



# 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



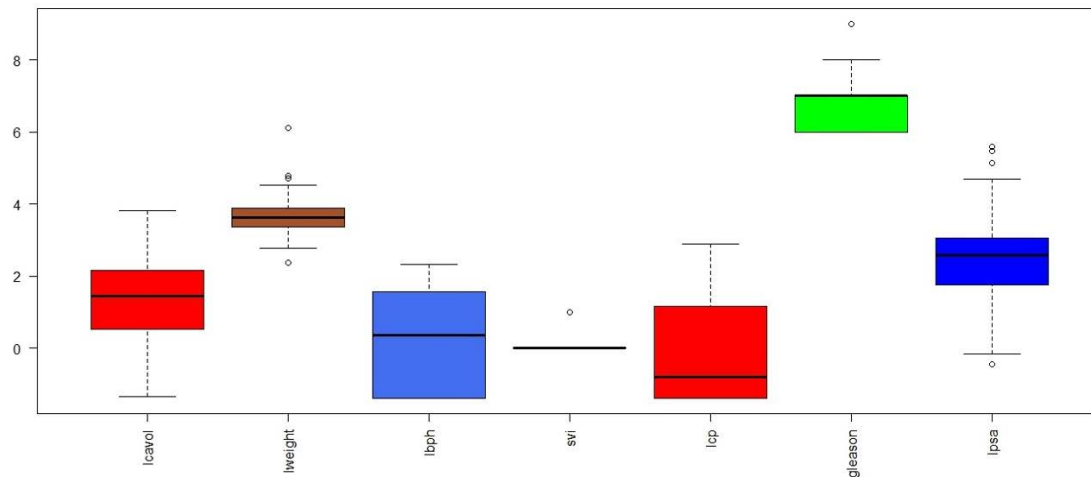
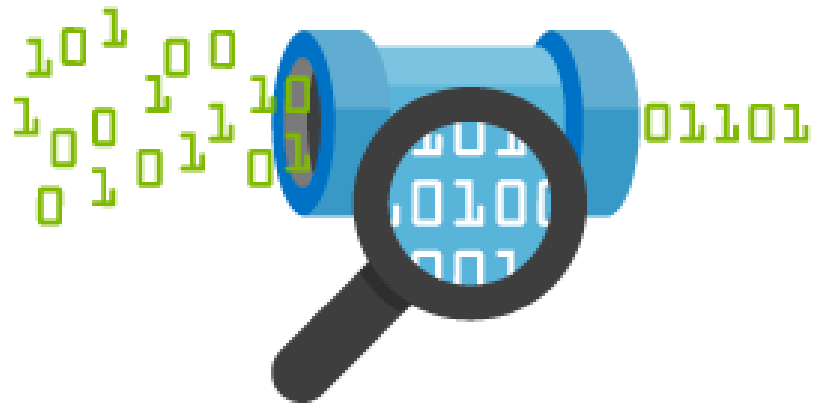
OSEMiN

