Introduction to Legal Data Science

Seán Fobbe

Ludwig-Maximilians-Universität München

29 August 2024

Student Chapter Roundtable

Liquid Legal Institute

Structure of the Presentation

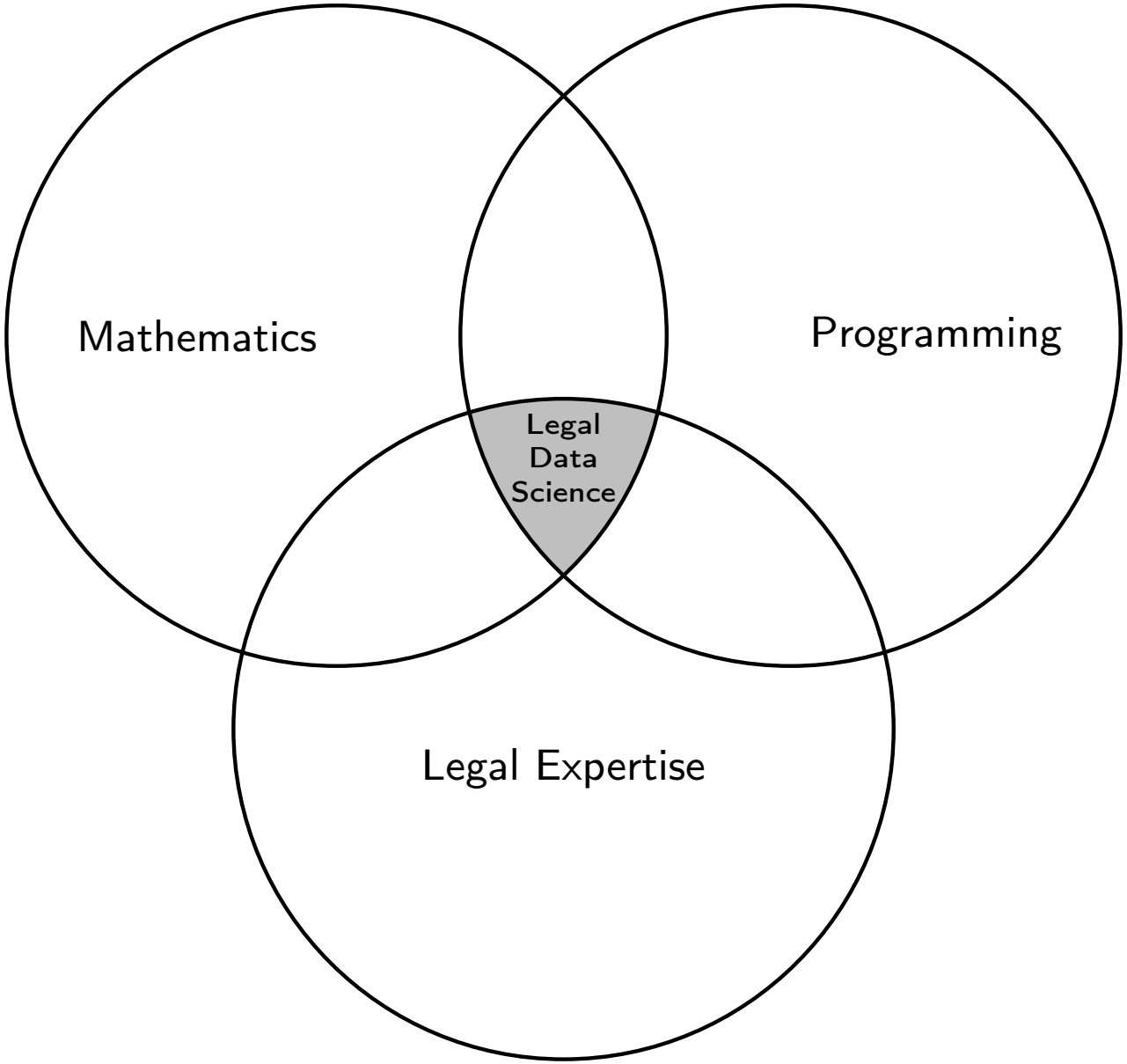
1. What is Legal Data Science?
2. Why Legal Data Science?
3. How to Get Started with Legal Data Science?

