Assignment 2

Vu The Doan (12918687), Aljer Lee Zhen Yee (12563412)

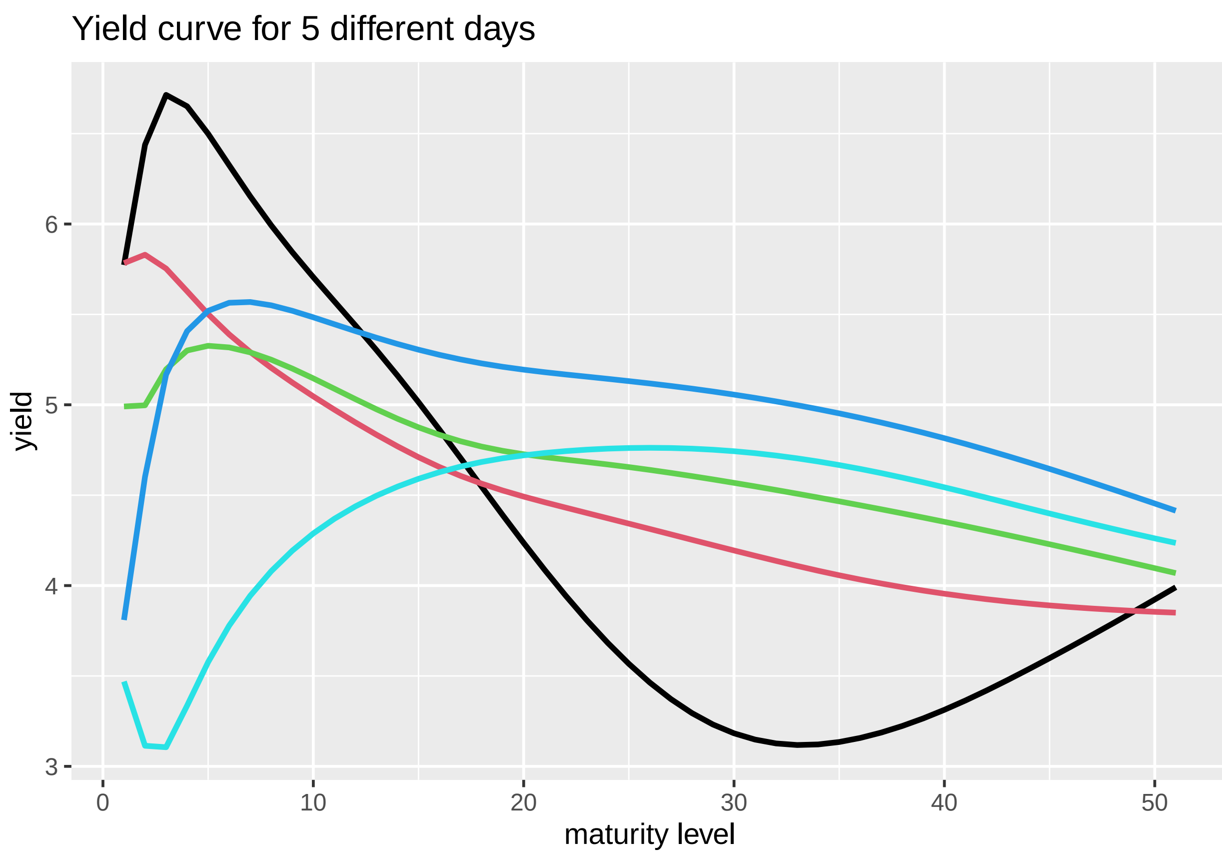
6/26/2021

1.

```
Data=data.matrix(read.table("InterestRates.txt", header=FALSE))
```

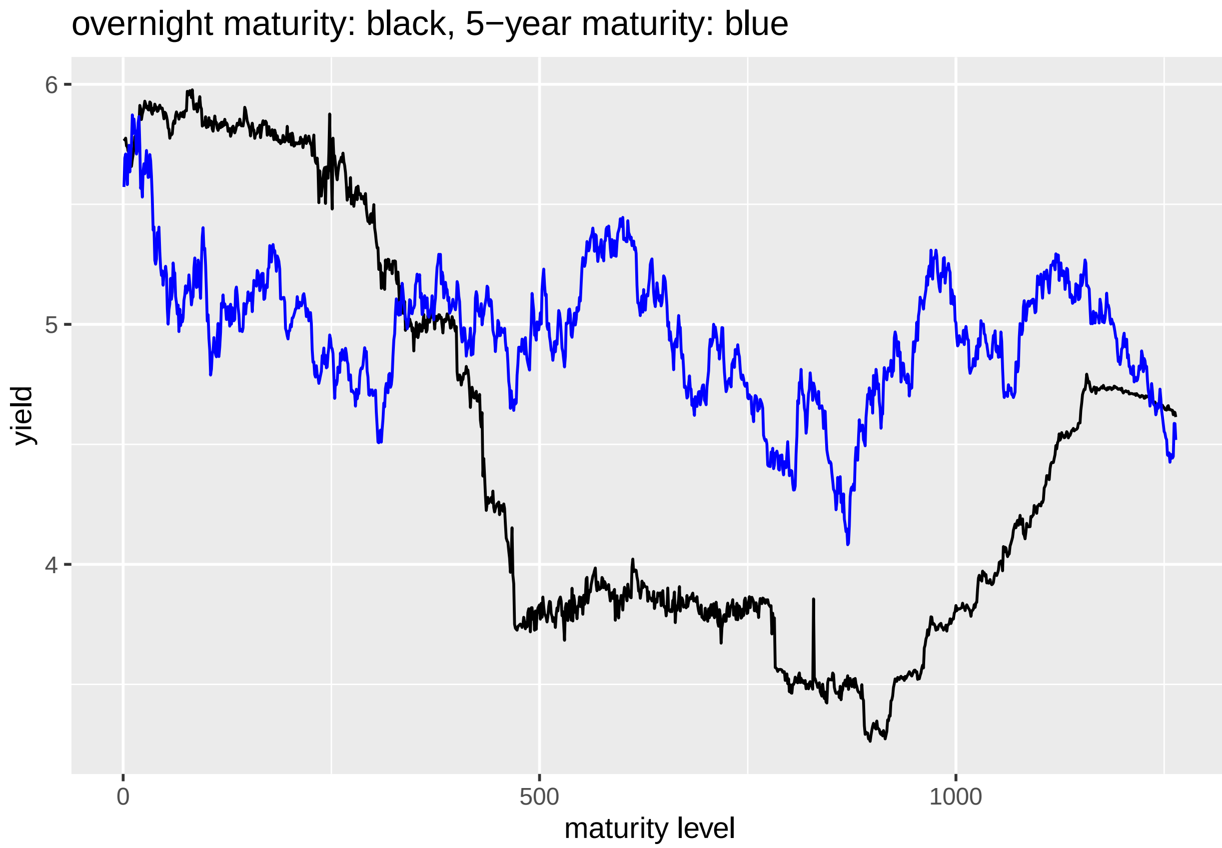
Plot of the yield curve for day 1, day 200, day 400, day 600 and day 800:

```
library(ggplot2)
d1 <- NULL
for (i in 1:ncol(Data)) {
  d1[i] <- Data[1,i]
}
d200 <- NULL
for (i in 1:ncol(Data)) {
  d200[i] <- Data[200,i]
}
d400<- NULL
for (i in 1:ncol(Data)) {
  d400[i] <- Data[400,i]
}
d600 <- NULL
for (i in 1:ncol(Data)) {
  d600[i] <- Data[600,i]
}
d800 <- NULL
for (i in 1:ncol(Data)) {
  d800[i] <- Data[800,i]
}
ggplot() +
  geom_line(mapping = aes(x=1:51, y=d1), col = 1, size = 1) +
  geom_line(mapping = aes(x=1:51, y=d200), col = 2, size = 1) +
  geom_line(mapping = aes(x=1:51, y=d400), col = 3, size = 1) +
