

Visualizing Deanononymization

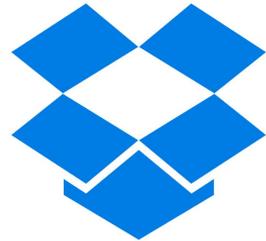
Application of VAST Challenge to Data Breaches

CPSC 547
Rosalyn Carr

Data Breaches: An Overview

- Illegal use or disclosure of confidential information
- 1,862 data breaches in 2021, surpassing 2020's total of 1,108
- 83% of these breaches included sensitive information such as Social Security Numbers

medibank



Dropbox



Current Approaches

Academia

- Use complicated techniques that are effective but not easily understood by lay-people
- Utilizes real data without permission, possible ethical concerns*

Media

- Text heavy articles repeating similar advice
- Often not sought out until a breach has occurred, yet often limited to preventative measures

*L. Liu, M. Han, Y. Wang, and Y. Zhou, 'Understanding Data Breach: A Visualization Aspect', in *Wireless Algorithms, Systems, and Applications*, Cham, 2018, pp. 883–892. doi: 10.1007/978-3-319-94268-1_81

Goal

- Visualize and walk through a simple but believable deanonymization process in a way that allows users to understand the severity of data breaches even when less sensitive data is exposed
- Use a representative but artificial dataset to preserve privacy

VAST 2021 Mini Challenge 2

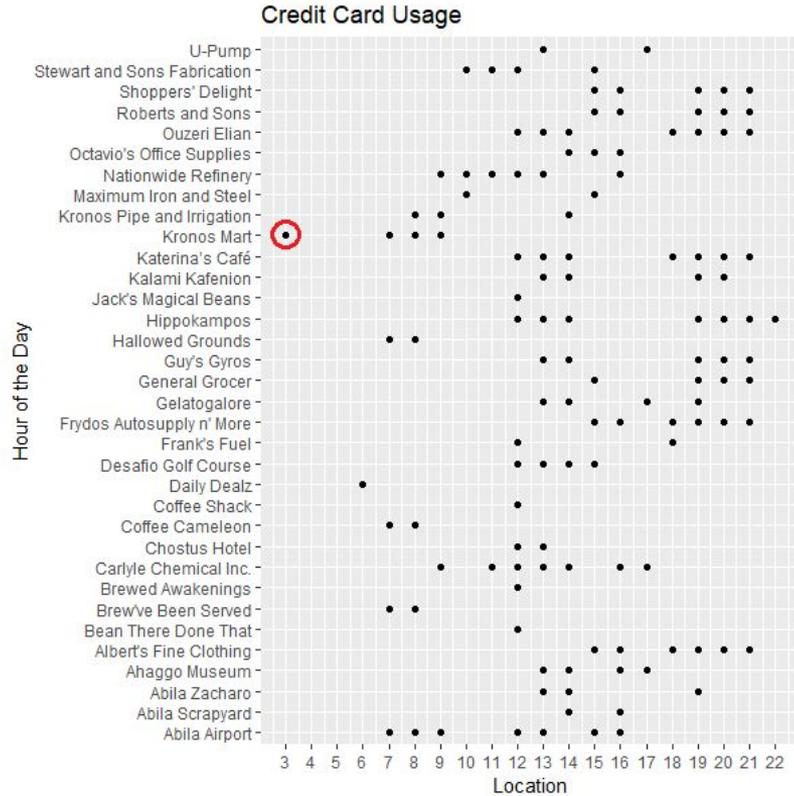
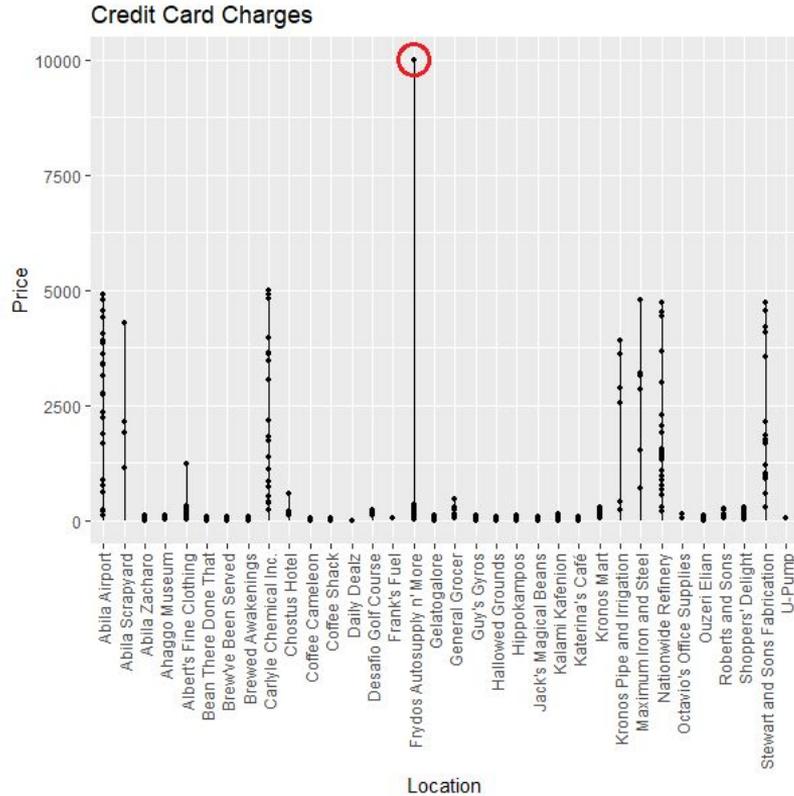
- Original challenge was “surveillance” themed
- Patterns of employee behaviour could be used as surrogate breach data



