

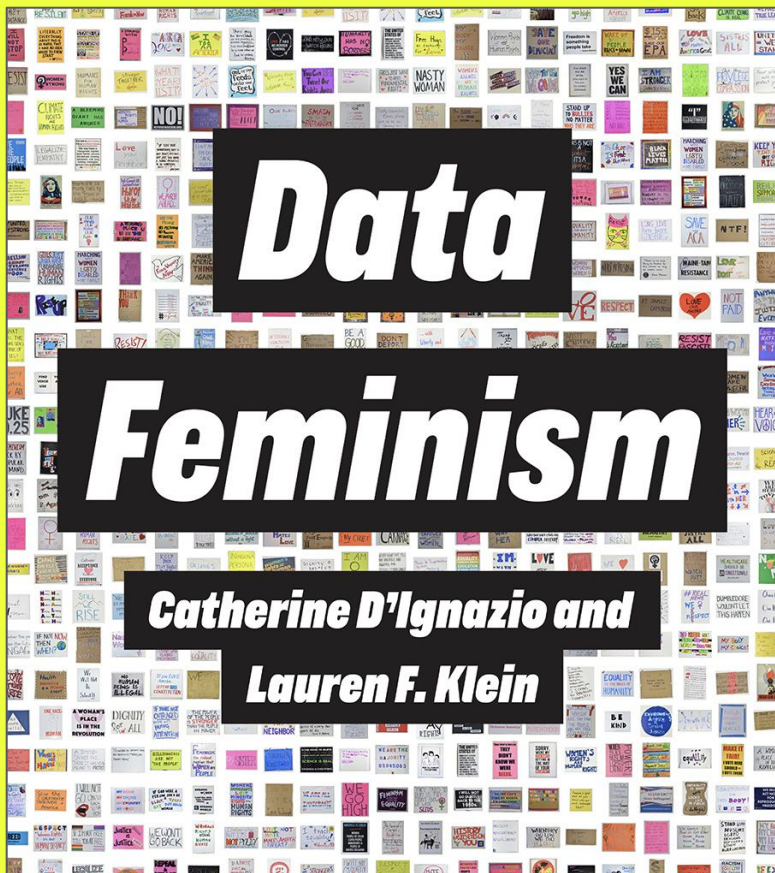
What is feminist data science?

XXII International Meeting on Gender Statistics

October 2021



Catherine D'Ignazio, Director
Data + Feminism Lab
Massachusetts Institute of Technology



Data Feminism is open access at datafeminism.io



Catherine D'Ignazio, Assistant Professor of Urban Science & Planning
Director, Data + Feminism Lab, MIT
[@kanarinka](https://twitter.com/kanarinka)



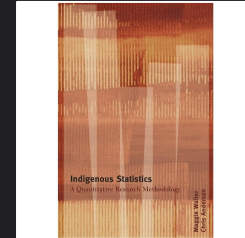
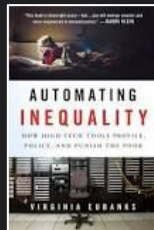
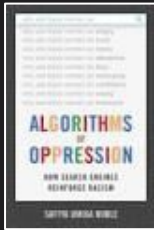
Lauren Klein, Associate Professor of English and Quantitative Theory & Methods
Director, Digital Humanities Lab, Emory University
[@laurenfklein](https://twitter.com/laurenfklein)

“Data is the new oil.”

- The Economist, Intel CEO, Reliance Industries CEO, UAE Minister of Artificial Intelligence, Google execs, etc, mainly elite white men

“Data is the same old oppression.”

- BIWOC, white women, Indigenous people, immigrant communities, LGBTQ+ individuals, + more...



Feminism is...

Belief

A belief in equal rights for all genders.

Political Action

Action to realize that belief in the world.

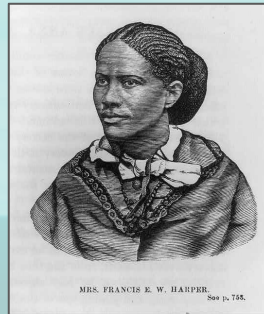
Intellectual Heritage

A set of theories and ideas that help us understand and challenge unequal power.

Intersectional feminism is not only about women and gender.

It is about power - who has it and who doesn't.

Intersectionality is a concept out of Black feminism, which draws on ideas from Frances Harper (1825-1911), the Combahee River Collective (1974-1981) and Kimberlé Crenshaw (1989), among others.





In today's world, data is power.

Intersectional feminism, when applied to data science, can help challenge and change that power.

