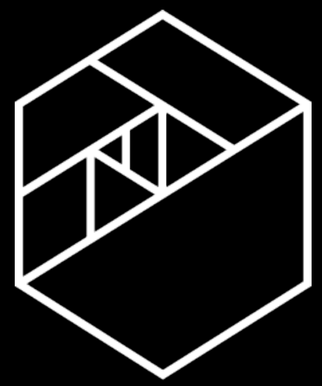
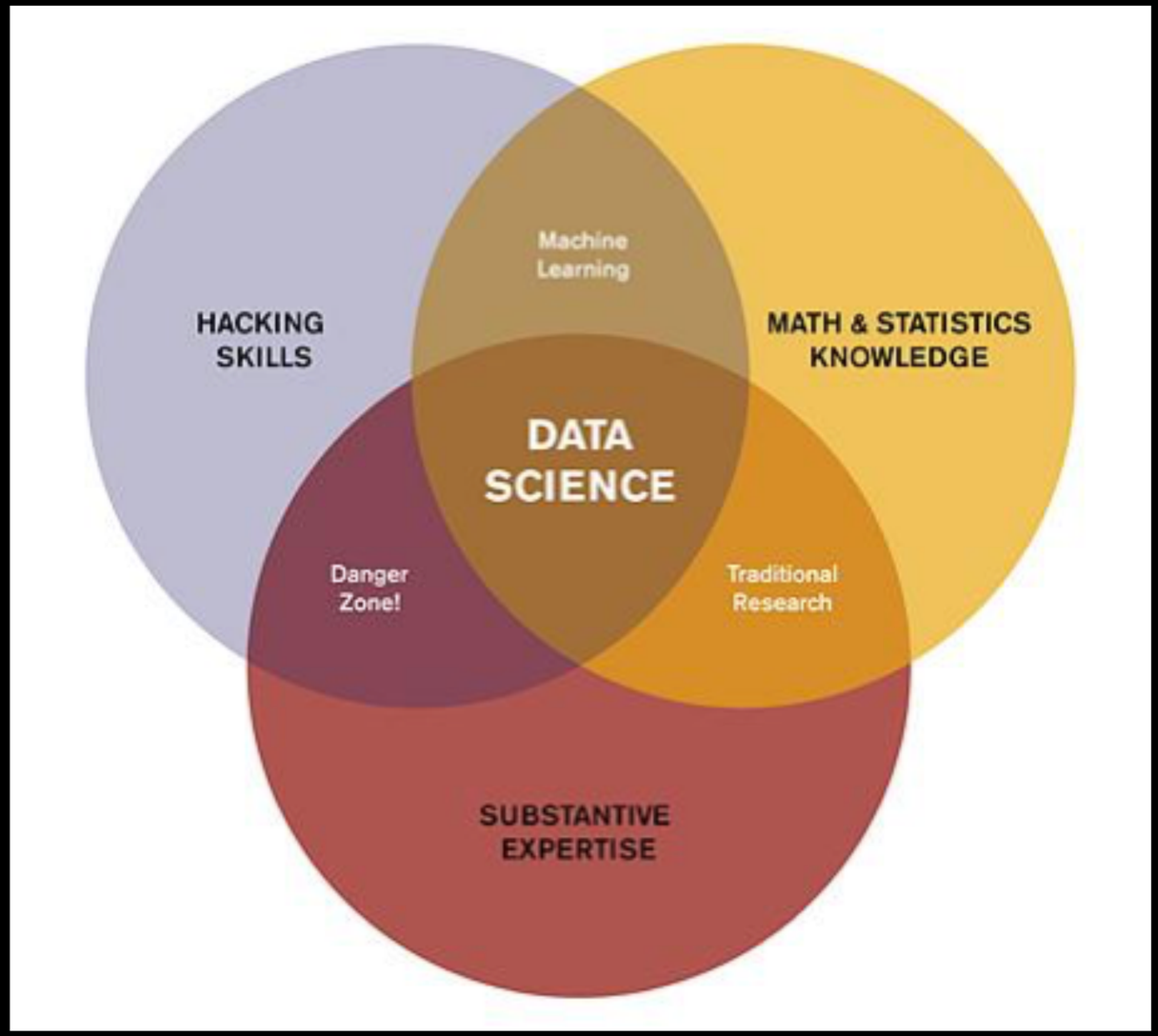
What is a Data Scientist

And how will you become one?



