

서울시(여의도, 상암)

공용자전거 사용자 분류 및

운동량 예측

따릉이



"따릉이 타고 어디까지 가봤어?" 시민 사랑에 이용건수 1억 돌파

머니투데이 | 기성훈 기자

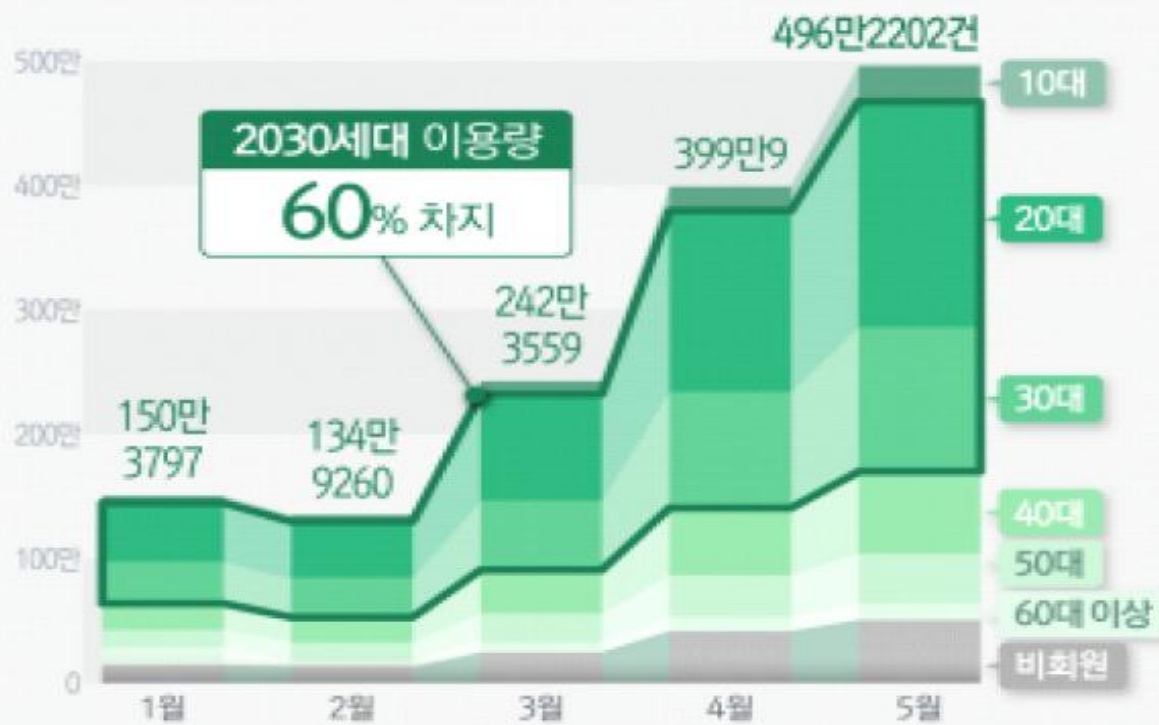
2022.04.27 06:00

| 서울시민 3명 중 1명은 '따릉이' 회원



N 네이버 메인에서 머니투데이 구독
Ch 카카오톡에서 머니투데이 채널 추가

연령대별 이용량 추이 ※2022년 1~5월 기준 단위 건 자료 서울시



자료 서울시, 서울시설공단 그래픽 이해영

따릉이 운영 현황

※서울시 전역(25개 자치구) 기준 *2022년말 예정



01 데이터 설명



서울시 공공데이터(www.data.seoul.go.kr)

서울시 공공자전거 이용정보(시간대별)

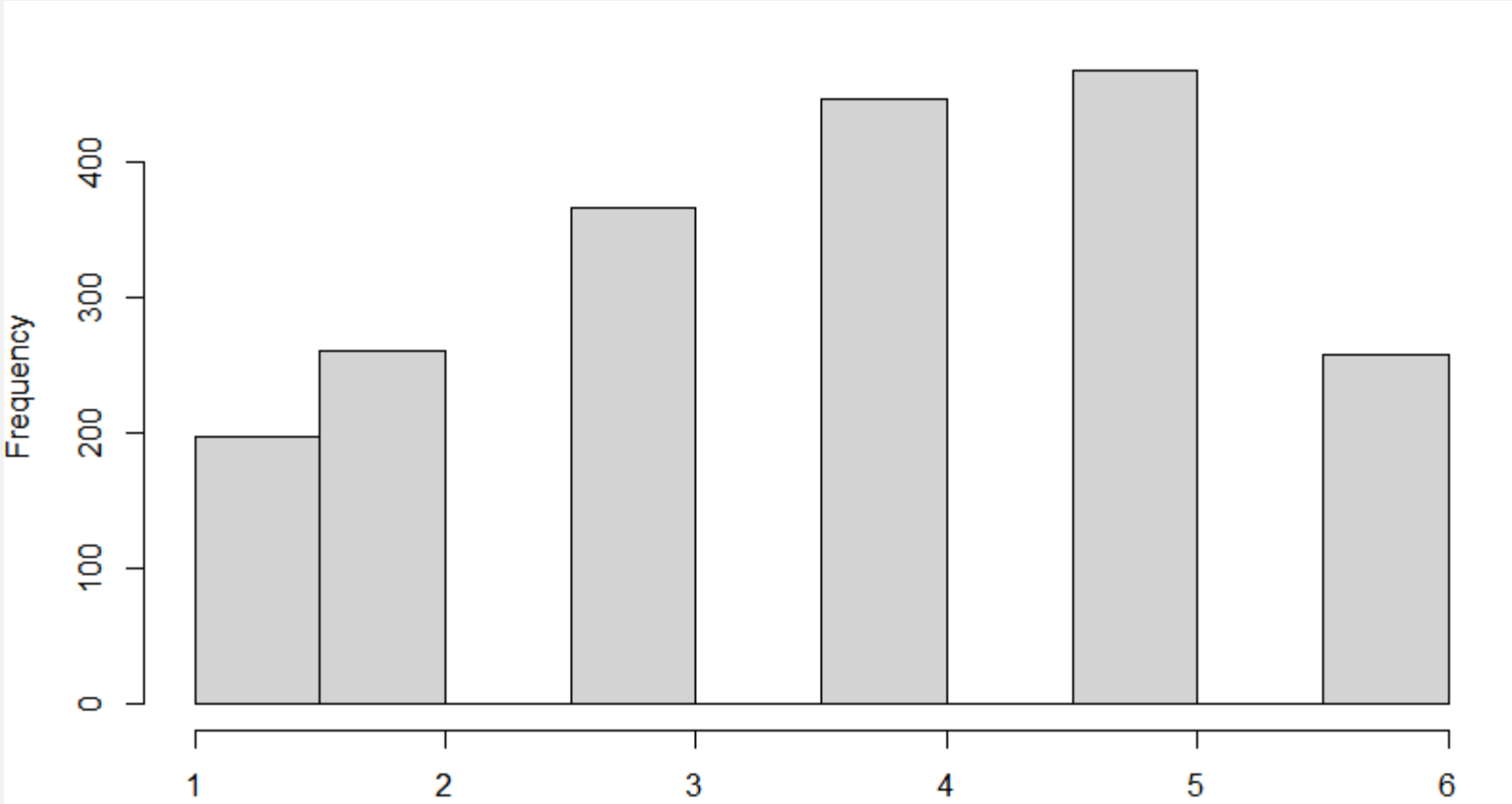
개수	3704328개
시기	2022-06-01 ~ 2022-06-30
RENT_DT	대여일자
RENT_ID	대여소번호
RENT_NM	대여소명
RENT_HR	대여시간
RENT_TYPE	대여구분코드
GENDER_CD	성별
AGE_TYPE	나이
USE_CNT	이용건수
EXER_AMT	운동량
CARBON_AMT	탄소 절감량
MOVE_METER	이동 거리(M)
MOVE_TIME	이동 시간(분)

