

출퇴근 맞춤버스 다람쥐버스



데이터마이닝 TEAM PROJECT

고명지 박정민 유채연

Contents



이번 정류소는

입니다



ID: 01-001
다람쥐 버스란?



ID: 02-002
DATA 설명서 & EDA



ID: 03-003
MODELING



ID: 04-004
PROPOSITION



다람쥐 버스란?



출퇴근 맞춤버스로 차내 혼잡이 주로 특정 구간과 특정 시간대에 극심하게 발생한다는 점에 착안해, 다람쥐 쳇바퀴 돌 듯 짧은 구간을 반복 운행하는 순환형 셔틀버스이다.

2017년 6월부터 시행

버스 노선	기.종점	대수	거리	배차간격	횟수	운행시간
8761	광흥창역 - 국회의사당역	4	8.5km	9-12분	12	07:00 ~ 09:00 17:30 ~ 19:30
8771	구산중학교 - 녹번역	4	7.7km	10-11분	12	07:00 ~ 09:00
8551	봉천역 - 노량진역	5	12.3km	10-12분	11	07:00 ~ 09:00
8331	마천사거리 - 잠실역	5	12.1km	10-11분	12	07:00 ~ 09:00

다람쥐버스 출퇴근 노선



서울 '다람쥐버스', 최우수정책상 수상

박대로 2018.12.20. 11:15



© 뉴시스 【서울=뉴시스】출·퇴근 맞춤버스(다람쥐버스) 운행 현황. 2018.12.20. (사진=서울시 제공)

【서울=뉴시스】박대로 기자 = 서울시는 국토교통부(국토부)가 주관한 '2018 지속가능 교통도시평가'에서 교통정책 우수사례부문 최우수정책으로 '다람쥐버스'가 선정됐다고 20일 밝혔다.

다람쥐버스(출·퇴근 맞춤버스)는 특정시간대·특정구간에 반복적으로 발생하는 차내 혼잡을 해소하기 위해 다람쥐 쳇바퀴 돌 듯 해당 구간만 반복 운행하는 버스다.

다람쥐버스는 지난해 6월 4개 노선 시범운영 개시 후 시민 만족도 조사결과를 바탕으로 올해 3월부터 7개 노선으로 확대됐다.

다람쥐버스 투입 구간 내 버스정보안내단말기의 혼잡 표출빈도가 평균 39.5% 감소해 기존 노선의 차내 혼잡 현상이 완화됐다고 시는 밝혔다.

다람쥐버스 운행 이후 혼잡이 줄었지만 해당 구간 전체 버스 탑승객은 평균 16.5% 늘어났다고 시는 설명했다. 기존 노선 혼잡도가 완화되자 이용객이 늘어난 것으로 풀이된다.

이용시민 설문조사 결과 '기존 노선의 혼잡도 개선에 도움이 됐다(96.2%)', '타 지역 노선 신설에 찬성한다(95.8%)' 등 이용 경험이 있는 **시민 다수가 다람쥐버스에 긍정적으로 평가했다.**

하루 평균 이용자 수는 지난해 6월 1692명에서 올 11월 4823명으로 늘었다. 지난달 말 기준 누적 이용객은 123만명을 기록했다.



전체 대중교통
이용자 수

기존 노선 버스
이용객수

평균적으로
39.5% down

16.5% up

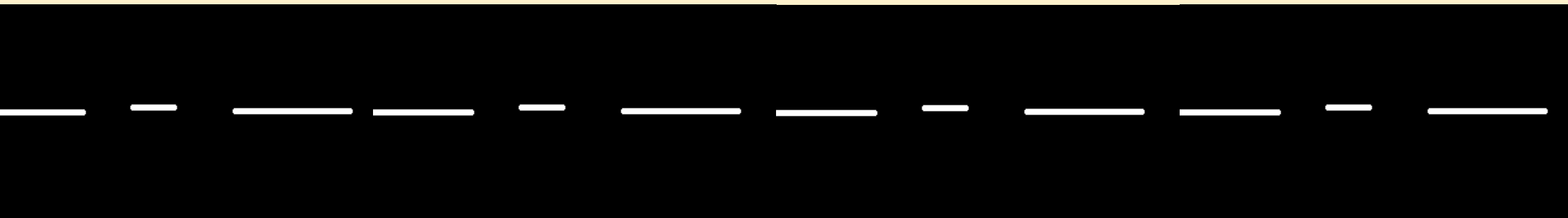
다람쥐버스
92.1점



일반버스
81점

분석목표

다람쥐 버스의 순승차인원을 예측하여 버스 노선 별 혼잡도를 계산한 후,
새로운 다람쥐 버스 노선을 제안



이번 정류소는

입니다



ID: 01-001
다람쥐 버스란?



ID: 03-003
MODELING



ID: 04-004
PROPOSITION

ID: 02-002
DATA 설명서 & EDA



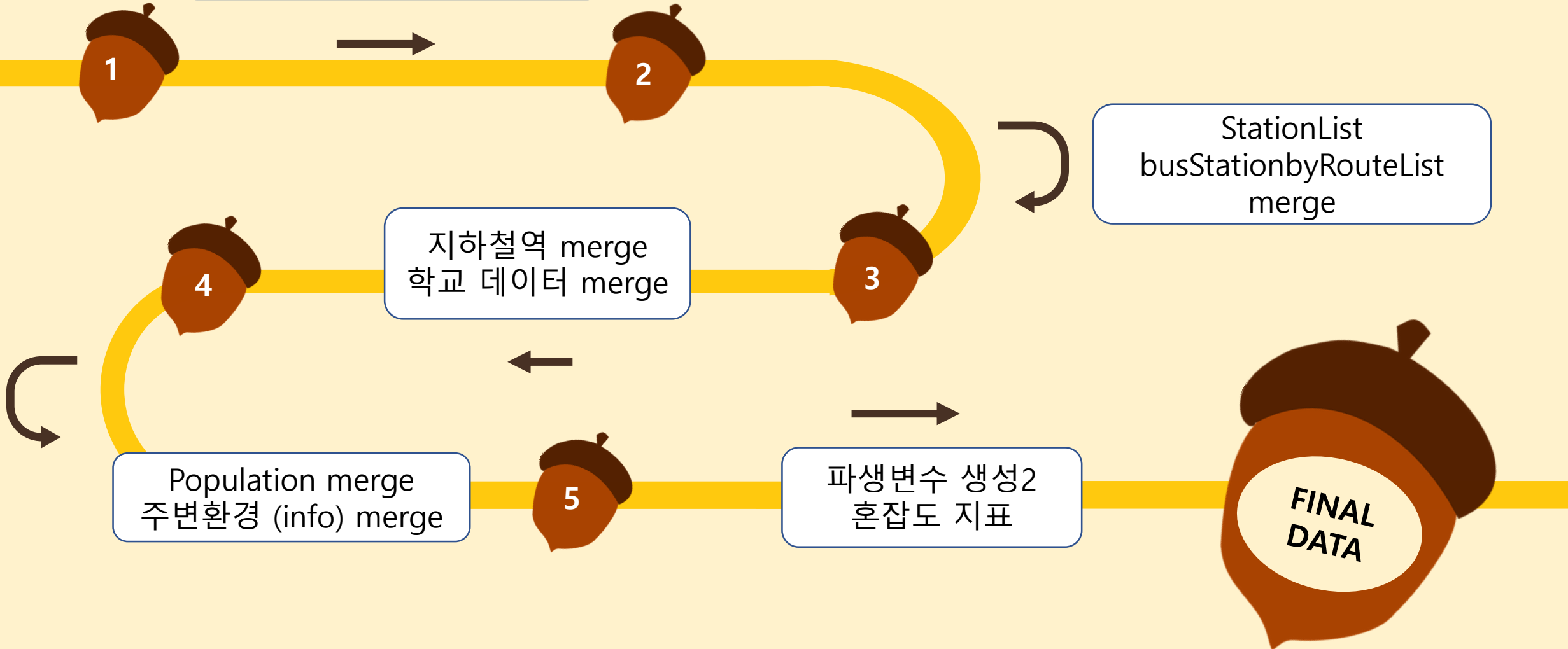
데이터 INTRODUCTION

Data Split	사용 연도
Train	201707 – 201712
Test	201801 - 201803

데이터명	Source	비고
서울시 버스노선별 정류장별 시간대별 승하차 인원 정보	공공데이터포털	기본 데이터
stationList	TOPIS	정류장별 위경도 자료 사용
busStationbyRouteList	공공데이터 API	노선 및 정류장 별 정보 이용
서울시 사업체 및 종사자 밀도(동별) 통계	서울열린데이터광장	정류장 주변 인구 밀도 정보
구/동 별 주거 지역, 녹지 지역, 상업 지역 (info)	서울열린데이터광장	정류장 주변 환경 정보
지하철 역 별 승하차 인원수	공공데이터포털	가까운 지하철역 승하차 수 및 지하철역의 개수 사용
학교 정보 – 중학교, 고등학교, 대학교, 대학원	공공데이터포털	등하교 학생 수

데이터 핸들링

파생변수 생성1
출근승.하차인원(07시 - 10시)
퇴근승.하차인원(17시 - 21시)
FOR BUS & SUBWAY



데이터 핸들링

병합

busStationbyRouteList & StationList by arslid

지하철역: 위.경도 가장 가까운 지하철역으로 merge

Population 데이터 merge by 구.동

학교: 위.경도 가장 가까운 학교의 학생 수 merge

주변환경(info): 위.경도 반경 500m 이내의 지구 수 merge

파생변수1

버스 출근 및 퇴근 순승하차 인원

지하철 출근 및 퇴근 순승하차 인원

파생변수2

혼잡도 지표

결측 및 필터링

현재 이용되지 않는 정류소 및 버스노선 제거

서울 이외 지역 제거 (Population 데이터가 붙지 않은 행 제거)

혼잡도 지표

morn_full =

출근 승차 인원 - 출근 하차 인원

even_full =

퇴근 승차 인원 - 퇴근 하차 인원

혼잡도지표 =

버스 별 누적 탑승인원 / 60 * 100

Variable	Class	데이터 설명서
busRouteNm	Factor	버스노선번호
arsId	Integer	정류장번호
lng, lat	Numeric	정류장의 위, 경도
morn_ride, morn_off	Numeric	출근 승하차 인원 평균
even_ride, even_off	Numeric	퇴근 승하차 인원 평균
stStationNm, edStationNm	Factor	기점 정류소 이름, 종점 정류소 이름
term	Integer	배차 간격
direction	Factor	방향
transYn	Factor	회차 여부
fullSectDist	Integer	정류소 간 거리
busnum	Numeric	시간 당 평균 운행 대수
seq	Integer	노선 순서
morn_full	Numeric	출근 승차인원 - 출근 하차 인원
even_full	Numeric	퇴근 승차인원 - 퇴근 하차 인원
cnt	Integer	정류소 당 지나가는 버스노선 개수
length	Numeric	전체 경로 거리
cong_morn	Numeric	출근 혼잡 지표
cong_even	Numeric	퇴근 혼잡 지표

Variable	Class	데이터 설명서
gu, dong	Factor	정류소가 위치하는 구, 동
사업체수. 밀도	Numeric	정류소가 위치하는 구, 동의 사업체 수 및 밀도
종사자수. 밀도	Numeric	정류소가 위치하는 구, 동의 종업자 수 및 밀도
인구수. 밀도	Numeric	정류소가 위치하는 구, 동의 인구수 수 및 밀도
subway	Factor	정류소에서 가장 가까운 지하철역
n_subway	Integer	500m 이내 지하철역 수
mean.smr	Numeric	지하철 출근
mean_smo	Numeric	지하철 출근 하차 인원 수
mean_ser	Numeric	지하철 퇴근 승차 인원 수
mean_seo	numeric	지하철 퇴근 하차 인원 수
info 1	Integer	녹지 지역 count
info 2	Integer	상업 지역 count
info 3	Integer	주거 지역 count



EDA

이용객수

기존 버스
vs
다람쥐버스
월별 이용 수
비교

승하차

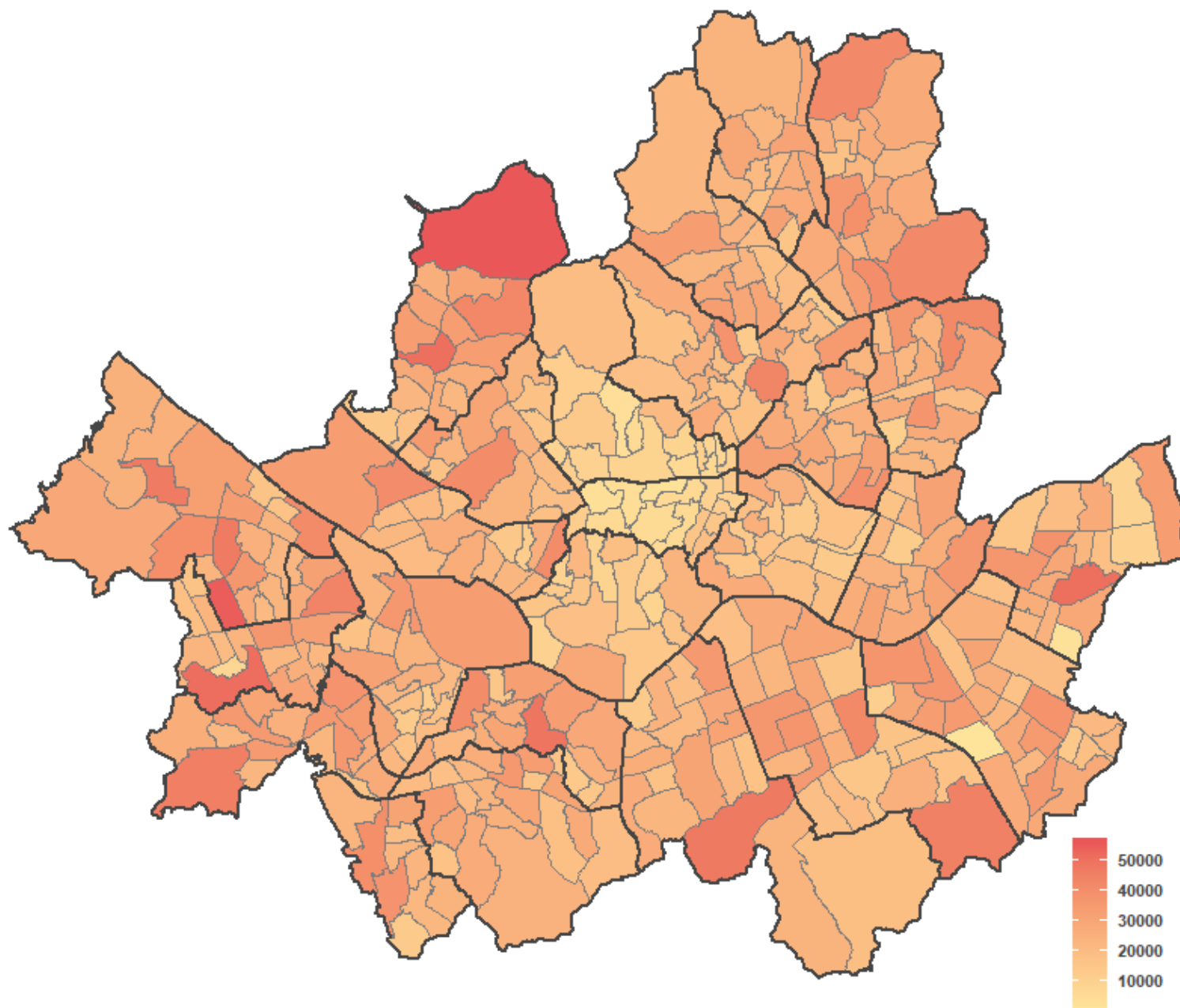
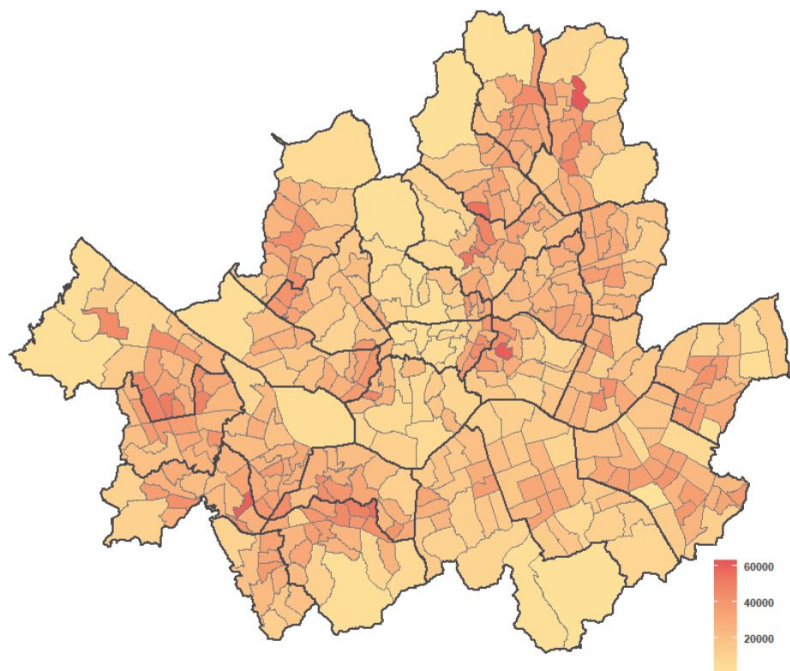
기존 버스
정류소별
승하차 인원
비교

