



UNIVERSITAT DE
BARCELONA

Introduction to Data Science

Eloi Puertas Prats

Data

We are rendering into **data** many aspects of the world that have never been quantified before.

Digital Transformation in many aspects in our lifes and work places enhances this process of **datification**.



Data everywhere

Where **Information** comes from?

- Corporate Data Bases (structured information).
- Unstructured information in documents, Wikipedia, textbooks, journals, blogs, tweets, etc.
- Images in the web, public cameras, phones, TV, YouTube, etc.
- Public APIs: smart cities, government, search engines, etc.
- Sensor Data: GPS, accelerometer, physicochemical sensors, sociometric sensors, supercolliders, telescopes, etc.

How to handle such amount of data?

Big Data

- **Big Data** is the **technology** to process such **massive** amounts of data.



Big Data

- Think in data like crude oil



*“**Big Data** is about extracting the ‘crude oil’, transporting it in tanks, siphoning it through pipelines, and storing it in massive silos”*

- But you need to refine the ‘crude oil’

“*by means of DATA SCIENCE*”

Data Science

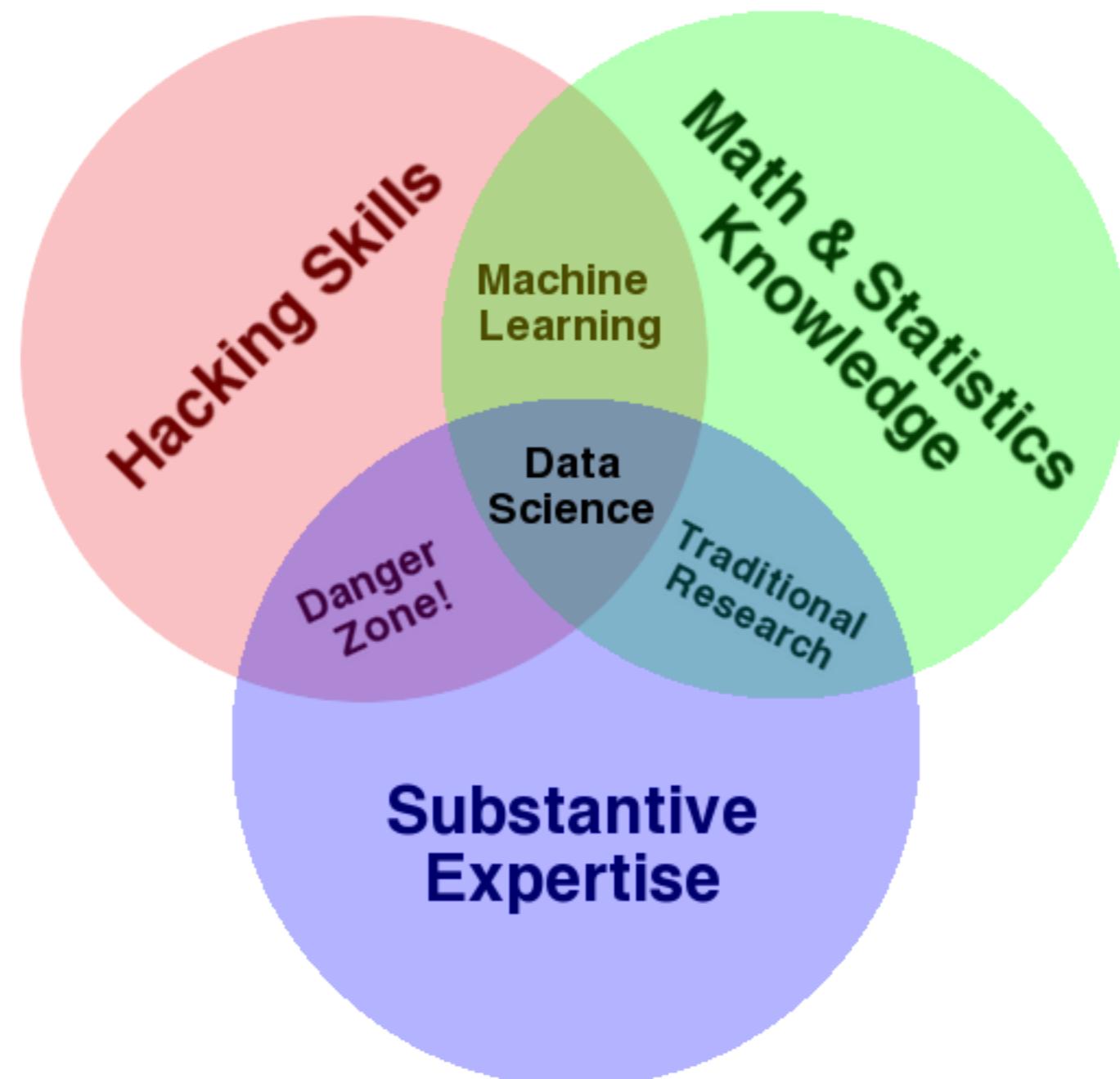
- Data Science** is not a science but a methodology based on multidisciplinar knowledge.
- Data Science** is a new data processing model focused on turning data into actions.
- Data science** is about the design and development of solutions to extract insights from data (structured and unstructured) using machine learning and predictive analytics techniques and tools..