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**

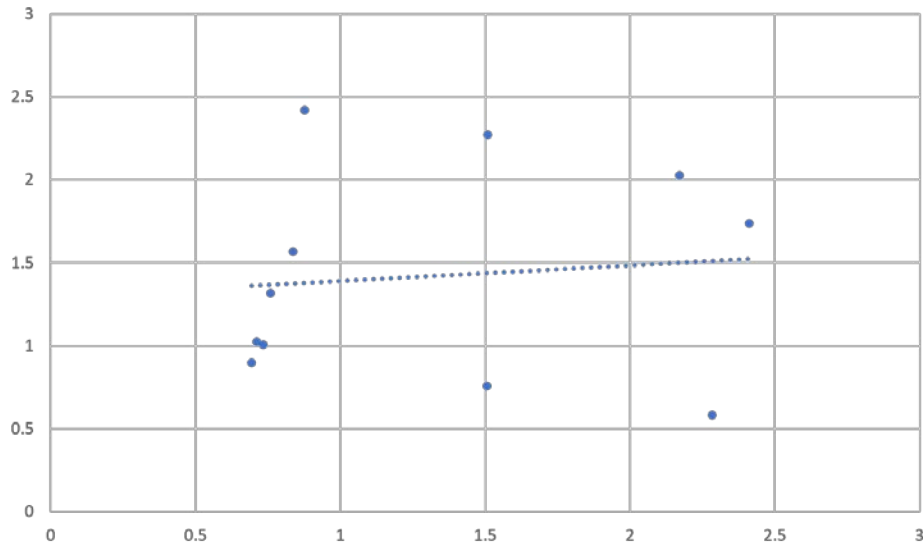
# Obtain – Scrub – Explore





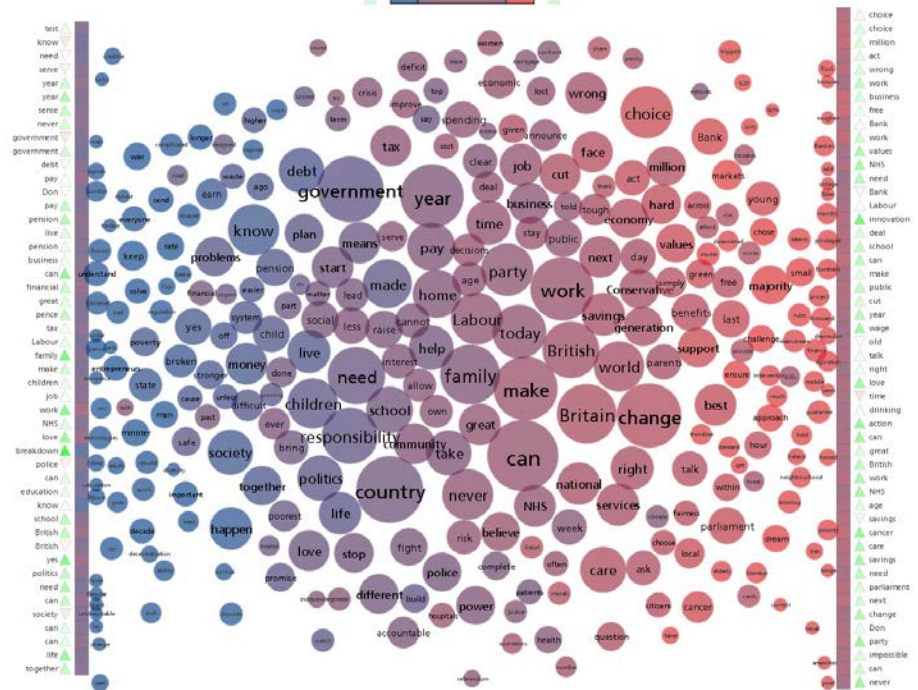
# Model – Interpret

*Is there a relationship?*

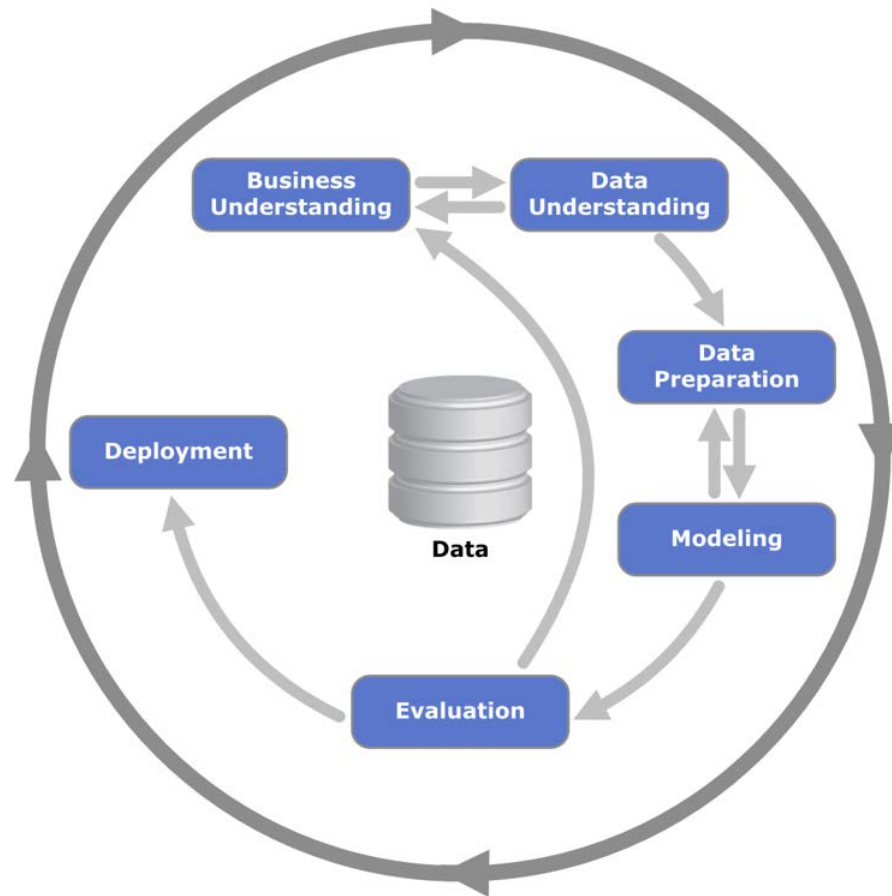


David Cameron's speech to the Conservative Party Conference, October, 2009

Gordon Brown's speech to the Labour Party Conference, September, 2009

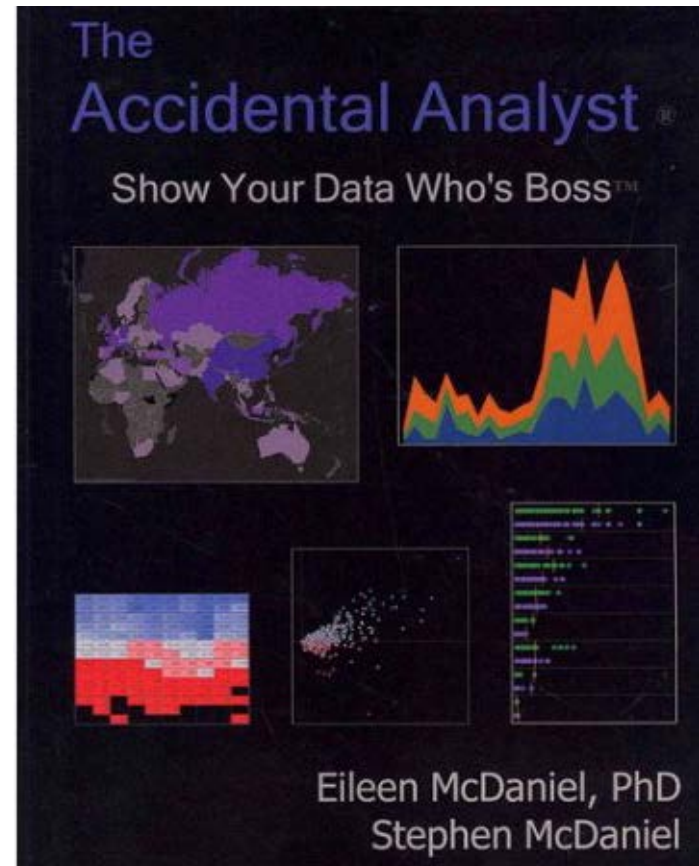


# Alternatives: Knowledge Discovery



# 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







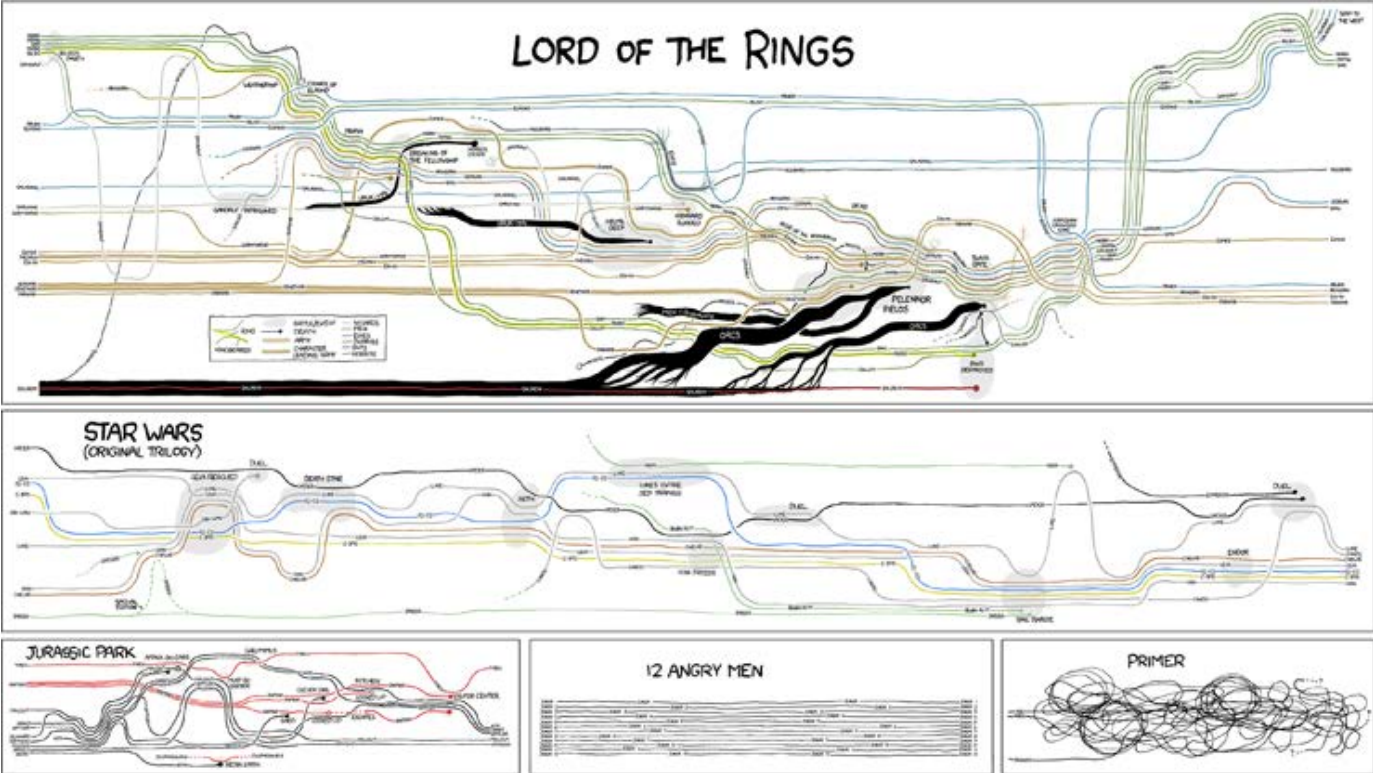
