

DATASAIL

Kick-off Meeting

Thursday, November 5th, 2015

Piotr Szczurek, Ph.D.

Assistant Professor

Director of Master of Science in Data Science

Lewis University

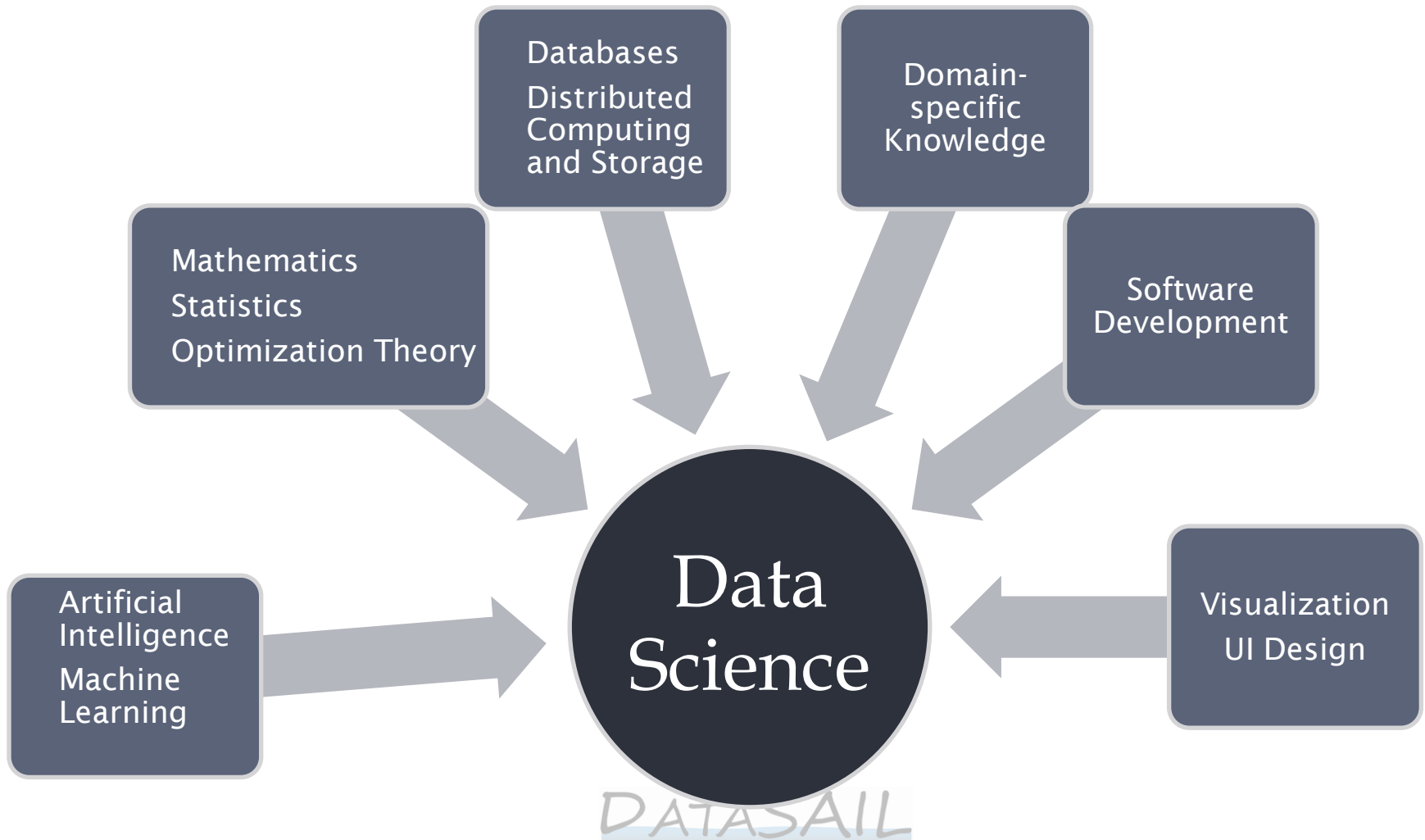
Agenda

1. Introduction: mission and vision
2. Faculty introductions
3. Projects