Part I

What is Legal Data Science?

Three Skillsets of a Legal Data Scientist

Adapted from Drew Conway's 'Data Science Venn Diagram'



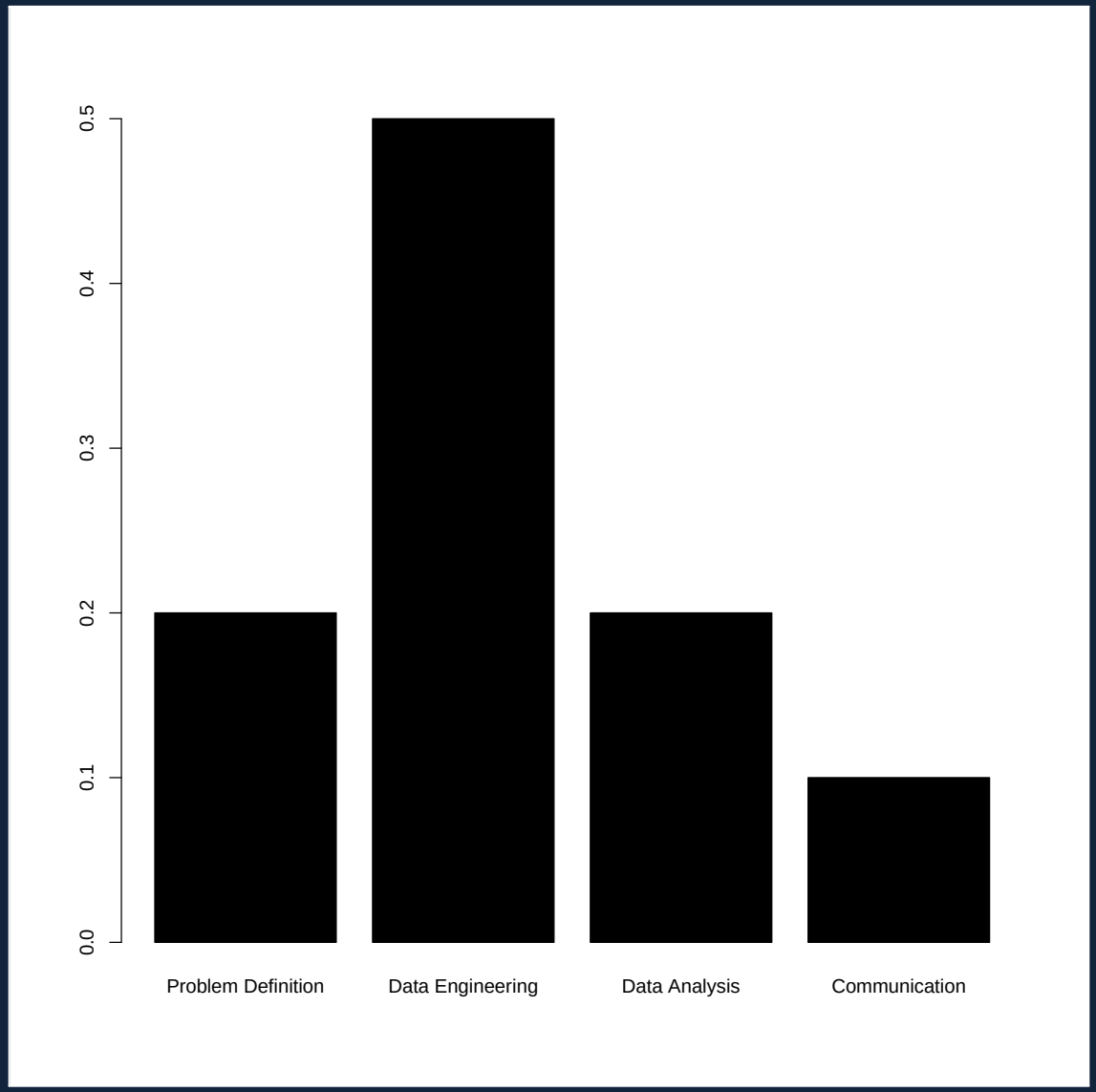
Four Components of a Data Science Project

