1조 조별과제

최재혁,오현지,최호경,오인혜

2020 5 29

# data 읽기

library( rJava )   
library( xlsx )   
  
setwd("C:\\Workspace\\WorkspaceR")  
Hot.xlsx <- read.xlsx( file = "전국Hot\_Place.xlsx", # 파일 저장 경로 설정   
 sheetIndex = 1,  
 encoding = "UTF-8" )  
Hot.xlsx

## Hot\_Place 유동인구 미용업 병원 숙박업 영화관 대규모점포  
## 1 대전광역시\_서구\_둔산동 355734 264 21 25 1 8  
## 2 서울특별시\_강남구\_역삼동 758043 495 6 70 4 5  
## 3 부산광역시\_부산진구\_부전동 822972 386 6 148 3 20  
## 4 인천광역시\_부평구\_부평동 500347 643 12 142 3 13  
## 5 대구광역시\_수성구\_범어동 293824 243 4 8 1 3  
## 6 광주광역시\_서구\_치평동 254355 173 19 51 3 7  
## 7 울산광역시\_남구\_삼산동 404668 243 6 74 2 8  
## 8 제주특별자치도\_제주시\_노형동 337848 280 0 30 1 4  
## 일반음식점 노래방 PC방 제과점  
## 1 1102 144 13 45  
## 2 2926 154 52 76  
## 3 1754 145 26 33  
## 4 1793 174 45 48  
## 5 814 55 15 31  
## 6 769 51 40 27  
## 7 855 106 40 32  
## 8 853 39 24 55

# data 구조 확인

class(Hot.xlsx) # data.frame

## [1] "data.frame"

str(Hot.xlsx) # 8 obs, 11 variables

## 'data.frame': 8 obs. of 11 variables:  
## $ Hot\_Place : chr "대전광역시\_서구\_둔산동" "서울특별시\_강남구\_역삼동" "부산광역시\_부산진구\_부전동" "인천광역시\_부평구\_부평동" ...  
## $ 유동인구 : num 355734 758043 822972 500347 293824 ...  
## $ 미용업 : num 264 495 386 643 243 173 243 280  
## $ 병원 : num 21 6 6 12 4 19 6 0  
## $ 숙박업 : num 25 70 148 142 8 51 74 30  
## $ 영화관 : num 1 4 3 3 1 3 2 1  
## $ 대규모점포: num 8 5 20 13 3 7 8 4  
## $ 일반음식점: num 1102 2926 1754 1793 814 ...  
## $ 노래방 : num 144 154 145 174 55 51 106 39  
## $ PC방 : num 13 52 26 45 15 40 40 24  
## $ 제과점 : num 45 76 33 48 31 27 32 55

dim(Hot.xlsx)

## [1] 8 11

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# 결측치, 특이값 확인

sum(is.na( Hot.xlsx ) ) # 결측치 : 0

## [1] 0

boxplot.stats( Hot.xlsx$유동인구 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$PC방 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$미용업 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$병원 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$숙박업 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$영화관 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$대규모점포 )$out # 특이값 1개 : 20

## [1] 20

boxplot.stats( Hot.xlsx$일반음식점 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$노래방 )$out

## numeric(0)

boxplot.stats( Hot.xlsx$제과점 )$out

## numeric(0)

# 변수별 막대그래프

# 유동인구

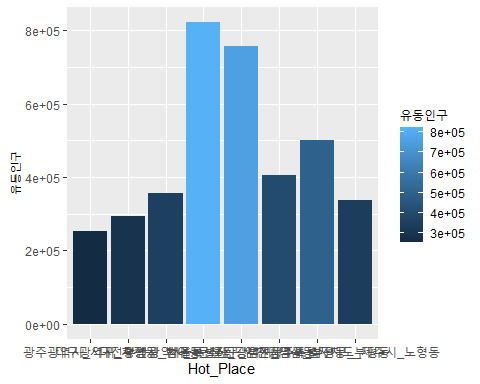
pop <- Hot.xlsx %>% select(Hot\_Place,유동인구)  
pop

