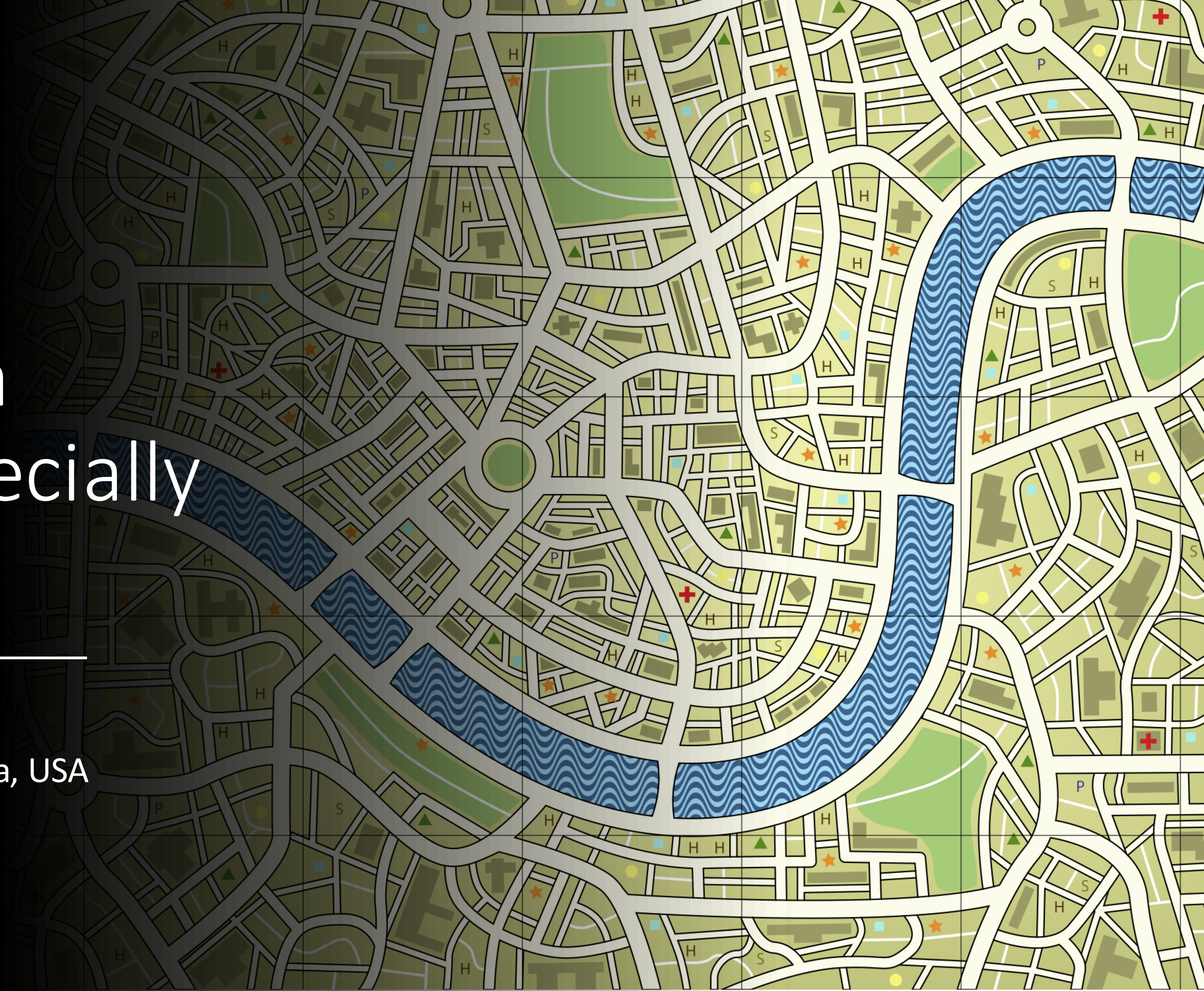
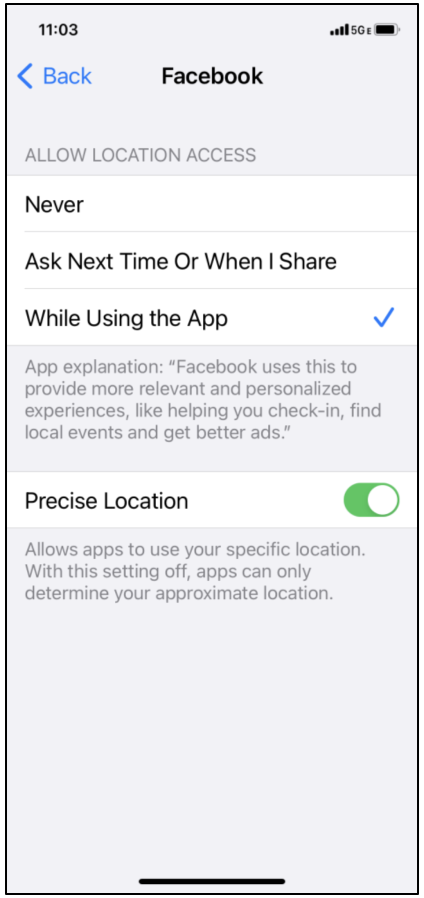
Personal Data Privacy – Especially Location

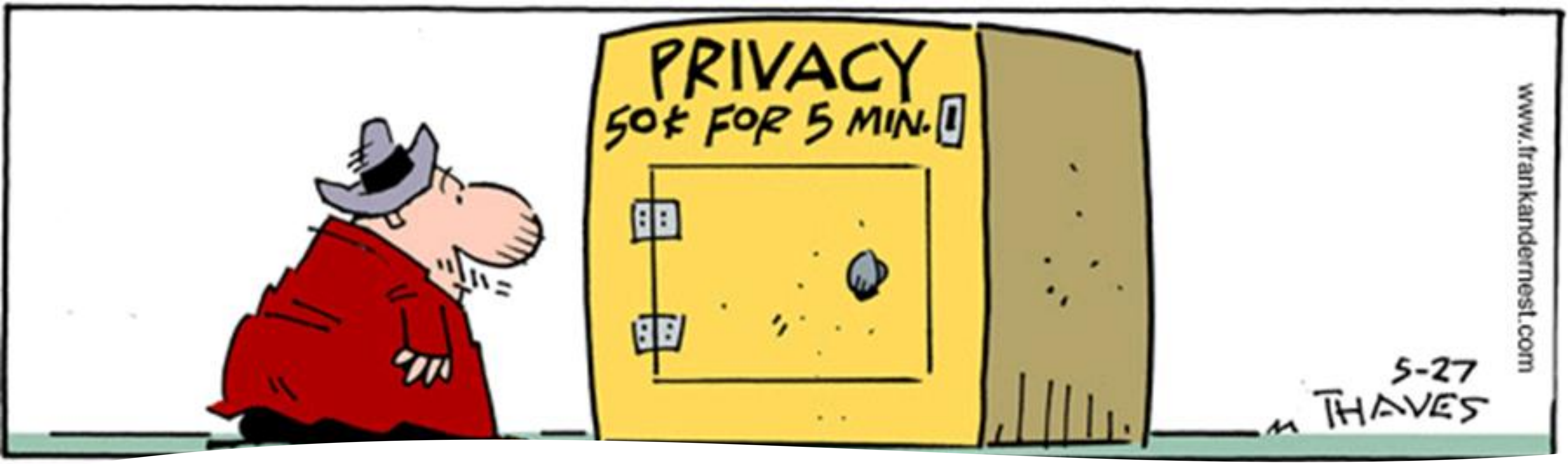
John Krumm
University of Southern California, USA



Reactions to this Phone Setting



- Your location data has value
- What could happen if I answer "yes"?
- Does Apple's differential privacy help me?
- Does it help to disallow "Precise Location"?



Outline

1. Privacy leaks
2. Can we boost peoples’ privacy sensitivity?
3. Compute value of personal data so regular people can sell it

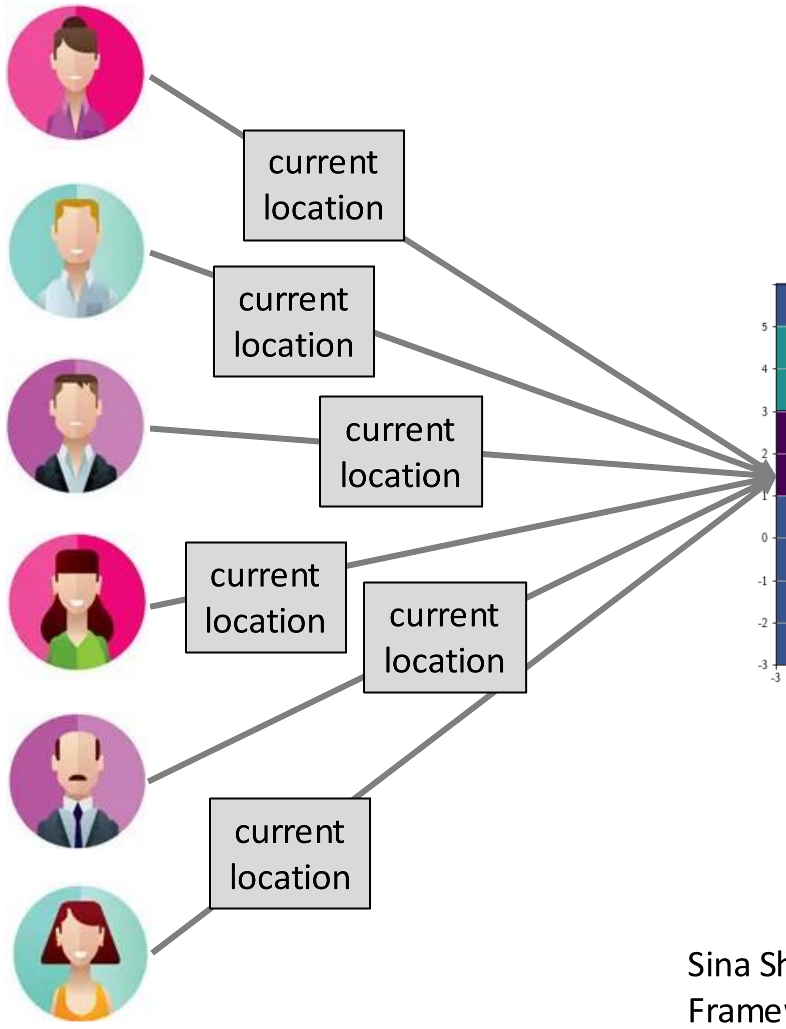
Personal Data Privacy Leaks

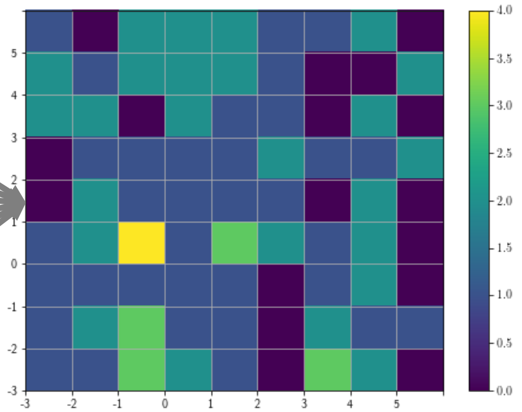
Joke

About television in North Korea.

laughter

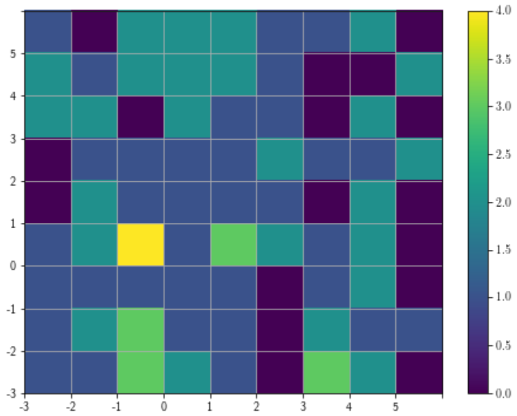
Is Differential Privacy the Answer?





exact histogram

add noise to cell counts

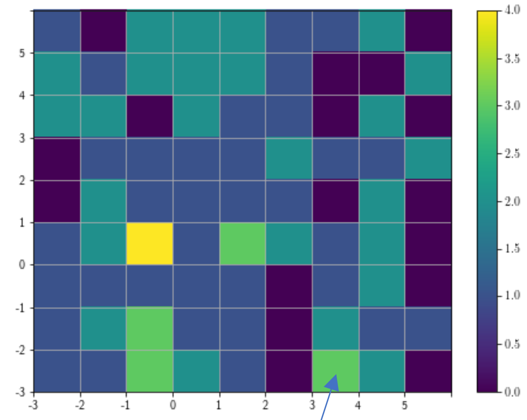


differentially private histogram to release to public

Sina Shaham, Gabriel Ghinita, Ritesh Ahuja, John Krumm, Cyrus Shahabi, “HTF: Homogeneous Tree Framework for Differentially-Private Release of Location Data”, October 2022, Transactions on Spatial Algorithms and Systems (TSAS).

Differential Privacy for Location

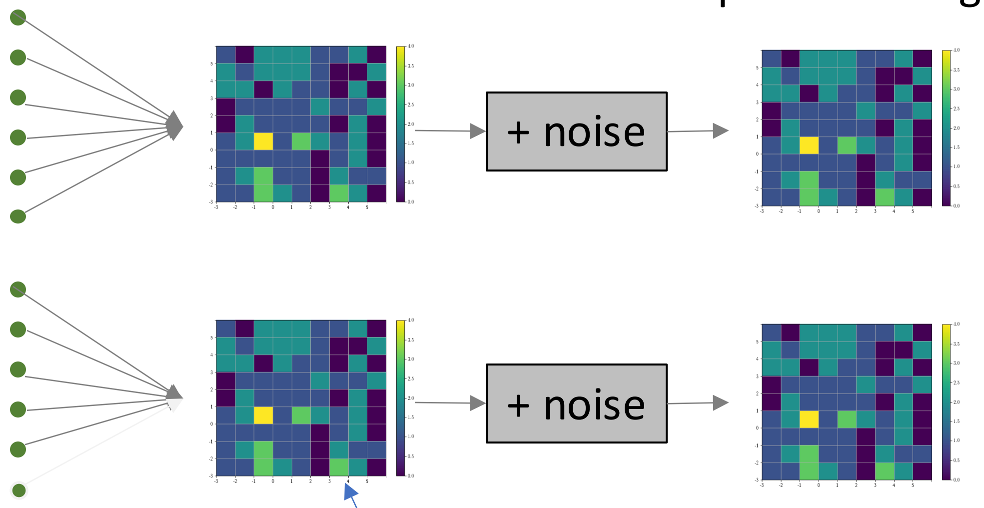
- Adding noise reduces the probability of knowing the presence of someone in the data.
- Differential privacy quantifies this and shows how much noise is necessary.



Sensitive location, e.g. casino, marijuana store, cancer clinic.

exact histogram

noisy differentially private histogram

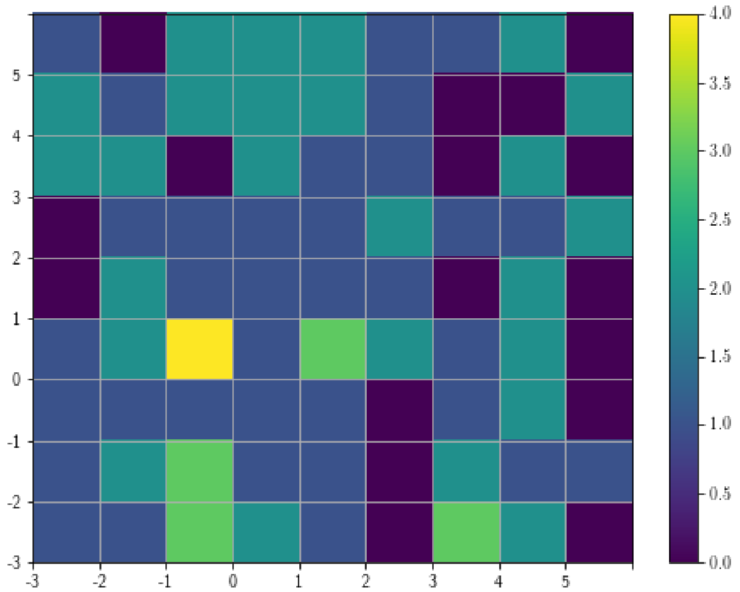


Drop this person from histogram

Count in this cell decreases by one, so I know who went there.