Data science projects can be broken down into four components:

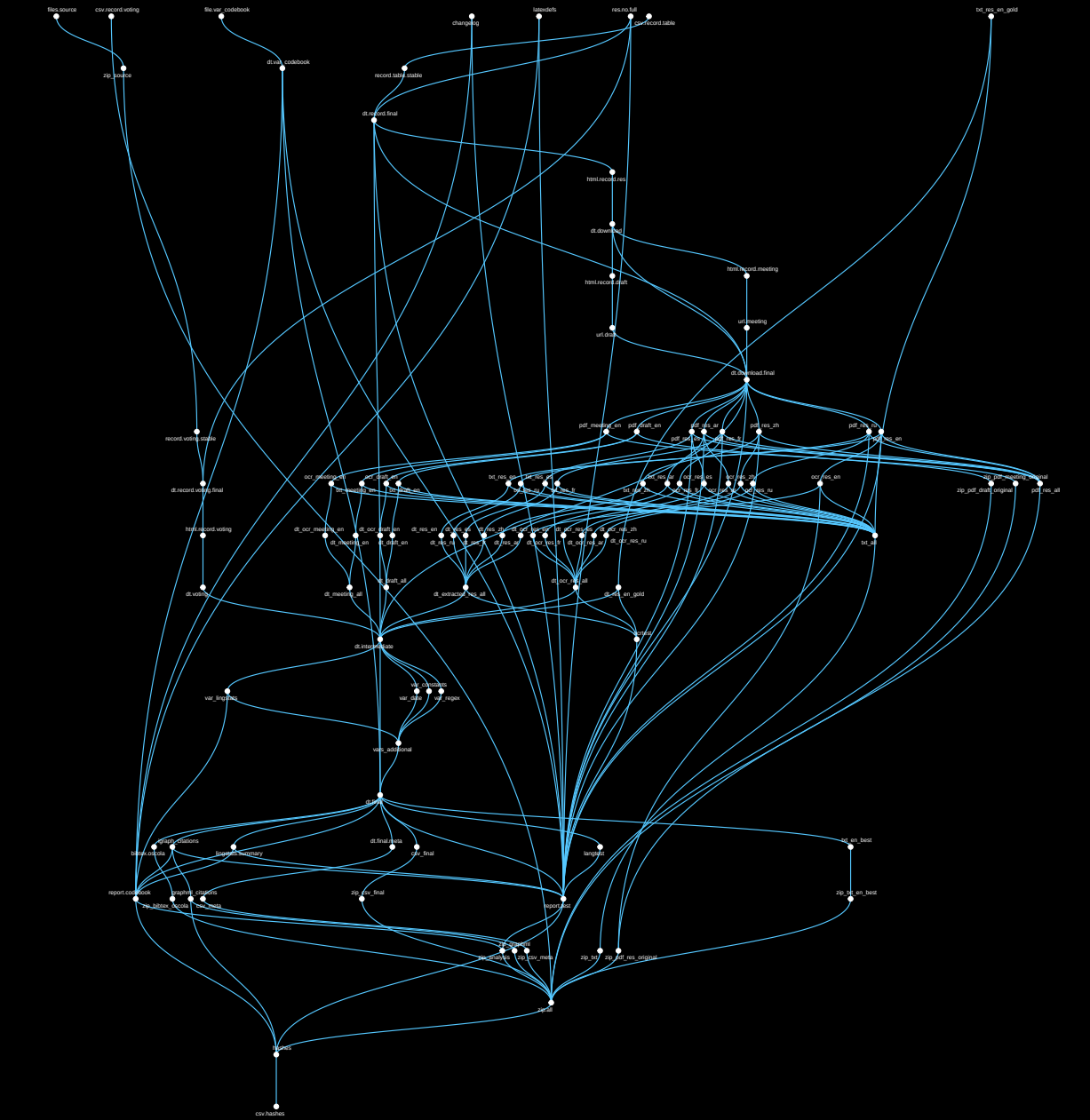
1. Figure out what you want to do and why (Problem Definition)
2. Acquire, clean and transform data (Data Engineering)
3. Explore, analyze, model and summarize data (Data Analysis)
4. Visualize and communicate results (Communication)