## Hot\_Place 유동인구  
## 1 대전광역시\_서구\_둔산동 355734  
## 2 서울특별시\_강남구\_역삼동 758043  
## 3 부산광역시\_부산진구\_부전동 822972  
## 4 인천광역시\_부평구\_부평동 500347  
## 5 대구광역시\_수성구\_범어동 293824  
## 6 광주광역시\_서구\_치평동 254355  
## 7 울산광역시\_남구\_삼산동 404668  
## 8 제주특별자치도\_제주시\_노형동 337848

summary( pop )

## Hot\_Place 유동인구   
## Length:8 Min. :254355   
## Class :character 1st Qu.:326842   
## Mode :character Median :380201   
## Mean :465974   
## 3rd Qu.:564771   
## Max. :822972

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 유동인구,  
 fill = 유동인구)) + geom\_bar(stat = 'identity')



# 미용업

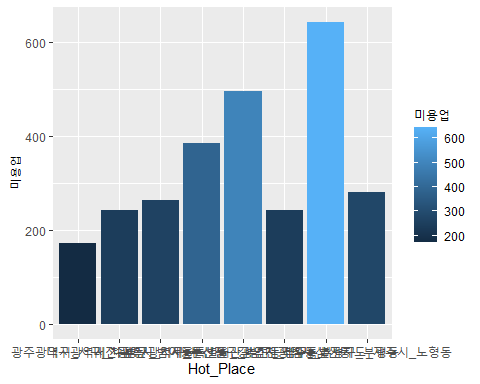
hair <- Hot.xlsx %>% select(Hot\_Place,미용업)  
hair

## Hot\_Place 미용업  
## 1 대전광역시\_서구\_둔산동 264  
## 2 서울특별시\_강남구\_역삼동 495  
## 3 부산광역시\_부산진구\_부전동 386  
## 4 인천광역시\_부평구\_부평동 643  
## 5 대구광역시\_수성구\_범어동 243  
## 6 광주광역시\_서구\_치평동 173  
## 7 울산광역시\_남구\_삼산동 243  
## 8 제주특별자치도\_제주시\_노형동 280

summary( hair )

## Hot\_Place 미용업   
## Length:8 Min. :173.0   
## Class :character 1st Qu.:243.0   
## Mode :character Median :272.0   
## Mean :340.9   
## 3rd Qu.:413.2   
## Max. :643.0

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 미용업,  
 fill = 미용업)) + geom\_bar(stat = 'identity')



# 병원

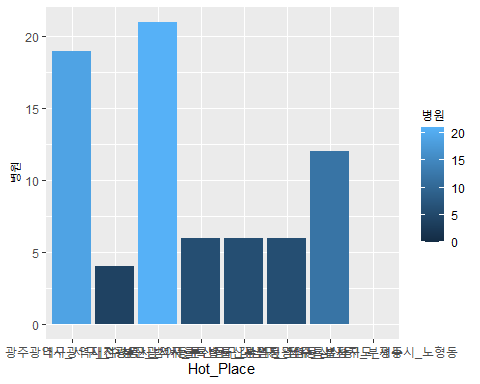
hos <- Hot.xlsx %>% select(Hot\_Place,병원)  
hos

## Hot\_Place 병원  
## 1 대전광역시\_서구\_둔산동 21  
## 2 서울특별시\_강남구\_역삼동 6  
## 3 부산광역시\_부산진구\_부전동 6  
## 4 인천광역시\_부평구\_부평동 12  
## 5 대구광역시\_수성구\_범어동 4  
## 6 광주광역시\_서구\_치평동 19  
## 7 울산광역시\_남구\_삼산동 6  
## 8 제주특별자치도\_제주시\_노형동 0

summary( hos )

