Assignment1

MengZhang

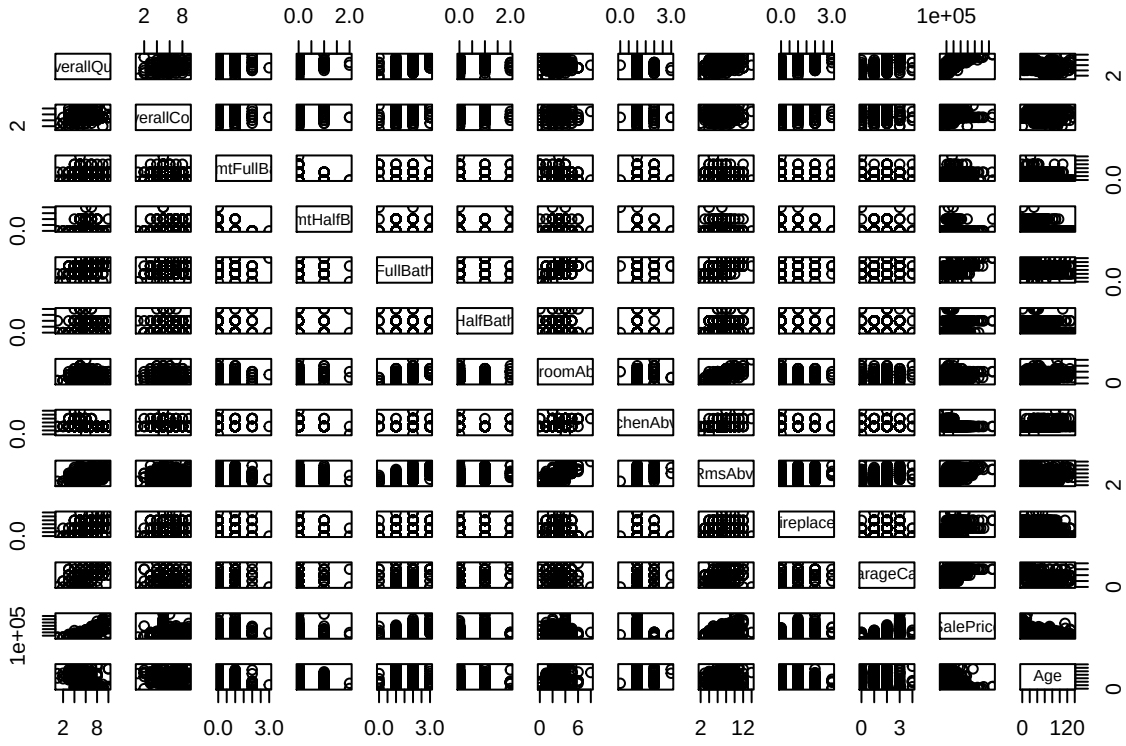
2019/2/18

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
price = read.csv("austin_house_price.csv")
pairs(price)
```




