STAT547 Homework 1

Xingche Guo

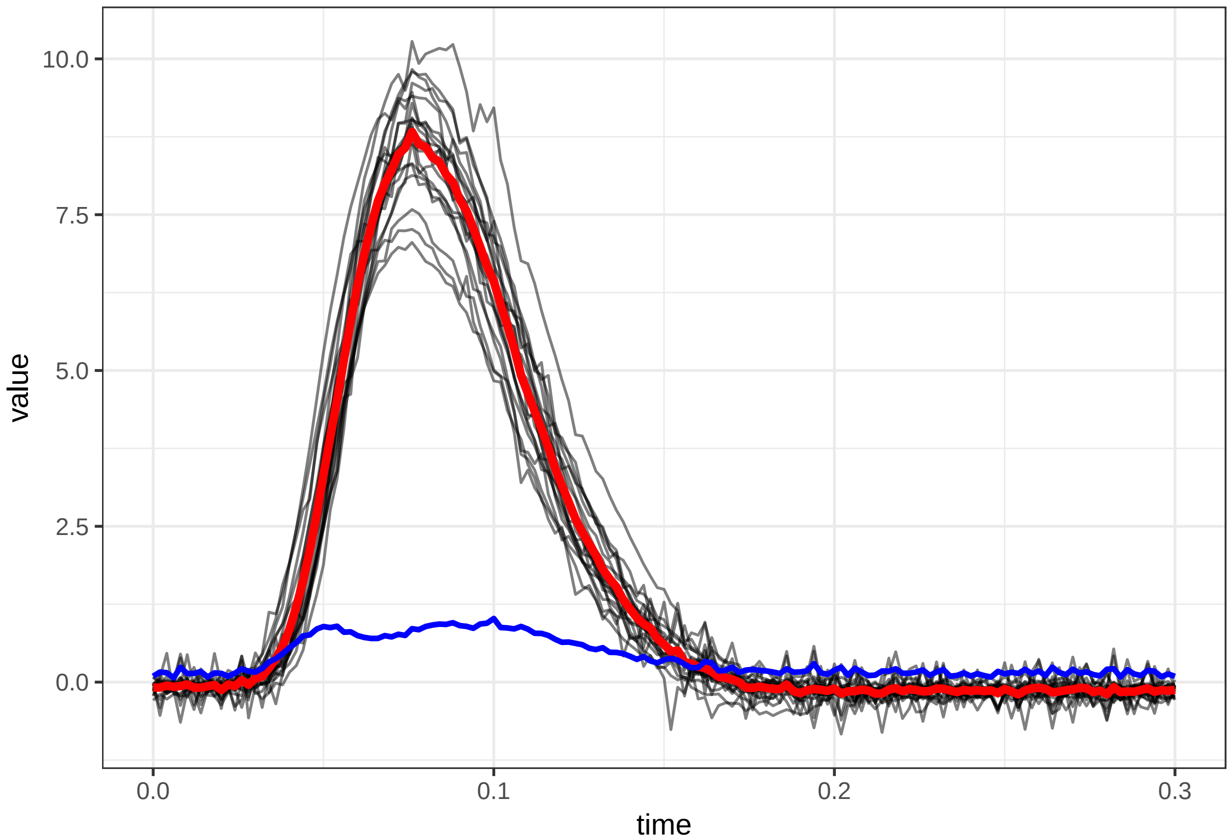
9/8/2019

Problem 1

```
data(pinch)
dim(pinch)
```

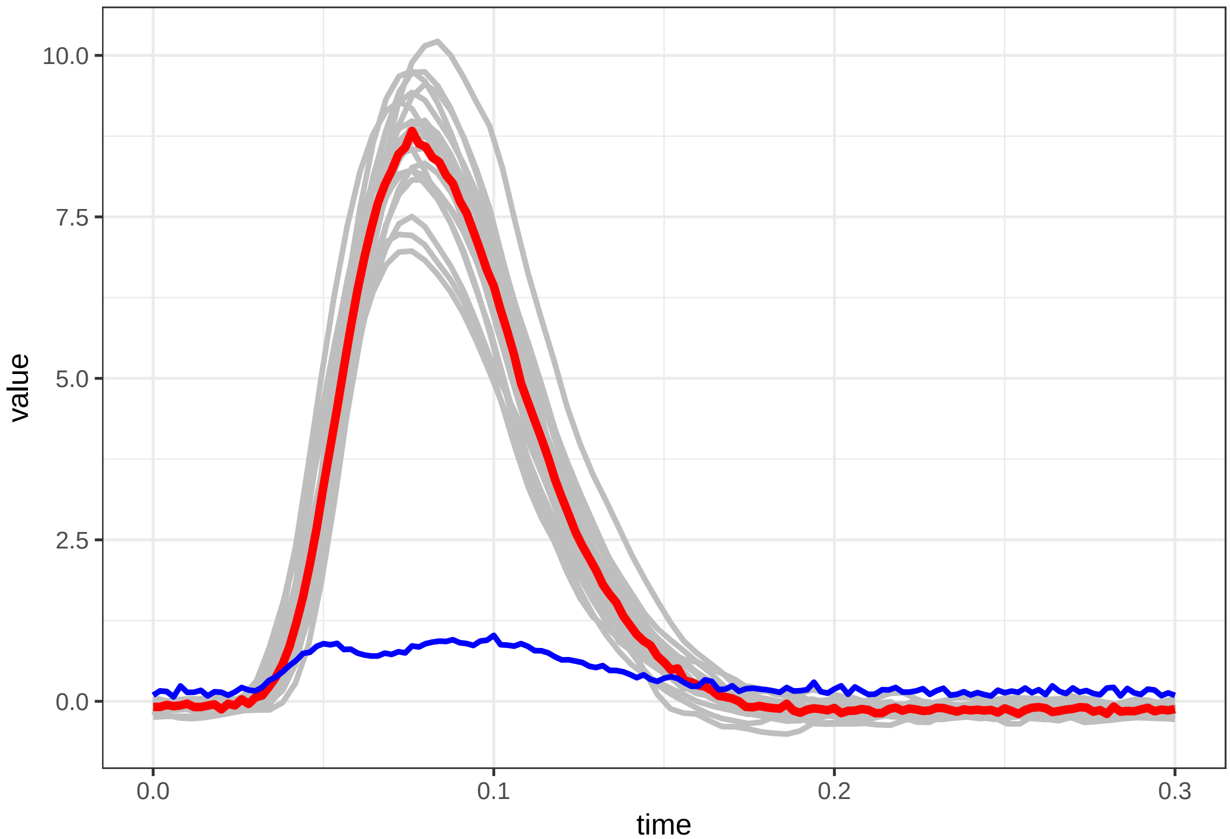
```
## [1] 151  20
```

```
##### (b) #####
pinch_mu <- apply(pinch, 1, mean)
pinch_sd <- apply(pinch, 1, sd)

r_pinch <- melt(pinch)
names(r_pinch) <- c("time", "rep", "value")
r_pinch$time <- rep(pinchtime, 20)
r_pinch$rep <- as.factor(r_pinch$rep)

## raw
ggplot(data = r_pinch) +
  geom_line(aes(x = time, y = value, group = rep),
            alpha = 0.5) +
  geom_line(data = data.frame(x = pinchtime, y = pinch_mu),
