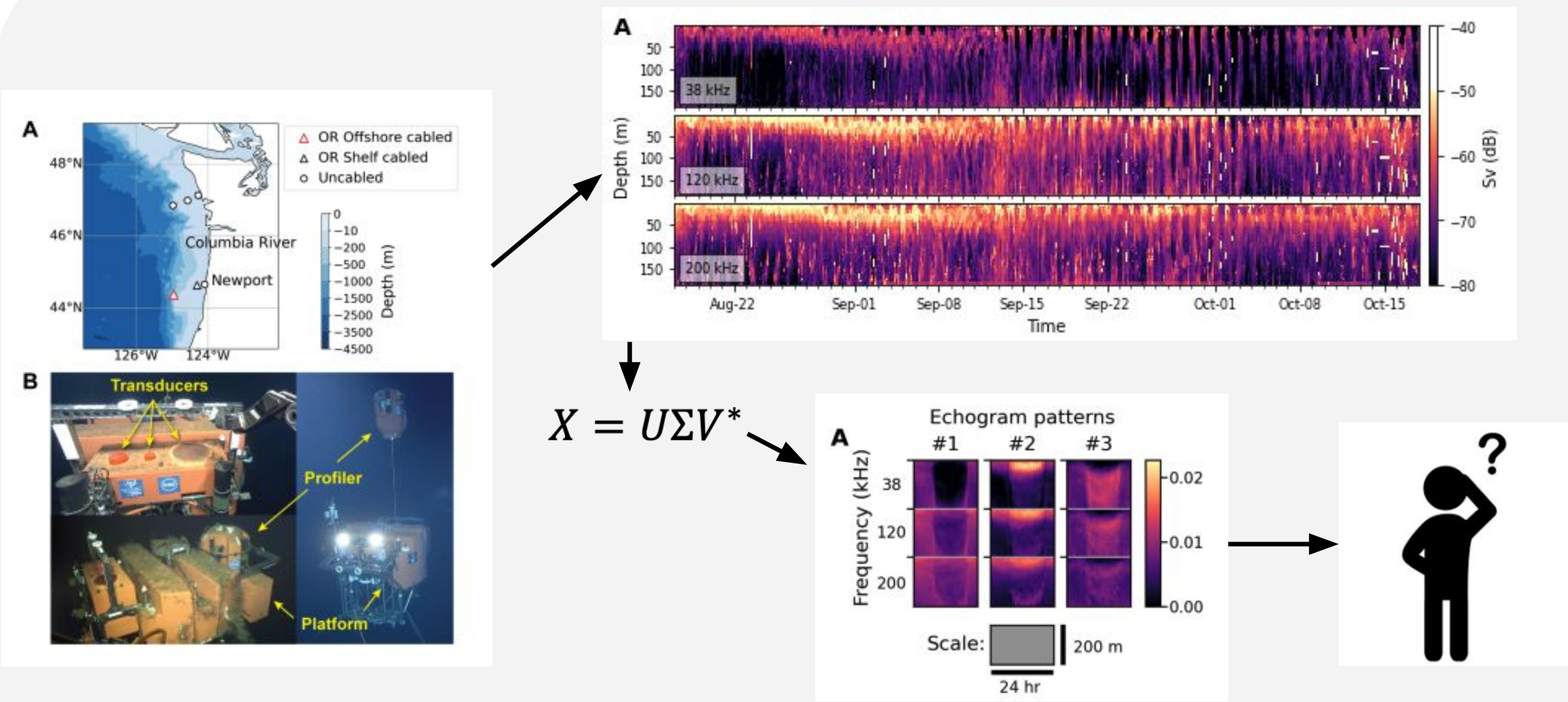
A Tool for building Intuition about Principal Component Analysis

