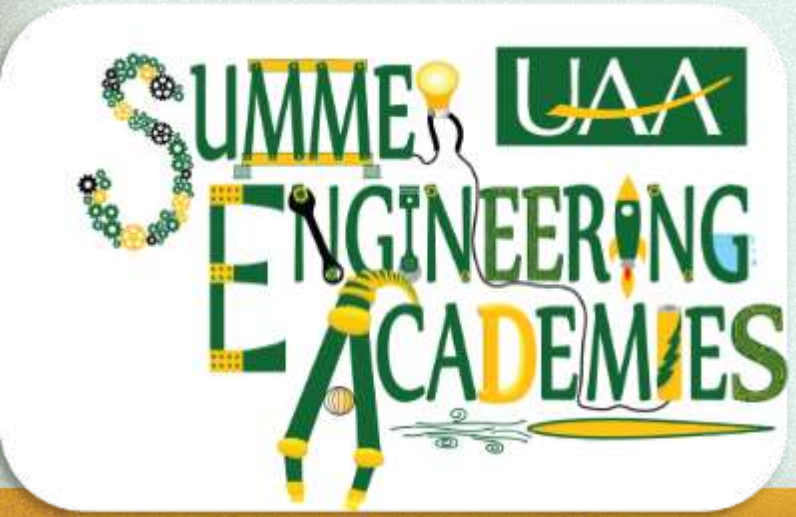
Text-Based Coding: Python

Pradeeban Kathiravelu, Ph.D.
Assistant Professor of Computer Science,
Department of Computer Science and Engineering,
University of Alaska Anchorage.

June 2nd – June 6th, 2025.



Day 1

Class Norms

Be smarter than your phone

– Put away your phone unless you are using it to assist you in learning.

- Questions, Questions, Questions.
- (Modified) Vegas Rule when it comes to peer discussions.
  – What your peers said here stays here and what is learned here leaves here.
- LOL. ☺
- Share the Airtime.
  – In peer discussions, let others in your group talk too.
- Reserve the right to change your mind.

[Adopted from the Safe Zone Project https://thesafezoneproject.com/]

Agenda

1. Introductions (to ourselves and the delivery of this program): Today
2. What is programming?: Today
3. Python Programming Language: Today
4. Learning to write Python programs: Today – Wednesday noon.
5. Make teams of 3 and writing your own applications in Python: Wednesday afternoon – Friday noon.
6. Present your works to the audience (incl. your parents/guardians): Friday afternoon.
   1. Presentation and a demo.

Timeline (Morning and Afternoon Sessions)

- Monday: Leave 12:15pm for lunch in the EIB Solarium, recess, return 1:30pm
- Tuesday: Leave 12:15pm for lunch in the EIB Solarium, recess, return 1:30pm
- Wednesday: Leave at 11:40am for PVT - show starts at noon, lunch in the Commons, return by 1:40 pm.
- Thursday: Leave 12:15pm for lunch in the EIB Solarium, recess, return 1:30pm
- Friday: Leave noon for lunch in the Student Union Den for lunch with parents, no recess, return 1:15pm. Parents and Guardians join the student presentations, 1:15 pm – 3:00 pm.

Introduction to the Instructor

- UAA Since August 2023.
  - CSCE A365 (Falls. 2023 onwards)
  - CSCE A465 (Springs. 2024 onwards)
  - CSCE A490 Distributed Computing (Falls. 2024 onwards)
  - CSCE A462 Data Mining (Springs. 2025 onwards).
  - CSCE A360 Database Systems (Falls. 2025 onwards).

- Postdoc at Emory University, Atlanta, GA (2019 Aug – 2023 Aug).
- PhD, ULisboa, Portugal and UcLouvain, Belgium (2014 Sep – 2019 Aug).
  - Erasmus Mundus Joint Doctorate in Distributed Computing.
- MSc at ULisboa, Portugal and KTH, Sweden (2012 Aug – 2014 Sep).
  - Erasmus Mundus European Master in Distributed Computing.
- BSc (CS&E), University of Moratuwa, Sri Lanka (2006 Jun – 2010 Sep).

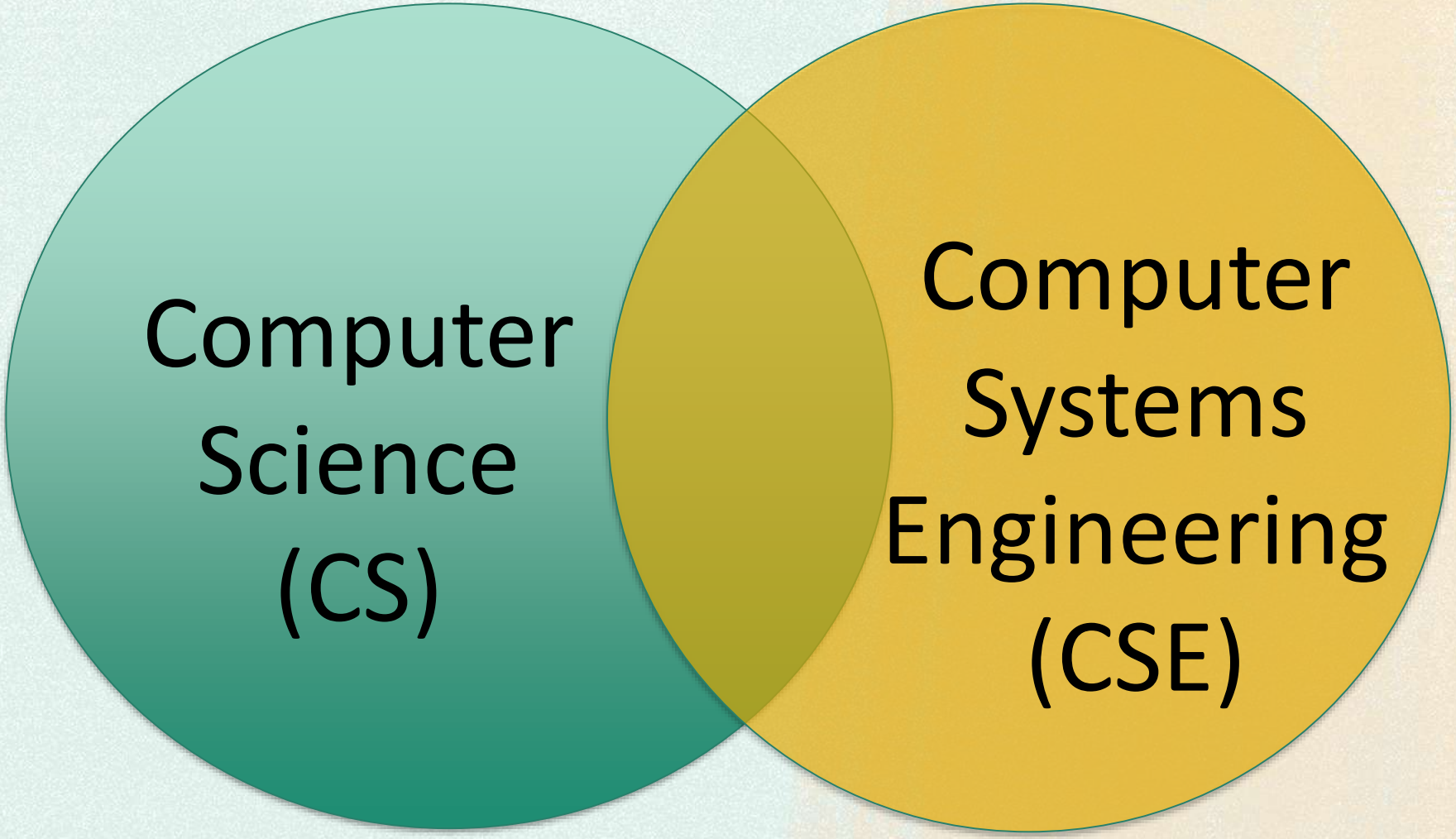
Introduction to the Team

Student Introductions

- My name is: _____
- My experience with the Python programming language is: _____
- My overall experience with coding is: _____
- In the future, I want to be: _____

Introduction to the Department of Computer Science and Engineering

Two Degree Programs In Computer Science and Engineering



In both programs, you must learn about computer programming, hardware concepts, algorithms, and network security.

Computer Science vs. Computer Systems Engineering

- Computer scientists work *primarily* with software while computer engineers work *primarily* with hardware
  - computer science is the study of computer systems and programming, while computer engineering is the application of electrical engineering principles to computer technologies
- A lot of crossover
  - employers in the computer technology industry expect workers to be comfortable with both software and hardware

CSCE Faculty At UAA

Kenrick Mock (Dean, CoEngg)

Masoumeh Heidari

Frank Witmer (Dept. Chair)

Sebastian Neumayer

Pradeeban Kathiravelu

Kamran Siddique

Computer Science and Engineering

- Robots, biometric security, autonomous cars, virtual reality and artificial intelligence are no longer just science fiction - they're part of our daily lives. Computers run all of these and more, ensuring a continuing high demand for Computer Systems Engineers and Computer Science professionals
- Burgeoning array of professional and technical careers in business, industry and government which require knowledge of computer hardware and software

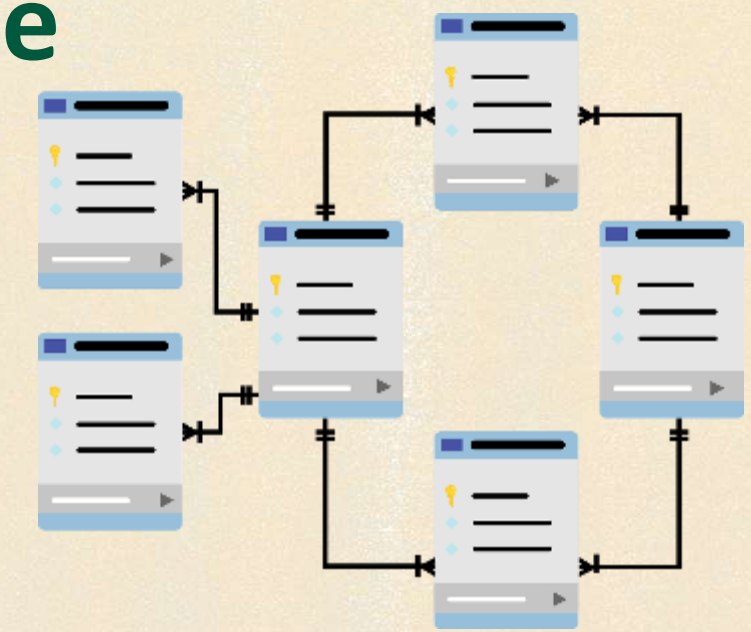
https://www.uaa.alaska.edu/academics/college-of-engineering/departments/computer-science-and-engineering/

Computer Science vs. Computer Systems Engineering

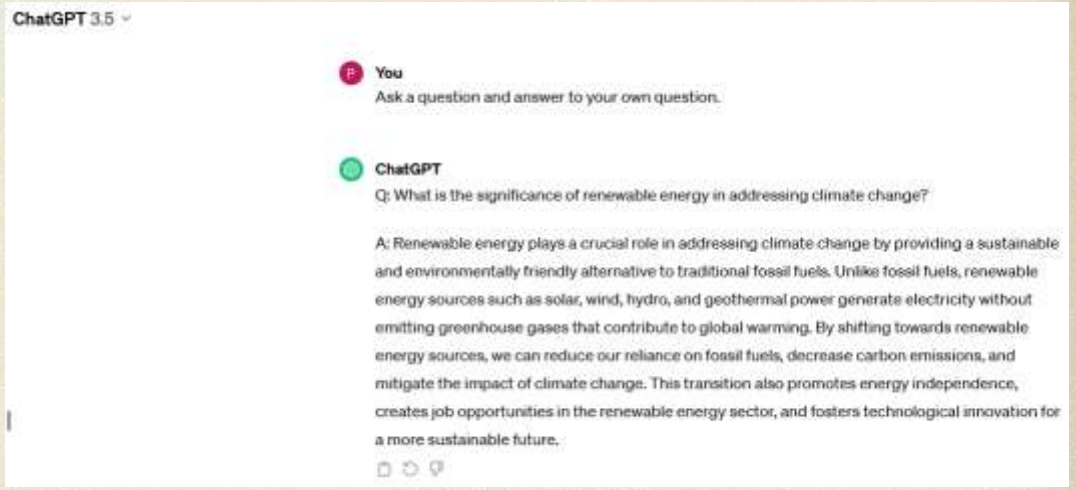
- Computer scientists sometimes work as part of a team and sometimes work alone
  - usually work in an office environment, though computer scientists who become web developers, computer programmers, and software developers may have the option to work from home
- Computer engineers typically work as part of a team in an office or a laboratory
  - generally do not have the option to work from home

Computer Science

- Software engineering
- Databases
- Data science

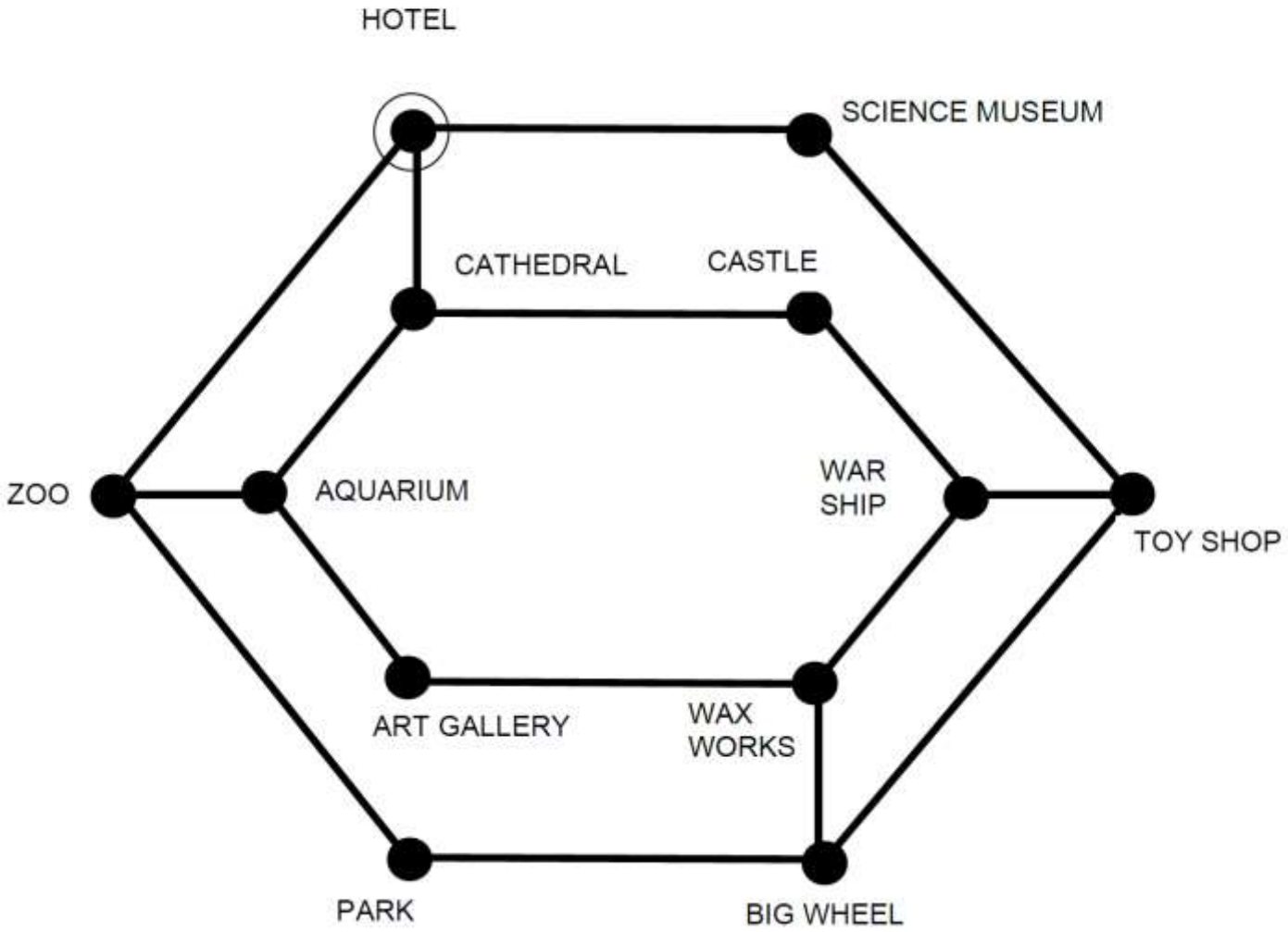


```
File  Edit  View  Search  Terminal  Help
#include <iostream>
#include <string>

void print_message(const char* arg_message, int arg_repeat)
{
    for(int l_repeat = 0; l_repeat <= arg_repeat; l_repeat++)
        write(1, arg_message, strlen(arg_message));
}

auto main(int argc, char* argv[]) -> decltype(0){
    print_message("CProgramming\n", 5);

    return 0;
}
```

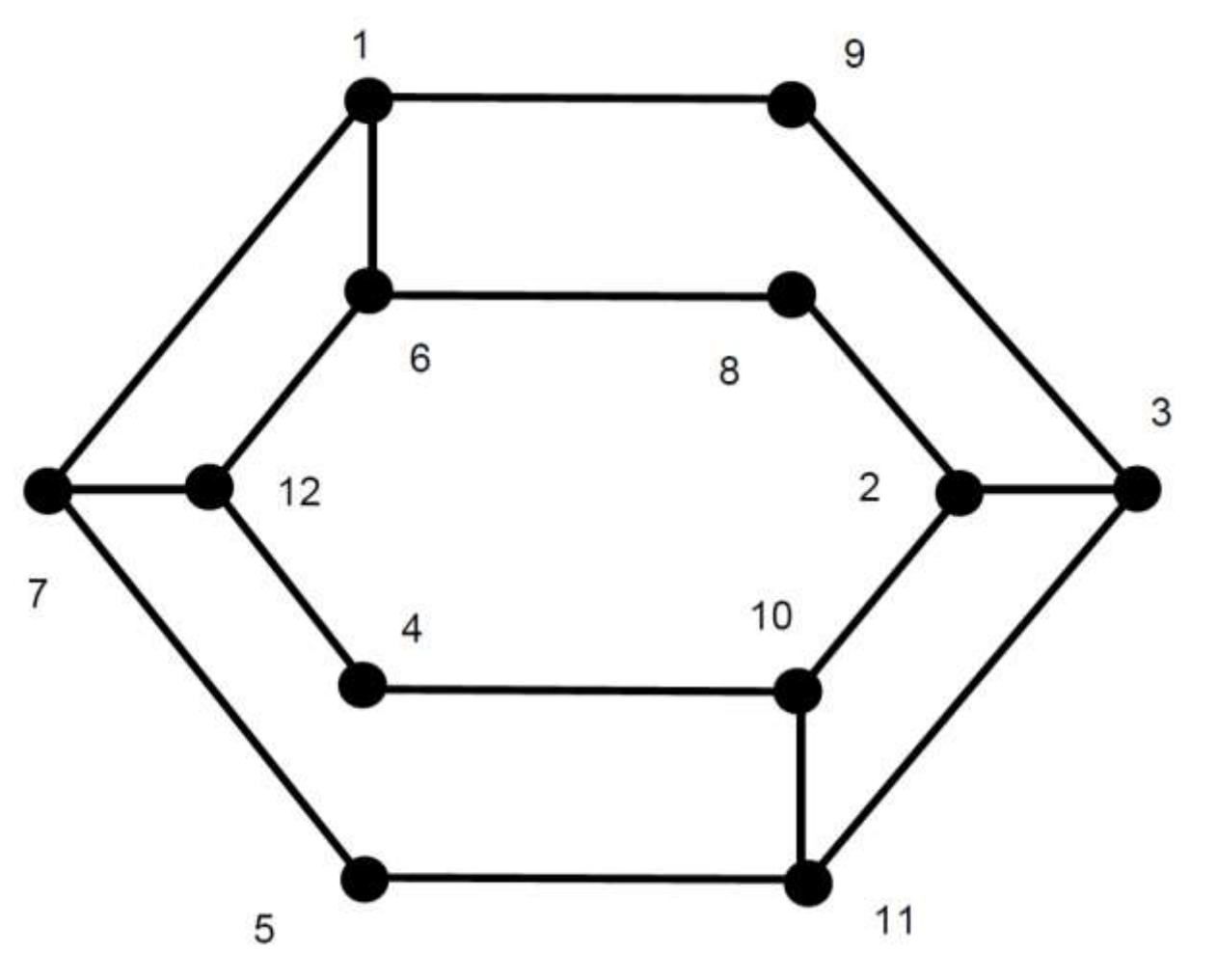


UAA 2025

Guided Tour
City Tube Map



City As A Graph



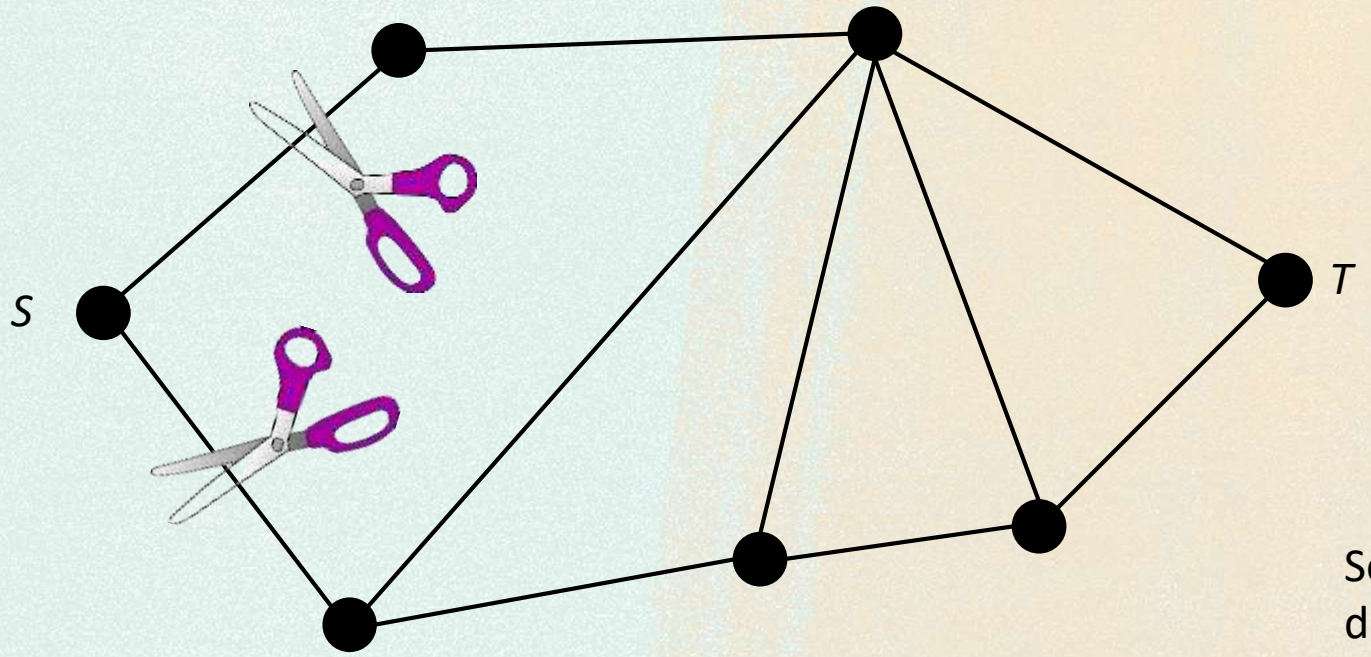
Exact locations and how far they each are apart doesn't matter

Graph, consisting of nodes and edges, is an **abstraction** of the real city.

What else can be modeled as a graph?

Minimum Cut

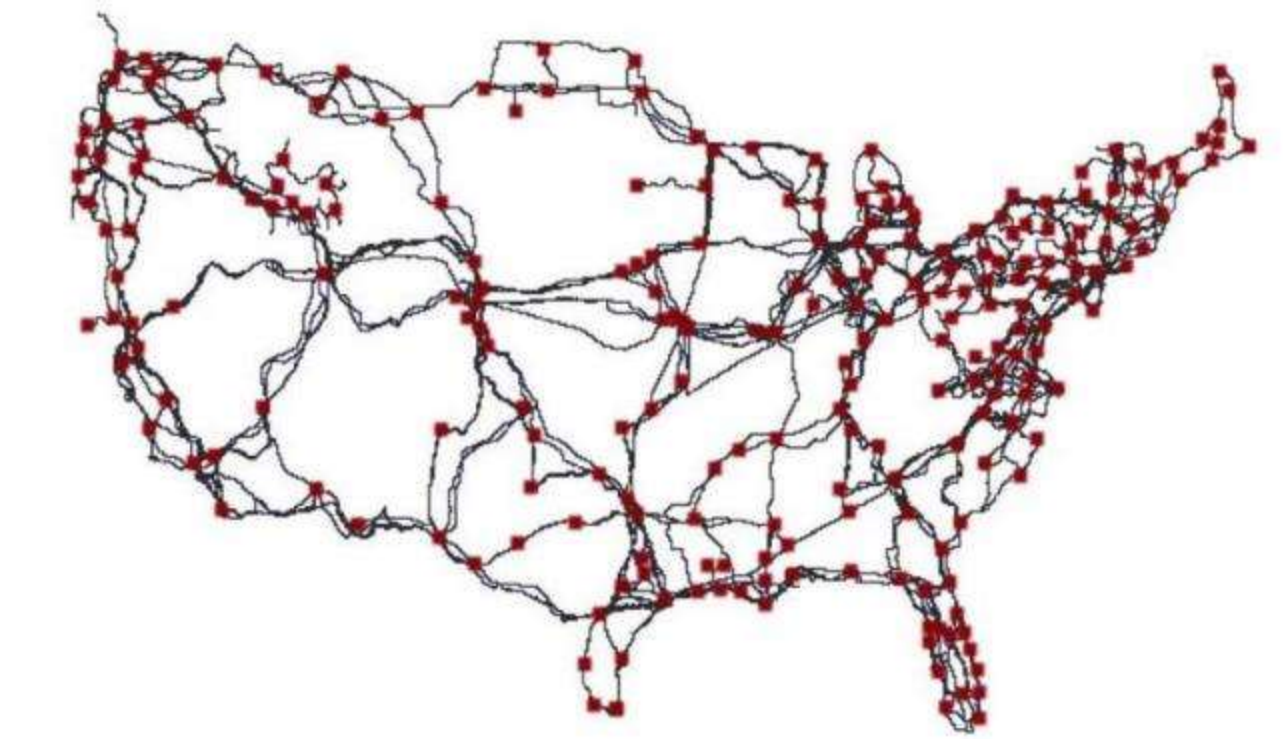
- What is the minimum number of link removals to disconnect two particular nodes on a graph?



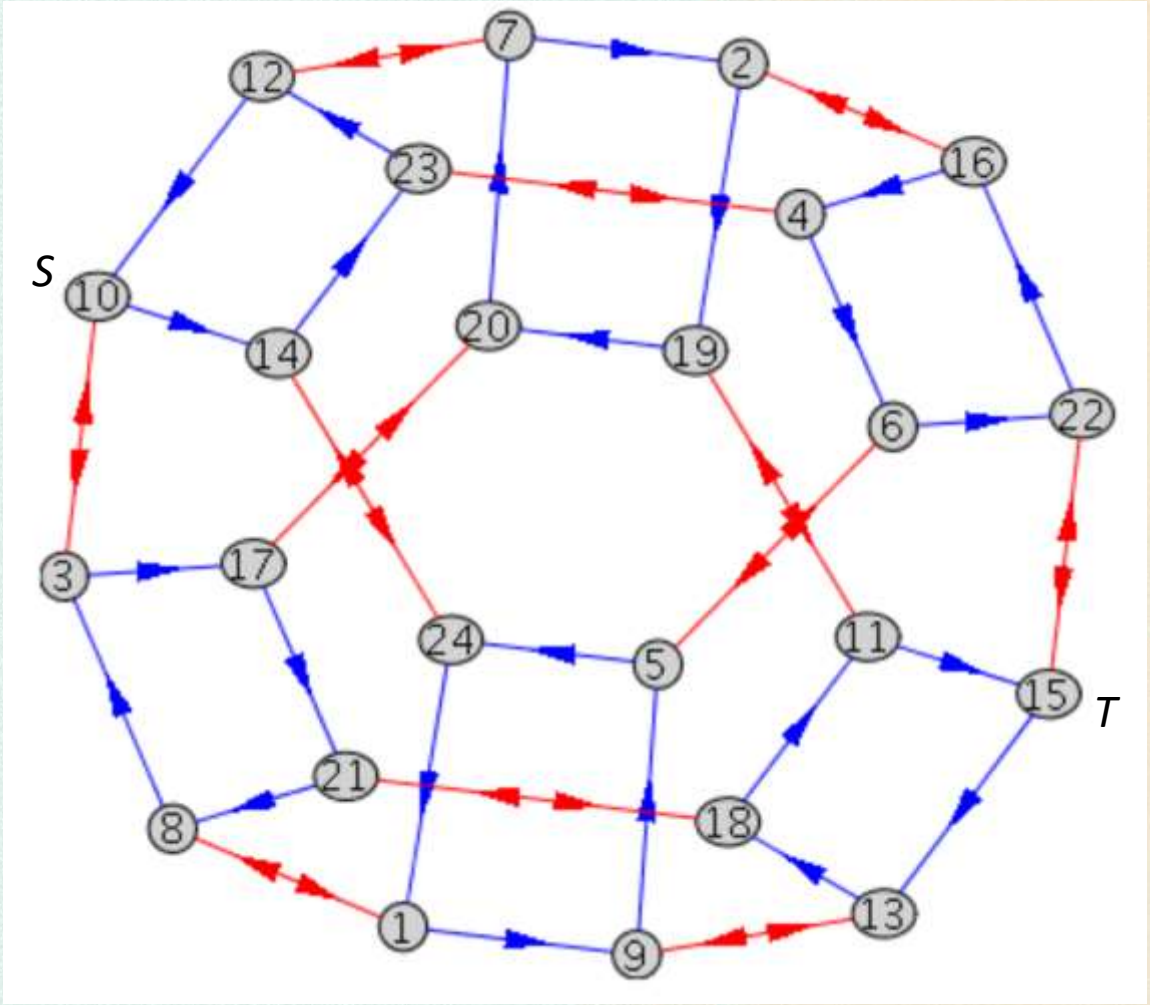
Set of links which disconnects S and T called a “cut set”

Motivation: Reliable Networks

- How many simultaneous link failures would disconnect Seattle from Washington D.C.?



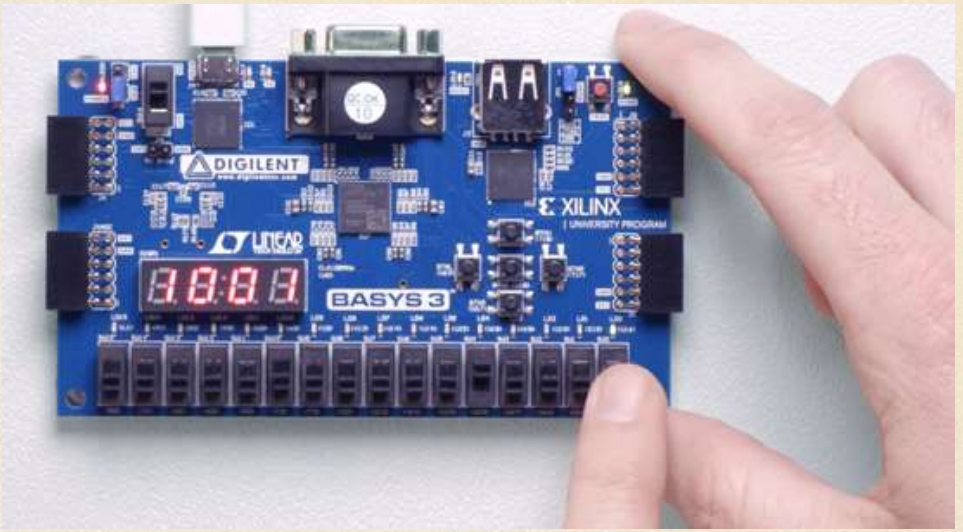
Algorithm For Finding Minimum Cut?