7 Principles of Data Feminism

- ★ Examine power
- ★ Challenge power
- ★ Rethink binaries and hierarchies
- ★ Elevate emotion and embodiment
- ★ Embrace pluralism
- ★ Consider context
- ★ Make labor visible



An Incomplete List of Missing Data Sets

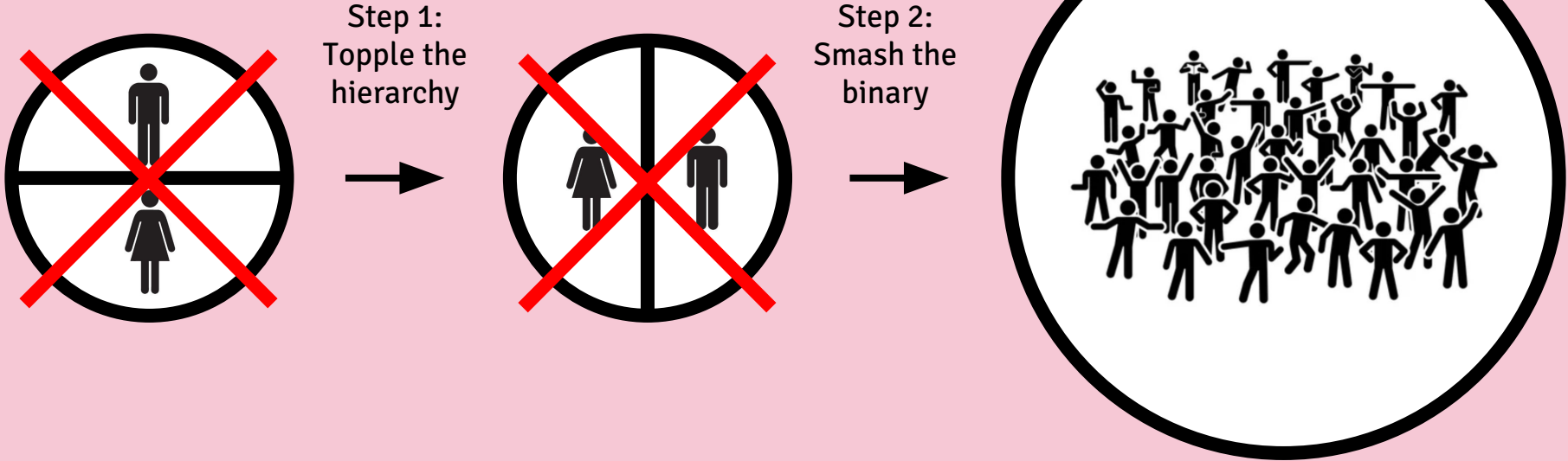
This list will always be incomplete, and is designed to be illustrative rather than comprehensive, from the perspective of the U.S, though the complete list of datasets featured here includes

- Civilians killed in encounters with police or law enforcement agencies [1]
- Sales and prices in the art world (and relationships between artists and galleries)
- People excluded from public housing because of criminal records
- Trans people killed or injured in instances of hate crime (note: existing data is incomplete)
- Poverty and employment statistics that include people who are behind on rent
- Muslim mosques/communities surveilled by the FBI/CIA
- Mobility for older adults with physical disabilities or cognitive impairments
- LGBT older adults discriminated against in housing
- Undocumented immigrants currently incarcerated and/or underpaid
- Undocumented immigrants for whom prosecutorial discretion has been exercised
- Measurements for global web users that take into account shared devices
- Firm statistics on how often police arrest women for making false rape accusations
- Master database that details if/which Americans are registered to vote
- Total number of local and state police departments using stingray phone location tracking
- How much Spotify pays each of its artists per play of song
-

The Library of Missing Datasets by Mimi Onuoha, 2016.

Principle: Examine Power

How to dismantle binary structures of power in two basic steps:

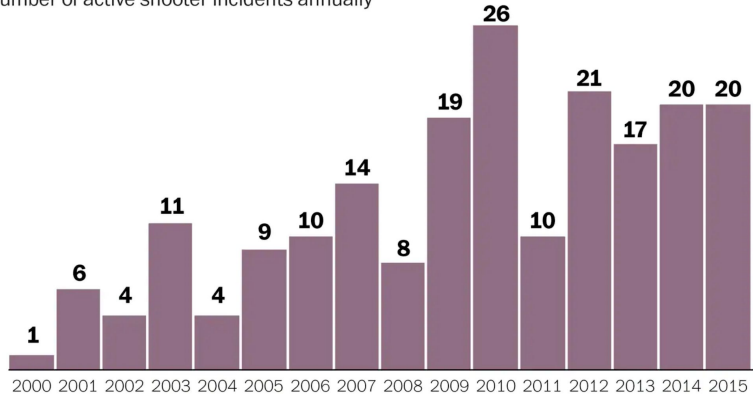


Principle: Rethink Binaries and Hierarchies

Data visualizations have an aura of objectivity

The era of “active shooters”

