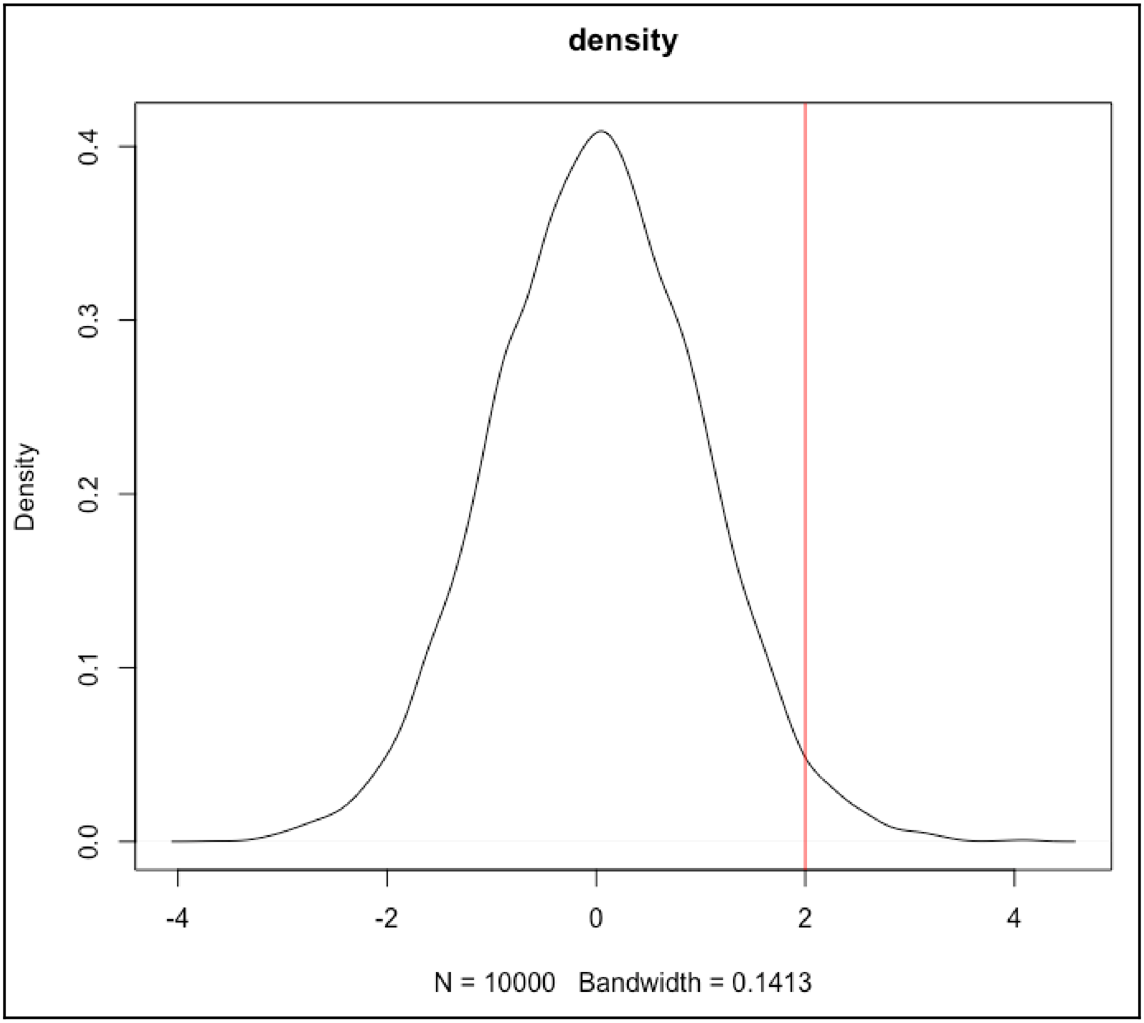
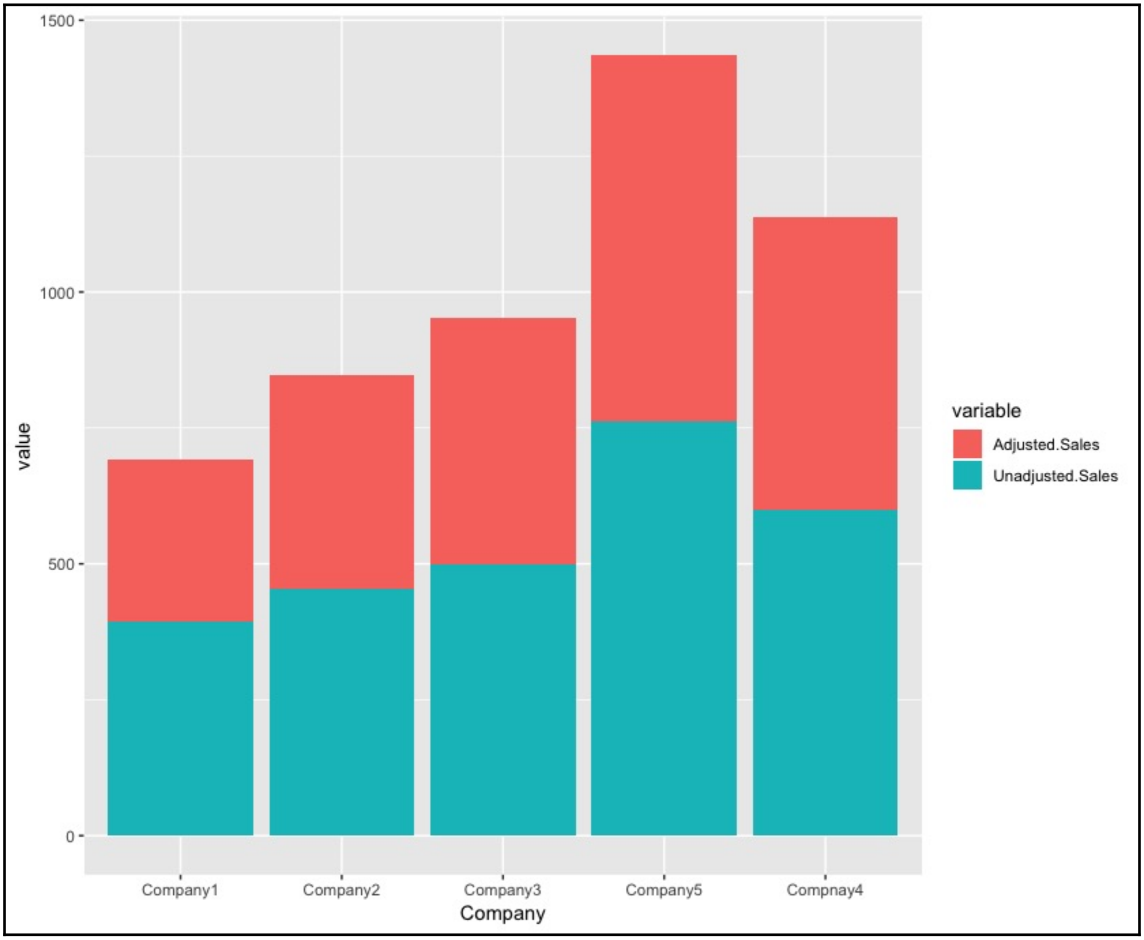
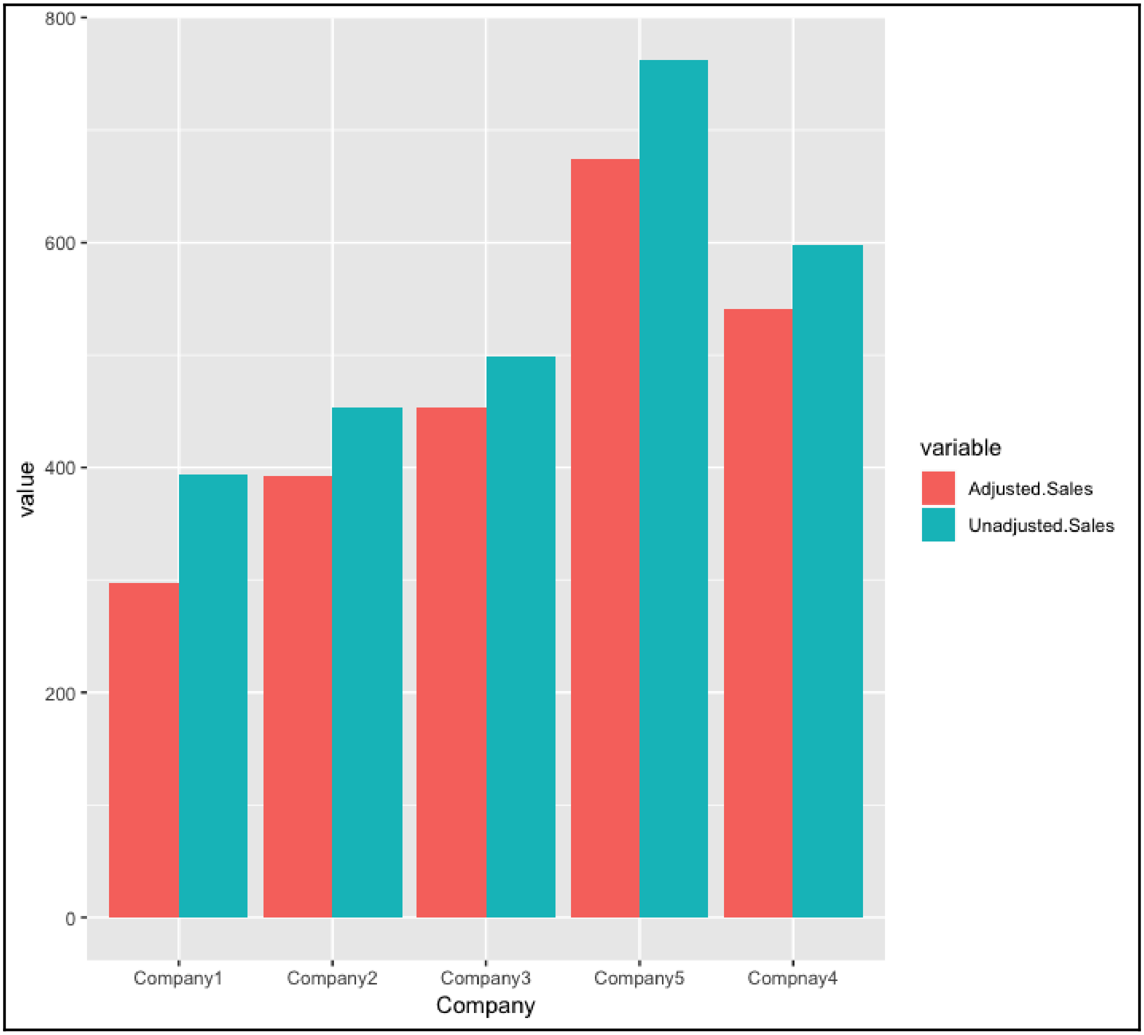
Chapter 1: Getting Started with R and Statistics

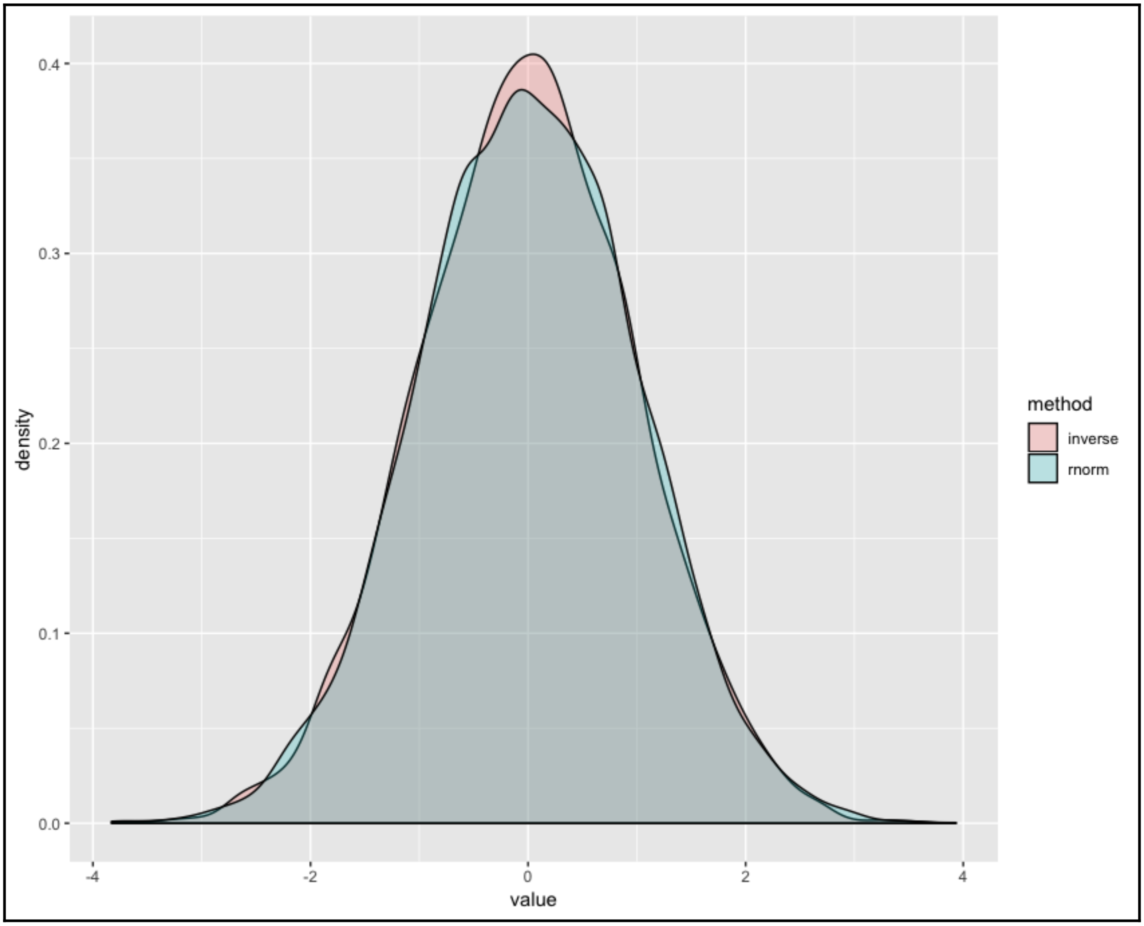
```
> print(paste("Area to the left of x=2",pnorm(2,0,1)))
[1] "Area to the left of x=2 0.977249868051821"
> print(paste("Area to the right of x=2",1-pnorm(2,0,1)))
[1] "Area to the right of x=2 0.0227501319481792"
> print(paste("90th Quantile: x value that has 90% to the left",qnorm(0.9772,0,1)))
[1] "90th Quantile: x value that has 90% to the left 1.99907721497177"
```

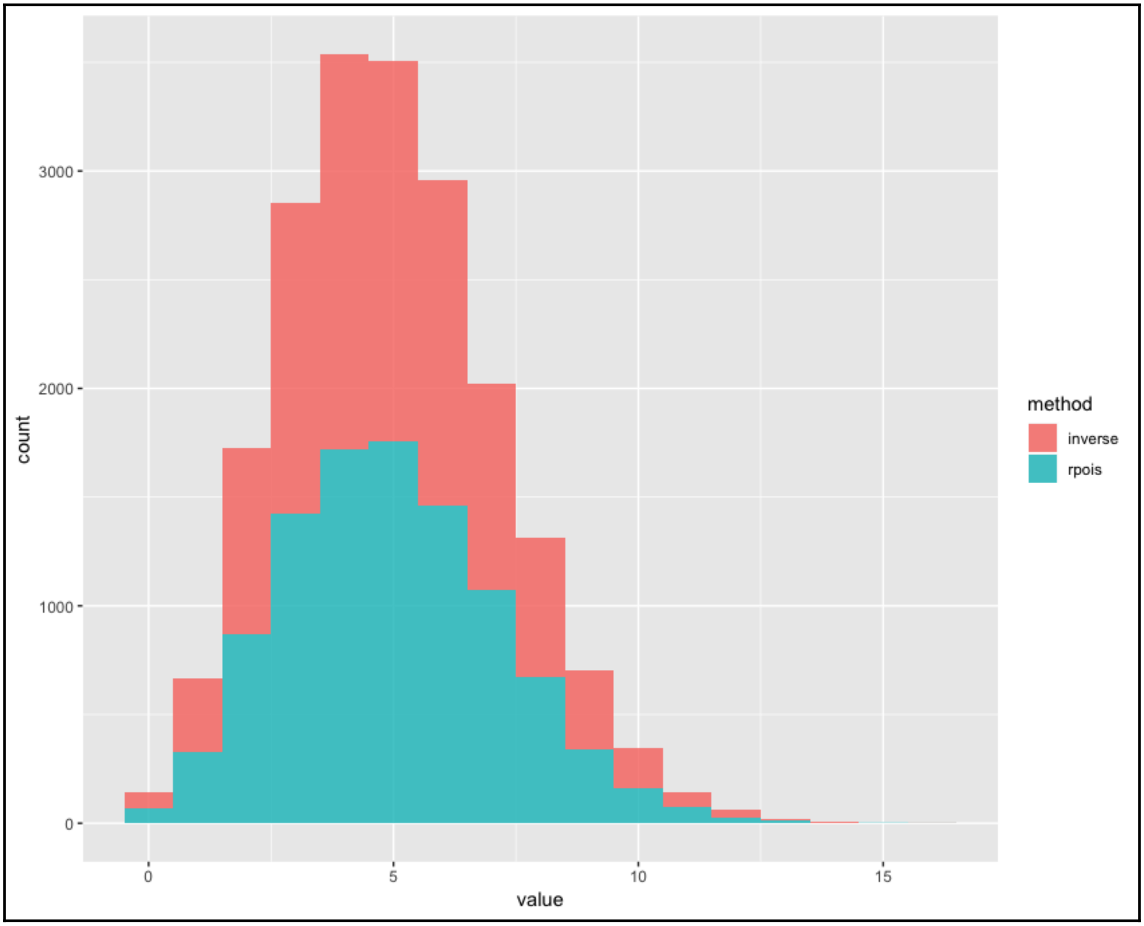



```
> print(paste("Area to the left of x=3",pchisq(3,33)))
[1] "Area to the left of x=3 2.2915465503982e-12"
```

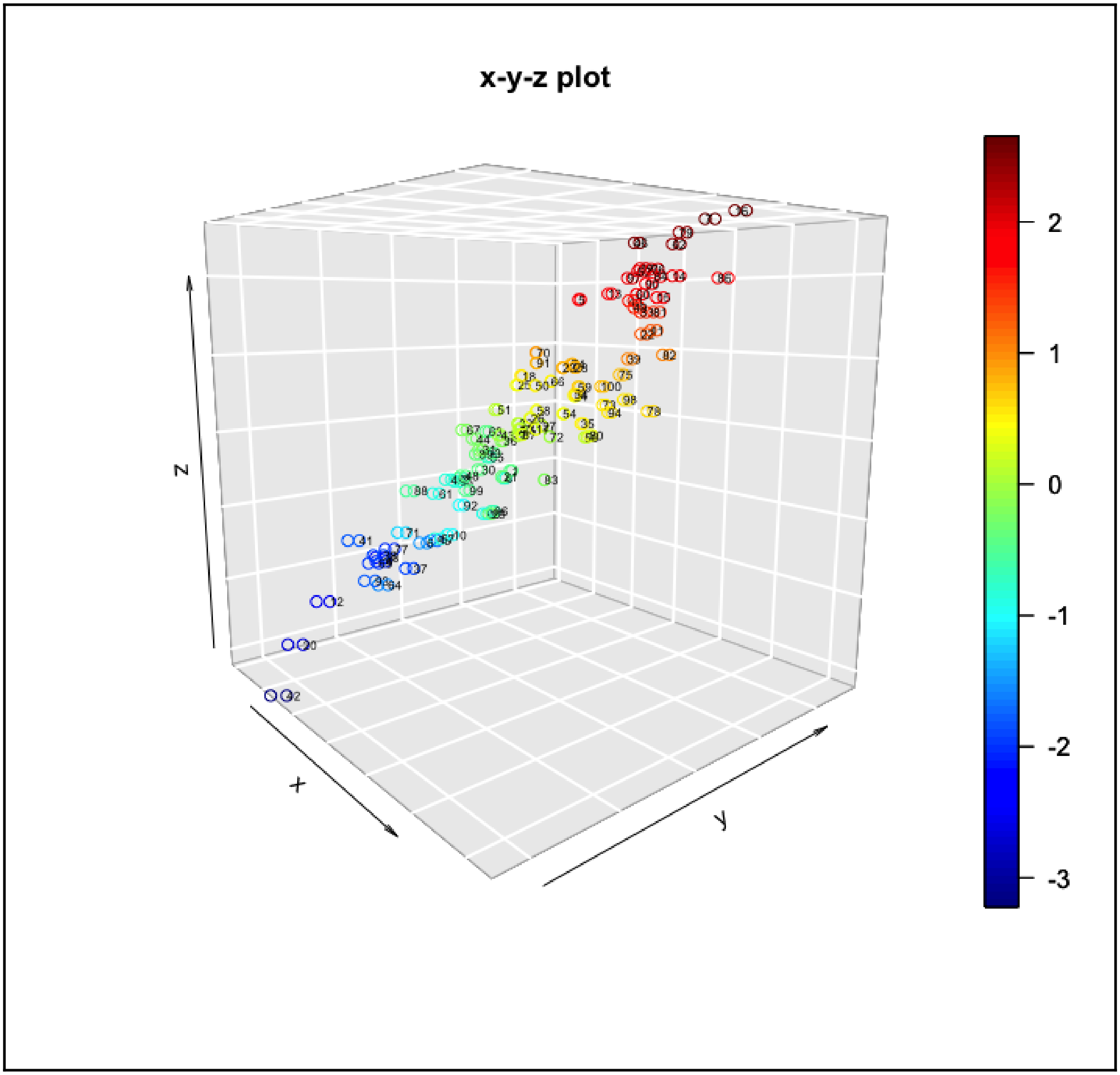

1500
1000
500
0
value
Company1
Company2
Company3
Company5
Compnay4
Company
variable
Adjusted.Sales
Unadjusted.Sales


800
600
400
200
0
value
Company1
Company2
Company3
Company5
Compnay4
Company
variable
Adjusted.Sales
Unadjusted.Sales


0.4
0.3
0.2
0.1
0.0
density
-4
-2
0
2
4
value
method
inverse
rnorm

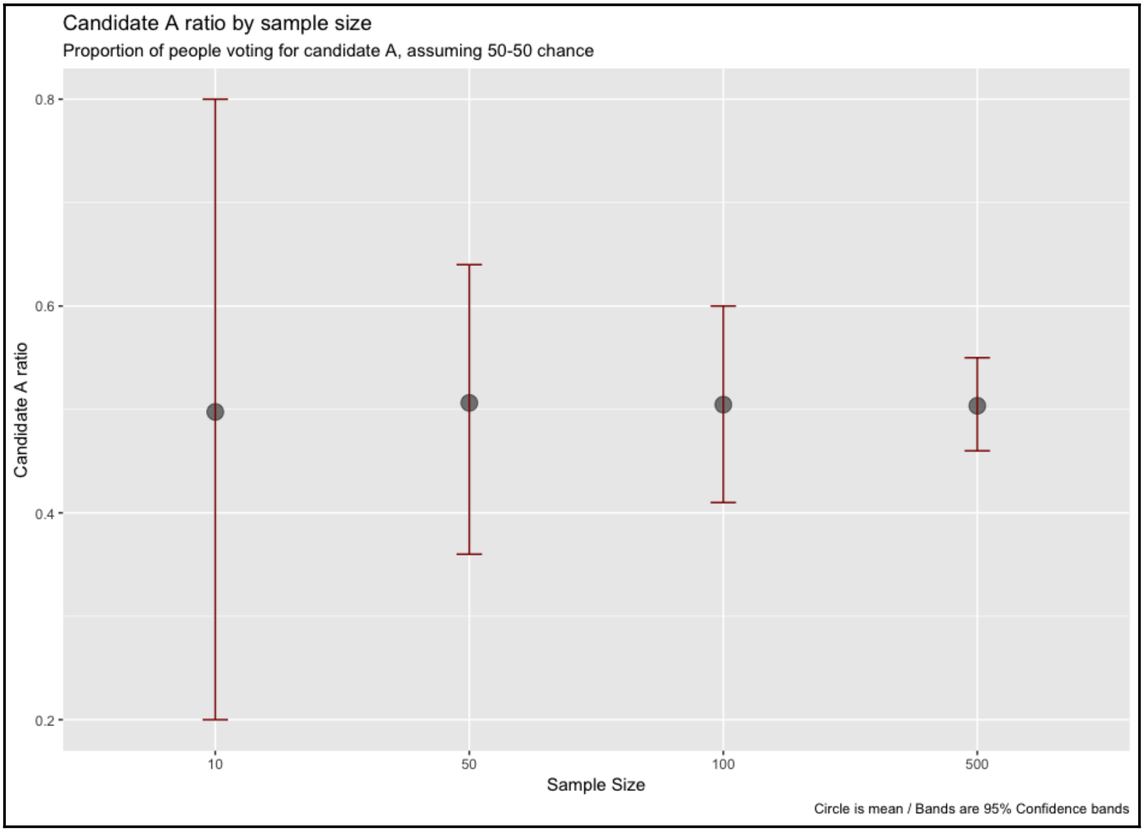


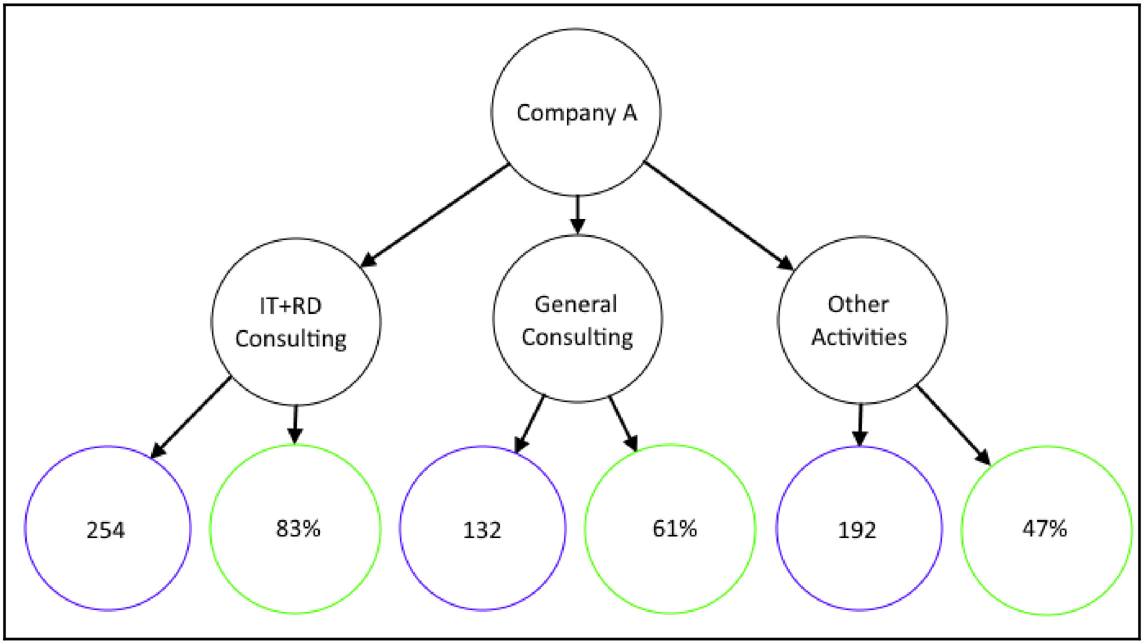
3000
2000
1000
0
count
0
5
10
15
value
method
inverse
rpois


x-y-z plot
z
x
y
2
1
0
-1
-2
-3

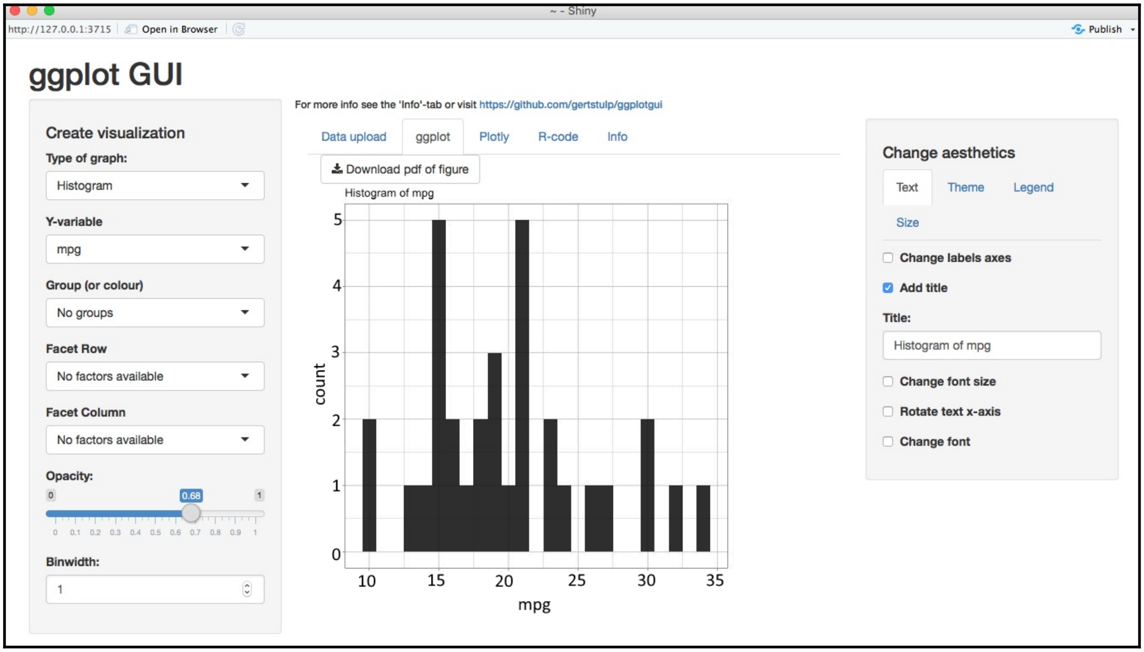
Career	Age	Freq
Egineer	+45	0
Engineer	+45	2
Lawyer	+45	0
Pilot	+45	0
Sales	+45	0
Teacher	+45	1
Egineer	20-30	0
Engineer	20-30	0
Lawyer	20-30	2
Pilot	20-30	0
Sales	20-30	1
Teacher	20-30	1
Egineer	30-45	1
Engineer	30-45	0
Lawyer	30-45	0
Pilot	30-45	1
Sales	30-45	2
Teacher	30-45	1

