Topic 1. Introduction to *Environmental Data Analysis*

ENS-215

Winter 2025

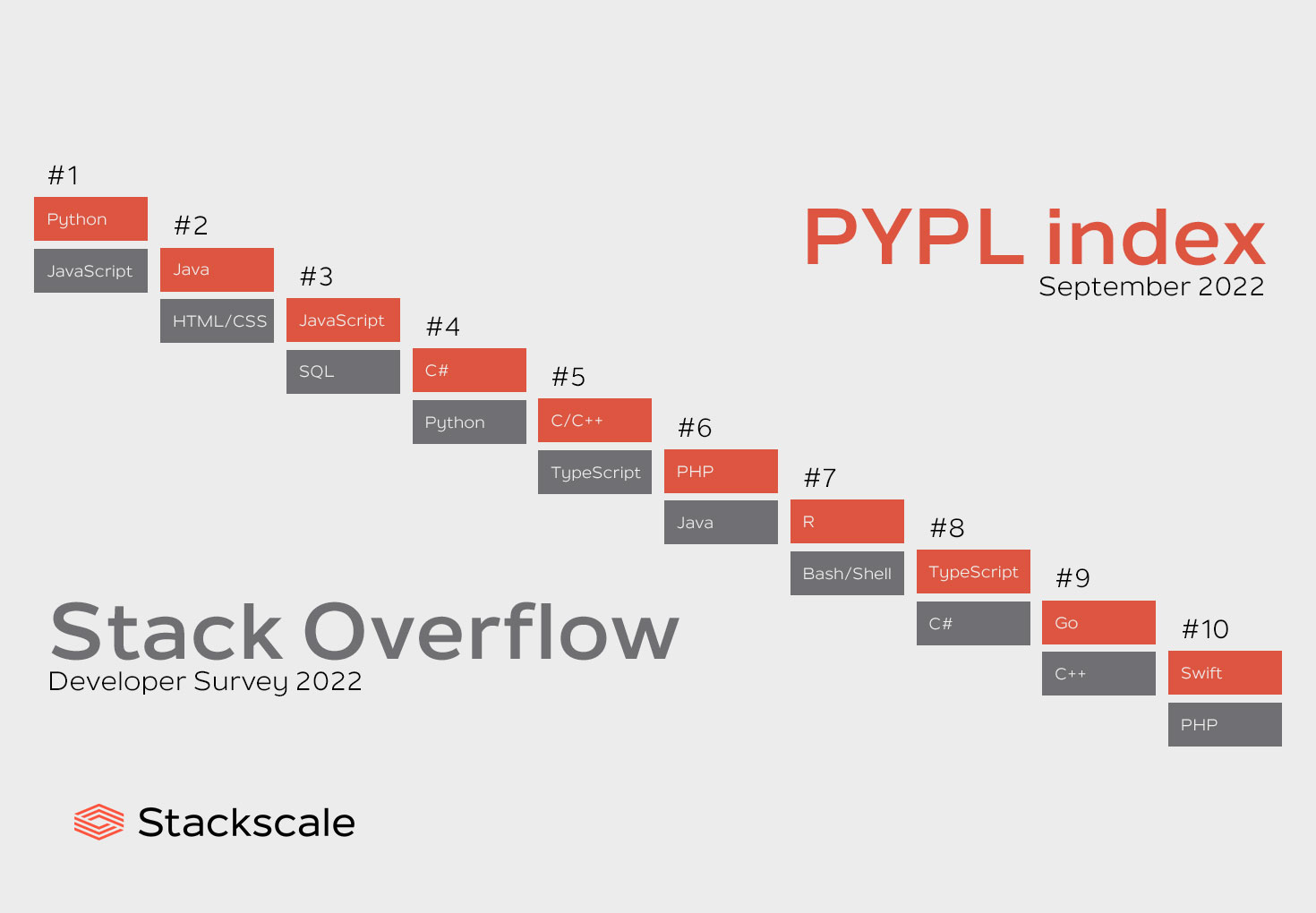
Table of Contents

## Why learn data science?

*It is pretty clear that almost every field today needs data-savvy researchers. It has become something as basic as driving a car in our intellectual work*  
– Saul Perlmutter (2011 Nobel Laureate in Physics)

Data science is a rapidly growing field and is becoming an integral part of nearly all fields of study, including the Earth and environmental sciences. In the past few years, programs specifically focused on data analytics for Earth and environmental sciences have been created or expanded at many prominent institutions (Stanford, Berkeley,U. Chicago, and UC Boulder to name a few).

