

# DATA 6 lecture - Exploring Human Contexts and Ethics of Data with the Bay Area Air Quality Management District

Cathryn Carson  
UC Berkeley  
8/4/22



# What's today for?

- Introductions
- Case study: Bay Area Air Quality Management District
  - What's in the Public Data Center?
  - “Making” data
  - Representation
- Human Contexts and Ethics (HCE)
  - Toolkit
  - Lifecycle (time permitting)

# Introductions



Cathryn Carson

## Air Quality Forecast Map



Enter address...

### AIR QUALITY FORECAST

#### Daily Forecast

Last Updated: Wednesday, 8/3/2022 at 03:11 AM

08/03

08/04

08/05

08/06

08/07



#### Air Quality Index

(0-50)  
Good

(51-100)  
Moderate

(101-150)  
Unhealthy for  
Sensitive  
Groups

(151-200)  
Unhealthy

(201-300)  
Very Unhealthy

(301-500)  
Hazardous

### Five Day Forecast

### More AQI Information



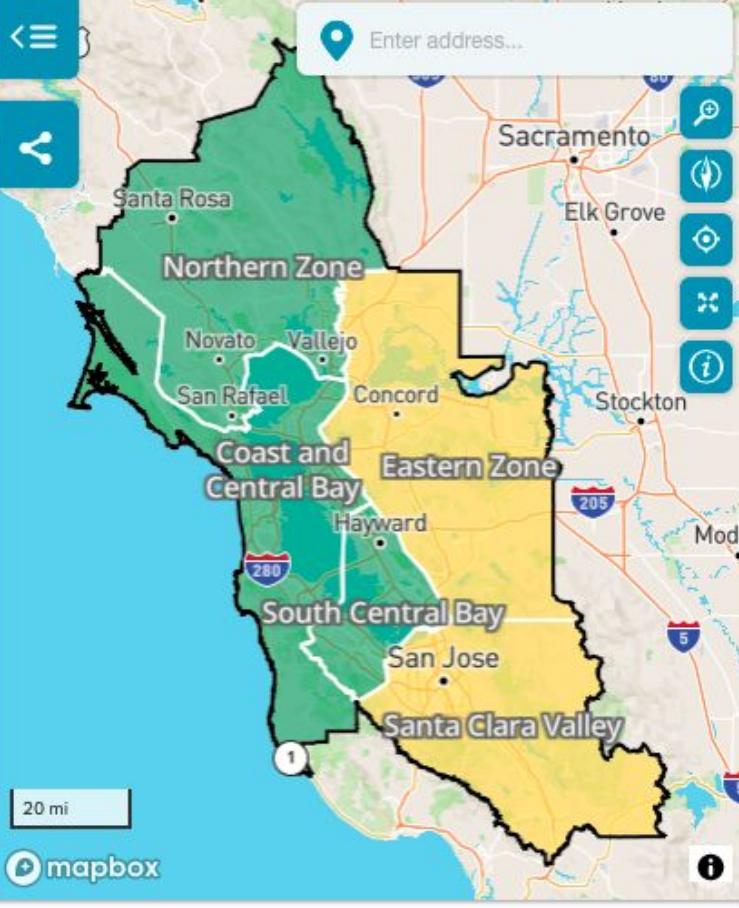
#### BOUNDARIES

20 mi

mapbox



AQI Reporting Zones





## What you breathe Air Quality in the Bay Area

Current Air Quality →  
Air Quality Forecast Map →  
Open Burn Map →  
Air Quality Summaries →

## Who we regulate Permits, Facility, and Rules information

Permit Applications Received →  
Public Notices on Permit Applications →  
Synthetic Minor Facilities →  
Title V Permits →  
Public Notices on Facility Risk Reduction →  
Facilities Maps →  
Current Rules →

## What we measure Emissions, Standards, and other Data

101  
001  
101  
110  
Emission Reduction Credit Bank Status →  
Data on Emissions Inventories →  
Air Quality Standards and Attainment Status →  
GHG Data →  
Flare Data →  
CARE Data →  
Air Monitoring Maps →

## How to get involved Interact with the Air District

Make an Air Quality Complaint →  
Subscriptions and Notifications →  
Board Meetings →  
Annual Budget →

# Starting questions

<https://www.baaqmd.gov/public-data-center>

1. What data can I find here?
2. What questions would be interesting to explore with this data?
3. What contextual information would be helpful for me to understand why it's there?

Anything else to add? Observations? Questions? For discussion?

