STAT 351 Homework 5

Charles Hwang

3/27/2020

Charles Hwang

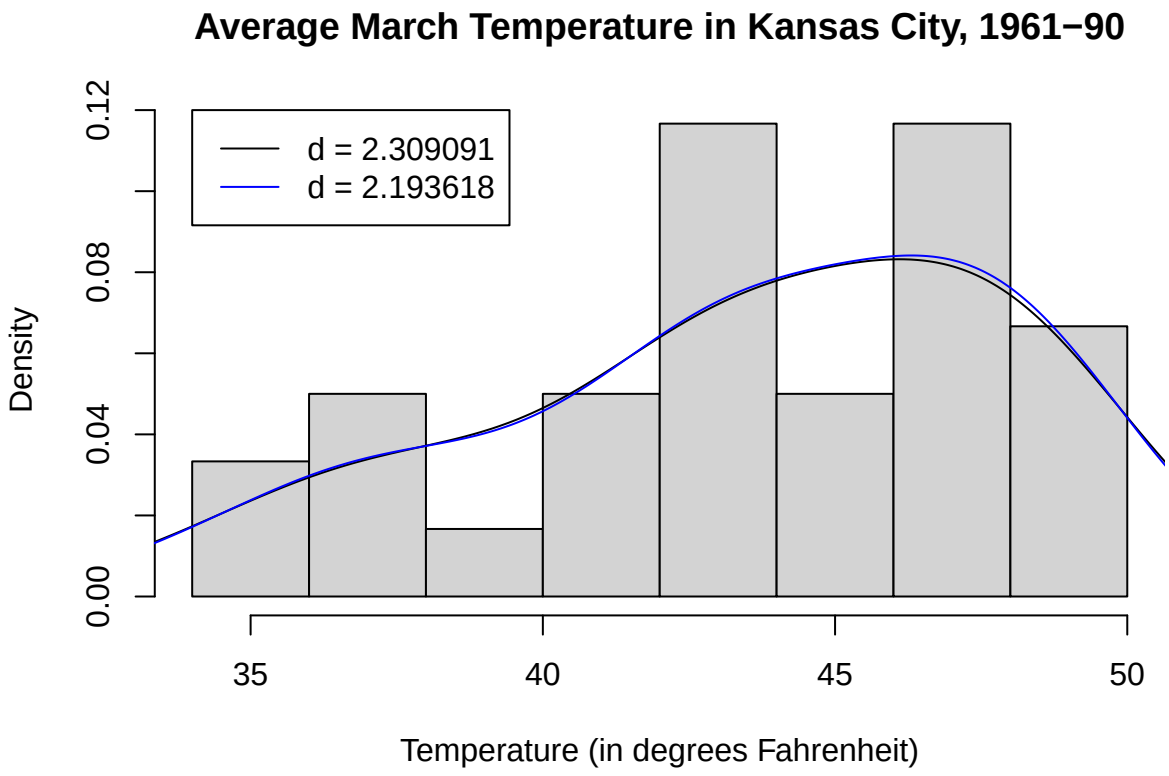
Professor Matthews

STAT 351-001

30 March 2020

Problem 1

```
rm(list=ls())
kcTemp <- c(43.8,40.1,49.2,41.8,34.0,49.1,47.8,48.1,37.6,42.0,43.7,47.1,47.7,46.9,36.5,45.0,48.0,37.6,42
hist(kcTemp,freq=FALSE,main="Average March Temperature in Kansas City, 1961-90",xlab="Temperature (in de
legend(34,.12,c("d = 2.309091","d = 2.193618"),lwd="1",col=c("black","blue"))
points(density(kcTemp,bw=1.06*sd(kcTemp)/length(kcTemp)^.2)$x,density(kcTemp,bw=1.06*sd(kcTemp)/length(k
points(density(kcTemp,bw=1.06*IQR(kcTemp)/(1.34*length(kcTemp)^.2))$x,density(kcTemp,bw=1.06*IQR(kcTemp)
```




```
cat("Either one of Hardle's two standards for d works well for this histogram.")
```

