

Who goes there?

Profiling audience street by street with Telco data

Kael Huerta

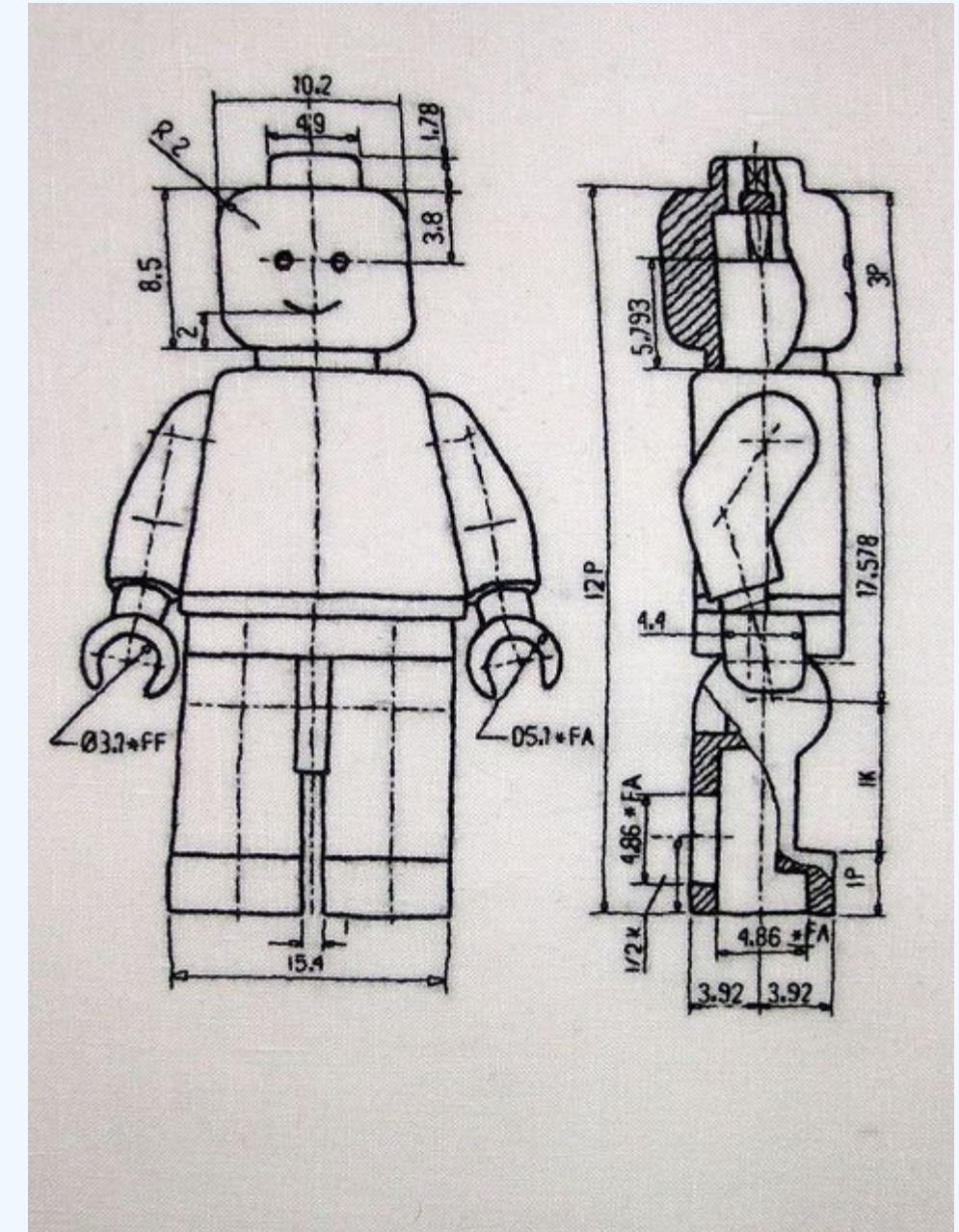
Data Scientist

kael.huerta@toptal.com

 @kaelhuerta

Today's plan

- Introduction
- Dataset
- Models
- Final data product



The problem

- Telcos do not know their customers



The dataset

- SQL database
- Hadoop cluster
- Flat files

The dataset

- **SQL database**
- Hadoop cluster
- Flat files



The dataset

- SQL database
- **Hadoop cluster**
- Flat files



The dataset

- SQL database
- Hadoop cluster
- **Flat files**



The models

Interest
Profiling



Scoring
Segmentation



Gender
Prediction



Age
Prediction



Interest Profiling



Interest Profiling



Open Catalog of Web Domains



Interest Profiling



Interest Profiling



Interest Profiling

Life tip:

Research for what is done already



Scoring Segmentation



Scoring Segmentation



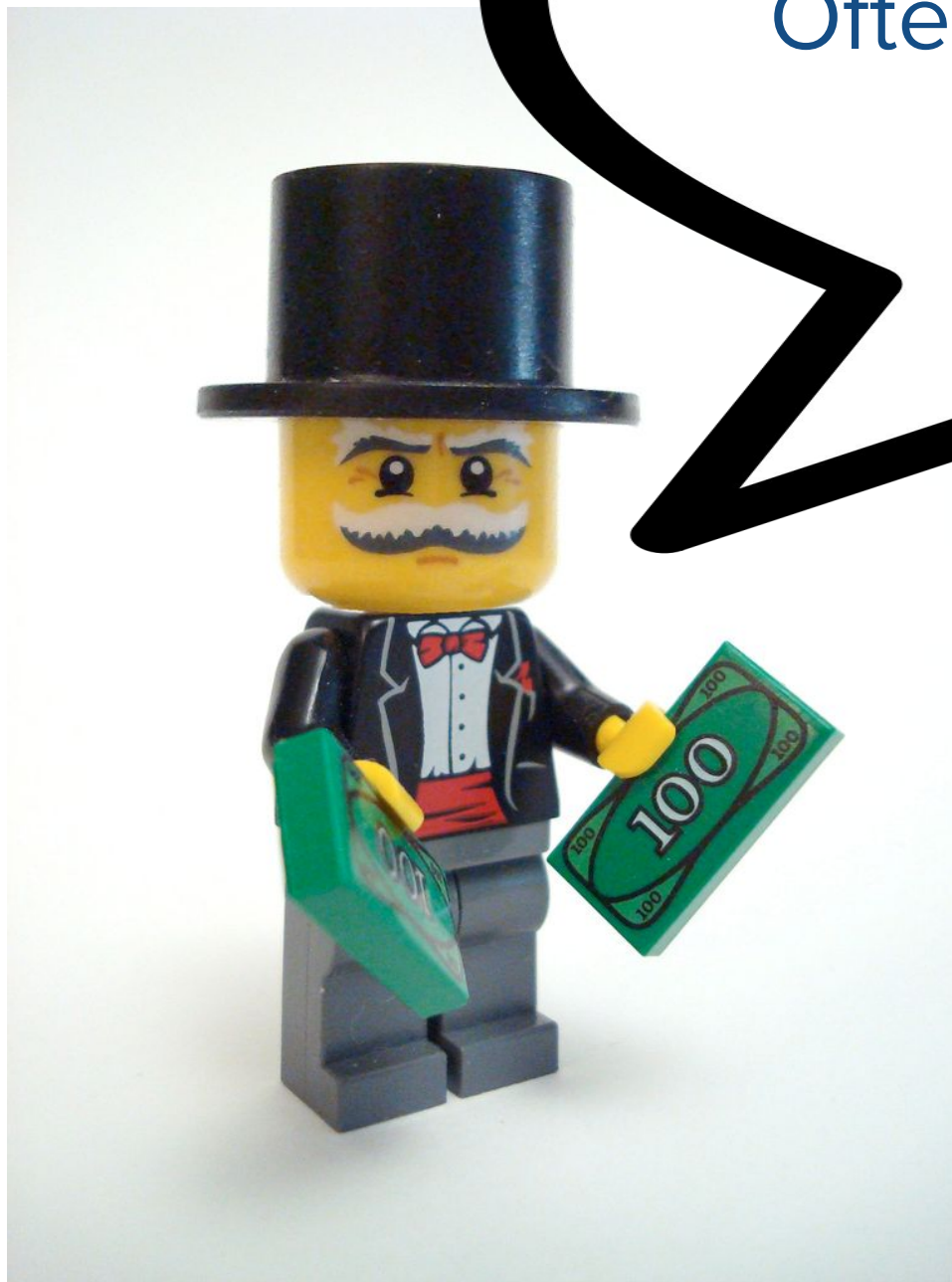
Fuzzy C-Means from [e1071](#)



Scoring Seq

Life tip:

Often, a *good-enough* model is enough



Hire the top 3% of freelance talent

www.toptal.com

@kaelhuerta

Gender Prediction



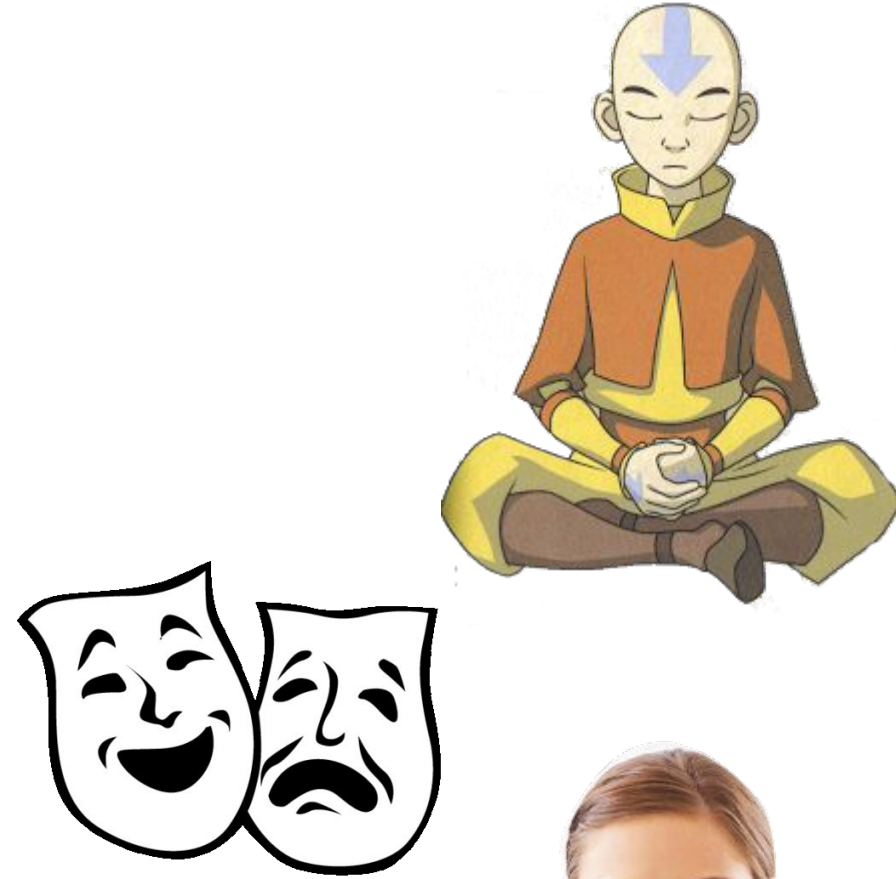
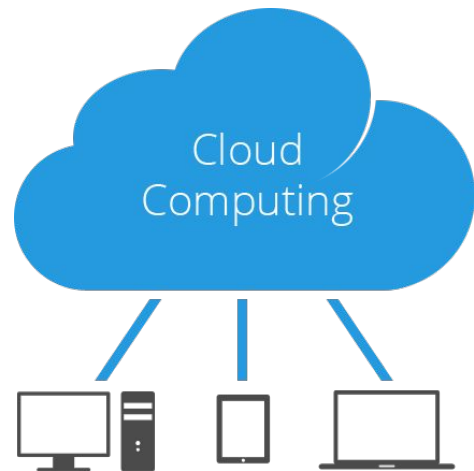
Gender Prediction



