eda-lab-assignment-in-r.R

nzcoo

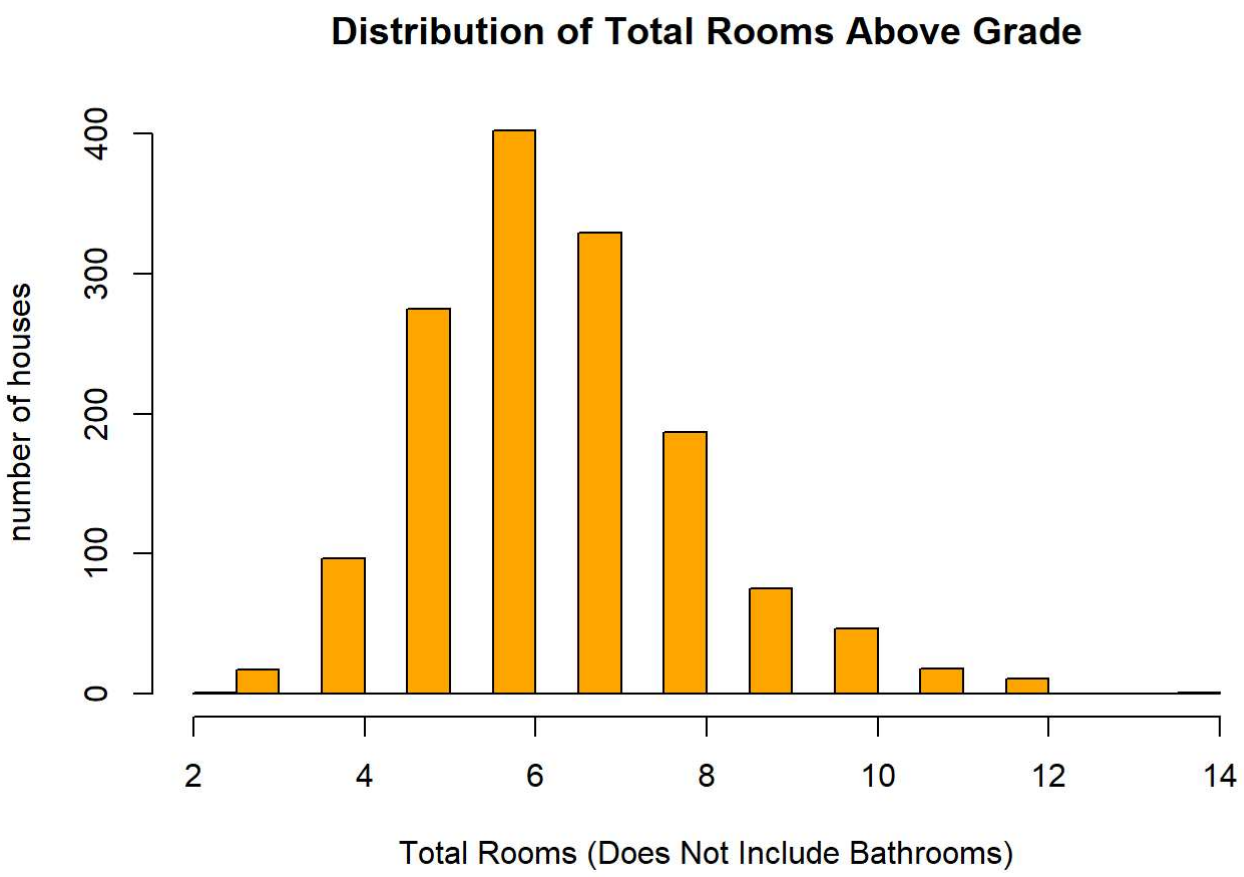
2022-05-15

```
ames <- read.table(file = "clipboard",
                      sep = "\t", header=TRUE)
hist(ames$SalePrice,main="Distribution of Sale Prices", xlab="Sale Price",ylab = "number of hous
es", col="orange", breaks= 40)
```


```
mean(ames[["SalePrice"]]) -> spm
median(ames[["SalePrice"]]) -> spmd
sd(ames[["SalePrice"]]) -> sps
print("mean 181k, median 163k, standard deviation 79k.")
```

```
## [1] "mean 181k, median 163k, standard deviation 79k."
```