Number of active shooter incidents annually



WAPO.ST/WONKBLOG

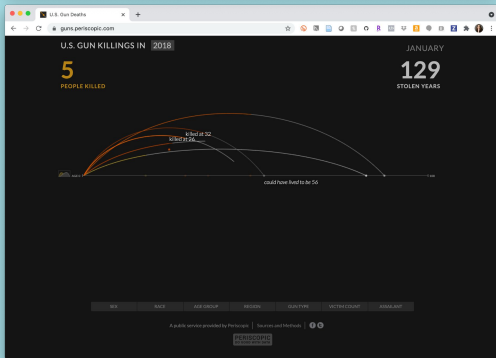
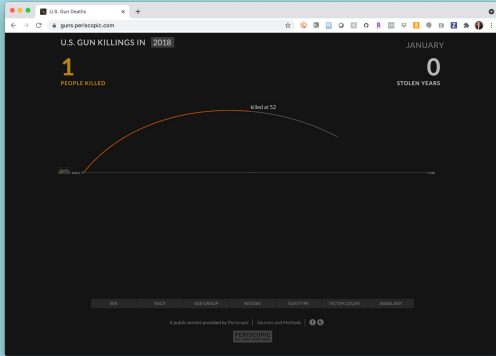
Source: FBI

“We focus on four conventions which imbue visualisations with a sense of objectivity, transparency and facticity. These include: a) two-dimensional viewpoints; b) clean layouts; c) geometric shapes and lines; d) the inclusion of data sources.”

From Helen Kennedy, Rosemary Lucy Hill, Giorgia Aiello, and William Allen. (2016) The work that visualisation conventions do. *Information, Communication and Society*, 19 (6). pp. 715-735. ISSN 1369-118X

Principle: Elevate Emotion and Embodiment

U.S. Gun Deaths, by
Periscope (2013-2018)
<http://guns.periscope.com>



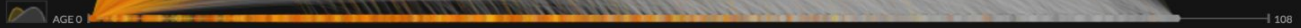
U.S. GUN KILLINGS IN 2018

11,356

PEOPLE KILLED

472,332

STOLEN YEARS



SEX

RACE

AGE GROUP

REGION

GUN TYPE

VICTIM COUNT

ASSAILANT

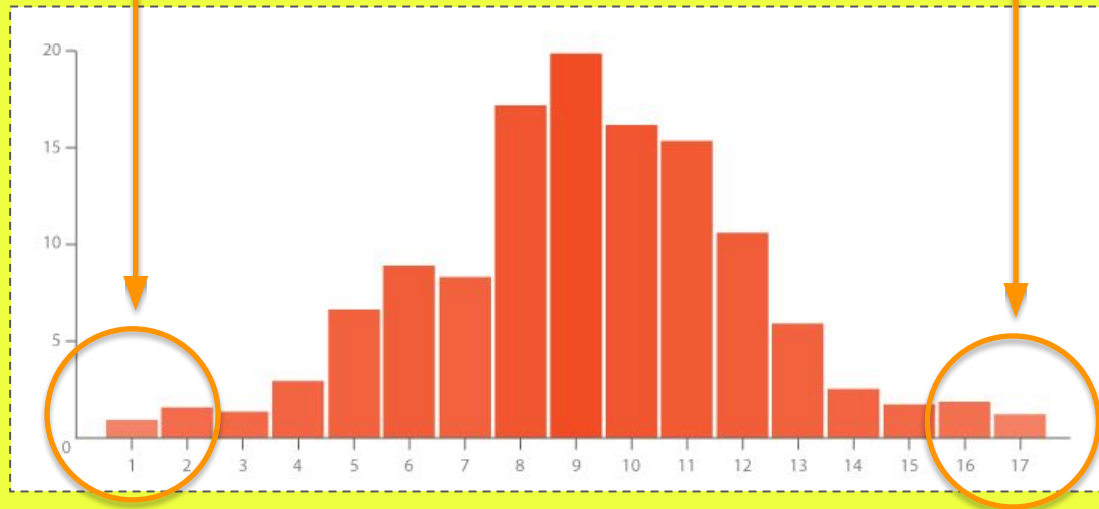
Principle: Elevate Emotion and Embodiment

Participatory process for data science

Feminist design prioritizes the participation of people who have been most marginalized by the system

