Human Contexts and Ethics Guest Lecture

Fairness in Housing Appraisal

Contextualizing the Cook County Assessor’s Office Open Data Initiative

Ari Edmundson

Data 100, Spring 2023 @ UC Berkeley

3652993

slido

Join at slido.com
#3652993

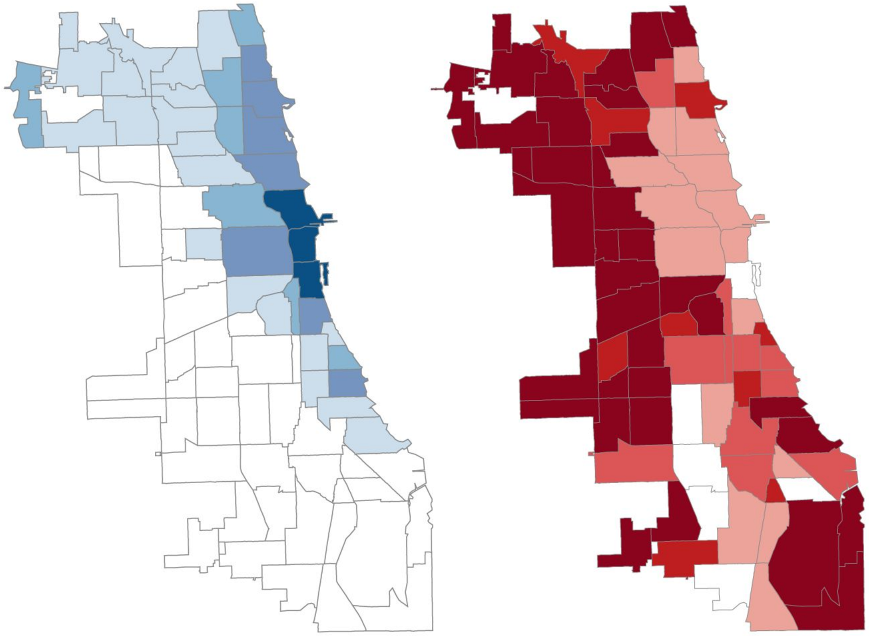
ⓘ Start presenting to display the joining instructions on this slide.

