Statistical Computing with Python

Professor L. Jason Anastasopoulos
Web: http://anastasopoulos.io
Email: ljasonanastas@gmail.com

Course Preliminaries

Before the first day of class please:

➔ Download and Install the Latest Version of Anaconda Python 3.7+ Individual Edition: https://www.anaconda.com/products/individual

➔ Familiarize yourself with Google Colaboratory Python Notebooks: https://colab.research.google.com/notebooks/intro.ipynb

➔ Join our Slack Channel: https://statistical-computing.slack.com/

Google Drive Materials*

https://drive.google.com/drive/folders/1V8JXoNY4J2tPAbM7Jx4EOcuP9iHhFU5X?usp=sharing

*Check each day before class for updates to code and slides.

Course Schedule

Thursday, January 6: Day 1 - Semi-structured Data: Web-scraping and Databases.

Recommended Reading Before Class (Available Online)

- Web scraping:
  https://www.datacamp.com/community/tutorials/web-scraping-using-python
- MongoDB COVID-19 Data Demo:
  https://colab.research.google.com/drive/1M_wu0QqAWnml5_WWYSeMESuJp5pUjtx-?usp=sharing

10:00pm-2:00pm
Live Session: Lecture (Online, Recorded)

   I. Semi-structured data:
       ○ HTML parsing.
       ○ JSON parsing.
  II. Database creation and extraction:
       ○ Introduction to SQL.
       ○ Introduction to MongoDB.
       ○ Using MongoDB and SQL to store and retrieve data.
 III. Web-scraping and Databases Assignment and Review

2:00pm-4:00pm
Break

4:00pm-5:00pm
Lab 1: Review of Day 1 Assignment (Online, Recorded)

Friday, January 7th: Day 2 - Natural Language Processing and Machine Learning

Recommended Reading Before Class (Available Online)

- *Machine learning with scikit-learn:*
  https://jakevdp.github.io/PythonDataScienceHandbook/05.00-machine-learning.html
- *Natural language processing and sentiment analysis:*
  https://colab.research.google.com/drive/1Y-vJ49-Hw6zFkUx1ON1OYB4euC4vNKKJ?usp=sharing

10:00am-2:00pm (Online, Recorded)
Live Session: Lecture (Online, Recorded)

   I.   Collecting social media data with APIs.
  II.   Intro to natural language processing.
 III.   Applied natural language processing with social media data.
 III.   Sentiment analysis and machine learning with text data.

2:00pm-4:00pm
Break

4:00pm-5:00pm

Lab 2: Review of Day 2 Assignment (Online, Recorded)

Saturday, January 8th: Advanced Data Analysis and Visualization

Recommended Reading Before Class (Available Online)

- Python Data Science Handbook (Free Google Colab Notebooks) by Jake VanderPlas

    - Data manipulation with Pandas:
      https://jakevdp.github.io/PythonDataScienceHandbook/03.00-introduction-to-pandas.html
    - Big data with Spark in Google Colab:
      https://medium.com/@rmache/big-data-with-spark-in-google-colab-7c046e24b3
    - Introduction to Geopandas:
      https://colab.research.google.com/github/shakasom/GDS/blob/master/Part1%20-%20Introduction.ipynb

10:00am-2:00pm (Online, Recorded)
Live Session: Lecture (Online, Recorded)

I. General statistical inference (review)
    - Linear regression.
    - Generalized Linear Models.
    - Time series analysis.
II. Big data analysis and inference
    - Import/export of massive data sets.
    - Statistical inference with massive data sets.
III. Advanced topics in data visualization
    - Making beautiful plots with Seaborn.
    - Geospatial data visualization.