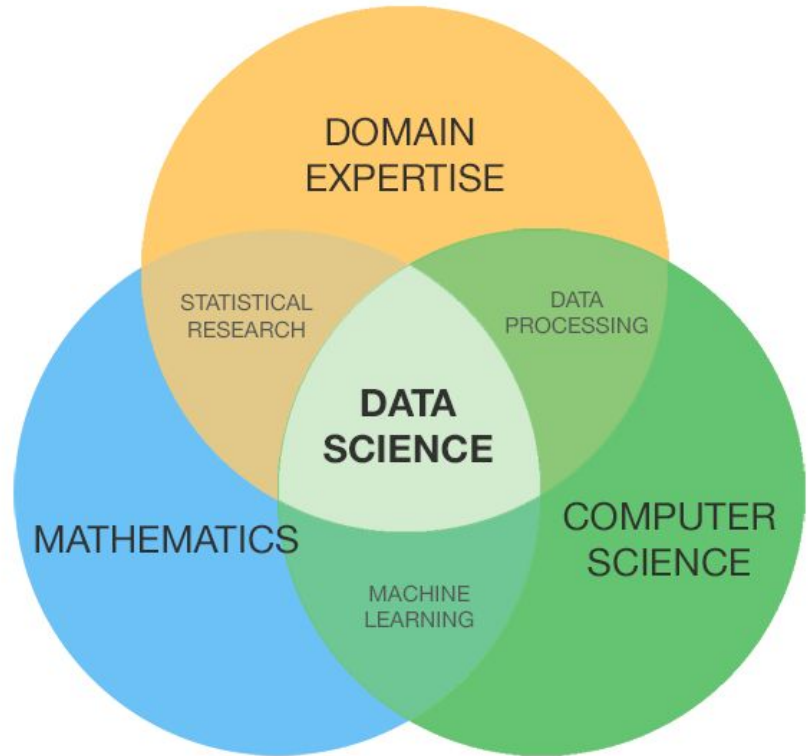




The Fun and Utility of Data Science

What Is Data Science?

Extraction of knowledge from data.



Source: Palmer, Shelly. *Data Science for the C-Suite*.
New York: Digital Living Press, 2015. Print.

What Does Data Science Do?

1. Collect
2. Clean up
3. Describe
4. Explain
5. Infer
6. Predict
7. Present
8. Defend

Example:

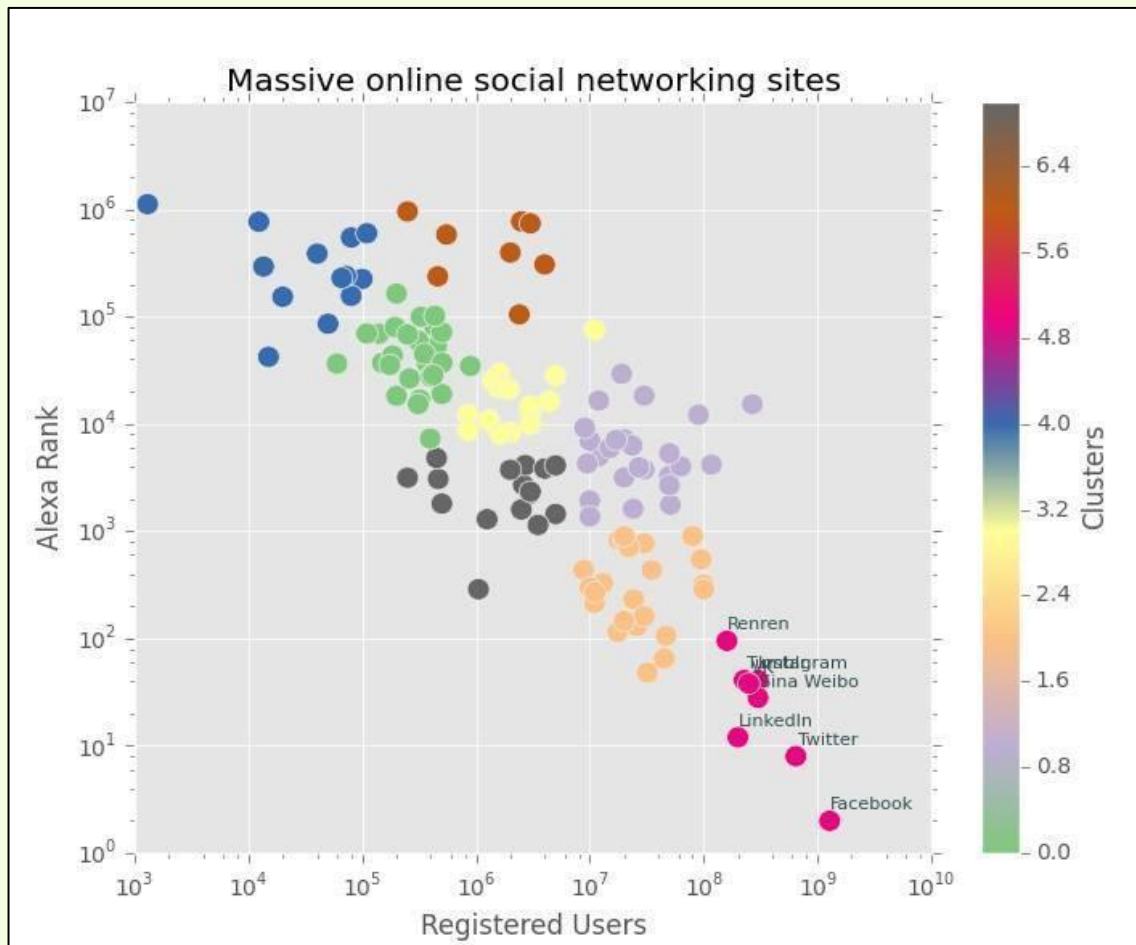
Massive online social
networking sites

Source:

Wikipedia

Result:

Higher-ranked sites have
more users



Data Science Topics

- Numeric data analysis
- Signal processing
- Text data analysis (natural language processing)
- Statistical inference
- Information integration
- Complex network analysis
- Data visualization
- Machine learning (including classification)
- Data mining
- Big data

Example:

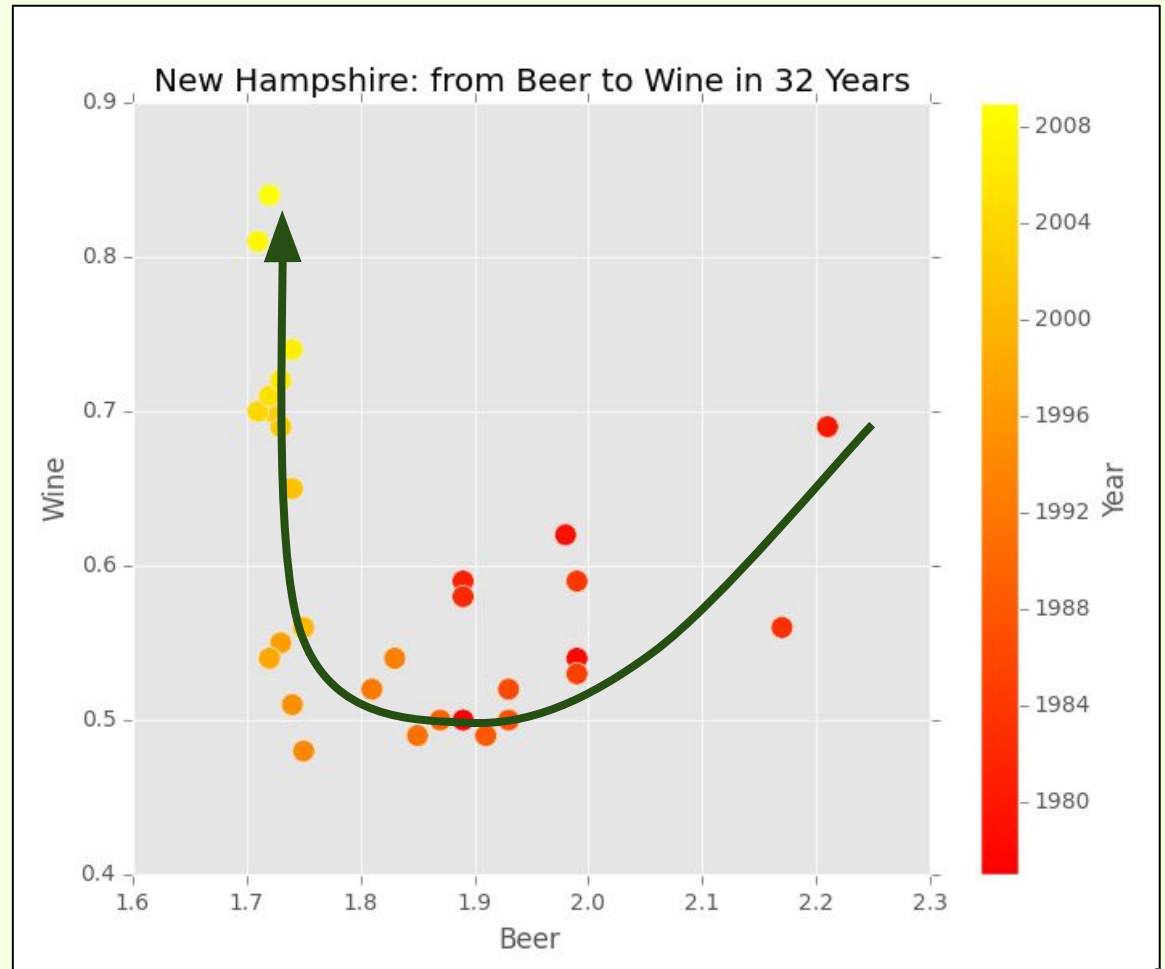
Beer consumption in the
USA

Source:

NIAAA report

Result:

New Hampshire switched
from beer to wine in 32
years



Data Science vs Business Intelligence

Business Intelligence engineers traditionally **make tools** for others to analyze data with. BI engineers **do not analyze the data**. Data scientists will both make and analyze using what they made.

Example:

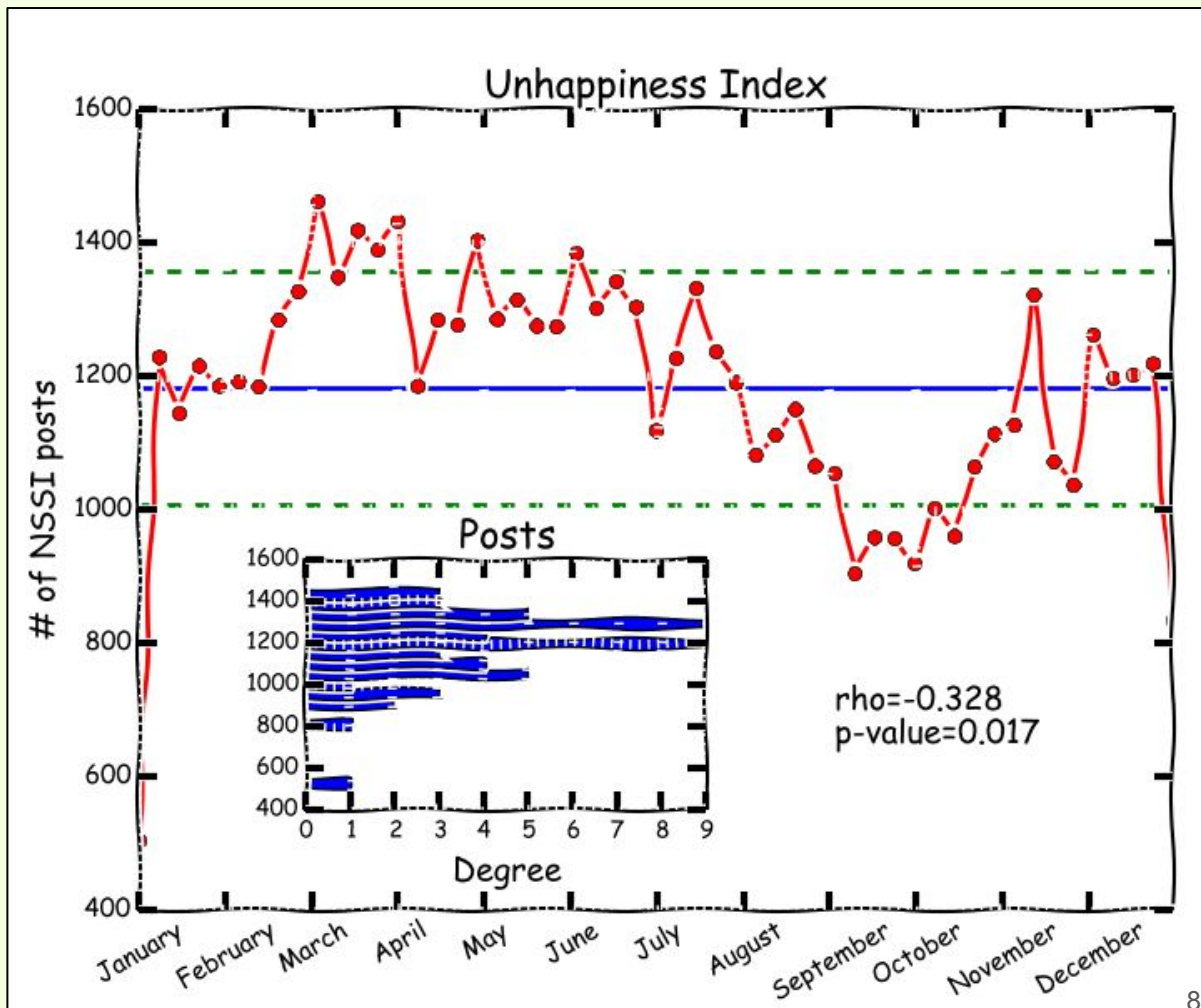
Unhappiness Index

Source:

LiveJournal blogging
platform

Result:

People are unhappy in
March – July



Data Science vs Statistics

Statisticians are traditionally content with the assumption (condition) that **all their data will fit in main memory at the same time**. Statisticians traditionally **used math** to squeeze as much information as possible from small numbers of observations or features, and then do **inference**. Data scientists use and create new software engineering tools to handle very large datasets.

Example:

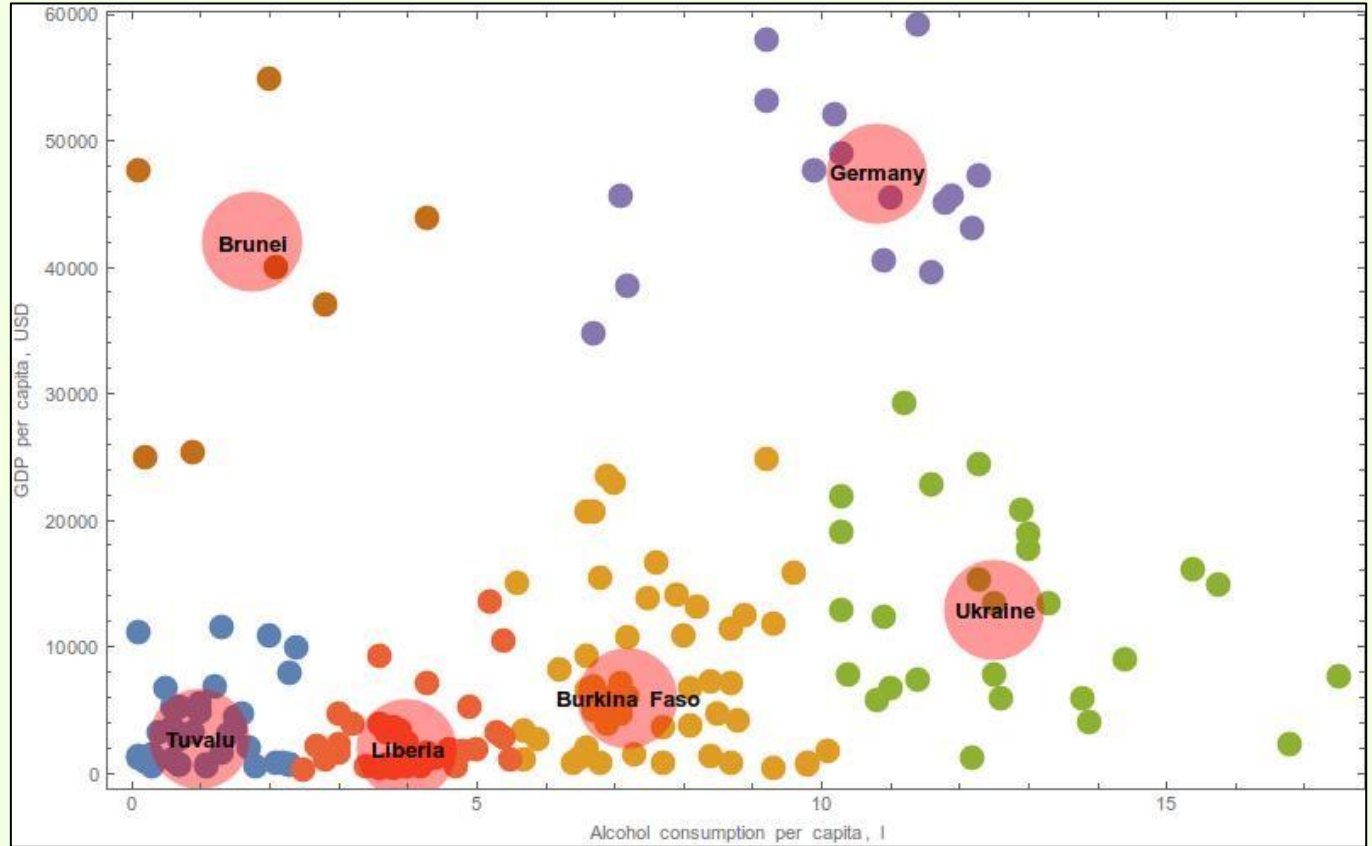
Alcohol
consumption vs
GDP

Source:

Wikipedia

Result:

I'd rather lived in
Germany!



