

Introduction to Data Programming Mid Project

좌석 개선을 통한 등교 시 대중교통 이용 편의성 증대 방안 연구

강성운, 정민우

Agenda

1

문제 정의 및 연구 주제

2

선행 연구 조사

3

데이터 분석 과정

4

결론 및 기대효과

문제 정의 및 연구 주제

```
#install.packages("readr")
#install.packages("dplyr")

# csv 파일을 읽기
library(readr)
library(dplyr)

file_path <- 'C:/Users/WOON/Downloads/Transportation Data.csv'
df <- read_csv(file_path)

# 열 합쳐 '버스를 이용하는 이동'과 '이용하지 않는 이동' 데이터 작성
selected_columns_df <- df %>%
  select(버스, `지하철(철도)`, `버스 + 지하철`)

# 결측치 제거
filtered_df <- selected_columns_df %>%
  filter(
    between(버스, mean(버스, na.rm = TRUE) - sd(버스, na.rm = TRUE),
            mean(버스, na.rm = TRUE) + sd(버스, na.rm = TRUE)),
    between(`지하철(철도)`, mean(`지하철(철도)`, na.rm = TRUE) -
            sd(`지하철(철도)`, na.rm = TRUE),
            mean(`지하철(철도)`, na.rm = TRUE) + sd(`지하철(철도)`, na.rm = TRUE)),
    between(`버스 + 지하철`, mean(`버스 + 지하철`, na.rm = TRUE) -
            sd(`버스 + 지하철`, na.rm = TRUE),
            mean(`버스 + 지하철`, na.rm = TRUE) + sd(`버스 + 지하철`, na.rm = TRUE))
  )

# 'Using Bus'와 'Not Using Bus' 열을 만들고
# 첫 번째 열에 행 번호 추가
final_df <- filtered_df %>%
  mutate(ID = row_number(), # 행 번호 추가
         `Using Bus` = 버스 + `버스 + 지하철`,
         `Not Using Bus` = `지하철(철도)` %>%
           select(ID, `Using Bus`, `Not Using Bus`) # 열 순서 조정

# 각 열의 평균과 표준편차 계산
summary_stats <- final_df %>%
  summarise(
    Mean_Using_Bus = mean(`Using Bus`, na.rm = TRUE),
    SD_Using_Bus = sd(`Using Bus`, na.rm = TRUE),
    Mean_Not_Using_Bus = mean(`Not Using Bus`, na.rm = TRUE),
    SD_Not_Using_Bus = sd(`Not Using Bus`, na.rm = TRUE)
  )

# 마지막 행 평균, 표준편차 추가
final_df <- rbind(final_df,
                  c("Mean", summary_stats$Mean_Using_Bus,
                    summary_stats$Mean_Not_Using_Bus),
                  c("S.D", summary_stats$SD_Using_Bus,
                    summary_stats$SD_Not_Using_Bus))

# 데이터 프레임 열 이름 설정
colnames(final_df) <- c("ID", "Using Bus", "Not Using Bus")

# 결과를 출력 및 CSV 파일 저장
print(final_df)
write_csv(final_df, 'C:/Users/WOON/Downloads/final_data.csv')
```



좌석 개선을 통한 대중교통 이용
편의성을 효과적으로 증대 하려면
버스의 좌석의 개선도 필수적이다.

R 이용, csv 파일 데이터를 1시그마 이내의 데이터만으로 **전처리** 및 **평균, 표준편차** 계산하여 새로 저장하는 코드를 활용하여 '버스'에 대한 개선과정도 필수적이라는 결론 도출

기존 지하철 좌석개선의 효과 분석

※ 객실의자 개량 시 기대효과

구 분		현 행	(개선) 7인석만 제거 (42명 확보)	
			탑승 인원	혼잡율
정원 (160명)		혼잡율 100%	160명	79.2% (△20.8%)
혼잡 (320명)		혼잡율 200%	320명	158.4% (△41.6%)
효과	4호선	혼잡율 193.4% (310명)	310명	153.4% (△40.0%)
	7호선	혼잡율 164.2% (263명)	263명	130.1% (△34.1%)

○ 4호선과 7호선은 '23년 3분기 기준 최고 혼잡도가 각각 193.4% 164.2%로 여전히 150%를 초과한다.

연 도	1호선	2호선	3호선	4호선	5호선	6호선	7호선	8호선
'22.4분기	105.6%	172.3%	154.8%	185.5%	140.9%	113.8%	160.6%	136.8%
'23.3분기	108.2%	148.6%	144.2%	193.4%	130.9%	111.6%	164.2%	140.0%

객실의자 제거를 통한 지하철 혼잡율 개선은 비교적 단순한 계산으로 1량당 6개인 7인석 제거를 통해 얻을 수 있는 효용은 승차 인원을 42명 추가 확보가 가능하다는 점에서 기존 정원 대비 정원이 202명으로 증가할 것으로 보도 자료를 내보내었다.

