Development of a BS degree in Data Science

Joseph Reid & Rosanna Overholser

Outline

- What is Data Science?
- Market Analysis
- Budget Analysis

What is Data Science?

"…[e]xtracting knowledge or insights from large quantities of complex data for use in a broad range of applications."

from the National Academy of Sciences Report *Data Science for Undergraduates: Opportunities and Options*

For example…

Let’s say we are designing a new program for data science and want to answer the following question:

*What **skills** does industry most want in a data scientist?*

The “traditional” solution

*What **skills** does industry most want in a data scientist?*

A statistician might suggest **surveying** a representative sample of hiring managers or current data scientists

But this solution is time consuming and depends upon the willingness of respondents to participate in the study.

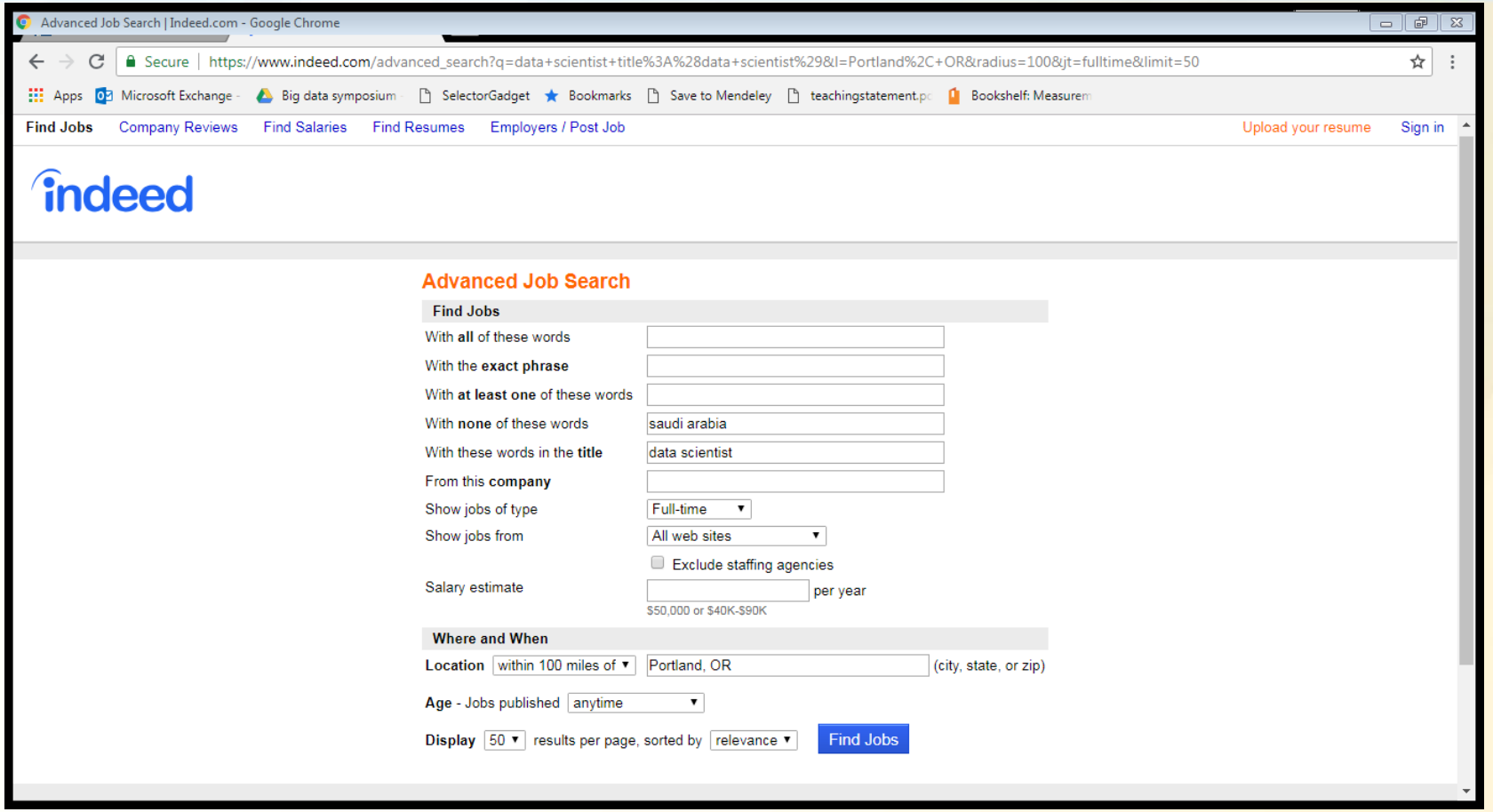
The “modern” solution

*What **skills** does industry most want in a data scientist?*

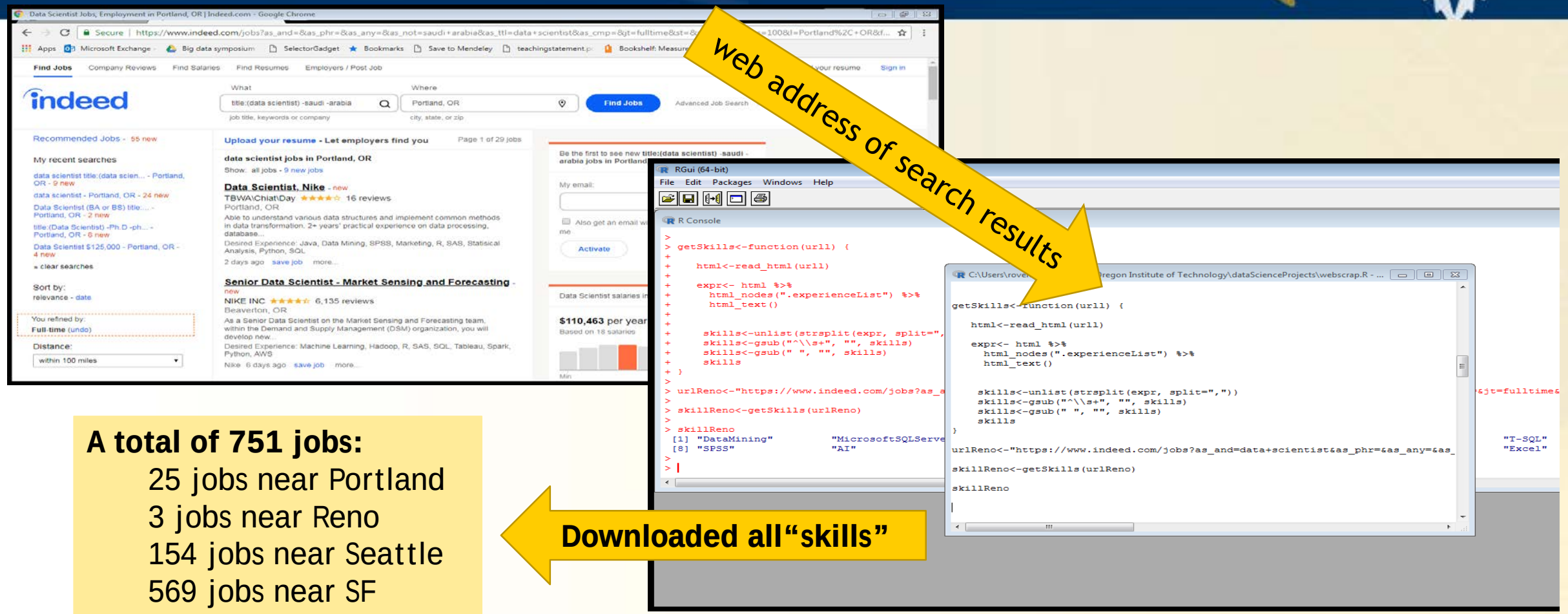
A data scientist might do some **web scraping**.

For example, we could use computer programming to scrape the “desired skills” part of “data scientist” job advertisements.

Step 1: Search for job ads for “data scientist”



Step 2: “scrape” the search results



A total of 751 jobs:

- 25 jobs near Portland
- 3 jobs near Reno
- 154 jobs near Seattle
- 569 jobs near SF