Crawl

School of Information Studies  
Syracuse University

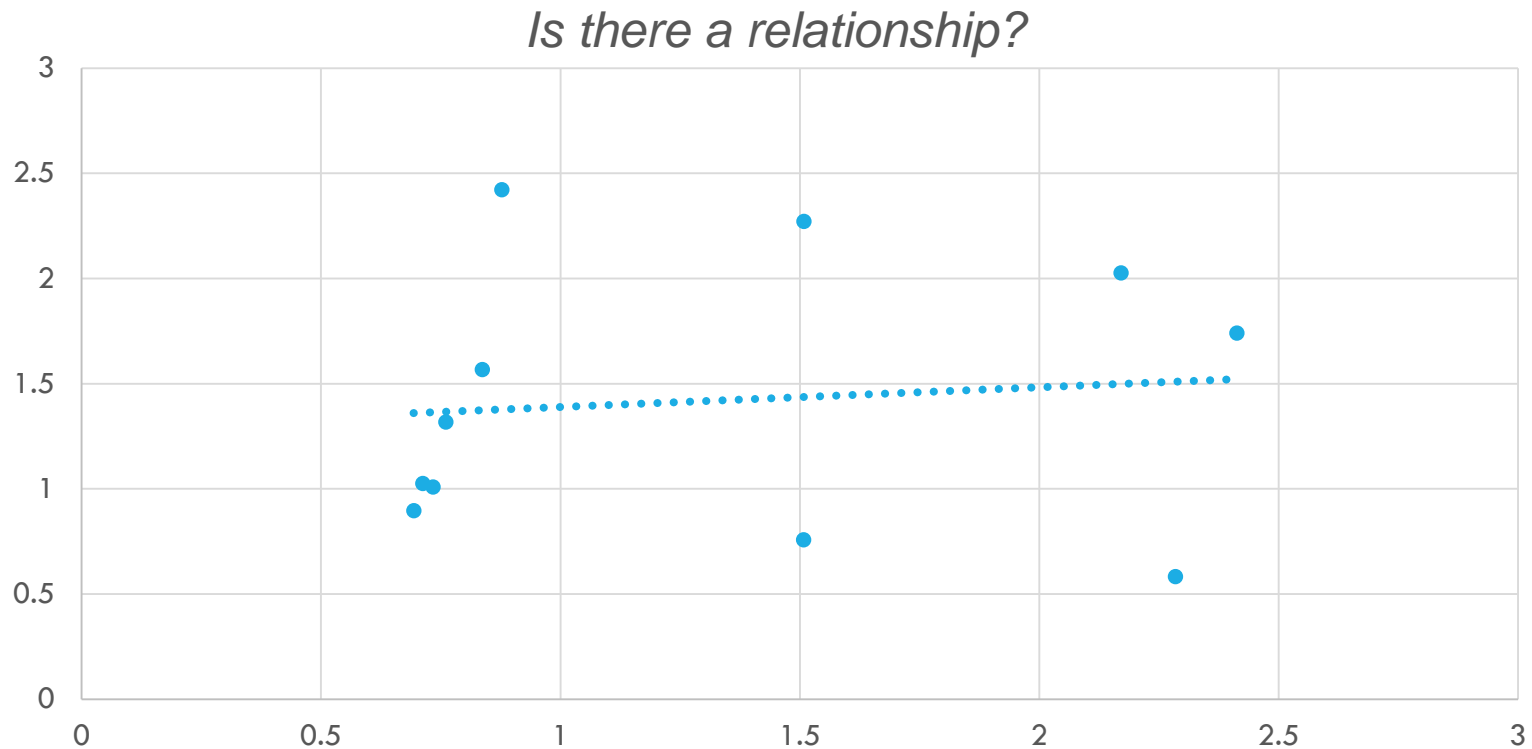


1

THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.  
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE  
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.



# Modeling



[Home](#)
[My Account](#)

**MODELING TECHNIQUES IN PREDICTIVE ANALYTICS**

**BUSINESS PROBLEMS AND SOLUTIONS WITH R**

REVISED AND EXPANDED EDITION

Sold & shipped by Walmart

**Free 2-Day Shipping**  
Arrives by Friday, Jul 28 [Shipping options](#)

