Assignment 6

Chitresh Kumar

if(!require("pacman")) install.packages("pacman")  
pacman::p\_load(tidyverse, reshape, gplots, ggmap, RStata,haven,  
 data.table,margins,pastecs,MASS,tinytex)  
search()  
theme\_set(theme\_light())

#getwd()  
br<-read\_dta('br2.dta')  
head(br)

## # A tibble: 6 x 10  
## price sqft bedrooms baths age owner pool traditional fireplace waterfront  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 66500 741 1 1 18 1 1 1 1 0  
## 2 66000 741 1 1 18 0 1 1 0 0  
## 3 68500 790 1 1 18 1 0 1 1 0  
## 4 102000 2783 2 2 18 1 0 1 1 0  
## 5 54000 1165 2 1 35 0 0 1 0 0  
## 6 143000 2331 2 2 25 1 0 1 1 0

## PART A

options(scipen = 1)  
summary(br)

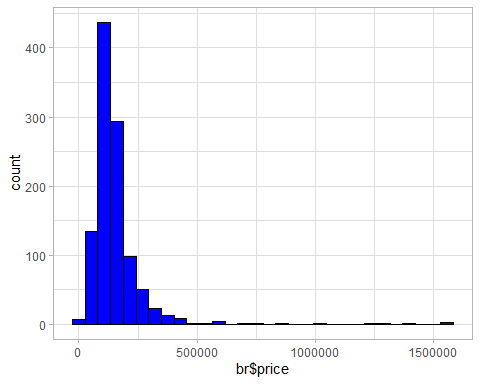
## price sqft bedrooms baths   
## Min. : 22000 Min. : 662 Min. :1.00 Min. :1.000   
## 1st Qu.: 99000 1st Qu.:1604 1st Qu.:3.00 1st Qu.:2.000   
## Median : 130000 Median :2186 Median :3.00 Median :2.000   
## Mean : 154863 Mean :2326 Mean :3.18 Mean :1.973   
## 3rd Qu.: 170163 3rd Qu.:2800 3rd Qu.:4.00 3rd Qu.:2.000   
## Max. :1580000 Max. :7897 Max. :8.00 Max. :5.000   
## age owner pool traditional   
## Min. : 1.00 Min. :0.0000 Min. :0.00000 Min. :0.0000   
## 1st Qu.: 5.00 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000   
## Median :18.00 Median :0.0000 Median :0.00000 Median :1.0000   
## Mean :19.57 Mean :0.4889 Mean :0.07963 Mean :0.5389   
## 3rd Qu.:25.00 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:1.0000   
## Max. :80.00 Max. :1.0000 Max. :1.00000 Max. :1.0000   
## fireplace waterfront   
## Min. :0.000 Min. :0.00000   
## 1st Qu.:0.000 1st Qu.:0.00000   
## Median :1.000 Median :0.00000   
## Mean :0.563 Mean :0.07222   
## 3rd Qu.:1.000 3rd Qu.:0.00000   
## Max. :1.000 Max. :1.00000

stat.desc(br)

