MA678 Final Project

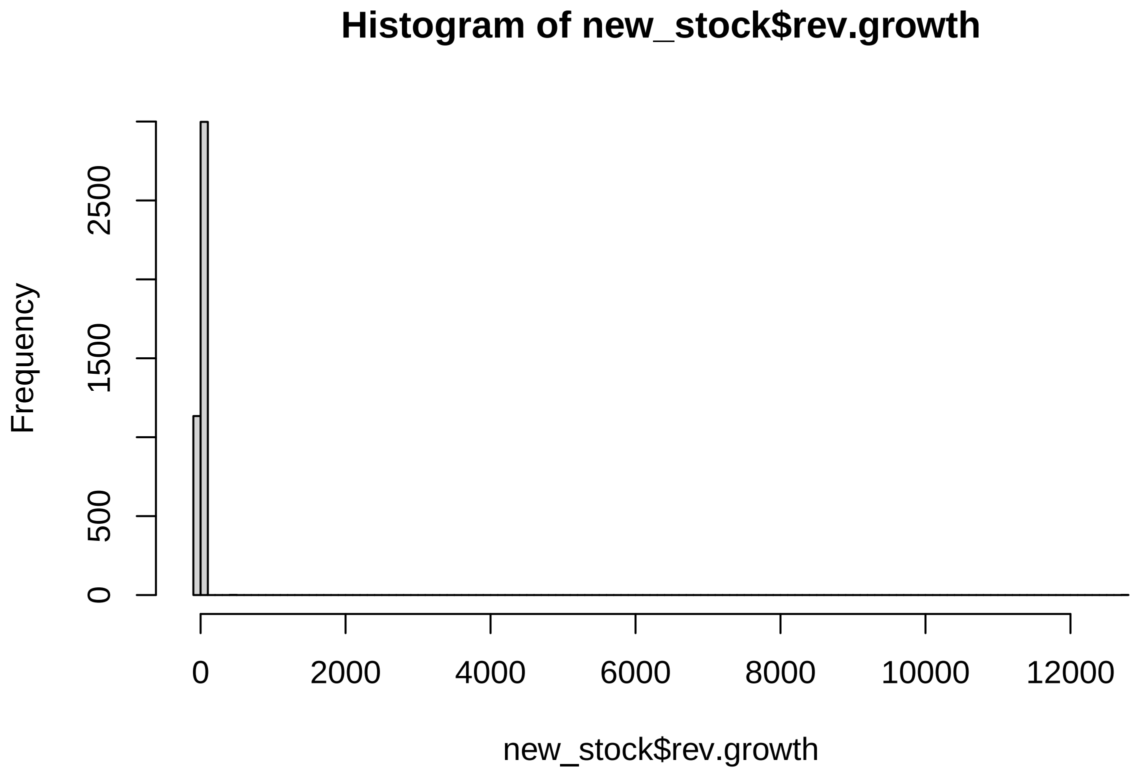
Kaiwei Xiao

2022-12-11

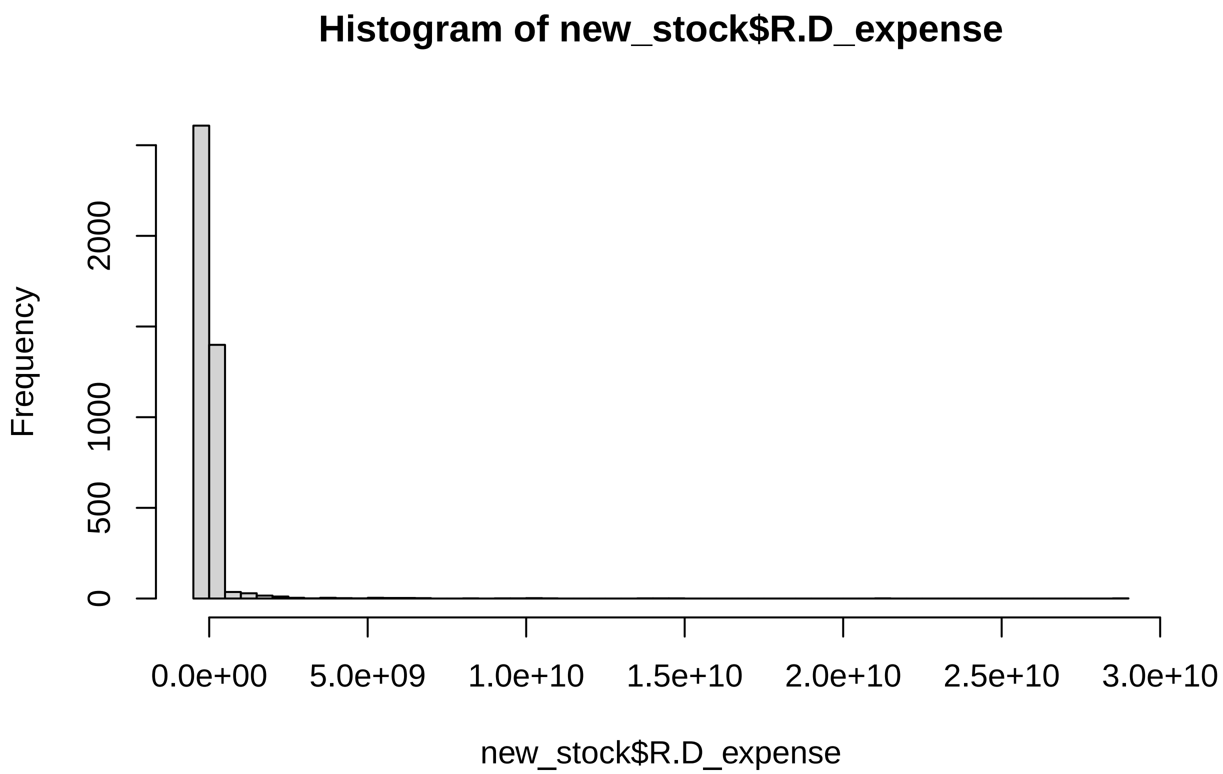
```
stock = read.csv("/Users/j/MA678/final project/archive/2018_Financial_Data.csv"
                 ,header = T)
```

```
#clean the data and rename the columns