          aes(x=x, y=y), colour = "red", size = 1.5) +
  geom_line(data = data.frame(x = pinchtime, y = pinch_sd),
          aes(x=x, y=y), colour = "blue", size = 1) +
  theme_bw()
```

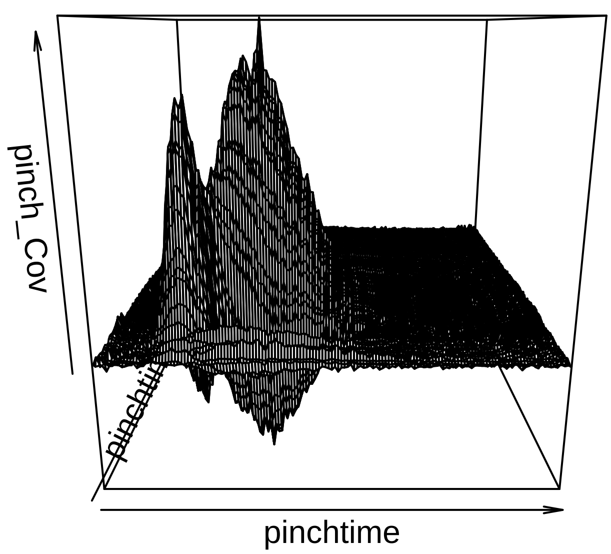



```
## smoothed
ggplot(data = r_pinch) +
  geom_smooth(aes(x = time, y = value, group = rep),
              alpha = 0.5, span = 0.1, se = FALSE, col = "gray") +
  geom_line(data = data.frame(x = pinchtime, y = pinch_mu),
            aes(x=x, y=y), colour = "red", size = 1.5) +
  geom_line(data = data.frame(x = pinchtime, y = pinch_sd),
            aes(x=x, y=y), colour = "blue", size = 1) +
  theme_bw()
```