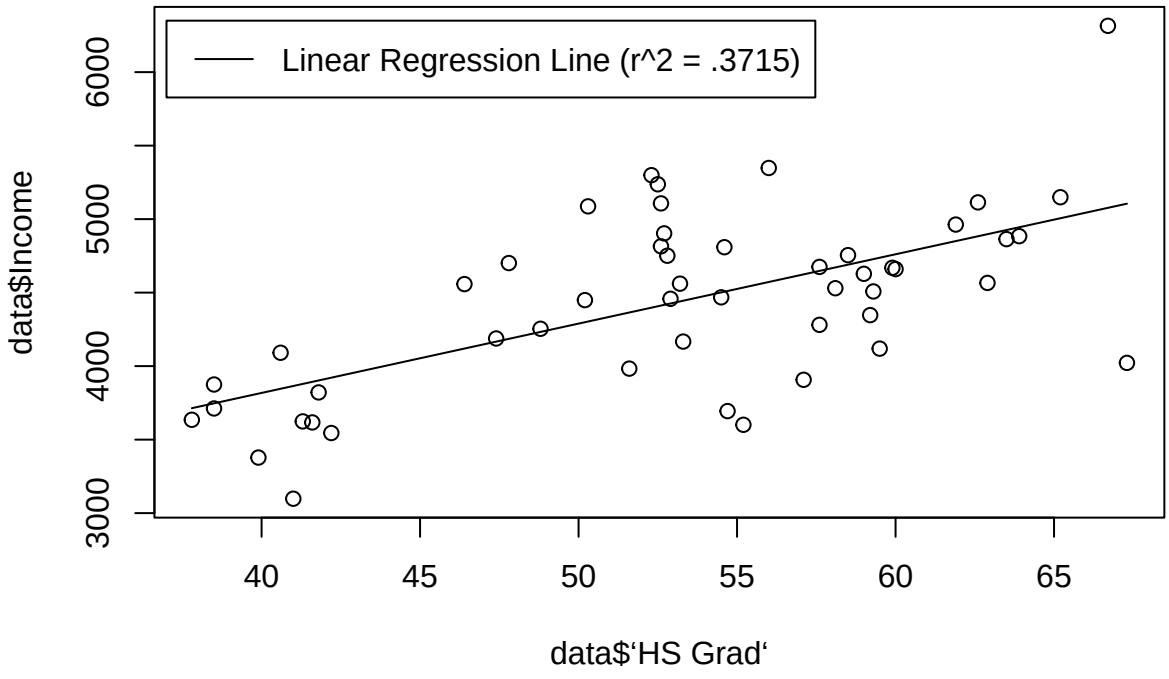
```
## Either one of Hardle's two standards for d works well for this histogram.
```

Problem 2

```
rm(list=ls())
data <- as.data.frame(state.x77)
plot(data$Income~data$`HS Grad`)                              # Problem 2a
summary(lm(data$Income~data$`HS Grad`))                       # Problem 2b
```

```
##
## Call:
## lm(formula = data$Income ~ data$`HS Grad`)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1083.13  -277.41   -34.15   241.46  1238.17
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1931.105    462.739   4.173 0.000125 ***
## data$`HS Grad`   47.162      8.616   5.474 1.58e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 487.1 on 48 degrees of freedom
## Multiple R-squared:  0.3843, Adjusted R-squared:  0.3715
## F-statistic: 29.96 on 1 and 48 DF,  p-value: 1.579e-06
```