Step 3: Display the results

*What **skills** does industry most want in a data scientist?*

Using 751 job ads from Portland, Seattle, SF and Reno, we see…

Python	42%
Machine Learning	36%
SQL	27%
Hadoop	18%
Java	15%
Spark	15%

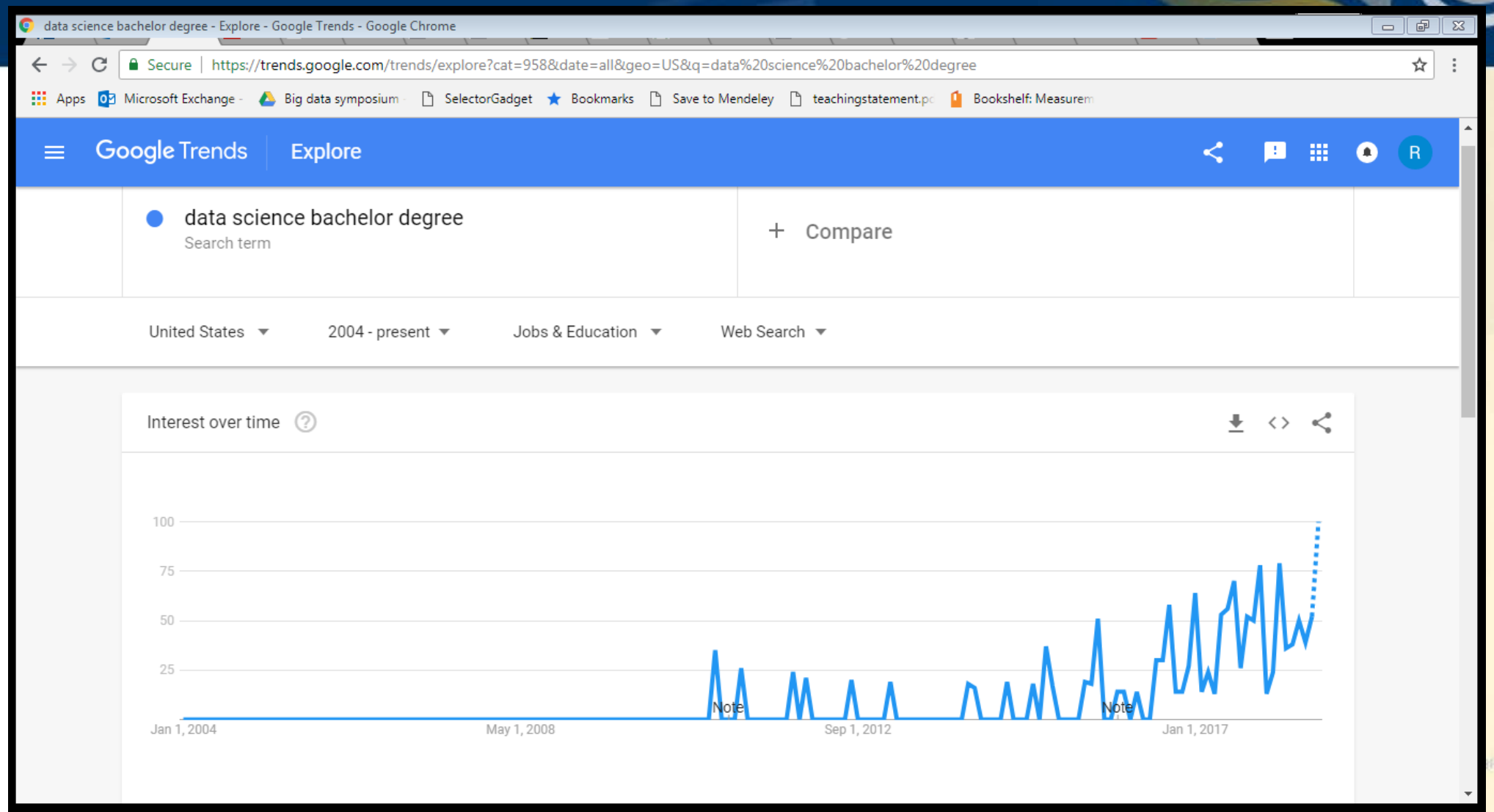
Another example of data science

The tools of data science make it quick and easy to figure out what skills our graduates should have to fit industry's needs.

