Big data, AI, and data privacy for transport planning

Presentation

Author(s):
Anda, Cuauhtémoc

Publication date:
2018-10-22

Permanent link:
https://doi.org/10.3929/ethz-b-000299710

Big Data, AI and Data Privacy for Transport Planning

Tongji University

Cuauhtemoc Anda
anda@arch.ethz.ch

22.Oct.2018

(FCL) FUTURE CITIES LABORATORY

ETH zürich

Big Data, AI and Data Privacy for Transport Planning

Tongji University

Cuauh**temo**c Anda
anda@arch.ethz.ch

22.Oct.2018

(FCL) FUTURE CITIES LABORATORY

ETH zürich

Big Data, AI and Data Privacy for Transport Planning

Tongji University

Cuauhtemoc **Anda**
anda@arch.ethz.ch

22.Oct.2018

(FCL) FUTURE CITIES LABORATORY

ETH zürich

(SEC) SINGAPORE-ETH CENTRE 新加坡－ETH研究中心

(FCL) FUTURE CITIES LABORATORY

CREATE
Campus for Research Excellence And Technological Enterprise

ETH zürich

EPFL

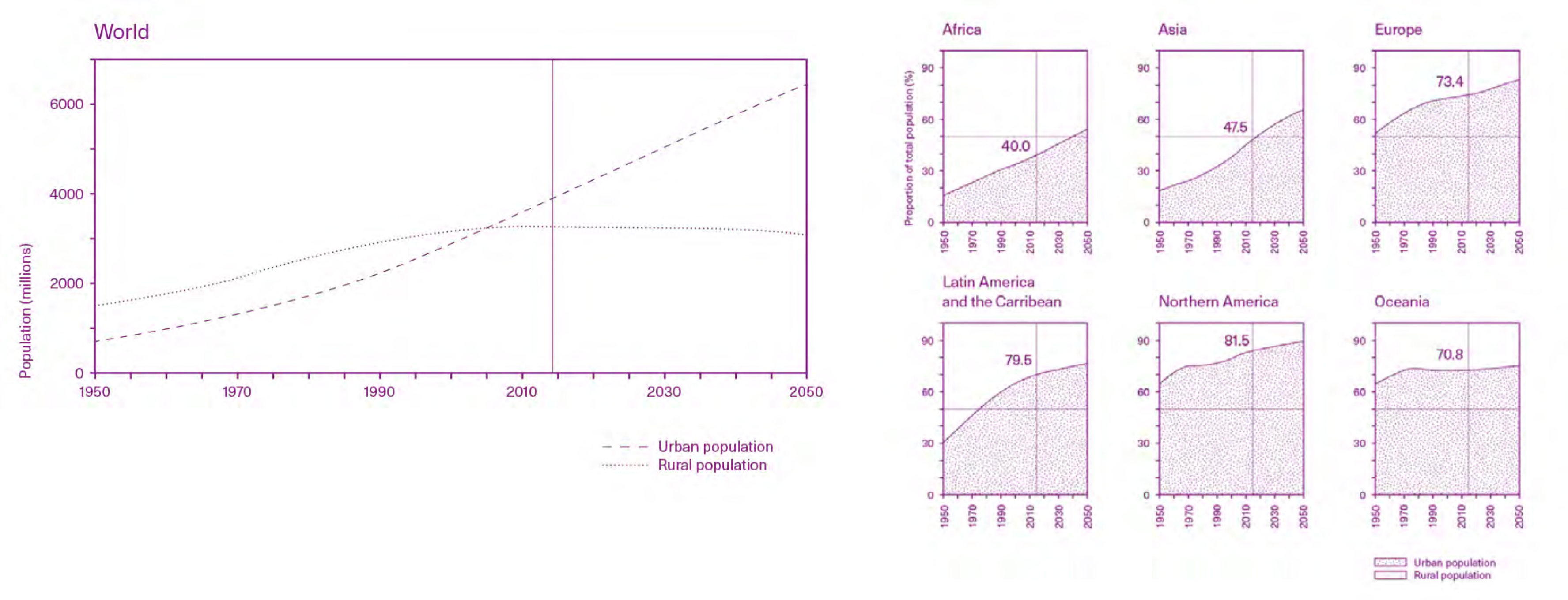
NANYANG TECHNOLOGICAL UNIVERSITY

NUS National University of Singapore

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

[FCL] Future Cities Laboratory

Sustainable Future Cities: Through Science, By Design, In Place



Rationale

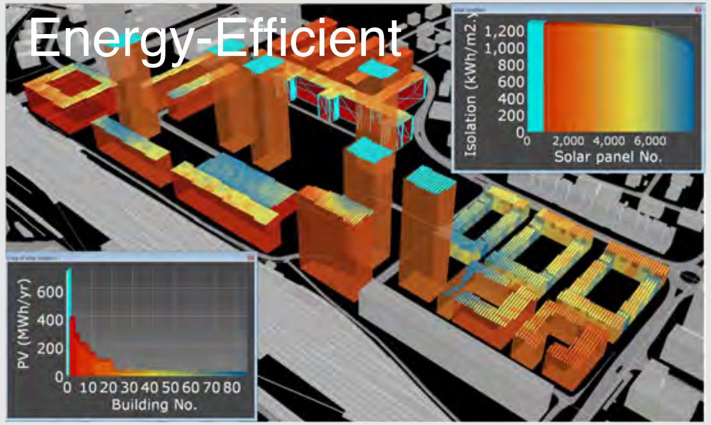
Urbanisation

Global phenomenon with unique local features

Presenting new threats and possible futures

Requiring new forms of description, analysis and action at global and local scales

Urban and rural population of the world, 1950–2050. Urban and rural population as proportion of total population, by major region, 1950–2050 (UN DESA 2014)



What Future City?

Developing a roadmap to inform future city-making and urban policies through 'Transformative Research':

- Innovative research approaches, tools and methods
- Design exploration
- Scenario testing
- Technology platforms

Programme Structure

Inter-Disciplinary Scenarios

High-Density Mixed-Use Cities

The *Grand Projet*
Ecosystem Services
Multi-Scale Energy Systems
Dense and Green

Responsive Cities

BigData-Informed Urban Design
Cyber Civil Infrastructure
Engaging Mobility
Cognition, Perception and Behaviour

Archipelago Cities

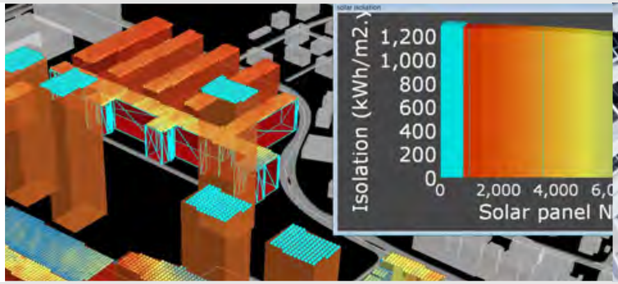
SIJORI and Extended Urbanisation
Urban-Rural Systems
Alternative Construction Materials
Tourism and Cultural Heritage

Programme Structure

Inter-Disciplinary Scenarios

High-Density Mixed-Use Cities

The *Grand Projet*
Ecosystem Services
Multi-Scale Energy Systems
Dense and Green



Responsive Cities

BigData-Informed Urban Design
Cyber Civil Infrastructure
Engaging Mobility
Cognition, Perception and Behaviour

Archipelago Cities

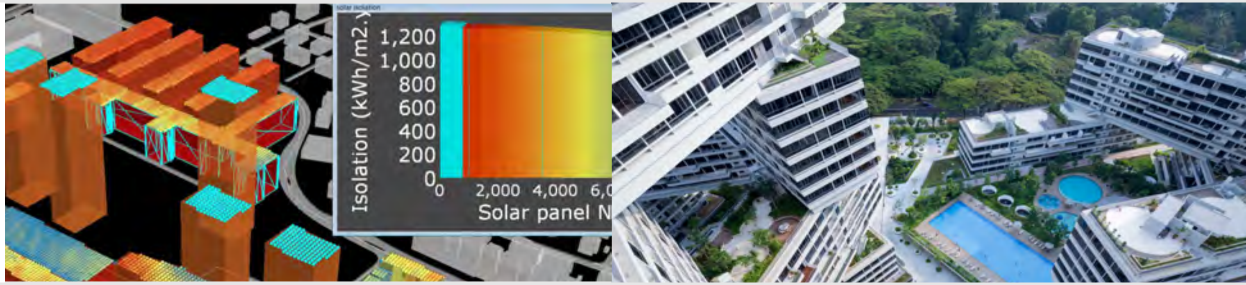
SIJORI and Extended Urbanisation
Urban-Rural Systems
Alternative Construction Materials
Tourism and Cultural Heritage

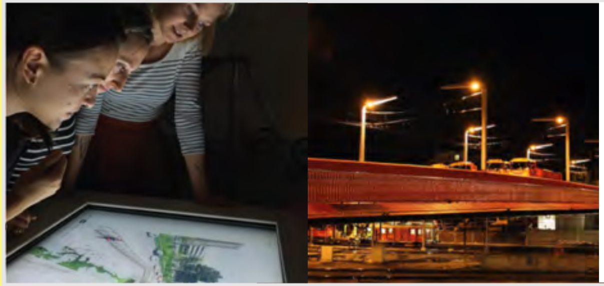
Programme Structure

Inter-Disciplinary Scenarios

High-Density Mixed-Use Cities

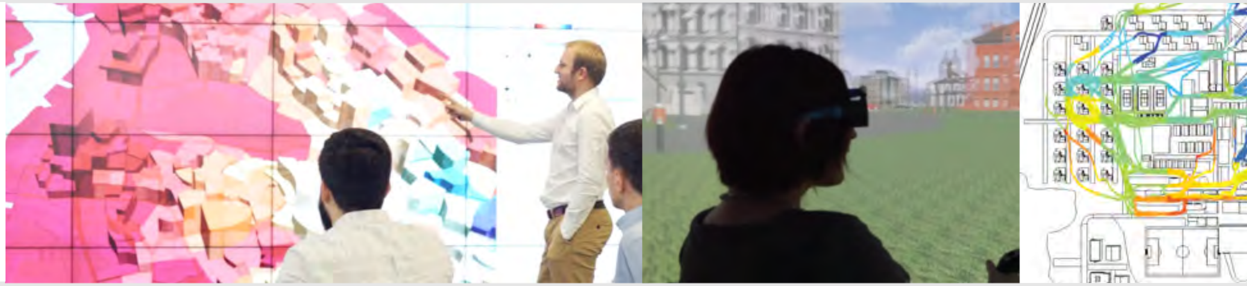
The *Grand Projet*
Ecosystem Services
Multi-Scale Energy Systems
Dense and Green





Responsive Cities

BigData-Informed Urban Design
Cyber Civil Infrastructure
Engaging Mobility
Cognition, Perception and Behaviour



Archipelago Cities

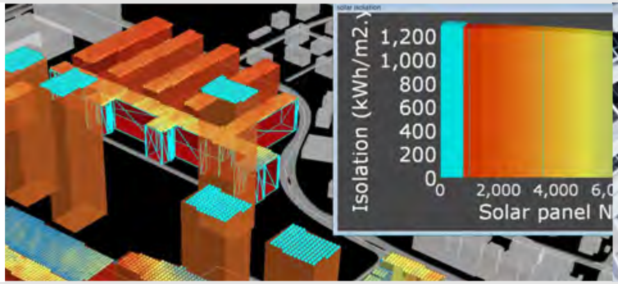
SIJORI and Extended Urbanisation
Urban-Rural Systems
Alternative Construction Materials
Tourism and Cultural Heritage

Programme Structure

Inter-Disciplinary Scenarios

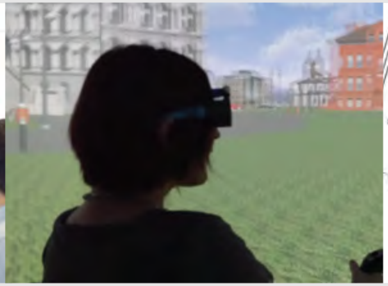
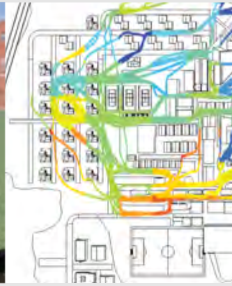

High-Density Mixed-Use Cities

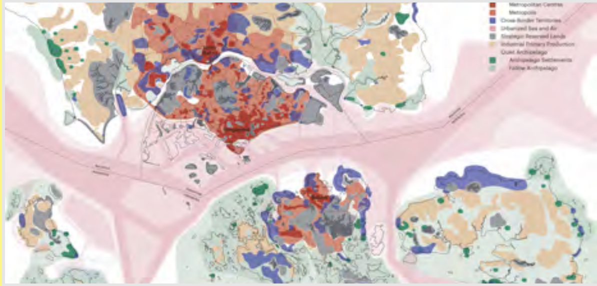
The *Grand Projet*
Ecosystem Services
Multi-Scale Energy Systems
Dense and Green




Responsive Cities

BigData-Informed Urban Design
Cyber Civil Infrastructure
Engaging Mobility
Cognition, Perception and Behaviour

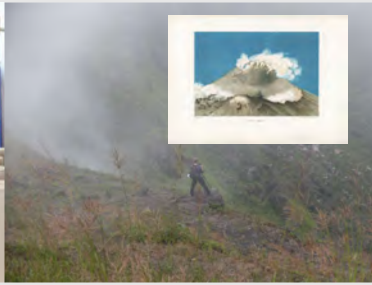






Archipelago Cities

SIJORI and Extended Urbanisation
Urban-Rural Systems
Alternative Construction Materials
Tourism and Cultural Heritage





Programme Structure

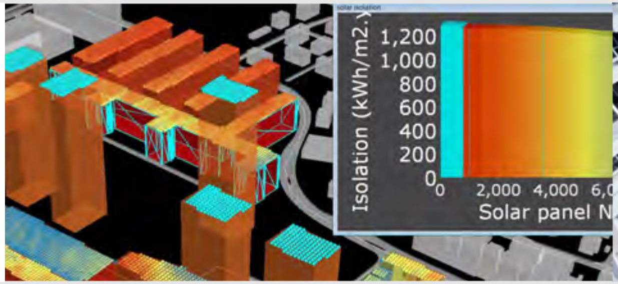
Disciplines

Inter-Disciplinary Scenarios

- Architecture, Planning & Design
- Mobility & Transport Planning
- Psychology & Social Sciences

High-Density Mixed-Use Cities

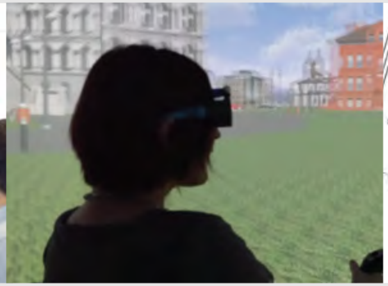
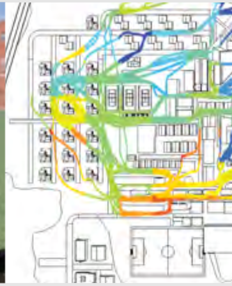
The *Grand Projet*
Ecosystem Services
Multi-Scale Energy Systems
Dense and Green



- Arts & Humanities
- Energy Systems
- Lanscape & Ecosystems

Responsive Cities

BigData-Informed Urban Design
Cyber Civil Infrastructure
Engaging Mobility
Cognition, Perception and Behaviour

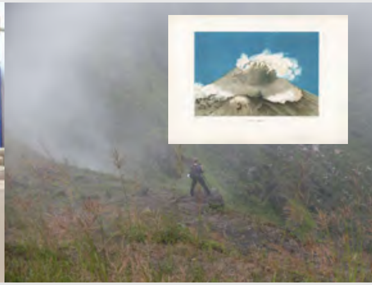




- Engineering & Materials Science
- Information Technology
- Computer Science

Archipelago Cities

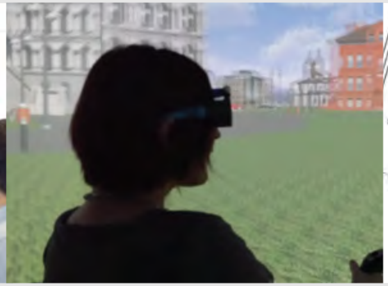
SIJORI and Extended Urbanisation
Urban-Rural Systems
Alternative Construction Materials
Tourism and Cultural Heritage

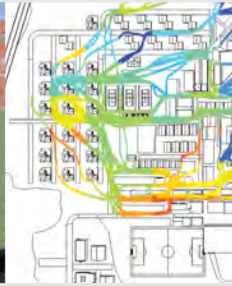


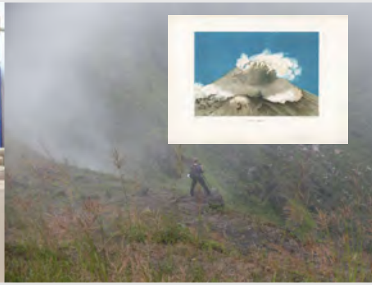


Programme Structure

Disciplines	Inter-Disciplinary Scenarios	Network
Architecture, Planning & Design Mobility & Transport Planning Psychology & Social Sciences	**High-Density Mixed-Use Cities** The *Grand Projet* Ecosystem Services Multi-Scale Energy Systems Dense and Green	Government: URA, HDB, LTA Universities: NUS, NTU, SUTD Universities: UI, Chulalongkorn, ITB
Arts & Humanities Energy Systems Lanscape & Ecosystems	**Responsive Cities** BigData-Informed Urban Design Cyber Civil Infrastructure Engaging Mobility Cognition, Perception and Behaviour	Universities: MIT, Harvard, LSE Industry: Schindler, Siemens, Shell Industry: Sika, Holcim, Veolia
Engineering & Materials Science Information Technology Computer Science	**Archipelago Cities** SIJORI and Extended Urbanisation Urban-Rural Systems Alternative Construction Materials Tourism and Cultural Heritage	Development Banks: ADB, WB NGOs: Rockefeller, Mercy Corps NGOs: KPC, Habitat for Humanity







What transport policy and design decisions can foster more liveable and sustainable cities in the future?

Engaging Active Mobility
How can we assess and evaluate cycling infrastructure designs based on behavioural reactions of cyclists?

Engaging Big Data
How can newly available large streams of big data be leveraged to improve transportation simulation models?

Planning for AV
How can we understand and evaluate impacts of AVs on Urban form and transport supply in Singapore?

Measuring Pedestrian Comfort
How can we enhance pedestrian comfort and mitigate negative impacts of crowding through better design?

We are making steady progress on our four ongoing projects.

Sources:
Top Left, Bottom right: Lina Meisen
Bottom left: Michael v Eggermond
Top right: Tanvi Maheshwari

Team

Dr. Pieter Fourie
Project Leader
Simulation

Dr. Alex Erath
Project Advisor

Prof. Dr. K. Axhausen
Co-PI
Transport Planning

Prof. Dr. C. Hölscher
Co-PI
Cognitive Psychology

Dr. Michael v. Eggermond
Project Coordinator
Senior Researcher
Active Mobility

Dr. Sergio Ordonez
Senior Researcher
Computer Science

Shuchen Xu
Researcher
Architect

Biyu Wang
MATSim developer

Cuauhtémoc Anda
PhD Researcher
Data Science

Mohsen Nazemi
PhD Researcher
Active Mobility

Tanvi Maheshwari
PhD Researcher
Urban Design

Michael Joos
Senior Software Engineer
Gaming Developer

URBAN REDEVELOPMENT AUTHORITY · Land Transport Authority · DataSpark · MATSim Multi-Agent Transport Simulation · PTV GROUP · NHTV Breda University of Applied Sciences · conveyal

Big Data, AI and Data Privacy for Transport Planning

Tongji University

Cuauhtemoc Anda
anda@arch.ethz.ch

22.Oct.2018

(FCL) FUTURE CITIES LABORATORY

ETH zürich

What is MATSim?

MATSim

Multi-Agent
Transport Simulation

- Tool for Transport Planning
- Make inform decisions
- Predicts and evaluates future mobility scenarios
- Co-evolutionary algorithm (AI)

[policy | infrastructure | land-use | new mobility systems]

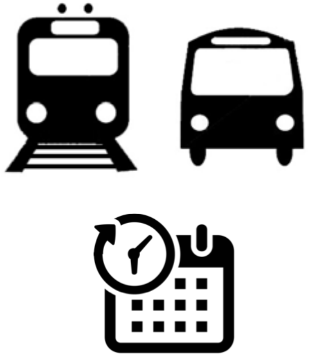
Up. Mobile Phone
Data MATSim
(Anda, et al., 2017)

What is needed?

What is needed to run a simulation?

