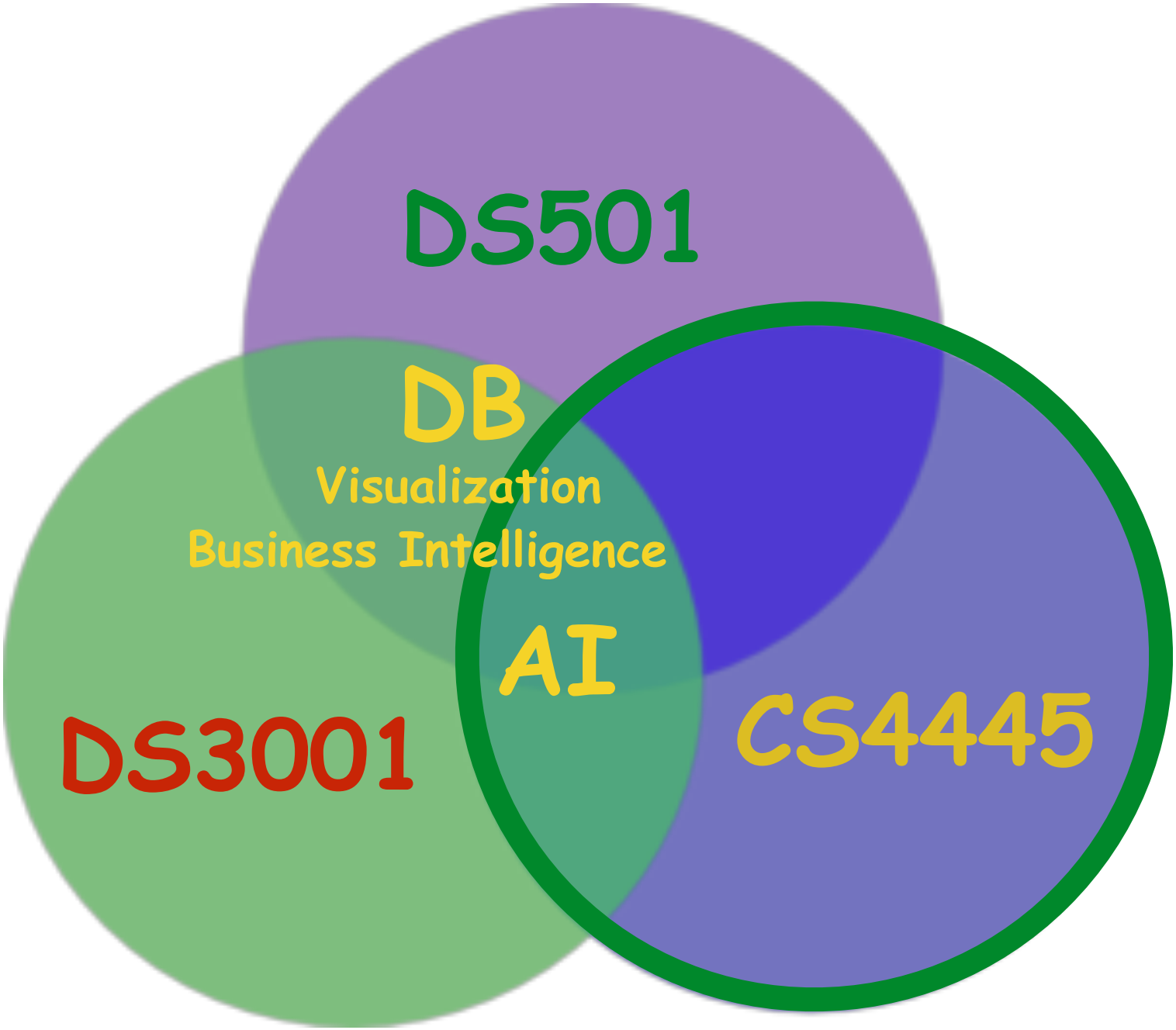
CS4445, CS4803, BCB4003 Data Mining & Knowledge Discovery