new_stock = stock[, c('Revenue.Growth', 'R.D.Expenses',
                      'X2019.PRICE.VAR....', 'Sector')]

new_stock = na.omit(data.frame(new_stock))
colnames(new_stock) =  c("rev.growth","R.D_expense","var","sector")
```

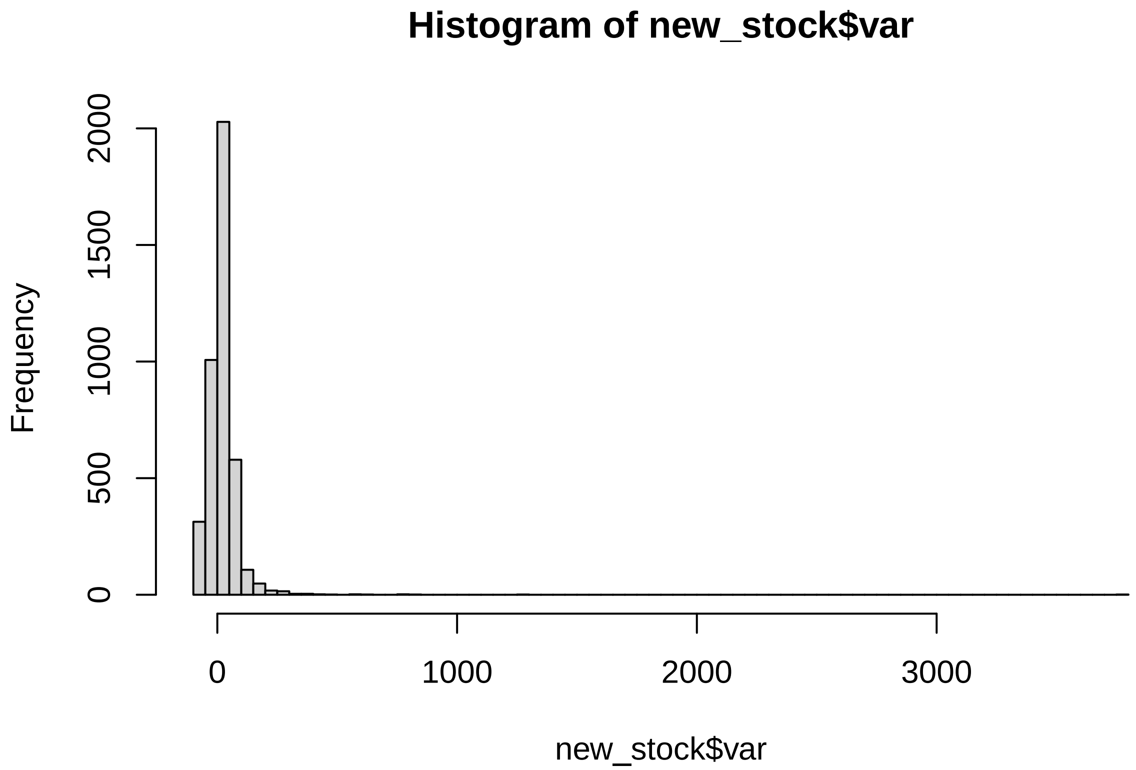
```
hist(new_stock$rev.growth, breaks = 100)
```