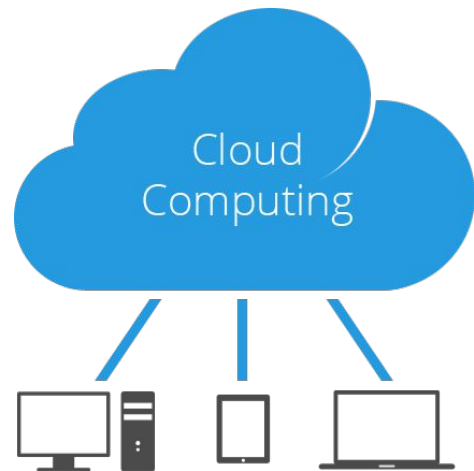
Logistic Lasso Regression from
`glmnet`



Gender Prediction

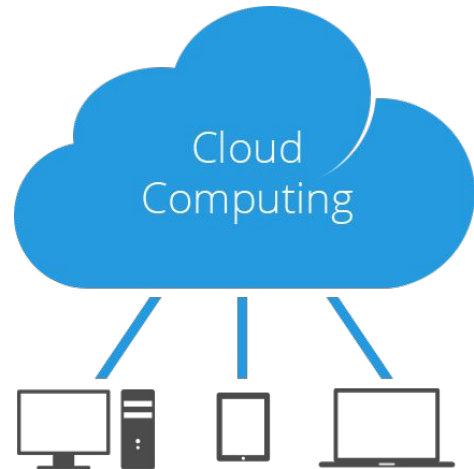


Gender Prediction



Tech tip:
`data.table`

Gender Prediction



Tech tip:

```
data.table  
set.index(dt, cols, ...)
```



Hire the top 3% of freelance talent

www.toptal.com