*Please compare notes at your table:*

*Work in a Google doc or slide deck for your table, put it on the screen  
Put your names and intended majors at the top*





Ozone (O<sub>3</sub>)



Fine Particulate Matter  
(PM<sub>2.5</sub>)



Black Carbon (BC)



Carbon Monoxide  
(CO)



Hydrogen Sulfide  
(H<sub>2</sub>S)



Nitric Oxide (NO)



Nitrogen Dioxide  
(NO<sub>2</sub>)



Oxides of  
Nitrogen (NO<sub>x</sub>)



Sulfur Dioxide  
(SO<sub>2</sub>)



Oxides of Nitrogen (NO<sub>x</sub>)

Measurement Units: parts per billion (ppb)

[learn more](#)

Location:

All Zones



Start Date:

9/6/2007

Interval:

Monthly

Weekly

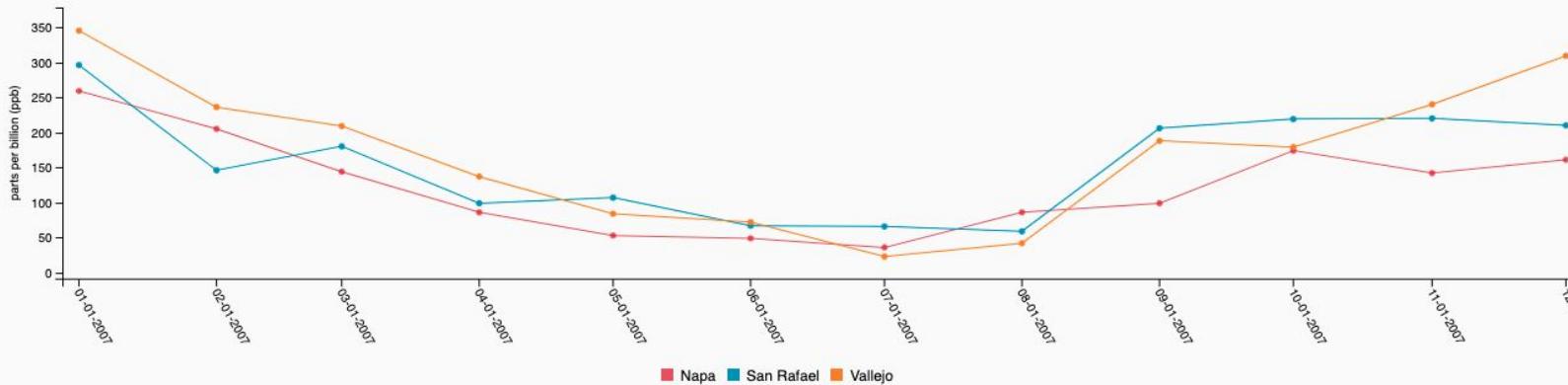
Daily

Hourly

Northern Zone

Monthly

< 2007 >



YEARLY

Napa  
High 259  
Avg 20

San Rafael  
High 296  
Avg 27

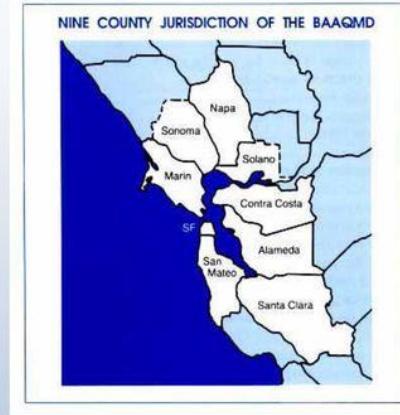
Vallejo  
High 345  
Avg 24



# BAAQMD Background

## The Air District

- Established in 1955
- Serves the nine Bay Area Counties
- Seven million residents
- 5,340 square miles



## Mission

To protect and improve public health, air quality, and the global climate



check out <https://www.baaqmd.gov/about-the-air-district/history-of-air-district>

### ABOUT AIR QUALITY

#### AIR QUALITY SUMMARIES

#### CURRENT AIR QUALITY

#### AIR QUALITY MEASUREMENT

#### INTERACTIVE DATA MAPS

### RESEARCH & DATA

#### Air Quality Standards & Attainment Status

#### Emissions Inventories

- Criteria Air Pollutants
- Production-Based GHG Emissions Inventory
- Consumption-Based GHG Emissions Inventory
- Toxic Air Contaminants
- Local Studies

