```
# VIDEO 1

# Read in the data
NBA = read.csv("NBA_train.csv")
str(NBA)


# VIDEO 2

# How many wins to make the playoffs?
table(NBA$W, NBA$Playoffs)

# Compute Points Difference
NBA$PTSdiff = NBA$PTS - NBA$oppPTS

# Check for linear relationship
plot(NBA$PTSdiff, NBA$W)

# Linear regression model for wins
WinsReg = lm(W ~ PTSdiff, data=NBA)
summary(WinsReg)


# VIDEO 3

# Linear regression model for points scored
PointsReg = lm(PTS ~ X2PA + X3PA + FTA + AST + ORB + DRB + BLK + TOV
 + STL, data=NBA)
summary(PointsReg)

# Sum of Squared Errors
PointsReg$residuals
SSE = sum(PointsReg$residuals^2)
SSE

# Root mean squared error
RMSE = sqrt(SSE/nrow(NBA))
RMSE

# Remove insignifcant variables
PointsReg3 = lm(PTS ~ X2PA + X3PA + FTA + AST + ORB + STL, data=NBA)
summary(PointsReg3)

# Compute SSE and RMSE again
SSE = sum(PointsReg3$residuals^2)
SSE
RMSE = sqrt(SSE/nrow(NBA))
RMSE
```

```r
# Check for correlations
RegVar = NBA[c("X2PA", "X3PA", "FTA", "AST", "ORB", "STL")]
cor(RegVar)


# VIDEO 4

# Read in test set
NBA_test = read.csv("NBA_test.csv")

# Make predictions on test set
PointsPredictions = predict(PointsReg, newdata=NBA_test)

# Compute out-of-sample R^2
SSE = sum((PointsPredictions - NBA_test$PTS)^2)
SST = sum((mean(NBA$PTS) - NBA_test$PTS)^2)
R2 = 1 - SSE/SST
R2

# Compute the RMSE
RMSE = sqrt(SSE/nrow(NBA_test))
RMSE
```