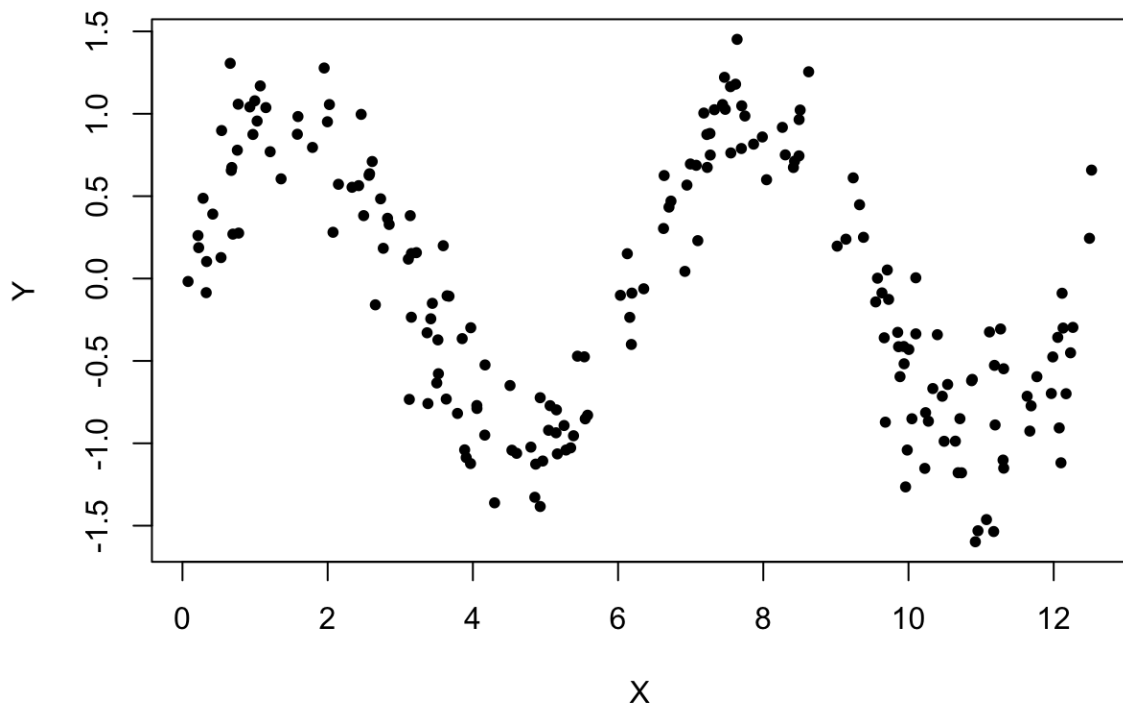
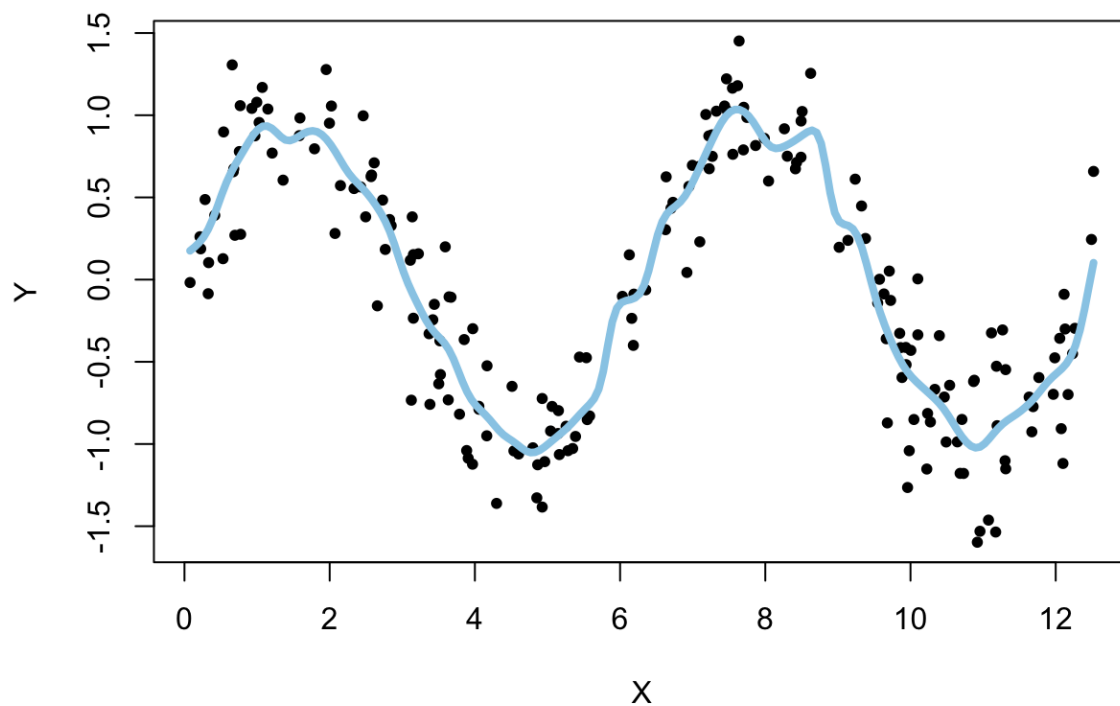