```
cor(price)
```

```
##              OverallQual OverallCond BsmtFullBath BsmtHalfBath    FullBath
## OverallQual   1.00000000 -0.09193234   0.11109779  -0.04015016  0.55059971
## OverallCond  -0.09193234  1.00000000  -0.05494152   0.11782092 -0.19414949
## BsmtFullBath  0.11109779 -0.05494152   1.00000000  -0.14787096 -0.06451205
## BsmtHalfBath -0.04015016  0.11782092  -0.14787096   1.00000000 -0.05453581
## FullBath      0.55059971 -0.19414949  -0.06451205  -0.05453581  1.00000000
## HalfBath      0.27345810 -0.06076933  -0.03090496  -0.01233990  0.13638059
## BedroomAbvGr  0.10167636  0.01298006  -0.15067281   0.04651885  0.36325198
## KitchenAbvGr -0.18388223 -0.08700086  -0.04150255  -0.03794435  0.13311521
## TotRmsAbvGrd  0.42745234 -0.05758317  -0.05327524  -0.02383634  0.55478425
## Fireplaces    0.39676504 -0.02381998   0.13792771   0.02897559  0.24367050
## GarageCars    0.60067072 -0.18575751   0.13188122  -0.02089106  0.46967204
```

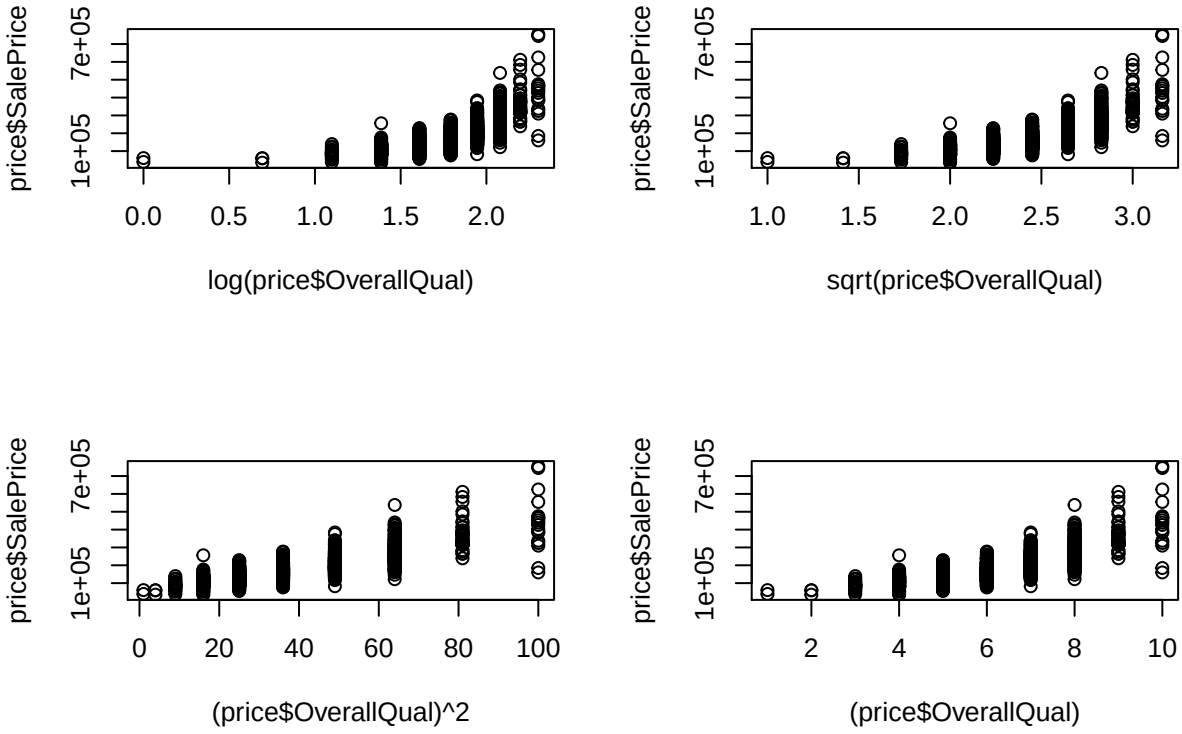
```
## SalePrice     0.79098160 -0.07785589   0.22712223  -0.01684415  0.56066376
## Age          -0.57262947  0.37732550  -0.18436183   0.03605963 -0.46840292
##                 HalfBath BedroomAbvGr KitchenAbvGr TotRmsAbvGrd
## OverallQual   0.27345810   0.10167636  -0.18388223   0.42745234
## OverallCond  -0.06076933   0.01298006  -0.08700086  -0.05758317
## BsmtFullBath -0.03090496  -0.15067281  -0.04150255  -0.05327524
## BsmtHalfBath -0.01233990   0.04651885  -0.03794435  -0.02383634
## FullBath      0.13638059   0.36325198   0.13311521   0.55478425
## HalfBath      1.00000000   0.22665148  -0.06826255   0.34341486
## BedroomAbvGr  0.22665148   1.00000000   0.19859676   0.67661994
## KitchenAbvGr -0.06826255   0.19859676   1.00000000   0.25604541
## TotRmsAbvGrd  0.34341486   0.67661994   0.25604541   1.00000000
## Fireplaces    0.20364851   0.10756968  -0.12393624   0.32611448
## GarageCars    0.21917815   0.08610644  -0.05063389   0.36228857
## SalePrice     0.28410768   0.16821315  -0.13590737   0.53372316
## Age          -0.24272773   0.06895972   0.17591841  -0.09695522
##                Fireplaces  GarageCars   SalePrice         Age
## OverallQual   0.39676504  0.60067072  0.79098160 -0.57262947
## OverallCond  -0.02381998 -0.18575751 -0.07785589  0.37732550
## BsmtFullBath  0.13792771  0.13188122  0.22712223 -0.18436183
## BsmtHalfBath  0.02897559 -0.02089106 -0.01684415  0.03605963
## FullBath      0.24367050  0.46967204  0.56066376 -0.46840292
## HalfBath      0.20364851  0.21917815  0.28410768 -0.24272773
## BedroomAbvGr  0.10756968  0.08610644  0.16821315  0.06895972
## KitchenAbvGr -0.12393624 -0.05063389 -0.13590737  0.17591841
## TotRmsAbvGrd  0.32611448  0.36228857  0.53372316 -0.09695522
## Fireplaces    1.00000000  0.30078877  0.46692884 -0.14854356
## GarageCars    0.30078877  1.00000000  0.64040920 -0.53872739
## SalePrice     0.46692884  0.64040920  1.00000000 -0.52335042
## Age          -0.14854356 -0.53872739 -0.52335042  1.00000000
```

