HW4

Ming Zhong mz2692

10/20/2018

```
library("plyr")
debt <- read.csv("/Users/zhongming/Downloads/debt.csv", as.is = TRUE)
dim(debt)
```

```
## [1] 1171    4
```

```
head(debt)
```

	Country <chr>	Year <int>	growth <dbl>	ratio <dbl>
1	Australia	1946	-3.557951	190.41908
2	Australia	1947	2.459475	177.32137
3	Australia	1948	6.437534	148.92981
4	Australia	1949	6.611994	125.82870
5	Australia	1950	6.920201	109.80940
6	Australia	1951	4.272612	87.09448

6 rows

1.

```
#Construct the function
mean.growth<-function(df){
  return(mean(df$growth))
}
#Using daply to get the average growth of different country
growth.country=daply(debt,.(Country),mean.growth)
#Print the value
round(c(growth.country["Australia"],growth.country["Netherlands"]),digits=2)
```

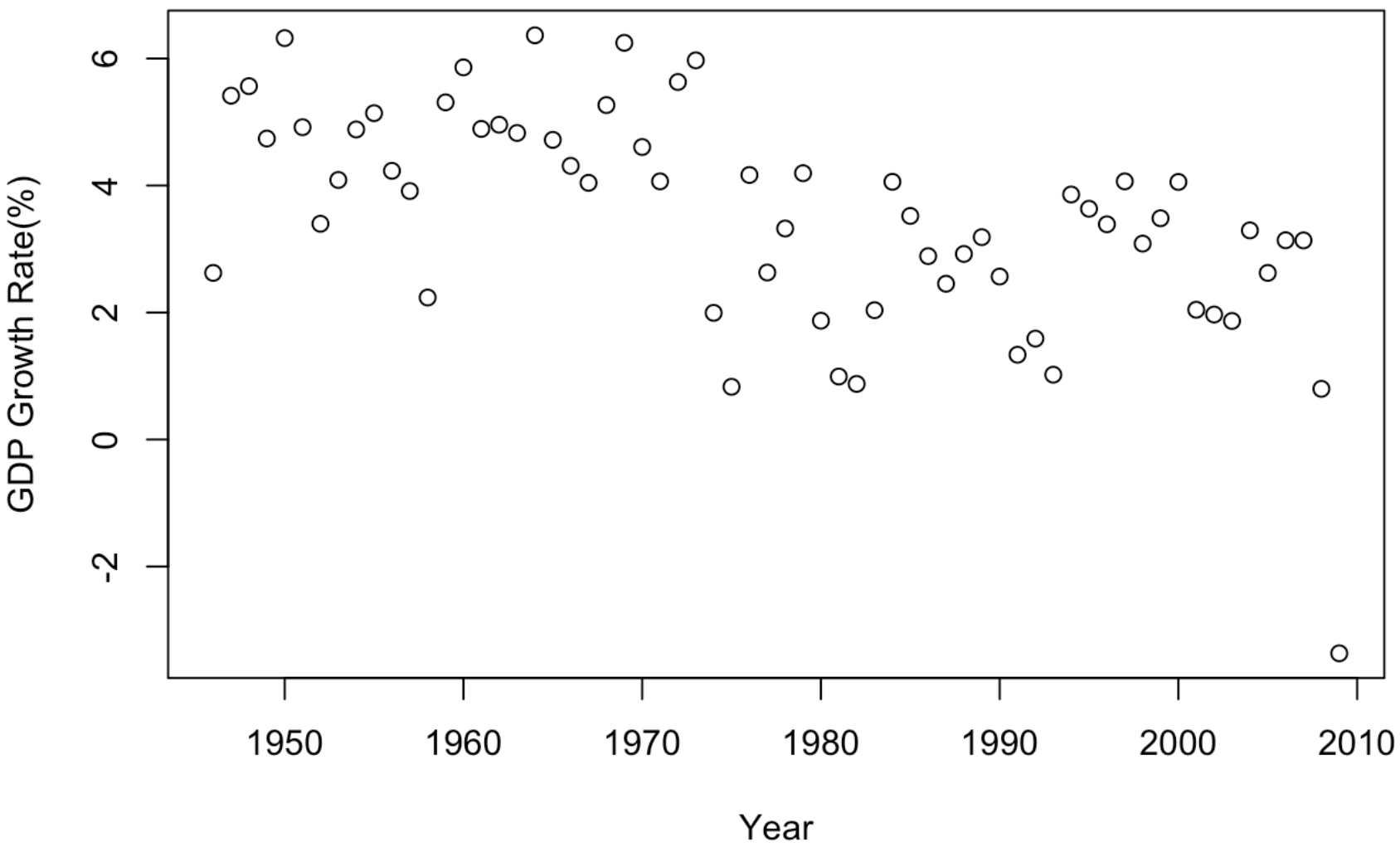
```
##   Australia Netherlands
##        3.72        3.03
```

2.

```
#get the yearly average growth rate
growth.year=daply(debt,.(Year),mean.growth)
#print out the targeted yearly growth rate
round(c(growth.year["1972"],growth.year["1989"]),digits=2)
```

```
## 1972 1989
## 5.63 3.19
```

```
#plot the yearly growth rate
plot(names(growth.year),growth.year,xlab="Year",ylab="GDP Growth Rate(%)")
```