4. Why and how to join?
5. Next steps

1. Introduction: Mission and Vision

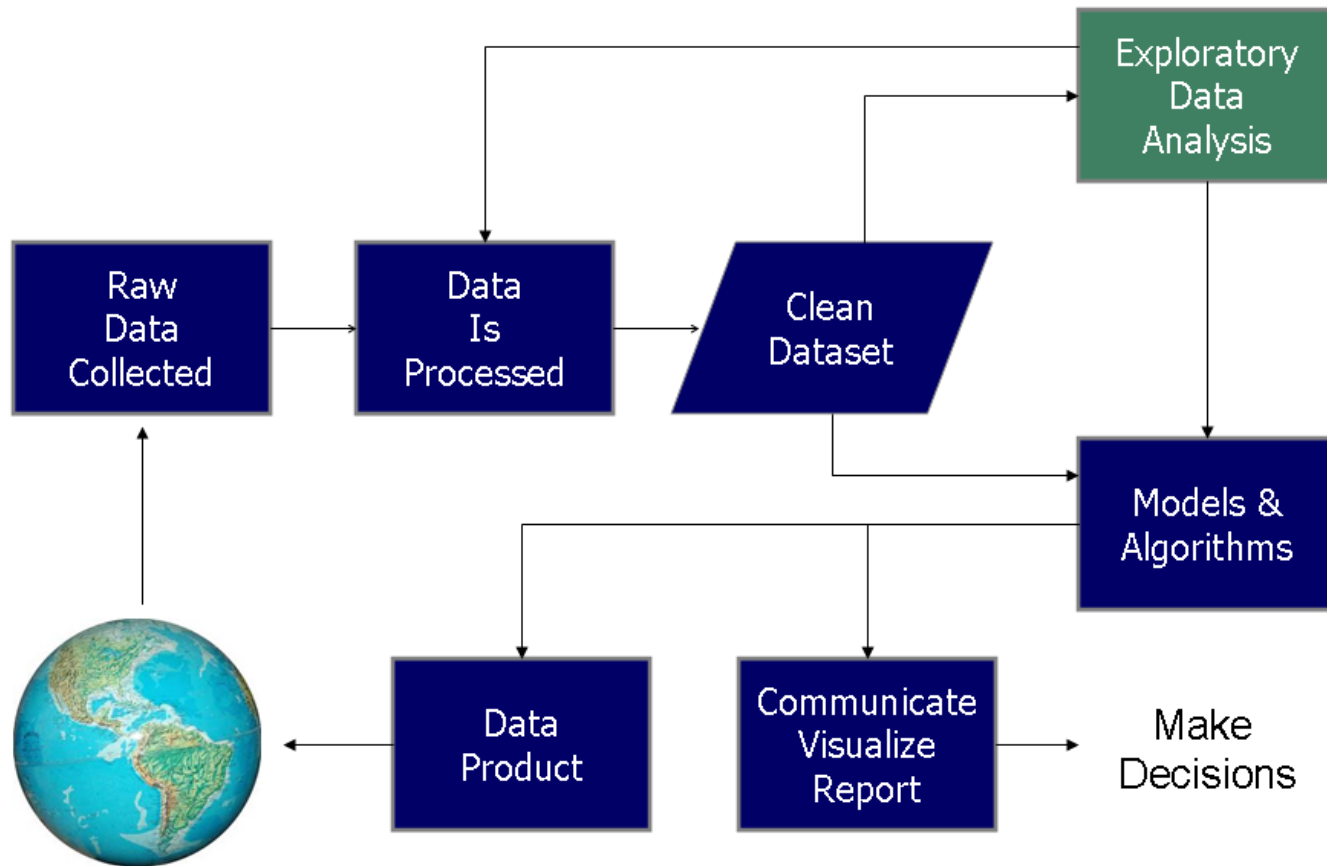
What is Data Science?

“**Data science** is the study of the generalizable extraction of knowledge from data” (Dhar, V. (2013). Data science and prediction. *Commun. ACM*, 56, 64-73.)



