



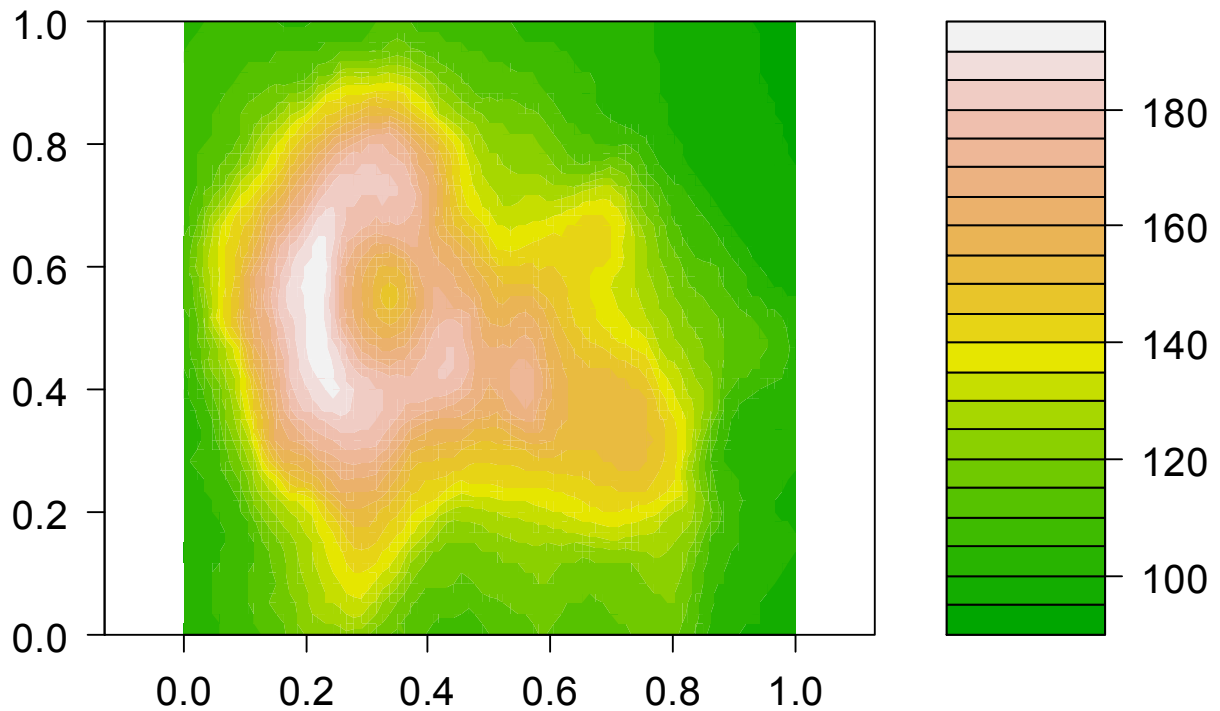
Projekt „*Nowa oferta edukacyjna Uniwersytetu Wrocławskiego odpowiedzią na współczesne potrzeby rynku pracy i gospodarki opartej na wiedzy*”

Dane:
Eksploracja (mining)

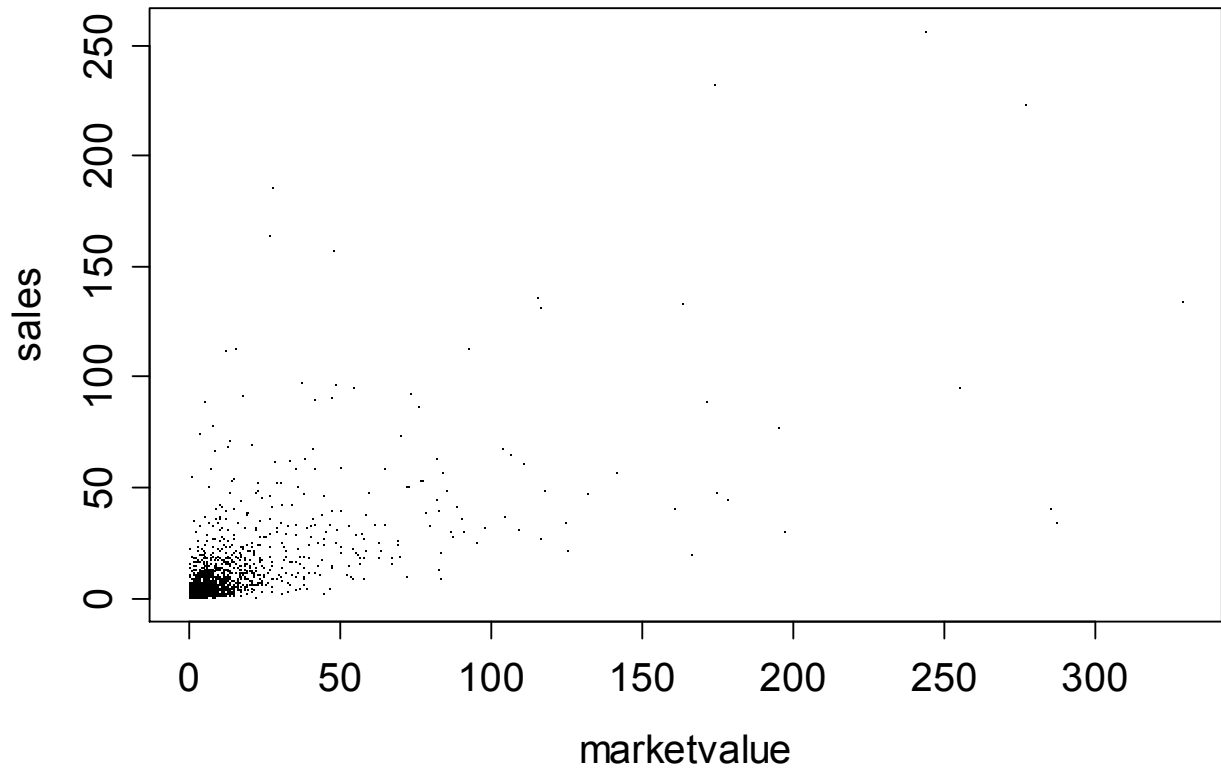
Relacje między parami zmiennych

```
filled.contour(volcano, color.palette = terrain.colors, asp = 1)  
title(main = "volcano data: filled contour map")
```