혼잡율에 대한 용어 정리



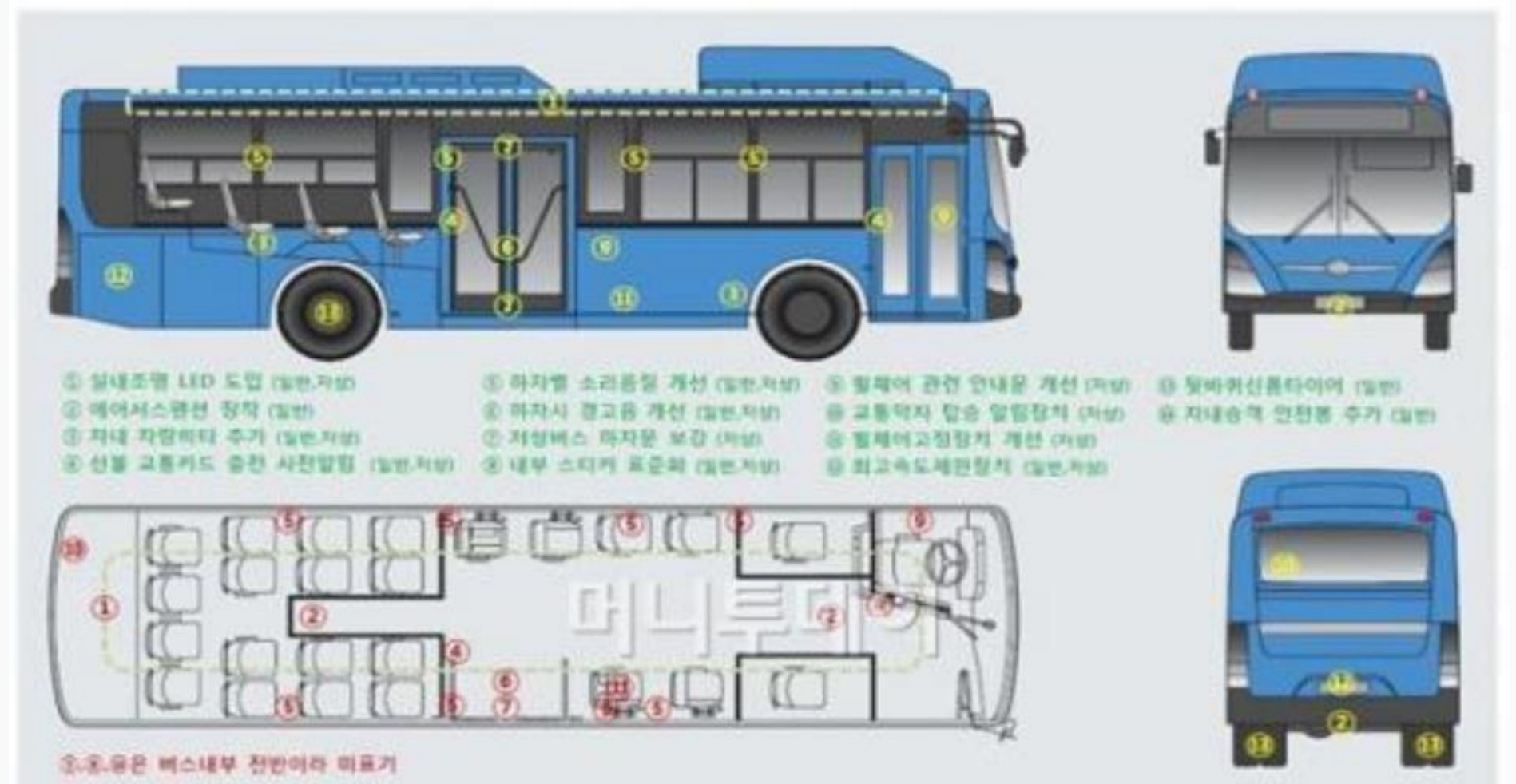
혼잡율(193.4%) = 310명/160명



혼잡율(153.4%) = 310명/202명

혼잡율 = 탑승 인원/정원

지하철과 버스의 공통점 및 차이점 분석



공통점

- 대중교통 수단 : 두 방식 모두 많은 사람들이 이용하는 대중교통 수단
- 정해진 노선 운행 : 버스와 지하철 모두 정해진 노선을 따라 운행
- 스케줄에 따른 운행 : 일정한 시간표에 따라 운행되며, 승객들은 이를 참고하여 이동

지하철과 버스의 공통점 및 차이점 분석

차내 혼잡도 (상태정의)	버스유형 (차량길이)	표시 아이콘	대형버스 (10m초과)	중형버스 (8~10m)	소형버스 (8m이하)
여유 좌석에 앉을 수 있는 수준			25명 이하	20명 이하	10명 이하
보통 신체접촉 없음, 쾌적한 입석 가능			26명 ~ 40명	21명 ~ 35명	11명 ~ 20명
혼잡 신체접촉 일부 발생, 일부 탑승자 불쾌			41명 ~ 55명	36명 ~ 50명	21명 ~ 25명
매우 혼잡 신체접촉 크게 발생, 대부분 탑승자 불쾌			56명 이상	51명 이상	26명 이상

내 손안에 서울		
구분	버스 차내 혼잡도	
여유		좌석에 앉을 수 있는 정도
보통		입석 승객이 손잡이를 하나씩 잡고 서 있을 수 있는 정도
혼잡		입석 승객들 사이 통로에 까지 승객이 서 있고, 입석 승객의 몸이 맞닿는 정도(이상)

차이점

버스

정원 40명 좌석 25개 (좌석비율 62.5%)
주변 교통환경이나 운행환경에 영향을 받음

지하철

정원 160명 좌석 42개 (좌석비율 26.5%)
주변 교통환경이나 운행환경에 영향을 받지 않음

*버스 종류에 따른 차이가 있지만, 주로 사용되는 모델인 "현대슈퍼에어로시티"의 사례를 통해 비교

데이터 분석 과정

01

연구 가정 및 전처리

02

분석 과정

03

결과 도출

데이터분석 - 연구 가정

분석 대상

동국대학교 주변을 하차하는 104번, 105번, 140번, 144번, 301번, 421번, 463번, 470번, 472번, 507번, 604번, 741번, 7011번, 7212번 버스의 데이터

분석 기간

10/23~10/27

버스 이용 인원은 공휴일 제외하면 큰 편차가 없기 때문에 공휴일이 없는 주간의 데이터를 수집

분석 방법

해당 기간동안의 재차인원을 산술평균으로 계산
(재차인원: 현재 탑승중인 인원 중 해당 정류장에서 하차하지 않는 인원)

데이터분석 - 전처리

교통카드빅데이터
통합 정보 시스템

로그인

주제별 분석 > 노선-정류장 지표 > 노선별 이용량

노선별 차내 재차인원

(단위: 명)