Xinran Wang, Eason Zhang, Ruican Zhong, and Sophia Jannetty

UW CSE512: Data Visualization – Spring '22

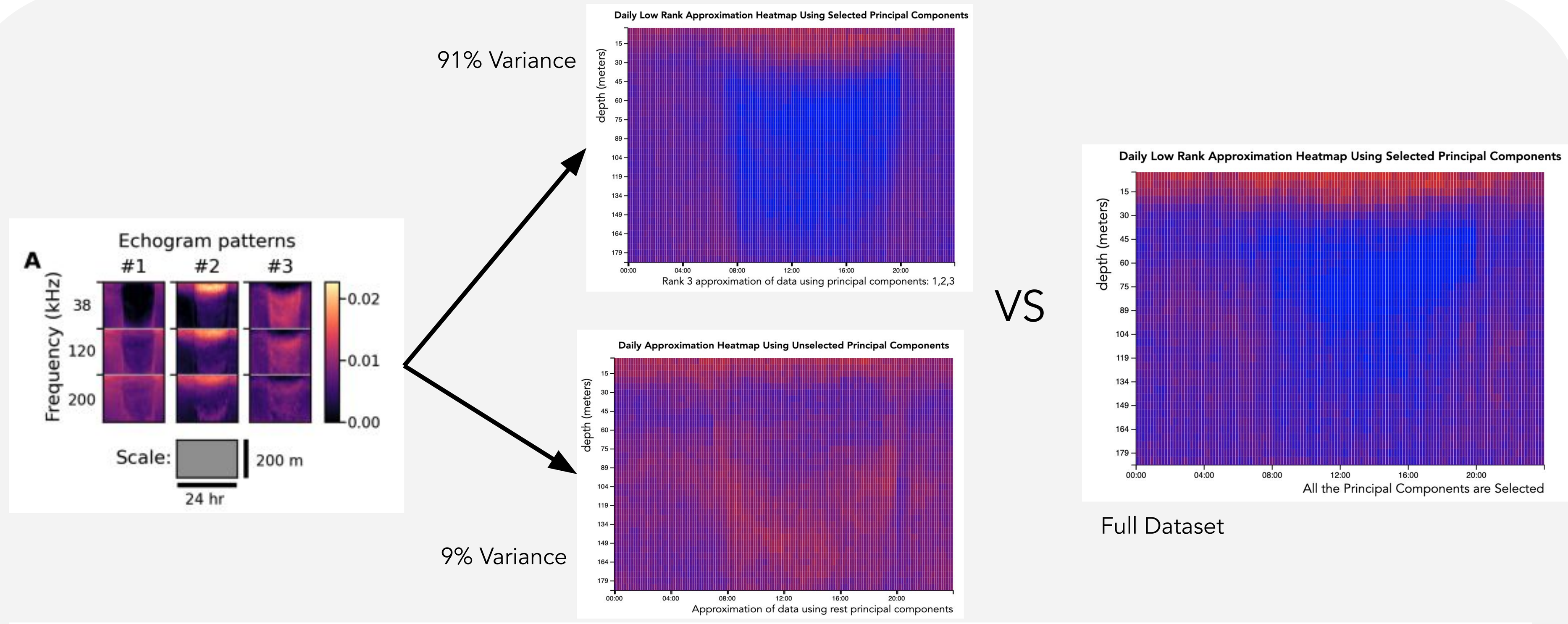
Motivation



Lee, Wu-Jung, and Valentina Staneva. 2020. "Compact Representation of Temporal Processes in Echosounder Time Series via Matrix Decomposition." The Journal of the Acoustical Society of America 148 (6). United States: 3429–42. doi:10.1121/10.0002670.