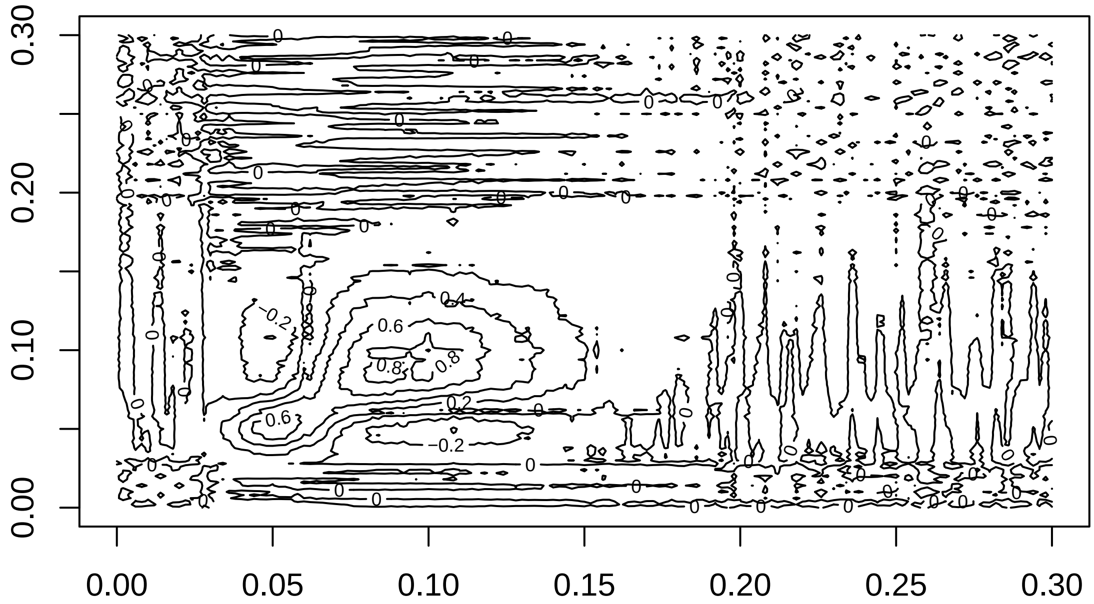
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
##### (c) #####
pinch_Cov <- cov(t(pinch))
persp(x=pinchtime, y=pinchtime, z=pinch_Cov)
```



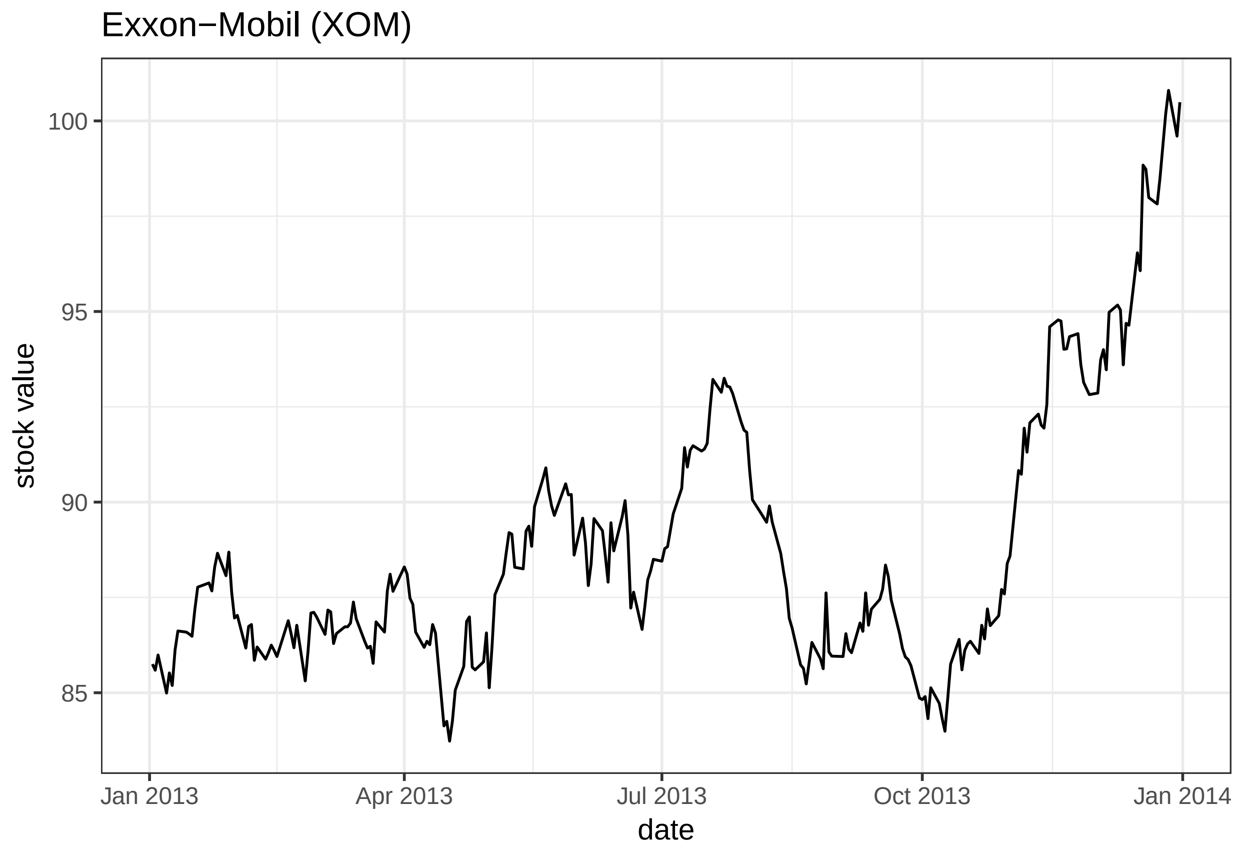

