

# INTRO to DATA SCIENCE

## LECTURE 2: MACHINE LEARNING

Jason Dolatshahi  
Data Scientist, EveryScreen Media

## **LAST TIME:**

- FIRST LOOK AT DATA SCIENCE & THE DATA MINING WORKFLOW**
- DATA EXPLORATION WITH UNIX**
- DATA VISUALIZATION WITH R & GGPLOT2**

## **QUESTIONS?**

## What's big data?

The practical viewpoint:

- ①  $O(n^2)$  algorithm feasible: small data
- ② Fits on one machine: medium data
- ③ Doesn't fit on one machine: big data

**I. WHAT IS MACHINE LEARNING?**

**II. MACHINE LEARNING PROBLEMS**

**EXERCISES:**

**III. MULTIPLE REGRESSION & FEATURE EXTRACTION**

# **I. WHAT IS MACHINE LEARNING?**

---

## WHAT IS MACHINE LEARNING?

---

6

from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can *learn from data*.”

*source: [http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning)*

---

## WHAT IS MACHINE LEARNING?

---

7

from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can *learn from data*.”

“The core of machine learning deals with *representation* and *generalization*...”

source: [http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning)

from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can *learn from data*.”

“The core of machine learning deals with *representation* and *generalization*...”

- *representation* – extracting structure from data

source: [http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning)



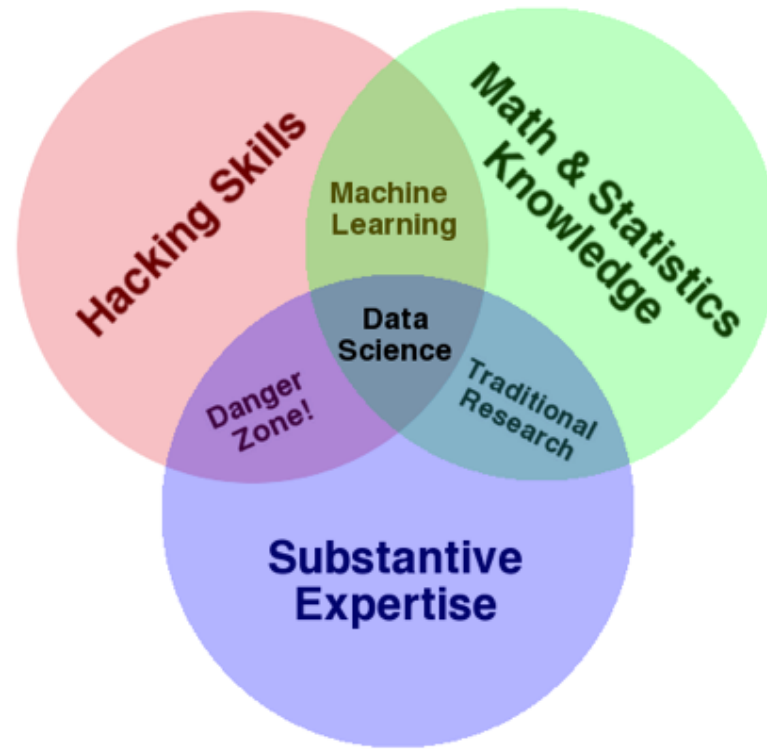
from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can *learn from data*.”

