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: **Student ID:**

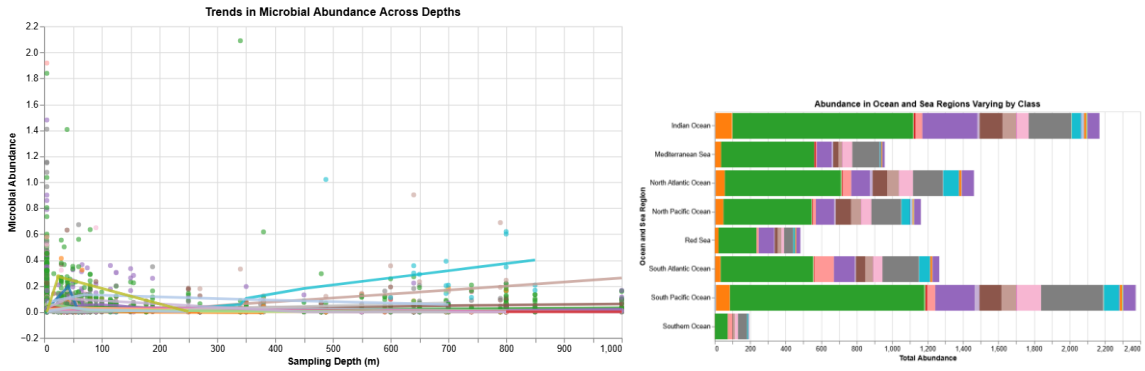
Part 1 – Interactive visualization using multiple coordinated views	**Your comments and answers**
Fit to user: Describe and justify the steps that have been taken to ensure that the visualization is appropriate for the intended user.	The visualization is designed specifically for **marine microbiologists** and oceanographers to explore microbial abundance in different ocean regions, depths, and over time. **Interactive selection** (class selection) enables dynamic updates across all views, improving usability. A **single universal legend** ensures color consistency across all views, reducing cognitive load and improving interpretability. The visualization follows accessibility principles, **using distinguishable colors** and maintaining clarity for different user expertise levels.
Visualization design: Describe and justify how you have made use of Gestalt theory and visualization design principles.	**Proximity & Similarity**: Related charts are placed near each other with consistent colors for microbial classes. **Continuity & Pre-attentive Processing**: Color encoding help users quickly grasp trends. **Use of appropriate chart types**: Stacked area charts, maps, and scatter plots provide **high-level summaries** and allow **detailed exploration**.
Visualization design: Describe the interactive features used in your visualization, and how they facilitate exploration.	Class Selection: Clicking on a microbial class updates all visualizations simultaneously. Hover Tooltips: Provides contextual information without cluttering the interface. Scatter Plot: Allows users to explore microbial classes externally by clicking for additional information. Geographical Mapping: Users can visualize where specific microbes are more abundant globally.

	Time-Series Interactivity: Enables tracking abundance changes across different years.
Visualization design: Describe the design of the multiple coordinated views (dashboard) visualization, and how it facilitates exploration.	Abundance Trend Chart & Bar Chart → This enables users to compare microbial depth trends and geographic distributions side by side. Geographic Distribution Map→ This gives a spatial perspective of microbial presence throughout ocean regions. Time Series & Scatter Plot → Enables tracking of abundance changes over time while connecting to external data through clickable points. All visualizations are used for a specific purpose but are connected through class selection. Reasonable positioning of views makes it easy for users to move from overall trends (top) to location-based insights (middle) to close class-specific exploration (bottom).
Part 2 - Uncertainty	
Uncertainty sources: Describe potential sources of uncertainty that may exist in the data.	Sampling Bias: Certain ocean regions may have more sampling coverage than others, affecting data representation. Measurement Variability: Different instruments and techniques may introduce fluctuations in microbial abundance measurements. Missing Data: Some microbial classes have incomplete records due to environmental or technical limitations.
Uncertainty visualization: Describe how you could adapt your visualization in part 1 to represent this uncertainty.	Shaded Confidence Intervals: Add uncertainty bands to microbial abundance trends in line and scatter plots. Error Bars: Show possible variability in scatter plots and bar charts. Opacity-Based Encoding: Reduce opacity for data points with low confidence or missing values. Tooltips with Data Completeness: Indicate % completeness for each microbial class.
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.	**Satisfies This Heuristic:** Interactive legend enables the isolation of microbial classes and examination of abundance patterns. Tooltip hover effect provides timely context, limiting external explanation requirements. **How to Improve:**