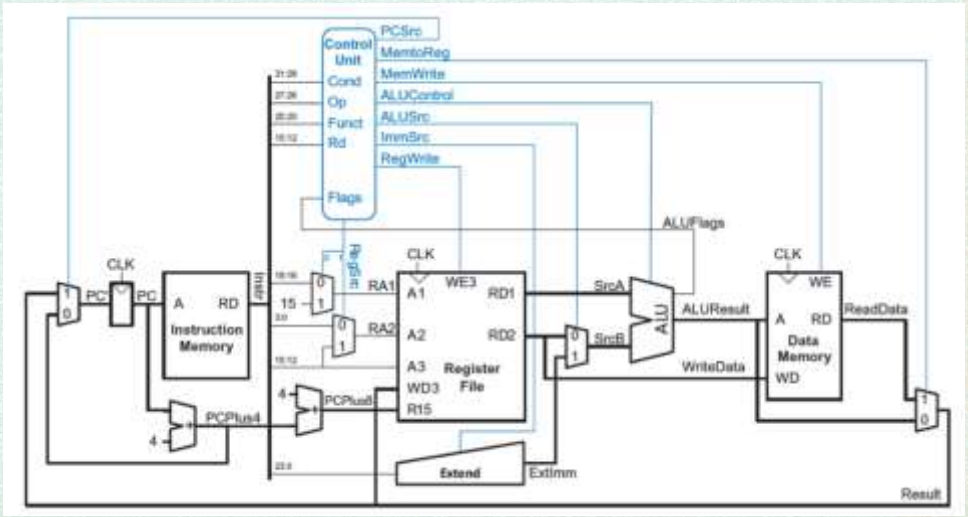


And how to code it?

Computer Systems Engineering

- Some crossover with Electrical Engineering
- Computer hardware and design
- Computer architecture
- Digital circuit design

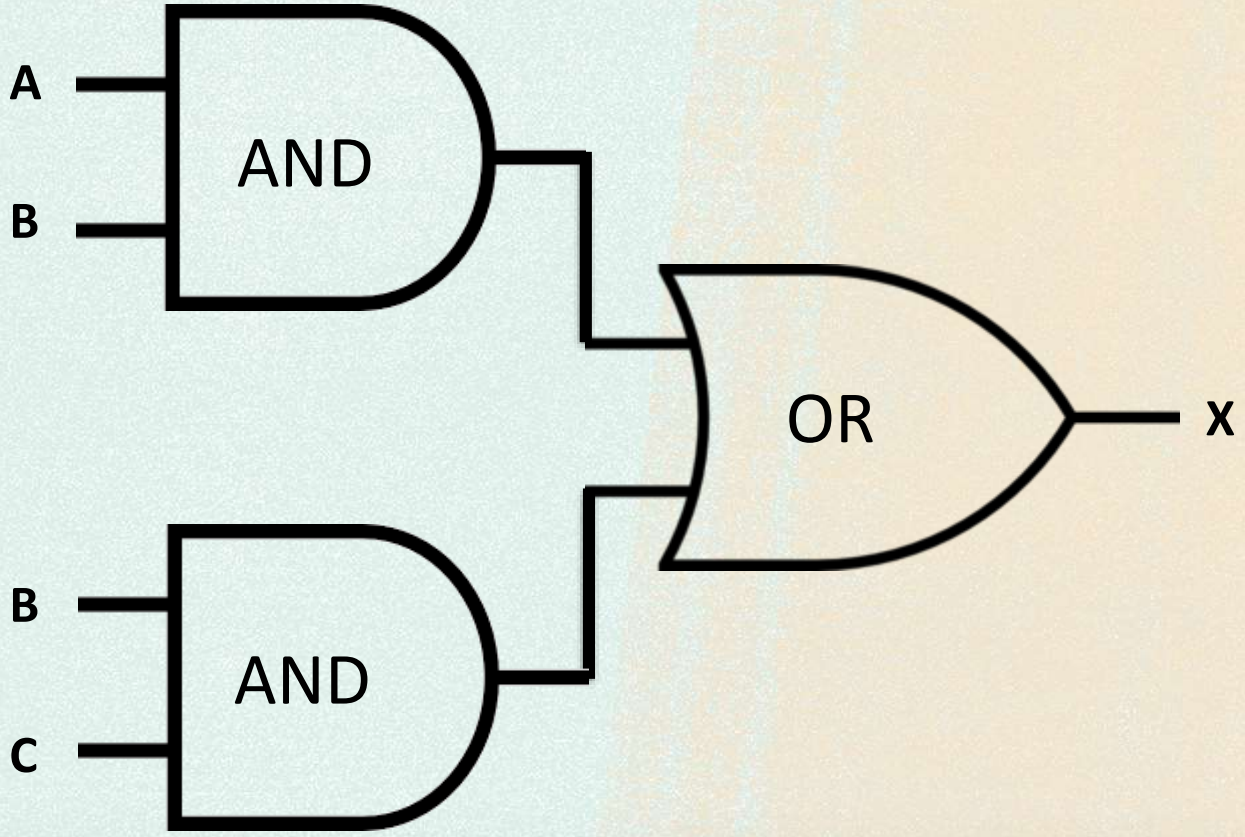




UAA 2025

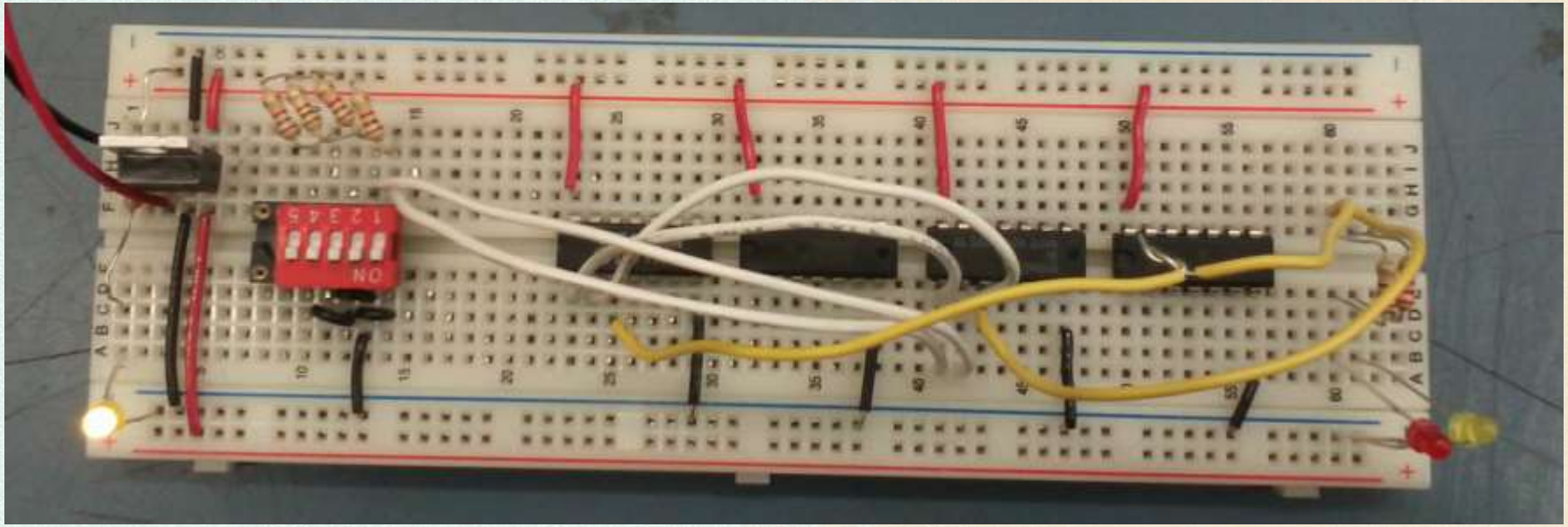
A Boolean Equation

- X = AB + BC



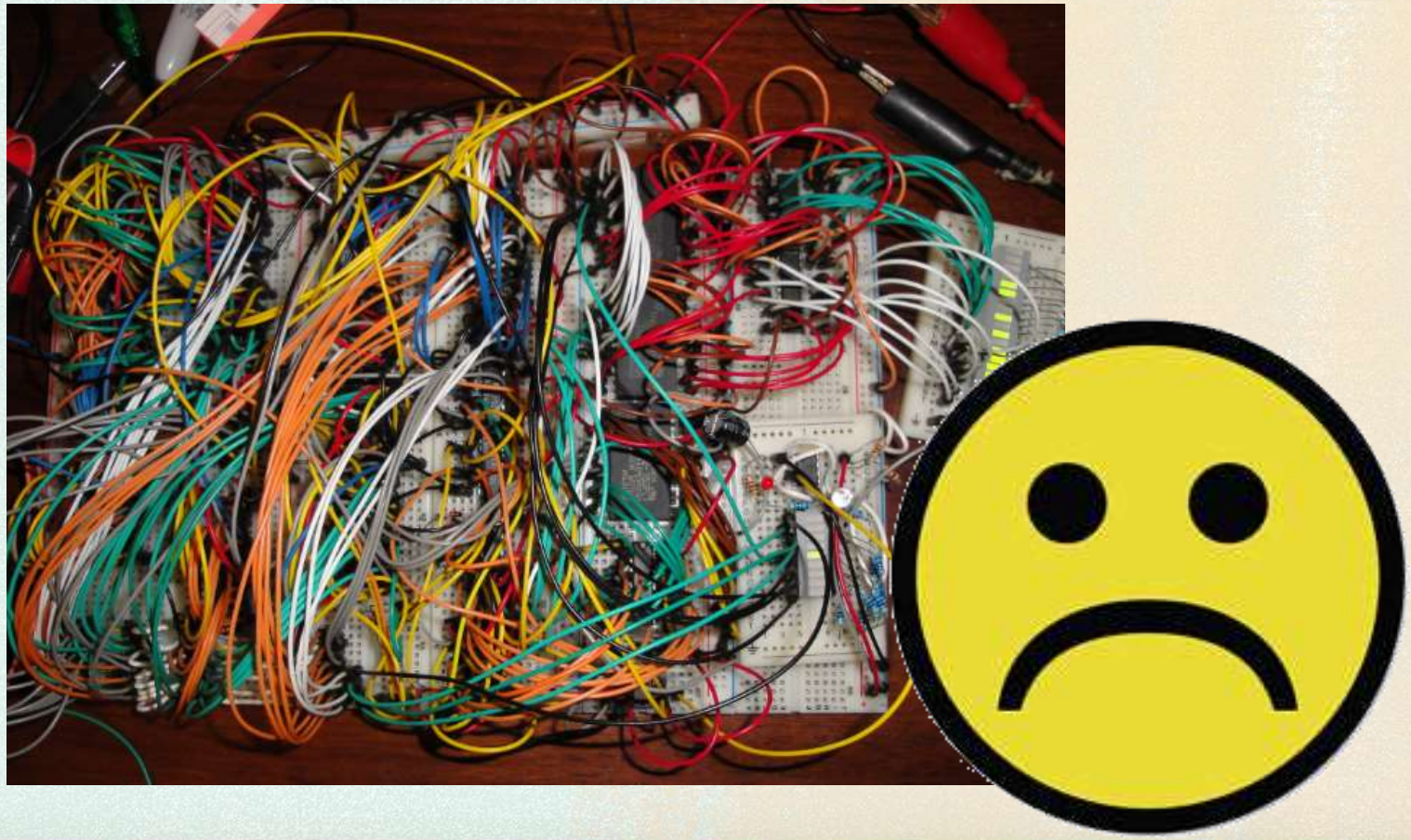
A possible implementation

- Breadboard, switches, resistors, AND chips, OR chips, LEDs, power source, lots of wires

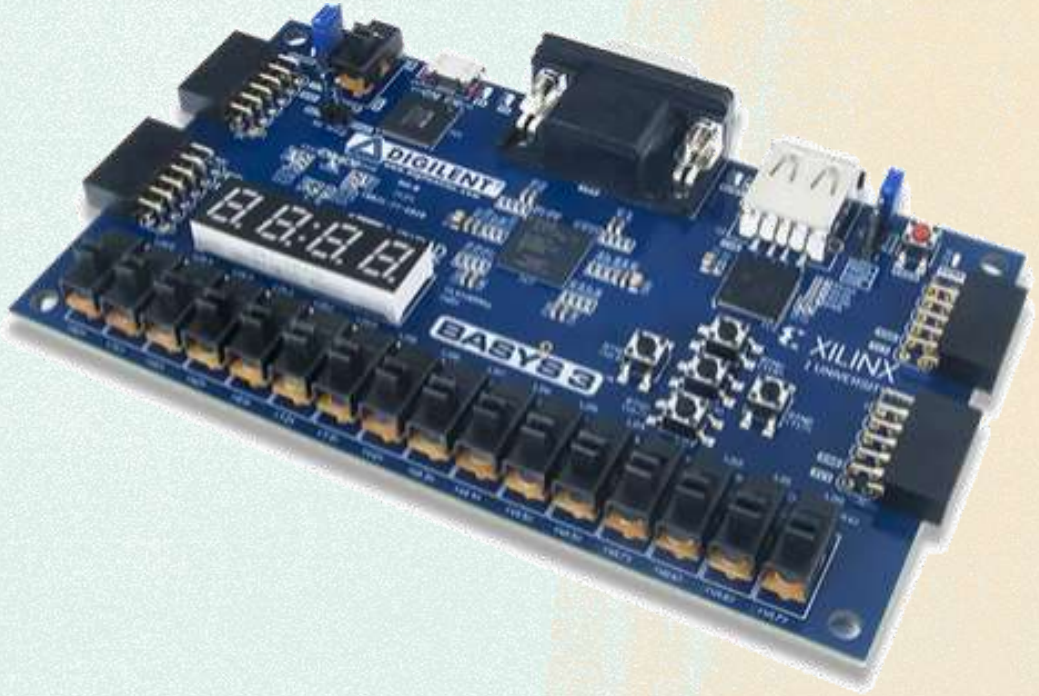


Changing designs may be problematic

- Hand built designs are error prone



Field Programmable Gate Arrays (FPGAs)



- "Programmable hardware"
  - Implementation done on the hardware level
  - Busywork with breadboard is eliminated!

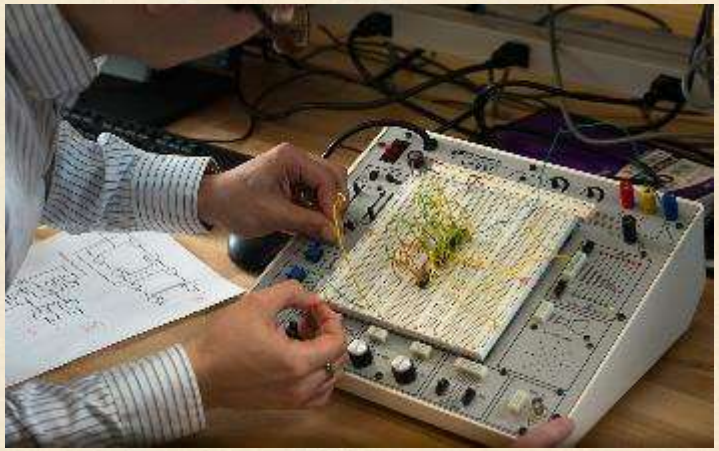
Career Options

Computer Science Degree

- Computer Programmer
- Database Administrator
- Software Developer
- Web Developer
- Artificial Intelligence (AI) and Machine Learning (ML) Expert

Either Degree

- Computer and Information System Managers
- Computer Systems Analyst
- Network and Computer System Administrator
- Information Security Analyst



Computer Systems Engineering Degree

- Computer Hardware Engineer
- Computer Network Architect
- Robotics Engineer

Common Job Duties

- Analyzing needs and developing software or hardware to meet those needs
- Ensuring the security of computer networks or databases
- Writing programs in various languages, like Java and Python
- Giving presentations to share information or persuade consumers
- Implementing new software or hardware into existing systems

CSCE At UAA

- Great club opportunities
  - Computer Science club
  - Robotics Club

UAA Robotics Club

Research Projects At UAA

- Students can work directly with CS&E faculty on funded research or apply for summer research opportunities with such organizations as the Air Force Research Laboratory or NASA. Recent examples include:
  - Analyze the reliability of Alaskan power and communication networks to disasters with a large area of impact, such as earthquakes (ConocoPhillips Arctic Science and Engineering Endowment Award)
  - Support salmon monitoring efforts by performing data analysis on historical daily salmon counts over several decades (U.S. Army Corps of Engineers)
  - Tick habitat research using geospatial data and agent-based modeling and simulation (Alaska INBRE)
  - Automatic coastline extraction from satellite imagery and erosion modeling (NOAA Alaska Sea Grant)
  - Great Lakes ice monitoring as part of the Arctic Domain Awareness Center (ADAC)
  - Intelligent compression and reconstruction of deep space images for NASA

Remote Open-Source Summer Programming Internships

- Google Summer of Code (GSoC)
- Outreachy
- Season of KDE (SoK)
- Google Season of Docs (GSoD) – for documentation.
- Hyperledger Mentorship Program
- The X.Org Endless Vacation of Code (EVoC)

- More FOSS?
- Annual?
- Stipend?
- Student?
- Country/Location?

Higher Education

- Research mobility – switch to application domains for your MSc/PhD.
  - Biomedical Informatics
  - Biomedical Engineering
  - Telecommunications.
- Explore countries and cultures.

How Much Do Computer Scientists and Engineers Make?

- The salary of a computer scientist or computer engineer depends on their career choices. The US Bureau of Labor Statistics says that $53,470 is the median wage for Computer Support Specialists—an entry-level job right out of college—but the median wage is $105,590 for Software Developers and $109,020 for Computer Network Architects.
- The median wage of all professions in the computer technology industry is $86,320, which is more than double the national average. According to the Alaska Department of Labor and Workforce Development, the median is just under $82,000 for computer professionals in Alaska.
- Overall, the job outlook over the next 10 years is very positive for computer scientists and engineers. This industry is expected to grow by 13%, which is almost double the national average. The need for Information Security Analysts and Software Developers, in particular, is expected to increase by almost 30%. However, because many companies are looking to outsource labor to cheaper markets, the number of jobs for Computer Programmers is expected to decline by 7%.

https://www.uaa.alaska.edu/academics/college-of-engineering/k-12-outreach/engineering-exploration/computer-science-and-systems-engineering.cshtml

What Do Our Grads Do?

- Emsi Labor Markets Analytics Reports that 75% of CS grads remain in Alaska to work, with 67% in the Anchorage area
  - local employers include: Alaska USA Federal Credit Union, GCI, Alaska Department of Fish and Game, and Resource Data Inc., Honeywell, Anchorage Water & Wastewater Utility (AWWU), Resource Data Inc., Alaska Railroad, Southcentral Foundation, State of Alaska: Fish and Game, Department of Transportation, Anchorage School District, PangoMedia
- Some graduates have gone on to CS graduate programs at Carnegie-Mellon University, the University of Southern California, UCLA, Clemson, University of Utah, UAF, the University of Colorado Boulder, and others

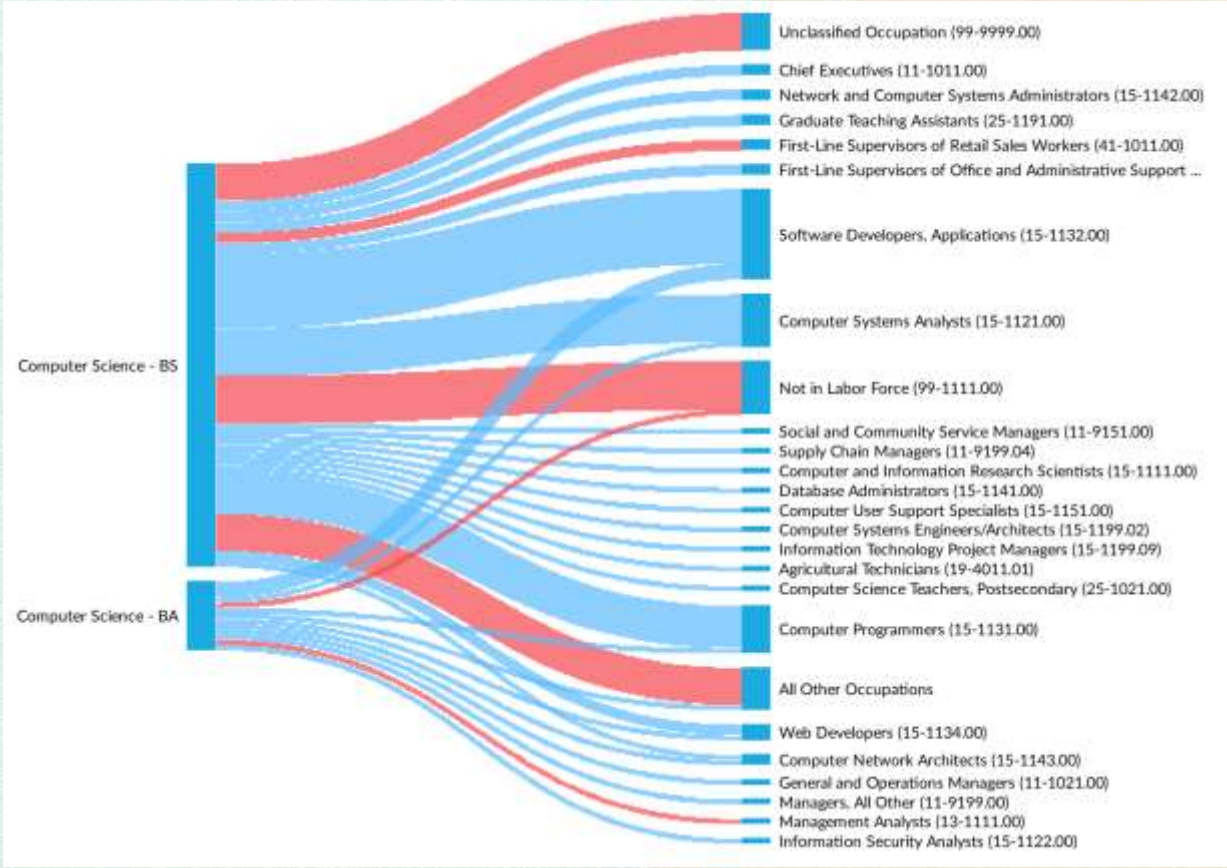
GCI

ALASKA DEPARTMENT OF FISH AND GAME

UAA 2025

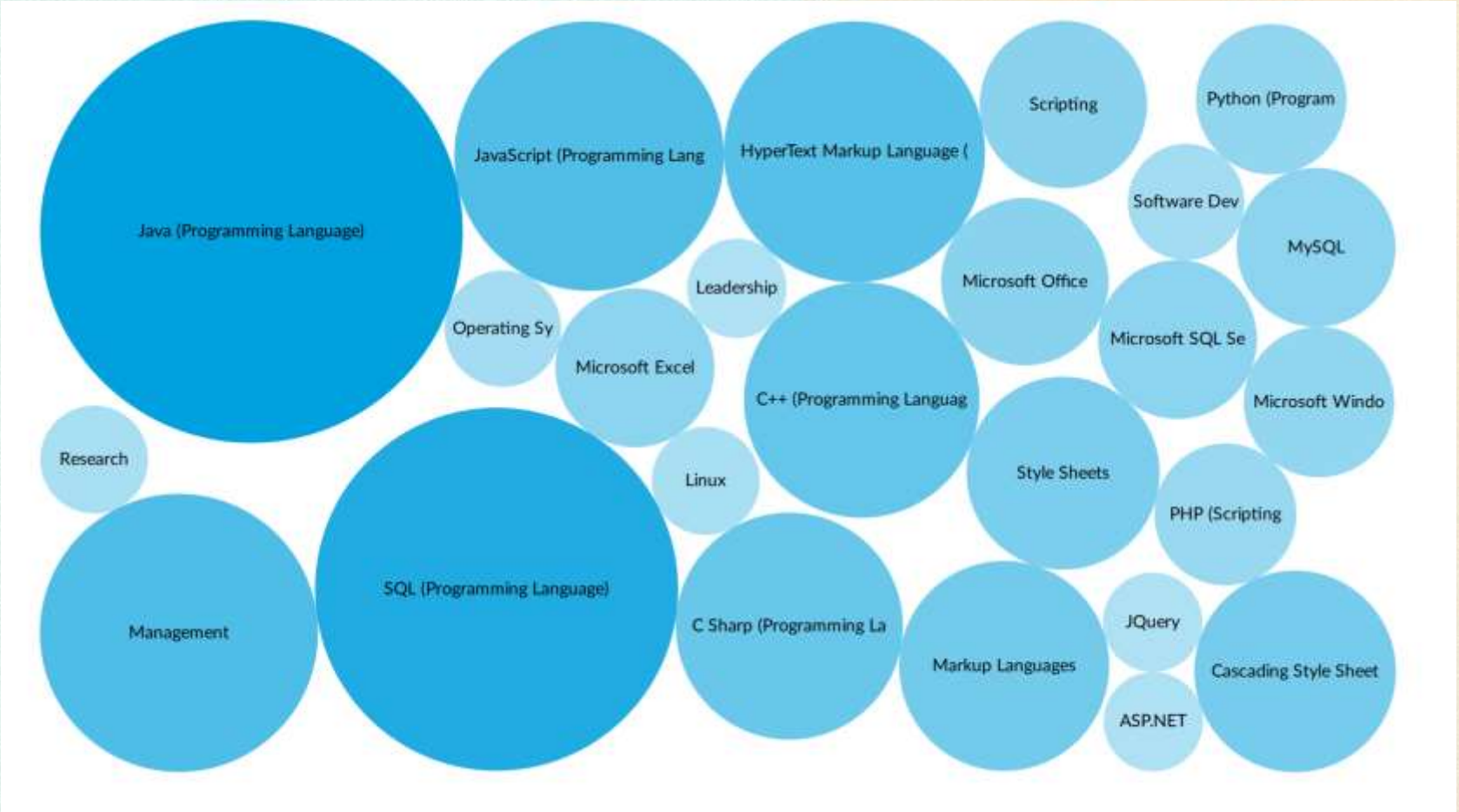
What Do Our Grads Do?

- From the Emsi Labor Market Analytics report on the Computer Science program (red indicates grads who are working outside the field of CS, and blue indicates grads working in the field of CS)



Skills Our Grads Use

- From the Emsi Labor Market Analytics report on the CS program



Programming!

This session won't be lecture-heavy

- Except for today.

- We start big today. Slow down and get into internals tomorrow.
- Focus on building projects Wednesday onwards.

Our goal is…

- To give you an overall idea of what it means to be a software engineer, and specifically a Python programmer.
- **Project-based learning.**
- **Team-based learning.**
- **Individualized.**

- https://github.com/KathiraveluLab/text-based-coding

What is programming?

What is Python?

What is Python?

- A modern high-level programming language.
- General-Purpose.

Why Python?

Why Python?

- Relatively easy to learn.
- Still… Potential to implement complex programs.
- Lots of existing libraries.
  - Use them as building blocks.
    - Easy to code complex applications.

Quick Question

- Will ChatGPT replace all the software engineers?

What does it mean to be a “software engineer?”

What can a Python program do for me?

- In other words, what will you present for your audience on Friday 1:15 pm – 3:00 pm as teams of three?

Coding in Python

Example Application:
Healthcare Data Mining in Python

Potential Domain: Data Mining

Explore United States of America

These popular destinations have a lot to offer

Las Vegas
566 properties

Myrtle Beach
2,642 properties

New York
1,381 properties

Orlando
3,683 properties

Panama City Beach
2,080 properties

Destin
2,341 properties

Deals for the weekend

Save on stays for January 17 - January 19

Genius
Aviator Hotel Anchorage
Anchorage, United States of America
8.4 Very Good · 296 reviews
2 nights ~~$428~~ $274

Genius
Breeze Inn Hotel
Seward, United States of America
7.3 Good · 833 reviews
2 nights ~~$338~~ $214

Genius
American Hotel Hostel
Seattle, United States of America
7.7 Good · 1,945 reviews
2 nights ~~$250~~ $233

Maria's Creekside B&B
Anchorage, United States of America
9.5 Exceptional · 170 reviews
2 nights ~~$250~~ $200

Trending destinations

Most popular choices for travelers from the United States

YouTube Mix

Mixes are playlists Youtube makes for you

Avicii - Wake Me Up (Official Video)
Avicii

Alan Walker - Alone
Alan Walker

Alan Walker - Faded
Alan Walker

Customers also bought

Based on products customers bought together

+

+

This item: Playskool Heroes Transformers Rescue Bots Academy Rescue Team Pack, 4...
4.7 ★★★★½ 5,796
-13% $38.99
Typical: ~~$44.90~~

Sponsored
Transformers Rescue Bots Griffin Rock Team Action Figures (Amazon Exclusive)
4.7 ★★★★½ 10,489
$38.99
✓prime

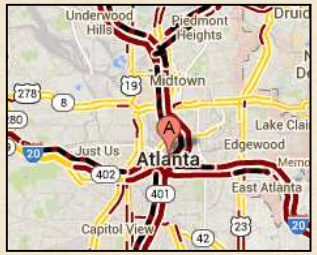
Sponsored
Spidey and his Amazing Friends Marvel Dino-Webs Crawler Toy Car Playset, Spider-Man & Gobl...
4.7 ★★★★½ 101
$34.99
✓prime

Large-scale Data is Everywhere!

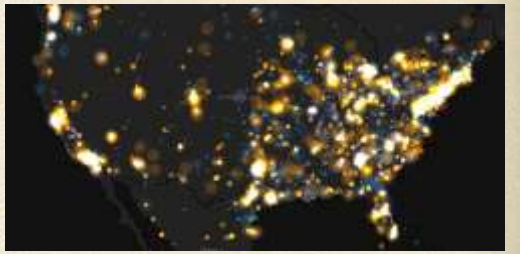
- Enormous data growth in both commercial and scientific databases
  - Advances in data generation and curation.
- Gather whatever data you can whenever and wherever possible.
  - Data is the new oil.
- Expectations
  - Gathered data will have value either for the purpose collected or for a purpose not envisioned.

