Data HW1_Jiaqi Li

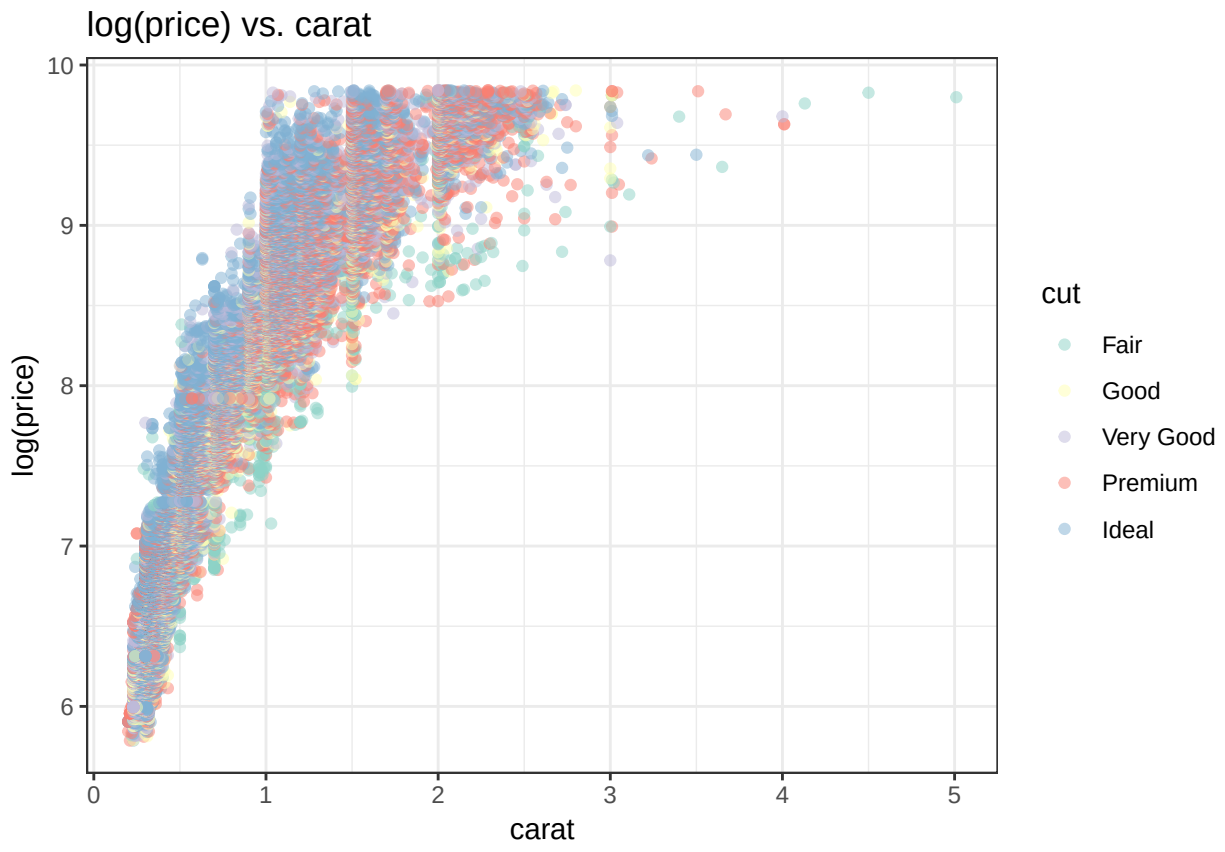
Jiaqi Li

April 7, 2019

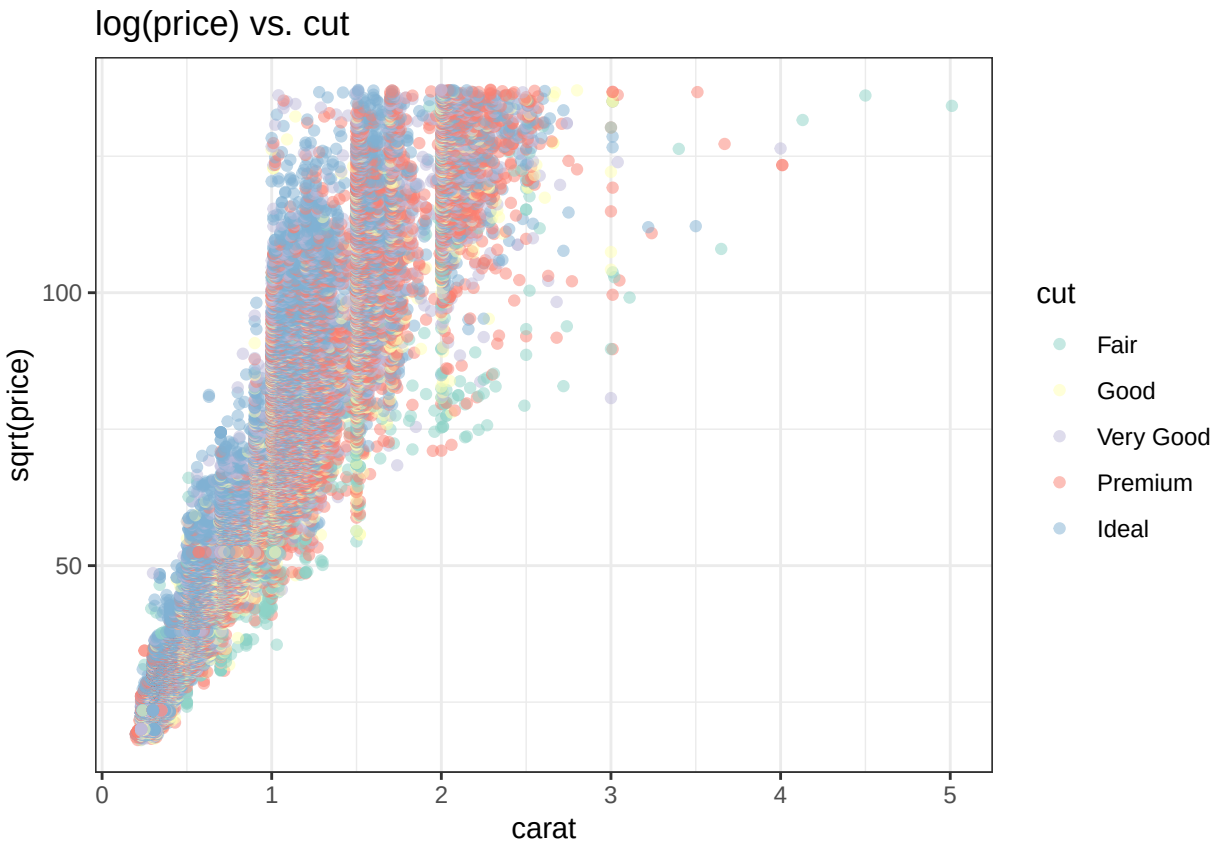
Question 1

1

```
options(warn = -1)
#----------------part 1---------------------#
library(ggplot2)
data(diamonds)
qplot(carat,log(price),data = diamonds, col = cut, alpha = I(1/2),
      main = "log(price) vs. carat") + theme_bw() +
  scale_color_brewer(palette="Set3")
```




```
qplot(carat,sqrt(price),data = diamonds, col = cut, alpha = I(1/2),
      main = "log(price) vs. cut") + theme_bw() +
  scale_color_brewer(palette="Set3")
```



