

NIST



PSCR

PUBLIC SAFETY
COMMUNICATIONS
RESEARCH

hero^x

DRIVEN DATA



Differential Privacy Temporal Map Challenge

Algorithm Contest Webinar – Sprint 2

January 14, 2021

Gary Howarth (NIST), Christine Task (Knexus Research), Greg Lipstein (DrivenData)

Agenda

- ❖ Background
- ❖ Challenge overview
- ❖ How to participate
- ❖ Q&A

PSCR Overview

Public Safety Communications Research Division of the National Institute of Standards and Technology is the primary federal laboratory conducting research, development, testing, and evaluation for public safety communications technologies.



5 Key Research Areas

LMR to LTE

User Interface User Experience
Mission Critical Voice

Location-Based Services
Public Safety Analytics

Security
Resilient Systems

Cross Cutting Research Areas

Why the Challenge?

- The Public Safety Communications Research Division (PSCR) of the National Institute of Standards and Technology (NIST) is sponsoring this exciting data science competition to help advance research for public safety communications technologies for America's First Responders
- As first responders utilize more advanced communications technology, there are opportunities to use data analytics to gain insights from public safety data, inform decision-making and increase safety.
- **But... we must assure data privacy.**



What's the Problem?

Public Safety As Data Generators

- As Public Safety entities make enormous gains in cyber and data infrastructure leading to the routine collection of many large datasets.
- Governments and the public are demanding greater protections on individual privacy and the privacy of individual records.
- Open data initiatives are pushing for the release of more information.

Public Safety Generates Sensitive Information

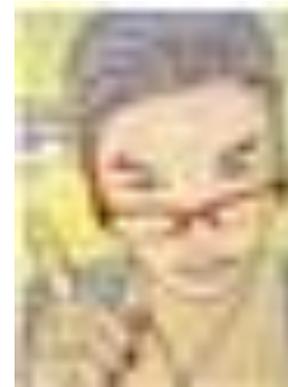
- Included in the data is personally identifiable information (PII) for police officers, victims, persons of interest, witnesses, suspects, etc.
- Studies have found that a combination of just 3 “quasi-identifiers” (date of birth, 5 digit postal code, and gender) uniquely identifies 87% of the population.

What do we mean by Privacy?

Privacy-preserving data-mining algorithms allow trusted data-owners to release useful, aggregate information about their data-sets (such as common user behavior patterns) while at the same time protecting individual-level information.

Intuitively, the concept of making large patterns visible while protecting small details makes sense.

You just 'blur' things a bit:



<http://fryeart1.weebly.com/journals.html>

If we refine this idea into a mathematically formal definition, we can create a standard for individual privacy.

Differential Privacy Prescreening--and Pitfalls to Avoid

Get Prescreened!



LEADERBOARD

DATA DOWNLOAD

SUBMISSIONS

PRE-SCREEN SUBMISSION

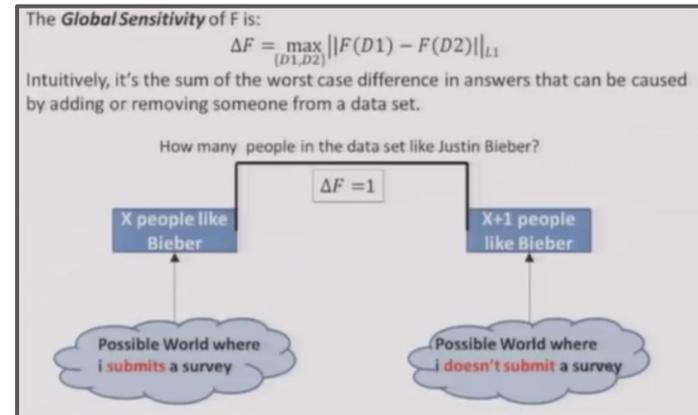
On the **Open Arena** the rules are made up and the points don't matter... especially if you have a perfect score.

Get your proof checked, and get into the **Prescreened Arena** today!

In order to be eligible for final scoring, participants must have their approach pre-screened as differentially private. To be considered for pre-screening, submit a brief PDF document below including a written explanation of your algorithm, any additional data sources used other than the provided data set(s), and a clear, correct mathematical proof that the solution satisfies differential privacy. An example pre-screening submission is provided in the competitor pack.

This document will be reviewed and validated by NIST staff or their delegates. Participants will receive "Prescreened" status if their written explanation proves that they have an essentially correct understanding of differential privacy as applied to their submission, or a brief explanation why their algorithm or proof is incorrect.

Watch your Sensitivity!



In privacy, every time you reference the ground truth data in any fashion, you're incurring a privacy cost. Don't get into debt! Make sure to keep your accounting straight and pay up with sufficient privacy noise!

Remember, **the maximum records per individual is 7 in sprint 2**. That means a simple count of the number of records in a given PUMA in a given year has a sensitivity of 7. Check out the privacy resources on the drivendata website, and feel free to ask questions on the forum.

Individuals can contribute to multiple time/map segments, which is tricky to keep in mind. **Be extra careful when doing parallel composition!**

Objective

In the Differential Privacy Temporal Map Challenge (DeID2) your task is to **develop algorithms that preserve data utility as much as possible while guaranteeing individual privacy is protected.**

Submissions will be assessed based on

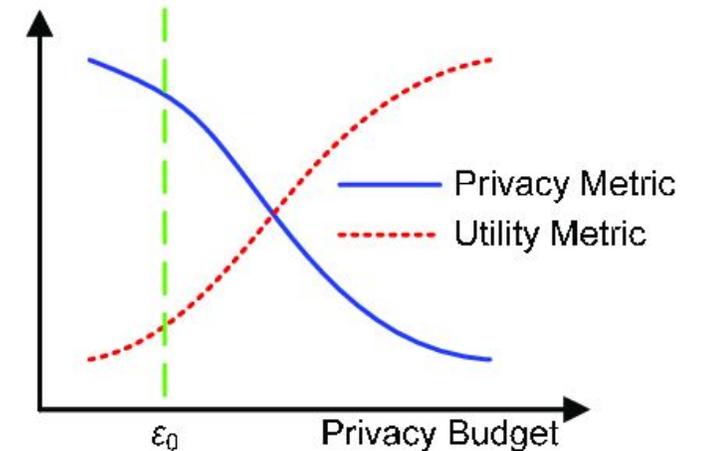
1. their ability to **prove they satisfy differential privacy**; and
2. the **accuracy of output data** as compared with ground truth.

