# **Data Science Education Program:** Curriculum Team

# Intro to Data Science Module

***Initial Outline***

# Introduction

### What is data science?

* **The beauty of data science**

### What this intro module is/is not

* **What you will take away from this module** (by the end of this… you will be able to...)
* **Introduction to iPython notebooks** -> how the formatting works
  + Try to make this a little relatable. Add images or famous/funny quotes.
* **Using NumPy**
  + *Make an animation?*
  + Think of it as your friend that will interact with objects and entities in the abstract realm of data science!

# Expressions

### Data Types

* + Ints, strings, floats (mention differences and their properties)

### Arithmetic

* + +, -, \*, /, //, \*\*, order of operations

### Variables

* + Variable declaration and assignment
  + Variable arithmetic

### Functions

* + How to call functions
  + Examples of functions (abs, min, max, etc)
  + Math module operators (sqrt, pow, etc)
  + Converting between data types (str(), int(), float())
  + Defining your own functions
  + Errors from functions

# Arrays/Lists

### Arrays

* + Creating arrays
  + Array properties and methods (getitem, size, abs, min, etc; [good list of functions here](https://www.inferentialthinking.com/chapters/04/4/arrays.html))
  + 0 indexing

### Ranges

* + Creating ranges (start, end, step parameters)
  + A range creates a new array
* **Lists**
  + Creating lists
  + List properties and methods (getitem, len, min, append, etc)

# Strings

* .split
* regex

# Tables

### Why tables?

* + Tables are used to organize/categorize data so it can be processed

### Basics

* + Creating tables (from scratch, loading csv)
  + General structure of table (columns and rows, headers, etc)
  + Properties of tables (num\_rows, num\_columns, labels, etc)

### Adding to/taking from tables

* + Adding new rows and columns
  + Selecting/dropping columns
  + Taking a certain number of rows
  + Relabeling columns
  + .where method (also include predicates)

# Data Manipulation

Done in ds8 library and pandas

### Advanced methods

* + .group
  + .where
  + To get a specific value, index for column then row
  + When to use each, and why some are better than the others.

# Data Visualization

# (SAVE THIS FOR THE MOD TEAM)

### Bar graphs

* + Initial examples
  + .barh method
  + Practical uses

### Line plot

* + Initial examples

### Histogram

* + .hist method
  + Bins, and what to do with them

### When to use each type of visualization

* 3D visualization!
  + Very basic example, maybe just a line of code
  + Show them that cool stuff is possible too!

# Hypothesis Testing

### Hypothesis

* + Null/Alternative
  + Some examples with situations of each

### Test Statistic

* + How to derive one

### P-value

* + Significance of cutoffs
  + Errors

### Sampling

* + Generating random numbers
  + np.random

### Simulations

* + Bootstrapping
  + For loops

### Conclusions

* + Rejecting and accepting hypotheses

# Confidence Intervals

* Example of the practical use of confidence intervals
* Percentile function
* Application of quartiles with the percentile function
* Left and right (97.5 and 2.5 percentiles) to make a confidence interval of 95%

# Linear Regression

* Refer to current data 8 module - it’s done pretty well over there.