Intro to arrest data analysis
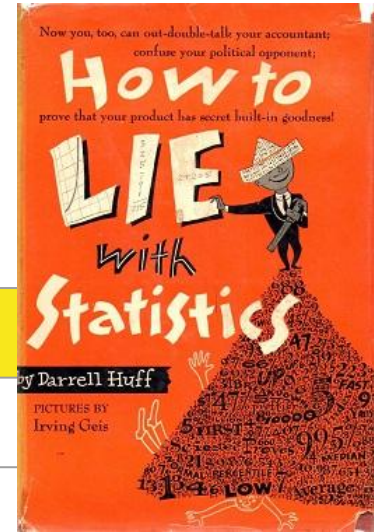
Josie Oakley
SpokesHub, February 2025

What is data "good for" in advocacy?

What is data good for?

— — —

Good For	Not so "good"???
"how many?" / "how often?" / "what?"	"why"
support truths	twist truths
may lend credibility	easy to manipulate/abuse for credibility "3 out of 4 experts recommend…!"
summarizing	storytelling
describing a population	describing an individual or a culture
can attract new audiences	can alienate people with lived experience

Now you, too, can out-double-talk your accountant; confuse your political opponent; prove that your product has secret built-in goodness!

How to LIE with Statistics

by Darrell Huff

PICTURES BY Irving Geis

Honoring data & people

———

- Intentional language ("person" vs "data point")
- Respect the privacy of individuals
- Practices for staying humanized
  - Imagine the individual person behind a row of a spreadsheet
  - Give gratitude for access to data, for what people sacrificed
  - Consider what's missing, what's incorrect
- If you can, share data with people actually in the dataset

Mimi Onuoha's "Library of Missing Datasets"

Mimi Onuoha's "Library of Missing Datasets V2"

Mimi Onuoha's "Library of Missing Datasets V3" – PRIVATE data

"better for those featured in them to protect"

What is arrest data "good for"?

___

- Inside the movement
  - Should we focus on decrim? Arrest support?
  - Where in the city should we focus our outreach?
  - Keeping workers informed of trends
  - How many people are missing our support?
- Outside the movement
  - Advocating for decrim
  - Advocating for grants, city resources
  - Holding police/city accountable
  - Making the old arguments in a new way

About Josie

___

- Former academic researcher
- Former techie
- Privacy expert
- Started in SW advocacy ~2018
  - "spreadsheet bitch"
- Pittsburgher
- gardener, dancer
- she/her

What we'll cover today

———

- What is data "good for"?
- Get to know arrest data
- Finding data
- Case Study: Baltimore
- Case Study: Your city? Atlanta?
- What can you do with arrest data?
- Hands-on Options
  - Option A: Dig in to a spreadsheet together
  - Option B: Hunt for data in your region together
  - Option C: Figure out how to identify stings together
- Summary

WHY DID YOU COME TODAY?
IS THERE ANYTHING SPECIFIC YOU WANT TO LEARN?

Get to know arrest data

Arrest data - RAW - let's look at a spreadsheet!
___

- Open "Pittsburgh arrest data sample - RAW" tab
- What do the columns mean?
- What do you notice?
- What's missing?

- What are some questions we could ask this data?

Arrest data - Commonly included
———

- ID numbers
- Offense
- Date
- Age
- Location or neighborhood

Commonly included

- "Severity" rating
- List of offenses
- Sex
- Race
- Address or coordinates

Sometimes

- Time spent in jail
- Names of people
- Evidence, convictions, penalties, fines, jailtime
- Context, story, truth

Not in arrest data

- UCR — Uniform Crime Reporting
- NIBRS — National Incident-Based Reporting System

Arrest data - Caveats

———

🔔**EVERY POLICE DEPT IS DIFFERENT**

- ID numbers
- Offense ─────────────────→ city, state, or fed laws?
- Date
- Age
- Location or neighborhood ─→ arrest or incident?
- "Severity" rating
- List of offenses
- Sex ─┐
       ├──────────────────→ who is deciding? / what if trans?
- Race ┘
- Address or coordinates ──┐
- Time spent in jail   └→ usually approx. sometimes
- Names of people   removed for victim safety
- Evidence, convictions, penalties, fines, jailtime
- Context

Formulating questions

———

- What are some questions we could ask this data?
    - Are arrests for prostitution increasing or decreasing?
    - What race/gender/age are arrested most commonly?
    - What neighborhoods have the most arrests?
    - How many people are arrested for prostitution vs patronizing?
- What are some questions this data cannot respond to?
    - How many trans people were arrested?
    - Who was arrested indoors vs on the street?
    - Was the situation coercive or violent for the provider?
    - Who snitched?