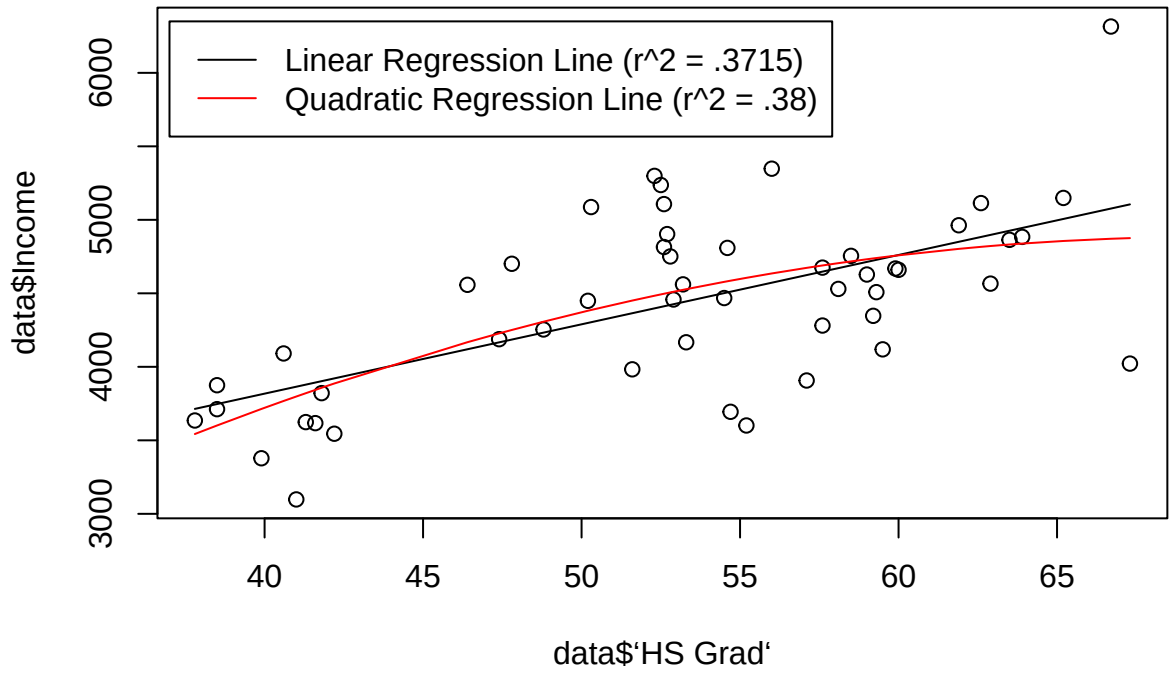
```
plot(data$Income~data$`HS Grad`)
legend(37,6350,"Linear Regression Line (r^2 = .3715)",lwd="1")
lines(sort(data$`HS Grad`),fitted(lm(data$Income~data$`HS Grad`))[order(data$`HS Grad`)])
```



```
summary(lm(data$Income~data$`HS Grad`+I(data$`HS Grad`^2))) # Problem 2c
```

```
##
## Call:
## lm(formula = data$Income ~ data$`HS Grad` + I(data$`HS Grad`^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1004.45  -249.87   -25.59   288.05  1443.79
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1505.424   2708.032  -0.556   0.5809
## data$`HS Grad`        183.196    105.988   1.728   0.0905 .
## I(data$`HS Grad`^2)    -1.313      1.020  -1.288   0.2042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 483.8 on 47 degrees of freedom
## Multiple R-squared:  0.4053, Adjusted R-squared:   0.38
## F-statistic: 16.02 on 2 and 47 DF,  p-value: 4.967e-06
```

```
plot(data$Income~data$`HS Grad`)
legend(37,6350,c("Linear Regression Line (r^2 = .3715)","Quadratic Regression Line (r^2 = .38)"),lwd="1
lines(sort(data$`HS Grad`),fitted(lm(data$Income~data$`HS Grad`))[order(data$`HS Grad`)])
lines(sort(data$`HS Grad`),fitted(lm(data$Income~data$`HS Grad`+I(data$`HS Grad`^2)))[order(data$`HS Gra
```