Network codification

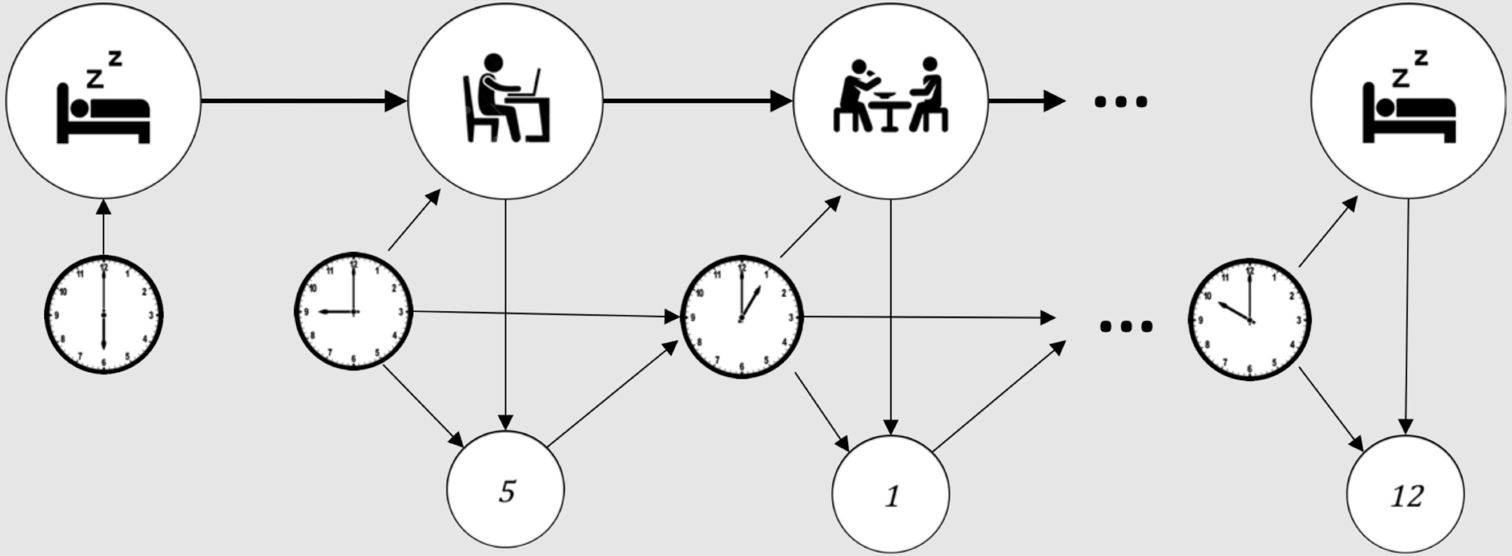
Transport mode availability



Facilities information

Description of the demand

Generating the population of agents



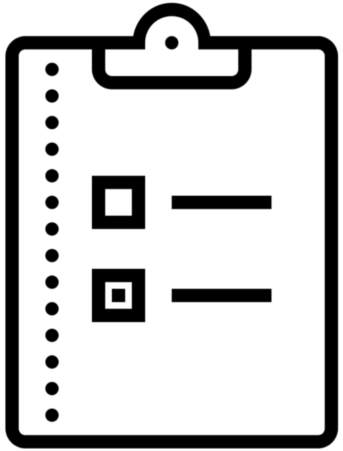
[activities | trips | schedules | socio-demographics | transport mode]

Description of the demand

How to get this information?

Data input: **Household Travel Survey**

- Rich data: destinations, activities, mode of transport, household structure, sociodemographic info, etc.



But,

- Covers only 1% of the population
- Expensive
- 4-5 years update
- Based on what people report

What if …
there were **mobility data** that **covers 50% of the population**. It is **collected everyday**.
Not reported but **measured data**. Continuously sensing **position and trajectories**.
You **don’t need** to invest in **additional infrastructure**.

Mobile phone *telco* data

What is it?

- Mobile phone antennas are scattered around Shanghai. Your mobile phone gets connected to the closest one, and leaves a timestamp.

How does it look like?

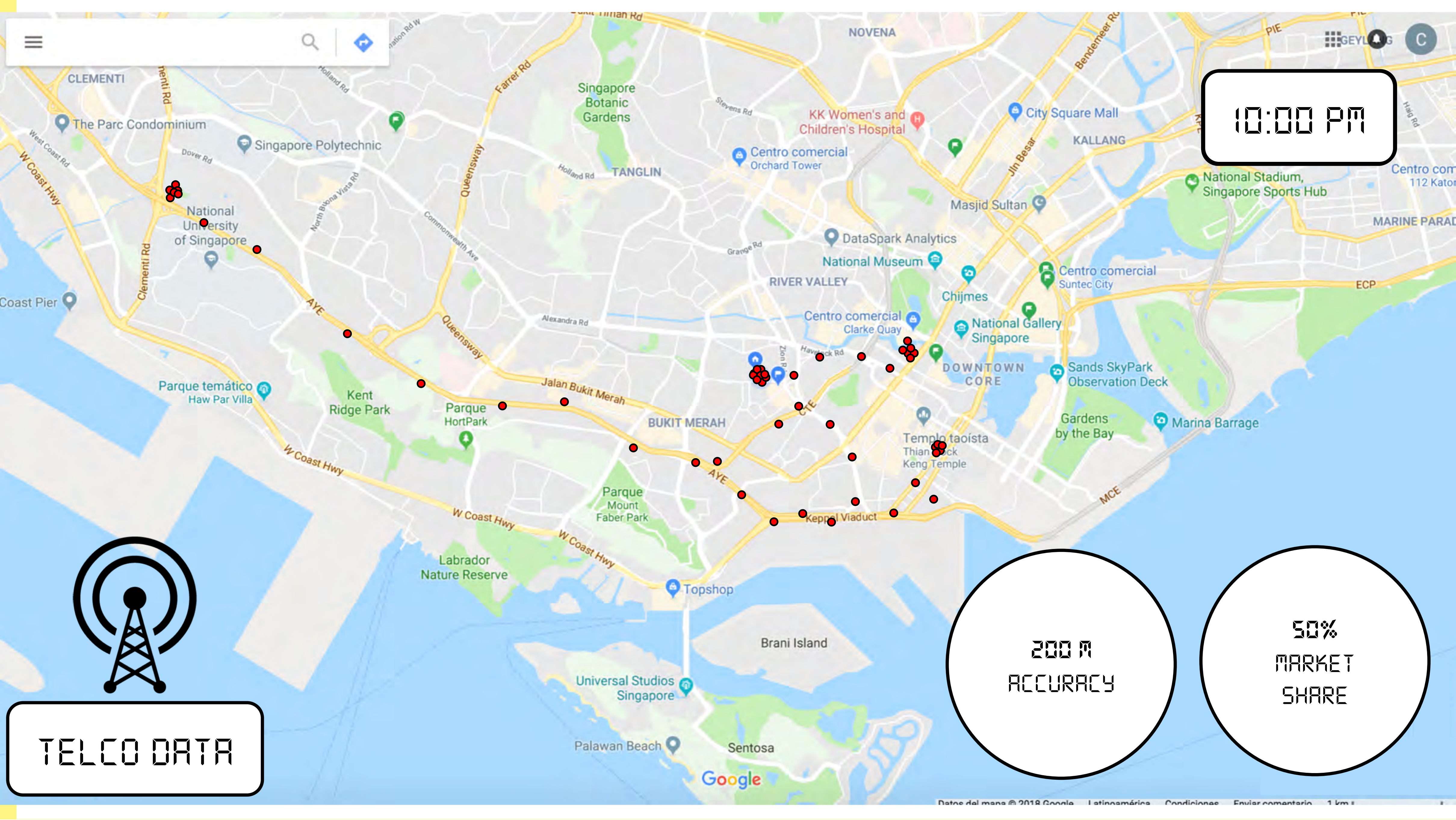
Name	Mobile	Connections
Dr. SHI Cheng	+86 1601..	CT45_20180112_07:51:23:2345, CT45_20180112_07:54:54:4351, CT47_20 CT51_20180112_08:33:25:2432...
Temo Anda	+65 8732..	CT23_20180112_06:30:44:6322, CT23_20180112_06:32:41:1211, CT23_20 CT20_20180112_07:55:33:5254...
Lu Han	+86 2021...	CT89_20180112_06:45:43:4425, CT90_20180112_06:45:59:9888, CT89_20 CT90_20180112_07:23:55:6329...

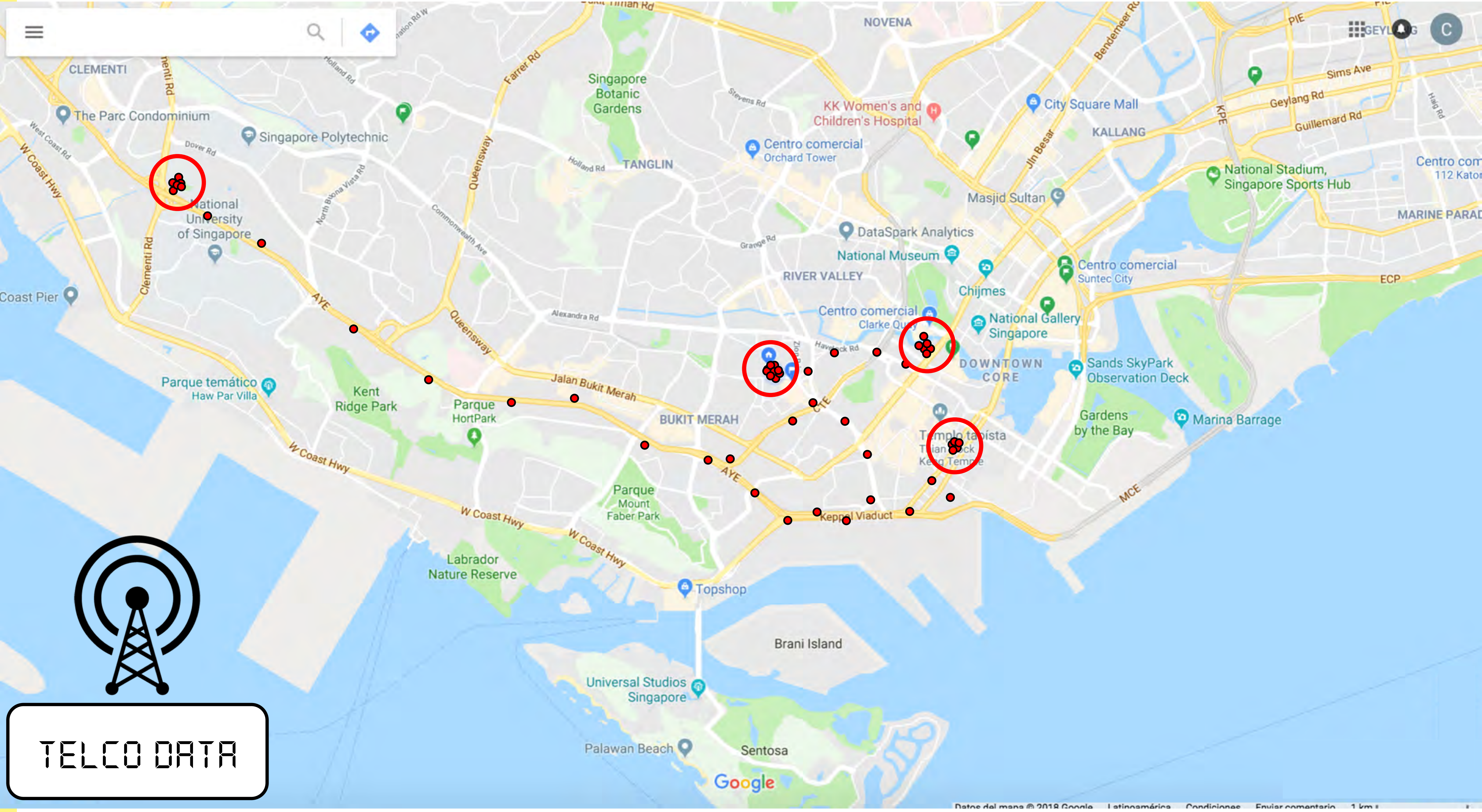
What if we anonymize it?

Id	Connections
X11234FFD8	CT45_20180112_07:51:23:2345, CT45_20180112_07:54:54:4351, CT47_20180112_08:11:02:3421, CT51_20180112_08:33:25:2432...
X95840908F	CT23_20180112_06:30:44:6322, CT23_20180112_06:32:41:1211, CT23_20180112_06:40:37:8678, CT20_20180112_07:55:33:5254...
X85040VDY3	CT89_20180112_06:45:43:4425, CT90_20180112_06:45:59:9888, CT89_20180112_06:51:63:6895, CT90_20180112_07:23:55:6329...

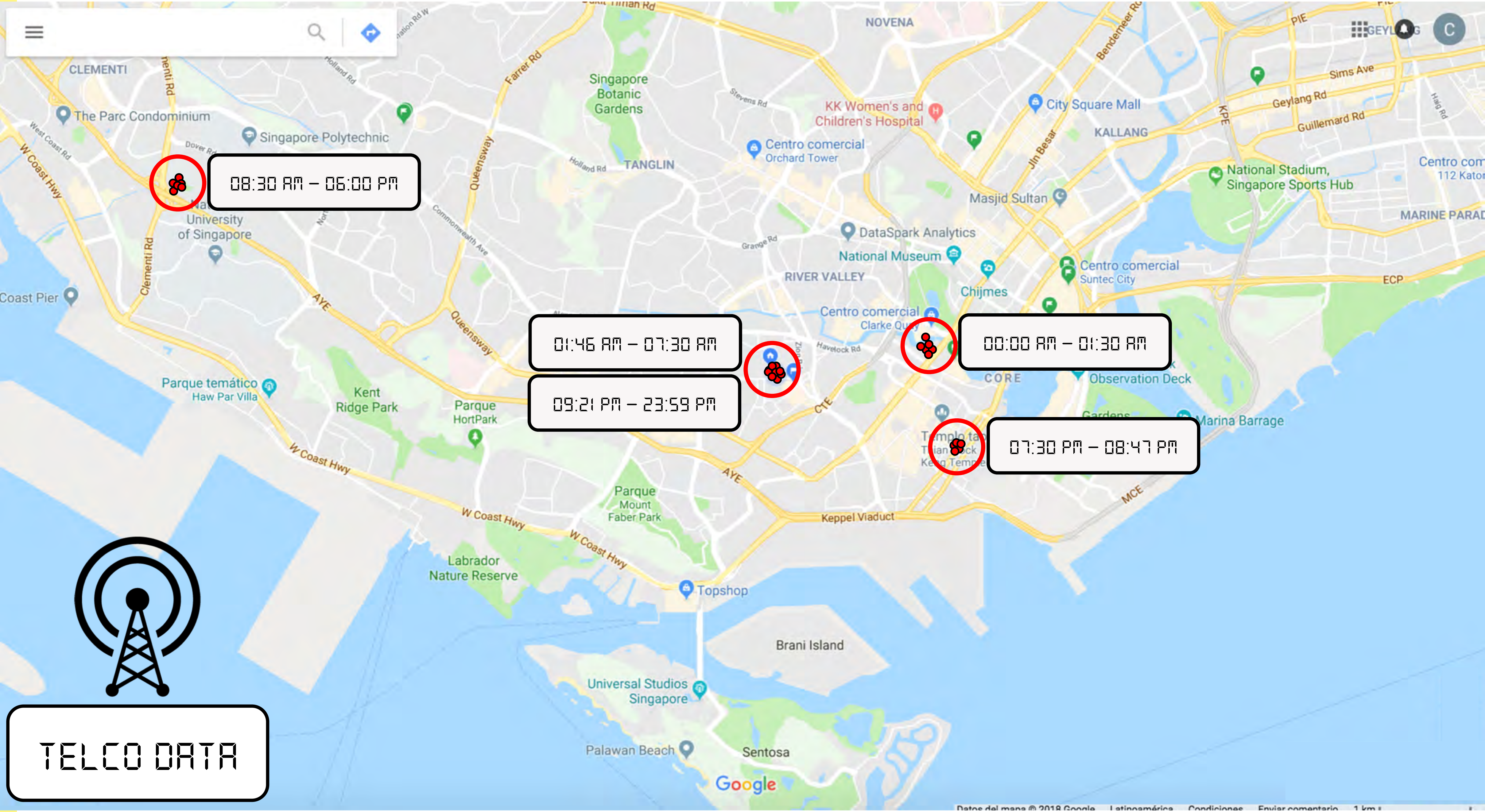
Is it safe?

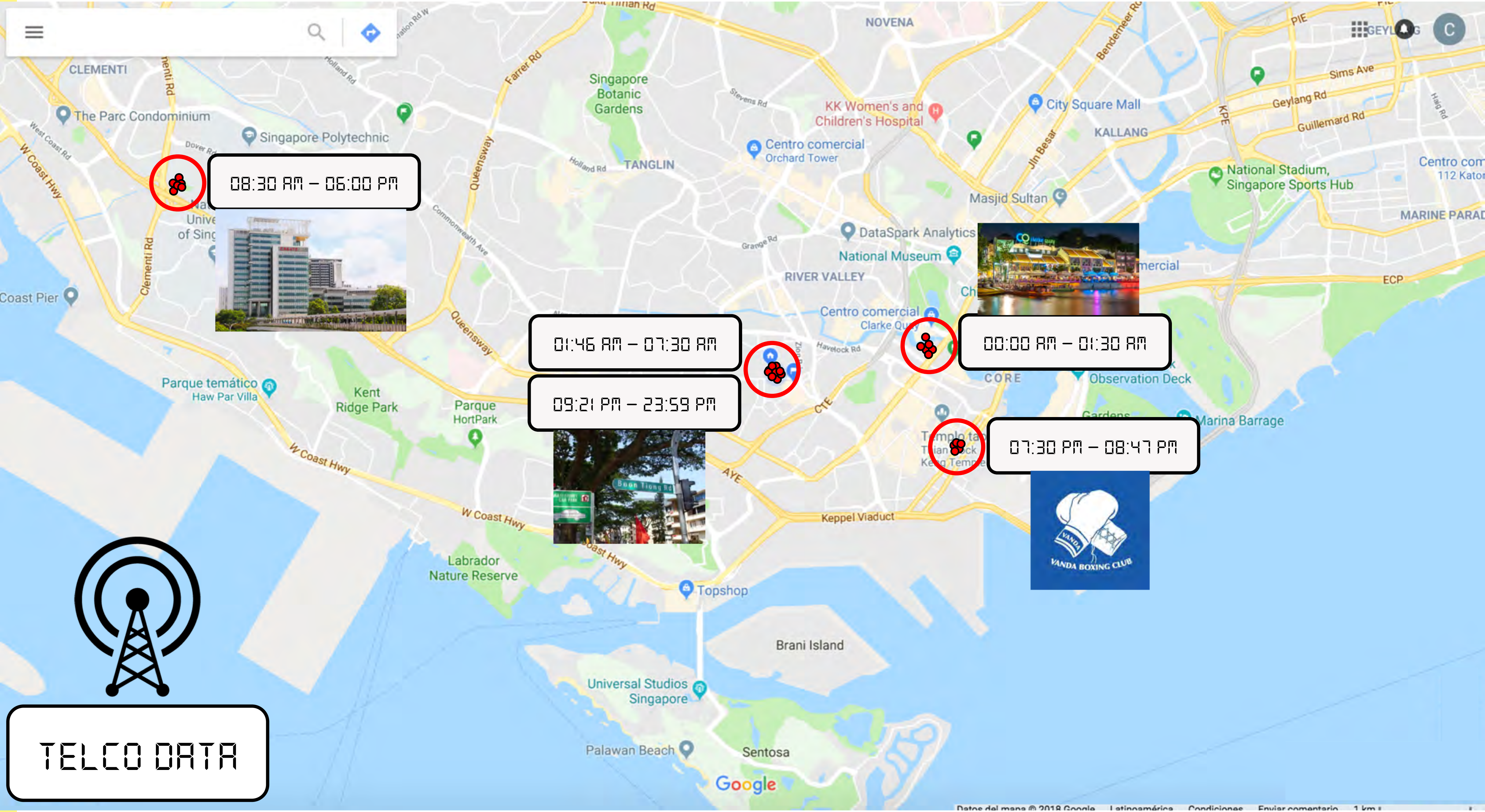
Mobile phone data:
Let’s take a closer look