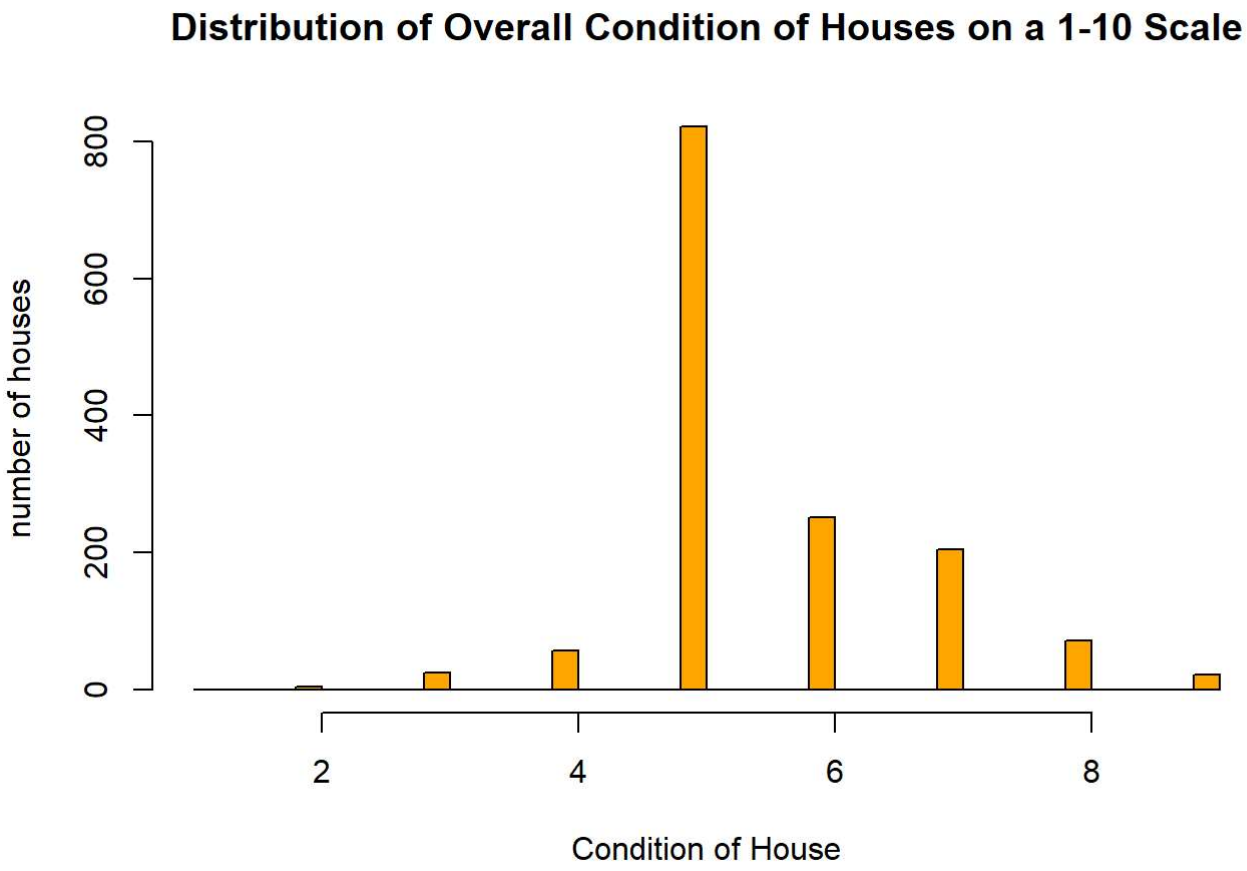
```
hist(ames$TotRmsAbvGrd,main="Distribution of Total Rooms Above Grade", xlab="Total Rooms (Does N
ot Include Bathrooms)",ylab = "number of houses", col="orange", breaks= 40)
```




```
mean(ames[["TotRmsAbvGrd"]]) -> tragm
median(ames[["TotRmsAbvGrd"]]) -> tragmd
sd(ames[["TotRmsAbvGrd"]]) -> trags
print("mean 181k, median 163k, standard deviation 79k.")
```

```
## [1] "mean 181k, median 163k, standard deviation 79k."
```

```
hist(ames$OverallCond,main="Distribution of Overall Condition of Houses on a 1-10 Scale", xlab=
"Condition of House",ylab = "number of houses", col="orange", breaks= 40)
```




```
mean(ames[["OverallCond"]]) -> ocm
median(ames[["OverallCond"]]) -> ocmd
sd(ames[["OverallCond"]]) -> ocs
print("mean 6.52, median 6, standard deviation 1.63.")
```

```
## [1] "mean 6.52, median 6, standard deviation 1.63."
```

```
belowavg <- subset(ames,OverallCond <5)
avgavg <- subset(ames,OverallCond ==5)
aboveavg <- subset(ames,OverallCond >5)


plot(density(belowavg$SalePrice),col='red')
lines(density(avgavg$SalePrice),col='blue')
lines(density(aboveavg$SalePrice),col='green')

print("In the density graph we can see that the below average conditioned houses (red) mostly se
ll for a lower amount than the other groups.  This group has been sold more than the above avera
ge(green) and average (blue) groups.  There are less average homes sold than any other type of h
ouse. it is suprising that there are not more average homes sold.")
```

```
## [1] "In the density graph we can see that the below average conditioned houses (red) mostly s
ell for a lower amount than the other groups.  This group has been sold more than the above aver
age(green) and average (blue) groups.  There are less average homes sold than any other type of
house. it is suprising that there are not more average homes sold."
```

```
library("dplyr")
```