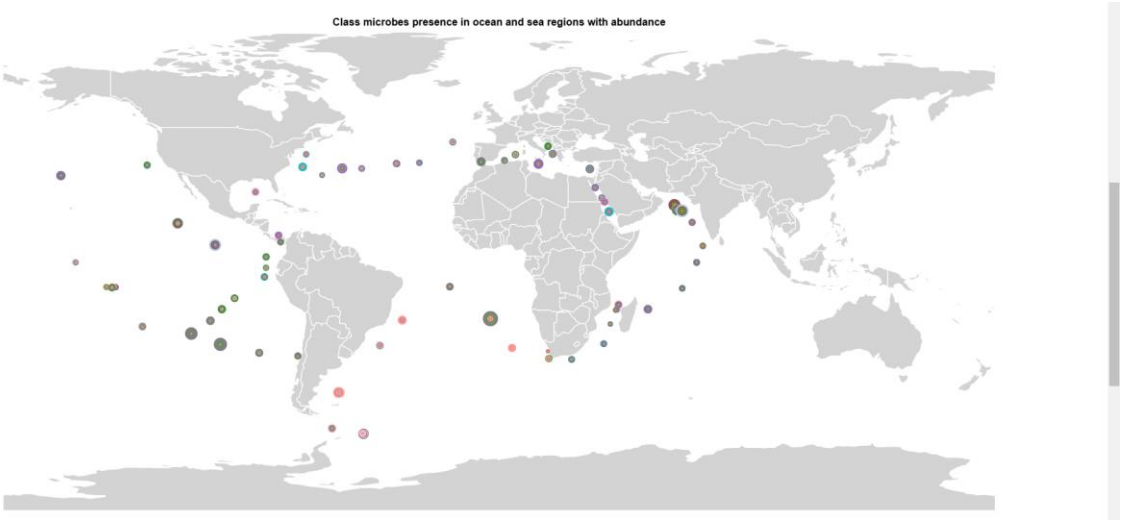
CSC8636 – Summative Assessment: Submission Table

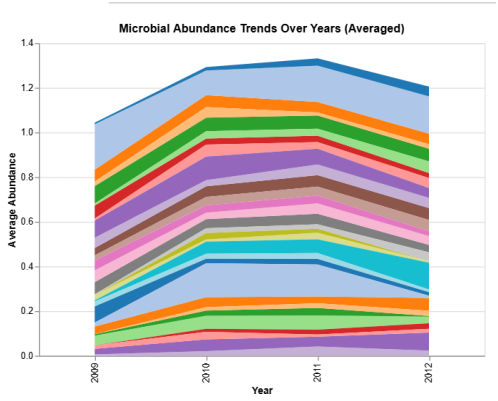
	Add a search input for microbial classes to facilitate instant retrieval of microbes. Use annotated highlights on master charts to guide users to intriguing results.
Evaluation: The visualization provides a big picture perspective of the data.	**Satisfies This Heuristic:** The stacked bar chart and time-series views allow users to understand microbial distribution over time and space. The map view relates abundance to geography, providing a worldwide outlook. **How to Improve:** Introduce summary statistics (e.g., median abundance) in tooltips to reinforce macro-level insights. Allow users to toggle between absolute and relative abundance in visualizations.
Evaluation: The visualization helps avoid making incorrect inferences.	**Satisfies This Heuristic**: The Loess smoothing trend line discourages users from misinterpreting random variation as meaningful trends. The normalized stacked area chart in the time-series prevents misinterpretation of raw abundance values. **Improvement Steps**: Add a warning note near the class legend stating that there is missing taxonomic data for certain microbial classes. Add tooltips explaining data collection limitations, reducing the possibilities of incorrect assumptions.