데이터 전처리

대여 일자	퍼센트
2022-06-01	14.3
2022-06-02	17.1
2022-06-03	15.0
2022-06-04	13.3
2022-06-05	9.8
2022-06-06	11.4
2022-06-07	13.6
2022-06-08	5.6

- 2022-06-01~2022-06-08
- NOT RAINY DAY

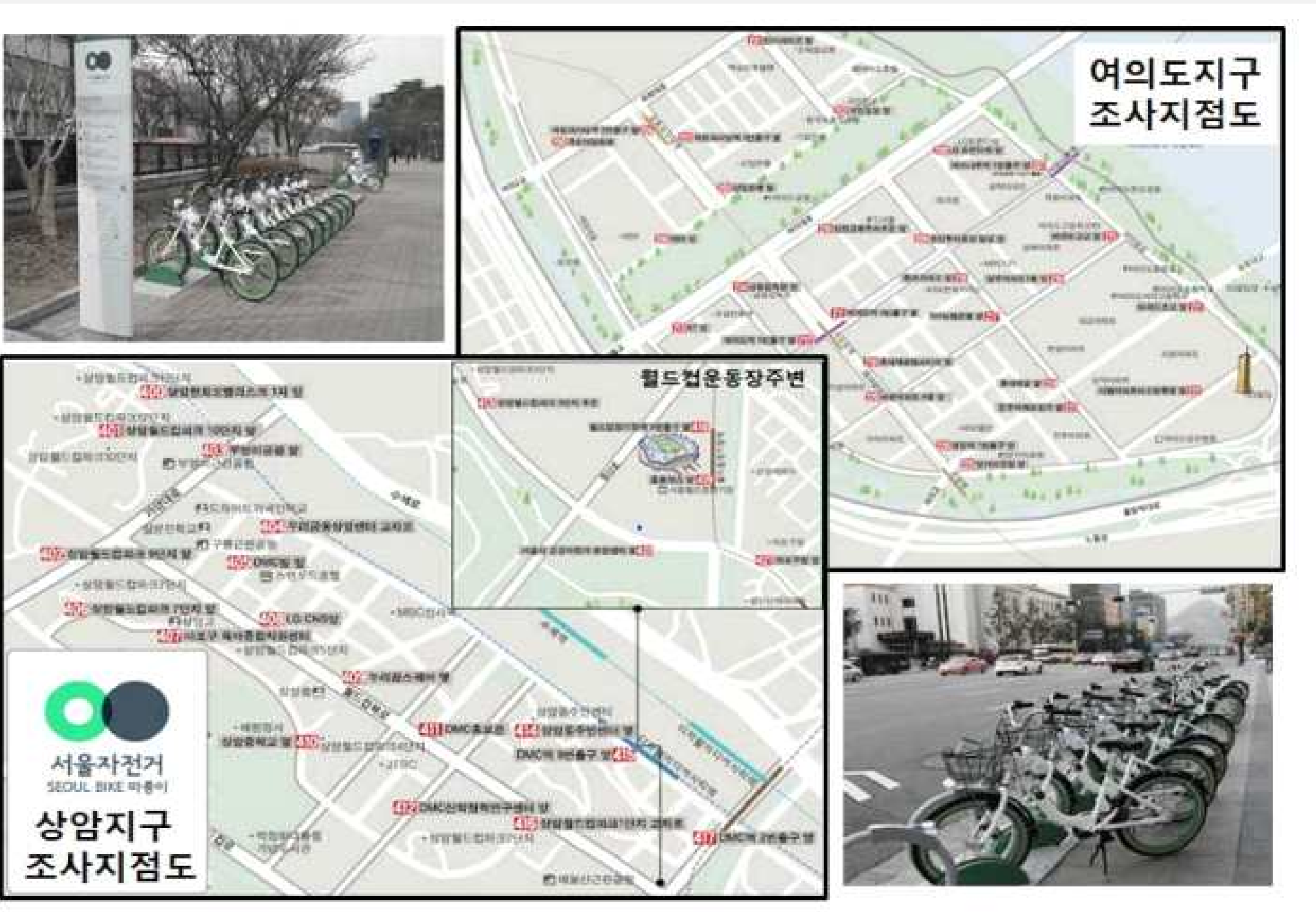
대여 일자



- 00:00~05:59 X
- 06:00~08:59(=1)
- 09:00~11:59(=2)
- 12:00~14:59(=3)
- 15:00~18:59(=4)
- 19:00~21:59(=5)
- 22:00~23:59(=6)

시간대

데이터 전처리



대여소명, 대여서번호 이용

업무(=1)

지하철(=2)

주거(=3)

