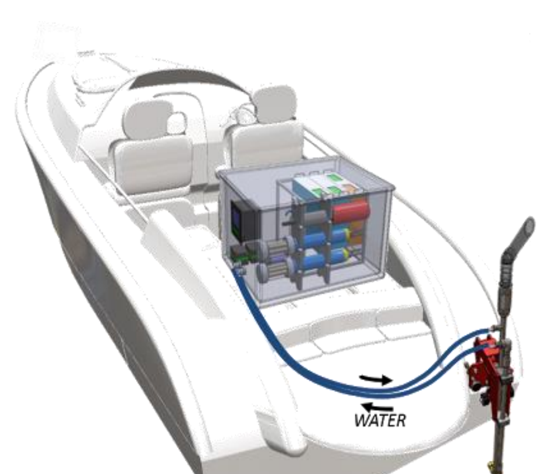
BigRiver: Developing a toolset to manipulate and visualize high-resolution spatiotemporal GHG data

Changming Feng, William C Gagne-Maynard, Robin Gold, Catherine Kuhn

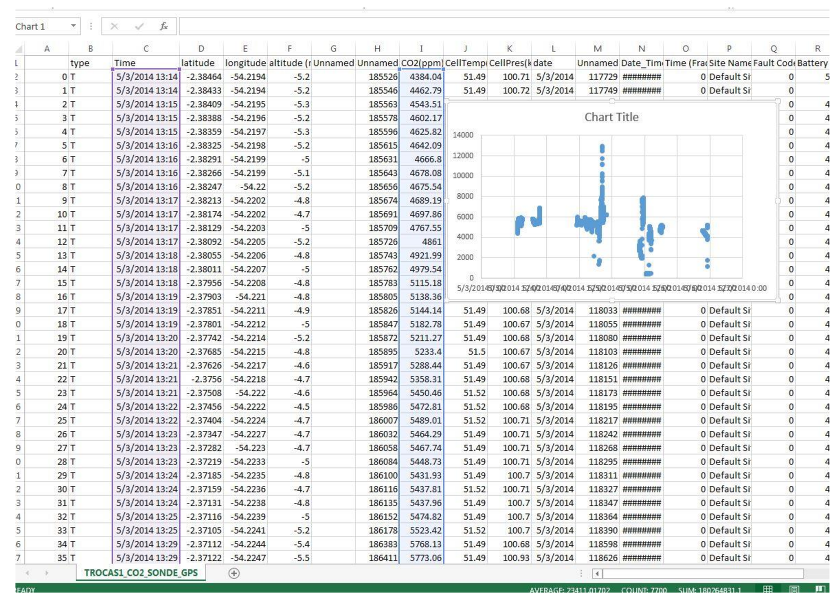
Background

Rivers, lakes and wetlands are generally known to be net sources of greenhouse gases(GHGs) to the atmosphere as they regularly emit CO_2 and CH_4. Understanding the dynamics that drive the outgassing and spatial variations of these GHGs is a key area of study.

Traditionally, scientists have gathered data from these systems by manually collecting headspace samples, resulting in a dataset of dozens to hundreds of samples at most. However, advances in high-frequency sampling techniques have allowed researchers to collect samples at high rates(1/s) over enormous spatial areas. Data quality and the spatial extent and resolution of this data present issues that can be problematic to solve without specialized programming skill.



