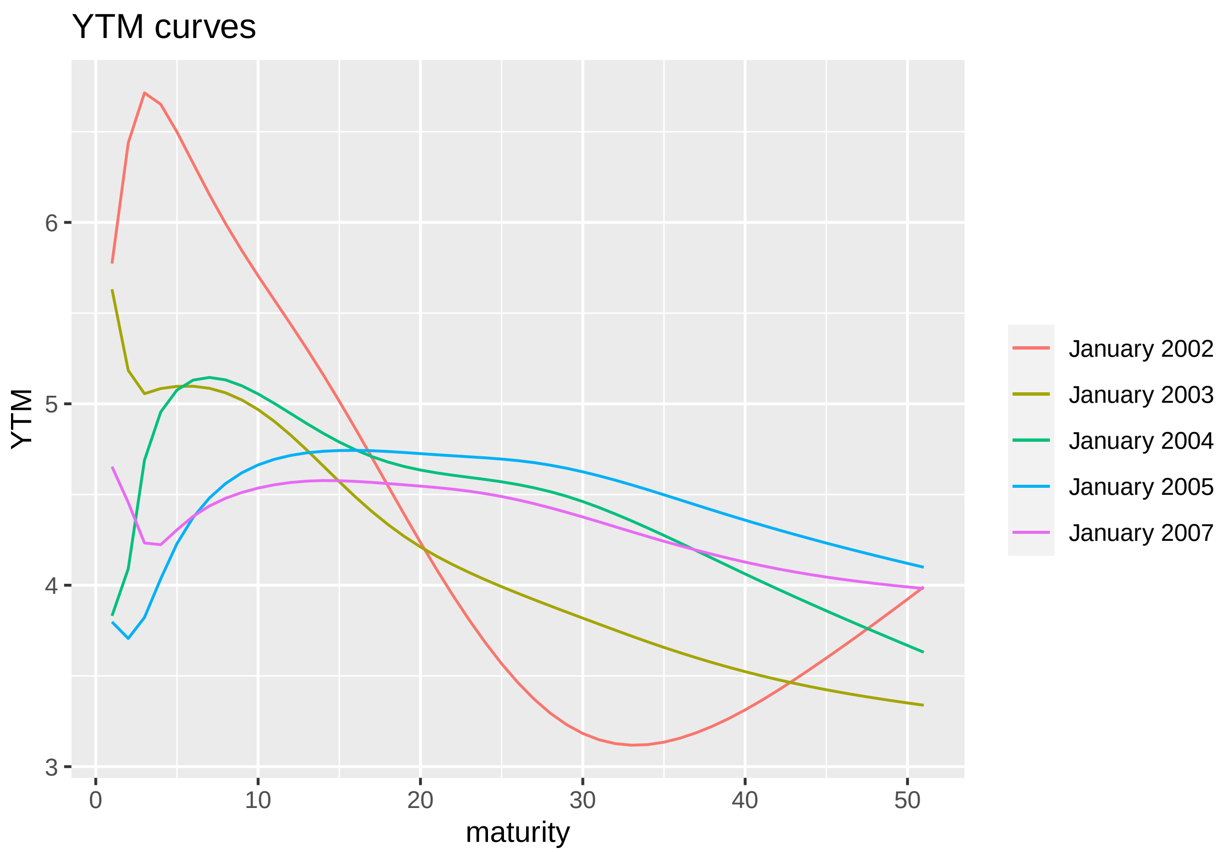
IDS — Assignment 1

Albin Lindqvist (13986236), Zeynep Mersinlioğlu (13616145)

2022-06-25

Question 1

```
library(ggplot2)
Data=data.matrix(read.table("InterestRates.txt", header=FALSE))
# head(Data)
ggplot(as.data.frame(t(Data)),aes(x=1:length(Data[1,]))) + geom_line(aes(y=V1,
↪  col="January 2002")) + geom_line(aes(y=V250, col="January 2003")) +
↪  geom_line(aes(y=V500,col="January 2004")) + geom_line(aes(y=V750, col="January
↪  2005"))  + geom_line(aes(y=V1250, col="January 2007")) +
↪  labs(x="maturity",y="YTM",title="YTM curves",col="")
```



The dates on the legend are approximations based on the following calculation: $\frac{1264}{5} = 252.8$ so, each year has on average $252.8 \approx 250$ observations.