다운로드

노선	기종점	순번	정류장명	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	00	01	02	03
104	선일교통(가상) - 선일교 통종점(가상)	1	선일교통(가상)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
		2	강북청소년수련관	4	1	3	7	5	4	3	3	2	4	4	4	5	7	6	8	6	1	0	0	0	0	0	0
		3	무량사입구	10	1	1	13	7	5	4	2	5	4	5	5	7	8	7	9	7	2	0	0	0	0	0	0
		4	국립4.19민주묘지	12	2	3	14	10	5	4	4	7	5	7	5	9	7	3	7	3	2	1	0	0	0	0	0
		5	국립4.19민주묘지입구사거리	21	2	4	15	13	7	6	6	8	6	6	6	12	7	8	5	2	2	0	0	0	0	0	0
		6	인수동장미원	26	7	8	20	18	11	7	7	7	9	7	10	12	10	9	5	6	2	3	0	0	0	0	0
		7	가오리역	30	11	10	20	18	13	8	9	9	9	10	11	10	16	13	7	9	3	4	3	0	0	0	0
		8	우이초등학교	30	13	12	28	25	16	9	12	10	10	9	12	13	17	12	8	4	5	2	0	0	0	0	0
		9	수유보건지소	31	13	13	31	23	15	12	13	12	11	11	9	12	10	15	6	8	6	3	2	0	0	0	0
		10	불래골입구.수유1동주민센터	33	16	18	33	26	21	14	12	11	13	11	12	12	19	16	10	5	5	4	0	0	0	0	0
		11	롯데마트삼양점.래미안1차아파트	37	18	19	32	28	23	14	15	11	12	12	13	11	19	17	10	7	5	6	3	0	0	0	0
		12	삼양초등학교	38	21	19	34	30	23	16	14	12	12	9	16	9	17	19	9	3	5	0	4	0	0	0	0
		13	강북청소년문화정보도서관입구	34	22	20	34	32	23	18	14	12	13	11	14	17	15	20	12	8	5	5	5	0	0	0	0
		14	삼양동사거리	43	24	23	34	33	28	22	15	13	19	13	16	16	18	20	13	11	4	6	6	0	0	0	0

기간선택

일

2023-10-23

공간선택

시도

시군구

읍면동

서울특별시

노선조회결과

선택된 노선 수: 14건

741

검색

105_7단지영업소 - 7단지영업소... ✕
140_도봉산역광역환승센터 - 도봉... ✕
144_우이등차고지(기점가상) - ... ✕
301_서울버스종점 - 서울버스종점... ✕
7011_은평공영차고지 - 은평공영... ✕
7212_은평공영차고지 - 은평권역... ✕
470_상암차고지(가상) - 상암차... ✕
472_개포동차고지(가상) - 개포... ✕
507_시흥동차고지(기점가상) - ... ✕

전체삭제

일자, 기종점, 노선, 정류장 순서, 정류장명,
시간대 별 재차인원이 하루 단위로
구성된 재차인원 데이터



목표 기간동안의 산술평균 자료

데이터분석 - 전처리

```
library(dplyr)

board_1023 <- read.csv("1023_재차인원.csv")
board_1024 <- read.csv("1024_재차인원.csv")
board_1025 <- read.csv("1025_재차인원.csv")
board_1026 <- read.csv("1026_재차인원.csv")
board_1027 <- read.csv("1027_재차인원.csv")

file_paths <- c("1023_재차인원.csv", "1024_재차인원.csv",
               "1025_재차인원.csv", "1026_재차인원.csv", "1027_재차인원.csv")

board_mean <- read.csv(file_paths[1], header = TRUE)

# 나머지 파일 읽어서 같은 행의 문자열 값은 1번 파일과 동일하게 설정하고, 같은 행/열의 숫자 값에 대해 합 계산
for (file_path in file_paths[-1]) {
  data <- read.csv(file_path, header = TRUE)

  # 같은 행의 문자열 값은 1번 파일과 동일하게 설정
  data[, 1:4] <- board_mean[, 1:4]

  # 같은 행/열의 숫자 값에 대해 합 계산
  board_mean[, 5:ncol(board_mean)] <- (board_mean[, 5:ncol(board_mean)] + data[, 5:ncol(data)])
}

board_mean[, 5:28] <- board_mean[, 5:28]/5 # 평균으로 변환
write.csv(board_mean, "평균_재차인원.csv", fileEncoding="UTF-8", row.names=FALSE)
```

