Requirements

School of Information Studies
Syracuse University

Graded – Exercises

- Practical exercises that build on readings
- A total of three (3) exercises, with each exercise being worth 15 points
- Exercises will increase in complexity
- Exercises will align with the course readings (generally)
- Exercises may be (and are encouraged to be) discussed among students and during sessions
- Exercises are individual work

Graded – Course Participation

- One of the coolest things about information analytics is the opportunity to learn as you go
- Each week, the async sessions provide you – the student – a chance to explore the theory and practice of analytics
- A total of ten (10) weekly sessions with each week/module being worth 1 point
- Maximum points are possible if the student completes the questions and activities prior to the weekly group discussion

Graded – Course Project

- The most extensive exercise of the course
- One course project worth 35 points
  - Project proposal worth 3 points
  - Project status worth 2 points
  - Final project worth 30 points
- Aligned with overall course goal:
  - *Be able to match available analytical methodologies to the information needs of clients and users, and communicate the results in a meaningful way.*

Graded – Discussions

- Weekly discussions on topics related to information analytics
- A total of ten (10) discussion posts, with each week/discussion being worth 1 point
- Discussions emphasize collaborative learning between peers, similar to the case method
- Discussions will align with the course readings and assigned case studies (generally)
- In the weekly session, there is not “a correct answer,” but not participating is an incorrect answer

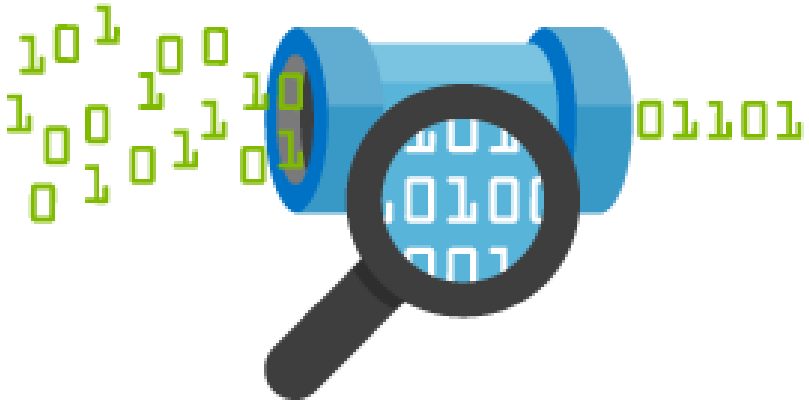
OSEMiN

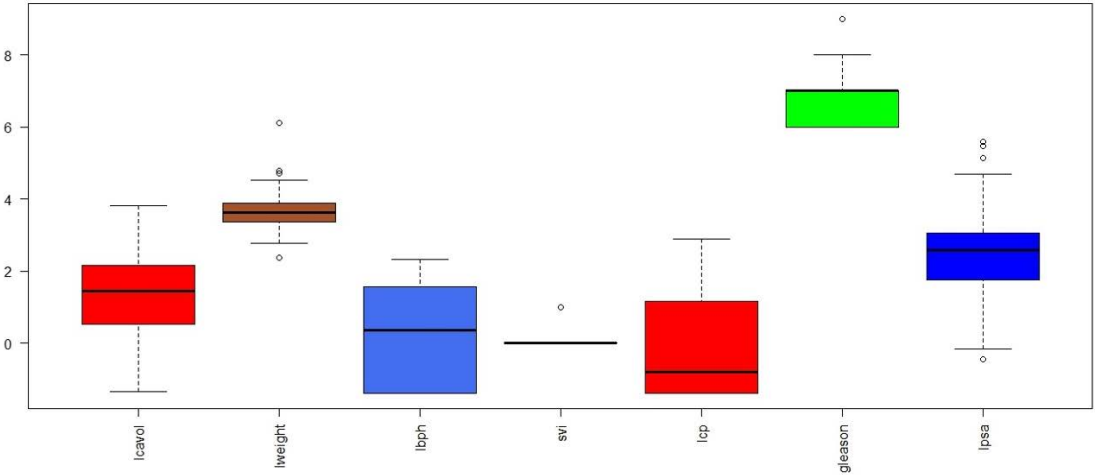
School of Information Studies
Syracuse University

Doing Data Science

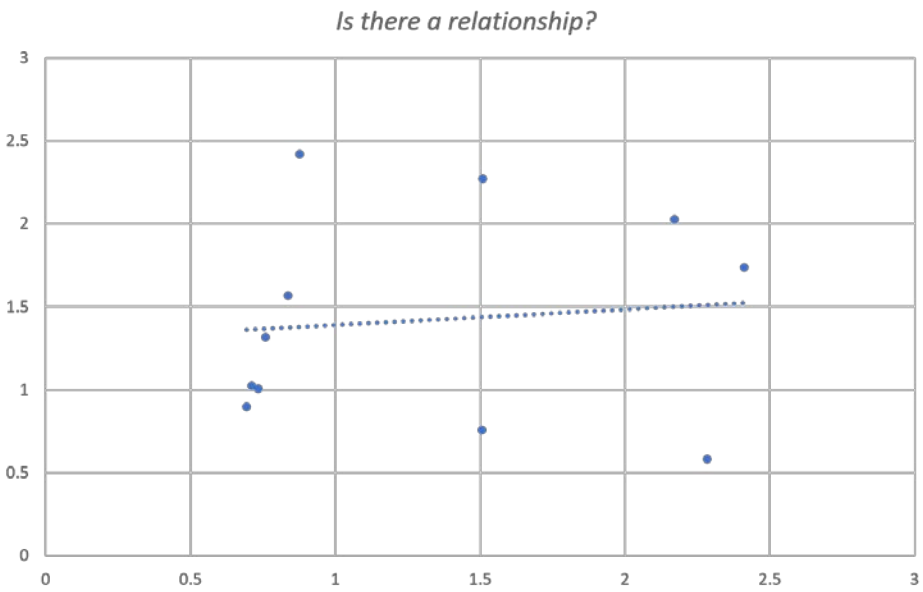
- Data analytics is an approach for turning data into information
- Frawley et al (1992) define knowledge discovery as the nontrivial extraction of information from data
- Several steps in a data analytic approach
  - Collection, preparation, **analysis**, visualization, management, and preservation
- O'Neil & Schutt (2013) describe this process as **OSEMiN**

