상점

## 라이선스

자료의 라이선스에 따라, 이 문서에는 특별히 [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/) 라이선스가 적용됩니다.

## 자료

* [Supermarket store branches sales analysis](https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data)
* [Big Mart Sales Prediction Datasets](https://www.kaggle.com/datasets/shivan118/big-mart-sales-prediction-datasets)
* [Coles Supermarket Sales](https://www.kaggle.com/datasets/ankitab18/coles-supermarket-sales)

summary(stores)

## Store.ID Store\_Area Items\_Available Daily\_Customer\_Count  
## Min. : 1.0 Min. : 775 Min. : 932 Min. : 10.0   
## 1st Qu.:224.8 1st Qu.:1317 1st Qu.:1576 1st Qu.: 600.0   
## Median :448.5 Median :1477 Median :1774 Median : 780.0   
## Mean :448.5 Mean :1485 Mean :1782 Mean : 786.4   
## 3rd Qu.:672.2 3rd Qu.:1654 3rd Qu.:1983 3rd Qu.: 970.0   
## Max. :896.0 Max. :2229 Max. :2667 Max. :1560.0   
## Store\_Sales   
## Min. : 14920   
## 1st Qu.: 46530   
## Median : 58605   
## Mean : 59351   
## 3rd Qu.: 71872   
## Max. :116320

summary(big\_mart)

## Item\_Identifier Item\_Weight Item\_Fat\_Content Item\_Visibility   
## Length:8523 Min. : 4.555 Length:8523 Min. :0.00000   
## Class :character 1st Qu.: 8.774 Class :character 1st Qu.:0.02699   
## Mode :character Median :12.600 Mode :character Median :0.05393   
## Mean :12.858 Mean :0.06613   
## 3rd Qu.:16.850 3rd Qu.:0.09459   
## Max. :21.350 Max. :0.32839   
## NA's :1463   
## Item\_Type Item\_MRP Outlet\_Identifier   
## Length:8523 Min. : 31.29 Length:8523   
## Class :character 1st Qu.: 93.83 Class :character   
## Mode :character Median :143.01 Mode :character   
## Mean :140.99   
## 3rd Qu.:185.64   
## Max. :266.89   
##   
## Outlet\_Establishment\_Year Outlet\_Size Outlet\_Location\_Type  
## Min. :1985 Length:8523 Length:8523   
## 1st Qu.:1987 Class :character Class :character   
## Median :1999 Mode :character Mode :character   
## Mean :1998   
## 3rd Qu.:2004   
## Max. :2009   
##   
## Outlet\_Type Item\_Outlet\_Sales   
## Length:8523 Min. : 33.29   
## Class :character 1st Qu.: 834.25   
## Mode :character Median : 1794.33   
## Mean : 2181.29   
## 3rd Qu.: 3101.30   
## Max. :13086.97   
##

summary(coles\_sales)

## X Coles\_StoreIDNo Expec\_Revenue Gross\_Sale   
## Min. : 1.0 Length:682 Min. :20.00 Min. :25.00   
## 1st Qu.:171.2 Class :character 1st Qu.:28.00 1st Qu.:34.00   
## Median :341.5 Mode :character Median :36.00 Median :42.00   
## Mean :341.5 Mean :35.17 Mean :42.87   
## 3rd Qu.:511.8 3rd Qu.:43.00 3rd Qu.:51.75   
## Max. :682.0 Max. :50.00 Max. :89.00   
##   
## Sales\_Cost Targeted\_Quarter Coles\_Forecast   
## Min. : 2 Length:682 Length:682   
## 1st Qu.: 8 Class :character Class :character   
## Median :10 Mode :character Mode :character   
## Mean :10   
## 3rd Qu.:12   
## Max. :20   
## NA's :15

summary(coles\_store)