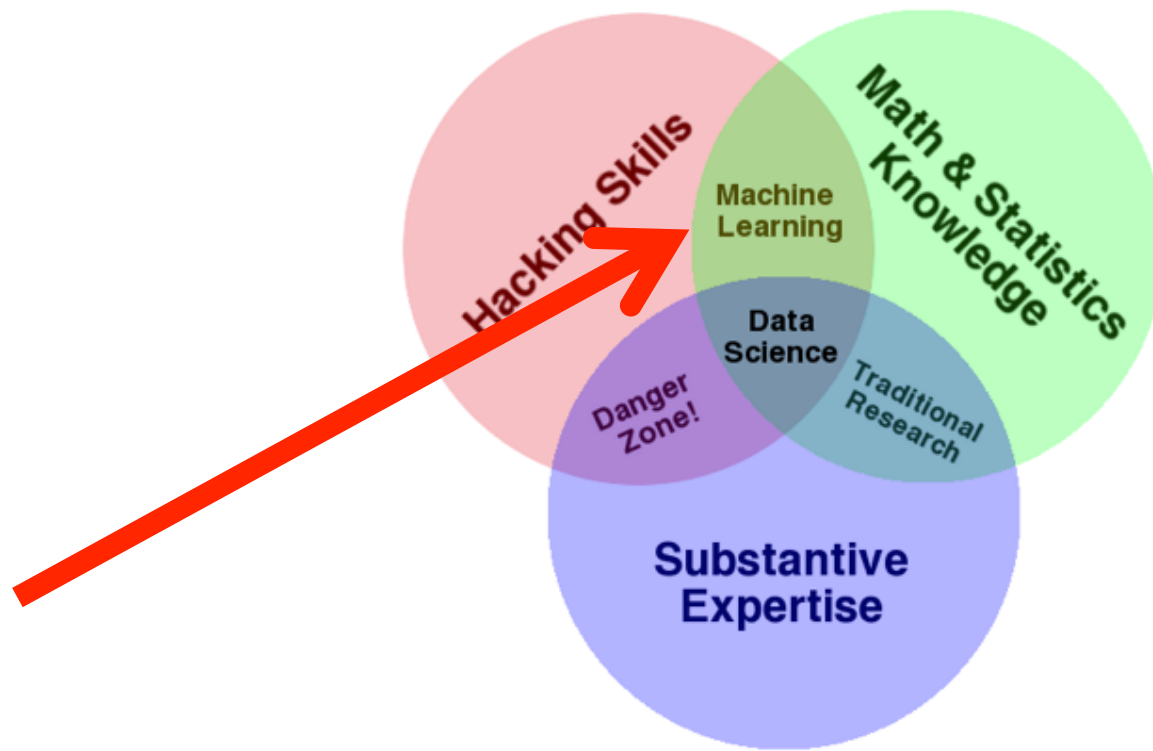
“The core of machine learning deals with *representation* and *generalization*...”

- *representation* – extracting structure from data
- *generalization* – making predictions from data

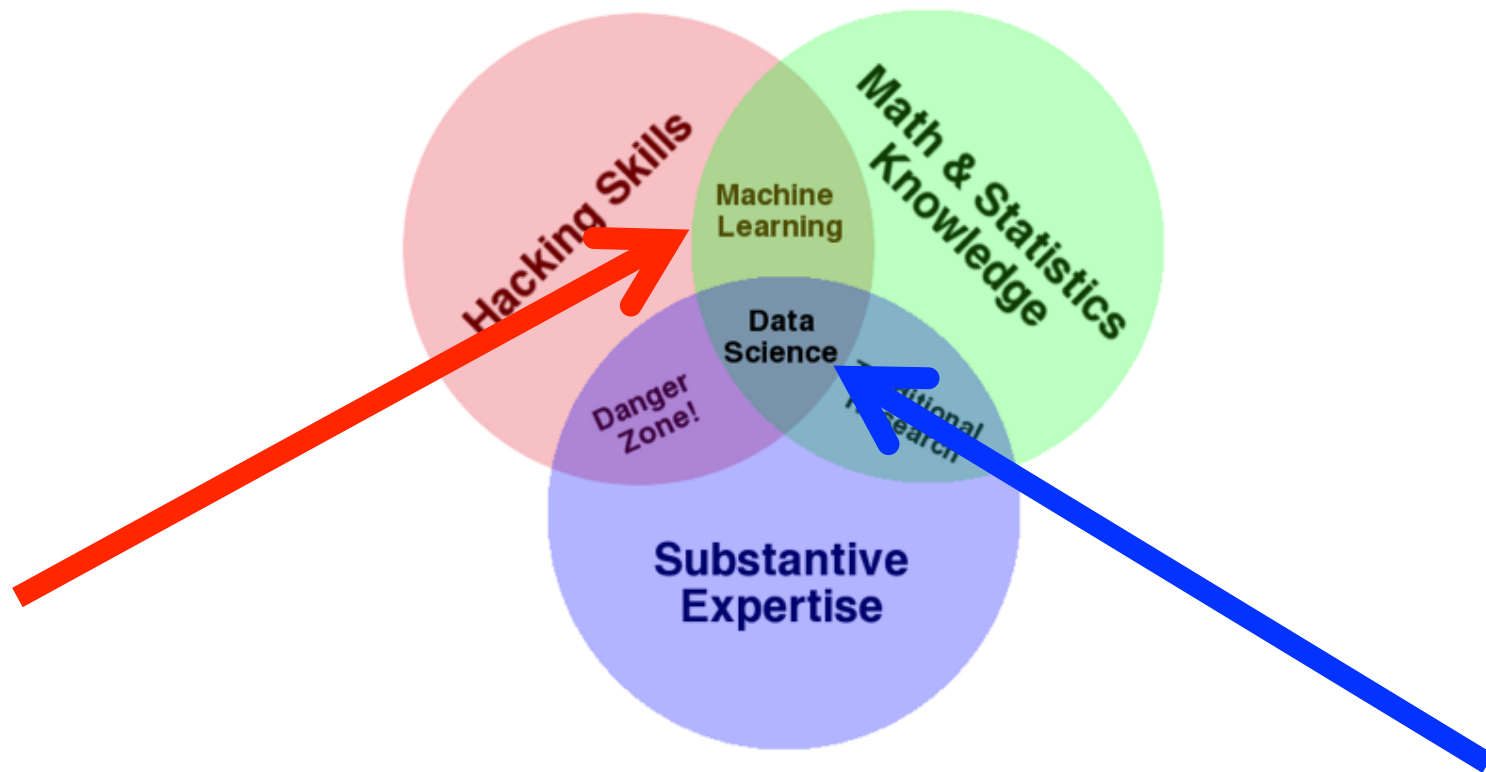
source: [http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning)