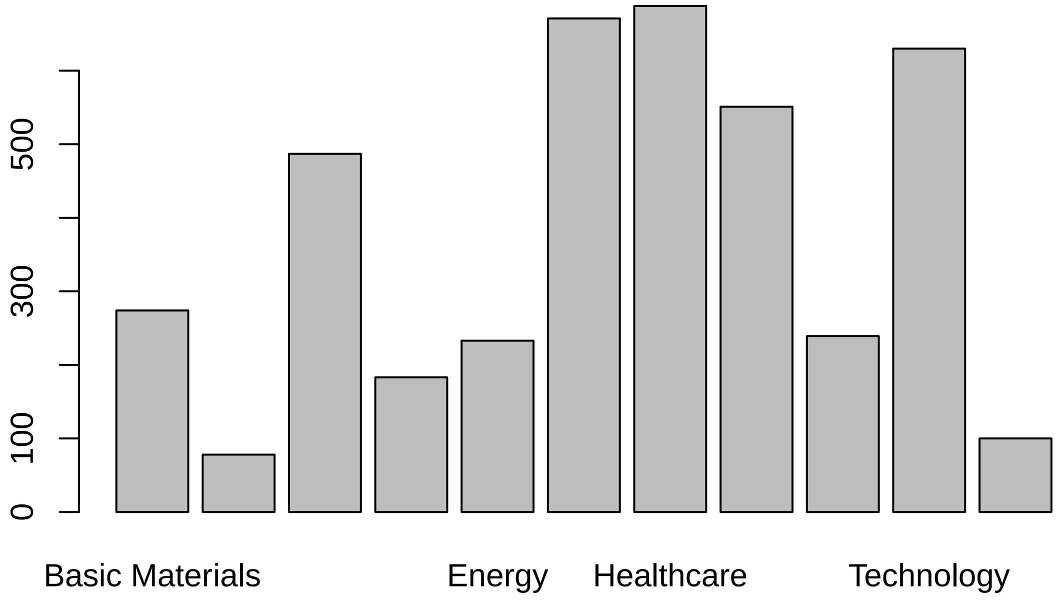
```r
hist(new_stock$R.D_expense, breaks = 100)
```



```r
hist(new_stock$var, breaks = 100)
```



```
barplot(table(new_stock$sector))
```



```
levels(factor(new_stock$sector))
```

```
##  [1] "Basic Materials"        "Communication Services" "Consumer Cyclical"
##  [4] "Consumer Defensive"     "Energy"                 "Financial Services"
##  [7] "Healthcare"             "Industrials"            "Real Estate"
## [10] "Technology"             "Utilities"
```