Data Science

Data Science is a **methodology** to define:

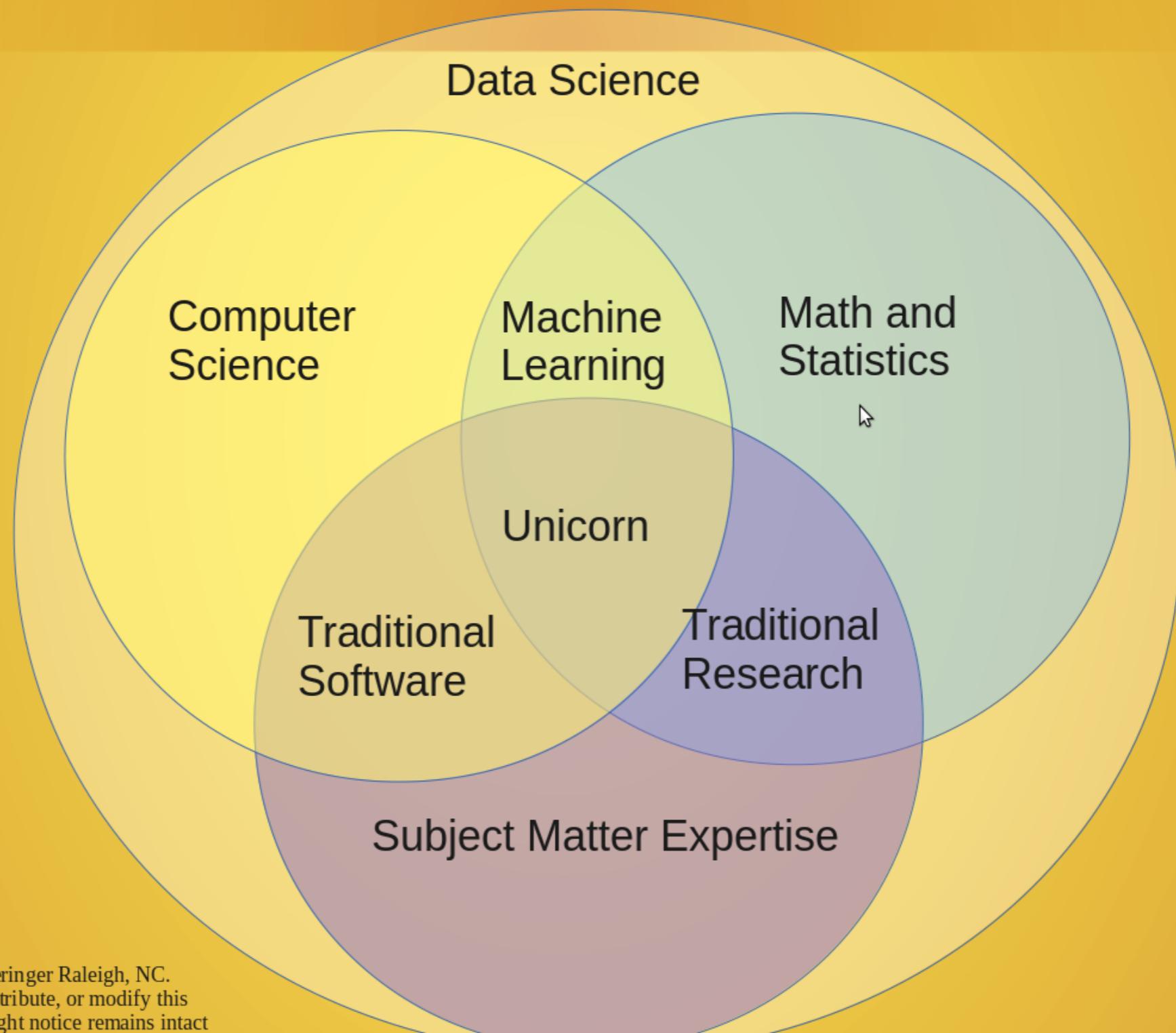
- what we want to do with data,
- how do we evaluate our actions,
- what decisions can be grounded on data,
- how do we combine evidences from several sources.