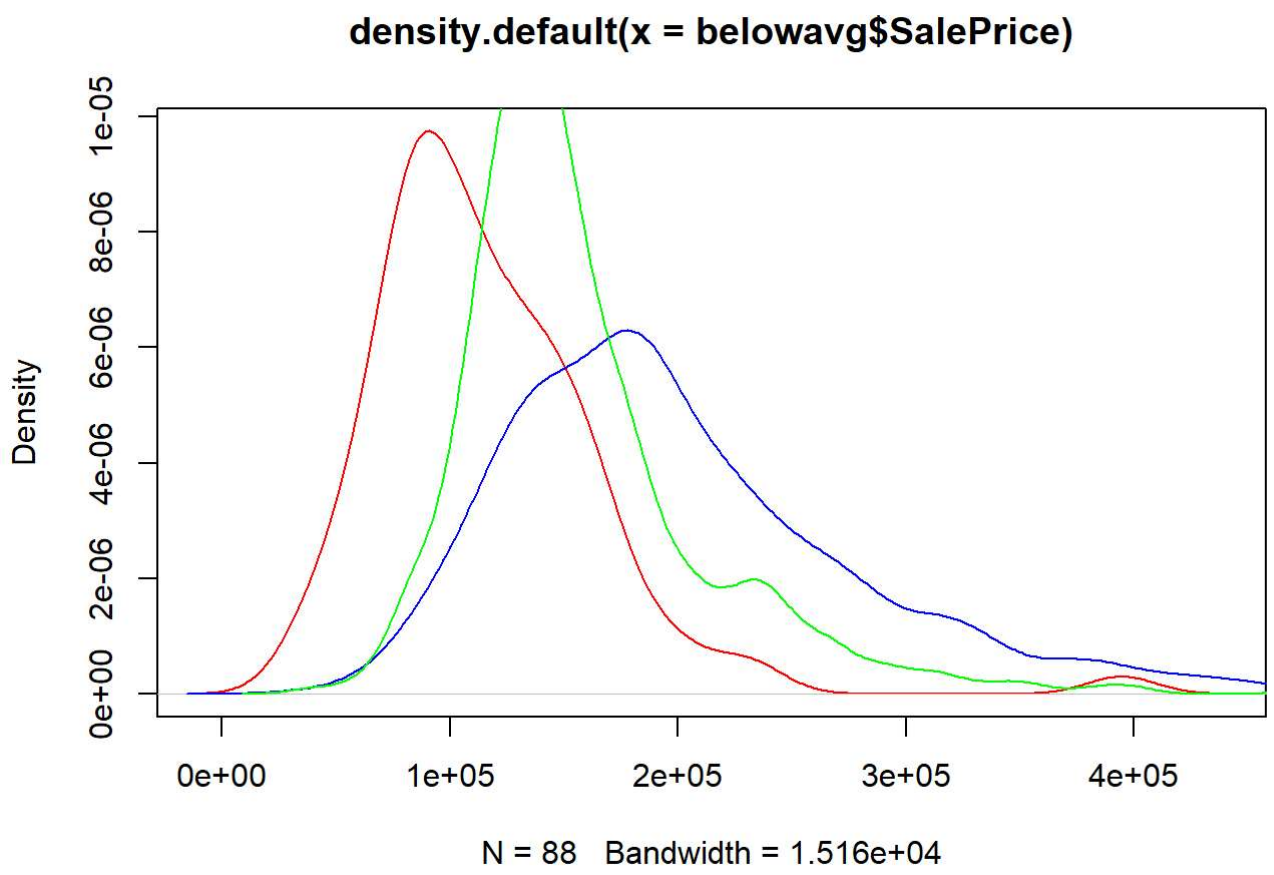
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