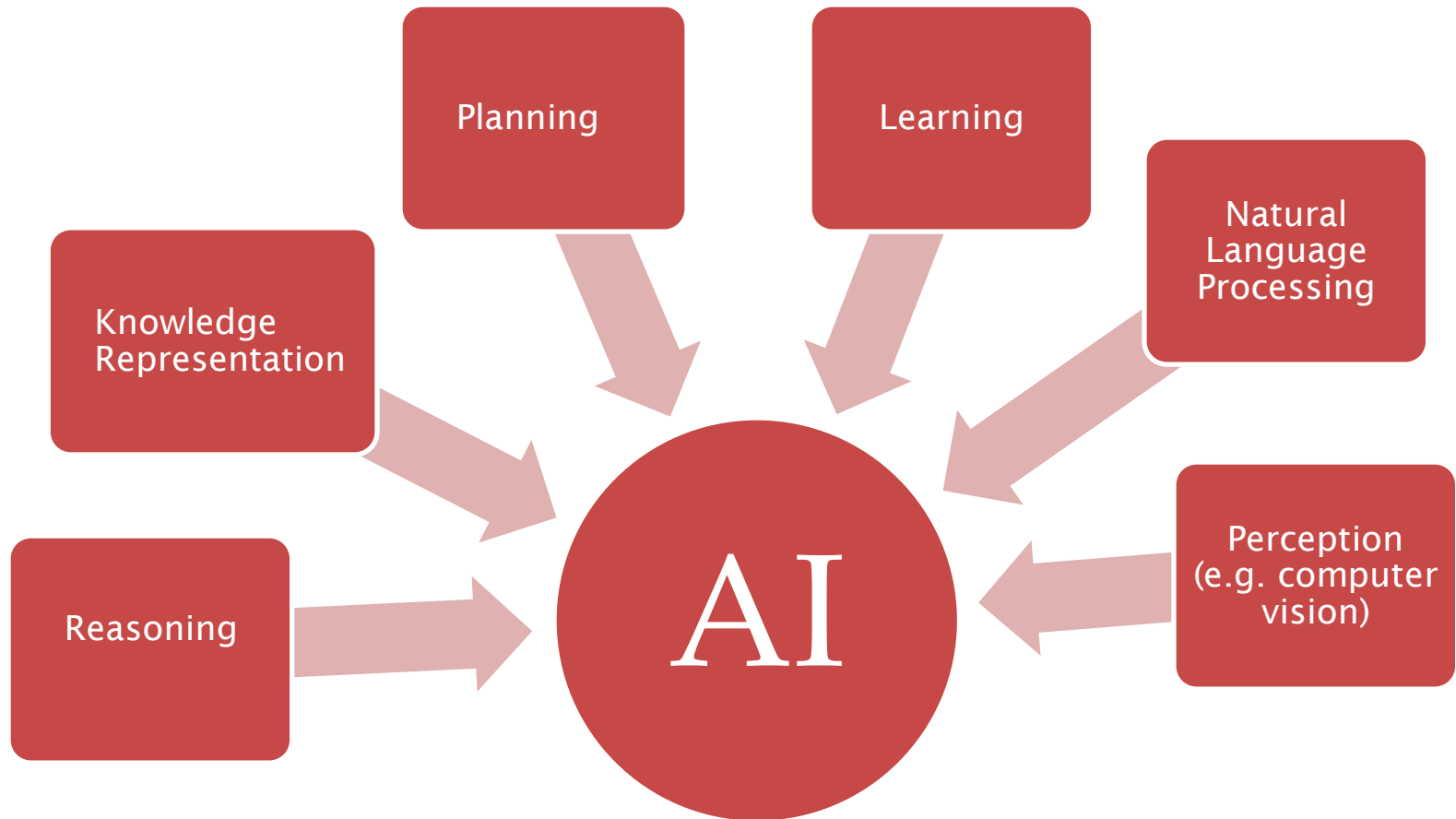
What is Data Science?

Data Science Process



What is Data Science?

Artificial Intelligence (AI) is the study of designing intelligent agents.



Mission Statement

The mission of the **Data Science** and **Artificial Intelligence** Laboratory (**DataSAIL**) is

to help foster the collaboration of students and faculty members to work on data science or artificial intelligence related problems,

which are of importance to the society, the community, or the University.

