Big Data and Classification
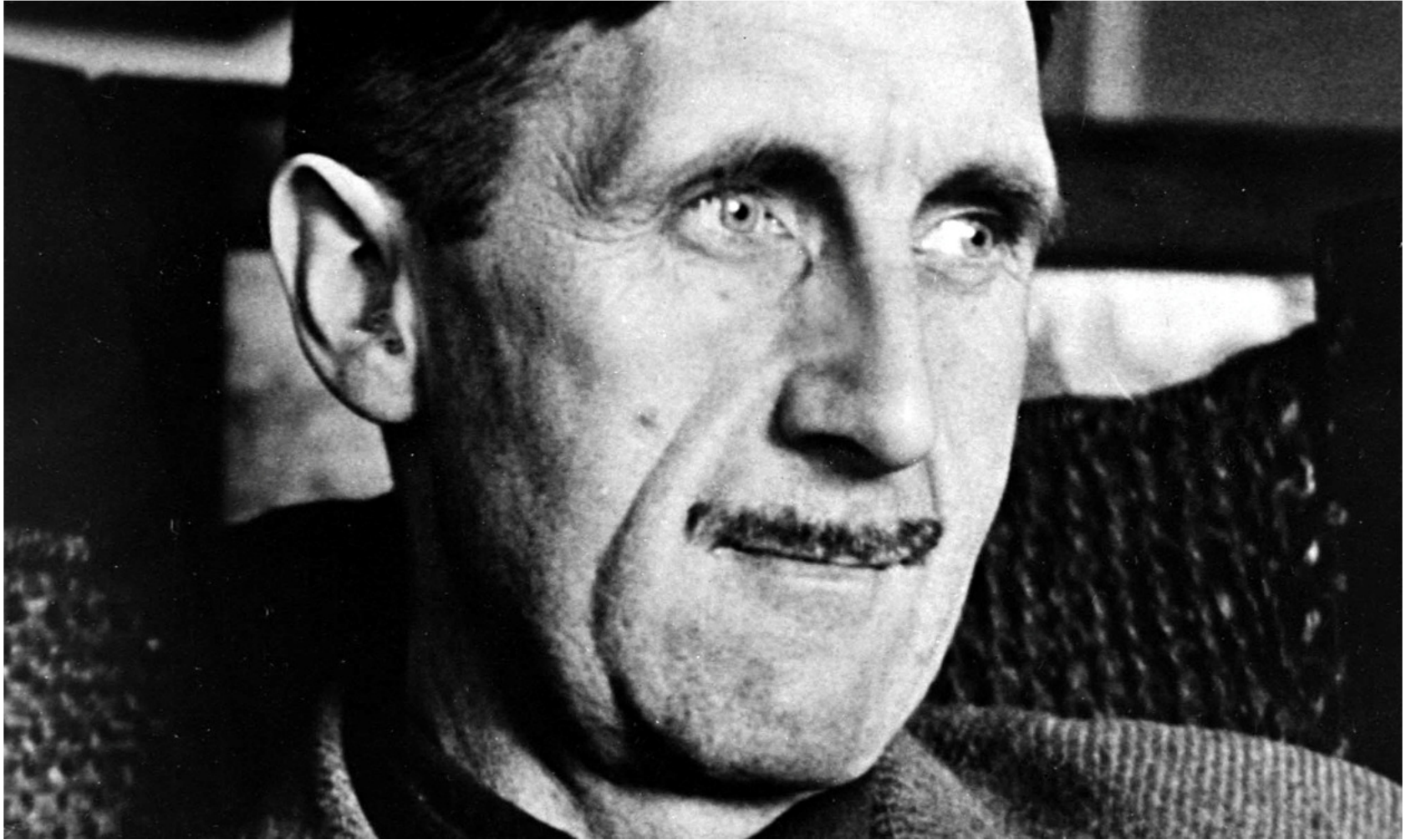

Paul Balas

Content Architect

303Computing

A Dystopian Future

George Orwell feared those who would deprive us of information.
He feared the truth would be concealed from us.