2

```
#----------------part 2---------------------#
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

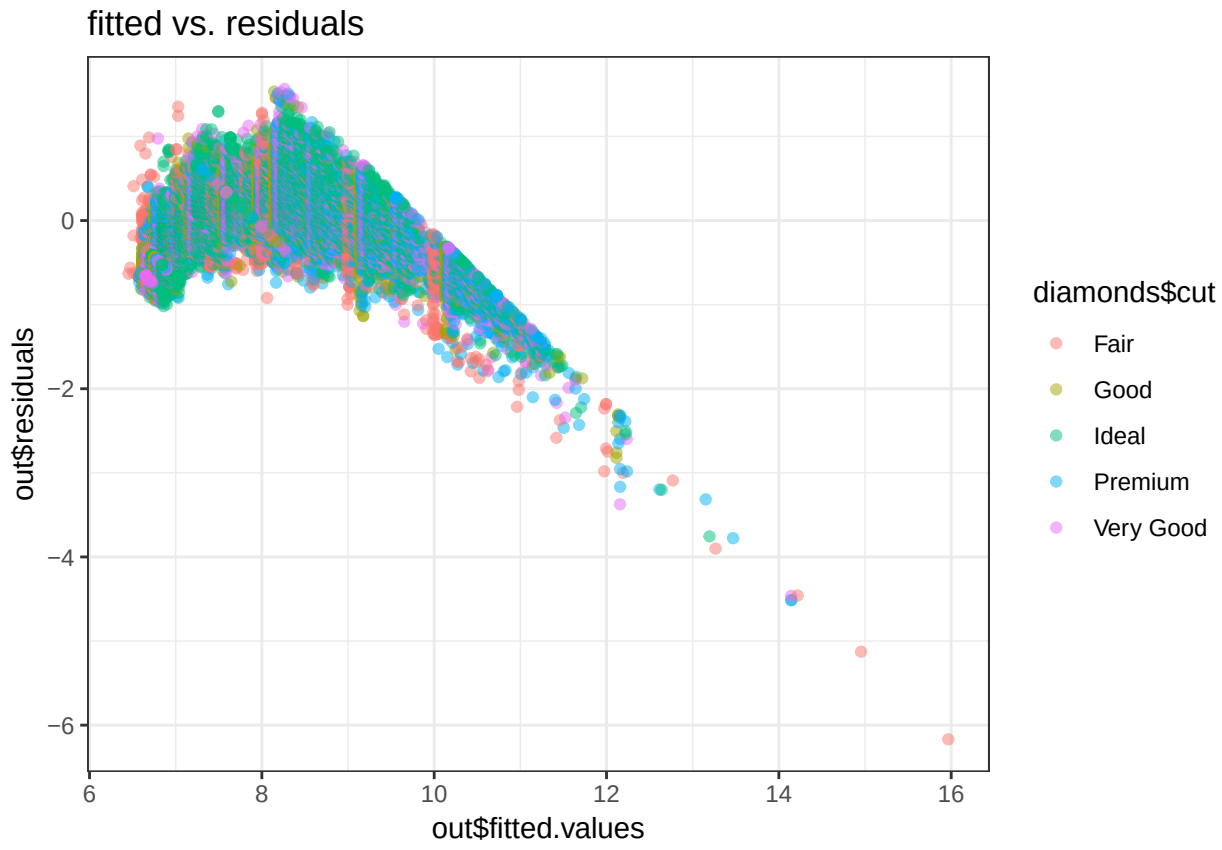
```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

