HW2

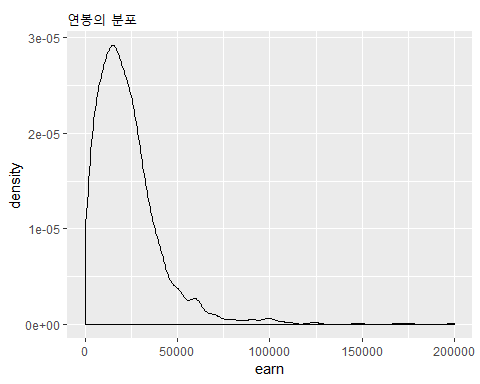
박효선 1585063

2018년 9월 30일

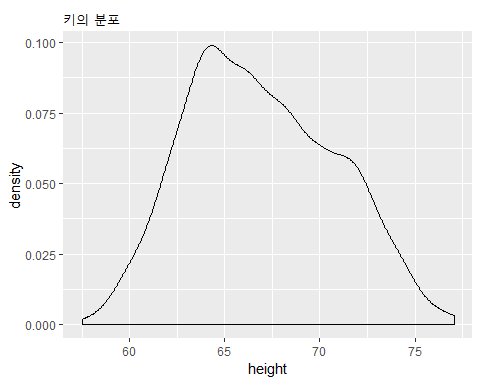
## 1번

### 연봉, 키, 성별, 교육정도, 나이, 인종 변수 각각의 분포

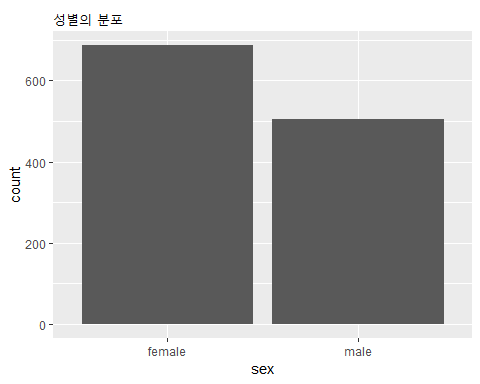
ggplot(Height, aes(earn))+geom\_density()+ggtitle( "연봉의 분포") #연봉의 분포



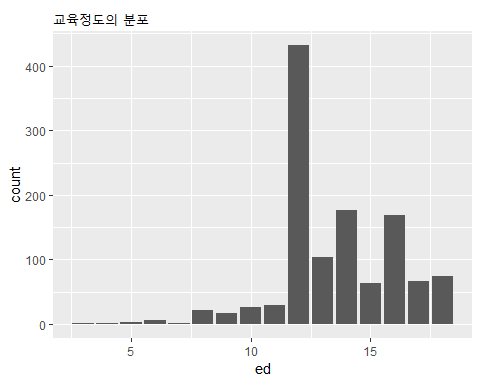
ggplot(Height, aes(height))+geom\_density()+ggtitle("키의 분포") #키의 분포



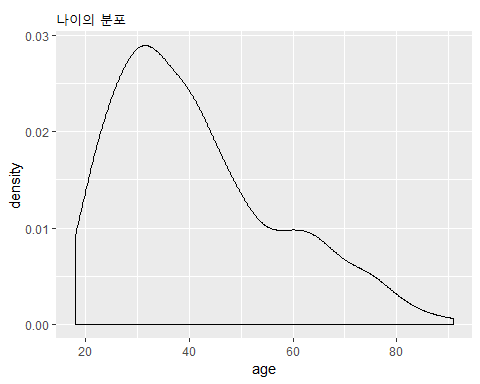
ggplot(Height, aes(sex))+geom\_bar()+ggtitle("성별의 분포") #성별의 분포



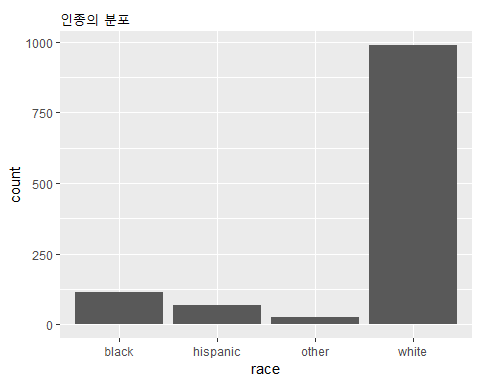
ggplot(Height, aes(ed))+geom\_bar()+ggtitle("교육정도의 분포") #교육정도의 분



ggplot(Height, aes(age))+geom\_density()+ggtitle("나이의 분포") #나이의 분포



ggplot(Height, aes(race))+geom\_bar()+ggtitle("인종의 분포") #인종의 분포

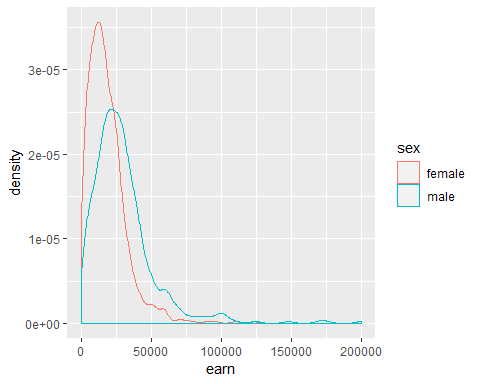


## 2번

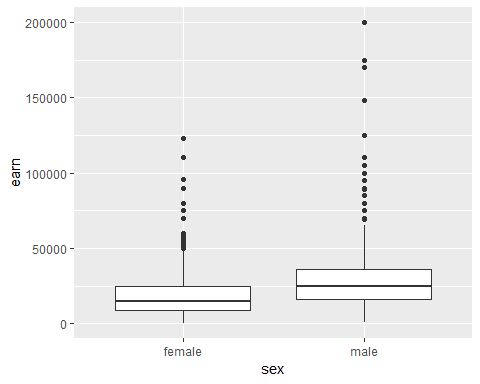
### 성별에 따라 연봉의 차이가 있는가?