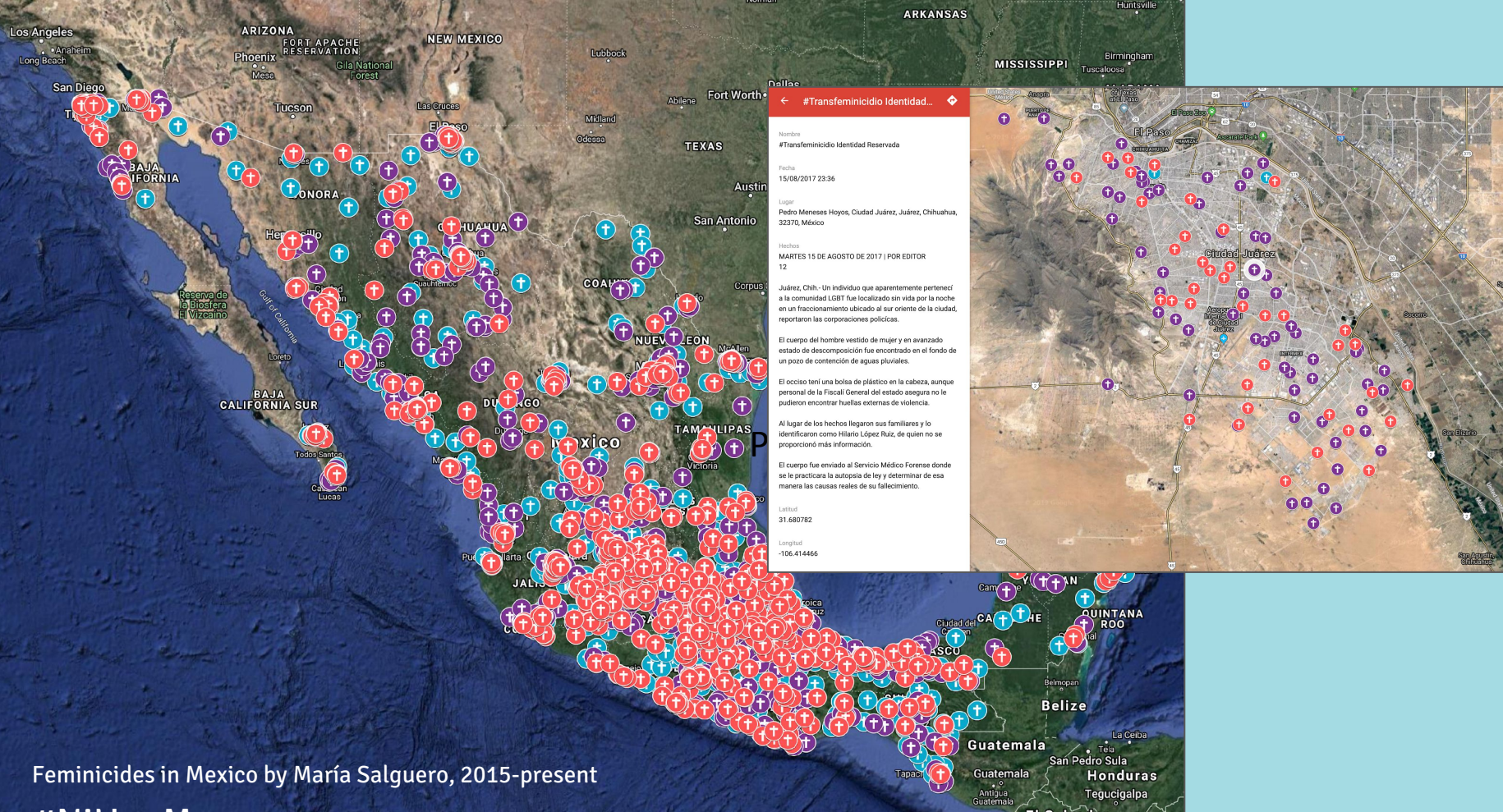
Feminist design

Feminist design



We center the voices of those who are directly impacted by the outcomes of the design process. – Design Justice Network

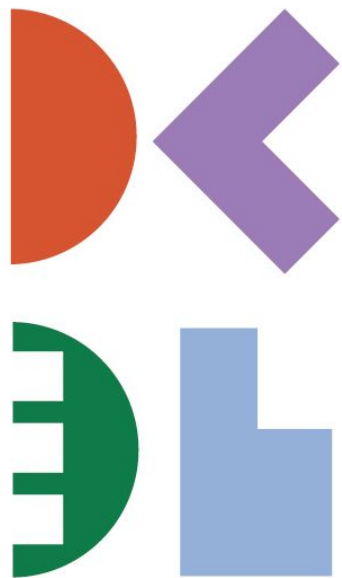
Principle: Embrace Pluralism



Femicides in Mexico by María Salguero, 2015-present

#NiUnaMenos

Principle: Challenge Power



DATOS CONTRA EL FEMINICIDIO

América Latina y el Caribe

Data Against Femicide

Team

Co-Leads

Catherine D'Ignazio, Data + Feminism Lab

Helena Suárez Val, Femicidio Uruguay

Silvana Fumega, Latin American Initiative for Open Data (ILDA)