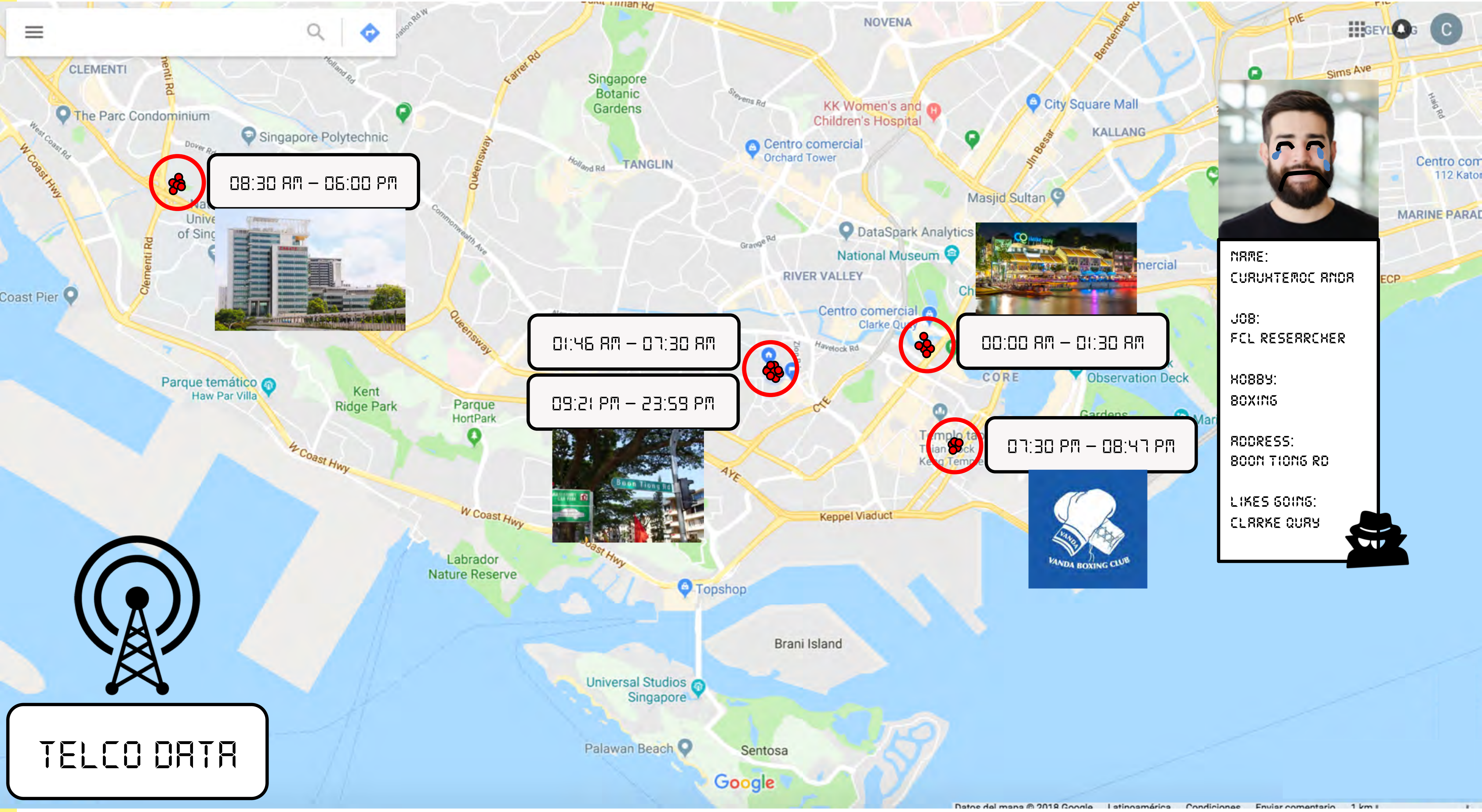

10:00 PM
200 M
ACCURACY
50%
MARKET
SHARE
TELCO DATA
CLEMENTI
The Parc Condominium
Singapore Polytechnic
National University of Singapore
Singapore Botanic Gardens
TANGLIN
NOVENA
KK Women's and Children's Hospital
Centro comercial Orchard Tower
City Square Mall
KALLANG
National Stadium, Singapore Sports Hub
Masjid Sultan
DataSpark Analytics
National Museum
RIVER VALLEY
Chijmes
Centro comercial Suntec City
Centro comercial Clarke Quay
National Gallery Singapore
DOWNTOWN CORE
Sands SkyPark Observation Deck
Gardens by the Bay
Marina Barrage
MARINE PARADE
Templo taoísta Thian Hock Keng Temple
BUKIT MERAH
Jalan Bukit Merah
Kent Ridge Park
Parque HortPark
Parque temático Haw Par Villa
Parque Mount Faber Park
Labrador Nature Reserve
Keppel Viaduct
Topshop
Brani Island
Universal Studios Singapore
Palawan Beach
Sentosa
Coast Pier
AYE
CTE
ECP
MCE
PIE
W Coast Hwy
Queensway
Farrer Rd
Alexandra Rd
Commonwealth Ave
Holland Rd
Dover Rd
Clementi Rd
Google
Datos del mapa ©2018 Google



CLEMENTI
The Parc Condominium
Singapore Polytechnic
Dover Rd
West Coast Rd
W Coast Hwy
National University of Singapore
Clementi Rd
Coast Pier
AYE
Farrer Rd
Holland Rd
Singapore Botanic Gardens
TANGLIN
Queensway
North Buona Vista Rd
Commonwealth Ave
Stevens Rd
NOVENA
KK Women's and Children's Hospital
Centro comercial Orchard Tower
Grange Rd
DataSpark Analytics
National Museum
RIVER VALLEY
Alexandra Rd
Centro comercial Clarke Quay
Chijmes
National Gallery Singapore
Havelock Rd
CTE
DOWNTOWN CORE
Bendemeer Rd
City Square Mall
Jln Besar
Masjid Sultan
KALLANG
Sims Ave
Geylang Rd
Guillemard Rd
KPE
PIE
National Stadium, Singapore Sports Hub
MARINE PARADE
Centro comercial Suntec City
ECP
Sands SkyPark Observation Deck
Gardens by the Bay
Marina Barrage
MCE
Parque temático Haw Par Villa
Kent Ridge Park
Parque HortPark
Jalan Bukit Merah
BUKIT MERAH
Parque Mount Faber Park
Keppel Viaduct
Labrador Nature Reserve
Topshop
Brani Island
Universal Studios Singapore
Palawan Beach
Sentosa
Google
Datos del mapa © 2018 Google
Latinoamérica
Condiciones
Enviar comentario
1 km
TELCO DATA


08:30 AM – 06:00 PM
01:46 AM – 07:30 AM
09:21 PM – 23:59 PM
00:00 AM – 01:30 AM
07:30 PM – 08:47 PM
TELCO DATA
CLEMENTI
The Parc Condominium
Singapore Polytechnic
University of Singapore
Singapore Botanic Gardens
TANGLIN
NOVENA
KK Women's and Children's Hospital
Centro comercial Orchard Tower
City Square Mall
KALLANG
Masjid Sultan
National Stadium, Singapore Sports Hub
DataSpark Analytics
National Museum
RIVER VALLEY
Chijmes
Centro comercial Suntec City
Centro comercial Clarke Quay
CORE
Observation Deck
Marina Barrage
Parque temático Haw Par Villa
Kent Ridge Park
Parque HortPark
Parque Mount Faber Park
Labrador Nature Reserve
Keppel Viaduct
Topshop
Brani Island
Universal Studios Singapore
Palawan Beach
Sentosa
Google


08:30 AM – 06:00 PM
01:46 AM – 07:30 AM
09:21 PM – 23:59 PM
00:00 AM – 01:30 AM
07:30 PM – 08:47 PM
TELCO DATA
VANDA BOXING CLUB
CLEMENTI
The Parc Condominium
Singapore Polytechnic
Singapore Botanic Gardens
TANGLIN
NOVENA
KK Women's and Children's Hospital
Centro comercial Orchard Tower
City Square Mall
KALLANG
National Stadium, Singapore Sports Hub
Masjid Sultan
DataSpark Analytics
National Museum
RIVER VALLEY
Centro comercial Clarke Quay
CORE
Observation Deck
Marina Barrage
MARINE PARADE
Parque temático Haw Par Villa
Kent Ridge Park
Parque HortPark
Labrador Nature Reserve
Topshop
Keppel Viaduct
Brani Island
Universal Studios Singapore
Palawan Beach
Sentosa
Google


08:30 AM – 06:00 PM
01:46 AM – 07:30 AM
09:21 PM – 23:59 PM
00:00 AM – 01:30 AM
07:30 PM – 08:47 PM
NAME:
CUAUHTEMOC ANDA
JOB:
FCL RESEARCHER
HOBBY:
BOXING
ADDRESS:
BOON TIONG RD
LIKES GOING:
CLARKE QUAY
TELCO DATA
VANDA BOXING CLUB
Boon Tiong Rd

Mobile phone telco data

And Privacy

"… in a dataset where the **location of an individual** is specified **hourly**, and with a **spatial resolution** equal to that given by the **carrier's antennas**, **four** **spatio-temporal points** are enough to **uniquely** **identify 95%** of the **individuals**."

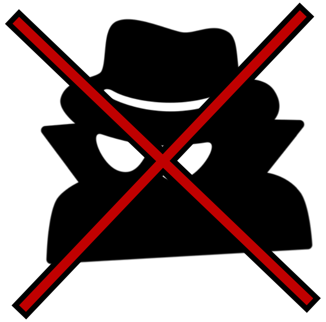
de Montjoye, Y.-A., C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific reports*, Vol. 3, 2013, p. 1376. https://doi.org/10.1038/srep01376.

What to do then?

Mobile phone telco data is very **valuable** for transport planning…

… Can we extract useful information without compromising users' privacy?

(Easy-sharing) (Less Cost)



Big question for Big Data

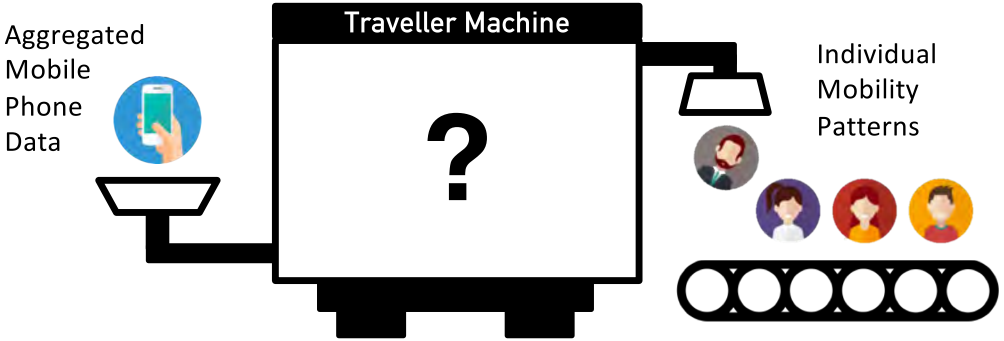
The Doppelgänger idea

The Doppelgänger idea

Keanu Reaves
(Matrix)

Adam Driver
(Star Wars)

We are interested in a detailed description of agents for MATSim…



<u>Task</u>

Without consulting **individual data generate** realistic individual **mobility** patterns

Generating a doppelgänger population

Generative model of Urban Mobility

An intuitive example

Real population weight

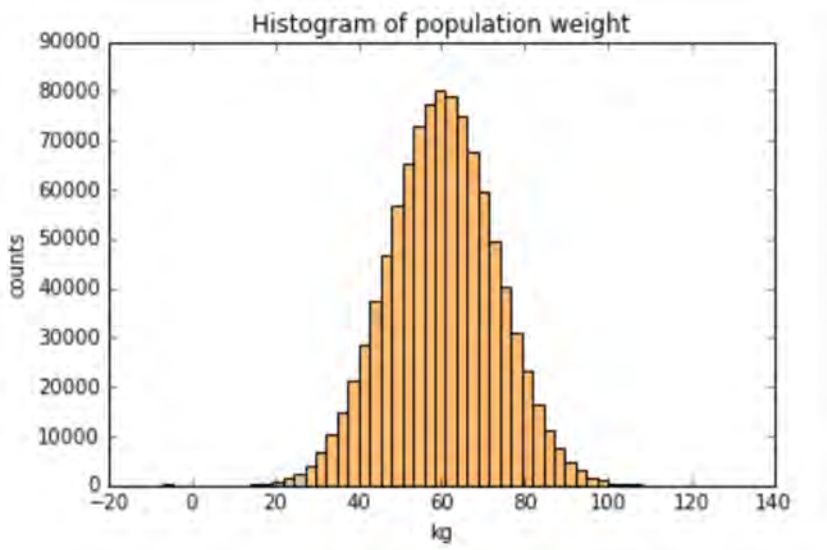
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...



Intuitive example:

Weight of citizens of Shanghai

Real population weight

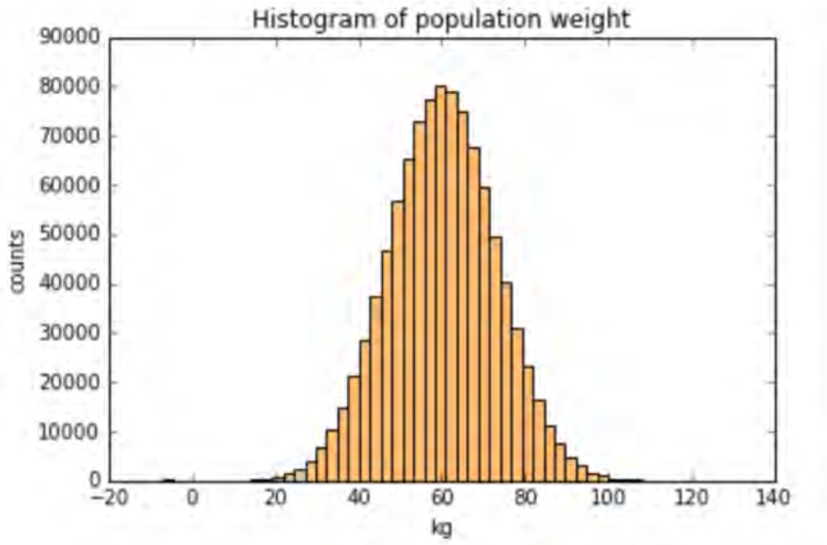
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...




```
In [ ]: np.random.normal()
```

Intuitive example:

Weight of citizens of Shanghai

Real population weight

50.11532168513191

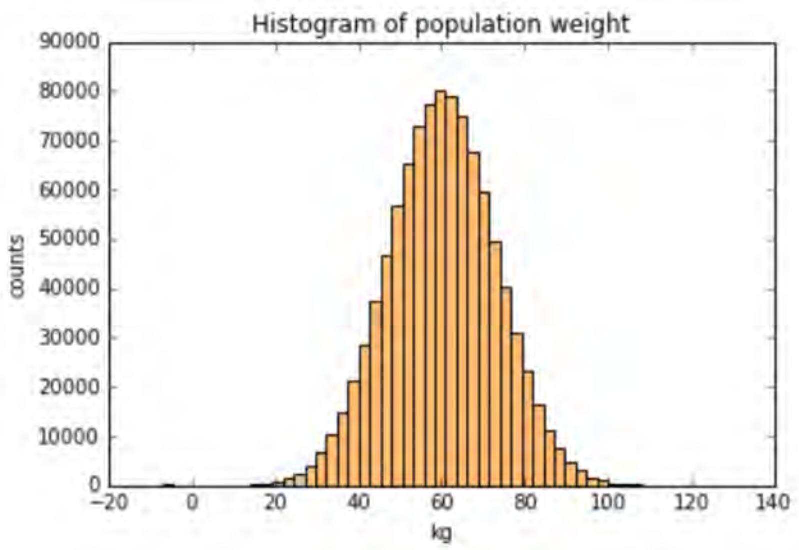
48.65649004247139

61.403481911621405

59.1568694908791

...

```
In [2]: np.random.normal()
Out[2]: 63.29561731708119
```



Intuitive example:

Weight of citizens of Shanghai

Real population weight

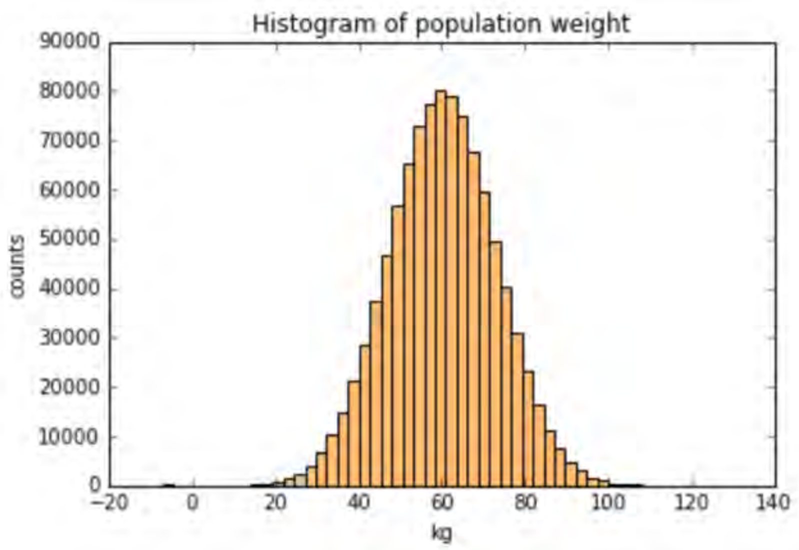
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...




```
In [3]: np.random.normal()
Out[3]: 67.25242342967776
```

Intuitive example:

Weight of citizens of Shanghai

Real population weight

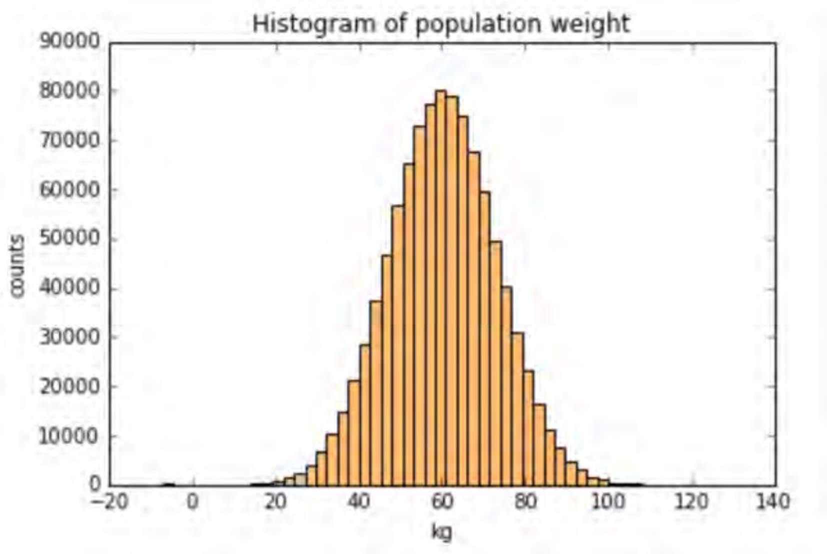
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...




```
In [4]: np.random.normal()
Out[4]: 52.30204118306012
```

Intuitive example:

Weight of citizens of Shanghai

Real population weight

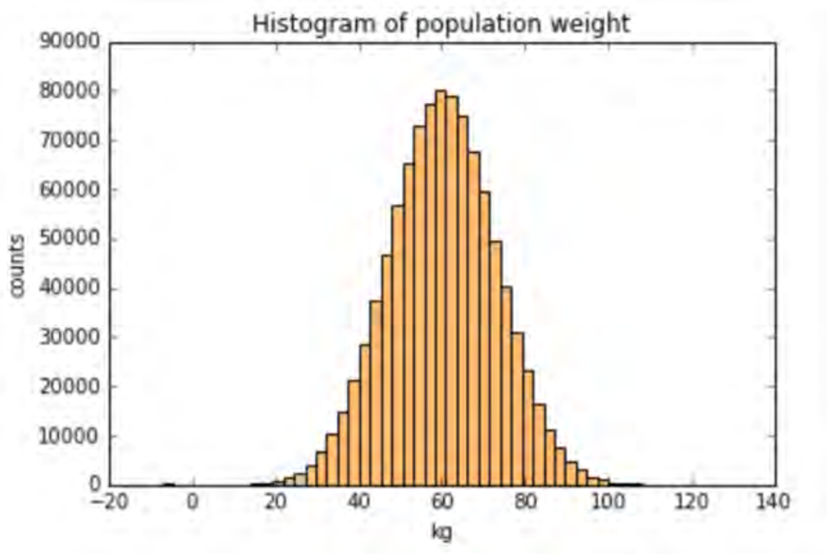
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...



Intuitive example:

Weight of citizens of Shanghai

```
In [5]: np.random.normal()
Out[5]: 46.52883437707794
```

Real population weight

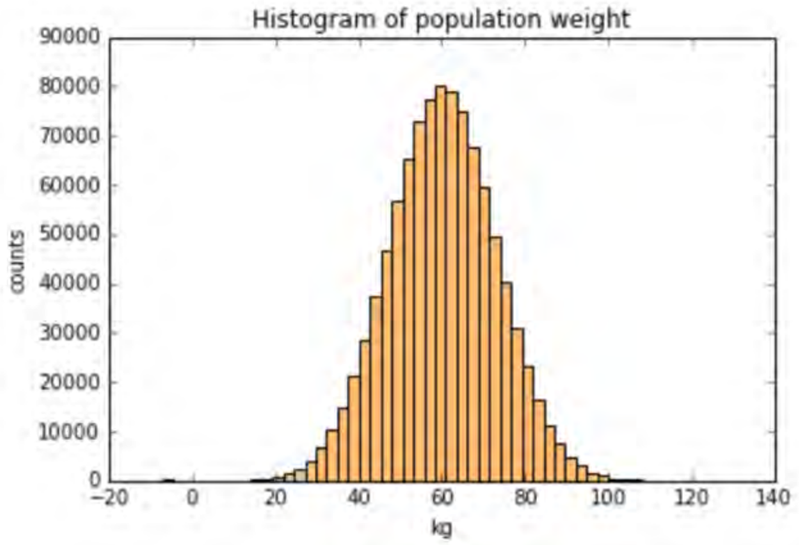
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...