**Free pickup at Bentonville**  
140019 Walmart Blvd [Pickup options](#)

Qty:

[Add to List](#) [Add to Registry](#)

**Introducing *No membership fee!* FREE 2-Day Shipping** **Shop now** [New members only](#)

**Highlights**

Hardcover, Prentice Hall, 2014, ISBN13: 9780133880416, ISBN10: 0133880418

[Read more...](#)

[Tell us if something is incorrect](#)

**Check out these related products**

150.96

141.35

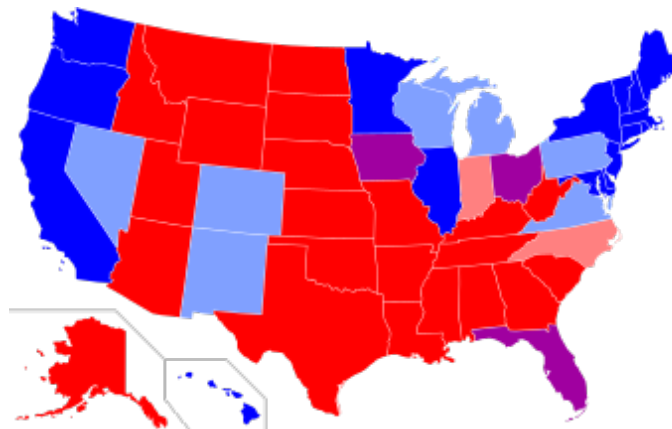
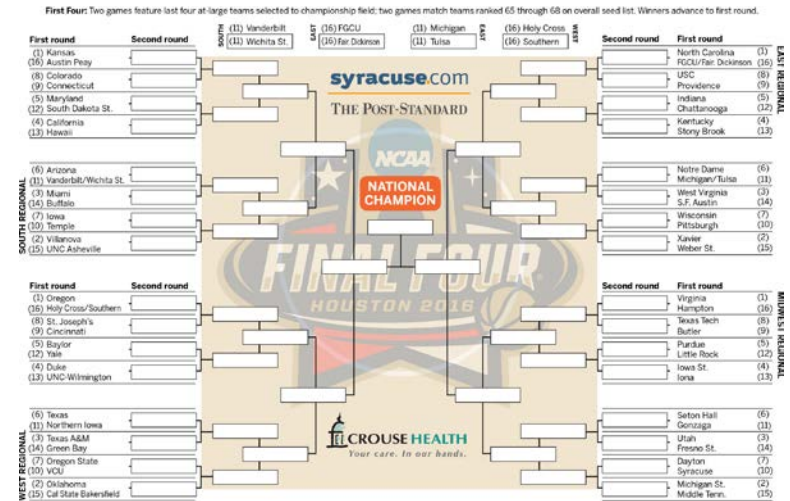
139.70 164 pages (2014)

