Crime

1

2022 5 12

# 2020년도 서울시 자치구별 범죄

2020년도 서울시 자치구별로 범죄가 발생한 곳의 시설물 및 시간 등의 요소들을 파악하여  
범죄발생이 어떤 요인에 의해 감소할 수 있는지 알아보기 위해 분석을 하게 되었다.

### 가설설정

1. CCTV가 많은 지역은 범죄율이 낮을 것이다.
2. 가로등이 많은 지역은 범죄율이 낮을 것이다.
3. 안심벨이 많은 지역은 범죄율이 낮을 것이다.
4. 집값이 높은 지역은 범죄율이 낮을 것이다.
5. 소득이 많은 지역은 범죄율이 낮을 것이다.
6. 주요 기관(경찰서, 소방서 등)이 많은 지역은 범죄율이 낮을 것이다.
7. 안심스카우트 이용이 많은 지역은 범죄율이 낮을 것이다.
8. 유동인구가 많은 지역은 범죄율이 낮을 것이다.

### 데이터 수집 경로

[통계지리정보서비스](https://sgis.kostat.go.kr/view/thematicMap/thematicMapMain?stat_thema_map_id=HvIIMFGspL202012070913137699HpnFwqwry&theme=CTGR_005&mapType=06&CTGRS=CTGR_001:recommend,CTGR_002:recommend,CTGR_003:recommend,CTGR_004:recommend,CTGR_005:recommend,CTGR_006:recommend)

[kosis](https://kosis.kr/statHtml/statHtml.do?orgId=110&tblId=DT_156006_001&vw_cd=MT_ZTITLE&list_id=C_11&scrId=&seqNo=&lang_mode=ko&obj_var_id=&itm_id=&conn_path=MT_ZTITLE&path=%252FstatisticsList%252FstatisticsListIndex.do)

[공공데이터 포털](https://www.data.go.kr/data/15029567/fileData.do)

[서울열린데이터 광장](http://data.seoul.go.kr/dataList/datasetList.do)

[경찰청](https://www.police.go.kr/user/search/ND_searchResult.do)

### 패키지 & 파일 불러오기

# 패키지 모음   
library(readxl)  
library(dplyr)  
library(ggplot2)  
library(rgdal)  
library(ggmap)  
library(rgeos)  
library(maptools)  
library(raster)  
library(reshape)  
library(reshape2)  
  
  
# 원본데이터 --------------------------------------------------------------  
crime\_data <- read\_excel("./Project/USE\_DATA/사용할꺼/seoulCrimeStatistic.xls")   
dist\_id <- read.csv("./Project/USE\_DATA/사용할꺼/district\_id.csv")  
map <- readOGR("./Project/USE\_DATA/사용할꺼/TL\_SCCO\_SIG.shp")

## OGR data source with driver: ESRI Shapefile   
## Source: "C:\rrrrrr\Project\USE\_DATA\사용할꺼\TL\_SCCO\_SIG.shp", layer: "TL\_SCCO\_SIG"  
## with 250 features  
## It has 3 fields

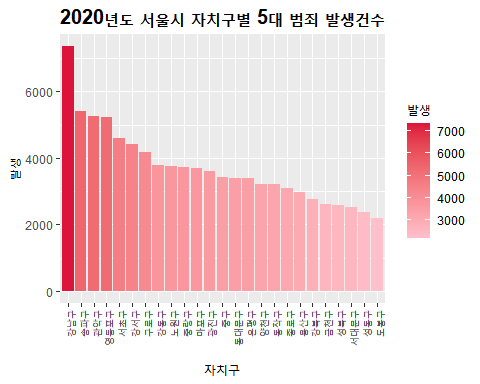
district\_center <- read.csv("./Project/USE\_DATA/사용할꺼/seoul\_district\_center.csv", header = T)  
victim\_data <- read\_excel("./Project/USE\_DATA/사용할꺼/victim\_sex\_age.xlsx")   
crime\_time <- read\_excel("./Project/USE\_DATA/사용할꺼/crime\_time.xlsx")  
lamp\_data <- read\_excel("./Project/USE\_DATA/사용할꺼/streetlight.xlsx")   
crime\_cctv\_data <- read\_excel("./Project/USE\_DATA/사용할꺼/crime\_cctv\_data.xlsx")  
bell\_data <- read\_excel("./Project/USE\_DATA/사용할꺼/bell.xlsx")  
realty\_raw <- read.csv("./Project/USE\_DATA/사용할꺼/non\_realty.csv")  
tax\_raw <- read\_excel("./Project/USE\_DATA/사용할꺼/non\_tax.xlsx")   
gorvernment\_data <- read\_xls("./Project/USE\_DATA/사용할꺼/gorvernment.xls")   
scout\_data <- read\_xlsx("./Project/USE\_DATA/사용할꺼/seoul\_scout.xlsx")   
crime\_n\_time <- read.csv("./Project/USE\_DATA/Day\_time\_Crime.csv")  
gu\_total <- read.csv("./Project/USE\_DATA/gu\_tatal.csv")

### 지도 불러오기 & 서울시 지도 만들기

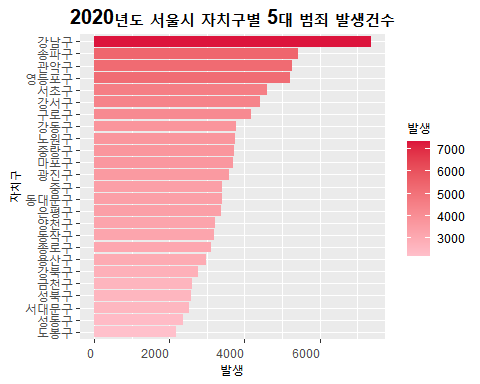
#좌표계 변환   
map\_1 <- spTransform(map, CRSobj = CRS('+proj=longlat +ellps=WGS84 +datum=WGS84 +no\_defs'))  
#map 데이터셋의 SIG\_CD변수들을 id로 변환  
new\_map <- fortify(map\_1, region = 'SIG\_CD')  
# id가 11740 이하가 서울시 구에 해당  
new\_map$id <- as.numeric(new\_map$id)  
# 서울지도 완성  
seoul\_map <- new\_map[new\_map$id <= 11740, ]  
  
# 각 변수 지도에 표시할 범죄통계  
code\_seoul\_b <- dplyr::rename(dist\_id, district = 시군구명)  
dis2\_1 <- dplyr::rename(district\_center, long = X, lat = Y, id = "코드", district = "시군구명")  
b <- crime\_data  
b <- dplyr::rename(b, district = "자치구", cnt = "발생")  
b <- cbind(code\_seoul\_b$id, b)  
b <- cbind(b, dis2\_1$long, dis2\_1$lat)  
b <- dplyr::rename(b, long = "dis2\_1$long", lat = "dis2\_1$lat")  
cntMax <- max(b$cnt)

### 자치구별 범죄 통계

# 범죄 통계 불러오기  
crime\_1 <- crime\_data  
# 막대그래프 시각화  
# 범죄 발생 순위별로 나열  
ggplot(data=crime\_1, aes(x = reorder(자치구,-발생), y = 발생, fill = 발생)) +   
 geom\_bar(stat = "identity") +  
 scale\_fill\_gradient(low = "#FFC0CB", high = "#DC143C") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5)) +  
 labs(title = "2020년도 서울시 자치구별 5대 범죄 발생건수", x = "자치구", y = "발생")+  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))

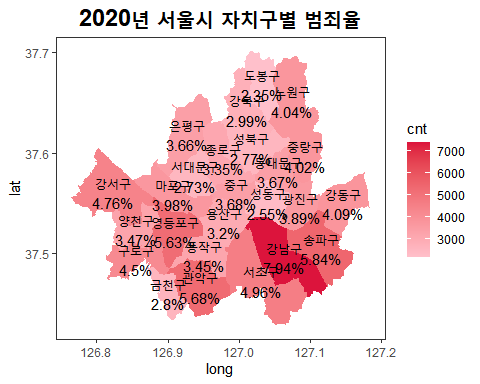


# x축 y축 변경 후, 나열  
ggplot(data=crime\_1, aes(x = reorder(자치구,+발생), y = 발생, fill = 발생)) +   
 geom\_bar(stat = "identity") +  
 scale\_fill\_gradient(low = "#FFC0CB", high = "#DC143C") +  
 theme(axis.text.x = element\_text(angle = 0, hjust = 1, vjust = 0.5)) +  
 coord\_flip() +  
 labs(title = "2020년도 서울시 자치구별 5대 범죄 발생건수", x = "자치구", y = "발생")+  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* 2020년도 서울시 자치구별 5대 범죄 발생
* top5 <강남구, 송파구, 관악구, 영등포구, 서초구>
* bottom5 <도봉구, 성동구, 서대문구, 성북구, 강북구>

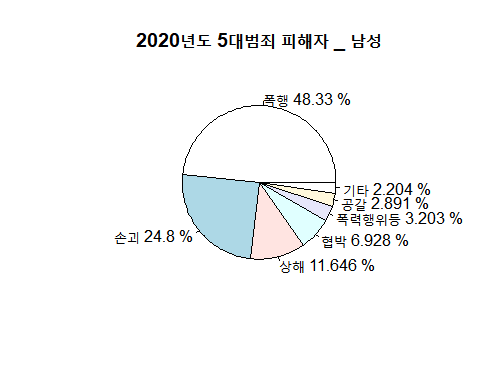
#지도시각화  
crime\_1 <- dplyr::rename(crime\_1, district = "자치구", cnt = "발생")  
code\_seoul\_1 <- dist\_id  
  
# 사용할 데이터 셋 만들기  
code\_seoul\_1 <- dplyr::rename(code\_seoul\_1, district = "시군구명")  
P <- merge(x = code\_seoul\_1, y = crime\_1, by = 'district')  
P\_merge <- merge(seoul\_map, P, by= 'id')  
  
center\_1 <- dplyr::rename(district\_center, long = X, lat = Y, id = "코드", district = "시군구명")  
center\_1 <- cbind(center\_1, crime\_1$cnt)  
center\_1 <- dplyr::rename(center\_1, cnt = "crime\_1$cnt")  
  
#서울시 범죄 발생건을 퍼센트로 바꾸기  
sum<-sum(center\_1$cnt)  
center\_1$pct<-(center\_1$cnt/sum) \* 100  
  
  
#서울시 지도 단계 구분 시각화   
ggplot() + geom\_polygon(data = P\_merge, aes(x = long, y = lat, group = group, fill = cnt)) +  
 scale\_fill\_gradient(low = "#FFC0CB", high = "#DC143C") +   
 theme\_bw() +  
 labs(title = "2020년 서울시 자치구별 범죄율") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_text(data = center\_1, aes(x = long, y = lat, label = paste(district, paste0(round(pct, 2), "%"), sep = "\n")), col = "black")



* 7.94%로 강남구가 범죄율이 가장 높다
* 2.35%로 도봉구가 범죄율이 가장 낮다

### 범죄 발생 요소(성별 & 시간)

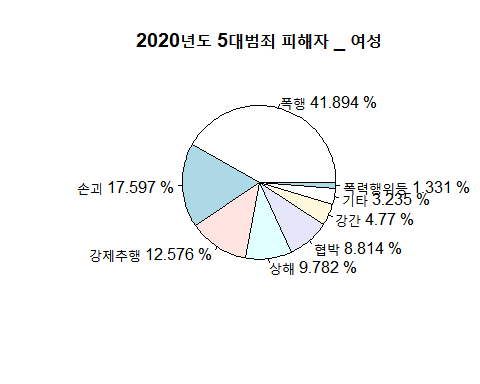
# 범죄 피해자 성별통계 및 범죄 시간 파악하기  
victime\_data\_2 <- victim\_data  
  
# 컬럼명 변경  
victime\_data\_2 <- reshape::rename(victime\_data\_2,  
 c(...2 = "구분",  
 남자...4 = "male1",  
 남자...5 = "male2",  
 남자...6 = "male3",  
 남자...7 = "male4",  
 남자...8 = "male5",  
 남자...9 = "male6",  
 남자...10 = "male7",  
 남자...11 = "male8",  
 남자...12 = "male9",  
 남자...13 = "male10",  
 남자...14 = "male11"))  
victime\_data\_2 <- reshape::rename(victime\_data\_2,  
 c(여자...15 = "female1",   
 여자...16 = "female2",  
 여자...17 = "female3",  
 여자...18 = "female4",  
 여자...19 = "female5",  
 여자...20 = "female6",  
 여자...21 = "female7",  
 여자...22 = "female8",  
 여자...23 = "female9",  
 여자...24 = "female10",  
 여자...25 = "female11"))  
  