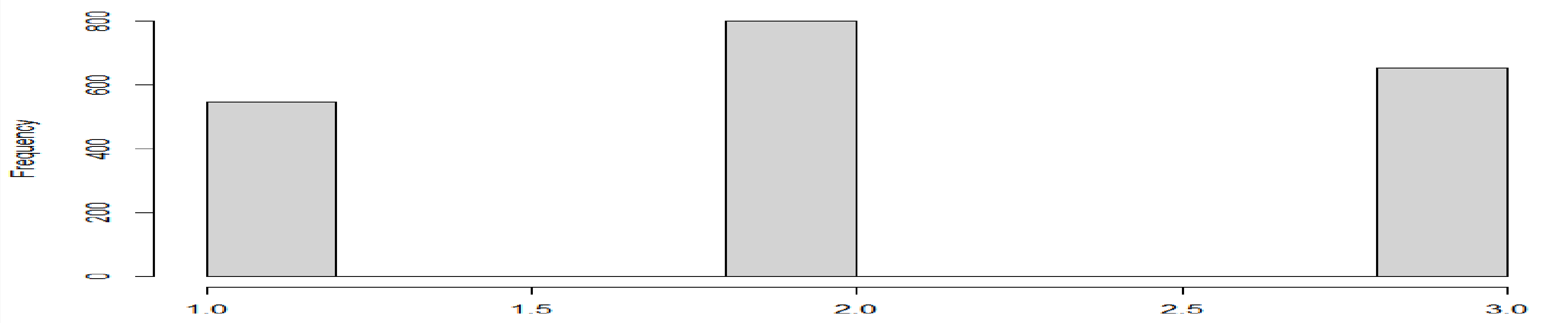
RENT_SPOT

출처: 서울시 공공 자전거 이용 특성에 관한 연구 (장재민 김태형 이무영 2016)

RENT_SPOT

여의도 지구 대여소	지역구분
200 국회위원회관	업무
201 진미파라곤 앞	업무
202 국민일보 앞	업무
203 국회의사당역 3번출구 옆	지하철
204 국회의사당역 5번출구 옆	지하철
205 산업은행 앞	업무
222 시범아파트버스정류장 옆	주거
223 진주아파트상가 앞	주거
224 롯데캐슬 앞	주거
225 양카라공원 앞	지하철
226 샛강역 1번출구 앞	지하철

상암지구 대여소	지역구분
400 상암 한화오벨리스크 1차 앞	주거
401 상암월드컵파크 10단지 앞	주거
402 상암월드컵파크 9단지 앞	주거
403 부영이공원 앞	주거
405 DMC빌 앞	업무
417 DMC역 2번출구 옆	지하철
418 월드컵경기장역 3번출구 옆	지하철
419 홈플러스 앞	지하철
420 서울시 공공자전거 운영센터 옆	업무
421 마고구청 앞	주거



park

여의도 지구 대여소	공원인접	상암지구 대여소	공원 인접
200 국회위원회관	X	400 상암 한화오벨리스크 1차 앞	X
201 진미파라곤 앞	X	401 상암월드컵파크 10단지 앞	X
202 국민일보 앞	0	402 상암월드컵파크 9단지 앞	X
203 국회의사당역 3번출구 옆	X	403 부영이공원 앞	X
204 국회의사당역 5번출구 옆	X	405 DMC빌 앞	X
205 산업은행 앞	0	417 DMC역 2번출구 옆	x
222 시범아파트버스정류장 옆	X	418 월드컵경기장역 3번출구 옆	0
223 진주아파트상가 앞	X	419 홈플러스 앞	0
224 롯데캐슬 앞	X	420 서울시 공공자전거 운영센터 옆	0
225 양카라공원 앞	0	421 마고구청 앞	0
226 샛강역 1번출구 앞	0		

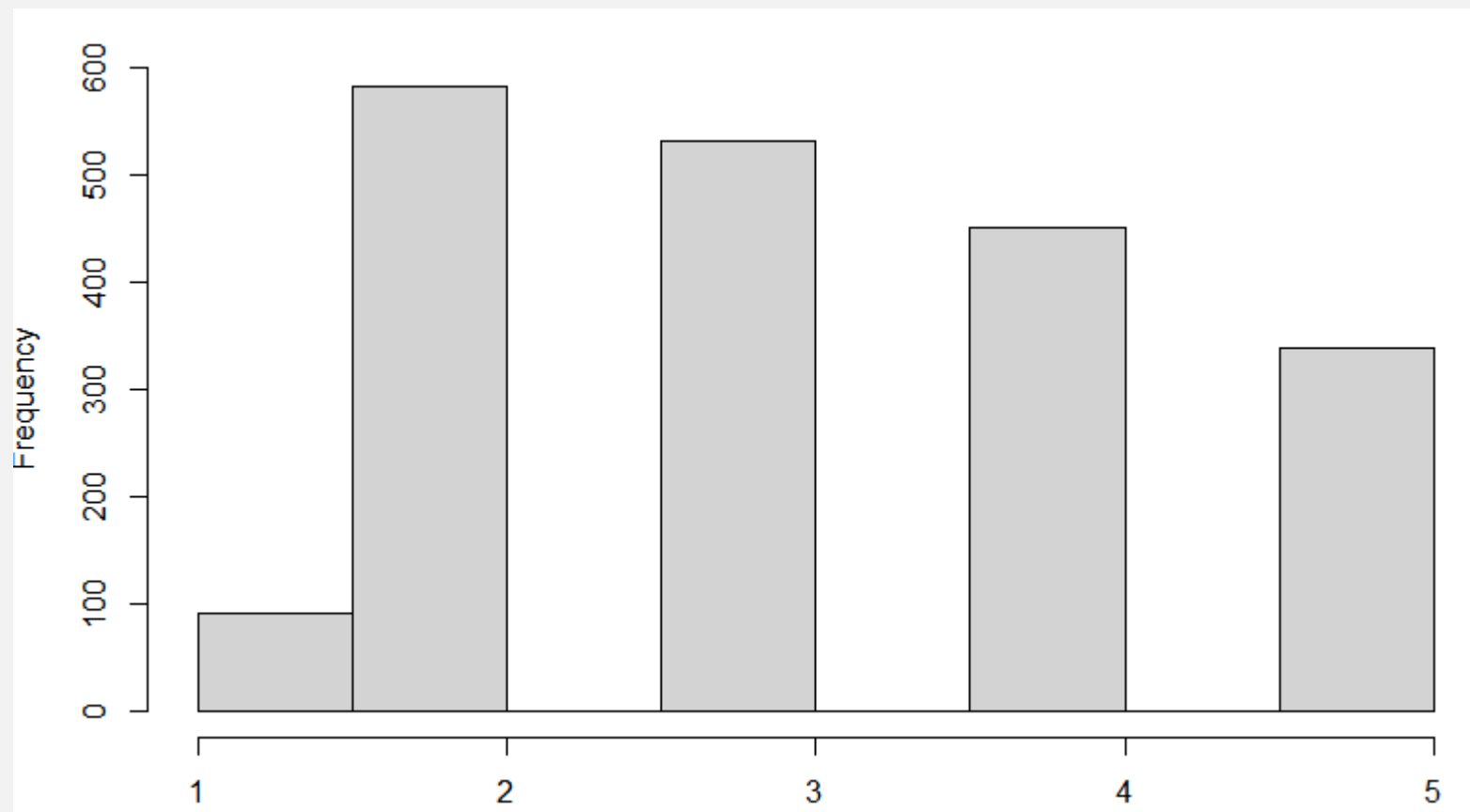
공원 인접	개수
0	1084
x	916

GENDER_CD, AGE_TYPE

성별	개수
남성	1168
여성	832

- 남성(=1)
- 여성(=0)