Differential Privacy is for Aggregating Data

DP protects your presence in aggregated data

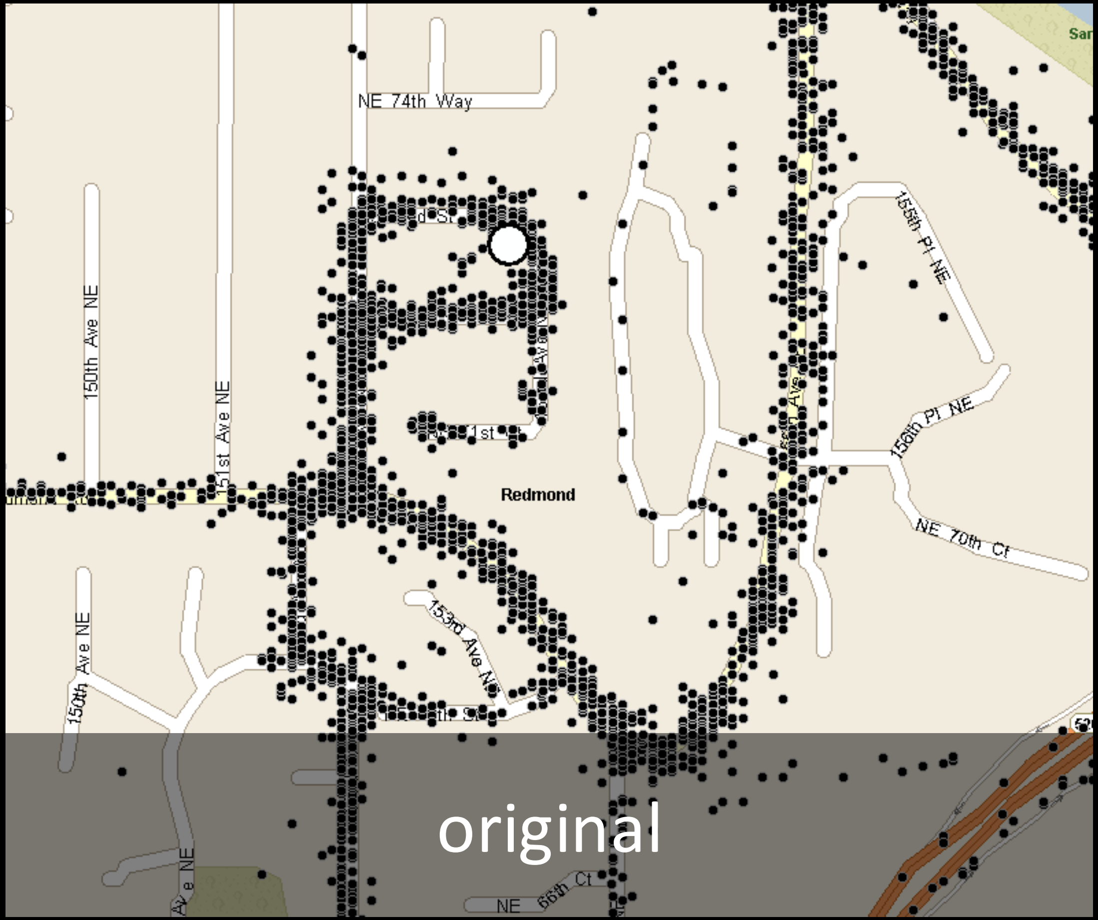


Location histogram

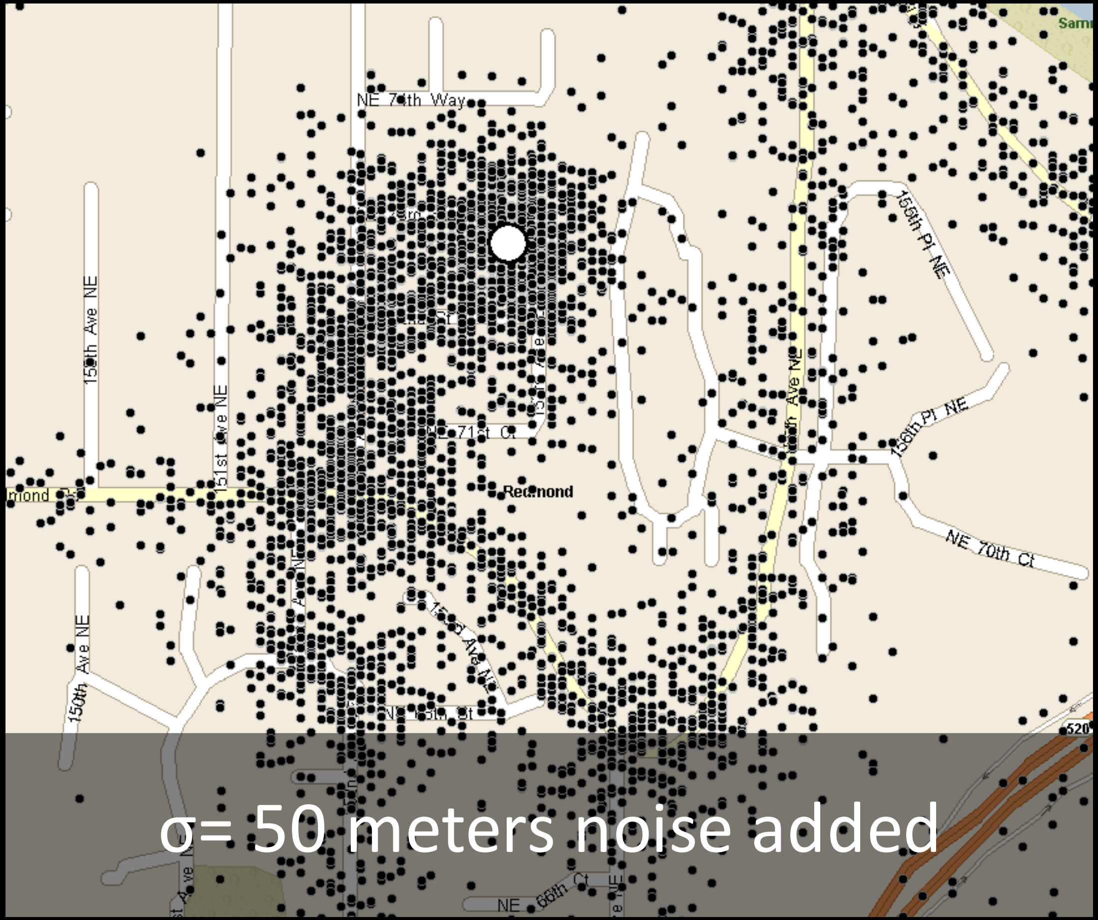
DP does not prevent the recipient from making inferences about you

“Siri, show snacks nearby.”

Can We Just Add Noise to Locations?



original



σ= 50 meters noise added

Inference Attacks on Location Tracks

Pseudonomized GPS tracks

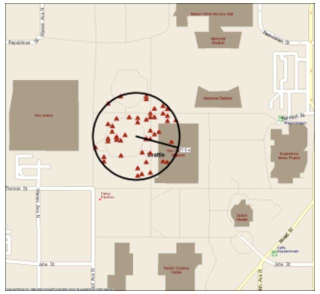
Infer home location

Reverse white pages for identity

- About 2 weeks of GPS tracks per subject
- 226 subjects
- Could correctly identify about 12% of home addresses and 5% of user names this way
- Simple algorithms

Krumm, John. "Inference attacks on location tracks." In Pervasive Computing: 5th International Conference, PERVASIVE 2007, Toronto, Canada, May 13-16, 2007. Proceedings 5, pp. 127-143. Springer Berlin Heidelberg, 2007.(10-year impact award 2017)

Location Privacy Technology: Obfuscation



Mix with nearby others' data

Delete around home

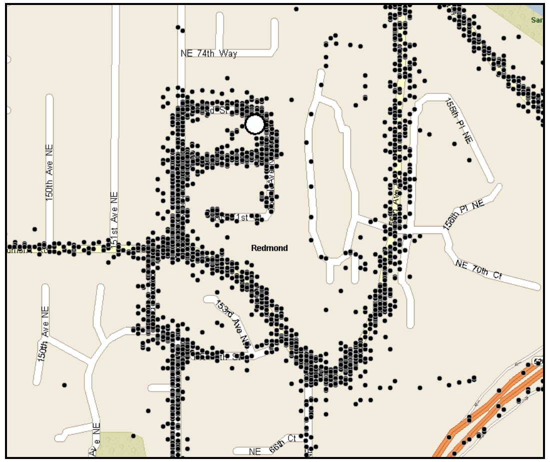
Add random noise

Discretize

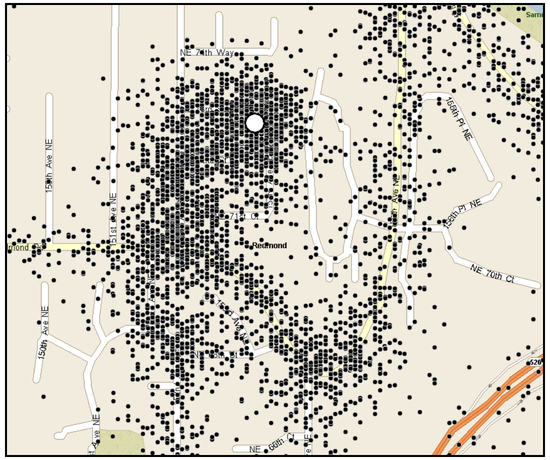
Subsample in time

This research says you have to obfuscate so much that location data becomes almost useless.

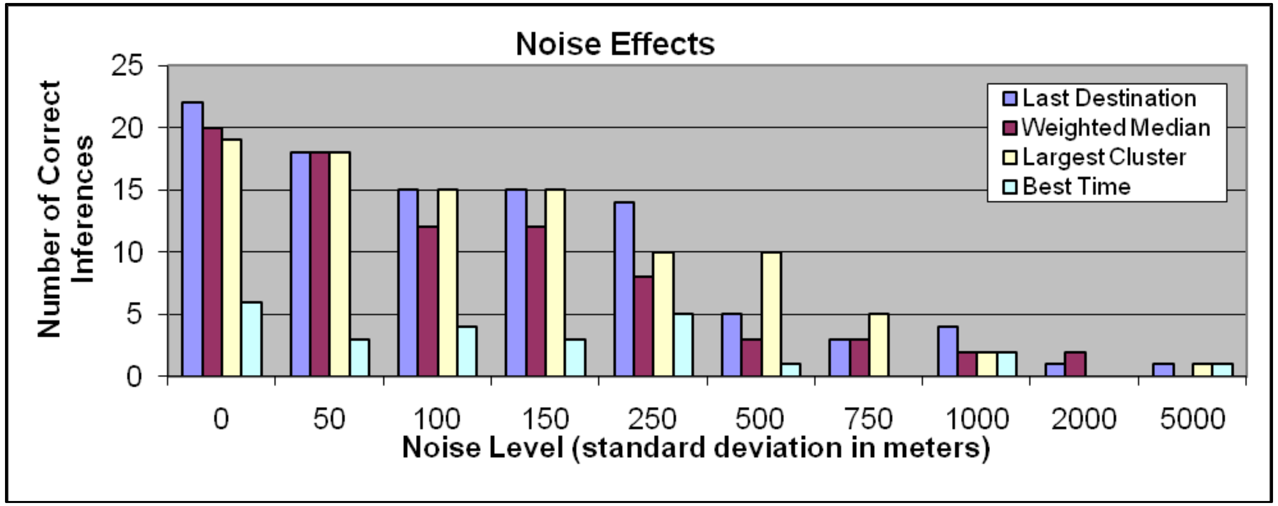
Example: Add Noise to Obfuscate



original



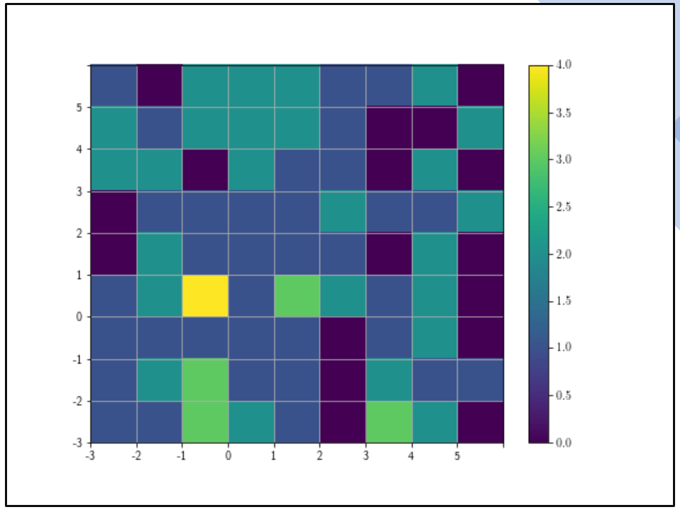
σ= 50 meters noise added

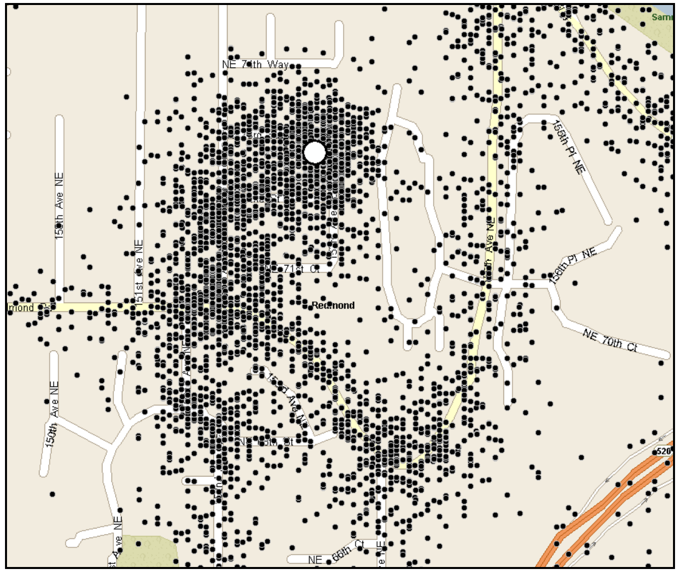


Effect of added noise on address-finding rate

Possible Solutions:

- Differential privacy – works for aggregating data, but not for protecting individual data
- Obfuscation – common methods don’t work for location data





Next: Which sensitive information can an attacker infer from your data?

What Can an Attacker Infer From Your Data?

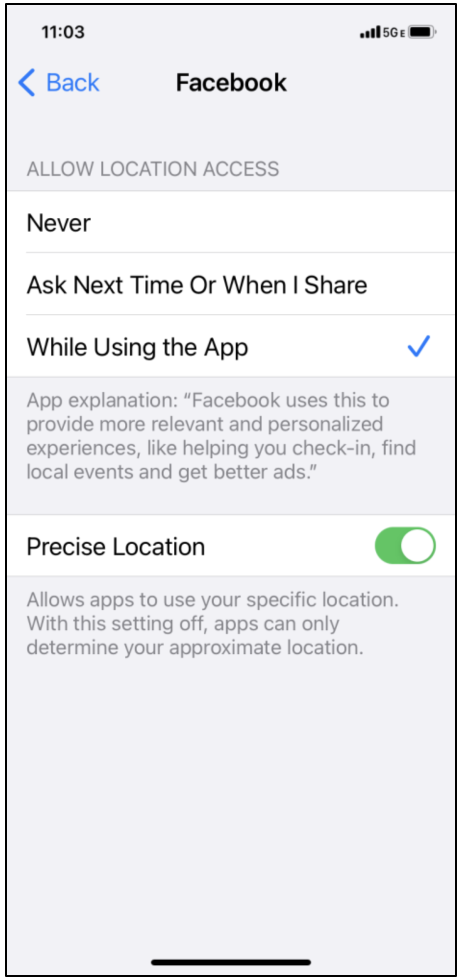
Joke

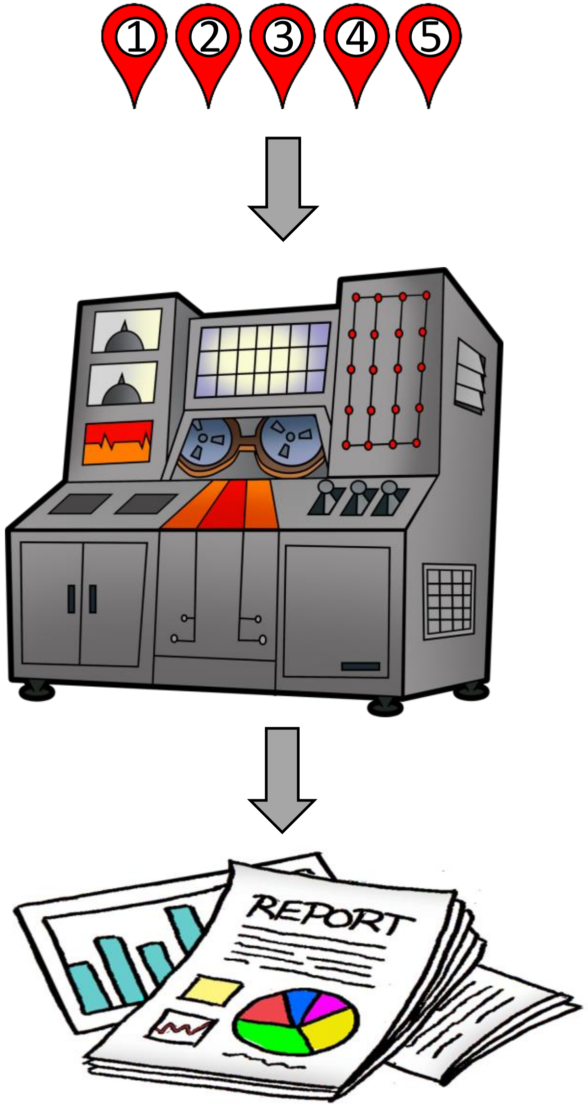
About a Roman walking into a bar.