# 성별 구분  
male\_2 <- victime\_data\_2 %>%   
 dplyr::select(...1, 구분, male1, male2, male3, male4, male5, male6, male7, male8, male9, male10, male11)  
female\_2 <- victime\_data\_2 %>%   
 dplyr::select(...1, 구분, female1, female2, female3, female4, female5, female6, female7, female8, female9, female10, female11)  
  
# 범죄 피해 통계 시각화 \_ 파이차트 \_ 데이터 정제  
# 남  
male\_2 <- male\_2 %>% dplyr::select(구분, male1)  
male\_2 <- male\_2[-c(1:3, 12, 13),] # 소계 제거  
male\_2$male1 <- as.numeric(male\_2$male1)  
# 여  
female\_2 <- female\_2 %>% dplyr::select(구분, female1)  
female\_2 <- female\_2[-c(1:3, 12, 13),] #소계 제거  
female\_2$female1 <- as.numeric(female\_2$female1)  
  
#### 퍼센트 추가하기  
#남  
sum\_2 <- sum(male\_2$male1)  
male\_2$pct\_1 <- (male\_2$male1 / sum\_2) \* 100  
  
male\_2\_1 <- male\_2 %>% arrange(desc(pct\_1)) %>% head()  
ar\_mal <- male\_2 %>% arrange(desc(pct\_1))  
etc <- ar\_mal[-1:-6,2:3]  
  
male\_2\_1 <- rbind(male\_2\_1, c("기타", sum(etc$male1), sum(etc$pct\_1)))  
male\_2\_1$male1 <- as.numeric(male\_2\_1$male1)  
male\_2\_1$pct\_1 <- as.numeric(male\_2\_1$pct\_1)  
# 남성\_ 파이차트 그리기  
pie(x = male\_2\_1$pct\_1, labels = paste(male\_2\_1$구분, round(male\_2\_1$pct\_1, 3), "%")) +  
 title("2020년도 5대범죄 피해자 \_ 남성")



## integer(0)

* 1위는 폭행(48.33%), 2위는 손괴 (24.8%), 3위는 상해(11.646%)

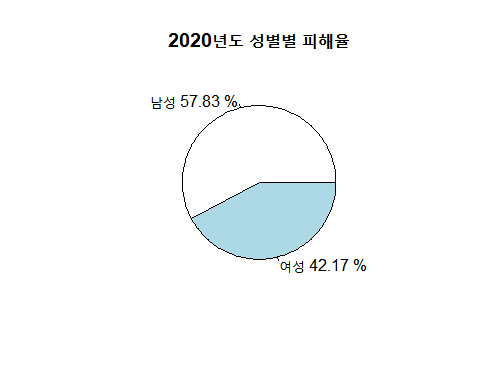
#여  
sum\_2 <- sum(female\_2$female1)  
female\_2$pct\_2 <- (female\_2$female1 / sum\_2)\*100  
  
female\_2\_1 <- female\_2 %>% arrange(desc(pct\_2)) %>% head(7)  
ar\_mal <- female\_2 %>% arrange(desc(pct\_2))  
  
etc <- ar\_mal[-1:-7,2:3]  
female\_2\_1 <- rbind(female\_2\_1, c("기타", sum(etc$female1), sum(etc$pct\_2)))  
female\_2\_1$female1 <- as.numeric(female\_2\_1$female1)  
female\_2\_1$pct\_2 <- as.numeric(female\_2\_1$pct\_2)  
female\_2\_1 <- female\_2\_1 %>% arrange(desc(pct\_2))  
# 여성\_ 파이차트 그리기  
pie(x = female\_2\_1$pct\_2, labels = paste(female\_2\_1$구분, round(female\_2\_1$pct\_2, 3), "%")) +  
 title("2020년도 5대범죄 피해자 \_ 여성")



## integer(0)

* 1위는 폭행(41.894%), 2위는 손괴 (17.597%), 3위는 강제추행(12.576%)
* **범죄 피해의 구분이 성별에 따라 조금씩 다르다.**

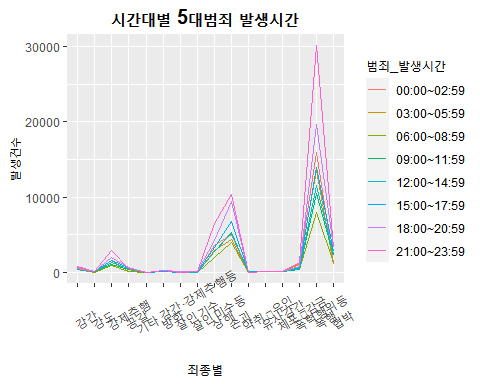
#여성,남성 범죄율 파이차트 그리기  
man\_tot <- sum(male\_2$male1)  
woman\_tot <- sum(female\_2$female1)  
tot <- sum(man\_tot, woman\_tot)  
m\_pc <- (man\_tot / tot) \* 100  
w\_pr <- (woman\_tot / tot) \* 100  
pcq <- data.frame(c(m\_pc, w\_pr))  
dimnames(pcq) = list(row = c("남성", "여성"), col = "피해율")  
  
pie(x = pcq$피해율, labels = paste(row.names(pcq),round(pcq$피해율, 2),"%")) +  
 title("2020년도 성별별 피해율")



## integer(0)

* 남성 피해율이 57.83% 여성 피해율이 42.17%로 남성의 피해율이 더 높다.

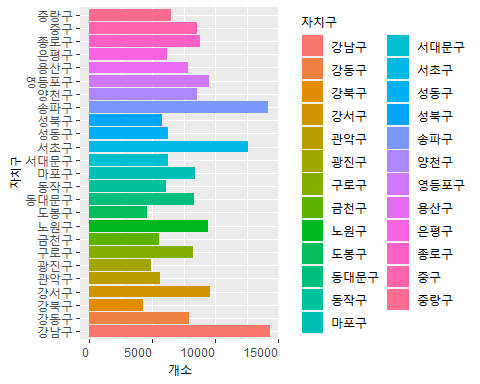
# 범죄 발생 시간  
time <- crime\_time  
  
# 필요한 부분만 추출  
time <- time[,-c(1,3)]  
time <- time[-c(1:2, 11, 12),]  
time <- dplyr::rename(time, 죄종별 = "죄종별...2")  
  
## 중첩 시각화  
lengends <- reshape2::melt(time, id.vars = "죄종별", na.rm = TRUE)  
lengends <- dplyr::rename(lengends, 범죄\_발생시간 = variable)  
ggplot(lengends) + geom\_line(mapping = aes(x = 죄종별, y = value, colour = 범죄\_발생시간, group = 범죄\_발생시간)) +  
 theme(axis.text.x = element\_text(angle = 30, hjust = 0.1, vjust = 0.5)) +  
 labs(title = "시간대별 5대범죄 발생시간", y = "발생건수") +   
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



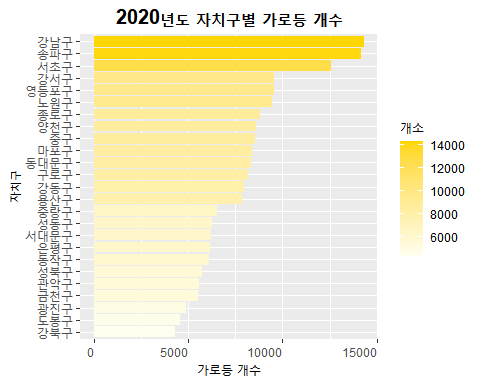
* 21:00~23:59분에 범죄가 제일 많이 일어나는 것으로 보여지고 있다.
* 두 번째로 많이 일어나는 시간은 18:00~20:59분으로 주로 저녁~밤 시간에 범죄가 많이 발생하는 것으로 보인다.

### 자치구별 범죄와 가로등의 관계

##가로등데이터 가져오기  
lamp\_data\_3 <- lamp\_data  
  
##범죄데이터 가져오기  
crime\_data\_3 <- crime\_data  
  
##상관분석을 위해 범죄데이터 + 가로등 데이터 구 기준 결합  
#가로등 변수 변경   
lamp\_data\_3 <- dplyr::rename(lamp\_data\_3, 자치구 = 구분)  
#결합  
crime\_lamp\_data <- left\_join(crime\_data\_3, lamp\_data\_3, by = "자치구")  
  
#fill=자치구  
ggplot(crime\_lamp\_data, aes(x=자치구, y=개소, fill=자치구))+geom\_bar(stat = "identity")+  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip()



#정렬, fill = 가로등 개수  
ggplot(crime\_lamp\_data, aes(x=reorder(자치구,개소), y=개소, fill=개소))+geom\_bar(stat = "identity")+  
 scale\_fill\_gradient(low = "#FFFFF0", high = "#FFD700") +  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +   
 coord\_flip() +   
 labs(title="2020년도 자치구별 가로등 개수", x="자치구", y="가로등 개수") +  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* top5 <강남구, 송파구, 서초구, 강서구, 영등포구>
* bottom5 <강북구, 도봉구, 광진구, 관악구, 성북구>

###상관분석###  
cor.test(crime\_lamp\_data$발생, crime\_lamp\_data$개소)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_lamp\_data$발생 and crime\_lamp\_data$개소  
## t = 5.3951, df = 23, p-value = 1.76e-05  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4998898 0.8820417  
## sample estimates:  
## cor   
## 0.747394

* t = 5.3951, df = 23, p-value = 1.76e-05
* cor 0.747394
* p-value이 0.05미만이므로 유의미하다.
* cor값이 0.747 이므로 높은 양의 관계를 가지고 있다고 판단된다.

### 절편과 기울기 구하기

#가로등 수를 독립변수, 범죄발생수를 종속변수  
lm(발생 ~ 개소, data = crime\_lamp\_data)

##   
## Call:  
## lm(formula = 발생 ~ 개소, data = crime\_lamp\_data)  
##   
## Coefficients:  
## (Intercept) 개소   
## 1130.222 0.325

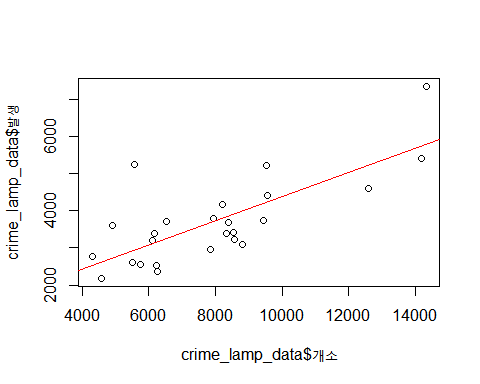
#(Intercept) 개소   
# 1130.222 0.325   
  
#y절편 : 1130.222  
#기울기 : 0.325  
  
#회귀분석  
summary(lm(발생 ~ 개소, data = crime\_lamp\_data))

##   
## Call:  
## lm(formula = 발생 ~ 개소, data = crime\_lamp\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -893.8 -497.7 -321.8 253.7 2314.8   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.130e+03 5.035e+02 2.245 0.0347 \*   
## 개소 3.250e-01 6.023e-02 5.395 1.76e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 796.7 on 23 degrees of freedom  
## Multiple R-squared: 0.5586, Adjusted R-squared: 0.5394   
## F-statistic: 29.11 on 1 and 23 DF, p-value: 1.76e-05

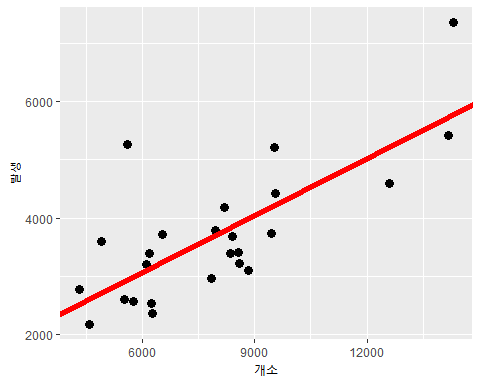
* 결정계수가 0.5586으로 설명력이 높지도 낮지도 않은 회귀직선이다.
* p-value가 1.76e-05로 0.001에서도 유의하다.