Question 2

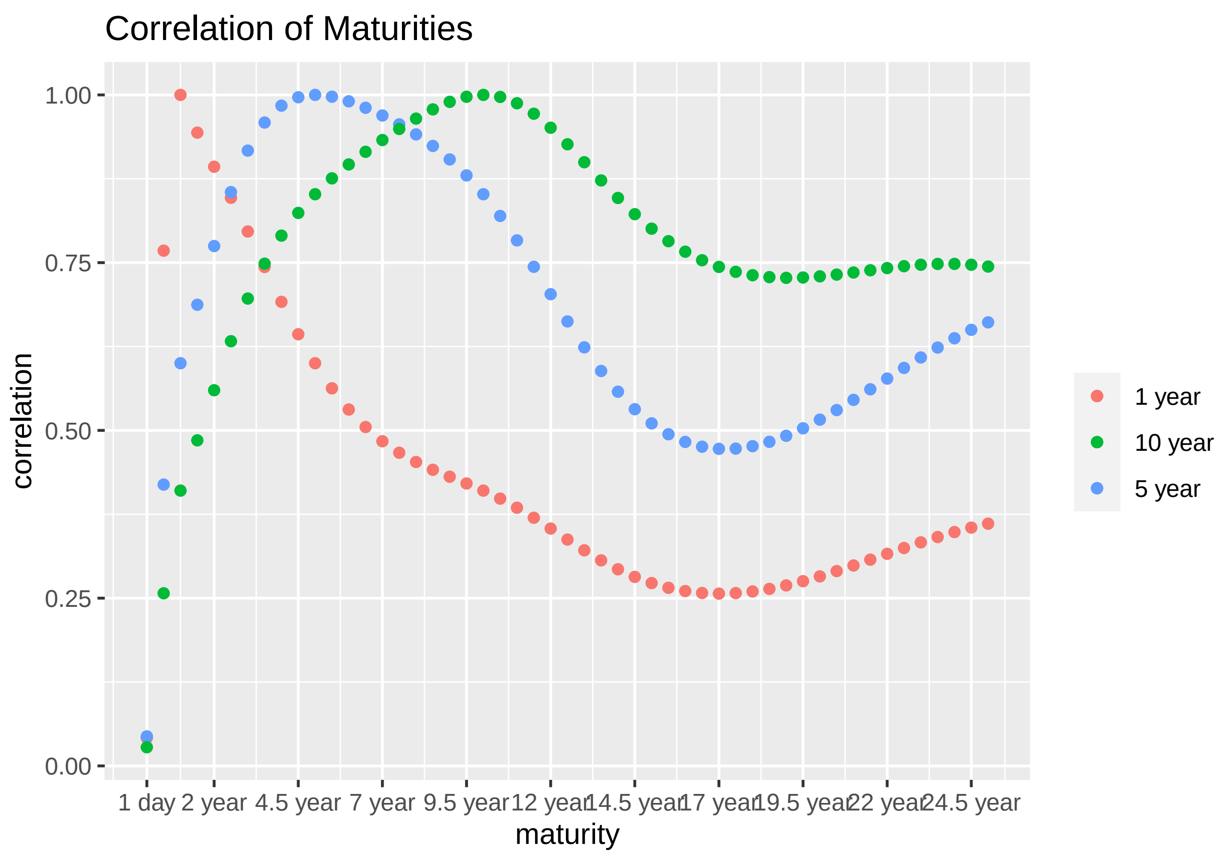
Question 3

```
# Subtracting the previous days yield to maturity with the current days, to get the
↪  difference.
# Data size will be ||Data|| - 1 because there is nothing to subtract the first entry
↪  with.
Data_diff = sapply(1:(length(Data[,1])-1), function(i)Data[i+1,]-Data[i,])

# Transposing the data to get maturity as rows and that days ytm as columns.
Data_diff_t = t(Data_diff)
Data_diff_t_cor = cor(Data_diff_t)

# Creating x tick labels with correct name of the maturity
labels = c()
breaks = c()
for(i in 1:length(Data_diff_t[1,])){
  if (i==1) {
    labels = append(labels, "1 day")
    breaks = append(breaks, i)
  } else if (i %% 5 != 0){
    next
  } else {
    k = (i-1)/2
    labels = append(labels, paste(as.character(k), "year", sep=" "))
    breaks = append(breaks, i)
  }
}

# Visualizing the correlation of the maturities
ggplot(as.data.frame(Data_diff_t_cor), aes(x=1:51)) + geom_point(aes(y=V3, colour="1
↪  year")) + geom_point(aes(y=V11, colour="5 year")) + geom_point(aes(y=V21, colour="10
↪  year")) + labs(x="maturity",y="correlation",title="Correlation of Maturities",col="")
↪  + scale_x_continuous(breaks = breaks, labels=labels)
```



Looking at the produced graph it can be noted that the correlation of the maturities increases the closer it comes to the maturity of that bond. For example, looking at the 10 year maturity one can see that it, of course, reaches 1 at 10 year, but for maturities close to 10 they are also close to 1. This means that there exists a high correlation between the maturities.

It can also be noted that the longer maturities seem to have a strong correlation with the longest maturities. Looking at the 10 year maturity it is clear that it have a strong correlation even from the 4 year maturity to the longest 25 maturity. The one year maturity on the other hand only have a weak correlation with the longer maturities. It is interesting, however, that the one year's maturity correlation increases with maturities longer than 17 year.