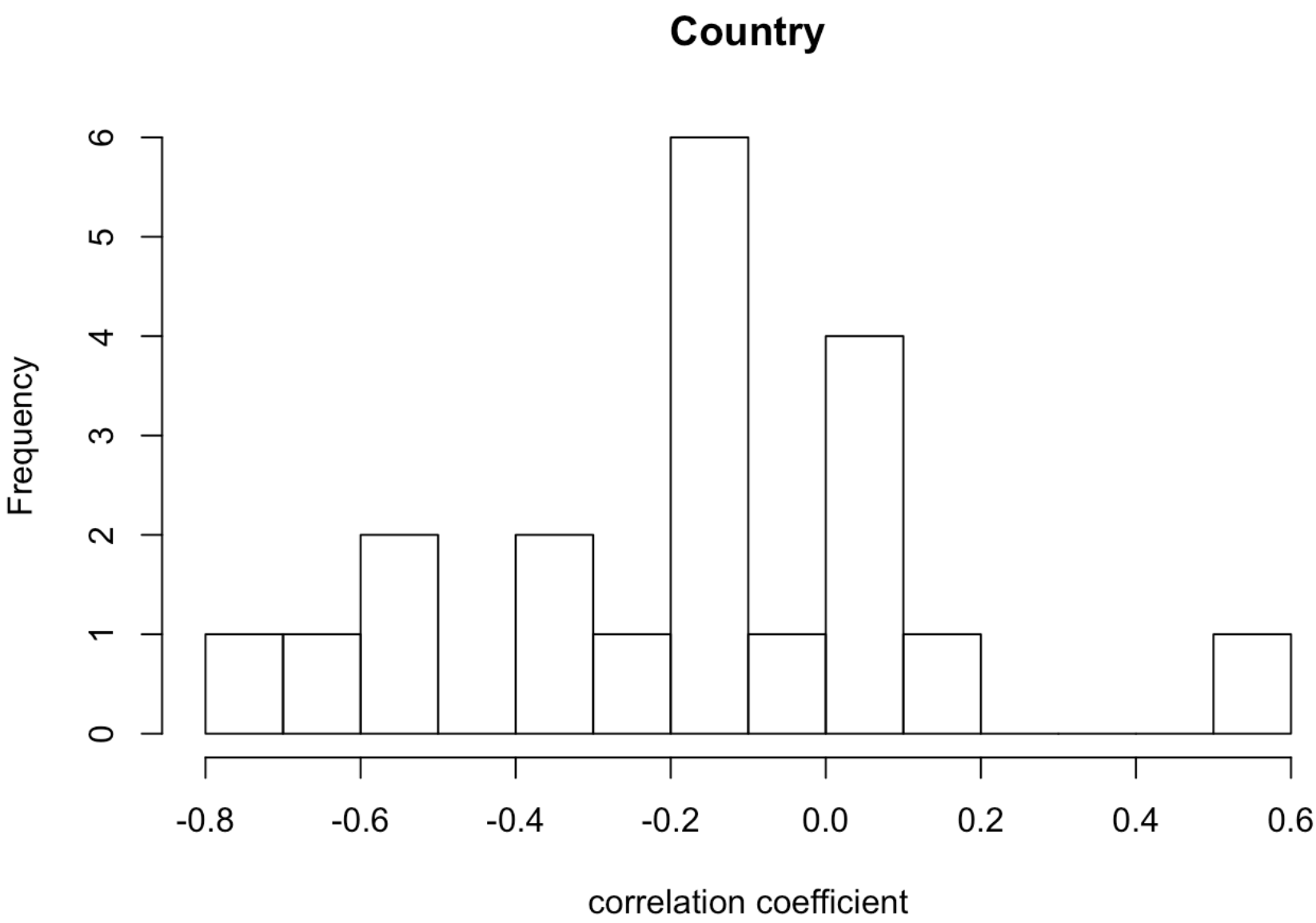


3.