@kaelhuerta

Age Prediction



Age Prediction

Factor Analysis from **FactoMineR**



18⁺

Age Prediction

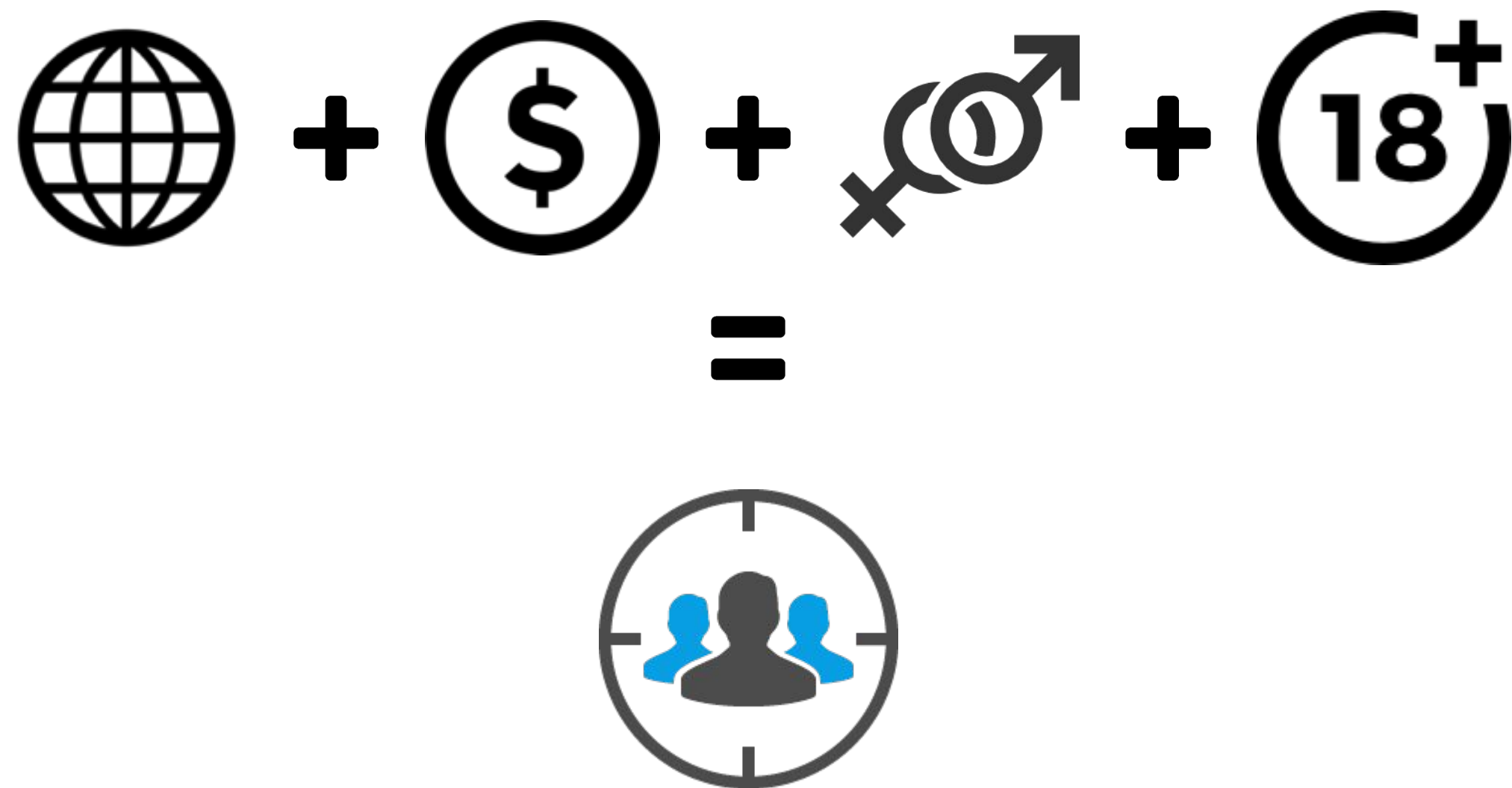


Life tip:
Share relatable examples

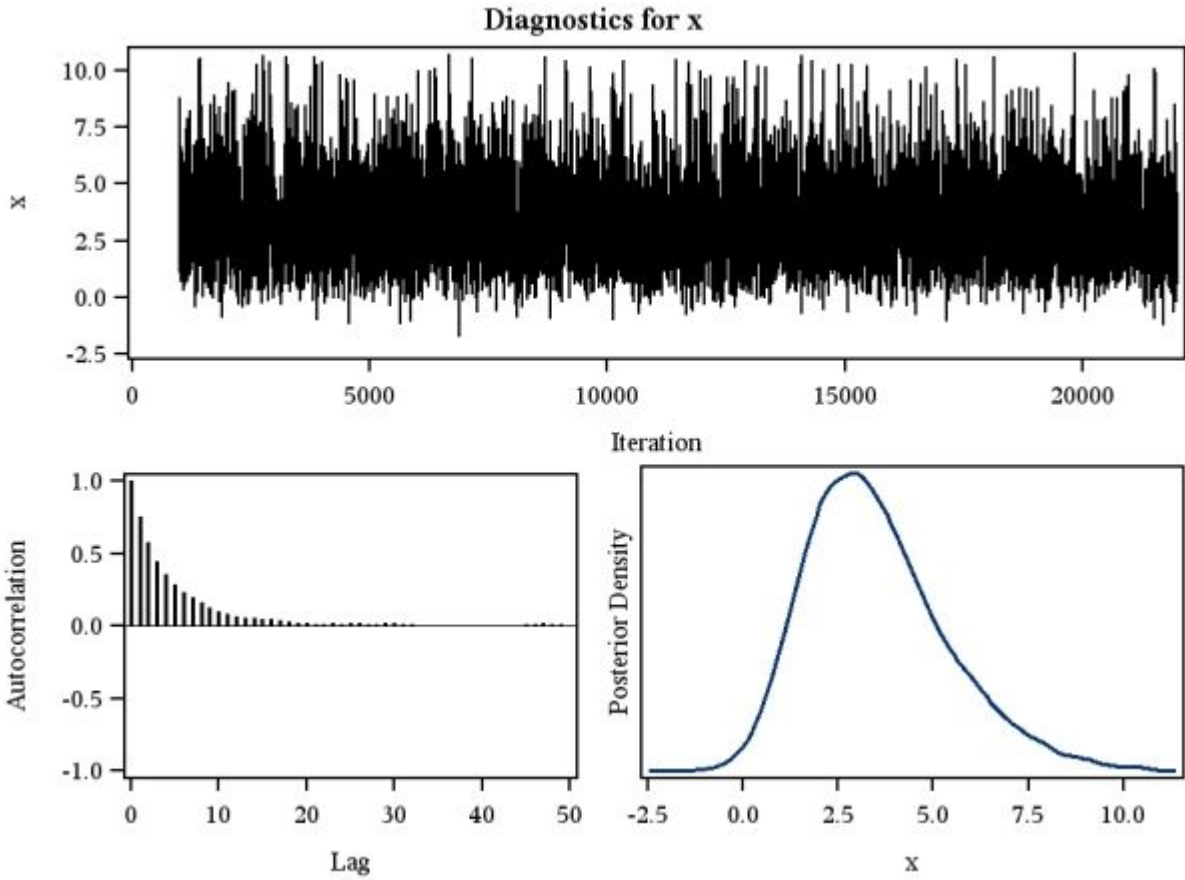
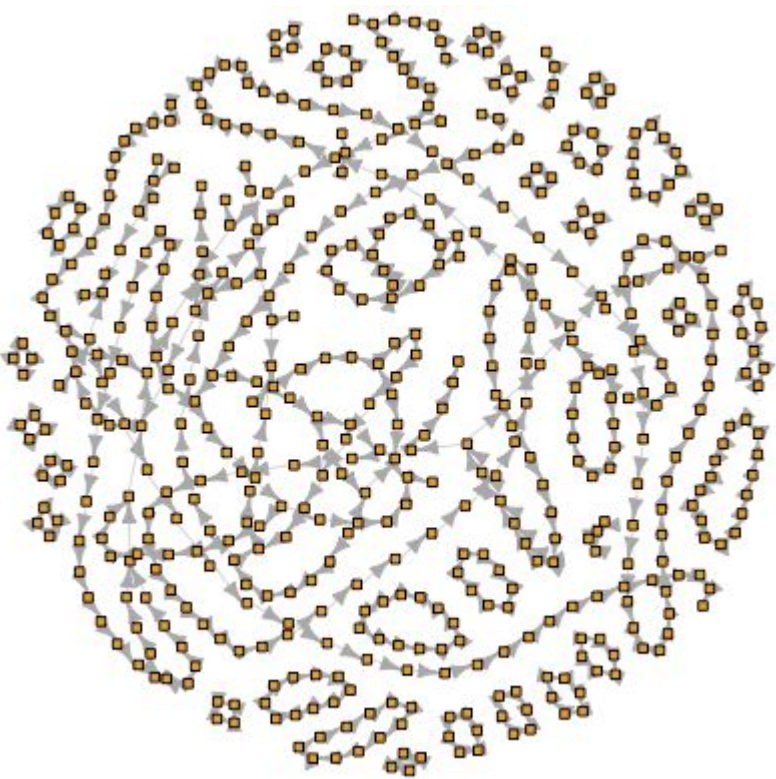
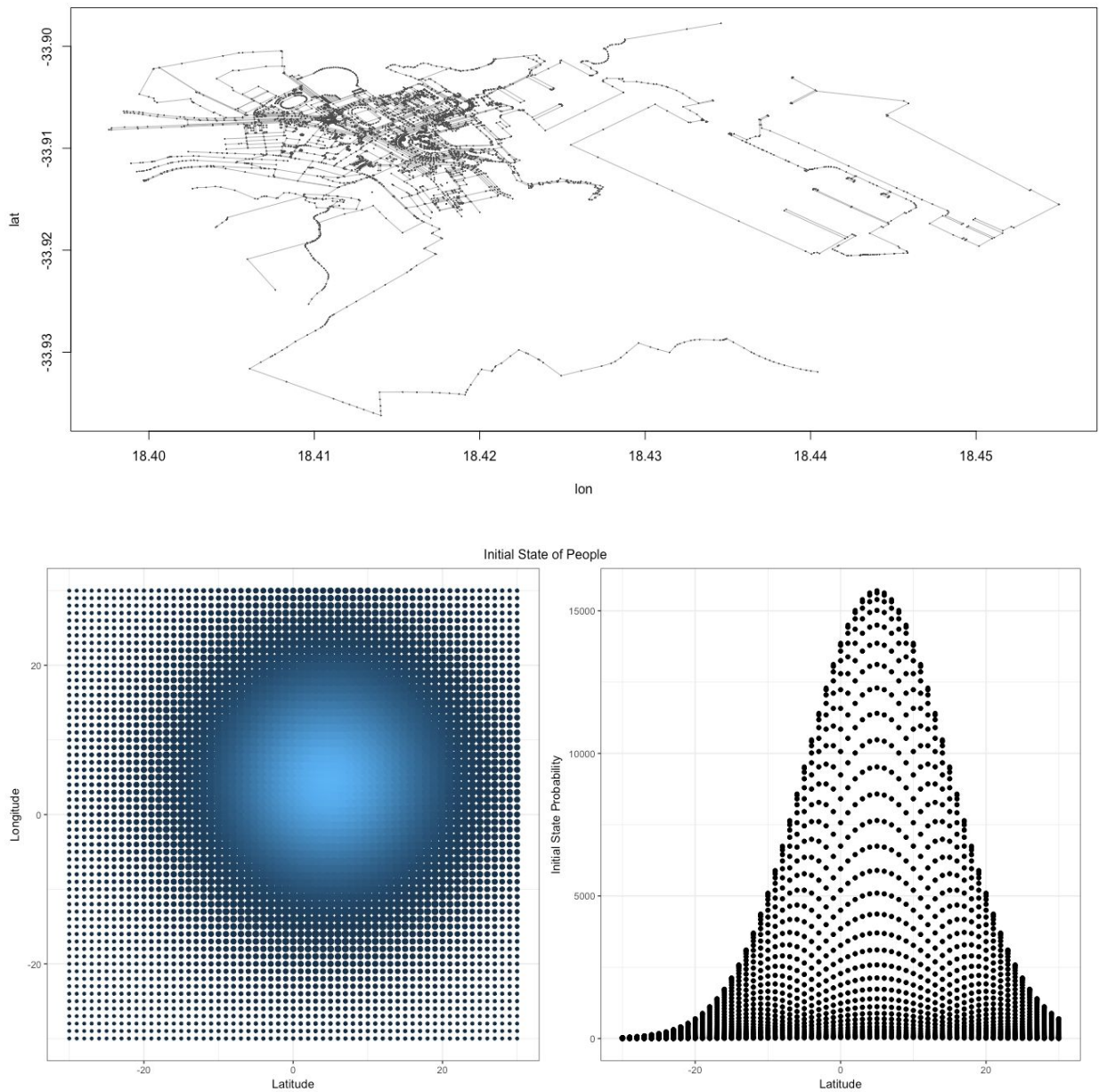
MineR

