Multivariate Analysis

The Application and Prediction of Public Bicycle System

Group 3

陳晏鵬

林家毅

李易修

鄧兆延

1 Motivation

Problems with Urban Traffic

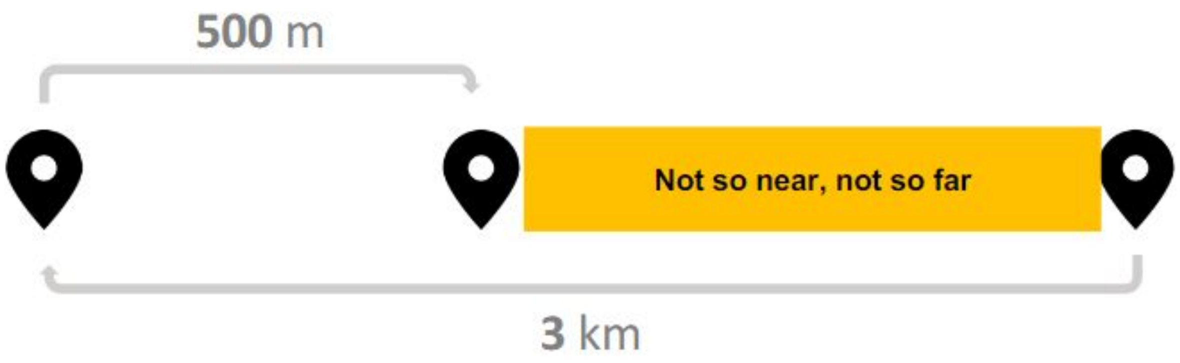
- 1: Traffic congestion, high volume of carbon emission and environmental pollution caused by automobiles

Congestion

Carbon Emissions

Pollution

2 : High cost of short–distance transportation



1 Background

Personal private bicycle

Public bicycle system

Bike sharing system

1 Background

Personal private bicycle

Public bicycle system

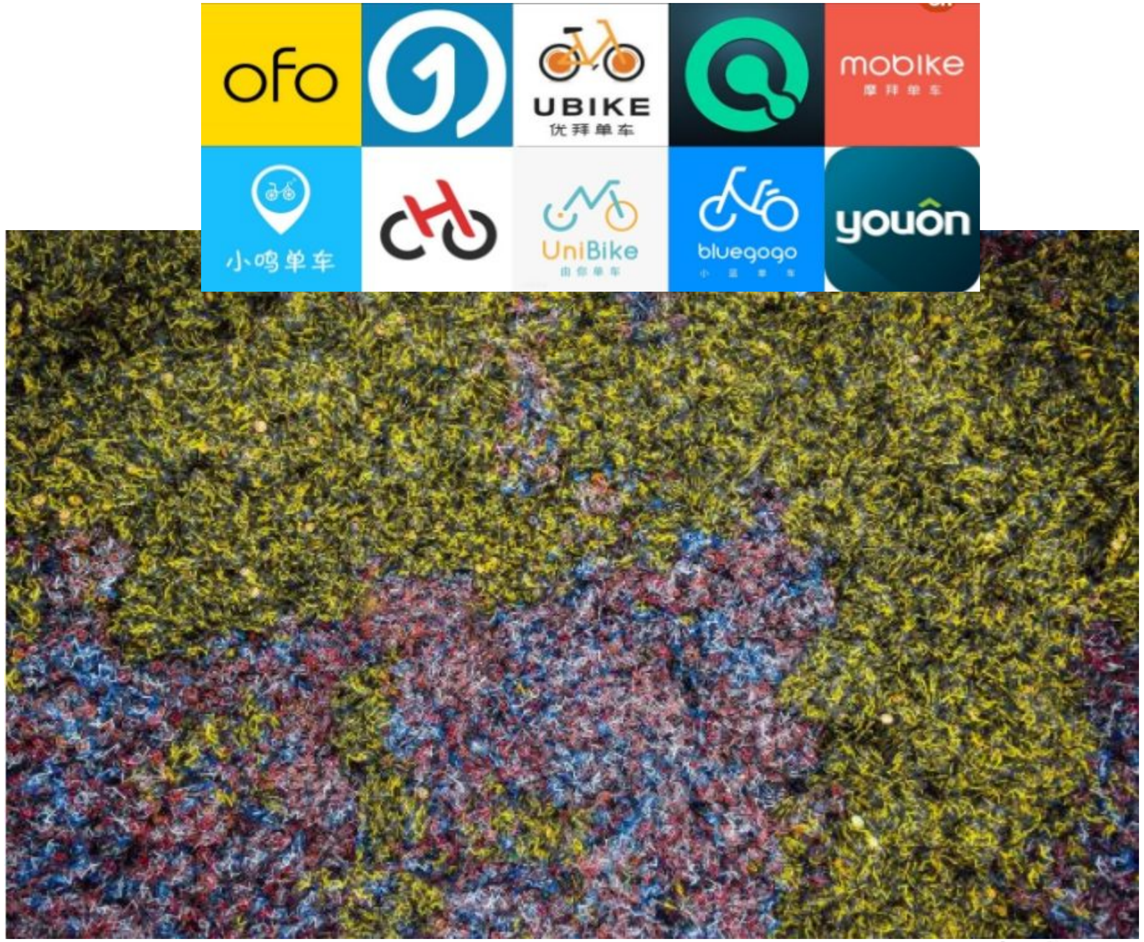
Bike sharing system

1 Background

Personal private bicycle

Public bicycle system

Bike sharing system

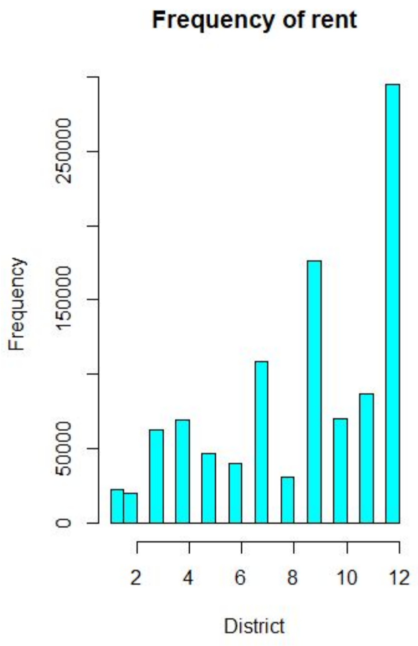


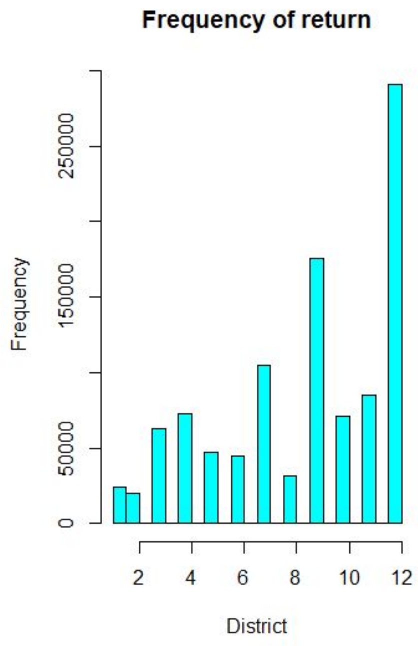
2 Research Question

Focus on how to predict the bikes of the station and the related application

Data Description

Making histogram that riders rented and returned in each district.





