



AI FOR GOOD
FOUNDATION

Bridging Science, Government, and Industry for a Better Future

James Hodson
CEO, AI for Good Foundation



AI FOR GOOD
FOUNDATION

Teaching People to Think with Data

The United Nation's goals for a better world by 2030 seem unlikely to be reached.

Can data and computation be **the engine to help us close the research and implementation gap on the biggest challenges of our generation?**

We think so.

My Experience

A scenic view of San Francisco, California, featuring the Golden Gate Bridge on the right side of the frame. The bridge's iconic red-orange towers and suspension cables are prominent against a blue sky with wispy white clouds. In the background, the San Francisco city skyline is visible across the water, including the Transamerica Pyramid. The foreground consists of a steep, grassy hillside with some yellow wildflowers and green shrubs. The overall atmosphere is bright and clear.

Studies

Summer Schools & Short Courses

Invited Talks

Europe & US Funded Projects

Hackathons, Challenges, Competitions

Industry Research Lab

Creation of Data Science Institutes

Data Science

A set of Algorithmic Methods and Engineering Practices that need a channel for development and adoption within empirical research.



The requisite Data Literacy to allow industry and society to harness the value of data to aid in solutions to a wide variety of challenges.

What We Want to Achieve



DATA DRIVEN SCIENCE

OPEN INFRASTRUCTURE, DATA, METHODS

DATA LITERACY ACROSS CAMPUS

DATA INNOVATION HUBS

**CROSS-DISCIPLINARY RESEARCH
+ TEACHING RESOURCES**

**DATA SCIENCE FOR
SOCIAL GOOD**

**ENCOURAGE
DIVERSITY**

The background image is a photograph of a rustic, old classroom. The walls and ceiling are made of dark, weathered wood. A large blackboard is mounted on the back wall. In the foreground and middle ground, there are several rows of simple wooden desks and benches. Two windows with multiple panes are visible on the left and right sides, letting in natural light.

Main Issues

Teach Methods Independently from the Research Process.

Seldom Teach how to Evaluate Data Collection Processes, Cleaning, Merging.

Teaching Applied Data Science in a Laboratory Setting is too Short, too Stylised, and has No Impact.

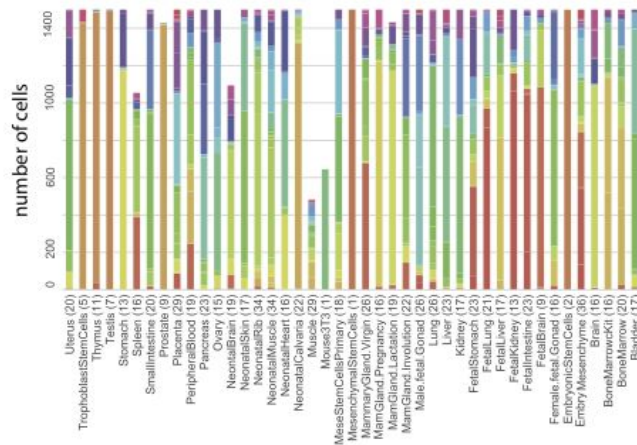
Challenge

Reasoning About the Unseen

Human Cell Atlas

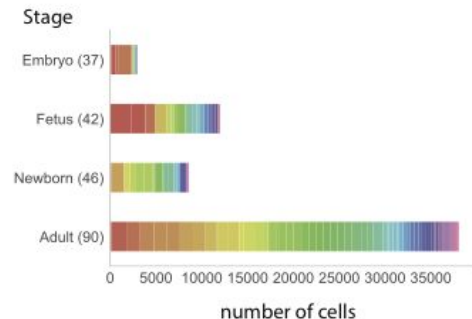
How Many Cell Types Have We Not Seen Yet?

