

Working with Python Notebooks

Tutorial Overview

This tutorial introduces you to using Python code in a Jupyter Notebook, an open source web application that enables you to create and share documents that contain rich text, equations and multimedia, alongside executable code and visualization of analysis outputs. The tutorial begins by stepping through the basics of setting up and being productive with Python notebooks. You will be introduced to ArcGIS Notebooks, which are Python Notebooks that are well-integrated within the ArcGIS platform. Finally, you will be guided through a series of ArcGIS Notebooks that illustrate how to create compelling data notebooks for data science that integrate your own Python scripts using the ArcGIS API for Python and ArcPy in combination with thousands of open source Python libraries to enhance your analysis and visualization.

Learning Outcomes

By completing this tutorial, you will become comfortable with the following skills:

- Setting up Jupyter Notebook server and different variations of ArcGIS Notebooks,
- Basic interaction with a notebook, and techniques for using them effectively, and
- Composing effective notebooks for data science.

Time Required

The following time is required to complete this tutorial:

- Two hours

Materials Required

Technology:

- *Required:* ArcGIS Pro 2.5+
- *Recommended:* ArcGIS Online subscription account with permissions to use Notebooks and GeoEnrichment
- *Optional:* Notebook Server for ArcGIS Enterprise 10.7.1+

Notebook Server for ArcGIS installation guide for administrators:

<https://enterprise.arcgis.com/en/notebook/latest/install/windows/welcome-to-the-arcgis-notebook-server-install-guide.htm>

Data:

- Sample data for this tutorial are included as part of the download.

Data Sources

- Toronto Police Service: Major Crime Indicators: <http://data.torontopolice.on.ca/datasets/mci-2014-to-2019>
- NOAA National Centers for Environmental Information: NCDC International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 3: <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00834>

Production Date

The Education and Research Group at Esri Canada makes every effort to present accurate and reliable information. The Web sites and URLs used in this tutorial are from sources that were current at the time of production, but are subject to change without notice to Esri Canada.

- Production Date: May 2020

Background Information

The Python programming language was first introduced to the ArcGIS community in 2004 with the release of version 9.0, as one of the scripting languages that provides access to the ArcGIS geoprocessing framework. Since then, Python has become the language of choice for scripting and automation among ArcGIS users, and indeed many other applications¹. At the same time, the ecosystem of Python libraries has continued to grow (e.g., NumPy or SciPy for advanced mathematical and scientific processing, or more specialized modules for integrating deep learning and machine learning tools such as Tensorflow). With the introduction of the notebook interface, initially using IPython, and more recently with the Jupyter Notebook server, the Python language and specifically Python Notebooks have become highly valuable tools for data science.

While using Python notebooks with ArcGIS has been possible for many years, it has become a much more integrated experience with the release of the ArcGIS API for Python in 2016, and more recently Notebook Server for ArcGIS Enterprise (2019), ArcGIS Notebooks in ArcGIS Pro (2020), and ArcGIS Notebooks in ArcGIS Online (currently in beta).

By completing this tutorial, you will learn how to set up and interact with ArcGIS Notebooks, and how to load, analyze, and visualize GIS datasets in a Python Notebook to create a compelling document that communicates your GIS data processes or data science workflows for other developers.















References and Reading

- Introducing ArcGIS Notebooks
<https://www.esri.com/arcgis-blog/products/arcgis-enterprise/analytics/introducing-arcgis-notebooks/>

¹ <https://www.wired.com/story/python-language-more-popular-than-ever/>
<https://redmonk.com/sograd/2020/02/28/language-rankings-1-20/> (referenced May 8, 2020)

Tutorial Folder Contents

The following table summarises the folders and files provided with this tutorial – copy these files to your hard drive. The location should have a simple and easy to find path that does not include spaces or special characters. With these files copied to your hard drive, and with ArcGIS Pro optionally licensed for offline use, you will be able to complete the entire tutorial without a connection to the Internet (portions that require ArcGIS Online services can be skipped). All references to files later in this document will be relative to the top `arcgis_notebooks` folder:


Folder Name / File Name	Description
 <code>arcgis_notebooks/</code>	
 <code>documents/</code>	Related tutorial documents
 <code>hurricane_analysis/</code>	Sample Python notebooks - hurricane track analysis
 <code>data/</code>	Pre-packaged sample data and required files
 <code>part1_prepare_hurricane_data.ipynb</code>	Part 1: Data wrangling
 <code>part2_explore_hurricane_data.ipynb</code>	Part 2: Data exploration and analysis
 <code>part3_analyze_hurricane_data.ipynb</code>	Part 3: Statistical correlation analysis
 <code>notebook_basics/</code>	Sample Python notebooks
 <code>data/</code>	Pre-packaged sample data and required files
 <code>starter_notebook.ipynb</code>	Basic overview of working with Jupyter Notebook
 <code>notebook_basics.aprx</code>	ArcGIS Pro project configured for working with the
	 <code>arcgispro_notebook.ipynb</code> notebook
 <code>arcgispro_notebook.ipynb</code>	Overview of notebooks integrated into ArcGIS Pro
 <code>hosted_notebook.ipynb</code>	Overview of notebooks in ArcGIS Online/Enterprise

Part A: Getting Started

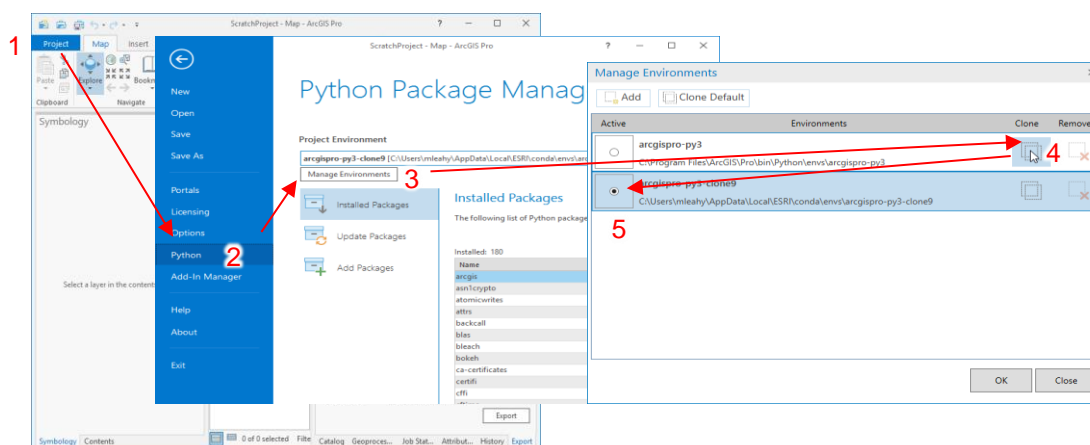
First, ensure that you have the following software installed:

- **ArcGIS Pro 2.5+**
 - *If necessary:* obtain a trial copy of ArcGIS Pro (<https://www.esri.com/en-us/arcgis/trial>)
 - *Optional:* authorize ArcGIS Pro for offline use (https://pro.arcgis.com/en/pro-app/get-started/start-arcgis-pro-with-a-named-user-license.htm#ESRI_SECTION1_15AD453E27C446CE9B51D45C021E8067)

With ArcGIS Pro 2.5 installed and licensed, prepare a clone of the default Python environment installed with ArcGIS Pro. To do this use the following steps:

1. Open the ArcGIS Pro settings (or when working on a project in Pro, click the **Project** tab)
2. Select **Python** in the left-hand menu to access the Python Package Manger interface
3. Click **Manage Environments**
4. Click the clone button () next to the default environment named **arcgispro-py3**

- Wait for the clone operation to complete, ensure that your new cloned environment is set as active, and close ArcGIS Pro.



- Now that you have your ArcGIS Pro Python environment cloned, you are free to modify it as you see fit. This tutorial will make use of several packages that are not installed by default. To install these, open the **Python Command Prompt** installed with ArcGIS Pro (Windows Start Menu > ArcGIS > Python Command Prompt). When the command prompt opens, you should see the name of the Python environment you just cloned in ArcGIS Pro to the left of the cursor and current path. Ensure your PC is connected to the Internet, then execute the following two commands:

```
conda install dask graphviz python-graphviz seaborn
```

```
conda install -c conda-forge jupyter_contrib_nbextensions
jupyter_nbextensions_configurator
```

The first command above will install a set of packages that are required for some of the sample notebooks included in this tutorial. The second command will install some additional packages that enable you to configure and use extensions that enhance the Jupyter Notebook system.

This installation process will take some time to complete, depending on your Internet connection and PC's hardware capacity. If you see any errors/warnings, you can safely ignore them. Do not close the command prompt until the process is completed, and the cursor has returned. When the process is completed, you can open the Jupyter Notebook server.

