Root of it all: Data.
"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.

- Computer Science
- Maths & Statistics
- Traditional Software
- Data Analysis
- Business / Domain Expertise

Machine Learning
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”)

Andrey Skripnikov  (University of Houston)  All Things ”Data”: Data Science, Data Ana... June 26, 2018  6 / 13
Might not be a real field, but has REAL JOBS.

Data Scientists Work in Many Industries

Percent of Respondents

Information Technology: 26%
Education / Science: 14%
Consulting: 13%
Financial Services: 11%
Healthcare & Medical: 9%
Government: 5%
Consumer Products: 4%
Professional Services: 2%
Communications: 2%
Engineering: 1%
Oil & Gas: 1%
Manufacturing / Ops: 1%
Travel & Transportation: 1%
Aerospace & Defense: 1%
Management: 1%
HR: 1%
Oil & Gas: 1%
Construction: 0%
Accounting: 0%
Utilities: 0%
Legal: 0%
Hospitality & Tourism: 0%
Trades & Services: 0%
Natural Resources: 0%
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:

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

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

Andrey Skripnikov (University of Houston) All Things "Data": Data Science, Data Analytics

June 26, 2018 10 / 13
Data Analyst:

- Acquire, process, and summarize data
- Design and create data reports using various reporting tools

Programming

Data Visualization & Communication

Statistics

Data Preparation

Machine Learning

Software & Databases

Calculus & Linear Algebra

Big data tools (Hive, Pig)

Knowledge of mathematical statistics
Data Scientist: **Skills/responsibilities**.

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

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

---

**The Ten Most Common Data Science Skills in Job Postings**

<table>
<thead>
<tr>
<th>Skill</th>
<th>Percentage of Job Listings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python</td>
<td>72%</td>
</tr>
<tr>
<td>R</td>
<td>64%</td>
</tr>
<tr>
<td>SQL</td>
<td>51%</td>
</tr>
<tr>
<td>Hadoop</td>
<td>39%</td>
</tr>
<tr>
<td>Java</td>
<td>33%</td>
</tr>
<tr>
<td>SAS</td>
<td>30%</td>
</tr>
<tr>
<td>Spark</td>
<td>27%</td>
</tr>
<tr>
<td>Matlab</td>
<td>20%</td>
</tr>
<tr>
<td>Hive</td>
<td>17%</td>
</tr>
<tr>
<td>Tableau</td>
<td>14%</td>
</tr>
</tbody>
</table>

**Skills & Talents**

- Distributed computing
- Predictive modeling
- Story-telling and visualizing
- Math, Stats, Machine Learning

Source: Glassdoor Economic Research.
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).