  geom_line(mapping = aes(x=1:51, y=d600), col = 4, size = 1) +
  geom_line(mapping = aes(x=1:51, y=d800), col = 5, size = 1) +
  labs(x="maturity level", y="yield") +
  ggtitle("Yield curve for 5 different days")
```



2. Plot of the overnight yield and 5-year maturity yield:

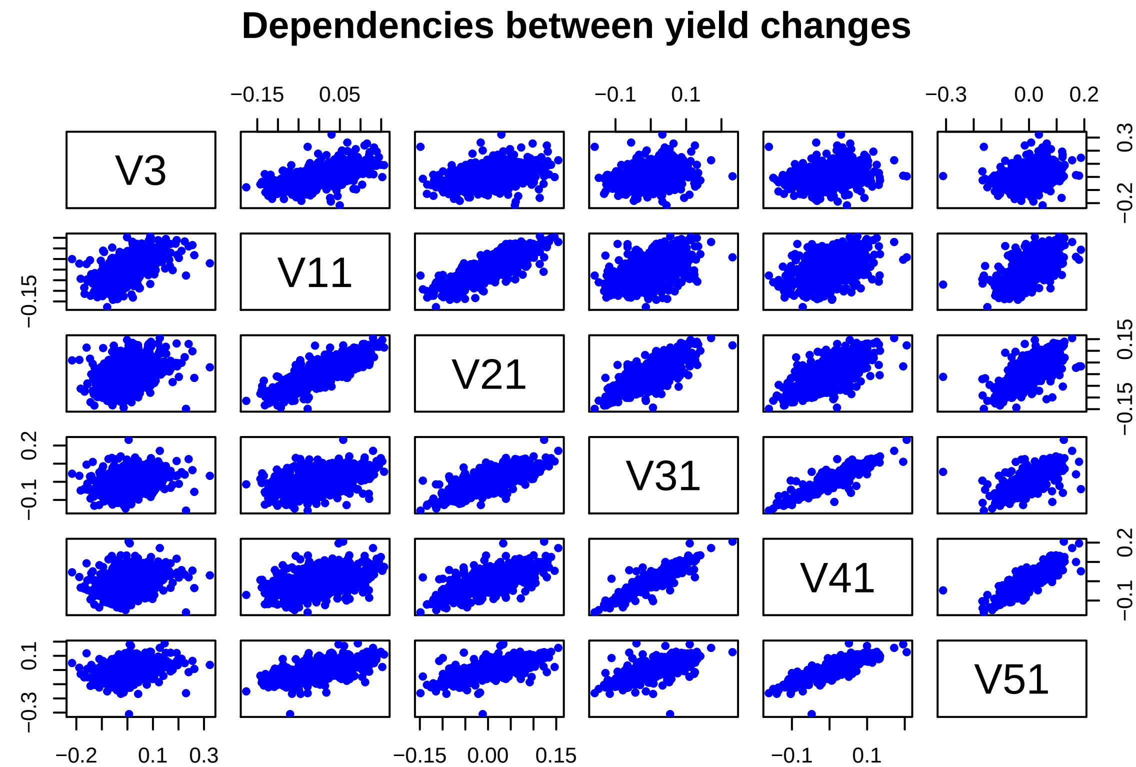
```
v1 <- NULL
for (i in 1:nrow(Data)) {
  v1[i] <- Data[i, 1]
}
v11 <- NULL
for (i in 1:nrow(Data)) {
  v11[i] <- Data[i, 11]
}
ggplot() +
  geom_line(mapping = aes(x=1:nrow(Data), y=v1), col = "black") +
  geom_line(mapping = aes(x=1:nrow(Data), y=v11), col = "blue") +
  labs(x="maturity level", y="yield") +
  ggtitle("overnight maturity: black, 5-year maturity: blue")
```



The overnight yield fluctuates heavily. It decreases from 5.773 to a minimum at 3.263, then increases to 4.614 in the end. The 5-year yield decreases overall, with the starting value of 5.572 and the end value of 4.518.