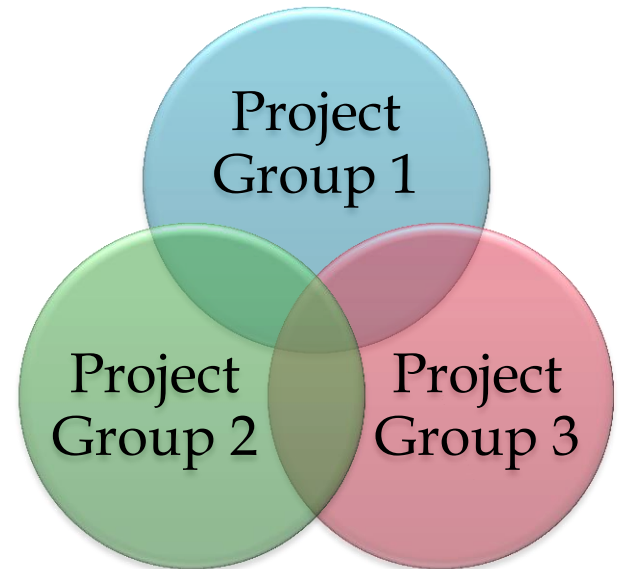
Vision

- Faculty + students working together on interesting problems in data science or artificial intelligence.
- Members of DataSAIL will meet regularly to
 - propose new projects
 - discuss existing ones, and
 - work on solving problems related to the projects.
- Goals of this meeting:
 - invite students to apply
 - form groups to work on projects

Vision

Biweekly meetings of all DataSAIL participants

- discuss what everyone is working on
- present tools/methods/languages
- share ideas/techniques/data
- invite guest speakers
- find collaborators/students to work on individual projects (**form project groups**)



Vision

Work on individual projects

(within project groups)

- faculty+students that participate on a given project would work on it continuously
 - regular group meetings and communication
 - report on progress during biweekly DataSAIL meetings
 - present work in relevant workshops, conferences, or scientific journals
- (or Celebration of Scholarship)

2. Faculty Introductions

Faculty

Dr. Piotr Szczurek

Assistant Professor, Director of MSDS

Education

- Ph.D., Computer Science, University of Illinois at Chicago (UIC), 2012
- B.S., Computer Science, University of Illinois at Chicago (UIC), 2005

Academic/research Interests

- Broad: artificial intelligence, machine learning, mobile databases
- Specific:
 - Intelligent transportation systems
 - Information relevance estimation
 - Applications of machine learning
 - Computer vision problems
 - Gaming AI

Faculty

Dr. Fatih Koksal

Assistant Professor

Education

- Ph.D., Mathematics, Texas Tech University, 2015
- Ph.D., Computer Engineering, Bogazici University, 2007
- M.S., Computer Engineering, Bogazici University, 2000
- B.S., Computer Engineering, Bogazici University, 1998

Academic/research Interests

- machine learning
- artificial intelligence
- fiber optic networking
- genetics
- homological algebra of rings

Faculty

Dr. Amanda Harsy

Assistant Professor

Education

- Ph.D., Mathematics, IUPUI, Indianapolis, IN, 2014
- M.S., Mathematics, IUPUI, Indianapolis, IN, 2011
- M.A., Mathematics, University of Kentucky, Lexington, KY, 2009
- B.A., Mathematics, Coaching Certificate, Taylor University, Upland, IN, 2007

Academic/research Interests

- mathematics education
- The Scholarship of Teaching and Learning
- geometric group theory
- machine learning.

Faculty

Dr. Jason Perry

Assistant Professor

Education

- Ph.D. Computer Science, Rutgers University, 2015
- M.A. Computer Science, Princeton University, 2004
- B.S. Computer Science, University of Kentucky, 1999

Academic/research Interests

- Data-driven security analysis
- Secure computation protocols
- Cryptography
- Natural Language Processing
- Theory of Programming Languages

Faculty

Daniel Ayala

Assistant Professor

Education

- Ph. D., Computer Science, University of Illinois at Chicago, exp. 2015
- M.C.S., Computer Science, University of Illinois at Urbana-Champaign, 2008
- B.S., Computer Science, University of Puerto Rico - Río Piedras, 2003

