Assignment 2

Ngoc Nguyen, Reyna Prijohutomo

6/25/2021

the data set

```
setwd("C:/Users/Asus/Desktop/Intro to Data Science/assignment 2")
df=data.matrix(read.table("InterestRates.txt", header=FALSE))
#Dimension of the data matrix:
dim(df)
```

```
## [1] 1264   51
```

```
x=data.frame(c(1:1264))
dfx=cbind(x,df)
```

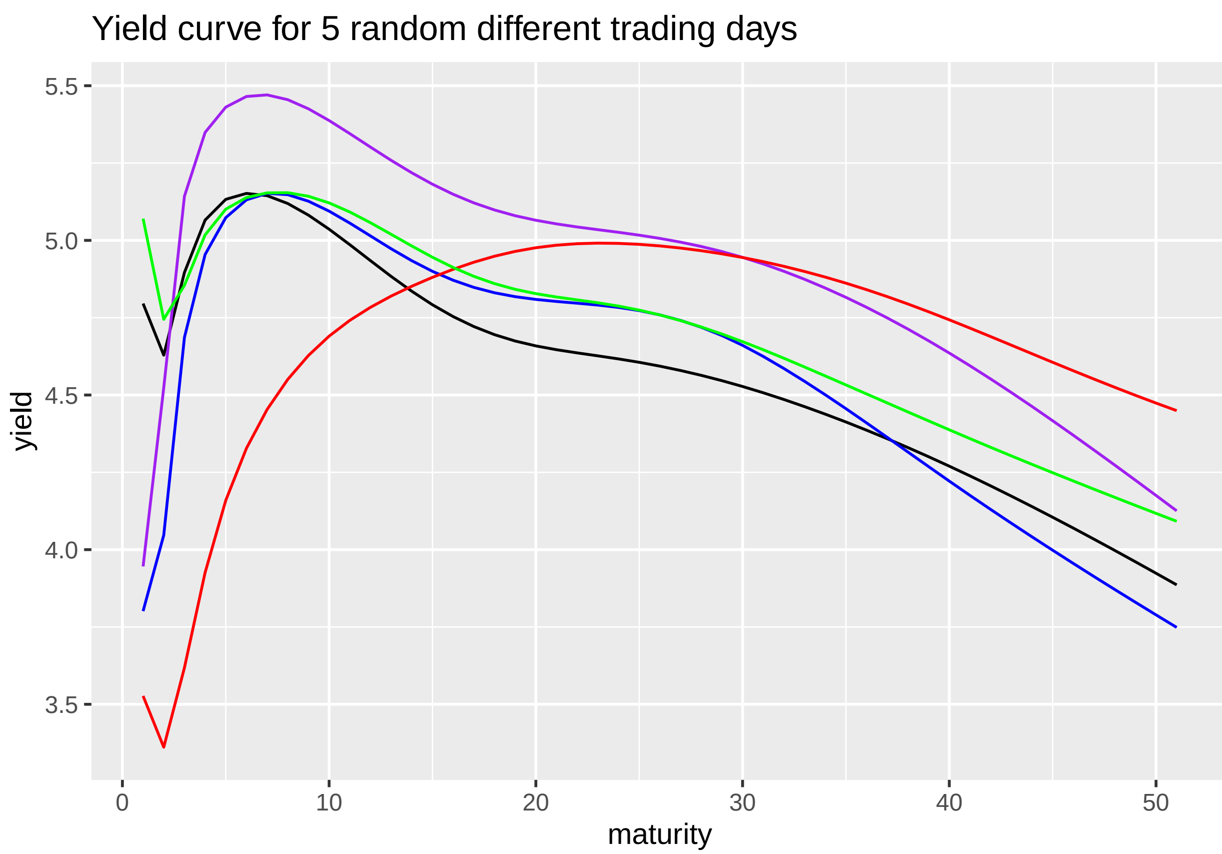
1

Make a plot with the yield curve for 5 different trading days.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