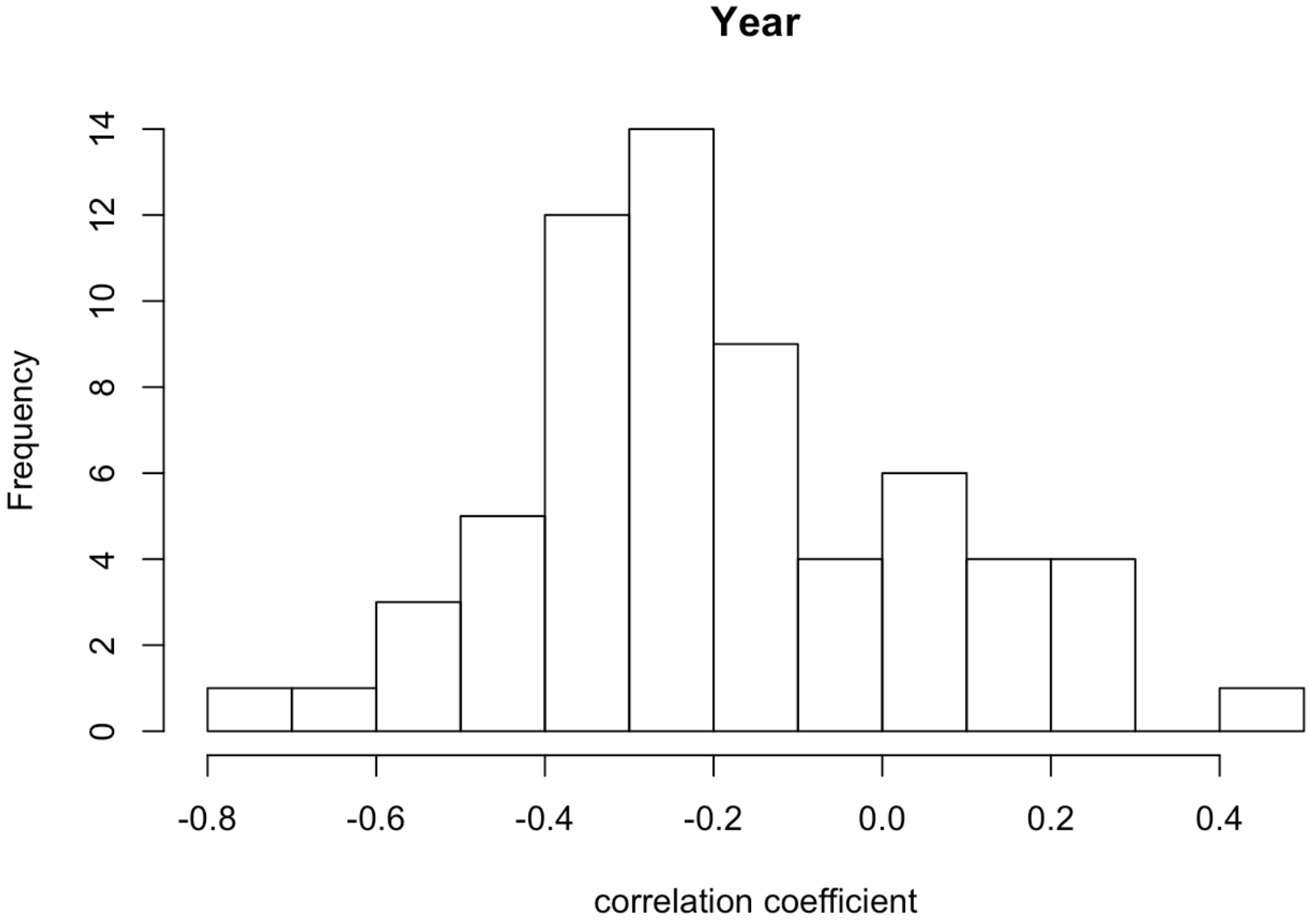
```
#get the correlation between growth rate and ratio
round(cor(debt$growth,debt$ratio),digits = 4)
```

```
## [1] -0.1995
```

```
corr<-function(df){return(round(cor(df$growth,df$ratio),digits = 4))}
#compute correatlion for each country
corr.country=daply(debt,.(Country),corr)
#plot histogram
hist(breaks=10,corr.country,xlab="correlation coefficient",main="Country")
```




```
#compute correlation for each year
corr.year=daply(debt,.(Year),corr)
#plot histogram
hist(breaks=10,corr.year,xlab="correlation coefficient",main="Year")
```

