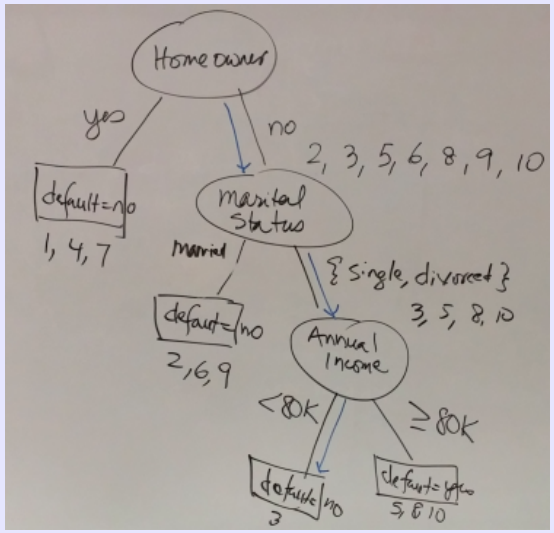
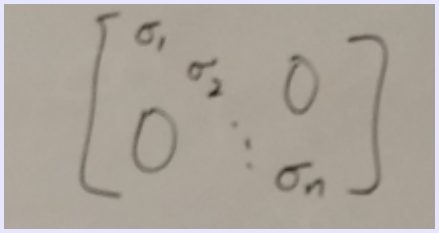
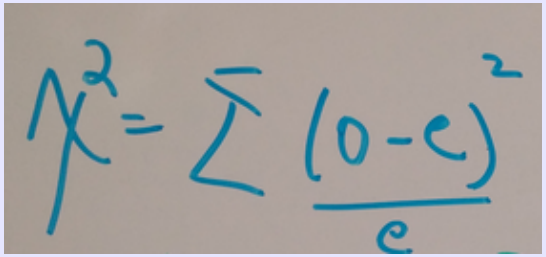
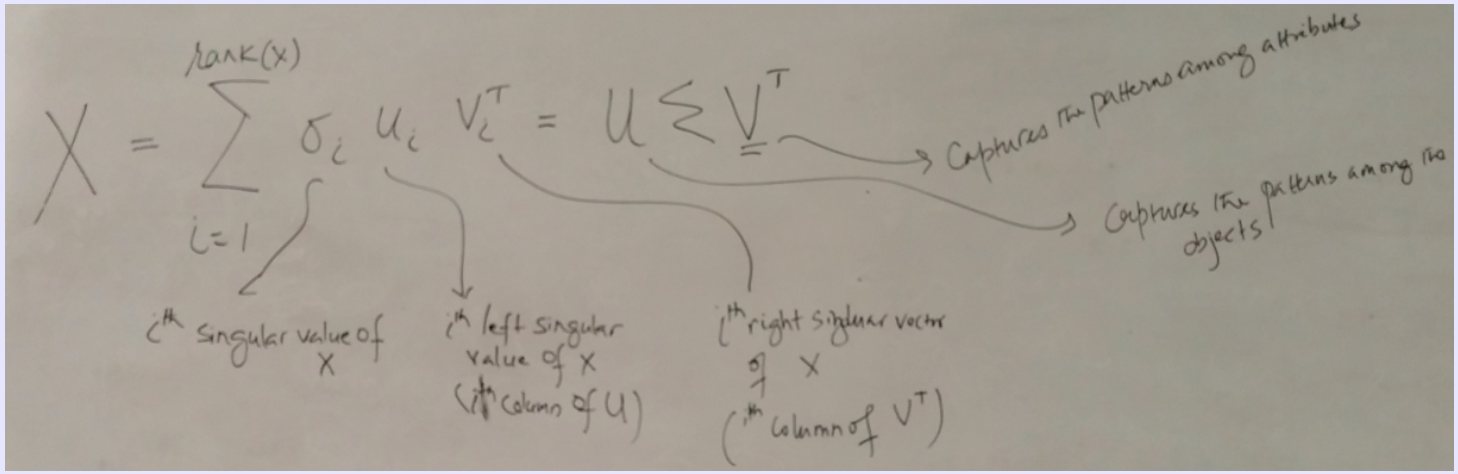
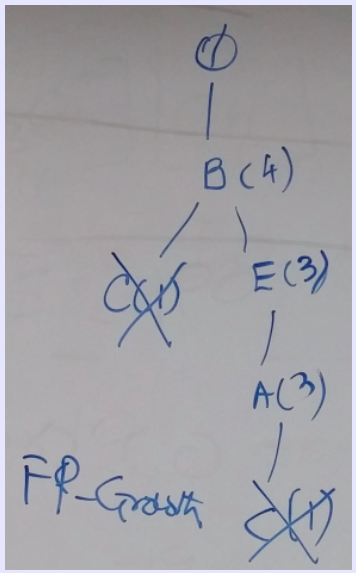
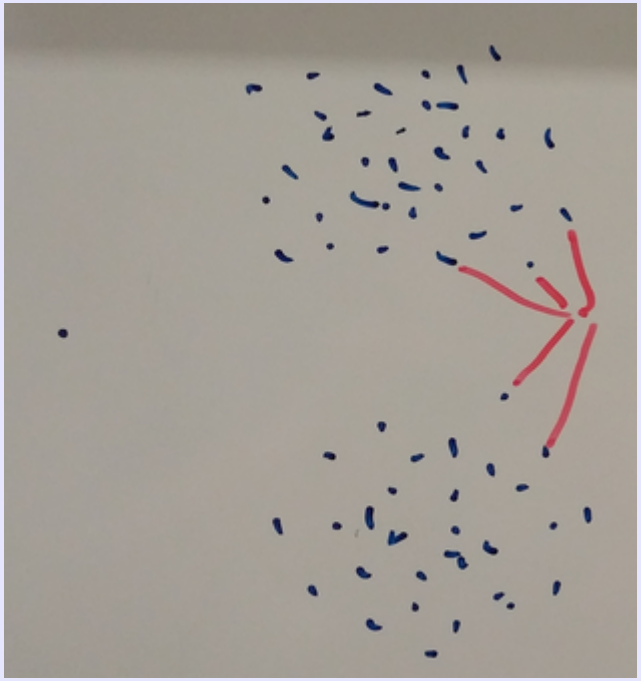












