Fidelity_Project

Weiling Li

10/17/2019

Sector Fund

For this project sector fund Fidelity Select Technology Portfolio (FSPTX) is chosen as the target fund.

For comparison, `Russel2000(^RUT)`, `NASDAQ(^IXIC)`, `S&P500(^GSPC)`, `S&PMidCap(^MID)` and `S&PSmlCap(^SML)` was selected as initial indexes to be campared with.

for comparison, Vanguard's similar index funds are also loaded as `IT ETF(VGT)`,`LargeCap ETF(VIGAX)` and `TotalMarket ETF(VTSAX)`

Before loading data, we will define some useful function to ease the data cleaning process, and claim some variables 1st.

```
## this function requires a dataframe input that has the daily close price named:Closed and a date colu
## We trimmed the data from 1990-10-07 because 1990-10-08 is a Monday and Stock market closed during we
dailynlogReturn <- function(Date1,DataFrame){
  DataFrame = mutate(DataFrame, dailyReturn = (Close-lag(Close))/Close)%>%mutate(perc_dailyRe = round(da
}
## This function returns a projection value of the fund from the start date and assuming 10k investment
getProjectionValue <- function(DF){
  P0 = pull(filter(DF, Date == pull(top_n(DF["Date"],-1)))%>%select(Close))
  DF = mutate(DF,ProjValper10k = (Close*10000)/P0)
}
## Calculate Euclidean distances between two sets of data
sqerr <- function(x,y){
  z = x - y
  z = sqrt(dot(z,z)/length(y))
  return(z)
}

## function for standarize NAV
standardizedNAV = function(DF){
  return(mutate(DF,Close.z = (Close-mean(Close))/sd(Close)))
}

## restrict ourselves to study data after 2014-01-01
StartDate = as.Date("2014-01-01")
```

Load and clean the data, then some EDA

Load the data:

```
FSPTX = dailynlogReturn(StartDate,read_csv("FSPTX.csv"))
NASDAQ = dailynlogReturn(StartDate,read_csv("^IXIC.csv"))
SnP500 = dailynlogReturn(StartDate,read_csv("^GSPC.csv"))
SnPMID = dailynlogReturn(StartDate,read_csv("^MID.csv"))
SnPSML = dailynlogReturn(StartDate,read_csv("^SML.csv"))
RUSSELL2000 = dailynlogReturn(StartDate,read_csv("^RUT.csv"))
VGT = dailynlogReturn(StartDate,read_csv("VGT.csv"))
```

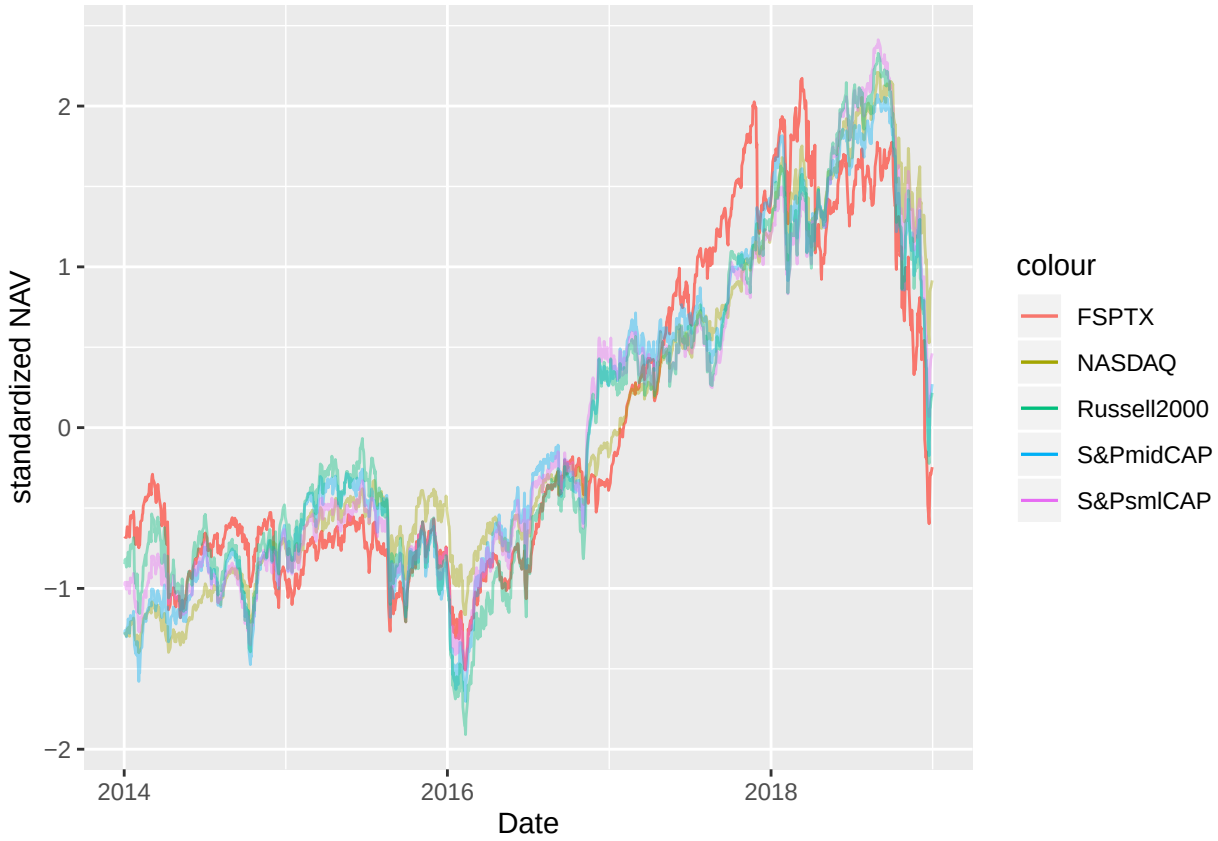
```
VIGAX = dailynlogReturn(StartDate,read_csv("VIGAX.csv"))
VTSAX = dailynlogReturn(StartDate,read_csv("VTSAX.csv"))
```

Standardize NAV

```
FSPTX = standardizedNAV(FSPTX)
NASDAQ = standardizedNAV(NASDAQ)
SnP500 = standardizedNAV(SnP500)
SnPMID = standardizedNAV(SnPMID)
SnPSML = standardizedNAV(SnPSML)
RUSSELL2000 = standardizedNAV(RUSSELL2000)
VGT = standardizedNAV(VGT)
VIGAX = standardizedNAV(VIGAX)
VTSAX = standardizedNAV(VTSAX)
```

Plot the NAVs:

```
ggplot(FSPTX)+geom_line(mapping = aes(x = Date,y = Close.z,color = "FSPTX"))+geom_line(mapping = aes(x =
```