The demand in the academic, public, and private sectors for environmental scientists with skills is data analytics has been growing and job and research opportunities are strong. The articles below provide a nice overview on the rising importance of data science in the environmental fields.

* <https://eos.org/opinions/training-the-next-generation-of-physical-data-scientists>
* <https://earth.stanford.edu/news/21st-century-earth-science-computer-intensive-and-data-driven>
* <https://www.earthdatascience.org/blog/earth-data-scientist-demand/>
* **Why learn programming?**
  + Allows you to analyze, model, and interpret data that is often impossible (or exceedingly difficult) to do otherwise
  + Streamlines your workflow, allowing you to conduct analysis in a documented, reproducible, and efficient manner
  + Opens the door to answering a whole new set of questions in your research
  + Highly marketable skill set in science/engineering careers as well as non-science careers
* **Why learn R?**
  + One of the most popular programming languages for data analysis and statistics
  + User friendly with lots of available support material and large user community
  + Tens of thousands of packages available to facilitate a wide range of analytical and visualization tasks
  + Allows for efficient and automated access to thousands of databases/datasets across the environmental and geosciences (as well as across many other fields)
  + Widely used by many leading environmental and geoscience organization, with widespread usage at the USGS
  + Highly desired programming language when applying for jobs and grad school
  + You can create cool figures, maps, and visualizations
* 
* **R is rapidly growing in popularity**  
  source (<https://www.stackscale.com/blog/most-popular-programming-languages>)

#### **Interactive figures are easy to make**

## Programming in R

### RStudio IDE

We will write and run our code within the RStudio Integrated Development Environment (IDE). RStudio allows us to write, excute (run), and debug code, along with view output and plots, within a single integrated environment.

### Getting help in R

There are lots of resources for [getting help in R](https://stahlm.github.io/ENS_215/Winter_2025/Syllabus.html#useful_linksresources). In addition to me and your classmates, there is a massive amount of help available freely online (a quick Google search typically yields an answer to just about any R question). There are also many websites devoted to teaching R.

Visit Union College’s very own [Center for Data Analytics](https://muse.union.edu/cda/)! It is located in **Wold 010** and hosts a data science help desk, where expert students and faculty who staff the help desk can help answer your questions. The Center for Data Analytics also host regular training workshops and guest speakers.

You can also get help directly in R Studio by typing ?term\_of\_interest\_here in the R **Console** or by searching in the **Help** bar on the right-hand side of your RStudio window.

Your textbooks [*R for Data Science*](https://r4ds.hadley.nz/) and [*ModernDive*](https://moderndive.com/v2/index.html) are also amazing resources and freely available online in a nice searchable format.

[R Cheatsheets](https://posit.co/resources/cheatsheets/) are also amazing resources that succinctly summarize many different aspects of R. I will hand these out throughout the term, though you can find them freely available here.

## Syllabus and course logistics

* **Course logistics**
  + You will find all course related materials (e.g. schedule/syllabus, assignments, notes, course policies,…) at our [class webpage](https://stahlm.github.io/ENS_215/Winter_2025/Home.html)
  + Assignments will be submitted through Nexus. You MUST use the specified naming convention for the files you submit.
  + Attendance and on-time arrival is critical and your grade will be affected by unexcused absences and late arrivals! You will quickly fall behind if you miss class or arrive late.
  + We will use computers during class and lab, however you should ONLY use your computer for class related activities. This means that you should close unrelated material prior to the start of class and that your should not check email, news, Instagram, Twitter,… during class.
  + Carefully read the syllabus. It has important information about grading, course policies, and tips for success in the class.
  + In each class we will work through code examples in R. I will post my R Notebooks (which act as lecture notes/slides) to the course website. DO NOT simply copy and paste the code from these notebooks. You will not learn the material by doing this. You should type out the code, think carefully about what it means, run the code, and examine the output before moving on to the next step.
  + Get a binder to hold class notes/handouts. This binder will act as a course specific textbook.
  + You need to be an active learner in this class. Learning programming and data analysis cannot be done passively.
  + You should recognize that there will be challenges and bumps as you are learning how to code. You are not alone in this (everyone even experienced programmers hit roadblocks when they work). There are TONS of resources out there to help you. A quick Google search of whatever problem you’ve hit or something new that you’d like to see an example for, will almost always yield tons of information directly related to your question.