Open Data and Shared Workflows for Criminal Legal Research: A Texera Case Study

Faculty leads: Naomi F. Sugie (UCLA), Kristin Turney (UC Irvine)

Graduate students: Rachel Bauman (UC Irvine), Stephanie Zhang (UCLA)

Support from National Science Foundation Integrated Data Systems and Services award (2609582)

Challenges in criminology research

- Difficult to reproduce existing scientific knowledge
- Data fragmentation prevents new scientific knowledge

Replication and reproduction of existing knowledge

ANNUAL REVIEWS

Annual Review of Sociology

Reproducibility in the Social Sciences

James W. Moody,[1,2] Lisa A. Keister,[1,2,3] and Maria C. Ramos[4]

[1]Department of Sociology, Duke University, Durham, North Carolina, USA; email: jmoody77@duke.edu

[2]Duke Network Analysis Center, Duke University, Durham, North Carolina, USA

[3]Sanford School of Public Policy, Duke University, Durham, North Carolina, USA

[4]Interdisciplinary Social Science Program, Florida State University, Tallahassee, Florida, USA

Harvard Data Science Review • Issue 5.4, Fall 2023

Crisis? What Crisis? Sociology's Slow Progress Toward Scientific Transparency

Kim A. Weeden[1]

[1]Department of Sociology, Cornell University, Ithaca, New York, United States of America

Data fragmentation and lack of access prevents new knowledge

- Data are project and PI-based

- Data are not used beyond the project

How Texera can address these issues

- Open access to data and analytic pipelines
  - Addresses data fragmentation, accessibility, and transparency
  - Facilitates replication

- Integration across projects
  - Generate new research insights
  - Broader scope to advance policy-relevant knowledge

How and why would a social science researcher use Texera?

- Interviews with social science researchers (n=20)

- Focus on 5 primary aims of Texera
  - Archiving data and analytic pipelines to support replication
  - Harmonizing data across projects to enable integration
  - Using an AI chatbot interface to assist in producing analysis
  - Providing a space to share machine learning models and other resources
  - Running as a high-performance computing resource for data-intensive projects

Case study: deaths in custody

Context

- Deaths in custody: deaths that occur in prisons and jails (as well as at arrest, by law enforcement)

- Need: federal agencies are not collecting and disseminating in a timely manner despite a congressional mandate (Death in Custody Reporting Act, 2013)

- Consequently: lack of knowledge about deaths in custody and how they might be prevented in the future.

CRIMINAL JUSTICE

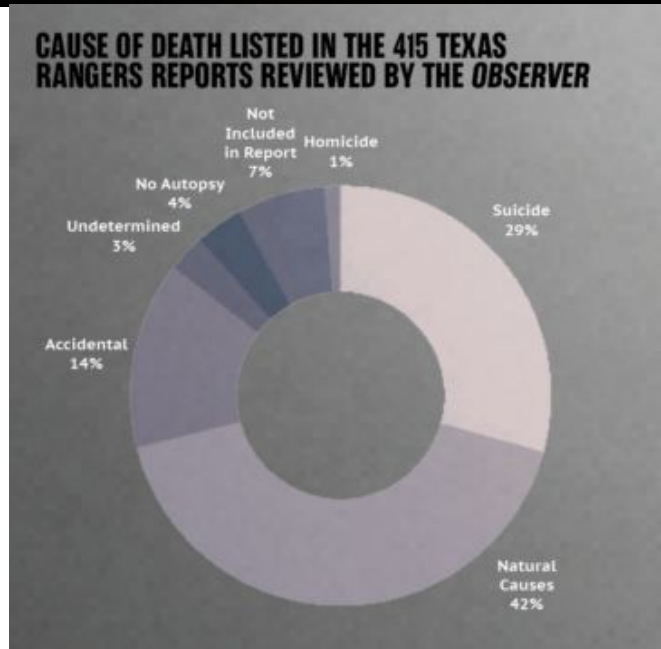
LOCKED UP AND LEFT TO DIE

In Texas, dying in jail is "par for the course."

