

Data Science Initiative

WORKSHOP KICK-OFF Session

DATA SCIENCE . CSUCHICO . EDU

DR. ROBIN DONATELLO * DR. RICK HUBBARD * DR. ESSIA HAMOUDA

24 MARCH 2017

W E L C O M E !

RICK HUBBARD (BIS DEPT)

W H Y ?

Why a Data Science Initiative at CSU, Chico?

Why DSI-Workshops?

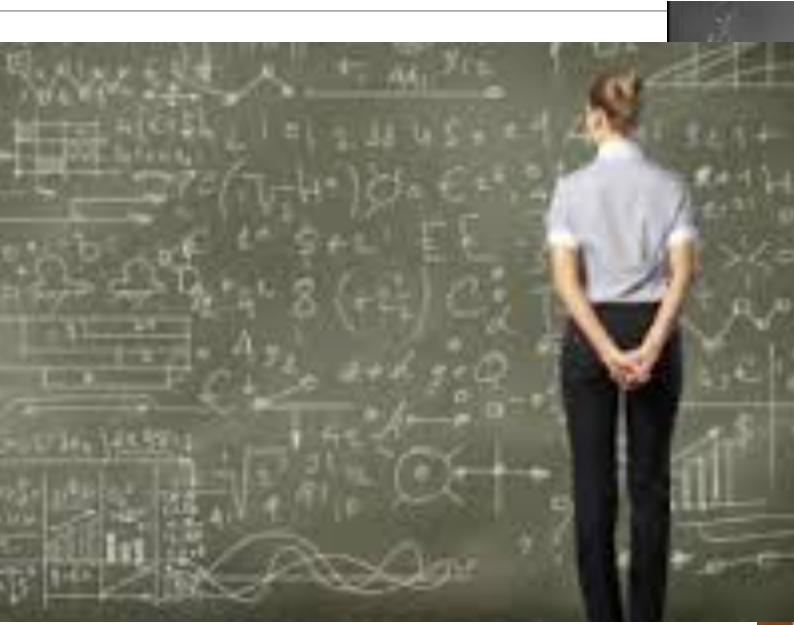
What's the Big Deal About Big Data, Data Science, and Analytics?