Partner

Rahul Bhargava, Data Culture Lab/Media Cloud, Northeastern University

Data + Feminism Lab Researchers

Natasha Ansari

Amelia Dogan

Isadora Cruxên

Mariel García-Montes

Alexis Hope

Ángeles Martínez de Cuba

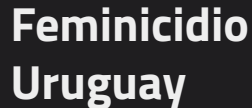
Luciana Ribeiro da Silva

Wonyoung So

Harini Suresh

Melissa Teng

Thuận Tran

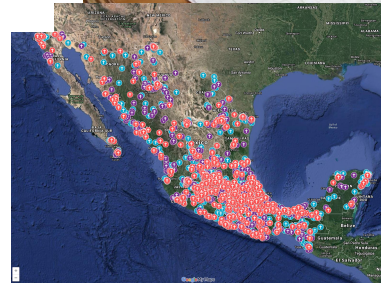


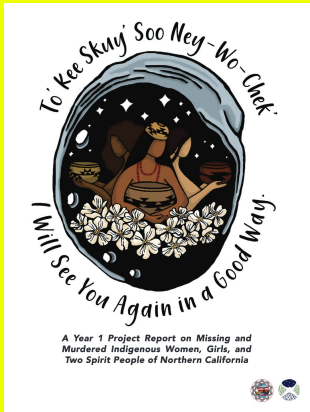


DATA AGAINST FEMINICIDE

Latin America, the Caribbean and beyond

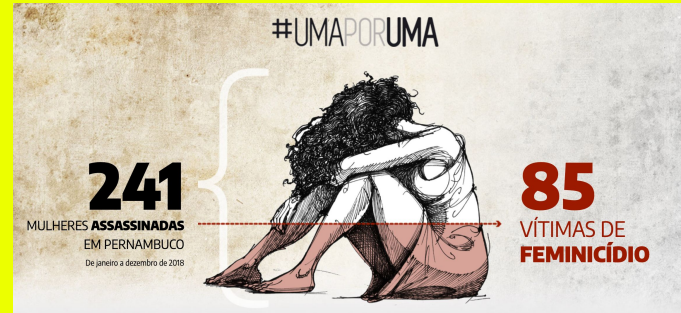
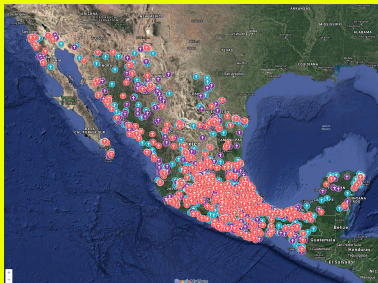
- A **qualitative research project** to understand activist counterdata & information practices
- A **participatory design project** to co-design digital tools to support monitoring groups
- A growing, global **community of practice** for people working on femicide & data





There are many grassroots efforts that collect data about gender-based killings and monitor cases.

We have found over 150 activist & civil society projects from around the world.



Qualitative & Co-design Research

23

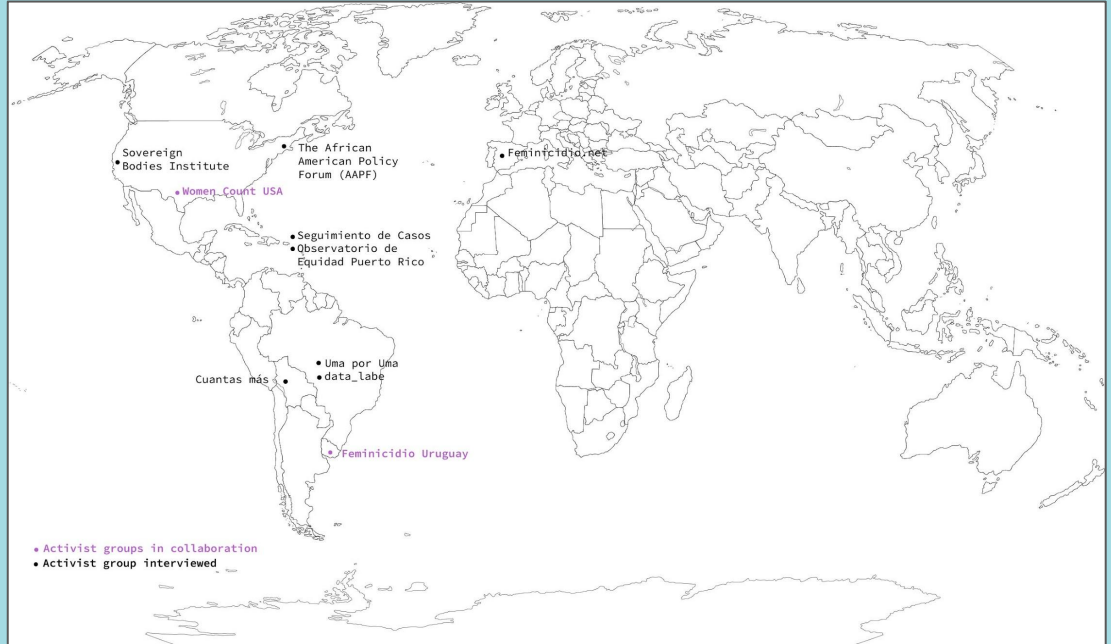
Activist groups

35+

Interviewees

12

Countries



Research Questions

Understanding activist motivations, context, data and information practices.

- Why do activists begin their monitoring projects?
- What sources do they use for information? What software do they use to log femicide and gender-based killings?
- What categories do they use to classify cases in their databases and why?
- What challenges do they face in collecting, analyzing and using the data they produce?