	Rent	Return		
1	22525	23845	1	文山區 Wenshan
2	20518	20182	2	南港區 Nangang
3	62627	63132	3	大同區 Datong
4	69937	73025	4	士林區 Shilin
5	46671	46680	5	北投區 Beitou
6	40199	45198	6	萬華區 Wanhua
7	108519	104964	7	中正區 Zhongzheng
8	33197	31284	8	內湖區 Neihu
9	176124	175707	9	信義區 Xinyi
10	70769	71299	10	松山區 Songshan
11	87331	84740	11	中山區 Zhongshan
12	294314	290475	12	大安區 Daan

Data Acquisition

1. Crawling the distance data between MRT stations and each stations from Google maps.

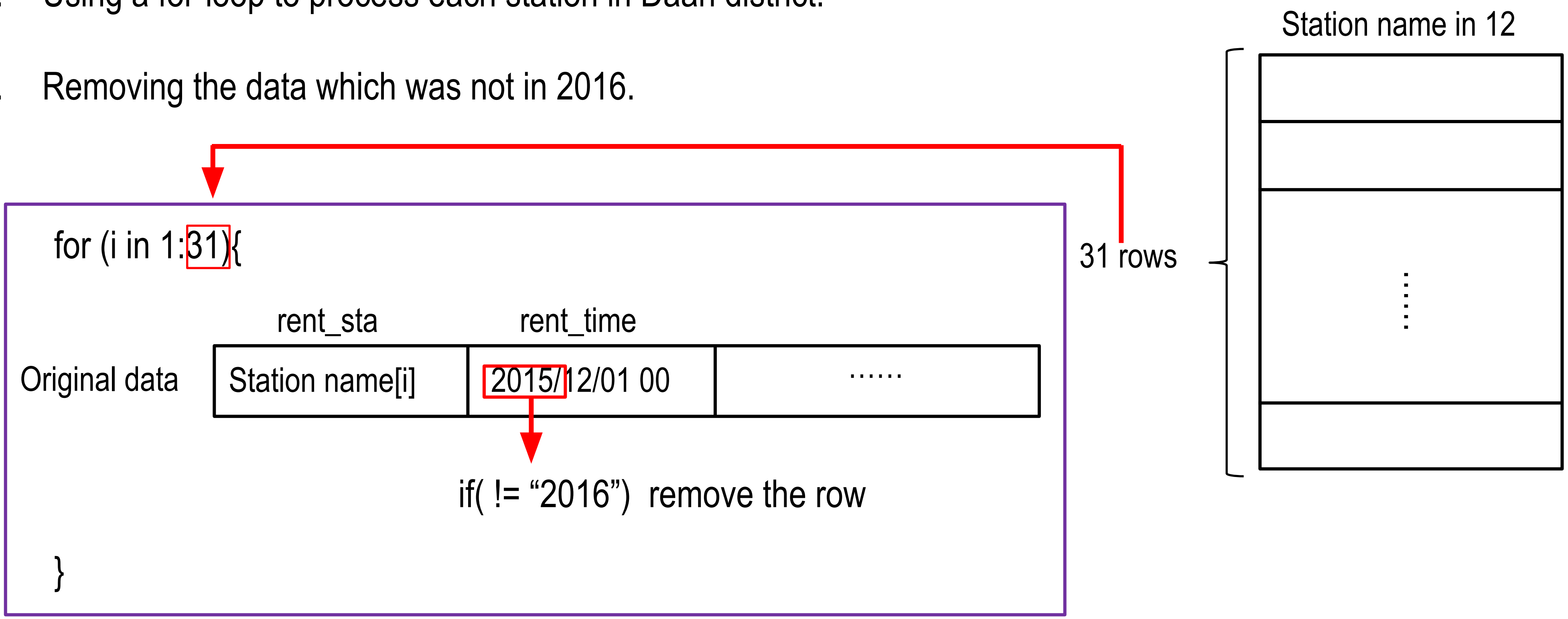
2. Choosing the minimum distance between station and the nearest MRT station of each station, and counted the number of MRT station which distance was less than 1 km from the station.

3. Downloading the temperature and hourly precipitation in Taipei from the CWB Observation Data Inquire System [1].

Data Preprocessing (Rent Data)

1. Using a for-loop to process each station in Daan district.

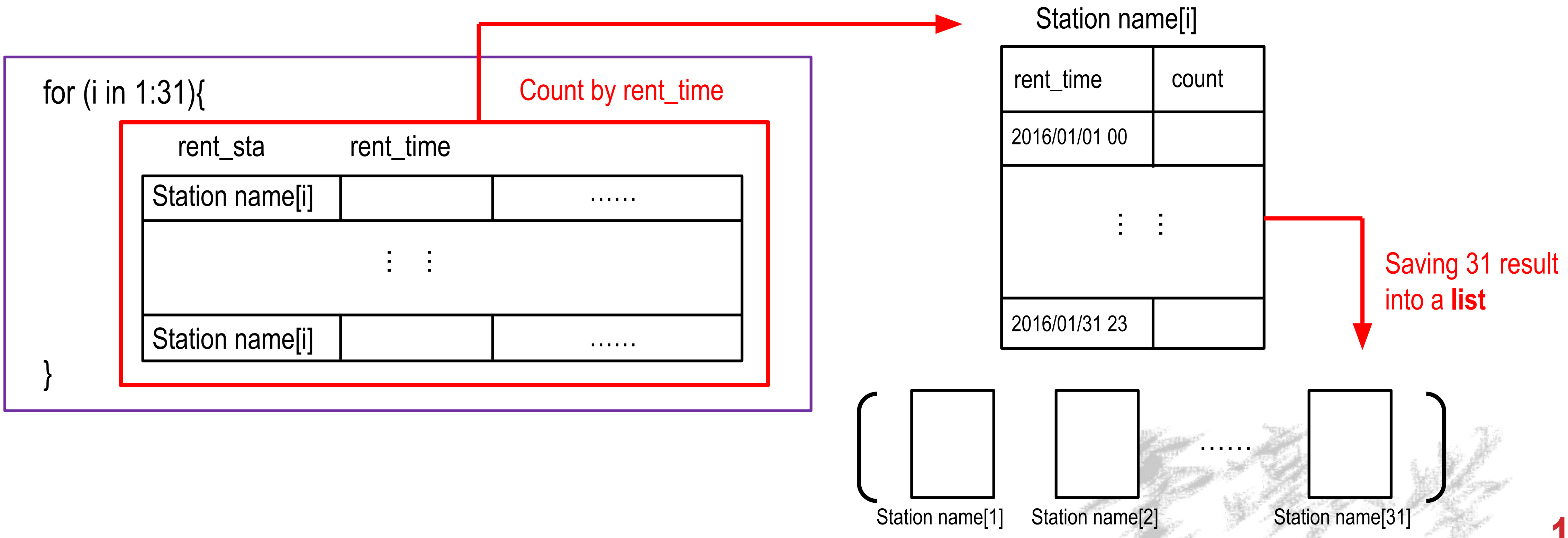
2. Removing the data which was not in 2016.



Data Preprocessing (Rent Data)

3. Counting the number of renting by the time when the rider rented the bike (rent_time).

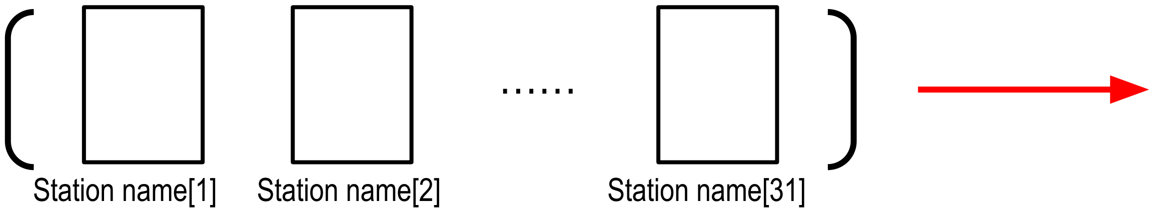
4. Saving the result of counting in each station into a **list**, so there were 31 little dataframe in the list.