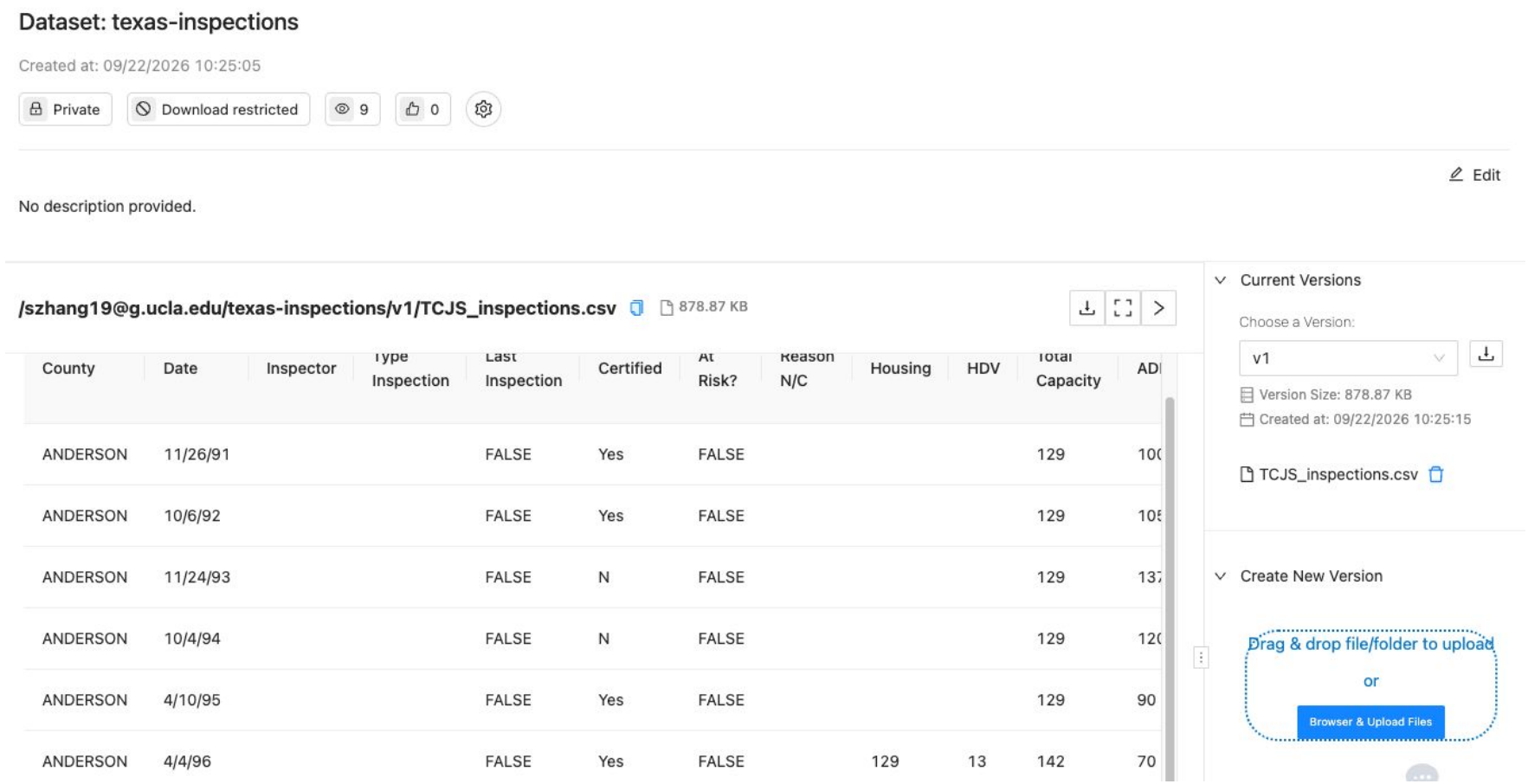
by MICHAEL BARAJAS *and* SOPHIE NOVACK

NOVEMBER 1, 2021, 8:19 AM, CDT

- Data: Obtained via public records requests to the Texas Commission on Jail Standards and made public by journalists Michael Barajas and Sophie Novack on:
  - Death investigations, including indications of falsified records
  - Agency inspection of jail facilities