For comparison, plot Vanguard's fund too

```
ggplot(VGT)+geom_line(mapping = aes(x = Date,y = Close.z,color = "VGT"))+geom_line(mapping = aes(x = NA
```


```
ggplot(VIGAX)+geom_line(mapping = aes(x = Date,y = Close.z,color = "VIGAX"))+geom_line(mapping = aes(x =
```


```
ggplot(VTSAX)+geom_line(mapping = aes(x = Date,y = Close.z,color = "VTSAX"))+geom_line(mapping = aes(x =
```

It seems that the Vanguard's ETF's are more correlated to the indexes. Another way to look at this is the correlation matrix:

```
cor(FSPTX$Close.z,NASDAQ$Close.z)
```

```
## [1] 0.9332307
```

```
cor(FSPTX$Close.z,RUSSELL2000$Close.z)
```

```
## [1] 0.9390869
```

```
cor(FSPTX$Close.z,SnPMID$Close.z)
```

```
## [1] 0.9410625
```

```
cor(FSPTX$Close.z,SnPSML$Close.z)
```

```
## [1] 0.9290002
```

```
cordat = cbind(FSPTX$Close.z,NASDAQ$Close.z,SnP500$Close.z,RUSSELL2000$Close.z,SnPMID$Close.z,SnPSML$Cl
colnames(cordat) = c("FSPTX","NASDAQ","SnP500","RUSSELL2000","SnPMID","SnPSML","VGT(IT ETF)","VIGAX(Lar
cordat = data.frame(cordat)
cor(cordat,cordat)
```

```
##                       FSPTX    NASDAQ    SnP500 RUSSELL2000    SnPMID
## FSPTX             1.0000000 0.9332307 0.9352564   0.9390869 0.9410625
## NASDAQ            0.9332307 1.0000000 0.9939862   0.9643265 0.9775877
## SnP500            0.9352564 0.9939862 1.0000000   0.9657109 0.9863761
## RUSSELL2000       0.9390869 0.9643265 0.9657109   1.0000000 0.9854988
## SnPMID            0.9410625 0.9775877 0.9863761   0.9854988 1.0000000
## SnPSML            0.9290002 0.9795196 0.9805100   0.9898554 0.9908919
```

```
## VGT.IT.ETF.         0.9284701 0.9964530 0.9911940   0.9564776 0.9684751
## VIGAX.LargeCAP.     0.9295225 0.9979172 0.9951918   0.9612092 0.9762120
## VTSAX.TotalMarket. 0.9389925 0.9937535 0.9990944   0.9749236 0.9900590
##                        SnPSML VGT.IT.ETF. VIGAX.LargeCAP.
## FSPTX               0.9290002   0.9284701       0.9295225
## NASDAQ              0.9795196   0.9964530       0.9979172
## SnP500              0.9805100   0.9911940       0.9951918
## RUSSELL2000         0.9898554   0.9564776       0.9612092
## SnPMID              0.9908919   0.9684751       0.9762120
## SnPSML              1.0000000   0.9753852       0.9751635
## VGT.IT.ETF.         0.9753852   1.0000000       0.9949407
## VIGAX.LargeCAP.     0.9751635   0.9949407       1.0000000
## VTSAX.TotalMarket. 0.9851894   0.9902647       0.9948887
##                     VTSAX.TotalMarket.
## FSPTX                        0.9389925
## NASDAQ                       0.9937535
## SnP500                       0.9990944
## RUSSELL2000                  0.9749236
## SnPMID                       0.9900590
## SnPSML                       0.9851894
## VGT.IT.ETF.                  0.9902647
## VIGAX.LargeCAP.              0.9948887
## VTSAX.TotalMarket.           1.0000000
```

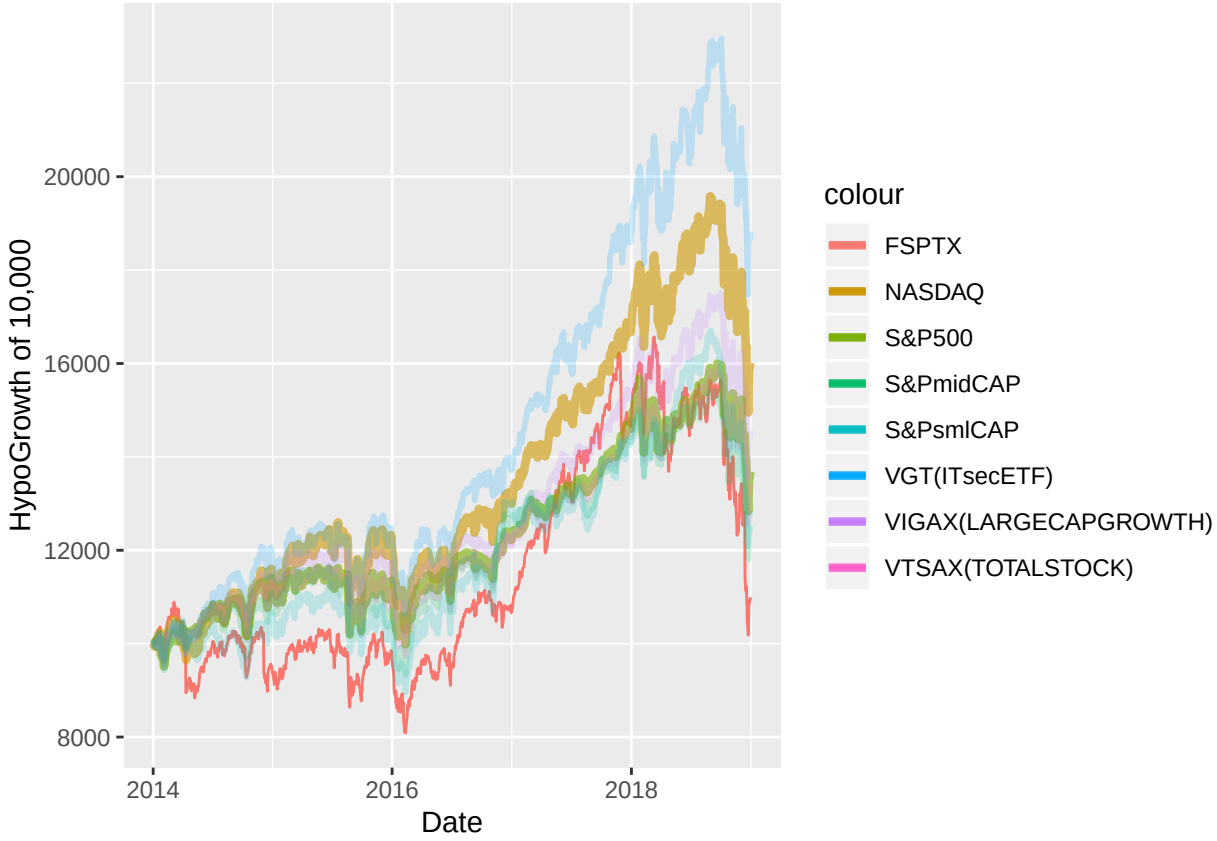
It seems that the Fidelity fund is not really matching the indexes we chose.