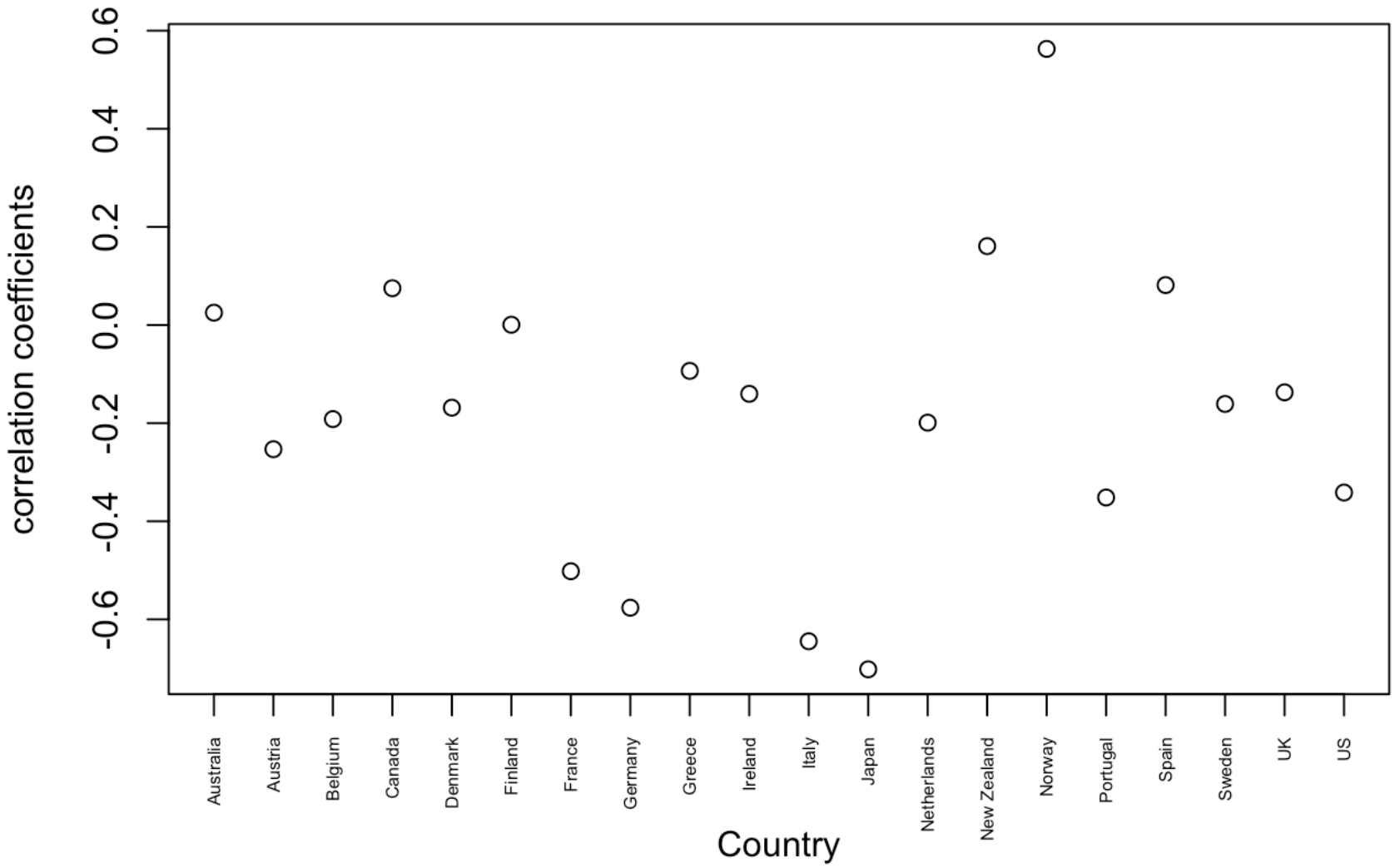



```
#mean of correlation,which is around -0.1906
round(mean(corr.year),digits=5)
```

