

R e non solo

Pillole formative per operatori statistici locali

Diego Giuliani, Flavio Santi

StatCities 2017 - Trento

Lettura dei dati e struttura del dataset

```
> X<- read.csv(file='risposte.csv')
```

```
> str(X)
```

```
'data.frame':    250 obs. of  12 variables:
 $ id           : int  1 2 3 4 5 6 7 8 9 10 ...
 $ genere       : Factor w/ 2 levels "M","F": 1 2 1 2 2 1..
 $ eta          : num  46 69 33 64 60 46 56 59 50 60 ...
 $ cittadinanza : Factor w/ 3 levels "Italiana","UE",...: ..
 $ medico       : Ord.factor w/ 5 levels "Mai"<"Di rado"<..
 $ farmaci      : Factor w/ 3 levels "Non so","No",...: 2 ..
 $ soddisfazione: Ord.factor w/ 5 levels "Molto insoddis"..
 $ tono         : int  2 5 1 4 4 3 4 4 3 4 ...
 $ vista        : int  2 5 1 4 4 3 4 4 3 4 ...
 $ udito        : int  2 4 1 3 3 3 4 3 3 3 ...
 $ olfatto      : int  3 4 2 3 3 3 4 3 3 3 ...
 $ tatto        : int  4 2 4 3 3 3 4 3 3 3 ...
```

Struttura del dataset

```
> str(X$eta)
num [1:250] 46 69 33 64 60 46 56 59 50 60 ...
> summary(X$eta)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 29.00  44.00   52.00   53.48  62.00   85.00
> str(X$medico)
Ord.factor w/ 5 levels "Mai"<"Di rado"<...: 3 1 5 1 2 3 2...
> levels(X$medico)
[1] "Mai"      "Di rado"  "A volte"  "Spesso"   "Sempre"
> summary(X$medico)
   Mai Di rado A volte  Spesso  Sempre
   71    34    75    59    11
```

Distribuzione di frequenze I

```
> table(X$medico)
```

Mai	Di rado	A volte	Spesso	Sempre
71	34	75	59	11

```
> prop.table(table(X$medico))
```

Mai	Di rado	A volte	Spesso	Sempre
0.284	0.136	0.300	0.236	0.044

Distribuzione di frequenze II

```
> freq_ass<- table(X$medico)
> cbind(freq_ass,'freq_cum'=cumsum(freq_ass),
+       'freq_rel'=prop.table(freq_ass),
+       'freq_rel_cum'=cumsum(prop.table(freq_ass)))
```

	freq_ass	freq_cum	freq_rel	freq_rel_cum
Mai	71	71	0.284	0.284
Di rado	34	105	0.136	0.420
A volte	75	180	0.300	0.720
Spesso	59	239	0.236	0.956
Sempre	11	250	0.044	1.000

Distribuzione congiunta di frequenze

```
> table(X$genere,X$medico)
```

	Mai	Di rado	A volte	Spesso	Sempre
M	26	14	52	35	7
F	45	20	23	24	4

```
> prop.table(table(X$genere,X$medico))
```

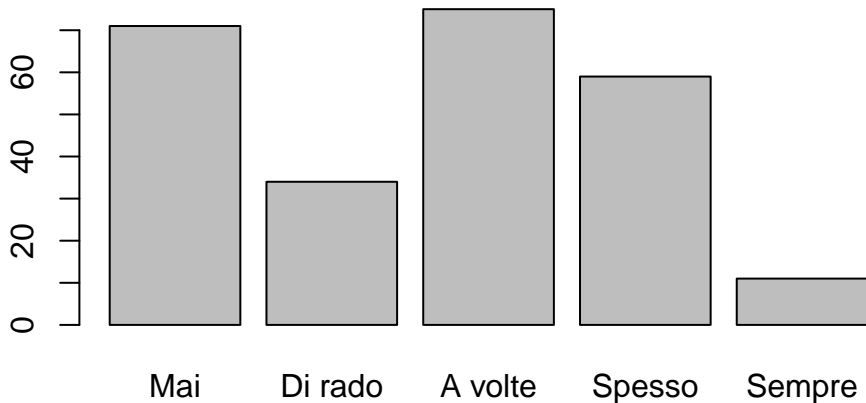
	Mai	Di rado	A volte	Spesso	Sempre
M	0.104	0.056	0.208	0.140	0.028
F	0.180	0.080	0.092	0.096	0.016

```
> prop.table(table(X$genere,X$medico),1)
```

