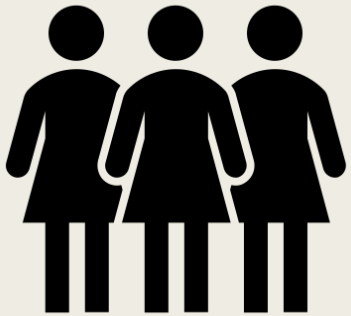
DELIVERY & TIPS

Chunhua Yu



RESEARCH QUESTIONS

- **Find the relationship between the PRICE of the order and TIPS.**
- **What kind of PEOPLE would like to give more tips.**
- **Is the DAY significant on the amount of delivered order ?**
- **Is the Time significant on the amount of delivered order ?**

VARIABLES

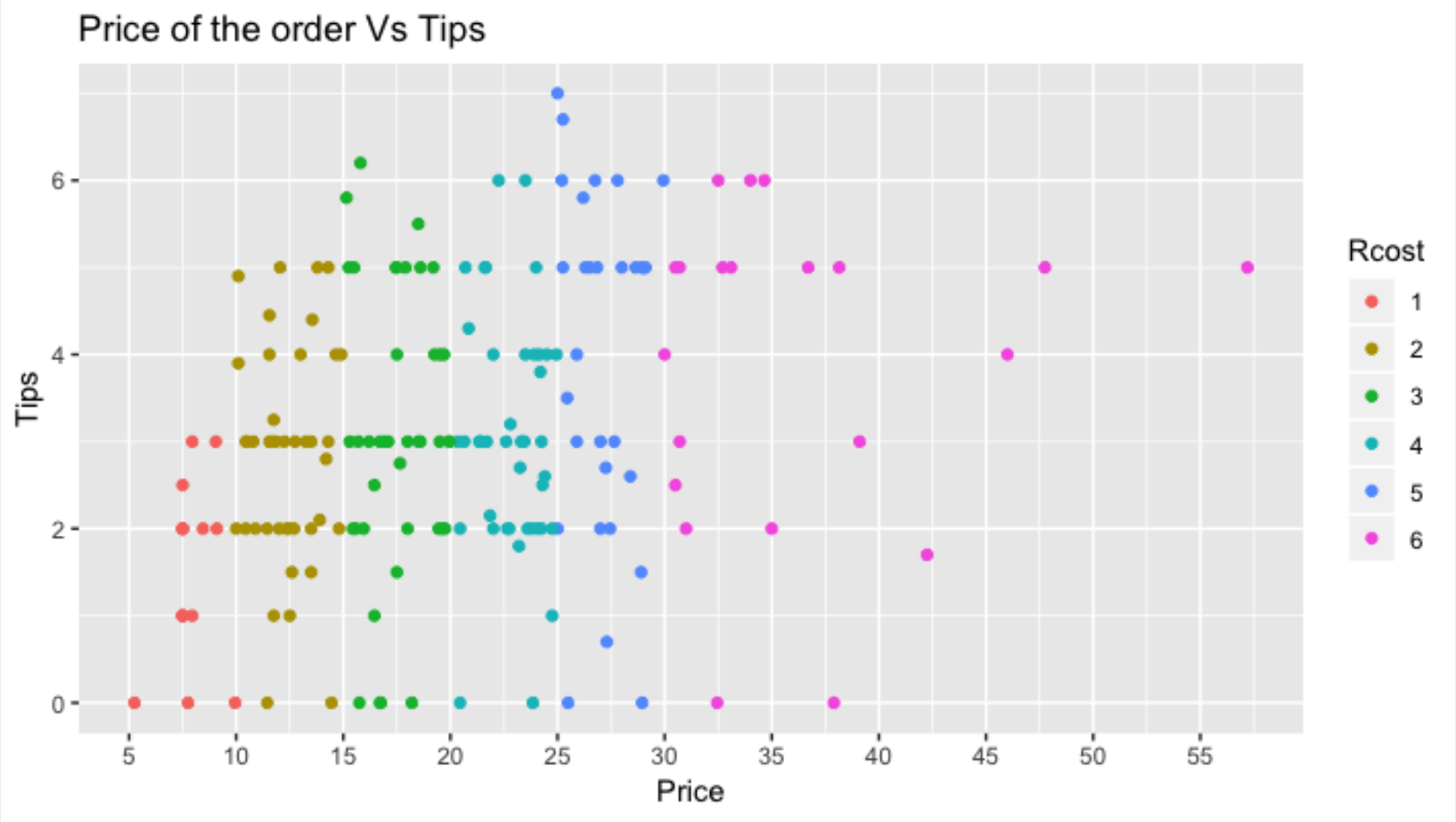
Response	• Tips
The information of the order	• Price, Address, Time, Day
The information of the people	• Gender, Age, Race, House

DATA CODING

	Gender	Age	Race	Time	House	Day
0	Female	Young	White	11am~6pm	Apartment	Monday-1
1	Male	Middle	Africa American	6pm~11pm	Small house	Tuesday-2
2	-	Older	Other		Big house	…
						Sunday-7