Advanced questions

———

- How often are multiple workers arrested on the same day?
- Are arrests seasonal? Do they match with sporting events?
- Are Black workers arrested disproportionately?

Might require advanced techniques or mixing datasets

What we've done in Pittsburgh

471 people were arrested for prostitution in Pittsburgh, 2016-2022

In total, there were 695 arrests related to prostitution or sex work.

FOR SAFER SEX WORK PITTSBURGH

Only 3% of arrests relate to violence or trafficking

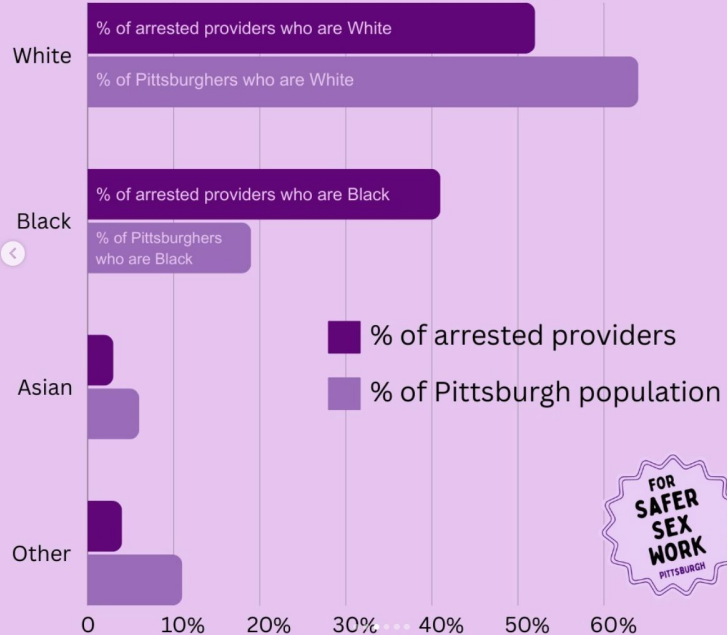| 10 arrests EXPLOITING MINORS | 1 arrest TRAFFICKING | 22 arrests VIOLENCE OR COERCION |

out of **695 arrests** related to prostitution

Takeaway: We strongly support the use of funding to combat violent scenarios in the sex industry through trauma-informed interventions. Unfortunately, only a tiny number of arrests (3%) include any charges related to trafficking or violence, suggesting that the current system of policing isn't reaching people in actual need.

FOR SAFER SEX WORK PITTSBURGH

What we've done in Pittsburgh

What we've done in Pittsburgh

_ _ _

What we've done in Pittsburgh - BREAKDOWN

———

1. Make a list of all the charges related to sex work ("sex industry" charges)
2. Count total sex industry charges
3. Count total prostitution charges

471 people were arrested for prostitution in Pittsburgh, 2016-2022

In total, there were 695 arrests related to prostitution or sex work.

FOR SAFER SEX WORK PITTSBURGH

What we've done in Pittsburgh - BREAKDOWN
———

1. Make a list of any charges related to violence or coercion
2. Make a list related to minors
3. Make a list related to trafficking
4. Count how many sex industry charges are *also* charges involving minors, etc

This is the most useful stat we have. It shows that arrests usually don't involve safety.