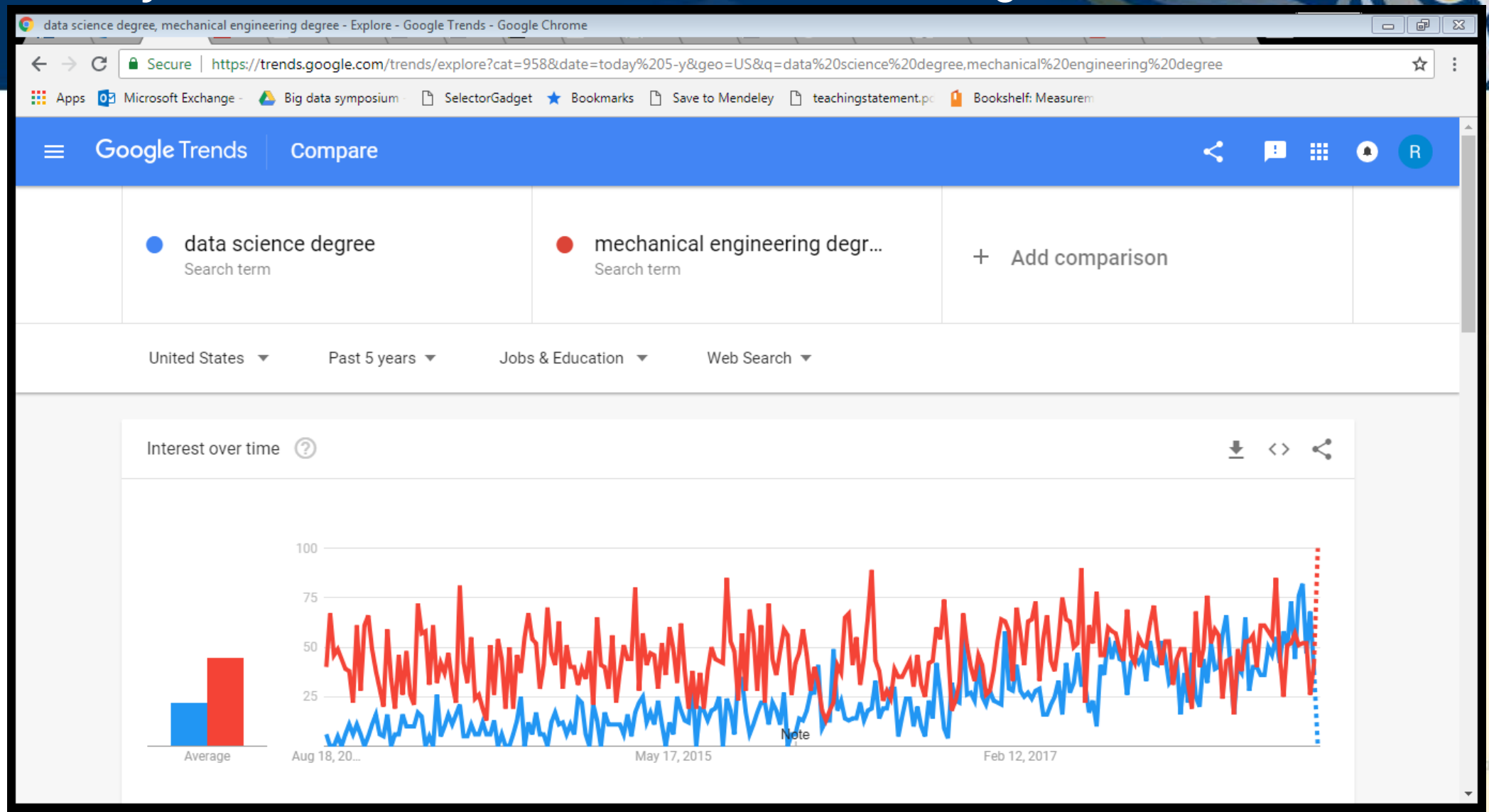
We can get a quick answer to related question…

Are people interested in earning a degree in data science?

Popularity of search terms over time via Google Trends



Popularity of search terms over time via Google Trends



Why Oregon Tech?