163.98 164 pages (2014)

155.24

134.43 164 pages (2014)

149.95





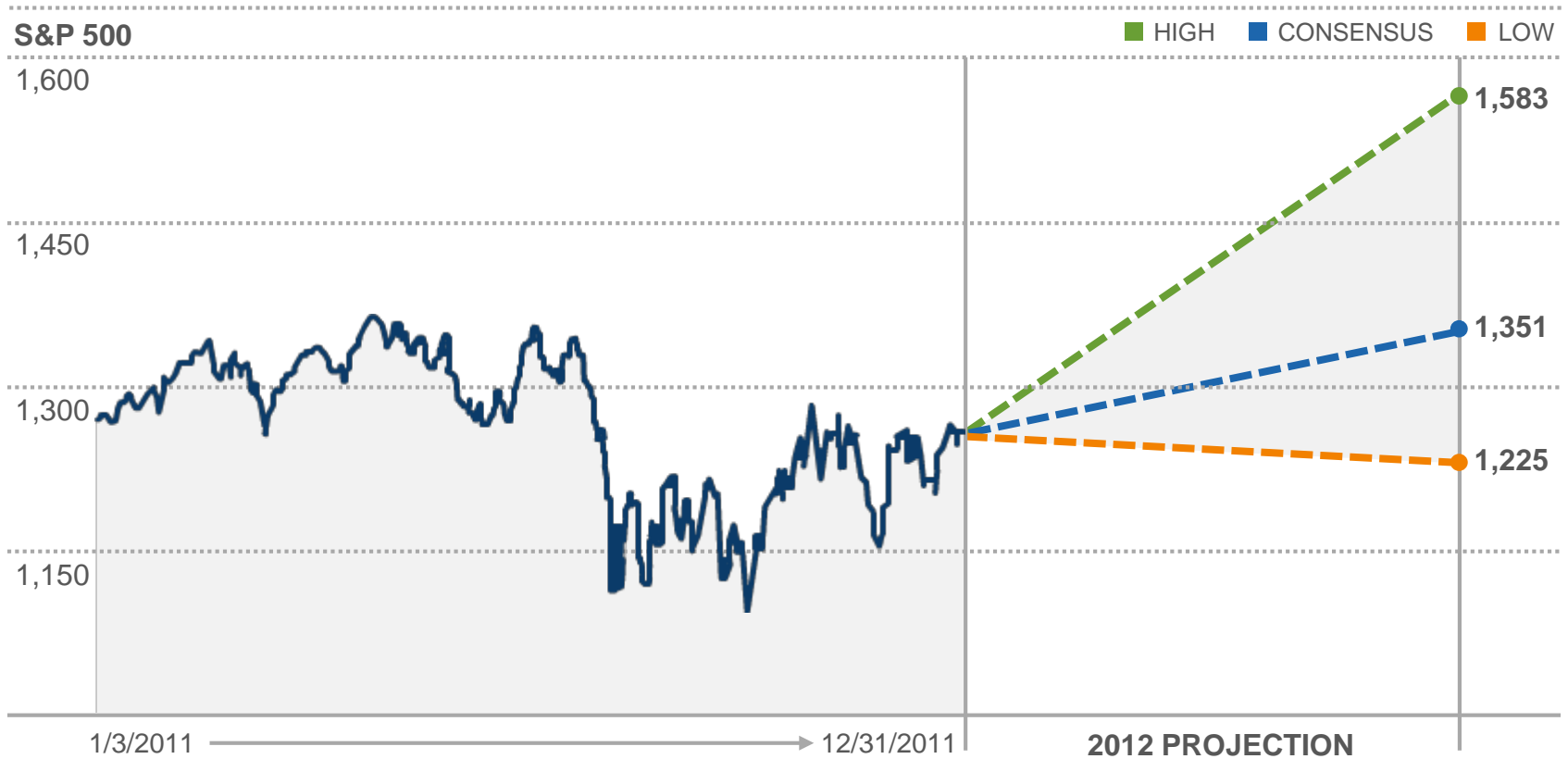


Walk

School of Information Studies  
Syracuse University

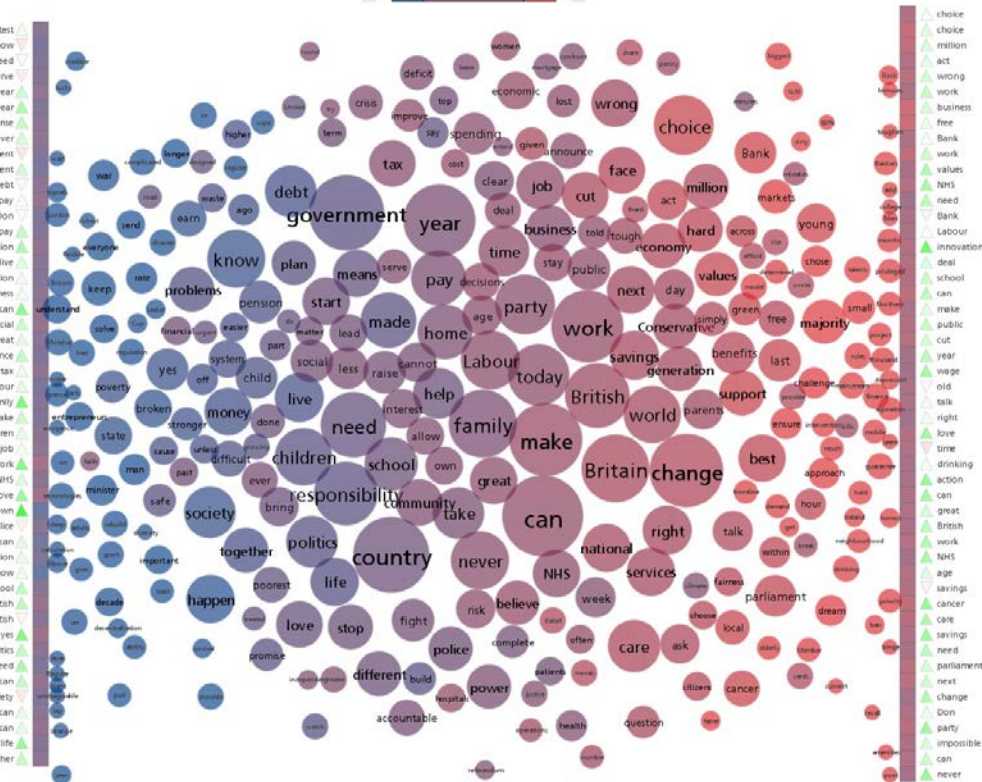


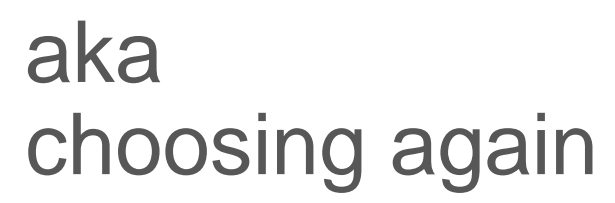
# Forecasting



SOURCE: REALTICK, CNNMONEY SURVEY