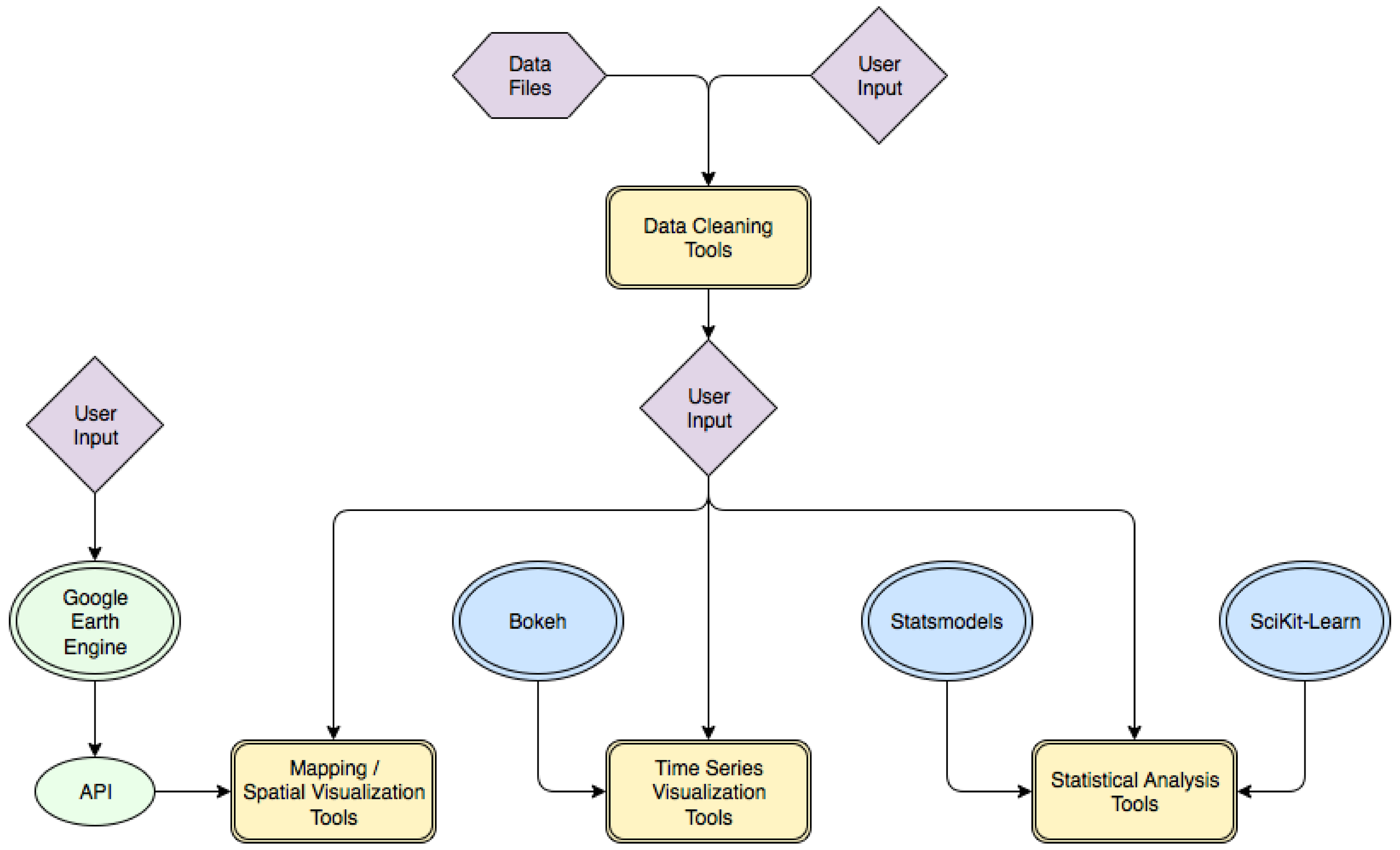
The Flame
USGS Sampling Platform



Objective

The motivation behind this collaborative effort is to build an online, open-source tool for synthesizing remotely sensed watershed-related geospatial data with in-situ field observations of water chemistry. We have developed tools that automates data ingestion, fusion and exploration to help understand the patterns and controls driving changes in aquatic GHGs to help inform regional and global carbon budgets.