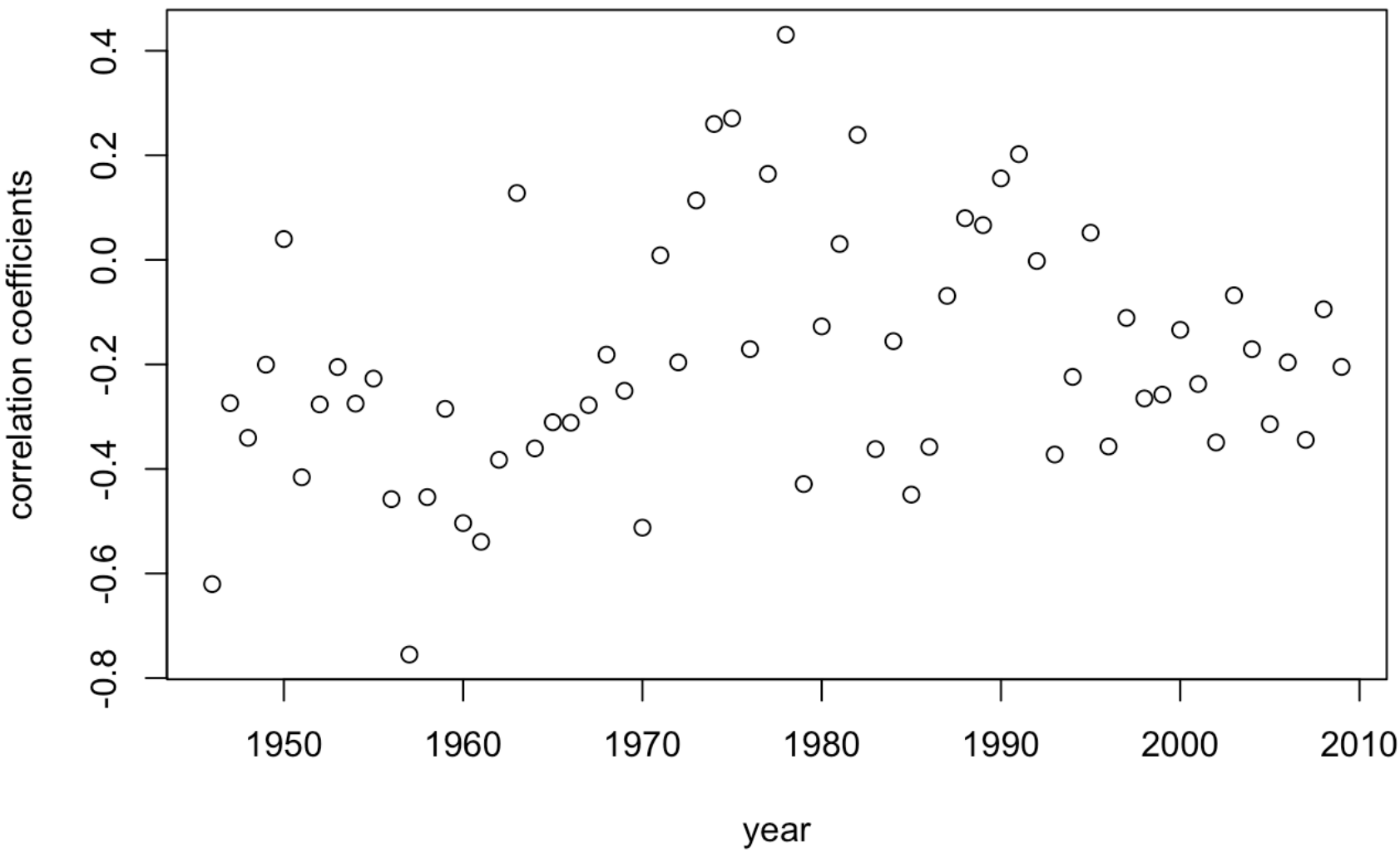
```
## [1] -0.19055
```

Are there any countries or years where the correlation goes against the general trend?

```
plot(corr.country,xlab="Country",ylab="correlation coefficients",xaxt = "n")
axis(side=1,at=1:20,labels=names(corr.country),las=2,cex.axis=0.5)
```




```
plot(x=as.numeric(names(corr.year)),y=corr.year,xlab="year",ylab="correlation coef
ficients")
```



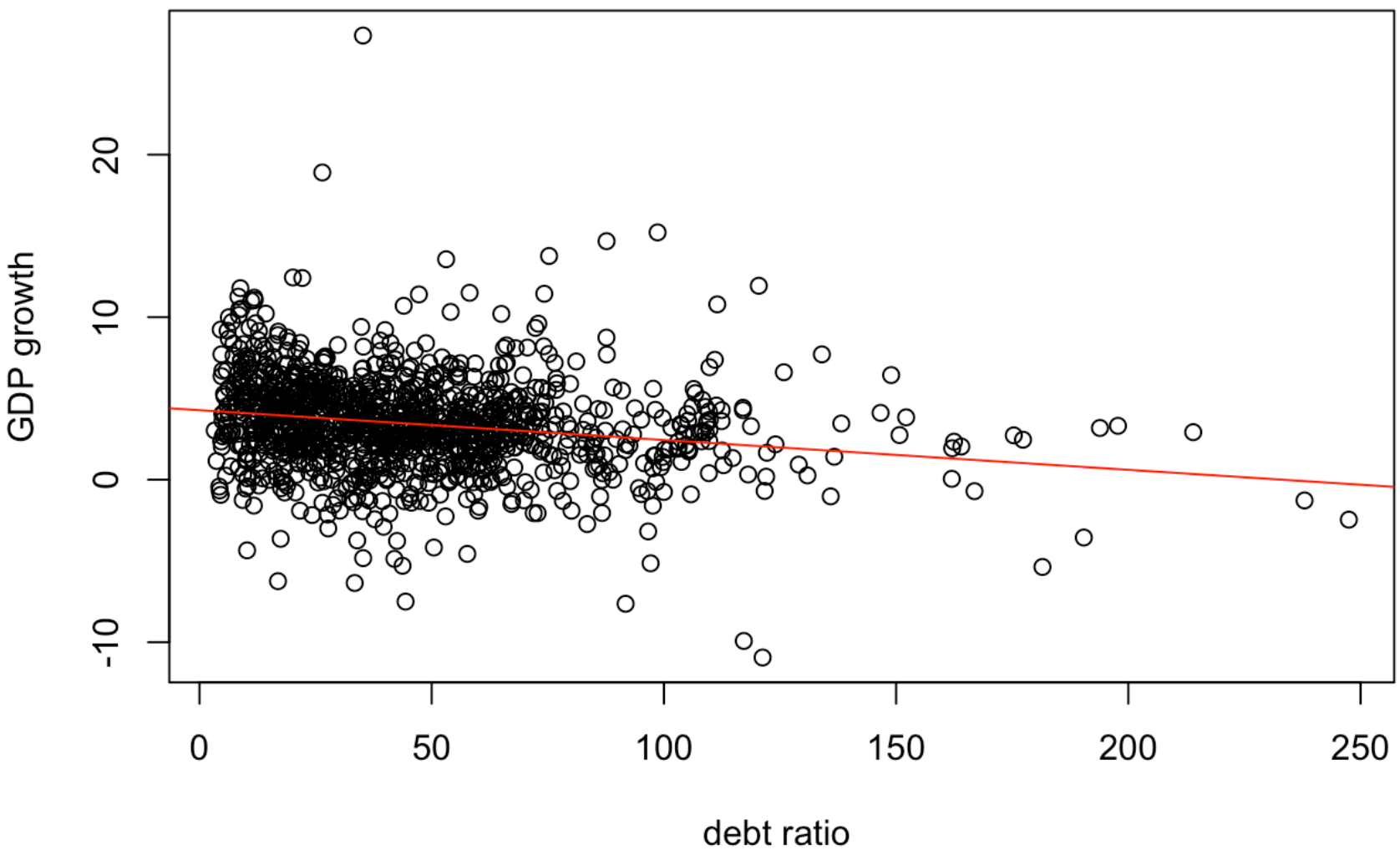
Ans: It seems that Norway is goes against the general trend and the correlation around 1980 also goes against the trend.