Only 3% of arrests relate to violence or trafficking

10 arrests	1 arrest	22 arrests
EXPLOITING MINORS	TRAFFICKING	VIOLENCE OR COERCION

out of **695 arrests** related to prostitution

FOR SAFER SEX WORK PITTSBURGH

Takeaway: We strongly support the use of funding to combat violent scenarios in the sex industry through trauma-informed interventions. Unfortunately, only a tiny number of arrests (3%) include any charges related to trafficking or violence, suggesting that the current system of policing isn't reaching people in actual need.
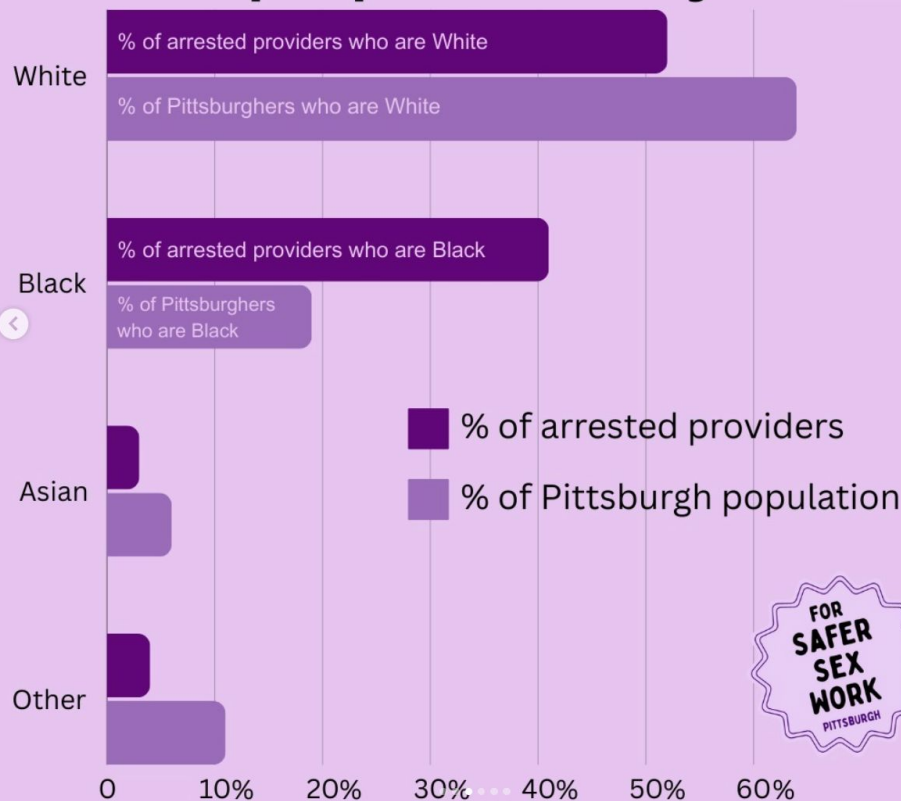
What we've done in Pittsburgh - BREAKDOWN

— — —

1. Find census data for Pittsburgh that includes race
2. Match census race categories to police race categories
3. Make a graph that shows census data and prostitution arrests by race

How do we teach social justice orgs to care about whores? Demographics stats are useful for drawing in ally orgs focused on racial & gender justice.
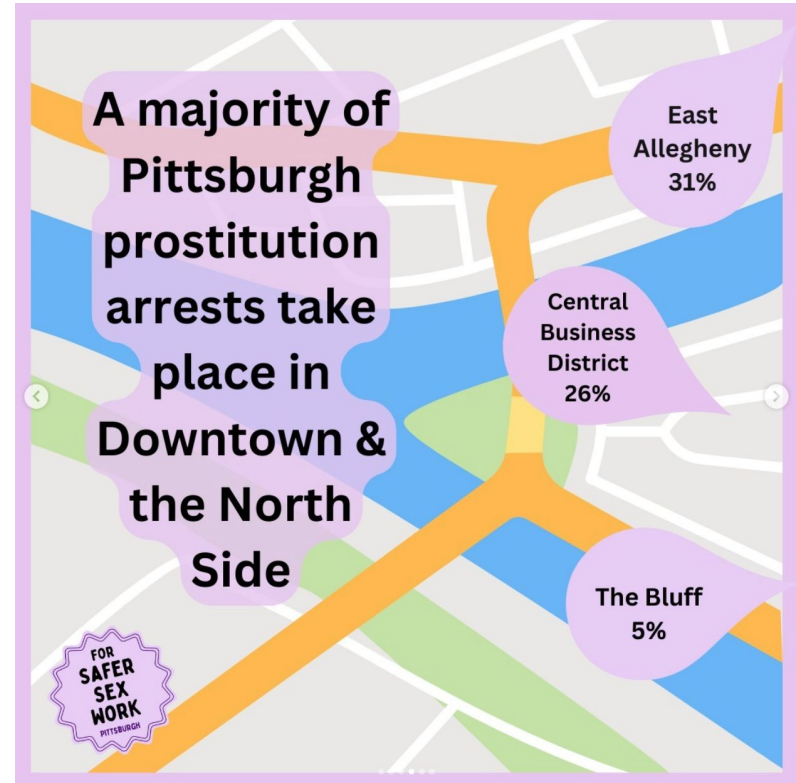
What we've done in Pittsburgh - BREAKDOWN