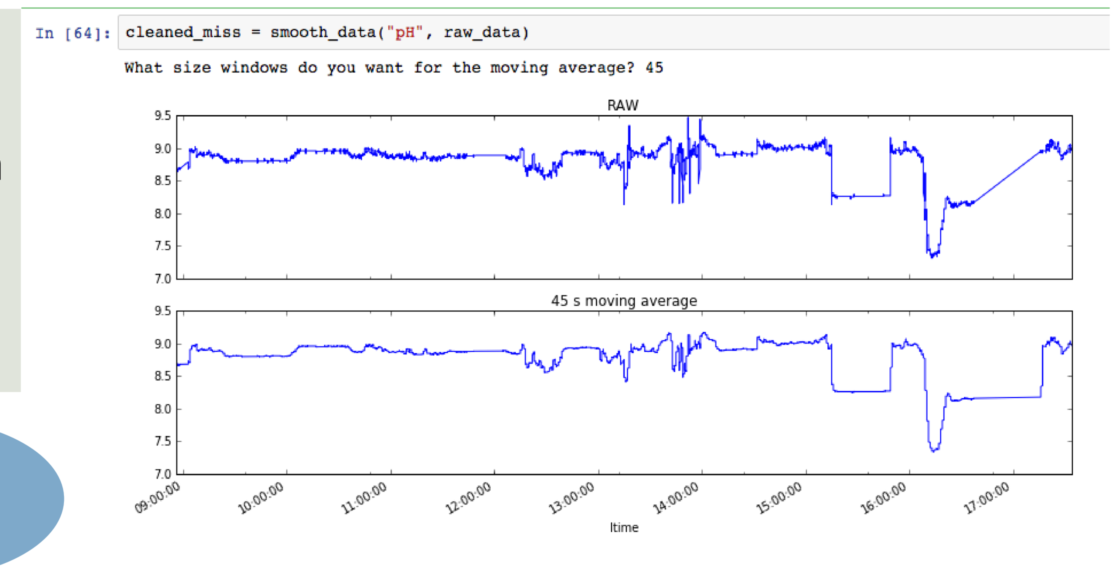
Project Architecture



Data Cleaning

Tools to:
- Import Data
- Apply smoothing function to data
- Resample data
- Remove all null values

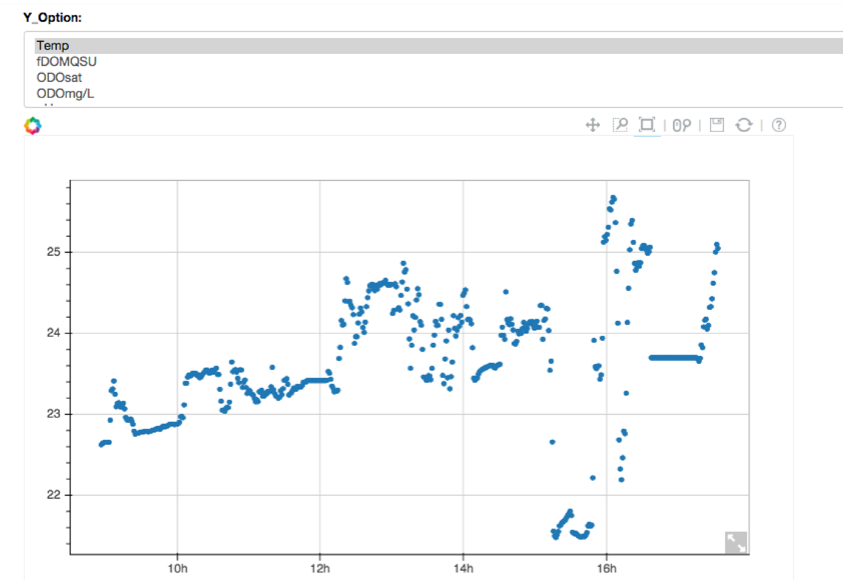
Pandas
sklearn



Time Series Plotting

Allows the user to interactively plot any column as a time series to explore the data