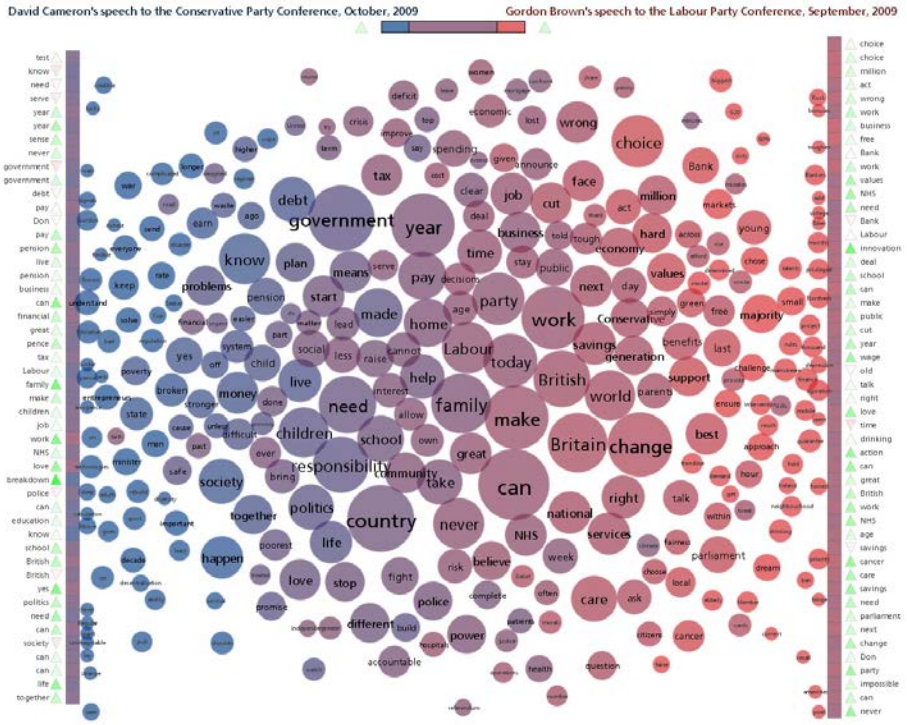
Obtain – Scrub – Explore





Model – Interpret



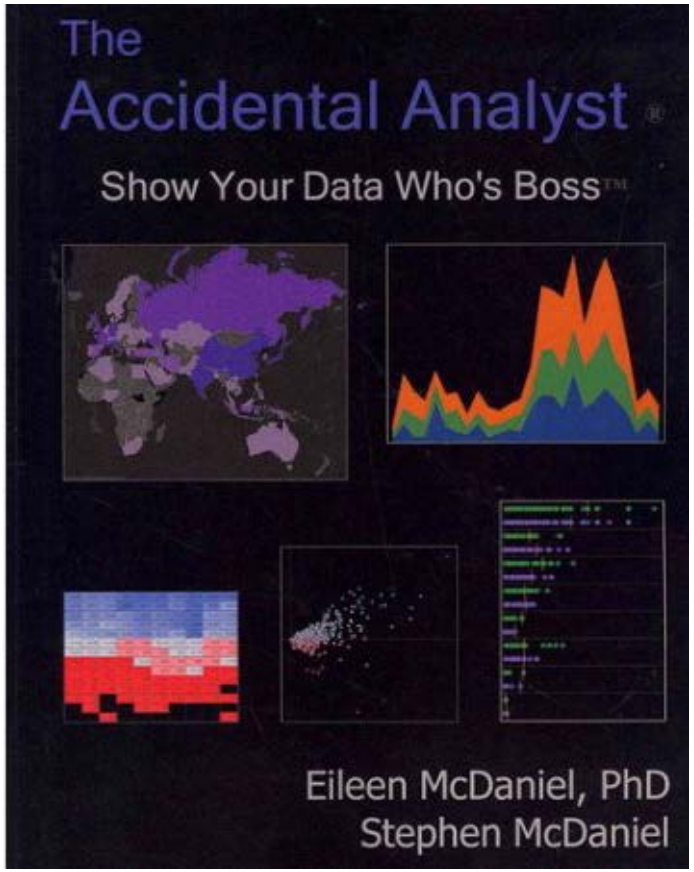


School of Information Studies
Syracuse University

Alternatives: Knowledge Discovery

Alternatives: The 7 C's

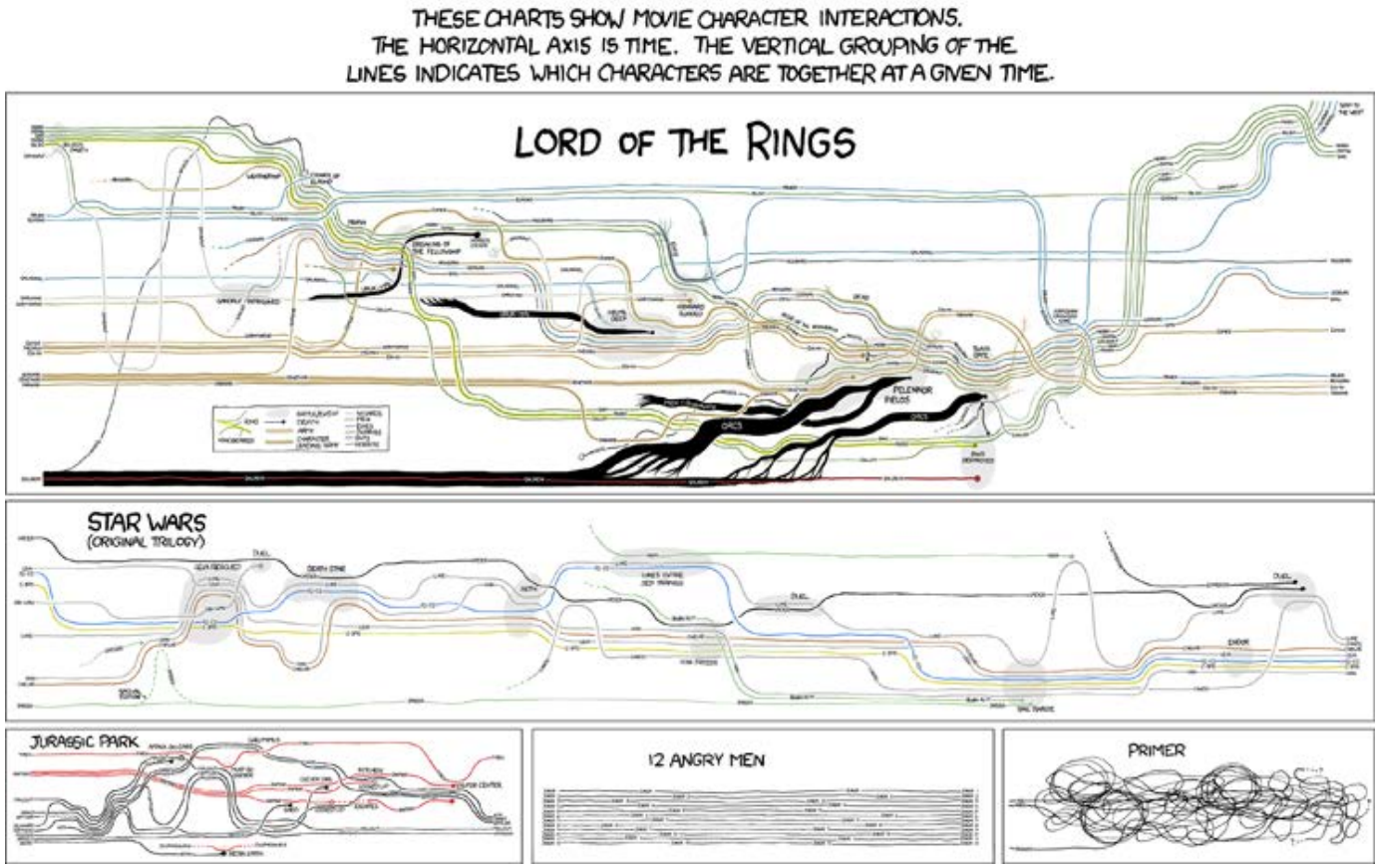
- Choose your questions
- Collect your data
- Check out your data
- Clean up your data
- Chart your analysis
- Customize your analysis
- Communicate your results