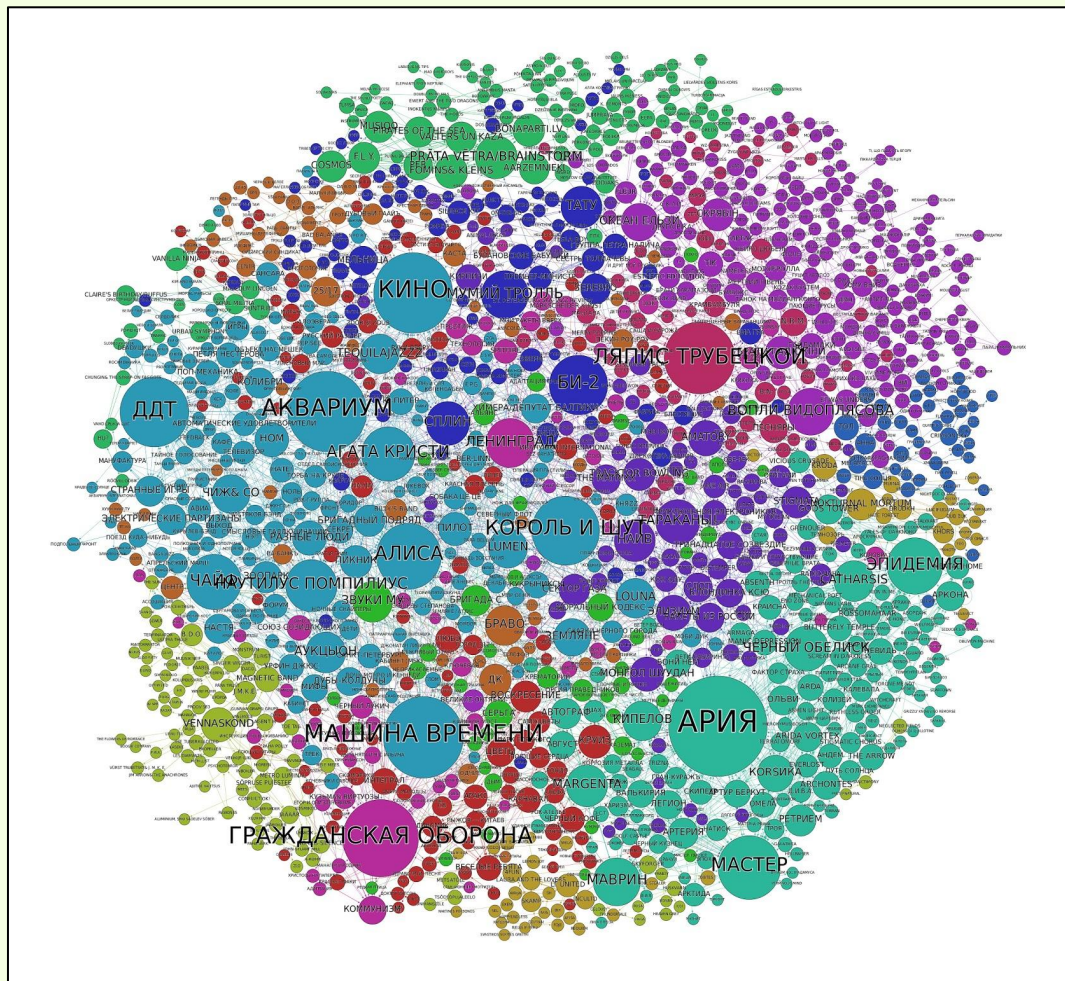
Data Science and Databases

Database programmers and administrators bring useful skills to data science but they are traditionally **focused on one data model: relational**. Handling graphs' nodes and edges (e.g., pagerank), images, video, text, as well as SQL when appropriate, are more like data science.

Russian non-academic music groups

Wikipedia

Success depends on sharing



Data Science and Visualization & Business Analytics

Visualization experts and business analysts bring skills but are traditionally **not concerned with massive scale** like hundreds or thousands of machines.

Example:

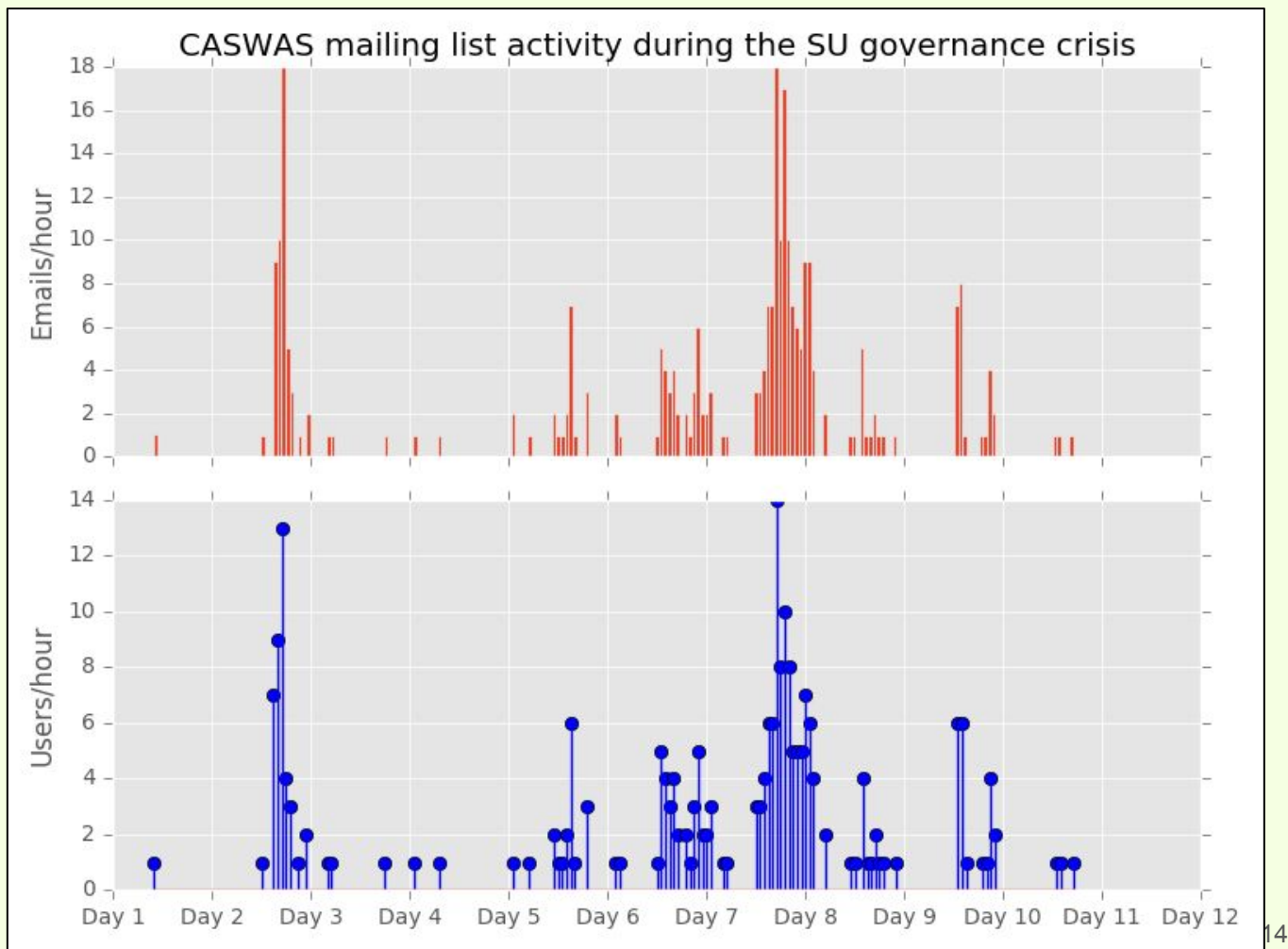
SU governance
crisis 2015 – 2016

Source:

Faculty mailing list

Result:

People sleep at
night



Data Science and Machine Learning

Machine learning is similar to data science but it's **a small fraction of it**. Data science is also about getting of data, cleaning, exploring, and making interactive visualizations and data products for yourself and for others to use.

Example:

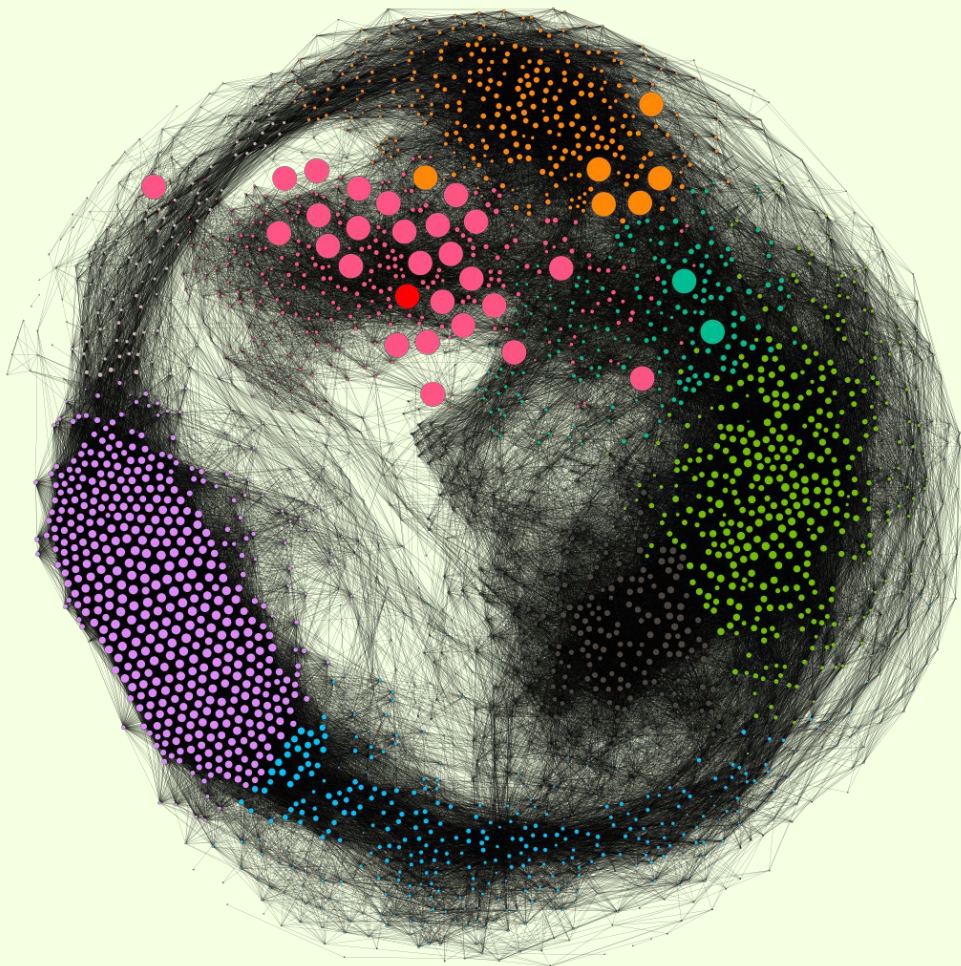
All SU course map

Source:

SU course catalogs

Result:

Courses form content clusters
(clickable!)



Jobs

- 700 data science positions on monster.com in Boston area
- The average salary for a data scientist is \$118,709 versus \$64,537 for a skilled programmer (glassdoor.com)