Shopping

## Read in Data

Data<-read.csv("Shopping.csv")  
summary(Data)

## X Spend Age Sex Height   
## Min. : 1.0 Min. : 39.97 Min. :20.07 F:490 Min. :147.3   
## 1st Qu.: 250.8 1st Qu.: 80.61 1st Qu.:31.01 M:510 1st Qu.:162.7   
## Median : 500.5 Median : 94.53 Median :43.19 Median :168.6   
## Mean : 500.5 Mean : 94.68 Mean :44.00 Mean :169.3   
## 3rd Qu.: 750.2 3rd Qu.:109.04 3rd Qu.:56.72 3rd Qu.:175.2   
## Max. :1000.0 Max. :174.61 Max. :69.98 Max. :198.3   
## Loyalty FamilySize MacBeth   
## No :285 Min. :1.000 No :810   
## Yes:715 1st Qu.:2.000 Yes:190   
## Median :3.000   
## Mean :2.989   
## 3rd Qu.:4.000   
## Max. :9.000

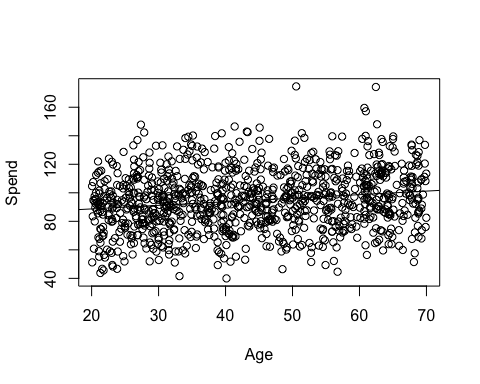
## Simple One Variable Linear Regression

### Spend as a function of Age

SpendAge = lm(Spend ~ Age, data = Data)   
  
summary(SpendAge)

##   
## Call:  
## lm(formula = Spend ~ Age, data = Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -53.750 -14.119 -0.349 13.229 78.293   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 83.70443 2.07309 40.38 < 2e-16 \*\*\*  
## Age 0.24951 0.04471 5.58 3.09e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.66 on 998 degrees of freedom  
## Multiple R-squared: 0.03026, Adjusted R-squared: 0.02929   
## F-statistic: 31.14 on 1 and 998 DF, p-value: 3.09e-08

plot(Spend ~ Age, data = Data)   
abline(lm(Spend~Age, data = Data))

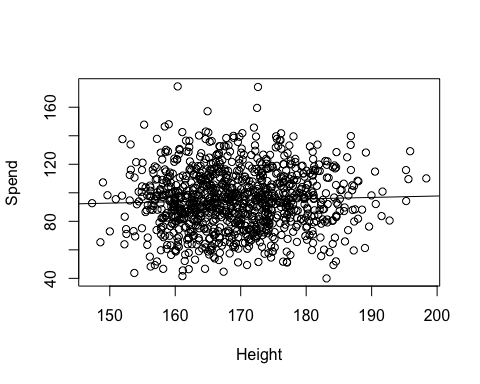


### Spend as a function of Height

SpendHeight = lm(Spend ~ Height, data = Data)   
  
summary(SpendHeight)

##   
## Call:  
## lm(formula = Spend ~ Height, data = Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -56.09 -14.23 -0.08 13.81 80.82   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 77.79789 13.04505 5.964 3.42e-09 \*\*\*  
## Height 0.09973 0.07694 1.296 0.195   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.96 on 998 degrees of freedom  
## Multiple R-squared: 0.00168, Adjusted R-squared: 0.0006801   
## F-statistic: 1.68 on 1 and 998 DF, p-value: 0.1952

plot(Spend ~ Height, data = Data)   
abline(lm(Spend~Height, data = Data))



## Two Variable Linear Regression

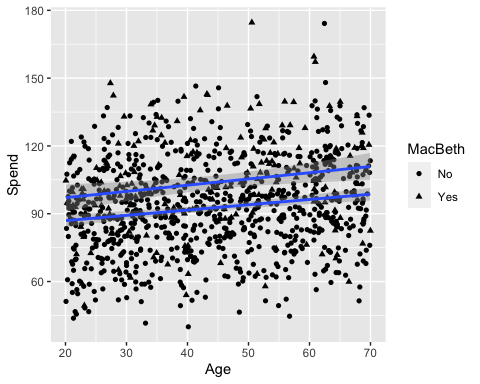
### Spend as a function of Age and MacBeth

library(ggplot2)  
SpendMacBeth = lm(Spend~Age+MacBeth,data=Data)  
summary(SpendMacBeth)

##   
## Call:  
## lm(formula = Spend ~ Age + MacBeth, data = Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -51.651 -12.943 -0.109 12.747 77.174   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 81.89428 2.04370 40.071 < 2e-16 \*\*\*  
## Age 0.24231 0.04372 5.542 3.83e-08 \*\*\*  
## MacBethYes 11.19535 1.62873 6.874 1.10e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.2 on 997 degrees of freedom  
## Multiple R-squared: 0.07414, Adjusted R-squared: 0.07228   
## F-statistic: 39.92 on 2 and 997 DF, p-value: < 2.2e-16

ggplot(Data, aes(x = Age, y = Spend,shape=MacBeth)) +   
 geom\_point() +  
 stat\_smooth(method = "lm")

## `geom\_smooth()` using formula 'y ~ x'



## Multiple Variable Linear Regression

SpendAll<-lm(Spend~Age+MacBeth+Loyalty+factor(FamilySize),data=Data)  
summary(SpendAll)

##   
## Call:  
## lm(formula = Spend ~ Age + MacBeth + Loyalty + factor(FamilySize),   
## data = Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -26.2969 -9.9341 -0.1485 9.9233 26.1855   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 51.37595 1.64911 31.154 < 2e-16 \*\*\*  
## Age 0.19510 0.02592 7.526 1.17e-13 \*\*\*  
## MacBethYes 10.09037 0.96507 10.456 < 2e-16 \*\*\*  
## LoyaltyYes 20.54005 0.83833 24.501 < 2e-16 \*\*\*  
## factor(FamilySize)2 7.66332 1.29516 5.917 4.52e-09 \*\*\*  
## factor(FamilySize)3 18.04657 1.28465 14.048 < 2e-16 \*\*\*  
## factor(FamilySize)4 28.39529 1.40907 20.152 < 2e-16 \*\*\*  
## factor(FamilySize)5 38.46692 1.72345 22.320 < 2e-16 \*\*\*  
## factor(FamilySize)6 47.47276 2.28780 20.750 < 2e-16 \*\*\*  
## factor(FamilySize)7 59.82409 3.75999 15.911 < 2e-16 \*\*\*  
## factor(FamilySize)8 58.32600 6.07000 9.609 < 2e-16 \*\*\*  
## factor(FamilySize)9 75.39242 8.51721 8.852 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.94 on 988 degrees of freedom  
## Multiple R-squared: 0.6795, Adjusted R-squared: 0.6759   
## F-statistic: 190.4 on 11 and 988 DF, p-value: < 2.2e-16

ggplot(Data, aes(x = Age, y = Spend,shape=MacBeth,color=FamilySize,fill=Loyalty,alpha=Loyalty)) +   
 geom\_point() +  
 stat\_smooth(method = "lm")

## Warning: Using alpha for a discrete variable is not advised.

## `geom\_smooth()` using formula 'y ~ x'