Case Study: Cook County Assessor’s Office and Property Appraisal

3652993

Cook County Assessor’s Office (CCAO) - Chicago, IL and surrounding townships

Charged with assessing property values in order to determine property taxes



COOK COUNTY GOVERNMENT

3652993

The Problem

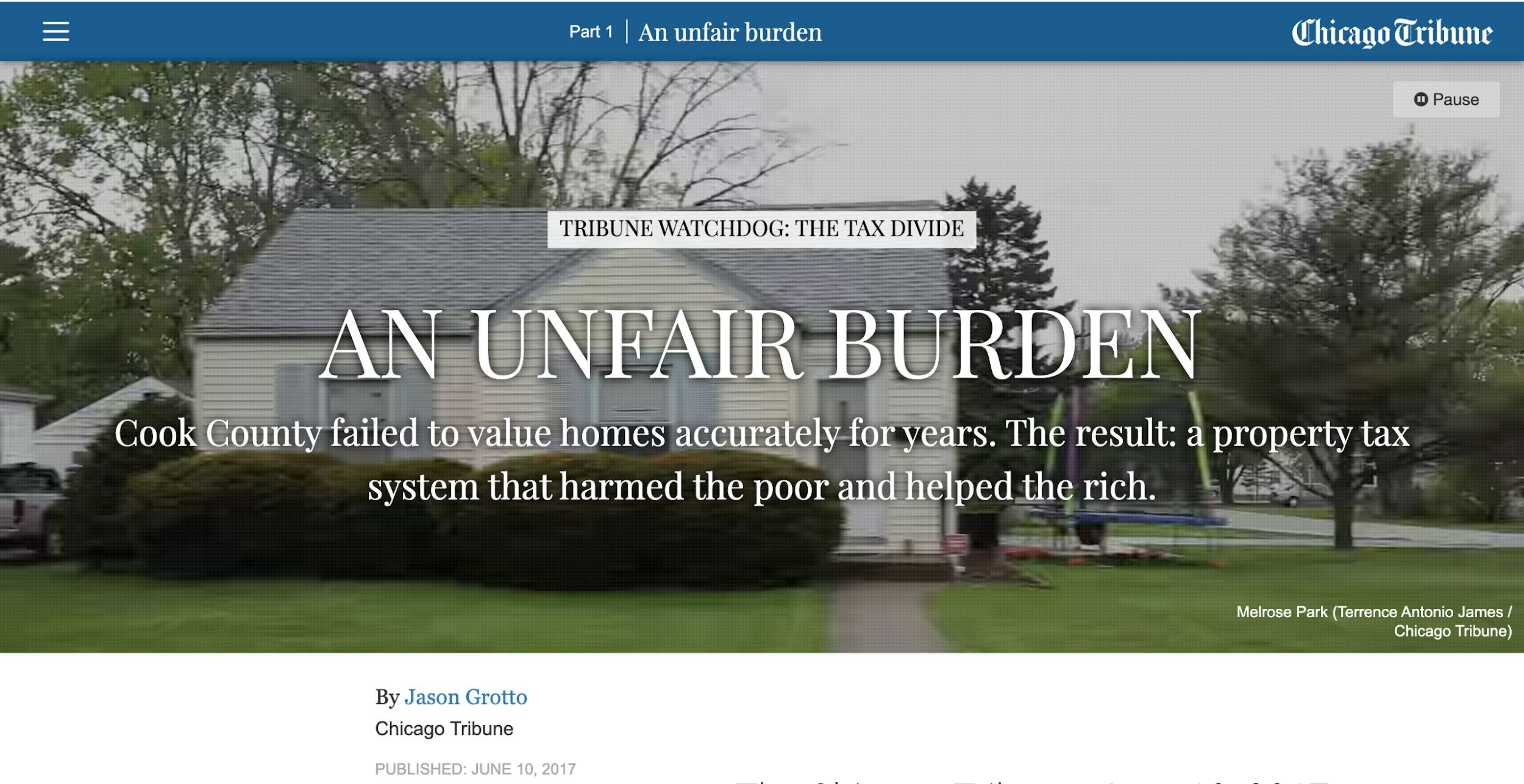
The Response

Key Takeaways

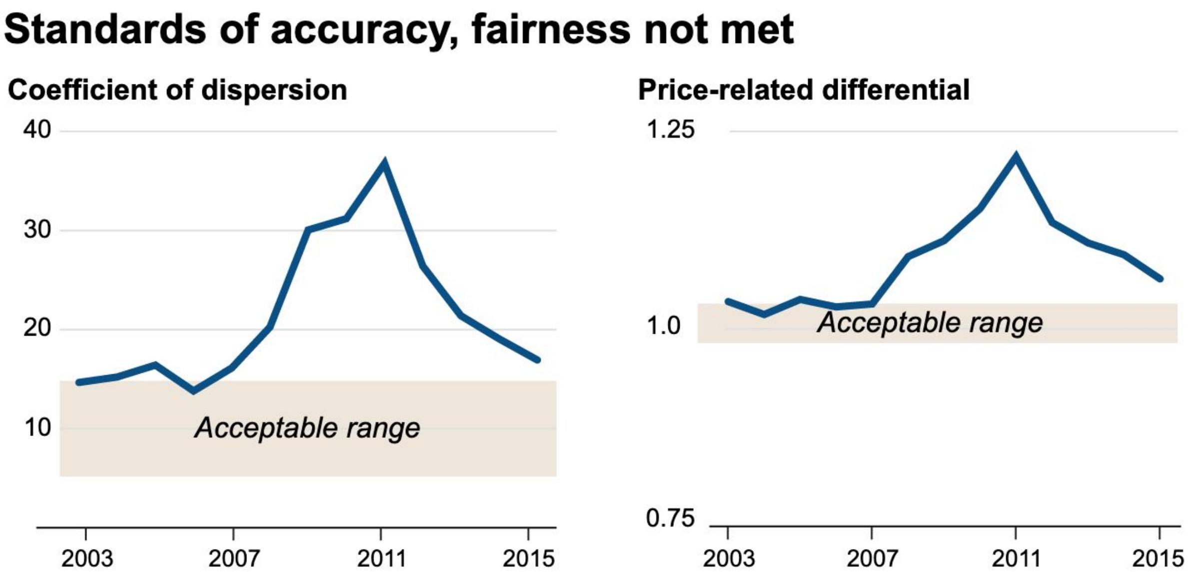
Lessons for Data Science Practice

3652993

The Problem



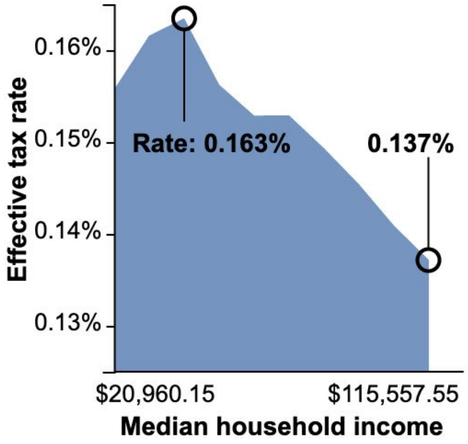
[The Chicago Tribune](), June 10, 2017



Sources: Cook County assessor's office, Illinois Department of Revenue, Tribune analysis

As income level rises, effective tax rates decline

These rates represent the percentage of homes' value that owners pay toward two taxing districts that cover all of Cook County. They are a small fraction of the overall tax rate but allow for comparisons between communities with widely differing tax bases.

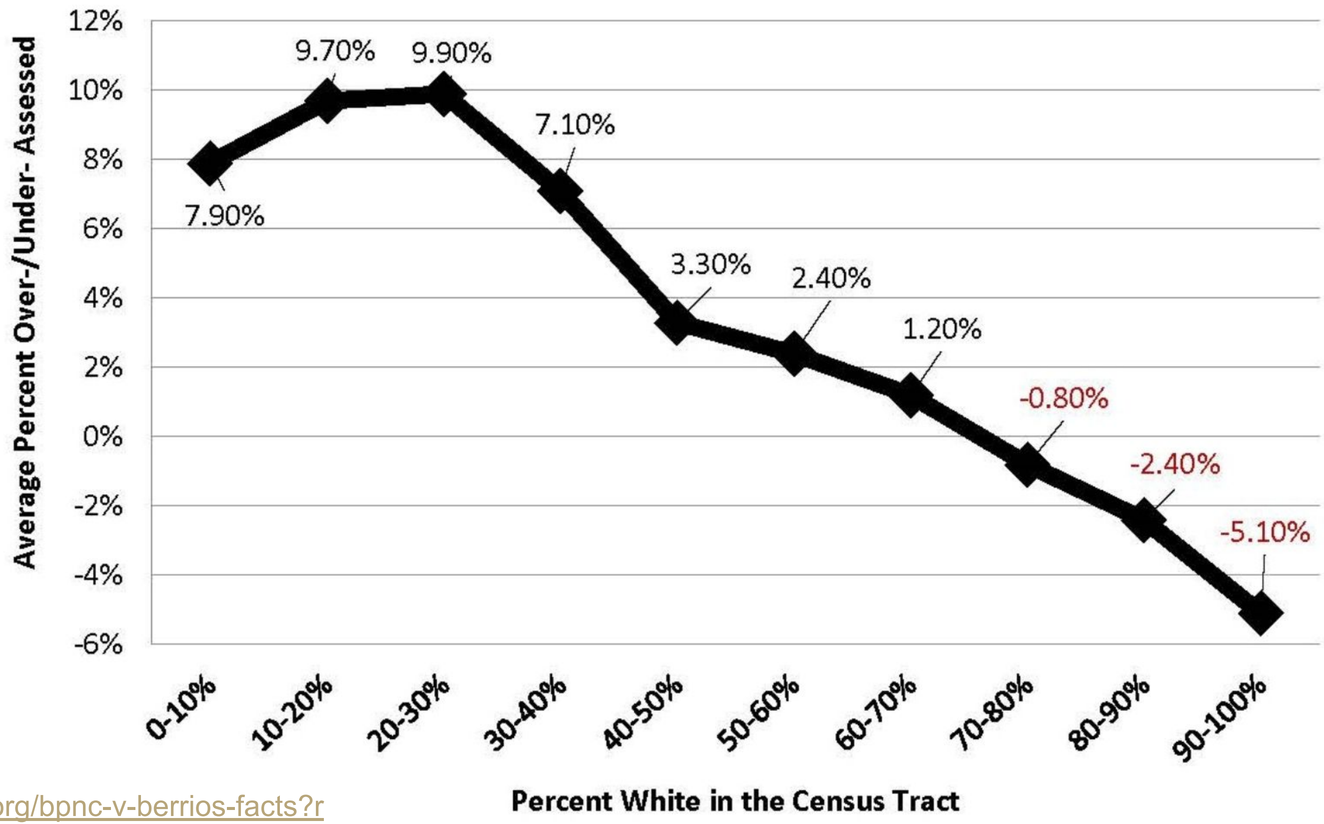


Sources: Cook County assessor's office, Cook County treasurer's office, Tribune analysis