성별



- 10대 이하(=1)
- 20대(=2)
- 30대(=3)
- 40대(=4)
- 50대 이상(=5)

연령대

USE_CNT, holiday

이용건수	개수
1	1691
>1	309

이용건수	개수
1	1476
0	524

이용건수	개수
1	974
0	1026

- 1(=1)
- >1(=0)

이용건수

- 정기권
(=1)
- 단체권, 일일권, 일일권(비회원)
(=0)

이용권

- 1일 4일 5일 6일
(=1)
- 2일 3일 7일 8일
(=0)

휴일

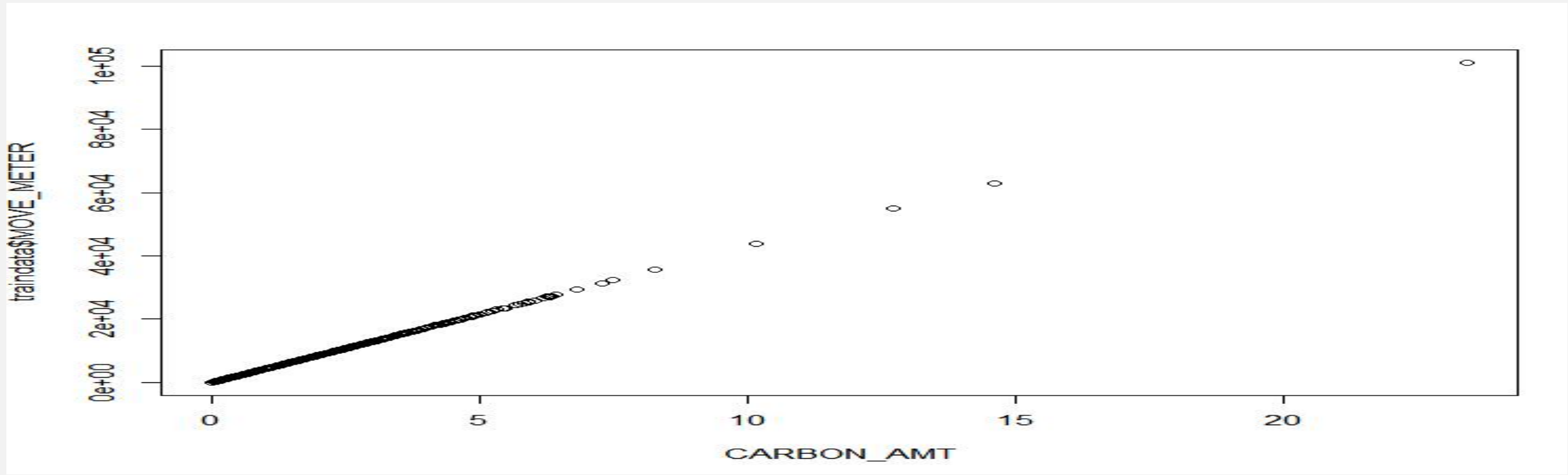
CARBON_AMT, MOVE_TIME

Min	1 ST QU	MEDIAN	MEAN	3 RD QU	MAX
0	0.26	0.58	1.011	1.250	23.430

