**Course Description**

A short course introducing Python programming for users of R. Throughout the course we will highlight syntactic similarities and differences between the two languages. As much as possible, we will make the course beginner friendly.

By the end of the course, participants will

* Understand the basics of Python syntax and semantics.
* Use Python to load, manipulate, summarize, and visualize data.
* Use Python to perform commonly used statistical tests.
* Use Python to build linear models and display results.

**Prerequisite:**

This course assumes basic knowledge of programming concepts such as variable assignment, using functions, vector/array construction and indexing. While the preferred prerequisite is a working knowledge of R or other statistical computing languages (e.g., SAS, Matlab), those with no prior background in programming are also welcome. We will make a concerted effort to make it beginner friendly. For those who do not have previous programming background in any language, we recommend reading the first 5 chapters of “Think Python” by Allen B. Downey. This book is freely accessible at [Think Python Link](https://greenteapress.com/wp/think-python/), or [Think Python PDF Version](about:blank).

|  |  |
| --- | --- |
| **Tentative Schedule for the day - Python for R Users** | |
| 8:30 AM - 9:30 AM | Registration, breakfast and assistance with software loading |
| 9:30 AM - 11:00 AM | Section 1: Introduction to Python |
| 11:00 AM - 11:15 AM | Break |
| 11:15 AM - 12:45 PM | Section 2: Data loading, cleaning and exploration |
| 12:45 PM - 1:15 PM | Lunch |
| 1:15 PM - 2:45 PM | Section 3: Data summarization and visualization |
| 2:45 PM - 3:00 PM | Break |
| 3:00 PM - 4:30 PM | Section 4: Statistical tests & models and Closeout |

**Section 0:**  Downloading and installing Python and installing libraries that we will use throughout the course. **Link to a pre-recorded video:** <https://youtu.be/C7ndPsUyTEY>**.**

**If you need additional support, please arrive 30 minutes to 1 hour before for assistance.**

**Python libraries of focus:** NumPy, Pandas, SciPy, StatModels, Scikit-learn, Matplotlib, and Seaborn.

**Section 1: Introduction to Python**

* + Mathematical operators in Python
  + Importing/loading packages
  + Installing packages
  + Python variable assignments
  + Valid and invalid variable names
  + Python data types
    - Strings
    - Numeric data types
      * int
      * float
      * complex
    - Booleans
    - Lists and tuples
    - Ranges
    - None type
  + Comparing values
  + NumPy arrays:
    - Creating, slicing, sub-setting, manipulating, lists and arrays.
    - Multidimensional arrays
  + Python functions
    - Constructing Functions and methods. Differences between functions and methods
  + Control flow:
    - Conditional evaluation: if else; if elif else
    - Loops (for, and while loops)
    - List comprehension
  + Python dictionaries

**Section 2: Data loading, cleaning, joining, and exploration**

* + Import and clean data.
  + Data manipulation with the Pandas Package (iloc, loc, selecting columns, adding columns, removing columns, filtering, and summarizing by a categorical column)
  + Joining/merging data using Pandas

**Section 3: Data summarization, and visualization**

* + Summary statistics
  + Scatterplots, line graphs, and bar graphs using matplotlib library
  + Statistical data visualization using seaborn library

**Section 4**: **Statistical tests & models**

* + Performing t-tests, proportion tests, and chi-square tests using Python
  + Linear models using StatModels library.

**Speaker short bio:**

Bereket Kindo, PhD has over 10 years of experience in the data science and statistics profession. In the early part of his data science career, he mainly used R and SAS for statistical analysis and development of machine learning models, with occasional use of Python. In the more recent 5 years, Bereket has been using Python as the primary programming language for his Lead Data Scientist role at Humana.

Bereket has a doctoral degree in Statistics with a focus in Bayesian machine learning, a masters degree in Actuarial science and a Bachelor's degree in Mathematics.

In addition to his technical skills, Bereket is a strong communicator, often explaining and illustrating complex programming and statistical concepts clearly.