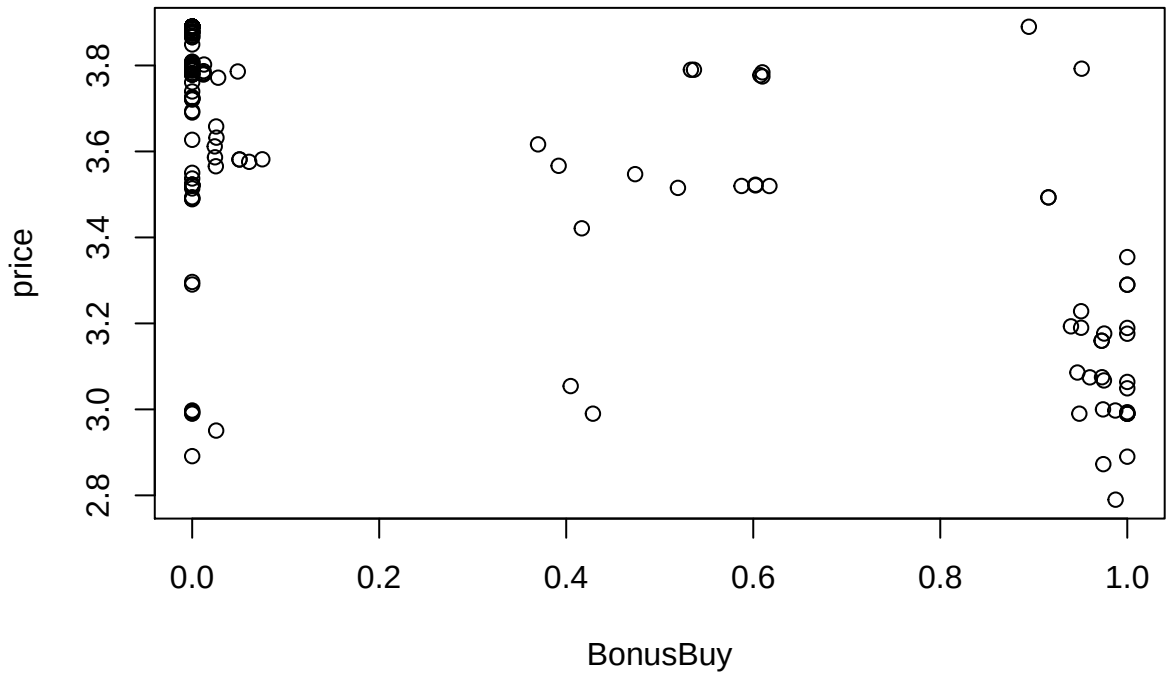
Shadata1995

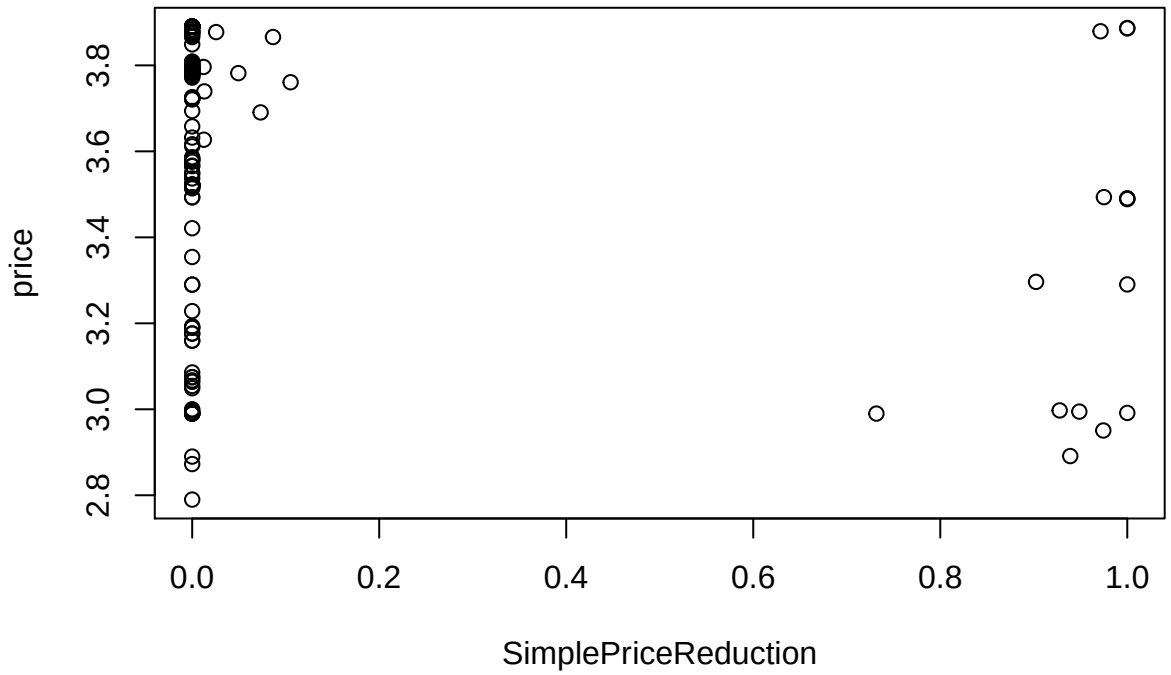
Dieudonne Ouedraogo

3/15/2018

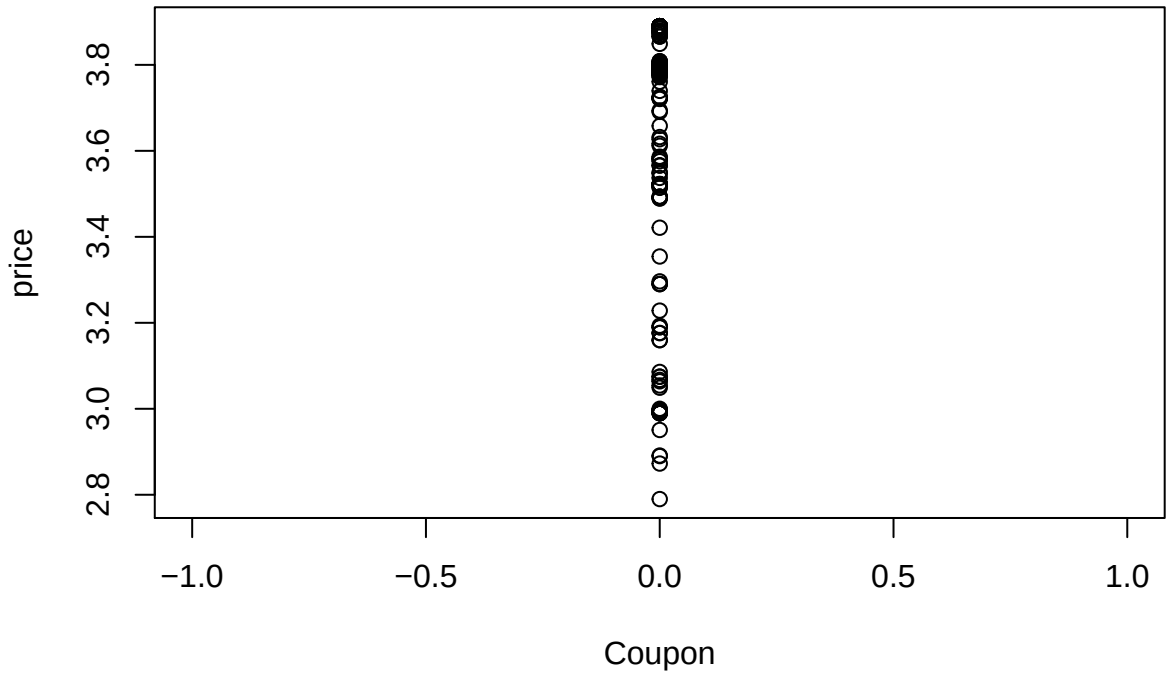
```
library(sas7bdat)
library(knitr)
library(pander)
#library(sqldf)
library(dplyr)
X=read.sas7bdat("~/Downloads/shadata.sas7bdat",debug=TRUE)
#pander(head(X,5))
#sha1995=filter(X,Year=='1995')
#kable(head(sha1995,15))
#kable(paste("number of row",nrow(sha1995)))
#pander(head(sha1995))
shaBestUpc=filter(X,UPC=='3700000089')
shadata <- read.csv("~/shadata.csv")
#shaBestUpc
plot(data=shaBestUpc,price~BonusBuy)
```




```
plot(data=shaBestUpc,price~SimplePriceReduction)
```




```
plot(data=shaBestUpc,price~Coupon)
```




```
#Coupon has nothing to do with this product
fit=lm(data = shaBestUpc,price~SimplePriceReduction+BonusBuy)
fit
```

```
##
## Call:
## lm(formula = price ~ SimplePriceReduction + BonusBuy, data = shaBestUpc)
##
## Coefficients:
##          (Intercept)  SimplePriceReduction              BonusBuy
```

```
##                3.7944                -0.4648                -0.6324
```

