CORNELL CENTER for
SOCIAL SCIENCES

# Understanding the Python Package Ecosystem

Python Workshop Series Part 2

Spring 2022

# Land Acknowledgement

Cornell University is located on the traditional homelands of the Gayogo̲hó꞉nǫ' (the Cayuga Nation). The Gayogo̲hó꞉nǫ' are members of the Haudenosaunee Confederacy, an alliance of six sovereign Nations with a historic and contemporary presence on this land. The Confederacy precedes the establishment of Cornell University, New York state, and the United States of America. We acknowledge the painful history of Gayogo̲hó꞉nǫ' dispossession, and honor the ongoing connection of Gayogo̲hó꞉nǫ' people, past and present, to these lands and waters.

Learn more about land-grab universities, Cornell's history with indigenous dispossession, and the movement to return stolen lands to indigenous people. More resources can be found at Cornell's American Indian & Indigenous Studies Program (AIISP).

# Community Norms

The Cornell Center for Social Sciences provides a welcoming environment for everyone embracing all backgrounds or identities. All instructors and attendees agree to abide by our community norms. We encourage the following behaviors in our workshops:

- Respect differing viewpoints and ideas
- Share your own perspectives and ask any questions
- Accept constructive criticism
- Use welcoming and inclusive language
- Show courtesy and respect for all instructors and attendees

If you believe that an instructor or attendee has violated the code of conduct, please report the violation to **ccss-researchsupport@cornell.edu**. We take all reported incidents seriously.

# Today's Learning Objectives

- Understand what packages are and how to use them
- Understand the process of installing new packages through PyPI and pip
- Familiarize yourself with the list of commonly used data science packages, and know which package to use for each specific task you might need in your own research
- Get hands-on experience with numpy, the package that forms the basis of most of the Python data science ecosystem

# The Right Tool for the Job

| *If you need to do…* | *Then you should use…* |
|---|---|
| Processing numerical data | numpy |
| Advanced statistics | scipy, statsmodels |
| Data visualization | matplotlib |
| Tabular data management (like spreadsheets) | pandas |
| Network analysis (e.g., social graphs) | networkx |
| Machine learning | scikit-learn |
| Deep learning | pytorch, tensorflow |
| Natural language processing | nltk |