Check the dailyreturn and project value

```
FSPTX = getProjectionValue(FSPTX)
NASDAQ = getProjectionValue(NASDAQ)
SnP500 = getProjectionValue(SnP500)
VGT = getProjectionValue(VGT)
VIGAX = getProjectionValue(VIGAX)
VTSAX = getProjectionValue(VTSAX)
SnPMID = getProjectionValue(SnPMID)
SnPSML = getProjectionValue(SnPSML)
RUSSELL2000 = getProjectionValue(RUSSELL2000)
```

Plot them:

```
ggplot()+geom_line(mapping = aes(x = FSPTX$Date,y = FSPTX$ProjValper10k,color = 'FSPTX'))+geom_line(mapp
```



Still not a good fit...

Below is just some misc works: Mainly exploring the daily return

```
## Check distance between lines
print("pointwise variance between taget fund and NASDAQ")
```

```
## [1] "pointwise variance between taget fund and NASDAQ"
```

```
print(sqerr(FSPTX$ProjValper10k,NASDAQ$ProjValper10k))
```

```
## [1] 2189.376
```

```
#ks.test()
print("pointwise variance between taget fund and S&P500")
```

```
## [1] "pointwise variance between taget fund and S&P500"
```

```
print(sqerr(FSPTX$ProjValper10k,SnP500$ProjValper10k))
```

```
## [1] 1198.561
```

```
print("pointwise variance between taget fund and IT sector ETF")
```

```
## [1] "pointwise variance between taget fund and IT sector ETF"
```

```
print(sqerr(FSPTX$ProjValper10k,VGT$ProjValper10k))
```

```
## [1] 3437.128
```

```
print("pointwise variance between taget fund and Large cap growth index fund")
```

```
## [1] "pointwise variance between taget fund and Large cap growth index fund"
```

```
print(sqerr(FSPTX$ProjValper10k,VIGAX$ProjValper10k))
```

```
## [1] 1489.18
```

```
print("pointwise variance between taget fund and Total Stock market index fund")
```

```
## [1] "pointwise variance between taget fund and Total Stock market index fund"
```

```
print(sqerr(FSPTX$ProjValper10k,VTSAX$ProjValper10k))
```

```
## [1] 1139.656
```

```
#ggplot(FSPTX)+aes(x = Date , y=perc_dailyRe) + geom_line()

## Compare daily returns
dailyReturnComp = cbind(as.Date(FSPTX$Date),FSPTX$dailyReturn,NASDAQ$dailyReturn,SnP500$dailyReturn,VGT
colnames(dailyReturnComp) = c("Date","FSPTX","NASDAQ","SnP500","VGT","VIGAX","VTSAX")
epsilon = 0.000000000000000001
dailyReturnComp = data.frame(dailyReturnComp)%>%mutate(Date = as_date(Date),vsNASDAQ = ifelse(NASDAQ*NA

print("Average ratio of dailyReturn(FSPTX/NASDAQ")
```

```
## [1] "Average ratio of dailyReturn(FSPTX/NASDAQ"
```

```
print(sum(sqrt(dailyReturnComp$vsNASDAQ*dailyReturnComp$vsNASDAQ)/length(dailyReturnComp$vsNASDAQ)))
```

```
## [1] 2.707788
```

```
print("Average ratio of dailyReturn(FSPTX/S&P500")
```

```
## [1] "Average ratio of dailyReturn(FSPTX/S&P500"
```

```
print(sum(sqrt(dailyReturnComp$vsSnP500*dailyReturnComp$vsSnP500)/length(dailyReturnComp$vsSnP500)))
```

```
## [1] 5.321327
```

```
print("Average ratio of dailyReturn(FSPTX/VGT")
```

```
## [1] "Average ratio of dailyReturn(FSPTX/VGT"
```

```
print(sum(sqrt(dailyReturnComp$vsVGT*dailyReturnComp$vsVGT)/length(dailyReturnComp$vsVGT)))
```

```
## [1] 2.014087
```

```
print("Average ratio of dailyReturn(FSPTX/VIGAX")
```

```
## [1] "Average ratio of dailyReturn(FSPTX/VIGAX"
```

```
print(sum(sqrt(dailyReturnComp$vsVIGAX*dailyReturnComp$vsVIGAX)/length(dailyReturnComp$vsVIGAX)))
```

```
## [1] 2.495098
```

```
print("Average ratio of dailyReturn(FSPTX/NASDAQ")
```

```
## [1] "Average ratio of dailyReturn(FSPTX/NASDAQ"
```

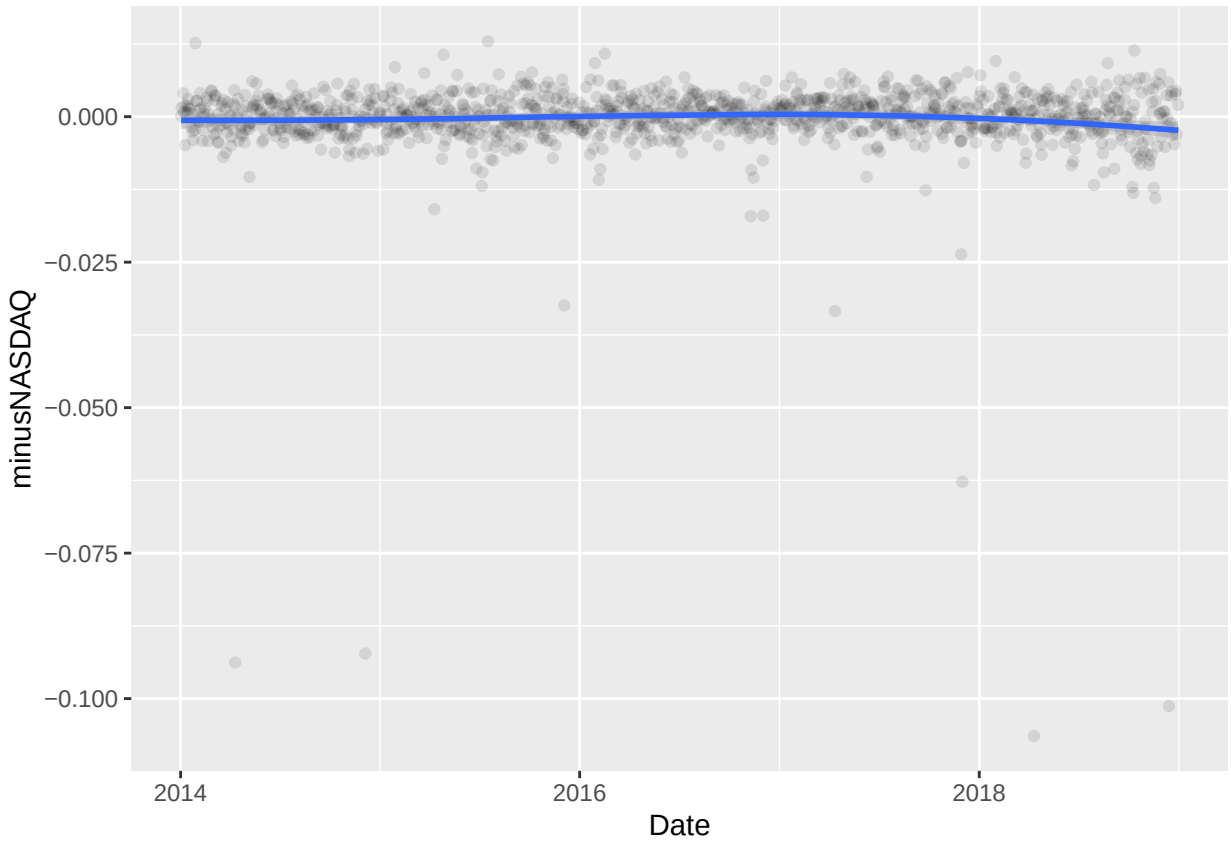
```
print(sum(sqrt(dailyReturnComp$vsVTSAX*dailyReturnComp$vsVTSAX)/length(dailyReturnComp$vsVTSAX)))
```

```
## [1] 3.13541
```