He never imagined "Big Data"

A glut of information that would conceal understanding

In 9.5 minutes or less…

Convince you that without classification

BIG DATA FAILS

clas·si·fi·ca·tion

/ˌklasəfəˈkāSH(ə)n/

noun

the action or process of classifying something according to shared qualities or characteristics.
"the classification of disease according to symptoms"
synonyms: categorization, categorizing, classifying, grouping, grading, ranking, organization, sorting, codification, systematization
"the classification of diseases"

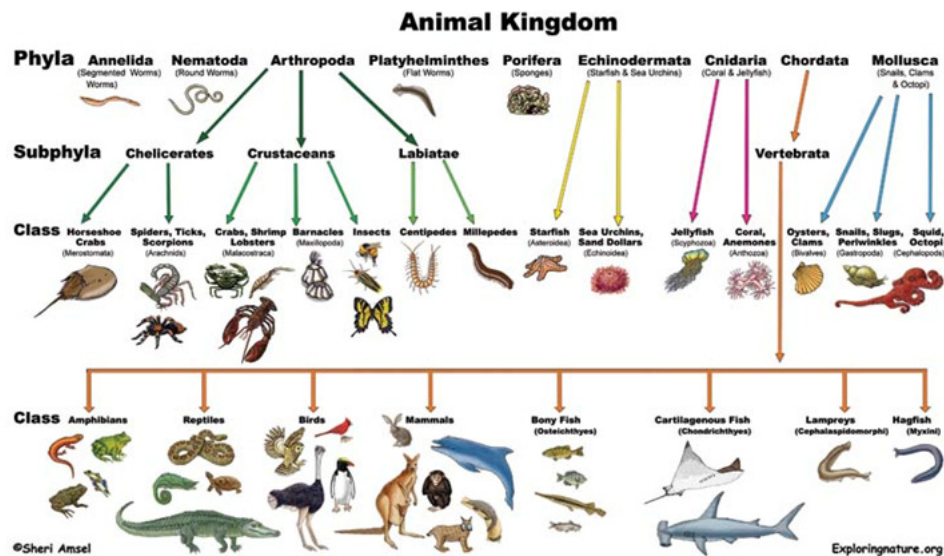- another term for taxonomy.

- a category into which something is put.
  plural noun: **classifications**
  synonyms: category, class, group, grouping, grade, grading, ranking
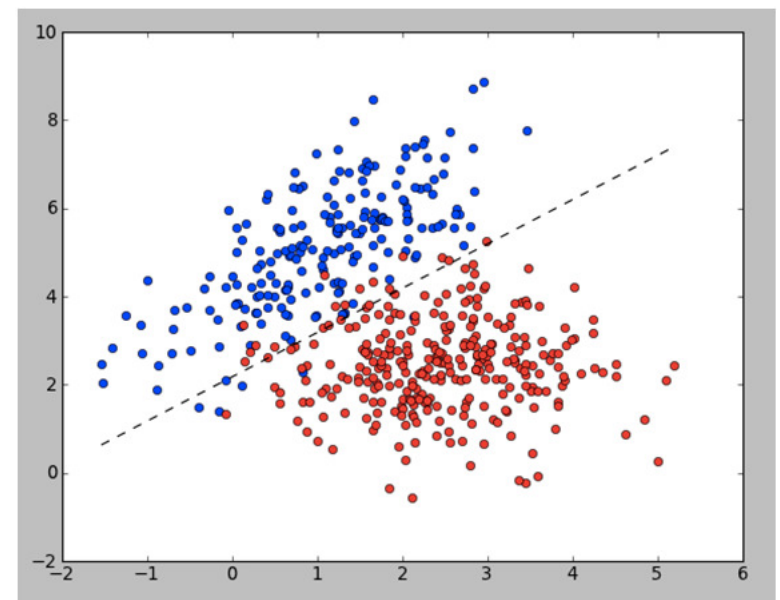 "a series of classifications"