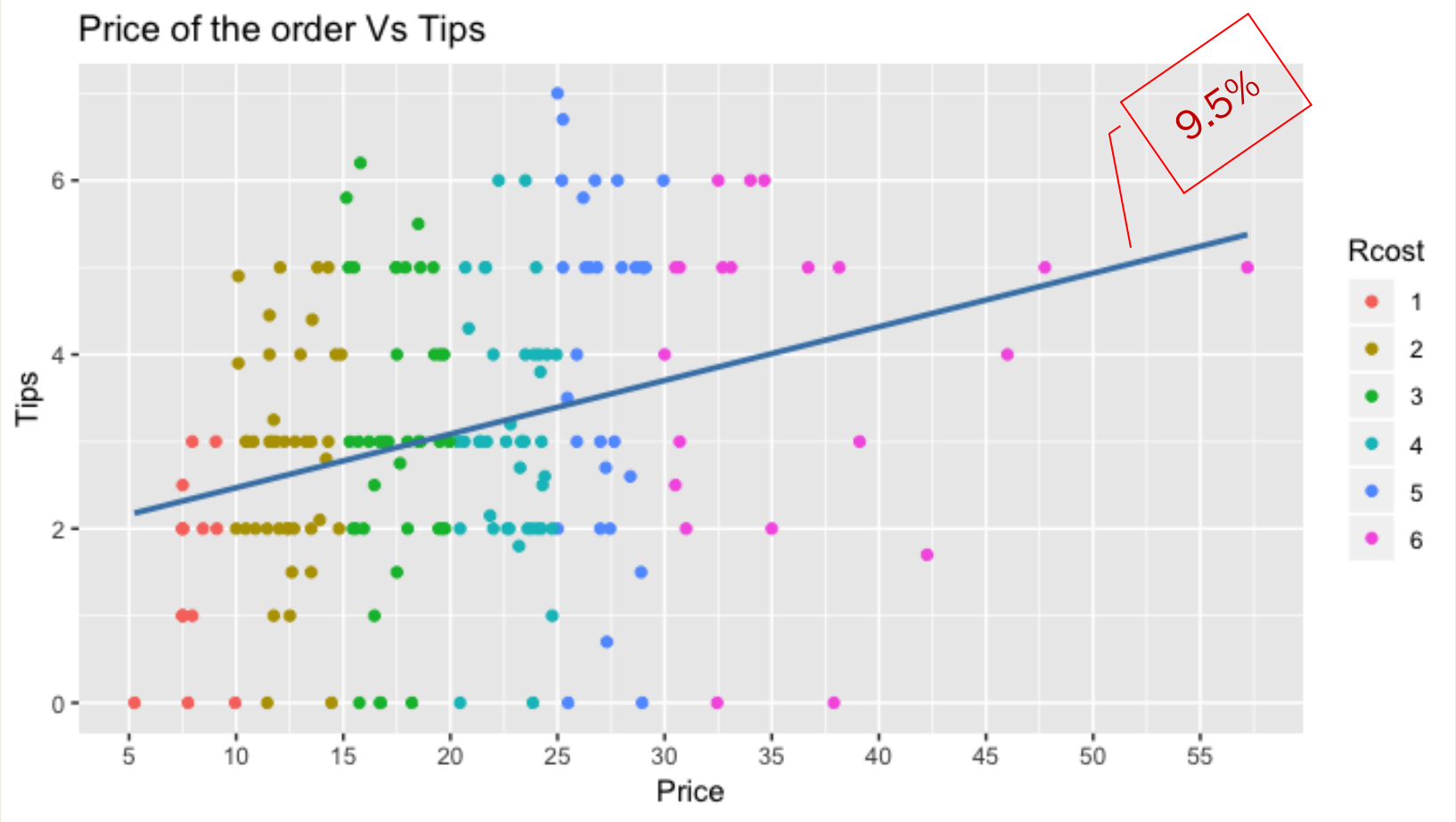
```
> head(data.r)
  Price Tip Gender Age Race Time house Day
1 57.20 5.0      0   1    0    1     2   1
2 47.75 5.0      1   1    2    1     1   5
3 46.00 4.0      0   0    0    0     1   7
4 42.25 1.7      1   1    0    1     2   4
5 39.10 3.0      0   0    0    1     1   2
6 38.15 5.0      0   2    0    1     2   6
>
```

RQ1: Find the relationship between the PRICE of the order and TIPS.



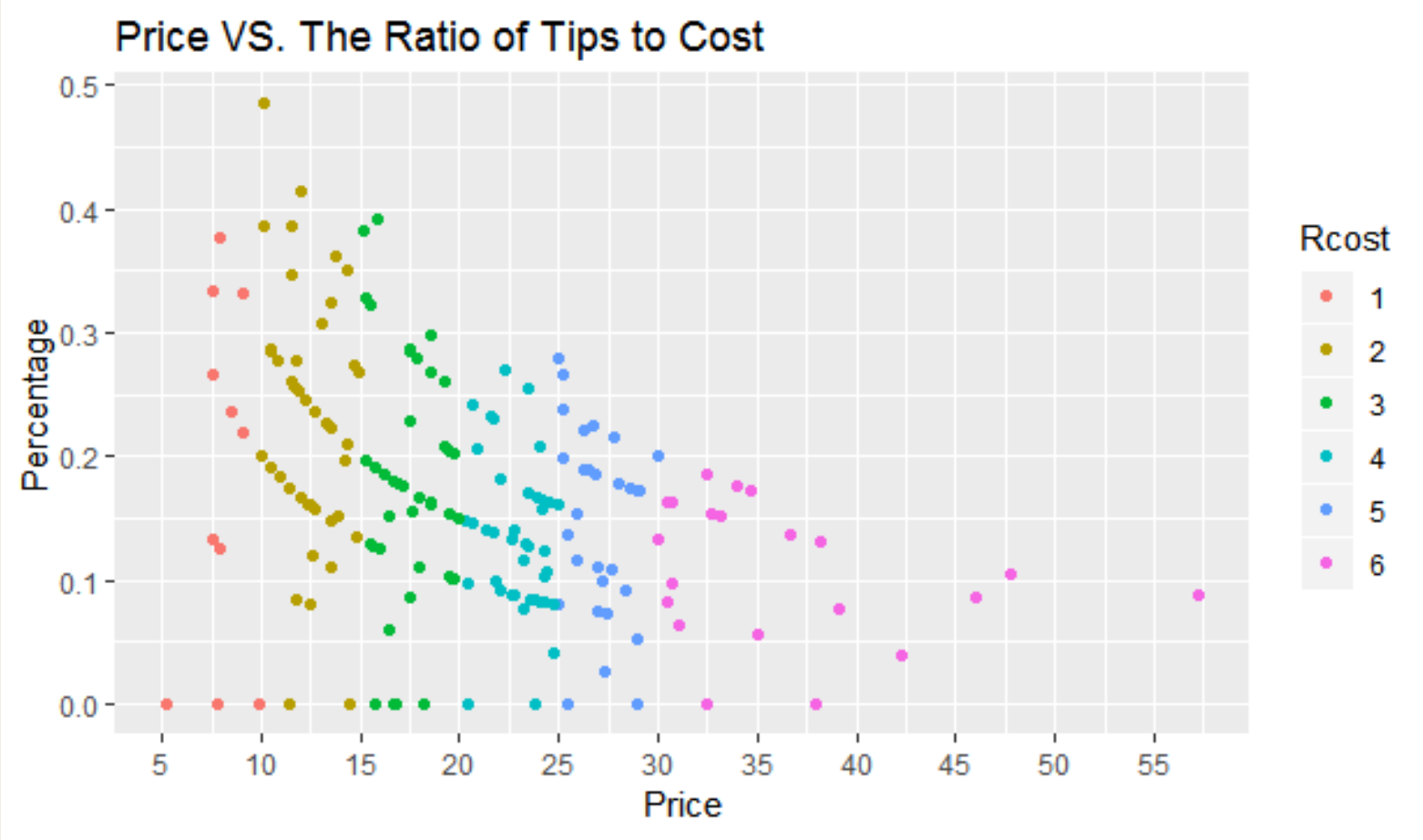
It is supposed to be a linear relatisonship.