The Chicago Tribune, June 10, 2017

3652993

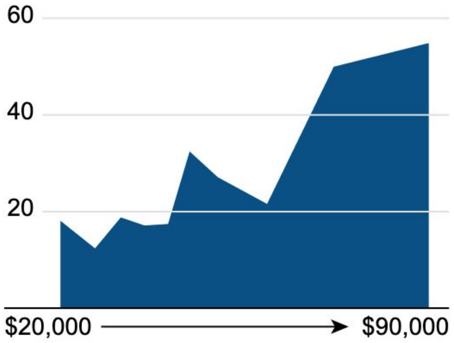
Chart 2: Average Percent Over/Under Assessment by Percent White, 2011-2015



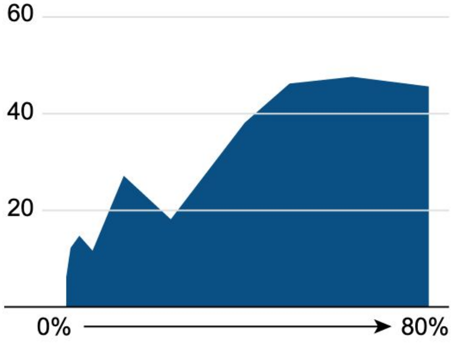
https://www.clccrul.org/bpnc-v-berrios-facts?rg=berrios