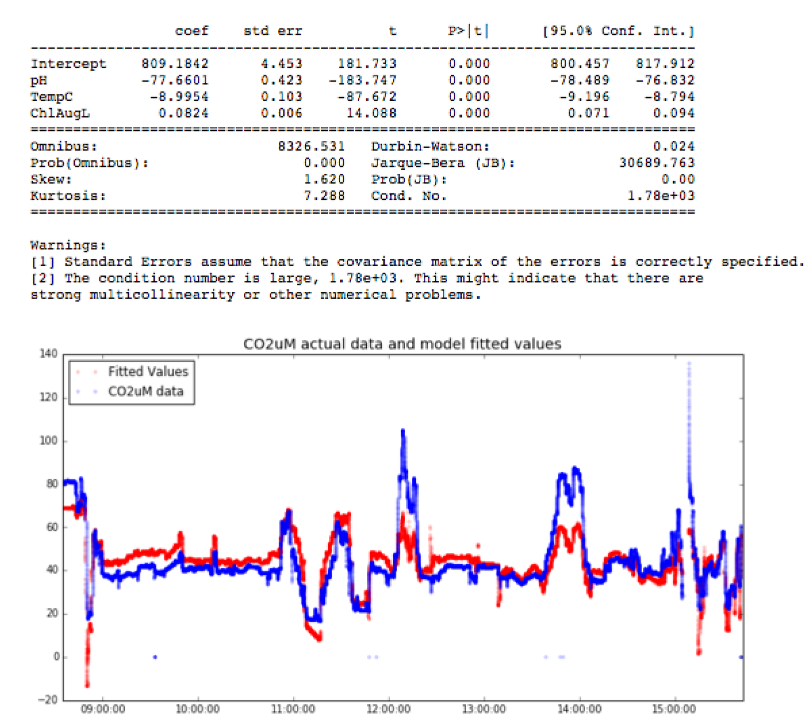
Bokeh

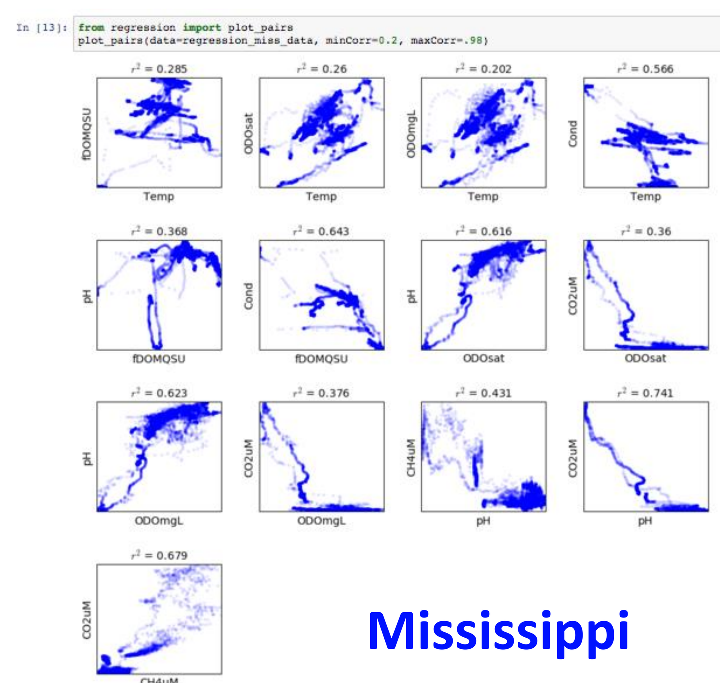


Statistics

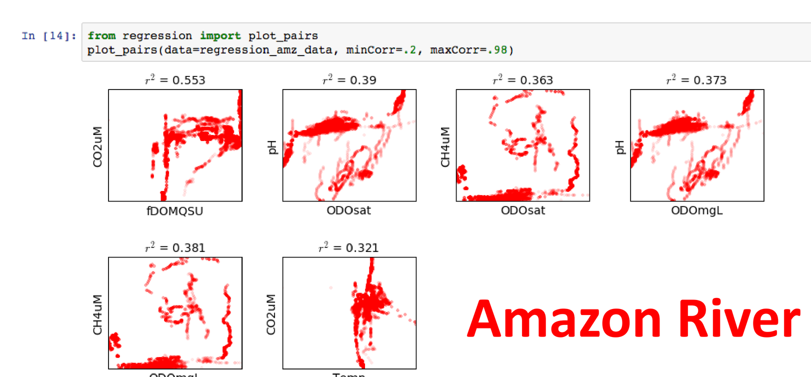
Tools to:
- Compare random OLS model runs
- Create a user specified OLS regression model
- Plot correlated covariates against each othe