RQ1: Find the relationship between the PRICE of the order and TIPS.



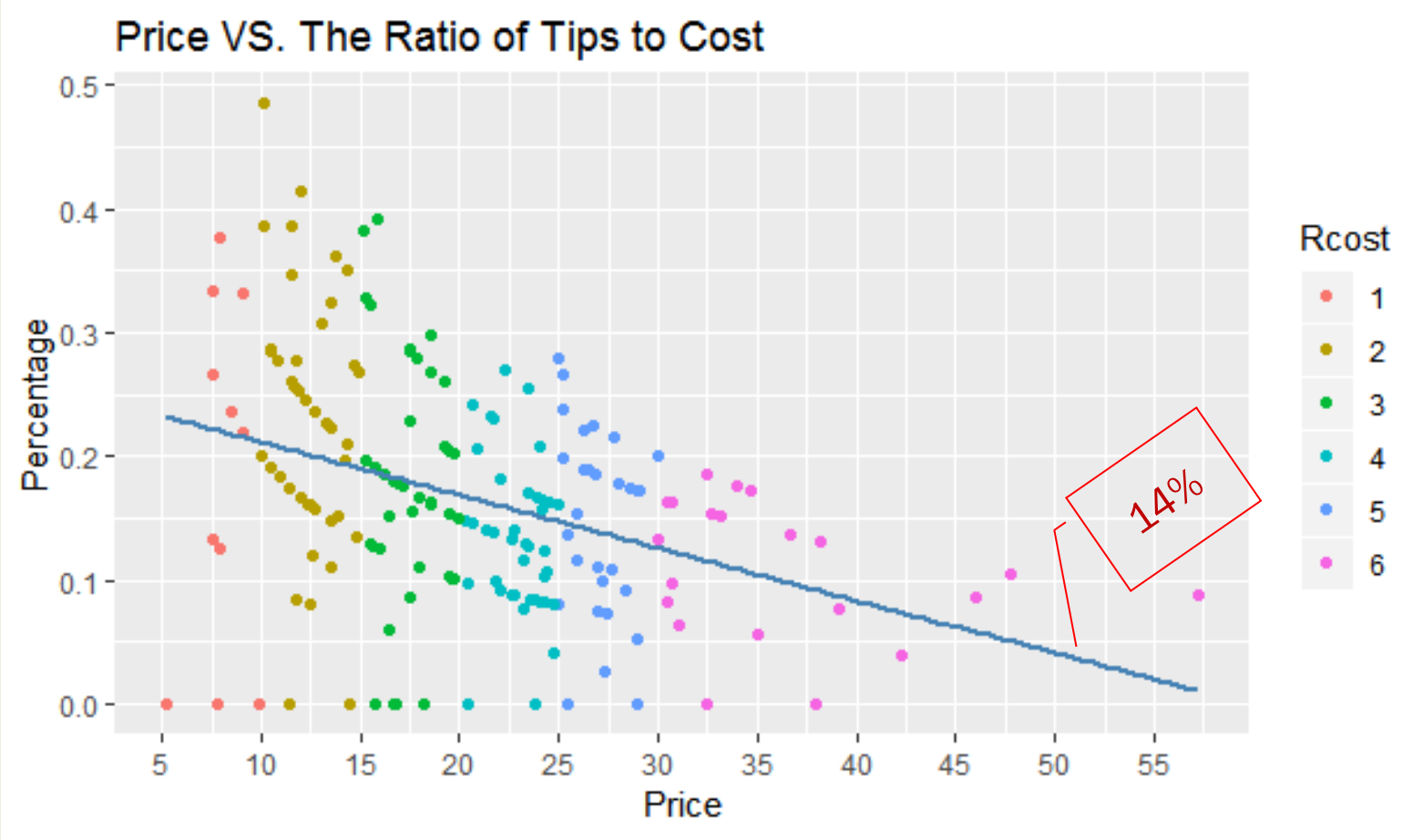
It is supposed to be a linear relatisonship.

RQ2: Find the relationship between the PRICE of the order and PERCENTAGE.



It is supposed to be a linear relatisonship.

