

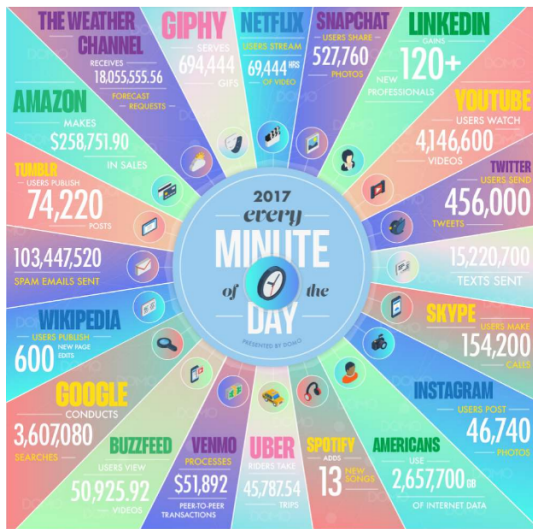
# All Things "Data": Data Science, Data Analytics, Big Data ...

Andrey Skripnikov

Department of Mathematics  
University of Houston

June 20, 2018

# Root of it all: Data.

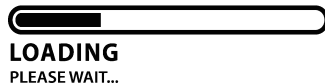


# Big Data: How Big is "Big"?

- **"Big"** is a **moving target**.

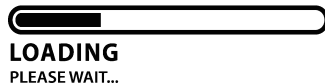
# Big Data: How Big is "Big"?

- "Big" is a **moving target**.
- "Big" - when size becomes **a challenge**.



# Big Data: How Big is "Big"?

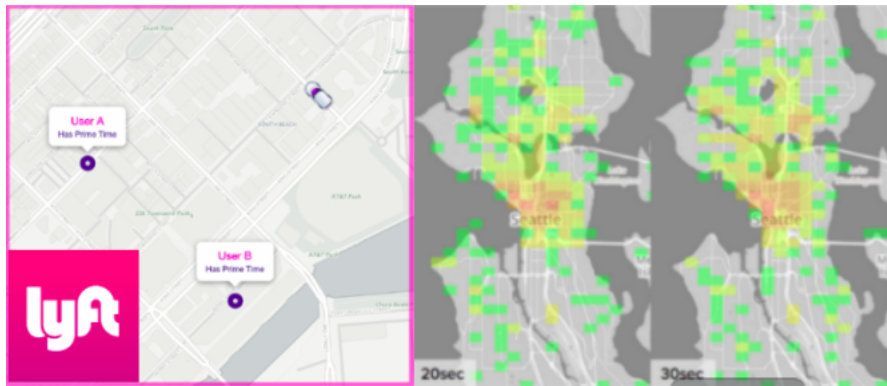
- "Big" is a **moving target**.
- "Big" - when size becomes **a challenge**.



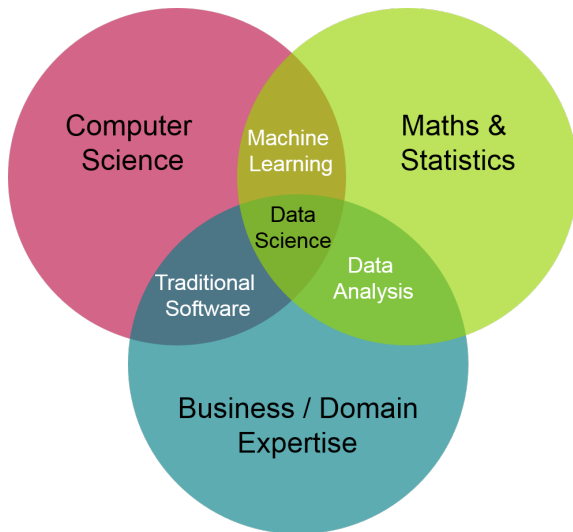
have to learn a **new host of tools**.

# Tech Companies: Uber/Lyft.

- **A/B testing.**
- **Statistical simulation.**

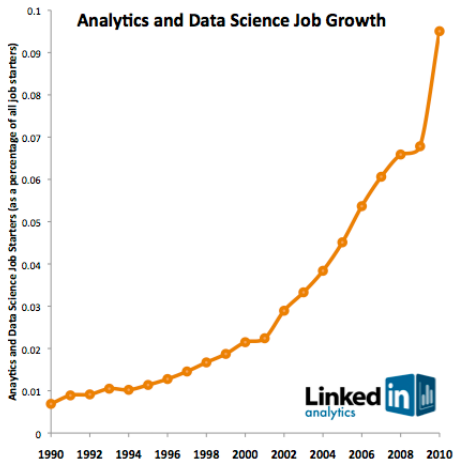


# What is Data Science? Venn Diagram.



Might not be a real field, but has REAL JOBS.