1







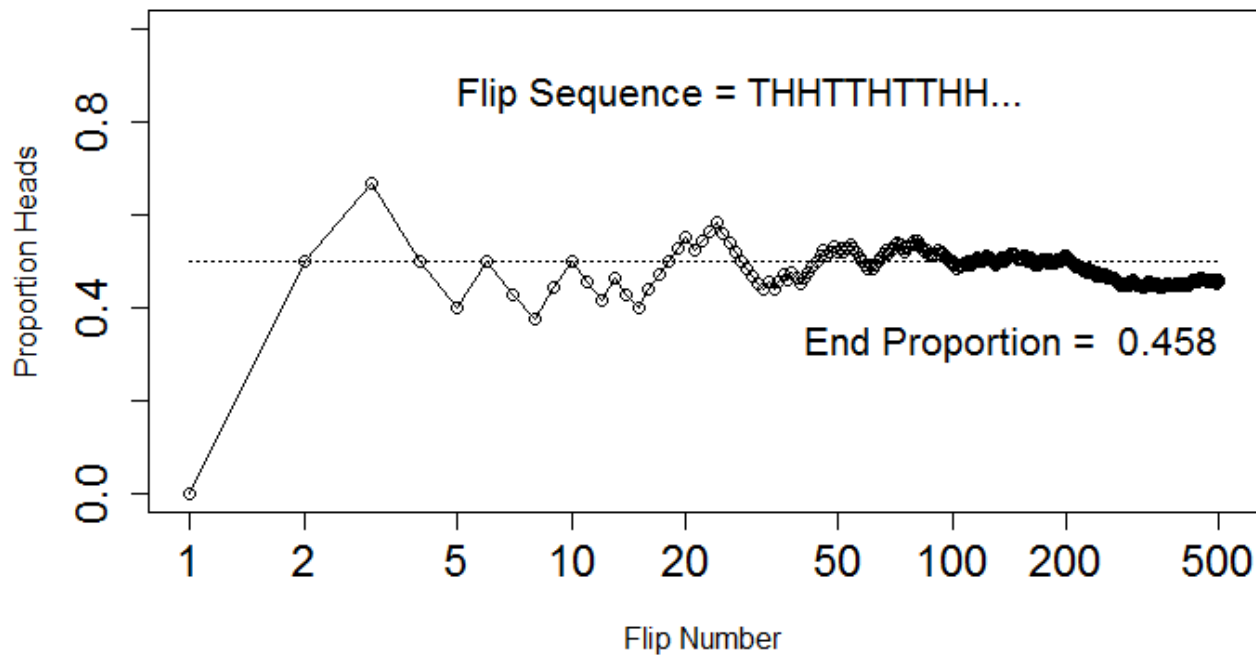


Run

School of Information Studies  
Syracuse University



# Choosing Again



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

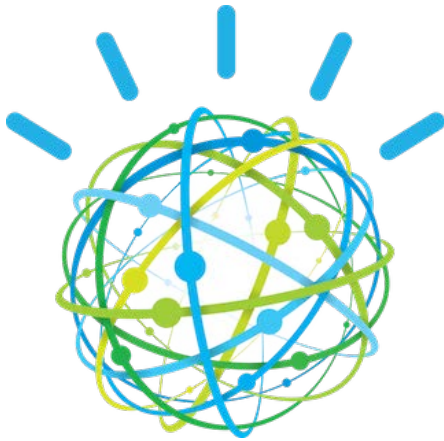
Flip Sequence = HTTTTTTTTTTTTTTTTTTT...