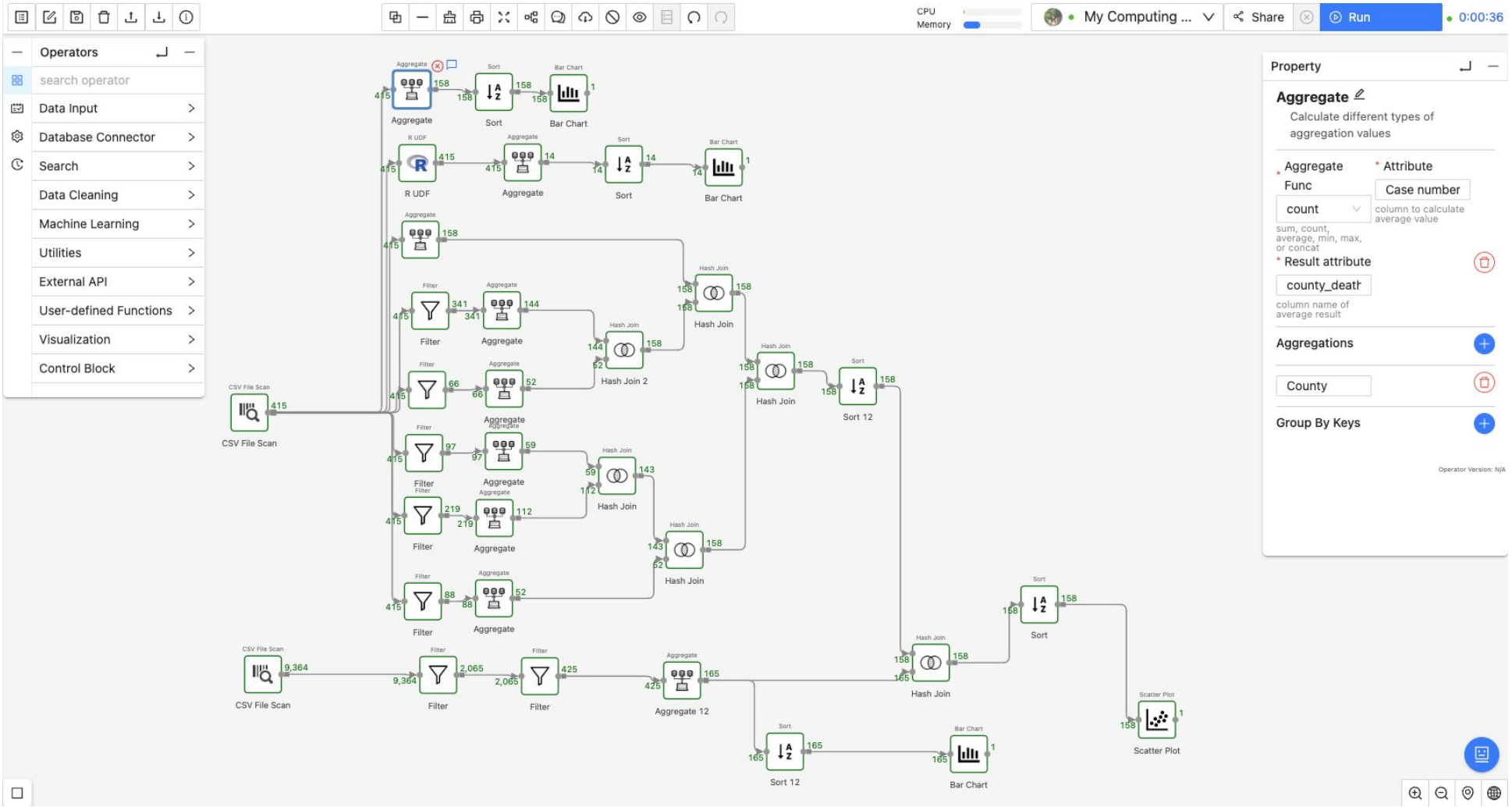
Data: jail inspections



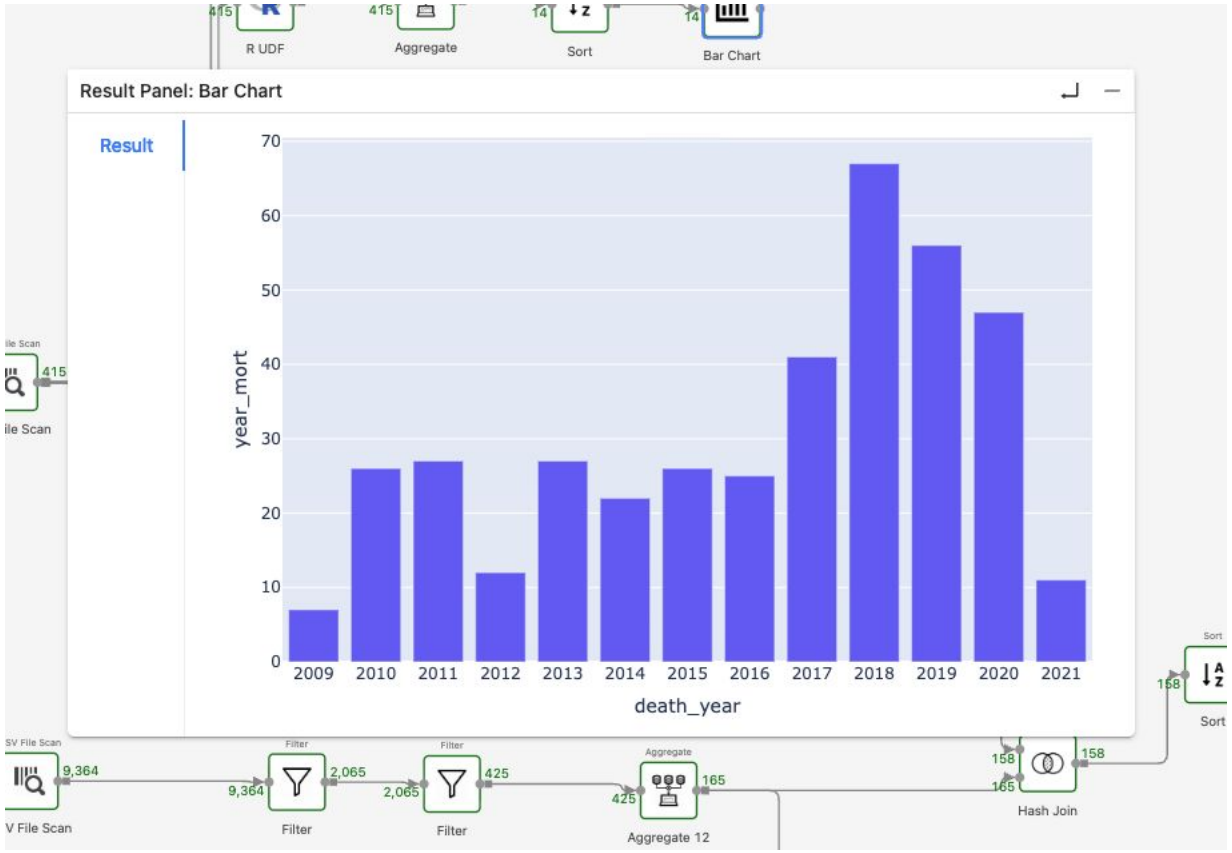
Data: deaths in custody

Case number	Deceased name	Pages	DOB	Race/ethnicity	Gender	County	Private contractor	DOD	Investigation outcome	Manner of death	Cause of death from autopsy	Suicide risk	Ignoring outcries or illness	Evidence of falsified cell checks	Evidence of falsified treatment records	Case summary
2017I-TRE-50015142	Garcia, Isaac Wiszar	17	11/13/1965	Hispanic	Male	Tom Green		7/14/2017	Closed, nothing followed Rangers investigation		Died as a result of Sepsis due to acute pyelonephritis. Sepsis due to a urinary tract infection. Additionally, Garcia was found to have Anasarca, Liver cirrhosis with a tumor, Hypertensive cardiovascular disease, and Intestinal necrosis.		Cellmates told the Rangers that jailers accused Garcia of faking his pain. One cellmate said jailer remarked that Garcia was "just whining."			