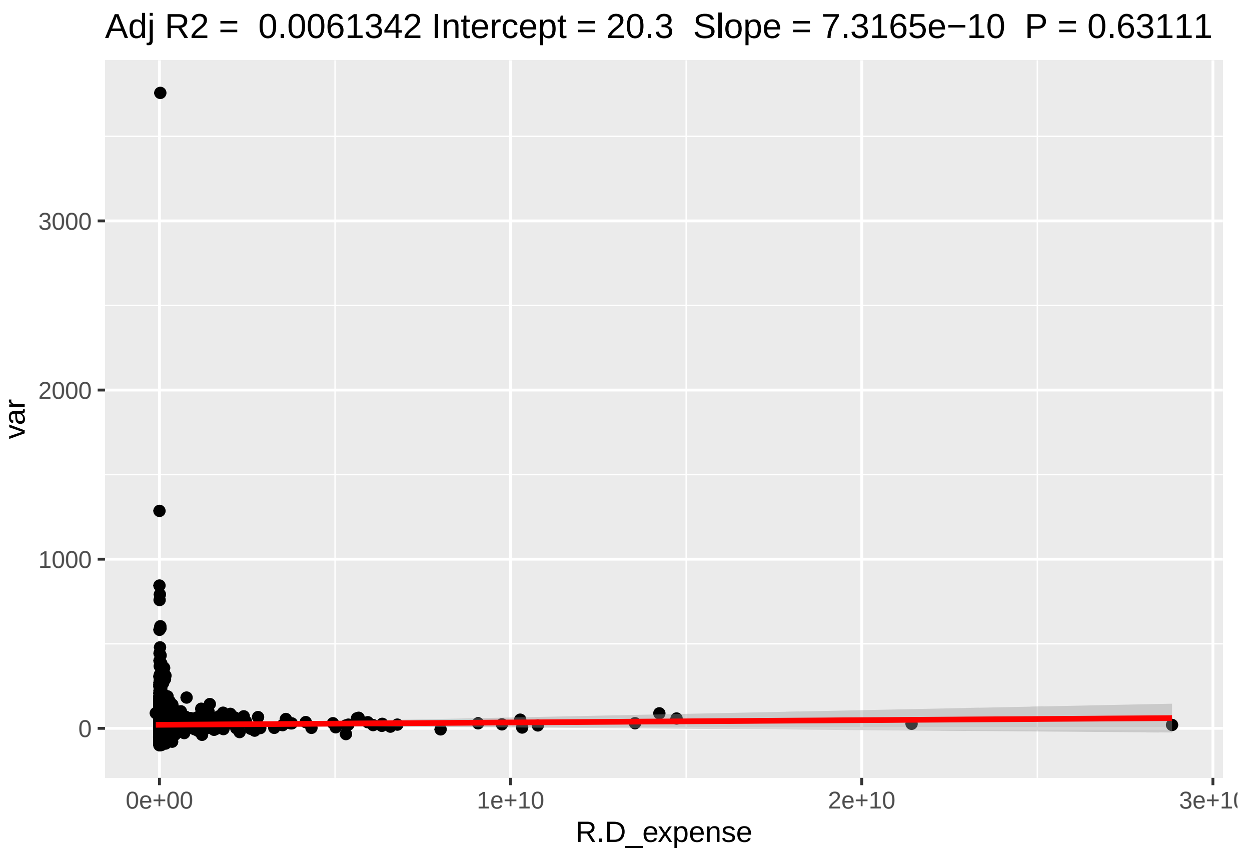
```
slr = lm( new_stock$var ~ new_stock$rev.growth + new_stock$R.D_expense +
            new_stock$sector)
summary(slr)
```

```
##
## Call:
## lm(formula = new_stock$var ~ new_stock$rev.growth + new_stock$R.D_expense +
##     new_stock$sector)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -127.7  -29.6   -2.8   20.2 3730.8
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           2.030e+01  5.114e+00   3.969 7.33e-05
## new_stock$rev.growth                 -2.094e-03  6.647e-03  -0.315  0.75270
## new_stock$R.D_expense                 7.317e-10  1.524e-09   0.480  0.63111
## new_stock$sectorCommunication Services -1.583e+01  1.086e+01  -1.457  0.14520
## new_stock$sectorConsumer Cyclical      -4.411e+00  6.393e+00  -0.690  0.49023
## new_stock$sectorConsumer Defensive      7.999e-01  8.082e+00   0.099  0.92116
## new_stock$sectorEnergy                 -2.594e+01  7.544e+00  -3.439  0.00059
## new_stock$sectorFinancial Services      1.961e-01  6.069e+00   0.032  0.97423
## new_stock$sectorHealthcare              5.615e+00  6.057e+00   0.927  0.35394
## new_stock$sectorIndustrials             2.430e+00  6.258e+00   0.388  0.69775
## new_stock$sectorReal Estate             5.003e+00  7.493e+00   0.668  0.50431
## new_stock$sectorTechnology              8.643e+00  6.150e+00   1.405  0.16000
## new_stock$sectorUtilities               4.627e+00  9.890e+00   0.468  0.63996
##
## (Intercept)                            ***
## new_stock$rev.growth
## new_stock$R.D_expense
## new_stock$sectorCommunication Services
## new_stock$sectorConsumer Cyclical
## new_stock$sectorConsumer Defensive
## new_stock$sectorEnergy                 ***
## new_stock$sectorFinancial Services
## new_stock$sectorHealthcare
## new_stock$sectorIndustrials
## new_stock$sectorReal Estate
## new_stock$sectorTechnology
## new_stock$sectorUtilities
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 84.65 on 4121 degrees of freedom
## Multiple R-squared:  0.00902,    Adjusted R-squared:  0.006134
## F-statistic: 3.126 on 12 and 4121 DF,  p-value: 0.0001935
```

```
ggplotRegression <- function (fit) {

require(ggplot2)

ggplot(fit$model, aes_string(x = names(fit$model)[2]
                             , y = names(fit$model)[1])) +
  geom_point() +
  stat_smooth(method = "lm", col = "red") +
  labs(title = paste("Adj R2 = ",signif(summary(fit)$adj.r.squared, 5),
                     "Intercept =",signif(fit$coef[[1]],5 ),
                     " Slope =",signif(fit$coef[[2]], 5),
                     " P =",signif(summary(fit)$coef[2,4], 5)))
}
ggplotRegression(lm(var ~ R.D_expense + rev.growth + sector, data = new_stock))
```