- Once you have completed the installation of the above packages, you will want to change the current drive and directory to the folder on your hard drive that contains the files for this tutorial copied earlier, and then launch the Jupyter Notebook server. You can do this with the following commands (adjust the drive letter and path in the first two commands as appropriate):

```
D:\
cd D:\arcgis_notebooks
jupyter notebook
```

The last command above will launch the Jupyter web interface, and you will see a view of the contents of the `arcgis_notebooks` directory.

8. Consider enabling the table of contents extension in Jupyter by navigating to the 'Nbextensions' tab in the Jupyter web interface, and clicking the checkbox next to the 'Table of Contents (2)' item (if this checkbox is greyed-out, you may need to uncheck the option near the top labelled 'disable configuration for nbextensions without explicit compatibility').

With this extension enabled, you can easily navigate within your notebooks with a table of contents that displays and synchronizes with headers in Markdown cells of your notebook.

Part B: Python Notebook Basics

For this portion of the tutorial, you will start by reviewing basic aspects of the Python Notebook environment, how to compose rich text and multimedia, and how to interact with code to execute Python commands and visualize outputs.

Using the Jupyter Notebook Web interface:

1. Within the web interface launched at the end of the steps in Part 1 (above), navigate into the `notebook_basics` folder, and open the `starter_notebook.ipynb` file.
 - At this point, you may choose to open the `documents/Jupyter_Cheatsheet.pdf` document, and keep it open for reference as you work with your notebooks in the Jupyter environment.
 - For a tour of the notebook interface, you can select 'User Interface Tour' in the Help menu.
2. To start the notebook from scratch, you can choose "Restart & Clear Output" from the Kernel menu.
3. You can follow the notebook by running code inside each code block in one of the following ways:
 - Select a cell with Python code in it and do one of the following:
 - i. Click the 'Run' button in the main toolbar.
 - ii. Press CTRL+Enter
 - iii. Press SHIFT+Enter (this automatically selects the next cell; repeat the SHIFT+Enter keystroke to execute multiple cells in sequential order)
 - iv. In the main Cell menu at the top, select one of the 'Run ...' options.

Note: one of the code cells in the `starter_notebook.ipynb` notebook will deliberately generate an error, so this notebook will not successfully run if you choose the 'Restart & Run All' option from the 'Kernel' menu.

Using ArcGIS Pro:

4. Open the `notebook_basics.aprx` project file located in the `notebook_basics` folder in ArcGIS Pro. In the Catalog pane, expand the Folders item, then the `notebook_basics`, then right-click on the `arcgispro_notebook.ipynb` file and choose 'Add To Project'. Expand the 'Notebooks' item in the Catalog pane, then double-click on the notebook file to open it.
5. Once the notebook is opened in ArcGIS Pro, you can follow its workflow by executing the code cells as described in step 3 above, or optionally choose 'Restart & Run All' option from the 'Kernel' menu.

Using ArcGIS Enterprise or ArcGIS Online:

6. Open a web browser and login to your ArcGIS Online organization with an account has access to advanced notebooks (currently in beta for ArcGIS Online), or login to your ArcGIS Enterprise portal that has a Notebook Server enabled with it (this must be installed and configured by your ArcGIS Enterprise administrator).

7. Navigate to the Content section, click 'Add Item' -> 'From my computer'. In the dialog that appears, choose the `hosted_notebook.ipynb` file from the `notebook_basics` folder, enter one or more tags, and click 'Add Item'.
8. When your notebook has been uploaded as a new item in your content, you will be taken to its details page. Click the 'Open Notebook' button near the top of the page to launch the hosted notebook in your web browser.
9. Follow this notebook by reading the Markdown text cells and executing code cells as described in step 3 above, or optionally by choosing the 'Restart & Run All' option from the 'Kernel' menu.

Note: some of the code in these notebooks includes processes that consume credits with ArcGIS Online services. If you need to skip any of these steps, do not choose the 'Restart & Run All' option from the 'Kernel' menu, and refer to the corresponding instructions included in the notebooks.