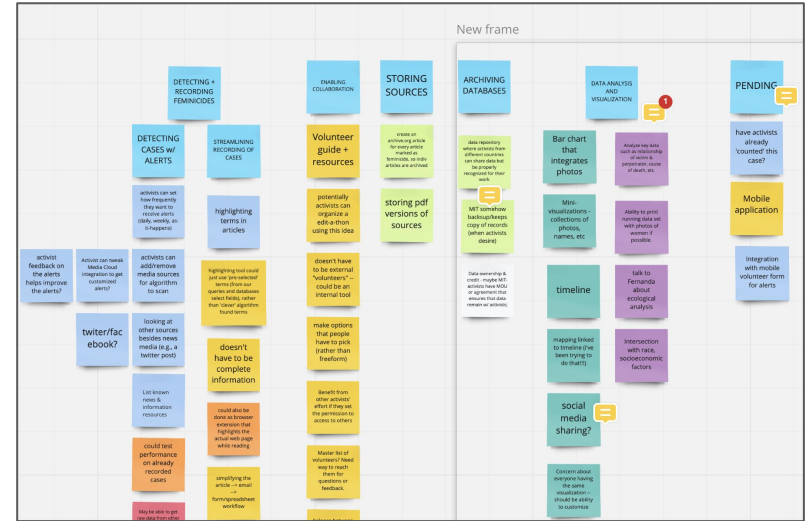
Design Implications that flow from this understanding of activist practices.

- What are ways that interactive tools, machine learning and AI can support and sustain civil society efforts to collect femicide data?
- What are ways that we can build community and connection to sustain this data activism?

Participatory design & interactive tools

Six co-design sessions with Women Count USA and Femicidio Uruguay

Spring pilot with 4 groups in the U.S. and 3 groups in Argentina/Uruguay




Miro board from one of our co-design sessions

Co-designed tool #1

The Data Highlighter

teenVOGUE STYLE POLITICS CULTURE IDENTITY SUMMIT SIGN IN NEWSLETTER Q



GETTY IMAGES

Identity

Femicide Is a Growing Issue in the United States

The killings of Vanessa Guillén, Oluwatoyin Salau, Nina Pop, and many more are part of a global problem.

BY NIDIA BAUTISTA
AUGUST 28, 2020





The killing of 20-year-old Vanessa Guillén, a Mexican-American Army specialist who disappeared from Fort Hood, Texas on April 22, whose

WATCH




DATA AGAINST FEMICIDE DATA HIGHLIGHTER

Open My Database ↗


Highlighting

- names (auto-detected) 
- places (auto-detected) 
- dates (auto-detected) 
- custom 

update words

Co-designed tool #2: Email Alerts



WOMEN COUNT USA: FEMICIDE ACCOUNTABILITY PROJECT

By Women Count USA — Delivering **once a week**

[Edit Setting](#)

News Articles

Search Collected Date From To

| Email to Friend | Title | Publish Date | Collected Date | Count |
|--------------------------|--|--------------|----------------|-------|
| <input type="checkbox"/> | Angela Weisheit's body was found in 2020 in Brown County. Now 2 people have been charged. | 09/02/2021 | 09/03/2021 | 2 |
| <input type="checkbox"/> | Woman once accused of getting Botox without paying now charged with murder in nurse's death | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | DC man accused of killing mother also charged in three cold case slayings, one nearly 30 years old | 09/02/2021 | 09/03/2021 | 3 |
| | WLDY | 09/02/2021 | | |
| | foxbusiness.com | 09/02/2021 | | |
| | Fox News | 09/02/2021 | | |
| <input type="checkbox"/> | Police: Man recently arrested for mother's death connected to multiple cold case homicides in DC | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Sierra Vista woman accused in death of her 2-year-old son | 09/02/2021 | 09/03/2021 | 6 |
| <input type="checkbox"/> | Former Kentucky pilot sentenced to life in prison for 2015 murder of his 3 neighbors | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Houston man wanted in death of missing woman arrested in Oregon, police say | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Man convicted in 2016 murder of ex-girlfriend | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Derby area man charged in uncle's stabbing death | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Appeals court upholds life sentence for Glen Ellyn murder | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Arrests made in investigation of woman found murdered in Brown County | 09/02/2021 | 09/03/2021 | 1 |
| <input type="checkbox"/> | Indianapolis man and woman arrested in Brown County murder investigation | 09/02/2021 | 09/03/2021 | 1 |

Projects

- African American Policy Forum
 - AAPF Test Project
 - femicide
- Data + Feminism Lab
 - AAPF Test Project, US Black Media Only
 - AAPF w/ new model
 - Another Test Project
 - Catherine's Massachusetts Project
 - English lang test
 - espanoltest
 - Femicidios Peru 2021
 - Maril
 - NewTestProject
 - Philadelphia Femicides
 - Proyecto del Demo
 - Rosa Test
 - SBI - LATAM Monitoring
 - SBI Test with Default Query, New Model
 - SBI Test with Medium Query, New Model
 - SBI Test With MMIW Query, New Model
 - SBI Test With MMIW Query, No Model
 - SBI Test with only Indigenous Media
 - SBI Test with US Tribal Names, New Model
 - Spanish Lang Test
 - Test: Model 1 w/ AAPF Query
 - test project33
- Mumala
 - Ataques de Odio Argentina
 - Femicidios CABA

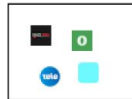
Media Cloud Database



US State & Local



Argentina State & Local



Uruguay State & Local

Media Cloud Story Processor

Each day, the Media Cloud downloads news articles from 50K+ different news sources

Project A Config

Project B Config

Project C Config

English Femicide Classifier

Spanish Femicide Classifier

Spanish Femicide Classifier

List of Articles
....