———

1. Make a graph that shows arrest counts by neighborhood
2. Turn it into a percentage

People (especially politicians) always want to know about their district.

We also looked closer at the most common locations to try and discern encampments, hotels, parlors, etc.

What we've done in Pittsburgh - BREAKDOWN

———

1. Filter arrests for charges with both "prostitution" and "instrument of crime"

Pittsburgh promised to stop using condoms as instruments of crime. It's not clear that they've actually held to that promise.

Useful for drawing in public health allies.

Formulating questions

- What other questions would you like to answer?

- Other questions we worked on
  - How often are providers arrested vs third parties vs johns?
  - How often are prostitution arrests also drug arrests?
  - What locations have multiple arrests on a single day? (possible round-ups)
  - How many charges are usually placed when someone is arrested for prostitution?
  - What are demographics of johns arrested?
  - Frequency of prostitution arrests compared to other better-resourced topics (e.g., domestic violence) – to advocate for fundings/priority

Finding your data

the
hardest
part!

Finding the basics - FBI Crime Data Explorer

— — —

Finding the basics - FBI Crime Data Explorer

– – –

- 👇Often missing data (participation is voluntary for police depts)
- 👇Doesn't include specific location data
- 👍Has basic demographics (sometimes)
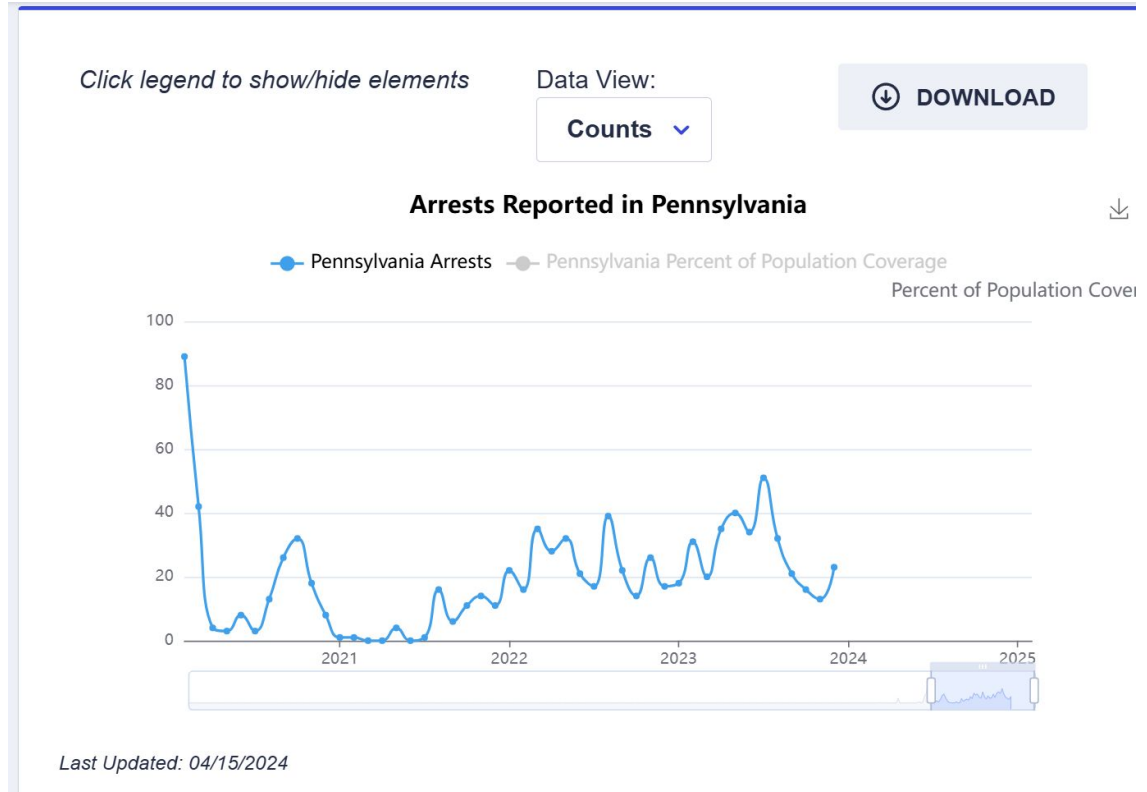
Click legend to show/hide elements