성별에 따라 연봉의 차이가 있다. 대체로 여성보다 남성의 소득이 높은 곳에 분포한다..

ggplot(Height, aes(earn, color=sex))+geom\_density()#+facet\_wrap(~sex)



ggplot(Height, aes(sex,earn))+geom\_boxplot()



## 3번

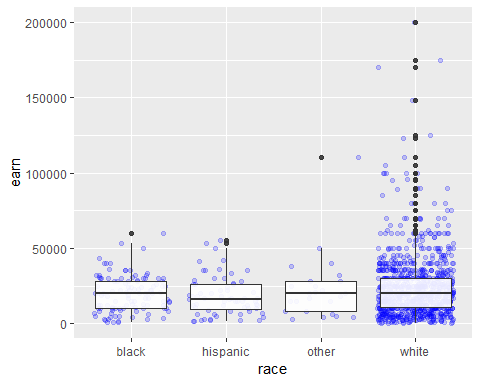
### 인종별로 연봉의 차이가 있는가?

연봉 0~50000 범위에서는 인종의 차이가 크게 두드러지지 않는다. 하지만 연봉 50000 이상의 고소득자의 연봉을 살펴보면 인종의 차이가 있다. 고소득자의 대부분은 white이다.

ggplot(Height, aes(earn, color=race))+geom\_density()#+facet\_wrap(~race)



ggplot(Height, aes(race,earn))+geom\_jitter(alpha=0.2, color='blue')+geom\_boxplot(alpha=0.9)

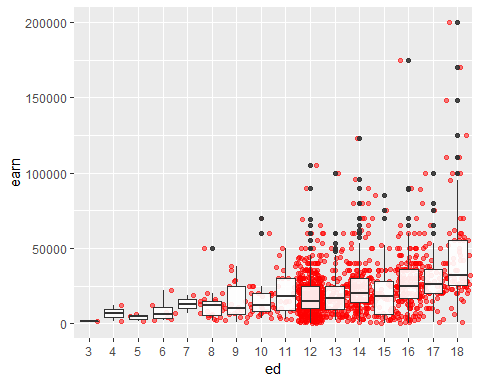


## 4번

### 교육받은 정도와 연봉이 상관성이 있는가?

상관성이 있다. 교육받은 연도가 길수록 연봉이 높다.

Height$ed<-as.factor(Height$ed)  
ggplot(Height, aes(ed, earn))+geom\_jitter( width=0.5, alpha=0.5, color='red')+geom\_boxplot(alpha=0.9)



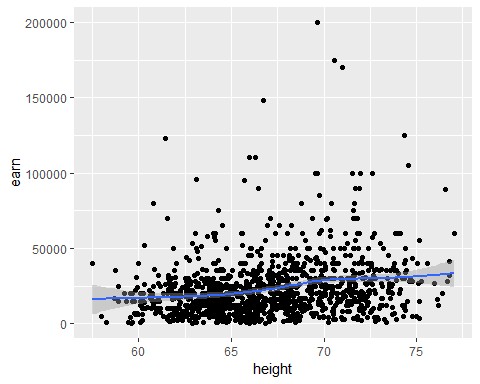
## 5번

### 연봉과 키가 상관성이 있는가?

상관성이 없다.

ggplot(Height, aes( height, earn))+geom\_point()+geom\_smooth()

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

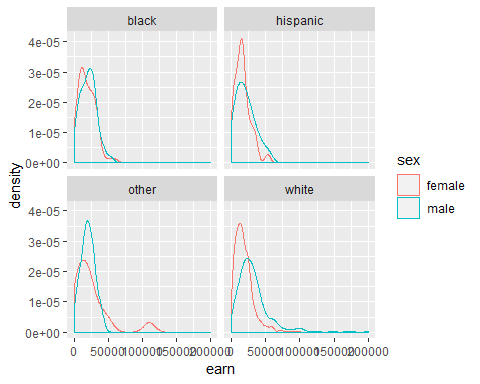


## 6번

### 인종별로 성별에 따라 연봉차이가 있는가?

모든 인종에서 성별에 따른 연봉차이가 있다.

ggplot(Height, aes( earn, color=sex))+geom\_density()+facet\_wrap(~race)



## 7번

### 조사된 자료에서 인종별로 나이의 분포가 같은가?

다르다. 대체로 hispanic 은 다른 인종에 비해 왼쪽…에 분포하고 있다. 다른연령과 다르게 50대의 비율이 낮고, 60대의 비율이 높다. white 는 다른 연령과 달리 75세 이상의 인구가 많다..

ggplot(Height, aes(age, color=race))+geom\_density()+xlim(0,120)