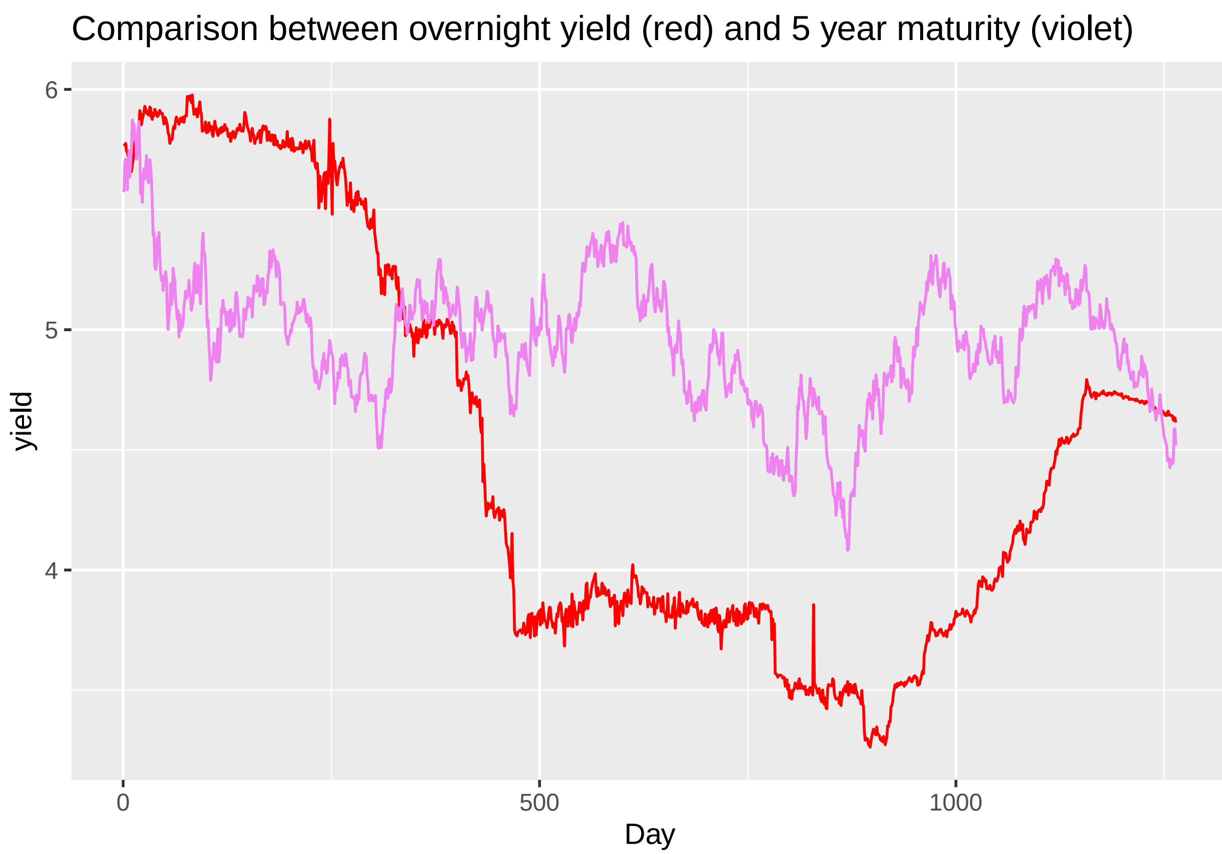
```
xrows=data.frame(c(1:51))
df5 = data.frame(df[sample(nrow(df),5),])
ggplot(mapping=aes(x=xrows$c.1.51.))+
  geom_line(aes(y=as.numeric(df5[1,])),col="black")+
  geom_line(aes(y=as.numeric(df5[2,])),col="purple")+
  geom_line(aes(y=as.numeric(df5[3,])),col="blue")+
  geom_line(aes(y=as.numeric(df5[4,])),col="red")+
  geom_line(aes(y=as.numeric(df5[5,])),col="green")+
  xlab("maturity") +
  ylab("yield")+
  ggtitle("Yield curve for 5 random different trading days")
```