```r
mlr = lm(SalePrice ~ ., data = price)
summary(mlr)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = price)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -274626  -21629   -3288   17476  374855
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -83029.36   10280.94  -8.076 1.40e-15 ***
## OverallQual   23140.58    1197.68  19.321  < 2e-16 ***
## OverallCond    4340.82    1035.88   4.190 2.95e-05 ***
## BsmtFullBath  21740.63    2130.05  10.207  < 2e-16 ***
## BsmtHalfBath  10236.97    4429.58   2.311    0.021 *
## FullBath      13417.14    2825.20   4.749 2.25e-06 ***
## HalfBath        239.56    2329.34   0.103    0.918
## BedroomAbvGr  -9599.12    1841.24  -5.213 2.12e-07 ***
## KitchenAbvGr -30303.01    5344.86  -5.670 1.73e-08 ***
## TotRmsAbvGrd  15129.23    1159.76  13.045  < 2e-16 ***
## Fireplaces    12668.63    1836.60   6.898 7.87e-12 ***
```

```
## GarageCars    16766.94    1873.62   8.949  < 2e-16 ***
## Age            -248.83      53.59  -4.643 3.75e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39330 on 1447 degrees of freedom
## Multiple R-squared:  0.7569, Adjusted R-squared:  0.7549
## F-statistic: 375.4 on 12 and 1447 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2, 2))
plot(log(price$OverallQual), price$SalePrice)
plot(sqrt(price$OverallQual), price$SalePrice)
plot((price$OverallQual)^2, price$SalePrice)
plot((price$OverallQual), price$SalePrice)
```




```
mlr = lm(SalePrice ~ log(OverallQual), data = price)
summary(mlr)
```

```
##
## Call:
## lm(formula = SalePrice ~ log(OverallQual), data = price)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -146807  -30758   -6069   21244  448193
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -248697      10538  -23.60   <2e-16 ***
## log(OverallQual)  241252       5864   41.14   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 54060 on 1458 degrees of freedom
## Multiple R-squared:  0.5372, Adjusted R-squared:  0.5369
## F-statistic:  1693 on 1 and 1458 DF,  p-value: < 2.2e-16
```