Data Preprocessing (Rent Data)

5. There were some missing data, and we assigned them as 0.

   Then, putting the 31 little matrices into a temporary matrix.



	2016/01/01 00	2016/01/01 01	……	2016/01/31 23
Station name[1]	Count(1,01 00)	Count(1,01 01)	……	Count(1,31 23)
Station name[2]	Count(2,01 00)	Count(2,01 01)	……	Count(2,31 23)
⋮ ⋮	⋮ ⋮	⋮ ⋮	……	⋮ ⋮
Station name[31]	Count(31,01 00)	Count(31,01 01)	……	Count(31,31 23)

Station name[i]

rent_time	count
2016/01/01 00	
2016/01/01 03	
⋮ ⋮	
2016/01/31 23	

2016/01/01 01 and 2016/01/01 03 do not exist, so we assigned them as 0.

Data Preprocessing (Rent Data)

6. Putting this matrix into a new matrix as a panel data, and combine the data we collected from web crawling.

	2016/01/01 00	2016/01/01 01	……	2016/01/31 23
Station name[1]	Count(1,01 00)	Count(1,01 01)	……	Count(1,31 23)
Station name[2]	Count(2,01 00)	Count(2,01 01)	……	Count(2,31 23)
⋮ ⋮	⋮ ⋮	⋮ ⋮	……	⋮ ⋮
Station name[31]	Count(31,01 00)	Count(31,01 01)	……	Count(31,31 23)

Station	ObsTime	Rent	Tem	Prep	MinDis	NumMRT
Station name[1]	2016/01/01 00					
Station name[1]	2016/01/01 01					
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Station name[9]	2016/01/21 15					
Station name[9]	2016/01/21 16					
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Station name[31]	2016/01/31 22					
Station name[31]	2016/01/31 23					

Data Preprocessing (Return Data)

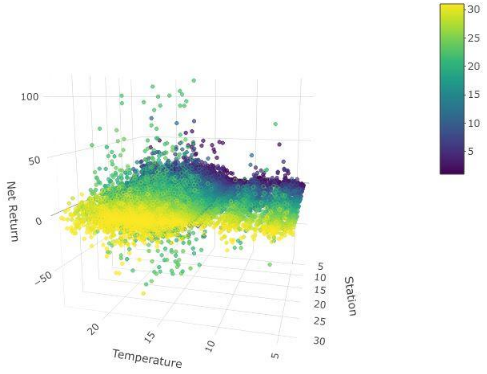
The same steps as Rent Data.

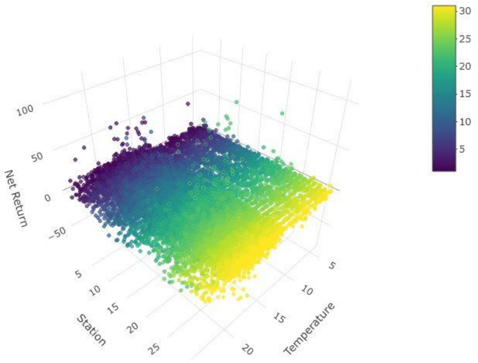
Station	ObsTime	Return	Tem	Prep	MinDis	NumMRT
Station name[1]	2016/01/01 00					
Station name[1]	2016/01/01 01					
⋮ ⋮	⋮	⋮	⋮	⋮	⋮	⋮
Station name[9]	2016/01/21 15					
Station name[9]	2016/01/21 16					
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Station name[31]	2016/01/31 22					
Station name[31]	2016/01/31 23					

Data Preprocessing (Net Return Data:)

Net return was the number of times returning minus renting.

Transformed Station and ObsTime column into categorical indices.





Station	ObsTime	Return	Tem	Prep	MinDis	NumMRT	Net Return
Station name[1]	2016/01/01 00						Return - Rent
Station name[1]	2016/01/01 01						Return - Rent
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Station name[9]	2016/01/21 15						Return - Rent
Station name[9]	2016/01/21 16						Return - Rent
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Station name[31]	2016/01/31 22						Return - Rent
Station name[31]	2016/01/31 23						Return - Rent

Building Models

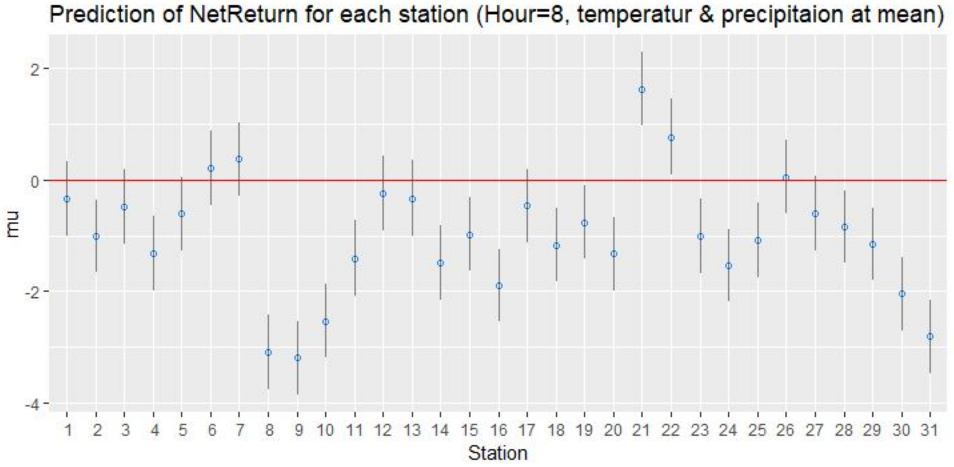
Review the dataset (23064 rows, 11 columns):

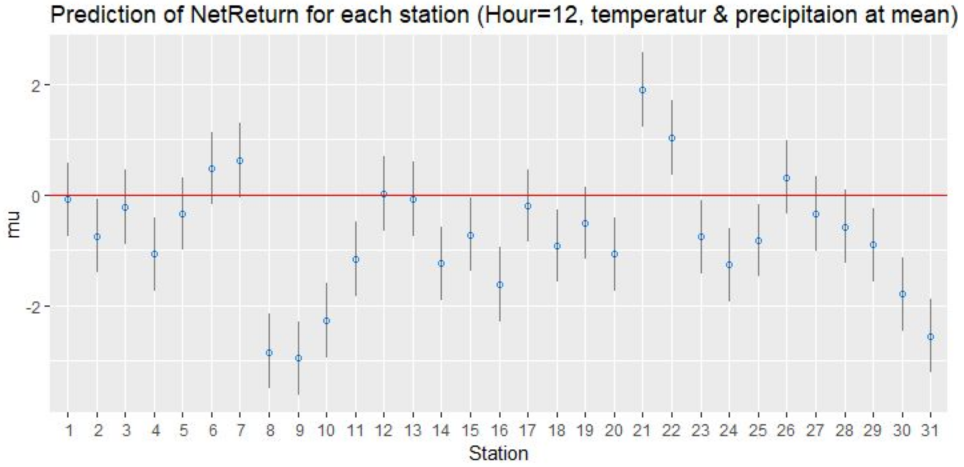
a) Station: name of each Youbike station
b) ObsTime: date and hour of observations (hours from 0 to 23)
c) Temperature: the temperature at each hour of a day
d) Precipitation: the precipitation at each hour of a day
e) MinDisMRT: the distance between Youbike station and the nearest metro station.
f) NumMRT: number of metro stations whose distance is less than 1km from the Youbike station
g) StationIndex: transform Station into categorical indices (from 1 to 31)
h) Hour: transform ObsTime into categorical indices (from 1 to 24)
i) Rent: number of bikes rented at the hour
j) Return: number of bikes returned at the hour
k) NetReturn: the difference of Return from Rent (Return - Rent)