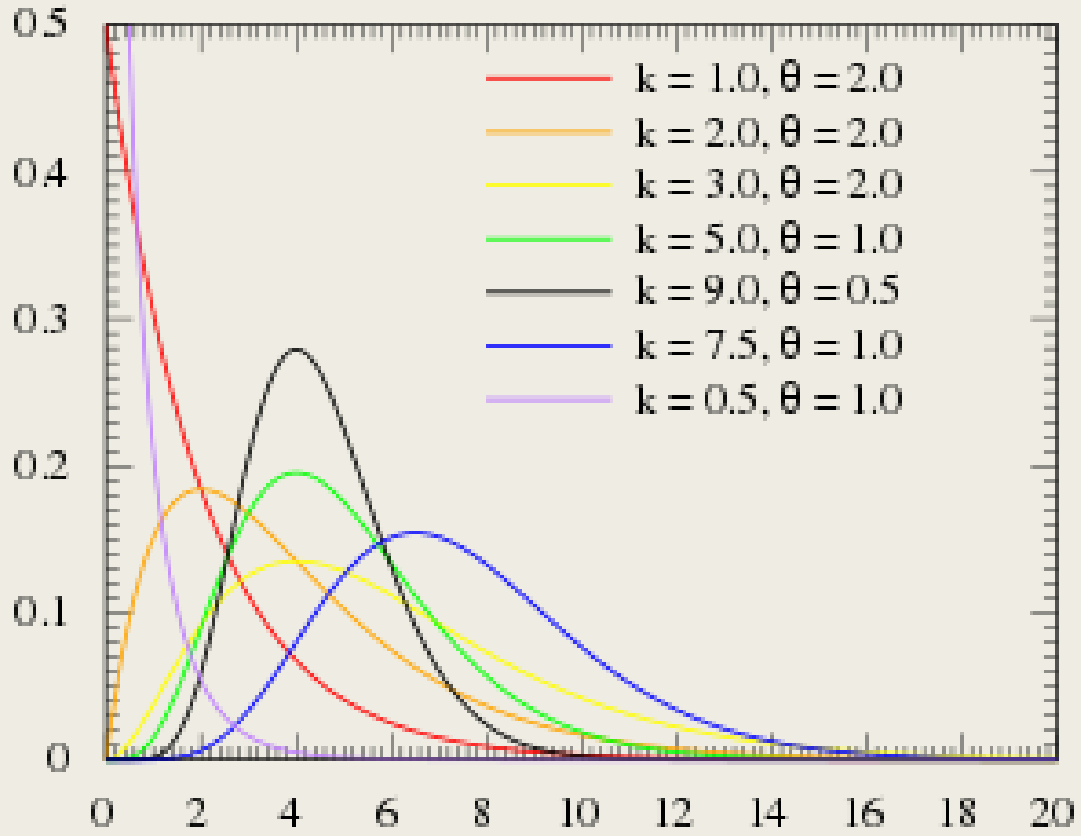
RQ2: Find the relationship between the PRICE of the order and PERCENTAGE.



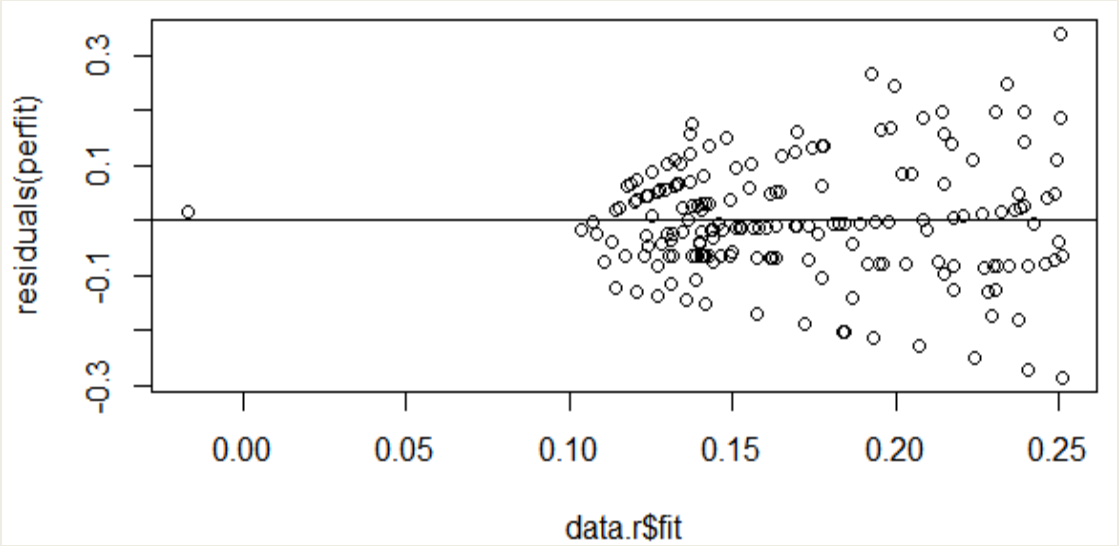
It is supposed to be a linear relatisonship. Nonlinear?