```
mlr = lm(SalePrice ~ sqrt(OverallQual), data = price)
summary(mlr)
```

```
##
## Call:
## lm(formula = SalePrice ~ sqrt(OverallQual), data = price)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -173134  -30139   -3786   21321  421866
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -345775      11584  -29.85   <2e-16 ***
## sqrt(OverallQual)  214690       4690   45.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50910 on 1458 degrees of freedom
## Multiple R-squared:  0.5897, Adjusted R-squared:  0.5894
## F-statistic:  2095 on 1 and 1458 DF,  p-value: < 2.2e-16
```

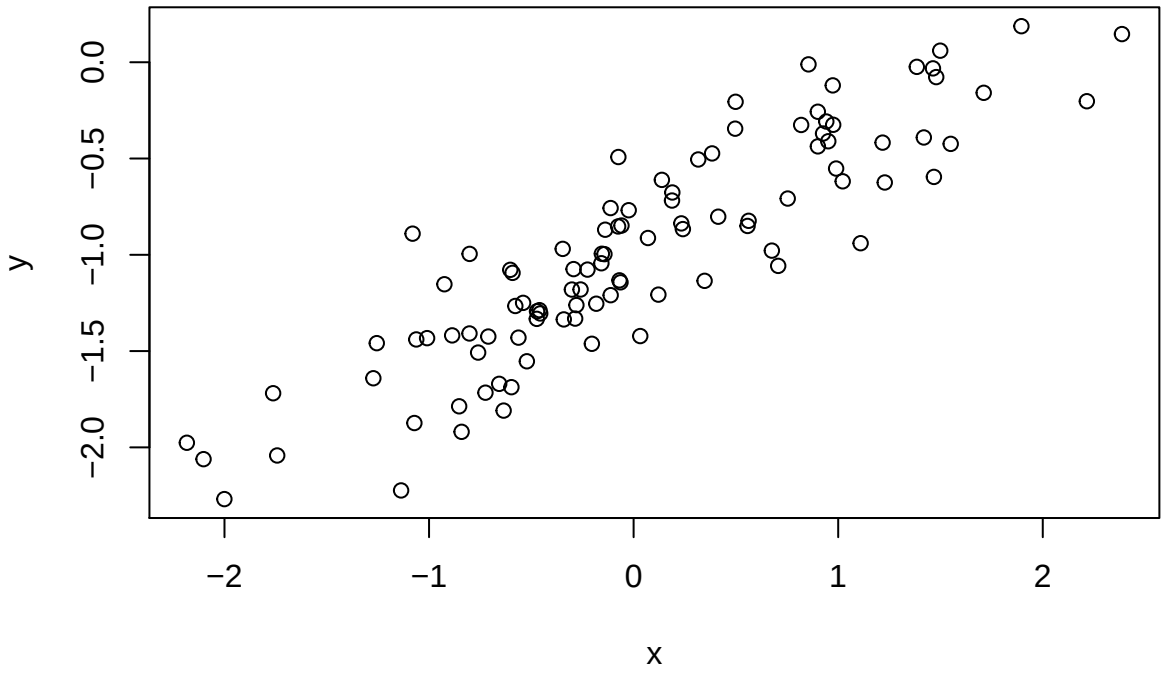
```
mlr = lm(SalePrice ~ (OverallQual)^2, data = price)
summary(mlr)
```

```
##
## Call:
## lm(formula = SalePrice ~ (OverallQual)^2, data = price)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -198152  -29409   -1845   21463  396848
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -96206.1     5756.4  -16.71   <2e-16 ***
## OverallQual  45435.8      920.4   49.36   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48620 on 1458 degrees of freedom
## Multiple R-squared:  0.6257, Adjusted R-squared:  0.6254
## F-statistic:  2437 on 1 and 1458 DF,  p-value: < 2.2e-16
```

```
set.seed(5)
x = rnorm(100, mean= 0, sd =1)
eps = rnorm(100, mean =0, sd = 0.25)
y = -1+0.5*x+eps
length(y)
```

