country of the world, the visualization can be too cluttered making it hard to answer questions. Still, the starting point of the Y-axis and interactiveness try to account for that by enabling the user to filter and zoom in on data to make it more viewable. The interactive nature of the visualisation encourages exploration and the discovery of unexpected insights. Scatter plots reveal clusters of values, indicating trends and providing opportunities for serendipitous discoveries. The visualisation offers a comprehensive perspective by aggregating data at the continent level and using maps to display global patterns. It integrates various economic, health, and environmental metrics, offering a holistic view of each country's characteristics within a global context. Individual graphs provide insights into specific attributes, and when viewed together, they offer a broader perspective. Filtering by continent helps focus on relevant data, while aggregation per continent highlights the big-picture country data. Although the proposed heuristics were achieved, there are still areas for improvement. Minor implementation issues, such as non-

	seamless selection, a scatter plot legend not representing all shown colours, data too cluttered in the visualization and other small "bugs," need to be addressed to enhance performance.
Reflect on and describe how your visualization could be improved to better 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.	First, other and more advanced filtering options, such as sliders for numerical data and interval selection, would allow users to refine their queries and focus on specific subsets of data. First, incorporating more advanced filtering options, such as sliders for numerical data and interval selection, would allow users to refine their queries and focus on specific data subsets. Implementing a search function to locate countries among all data points would enable users to quickly find specific information. Adding dynamic annotations to highlight key insights or trends could guide users to relevant observations as they interact with the visualisation. Additionally, specific issues need addressing, such as ensuring all colours are included in the scaling legend for clarity, providing an indicator for countries excluded due to insufficient data and thinking of a way to make the data less cluttered for better visualization. To encourage serendipitous discoveries, exploratory prompts or questions could suggest new ways to interact with the views, motivating users to explore different dimensions and discover new ideas. For example, a statement about poorer countries contributing less to global air pollution could, when clicked, highlight the relevant countries. Finally, to improve data full comprehension, creating more analytical and summarization of the data showing key metrics and trends would provide users with an immediate and bigger picture overview of the landscape. Also, adding narrative storytelling elements and other extra information can walk users through the large amount of data present on the dashboard, highlighting major patterns and insights to ensure they understand the broader context. Additionally, comparative views of data through time can help the users gain better insights and see the big picture across different periods; also helping the users to understand how various factors interrelate globally.

References:

1 Bertin, Jacques. (2011). General Theory, from Semiology of Graphics. The Map Reader: Theories of Mapping Practice and Cartographic Representation. 8-16. 10.1002/9780470979587.ch2.

2 Cherry, K. (2024) *Gestalt psychology: A holistic approach*, *Verywell Mind*. Available at: https://www.verywellmind.com/what-is-gestalt-psychology-2795808 (Accessed: 01 December 2024).

3 Fung, H. (2022) *Data storytelling: Making sense of complex, multi-dimensional data with parallel coordinates plots*, *Medium*. Available at: https://medium.com/@hokifung_/data-storytelling-making-sense-of-complex-multi-dimensional-data-with-parallel-coordinates-plots-e73ddb1eb7f4 (Accessed: 02 December 2024).

4 Koffka, K. (1935). Principles Of Gestalt Psychology (1st ed.). Routledge. https://doi.org/10.4324/9781315009292 (Acessed: 04 December 2024)

5 Midway, S.R. (2020) 'Principles of Effective Data Visualization', *Patterns*, 1(9), p. 100141. doi:10.1016/j.patter.2020.100141. (Acessed: 04 December 2024)
6 *Style Guide for Dashboards* (no date) *MySidewalk*. Available at: https://dashboards.mysidewalk.com/style-guide-for-dashboards/color (Accessed: 02 December 2024).

7 Wall, E., Agnihotri, M., Matzen, L., Divis, K., Haass, M., Endert, A., & Stasko, J. (2019). A heuristic approach to Value-Driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 491–500. https://doi.org/10.1109/tvcg.2018.2865146

8 Wang C, Shen H-W. Information Theory in Scientific Visualization. *Entropy*. 2011; 13(1):254-273. https://doi.org/10.3390/e13010254 (Accessed: 30 November 2024).

9 *What are the gestalt principles? - updated 2024* (2024) *The Interaction Design Foundation*. Available at: https://www.interaction-design.org/literature/topics/gestalt-principles#:~:text=Gestalt%20principles%20or%20laws%20are,the%20separate%20simpler%20elements%20involved (Accessed: 02 December 2024).