Gamma PDF Curve

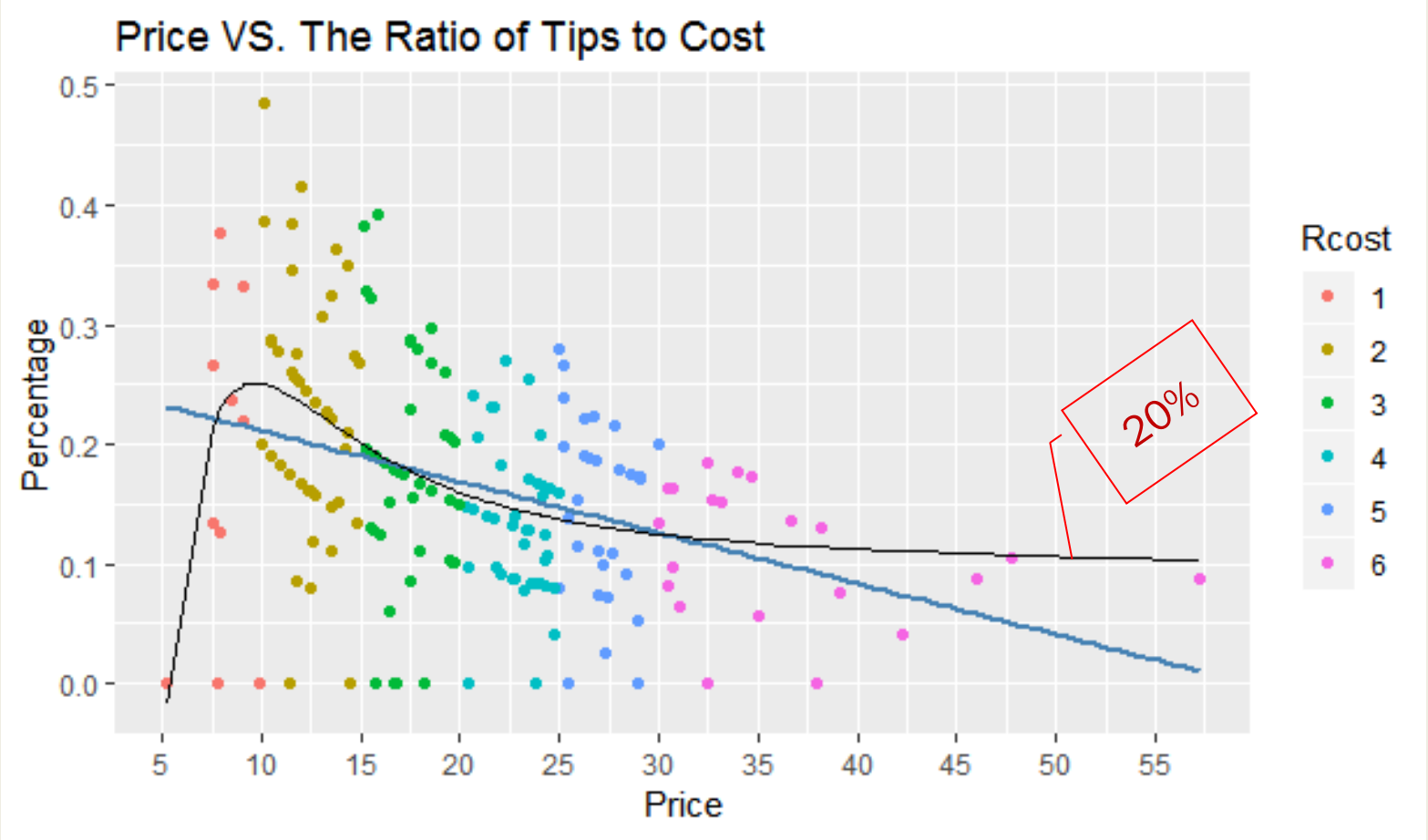
$$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$



```
> data.r$expp=exp(data.r$Percentage)
> data.r$recp=data.r$Price^-2
> data.r$expc=exp(-0.5*data.r$Price)
> perfit=lm(expp~recp+expc,data.r)
> summary(perfit)

Call:
lm(formula = expp ~ recp + expc, data = data.r)

Residuals:
     Min       1Q   Median       3Q      Max
-0.28540 -0.06589 -0.00764  0.06143  0.33962

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.10053    0.01437  76.584  < 2e-16 ***
recp         29.52541    4.23233   6.976 4.96e-11 ***
expc        -16.40825    2.70772  -6.060 7.24e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1034 on 189 degrees of freedom
Multiple R-squared:  0.2083,    Adjusted R-squared:  0.1999
F-statistic: 24.86 on 2 and 189 DF,  p-value: 2.595e-10

> shapiro.test(residuals(perfit))

        Shapiro-Wilk normality test

data:  residuals(perfit)
W = 0.98975, p-value = 0.1857
```

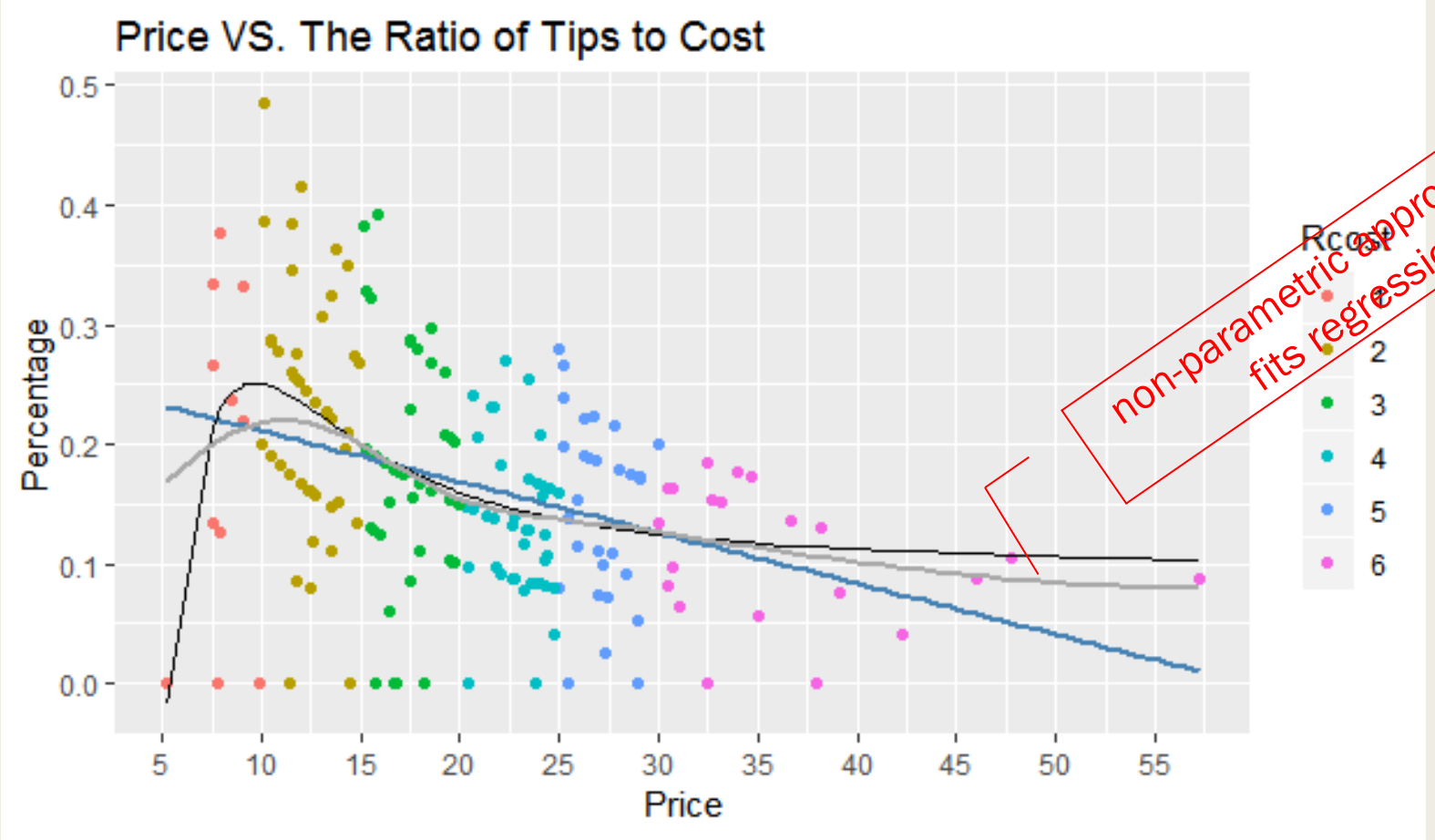


RQ2: Find the relationship between the PRICE of the order and Ratio of Tips to Price.