일별 재차인원에 대한 정보를 기간동안의 산술 평균 자료로 변환하여 저장하는 코드

데이터분석 - 분석

```

degree_crowd <- data.frame(
  노선=board_mean$노선,
  기종점=board_mean$기종점,
  순번=board_mean$순번,
  정류장명=board_mean$정류장명
) # 혼잡도 데이터프레임 생성

degree_crowd[, 5:ncol(board_mean)] <- lapply(board_mean[, 5:ncol(board_mean)], function(x) x / 40)
# 혼잡도 계산

write.csv(degree_crowd,"혼잡도.csv",fileEncoding="UTF-8",row.names=FALSE)
library(dplyr)

data <- read.csv("혼잡도.csv")

average_values <- data %>%
  summarise(across(5:28, mean, na.rm = TRUE))

sorted_columns <- names(average_values)[order(-unlist(average_values))]

head(sorted_columns)

# 7,8,9,16,17,18 을 출퇴근시간으로 선정

write.csv(sorted_columns,"시간별 평균 혼잡도(내림차순).csv",fileEncoding="UTF-8",row.names=FALSE)
data <- data %>%
  mutate(max=pmax(!!!select(.,5:28),na.rm=TRUE))
data <- data %>%
  arrange(desc(max))

desc_crowd <- data %>%
  distinct(노선,.keep_all=TRUE)

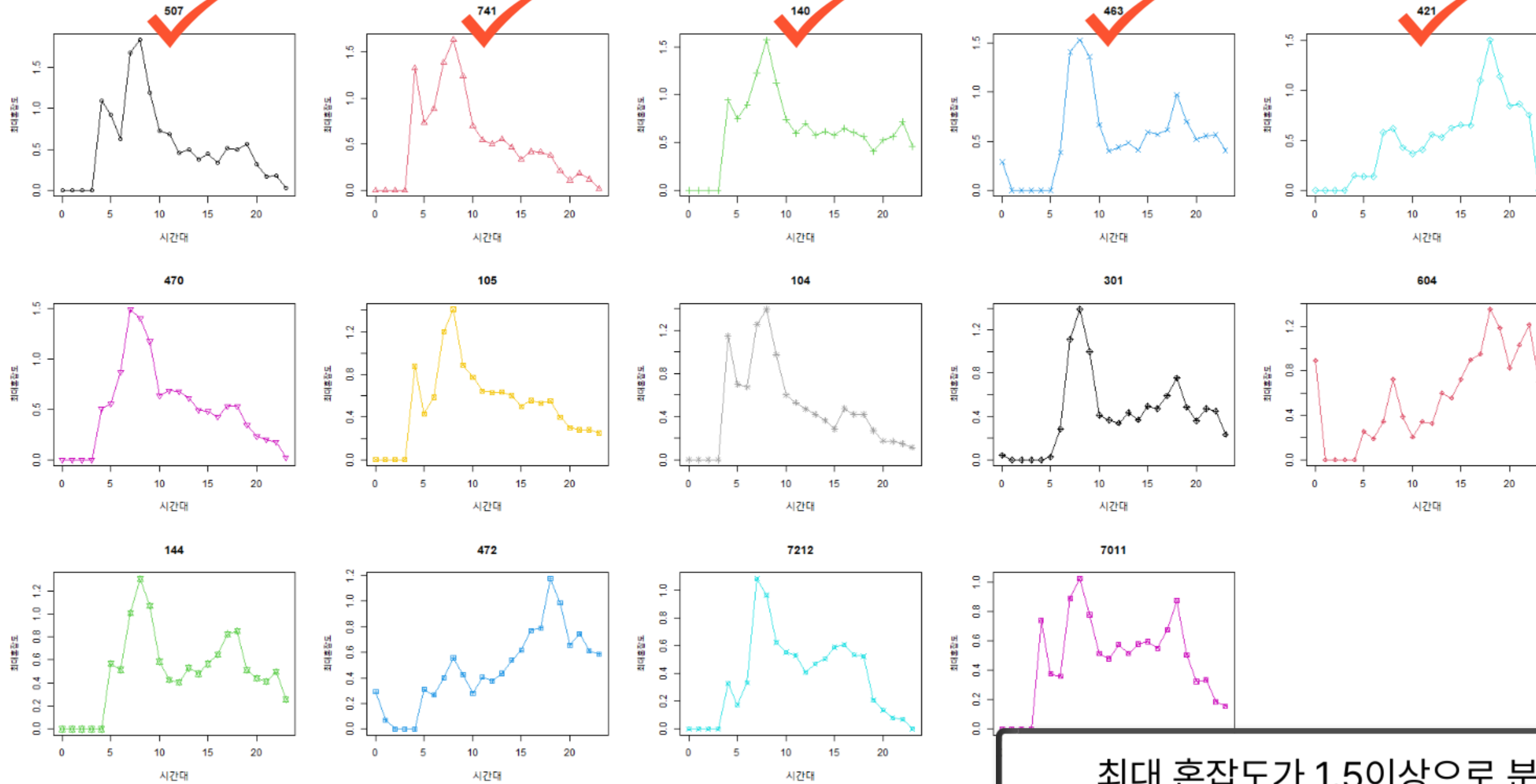
head(desc_crowd[,1])

write.csv(desc_crowd,"혼잡도 정렬.csv",fileEncoding="UTF-8",row.names=FALSE)