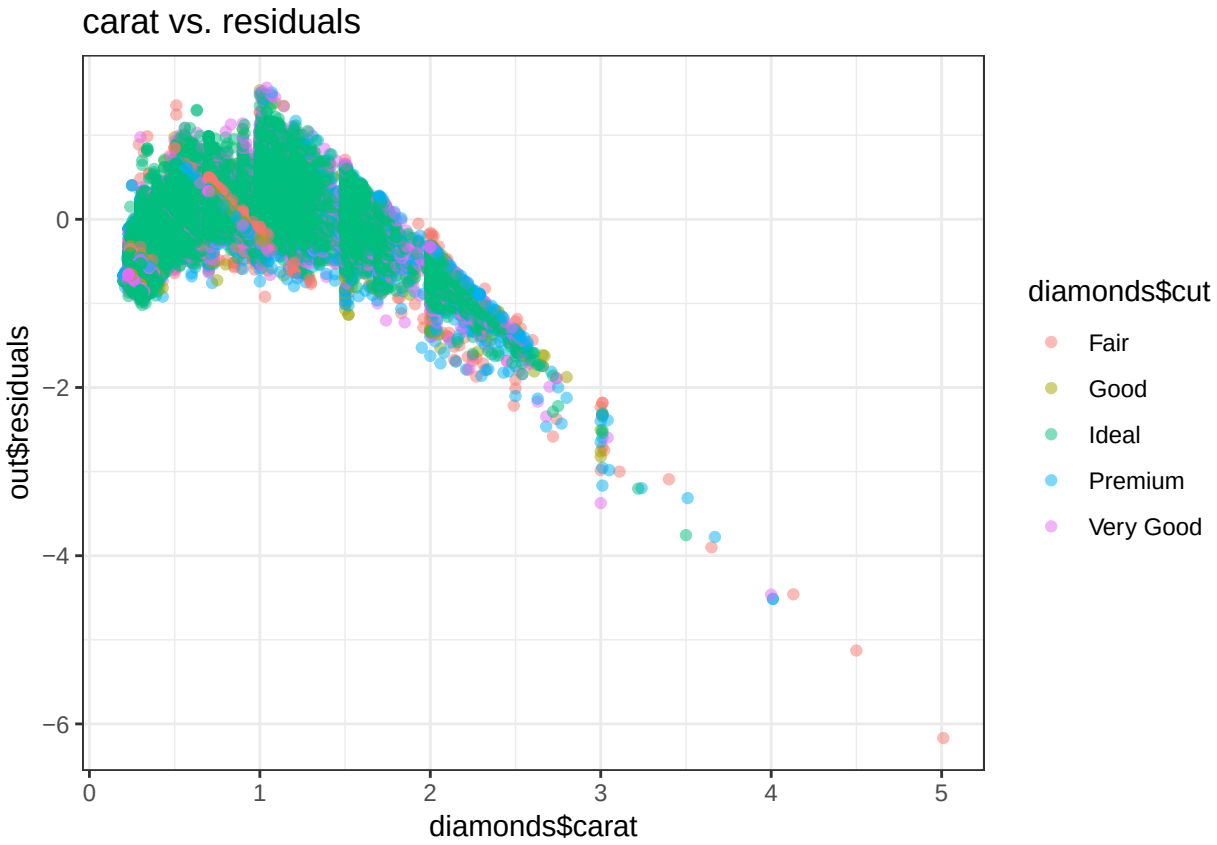
```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
cutf = as.character(diamonds$cut) %>% as.factor()
diamonds$cut = cutf
out = lm(log(price)~carat+cut,data = diamonds)
qplot(out$fitted.values,out$residuals, alpha = I(1/2),
      col = diamonds$cut, main = "fitted vs. residuals")+
  theme_bw()
```




```
qplot(diamonds$carat,out$residuals, alpha = I(1/2),
      col = diamonds$cut, main = "carat vs. residuals")+
  theme_bw()
```



Based on the observation, we can tell the residules plots show that at small fitted values, the linear regression line fits the data farily, but at the large fitted value, the linear regression line fits the data poorly. As the price increases, the residules also increase. This inplies that the linear regression fit the data poorly since the residules shows a trend.