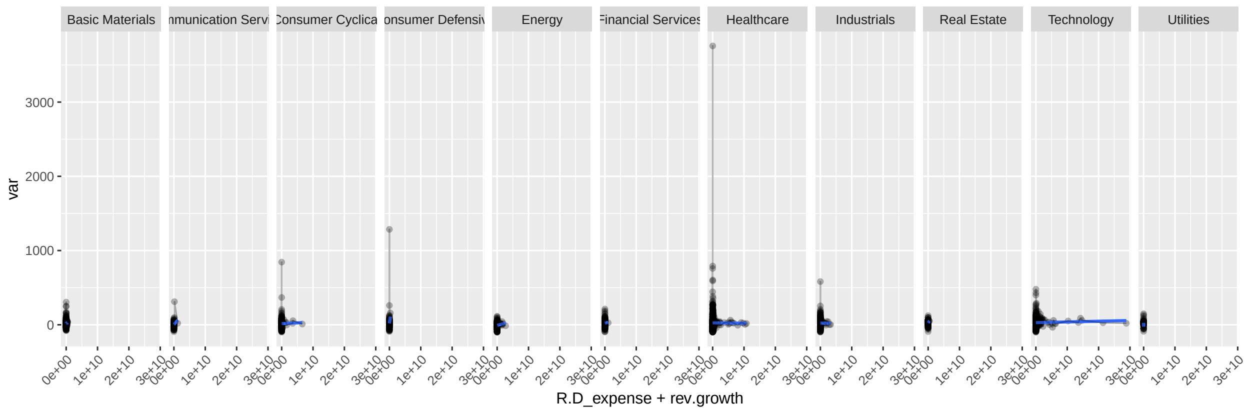
```
## Loading required package: ggplot2
```

```
## `geom_smooth()` using formula 'y ~ x'
```