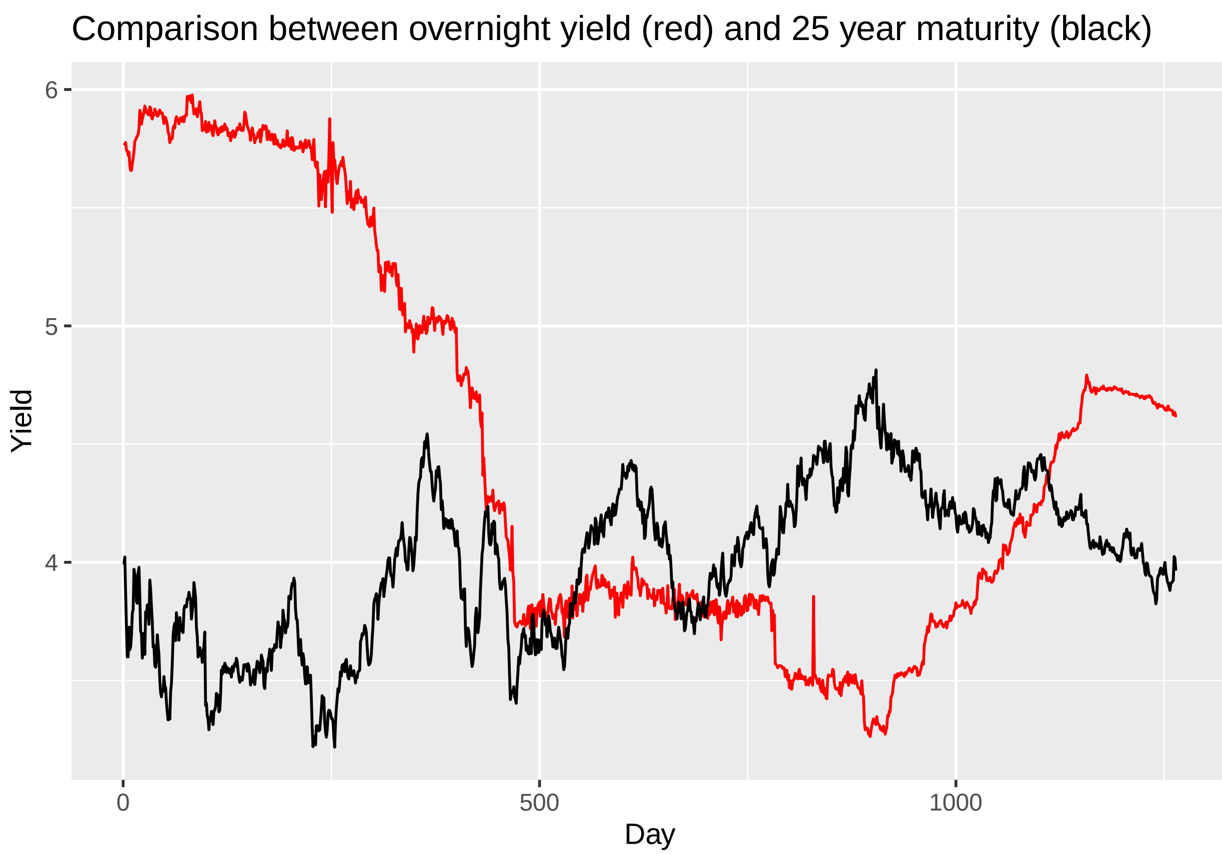
2

Make a plot with the overnight yield over time. Add another maturity to this plot. What do you see?

```
library(ggplot2)
xcols=data.frame(c(1:1264))

ggplot(mapping=aes(x=xcols$c.1.1264))+
  geom_line(aes(y=as.numeric(df[,1])),col="red")+
  geom_line(aes(y=as.numeric(df[,11])),col="violet")+
  xlab("Day") +
  ylab("yield")+
  ggtitle("Comparison between overnight yield (red) and 5 year maturity (violet)")
```




```
ggplot(mapping=aes(x=xcols$c.1.1264))+
  geom_line(aes(y=as.numeric(df[,1])),col="red")+
  geom_line(aes(y=as.numeric(df[,51])),col="black")+
  xlab("Day") +
  ylab("Yield")+
  ggtitle("Comparison between overnight yield (red) and 25 year maturity (black)")
```



#Note: Explain why we chose a certain yield to compare For the graph of the overnight yield (the red line), we see drastic fluctuation. Yield was reducing for the first 800 days, hitting its lowest point at 3.26% before increasing to around 4.65% and maintaining at that level in the last 100 days

For the graph of the yield after 5 years (the violet line), it was fluctuating with a downward trend, with the starting point and ending point at 5.5% and 4.5% consecutively.

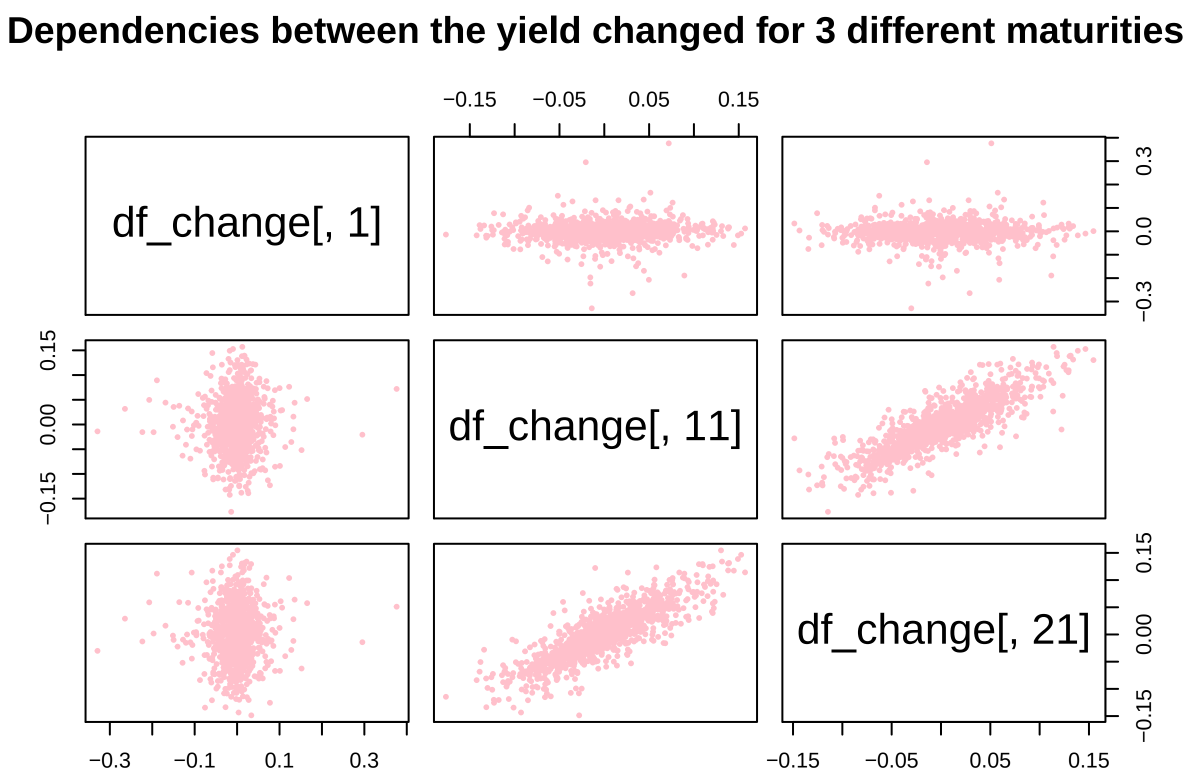
For the graph yield after 25 years (the black line), it was also fluctuating but the starting point and the ending point are both at the 4.0% level.