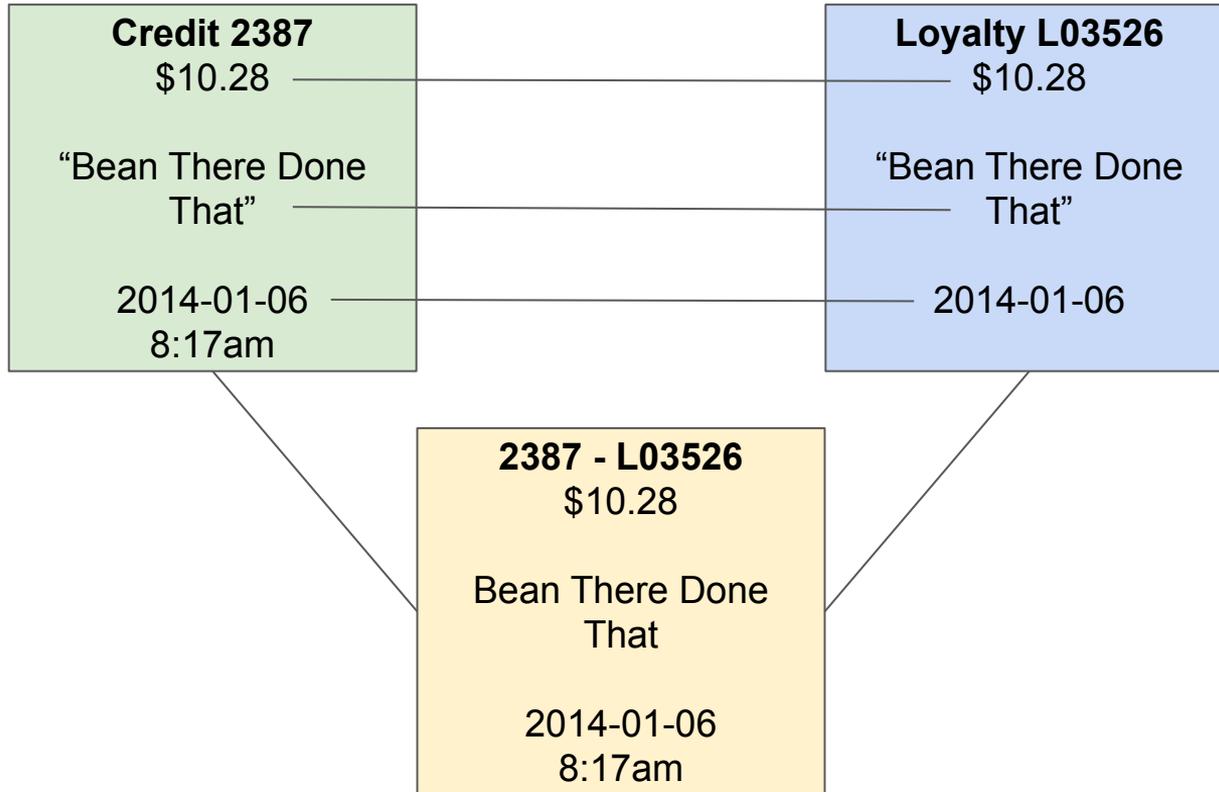
Data

- Credit Card Log
 - Price, Time and Date, Credit Card Number, Business Name (Location)
- Loyalty Card Log
 - Price, Date, Loyalty Card Number, Business Name (Location)
- Car Assignments
 - Car ID Number, Employee Name, Title and Job Classification
- GPS Log
 - Date and Time, Car ID Number, Latitude, Longitude
- Geospatial Maps
 - ESRI Shapefiles and Projects*
- Tourist Map Image

Solution: Explainer Article



Task 1: Matching Credit and Loyalty Cards



Purchase History

Location:

Credit Card:

Loyalty Card:

Show entries

Search:

	Price	Date	Location	Credit.Card	Loyalty.Card
131	20.33	2014-01-13	Kalami Kafenion	5368	L2247
132	36.24	2014-01-13	Gelatogalore	8202	L2343
133	38.07	2014-01-13	Guy's Gyros	7117	L6417
134	26.96	2014-01-13	Hippokampos	4948	L9406
135	188.57	2014-01-13	Frydos Autosupply n' More	8129	L8328
136	64.6	2014-01-13	Frydos Autosupply n' More	8411	L6110
137	21.83	2014-01-13	Ouzeri Elian	2681	L1107
138	19.77	2014-01-13	Guy's Gyros	5368	L2247
139	32.29	2014-01-13	Hippokampos	7117	L6417
140	202.05	2014-01-13	Frydos Autosupply n' More	2418	L9018

Showing 131 to 140 of 1,087 entries

Previous 1 ... 13 **14** 15 ... 109 Next

Purchase History

Location:

Credit Card:

Loyalty Card:

Show entries

Search:

	Price	Date	Location	Credit.Card	Loyalty.Card
232	10.28	2014-01-06	Bean There Done That	1415	L7783
233	15.44	2014-01-06	Guy's Gyros	1415	L7783
234	21.33	2014-01-06	Hippokampos	1415	L7783
308	13.52	2014-01-07	Abila Zacharo	1415	L7783
309	11.56	2014-01-07	Hippokampos	1415	L7783
367	14.41	2014-01-08	Abila Zacharo	1415	L7783
368	17.45	2014-01-08	Bean There Done That	1415	L7783
369	10.52	2014-01-08	Hippokampos	1415	L7783
441	13	2014-01-09	Bean There Done That	1415	L7783
516	15.39	2014-01-10	Bean There Done That	1415	L7783