METIS

Skills of a Data Scientist



Skill domains of a data scientist

Statistics / Machine Learning
Programming
Communication
Design

Who can be a data scientist?

- Personality Traits

Who can be a data scientist?

Personality Traits

Curiosity
Creativity
Grit

Who can be a data scientist?

Personality Traits

Continuous drive to learn
Creativity
Grit

Who can be a data scientist?

Personality Traits

Continuous drive to learn
Tons of different angles & ideas
Grit

Who can be a data scientist?

Personality Traits

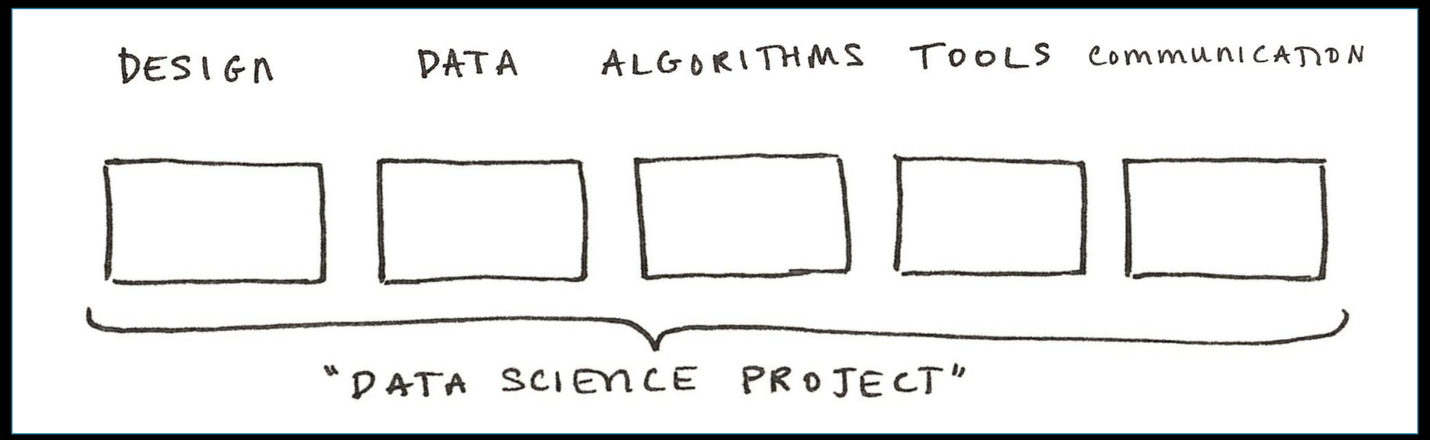
Continuous drive to learn
Tons of different angles & ideas
Ability to push through walls

Bootcamp model

Training
Repetition
Culture/community

Training

Project-based model



Training

Obtain

Scrub

Explore

Model

Interpret

Communicate

Training

Obtain	Data acquisition
Scrub	Data exploration
Explore	Domain awareness
Model	Machine learning
Interpret	Statistics
Communicate	Visualization & Web

Training

Tool selection

Comfort & facility with a core toolbox
Awareness of options
Some curveballs

Training

Tools

Python when possible

Training

Tools

Python, ipython notebook, git/github

Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron

Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flat files), PostgreSQL, mongoDB

Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flat files), PostgreSQL, mongoDB
numpy, scipy, matplotlib, pandas

Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flat files), PostgreSQL, mongoDB
numpy, scipy, matplotlib, pandas
statsmodels, scikit-learn

Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flat files), PostgreSQL, mongoDB
numpy, scipy, matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, JS, d3.js

Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flat files), PostgreSQL, mongoDB
numpy, scipy, matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, JS, d3.js
Amazon Web Services, Flask