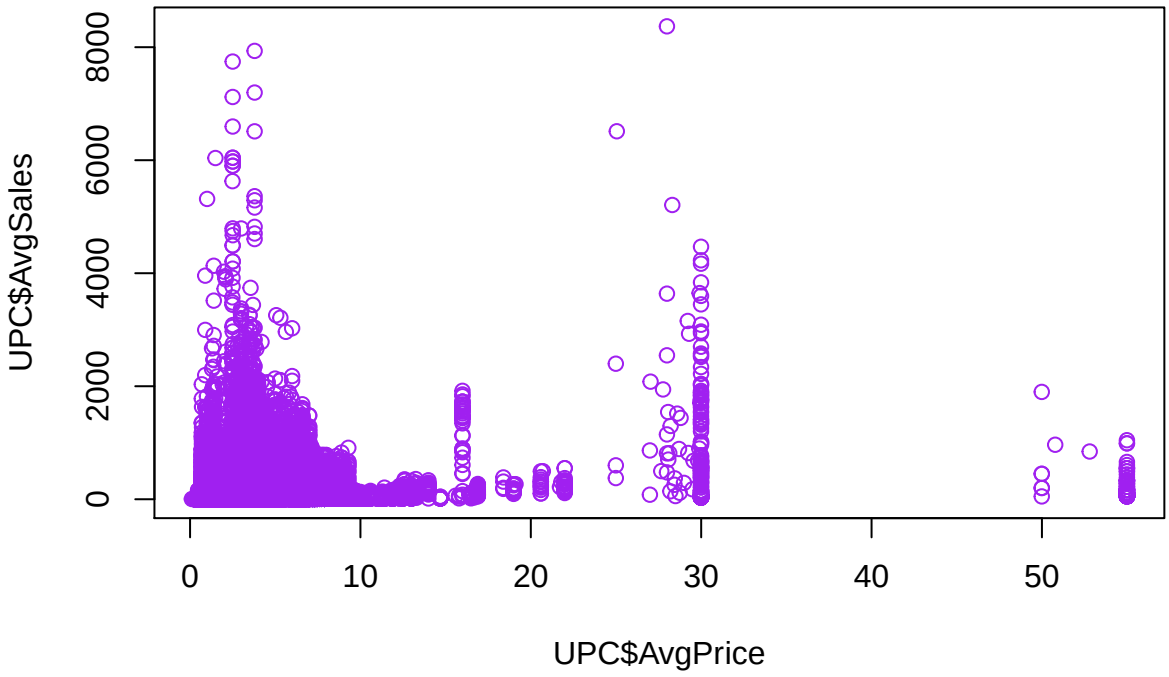
```
#plot(fit)

byUPC<-shadata%>%group_by(UPC,WEEK)
UPC=summarize(byUPC,count=n(),AvgPrice= mean(price, na.rm = T),AvgBonus= mean(BonusBuy, na.rm=T),AvgSim
#write.csv(UPC,file='UPC_WEEK.csv')
#nrow(UPC)

byUPC<-shadata%>%group_by(UPC,WEEK)
UPC=summarize(byUPC,count=n(),AvgSales= mean(sales, na.rm = T),AvgPrice= mean(price, na.rm = T),AvgBonus
#write.csv(UPC,file='UPC_WEEK.csv')
nrow(UPC)
```

```
## [1] 217078
```

```
kable(head(UPC,10))
```

UPC	WEEK	count	AvgSales	AvgPrice	AvgBonus	AvgSimplePrice	AvgCoupon
5690310	351	1	25.00	2.50	0.125	0.875	0
5690400	351	1	15.00	2.50	0.250	0.750	0
370071913	147	1	1.89	1.89	0.000	0.000	0
370071913	148	1	1.49	1.49	0.000	1.000	0
370071913	149	1	3.78	1.89	0.000	0.000	0
521328700	128	1	1.69	1.69	0.000	0.000	0
521328700	131	1	0.99	0.99	0.000	1.000	0
521328700	132	1	1.69	1.69	0.000	0.000	0
521328700	137	1	1.39	1.39	0.000	0.000	0
521328700	156	1	1.59	1.59	0.000	0.000	0