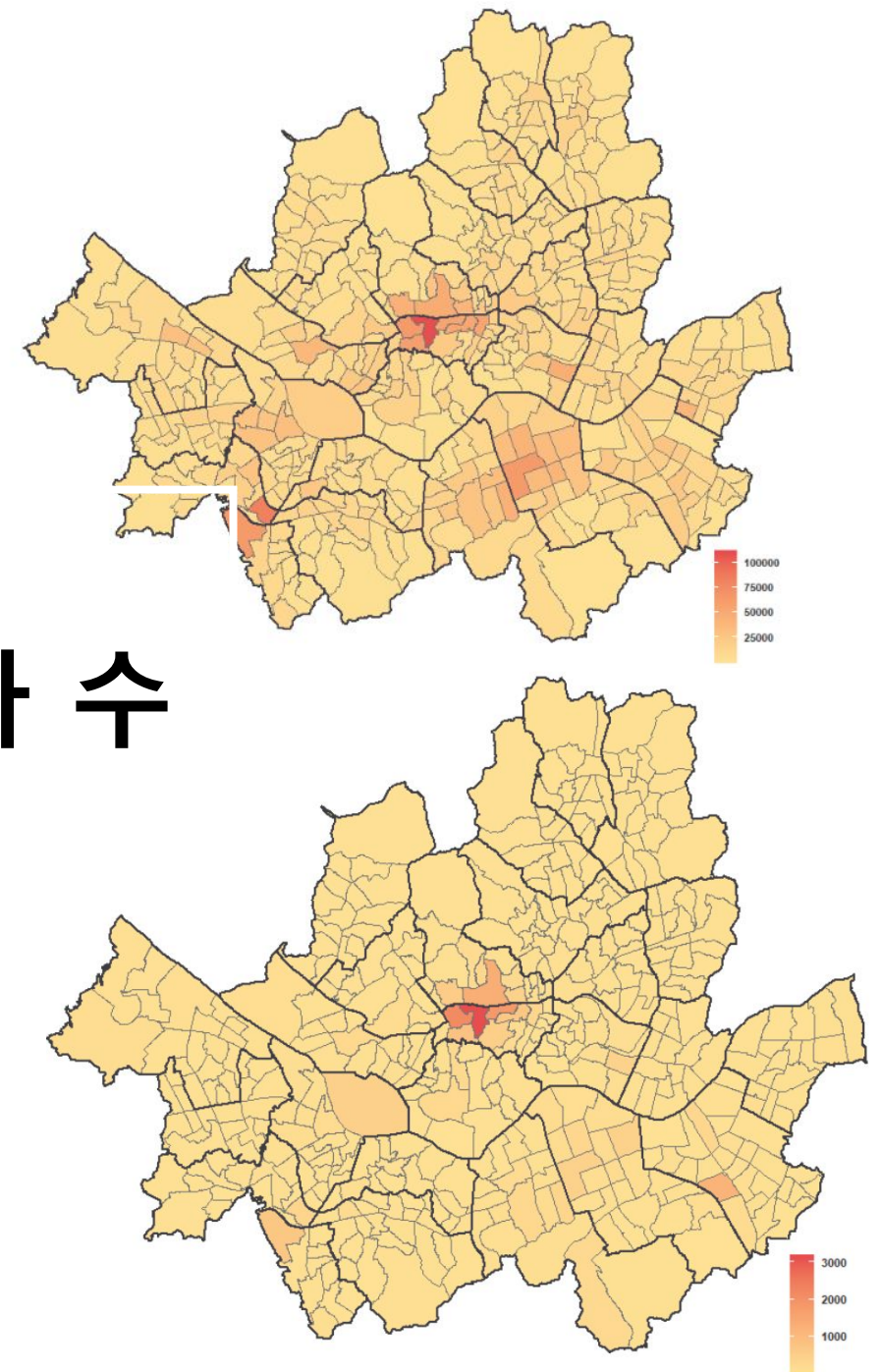
동별 통계

동별 잠재
버스 이용
인원

서울시 동 별 Total Population

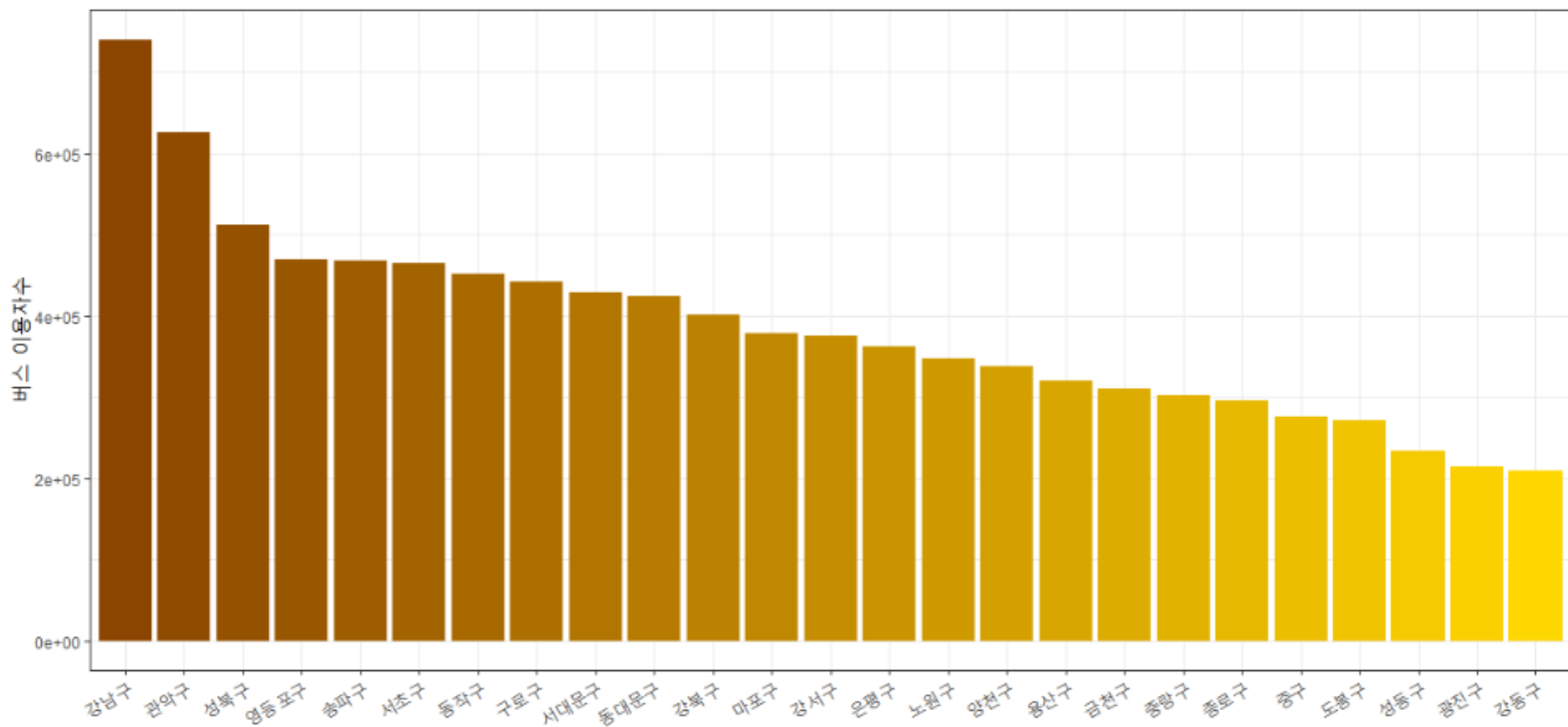
서울시 동별 인구밀도-B.

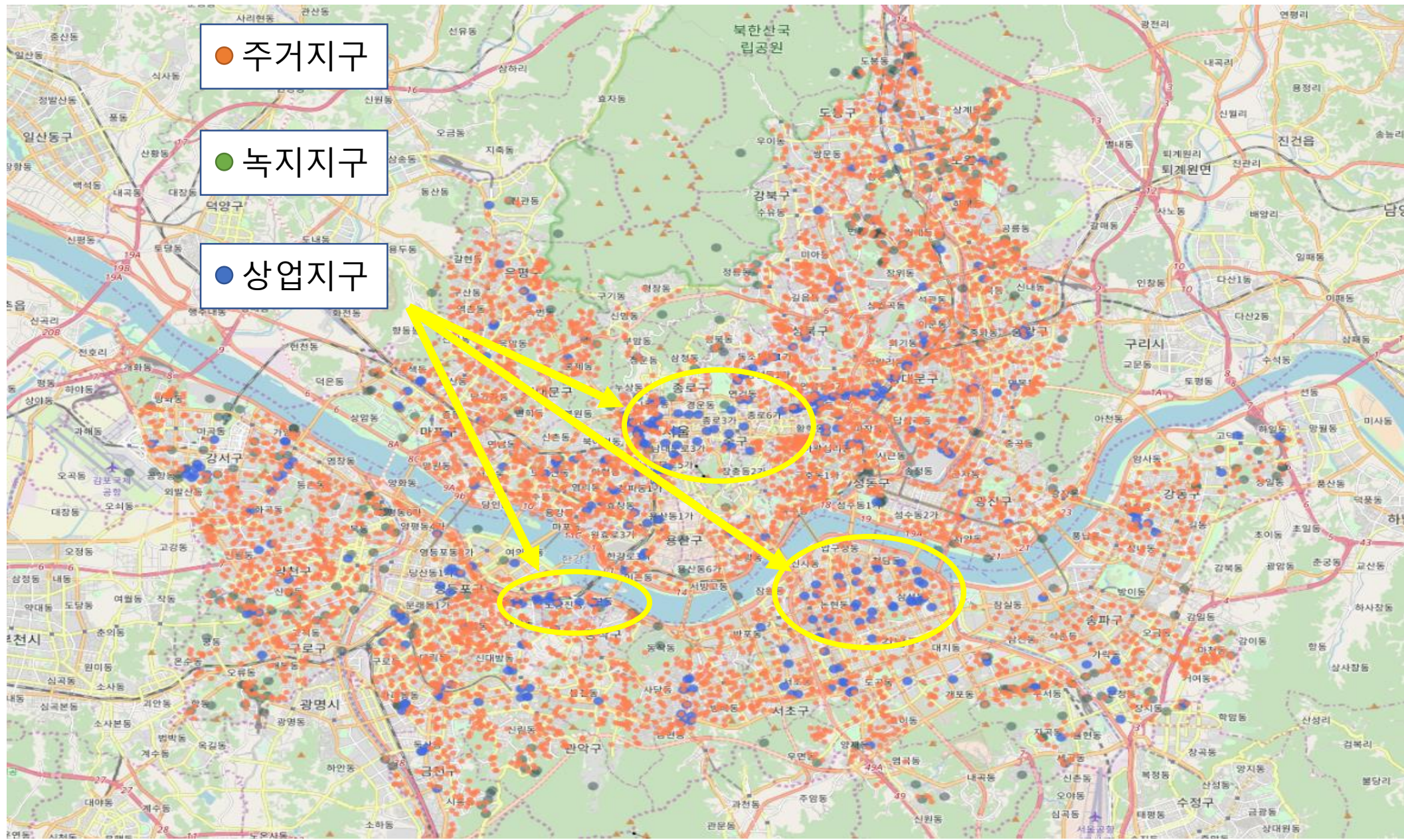




구별 버스 이용자 수

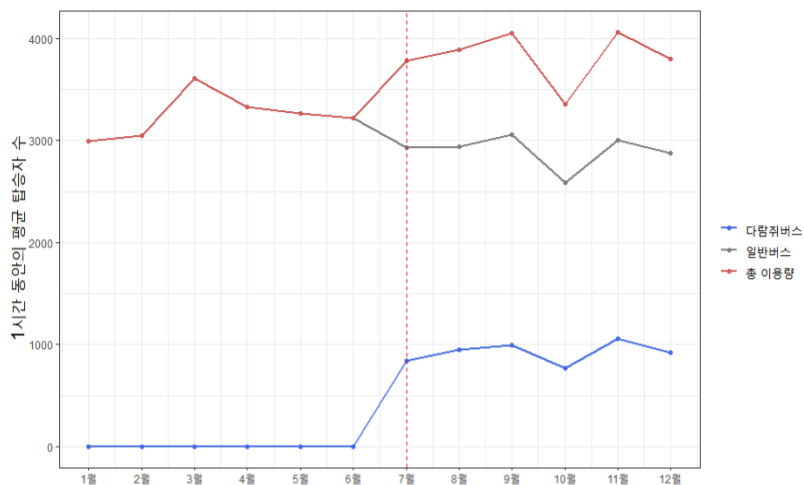
구별 하루동안의 버스 이용자 수





500

8551



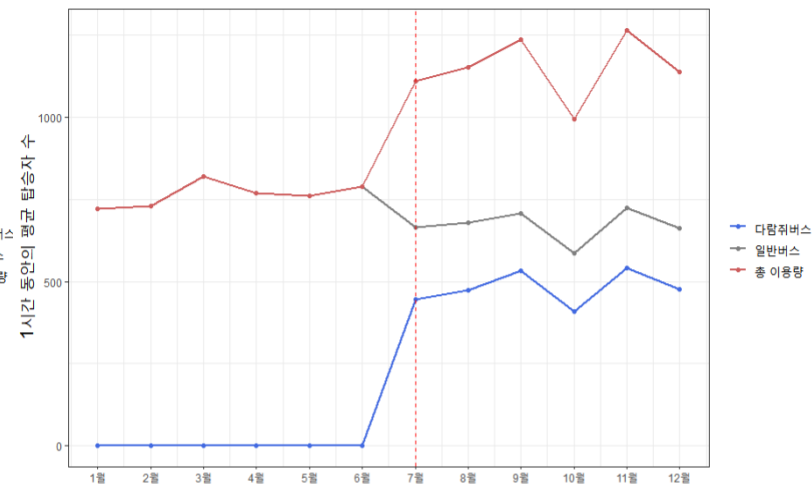
5535

8551



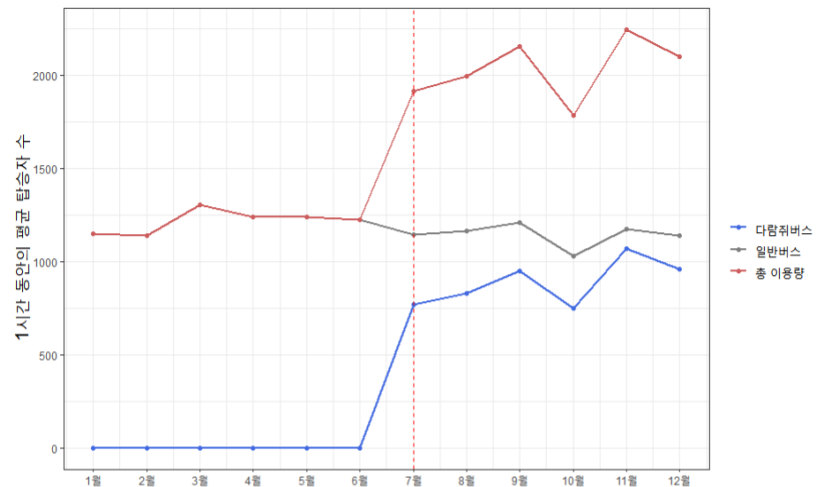
153

8761



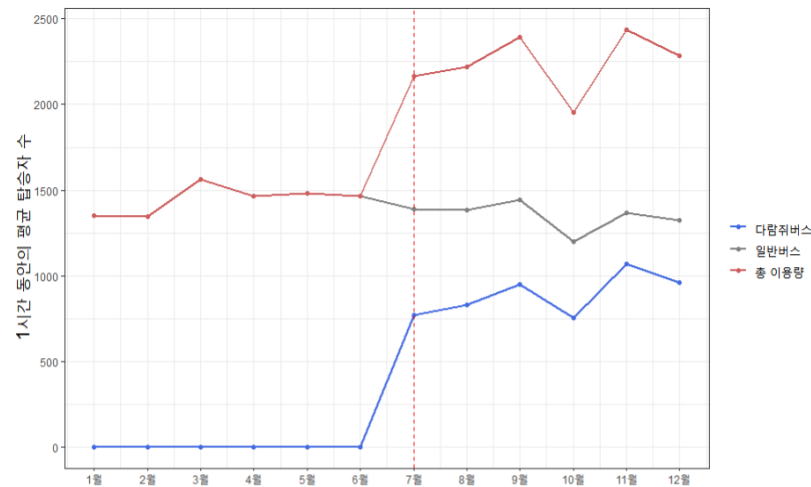
702A

8771



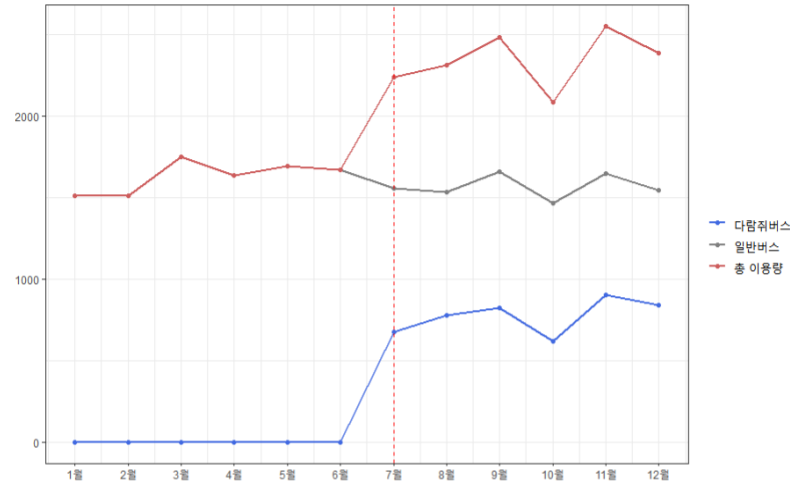
702B

8771



3315

8331



[illegible]