Training

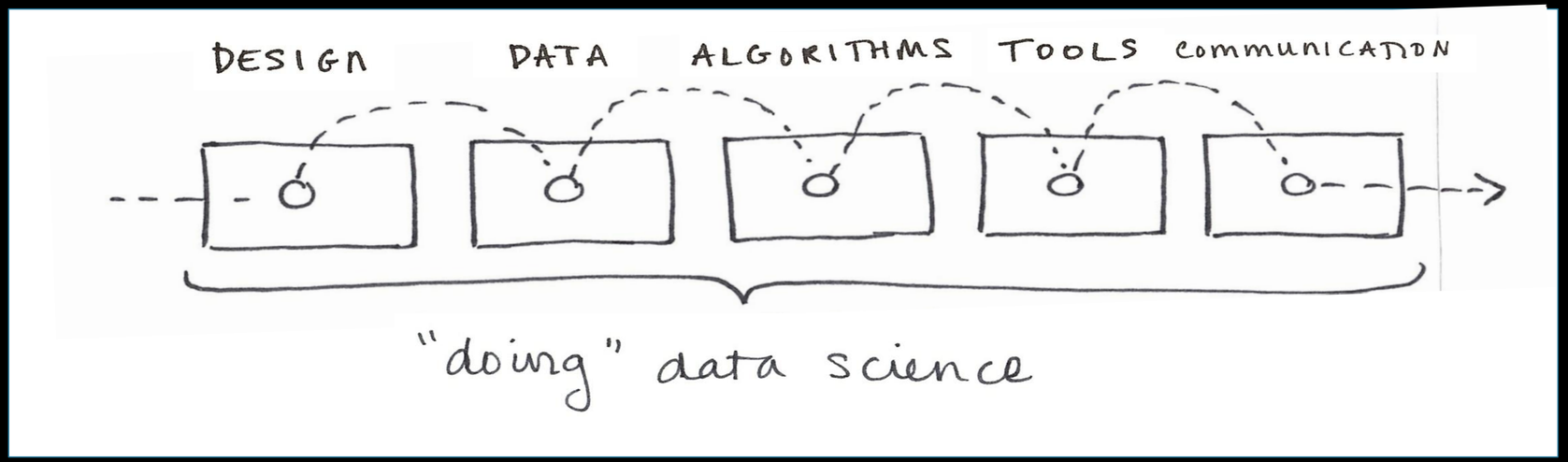
Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flat files), PostgreSQL, mongoDB
numpy, scipy, matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, JS, d3.js
Amazon Web Services, Flask
Google