volcano data: filled contour map

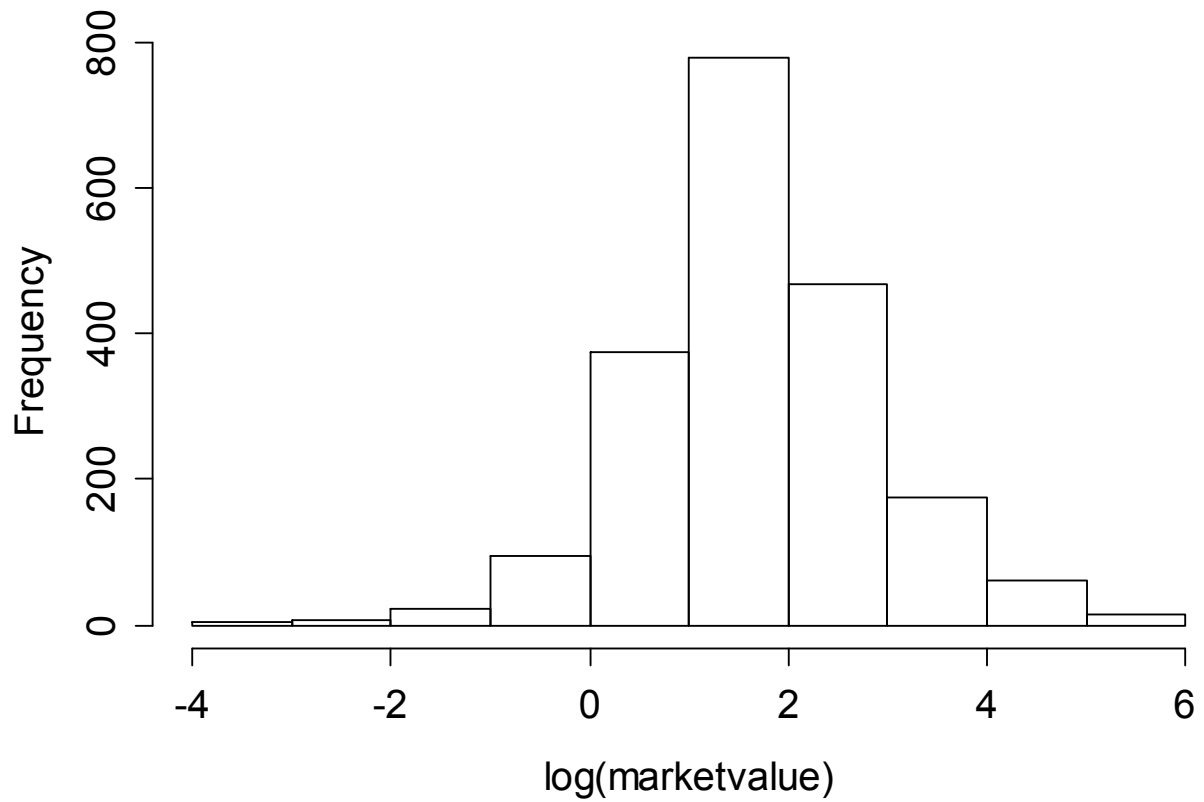


```
plot(sales~marketvalue,pch=".")
```

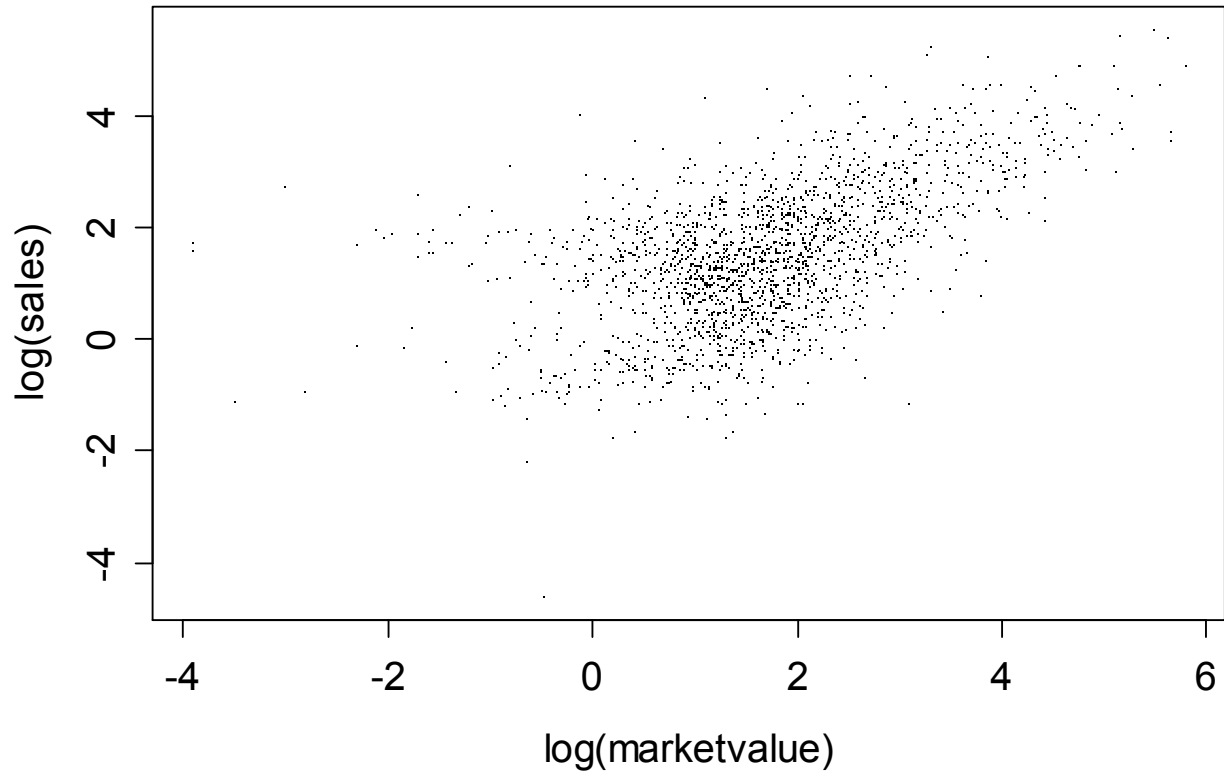


```
hist(marketvalue)
hist(log(marketvalue))
```