CS 422: Data Mining
Vijay K. Gurbani, Ph.D.,
Illinois Institute of Technology

Lecture 1: **Introduction**

Introduction: Placing data mining in context

- Data mining vs. machine learning vs. …

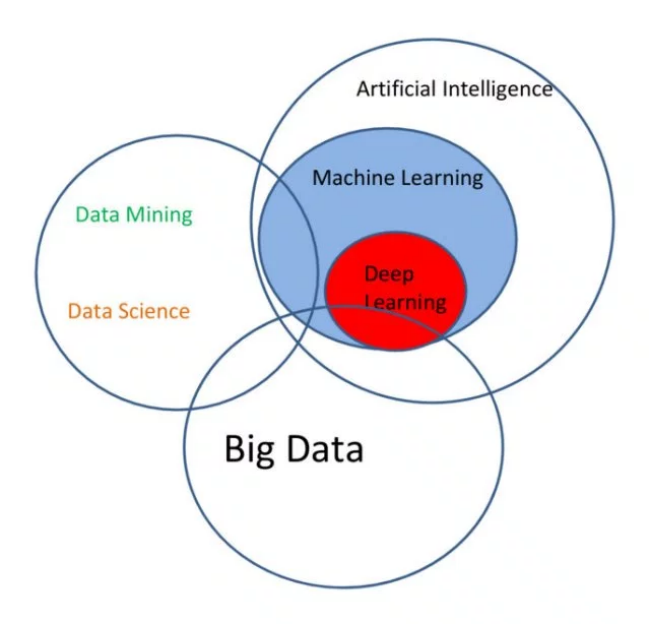


Image source: https://whatsthebigdata.com/2016/10/17/visually-linking-ai-machine-learning-deep-learning-big-data-and-data-science/