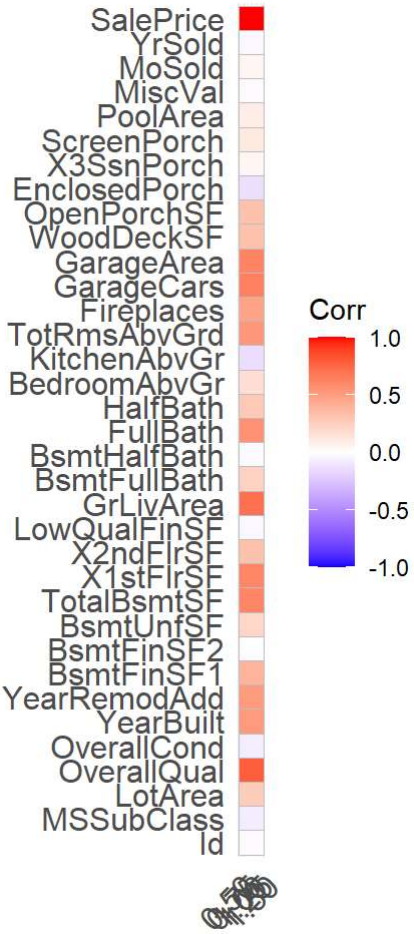
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
numbers <- select_if(ames, is.numeric)

correlation <-cor(numbers$SalePrice,numbers)