## X Coles\_StoreID Store\_Location Customer\_Count   
## Min. : 1.0 Length:682 Length:682 Min. : 6006   
## 1st Qu.:171.2 Class :character Class :character 1st Qu.: 7156   
## Median :341.5 Mode :character Mode :character Median : 8402   
## Mean :341.5 Mean :12648   
## 3rd Qu.:511.8 3rd Qu.: 9593   
## Max. :682.0 Max. :59969   
## NA's :57   
## Staff\_Count Store\_Area   
## Min. :15.00 Min. :1700   
## 1st Qu.:18.00 1st Qu.:1700   
## Median :22.00 Median :1700   
## Mean :22.99 Mean :1918   
## 3rd Qu.:26.00 3rd Qu.:1900   
## Max. :39.00 Max. :2700   
##

## 계획

### 정보

Store\_Area와 Items\_Available, Daily\_Customer\_Count, Store\_Sales의 관계를 확인합니다.

### 지혜

자료에 있는 상점의 면적이 그 상점의 품목 개수나 일간 방문객 수, 월 매출과 일정한 관계를 보이는지 확인합니다.

### 예상 효과

* 상점 창업 부지 면적을 합리적으로 결정할 수 있습니다.
* 상점이 대체로 얼마나 빽빽하게 상품을 진열하는지 알 수 있습니다.
* 이상하지 않은 범위에서 좋은(단위면적 당 임대료 비쌈) 입지의 좁은(면적) 상점과 나쁜(단위면적 당 임대료 쌈) 입지의 넓은(면접) 상점 중 더 효과적인 창업 전략을 선택할 수 있습니다.

### 2학기에 추가하는 내용

* 새 자료 추가
* 자료 설명 축소
* 예상 효과 개선
* 전처리 개선
* 새 시각화 추가
* 결과 개선

## 실행

### 전처리

* stores
  + 시각화 속도를 개선하기 위해 미리 Store\_Area를 기준으로 오름차순 정렬합니다.
  + 수치와 무관한 식별자 Store.ID를 제거합니다.
* big\_mart
  + Outlet\_Size를 알 수 없는 행을 제거합니다.
  + Outlet\_Size를 수치화합니다.
  + Item\_Outlet\_Sales를 상점마다 집계합니다.
  + stores와 대응하는 열만 사용하고 열 이름을 일치시킵니다.
  + 시각화 속도를 개선하기 위해 미리 Store\_Area를 기준으로 오름차순 정렬합니다.
* coles
  + coles\_sales와 coles\_store를 병합합니다.
  + stores와 대응하는 열만 사용하고 열 이름을 일치시킵니다.
  + 시각화 속도를 개선하기 위해 미리 Store\_Area를 기준으로 오름차순 정렬합니다.
* 새로운 파일로 저장합니다.

stores <- stores[order(stores$Store\_Area), ]  
stores$Store.ID <- NULL  
  
write.csv(stores, "preprocessed-fall-semester.csv")  
  
big\_mart$Outlet\_Size <-  
 as.numeric(  
 factor(  
 big\_mart$Outlet\_Size,  
 levels = c("Small", "Medium", "High"), labels = c(0, 1, 2)  
 )  
 )  
  
big\_mart <- big\_mart[!is.na(big\_mart$Outlet\_Size), ]  
  
big\_mart\_ <- aggregate(  
 x = big\_mart$Item\_Outlet\_Sales,  
 by = list(big\_mart$Outlet\_Identifier),  
 FUN = sum  
)  
  
colnames(big\_mart\_) <- c("Outlet\_Identifier", "Store\_Sales")  
  
big\_mart <- merge(big\_mart\_, big\_mart, by = "Outlet\_Identifier")  
  
big\_mart <- big\_mart[, c("Outlet\_Identifier", "Outlet\_Size", "Store\_Sales")]  
  
big\_mart <- big\_mart[!duplicated(big\_mart$Outlet\_Identifier), ]  
  
colnames(big\_mart) <- c("Store.ID", "Store\_Area", "Store\_Sales")  
  
big\_mart$Store.ID <- NULL  
  
big\_mart <- big\_mart[order(big\_mart$Store\_Area), ]  
  