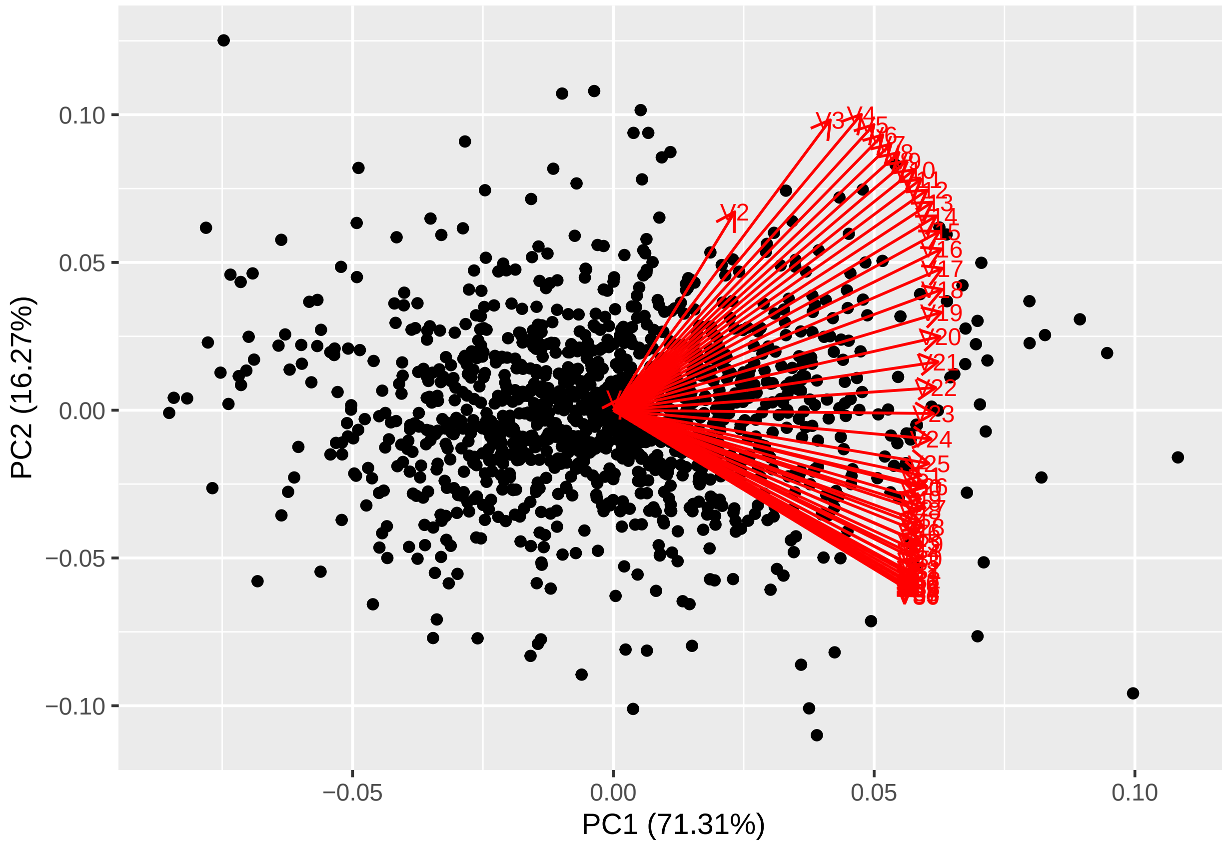
Question 4

We do not need to standardize the data because our data all have the same size. This means that there are no effect of standardizing the data. The reason one wants to standardize the data usually is because the principal component analysis becomes biased. That is, some data that have a larger (distance) spread gets a larger variance, whilst other data that don't have the same distance of spread gets a lower variance. The both data could have the same standardized spread, which means their variance is the same, thus the principal component analysis becomes proportionate with standardization. However, as mentioned earlier, our data does not need to be standardized because it all has the same strength. So a standardization would be redundant.

```
library("ggfortify")

# Principal component analysis
analysis = prcomp(Data_diff_t)
```

```
autoplot(analysis, loadings = TRUE, loadings.label=TRUE, loadings.label.size= 3)
```




```
summary(analysis)
```