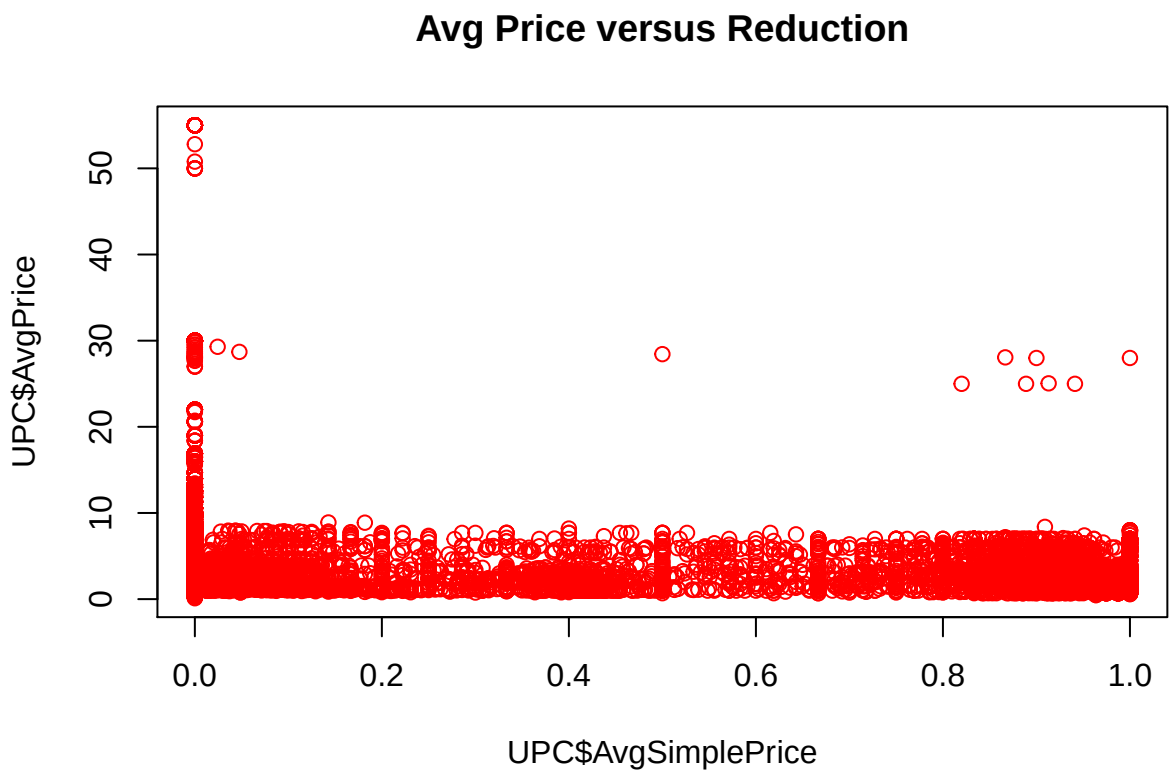
```
plot(UPC$AvgPrice,UPC$AvgSales,col="purple",main="Avg Price versus Sales")
```