Introduction: Placing data mining in context

- Data mining vs. machine learning vs. …

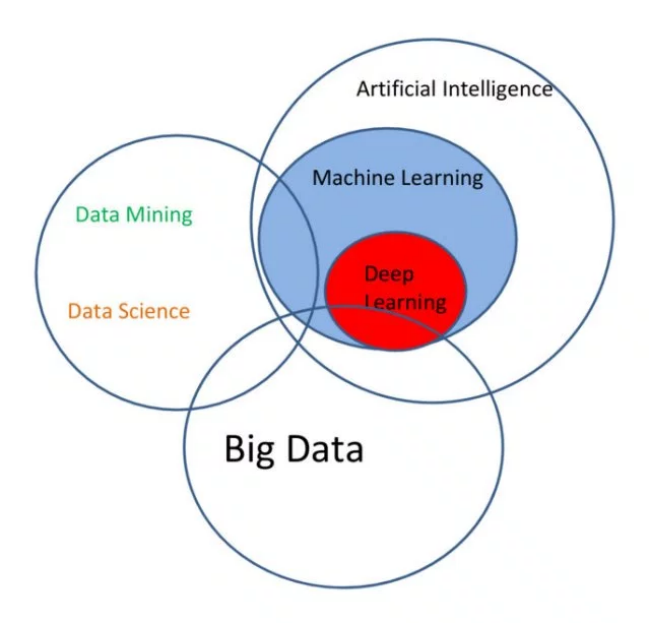


Image source: https://whatsthebigdata.com/2016/10/17/visually-linking-ai-machine-learning-deep-learning-big-data-and-data-science/

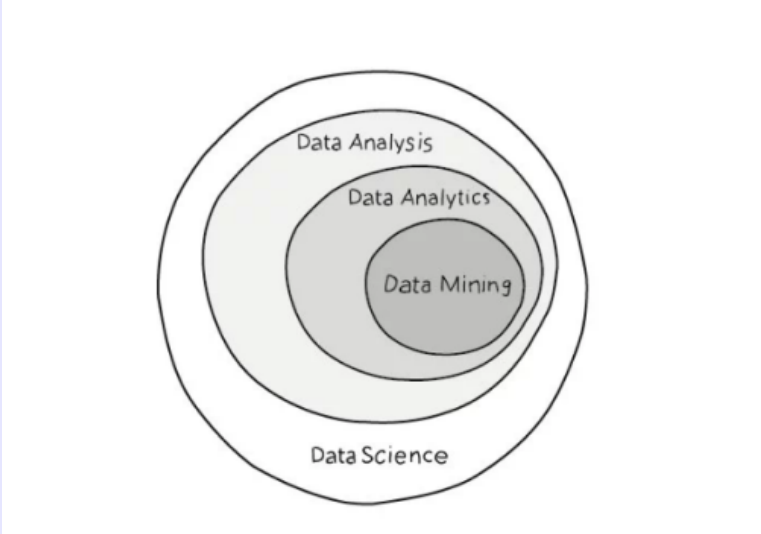


Image source: https://www.mo-data.com/what-is-the-difference-between-data-analytics-data-analysis-data-mining-data-science-machine-learning-big-data-and-predictive-analytics/

Introduction: Placing data mining in context

- Data mining vs. machine learning vs. …

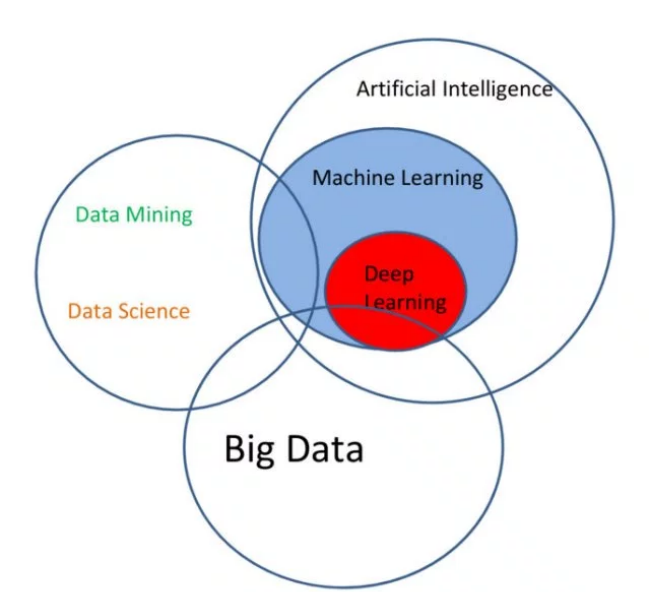


Image source: https://whatsthebigdata.com/2016/10/17/visually-linking-ai-machine-learning-deep-learning-big-data-and-data-science/