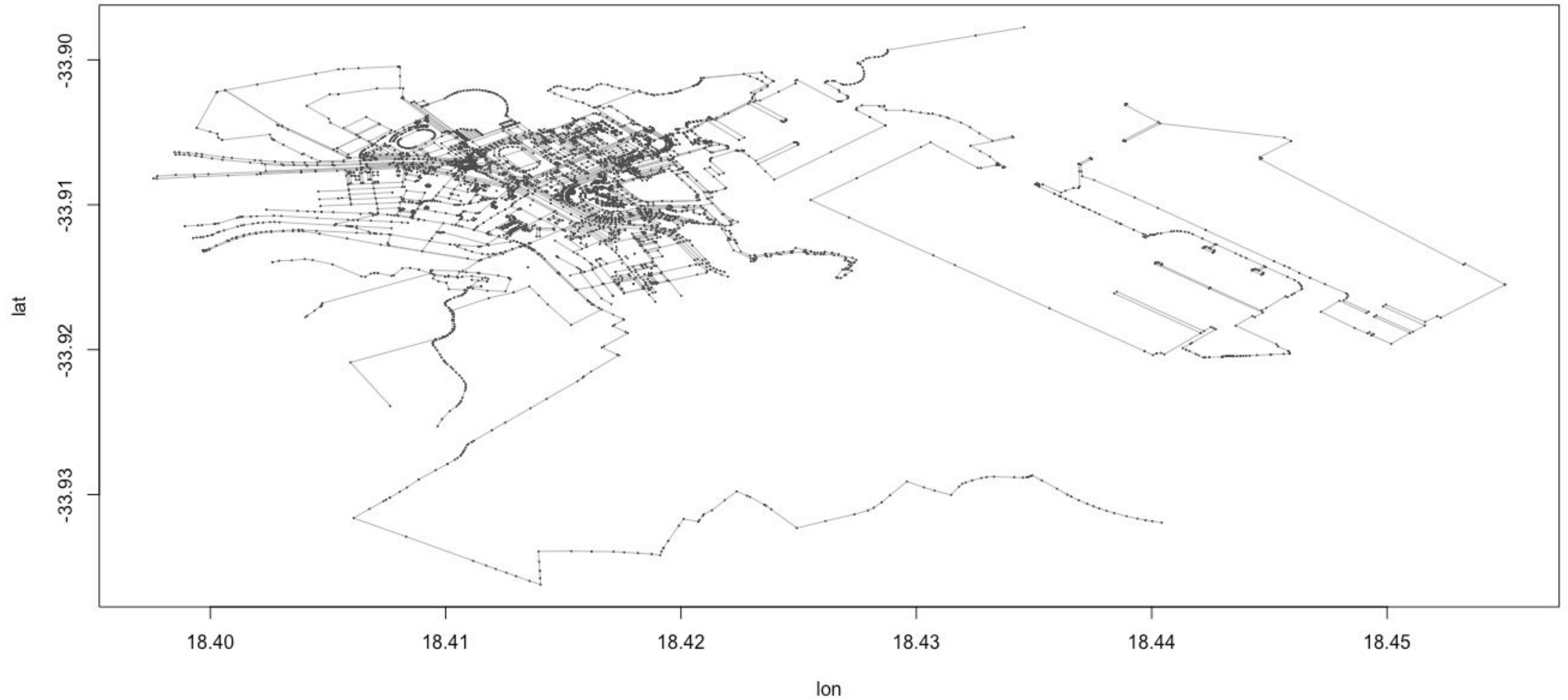
18⁺

Final Data Product

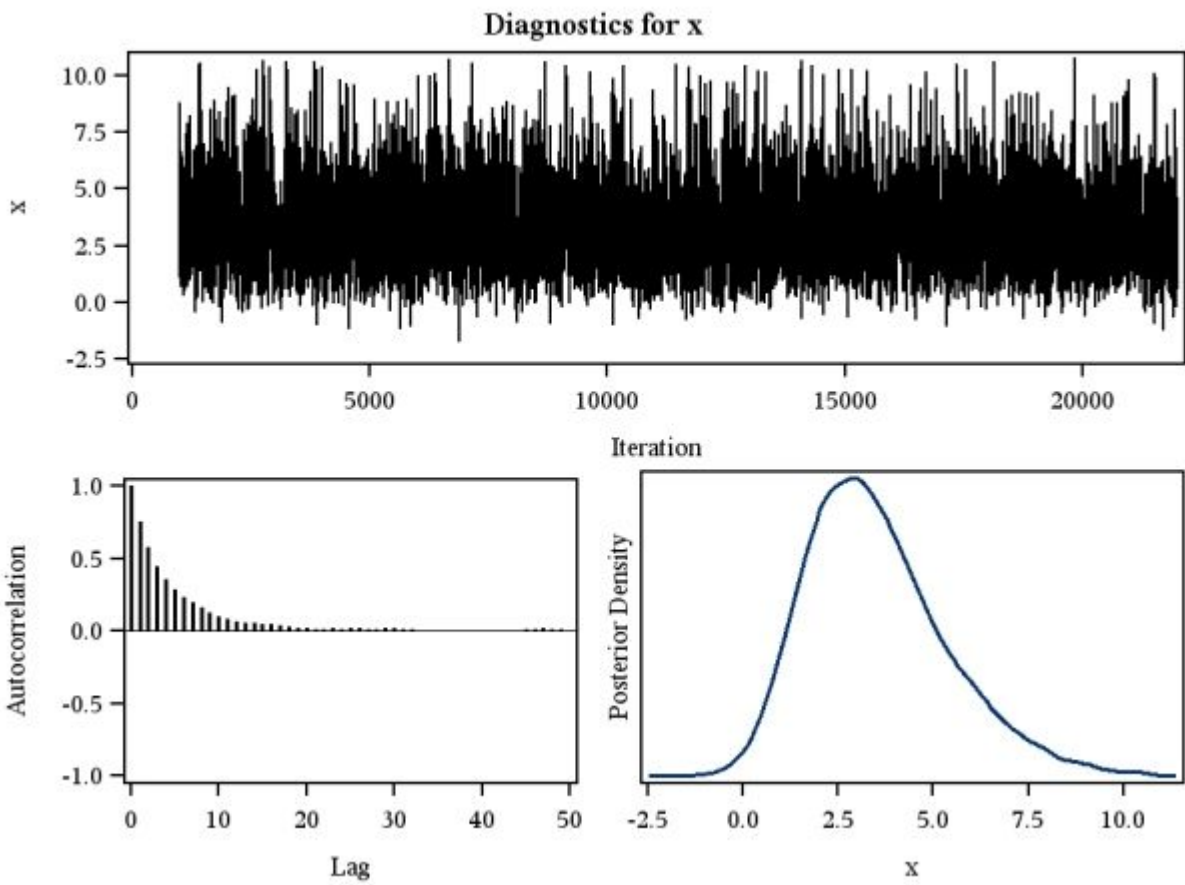
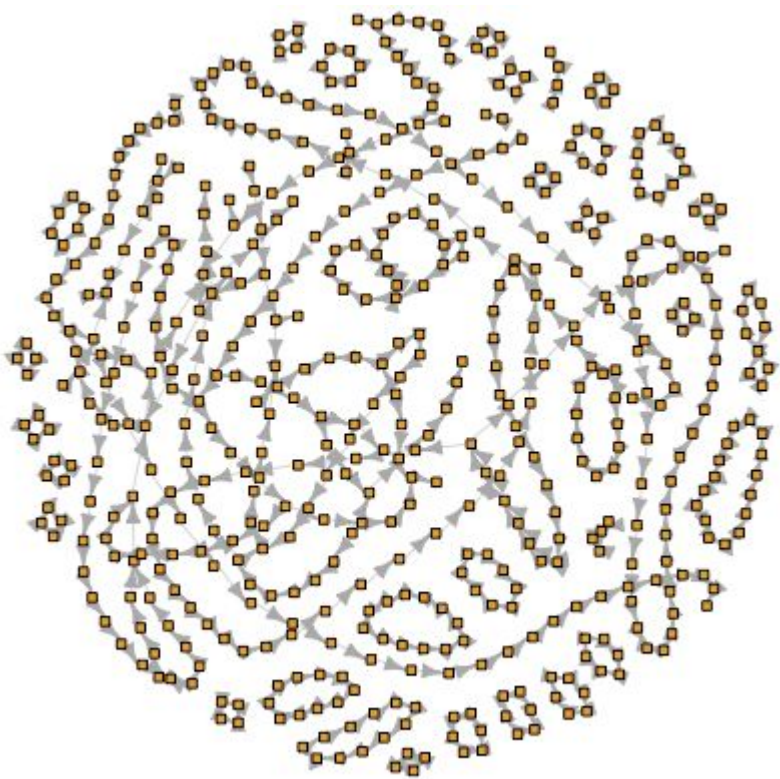
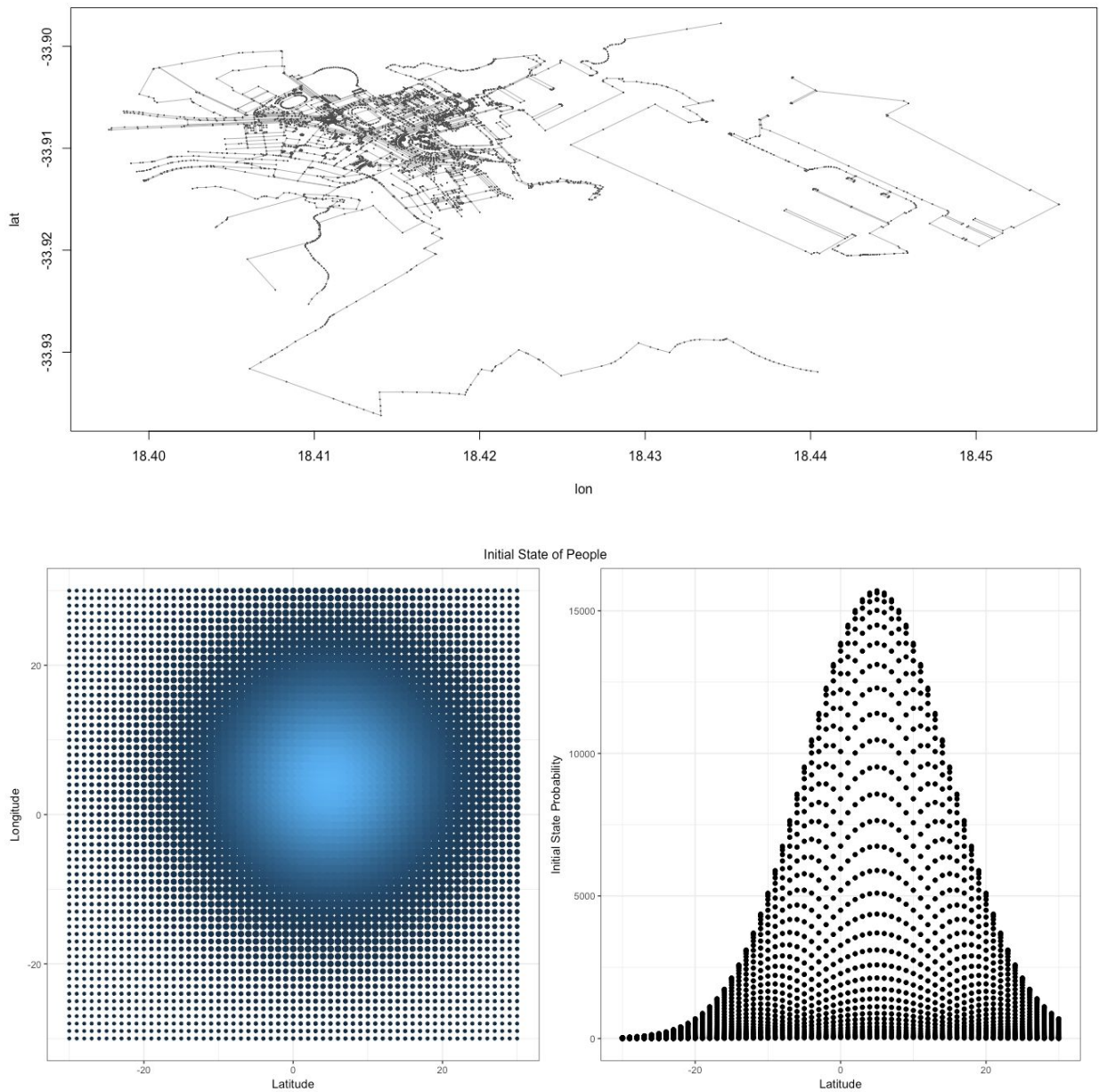