Big Data in the Past

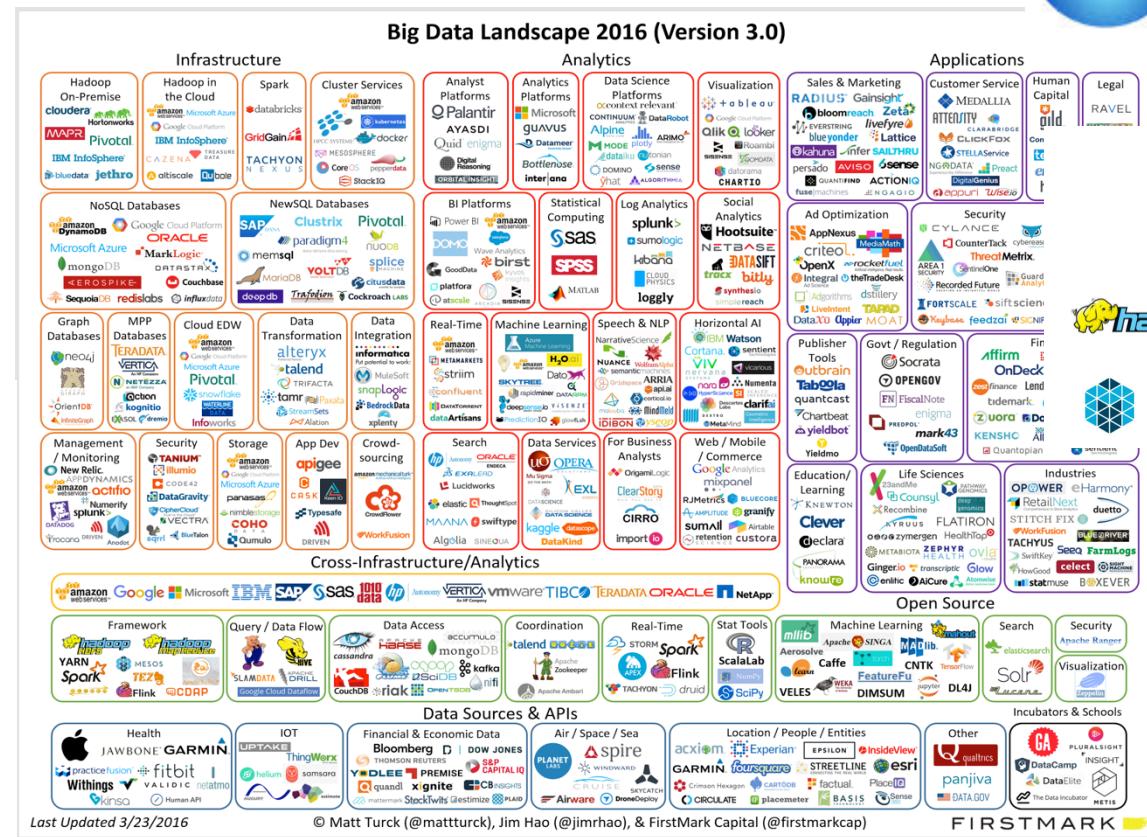
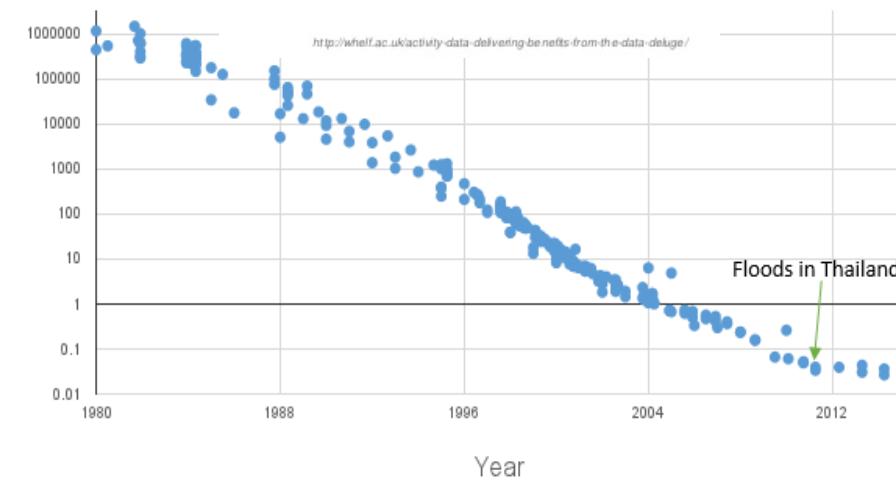
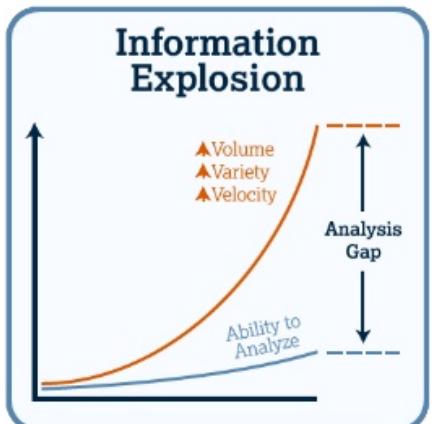