## Hot\_Place 병원   
## Length:8 Min. : 0.00   
## Class :character 1st Qu.: 5.50   
## Mode :character Median : 6.00   
## Mean : 9.25   
## 3rd Qu.:13.75   
## Max. :21.00

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 병원,  
 fill = 병원)) + geom\_bar(stat = 'identity')



# 숙박업

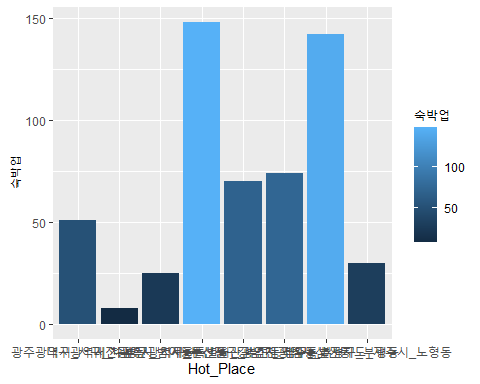
room <- Hot.xlsx %>% select(Hot\_Place,숙박업)  
room

## Hot\_Place 숙박업  
## 1 대전광역시\_서구\_둔산동 25  
## 2 서울특별시\_강남구\_역삼동 70  
## 3 부산광역시\_부산진구\_부전동 148  
## 4 인천광역시\_부평구\_부평동 142  
## 5 대구광역시\_수성구\_범어동 8  
## 6 광주광역시\_서구\_치평동 51  
## 7 울산광역시\_남구\_삼산동 74  
## 8 제주특별자치도\_제주시\_노형동 30

summary( room )

## Hot\_Place 숙박업   
## Length:8 Min. : 8.00   
## Class :character 1st Qu.: 28.75   
## Mode :character Median : 60.50   
## Mean : 68.50   
## 3rd Qu.: 91.00   
## Max. :148.00

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 숙박업,  
 fill = 숙박업)) + geom\_bar(stat = 'identity')



# 영화관

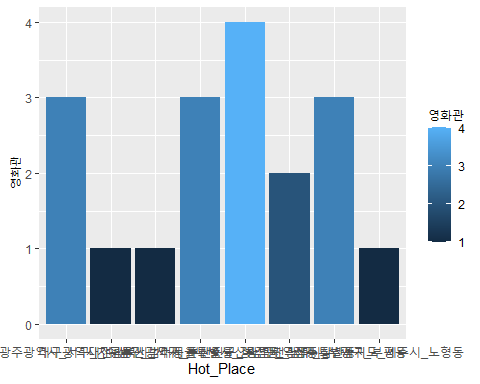
movie <- Hot.xlsx %>% select(Hot\_Place,영화관)  
movie

## Hot\_Place 영화관  
## 1 대전광역시\_서구\_둔산동 1  
## 2 서울특별시\_강남구\_역삼동 4  
## 3 부산광역시\_부산진구\_부전동 3  
## 4 인천광역시\_부평구\_부평동 3  
## 5 대구광역시\_수성구\_범어동 1  
## 6 광주광역시\_서구\_치평동 3  
## 7 울산광역시\_남구\_삼산동 2  
## 8 제주특별자치도\_제주시\_노형동 1

summary( movie )

## Hot\_Place 영화관   
## Length:8 Min. :1.00   
## Class :character 1st Qu.:1.00   
## Mode :character Median :2.50   
## Mean :2.25   
## 3rd Qu.:3.00   
## Max. :4.00

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 영화관,  
 fill = 영화관)) + geom\_bar(stat = 'identity')