The Economist | Topics | Current edition | More

Fuel of the future

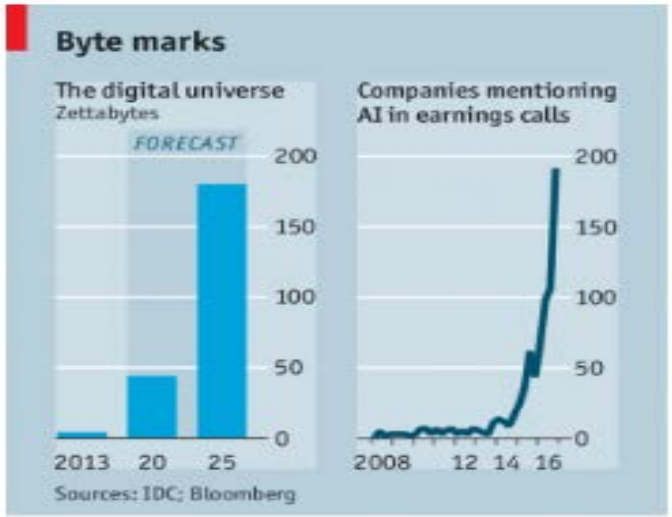
Data is giving rise to a new economy

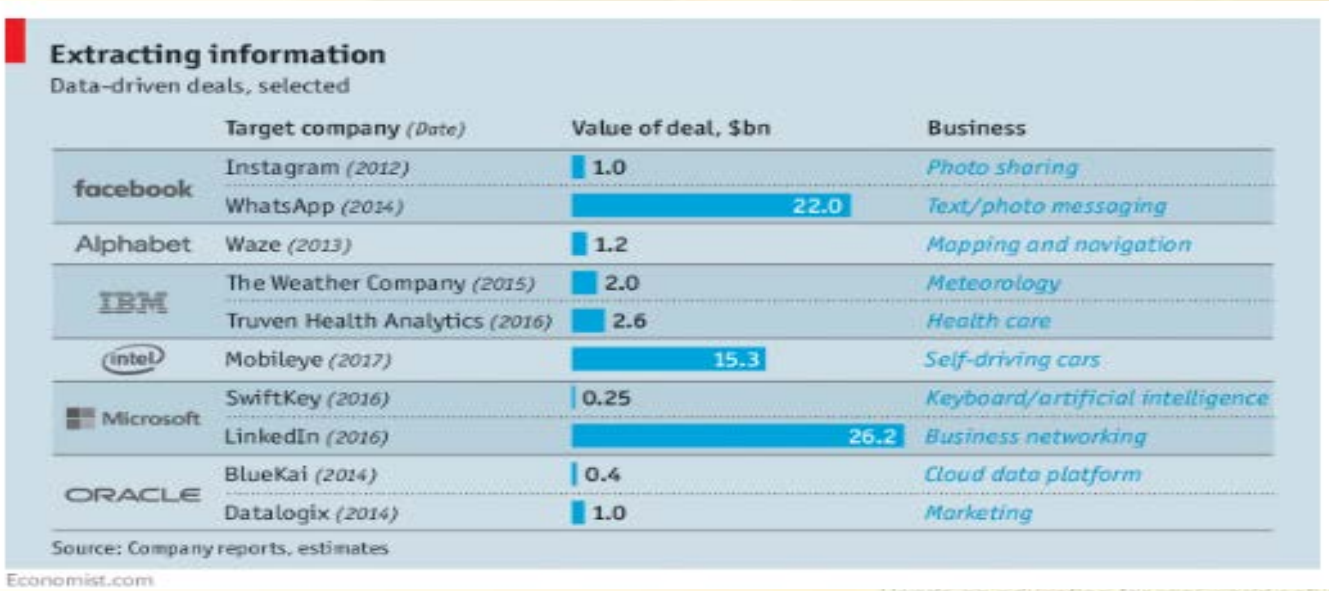
How is it shaping up?

Print edition | Briefing >
May 6th 2017

We are Industry's University, founded to support the postwar industrial development of Oregon and the Pacific Northwest.

The foundation of tomorrow's industry is **data.**





Why Oregon Tech?

It is absolutely essential that there is a strong data science presence at Oregon Tech.

Without opportunities to learn data science, other programs will weaken and fall behind the times.

Ex: GIS, Public Health, Energy Industry, Bio-informatics, Robotics (AI), etc…

Data and algorithms are increasingly a part of everyday life. All of our graduates deserve to have a basic understanding of how their personal data is being used to create credit scores, purchase recommendations, etc.

BS in Data Science

Initially hosted by the math department at KF campus, emphasizing "hands on learning for real world achievement"

An essential part of a data scientist's training is the completion of realistic projects. Our campus has a wealth of opportunities for projects and our small class sizes will help to ensure student success.

What will the program look like?