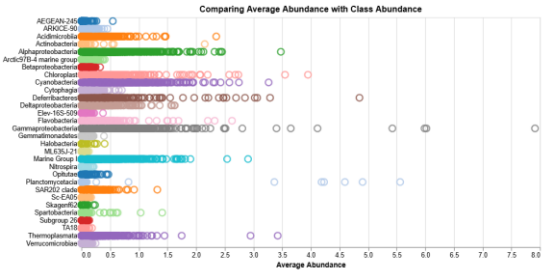
Below are my visualization:
Note: Loading dashboard might take few minutes.



CSC8636 – Summative Assessment: Submission Table







1. Project Overview

This project presents an **interactive dashboard** that visualizes microbial abundance patterns in oceanic regions. The dashboard provides insights into microbial diversity, spatial distribution, and temporal trends using **multiple coordinated views**.
The dashboard is built using **Altair**, a declarative visualization library in Python, and integrates interactivity for enhanced exploration of the Tara Ocean dataset.

2. Features & Functionalities

Key Features:

- Multiple Coordinated Views: Different visualizations are linked via a universal class selection legend.
- Dynamic Interactivity: Clicking on a microbial class updates all views.
- Geospatial Analysis: A world map representation shows the microbial distribution across the oceans.
- Temporal Trends: A stacked area chart visualizes microbial abundance trends over the years.
- Scatter Plots with Tooltips: Provides point-level insights.
- Smooth Trend Lines: Loess smoothing captures patterns in microbial abundance across depths.

3. Data Processing & Preprocessing

Data Preparation Steps:

- Convert relevant columns to numerical data (samplingdepth, abundance).
- Handle missing values and normalize data where required.
- Aggregate data for regional and temporal analysis.
- Assign meaningful color scales for microbial class representation.

4. Visualization Design Principles

The dashboard helps to Gestalt Theory and visualization best practices:

Principle	Application
Proximity	Related views are placed near each other for logical flow.
Color Consistency	A single legend controls all views, preventing confusion.
Interactivity	Users can filter by microbial class dynamically.
Hierarchy	The most important views appear at the top for easy access.

5. User Interactivity & Controls

Legend Selection

- Users can click on a microbial class to update all visualizations.
- The legend is unified across all charts to maintain a consistent color scheme.

Tooltip Hover Insights

- Provides detailed information on individual points.
- Applied to scatter plots, bar charts, and the geospatial map.

Projection in Map View

- Users can explore different ocean regions dynamically.

6. Visualizations in the Dashboard

Bar Chart (Abundance in Ocean & Sea Regions)

- Displays the total microbial abundance by region.
- Uses stacked bars to compare different microbial classes.

Geographic Distribution (World Map)

- Represents microbial presence spatially.
- Color-coded by microbial class, with hover tooltips.

Scatter Plot (Depth vs. Abundance)

- Shows microbial abundance at different ocean depths.
- Includes Loess smoothed trend lines to highlight patterns.

Normalized Stacked Area Chart (Time Series)

- Displays relative microbial abundance over time.
- Helps in identifying long-term patterns.

Interactive Scatter Plot

- Each point represents a microbial class.
- Clicking a point opens external sources for further study.

7. Limitations & Future Improvements

Current Limitations:

- Some microbial classes have low representation, making trends less visible.

- Certain uncertainties in sampling data are not explicitly visualized.

Potential Enhancements:

Uncertainty Visualization: Add confidence intervals to indicate reliability.
More Interactivity: Allow filtering by multiple parameters.
Expanded Data Insights: Add statistical summaries in tooltips.

8. Conclusion

This interactive dashboard serves as a powerful tool for marine microbiologists and researchers, enabling them to explore microbial abundance across depth, location, and time. By leveraging coordinated views and interactivity, the dashboard provides actionable insights while maintaining an intuitive and engaging user experience.

List of references:

1. Shneiderman, **"The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations"**, IEEE Symposium on Visual Languages, 1996.
2. D. Sarkar, **"The Art of Effective Visualization of Multi-dimensional Data"**, Available: https://medium.com
3. https://altair-viz.github.io/gallery/index.html
4. Wilke, **"Fundamentals of Data Visualization"**, O'Reilly Media, Available: https://clauswilke.com/dataviz/