### 그래프 & 지도 시각화

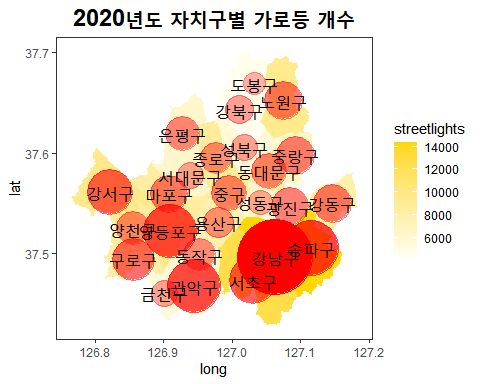
#plot, abline 이용 방법  
plot(crime\_lamp\_data$개소, crime\_lamp\_data$발생)  
abline(lm(발생 ~ 개소, data = crime\_lamp\_data), col = 'red')



#ggplot 이용방법  
ggplot(crime\_lamp\_data, aes(x = 개소, y = 발생)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 1130.222, slope = 0.325, col = 'red', size = 2)



#양의 상관관계를 보임  
  
###지도 시각화###  
# 사용할 데이터 셋 만들기  
crime\_lamp <- dplyr::rename(crime\_lamp\_data, district = "자치구", cnt = 발생, streetlights = "개소")  
crime\_lamp\_1 <- merge(x = code\_seoul\_b, y = crime\_lamp, by = 'district')  
lamp\_merge <- merge(seoul\_map, crime\_lamp\_1, by= 'id')  
  
ggplot() + geom\_polygon(data = lamp\_merge, aes(x = long, y = lat, group = group, fill = streetlights)) +  
 scale\_fill\_gradient(low = "#FFFFF0", high = "#FFD700") +  
 theme\_bw() +  
 labs(title = "2020년도 자치구별 가로등 개수") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "black", size = 5)

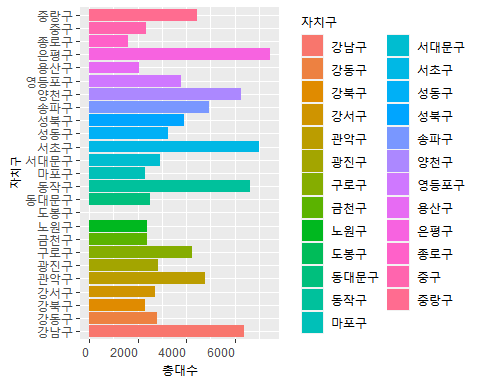


#### **각 빨간 원들은 자치구의 범죄건수를 나타내주며, 범죄건수가 크면 원의 크기가 커지고 원의 투명도는 범죄건수의 최대값으로부터 낮아질수록 투명해진다.**

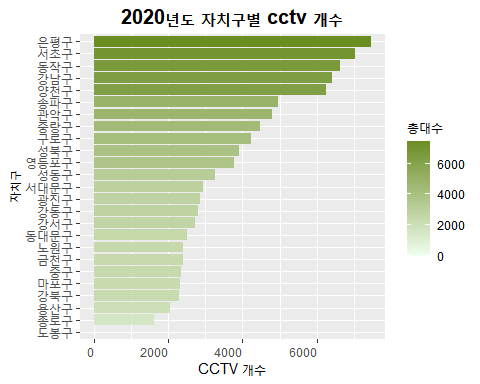
* 강남구를 제외하면 가로등 개수가 많은 곳은 범죄율이 낮다.

### CCTV와 범죄율의 관계

cctv <- crime\_cctv\_data # 상관분석에 사용할 예정(도봉구 삭제)  
cctv\_1 <- cctv # 차트표에 사용 (도봉구 0)  
cctv\_2 <- cctv # 지도시각화에 사용 (도봉구 NA)  
  
cctv <- cctv[-10,]  
cctv\_1[10,3] <- 0  
  
#  
ggplot(cctv\_1, aes(x=자치구, y=총대수, fill=자치구))+geom\_bar(stat = "identity")+  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip()



#정렬, fill = cctv 개수  
ggplot(cctv\_1, aes(x=reorder(자치구,총대수), y=총대수, fill=총대수))+geom\_bar(stat = "identity")+  
 scale\_fill\_gradient(low = "#F0FFF0", high = "#6B8E23") +  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip() +  
 labs(title="2020년도 자치구별 cctv 개수", x="자치구", y="CCTV 개수") +   
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* top5 <은평구, 서초구, 동작구, 강남구, 양천구>
* bottom5 <종로구, 용산구, 강북구, 마포구, 중구>

cor.test(cctv$범죄발생수, cctv$총대수)

##   
## Pearson's product-moment correlation  
##   
## data: cctv$범죄발생수 and cctv$총대수  
## t = 2.1447, df = 22, p-value = 0.04328  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.0149489 0.7015511  
## sample estimates:  
## cor   
## 0.415838

# cor   
# 0.415838   
  
lm(범죄발생수 ~ 총대수, cctv)

##   
## Call:  
## lm(formula = 범죄발생수 ~ 총대수, data = cctv)  
##   
## Coefficients:  
## (Intercept) 총대수   
## 2722.9250 0.2725

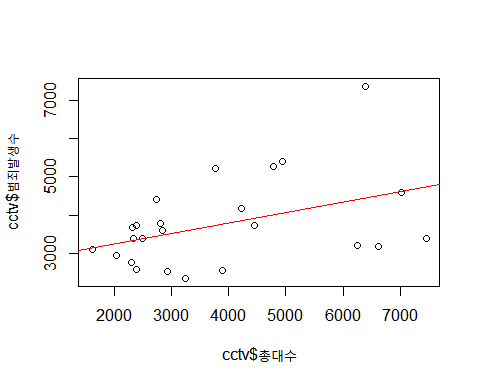
# (Intercept) 총대수   
# 2722.9250 0.2725   
# y 절편 : 2722.9250  
# 기울기 : 0.2725  
  
#회귀분석  
summary(lm(범죄발생수 ~ 총대수, cctv))

##   
## Call:  
## lm(formula = 범죄발생수 ~ 총대수, data = cctv)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1363.60 -830.44 -18.85 341.46 2890.12   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2722.9250 535.4716 5.085 4.29e-05 \*\*\*  
## 총대수 0.2725 0.1270 2.145 0.0433 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1073 on 22 degrees of freedom  
## Multiple R-squared: 0.1729, Adjusted R-squared: 0.1353   
## F-statistic: 4.6 on 1 and 22 DF, p-value: 0.04328

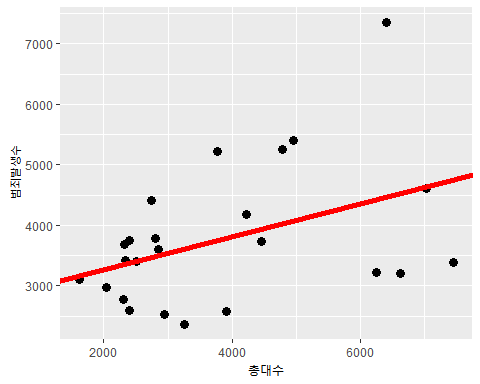
* 결정계수가 0.1729로 설명력이 매우 낮다.
* p-value는 0.04328로 p값0.05에서 유의하다.

### 그래프 & 지도 시각화

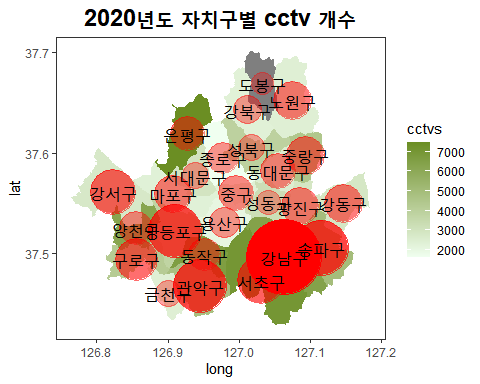
#plot, abline 이용 방법  
plot(cctv$총대수, cctv$범죄발생수)  
abline(lm(범죄발생수 ~ 총대수, cctv), col = 'red')



#ggplot 이용방법  
ggplot(cctv, aes(x = 총대수, y = 범죄발생수)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 2722.9250, slope = 0.2725, col = 'red', size = 2)



###지도 시각화###  
# 사용할 데이터 셋 만들기  
crime\_cctv <- dplyr::rename(cctv\_2, district = "자치구", cnt = 범죄발생수, cctvs = "총대수")  
crime\_cctv\_1 <- merge(x = code\_seoul\_b, y = crime\_cctv, by = 'district')  
cctv\_merge <- merge(seoul\_map, crime\_cctv\_1, by= 'id')  
  
ggplot() + geom\_polygon(data = cctv\_merge, aes(x = long, y = lat, group = group, fill = cctvs)) +  
 scale\_fill\_gradient(low = "#F0FFF0", high = "#6B8E23") +  
 theme\_bw() +  
 labs(title = "2020년도 자치구별 cctv 개수") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "black", size = 5)



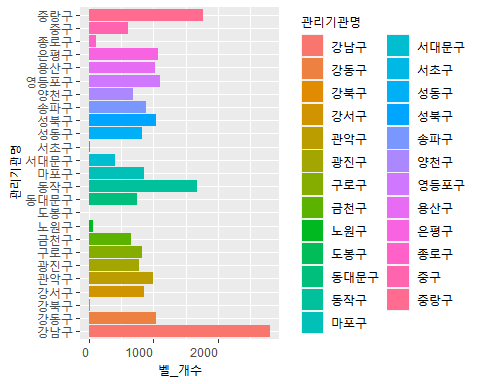
* 강남구를 제외하면 cctv대수가 많은 곳은 범죄율이 낮다.

### 안전비상벨과 범죄율의 관계

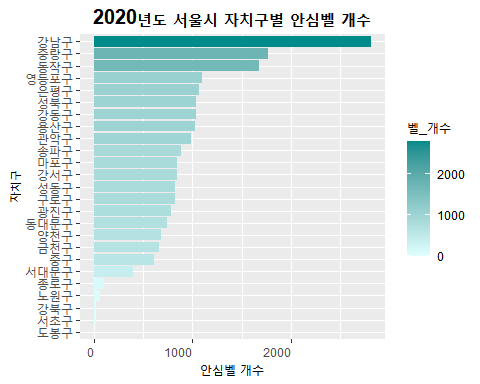
bell <- bell\_data  
  
unique(bell$관리기관명)

## [1] "중랑구" "성동구" "송파구" "관악구" "은평구" "동작구"   
## [7] "서초구" "강북구" "종로구" "성북구" "강남구" "구로구"   
## [13] "양천구" "금천구" "영등포구" "서대문구" "중구" "노원구"   
## [19] "용산구" "광진구" "강동구" "동대문구" "강서구" "마포구"

bell\_num <- bell %>% dplyr::select(관리기관명) %>% group\_by(관리기관명) %>% summarise(벨\_개수 = n())  
bell\_num\_1 <- bell\_num # 차트표에 사용  
bell\_num\_2 <- bell\_num # 지도시각화에 사용  
  
bell\_num\_1[25, ]$관리기관명 <- "도봉구"  
bell\_num\_2[25, ]$관리기관명 <- "도봉구"  
bell\_num\_1[25, ]$벨\_개수 <- 0  
  
#fill = 자치구  
ggplot(bell\_num\_1, aes(x=관리기관명, y=벨\_개수, fill=관리기관명))+geom\_bar(stat = "identity")+  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip()



#정렬, fill = 벨 개수  
ggplot(bell\_num\_1, aes(x=reorder(관리기관명,벨\_개수), y=벨\_개수, fill=벨\_개수))+geom\_bar(stat = "identity")+  
 scale\_fill\_gradient(low = "#E0FFFF", high = "#008B8B") +  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip() +   
 labs(title="2020년도 서울시 자치구별 안심벨 개수", x="자치구", y="안심벨 개수") +  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* top5< 강남구, 중랑구, 동작구, 영등포구, 은평구>
* bottom5< 서초구, 강북구, 노원구, 종로구, 서대문구>

crime\_data\_5 <- dplyr::rename(crime\_data, 관리기관명 = 자치구)  
  
#범죄율 데이터 도봉구 제거  
crime\_data\_5\_1 <- crime\_data\_5[-10, ]  
  
crime\_bell\_data <- left\_join(crime\_data\_5\_1, bell\_num, by = "관리기관명") # 상관관계에 사용  
crime\_bell\_data\_2 <- left\_join(crime\_data\_5, bell\_num\_2, by = "관리기관명") # 지도 시각화에 사용용  
  
cor.test(crime\_bell\_data$발생, crime\_bell\_data$벨\_개수)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_bell\_data$발생 and crime\_bell\_data$벨\_개수  
## t = 2.8545, df = 22, p-value = 0.009217  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.1473900 0.7632146  
## sample estimates:  
## cor   
## 0.5198766

# cor   
# 0.5198766  
  
lm(발생 ~ 벨\_개수, crime\_bell\_data)

##   
## Call:  
## lm(formula = 발생 ~ 벨\_개수, data = crime\_bell\_data)  
##   
## Coefficients:  
## (Intercept) 벨\_개수   
## 2915.1932 0.9864

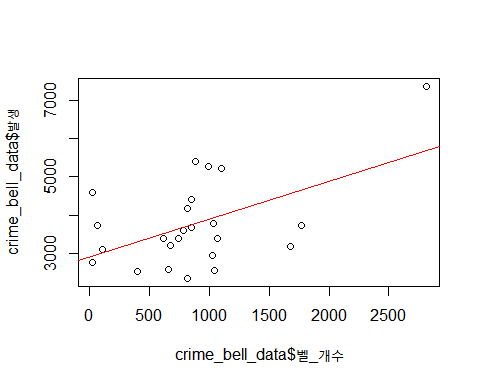
# (Intercept) 벨\_개수   
# 2915.1932 0.9864   
# y 절편 : 2915.1932  
# 기울기 : 0.9864   
  
#회귀분석  
summary(lm(발생 ~ 벨\_개수, crime\_bell\_data))

##   
## Call:  
## lm(formula = 발생 ~ 벨\_개수, data = crime\_bell\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1376.0 -817.6 -129.5 689.4 1668.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2915.1932 363.5836 8.018 5.69e-08 \*\*\*  
## 벨\_개수 0.9864 0.3456 2.855 0.00922 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1008 on 22 degrees of freedom  
## Multiple R-squared: 0.2703, Adjusted R-squared: 0.2371   
## F-statistic: 8.148 on 1 and 22 DF, p-value: 0.009217

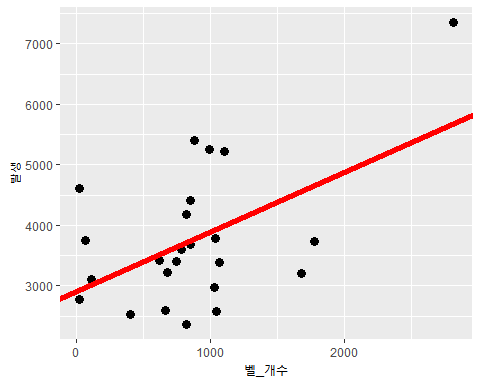
* 결정계수0.2703으로 설명력이 매우 낮다.
* p-value값은 0.009217로 0.01에서 유의하다.