Cyber Security

E-Commerce

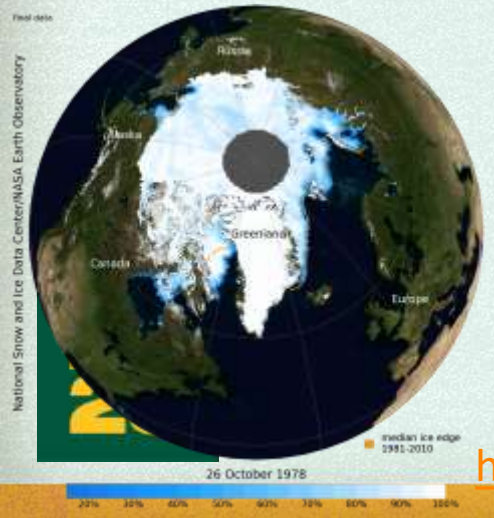


Traffic Patterns

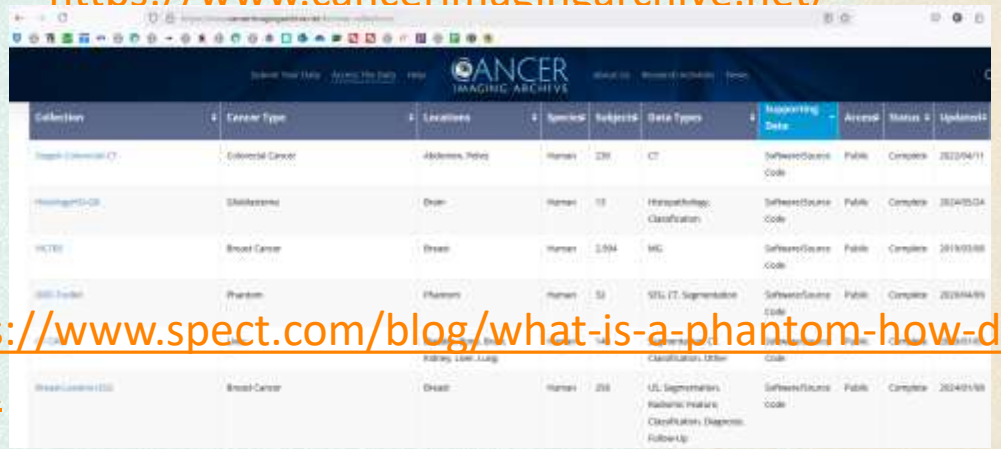


Social Networking: Twitter

Polar Data, Daily, since 1978

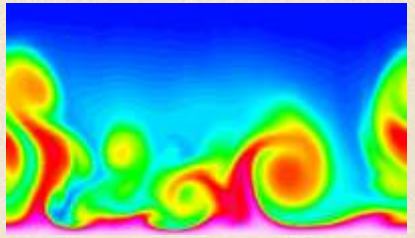


The Cancer Imaging Archive (TCIA)
https://www.cancerimagingarchive.net/



https://www.spect.com/blog/what-is-a-phantom-how-does-it-work

Sensor Networks



Computational Simulations

https://noaadata.apps.nsidc.org/NOAA/G02135/north/daily/images/

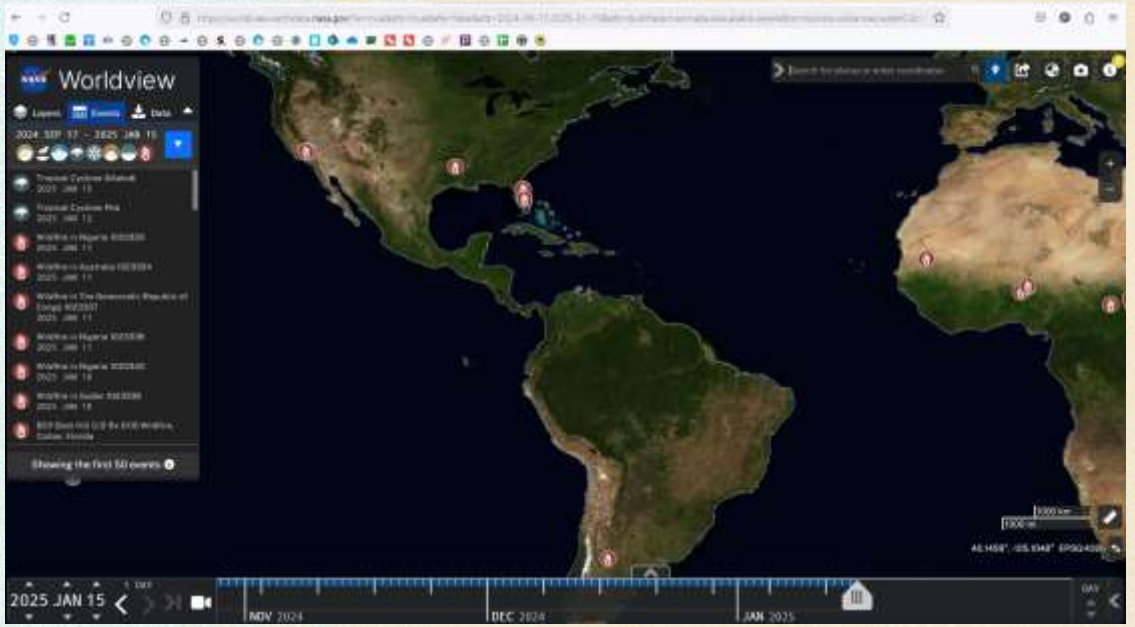
Why Data Mining?
(1 of 2) Commercial Viewpoint

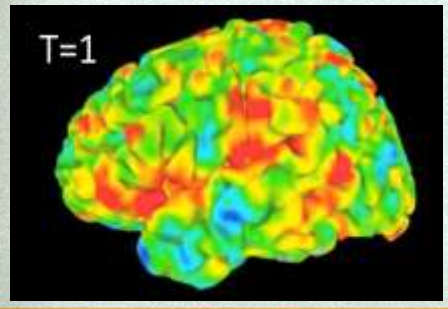
- Lots of data is being
  - Web data
    - Google has Petabytes of web data
    - Facebook has billions of active users
  - purchases at the department/ grocery stores, e-commerce
    - Amazon handles millions of visits/day
  - Bank/Credit Card transactions

- Computers have become cheaper and more powerful
- Competitive Pressure is Strong
  - Provide better, customized services for an edge (e.g. in Customer Relationship Management)

Why Data Mining?
(2 of 2) Scientific Viewpoint

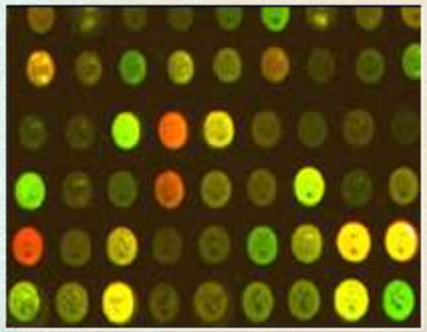
- Data collected and stored at enormous speeds
  - remote sensors on a satellite
    - NASA EOSDIS archives over petabytes of earth science data / year
  - telescopes scanning the skies
    - Sky survey data
  - High-throughput biological data
  - scientific simulations
    - terabytes of data generated in a few hours
- Data mining helps scientists
  - in automated analysis of massive datasets
  - In hypothesis formation

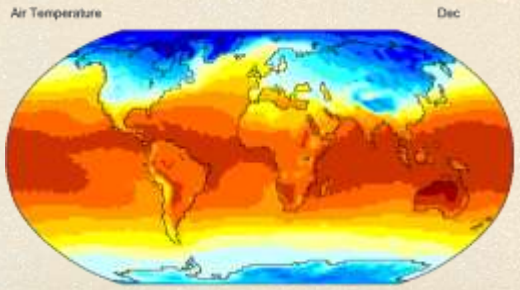




fMRI Data from Brain


Sky Survey Data


Gene Expression Data

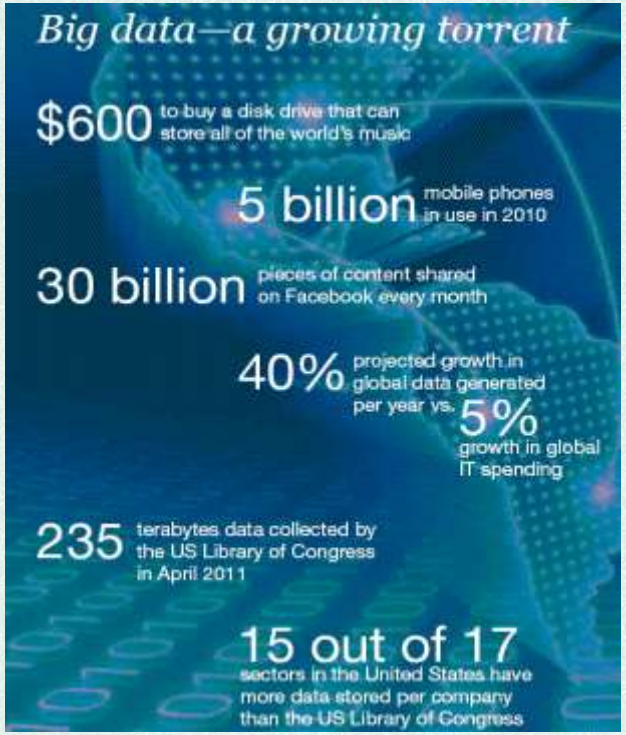


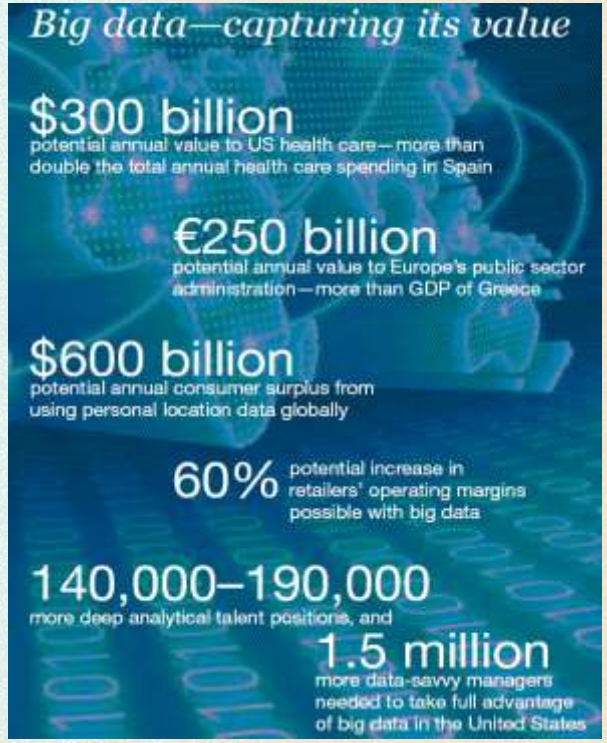
Surface Temperature of Earth

Great opportunities to improve productivity in all walks of life

McKinsey Global Institute

Big data: The next frontier for innovation, competition, and productivity

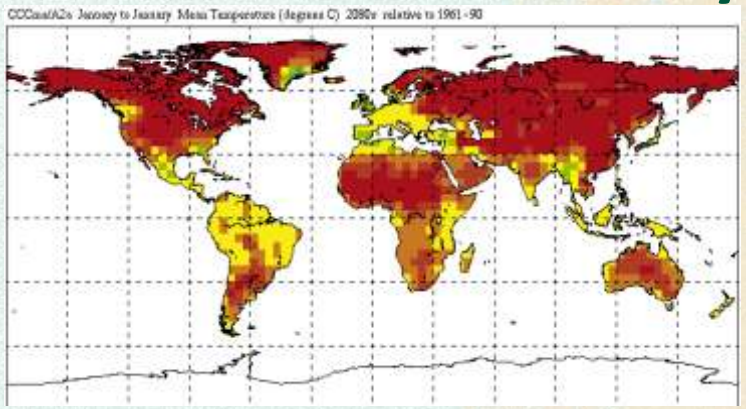




UAA 2025

Opportunities to Solve Society's Major Problems

Improving health care and reducing costs



Predicting the impact of climate change

Preventing wildfires

https://portugalchama.pt/

Arctic Coastal Erosion: https://toolkit.climate.gov/case-studies/relocating-kivalina

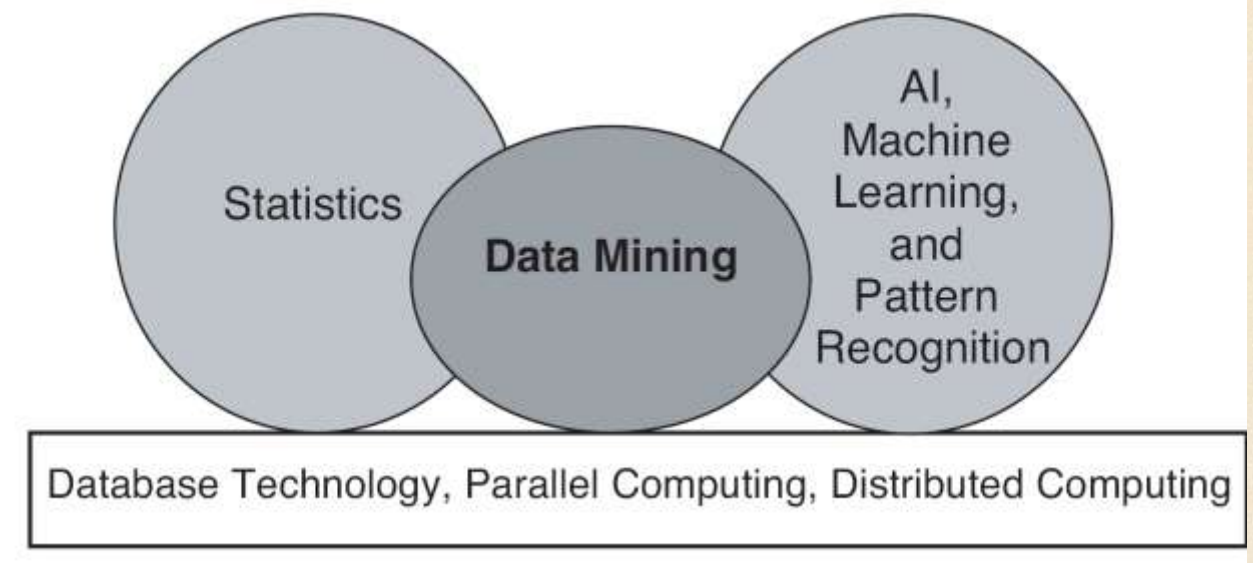
Finding alternative/ green energy sources

Reducing hunger and poverty by increasing agriculture production

Origins of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems

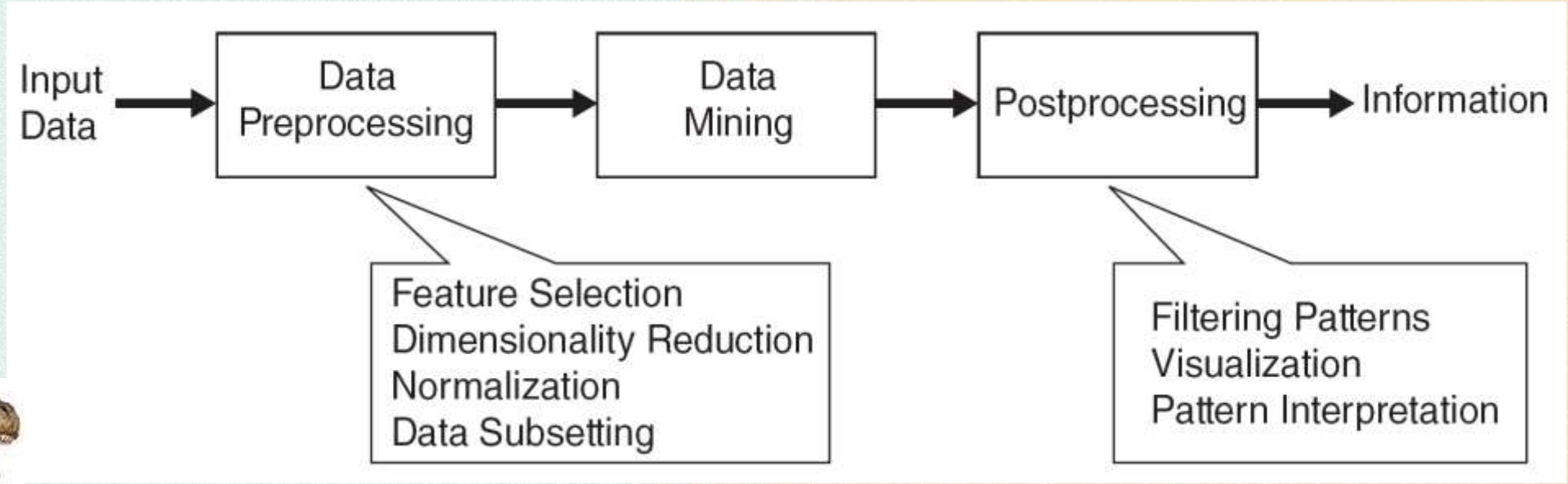
- Traditional techniques may be unsuitable due to data that is
  - Large-scale
  - High-dimensional
  - Heterogeneous
  - Complex
  - Distributed



- A key component of the emerging field of data science and data-driven discovery

What is Data Mining?

- Many Definitions
  - Non-trivial extraction of implicit, previously unknown, and potentially useful information from data
  - Exploration & analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns



Data Mining and Knowledge Discovery in Databases

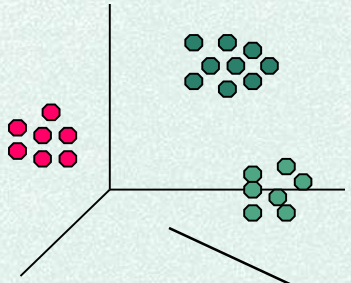
A. Data Preprocessing

- Preparing the data.
  - Data Quality.
  - Data Cleaning.
  - Near Duplicate Detection.

B. Data Mining

- Prediction Methods
  - Use some variables to predict unknown or future values of other variables.
- Description Methods
  - Find human-interpretable patterns that describe the data.

Data Mining Tasks ...

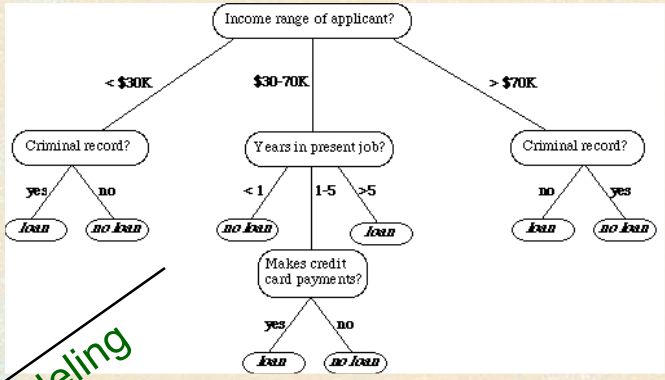


Clustering

Data

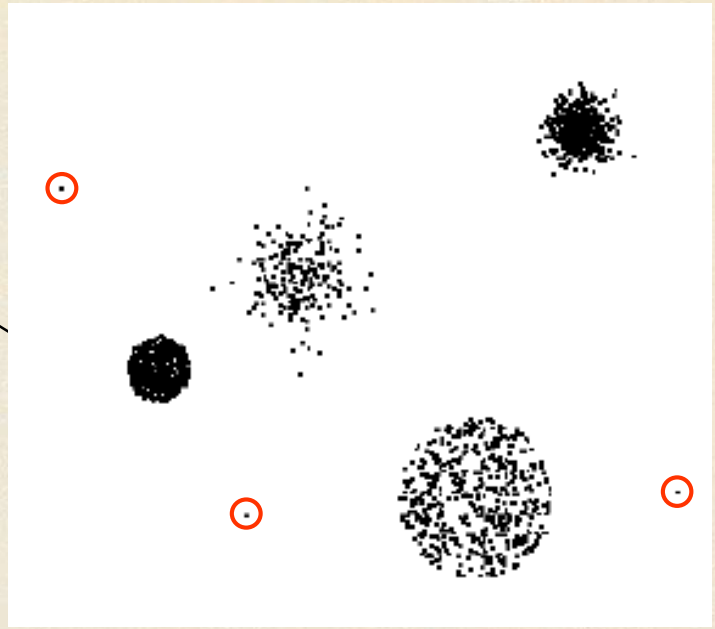
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes
11	No	Married	60K	No
12	Yes	Divorced	220K	No
13	No	Single	85K	Yes
14	No	Married	75K	No
15	No	Single	90K	Yes

Predictive Modeling



Association Rules

Anomaly Detection



Predictive Modeling

- Classification: discrete target variable(s).
- Regression: continuous target variable(s).

Predictive Modeling:

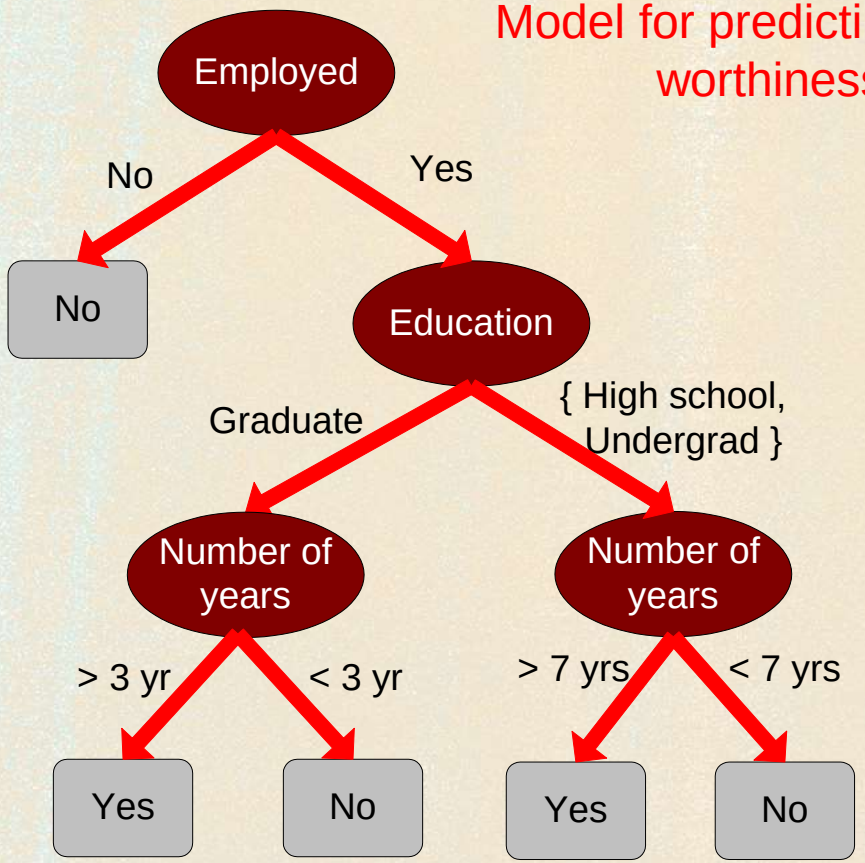
(A) Classification: for discrete target variables.

- Find a model for class attribute as a function of the values of other attributes

Class

Tid	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Graduate	5	Yes
2	Yes	High School	2	No
3	No	Undergrad	1	No
4	Yes	High School	10	Yes
…	…	…	…	…

Model for predicting credit worthiness

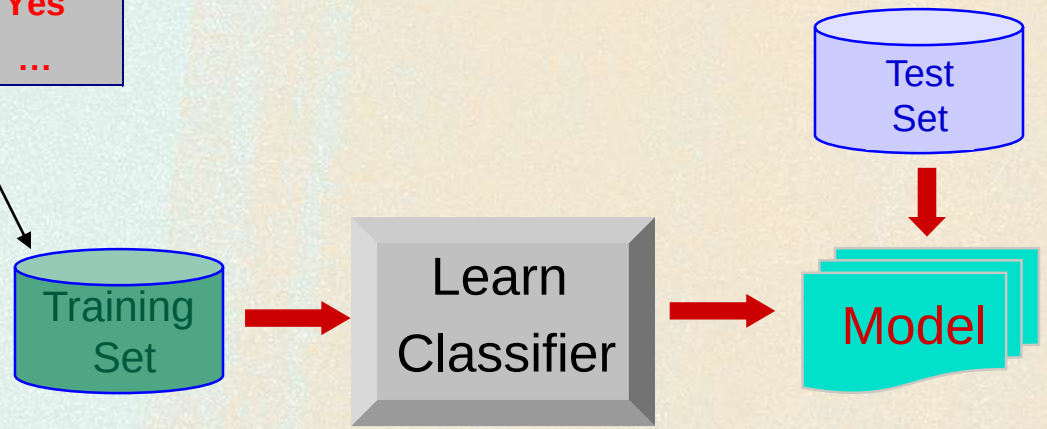


Classification Example

categorical categorical quantitative class

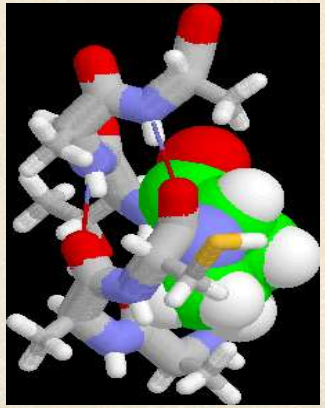
Tid	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Graduate	5	Yes
2	Yes	High School	2	No
3	No	Undergrad	1	No
4	Yes	High School	10	Yes
…	…	…	…	…

Tid	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Undergrad	7	?
2	No	Graduate	3	?
3	Yes	High School	2	?
…	…	…	…	…



Examples of Classification Task

- Classifying credit card transactions as legitimate or fraudulent
- Classifying land covers (water bodies, urban areas, forests, etc.) using satellite data
- Categorizing news stories as finance, weather, entertainment, sports, etc
- Identifying intruders in the cyberspace
- Predicting tumor cells as benign or malignant
- Classifying secondary structures of protein as alpha-helix, beta-sheet, or random coil



Classification: Sample Applications

- (1 of 5) Fraud Detection
  - **Goal:** Predict fraudulent cases in credit card transactions.
  - **Approach:**
    - Use credit card transactions and the information on its account-holder as attributes.
      - When does a customer buy, what do they buy, how often they pay on time, where do they buy, etc
    - Label past transactions as fraud or fair transactions. This forms the class attribute.
    - Learn a model for the class of the transactions.
    - Use this model to detect fraud by observing credit card transactions on an account.

(2 of 5) Churn prediction for telephone customers

- **Goal:** To predict whether a customer will likely be lost to a competitor.
- **Approach:**
  - Use detailed record of transactions with each of the past and present customers, to find attributes.
    - How often the customer calls, where he calls, what time-of-the day he calls most, his financial status, marital status, etc.
  - Label the customers as loyal or disloyal.
  - Find a model for loyalty.

From [Berry & Linoff] Data Mining Techniques, 1997

(3 of 5) Sky Survey Cataloging

- **Goal:** To predict class (star or galaxy) of sky objects, especially visually faint ones, based on the telescopic survey images (from Palomar Observatory).
  - 3000 images with 23,040 x 23,040 pixels per image.
- **Approach:**
  - Segment the image.
  - Measure image attributes (features) - 40 of them per object.
  - Model the class based on these features.
  - Success Story: Could find 16 new high red-shift quasars, some of the farthest objects that are difficult to find!

From [Fayyad, et.al.] Advances in Knowledge Discovery and Data Mining, 1996

(4 of 5) Binary classification of image quality

2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)

Classifying and Quantifying MRI Image Quality from DICOM Data at the Edge

Nishchal Singi* Puneet Sharma† Rochan Singh* Pradeeban Kathiravelu‡
*Keshav Memorial Institute of Technology, Hyderabad, TS 500029, India.
†Emory University, Atlanta, GA 30306, USA.
‡University of Alaska Anchorage, Anchorage, AK 99508, USA

CCECE) | 979-8-3503-7162-8/24$31.00 ©2024 IEEE | DOI: 10.1109/CCECE59415.2024.10667219

***Abstract*—Image quality is a critical aspect of diagnosis in radiology for the radiographic images produced by various scanner modalities. A radiologist can view and describe an image of good quality. Optimizing the image quality while minimizing the scan duration is crucial for operational efficiency and patient comfort. This paper presents a systematic approach to classifying MRI images and presenting them with quantitative attributes derived from the metadata of the DICOM (Digital Imaging and Communications in Medicine) images. We build *MriqNet*, a framework that executes deep learning algorithms on the DICOM images in a computing environment to determine the image quality and display the classified results in conjunction with DICOM header attributes. *MriqNet* quantifies the image quality, image scan duration, and other DICOM series-level metrics. We deploy and test *MriqNet* with our open-source software framework, *Niffler*, to run the image quality pipelines in environments like the cloud and the edge. Evaluations on *MriqNet* with MRI images show its accuracy and performance in classifying and quantifying image quality.**

***Index Terms*—Machine Learning, Digital Imaging and Communications in Medicine (DICOM), Image Quality, Binary Classification.**

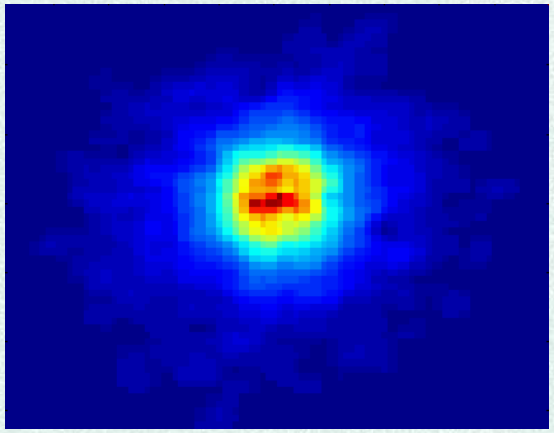
must recognize suboptimal image quality as it occurs and decide whether to repeat the acquisition. Due to competing exam workflow and patient care demands, this may be challenging in busy MRI departments. As a result, visual inspection of images is often done after MR exams or missed entirely. An automated, deep learning-based classification system would greatly benefit fast image quality assessment. Such a framework also facilitates retrospective image review of the site- or protocol-specific quality performance on a large scale for quality improvement purposes.

The image quality depends on several factors, some static and others that the radiographer can control and optimize. Some of these factors are the resolution (i.e., matrix, the field of view, and slice thickness), Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Contrast to noise ratio (CNR), Artifacts, and MRI hardware specifications. SNR is an essential, albeit non-specific, parametric indicator of image quality and is highly controlled by the user. On images, SNR is equal to the ratio of the average signal intensity over the

(5 of 5) Classifying Galaxies

http://aps.umn.edu

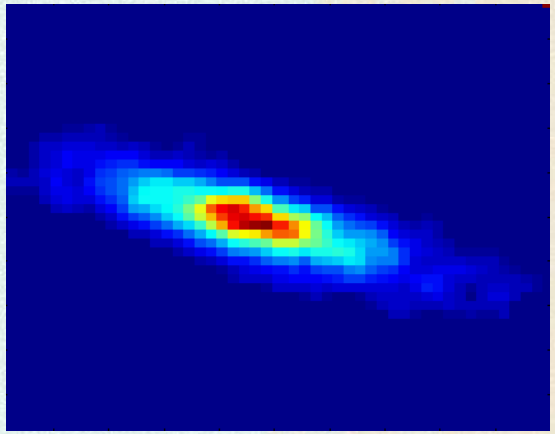
Early



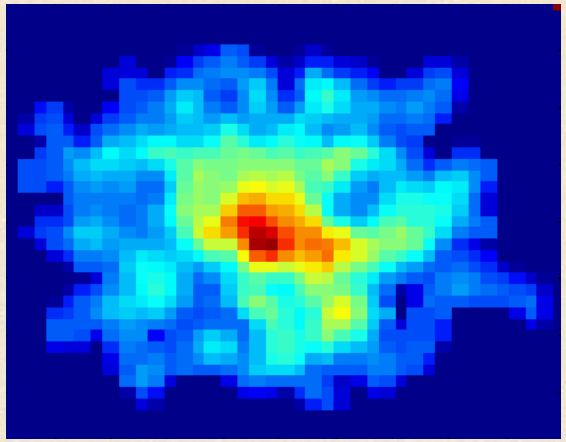
Class:
- Stages of Formation

Attributes:
- Image features,
- Characteristics of light waves received, etc.

Intermediate



Late



Data Size:
- 72 million stars, 20 million galaxies
- Object Catalog: 9 GB
- Image Database: 150 GB

UAA 2025

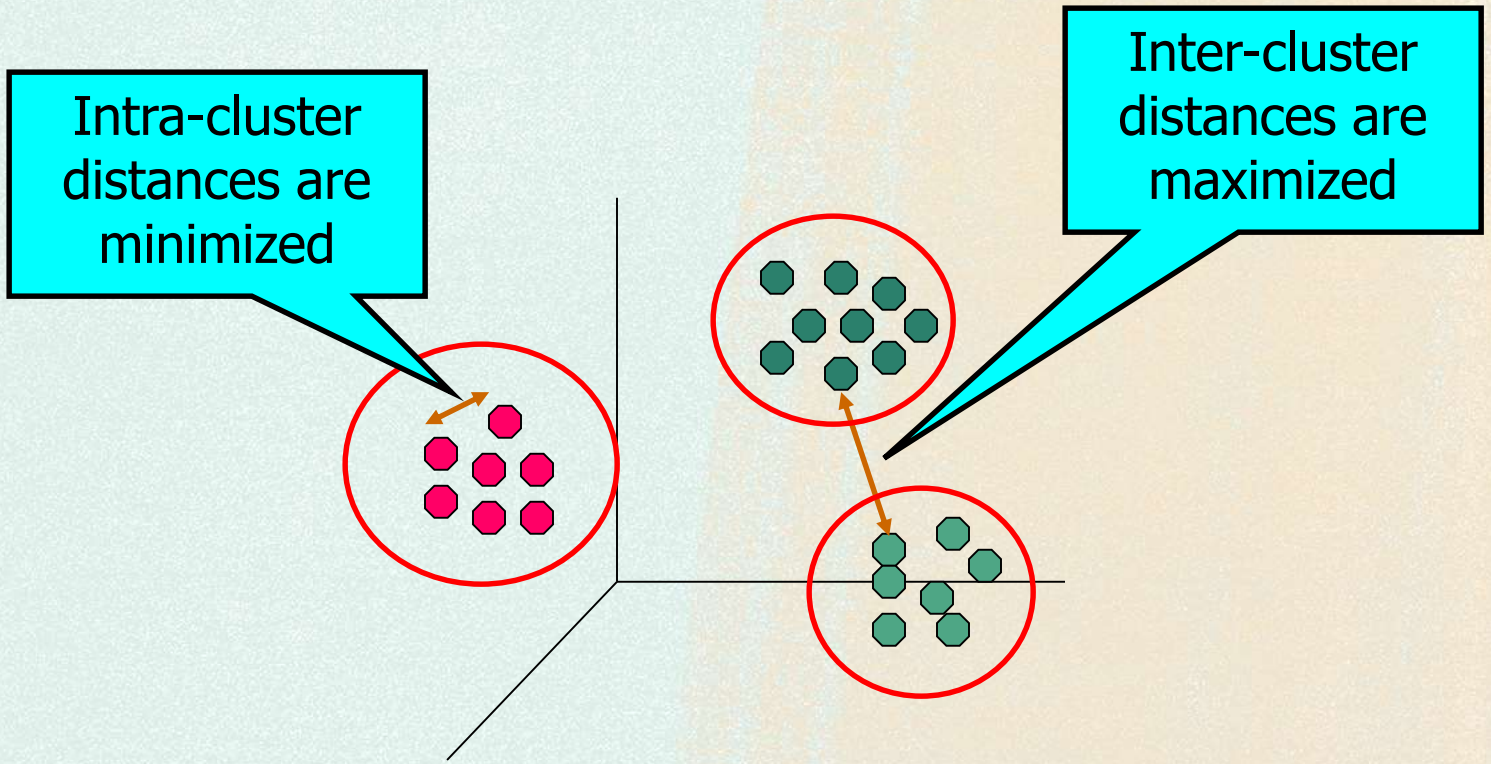
Predictive Modeling:

(B) Regression

- Predict a value of a given continuous target variable based on the values of other variables, assuming a linear or nonlinear model of dependency.
- Extensively studied in statistics and neural network fields.
- Examples:
  - Predicting sales amounts of new products based on advertising expenditure.
  - Predicting wind velocities as a function of temperature, humidity, air pressure, etc.
  - Time series prediction of stock market indices.