It can be seen that the more current the yield is the more stable it is. Current yield is also lower compare to older ones.

3

Investigate the dependencies between the yield changed for different maturities. Visualize the dependence of the change in the one year, 5 year and 10 year yield with yield changes in the other maturities.

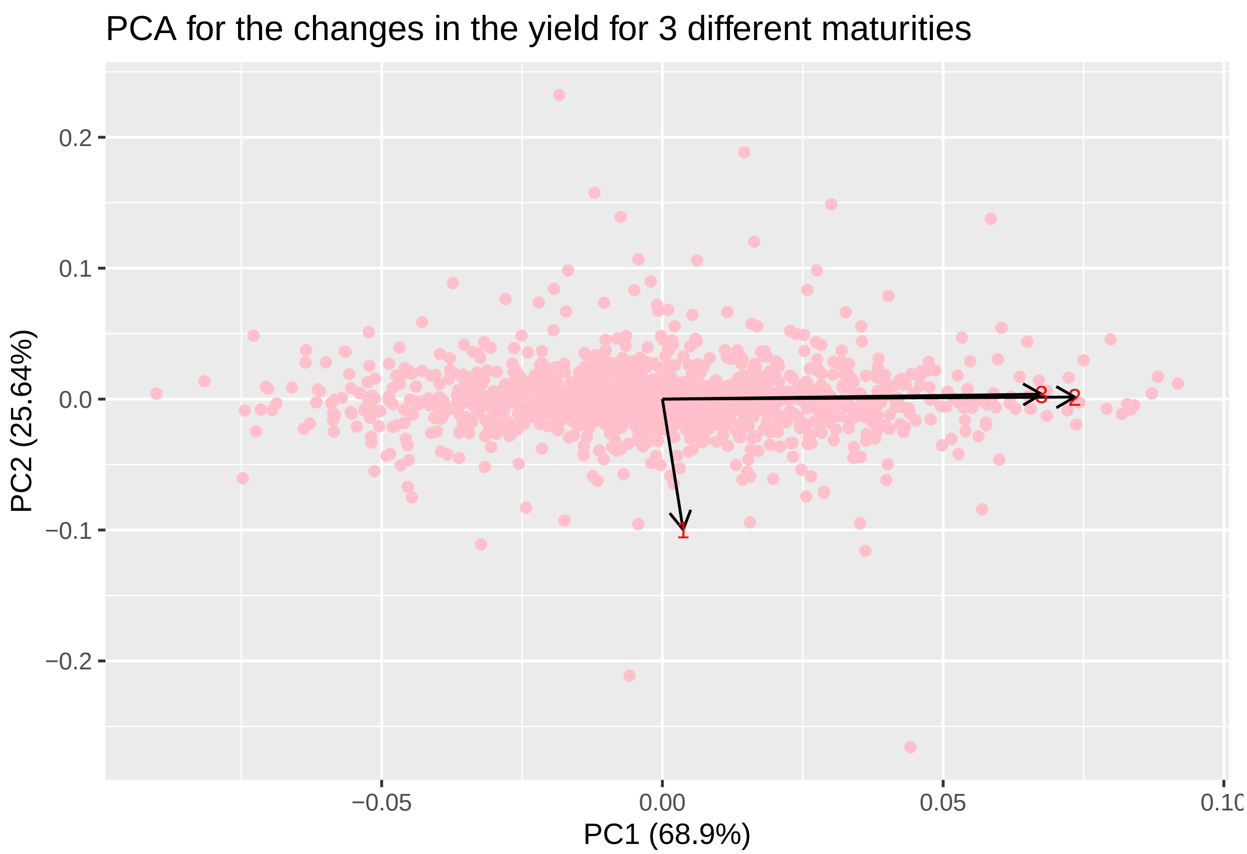
```
df_change <- df[2:1264,] - df[1:1263,]

dfq3 = cbind(df_change[,1],df_change[,11],df_change[,21])

pairs(~df_change[,1] + df_change[,11] + df_change[,21], data=dfq3, pch=20, col="pink",cex = 0.6, main="]
```




```r
library(ggfortify)
```

```
## Warning: package 'ggfortify' was built under R version 4.0.5
```

```r
library(ggplot2)
df.pca = prcomp(dfq3)
autoplot(df.pca,  loadings = TRUE, loadings.colour = 'black',
         loadings.label = TRUE, col="pink",loadings.label.size = 3,
         loading.label.color = 'black',loadings.label.repel=T,main="PCA for the changes in the yield fo
```



The plot of the data is not scattered, so each variable can be used to explained other ones.

The relationship between the yield curve in 5 years and in 10 years is strong, positive and linear.

There seems to be no correlation between the overnight yield and the yield curve in 5 years.

Similarly, there seems to be no correlation between the overnight yield and in 10 years.

Since the overnight yield does not contain information about interest rates or maturity, we can hardly draw prediction from it. However, for the maturity in 5 years and the maturity in 10 years, they have past information that can be used create a certain pattern that the yield follows, and that explains how the the maturity in 5 years and the maturity in 10 years is strongly correlated while the overnight yield is independent from both the maturity in 5 years and the maturity in 10 years.