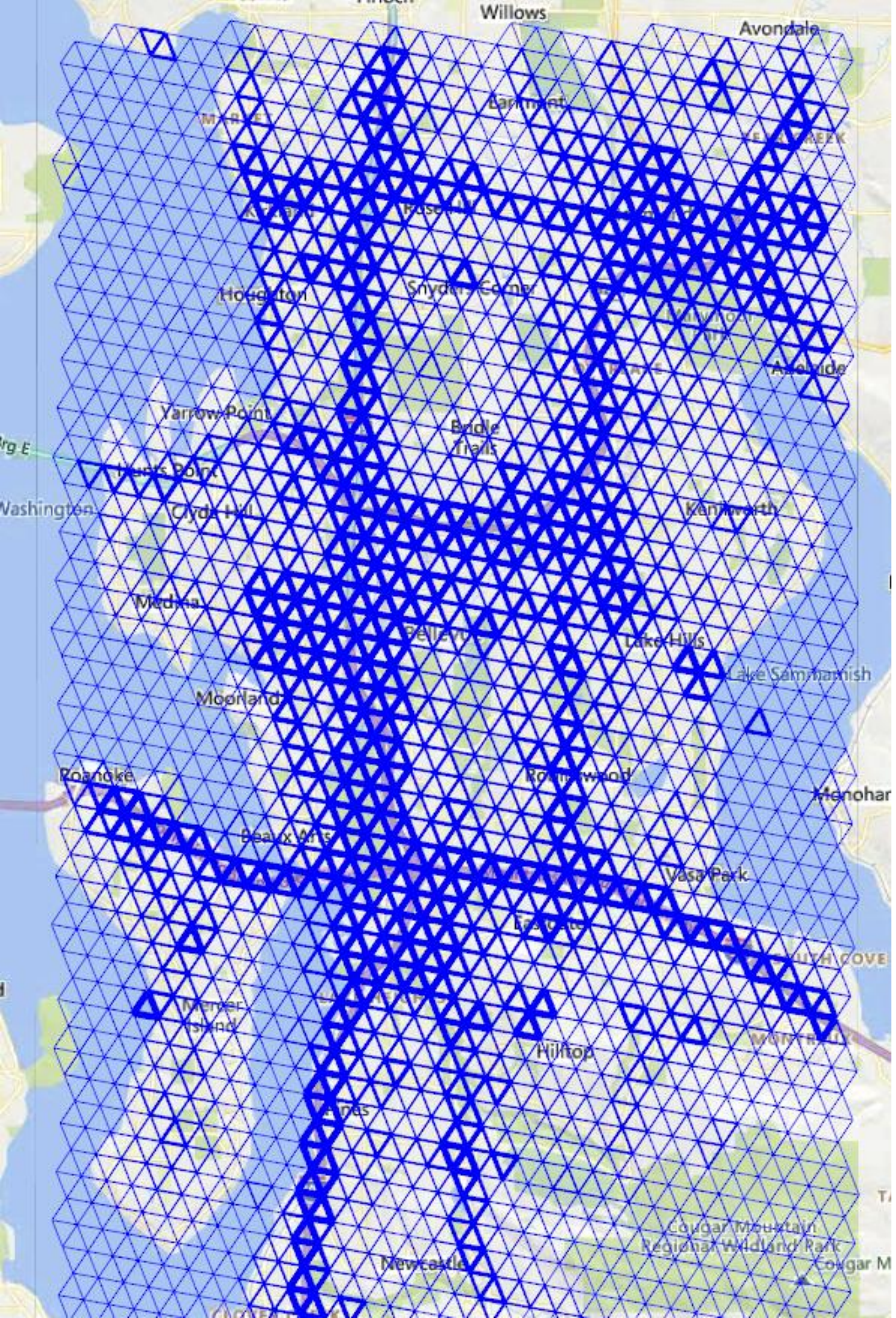
laughter

- Infer where you'll go from just a few location points
- Infer where you went between location points
- Infer your membership in a sensitive group from your web browsing

What happens after you answer yes?







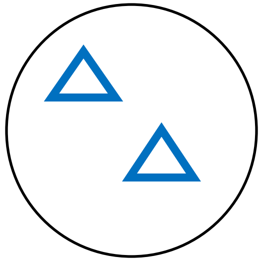
Analyze inference accuracy from small location disclosure

- 101,507 people
- 2,796,346 triangle visits
- Median 12 visits per person

What can we infer by looking at only a few (1-5) of a person’s visits?

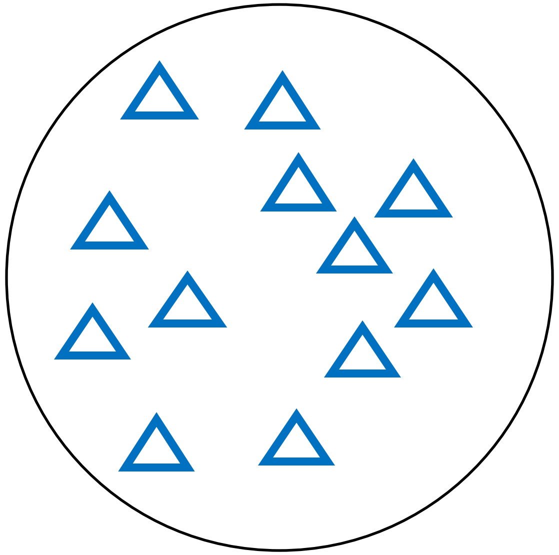
Krumm, John. "Sensitivity Analysis of Personal Location Disclosure." In 2022 23rd IEEE International Conference on Mobile Data Management (MDM), 2022. (Best paper award)

Infer other triangle visits

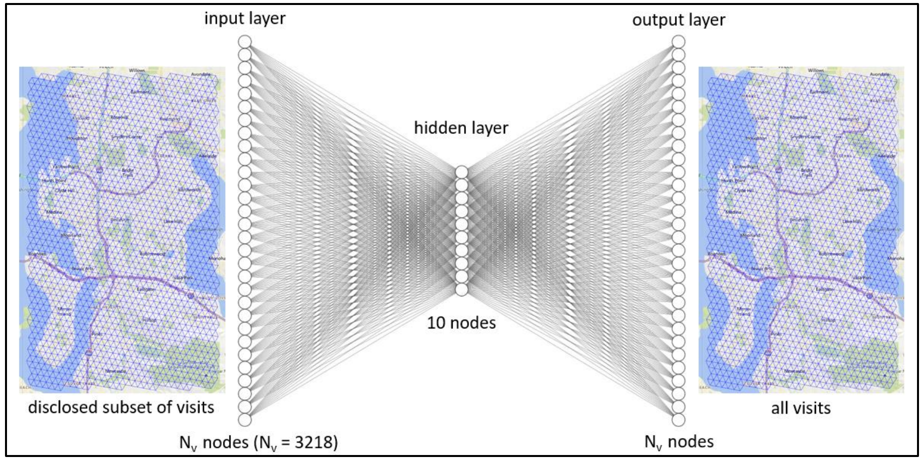


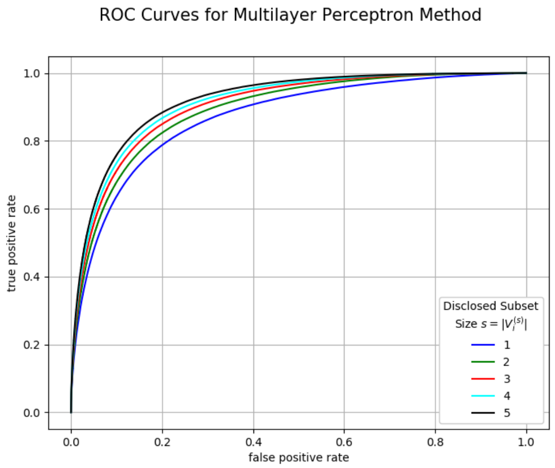
Input: 1-5 triangle visits

Simple inference algorithm



Output: all triangle visits

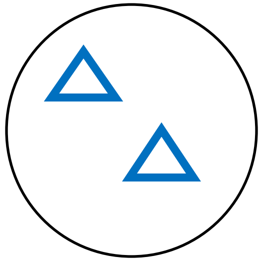




Disclosure Size	Copy	Prior	Joint	kNN	MLP
1	0.518	0.843	**0.880**	0.879	0.869
2	0.536	0.843	0.879	**0.892**	0.890
3	0.554	0.843	0.826	0.900	**0.903**
4	0.572	0.843	0.809	0.909	**0.912**
5	0.590	0.843	0.766	0.913	**0.920**

Table 1: AUC values for tested methods. The independent variable, "Disclosure Size", is s, which is the size of the disclosed subset. MLP is the neural net.

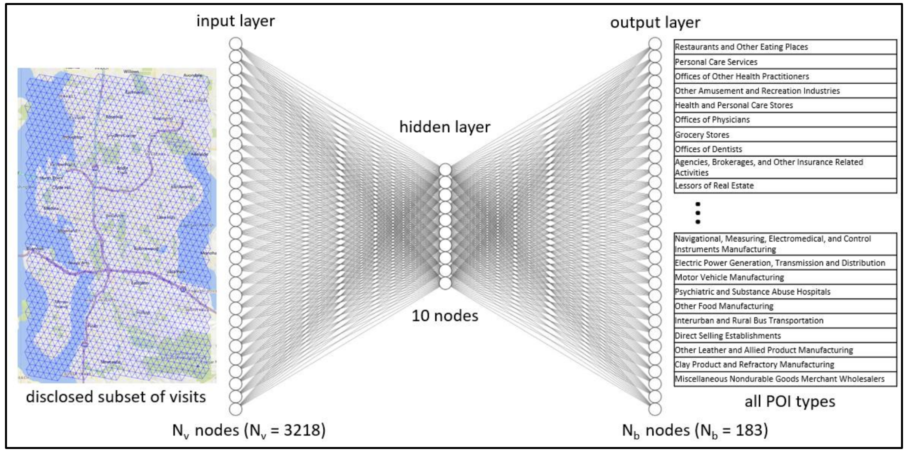
Infer preferred point-of-interest categories

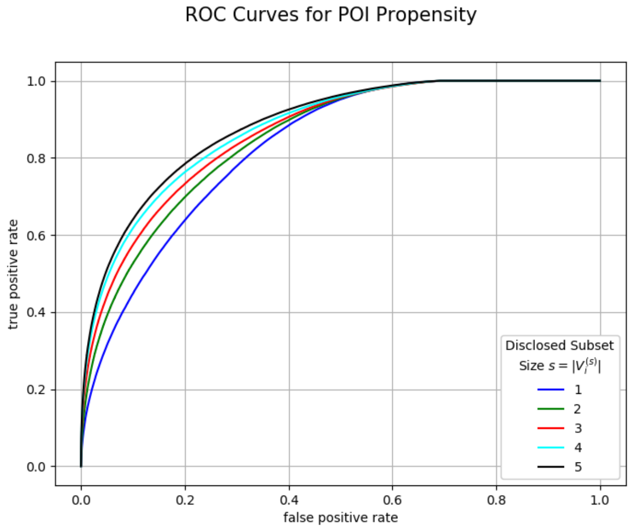


Input: 1-5 triangle visits

Simple inference algorithm

Preferred visit categories





Dis-closure Size	Inferred Location → POI AUC	Disclosure →POI AUC	Disclosure →POI FPR	Disclosure →POI TPR
1	0.740	0.828	0.295	0.773
2	0.752	0.849	0.259	0.771
3	0.630	0.863	0.242	0.780
4	0.653	0.876	0.221	0.784
5	0.809	0.884	0.205	0.790

Table 3: For inferring unusually high visits to POI categories, these are the AUC values for the method that infers POI from inferred visits (Disclosure →Location →POI) and directly from disclosed visits (Disclosure →POI). The direct method consistently outperforms. The last two columns show the best false positive rate (FPR) and true positive rate (TPR) of the direct method, all as a function of the size of the location disclosure.

Ad Delivery Decision Theory

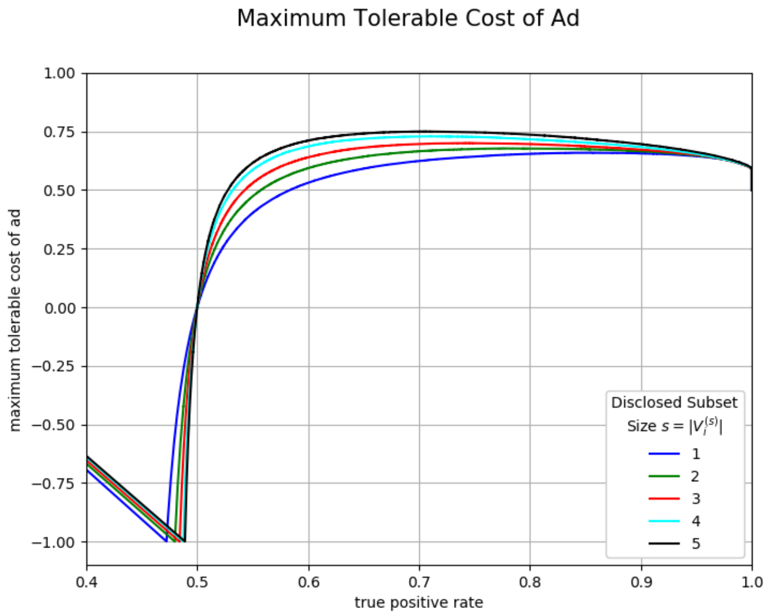
Table 4: The payoff matrix for ad delivery. The values on the right side of the equal signs are used for our model example. The cost of buying an ad is α.

		POI Category Propensity	
		no propensity	high propensity
Ad	do not deliver	$b_{11} = 0$	$b_{12} = -1.0 + \alpha$
	deliver	$b_{21} = -\alpha$	$b_{22} = 1.0 - \alpha$

- Payoff matrix for delivering an ad for a certain POI category
- Cost of buying ad is α

$$\mathbb{E}[P] = b_{11} \cdot \text{TNR} + b_{12} \cdot \text{FNR} + b_{21} \cdot \text{FPR} + b_{22} \cdot \text{TPR} \qquad (6)$$

Expected payoff based on true positive rate (TPR), etc. of preferred POI classifier



Summary of Sensitivity Analysis

- Even a small location disclosure can reveal future visits (both location and category)
- Quantify how inferences get better with more data
- Direct path to making ad delivery decisions

Experiment to see what can be inferred from small location disclosure

Interpolating Where You Went

Maximum Entropy Bridgelets for Trajectory Completion

Krumm, John. "Maximum entropy bridgelets for trajectory completion." In Proceedings of the 30th International Conference on Advances in Geographic Information Systems, 2022. (Best paper award)

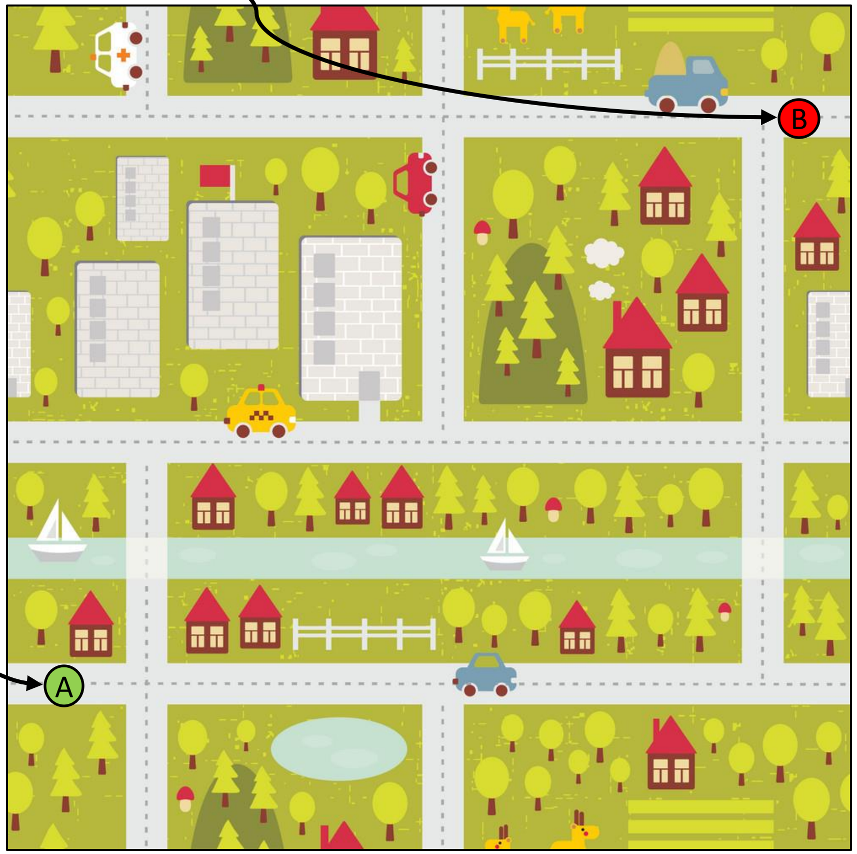
Joke

About the hamburger walking into a bar.

laughter

Trajectory Completion

Where was the entity between points A and B?



