

LEWIS UNIVERSITY



**DATASAIL**

**DATATHON**

April 8<sup>th</sup>, 2016

**Information Session**

**Wednesday, March 9<sup>th</sup>, 2016**

**Piotr Szczurek, Ph.D.**

Assistant Professor, Department of Computer and Mathematical Sciences

Director of Master of Science in Data Science

Lewis University

## Agenda

1. What is a DATATHON?
2. Who can compete?
3. Competition rules
4. Prizes
5. Questions

# 1. What is a DATATHON?

## What is a DATATHON?

"A **datathon** is an intense 24-hour workshop that asks researchers to do their best to turn information into knowledge.

It's a format modeled after *hackathons*.

The difference is that **datathons** use research questions and datasets to advance knowledge, not to launch apps."

(<https://ipk.nyu.edu/initiatives/datathons>)

## What is a DATATHON?

### Goal:

- Use data to extract knowledge
- Results need to answer given *framing questions* - given at the start of the competition
- Form a presentation of results (PowerPoint slides, video, web page) and present to judges

## What is a DATATHON?

### Past Datathons:

**UC-Berkeley D-Lab:** "Big cities, big data: Big opportunity for computational social science"

- Goal: "to examine contemporary urban issues - especially around housing - with municipal data from cities including San Francisco, New York, Seattle, Boston, Austin, and Chicago."

**NYU:** Social and Meteorological Data

- Goal: "to investigate the impact of climate change on New York City"

## What is a DATATHON?

Example Datathon topic (NYU):

**"The prompt:** Climate change will increasingly shape everyday lives around the world and become a focal lens for social science researchers. Large publicly available datasets full of social and meteorological data exist going back for more than a hundred years."

## What is a DATATHON?

Example of **framing questions** (NYU):

- How can these datasets be brought together to meaningfully examine the impact of climate change on cities?
- Should social scientists and/or climate scientists call for new types of data to be collected to meet the challenges of climate change research?
- What advantages are offered by focusing on the city as a unit of analysis?
- What technological, political, and social challenges will need to be addressed as climate change grows as a focal lens in research communities?



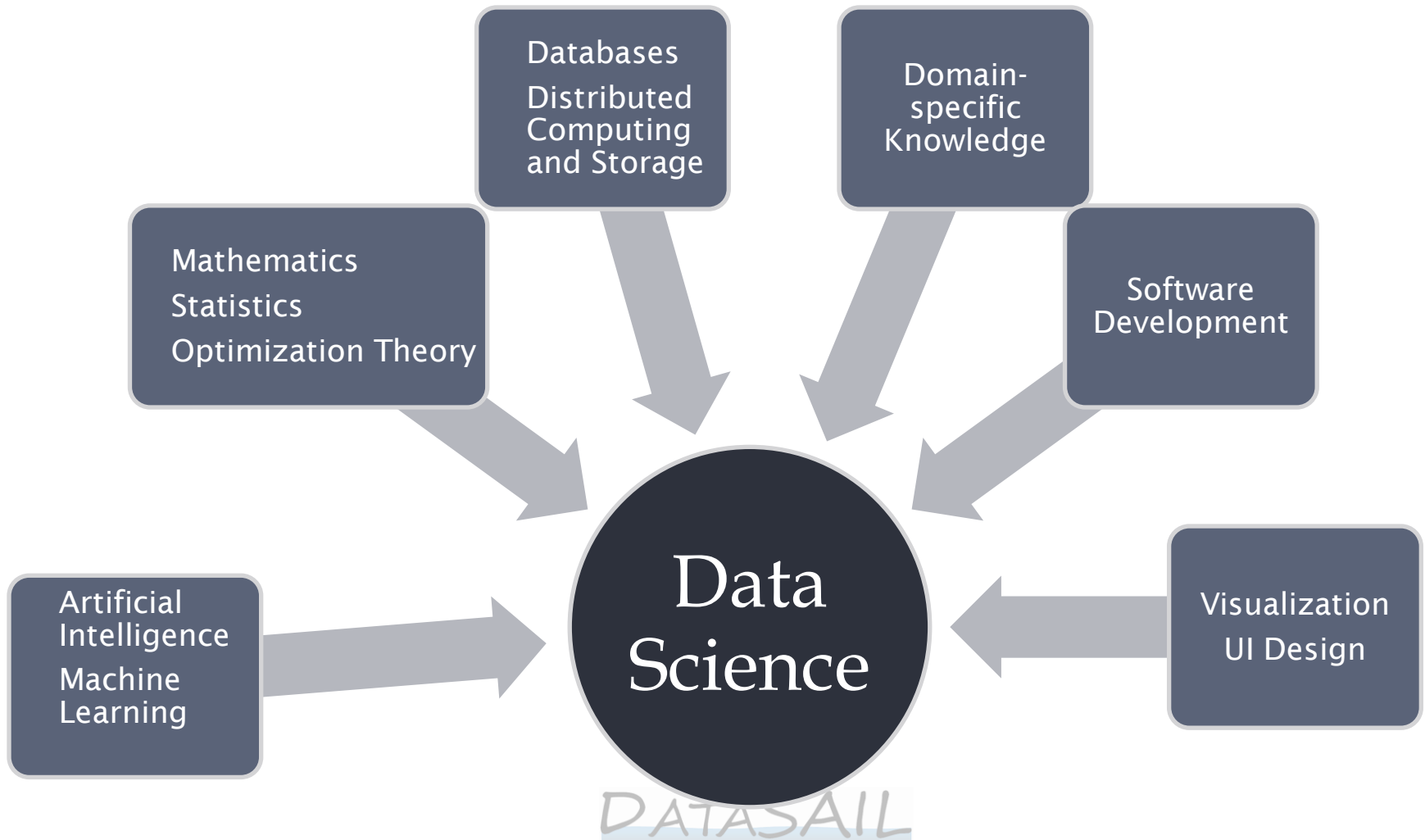
## What is a DATATHON?

