

# Eric Pitman Summer Workshop in Computational Science

## Intro to

...an introduction to R, statistics, programming, and  
getting to know datasets

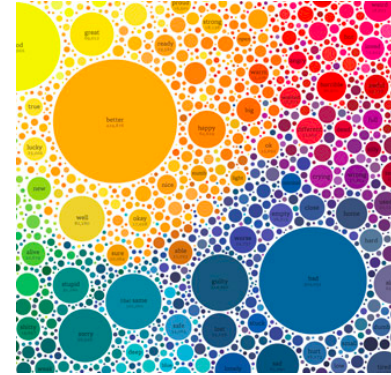


CENTER FOR **COMPUTATIONAL RESEARCH**

 **University at Buffalo**  
*The State University of New York*

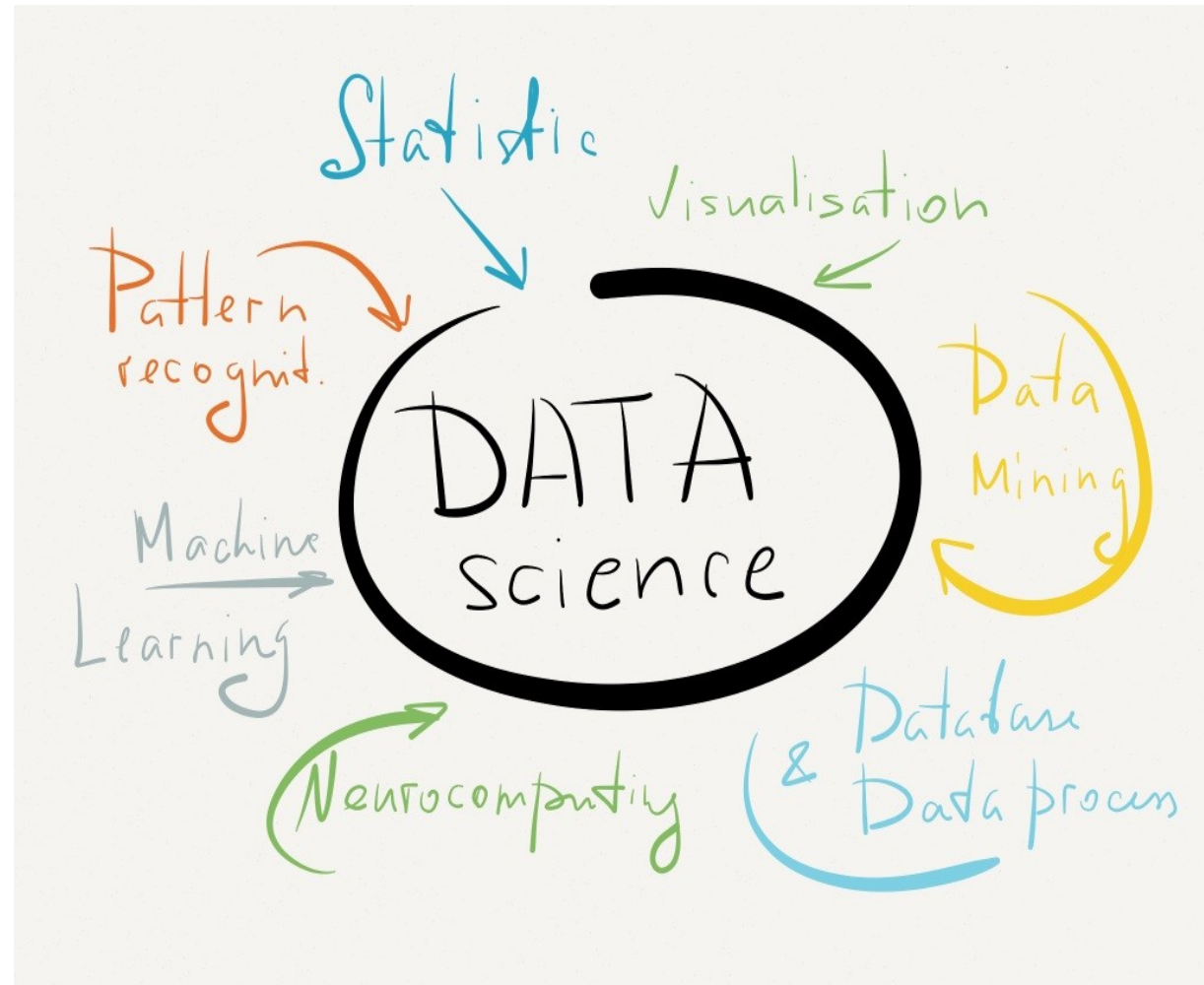
# Our Plan for the Workshop

- Introduce the R language
- Do some programming
- Tours, demos, guest lectures: sample some science & engineering
- Examine, model, and visualize datasets
- Read about, discuss data science
- Project: explore materials science design questions

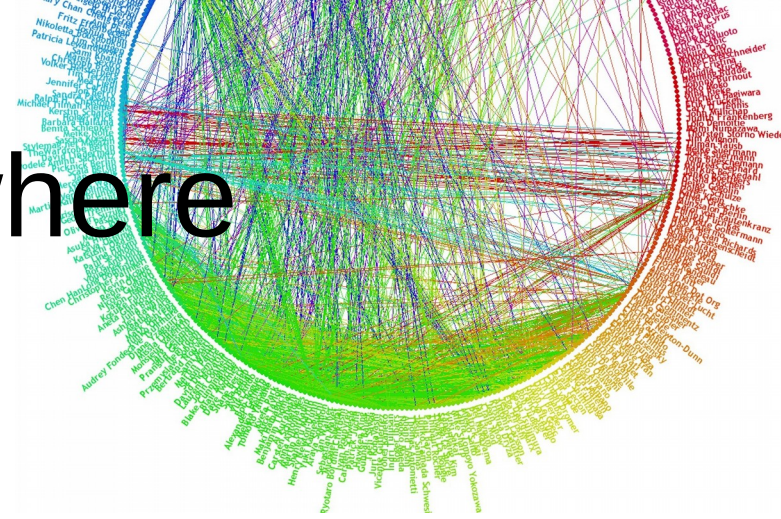


# Our (Informal) Plan for the Workshop

- No grades
- Collaborate
- Respect
- Learn something!
- Life is open internet (so is the workshop)



# Data is Everywhere



- For example:
  - Science/engineering/medicine
  - Environmental science/Social science/Law enforcement
  - Finance and marketing
  - Social media
- How do we come to an understanding of what a dataset contains?
- Can we draw conclusions from a dataset?
- Let's taste the complexities for ourselves

# “Big Data” means Three Things

*Volume*: lots of data

*Velocity*: coming at you fast—Twitter, 7+TB/day

*Variety*: text, pictures, video, etc.

Statistic

Visualisation

Pattern  
recognit.

DATA  
science

Data  
Mining

Machine  
Learning

Neurocomputing

& Database  
Data process