Academic/research Interests

- mobile data management
- intelligent transportation systems
- machine learning

Faculty

Dr. Sarah Powers

Assistant Professor

Education

- Ph. D., Immunology, University of Chicago, 2011
- B.A., Biological Sciences, University of Chicago, 2004

Academic/research Interests

- transcriptome analysis of human cancers bearing cyclin D3 mutations as well as structural changes within the protein caused by these mutations

Faculty

Dr. Cindy Howard

Associate Professor

Education

- Ph. D., Computer Science, University of Illinois at Chicago, 2010
- M.S., Computer Science, Governors State University, 2001
- B.B.A., Accounting and Information Systems, University of Wisconsin, 1985

Academic/research Interests

- mobile applications
- natural language processing
- intelligent tutoring systems

Faculty

Dr. Ray Klump

Professor and Chair of Computer and Mathematical Sciences

Education

- Ph.D., Electrical Engineering, University of Illinois at Urbana-Champaign, 2000
- M.S., Electrical Engineering, University of Illinois at Urbana-Champaign, 1995
- B.S., Electrical Engineering, University of Illinois at Urbana-Champaign, 1993

Academic/research Interests

- electric power system analysis
- computational techniques
- data visualization
- cyber security of critical infrastructures

4. Projects

Projects

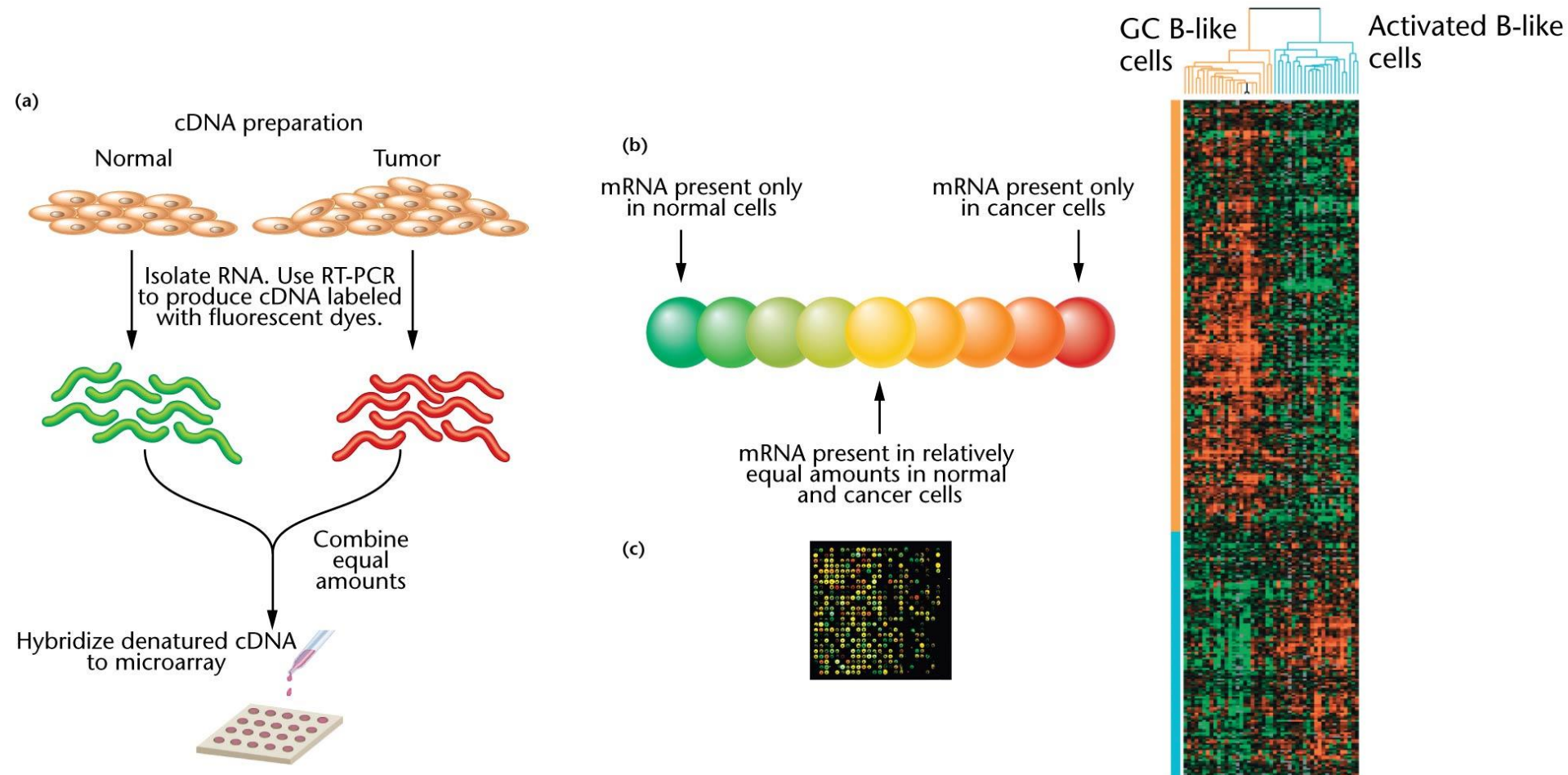
Current Projects / Project Ideas

- Analysis of Microarray Data from Cancers with Mutations in D-Type Cyclins
- College Enrollment Prediction
- Predicting Student Success
Using Machine Learning and Ranking methods
- Intrusion Detection
- Pedestrian Flows

Analysis of Microarray Data from Cancers with Mutations in D-Type Cyclins

Dr. Sarah Powers
Biology Department

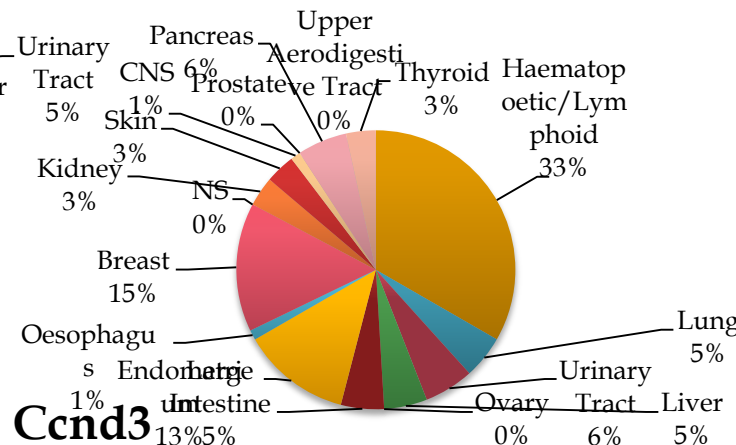
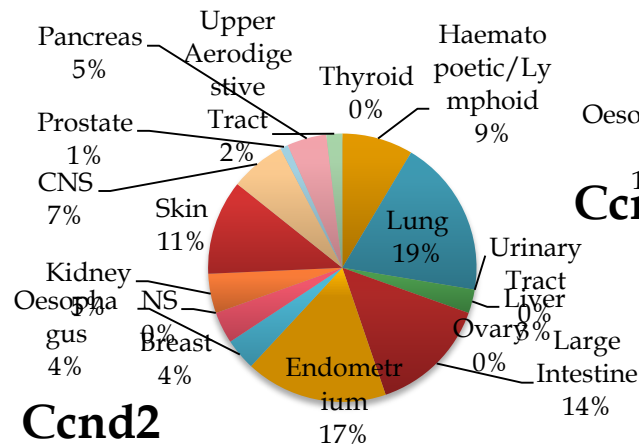
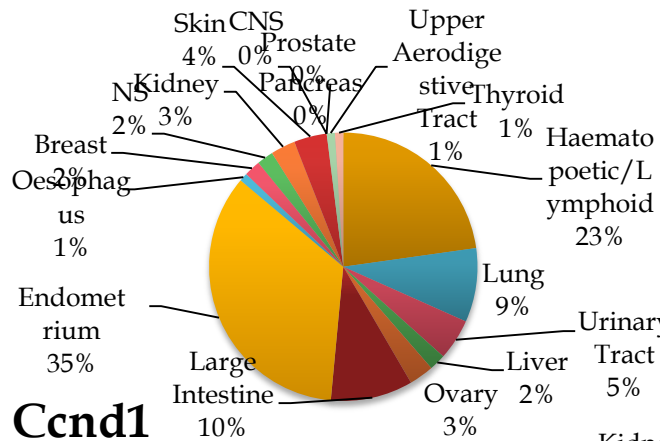
What is a microarray data set?



