# **Teacher Notes for Introduction to Jupyter Notebooks**

## **Motivation and Essential Understandings**

Notebook applications are software tools designed to support exploratory data analysis and visualization. Unlike traditional software programs or analytical libraries found in Java, Python, SAS, Stata, or other toolsets, the notebook model incorporates computation, documentation, and visualization into a single platform. The resulting artifact, often called a “notebook”, can be shared among researchers, and readily run on and extended with little technical effort. For this reason, notebooks are becoming popular in the academic community.

Students of data science should be aware of notebook platforms as another tool for their data analysis toolbox. This lesson provides a basic introduction to notebook technology and can be used as the foundation for additional explorations into machine learning and classification algorithms.

## **Context and Dataset**

Students will use the classic Iris data set containing 150 observations of 3 different species of Iris flowers. This data set is considered a classic in the machine learning community. The data set is built-in to the Python scikit-learn package, so it is readily available with the Anaconda distribution.

## **Learning Objectives**

Students will be able to:

## Import a classic machine learning data science data set into a notebook

## Use Python data science tools to perform descriptive statistics on the data set

## Use Python tools to create scientific plots and graphs from the data in the notebook

## Answer botanical questions by using descriptive statistical results and scientific graphs rendered in the notebook

## **Data Science Concepts and Skills**

1. Jupyter notebook facility
2. Descriptive statistics
3. Exploratory data analysis and data visualization

## **Students**

This lesson is for early undergraduate students, although it has been used in an introductory graduate setting successfully as well. Students should be familiar with statistical concepts, basic data visualizations, and have worked with the Python programming language. Students do not need to have Python programming experience; all the Python code has been included in the notebook and does not need to be modified, although advanced students may wish to make modifications if desired.

## **Time to Teach this Lesson**

This lesson can be taught in **1 session**. It consists of a worksheet with botanical questions and a predefined Jupyter notebook with Python code embedded inside the notebook cells.

**First Week**: 5-10 minutes orientation; 20-30 minutes for the lab; 10-20 minutes sharing results

## **Lesson Materials**

You will find all the lesson materials in the GenAI GitHub repository. The Jupyter notebook is necessary to teach this lesson.

|  |  |  |
| --- | --- | --- |
| **Materials** | **File** | **Description** |
| Worksheet | Worksheet\_Intro\_to\_Jupyter\_Notebooks\_2020.docx | Worksheet containing overview of the lab and set of botanical questions to be answered |
| Dataset | Built-into the Python scikit-learn package. Instantiated using the “load\_iris()” command from within the Notebook. | Classic Iris data set |
| Jupyter notebook | [Intro\_to\_Jupyter\_Notebooks.ipynb](https://github.com/generationai/GMU-Intro-to-Jupyter-Notebooks-2019/blob/master/git_lesson/Intro_to_Jupyter_Notebooks.ipynb) | Jupyter notebook containing documentation and Python code for statistical analysis and visualizations |
| Template | Lesson\_Template\_Intro\_to\_Jupyter\_Notebooks\_2020.docx | Lesson planner with links to resources |

## **Teaching Strategies**

* Review Descriptive Statistics
* Review basic data visualization
* Pose **questions** to allow students to interpret statistics and visualizations to answer botanical questions.

## **Lesson Narrative**

|  |
| --- |
| **Module 0: Pre-lesson** |

No pre lesson prep is required. The following may be useful depending on the background of the students.

Review ideas of descriptive statistics (mean, median, mode, standard deviation)

Review ideas from data visualization (box plot, histogram, scatter plot)

|  |
| --- |
| **Module 1: Datasets and Exploratory Data Analysis** |

**Introduce the Iris dataset**

Explain the history and contents of the Iris data set.

Explain what the parts of an Iris flower are for non-botanists.

**Review descriptive statistics**

Review descriptive statistics tools available in Jupyter Notebooks provided via Python

**Review basic data visualization tools**

Review ggplot() and basic data visualization tools available in Jupyter Notebooks provided via Python

|  |
| --- |
| **Module 2: Explain Predictive Modeling** |

Present the assignment.

* Students will have the botanical questions on the worksheet and the Jupyter notebook available to run.

|  |
| --- |
| **ASSIGNMENT:** Ask students to review their worksheets, run the notebook, and provide answers to the questions. |

|  |
| --- |
| **Module 3: Close Out** |

Review results with the students.

* Did everybody come to the same conclusions? Explain.