```
contour(x=pinchtime, y=pinchtime, z=pinch_Cov)
```



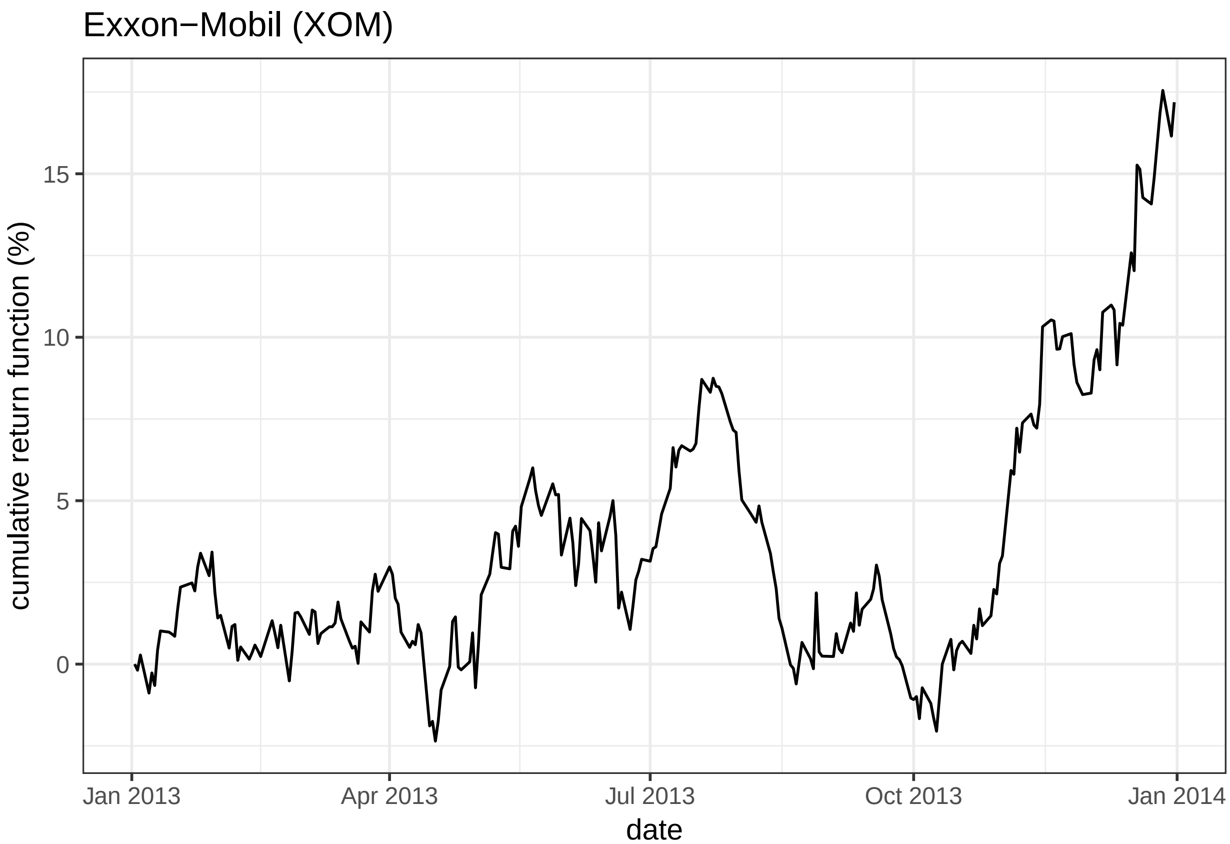
```r
#plot_ly(x=pinchtime, y=pinchtime, z=pinch_Cov) %>% add_surface(
#  contours = list(
#    z = list(
#      show=TRUE,
#      usecolormap=TRUE,
#      highlightcolor="#ff0000",
#      project=list(z=TRUE)
#    )
#  )
#)
```

Problem 2

```r
######################## Problem 2 #######################
D2 <- read.csv("/Users/apple/Desktop/ISU 2019 fall/STAT547/data/DataSets/Dow_companies_data.csv")

##### (a) #####
P1 <- D2$XOM[1]
date <- as.Date(as.character(D2$Date), format = "%m/%d/%y")

D_XOM <- data.frame(date = date, v1 = D2$XOM,
                    v2 = 100*((D2$XOM/P1)-1))