Question 2

1

```
#----------------part 1---------------------#
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

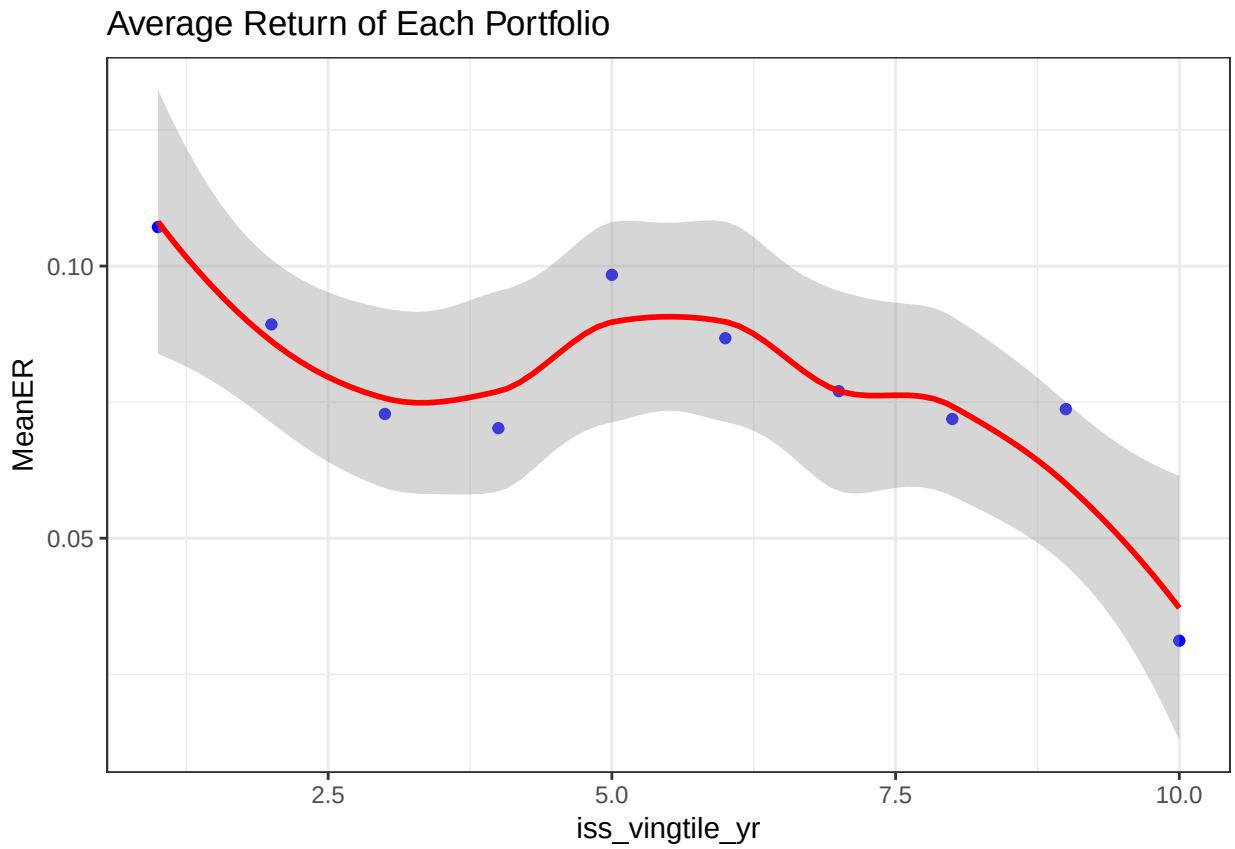
```
library(haven)
StockR = read_dta("StockRetAcct_insample.dta")
StockR = as.data.table(StockR)
StockR[,ExRet:=exp(lnAnnRet)-exp(lnRf)]
StockR[,lnIssue:=jitter(lnIssue,amount=0)]
for(i in 1981:2014){
```

```
  StockR[year == i, iss_vingtile_yr:=cut(StockR[year == i,]$lnIssue,
         breaks = quantile(StockR[year == i,]$lnIssue,probs = c(0:10)/10,na.rm = T),
         include.lowest=T,labels=F)]
}
Stock_R_Portf_yr = StockR[,list(MeanER_Yr = weighted.mean(ExRet,MEwt)),
                    by = list(iss_vingtile_yr,year)]
Stock_R_Portf = na.omit(Stock_R_Portf_yr[,list(MeanER = mean(MeanER_Yr)),
                by = iss_vingtile_yr])
```

2

```
#----------------part 2---------------------#
qplot(iss_vingtile_yr, MeanER, data = Stock_R_Portf,col=I("blue"),
      na.rm = T, main = "Average Return of Each Portfolio") +
  geom_smooth(col=I("red"))+theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The pattern does not seems very linear, but it does show a slightly linear trend with negative slope.