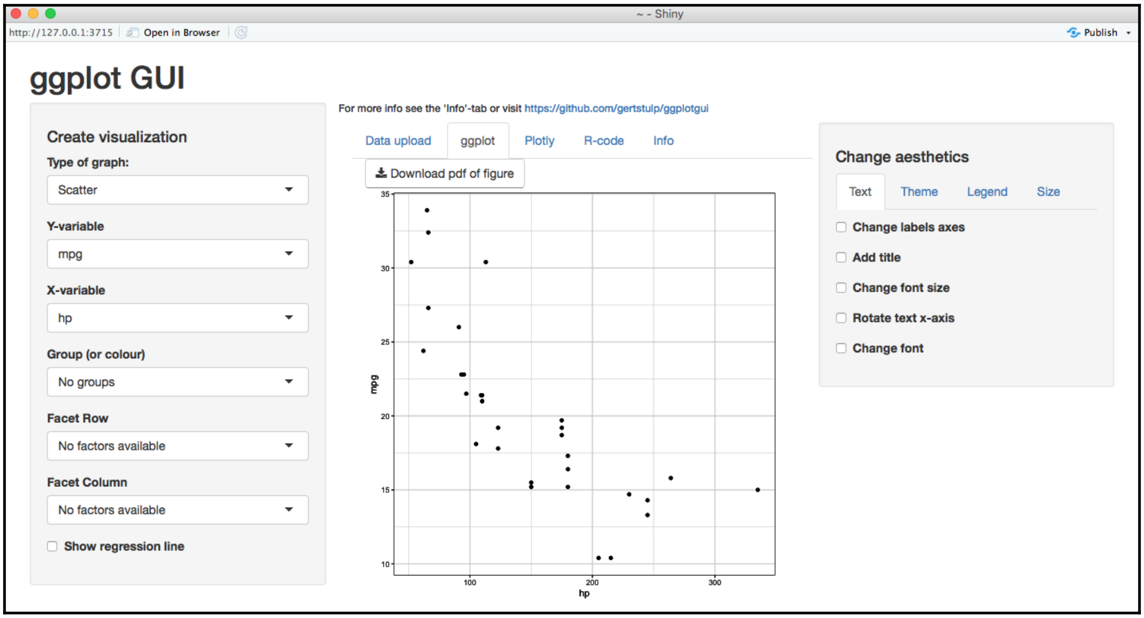
Person	Career	Age	Salary	Contacted
John	Engineer	+45	10	✔ Yes
Mike	Engineer	+45	13	✖ No
Lisa	Teacher	30-45	12	✔ Yes
Marcel	Sales	30-45	18	✔ Yes
Max	Sales	30-45	12	✔ Yes
Paul	Teacher	+45	13	✖ No
Paul	Egineer	30-45	18	✖ No
Anthony	Sales	20-30	13	✖ No
Tom	Lawyer	20-30	19	✖ No
Lisa	Lawyer	20-30	18	✔ Yes
Alex	Teacher	20-30	21	✔ Yes
Tom	Pilot	30-45	23	✔ Yes

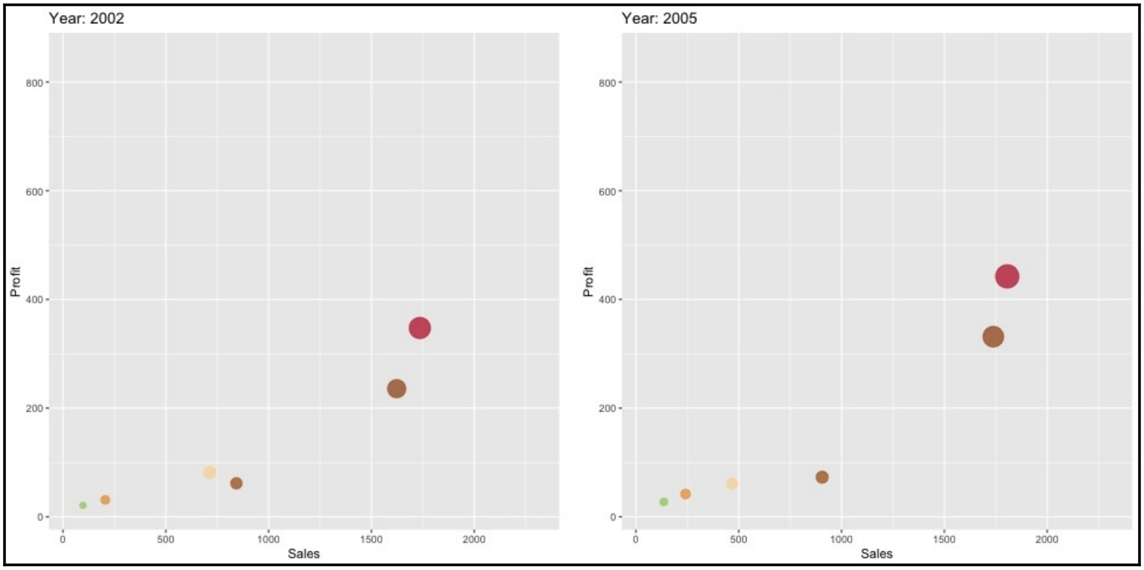

Candidate A ratio by sample size
Proportion of people voting for candidate A, assuming 50-50 chance
0.8
0.6
0.4
0.2
Candidate A ratio
10
50
100
500
Sample Size
Circle is mean / Bands are 95% Confidence bands




```
> start_time <- Sys.time()
> bring_element(some__vector,some__matrix)
Process starting
Process ended
[1] 522081
> end_time <- Sys.time()
> print(end_time - start_time)
Time difference of 0.106581 secs
>
>
> start_time <- Sys.time()
> Rfunc(some__vector,some__matrix)
[1] 522081
> end_time <- Sys.time()
> print(end_time - start_time)
Time difference of 0.2182128 secs
```

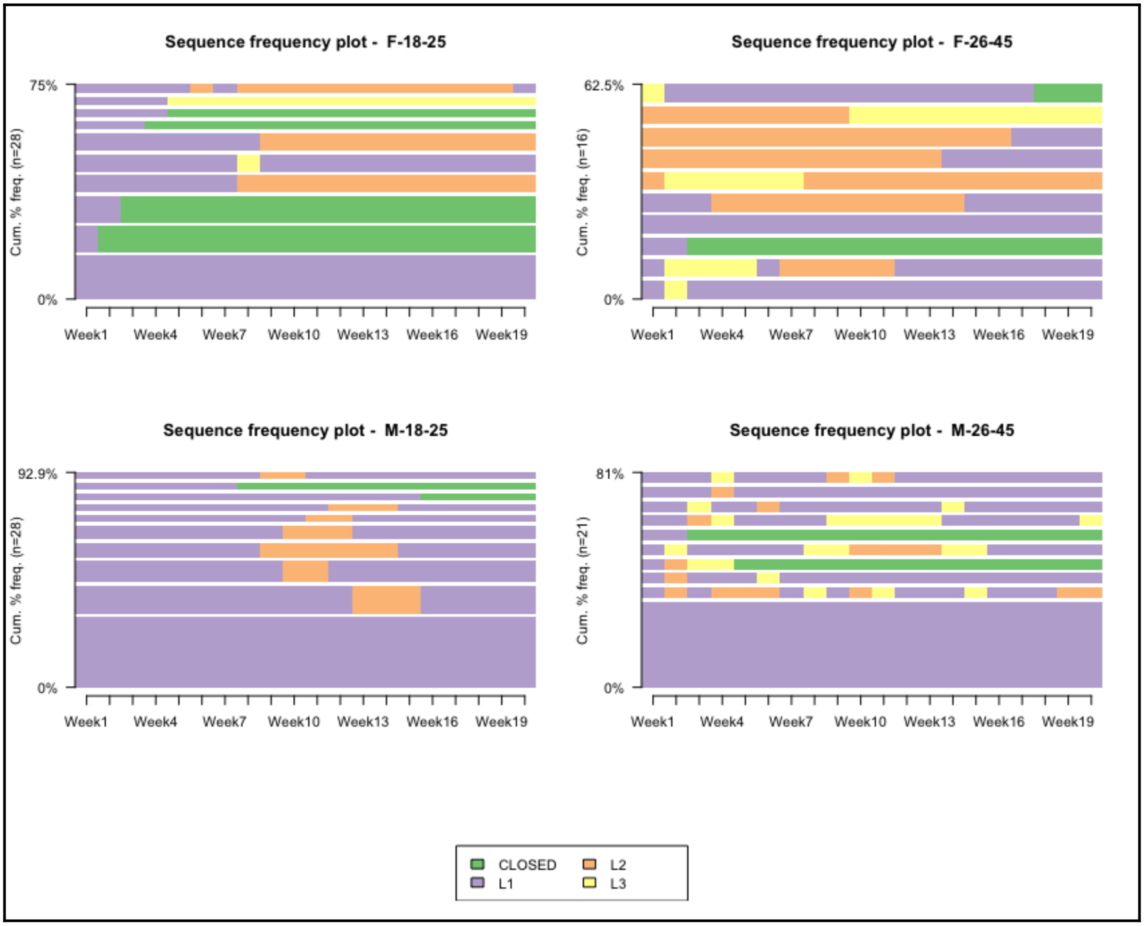

~ - Shiny
http://127.0.0.1:3715
Open in Browser
Publish
ggplot GUI
For more info see the 'Info'-tab or visit https://github.com/gertstulp/ggplotgui
Create visualization
Type of graph:
Histogram
Y-variable
mpg
Group (or colour)
No groups
Facet Row
No factors available
Facet Column
No factors available
Opacity:
0.68
Binwidth:
1
Data upload
ggplot
Plotly
R-code
Info
Download pdf of figure
Histogram of mpg
count
mpg
Change aesthetics
Text
Theme
Legend
Size
Change labels axes
Add title
Title:
Histogram of mpg
Change font size
Rotate text x-axis
Change font

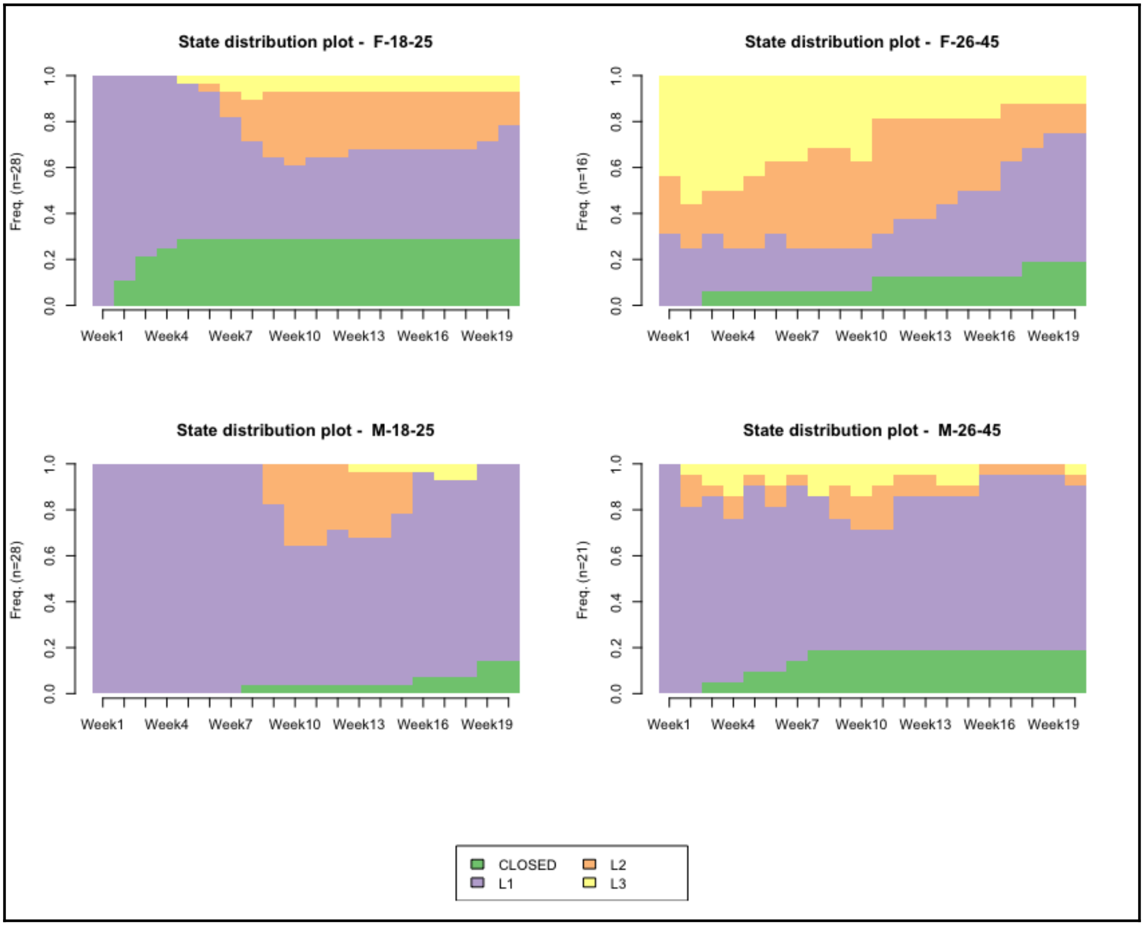

~ - Shiny
http://127.0.0.1:3715
Open in Browser
Publish
ggplot GUI
For more info see the 'Info'-tab or visit https://github.com/gertstulp/ggplotgui
Create visualization
Type of graph:
Scatter
Y-variable
mpg
X-variable
hp
Group (or colour)
No groups
Facet Row
No factors available
Facet Column
No factors available
Show regression line
Data upload
ggplot
Plotly
R-code
Info
Download pdf of figure
mpg
hp
Change aesthetics
Text
Theme
Legend
Size
Change labels axes
Add title
Change font size
Rotate text x-axis
Change font

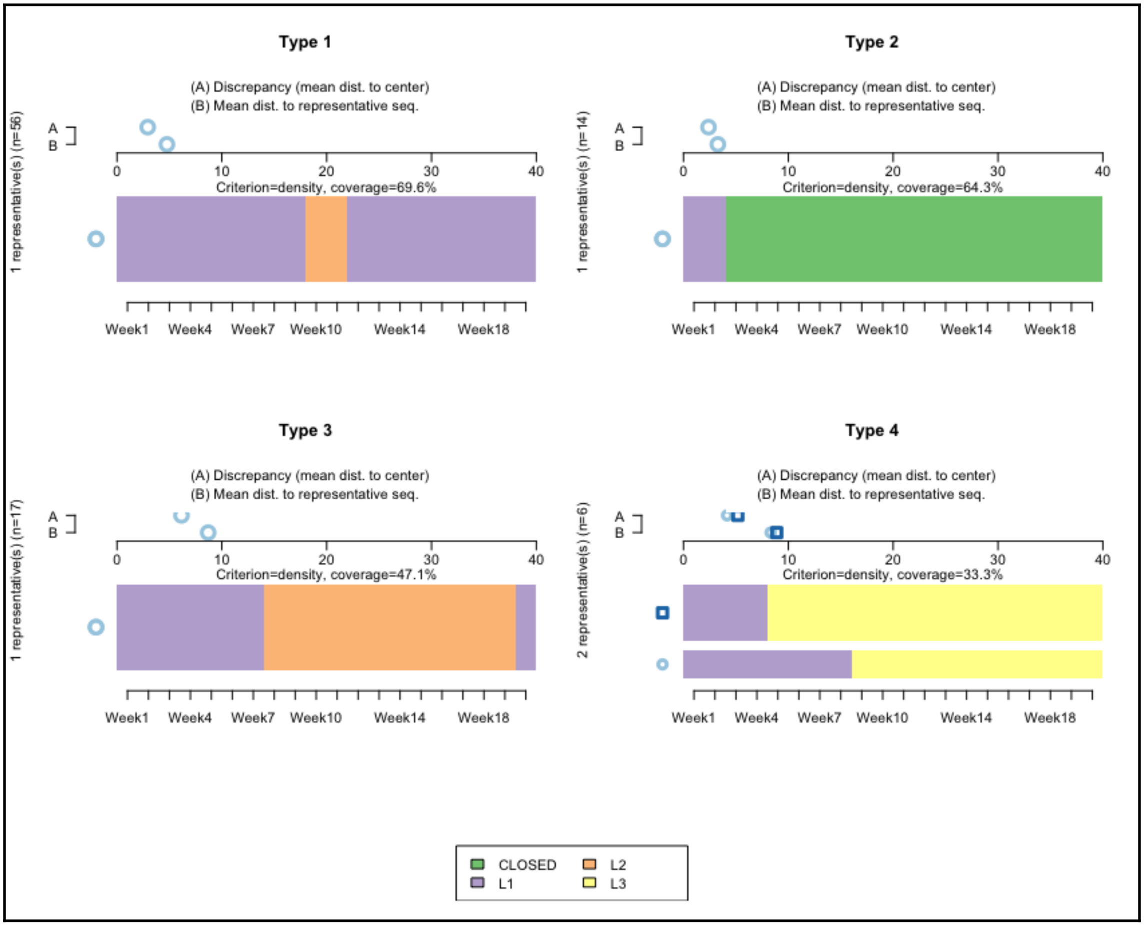


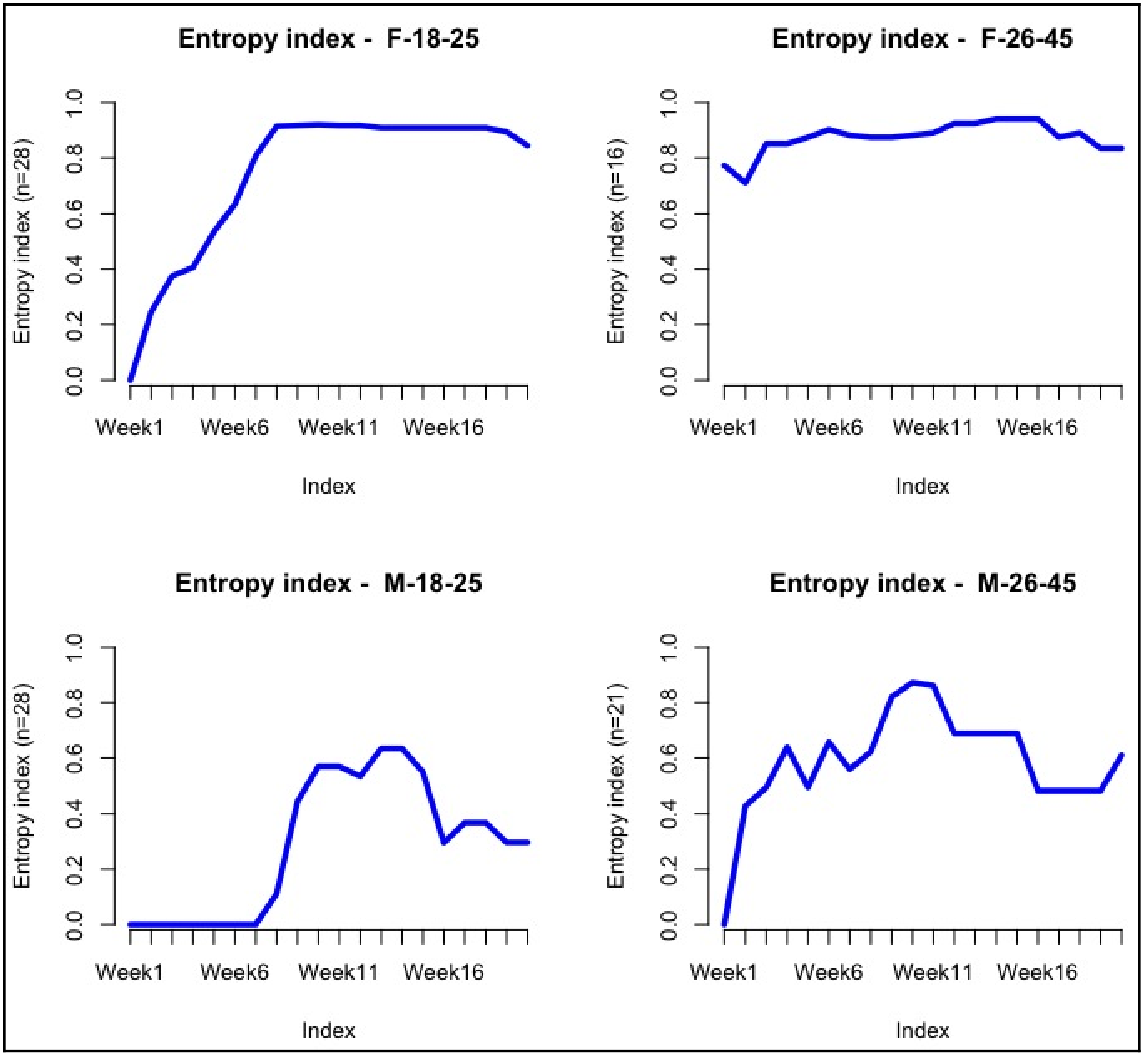

```
> list_of_customers[[1]]$full_print()
[1] "------------------------------------"
[1] "Customer name -> Michaela"
[1] "Customer city -> London"
[1] "------------------------------------"
>
```

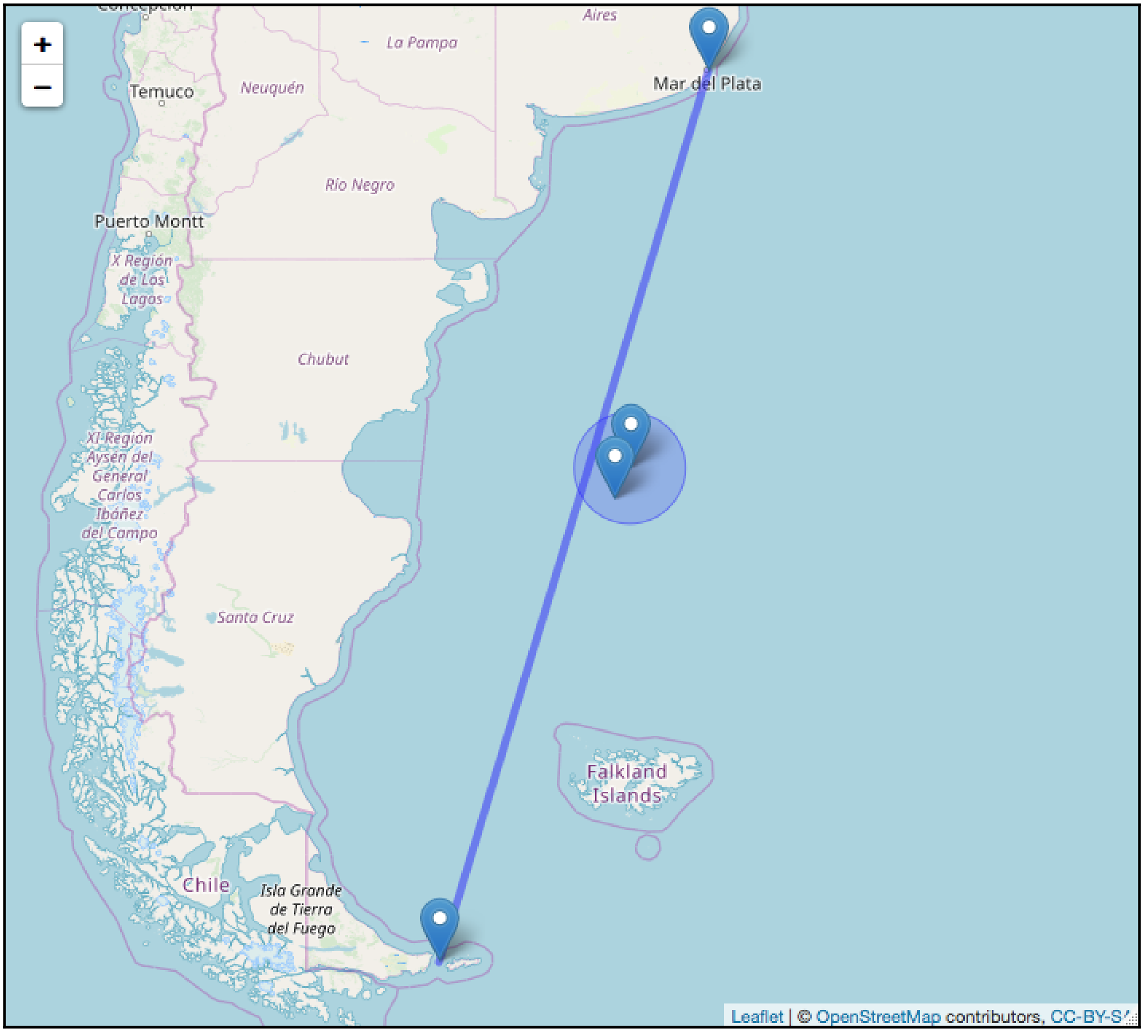
```
> list_of_customers[[1]]$full_print()
[1] "------------------------------------"
[1] "Customer name -> Michaela"
[1] "Customer city -> London"
[1] "Missing prod  -> TV"
[1] "Missing since -> 01/01/2018"
[1] "------------------------------------"
>
```