Process:

1. Find a team of researchers
2. Get dataset(s)
3. Try your best to analyze the data and answer the framing questions (using all possible tools)
4. Make a presentation
5. Present to judges

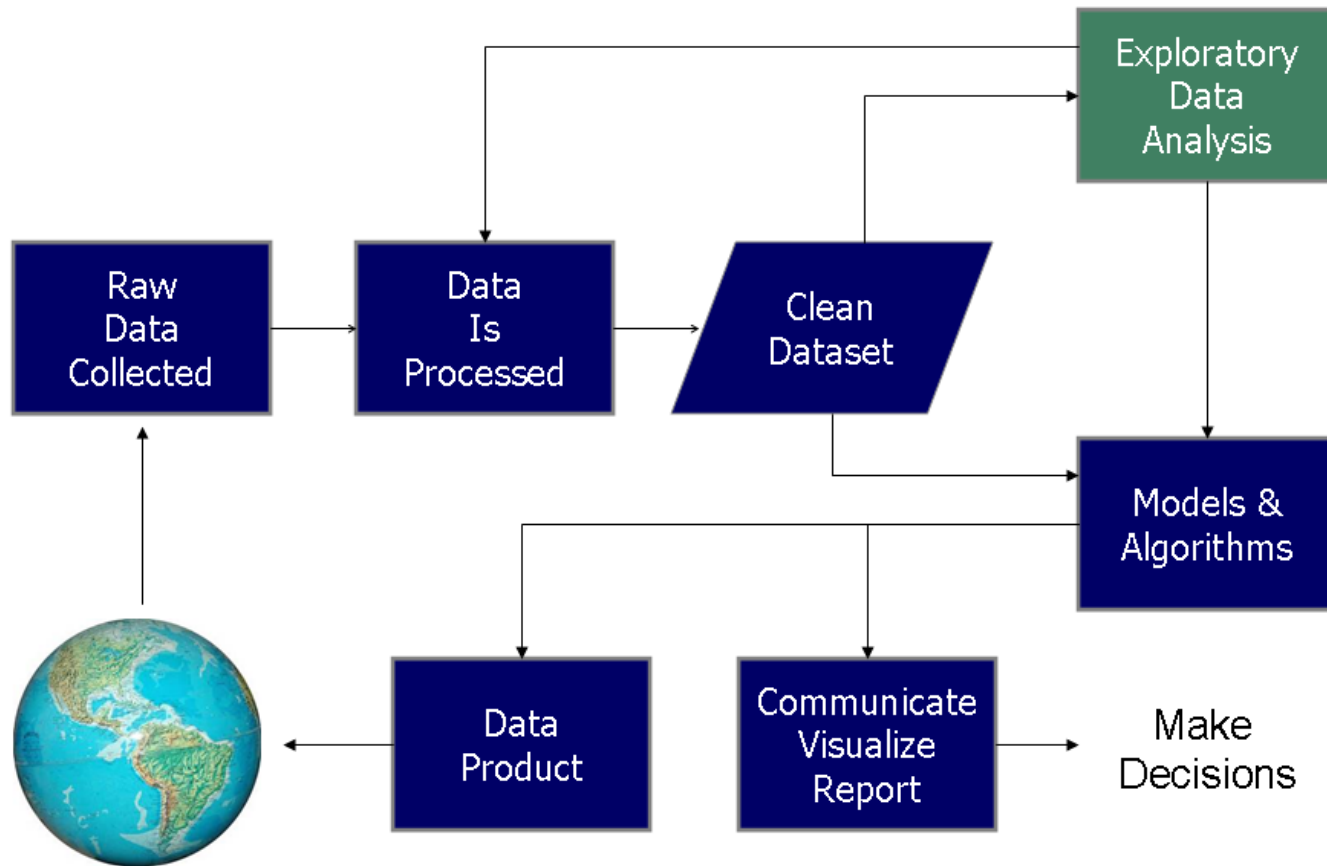
# What is Data Science?