Final Data Product

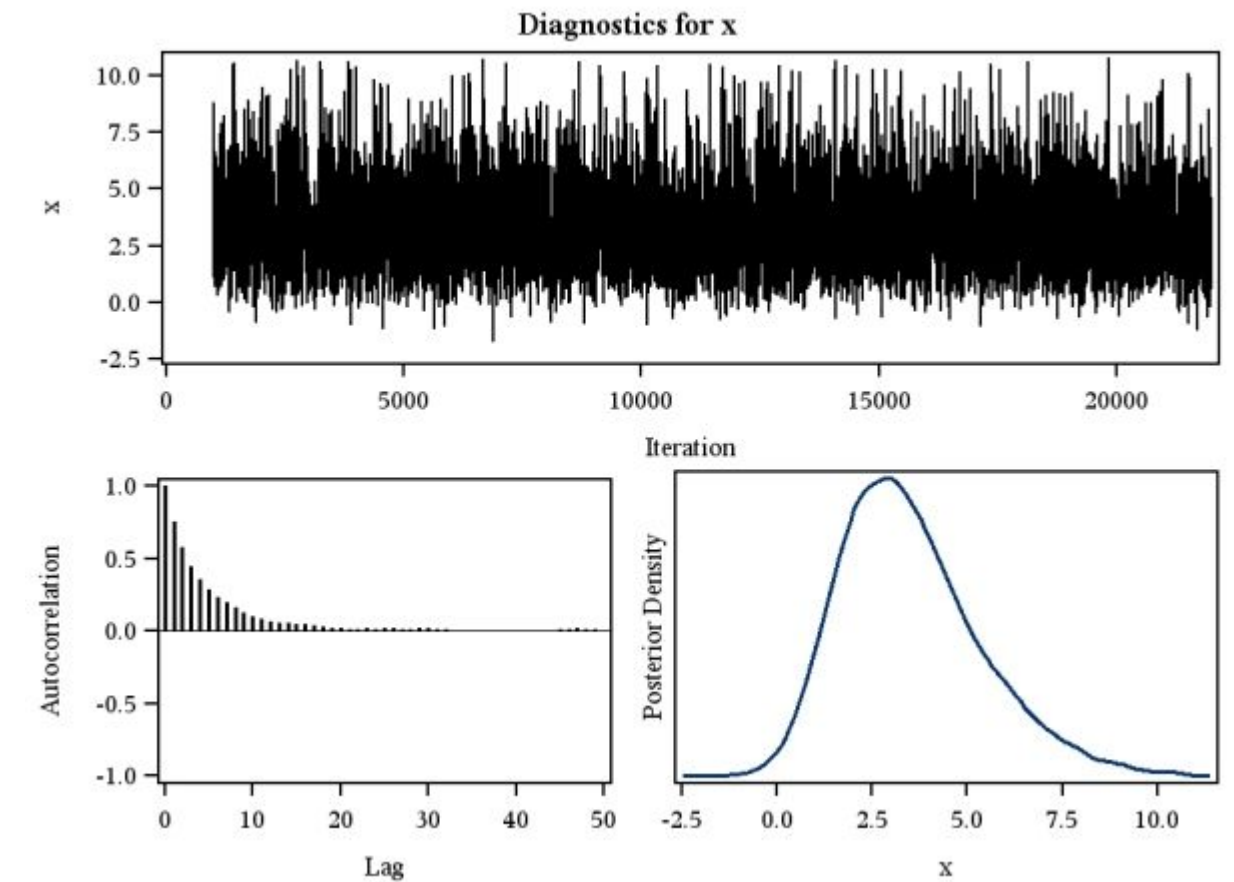
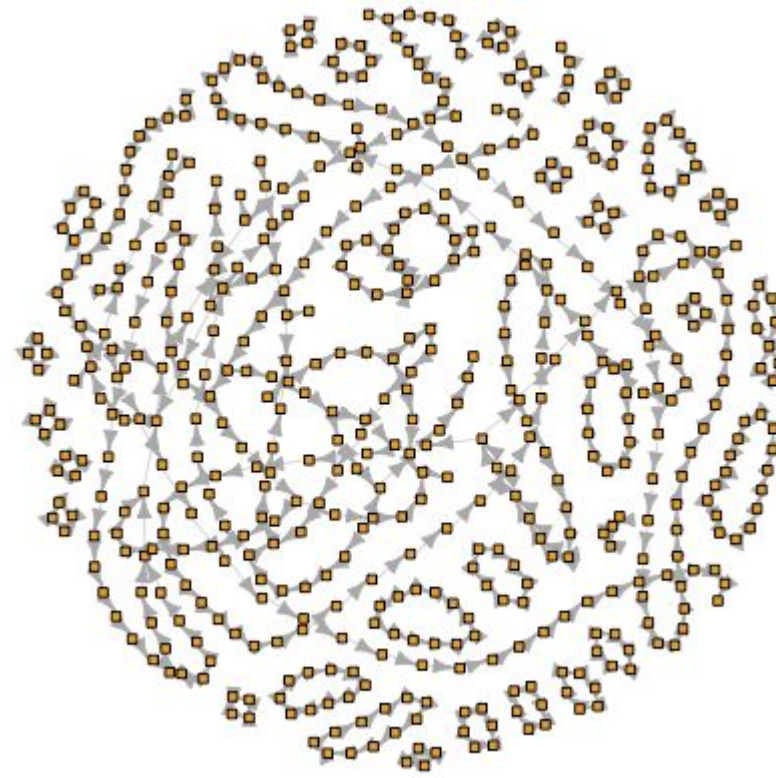
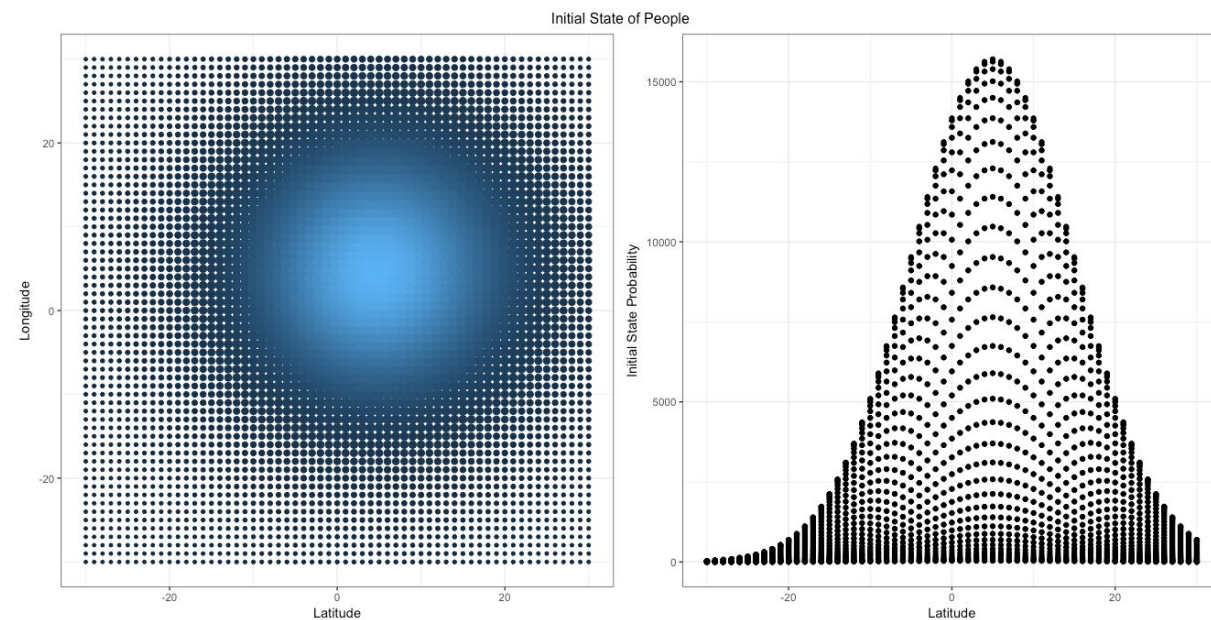
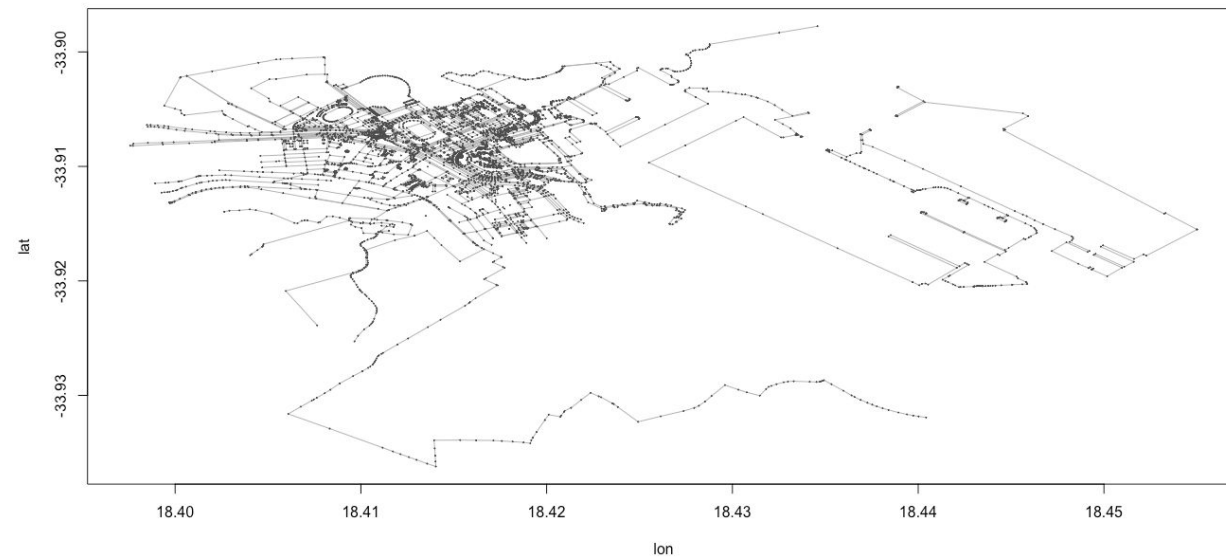