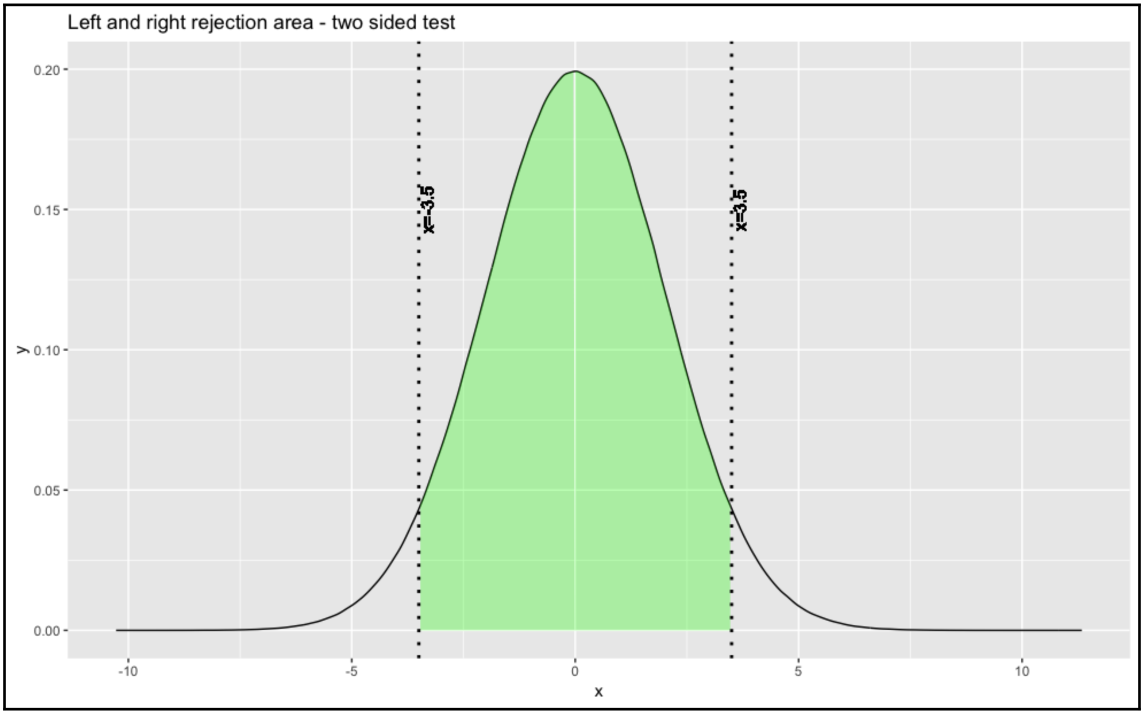

Sequence frequency plot - F-18-25
Sequence frequency plot - F-26-45
Sequence frequency plot - M-18-25
Sequence frequency plot - M-26-45
Cum. % freq. (n=28)
Cum. % freq. (n=16)
Cum. % freq. (n=28)
Cum. % freq. (n=21)
75%
62.5%
92.9%
81%
0%
Week1 Week4 Week7 Week10 Week13 Week16 Week19
CLOSED
L1
L2
L3

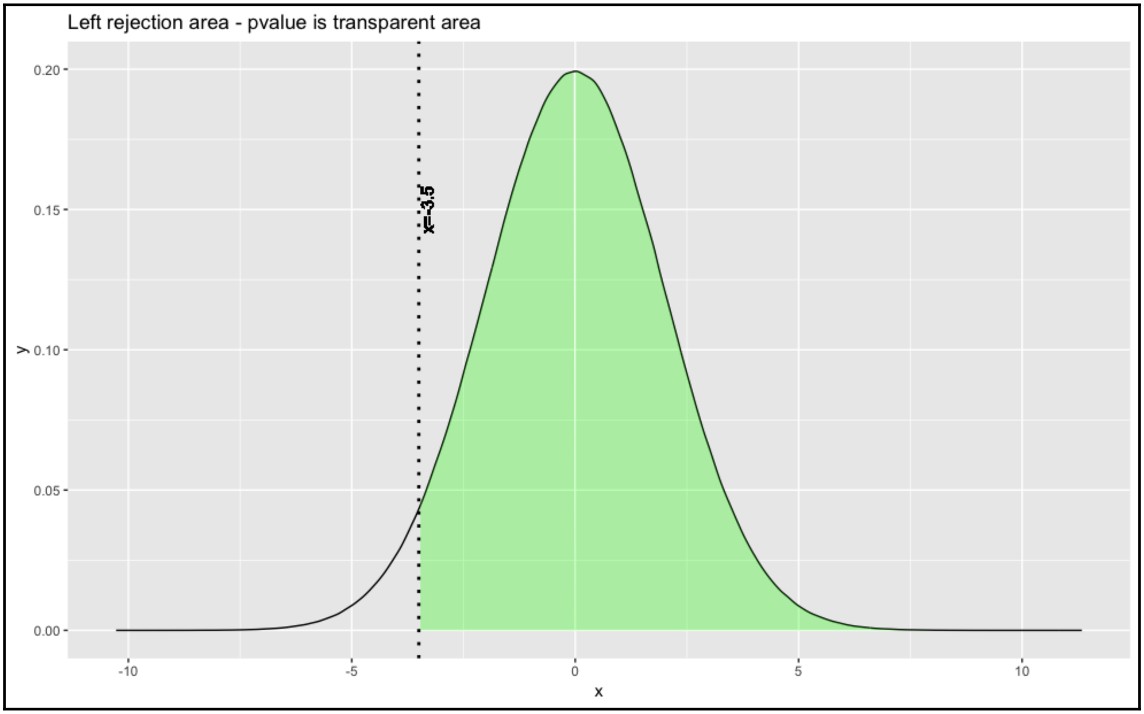

State distribution plot - F-18-25
State distribution plot - F-26-45
State distribution plot - M-18-25
State distribution plot - M-26-45
Freq. (n=28)
Freq. (n=16)
Freq. (n=28)
Freq. (n=21)
0.0
0.2
0.4
0.6
0.8
1.0
Week1
Week4
Week7
Week10
Week13
Week16
Week19
CLOSED
L1
L2
L3

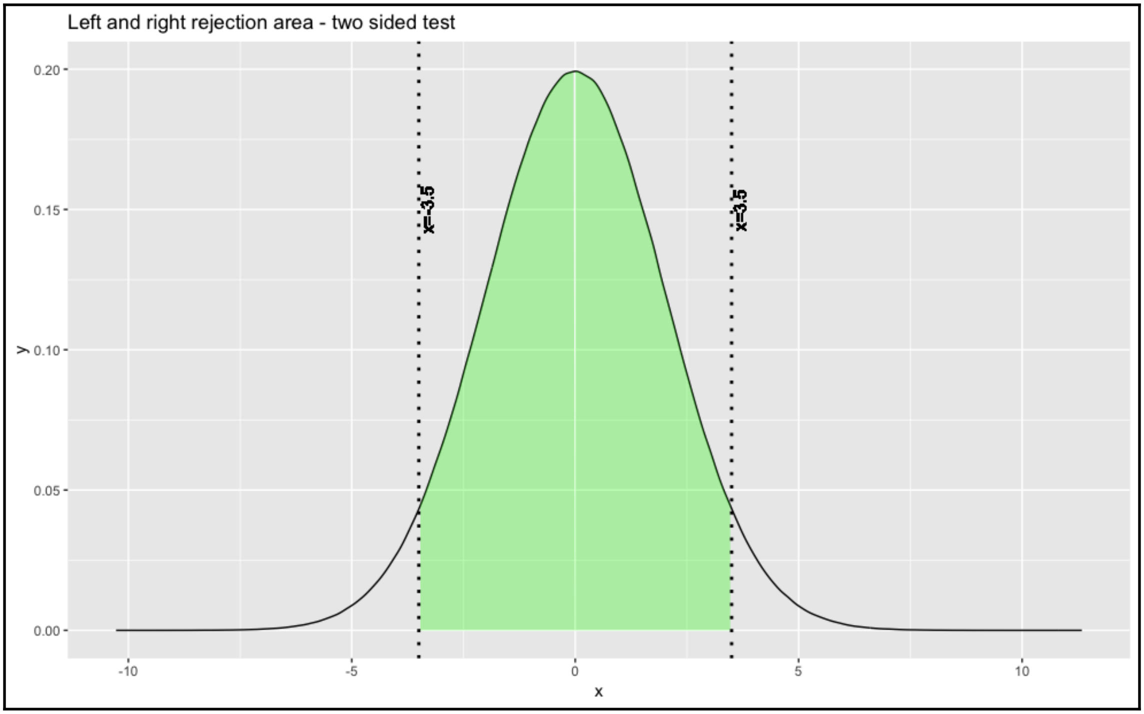

Type 1
(A) Discrepancy (mean dist. to center)
(B) Mean dist. to representative seq.
1 representative(s) (n=56)
Criterion=density, coverage=69.6%
Type 2
(A) Discrepancy (mean dist. to center)
(B) Mean dist. to representative seq.
1 representative(s) (n=14)
Criterion=density, coverage=64.3%
Type 3
(A) Discrepancy (mean dist. to center)
(B) Mean dist. to representative seq.
1 representative(s) (n=17)
Criterion=density, coverage=47.1%
Type 4
(A) Discrepancy (mean dist. to center)
(B) Mean dist. to representative seq.
2 representative(s) (n=6)
Criterion=density, coverage=33.3%
Week1 Week4 Week7 Week10 Week14 Week18
CLOSED
L1
L2
L3


Entropy index - F-18-25
Entropy index (n=28)
0.0
0.2
0.4
0.6
0.8
1.0
Week1
Week6
Week11
Week16
Index
Entropy index - F-26-45
Entropy index (n=16)
Week1
Week6
Week11
Week16
Index
Entropy index - M-18-25
Entropy index (n=28)
Week1
Week6
Week11
Week16
Index
Entropy index - M-26-45
Entropy index (n=21)
Week1
Week6
Week11
Week16
Index


Temuco
Neuquén
La Pampa
Aires
Mar del Plata
Río Negro
Puerto Montt
X Región de Los Lagos
Chubut
XI Región Aysén del General Carlos Ibáñez del Campo
Santa Cruz
Falkland Islands
Chile
Isla Grande de Tierra del Fuego
Leaflet | © OpenStreetMap contributors, CC-BY-S


Left and right rejection area - two sided test
0.20
0.15
0.10
0.05
0.00
y
x=-3.5
x=3.5
-10
-5
0
5
10
x


Left rejection area - pvalue is transparent area
x=-3.5
0.20
0.15
0.10
0.05
0.00
y
-10
-5
0
5
10
x




```
> leveneTest(Height ~ Sample,data)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  0.6526 0.4201
      197
```

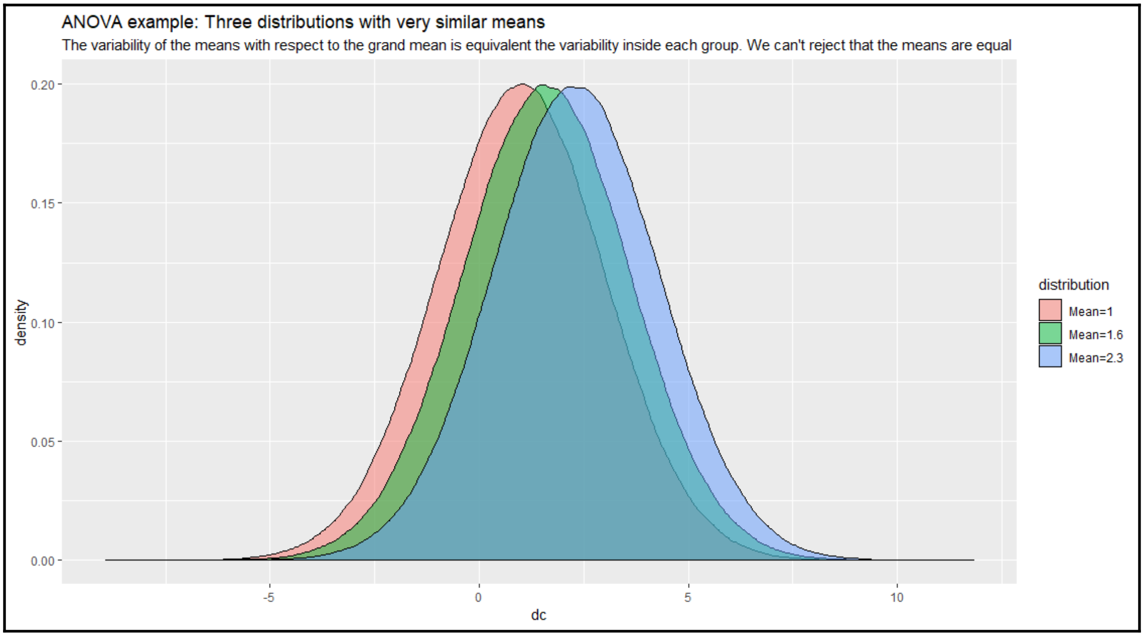
```
> t.test(sample1,sample2,var.equal=TRUE,conf.level = .95,alternative="two.sided")

	Two Sample t-test

data:  sample1 and sample2
t = -0.14195, df = 197, p-value = 0.8873
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.096543  2.680704
sample estimates:
mean of x mean of y
 179.5081  179.7160
```

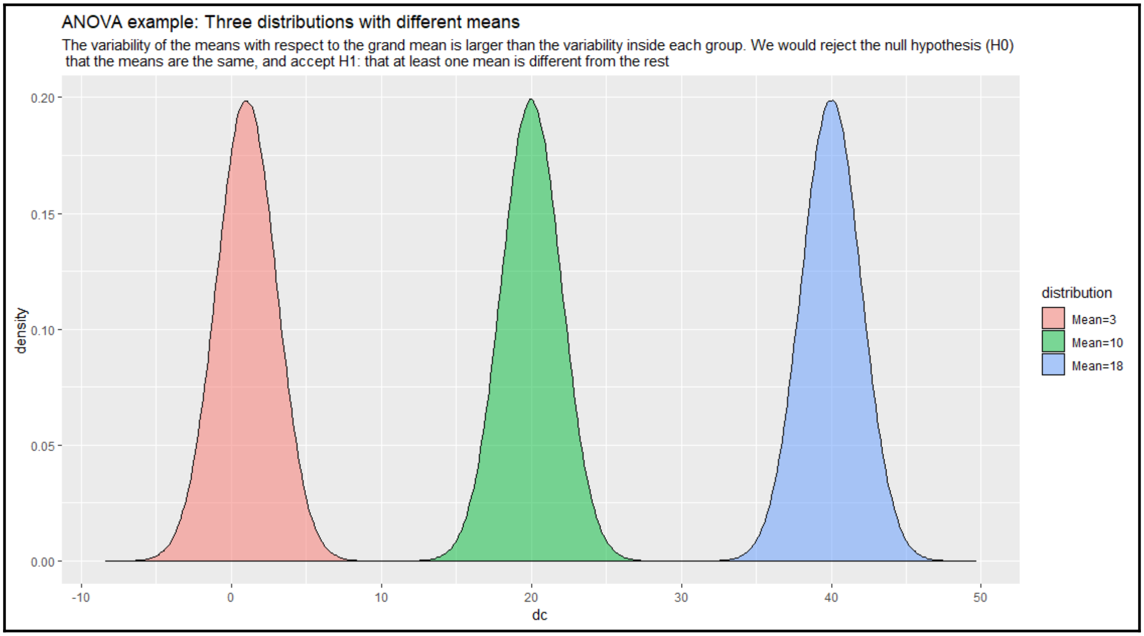
```
> print(paste("n=10 / sd1=2 / sd2=5 / effective alpha=",calc_effective_alphas(10,2,5,TRUE)))
[1] "n=10 / sd1=2 / sd2=5 / effective alpha= 0.05647"
> print(paste("n=10 / sd1=2 / sd2=20/ effective alpha=",calc_effective_alphas(10,2,20,TRUE)))
[1] "n=10 / sd1=2 / sd2=20/ effective alpha= 0.06595"
> print(paste("n=10 / sd1=2 / sd2=5 / effective alpha=",calc_effective_alphas(10,2,5,FALSE)))
[1] "n=10 / sd1=2 / sd2=5 / effective alpha= 0.04978"
> print(paste("n=10 / sd1=2 / sd2=20/ effective alpha=",calc_effective_alphas(10,2,20,FALSE)))
[1] "n=10 / sd1=2 / sd2=20/ effective alpha= 0.05031"
```

```
> print(paste("n=10 / sd1=2 / sd2=20/ effective power=",calc_power(10,2,2,TRUE)))
[1] "n=10 / sd1=2 / sd2=20/ effective power= 0.5658"
> print(paste("n=10 / sd1=2 / sd2=20/ effective power=",calc_power(10,2,2,FALSE)))
[1] "n=10 / sd1=2 / sd2=20/ effective power= 0.55787"
```

```
> t.test(data$post_bonus,data$pre_bonus,conf.level = .95,alternative="greater",paired=TRUE)

	Paired t-test

data:  data$post_bonus and data$pre_bonus
t = 3.676, df = 79, p-value = 0.0002151
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.265567      Inf
sample estimates:
mean of the differences
               2.312674
```

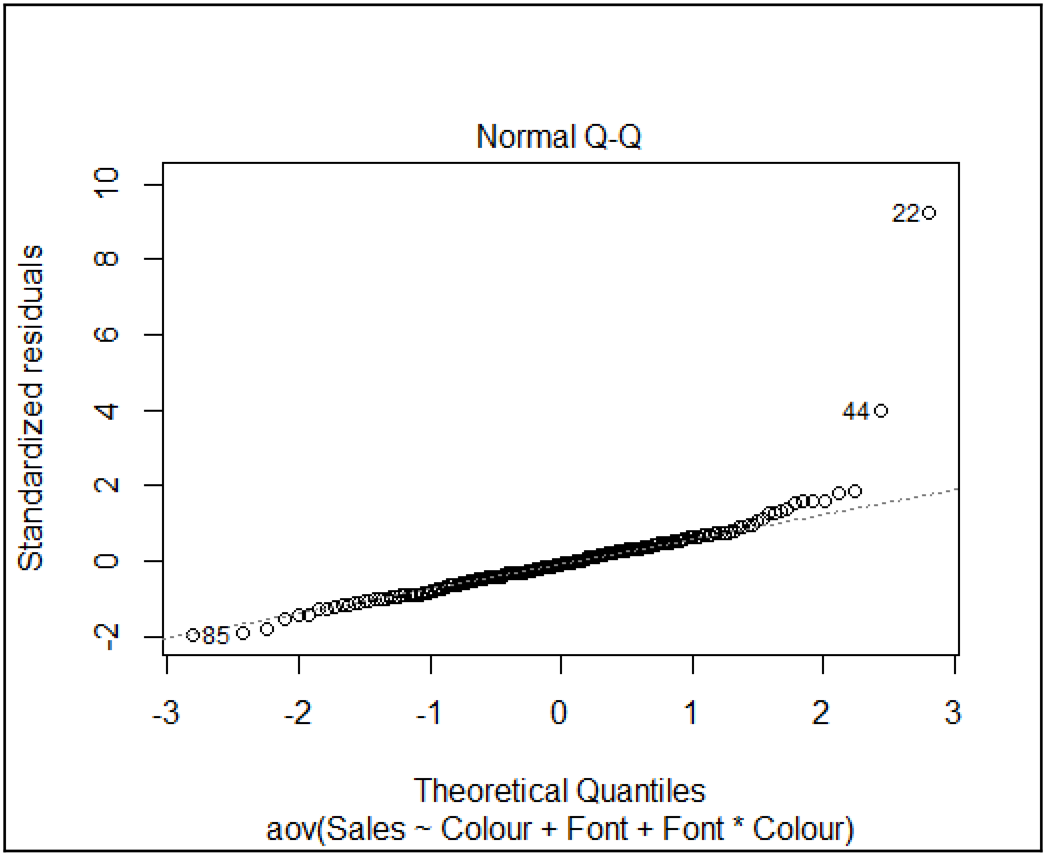
```
> t.test(data$post_bonus,data$pre_bonus,conf.level = .95,alternative="less",paired=TRUE)

	Paired t-test

data:  data$post_bonus and data$pre_bonus
t = 3.676, df = 79, p-value = 0.9998
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 3.359781
sample estimates:
mean of the differences
               2.312674
```



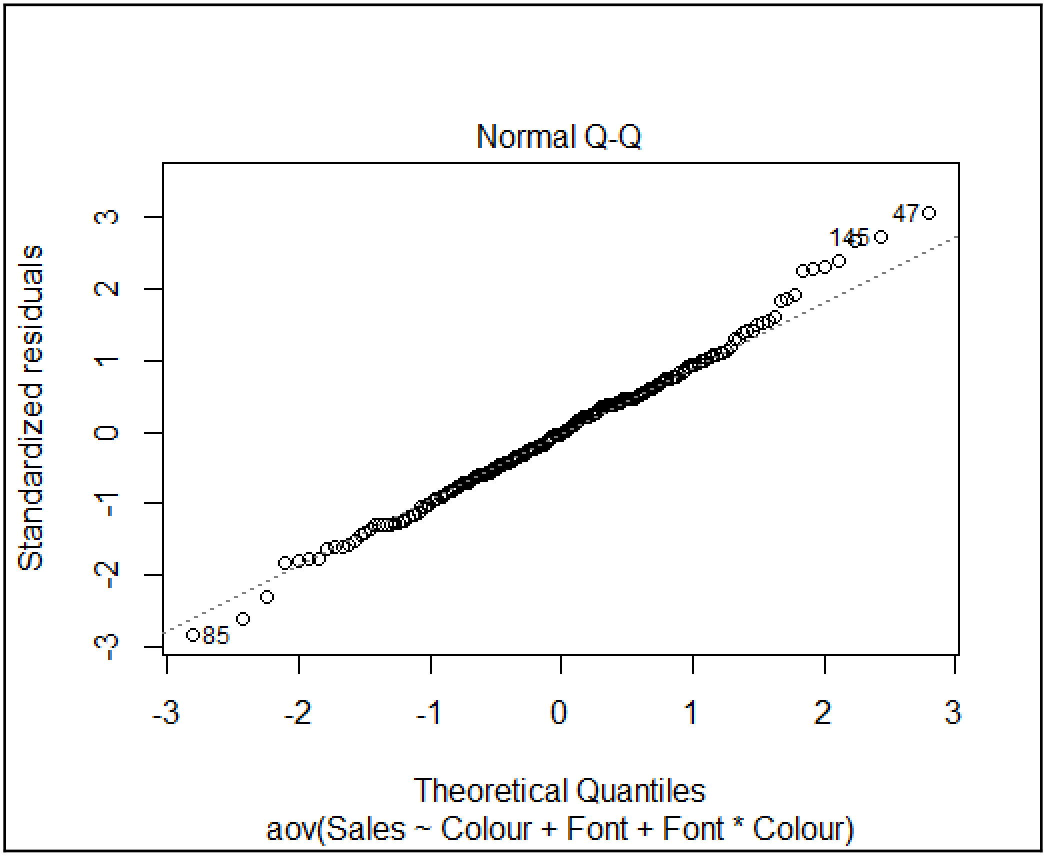



```
> print(paste("SS(ERROR) = ",SS_ERROR))
[1] "SS(ERROR) =  5124.39647632035"
> print(paste("SS(LOT) =",SS_LOT,"/F(LOT) = ",FF_LOT,"pvalue = ",pval_LOT))
[1] "SS(LOT) = 573.680971022944 /F(LOT) =  6.2692522964877 pvalue =  0.0152258042235658"
> print(paste("SS(FOODTYPE) =",SS_FOODTYPE,"/F(FOODTYPE) = ",FF_FOODTYPE,"pvalue = ",pval_FOODTYPE))
[1] "SS(FOODTYPE) = 4038.80007335211 /F(FOODTYPE) =  22.0682381967178 pvalue =  8.56406995319858e-08"
```

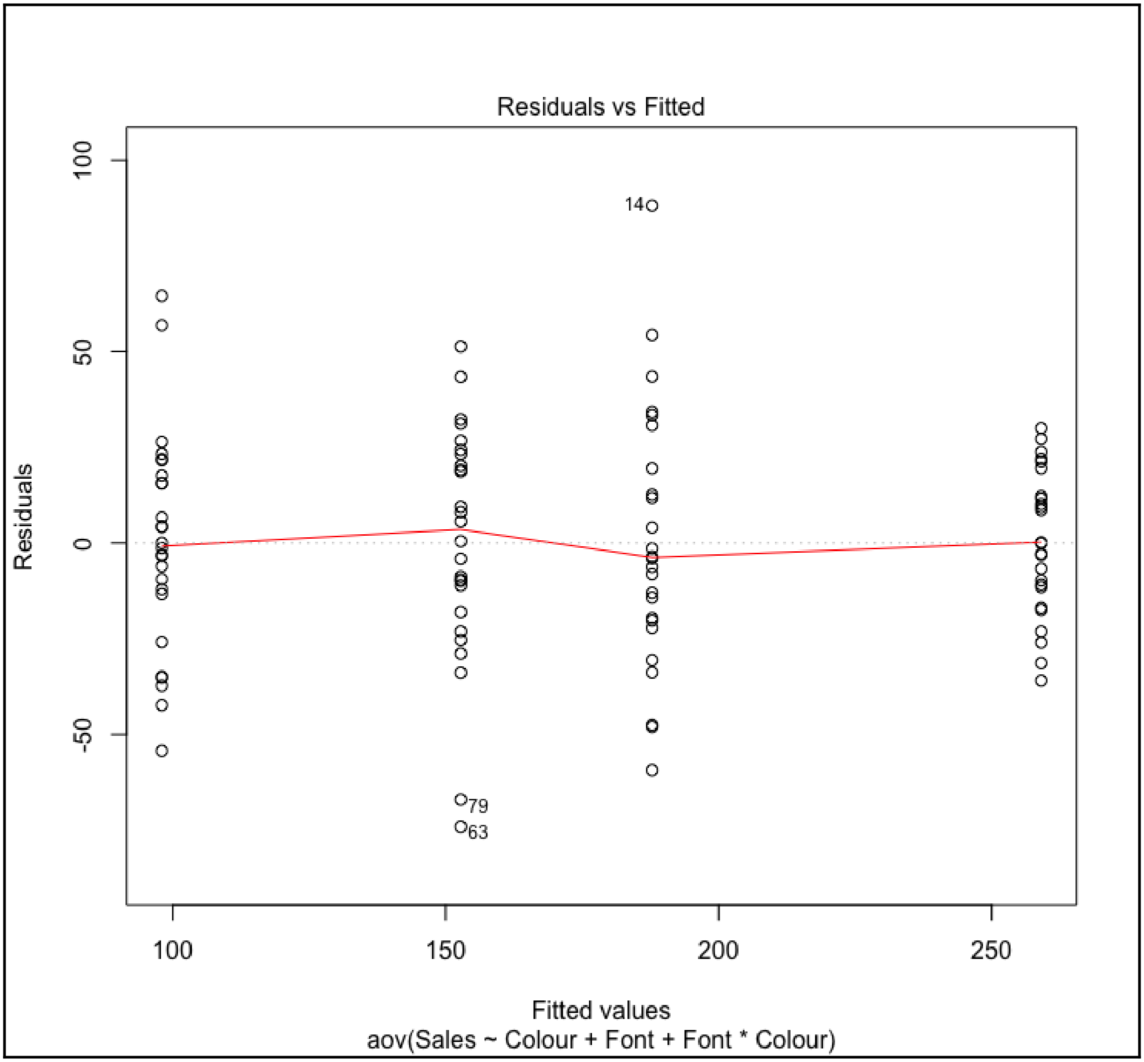
```
> anova(result)
Analysis of Variance Table

Response: Result
          Df Sum Sq Mean Sq F value    Pr(>F)    
Lot        1  573.7  573.68  6.2693   0.01523 *  
Food.Type  2 4038.8 2019.40 22.0682 8.564e-08 ***
Residuals 56 5124.4   91.51                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```




```
> shapiro.test(residuals(d))

        Shapiro-Wilk normality test

data:  residuals(d)
W = 0.7372, p-value < 2.2e-16
```




```
> shapiro.test(residuals(d))

        Shapiro-Wilk normality test

data:  residuals(d)
W = 0.9926, p-value = 0.4141
```


Residuals vs Fitted
Residuals
100
50
0
-50
14
79
63
100
150
200
250
Fitted values
aov(Sales ~ Colour + Font + Font * Colour)

```
> anova(d)
Analysis of Variance Table

Response: Sales
              Df Sum Sq Mean Sq  F value    Pr(>F)    
Colour         1  67573   67573 173.9832 < 2.2e-16 ***
Font           1  18751   18751  48.2786 5.399e-11 ***
Colour:Font    1    280     280   0.7209    0.3969    
Residuals    195  75735     388                       
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> rbind(TukeyHSD(d)$Colour,TukeyHSD(d)$Font)
                    diff       lwr       upr        p adj
Red-Blue       -36.85853 -42.36961 -31.34746 0.000000e+00
Modern-Classic  19.42561  13.91175  24.93947 5.391809e-11
```