### 그래프 & 지도 시각화

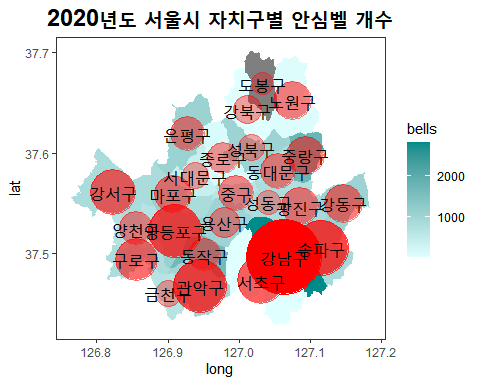
#plot, abline 이용 방법  
plot(crime\_bell\_data$벨\_개수, crime\_bell\_data$발생)  
abline(lm(발생 ~ 벨\_개수, crime\_bell\_data), col = 'red')



#ggplot 이용방법  
ggplot(crime\_bell\_data, aes(x = 벨\_개수, y = 발생)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 2915.1932, slope = 0.9864, col = 'red', size = 2)



###지도 시각화###  
# 사용할 데이터 셋 만들기  
crime\_bell\_data\_2 <- dplyr::rename(crime\_bell\_data\_2, district = "관리기관명", cnt = 발생, bells = "벨\_개수")  
  
bell\_1 <- merge(x = code\_seoul\_b, y = crime\_bell\_data\_2, by = 'district')  
bell\_merge <- merge(seoul\_map, bell\_1, by= 'id')  
  
ggplot() + geom\_polygon(data = bell\_merge, aes(x = long, y = lat, group = group, fill = bells)) +  
 scale\_fill\_gradient(low = "#E0FFFF", high = "#008B8B") +  
 theme\_bw() +  
 labs(title = "2020년도 서울시 자치구별 안심벨 개수") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "black", size = 5)



* 강남구를 제외하면 안심벨 개수가 많은 곳은 범죄율이 낮다.

### 서울시 자치구별 집값과 범죄율의 관계

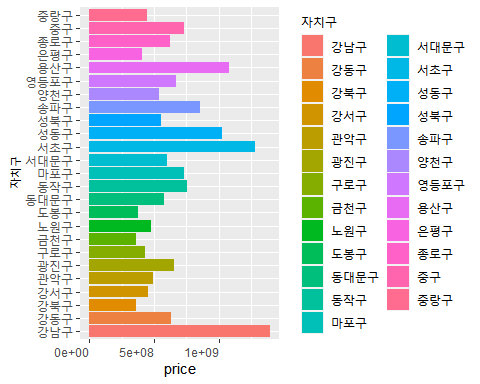
str(realty\_raw)

## 'data.frame': 176001 obs. of 4 variables:  
## $ 자치구명 : chr "성북구" "성북구" "구로구" "용산구" ...  
## $ 건물주용도: chr "아파트" "아파트" "아파트" "아파트" ...  
## $ 물건금액 : num 7.49e+08 1.20e+09 1.05e+08 8.30e+08 7.35e+08 1.43e+09 4.00e+08 8.75e+08 3.13e+08 7.40e+08 ...  
## $ 건물명 : chr "래미안 장위포레카운티" "롯데캐슬 클라시아 " "궁전아트빌라" "리버힐삼성" ...

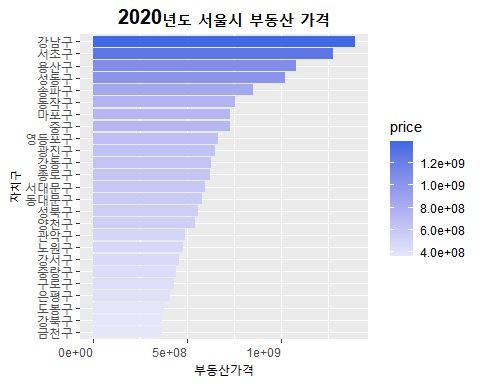
realty <- realty\_raw  
realty$물건금액 <- as.integer(realty$물건금액)

## Warning: NAs introduced by coercion to integer range

realty <- realty\_raw %>%dplyr::select(자치구명,물건금액)%>%group\_by(자치구명)%>%summarise(mean(물건금액))  
  
#변수명 변경  
realty <- dplyr::rename(realty, price="mean(물건금액)")  
realty <- dplyr::rename(realty, 자치구="자치구명")  
  
#fill=자치구  
ggplot(realty, aes(x=자치구, y=price, fill=자치구))+geom\_bar(stat = "identity")+  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip()



#정렬, fill=price / 파란색 bar  
ggplot(realty, aes(x=reorder(자치구,price), y=price, fill=price))+geom\_bar(stat = "identity")+  
 scale\_fill\_gradient(low = "#E6E6FA", high = "#4169E1") +  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip() +  
 labs(title="2020년도 서울시 부동산 가격", x="자치구", y="부동산가격") +  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* top5<강남구, 서초구, 용산구, 성동구, 송파구>
* bottom5<금천구, 강북구,도봉구, 은평구, 구로구>

crime\_realty\_data <- left\_join(crime\_data, realty, by = "자치구")  
  
#상관관계 분석  
cor.test(crime\_realty\_data$발생, crime\_realty\_data$price)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_realty\_data$발생 and crime\_realty\_data$price  
## t = 2.4769, df = 23, p-value = 0.02104  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.07787828 0.72293575  
## sample estimates:  
## cor   
## 0.4588882

# t = 2.4769, df = 23, p-value = 0.02104  
# cor   
# 0.4588882   
  
lm(발생 ~ price, crime\_realty\_data)

##   
## Call:  
## lm(formula = 발생 ~ price, data = crime\_realty\_data)  
##   
## Coefficients:  
## (Intercept) price   
## 2.422e+03 1.947e-06

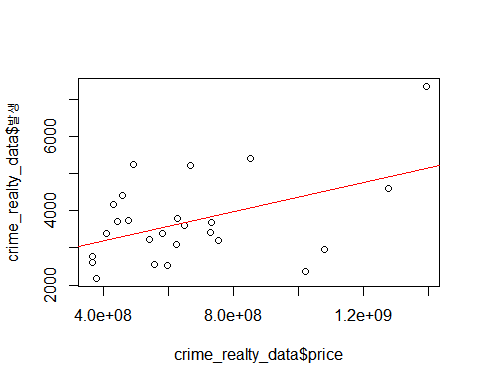
# (Intercept) price   
# 2.422e+03 1.947e-06   
# y 절편 : 2.422e+03  
# 기울기 : 1.947e-06   
  
#회귀분석  
summary(lm(발생 ~ price, crime\_realty\_data))

##   
## Call:  
## lm(formula = 발생 ~ price, data = crime\_realty\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2049.8 -534.6 -157.4 440.3 2221.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.422e+03 5.609e+02 4.318 0.000255 \*\*\*  
## price 1.947e-06 7.861e-07 2.477 0.021035 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1065 on 23 degrees of freedom  
## Multiple R-squared: 0.2106, Adjusted R-squared: 0.1763   
## F-statistic: 6.135 on 1 and 23 DF, p-value: 0.02104

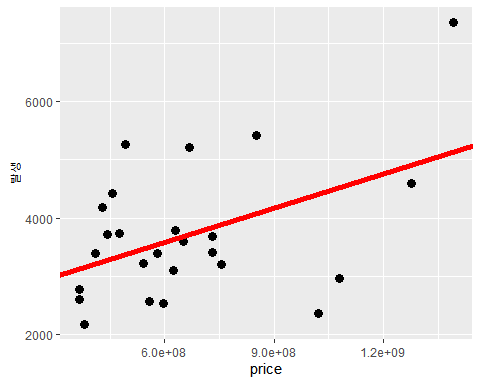
* 결정계수가 0.2106으로 설명력이 매우 낮은 회귀직선이다.
* p-value가 0.02104로 0.05에서 유의하다.

### 그래프 & 지도 시각화

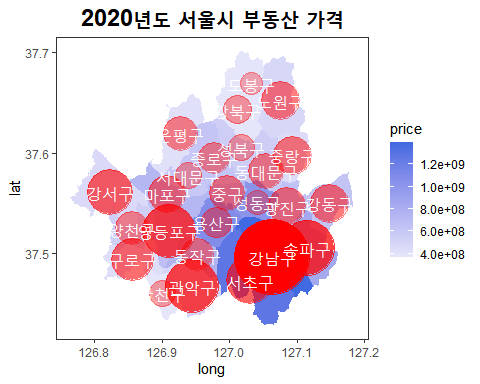
#plot, abline 이용 방법  
plot(crime\_realty\_data$price, crime\_realty\_data$발생)  
abline(lm(발생 ~ price, crime\_realty\_data), col = 'red')



#ggplot 이용방법  
ggplot(crime\_realty\_data, aes(x = price, y = 발생)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 2.422e+03, slope = 1.947e-06, col = 'red', size = 2)



###지도 시각화###  
# 사용할 데이터 셋 만들기  
crime\_realty\_data <- dplyr::rename(crime\_realty\_data, district = "자치구", cnt = 발생)  
realty\_1 <- merge(x = code\_seoul\_b, y = crime\_realty\_data, by = 'district')  
realty\_merge <- merge(seoul\_map, realty\_1, by= 'id')  
  
ggplot() + geom\_polygon(data = realty\_merge, aes(x = long, y = lat, group = group, fill = price)) +  
 scale\_fill\_gradient(low = "#E6E6FA", high = "#4169E1") +  
 theme\_bw() +  
 labs(title = "2020년도 서울시 부동산 가격") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "white", size = 5)



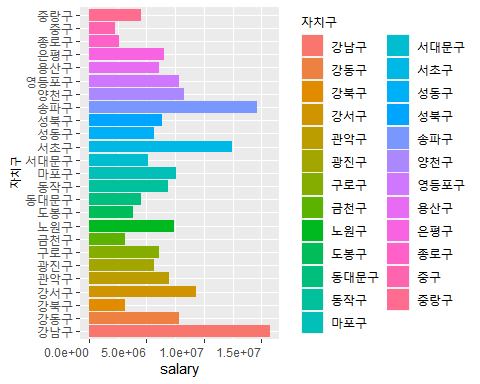
* 강남구를 제외하면 부동산 가격이 높은 곳이 범죄율도 높다.