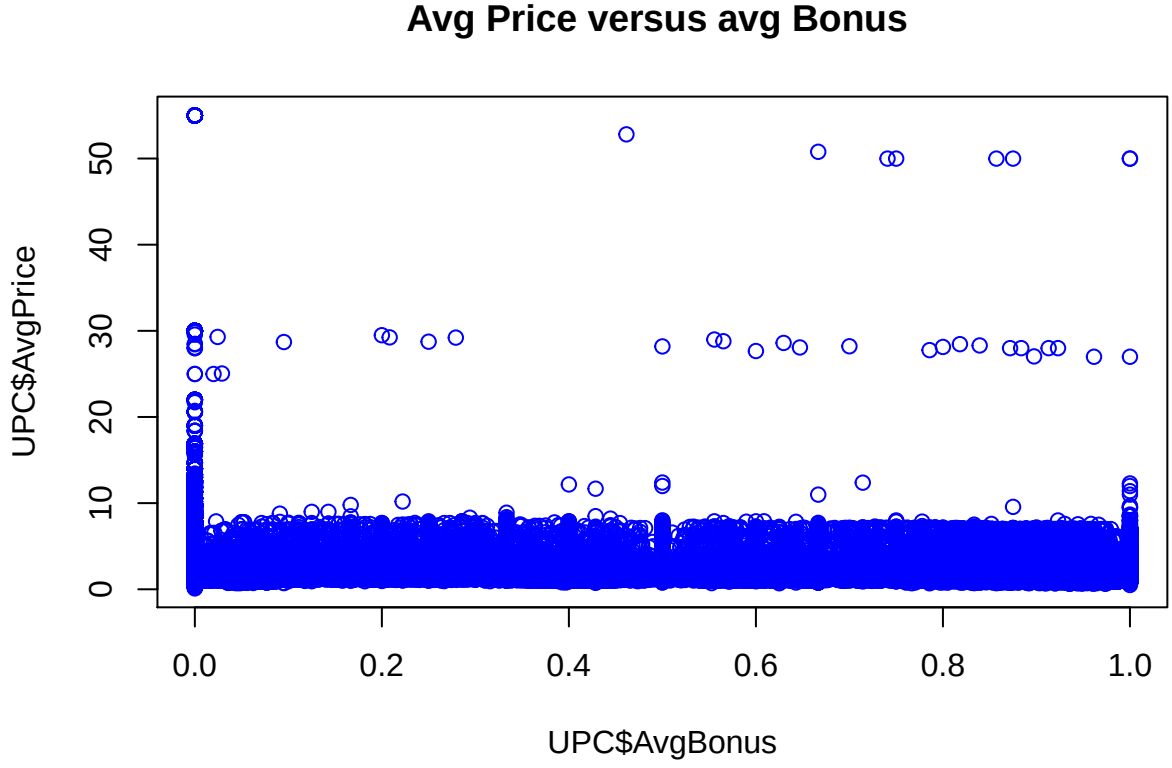

```
kable(head(UPC,10))
```

UPC	WEEK	count	AvgSales	AvgPrice	AvgBonus	AvgSimplePrice	AvgCoupon
5690310	351	1	25.00	2.50	0.125	0.875	0
5690400	351	1	15.00	2.50	0.250	0.750	0
370071913	147	1	1.89	1.89	0.000	0.000	0
370071913	148	1	1.49	1.49	0.000	1.000	0
370071913	149	1	3.78	1.89	0.000	0.000	0
521328700	128	1	1.69	1.69	0.000	0.000	0
521328700	131	1	0.99	0.99	0.000	1.000	0
521328700	132	1	1.69	1.69	0.000	0.000	0
521328700	137	1	1.39	1.39	0.000	0.000	0
521328700	156	1	1.59	1.59	0.000	0.000	0

```
plot(UPC$AvgSimplePrice,UPC$AvgPrice,col="red",main="Avg Price versus Reduction")
```




```
plot(UPC$AvgBonus,UPC$AvgPrice,col="blue",main="Avg Price versus avg Bonus ")
```



By quarter and year

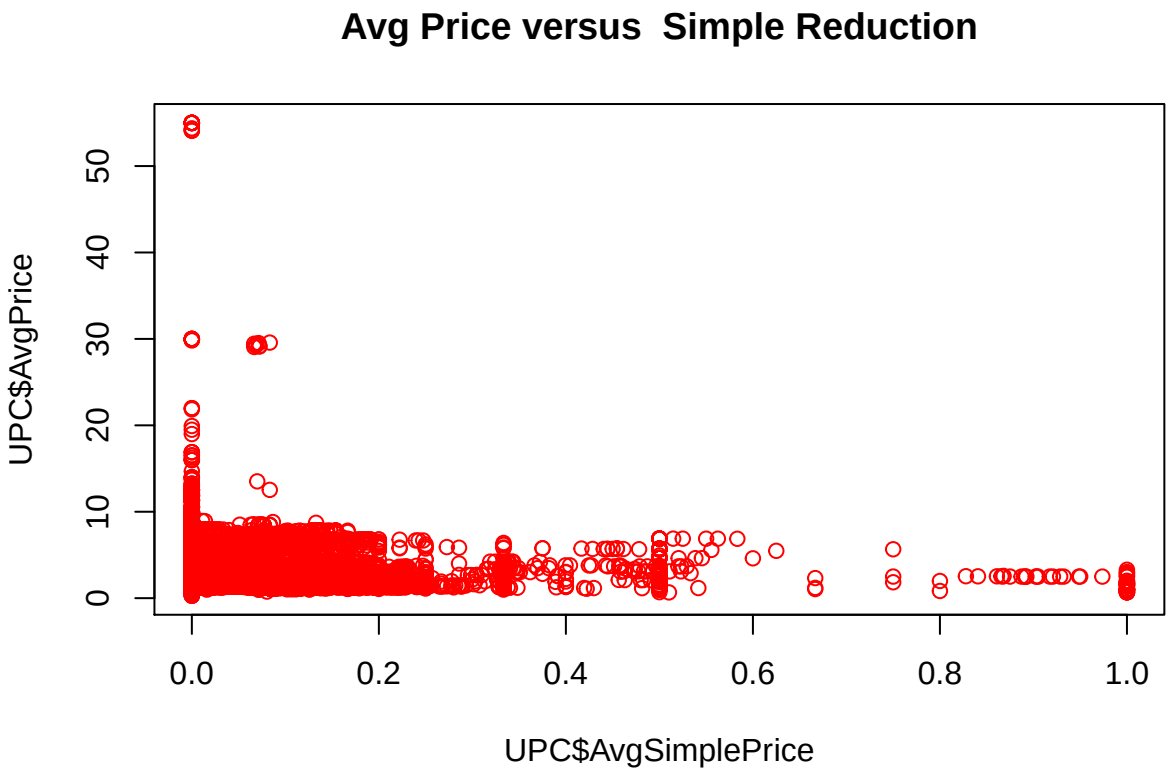
```
byUPC<-shadata%>%group_by(UPC,Quarter,Year)
UPC=summarize(byUPC,count=n(),AvgPrice= mean(price, na.rm = T),AvgBonus= mean(BonusBuy, na.rm =T),AvgSi
write.csv(UPC,file='UPC_Quarter_Year.csv')
nrow(UPC)
```

```
## [1] 23820
```