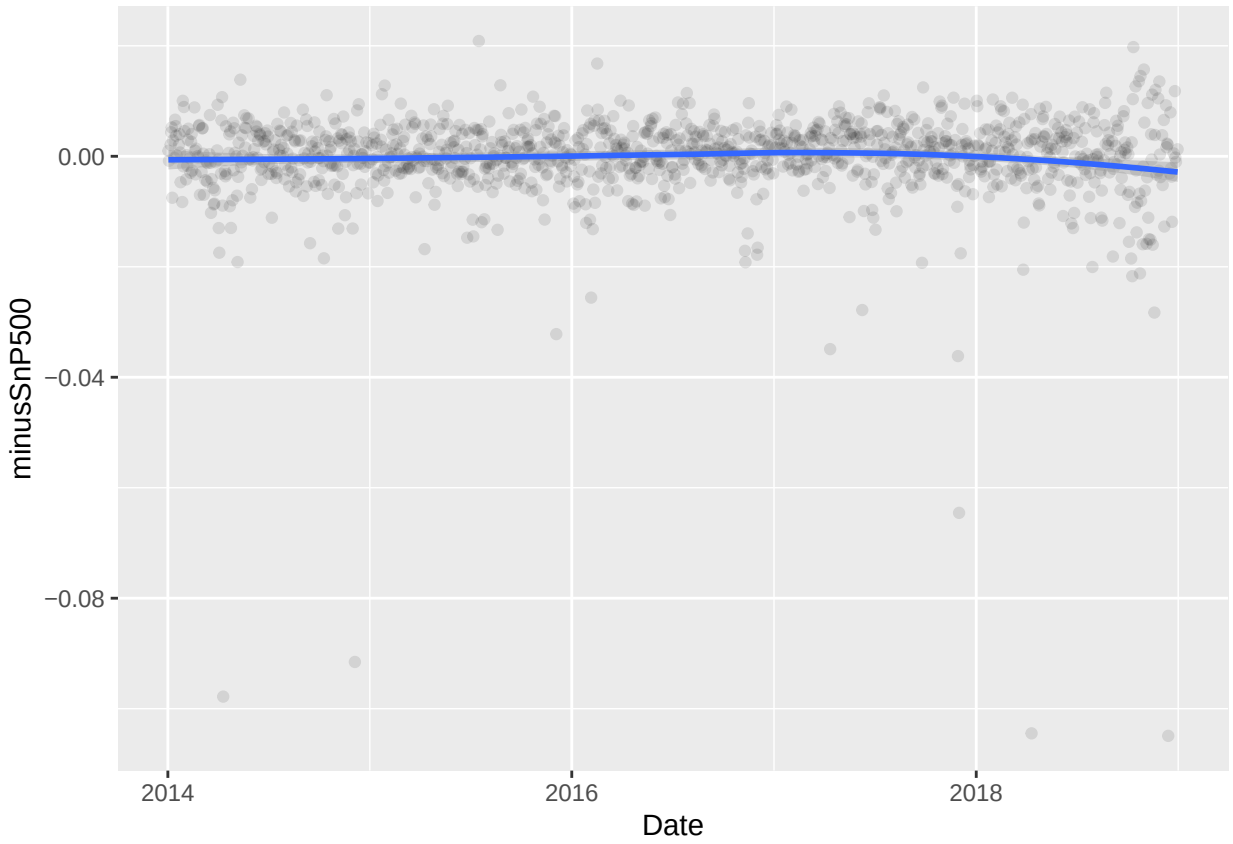
```
dailyReturnSTD = select(dailyReturnComp,contains("minus"))%>%summarise_all(funs(sd))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

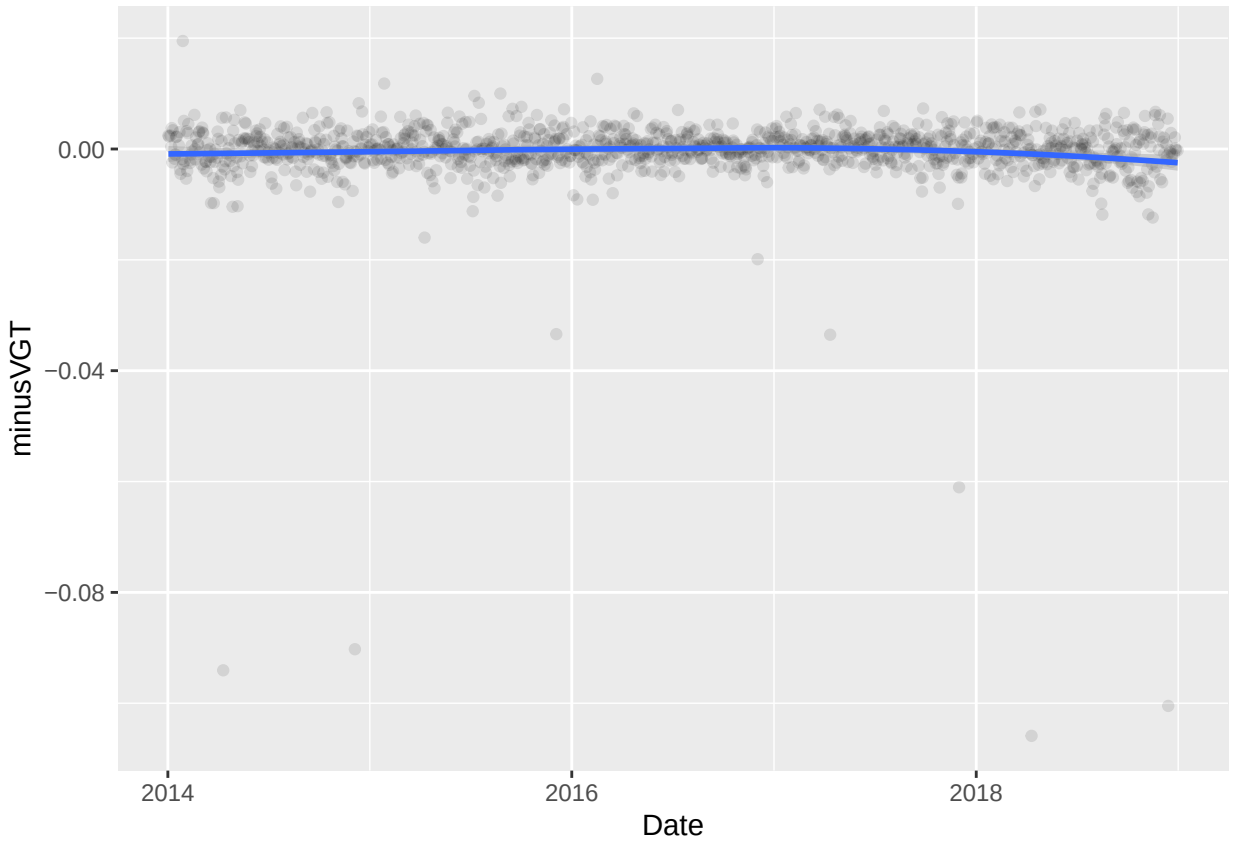
```
ggplot(dailyReturnComp)+aes(x = Date,y = minusNASDAQ)+geom_point(alpha = .1)+geom_smooth(method = "loess
```



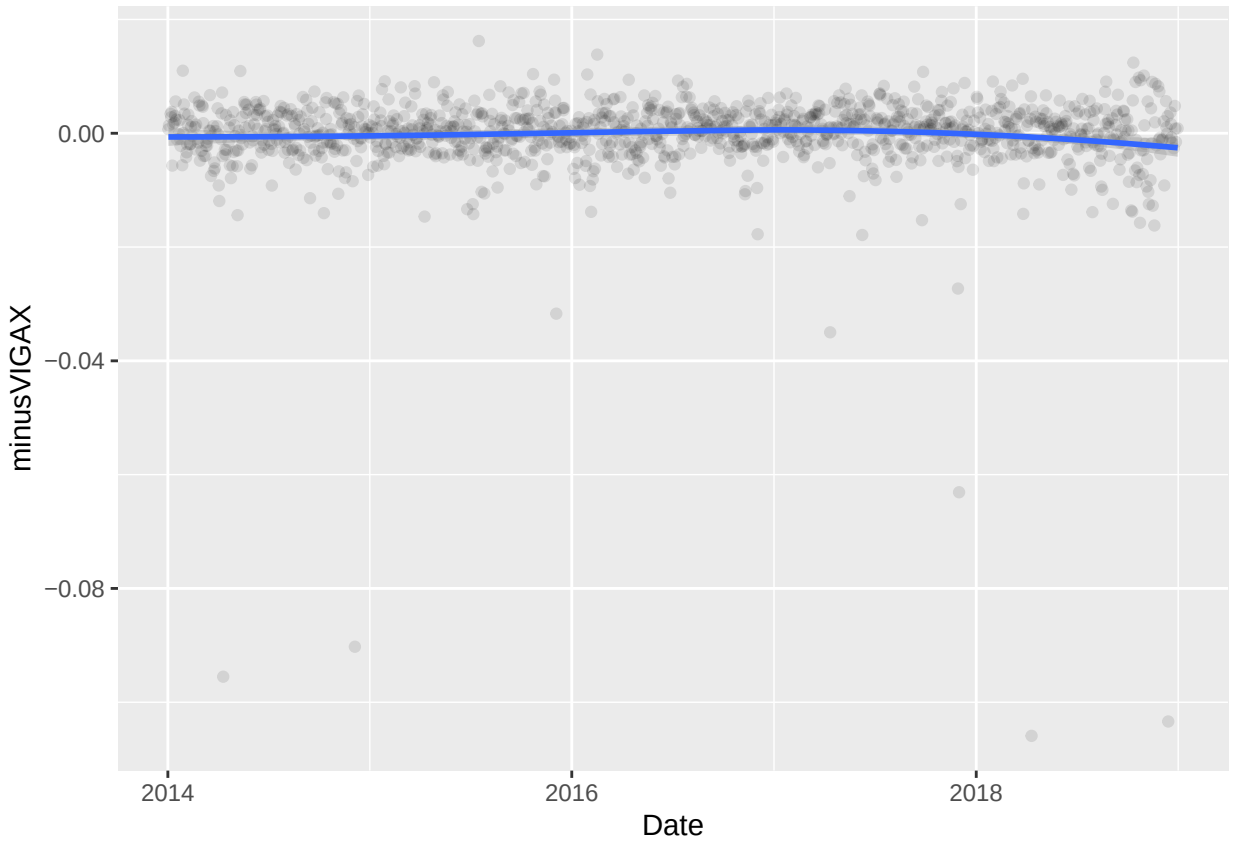