Garcia saw jail nurses several times for back pain, shoulder pain, shortness of breath, elevated blood pressure, elevated body temperature, trouble urinating, bloated abdomen, fever, and vomiting, and was placed in the medical holding cell for observation. On July 10, a nurse reported a call from his family complaining that he was not being treated for back and neck pain. The next day he was sent to the hospital, where he died a few days later. The report says Garcia had been fighting an infection and had numerous blood transfusions.[illegible]
2018I-TRC-50029700	Angell, Edwin Carl	7	1/3/1959	White	Male	Clay		10/6/2018	Closed, nothing followed Rangers investigation	Natural causes	[illegible]Died as a result of hepatic cirrhosis in conjunction with hepatitis C infection and chronic substance abuser. Hypertensive and atherosclerotic cardiovascular disease contributed to his death.[illegible]					The sheriff[illegible]s office obtained personal recognizance bonds to release Angell, who before his arrest had been in in-home hospice care for a year due to multiple illnesses, but the sheriff said the TDCJ Pardon and Parole Board would not authorize his release. A couple days later, he was found unresponsive in his cell and taken to a hospital, then to a hospice facility, where he died. Another inmate with the same last name (relation unspecified) told the Rangers that Angell had been coughing up blood and had been telling jail staff he needed to go to the hospital, and that Angell[illegible]s health declined quickly while incarcerated. According to the report, the DA said there was "no need" to present the case to a grand jury since the cause of death was "natural."