	Mai	Di rado	A volte	Spesso	Sempre
M	0.19402985	0.10447761	0.38805970	0.26119403	0.05223881
F	0.38793103	0.17241379	0.19827586	0.20689655	0.03448276

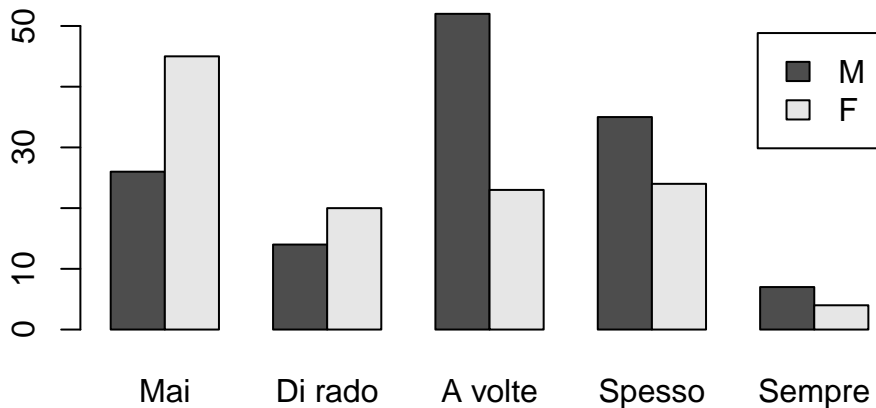
Grafici a barre I

```
> barplot(table(X$medico))
```



Grafici a barre II

```
> barplot(table(X$genere,X$medico),beside=TRUE,  
+         legend.text=TRUE)
```



Selezione dei soli questionari a cui è stata data risposta “Sì” o “No”:

```
> newX<- subset(X,farmaci!='Non so')
```

Stima del modello logistico:

```
> modello<- glm(formula=farmaci~genere+eta,  
+               data=newX,family='binomial')
```

Regressione logistica II

```
> summary(modello)
```

Call:

```
glm(formula = farmaci ~ genere + eta, family = "binomial", data = dati)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.39963	-0.49116	0.06637	0.44388	2.41647

Coefficients:

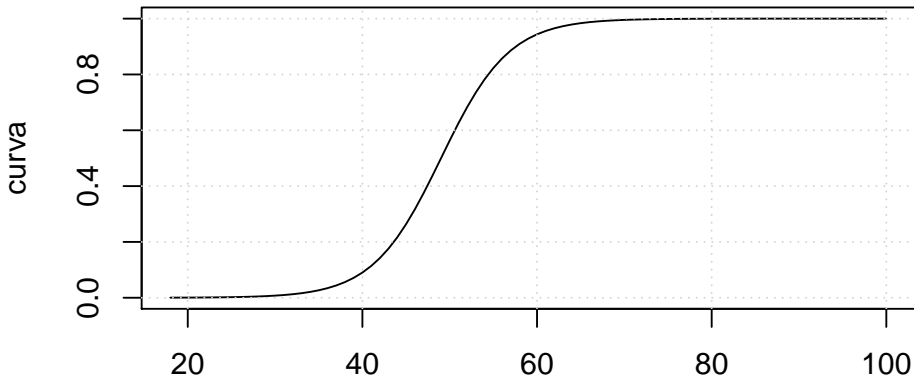
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-12.57502	1.81292	-6.936	4.02e-12	***
genereF	-0.04019	0.44847	-0.090	0.929	
eta	0.25661	0.03697	6.940	3.91e-12	***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Regressione logistica III