4

Perform a principal component analysis for the changes in the yield curve.

```
library(ggfortify)
library(ggplot2)

df.pca = prcomp(df_change)
df.PC = df.pca$rotation
#PC1
df.PC[,1]
```

```
##            V1            V2            V3            V4            V5            V6
```

```
## 0.003510332 0.056655860 0.101142785 0.115639743 0.121540928 0.125682493
##          V7          V8          V9         V10         V11         V12
## 0.129489675 0.133204568 0.136819626 0.140262144 0.143445330 0.146283400
##         V13         V14         V15         V16         V17         V18
## 0.148702051 0.150642294 0.152070438 0.152981568 0.153400217 0.153376850
##         V19         V20         V21         V22         V23         V24
## 0.152978710 0.152284670 0.151380523 0.150350076 0.149267301 0.148191531
##         V25         V26         V27         V28         V29         V30
## 0.147166701 0.146223240 0.145379157 0.144640670 0.144009814 0.143486390
##         V31         V32         V33         V34         V35         V36
## 0.143064270 0.142731550 0.142474258 0.142281935 0.142146372 0.142059275
##         V37         V38         V39         V40         V41         V42
## 0.142011366 0.141997660 0.142015094 0.142060228 0.142128038 0.142213924
##         V43         V44         V45         V46         V47         V48
## 0.142314550 0.142427200 0.142549721 0.142682087 0.142824682 0.142978128
##         V49         V50         V51
## 0.143143040 0.143320026 0.143509709
```

```r
#PC2
df.PC[,2]
```

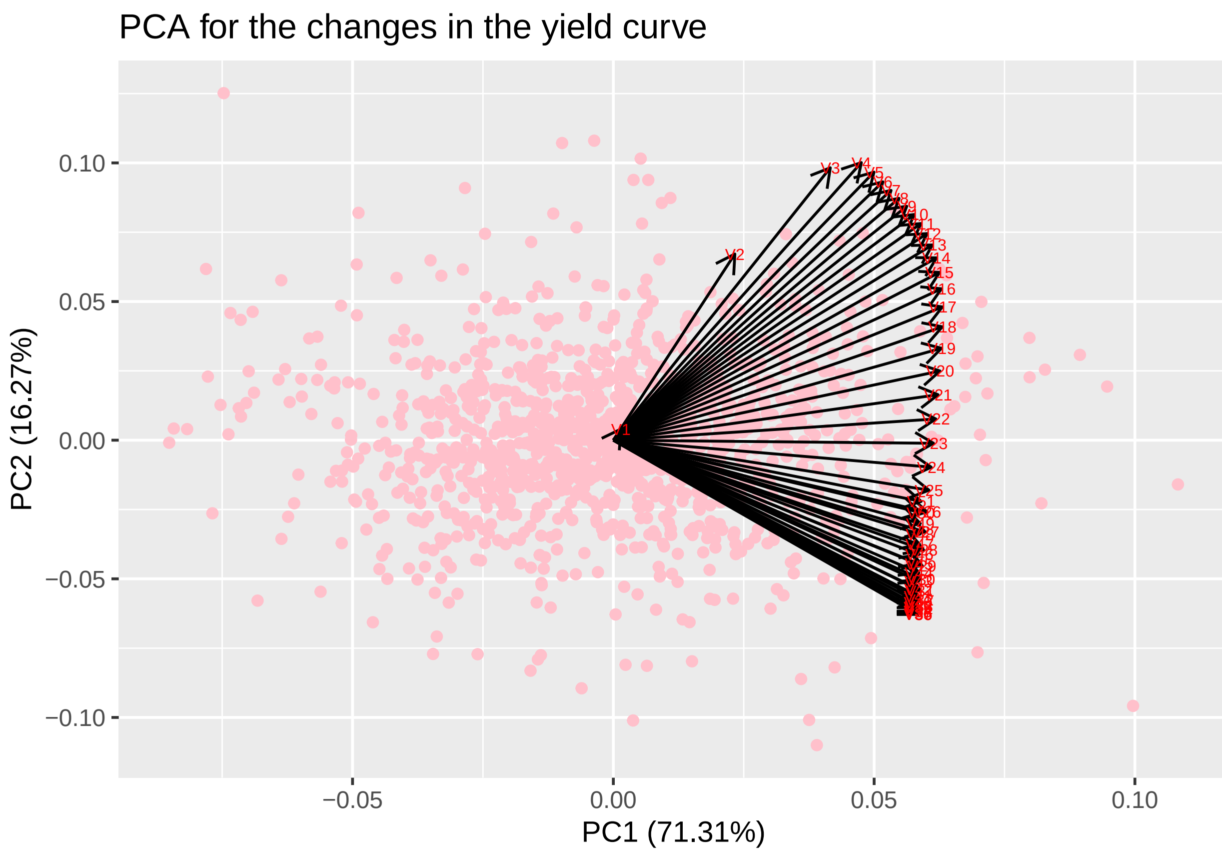
```
##           V1           V2           V3           V4           V5           V6
##  0.009726247  0.163267177  0.238914897  0.243456094  0.235098718  0.226563681
##           V7           V8           V9          V10          V11          V12
##  0.219032346  0.212069610  0.205163803  0.197917152  0.190010712  0.181180420
##          V13          V14          V15          V16          V17          V18
##  0.171207480  0.159904262  0.147125062  0.132784279  0.116867874  0.099435990
##          V19          V20          V21          V22          V23          V24
##  0.080629957  0.060676894  0.039880025  0.018609589 -0.002715267 -0.023672253
##          V25          V26          V27          V28          V29          V30
## -0.043846775 -0.062856253 -0.080376376 -0.096159741 -0.110047038 -0.121954254
##          V31          V32          V33          V34          V35          V36
## -0.131845594 -0.139729912 -0.145661595 -0.149736871 -0.152073753 -0.152796181
##          V37          V38          V39          V40          V41          V42
## -0.152032879 -0.149915493 -0.146579984 -0.142163854 -0.136808390 -0.130650297
##          V43          V44          V45          V46          V47          V48
## -0.123816902 -0.116413611 -0.108517480 -0.100185958 -0.091470697 -0.082418218
##          V49          V50          V51
## -0.073073392 -0.063481034 -0.053685435
```

```r
#PC3
df.PC[,3]
```

```
##           V1           V2           V3           V4           V5           V6
## -0.001115077  0.273137839  0.402224226  0.355810175  0.274742514  0.195850276
##           V7           V8           V9          V10          V11          V12
##  0.125003213  0.062353904  0.007093352 -0.041353531 -0.083251916 -0.118657570
##          V13          V14          V15          V16          V17          V18
## -0.147545723 -0.169903360 -0.185819864 -0.195521879 -0.199364521 -0.197852905
##          V19          V20          V21          V22          V23          V24
## -0.191598754 -0.181299456 -0.167702132 -0.151556881 -0.133589858 -0.114476184
##          V25          V26          V27          V28          V29          V30
## -0.094812351 -0.075130973 -0.055870065 -0.037361788 -0.019843564 -0.003491718
```

```
##          V31         V32         V33         V34         V35         V36
##  0.011550395 0.025185568 0.037365043 0.048080027 0.057341318 0.065186542
##          V37         V38         V39         V40         V41         V42
##  0.071667088 0.076855843 0.080836792 0.083696157 0.085528679 0.086437634
##          V43         V44         V45         V46         V47         V48
##  0.086520941 0.085864621 0.084542159 0.082623868 0.080177140 0.077263301
##          V49         V50         V51
##  0.073941700 0.070271673 0.066311970
```

```r
autoplot(df.pca,  loadings = TRUE, loadings.colour = 'black',         loadings.label = TRUE, col="pink"
         loading.label.color = 'black',
         loadings.label.repel=T,
         main="PCA for the changes in the yield curve")
```