Goal of this course

A bird's eye view of Data Mining

Overlaps in Course Material

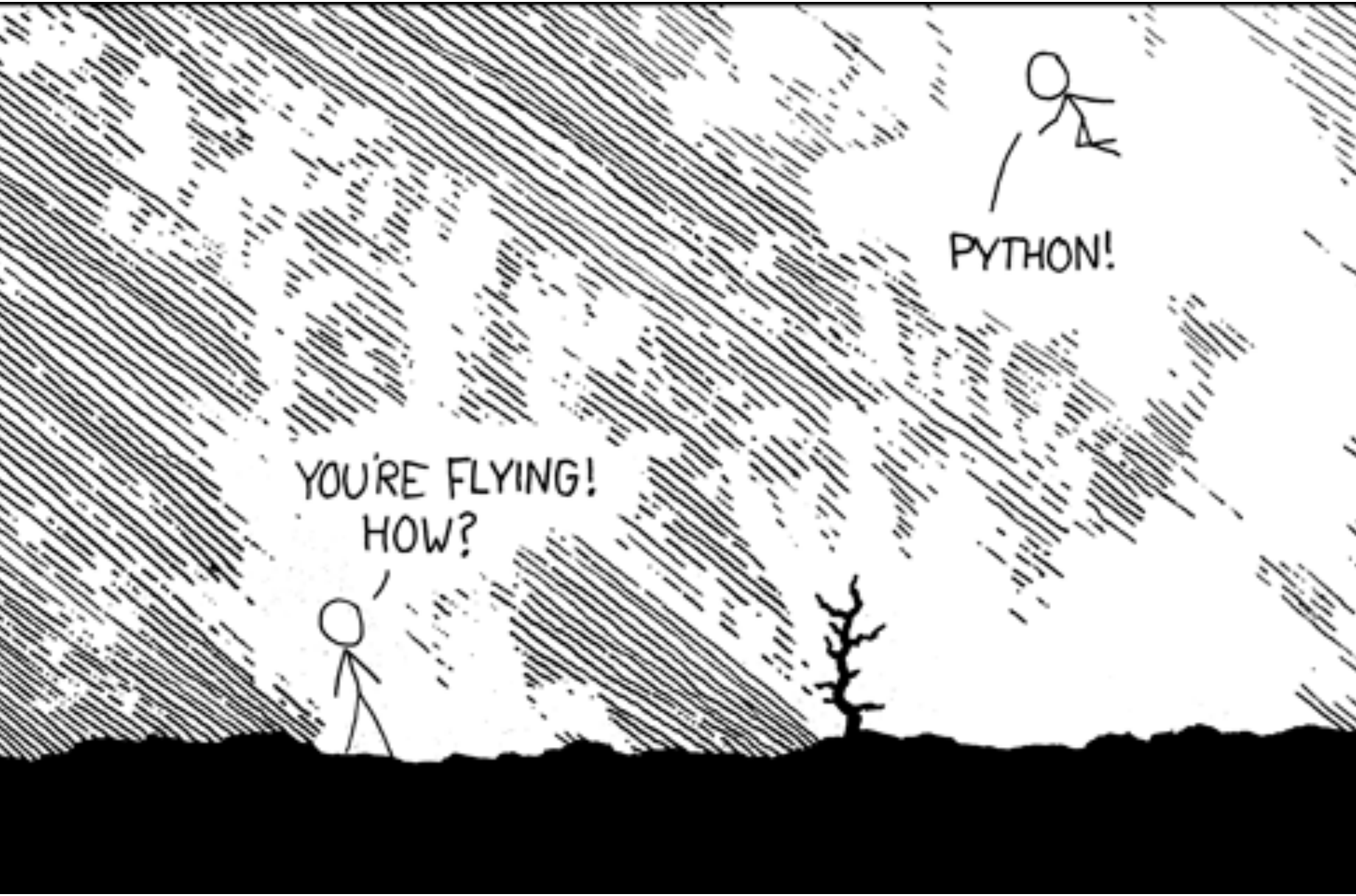


Gear up in the era of Big Data

Math + Tools

- Be comfortable with both

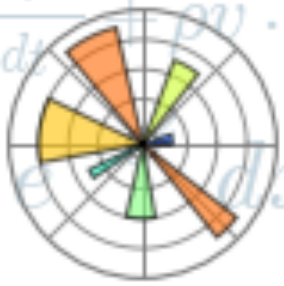
Official Language for CS4445



IP[y]: IPython
Interactive Computing

scikit learn

NumPy



matplotlib

SciPy.org

Sponsored By ENTHOUGHT

Scores

- Homework Assignments: 60%
- Quizzes: 40%

Homework (5X12%=60%)

- 6 Homework Assignments
- the lowest homework score will be dropped
- Need Python 3
- Homework 1 is released TODAY

Quizzes (10 X 4% = 40%)

- 12 Quizzes
- 2 lowest quiz scores will be dropped
- No exam

Course Calendar

Due Dates: homework, quiz

Course Summary:

Date	Details	
Thu Oct 26, 2017	Lecture 1	12pm to 1:50pm
Mon Oct 30, 2017	Lecture 2	12pm to 1:50pm
	Quiz 1	12pm to 12:10pm
Thu Nov 2, 2017	HW 1 due at 12pm	12am
	Lecture 3	12pm to 1:50pm
	Quiz2	12pm to 12:10pm
Mon Nov 6, 2017	Lecture 4	12pm to 1:50pm
	Quiz 3	12pm to 12:10pm
Tue Nov 7, 2017	Grading of HW1 due	12am
Thu Nov 9, 2017	HW2 due at 12pm	12am
	Lecture 5	12pm to 1:50pm
	Quiz 4	12pm to 12:10pm
Mon Nov 13, 2017	Lecture 6	12pm to 1:50pm
	Quiz 5	12pm to 12:10pm

What is Data Mining?