```
> curva<- predict(modello,data.frame(genere='M',eta=18:100),  
+                  type='response')  
> plot(18:100,curva,type='l')  
> grid()
```



Perché R?

- analisi replicabili ed automatizzabili
 - ▶ mediante l'uso di script
 - ▶ mediante la reportistica automatica
- ambiente altamente integrabile e flessibile
 - ▶ in input (file RData, xlsx, csv, tsv, ...; v. inoltre pacchetti “eurostat” e “wdi”)
 - ▶ in output (v. ad es. report automatici, monitoraggio attività, pagine interattive)
 - ▶ in fase di elaborazione con altri linguaggi ed applicativi
 - ▶ lato server
- ampia disponibilità di strumenti di analisi avanzati (11401 pacchetti disponibili al 14-09-2017 sul CRAN)
- documentazione gratuita molto ricca e di qualità
- è un software open source, quindi è **gratuito** e **sicuro**

- Espa, G. e R. Micciolo (2012). *Analisi esplorativa dei dati con R*, Apogeo.
- Espa, G. e R. Micciolo (2014). *Problemi ed esperimenti di statistica con R*, Apogeo.
- Espa, G., R. Micciolo e L. Canal (2013). *Ricerca con R. Metodi di inferenza statistica*, Apogeo.
- Gandrud, C. (2016) *Reproducible Research with R and R Studio*, Seconda edizione, CRC Press.
- Wickham, H. e G. Grolemund (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, O'Reilly.

Rapporto mensile

Pillole formative per operatori statistici locali - R e non solo

SmartCities - Trento, 14 settembre 2017

Strategie d'investimento

