# POD 4 Bootcamp Curriculum 2024/2025

TAs: Johanna Bayer, Amanda Gahlot

***PLEASE NOTE: that these topics and schedules are subject to change***

| WEEK 1: Sept 30th | |
| --- | --- |
| October 4th 12-1pm PT: Link to Zoom recording   * Introductions for TAs and learners * Set curriculum for bootcamp * Set Pod meeting schedule | |
|  | |
| Option 1: MONDAYS 2PM EST | Option 2: FRIDAYS 11AM EST |
| WEEK 3: Oct 14th | |
| **Setting Up Reproducible Markdown Files Across Platforms**   * **Description**: This class introduces markdown files for reproducibility across RStudio, Jupyter, Google Colab, and MATLAB, with the goal of integrating code, text, and outputs. | **Working with Python in a Research Setting:**   * **Description**: In this session, we will explore various tools and setups for using Python's machine learning libraries, like scikit-learn (sklearn), in a research context. We'll discuss different development environments such as VSCode, Jupyter Notebooks, and Python-specific IDEs. Additionally, we'll walk through the creation and activation of virtual environments, along with tips for debugging code effectively. We will also have a more conceptual look at machine learning. |
| WEEK 4: Oct 21th | |
| **Data Storage, Access, and Import with Data Wrangling**   * **Description**: Covers data storage and importing data, and includes data wrangling with dplyr and tidyverse. Also covers basic outlier detection and visualization. | **Classification (Unsupervised Learning)**   * **Description:** This week, we'll dive into classification techniques, focusing on methods such as k-Nearest Neighbors (KNN) and Logistic Regression. We'll apply these techniques to the Iris dataset to explore how classification models function in practice. |
| WEEK 5: Oct 28th | |
| **Longitudinal Data Handling and Analysis**   * **Description**: This class focuses on reshaping and analyzing longitudinal data, covering wide and long data formats and analysis with nlme | **Regression (Supervised Machine Learning)**   * **Description:** We will introduce linear regression by fitting a model to real-world data. We'll then evaluate the model’s performance, discussing key metrics and considerations in supervised learning. |
| WEEK 6: Nov 4th | |
| **Creating Dynamic Tables and Visualization with Advanced R Notebook Features**   * **Description**: Enhancing R Notebooks with creating loops in R, interactive tables and correlation plots, including markdown customization for reproducible reports | **Cross-Validation and Grid Search**   * **Description**: We'll cover the fundamentals of cross-validation, a crucial technique for tuning model parameters. Additionally, we will implement grid search for parameter optimization and feature selection, ensuring our models are fine-tuned for optimal performance. |
| WEEK 7: Nov 11th | |
| **Commit to Git - Starting with GitHub Across Platforms**   * **Description**: Introduces GitHub basics, including account creation and code commits from RStudio, MATLAB, Jupyter, and Google Colab | **Metrics and Categorical Features**   * **Description:** In this session, we'll discuss evaluation metrics such as the Confusion Matrix and Area Under the Curve (AUC). We will also introduce pipelines to handle categorical features efficiently, ensuring smooth integration in machine learning workflows. |
| WEEK 8: Nov 18th | |
| * Learners to present knowledge gained through bootcamp and discuss plan for self-guided projects for next 4 months (5 minutes each) | |