statsmodels
sklearn





Mississippi

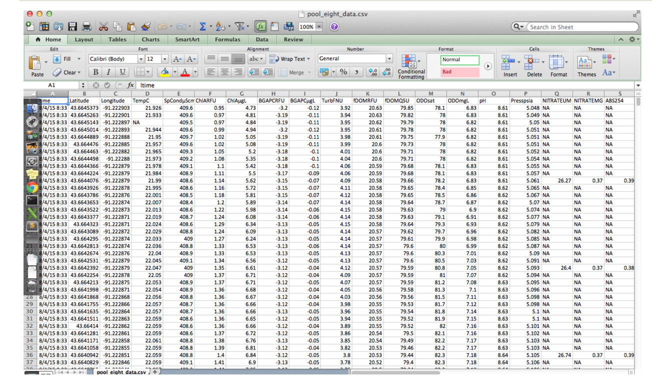


Amazon River

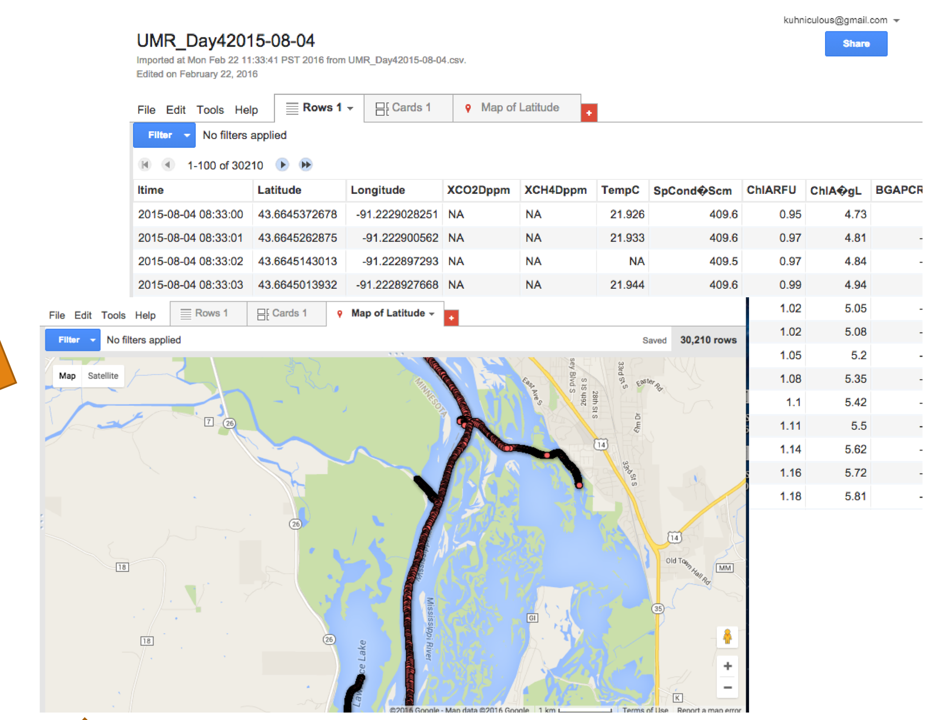
Mapping the Data

To visualize this data, we chose to use Google Earth Engine, a planetary-scale platform for geospatial data and analysis. Google Earth Engine gives us an open-source, cloud-based way to visualize our data, as well as combine our vector data with remotely sensed raster datasets held in Earth Engine. Because it is based in the cloud, Earth Engine gives us the potential to perform analyses and visualize our datasets faster and at a greater spatial resolution than would have been possible before.

Earth Engine also includes a Python API that can be called from an iPython notebook, allowing us to export our visualizations into the same notebook that we would use for all prior data transformations and visualizations.