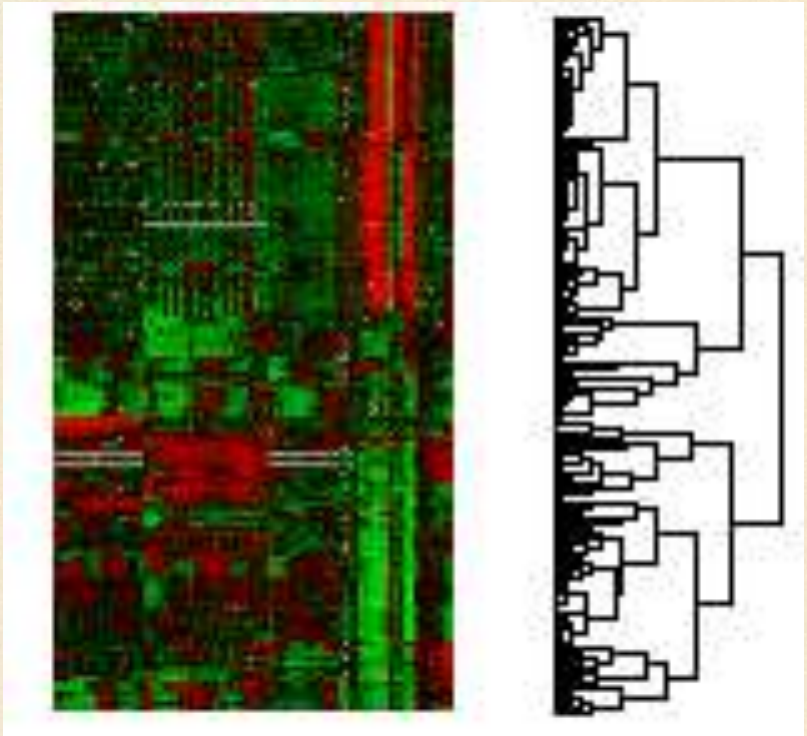
Clustering

- Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups

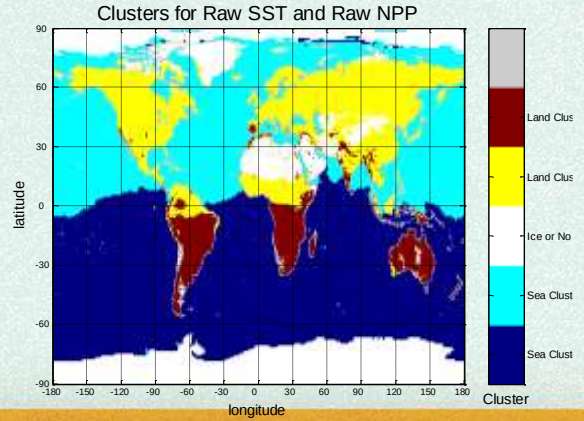


Applications of Cluster Analysis

- **Understanding**
  - Custom profiling for targeted marketing
  - Group-related documents for browsing
  - Group genes and proteins that have similar functionality
  - Group stocks with similar price fluctuations
- **Summarization**
  - Reduce the size of large data sets



Courtesy: Michael Eisen



Use of K-means to partition Sea Surface Temperature (SST) and Net Primary Production (NPP) into clusters that reflect the Northern and Southern Hemispheres.

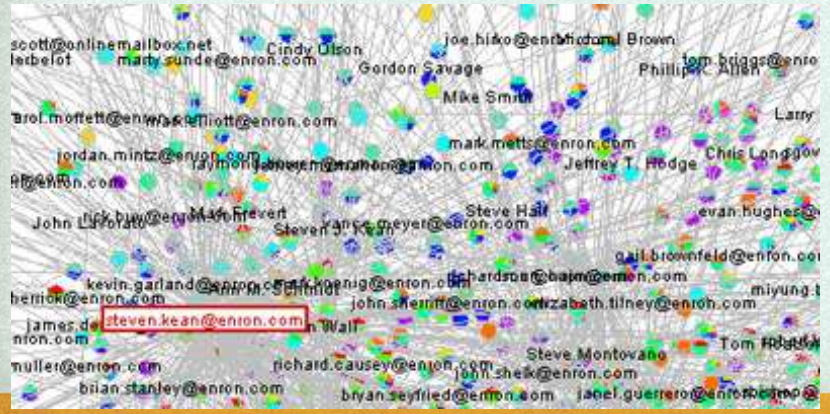
Clustering: Sample Applications

- (1 of 3) Market Segmentation:
  - **Goal:** subdivide a market into distinct subsets of customers where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix.
  - **Approach:**
    - Collect different attributes of customers based on their geographical and lifestyle related information.
    - Find clusters of similar customers.
    - Measure the clustering quality by observing buying patterns of customers in same cluster vs. those from different clusters.

(2 of 3) Document Clustering

- **Goal:** To find groups of documents that are similar to each other based on the important terms appearing in them.
- **Approach:** To identify frequently occurring terms in each document. Form a similarity measure based on the frequencies of different terms. Use it to cluster.

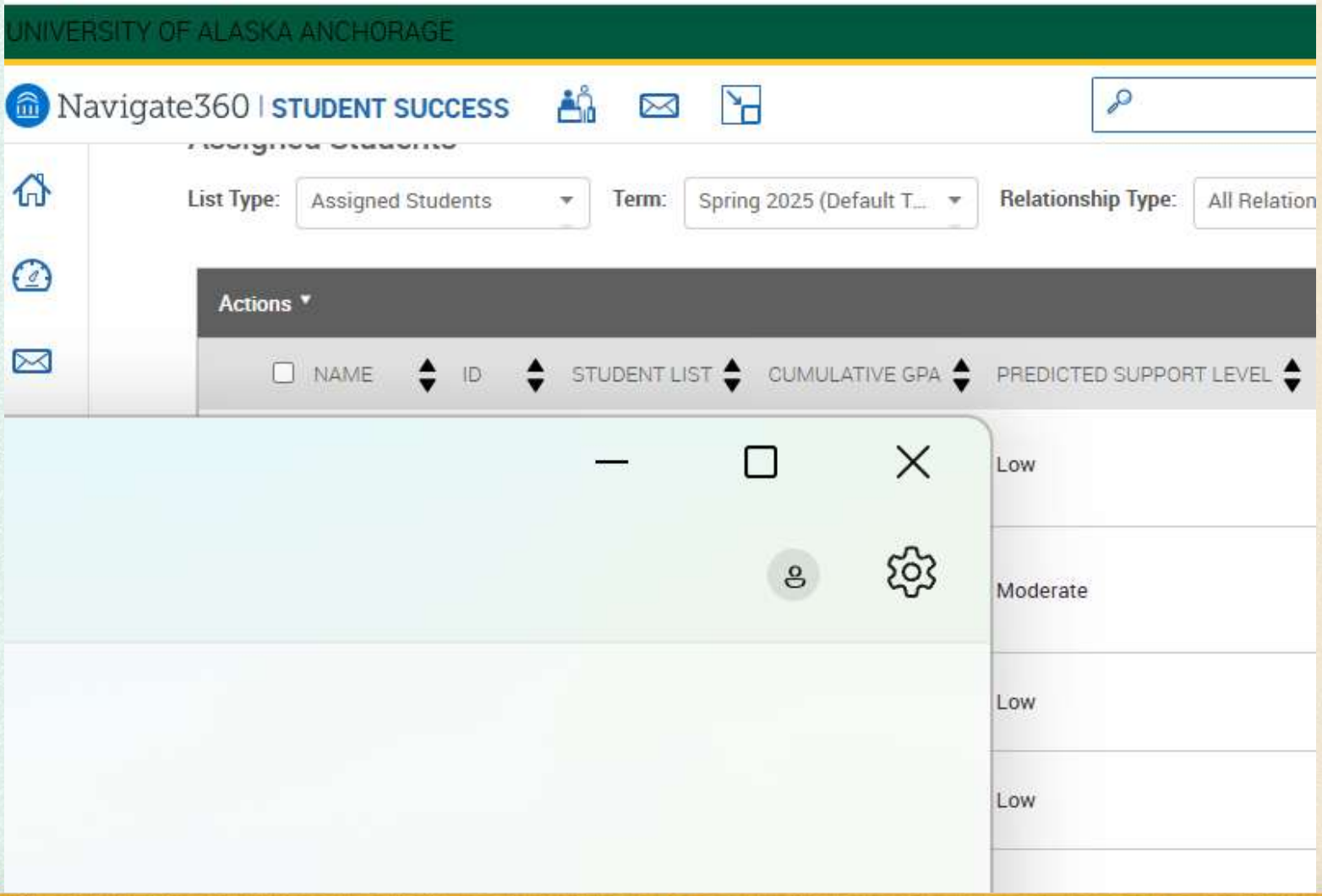
Enron email dataset



https://www.kaggle.com/datasets/wcukierski/enron-email-dataset

The Enron email dataset contains approximately 500,000 emails generated by employees of the Enron Corporation. It was obtained by the Federal Energy Regulatory Commission during its investigation of Enron's collapse.

(3 of 3) Predict required Student Support based on their past success (GPA, interactions, withdrawals)



Association Rule Discovery: Definition

- Given a set of records each of which contain some number of items from a given collection
  - Produce dependency rules which will predict occurrence of an item based on occurrences of other items.

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

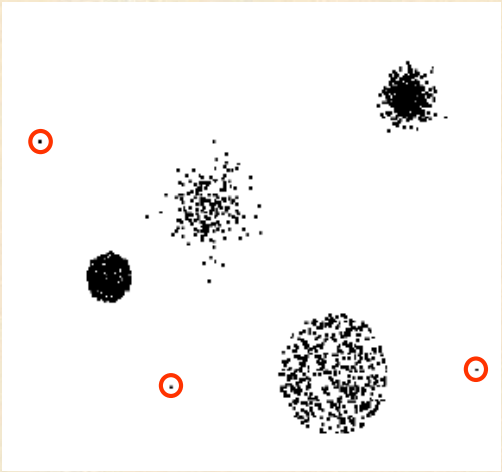
Rules Discovered:
{Milk} --> {Coke}
{Diaper, Milk} --> {Beer}

Association Analysis: Applications

- Market-basket analysis
  - Rules are used for sales promotion, shelf management, and inventory management
- Telecommunication alarm diagnosis
  - Rules are used to find combination of alarms that occur together frequently in the same time period
- Medical Informatics
  - Rules are used to find combination of patient symptoms and test results associated with certain diseases

Deviation/Anomaly/Change Detection

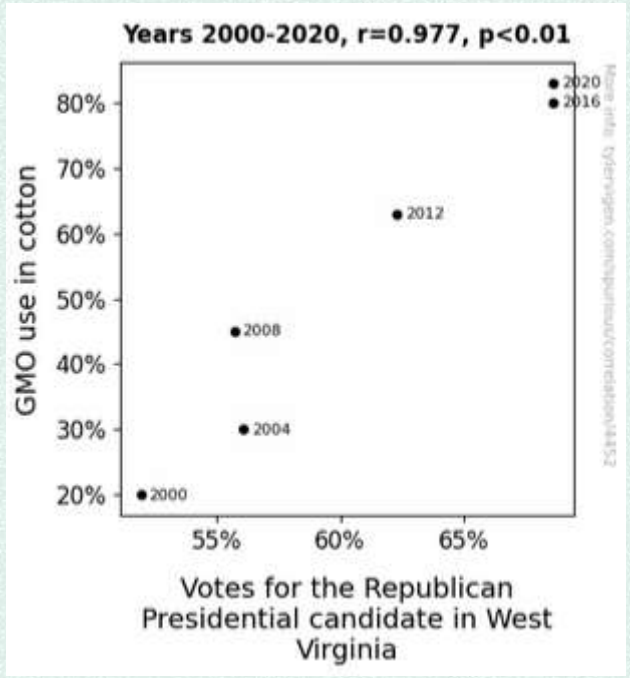
- Detect significant deviations from normal behavior
- Applications:
  - Credit Card Fraud Detection
  - Network Intrusion Detection.
  - Identify anomalous behavior from sensor networks for monitoring and surveillance.
  - Abnormal changes in heartbeat or oxygen level.

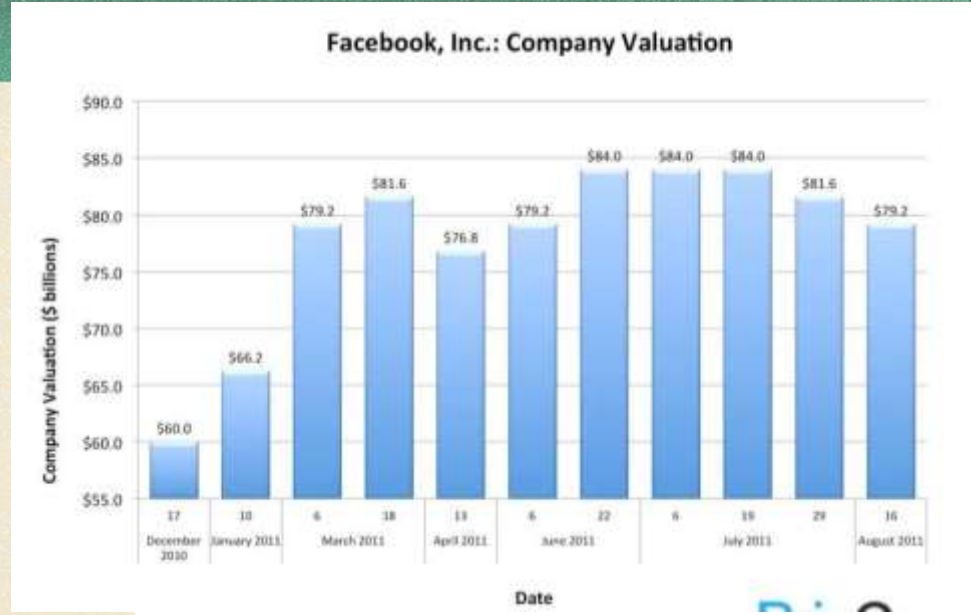


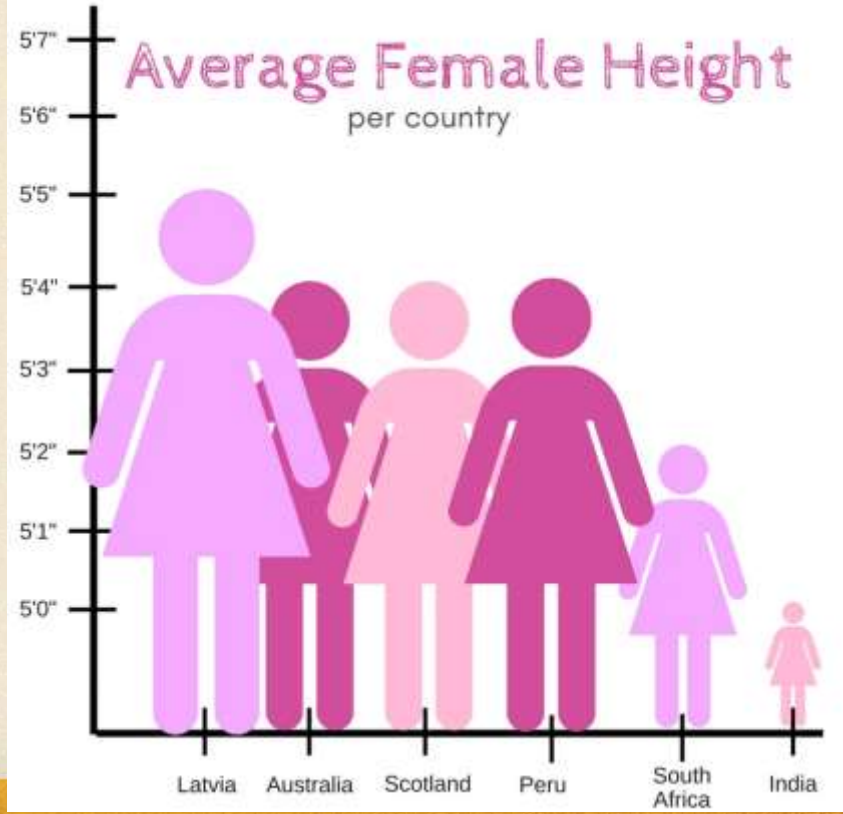
Motivating Challenges

- Scalability
- High Dimensionality
- Heterogeneous and Complex Data
- Data Ownership and Distribution

C. Postprocessing. Visualization: Present your findings properly







https://tylervigen.com/spurious/correlation/4452_gmo-use-in-cotton_correlates-with_votes-for-the-republican-presidential-candidate-in-west-virginia#datadetail

Data Software Engineer – Corporate Technology Data Engineering & Analytics
RemoteWorker US
New York, NY • via LinkedIn
6 hours ago · Full-time

G

Senior Software Engineer - Mobile
General Motors
Mountain View, CA (+2 others) • via General Motors Careers
3 days ago · Full-time · No degree mentioned

Senior Computer Programmer Analyst
PG&E Corporation
Oakland, CA • via PG&E
7 days ago · Full-time

I

Senior Computer Programmer
ITC Corp
United States • via Indeed
Full-time · No degree mentioned

K

Staff Software Engineer, Runtime Interpreter (Scheme/Rust)
Kong
Anywhere • via Indeed
Work from home · Full-time · No degree mentioned

Python Software Engineer
Russell Tobin
Boston, MA • via LinkedIn
5 days ago · Contractor · Health insurance · Dental insurance

AMAZING STORIES BEING WRITTEN EVERY DAY.

Software Engineer / Programmer Jobs

Data Mining Related Jobs.

Data Scientist- School of Medicine, Radiology and Imaging Sciences

Job Number 140336 **Job Type** Regular Full-Time **Division** School Of Medicine **Department** SOM: Rad: Interventional **Job Category** Information Technology US-GA-Atlanta

Emory Campus-Clifton Corridor **Remote Work Classification** Primarily On Campus **Health and Safety Information** Position involves clinical patient contact

Discover Your Career at Emory University

Emory University is a leading research university that fosters excellence and attracts world-class talent to innovate today and prepare leaders for the future. We welcome candidates who can contribute to the diversity and excellence of our academic community.

Description https://staff-emory.icims.com/jobs/140336/data-scientist--school-of-medicine%2c-radiology-and-imaging-sciences/job

JOB DESCRIPTION:

- Data Scientist is responsible for mining and analyzing data from various internal and external data sources to drive optimization and improvement of operations, clinical outcomes, and business strategies.
- Works with stakeholders throughout the organization to drive business solutions and are adept at using large data sets and models to test the effectiveness of different courses of action using predictive algorithms.
- They assess the effectiveness and accuracy of new data sources and data gathering techniques.
- Data Scientist is an expert in at least 1 programming language (e.g. R, SAS, Python), has proficiency in statistical theory and underlying mathematical distributions, and expertise in querying relational databases with SQL.
- They have experience with of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and can demonstrate real-world application of these techniques.

ADDITIONAL JOB DESCRIPTION:

- This candidate will be repsonsible for data extraction from PACS and EHR, data curation, and managing large datasets developed by the lab.

MINIMUM QUALIFICATIONS:

- Master's or PhD in Statistics, Mathematics, Computer Science or another quantitative field and 3 years of relevant experience.

PREFERRED QUALIFICATIONS:

- MS in data science, health informatics, engineering, or computer science

NOTE: Position tasks are generally required to be performed in-person at an Emory University location. Remote work from home day options may be granted at department discretion. Emory reserves the right to change remote work status with notice to employee.

Data Scientist - School of Medicine, Pulmonary

Job Number 140538 | **Job Type** Regular Full-Time | **Division** School Of Medicine | **Department** SOM: Medicine: Pulmonary | **Job Category** Information Technology | US-GA-Atlanta | Emory Campus-Clifton Corridor | **Remote Work Classification** Hybrid | **Health and Safety Information** Not Applicable

Discover Your Career at Emory University

Data Mining Related Jobs.

Emory University is a leading research university that fosters excellence and attracts world-class talent to innovate today and prepare leaders for the future. We welcome candidates who can contribute to the diversity and excellence of our academic community.

Description

Our lab specializes in developing precision medicine algorithms for critically ill patients using EHR data. We are looking for a full-time data scientist with proficiency in R and python, and the ability to take machine learning projects from beginning to end, from data cleaning to pre-processing to development and validation of supervised and unsupervised learning algorithms.

KEY RESPONSIBILITIES:

- Data Scientist is responsible for mining and analyzing data from various internal and external data sources to drive optimization and improvement of operations, clinical outcomes, and business strategies.
- Works with stakeholders throughout the organization to drive business solutions and are adept at using large data sets and models to test the effectiveness of different courses of action using predictive algorithms.
- They assess the effectiveness and accuracy of new data sources and data gathering techniques.
- Data Scientist is an expert in at least 1 programming language (e.g. R, SAS, Python), has proficiency in statistical theory and underlying mathematical distributions, and expertise in querying relational databases with SQL.
- They have experience with of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and can demonstrate real-world application of these techniques.

MINIMUM QUALIFICATIONS:

- Master's or PhD in Statistics, Mathematics, Computer Science or another quantitative field and 3 years of relevant experience.

PREFERRED QUALIFICATIONS:

- Experience with Epic data extraction and handling.

https://staff-emory.icims.com/jobs/140538/data-scientist---school-of-medicine%2c-pulmonary/job

NOTE: This role will be granted the opportunity to work from home regularly but must be able to commute to Emory University on a flexible weekly schedule based upon business needs. Schedule is based on agreed upon guidelines of department. This role requires residency in the state of GA. Emory reserves the right to change remote work status with notice to employee.

Data Mining in Healthcare

- Use case: Classifying and Quantifying MRI Image Quality from DICOM Data

2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)

Classifying and Quantifying MRI Image Quality from DICOM Data at the Edge

Nishchal Singi* Puneet Sharma† Rochan Singh* Pradeeban Kathiravelu‡
*Keshav Memorial Institute of Technology, Hyderabad, TS 500029, India.
†Emory University, Atlanta, GA 30306, USA.
‡University of Alaska Anchorage, Anchorage, AK 99508, USA

CCECE) | 979-8-3503-7162-8248-31.00 ©2024 IEEE | DOI: 10.1109/CCECE59415.2024.10667219

***Abstract*—Image quality is a critical aspect of diagnosis in radiology for the radiographic images produced by various scanner modalities. A radiologist can view and describe an image of good quality. Optimizing the image quality while minimizing the scan duration is crucial for operational efficiency and patient comfort. This paper presents a systematic approach to classifying MRI images and presenting them with quantitative attributes derived from the metadata of the DICOM (Digital Imaging and Communications in Medicine) images. We build *MriqNet*, a framework that executes deep learning algorithms on the DICOM images in a computing environment to determine the image quality and display the classified results in conjunction with DICOM header attributes. *MriqNet* quantifies the image quality, image scan duration, and other DICOM series-level metrics. We deploy and test *MriqNet* with our open-source software framework, *Niffler*, to run the image quality pipelines in environments like the cloud and the edge. Evaluations on *MriqNet* with MRI images show its accuracy and performance in classifying and quantifying image quality.**

***Index Terms*—Machine Learning, Digital Imaging and Communications in Medicine (DICOM), Image Quality, Binary Classification.**

must recognize suboptimal image quality as it occurs and decide whether to repeat the acquisition. Due to competing exam workflow and patient care demands, this may be challenging in busy MRI departments. As a result, visual inspection of images is often done after MR exams or missed entirely. An automated, deep learning-based classification system would greatly benefit fast image quality assessment. Such a framework also facilitates retrospective image review of the site- or protocol-specific quality performance on a large scale for quality improvement purposes.

The image quality depends on several factors, some static and others that the radiographer can control and optimize. Some of these factors are the resolution (i.e., matrix, the field of view, and slice thickness), Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Contrast to noise ratio (CNR), Artifacts, and MRI hardware specifications. SNR is an essential, albeit non-specific, parametric indicator of image quality and is highly controlled by the user. On images, SNR is equal to the ratio of the average signal intensity over the

Radiographic Images

- Standardized across various modalities of scanners for **storage** and **transfer**.
  - X-Ray, MRI, CT, Ultrasound, …
- **DICOM (Digital Imaging and Communications in Medicine).**
  - A storage and messaging protocol and standard for radiographic images.

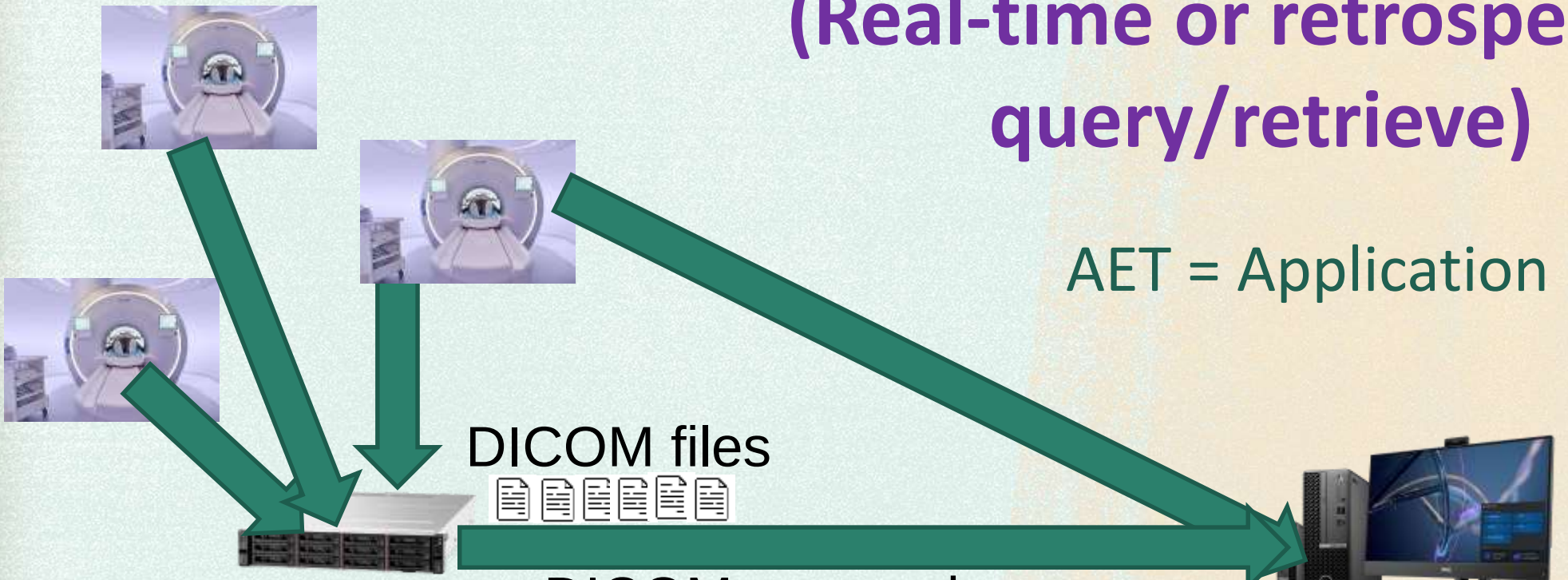
DICOM

- Extends TCP (Digital Imaging and Communications in Medicine) as a radiology messaging protocol
  - to send data between
    - scanners,
    - PACS (Picture Archiving and Communication System),
    - and other DICOM systems.
- Store and query/retrieve images of diverse modalities.