ggplot(data = D_XOM) +
  geom_line(aes(x = date, y = v1)) +
  ylab("stock value") +
  ggtitle("Exxon-Mobil (XOM)") + theme_bw()
```



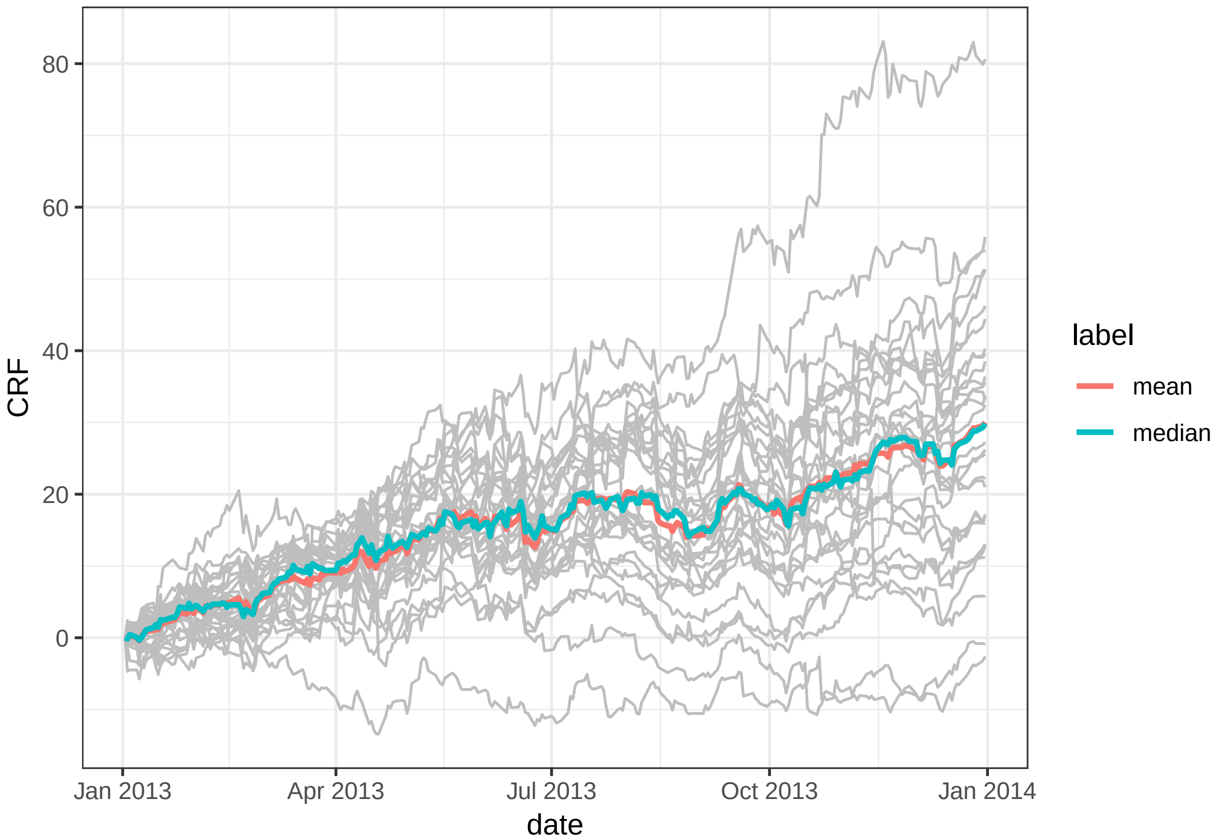

```
ggplot(data = D_XOM) +
  geom_line(aes(x = date, y = v2)) +
  ylab("cumulative return function (%)") +
  ggtitle("Exxon-Mobil (XOM)") + theme_bw()
```



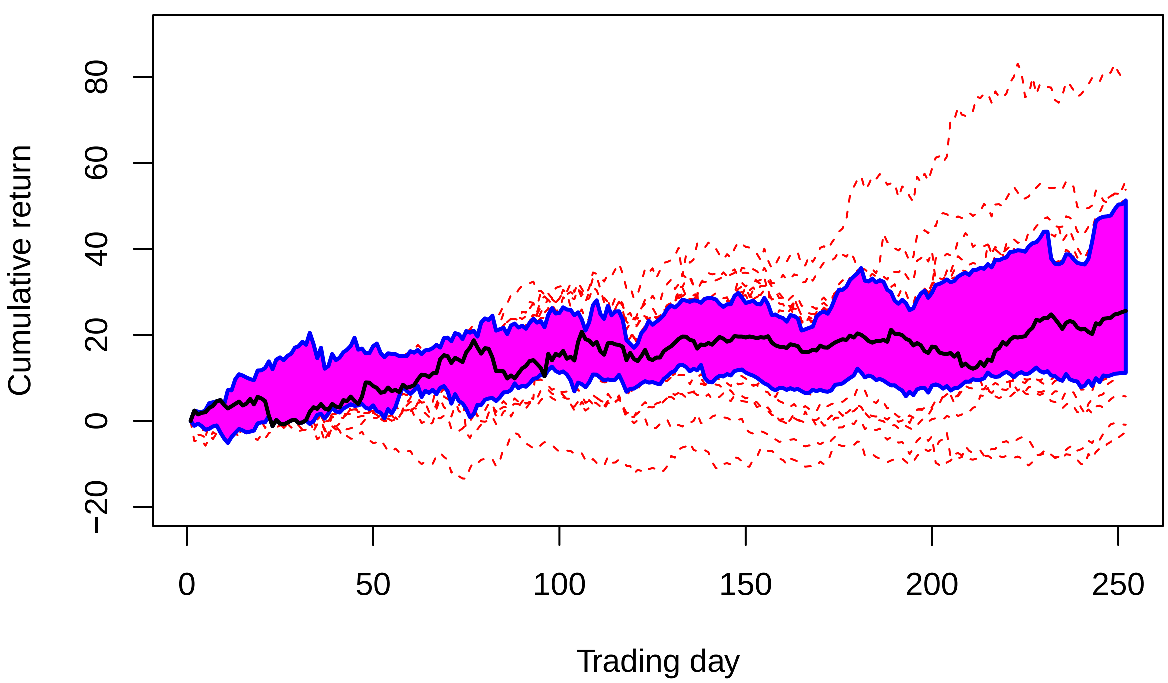

