

Data Science Intro Workshop

Charles S. Givre

thedataist.com

@cgivre

<https://www.linkedin.com/in/cgivre/>

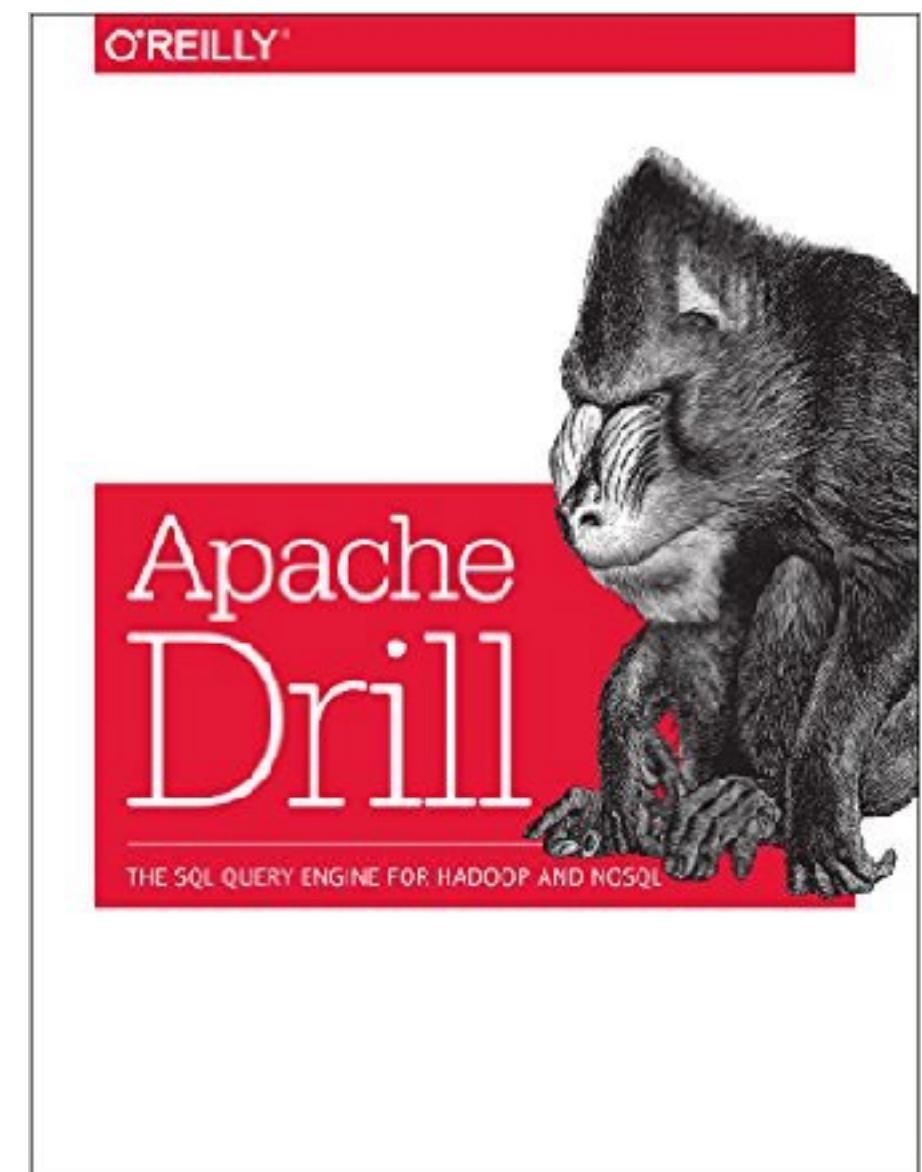
Who are you?

- Your name (quest, favorite color)
- Where you work
- Your experience with Python
- Your experience with data science

Who am I?

Charles Givre

- Senior Lead Data Scientist @ Booz Allen
- 5 Years @ CIA
- Working on Apache Drill Book
- Masters Degree in Middle Eastern Studies
- Undergraduate in Comp.Sci & Music



Booz | Allen | Hamilton
— 100 —
YEARS

@cgivre

www.thedataist.com

charles.givre@gtkcyber.com

Goals for Today

- Introduce you to Data Science
- Introduce you to the Python Data Science ecosystem
- Set up your computer
- Set up Jupyter Notebook



Expectations

- Please participate and **ask questions.**
- Please follow along and **TRY OUT** the examples yourself **during the class**
- All the answers are in the slide decks or GitHub repository, but please try to complete any exercises **without looking at the answers!**
- Have fun!

What is data science?

Data Science is the art of **turning data into actions**. This is accomplished through the creation of data products, which provide actionable information without exposing decision makers to the underlying data or analytics