#### Maps Data and Documents

#### Research & Modeling

#### Forecasting & Data Analysis

#### Refinery Flare Monitoring

#### Community Air Risk Evaluation (CARE) Program

### OPEN BURN STATUS

### AIR QUALITY FORECAST

### INCIDENTS AND ADVISORIES

#### SPARE THE AIR

#### WHAT YOU CAN DO

#### GLOSSARY

#### WILDFIRE AIR QUALITY RESPONSE PROGRAM

## AIR QUALITY FORECAST

### INCIDENTS AND ADVISORIES

### SPARE THE AIR

### WHAT YOU CAN DO

### GLOSSARY

### WILDFIRE AIR QUALITY RESPONSE PROGRAM

### WOOD SMOKE

### LAB & SOURCE TEST

### SPECIAL AIR MONITORING PROJECTS



## Flare Refinery Archives

Year: **2022** 2021 2020 2019 2018 2017 2016 2015

### — Chevron Richmond

Report by Month - 2022

Alky-Poly	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
D&R	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Fluidized Catcracker	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Hydrogen H2	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Low Sulfur Fuel Oil	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
North Isomax	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Richmond Lube Oil Project	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC
South Isomax	JAN	FEB	MAR	APR	<b>MAY</b>	JUN	JUL	AUG	SEP	OCT	NOV	DEC

### ⊕ Phillips 66 Rodeo

Report by Month - 2022

### ⊕ Shell Martinez

Report by Month - 2022

### ⊕ Tesoro Martinez

Report by Month - 2022

### ⊕ Valero Benicia

Report by Month - 2022

## Refinery Stakeholders

spare the air

# BAAQMD Refinery Flare Emission Report



Report Settings ▾

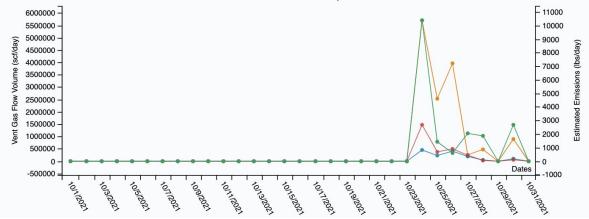
Chevron Richmond - North Isomax

October - 2021

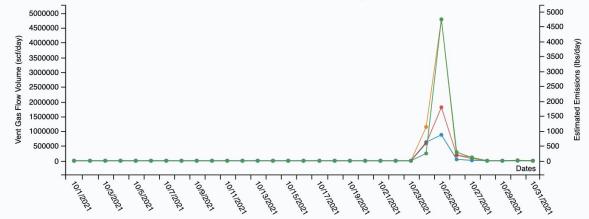
[Download CSV](#)

Date (mo/day/yr)	Vent Gas Flow (volume in scf) See notes 6,7	Methane (lbs)	NMHC (lbs)	Sulfur Dioxide (lbs)
10/1/2021	0.00	0.00	0.00	0.00
10/2/2021	0.00	0.00	0.00	0.00
10/3/2021	0.00	0.00	0.00	0.00
10/4/2021	0.00	0.00	0.00	0.00
10/5/2021	0.00	0.00	0.00	0.00
10/6/2021	0.00	0.00	0.00	0.00
10/7/2021	0.00	0.00	0.00	0.00
10/8/2021	0.00	0.00	0.00	0.00
10/9/2021	0.00	0.00	0.00	0.00
10/10/2021	0.00	0.00	0.00	0.00
10/11/2021	0.00	0.00	0.00	0.00
10/12/2021	0.00	0.00	0.00	0.00
10/13/2021	0.00	0.00	0.00	0.00
10/14/2021	0.00	0.00	0.00	0.00
10/15/2021	0.00	0.00	0.00	0.00
10/16/2021	0.00	0.00	0.00	0.00
10/17/2021	0.00	0.00	0.00	0.00
10/18/2021	0.00	0.00	0.00	0.00
10/19/2021	0.00	0.00	0.00	0.00
10/20/2021	0.00	0.00	0.00	0.00
10/21/2021	0.00	0.00	0.00	0.00
10/22/2021	0.00	0.00	0.00	0.00
10/23/2021	0.00	0.00	0.00	0.00
10/24/2021	5,701,643.79	827.73	2,683.92	10,379.10
10/25/2021	2,531,180.70	424.37	684.98	1,434.49
10/26/2021	3,962,224.05	740.40	911.15	609.12
10/27/2021	256,686.00	351.93	441.84	2,051.86
10/28/2021	476,387.27	96.49	52.16	1,868.24
10/29/2021	0.00	0.00	0.00	0.00
10/30/2021	895,142.38	181.40	105.44	2,681.75
10/31/2021	0.00	0.00	0.00	0.00
<b>October Totals</b>	<b>13,823,264.19</b>	<b>2,622.32</b>	<b>4,879.49</b>	<b>19,024.56</b>

