[GUN VIOLENCE] Incidents\_by\_state

1585107 이지인

2018년 12월 3일

# "소득이 낮은 지역일수록 총기사고가 더 많이 일어날 것이다."  
# 2017년 미국 주별 연간소득 데이터 불러오기  
names(income)

## [1] "state" "linecode" "Description" "income"

( income <- income %>%   
 filter( linecode == "3") %>%  
 select(state, income) )

## # A tibble: 51 x 2  
## state income  
## <chr> <dbl>  
## 1 Alabama 39976  
## 2 Alaska 56042  
## 3 Arizona 41633  
## 4 Arkansas 40791  
## 5 California 58272  
## 6 Colorado 53504  
## 7 Connecticut 70121  
## 8 Delaware 49125  
## 9 District of Columbia 76986  
## 10 Florida 46858  
## # ... with 41 more rows

# b - state code lat long pop acc acc\_pop  
b

## # A tibble: 51 x 7  
## state code lat long pop acc acc\_pop  
## <chr> <int> <dbl> <dbl> <int> <int> <dbl>  
## 1 Alabama 1 32.8 -86.8 4874747 3804 78  
## 2 Alaska 2 61.4 -152. 739795 1142 154  
## 3 Arizona 4 33.7 -111. 7016270 1621 23  
## 4 Arkansas 5 35.0 -92.4 3004279 2077 69  
## 5 California 6 36.1 -120. 39536653 11439 29  
## 6 Colorado 8 39.1 -105. 5607154 2409 43  
## 7 Connecticut 9 41.6 -72.8 3588184 2320 65  
## 8 Delaware 10 39.3 -75.5 961939 1083 113  
## 9 District of Columbia 11 38.9 -77.0 693972 2177 314  
## 10 Florida 12 27.8 -81.7 20984400 11059 53  
## # ... with 41 more rows

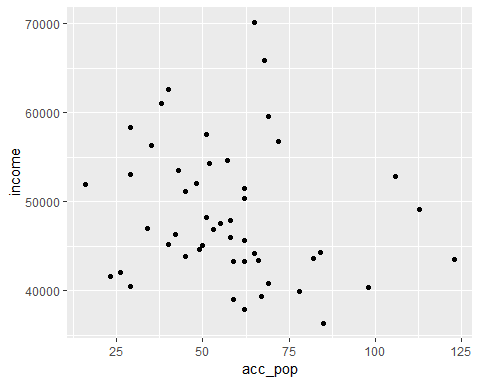
( acc\_inc <- full\_join(b, income, by="state") %>%  
 select( state, acc\_pop, income) )

## # A tibble: 51 x 3  
## state acc\_pop income  
## <chr> <dbl> <dbl>  
## 1 Alabama 78 39976  
## 2 Alaska 154 56042  
## 3 Arizona 23 41633  
## 4 Arkansas 69 40791  
## 5 California 29 58272  
## 6 Colorado 43 53504  
## 7 Connecticut 65 70121  
## 8 Delaware 113 49125  
## 9 District of Columbia 314 76986  
## 10 Florida 53 46858  
## # ... with 41 more rows

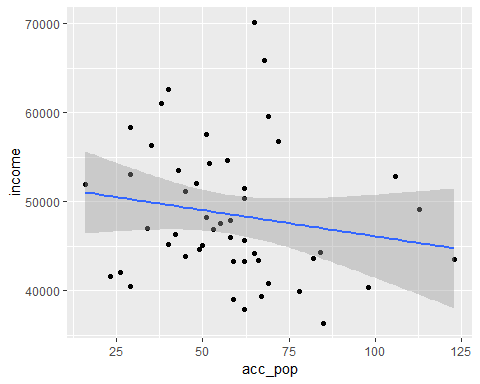
# income 소득과 acc\_pop 인구대비 사건수 상관관계를 확인해보기 위해 산점도를 그려보았다.   
summary(b$acc\_pop)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 16.00 44.00 58.00 64.75 68.50 314.00

acc\_inc %>% filter(acc\_pop<150) %>%  
 ggplot(aes(acc\_pop, income)) +   
 geom\_point()



acc\_inc %>% filter(acc\_pop<150) %>%  
 ggplot(aes(acc\_pop, income)) +   
 geom\_point() +geom\_smooth(method = "lm")



# 산점도와 lm 곡선을 그려본 결과,  
# 소득과 인구대비 총기사고 수간의 상관관계가 없어 보인다.  
# 추가적으로 회귀분석 결과를 확인해보자.  
cor(acc\_inc$income, acc\_inc$acc\_pop, method = "pearson")

## [1] 0.3387797

income.reg <- lm(acc\_inc$acc\_pop ~ acc\_inc$income)  
summary(income.reg)

##   
## Call:  
## lm(formula = acc\_inc$acc\_pop ~ acc\_inc$income)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -53.41 -24.44 -7.35 14.74 200.59   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.182e+01 3.484e+01 -0.626 0.534   
## acc\_inc$income 1.757e-03 6.969e-04 2.521 0.015 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 41.81 on 49 degrees of freedom  
## Multiple R-squared: 0.1148, Adjusted R-squared: 0.09671   
## F-statistic: 6.353 on 1 and 49 DF, p-value: 0.01502

# 회귀 분석 결과 결정계수가 0.097로 매우 작으므로  
# 둘 간의 상관관계가 없다고 할 수 있다.