3. The pairwise scatterplot for the changes in the yield for the 1-year, 5-year, 10-year, 15-year, 20-year and 25-year maturity:

```
Change <- Data[2:1264,] - Data[1:1263,]
pairs(~V3 + V11 + V21 + V31 + V41 + V51, data=Change, pch=20, col="blue",
      main = "Dependencies between yield changes")
```



Dependency can be explained by the pairwise scatterplot, and based on the scatterplot, the linear relation to each variable to the subsequent Year 5, 10, 15, 20 and 25 can be observed. The dependency of Year 5 to the subsequent years is shown above and a pattern can be observed, where the dependency of the Year 5 to Year 10 is strongest and the subsequent years decreases as it approaches Year 25, which is depicted by the linear relationship of the scatterplot. On the other hand, there are further findings that can be observed:

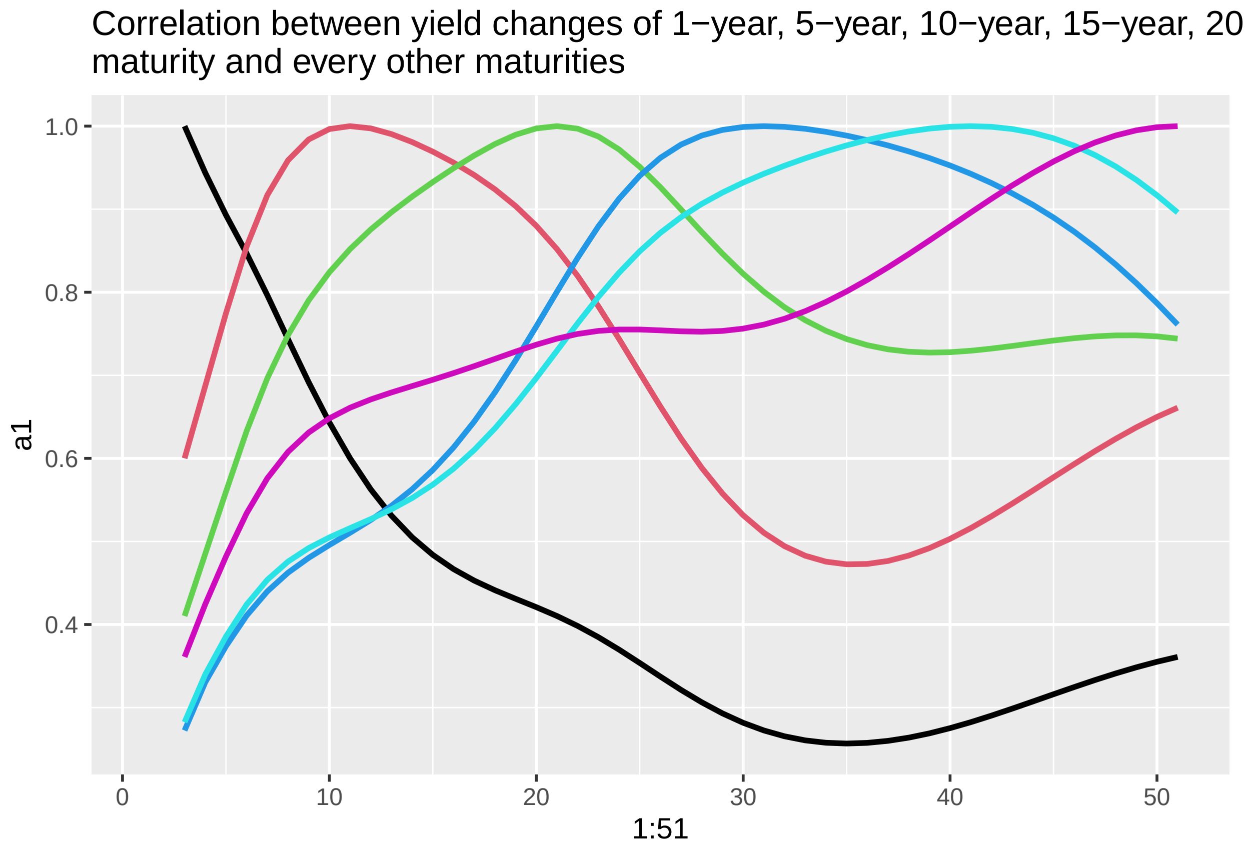
- Observation 1: Year 1 has a stronger dependency to Year 5, as compared to the subsequent years later as the dependency decreases, which can be observed from the linearity of the scatterplot.
- Observation 2: Year 5 has a stronger dependency to Year 10, as compared to the subsequent years later and its dependency is much more prominent that Observation 1.
- Observation 3: Year 15 has the strongest dependency to Year 20, as compared to the previous two observations.