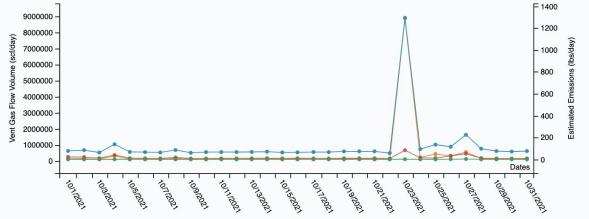
## Chevron Richmond North Isomax Flare Report 10/1/2021 - 10/31/2021



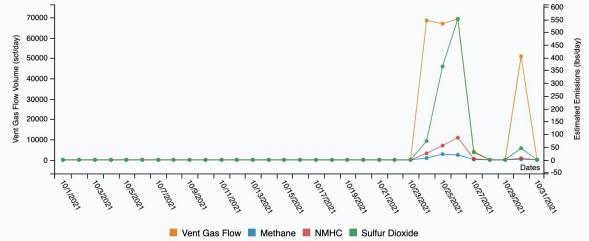
## Chevron Richmond Fluidized Catcracker Flare Report 10/1/2021 - 10/31/2021



## Chevron Richmond Hydrogen H2 Flare Report 10/1/2021 - 10/31/2021



## Chevron Richmond South Isomax Flare Report 10/1/2021 - 10/31/2021



## Flare Causal Reports

Air District / About Air Quality / Research & Data / Refinery Flare Monitoring / Flare Causal Reports



View Flare Causal Analysis Reports for flaring activity at Bay Area refineries.

## FILTERS ▾ Expand to show search or filter options

[◀](#) [◀](#) [1](#) [2](#) [3](#) [4](#) [5](#) ... [▶](#) [▶](#)

Items per Page: [10](#)

120 items in 12 pages

Start Date	End Date	Causal Report	Flare	Refinery
11/11/2021	11/11/2021	<a href="#">Report</a> (639 Kb PDF, 3 pgs, posted 1/5/2022)	Main	Phillips 66 Rodeo
10/30/2021	10/30/2021	<a href="#">Report</a> (182 Kb PDF, 5 pgs, posted 1/5/2022)	North Isomax	Chevron Richmond
10/27/2021	10/27/2021	<a href="#">Report</a> (162 Kb PDF, 4 pgs, posted 1/5/2022)	Hydrogen H2	Chevron Richmond
10/27/2021	10/28/2021	<a href="#">Report</a> (186 Kb PDF, 5 pgs, posted 1/5/2022)	North Isomax	Chevron Richmond
10/25/2021	10/25/2021	<a href="#">Report</a> (2 Mb PDF, 3 pgs, posted 1/5/2022)	Light Oil Products	Martinez Refining

# Chevron Refinery Malfunction During Storm Shut Down Processing Units, Causing Fire and Toxic Flaring

By Ted Goldberg Oct 28, 2021 [Save Article](#)



The Chevron refinery in Richmond. (Justin Sullivan/Getty Images)

A loss of steam production at Chevron's Richmond refinery during a major rainstorm on Sunday morning triggered the shutdown of several processing units, leading to a fire at the facility and several days of flaring, the oil giant told regulators this week.

Refinery officials say close to 17 tons of sulfur dioxide were released over two days as the facility sent gases to its flares, a safety technique often used by refineries to ease pressure and stabilize operations.

The accident prompted dozens of complaints from nearby residents, and is now being investigated by the regional air district and county public health officials to determine whether the odor led to the closure of several local schools. The incident, along with another malfunction at the PBF Refinery in nearby Martinez, may also temporarily increase the average cost of gasoline in California, according



The Air District is closely monitoring the current incident at the **#Chevron** Refinery in Richmond. Air District inspectors are on scene investigating and we are working to find the root cause of the smoke in addition to any potential air quality violations.



3:21 PM · Aug 10, 2021 · Twitter Web App

## Flaring activity at Chevron Richmond Refinery sends off huge plume of black smoke

Tuesday, August 10, 2021



# What's in the data?

- Overall air quality measurements
  - Time series of air quality indices and “criteria pollutants”
  - At particular, specifically selected locations
- Information about refinery flares
  - Time series for particular pollutants at individual refineries
  - Time series of estimated emissions/pounds per day
  - Causal Analysis Reports for incidents
- Complaint data
- And more

# What's not in the data?



## THE West Berkeley Alliance for Clean Air and Safe Jobs HISTORY

This is a history of Pacific Steel Casting Company, the industry's local and regional regulators, and the East Bay community their actions impact.