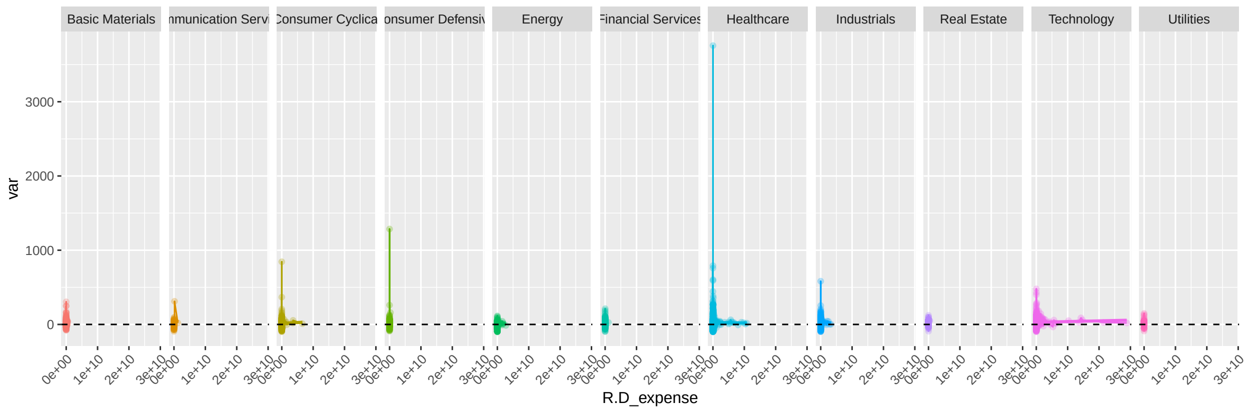


```
## Warning: Ignoring unknown aesthetics: sector
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
## `geom_smooth()` using formula 'y ~ x'
```



```
library(lme4)
```

```
## Loading required package: Matrix
```

```
new_stock$fsector = factor(new_stock$sector)
new_stock$pvar =1+ new_stock$var/100
new_stock$growth = new_stock$rev.growth +1




#define Min-Max normalization function
min_max_norm <- function(x) {
    (x - min(x)) / (max(x) - min(x))
  }

#apply Min-Max normalization to  columns in  dataset

nstock <- as.data.frame(lapply(new_stock[,c(1,2,3)], min_max_norm))
nstock[,"sector"] = new_stock$fsector

head(nstock)
```

```
##     rev.growth R.D_expense        var            sector
## 1 0.0002804011 0.003600404 0.03439817  Consumer Cyclical
## 2 0.0002741621 0.003600404 0.03641900             Energy
## 3 0.0002817666 0.471549210 0.03375018         Technology
## 4 0.0003105365 0.077577986 0.04254507         Technology
## 5 0.0002738874 0.003600404 0.03750021        Industrials
## 6 0.0002751509 0.003600404 0.03738564 Financial Services
```

```
mix = lmer(var ~ rev.growth + R.D_expense + ((1 + R.D_expense + rev.growth)|sector)
           , data=nstock )
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(mix)
```