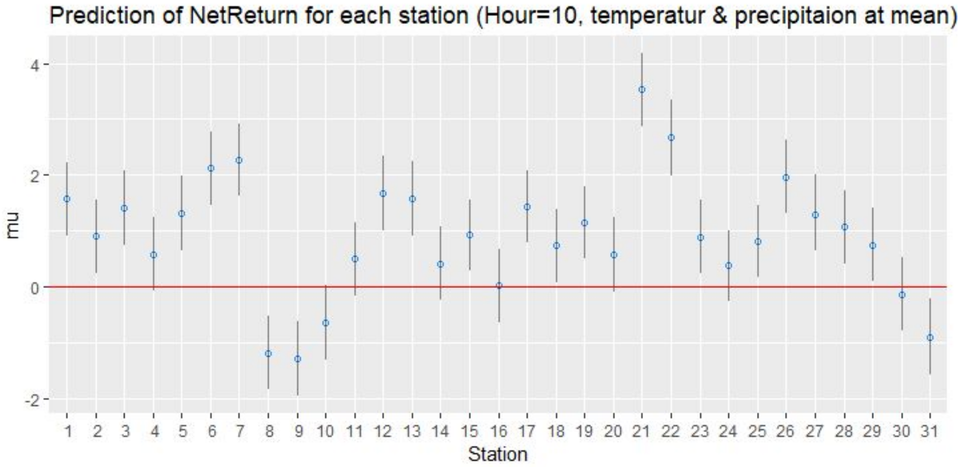
Our model should look like this:

$$NetReturn \sim normal(mu, sigma)$$
$$mu = \alpha\_S_{station[i]} + \alpha\_H_{hour[j]} + \beta t_{hour[j]} * Temperature + \beta p_{hour[j]} * Precipitation$$
$$\alpha\_S_{station[i]} \sim normal(0, 1)$$
$$\alpha\_H_{hour[j]} \sim normal(0, 1)$$
$$\beta t_{hour[j]} \sim normal(0, 1)$$
$$\beta p_{hour[j]} \sim normal(0, 1)$$
$$Sigma \sim exponential(1)$$
$$\forall i = 1, ..., 31$$
$$\forall j = 1, ..., 24$$

Net return values for each station in the morning (Hour = 8, 10, 12)

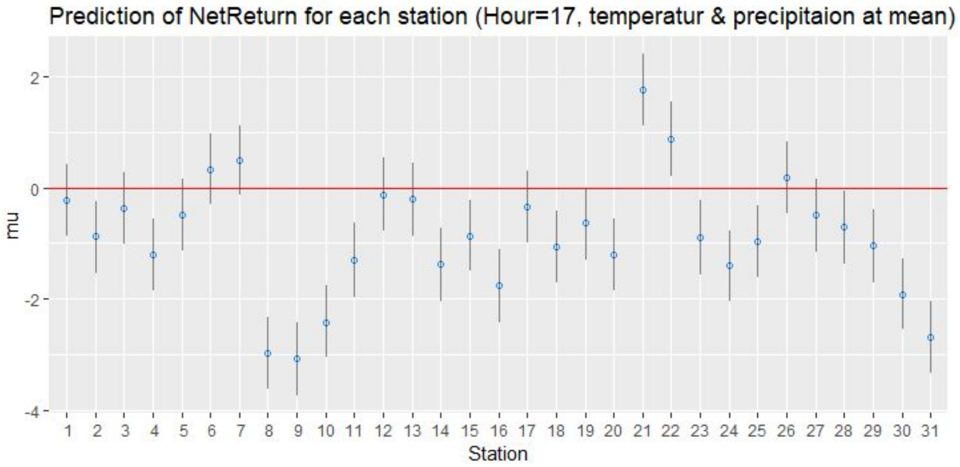


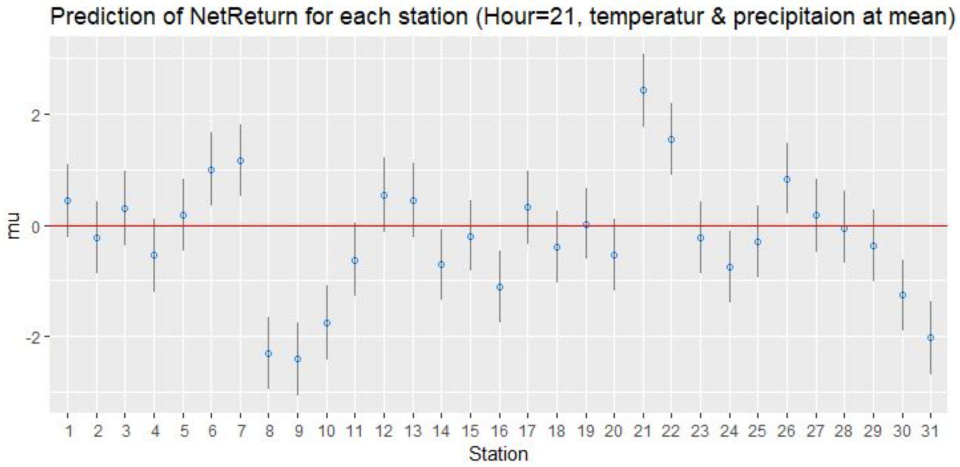


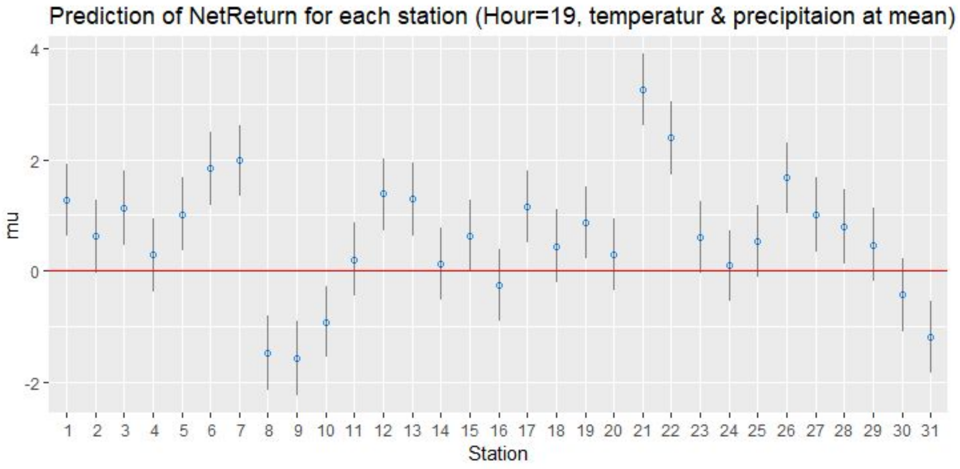


5 Results and Counterfactual Plots

Net return values for each station at night (Hour = 17, 19, 21)