```
ggplot(dailyReturnComp)+aes(x = Date,y = minusSnP500)+geom_point(alpha = .1)+geom_smooth(method = "loess
```



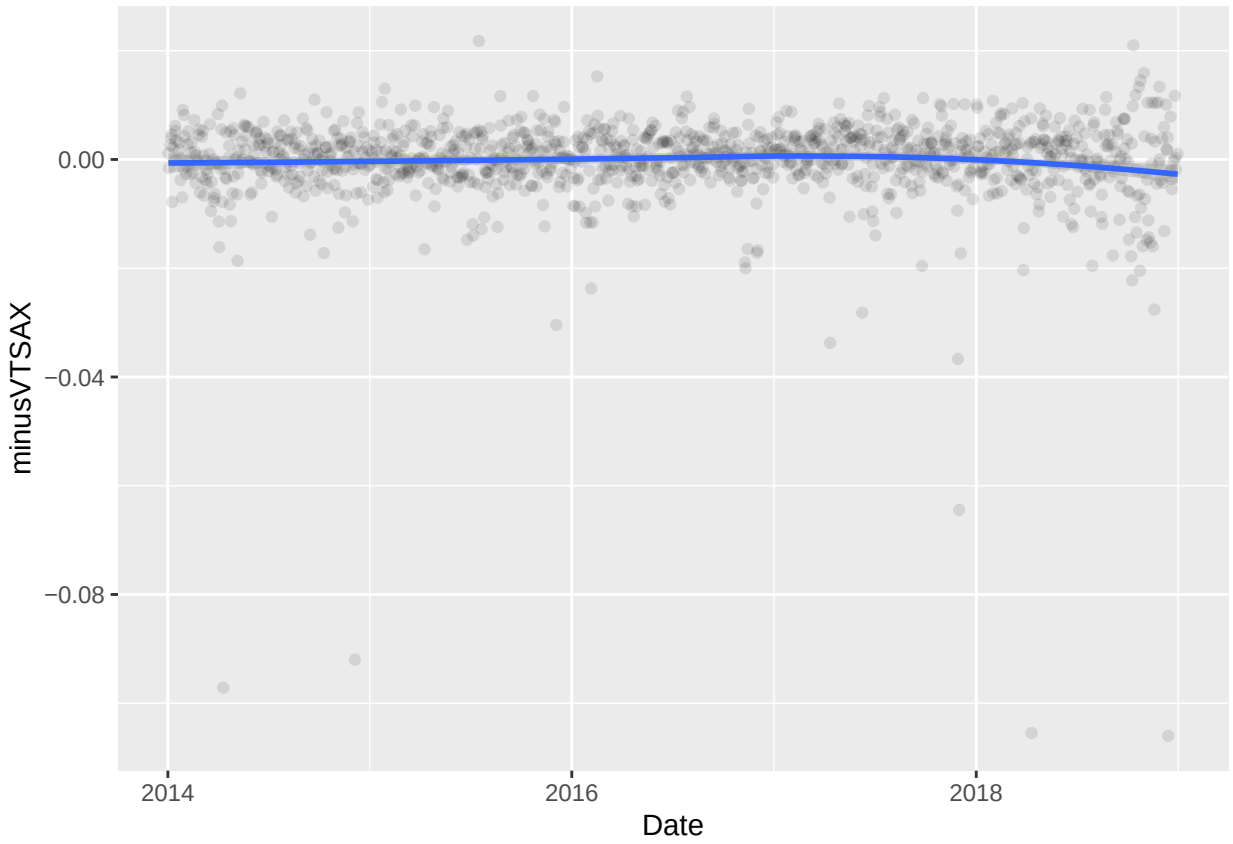

