


3. 데이터 가공




(예시 2) 서울시 공유자전거 수요 데이터 (2020년2월)

<https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>

 UC Irvine
Machine Learning
Repository

Datasets Contribute Dataset About Us

 **Seoul Bike Sharing Demand**
Donated on 2/29/2020

The dataset contains count of public bicycles rented per hour in the Seoul Bike Sharing System, with corresponding weather data and holiday information

| | | |
|--|---------------------------------|---------------------------------------|
| Dataset Characteristics Multivariate | Subject Area Business | Associated Tasks Regression |
| Feature Type Integer, Real | # Instances 8760 | # Features 13 |

서울시 공유자전거 대여 데이터, 날씨와 휴일 정보 포함

n=8760(2017.12.1-2018.11.30, 1년 데이터), 13개 변수,

관련 분석기법 : 회귀분석

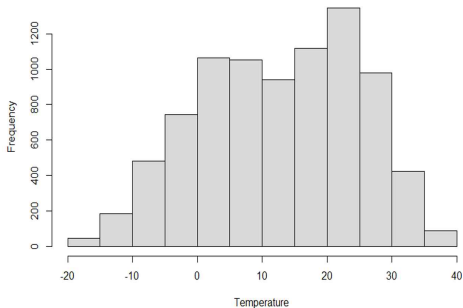


3. 데이터 가공 : 공공데이터

(예시 2) 서울시 공유자전거 수요 데이터 (2020년2월)



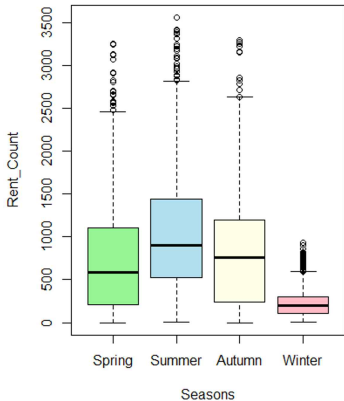
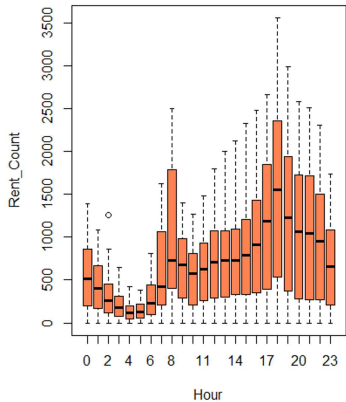
<https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>



Date (날자) : year-month-day
Rented Bike count (대여자전거수)
Hour (시간) – Hour of the day
Temperature (온도)-Temperature in Celsius
Humidity (습도) - %
Windspeed (풍속) – m/s
Visibility (가시도) – 10m
Dew point temperature (이슬점) – Celsius
Solar radiation – MJ/m2
Rainfall (강수량) – mm
Snowfall (적설량) - cm
Seasons (계절) - Winter, Spring, Summer, Autumn
Holiday (공휴일/평일) - Holiday/No holiday

3. 데이터 가공 : 공공데이터

(예시 2) 서울시 공유자전거 수요 데이터 (2020년2월)



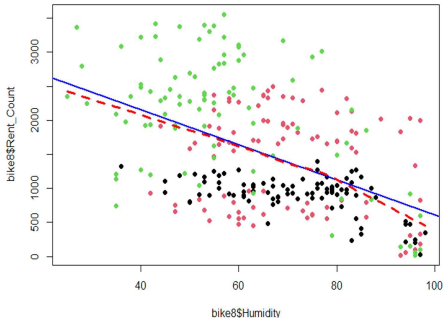
Question!

1. 어느 시간대 수요가 가장 많은가?
2. 자전거 대여 수요에 계절 영향이 있나?

3. 데이터 가공 : 공공데이터

(예시 2) 서울시 공유자전거 수요 데이터 (2020년2월)

Summer, 0시, 8시, 18시 선택,
검정색 : 0시, 빨간색 : 8시, 초록색 : 18시



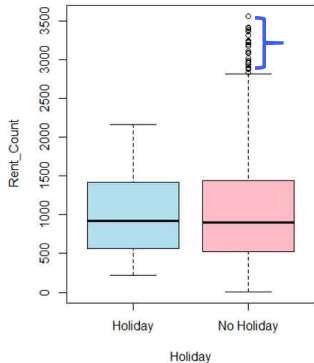
여름 : 습도와 자전거 대여수의 관계

Question!

3. 여름의 습도와 대여수와의 관계는?

3. 데이터 가공 : 공공데이터

(예시 2) 서울시 공유자전거 수요 데이터 (2020년2월)



- (1) 자전거 대여 수요량에 주말/주중 영향이 있을까요?
- (2) Holiday 공휴일에 주말이 포함되었나?
- (3) 데이터를 생산하는 기관이라면 이 변수를 어떻게 작성하는게 좋을까요?

Holiday (공휴일/평일) - Holiday/No holiday



Holiday (공휴일/평일) - Holiday/No holiday/Weekend