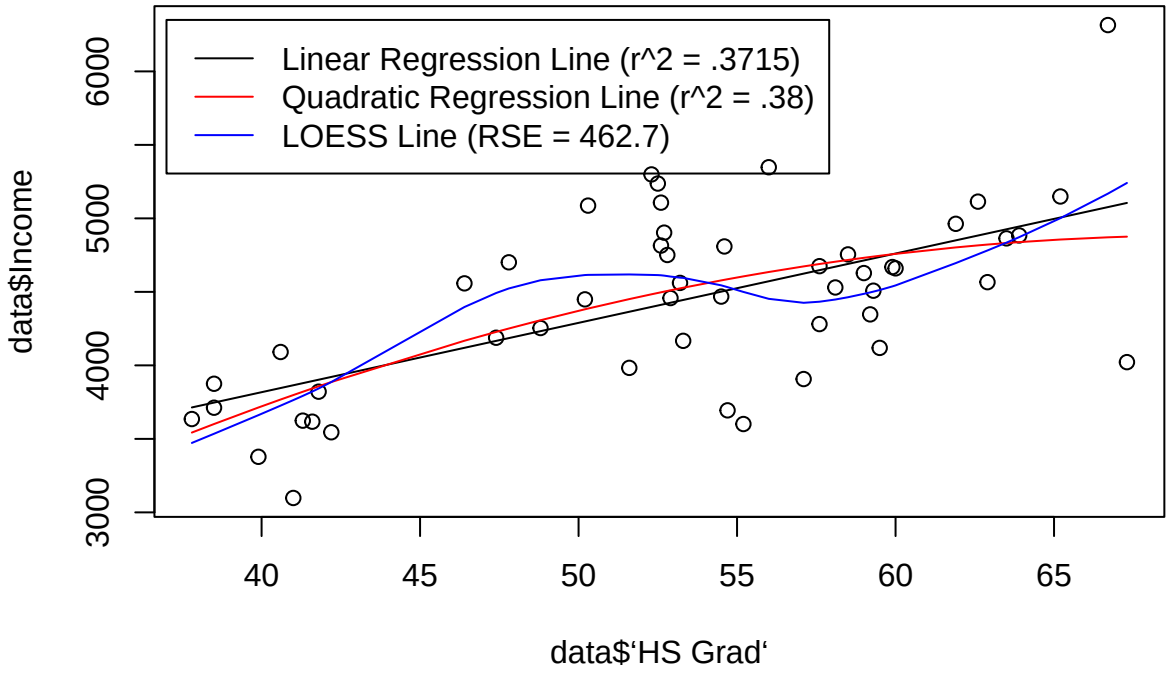



```
summary(loess(data$Income~data$`HS Grad`))              # Problem 2d
```

```
## Call:
## loess(formula = data$Income ~ data$`HS Grad`)
##
## Number of Observations: 50
```

```
## Equivalent Number of Parameters: 4.68
## Residual Standard Error: 462.7
## Trace of smoother matrix: 5.13  (exact)
##
## Control settings:
##   span     :  0.75
##   degree   :  2
##   family   :  gaussian
##   surface  :  interpolate      cell = 0.2
##   normalize:  TRUE
##  parametric:  FALSE
## drop.square:  FALSE
```

```
plot(data$Income~data$`HS Grad`)
legend(37,6350,c("Linear Regression Line (r^2 = .3715)","Quadratic Regression Line (r^2 = .38)","LOESS
lines(sort(data$`HS Grad`),fitted(lm(data$Income~data$`HS Grad`))[order(data$`HS Grad`)])
lines(sort(data$`HS Grad`),fitted(lm(data$Income~data$`HS Grad`+I(data$`HS Grad`^2)))[order(data$`HS Gra
lines(sort(data$`HS Grad`),fitted(loess(data$Income~data$`HS Grad`))[order(data$`HS Grad`)],col="blue")
```




```
# "lines(sort(data$`HS Grad`),loess(data$Income~data$`HS Grad`)$fitted[order(data$`HS Grad`)],col="blue
cat("The linear and quadratic lines are very similar and fit the data with weak to fair strength. The L
```