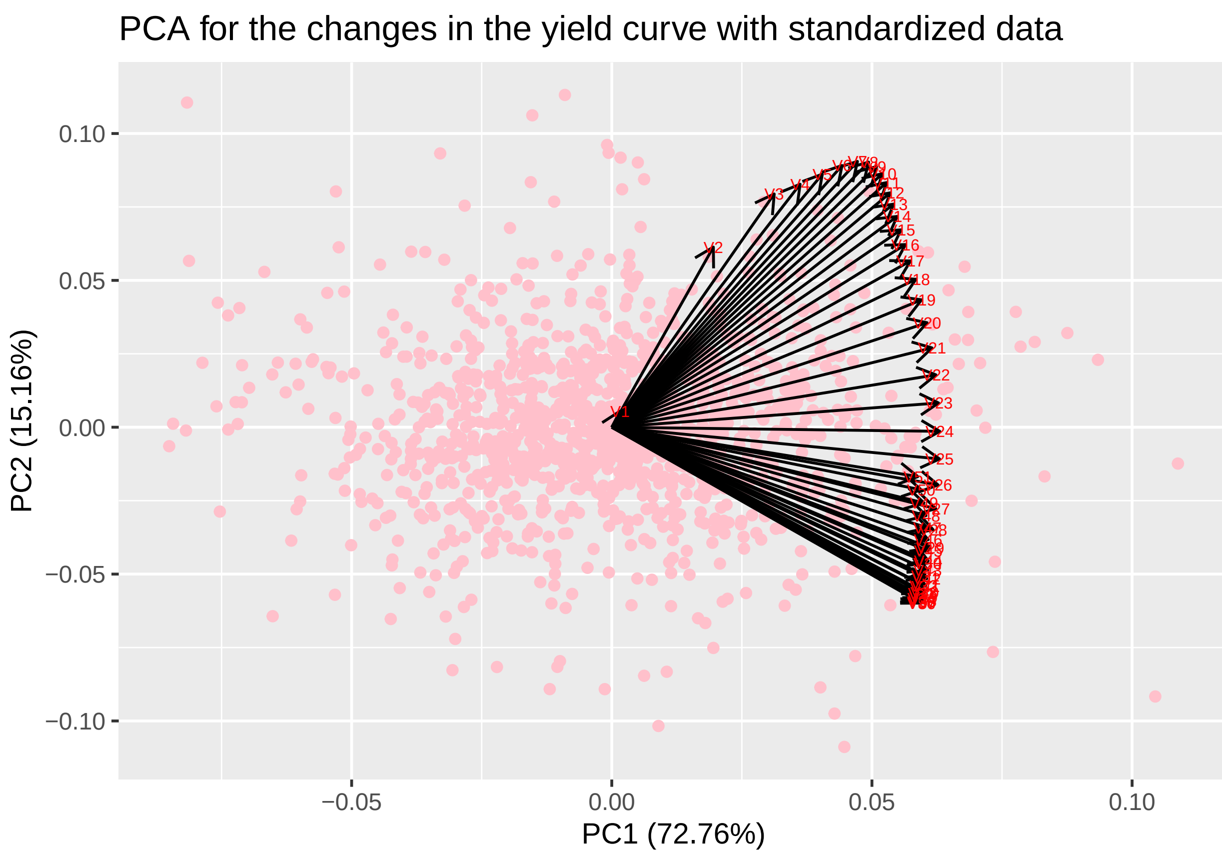
The first principal component is influenced equally by all the yield except for the overnight change in the yield and the changed yield after 6 months. This might be due to the fact that we have not collected enough information in the first 6 months to come up with a pattern for the yield.