Booz Allen Hamilton, Field Guide to Data Science, Pg. 17

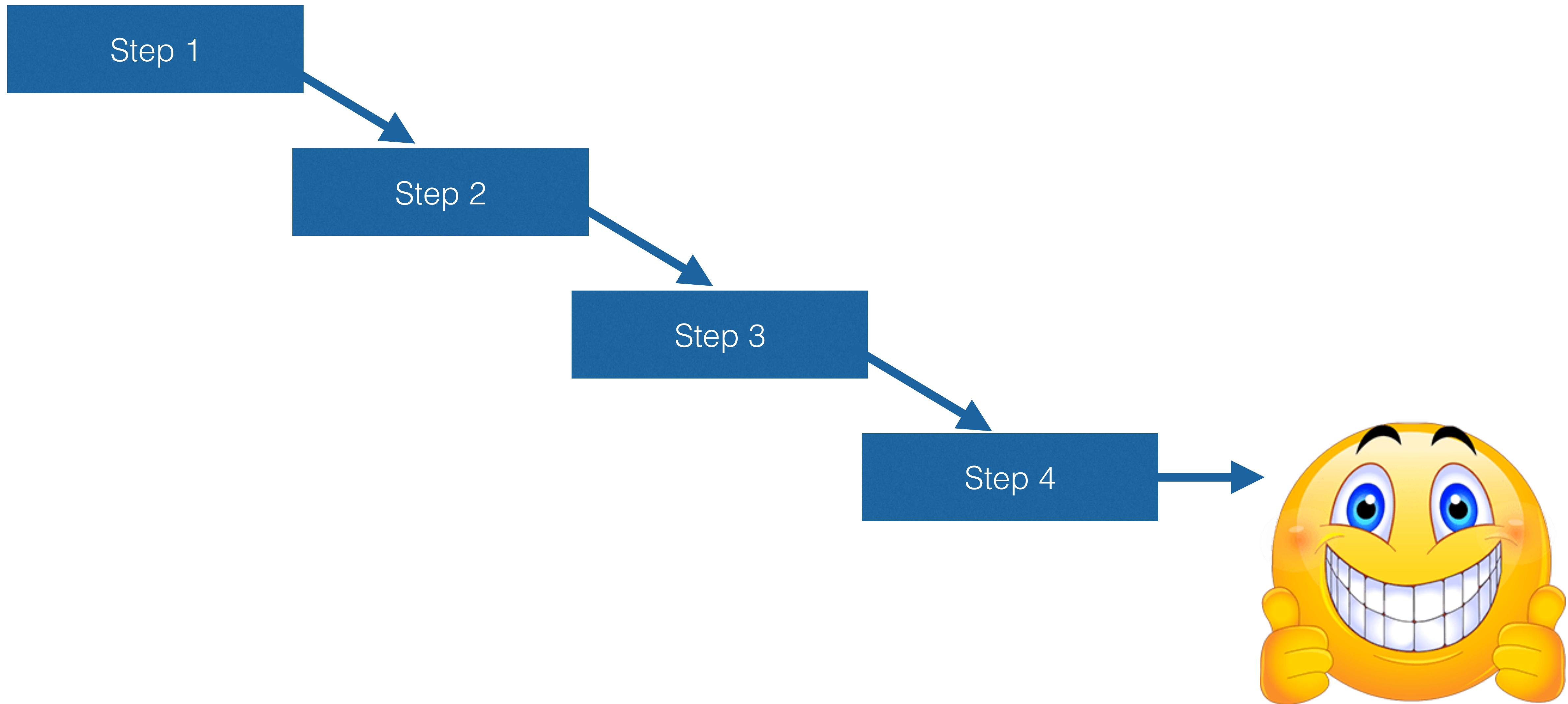
Analyst



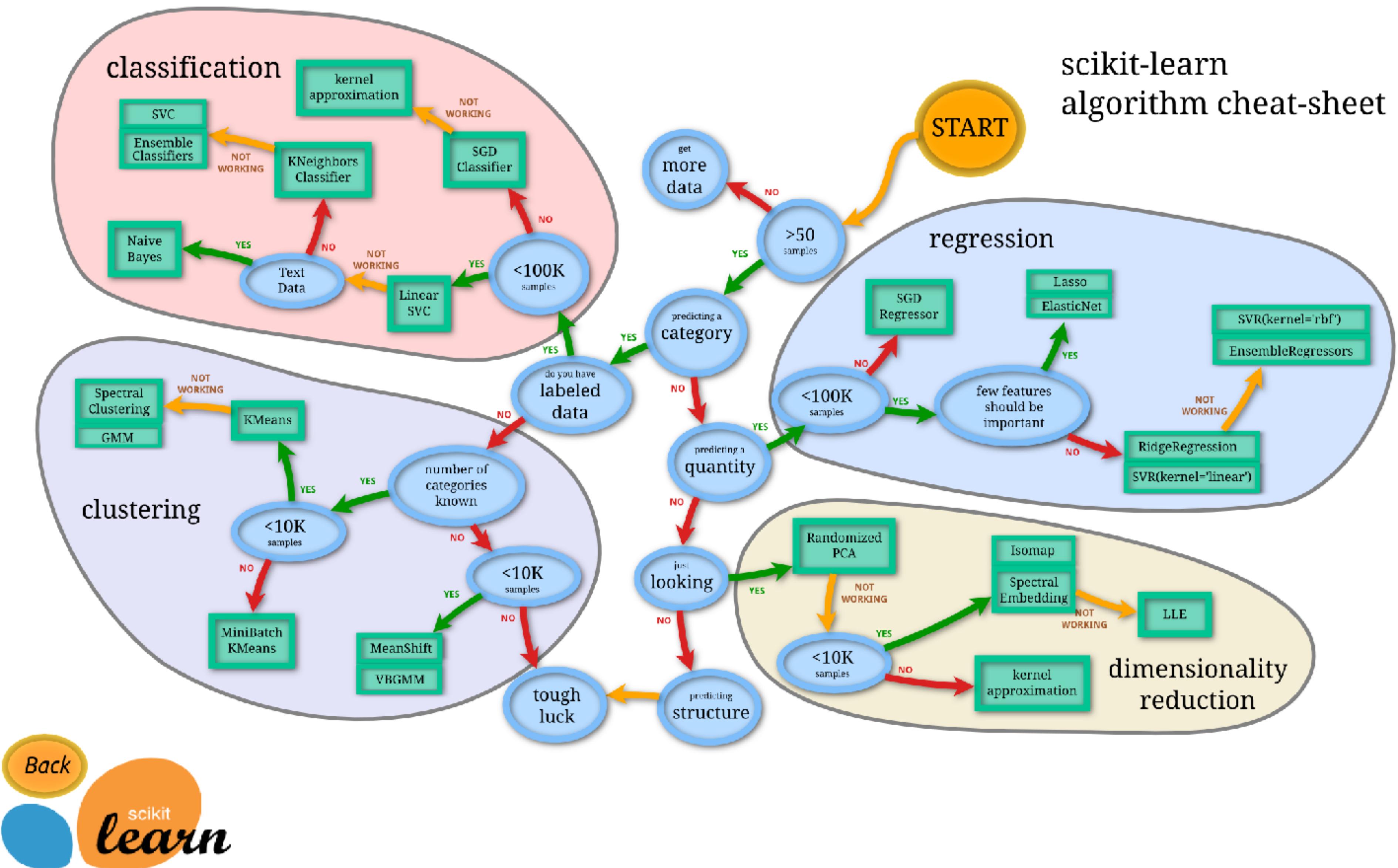
Developer

Analyst + Developer

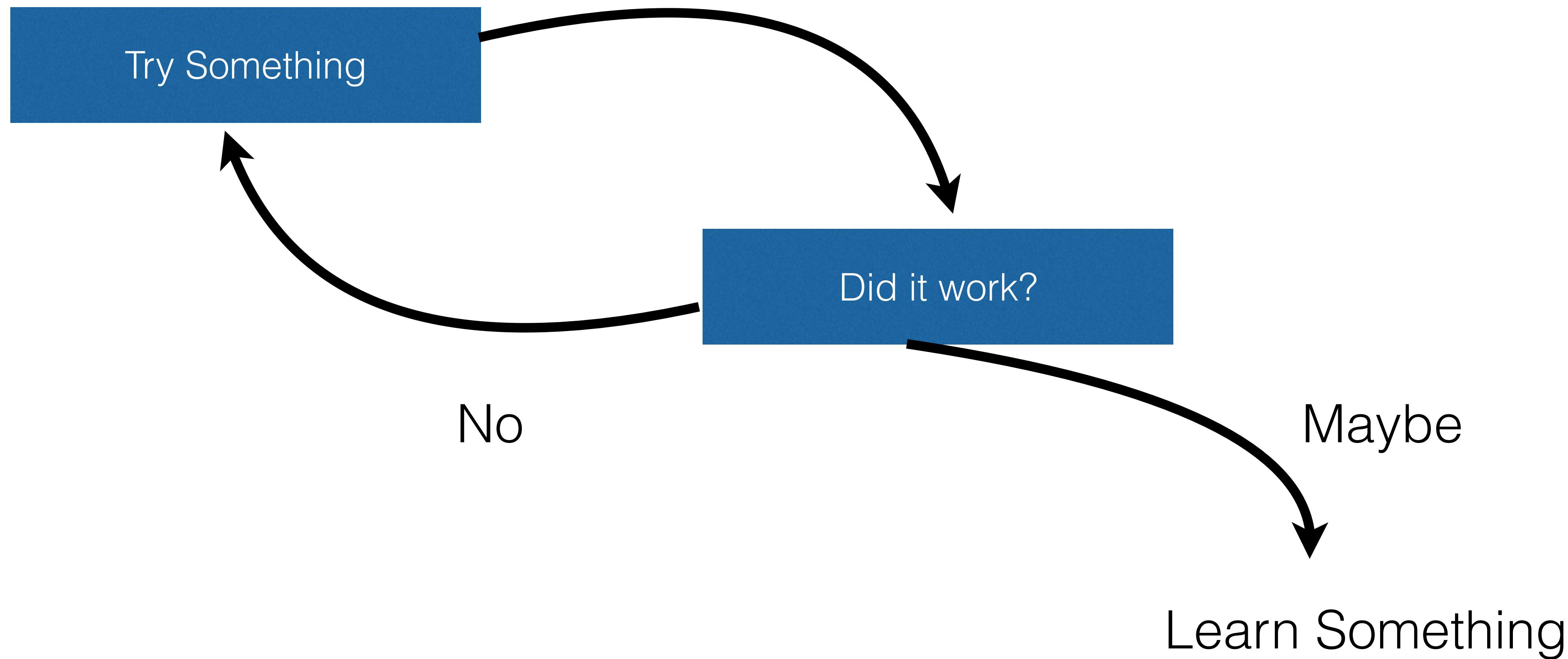
What Data Science is Not



What Data Science is Not

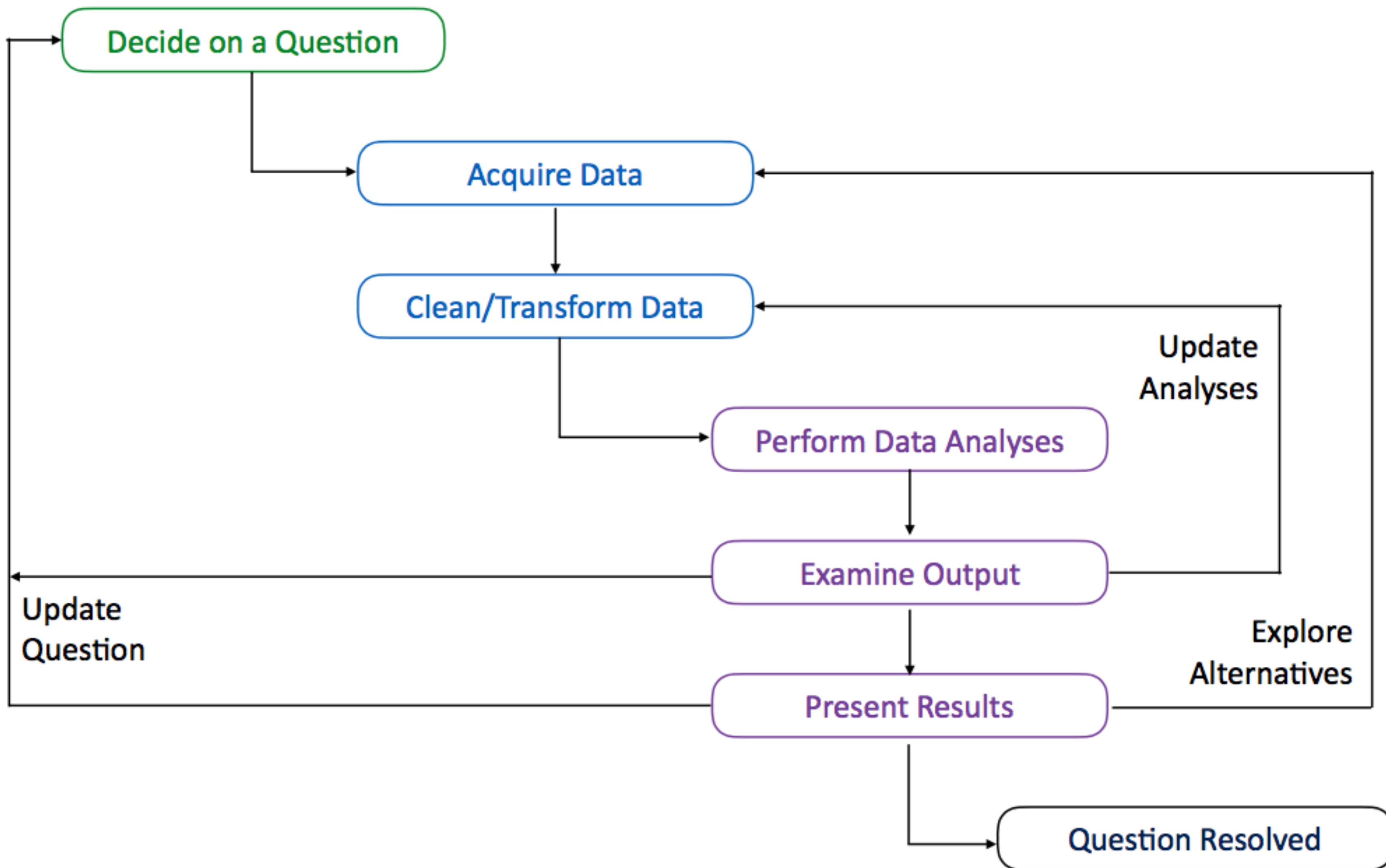


What Data Science is





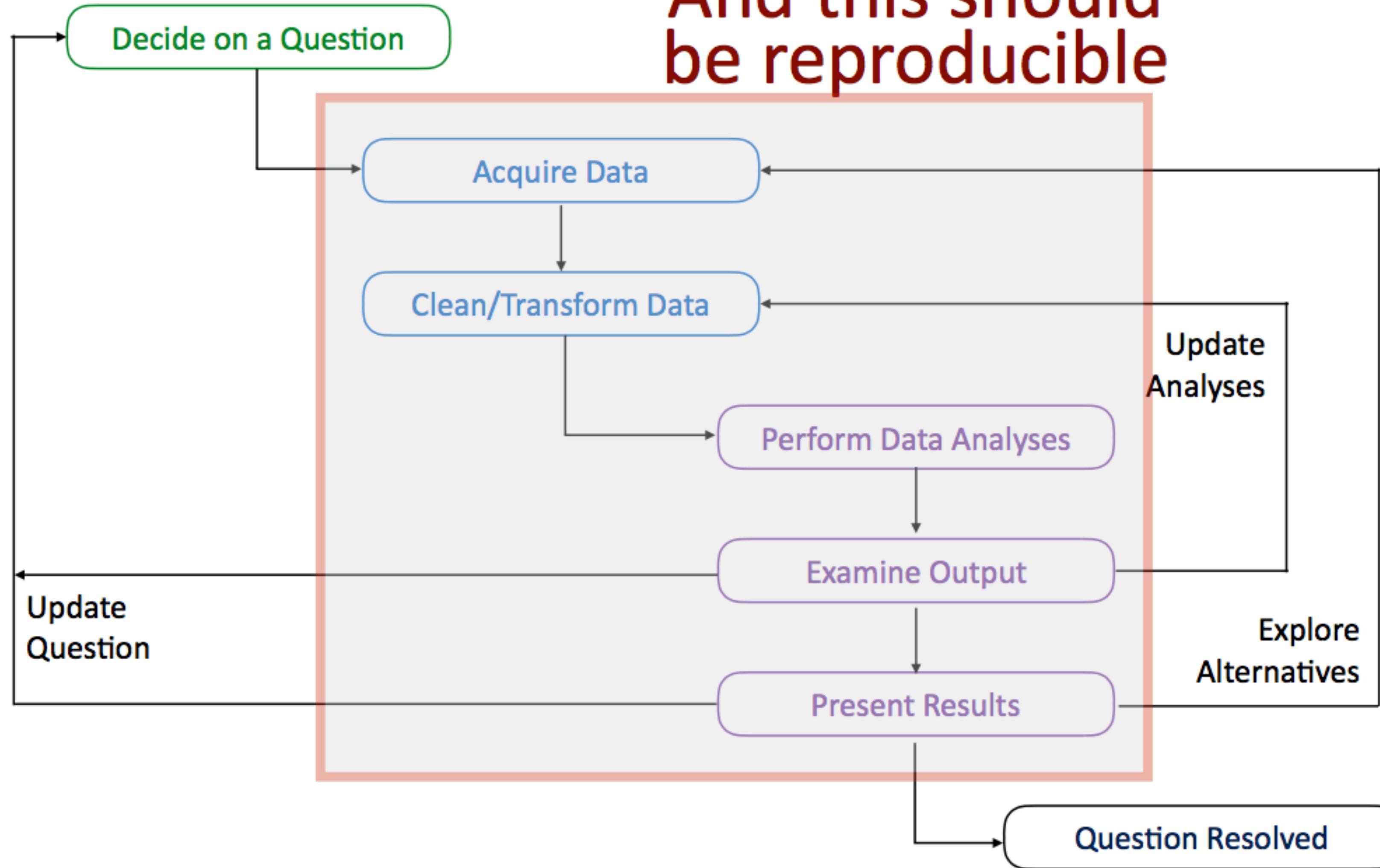
Research Process





Research Process

And this should
be reproducible





“The term "data scientist" will subside and may well sound dated five years from now. **The skills will become more commonplace and commoditized. When that happens, the real boom will begin**, because the technology will become widely adopted and thus more useful. ... **Instead, we need self-service tools that empower smart and tenacious business people to perform Big Data analysis themselves.**