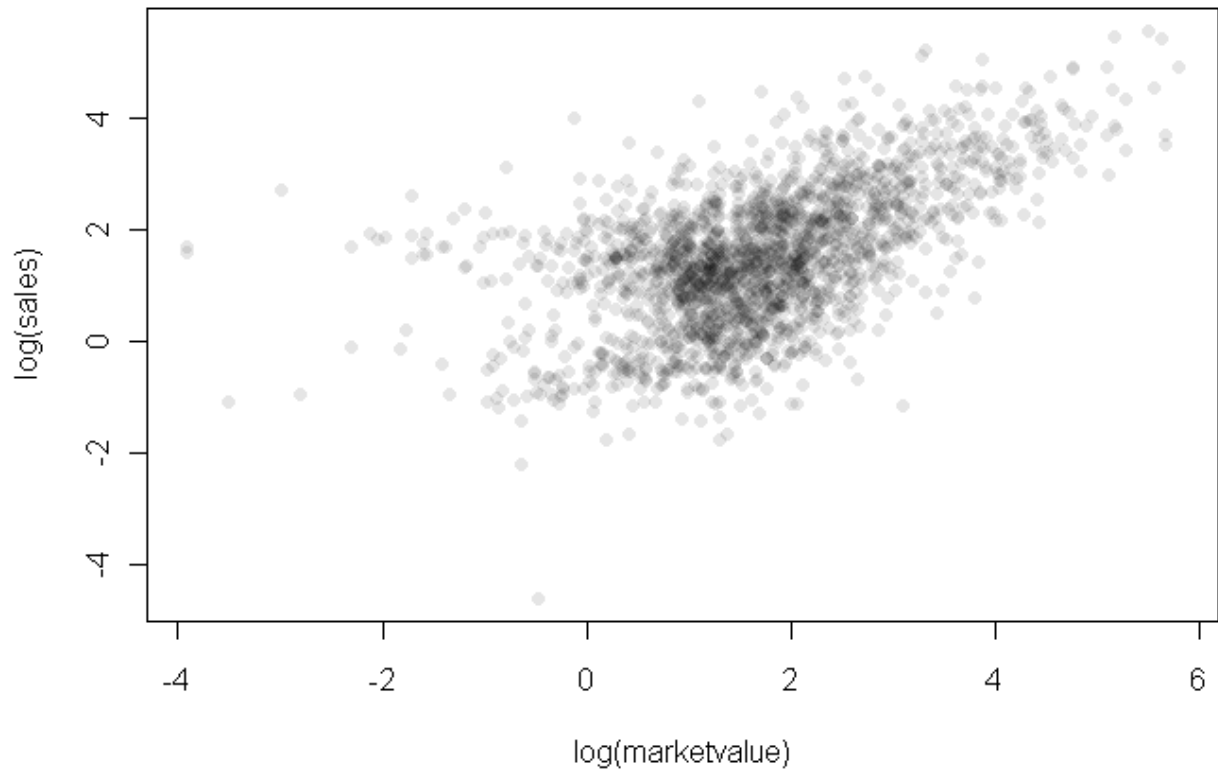
Histogram of log(marketvalue)



```
plot(log(sales)~log(marketvalue),pch=".")
```



```
plot(log(sales)~log(marketvalue),col=rgb(0,0,0,0.1),pch=16)
```



Drabina Tukeya

```
ladderTukey <- function(x,y){  
  qy <- fivenum(y)[1:3];qx <- fivenum(x)[1:3]  
  b1 <- (qy[2]-qy[1])/(qx[2]-qx[1])  
  b2 <- (qy[3]-qy[2])/(qx[3]-qx[2])  
  blad <- (b1-b2)/(b1+b2)  
  c(blad,b1,b2)  
}
```

```
ladderTukey(marketvalue,sales)  
-0.1313034 0.7425926 0.9670782
```

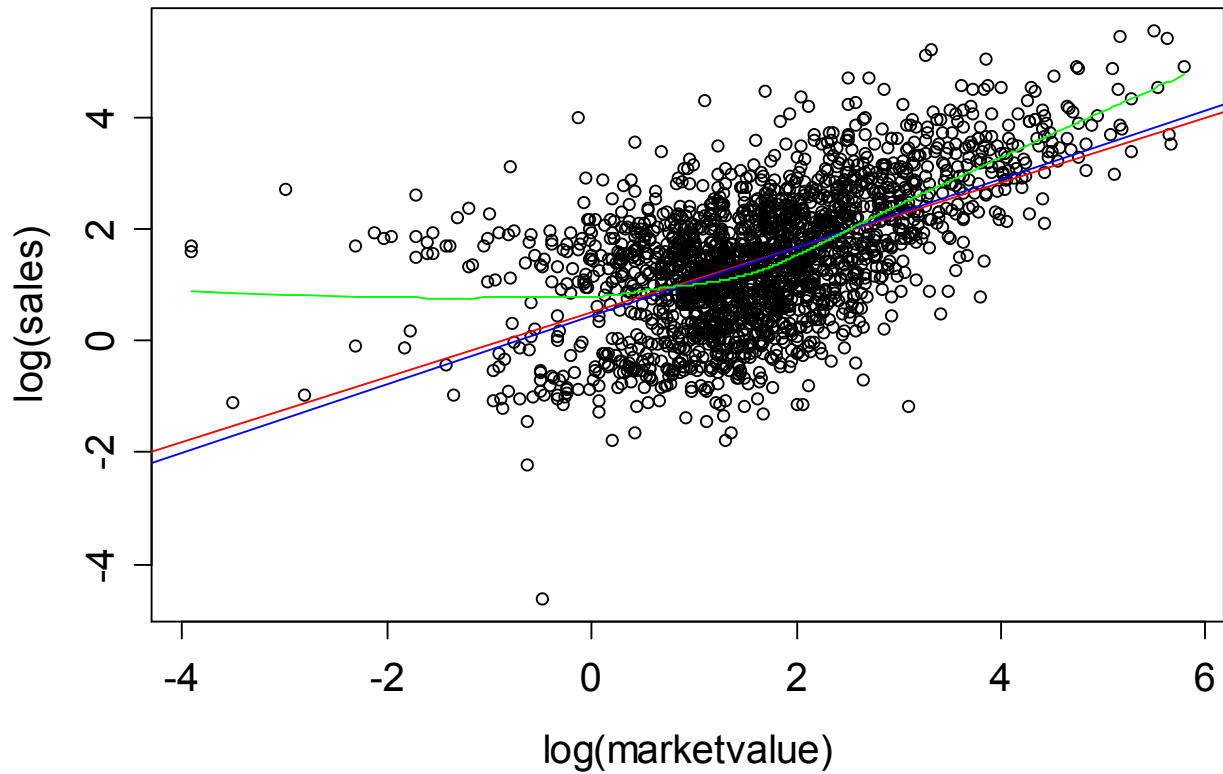
```
ladderTukey(marketvalue,log(sales))  
0.7213510 1.9651060 0.3181076
```

```
ladderTukey(log(marketvalue),log(sales))  
-0.05713144 1.08002423 1.21090852
```