### 서울시 지역별 평균소득과 범죄율의 관계

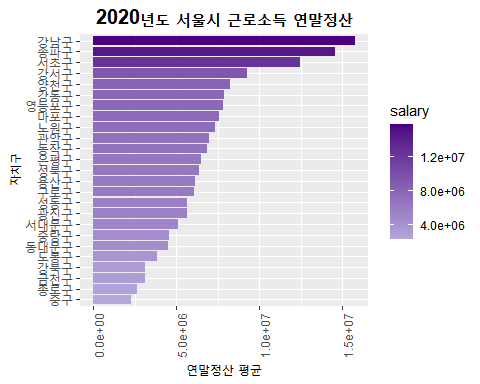
str(tax\_raw)

## tibble [25 x 3] (S3: tbl\_df/tbl/data.frame)  
## $ ...1 : chr [1:25] "송파구" "강서구" "관악구" "강남구" ...  
## $ 인원 : num [1:25] 280206 247909 217065 210941 191648 ...  
## $ 금액(백만원): num [1:25] 14594561 9287969 6961669 15754700 7363982 ...

tax <- tax\_raw  
tax <- dplyr::rename(tax, salary="금액(백만원)")  
tax <- dplyr::rename(tax, 자치구 = ...1)  
tax <- tax%>%mutate(salary/인원)  
  
  
  
#salary/인원 시각화  
ggplot(tax, aes(x=자치구, y=salary, fill=자치구))+geom\_bar(stat="identity")+  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip()



#정렬 / fill=salary  
ggplot(tax, aes(x=reorder(자치구,salary), y=salary, fill=salary)) + geom\_bar(stat="identity") +  
 scale\_fill\_gradient(low = "#B3A7DC", high = "#4B0082") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1)) +   
 coord\_flip() +  
 labs(title="2020년도 서울시 근로소득 연말정산", x="자치구", y="연말정산 평균") +  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* top5 < 강남구, 송파구, 서초구, 강서구, 양천구 >
* bottom5 <중구, 종로구, 금천구, 강북구, 도봉구 >

crime\_tax\_data <- left\_join(crime\_data, tax, by = "자치구")  
  
#상관관계 분석  
cor.test(crime\_tax\_data$발생, crime\_tax\_data$salary)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_tax\_data$발생 and crime\_tax\_data$salary  
## t = 6.0719, df = 23, p-value = 3.416e-06  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.5647158 0.9006142  
## sample estimates:  
## cor   
## 0.784742

# t = 6.0719, df = 23, p-value = 3.416e-06  
# cor 0.784742   
  
crime\_tax\_data <- dplyr::rename(crime\_tax\_data, mean\_salary = `salary/인원`)  
  
lm(발생 ~ mean\_salary, crime\_tax\_data)

##   
## Call:  
## lm(formula = 발생 ~ mean\_salary, data = crime\_tax\_data)  
##   
## Coefficients:  
## (Intercept) mean\_salary   
## 1796.52 43.72

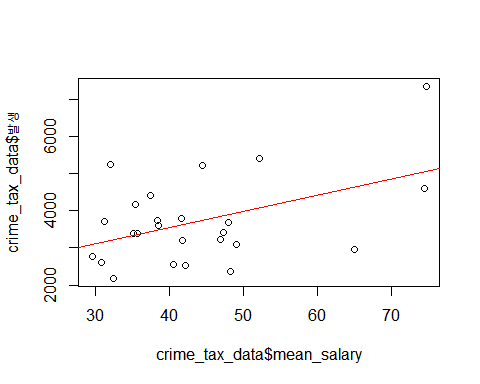
# (Intercept) mean\_salary   
# 1796.52 43.72   
# y 절편 : 1796.52  
# 기울기 : 43.72  
  
#회귀분석  
summary(lm(발생 ~ mean\_salary, crime\_tax\_data))

##   
## Call:  
## lm(formula = 발생 ~ mean\_salary, data = crime\_tax\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1665.8 -628.0 -207.4 565.6 2294.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1796.52 803.13 2.237 0.0353 \*  
## mean\_salary 43.72 17.72 2.468 0.0215 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1066 on 23 degrees of freedom  
## Multiple R-squared: 0.2093, Adjusted R-squared: 0.1749   
## F-statistic: 6.089 on 1 and 23 DF, p-value: 0.02147

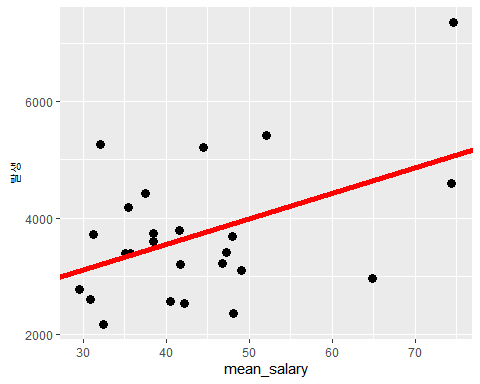
* 결정계수가 0.2093으로 설명력이 매우 낮은 회귀직선이다.
* p-value는 0.02147로 0.05에서 유의하다.

### 그래프 & 지도 시각화

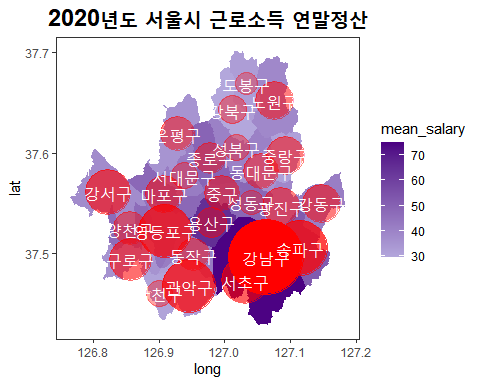
#plot, abline 이용 방법  
plot(crime\_tax\_data$mean\_salary, crime\_tax\_data$발생)  
abline(lm(발생 ~ mean\_salary, crime\_tax\_data), col = 'red')



#ggplot 이용방법  
ggplot(crime\_tax\_data, aes(x = mean\_salary, y = 발생)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 1796.52, slope = 43.72, col = 'red', size = 2)



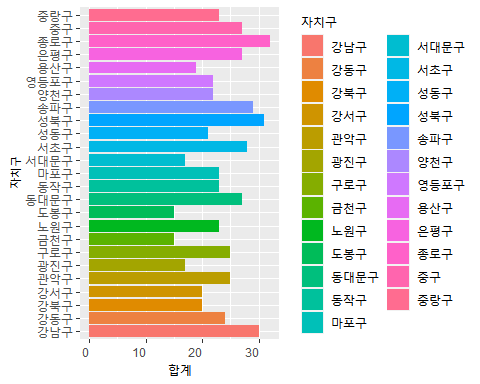
###지도 시각화###  
# 사용할 데이터 셋 만들기  
crime\_tax\_data <- dplyr::rename(crime\_tax\_data, district = "자치구")  
tax\_1 <- merge(x = code\_seoul\_b, y = crime\_tax\_data, by = 'district')  
tax\_1 <- dplyr::rename(tax\_1, cnt = 발생)  
tax\_merge <- merge(seoul\_map, tax\_1, by= 'id')  
  
ggplot() + geom\_polygon(data = tax\_merge, aes(x = long, y = lat, group = group, fill = mean\_salary)) +  
 scale\_fill\_gradient(low = "#B3A7DC", high = "#4B0082") +  
 theme\_bw() +  
 labs(title = "2020년도 서울시 근로소득 연말정산") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "white", size = 5)



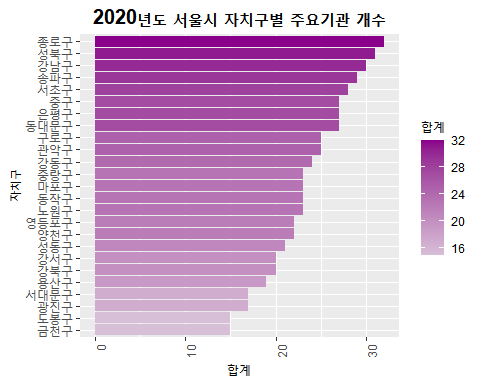
* 강남구를 제외하면 소득이 높은 곳은 범죄율이 낮다.

### 서울시 주요공공기관(경찰 & 소방서)개수와 범죄율의 관계

#공공기관데이터 가져오기  
gorvernment\_1 <- gorvernment\_data  
  
#  
ggplot(gorvernment\_1, aes(x=자치구, y=합계, fill=자치구))+geom\_bar(stat="identity")+  
 theme(axis.text.x = element\_text(angle=0, hjust=1)) +  
 coord\_flip()



#정렬 / fill = 공공기관 개수  
ggplot(gorvernment\_1, aes(x=reorder(자치구,합계), y=합계, fill=합계))+geom\_bar(stat="identity")+  
 scale\_fill\_gradient(low = "#D8BFD8", high = "#8B008B") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1)) +   
 coord\_flip() +  
 labs(title="2020년도 서울시 자치구별 주요기관 개수", x="자치구", y="합계") +  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



##범죄데이터 가져오기  
crime\_data\_8 <- crime\_data  
  
##상관분석을 위해 범죄데이터 + 주요기관 데이터 구 기준 결합  
#공공기관 변수 변경   
gorvernment\_1<- dplyr::rename(gorvernment\_1, sum= 합계)  
#결합  
crime\_gorvernmnet <- left\_join(crime\_data\_8, gorvernment\_1, by = "자치구")  
  
###상관분석###  
cor.test(crime\_gorvernmnet$발생, crime\_gorvernmnet$sum)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_gorvernmnet$발생 and crime\_gorvernmnet$sum  
## t = 2.3977, df = 23, p-value = 0.02501  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.06322413 0.71583012  
## sample estimates:  
## cor   
## 0.4471836

#t = 2.3977, df = 23, p-value = 0.02501  
#cor 0.4471836  
  
#결과 분석  
#p-value이 0.05미만이므로 유의미하다.  
#cor값이 0.447이므로 거의 관계가 없다고 판단된다.  
  
##절편과 기울기 구하기  
#주요기관 수를 독립변수, 범죄발생수를 종속변수  
lm(발생 ~ sum, data = crime\_gorvernmnet)

##   
## Call:  
## lm(formula = 발생 ~ sum, data = crime\_gorvernmnet)  
##   
## Coefficients:  
## (Intercept) sum   
## 1136.6 109.9

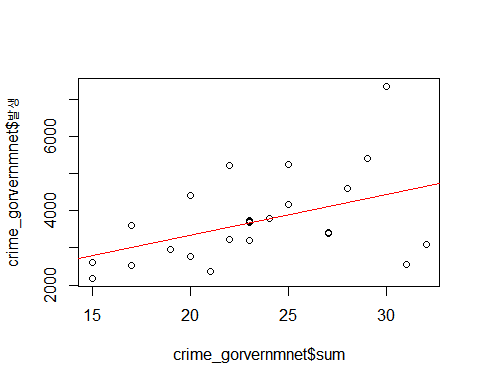
#(Intercept) sum   
# 1136.6 109.9   
  
#y절편 : 1136.6  
#기울기 : 109.9  
  
#회귀직선  
summary(lm(발생 ~ sum, data = crime\_gorvernmnet))

##   
## Call:  
## lm(formula = 발생 ~ sum, data = crime\_gorvernmnet)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1975.0 -605.4 -186.4 388.5 2923.8   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1136.60 1093.34 1.040 0.309   
## sum 109.85 45.82 2.398 0.025 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1073 on 23 degrees of freedom  
## Multiple R-squared: 0.2, Adjusted R-squared: 0.1652   
## F-statistic: 5.749 on 1 and 23 DF, p-value: 0.02501

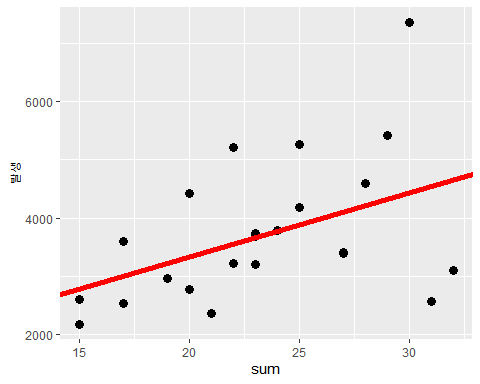
* 결정계수0.2로 설명력이 매우 낮은 회귀직선이다.
* p-value는 0.02501로 0.05에서 유의하다.