```
> summary(type1)
              Df Sum Sq Mean Sq F value  Pr(>F)    
Colour         1  67573   67573 173.983 < 2e-16 ***
Font           1  18751   18751  48.279 5.4e-11 ***
Colour:Font    1    280     280   0.721   0.397    
Residuals    195  75735     388                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

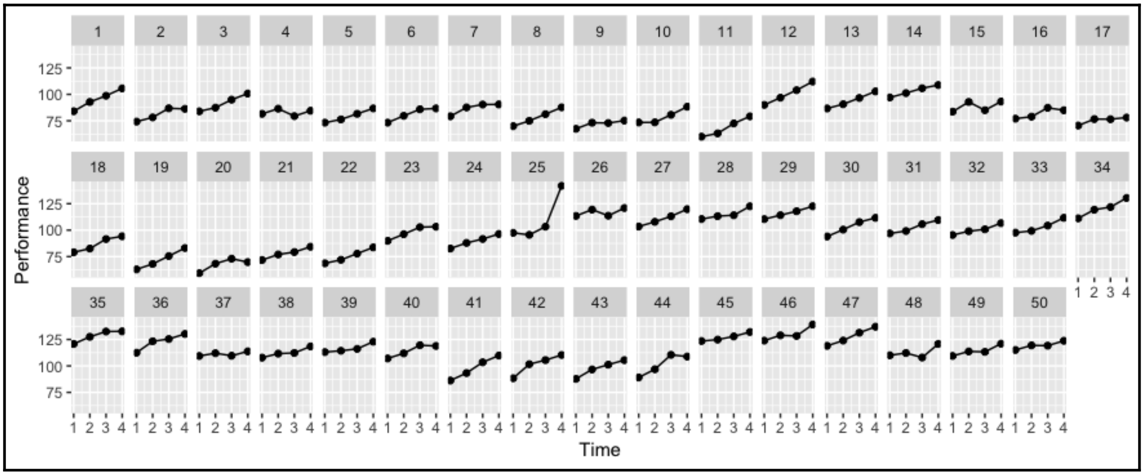
```
> type2
Anova Table (Type II tests)

Response: Sales
          Sum Sq  Df F value    Pr(>F)    
Colour     67967   1 175.248 < 2.2e-16 ***
Font       18751   1  48.347 5.187e-11 ***
Residuals  76015 196                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> type3
Anova Table (Type III tests)

Response: Sales
              Sum Sq  Df   F value    Pr(>F)    
(Intercept) 2105397   1 5420.8923 < 2.2e-16 ***
Colour        68190   1  175.5741 < 2.2e-16 ***
Font          18817   1   48.4498 5.034e-11 ***
Colour:Font     280   1    0.7209    0.3969    
Residuals     75735 195                         
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> summary(E)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: Sales ~ -1 + Strategy + (1 | Client) + (1 | Salesman)
   Data: clients

REML criterion at convergence: 1627.8

Scaled residuals: 
     Min       1Q   Median       3Q      Max 
-2.33945 -0.49886 -0.07755  0.65161  2.40887 

Random effects:
 Groups   Name        Variance Std.Dev.
 Client   (Intercept)  61.47    7.840  
 Salesman (Intercept) 243.51   15.605  
 Residual              73.90    8.596  
Number of obs: 212, groups:  Client, 80; Salesman, 10

Fixed effects:
                        Estimate Std. Error     df t value Pr(>|t|)    
StrategyDiscount_driven   83.712      5.236 10.812   15.99 7.22e-09 ***
StrategyModern            57.135      5.257 10.974   10.87 3.26e-07 ***
StrategyTraditional       98.612      5.131  9.978   19.22 3.27e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            StrtD_ StrtgM
StratgyMdrn 0.885        
StrtgyTrdtn 0.952  0.903 
```

```
> anova(E)
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
Strategy  45775   15258     3 21.537  206.48 5.101e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> print(result_lsmeans)
$`lsmeans`
 Strategy        lsmean   SE   df lower.CL upper.CL
 Discount_driven   83.7 5.24 10.9     72.2     95.3
 Modern            57.1 5.26 11.0     45.6     68.7
 Traditional       98.6 5.13 10.0     87.2    110.0

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast                      estimate   SE    df t.ratio p.value
 Discount_driven - Modern          26.6 2.52 110.5  10.539 <.0001
 Discount_driven - Traditional    -14.9 1.61 148.4  -9.266 <.0001
 Modern - Traditional             -41.5 2.29  86.1 -18.084 <.0001

P value adjustment: tukey method for comparing a family of 3 estimates
```



```
Linear mixed-effects model fit by REML
 Data: data_company
       AIC      BIC    logLik
  1175.053 1204.464 -578.5266

Random effects:
 Formula: ~1 | Employee
        (Intercept) Residual
StdDev:    8.121308 4.708467

Correlation Structure: AR(1)
 Formula: ~Time | Employee
 Parameter estimate(s):
      Phi
0.7055112
Fixed effects: Performance ~ x_Bonus + x_Sector + Time + Time:x_Bonus
                  Value Std.Error  DF  t-value p-value
(Intercept)    84.84882 1.4303763 145 59.31923  0.0000
x_Bonus1        9.82991 1.1033073 145  8.90949  0.0000
x_Sector1     -10.04146 1.8167875 145 -5.52704  0.0000
x_Sector2       0.17172 1.1674012 145  0.14710  0.8833
Time            5.04404 0.2592823 145 19.45387  0.0000
x_Bonus1:Time   0.43554 0.2527656 145  1.72310  0.0870
 Correlation:
              (Intr) x_Bns1 x_Sct1 x_Sct2 Time
x_Bonus1      -0.085
x_Sector1      0.021  0.044
x_Sector2      0.043 -0.027 -0.783
Time          -0.455  0.065  0.001 -0.119
x_Bonus1:Time  0.051 -0.573  0.000 -0.010 -0.113

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.40845899 -0.36621574 -0.01625146  0.42886226  1.70236869

Number of Observations: 200
Number of Groups: 50
```

```
> anova(fit)
             numDF denDF  F-value p-value
(Intercept)      1   145 6079.059  <.0001
x_Bonus          1   145  154.109  <.0001
x_Sector         2   145   19.749  <.0001
Time             1   145  391.135  <.0001
x_Bonus:Time     1   145    2.969   0.087
```

```
Linear mixed-effects model fit by REML
 Data: data_company
       AIC      BIC    logLik
  1195.103 1221.246 -589.5517

Random effects:
 Formula: ~1 | Employee
        (Intercept) Residual
StdDev:    8.580379 3.071642

Fixed effects: Performance ~ x_Bonus + x_Sector + Time + Time:x_Bonus
                 Value Std.Error  DF   t-value p-value
(Intercept)   84.97143 1.3371508 145  63.54663  0.0000
x_Bonus1      10.24409 1.0169992 145  10.07286  0.0000
x_Sector1    -10.00198 1.7639710 145  -5.67015  0.0000
x_Sector2      0.47204 1.1025057 145   0.42815  0.6692
Time           5.02791 0.2023747 145  24.84454  0.0000
x_Bonus1:Time  0.21025 0.1973925 145   1.06514  0.2886
 Correlation:
              (Intr) x_Bns1 x_Sct1 x_Sct2 Time
x_Bonus1      -0.077
x_Sector1      0.022  0.043
x_Sector2      0.046 -0.028 -0.804
Time          -0.381  0.052  0.002 -0.154
x_Bonus1:Time  0.040 -0.485  0.000 -0.014 -0.107

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-5.39626213 -0.46356930  0.02759754  0.55655047  2.84400914

Number of Observations: 200
Number of Groups: 50
```

```
$`lsmeans`
 x_Bonus      lsmean   SE df lower.CL upper.CL
 boost_bonus  108.4 1.50 49    105.4    111.4
 standard      86.6 1.62 49     83.3     89.8

Results are averaged over the levels of: x_Sector
d.f. method: containment
Confidence level used: 0.95

$contrasts
 contrast                 estimate   SE  df t.ratio p.value
 boost_bonus - standard       21.8 1.81 146 12.049  <.0001

Results are averaged over the levels of: x_Sector
```

```
Linear mixed-effects model fit by REML
 Data: data_company
       AIC      BIC    logLik
  1176.371 1212.317 -577.1855

Random effects:
 Formula: ~1 + Time | Employee
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev   Corr
(Intercept) 9.893662 (Intr)
Time        1.090883 -0.609
Residual    3.493198

Correlation Structure: AR(1)
 Formula: ~Time | Employee
 Parameter estimate(s):
      Phi
0.5151983
Fixed effects: Performance ~ x_Bonus + x_Sector + Time + Time:x_Bonus
                 Value Std.Error  DF  t-value p-value
(Intercept)   84.87072 1.5530358 145 54.64827  0.0000
x_Bonus1       9.55327 1.1877900 145  8.04289  0.0000
x_Sector1     -9.69168 1.7532483 145 -5.52784  0.0000
x_Sector2     -0.25139 1.1217456 145 -0.22411  0.8230
Time           5.04330 0.2719045 145 18.54805  0.0000
x_Bonus1:Time  0.56062 0.2613440 145  2.14516  0.0336
 Correlation:
              (Intr) x_Bns1 x_Sct1 x_Sct2 Time
x_Bonus1      -0.083
x_Sector1      0.022  0.033
x_Sector2      0.041 -0.020 -0.768
Time          -0.611  0.074 -0.005 -0.115
x_Bonus1:Time  0.058 -0.664  0.009 -0.016 -0.112

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-3.126789886 -0.395687859  0.004531282  0.459027846  2.082166827

Number of Observations: 200
Number of Groups: 50
```

```
> sapply(class1,mean)
    Math  History  Physics
6.824406 5.000136 4.893103
> sapply(class2,mean)
    Math  History  Physics
5.949069 5.873906 5.754169
```

```
> print(test_hotelling)
Test stat:  9.0882
Numerator df:  3
Denominator df:  96
P-value:  2.358e-05
```

```
> boxM(cbind(combined$Math,combined$History,combined$Sociology)~group,data = combined)

        Box's M-test for Homogeneity of Covariance Matrices

data:  Y
Chi-Sq (approx.) = 5.9204, df = 3, p-value = 0.1155
```

```
> summary(result)
           Df  Pillai approx F num Df den Df    Pr(>F)
class       1 0.73891   137.74      3    146 < 2.2e-16 ***
Residuals 148
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> summary.aov(result)
 Response History :
             Df  Sum Sq Mean Sq F value    Pr(>F)
class         1 10118.4 10118.4  396.98 < 2.2e-16 ***
Residuals   148  3772.3    25.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response Math :
             Df  Sum Sq Mean Sq F value Pr(>F)
class         1   4.488  4.4879  2.3414 0.1281
Residuals   148 283.685  1.9168

 Response Biology :
             Df  Sum Sq Mean Sq F value Pr(>F)
class         1   0.064 0.06407  0.0323 0.8576
Residuals   148 293.421 1.98258
```

```
> print(SS)
$class
            History        Math      Biology
History 10118.41365 213.0971610 -25.46234256
Math      213.09716   4.4878972  -0.53624541
Biology   -25.46234  -0.5362454   0.06407436

$Residuals
          History     Math  Biology
History 3772.2972 230.4865 229.5842
Math     230.4865 283.6854 160.5399
Biology  229.5842 160.5399 293.4213
```

```
> summary(result)
           Df  Pillai approx F num Df den Df    Pr(>F)    
class       1 0.73891   137.74      3    146 < 2.2e-16 ***
Residuals 148                                             
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Call:
lm(formula = depvar ~ sim_data1 + sim_data2, data = model_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-58.429 -13.158   0.645  13.329  64.036 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 38.21207    1.59951   23.89   <2e-16 ***
sim_data1    0.99035    0.03020   32.79   <2e-16 ***
sim_data2    1.01592    0.02082   48.79   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.08 on 997 degrees of freedom
Multiple R-squared:  0.9158,	Adjusted R-squared:  0.9157 
F-statistic:  5424 on 2 and 997 DF,  p-value: < 2.2e-16
```

```
> beta
              depvar
intercept 41.0854952
sim_data1  0.9928917
sim_data2  0.9958641
```

```
> head(predictions)
        depvar
[1,] 222.4471
[2,] 172.8185
[3,] 150.7092
[4,] 229.6940
[5,] 114.6576
[6,] 118.9390
```

```
> print(paste("Std Error:",diag(cov_matrix)))
[1] "Std Error: 1.78487972893079"    "Std Error: 0.0316733298440072" "Std Error: 0.023208226271635"
```

	Sales		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	22.92	19.87 – 25.97	**<0.001**
Offer_discount	-0.73	-2.77 – 1.31	0.485
Client	-0.28	-1.01 – 0.46	0.465
Michael	7.19	4.69 – 9.69	**<0.001**
Tom	13.76	11.26 – 16.26	**<0.001**
Observations	51		
R^2 / adjusted R^2	0.720 / 0.696		

	Sales		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	29.90	21.74 – 38.07	**<0.001**
Offer_discount	-0.74	-2.78 – 1.30	0.476
Client	-0.28	-1.01 – 0.46	0.463
Random Effects			
σ^2	13.77		
$\tau_{00\ \mathrm{Salesman}}$	46.56		
ICC $_{\mathrm{Salesman}}$	0.77		
Observations	51		
Marginal R^2 / Conditional R^2	0.005 / 0.773		

Clients dataset

Client	*Salesman*	*Strategy*	*Sales*
1	John	Classic	21.2901894915695
2	John	Classic	28.0594553276812
2	John	Classic	21.1247966461804
3	John	Classic	22.7955063753133
3	John	Classic	12.8519131030086
4	John	Classic	22.1577007145641
4	John	Classic	24.3990014648445
5	John	Classic	27.4734552070313
1	Michael	Classic	28.0201688368905
1	Michael	Classic	30.4597384141886
2	Michael	Classic	30.4698635798089
2	Michael	Classic	28.7626171676902

```
> findLinearCombos(X)
$linearCombos
$linearCombos[[1]]
[1] 6 4

$linearCombos[[2]]
[1] 7 3 4 5



$remove
[1] 6 7
```

```
> det(t(X) %*% X)
[1] 0
```

```
Call:
lm(formula = Sales ~ women_apparel_price + male_apparel_price +
    shoes_female_price + shoes_male_price + shoes_kids_prices +
    shoes_male_price_b + prices_shoes, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-44.066 -11.621  -0.917  13.037  32.518

Coefficients: (2 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -14.6255    14.9773  -0.977  0.33208
women_apparel_price   0.8027     0.3428   2.342  0.02197 *
male_apparel_price    1.2161     0.3420   3.556  0.00067 ***
shoes_female_price    5.5924     1.0756   5.199  1.8e-06 ***
shoes_male_price     11.3961     0.3525  32.332  < 2e-16 ***
shoes_kids_prices     6.2991     1.0405   6.054  5.8e-08 ***
shoes_male_price_b        NA         NA      NA       NA
prices_shoes              NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.74 on 72 degrees of freedom
Multiple R-squared:  0.9818,    Adjusted R-squared:  0.9805
F-statistic: 775.3 on 5 and 72 DF,  p-value: < 2.2e-16
```

```
> det(t(X[,c(-6,-7)]) %*% X[,c(-6,-7)])
[1] 2.182987e+19
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-85.359 -25.993  -3.738  20.830  88.742 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         37.82833   15.48058   2.444   0.0164 *  
women_apparel_price  0.27156    0.15078   1.801   0.0749 .  
male_apparel_price   0.08271    0.14102   0.587   0.5589    
shoes_female_price   0.09460    0.01341   7.052 3.07e-10 ***
shoes_male_price     0.07854    0.01358   5.782 9.81e-08 ***
shoes_kids_prices    0.11604    0.01611   7.203 1.51e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.66 on 93 degrees of freedom
Multiple R-squared:  0.712,	Adjusted R-squared:  0.6965 
F-statistic: 45.98 on 5 and 93 DF,  p-value: < 2.2e-16
```

```
> vif(fixedmodel)
women_apparel_price  male_apparel_price  shoes_female_price    shoes_male_price   shoes_kids_prices 
          26.375024           26.358965           10.229845            1.065036           10.287159 
```

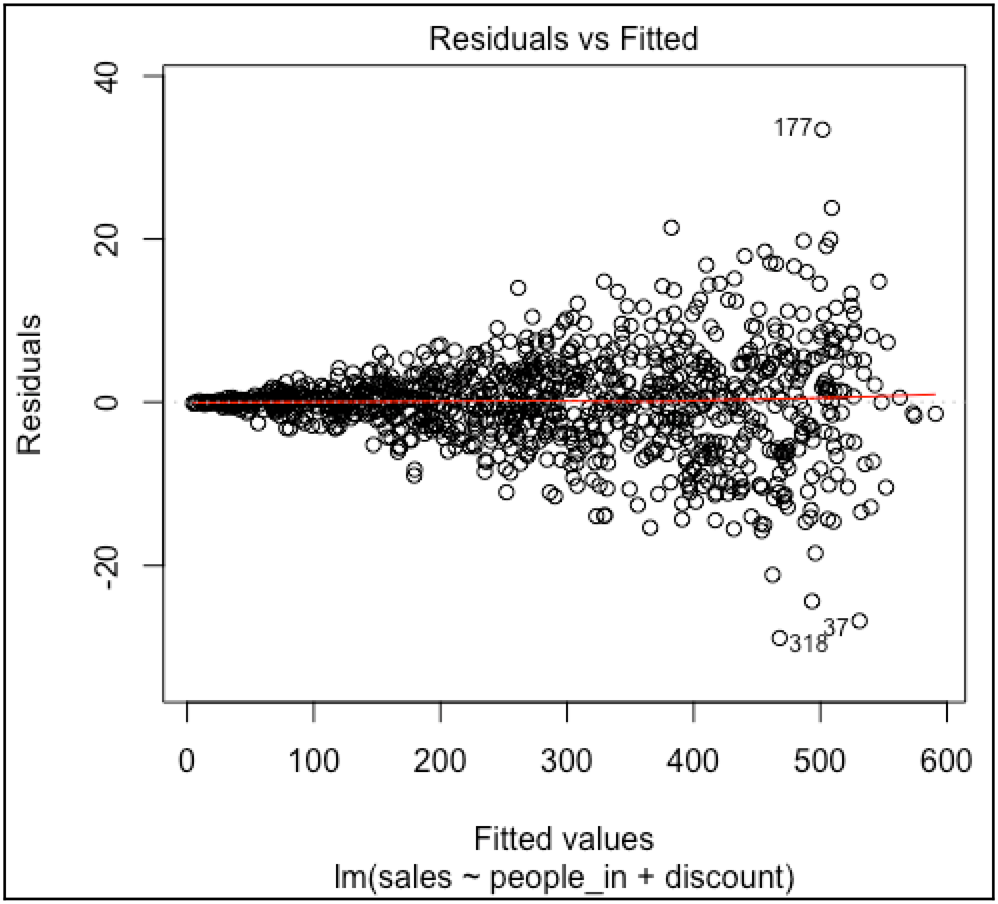
```
Call:
lm(formula = Sales ~ aggregated_apparel + shoes_male_price +
    aggregated_femalekids, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-43.195 -12.076  -0.985  12.703  33.518

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -8.36624   11.73747  -0.713    0.478
aggregated_apparel     1.01046    0.03316  30.472   <2e-16 ***
shoes_male_price      11.39867    0.34866  32.692   <2e-16 ***
aggregated_femalekids  5.94341    0.17072  34.813   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.56 on 74 degrees of freedom
Multiple R-squared:  0.9816,	Adjusted R-squared:  0.9809
F-statistic:  1319 on 3 and 74 DF,  p-value: < 2.2e-16
```

```
> vif(finalmodel)
   aggregated_apparel      shoes_male_price aggregated_femalekids
             1.000553              1.064195              1.064043
```

```
Call:
lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-18.4347  -4.6414  -0.0136   4.4449  23.6712

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -16.7063     2.5520  -6.546 1.54e-10 ***
size               5.6270     0.1068  52.666  < 2e-16 ***
number.bathrooms   1.1513     0.1457   7.904 1.91e-14 ***
number.bedrooms    1.4115     0.1758   8.028 7.86e-15 ***
number.entrances   0.6621     0.2079   3.186  0.00154 **
size_balcony       1.3693     0.2137   6.408 3.56e-10 ***
size_entrance      0.4055     0.0906   4.476 9.54e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.491 on 474 degrees of freedom
Multiple R-squared:  0.8612,    Adjusted R-squared:  0.8594
F-statistic: 490.1 on 6 and 474 DF,  p-value: < 2.2e-16
```

```
> summary(glht(model,linfct = c("number.bathrooms + number.entrances + number.bedrooms + size_balcony + size_entrance - size  = 0")))

         Simultaneous Tests for General Linear Hypotheses

Fit: lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data)

Linear Hypotheses:
                                                                                           Estimate Std. Error t value Pr(>|t|)
number.bathrooms + number.entrances + number.bedrooms + size_balcony + size_entrance - size == 0  -0.6272     0.3845  -1.631    0.104
(Adjusted p values reported -- single-step method)
```

```
> summary(glht(model,linfct = c("number.entrances + number.bathrooms - size_balcony - size_entrance = 0")))

	 Simultaneous Tests for General Linear Hypotheses

Fit: lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data)

Linear Hypotheses:
                                                                    Estimate Std. Error t value Pr(>|t|)
number.entrances + number.bathrooms - size_balcony - size_entrance == 0  0.03864    0.33817   0.114    0.909
(Adjusted p values reported -- single-step method)
```



```
> bptest(model)

        studentized Breusch-Pagan test

data:  model
BP = 149.49, df = 2, p-value < 2.2e-16
```


```
> bptest(model)

        studentized Breusch-Pagan test

data:  model
BP = 149.49, df = 2, p-value < 2.2e-16
```

```
Call:
lm(formula = sales ~ people_in + discount, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-28.885  -2.557   0.108   2.635  33.422

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.13544    0.43329   0.313    0.755
people_in    10.00003    0.01281 780.793   <2e-16 ***
discount    179.33456    1.31497 136.379   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.876 on 997 degrees of freedom
Multiple R-squared:  0.9984,    Adjusted R-squared:  0.9984
F-statistic: 3.129e+05 on 2 and 997 DF,  p-value: < 2.2e-16
```

```
t test of coefficients:

              Estimate Std. Error  t value Pr(>|t|)
(Intercept)   0.135440   0.313656   0.4318    0.666
people_in    10.000028   0.014173 705.5933   <2e-16 ***
discount    179.334557   1.324899 135.3571   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
  mindex  n                                                    predictors   rsquare      adjr   predrsq       cp      aic     sbic      sbc     msep
1   8100 11          crim zn chas nox rm dis rad tax ptratio black lstat 0.7405823 0.7348058 0.7214716 10.11455 3023.726 1588.436 3078.671 22.97774
2   8178 12    crim zn indus chas nox rm dis rad tax ptratio black lstat 0.7406412 0.7343282 0.7207760 12.00275 3025.611 1590.384 3084.783 23.06590
3   8179 12      crim zn chas nox rm age dis rad tax ptratio black lstat 0.7405837 0.7342694 0.7196558 12.11176 3025.724 1590.490 3084.895 23.07102
4   8191 13 crim zn indus chas nox rm age dis rad tax ptratio black lstat 0.7406427 0.7337897 0.7189544 14.00000 3027.609 1592.438 3091.007 23.15973
5   7814 10               crim zn nox rm dis rad tax ptratio black lstat 0.7352631 0.7299149 0.7187463 18.20493 3031.997 1596.197 3082.715 23.35414
6   8101 11         crim zn indus nox rm dis rad tax ptratio black lstat 0.7354930 0.7296032 0.7181603 19.76882 3033.557 1597.793 3088.502 23.42852
       fpe       apc        hsp
1 22.96389 0.2720210 0.04550083
2 23.04966 0.2730369 0.04567542
3 23.05476 0.2730974 0.04568554
4 23.14088 0.2741175 0.04586121
5 23.34226 0.2765029 0.04624618
6 23.41440 0.2773575 0.04639347
```

```
                          Model Summary
-----------------------------------------------------------------
R                         0.861       RMSE                 4.736
R-Squared                 0.741       Coef. Var           21.019
Adj. R-Squared            0.735       MSE                 22.432
Pred R-Squared            0.721       MAE                  3.272
-----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
```

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	36.341	5.067		7.171	0.000	26.385	46.298
lstat	-0.523	0.047	-0.406	-11.019	0.000	-0.616	-0.429
rm	3.802	0.406	0.290	9.356	0.000	3.003	4.600
ptratio	-0.947	0.129	-0.223	-7.334	0.000	-1.200	-0.693
dis	-1.493	0.186	-0.342	-8.037	0.000	-1.858	-1.128
nox	-17.376	3.535	-0.219	-4.915	0.000	-24.322	-10.430
chas	2.719	0.854	0.075	3.183	0.002	1.040	4.397
black	0.009	0.003	0.092	3.475	0.001	0.004	0.015
zn	0.046	0.014	0.116	3.390	0.001	0.019	0.072
crim	-0.108	0.033	-0.101	-3.307	0.001	-0.173	-0.044
rad	0.300	0.063	0.284	4.726	0.000	0.175	0.424
tax	-0.012	0.003	-0.216	-3.493	0.001	-0.018	-0.005

Selection Summary

Step	Variable Entered	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	lstat	0.5441	0.5432	362.7530	3288.9750	6.2158
2	rm	0.6386	0.6371	185.6474	3173.5423	5.5403
3	ptratio	0.6786	0.6767	111.6489	3116.0973	5.2294
4	dis	0.6903	0.6878	91.4853	3099.3590	5.1386
5	nox	0.7081	0.7052	59.7536	3071.4386	4.9939
6	chas	0.7158	0.7124	47.1754	3059.9390	4.9326
7	black	0.7222	0.7183	37.0589	3050.4384	4.8818
8	zn	0.7266	0.7222	30.6240	3044.2750	4.8474
9	crim	0.7288	0.7239	28.4179	3042.1546	4.8326
10	rad	0.7342	0.7288	20.2658	3034.0687	4.7895
11	tax	0.7406	0.7348	10.1145	3023.7264	4.7362

```
Final Model Output
------------------

                        Model Summary
---------------------------------------------------------------
R                       0.861       RMSE                4.736
R-Squared               0.741       Coef. Var          21.019
Adj. R-Squared          0.735       MSE                22.432
Pred R-Squared          0.721       MAE                 3.272
---------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
```

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	36.341	5.067		7.171	0.000	26.385	46.298
crim	-0.108	0.033	-0.101	-3.307	0.001	-0.173	-0.044
zn	0.046	0.014	0.116	3.390	0.001	0.019	0.072
chas	2.719	0.854	0.075	3.183	0.002	1.040	4.397
nox	-17.376	3.535	-0.219	-4.915	0.000	-24.322	-10.430
rm	3.802	0.406	0.290	9.356	0.000	3.003	4.600
dis	-1.493	0.186	-0.342	-8.037	0.000	-1.858	-1.128
rad	0.300	0.063	0.284	4.726	0.000	0.175	0.424
tax	-0.012	0.003	-0.216	-3.493	0.001	-0.018	-0.005
ptratio	-0.947	0.129	-0.223	-7.334	0.000	-1.200	-0.693
black	0.009	0.003	0.092	3.475	0.001	0.004	0.015
lstat	-0.523	0.047	-0.406	-11.019	0.000	-0.616	-0.429

Elimination Summary

Step	Variable Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	age	0.7406	0.7343	12.0027	3025.6114	4.7405
2	indus	0.7406	0.7348	10.1145	3023.7264	4.7362

```
Final Model Output
------------------

                        Model Summary
--------------------------------------------------------------
R                       0.861       RMSE                4.736
R-Squared               0.741       Coef. Var          21.019
Adj. R-Squared          0.735       MSE                22.432
Pred R-Squared          0.721       MAE                 3.272
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
```