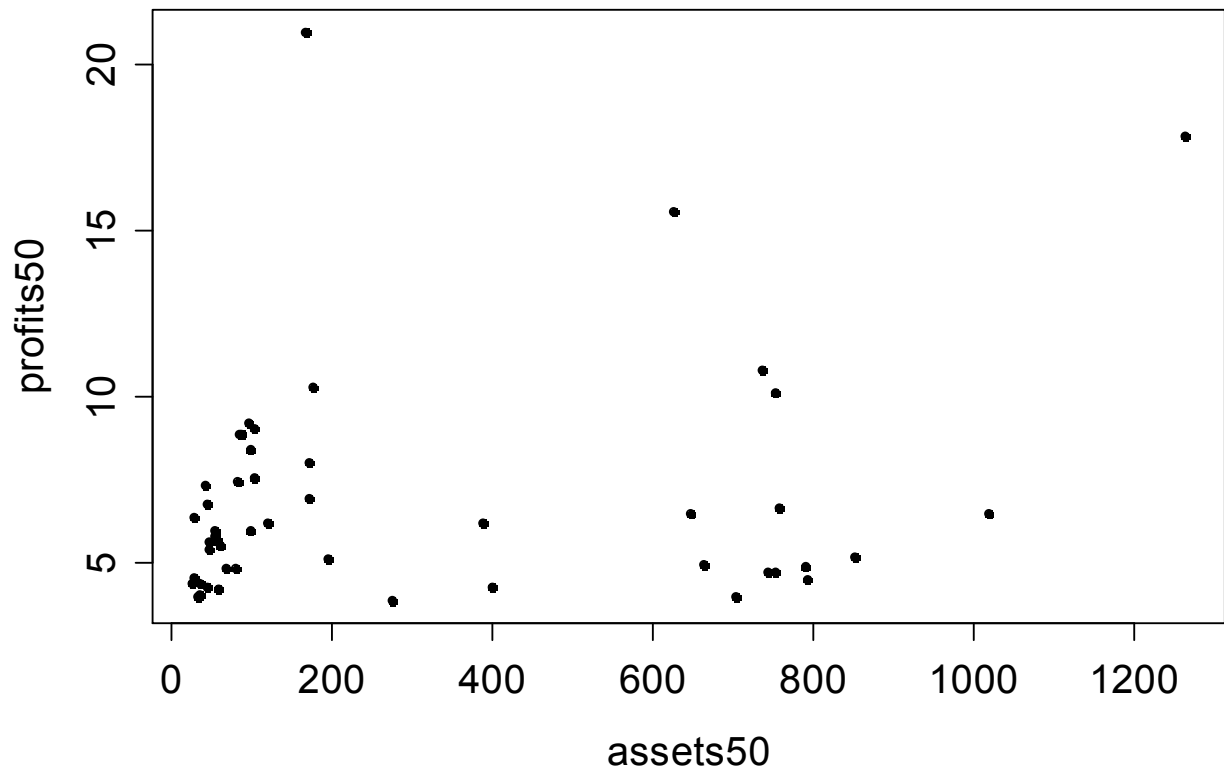
```
summary(lm(log(sales)~log(marketvalue)))
```

```
Call:  
lm(formula = log(sales) ~ log(marketvalue))  
  
Residuals:  
    Min       1Q   Median       3Q      Max   
-4.8400 -0.7255  0.0232  0.6890  3.9379   
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)      
(Intercept)    0.51153    0.03946   12.96  <2e-16 ***  
log(marketvalue) 0.57885    0.01907   30.35  <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 1.018 on 1998 degrees of freedom  
Multiple R-squared:  0.3155,    Adjusted R-squared:  0.3152   
F-statistic: 920.9 on 1 and 1998 DF,  p-value: < 2.2e-16
```

```
library(MASS)
plot(log(marketvalue),log(sales))
abline(coefficients(lm(log(sales)~log(marketvalue))),col="red")
abline(coefficients(rlm(log(sales)~log(marketvalue))),col="blue")
lines(lowess(x = log(marketvalue), y = log(sales)), col = "green", lwd = 1)
```