1

Privacy write-ups
Confirmed by subject
matter experts

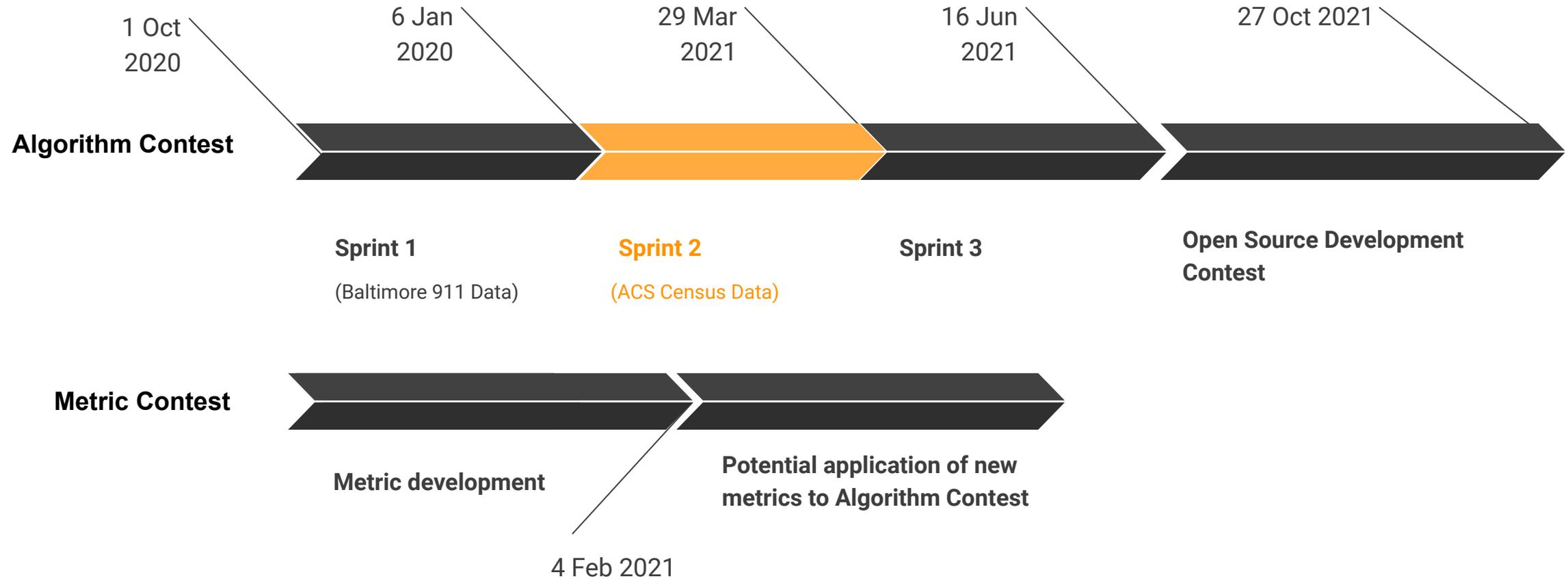
2

Algorithm submissions
Evaluated by published
performance metric



Sample illustration of the privacy-utility tradeoff.
From Liu et al. "Privacy-Preserving Monotonicity of
Differential Privacy Mechanisms." 2018.

Challenge Timeline



Prize Awards

Sprint 1 Oct - Dec 2020	Sprint 2 Jan - Mar 2021	Sprint 3 Apr - Jun 2021
1st Place: \$10,000 2nd Place: \$7,000 3rd Place \$5,000 4th place \$2,000 5th place \$1,000	1st Place: \$15,000 2nd Place: \$10,000 3rd Place \$5,000 4th place \$3,000 5th place \$2,,000	1st Place: \$25,000 2nd Place: \$20,000 3rd Place \$15,000 4th place \$10,000 5th place \$5,000
progressive prize: 4 @\$1,000	progressive prize: 4 @\$1,000	progressive prize: 4 @\$1,000
Total: \$29,000	Total: \$39,000	Total: \$79,000
Metric Paper Prizes: \$29,000 Open Source Development Prizes: \$100,000 Total Prize Purse for Differential Privacy Temporal Map Challenge: \$276,000		

Algorithm Sprint Structure

This contest sprint will proceed in two phases:

- **Development Phase (Jan 6 - Feb 15, 2021)**

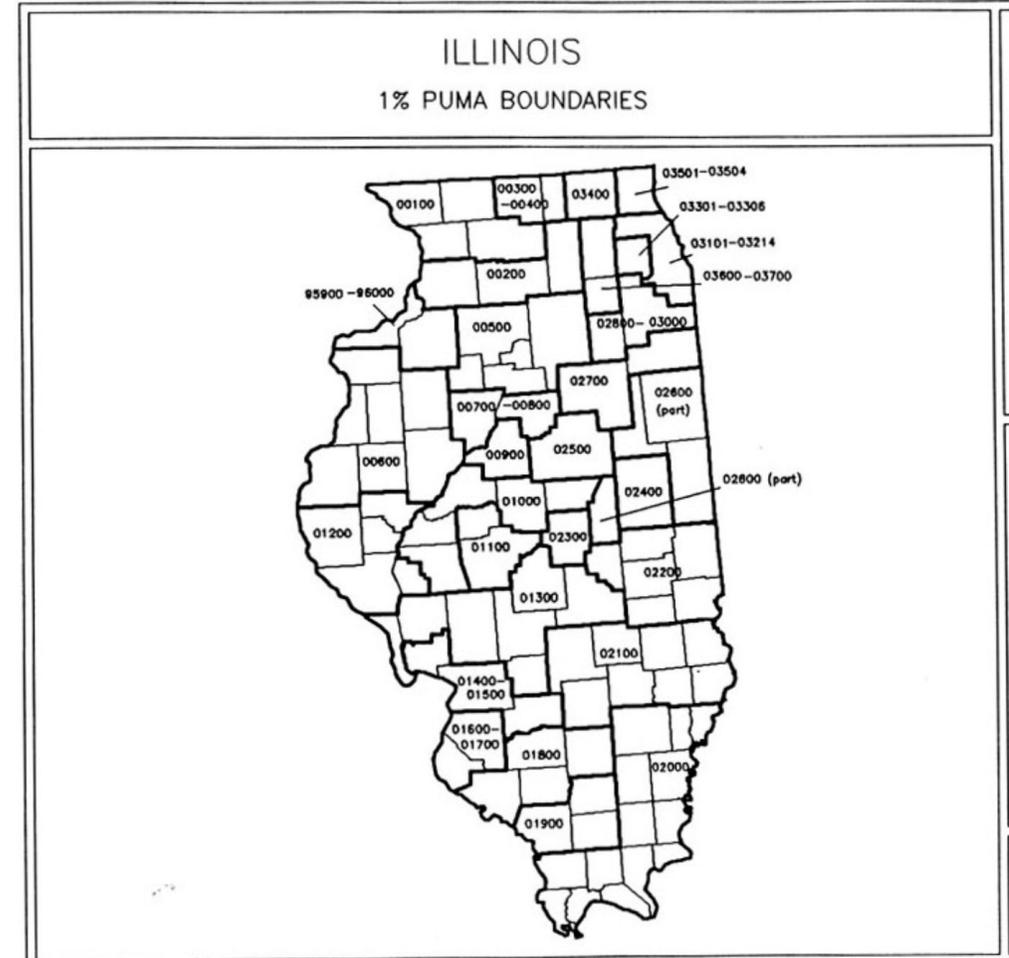
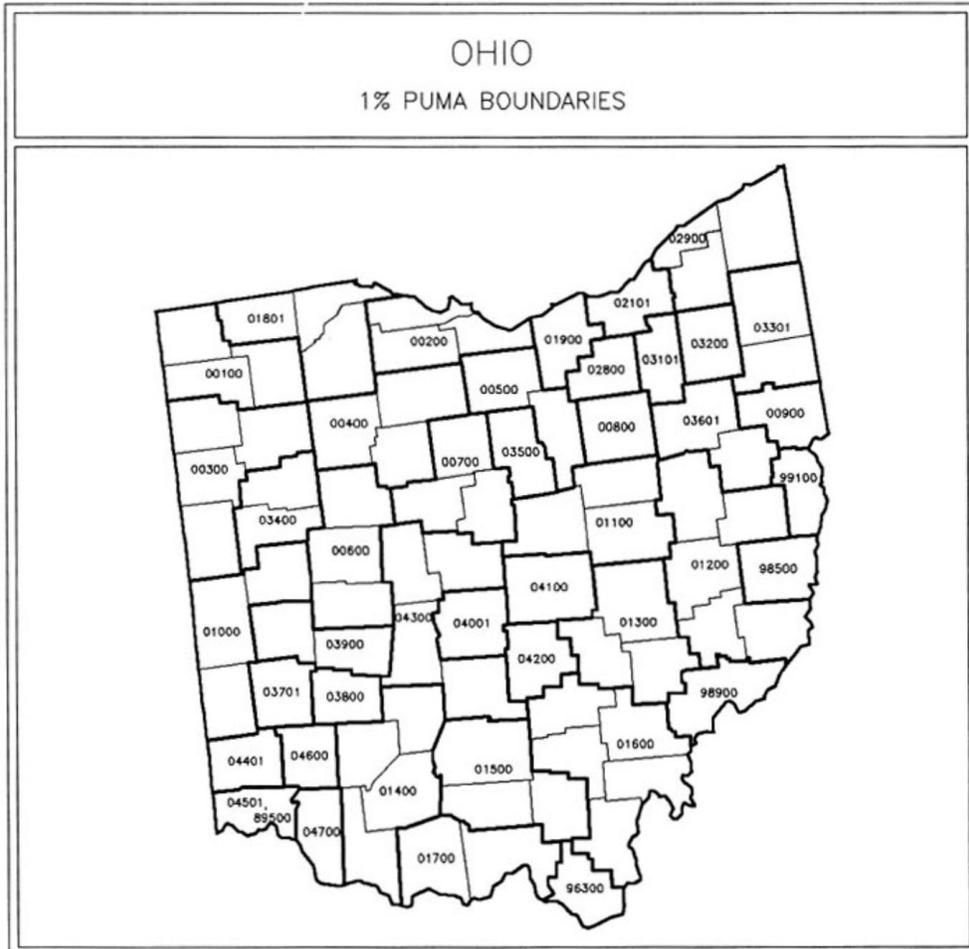
- Competitors use 'publicly available' data to evaluate and refine their algorithms without privacy loss.
- Competitors submit privatized data to be scored on public facing Leaderboard.
- Algorithm write-ups and privacy proofs can be submitted to be **pre-screened as differentially private**.
- Prescreened participants can enter the Prescreened Arena, make executable code submissions to the containerized test harness, and qualify for Final Scoring.
- **Progressive prizes** reward best scores by **Jan 25th**, giving precedence to the Prescreened Leaderboard.

