

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

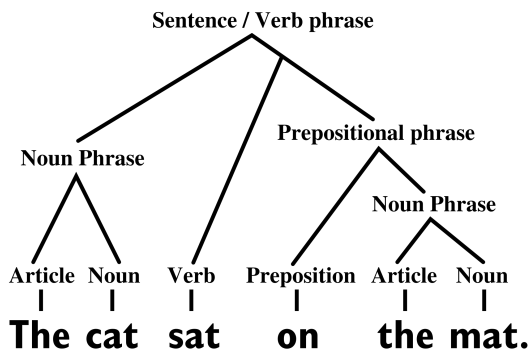
Password: rstudio

Key Questions for the Morning

What is python?	How is it useful in analysis?

What is Syntax?

Basic constituent structure analysis of a sentence:



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

The screenshot shows the Jupyter web interface. At the top, there's a "Control Panel" and "Logout" button. Below that, there are tabs for "Files", "Running", and "Clusters". A blue callout bubble points to the "New" button, saying "Click to create a new Notebook". The "New" dropdown menu is open, showing options: "Text File", "Folder", "Terminal", "Notebooks", "Julia 0.4.6", "Python 3" (highlighted with a red box), and "R". A blue callout bubble points to the file list, saying "List of Files". The file list includes: "R", "ShinyApps", "311_exercise.ipynb", "OldFaithful.ipynb", "311_a.R", "311_b.R", "faithful_a.R", "new_notebook.png", and "Welcome.R".

The screenshot shows the Jupyter code editor. At the top, there's a "Run code (or press Shift + Enter)" button. Below that, there's a menu bar with "File", "Edit", "View", "Insert", "Cell", "Kernel", and "Help". The "Cell" menu is open, showing options: "Run", "Insert", "Delete", "Duplicate", "Copy", "Paste", "Clear", "Code", and "CellToolbar". A blue callout bubble points to the "Run" button, saying "Run code (or press Shift + Enter)". The code editor shows three code cells:

```
In [ ]: import pandas as pd

In [ ]: df = pd.read_csv('http://training.datapolitan.com/data-analysis-python/data

In [ ]: df.head()
```

A blue callout bubble points to the first code cell, saying "Insert a line". Another blue callout bubble points to the second code cell, saying "Block of code".

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()`

Key Questions for the Afternoon

How is Python/pandas helpful for me?

What key tasks do I need to learn/practice?

Learning More

- **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
- **Style Guide for Python Code**- <https://www.python.org/dev/peps/pep-0008/>
- George Seif “23 great Pandas codes for Data Scientists” (<https://towardsdatascience.com/23-great-pandas-codes-for-data-scientists-cca5ed9d8a38>)
- Peter Gleeson “An A-Z of useful Python tricks” (<https://medium.freecodecamp.org/an-a-z-of-useful-python-tricks-b467524ee747>)

Other Resources

- Stack Overflow (<http://stackoverflow.com/>)- One of the best Q&A sites for various technical questions
- Datapolitan training classes <http://training.datapolitan.com/>

Contact Information

Email: training@datapolitan.com

Twitter: [@datapolitan](https://twitter.com/datapolitan)

Website: <http://www.datapolitan.com>

Your Notes