Neighborhood demographics affect appeals rate

By median income in 2015

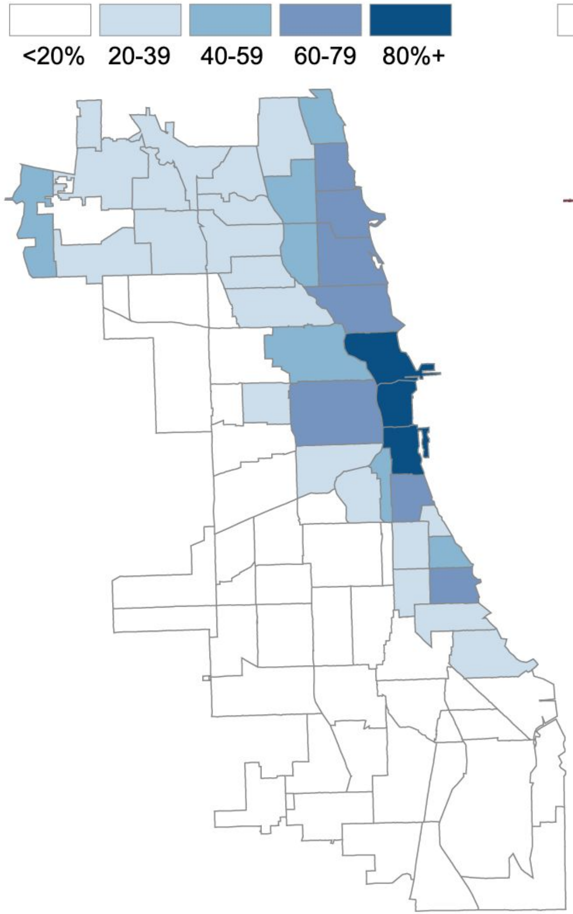


By percentage white in 2015

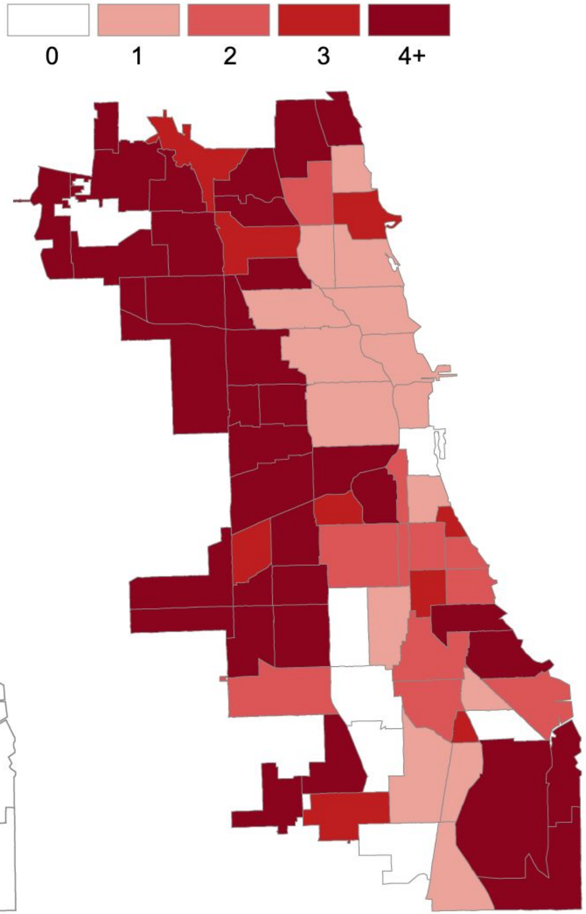


Sources: Cook County assessor's office, U.S. Census Bureau, Tribune analysis

Percentage of homeowners who appealed, 2011-15



Number of years homes were overvalued, 2011-15



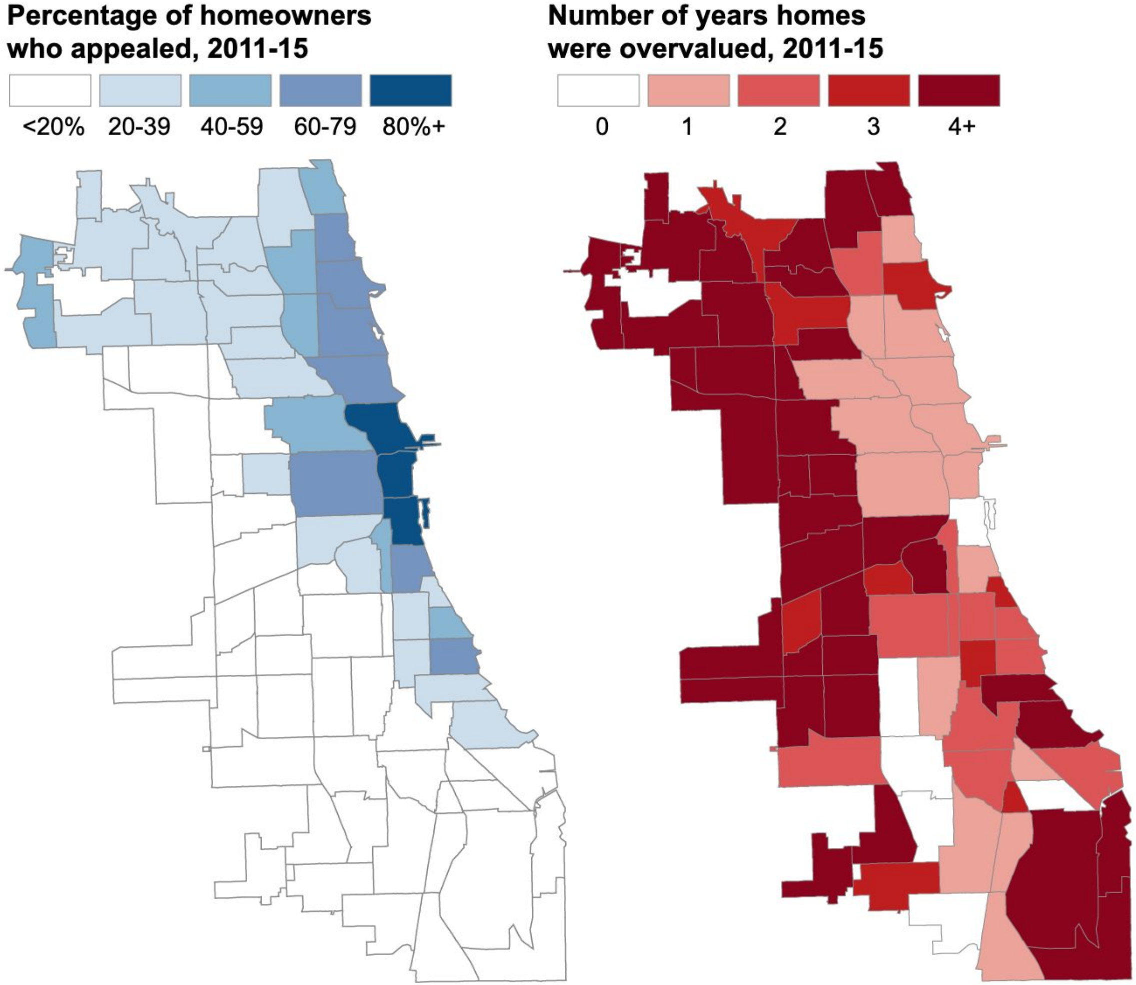
Sources: Cook County assessor's office, Illinois Department of Revenue, Tribune analysis

The role of the appeals process in producing inequity.

“Appeals are a good thing,” Thomas Jaconetty, deputy assessor for valuation and appeals, said in an interview. “The goal here is fairness. We made the numbers. We can change them.”

Fairness as equal access: “anyone can appeal” - but that’s not really the case:

Part of a deeper, institutional pattern, potential corruption



Sources: Cook County assessor’s office, Illinois Department of Revenue, Tribune analysis

Human impacts

3652993

In 2011 Braxton-Williams learned the assessor's office had valued the house at $147,550. "I love my house, but I know it's not worth that much," she said. "And they know it's not worth that much." (Terrence Antonio James/Chicago Tribune)

The Chicago Tribune, June 10, 2017

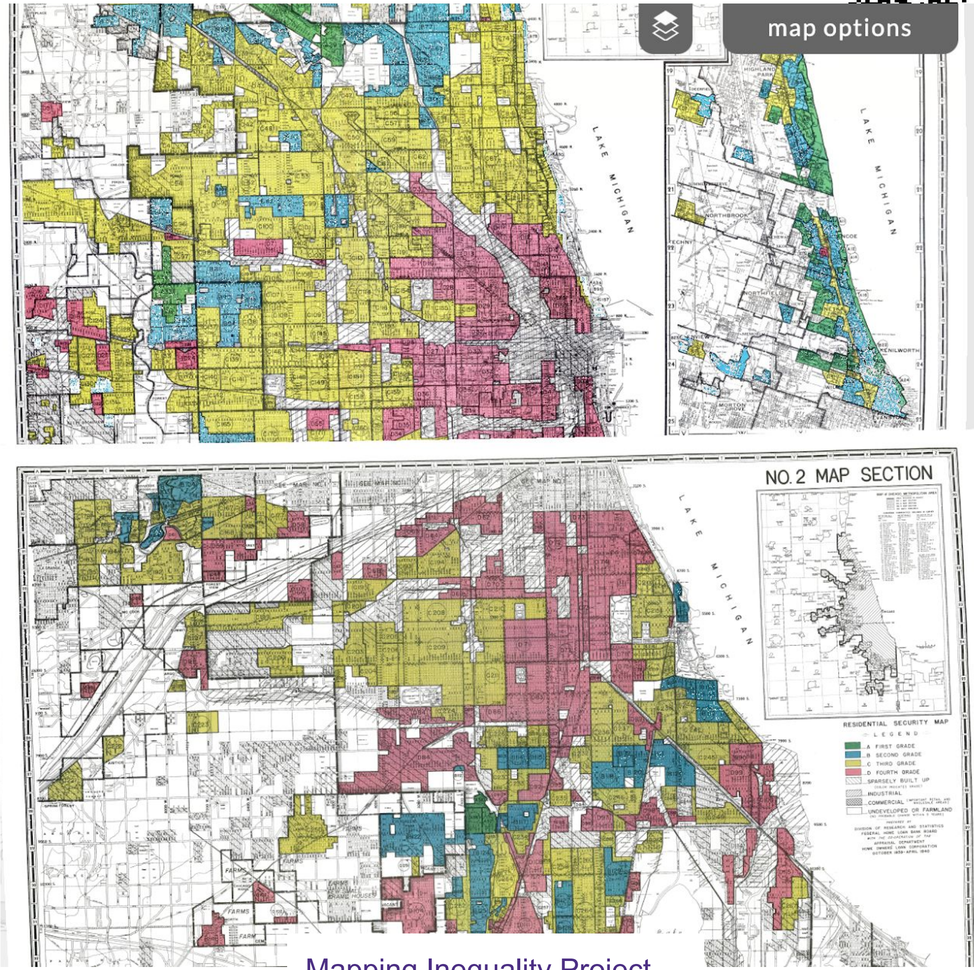
Real estate and racial inequality in the United States

Housing has been a key motor of racial inequality in modern US History

Segregation and credit-market racism

Redlining: making it difficult or impossible to get a federally-backed mortgage to buy a house in specific neighborhoods coded as “risky” (red).

What made them “risky” according to the makers of these maps? Their racial composition...



[Mapping Inequality Project]

3652993

Real estate and racial inequality in the United States

Segregation was not only a result of federal policy, but developed by real estate professionals

Real estate industry "professionalized" in the 1920's and 1930's by aspiring to become a science guided by strict methods and principles.