Crawl

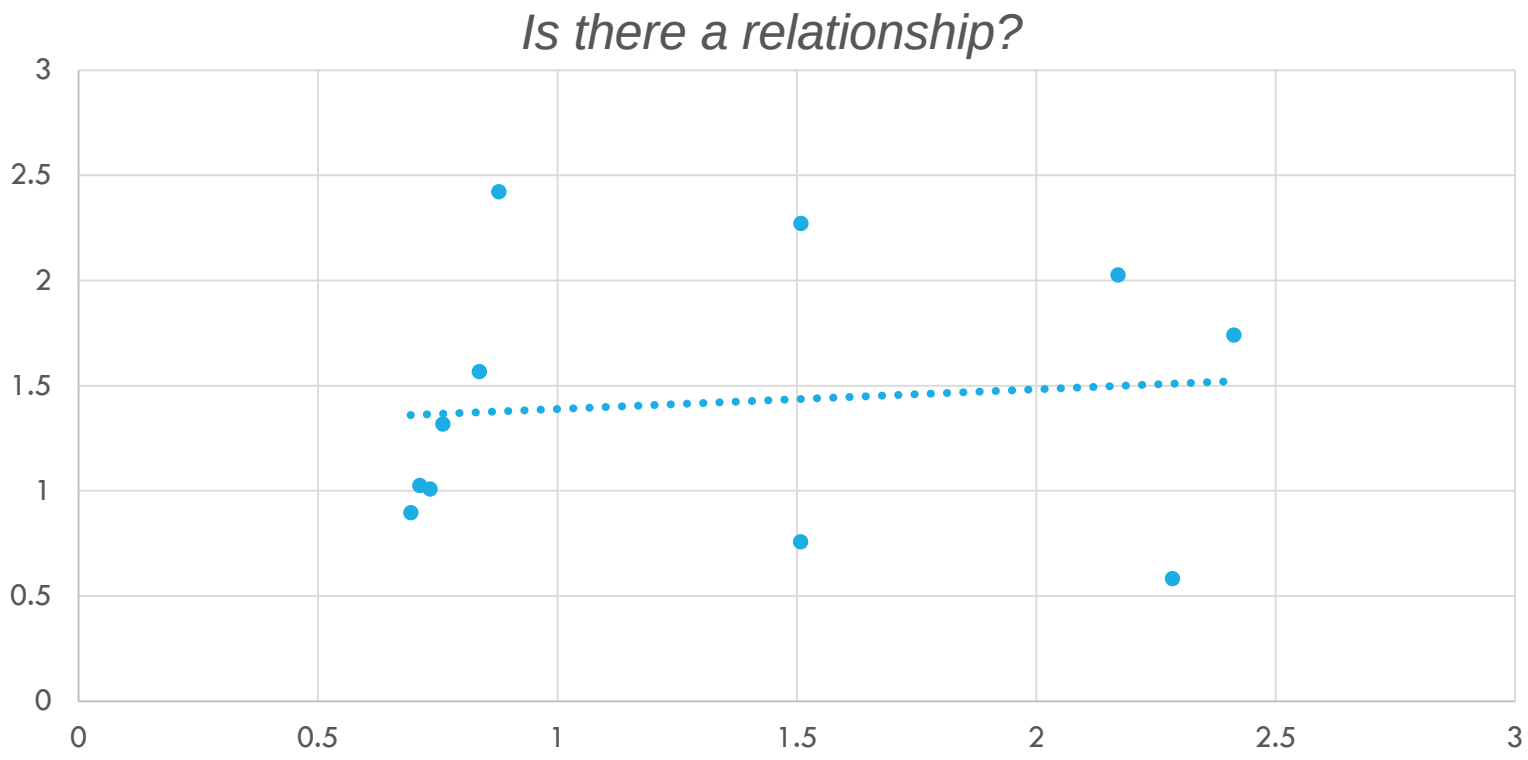
School of Information Studies
Syracuse University

