Assignment #3: Test MSE and Model Complexity

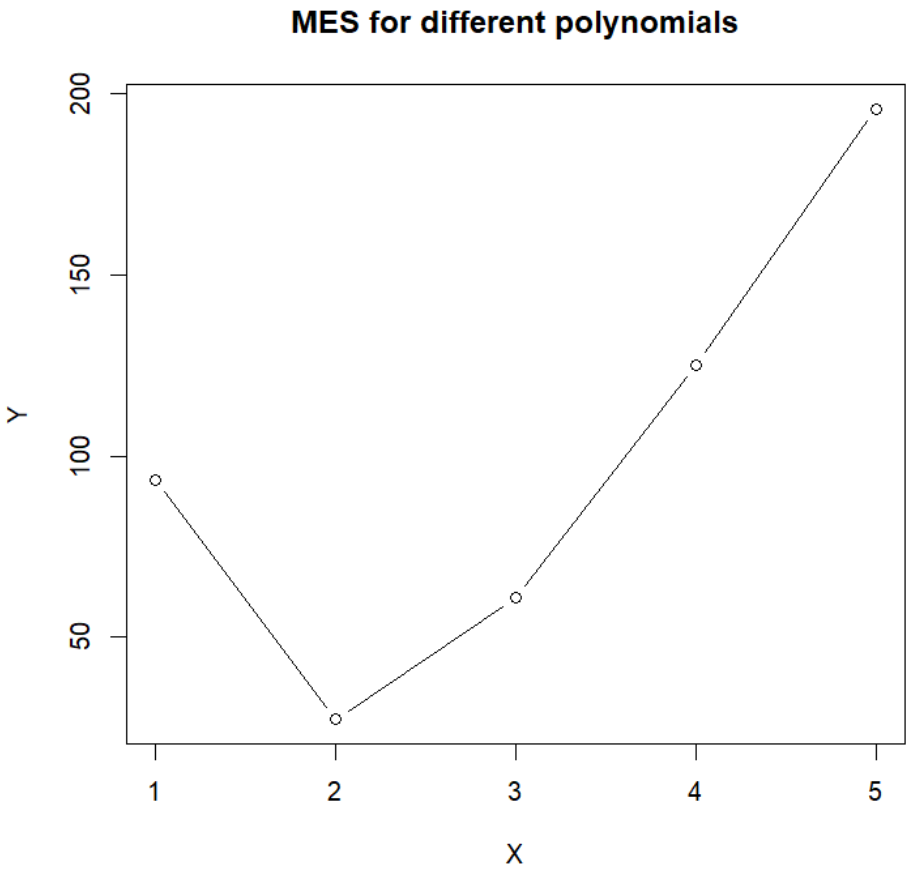
Coding :

```
#Remember to change working directory to where you are storing Assignment3Data.RData
setwd("C:/Users/doncorleron/Documents/coding_r/413hw/hw3")
#Be sure to keep Rdata file in the specified working directory
load("Assignment 3.RData")

### display the dataset, check the dimension of the dataset
train
test
View(train)
View(test)
```

```
Y = c() #make a Y vector for storing value of MES for 5 different polynomials
for( i in 1:5){ #making a for loop to calculate different polynomials
model <- lm(y ~ I(x^i),data=train)

### you can look up for details of the model
#summary(model)



### use this code to fit data in test set
pred=predict(model,test)

### compute the test error
msetest=mean((test$y-pred)^2)
Y <- append(Y,msetest)
msetest}
```

```
X = c(1,2,3,4,5)#making X vector for MES
plot(Y~X, type = 'b', main = "MES for different polynomials")
#plot picture for test MES
```

Plot:



Analysis: As a data scientist, I would prefer with polynomials of degree 2. Because when the degree is 2, it has lowest test mean squared error. And for the best prediction performance, we will select the method that will lead lowest test MSE. Therefore when the degree is 2, the prediction will be the most accuracy one.