Data-Driven Problem Solving in Mechanical Engineering

Masoud Masoumi

ME 364 - Spring 2022

Department of Mechanical Engineering
The Cooper Union for the Advancement of Science and Art

January, 2022

Data-Driven *vs* First-Principle

First-Principle

- We know how the system works

- Can be based on theoretical explanations: Newton's laws, Hook's law, Bernoulli's Principle, Conservation of Energy, etc

- Reasoning based on the knowledge of domain: voters are not happy if the economy is bad so they vote for a specific candidate

Data-Driven

- Based on an observed correlation between input parameters and output variables

- Predicting tomorrow's weather or the price of a stock

- Possible to build a model on a domain we don't know anything about, if we are given enough good data

However, **best models are a mixture of both theory and data**

Engineers *vs* Scientists

Engineers

- Mainly hypothesis-driven and don't respect the data as much. Tend to make their own clean and organized virtual world.

- Traditionally have been taught that engineering/physics laws are the main things.

- Bad engineers worry about producing plausible numbers. Less invested in what can be learned from the data and analysis, as opposed to get it done quickly and efficiently.

Scientists

- Try to understand the natural world, which is complicated and messy! They are obsessed with discovering things.

- They are data-driven and spend a lot of time and effort collecting data to answer their questions.

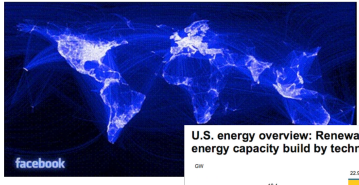
- They care about the answers and care about how the real world works. They are comfortable with the idea that data has errors.

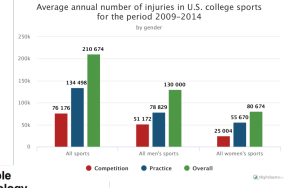
Why Data-Driven Now?

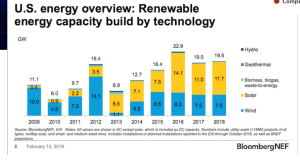
- A vast amounts of data from sensors and loggings

- Capability to analyze large data sets on scales, specially using accessible cloud computing

- Power of modern data analysis. Examples are Google, Facebook, hedge funds, energy companies, engineering firms, sports management, election forecasting, ...







Where to get Data?

It really depends on what type of data you are looking for

- The U.S. Census Bureau: https://data.census.gov
- U.S. government: https://www.data.gov
- Repository for Machine Learning: https://archive.ics.uci.edu/ml/index.php
- Kaggle: https://www.kaggle.com/datasets
- NHTSA: https://one.nhtsa.gov/Data
- NASA: https://nasa.github.io/data-nasa-gov-frontpage or https://api.nasa.gov
- New York City data: https://opendata.cityofnewyork.us
- Google Ngrams: https://books.google.com/ngrams
- IMDB: https://www.imdb.com/interfaces
- Federal Reserve Economic Data: https://fred.stlouisfed.org
- **Produce your own data to answer a specific question!**

Properties of Data

(a) Structured *vs* Unstructured Data

- **Structured data** examples: tables in a database or spreadsheet programs. These are typically represented in the form of matrices or dataframes.
- **Unstructured data** examples: Large text corpus with images and links like Wikipedia, personal medical records complied of notes and tests results, collection of tweets from Twitter.

(b) Quantitative *vs* Categorical Data

- **Quantitative data** consists of numerical values like height and weight
- **Categorical data** describe the properties such as hair color, gender, occupation.

(c) Big Data *vs* Little Data

Big Data is typically the product of some loggin process that records discrete events, or distributed contributions from millions of people over social media. Examples: https://www.internetlivestats.com

Data Science in Mechanical Engineering

"Today, data is abundant and abundantly collected in each single experiment at a very small cost. Data-driven modeling and scientific discovery is a change of paradigm on how many problems, both in science and engineering, are addressed. " ("Data-driven modeling and learning in science and engineering." Comptes Rendus Mécanique 347.11 (2019): 845-855.)

- **Fluid Mechanics**: Turbulence modeling, flow control, optimal design of aerodynamic structures

- **Solid Mechanics**: Complex composite materials, nonlinear solids, biological materials, assimilating data from testings and observations, fatigue cracks

- **Biomechanics**: Characterizing soft tissues, cells, and their behavior

- **Vibrations & Control**: Structural health monitoring, system identification, nonlinear control design

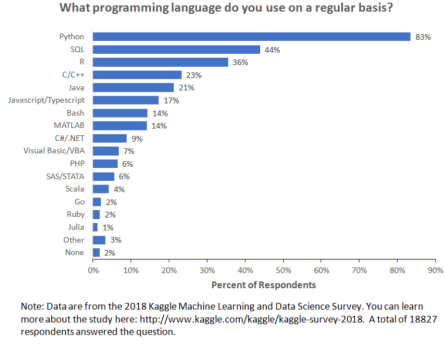
- **Renewable Energies**: Renewable energy system modeling, solar/wind/ocean wave energy resource mapping, renewable energy forecasting

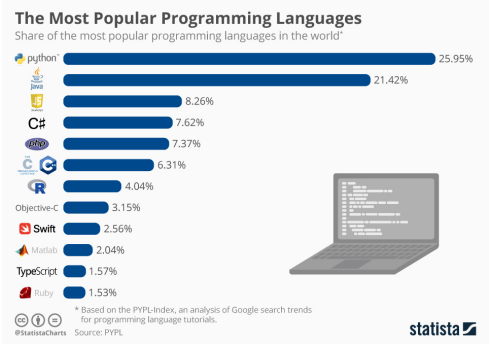
Why Python Notebook for Data Analysis?

- Python contains variety of libraries to easily perform data analysis tasks, from scraping to visualization, to linear algebra, and machine learning.

- We can provide a computable notebook environment, including the code, data, computational results, and written analysis and interpretations.

- It is the most common programming language for data analysis and machine learning





References for the Course

Theory & Concepts:

-"The Data Science Design Manual" by *Steven S. Skiena* (ISBN: 978-3319554433) [Website: http://www.data-manual.com]

-"Learning from Data: A Short Course" by *Abu-Mostafa, Yaser S., Malik Magdon-Ismail, and Hsuan-Tien Lin* (ISBN: 978-1600490064) [Website: http://www.AMLbook.com]

Programming:

-"Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython" by *Wes McKinney*, 2^{nd} Edition (ISBN: 978-1491957660) [Website: https://github.com/wesm/pydata-book]

-"Python Data Science Handbook. Essential Tools for Working with Data" by *Jake VanderPlas* (ISBN: 978-1491912058) [Website: https://github.com/jakevdp/PythonDataScienceHandbook/tree/master/notebooks]