Repetition

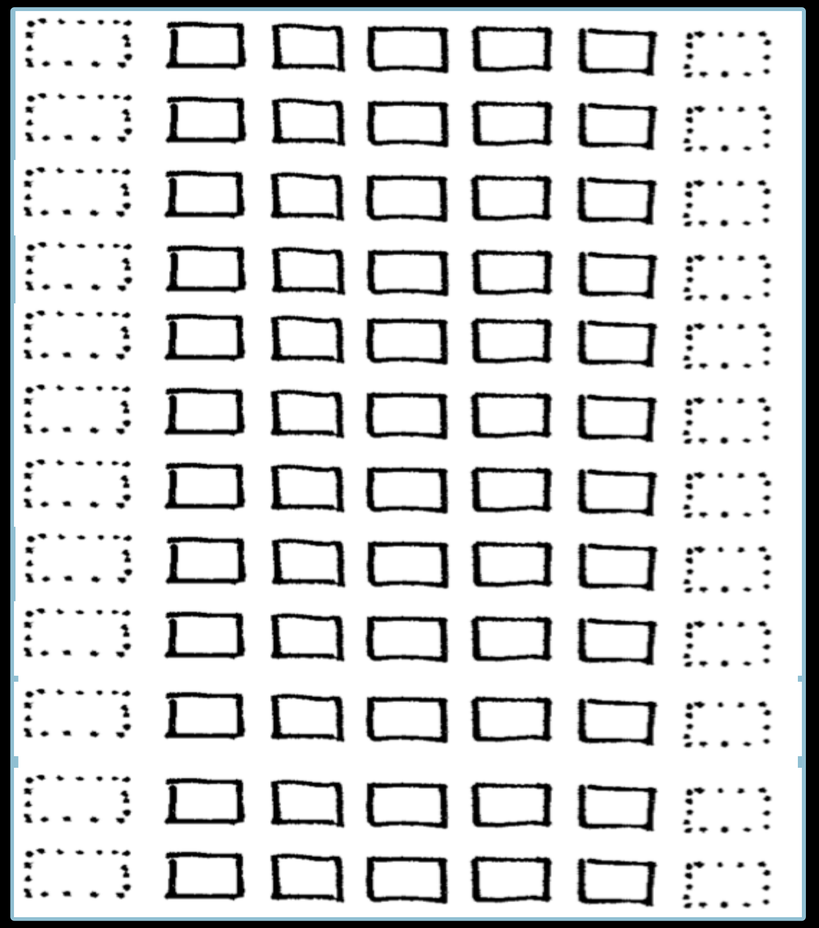
Project-based model

Unit of repetition should be *the practice of doing data science*



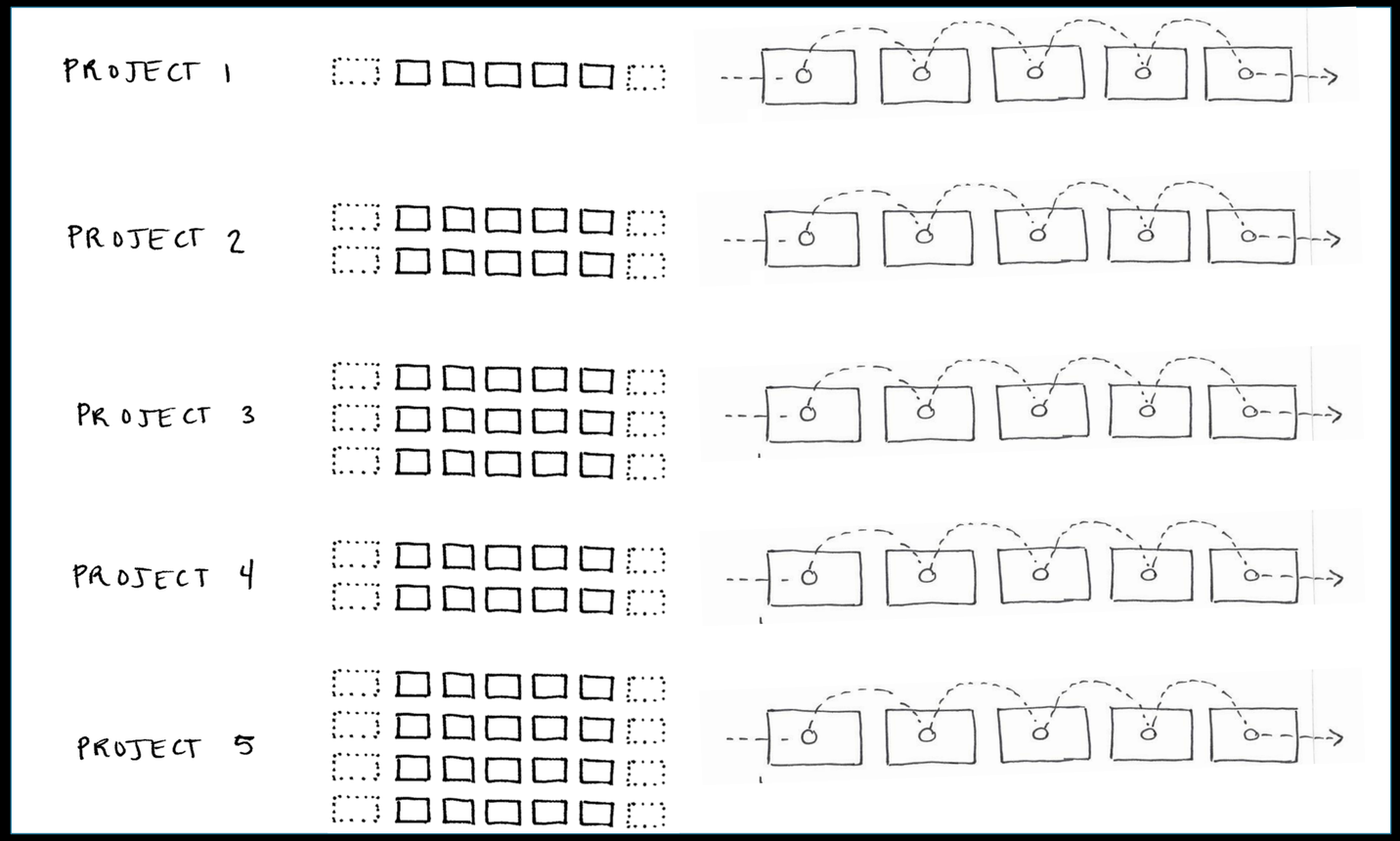
Repetition

Partitioning 12 weeks



Repetition

Partitioning 12 weeks



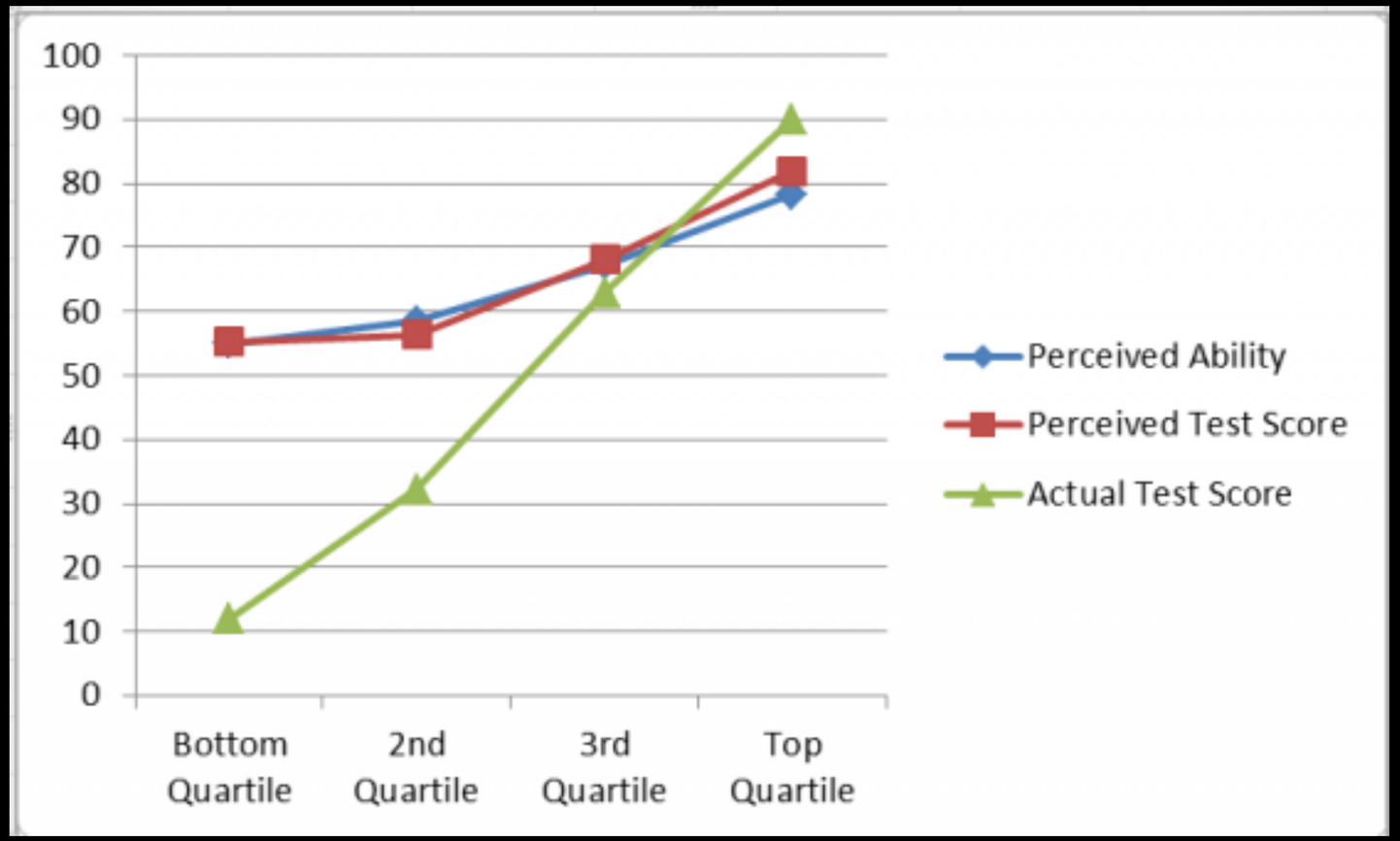
Culture & Community

Must overcome:

Impostor syndrome

Perfectionism

Fighting the Dunning-Kruger effect



Culture & Community

Must overcome:

Impostor syndrome
group work
communication

Culture & Community

Must overcome:

Perfectionism
- unfairly short deadlines
- jumping into the unfamiliar
- presenting projects

Culture & Community

Becoming data scientists

Mixture of group and individual work

Guest speakers and meetups

Hiring partners