library(ggcorrplot)
```

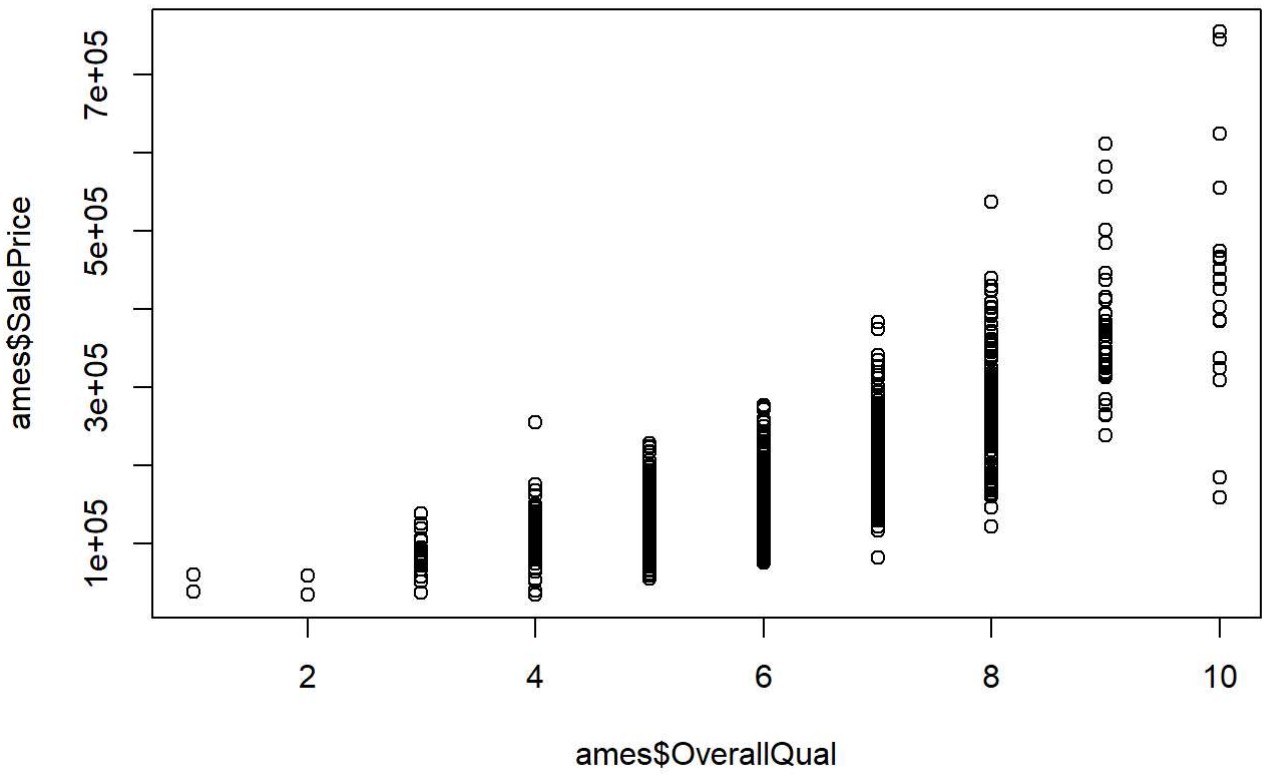
```
## Loading required package: ggplot2
```



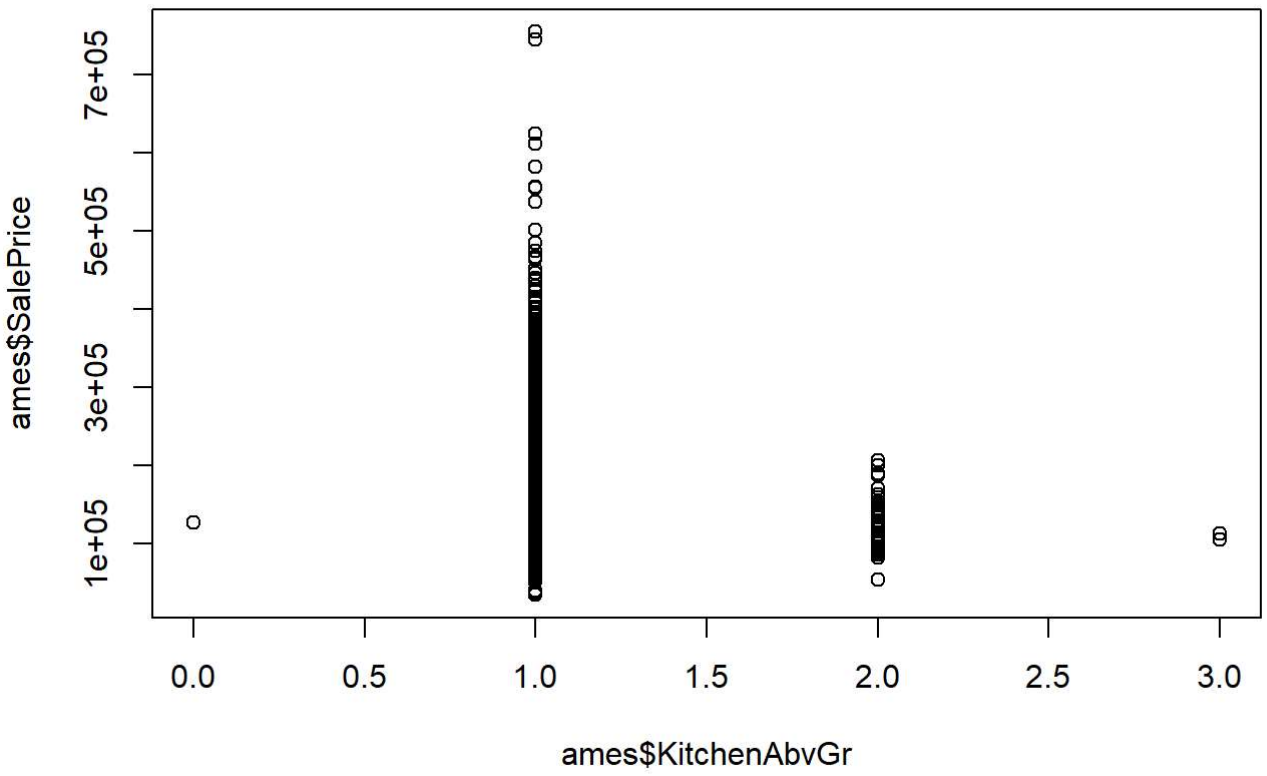