- **Final Scoring Phase (Feb 15 - 22, 2021)**

- Participants who have passed pre-screening are invited to submit their **final code** (docker container executables) and **write-ups** (privacy proofs, source code and code guides) for final scoring.
- SME panel performs **Final DP Validation** on proofs and source code. Competitors may be asked to fix minor errors and resubmit.
- **Final Scoring** occurs on multiple 'private' data sets with the same schema as the 'publicly available' data, at multiple epsilon values.
- 1st through 5th Place prizes are based on final scoring.

About the Sprint 2 Data: American Community Survey

There are 36 total columns in the ground truth data, including `PUMA`, `YEAR`, 33 additional survey features, and an ID denoting simulated individuals. More information is available on the challenge website and in the `parameters.json`



About the Sprint 2 Data: American Community Survey

There are 36 total columns in the ground truth data, including `PUMA`, `YEAR`, 33 additional survey features, and an ID denoting simulated individuals. More information is available on the challenge website and in the `parameters.json`

- `PUMA` (str) — Identifies the Public Use Microdata Area (PUMA) where the housing unit was located.
- `YEAR` (uint32) — Reports the four-digit year when the household was enumerated or included in the survey.
- `HHWT`, `PERWT`, `GQ` Meta/Survey Data
- `SEX`, `AGE`, `MARST`, `RACE`, `HISPAN`, `CITIZEN`, `SPEAKENG` - Demographic Features
- `HCOVANY`, `HCOVPRIV`, `HINSEMP`, `HINSCAID`, `HINSCARE` (uint8) — Indicates whether persons had any health insurance coverage at the time of interview, and whether they had private, employer-provided, Medicaid or other government insurance, or Medicare coverage, respectively.
- `EDUC` (uint8) — Indicates respondents' educational attainment, as measured by the highest year of school or degree completed.
- `EMPSTAT`, `EMPSTATD`, `LABFORCE`, `WRKLISTWK`, `ABSENT`, `LOOKING`, `AVAILBLE`, `WRKRECAL`, `WORKEDYR` (uint8) — Indicates whether the respondent was a part of the labor force (working or seeking work), whether the person was currently unemployed, their work-related status in the last week, whether they were informed they would be returning to work (if not working in the last week), and whether they worked during the previous year.
- `INCTOT`, `INCWAGE`, `INCWELFR`, `INCINVST`, `INCEARN` (int32) — Reports each respondent's total pre-tax personal income or losses from all sources for the previous year, as well as the income from wages, welfare, investment, and wages or a person's own business or farm, respectively.
- `POVERTY` (uint32) — Expresses each family's total income for the previous year as a percentage of the poverty threshold.
- `DEPARTS`, `ARRIVES` (uint32) — Reports the time that the respondent usually left home for work and arrived at work last week, measured using a 24-hour clock.
- `sim_individual_id` (int) — Unique, synthetic ID for the notional person to which this record was attributed. The largest number of records attributed to a single simulated resident is provided in the `parameters.json` file as `max_records_per_individual`.

About the Sprint 2 Data: American Community Survey

There are 36 total columns in the ground truth data, including `PUMA`, `YEAR`, 33 additional survey features, and an ID denoting simulated individuals. More information is available on the challenge website and in the `parameters.json`

- `PUMA` (str) — Identifies the Public Use Microdata Area (PUMA) where the housing unit was located.
- `YEAR` (uint32) — Reports the four-digit year when the household was enumerated or included in the survey.
- `HHWT`, `PERWT`, `GQ` Meta/Survey Data
- `SEX`, `AGE`, `MARST`, `RACE`, `HISPAN`, `CITIZEN`, `SPEAKENG` - Demographic Features
- `HCOVANY`, `HCOVPRIV`, `HINSEMP`, `HINSCAID`, `HINSCARE` (uint8) — Indicates whether persons had any health insurance coverage at the time of interview, and whether they had private, employer-provided, Medicaid or other government insurance, or Medicare coverage, respectively.
- `EDUC` (uint8) — Indicates respondents' educational attainment, as measured by the highest year of school or degree completed.
- `EMPSTAT`, `EMPSTATD`, `LABFORCE`, `WRKLISTWK`, `ABSENT`, `LOOKING`, `AVAILBLE`, `WRKRECAL`, `WORKEDYR` (uint8) — Indicates whether the respondent was a part of the labor force (working or seeking work), whether the person was currently unemployed, their work-related status in the last week, whether they were informed they would be returning to work (if not working in the last week), and whether they worked during the previous year.
- `INCTOT`, `INC` — Total and pre-tax personal income or losses from all sources for the previous year, as well as income from business or farm, respectively.
- `POVERTY` (uint8) — Percentage of the poverty threshold.
- `DEPARTS`, `ARRIVES` — Whether the respondent departed from home and arrived at work last week, measured using a 24-hour clock.
- `sim_individual_id` (int) — Unique, synthetic ID for the notional person to which this record was attributed. The largest number of records attributed to a single simulated resident is provided in the `parameters.json` file as `max_records_per_individual`.