Fit nonlinear regression exp

RQ2: Find the relationship between the PRICE of the order and Ratio of Tips to Price.



Fit nonlinear regression exp

RQ1: Find the relationship between the PRICE of the order and Tips.

Although the increasing trend of percentage is going to be flat, tips is increasing along the increasing price of order.

RQ3: People vs Percentage ?

```
> perfit.full=lm(expp~recp+expc+Age+Race+Gender+House,data.r)
> summary(perfit.full)

Call:
lm(formula = expp ~ recp + expc + Age + Race + Gender + House,
    data = data.r)

Residuals:
      Min        1Q    Median        3Q       Max
-0.290465 -0.061844  0.003994  0.062692  0.315149

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.01264    0.02239  45.237  < 2e-16 ***
recp         31.49536    4.05203   7.773 5.47e-13 ***
expc        -17.02158    2.58517  -6.584 4.74e-10 ***
Age1          0.05553    0.01728   3.214  0.00155 **
Age2          0.02542    0.01836   1.384  0.16803
Race1        -0.01043    0.02650  -0.394  0.69435
Race2         0.02447    0.02091   1.170  0.24350
Gender1       0.02564    0.01469   1.745  0.08263 .
House1        0.05187    0.01690   3.070  0.00247 **
House2        0.06021    0.02180   2.762  0.00634 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09701 on 182 degrees of freedom
Multiple R-squared:  0.329,     Adjusted R-squared:  0.2958
F-statistic: 9.913 on 9 and 182 DF,  p-value: 2.549e-12
```

RQ3: People vs Percentage ?