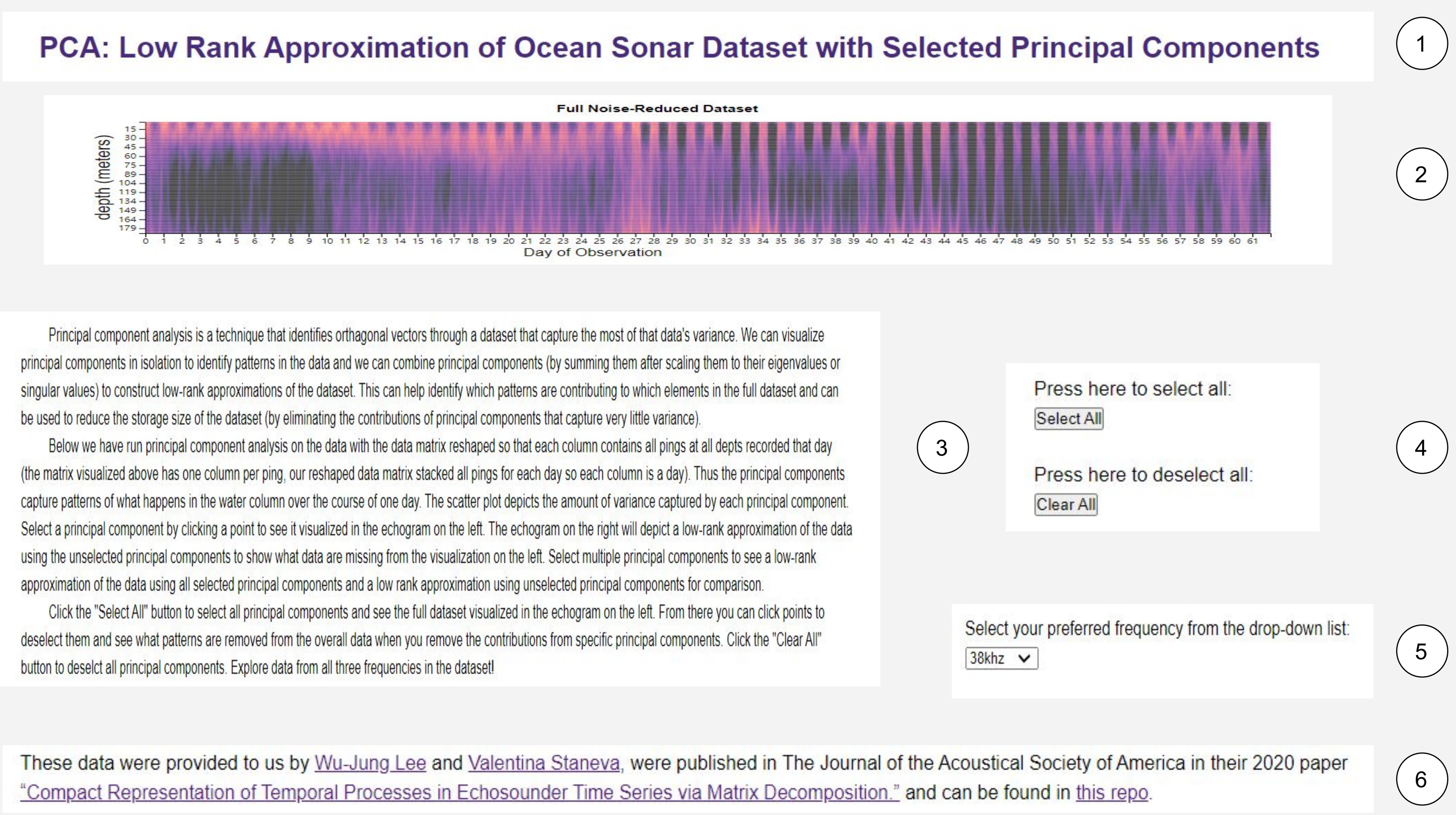
Principal component analysis is a powerful dimensionality reduction technique that is useful for extracting temporal patterns in ocean sonar time series data. Current visualization methods visualize components in isolation or low-rank approximations of datasets. This makes it difficult for researchers to get a full picture of what is happening in the dataset and makes it difficult for students to understand the meaning of individual components.

Design Goals



We wanted to create a visualization that helps ***students*** understand how principal components combine to make low rank approximations of data, and how low rank approximations constructed from principal components that capture most variance in the data more closely resemble the full dataset than approximations from principal components capturing less variance. We also wanted to let ***ocean researchers*** explore the principal components of daily data in Wu-Jung and Staneva's echosounder water column data.