These methods centered on creating objective rating systems (information technologies) for the appraisal of property values…

which encoded race as a factor of valuation

and which, in turn, influenced federal policy and practice

Additions	Per Cent
Site	15
Type of neighborhood and social factors	20
View and climate	15
Public utilities and schools	5
Streets and alleys; distance to work in city	10
Contour and soil	5
Physical environment	20
Restrictions and planning	10

A table of common deductions reducing the above percentages of value follows:

Deductions	Per Cent
Noise and dirt, up to	25
Racial and foreign neighbors, up to	60
Adjacent vacancy, up to	20
Poor architecture, up to	20
Obsolescence, up to	70
Distances from city, work, schools, etc., up to	100
Nuisances (funerals, freight, trucks, etc.), up to	100
Dead-end streets, up to	15

Table of common deductions from a 1937 Appraising Manual (image from Colin Koopman, *How We Became Our Data* (2019) p. 137)

The Response

3652993

CCAO’s mandate under new Assessor, Fritz Kaegi

- Distributional equity in property taxation = properties of same value treated alike during assessments
- Creates new Office of Data Science,

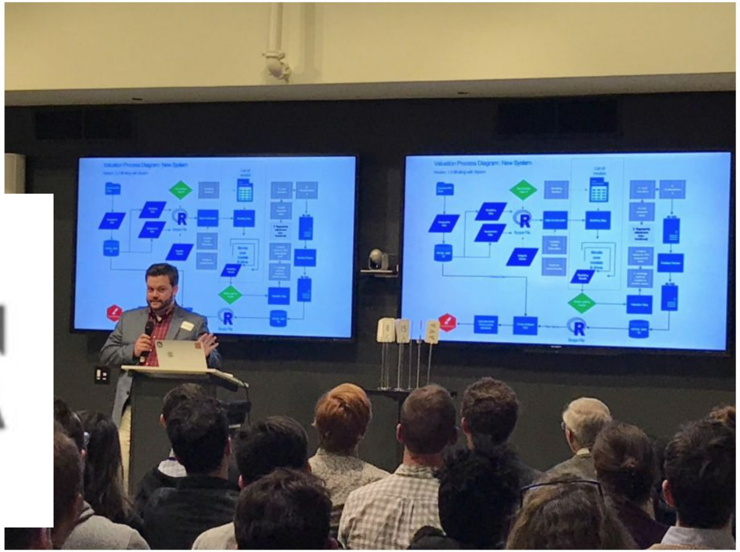
COOK COUNTY GOVERNMENT | OPEN DATA

Top highlig

Why the Cook County Assessor’s Office made its residential assessment code and data public — voluntarily

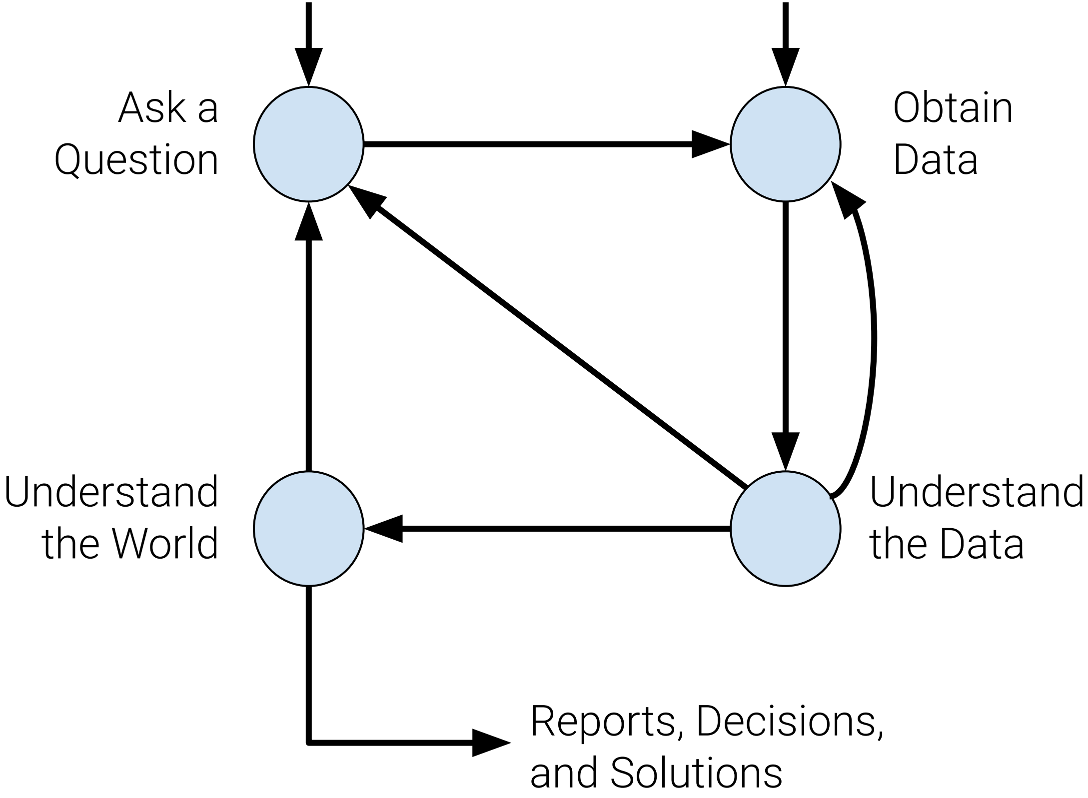
Cook County Assessor Apr 17, 2019 · 4 min read

By Robert Ross
Chief Data Officer, Cook County Assessor’s Office



Rob Ross, Chief Data Officer with the Cook County Assessor’s Office discusses the residential assessment modeling used by the CCAO during a presentation at Chi Hack Night on April 16th, 2019.

Data science lifecycle



The data science lifecycle is a **high-level description** of the data science workflow.

Note the two distinct entry points!

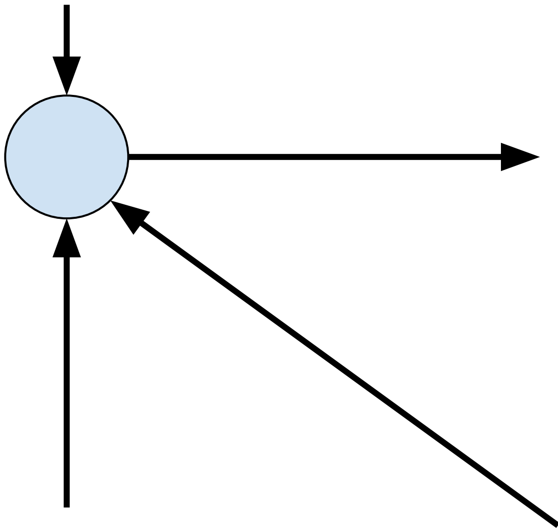
3652993

3652993

1. Question/Problem Formulation

- What do we want to know?
- What problems are we trying to solve?
- What are the hypotheses we want to test?
- What are our metrics for success?