Time Spent in Good Data Science Projects

An Educated Guess

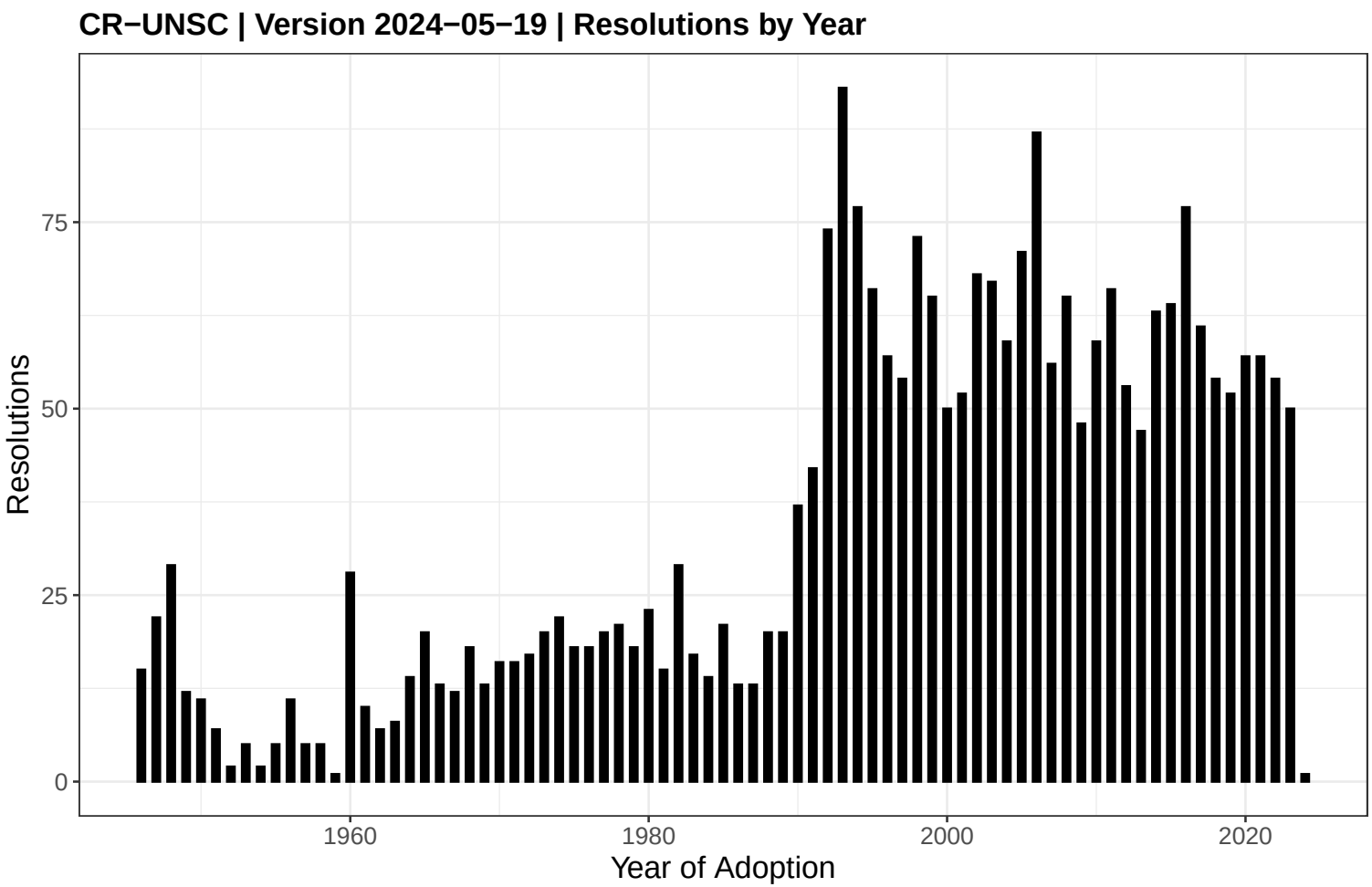


CR−UNSC | Version 2024−05−19 | Data set compilation process



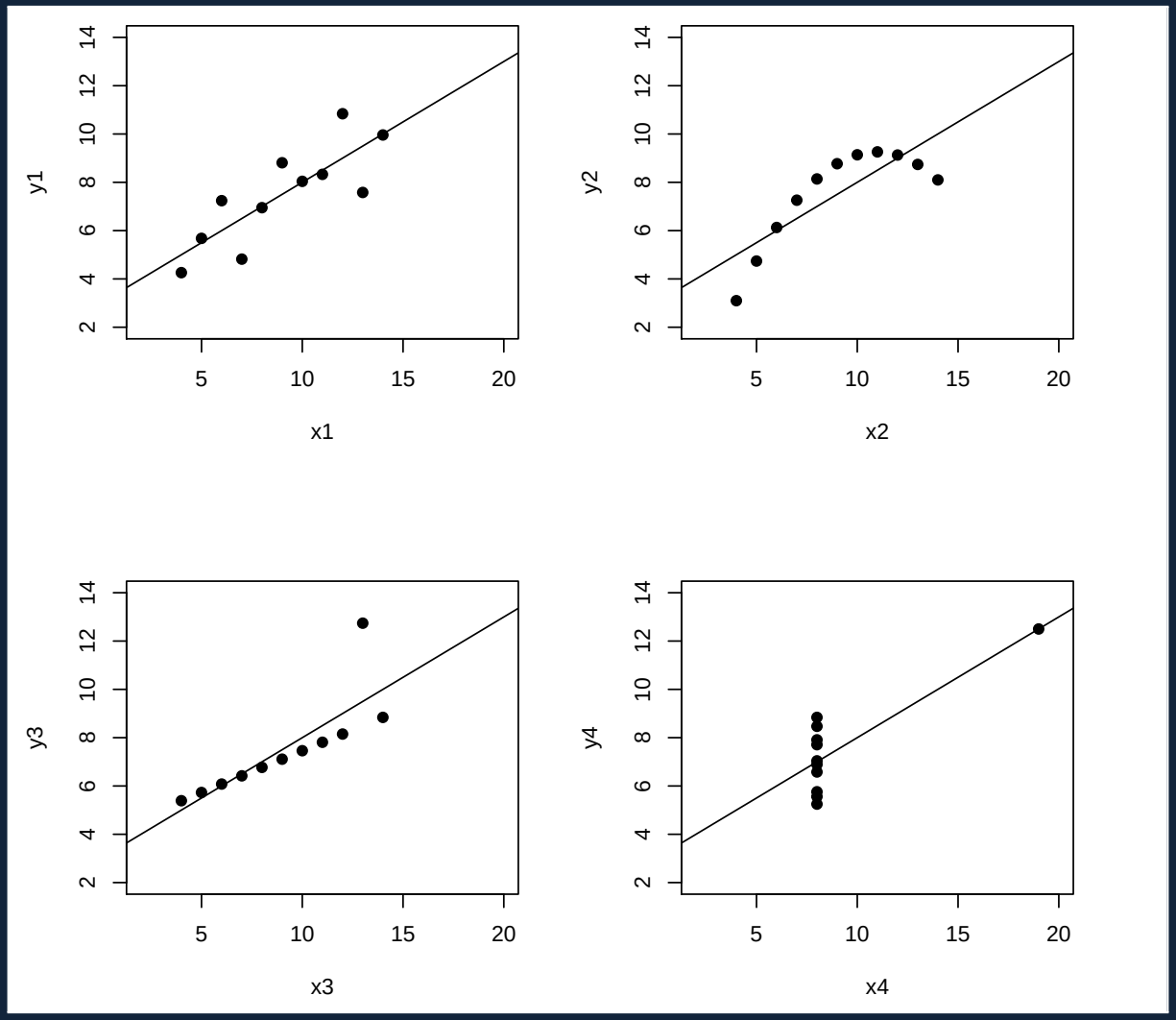
Fobbe/Gasbarri/Ridi | DOI: 10.5281/zenodo.11212056