DICOM Data Transfer

(Real-time or retrospective query/retrieve)

AET = Application Entity Title



Niffler for DICOM image and RIS clinical data transfer and processing

Journal of Digital Imaging (2021) 34:1005–1013
https://doi.org/10.1007/s10278-021-00491-w

ORIGINAL PAPER

A DICOM Framework for Machine Learning and Processing Pipelines Against Real-time Radiology Images

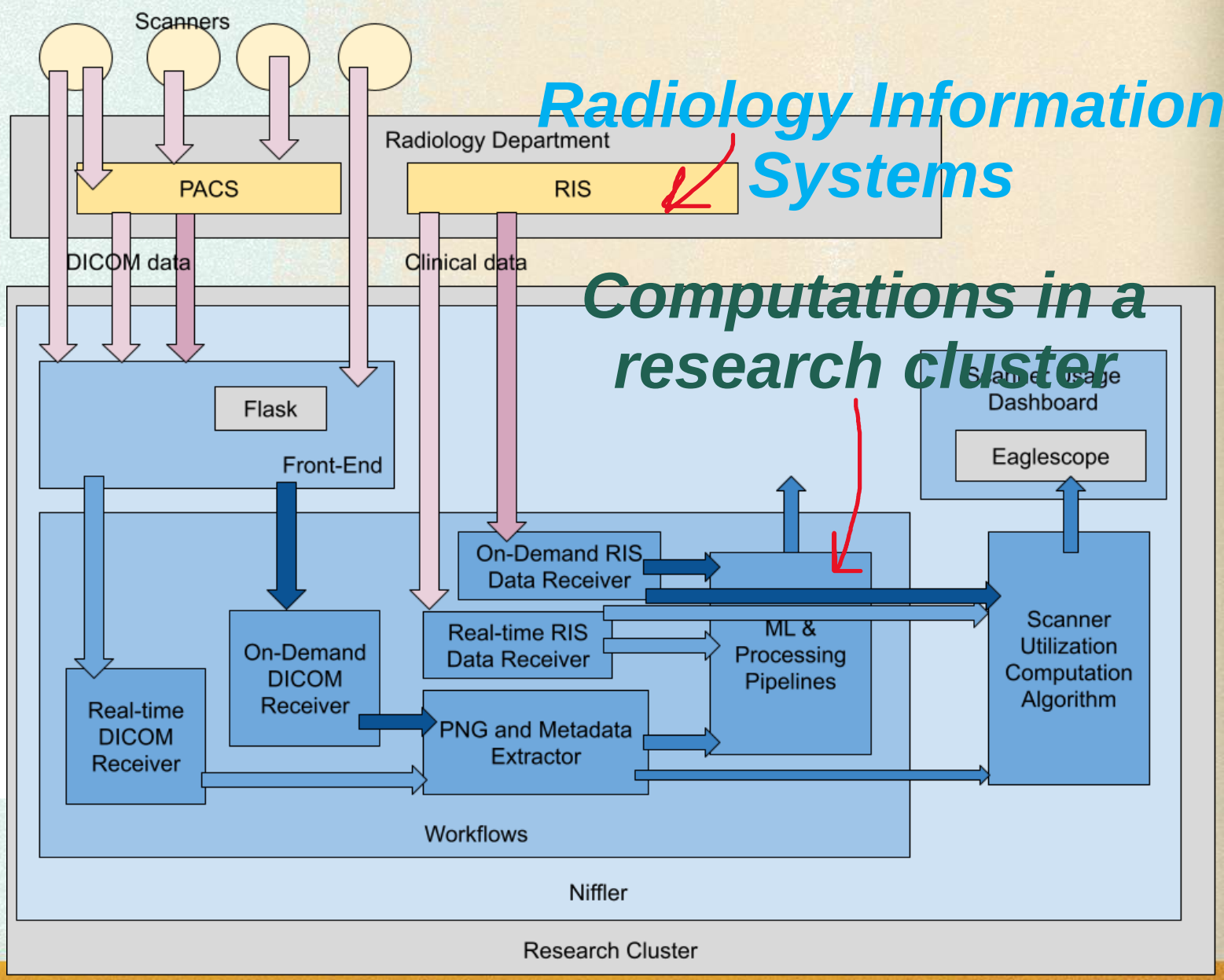
Pradeeban Kathiravelu[1] · Puneet Sharma[1] · Ashish Sharma[1] · Imon Banerjee[1] · Hari Trivedi[1] · Saptarshi Purkayastha[2] · Priyanshu Sinha[3] · Alexandre Cadrin-Chenevert[4] · Nabile Safdar[1] · Judy Wawira Gichoya[1]

Received: 26 August 2020 / Revised: 29 April 2021 / Accepted: 5 July 2021 / Published online: 17 August 2021

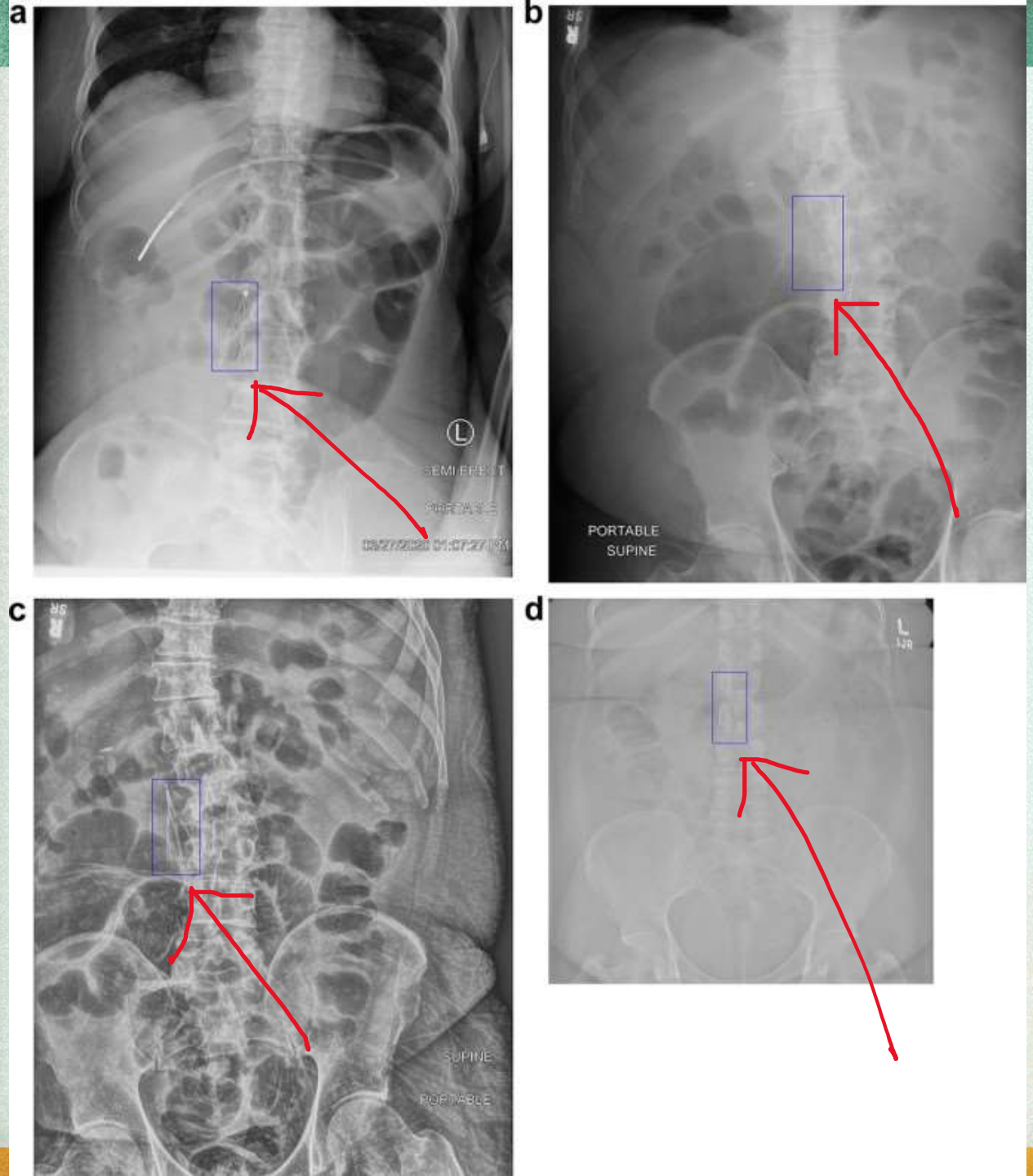
Abstract
Real-time execution of machine learning (ML) pipelines on radiology images is difficult due to limited computing resources in clinical environments, whereas running them in research clusters requires efficient data transfer capabilities. We developed Niffler, an open-source Digital Imaging and Communications in Medicine (DICOM) framework that enables ML and processing pipelines in research clusters by efficiently retrieving images from the hospitals' PACS and extracting the metadata from the images. We deployed Niffler at our institution (Emory Healthcare, the largest healthcare network in the state of Georgia) and retrieved data from 715 scanners spanning 12 sites, up to 350 GB/day continuously in real-time as a DICOM data stream over

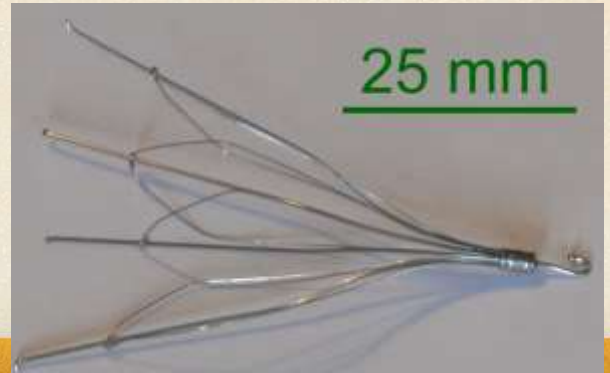
UAA 2025

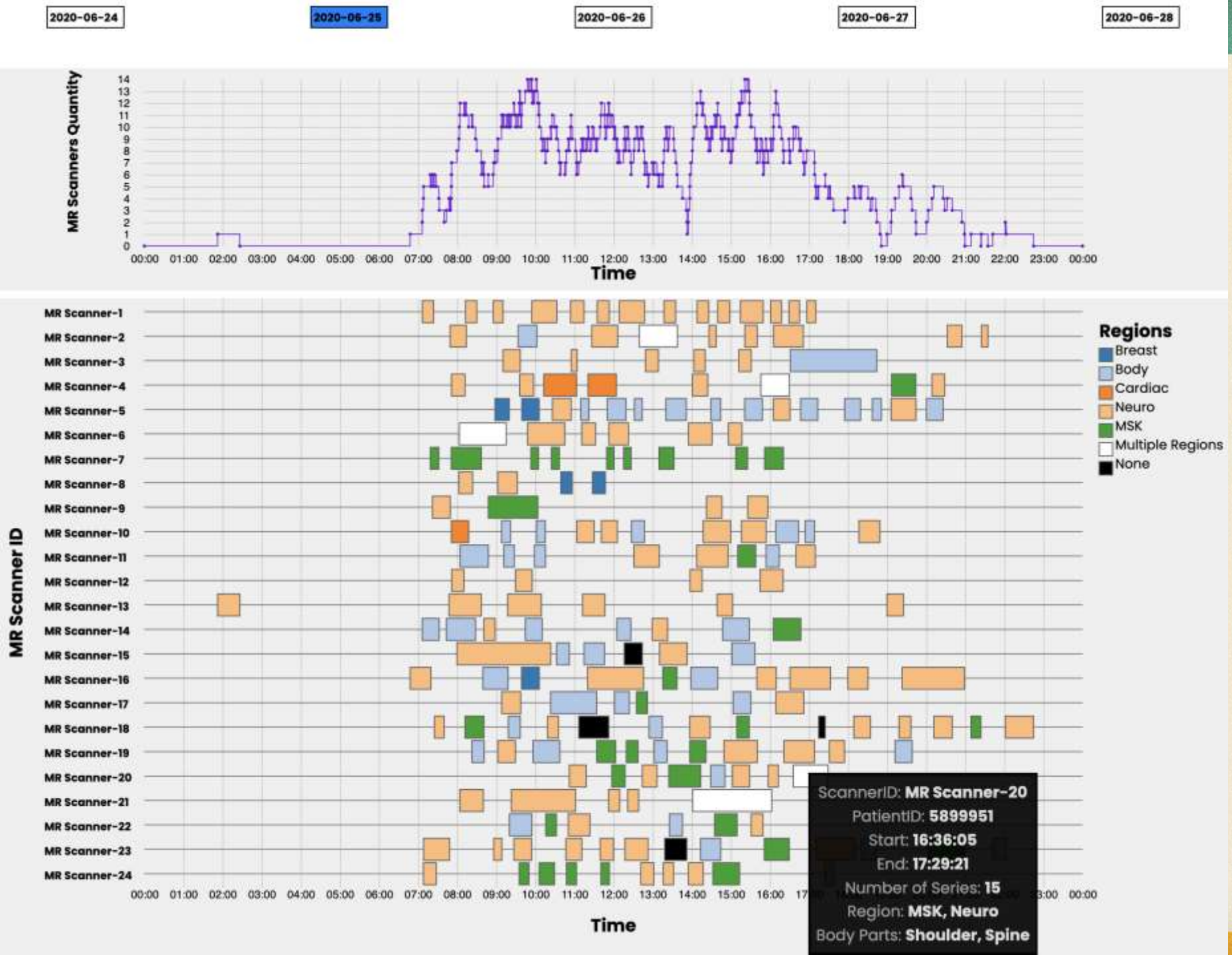
https://github.com/Emory-HITI/Niffler/



IVC (Inferior Vena Cava) Filter Detection in real-time from CXR (Chest X-ray) images with 96% accuracy

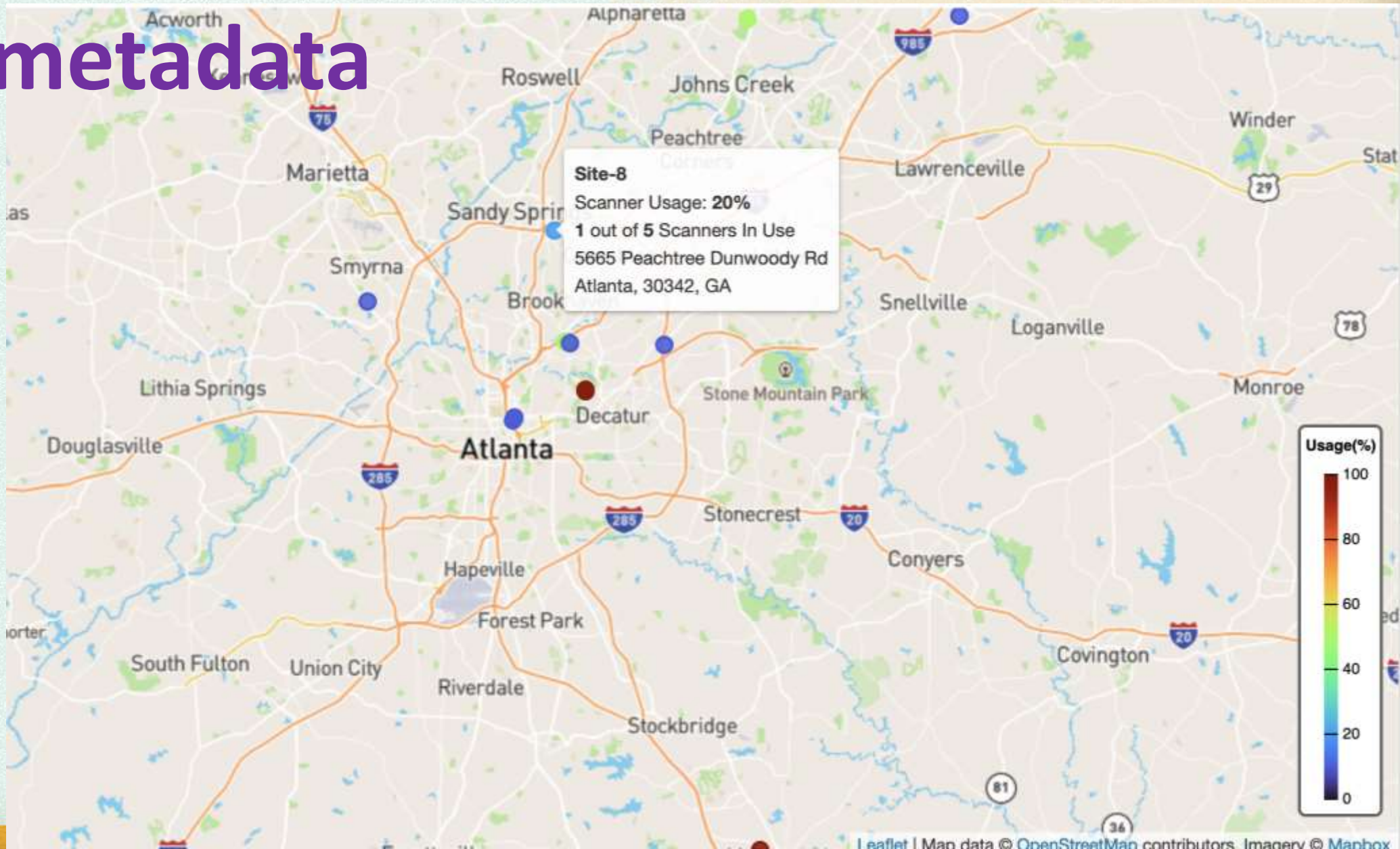






MRI Scanner utilization from DICOM and RIS data

MRI Scanner Usage in real-time from DICOM metadata



Demonstrated use cases…

- ML pipelines on DICOM images (received real-time or on-demand from PACS or scanners) in a research cluster.
- Scanner usage computations from DICOM metadata.
  - Operational efficiency and patient comfort.
    - Minimal scan time and minimal wait time.
  - Transfer of patients to other sites of the healthcare network with the scanner modality available.
- Ability to transfer images to a cloud or an edge node.
  - **Edge-to-Cloud transfers from the PACS or scanners via Niffler.**

Image Quality Matters!

- Certain challenges in maintaining image quality.
  - Patient movement during the scan.
  - Equipment and operator errors…

Can we use Data Mining to Classify Images?

- **Binary classification** of the quality: acceptable or not
- Of course, we cannot run these ML tasks on scanners!
- But we have Niffler.

Motivation

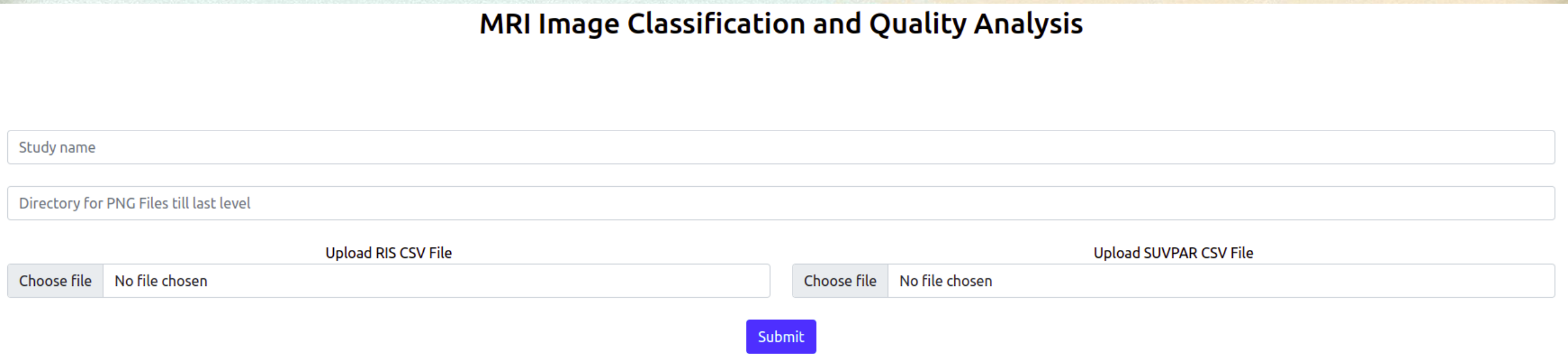
- A systematic approach to classifying MRI images using quantitative attributes from DICOM images and their metadata.

Proposed Solution: MriqNet

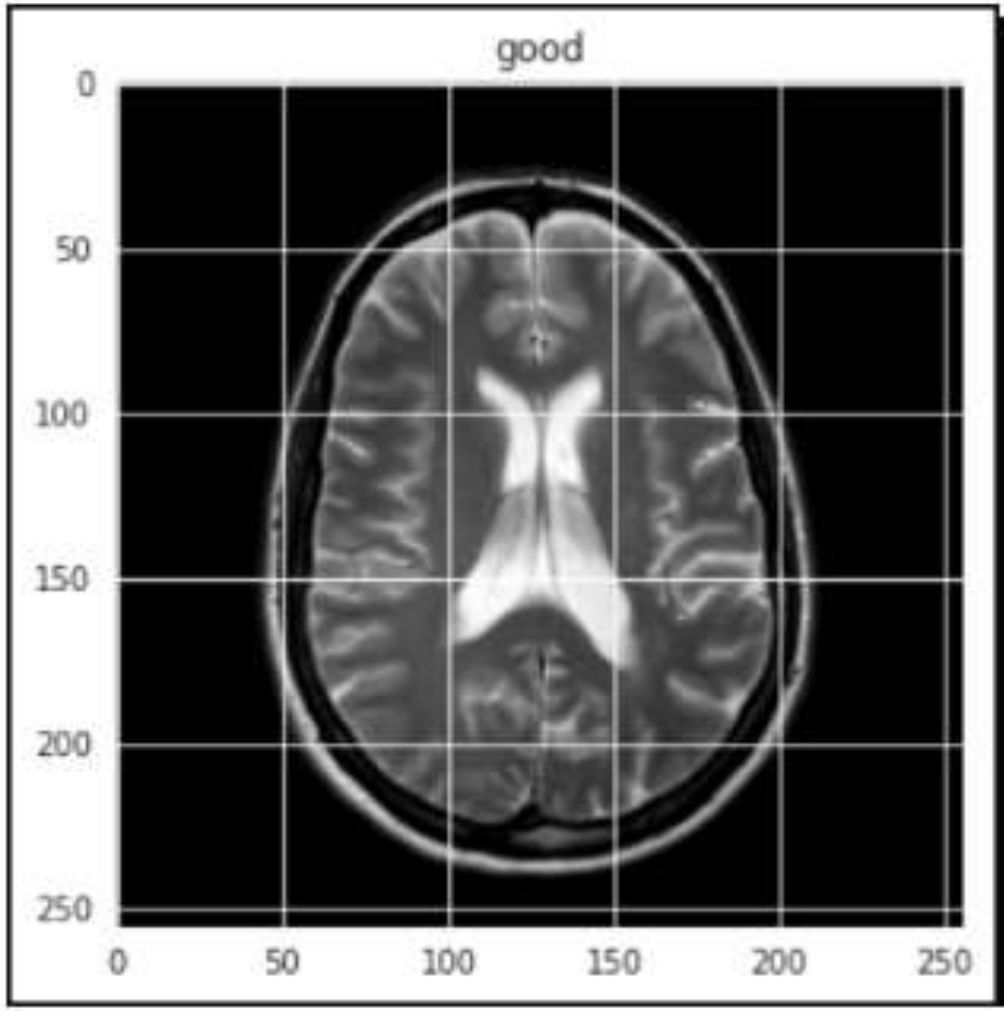
- A framework employing deep learning algorithms to assess image quality and display results alongside DICOM header attributes.
- MriqNet evaluates image quality, scan duration, and other metrics.
  - Deployed and tested with the opensource Niffler framework
  - Ability to run in the cloud and edge servers.
- Evaluations show MriqNet's accuracy and performance in classifying and quantifying image quality.

MriqNet front-end

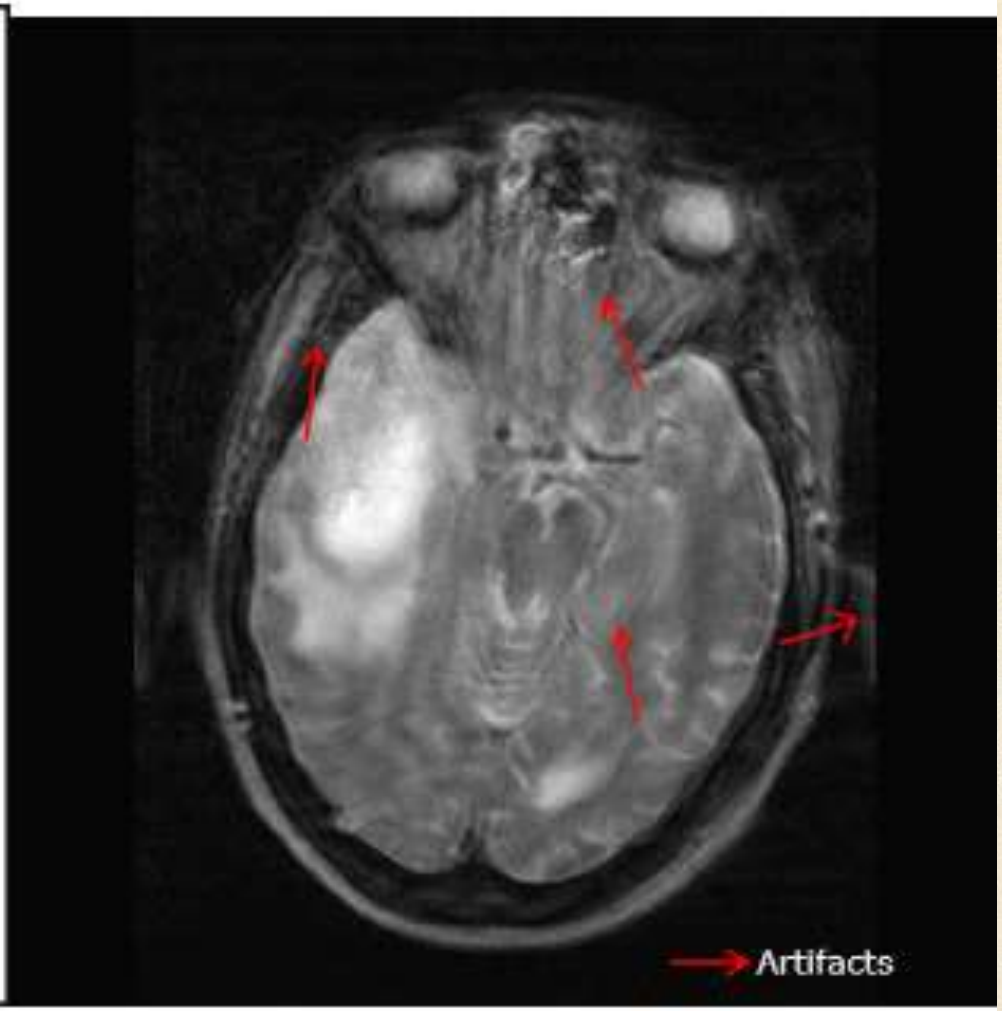
- Easy image uploads through a GUI and an API.



The binary classification of the images



(a) Classified as good. Clear. No artifacts are present.

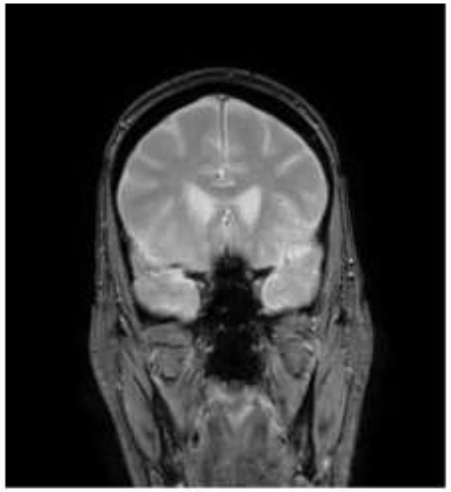


(b) Classified as bad. Visible artifacts are present. The red arrows indicate the artifacts.

Dashboard displaying the images classified as good.

Series Date - 20000423 | Series Time - 35332.22 | Series Duration - 184.54 sec | Ordered Item - MRI Brain w/o Contrast | Body Part - Brain

GOOD



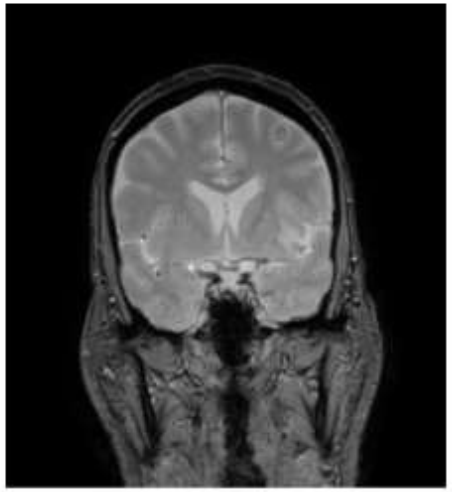
10f346cb8abc785d0907a9ef3722de1bfcbf9faae06152e368763c06.png

METRICS

SNR: 12.572

PSNR: 10.578653814386808

GOOD



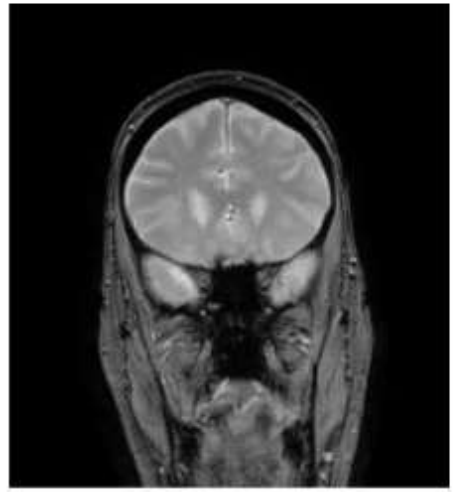
2425e9db7d7f532092d40d5ca1392889e00b882a71eb1d56c01dfe62.png

METRICS

SNR: 11.973

PSNR: 11.125848594517754

GOOD



df8a1e61f5dc388d8fea767530a17c64a1e981fe3dd6fe49eb090797.png

METRICS

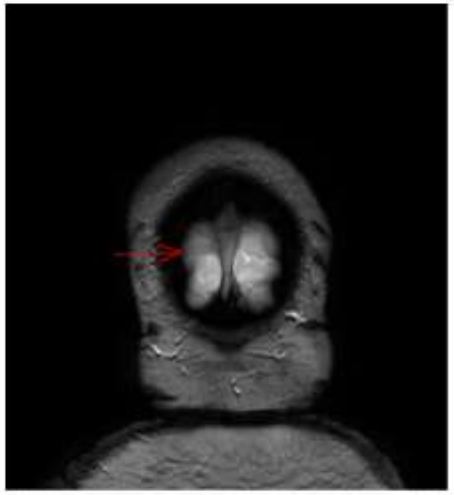
SNR: 12.965

PSNR: 11.296159436361686

Dashboard displaying the images classified as bad.

Series Date - 20000423 | Series Time - 35332.22 | Series Duration - 184.54 sec | Ordered Item - MRI Brain w/o Contrast | Body Part - Brain

BAD



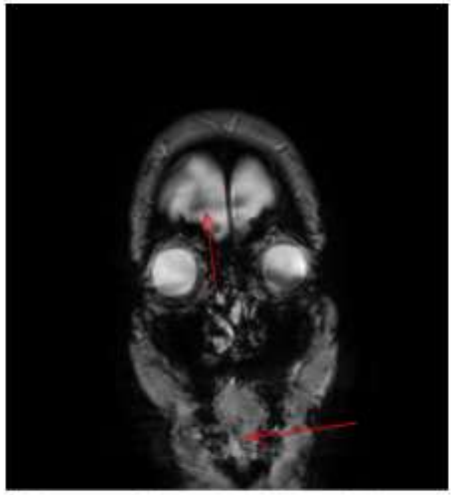
a80bbb0fcb135cb57352a6ba14ee61d6c9bfeec76eaba81a25f9d9f0.png

METRICS

SNR: 11.615

PSNR: 15.370512774775072

BAD



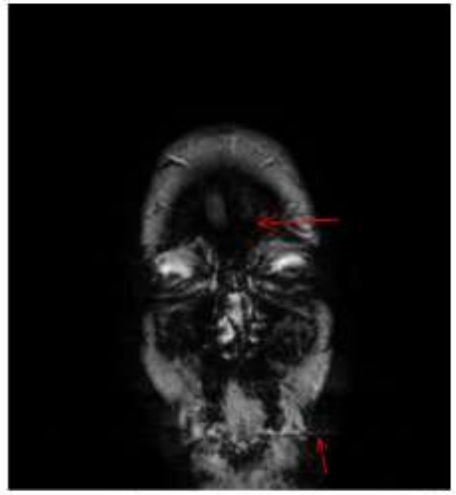
2d10b7656bd1ac9368450fb622c30b24dcbd6418a5d69b786a8d5cdd.png

METRICS

SNR: 12.283

PSNR: 15.275080429759601

BAD



b7b1b29453691e6984ab43419e67b262083e732357921a2cb2db4e35.png

METRICS

SNR: 11.999

PSNR: 17.45993712308109

→ Artifacts/Noise

How MriqNet Works?

- MriqNet provides a custom ten-layer deep learning architecture adapted for MR image classification to prevent overfitting.
- With Data augmentation, we additionally reduced overfitting by increasing the diversity and quantity of training data.
- The activation function used in the convolutional layers is ReLU (Rectified linear activation function). As MriqNet classifies images as ’good’ or ’bad’ (i.e., either 1 or 0), we identified ReLU as the most efficient choice.
- The total number of trainable parameters of the MriqNet model is 977,953.

Dataset used in this paper…

- **Brain MRI images** for training, validation, and testing.
- Image Source: **The Cancer Imaging Archive (TCIA)**
  - a public repository consisting of anonymized DICOM images.