탄소절감량

Min	1 ST QU	MEDIAN	MEAN	3 RD QU	MAX
0	1130	2490	4360	5390	101099

이동 거리



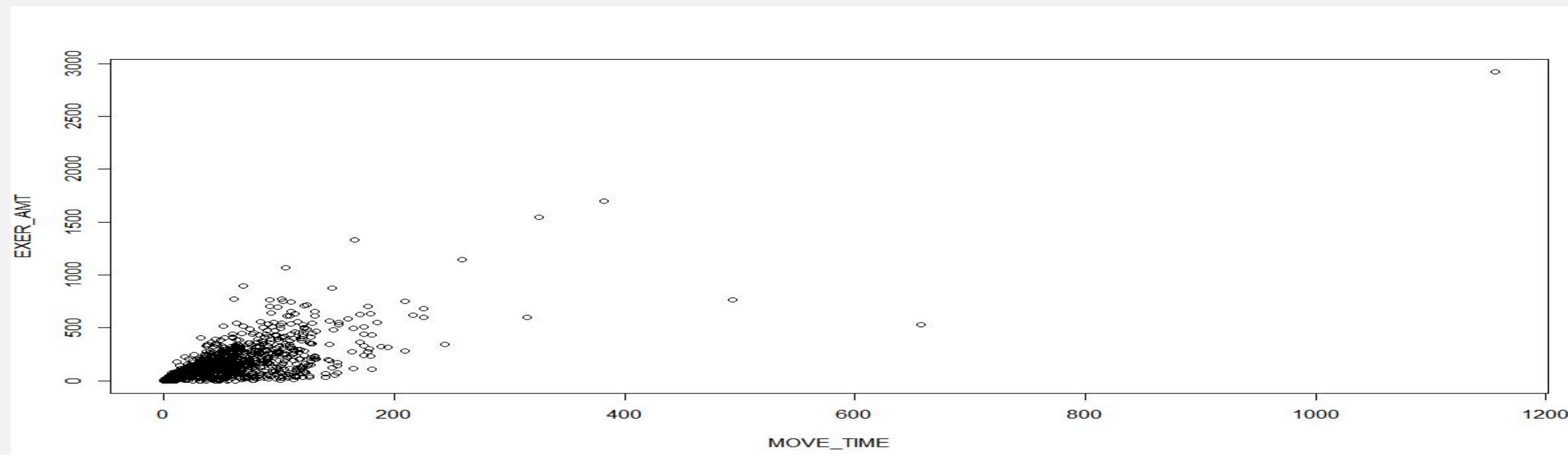
MOVE_TIME&EXER_AMT

이용시간

Min	1 ST QU	MEDIAN	MEAN	3 RD QU	MAX
0	9	23	38.54	53	1156

운동량

Min	1 ST QU	MEDIAN	MEAN	3 RD QU	MAX
0	23.69	64.90	115.09	142.37	2922.00



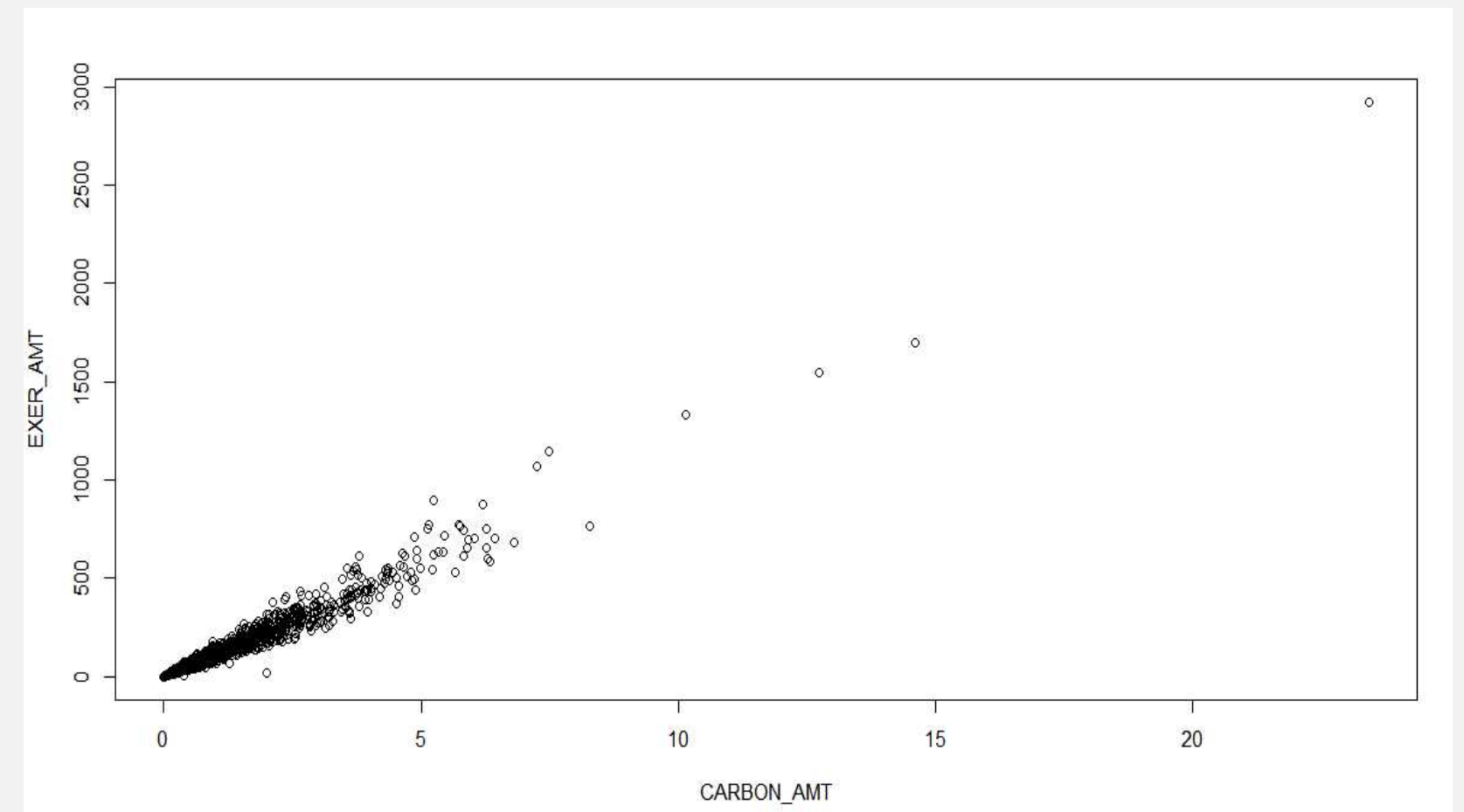
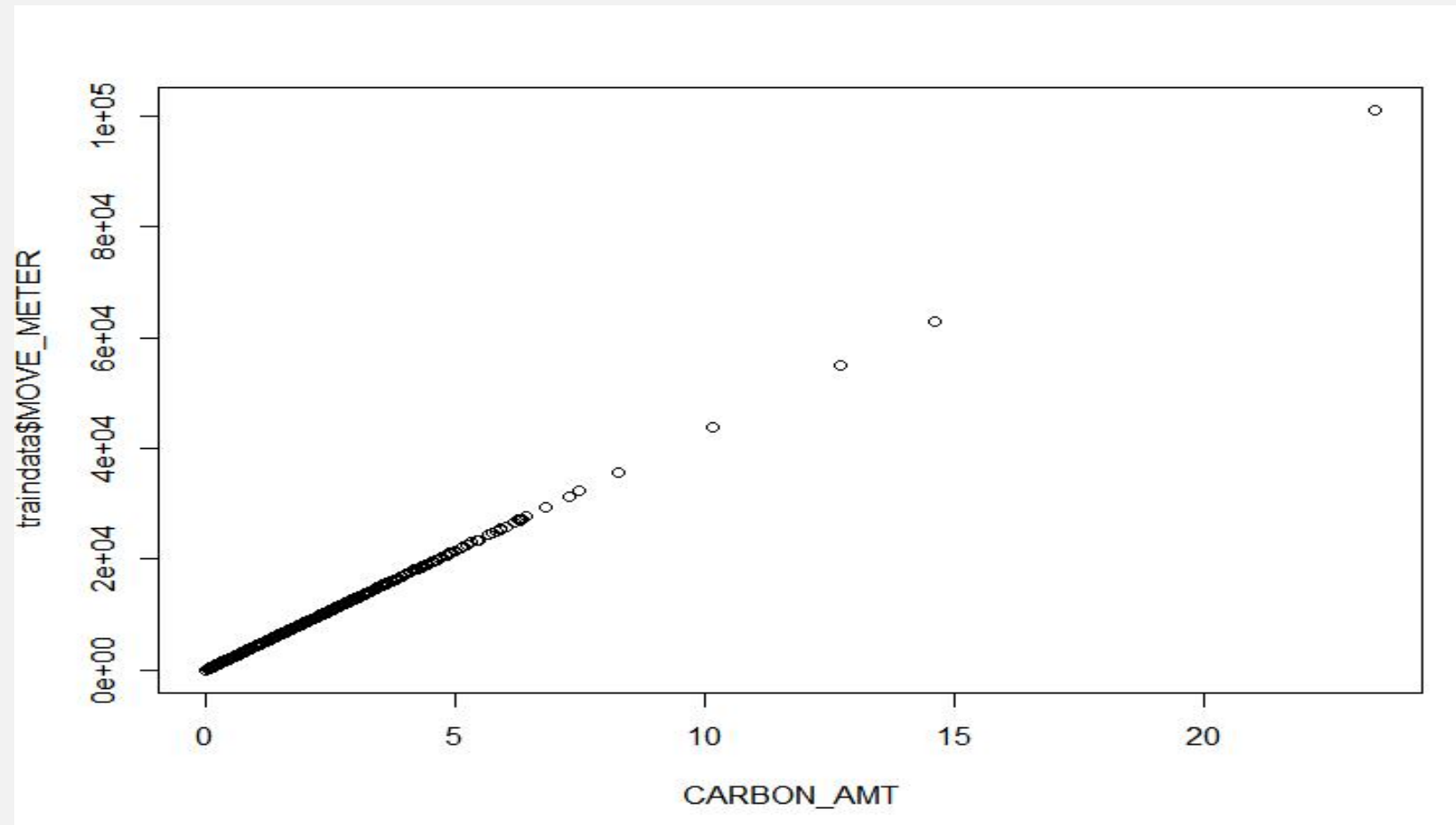
02 구현할 모델