Data Science Skills V1.0



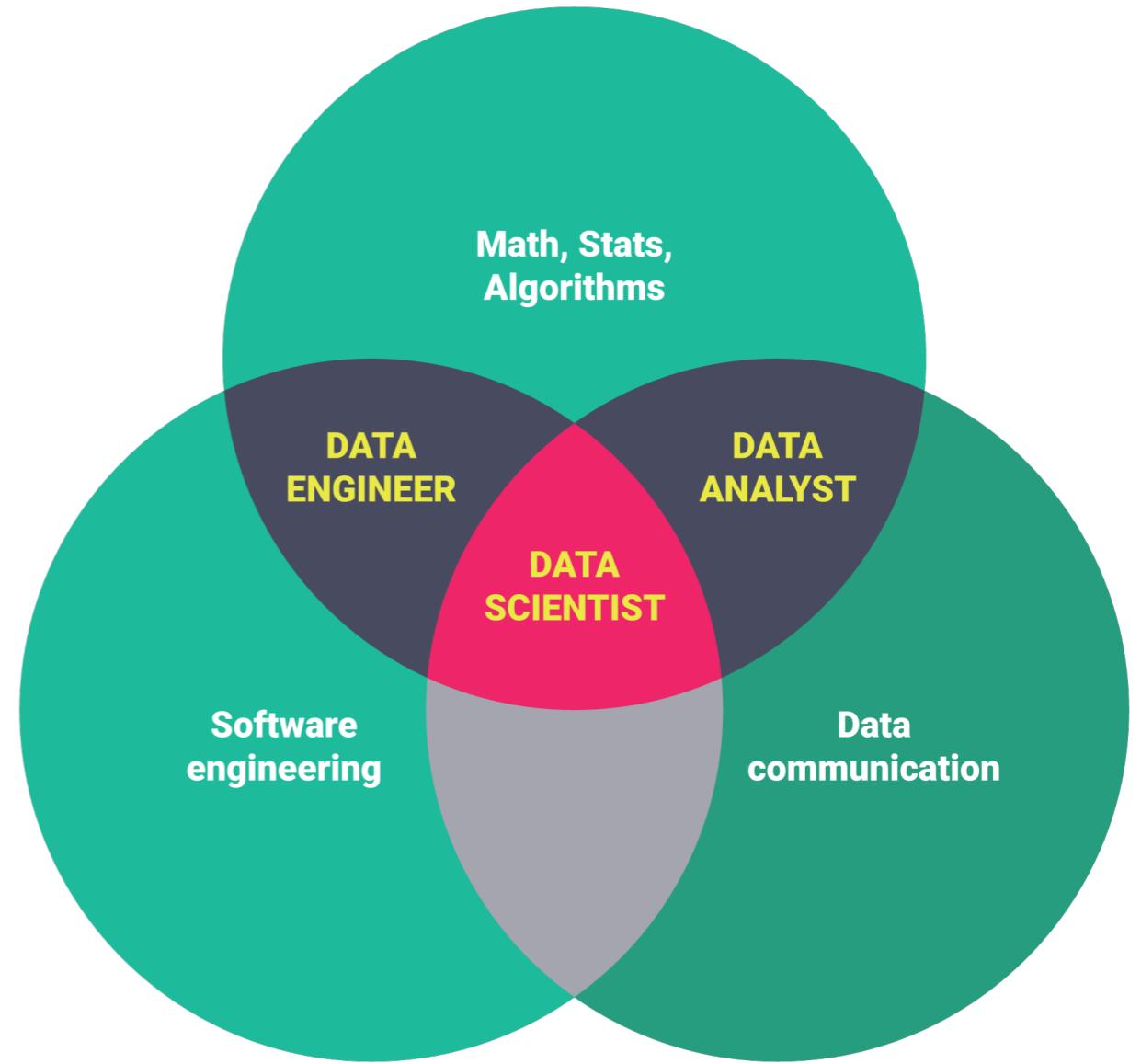
Drew Conway's Data Science Venn Diagram

Data Science V2.0



Data Science as a Team

- **Data Scientist** (A.k.a Statisticians, Data Managers)
- **Data Engineer** (A.k.a Data Managers, Database Administrators)
- **Data Analyst** (A.k.a Bussiness analysts)