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	36.341	5.067		7.171	0.000	26.385	46.298
lstat	-0.523	0.047	-0.406	-11.019	0.000	-0.616	-0.429
rm	3.802	0.406	0.290	9.356	0.000	3.003	4.600
ptratio	-0.947	0.129	-0.223	-7.334	0.000	-1.200	-0.693
dis	-1.493	0.186	-0.342	-8.037	0.000	-1.858	-1.128
nox	-17.376	3.535	-0.219	-4.915	0.000	-24.322	-10.430
chas	2.719	0.854	0.075	3.183	0.002	1.040	4.397
black	0.009	0.003	0.092	3.475	0.001	0.004	0.015
zn	0.046	0.014	0.116	3.390	0.001	0.019	0.072
crim	-0.108	0.033	-0.101	-3.307	0.001	-0.173	-0.044
rad	0.300	0.063	0.284	4.726	0.000	0.175	0.424
tax	-0.012	0.003	-0.216	-3.493	0.001	-0.018	-0.005

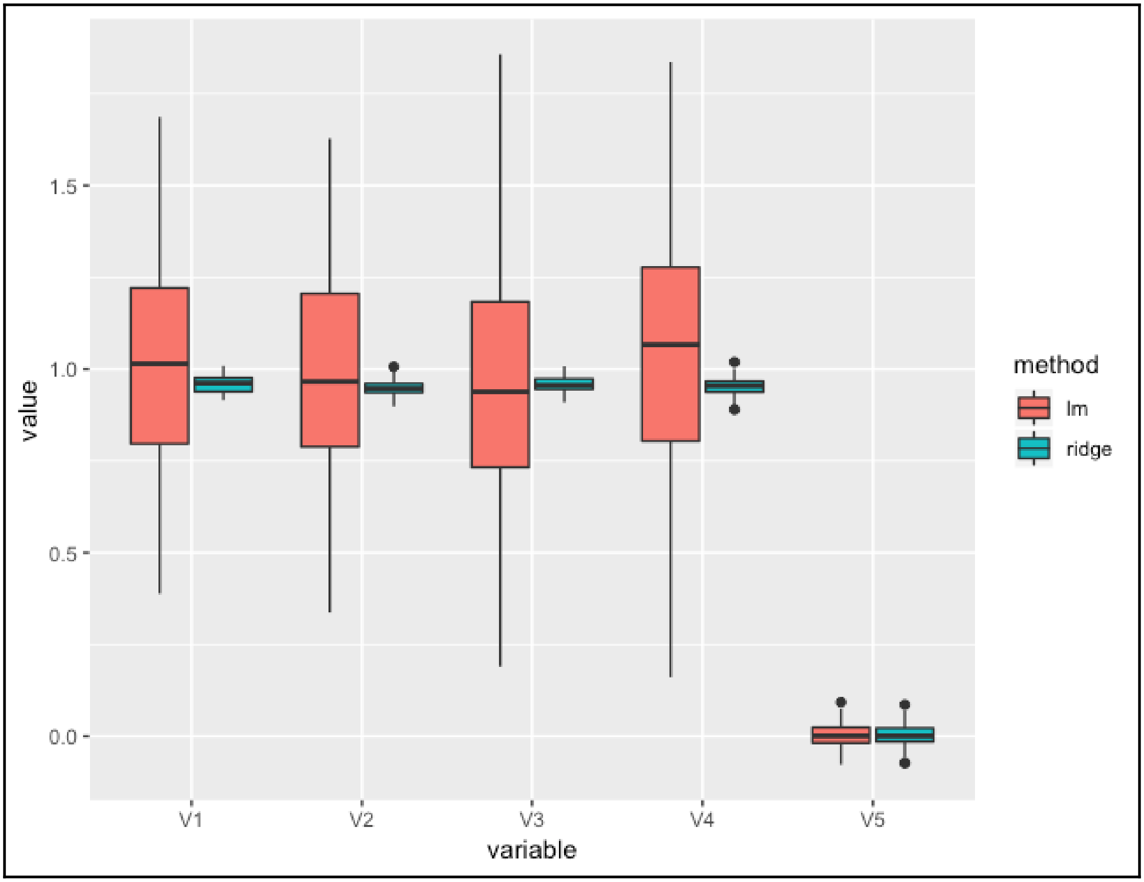
Stepwise Selection Summary

Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	lstat	addition	0.544	0.543	362.7530	3288.9750	6.2158
2	rm	addition	0.639	0.637	185.6470	3173.5423	5.5403
3	ptratio	addition	0.679	0.677	111.6490	3116.0973	5.2294
4	dis	addition	0.690	0.688	91.4850	3099.3590	5.1386
5	nox	addition	0.708	0.705	59.7540	3071.4386	4.9939
6	chas	addition	0.716	0.712	47.1750	3059.9390	4.9326
7	black	addition	0.722	0.718	37.0590	3050.4384	4.8818
8	zn	addition	0.727	0.722	30.6240	3044.2750	4.8474
9	crim	addition	0.729	0.724	28.4180	3042.1546	4.8326
10	rad	addition	0.734	0.729	20.2660	3034.0687	4.7895
11	tax	addition	0.741	0.735	10.1150	3023.7264	4.7362

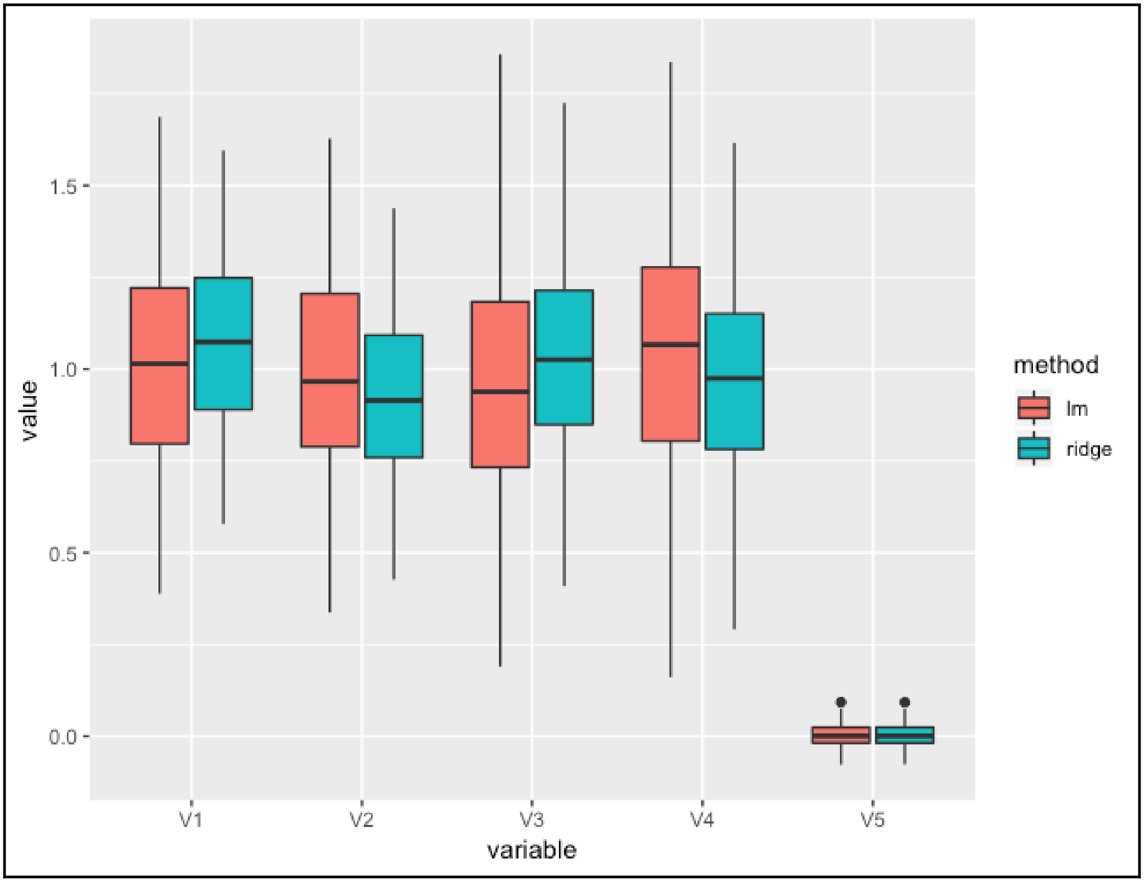
```
No more variables to be added.
                         Selection Summary
-----------------------------------------------------------------------
Variable        AIC         Sum Sq         RSS        R-Sq    Adj. R-Sq
-----------------------------------------------------------------------
lstat        3288.975    23243.914    19472.381    0.54415      0.54324
rm           3173.542    27276.986    15439.309    0.63856      0.63712
ptratio      3116.097    28988.310    13727.985    0.67862      0.67670
dis          3099.359    29487.388    13228.908    0.69031      0.68784
nox          3071.439    30246.951    12469.344    0.70809      0.70517
chas         3059.939    30575.223    12141.073    0.71577      0.71236
black        3050.438    30848.060    11868.236    0.72216      0.71826
zn           3044.275    31037.996    11678.299    0.72661      0.72221
crim         3042.155    31132.708    11583.588    0.72883      0.72390
rad          3034.069    31361.312    11354.983    0.73418      0.72881
tax          3023.726    31634.931    11081.364    0.74058      0.73481
-----------------------------------------------------------------------
```

```
Variables Removed:

✖ age
✖ indus

No more variables to be removed.

                    Backward Elimination Summary
-----------------------------------------------------------------------
Variable         AIC         RSS        Sum Sq       R-Sq    Adj. R-Sq
-----------------------------------------------------------------------
Full Model    3027.609    11078.785    31637.511    0.74064     0.73379
age           3025.611    11078.846    31637.449    0.74064     0.73433
indus         3023.726    11081.364    31634.931    0.74058     0.73481
-----------------------------------------------------------------------
```

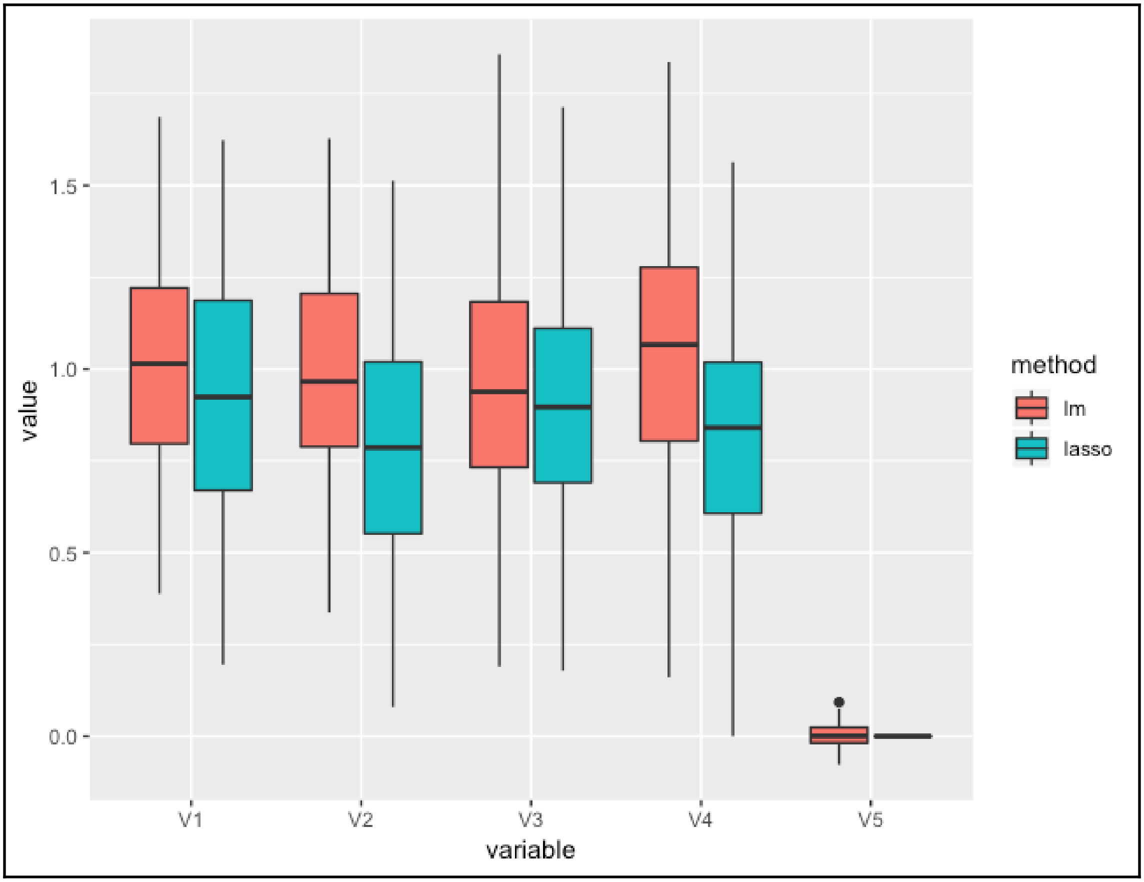
```
No more variables to be added or removed.

                              Stepwise Summary
------------------------------------------------------------------------------------
Variable     Method       AIC          RSS         Sum Sq       R-Sq      Adj. R-Sq
------------------------------------------------------------------------------------
lstat       addition    3288.975    19472.381    23243.914    0.54415      0.54324
rm          addition    3173.542    15439.309    27276.986    0.63856      0.63712
ptratio     addition    3116.097    13727.985    28988.310    0.67862      0.67670
dis         addition    3099.359    13228.908    29487.388    0.69031      0.68784
nox         addition    3071.439    12469.344    30246.951    0.70809      0.70517
chas        addition    3059.939    12141.073    30575.223    0.71577      0.71236
black       addition    3050.438    11868.236    30848.060    0.72216      0.71826
zn          addition    3044.275    11678.299    31037.996    0.72661      0.72221
crim        addition    3042.155    11583.588    31132.708    0.72883      0.72390
rad         addition    3034.069    11354.983    31361.312    0.73418      0.72881
tax         addition    3023.726    11081.364    31634.931    0.74058      0.73481
------------------------------------------------------------------------------------
```


value
1.5
1.0
0.5
0.0
V1
V2
V3
V4
V5
variable
method
lm
ridge

```
   variable method   median
   <chr>    <fct>     <dbl>
 1 V1       lm      1.01
 2 V1       ridge   0.962
 3 V2       lm      0.966
 4 V2       ridge   0.947
 5 V3       lm      0.938
 6 V3       ridge   0.956
 7 V4       lm      1.07
 8 V4       ridge   0.954
 9 V5       lm      0.000713
10 V5       ridge   0.000578
```

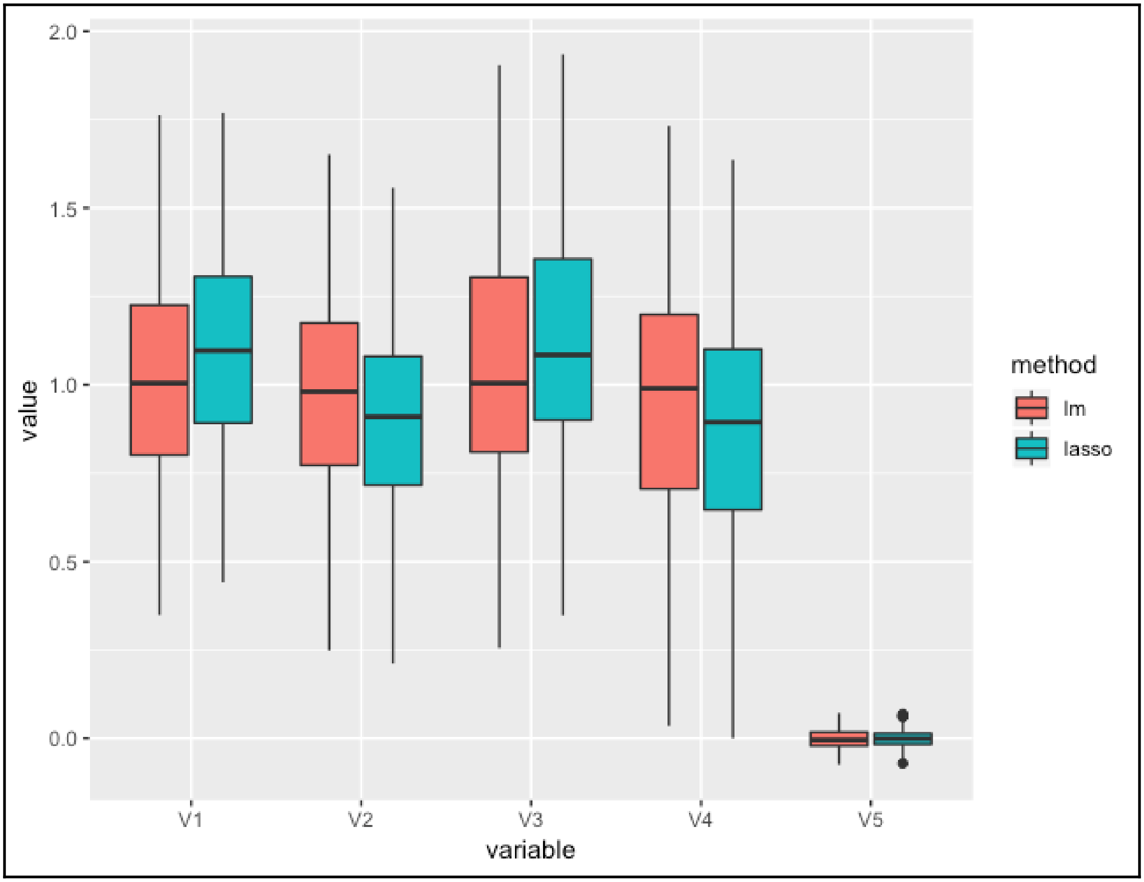

1.5
1.0
0.5
0.0
value
V1
V2
V3
V4
V5
variable
method
lm
ridge

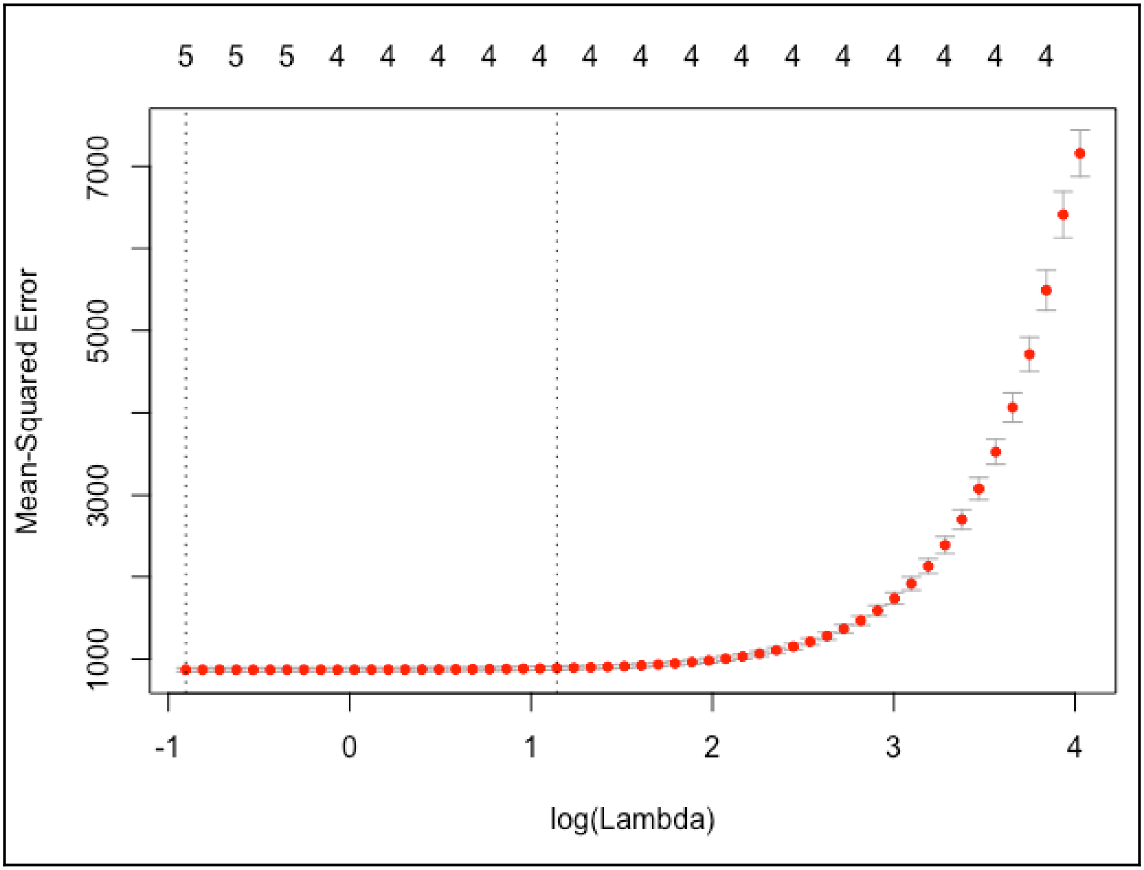
```
   variable method   median
   <chr>    <fct>     <dbl>
 1 V1       lm      1.01
 2 V1       ridge   1.07
 3 V2       lm      0.966
 4 V2       ridge   0.914
 5 V3       lm      0.938
 6 V3       ridge   1.03
 7 V4       lm      1.07
 8 V4       ridge   0.974
 9 V5       lm      0.000713
10 V5       ridge   0.000570
```



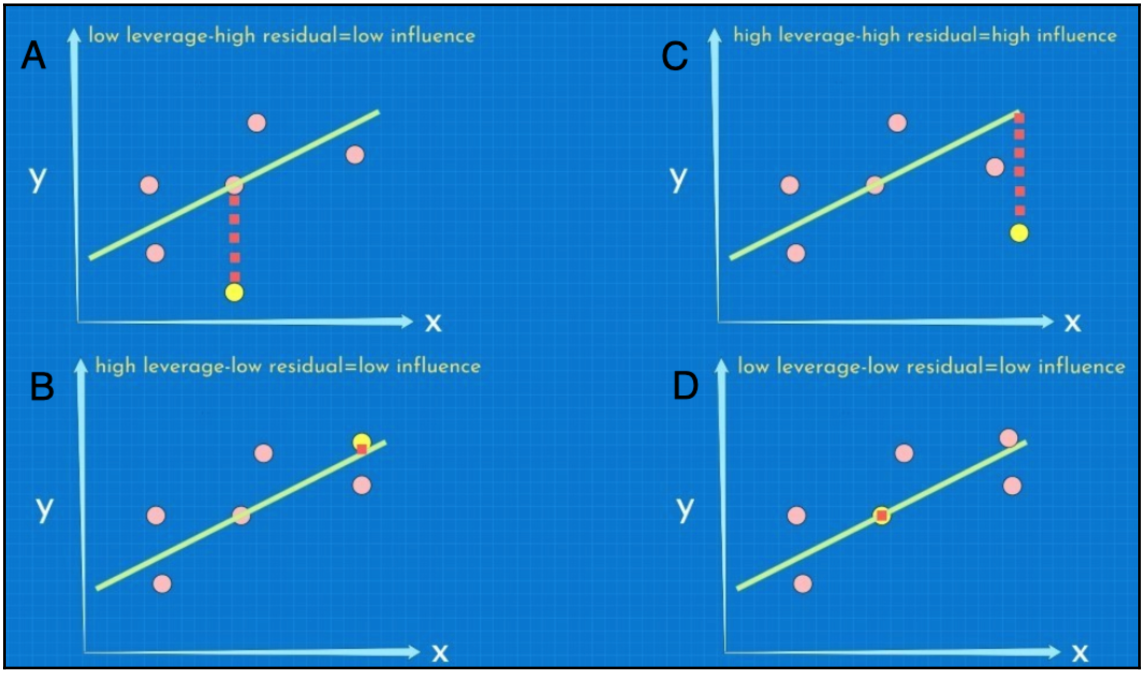
1.5
1.0
0.5
0.0
value
V1
V2
V3
V4
V5
variable
method
lm
lasso

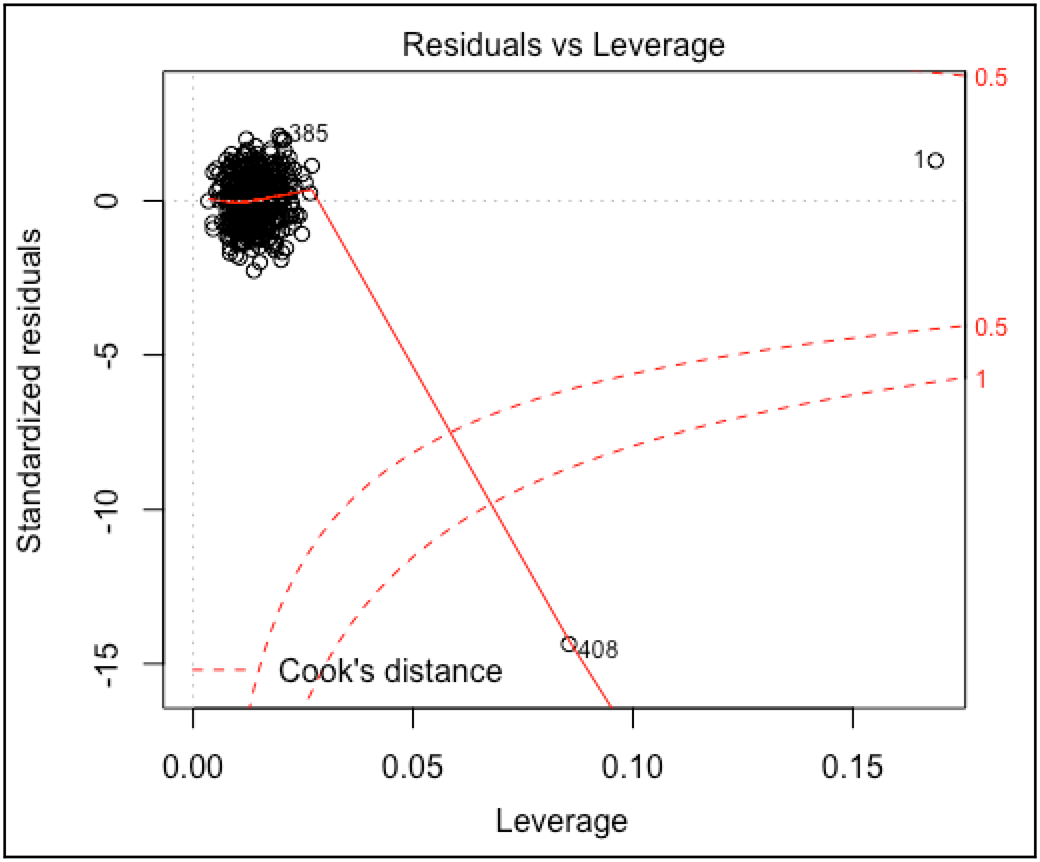
```
   variable method   median
   <chr>    <fct>     <dbl>
 1 V1       lm     1.01
 2 V1       lasso  0.924
 3 V2       lm     0.966
 4 V2       lasso  0.786
 5 V3       lm     0.938
 6 V3       lasso  0.896
 7 V4       lm     1.07
 8 V4       lasso  0.841
 9 V5       lm     0.000713
10 V5       lasso  0
```