Gaussian Mixture Model(GMM)



$$f(x; \theta) = \sum_{j=1}^k \frac{p_j}{(2\pi)^{p/2} |\Sigma_j|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu_j)' \Sigma_j^{-1} (x - \mu_j) \right)$$

X : CARBON_AMT & MOVE_TIME




Gaussian Mixture Model(GMM)



	K=2	K=3	K=4
Log-likelihood	-10657.99	-10122.29	-9835.712
df	11	17	23
BIC	-21399.58	-20373.8	-19846.24
CLUSTERING	1:1114, 2:886	1:804, 2: 934, 3: 262	1: 596, 2:596, 3:66, 4:742

->According to bic criterion, select the k=2 model

Gaussian Mixture Model(GMM)



```
$pro  
[1] 0.5433096 0.4566904
```

```
$mean  
  
          [,1]      [,2]  
CARBON_AMT 0.3954549 1.744279  
MOVE_TIME  12.2312660 69.841926
```

```
$variance$sigma  
, , 1
```

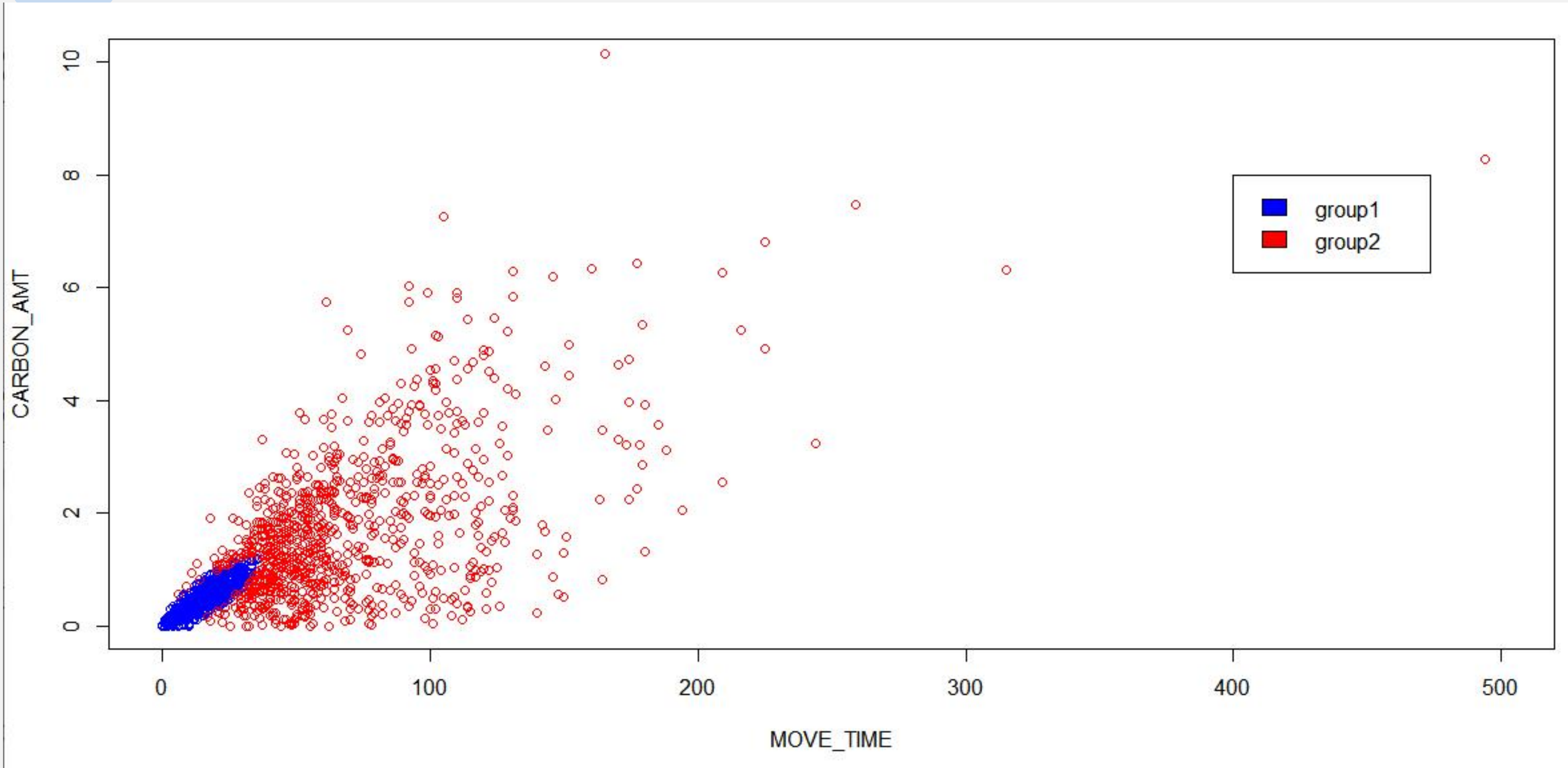
	CARBON_AMT	MOVE_TIME
CARBON_AMT	0.07257567	1.926341
MOVE_TIME	1.92634072	63.096600

```
, , 2
```