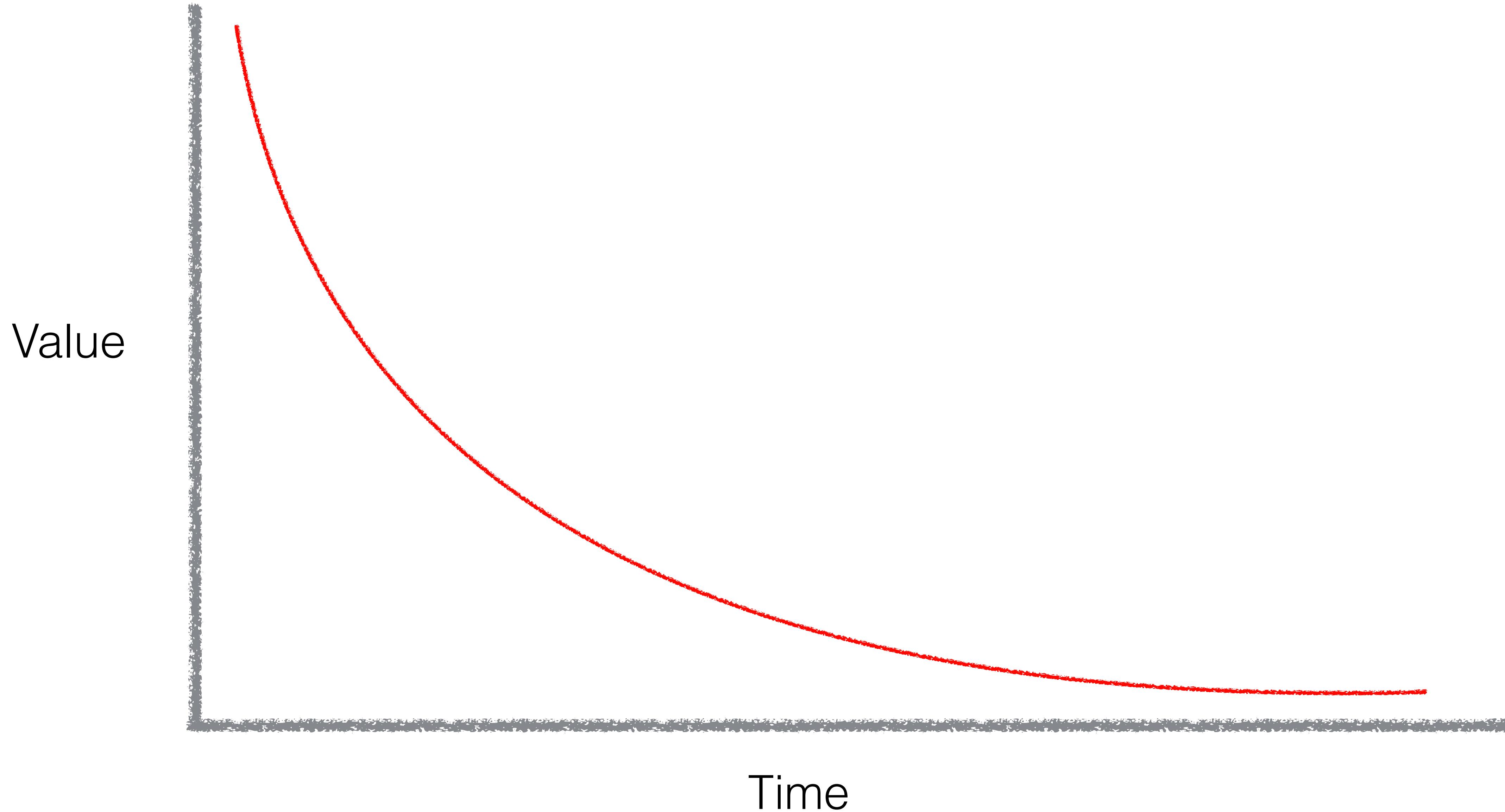
–Andrew Brust, “Data scientists don't scale”, <http://www.zdnet.com/article/data-scientists-dont-scale/>