Applications

- Did the person see an advertisement?
- Was the person exposed to a disease?
- Did the person witness an event?
- Did the person commit a crime?
- What are privacy losses? If you have a long time between location samples, what can still be inferred about you?

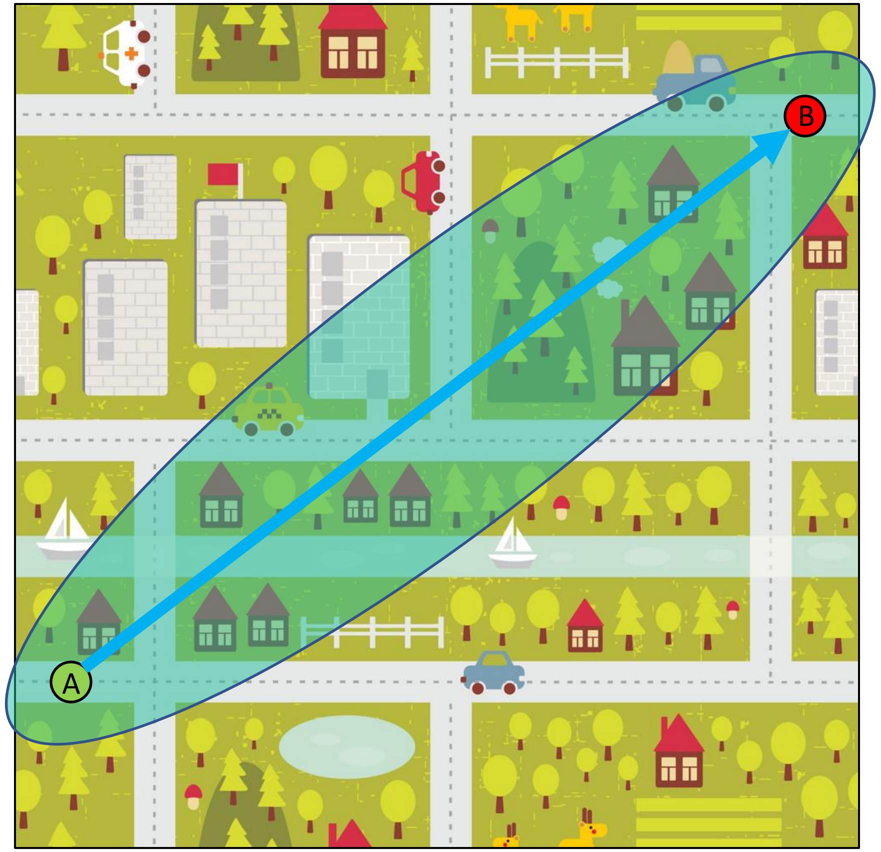
How To Do It Better

How we do this now:

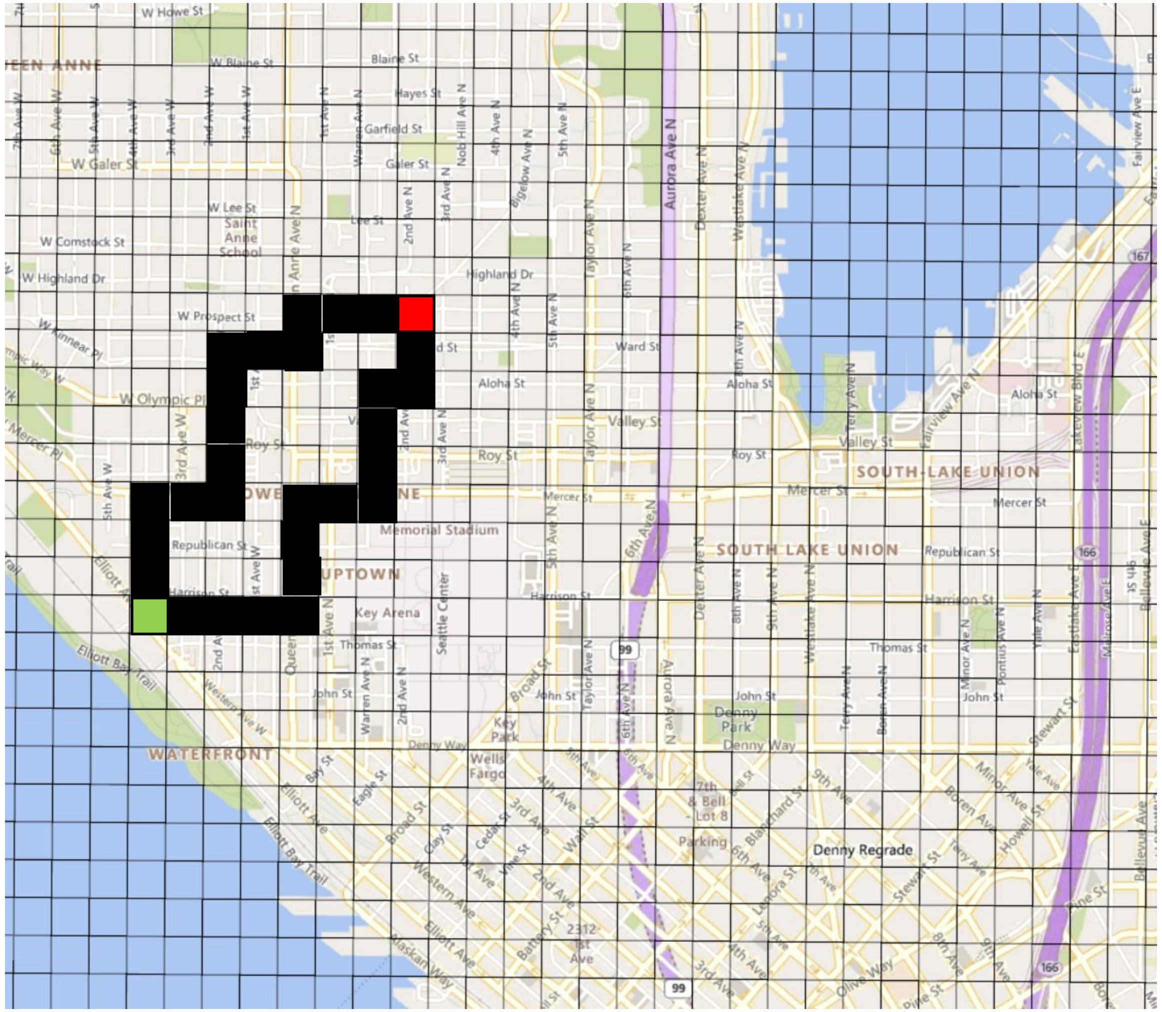
- Linear interpolation
- Space-time prism
- Gaussian process
- Brownian bridge

How we want to do this:

- Probabilistic – represent uncertainty
- Learn from trajectory data
- Minimal assumptions about movement



Bridgelet Gives All the “Walks” on a Grid



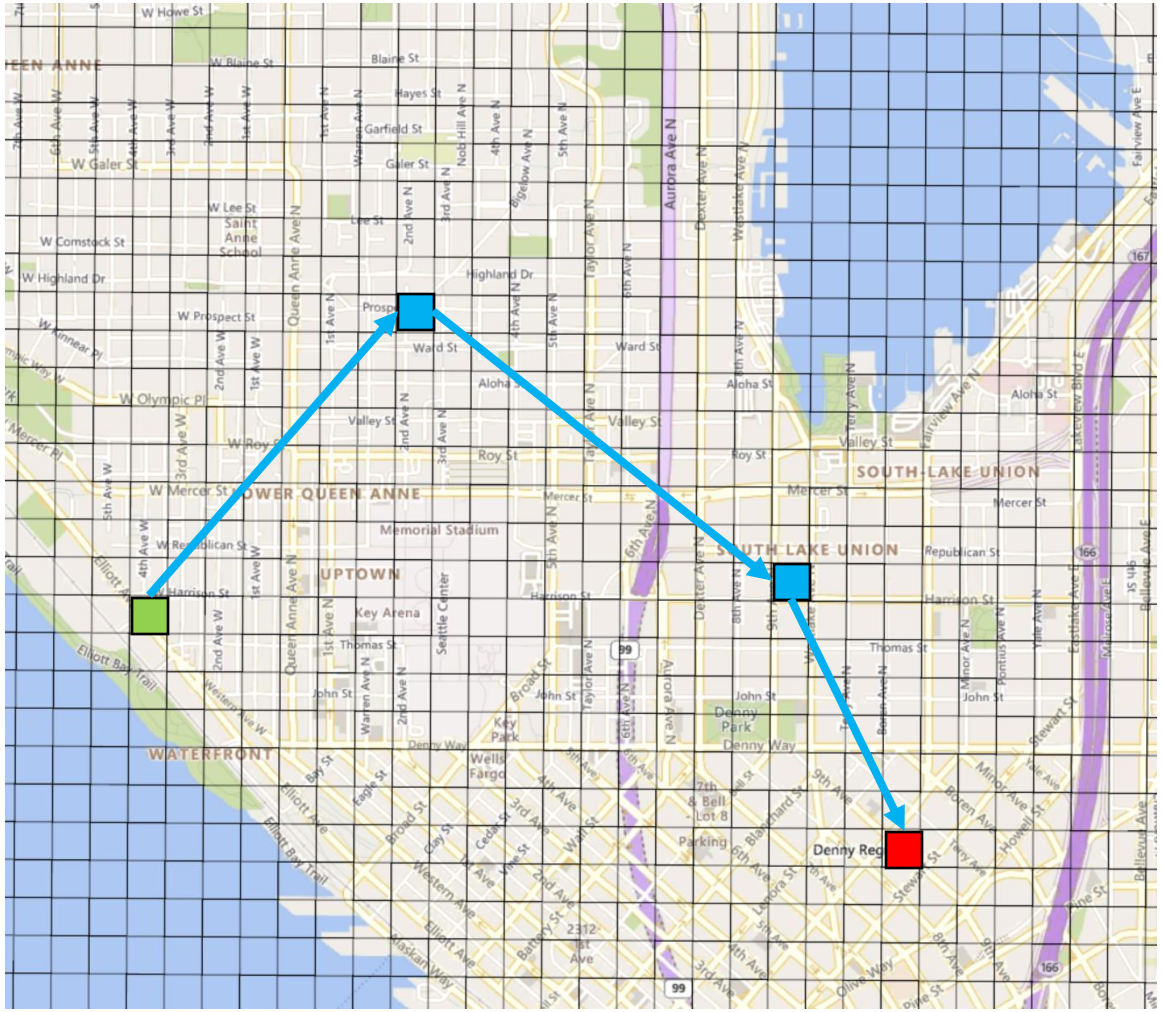
A bridgelet is the set of all possible walks on a grid between two points.

For each cell in bridgelet, compute

- Visit probability
- Dwell time probability

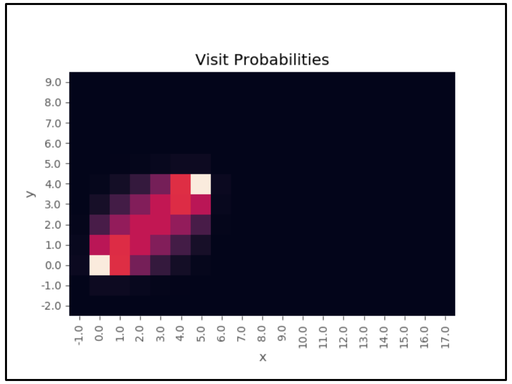
This gives a proper accounting of our uncertainty about where the person went.

Bridgelets Between Cells in Trajectory

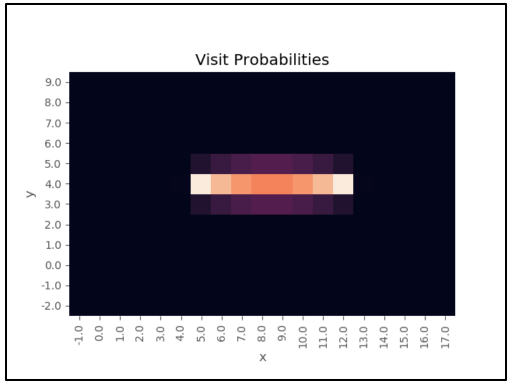


Bridgelets combine to make a bridge

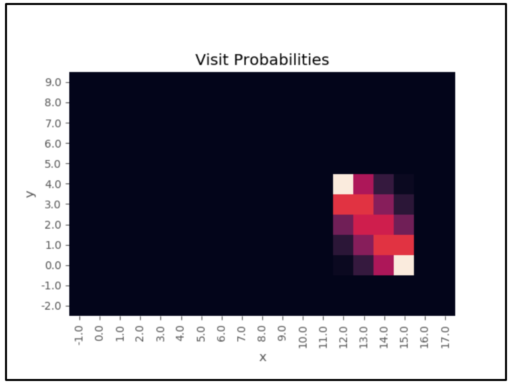
Combine Bridgelets into Bridge



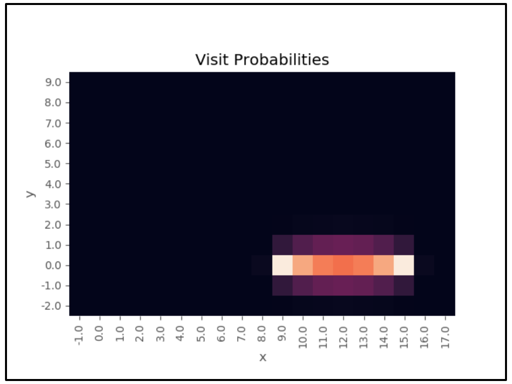
$W_{X,Y,T} = W_{5,4,12}$



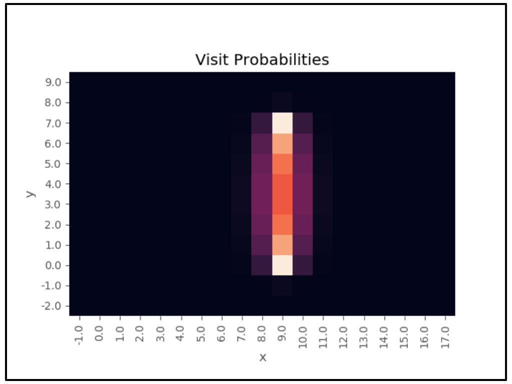
$W_{X,Y,T} = W_{7,0,10}$



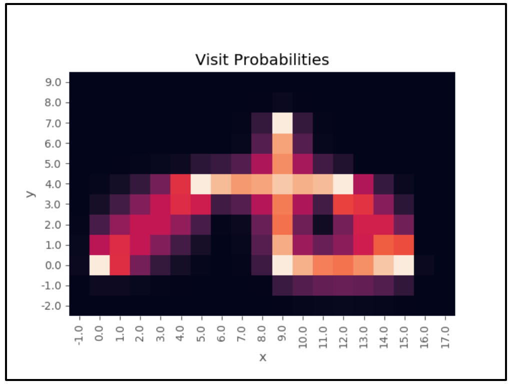
$W_{X,Y,T} = W_{3,-4,8}$



$W_{X,Y,T} = W_{-6,0,10}$



$W_{X,Y,T} = W_{0,7,12}$

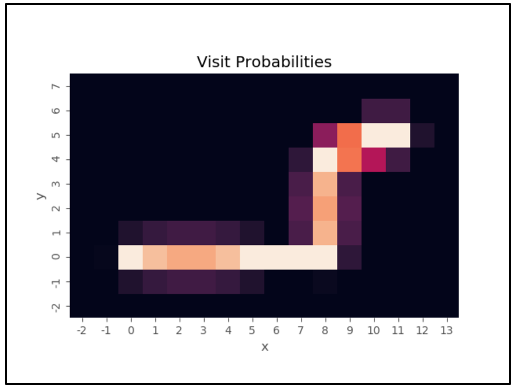


$$P(v|W_1 W_2 \ldots W_{N-1}) = 1 - \prod_{i=1}^{N-1} \left(1 - P(v|W_i)\right) \qquad (3)$$

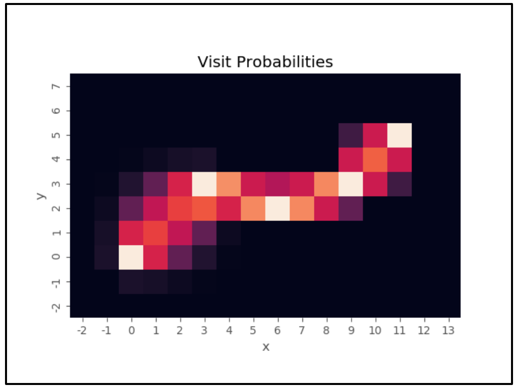
Multiple Bridges from Different Trajectories

- Each observed trajectory makes a bridge
- Combine into one aggregate bridge
- Represents peoples' preferences for travel between two distant cells