Describing

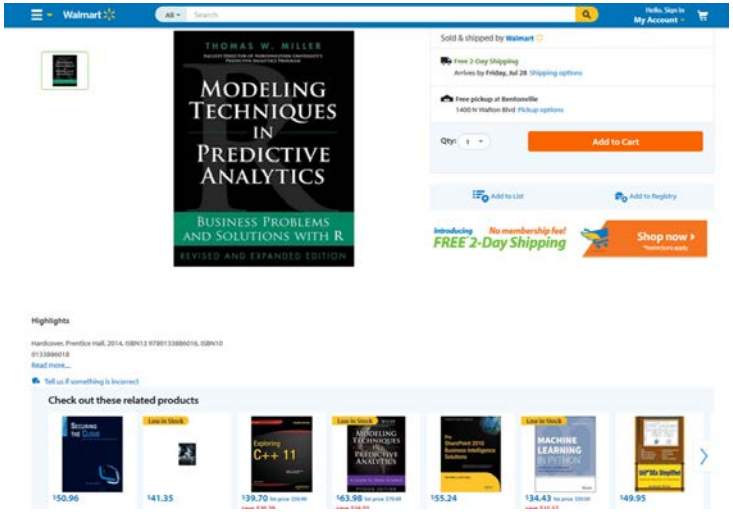


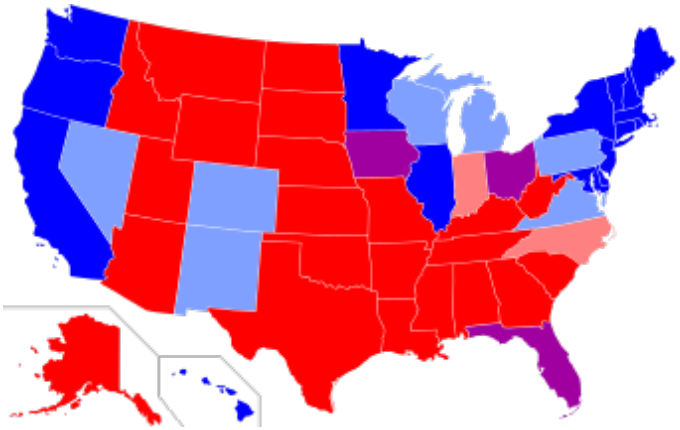
School of Information Studies
Syracuse University