# 대규모점포

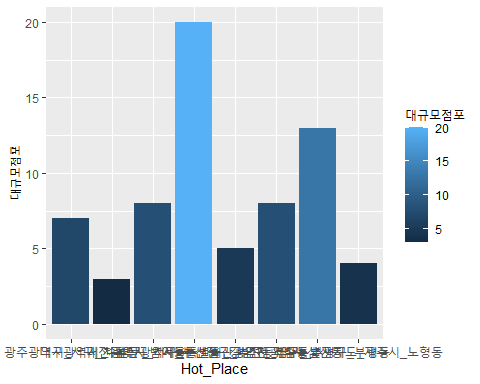
super <- Hot.xlsx %>% select( Hot\_Place, 대규모점포)  
super

## Hot\_Place 대규모점포  
## 1 대전광역시\_서구\_둔산동 8  
## 2 서울특별시\_강남구\_역삼동 5  
## 3 부산광역시\_부산진구\_부전동 20  
## 4 인천광역시\_부평구\_부평동 13  
## 5 대구광역시\_수성구\_범어동 3  
## 6 광주광역시\_서구\_치평동 7  
## 7 울산광역시\_남구\_삼산동 8  
## 8 제주특별자치도\_제주시\_노형동 4

summary( super )

## Hot\_Place 대규모점포   
## Length:8 Min. : 3.00   
## Class :character 1st Qu.: 4.75   
## Mode :character Median : 7.50   
## Mean : 8.50   
## 3rd Qu.: 9.25   
## Max. :20.00

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 대규모점포,  
 fill = 대규모점포)) + geom\_bar(stat = 'identity')



# 일반음식점

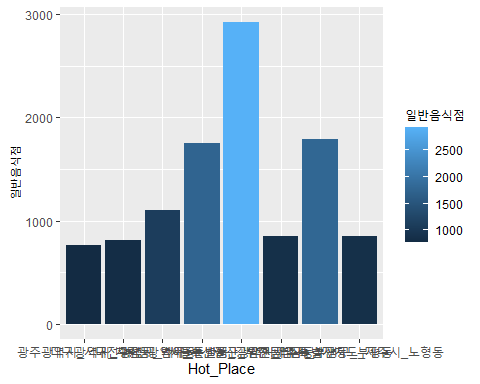
food <- Hot.xlsx %>% select( Hot\_Place, 일반음식점)  
food

## Hot\_Place 일반음식점  
## 1 대전광역시\_서구\_둔산동 1102  
## 2 서울특별시\_강남구\_역삼동 2926  
## 3 부산광역시\_부산진구\_부전동 1754  
## 4 인천광역시\_부평구\_부평동 1793  
## 5 대구광역시\_수성구\_범어동 814  
## 6 광주광역시\_서구\_치평동 769  
## 7 울산광역시\_남구\_삼산동 855  
## 8 제주특별자치도\_제주시\_노형동 853

summary( food )

## Hot\_Place 일반음식점   
## Length:8 Min. : 769.0   
## Class :character 1st Qu.: 843.2   
## Mode :character Median : 978.5   
## Mean :1358.2   
## 3rd Qu.:1763.8   
## Max. :2926.0

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 일반음식점,  
 fill = 일반음식점)) + geom\_bar(stat = 'identity')



# 노래방

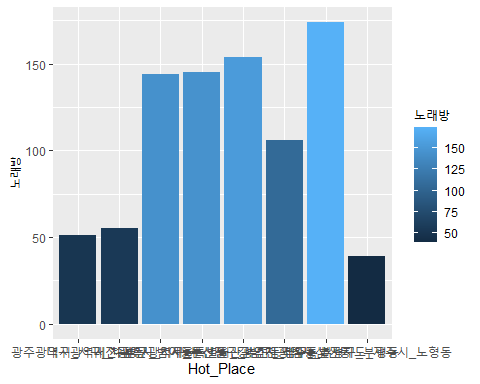
song <- Hot.xlsx %>% select( Hot\_Place, 노래방)  
song