1주차 숙제_Part1

```
### (a) Data generation
X = sort(runif(200,min=0,max=4*pi))
Y = sin(X) + rnorm(200,sd=0.3)
plot(X,Y,pch=20)
```

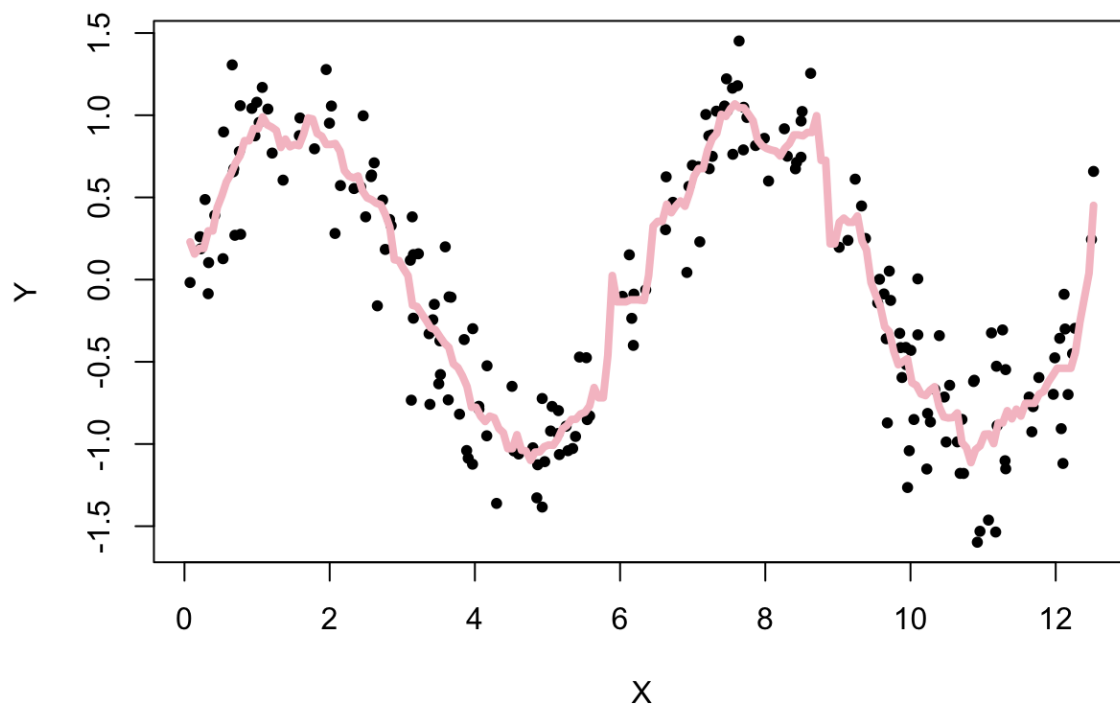


```
Kreg = ksmooth(x=X,y=Y,kernel = 'normal',bandwidth=0.5)
plot(X,Y,pch=20)
lines(Kreg,lwd=4,col='skyblue')
```



Kernel function : Gaussian Kernel

```
## (a) Difference between Uniform Kernel and Gaussain Kernel
Kreg = ksmooth(x=X,y=Y,kernel = 'box',bandwidth=0.5)
plot(X,Y,pch=20)
lines(Kreg,lwd=4,col='pink')
```