Modeling



Choosing



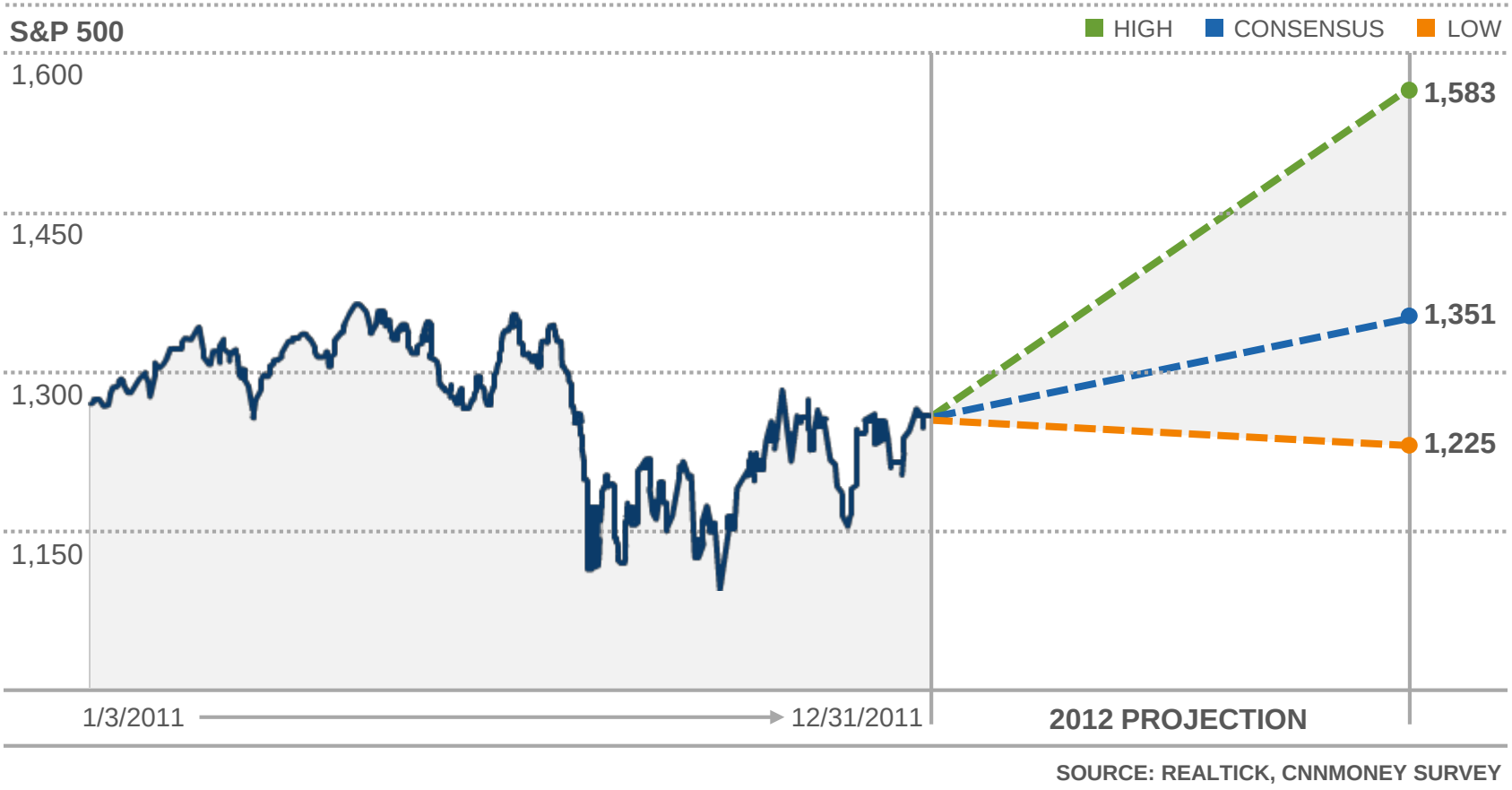


School of Information Studies
Syracuse University

Walk

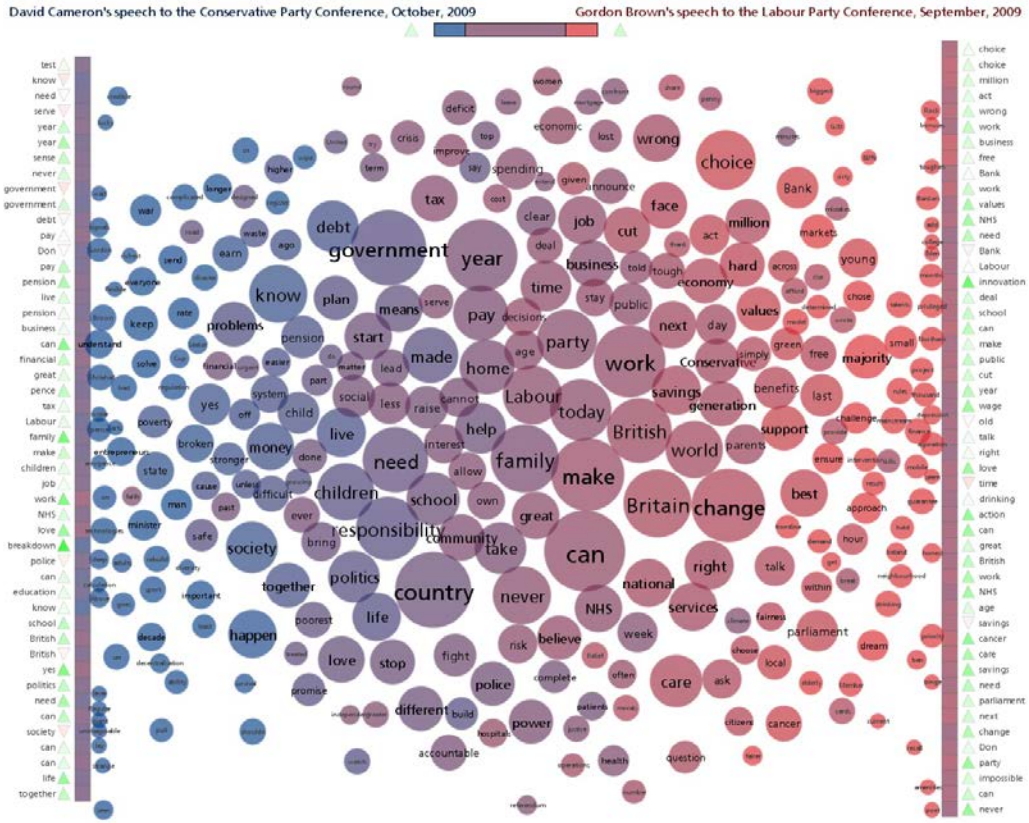
School of Information Studies
Syracuse University

Forecasting



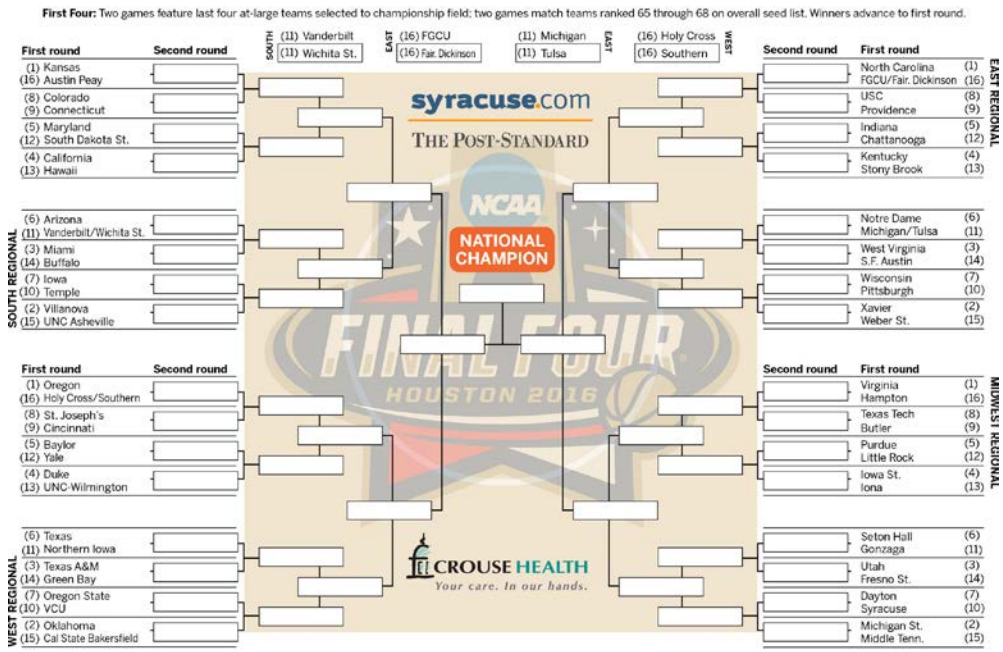
School of Information Studies
Syracuse University