###그래프 & 지도 시각화

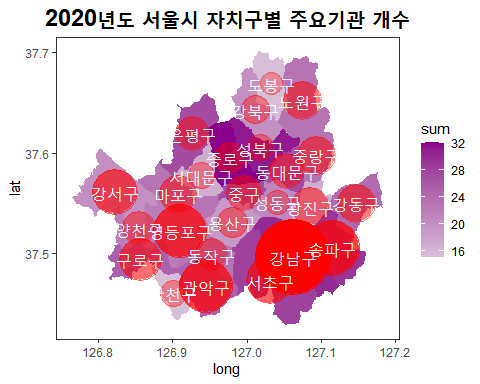
#plot, abline 이용 방법  
plot(crime\_gorvernmnet$sum, crime\_gorvernmnet$발생)  
abline(lm(발생 ~ sum, data = crime\_gorvernmnet), col = 'red')



#ggplot 이용방법  
ggplot(crime\_gorvernmnet, aes(x = sum, y = 발생)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 1136.6, slope = 109.9, col = 'red', size = 2)



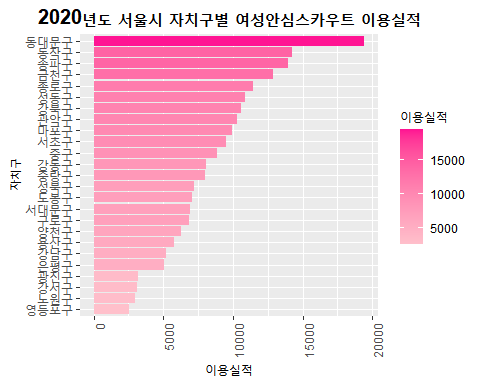
crime\_gorvernmnet <- dplyr::rename(crime\_gorvernmnet, district = "자치구", cnt = "발생")  
gorvernment\_1 <- merge(x = code\_seoul\_b, y = crime\_gorvernmnet, by = 'district')  
gorvern\_merge <- merge(seoul\_map, gorvernment\_1, by= 'id')  
  
ggplot() + geom\_polygon(data = gorvern\_merge, aes(x = long, y = lat, group = group, fill = sum)) +  
 scale\_fill\_gradient(low = "#D8BFD8", high = "#8B008B") +  
 theme\_bw() +  
 labs(title = "2020년도 서울시 자치구별 주요기관 개수") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "white", size = 5)



* 강남구를 제외하면 자치구별 주요기관 개수가 많을 수록 범죄율이 낮다.

### 안심 스카우트 이용건수와 범죄율의 관계

# 2020년도 서울시 자치구별 안심스카우트 이용 건수  
#2020년만 추출  
a\_2020 <- scout\_data %>%   
 dplyr::select('자치구', '2020', '...14')  
  
# 불필요한 행 제거 및 컬럼명 수정  
a\_2020 <- a\_2020[-c(1:2),]  
names(a\_2020) <- c("district", "이용실적", "스카우트 인원")  
  
# 문자형에서 숫자형으로 변환  
a\_2020 <- as.data.frame(a\_2020)  
a\_2020$이용실적 <- as.numeric(a\_2020$이용실적)  
a\_2020$'스카우트 인원' <- as.numeric(a\_2020$'스카우트 인원')  
  
  
#안심스카우트와 범죄발생 합치기  
a\_2020 <- cbind(a\_2020, crime\_data$발생)  
a\_2020 <- dplyr::rename(a\_2020, cnt = "crime\_data$발생")  
  
  
#정렬 / fill = 스카우트 이용실적  
ggplot(a\_2020, aes(x=reorder(district,이용실적), y=이용실적, fill=이용실적))+geom\_bar(stat="identity")+  
 scale\_fill\_gradient(low = "#FFC0CB", high = "#FF1493") +  
 theme(axis.text.x = element\_text(angle=90, hjust=1)) +   
 coord\_flip() +  
 labs(title="2020년도 서울시 자치구별 여성안심스카우트 이용실적", x="자치구", y="이용실적") +  
 theme(plot.title = element\_text(face = "bold", size = 15, hjust = 0.5))



* top5 < 동대문구, 동작구, 송파구, 금천구, 종로구 >
* bottom5 <영등포구, 노원구, 강서구, 광진구, 은평구 >

#상관관계  
cor.test(a\_2020$이용실적, a\_2020$cnt)

##   
## Pearson's product-moment correlation  
##   
## data: a\_2020$이용실적 and a\_2020$cnt  
## t = -0.8865, df = 23, p-value = 0.3845  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.5382419 0.2298720  
## sample estimates:  
## cor   
## -0.1817688

#상관계수가 -0.1817688로 관계가 거의 없다고 볼 수 있다.   
  
lm(cnt ~ 이용실적, a\_2020)

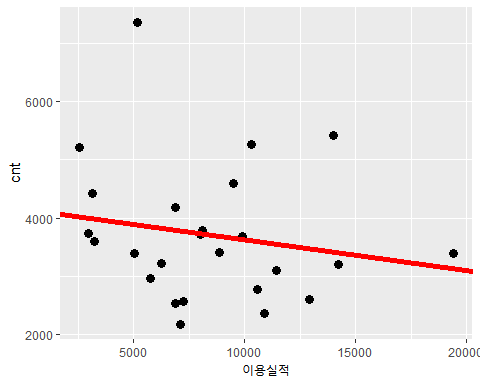
##   
## Call:  
## lm(formula = cnt ~ 이용실적, data = a\_2020)  
##   
## Coefficients:  
## (Intercept) 이용실적   
## 4150.29889 -0.05268

#회귀분석  
summary(lm(cnt ~ 이용실적, a\_2020))

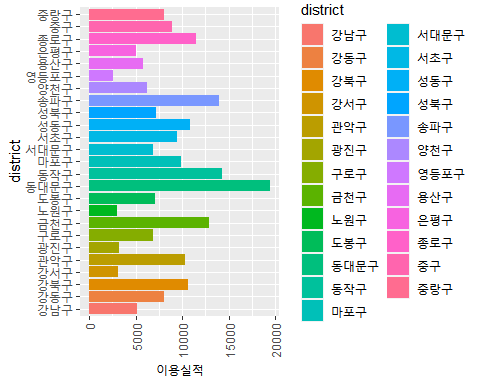
##   
## Call:  
## lm(formula = cnt ~ 이용실적, data = a\_2020)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1597.7 -823.3 -251.0 387.3 3479.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4150.29889 552.70865 7.509 1.25e-07 \*\*\*  
## 이용실적 -0.05268 0.05943 -0.887 0.385   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1179 on 23 degrees of freedom  
## Multiple R-squared: 0.03304, Adjusted R-squared: -0.009002   
## F-statistic: 0.7859 on 1 and 23 DF, p-value: 0.3845

* 결정계수가 0.03304로 설명력이 아주 낮다.
* p-value값이 0.3845로 유의하지 않다.

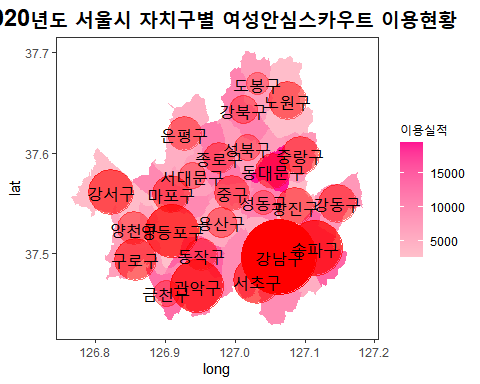
#ggplot 이용하여 산점도에 회귀직선 그리기  
ggplot(a\_2020, aes(x = 이용실적, y = cnt)) +  
 geom\_point(size = 3) +  
 geom\_abline(intercept = 4150.29889, slope = -0.05268, col = 'red', size = 2)



# 2020 이용실적 내림차순 정렬  
a\_2020 <- a\_2020 %>% arrange(desc(a\_2020$`이용실적`))  
  
##  
ggplot(data = a\_2020, aes(x = district, y = 이용실적, fill = district)) +  
 geom\_bar(stat = "identity") +  
 coord\_flip() +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5))



# 시각화  
# 사용할 데이터 셋 만들기  
ansim <- merge(x = code\_seoul\_b, y = a\_2020, by = 'district')  
  
ansim\_merge <- merge(seoul\_map, ansim, by= 'id')  
  
ggplot() + geom\_polygon(data = ansim\_merge, aes(x = long, y = lat, group = group, fill = 이용실적)) +  
 scale\_fill\_gradient(low = "#FFC0CB", high = "#FF1493") +  
 theme\_bw() +  
 labs(title = "2020년도 서울시 자치구별 여성안심스카우트 이용현황") +  
 theme(panel.grid.major.x = element\_blank(), panel.grid.minor.x = element\_blank(),   
 panel.grid.major.y = element\_blank(), panel.grid.minor.y = element\_blank(),  
 plot.title = element\_text(face = "bold", size = 18, hjust = 0.5)) +  
 geom\_point(data = b, aes(x = long, y = lat), size = b$cnt/ 280, color = "red", alpha = b$cnt/cntMax) +  
 geom\_text(data = b, aes(x = long, y = lat), label = b$district, color = "black", size = 5)



* 여성 안심스카우트 이용실적이 높을 수록 범죄율이 낮다.

### 유동인구와 범죄율의 관계

#유동인구-----------------------------------  
#상관관계 전체 유동인구(x)와 범죄율(y)  
cor.test(crime\_n\_time$tot\_mean, crime\_n\_time$crime)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_n\_time$tot\_mean and crime\_n\_time$crime  
## t = 0.61127, df = 6, p-value = 0.5634  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.5577025 0.8088012  
## sample estimates:  
## cor   
## 0.2421251

#상관계수가 0.2421251로 관계가 매우 낮다  
  
  
#상관관계 성별 유동인구(x)와 범죄율(y)  
cor.test(crime\_n\_time$man\_mean, crime\_n\_time$crime)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_n\_time$man\_mean and crime\_n\_time$crime  
## t = 0.73663, df = 6, p-value = 0.4891  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.5227770 0.8251969  
## sample estimates:  
## cor   
## 0.2879862

#상관계수가 0.2879762로 관계가 매우 낮다  
  
cor.test(crime\_n\_time$woman\_mean, crime\_n\_time$crime)

##   
## Pearson's product-moment correlation  
##   
## data: crime\_n\_time$woman\_mean and crime\_n\_time$crime  
## t = 0.50503, df = 6, p-value = 0.6316  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.5861497 0.7936677  
## sample estimates:  
## cor   
## 0.2019285

#상관계수가 0.2019825로 관계가 매우 낮다

* **상관계수가 모두 낮으므로 유동인구의 변수는 제거하기로 한다.**

### 전체 다중회귀분석

#전체 다중회귀분석----------------  
lm(범죄발생수~ccty\_개수+가로등\_개수+벨\_개수+공공기관\_개수+평균소득+평균부동산+안심스카우트\_이용실적,data=gu\_total)

##   
## Call:  
## lm(formula = 범죄발생수 ~ ccty\_개수 + 가로등\_개수 + 벨\_개수 +   
## 공공기관\_개수 + 평균소득 + 평균부동산 + 안심스카우트\_이용실적,   
## data = gu\_total)  
##   
## Coefficients:  
## (Intercept) ccty\_개수 가로등\_개수   
## 2.310e+03 6.896e-02 3.820e-01   
## 벨\_개수 공공기관\_개수 평균소득   
## 4.690e-01 1.634e+01 -8.647e+01   
## 평균부동산 안심스카우트\_이용실적   
## 2.602e-06 -7.212e-02

summary(lm(범죄발생수~ccty\_개수+가로등\_개수+벨\_개수+공공기관\_개수+평균소득+평균부동산+안심스카우트\_이용실적,data=gu\_total))

##   
## Call:  
## lm(formula = 범죄발생수 ~ ccty\_개수 + 가로등\_개수 + 벨\_개수 +   
## 공공기관\_개수 + 평균소득 + 평균부동산 + 안심스카우트\_이용실적,   
## data = gu\_total)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1004.24 -386.21 -85.28 232.58 1851.47   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.310e+03 1.005e+03 2.299 0.035302 \*   
## ccty\_개수 6.896e-02 9.393e-02 0.734 0.473455   
## 가로등\_개수 3.820e-01 8.550e-02 4.468 0.000388 \*\*\*  
## 벨\_개수 4.690e-01 3.009e-01 1.559 0.138605   
## 공공기관\_개수 1.634e+01 3.975e+01 0.411 0.686503   
## 평균소득 -8.647e+01 4.835e+01 -1.788 0.092652 .   
## 평균부동산 2.602e-06 1.973e-06 1.319 0.205743   
## 안심스카우트\_이용실적 -7.212e-02 4.099e-02 -1.760 0.097588 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 673.5 on 16 degrees of freedom  
## Multiple R-squared: 0.7631, Adjusted R-squared: 0.6595   
## F-statistic: 7.365 on 7 and 16 DF, p-value: 0.0004879

* 결정개수가 0.7631로 설명력이 높은 회귀직선이다.