Example: Simple Frequency Table

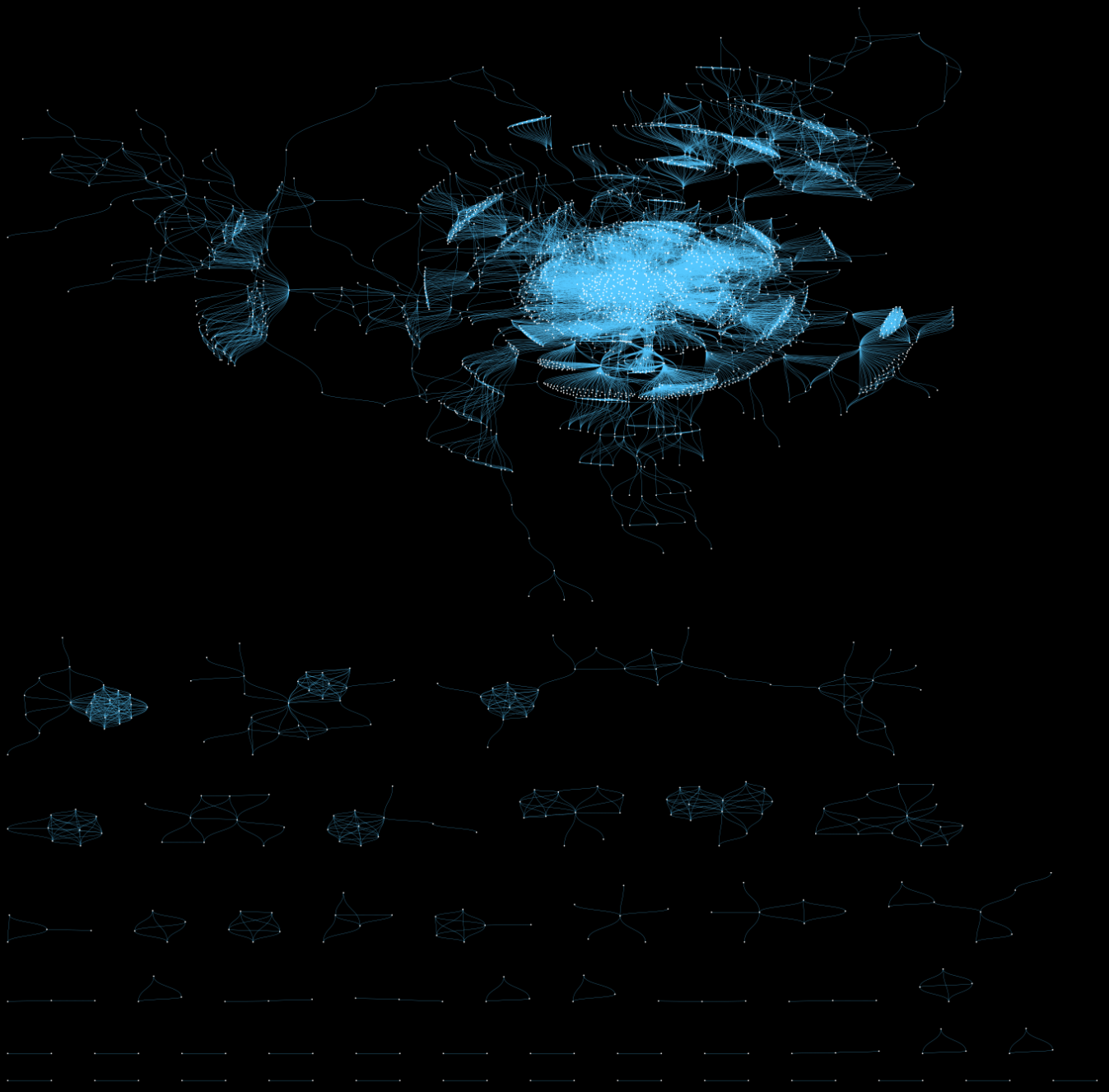


Example: Predictive Modeling

Anscombe's Quartet



CR-UNSC | Version 2024-05-19 | The