중학교	경영학과	교육학과
노랑중학교	10	10
노랑중학교	100	80
동국대학교노랑진조동중학교	20	20
동국대학교노랑진조동중학교	10	10
장성고등학교	20	20
장성고등학교	10	10
7호선 장승배기역 상도2지간 중학교	100	80
7호선 장승배기역 상도2지간 중학교	100	80
상도2동아이마루 포스코대산상도중학교	10	10
상도2동아이마루 포스코대산상도중학교	10	10
신상도중학교	100	80
신상도중학교	100	80
상도4동역수원선	200	180
경향엑스플렉스동대구청	100	80
경향엑스플렉스동대구청	100	80
상도8동역수원선	20	20
상도8동역수원선	20	20
구암중학교	100	80
은천동혁신마을 중앙아파트 105동	10	10
현대시장	100	80
현대시장	100	80

버스정류장	공공교통	민간버스
광종장역	100	100
서강동주민센터	400	350
공종장역	100	100
서강동주민센터	100	100
창천현대플라워	100	100
창천현대플라워	100	100
신촌로터리	100	100
국회의사당	100	100
국회의사당역	200	200
국민은행	100	100
국회의사당역	100	100
국회의사당	400	400

버스정류장	초방문자 (First-time visitors)	반복방문자 (Repeat visitors)
중앙역신사오거리	140	120
신진교대기동교교도이마트은행	140	140
서부병원	90	90
서대문내무서출입	70	70
은행구청	40	40
구산중학교구산교회	150	140
시립서부병원세원교회앞	190	170
서울기독대학교입구	150	140
유림교동	140	130
신사동교도은행지하차량출입	80	90
녹번역	40	40

버스정류장	남성 (Male)	여성 (Female)
양양역산시오거리	150	-150
신진과학기술고교	150	-150
이마로종합	160	-10
사북병원	110	-50
서대문세무서본관	90	-50
운봉구청	50	-150
구산중학교	160	-10
구산고교	220	-20
시립사북병원세원고	180	-10
서울기독교대학고급부	170	-10
우남고등학교	170	-10
신사중고교	100	-100
은평지역자활센터	100	-100
녹번역	50	-250

버스종류장	운행 전 (분)	운행 후 (분)
상설의원비둘기	70	-85
방이동사거리	50	-20
송파구청KT종합지사	40	-45
방이우체국	70	-10
한양삼의아파트	100	-20
오금동현대아파트	70	-45
오금공립앞	10	-25
오금동상아아파트	50	-15
보인고교 오금동삼성아파트	50	-35
송파사회복지관	70	-20
매천동사거리	60	50
매천2동운전별모신동아아파트	110	105
서울방이동교본초	40	-15

이번 정류소는

입니다



ID: 01-001
다람쥐 버스란?

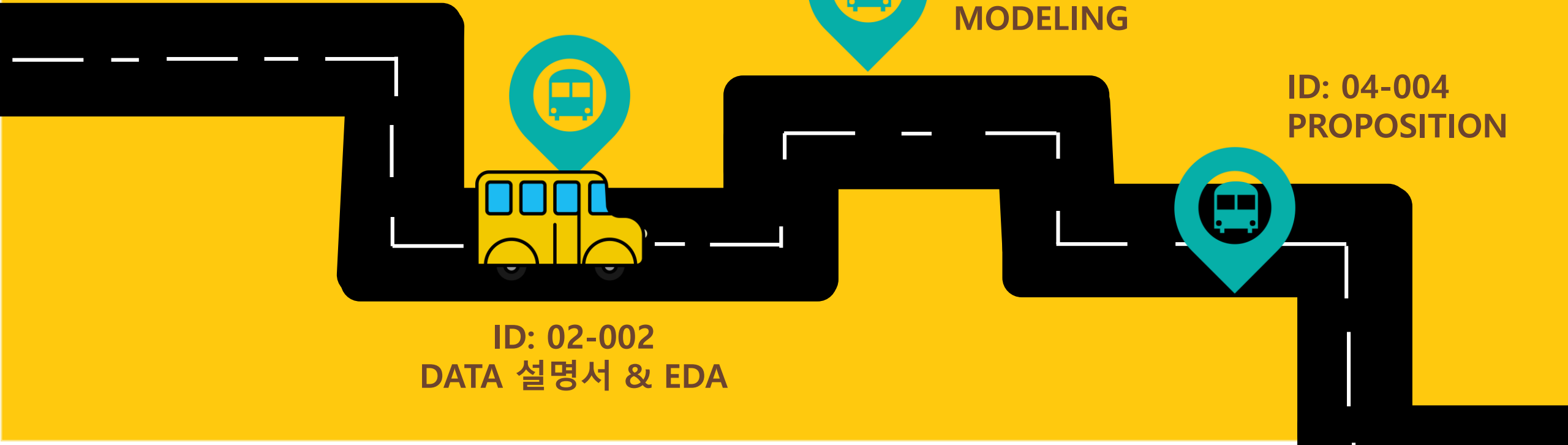


ID: 03-003
MODELING

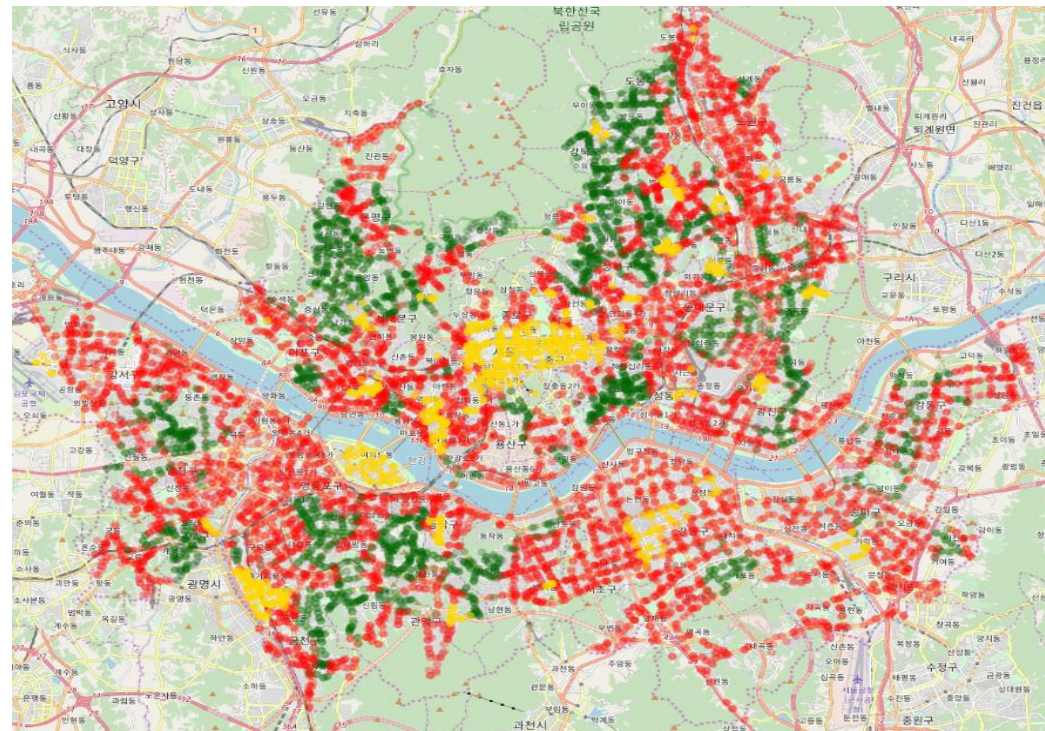
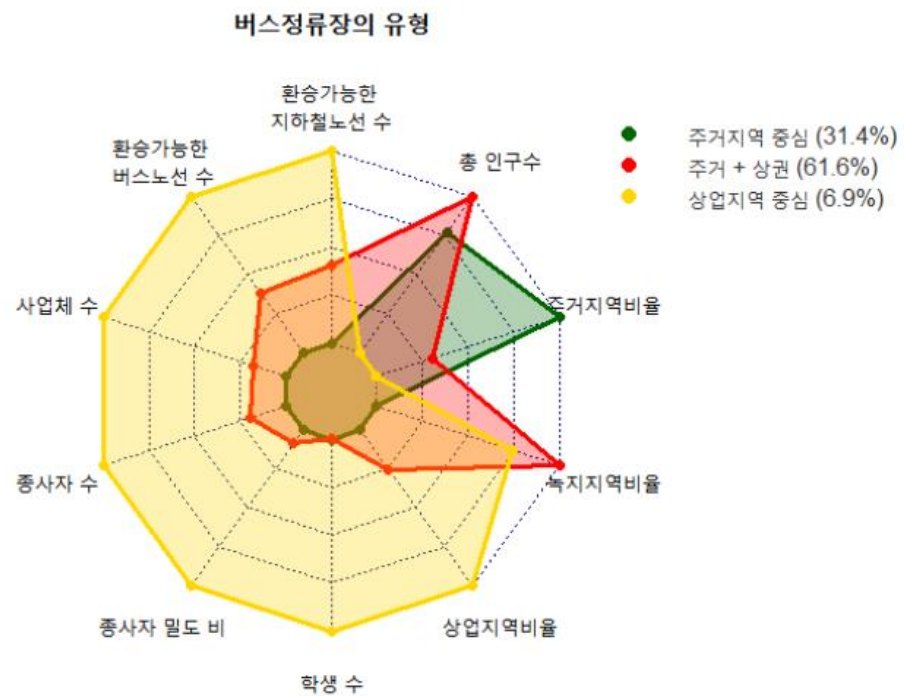


ID: 02-002
DATA 설명서 & EDA

ID: 04-004
PROPOSITION



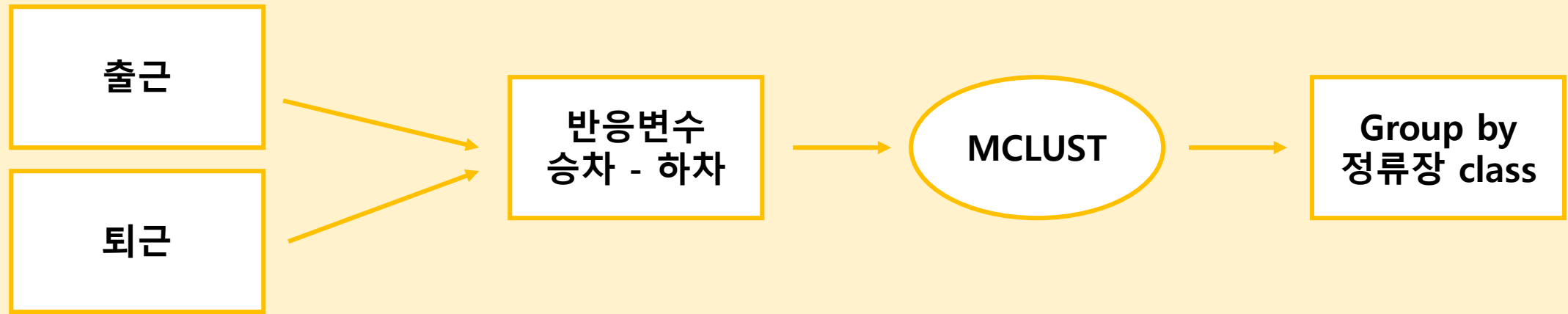
MCLUST



class	n_student	사업체수	종사자수	totpop	n_subway	cnt	종사자밀도비	info1	info2	info3
1	870	1283	4674	25722	1	2.60	18.3	0	0.437	15.2
2	825	2242	15542	27629	1.25	3.14	61.7	0.8	0.917	13.1
3	6426	6573	59899	19178	1.60	4.03	532	0.595	2.31	12.1

2018 신규 정류장 class

arsId	n_student	사업체수	종사자수	totpop	n_subway	cnt	종사자밀도비	info1	info2	Info3	Class
5231	656	783	5705	11889	1	2	48.19	0	1	16	1
9277	850	1962	8026	31274	1	2	26.50	0	2	31	1
12821	1013	1844	4934	26632	1	1	18.52	0	1	18	1
23387	270	448	8210	24927	1	2	32.80	2	0	16	1
5187	759	1711	8543	24738	1	3	34.94	0	1	13	2
8348	1098	921	4242	27347	1	2	15.47	3	0	11	2
10292	917	1277	7384	28694	1	1	25.56	2	0	7	2
16294	942	2657	23388	33991	1	2	68.38	2	0	15	2
18993	667	4085	24308	395456	1	1	61.66	0	0	4	2
23285	811	15590	146000	36499	1	4	399.64	0	3	8	3
37501	896	5271	49006	23876	1	2	205.56	1	0	4	3



Linear Regression

lm(y~1)

기본	1	2	3
Morn	68.18492	71.42654	64.87197
Even	55.35715	57.0989	51.55706

	model	rmse	rmse_test
1	morn_class1	68.52381	65.5347
2	morn_class2	72.60339	69.59333
3	morn_class3	63.29077	61.56741
4	even_class1	54.91193	53.41604
5	even_class2	56.70446	55.44192
6	even_class3	53.48011	49.96264

Used Variables for Linear Regression

lng	lat	gu	totpop
사업체수	종사자수	평균종사자	사업체밀도
종사자밀도.A.	인구밀도.B.	종사자밀도비.A.B.100.	n_subway
mean.smr	mean.smo	n_student	info1
info2	info3	length	routeType
term	seq	transYn	fullSectDist
busnum	morn_full	cnt	

Ridge		model	rmse	rmse_test	Lasso		model	rmse	rmse_test
	1	morn_class1	66.19194	63.11304		1	morn_class1	67.8731	64.80886
	2	morn_class2	69.609	66.94496		2	morn_class2	69.5668	66.91397
	3	morn_class3	58.46688	56.94269		3	morn_class3	57.24589	55.80033
	4	even_class1	54.09342	52.51536		4	even_class1	54.39652	52.84271
	5	even_class2	55.49056	54.16375		5	even_class2	55.8341	54.54094
	6	even_class3	52.34504	48.48037		6	even_class3	51.50364	47.41875

Used Variables for LASSO & RIDGE			
lng	lat	gu	totpop
사업체수	종사자수	평균종사자	사업체 밀도
종사자 밀도.A.	인구 밀도.B.	종사자 밀도비.A.B.100.	n_subway
mean.smr	mean.smo	n_student	info1
info2	info3	length	routeType
term	seq	transYn	fullSectDist
busnum	morn_full	cnt	

RandomForest

	model	mtry	ntree	rmse	rmse_test
1	morn_class1	15	200	57.1346	26.72477
2	morn_class2	15	200	61.21784	28.54134
3	morn_class3	15	200	52.78453	26.91519
4	even_class1	15	100	45.94371	22.19747
5	even_class2	15	100	48.32353	23.66954
6	even_class3	15	200	44.31618	23.47896

	mtry	ntree
1	5	50
2	10	50
3	15	50
4	5	100
5	10	100
6	15	100
7	5	200
8	10	200
9	15	200

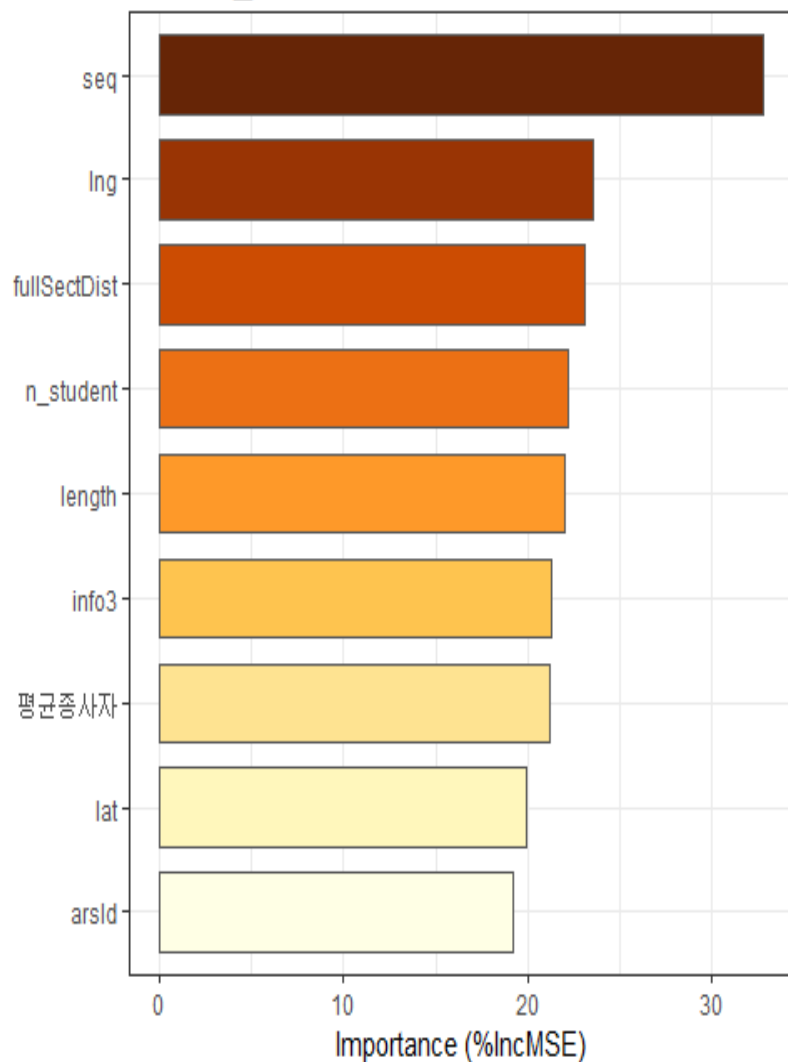
Bagging

	model	ntree	rmse	rmse_test
1	morn_class1	200	57.19326	28.01503
2	morn_class2	200	61.08975	29.9069
3	morn_class3	100	54.91622	28.40684
4	even_class1	200	45.54813	23.87543
5	even_class2	200	50.07567	25.69611
6	even_class3	100	47.19759	25.3292