```
ggplot(dailyReturnComp)+aes(x = Date,y = minusVGT)+geom_point(alpha = .1)+geom_smooth(method = "loess",s
```



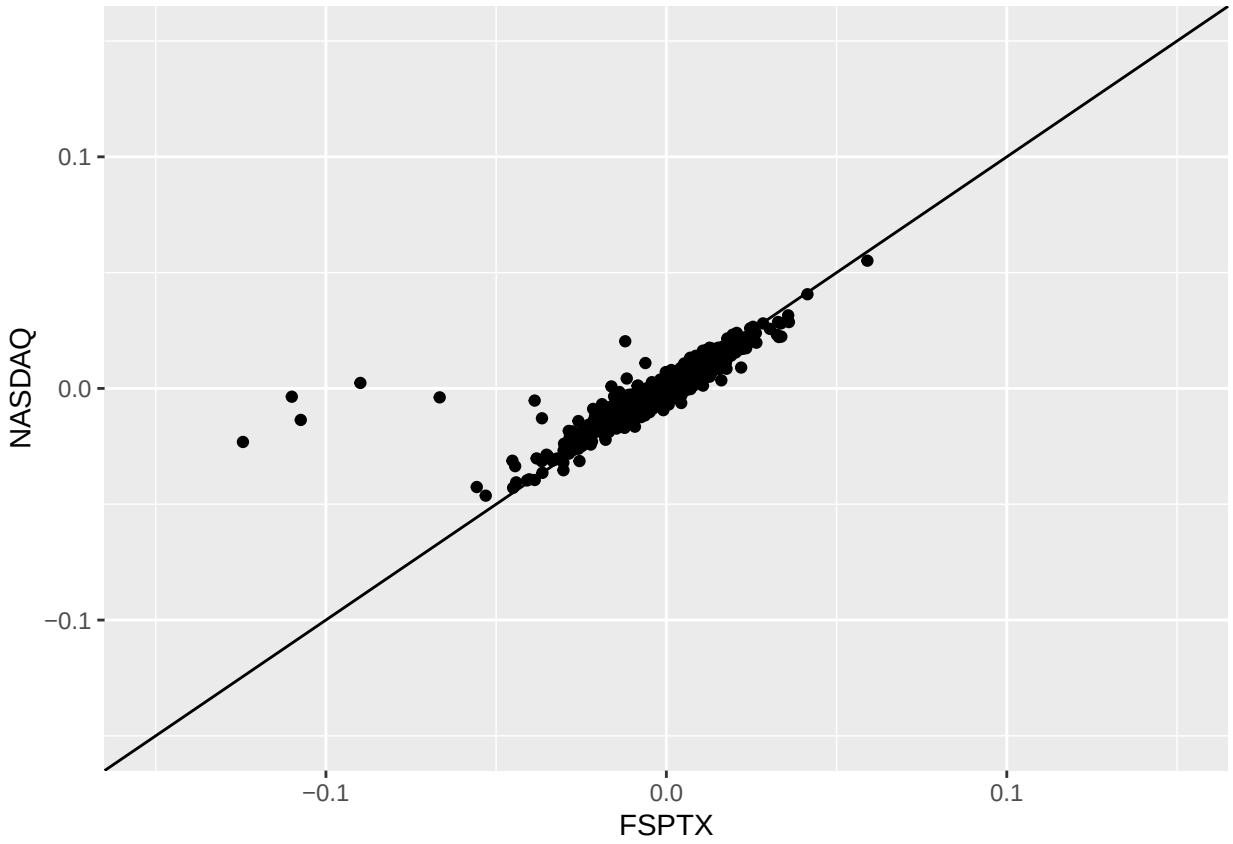

