Building a reproducible climate data processing pipeline for social and environmental science applications

Undergraduate Summer Research Project 2022

Project Background

Global, gridded climate datasets are critical for environmental science, health, and justice research today, allowing us to draw associations between changes in climate and outcomes spanning from crop yield to human health. With climate change already beginning to show its effects on our lives and the ecosystems we rely on, it is critical to ensure the swift and robust production of high quality research linking socioeconomic outcomes to changes in the climate. However, in order to obtain a complete and robust dataset to analyze relationships between climatic variables and outcomes of interest, each researcher must follow a series of processing steps. Generating a pipeline to ensure proper data processing is time consuming and presents itself as a barrier to researchers wishing to conduct such research, particularly for those working in social sciences who are not familiar with the intricacies of climate datasets. Researchers at the Environmental Markets Lab (emLab) have developed a flexible pipeline in the programming language R to process gridded climate data as required before statistical analysis can be performed. In order to facilitate research relying on such data and expedite the course of environmental data science, we are interested in publishing this pipeline for use by all, in the form of an R package and a peer-reviewed research publication. emLab is searching for one (1) motivated undergraduate student with experience and interest in programming and/or environmental data science to join this project for the summer of 2022.

Responsibilities

The chosen candidate will be responsible for:

- Converting the existing data pipeline into a publishable R package
- Drafting, with support from other team members, a manuscript for submission to a peer-reviewed journal, with potential for co-authorship for the motivated student.

For reference, an ideal project timeline has been included below.

Qualifications

- Registered UCSB undergraduate student in good academic standing
- Proficiency in programming, preferably in R. However, experience in other languages may be sufficient if the applicant is confident in their programming abilities and ability to apply their knowledge to R.
- Proficiency using GitHub
- Strong reading and writing skills, preferably from previous research experience

emLab is committed to sustaining an equitable and inclusive workplace and to the goal of reflecting the rich diversity of our campus community within our staff. We encourage applications from students of all backgrounds and majors. We strive to create an adaptive, supportive environment, especially during circumstances caused by the COVID-19 pandemic.

How to Apply

Apply for this position via Handshake with either the: a) job number – 6406175; b) job title – Reproducible Climate Data Processing Pipeline Student Researcher; or c) employer – Marine Science Institute. In your application package, please include the following:

- CV or resume
- Unofficial transcript
- A brief (1 paragraph) description of why you are interested in the project and why you would be a good fit.

The selected student will work under the guidance of PhD student Anna Boser and PI Dr. Tamma Carleton for 10-12 weeks, starting in June. The position can be full or part-time and compensation will be $16/hr. While we prefer in-person work at emLab in the Bren School, remote work is also possible.

Please contact annaboser@ucsb.edu with any questions related to this position.

Application Deadline: May 15th

Timeline

Week	Goals
1	Familiarize yourself with the existing pipeline. Try running a few examples through the pipeline.
2	Familiarize yourself with the format necessary to publish an R package, and identify necessary changes.
3	Implement other necessary changes to the pipeline.
4	Implement other necessary changes to the pipeline. Write the outline for a short tutorial.
5	Implement final changes and publish the package. Generate code and figures for the short tutorial.
6	Write out and publish the tutorial on how to use the package.
7	Write the introduction to the manuscript (see this paper for an example).
8	Using the tutorial as a starting point, write a more comprehensive guide to the package.

9	Write the methods, conclusion, and abstract to the manuscript.
10	Revise introduction and guide
11	Revise methods, conclusion, and abstract.
12	Manuscript edits and prepare for submission.