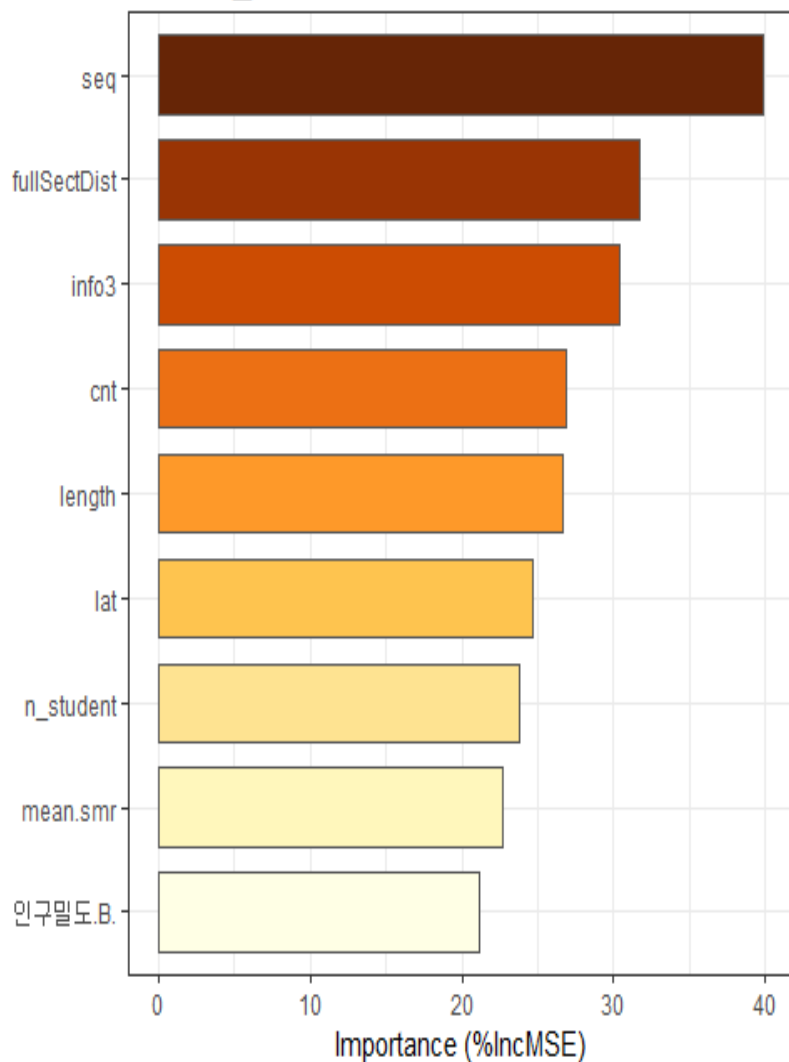
	ntree
1	50
2	100
3	200

RF Morning Variable Importance

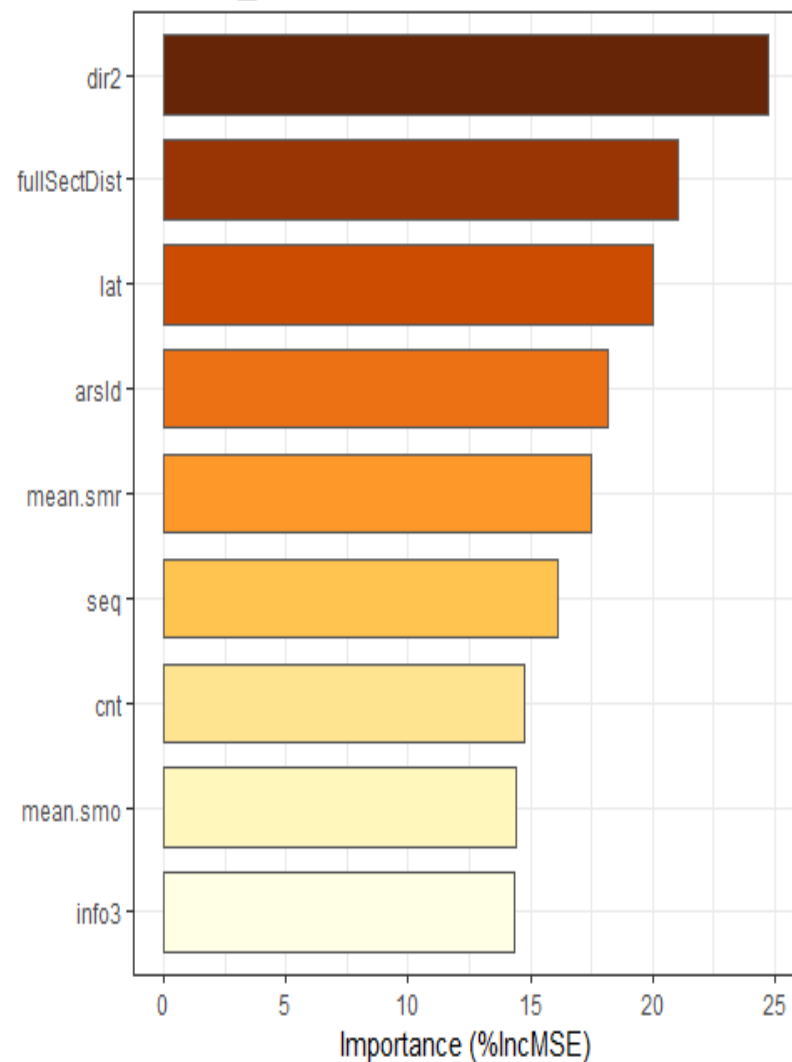
Class1_출근 순 승차수



Class2_출근 순 승차수

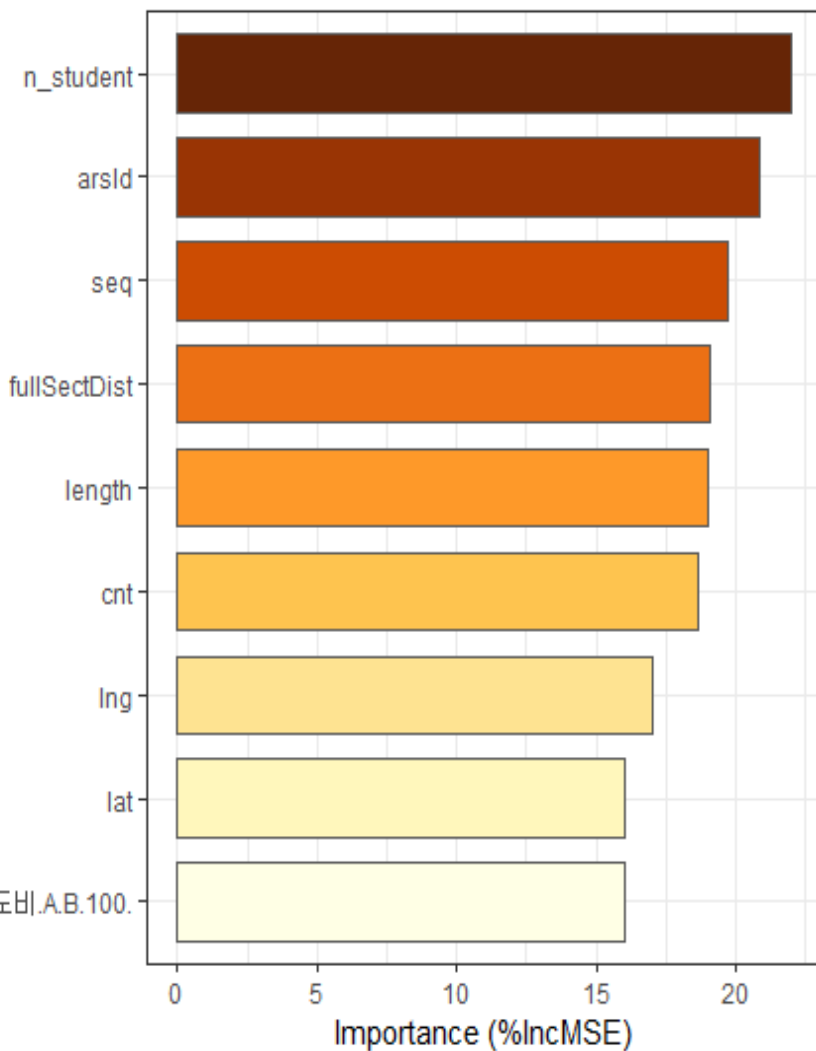


Class3_출근 순 승차수

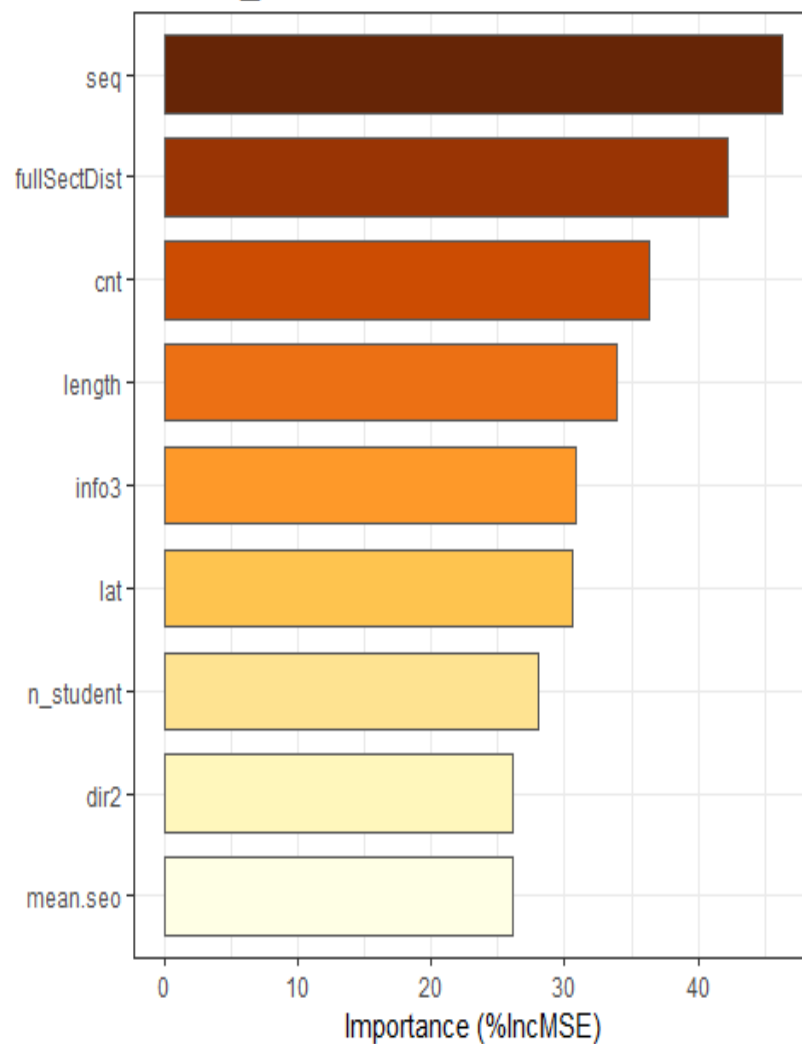


RF Evening Variable Importance

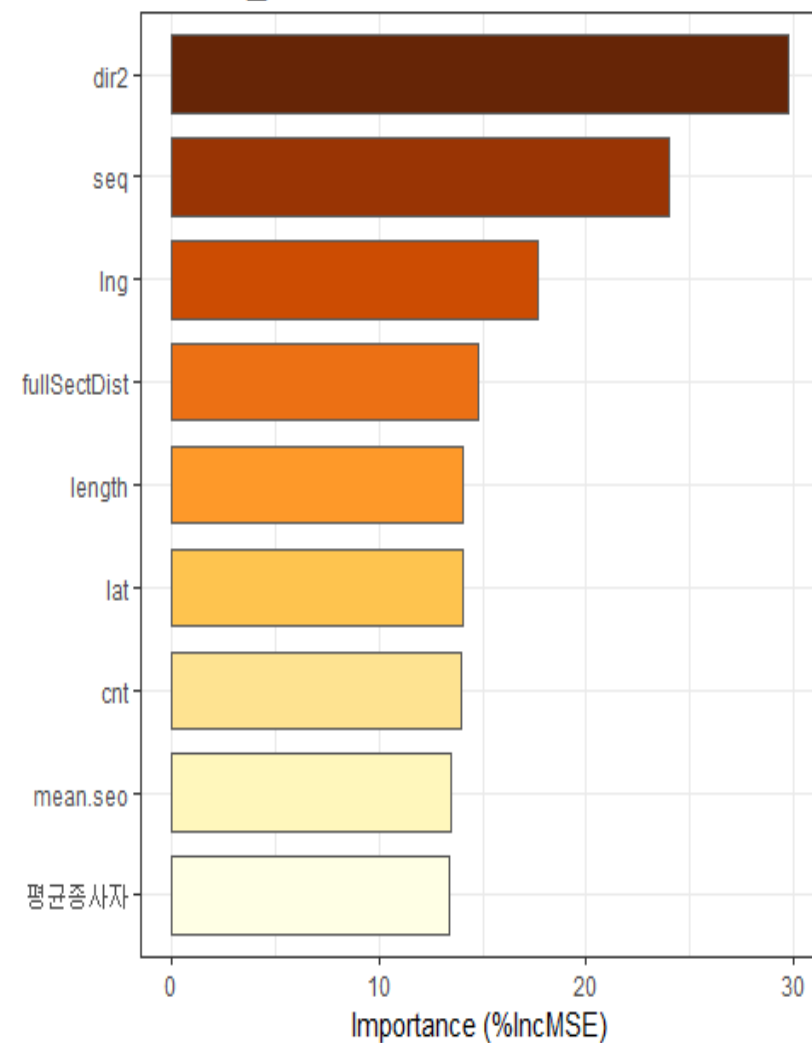
Class1_퇴근 순 승차수



Class2_퇴근 순 승차수

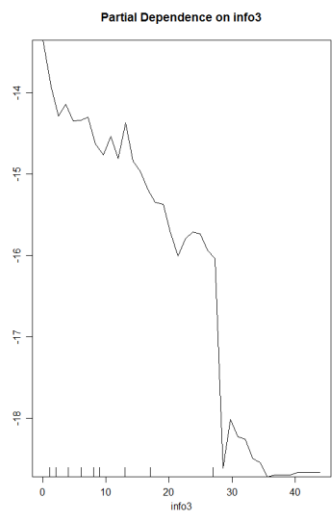


Class3_퇴근 순 승차수

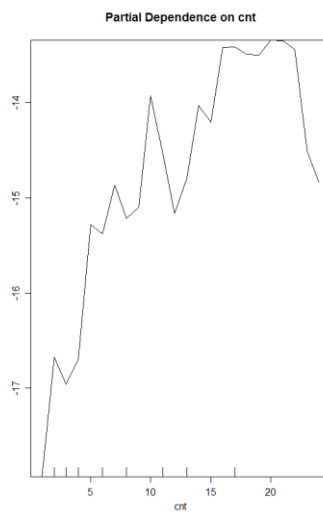


Morning

Class = 1

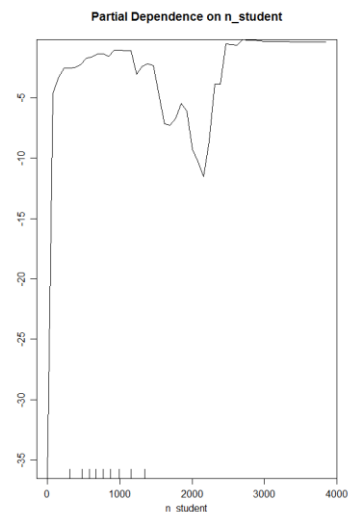


info3

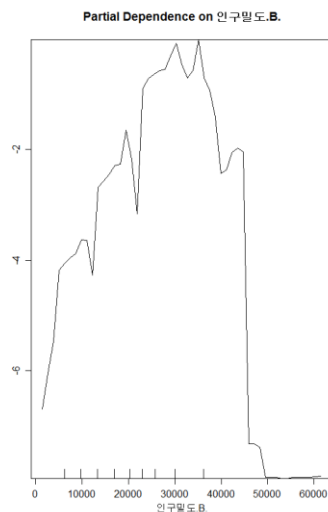


cnt

Class = 2

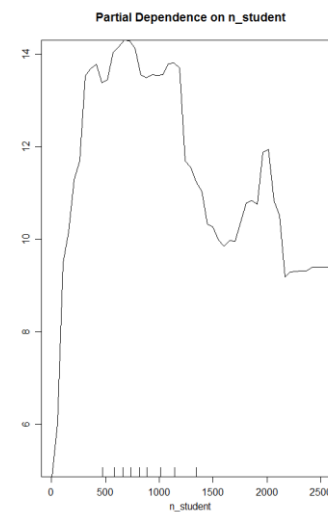


N_student

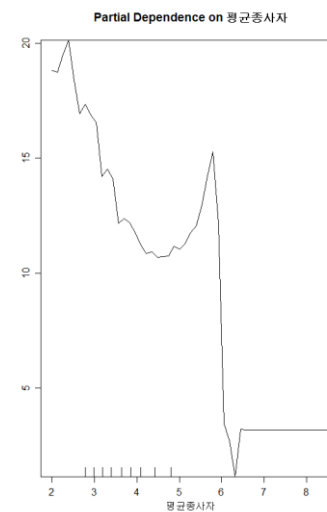


인구밀도.B.