Pacific Steel Casting Company (PSC), the City of Berkeley, and the Bay Area Air Quality Management District (BAAQMD) have been at odds with neighbors for years. This is partially because of PSC's uncooperativeness and foul emissions, partly because of the City of Berkeley's lack of meaningful action, and partly because of BAAQMD's impenetrable regulatory bureaucracy and nearly unusable complaints policy.

Neighbors have long known that PSC's burning pot handle/burning brake odor is attributed to phenol and formaldehyde, as well as various other particulates and gases. These substances include recognized and suspected carcinogens, suspected cardiovascular or blood toxicants, developmental toxicants, immunotoxicants, kidney toxicants, gastrointestinal or liver toxicants, neurotoxicants, reproductive toxicants, respiratory toxicants, and skin or sense organ toxicants. Although they have complained and organized to have the emissions filtered out of the air, the odors are a recurring nightmare for residents in El Cerrito, Albany, Berkeley, and Kensington for the past 25 or more years. Neighbors won some partial victories, with PSC installing a carbon adsorption system on Plant #2 in 1985, and on Plant #1 in 1991. A carbon adsorption system in Plant #3 became operational in October of 2006. However, the problems were not fully addressed and the odor nuisance continues.

# Making data at the Air District

- What's "the phenomenon" in the world?
- What work goes into making the data?
- Why is it getting made?
- Who gets to define it?
- Who participates? How?
- Why ask these particular questions? Who cares?

Making data, not just collecting ...



# Data takes work. It does work.

Data is *made*. By *people*. To *do work* in the world. In *contexts* that shape what they do.

Data *travels* with hidden effects of how it was made - the *work* that went into making it - the *choices* that were made about how to have it *represent* the world.

# Data **represents** the world

Making data begins with an act of seeing and recording -- seeing and recording something that was previously, hidden, nameless, even unseen.

Arbitrary observations are not data. A dataset is not just an arbitrary jumble.

Data, as data, must have meaning as *standing for* a phenomenon in the world.

# Representation is selective

When people make data, they make choices about

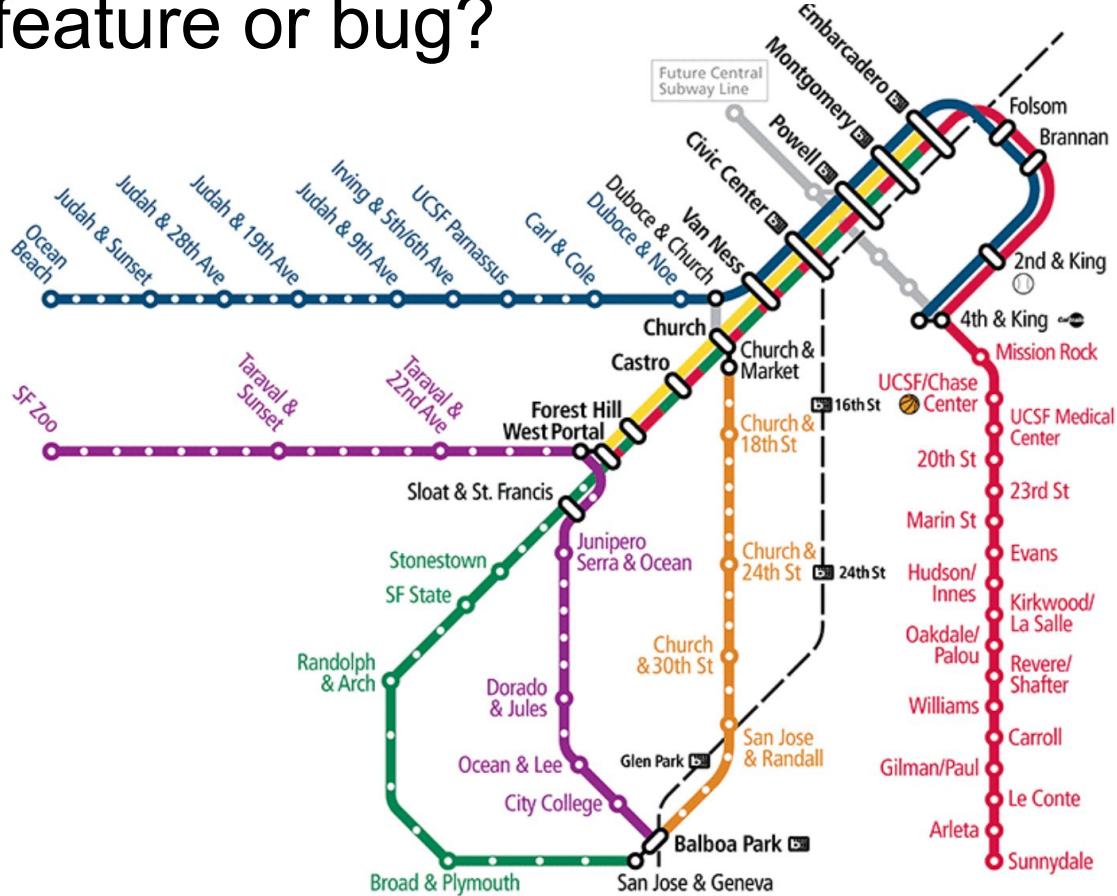
- What phenomena to know about
- What data to try to capture it with
- What aspects to feature
- By what means, with what instruments
- How to present the data to make it actionable