Data Mining:
Why should we care?

Data in Silicon Valley

http://ed.ted.com/on/MYodUMYe

YAHOO!

Microsoft

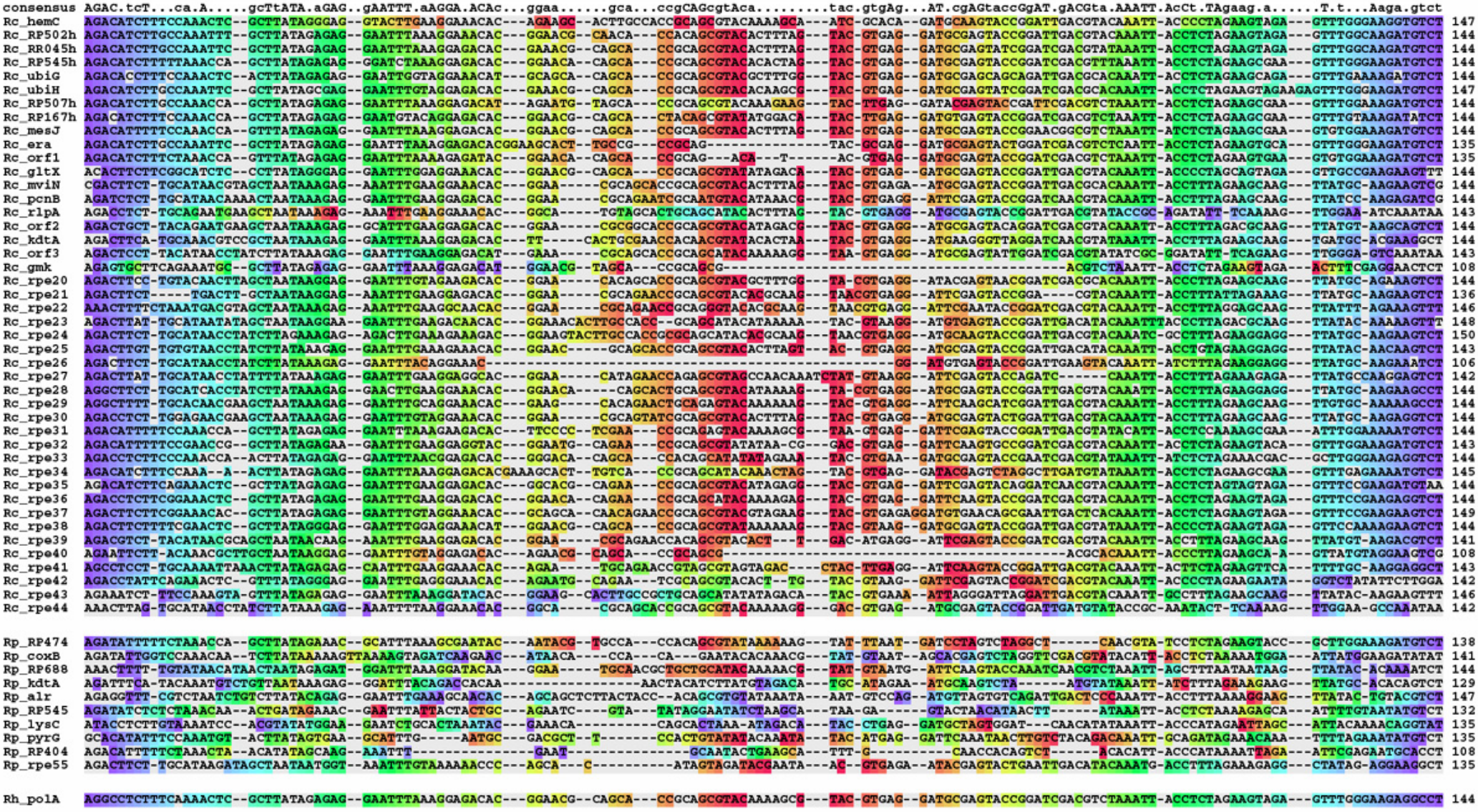
in

a

facebook

Data in Wall Street

Data in Biomedical Research



https://www.youtube.com/watch?v=FzcTgrxMzZk

BBC documentary

- The Age of Big Data
  - Crime Prevention
  - …