Part C: Data Science with ArcGIS and Python Notebooks

In this section of the tutorial, you step through a series of three notebooks that demonstrate how different stages of an end-to-end data science workflow can be created using ArcGIS Notebooks. These notebooks have been adapted from a set of [sample notebooks](#) prepared by Esri that prepare and analyze hurricane track data using GeoAnalytics in ArcGIS Enterprise. For this tutorial, they have been adapted to use the GeoAnalytics toolbox in ArcGIS Pro, enabling the analysis to be completed on a desktop PC with ArcGIS Pro installed.

The stages of the data science workflow demonstrated in this series of notebooks begin with data preparation (or 'data wrangling') in the first notebook. This is followed by data exploration in the second notebook, and finally correlation analysis in the third notebook.

Complete the following steps to work through the three notebooks for this section of the tutorial:

1. Launch the Jupyter Notebook web interface (by opening the Python Command Prompt, changing to the `arcgis_notebooks` folder, and executing the `jupyter notebook` command – as described in Part A of this tutorial), navigate into the `hurricane_analysis` folder, and open the `part1_prepare_hurricane_data.ipynb` file.
2. Optionally, restart the kernel and clear all output in the notebook (if you do not, previous outputs remain visible).
3. Follow the instructions within the notebook as you did in the previous Python Notebook Basics section of the tutorial.
4. Repeat steps 1-3 above for the `part2_explore_hurricane_tracks.ipynb` and `part3_analyze_hurricane_tracks.ipynb` notebooks.

Note: some of the code in these notebooks include long-running processes that require significant Internet bandwidth and consume credits with ArcGIS Online services. If you wish to skip any of these steps, do not choose the 'Restart & Run All' option from the 'Kernel' menu, and refer to the corresponding instructions included in the notebooks.

Summary

After completing this tutorial, you have learned how to work effectively with Jupyter Notebook to create documents that enable you to write and execute Python code interactively, visualize analysis outputs, and provide detailed narration using rich text and multimedia written in Markdown syntax. In Part B of the tutorial, you gained basic skills for using Jupyter Notebook server as a standalone application in Python, or as an embedded interface within ArcGIS Pro and ArcGIS Online / ArcGIS Enterprise. You also learned

how to integrate the capabilities of the ArcPy and ArcGIS API for Python libraries, alongside the capabilities of ArcGIS Pro and ArcGIS Online to create interactive and engaging ArcGIS Notebooks.

In Part C, you worked through a series of notebooks that demonstrate a complete end-to-end data science workflow for analyzing historical data for hurricanes. In these notebooks, you used ArcPy, the ArcGIS API for Python, open source Python libraries, and ArcGIS Online within a cohesive 3-part series of notebooks that demonstrate and articulate the data processing and analysis outputs throughout the exercise.

Future Considerations

You now have a good understanding of how to create ArcGIS Notebooks, and how they can be an effective medium for presenting detailed data processing and analysis workflows for data science. The next step is to consider how you can apply these skills to your current or future work. If you are processing or analyzing data for a project and need to document your workflows to share the results to others, consider adapting your work into the Jupyter Notebook environment. This will allow you easily to create and share notebook documents that show the exact methods you used so that it is easily repeatable for your own purposes as well as by others (i.e., the python code), with presentation of the results, alongside your explanation/interpretation with rich text and multimedia.

There are many additional examples and documentation sources available for you to learn more. The resources listed at the end of this document are a good starting point for reference as you continue developing your skills to work effectively with ArcGIS Notebooks.

Resources

- ArcGIS API for Python (API reference, guides, sample notebooks)
<https://developers.arcgis.com/python/>
- Hosted ArcGIS Notebooks Samples by Esri:
<https://www.arcgis.com/home/search.html?q=owner%3A%22ArcGISPyAPIBot%22>
- ArcGIS Pro Python reference (ArcPy)
<https://pro.arcgis.com/en/pro-app/arcpy/main/arcgis-pro-arcpy-reference.htm>
- Jupyter Notebook project (community, documentation, resources):
<https://jupyter.org/>
- Coding Standards for Jupyter Notebook
<https://www.esri.com/about/newsroom/arcuser/coding-standards-for-jupyter-notebook/>

© 2020 Esri Canada. All rights reserved. Trademarks provided under license from Environmental Systems Research Institute Inc. Other product and company names mentioned herein may be trademarks or registered trademarks of their respective owners. Errors and omissions excepted.

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/). The Education and Research Group at Esri Canada makes every effort to present accurate and reliable information. The Web sites and URLs used in this tutorial are from sources that were current at the time of production but are subject to change without notice to Esri Canada.