2.0
1.5
1.0
0.5
0.0
value
V1
V2
V3
V4
V5
variable
method
lm
lasso

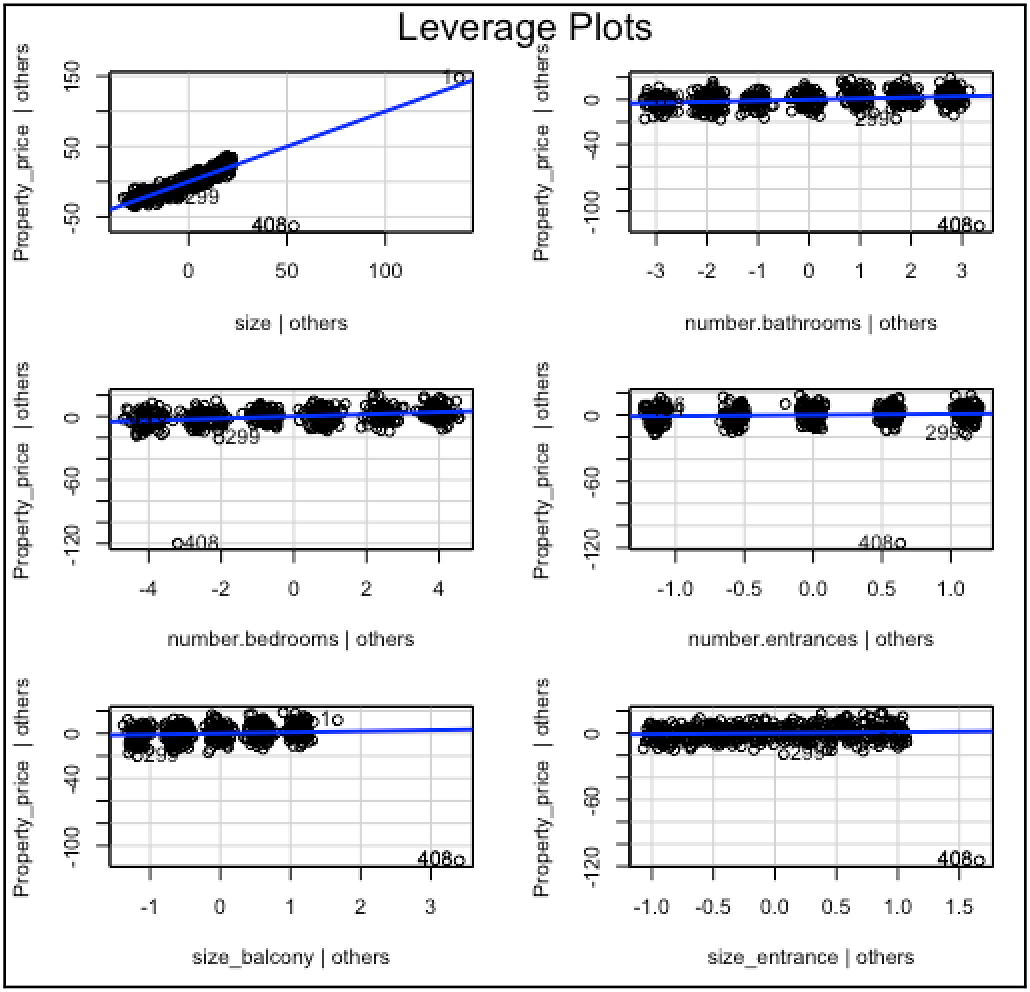



```
$lambda.min
[1] 0.4050034

$lambda.1se
[1] 3.135794
```

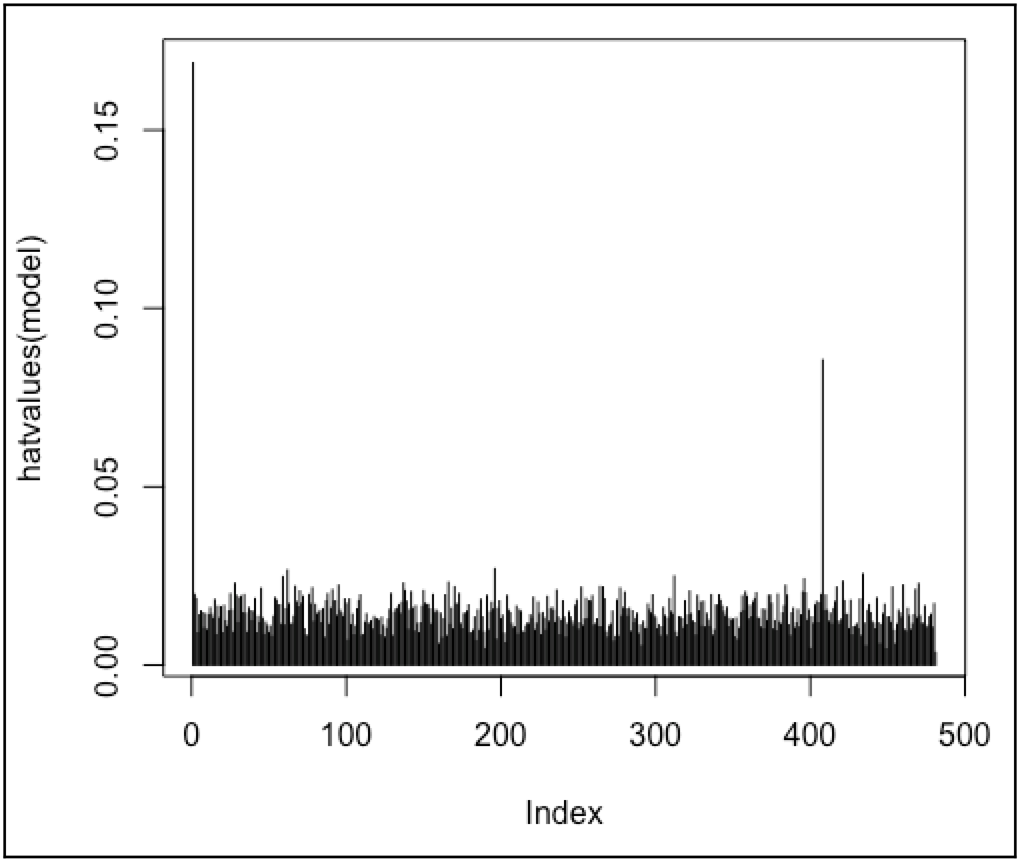

A
low leverage-high residual=low influence
y
x
C
high leverage-high residual=high influence
y
x
B
high leverage-low residual=low influence
y
x
D
low leverage-low residual=low influence
y
x

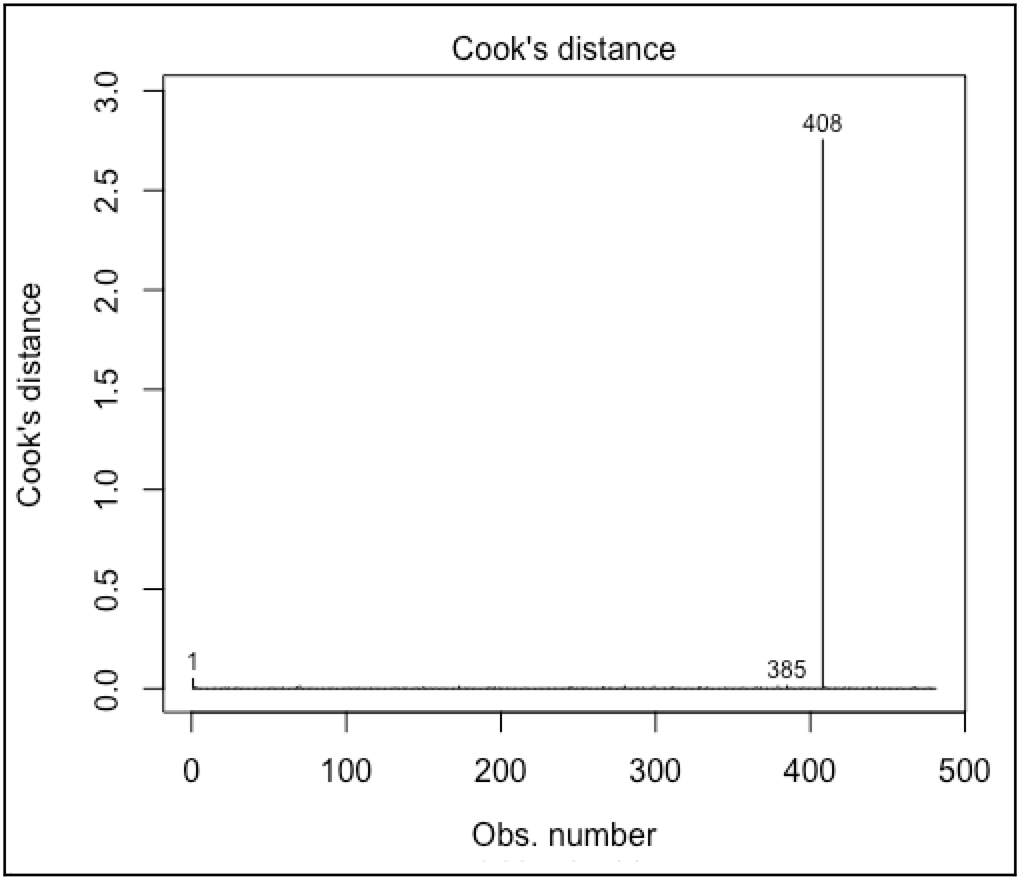


Residuals vs Leverage
Standardized residuals
Leverage
0
-5
-10
-15
0.00
0.05
0.10
0.15
385
1
408
Cook's distance
0.5
0.5
1



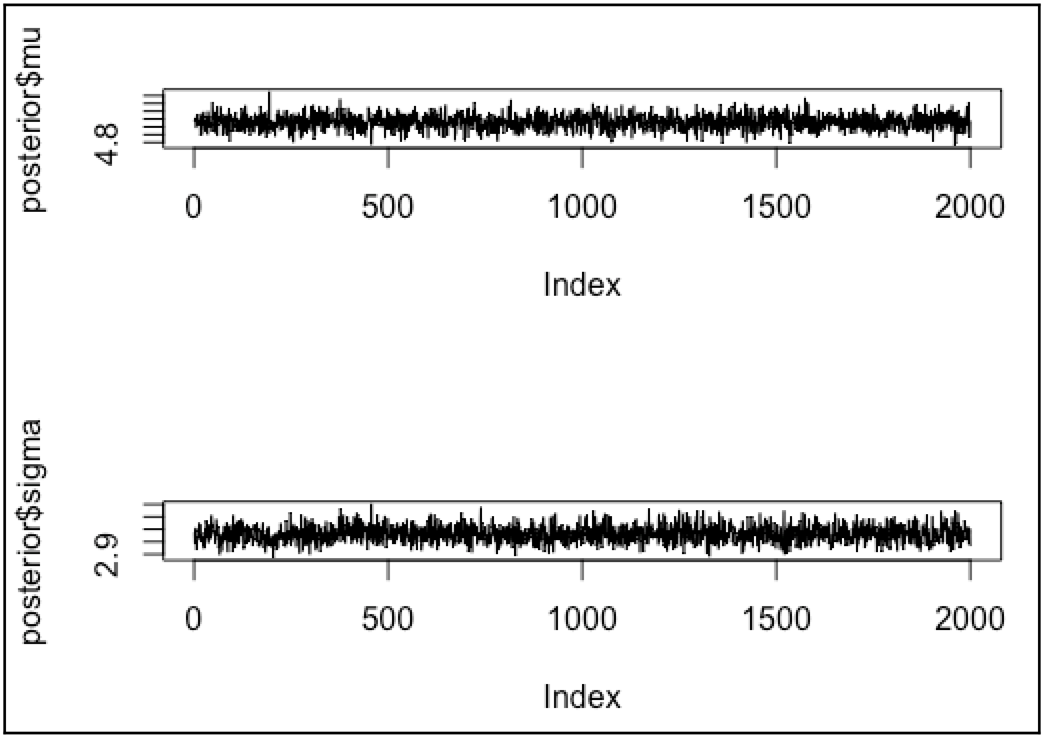

```
> outlierTest(model)
     rstudent unadjusted p-value Bonferonni p
408 -19.08144         1.2928e-60   6.2183e-58
```

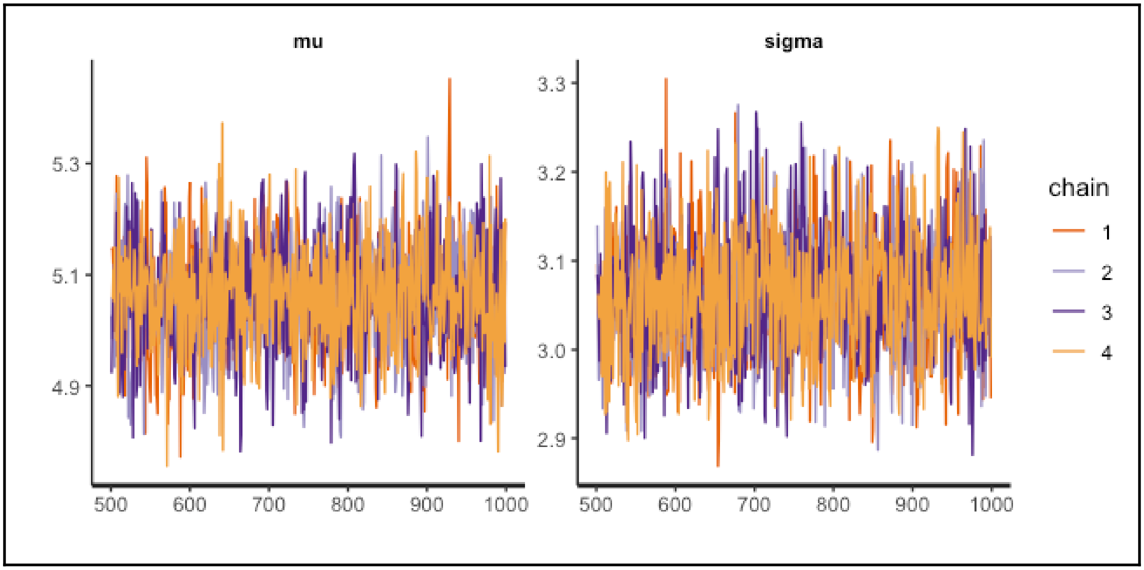

hatvalues(model)
0.00
0.05
0.10
0.15
0
100
200
300
400
500
Index

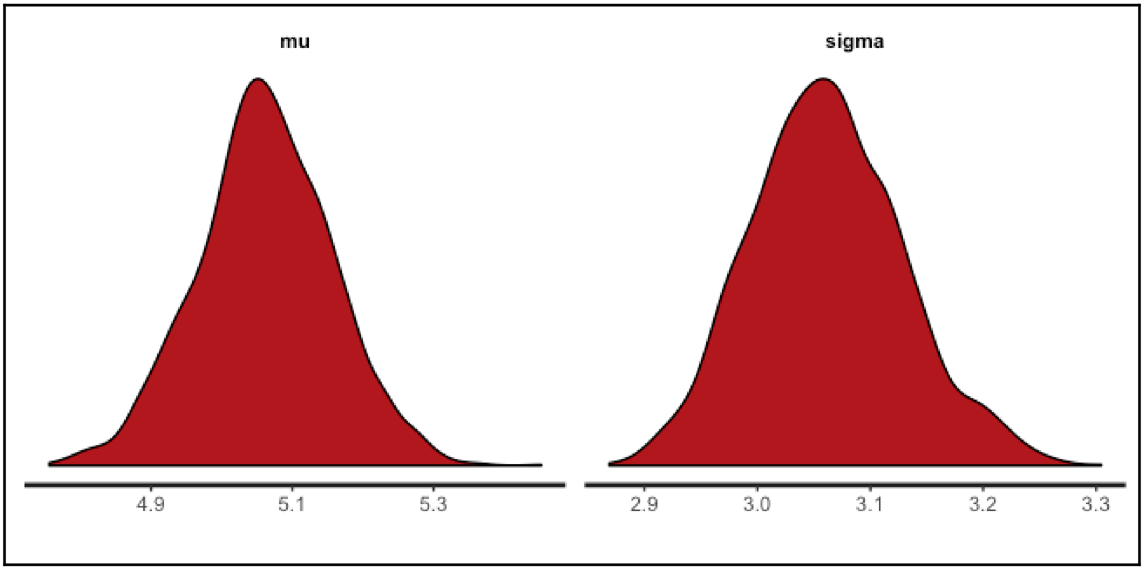



```
Call:
lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data)

Coefficients:
     (Intercept)              size  number.bathrooms   number.bedrooms  number.entrances      size_balcony     size_entrance
         -4.8841            5.3253            0.9583            1.6219            0.5660            0.5791            0.1908

> model2

Call:
lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data[-c(408),
    ])

Coefficients:
     (Intercept)              size  number.bathrooms   number.bedrooms  number.entrances      size_balcony     size_entrance
        -16.8767            5.6124            1.1817            1.4371            0.7132            1.3564            0.4043
```


prior distribution
data
posterior distribution

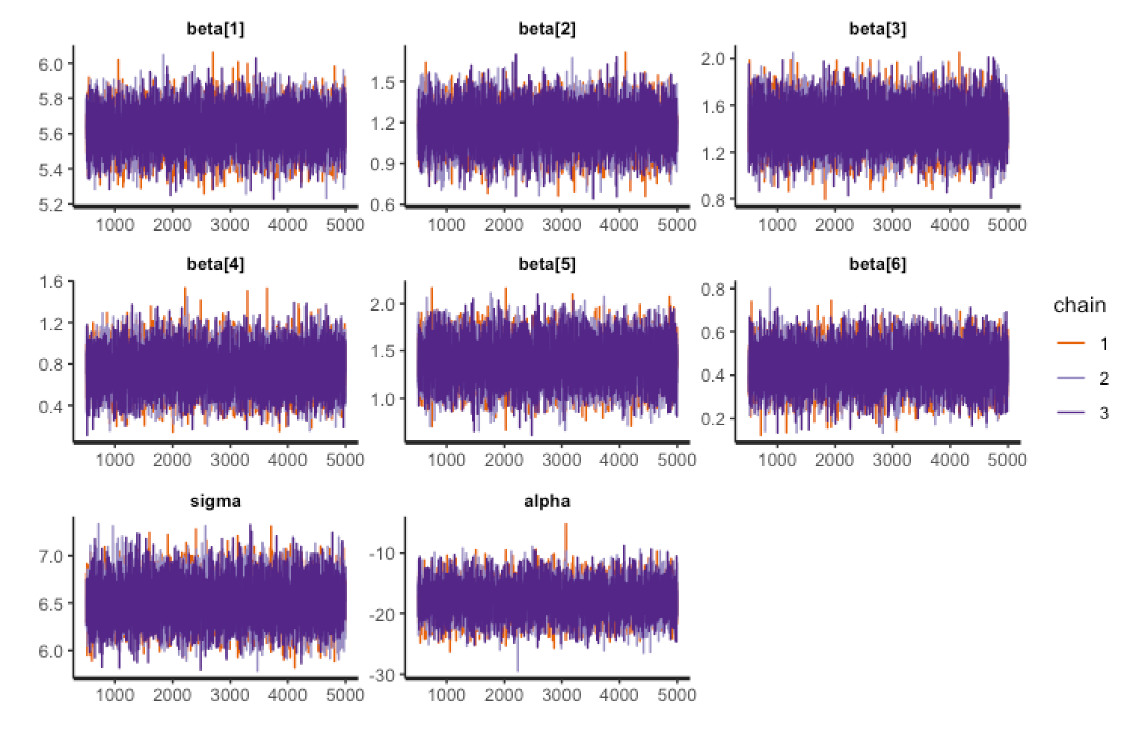

posterior$mu
4.8
0
500
1000
1500
2000
Index
posterior$sigma
2.9
0
500
1000
1500
2000
Index



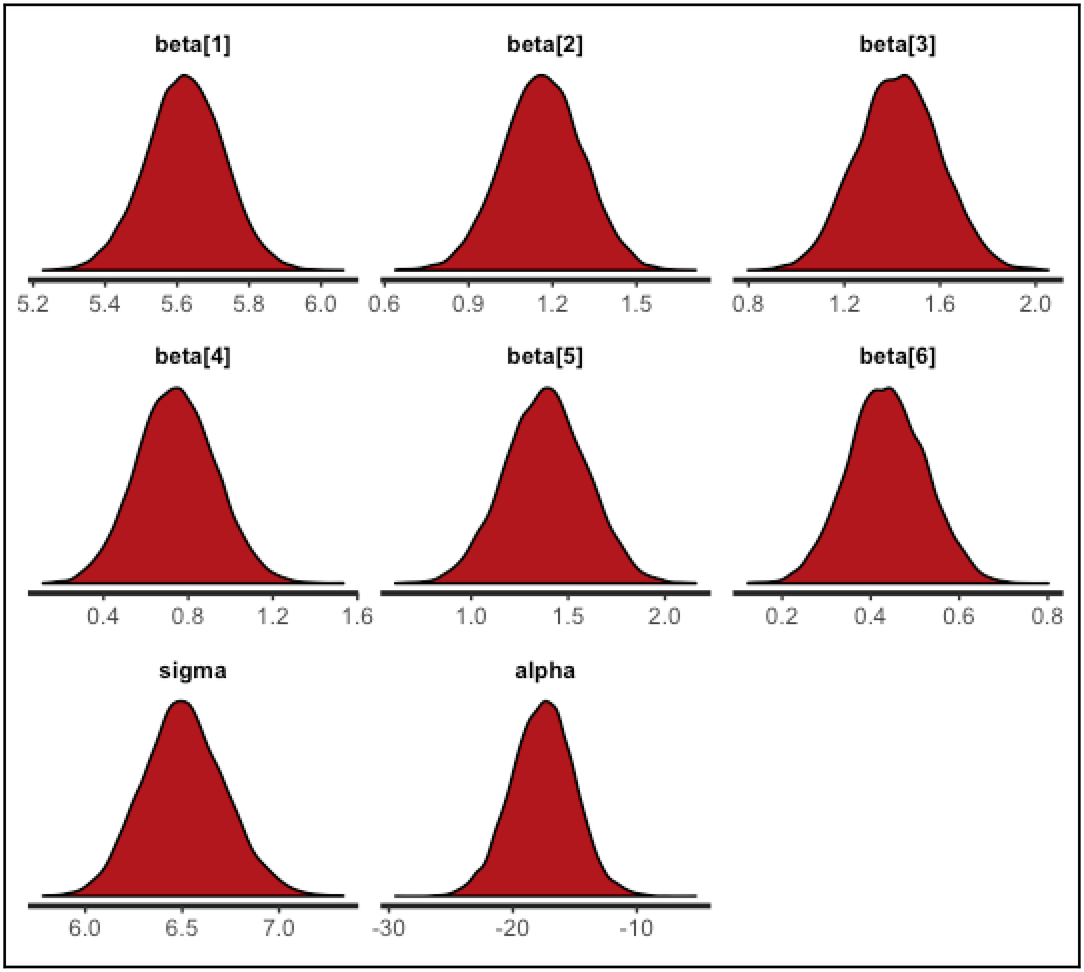



```
$summary
             mean     se_mean         sd         2.5%          25%          50%          75%        97.5%     n_eff      Rhat
mu       5.061284 0.002299197 0.09518200     4.875067     5.002210     5.060738     5.125745     5.245615 1713.7884 0.9996033
sigma    3.061439 0.001555272 0.06721101     2.937593     3.015324     3.058954     3.106601     3.204593 1867.5325 1.0012654
lp__  -1619.014663 0.031237429 0.96577384 -1621.529914 -1619.400912 -1618.724423 -1618.319770 -1618.075128  955.8733 1.0001530
```

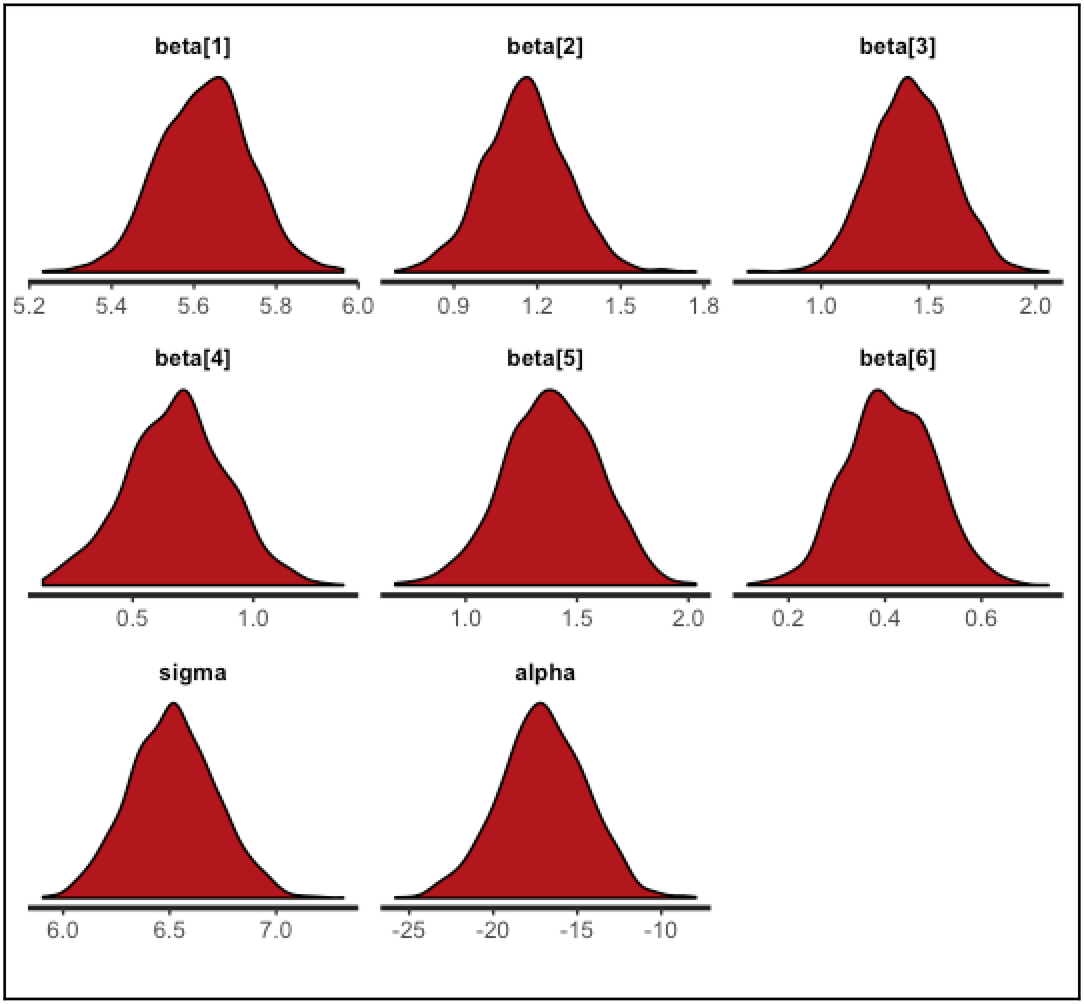
```
Call:
lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data)

Coefficients:
     (Intercept)              size  number.bathrooms   number.bedrooms  number.entrances      size_balcony     size_entrance
        -16.7063            5.6270            1.1513            1.4115            0.6621            1.3693            0.4055
```