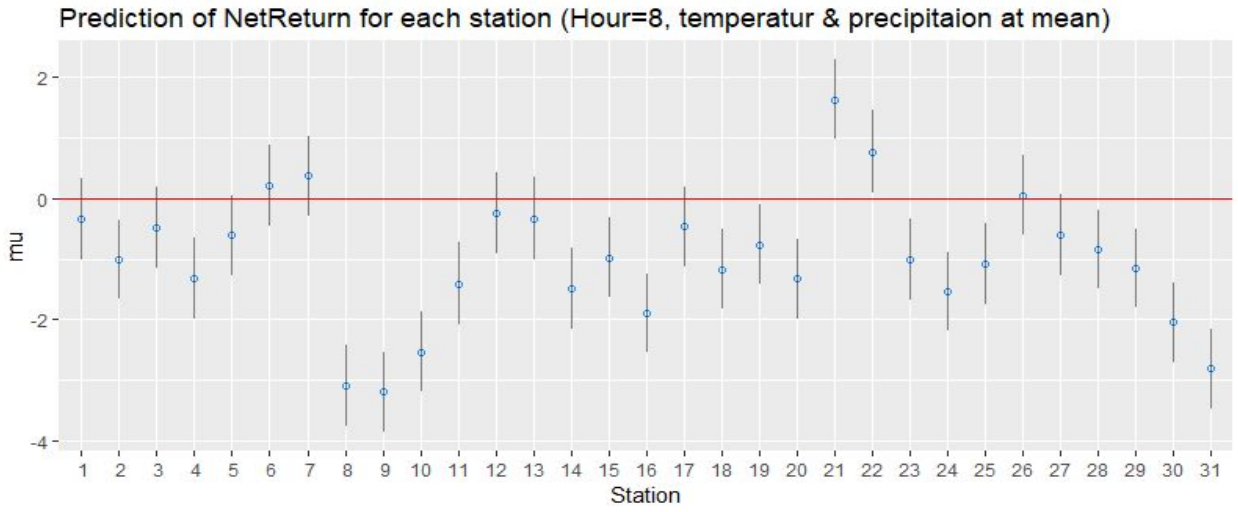


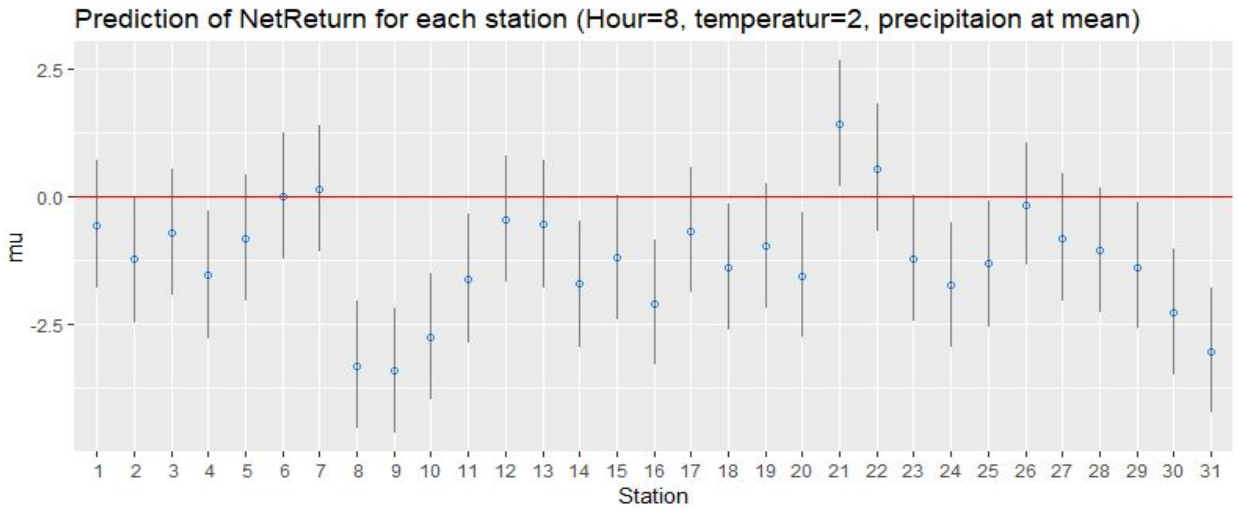




5 Results and Counterfactual Plots

Second, we also want to know how temperature affect the net return values.

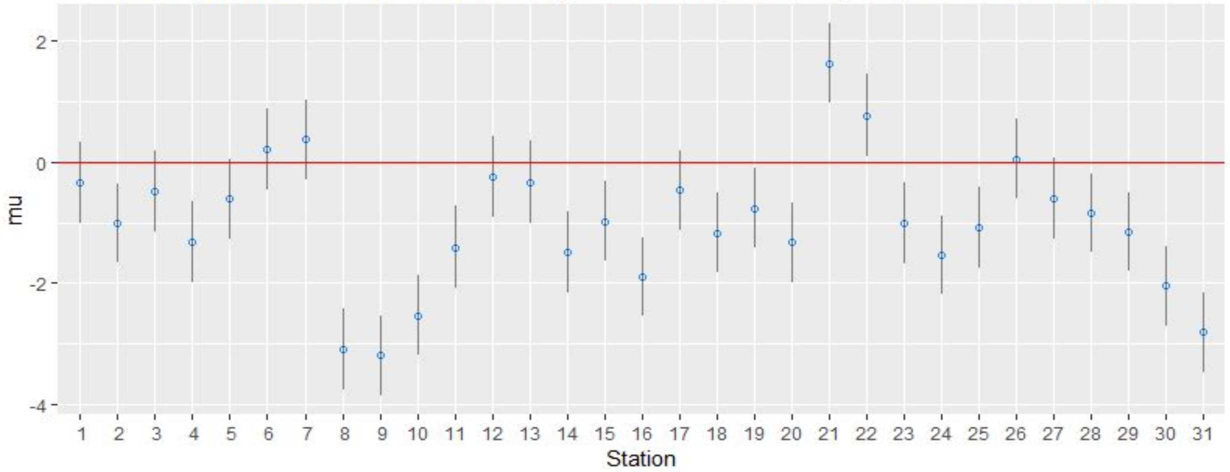




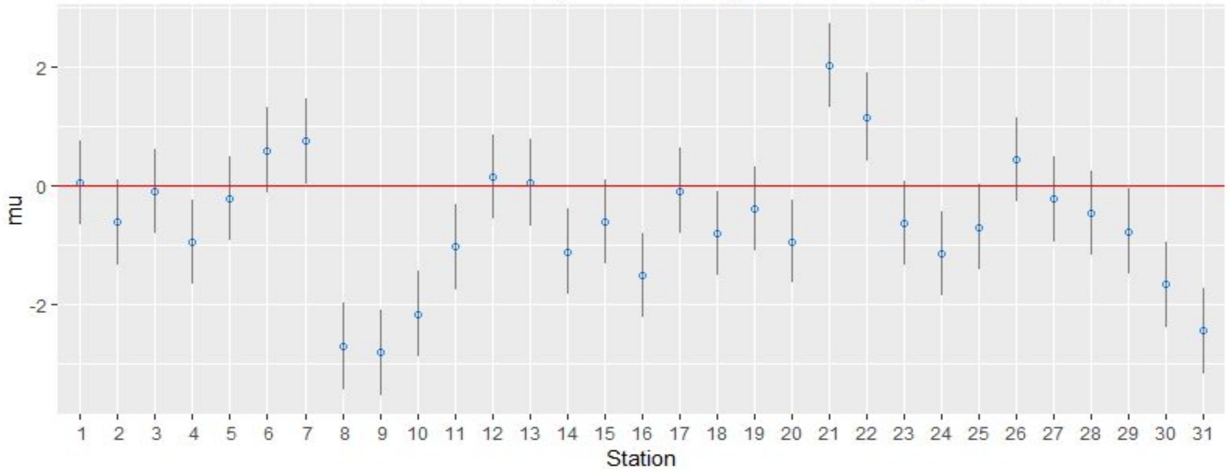
5 Results and Counterfactual Plots

Third, precipitation may also affect the net return values.