```
ggplot(dailyReturnComp)+aes(x = Date,y = minusVIGAX)+geom_point(alpha = .1)+geom_smooth(method = "loess
```



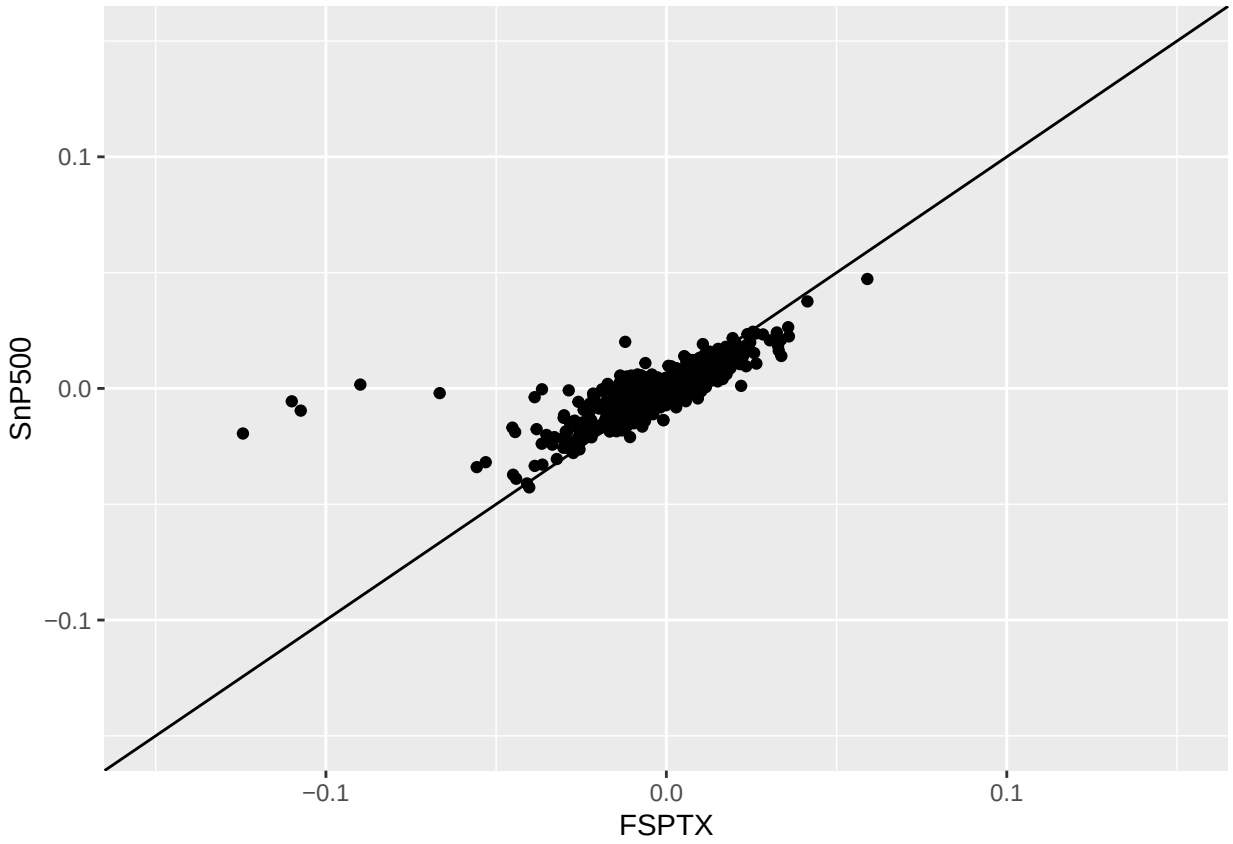

```
ggplot(dailyReturnComp)+aes(x = Date,y = minusVTSAX)+geom_point(alpha = .1)+geom_smooth(method = "loess
```



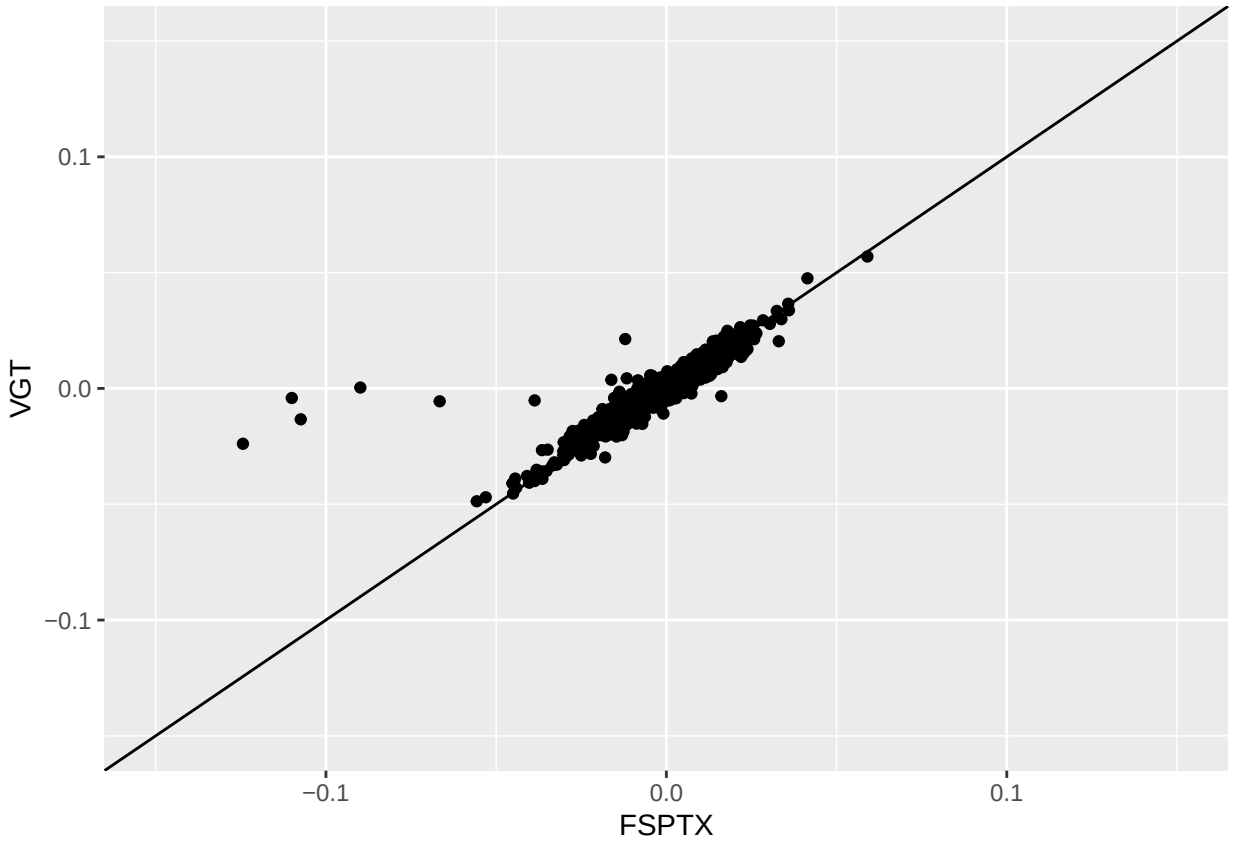