Simulated individuals were made by matching records across years.

All simulated individuals have between 4 and 7 records.

Demographics and education were used as hard matching criteria,

Income and PUMA were used as softer criteria.

About the Problem: Pile of Individual Records



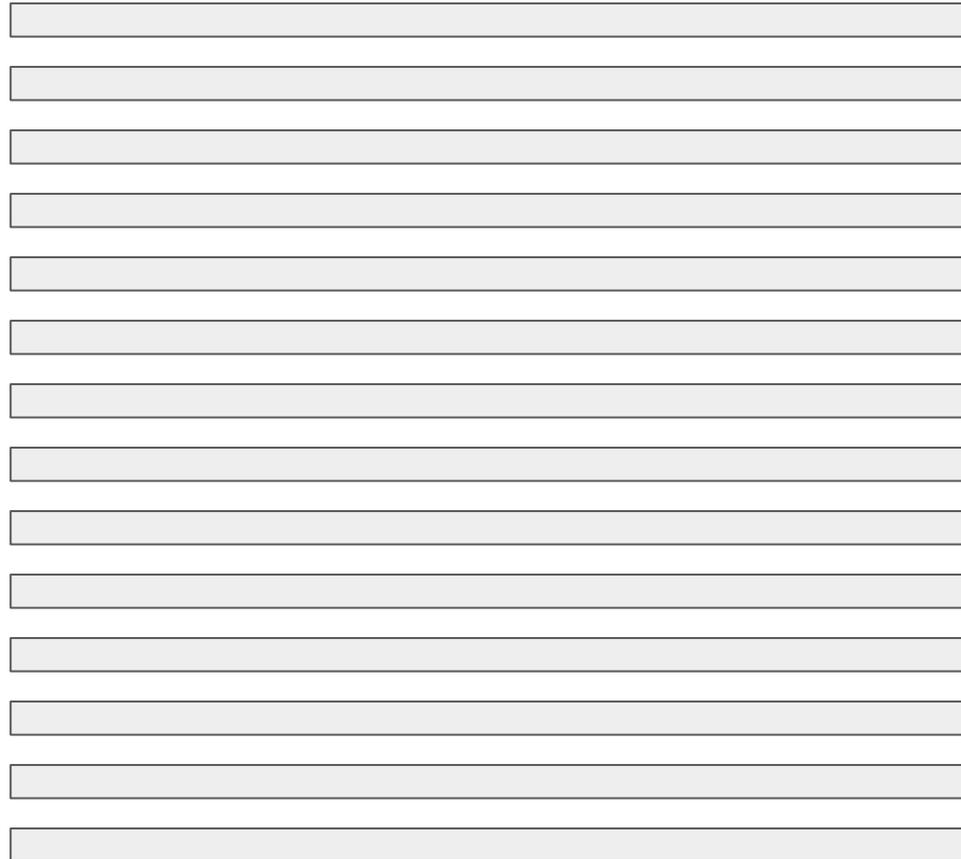
As compared to Sprint #1, this is a much larger feature space. This means output needs to be Synthetic Data.

All individuals have between 4-7 records.

Epsilon values are more challenging, at 10, 1, 0.1

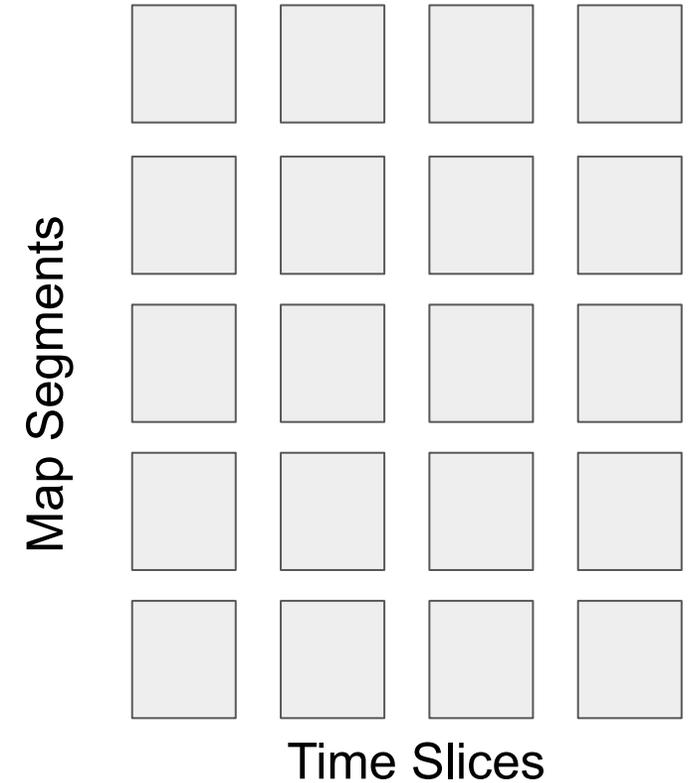
Survey Records:

Time stamp, Map Segment, Features, Person ID



Evaluation Space:

Synthetic Records by Time Slice and Map Segment



About the Problem: The Sprint 2 Baseline Solution



You may notice that the new baseline solution doesn't actually make use of the input epsilon values, or the Laplace random number generator. In fact, it doesn't make use of the input data at all.

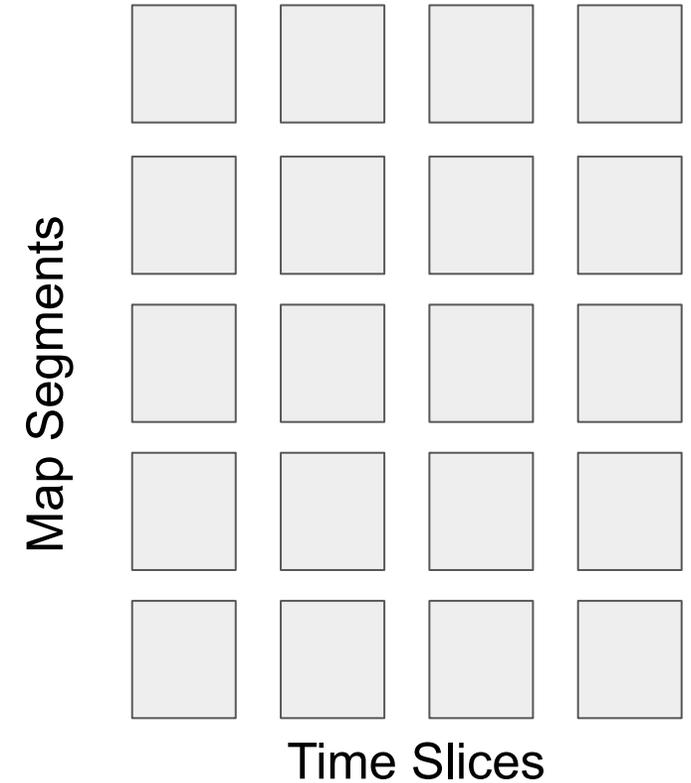
Instead, it generates complete, correctly formatted records by uniformly randomly selecting values for each feature using **the list of possible values in the schema provided in the parameters.json file.**