Prediction of NetReturn for each station (Hour=8, temperatur & precipitaion at mean)

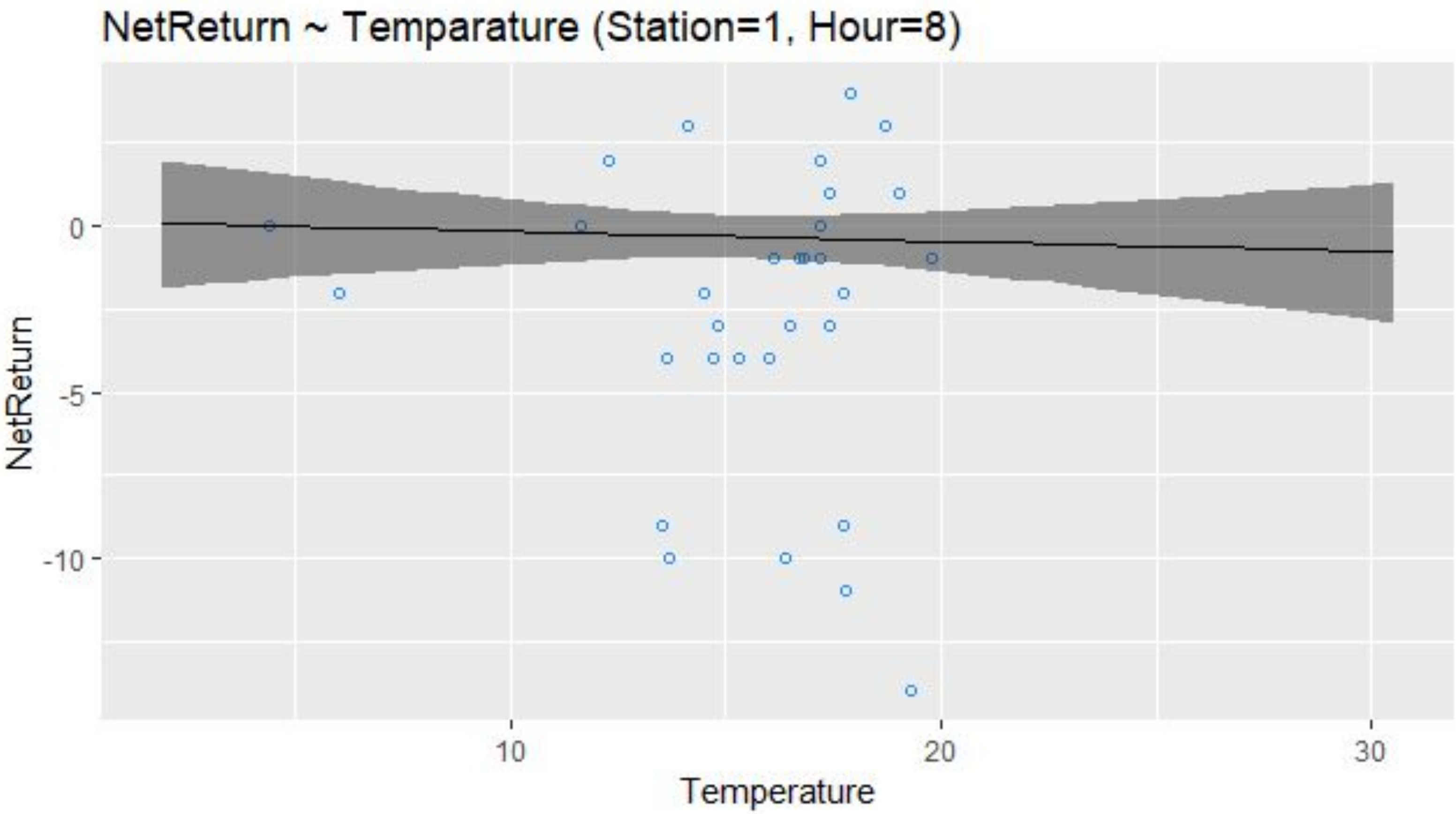


Prediction of NetReturn for each station (Hour=8, precipitaion=2, temperatur at mean)



5 Results and Counterfactual Plots

How about Station = 1 with Hour = 8, where unscaled temperature = 0 to 30?

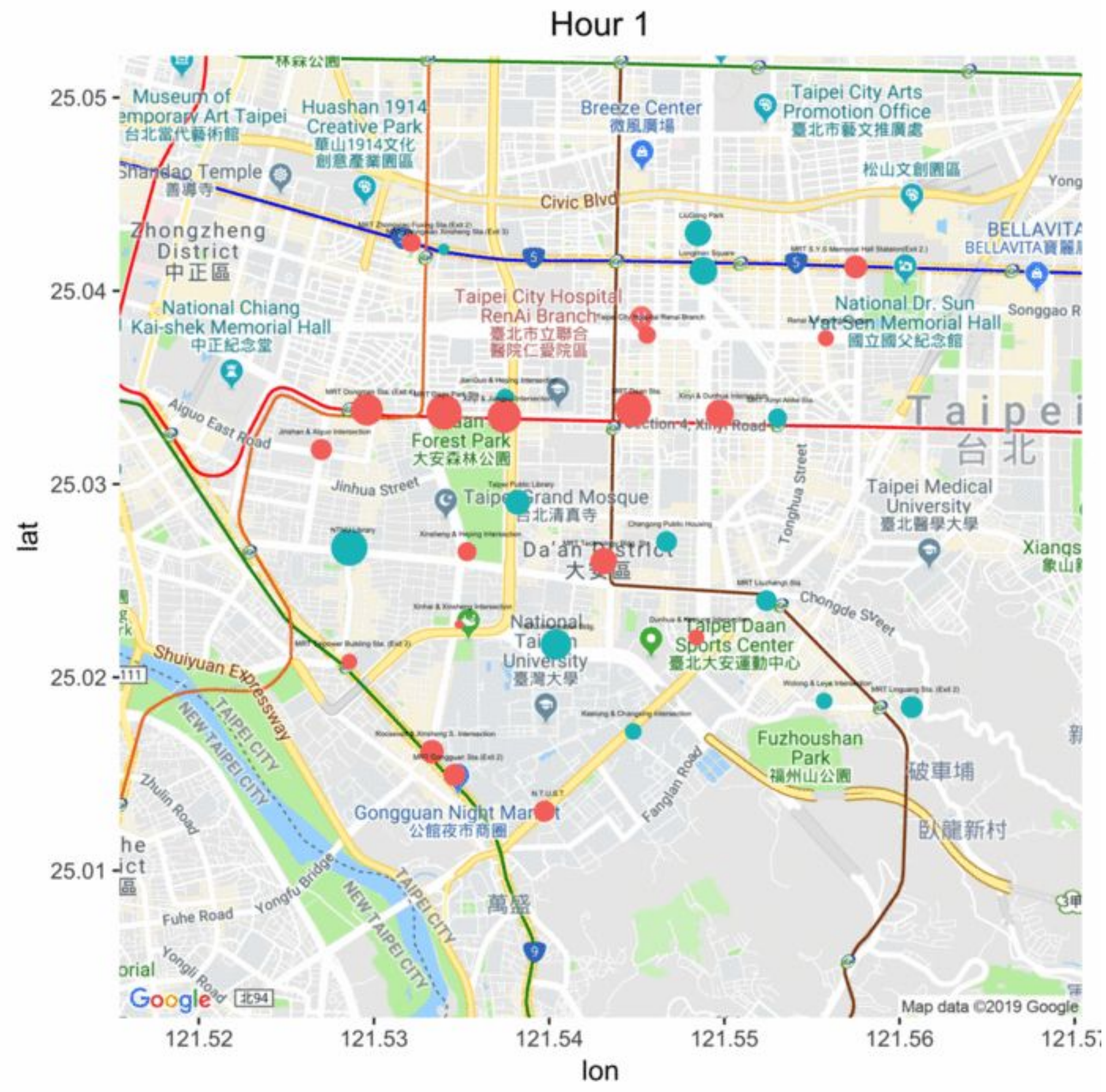


6 Results and ggmap

We want to explore some potential insights regarding net return value and the stations' location and surroundings.