Final Data Product

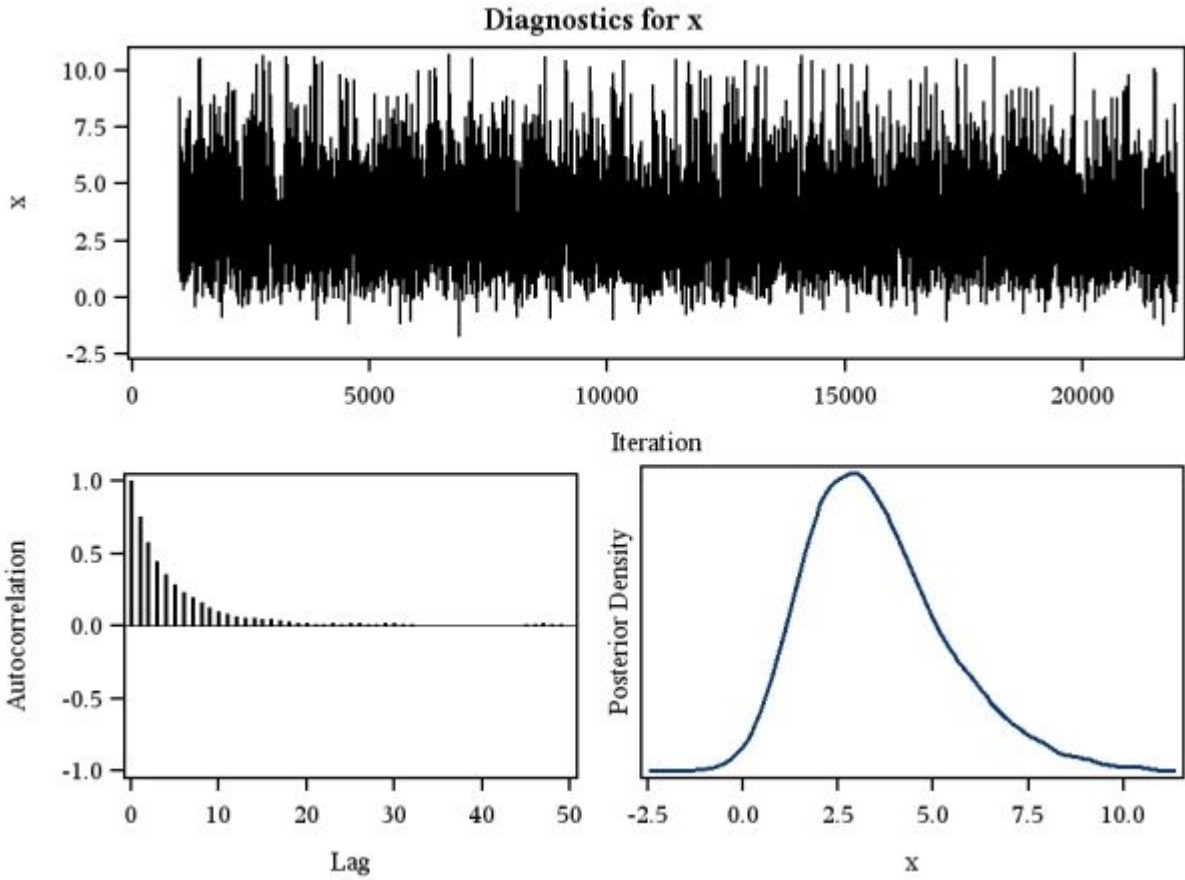
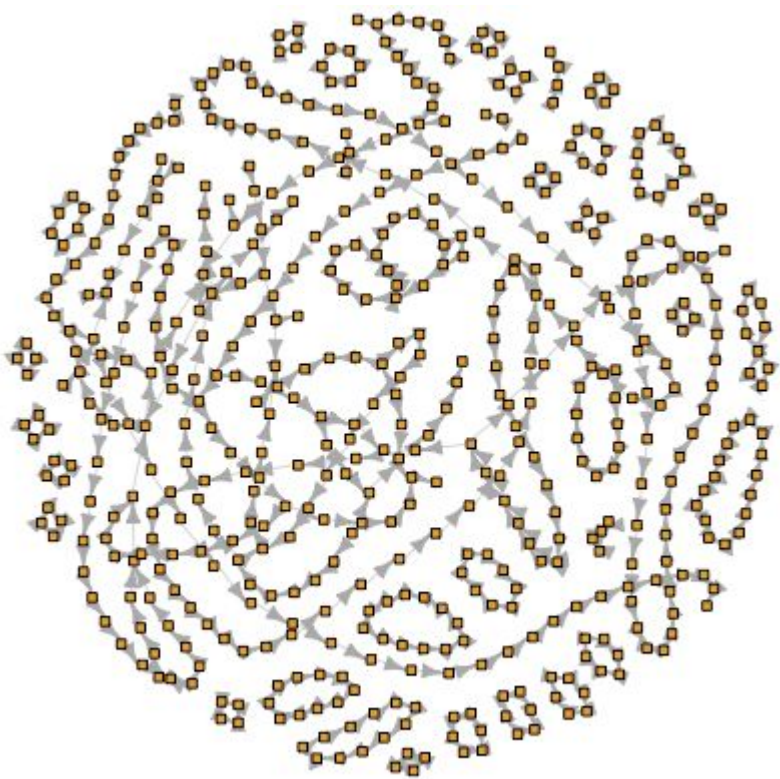
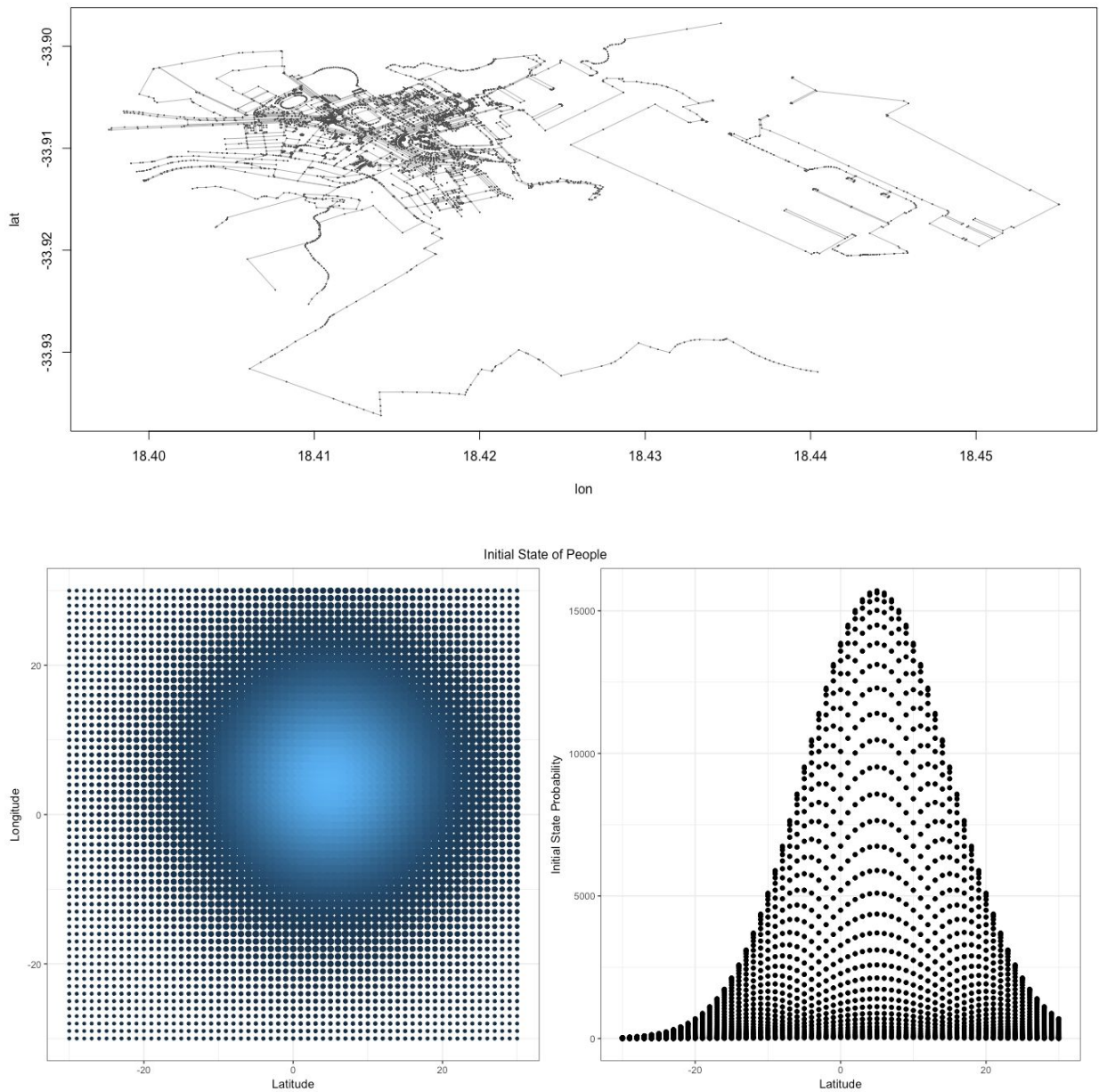