Selection is often better thought of as *perspective* than as *bias*.

# Representation -- feature or bug?

Think of a map ...

**Selection, abstraction, mediation, and perspective** are in the very nature of representation.



# Representation, legitimacy, and authority go together



PEOPLE POWER 'Collaboration is where power is,' said the Sierra Club's Jacob Klein. 'Without the network, the refineries would have won,' seconded Steve Rosenbaum, 350 Bay Area member and a volunteer for the BAAQMD Network.

**People Power: How networking won a vital air pollution fight**

By Janis Hashe Aug 18, 2021

# Representation

= The way in which one thing is made to “stand for” something else.

Who is represented in data? Whose experience is represented?

Representation is two-sided. It deals with both *knowledge* and *power* (authority, legitimacy, “what counts”).

Data is a technology of representation. It creates representations of social and natural phenomena and of people. These representations do work in the world. They take on lives of their own.

# How do we know what to count? What “counts”?

For the Air District:

- What count as pollutants to measure?
- What's not included? What “doesn't count”?
- What counts as a site to be monitored?
- What counts as a flare?



# Transition

What subjects would you need to study  
in order to think through

- How to use the Air District's data?
- How to serve the goal of reducing harm from pollution?

# Why HCE?

Data science is deeply social

Power and ethical action are at stake

Responsible data science involves more than technical training

# HUMAN CONTEXTS & ETHICS TOOLKIT



# The Human Contexts and Ethics Toolkit

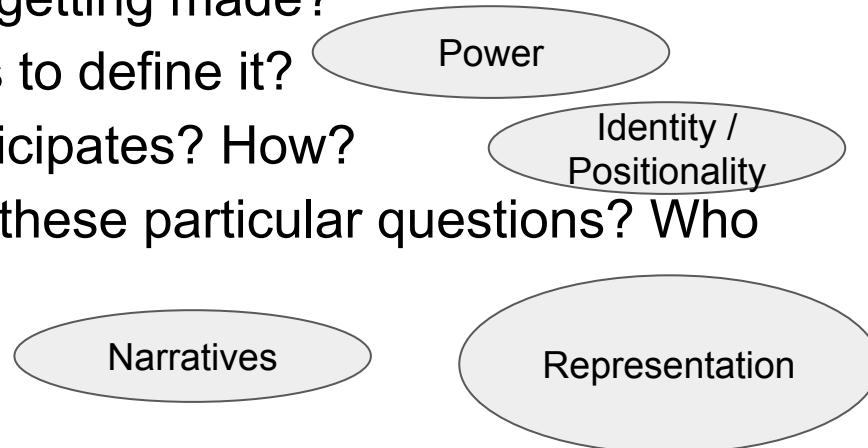
A set of concepts and methods from social science and humanities selected to build your understanding of the datafied world, helping you identify where human power structures and choices get built into technical work, and empowering you to discover how, when, and where you can responsibly and effectively intervene.

Learn more about the toolkit on our [website](#)

HCE Toolkit	
Power Sociotechnical Systems	Narratives Identity/Positionality
Agency Classification Co-production Expertise Institutions	Labor Materiality Performativity Representation Sociotechnical Imaginaries

# Making data at the Air District

- What's "the phenomenon" in the world?
- What work goes into making the data?
- Why is it getting made?
- Who gets to define it?
- Who participates? How?
- Why ask these particular questions? Who cares?