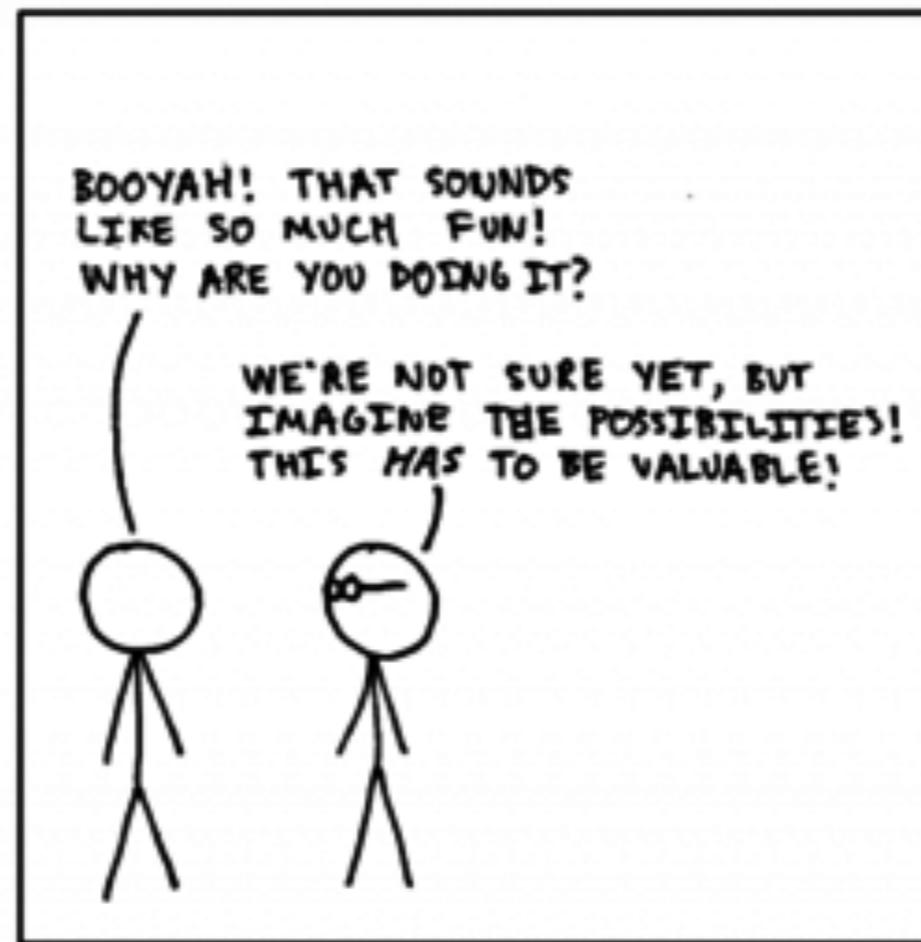
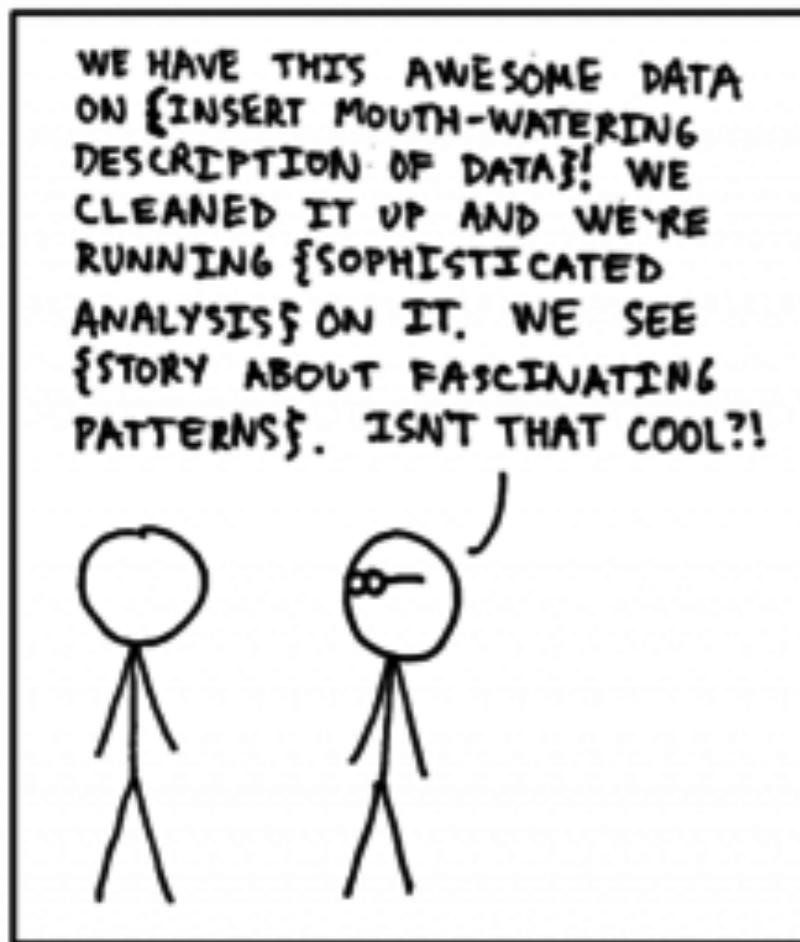
<https://www.springboard.com/blog/data-science-career-paths-different-roles-industry/>

