

Lab 08

Managing Data

Mid-semester check-in comments

- Most appreciated aspects of the class: **Interactive & Collaborative Learning** (labs, Friday lectures, collaborative work)
- You would like more: walkthroughs, code demos; more time to ask questions and practice but also labs to be shorter. Go through lab answers. More group work and centralized resources (check the course's website).

Data Dictionaries

Police Transparency Hub - Data Portal:

<https://bpd-transparency-initiative-berkeleypd.hub.arcgis.com/search>

Data Dictionaries

In this lab you will have to create data dictionaries (*aka* codebooks) for the two data sets involved in Project 2

- **Berkeley PD Stop Data (2024 - Current)**
- **Collision Data (2011 to Current)**

**These codebooks are part of your Project 2 submission; the sooner you complete this task, the earlier you can begin your reproducibility analysis.*

Stop Data

About Stop Data

Conduct a web search to know more about “**Stop Data**” from police departments.

- What is Stop Data?
- What information is typically recorded during police stops?

These questions will help you understand what you are working with and in any case, eventually write your report.

About Stop Data: Example

Stop data comprises records collected by law enforcement officers regarding discretionary and dispatched pedestrian or vehicle stops. It tracks the details of police-community interactions to promote constitutional policing, transparency, and accountability.

What is Collected?

When an officer stops or detains someone (such as during a traffic stop or street check), they record specific information about the interaction, including:

- **Time and Location:** Date, time, duration, and geographic location of the stop.
- **Perceived Demographics:** The officer's perception of the person's race or ethnicity, age, gender identity, LGB+ status, limited English fluency, and disability status.
- **Reason for the Stop:** Whether the contact was a call for service (e.g., a 911 dispatch) or officer-initiated (e.g., a traffic violation or suspected criminal activity).
- **Actions Taken:** Any steps the officer took, such as ordering someone out of a vehicle, placing them in handcuffs, searching the person or property, or seizing evidence/contraband.
- **Outcome:** The result of the encounter, such as a verbal warning, a traffic citation, a field interview card, or a custodial arrest.

Berkeley PD Stop Data (2024 - Current)

- Download CSV file from BPD's Data Portal
- Extract the name of the columns
- In a text file or document, create your data dictionary
- Include, *at least*, the following:
 - Name of variable (as it appears in the CSV file; this is usually an abbreviation)
 - Full name (longer or complete name of variable)
 - Description of variable (what the variable represents)
 - Recommended data type (boolean, integer, float, string, date/time, etc)
 - How missing values are encoded (if applicable)
 - Details or comments (e.g. what numbers mean if a categorical variable is encoded numerically)
- You may need to do a web/AI search to know more of a certain variable

Mini example: First variable in Stop Data

- **Column:** LEA_Record_ID
- **Name:** Law Enforcement Agency Record Identifier
- **Description:** Unique internal tracking number or report identifier generated by the local agency for that specific police stop
- **Data type:** string or character
- **Details:** None

Collision Data

Collision Data (2011 - Current)

- Download CSV file from BPD's Data Portal
- Extract the name of the columns
- In a text file or document, create your data dictionary
- Include, *at least*, the following:
 - Name of variable (as it appears in the CSV file; this is usually an abbreviation)
 - Full name (longer or complete name of variable)
 - Description of variable (what the variable represents)
 - Recommended data type (boolean, integer, float, string, date/time, etc)
 - How missing values are encoded (if applicable)
 - Details or comments (e.g. what numbers mean if a categorical variable is encoded numerically)
- You may need to do a web/AI search to know more of a certain variable

Each team member should do their first: (clone) -> commit -> push workflow to test their access and as a way of acknowledging they are both aware of the Project.

(Reading) Checkpoint #7

You'll return it until the end of class

Start your Project 2!