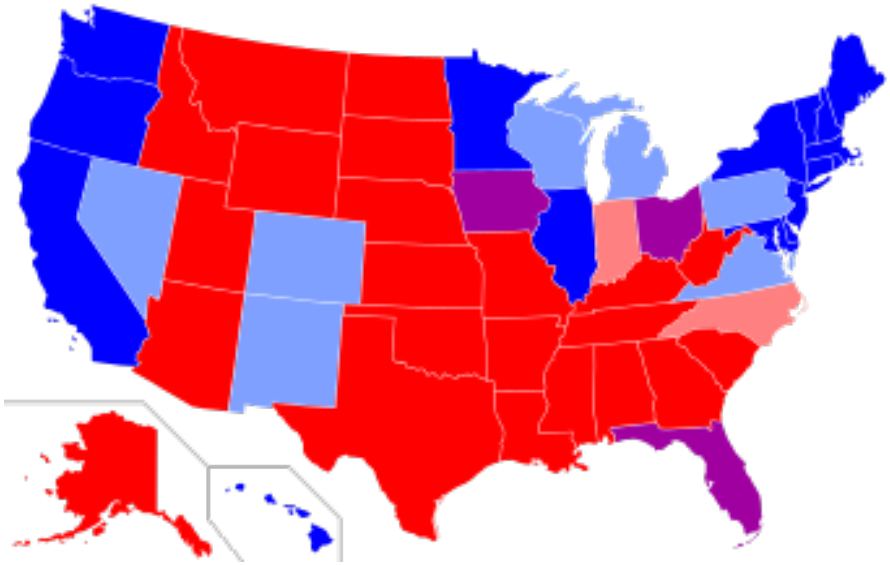
Inferring



School of Information Studies
Syracuse University

Picking





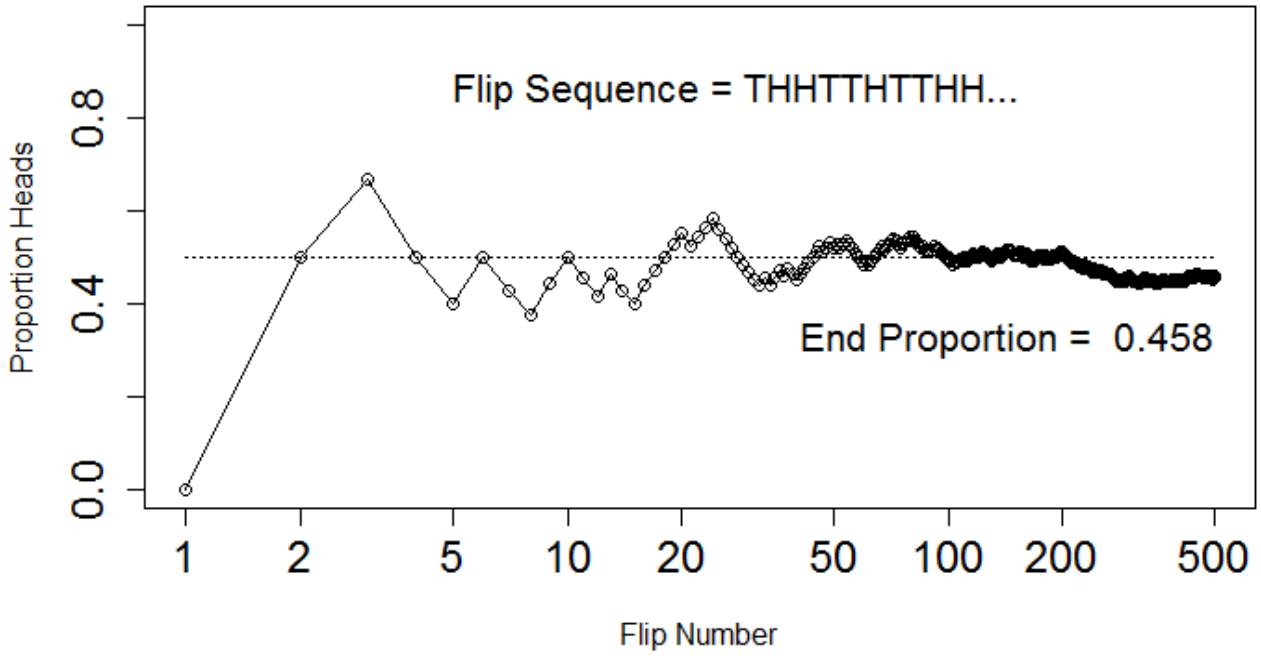
aka
choosing again

School of Information Studies
Syracuse University

Run

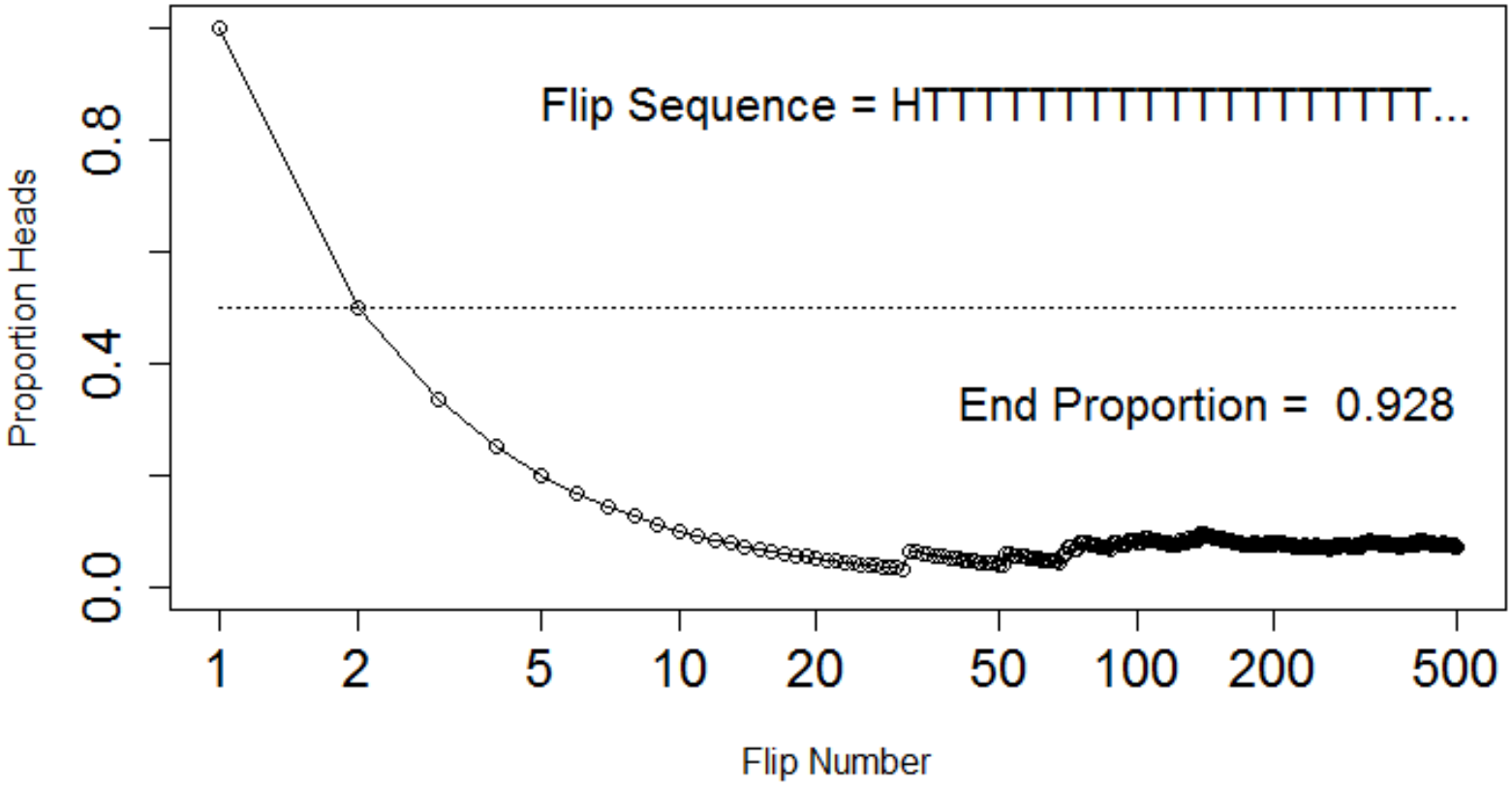
School of Information Studies
Syracuse University