## Hot\_Place 노래방  
## 1 대전광역시\_서구\_둔산동 144  
## 2 서울특별시\_강남구\_역삼동 154  
## 3 부산광역시\_부산진구\_부전동 145  
## 4 인천광역시\_부평구\_부평동 174  
## 5 대구광역시\_수성구\_범어동 55  
## 6 광주광역시\_서구\_치평동 51  
## 7 울산광역시\_남구\_삼산동 106  
## 8 제주특별자치도\_제주시\_노형동 39

summary( song )

## Hot\_Place 노래방   
## Length:8 Min. : 39.0   
## Class :character 1st Qu.: 54.0   
## Mode :character Median :125.0   
## Mean :108.5   
## 3rd Qu.:147.2   
## Max. :174.0

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 노래방,  
 fill = 노래방)) + geom\_bar(stat = 'identity')



# PC방

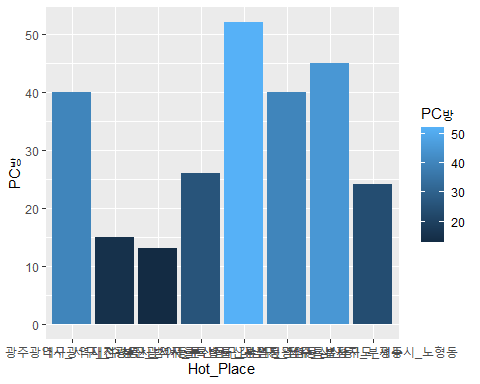
pc <- Hot.xlsx %>% select( Hot\_Place, PC방)  
pc

## Hot\_Place PC방  
## 1 대전광역시\_서구\_둔산동 13  
## 2 서울특별시\_강남구\_역삼동 52  
## 3 부산광역시\_부산진구\_부전동 26  
## 4 인천광역시\_부평구\_부평동 45  
## 5 대구광역시\_수성구\_범어동 15  
## 6 광주광역시\_서구\_치평동 40  
## 7 울산광역시\_남구\_삼산동 40  
## 8 제주특별자치도\_제주시\_노형동 24

summary( pc )

## Hot\_Place PC방   
## Length:8 Min. :13.00   
## Class :character 1st Qu.:21.75   
## Mode :character Median :33.00   
## Mean :31.88   
## 3rd Qu.:41.25   
## Max. :52.00

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = PC방,  
 fill = PC방)) + geom\_bar(stat = 'identity')



# 제과점

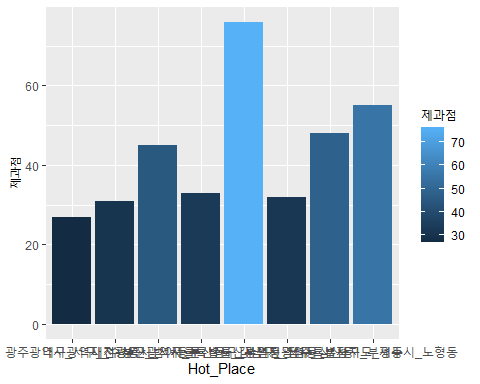
bread <- Hot.xlsx %>% select( Hot\_Place, 제과점)  
bread

## Hot\_Place 제과점  
## 1 대전광역시\_서구\_둔산동 45  
## 2 서울특별시\_강남구\_역삼동 76  
## 3 부산광역시\_부산진구\_부전동 33  
## 4 인천광역시\_부평구\_부평동 48  
## 5 대구광역시\_수성구\_범어동 31  
## 6 광주광역시\_서구\_치평동 27  
## 7 울산광역시\_남구\_삼산동 32  
## 8 제주특별자치도\_제주시\_노형동 55

summary( bread )

## Hot\_Place 제과점   
## Length:8 Min. :27.00   
## Class :character 1st Qu.:31.75   
## Mode :character Median :39.00   
## Mean :43.38   
## 3rd Qu.:49.75   
## Max. :76.00

ggplot(data = Hot.xlsx, aes(x = Hot\_Place, y = 제과점,  
 fill = 제과점)) + geom\_bar(stat = 'identity')