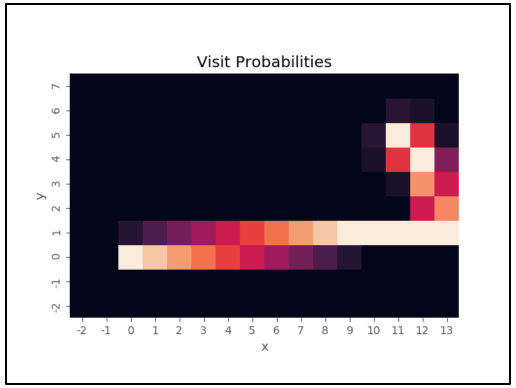
Combine Bridges from Trajectory Data



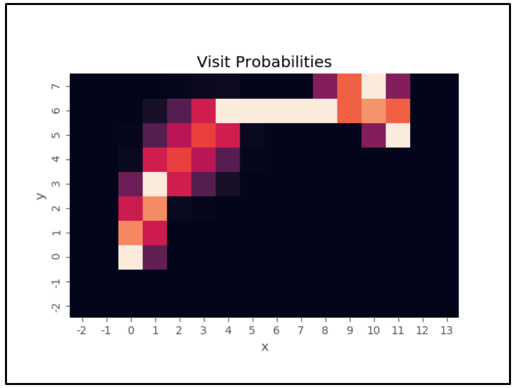
trajectory 1



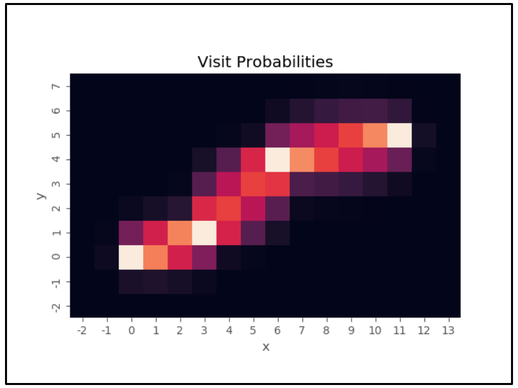
trajectory 2



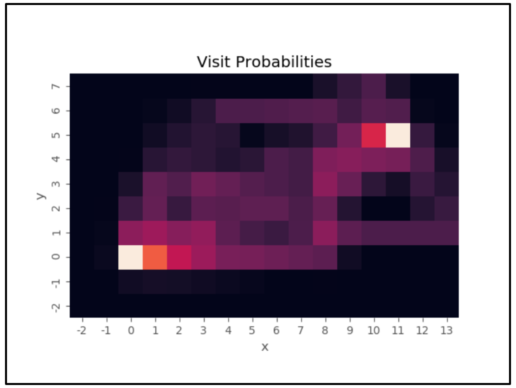
trajectory 3



trajectory 4



trajectory 5



final bridge

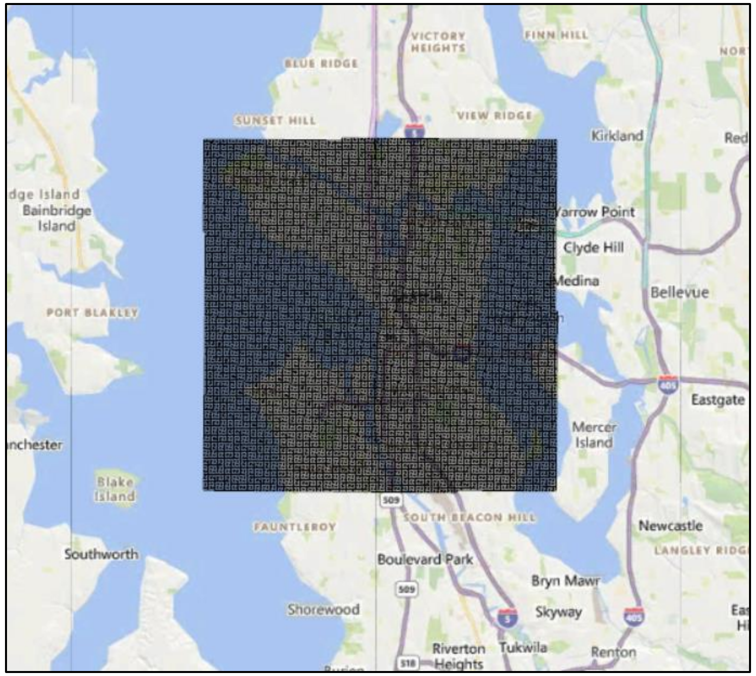
Experiments

Safegraph trajectory data

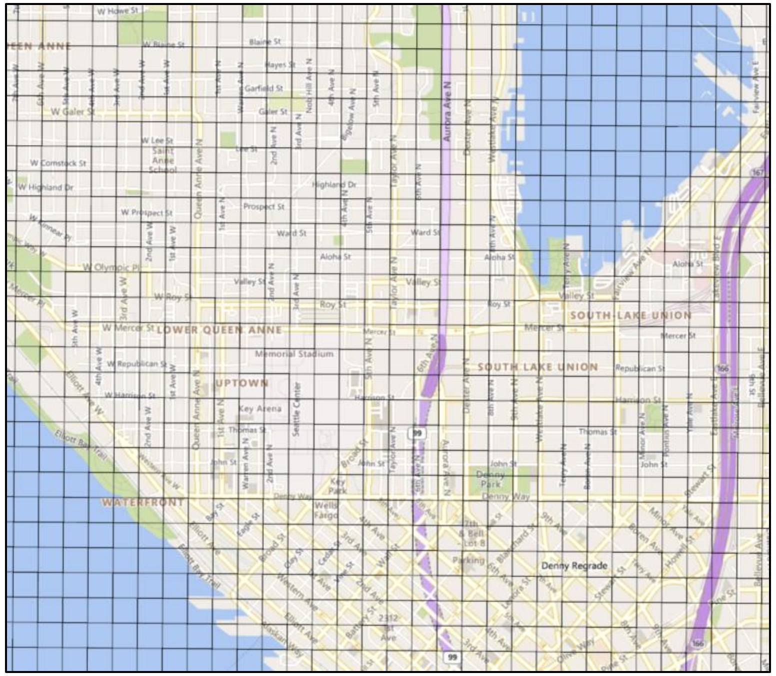
- First week of April 2022
- 1.7M distinct users
- 79.5M trajectories

Bridges

- 100m X 100m grid
- ΔT = 5s
- Bridgelets T_{max} = 15
- Bridgelets ≤ 15 x 5s = 75s



Trajectories from Seattle

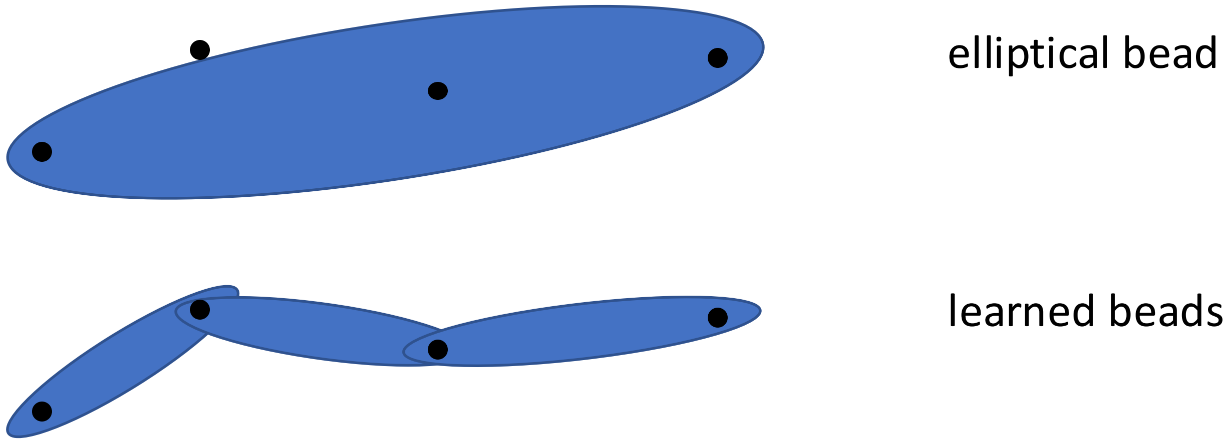


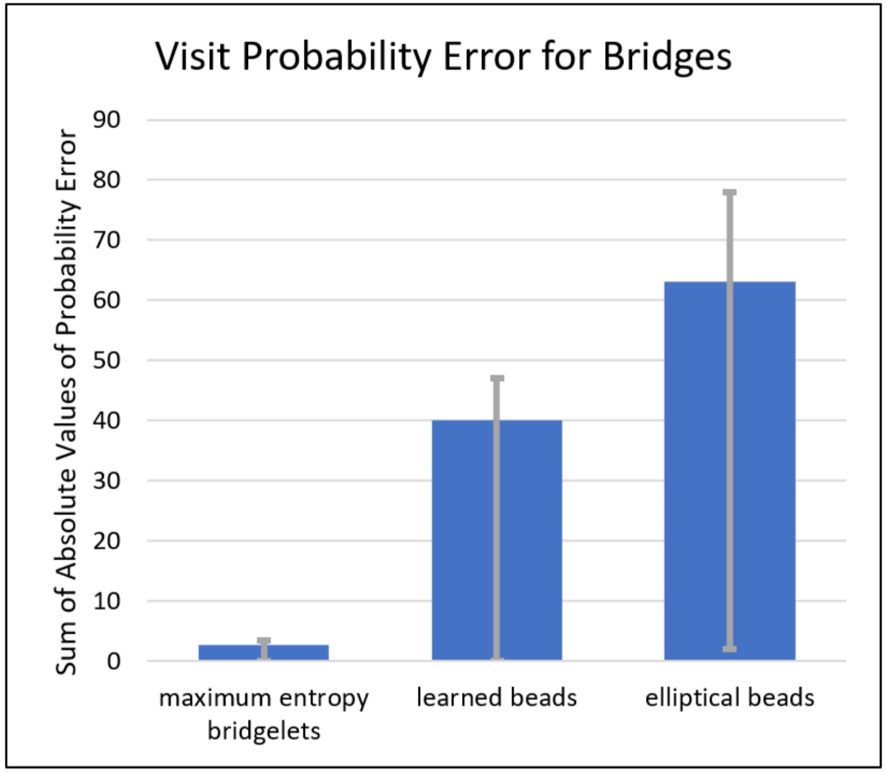
100m x 100m Military Grid Reference System (MGRS)

Experiments

Accuracy Test

- 75K trajectories
- Baseline: two versions of traditional beads
- Sum of absolute visit probability error over all trajectories





Implication: Large improvement in inferring where someone goes between location measurements

Infer Sensitive Group Membership from Online Advertising Profiles

With Kyle Crichton (Carnegie Mellon University) and Sid Suri (Microsoft Research)

The New Yorker Magazine (1993)

Joke

About the airplane with engine problems

laughter

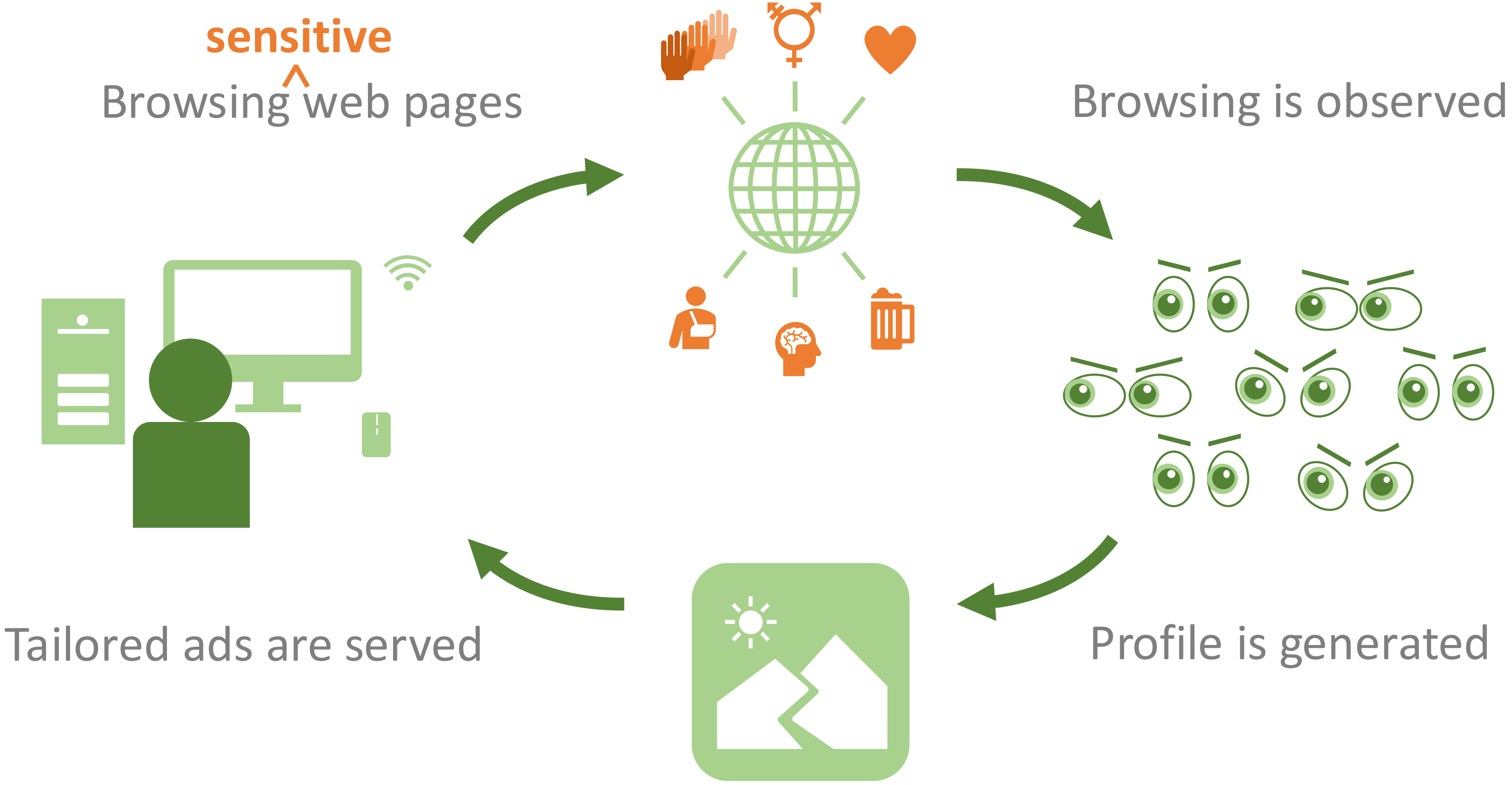
Privacy harm arising from web tracking

- Predatory and discriminatory advertisements on the basis of…
  - Race[1,2]
  - Gender[1,3]
  - Sexual orientation[4]
  - History of substance abuse[5]

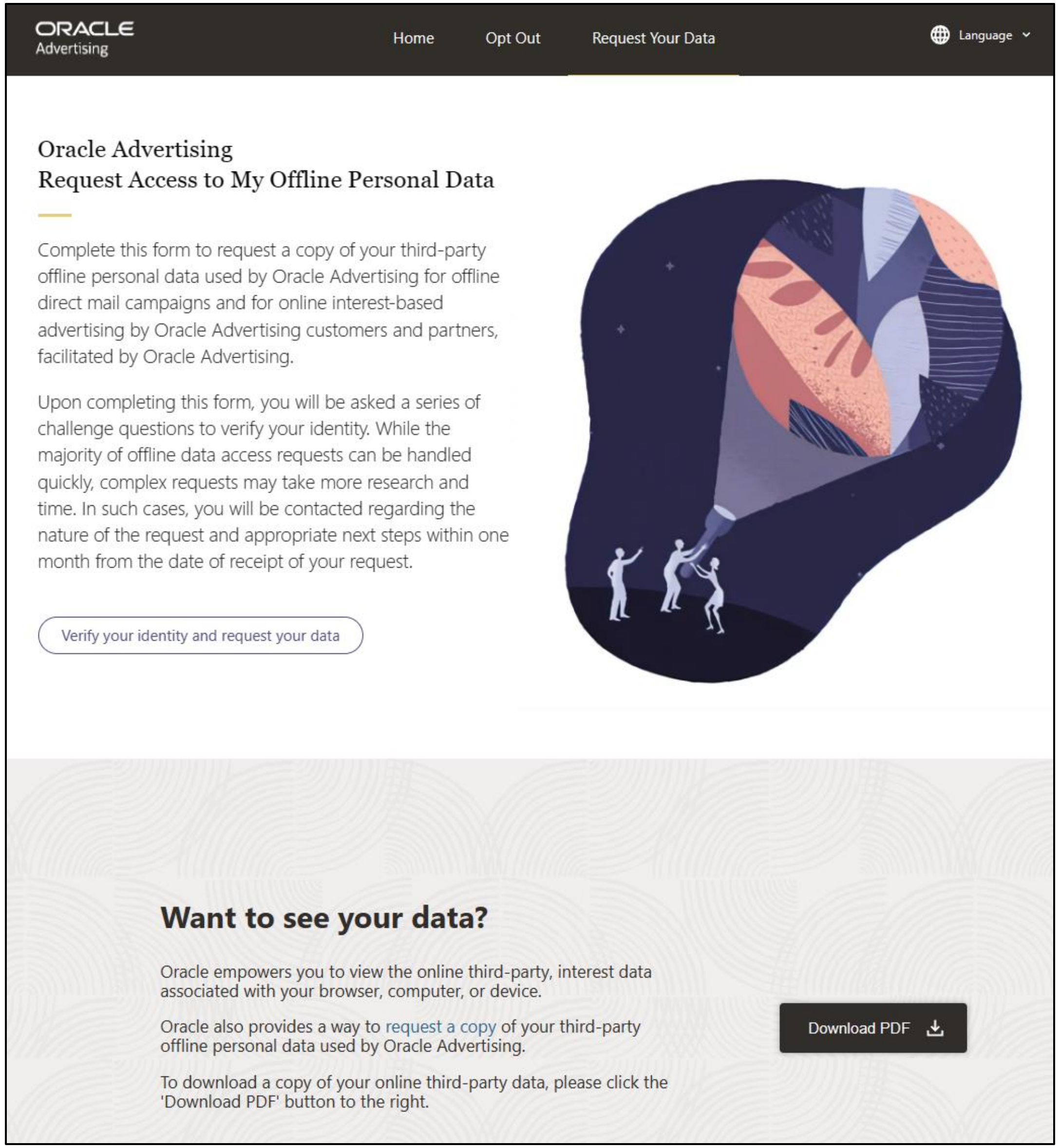
[1] Muhammad Ali et al. Discrimination through optimization: How facebook's ad delivery can lead to skewed outcomes. CoRR, abs/1904.02095, 2019.
[2] Latanya Sweeney. Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. Mar 2013.
[3] Anja Lambrecht and Catherine Tucker. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. Management Science, 65(7):2966–2981, 2019.
[4] Craig E. Wills and Can Tatar. Understanding what they do with what they know. WPES '12, 2012.
[5] Amit Datta et al. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. CoRR, abs/1408.6491, 2014.