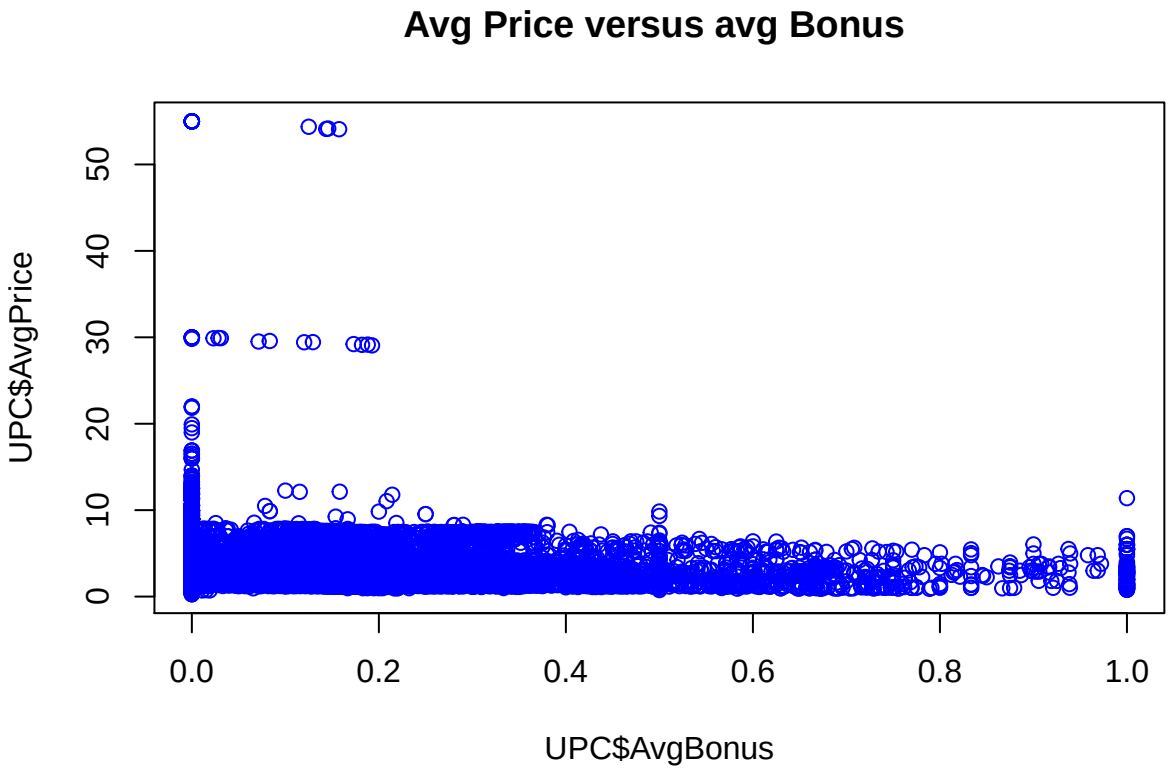
```
#UPC<- summarize(byUPC, count = n(), AvgSales = mean(sales, na.rm = F),TotalSales=sum(sales, na.rm = F)
kable(head(UPC,10))
```