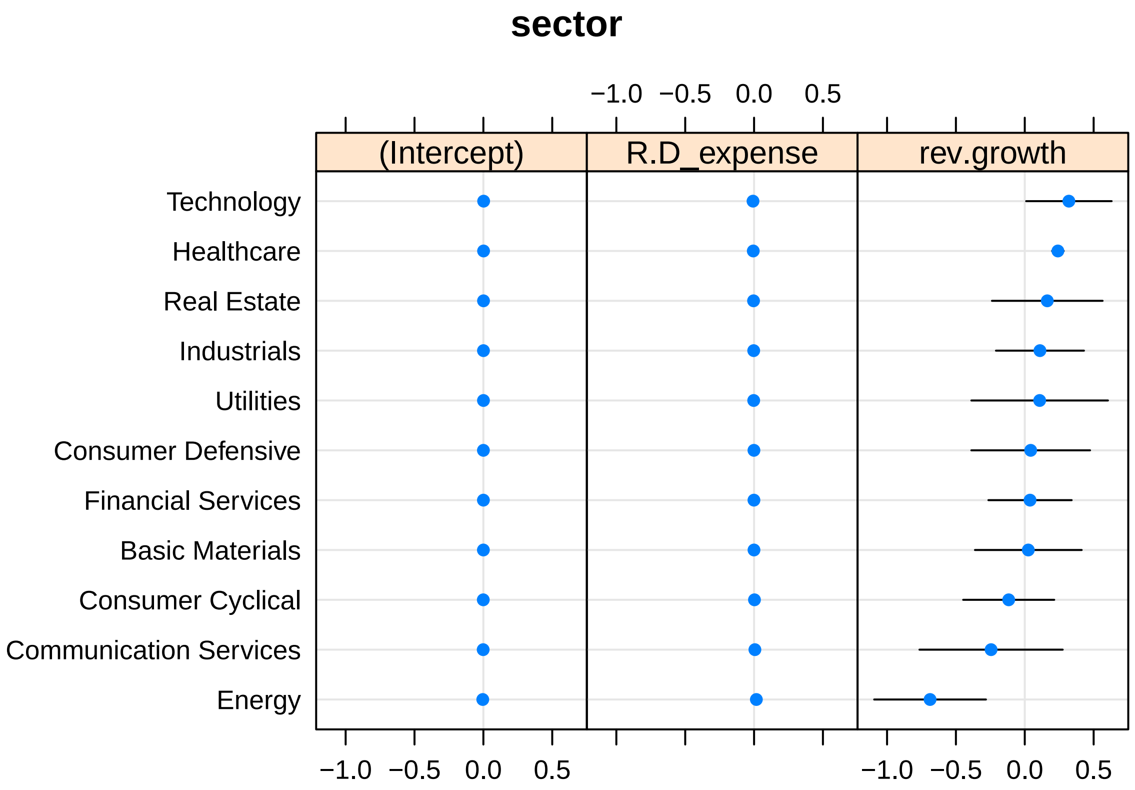
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: var ~ rev.growth + R.D_expense + ((1 + R.D_expense + rev.growth) |
##     sector)
##    Data: nstock
##
## REML criterion at convergence: -19805.5
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.490 -0.352 -0.034  0.238 44.077
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  sector   (Intercept) 4.228e-06 0.002056
##           R.D_expense 6.247e-05 0.007904 -1.00
##           rev.growth  1.151e-01 0.339240  0.96 -0.96
##  Residual             4.819e-04 0.021951
## Number of obs: 4134, groups:  sector, 11
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  0.0309275  0.0007414  41.714
## rev.growth  -0.2471375  0.1860099  -1.329
## R.D_expense  0.0130297  0.0115923   1.124
##
## Correlation of Fixed Effects:
##             (Intr) rv.grw
## rev.growth   0.482
## R.D_expense -0.304 -0.111
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```r
coef(mix)
```

```
## $sector
##                         (Intercept)   rev.growth R.D_expense
## Basic Materials          0.03109236 -0.221428763 0.012396001
## Communication Services   0.02937675 -0.491735228 0.018990375
## Consumer Cyclical        0.03019616 -0.363863972 0.015840764
## Consumer Defensive       0.03119996 -0.204185687 0.011982425
## Energy                   0.02658465 -0.934760780 0.029722481
## Financial Services       0.03117387 -0.208978990 0.012082704
## Healthcare               0.03242517 -0.006749558 0.007273019
## Industrials              0.03163100 -0.136970271 0.010325628
## Real Estate              0.03196044 -0.083836648 0.009059327
## Technology               0.03295192  0.073060860 0.005248346
## Utilities                0.03161004 -0.139063459 0.010406163
##
## attr(,"class")
## [1] "coef.mer"
```

```r
#install.packages("lattice")
library(lattice)
dotplot(ranef(mix, condVar=T))
```

```
## $sector
```



```
#install.packages("glmmTMB")
library(glmmTMB)
plot_model(mix, type = "re", show.values = TRUE)
```

```
plot_model(mix, type = "diag", show.values = TRUE)
```