Online Advertising

Online Advertising



Download Your Advertising Profile



https://datacloudoptout.oracle.com/request-your-data

Profile labels are relatively benign

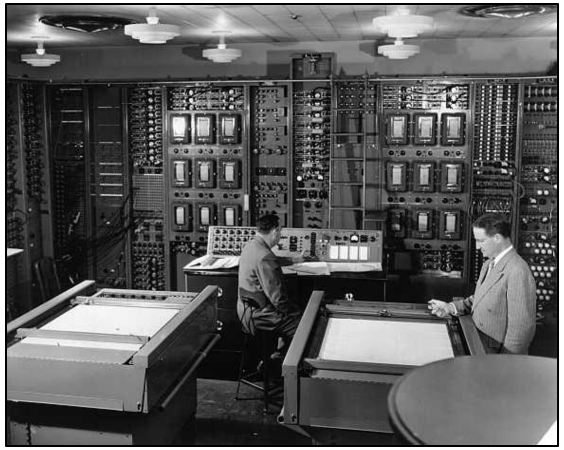
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Exercise
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Exercise > Gyms
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Exercise > Running and Jogging
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Wellness
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Wellness > Dieting and Weight Loss
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Arts and Crafts
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Collecting
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Games
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Games > Board Games
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Games > Card Games and Trading Cards (CCG)
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Reading

Scenario

Browsing sites including drug addition recovery (from 0% to 100%)

Advertising profile (innocuous categories)

exercise	TRUE
Italy	FALSE
board games	FALSE
Toyotas	TRUE
baseball	TRUE
…	
exotic birds	FALSE



AI/machine learning

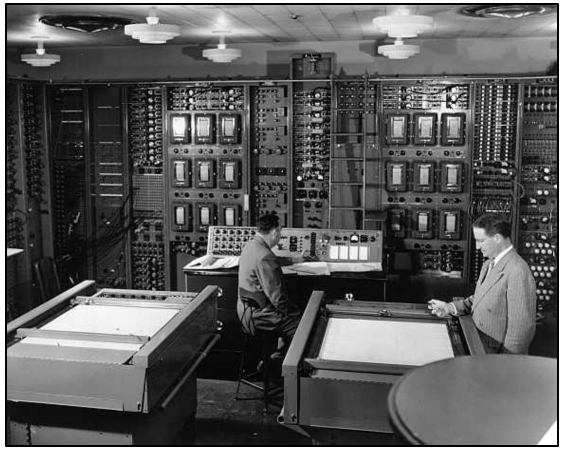
Probability of drug addiction recovery

TABLE 2. DECISION TREE CLASSIFIER RESULTS

	Group	Precision	Recall	F1 Score
	Baseline	0.270	0.732	0.388
Privacy-Sensitive	Child	0.900	0.886	0.890
	Financial Distress	0.908	0.838	0.868
	Asian	0.866	0.848	0.854
	Grief	0.898	0.814	0.854
	Deaf	0.906	0.804	0.852
	Smoking Recovery	0.870	0.836	0.850
	Physical Disability	0.850	0.836	0.842
	Hispanic	0.866	0.796	0.832
	Immigrant	0.870	0.792	0.828
	Heart Disease	0.822	0.806	0.814
	Poverty	0.816	0.822	0.814
	Divorce	0.844	0.786	0.812
	Black	0.842	0.782	0.810
	Right	0.830	0.790	0.808
	Foot Fetish	0.852	0.766	0.806
	Senior	0.834	0.786	0.806
	Drug Recovery	0.822	0.796	0.802
	Sexual Assault	0.862	0.744	0.798
	BDSM	0.860	0.730	0.790
	PTSD	0.830	0.744	0.784
	Transgender	0.806	0.762	0.782
	Domestic Violence	0.804	0.752	0.778
	Furry	0.822	0.734	0.776
	Male	0.766	0.786	0.776
	HIV	0.764	0.786	0.774
	Cancer	0.752	0.792	0.770
	Pregnancy	0.748	0.776	0.760
	Abortion	0.726	0.798	0.758
	Depression	0.780	0.726	0.752
	Left	0.794	0.720	0.752
	Suicide	0.742	0.724	0.732
	Female	0.710	0.752	0.730
	Gay	0.734	0.724	0.728
	Lesbian	0.724	0.716	0.722
	Diabetes	0.728	0.700	0.714
	Alcohol Recovery	0.686	0.686	0.680
	Porn Recovery	0.676	0.676	0.674
Echo Chamber	Climate Denial	0.904	0.860	0.880
	Smoking Addiction	0.866	0.834	0.850
	Alcohol Addiction	0.820	0.854	0.838
	Far Left	0.854	0.802	0.826
	Anti-vax	0.854	0.780	0.814
	White Supremacy	0.810	0.774	0.792
	Election Fraud	0.796	0.760	0.778
	Far Right	0.822	0.722	0.764
	Islamic Extremism	0.774	0.750	0.760
	Porn Addiction	0.626	0.698	0.648
	Mean	**0.798**	**0.774**	**0.781**
	Standard Deviation	**0.100**	**0.048**	**0.078**

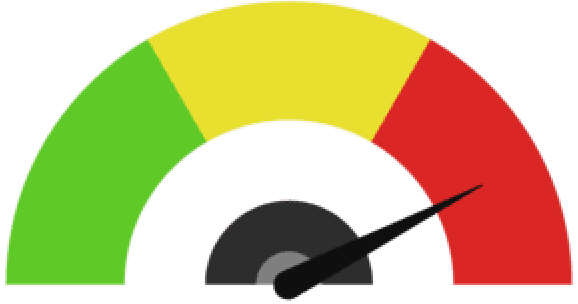
47 sensitive groups + 1 baseline group

Simulated 24,000 people browsing across all groups



Decision tree classifier

Skipping details:
- How choose sites to browse
- Mix of sensitive + regular sites
- Variety of IP addresses



77% classification accuracy into sensitive group

Implications

Even with innocuous advertising labels, advertisers can infer membership in sensitive groups from web browsing.

- Sensitive characteristic(s) could be exposed
- Predatory, dangerous advertising could result

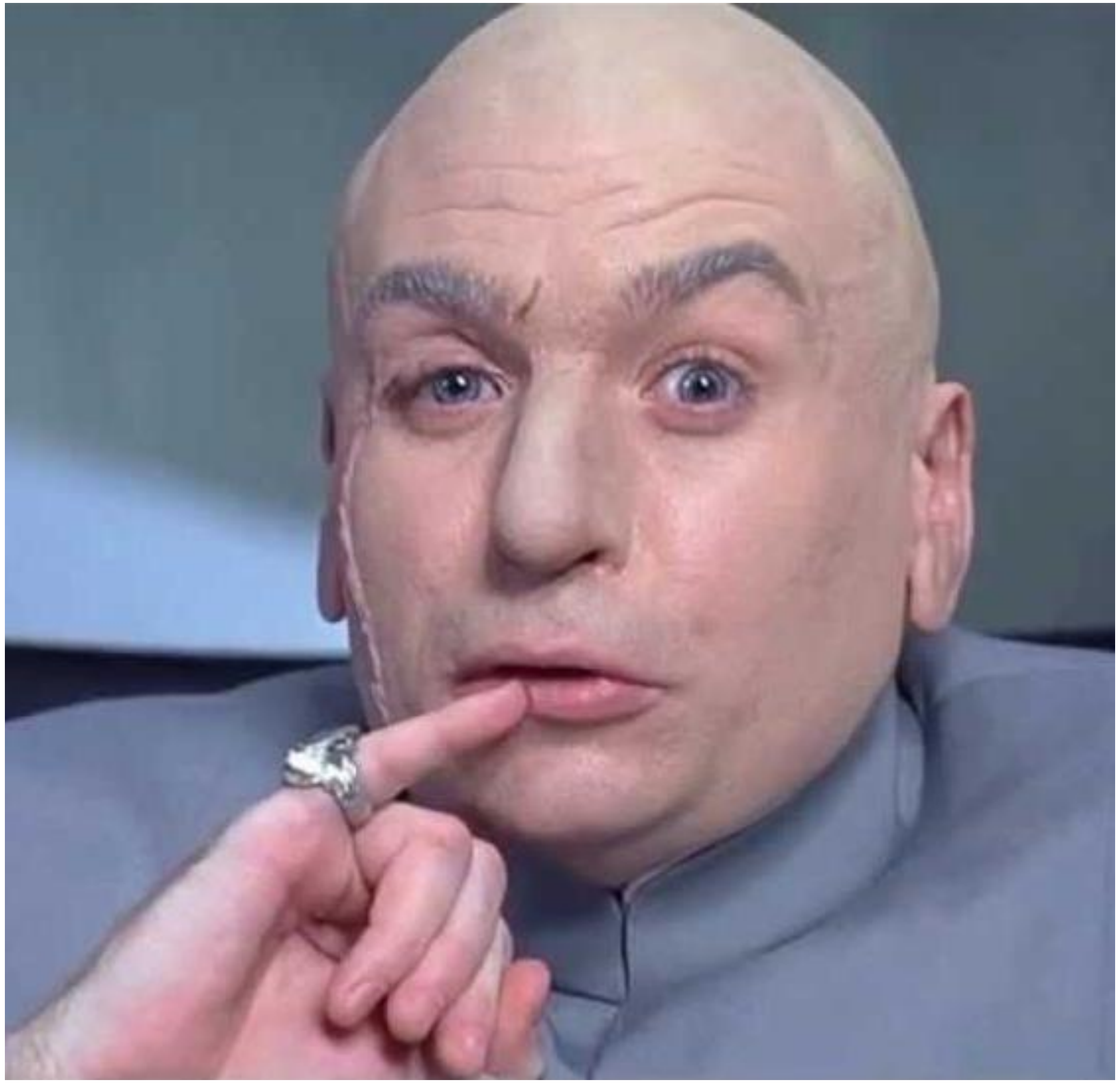
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Exercise
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Exercise > Gyms
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Exercise > Running and Jogging
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Wellness
Audiences by Oracle > Hobbies and Interests (Affinity) > Health and Fitness > Wellness > Dieting and Weight Loss
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Arts and Crafts
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Collecting
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Games
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Games > Board Games
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Games > Card Games and Trading Cards (CCG)
Audiences by Oracle > Hobbies and Interests (Affinity) > Hobbies > Reading

TABLE 2. DECISION TREE CLASSIFIER RESULTS

	Group	Precision	Recall	F1 Score
	Baseline	0.270	0.732	0.388
Privacy-Sensitive	Child	0.900	0.886	0.890
	Financial Distress	0.908	0.838	0.868
	Asian	0.866	0.848	0.854
	Grief	0.898	0.814	0.854
	Deaf	0.906	0.804	0.852
	Smoking Recovery	0.870	0.836	0.850
	Physical Disability	0.850	0.836	0.842
	Hispanic	0.866	0.796	0.832
	Immigrant	0.870	0.792	0.828
	Heart Disease	0.822	0.806	0.814
	Poverty	0.816	0.822	0.814
	Divorce	0.844	0.786	0.812
	Black	0.842	0.782	0.810
	Right	0.830	0.790	0.808
	Foot Fetish	0.852	0.766	0.806
	Senior	0.834	0.786	0.806
	Drug Recovery	0.822	0.796	0.802
	Sexual Assault	0.862	0.744	0.798
	BDSM	0.860	0.730	0.790
	PTSD	0.830	0.744	0.784
	Transgender	0.806	0.762	0.782
	Domestic Violence	0.804	0.752	0.778
	Furry	0.822	0.734	0.776
	Male	0.766	0.786	0.776
	HIV	0.764	0.786	0.774
	Cancer	0.752	0.792	0.770
	Pregnancy	0.748	0.776	0.760
	Abortion	0.726	0.798	0.758
	Depression	0.780	0.726	0.752
	Left	0.794	0.720	0.752
	Suicide	0.742	0.724	0.732
	Female	0.710	0.752	0.730
	Gay	0.734	0.724	0.728
	Lesbian	0.724	0.716	0.722
	Diabetes	0.728	0.700	0.714
	Alcohol Recovery	0.686	0.686	0.680
	Porn Recovery	0.676	0.676	0.674
Echo Chamber	Climate Denial	0.904	0.860	0.880
	Smoking Addiction	0.866	0.834	0.850
	Alcohol Addiction	0.820	0.854	0.838
	Far Left	0.854	0.802	0.826
	Anti-vax	0.854	0.780	0.814
	White Supremacy	0.810	0.774	0.792
	Election Fraud	0.796	0.760	0.778
	Far Right	0.822	0.722	0.764
	Islamic Extremism	0.774	0.750	0.760
	Porn Addiction	0.626	0.698	0.648
	Mean	**0.798**	**0.774**	**0.781**
	Standard Deviation	**0.100**	**0.048**	**0.078**

What Can an Attacker Infer From Your Data?

- From just a few samples of your location, can infer future locations
- From widely spaced GPS points, can interpolate over gaps
- From web browsing, can infer membership in a sensitive group



Boost Privacy Concern with Inferences

Show people what can be inferred from their personal data.

With Eleanor Schille-Hudson (Indiana University) and Sid Suri (Microsoft Research)

Joke

About the guy buying a new car in Russia.

laughter

Change Your Mind About Privacy?

Experiment/survey of 928 people on Mechanical Turk

How much do you value the privacy of your age, gender and ZIP code data?

Enter your age, gender and ZIP code

Display inferences from age, gender and ZIP code

Now how much do you value the privacy of your age, gender and ZIP code data?

Compute difference

Do people change their attitude when they know what can be inferred?

Example Custom Inference

(50-59 years, male, 35016)

You told us you are a man. About 49.2% of Americans are men. You told us your age is 50 to 59 years. You are among 21.2 million American men, or about 6.55% of the total population, in that age range. You also told us you are living in ZIP code 35016. Based on your gender, age, and ZIP code, we can narrow down your identity to 0.0084% of the US population, or about 27,912 people out of the U.S. population of around 332 million.

Your annual income from wages or salary is likely around $45,000 per year, compared to the US average of $24,000. Your income is affected by your highest level of schooling, which we estimate is high school. We think the property you live on is worth about $180,000, which is different than the US average of $210,000. People like you most likely have 2 vehicles available in their household to drive.