List of Articles
....

List of Articles
....

Email Alerts - Projects

Fetch the config file to the Media Cloud Story Processor

Send the results to the Email Alert

Project A

((murder OR ...) AND (woman OR...)) AND language:en

Query

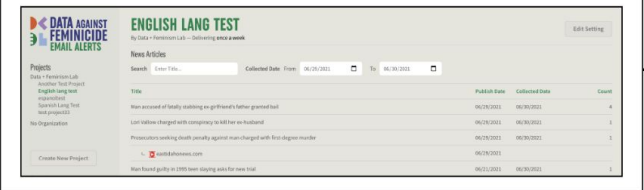
English Femicide Classifier

ML Classifier

/projectA/articles.json POST

Endpoint

List of Articles



Project B

Project C

Email Delivery

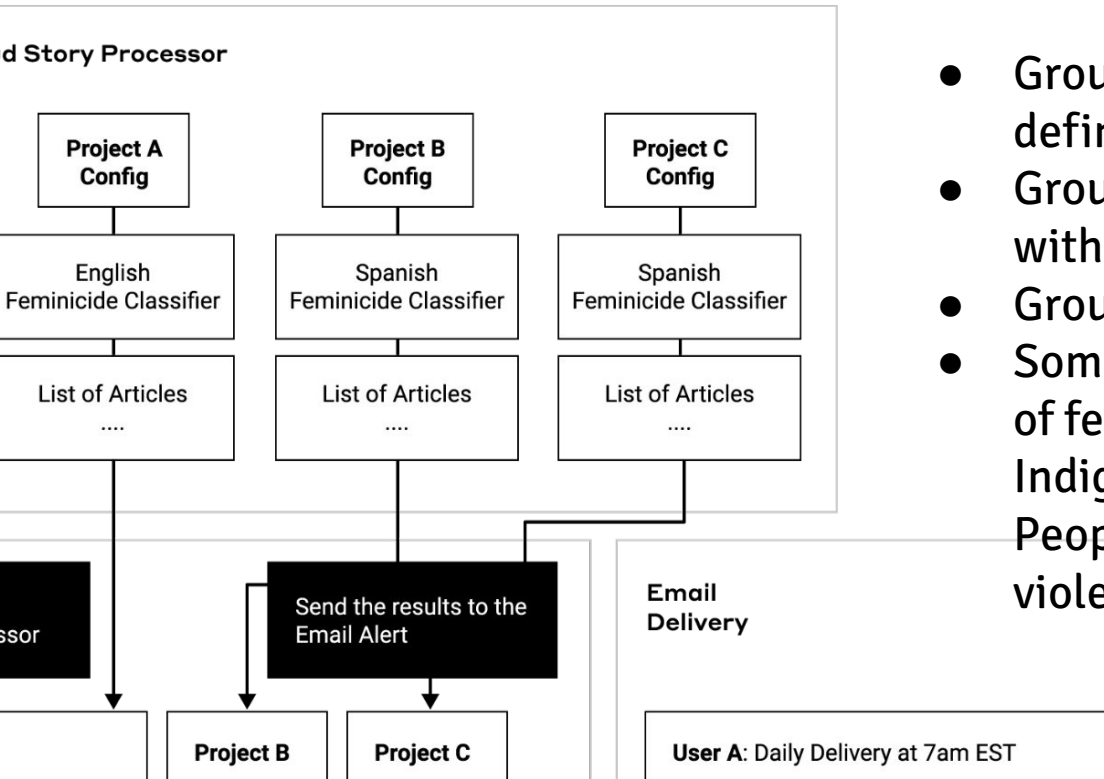
User A: Daily Delivery at 7am EST



User B: Weekly Delivery at 9am EST

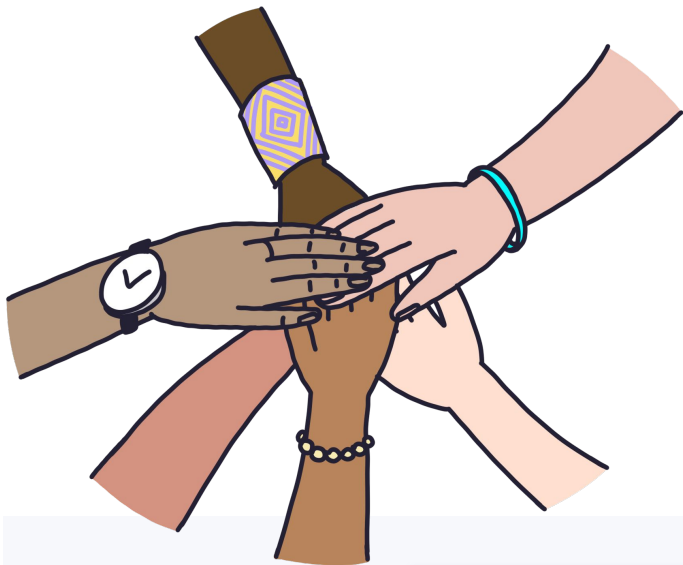
User C: Daily Delivery at 7am EST

Considering Context



- Groups do not all use the same definition of femicide
- Groups monitor different geographies, with different media ecologies
- Groups work in different languages
- Some groups monitor racialized types of femicide like Missing & Murdered Indigenous Women, Girls & Two-Spirit People or Black women killed in police violence

Community of Practice



Workshops

16 to 19 November 2020

An intensive week of workshops and a discussion about the femicide data standardization project

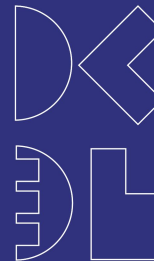
Webinar- Femicide Data Standardization- ILDA/BID



Femicide Data + R



DATA AGAINST FEMICIDE
Technology in Action



Save the date
Data Against Femicide
2nd edition - November 24

Open discussions—tools—workshop

Organized by: Data + Feminism Lab (MIT),
ILDA and Femicidio Uruguay

More details soon



Data feminism in action

Examine power Attend to the ways that feminicide data are missing and neglected by authorities. How the media perpetuate harmful stereotypes and data gaps.

Challenge power Design tools to support counterdata collection so that activists may improve, streamline, and multiply their abilities to use data to challenge powerful institutions.

Elevate emotion and embodiment Attend to ways that interactive technologies may aid storytelling and data visualization to gather public attention about feminicide.

Rethink binaries and hierarchies Push back against all data systems that code gender as a binary. Center the counterdata collection efforts of Indigenous women, Black & Afro-descendent women, low-income women, and/or LGBTQ+ people. De-center the Global North – its tools, technologies and colonizing computing assumptions.