The correlation between the yield changes of 1-year (black), 5-year (red), 10-year (green), 15-year (blue), 20-year (teal) and 25-year (teal) maturity and every other maturities:

```
a1<- NULL
for (i in 3:51) {
  a1[i] <- cor(Change[,3], Change[,i])
}
a11<- NULL
for (i in 3:51) {
  a11[i] <- cor(Change[,11], Change[,i])
}
a21<- NULL
```

```r
for (i in 3:51) {
  a21[i] <- cor(Change[,21], Change[,i])
}
a31<- NULL
for (i in 3:51) {
  a31[i] <- cor(Change[,31], Change[,i])
}
a41<- NULL
for (i in 3:51) {
  a41[i] <- cor(Change[,41], Change[,i])
}
a51<- NULL
for (i in 3:51) {
  a51[i] <- cor(Change[,51], Change[,i])
}
ggplot() +
  geom_line(mapping = aes(x=1:51, y=a1), col = 1, size = 1) +
  geom_line(mapping = aes(x=1:51, y=a11), col = 2, size = 1) +
  geom_line(mapping = aes(x=1:51, y=a21), col = 3, size = 1) +
  geom_line(mapping = aes(x=1:51, y=a31), col = 4, size = 1) +
  geom_line(mapping = aes(x=1:51, y=a41), col = 5, size = 1) +
  geom_line(mapping = aes(x=1:51, y=a51), col = 6, size = 1) +
  ggtitle("Correlation between yield changes of 1-year, 5-year, 10-year, 15-year, 20-year and 25-year \r
```

```
## Warning: Removed 2 row(s) containing missing values (geom_path).

## Warning: Removed 2 row(s) containing missing values (geom_path).

## Warning: Removed 2 row(s) containing missing values (geom_path).

## Warning: Removed 2 row(s) containing missing values (geom_path).

## Warning: Removed 2 row(s) containing missing values (geom_path).

## Warning: Removed 2 row(s) containing missing values (geom_path).
```



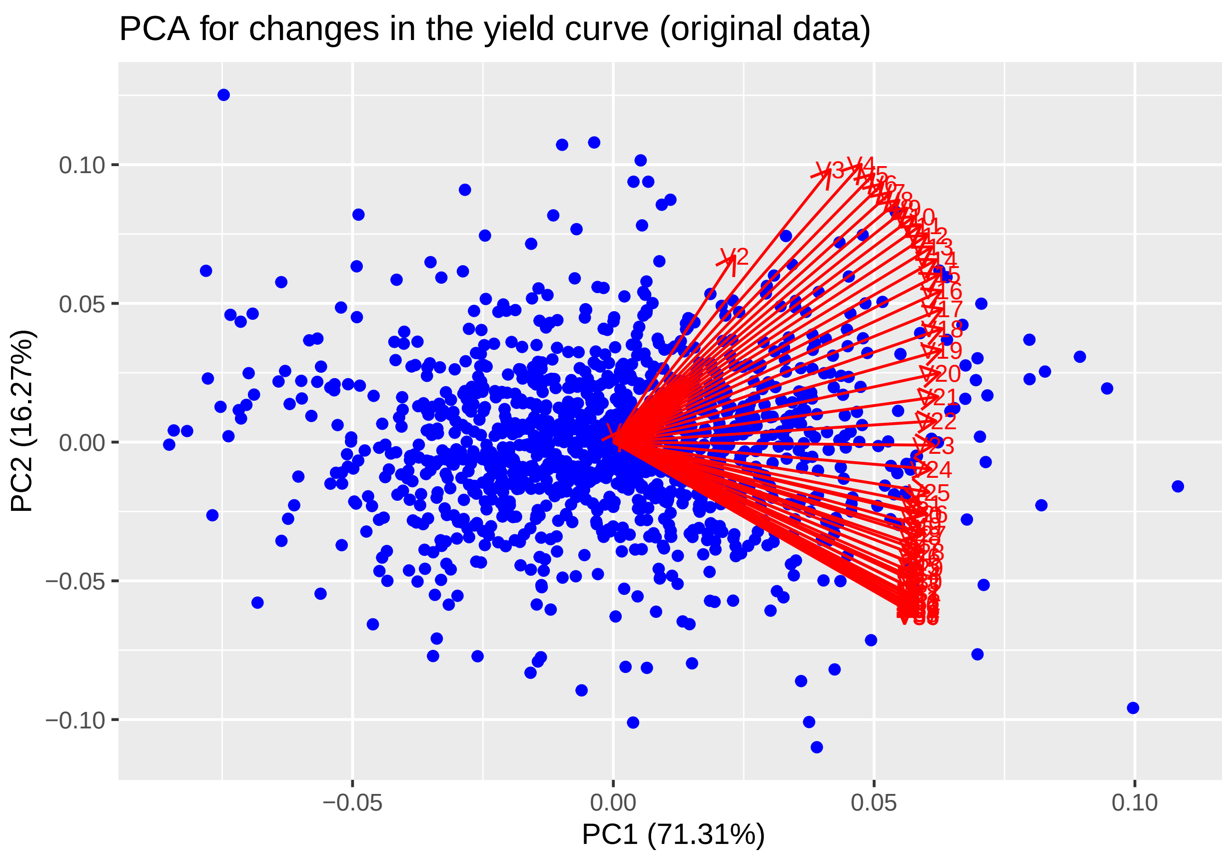
Based on the line graph, these are the findings that can be observed:

- Observation 1: With a comparison from Year 1 to the subsequent years, it can be seen that the correlation decreases to below 0.5 in Year 25.
- Observation 2: With a comparison from Year 15 to subsequent years, it can be seen that the correlation is decreasing as well, however, it's higher than Observation 1 which is 0.75 at Year 25.
- Observation 3: With a comparison from Year 20 to subsequent years, it can be seen that the correlation decreases, however, it's higher than Observation 1 and Observation 2 at 0.9 at Year 25.