Doppelgängers weight

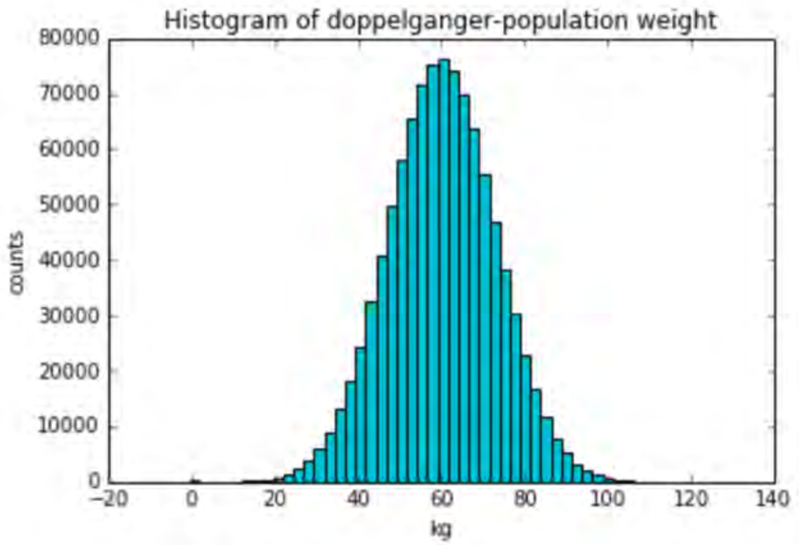
63.29561731708119

67.25242342967776

52.30204118306012

46.52883437707794

...



Intuitive example:

Weight of citizens of Shanghai

Real population weight

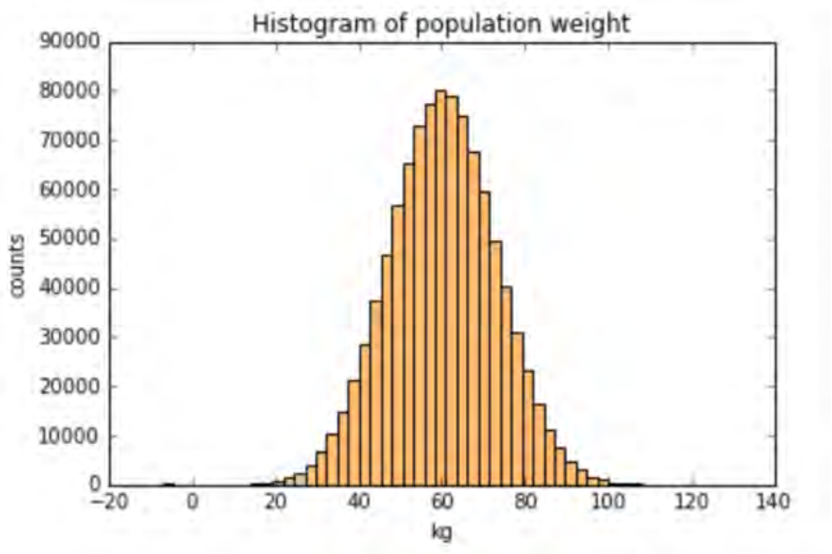
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...




```
In [ ]: GenerateAgentLocations()
```

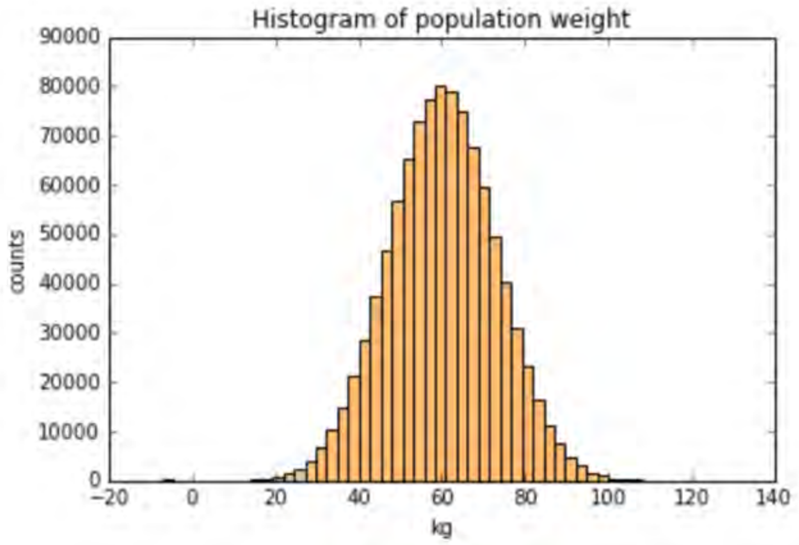
Generative model of urban mobility

Generative model of urban mobility

Real population weight

```
50.11532168513191
48.65649004247139
61.403481911621405
59.1568694908791
```

...




```
In [11]: GenerateAgentLocations(1)
```

Out[11]:

	User	Start Time	Subzone	Duration	End Time
0	u_1	00:00:00	KEAT HONG	8	08:00:00
1	u_1	08:30:00	CENTRAL SUBZONE	14	22:30:00
2	u_1	23:30:00	KEAT HONG	12	36:00:00

Generative model of urban mobility

Real population weight

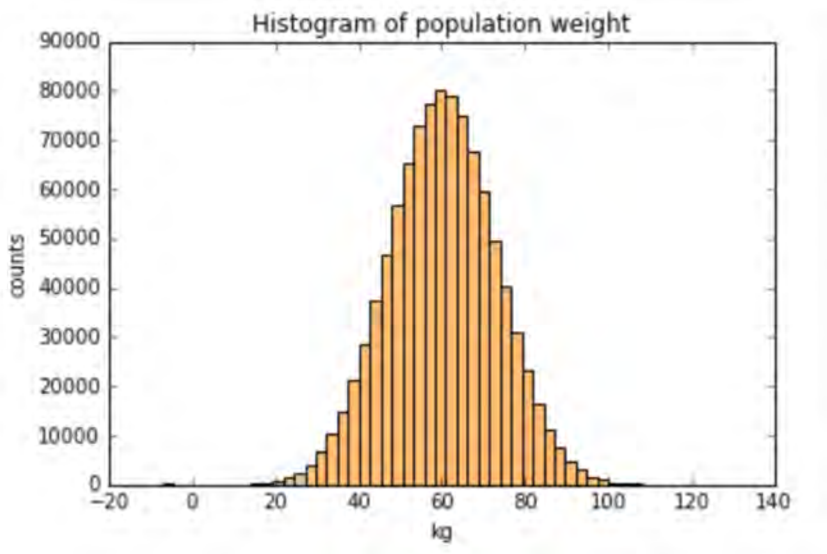
50.11532168513191

48.65649004247139

61.40348191162140 5

59.156869490879 1

...




```
In [12]: GenerateAgentLocations(1)
```

Out[12]:

	User	Start Time	Subzone	Duration	End Time
0	u_1	00:00:00	YUNNAN	7	07:00:00
1	u_1	07:30:00	TUAS NORTH	2	09:30:00
2	u_1	10:30:00	TAI SENG	1	12:00:00
3	u_1	13:00:00	YUNNAN	16	29:30:00

Real population weight

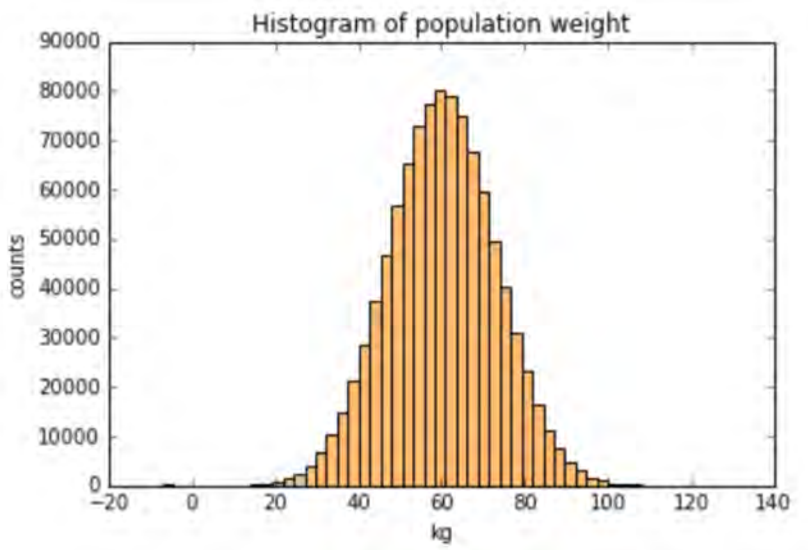
50.11532168513191

48.65649004247139

61.403481911621405

59.1568694908791

...




```
In [13]: GenerateAgentLocations(1)
```

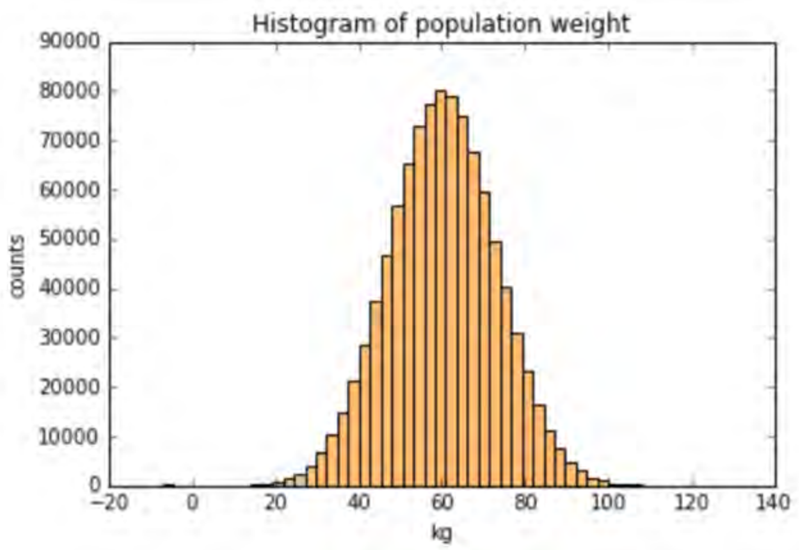
Out[13]:

	User	Start Time	Subzone	Duration	End Time
0	u_1	00:00:00	FLORA DRIVE	4	04:00:00
1	u_1	05:00:00	CHANGI AIRPORT	1	06:00:00
2	u_1	07:00:00	MARINA EAST	3	10:30:00
3	u_1	11:30:00	KALLANG BAHRU	6	18:00:00
4	u_1	19:00:00	FLORA DRIVE	11	30:30:00

Generative model of urban mobility

Real population weight

```
50.11532168513191
48.65649004247139
61.40348191162140 5
59.156869490879 1
...
```

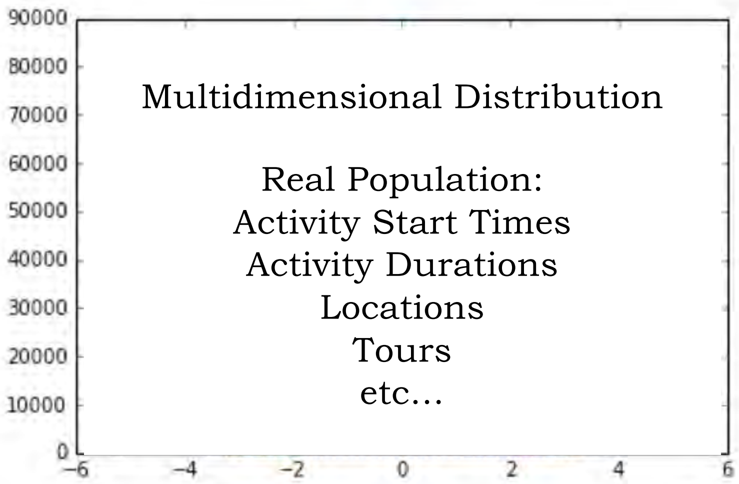



```
In [14]: GenerateAgentLocations(1)
Out[14]:
```

	User	Start Time	Subzone	Duration	End Time
0	u_1	00:00:00	GEYLANG EAST	7	07:00:00
1	u_1	07:30:00	SINGAPORE GENERAL HOSPITAL	11	18:30:00
2	u_1	19:00:00	GEYLANG EAST	3	22:30:00
3	u_1	22:45:00	GEYLANG EAST	13	36:15:00

Generative model of urban mobility

Generative model of urban mobility




```
In [14]: GenerateAgentLocations(1)
Out[14]:
```

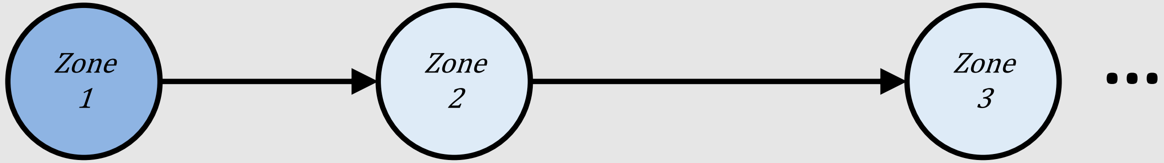
	User	Start Time	Subzone	Duration	End Time
0	u_1	00:00:00	GEYLANG EAST	7	07:00:00
1	u_1	07:30:00	SINGAPORE GENERAL HOSPITAL	11	18:30:00
2	u_1	19:00:00	GEYLANG EAST	3	22:30:00
3	u_1	22:45:00	GEYLANG EAST	13	36:15:00

Generative model of individual mobility patterns

[Dynamic Bayesian Networks]

Generative Model for Sequential Data **Dynamic Bayesian Networks**

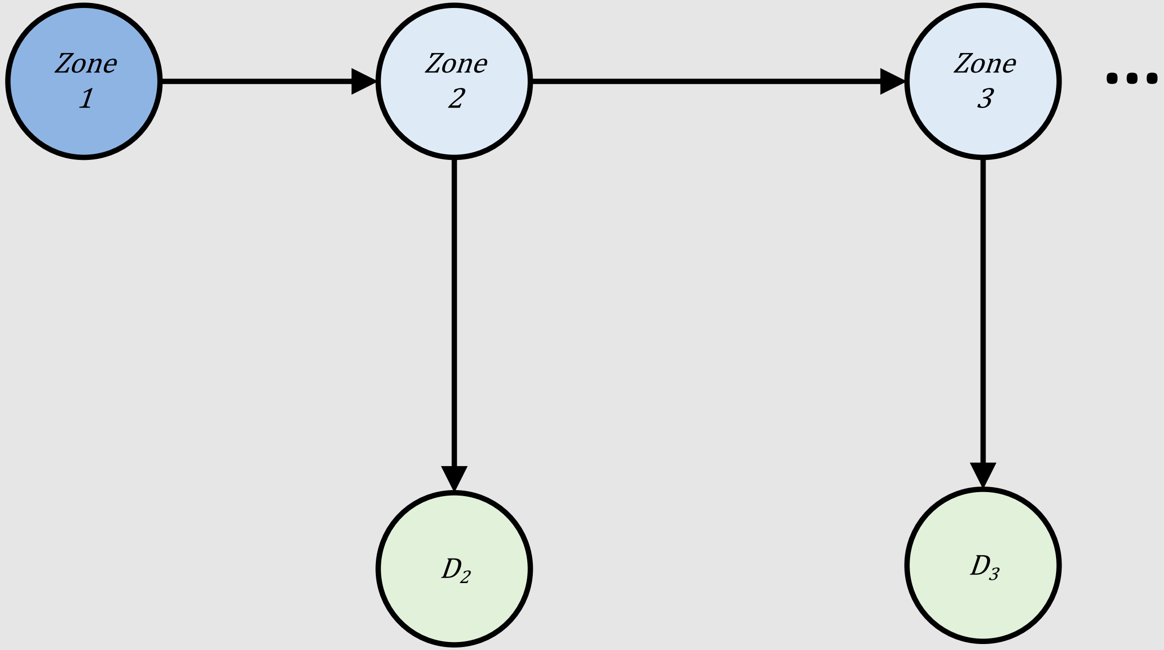
Markov Chain



$$\boldsymbol{P}(\boldsymbol{Z}_{\mathbf{1}:\boldsymbol{N}}) = P(Z_1)\prod_{k=2}^{N} P(Z_k|Z_{k-1})$$

Generative Model for Sequential Data
Dynamic Bayesian Networks

Hidden Markov Model

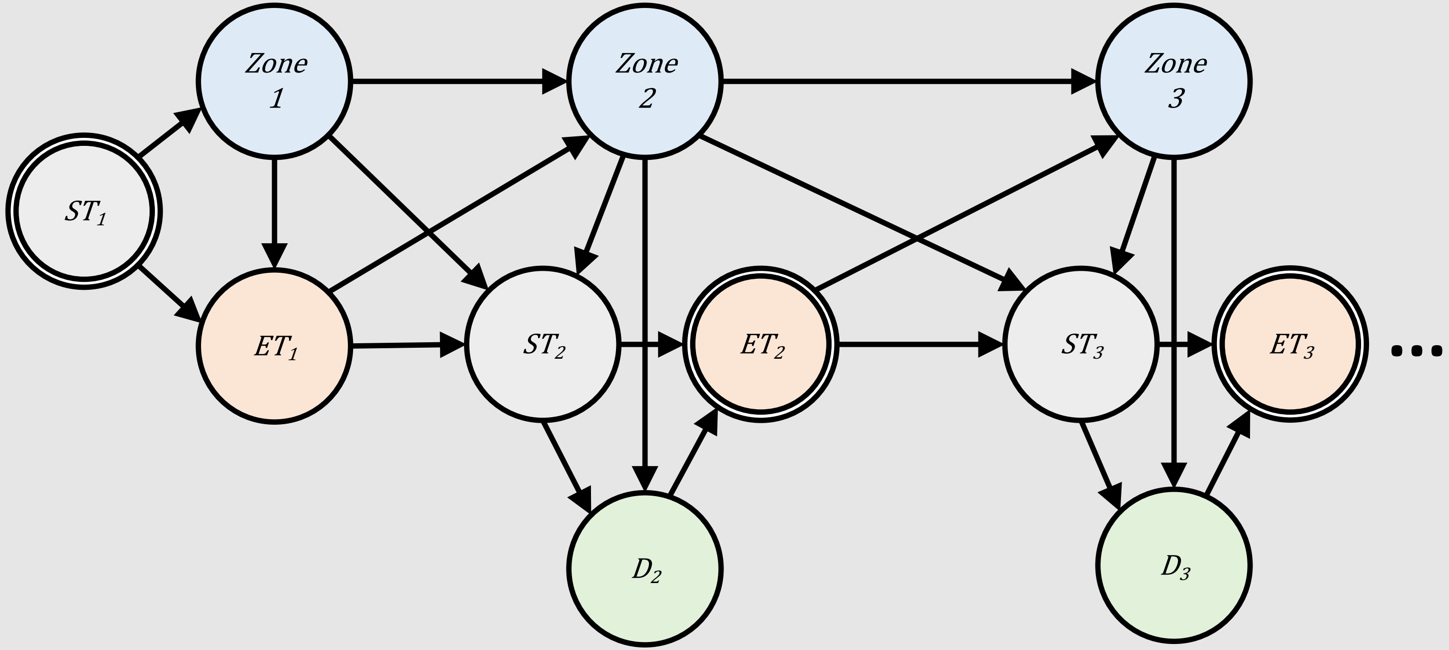


$$\boldsymbol{P}(\boldsymbol{Z}_{\mathbf{1}:\boldsymbol{N}}, \boldsymbol{D}_{\mathbf{2}:\boldsymbol{N}}) = P(Z_1) \prod_{k=2}^{N} P(Z_k|Z_{k-1})P(D_k|Z_k)$$