```

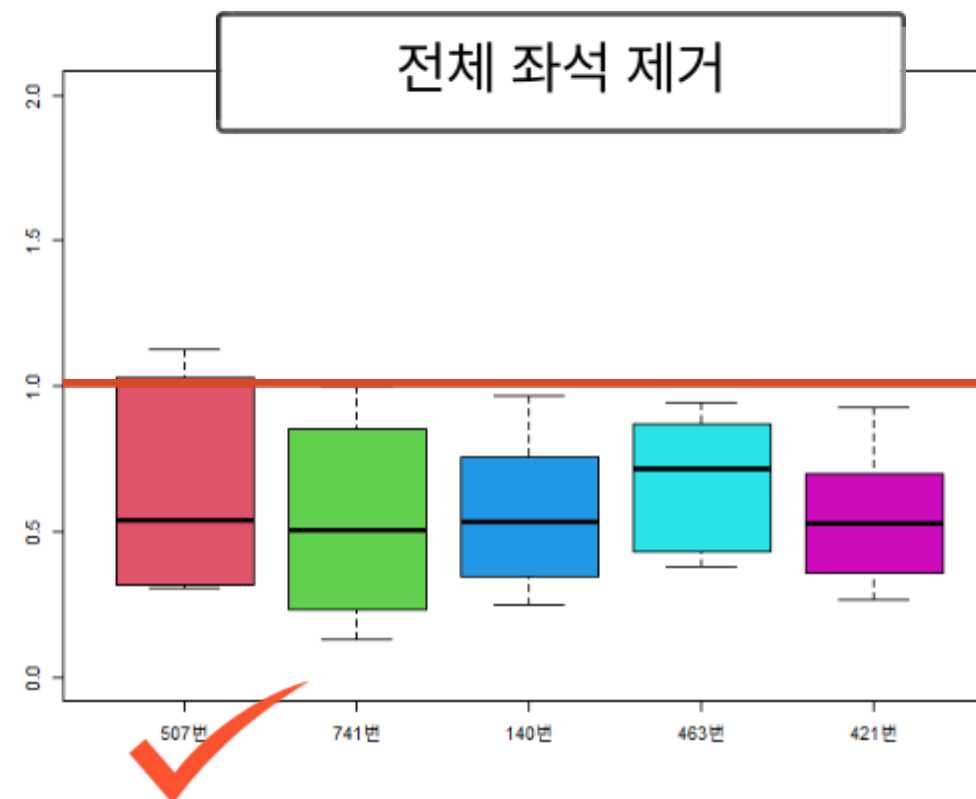
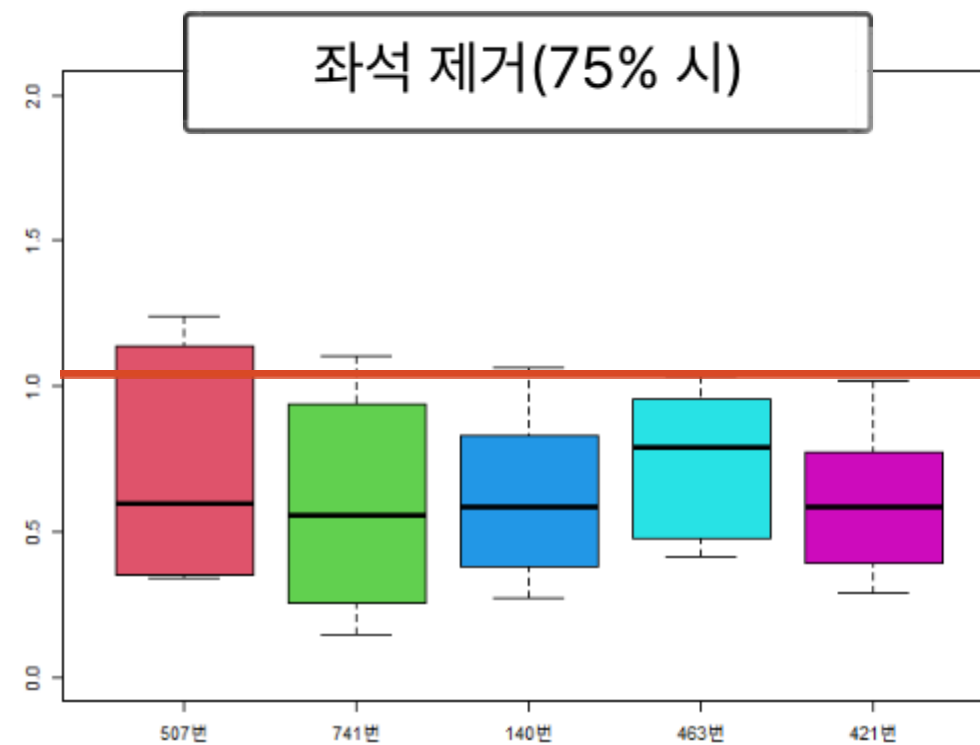
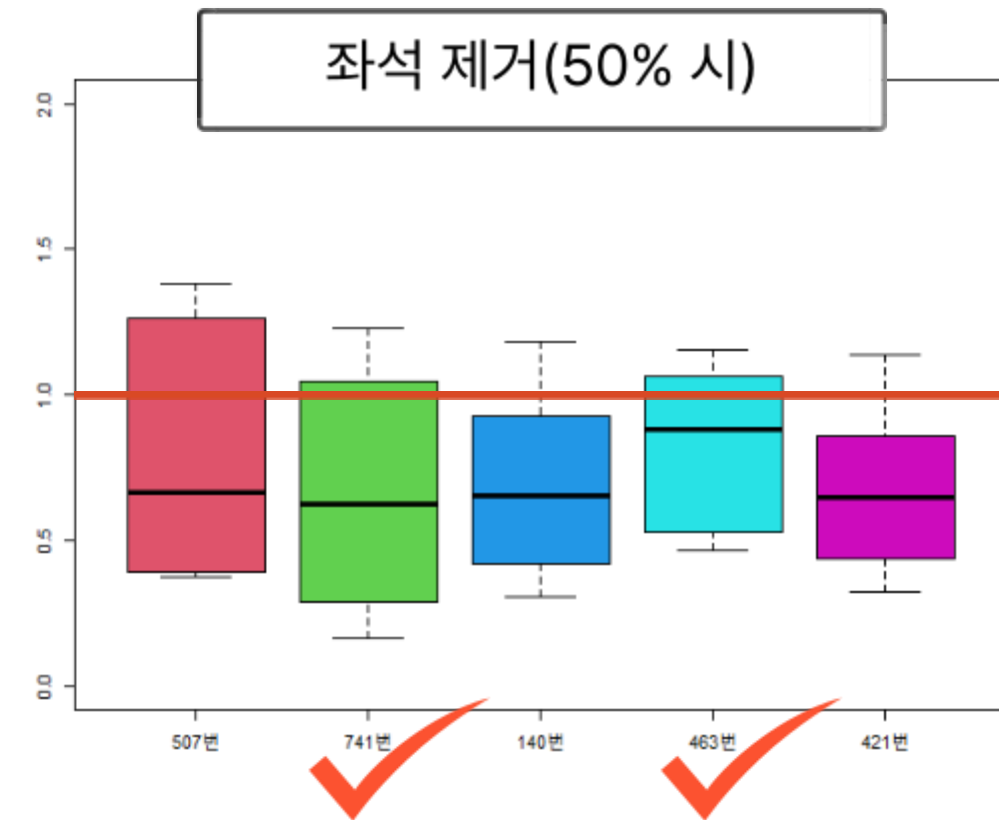