```
ggcorrplot(correlation)
```




```
plot(ames$OverallQual,ames$SalePrice)
```




```
plot (ames$KitchenAbvGr,ames$SalePrice)
```



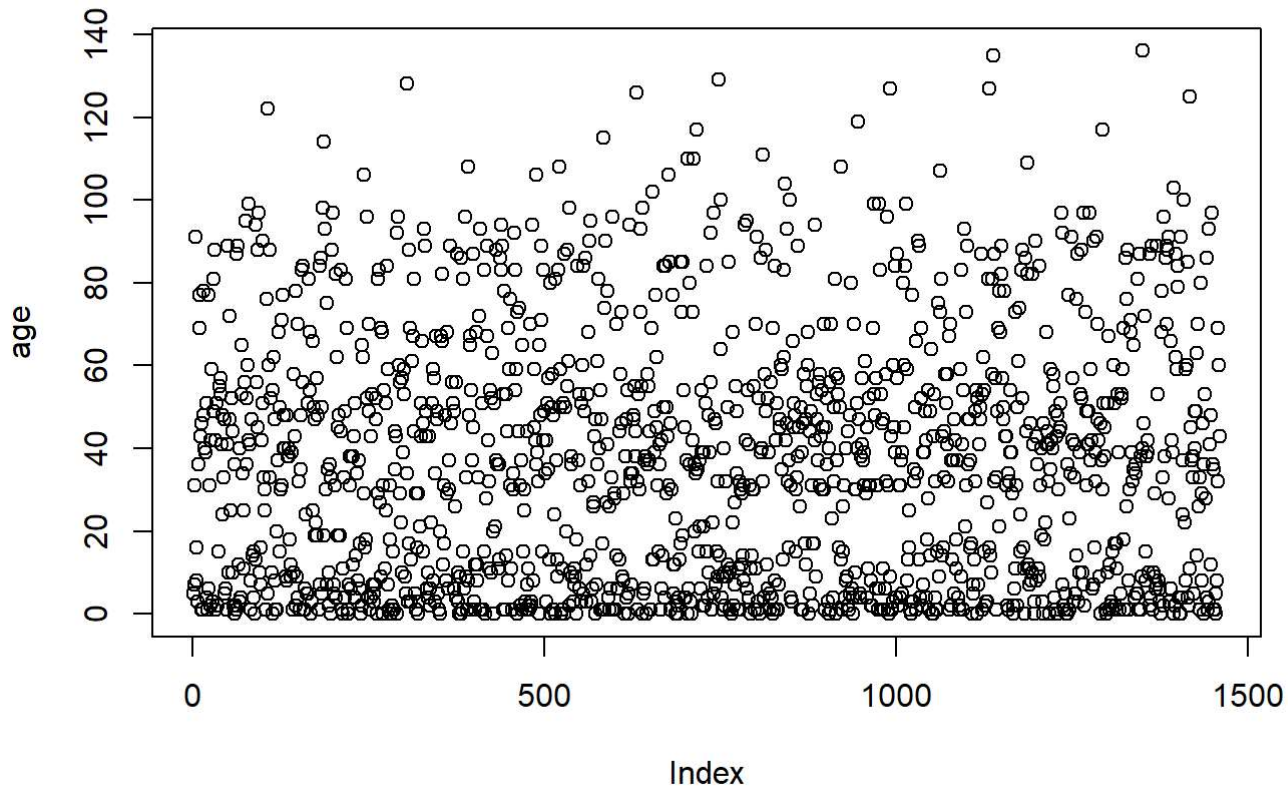

```
print ("the most correlated is overallquality, the least correlated is kitchen above ground")
```

```
## [1] "the most correlated is overallquality, the least correlated is kitchen above ground"
```

```
age <- ames$YrSold - ames$YearBuilt

print (head(age))
```

```
## [1]  5 31  7 91  8 16
```

```
plot(age)
```




```
print ("age and price of the house are positively correlated")
```

```
## [1] "age and price of the house are positively correlated"
```