3

```
#----------------part 3---------------------#
StockR[iss_vingtile_yr == 1, transformed_iss:=-1]
StockR[iss_vingtile_yr == 10, transformed_iss:=1]
for(i in 2:9){
  StockR[iss_vingtile_yr == i, transformed_iss:=0]
}
out2 = lm(ExRet ~ transformed_iss, data = StockR)
summary(out2)
```

```
##
## Call:
## lm(formula = ExRet ~ transformed_iss, data = StockR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1684 -0.2612 -0.0399  0.1967 10.0304
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.087989   0.001870  47.058   <2e-16 ***
## transformed_iss -0.036987   0.004176  -8.857   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4661 on 62140 degrees of freedom
##   (8614 observations deleted due to missingness)
## Multiple R-squared:  0.001261,   Adjusted R-squared:  0.001245
## F-statistic: 78.45 on 1 and 62140 DF,  p-value: < 2.2e-16
```

```
#this implies short stocks in decile 10 and long stocks in decile 1
```

Based on the coefficient of the transformed issuance characteristic, we will earn an excess return if the characeristic is -1 and negative excess return is the characteristic is 1. Thus, we should long the portfolio contains stocks in Decile 1 and short the portfolio contains stocks in Decile 10.

Question 3

1

```
#----------------part 1---------------------#
StockR = read_dta("StockRetAcct_insample.dta")
StockR = as.data.table(StockR)
StockR[,ExRet:=exp(lnAnnRet)-exp(lnRf)]
StockR[,lnBM:=jitter(lnBM,amount=0)]
StockR[,lnME:=jitter(lnME,amount=0)]
for(i in 1981:2014){
  StockR[year == i, bm_quintile_yr := cut(StockR[year == i,]$lnBM,
         breaks = quantile(StockR[year == i,]$lnBM, probs = c(0:5)/5,
         na.rm = T), include.lowest = T, labels = F)]
```

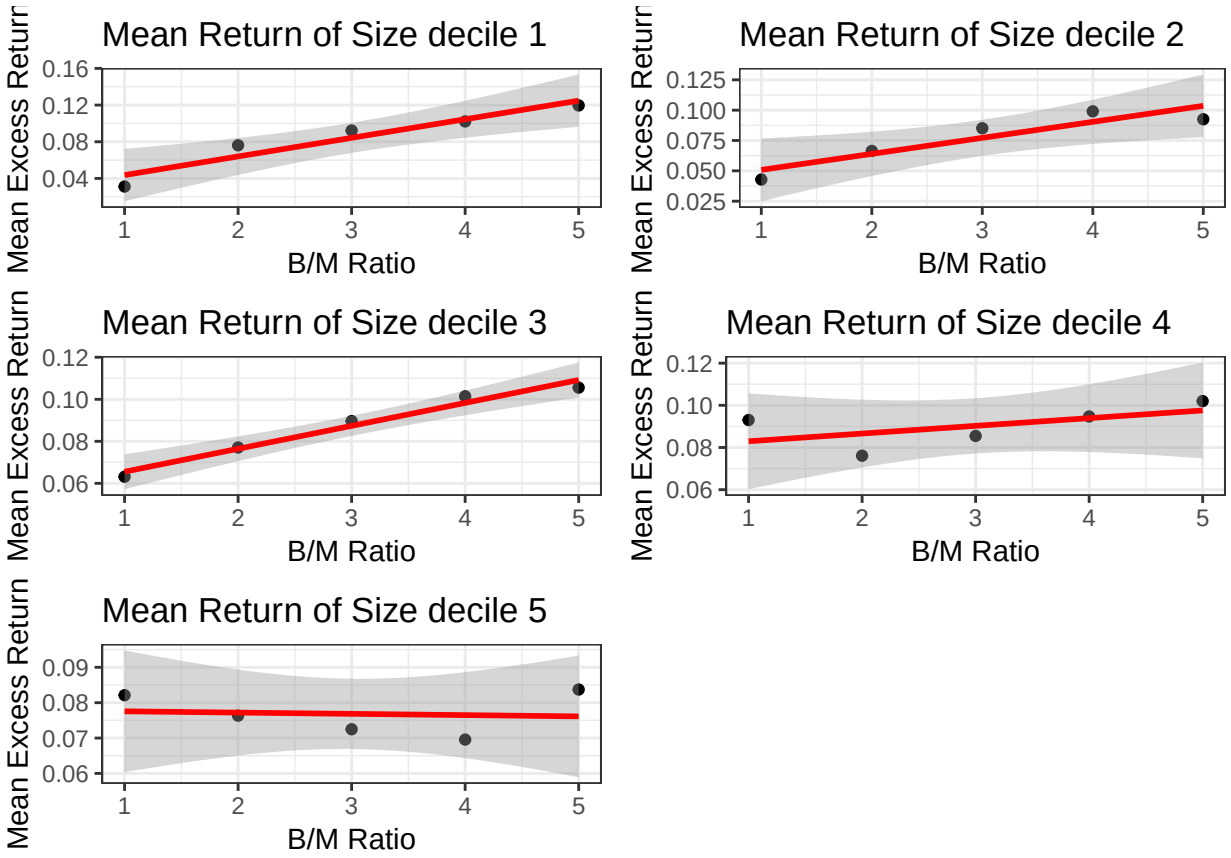
```
  StockR[year == i, me_quintile_yr := cut(StockR[year == i,]$lnME,
         breaks = quantile(StockR[year == i,]$lnME, probs = c(0:5)/5,
         na.rm = T), include.lowest = T, labels = F)]
}
```