Therefore, based on the observation, a pattern can be concluded. The closer the base year we used for comparison, to the year of maturity, the higher the correlation is.

4. Principal component analysis on the original "Change" dataset:

```
Change.PCA <- prcomp(Change)
library(ggfortify)
autoplot(Change.PCA, loadings = TRUE, loadings.label = TRUE, loadings.label.size = 3, col = "blue",
         main = "PCA for changes in the yield curve (original data)")
```




```r
summary(Change.PCA)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     0.2837 0.1355 0.08059 0.05801 0.04499 0.03599 0.02471
## Proportion of Variance 0.7131 0.1627 0.05753 0.02980 0.01793 0.01148 0.00541
## Cumulative Proportion  0.7131 0.8758 0.93332 0.96312 0.98105 0.99253 0.99794
##                            PC8      PC9     PC10     PC11      PC12      PC13
## Standard deviation     0.01245 0.007596 0.003981 0.001753 0.0009451 0.0005199
## Proportion of Variance 0.00137 0.000510 0.000140 0.000030 0.0000100 0.0000000
## Cumulative Proportion  0.99931 0.999820 0.999960 0.999990 1.0000000 1.0000000
##                             PC14      PC15      PC16     PC17     PC18    PC19
## Standard deviation     0.0002852 0.0001816 0.0001124 8.21e-05 6.92e-05 5.7e-05
## Proportion of Variance 0.0000000 0.0000000 0.0000000 0.00e+00 0.00e+00 0.0e+00
## Cumulative Proportion  1.0000000 1.0000000 1.0000000 1.00e+00 1.00e+00 1.0e+00
##                             PC20      PC21      PC22      PC23      PC24
## Standard deviation     4.299e-05 3.097e-05 2.243e-05 1.763e-05 1.271e-05
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC25      PC26      PC27      PC28      PC29
## Standard deviation     1.019e-05 8.428e-06 6.523e-06 4.986e-06 3.655e-06
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC30      PC31      PC32      PC33      PC34
## Standard deviation     3.389e-06 2.886e-06 2.834e-06 2.189e-06 1.972e-06
```

```
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC35      PC36      PC37      PC38      PC39
## Standard deviation     1.703e-06 1.524e-06 1.411e-06 1.337e-06 1.299e-06
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC40      PC41      PC42      PC43      PC44
## Standard deviation     1.092e-06 8.902e-07 7.583e-07 5.866e-07 5.147e-07
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC45      PC46      PC47      PC48      PC49
## Standard deviation     4.617e-07 4.099e-07 3.164e-07 2.243e-07 9.994e-08
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC50      PC51
## Standard deviation     8.623e-08 4.153e-08
## Proportion of Variance 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00
```

According to the summary, we should keep the first 4 principal components to explain 96.31% of the variance.

```
PC <- Change.PCA$rotation
round(PC[,1], digits = 3)
```

```
##    V1    V2    V3    V4    V5    V6    V7    V8    V9   V10   V11   V12   V13
## 0.004 0.057 0.101 0.116 0.122 0.126 0.129 0.133 0.137 0.140 0.143 0.146 0.149
##   V14   V15   V16   V17   V18   V19   V20   V21   V22   V23   V24   V25   V26
## 0.151 0.152 0.153 0.153 0.153 0.153 0.152 0.151 0.150 0.149 0.148 0.147 0.146
##   V27   V28   V29   V30   V31   V32   V33   V34   V35   V36   V37   V38   V39
## 0.145 0.145 0.144 0.143 0.143 0.143 0.142 0.142 0.142 0.142 0.142 0.142 0.142
##   V40   V41   V42   V43   V44   V45   V46   V47   V48   V49   V50   V51
## 0.142 0.142 0.142 0.142 0.142 0.143 0.143 0.143 0.143 0.143 0.143 0.144
```

For PC1, all the variable has the same direction, however, V1 (overnight) and V2 (6-month) has the lowest magnitude, while the rest has a similar magnitude that ranges from 0.1 to 0.15.

```
round(PC[,2], digits = 3)
```

```
##     V1     V2     V3     V4     V5     V6     V7     V8     V9    V10    V11
##  0.010  0.163  0.239  0.243  0.235  0.227  0.219  0.212  0.205  0.198  0.190
##    V12    V13    V14    V15    V16    V17    V18    V19    V20    V21    V22
##  0.181  0.171  0.160  0.147  0.133  0.117  0.099  0.081  0.061  0.040  0.019
##    V23    V24    V25    V26    V27    V28    V29    V30    V31    V32    V33
## -0.003 -0.024 -0.044 -0.063 -0.080 -0.096 -0.110 -0.122 -0.132 -0.140 -0.146
##    V34    V35    V36    V37    V38    V39    V40    V41    V42    V43    V44
## -0.150 -0.152 -0.153 -0.152 -0.150 -0.147 -0.142 -0.137 -0.131 -0.124 -0.116
##    V45    V46    V47    V48    V49    V50    V51
## -0.109 -0.100 -0.091 -0.082 -0.073 -0.063 -0.054
```