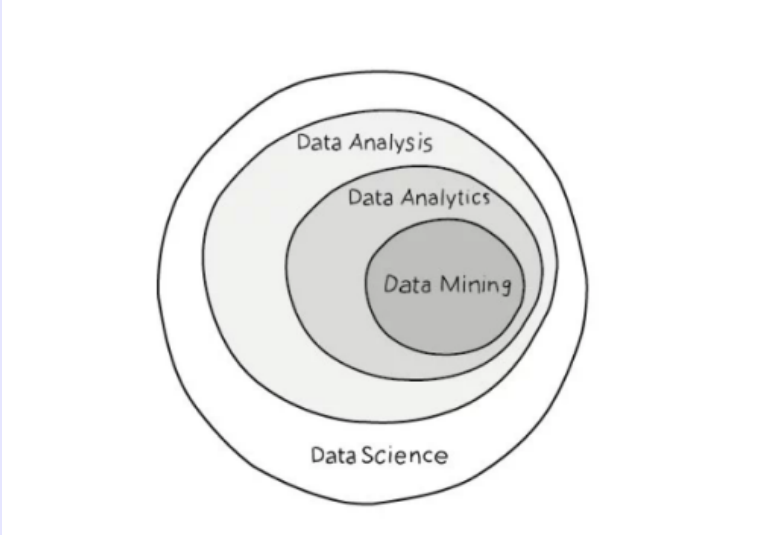
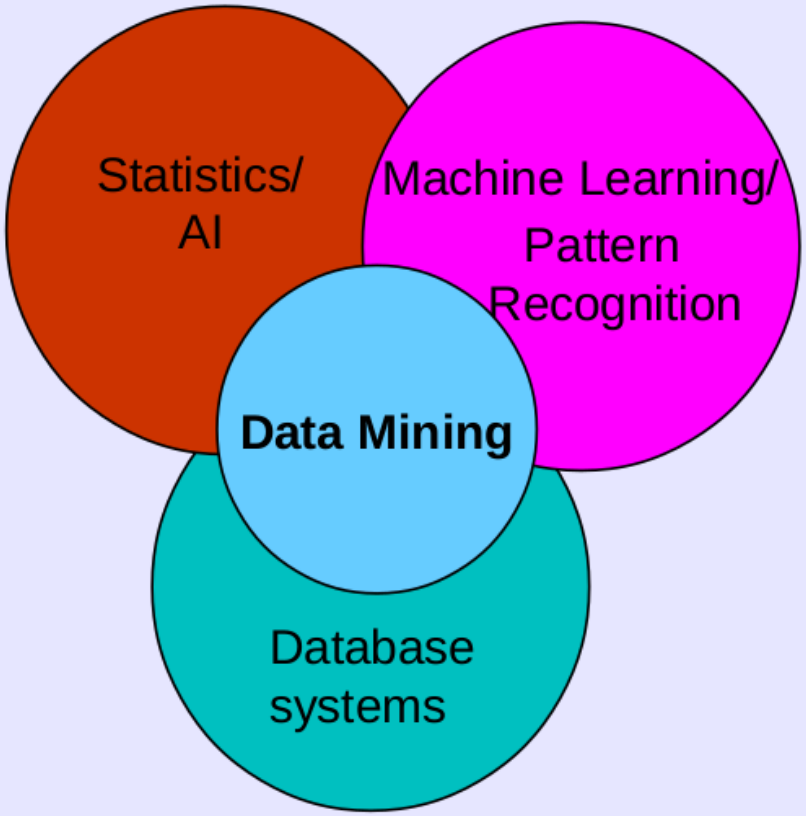


Image source: https://www.mo-data.com/what-is-the-difference-between-data-analytics-data-analysis-data-mining-data-science-machine-learning-big-data-and-predictive-analytics/

Insert your favourite picture here

Introduction: Placing data mining in context

- **Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems**
- **Traditional techniques may be unsuitable due to**
  - **Enormity of data**
  - **High dimensionality of data**
  - **Heterogeneous, distributed nature of data**



Introduction: AI

- The field of AI research was found at a workshop in Dartmouth College in 1956.
  - Allen Newell (CMU), Herbert Simon (CMU), John McCarthy (MIT), Marvin Minsky (MIT)
    - Alan Turing (not part of the workshop but had seminal impact on AI through the “Turing test”).
  - The field produced programs that:
    - Played checkers (and beating average humans circa 1959),
    - Solved word problems in Algebra,
    - Proved logic theorems,
    - Spoke English (Eliza).
- 1960’s: US DoD funds AI projects.
  - Herbert Simon: "machines will be capable, within twenty years, of doing any work a man can do".
  - Marvin Minsky: "within a generation ... the problem of creating 'artificial intelligence' will substantially be solved".