Of course, that blind sampling isn't likely to provide much utility. But if you're just starting out and you're looking for ideas, you might think about queries that you could privatize that would let you improve on that procedure.

```
"YEAR": {
  "dtype": "uint32"
  "values": [
    2012,
    2013,
    2014,
    2015,
    2016,
    2017,
    2018
  ]
},
"HHWT": {
  "dtype": "float"
},
"GQ": {
  "values": [
    0,
    1,
    2,
    3,
    4,
    5,
    6
  ],
  "dtype": "uint8"
},
"PERWT": {
  "dtype": "float"
},
"AGE": {
  "min": 0,
  "max": 135,
  "dtype": "uint8"
},
"MARST": {
  "values": [
    1,
    2,
    3,
    4,
    5,
    6
  ],
  "dtype": "uint8"
},
"RACE": {
  "values": [
    1,
    2,
    3,
    4,
    5,
    6,
    7,
    8,
    9
  ],
  "dtype": "uint8"
}
```

Evaluation Space:

Synthetic Records by Time Slice and Map Segment

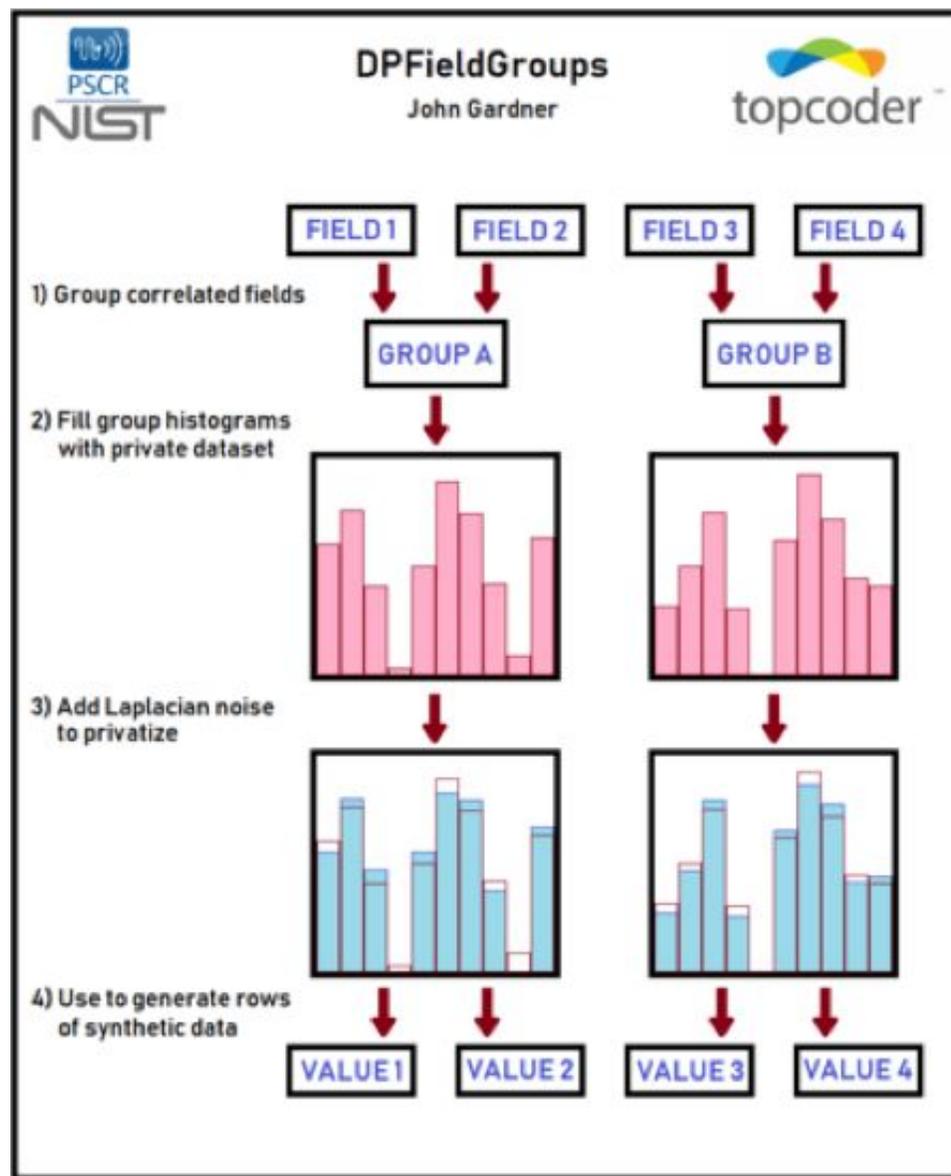


About the Problem: Differential Private Synthetic Data



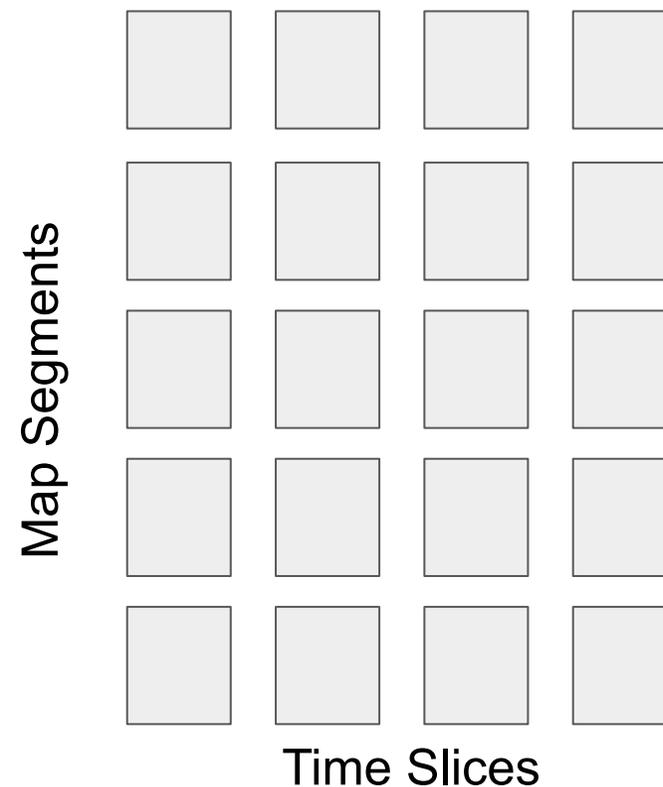
There's lots of great techniques for producing differentially private synthetic data, using marginals, probabilistic graphical models, GANS, or other approaches.

Check out the results of the **2018 NIST Differential Privacy Synthetic Data Challenge**



Evaluation Space:

Synthetic Records by Time Slice and Map Segment



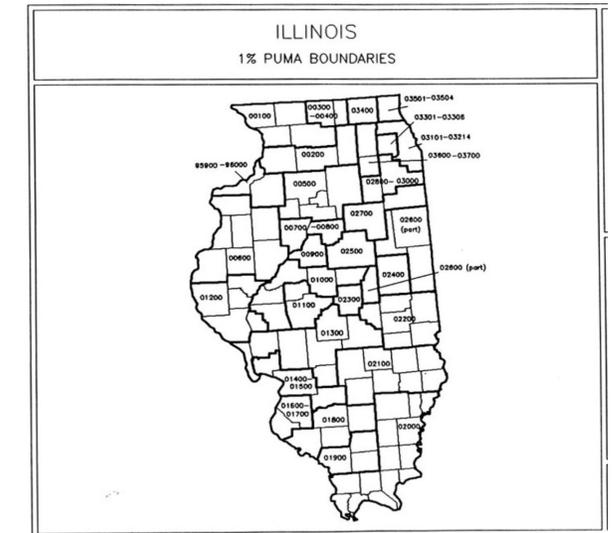
A Few Notes on the Public vs Sequestered Data

In this Sprint, we'll be keeping the years the same for final judging, but changing which states you're running on.