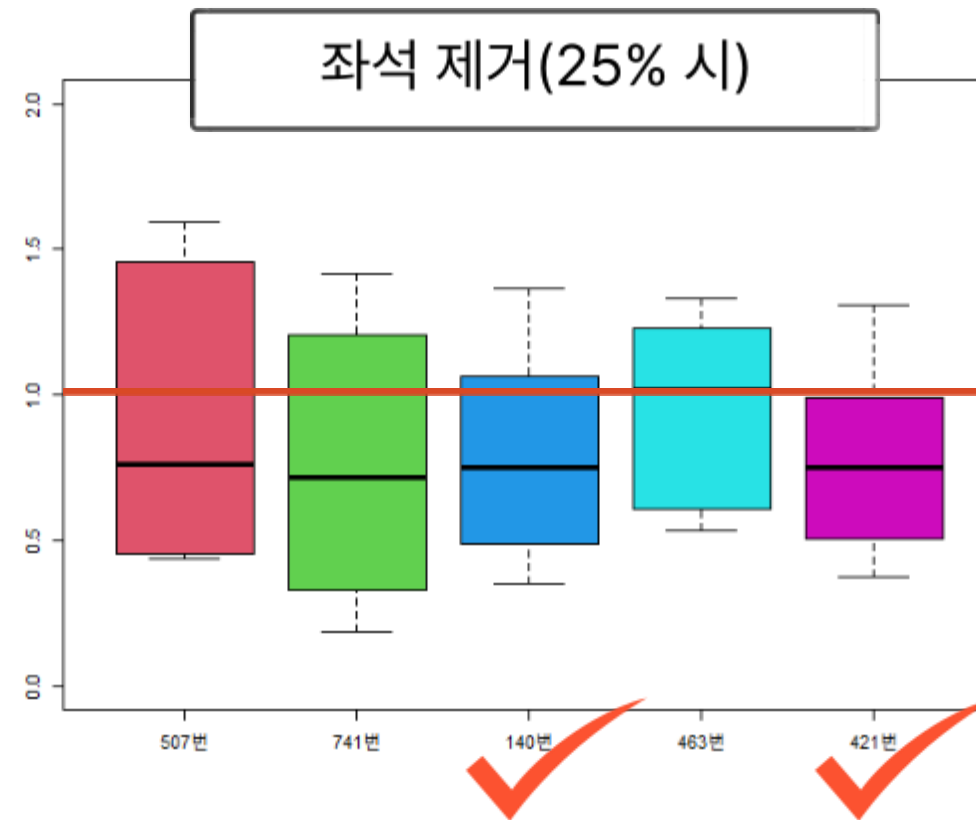
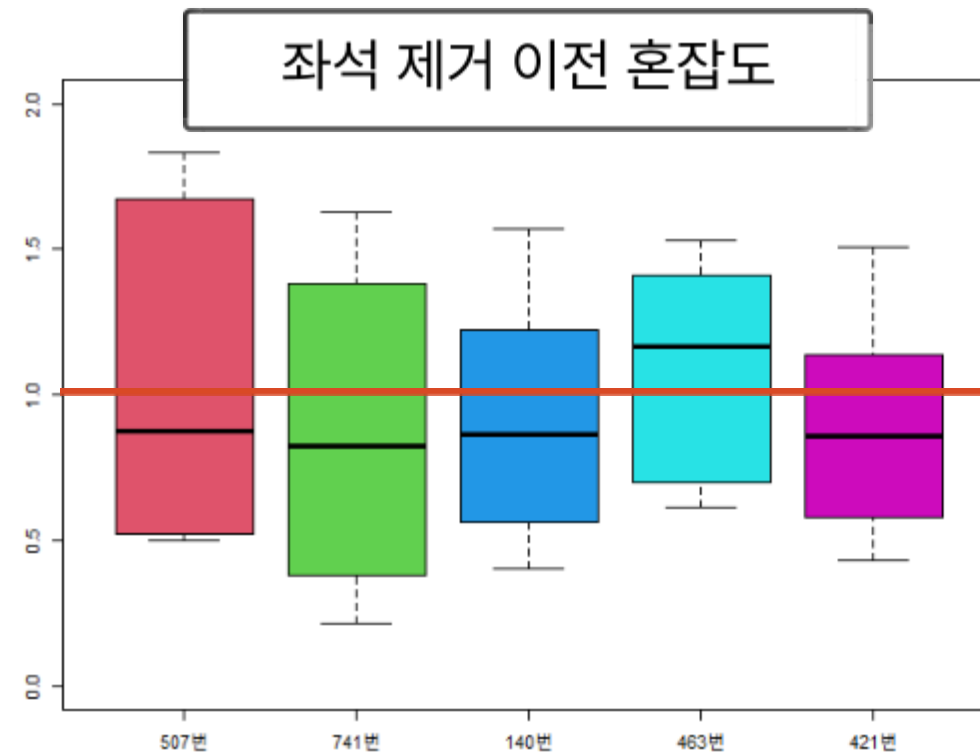
목표 기간동안의 산술 평균 자료를 분석하여
가장 혼잡한 시간대의 정보를 추출한다.