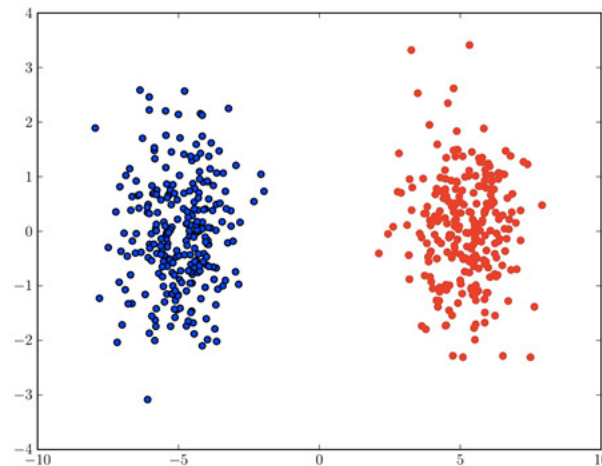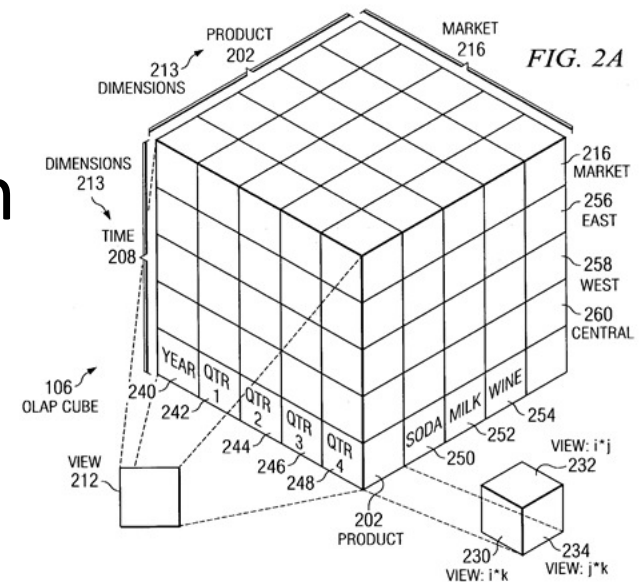# Methods for Classification

Taxonomies

Data Mining Classification

Human-driven

Machine-driven

Classification Helps!

- Group information by common attributes

- Easily compare similarities and differences

People Classify

Not Machines

Not Algorithms

All classification is done by humans at some point in the life of a datum

Without Classification, finding information is like finding a needle in a haystack...