Data Science Steps

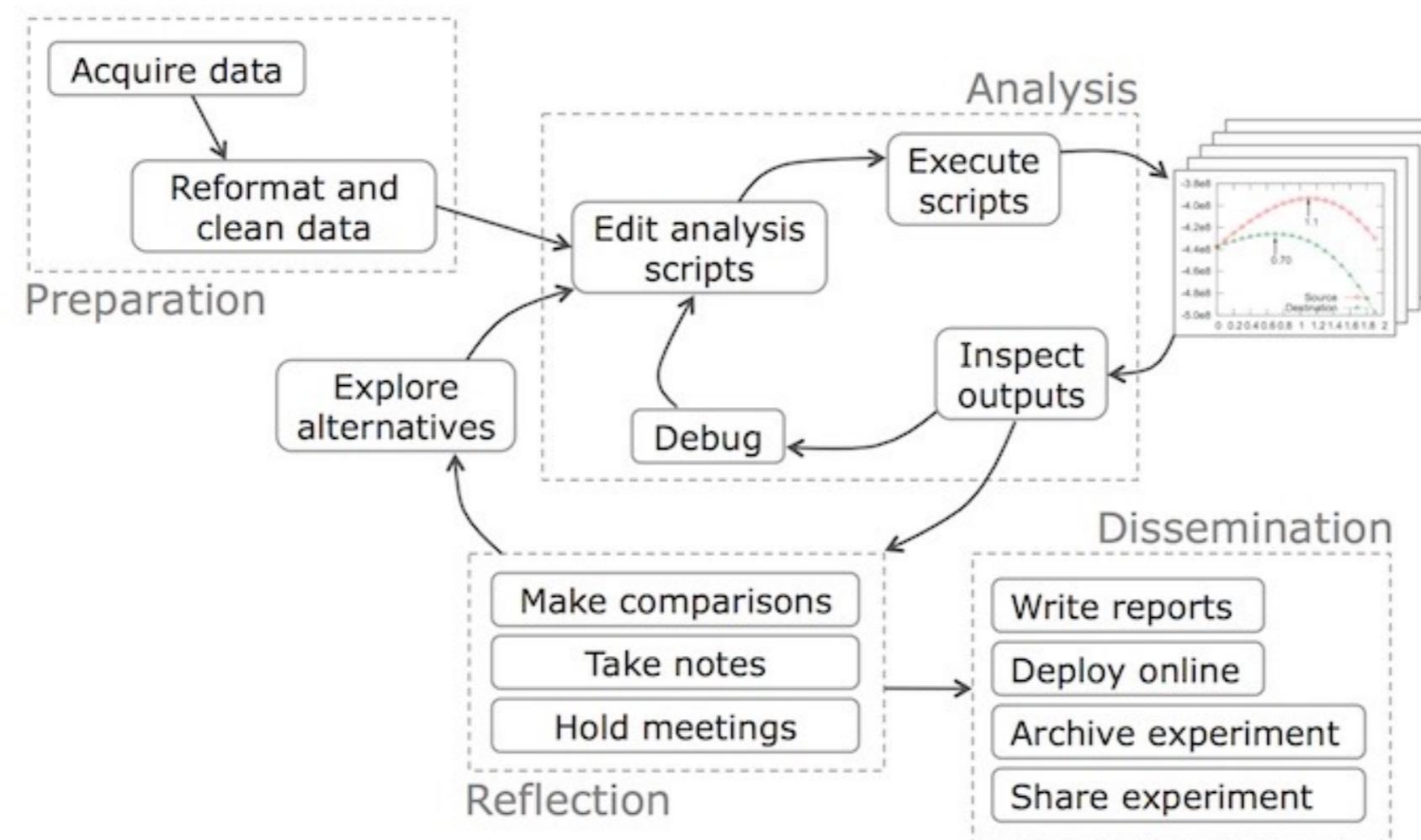
Steps:

1. Ask a question.
2. Get the data. They can be heterogeneous and non structured
3. Data Processing (cleaning, munging, ETL.).
4. Data Analysis (computer science, linguistics, economy, sociology, etc.).
5. Take a decision and act.

Ask a question



Data Science workflow



Homework

- Think about a question to be asked with data
- Think about how to get such data
- Find a team! (Deadline 14/10)
- Define your project proposal! (Deadline 28/10)