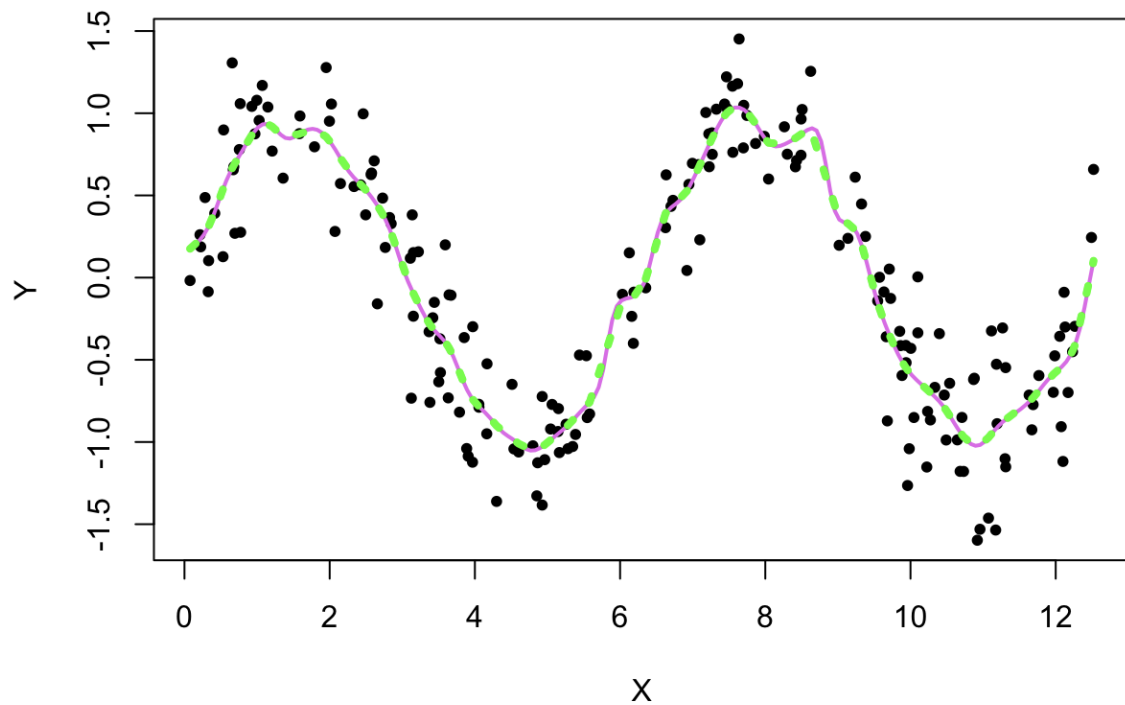


→ Blue(Gaussian Kernel) is more smoother than Pink(uniform kernel) one.

```
## (b) Your own Kernel Function
ksmooth.train = function(x.train, y.train, bandwidth=0.5){
  sigma = 0.25*bandwidth/qnorm(0.75,0,1)
  kern = function(x) dnorm(x,0,sigma)
  yhat.train = numeric(length(x.train))
  for (i in 1:length(x.train)){
    yhat.train[i] = sum(y.train*kern(x.train[i]-x.train))/sum(kern(x.train[i]-x.train))
  }
  ksmooth.train.out = cbind(x.train,yhat.train)

  return(ksmooth.train.out)
}
```

```
## (c) Check
Kreg = ksmooth(x=X,y=Y,kernel='normal',bandwidth=0.5)
myKreg = ksmooth.train(x.train=X,y.train=Y,bandwidth=0.5)
plot(X,Y,pch=20)
lines(Kreg,lwd=2,col='violet')
lines(myKreg,lty=3,lwd=4,col='green')
```



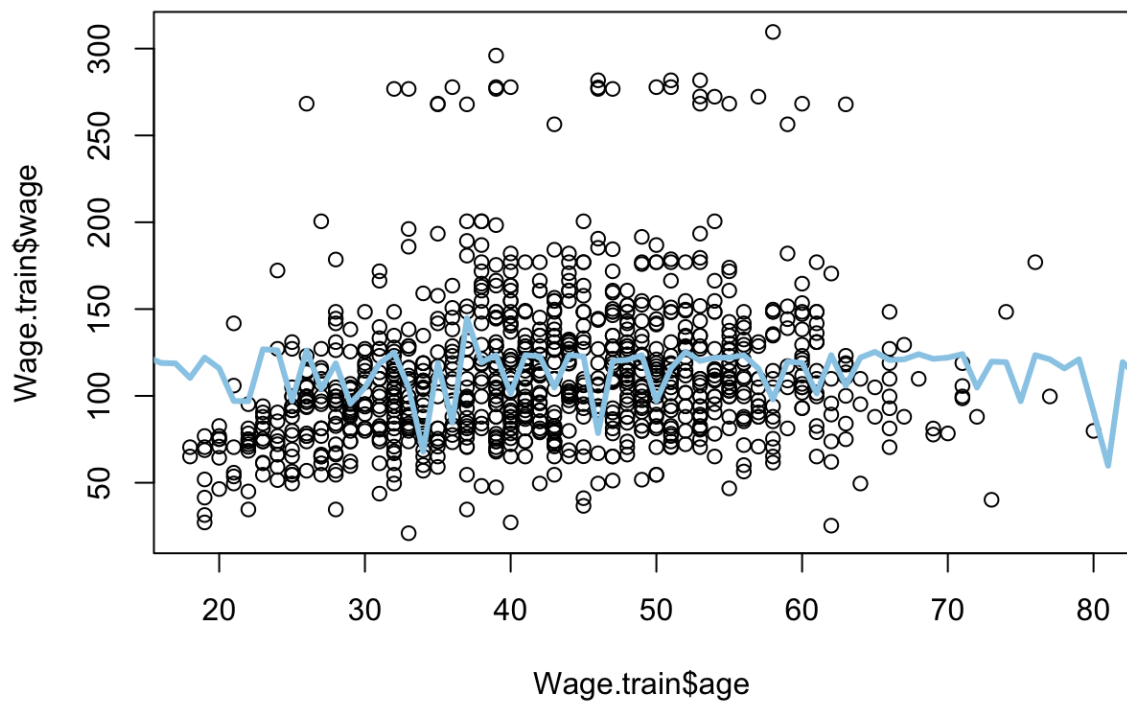
: Gaussian Kernel 을 사용하고 생성한 값들을 yhat에 저장하고 scatter

```
## (d) Application
source('home1-part1-data.R')
smooth = ksmooth.train(Wage.train$age,Wage.train$wage,bandwidth=3)
age.train=smooth[,1]
wage.train = smooth[,2]
RSS.train = sum((Wage.train$wage - wage.train)^2)
cat("RSS.train : ", RSS.train)

RSS.train : 1625121
```