Generative Model for Sequential Data **Dynamic Bayesian Networks**

Our architecture 1.0

1st order Markov constraint

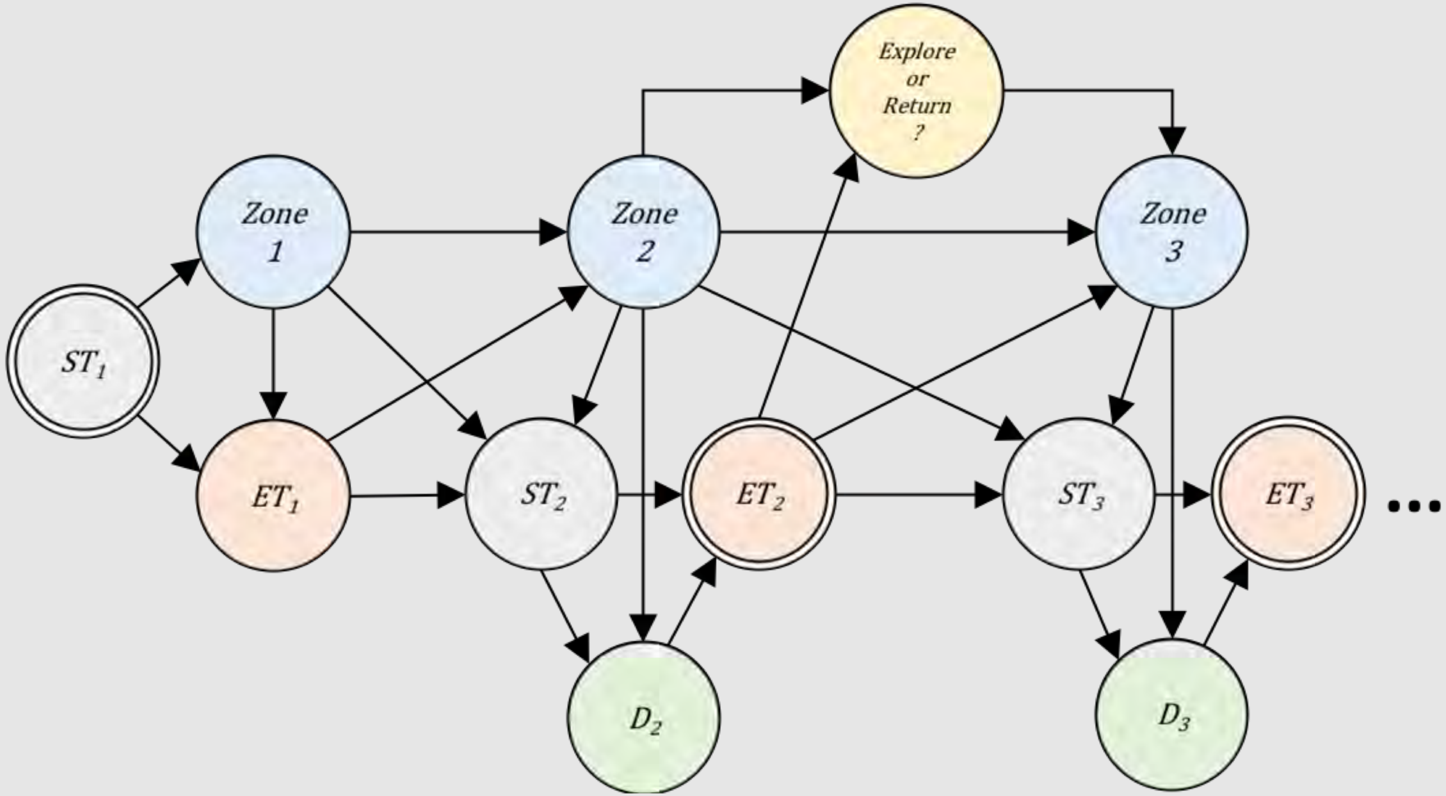


$$\boldsymbol{P(Z_{1:N}, ET_{1:N}, ST_{2:N}, D_{2:N})} =$$

$$P(ST_1)P(Z_1|ST_1\ )P(ET_1|ST_1, Z_1)\prod_{k=2}^{N} P(Z_k|ET_{k-1}, Z_{k-1})P(D_k|Z_k, ST_k)P(ST_k|Z_{k-1}, ET_{k-1})$$

Generative Model for Sequential Data **Dynamic Bayesian Networks**

Our architecture 2.0



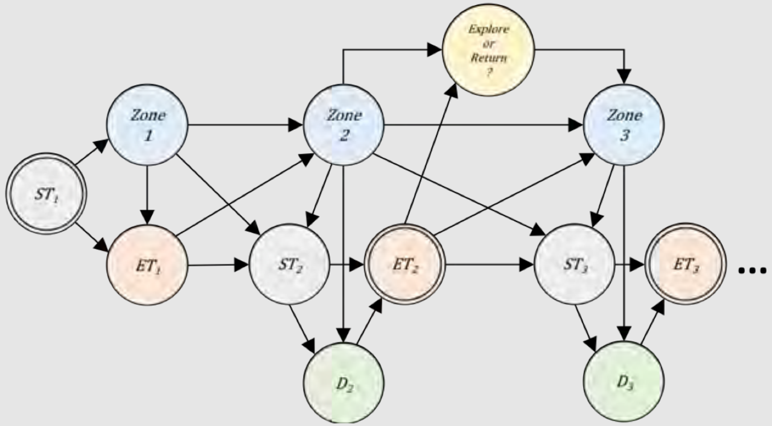
$$\boldsymbol{P(Z_{1:N}, ET_{1:N}, ST_{2:N}, D_{2:N}, ER_{2:N}) =}$$

$$P(ST_1)P(Z_1|ST_1\ )P(ET_1|ST_1,Z_1)\prod_{k=2}^{N}P(Z_k|ET_{k-1},Z_{k-1},ER_{k-1})P(D_k|Z_k,ST_k)P(ST_k|Z_{k-1},ET_{k-1})P(ER_k|Z_k,ET_k)$$

Time-space model of Urban Mobility **Variables**

Random Variables

315	**Subzones** (Discrete)	Aljunied Raffles Place Tiong Bahru MRT Changi Airport Bedok Cecil Woodlands …
24	**Start Times** (Discrete)	00:00 01:00 … 23:00
24	**End Time** (Discrete)	00:00 01:00 … 23:00
17	**Durations** (Discrete)	0 hr 1 hr … 16 hr
2	**Explore or Return** (Discrete)	Explore Return



Model parameters

Symbol	Description	# parameters
$\boldsymbol{z1}_i$	probability of starting at zone *i*	315
$\boldsymbol{et1}_k$	probability of first end time *k*	24
$\boldsymbol{z}_{i,j,k}$	probability of transition to zone *i* from zone *j* at time *k* subject to *explore/return* filter	315x315x24
$\boldsymbol{st}_{p,j,k}$	probability of start time *p* given zone *j* at time *k*	24x315x24
$\boldsymbol{d}_{q,i,p}$	probability of duration *q* at zone *i* and time *p*	17x315x24
$\boldsymbol{r}_{q,i,p}$	probability of return *r* from zone *j* and time *k*	2x315x24
	Total # of parameters	**2,706,819**

Time-space model of Urban Mobility **Learning**

Maximum Likelihood Estimation

1. Obtain **Likelihood Function**

   $Likelihood(\boldsymbol{\Theta}) = P(\boldsymbol{Data}|Model(\boldsymbol{\Theta}))$

2. **Minimize negative log-likelihood** subject to parameters of random variables sum to one

3. For **categorical** and fully **observable** random variables a **closed-form solution** is obtained.

4. Learning is the counts of **occurrences** in the data (i.e. frequencies).

Time-space model of Urban Mobility **Learning**

5 Histograms or distributions **needed from the mobile phone operator**

Resolution
Temporal:
1 Hour
Spatial:
Subzones

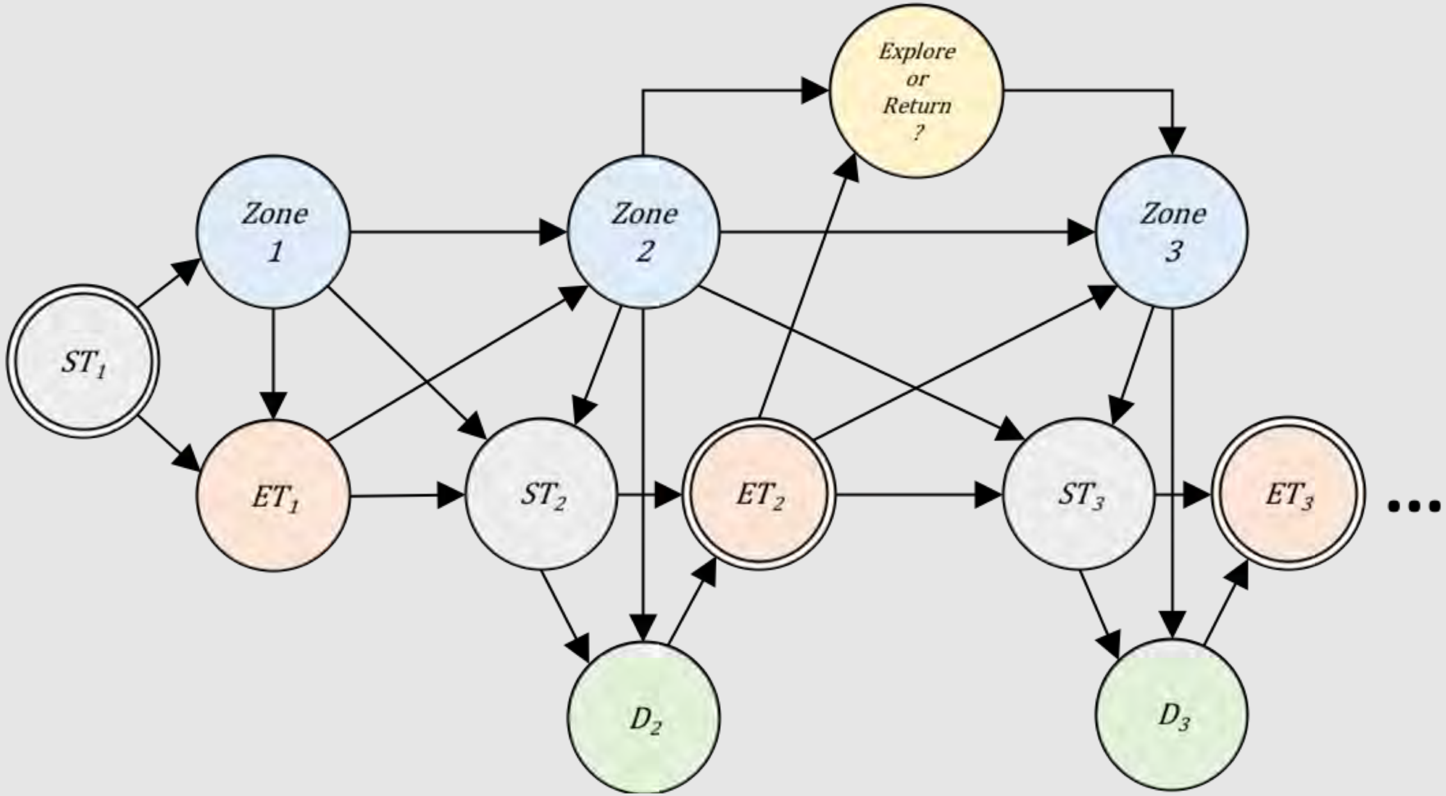
Origin Destination matrices

Initial count of users in each subzone

Activity **Duration** histograms

Time of first departure histogram

Explore or Return histograms

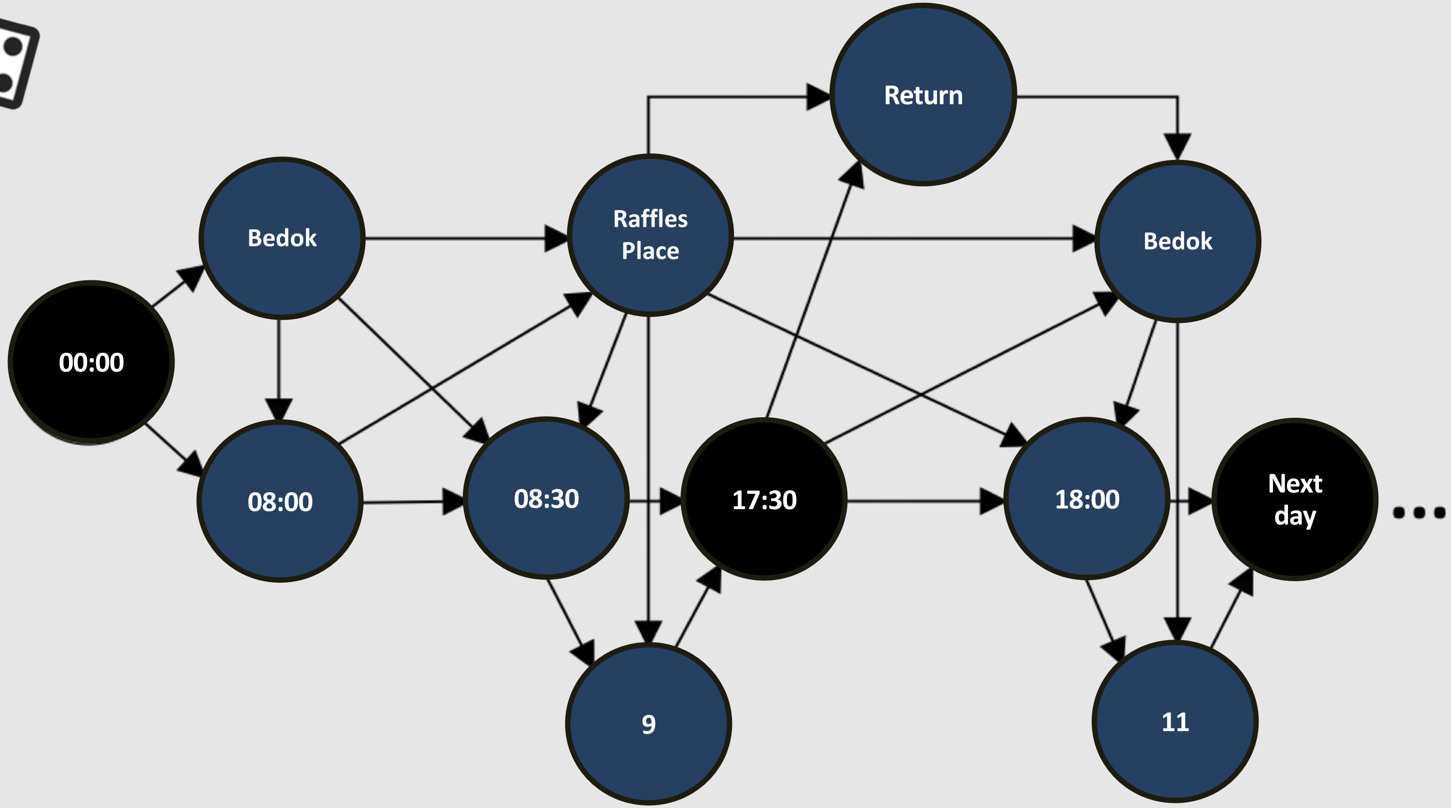


Time-space model of Urban Mobility
Sampling

Prior (Forward) Sampling

Time-space model of Urban Mobility **Sampling**

Prior (Forward) Sampling

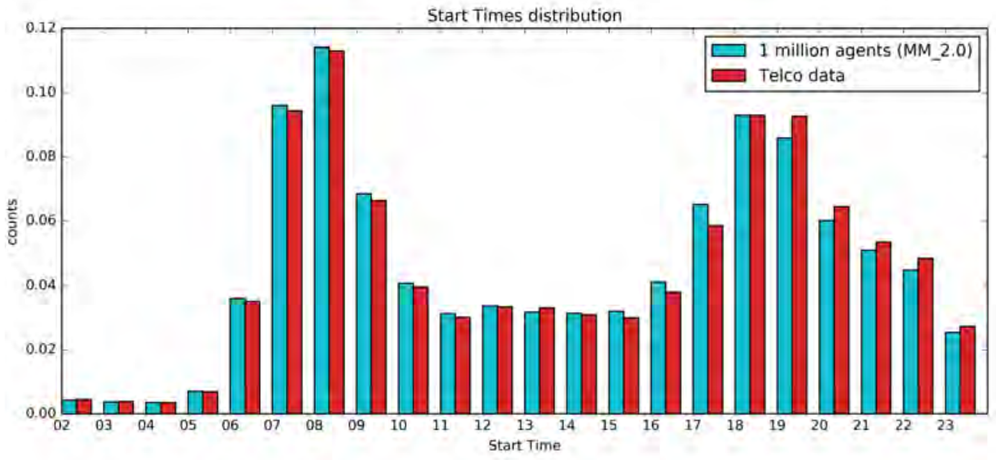


[Generating 1 million agents]