The second principal component weakly depends on the overnight change of the yield, and is negatively affected by the changes in the yield from year 11 onward.

The third principal components heavily depends on the changes in the year 1 to year 3 compared to other changes. The effect of the changes from year 1 to year 3 is opposite than the effect of other changes.

```r
df.q4 = scale(df_change)
df.q4.pca = prcomp(df.q4)
autoplot(df.q4.pca,  loadings = TRUE, loadings.colour = 'black',
```

```
        loadings.label = TRUE, col="pink",loadings.label.size = 2,
        loading.label.color = 'black',loadings.label.repel=T,main="PCA for the changes in the yield cu
```



In this case, there is no need to standardize the data because:

Reason 1: The graph of PCA for the changes in the yield curve and the graph of PCA for the changes in the yield curve with standardized data look similar.

Reason 2: Since the data set "df_change" only has observations between -1 and 1, so it's similar to the scaled data set already.

5

Use the principal component analysis to explain the changes in the yield curve over time.

```
df.PCA = prcomp(df)
PC = df.PCA$rotation
#PC1
PC[,1]
```

```
##           V1           V2           V3           V4           V5           V6
## -0.268983741 -0.289377580 -0.264582429 -0.225487782 -0.188563636 -0.156632109
##           V7           V8           V9          V10          V11          V12
## -0.129084415 -0.105035317 -0.083691936 -0.064296476 -0.046187523 -0.028871070
##          V13          V14          V15          V16          V17          V18
```

```
## -0.012058921  0.004333145  0.020252413  0.035564384  0.050123998  0.063815015
##          V19          V20          V21          V22          V23          V24
##  0.076579219  0.088420202  0.099373375  0.109490936  0.118824485  0.127390379
##          V25          V26          V27          V28          V29          V30
##  0.135177551  0.142151360  0.148267436  0.153464293  0.157701002  0.160968493
##          V31          V32          V33          V34          V35          V36
##  0.163276990  0.164639514  0.165070848  0.164623369  0.163372764  0.161405078
##          V37          V38          V39          V40          V41          V42
##  0.158807466  0.155653750  0.152017246  0.147967760  0.143573865  0.138902671
##          V43          V44          V45          V46          V47          V48
##  0.134021209  0.128991671  0.123851602  0.118620685  0.113315023  0.107949124
##          V49          V50          V51
##  0.102537093  0.097093051  0.091631175
```

```
#PC2
PC[,2]
```

```
##          V1          V2          V3          V4          V5          V6
## -0.04736740 -0.22698243 -0.30503982 -0.30045302 -0.27944671 -0.25770364
##          V7          V8          V9         V10         V11         V12
## -0.23741449 -0.21883634 -0.20201952 -0.18704590 -0.17395905 -0.16264552
##         V13         V14         V15         V16         V17         V18
## -0.15290401 -0.14450050 -0.13721260 -0.13082817 -0.12519946 -0.12020382
##         V19         V20         V21         V22         V23         V24
## -0.11574216 -0.11168374 -0.10792289 -0.10439713 -0.10107566 -0.09795765
##         V25         V26         V27         V28         V29         V30
## -0.09509219 -0.09252779 -0.09032900 -0.08855201 -0.08719034 -0.08622429
##         V31         V32         V33         V34         V35         V36
## -0.08563041 -0.08536920 -0.08537794 -0.08558986 -0.08593397 -0.08635494
##         V37         V38         V39         V40         V41         V42
## -0.08680541 -0.08723201 -0.08757815 -0.08781126 -0.08791242 -0.08786412
##         V43         V44         V45         V46         V47         V48
## -0.08765062 -0.08726438 -0.08669968 -0.08595940 -0.08505311 -0.08399496
##         V49         V50         V51
## -0.08279985 -0.08148217 -0.08005287
```

```
#PC3
PC[,3]
```

```
##            V1            V2            V3            V4            V5
## -0.5831706864 -0.3732353869 -0.1670073999 -0.0411352169  0.0376032008
##            V6            V7            V8            V9           V10
##  0.0863509303  0.1157978722  0.1334421968  0.1439319785  0.1500389699
##           V11           V12           V13           V14           V15
##  0.1533028194  0.1545497976  0.1540221532  0.1517174843  0.1474857657
##           V16           V17           V18           V19           V20
##  0.1411483141  0.1325671617  0.1218125686  0.1090900861  0.0947498827
##           V21           V22           V23           V24           V25
##  0.0792668132  0.0631640547  0.0469333175  0.0309637566  0.0154807930
##           V26           V27           V28           V29           V30
##  0.0005977429 -0.0135593805 -0.0269786444 -0.0396179218 -0.0514044671
##           V31           V32           V33           V34           V35
## -0.0621946710 -0.0718834716 -0.0804504401 -0.0879255905 -0.0943635980
```