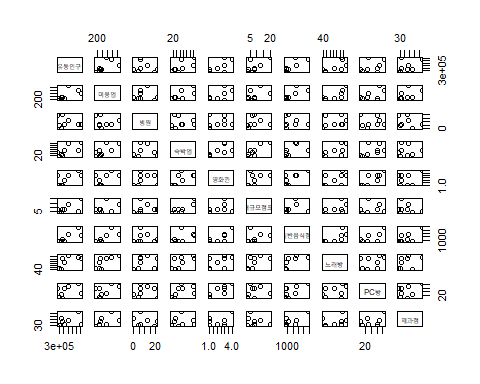


# 다중산점도

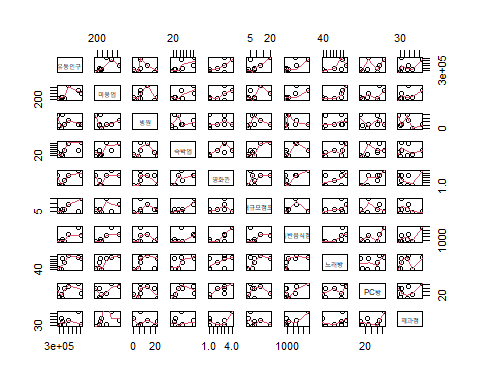
group <- Hot.xlsx[,c("유동인구","미용업","병원","숙박업",  
 "영화관","대규모점포","일반음식점",  
 "노래방","PC방","제과점")]   
group

## 유동인구 미용업 병원 숙박업 영화관 대규모점포 일반음식점 노래방 PC방 제과점  
## 1 355734 264 21 25 1 8 1102 144 13 45  
## 2 758043 495 6 70 4 5 2926 154 52 76  
## 3 822972 386 6 148 3 20 1754 145 26 33  
## 4 500347 643 12 142 3 13 1793 174 45 48  
## 5 293824 243 4 8 1 3 814 55 15 31  
## 6 254355 173 19 51 3 7 769 51 40 27  
## 7 404668 243 6 74 2 8 855 106 40 32  
## 8 337848 280 0 30 1 4 853 39 24 55

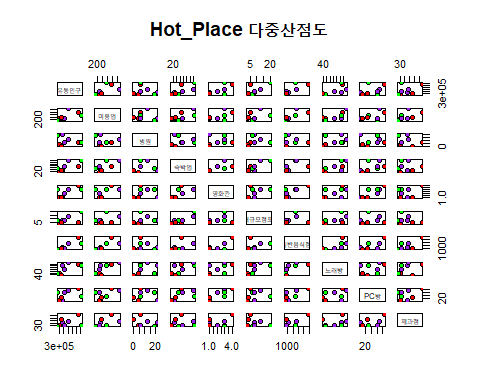
pairs( group[ , ] )



pairs(group, panel=panel.smooth)



pairs(~유동인구+미용업+병원+숙박업+영화관+대규모점포+일반음식점+노래방+PC방+제과점  
 , main = "Hot\_Place 다중산점도", pch = 21, bg = c("Purple","Red","green" ), data = group)



# 상관관계 계수 구하기( 그룹변수 )

library(corrplot)

## corrplot 0.84 loaded

group.cor <- cor(group)  
  