Results

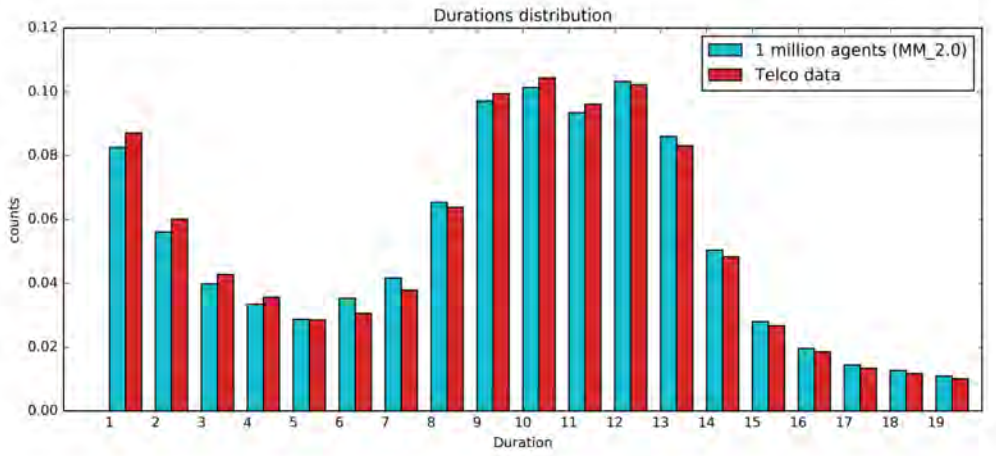
Temporal distributions

Start time of activities



4.18% error

Duration of activities

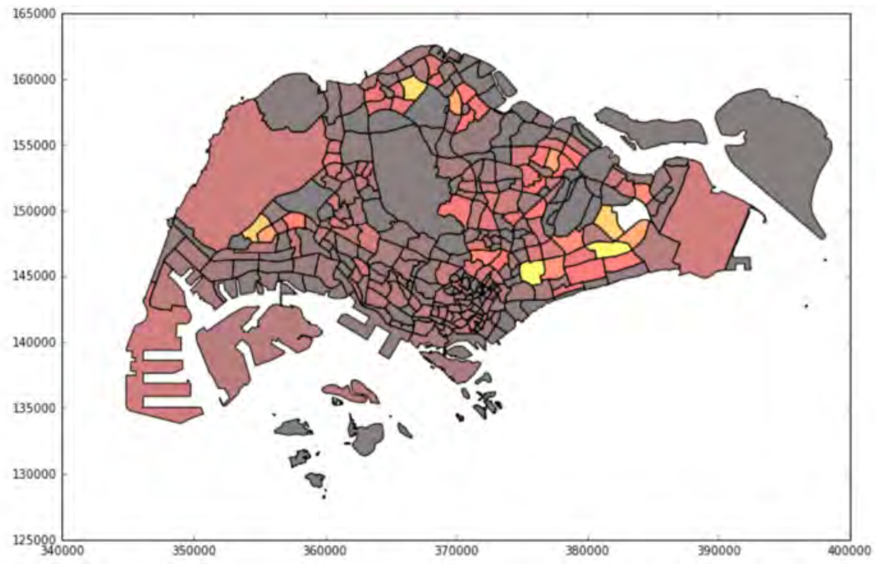


4.32% error

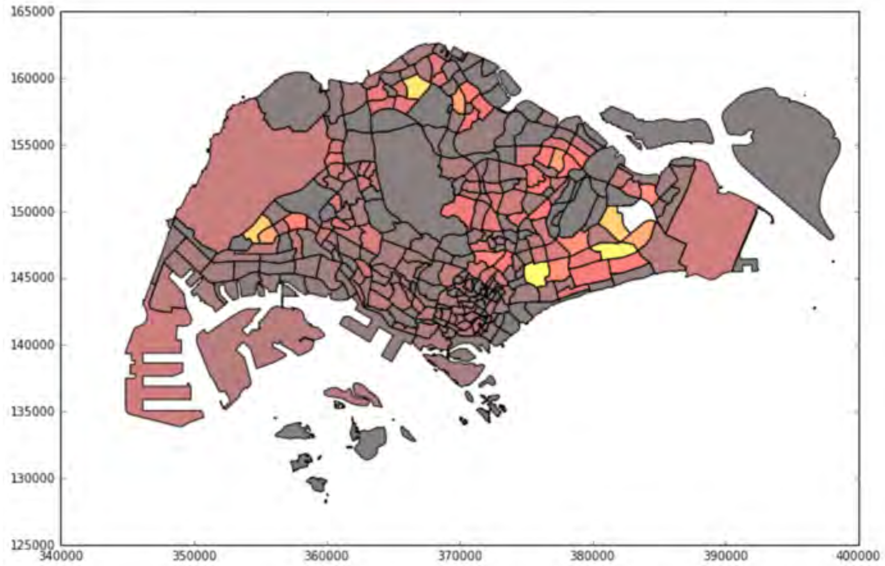
Results

Spatial distributions

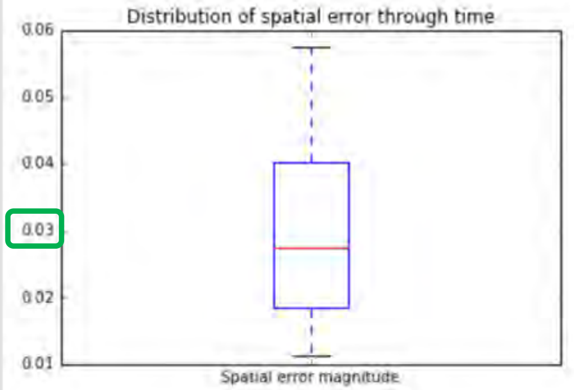
7 am density of **telco users**



7 am density of **agents generated**



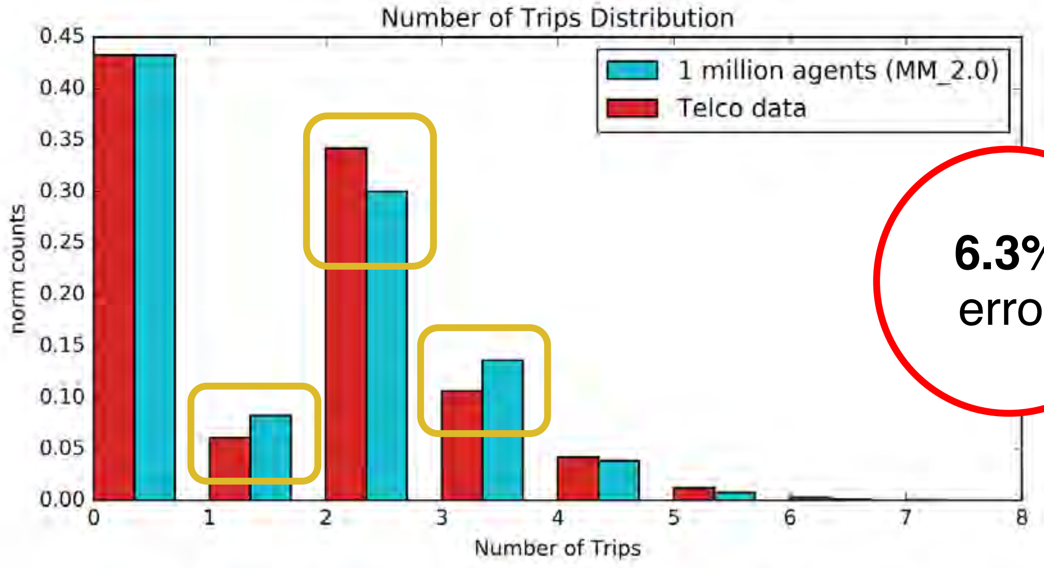
2.92% error



Results

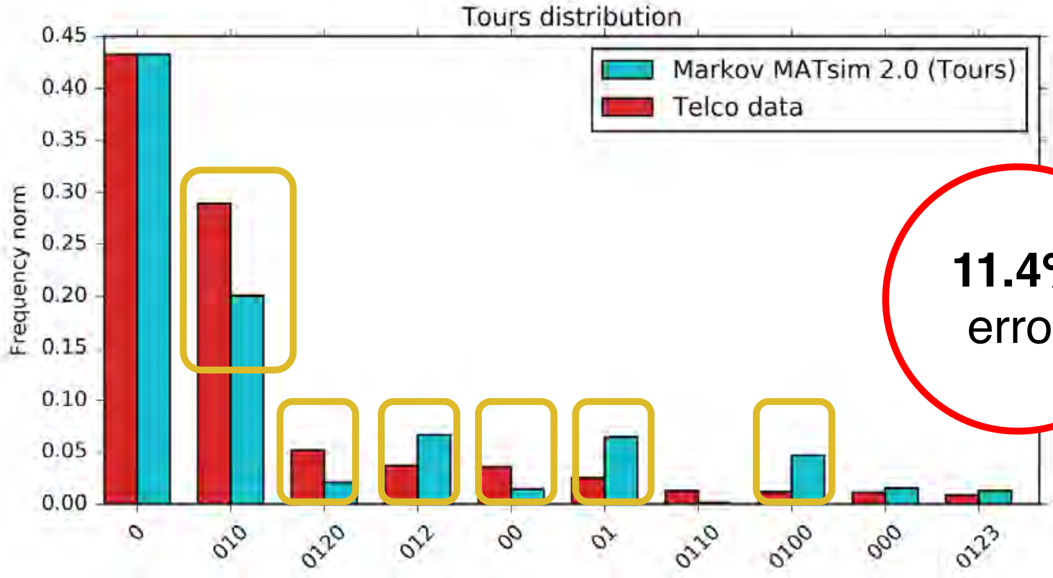
User-specific distributions

Number of trips per person



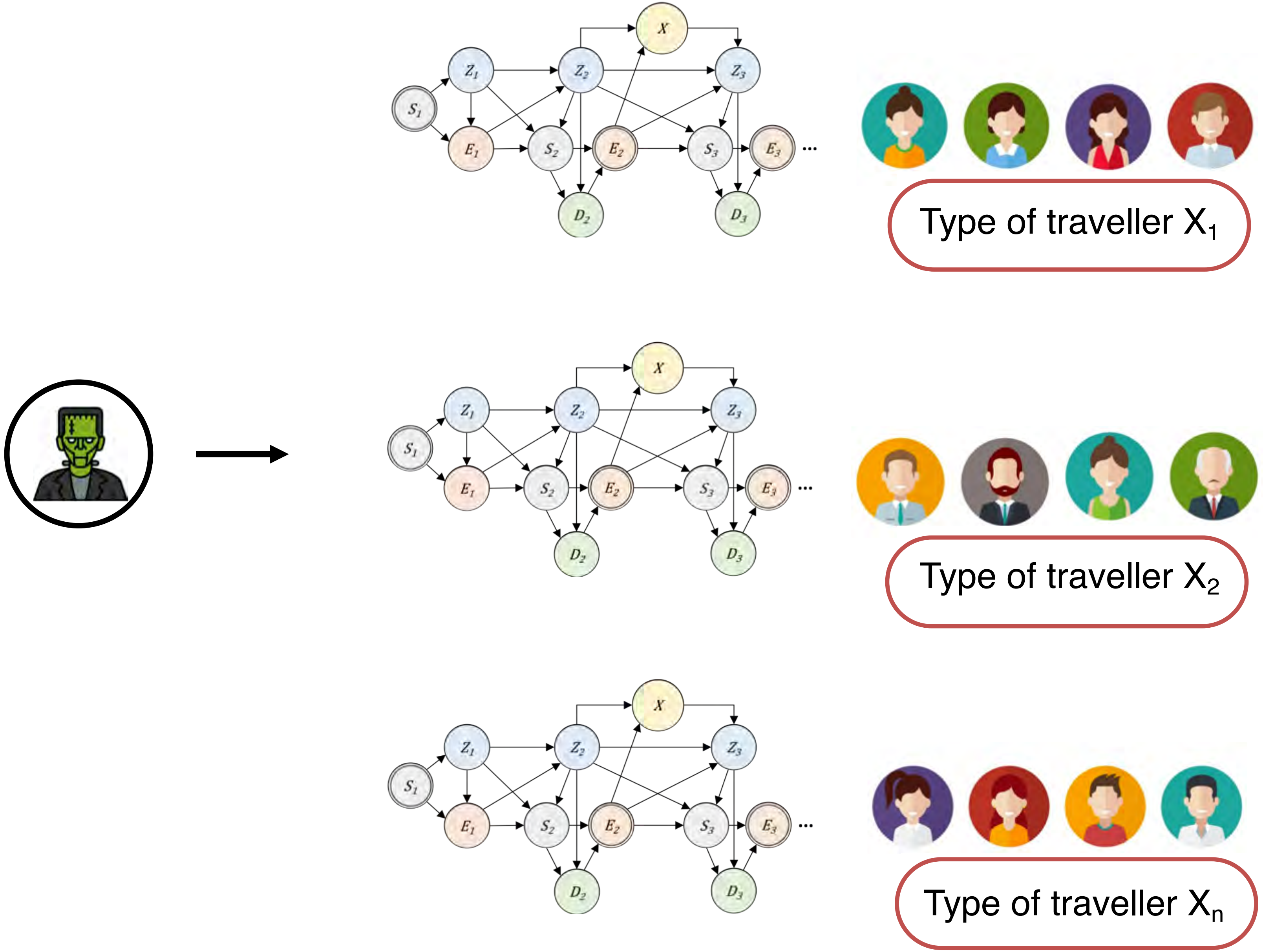
6.3% error

Top 10 daily tour configurations



11.4% error

Problem explanation



Archetypes of Urban Travellers
[in Singapore]

In transport studies, travellers are segmented into groups:

Workers

Students

Non workers

However,

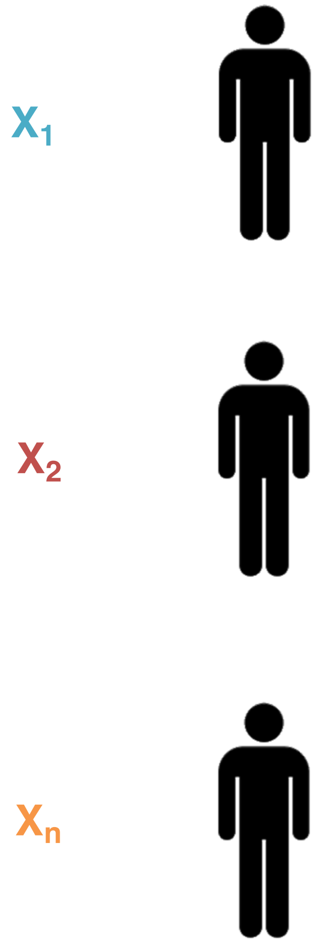
No context / socio-demographics in **mobile phone telco data** (CDR)

In transport studies, travellers are segmented into groups:

However,

No context / socio-demographics in **mobile phone telco data** (CDR)

By **only looking** at the **digital trace** left by mobile phones, how can we **group** the different types of **travellers**?



The idea

1. **Design** a set of **features** that describe the **travel behavior** observed in mobile phone stay locations data

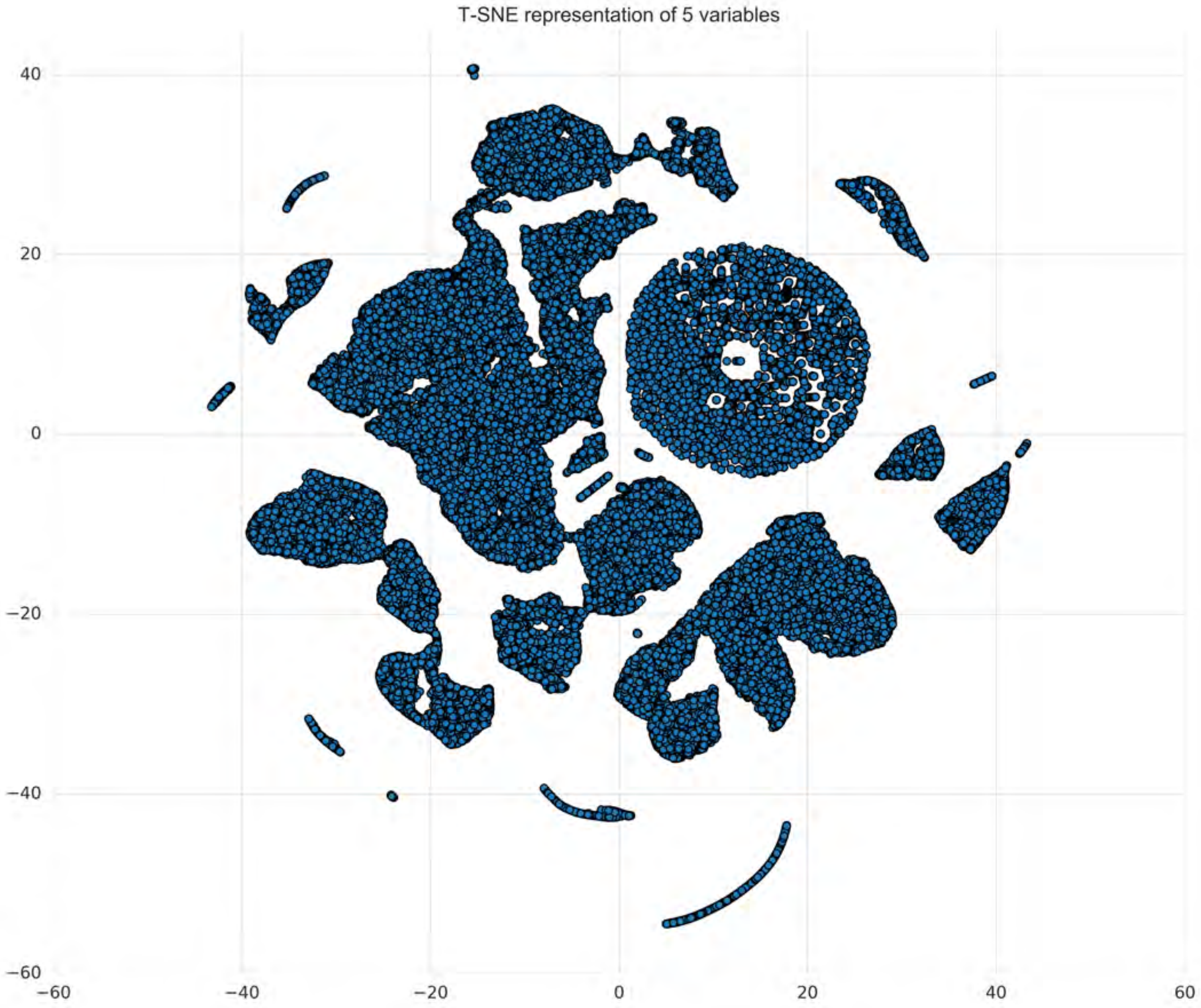
2. Use an appropriate **clustering** algorithm to find the emerging clusters of travellers

3. For **each** of the **clusters** found create a **generative model**, hopefully, the validation histograms will have a better match.

Feature Engineering

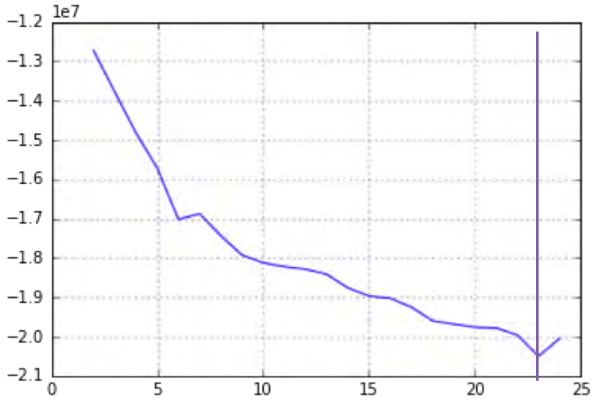
1	Mean of stay durations	The average of all user's stay durations. It gives a notion of the number of trips done by the user	[min]
2	Standard deviation of stay durations	Standard deviation of user's stay durations. Differentiate users with homogenous and non-homogenous activity durations	[min]
3	Bias morning / night	Mean of durations before 12pm – Mean of durations after 12pm Tells if user is a mornig or late traveler	[min]
4	First departure	Time of the day when user makes his/her first trip	[min]
5	Last arrival	Time of the day when user makes his/her last arrival	[min]

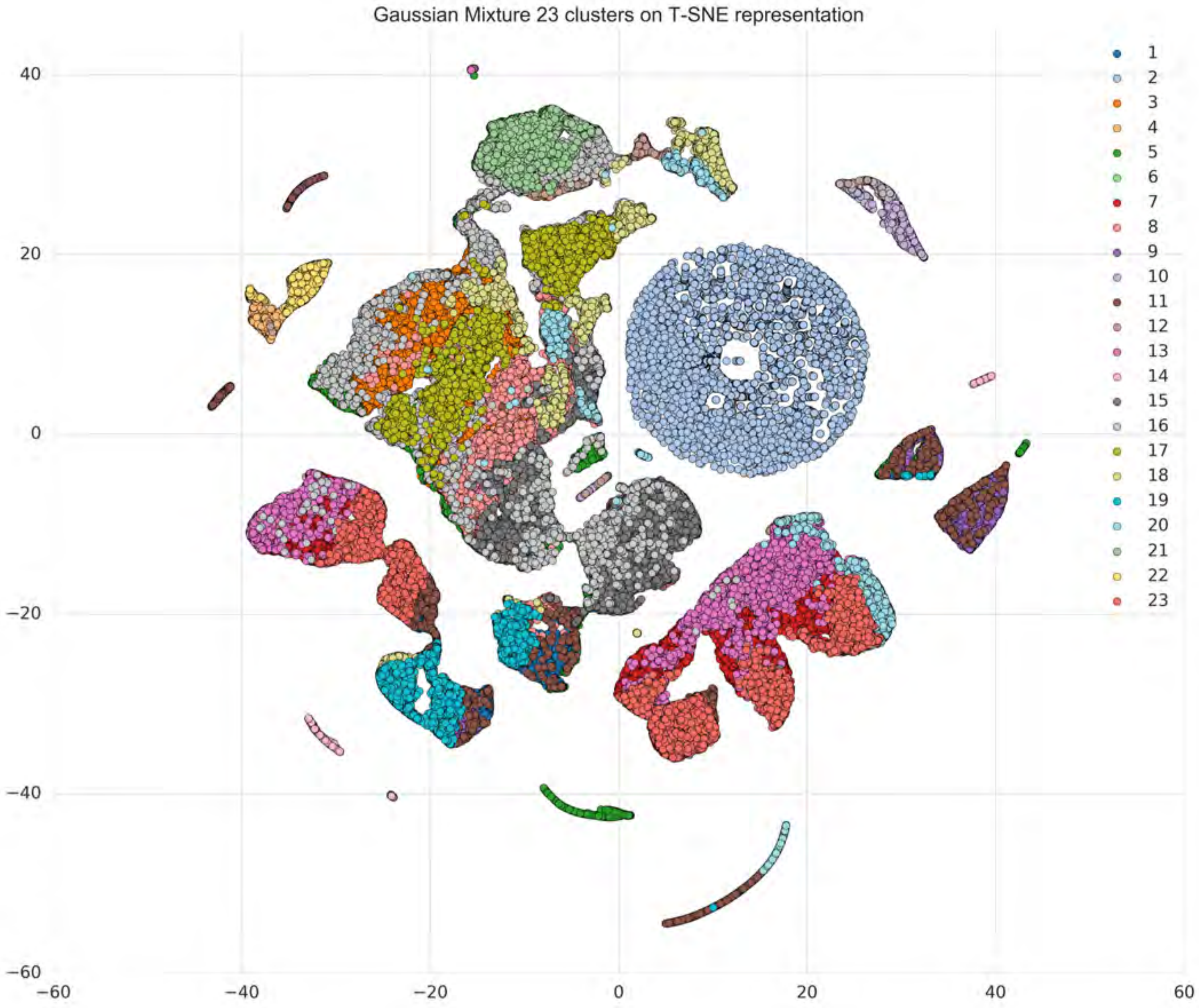
T-SNE representation
(dimensionality reduction)



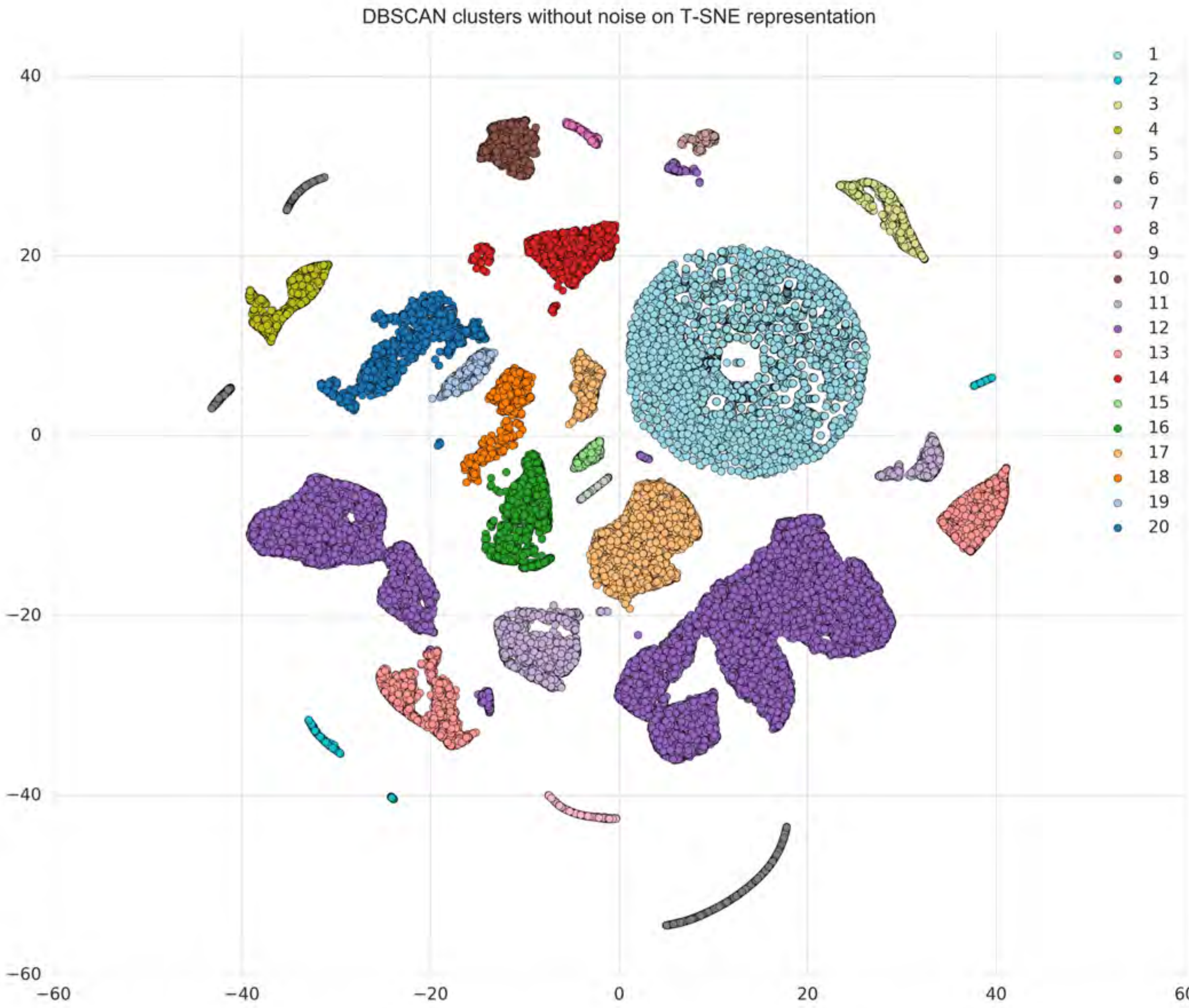
Clustering: Gaussian Mixture

Bayesian Information Criterion (BIC)