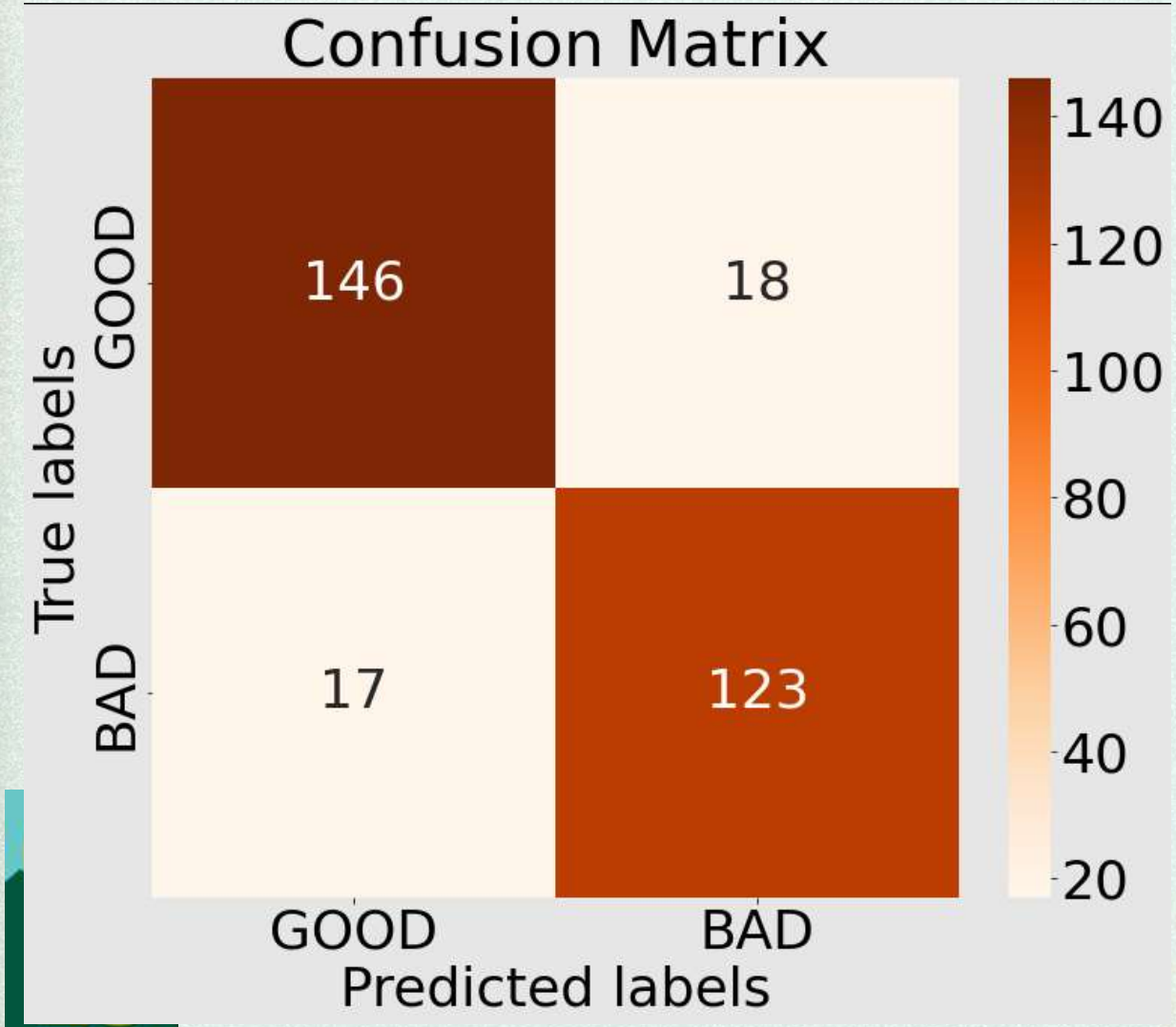
- For training and validation dataset – 70:30 split.
- Then we tested the model on a totally unseen new dataset

- For 100 epochs
  - An accuracy of 90% for training data and 86% for validation data.

Evaluation

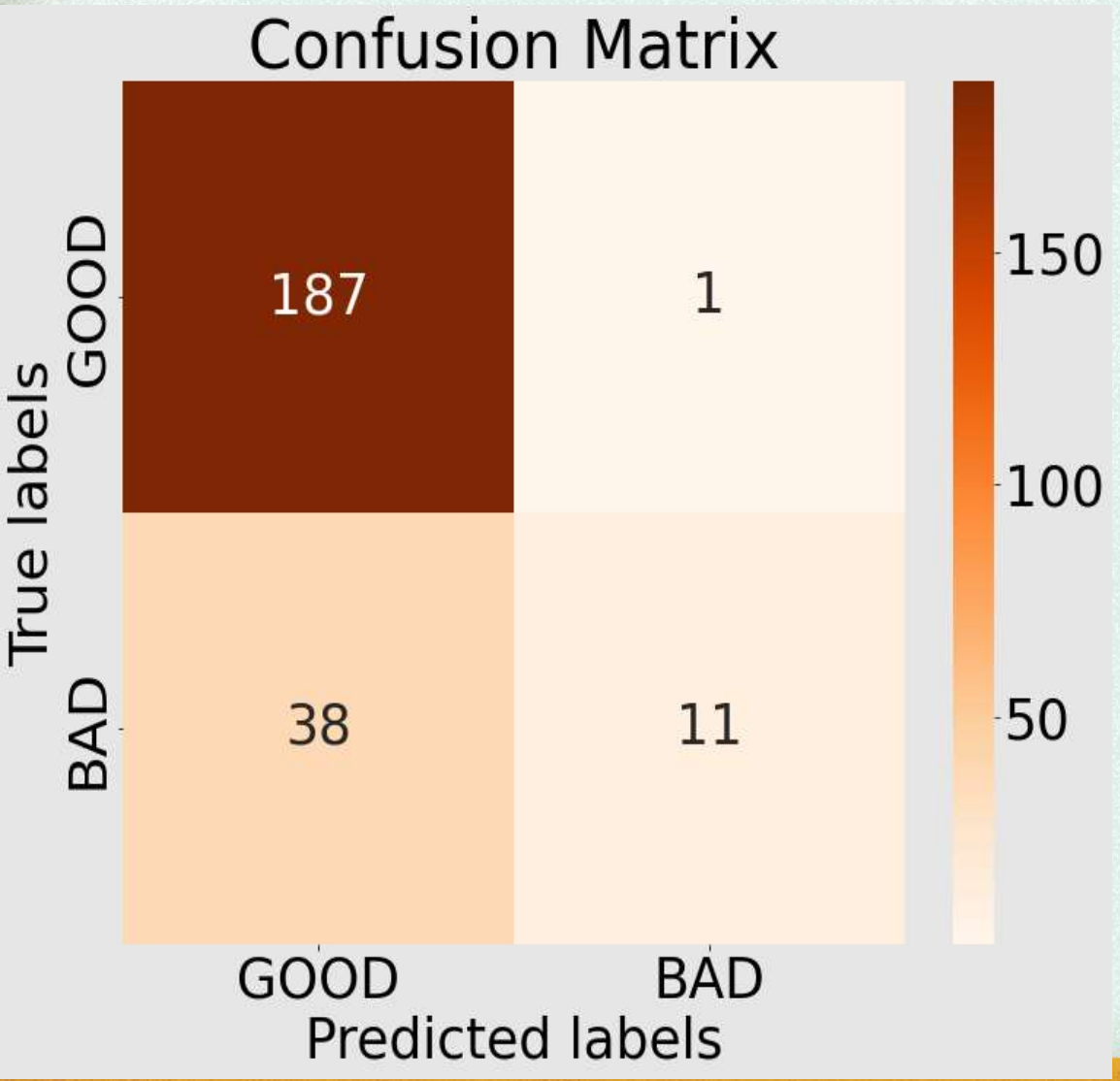
- Benchmark MriqNet against ResNet

Confusion Matrix of MriqNet - Validation Data



- The overall **accuracy** is **~90%.**

Confusion Matrix of MriqNet - Unseen Test Data



- The overall **accuracy** is **~83.5%.**

Conclusion

- MriqNet provides a binary classification of DICOM images with deep learning
  - images transferred to a research node in the edge or cloud with Niffler.
- **Future Work**
  - Expanding to other body parts and modalities.
  - Using real-time and retrospective images, anonymized, from our own clinical PACS.
  - **TCIA images are curated, and finding "bad" images is challenging.**

How to make good tech/research presentations?

Why do people make bad presentations? (1 of 2)

- 1) Lack of preparation.
  - Not enough time, interest, or practice.
- **2) "Creativity."**
  - Creativity is good. But if there is a standard, there could be a reason.
- 3) Language barriers.
  - Needs more time/effort to prepare (and attempting to memorize).
- 4) Not knowing the subject/topic well.
  - I present my (especially first-author or last-author) papers better than when I present someone else's.
- 5) Not timing it correctly.
  - Class presentations have the luxury to carry it forward to the next class.

Why do people make bad presentations? (2 of 2)

- Timezone difference.
  - for conference presentations.
- Tiredness, Lack of sleep, and other external/social factors.
- Not knowing the audience.
  - Boring talk/topic and uninterested audience.
- Shy/nervousness.
  - Fear of judgment/audience.

Some Pointers

- Practice (repeatedly, if you can!).
- How important is this presentation for you?
  - A small class assignment vs. conference presentation vs. career-defining
- Feel free to drop the complex matters out of the presentation.
  - But don't drop everything. Then, your presentation lacks merit.
- Localize the content to fit the audience and place.
- Update the content to fit the time (if the topic/presentation is old!).

How To Present a Paper (book chapter)

Šarūnas Girdzijauskas, KTH

Your Presentations

This presentation assumes you to have 30 – 35 minutes of presentation and totally 45 minutes in total, including Q&A.

- 45 min in total (target 30-35 min for your presentation)

- Student presentations are evaluated on:
  - Slides;
  - Quality of Presentation;
  - Paper understanding;
  - Question Answering.

What to Present?

- **You do not need to retell each and every part of the paper (book chapter)!!!**

- Motivation
- The Ideas/Main Concepts
- Results

- Unveil both positive and negative sides of the paper
  - negative sides are usually well hidden
  - Can be a bit different for a book chapter

How to Present?

- Know Your Audience!!!

- While referring to **a concept** check if…
  - …it is a "**common knowledge**" or needs a special introduction
    - maybe not included in the paper (your book chapter) itself!
  - …it **has been introduced** in the presentation before

Never proceed if you think some concept might not be clear for your audience!

How to Present? (cont.)

- If you think you won't fit into 30-35 min with all your concepts…
- …order them by importance
- …throw the least important away
- …iterate until you fit into 30-35 min ;-)

- **1 slide ~ 2 min, so do not have too many slides!**

- Make few Hidden Slides for Q&A

Slides

- In summary the primary contributions of our work are the following:
  - We introduce the architecture of a scalable RDF database system that leverages best-of-breed single node RDF-stores and parallelizes execution across them using the Hadoop framework.
  - We describe date partitioning and placement techniques that can dramatically reduce the amount of network communication at query time.
  - We provide an algorithm for automatically decomposing queries into parallelizable chunks. We experimentally evaluate our system to understand the effect of various system parameters and compare against other currently available RDF stores.
  - Many of our techniques are applicable to general graph data management applications, especially those for which subgraph matching is an important task.
- Furthermore, while our system is primarily designed for the largest RDF data sets, we will also show that even for smaller data sets, the techniques introduced in this paper are highly relevant.

Slides (alternative)

- Contributions:
  - Scalable RDF database on Hadoop
  - Data Partitioning & Query Decomposition
  - General Applicability; Small & Big Data Sets

Slides (alternative)

- Contributions:
  - **Scalable** RDF database on Hadoop
  - **Data Partitioning** & Query Decomposition
  - General Applicability; **Small & Big Data Sets**

Another (Bad) Example

C. FORMAL DEFINITION OF THE UNDIRECTED N-HOP GUARANTEE

DEFINITION 2. *Let $G = \{V, E\}$ be a graph; and $W \subset V$. P_n, the undirected n-hop guarantee for W, $P_n \subset G$ is defined recursively as follows:*

(1) $P_0 = \{V_0, E_0\}$ where $V_0 = W$ and $E_0 = \emptyset$ is the undirected 0-hop guarantee for W [4].

(2)If $P_n = \{V_n, E_n\}$ is the undirected n-hop guarantee for W, then $P_{n+1} = \{V_{n+1}, E_{n+1}\}$ is the undirected (n+1)-hop guarantee for W where

$$V_{n+1} = \{v | (v, v') \in E \text{ or } (v', v) \in E, v' \in V_n\} \cup V_n,$$

$$E_{n+1} = \{(v, v') | (v, v') \in E, v \in V_n, v' \in V_{n+1}\} \cup \{(v', v) | (v', v) \in E, v \in V_n, v' \in V_{n+1}\} \cup E_n.$$

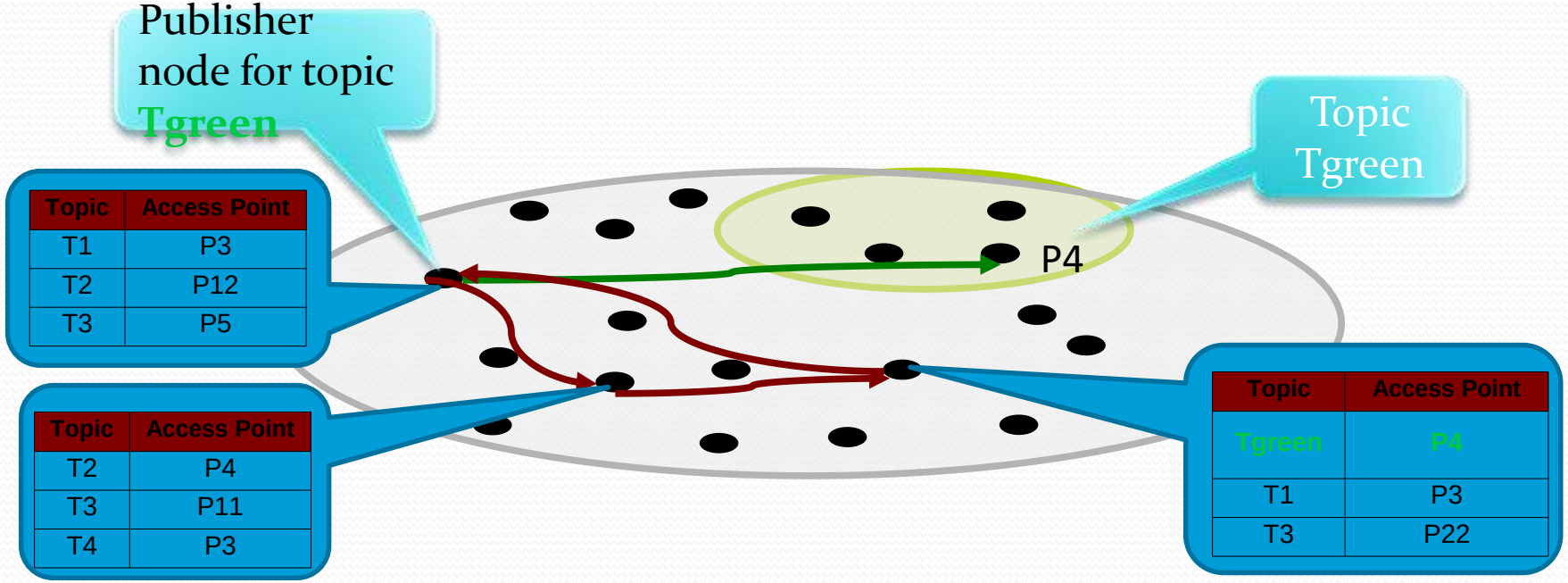
Pseudocodes

Algorithm 4 Select Neighbors

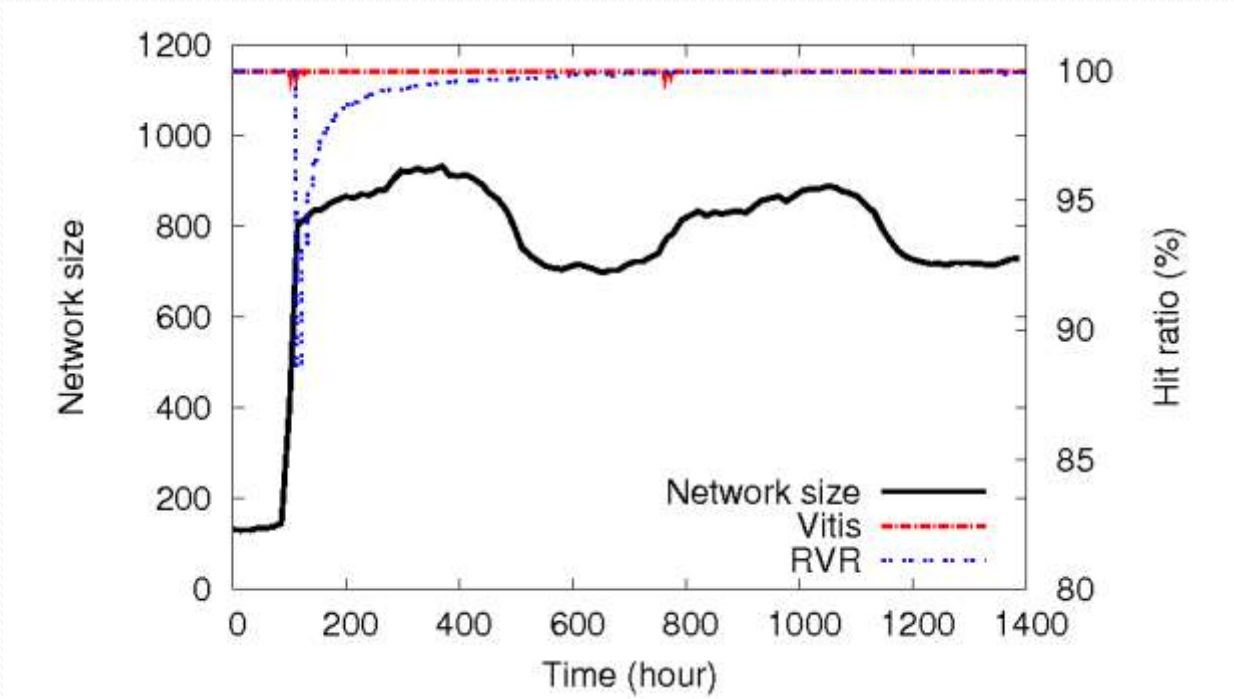
```
1:  procedure SELECTNEIGHBORS(buffer)
2:      successor ← findSuccessor(buffer)
3:      buffer.remove(successor)
4:      selectedNeighbors.add(successor)
5:      predecessor ← findPredeccessor(buffer)
6:      buffer.remove(predecessor)
7:      selectedNeighbors.add(predecessor)
8:      sw-neighbor ← buffer.select-sw-neighbor(RANDOM-DISTANCE)
9:      buffer.remove(sw-neighbor)
10:     selectedNeighbors.add(sw-neighbor)
11:     for all node in buffer do
12:         utility[node] ← calculateUtility(node, self)
13:     end for
14:     sortedNeighbors ← utility[].sort()
15:     friends ← sortedNeighbors.top(RT-SIZE −3)
16:     selectedNeighbors.add(friends)
17:     return selectedNeighbors
18: end procedure
```

Slides (cont)

- Animations as an alternative to text, pseudocodes and formulas



Figures



- **Always** introduce X and Y axes as well as general settings before explaining!

Title Slide

Presented by: Your name
The title of the paper/project you are presenting.
Event: CSCE A490 Class Activity.
Date: DATE.

Agenda

!SKIP!

Introduction

The global state of matters.

- UberEats.
- The world is still hungry.

- There are clouds everywhere. Still, we have high latency!

UAA 2025

State of the art

Other people must be doing something about this issue. Right?

Motivation

Do you think you have a better approach?

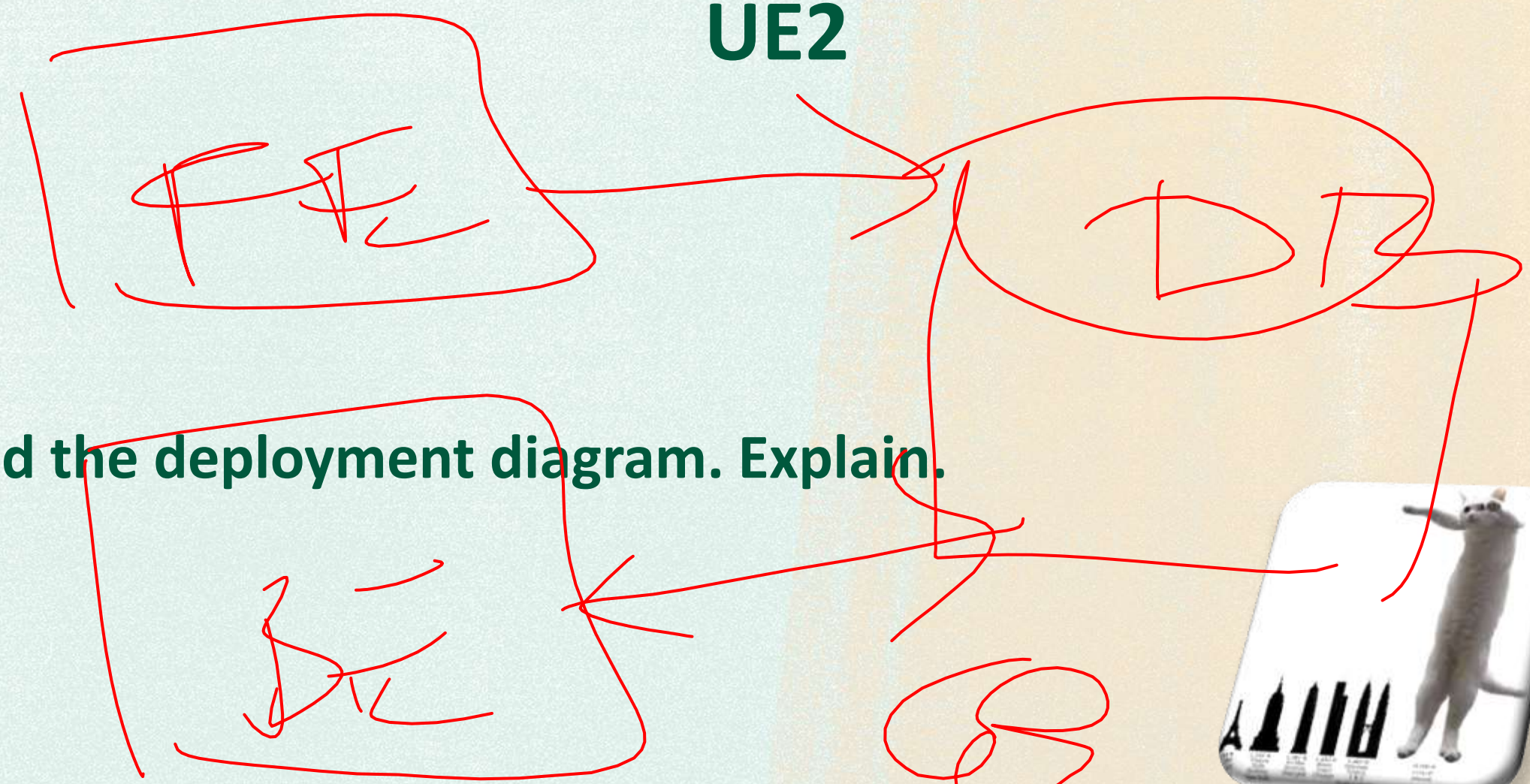
Solution

UE2!

Or L4S, …., cloud-to-edge continuum, community cloud, … whatever the paper is proposing.

UAA 2025

UE2



Add the deployment diagram. Explain.

UE2

Add the major algorithm/brain. Explain.

UE2

Add the use cases

UE2

Add the architecture diagram and explain the implementation.

UE2

Prototype Implementation and Evaluation. Add the plots.

UE2

Discussion. Why your solution is better (or justify why it is not better and still deserve publication).

UAA 2025

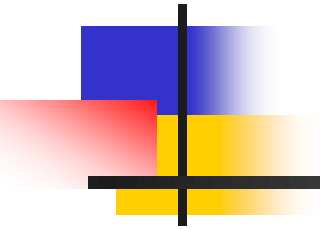
UE2

Conclusion and future work

UE2

Your own reflections, if this is someone else's published work.

How about some example bad presentations?



Chicken Chicken Chicken: Chicken Chicken

Doug Zongker
University of Washington

Chicken

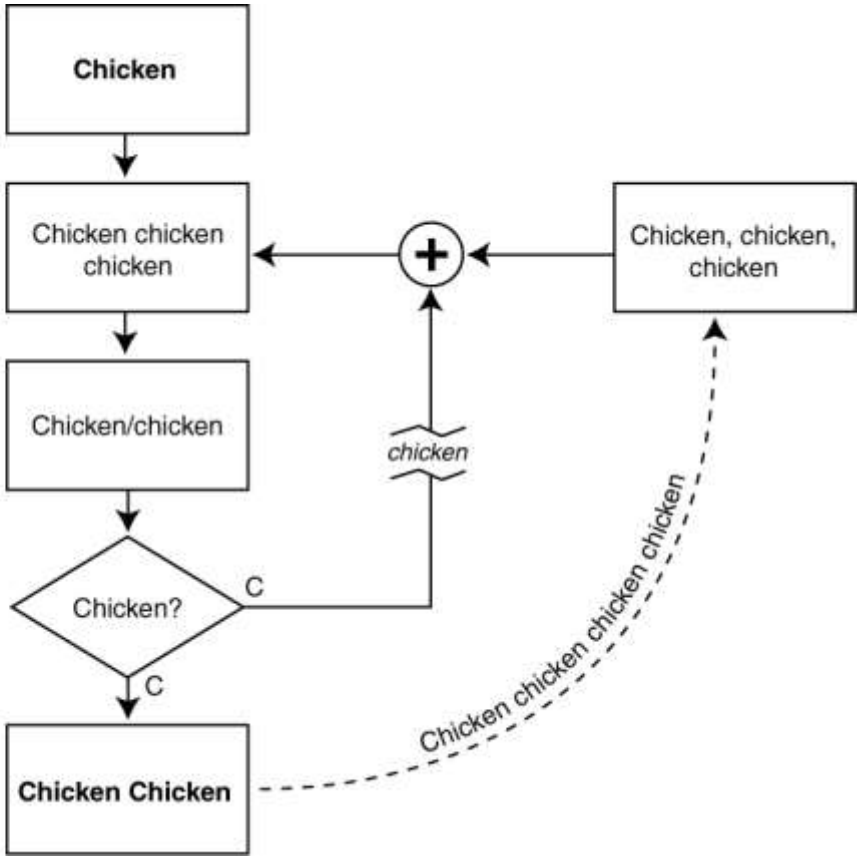
- **Chicken**

Chicken chicken

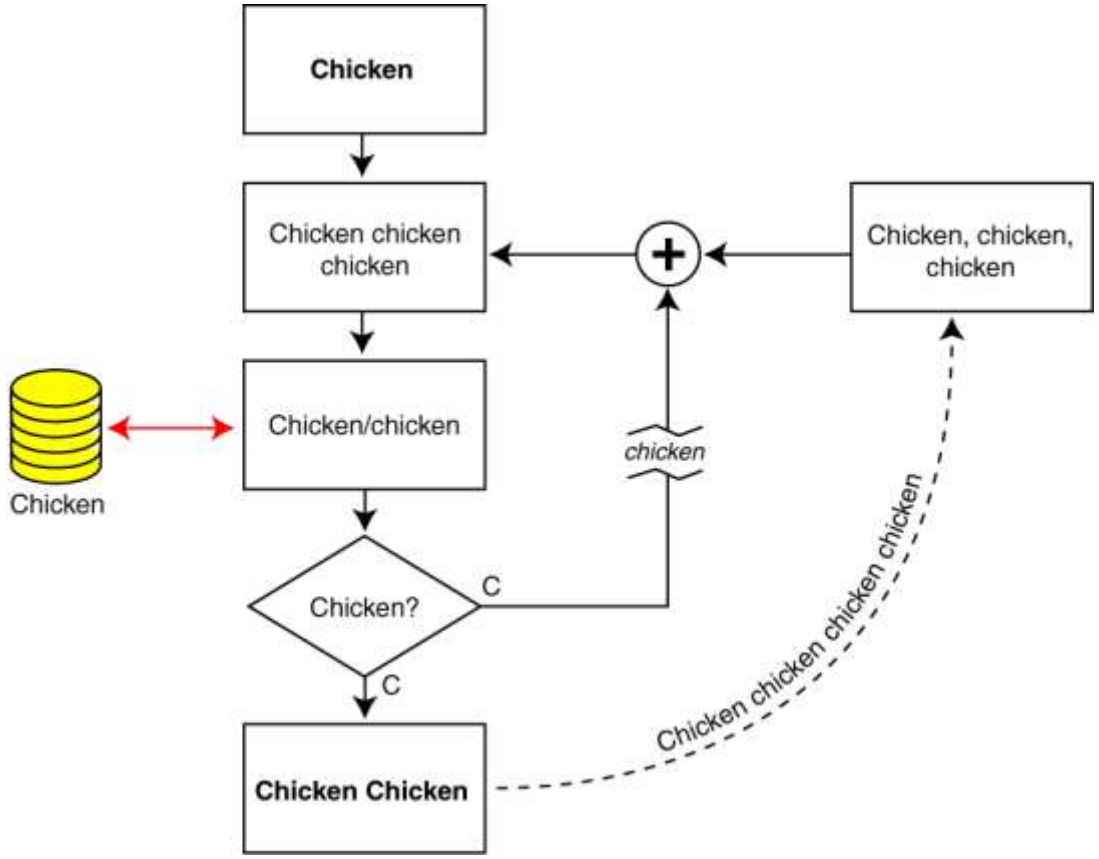
Chicken

Chicken chicken chicken

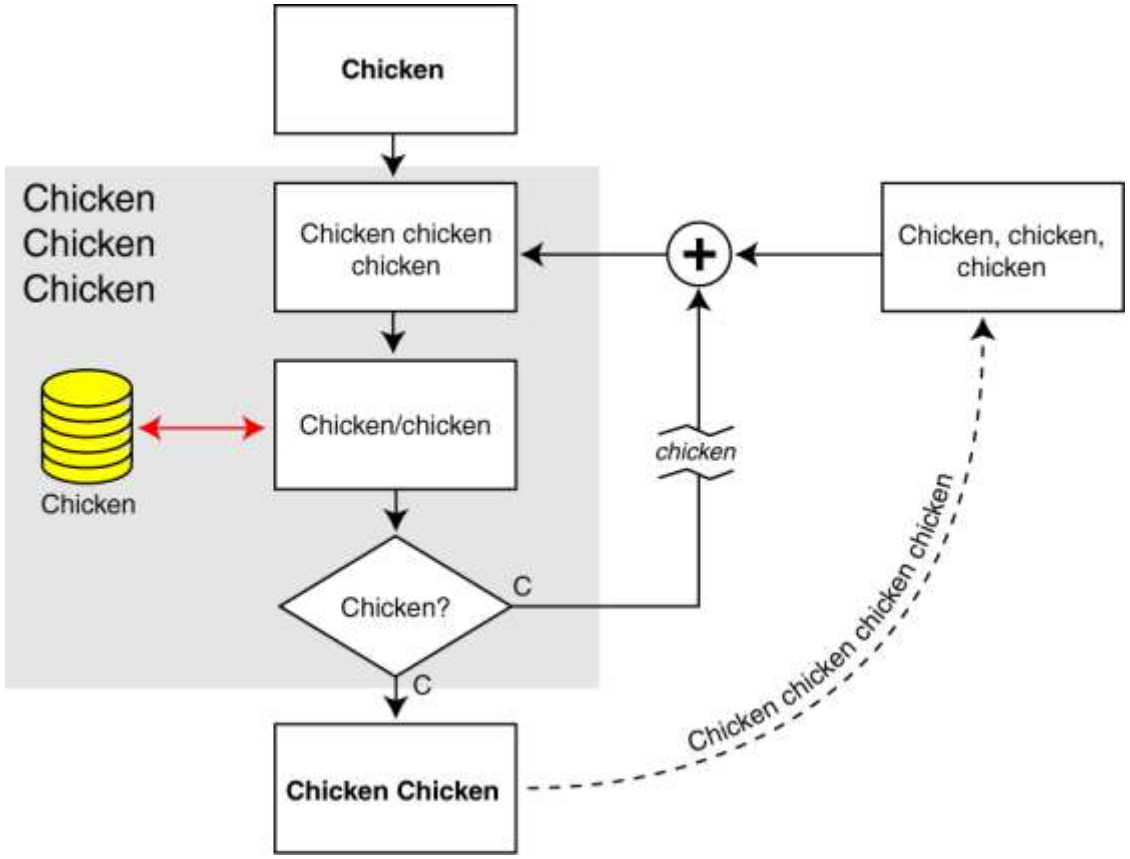
Chicken chicken



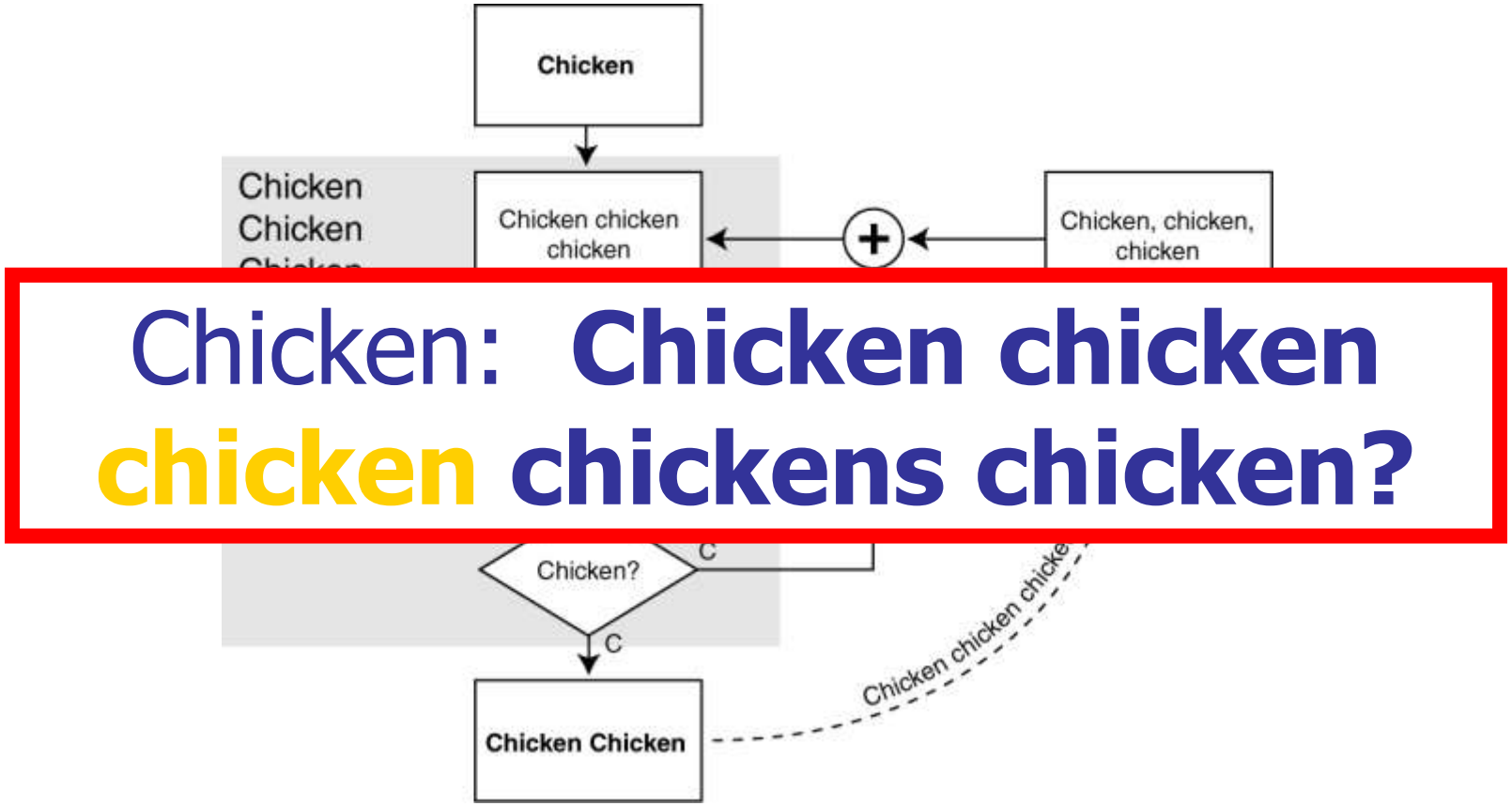
Chicken chicken



Chicken chicken



Chicken chicken



Chicken: **Chicken chicken chicken chickens chicken?**

Chicken

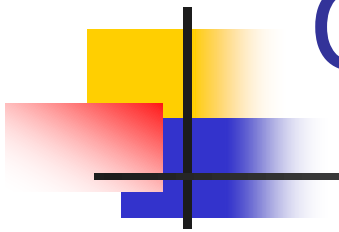
Chicken

- **Chicken chicken**

Chicken

Chicken chicken chicken

Chicken chicken



- Chicken chicken/chickens
  - Chicken-chicken [Ch ’95]
  - “Chickens” chickens [C&C ’97a]
  - Chickens chickens chickens chickens [Ch ’00]
- Chickens chickens chickens
  - Chickens/chickens [Ch ’01]
  - Chicken [C. Ch ’97b]

