Datapolitan

Data Solutions for the Modern Metropolis

tiny PANTHER

Data Analysis with Python

Follow along at: http://bit.ly/data-analysis-python

See the code at: http://bit.ly/data-analysis-python-code

Open your browser to:

http://student___.datapolitan.com/julia

Username: **rstudio** Password: **rstudio**

Richard Dunks

- Email: richard@datapolitan.com
- Website: http://www.datapolitan.com
- Twitter: @datapolitan

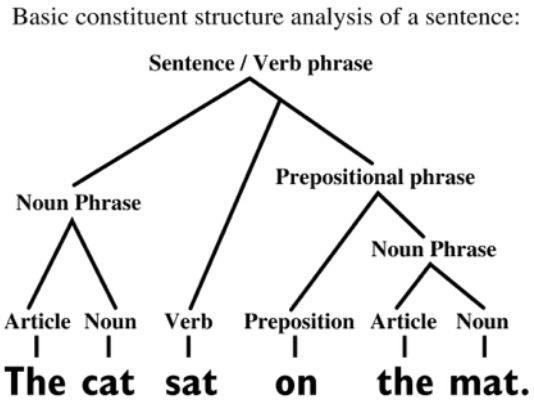
Julia Marden

- Email: julia@tinypanther.com
- Website: http://tinypanther.com
- Twitter: @juliaem

Resources

- Python for Data Analysis (http://shop.oreilly.com/product/0636920023784.do) - The textbook on using pandas for data analysis (2nd edition coming soon)
- Beginner's Python Tutorial (https://en.wikibooks.org/wiki/A_Beginner%27s_Python_Tutorial)- A good way to get started with basic tasks
- Stack Overflow (http://stackoverflow.com/)- One of the best Q&A sites for technical questions of all kinds

What is Syntax?



Python Syntax

- Variables hold some value
- We create variables and assign a value using the = sign
- We can perform operations with mathematical operators
- We can use built-in functions for operations
- Reference a particular column like df['Column Name']
- Use a dot (.) to call a function on an object

