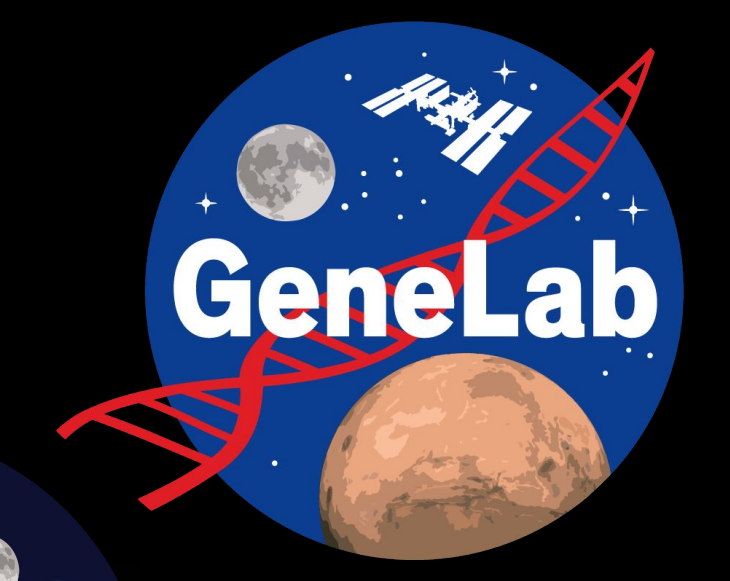


Who, What, Where, When, and Why: Analyzing Metrics for OSDR Website and Datasets to Determine Engagement