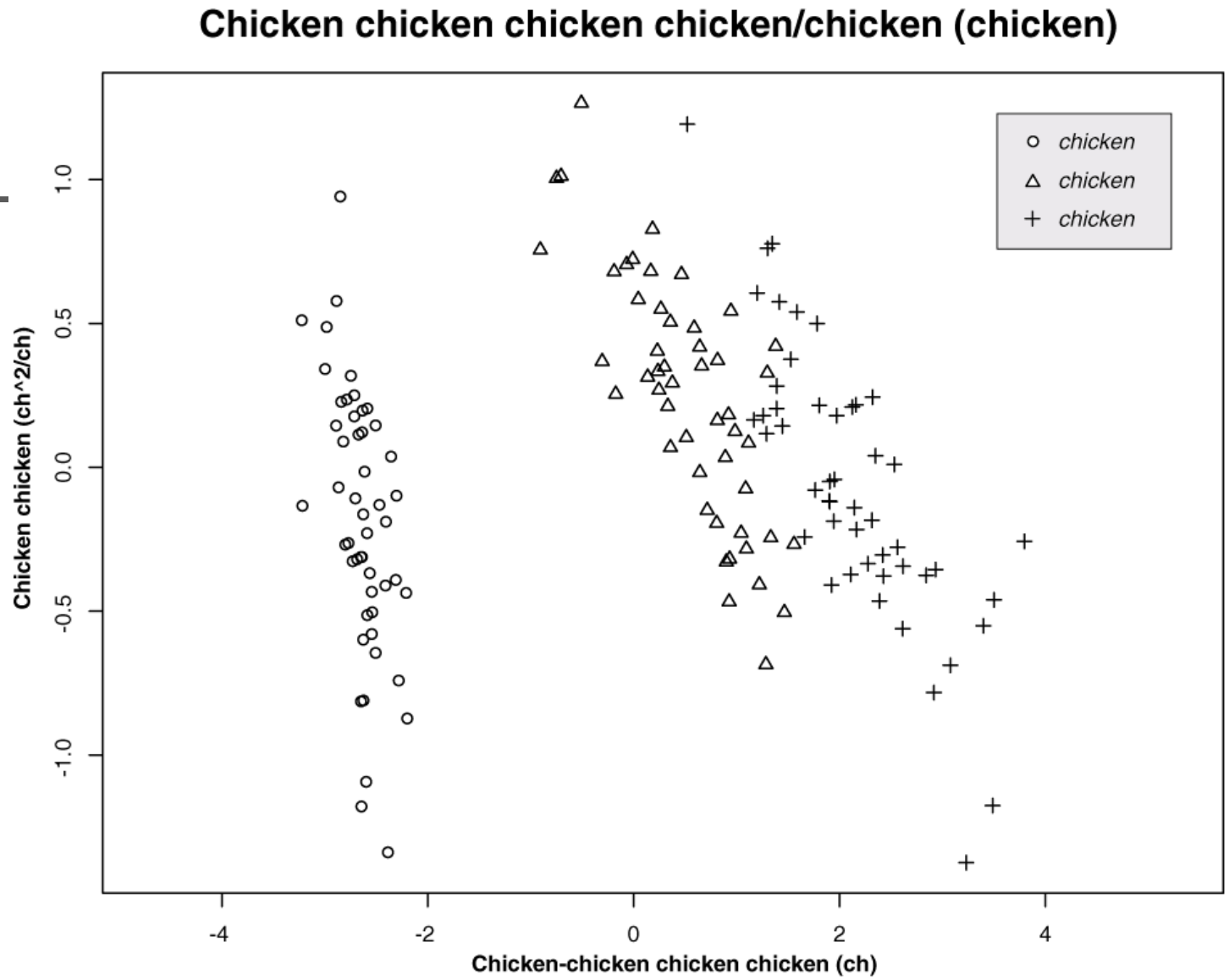
Chicken

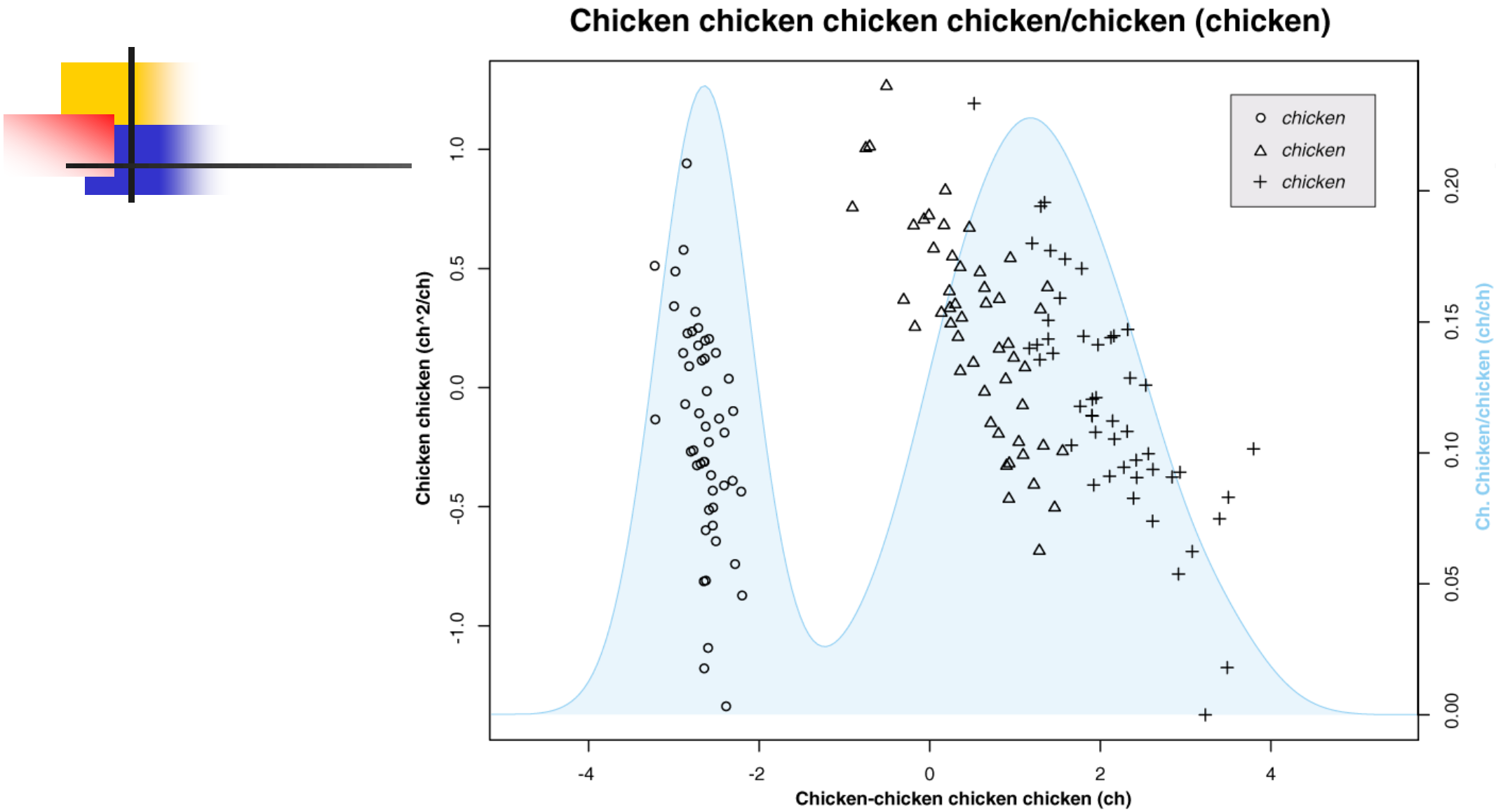
Chicken

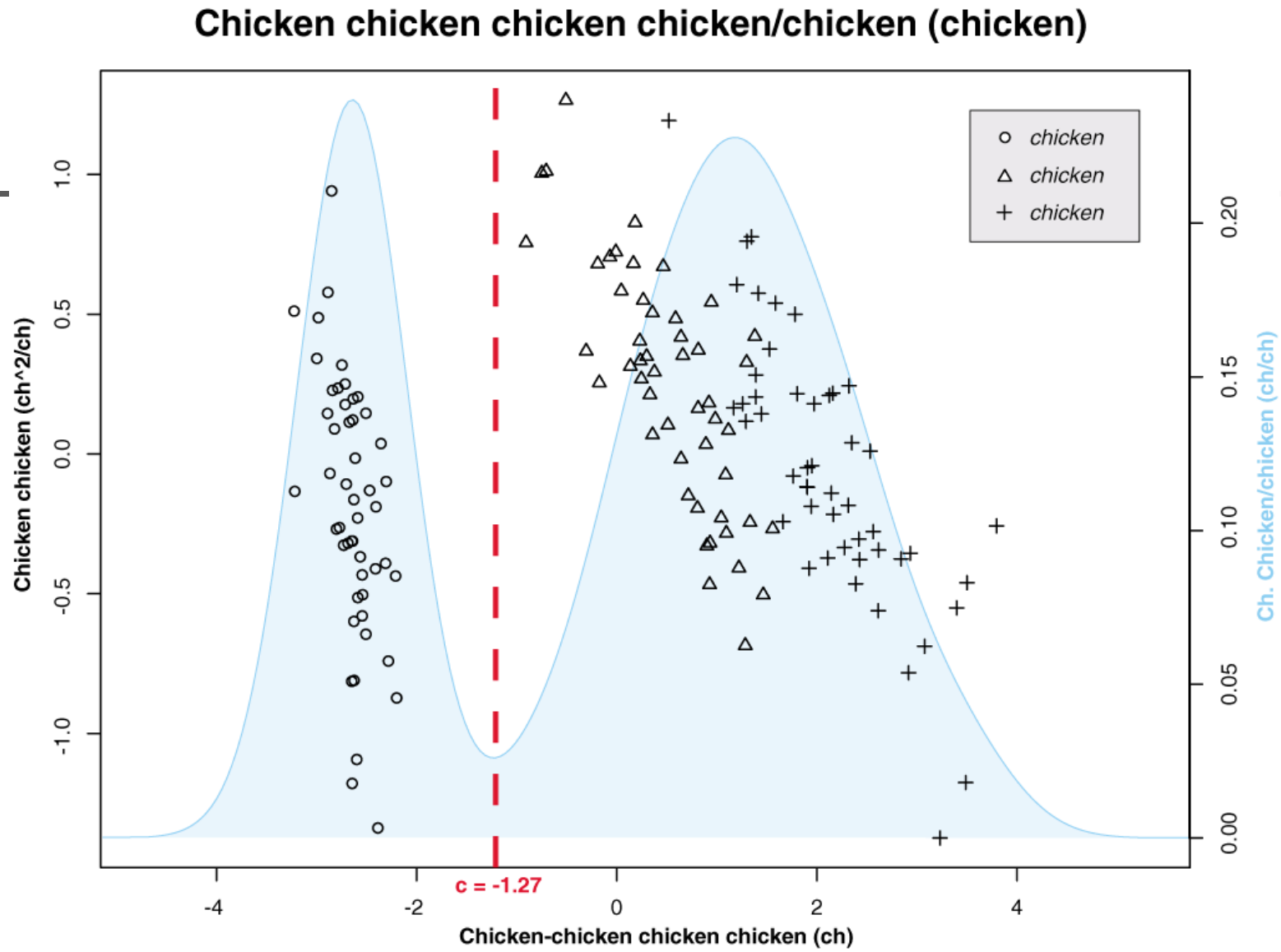
Chicken chicken

- **Chicken**

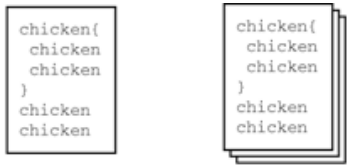
Chicken chicken chicken

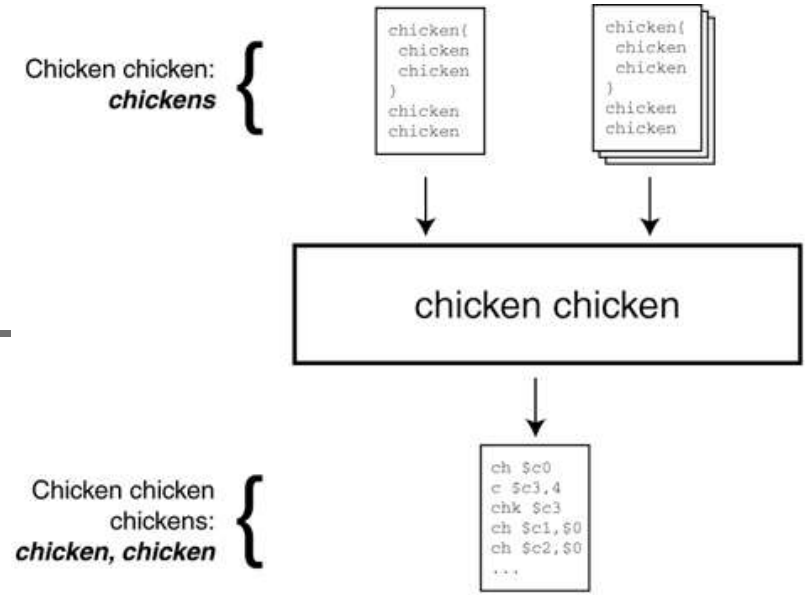

Chicken chicken chicken chicken/chicken (chicken)
chicken
chicken
chicken
Chicken chicken (ch^2/ch)
1.0
0.5
0.0
-0.5
-1.0
-4
-2
0
2
4
Chicken-chicken chicken chicken (ch)

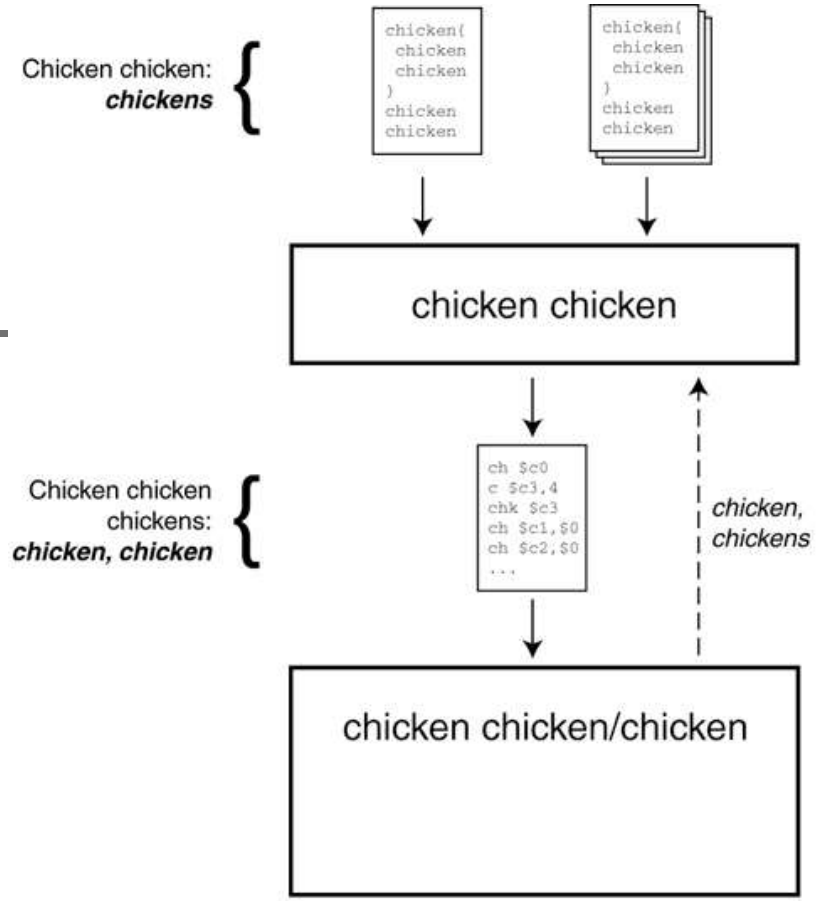

Chicken chicken chicken chicken/chicken (chicken)
chicken
chicken
chicken
Chicken chicken (ch^2/ch)
Ch. Chicken/chicken (ch/ch)
Chicken-chicken chicken chicken (ch)
1.0
0.5
0.0
-0.5
-1.0
0.20
0.15
0.10
0.05
0.00
-4
-2
0
2
4

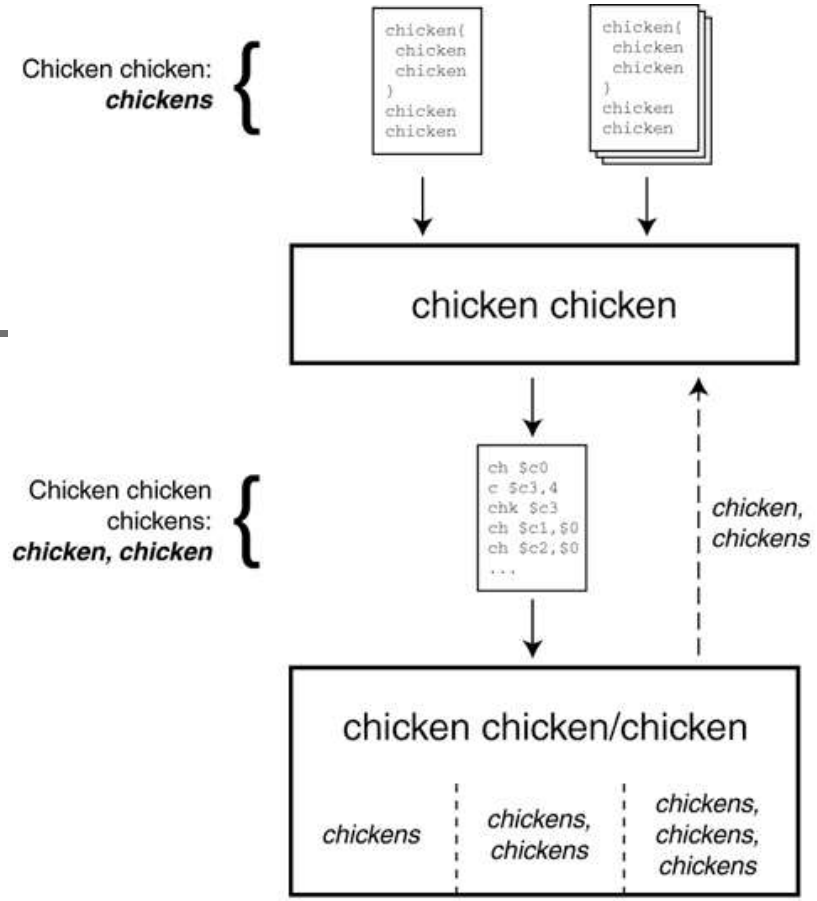

Chicken chicken chicken chicken/chicken (chicken)
chicken
chicken
chicken
Chicken chicken (ch^2/ch)
Ch. Chicken/chicken (ch/ch)
c = -1.27
Chicken-chicken chicken chicken (ch)
1.0
0.5
0.0
-0.5
-1.0
0.20
0.15
0.10
0.05
0.00
-4
-2
0
2
4

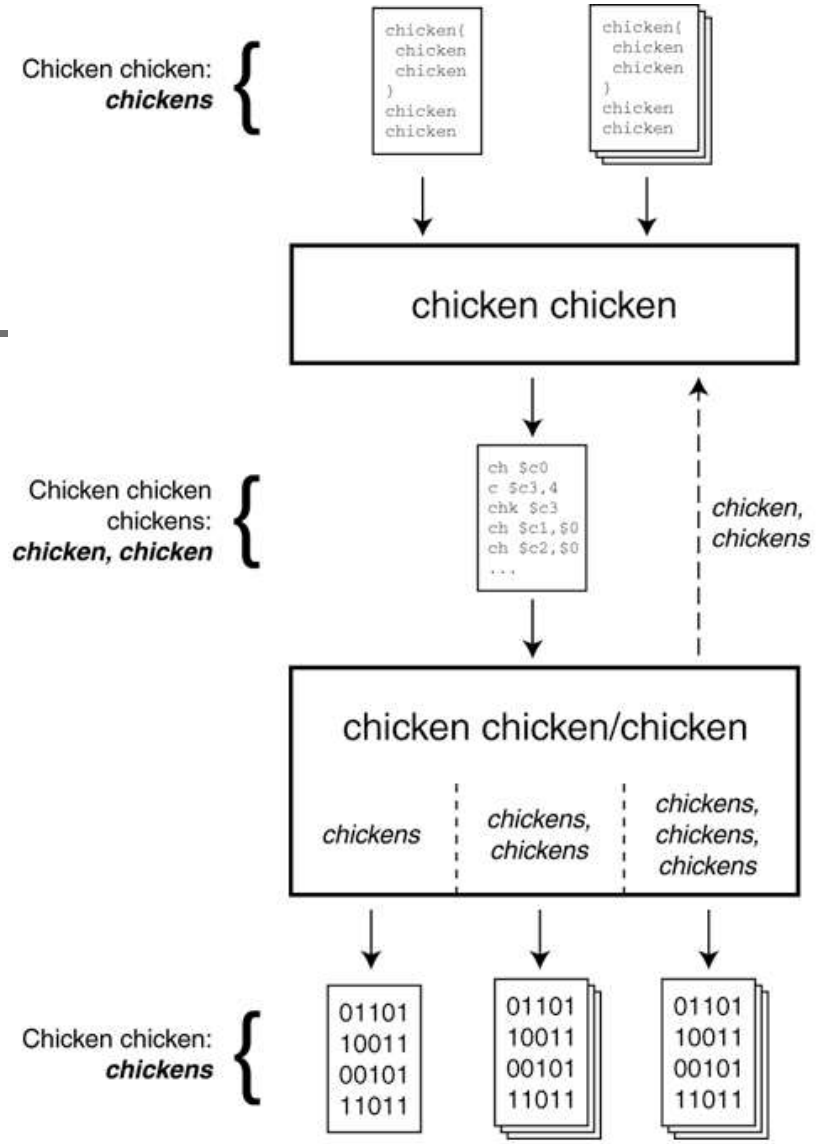
Chicken chicken:
chickens

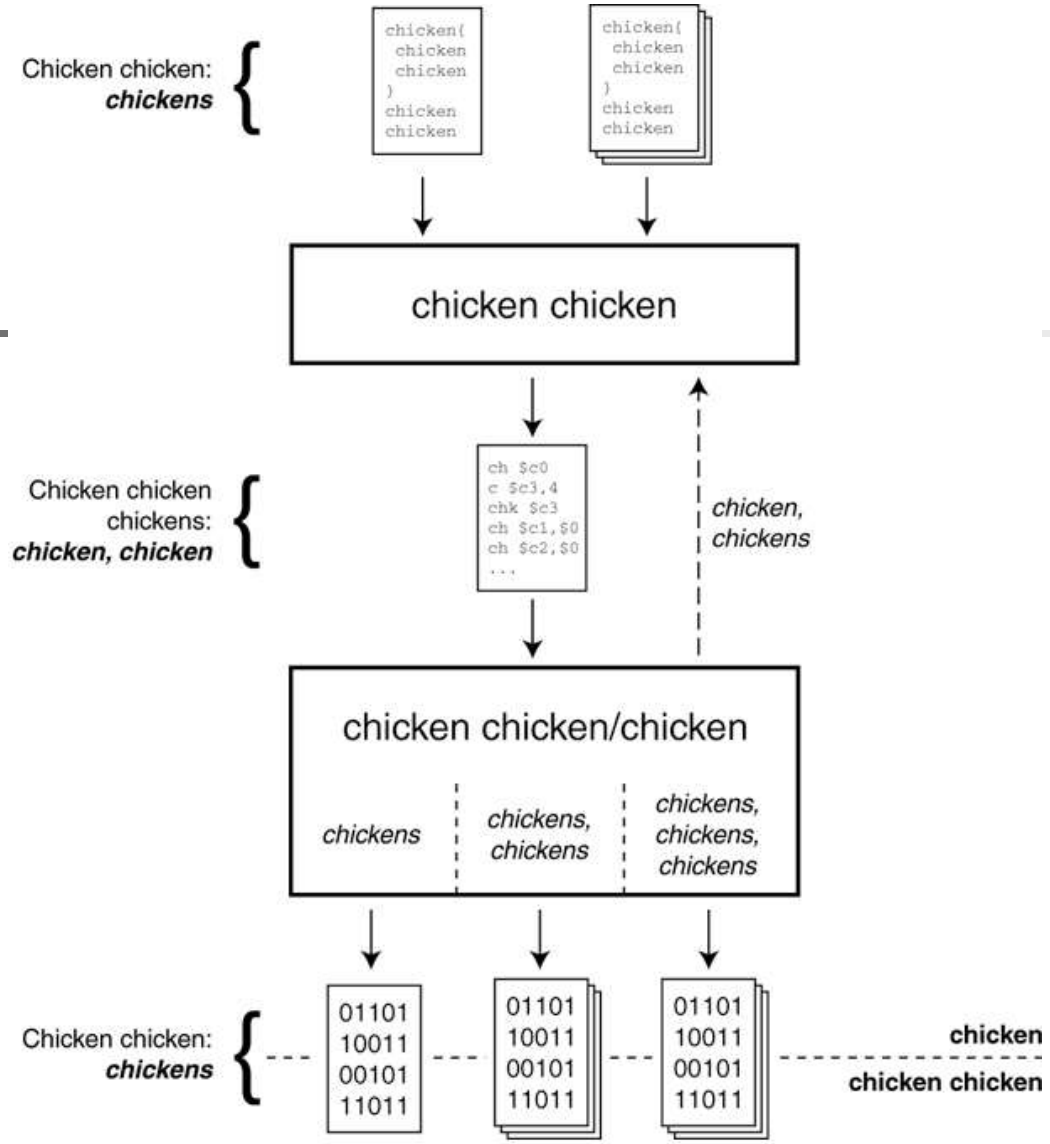


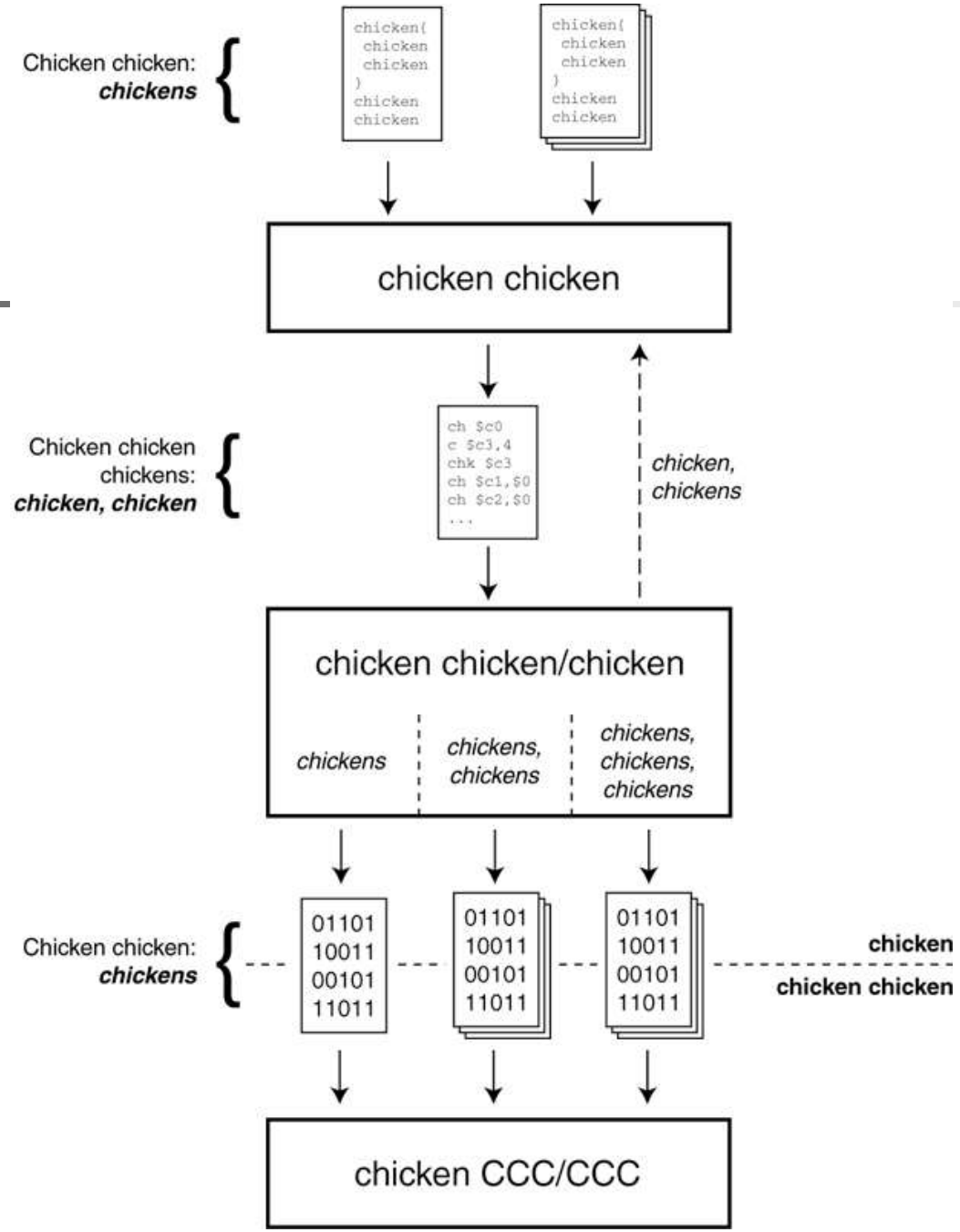

Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...