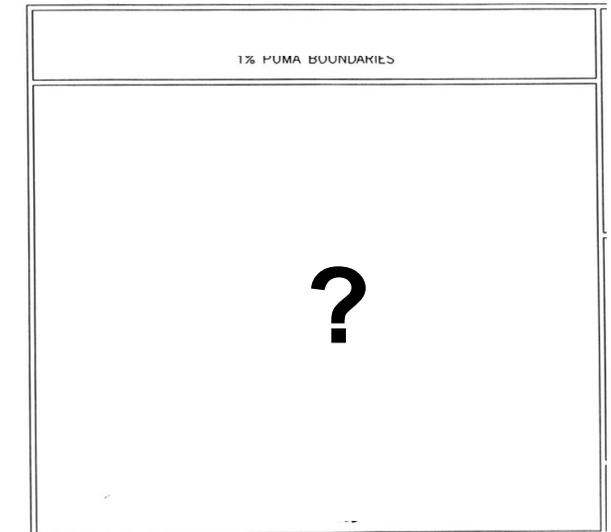
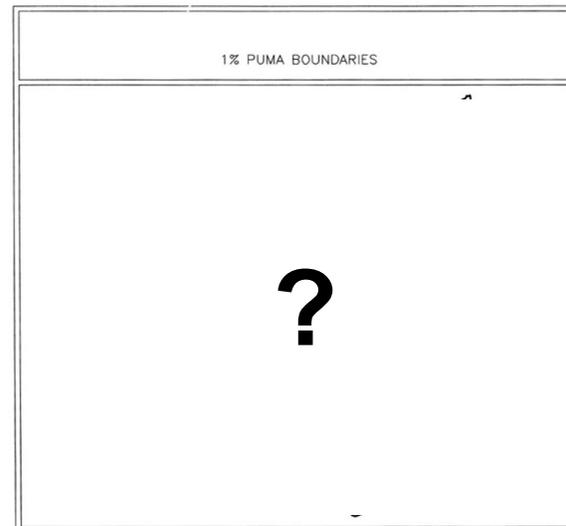
That means that things like **general economic trends/ranges** and **correlation patterns between features** can be learned from the public data and applied to the private data...

But you cannot learn about patterns in the public data maps (PUMA's) and apply them to the private data.

Public Data:



Final Judging Data:



A Few Notes on Gentrification and Civic Planning

Recall that simulated individuals were made by solving a matching problem that attempted to maintain individuals in their PUMAs and in their income brackets.

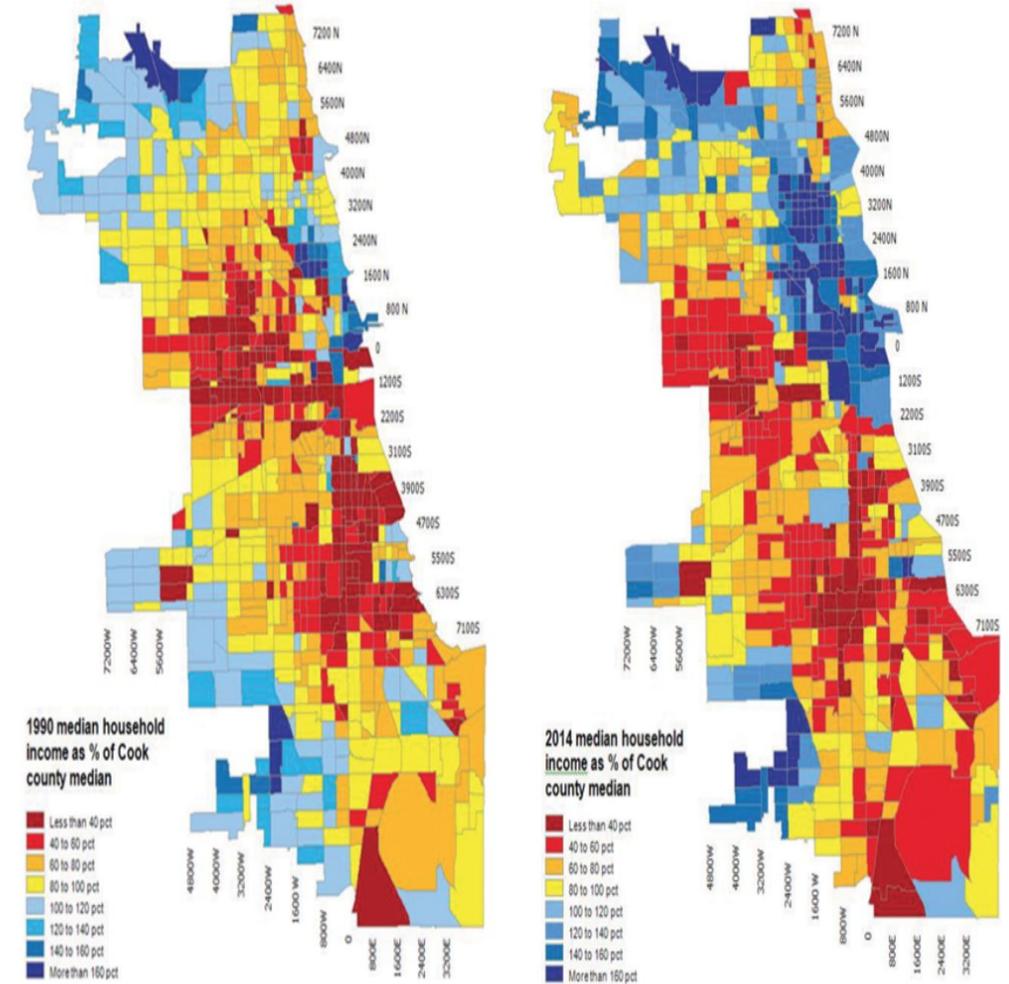
Individuals that we couldn't match for at least 4 records were dropped from the data.

What happens when a PUMA experiences a significant change over the 7 years of data? Some PUMA populations increase or decrease, become wealthier or poorer, or their demographics change.

You'll clearly see that impact in the data. **Those PUMAs may be the most interesting and the most challenging to model.**

- **PUMA** (str) — Identifies the Public Use Microdata Area (PUMA) where the housing unit was located.
- **YEAR** (uint32) — Reports the four-digit year when the household was enumerated or included in the survey.
- **SEX**, **AGE**, **MARST**, **RACE**, **HISPAN**, **CITIZEN**, **SPEAKENG** - Demographic Features
- **INCTOT**, (int32) — Reports each respondent's total pre-tax personal income or losses from all sources for the previous year,

CHICAGO



Source: **Chicago Central Area Committee**

About Sprint 2 Scoring: K-Marginal Scoring

Baseline 3-Marginal Score:

The k-marginal evaluation tool is a randomized heuristic that measures similarity between two high dimensional data sets, by considering all correlations of k or fewer variables (we often use $k = 3$). It was developed to score solutions in the NIST DeID1 Synthetic Data Challenge.