Nota metodologica. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur tempor luctus tortor, non porttitor ipsum molestie nec. Vivamus et augue mollis dolor imperdiet posuere. Maecenas laoreet porta est eget sodales. Aenean in pulvinar lectus. Sed ullamcorper porttitor nunc sed malesuada. Proin pulvinar neque risus, at hendrerit nunc maximus ut. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris eu dui non purus facilisis tincidunt.

Composizione per unità

Portfolio	ENI	FCA	UCG
PTFA	30	100	400
PTFB	100	250	200
PTFC	300	200	40

Composizione per valore

	ENI	FCA	UCG	Value
PTFA	451.31	862.91	5480.32	6794.54
PTFB	1504.38	2157.27	2740.16	6401.82
PTFC	4513.15	1725.82	548.03	6787.00

Composizione per quote

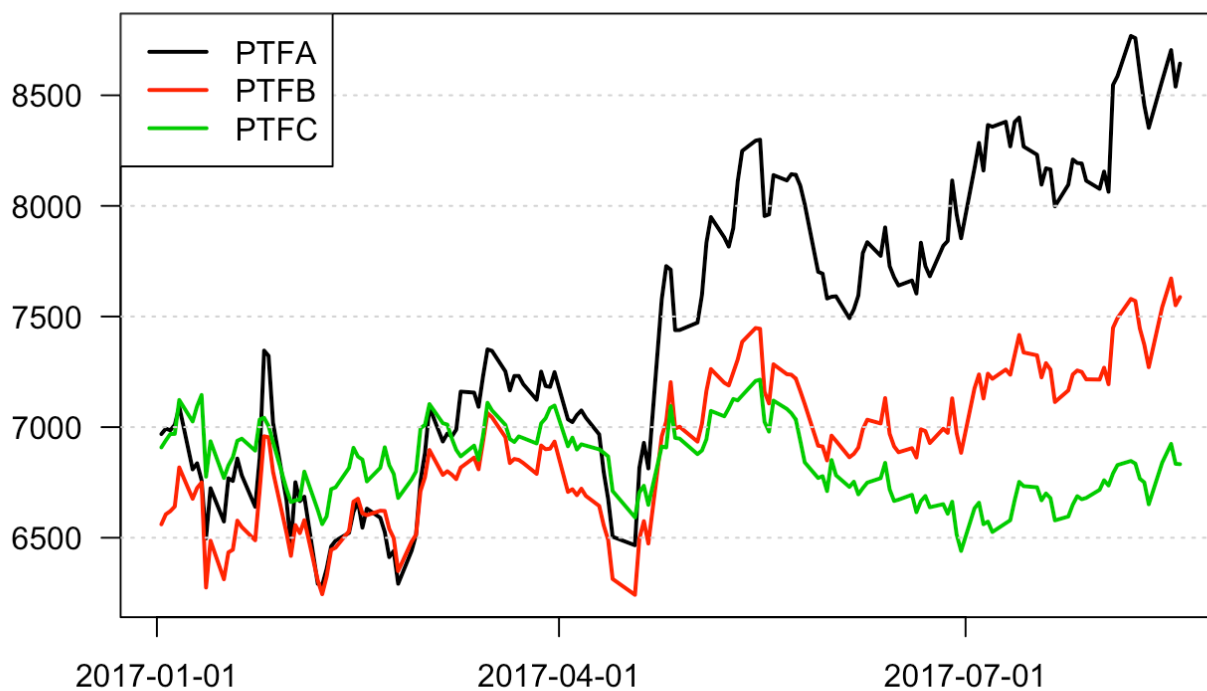
	ENI	FCA	UCG
PTFA	6.64	12.70	80.66
PTFB	23.50	33.70	42.80
PTFC	66.50	25.43	8.07

Rendimenti

Analisi e commenti. Vivamus nec consectetur enim, vel consectetur magna. Aliquam at risus orci. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. In nibh quam, iaculis et iaculis at, ullamcorper eu ante. Phasellus nunc dolor, hendrerit nec interdum nec, volutpat at risus. Vestibulum mollis sapien quis ullamcorper rutrum. Proin sed laoreet sem. In hac habitasse platea dictumst. Phasellus risus velit, lacinia vitae elementum sed, sagittis sit amet dolor. Suspendisse quis sodales tellus, in faucibus tortor.

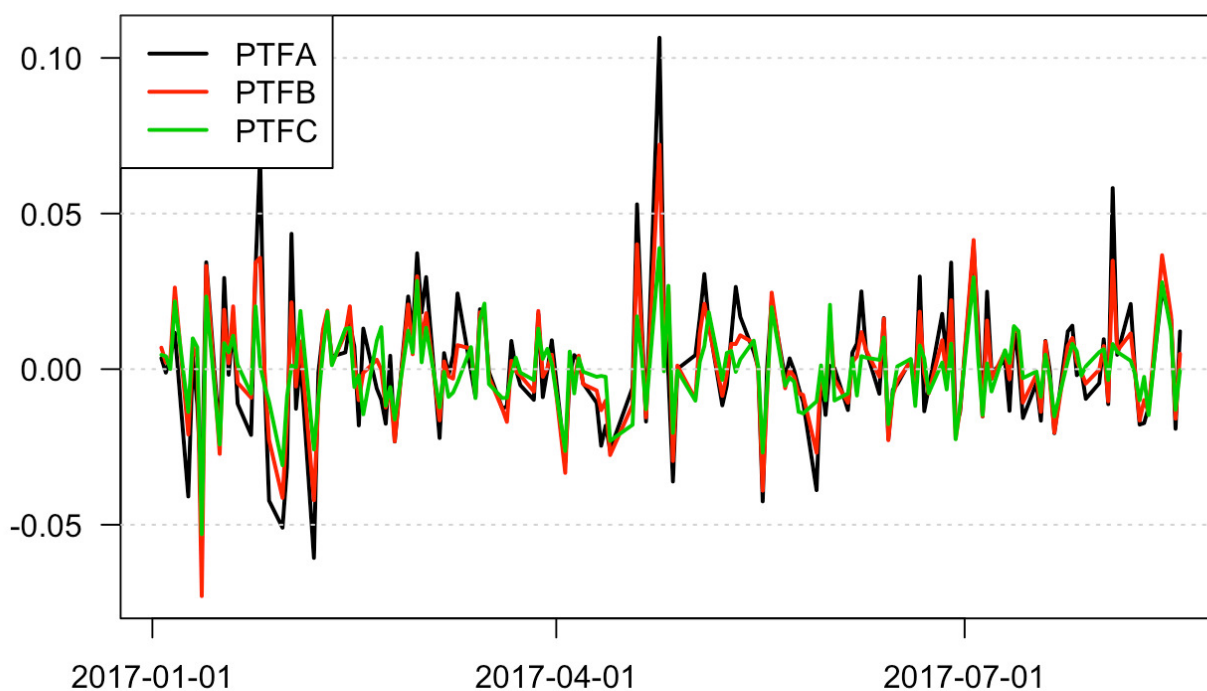
Prezzi