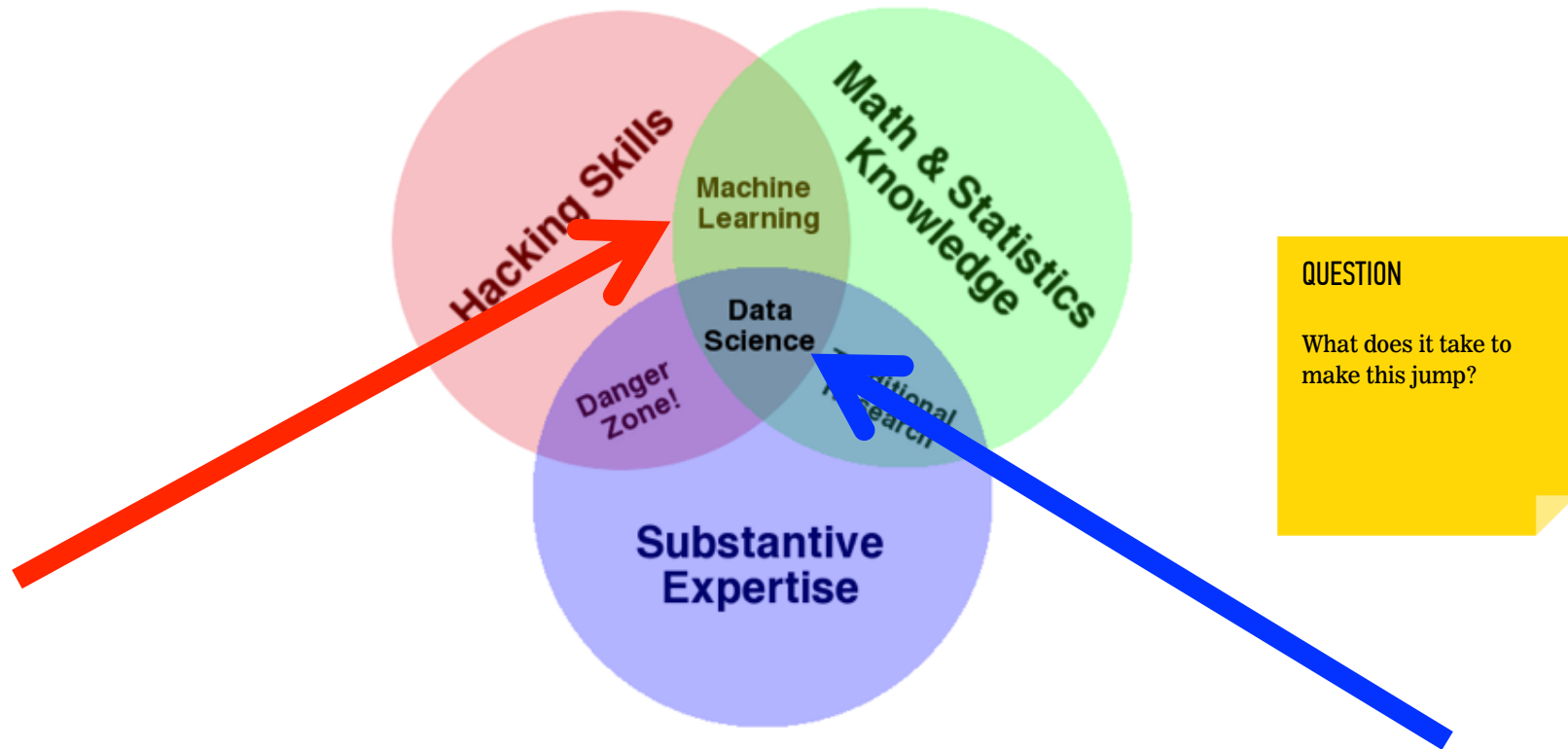
source: <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>