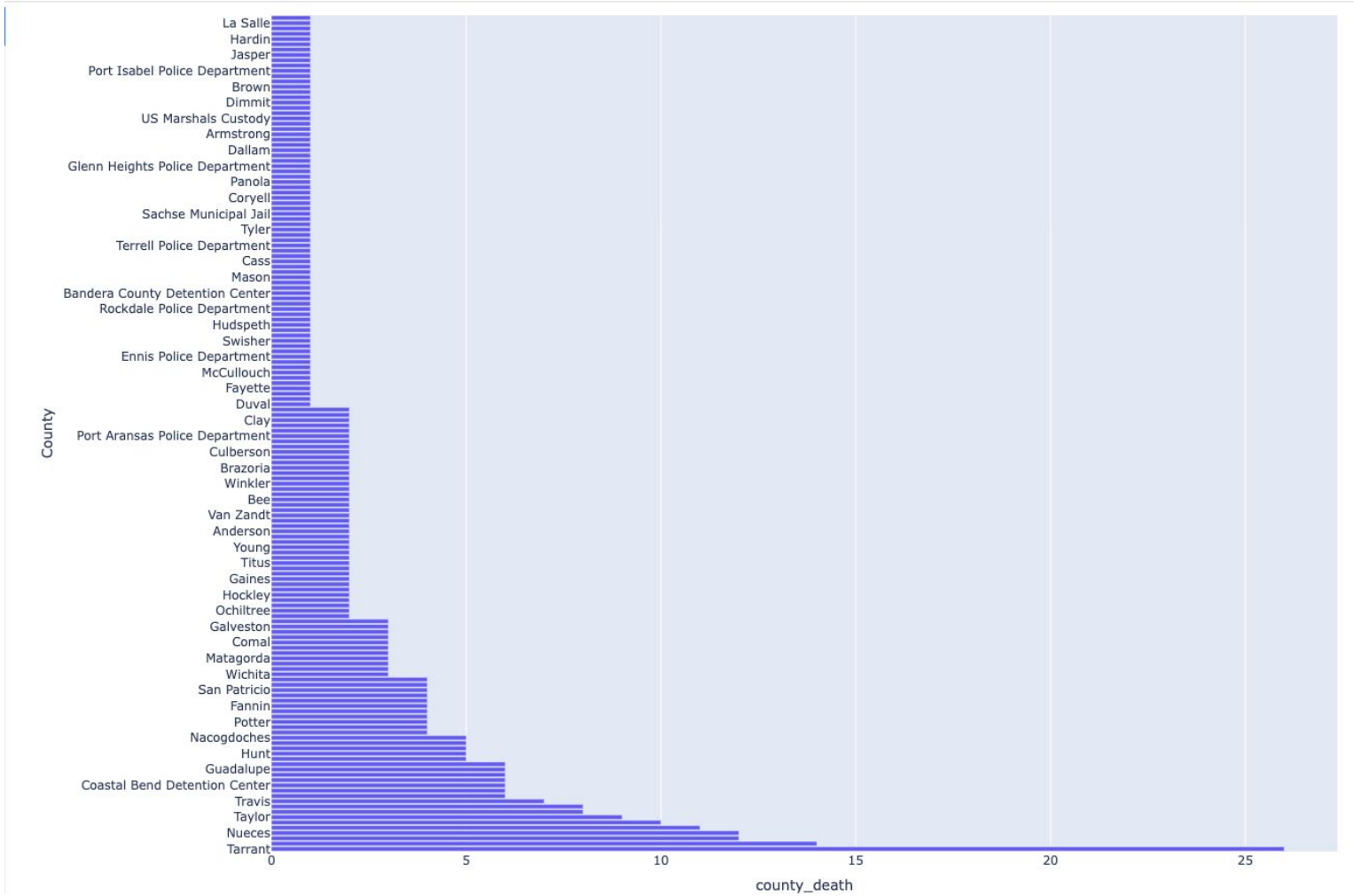
Workflow


Operators
search operator
Data Input
Database Connector
Search
Data Cleaning
Machine Learning
Utilities
External API
User-defined Functions
Visualization
Control Block
CPU
Memory
My Computing ...
Share
Run
0:00:36
Property
Aggregate
Calculate different types of aggregation values
Aggregate Func
count
sum, count, average, min, max, or concat
Attribute
Case number
column to calculate average value
Result attribute
county_death
column name of average result
Aggregations
County
Group By Keys
Operator Version: N/A
CSV File Scan
Filter
Aggregate
Hash Join
Hash Join 2
Sort
Sort 12
Bar Chart
R UDF
Aggregate 12
Scatter Plot

Output: deaths by year, 2009-2021



Output: deaths by county, 2009-2021



Next steps

- Continue to add project- and region-based data on deaths in custody

- Solicit user feedback from diverse audiences
  - Social science researchers
  - Legal experts and professors
  - Journalists

- Analyze across datasets to identify new insights about deaths in custody, factors relating to deaths, and recommendations to prevent deaths in custody

Thanks!

Naomi Sugie, nsugie@soc.ucla.edu

UCLA

Support from National Science Foundation Integrated Data Systems and Services award (2609582)

UC Irvine