2

```
#---------------part 2--------------------#
Stock_R_Portf_double_yr = StockR[,list(MeanER_Yr = weighted.mean(ExRet,MEwt)),
  by = list(bm_quintile_yr,me_quintile_yr,year)]
Stock_R_Portf_double = na.omit(Stock_R_Portf_double_yr[,list(MeanER = mean(MeanER_Yr)),
  by = list(bm_quintile_yr,me_quintile_yr)])
setkey(Stock_R_Portf_double,me_quintile_yr)

for(i in 1:5){
  assign(paste("Plot",i,sep=""),
         qplot(bm_quintile_yr, MeanER,data = Stock_R_Portf_double[.(i)], na.rm = T,
         main = paste("Average Return of Portfolio with Size belongs to decile",i,sep=""),
         xlab="B/M Ratio", ylab = "Mean Excess Return") +
          geom_smooth(method = "lm",col=I("red"))+theme_bw())
}

p1= qplot(bm_quintile_yr, MeanER,data = Stock_R_Portf_double[.(1)], na.rm = T,
      main = "Mean Return of Size decile 1",
      xlab="B/M Ratio", ylab = "Mean Excess Return") +
  geom_smooth(method = "lm",col=I("red"))+theme_bw()
p2= qplot(bm_quintile_yr, MeanER,data = Stock_R_Portf_double[.(2)], na.rm = T,
      main = "Mean Return of Size decile 2",
      xlab="B/M Ratio", ylab = "Mean Excess Return") +
  geom_smooth(method = "lm",col=I("red"))+theme_bw()
p3= qplot(bm_quintile_yr, MeanER,data = Stock_R_Portf_double[.(3)], na.rm = T,
      main = "Mean Return of Size decile 3",
      xlab="B/M Ratio", ylab = "Mean Excess Return") +
  geom_smooth(method = "lm",col=I("red"))+theme_bw()
p4= qplot(bm_quintile_yr, MeanER,data = Stock_R_Portf_double[.(4)], na.rm = T,
      main = "Mean Return of Size decile 4",
      xlab="B/M Ratio", ylab = "Mean Excess Return") +
  geom_smooth(method = "lm",col=I("red"))+theme_bw()
p5= qplot(bm_quintile_yr, MeanER,data = Stock_R_Portf_double[.(5)], na.rm = T,
      main = "Mean Return of Size decile 5",
      xlab="B/M Ratio", ylab = "Mean Excess Return") +
  geom_smooth(method = "lm",col=I("red"))+theme_bw()
grid.arrange(p1,p2,p3,p4,p5)
```



The assumption in linearity seems ok based on the 5 plots above. Only the portfolio of size in decile 4 and 5 shows a little bit non-linear trend, but overall, the average excess return does show a linear correlation with B/M ratio.