4.

```
#get the regreesino result
summary(lm(data=debt,growth~ratio))
```

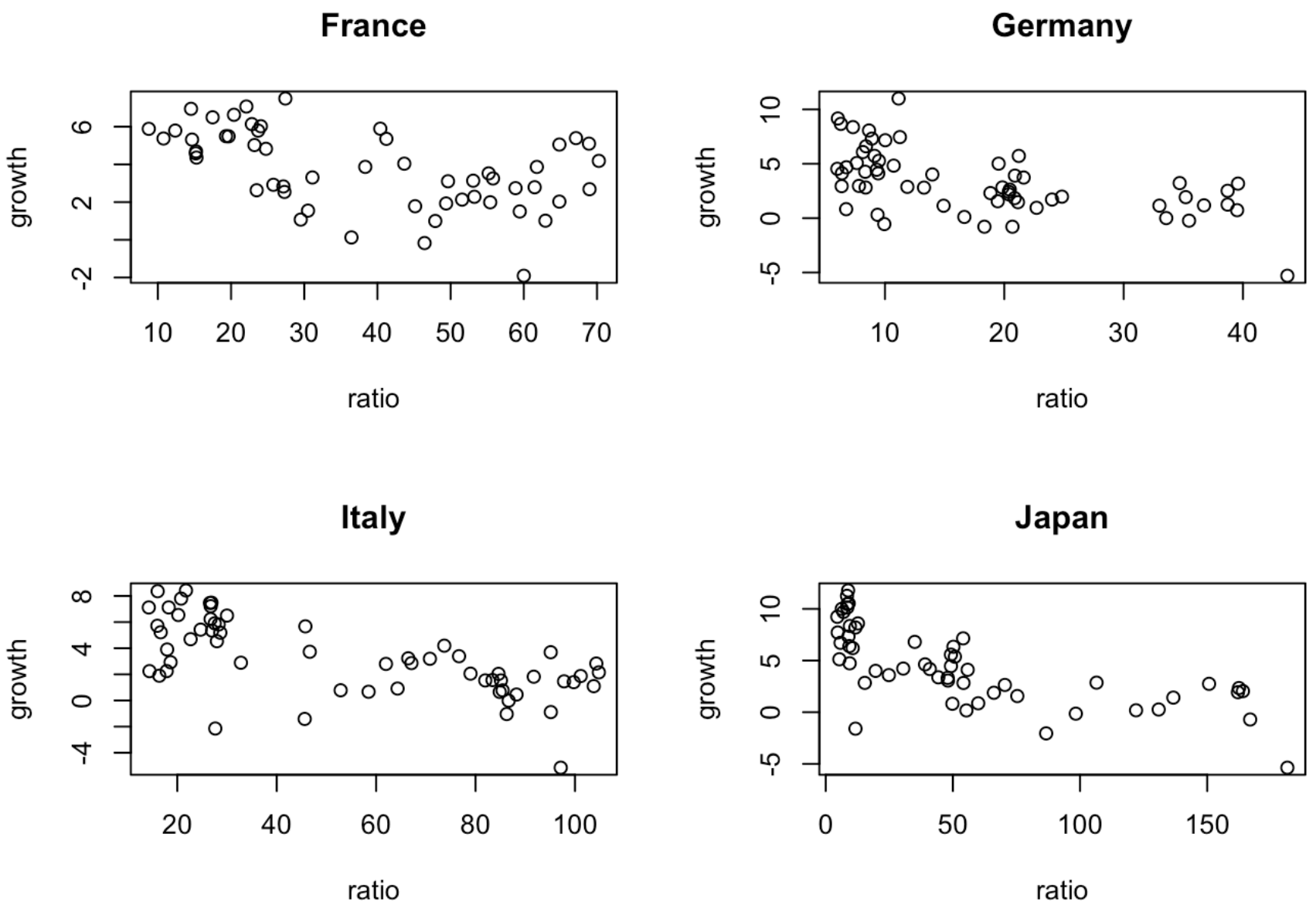
```
##
## Call:
## lm(formula = growth ~ ratio, data = debt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9958  -1.5200  -0.0774   1.5707  23.6960
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.279290   0.148970   28.73  < 2e-16 ***
## ratio       -0.018355   0.002637   -6.96 5.67e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.922 on 1169 degrees of freedom
## Multiple R-squared:  0.03979,    Adjusted R-squared:  0.03897
## F-statistic: 48.44 on 1 and 1169 DF,  p-value: 5.666e-12
```

```
debt=as.data.frame(debt)
#plot debt ratio against debt growth
plot(debt$ratio,debt$growth,xlab="debt ratio ",ylab="GDP growth")
abline(lm(data=debt,growth~ratio),col="red")
```



5.

```
#print out the name of the the 4 countries
country4=names(corr.country[corr.country<(-0.5)])
#plot 4 graphs in one figure
par(mfrow=c(2,2))
for (this.country in country4){
  country.dat<-subset(debt,Country==this.country)
  with(data=country.dat,plot(ratio,growth,main=this.country))
}
```