Ask a Question



COOK COUNTY GOVERNMENT | OPEN DATA

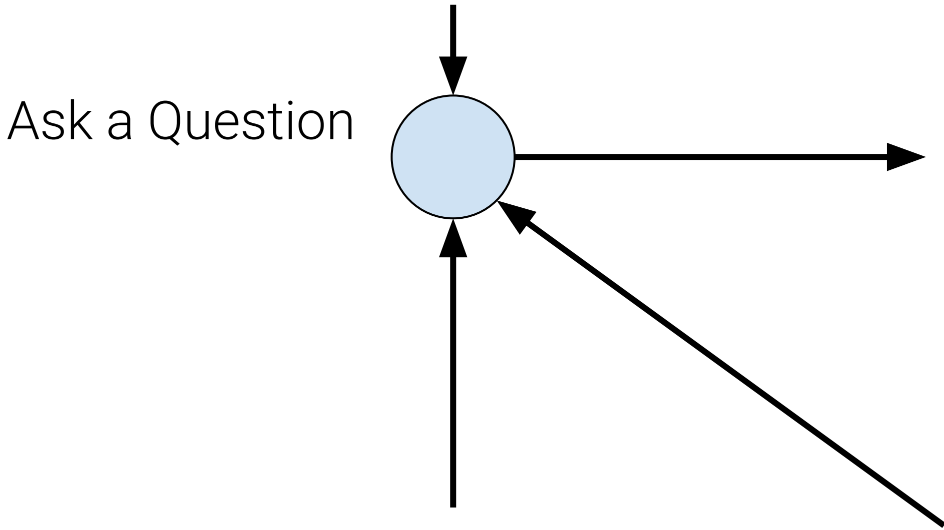
3652993

1. Question/Problem Formulation

What problems are we trying to solve?

1. Accurately, uniformly, and impartially assess the value of a home
   a. → accurately predict the sale price of a home within the next year
2. Create a system that assesses at scale, and and is fair to all people, across perceived racial and income differences

What are our metrics for success?

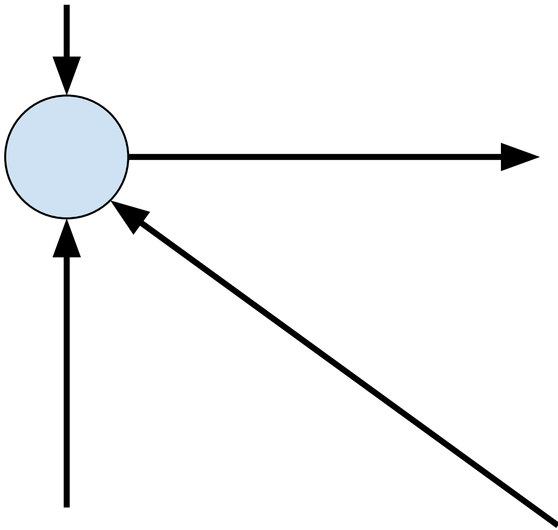


1. Question/Problem Formulation

What are our metrics for success?

1. The model predicts market values **accurately** and **uniformly**
    - Following international **standards** (like coefficient of dispersion)
2. The system is **fair, accountable, and transparent**,
    - thereby disrupts the circuit of corruption (Board of Review appeals process)
    - Eliminates regressivity
    - And engenders **trust** in the system among all stakeholders

Ask a Question



3652993

Fairness

Defined as: the ability of our pipeline to accurately assess all residential property values, accounting for disparities in geography, information, etc.

Transparency

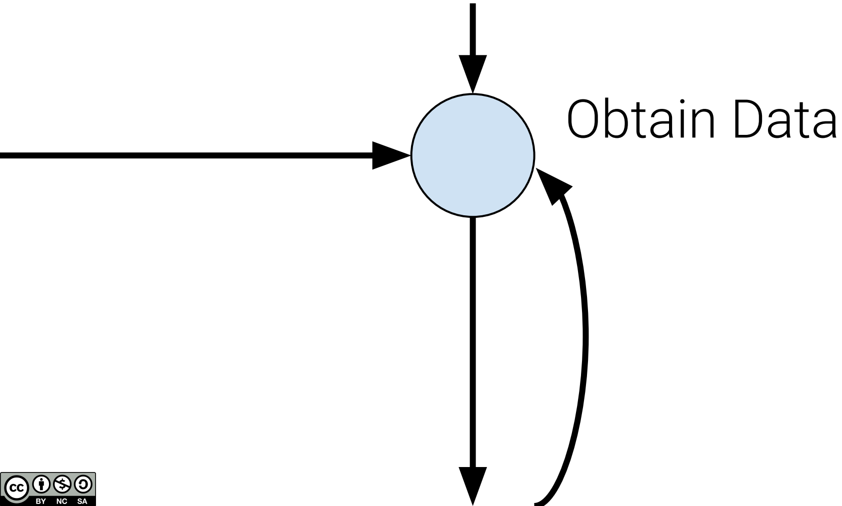
Defined as: the ability of the data science department to share and explain pipeline results and decisions to both internal and external stakeholders

https://gitlab.com/groups/ccao-data-science---modeling/-/epics/1

2. Data Acquisition and Cleaning

3652993

- What data do we have and what data do we need?
- How will we sample more data?
- Is our data representative of the population we want to study?



What’s in the data?

1. Sales data
   a. All recorded sales data 2013-2019
2. Property characteristics
   a. Property Identification Number
   b. Physical characteristics (Age, Bedroom, Baths, Square feet, Neighborhood, Site Desirability, etc.)

How was this data collected? When? By whom? For what purposes? How and why were particular categories created?

Feature Name	Category	Type
Age	Characteristic	numeric
Central Air Conditioning	Characteristic	categorical
Apartments	Characteristic	categorical
Attic Finish	Characteristic	categorical
Attic Type	Characteristic	categorical
Bedrooms	Characteristic	numeric
Building Square Feet	Characteristic	numeric
Basement	Characteristic	categorical
Basement Finish	Characteristic	categorical
Wall Material	Characteristic	categorical
Full Baths	Characteristic	numeric
Fireplaces	Characteristic	numeric
Garage 1 Area	Characteristic	categorical
Garage 1 Attached	Characteristic	categorical
Garage 1 Material	Characteristic	categorical

What’s in the data?

Are these attributes differentially reported?