Class = 3



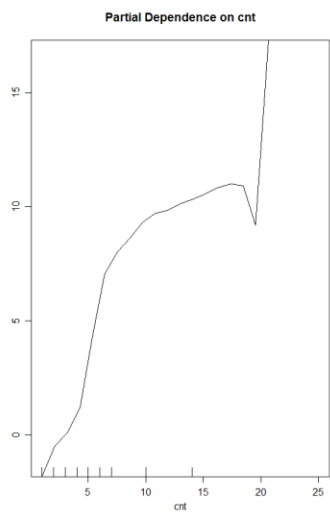
n_student



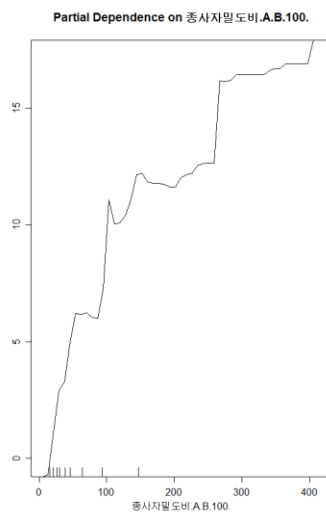
평균종사자

Evening

Class = 1

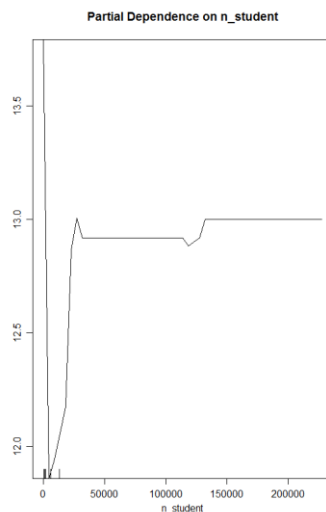


cnt

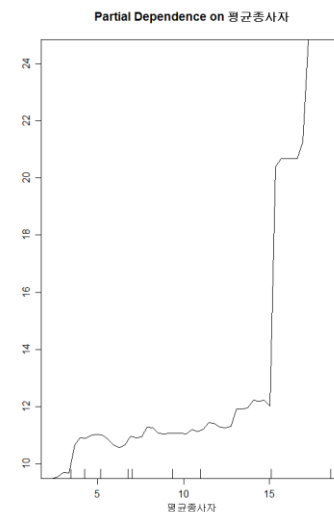


종사자밀도

Class = 2

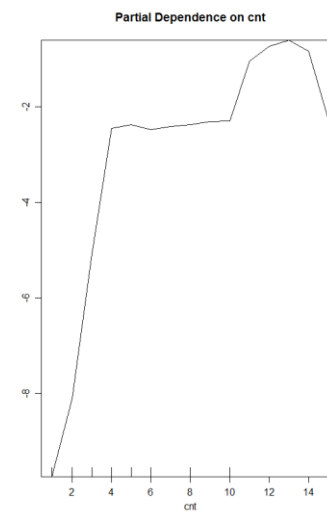


N_student

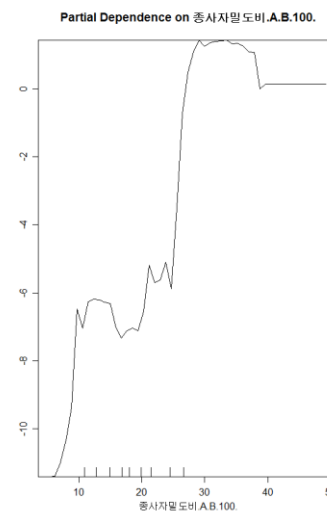


평균종사자

Class = 3



cnt



종사자밀도

XGBoost

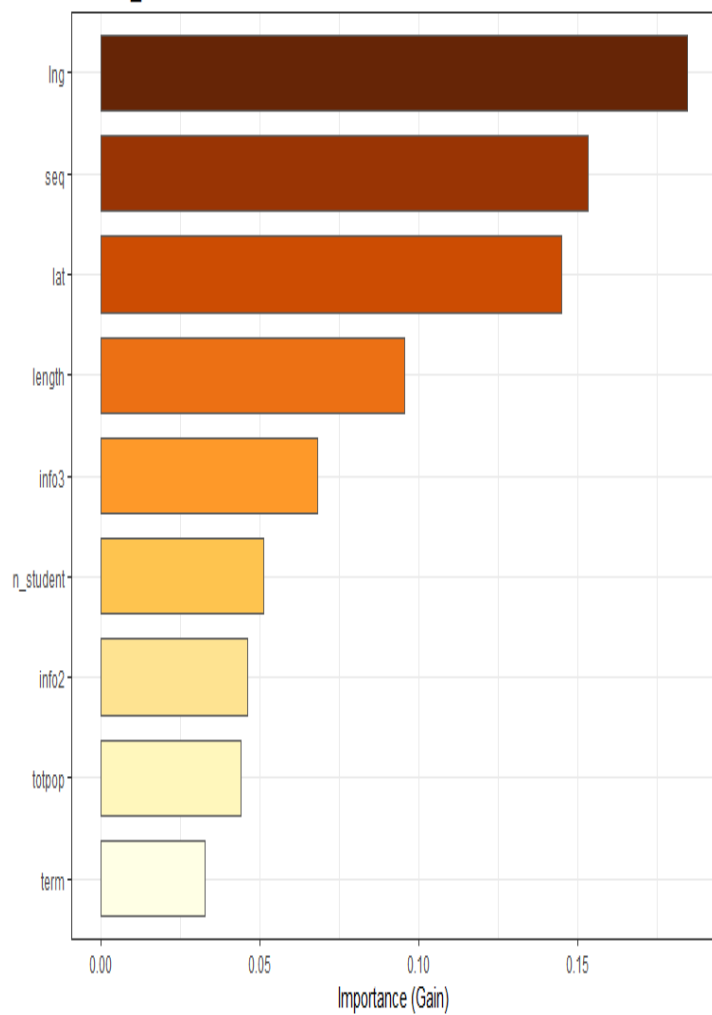
사용변수	lng	lat	사업체수	종사자수	평균종사자
사업체밀도	종사자밀도.A.	인구밀도.B.	종사자밀도비.A.B.100	n_subway	Mean.smr
Mean.smo	nstudent	info1	info2	info3	length
term	seq	busnum	cnt	routeType	totpop

	model	Max.depth	eta	rmse	rmse_test
1	morn_class1	30	0.2	4.114139198	12.45228069
2	morn_class2	50	0.5	0.206964311	16.6225823
3	morn_class3	30	0.1	0.904004724	16.5686303
4	even_class1	10	0.4	5.142094328	14.52295622
5	even_class2	50	0.1	0.328105589	11.21505685
6	even_class3	10	0.3	3.344717205	16.18162585

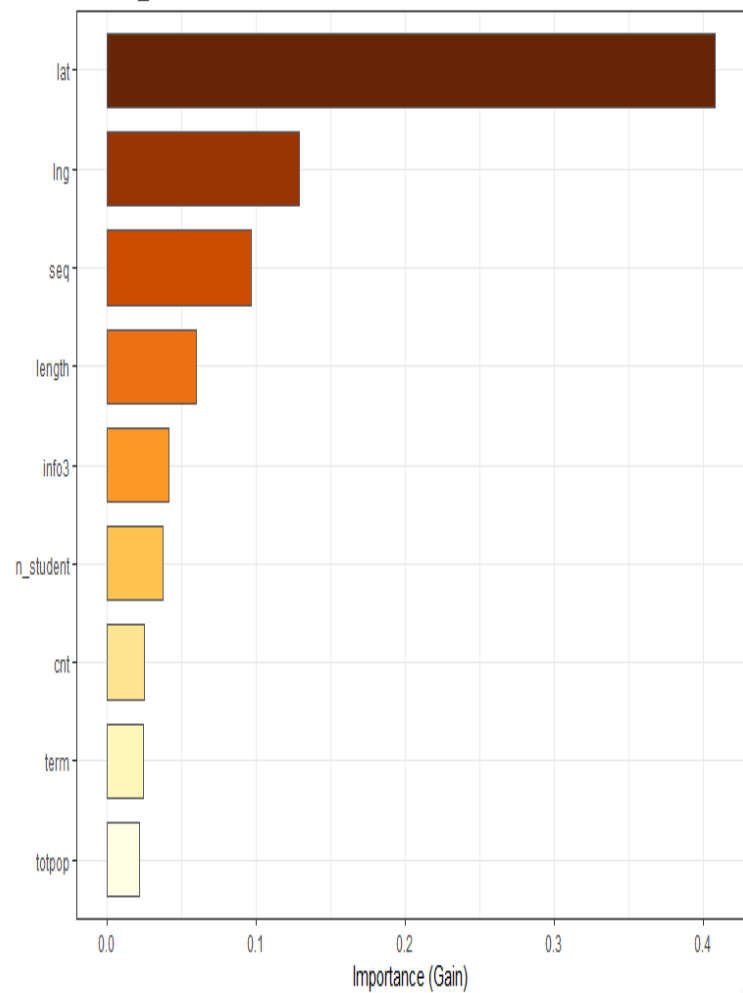
	Max.depth	eta
1	10	0.0
2	30	0.0
3	50	0.0
4	10	0.1
5	30	0.1
6	50	0.1
7	10	0.2
8	30	0.2
9	50	0.2
⋮	⋮	⋮
22	10	0.8
23	30	0.8
24	50	0.8
25	10	0.9
26	30	0.9
27	50	0.9
28	10	1.0
29	30	1.0
30	50	1.0