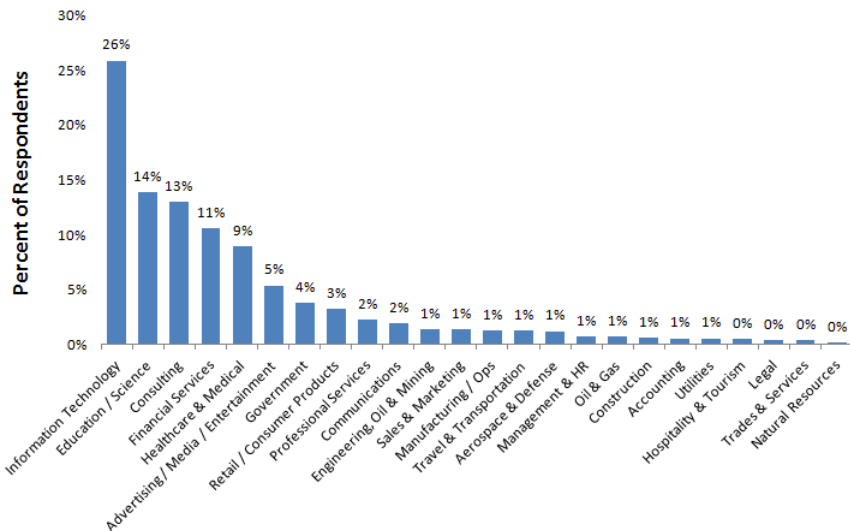
**"So even if data science isn't a 'real field', it has REAL JOBS."**  
(R. Schutt, C. O'Neil, "Doing Data Science")





Might not be a real field, but has REAL JOBS.

## Data Scientists Work in Many Industries



# Big Three: Data Science, Big Data, Data Analytics.



# Big Three: What your **salary** will look like?

## #5. What your salary will look like



Data Science  
Professional

**\$123,000**



Big Data  
Professional

**\$88,000**




Data Analytics  
Professional

**\$61,000**

# Big Data Specialist, Data Engineer.

**Big Data Specialist**, else called **(Big) Data Engineer**:





**Languages**

SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C++, Perl

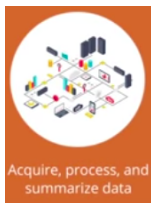
**Skills & Talents**

- ✓ Database systems (SQL & NO SQL based)
- ✓ Data modeling & ETL tools
- ✓ Data APIs
- ✓ Data warehousing solutions

	Data Engineer
Programming	●
Data Visualization & Communication	●
Statistics	●
Data Preparation	●
Machine Learning	●
Software & Databases	●
Calculus & Linear Algebra	●

# Data Analyst.

## Data Analyst:



Big data tools  
(Hive, Pig)



Knowledge of mathematical  
statistics

	Data Analyst
Programming	
Data Visualization & Communication	
Statistics	
Data Preparation	
Machine Learning	
Software & Databases	
Calculus & Linear Algebra	

# Data Scientist: Skills/responsibilities.

**Data Scientist** (on **TOP** of **Data Analyst** skills & duties):

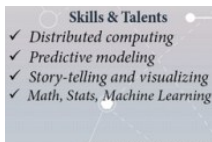


The Ten Most Common  
Data Science Skills in Job Postings

Skill	Percentage of Job Listings
Python	72%
R	64%
SQL	51%
Hadoop	39%
Java	33%
SAS	30%
Spark	27%
Matlab	20%
Hive	17%
Tableau	14%

Source: Glassdoor Economic Research

glassdoor



	Data Scientist
Programming	●
Data Visualization & Communication	●
Statistics	●
Data Preparation	●
Machine Learning	●
Software & Databases	●
Calculus & Linear Algebra	●

# Final Thoughts.

- ① **Data Science** can be used **everywhere**.
- ② Not a specialization, but a **state of mind**.
- ③ **Ask questions**, use **data**, do a **project**.

# Final Thoughts.

- ① **Data Science** can be used **everywhere**.
- ② Not a specialization, but a **state of mind**.
- ③ **Ask questions**, use **data**, do a **project**.

If you do a project:

- ① **Mess up**. Mess up **A LOT**. **Learn** from it.
- ② **Own up**.
- ③ **Seek the truth** (I mean, **Google** it).