

PSTAT100



Mini project 2: primary productivity in coastal waters

In this project you're again given a dataset and some questions. The data for this project come from the [EPA's National Aquatic Resource Surveys](#), and in particular the National Coastal Condition Assessment (NCCA); broadly, you'll do an exploratory analysis of primary productivity in coastal waters.

By way of background, chlorophyll A is often used as a proxy for [primary productivity in marine ecosystems](#); primary producers are important because they are at the base of the food web. Nitrogen and phosphorus are key nutrients that stimulate primary production.

In the data folder you'll find water chemistry data, site information, and metadata files. It might be helpful to keep the metadata files open when tidying up the data for analysis. It might also be helpful to keep in mind that these datasets contain a considerable amount of information, not all of which is relevant to answering the questions of interest. Notice that the questions pertain somewhat narrowly to just a few variables. It's recommended that you determine which variables might be useful and drop the rest.

As in the first mini project, there are accurate answers to each question that are mutually consistent with the data, but there aren't uniquely correct answers. You will likely notice that you have even more latitude in this project than in the first, as the questions are slightly broader. Since we've been emphasizing visual and exploratory techniques in class, you are encouraged (but not required) to support your answers with graphics.

The broader goal of these mini projects is to cultivate your problem-solving ability in an unstructured setting. Your work will be evaluated based on the following: - approach used to answer questions; - clarity of presentation; - code style and documentation.

Please write up your results separately from your codes; codes should be included at the end of the notebook.

Part 1: data description

Merge the site information with the chemistry data and tidy it up. Determine which columns to keep based on what you use in answering the questions in part 2; then, print the first few rows here (but *do not include your codes used in tidying the data*) and write a brief description (1-2 paragraphs) of the dataset conveying what you take to be the key attributes. You do not need to describe preprocessing steps. Direct your description to a reader unfamiliar with the data; ensure that in your data preview the columns are named intelligibly.

Suggestion: export your cleaned data as a separate `.csv` file and read that directly in below, as in:

```
pd.read_csv('YOUR DATA FILE').head()
```

```
# show a few rows of clean data
```

Write your description here.

Part 2: exploratory analysis

Answer each question below and provide a graphic or other quantitative evidence supporting your answer. A description and interpretation of the graphic/evidence should be offered.

- i. What is the apparent relationship between nutrient availability and productivity? *Comment:* it's fine to examine each nutrient – nitrogen and phosphorus – separately, but do consider whether they might be related to each other.
- ii. Are there any notable differences in available nutrients among U.S. coastal regions?
- iii. Based on the 2010 data, does productivity seem to vary geographically in some way? If so, explain how; If not, explain what options you considered and why you ruled them out.
- iv. How does primary productivity in California coastal waters change seasonally in 2010, if at all? Does your result make intuitive sense?
- v. Pose and answer one additional question.

Write up your answers here.

Code appendix

```
import pandas as pd
import numpy as np
import altair as alt

ncca_raw = pd.read_csv('data/assessed_ncca2010_waterchem.csv')
ncca_sites = pd.read_csv('data/assessed_ncca2010_siteinfo.csv')
```