```
D_XOM$v2[length(D_XOM$v2)]
```

```
## [1] 17.1895
```

```
##### (b) #####
D2_1 <- D2[,-1]

D2_2 <- apply(D2_1, 2, FUN = function(x){
  return( 100*(x - x[1])/x[1] )
})

r_D2_2 <- melt(D2_2)
names(r_D2_2) <- c("date", "stock", "CRF")
r_D2_2$date <- rep(date, 30)

stock_mean <- apply(D2_2, 1, mean)
stock_med <- apply(D2_2, 1, median)

DD <- data.frame(date = rep(date,2),
                 CRF = c(stock_mean, stock_med),
                 label = rep(c("mean", "median"), each = 252))


ggplot(data = r_D2_2) +
  geom_line(aes(x = date, y = CRF, group = stock), colour = "gray") +
  geom_line(data = DD, aes(x = date, y = CRF, colour = label), size = 1) +
  theme_bw()
```



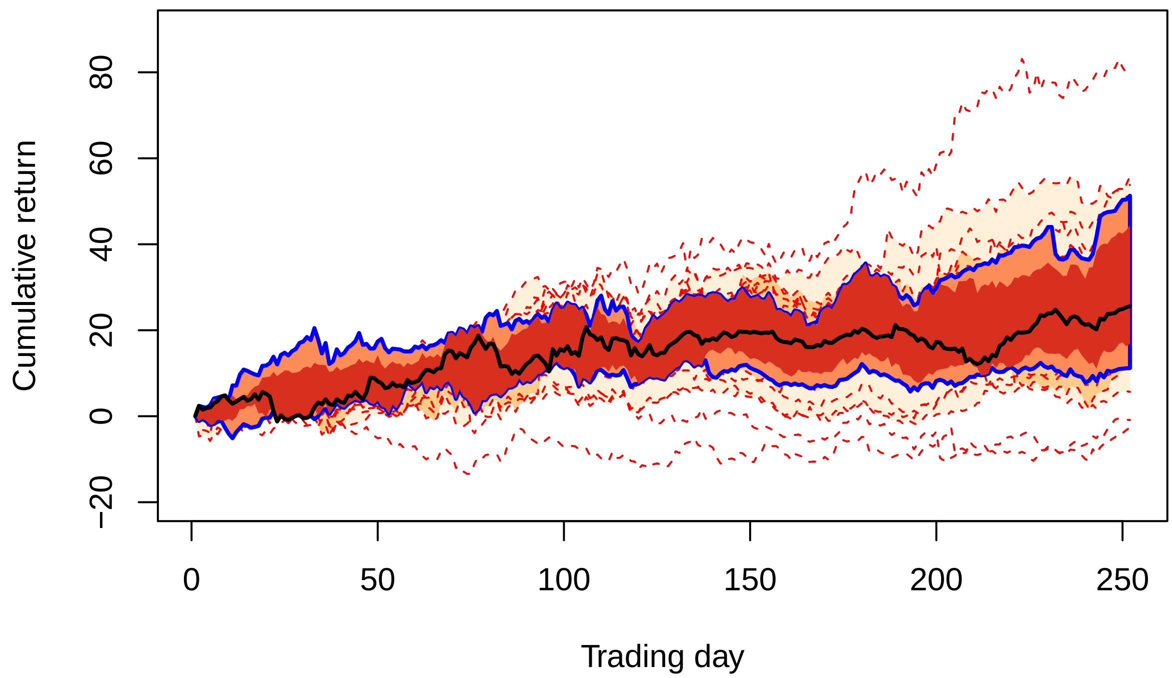

```
##### (c) #####
fbplot(fit = D2_2, ylim = c(-20,90), xlab="Trading day",
       ylab = "Cumulative return")
```




```
## $depth
##        MMM        AXP          T         BA        CAT        CVX       CSCO
```

```
## 0.4152915 0.3129926 0.2976852 0.1558498 0.1287926 0.4631477 0.4010787
##        KO        DD       XOM        GE        GS        HD      INTC
## 0.3526204 0.3906062 0.2521096 0.4612320 0.4046730 0.4735837 0.3979771
##       IBM       JNJ       JPM       MCD       MRK      MSFT       NKE
## 0.2310368 0.3479132 0.4568167 0.4265850 0.4872856 0.3478768 0.3127919
##       PFE        PG       TRV       UNH       UTX        VZ         V
## 0.4678366 0.4624726 0.4250160 0.3453772 0.4552842 0.4319308 0.4645525
##       WMT       DIS
## 0.4375319 0.3458881
##
## $outpoint
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
## [24] 24 25 26 27 28 29 30
##
## $medcurve
## MRK
##  19
```