Time to Insight

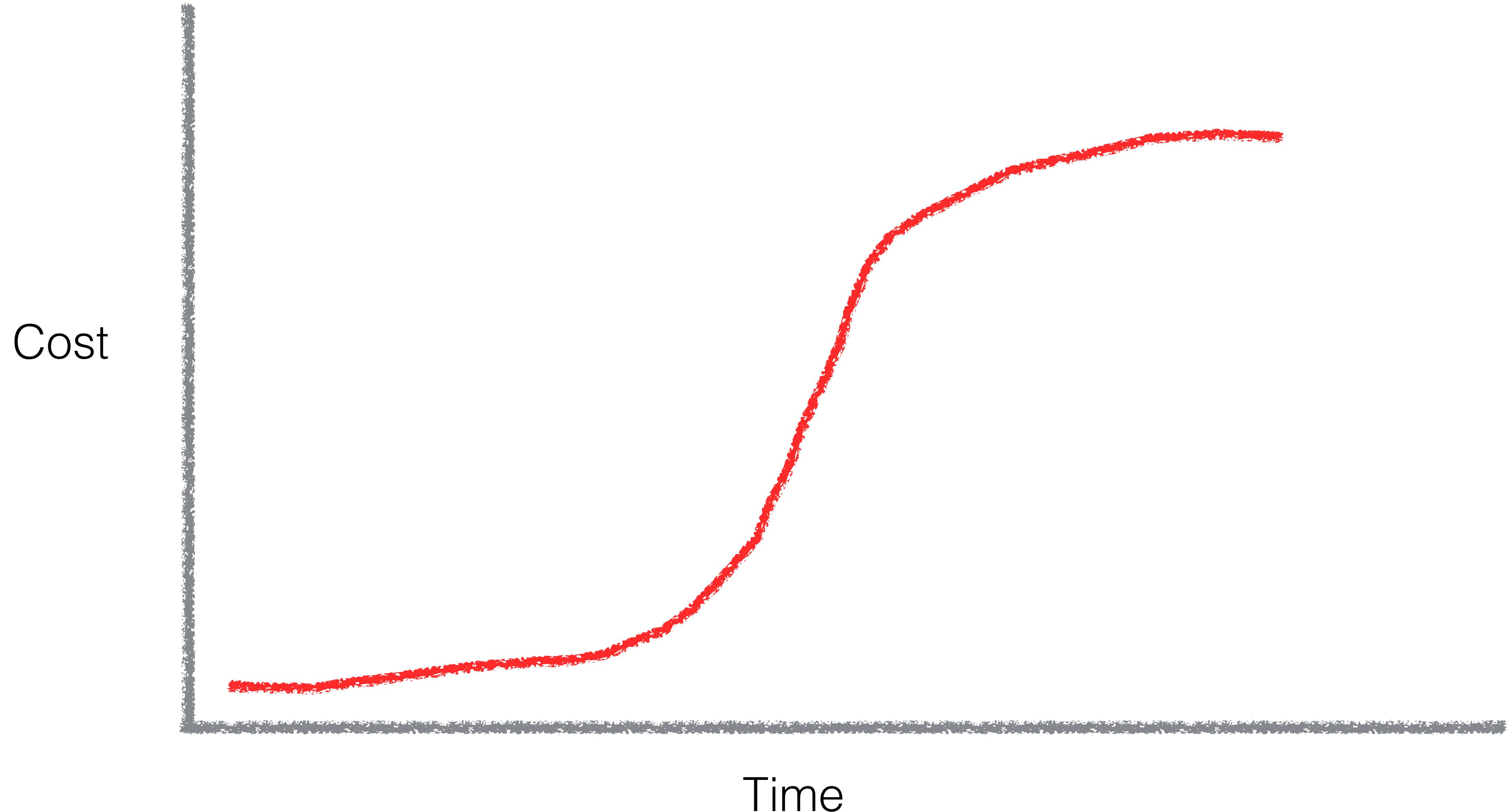
Time to Insight

Time = \$\$

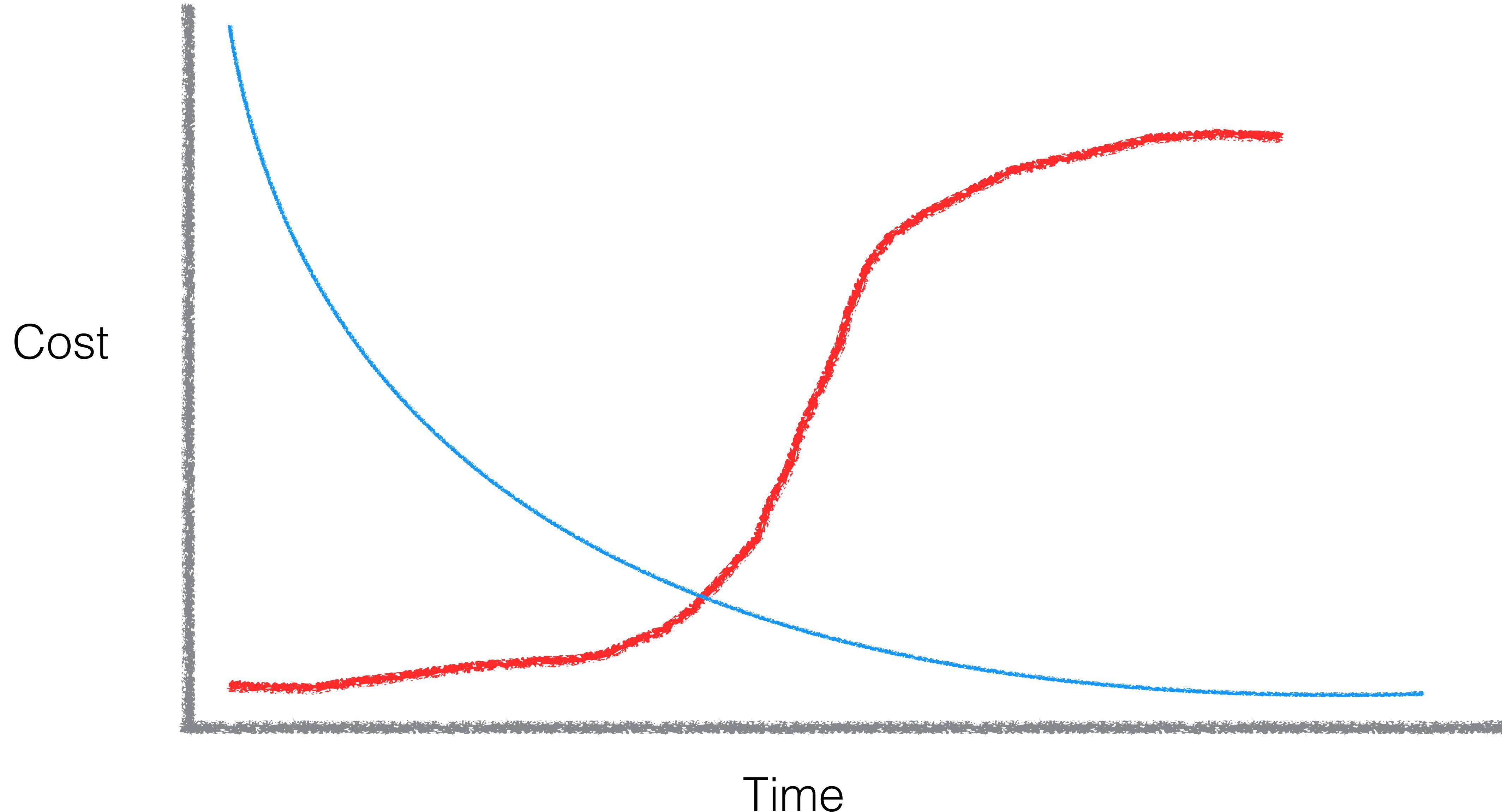
Value of Insights over Time



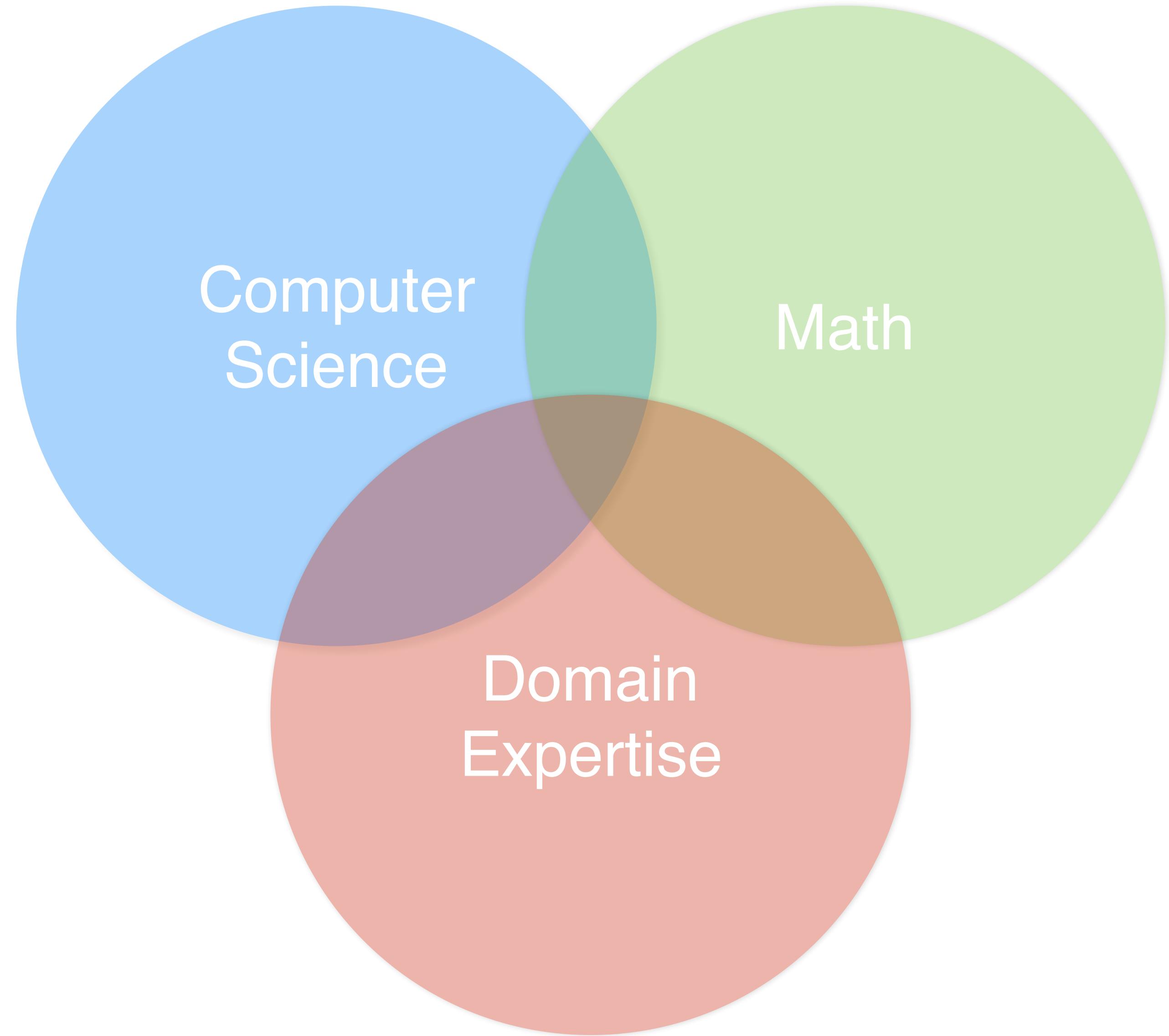
Cost of Insights over Time

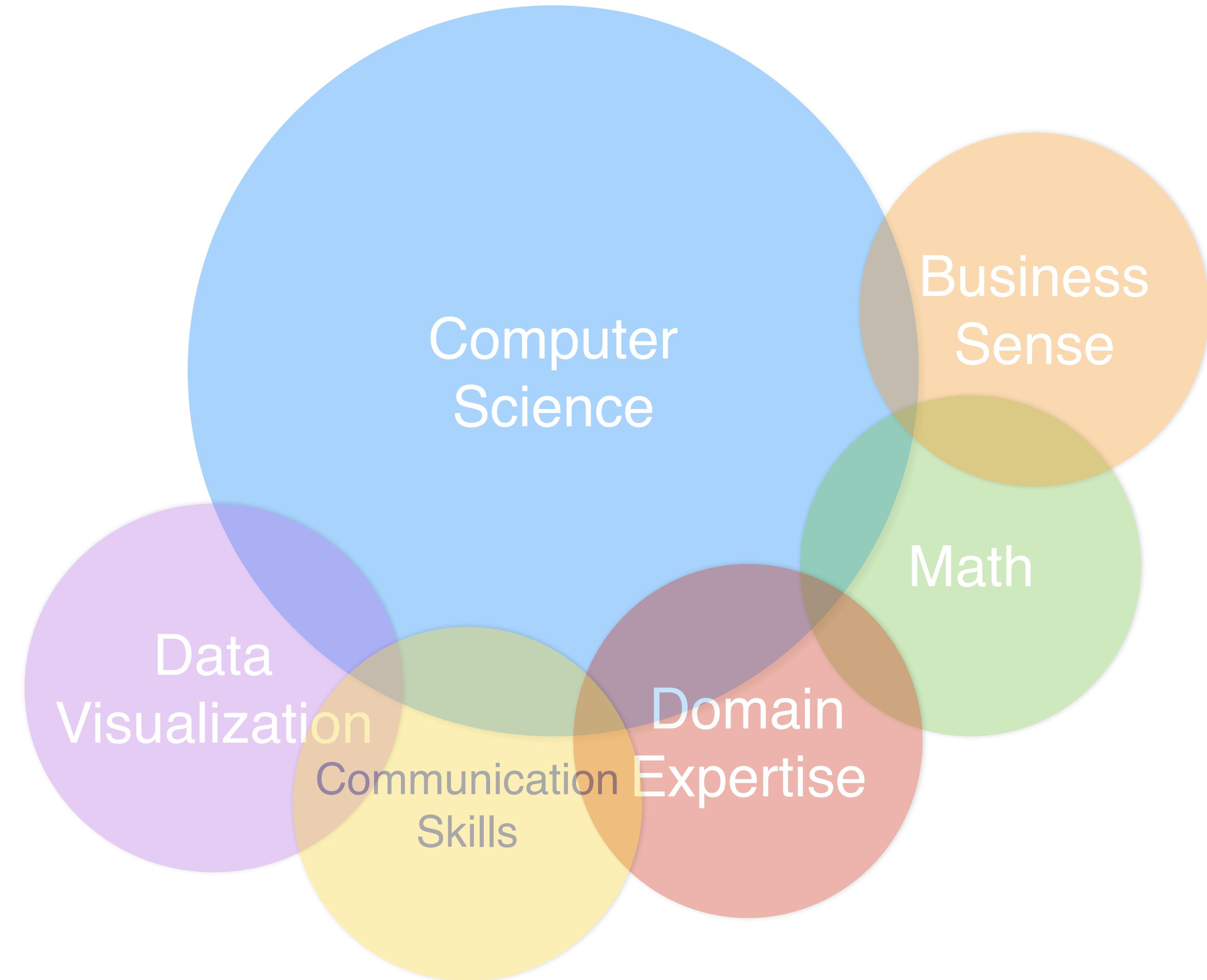


Cost of Insights over Time

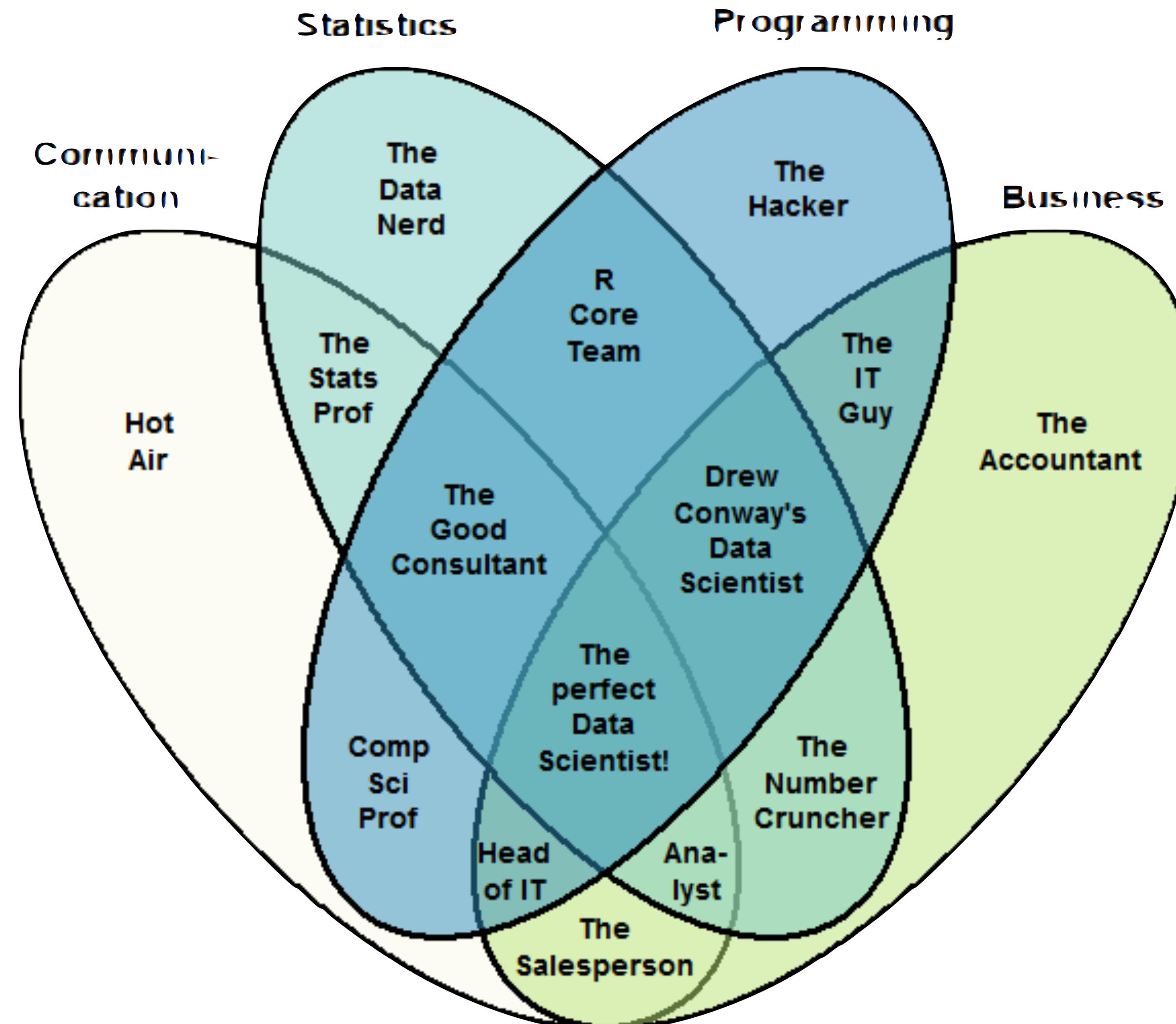


What skills does a data scientist