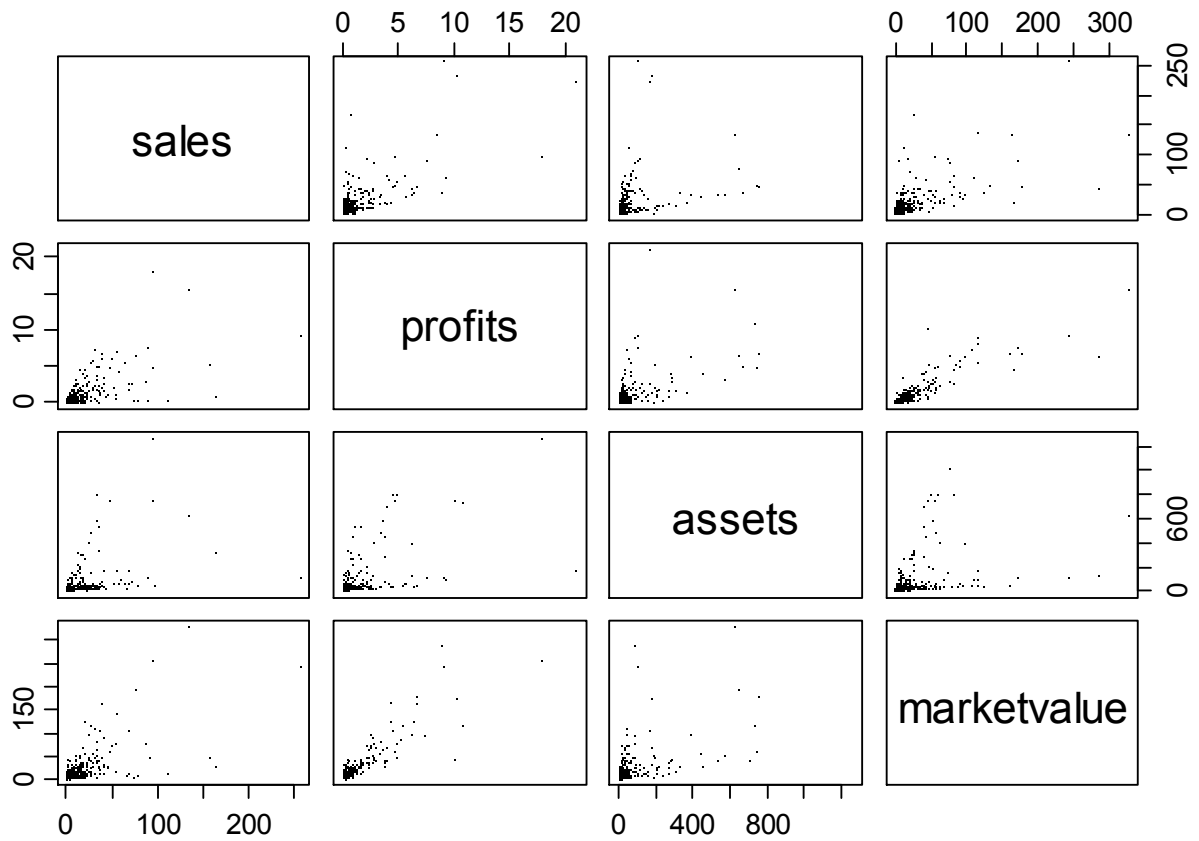


```
library(calibrate)
orderProfits <- rev(order(profits))
profits50 <- profits[orderProfits][6:55]
assets50 <- assets[orderProfits][6:55]
abbreviateCountry <- abbreviate(country)
abbCountry50 <- abbreviateCountry[orderProfits][6:55]
plot(assets50,profits50,pch=20)
```

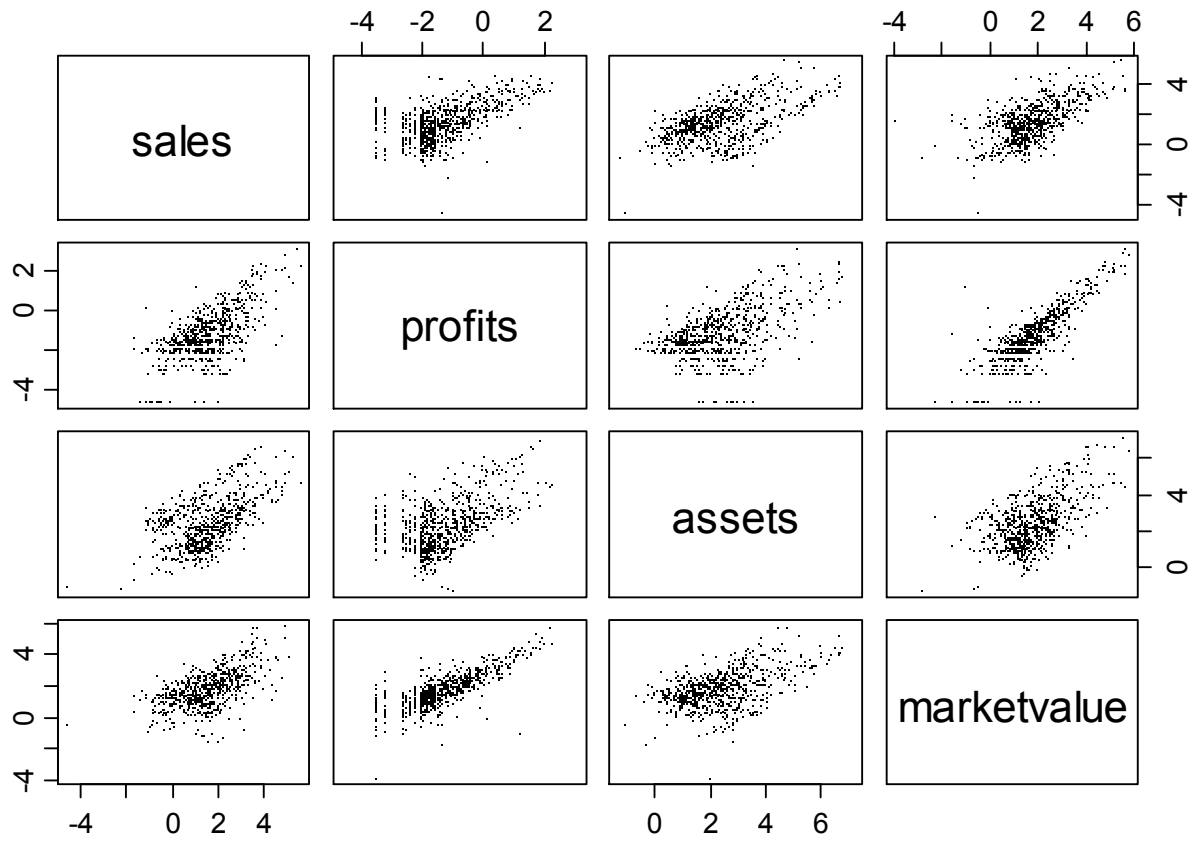



```
#fragment danych bez brakujących
```

```
which(apply(Forbes2000,2,is.na)==T)  
frb<-Forbes2000[!naProfits,c("sales","profits","assets","marketvalue")]  
frbPlus <- frb[frb$profits>0,c("sales","profits","assets","marketvalue")]  
pairs(frbPlus,pch=".")
```