XGBoost Morning Class

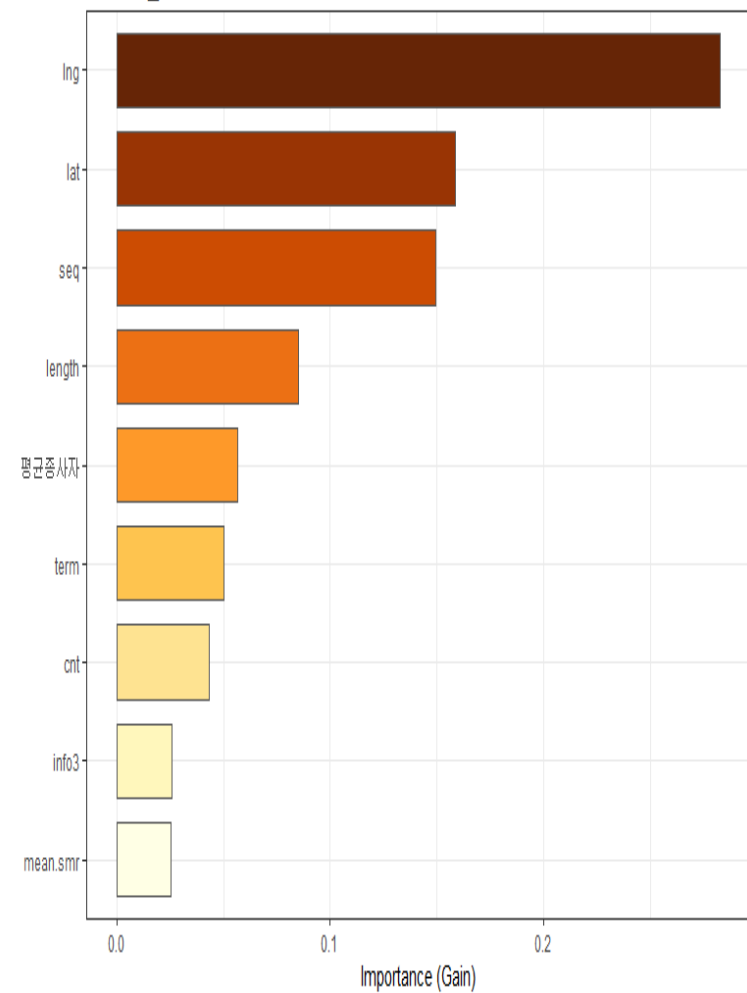
Class1_출근 순 승차수



Class2_출근 순 승차수

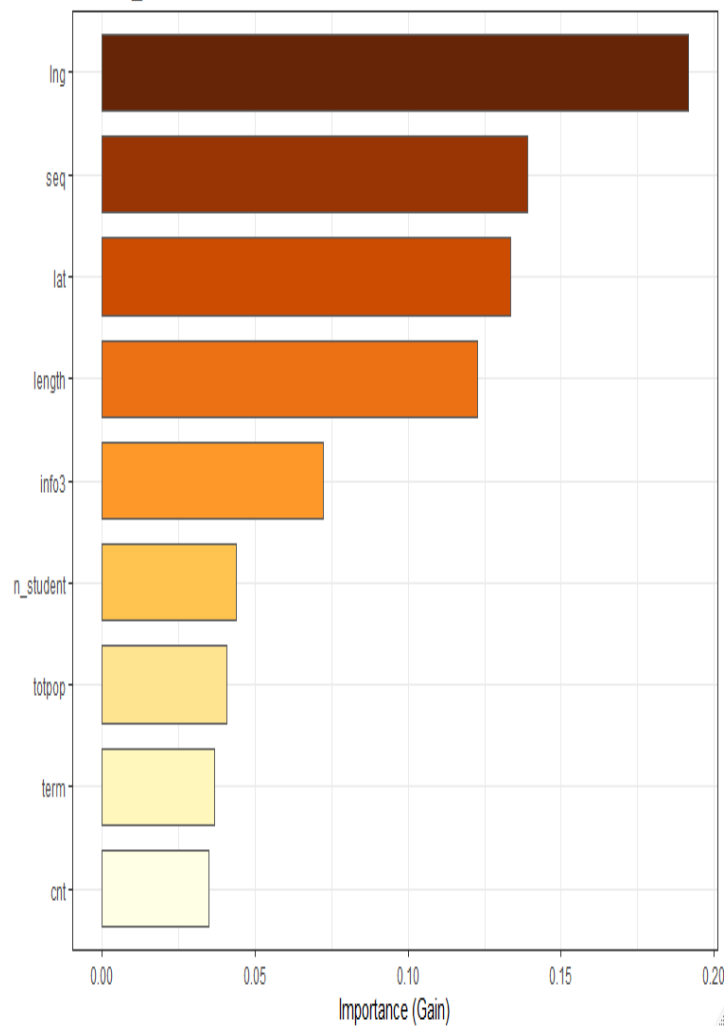


Class3_출근 순 승차수

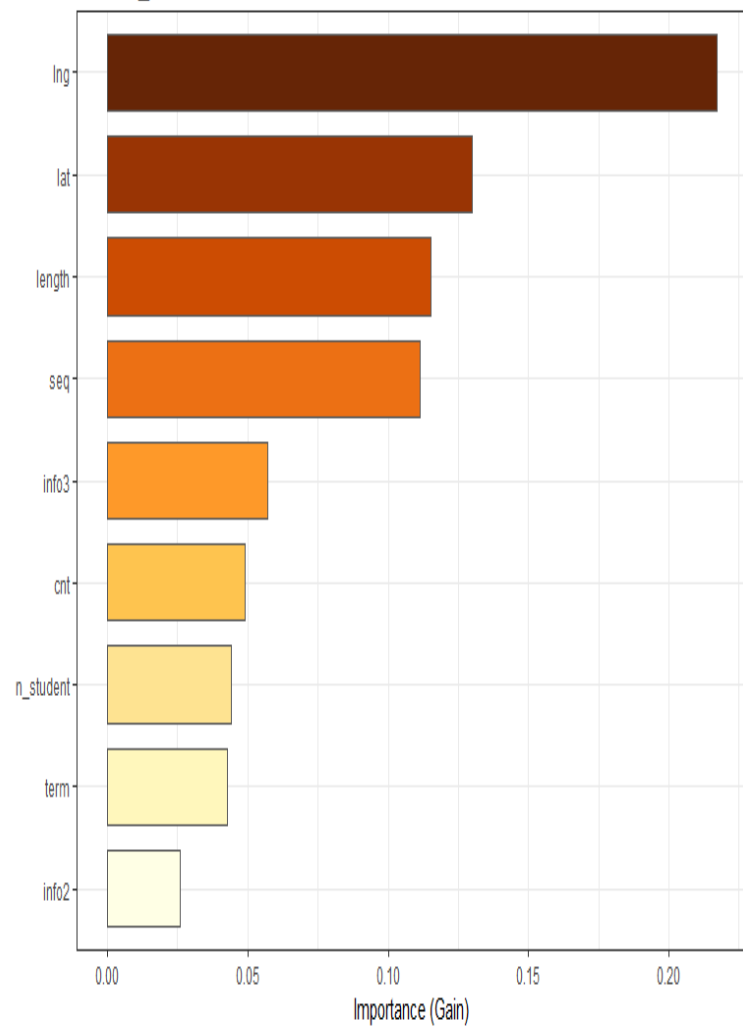


XGBoost Evening Class

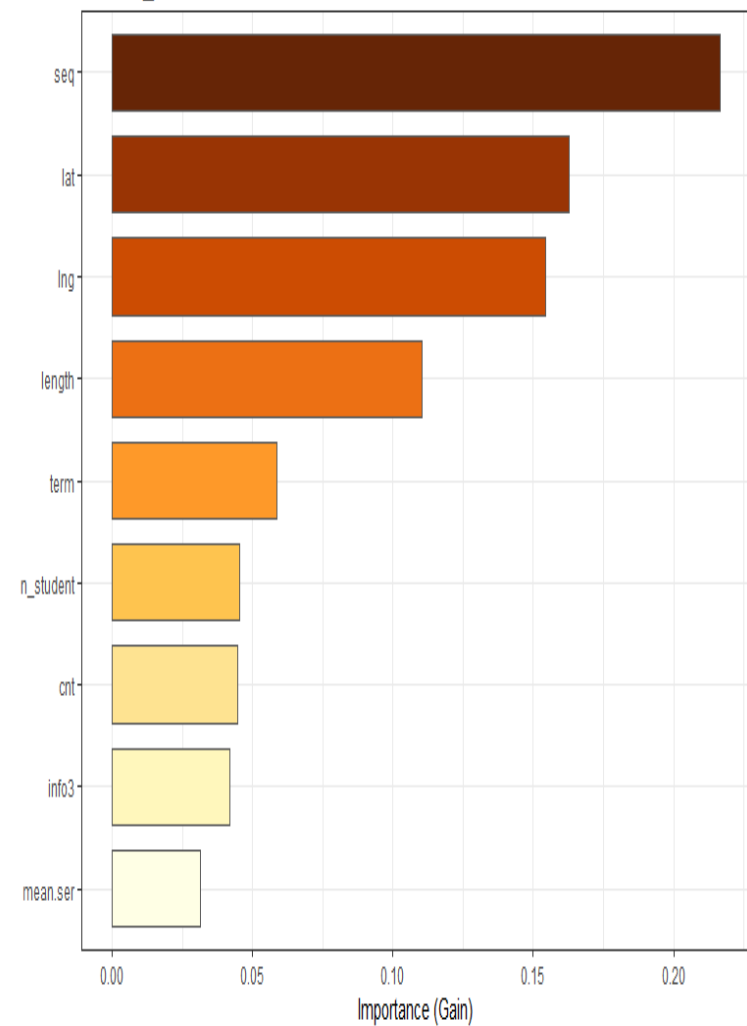
Class1_퇴근 순 승차수



Class2_퇴근 순 승차수

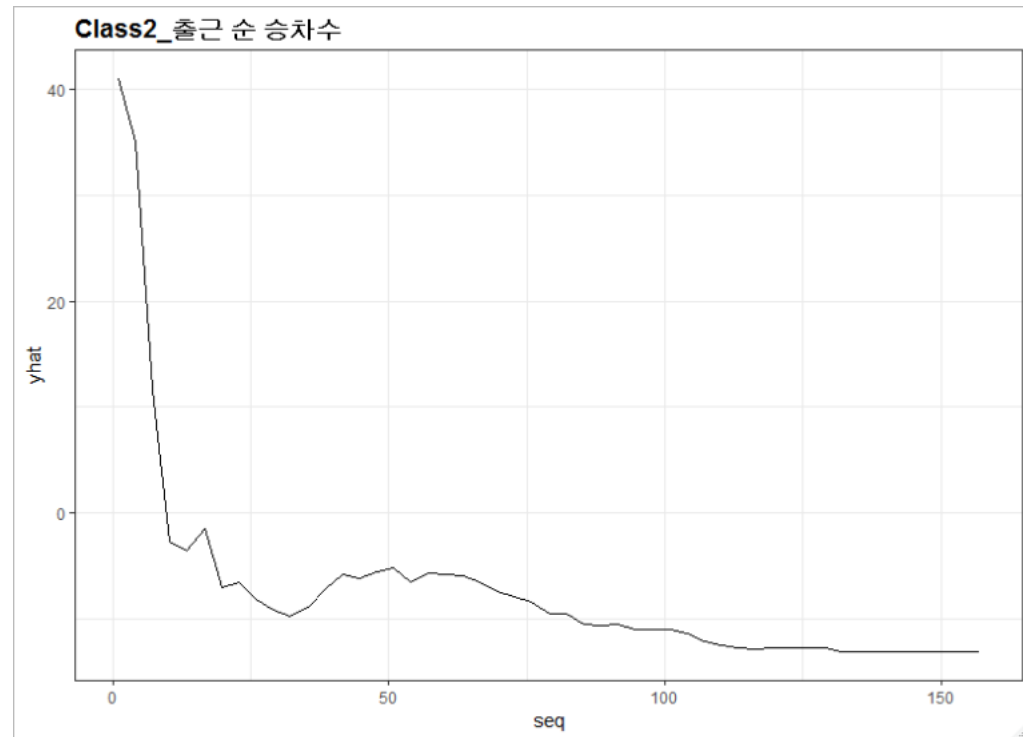


Class3_퇴근 순 승차수

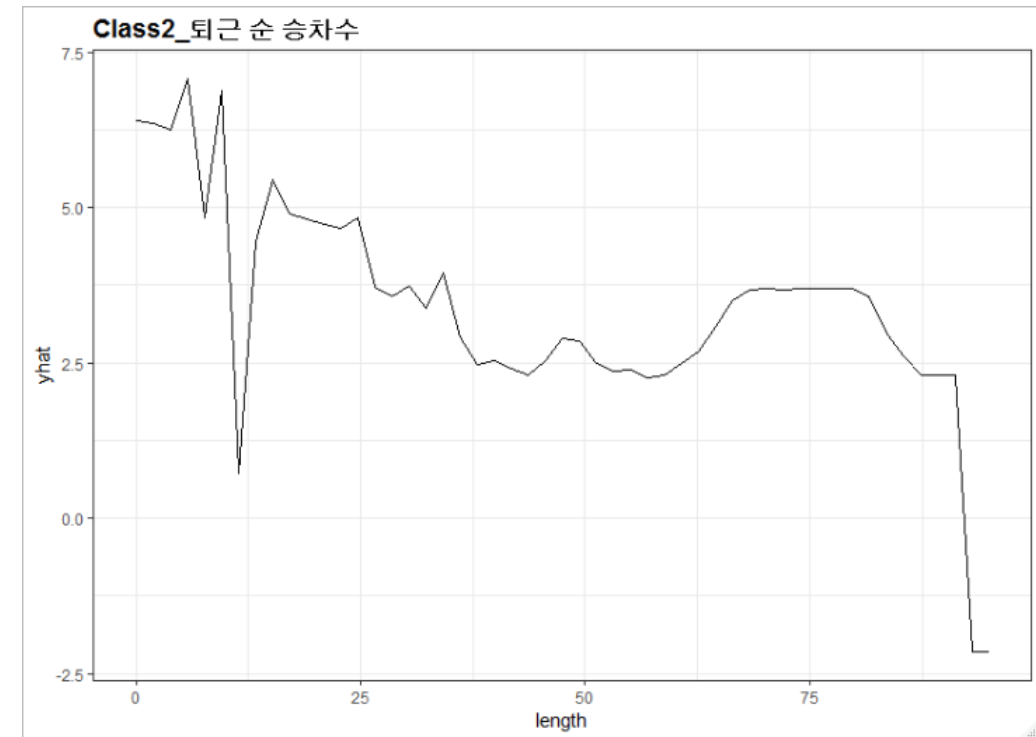


CLASS 2

출근

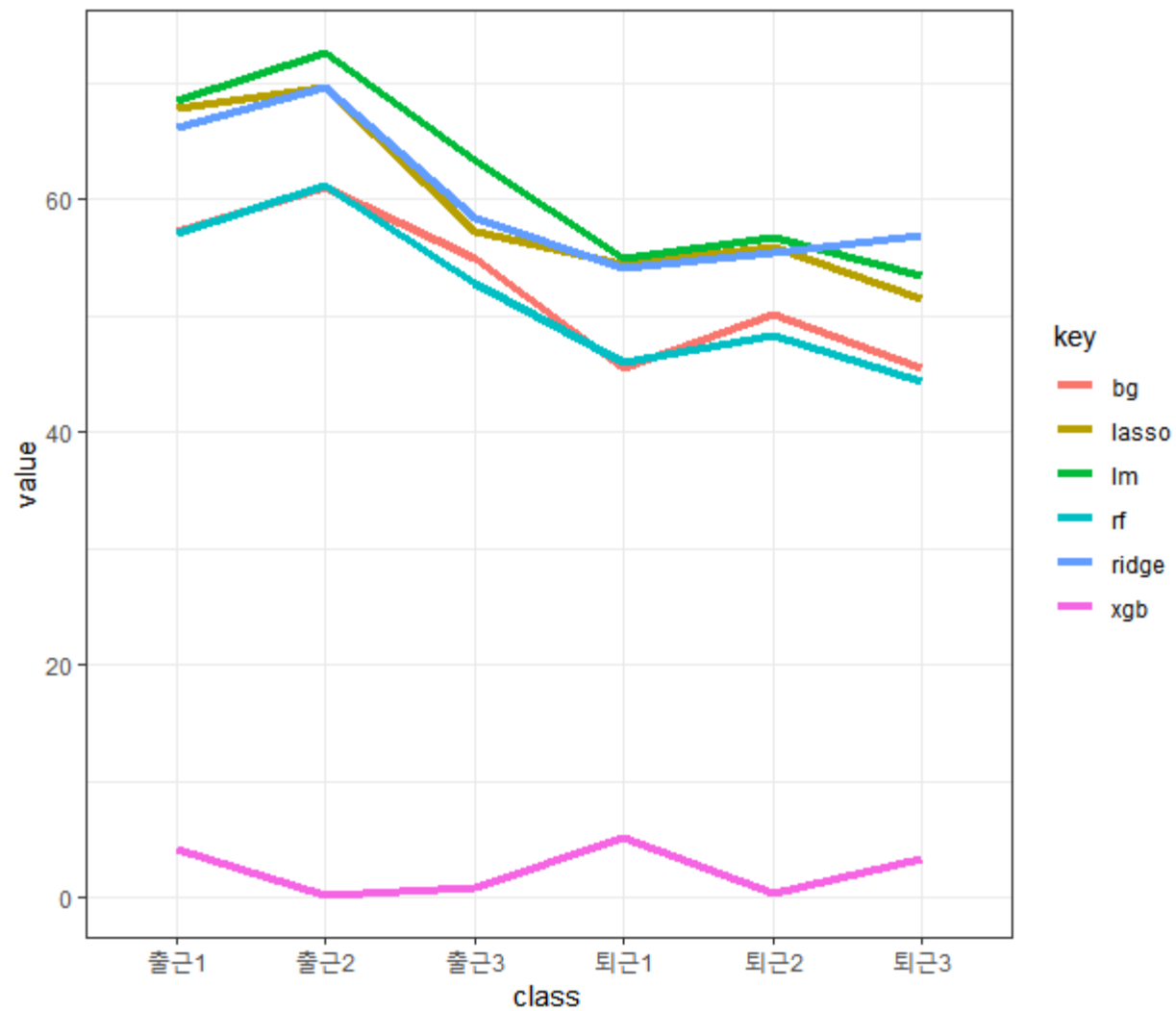


퇴근

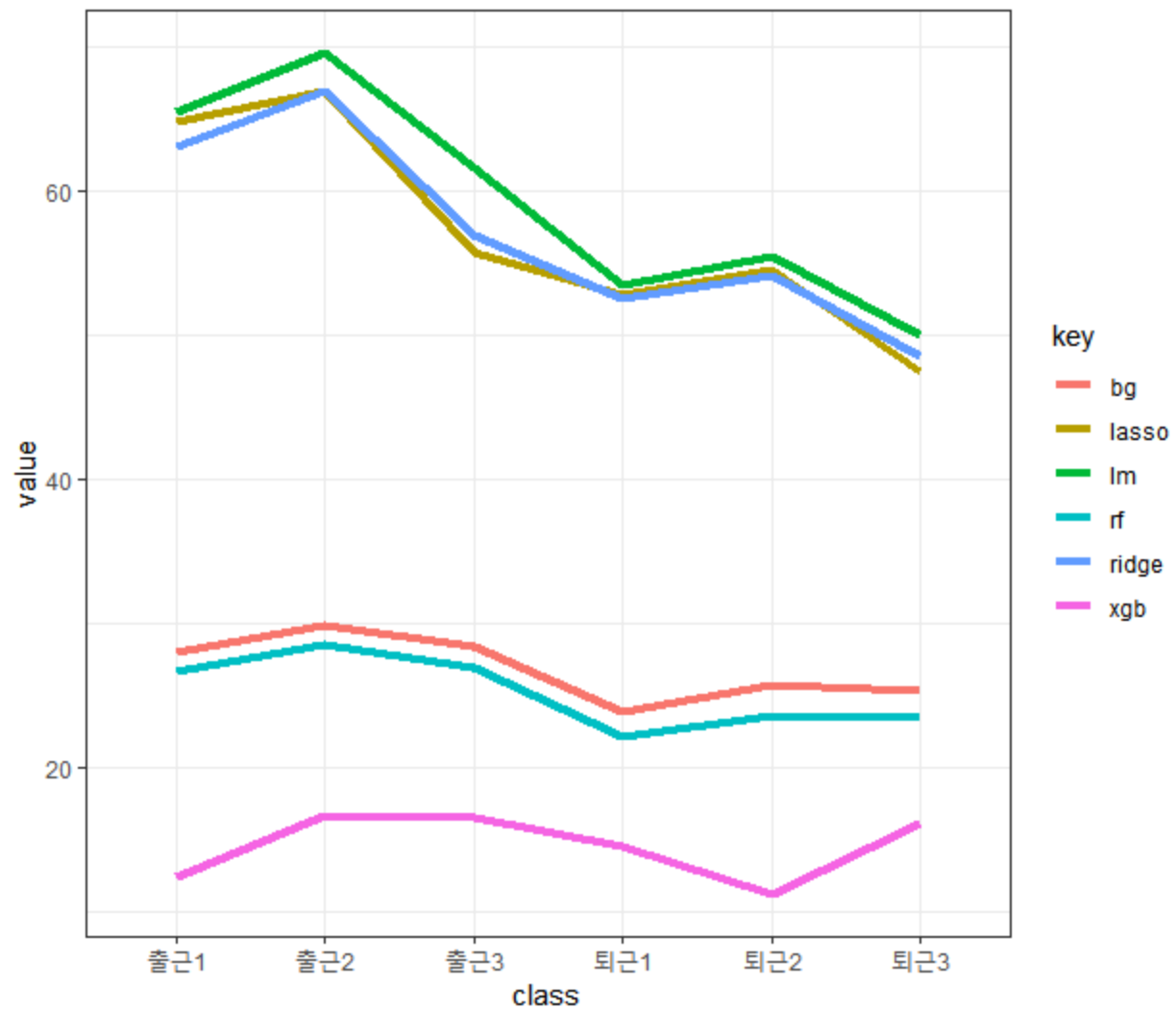


Model Comparison

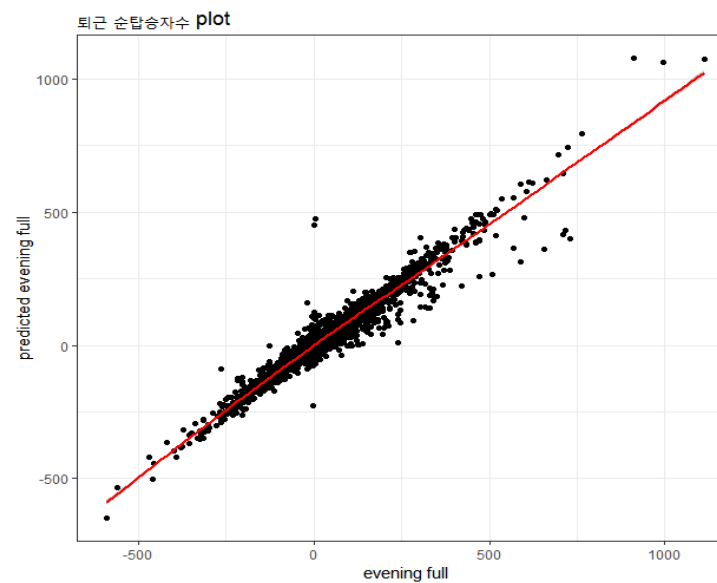
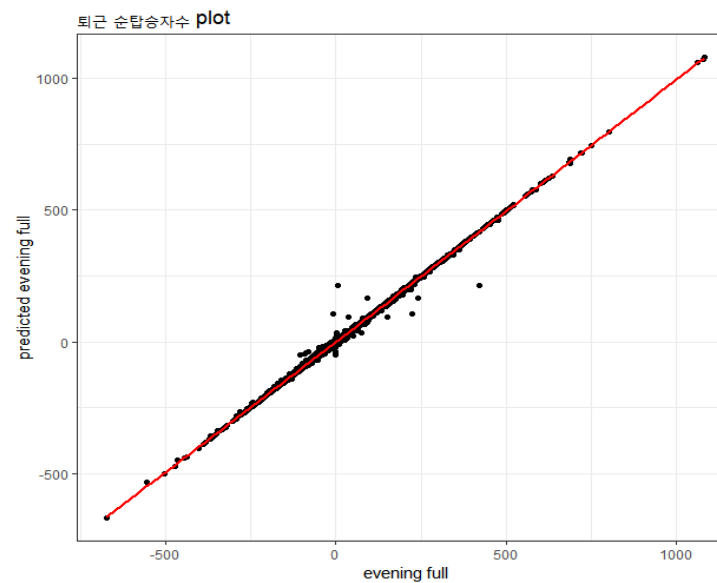
train RMSE