“**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



## 2. Who Can Compete?

## Who Can Compete?

### **Open to all Lewis University students**

- This includes online students  
(through Collaborate or join.me tool for presentations)

### **All majors are encouraged to form teams**

- It's useful to have a variety of skills and knowledge

## Who Can Compete?

### **Knowledge:**

prob & stats, database management, programming languages, data mining/machine learning, ...

### **Skills:**

programming and using various tools, e.g.,:  
MS Excel/Access, SAS, R, Weka, Python, Tableau, ...

## 3. Competition Rules

## Competition Rules

1. Teams will have a maximum of four persons.
2. The datathon will be held April 8<sup>th</sup>-9<sup>th</sup> over a 24 hour period of time (3:30pm - information, then contest runs from 4pm Friday to 4pm Saturday).
3. Teams can work in the lab or work outside. They would report results to judges after 24 hours.
4. Conclusions must come from given dataset. No outside help for analysis.
5. The teams can use any tools necessary.



## Competition Rules

6. Results must answer *framing questions*. These questions will be supplied before the start of the Datathon.
7. The results must be presented by the end of the Datathon to the judges. The form of the results can be any or all one of the following: PowerPoint presentation, video demonstration, or a presentation of developed software.
8. Each team will have 15 minutes to present their work, followed by 3 minutes of questions from the judges.

## Competition Rules

The results will be judged by the following categories (10 point scale for each):

- **novelty** – how original and surprising the results are?
- **practicality** – can the results be useful for solving real-life problems?
- **demonstrated depth of understanding** – do teams show their knowledge of the methodology which was employed for getting the results?
- **mobilization of existing datasets** – was all of the data utilized and were there any external datasets used?
- **communication** – were the results presented effectively?

## Competition Rules

### **Judges:**

- Dr. Piotr Szczurek, CaMS
- Dr. Ray Klump, CaMS
- Dr. Sarah Powers, Biology
- Dr. Fatih Koksak, CaMS
- Dr. Jason Perry, CaMS
- Dr. Amanda Harsy, CaMS

## 4. Prizes

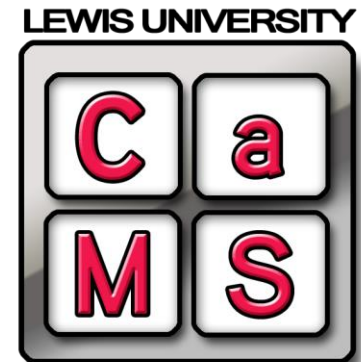
## Prizes

**Sponsored by the  
Department of Computer and Mathematical  
Sciences**

**1<sup>st</sup> place - \$500**

**2<sup>nd</sup> place - \$250**

**3<sup>rd</sup> place - \$100**



**Note: judges reserve the right not to award some or all of the prizes if the results do not answer the framing questions or do so in a trivial way.**

## 5. Next Steps and Questions

## Next Steps and Questions

- Slides will be posted to DataSAIL website:  
<http://cs.lewisu.edu/datasail/>
- Announcement will be made about registering teams
- Questions?

**THANK YOU FOR PARTICIPATING !!!**

DATASAIL