write.csv(big\_mart, "preprocessed-fall-semester-big-mart.csv")  
  
coles\_store$Coles\_StoreIDNo <- coles\_store$Coles\_StoreID  
  
coles <- merge(coles\_sales, coles\_store, by = "Coles\_StoreIDNo")  
  
coles <- coles[, c("Store\_Area", "Gross\_Sale")]  
  
colnames(coles) <- c("Store\_Area", "Store\_Sales")  
  
big\_mart <- big\_mart[order(big\_mart$Store\_Area), ]  
  
write.csv(coles, "preprocessed-fall-semester-coles.csv")

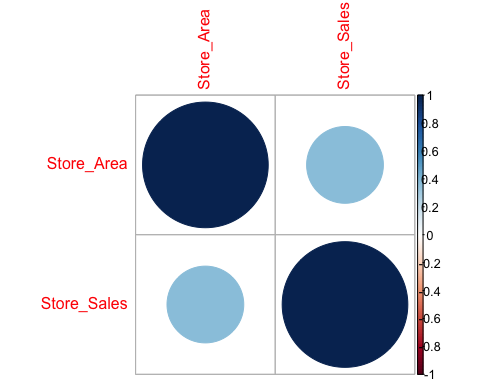
### 시각화

#### 상관관계

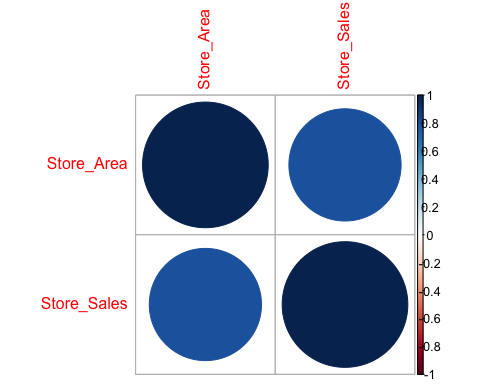
corrplot(cor(stores))



corrplot(cor(big\_mart))

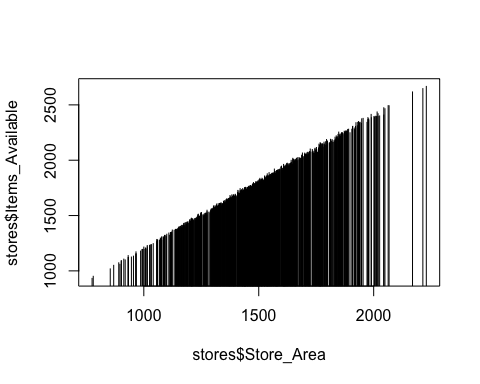


corrplot(cor(coles))



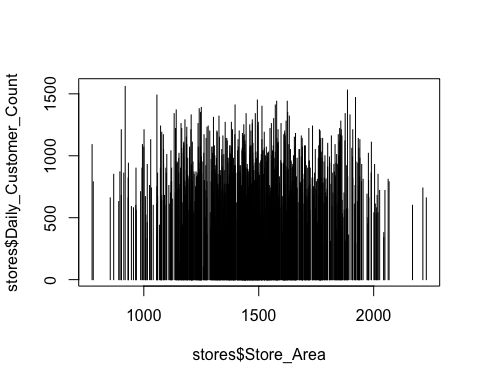
#### 면적에 따른 품목 개수

plot(stores$Store\_Area, stores$Items\_Available, type = "h")



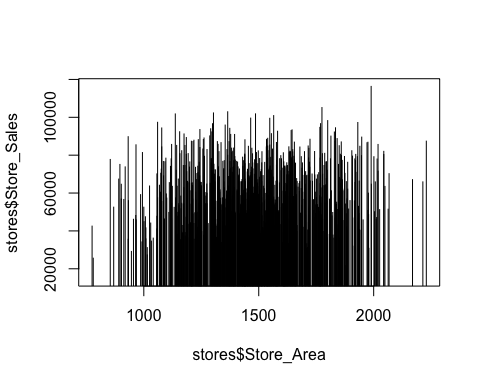
#### 면적에 따른 일간 방문객 수

plot(stores$Store\_Area, stores$Daily\_Customer\_Count, type = "h")