Evaluation/Feedback

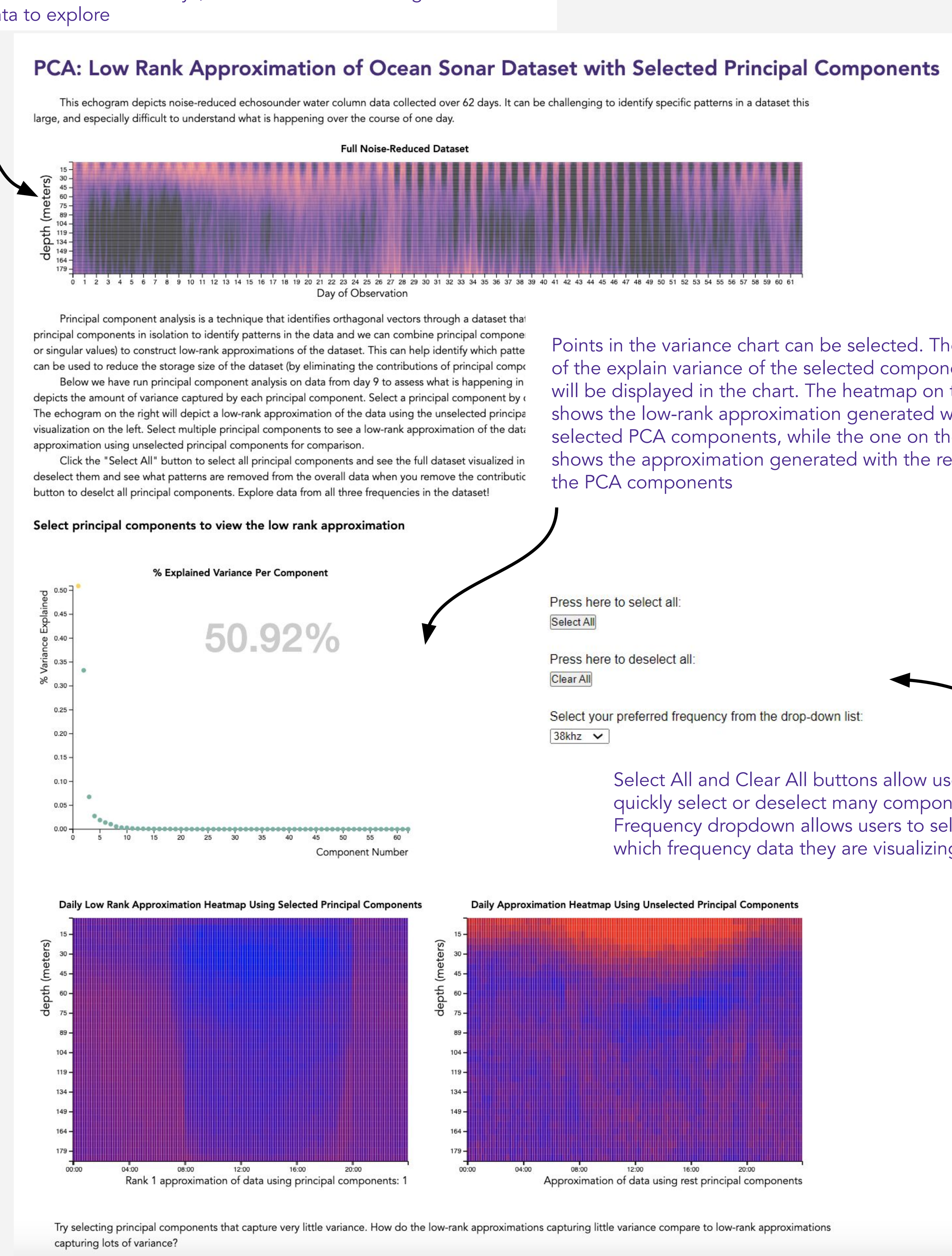


In response to feedback from Dr. Staneva we made the following changes:

1. Included "ocean sonar" or "echosounder water column" in the title to be more explicit about what is used in the scientific community
2. Added a static visualization displaying all data across the 62 days to provide an overview
3. Added more explanations to clarify what data is being visualized
4. Implemented the "Select All" and "Clear All" buttons to ensure easy selection. Specifically, the "Select All" button provides the option to quickly select and visualize the entire dataset
5. Included dropdown menu to select and visualize different frequencies
6. Added references to the ooi-echo-matrix-decomposition repository

Approach

The top graph displays the noise-reduced echosounder water column data over 62 days, but it is an overwhelming amount of data to explore



Points in the variance chart can be selected. The sum of the explain variance of the selected components will be displayed in the chart. The heatmap on the left shows the low-rank approximation generated with the selected PCA components, while the one on the right shows the approximation generated with the rest of the PCA components

Select All and Clear All buttons allow users to quickly select or deselect many components. Frequency dropdown allows users to select which frequency data they are visualizing

Thanks to Valentina Staneva Ph.D and Wu-Jung Lee Ph.D for supporting this project!

PAUL G. ALLEN SCHOOL

Link to the Interactive Visualization:
https://cse512-22sp.pages.cs.washington.edu/Sonar-Principal-Components/