Or, mistaking the haystack for a pile of needles

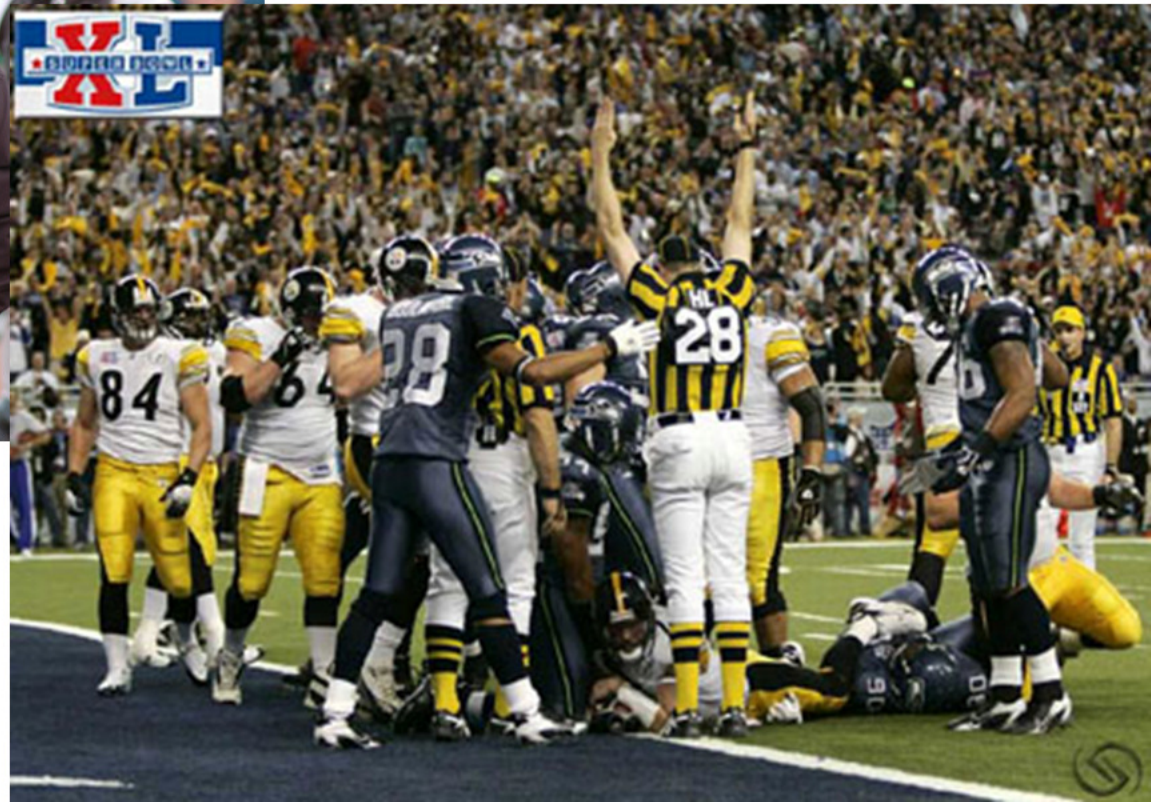With Big Data, the haystack is huge

People don't always agree

Scott Steinmann

Super Bowl XL

A Quiz for you…

On the next slide, I want you to tell me what these four types of data have in common

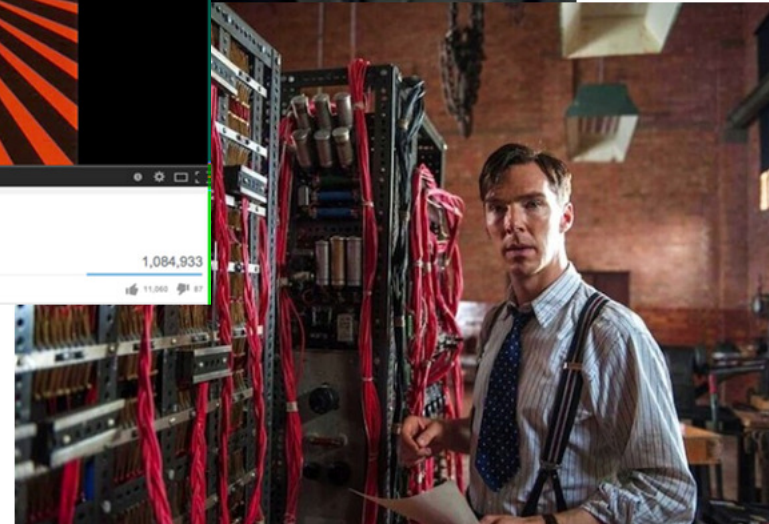Raise your hand when you get the answer…

(don't worry, I won't call on anyone)

Text

"A computer would deserve to be called intelligent if it could deceive a human into believing that it was human."

Video/Audio

Maps

Images

The Machine - Moons of Neptune

mustaba1234's channel

Subscribe 6,089

1,084,933

Did you get it right?

Alan Turing

The more data types we have
The harder the classification

Classification Cracked The Enigma Code

158,962,555,217,826,360,000 possibilities

Turing used Classification of the data to narrow the problem set

1st   A letter can never be itself
2nd Known Phrases - The weather report

Without Classification

There is no Correlation

Without Correlation

We are all out of jobs!

The 'Classification Food Chain'

Classification shapes data

Shaped data enables data quality

Data Quality delivers confidence in results
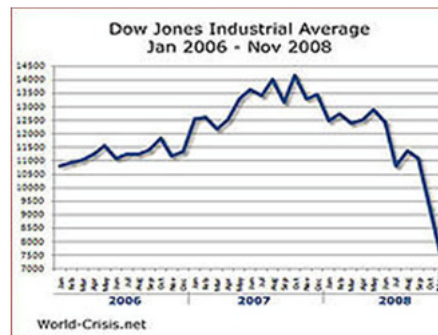
Bad Classification Has Bad Consequences

Elections are won

Shuttles explode

Financial Markets Meltdown

If you want to be confident in your Big Data results...

Invest in your classifications as they are critical to your success!

Thank You!

Paul Balas

303computing@gmail.com