D-Type Cyclins in Variety of Cancers

Questions:

- Clustering based on cyclin mutant vs normal?
- Clustering based on cancer type?
- Clustering based on location of mutation?
- Clustering based on level of expression change?
- Other ways to group?



College Enrollment Prediction

Dr. Piotr Szczurek, Daniel Ayala, Dr. Ray Klump
Computer and Mathematical Sciences Department

College Enrollment Prediction

Data:

- Student info from marketing campaign (**Royall**)
- CampusAnywhere (if student responds)
- ACT/SAT scores



Goal: we want to know if certain *strategies make sense*

- are we targeting the right students?
- what are the most effective recruitment strategies?

College Enrollment Prediction

Example Questions

- ❑ *If they send us their ACT test score, are they more likely to enroll?*

- ❑ *if they visit campus, they are more likely to enroll.*
 - does it matter when they visit?
 - does it matter when they apply?

- ❑ *Is there a profile that never ever come to Lewis?*
 - from certain high schools? with certain GPAs, interests?

- ❑ *Where should we buy names?*

- ❑ *Which name pools should we be paying the most attention to?*

College Enrollment Prediction

PROBLEM/TASKS

1. Data gathering - high school info, location info, major info
 - Some manual searching / querying
 - Making scripts which parse data from web / web services
2. Dealing with missing values
 - Research and experimentation problem
 - Write programs or use existing tools
 - Test performance

College Enrollment Prediction

PROBLEM/TASKS (cont.)

3. Determining whether a student responds to a campaign
4. Determining most likely status sequence for a student
5. Determining which students that responds end up enrolling
 - Prediction problem - using machine learning
 - Write programs or use existing tools
 - Testing
6. Examining student types (clustering problem)
7. Finding association rules

College Enrollment Prediction

PROBLEM/TASKS (cont.)

8. Examining trends in majors (distribution over time; most likely majors, etc.)

- Modeling/prediction/regression

Some other potential directions:

- Can monitoring of social network activity be used to detect students likely interested in Lewis ???
- Developing a framework for examining student enrollment
- Make an application that allows for finding enrollment estimates, likely majors, finding interested students, etc.

Predicting Student Success Using Machine Learning and Ranking methods

Dr. Amanda Harsy

General Questions

- What contributes to a student's perseverance in their degree?
- Can we predict what types of students will succeed in a particular major?
- What contributes to gainful employment after graduation?
- What type of student is likely to attend and finish their degree at Lewis?
- How much impact does a major change in curriculum have for student success and recruitment?