Introduction: AI

- By 1974, difficulty of the AI became apparent and progress slowed.
- AI entered an “AI Winter”, its first (1974-1980).
  - Lasts until early 1980’s, when symbolic computing becomes dominant in the form of **expert systems**
    - Lisp
    - Prolog.
- These expert systems proved to be difficult to maintain, and the industry collapsed again leading to the **second AI Winter** (1987-1993)**.**

- LISP: A symbolic computational language using first-order logic.
- **Facts**:
  - mother(trude, sally).
  - father(tom, sally).
  - father(tom, erica).
- **Rules**:
  - sibling(X, Y) :- parent(Z, X), parent(Z, Y).
  - parent(X, Y) :- father(X,Y).
  - parent(X, Y) :- mother(X, Y).
- **Inferences**:
  - ?- sibling(sally, erica).
  - Yes
  - ?- sibling(sally, sally).
  - ___

Introduction: AI

- **Modern AI**
  - **1993-2011**
    - Intelligent agents: work from decision theory and economics moves into AI through Bayesian networks, hidden Markov models, optimization, information theory and stochastic modeling.
  - **2011-present**
    - Deep learning: AI makes strides in computer vision, speech recognition and language processing.
  - Where modern AI falls short.
    - Pattern matching, not common sense inferences.
    - Needs large datasets to understand what patterns to match.
      - Human brain is more agile, does not need large datasets.
    - “Catastrophic Forgetfulness”
    - Things easy for humans to do (vision, language) are hard for AI to get right.
    - Things easy for AI to do (pattern matching across large datasets) are hard for humans to get right.
  - Future
    - Artificial General Intelligence
  - Are we at the cusp of another AI winter?
    - https://www.bbc.com/news/technology-51064369

Introduction: AI

- **AI: Challenges in Artificial General Intelligence.**

Once upon a time, before the fashionable rise of machine learning and “big data,” A.I. researchers tried to understand how complex knowledge could be encoded and processed in computers. This project, known as knowledge engineering, aimed not to create programs that would detect statistical patterns in huge data sets but to formalize, in a system of rules, the fundamental elements of human understanding, so that those rules could be applied in computer programs. Rather than merely imitating the results of our thinking, machines would actually share some of our core cognitive abilities.

That job proved difficult and was never finished. But “difficult and unfinished” doesn’t mean misguided. A.I. researchers need to return to that project sooner rather than later, ideally enlisting the help of cognitive psychologists who study the question of how human cognition manages to be endlessly flexible.

“A.I. is harder than you think”, Gary Marcus and Ernest Davis, New York Times Online, May 18 2018 (https://www.nytimes.com/2018/05/18/opinion/artificial-intelligence-challenges.html)
”Rebooting AI: Building artificial intelligence we can trust,” Gary Marcus and Earnest Davis, Pantheon Books, 2019.

Introduction: AI

- **AI: Challenges in Artificial General Intelligence.**
  - For AI to be truly useful, it needs to have the notion of "common sense".
  - Many projects have aimed to teach common sense to AI:
    - Project NELL (Never Ending Language Learner): Project started in 2011 by Tom Mitchell (CMU).
      - **Aim**: Learn everyday knowledge by crawling (or "scraping") the web.
    - Project ConceptNet: Started in 1999 at MIT Media Lab.
      - **Aim**: "Crowdsource" common sense information; volunteers enter information through a website.

Introduction: AI

- **AI: Challenges in Artificial General Intelligence.**
  - Many projects have aimed to teach common sense to AI (see example next slide on Cyc facts and rules):
    - Project Cyc: Largest project to teach common sense to AI. Launched in 1984 by Douglas Lenat.
      - **Aim**: Assemble a comprehensive ontology and knowledge base on how the world works.
  - It is perhaps interesting to see that many of these "common sense" AI projects are reverting back to first-order symbolic logic.

Introduction: AI

```
(#$isa #$BillClinton #$UnitedStatesPresident)
```

"Bill Clinton belongs to the collection of U.S. presidents."

```
(#$genls #$Tree-ThePlant #$Plant)
```

"All trees are plants."

```
(#$capitalCity #$France #$Paris)
```

"Paris is the capital of France."