A cell type distributions across mouse organs



B

cell type distributions across developmental stages

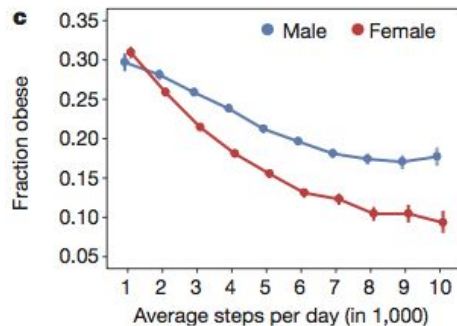


Challenge

Establishing Causal Relationships

Observational Data Alone Do Not Answer
“What If?” Questions

*“Do I walk too little because I am obese,
or am I obese because I walk too little?”*



Challenge

3

Ensuring Replicability of Results

Amgen could replicate only 6 of 53 studies considered landmarks in basic cancer science.

Pervasive; e.g. see Kaplan's report from FDA

Huge national and international concerns

- National Academy of Sciences
- Leading publications (*Science, Nature,...*)
- Leading professional societies
- President's Council of Advisers on S&T

Societal Concern: Public Confidence in Science?

Challenge

4

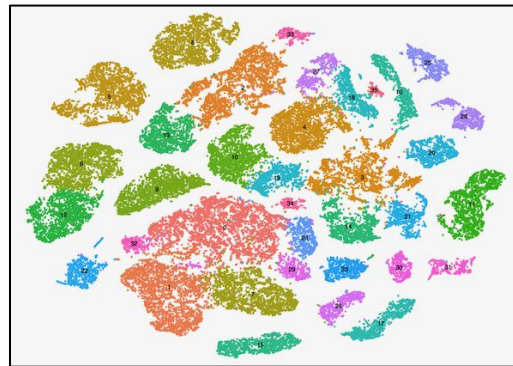
Visualizing Data Effectively

Datasets are Large, Unruly, & Difficult to Visualize.

Looking at data requires choosing one of many possible representations; this is not neutral or easy:

What cell types do we see in a tumor sample?

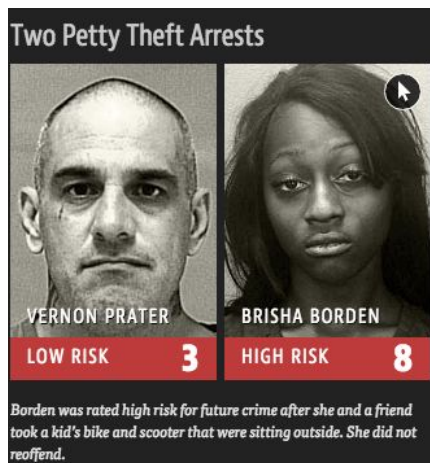
Visualization of single-cell mRNA-seq



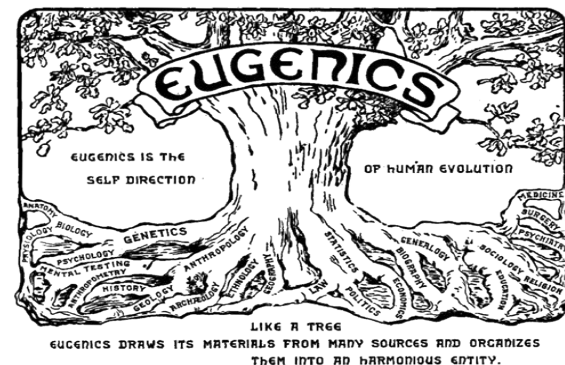
Challenge

Distinguishing Information from Structural Bias

Let us not reify our prejudices



Do recidivism prediction instruments produce discriminatory biases?



Why White Supremacists Are Chugging Milk (and Why Geneticists Are Alarmed)

Amy Harmon, *The New York Times*
October 17, 2018

Social Impact.



How does “Social Good” fit in?

A man with a beard is working on a large architectural drawing on a wooden table. He is wearing a light blue shirt and a watch. The drawing is spread out on the table, and he is using a ruler to measure it. The background is dark and out of focus.

Open Problems.

Open Data.

Open Source.

Open Infrastructure.

Open Innovation.

THE GLOBAL GOALS

For Sustainable Development



Bringing Social Impact into the Data Science Classroom



Courses need to last longer than a single semester in order to cover the full process.

Engaging with outside organisations that represent the social challenges.

Systems Analysis, Experimental Design, Data Collection, Causal Interpretation, Evaluation, Implementation, Monitoring, Social Review



Become a member.

Sign up to our Newsletter.

Partner with us.

<https://www.ai4good.org>
info@ai4good.org