6.

```
#a.create new data frame
debt.France=subset(debt,Country=="France")
dim(debt.France)
```

```
## [1] 54  4
```

```
year=debt.France$Year+1
#b.create new column
next.growth=rep(1,54)
for(i in 1:54){
  if(year[i] %in% debt.France$Year){
    next.growth[i]=debt.France[debt.France$Year==year[i],3]
  }
  else{
    next.growth[i]=NA
  }
}
debt.france=cbind(debt.France,next.growth)
#print out to test
round(debt.france[debt.france$Year==1971|debt.france$Year==1972,5],digits=3)
```

```
## [1] 5.886    NA
```

7.

```
#encapsulate question 6 into function
func<-function(df){
  y=df$Year+1
  next.growth=rep(NA,nrow(df))
  for(i in 1:nrow(df)){
    if(y[i] %in% df$Year){
      next.growth[i]=df[df$Year==y[i],3]
    }
    else{
      next.growth[i]=NA
    }
  }
  return(cbind(df,next.growth))
}
#get the data frame
df=ddply(debt,.(Country),func)
#doube check the France 2009 is NA
df[df$Country=="France"&df$Year==2009,"next.growth"]
```

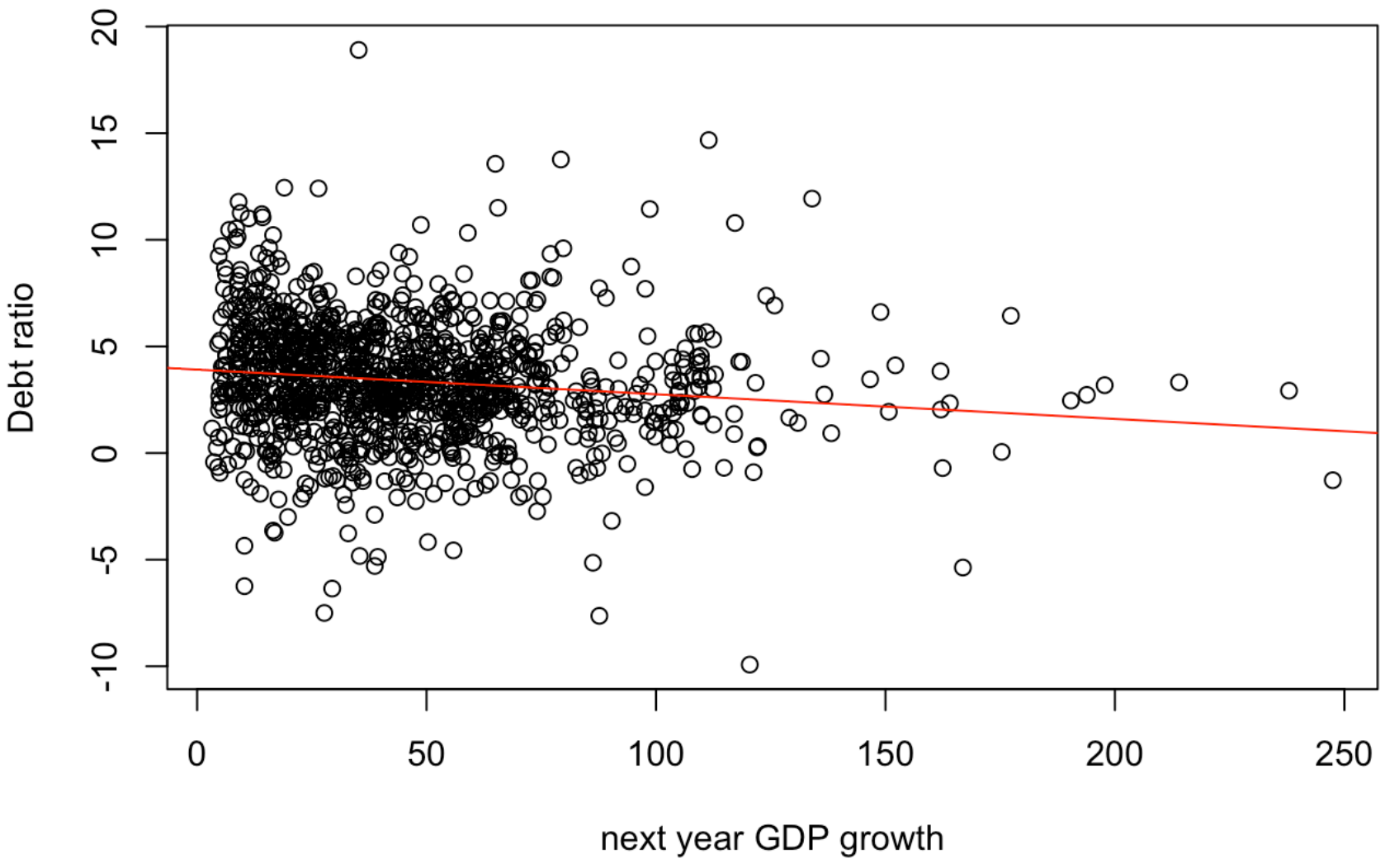
```
## [1] NA
```

8.

```
#plot next year's GDP growth against this year's debt ratio
plot(df$ratio,df$next.growth,xlab="next year GDP growth",ylab="Debt ratio")
#Linearly regress next year's growth rate on the current year's debt ratio
summary(lm(df$next.growth~df$ratio))
```

```
##
## Call:
## lm(formula = df$next.growth ~ df$ratio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4488  -1.4567  -0.0374   1.6331  15.3864
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.924722   0.143852  27.283  < 2e-16 ***
## df$ratio    -0.011608   0.002555  -4.544 6.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.789 on 1145 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.01771,    Adjusted R-squared:  0.01686
## F-statistic: 20.65 on 1 and 1145 DF,  p-value: 6.105e-06
```

```
#d add the line to the plot
abline(lm(df$next.growth~df$ratio),col="red")
```