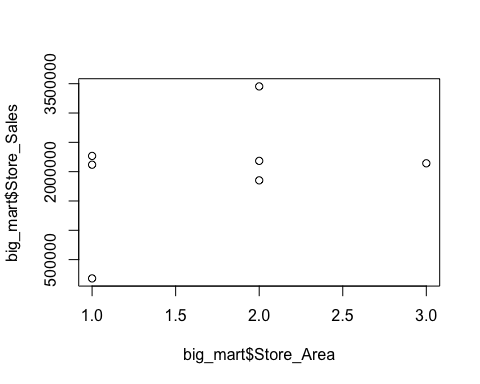


#### 면적에 따른 (월 등) 매출

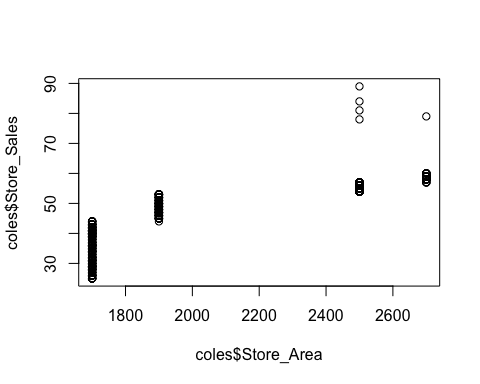
plot(stores$Store\_Area, stores$Store\_Sales, type = "h")



plot(big\_mart$Store\_Area, big\_mart$Store\_Sales)

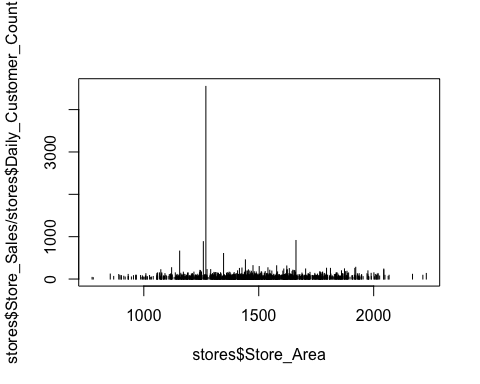


plot(coles$Store\_Area, coles$Store\_Sales)



#### 면적에 따른 일간 방문객 당 매출

plot(  
 stores$Store\_Area,  
 stores$Store\_Sales / stores$Daily\_Customer\_Count,  
 type = "h"  
)



## 결과

### 설명

* 자료에 있는 상점의 면적에 대한 품목 개수는 제곱미터 당 약 1.3개로 거의 정비례합니다.
* 하지만 면적(또는 품목 개수)과 일간 방문객 수는 상관관계가 거의 없습니다.
* 면적(또는 품목 개수)과 월 매출은 미묘한 상관관계가 있습니다.
  + 여전히 유의미하지 않지만 면적과 일간 방문객 수의 상관관계보다 높습니다.
* 덧붙이자면, 일간 방문객 수와 월 매출은 전혀 상관관계가 없습니다.

### 분석

* 상점이 넓으면 확실히 더 많은 상품을 판매할 수 있습니다.
* 하지만 의외로, 더 넓거나 판매 상품 개수가 더 많다고 반드시 고객이나 매출이 더 많지는 않습니다.
  + 새로운 자료는 원래 자료보다는 넓이와 매출 간의 상관관계가 훨씬 더 높습니다.
* 넓이 등과 무관하게 방문객이 많더라도 매출은 전혀 상응하지 않습니다.
* 가장 큰 수준의 상점은 매출로 중간 크기 상점의 평균과 비슷합니다.
* 방문객 당 매출이 중요할 수 있습니다.

### 개선 방안

* 면적과 매출의 상관관계를 더 자세히 분석하고 예측해볼 수 있습니다.