Making data, not just collecting ...



# What is the Data Science Ethos Lifecycle?

The Data Science Ethos Lifecycle is a framework that helps data science practitioners, and students examine the human contexts and ethical consequences of the tools and services they are developing at every stage of their work.

The Ethos Lifecycle is a new data science standards that integrates a toolset comprised of four specific humanistic and social science modes of inquiry into the data science lifecycle workflow.



Check out our recently published article in the Journal of Statistics and Data Science Education - "Data Science Ethos Lifecycle: Interplay of ethical thinking and data science practice"

# Thank you!

Cathryn Carson  
UC Berkeley  
8/4/22



# Additional slides

# What is the HCE Program?

## Dedicated courses

- Data 4AC: Data and Justice
- Data C104: Human Contexts and Ethics of Data

## Integrated course materials

- Data 100 modules

## General curricular materials

- HCE Toolkit
- Data Science Ethos Lifecycle

## DSUS Program Support

## Research and more

- Community-based partnerships (coming soon!)

# Dedicated HCE courses

Data 4AC: Data and Justice

Data 104: Human Contexts and Ethics



# Data 4AC: Data and Justice

Fulfills Berkeley's American Cultures requirement

How can data be involved in pursuit of social justice?

How has the meaning and practice of justice changed over time and in interaction with data?

What kinds of stories can you tell with data?  
Whose stories are they?

For more, see [data4ac.org](http://data4ac.org)



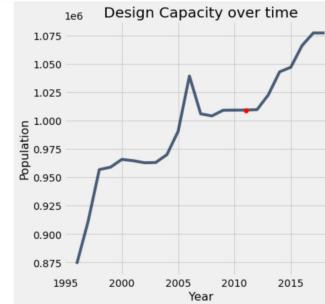
W.E.B. Du Bois - Data visualization from the 1900 Paris Exposition Universelle

# Data 4AC: Data and Justice

Soft immersion in Jupyter notebooks through a series of modules

1. Japanese-American Detainment during World War II
2. Prisons and Jails
3. Algorithmic Fairness in Recidivism Algorithms
4. Data has Boundaries (spatial data and environmental justice)
5. Health Risk Scores

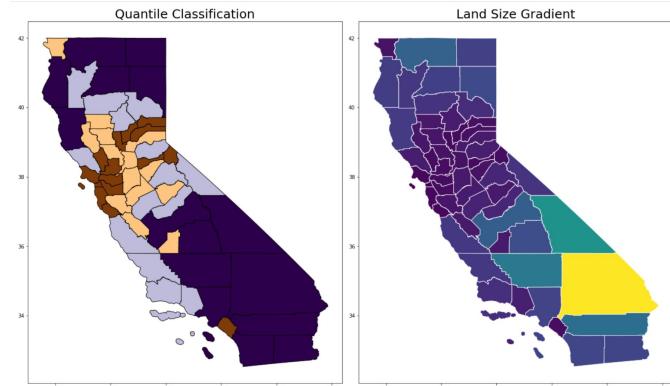
For more, see [data4ac.org](http://data4ac.org)



Question 2.1a: In looking at the graph produced, how does it reflect the systematic changes in California's potential prison population? In particular, name a court case that is related to the red marker in our graph.

Answer: YOUR ANSWER HERE

Question 2.1b: Let's analyze our graph further; what do you think happened in 2006 that caused the spike we see?



# Data 4AC: Data and Justice

Capstone project: Narratives of Justice

Group multimedia project combining stories, data, and visualizations

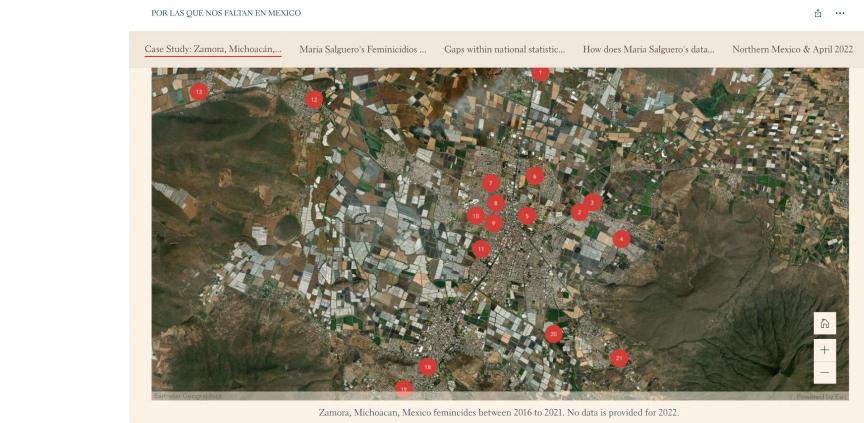
Peer learning

Community engagement

For more, see [data4ac.org](http://data4ac.org)

## Environmental Justice in Richmond, CA Resilience & Identity

How are community members and local leaders defining environmental justice for Richmond?