Dependent on Specialized Skills





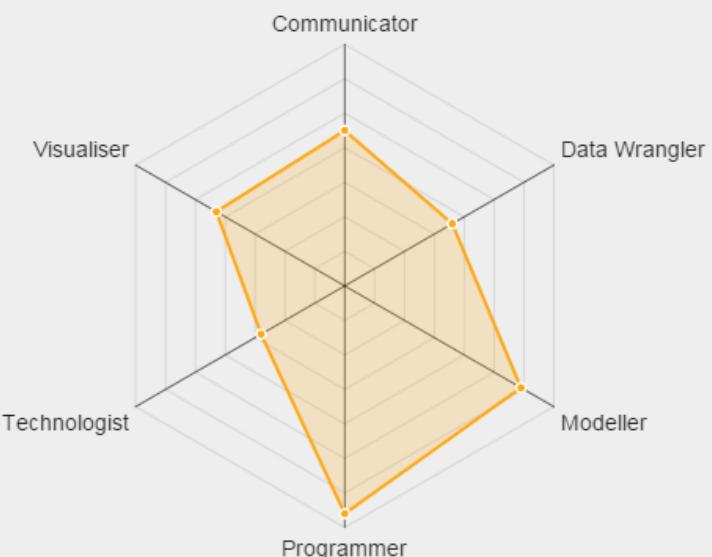
Things Have Changed



APACHE
HBASE



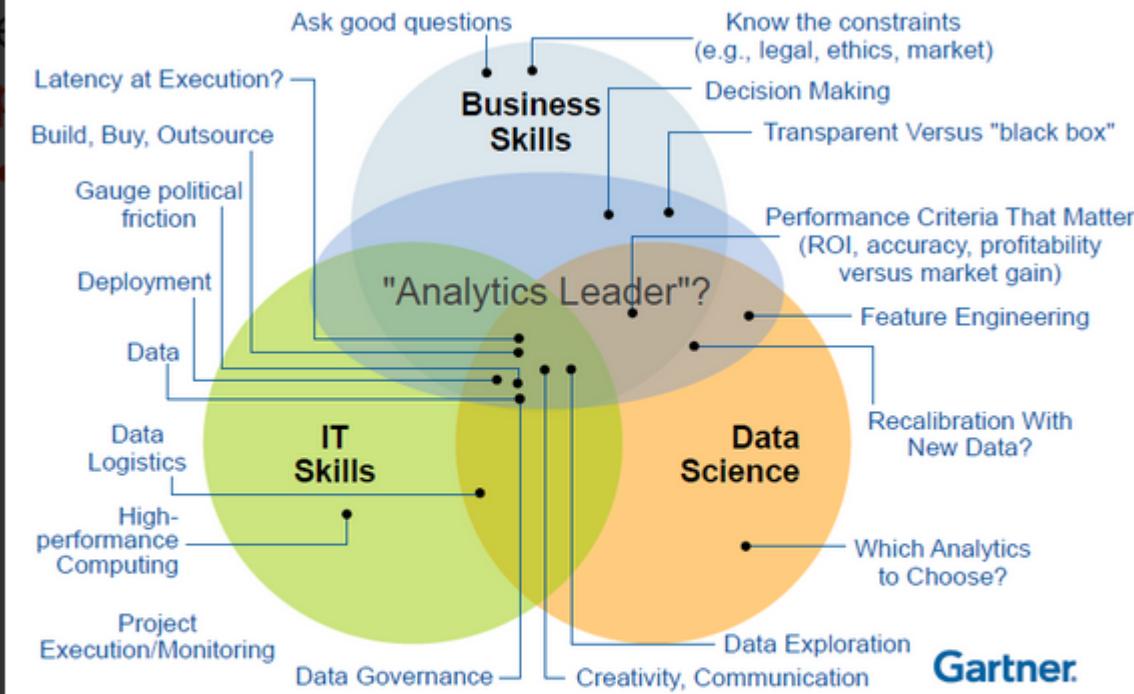
Today



Robin's Data Scientist Profile

When you are at the forefront of innovation, often the tools needed to solve a problem simply do not exist.
You may already be a master of multiple technical languages and enjoy adding languages to your skillset.
You combine carefully constructed analysis workflow, with robust pipelines to automate as much of the process as possible, but enjoy building applications from scratch.
You understand the value of planning, and know that thinking through an analysis is more efficient and less error prone than starting an analysis and seeing what sticks.
You always ensure that your customers are not misled by your work by providing thorough reporting, including lists of assumptions and descriptions of algorithms along with your findings.
You use software development approaches such as unit testing and version control to ensure that costly mistakes are not caused by your work.
Your cool head and rational approach serve as a great counterpoint to team members whose focus on the big picture can lead to key details being missed.

Driving the Success of Data Science Solutions: Skills, Roles and Responsibilities ...