```
fbplot(fit = D2_2, ylim = c(-20,90), prob = c(0.9, 0.6, 0.5, 0.3),
      color = brewer.pal(4, "OrRd"),
      xlab="Trading day",
      ylab = "Cumulative return")
```




```
## $depth
##       MMM       AXP         T        BA       CAT       CVX      CSCO
## 0.4152915 0.3129926 0.2976852 0.1558498 0.1287926 0.4631477 0.4010787
##        KO        DD       XOM        GE        GS        HD      INTC
## 0.3526204 0.3906062 0.2521096 0.4612320 0.4046730 0.4735837 0.3979771
##       IBM       JNJ       JPM       MCD       MRK      MSFT       NKE
## 0.2310368 0.3479132 0.4568167 0.4265850 0.4872856 0.3478768 0.3127919
##       PFE        PG       TRV       UNH       UTX        VZ         V
## 0.4678366 0.4624726 0.4250160 0.3453772 0.4552842 0.4319308 0.4645525
##       WMT       DIS
```

```
## 0.4375319 0.3458881
##
## $outpoint
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
## [24] 24 25 26 27 28 29 30
##
## $medcurve
## MRK
##  19
```

Problem 3

Suppose $\mathrm{Var}(\boldsymbol{x}) = \Sigma = P\Lambda P^T$, where Λ is the eigenvalue matrix such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, P is the corresponding eigenvector matrix, then $\mathrm{Var}(\boldsymbol{l}_1^T\boldsymbol{x}) = \boldsymbol{l}_1^T\Sigma\boldsymbol{l}_1 = \boldsymbol{y}_1^T\Lambda\boldsymbol{y}_1$, where $\boldsymbol{y}_1 = P^T\boldsymbol{l}_1$, and $\boldsymbol{y}_1^T\boldsymbol{y}_1 = \boldsymbol{l}_1^T\boldsymbol{l}_1 = 1$. Note that:

$$\boldsymbol{y}_1^T\Lambda\boldsymbol{y}_1 = \sum_{i=1}^{d}\lambda_i y_{1i}^2 \geq \lambda_1,$$

equality holds iff $\boldsymbol{y}_1 = \boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$. Therefore, $\boldsymbol{v}_1 = P\boldsymbol{e}_1 = \boldsymbol{p}_1$, where $\boldsymbol{p}_1$ is the eigenvector corresponding to λ_1.

Also note that $\boldsymbol{e}_1^T\boldsymbol{y}_2 = \boldsymbol{v}_1^T\boldsymbol{l}_2 = 0$, which means $y_{21} = 0$, therefore:

$$\boldsymbol{y}_2^T\Lambda\boldsymbol{y}_2 = \sum_{i=2}^{d}\lambda_i y_{2i}^2 = \lambda_2,$$

and $\boldsymbol{v}_2 = P\boldsymbol{e}_2 = \boldsymbol{p}_2$, where $\boldsymbol{p}_2$ is the eigenvector corresponding to λ_2.

Thus, $V := (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d) = P$.

Problem 4

(a)

$$\mathrm{E}(\xi_k) = \mathrm{E}(\boldsymbol{v}_k^T(\boldsymbol{x} - \boldsymbol{\mu})) = \boldsymbol{v}_k^T(\boldsymbol{\mu} - \boldsymbol{\mu}) = 0$$

(b)

$$\mathrm{Var}(\xi_k) = \boldsymbol{v}_k^T\Sigma\boldsymbol{v}_k = \boldsymbol{e}_k^T P^T P\Lambda P^T P\boldsymbol{e}_k = \lambda_k$$

(c)

$$\mathrm{Cov}(\xi_j, \xi_k) = \boldsymbol{v}_j^T\Sigma\boldsymbol{v}_k = \boldsymbol{e}_j^T\Lambda\boldsymbol{e}_k = \delta_{jk}\lambda_k$$

(d)

$$\mathrm{Corr}(X_j, \xi_k) = \frac{\mathrm{Cov}(X_j, \xi_k)}{\sigma_{jj}\sqrt{\lambda_k}} = \frac{\boldsymbol{e}_j^T\Sigma\boldsymbol{v}_k}{\sigma_{jj}\sqrt{\lambda_k}} = \frac{\boldsymbol{e}_j^T(\lambda_k\boldsymbol{v}_k)}{\sigma_{jj}\sqrt{\lambda_k}} = \frac{v_{jk}\sqrt{\lambda_k}}{\sigma_{jj}}$$

Problem 5

(a)

$$||ax|| =< ax, ax >^{1/2}= |a| < x, x >^{1/2}= |a| \ ||x||$$

