Visual Analysis of the Ocean Microbiome

This is the submission table to be used as a report for the summative assessment in CSC8636 – Complex Data Visualization. Fill in your comments and answers in the table below, as lined out in the Summative Assessment description. You should include a list of references to sources and literature you have used, and cite them appropriately in your answers. Submit this document in pdf format together with your visualization (as html page), Python code and any datasets that are loaded by the code, as a single zip file in Canvas. ***The submission deadline is Friday 21st February.***

Student name: Amey Madane **Student ID: 240698836**

Part 1 – Interactive visualization using multiple coordinated views	**Your comments and answers**
Fit to user: Describe and justify the steps that have been taken to ensure that the visualization is appropriate for the intended user.	To ensure the visualization is appropriate for the intended user, the dashboard is designed for a **1920x1080** resolution, ensuring accessibility for researchers and students working with microbiome datasets. It features a single-page layout to minimize navigation and enhance user experience. Key design choices include hierarchical taxonomic visualization for analysing microbial diversity, interactive filtering by ocean layers **(SRF, DCM, MES**), and accessibility features such as distinct colour schemes and readable labels. Additionally, the dashboard is built for scalability, efficiently handling large datasets without performance issues.
Visualization design: Describe and justify how you have made use of Gestalt theory and visualization design principles.	The visualization adheres to **Gestalt principles** by implementing: Related charts are grouped together to show connections clearly. Smooth transitions help reveal trends and patterns. A consistent colour scheme is used to distinguish different microbial groups and sample layers. Clear contrasts and interactive tooltips highlight important details, making it easier to focus on key insights.
Visualization design: Describe the interactive features used in your visualization, and how they facilitate exploration.	The visualization is interactive, making it easy to explore the data. When you hover over a chart, tooltips show exact values for better clarity. You can also filter the data by selecting specific microbial groups from the legend. Zooming and panning also allow you to explore the trends in more detail.These allow researchers to study microbial distribution in depth and pick up patterns more easily.
Visualization design: Describe the design of the multiple	The dashboard consists of **6 visualizations**, grouped into three rows: 1. **Abundance Distribution (Bar Chart):** Displays the most abundant microbial phyla across all samples.

coordinated views (dashboard) visualization, and how it facilitates exploration.	2. **Abundance Scatter Plot:** Shows how microbial abundance varies across ocean layers. 3. **Relative Taxonomic Distribution (Stacked Bar Chart):** Visualizes the composition of microbial communities by taxonomic group. 4. **Co-Occurrence Network:** Represents relationships between microbial phyla and ocean layers. 5. **Taxonomic Flow (Sankey Diagram Alternative):** Shows relative abundances of microbes across layers. 6. **Abundance Variation Plot:** Allows users to compare abundance distributions between samples. By placing these visualizations horizontally in rows, users can compare relationships across different data views without excessive scrolling 1. **Row 1**: Abundance Distribution & Abundance Scatter Plot (with dropdown filter) 2. **Row 2**: Relative Taxonomic Distribution & Microbial Co-Occurrence Network 3. **Row 3**: Taxonomic Flow & Enhanced Abundance Variation The layout facilitates coordinated exploration, ensuring that different perspectives on microbial abundance are easily accessible.
Part 2 - Uncertainty	
Uncertainty sources: Describe potential sources of uncertainty that may exist in the data.	There are several uncertainties in the dataset that can affect the analysis. Since the samples are collected from specific ocean locations, they may not reflect microbial life at all places. Microbial communities also vary naturally due to ocean conditions and seasons. Some microbes are yet to be discovered, hence placeholder names are listed in the data. Moreover, certain ocean regions and depths may be under-sampled, and sequencing methods may add variability to abundance estimates
Uncertainty visualization: Describe how you could adapt your visualization in part 1 to represent this uncertainty.	To better represent uncertainty in data visualization, certain techniques can be used. Placing error bars on top of abundance values can show measurement variability, and adjusting opacity levels can have uncertain data displayed as a transparent overlay. Violin plots can be employed to show variation in distribution, and filled areas can show anticipated uncertainty ranges in bar or scatter plots. Transparency can also be used to indicate confidence in taxonomic identification. Error bars may also reflect variability in measurements of abundance, and completeness markers for data may reflect proportions that are missing at different taxonomic levels. Annotations finally may serve to mark likely areas of bias or missingness.
Part 3 – Heuristics evaluation	**For each heuristic below, describe how your visualization in part 1: a) meets or does not meet the heuristic;**

	b) could be changed to better meet the heuristic.
Evaluation: The visualization facilitates answering questions about the data.	**Meets:** The visualization allows users to easily identify the most abundant microbes and spot patterns across different ocean layers. Coordinated views make it simple to compare taxonomic groups and sampling conditions. Additionally, filtering options and tooltips provide quick answers to specific questions about microbial abundance. **Improvement:** Including search functionality for filtering microbial taxa by name could enhance usability.
Evaluation: The visualization provides a big picture perspective of the data.	**Meets:** The multiple coordinated views offers a comprehensive overview of ocean microbiomes, highlighting the distribution patterns at different taxonomic levels**.** **Improvement:** Introducing summary statistics (e.g. mean abundance per layer) within tooltips can provide some additional high level insights.
Evaluation: The visualization helps avoid making incorrect inferences.	The tool addresses the requirements by applying log scale transformationsto avoid incorrect interpretation of asymmetric abundance data, and tooltips provide accurate numerical values to minimize errors in estimation. **Improvement:** adding uncertainty annotations would improve the tool by helping users better understand and avoid misinterpreting fluctuating data values.

List of references:

- Altair Documentation: https://altair-viz.github.io/
- Canvas practical exercises slides.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.