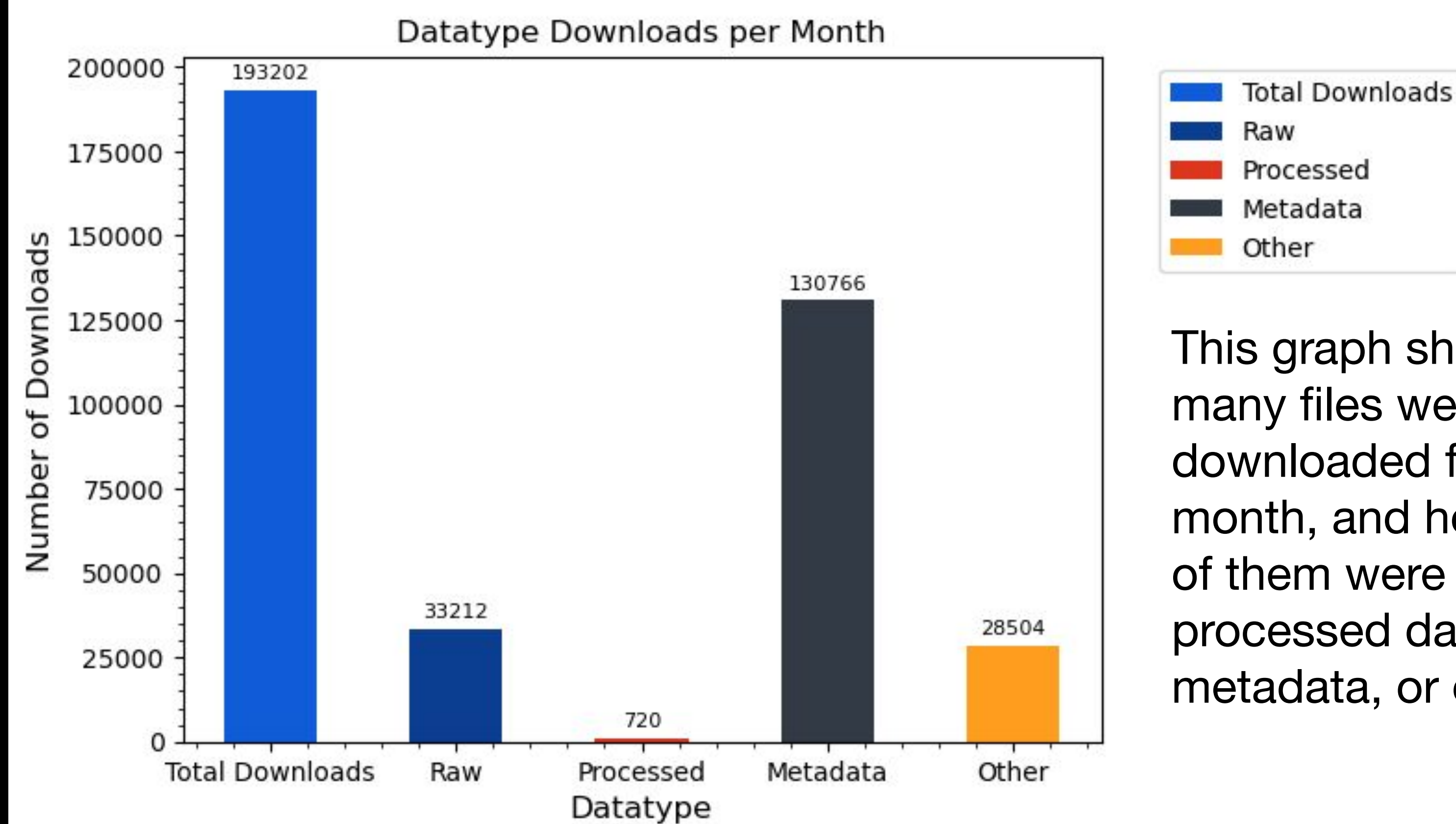


Nova Butte¹

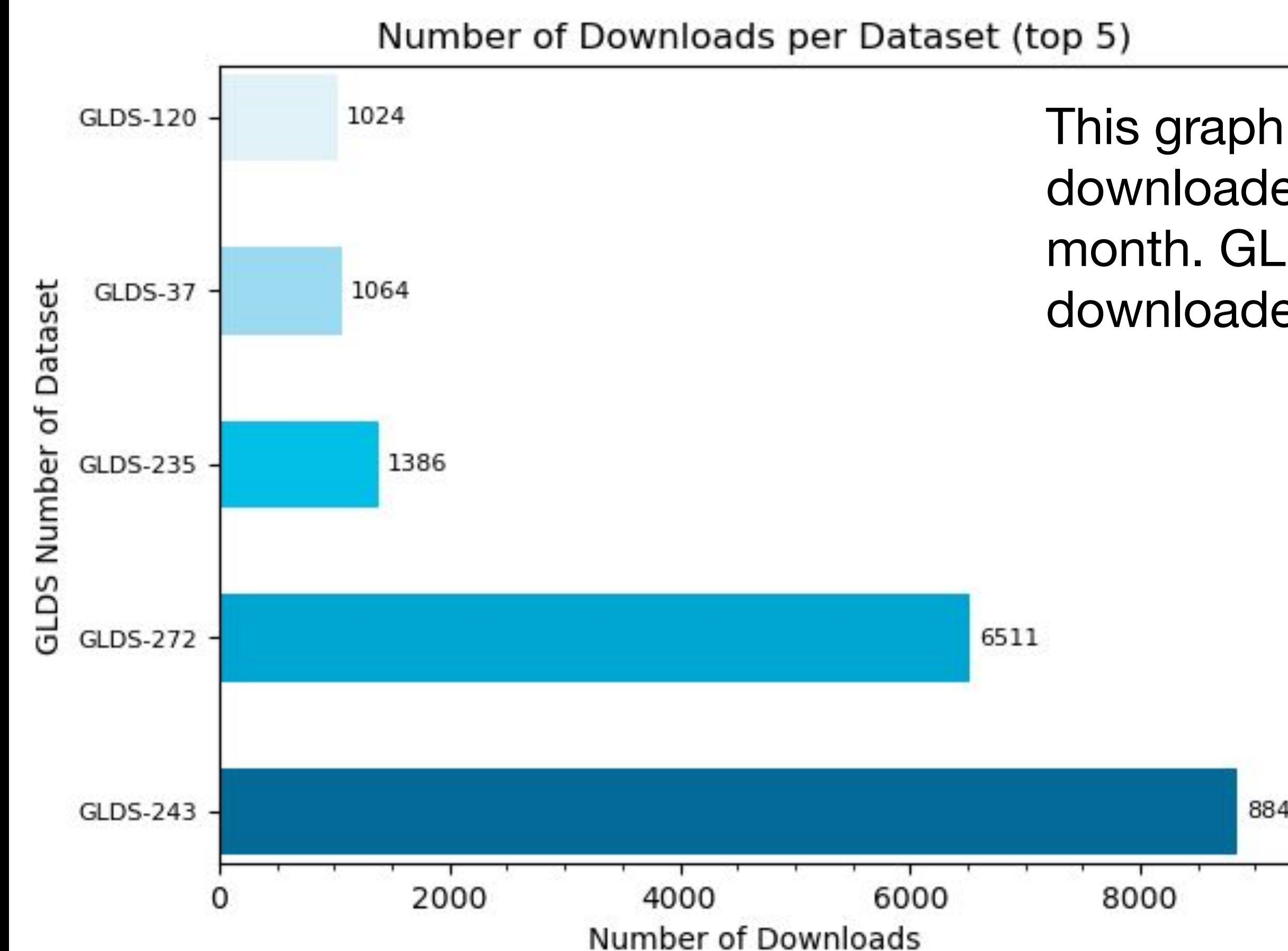
¹Blue Marble Space Institute of Science

Part 1: Monthly graphs from CSV files

The part 1 graphs are for the Feb 2021 data.