UPC	Quarter	Year	count	AvgPrice	AvgBonus	AvgSimplePrice	AvgCoupon
5690310	2	1996	1	2.500000	0.125	0.8750000	0
5690400	2	1996	1	2.500000	0.250	0.7500000	0
370071913	3	1992	3	1.756667	0.000	0.3333333	0
521328700	1	1992	3	1.456667	0.000	0.3333333	0
521328700	1	1993	2	0.750000	0.500	0.0000000	0
521328700	1	1995	1	0.690000	0.000	1.0000000	0
521328700	2	1992	1	1.390000	0.000	0.0000000	0
521328700	3	1992	2	1.290000	0.500	0.0000000	0
521328700	4	1992	2	1.240000	0.500	0.0000000	0
521328700	4	1994	1	0.690000	0.000	1.0000000	0

```
plot(UPC$AvgSimplePrice,UPC$AvgPrice,col="Red",main="Avg Price versus  Simple Reduction")
```




```
plot(UPC$AvgBonus,UPC$AvgPrice,col="blue",main="Avg Price versus avg Bonus ")
```

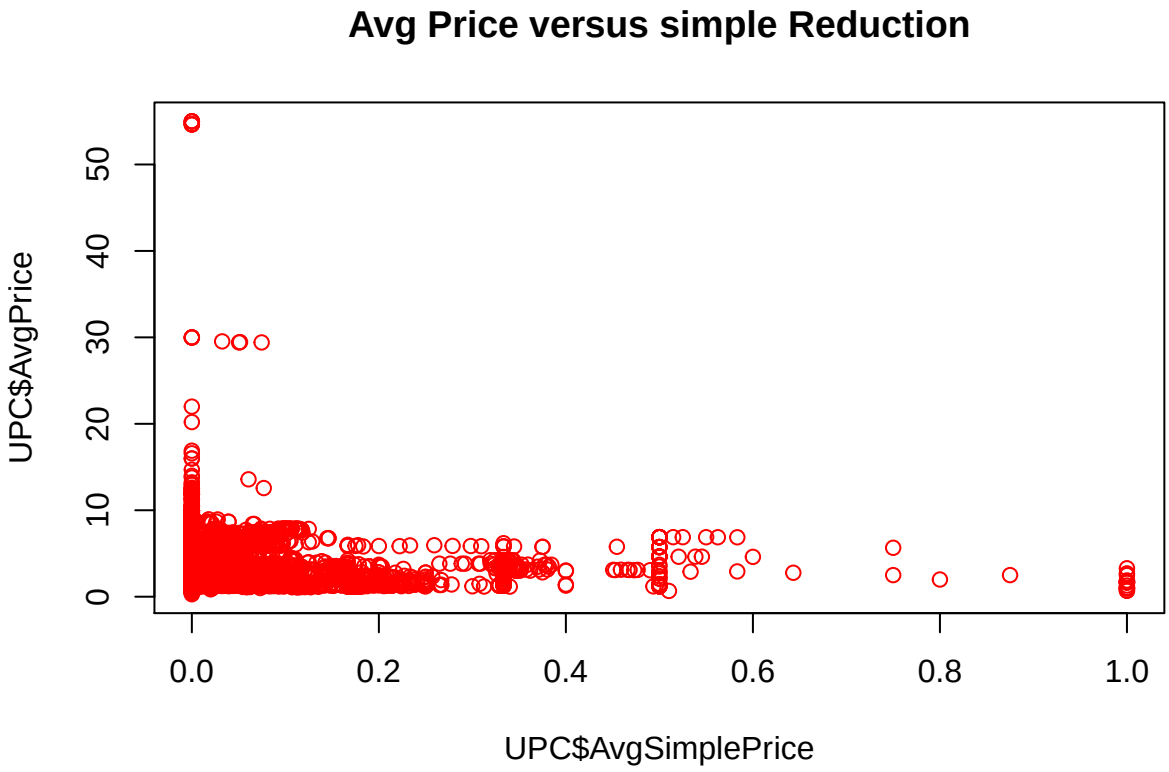


By year

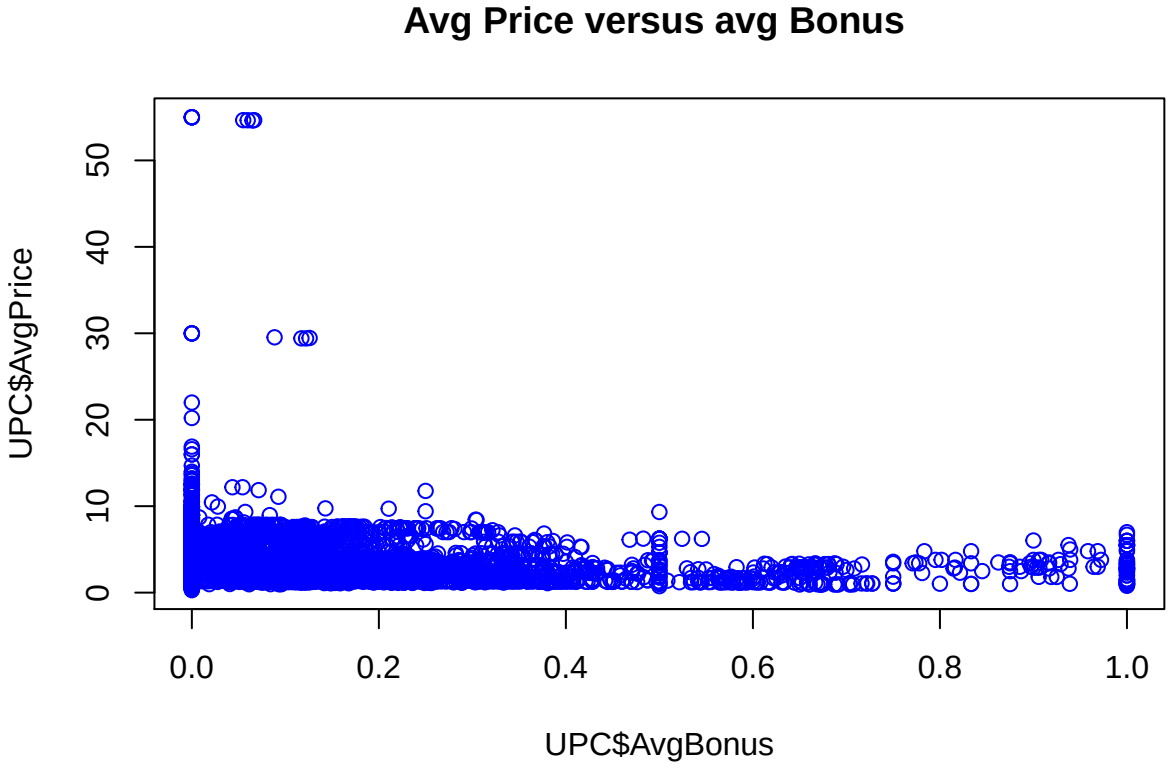
```
byUPC<-shadata%>%group_by(UPC,Year)
UPC=summarize(byUPC,count=n(),AvgPrice= mean(price, na.rm = T),AvgBonus= mean(BonusBuy, na.rm = T),AvgS

