

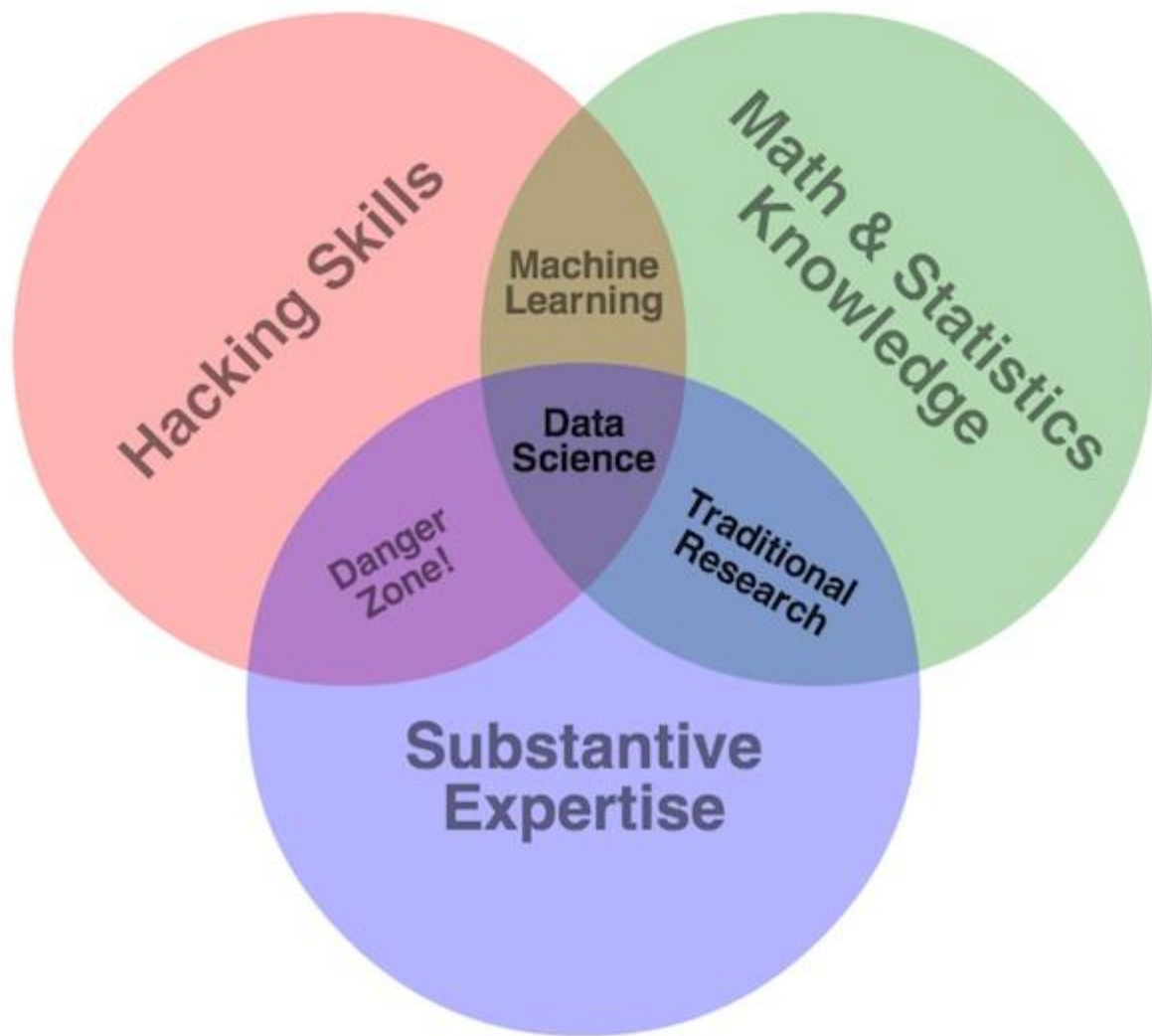
Getting Started with Data Science

Portia Burton
PLB Analytics
3/3/2015

Who am I?

- Founder of [PLB Analytics](#)
- Founder and organizer of <http://www.meetup.com/Portland-Data-Science-Group/>
- Data Science instructor

What is Data Science?



Data Science is a mixture of

- Programming
- Math and Statistics
- Domain Knowledge (like Ford cars, or Vietnamese coffee)

Major Fields of Data Science

- Natural Language Processing
- Data Visualization
- Machine Learning
- Analyzing Data

5. Text Mining / NLP

8. Data Ingestion

6. Visualization

4. Machine Learning

10. Toolbox

7. Big Data

2. Statistics

3. Programming

1. Fundamentals



Data Visualization

- [Matplotlib](#) or [ggplot](#) (Python and R)
- [D3.js](#) (JavaScript)
- [Plot.ly](#) (Visualization as a Service)

Analyzing Data

- Python ([Pandas](#))
- [R](#)

Python vs. R (The Great Debate)

R

- Used in academia
- More mature statistical environment
- Mostly used to analyze data

Pandas

- Used mostly in industry
- Younger platform
- Python is used to build applications

Machine Learning

- <http://scikit-learn.org/stable/>
- Used in find spam
- Evernote uses K-nearest neighbor algorithm to find where to place your notes

Natural Language Processing

- Natural Language Tokenization
- Sentiment Analysis
- Lexical Diversity

How to Get Started

Data Visualization: <https://github.com/mbostock/d3/wiki/Gallery>

Analyzing Civic Data: <https://github.com/jvns/pandas-cookbook>

Machine Learning:
<http://www.kaggle.com/>