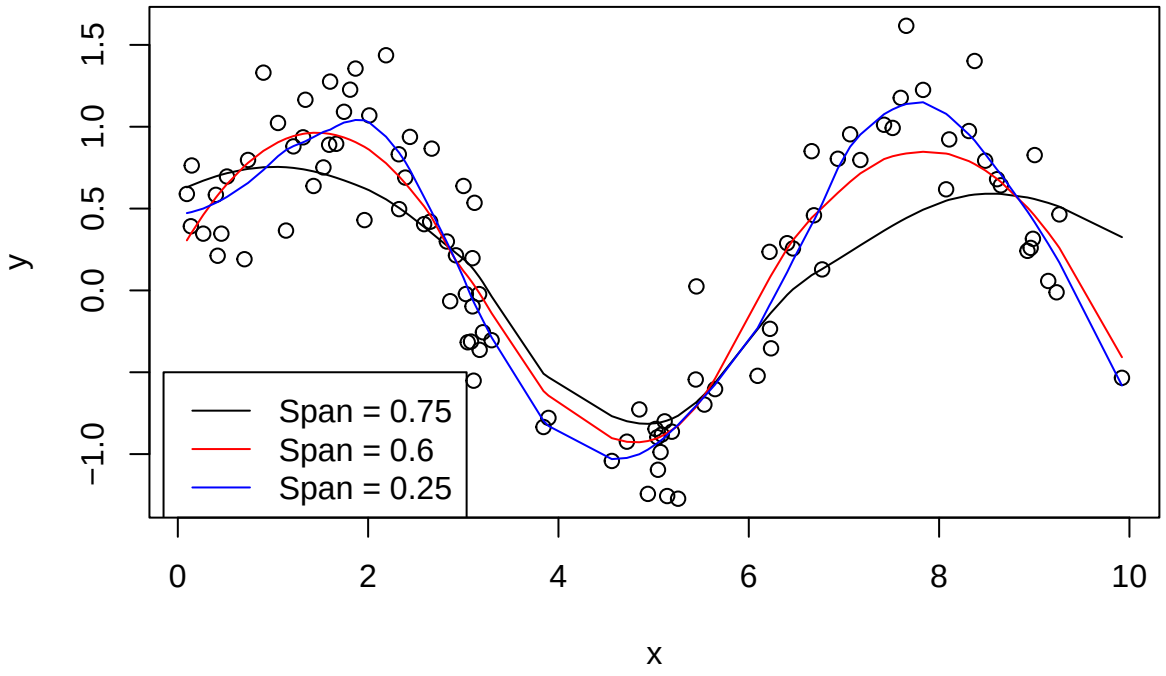
```
## The linear and quadratic lines are very similar and fit the data with weak to fair strength. The LOES
```

Problem 3

```
rm(list=ls())
set.seed(1234)
x <- runif(100,0,10)
y <- sin(x) + rnorm(100,0,.3)
plot(x,y)
summary(loess(y~x))
```

```
## Call:
## loess(formula = y ~ x)
##
## Number of Observations: 100
## Equivalent Number of Parameters: 4.31
## Residual Standard Error: 0.415
## Trace of smoother matrix: 4.7  (exact)
##
## Control settings:
##   span     :  0.75
##   degree   :  2
##   family   :  gaussian
##   surface  :  interpolate      cell = 0.2
##   normalize:  TRUE
##  parametric:  FALSE
## drop.square:  FALSE
```

```
plot(x,y)
legend(-.15,-.5,c("Span = 0.75","Span = 0.6","Span = 0.25"),lwd="1",col=c("black","red","blue"))
lines(sort(x),fitted(loess(y~x))[order(x)])
lines(sort(x),fitted(loess(y~x,span=.6))[order(x)],col="red")
lines(sort(x),fitted(loess(y~x,span=.25))[order(x)],col="blue")
```




```
cat("It appears that [0.25, 0.6] would be a good range of spans to construct a LOESS model for this data
```

```
## It appears that [0.25, 0.6] would be a good range of spans to construct a LOESS model for this data.
```