need?





The Data Scientist Venn Diagram



Data Scientists spend
50-90% of their time being...

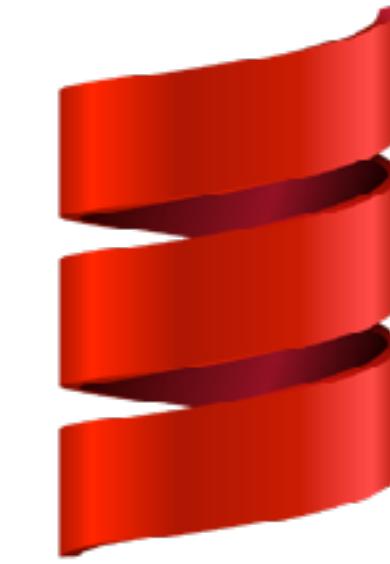
Data Janitors!



Questions?

Data Science Tools

Programming Languages





Why Python?



Why Python?

Python is easy to learn



Why Python?

Python is extensible



Why Python?

Python has many
available extensions



Python has many available extensions

- NumPy / SciPy: Statistics and Linear Algebra
- Pandas: Vectorized Computing
- Scikit-Learn: Machine Learning
- Matplotlib / Seaborn / Bokeh /Yellowbrick: Visualization

Set up your computer

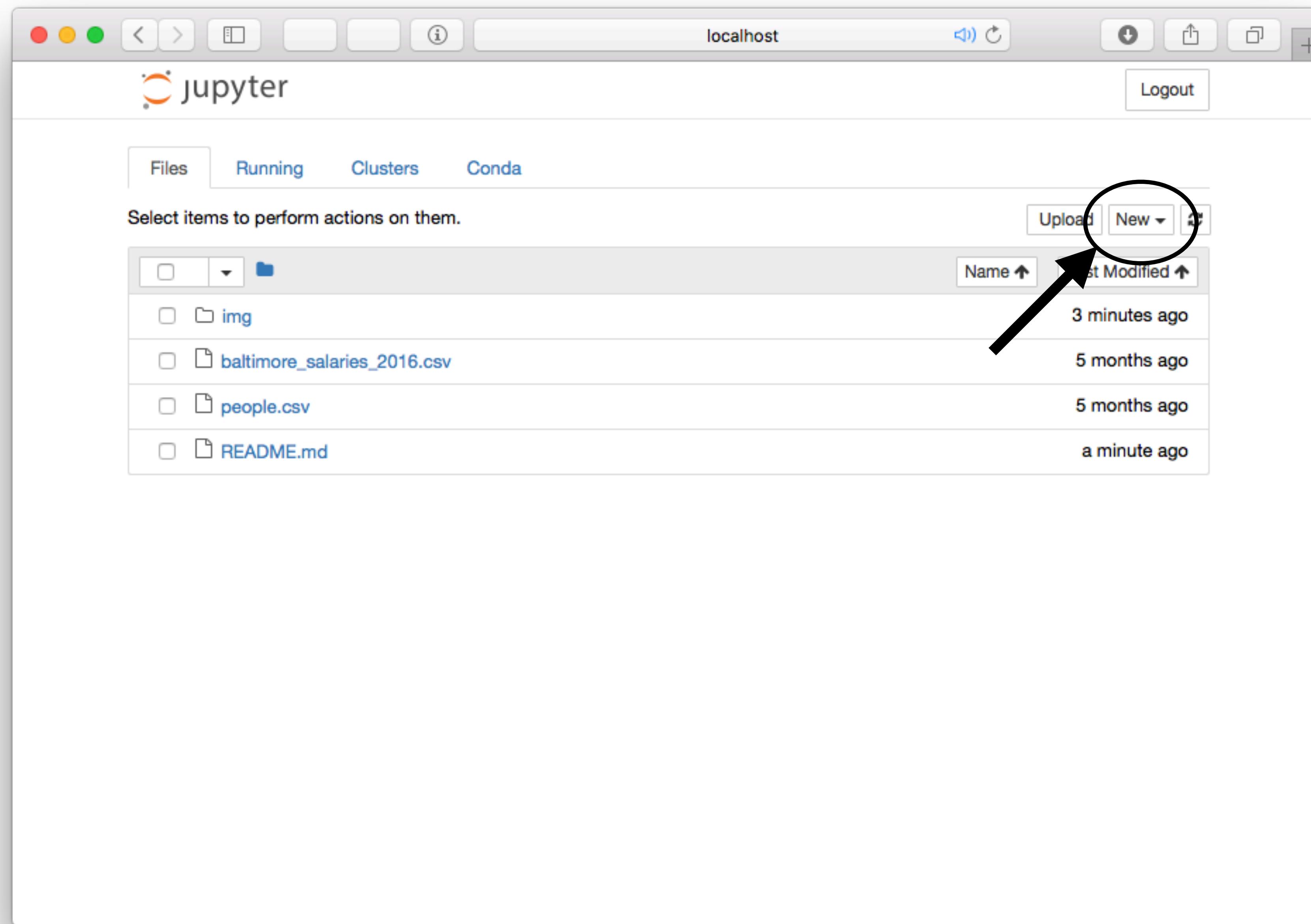
Please go to <https://github.com/cgivre/metis-data-science-intro/> and follow the instructions for setting up your computer