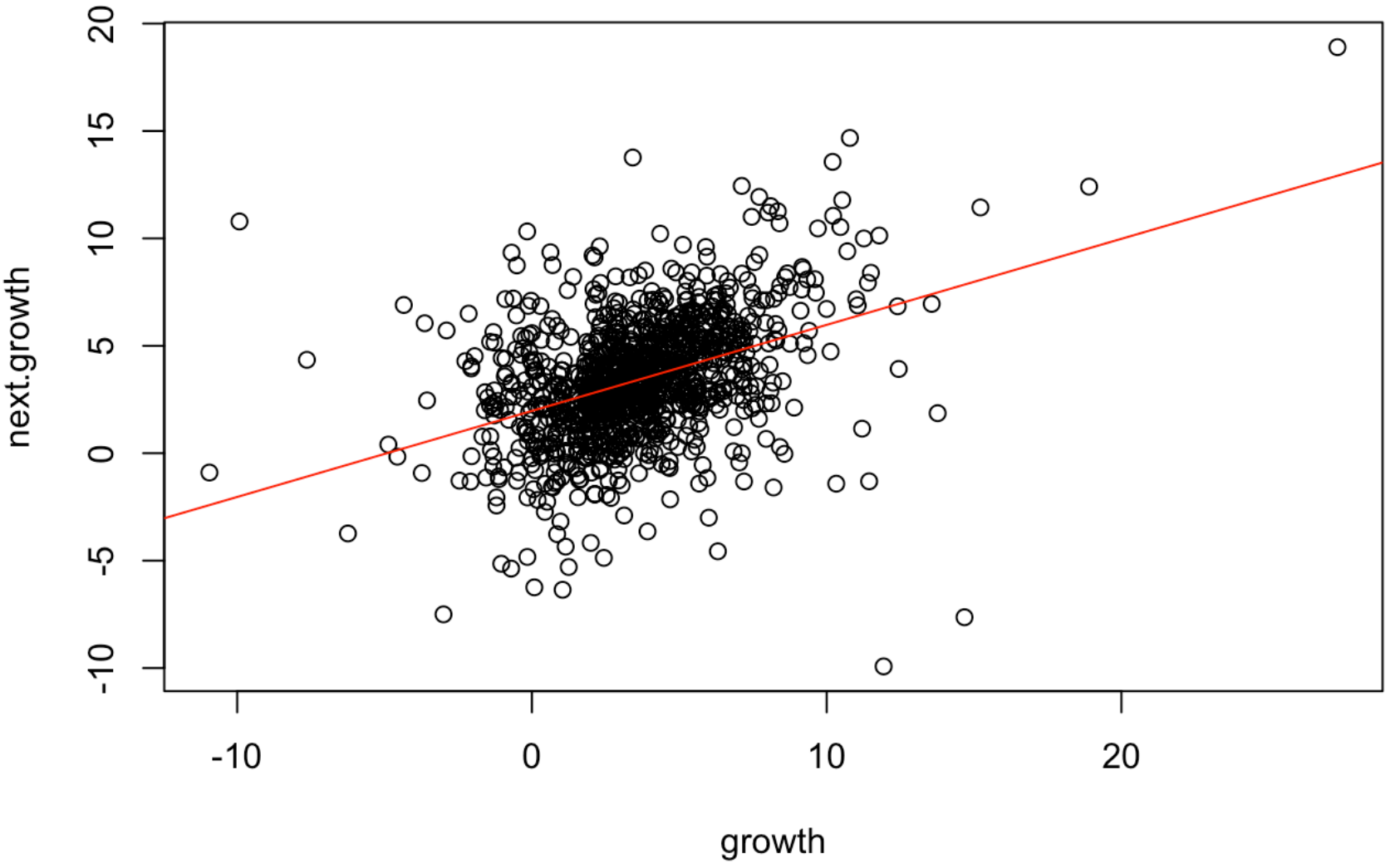
Ans:The regreesion result compared to the regression of the current year's growth on the current year's debt ratio is even worse. The R square of this result is smaller than the previous.

9.

```
#plot of next year's GDP growth against the current year's GDP growth
with(data=df,plot(growth,next.growth))
#Linearly regress next year's growth on this year's growth
summary(lm(df$next.growth~df$growth))
```

```
##
## Call:
## lm(formula = df$next.growth ~ df$growth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.6738  -1.3570   0.0401   1.3994  12.7917
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.97106    0.12040   16.37   <2e-16 ***
## df$growth    0.40065    0.02643   15.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.568 on 1145 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.1671, Adjusted R-squared:  0.1664
## F-statistic: 229.8 on 1 and 1145 DF,  p-value: < 2.2e-16
```

```
#add the line to the plot
abline(lm(df$next.growth~df$growth),col="red")
```



Ans:The current growth is a better predictor.