source: <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>



source: <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>



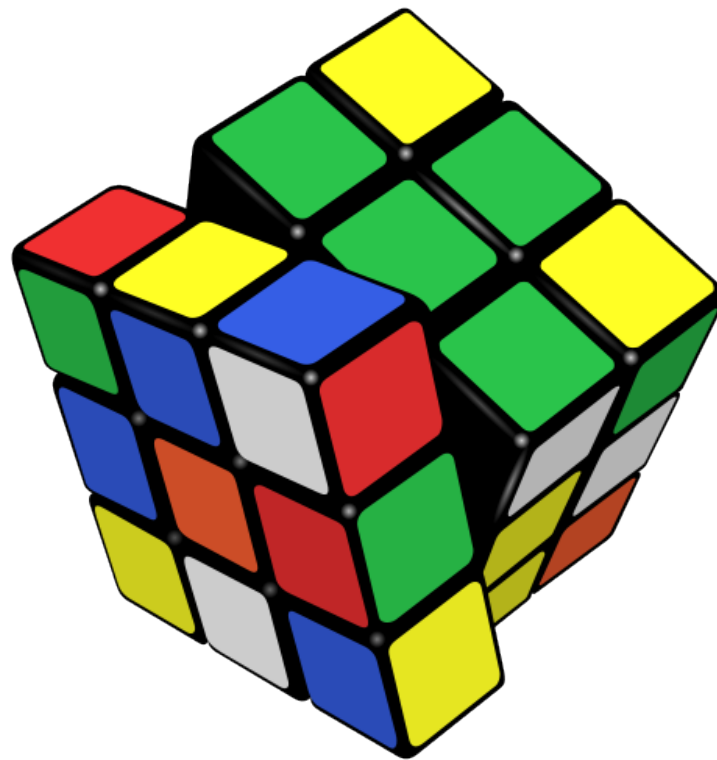
source: <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>