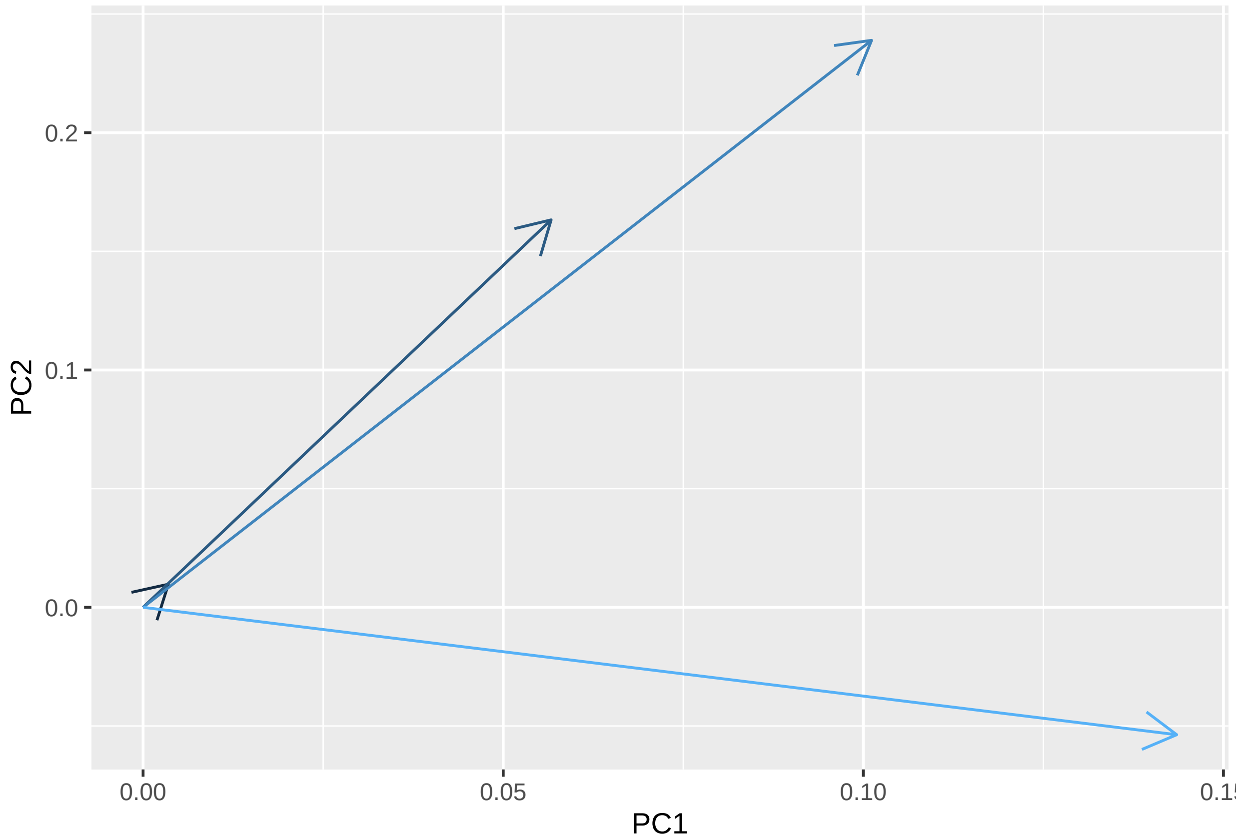
```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     0.2837 0.1355 0.08059 0.05801 0.04499 0.03599 0.02471
## Proportion of Variance 0.7131 0.1627 0.05753 0.02980 0.01793 0.01148 0.00541
## Cumulative Proportion  0.7131 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
```

Looking at the results from the principal component analysis it can be observed that with 12 principal components practically no information is lost. However, with only 4 principal components over 96% of the information is stored. Thus, an appropriate amount of principal components to keep would be 4 if we wanted to minimize it. On the other hand it can also be argued that 7 or 8 principal components is more appropriate as they contain 99.79% and 99.93% of the information respectively. It is however clear that there is a very high correaltion between the maturities.

Question 5

```
ggplot(analysis) + geom_segment(as.data.frame(analysis$rotation[c(1,2,3,51),1:2]),
↪  mapping=aes(x=0,y=0,xend=PC1,yend=PC2,colour=PC1),arrow=arrow(length =
↪  unit(0.5,"cm")), show.legend = FALSE) + xlab("PC1") + ylab("PC2")
```



Looking at the following bi-plot with the first and second principal component one can clearly see the correlation captured of the maturities. The first principal component contains the largest amount of information, which is also noted in the axis name, where 71.31% is the proportion of variance captured. This makes it quite clear that the data is positively correlated.

An interesting observation is however, that the V1 vector is much shorter than any other. Looking at the following graph it is easy to see that. In this graph we have the eigenvectors for V1, V2, V3, and V51 (overnight, half a year, one year, and 25 year maturity) plotted for principal components one and two. The lighter the colour the longer the maturity is, so light blue is 25 year maturity. One key take away from this graph is that the overnight maturity is neither contained in the first or second principal component, or at least not to a considerable extent.

It also appears that the half year maturity and the 25 year maturity is the farthest away (look also at previous graph), or have the largest angle between them, yet they are not orthogonal, which would indicate that they are uncorrelated.