```
## [1] 100
```

```r
plot(y~x)
```




```r
lr = lm(y~x)
summary(lr)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62059 -0.15387  0.02532  0.18585  0.68291
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.99693    0.02611  -38.18   <2e-16 ***
## x            0.53328    0.02775   19.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.261 on 98 degrees of freedom
## Multiple R-squared:  0.7903, Adjusted R-squared:  0.7882
## F-statistic: 369.4 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
polyrm = lm(y~poly(x,2))
summary(polyrm)
```

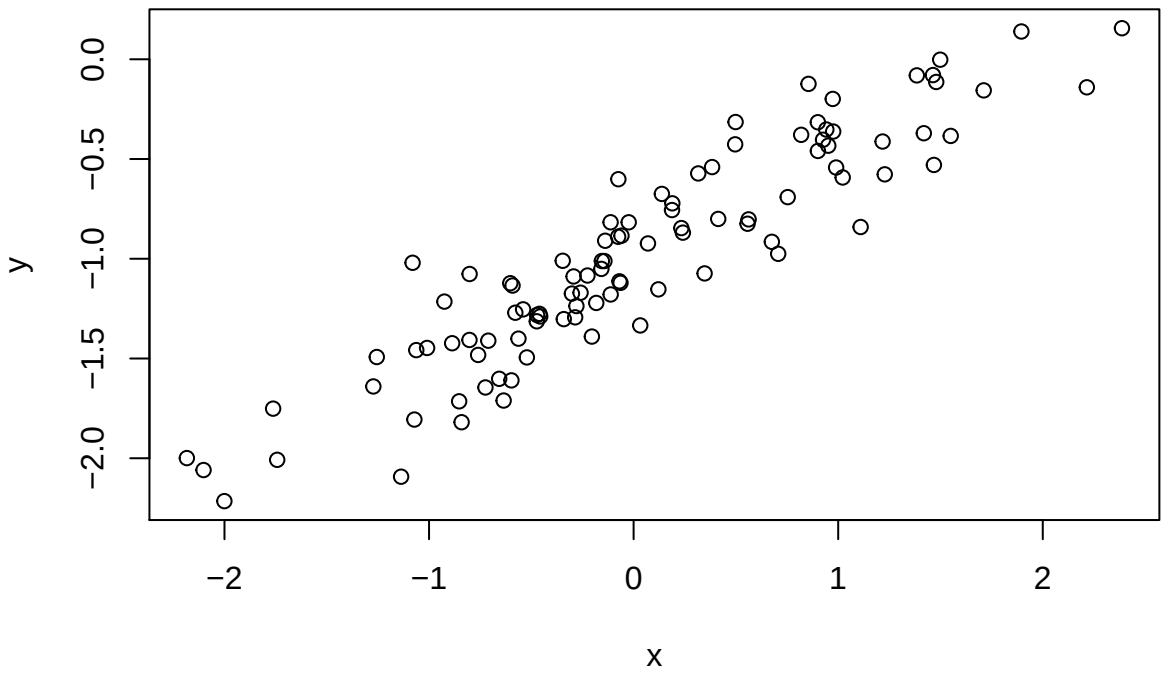
```
##
## Call:
## lm(formula = y ~ poly(x, 2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.61154 -0.16155  0.01887  0.20026  0.68997
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.98005    0.02617 -37.448   <2e-16 ***
## poly(x, 2)1  5.01572    0.26171  19.165   <2e-16 ***
## poly(x, 2)2 -0.17489    0.26171  -0.668    0.506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2617 on 97 degrees of freedom
## Multiple R-squared:  0.7913, Adjusted R-squared:  0.787
## F-statistic: 183.9 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
set.seed(5)
x = rnorm(100, mean= 0, sd =1)
eps = rnorm(100, mean =0, sd = 0.2)
y = -1+0.5*x+eps
length(y)
```

```
## [1] 100
```

```
plot(y~x)
```




```
lr = lm(y~x)
summary(lr)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49647 -0.12310  0.02026  0.14868  0.54633
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.99754    0.02089  -47.75   <2e-16 ***
## x            0.52662    0.02220   23.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2088 on 98 degrees of freedom
## Multiple R-squared:  0.8517, Adjusted R-squared:  0.8502
## F-statistic: 562.9 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
polyrm = lm(y~poly(x,2))
summary(polyrm)
```

```
##
## Call:
## lm(formula = y ~ poly(x, 2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48924 -0.12924  0.01509  0.16021  0.55197
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.98088    0.02094 -46.850   <2e-16 ***
## poly(x, 2)1  4.95312    0.20937  23.657   <2e-16 ***
## poly(x, 2)2 -0.13992    0.20937  -0.668    0.506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2094 on 97 degrees of freedom
## Multiple R-squared:  0.8524, Adjusted R-squared:  0.8493
## F-statistic: 280.1 on 2 and 97 DF,  p-value: < 2.2e-16
```