Showing 1 to 10 of 24 entries

Previous 2 3 Next

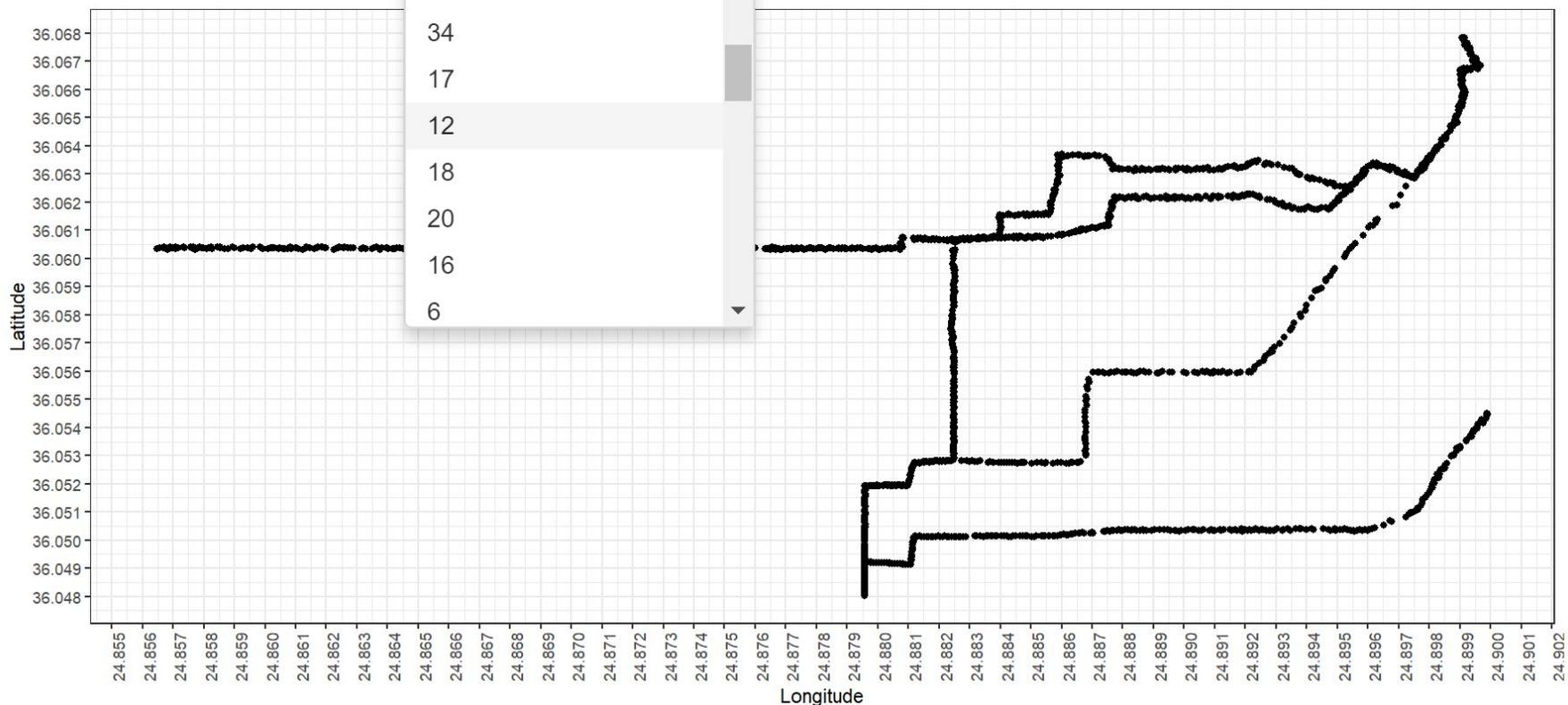
GPS Logs

Date: Car ID: Location Labels

2014-01-08

12

- 32
- 34
- 17
- 12
- 18
- 20
- 16
- 6



GPS Logs

Date:

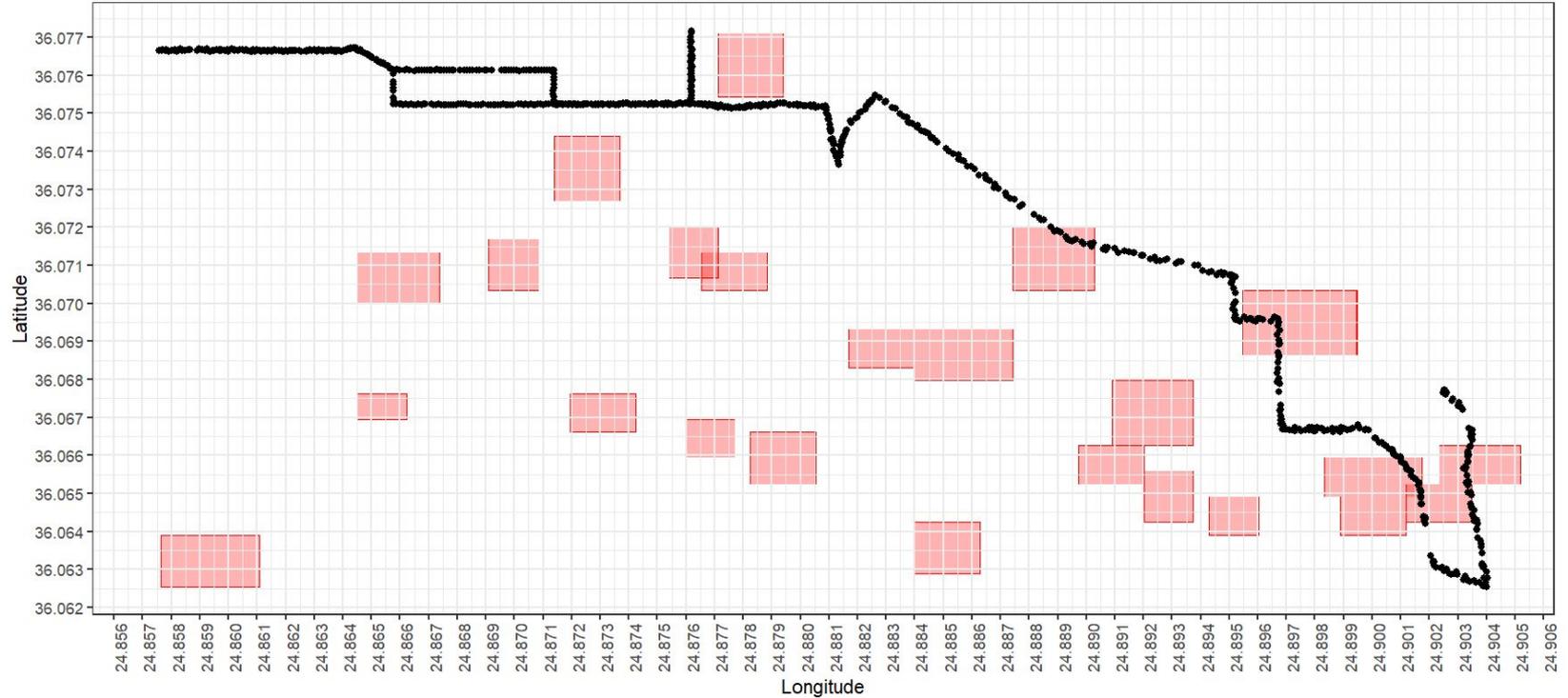
2014-01-11

Car ID:

34

Location

Labels



Deanonymized Purchase History

Name: Location: Credit Card: Loyalty Card:

Show entries

Search:

	Price	Date	Location	Credit Card	Loyalty Card	Name
1	22.64	2014-01-06	Gelatogalore	7889	L6119	Lucas Alcazar
2	14.05	2014-01-07	Hallowed Grounds	7889	L6119	Lucas Alcazar
3	22.27	2014-01-07	Katerina's Café	7889	L6119	Lucas Alcazar
4	35.29	2014-01-08	Gelatogalore	7889	L6119	Lucas Alcazar
5	36.6	2014-01-08	Katerina's Café	7889	L6119	Lucas Alcazar
6	13.59	2014-01-09	Guy's Gyros	7889	L6119	Lucas Alcazar
7	8.23	2014-01-09	Guy's Gyros	7889	L6119	Lucas Alcazar
8	10.34	2014-01-10	Hallowed Grounds	7889	L6119	Lucas Alcazar
9	12.55	2014-01-11	Katerina's Café	7889	L6119	Lucas Alcazar
10	8.54	2014-01-12	Hippokampos	7889	L6119	Lucas Alcazar

Showing 1 to 10 of 20 entries

Previous 2 Next

Lessons Learned

- What is conceptually simple can be computationally tedious
 - Data linking can become very complicated when introducing margins
- The tool you use matters
 - R Shiny doesn't deal with missing values very well
- Artificial data can create huge problems
 - ESRI no longer offers student licenses, so the geospatial data files were almost completely unworkable and required majority manual intervention
 - Data was designed for a specific challenge, and trying to expand on an already difficult set of tasks was a bit ambitious