End Proportion = 0.928

Proportion Heads

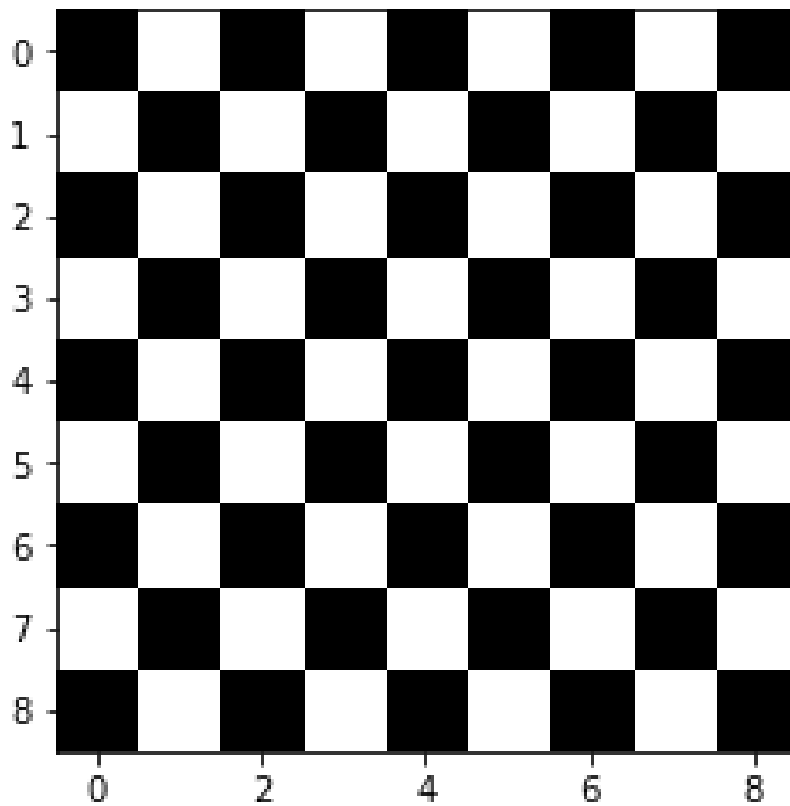
Flip Number

# Machine Learning

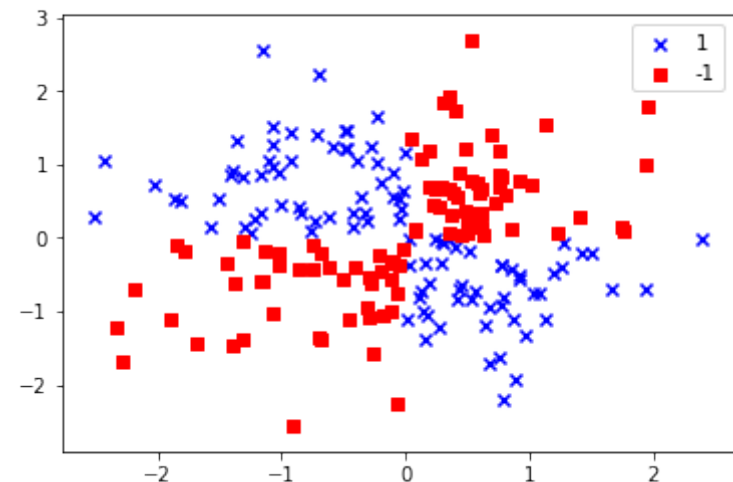


**H<sub>2</sub>O**

# Machine Learning

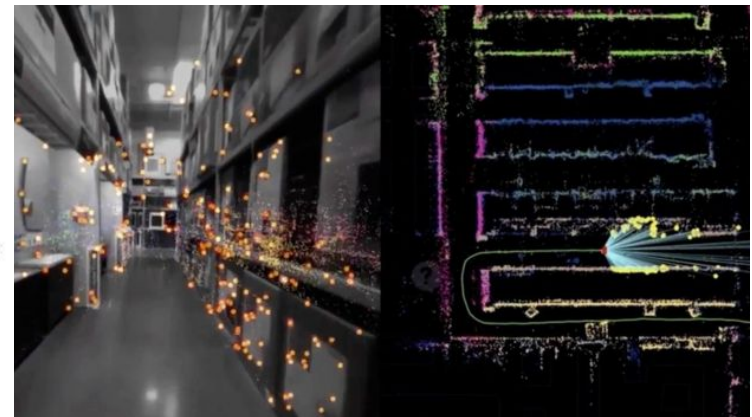
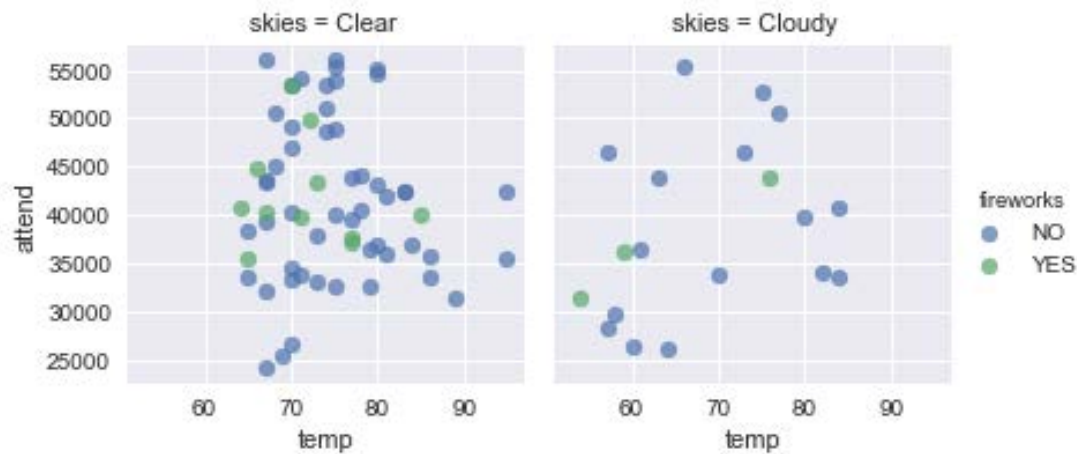
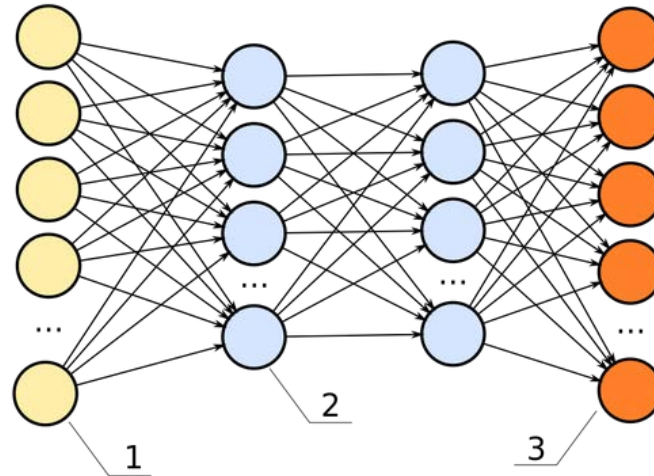


0 0 0 0 0 0 0 0 0 0 0 0 0 0  
1 1 1 1 1 1 1 1 1 1 1 1 1 1  
2 2 2 2 2 2 2 2 2 2 2 2 2 2  