```
ggplot(dailyReturnComp)+aes(x = FSPTX,y = NASDAQ)+geom_point()+xlim(c(-0.15,0.15))+ylim(c(-0.15,0.15))+g
```



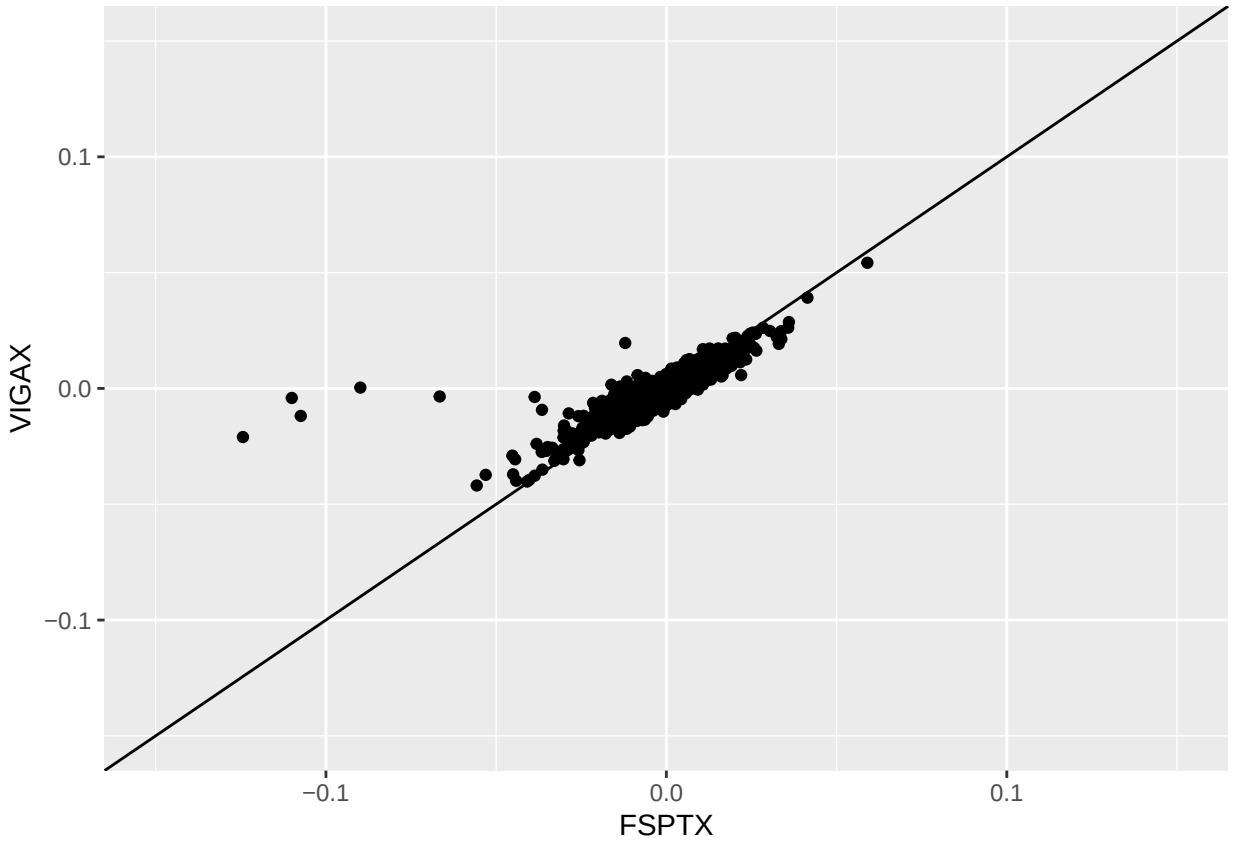

```
ggplot(dailyReturnComp)+aes(x = FSPTX,y = SnP500)+geom_point()+xlim(c(-0.15,0.15))+ylim(c(-0.15,0.15))+g
```



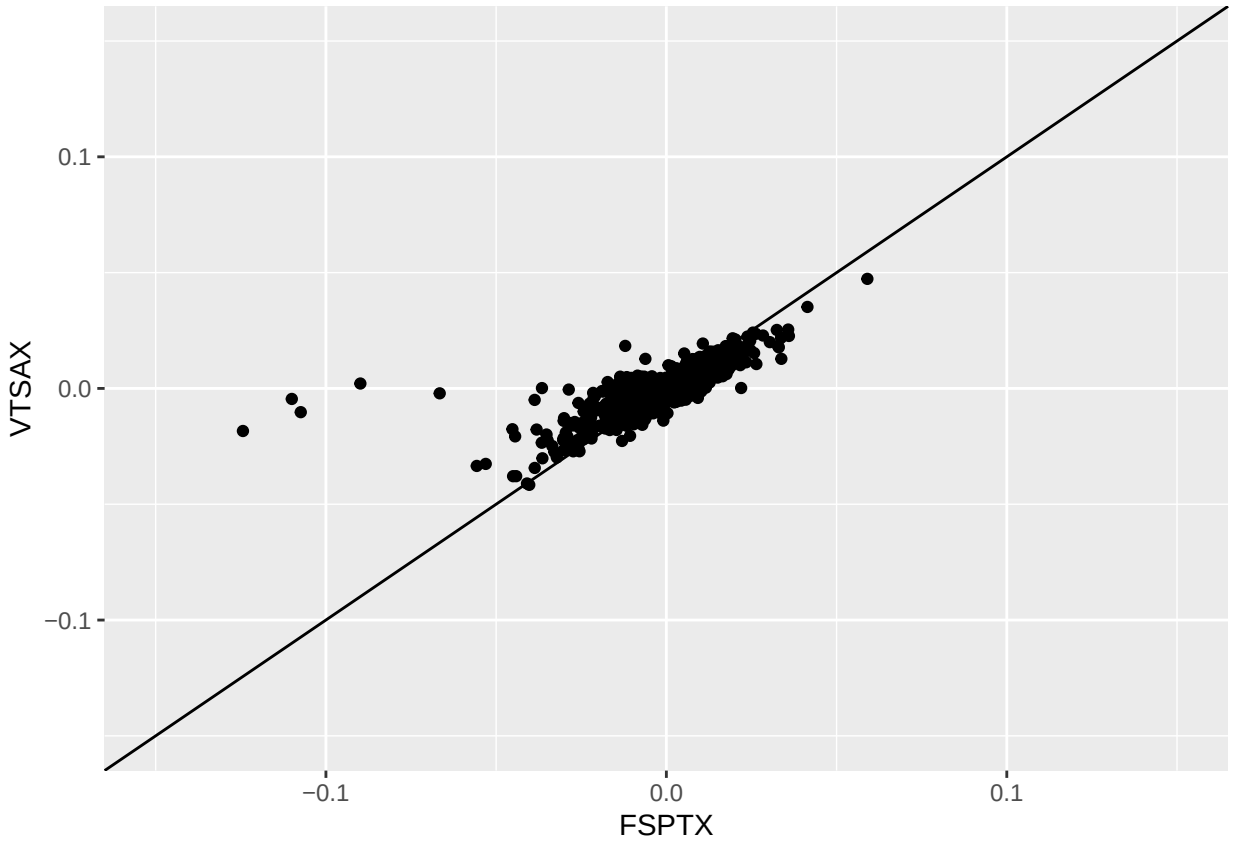

```
ggplot(dailyReturnComp)+aes(x = FSPTX,y = VGT)+geom_point()+xlim(c(-0.15,0.15))+ylim(c(-0.15,0.15))+geo
```




```
ggplot(dailyReturnComp)+aes(x = FSPTX,y = VIGAX)+geom_point()+xlim(c(-0.15,0.15))+ylim(c(-0.15,0.15))+g
```




```
ggplot(dailyReturnComp)+aes(x = FSPTX,y = VTSAX)+geom_point()+xlim(c(-0.15,0.15))+ylim(c(-0.15,0.15))+g
```


0.1
0.0
−0.1
VTSAX
−0.1
0.0
0.1
FSPTX