complete UNSC citation network



Fobbe/Gasbarri/Ridi | DOI: 10.5281/zenodo.11212056

Part II

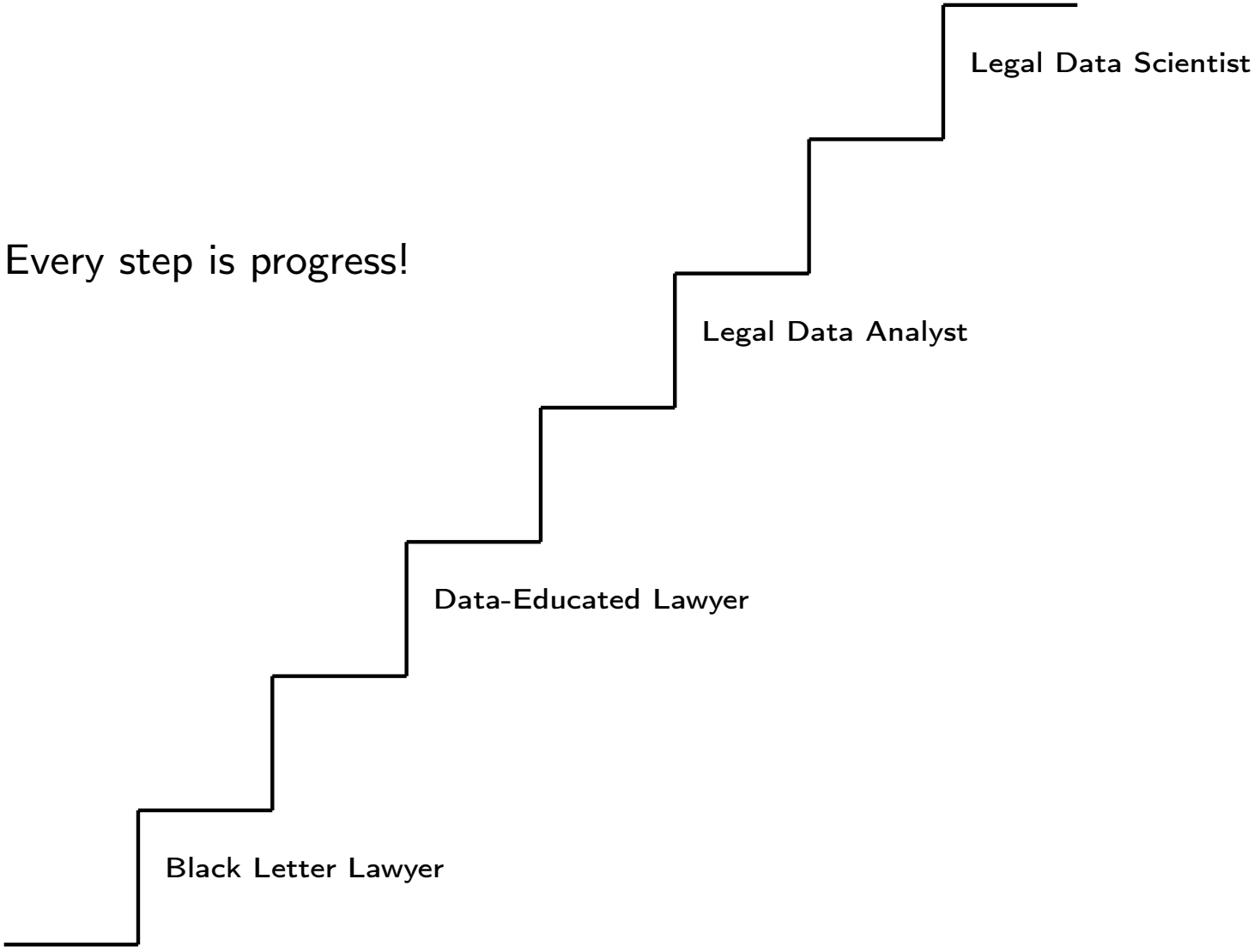
Why Legal Data Science?