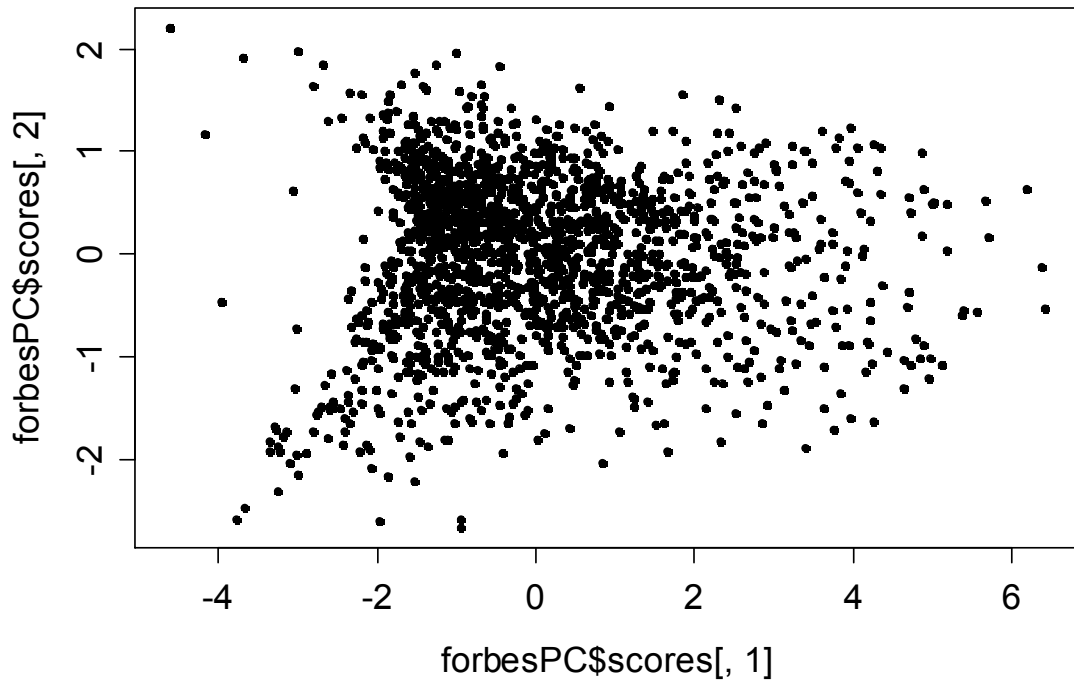


```
pairs(log(frbPlus),pch=".")
```



#składowe główne

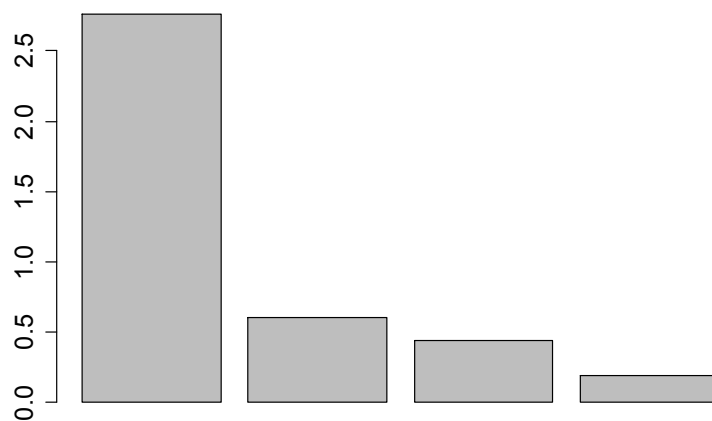
```
forbesPC <- princomp(log(frbPlus),cor=T)  
plot(forbesPC$scores[,2] ~forbesPC$scores[,1],pch=20)
```



```
> summary(forbesPC)
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.6613196	0.7750725	0.6660821	0.44228306
Proportion of Variance	0.6899957	0.1501844	0.1109164	0.04890358
Cumulative Proportion	0.6899957	0.8401801	0.9510964	1.00000000



Ważność składowych (wariancje)

```
> forbesPC$loadings
```

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4
sales	0.488	-0.186	0.850	
profits	0.523	0.466	-0.249	0.669
assets	0.446	-0.784	-0.431	
marketvalue	0.538	0.365	-0.172	-0.740