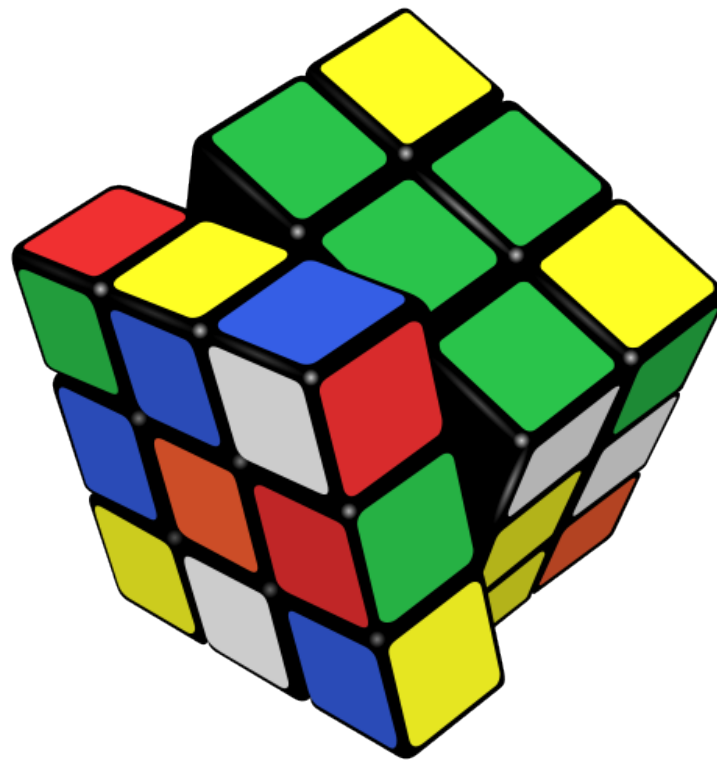
---

**ANSWER: PROBLEM SOLVING!**

---

**14**



**NOTE**

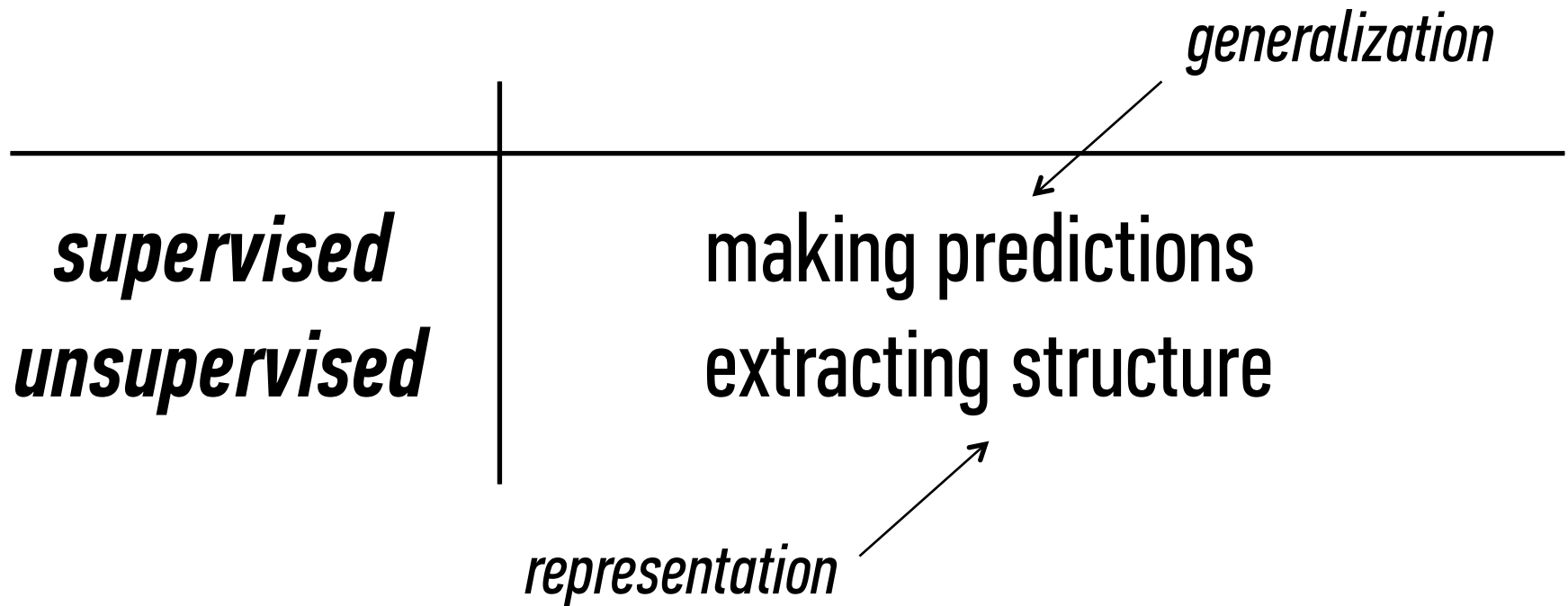
Implementing  
solutions to ML  
problems is the focus  
of this course!

# **II. MACHINE LEARNING PROBLEMS**



---

<i><b>supervised</b></i>	making predictions
<i><b>unsupervised</b></i>	extracting structure



	<i><b>continuous</b></i>	<i><b>categorical</b></i>
	<b>quantitative</b>	<b>qualitative</b>

*continuous*

*categorical*

quantitative

qualitative

NOTE

The space where data live is called the *feature space*.

Each point in this space is called a *record*.

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

## NOTE

We will implement solutions using *models* and *algorithms*.

Each will fall into one of these four buckets.