### 전체 COR값 확인하기 & 군집으로 나눠보기

#PCA 분석  
gu\_total4<-read.csv("./Project/USE\_DATA/사용할꺼/gu\_total4.csv")  
library(psych)

## Warning: 패키지 'psych'는 R 버전 4.1.3에서 작성되었습니다

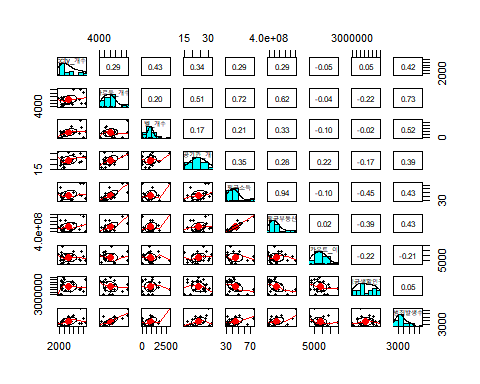
##   
## 다음의 패키지를 부착합니다: 'psych'

## The following object is masked from 'package:raster':  
##   
## distance

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

gu\_total4<-gu\_total4[,-1]  
pairs.panels(gu\_total4)

## Warning in breaks[-1L] + breaks[-nB]: 정수형 오버플로우에 의하여 생성된 NA입니다



* 안심 스카우트를 제외한 다른 변수들과 범죄율의 관계는 모두 양의 관계를 가지고 있다.

gu\_prin <- princomp(gu\_total4, cor=T)  
summary(gu\_prin)

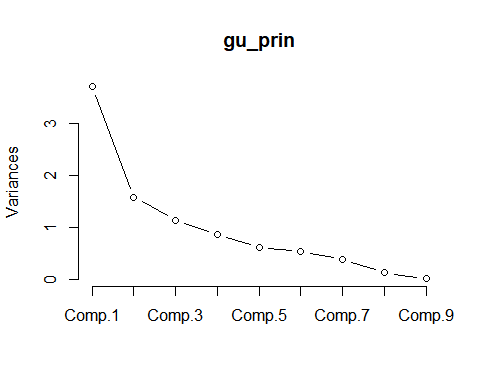
## Importance of components:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5  
## Standard deviation 1.9284917 1.2583449 1.0683393 0.93377498 0.78355182  
## Proportion of Variance 0.4132311 0.1759369 0.1268165 0.09688175 0.06821705  
## Cumulative Proportion 0.4132311 0.5891680 0.7159845 0.81286629 0.88108334  
## Comp.6 Comp.7 Comp.8 Comp.9  
## Standard deviation 0.73329215 0.61672037 0.36105321 0.147746714  
## Proportion of Variance 0.05974638 0.04226045 0.01448438 0.002425455  
## Cumulative Proportion 0.94082972 0.98309017 0.99757455 1.000000000

screeplot(gu\_prin, npsc=5, type="lines")

## Warning in plot.window(...): "npsc"는 그래픽 매개변수가 아닙니다

## Warning in plot.xy(xy, type, ...): "npsc"는 그래픽 매개변수가 아닙니다

## Warning in title(...): "npsc"는 그래픽 매개변수가 아닙니다



* summary(gu\_prin)값을 보면 주성분을 4개로 나눴을때 81.3%로 설명할 수있어서 주성분을 4개로 나눈다.

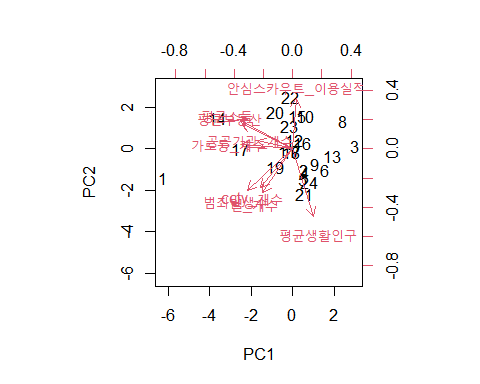
loadings(gu\_prin)

##   
## Loadings:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8  
## ccty\_개수 0.276 0.335 0.308 0.287 0.713 0.351   
## 가로등\_개수 0.441 -0.420 -0.164 0.325 0.668  
## 벨\_개수 0.260 0.373 0.108 0.631 -0.441 -0.114 -0.276 0.309  
## 공공기관\_개수 0.299 0.539 -0.363 0.104 -0.430 -0.527   
## 평균소득 0.447 -0.239 -0.299 0.197 0.176 -0.192   
## 평균부동산 0.434 -0.219 -0.241 0.176 0.381 -0.296 -0.266  
## 안심스카우트\_이용실적 -0.427 0.670 0.149 -0.226 0.502 0.159   
## 평균생활인구 -0.177 0.576 -0.337 0.607 -0.366   
## 범죄발생수 0.390 0.353 -0.208 -0.401 0.366 -0.609  
## Comp.9  
## ccty\_개수   
## 가로등\_개수 0.207  
## 벨\_개수   
## 공공기관\_개수   
## 평균소득 -0.740  
## 평균부동산 0.605  
## 안심스카우트\_이용실적 -0.132  
## 평균생활인구   
## 범죄발생수 -0.118  
##   
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9  
## SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000  
## Proportion Var 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111  
## Cumulative Var 0.111 0.222 0.333 0.444 0.556 0.667 0.778 0.889 1.000

gu\_prin$scores

## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## [1,] 6.3925798 1.46354115 -0.41619223 0.86245052 -1.063460119 -0.09917543  
## [2,] -0.5667286 1.06257174 0.05580880 -0.64172118 -0.472431440 0.86426251  
## [3,] -3.0786998 -0.14272330 0.20203926 -0.51605082 0.110447645 0.62727177  
## [4,] -0.6058367 1.15769671 -1.14699426 -0.75625928 -0.695209074 -0.38021333  
## [5,] -0.5403020 1.48691779 1.24314619 -0.06997474 -0.334075290 0.17563423  
## [6,] -1.6108369 1.03495407 -1.61935066 0.32770194 -0.001022142 0.33945394  
## [7,] -0.2379867 0.10950542 0.31128824 -0.06288480 -0.108856125 -1.37935530  
## [8,] -2.5167220 -1.38898382 -0.12621375 1.32262427 -0.708519815 -0.14116220  
## [9,] -1.1158530 0.75894347 -1.01280497 -1.99439481 0.139824862 -0.05403986  
## [10,] -0.6064367 -1.63043084 2.19747750 -0.13610364 -1.067934382 0.59163598  
## [11,] 0.2649267 0.12911948 1.48260010 2.15241794 0.501085251 0.55545047  
## [12,] -0.1019483 -0.43359213 -0.25846557 -0.25653202 -0.610962740 0.55975644  
## [13,] -1.9704255 0.34454127 -1.10571559 -0.15030010 0.484455465 1.20662481  
## [14,] 3.7271976 -1.53000876 -0.30335079 -0.84331902 2.004487701 0.75684185  
## [15,] -0.2558747 -1.60445935 -0.53339695 1.21674442 0.313756414 0.58124976  
## [16,] -0.4700223 -0.28403307 0.87385721 0.22272898 0.590389429 -1.42121563  
## [17,] 2.5857044 -0.01357414 1.39792819 -1.37433591 -0.496190869 1.03591033  
## [18,] 0.0821375 0.15465966 -0.19538074 0.38337965 1.353456259 -0.49891838  
## [19,] 0.8594701 0.90883394 -1.19197888 -0.16439949 -0.548136727 -0.90537333  
## [20,] 0.8681222 -1.79756140 -2.33572376 1.14482723 -0.173929339 -0.02760849  
## [21,] -0.5547857 2.25096644 1.08890305 0.14883507 1.731984800 -0.42275488  
## [22,] 0.1277548 -2.54297240 0.73073635 -1.23362769 -0.170328076 -1.25559235  
## [23,] 0.1405027 -1.13456140 -0.04941752 -0.53368608 -0.257359551 -0.48017628  
## [24,] -0.8159368 1.64064948 0.71120079 0.95187957 -0.521472137 -0.22850664  
## Comp.7 Comp.8 Comp.9  
## [1,] -0.36306116 -0.082638662 -0.08101268  
## [2,] -0.90947619 0.223054750 -0.09464465  
## [3,] 0.01259105 -0.467699388 -0.09702787  
## [4,] 0.57571526 0.350737893 0.06466320  
## [5,] 0.26971971 -1.170977847 -0.11681904  
## [6,] -0.22733355 -0.526819930 0.10359798  
## [7,] 0.80805487 -0.091579381 0.10775302  
## [8,] 1.38206330 0.147221206 -0.09559560  
## [9,] -0.02974077 0.261150190 0.17494683  
## [10,] 0.10954009 0.097723480 0.05522080  
## [11,] 0.11327506 0.128131637 0.03217421  
## [12,] -0.37706857 0.084436937 -0.17597440  
## [13,] -0.25787633 0.281226211 -0.13580393  
## [14,] 0.61617176 -0.386652496 0.01255589  
## [15,] -0.50871027 -0.095846400 0.48197149  
## [16,] -1.31750472 0.071602032 0.03052020  
## [17,] 0.59360064 0.453031919 0.09110429  
## [18,] 0.88655583 0.492960564 -0.18693709  
## [19,] 0.69790918 -0.254010950 0.12118578  
## [20,] -0.46773096 0.061901191 -0.20700159  
## [21,] -0.32114234 0.174307291 -0.02886754  
## [22,] -0.52726369 -0.128182880 -0.15804309  
## [23,] -0.49895227 -0.002386231 0.06461039  
## [24,] -0.25933596 0.379308863 0.03742340

gu\_pca <- prcomp(gu\_total4, scale. =T)  
biplot(gu\_pca, scale=0)



* 군집 1 (평균소득, 평균부동산, 공공기관\_개수, 가로등\_개수)
* 군집 2 (CCTV\_개수, 벨\_개수, 범죄발생률)
* 군집 3 (평균생활인구)
* 군집 4 (안심스카우트\_이용실적)
* **범죄발생률과 가장 관계가 있는 변수는 CCTV\_개수, 벨\_개수다.**

### 연도별 top5 & bottom5 지역구 비교

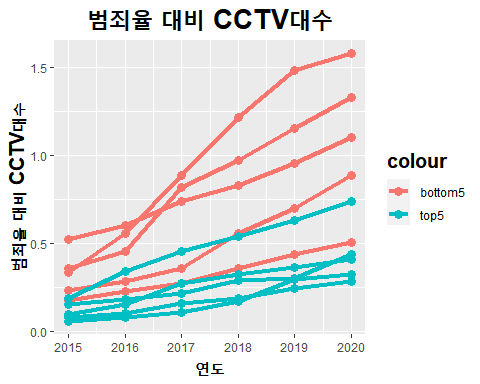
* **상위 5 - 강남, 송파, 강서, 영등포, 서초구**
* **하위 5 - 도봉, 성동, 서대문, 성북, 금천**

### 연도별 |범죄율 상관관계 가설설정

1. CCTV가 많은 지역은 범죄율이 낮을 것이다.
2. 가로등이 많은 지역은 범죄율이 낮을 것이다.
3. 평균집값이 높은 지역은 범죄율이 낮을 것이다.
4. 평균소득이 많은 지역은 범죄율이 낮을 것이다.
5. 안심스카우트 이용이 많은 지역은 범죄율이 낮을 것이다.

### CCTV & 범죄율 추세 파악하기

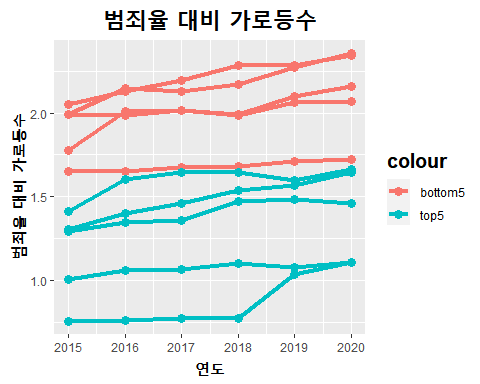
#CCTV 개수/범죄율 통합그래프  
  
#상위 5 - 강남, 송파, 강서, 영등포, 서초구  
#하위 5 - 도봉, 성동, 서대문, 성북, 금천  
  
library(readxl)  
  
cctv\_crime\_top5 <- read\_excel("./Project/USE\_DATA/사용할꺼/cctv\_crime\_top5.xlsx")  
cctv\_crime\_bottom5 <- read\_excel("./Project/USE\_DATA/사용할꺼/cctv\_crime\_bottom5.xlsx")  
  
  
total\_cctv <- cbind(cctv\_crime\_top5, cctv\_crime\_bottom5)  
total\_cctv <- total\_cctv[, -9]  
  
ggplot(total\_cctv, aes(x = 연도)) +   
   
 geom\_point(aes(y = 도봉구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 도봉구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성동구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성동구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 서대문구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 서대문구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성북구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성북구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 금천구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 금천구, colour = "bottom5"), size = 1.5) +  
   
 geom\_point(aes(y = 강남구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강남구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 송파구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 송파구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 강서구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강서구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 영등포구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 영등포구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 서초구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 서초구, colour = "top5"), size = 1.5) +  
   
 labs(x = '연도',  
 y = '범죄율 대비 CCTV대수',  
 title = '범죄율 대비 CCTV대수') +  
 theme(plot.title = element\_text(size = 20, face = "bold", hjust = 0.5),  
 legend.title = element\_text(size = 15, face = "bold"),  
 axis.title = element\_text(size = 15, face = "bold"))



* 범죄율이 높은 지역이 범죄율이 낮은 지역보다 상대적으로 CCTV대수가 적다.
* CCTV가 많은 지역일수록 범죄율이 낮다.