해당 train 에 대한 Residual sum of squares : 1625121

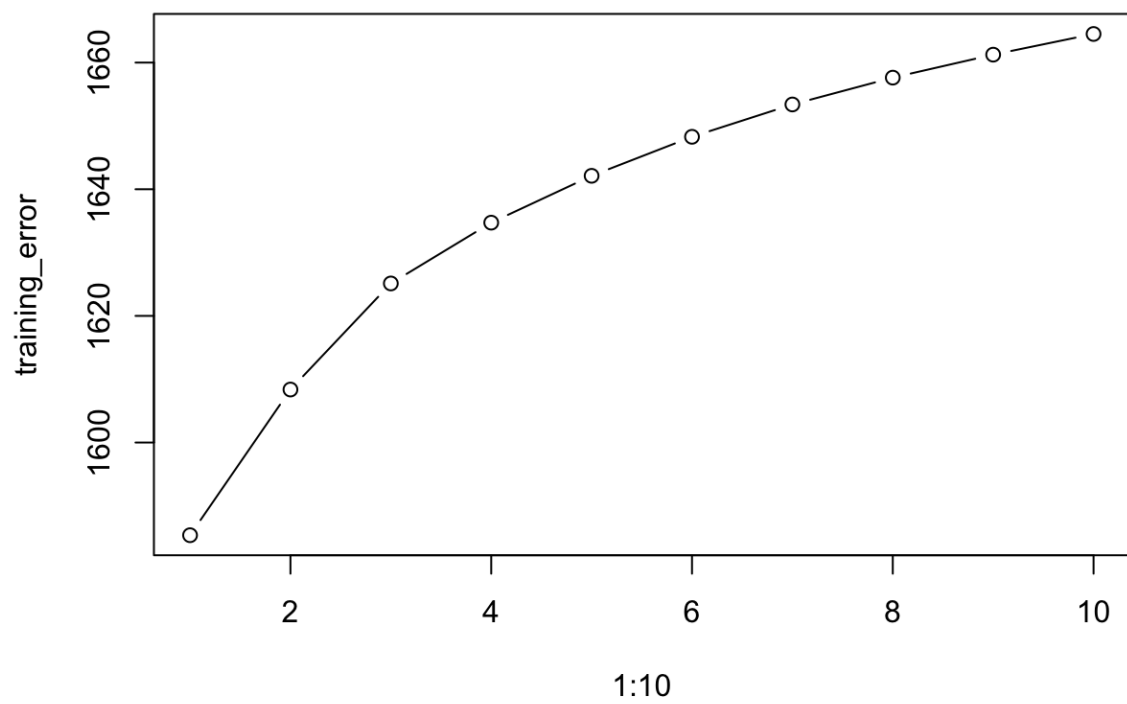
```
### Plotting
plot(Wage.train$age,Wage.train$wage)
lines(wage.train,col='skyblue',lwd=3)
```



```
### Training error Plot
training_error = numeric(10)
for (i in 1:10) {
  trained = ksmooth.train(Wage.train$age, Wage.train$wage, bandwidth=i)
  training_error[i] = sum((Wage.train$wage - trained[,2])^2)/length(trained[,2])
}
print(training_error)

[1] 1585.364 1608.370 1625.121 1634.722 1642.120 1648.282 1653.387 [8] 1657.624 1661.252 1664.519
```

```
plot(1:10, training_error, type='b')
```



: bandwidth 가 1에서 10으로 증가함에 따라 증가하는 training error