**Draft Outline for Python for BCPS Course Content**

**Part 1: Getting Started**

**Goals:**

1. Participants should be able to create a new jupyter notebook, open an old one, or run a simple python script from within an IDE (stretch goal would be from command line)
2. Participants can list the main datatypes and basic objects used in python
3. Participants can produce code for simple python tasks: math, looping, functions, logical statements

**Content:**

* What is Python and Why it Matters
* How to Python – Anaconda, Conda, Jupyter, IDEs **(Lindsay)**
  + Installing Python/Jupyter etc.
    - Make this ideally pre-work so they will have at least attempted to install the basics onto their computer before the workshop
  + Overview of working within given workspace
  + IDEs:
    - VSCode, PyCharm
* Python Fundamentals
  + Writing code
    - Python syntax and practice
    - Troubleshooting and help
    - Documenting/organizing work
  + Core data structure concepts **(Lindsay)**
    - Variables, Datatypes, Lists, Dicts, Tuples etc.
    - Methods/functions and how they work with data

**Part 2: Getting, Understanding, and Cleaning Data**

**Goals:**

1. Participants can read data into pandas dataframes from standard sources (csv, excel)
2. Participants can remove/replace NA data
3. Participants can summarize a dataset at a variety of levels of aggregation
4. Participants can describe the difference in different styles of augmenting datasets (left/inner joins, concatentations, etc)

**Content:**

* Getting data **(Lindsay)**
  + Using numpy and pandas
  + Reading data from existing sources
  + Creating data
  + Viewing data
* Data Cleaning
  + - Dealing with missing (e.g. NA) data
    - Changing data types
    - Changing values

**Part 3: Analyzing Data**

**Goals:**

1. Participants can explore data in their datasets to enable understanding contents.
2. Participants can produce report summary and statistics (and statistical tests) for different types of data
3. Participants can produce charts to display statistics, including bar, line, scatter

**Content:**

* Exploring data structures **(Stu)**
  + Same stuff they cover in the dplyr section of the R course:
    - Data overview (df.info / df.describe)
    - Selecting columns
    - Filtering
    - Grouping/summarizing -> .agg
    - Counts -> value\_counts
    - Mutating (assign)
    - ‘Piping’ into a single go to avoid temporary dataframes
* Summary stats / Statistical tests
* Graphical depictions of data **(Stu)**
  + e.g. Matplotlib and/or Seaborn

**Part 4: Other Topics**

**Goals:**

1. Participants can find and download data from the bcdata catalog
2. Participants can produce a basic html from an ipynb workbook
3. Participants can describe what the methods .fit(), .predict() will do in a basic sklearn pipeline
4. Participants understand the difference between regression and classification techniques, and can explain which are the typical scoring methods used for each

**Content:**

* Publishing/Reporting
* Fetching data from bcdata **(Lindsay)**
* Machine Learning use cases – Scikit-learn **(Stu)**