group.cor

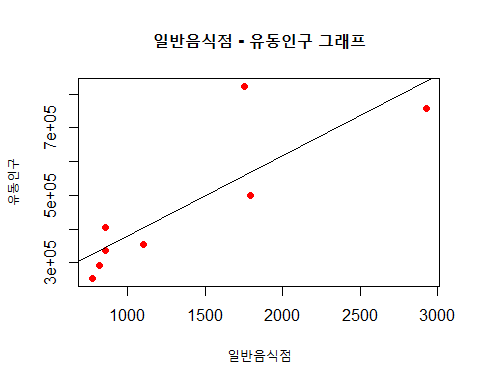
## 유동인구 미용업 병원 숙박업 영화관  
## 유동인구 1.0000000 0.6187029 -0.27749608 0.700991294 0.67285454  
## 미용업 0.6187029 1.0000000 -0.11741994 0.694455140 0.56249476  
## 병원 -0.2774961 -0.1174199 1.00000000 -0.032065341 0.09088426  
## 숙박업 0.7009913 0.6944551 -0.03206534 1.000000000 0.65974076  
## 영화관 0.6728545 0.5624948 0.09088426 0.659740757 1.00000000  
## 대규모점포 0.6176825 0.4043432 0.13108203 0.877801131 0.39553276  
## 일반음식점 0.8421826 0.7683398 -0.12129270 0.493814158 0.75760705  
## 노래방 0.6886641 0.7649985 0.25308500 0.673334517 0.51213511  
## PC방 0.3718135 0.5183292 -0.02376680 0.489976042 0.84655115  
## 제과점 0.4318298 0.5635435 -0.24156403 0.003086065 0.31598722  
## 대규모점포 일반음식점 노래방 PC방 제과점  
## 유동인구 0.61768250 0.8421826 0.6886641 0.37181353 0.431829791  
## 미용업 0.40434320 0.7683398 0.7649985 0.51832916 0.563543481  
## 병원 0.13108203 -0.1212927 0.2530850 -0.02376680 -0.241564030  
## 숙박업 0.87780113 0.4938142 0.6733345 0.48997604 0.003086065  
## 영화관 0.39553276 0.7576070 0.5121351 0.84655115 0.315987224  
## 대규모점포 1.00000000 0.2550376 0.5906415 0.07567374 -0.261511733  
## 일반음식점 0.25503762 1.0000000 0.7219787 0.57757739 0.745735645  
## 노래방 0.59064149 0.7219787 1.0000000 0.34492016 0.373745869  
## PC방 0.07567374 0.5775774 0.3449202 1.00000000 0.387475956  
## 제과점 -0.26151173 0.7457356 0.3737459 0.38747596 1.000000000

corrplot(group.cor, method="number") # 상관계수 시각화

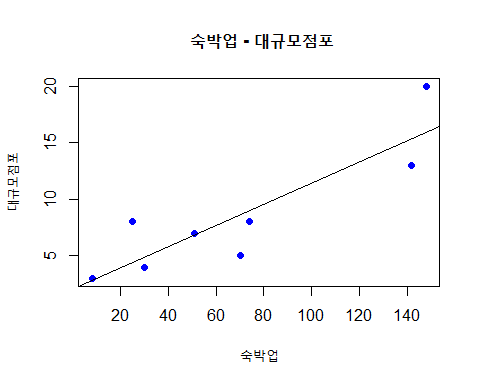


# 상관관계 높은 변수들 간의 상관도( 0.8 이상 )

plot( 유동인구~일반음식점, data = Hot.xlsx, # 그래프 상의 y축의 e+05 = 10^5   
 main = "일반음식점 - 유동인구 그래프",  
 xlab = "일반음식점",  
 ylab = "유동인구",  
 col = "Red",  
 pch = 19)  
  
res <- lm( 유동인구~일반음식점, data = Hot.xlsx)  
abline(res)



plot( 대규모점포~숙박업, data = Hot.xlsx,   
 main = " 숙박업 - 대규모점포",  
 xlab = "숙박업",  
 ylab = "대규모점포",  
 col = "Blue",  
 pch = 19)  
  
res <- lm( 대규모점포~숙박업, data = Hot.xlsx)  
abline(res)



plot( PC방~영화관, data = Hot.xlsx,   
 main = " 영화관 - PC방",  
 xlab = "영화관",  
 ylab = "PC방",  
 col = "Green",  
 pch = 19)  
  
res <- lm( PC방~영화관, data = Hot.xlsx)  
abline(res)