1. We expect to see some patterns or differences in net return values regarding each station's location.
2. For convenience, we set temperature and precipitation at their mean values and only change the "hour" parameter.

6 Results and ggmap



This gif is the visualization of net return prediction in Da-an district.

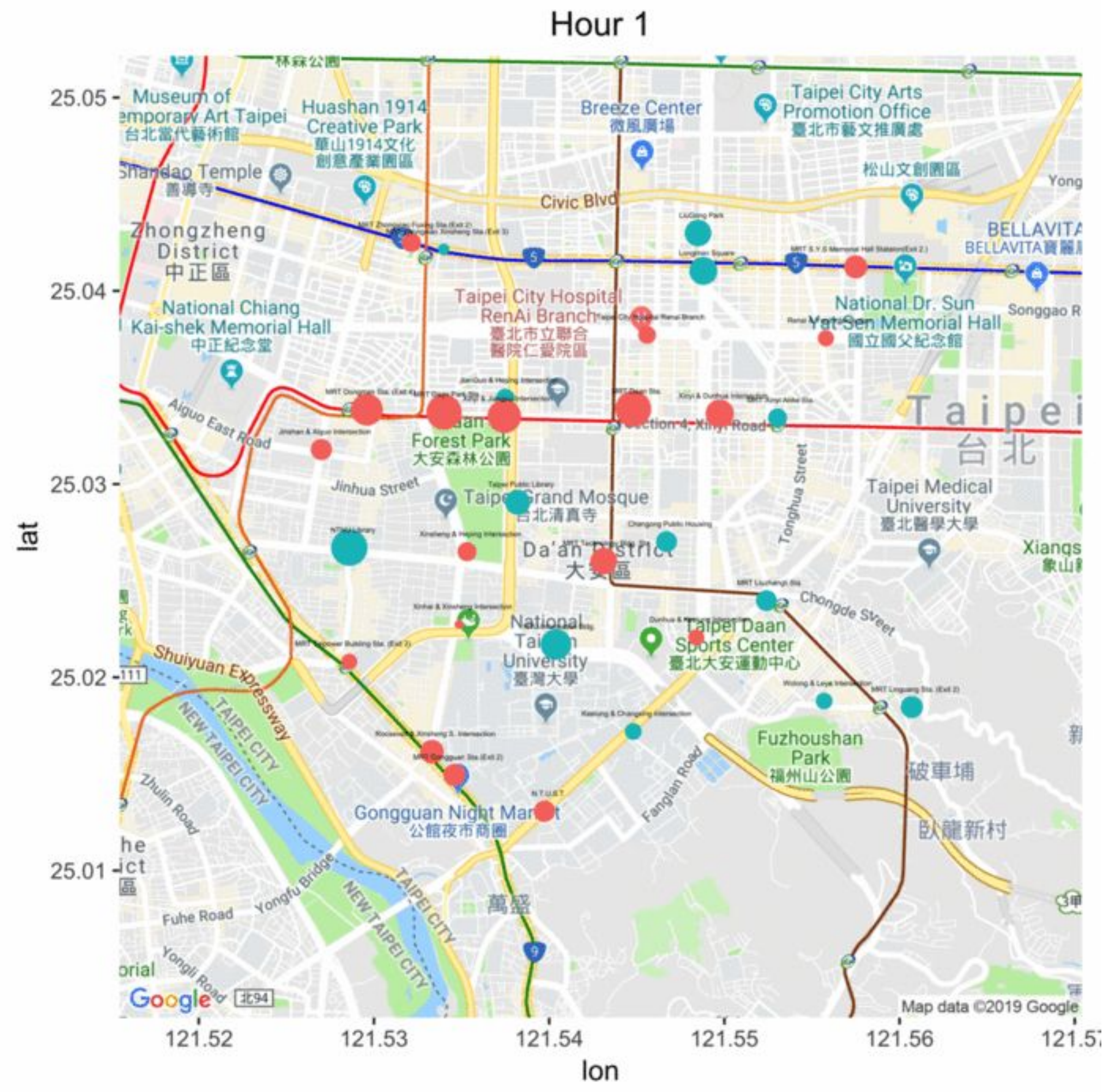
6 Results and ggmap

Observations:

1. The demand and supply of each station is constantly changing.

2. Youbike stations located on the MRT line tends to have negative net return value, especially the stations on the red line.