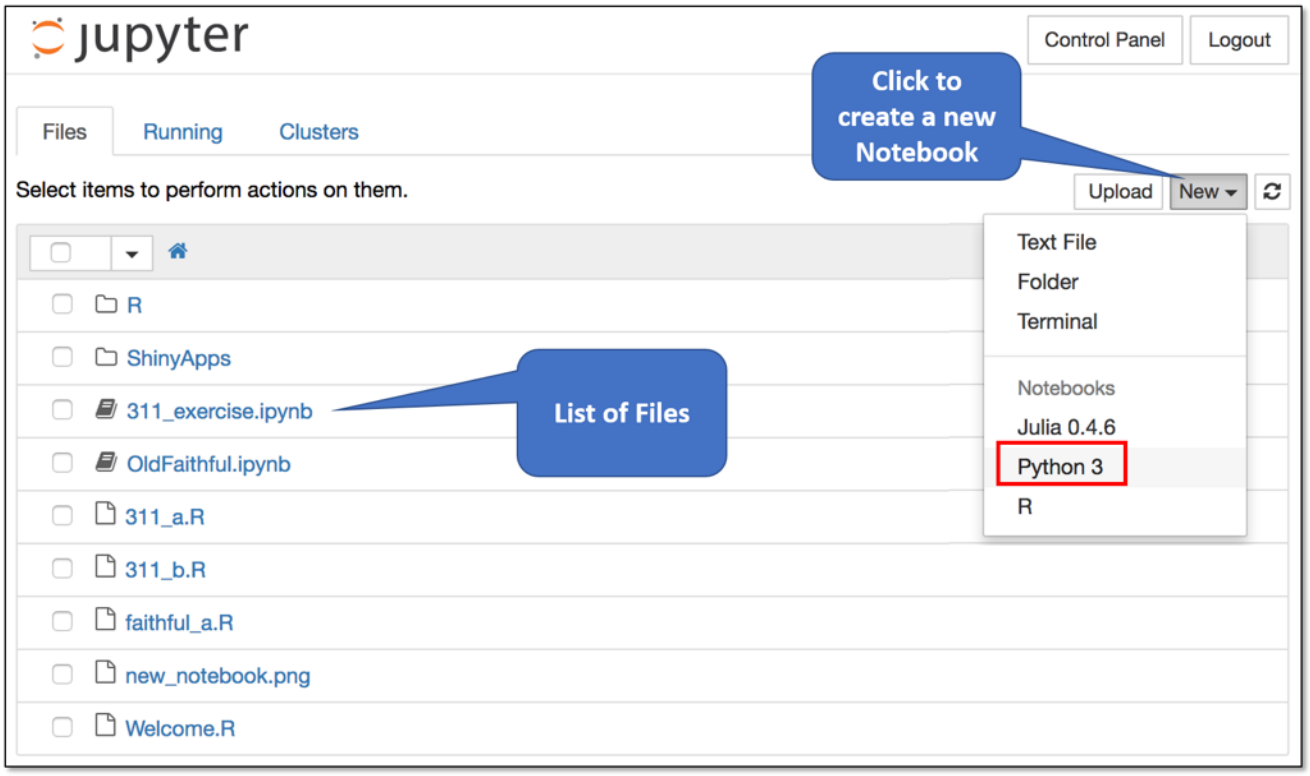
Key pandas Functions

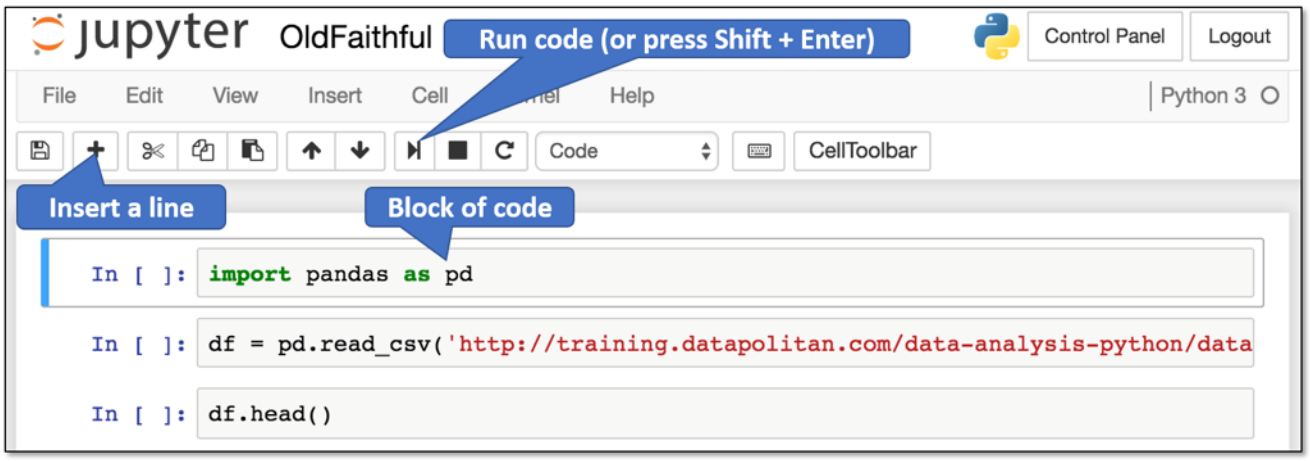
- **read_csv()** - import data from CSV into a DataFrame
- **read_excel()** - import data from .xls and .xlsx files into a DataFrame
- **head()** & **tail()** - first (head) and last (tail) 5 rows of DataFrame
- **count()** - count of all rows in a DataFrame column
- **max()** & **min()** - maximum and minimum values in a DataFrame column
- **mean()** & **median()** - mean and median values of numbers in a DataFrame column
- **describe()** - summary statistics for DataFrame
- **plot()** - plot data from a DataFrame
- **hist()** - create a histogram of values
- **groupby()** - group values together in data frame
- **sort_values()** - sort by values

Some Key Operations with Data

Selecting Multiple Columns -> df[[col1,col2,col3,...]] **Filtering Rows** -> df[df['Column Name']=='Value']

Grouping by Table Values -> df.groupby('Column you want to group')['Column you want to count'].count()


jupyter
Control Panel
Logout
Click to create a new Notebook
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
Text File
Folder
Terminal
Notebooks
Julia 0.4.6
Python 3
R
R
ShinyApps
311_exercise.ipynb
List of Files
OldFaithful.ipynb
311_a.R
311_b.R
faithful_a.R
new_notebook.png
Welcome.R


jupyter
OldFaithful
Run code (or press Shift + Enter)
Control Panel
Logout
File
Edit
View
Insert
Cell
Help
Python 3
Code
CellToolbar
Insert a line
Block of code
In [ ]: import pandas as pd
In [ ]: df = pd.read_csv('http://training.datapolitan.com/data-analysis-python/data
In [ ]: df.head()