## price sqft bedrooms baths  
## nbr.val 1.080000e+03 1.080000e+03 1080.00000000 1080.00000000  
## nbr.null 0.000000e+00 0.000000e+00 0.00000000 0.00000000  
## nbr.na 0.000000e+00 0.000000e+00 0.00000000 0.00000000  
## min 2.200000e+04 6.620000e+02 1.00000000 1.00000000  
## max 1.580000e+06 7.897000e+03 8.00000000 5.00000000  
## range 1.558000e+06 7.235000e+03 7.00000000 4.00000000  
## sum 1.672522e+08 2.512013e+06 3434.00000000 2131.00000000  
## median 1.300000e+05 2.186500e+03 3.00000000 2.00000000  
## mean 1.548632e+05 2.325938e+03 3.17962963 1.97314815  
## SE.mean 3.740118e+03 3.067544e+01 0.02158927 0.01862460  
## CI.mean.0.95 7.338728e+03 6.019028e+01 0.04236172 0.03654454  
## var 1.510756e+10 1.016262e+06 0.50338448 0.37462585  
## std.dev 1.229128e+05 1.008098e+03 0.70949593 0.61206687  
## coef.var 7.936865e-01 4.334157e-01 0.22313792 0.31019813  
## age owner pool traditional  
## nbr.val 1080.0000000 1080.00000000 1.080000e+03 1080.00000000  
## nbr.null 0.0000000 552.00000000 9.940000e+02 498.00000000  
## nbr.na 0.0000000 0.00000000 0.000000e+00 0.00000000  
## min 1.0000000 0.00000000 0.000000e+00 0.00000000  
## max 80.0000000 1.00000000 1.000000e+00 1.00000000  
## range 79.0000000 1.00000000 1.000000e+00 1.00000000  
## sum 21140.0000000 528.00000000 8.600000e+01 582.00000000  
## median 18.0000000 0.00000000 0.000000e+00 1.00000000  
## mean 19.5740741 0.48888889 7.962963e-02 0.53888889  
## SE.mean 0.5232045 0.01521781 8.241532e-03 0.01517545  
## CI.mean.0.95 1.0266135 0.02985984 1.617125e-02 0.02977674  
## var 295.6423300 0.25010812 7.335667e-02 0.24871795  
## std.dev 17.1942528 0.50010811 2.708444e-01 0.49871630  
## coef.var 0.8784197 1.02294841 3.401301e+00 0.92545293  
## fireplace waterfront  
## nbr.val 1080.00000000 1.080000e+03  
## nbr.null 472.00000000 1.002000e+03  
## nbr.na 0.00000000 0.000000e+00  
## min 0.00000000 0.000000e+00  
## max 1.00000000 1.000000e+00  
## range 1.00000000 1.000000e+00  
## sum 608.00000000 7.800000e+01  
## median 1.00000000 0.000000e+00  
## mean 0.56296296 7.222222e-02  
## SE.mean 0.01510040 7.880371e-03  
## CI.mean.0.95 0.02962947 1.546259e-02  
## var 0.24626369 6.706827e-02  
## std.dev 0.49624962 2.589754e-01  
## coef.var 0.88149604 3.585814e+00

*HISTOGRAM OF PRICE* #warning=False,message=False

ggplot(data=br, aes(x=br$price)) +   
 geom\_histogram(color='black', fill='blue')



**PART B**

PRICE=br$price/1000  
SQFT=br$sqft/100  
lm\_b<-lm(log(PRICE)~SQFT+age+bedrooms+baths+owner+pool+traditional+fireplace+waterfront,data=br)  
summary(lm\_b)

##   
## Call:  
## lm(formula = log(PRICE) ~ SQFT + age + bedrooms + baths + owner +   
## pool + traditional + fireplace + waterfront, data = br)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.13459 -0.12758 0.00656 0.14785 1.06650   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.9808326 0.0458947 86.738 < 2e-16 \*\*\*  
## SQFT 0.0299011 0.0014059 21.269 < 2e-16 \*\*\*  
## age -0.0062145 0.0005179 -11.999 < 2e-16 \*\*\*  
## bedrooms -0.0315060 0.0166109 -1.897 0.058135 .   
## baths 0.1901190 0.0205579 9.248 < 2e-16 \*\*\*  
## owner 0.0674654 0.0177460 3.802 0.000152 \*\*\*  
## pool -0.0042748 0.0315812 -0.135 0.892353   
## traditional -0.0560926 0.0170267 -3.294 0.001019 \*\*   
## fireplace 0.0842748 0.0190150 4.432 0.0000103 \*\*\*  
## waterfront 0.1099700 0.0333550 3.297 0.001010 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.27 on 1070 degrees of freedom  
## Multiple R-squared: 0.7373, Adjusted R-squared: 0.7351   
## F-statistic: 333.7 on 9 and 1070 DF, p-value: < 2.2e-16

**PART C**

lm\_c<-lm(log(PRICE)~SQFT+age+bedrooms+baths+owner+pool+traditional+fireplace+waterfront+waterfront : traditional ,data=br)  
summary(lm\_c)