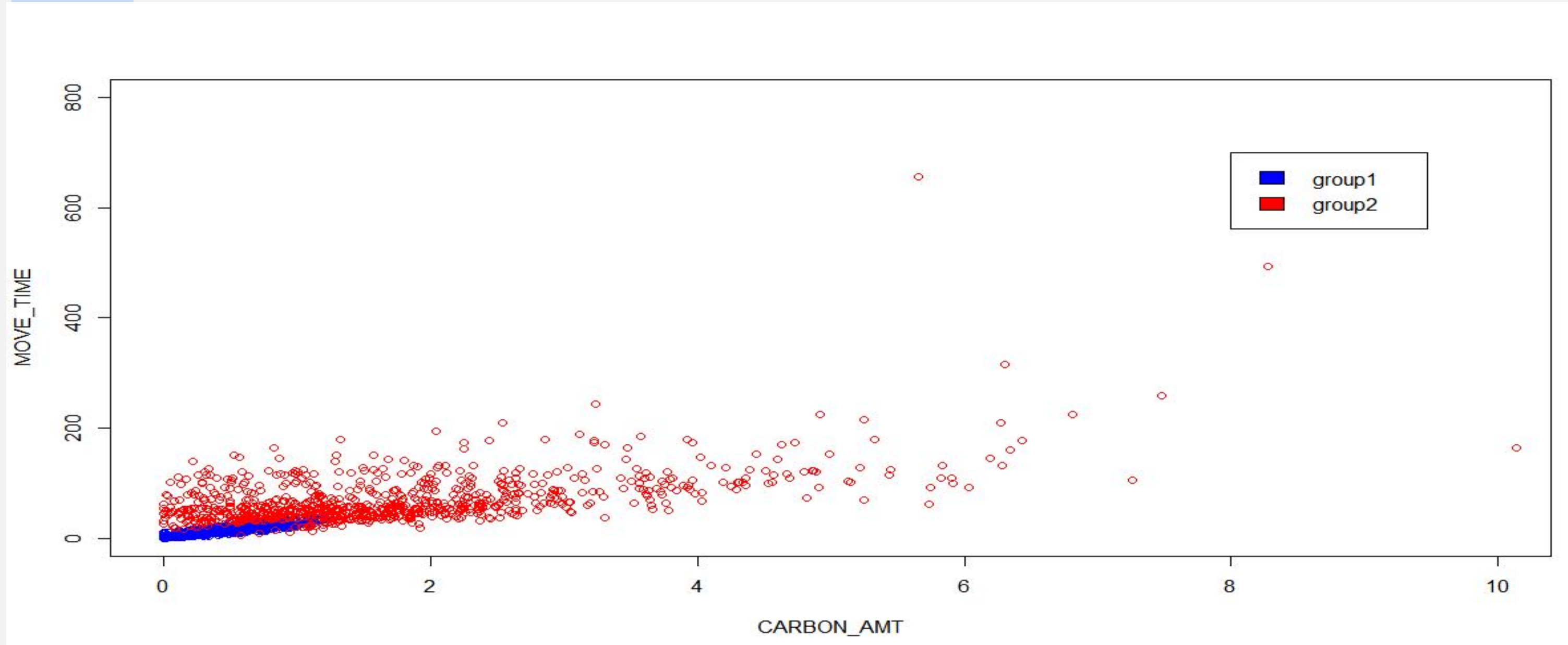
	CARBON_AMT	MOVE_TIME
CARBON_AMT	2.62635	67.18486
MOVE_TIME	67.18486	3507.56188

->EXER_AMT: GROUP2 > GROUP1

Gaussian Mixture Model(GMM)



Gaussian Mixture Model(GMM)



Latent Variable: SPEED, URGENT(PRESSING)

GROUP1: 긴박한 사람, 출근하는 사람 => 단기간에 빠른 이동을 해야하는 사람

GROUP2: 심리적 여유가 있는 사람, 여가활동을 하는 사람 => 상대적으로 천천히 많이 이동하는 사람

02 구현할 모델

Mixture of Experts(MoE)



$$\begin{aligned} p(y | \mathbf{x}) &= \sum_{k=1}^K \pi_k(\mathbf{x}; \boldsymbol{\alpha}) \phi_1(y; \beta_{0k} + \mathbf{x}^\top \boldsymbol{\beta}_k, \sigma_k^2) \\ &= \sum_{k=1}^K \underbrace{\frac{\exp(\alpha_{0k} + \mathbf{x}^\top \boldsymbol{\alpha}_k)}{\sum_{j=1}^K \exp(\alpha_{0j} + \mathbf{x}^\top \boldsymbol{\alpha}_j)}}_{p(z=k|\mathbf{x})} \underbrace{\phi_1(y; \beta_{0k} + \mathbf{x}^\top \boldsymbol{\beta}_k, \sigma_k^2)}_{p(y|\mathbf{x}, z=k)}, \end{aligned}$$

->Y: EXER_AMT

->GATING: AGE_TYPE + RENT_HR + RENT_TYPE + RENT_SPOT + park + holiday + GENDER_CD + USE_CNT

->EXPERT: CARBON_AMT, MOVE_TIME

Mixture of Experts(MoE)



	K=3	K=4	K=5
Log-likelihood	-7618.534	-7081.762	-6966.169
df	46	67	88
BIC	-155586.71	-14672.784	-14601.217
AIC	-15329.068	-14297.524	-14108.338
CLUSTERING	1:582, 2:865, 3: 553	1:573, 2: 698, 3: 260, 4: 469	1: 678, 2: 570, 3: 133, 4: 433, 5: 186

->According to bic criterion, select the k=3 model

Mixture of Experts(MoE)

Coefficients:

```
(Intercept) AGE_TYPE2 AGE_TYPE3 AGE_TYPE4 AGE_TYPE5 RENT_HR2 RENT_HR3 RENT_HR4 RENT_HR5 RENT_HR6
2  1.8231830 -0.4231997 -0.3373855 -0.5493095 -0.1454473 -0.34745963 -0.6641650 -0.4667648 -0.3340197 -0.66676430
3 -0.8814753  0.4404803  0.9031536  1.1422028  1.0327297  0.06997129  0.1549685  0.3602649  0.5662227  0.03333575
  RENT_TYPE1    park1  holiday1 RENT_SPOT2 RENT_SPOT3 GENDER_CD1  USE_CNT1
2  0.1315269 0.1711044 0.07746981 0.434139480  0.3624938 -0.6221183 -0.9029996
3 -0.2672219 0.2912679 0.05328579 0.007724881 -0.4494544  3.2720818 -2.9198973
```



Cluster1 :

Coefficients:

```
(Intercept) CARBON_AMT MOVE_TIME
0.025831221 110.984818549 -0.001192848
```

Cluster2 :

Coefficients:

```
(Intercept) CARBON_AMT MOVE_TIME
4.782200154 92.275657056 -0.002849478
```

Cluster3 :

Coefficients:

```
(Intercept) CARBON_AMT MOVE_TIME
3.57816047 126.76086739 -0.08268033
```