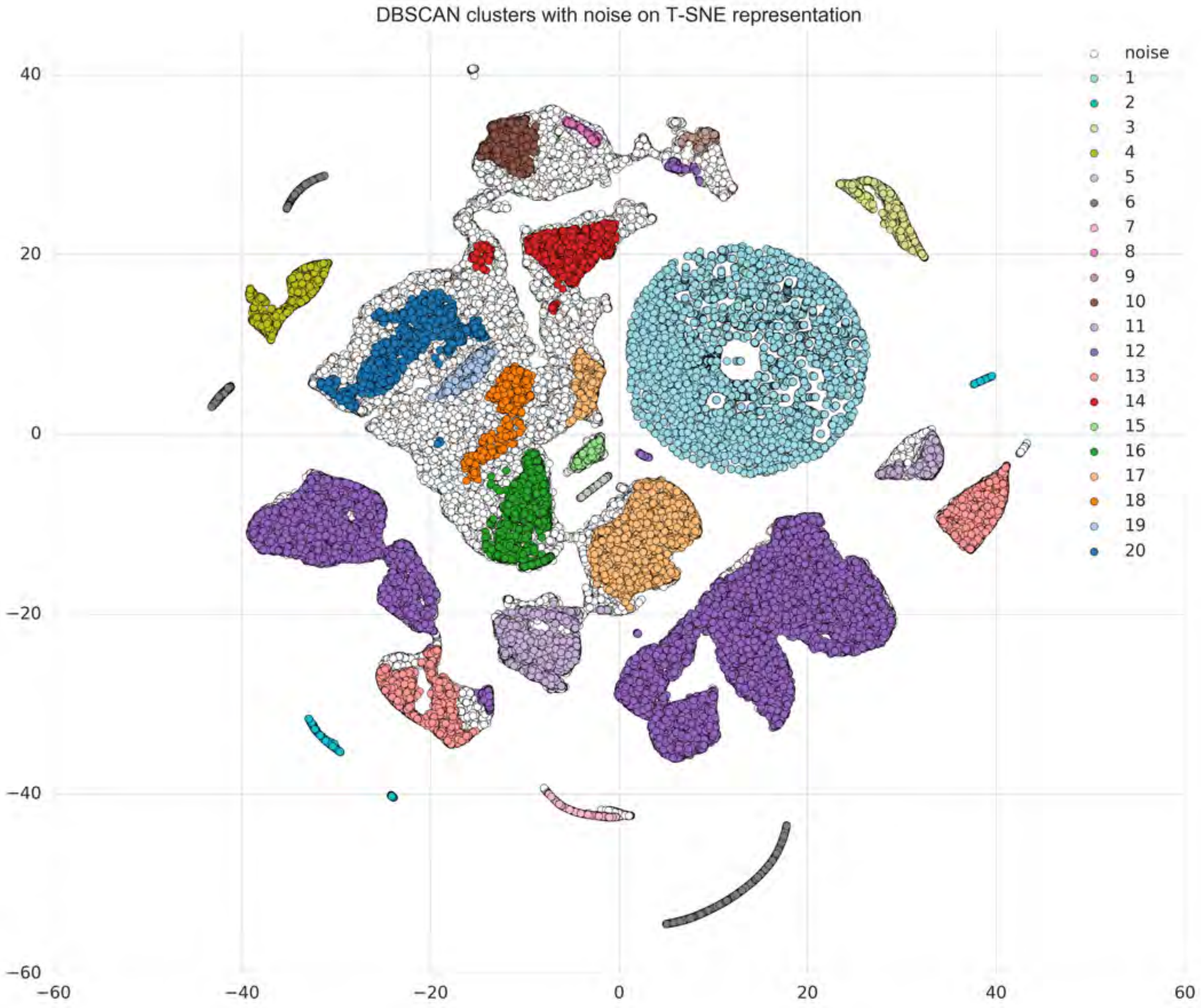




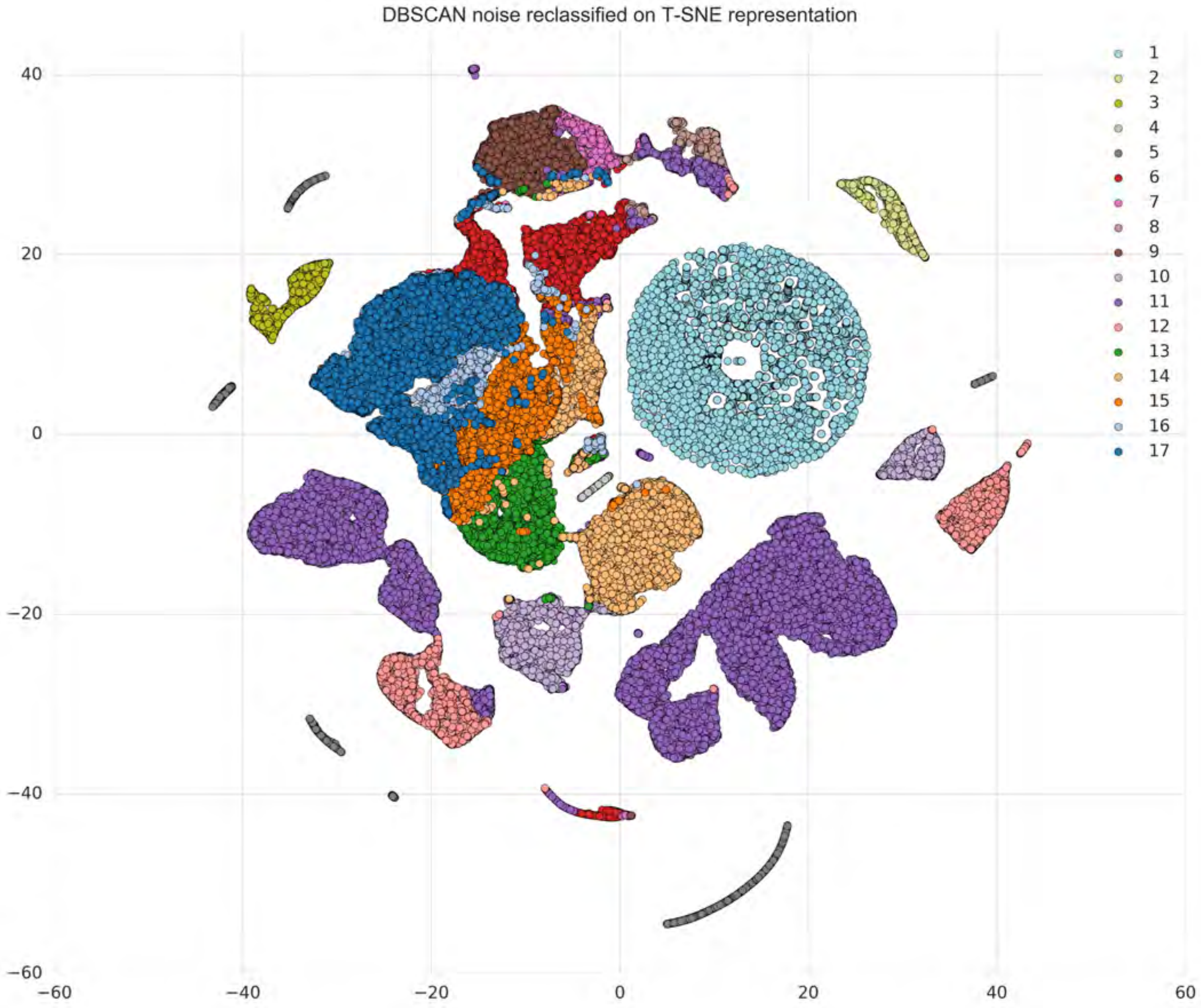
DBSCAN without noise



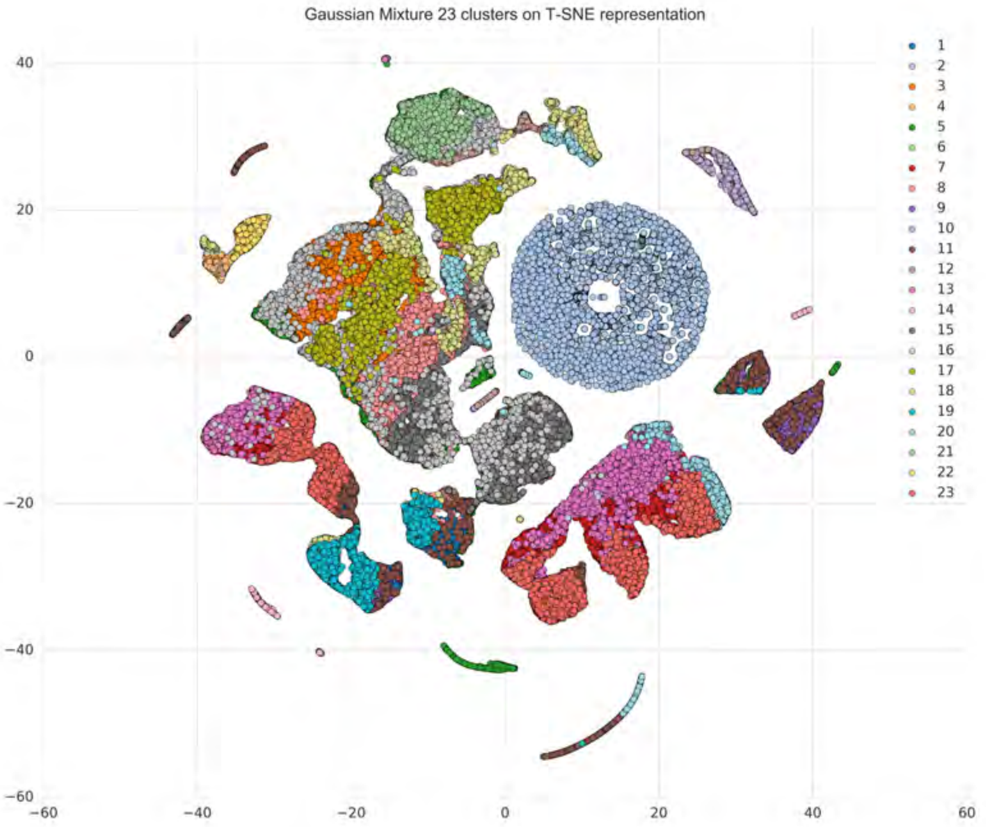
DBSCAN with noise



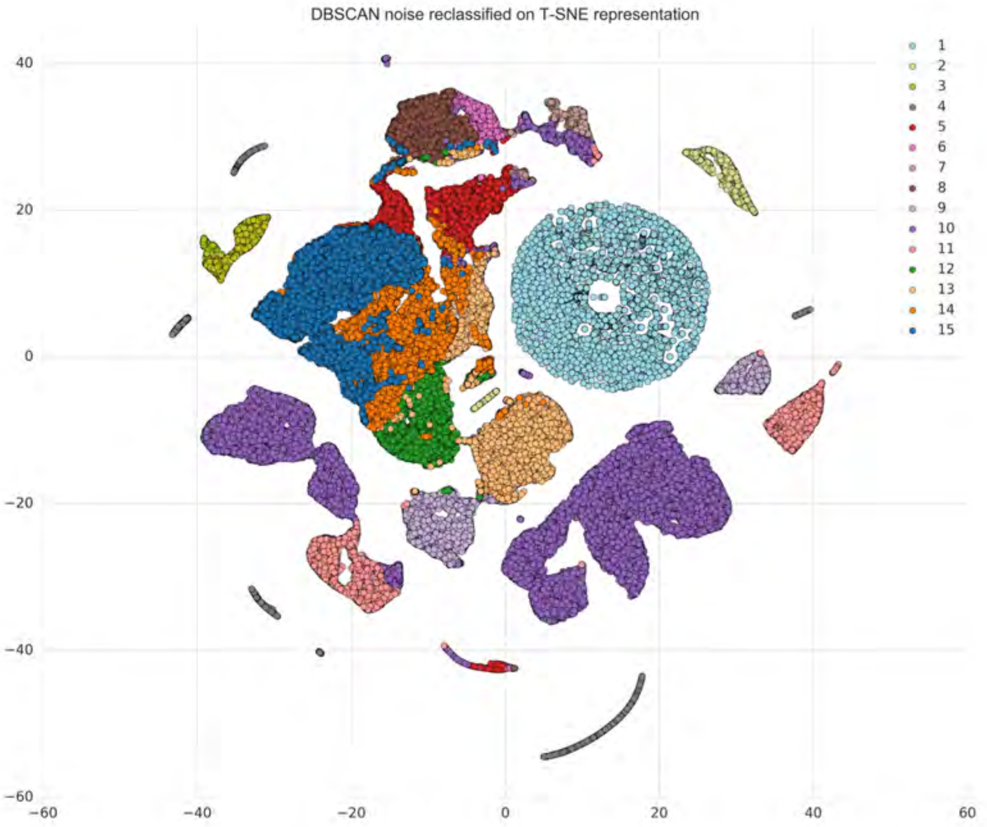
DBSCAN reclassification of noise (K-nearest neighbors)



Gaussian Mixture vs DBSCAN

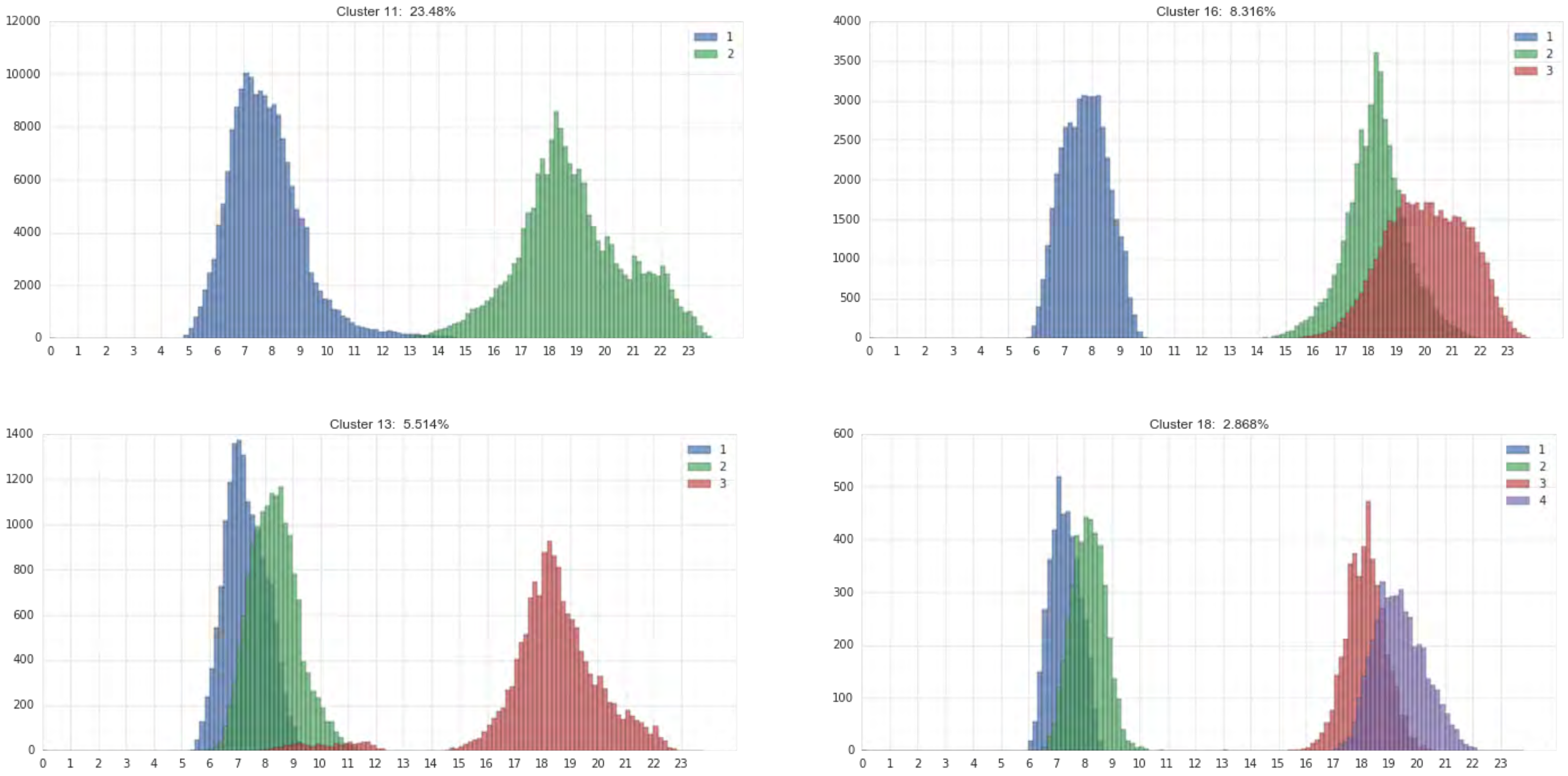


Gaussian Mixture ❌

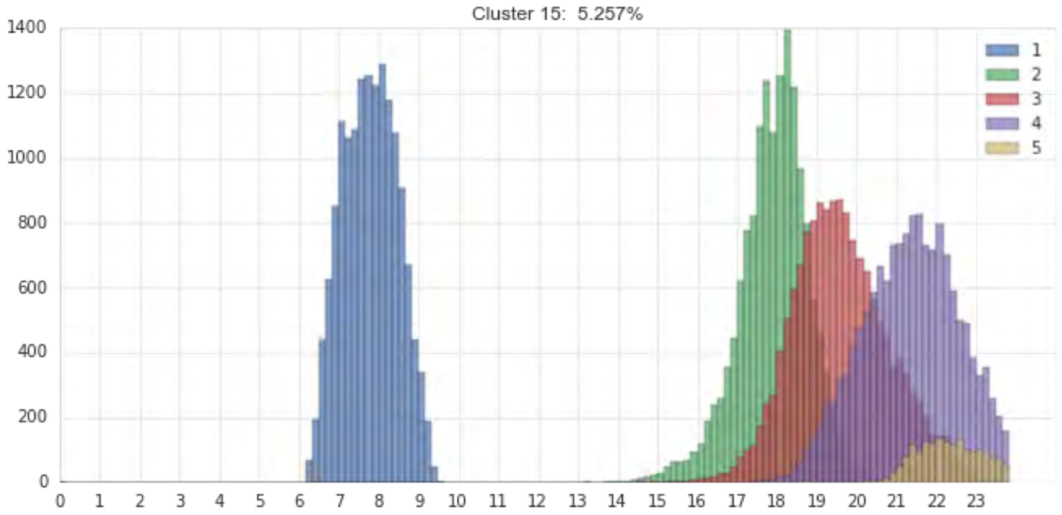


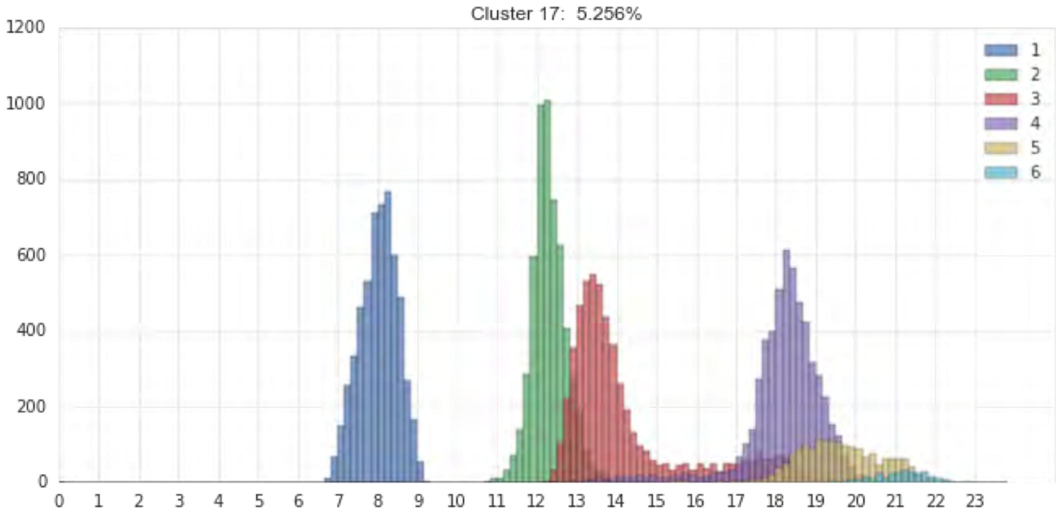
DBSCAN ✓

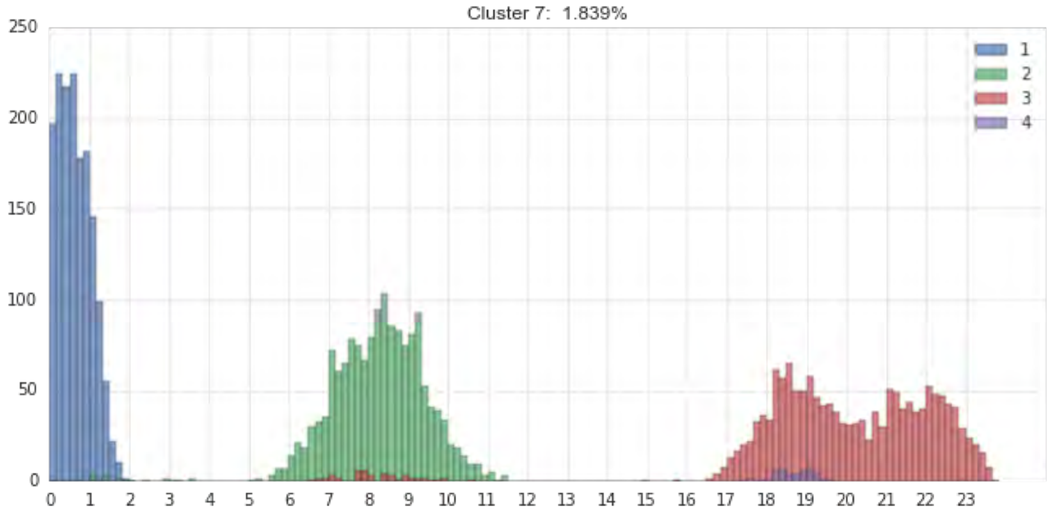
Commuter with extra stop(s)