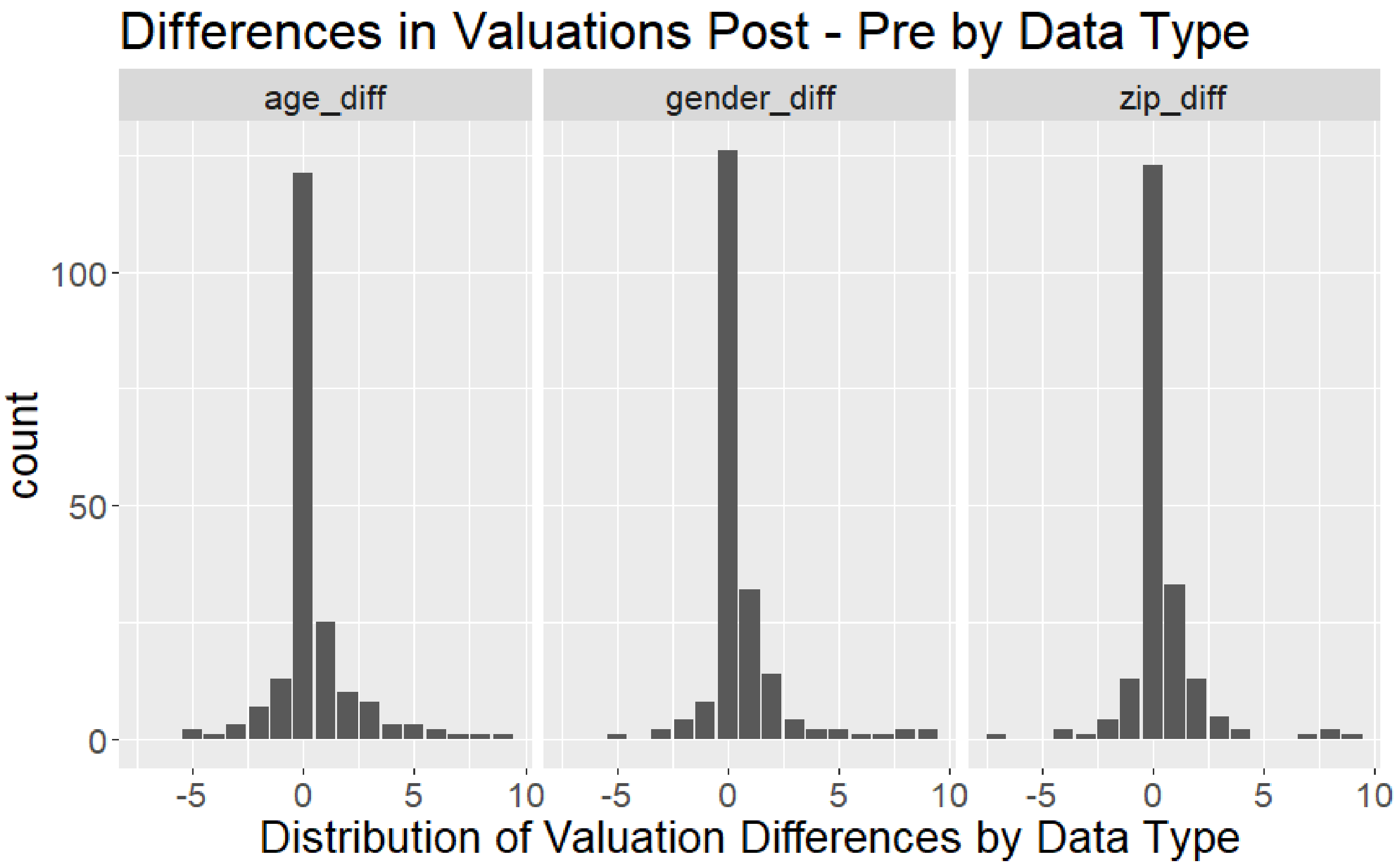
You very probably voted for Donald Trump in the 2020 presidential election.

Your chance of being obese is 35.6%, which is about the same as the U.S. average of 29.7%. We calculated that the chance of you being physically inactive is 32.4%, which is more than the U.S. average of 22.7%. Your chance of being a smoker is 22.2%, which is more than the U.S. average of 17.5%, and your chance of being an excessive drinker is 14.7%, which is less than the U.S. average of 19.3%.

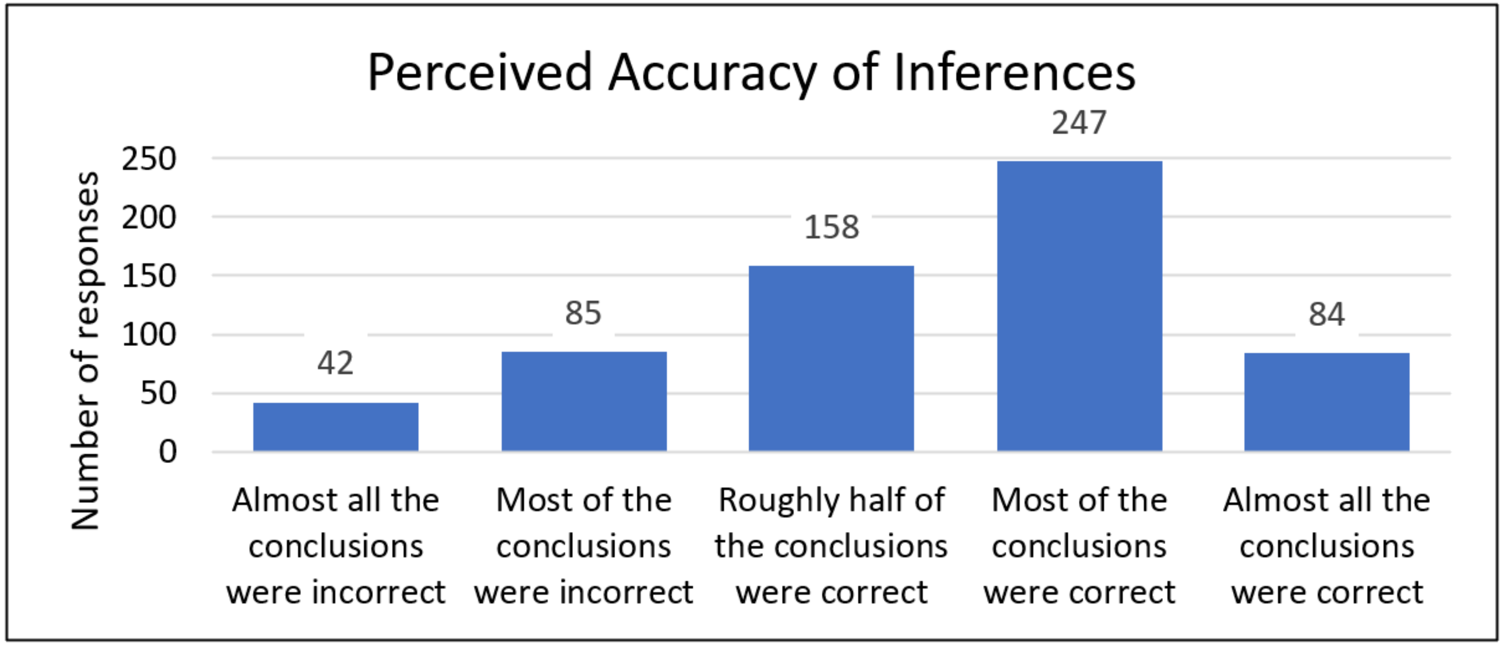
Results

We were able to shift people's valuations of age, gender, and ZIP code information by showing inferences.

Shifts were small, but statistically significant.



Inferences were generally correct



Personal Location Data Selling

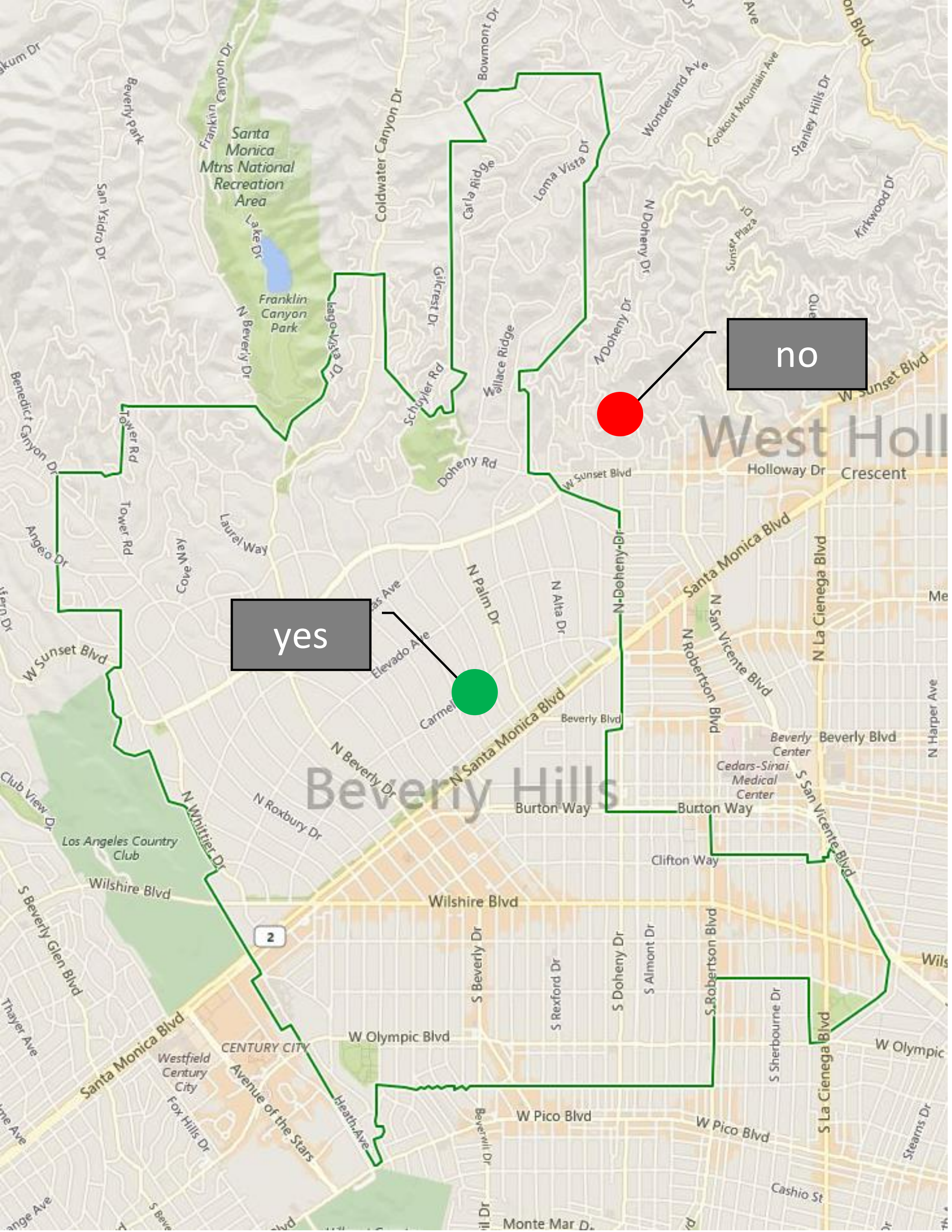
How much for one location point?

Buy one location point from one person

Joke

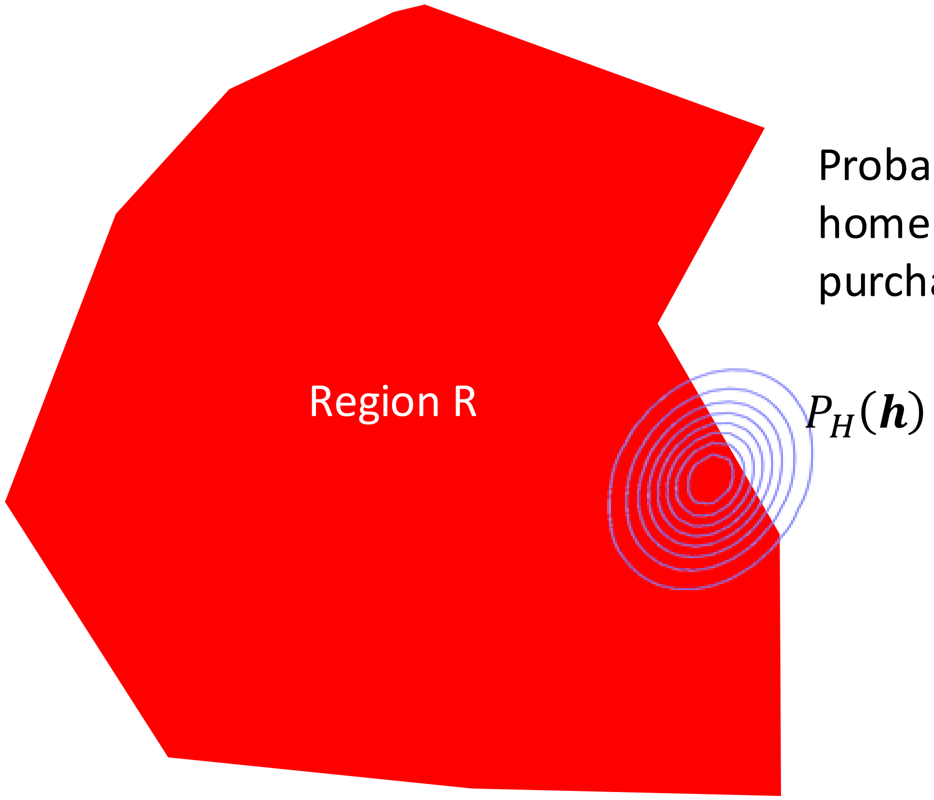
About the guy getting a call from his doctor.

laughter



Scenario: Business wants to deliver advertising to people with homes in given region.

- mobile dog grooming
- construction with license in region
- real estate in targeted neighborhood
- local stores and restaurants



Probability distribution of home location based on purchased lat/long points

$$\boldsymbol{h} = \begin{bmatrix} x \\ y \end{bmatrix}$$

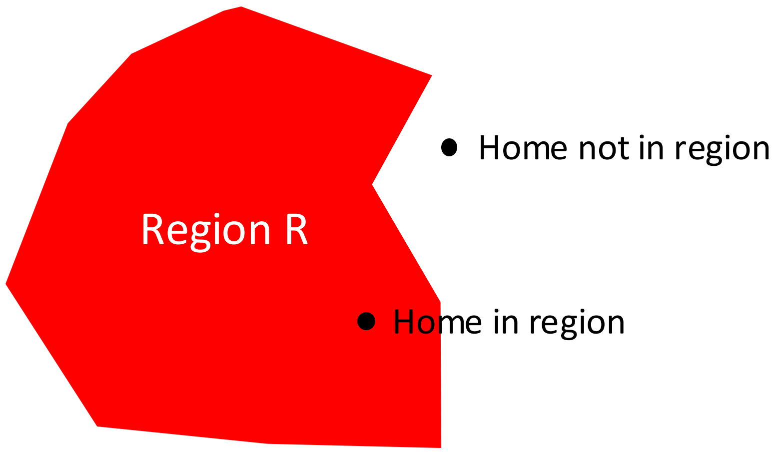
Probability of home in region R

$$p_R = \int_R P_H(\boldsymbol{h}) d\boldsymbol{h}$$

		Home location	
		Not in region	In region
Decision	Do not deliver ad	b_{11}	b_{12}
	Deliver ad	b_{21}	b_{22}

Payoff matrix

- Do not deliver ad to home **not in** region: No harm, no benefit, $b_{11} = 0$.
- Do not deliver ad to home **in** region: Could miss a sale, $b_{12} \leq 0$.
- Deliver ad to home **not in** region: Wasted ad, $b_{21} \leq 0$.
- Deliver ad to home **in** region: Good outcome, $b_{22} \geq 0$.

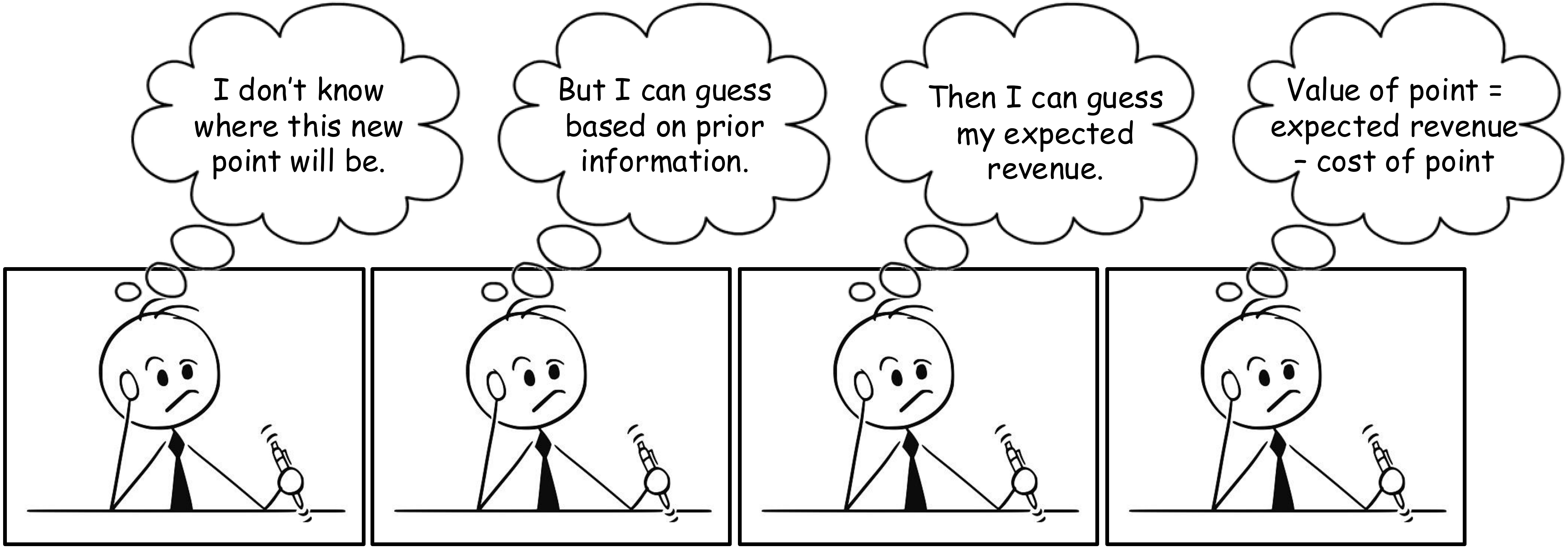


$E[V|no\ ad] = (1 - p_R)b_{11} + p_R b_{12}$ Expected return if *do not* deliver ad