For PC2, the variable V3 to V9 (1-year to 4-year) dominates in PC2 with an approximation of the same magnitude in the 0.2 scale, on the other hand, from V23 (11-year) onwards, it has an opposite direction compared to the previous years.

```r
round(PC[,3], digits = 3)
```

```
##     V1     V2     V3     V4     V5     V6     V7     V8     V9    V10    V11
## -0.001  0.273  0.402  0.356  0.275  0.196  0.125  0.062  0.007 -0.041 -0.083
##    V12    V13    V14    V15    V16    V17    V18    V19    V20    V21    V22
## -0.119 -0.148 -0.170 -0.186 -0.196 -0.199 -0.198 -0.192 -0.181 -0.168 -0.152
##    V23    V24    V25    V26    V27    V28    V29    V30    V31    V32    V33
## -0.134 -0.114 -0.095 -0.075 -0.056 -0.037 -0.020 -0.003  0.012  0.025  0.037
##    V34    V35    V36    V37    V38    V39    V40    V41    V42    V43    V44
##  0.048  0.057  0.065  0.072  0.077  0.081  0.084  0.086  0.086  0.087  0.086
##    V45    V46    V47    V48    V49    V50    V51
##  0.085  0.083  0.080  0.077  0.074  0.070  0.066
```

For PC3, the variable V2 to V5 (6-month to 2-year) dominates with the highest magnitude in the positive direction, and in the negative direction, V12 to V24 (5.5-year to 11.5-year) has the highest magnitude.

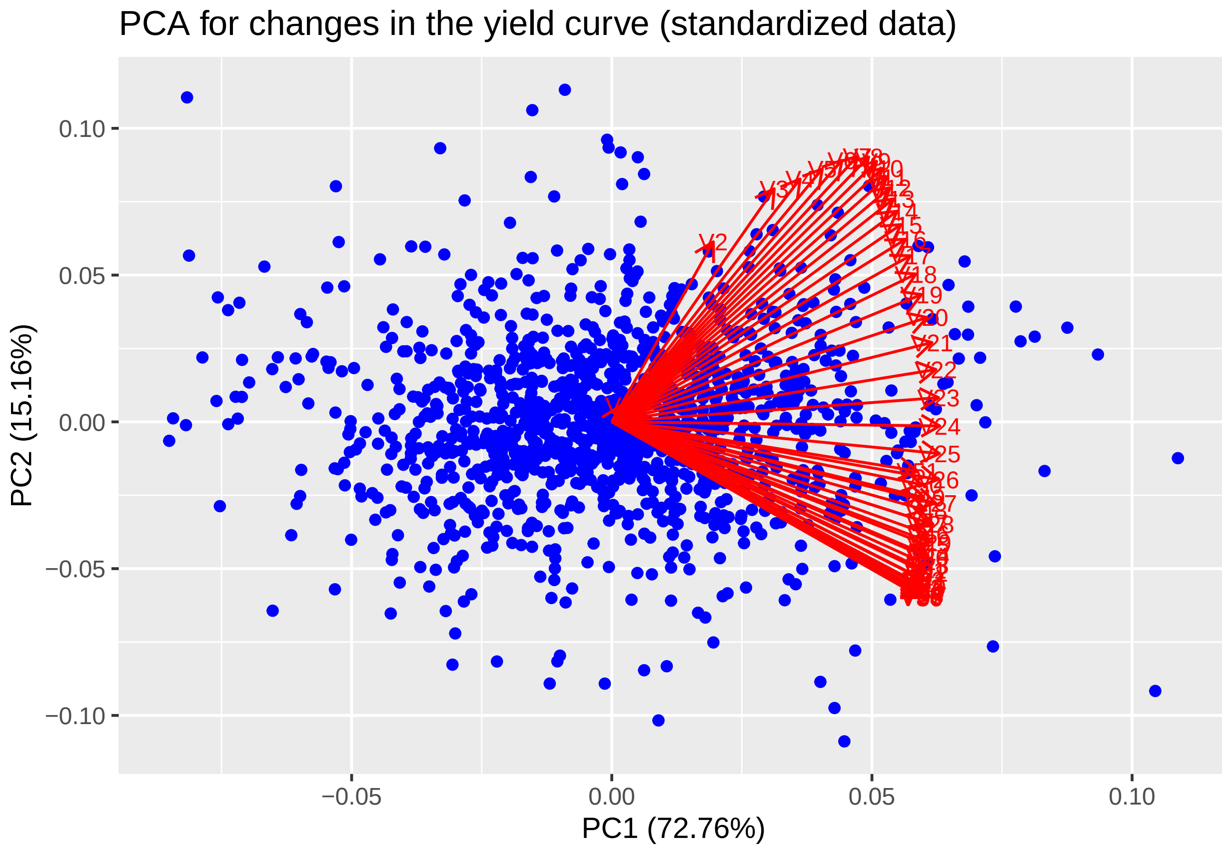
```r
round(PC[,4], digits = 3)
```

```
##     V1     V2     V3     V4     V5     V6     V7     V8     V9    V10    V11
## -0.013 -0.127 -0.198 -0.148 -0.087 -0.035  0.007  0.040  0.065  0.083  0.093
##    V12    V13    V14    V15    V16    V17    V18    V19    V20    V21    V22
##  0.098  0.095  0.087  0.072  0.053  0.029  0.003 -0.026 -0.056 -0.085 -0.112
##    V23    V24    V25    V26    V27    V28    V29    V30    V31    V32    V33
## -0.137 -0.158 -0.174 -0.186 -0.192 -0.194 -0.190 -0.182 -0.169 -0.154 -0.134
##    V34    V35    V36    V37    V38    V39    V40    V41    V42    V43    V44
## -0.113 -0.089 -0.065 -0.039 -0.013  0.014  0.041  0.067  0.094  0.120  0.145
##    V45    V46    V47    V48    V49    V50    V51
##  0.171  0.196  0.220  0.245  0.269  0.293  0.317
```

For PC4, the variable V47 to V51 (23-year to 25-year) dominates with a magnitude of 0.22 to 0.317 with a positive direction, however, there is a change of direction from V1 to V38 (overnight to 18.5-year).