```
> forbesPC$center
```

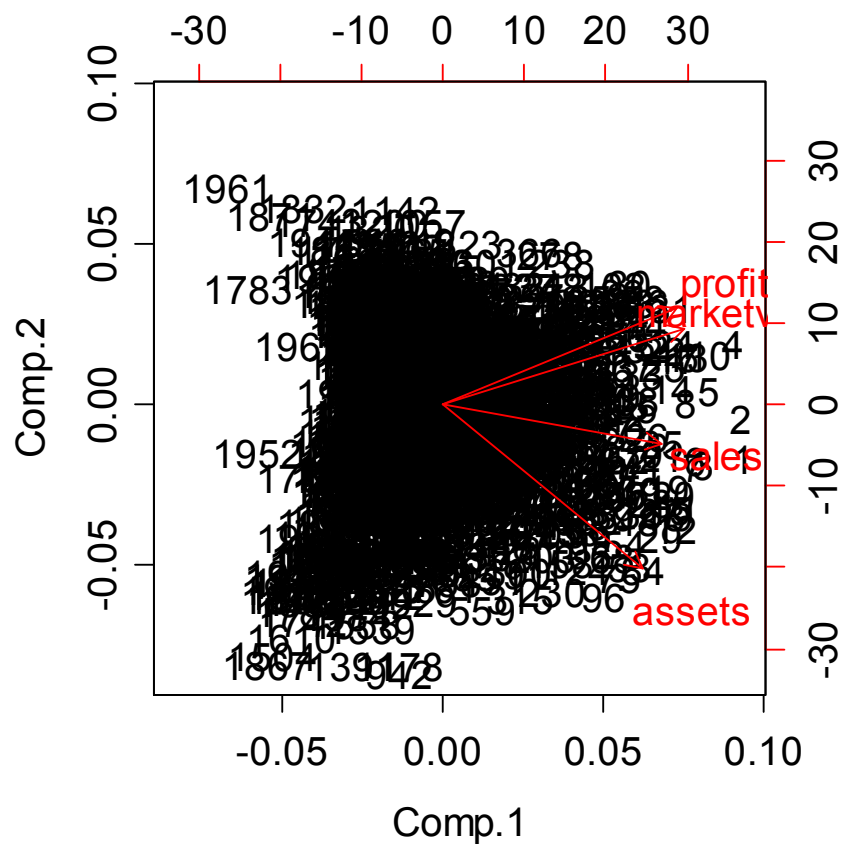
sales	profits	assets	marketvalue
1.457143	-1.320061	2.296328	1.763937

```
> forbesPC$scale
```

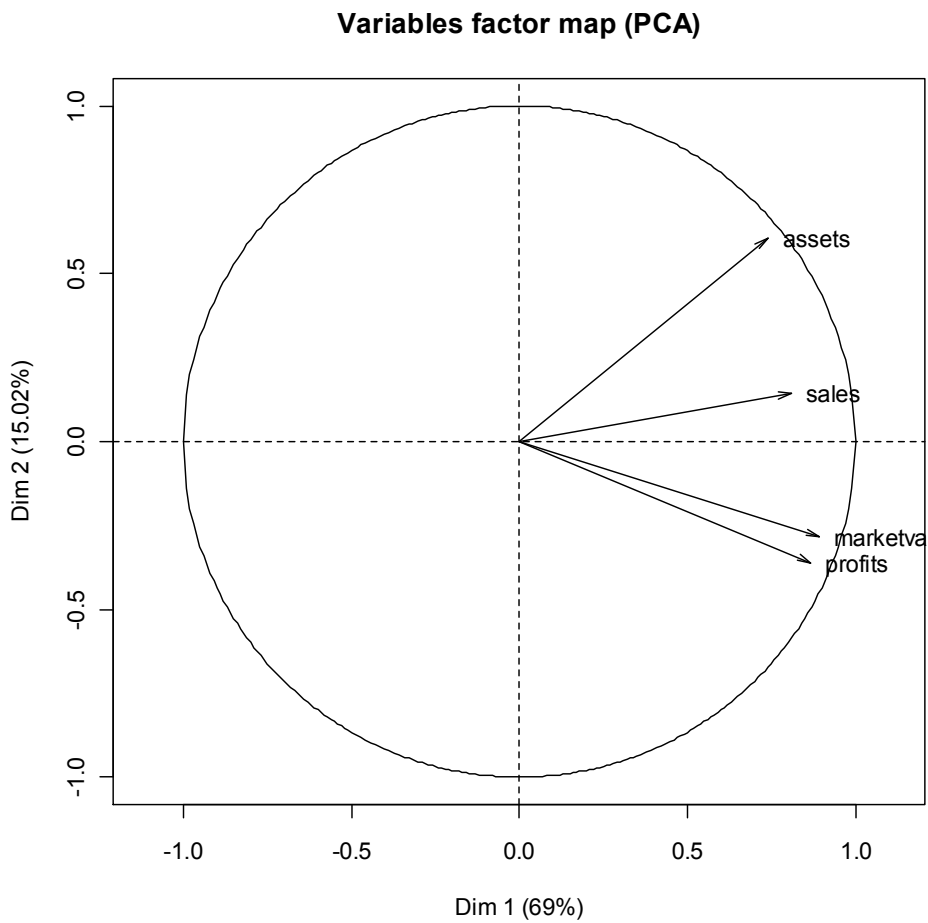
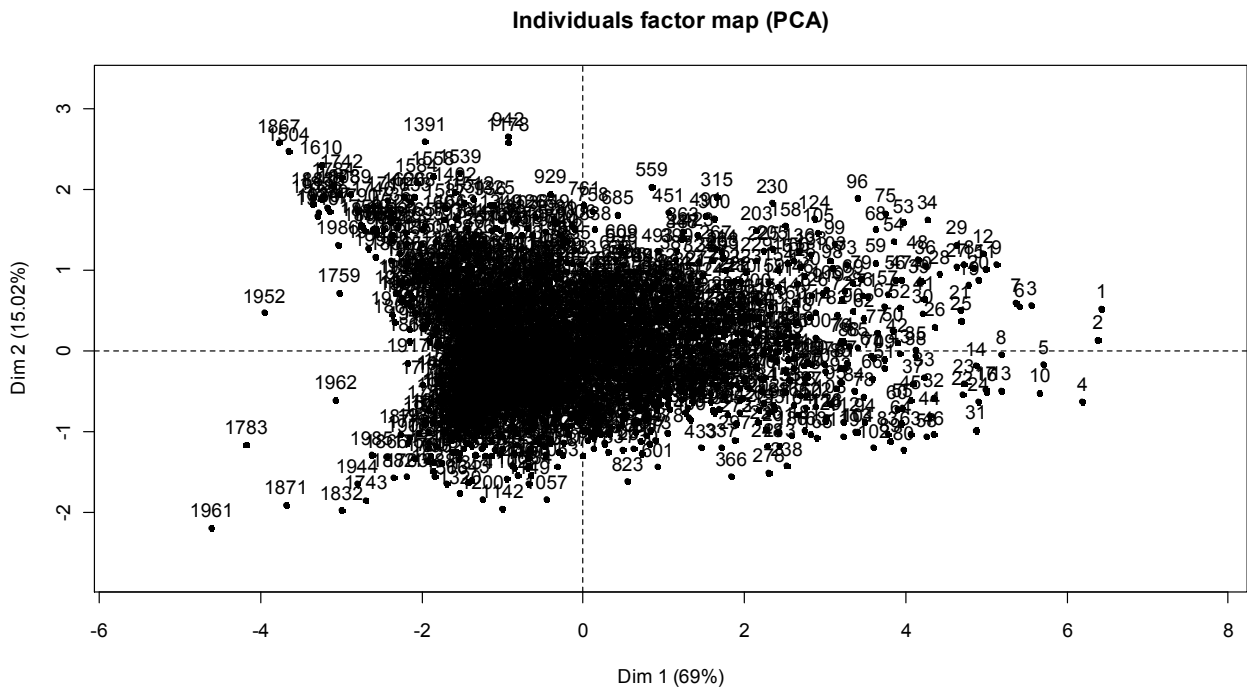
sales	profits	assets	marketvalue
1.234526	1.226756	1.326477	1.142203

