

### ENVIRONMENTAL DATA ANALYTICS: M11 – PYTHON FOR R USERS

Spring 2023

Nicholas School of the Environment - Duke University

# Agenda

- Questions on Projects; Presentations
- Python!
- Course wrap-up...

#### **Projects – Presentation**

One person in each team should have slides set up on a laptop.

When it's your team's turn to present, you'll share your screen to the Zoom session which will be projected on the classroom screen.

Questions & comments, time permitting...

#### **Projects – Presentation**

#### $\Box$ 5-7 minutes per group; ~5 slides

- Your central question(s) or hypotheses you will address
- The data sources you will use to address the question(s) and any exploration of the data you've done (or intend on doing)
- The analysis (including wrangling) of your data you have done/plan on doing to test your hypothesis
- How you plan on communicating your findings, in terms of visualizations, plots, tables etc.
- Any challenges you are anticipating or have discovered in successfully confirming or refuting your initial hypothesis

## R & Python

#### R vs Python: Key Differences

https://www.datacamp.com/blog/python-vs-r-for-data-science-whats-the-difference

Which language to learn? Both, of course

The future of R-studio: Posit "Where things just work..."

### Learning Python

□ A new repository:

https://github.com/ENV872/PythonForRUsers

Jupyter notebooks (on Duke's container)
 Click on Containers link at bottom of class website
 Click on Reserve Jupyter to create your Jupyter environment
 Open Jupyter....



- Navigating the Jupyter Lab interface...
  - File browser
  - Kernel manager
  - Git
- □ Writing code...
  - Create a new Python notebook
  - Code/markdown "cells"

# Cloning the material

#### □ Clone the repo:

#### https://github.com/ENV872/PythonForRUsers

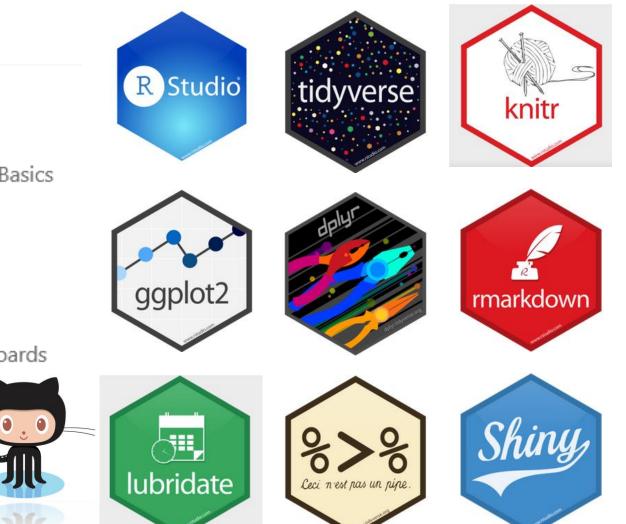
Notebook	Rmd counterpart	Topics
01-Getting-Started	-	Quick tour of JupyterLab
02-ReproducibilityCoding-Basics.ipynb	02_Reproducibility_CodingBasics.Rmd	Basics of Python
A-Basic-Python.ipynb	-	Basics of Python
03-Data-Exploration.ipynb	03_DataExploration.Rmd	Data Exploration
03-Data-Exploration_II.ipynb	03_DataExploration_Part2.Rmd	Exploration & Visualization
04-Data-Wrangling.ipynb	04_Part1_DataWrangling.Rmd	Data Wrangling
04-Data-Wrangling_II.ipynb	04_Part1_DataWrangling.Rmd	More Data Wrangling
B-Web-Services-APIs-Python.ipynb	-	Scraping data



### Recap

#### MODULES

- 0-Course Setup & Intro
- 1-Intro to Data Analytics
- 2-Reproducibility & Coding Basics
- 3-Data Exploration
- 4-Data Wrangling
- 5-Data Visualization
- 6-GLMs
- 7-Crafting Reports & Dashboards
- 8-Time Series Analysis
- 9-Spatial Analysis
- 10-Data Scraping
- 11-Python for R Users



# **DEI and Data Science**

- Reproducibility & Transparency
- Data can overcome bias....
- BUT data can be biased!
- □ Correlation ≠ causation
  <u>https://www.tylervigen.com/spurious-correlations</u>
- Visualizations can reach a wider audience...
- BUT visualizations can be deceptive!