Principal component analysis on the standardized “Change” dataset:

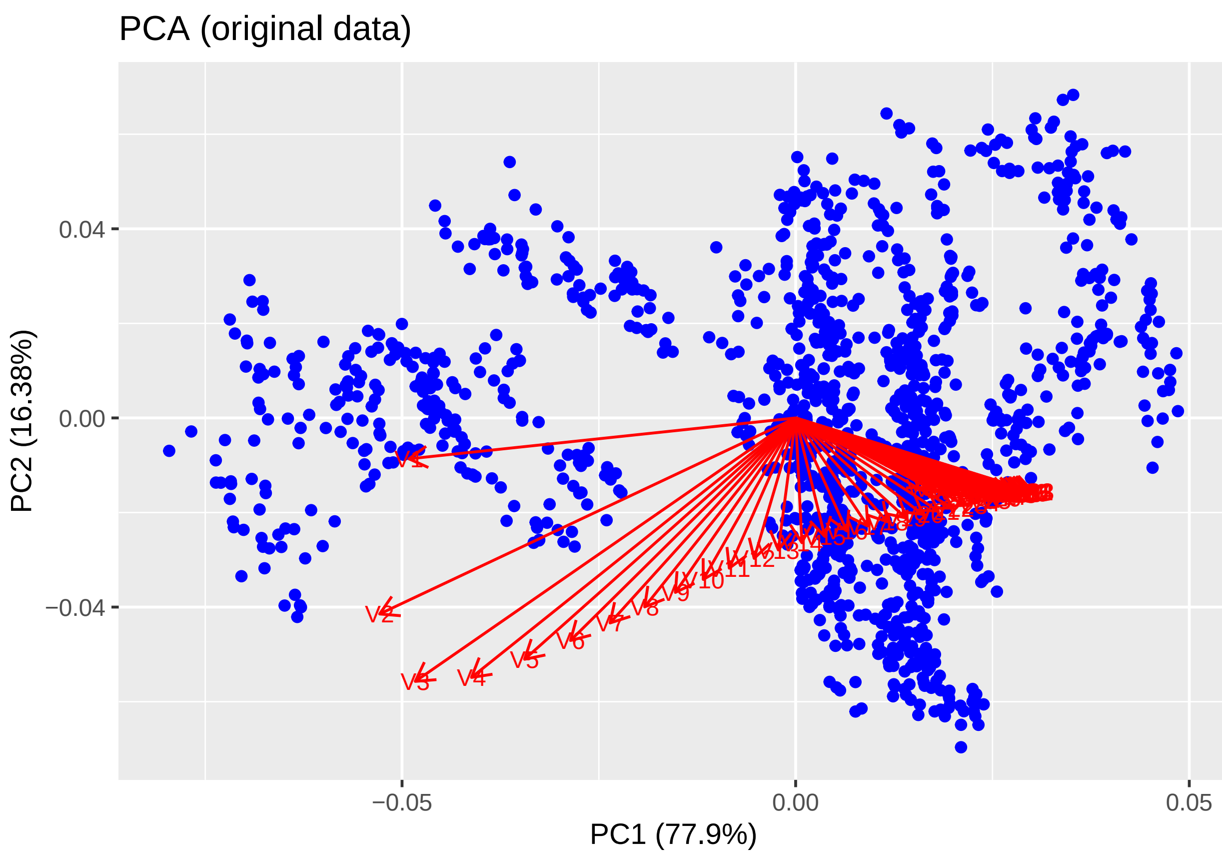
```r
Change_scaled <- scale(Change)
Change_scaled.PCA <- prcomp(Change_scaled)
autoplot(Change_scaled.PCA, loadings = TRUE, loadings.label = TRUE, loadings.label.size = 3,
         col = "blue", main = "PCA for changes in the yield curve (standardized data)")
```



Standardisation is not required because this transformation is done for data of different scale, however, the data that we currently have is of the same scale. Moreover, it can be seen that the biplot of the original data and the standardized data is similar. This further solidify that the data is standardised, and transformation is not required to standardise the given data.

5. Principal component analysis on the original dataset:

```
Data.PCA <- prcomp(Data)
autoplot(Data.PCA, loadings = TRUE, loadings.label = TRUE, loadings.label.size = 3, col = "blue",
         main = "PCA (original data)")
```




```r
summary(Data.PCA)
```

```
## Importance of components:
##                          PC1    PC2    PC3     PC4    PC5     PC6     PC7
## Standard deviation     2.854 1.3087 0.5401 0.42852 0.2999 0.16059 0.06611
## Proportion of Variance 0.779 0.1638 0.0279 0.01756 0.0086 0.00247 0.00042
## Cumulative Proportion  0.779 0.9428 0.9707 0.98825 0.9969 0.99932 0.99974
##                            PC8     PC9    PC10     PC11     PC12     PC13
## Standard deviation     0.04252 0.02679 0.01231 0.007842 0.004507 0.002125
## Proportion of Variance 0.00017 0.00007 0.00001 0.000010 0.000000 0.000000
## Cumulative Proportion  0.99991 0.99998 0.99999 1.000000 1.000000 1.000000
##                            PC14      PC15     PC16      PC17      PC18
## Standard deviation     0.001039 0.0008103 0.000538 0.0003512 0.0002524
## Proportion of Variance 0.000000 0.0000000 0.000000 0.0000000 0.0000000
## Cumulative Proportion  1.000000 1.0000000 1.000000 1.0000000 1.0000000
##                             PC19      PC20      PC21      PC22      PC23
## Standard deviation     0.0002044 0.0001516 0.0001084 6.661e-05 4.686e-05
## Proportion of Variance 0.0000000 0.0000000 0.0000000 0.000e+00 0.000e+00
## Cumulative Proportion  1.0000000 1.0000000 1.0000000 1.000e+00 1.000e+00
##                             PC24      PC25      PC26      PC27      PC28
## Standard deviation     4.439e-05 3.435e-05 2.998e-05 2.035e-05 1.692e-05
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                             PC29     PC30      PC31      PC32      PC33
## Standard deviation     1.225e-05 1.09e-05 9.527e-06 8.849e-06 6.782e-06
```

```
## Proportion of Variance 0.000e+00 0.00e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.00e+00 1.000e+00 1.000e+00 1.000e+00
##                            PC34      PC35      PC36      PC37    PC38
## Standard deviation     6.299e-06 5.631e-06 5.359e-06 5.166e-06 4.7e-06
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.0e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.0e+00
##                            PC39      PC40      PC41      PC42      PC43
## Standard deviation     3.651e-06 3.185e-06 2.684e-06 2.158e-06 2.002e-06
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                            PC44      PC45      PC46      PC47      PC48
## Standard deviation     1.898e-06 1.265e-06 1.077e-06 8.829e-07 6.081e-07
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
##                            PC49      PC50      PC51
## Standard deviation     2.609e-07 2.272e-07 1.015e-07
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00
```

