

```
# VIDEO 4
```

```
# Read in data
```

```
wine = read.csv("wine.csv")
```

```
str(wine)
```

```
summary(wine)
```

```
# Linear Regression (one variable)
```

```
model1 = lm(Price ~ AGST, data=wine)
```

```
summary(model1)
```

```
# Sum of Squared Errors
```

```
model1$residuals
```

```
SSE = sum(model1$residuals^2)
```

```
SSE
```

```
# Linear Regression (two variables)
```

```
model2 = lm(Price ~ AGST + HarvestRain, data=wine)
```

```
summary(model2)
```

```
# Sum of Squared Errors
```

```
SSE = sum(model2$residuals^2)
```

```
SSE
```

```
# Linear Regression (all variables)
```

```
model3 = lm(Price ~ AGST + HarvestRain + WinterRain + Age +  
  FrancePop, data=wine)
```

```
summary(model3)
```

```
# Sum of Squared Errors
```

```
SSE = sum(model3$residuals^2)
```

```
SSE
```

```
# VIDEO 5
```

```
# Remove FrancePop
```

```
model4 = lm(Price ~ AGST + HarvestRain + WinterRain + Age, data=wine)
```

```
summary(model4)
```

```
# Correlations
```

```
cor(wine$WinterRain, wine$Price)
```

```
cor(wine$Age, wine$FrancePop)
```

```
cor(wine)
```

```
# Remove Age and FrancePop
```

```
model5 = lm(Price ~ AGST + HarvestRain + WinterRain, data=wine)
```

```
summary(model5)
```

```
# VIDEO 6
```

```
# Read in test set
```

```
wineTest = read.csv("wine_test.csv")  
str(wineTest)
```

```
# Make test set predictions
```

```
predictTest = predict(model4, newdata=wineTest)  
predictTest
```

```
# Compute R-squared
```

```
SSE = sum((wineTest$Price - predictTest)^2)  
SST = sum((wineTest$Price - mean(wine$Price))^2)  
1 - SSE/SST
```