Data View:

Counts ⌄

⊕ DOWNLOAD

Arrests Reported in Pennsylvania

● Pennsylvania Arrests   ● Pennsylvania Percent of Population Coverage

Percent of Population Cover

Last Updated: 04/15/2024

Finding the basics - tips

———

- Google keywords
  - "crime map Pittsburgh"
  - "arrest data Pittsburgh"
- Consider jurisdiction (city, county, boroughs)
  - in Pittsburgh metro, there are 35+ police departments
  - larger cities have Statistic departments

Case Study: Baltimore

Case Study: Your city?

Finding data - advanced

———

- Ask local orgs for help
  - prison abolitionists, court watch, bail fund, harm reduction, domestic violence
- I swear every city has an "open data" guy somewhere
  - usually he's into public transit
- If crime maps exist, it means that the data has been collected *somewhere* by *someone*
  - You can ask them for the data
  - You could maybe steal it
  - You might be able to legally request the data from PD

Finding data - advanced

———

- using an API
- scraping data
  - find a tech friend or student

- Sending Right to Know requests
  - DIFFERENT in each state/city
  - AKA: FOIA, Open Records, Sunshine Law
  - Be incredibly specific
  - Patience, persistence
  - A lawyer can be helpful
  - I have low success with RTKs in Pennsylvania

OOR
Office of Open Records

RIGHT-TO-KNOW LAW ⌄   APPE

HOW TO FILE A REQUEST

The following guide is designed to help you file an effective Right-to-Know (RTK) request.

1. Determine Which Agency to Ask

The first step to filing an effective RTK request is to determine which agency is most likely to have the records you records about the cost of snow removal in your local municipality, you should contact that municipality. If you're system, the best agency would likely be the Department of Conservation and Natural Resources.

This step is absolutely vital. Asking the wrong agency for records will cost you time.

Because misdirected requests are so common, the OOR has produced a table of Commonly Misdirected Request commonly requested records. Please review this table before submitting a request.

2. Find the Agency Open Records Officer

Every agency, whether local or Commonwealth, is required by the Right-to-Know Law (RTKL) to appoint an Agency person as a Right-to-Know Officer, but the function is the same. This is the person you should address your requ

Often, the best way to find the AORO's contact information is on the agency's website. If you can't find it there, o Open Records' (OOR) database of AOROs. If you still can't find the AORO's contact information, simply address it main address.

3. Prepare and Submit the Request

Some agencies have their own RTK Request Forms. Whenever possible, it's best to use the agency's form. But yo is available here in PDF and Word versions:

You can file a RTK request in four ways:
- Email
- Fax

What we'll cover today

———

- What is data good for?
- Get to know arrest data
- Finding data
- Case Study: Baltimore
- Case Study: Your city? Atlanta?
- What can you do with arrest data?
- Hands-on Options
    - Option A: Dig in to a spreadsheet together
    - Option B: Hunt for data in your region together
    - Option C: Figure out how to identify stings together

WHY DID YOU COME TODAY?
IS THERE ANYTHING SPECIFIC YOU
WANT TO LEARN?