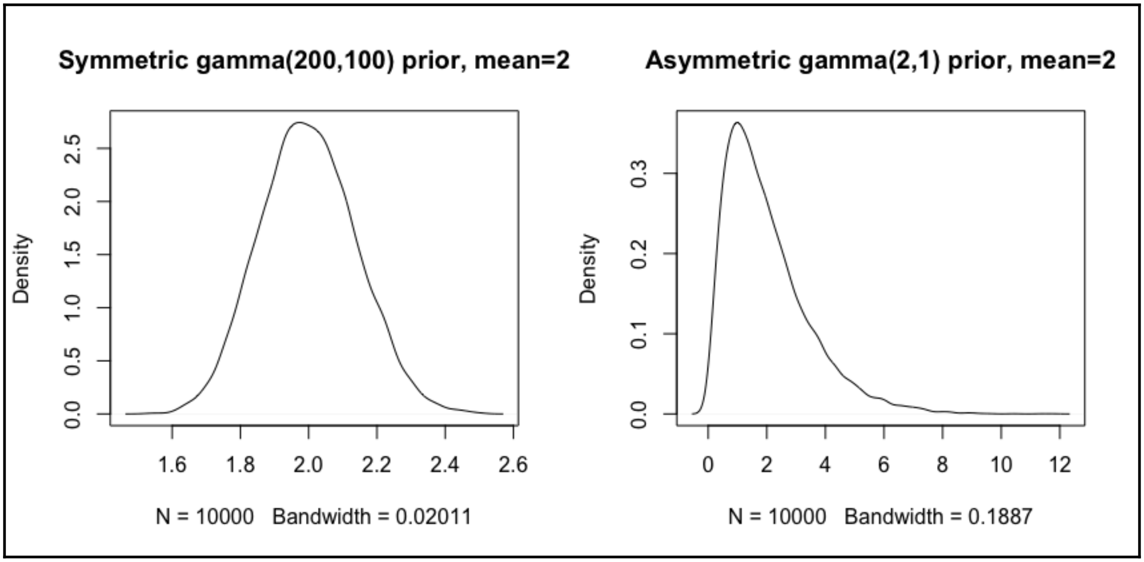


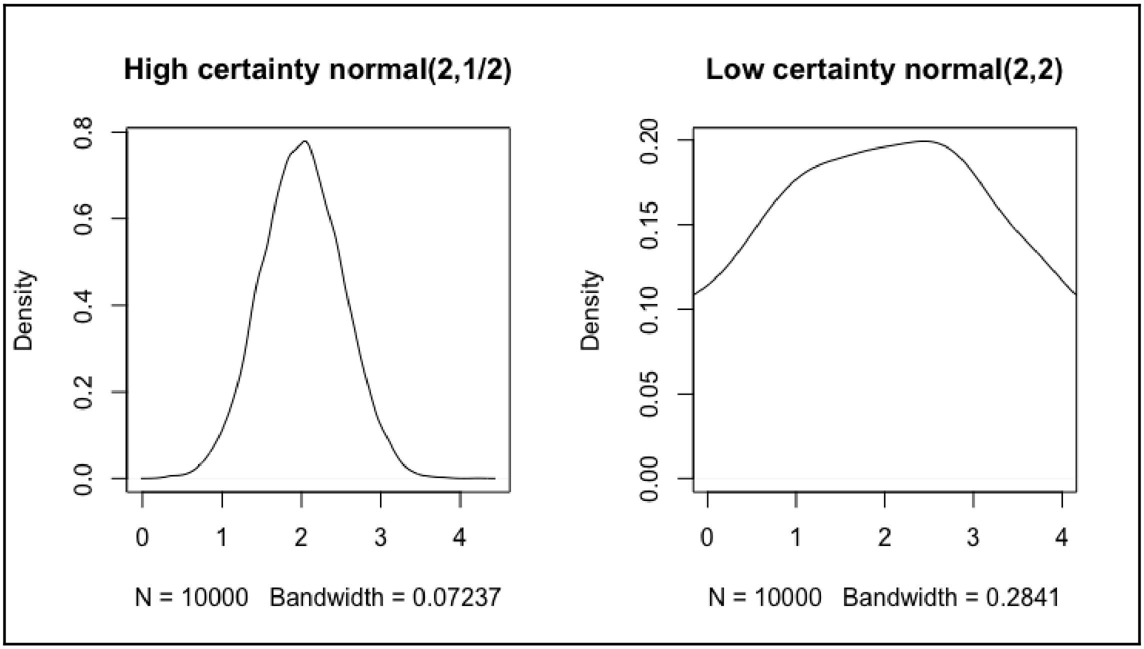

```
               mean      se_mean         sd          2.5%           25%           50%           75%         97.5%     n_eff      Rhat
beta[1]    5.6225788 0.0009766837 0.10683040     5.4111115     5.5517882     5.6226907     5.6952738     5.8314363 11964.149 0.9998872
beta[2]    1.1643991 0.0012442749 0.14226817     0.8850887     1.0686641     1.1644535     1.2604379     1.4415373 13073.228 0.9999466
beta[3]    1.4262629 0.0015232267 0.17442242     1.0898667     1.3086345     1.4263075     1.5452445     1.7667870 13112.199 1.0000970
beta[4]    0.7435934 0.0015865555 0.18663472     0.3824652     0.6162078     0.7406738     0.8678405     1.1139111 13838.032 1.0000081
beta[5]    1.3852341 0.0019617951 0.21042770     0.9742840     1.2438578     1.3845633     1.5274917     1.7951890 11505.316 0.9998713
beta[6]    0.4362908 0.0007741666 0.08766455     0.2666178     0.3767417     0.4353704     0.4962245     0.6096573 12822.691 1.0000954
sigma      6.5099420 0.0018426662 0.21313951     6.1080880     6.3651862     6.5041227     6.6512168     6.9438713 13379.336 1.0000498
alpha    -17.5840560 0.0271623199 2.52921088   -22.6663692   -19.2924450   -17.5431884   -15.8754292   -12.6919085  8670.345 1.0000525
lp__   -1148.7268327 0.0254192688 2.01624282 -1153.6471968 -1149.8241523 -1148.3994550 -1147.2645835 -1145.8007536  6291.578 0.9998735
```

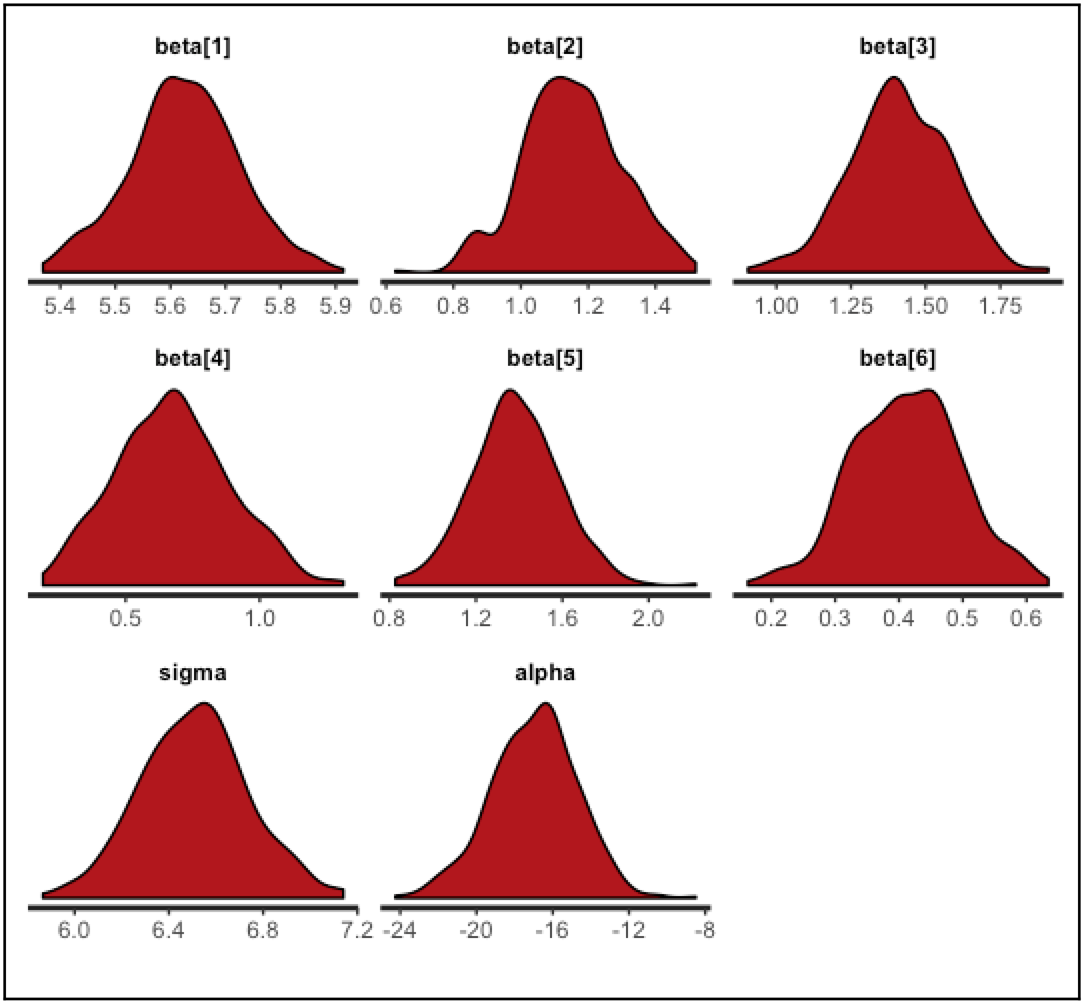


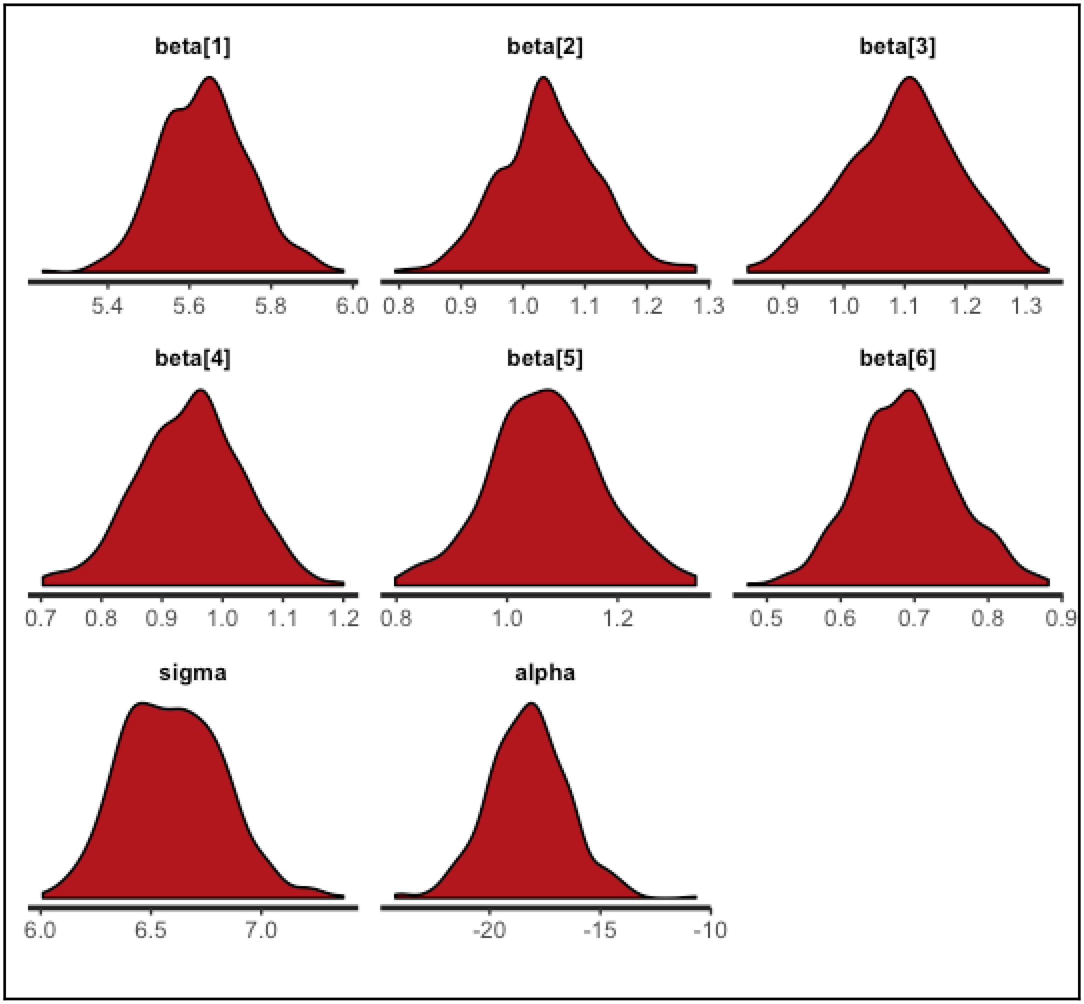
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	5.6255568	0.002820783	0.11001490	5.4159132	5.5475310	5.6279355	5.6985104	5.843097	1521.1205	1.0013970
beta[2]	1.1583504	0.004053775	0.14967539	0.8601551	1.0606004	1.1567897	1.2535384	1.448670	1363.2689	1.0007515
beta[3]	1.4199475	0.005144073	0.18548777	1.0690988	1.2909298	1.4175903	1.5455242	1.775674	1300.2186	1.0030240
beta[4]	0.6818515	0.005846059	0.20679897	0.2622836	0.5406722	0.6862651	0.8200143	1.085002	1251.3256	0.9992717
beta[5]	1.3868954	0.006095161	0.21271219	0.9596830	1.2404890	1.3883635	1.5363500	1.786665	1217.9079	1.0006140
beta[6]	0.4120919	0.002543454	0.09091023	0.2421430	0.3505592	0.4094255	0.4769115	0.591295	1277.5498	0.9998908
sigma	6.5155501	0.005030784	0.20372699	6.1263465	6.3762869	6.5130353	6.6489729	6.928737	1639.9315	1.0025792
alpha	-17.0314924	0.086909093	2.59657269	-22.2778510	-18.7123645	-17.0306260	-15.2446184	-12.158003	892.6281	0.9994337
lp__	-1144.0256355	0.074892136	2.06822589	-1148.7713547	-1145.2262920	-1143.6861299	-1142.4698515	-1140.978021	762.6469	1.0027350

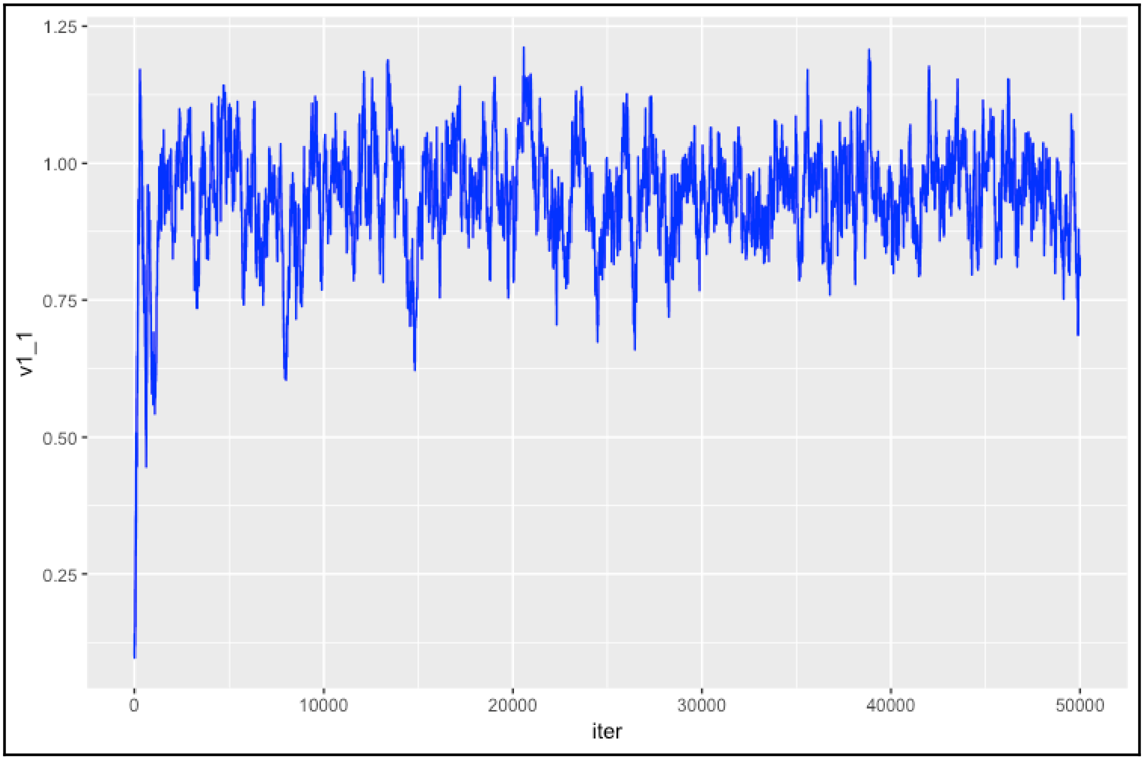

beta[1]
5.2
5.4
5.6
5.8
6.0
beta[2]
0.9
1.2
1.5
1.8
beta[3]
1.0
1.5
2.0
beta[4]
0.5
1.0
beta[5]
1.0
1.5
2.0
beta[6]
0.2
0.4
0.6
sigma
6.0
6.5
7.0
alpha
-25
-20
-15
-10

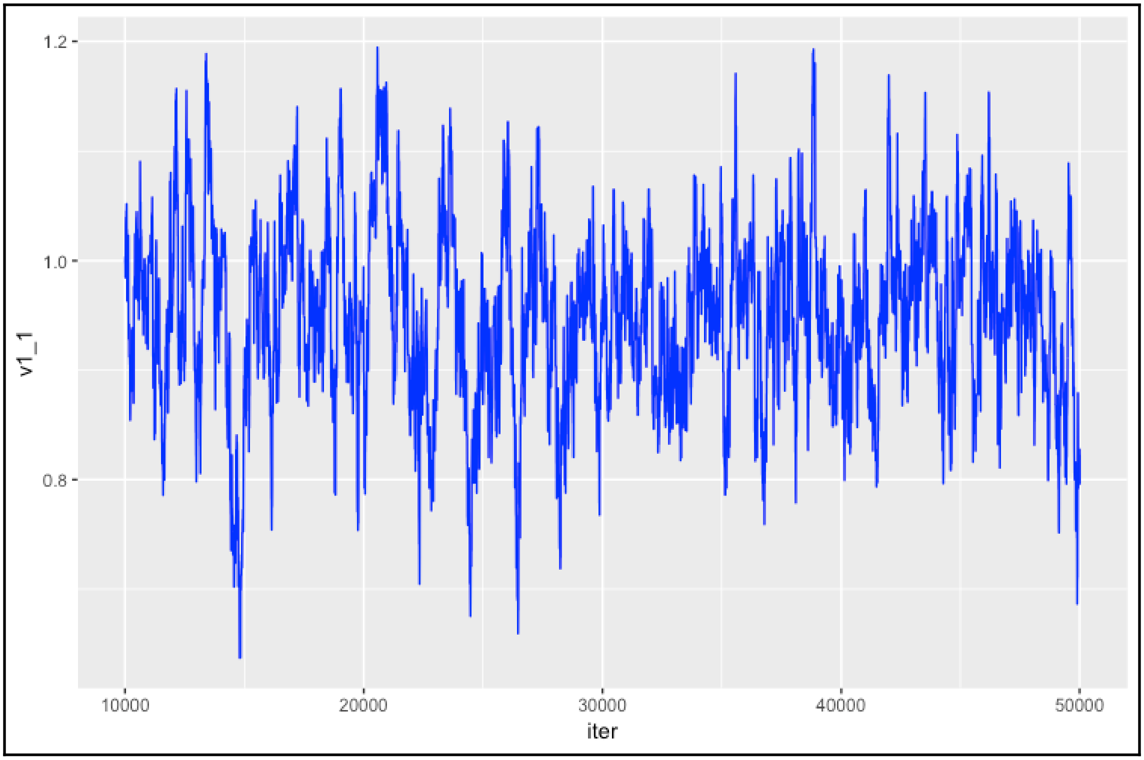

Symmetric gamma(200,100) prior, mean=2
Density
N = 10000 Bandwidth = 0.02011
Asymmetric gamma(2,1) prior, mean=2
Density
N = 10000 Bandwidth = 0.1887

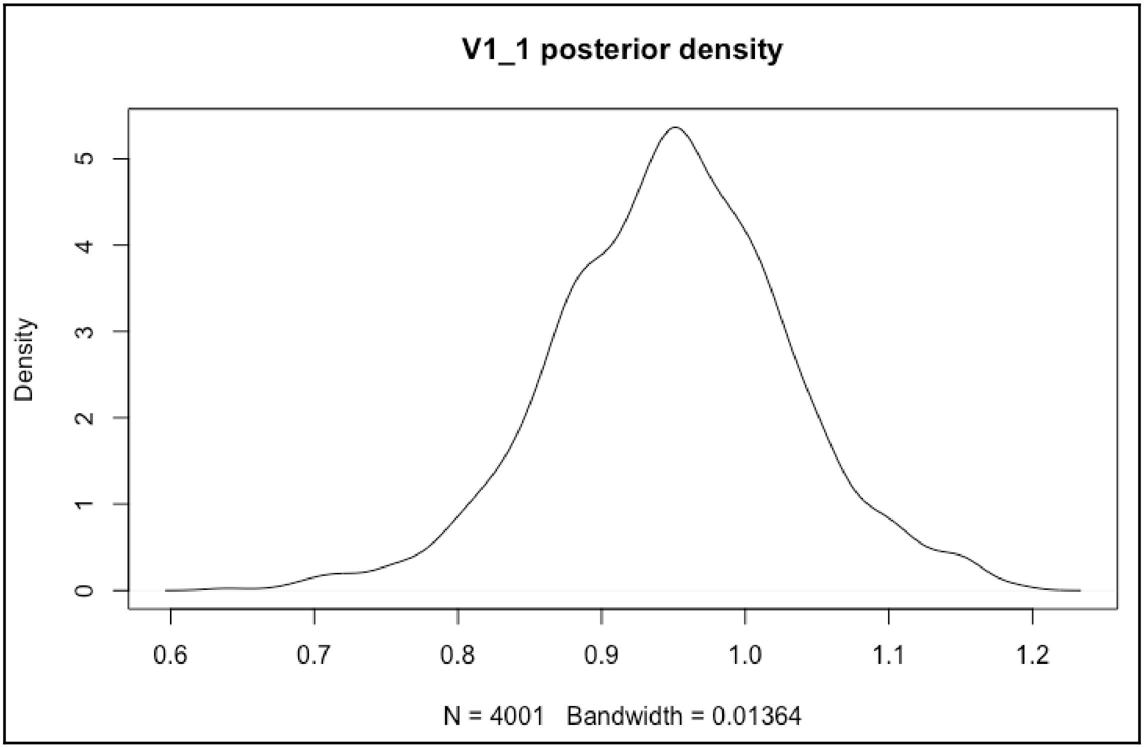

High certainty normal(2,1/2)
Density
N = 10000 Bandwidth = 0.07237
Low certainty normal(2,2)
Density
N = 10000 Bandwidth = 0.2841

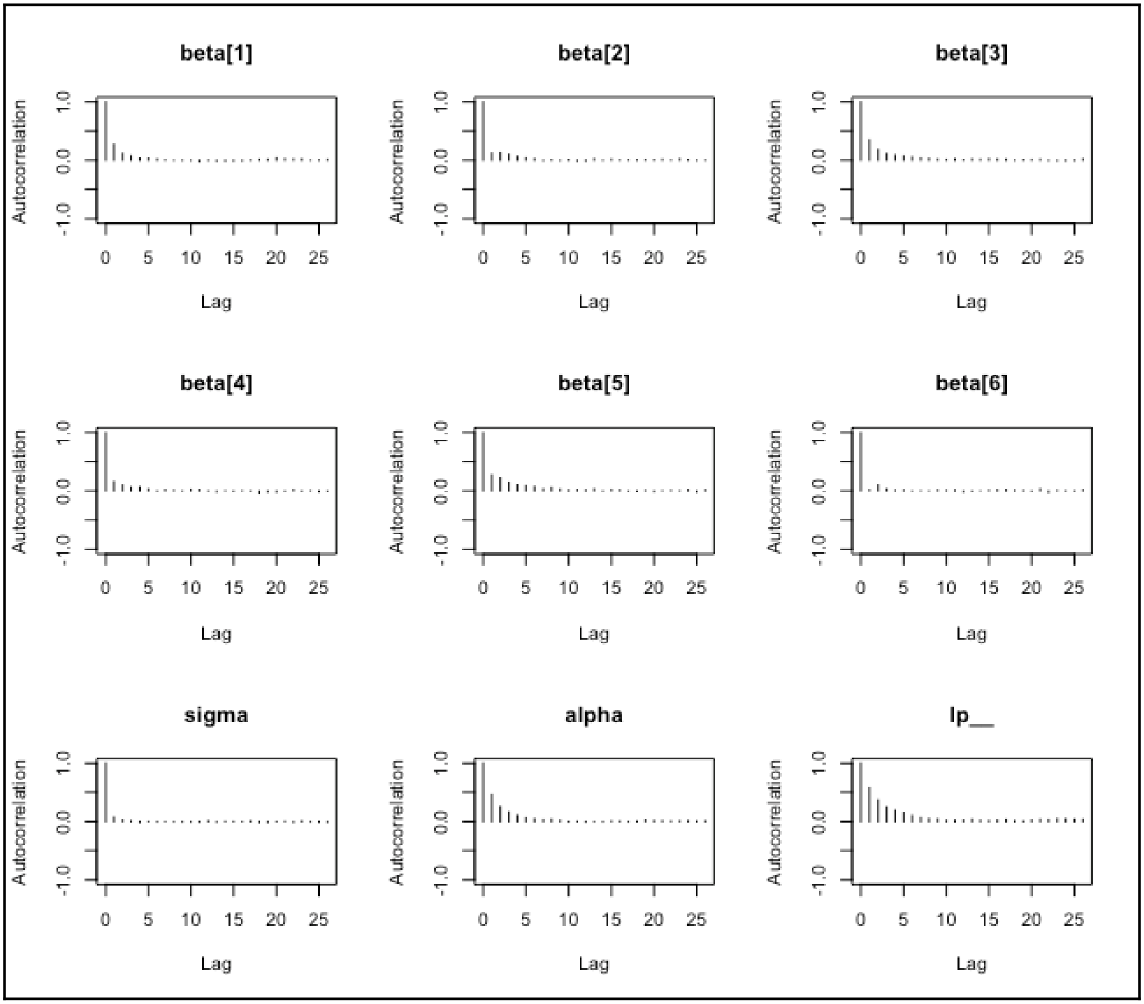

beta[1]
5.4 5.5 5.6 5.7 5.8 5.9
beta[2]
0.6 0.8 1.0 1.2 1.4
beta[3]
1.00 1.25 1.50 1.75
beta[4]
0.5 1.0
beta[5]
0.8 1.2 1.6 2.0
beta[6]
0.2 0.3 0.4 0.5 0.6
sigma
6.0 6.4 6.8 7.2
alpha
-24 -20 -16 -12 -8

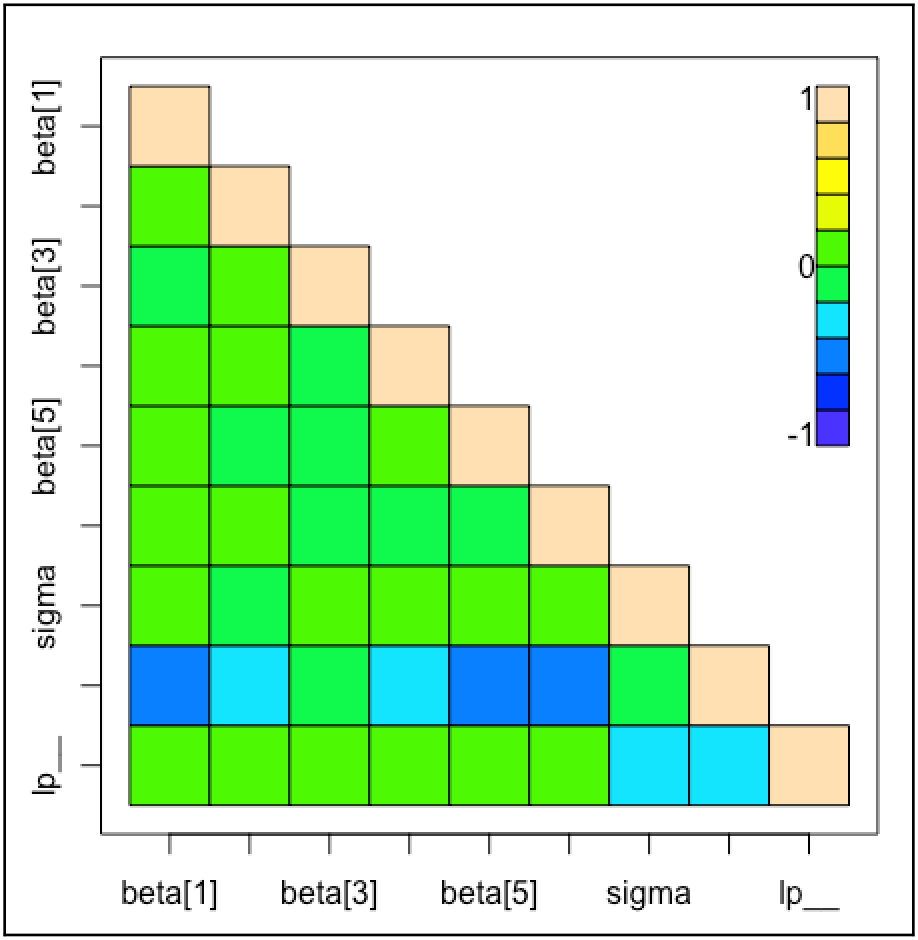

beta[1]
5.4 5.6 5.8 6.0
beta[2]
0.8 0.9 1.0 1.1 1.2 1.3
beta[3]
0.9 1.0 1.1 1.2 1.3
beta[4]
0.7 0.8 0.9 1.0 1.1 1.2
beta[5]
0.8 1.0 1.2
beta[6]
0.5 0.6 0.7 0.8 0.9
sigma
6.0 6.5 7.0
alpha
-20 -15 -10