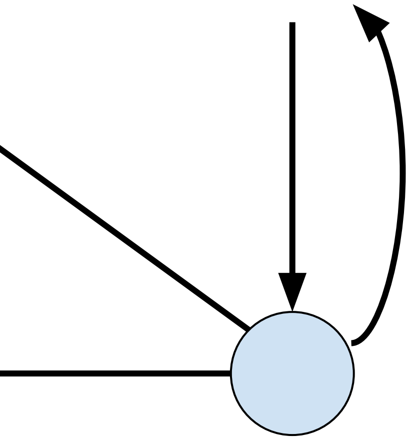
How are “improvements” (i.e. renovations) tracked and updated?

How might these attributes be differentially reported? Which data is missing, and for which neighborhoods or populations is it missing? And how do you know?

What other data sources might be valuable? Creating new attributes (flood plains, airport flight path)

Feature Name	Category	Type
Age	Characteristic	numeric
Central Air Conditioning	Characteristic	categorical
Apartments	Characteristic	categorical
Attic Finish	Characteristic	categorical
Attic Type	Characteristic	categorical
Bedrooms	Characteristic	numeric
Building Square Feet	Characteristic	numeric
Basement	Characteristic	categorical
Basement Finish	Characteristic	categorical
Wall Material	Characteristic	categorical
Full Baths	Characteristic	numeric
Fireplaces	Characteristic	numeric
Garage 1 Area	Characteristic	categorical
Garage 1 Attached	Characteristic	categorical
Garage 1 Material	Characteristic	categorical

3. Exploratory Data Analysis & Visualization



Understand the Data

- How is our data organized and what does it contain?
- Do we already have relevant data?
- What are the biases, anomalies, or other issues with the data?
- How do we transform the data to enable effective analysis?

What’s in the data?

Which attributes are most predictive of sales price?

Is the data uniformly distributed? Do all neighborhoods have up to date data? Same granularity? Or do some neighborhoods have missing or outdated data?

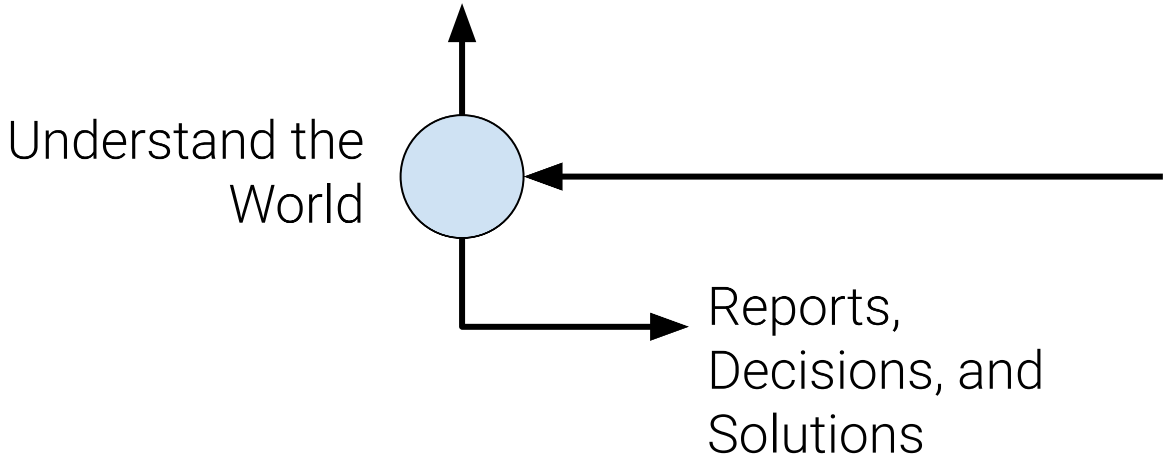
CCAO noticed that low income neighborhoods had disproportionately spottier data

Need to develop new data collection practices--including finding new sources of data

4. Prediction and Inference

3652993

- What does the data say about the world?
- Does it answer our questions or accurately solve the problem?
- How robust are our conclusions and can we trust the predictions?



Using machine learning for mass appraisal

3652993

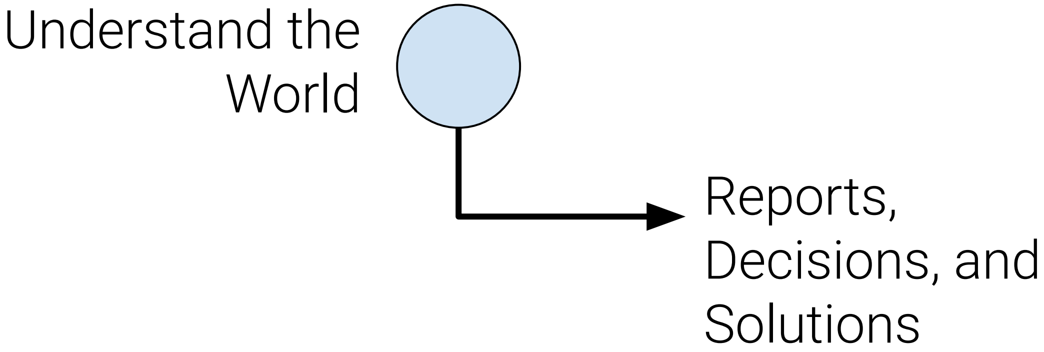
Predict sale price (“fair market value”) of unsold properties by discovering patterns in data sets containing known sale prices and characteristics of similar and nearby properties.

Compared to traditional mass appraisal, the CCAO’s approach is more granular and more sensitive to neighborhood variations

Uses a different model for each township

Reports, Decisions, and Conclusions

3652993

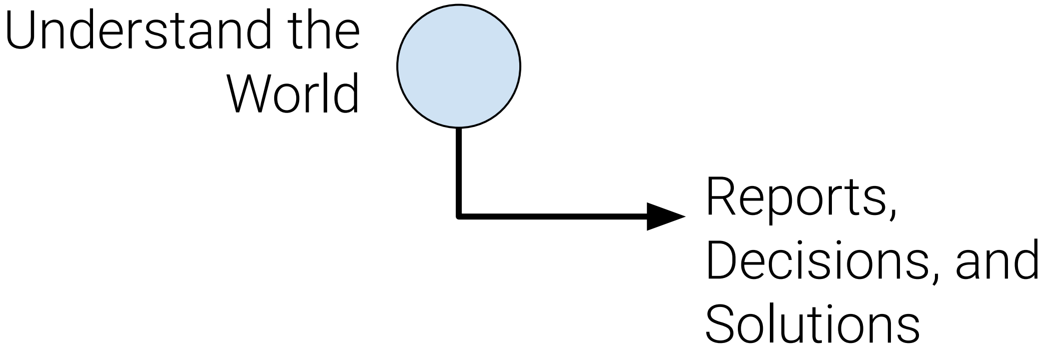


3652993

Reports, Decisions, and Conclusions

Linking the solutions to the two "problems"-- 1) accuracy/uniformity of the model, and 2) a fair and transparent system that eliminates regressivity and engenders trust

How successful is the system? And how do you know? Is the system fair and transparent?