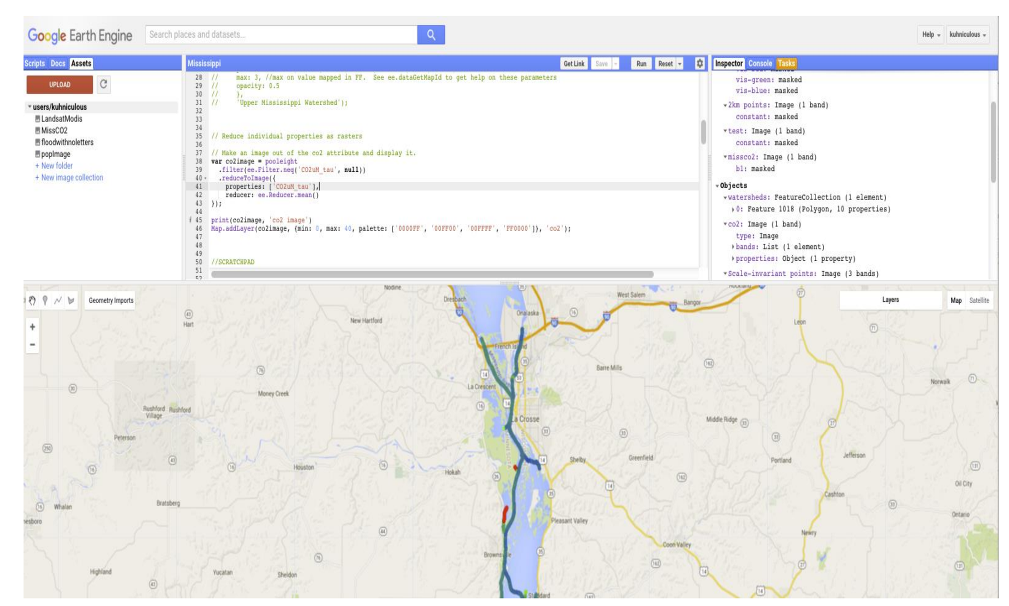


.CSV

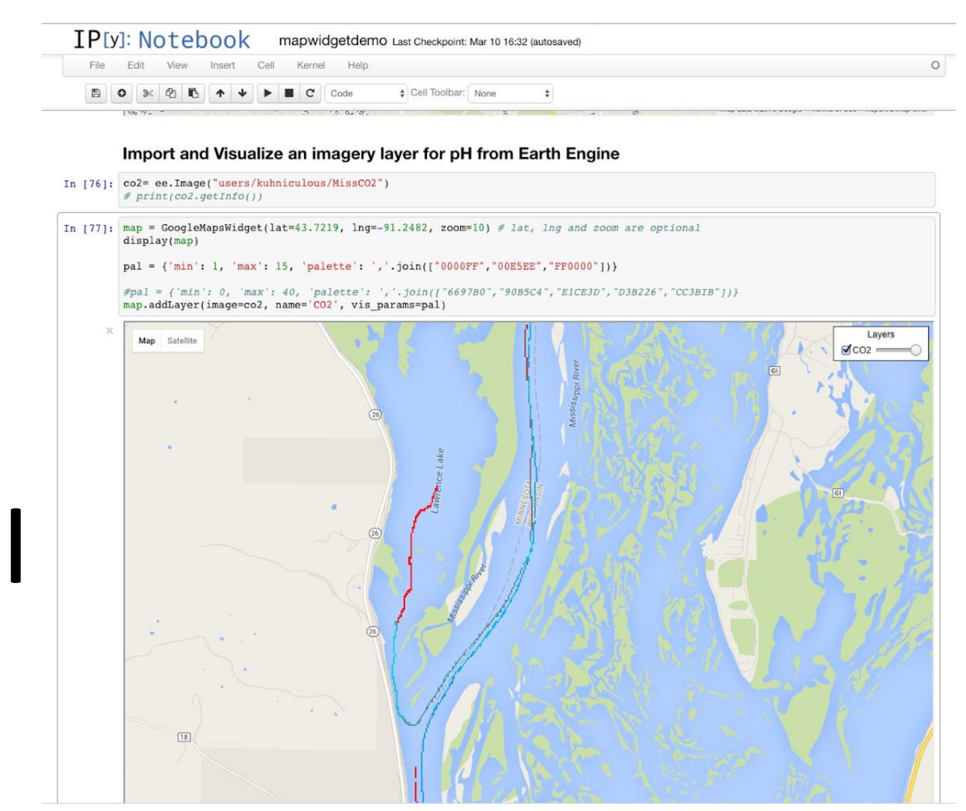


Fusion Tables

Javascript console



Google Earth Engine



iPython Notebook API