# Upload data and local regression library

training\_data <- read.csv(‘pbp\_model\_training\_data.csv’)

test\_data <- read.csv(‘pbp\_model\_test\_data.csv’)

library(‘locfit’)

# Fit local binomial logistic regression

locfit(formula = game\_result ~ lp(seconds\_remaining) + lp(score\_difference) +

lp(num\_home\_stars) + lp(num\_away\_stars) + possession, data = training\_data, family = "binomial")

# Add column to test data for predicted probabilities

pred\_test <- predict(fit, newdata=test\_data, se.fit=TRUE)

test\_data$win\_prob <- pred\_test

# Create data visualization

no\_star\_diff <- subset(test\_data, star\_difference==0)

no\_star\_diff <- subset(no\_star\_diff, possession==1)

down\_five <- subset(no\_star\_diff, score\_difference==-5)

tied <- subset(no\_star\_diff, score\_difference==0)

up\_five <- subset(no\_star\_diff, score\_difference==5)

library(‘ggplot2’)

plot <- ggplot() + geom\_line(data=down\_five, aes(seconds\_in, win\_prob), color='red') + geom\_line(data=up\_five, aes(seconds\_in, win\_prob), color='green') + geom\_line(data=tied, aes(seconds\_in, win\_prob), color='blue')

# For question 3a, I created hypothetical data in Excel to model the

situation given.

# Score difference > 10

scen\_a\_df <- read.csv("scenario\_a.csv")

scen\_a\_df$win\_prob <- pred\_a$fit

subs1 <- subset(test\_data, score\_difference > 10)

> minus\_one <- subset(subs1, star\_difference==-1)

> same\_stars <- subset(subs1, star\_difference==0)

> plus\_one <- subset(subs1, star\_difference==1)

> plot\_stars <- ggplot() + geom\_line(data=minus\_one, aes(seconds\_in, win\_prob), color='red') + geom\_line(data=same\_stars, aes(seconds\_in, win\_prob), color='green') + geom\_line(data=plus\_one, aes(seconds\_in, win\_prob), color='blue')

> plot\_stars

# Score difference > 20

> subs1 <- subset(test\_data, score\_difference > 20)

> plus\_one <- subset(subs1, star\_difference==1)

> same\_stars <- subset(subs1, star\_difference==0)

> minus\_one <- subset(subs1, star\_difference==-1)

> plot\_stars <- ggplot() + geom\_line(data=minus\_one, aes(seconds\_in, win\_prob), color='red') + geom\_line(data=same\_stars, aes(seconds\_in, win\_prob), color='green') + geom\_line(data=plus\_one, aes(seconds\_in, win\_prob), color='blue')

> plot\_stars

# 240 < seconds remaining < 360

subs2 <- subset(test\_data, seconds\_remaining < 360 & seconds\_remaining > 240)

minus\_one\_score <- subset(subs2, star\_difference==-1)

> even\_score <- subset(subs2, star\_difference==0)

> plus\_one\_score <- subset(subs2, star\_difference==1)

> plot\_score <- ggplot() + geom\_line(data=minus\_one\_score, aes(score\_difference, win\_prob), color='red') + geom\_line(data=even\_score, aes(score\_difference, win\_prob), color='green') + geom\_line(data=plus\_one\_score, aes(score\_difference, win\_prob), color='blue')

> plot\_score