3 3 3 3 3 3 3 3 3 3 3 3 3 3  
4 4 4 4 4 4 4 4 4 4 4 4 4 4  
5 5 5 5 5 5 5 5 5 5 5 5 5 5  
6 6 6 6 6 6 6 6 6 6 6 6 6 6  
7 7 7 7 7 7 7 7 7 7 7 7 7 7  
8 8 8 8 8 8 8 8 8 8 8 8 8 8  
9 9 9 9 9 9 9 9 9 9 9 9 9 9





# Machine Learning

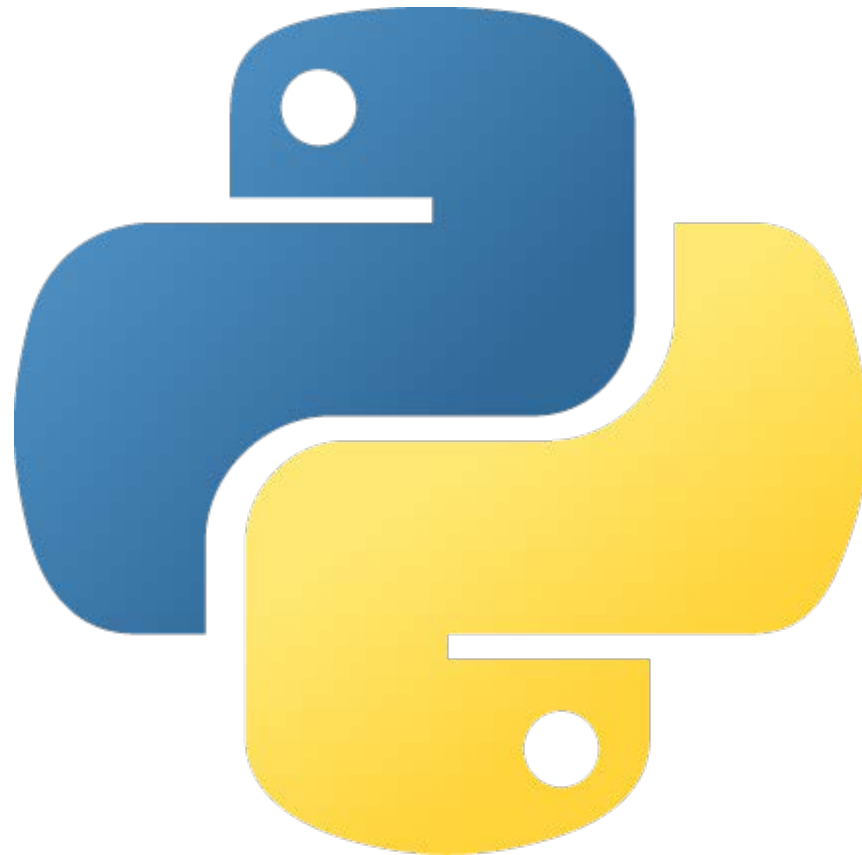




# Tools

School of Information Studies  
Syracuse University

# Python





# Anaconda



# Spark



# Others