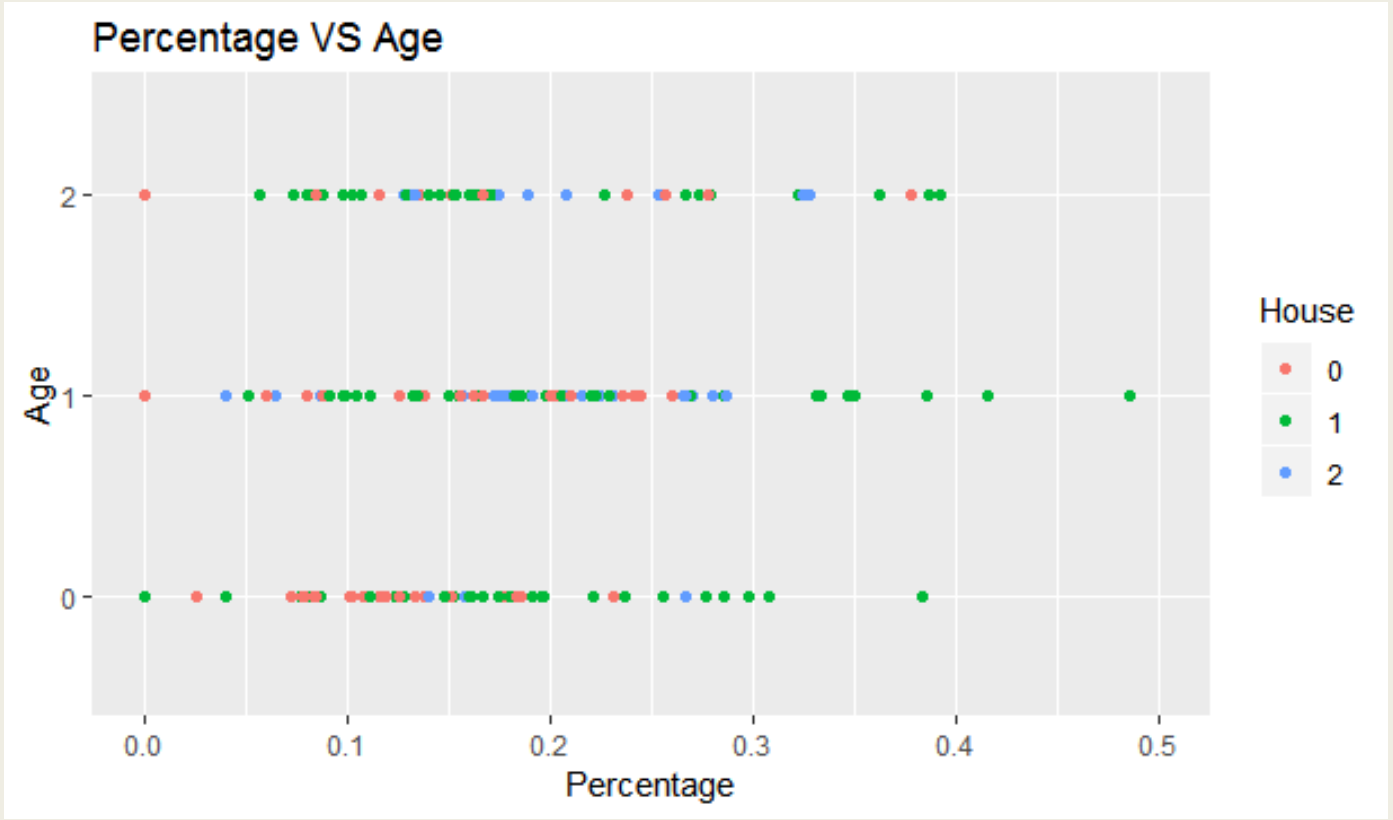
```
> anova(perfit.full)
Analysis of Variance Table

Response: expp
           Df  Sum Sq Mean Sq F value    Pr(>F)
recp        1 0.13905 0.13905 14.7749 0.0001673 ***
expc        1 0.39264 0.39264 41.7196 9.325e-10 ***
Age         2 0.14381 0.07191  7.6404 0.0006513 ***
Race        2 0.03749 0.01875  1.9919 0.1393987
Gender      1 0.02217 0.02217  2.3560 0.1265363
House       2 0.10452 0.05226  5.5529 0.0045617 **
Residuals 182 1.71288 0.00941
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

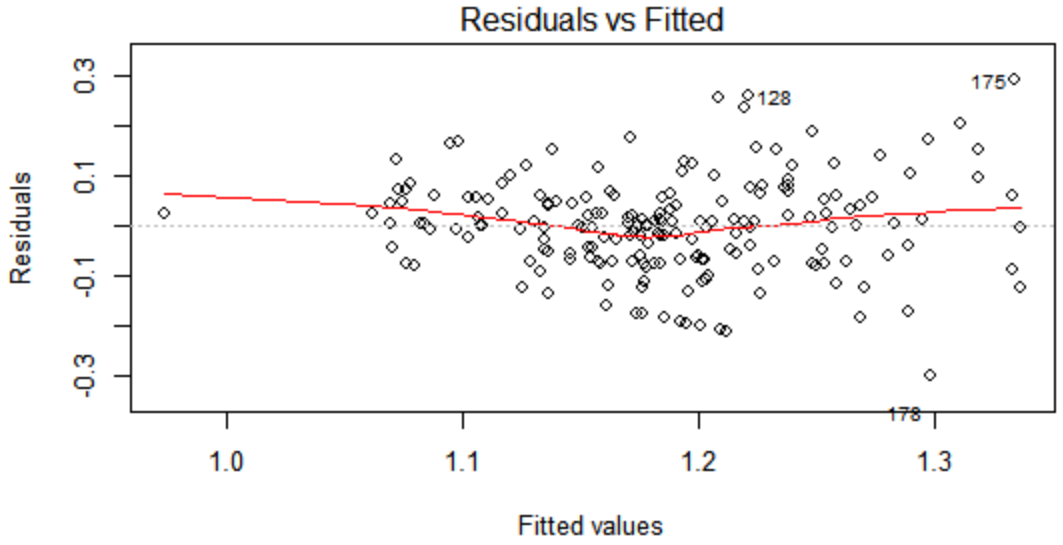
RQ3: People vs Percentage ?

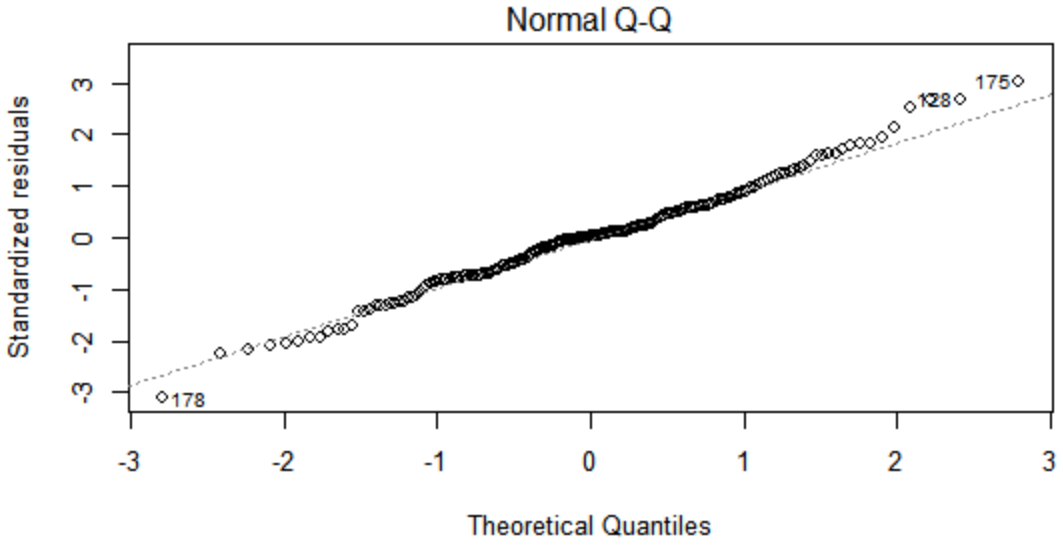
```
> perfit.red=lm(expp~recp+expc+Age+House,data.r)
> anova(perfit.red,perfit.full)
Analysis of Variance Table

Model 1: expp ~ recp + expc + Age + House
Model 2: expp ~ recp + expc + Age + Race + Gender + House
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    185 1.7637
2    182 1.7129  3  0.050772 1.7982 0.1491
> anova(perfit.red)
Analysis of Variance Table

Response: expp
           Df  Sum Sq Mean Sq F value    Pr(>F)
recp        1 0.13905 0.13905 14.5861 0.0001826 ***
expc        1 0.39264 0.39264 41.1865  1.13e-09 ***
Age         2 0.14381 0.07191  7.5427 0.0007095 ***
House       2 0.11342 0.05671  5.9484 0.0031354 **
Residuals 185 1.76366 0.00953
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