---

## QUESTION

---

***WHAT  
IS THE  
GOAL  
OF  
MACHINE LEARNING?***

---

<i><b>supervised</b></i>	making predictions
<i><b>unsupervised</b></i>	extracting structure

ANSWER

The goal is determined  
by the type of problem.



---

## QUESTION

---

***HOW  
DO YOU  
DETERMINE  
THE RIGHT  
APPROACH?***

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

## ANSWER

The right approach is determined by the desired solution.

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

ANSWER

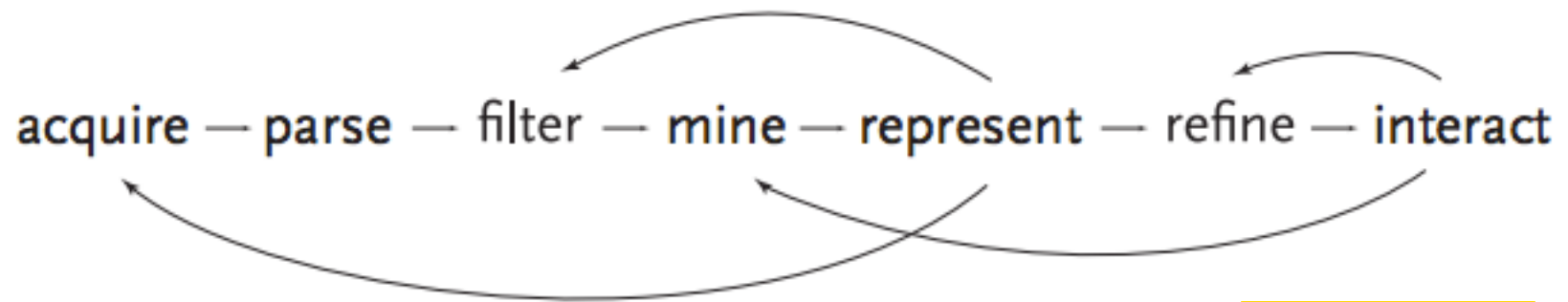
NOTE  
The is d  
des All of this depends on  
your data!

---

## QUESTION

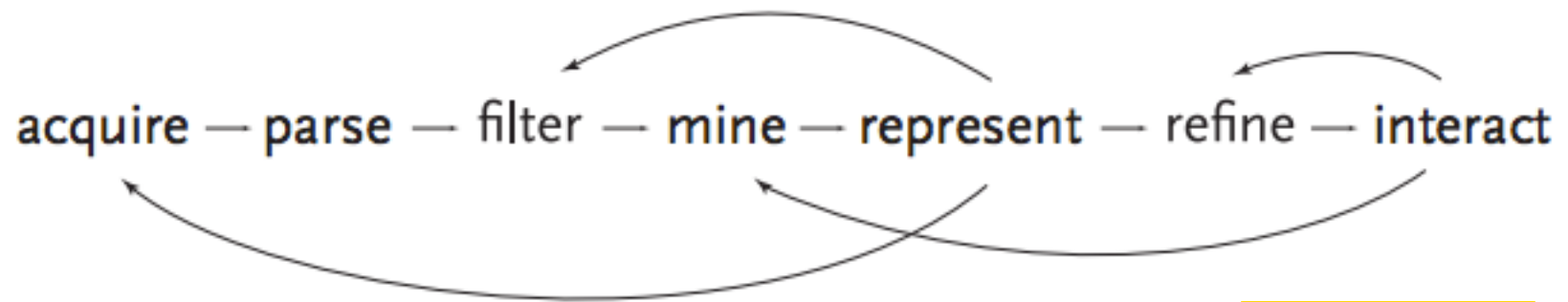
---

***WHAT  
DO YOU  
DO  
WITH YOUR  
RESULTS?***



### ANSWER

Interpret them and react accordingly.



ANSWER

In NOTE  
re

This also relies on your  
problem solving skills!

# **III. RELATIONSHIPS AMONG SEVERAL VARIABLES**

---

## EXERCISE – MULTIPLE REGRESSION (BACKWARD ELIMINATION)

---

32

### KEY OBJECTIVES

---

- Create a regression model using several independent variables
- Extract meaningful features

### TOOLS

---

- R (plot, lm, update)



---

**INTRO TO DATA SCIENCE**

---

**DISCUSSION**