

Expanding the OMOP Common Data Model in Accord with Federal Rules for Hospital Price Transparency and Transparency in Coverage

Dr. Jaan Altosaar Li
One Fact Foundation
University of Tartu University
jaan@onefact.org

Michele Tadiello
One Fact Foundation
michele@onefact.org

Rohan Bansal
Stanford University
rohan@onefact.org

BOOTCAMP DESCRIPTION. The material consequences of data and the artificial intelligence algorithms trained on data belie global hierarchies of power. For example, the field of algorithmic fairness may have emerged because of the economic advantages of providing advertising systems free of regulatory problems. Data-driven advertising practices, financial instrument design, credit scoring, tax planning, biomedical science, pharmaceutical research and development, housing, corporate real estate, and public infrastructure are all driven by data and the humans who interpret algorithms trained on this data and make decisions based on their interpretations that affect us all across these sectors of society. Becoming fluent in the abstractions necessary to represent the incentives and market dynamics of artificial intelligence is vital for learners who seek autonomy, and community organizations that seek to assess the impact they aim to achieve. Understanding and amplifying these goals of increased autonomy and impact for under-represented learners and community organizations they might work with requires understanding the human scale prior to building mental models of algorithms that scale. Human involvement is necessary across the discipline of data thinking, from data collection, curation, standardization, analysis, visualization, communication, and advertising, alongside other core data thinking skills. Decisions about capital allocation and human resources allocation must be made at each stage a data-to-decision journey. Many people prefer to live and work within countries and systems that prioritize the worst-off among a population. In this data thinking bootcamp for Tech Freedom Schools, we take this stance, and center the creation and delivery of educational materials on the emotional journey of a learner. Besides enabling autonomy and impact, this can help bridge educational gaps between under-represented community organizations and the PhD holders who build AI (both groups bear the consequences of algorithmic decision-making systems at scale). Through this collaboration between the One Fact Foundation and instructors the Ida B. Wells Just Data Lab in the Department of African American Studies and the Department of Sociology at Harvard University, we can help learners give informed consent from sociological, anthropological, and ethnographic lenses that are vital to understand the emotions, thoughts, and behaviors of people in power who deploy artificial intelligence---and truly give every learner a chance to decide if and when to subject themselves to algorithmic decision-making that may run the risk of ignoring some of these lenses.

WEEK 1: DESCRIPTIVISM AND PRESCRIPTIVISM IN LANGUAGE, HEALTH, PSYCHOLOGY, & CULTURE

The first week sets the stakes and exposes learners to difficult questions, and begins shaking the foundations for concepts that might traditionally be attended with psychological rigidity.

To start helping learners build psychological flexibility in approaching core data thinking skills, an experiential approach is necessary for exposure to difficult concepts and communication about these in semi-public spaces such as the learner team chat or GitHub issues where homework, reading, watching & listening exercises are assigned and experience reports recorded.

In English style and usage, there is no right answer. This is because practitioners of data operate on the principle of parsimony: shorter descriptions of things are easier to communicate about. For if a phenomenon *must* be communicated about a certain way (a 'prescriptive' stance), then the working memory of the practitioner would suffer from increased load: two things must be kept in mind instead of one, the first being a description of a phenomenon, and the second being the rules the phenomenon *must* be described with. But cognitive load prevents proper analysis of data and subsequent decision-making, and runs the risk of a practitioner omitting unobserved confounders from analysis and making false conclusions.

To illustrate these principles to learners, we ground our initial discussion on language, as a gentle introduction to the feeling of the rug being swept from under your feet as you realize that most things are a social construct subject to the forces of cultural evolution, collective behavior, history and so on: from race to gender, to the country you live in and the verbal events of thoughts, emotions, and feelings you use to communicate each day.

Topics. We illustrate the differences between prescriptivist and descriptivist stances toward the English language using the following readings:

Skills. The first week serves to begin exposing you to stimuli of your choice (such as embarking on this course! or choosing what to read).

- **Speaking on Zoom.** This helps socialize you to the common fear of public speaking, notice which other learners or instructors also share this fear, and expose ourselves collectively. Such exposure during the first week helps begin to dull the emotional response to the stimulus of public speaking and associated anxieties or fears, if they exist.
- **Signing up for new tools and paying attention to friction, pain points, cognitive processes.**
- **Writing.** Experience report.
- **Asking for help.**
- **Creating a visualization of 10 million 3-1-1 calls in New York City.**
- **Asking large language models for help reading.**

WEEK 2: MATHEMATICS, STATISTICS, AND THINKING LIKE A LIMITED PARTNER BUYING OUT

WEEK 4: VISUALIZATIONS AND METADATA

(Heer and Moritz 2023) will be used to visualize all of the team chat to date.

WEEK 5: ACTIVE LEARNING WITH A HUMAN IN THE LOOP

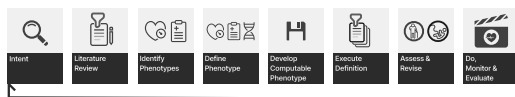


FIGURE 1. The phenotype workflow can help practice data thinking to best validate definitions of health and disease, intellectual property, the market dynamics, etc.

REFERENCES

Heer, J., & Moritz, D. (2023). Mosaic: An Architecture for Scalable & Interoperable Data Views. *Private communication and early access provided to one fact foundation - do not re-cite.*