# Data C104: Human Contexts and Ethics of Data

Fulfills HCE requirement for the DS major and minor

Offered every semester since Sp 2018

Enrolls nearly 1,000 students per year



First lecture of Fall 2019 semester of Data C104

For more, see [data104.org](http://data104.org)

# Data C104: Human Contexts and Ethics of Data

## Learning Activities

Weekly readings and writing reflections

Capstone research project: “HCE vignette”

- Practicing public-facing use of expertise (op-ed, policy paper, memo)
- Multiple drafting stages

For more, see [data104.org](http://data104.org)



First lecture of Fall 2019 semester of Data C104

# Peeling back the layers of the “datafied world”

Unit 1: Our datafied world

Unit 2: Responsible data

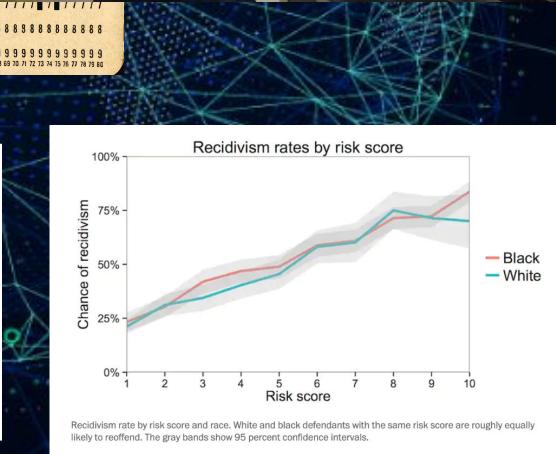
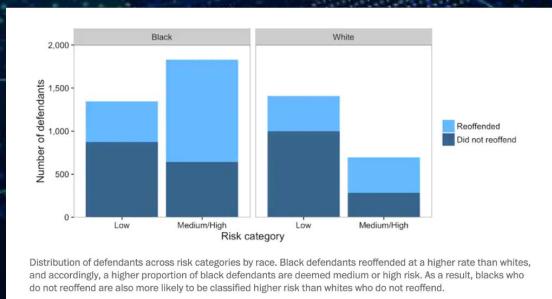
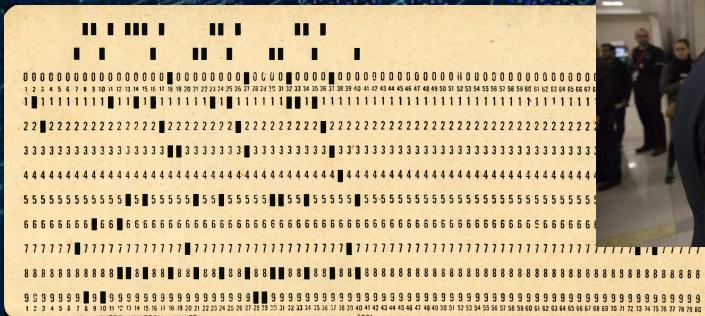
Unit 3: When data is personal

Unit 4: Collective life

Unit 5: Data and democracy

Unit 6: Capitalism, industry, and labor

Unit 7: Foundations of data science practice



Selina Liu (2023)

# Who is the HCE Program?

## Student Team (Spring 2022)



Anna  
Gueorguieva  
(2023)  
*Team Lead*



Sammy Raucher  
(2023)  
*Team Lead*



Carlos Ortiz  
(2022)



Aryana Far  
(2023)



Selina Liu (2023)



Anika Cruz (2021)

Not pictured: Louie Ortiz (2022) and Anu Thirunarayanan (2023)



We'd love to hear from you!

See our website at [data.berkeley.edu/hce](http://data.berkeley.edu/hce) for more information and to access our curricular materials

Contact us at:

[hce@berkeley.edu](mailto:hce@berkeley.edu)

Cathryn Carson - [clcarnson@berkeley.edu](mailto:clcarnson@berkeley.edu)

Ari Edmundson - [aedmundson@berkeley.edu](mailto:aedmundson@berkeley.edu)