#UPC<- summarize(byUPC, count = n(), AvgSales = mean(sales, na.rm = F),TotalSales=sum(sales, na.rm = F)
kable(head(UPC,10))
```

UPC	Year	count	AvgPrice	AvgBonus	AvgSimplePrice	AvgCoupon
5690310	1996	1	2.500000	0.1250000	0.8750000	0
5690400	1996	1	2.500000	0.2500000	0.7500000	0
370071913	1992	3	1.756667	0.0000000	0.3333333	0
521328700	1992	8	1.352500	0.2500000	0.1250000	0
521328700	1993	2	0.750000	0.5000000	0.0000000	0
521328700	1994	1	0.690000	0.0000000	1.0000000	0
521328700	1995	1	0.690000	0.0000000	1.0000000	0
521346000	1992	9	1.253556	0.2222222	0.1666667	0
521346000	1993	4	1.190000	0.5000000	0.2500000	0
1150900201	1992	46	6.039589	0.0712085	0.0000000	0

```
plot(UPC$AvgSimplePrice,UPC$AvgPrice,col="Red",main="Avg Price versus simple Reduction")
```




```
#write.csv(UPC,file='UPC_Year.csv')

plot(UPC$AvgBonus,UPC$AvgPrice,col="blue",main="Avg Price versus avg Bonus ")
```



```r
#nrow(UPC)

plot(UPC$AvgCoupon,UPC$AvgPrice,col="green",main="Avg Price versus Coupon ")
```