##   
## Call:  
## lm(formula = log(PRICE) ~ SQFT + age + bedrooms + baths + owner +   
## pool + traditional + fireplace + waterfront + waterfront:traditional,   
## data = br)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.13891 -0.12591 0.00672 0.14693 1.05734   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.9711130 0.0459460 86.430 < 2e-16 \*\*\*  
## SQFT 0.0300308 0.0014034 21.399 < 2e-16 \*\*\*  
## age -0.0061470 0.0005174 -11.881 < 2e-16 \*\*\*  
## bedrooms -0.0313330 0.0165702 -1.891 0.05890 .   
## baths 0.1882577 0.0205208 9.174 < 2e-16 \*\*\*  
## owner 0.0683701 0.0177061 3.861 0.00012 \*\*\*  
## pool -0.0023939 0.0315125 -0.076 0.93946   
## traditional -0.0449127 0.0175612 -2.557 0.01068 \*   
## fireplace 0.0873139 0.0190070 4.594 4.87e-06 \*\*\*  
## waterfront 0.1653741 0.0399505 4.139 3.75e-05 \*\*\*  
## traditional:waterfront -0.1721747 0.0687162 -2.506 0.01237 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2693 on 1069 degrees of freedom  
## Multiple R-squared: 0.7389, Adjusted R-squared: 0.7364   
## F-statistic: 302.5 on 10 and 1069 DF, p-value: < 2.2e-16

**PART D**

lm\_d1<-lm(log(PRICE)~SQFT+age+bedrooms+baths+owner+pool+fireplace+waterfront,data=br)  
summary(lm\_d1)

##   
## Call:  
## lm(formula = log(PRICE) ~ SQFT + age + bedrooms + baths + owner +   
## pool + fireplace + waterfront, data = br)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.15673 -0.12355 -0.00287 0.14356 1.03816   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.9701078 0.0459892 86.327 < 2e-16 \*\*\*  
## SQFT 0.0301592 0.0014101 21.387 < 2e-16 \*\*\*  
## age -0.0061907 0.0005203 -11.899 < 2e-16 \*\*\*  
## bedrooms -0.0405182 0.0164592 -2.462 0.013984 \*   
## baths 0.1894469 0.0206512 9.174 < 2e-16 \*\*\*  
## owner 0.0650077 0.0178117 3.650 0.000275 \*\*\*  
## pool 0.0007741 0.0316887 0.024 0.980516   
## fireplace 0.0911987 0.0189852 4.804 1.78e-06 \*\*\*  
## waterfront 0.1225762 0.0332869 3.682 0.000243 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2712 on 1071 degrees of freedom  
## Multiple R-squared: 0.7347, Adjusted R-squared: 0.7327   
## F-statistic: 370.7 on 8 and 1071 DF, p-value: < 2.2e-16

lm\_d2<-lm(log(PRICE)~SQFT+age+bedrooms+baths+owner+pool+fireplace+waterfront+traditional+SQFT:traditional+age:traditional+bedrooms:traditional+baths :traditional+ owner:traditional+pool:traditional+fireplace:traditional+waterfront:traditional,data=br)  
summary(lm\_d2)

##   
## Call:  
## lm(formula = log(PRICE) ~ SQFT + age + bedrooms + baths + owner +   
## pool + fireplace + waterfront + traditional + SQFT:traditional +   
## age:traditional + bedrooms:traditional + baths:traditional +   
## owner:traditional + pool:traditional + fireplace:traditional +   
## waterfront:traditional, data = br)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.1376 -0.1248 0.0045 0.1462 1.0578   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.0672688 0.0576957 70.495 < 2e-16 \*\*\*  
## SQFT 0.0324010 0.0018412 17.598 < 2e-16 \*\*\*  
## age -0.0054674 0.0007292 -7.498 1.37e-13 \*\*\*  
## bedrooms -0.0713737 0.0236571 -3.017 0.002614 \*\*   
## baths 0.1831139 0.0288594 6.345 3.29e-10 \*\*\*  
## owner 0.0388479 0.0258967 1.500 0.133884   
## pool 0.0021253 0.0419397 0.051 0.959594   
## fireplace 0.0578017 0.0296703 1.948 0.051662 .   
## waterfront 0.1729789 0.0406915 4.251 2.32e-05 \*\*\*  
## traditional -0.3350839 0.0944926 -3.546 0.000408 \*\*\*  
## SQFT:traditional -0.0052974 0.0028196 -1.879 0.060549 .   
## age:traditional -0.0012916 0.0010325 -1.251 0.211211   
## bedrooms:traditional 0.0989064 0.0335594 2.947 0.003277 \*\*   
## baths:traditional 0.0310767 0.0412135 0.754 0.450991   
## owner:traditional 0.0586870 0.0353000 1.663 0.096703 .   
## pool:traditional -0.0237596 0.0630941 -0.377 0.706566   
## fireplace:traditional 0.0650471 0.0386865 1.681 0.092982 .   
## waterfront:traditional -0.2069886 0.0710609 -2.913 0.003657 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2672 on 1062 degrees of freedom  
## Multiple R-squared: 0.7447, Adjusted R-squared: 0.7406   
## F-statistic: 182.2 on 17 and 1062 DF, p-value: < 2.2e-16

anova(lm\_d1,lm\_d2)