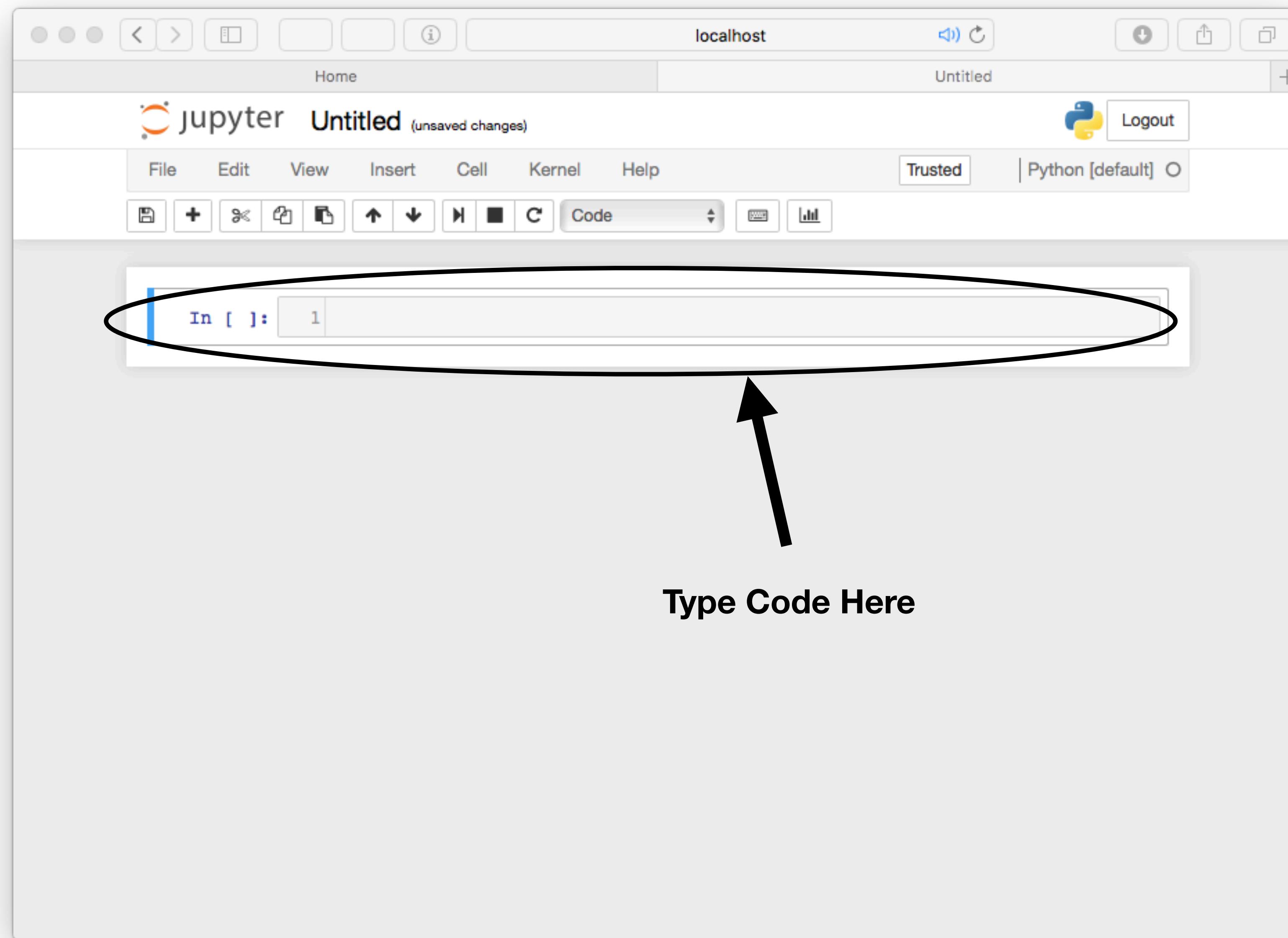
metis-data-science-intro — -bash — 80x24

Charless-MBP-2:metis-data-science-intro cgivre\$ jupyter notebook

http://localhost:8888



http://localhost:8888



Pandas and Vectorized Computing

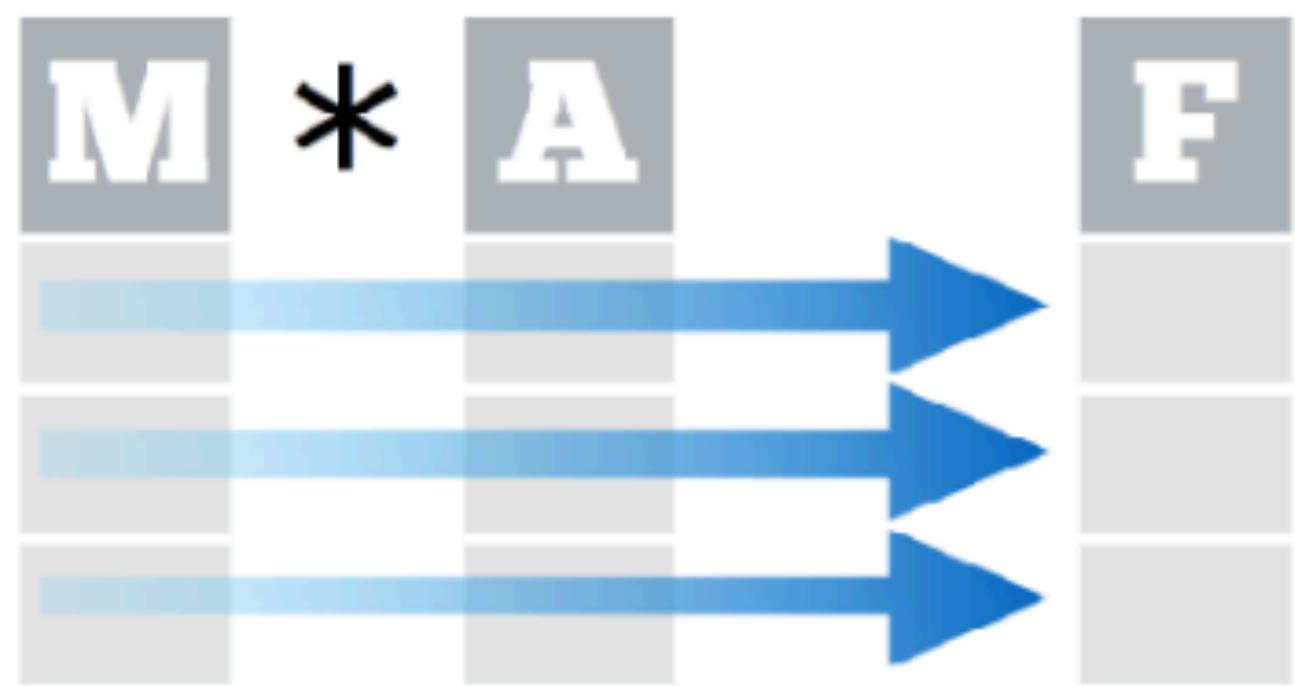
Vectorized Data Structures let you
perform operations on your data
all at once

```
for x in range(0, len( data ) ):  
    data[ x ] = data[ x ] + 1
```

```
for x in range(0, len( data )):  
    data[ x ] = data[ x ] + 1
```

odds = evens + 1

$$F = M * A$$



Each **variable** is saved
in its own **column**



&



Each **observation** is
saved in its own **row**

No loops





Dimensions	Name	Description
1	Series	Indexed 1 dimensional data structure
1	Timeseries	Series using timestamps as an index
2	DataFrame	A two dimensional table
3	Panel	A three dimensional mutable data structure. (You're making things too complicated)

In Class Exercise

Please open and complete the Pandas Worksheet

The Python Ecosystem lets you
focus on **what** you what to do,
not how to do it.

Questions?

Thank you!