데이터분석 - 분석결과 도출



최대 혼잡도가 1.5이상으로 분석되는 결과를
바탕으로 좌석개선의 효과를 분석

데이터분석 - 분석결과 도출

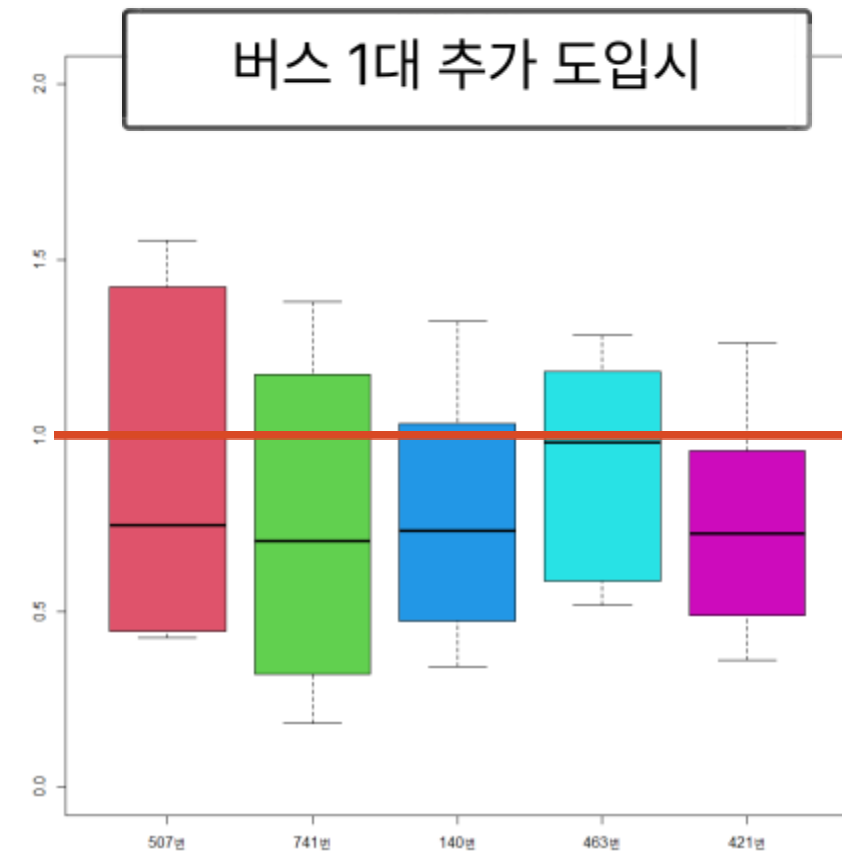
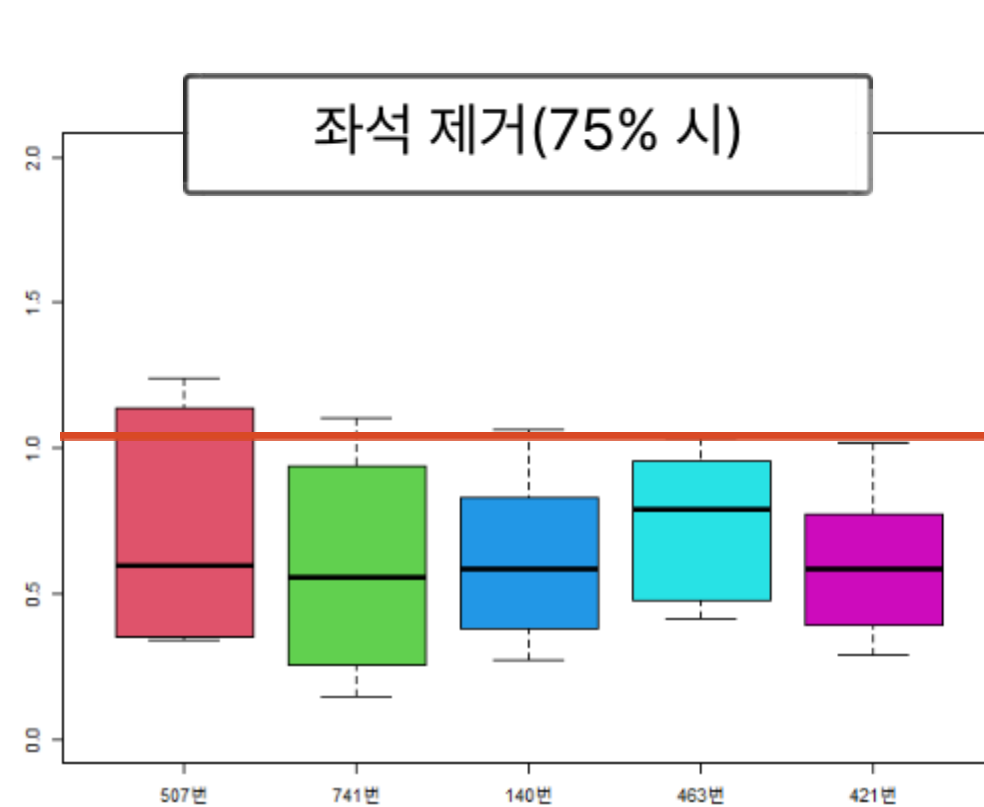
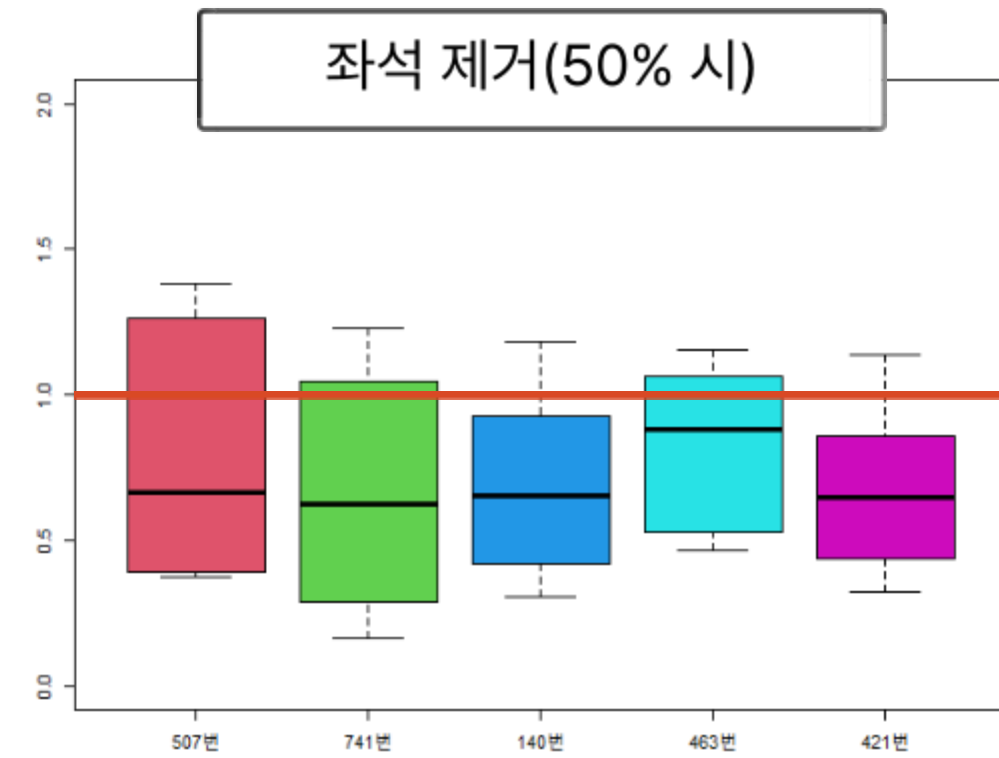
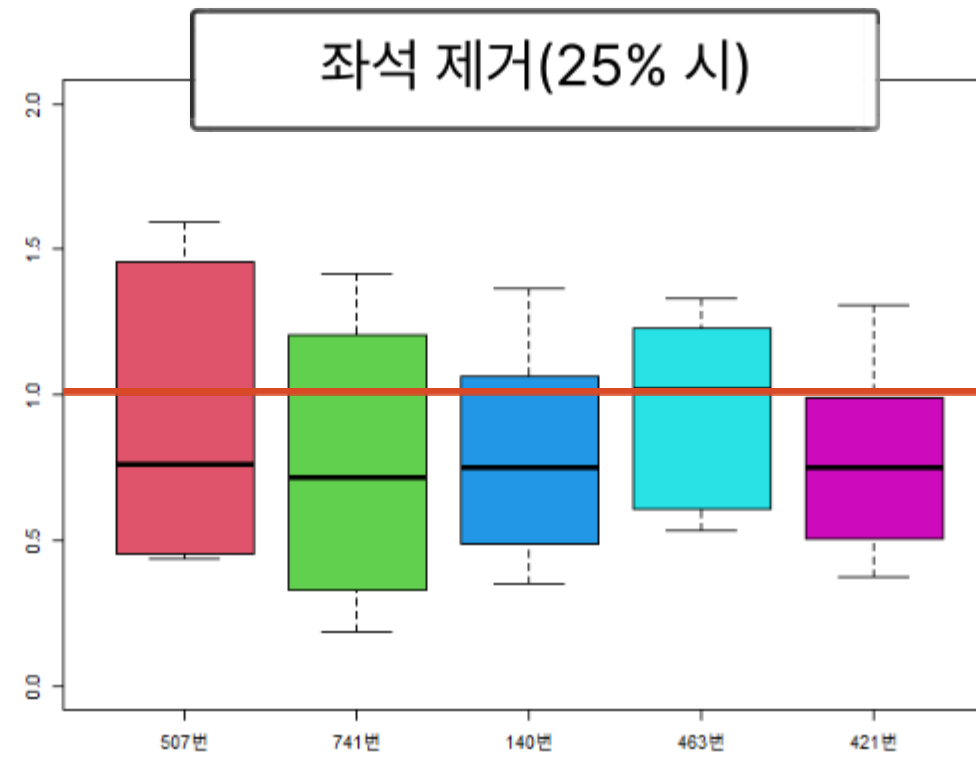
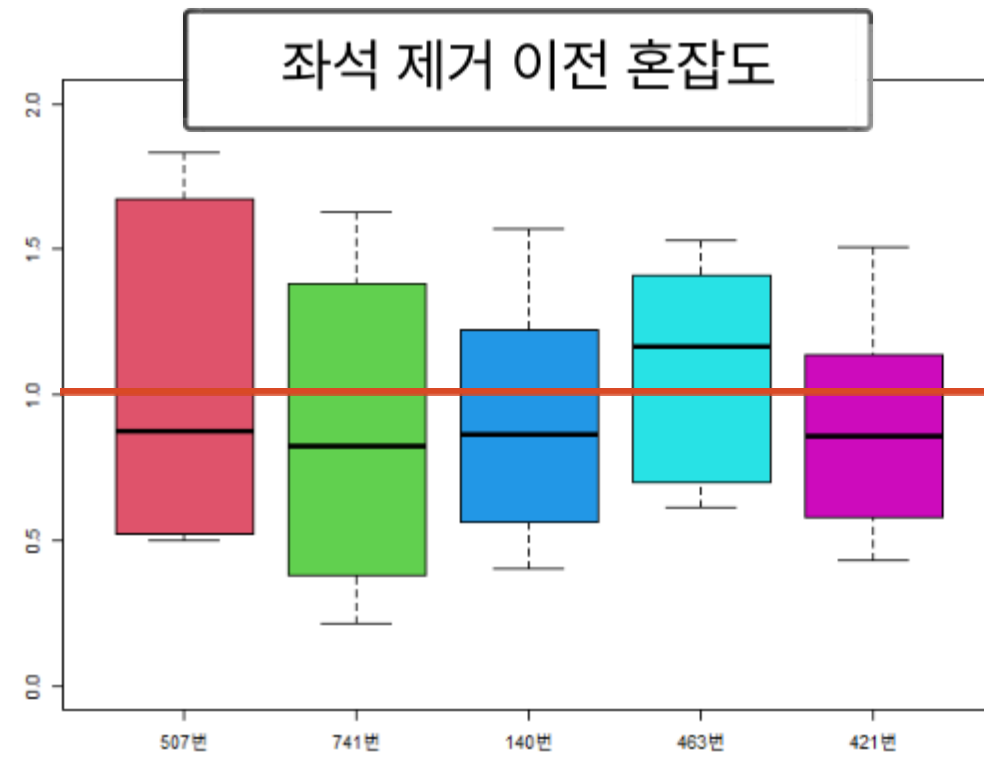


