

# INTRO to DATA SCIENCE

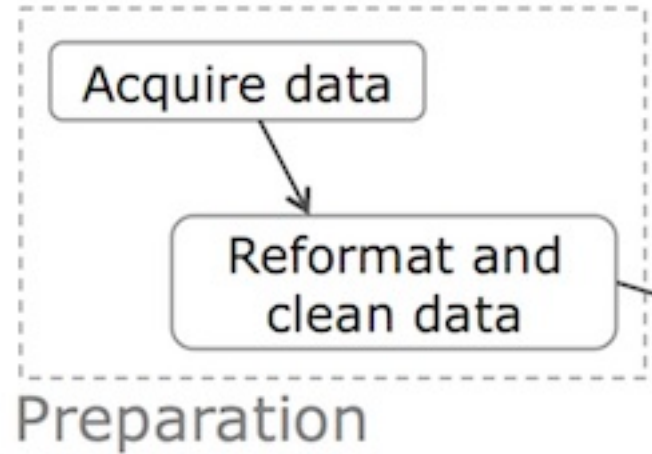
## RECAP

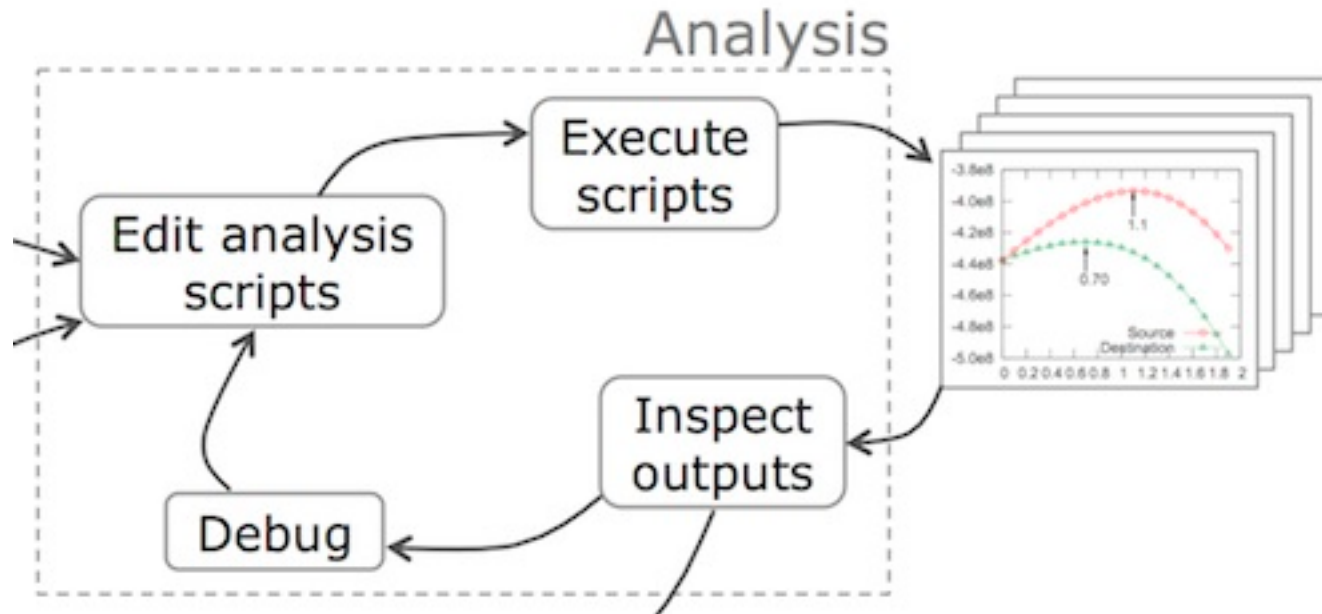
---

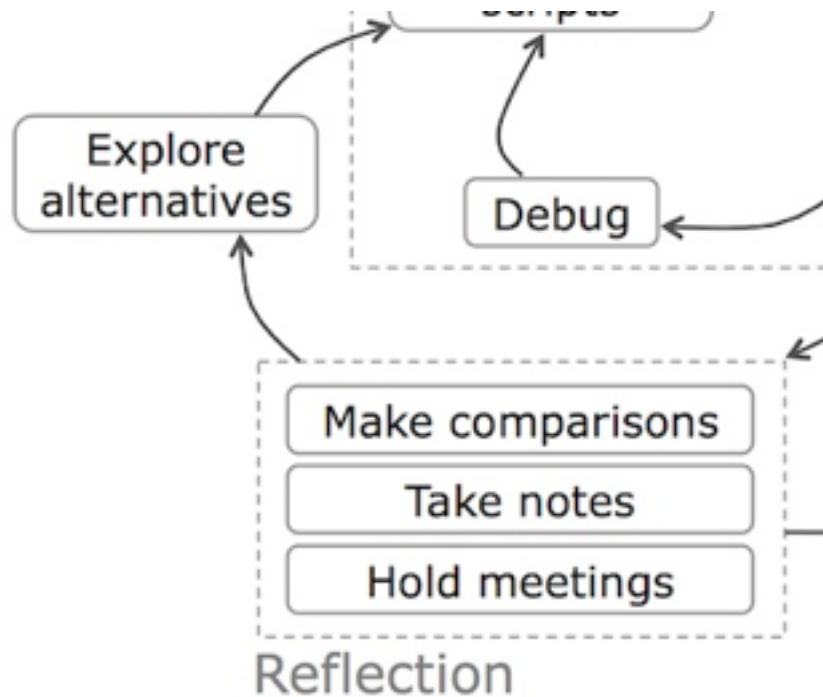
**INTRO TO DATA SCIENCE**

---

**DATA SCIENCE SKILLS**

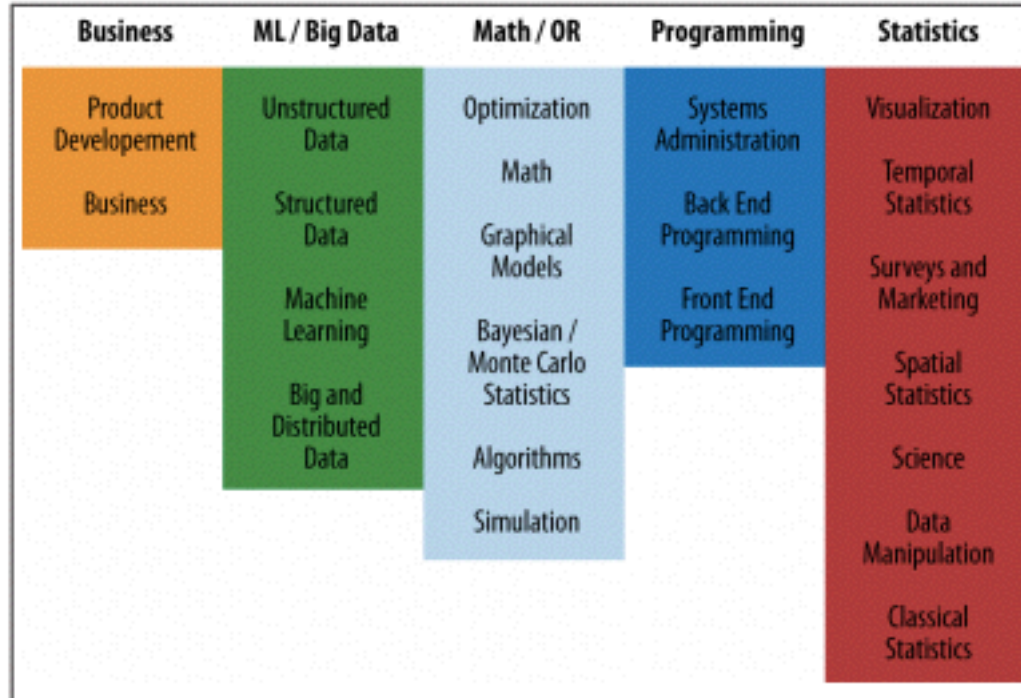






|                     |                    |                |              |
|---------------------|--------------------|----------------|--------------|
| Data Developer      | Developer          | Engineer       |              |
| Data Researcher     | Researcher         | Scientist      | Statistician |
| Data Creative       | Jack of All Trades | Artist         | Hacker       |
| Data Businessperson | Leader             | Businessperson | Entrepreneur |







---

**INTRO TO DATA SCIENCE**

---

**FURTHER TOPICS**

## Data Visualization

- FlowingData
- plot.ly
- Grammar of Graphics
- D3.js
- Tableau

## Data Processing

- Command line tools
- Connecting Python and the database

## Statistics

- Experimental design
- Monte Carlo simulation
- Causality testing

## Learning methods

- Neural Networks
- Support Vector Machines
- Latent variable models
- Non-linear dimensionality reduction/Manifold learning
- Recommender systems/Collaborative filtering
- Outlier detection

## Other Languages

- **R:** For the latest statistical models, better visualization
- **Julia:** Faster combination of R + Python (very new)
- **Scala:** Many utilities for Hadoop and Spark

---

**INTRO TO DATA SCIENCE**

---

**DATA SCIENCE IN NY**

### Meetups:

- [NYC Python](#)
- [Data Driven NYC](#)
- [Machine Learning](#)
- [Open Analytics](#)
- [NYC Data Engineering](#)
- [Hadoop NYC](#)
- [DataKind](#)



### Conferences:

- [Strata + HadoopWorld](#)
- [DataGotham](#)
- [PyData](#)
- KDD (in NYC this summer)