(b)

$$\because < x + ay, x + ay > \geq 0, \quad \forall a,$$
$$\therefore < y, y > a^2 + 2 < x, y > a + < x, x > \geq 0, \quad \forall a,$$
$$\therefore 4 < x, y >^2 - 4 < x, x >< y, y > \geq 0,$$
$$\therefore | < x, y > | \leq ||x|| \ ||y||.$$

(c)

$$\because ||x + y||^2 =< x + y, x + y >=< x, x > + < y, y > +2 < x, y > \leq ||x||^2 + ||y||^2 + 2||x|| \ ||y|| = (||x|| + ||y||)^2,$$
$$\therefore ||x + y|| \leq ||x|| + ||y||.$$

(d)

$$d(x, y) = ||x - y|| =< x - y, x - y >^{1/2}=< y - x, y - x >^{1/2}= ||y - x|| = d(y, x) > 0, \quad iff \ \ x \neq y,$$
$$d(x, x) = ||x - x|| = 0$$
$$d(x, y) = ||x - y|| = ||(x - z) + (z - y)|| \leq ||x - z|| + ||z - y|| = d(x, z) + d(z, y).$$

Problem 6

It's easy to find that l_2 space is a valid inner product space, we only need to verify that this space is complete. $\forall$ Cauchy sequence $\{x^{(n)}\}_{n=1}^{\infty}$, we have:

$$lim_{n,m\to\infty}||x^{(n)} - x^{(m)}|| = 0,$$

therefore, for large n_k, $\exists n_{k+1} > n_k$, such that:

$$||x^{(n_{k+1})} - x^{(n_k)}|| \leq \frac{1}{k^2}$$

$$\therefore \sum_{k=1}^{\infty} ||x^{(n_{k+1})} - x^{(n_k)}|| < \infty$$

Define $x = \liminf_{k\to\infty} x^{(n_k)}$, we only need to prove (i). $||x||^2 < \infty$, and (ii). $\lim_{n\to\infty} ||x - x^{(n)}|| = 0$. First note that:

$$||x||^2 = \sum_{i=1}^{\infty} |x_i|^2 = \sum_{i=1}^{\infty} \liminf_k |x_i^{(n_k)}|^2 \leq \liminf_k \sum_{i=1}^{\infty} |x_i^{(n_k)}|^2 = \liminf_k ||x^{(n_k)}||^2,$$

The inequality comes from Fatou's lemma, also:

$$\liminf_K ||x^{(n_K)}|| \leq ||x^{(n_1)}|| + \liminf_K \sum_{k=1}^{K-1} ||x^{(n_{k+1})} - x^{(n_k)}|| < \infty,$$

therefore $||x||^2 < \infty$.

Again, by Fatou's lemma:

$$\lim_{n\to\infty} ||x - x^{(n)}||^2 = \lim_{n\to\infty} \sum_{i=1}^{\infty} |x_i - x_i^{(n)}|^2 = \lim_{n\to\infty} \sum_{i=1}^{\infty} \liminf_k |x_i^{(n_k)} - x_i^{(n)}|^2 \leq \lim_{n,k\to\infty} \sum_{i=1}^{\infty} |x_i^{(n_k)} - x_i^{(n)}|^2$$

By the property of Cauchy sequence:

$$\lim_{n\to\infty} ||x - x^{(n)}||^2 \leq \lim_{n,k\to\infty} ||x^{(n_k)} - x^{(n)}||^2 = 0,$$

hence, $\lim_{n\to\infty} ||x - x^{(n)}|| = 0$. Thus, l_2 space to that end is an Hilbert Space.

Problem 7

$$\begin{aligned} \mathrm{E}||X - \sum_{k=1}^{K} < X, e_k > e_k||^2 &= \mathrm{E} \int X^2(t)dt - 2\sum_{k=1}^{K} \mathrm{E} < X, e_k >^2 + \sum_{k=1}^{K}\sum_{j=1}^{K} \mathrm{E} < X, e_j >< X, e_k > \delta_{jk} \\ &= \int \mathrm{E}X^2(t)dt - \sum_{k=1}^{K} \mathrm{E} < X, e_k >^2 \\ &= \int G(t,t)dt - \sum_{k=1}^{K} \int\int e_k(s)e_k(t)G(s,t)dsdt. \end{aligned}$$

Thus, minimize $\mathrm{E}||X - \sum_{k=1}^{K} < X, e_k > e_k||^2$ is equivalent to maximize $\sum_{k=1}^{K} \int\int e_k(s)e_k(t)G(s,t)dsdt$. By the theorem, for K orthonormal basis $\{e_k\}_{k=1}^{K}$, we have:

$$\max_{\{e_k\}_{k=1}^{K}} \sum_{k=1}^{K} \int\int e_k(s)e_k(t)G(s,t)dsdt = \sum_{k=1}^{K} \lambda_k,$$

and

$$\arg\max_{\{e_k\}_{k=1}^{K}} \sum_{k=1}^{K} \int\int e_k(s)e_k(t)G(s,t)dsdt = \{\phi_k\}_{k=1}^{K},$$

where (λ_k, ϕ_k) is the kth eigenvalue-eigenfunction pairs for NND function $G(s,t)$.