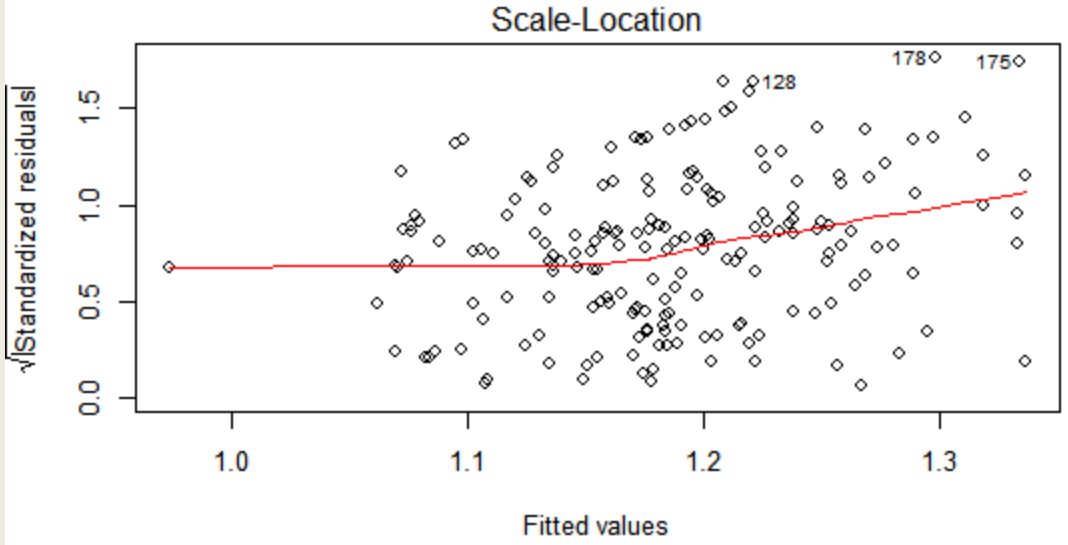
RQ3: People vs Percentage ?

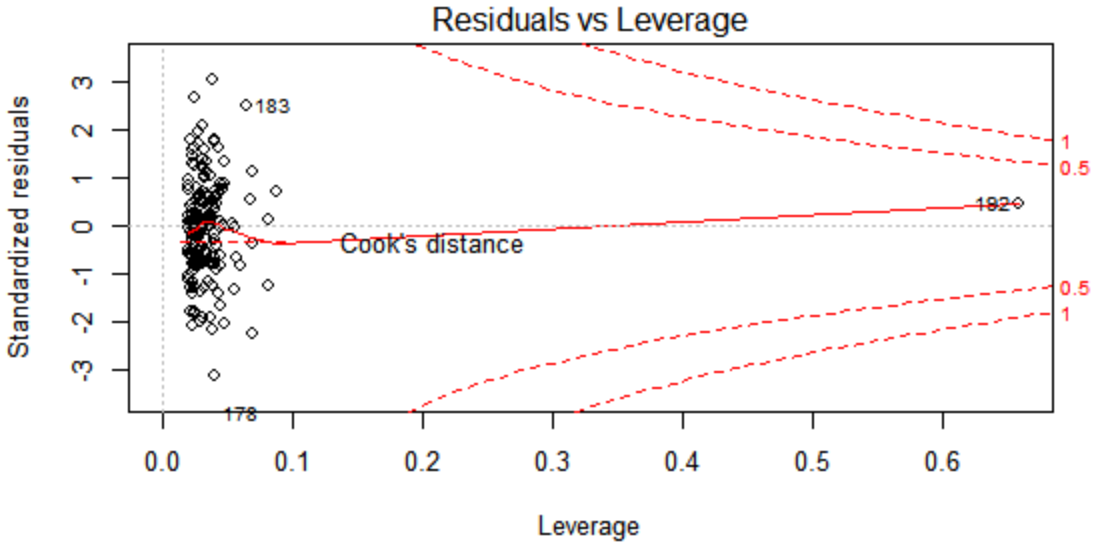
	House			
Age	0	1	2	Sum
0	24	34	6	64
1	18	36	19	73
2	12	30	13	55
Sum	54	100	38	192




Residuals vs Fitted
Residuals
Fitted values
128
175
178


Normal Q-Q
Standardized residuals
Theoretical Quantiles
128
175
178


Scale-Location
√|Standardized residuals|
Fitted values
128
175
178


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
183
182
178
1
0.5
0.5
1

RQ4: Is the Time (Day) significant on the amount of delivered order ?

```
> data.salaryT
   salaryT orderday ntime time date
1    51.00   214.25    12    0    1
2    69.90   291.80    18    0    2
3    61.05   282.95    20    0    4
4    76.40   321.70    17    0    6
5    78.70   394.95    17    0    7
6    78.30   450.60    19    1    1
7    59.40   298.34    16    1    2
8    77.55   448.70    20    1    4
9    79.15   434.19    18    1    6
10   53.30   289.65    14    1    7
```

```
> matrix(ntime,ncol=2)
     [,1] [,2]
[1,]   12   19
[2,]   18   16
[3,]   20   20
[4,]   17   18
[5,]   17   14
> chisq.test(matrix(ntime,ncol=2))

        Pearson's Chi-squared test

data:  matrix(ntime, ncol = 2)
X-squared = 1.9652, df = 4, p-value = 0.7422
```

RQ4: Is the Time (Day) significant on the amount of delivered order ? Poisson

```
> nfit=glm(ntime~time+date,family = "poisson" )
> summary(nfit)

Call:
glm(formula = ntime ~ time + date, family = "poisson")

Deviance Residuals:
       1         2         3         4         5         6         7         8
-0.85938   0.31368   0.07892  -0.04663   0.44567   0.78723  -0.31618  -0.07800
       9        10
 0.04565  -0.45495

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.723e+00  1.957e-01  13.912   <2e-16 ***
time1        3.509e-02  1.530e-01   0.229    0.819
date2        9.237e-02  2.483e-01   0.372    0.710
date4        2.549e-01  2.393e-01   1.065    0.287
date6        1.214e-01  2.466e-01   0.492    0.623
date7       -1.692e-16  2.540e-01   0.000    1.000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3.6019  on 9  degrees of freedom
Residual deviance: 1.9788  on 4  degrees of freedom
AIC: 60.737
```

RQ4: Is the Time (Day) significant on the salary?

```
> salaryfit=lm(salaryT~ ntime+orderday+time+date)
> anova(salaryfit)
Analysis of Variance Table

Response: salaryT
          Df Sum Sq Mean Sq F value  Pr(>F)
ntime      1 542.02  542.02  95.757 0.01028 *
orderday   1 328.53  328.53  58.042 0.01680 *
time       1 163.57  163.57  28.897 0.03291 *
date       4  89.93   22.48   3.972 0.21111
Residuals  2  11.32    5.66
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

LIMITATION

- RACE=c(149,16,27)
- The driver who collected information judged the data, biased data
- Full-load work for the driver.
- Some observations were from same person.

Thanks!!!