New math courses (41 units):

Data Science Methods I & II (8 units)

Statistical Machine Learning I & II (8 units)

Spatial Statistics (4 units)

Advanced Methods in Data Science (4 Units)

+

Junior Project Sequence (9 units)

+

Senior Project Sequence (8 units)

+

Existing math, CSET, management, geomatics and gen ed courses (147 units)

No accreditation agency *yet.*

Our program is designed according the guidelines for a BS in Data Science offered by a math department in the recent National Academy of Sciences Report *Data Science for Undergraduates: Opportunities and Options*

What will our graduates do?

- Graduate school

Computer science, statistics, data analytics, MBA, other

- Industry

*On August 13, 2018, there were **751 jobs** advertisements for "data scientists" within 100 miles of Reno, Portland, Seattle and SF.*

Who else is offering BS in Data Science?

- Found 51 four year programs in DS or very closely related fields nationally and internationally combined.
- Closest are in Univ of SF, UC Irvine, UC Berkeley, and BYU Idaho
- Masters in Data Analytics at Oregon State and focus at OHSU CSEE Masters

No 4 year program in any Oregon School

Industry Characteristics

- Education: Approx 51 programs at the BS level in the US and abroad (not counting IT and similar programs)
- Three categories of educational programs: Theoretical/Mathematical, Business Focused, Programming Focused
- Most programs are at the graduate level and expect students to have strong mathematical and programming skills (except for MBA programs which are more focused on Database and business presentation skills).

Industry Characteristics of Employers

- E-Commerce, Finance, Government, Healthcare, Science, Social Networking, Telecom, Utilities, Marketing, Insurance, etc…
- BLS expected change in employment between 2014 and 2024 is +54% growth.
- Structural Employment Gap is estimated between 140,000 and 190,000 jobs beyond the number of graduates produced for the 2016/2017 biennium.

Strengths

- Location: Only undergrad 4 year program in Oregon
- Market Segment: Students can focus in working with geographic, engineering, and health type of data while developing a strong background in business implementation.
- Students can specialize by obtaining additional training in related fields (example Pacific Power)
- Can fill the “applied role” for students to fill entry level jobs where they are well rounded in both theoretical and implementation of solutions.
  (DS is at the intersection of Applied Mathematics, Computer Science, and Business/IT. We are in unique position to offer a strong background in all three.)
- May partner with graduate programs in Oregon
- Small Campus: students won’t feel like just a number
- Strong applied math department and management/IT requires few new courses and leverages current inefficiencies in class enrolment.
- Encourages applied research at undergraduate level

Weaknesses

- Faculty turnover is high. The program can't be built on simply faith and must be supported.
- Direct competition in Northern California at research universities
- Location is not near many employment opportunities
- Workload in these courses is high for both faculty and students as each application requires the ability to create a specialized approach.

Opportunities

- Retention: May act as safety net for students who aren't interested in current CSET or IT focus areas
- Recruitment: Fast growing field with easy to sell products. (Machine Learning and AI)
- Placement: Mid Size companies needing well rounded data-scientists and Graduate Schools
- Partnerships: OSU? OHSU? Industry?
- Uniqueness: Cross training students with flexible program that can range from biology to engineering will allow competent, sector specific, graduates.

Threats

- Fast changing field
- Potential turnover in faculty
- Portland State University
- High mathematical/programming threshold for graduates
  - We won't encourage a weak program regardless of an increased draw.

Enrollment

- UC Berkley say explosive growth at the undergraduate level (from 300 to >1000 students in one year)
- University of SF is a campus of 6745 and had an enrolment of 85 students in the 2017 to 2018 cohort. (Approx 5% of total cohort)
- Other universities appear to average between 2% and 4% of total enrollment.
- At Klamath Falls, this is 40-80 students. Including potential dual majors would likely take this number to between 60-120 students in a four year period (cohorts of 15 to 30 students, sufficient to justify courses).
- This is supported by the trial run of machine learning.

Budget Analysis

	Year 0	Year 1	Year 2	Year 3	
Number of **new** students enrolled	-	5	5	10	
New Faculty (FTE = 1.28)	$0	$0	$0	$116,244	
Professional Development ($2500 per faculty)	$12,500	$12,500	$12,500	$15,000	
Upgrade faculty/lab computers	$5000	-	$25,000	-	
Marketing	$500	$1000	$1500	$2000	
Library Resources	$500	$100	$100	$100	
Profit	-$18,500	$48,835	$73,370	$77,996	**$181,701**