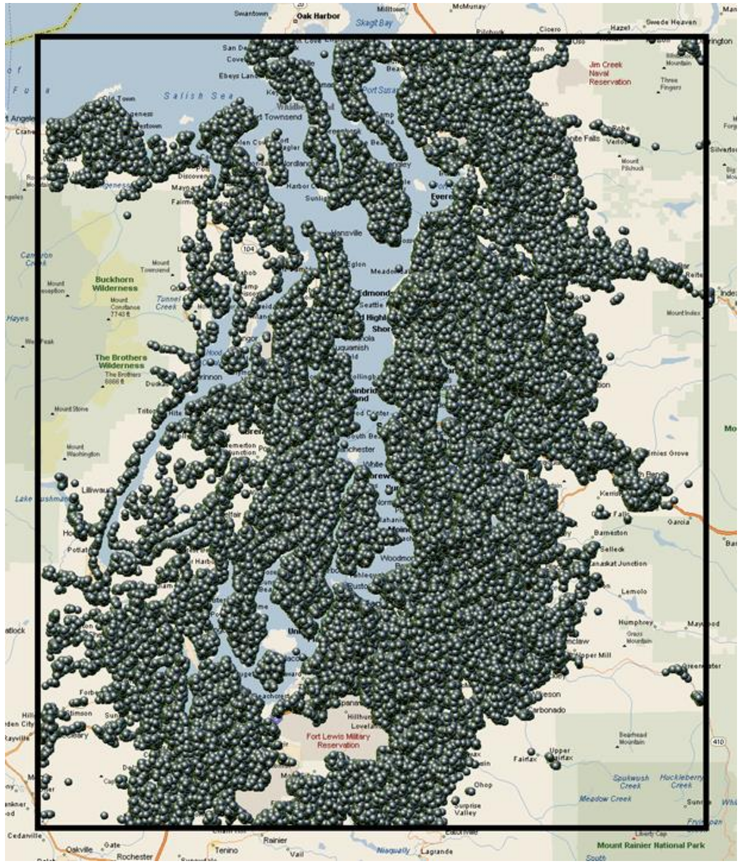
$E[V|ad] = (1 - p_R)b_{21} + p_R b_{22}$ Expected return if *do* deliver ad

Seller Offers Point for Sale

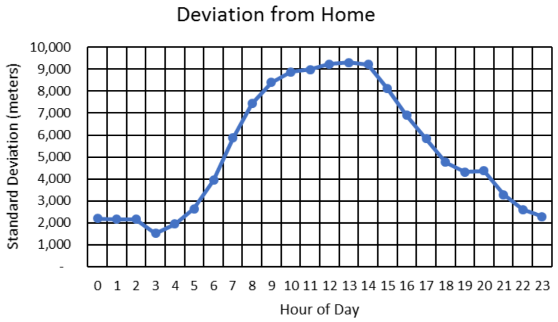
Buyer’s Reasoning About Point for Sale



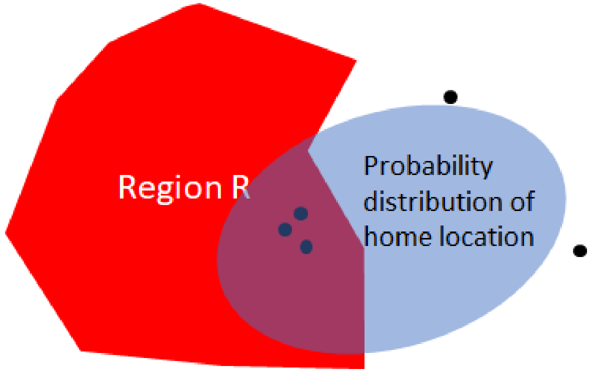
Heba Aly, John Krumm, Gireeja Ranade, and Eric Horvitz. "On the value of spatiotemporal information: Principles and scenarios." In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 179-188. 2018. (best paper runner up)



Known locations of all homes



Distance from home vs. time of day



Previously purchased location points

Prior Information

Economics of Buying New Point

$$VOI(l_{n+1}|L_1^n = l_1^n) = \mathrm{E}[V|L_1^{n+1} = l_1^{n+1}] - \mathrm{E}[V|L_1^n = l_1^n]$$

Value of information of new (unknown) point given previously purchased points

Expected revenue from new point along with all previous points

Expected revenue from just the previous points

$$EVOI(l_{n+1}|L_1^n = l_1^n) = \int VOI(l_{n+1}|L_1^n = l_1^n) \cdot P_{L_{n+1}}(l_{n+1})\,dl_{n+1}$$

EVOI tells what buyer is expected to gain by buying a new point, but without knowledge of the new point's location

Expected value of information of new (unknown) point given previously purchased points

Integrate VOI over anticipated distribution of new point's location

$$E\left[\text{Profit}\,\middle|\,L_1^{n+1} = l_1^{n+1}\right] = EVOI(l_{n+1}|L_1^n = l_1^n) - c_{n+1}$$

Expected profit is expected value of new point minus cost of new point

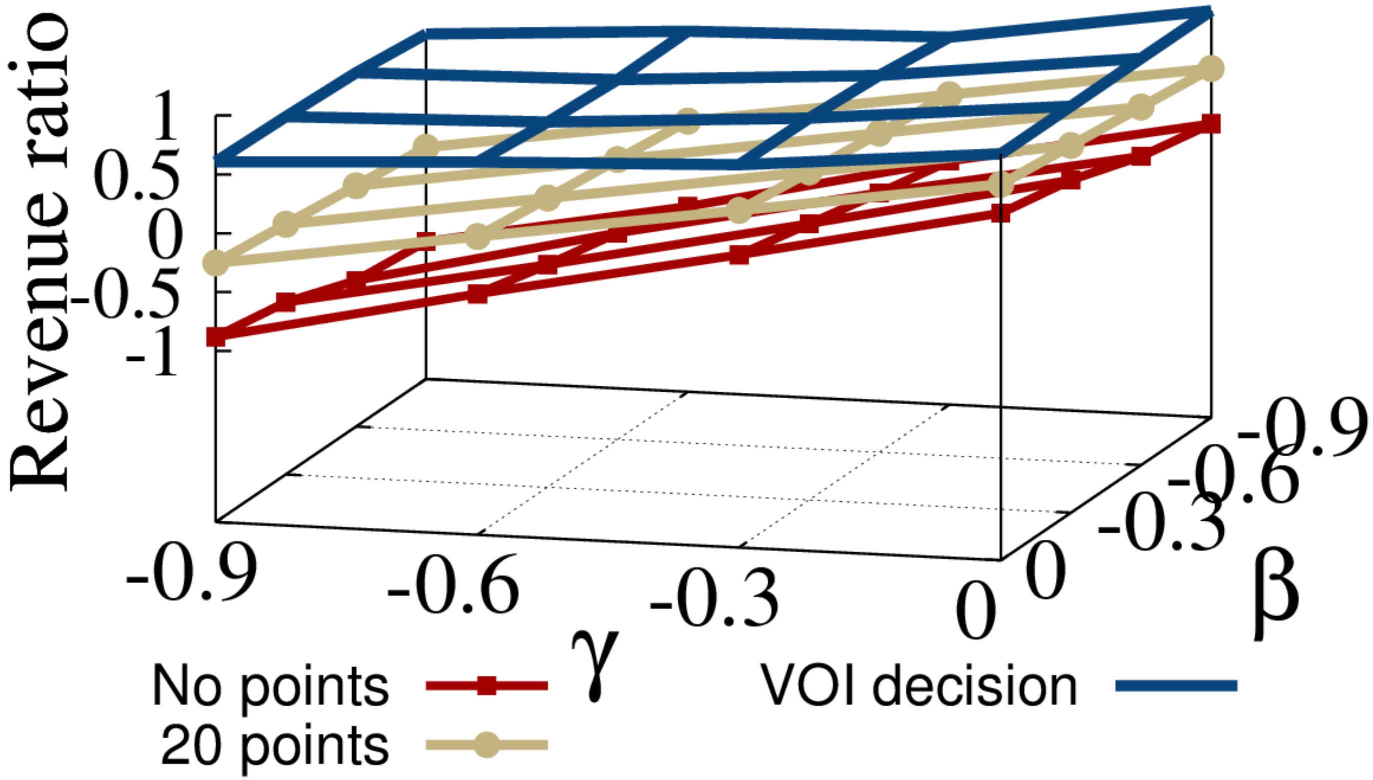
Test Data

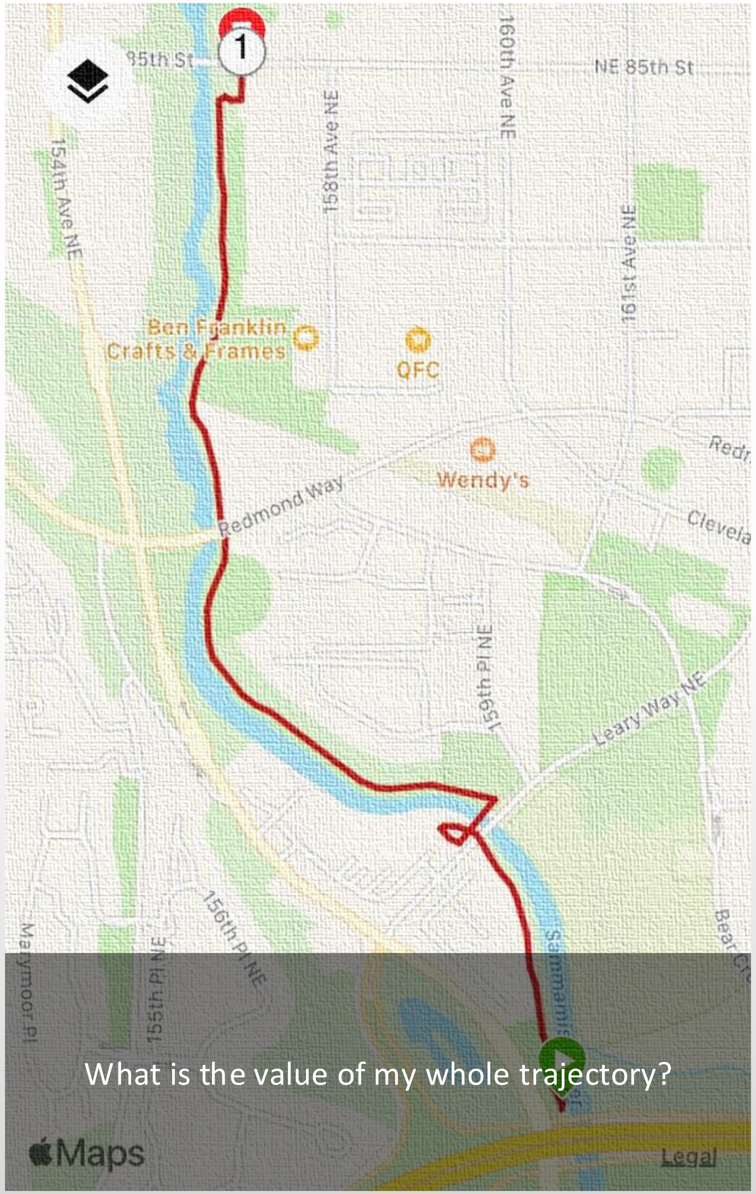
- GPS from 66 people
- Average 40 days each
- Advertising regions R_1, R_2, R_3

		Home location	
		Not in region	In region
Decision	Do not deliver ad	$b_{11} = 0$	$b_{12}=\beta$
	Deliver ad	$b_{21} = \gamma$	$b_{22}=1$

Payoff matrix

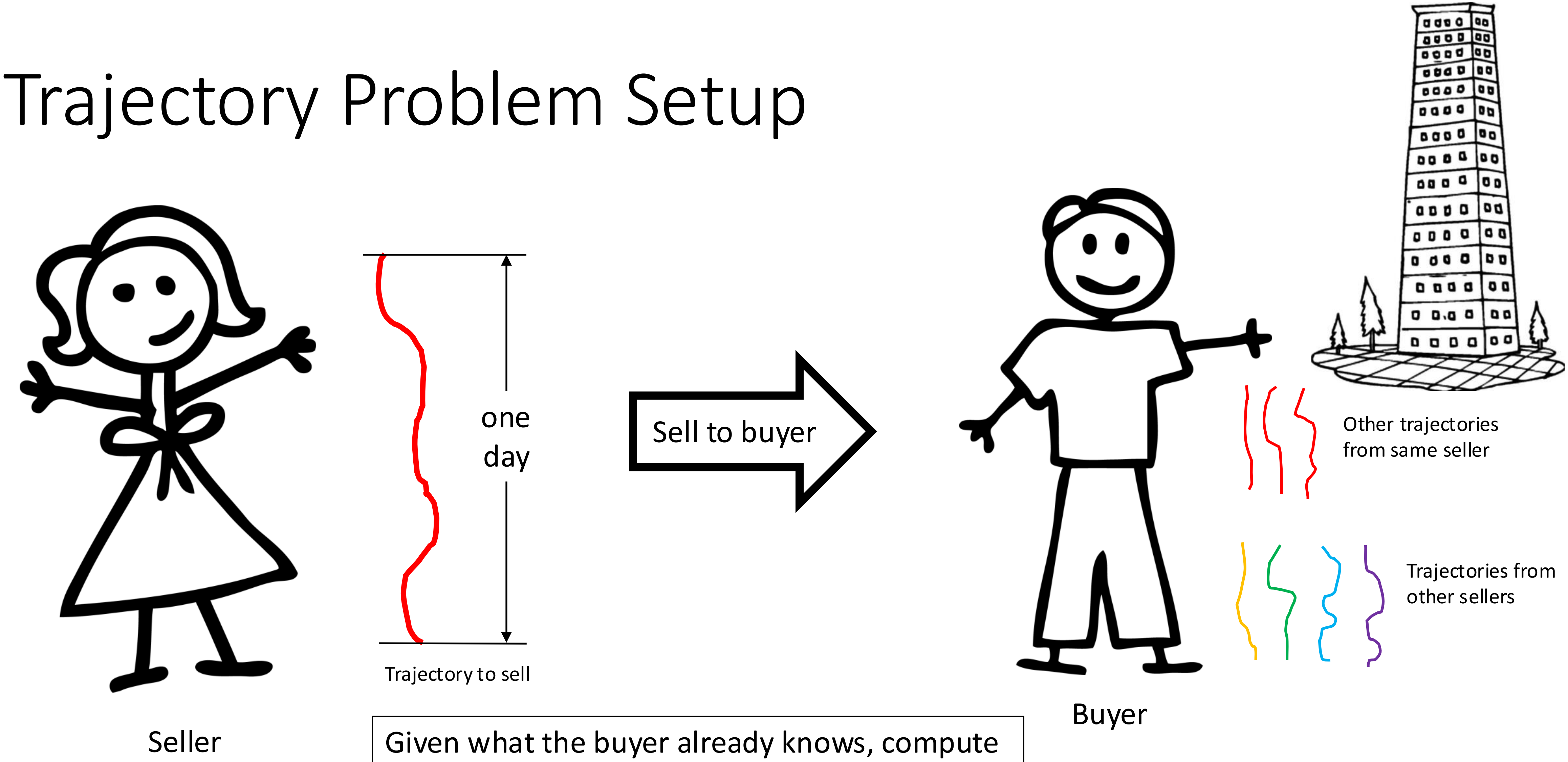
Revenue ratio = fraction of maximum possible profit with perfect decisions





Sell a whole trajectory

Trajectory Problem Setup



Given what the buyer already knows, compute the value of information in seller's trajectory.

Nguyen, Kien, John Krumm, and Cyrus Shahabi. "Quantifying Intrinsic Value of Information of Trajectories." In *proceedings of the 29th International Conference on Advances in Geographic Information Systems*, pp. 81-90. 2021.

Summary - Personal Data Privacy – Especially Location

Privacy Leaks

- Differential privacy is for aggregating data
- Adding noise to your location data is not the ultimate answer
- Inferences from 1-5 location points
- Infer visits between measurements
- Infer membership in sensitive group

Raise privacy awareness by exposing possible inferences

Compute monetary value of location data for sale

Summary - Jokes

- About television in North Korea
- About the Roman walking into a bar
- About the hamburger walking into a bar
- About the airplane with engine problems
- About the guy buying a new car in Russia
- About the guy getting a call from his doctor

applause

Jokes

1) Did you hear about the first restaurant to open on the moon? It had great food, but no atmosphere.
2) A Roman walks into a bar, holds up two fingers and says, “Five beers, please.”
3) A hamburger walked into a bar, and the bartender said, “Sorry, we don’t serve food here.”
4) A man answers the phone, and his doctor says “I have bad news and worse news”.
   “Oh dear, what's the bad news?” asks the man.
   The doctor replies, “You only have 24 hours to live.”
   “That's terrible”, said the man. “How can the news possibly be worse?”
   The doctor replies, “I've been trying to contact you since yesterday.”
5) A guy goes to a car dealership in in Russia to buy a new car. The salesman says, “Your new car will be available to pick up in 10 years.” The guys asks, “Morning or afternoon?” The salesman says, “It’s 10 years from now. Does it really matter?” And the guys says, “Well, the plumber is coming that morning.”

laughter

smattering of applause