Data feminism in action

Embrace pluralism Design counterdata collection tools in collaboration, coalition and solidarity with the people already doing the work. Use participatory methods.

Consider context Design tools that support variation in how groups define and classify femicide across diverse geographic, media and intersectional contexts. Recognize the deeply grounded expertise activists carry in detecting and verifying femicide cases in their locales.

Make labor visible Design tools to ease the emotional and physical labor of counterdata collection and support existing activist workflows. Recognize and support the care work that is at the center of counterdata collection towards justice. Credit the many students and researchers that have worked on the project.

Data Feminism requires an expanded definition of data science

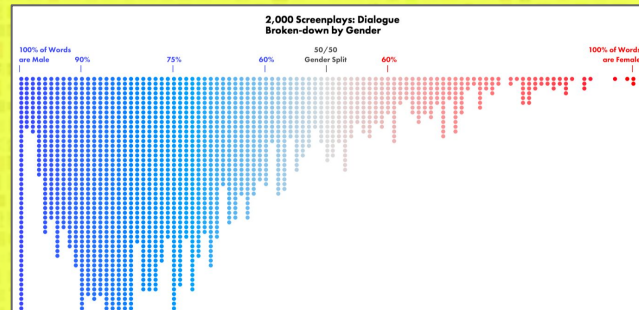
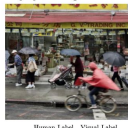
Seeing through the Human Reporting Bias: Visual Classifiers from Noisy Human-Centric Labels

Ishan Misra¹ * C. Lawrence Zitnick³ Margaret Mitchell² Ross Girshick³
¹Carnegie Mellon University ²Microsoft Research ³Facebook AI Research

Abstract

When human annotators are given a choice about what to label in an image, they apply their own subjective judgments on what to ignore and what to mention. We refer to these noisy "human-centric" annotations as exhibiting human reporting bias. Examples of such annotations include image tags and keywords found on photo sharing sites, or in datasets containing image captions. In this paper, we use these noisy annotations for learning visually correct image classifiers. Such annotations do not use consistent vocabulary, and miss a significant amount of the information present in an image; however, we demonstrate that the noise in these annotations exhibits structure and can be modeled

(a) A woman standing next to a bicycle with basket.
 (b) A city street filled with lots of people walking in the rain.



cs.CV | 12 Apr 2016







Download Data Feminism infographics in 5 languages at:
datafeminism.io


Thank you!

Catherine

kanarinka.com

dataplusfeminism.mit.edu


@kanarinka  


@kanarinkaprojects 

Lauren

lklein.com

dhlabs.lmc.gatech.edu

@laurenfklein 

laurenfklein / EmoryDHLab 

@EmoryDHLab 