Here's how it works:

- (1) Numerical features are grouped into range bins.
- (2) Select a set of k -marginals to score according to the picking strategy specified in the configuration (for example—choose marginals uniformly random at a sampling rate of 0.1).
- (3) For each selected k -marginal, compute absolute difference between the two densities
- (4) Average scores across all selected k -marginals to create a total score.
- (5) Compute the human-readable NIST score = $((2 - \text{raw_score})/2) \times 1000$.

		Features						
		A	B	C	D	E	F	G
Records	1	a4	b1	c2	d1	e1	f3	g1
	2	a1	b1	c1	d1	e1	f5	g2
	3	a2	b1	c4	d2	e2	f6	g1
	4	a1	b1	c6	d3	e2	f8	g2
	5	a3	b2	c5	d2	e1	f11	g1

	100	a4	b2	c7	d3	e2	f14	g2

An example dataset with three features selected: B, E, G.

		REAL			SYNTHETIC						
		Features			Features						
		B	E	G	Counts	Density	B	E	G	Counts	Density
Bins	1	b1	e1	g1	10	0.10	b1	e1	g1	8	0.08
	2	b1	e1	g2	11	0.11	b1	e1	g2	10	0.10
	3	b1	e2	g1	16	0.16	b1	e2	g1	17	0.17
	4	b1	e2	g2	15	0.15	b1	e2	g2	18	0.18
	5	b2	e1	g1	9	0.09	b2	e1	g1	8	0.08
	6	b2	e1	g2	8	0.08	b2	e1	g2	10	0.10
	7	b2	e2	g1	14	0.14	b2	e2	g1	13	0.13
	8	b2	e1	g2	17	0.17	b2	e1	g2	16	0.16

Marginal density on the real and synthetic data for selected features.

About Sprint 2 Scoring: K-Marginal Scoring

Baseline 3-Marginal Score:

The k-marginal evaluation tool is a randomized heuristic that measures similarity between two high dimensional data sets, by considering all correlations of k or fewer variables (we often use $k = 3$). It was developed to score solutions in the NIST DeID1 Synthetic Data Challenge.

Here's how it works:

- (1) Numerical features are grouped into range bins.
- (2) Select a set of k -marginals to score according to the picking strategy specified in the configuration (for example—choose marginals uniformly random at a sampling rate of 0.1).
- (3) For each selected k -marginal, compute absolute difference between the two densities
- (4) Average scores across all selected k -marginals to create a total score.
- (5) Compute the human-readable NIST score = $((2 - \text{raw_score})/2) \times 1000$.

REAL						SYNTHETIC						
Features						Features						
	B	E	G	Counts	Density		B	E	G	Counts	Density	
Bins	1	b1	e1	g1	10	0.10	1	b1	e1	g1	8	0.08
	2	b1	e1	g2	11	0.11	2	b1	e1	g2	10	0.10
	3	b1	e2	g1	16	0.16	3	b1	e2	g1	17	0.17
	4	b1	e2	g2	15	0.15	4	b1	e2	g2	18	0.18
	5	b2	e1	g1	9	0.09	5	b2	e1	g1	8	0.08
	6	b2	e1	g2	8	0.08	6	b2	e1	g2	10	0.10
	7	b2	e2	g1	14	0.14	7	b2	e2	g1	13	0.13
	8	b2	e1	g2	17	0.17	8	b2	e1	g2	16	0.16

Marginal density on the real and synthetic data for selected features.

Bin Number	Real Density	Synthetic Density	Difference	Absolute Value
1	0.10	0.08	0.02	0.02
2	0.11	0.10	0.01	0.01
3	0.16	0.17	-0.01	0.01
4	0.15	0.18	-0.03	0.03
5	0.09	0.08	0.01	0.01
6	0.08	0.10	-0.02	0.02
7	0.14	0.13	0.01	0.01
8	0.17	0.16	0.01	0.01
SUM				0.12

About Sprint 2 Scoring: K-Marginal Scoring

Baseline 3-Marginal Score:

The k-marginal evaluation tool is a randomized heuristic that measures similarity between two high dimensional data sets, by considering all correlations of k or fewer variables (we often use k = 3). It was developed to score solutions in the NIST DeID1 Synthetic Data Challenge.

Here's how it works:

- (1) Numerical features are grouped into range bins.
- (2) Select a set of k-marginals to score according to the picking strategy specified in the configuration (for example—choose marginals uniformly random at a sampling rate of 0.1).
- (3) For each selected k-marginal, compute absolute difference between the two densities
- (4) Average scores across all selected k-marginals to create a total score.
- (5) Compute the human-readable NIST score = $((2 - \text{raw_score})/2) \times 1000$.

		Features						
		A	B	C	D	E	F	G
Records	1	a4	b1	c2	d1	e1	f3	g1
	2	a1	b1	c1	d1	e1	f5	g2
	3	a2	b1	c4	d2	e2	f6	g1
	4	a1	b1	c6	d3	e2	f8	g2
	5	a3	b2	c5	d2	e1	f11	g1

	100	a4	b2	c7	d3	e2	f14	g2

NIST Score: 982

Continue across many randomly selected marginals, average results.

About Sprint 2 Scoring: K-Marginal Scoring

Baseline 3-Marginal Score:

The k-marginal evaluation tool is a randomized heuristic that measures similarity between two high dimensional data sets, by considering all correlations of k or fewer variables (we often use $k = 3$). It was developed to score solutions in the NIST DeID1 Synthetic Data Challenge.

Here's how it works:

- (1) Numerical features are grouped into range bins.
- (2) Select a set of k -marginals to score according to the picking strategy specified in the configuration (for example—choose marginals uniformly random at a sampling rate of 0.1).
- (3) For each selected k -marginal, compute absolute difference between the two densities
- (4) Average scores across all selected k -marginals to create a total score.
- (5) Compute the human-readable NIST score = $((2 - \text{raw_score})/2) \times 1000$.

Bias Penalty:

If the total number of individuals in a PUMA-YEAR varies from the Ground Truth by more than **250**, the score is 0 for that PUMA-YEAR

		Features						
		A	B	C	D	E	F	G
Records	1	a4	b1	c2	d1	e1	f3	g1
	2	a1	b1	c1	d1	e1	f5	g2
	3	a2	b1	c4	d2	e2	f6	g1
	4	a1	b1	c6	d3	e2	f8	g2
	5	a3	b2	c5	d2	e1	f11	g1

	100	a4	b2	c7	d3	e2	f14	g2

NIST Score: 982

Continue across many randomly selected k -marginals, average results.

Welcome Aboard to Sprint 2!

- ❖ The obstacles you'll run into on this challenge---sparse histograms, sparse map segments, longitudinal privacy, positive bias, ***complex and heterogeneous map segments***
- ❖ And the data properties you may leverage to help address them---***similarity of features***, trends across time, similarity of individuals, and many others...

Welcome Aboard to Sprint 2!

- ❖ The obstacles you'll run into on this challenge---sparse histograms, sparse map segments, longitudinal privacy, positive bias, ***complex and heterogeneous map segments***
- ❖ And the data properties you may leverage to help address them---***similarity of features***, trends across time, similarity of individuals, and many others...
- ❖ Are problem features that show up in critical real world differential privacy applications. And the standards you'll be held to are ***the standards of the stakeholders*** in those applications

Welcome Aboard to Sprint 2!