Specific Questions

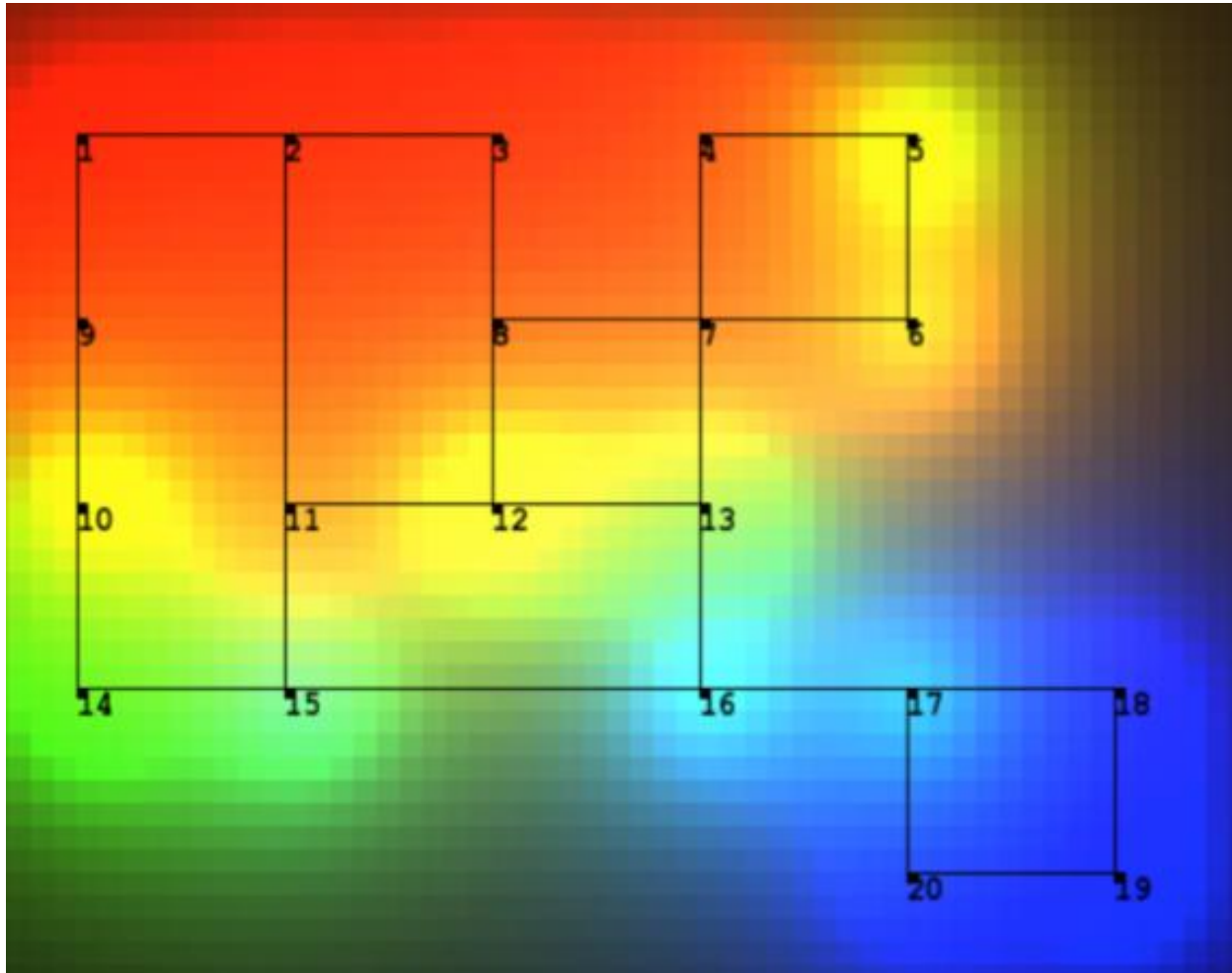
- Are there specific courses which can predict a student's success in the math major?
- Does commuting distance contribute to success in the math major?
- Is there an optimal number of math classes one should take per semester?
- How much of an impact do study habits have for a math major?
- Does your progression through the major influence your gpa?
- Does working while taking classes influence your success/perseverance in the major?
- What type of success are we looking for?
- Exploring different demographics in the major-male/female, commuter/resident

Other Ideas

Dr. Ray Klump

DATASAIL

Intrusion Detection Data



DATASAIL

4. Why and how to apply?

Advantages

1. Looks great on resume !!!
2. Gain experience
3. Meet other students who share your passions
4. Learn new knowledge
5. Collaborate with faculty, etc.
6. Can use research for senior seminar project, undergraduate capstone, independent study, Master's thesis or project, ...

Qualifications for Membership

Students who wish to participate should have

- At least one relevant course completed (machine learning, artificial intelligence, data mining, etc.)
- Some programming experience (python, java, ...)
- Time to work on projects and meet with group members

Both graduate and undergraduate students are welcome

How to Apply?

To apply:

- Send your resume, statement of purpose, and a letter of recommendation
- If you want to work on a specific project or with a specific faculty member - contact them directly
- Otherwise - contact me - szczurpi@lewisu.edu

If you don't have all qualifications, you may still be able to participate as a *provisional member*

5. Next Steps

Next Steps

1. Talk about current and future project ideas
2. Form groups - talk to professors and apply to join
3. Schedule DataSAIL next meeting