Choosing Again



$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

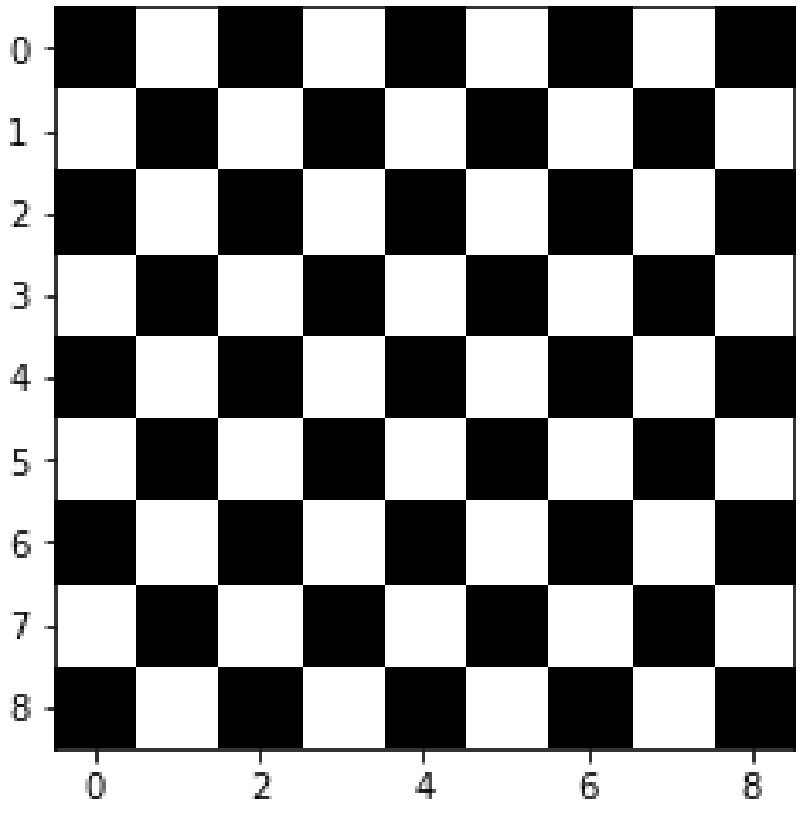
Biased Coin?

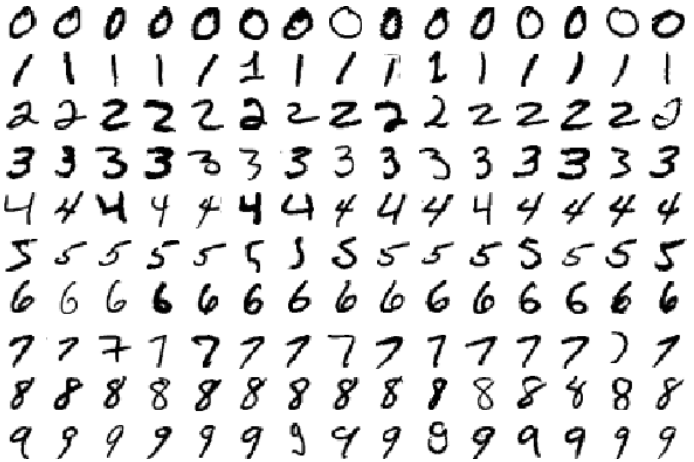

Flip Sequence = HTTTTTTTTTTTTTTTTTT...
End Proportion = 0.928
Proportion Heads
0.0
0.4
0.8
1
2
5
10
20
50
100
200
500
Flip Number