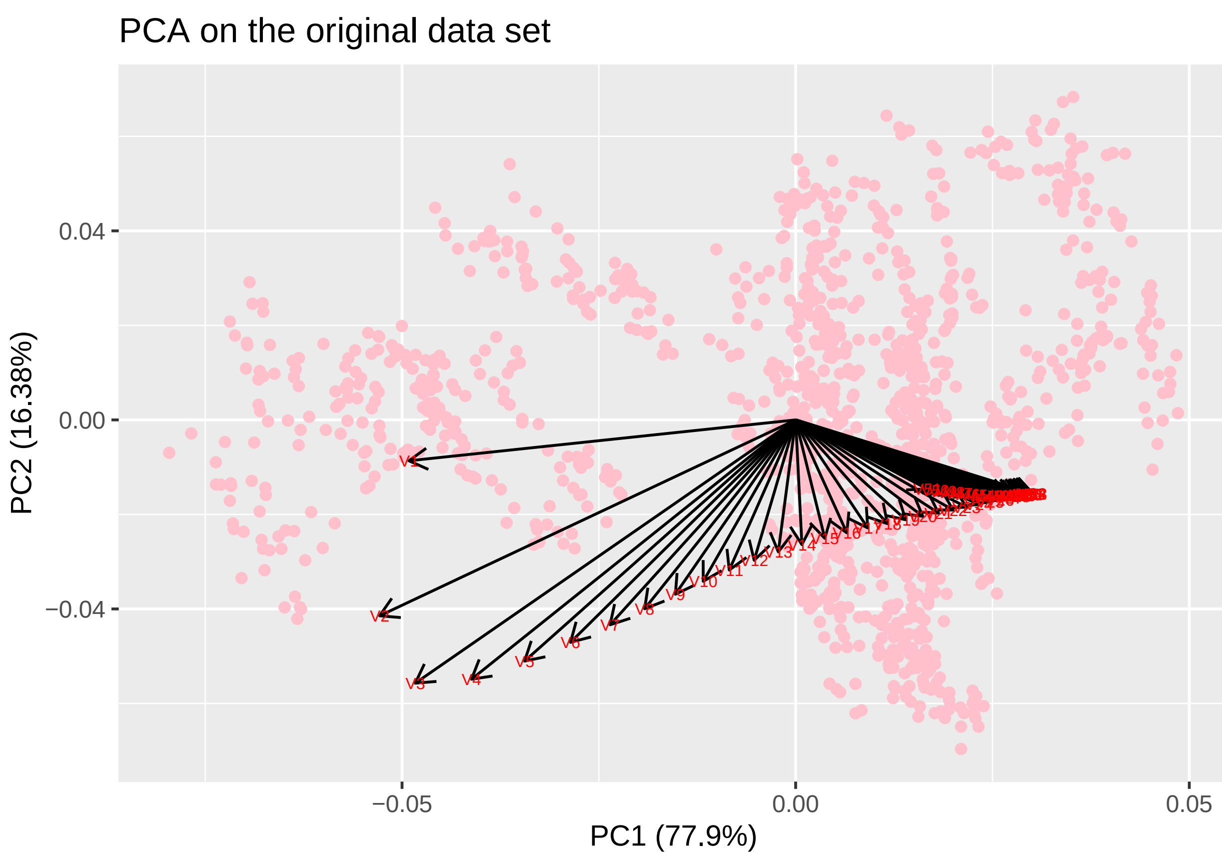
```
##            V36           V37           V38           V39           V40
## -0.0997464050 -0.1040404124 -0.1072334709 -0.1093394434 -0.1104301294
##            V41           V42           V43           V44           V45
## -0.1106117878 -0.1100242302 -0.1087906225 -0.1070022189 -0.1047403132
##            V46           V47           V48           V49           V50
## -0.1020555588 -0.0989939366 -0.0956101062 -0.0919636079 -0.0881149530
##            V51
## -0.0841291419
```

The first principal component of the original data set is influenced equally by the data from year 11 to year 25, and the remaining data is either close to 0 or have the opposite sign. It also has a wider range compared to the first principal components of the changed data set.

The second principal components only contains negative value. Therefore, the effect on the yield is in the opposite direction.

PCA on the original data set

```
library(ggfortify)
library(ggplot2)
autoplot(df.PCA,  loadings = TRUE, loadings.colour = 'black',
         loadings.label = TRUE, col="pink",loadings.label.size = 2,
         loading.label.color = 'black',loadings.label.repel=T, main="PCA on the original data set")
```



For the first half of the 25-year period, we can not draw a pattern from the yield in the original data set. That is because the plot for the original data set is scattered and the principal components have different directions. Also, there is no clear pattern in the first three principal components of the original data set. Therefore, the correlation between the yields in the first 10 years are insignificant. For the remaining part of the data set, there are more similarities between the direction of the principal components, so we can conclude that there exists dependency.