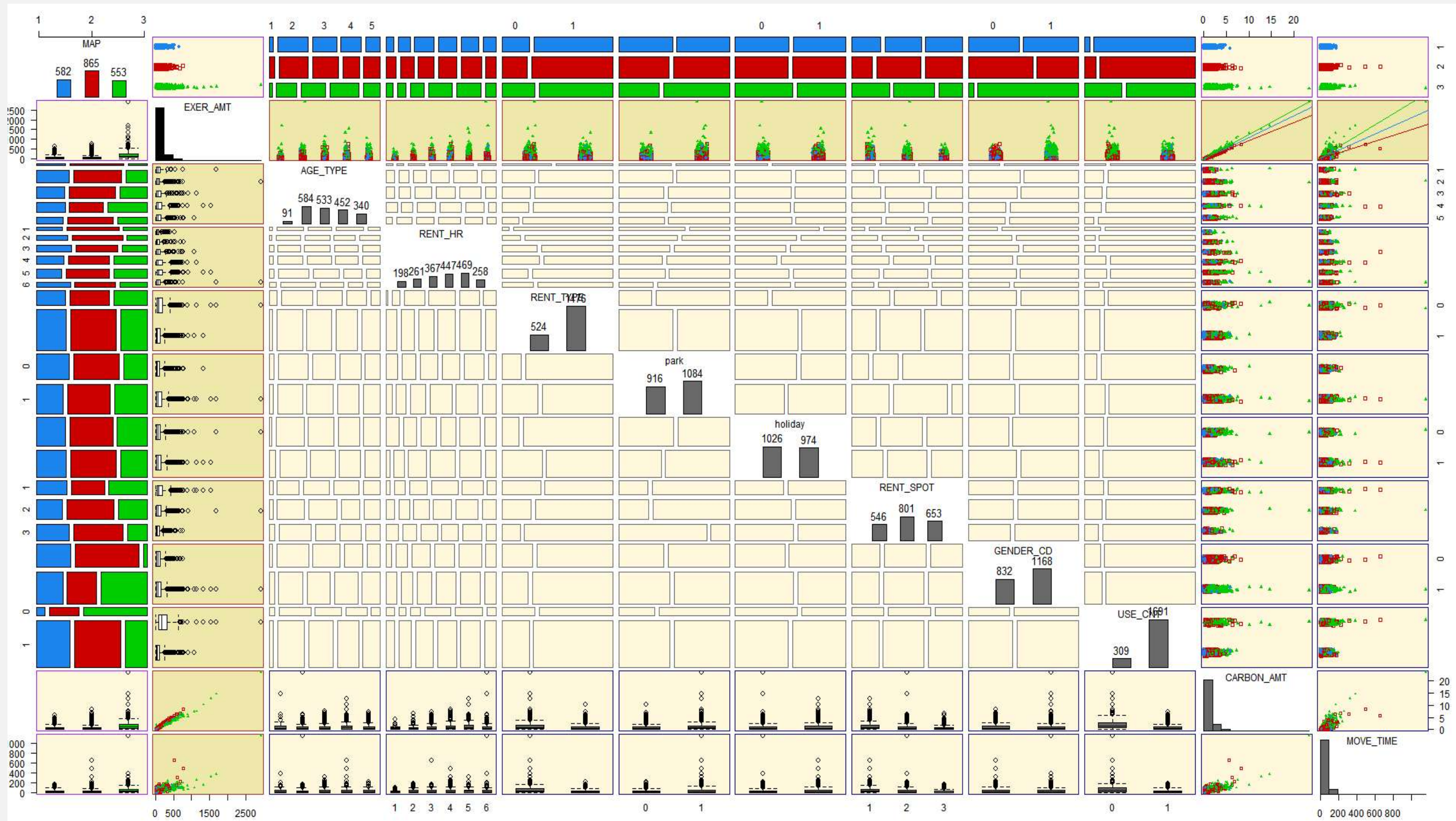
Formula: ~CARBON_AMT + MOVE_TIME

\$mean

	cluster1	cluster2	cluster3
[1,]	83.01738	80.8889	191.2407

->MOST EXER_AMT GROUP: CLUSTER 3

mixture of Experts(mixtools package)



Mixture of Experts(MoE)

	CLUSTER1					CLUSTER2					CLUSTER3				
MEAN(MOVE_TIME)	31.04811					32.12023					56.47197				
MEAN(CARBON_AMT)	0.7330241					0.7811445					1.66472				
AGE_TYPE	10대이하	20대	30대	40대	50대이상	10대이하	20대	30대	40대	50대이상	10대이하	20대	30대	40대	50대이상
	4%	32%	25%	21%	15%	5%	31%	27%	17%	17.5%%	3%	21%	25%	31%	17.7%
GENDER_CD	남		여			남		여			남		여		
	52%		47%			39%		60%			94%		6%		
RENT_SPOT	업무	지하철	주거			업무	지하철	주거			업무	지하철	주거		
	28.3%	35.5%	36.6%			20.4%	42.8%	36.6%			36.8%	40.3%	22.7%		



Mixture of Experts(MoE)



	CLUSTER1						CLUSTER2						CLUSTER3					
RENT_HR	06-09	09-12	12-15	15-18	18-21	21-00	06-09	09-12	12-15	15-18	18-21	21-00	06-09	09-12	12-15	15-18	18-21	21-00
	8.7%	13.7%	21.8%	20.9%	19.9%	14.7%	11.6%	15.0%	17.4%	20.6%	23.1%	12.0%	8.3%	9.2%	16.0%	26.4%	27.6%	12.2%
RENT_TYPE	정기권			그외			정기권			그외			정기권			그외		
	74.2%			25.7%			76.4%			23.5%			69.2%			30.7%		
holiday	O			X			O			X			O			X		
	49.3%			50.6%			50%			49.9%			45.9%			54%		
park	O			X			O			X			O			X		
	49.4%			50.5%			52.3%			47.6%			62%			37.9%		
USE_CNT	1			>1			1			>1			1			>1		
	95.3%			4.6%			89.7%			10.2%			65%			34.9%		



->Latent Variable : RIDING PURPOSE
CLUSTER1: LEISURE, SPARE TIME, FREE TIME
CLUSTER2: COMMUTE
CLUSTER3: EXERCISE

Cluster3

Latent Variable: Willpower(Desire) of riding



Most powerful Willpower(Desire) of riding

Cluster2

Latent Variable: Willpower(Desire) of riding



Least Willpower(Desire) of riding

Cluster1

Latent Variable: Willpower(Desire) of riding



Willpower(Desire) of riding :between Cluster 3 and Cluster 2

Model assessment



	Random Forest	Linear Model	Misture of Experts
RMSE	61.07248	30.85834	27.87566
MAE	28.00866	16.86678	9.811386

Problem

-> lack of meaningful variable for clustering

-> excessive supposition(discovery latent variable)

THANK YOU