According to the summary, we should keep the first 3 principal components to explain 97.07% of the variance.

```r
Data.PC <- Data.PCA$rotation
round(Data.PC[,1], digits = 3)
```

```
##     V1     V2     V3     V4     V5     V6     V7     V8     V9    V10    V11
## -0.269 -0.289 -0.265 -0.225 -0.189 -0.157 -0.129 -0.105 -0.084 -0.064 -0.046
##    V12    V13    V14    V15    V16    V17    V18    V19    V20    V21    V22
## -0.029 -0.012  0.004  0.020  0.036  0.050  0.064  0.077  0.088  0.099  0.109
##    V23    V24    V25    V26    V27    V28    V29    V30    V31    V32    V33
##  0.119  0.127  0.135  0.142  0.148  0.153  0.158  0.161  0.163  0.165  0.165
##    V34    V35    V36    V37    V38    V39    V40    V41    V42    V43    V44
##  0.165  0.163  0.161  0.159  0.156  0.152  0.148  0.144  0.139  0.134  0.129
##    V45    V46    V47    V48    V49    V50    V51
##  0.124  0.119  0.113  0.108  0.103  0.097  0.092
```

For PC1, the variable V21 to V51 (10-year to 25-year) dominates with the equal magnitude in the positive direction, and in the negative direction, V1 to V13 (overnight to 6-year) has the highest magnitude.

```r
round(Data.PC[,2], digits = 3)
```

```
##     V1     V2     V3     V4     V5     V6     V7     V8     V9    V10    V11
## -0.047 -0.227 -0.305 -0.300 -0.279 -0.258 -0.237 -0.219 -0.202 -0.187 -0.174
##    V12    V13    V14    V15    V16    V17    V18    V19    V20    V21    V22
## -0.163 -0.153 -0.145 -0.137 -0.131 -0.125 -0.120 -0.116 -0.112 -0.108 -0.104
##    V23    V24    V25    V26    V27    V28    V29    V30    V31    V32    V33
## -0.101 -0.098 -0.095 -0.093 -0.090 -0.089 -0.087 -0.086 -0.086 -0.085 -0.085
##    V34    V35    V36    V37    V38    V39    V40    V41    V42    V43    V44
## -0.086 -0.086 -0.086 -0.087 -0.087 -0.088 -0.088 -0.088 -0.088 -0.088 -0.087
##    V45    V46    V47    V48    V49    V50    V51
## -0.087 -0.086 -0.085 -0.084 -0.083 -0.081 -0.080
```

For PC2, all the values are negative. This implies a negative effect on the yield.

```r
round(Data.PC[,3], digits = 3)
```

```
##      V1     V2     V3     V4     V5     V6     V7     V8     V9    V10    V11
## -0.583 -0.373 -0.167 -0.041  0.038  0.086  0.116  0.133  0.144  0.150  0.153
##     V12    V13    V14    V15    V16    V17    V18    V19    V20    V21    V22
##  0.155  0.154  0.152  0.147  0.141  0.133  0.122  0.109  0.095  0.079  0.063
##     V23    V24    V25    V26    V27    V28    V29    V30    V31    V32    V33
##  0.047  0.031  0.015  0.001 -0.014 -0.027 -0.040 -0.051 -0.062 -0.072 -0.080
##     V34    V35    V36    V37    V38    V39    V40    V41    V42    V43    V44
## -0.088 -0.094 -0.100 -0.104 -0.107 -0.109 -0.110 -0.111 -0.110 -0.109 -0.107
##     V45    V46    V47    V48    V49    V50    V51
## -0.105 -0.102 -0.099 -0.096 -0.092 -0.088 -0.084
```

For PC3, the variable V1 (overnight) and V2 (6-month) dominates with the highest magnitude in the negative direction.

The PC1 explains roughly 77.9% of the variance, PC2 explains 16.38% of the variance, and PC3 explains 2.79% of the variance. In sum, the first 3 PC's explain 97.07% of the variance. As a consequence, the yield curve movements can be approximated linear combination of the first three loadings, with small relative error. In other words, a three or four factor model will do a very good job in fitting the time-series yield curve. By looking at the principal components, it can be seen that different maturities impact the yield in different extents. Therefore, there exists the changes in the yield over time.