```
> summary(forbesPC)
```

```
biplot(forbesPC)
```



```
library(FactoMineR)
result <- PCA(log(frbPlus)) # graphs generated automatically
```

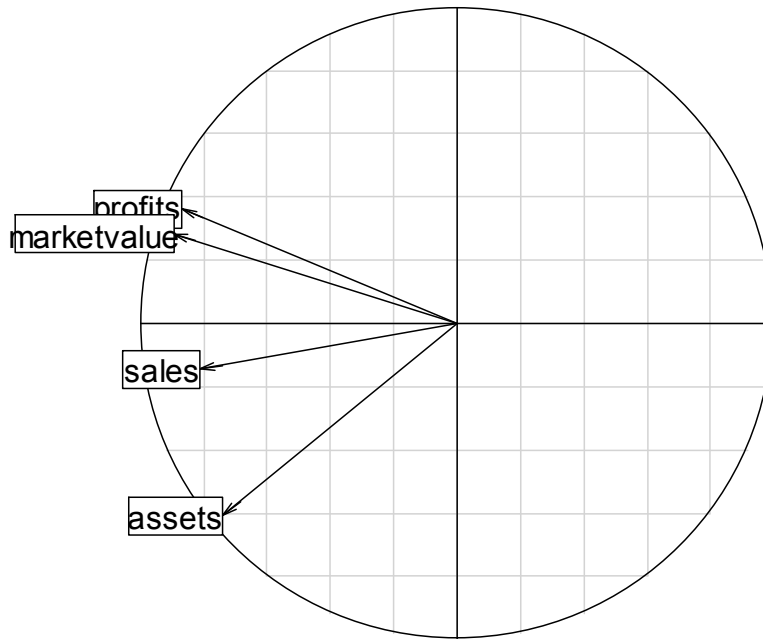


```
library(ade4)
forbesPCA<- dudi.pca(log(frbPlus), scan = T) # a normed PCA
```

```
> forbesPCA$c1
```

	CS1	CS2
sales	-0.4881373	-0.1864858
profits	-0.5228124	0.4661999
assets	-0.4457226	-0.7838260
marketvalue	-0.5382570	0.3653730

```
s.corcircle(forbesPCA$co, lab = names(frbPlus))
```



Pierwsza składowa główna a kolejność firm na liście Forbesa

```
> name[1:5]
[1] "Citigroup" "General Electric" "American Intl Group"
[4] "ExxonMobil" "BP"
```

```
> forbesPC$scores[,1][1:5]
      1      2      3      4      5
6.422928 6.386240 5.554190 6.188101 5.701150
```

```
> name[6:10]
[1] "Bank of America" "HSBC Group" "Toyota Motor" "Fannie Mae"
[5] "Wal-Mart Stores"
```

```
> forbesPC$scores[,1][6:10]
      6      7      8      9     10
5.401687 5.360522 5.177946 5.124337 5.668962
```

```
> name[1695:1705]
[1] "Kagoshima Bank" "Regal Entertainment Group"
[3] "Allergan" "Hong Leong Credit"
[5] "Washington Federal" "GTech Holdings"
[7] "Equifax" "Broadcom"
[9] "Dentsply Intl" "EMAP"
[11] "Neptune Orient Lines"
```

```
> forbesPC$scores[,1][1695:1705]
      1986      1987      1988      1989      1990      1991      1992      1995
-3.041736 -1.948617 -2.145160 -1.529307 -1.805009 -1.666984 -1.495070 -2.109997
      1996      1998      2000
-2.578421 -1.553430 -1.945621
```

Druga składowa główna a kolejność firm na liście Forbesa

```
> forbesPC$scores[,2][1:5]
```

```
      1      2      3      4      5  
-0.5251281 -0.1341675 -0.5707036  0.6288218  0.1653822
```

```
> forbesPC$scores[,2][6:10]
```

```
      6      7      8      9     10  
-0.54508177 -0.59806428  0.04054584 -1.07973652  0.52226436
```

```
> forbesPC$scores[,2][1695:1705]
```

```
      1986      1987      1988      1989      1990      1991      1992  
-1.3132939 -0.3329063 -0.8525001  0.5921074  0.3627939 -0.6005488  0.4087964  
      1995      1996      1998      2000  
-0.6535188 -1.1693902  0.5123731 -0.4417627
```


Analiza czynnikowa

```
forbesFA <- factanal(log(frbPlus), 1, rotation="varimax")
```

```
> print(forbesFA, digits=2, cutoff=.3, sort=TRUE)
```

Call:

```
factanal(x = log(frbPlus), factors = 1, rotation = "varimax")
```

Uniquenesses:

sales	profits	assets	marketvalue
0.54	0.25	0.66	0.16

Loadings:

```
[1] 0.68 0.86 0.58 0.92
```

	Factor1
SS loadings	2.38
Proportion Var	0.60

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 102.74 on 2 degrees of freedom.
The p-value is 4.91e-23

```
> load <- forbesFA$loadings[,1]
```

```
> plot(load,type="n") # set up plot
```

```
> text(load,labels=names(frbPlus),cex=.7) # add variable names
```