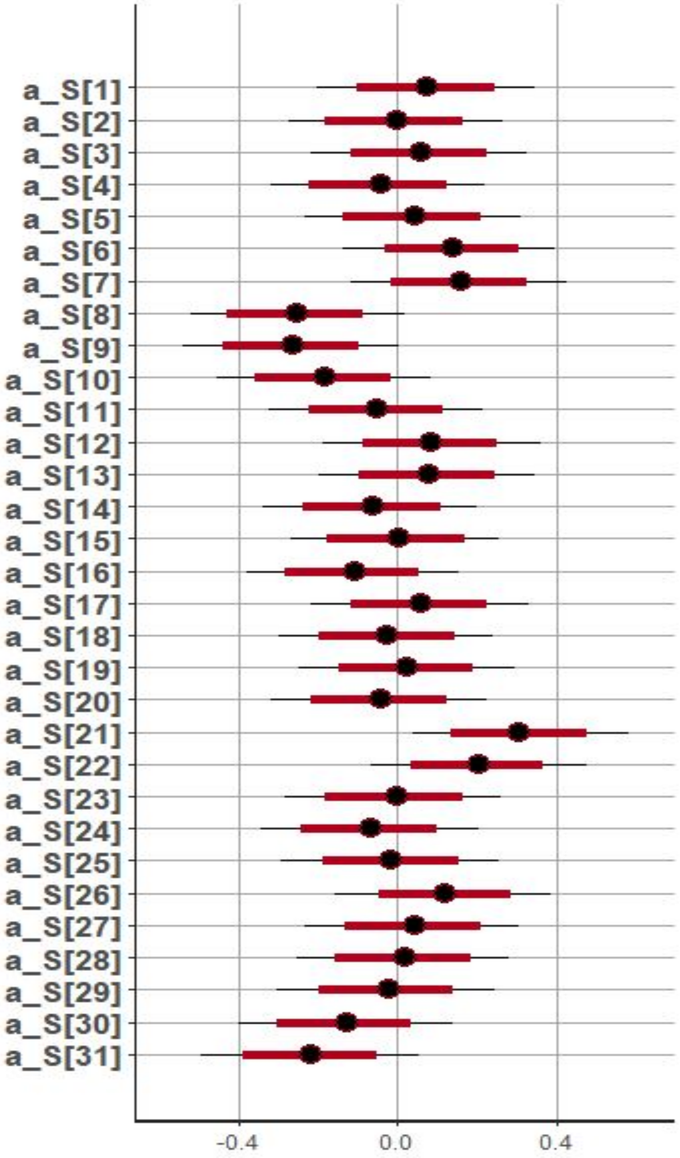
6 Results and ggmap

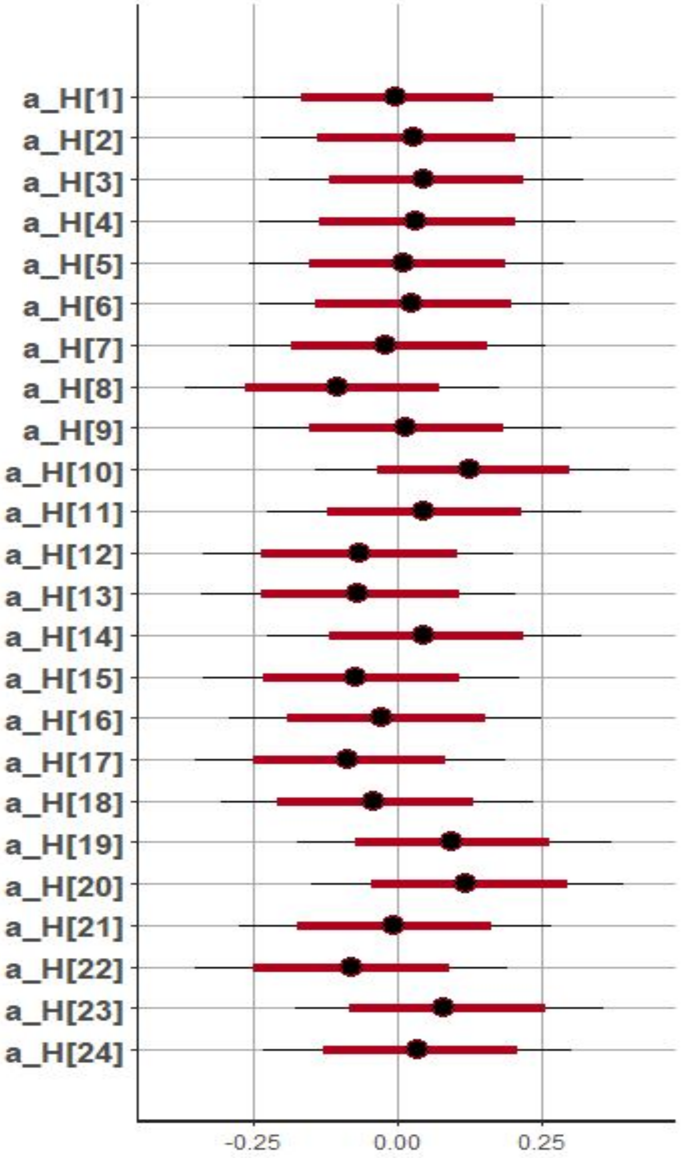


Insights:

1. The habit of resident in Da-an district is taking MRT to the closest station first, and then using youbike to reach their final destination.

If we can increase the number of people using youbike to travel from stations far from MRT to the stations on the MRT line, we might balance the net return value at each station and decrease the cost of operating staffed trucks for manual relocation.


a_S[1]
a_S[2]
a_S[3]
a_S[4]
a_S[5]
a_S[6]
a_S[7]
a_S[8]
a_S[9]
a_S[10]
a_S[11]
a_S[12]
a_S[13]
a_S[14]
a_S[15]
a_S[16]
a_S[17]
a_S[18]
a_S[19]
a_S[20]
a_S[21]
a_S[22]
a_S[23]
a_S[24]
a_S[25]
a_S[26]
a_S[27]
a_S[28]
a_S[29]
a_S[30]
a_S[31]
-0.4
0.0
0.4


a_H[1]
a_H[2]
a_H[3]
a_H[4]
a_H[5]
a_H[6]
a_H[7]
a_H[8]
a_H[9]
a_H[10]
a_H[11]
a_H[12]
a_H[13]
a_H[14]
a_H[15]
a_H[16]
a_H[17]
a_H[18]
a_H[19]
a_H[20]
a_H[21]
a_H[22]
a_H[23]
a_H[24]
-0.25
0.00
0.25

Second model:

$$NetReturn \sim normal\ (\ mu\ ,\ sigma\ )$$

$$mu = \alpha + ßm * MinDisMRT + ßn * NumMRT + \alpha\_H_{hour[j]} +$$
$$\beta t_{hour[j]} * Temperature + \beta p_{hour[j]} * Precipitation$$

$$\alpha \sim normal\ (\ 0\ ,\ 1\ )$$
$$ßm \sim normal\ (\ 0\ ,\ 1\ )$$
$$ßn \sim normal\ (\ 0\ ,\ 1\ )$$
$$\alpha\_H_{hour[j]} \sim normal\ (\ 0\ ,\ 1\ )$$
$$\beta t_{hour[j]} \sim normal\ (\ 0\ ,\ 1\ )$$
$$\beta p_{hour[j]} \sim normal\ (\ 0\ ,\ 1\ )$$
$$Sigma\ \sim exponential\ (\ 1\ )$$
$$\forall j = 1, \ldots, 24$$

Third model:

$$NetReturn \sim normal\,(\,mu\,,\ sigma\,)$$
$$mu = \alpha\_S_{station[i]} + \alpha\_H_{hour[j]}$$
$$\alpha\_S_{station[i]} \sim normal\,(\,0\,,\ 1\,)$$
$$\alpha\_H_{hour[j]} \sim normal\,(\,0\,,\ 1\,)$$
$$Sigma \sim exponential\,(\,1\,)$$
$$\forall i = 1, \ldots, 31$$
$$\forall j = 1, \ldots, 24$$

Comparing Models

By comparing the WAIC values of each model, we got the following results, which indicated that the third model would be the best model to be considered when predicting the Youbike net return values.

	elpd_diff	se_diff	elpd_waic	p_waic	waic
Model 3	0.0	0.0	-2609.0	3.9	5217.9
Model 1	-0.1	1.5	-2609.1	5.5	5225.9
Model 2	-4.0	2.5	-2613.0	3.6	5225.9

Operation Strategy

With the ability to predict supply and demand at each rental station given the time, temperature, and amount of rainfall, we can use this information to construct a transportation integer programming model to minimize the cost of relocating bicycles.

Operation Strategy

Cij : the cost of transport one bicycle from station i to j
Di : the predicted net return value at station I
Xij : the decision variable indicating number of bicycles needed to be transported from station i to station j
K : the negative utility of the system having unbalance net return value

$$min\ Z = \sum_i \sum_j C_{ij} X_{ij} + k \times \sum_i \left| \left(D_i - \sum_j X_{ij} + \sum_j X_{ji} \right) \right|$$

$s.t.$

$$\forall i,\ D_i \geq \sum_j X_{ij},\ if\ D_i \geq 0,\ D_i \leq \sum_j X_{ji},\ if\ D_i < 0$$

$$\forall ij, C_{ij} \geq 0, X_{ij} \geq 0, X_{ij} \in \mathbb{N}$$

$$k \geq 0$$

Operation Strategy

Possible improvements after solving the model:

1. Plan a better route for the youbike transportation truck
2. Cutdown the staff needed to relocate the bicycles
3. Increase the utilization rate of youbike
4. Improve youbike users' satisfaction by preventing situation where there is no bike to rent or no place to return

Pricing Strategy

We can solve the relocation problem using pricing strategy too.

If we can lower the price of certain route (from station i to station j) by giving special mission to users, we can give youbike users the incentive to relocate X_{ij} bikes for us.

⇒ eliminate the need of manual relocating the bicycles.
⇒ Both the users and youbike company will benefit from this.