### 가로등 & 범죄율 추세 파악하기

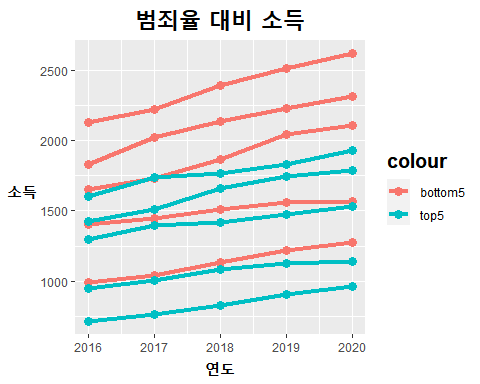
#가로등 개수/범죄율 통합그래프  
#상위 5 - 강남, 송파, 강서, 영등포, 서초구  
#하위 5 - 도봉, 성동, 서대문, 성북, 금천  
  
  
light\_crime\_top5 <- read\_excel("./Project/USE\_DATA/사용할꺼/light\_crime\_top5.xlsx")  
light\_crime\_bottom5 <- read\_excel("./project/USE\_DATA/사용할꺼/light\_crime\_bottom5.xlsx")  
  
  
total\_light <- cbind(light\_crime\_top5, light\_crime\_bottom5)  
total\_light <- total\_light[, -9]  
  
ggplot(total\_light, aes(x = 연도)) +   
   
 geom\_point(aes(y = 도봉구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 도봉구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성동구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성동구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 서대문구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 서대문구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성북구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성북구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 금천구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 금천구, colour = "bottom5"), size = 1.5) +  
   
 geom\_point(aes(y = 강남구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강남구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 송파구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 송파구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 강서구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강서구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 영등포구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 영등포구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 서초구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 서초구, colour = "top5"), size = 1.5) +  
   
 labs(x = '연도',  
 y = '범죄율 대비 가로등수',  
 title = '범죄율 대비 가로등수') +  
 theme(plot.title = element\_text(size = 20, face = "bold", hjust = 0.5),  
 legend.title = element\_text(size = 15, face = "bold"),  
 axis.title = element\_text(size = 15, face = "bold"))



* 범죄율이 높은 지역이 범죄율이 낮은 지역보다 상대적으로 가로등수가 적다.
* 가로등개수가 많은 지역일수록 범죄율이 낮다.

### 소득 & 범죄율 추세 파악하기

#소득/범죄율 통합그래프  
options(scipen = 100)  
  
##cctv  
tax <- read.csv("./Project/USE\_DATA/사용할꺼/연말정산.csv")  
  
#범죄율 데이터 가져오기  
crime <- read\_excel("./Project/USE\_DATA/사용할꺼/범죄율\_total.xls")  
  
tax\_crime <- read\_excel("./Project/USE\_DATA/사용할꺼/tax\_crime.xlsx")  
  
only\_text\_1 <- tax\_crime[1, 1:6]  
only\_text\_2 <- tax\_crime[5, c(1, 7:11)]  
  
  
ggplot(tax\_crime, aes(x = 연도)) +   
   
 geom\_point(aes(y = 도봉구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 도봉구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성동구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성동구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 서대문구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 서대문구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성북구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성북구, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 금천구, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 금천구, colour = "bottom5"), size = 1.5) +  
   
 geom\_point(aes(y = 강남구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강남구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 송파구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 송파구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 강서구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강서구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 영등포구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 영등포구, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 서초구, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 서초구, colour = "top5"), size = 1.5) +  
 labs(x = '연도', y = '소득', title = '범죄율 대비 소득')+  
 theme(plot.title = element\_text(size = 20, face = "bold", hjust = 0.5),  
 legend.title = element\_text(size = 15, face = "bold"),  
 axis.title = element\_text(size = 15, face = "bold"),  
 axis.title.y = element\_text(angle = 0, vjust = 0.5))



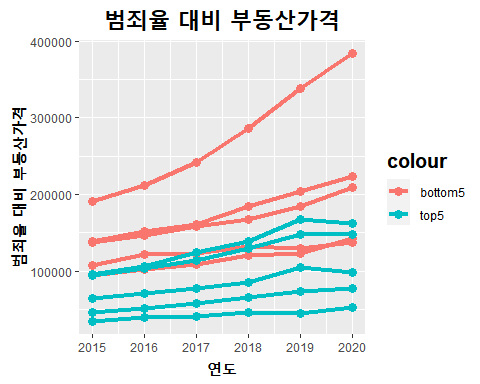
* 범죄율이 높은 지역이 범죄율이 낮은 지역보다 상대적으로 소득이 적다.
* 소득이 많은 지역일수록 범죄율이 낮다.

### 부동산(집값) & 범죄율 추세 파악하기

#부동산(집값)/범죄율 통합그래프  
realty\_crime <- read\_excel("./Project/USE\_DATA/사용할꺼/realty\_crime.xlsx")

## New names:  
## \* `` -> `...1`

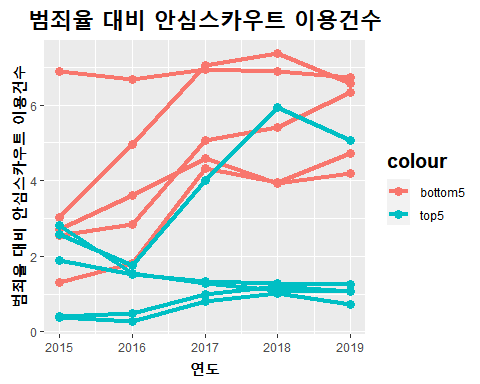
ggplot(realty\_crime, aes(x = 연도)) +   
   
 geom\_point(aes(y = 도봉, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 도봉, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성동, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성동, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 서대문, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 서대문, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성북, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성북, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 금천, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 금천, colour = "bottom5"), size = 1.5) +  
   
 geom\_point(aes(y =강남, colour = "top5"), size = 3) +  
 geom\_line(aes(y =강남, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 송파, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 송파, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 강서, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강서, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 영등포, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 영등포, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 서초, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 서초, colour = "top5"), size = 1.5)+   
 labs(x = '연도',  
 y = '범죄율 대비 부동산가격',  
 title = '범죄율 대비 부동산가격') +  
 theme(plot.title = element\_text(size = 20, face = "bold", hjust = 0.5),  
 legend.title = element\_text(size = 15, face = "bold"),  
 axis.title = element\_text(size = 15, face = "bold"))



* 범죄율이 높은 지역이 범죄율이 낮은 지역보다 상대적으로 부동산가격이 낮다.
* 부동산 가격이 높은 지역일수록 범죄율이 낮다.

### 안심스카우트 & 범죄율 추세 파악하기

#안심스카우트/범죄율 통합그래프  
  
ansim\_crime <- read\_excel("./Project/USE\_DATA/사용할꺼/ansim\_crime.xlsx")  
  
ggplot(ansim\_crime, aes(x = 연도)) +   
   
 geom\_point(aes(y = 도봉, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 도봉, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성동, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성동, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 서대문, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 서대문, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 성북, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 성북, colour = "bottom5"), size = 1.5) +  
 geom\_point(aes(y = 금천, colour = "bottom5"), size = 3) +  
 geom\_line(aes(y = 금천, colour = "bottom5"), size = 1.5) +  
   
 geom\_point(aes(y =강남, colour = "top5"), size = 3) +  
 geom\_line(aes(y =강남, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 송파, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 송파, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 강서, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 강서, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 영등포, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 영등포, colour = "top5"), size = 1.5) +  
 geom\_point(aes(y = 서초, colour = "top5"), size = 3) +  
 geom\_line(aes(y = 서초, colour = "top5"), size = 1.5)+   
 labs(x = '연도',  
 y = '범죄율 대비 안심스카우트 이용건수',  
 title = '범죄율 대비 안심스카우트 이용건수') +  
 theme(plot.title = element\_text(size = 20, face = "bold", hjust = 0.5),  
 legend.title = element\_text(size = 15, face = "bold"),  
 axis.title = element\_text(size = 15, face = "bold"))



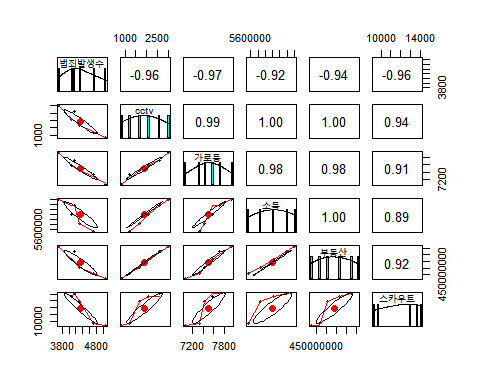
* 범죄율이 높은 지역이 범죄율이 낮은 지역보다 상대적으로 안심스카우트 이용건수가 적다.
* 안심스카우트 이용건수가 많은 지역일수록 범죄율이 낮다.

### 연도별 서울시 전체 요소 평균값(x) 범죄발생수(y) 선형회귀

##x-(cctv,가로등,소득,부동산,여성 스카우트)  
  
##cctv  
cctv <- read\_excel("./Project/USE\_DATA/사용할꺼/CCTV(12~21)\_fix.xlsx")  
cctv\_total <- cctv[,c(1,27)]  
cctv\_total <- dplyr::rename(cctv\_total, cctv = 평균)  
  
##가로등  
light <- read\_excel("./Project/USE\_DATA/사용할꺼/가로등.xlsx")  
light\_total <- light[,c(1,27)]  
light\_total <- dplyr::rename(light\_total, 가로등 = 평균)  
  
##소득  
tax <- read.csv("./Project/USE\_DATA/사용할꺼/tax\_add\_mean.csv")  
tax\_total <- tax[,c(1,27)]  
tax\_total <- dplyr::rename(tax\_total, 소득 = 평균)  
  
##부동산  
land <- read\_excel("./Project/USE\_DATA/사용할꺼/부동산\_평균.xlsx")  
land\_total <- land  
land\_total <- dplyr::rename(land\_total, 부동산 = 서울시)  
  
##여성 스카우트   
scoute <- read\_excel("./Project/USE\_DATA/사용할꺼/안심스카우트\_평균.xlsx")  
scoute\_total <- scoute  
scoute\_total <- dplyr::rename(scoute\_total, 스카우트 = 서울시)  
  
  
##범죄발생수  
crime <- read\_excel("./Project/USE\_DATA/사용할꺼/범죄율\_total.xls")  
crime\_total <- crime[,c(1,27)]  
crime\_total <- dplyr::rename(crime\_total, 범죄발생수 = 평균)  
  
  
### 16~19년도 통합  
total <- left\_join(crime\_total, cctv\_total, by = "연도")  
total <- left\_join(total, light\_total, by = "연도")  
total <- left\_join(total, tax\_total, by = "연도")  
total <- left\_join(total, land\_total, by = "연도")  
total <- left\_join(total, scoute\_total, by = "연도")  
  
  
total\_2 <- scale(total[2:7])  
total\_2 <- total\_2[2:5, ]   
  
total\_2 <- as.data.frame(total)  
  
lm(범죄발생수 ~ cctv + 가로등 + 소득 + 부동산 + 스카우트, data = total)

##   
## Call:  
## lm(formula = 범죄발생수 ~ cctv + 가로등 + 소득 + 부동산 + 스카우트,   
## data = total)  
##   
## Coefficients:  
## (Intercept) cctv 가로등 소득 부동산   
## -1022350.325 114.374 318.258 -0.261 NA   
## 스카우트   
## NA

library(psych)  
total\_2 <- total\_2[,-1]  
pairs.panels(total\_2)



* 범죄발생수와 요소들의 상관관계가 모두 음의 상관관계로 나타나 범죄발생수가 낮아질 수록 요소들의 수,값도 낮아진다.

## 결론도출

평균집값과 평균소득이 늘어남에 따라 범죄율이 줄기는 하지만 실질적으로 해결해 주기는 어렵다고 봅니다. 따라서 상대적으로 범죄율이 높은 곳에 cctv와 가로등과 같은 시설물들을 추가 설치 및 주기적인 장비관리를 통해 범죄예방에 도움을 줄 수있다 생각합니다. 또한, 안심스카우트와 같은 범죄 예방 정책을 확대하여 범죄로부터 안전한 거리와 일상을 만들었으면 합니다 .