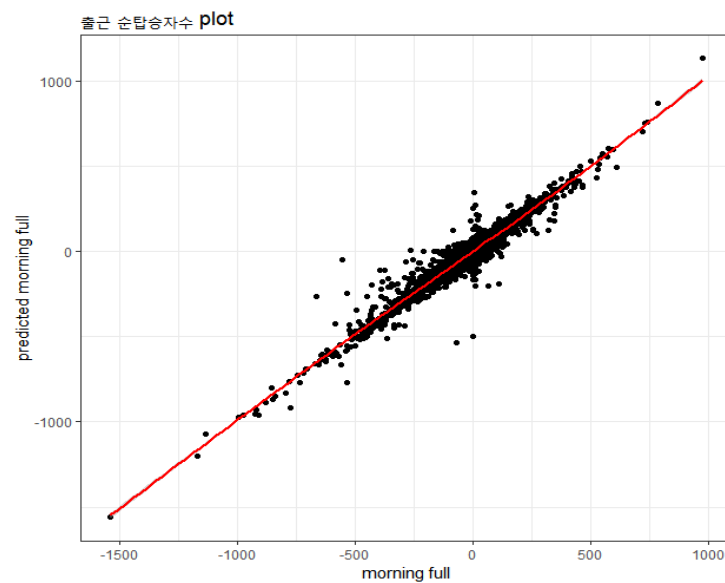
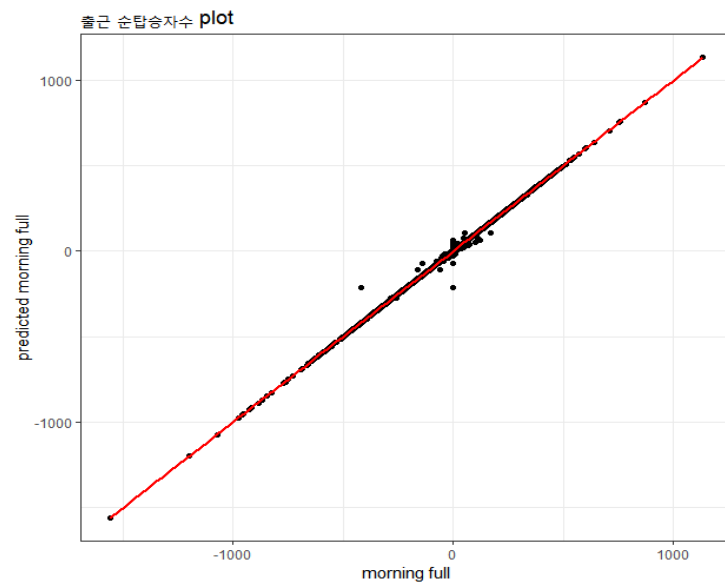
test RMSE



퇴근



출근



2017

2018

Fitted

이번 정류소는

입니다



ID: 01-001
다람쥐 버스란?



ID: 02-002
DATA 설명서 & EDA



ID: 03-003
MODELING



ID: 04-004
PROPOSITION

출근



노선예측



퇴근

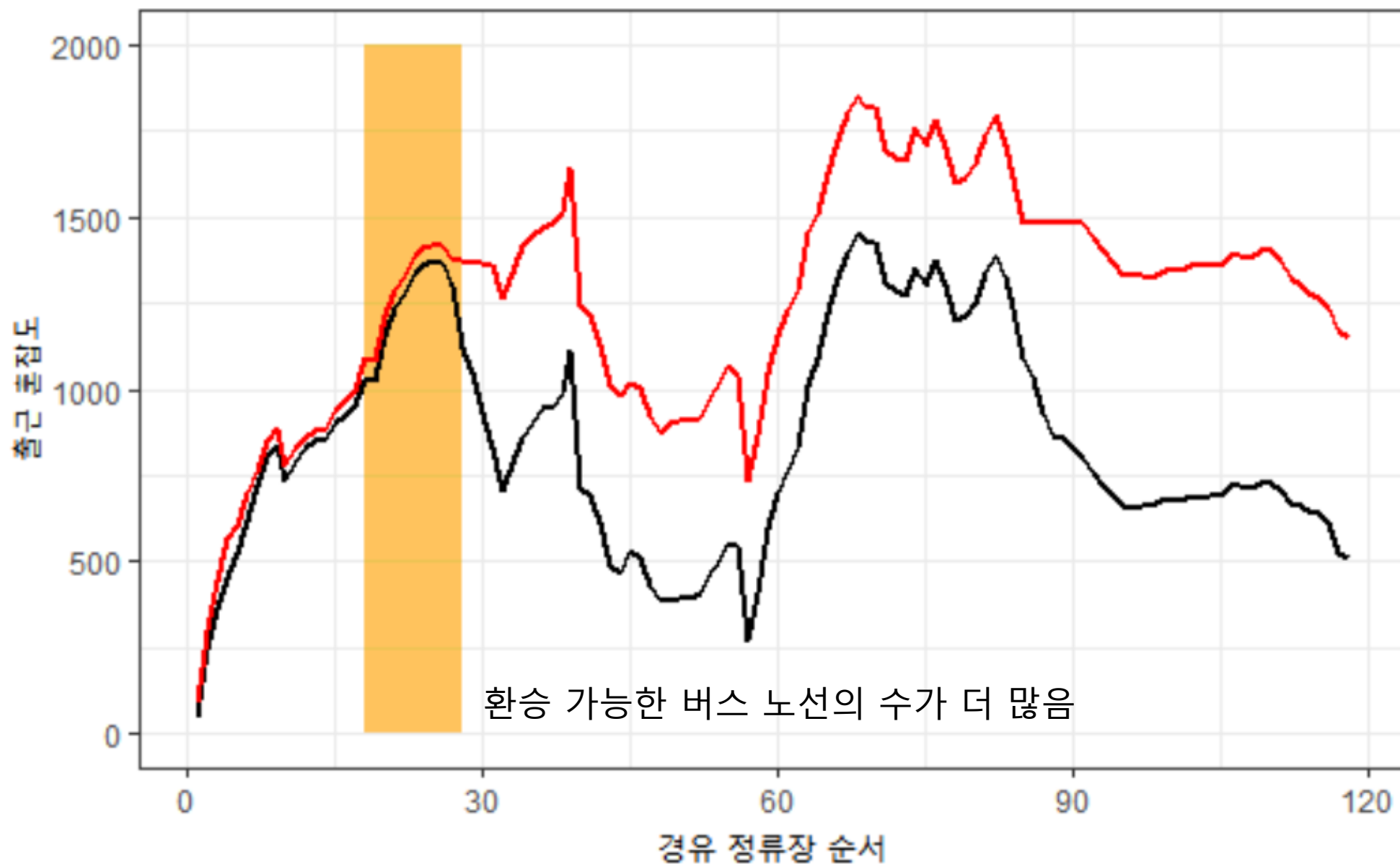
※혼잡 구간 예측 기준※

상위 95%의 혼잡도의 연속이 긴 구간 순
(출근: 1158.664, 퇴근: 834.1195)

출근	노선번호	연속구간
1	720	93
2	150	60
3	160	57
4	340	44

퇴근	노선번호	연속구간
1	7211	41
2	110B	39
3	2012	29
4	금천03	29
5	461	24
6	603	24
7	153	23
8	은평02	23
9	1154	20
10	340	20

160번

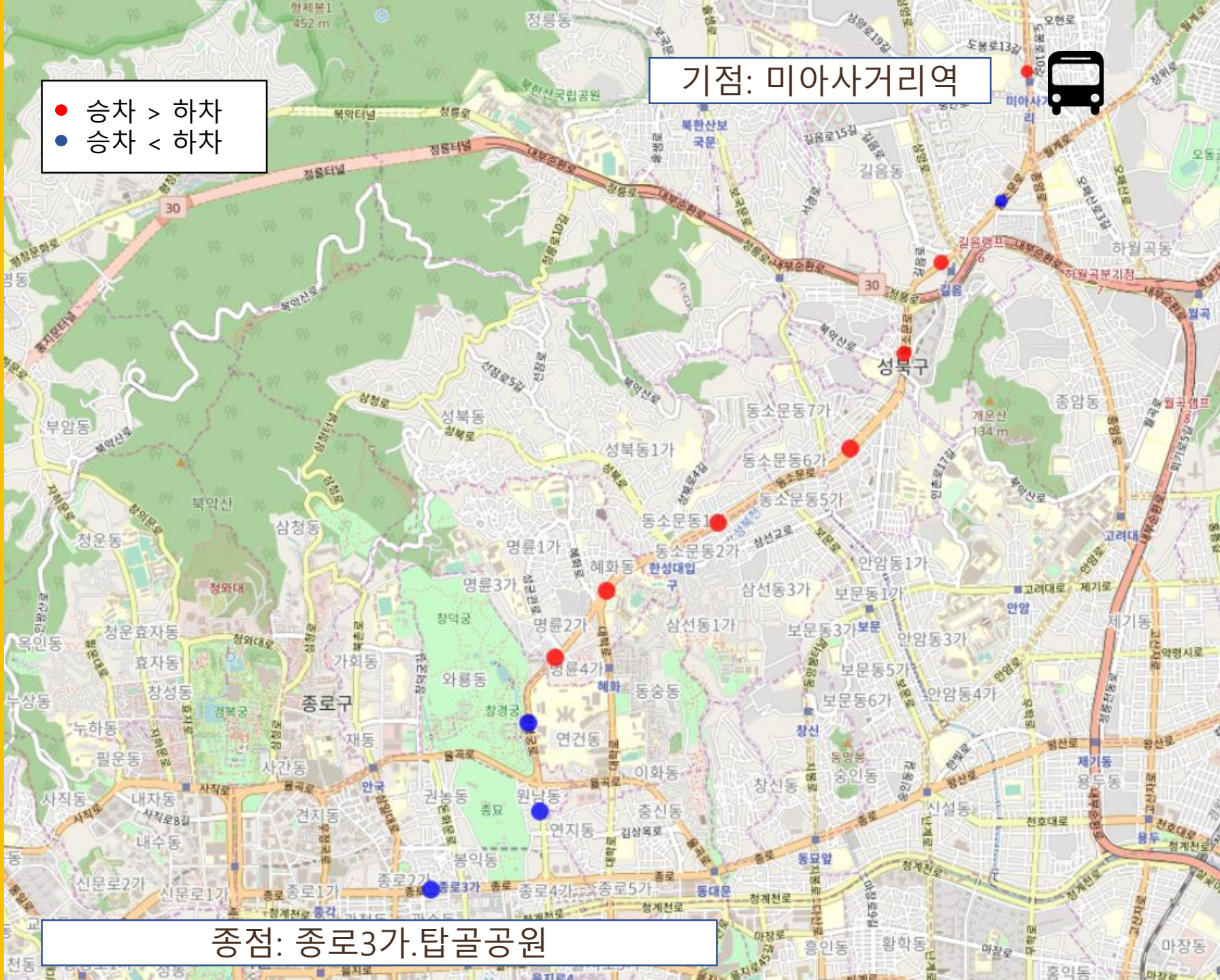


150, 160 출근

- 기점 ○ 미아사거리역
- 월곡뉴타운
- 길음뉴타운
- 미아리고개.미아리예술극장
- 돈암사거리.성신여대입구
- 삼선교.한성대학교
- 명륜3가.성대입구
- 창경궁.서울대학교병원
- 원남동
- 종점 ○ 종로3가.탑골공원