- ❖ The obstacles you'll run into on this challenge---sparse histograms, sparse map segments, longitudinal privacy, positive bias, ***complex and heterogeneous map segments***
- ❖ And the data properties you may leverage to help address them---***similarity of features***, trends across time, similarity of individuals, and many others...
- ❖ Are problem features that show up in critical real world differential privacy applications. And the standards you'll be held to are ***the standards of the stakeholders*** in those applications
- ❖ You can check out the Sprint #1 winners' announcement to see some of the creativity and intuition that was developed for dealing with this multi-record temporal data with Sprint #1. And you can look at the NIST Privacy Engineering Collaboration Space to see some of the solutions that were successful in the NIST Differential Privacy Synthetic Data Challenge.

Welcome Aboard to Sprint 2!

- ❖ The obstacles you'll run into on this challenge---sparse histograms, sparse map segments, longitudinal privacy, positive bias, ***complex and heterogeneous map segments***
- ❖ And the data properties you may leverage to help address them---***similarity of features***, trends across time, similarity of individuals, and many others...
- ❖ Are problem features that show up in critical real world differential privacy applications. And the standards you'll be held to are ***the standards of the stakeholders*** in those applications
- ❖ You can check out the Sprint #1 winners' announcement to see some of the creativity and intuition that was developed for dealing with this multi-record temporal data with Sprint #1. And you can look at the NIST Privacy Engineering Collaboration Space to see some of the solutions that were successful in the NIST Differential Privacy Synthetic Data Challenge.
- ❖ Let's see what you all can create for Sprint #2!

Important Dates

Sprint 2 starts	Jan 6, 2021
Deadline for pre-screening submissions to be reviewed before Progressive Prizes	Jan 19, 2021
Progressive Prize rankings determined	Jan 25, 2021
Development phase closes	Feb 15, 2021
Final code submissions and write-ups due	Feb 22, 2021
Winners announced	Mar 23, 2021

Competitor's Pack Contents

A [competitor pack](#) is provided to help participants get started! The contents include:

“Publicly available” survey records data

Parameters of the data, including schema and configuration inputs

Naive **baseline solution**

Sample **privacy write-up**

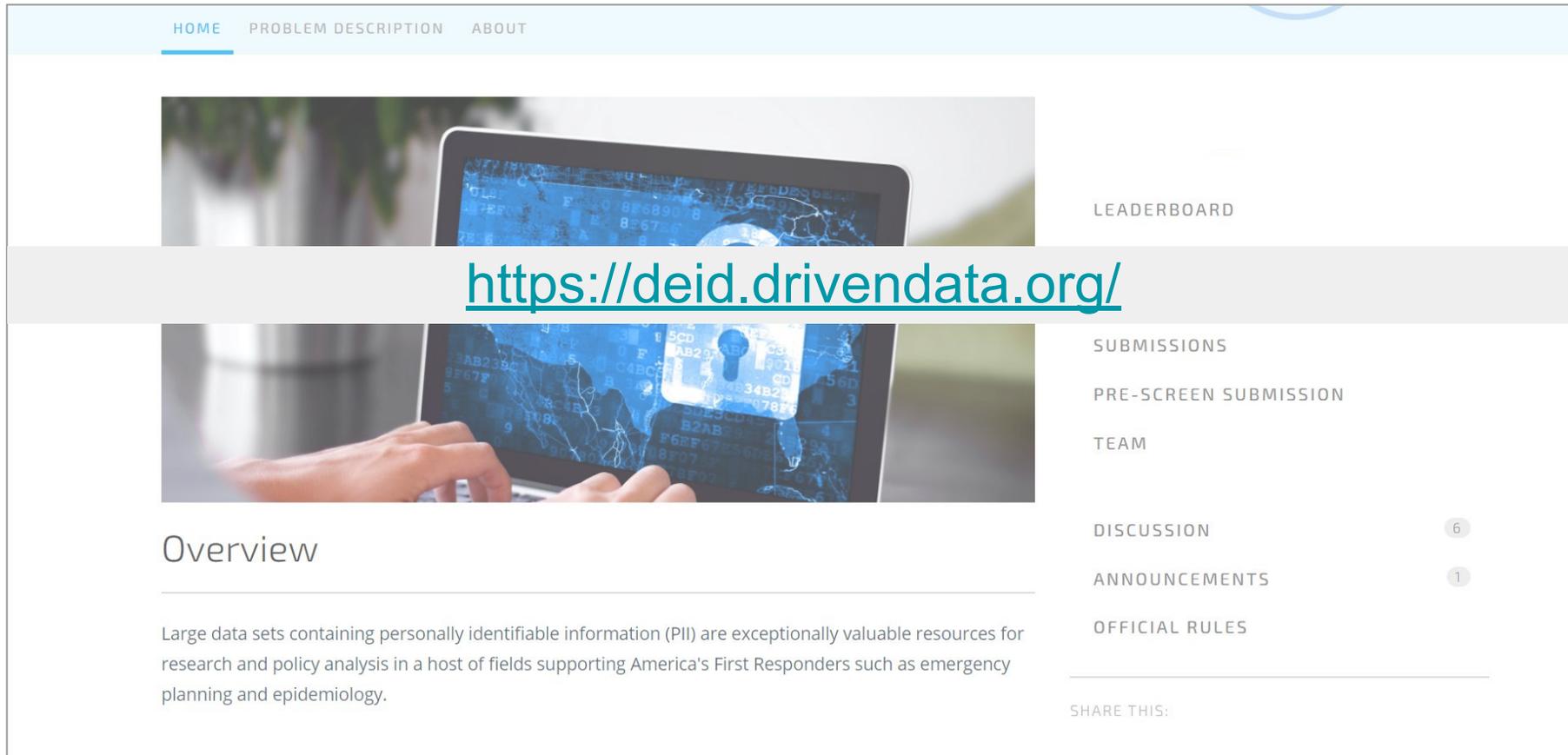
Scoring metric implementation for local testing

Simple visualizer of scoring outputs

Note on Eligibility

- ❖ The challenge is open to international participants, however to be eligible to receive a prize the Team Lead must be age 18 or older and a US citizen or permanent resident of the US or its territories.
- ❖ The Team Lead is the sole person who will accept the cash prizes on behalf of the team.
- ❖ Teams are welcome and encouraged to include participants from around the globe. The [forum](#) can be a great way to find teammates.
- ❖ Check out the [official rules on challenge.gov](#) for further information on eligibility.

Let's Get Going!



The screenshot shows the website <https://deid.drivendata.org/>. The navigation bar at the top includes links for HOME, PROBLEM DESCRIPTION, and ABOUT. The main content area features a large image of a laptop displaying data and a lock icon, with the URL <https://deid.drivendata.org/> overlaid. Below the image is the 'Overview' section, which states: 'Large data sets containing personally identifiable information (PII) are exceptionally valuable resources for research and policy analysis in a host of fields supporting America's First Responders such as emergency planning and epidemiology.' To the right, a sidebar menu lists: LEADERBOARD, SUBMISSIONS, PRE-SCREEN SUBMISSION, TEAM, DISCUSSION (6), ANNOUNCEMENTS (1), and OFFICIAL RULES. At the bottom of the sidebar is a 'SHARE THIS:' section.

Questions?

Competition Details and Official Rules

Challenge.gov

<https://www.challenge.gov/challenge/differential-privacy-temporal-map-challenge/>

DrivenData

<https://deid.drivendata.org/>

Challenge Questions

gary.howarth@nist.gov OR
[Challenge Forum](#)



Thank you!