Some Reasons for Legal Data Science

1. Walk the walk, don't just talk the talk
2. A little knowledge is dangerous, zero knowledge even more so
3. Maintain your autonomy in a digital world
4. Job opportunities
5. You want to found a startup (hint: you will probably not find any programmers)

The Legal Data Science Staircase

Every step is progress!



Some Motivation for the Average Lawyer

Why bother with Legal Data Science if you don't want to become a Legal Data Scientist?

- You don't have to become a full-fledged attorney to benefit from legal knowledge, nor do you have to become a full-fledged data scientist to benefit from data science skills
- Basic knowledge of tax law and employment law are useful for everyone
- Basic data literacy, statistics and dataviz are useful in many jobs (see Excel!)
- Working with data scientists and machine learning engineers requires common ground, i.e. some knowledge of their fields

Part III

How to Get Started with Legal Data Science?

How to Get Started

1. Read general data science literature
   - Many data science books are Open Access
   - Will teach you statistics and programming at the same time
2. Start your own computational projects as soon as possible
   - You cannot learn programming without programming things
   - You have to write a lot of bad code to write good code
   - Motivation is strongly associated with perceived usefulness
3. https://seanfobbe.com/how-to-get-started-with-legal-data-science/

Where to Get Started

1. Start with Natural Language Processing (NLP) if you are a lawyer — you already have lots of experience working with texts
2. Pick either Python or R as a programming language and focus on learning one language well
3. Descriptive statistics (e.g. mean, median, distributions) are valuable knowledge in any job
4. Data visualization is incredibly important, spend a lot of time learning to do it well
5. Predictive modeling and causal inference are hard, you need to put in a lot of work to do them right — not for beginners

Example NLP Tutorial:
https://seanfobbe.com/tutorials/kwic-lexical-dispersion/

Open Legal Data

More and more Open Legal Data is becoming available and saves you a lot of data engineering work if you want to start practicing right away!

1. https://seanfobbe.com/data/
   - 18 legal datasets
   - Public Domain and Open Access
   - International law and German law
   - Extensive Documentation

2. https://github.com/Liquid-Legal-Institute/Legal-Text-Analytics
   - Many jurisdictons
   - Many authors

Contact

Tutorials, Blog, Open Data, Code, Advice

---

www.seanfobbe.com