Option A: Intro Data Analysis

Intro to data analysis

---

1. Define your questions
2. Gather data (the *hardest* part)
3. Clean data (the *longest* part)
4. Explore & befriend the data
5. Visualize
6. Statistics
7. Communicate your findings

Types of data
___

- Numbers (whole, fraction, percentage)
- Words (city name, arresting officer name, strings)
- Categories (e.g., colors, arrest code)
- Time (year, date, hour, seconds since 1970)
- True/False (yes or no, boolean)
- Location (address, GPS coordinate)
- Void/Null (missing data)
- Lists (of any of the types above)

Sample Data!

I downloaded all available Pittsburgh arrest data that involved prostitution of trafficking.

1. Open "pittsburgh_sex-industry-arrests-2016-2022-cleaned" tab.
2. Make a copy (File > Make a Copy) or download

- Tell the spreadsheet what kind of data we have (Format > Text)
- The "pittsburgh_column_definitions" tab might be helpful

Spreadsheet Challenges!

___

➔ Basic: Count how many providers were arrested

➔ Basic: Make a pie chart showing the sex of people arrested

➔ Medium: Make a line graph plotting client arrests by year

➔ Hard: Make a pivot table that shows arrests by gender for providers vs third-parties vs clients

➔ Make You Cry: What is the average number of charges when someone is arrested?

Option B: Hunt for your data together

Finding the basics - tips

———

- Google keywords
  - "crime map Pittsburgh"
  - "arrest data Pittsburgh"
- Consider jurisdiction (city, county, boroughs)
  - in Pittsburgh metro, there are 35+ police departments
  - larger cities have Statistic departments
- Ask local orgs for help
  - prison abolitionists, court watch, bail fund, harm reduction, domestic violence
- Many cities have a token "open data" guy somewhere
  - usually he's into public transit
- If crime maps exist, it means that the data has been collected *somewhere* by *someone*
  - You can ask them for the data
  - You could maybe steal it
  - You might be able to legally request the data from PD

Option C: Figure out how to identify stings together

How would you use arrest data to identify possible prostitution stings?

How do we need to format the data?

-Start by looking at arrest dates (sort by date)

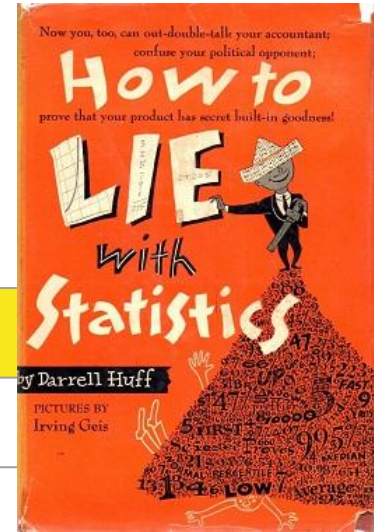-Then look at locations - did the arrests happen in one place?

Summary

tips for presenting data persuasively

———

- Always give audience the takeaway from a graph
- Combine a stat with storytelling
- For credibility, memorize just 3 numbers!
- Share findings in the "debunk" format
- "We don't have the data for that, but our experience suggests…"

REVIEW: What is data good for?

– – –

Good For	Not so "good"???
"how many?" / "how often?" / "what?"	"why"
support truths	twist truths
may lend credibility	easy to manipulate/abuse for credibility "3 out of 4 experts recommend…!"
summarizing	storytelling
describing a population	describing an individual or a culture
can attract new audiences	can alienate people with lived experience

REVIEW: Treat datasets like tarot

REVIEW: Arrest data can only answer some questions

___

- What are some questions we could ask this data?
  - Are arrests for prostitution increasing or decreasing?
  - What race/gender/age are arrested most commonly?
  - What neighborhoods have the most arrests?
  - How many people are arrested for prostitution vs patronizing?
- What are some questions this data cannot respond to?
  - How many trans people were arrested?
  - Who was arrested indoors vs on the street?
  - Was the situation coercive or violent for the provider?
  - Who snitched?

REVIEW: Finding your data

___

- This is often the hardest part!
- Start with FBI Crime Explorer
- Google keywords
  - "crime map Pittsburgh"
  - "arrest data Pittsburgh"
- Look for local friends who already know