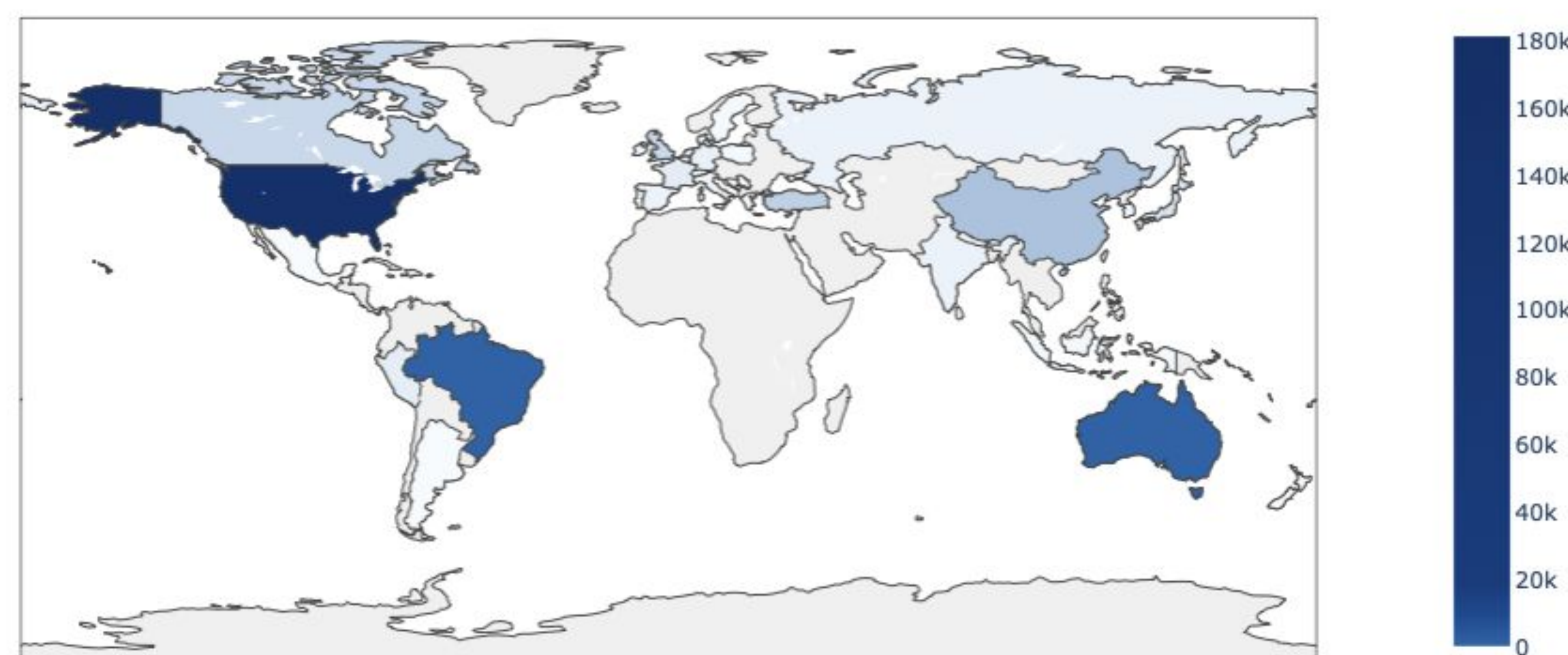
This graph shows how many files were downloaded for the month, and how many of them were raw data, processed data, metadata, or other.



This graph shows the top 5 most downloaded datasets for the month. GLDS-243 was the most downloaded dataset in Feb 2021.

The graph below shows what countries those who downloaded the datasets came from, and which country downloaded the most datasets.

Downloads per Country



About the Project

Project:

- analyzing metrics for the OSDR website, and for the OSDR datasets themselves

Objective:

- automate the analyses of the users of the OSDR database and determine what types of datasets people were downloading the most
- determine whether people downloaded raw or processed data files more, to determine whether it was worth the time to process the dataset if it helped more people become interested in working with the data and research in general

Overall Methods:

- Jupyter Notebook and Python
- numpy and pandas to make tables
- matplotlib and plotly to make the graphs

Part 1 Methods:

- received a monthly CSV file
- read that into numpy arrays, with an array for each column
- looked at what URL people were visiting, and whether that URL was a link to download a dataset
- if so what dataset was being downloaded

Part 2 Methods:

- retrieved the statistics of the datasets themselves as a CSV from the OSDR API
- used that data to make the pie charts showing the groups of the datasets

Acknowledgements: Thank you to Ryan Scott, Danielle Lopez, Samrawit Gebre, and Dr. Lauren Sanders for helping me with this project!

Part 2: Pie charts of datasets from API