## Analysis of Variance Table  
##   
## Model 1: log(PRICE) ~ SQFT + age + bedrooms + baths + owner + pool + fireplace +   
## waterfront  
## Model 2: log(PRICE) ~ SQFT + age + bedrooms + baths + owner + pool + fireplace +   
## waterfront + traditional + SQFT:traditional + age:traditional +   
## bedrooms:traditional + baths:traditional + owner:traditional +   
## pool:traditional + fireplace:traditional + waterfront:traditional  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 1071 78.772   
## 2 1062 75.799 9 2.9724 4.6272 5.037e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

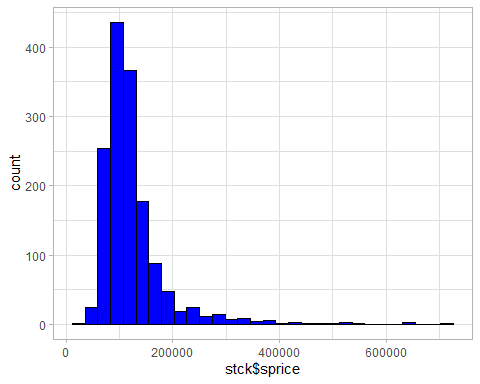
#Question 7.16

stck<-read\_dta('stckton4.dta')  
head(stck)

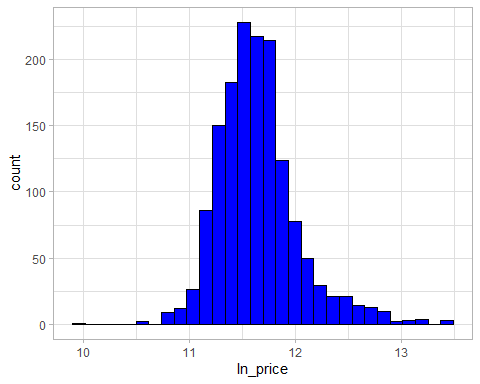
## # A tibble: 6 x 7  
## sprice livarea beds baths lgelot age pool  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 138000 17 3 2 1 97 0  
## 2 105700 21 4 2.5 0 18 0  
## 3 22000 7 2 1 0 49 0  
## 4 255000 30 3 3 1 23 0  
## 5 203000 21 4 2 1 18 0  
## 6 129178 16 3 2 0 2 0

**PART a** #Histogram

ggplot(data=stck, aes(x=stck$sprice)) +   
 geom\_histogram(color="black", fill="blue")

 #Histogram of Log(price)

ln\_price<-log(stck$sprice)  
ggplot(data=stck, aes(x=ln\_price)) +   
 geom\_histogram(color="black", fill="blue")



**PART b**

PRICE<-stck$sprice/1000  
lm\_b<-lm(log(PRICE)~livarea+beds+baths+age+lgelot+pool,data=stck)  
summary(lm\_b)

##   
## Call:  
## lm(formula = log(PRICE) ~ livarea + beds + baths + age + lgelot +   
## pool, data = stck)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29751 -0.11979 -0.00427 0.12671 2.00684   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.9859688 0.0373406 106.746 < 2e-16 \*\*\*  
## livarea 0.0539316 0.0017080 31.576 < 2e-16 \*\*\*  
## beds -0.0382209 0.0113593 -3.365 0.000786 \*\*\*  
## baths -0.0102729 0.0165268 -0.622 0.534309   
## age -0.0013113 0.0004601 -2.850 0.004433 \*\*   
## lgelot 0.2530908 0.0255382 9.910 < 2e-16 \*\*\*  
## pool 0.0786611 0.0230548 3.412 0.000662 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2128 on 1493 degrees of freedom  
## Multiple R-squared: 0.6884, Adjusted R-squared: 0.6871   
## F-statistic: 549.6 on 6 and 1493 DF, p-value: < 2.2e-16

**PART d**

lm\_d<-lm(log(PRICE)~livarea+beds+baths+age+lgelot+pool+lgelot:livarea,data=stck)  
summary(lm\_d)