부분적 좌석 개선으로 혼잡도를
꾸준히 1.0 이하로 개선 가능하다.

혼잡도 개선 시 기준인원을
좌석 1개당 1명 → 1개당 2명으로 계산

(서울시 - 좌석개선의 효용성 언론발표용 자료 참고)

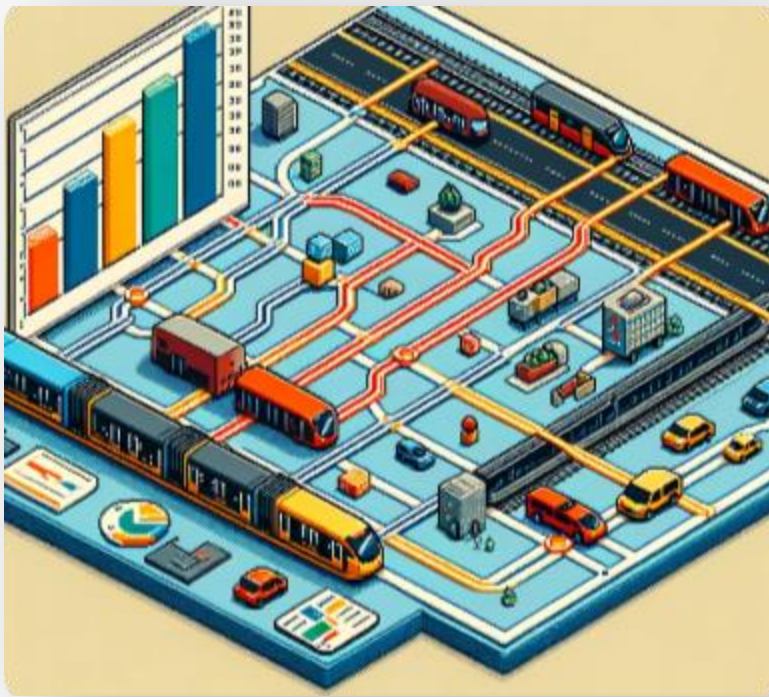
데이터분석 - 분석결과 도출



25%의 좌석 개선은 버스 노선 1대를 추가하는 효과와 유사하다.

비슷한 특성의 버스(출근시간에 바쁜 버스)간 높은 정밀도의 비례관계가 성립한다.

결론 및 기대효과



1. 서술한 비례관계를 향후 추가 배차가 필요한 노선에 활용 가능



2. 서술한 비례관계를 통해 경제적인 감차 가능



3. IoT 기술로 유동적인 좌석 조정을 통한 혼잡도 조정 가능
(배차 조정보다 유도리 있는 대응)

한계점 및 대응 방안



1. 좌석 제거 이후 운행 시의 안전 문제
→
천장 손잡이의 적극적 활용으로 대응



2. 특수좌석 문제
→
일부 일반 좌석만 개선



3. 기상 및 도로 문제로 유동적인 버스 수요
→
IoT기술을 접목한 실시간 대응 가능성

Introduction to Data Programming Mid Project

**THANK YOU
FOR WATCHING!**