```
plot(compareptf,col=1:ncol(compareptf),plot.type='single',xlab='',ylab='',las=1,lwd=2)
grid(NA,NULL)
legend('topleft',legend=colnames(compareptf),col=1:ncol(compareptf),lty=1,lwd=2,bg='white')
```



Rendimenti

```
plot(returns(compareptf),col=1:ncol(compareptf),plot.type='single',xlab='',ylab='',las=1,lwd=2)
grid(NA,NULL)
legend('topleft',legend=colnames(compareptf),col=1:ncol(compareptf),lty=1,lwd=2,bg='white')
```



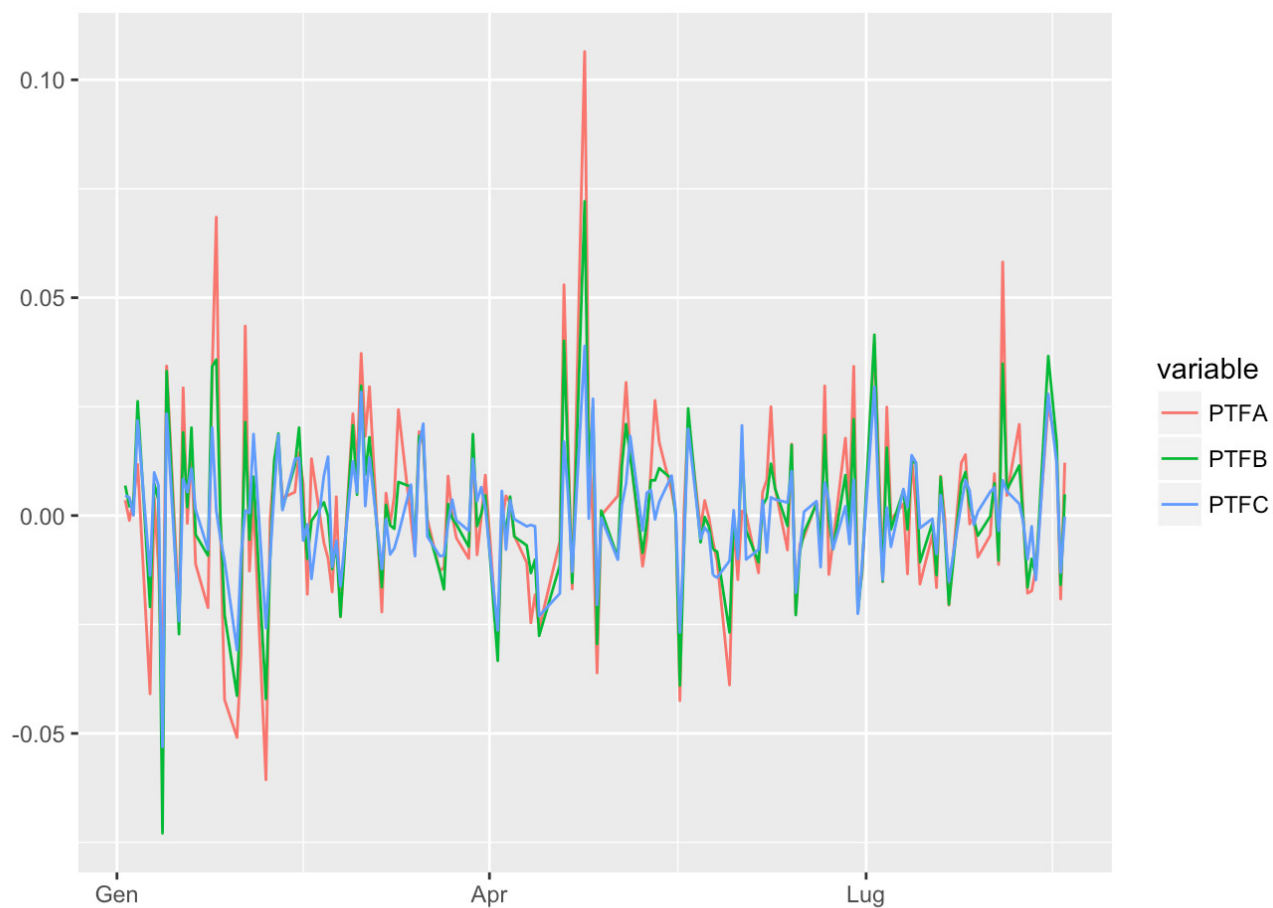
Performances

##	PTFA	PTFB	PTFC
## Min.	:-0.0606476	Min. :-0.0729155	Min. :-5.305e-02
## 1st Qu.	