##   
## Call:  
## lm(formula = log(PRICE) ~ livarea + beds + baths + age + lgelot +   
## pool + lgelot:livarea, data = stck)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.17288 -0.12284 -0.00263 0.12812 2.02143   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.964941 0.037033 107.064 < 2e-16 \*\*\*  
## livarea 0.058857 0.001864 31.582 < 2e-16 \*\*\*  
## beds -0.047996 0.011328 -4.237 2.41e-05 \*\*\*  
## baths -0.020062 0.016398 -1.223 0.221356   
## age -0.001612 0.000457 -3.527 0.000433 \*\*\*  
## lgelot 0.613440 0.063209 9.705 < 2e-16 \*\*\*  
## pool 0.085349 0.022795 3.744 0.000188 \*\*\*  
## livarea:lgelot -0.016125 0.002593 -6.217 6.55e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2102 on 1492 degrees of freedom  
## Multiple R-squared: 0.6962, Adjusted R-squared: 0.6948   
## F-statistic: 488.5 on 7 and 1492 DF, p-value: < 2.2e-16

**PART e**

#Restricted Model  
lm\_e1<-lm(log(PRICE)~livarea+beds+baths+age+pool,data=stck)  
summary(lm\_e1)

##   
## Call:  
## lm(formula = log(PRICE) ~ livarea + beds + baths + age + pool,   
## data = stck)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.16849 -0.13118 -0.01003 0.12675 2.00675   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.9794107 0.0385303 103.280 < 2e-16 \*\*\*  
## livarea 0.0606975 0.0016157 37.567 < 2e-16 \*\*\*  
## beds -0.0594013 0.0115137 -5.159 2.81e-07 \*\*\*  
## baths -0.0262415 0.0169748 -1.546 0.1223   
## age -0.0007805 0.0004716 -1.655 0.0981 .   
## pool 0.0989178 0.0236994 4.174 3.17e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2196 on 1494 degrees of freedom  
## Multiple R-squared: 0.6679, Adjusted R-squared: 0.6667   
## F-statistic: 600.8 on 5 and 1494 DF, p-value: < 2.2e-16

#Unrestricted model  
lm\_e2<-lm(log(PRICE)~livarea+beds+baths+age+lgelot+pool+lgelot:livarea+lgelot:beds+lgelot:baths+lgelot:age+lgelot:pool,data=stck)  
summary(lm\_e2)

##   
## Call:  
## lm(formula = log(PRICE) ~ livarea + beds + baths + age + lgelot +   
## pool + lgelot:livarea + lgelot:beds + lgelot:baths + lgelot:age +   
## lgelot:pool, data = stck)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.09828 -0.12100 -0.00141 0.12783 2.02787   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.982753 0.038386 103.755 < 2e-16 \*\*\*  
## livarea 0.060383 0.001925 31.365 < 2e-16 \*\*\*  
## beds -0.052190 0.011950 -4.368 1.34e-05 \*\*\*  
## baths -0.033442 0.017394 -1.923 0.054714 .   
## age -0.001598 0.000484 -3.301 0.000986 \*\*\*  
## lgelot 0.429324 0.140851 3.048 0.002344 \*\*   
## pool 0.069685 0.025131 2.773 0.005627 \*\*   
## livarea:lgelot -0.026640 0.004325 -6.159 9.39e-10 \*\*\*  
## beds:lgelot 0.043412 0.037391 1.161 0.245819   
## baths:lgelot 0.116104 0.051893 2.237 0.025409 \*   
## age:lgelot -0.000219 0.001447 -0.151 0.879738   
## lgelot:pool 0.056183 0.060423 0.930 0.352616   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2098 on 1488 degrees of freedom  
## Multiple R-squared: 0.6982, Adjusted R-squared: 0.696   
## F-statistic: 313 on 11 and 1488 DF, p-value: < 2.2e-16

anova(lm\_e1,lm\_e2)

## Analysis of Variance Table  
##   
## Model 1: log(PRICE) ~ livarea + beds + baths + age + pool  
## Model 2: log(PRICE) ~ livarea + beds + baths + age + lgelot + pool + lgelot:livarea +   
## lgelot:beds + lgelot:baths + lgelot:age + lgelot:pool  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 1494 72.063   
## 2 1488 65.471 6 6.5921 24.97 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