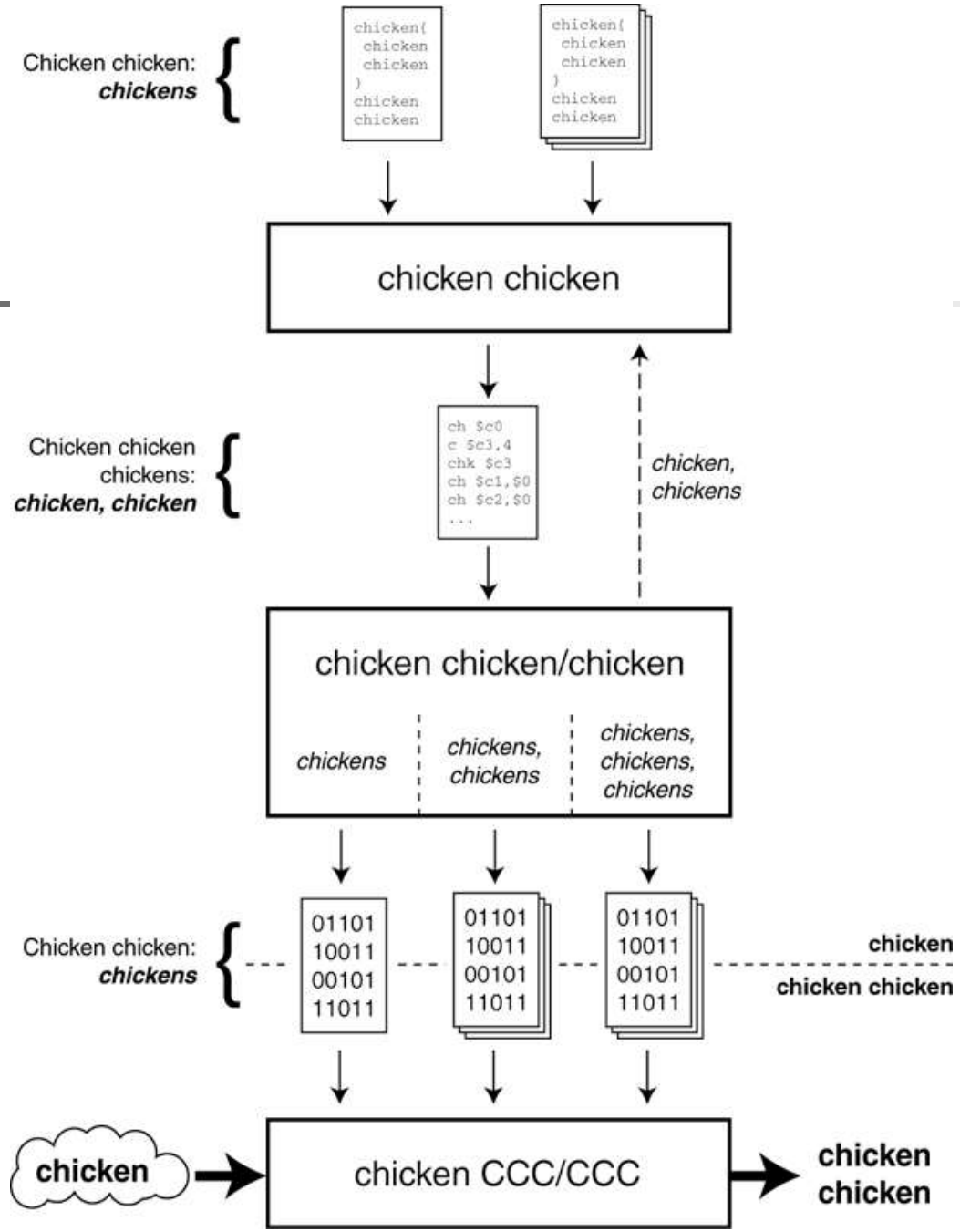

Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...
chicken,
chickens
chicken chicken/chicken


Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...
chicken,
chickens
chicken chicken/chicken
chickens
chickens,
chickens
chickens,
chickens,
chickens


Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...
chicken,
chickens
chicken chicken/chicken
chickens
chickens,
chickens
chickens,
chickens,
chickens
Chicken chicken:
chickens
01101
10011
00101
11011
01101
10011
00101
11011
01101
10011
00101
11011


Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...
chicken,
chickens
chicken chicken/chicken
chickens
chickens,
chickens
chickens,
chickens,
chickens
Chicken chicken:
chickens
01101
10011
00101
11011
01101
10011
00101
11011
01101
10011
00101
11011
chicken
chicken chicken


Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...
chicken,
chickens
chicken chicken/chicken
chickens
chickens,
chickens
chickens,
chickens,
chickens
Chicken chicken:
chickens
01101
10011
00101
11011
01101
10011
00101
11011
01101
10011
00101
11011
chicken
chicken chicken
chicken CCC/CCC


Chicken chicken:
chickens
chicken{
chicken
chicken
}
chicken
chicken
chicken{
chicken
chicken
}
chicken
chicken
chicken chicken
Chicken chicken
chickens:
chicken, chicken
ch $c0
c $c3,4
chk $c3
ch $c1,$0
ch $c2,$0
...
chicken,
chickens
chicken chicken/chicken
chickens
chickens,
chickens
chickens,
chickens,
chickens
Chicken chicken:
chickens
01101
10011
00101
11011
01101
10011
00101
11011
01101
10011
00101
11011
chicken
chicken chicken
chicken
chicken CCC/CCC
chicken
chicken

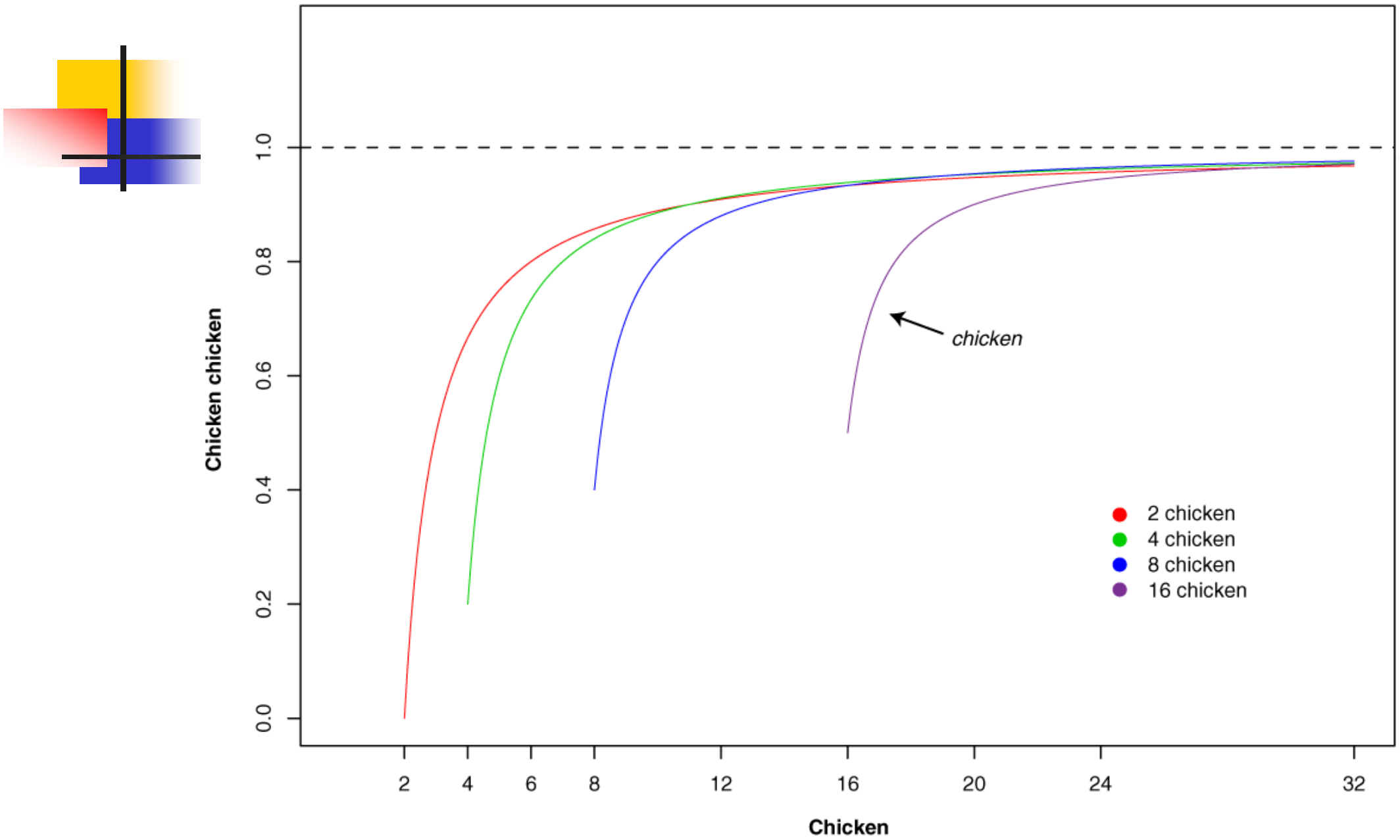
Chicken

Chicken

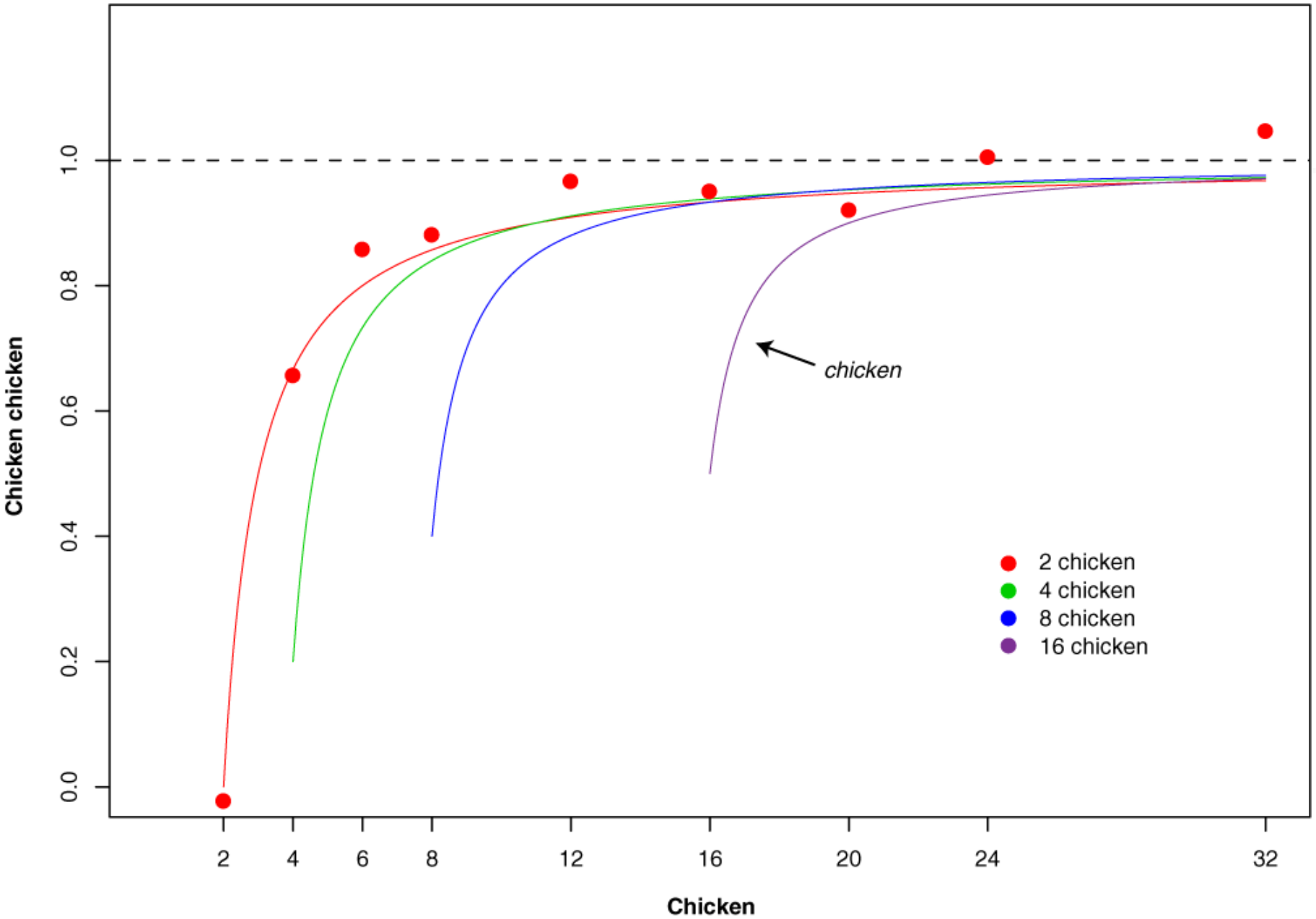
Chicken chicken

Chicken

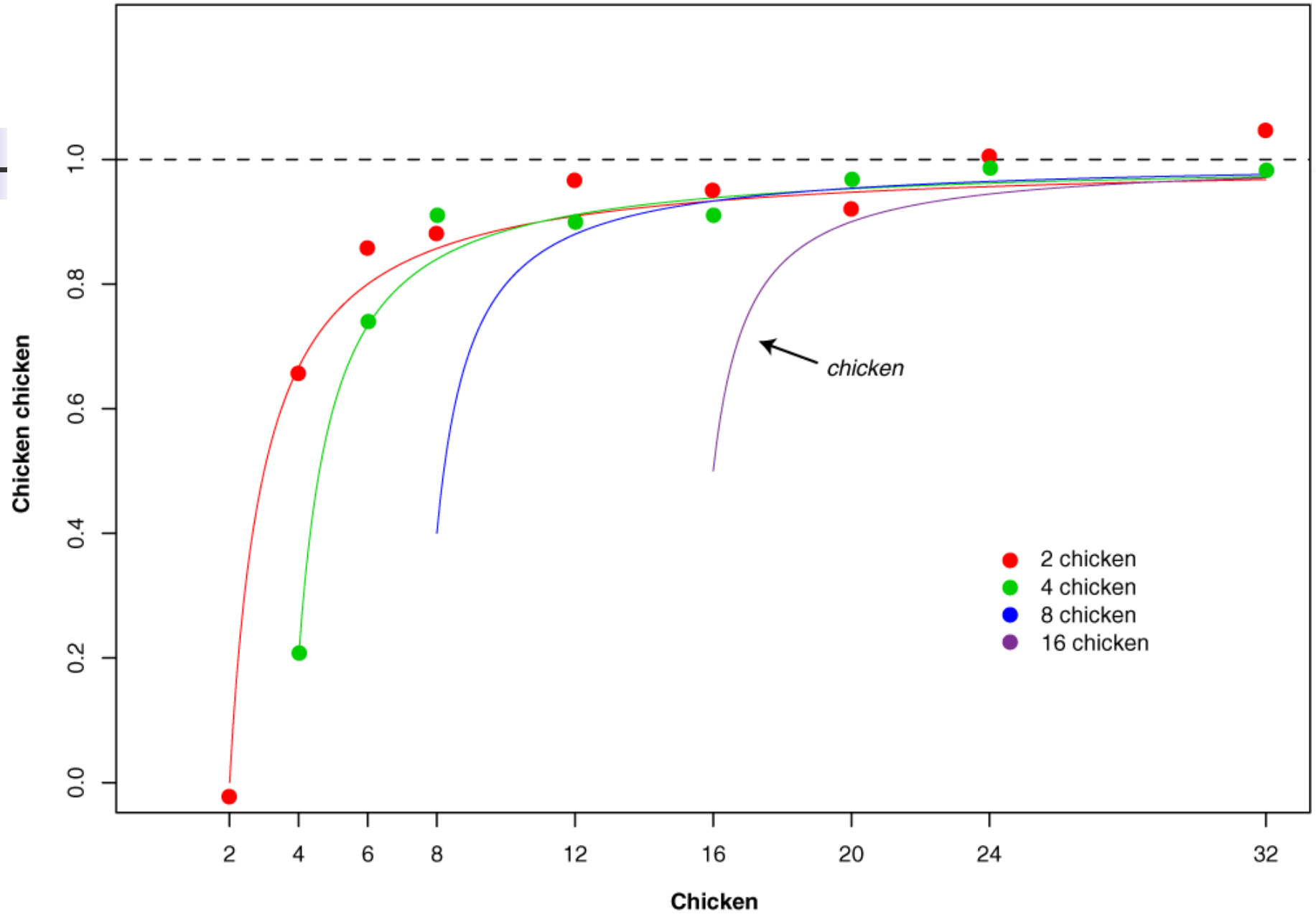
- **Chicken chicken chicken**


Chicken chicken chicken, chicken chicken
Chicken chicken
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
12
16
20
24
32
Chicken
chicken
2 chicken
4 chicken
8 chicken
16 chicken

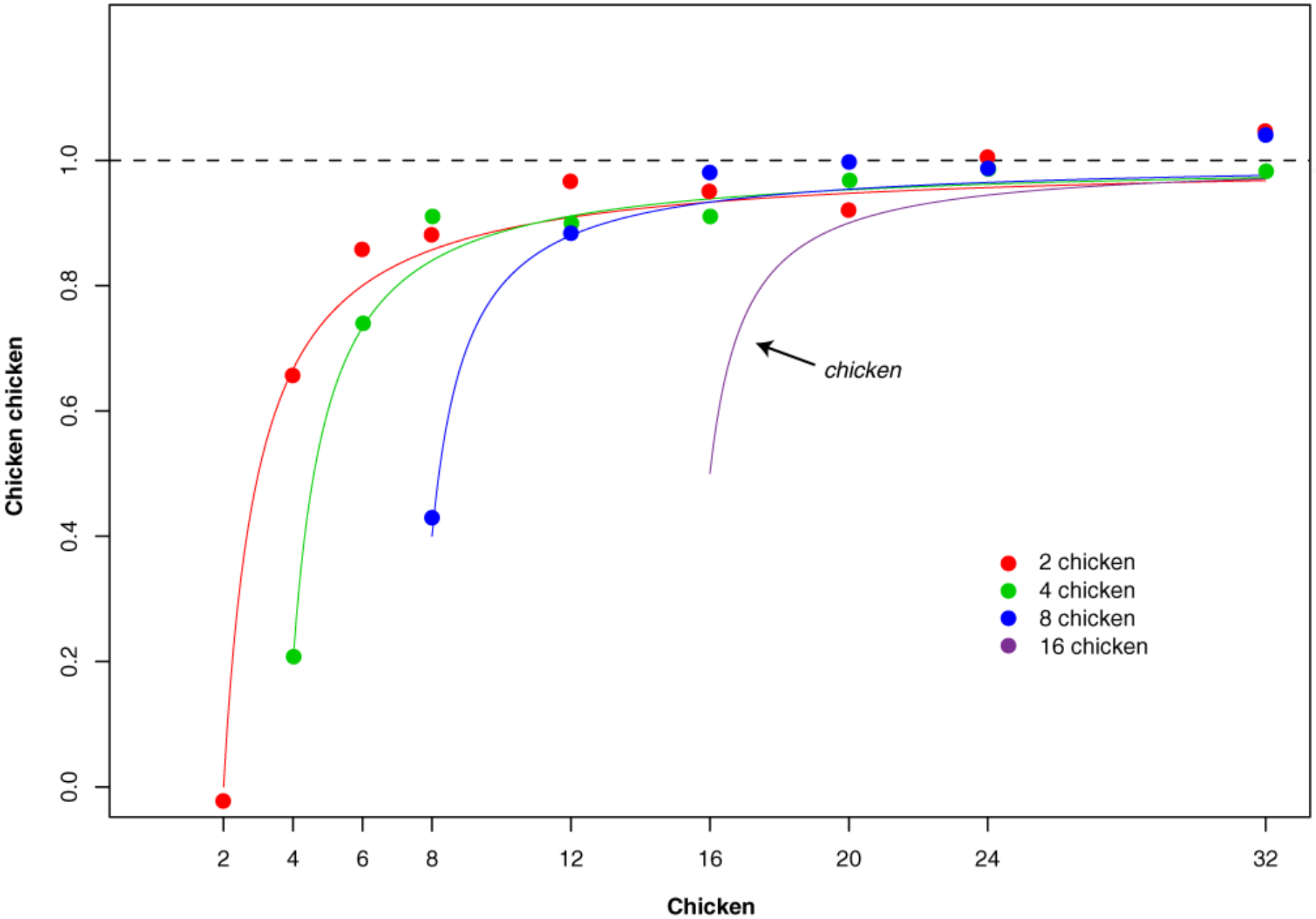
Chicken chicken chicken, chicken chicken



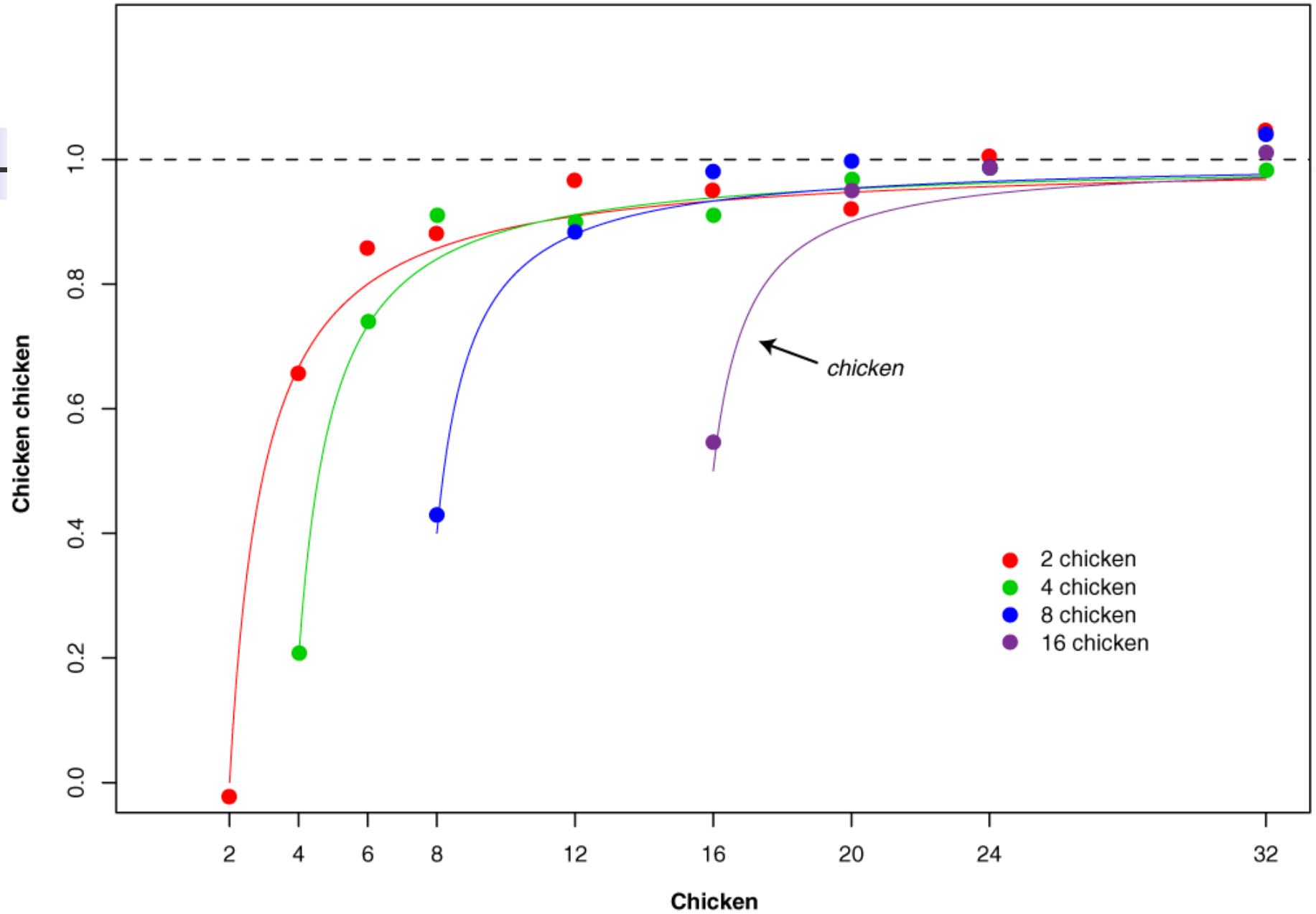
Chicken chicken chicken, chicken chicken



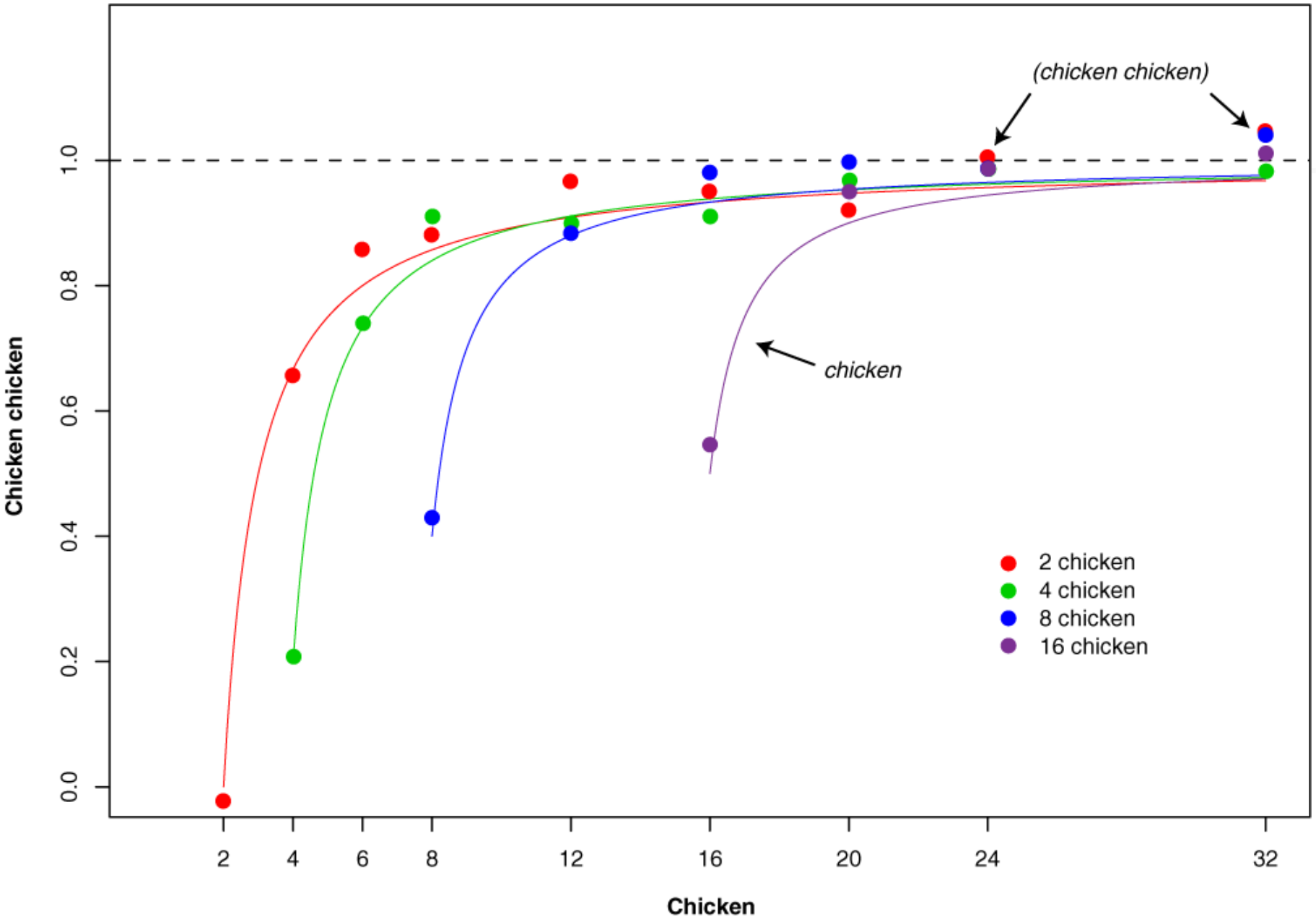
Chicken chicken chicken, chicken chicken



Chicken chicken chicken, chicken chicken



Chicken chicken chicken, chicken chicken



Chickens

- Chicken chicken:
  - Chickens chickens
  - Chicken
  - Chicken C-1049355
- Chicken chicken chickens:
  - cccc://chicken.chicken.chk/~chickens/

Chicken chicken, chicken chicken.

Chickens chickens chicken

- Chickens chicken:

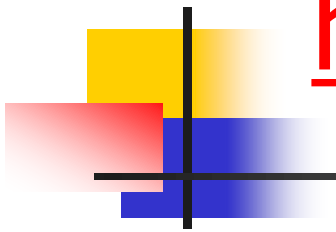
$$C(K) = \sum_{i=1}^{n} \Delta^2(K_i) = \sum_{i=1}^{n} ||E_i - K(H_i)||^2$$

chicken:

$$(c, h, i) \in \mathbf{S} \quad \longrightarrow \quad K(c, h, i) = \begin{bmatrix} K_i(c, h, i) \\ K_e(c, h, i) \end{bmatrix}$$

$$K(E) = \lambda_1 \mathbf{k}_c + \lambda_2 \mathbf{k}_{i_n} + \lambda_3 \mathbf{k}_{i_{n+1}}$$

Read the paper at https://isotropic.org/papers/chicken.pdf



- Find the recording at https://www.youtube.com/watch?v=yL_-1d9OSdk

- Find the full story at

https://news.cs.washington.edu/2013/08/14/chicken-chicken-chicken-chicken/

Another example?

Death by PowerPoint

Pradeeban Kathiravelu ප්‍රදීබන් කදිරවේලු

Xiao Chen 陈潇

Cantopop

粵語流行音樂


TWINS - 恋爱大过天 (KTV)
Cantonese Popular Music
Cantopop (Chinese: 粵語流行音樂) is a contraction of "Cantonese popular music". It is sometimes referred to as HK-pop, short for "Hong Kong popular music". The term "Cantopop" has its origin in "Cantorock", a term first used in 1974 to describe rock music in Hong Kong by Billboard correspondent Hans Ebert. Later in 1978 Ebert revised the term to "Cantopop" after noting a change in its style to something similar to British-American soft-rock. It now describes a contemporary category of popular music made primarily in Hong Kong in the Cantonese language since the 1970s, and the cultural context of its production and consumption.
想不想也日夜懷念
戀愛大過天
2:34 / 3:45

Cantopop Globally!

- 音樂作品新嘗試

- 邁進2010年，多位歌手推出以新主題及新曲風的唱片，對於粵語流行音樂市場是一種新的突破和嘗試。其中，樂壇天后鄭秀文復出後的首張專輯大碟大膽地打破以往的情歌以及電子舞曲路線，改以流行福音歌曲作為主題，為香港樂壇史上少有的並且是最成功和暢銷的福音專輯，成功為福音唱片於香港及亞洲的主流音樂市場開創了先河。同年亦有[illegible]，發燒音

Cantopop vs Mandopop

Two main categories of C-Pop.

- Cantopop is *mostly* **Hong Kong*** based.
- Mandopop is everywhere***

* May be Guangzhou and Shenzhen too? I don't know.

*** Xiao told me so.

Public Opinion: “Do you like Cantopop, Mandopop, or K-Pop?”

Cantopop:

- Xiao: “sometimes I will listen to Cantopop”
- Qi Qi: “What's Cantopop?!”

Mandopop:

- Leo: “It's different, cool. Interesting!”

K-Pop:

- Leo: “not really!”
- Dipesh: “oh .. i never heard”

Conclusion

- Cantopop is confirmed to be most popular among the global citizens according to the survey.

Thank you

- Questions?
- pkathiravelu@alaska.edu