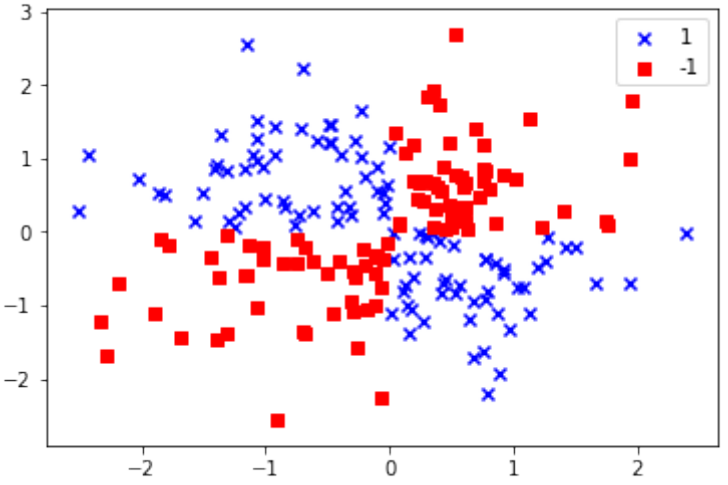
Machine Learning

Microsoft CNTK

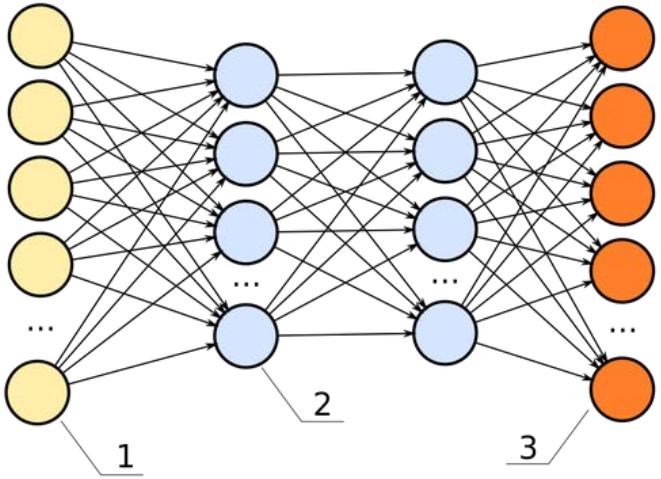
Machine Learning

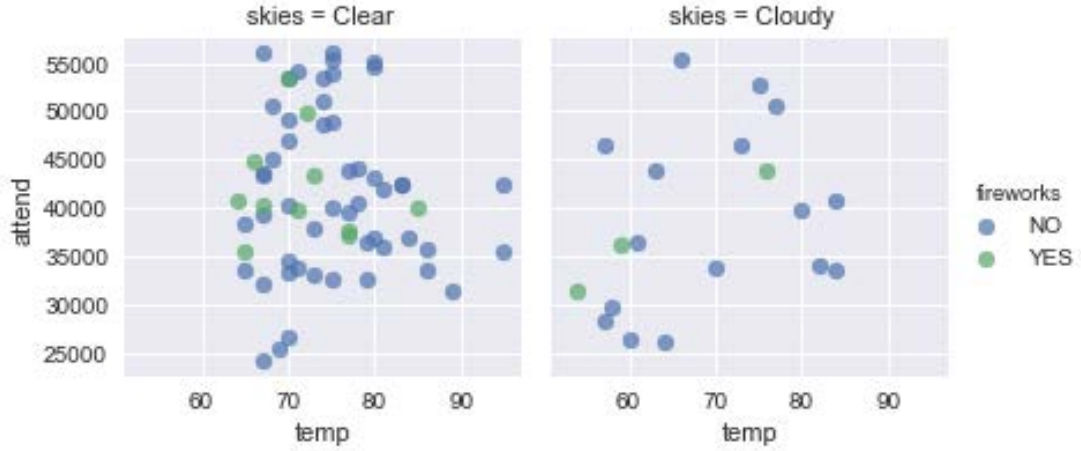


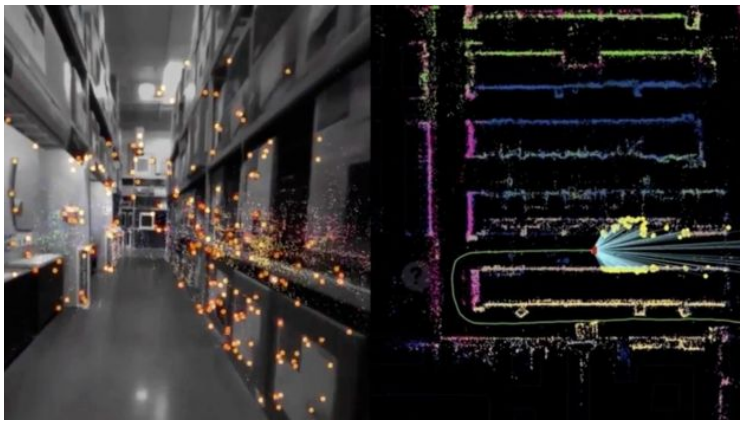




Machine Learning



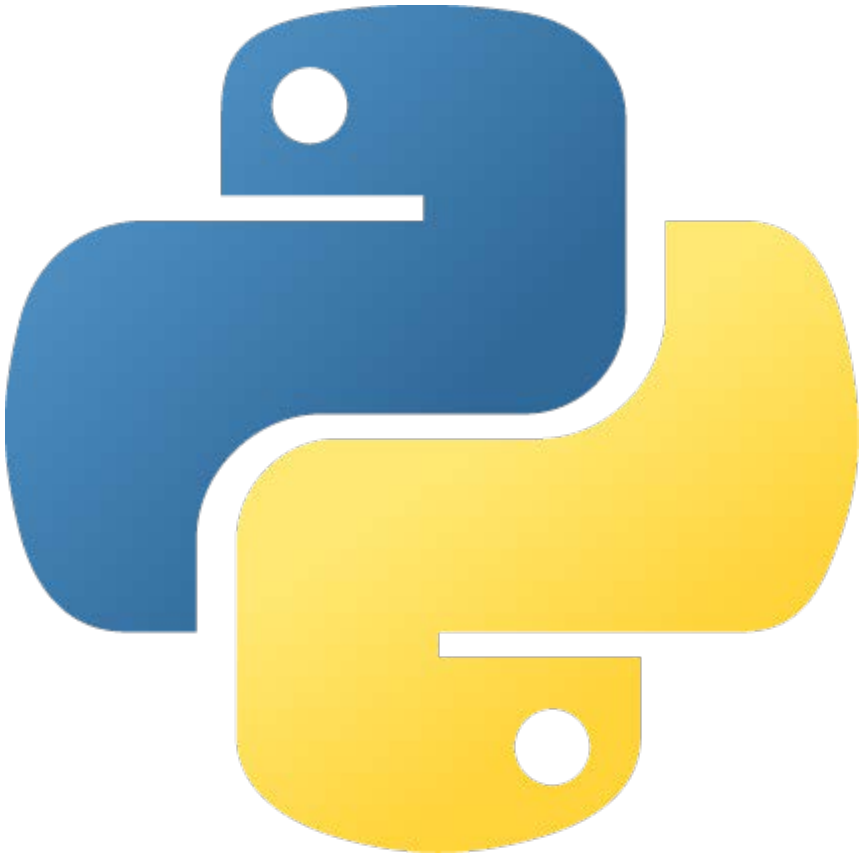




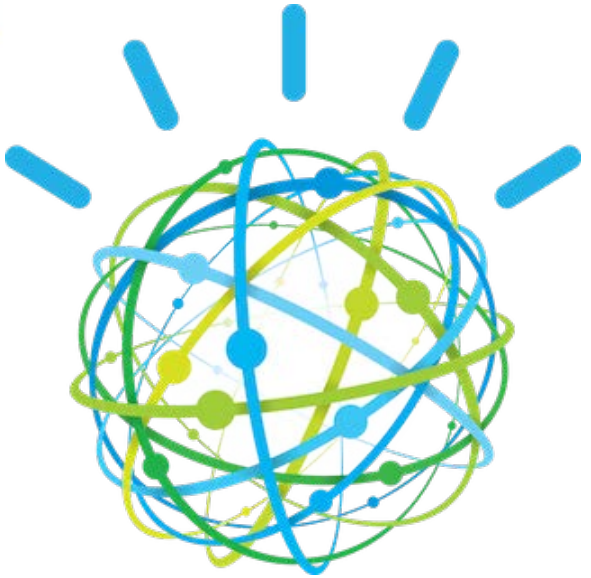
Tools

School of Information Studies
Syracuse University

Python



Anaconda

Spark

Others