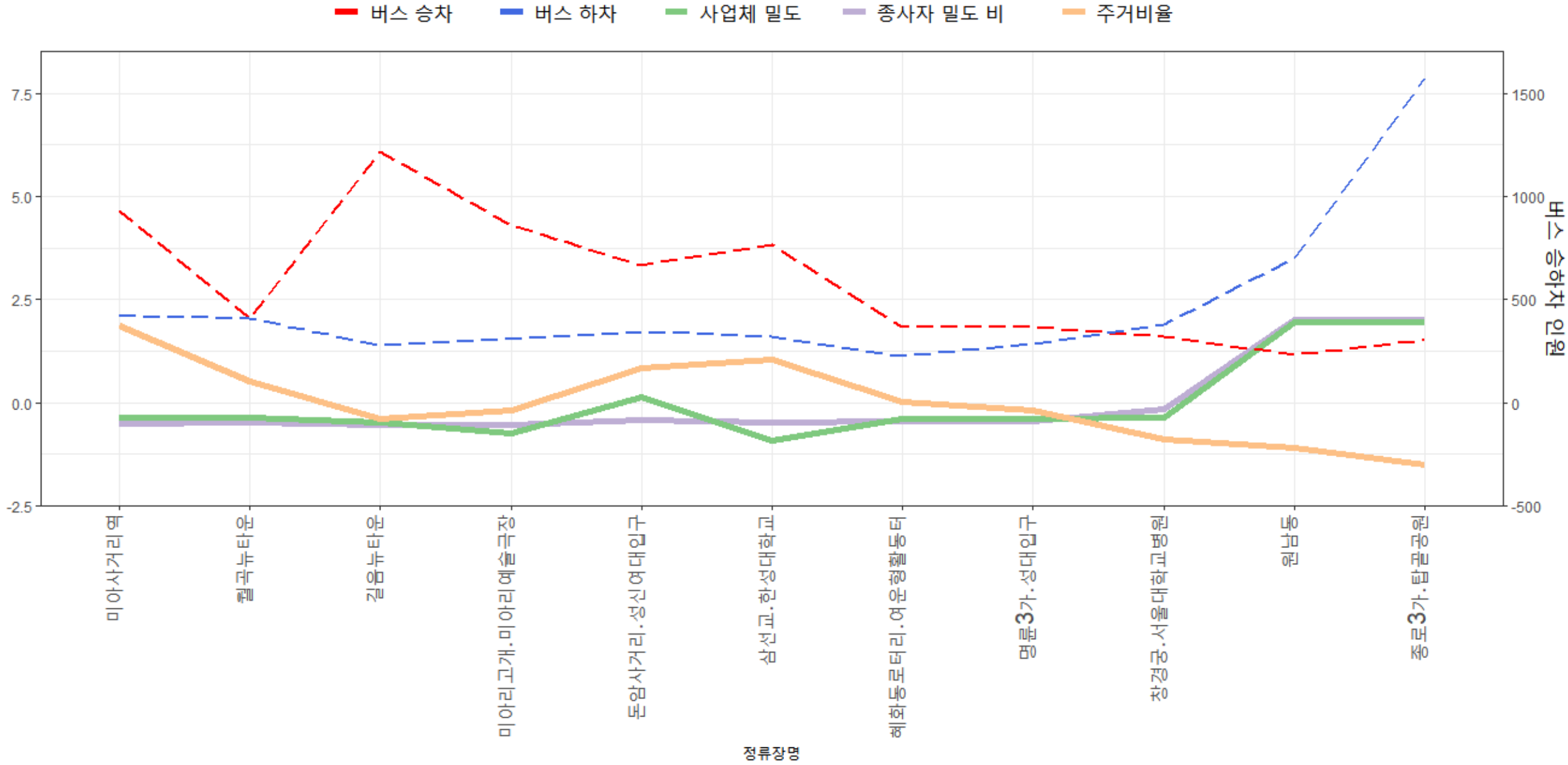
- 승차 > 하차
- 승차 < 하차

기점: 미아사거리역

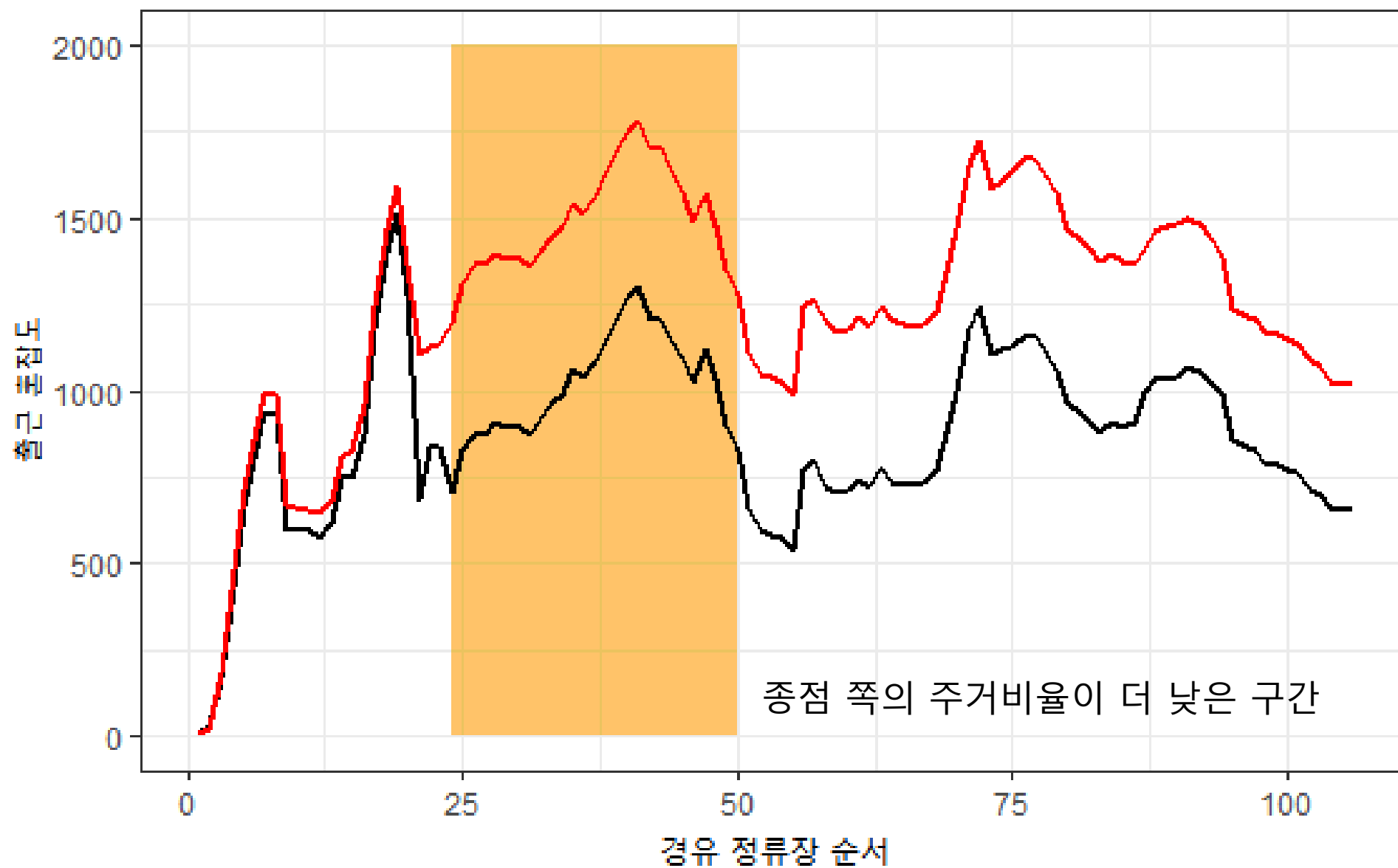


종점: 종로3가.탑골공원

출근시간대 160번 버스 노선의 혼잡도를 해소해 줄 새로운 다람쥐버스



340번 - 출근



340 출근

기점

천호역현대백화점

천호역.풍납시장

영파여고.강동구청역

현대아파트앞

풍납동극동.쌍용아파트

잠실파크리오아파트상가.올림픽공원

몽촌토성역.한성백제역

방이2동주민센터

송파나루역.방이시장

송파초등학교.송파동성당

송파여성문화회관

석촌동한솔아파트앞

석촌고분역.올림픽병원

송파구민회관구의회

삼전역

잠실우성4차현대아파트

잠실근린공원.성현교회

아주중학교

탄천교

램안하이스턴.대치순복음교회

은마아파트입구사거리

베스티안병원

한티역2번출구.서울강남고용노동지청

한티역7번출구

여감중학교.강남세브란스병원

도곡1차IPARK.모커리한방병원

종점

역삼럭키아파트.역삼월드메르디앙아파트

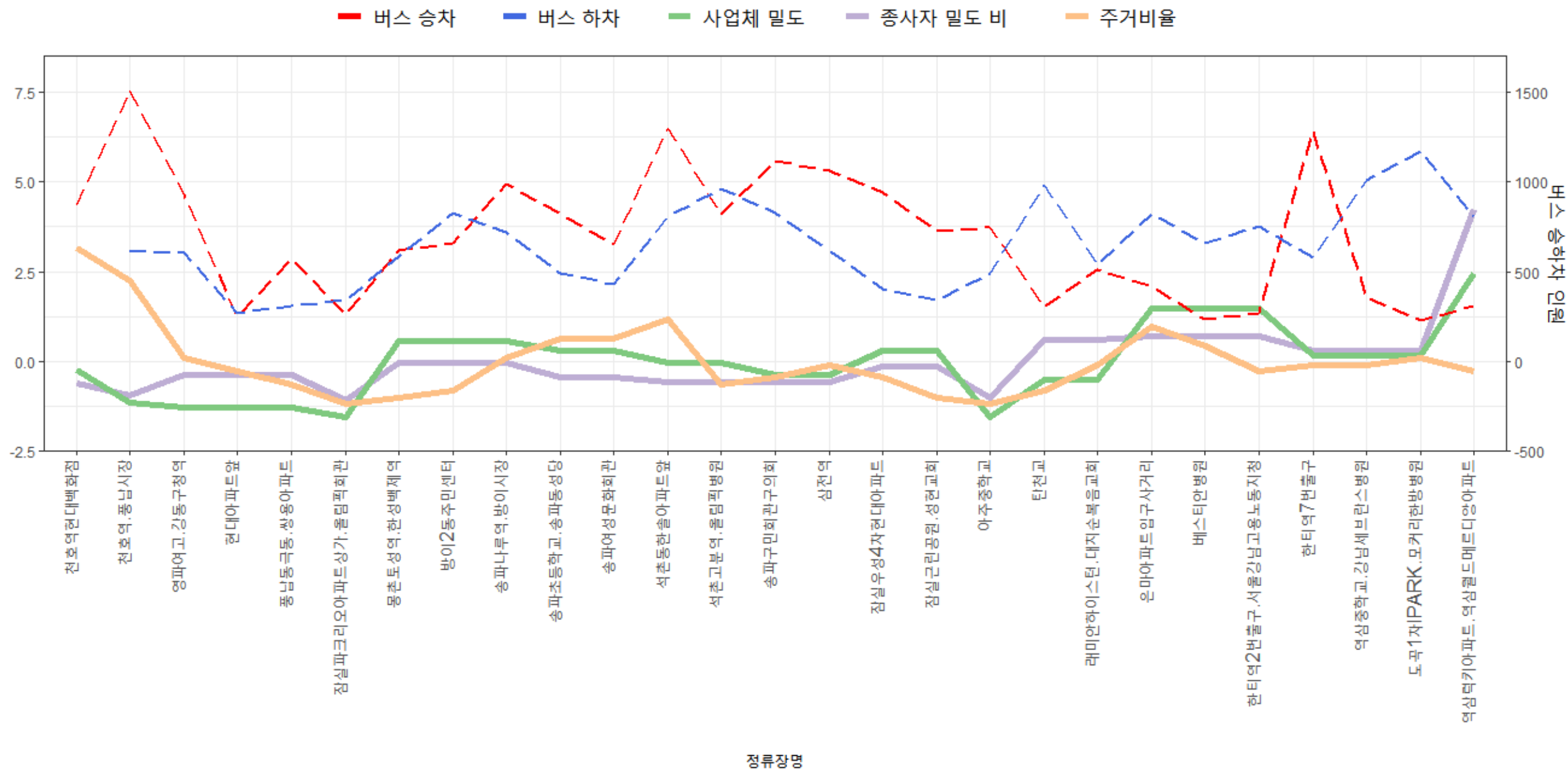
● 승차 > 하차
● 승차 < 하차

기점: 천호역현대백화점

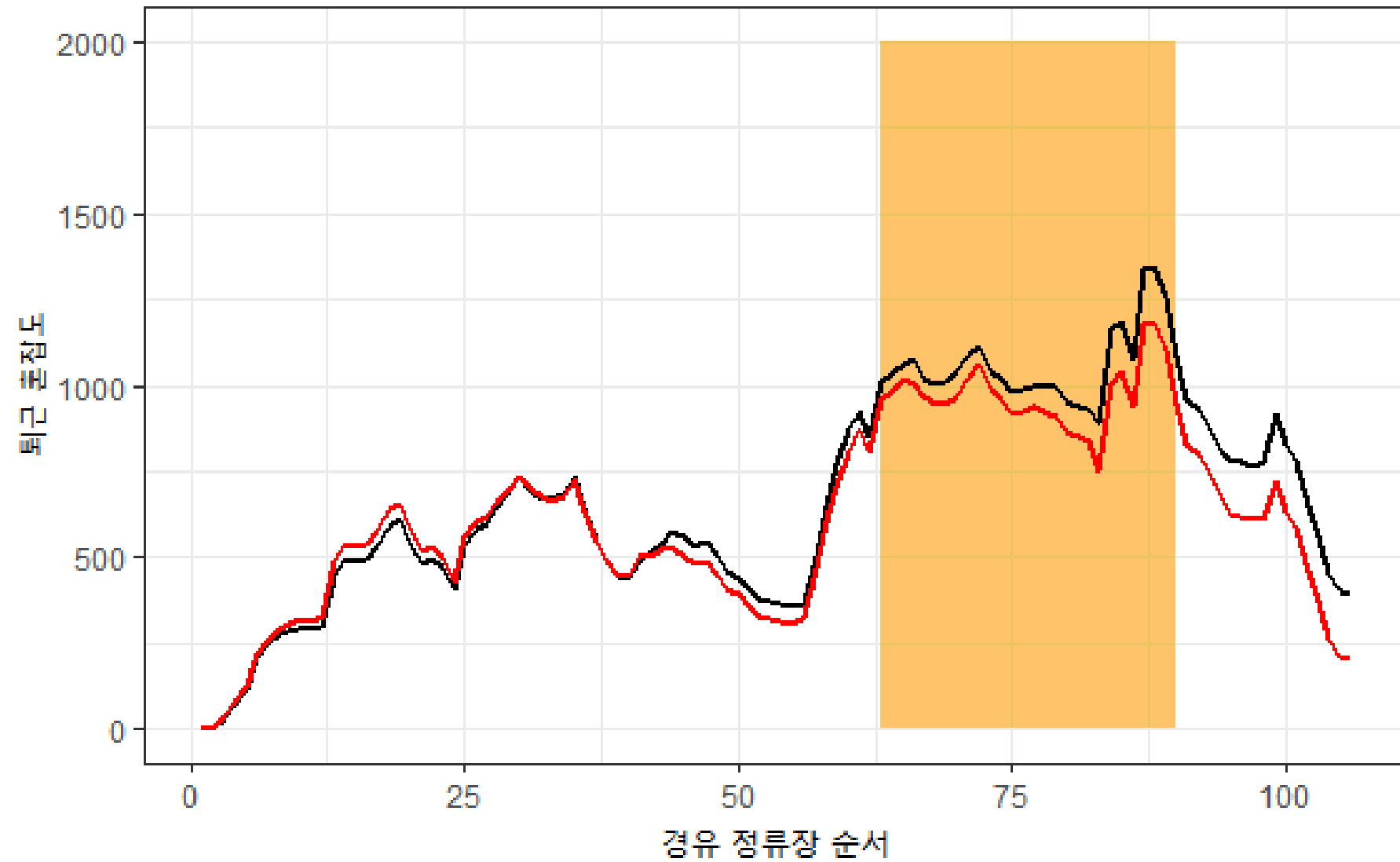


종점: 역삼럭키아파트.역삼월드메르디앙아파트

출근시간대 340번 버스 노선의 혼잡도를 해소해 줄 새로운 다람쥐버스



340번- 퇴근



340 퇴근

기점

한티역2번출구.서울강남고용노동지청

삼성아파트입구

은마아파트

우성아파트

아주중학교

아시아선수촌아파트삼거리.성현교회

잠실우성4차현대아파트

삼전역

송파구민회관구의회

석촌고분역.올림픽병원

석촌동한솔아파트

송파여성문화회관

송파초등학교.송파왕성당

송파나루역.방이시장

한성백제역.방이2동주민센터

몽촌토성역.한성백제역

올림픽회관

풍남동극동.쌍용아파트

강동구청역농협중앙회

강동구청역1번출구.영파여고

천호역

천호역현대백화점

천호동공원

암사역2번출구.암사종합시장

삼성광나루아파트.암사동유적

신암중학교

종점

강일중선사고.양지종합시장

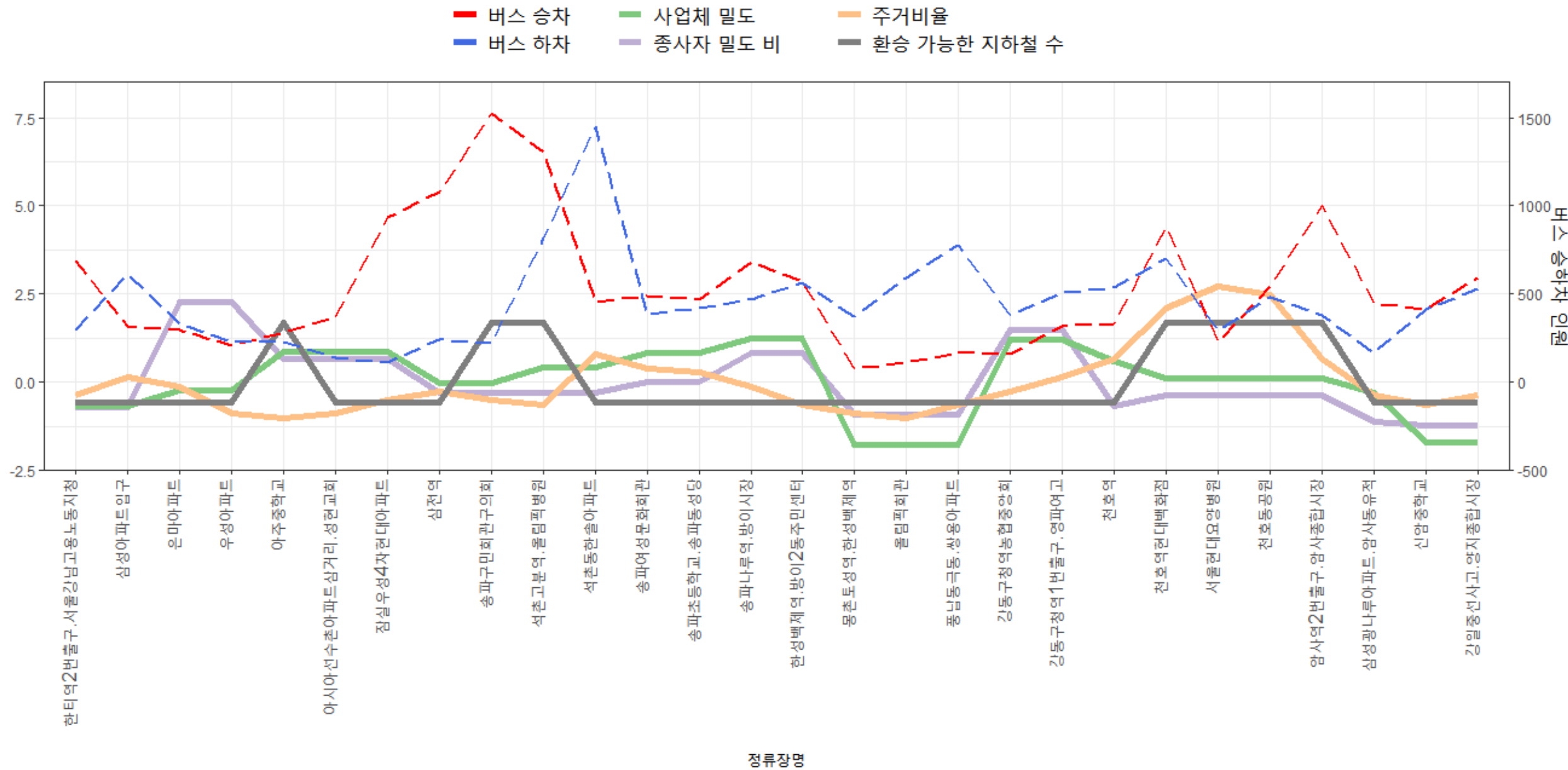
● 승차 > 하차
● 승차 < 하차



종점: 강일중선사고.양지종합시장

기점: 한티역2번출구.서울강남고용노동지청

퇴근시간대 340번 버스 노선의 혼잡도를 해소해 줄 새로운 다람쥐버스





Conclusion

의의: 출근 및 퇴근 시간의 버스 혼잡도 완화 및 대중교통 이용률 증가

한계 : 실제 혼잡도의 감소량은 예측 불가능

버스 노선	기.종점	대수	거리	배차간격	횟수	운행시간
8432(퇴)	한티역2번출구 ~ 강일중선사고	5	12.5km	10-12분	11	17:30 ~ 19:30
8432(출)	천호역현대백화점 ~ 역삼럭키아파트	4	11.6km	10-11분	12	07:00 ~ 09:00
8661	한티역2번출구 ~ 강일중선사고	5	7.5km	10-12분	11	07:00 ~ 09:00

종점입니다

모두 하차해주시길 바랍니다



ID: 01-001
다람쥐 버스란?



ID: 02-002
DATA 설명서 & EDA



ID: 03-003
MODELING



ID: 04-004
PROPOSITION