Justice: as *fairness* as *accuracy* as *predictive power* as *uniformity* as *impartiality*…

3652993

Fairness as algorithmic (quantitative, formal, objective, accurate, predictive, impartial) vs public appeals process with human “checks” in the loop?

Centering role of the “human” and “society” in and around the algorithm

Predictive power can come at the cost of “explainability” and “control”

- Appeal to “expertise” and the quality of outcomes as stopgap
    - What happens to transparency?
    - What happens to community engagement?
- Can transparency invite new kinds of corruption? Attempts to “game” the system by exploiting loopholes in the code?
- What forms of justice does it leave out?
- “Fair market value” as optimal target presumes fairness of markets

Key Takeaways

3652993

3652993

1. Accuracy is a necessary, but not sufficient, condition of a fair system.

2. Fairness and transparency are context-dependent and sociotechnical concepts

3. Learn to work with contexts, and consider how your data analysis will reshape them

3652993

4. Keep in mind the power, and limits, of data analysis

3652993

In 2011 Braxton-Williams learned the assessor's office had valued the house at $147,550. "I love my house, but I know it's not worth that much," she said. "And they know it's not worth that much." (Terrence Antonio James/Chicago Tribune)

The Chicago Tribune, June 10, 2017

What is a home "worth"?

Lessons for Data Science Practice

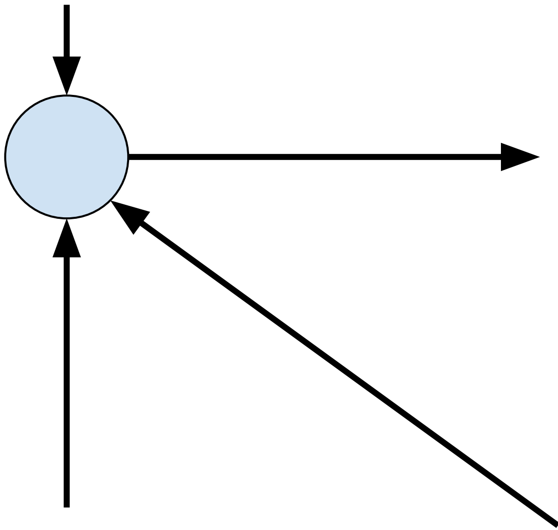
3652993

3652993

1. Question/Problem Formulation

- What do we want to know?
- What problems are we trying to solve?
- What are the hypotheses we want to test?
- What are our metrics for success?

Ask a Question

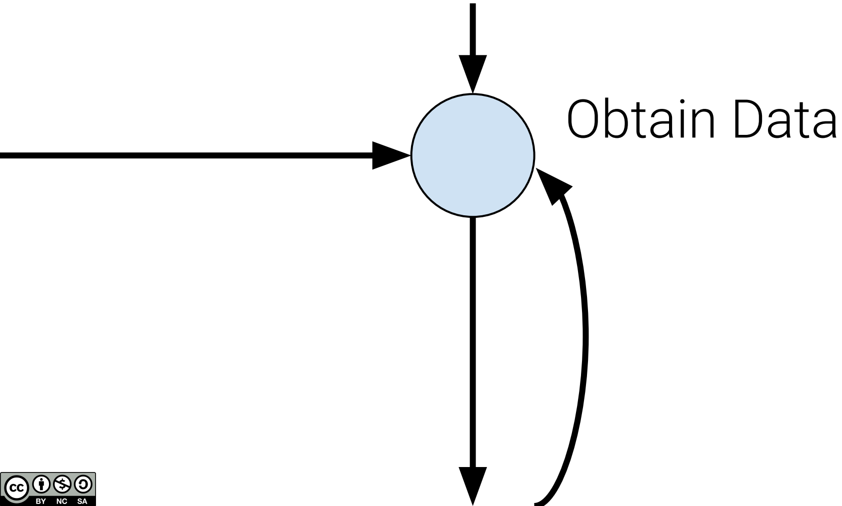


- Who is responsible for framing the problem?
- Who are the stakeholders? How are they involved in the problem framing?
- What do you bring to the table? How does your positionality affect your understanding of the problem?
- What are the narratives that you're tapping into?

2. Data Acquisition and Cleaning

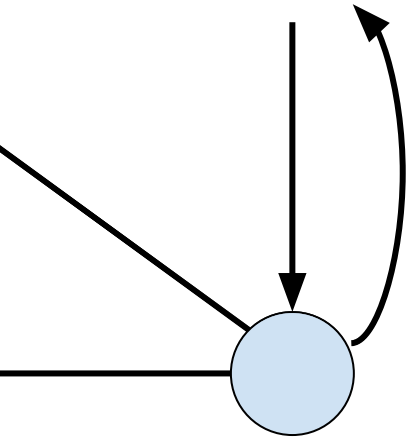
3652993

- What data do we have and what data do we need?
- How will we sample more data?
- Is our data representative of the population we want to study?



- Where does the data come from?
  - Who collected it? For what purpose?
- What kinds of collecting and recording systems and techniques were used?
- How has this data been used in the past?
- What restrictions are there on access to the data? What enables you to have access?

3. Exploratory Data Analysis & Visualization



Understand the Data

- How is our data organized and what does it contain?
- Do we already have relevant data?
- What are the biases, anomalies, or other issues with the data?
- How do we transform the data to enable effective analysis?

- What kind of personal or group identities have become salient in this data?
- Which variables became salient, and what kinds of relationship obtain between them?
- Do any of the relationships made visible lend themselves to arguments that might be potentially harmful to a particular community?

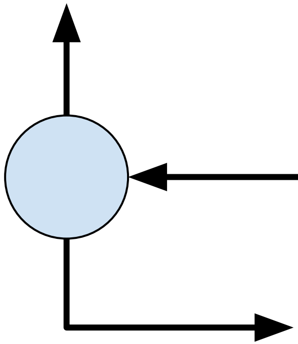
3652993

4. Prediction and Inference

- What does the data say about the world?
- Does it answer our questions or accurately solve the problem?
- How robust are our conclusions and can we trust the predictions?

- What does the prediction or inference *do* in the world?
- Are the results useful for the intended purposes?
- Are there benchmarks to compare the results?
- How are your predictions and inferences dependent upon the larger system in which your model works?

Understand the World



Reports, Decisions, and Solutions

3652993

Write down one take-away from today's lecture

ⓘ Start presenting to display the poll results on this slide.