Gartner



Karsten & Charles @twoARguys · Apr 2

#datascientists give a hell of a Venn diagram, hybrid skillsets like that won't come easy Train or buy? #GartnerBI pic.twitter.com/kT4NHG68Ua

Reply Retweet Favorite

Flag media

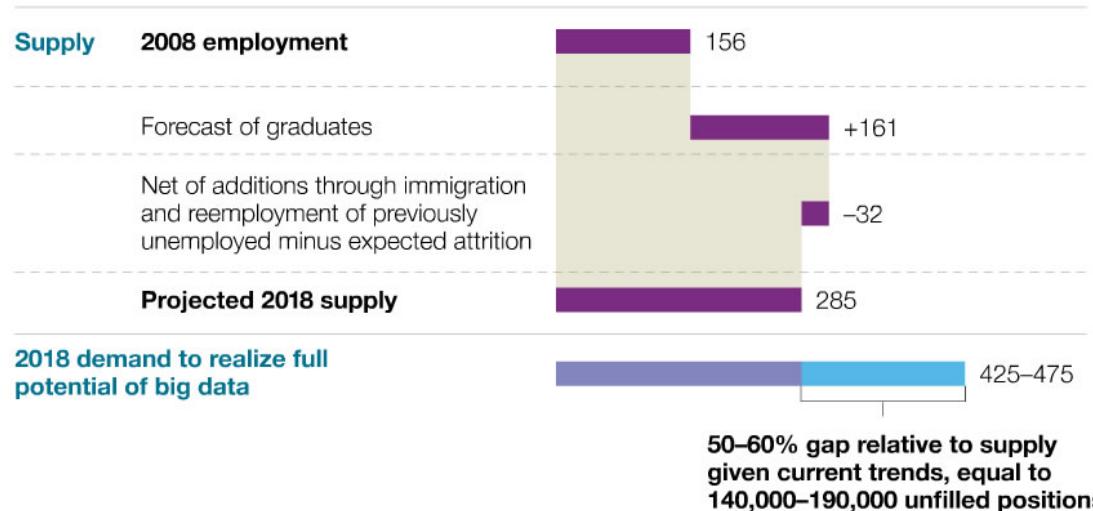
Tomorrow

In Analytics Economy, Demand for Data Scientists Outpaces Supply, McKinsey Says

By Benjamin Herold on December 8, 2016 4:27 PM

Demand in the United States for people with deep expertise in data analysis could be greater than its projected supply in 2018.

Deep analytical talent, thousands of FTEs¹



¹Deep analytical talent are people who have advanced training with statistics or machine learning.
FTE = full-time equivalent.

Source: Dun & Bradstreet; company interviews; US Bureau of Labor Statistics; US Census Bureau; McKinsey Global Institute analysis

Data Science Initiative at CSU, Chico

Prepare Students for Success in the Analytics Economy

Support Campus in Their Efforts
Faculty
Administration

Enable the Community to Benefit from Applied Data Science

Today's Agenda

- WHY?
- WHAT?
- HOW?
- Topical Request Exercise
- Hands on Intro to R

WHAT?

What is the Chico State Data Science Initiative?

*An interdisciplinary community that
provides support and training for Data Science*

ESSIA HAMOUDA (CS DEPT)

How we got here

- Concept of a Data Science program initiated by Provost Bell Wei
- Robin was hired (in part) to develop this program
- Tasked with building curriculum to train Data Scientists (or at least prepare them for grad. school in Data science)
- A cohort was formed, work began
- Encouragement and enthusiasm has come from all corners of campus and community
- A report on the current state of Data Science Interest at Chico State is available on our [website](#) (and still open to contribution!)

A Resource for Campus & Community

The DSI aims to help organize the Data Science community by

- Providing a centralized location for information sharing and networking
- Providing a well-known nexus to obtain, maintain, share and enhance relevant skills, knowledge and abilities
- Providing a comprehensive overview of data science related curriculum Connecting researchers and students
- Connecting community partners with campus resources



Who is the DSI for?