Final Data Product



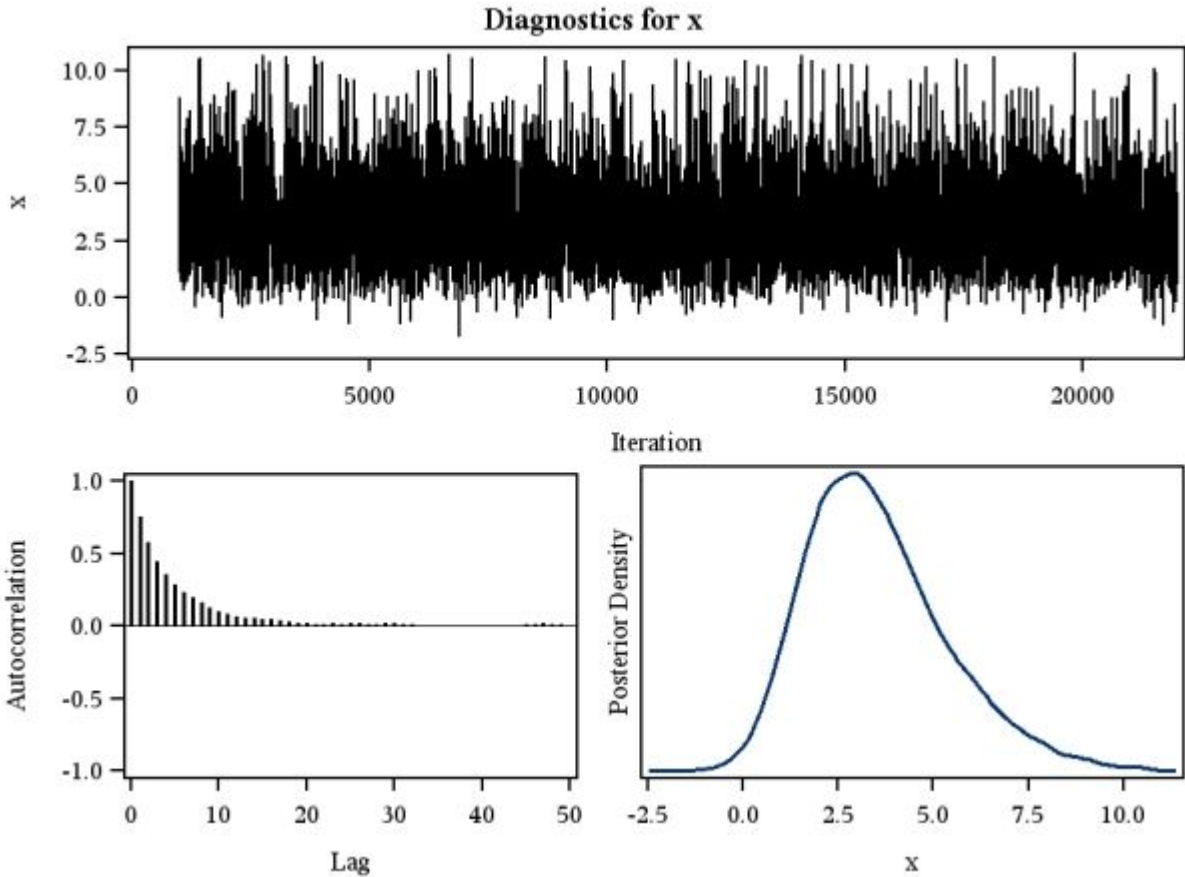
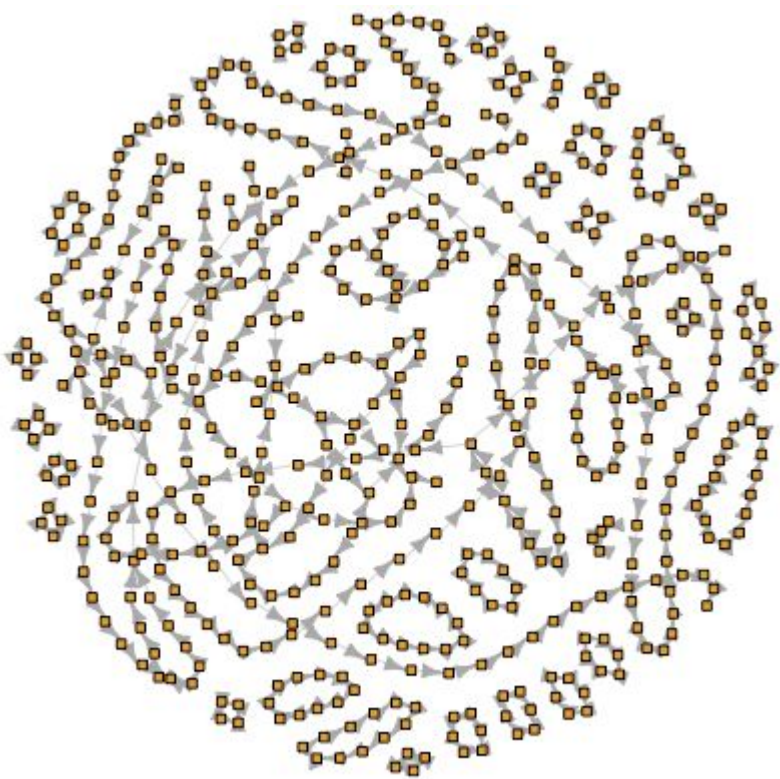
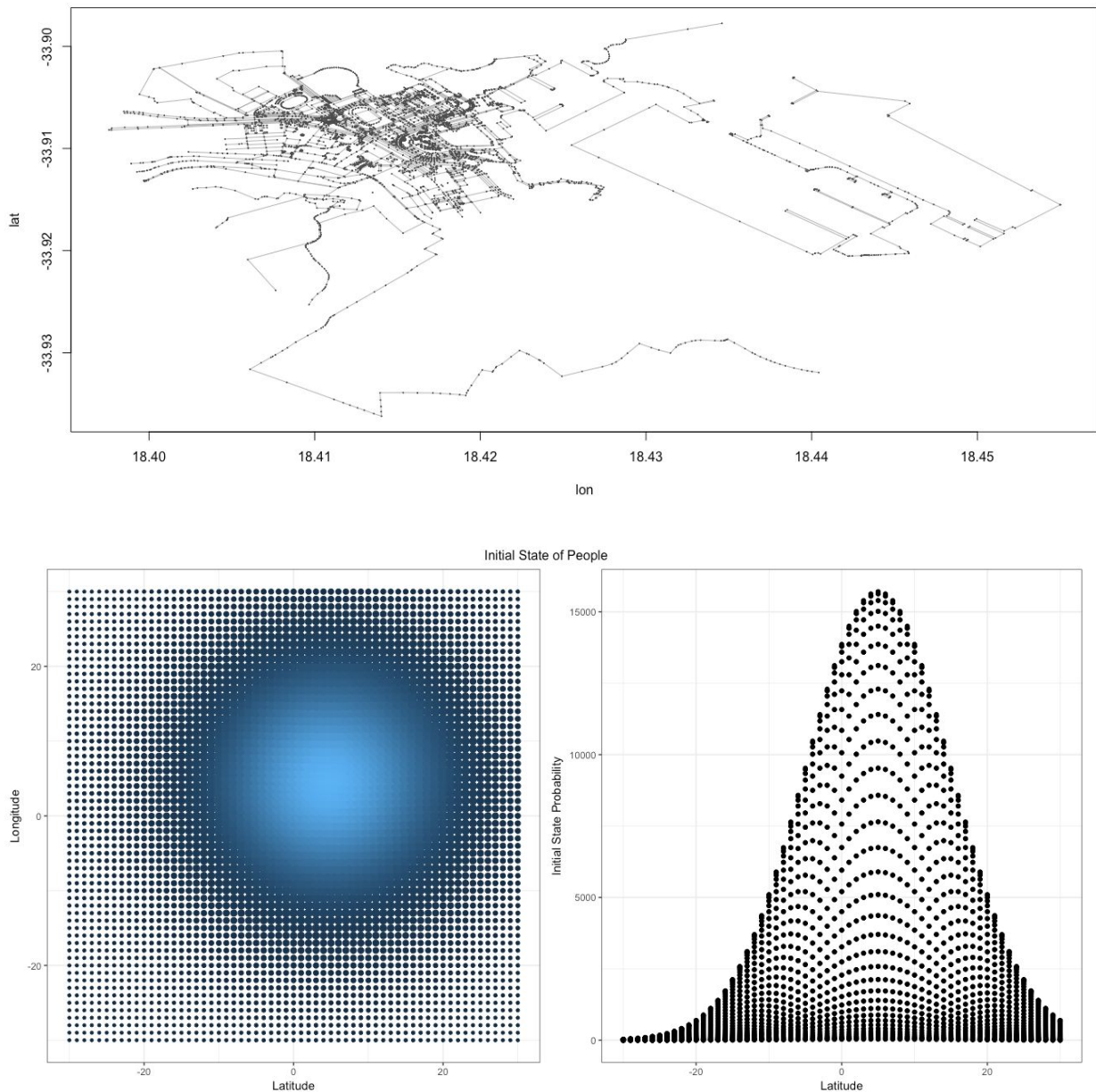
Package **osmar**

Final Data Product



Package `sp`

Final Data Product



Package `leaflet`

Applications



Applications



Applications



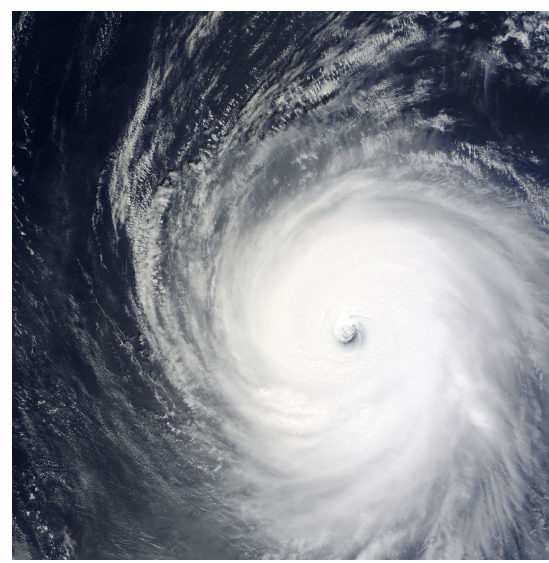
Applications



Applications



Applications



Reach out!

Kael Huerta

Data Scientist

kael.huerta@toptal.com

 @kaelhuerta