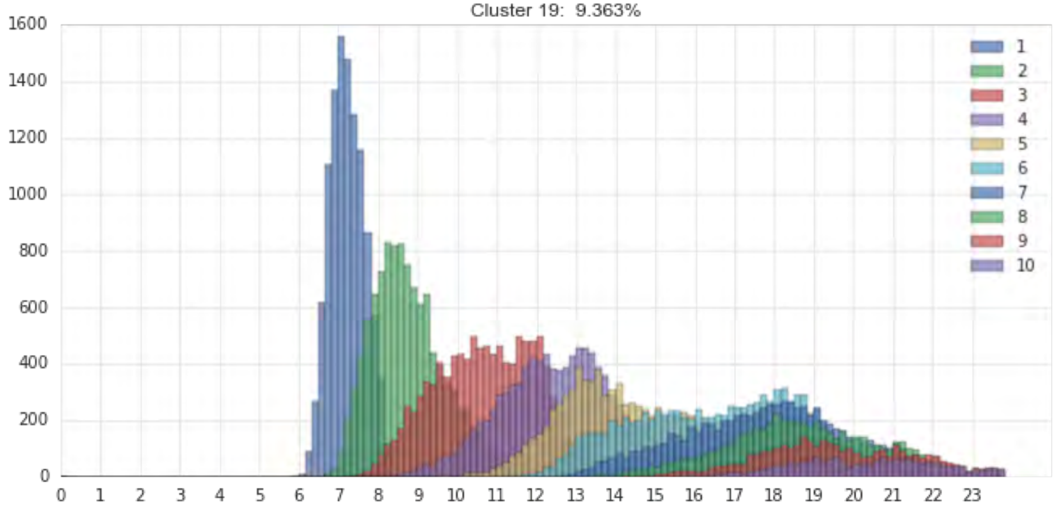
Commuter with additional trips

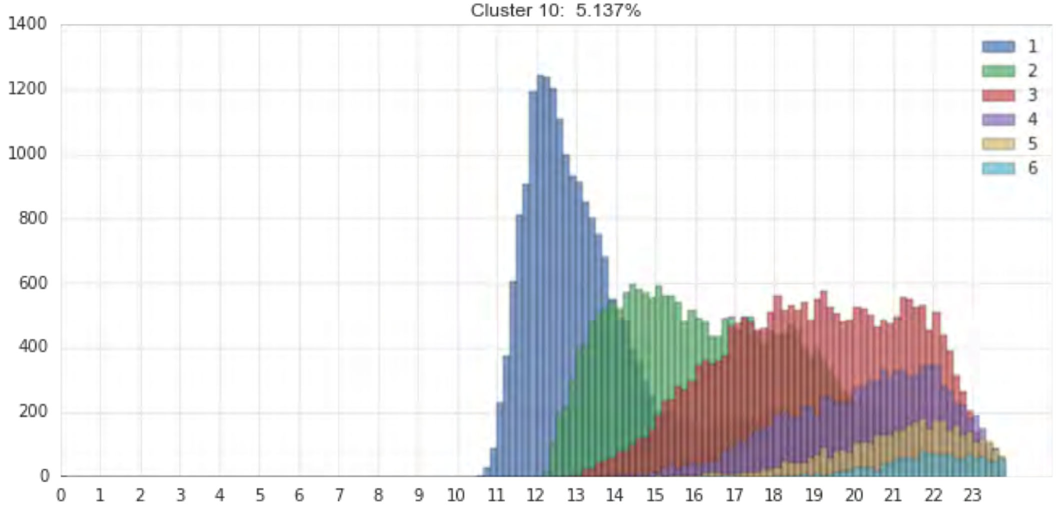


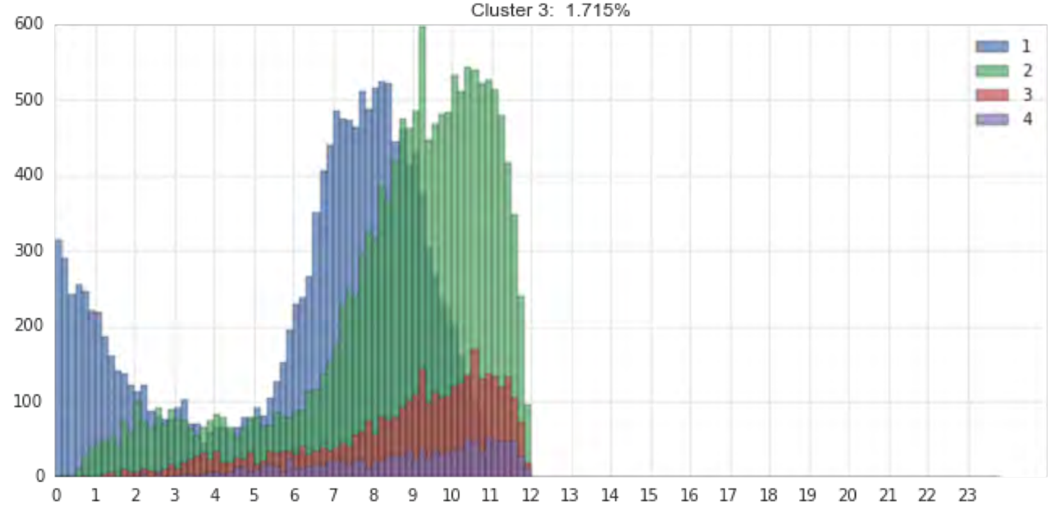


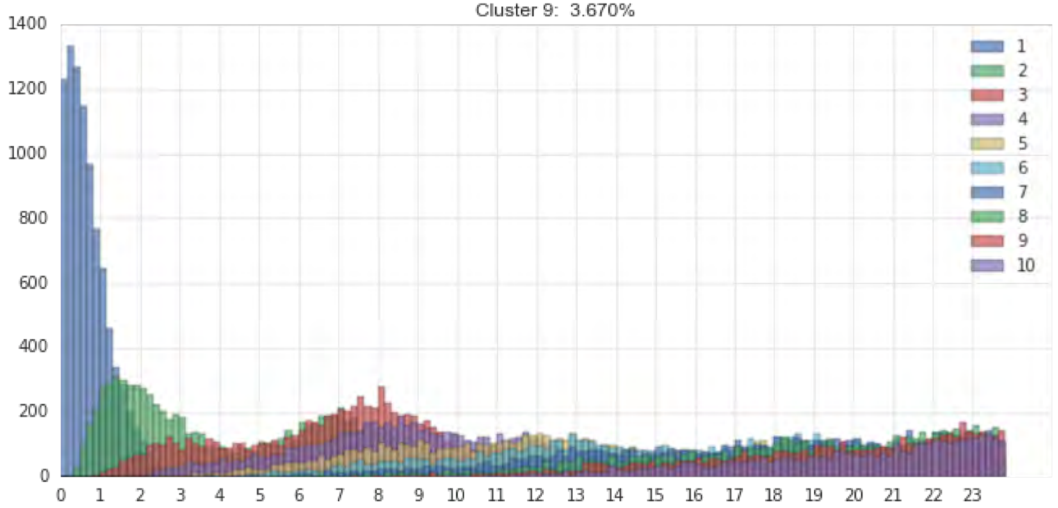


Non commuter with several trips

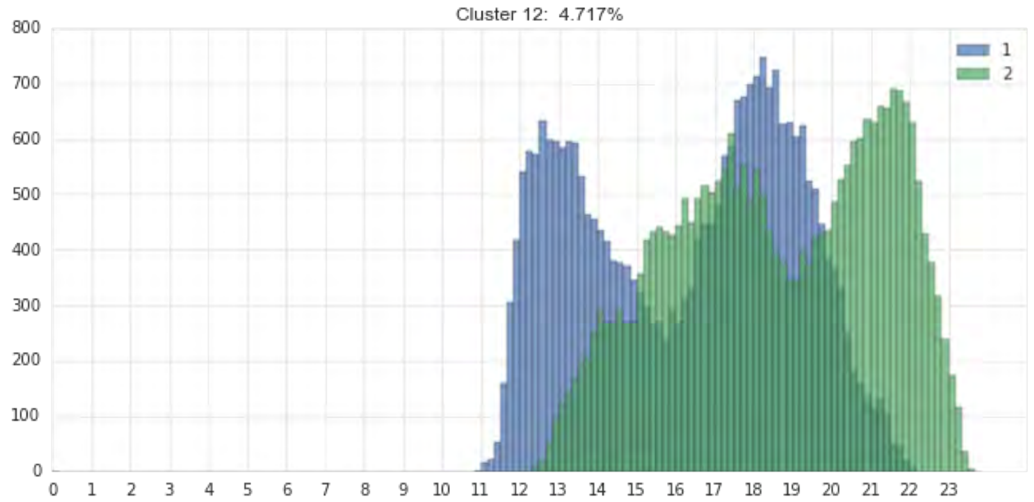


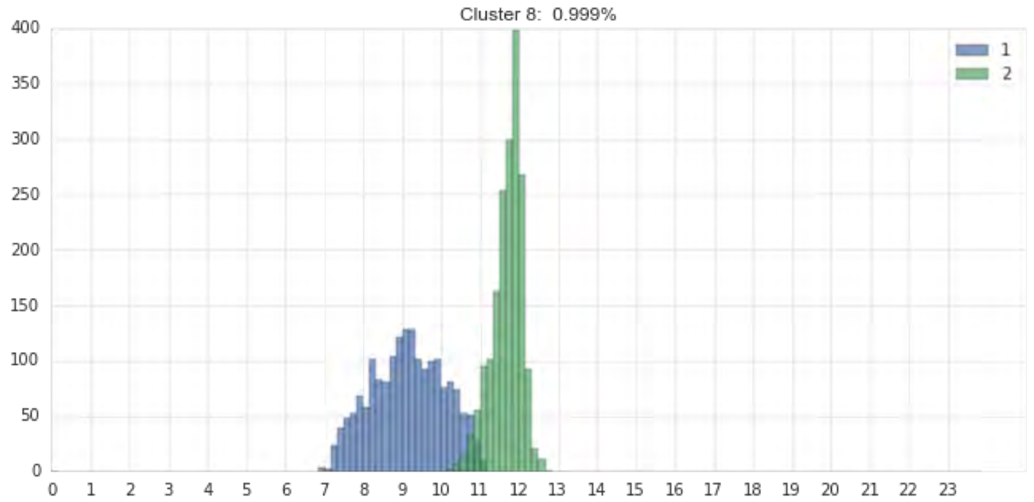


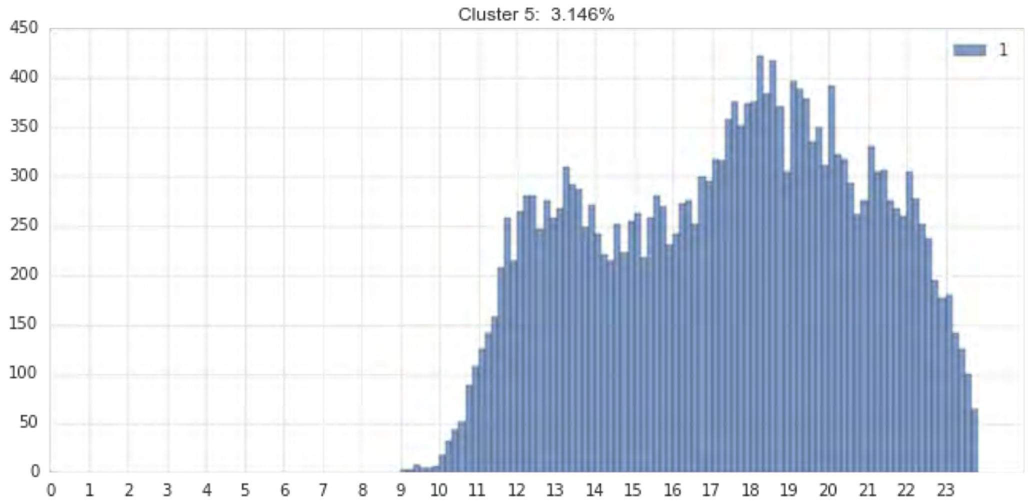


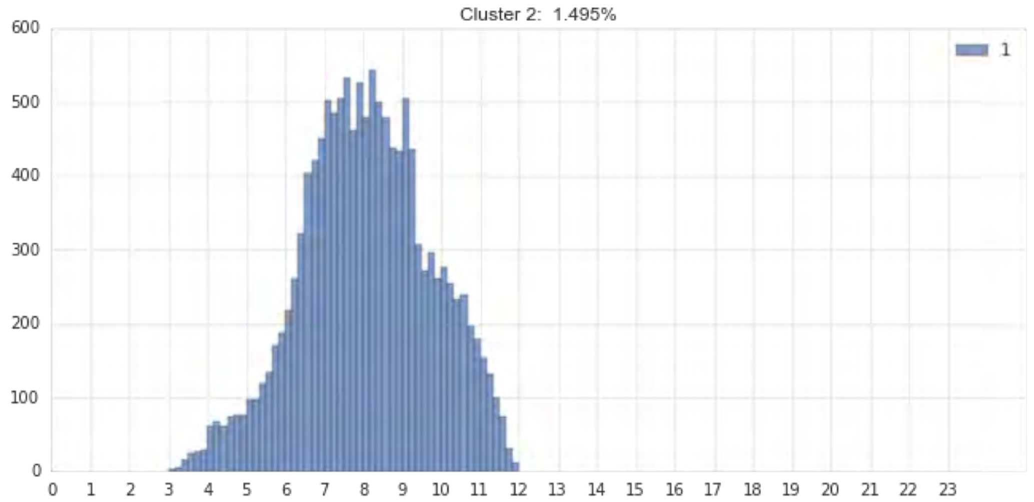


Non commuter with one activity/trip

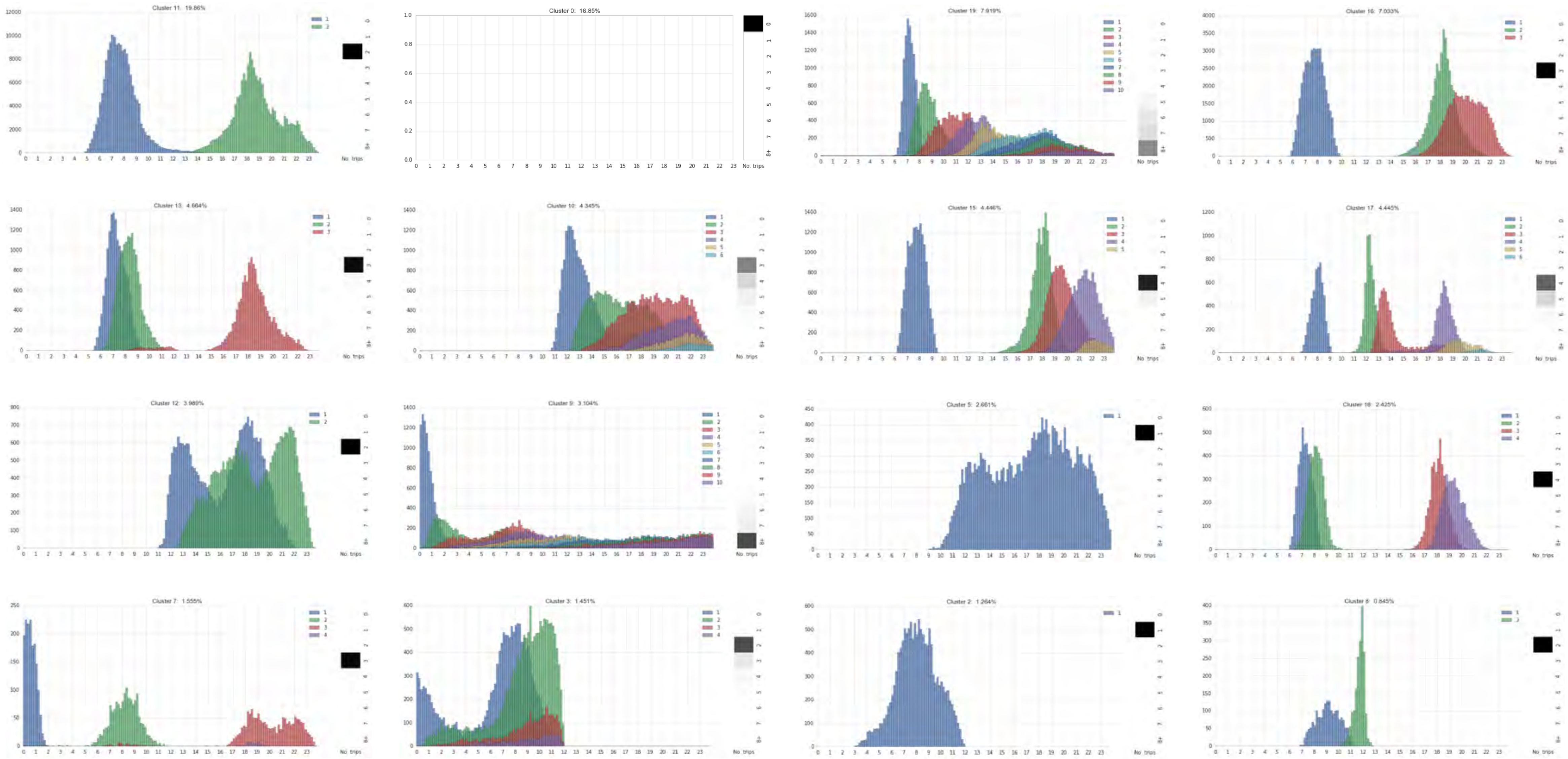






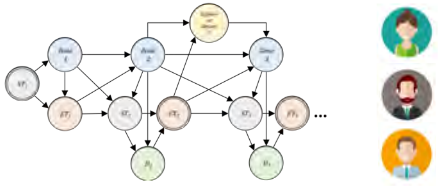


16 traveller archetypes in Singapore

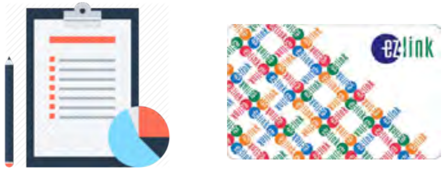


Future work

Future Work



Results of **generative model** of individual mobility patterns for each of the **clusters of travelers** found



Data Fusion. Link socio-demographics and mode of transport to individual mobility patterns. Public transport smart card + travel survey + mobile phone data

Generate population for alternative scenario Extrapolate parameters of the model to be able to generate an alternative population. Test in MATSim.

Takeaways

Takeaway

Background

Availability of **Big Data** for **urban mobility**

Big Data enables **new AI models** and applications

Big Data raises **privacy concerns**

Aim

Develop **new** Big Data-driven and privacy-by-design **models** for **urban mobility** and **transport planning**

Research question

How to use **aggregates** of Big Data to **generate** a **population** for agent based simulations?

Methodology

1. Find clusters of travel behaviour in Big Data

2. For each cluster, train a generative model (dynamic Bayesian network) which only requires aggregated data.

3. Sample from the model to generate population.

4. Use population in agent-based simulations to answer planning questions.

ENGAGING BIG DATA – ENGAGING MOBILITY - FCL

Cuauhtemoc Anda
PhD Researcher
Data Scientist

Dr. Sergio Ordonez
Senior Researcher
Computer Science

Dr. Pieter Fourie
Project Leader
Senior Researcher