:-0.0112539	1st Qu.:-0.0083504	1st Qu.:-7.832e-03
## Median	:-0.0001263	Median :-0.0002382	Median : 8.458e-04
## Mean	: 0.0013461	Mean : 0.0009092	Mean :-6.894e-05
## 3rd Qu.	: 0.0130991	3rd Qu.: 0.0097409	3rd Qu.: 6.163e-03
## Max.	: 0.1064748	Max. : 0.0721205	Max. : 3.890e-02

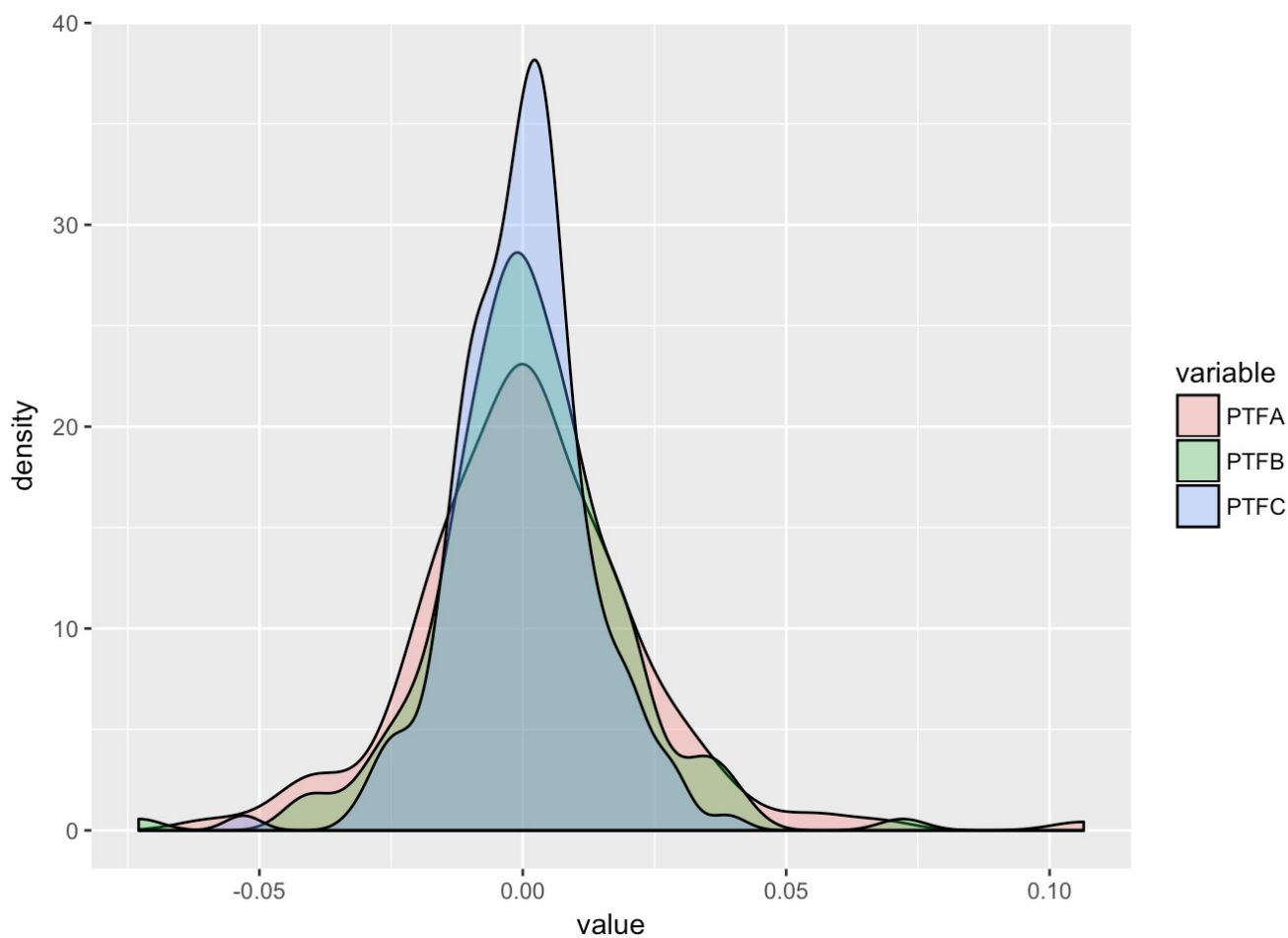
ggplot2

Rendimenti: serie storiche, distribuzioni e correlazioni

```
ggplot (ggdata, aes (x=date, y=value)) + geom_line (aes (colour=variable)) + xlab ('') + ylab ('')
```

```
ggplot (ggdata, aes (x=value, y=..density.., group=variable)) + geom_density (aes (fill=variable), alpha=0.3)
```



```
ggpairs(data.frame(returns(compareptf)))
```