**#EVERYONE
IS WELCOME**



You're
invited

**EVERY
ONE
BELONGS**

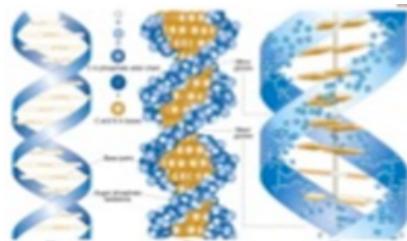
Four stylized human figures, each a simple outline of a person in a different color (orange, yellow, green, and blue) standing side-by-side.

Applications/Examples

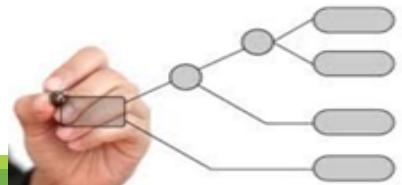
- Business Analytics / Business Intelligence



- Bioinformatics / Healthcare Informatics



- Data Modeling / Analytics Methods



- Geoinformatics and GIS
- AG-Informatics



- Data Capture/Data Wrangling/ Data preparation.



- Artificial Intelligence / Machine Learning



The DSI - Workshops

- To these ends, an important focus of the DSI are “DSI-Workshops” which offer:
 - Timely resources, practical advice and germane information regarding areas-of-interest to all Data Science users—from novice to expert practitioner
- These workshops are one small step for the DSI, but **one giant leap** for the community



What are the DSI-Workshops?

- Short events to enable people from across campus and community to learn about data science/analytic techniques and software
- Hands on - active participation
- Drop in as interested
- Variety of skill & knowledge levels
- Community Driven—All Volunteers
- Training by experts (you) in the field
- Open source training materials



What the DSI-Workshops Include...

- Learning: techniques/material/career info
- Doing: Walk through installation / analysis
- Problem-Solving
- Assistance
- Advice
- Exchange of Ideas:
 - Have an idea for a talk?
 - Want a place to demo your research?
- Community / Networking



Technical Topics (Examples)

- R & R Studio
- Python
- SQL
- Version Control
- Shell scripting
- Regular expressions
- Excel
- Tableau
- Amazon Web Service
- Dynamic web applications
- Scientific Computing
- Data Analysis
- Reproducible Research
- Academic website hosting using GitHub Pages
- Sentiment analysis

HOW?

Community building

Community organizing

Skills and competency development

ROBIN DONATELLO (MATH DEPT)

A Point in Space and Time (with Cookies!)

A shared collaborative space and consistent dedicated time for people to meet, learn, share and do Data Science

Open and social collaborative coding / analysis sessions

Bring your data and coding problems, we can solve them together.

TARDIS Parking Available



Formal Curriculum Development

Currently Offering

Introduction to R (MATH 130) - 1 unit

Introduction to Data Science (MATH 385 / CSCI 385) - 3 units

More in other areas around campus

In Progress

Advanced Data Science (MATH 485 / CSCI 485) - 3 units. (Sp 18)

Undergraduate Certificate in Data Science - (Fall 18)

Future Consideration

Option in Business Analytics

Masters Program / Graduate Certificate in Data Science

Spring '17 Workshop Series - Intro to R

- 3/24: Getting started with R and RStudio
- 4/7: Preparing Data for Analysis - How to structure data now to avoid losing your mind later (sometimes call ““Tidy Data””)
- 4/14: Exploring your data - Get your data into R and start learning
- 4/21: How to get help - aka translating horrible error messages and writing MWE
- 4/28: Common statistical analysis techniques using R
- 5/5: Visualizing your Data using the Grammar of Graphics (ggplot2)
- 5/12: Networking event

Agenda Driven by Community

What do you want to learn?

What can you contribute?

Topical Elicitation

Next Steps

Hands on walkthrough on how to get started with R and RStudio

Been there, done that? Hang out in the back and chat about collaboration possibilities

Getting Started with R and RStudio

How to become a useR

Tutorial Materials

<http://datascience.csuchico.edu/materials/2017-03-24.html>

Need a Laptop? We have a few.

Don't have R installed? Borrow a Laptop. Do now, Install later.

Volunteer TAs: Ricardo Aguilar, Aaron Shaffer

Need Guest Wireless access?