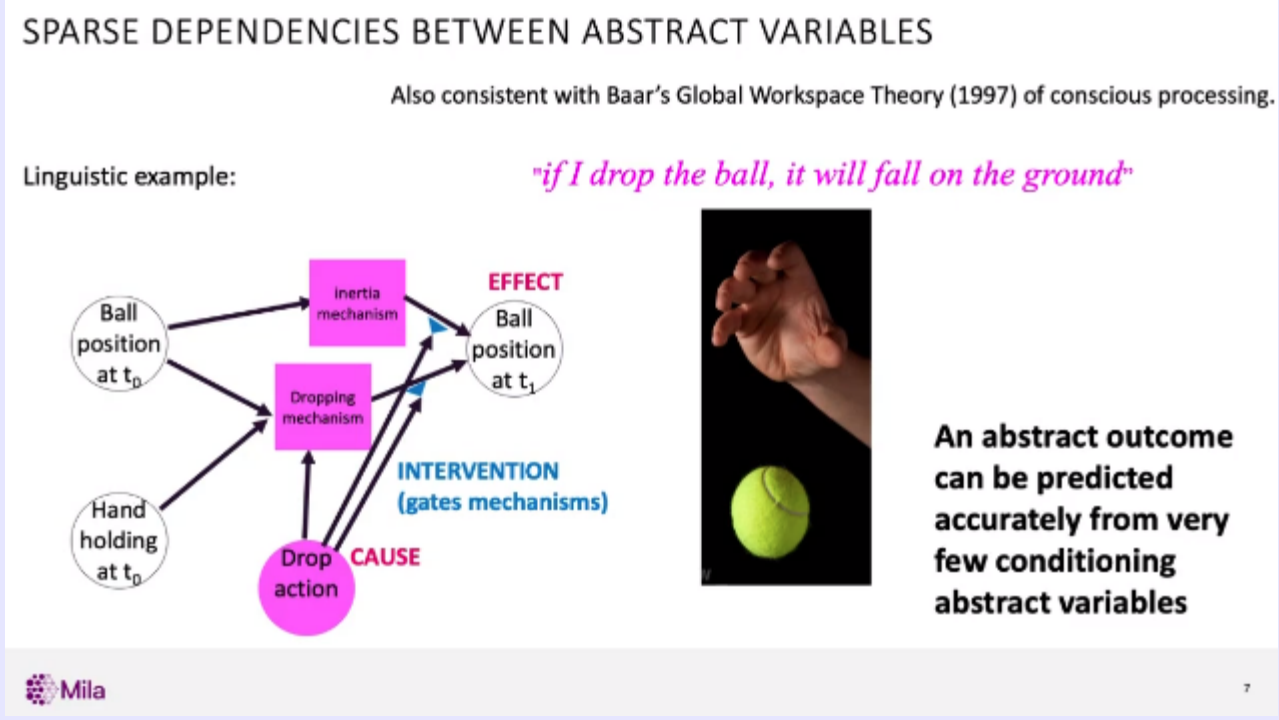
Sentences can also contain variables, strings starting with "?". These sentences are called "rules". One important rule asserted about the #$isa predicate reads:

```
(#$implies
   (#$and
     (#$isa ?OBJ ?SUBSET)
     (#$genls ?SUBSET ?SUPERSET))
   (#$isa ?OBJ ?SUPERSET))
```

Source: https://en.wikipedia.org/wiki/Cyc, visited Dec 22, 2019.

Introduction: AI

- **AI: Challenges in Artificial General Intelligence.**
  - Causality!



Slide courtesy Yoshua Bengio, NVidia GTC Conference, 2021.

Introduction: AI

- **AI: Challenges in Artificial General Intelligence.**
  - Causality!

https://comps.canstockphoto.com/woman-with-punctured-car-tyre-looking-picture_csp55707317.jpg

Resources

- R Tutorials
  - **https://ademos.people.uic.edu/index.html**
  - **http://www.cookbook-r.com/**
  - **https://github.com/matloff/fasteR**
  - http://www.r-tutor.com/r-introduction
  - https://cran.r-project.org/doc/manuals/R-intro.pdf
- R Programming Tutorial, Derek Banas https://www.youtube.com/watch?v=s3FozVfd7q4
- Will post recitation on R shortly.

- R Development Environment:
  - R https://www.r-project.org/
  - RStudio https://www.rstudio.com
  - RStudio Cloud https://rstudio.cloud/
  - R@Illinois Tech

    ~~Available on the Fusion cluster: https://linux1.cs.iit.edu:8787 Use your normal IIT login/password for authentication.)~~