1.25
1.00
0.75
0.50
0.25
v1_1
0
10000
20000
30000
40000
50000
iter


1.2
1.0
0.8
v1_1
10000
20000
30000
40000
50000
iter

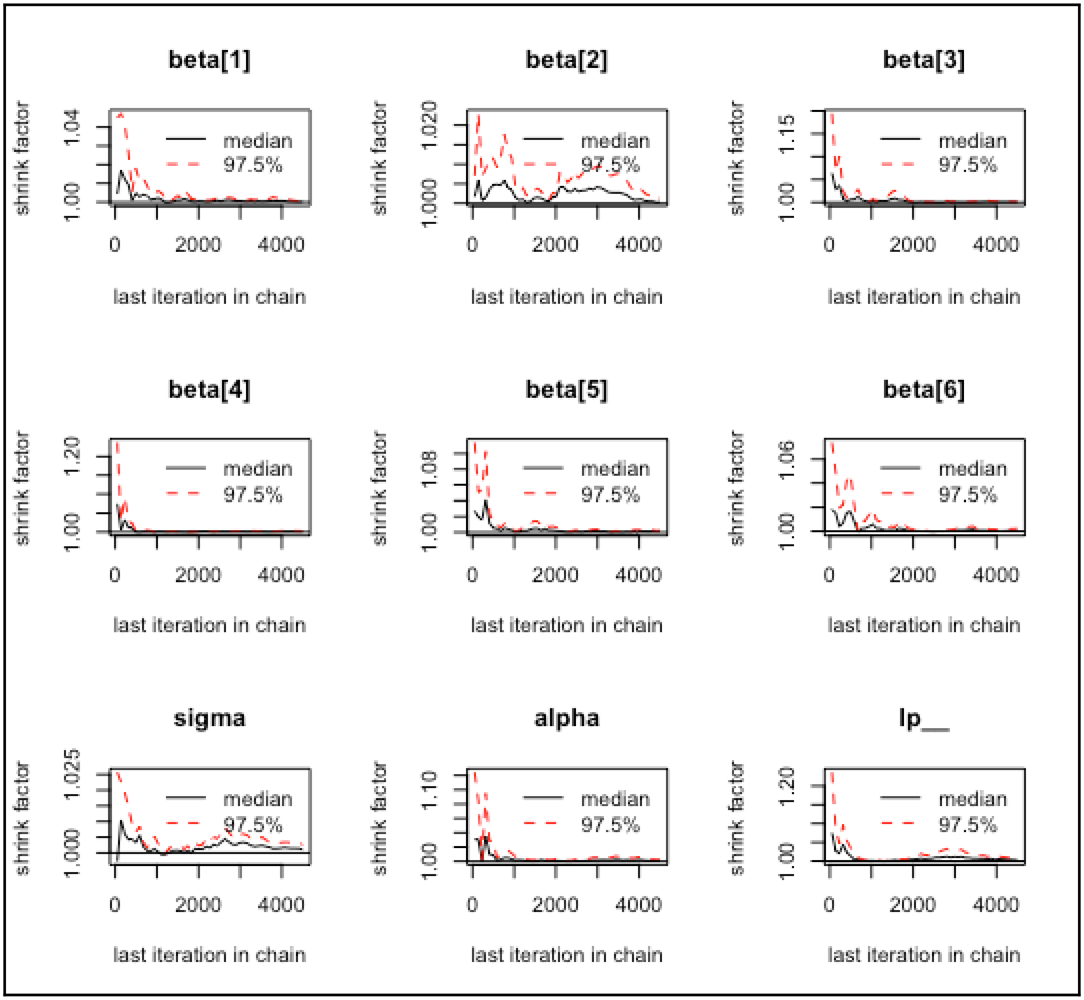

V1_1 posterior density
Density
0
1
2
3
4
5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
N = 4001   Bandwidth = 0.01364

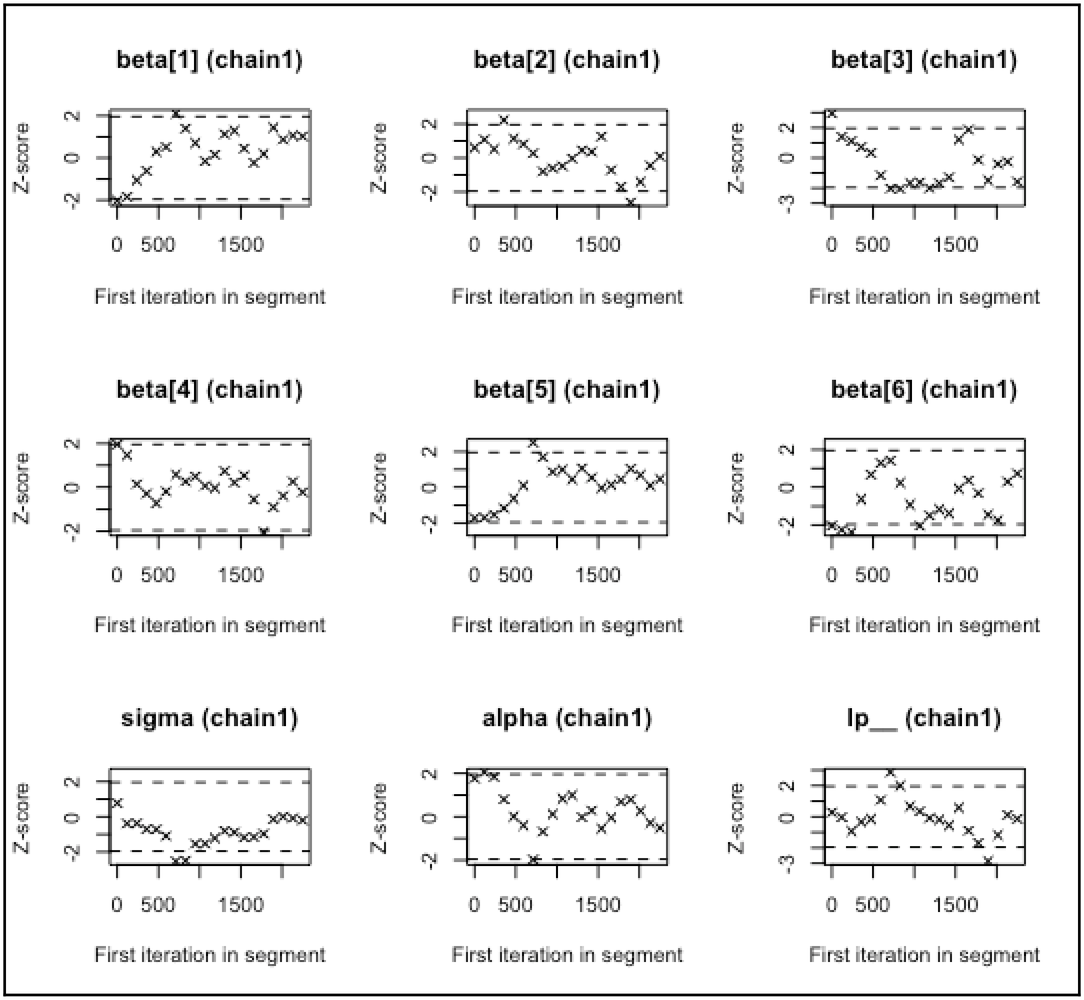

beta[1]
beta[2]
beta[3]
beta[4]
beta[5]
beta[6]
sigma
alpha
lp__
Autocorrelation
Lag
1.0
0.0
-1.0
0
5
10
15
20
25




```
> effectiveSize(coda__obj)
  beta[1]   beta[2]   beta[3]   beta[4]   beta[5]   beta[6]     sigma     alpha      lp__
10687.791 10298.065 10233.660 12026.854  7195.862 11783.860 15356.379  6442.785  4596.752
```


```
> effectiveSize(coda__obj)
  beta[1]   beta[2]   beta[3]   beta[4]   beta[5]   beta[6]     sigma     alpha      lp__
10687.791 10298.065 10233.660 12026.854  7195.862 11783.860 15356.379  6442.785  4596.752
```


beta[1]
beta[2]
beta[3]
beta[4]
beta[5]
beta[6]
sigma
alpha
lp__
shrink factor
last iteration in chain
median
97.5%


beta[1] (chain1)
beta[2] (chain1)
beta[3] (chain1)
beta[4] (chain1)
beta[5] (chain1)
beta[6] (chain1)
sigma (chain1)
alpha (chain1)
lp__ (chain1)
Z-score
First iteration in segment
0 500 1500

```
[[1]]

         Stationarity start      p-value
         test         iteration
beta[1] passed       1          0.5614
beta[2] passed       1          0.4287
beta[3] passed       1          0.3189
beta[4] passed       1          0.5609
beta[5] passed       1          0.4580
beta[6] passed       1          0.0979
sigma   passed       1          0.2446
alpha   passed       1          0.6065
lp__    passed       1          0.5820
```

```
[[1]]
                 lower        upper
beta[1]  5.251992e+00   6.00633618
beta[2]  3.408653e-01   1.41839437
beta[3]  6.126771e-01   1.87168109
beta[4]  1.178434e-03   0.83393762
beta[5]  8.508981e-02   1.54516980
beta[6]  1.377528e-02   0.55803580
sigma    5.285958e+00   6.85649074
alpha   -1.765699e+01  -0.09846626
lp__    -2.925712e+02 -283.47378452
attr(,"Probability")
[1] 0.95
```

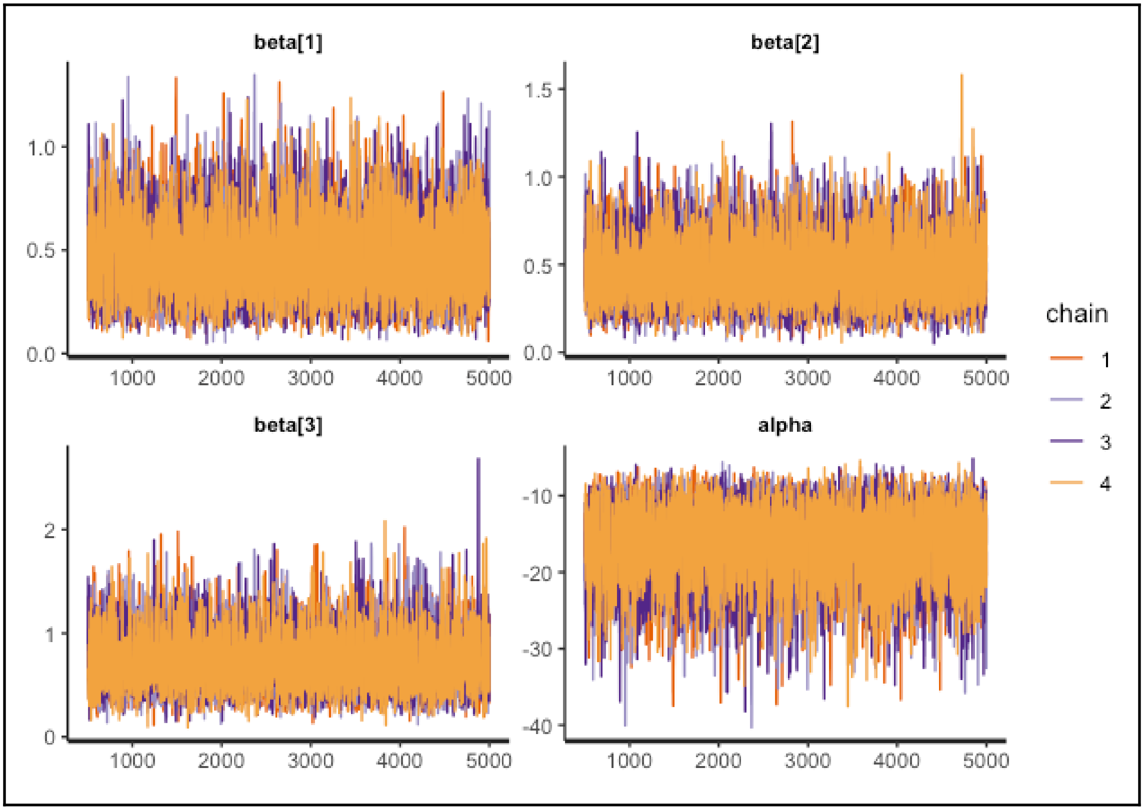
```
[[1]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

         Burn-in  Total Lower bound  Dependence
         (M)      (N)   (Nmin)       factor (I)
 beta[1] 5        5557  3746         1.48
 beta[2] 11       12191 3746         3.25
 beta[3] 11       11890 3746         3.17
 beta[4] 12       12504 3746         3.34
 beta[5] 28       28460 3746         7.60
 beta[6] 12       14714 3746         3.93
 sigma   4        4615  3746         1.23
 alpha   7        8087  3746         2.16
 lp__    10       12502 3746         3.34
```

```
Iterations = 5001:6000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

           Mean      SD Naive SE Time-series SE
b[1]   0.988601 0.05896 0.001865       0.015826
b[2]   0.964735 0.07764 0.002455       0.025873
b[3]   1.061319 0.05691 0.001800       0.014521
b[4]   0.032153 1.38067 0.043661       0.042851
b[5]  -0.020222 1.43672 0.045433       0.048452
b[6]   0.006282 1.41533 0.044757       0.046879
id[1]  1.000000 0.00000 0.000000       0.000000
id[2]  1.000000 0.00000 0.000000       0.000000
id[3]  1.000000 0.00000 0.000000       0.000000
id[4]  0.017000 0.12934 0.004090       0.010163
id[5]  0.016000 0.12554 0.003970       0.009451
id[6]  0.019000 0.13659 0.004319       0.008653

2. Quantiles for each variable:

         2.5%     25%        50%    75% 97.5%
b[1]   0.8861  0.9458  9.851e-01 1.0349 1.101
b[2]   0.8182  0.9103  9.533e-01 1.0170 1.124
b[3]   0.9550  1.0182  1.069e+00 1.1005 1.172
b[4]  -2.6413 -0.8223  8.321e-03 0.8902 2.668
b[5]  -2.7805 -0.9974  8.653e-03 0.9164 2.826
b[6]  -2.7208 -0.9623 -4.815e-05 0.9518 2.906
id[1]  1.0000  1.0000  1.000e+00 1.0000 1.000
id[2]  1.0000  1.0000  1.000e+00 1.0000 1.000
id[3]  1.0000  1.0000  1.000e+00 1.0000 1.000
id[4]  0.0000  0.0000  0.000e+00 0.0000 0.000
id[5]  0.0000  0.0000  0.000e+00 0.0000 0.000
id[6]  0.0000  0.0000  0.000e+00 0.0000 0.000
```

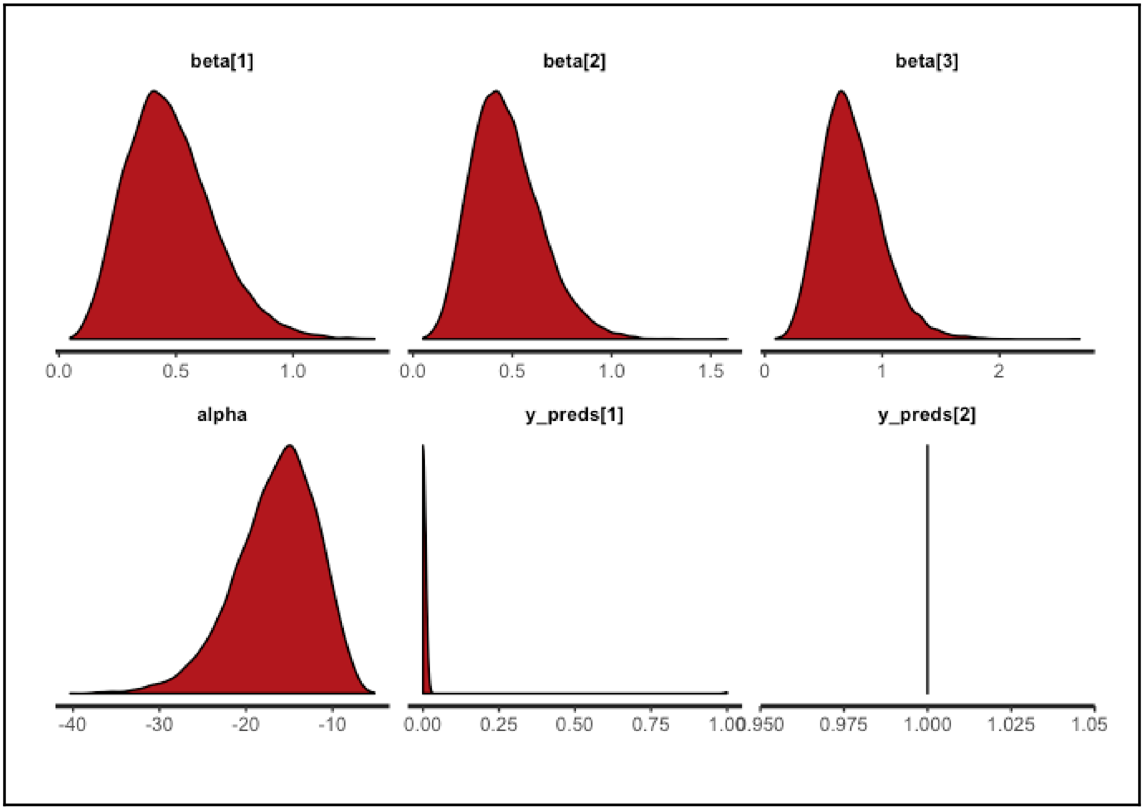
```
Iterations = 2001:3000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean      SD Naive SE Time-series SE
b[1]   1.00132 0.09977 0.003155        0.04589
b[2]   1.01713 0.10036 0.003174        0.04474
b[3]   1.05612 0.08919 0.002820        0.03660
b[4]  -0.15114 0.82688 0.026148        0.02646
b[5]   0.05981 1.09685 0.034685        0.03469
b[6]  -0.02838 1.38076 0.043663        0.04163
id[1]  1.00000 0.00000 0.000000        0.00000
id[2]  1.00000 0.00000 0.000000        0.00000
id[3]  1.00000 0.00000 0.000000        0.00000
id[4]  0.65300 0.47625 0.015060        0.34883
id[5]  0.28800 0.45306 0.014327        0.16553
id[6]  0.02700 0.16216 0.005128        0.01380

2. Quantiles for each variable:

         2.5%     25%      50%      75% 97.5%
b[1]   0.8053  0.9291  1.01673  1.06108 1.201
b[2]   0.8755  0.9436  0.99416  1.06865 1.288
b[3]   0.8850  0.9828  1.05490  1.13211 1.209
b[4]  -2.0891 -0.2335 -0.16598 -0.08432 1.918
b[5]  -2.3179 -0.4676  0.12773  0.53338 2.269
b[6]  -2.8256 -0.9137 -0.01822  0.93618 2.565
id[1]  1.0000  1.0000  1.00000  1.00000 1.000
id[2]  1.0000  1.0000  1.00000  1.00000 1.000
id[3]  1.0000  1.0000  1.00000  1.00000 1.000
id[4]  0.0000  0.0000  1.00000  1.00000 1.000
id[5]  0.0000  0.0000  0.00000  1.00000 1.000
id[6]  0.0000  0.0000  0.00000  0.00000 1.000
```

```
               mean      se_mean         sd          2.5%           25%           50%           75%         97.5%     n_eff      Rhat
beta[1]       5.6275059 0.0009294059 0.10789966     5.4185152     5.5546606     5.6269981     5.6997151     5.8401070 13478.119 0.9998874
beta[2]       1.1518596 0.0010883722 0.14589311     0.8665166     1.0538650     1.1507702     1.2512684     1.4381903 17968.618 0.9999088
beta[3]       1.4128635 0.0013060201 0.17515826     1.0702645     1.2953674     1.4124103     1.5301027     1.7587253 17987.115 0.9999276
beta[4]       0.6676620 0.0016614802 0.21000246     0.2529046     0.5266094     0.6679713     0.8072659     1.0855280 15975.646 0.9998832
beta[5]       1.3740549 0.0018720069 0.21455924     0.9597978     1.2291736     1.3722468     1.5192225     1.7936711 13136.497 0.9999206
beta[6]       0.4056779 0.0007691102 0.09220588     0.2247098     0.3441523     0.4058902     0.4679650     0.5854535 14372.759 0.9999855
sigma         6.5077550 0.0017798626 0.21202722     6.1074628     6.3597243     6.5015079     6.6468875     6.9351702 14190.911 1.0001393
alpha       -16.7700929 0.0268396435 2.55982821   -21.8520041   -18.5110227   -16.7427922   -15.0200079   -11.8247415  9096.371 0.9999690
y_preds[1]   61.3931065 0.0492257745 6.55173940    48.6690822    56.9603209    61.4012673    65.8555620    74.2613698 17714.468 0.9998845
y_preds[2]   40.5941728 0.0498854855 6.57418160    27.4423010    36.1674000    40.6039721    45.0557747    53.3749644 17367.407 0.9999803
lp__      -1141.1022314 0.0213353678 2.02075400 -1145.8784114 -1142.2271040 -1140.7874901 -1139.6091931 -1138.1672875  8970.706 1.0001144
```

```
> extract(fit)$y_preds

iterations        [,1]     [,2]
      [1,] 58.25559 54.86578
      [2,] 55.39696 34.32738
      [3,] 66.14404 28.76434
      [4,] 59.08469 46.10736
      [5,] 58.27192 49.48367
      [6,] 67.84411 40.79301
      [7,] 70.30991 42.59055
      [8,] 49.65133 37.09685
      [9,] 74.10806 44.17806
     [10,] 55.92604 40.12660
     [11,] 56.21416 40.60646
```



	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	0.4756442	0.0024730002	0.18501358	0.1740971	0.3422852	0.4566114	0.5881280	0.8926634	5597.046	1.0002977
beta[2]	0.4708353	0.0019771001	0.17429737	0.1871279	0.3452296	0.4509138	0.5762759	0.8647380	7771.849	1.0003868
beta[3]	0.7399918	0.0028128931	0.25691916	0.3188437	0.5578754	0.7104427	0.8925439	1.3216234	8342.314	0.9999817
alpha	-16.5769443	0.0652489584	4.78373421	-27.2921842	-19.4677494	-16.0367246	-13.1361982	-8.7693224	5375.105	1.0001868
y_preds[1]	0.0050000	0.0005303301	0.07053564	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	17689.868	1.0000827
y_preds[2]	1.0000000	NaN	0.00000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	NaN	NaN
lp__	-27.3735049	0.0204878662	1.49291405	-31.2031392	-28.1092573	-27.0202943	-26.2841327	-25.5254935	5309.775	1.0005955



beta[1]
beta[2]
beta[3]
0.0
0.5
1.0
0.0
0.5
1.0
1.5
0
1
2
alpha
y_preds[1]
y_preds[2]
-40
-30
-20
-10
0.00
0.25
0.50
0.75
1.00
0.950
0.975
1.000
1.025
1.05

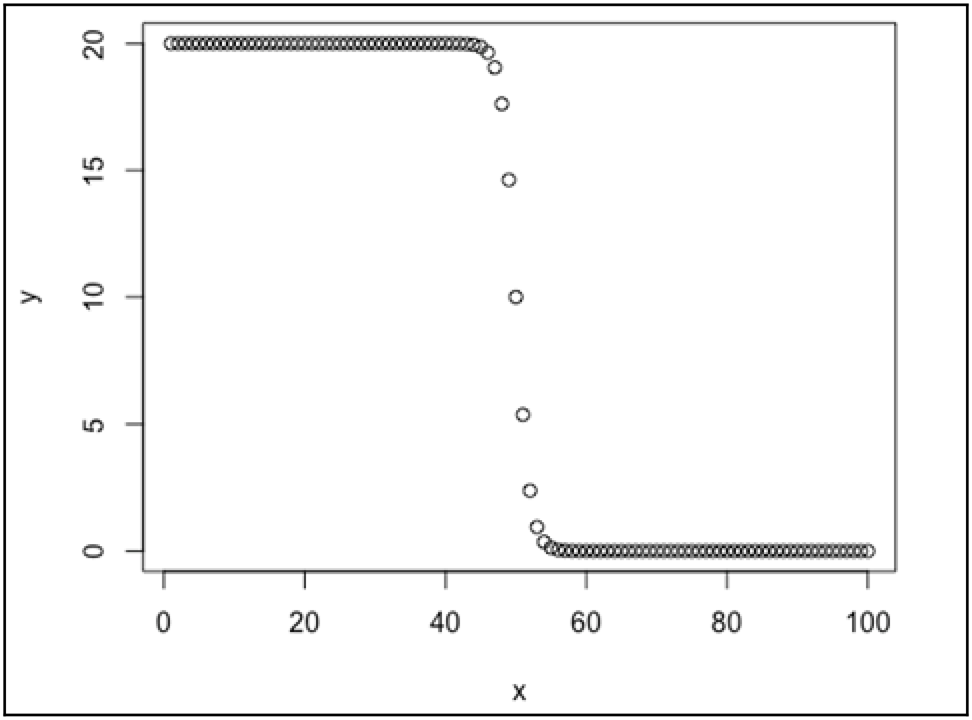
Chapter 5: Nonparametric Methods

```
> wilcox.test(Height ~ Sample,data=data)

	Wilcoxon rank sum test with continuity correction

data:  Height by Sample
W = 4868, p-value = 0.841
alternative hypothesis: true location shift is not equal to 0
```

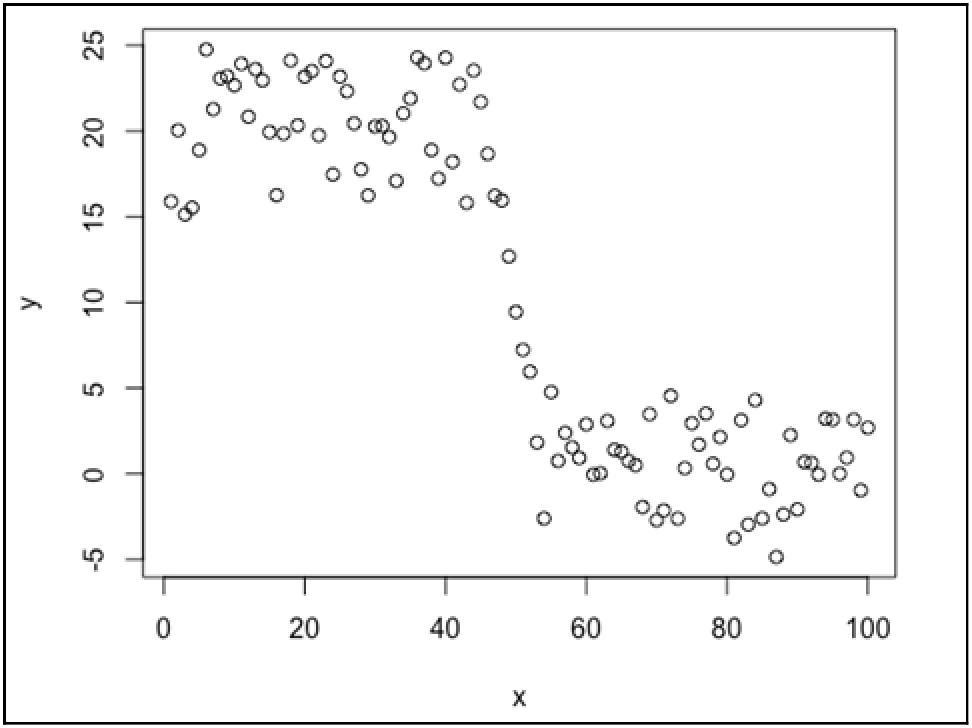
```
> wilcox.test(data$pre_bonus, data$post_bonus,paired=TRUE)

	Wilcoxon signed rank test with continuity correction

data:  data$pre_bonus and data$post_bonus
V = 921, p-value = 0.0008075
alternative hypothesis: true location shift is not equal to 0
```

```
> kruskal.test(Result ~ Food.Type,data=t)

	Kruskal-Wallis rank sum test

data:  Result by Food.Type
Kruskal-Wallis chi-squared = 24.629, df = 2, p-value = 4.487e-06
```

```
  Comparison        Z      P.unadj        P.adj
1      A - B 3.612376 3.034039e-04 6.068078e-04
2      A - C 4.753127 2.002948e-06 6.008843e-06
3      B - C 1.140750 2.539738e-01 2.539738e-01
```


y
20
15
10
5
0
0
20
40
60
80
100
x

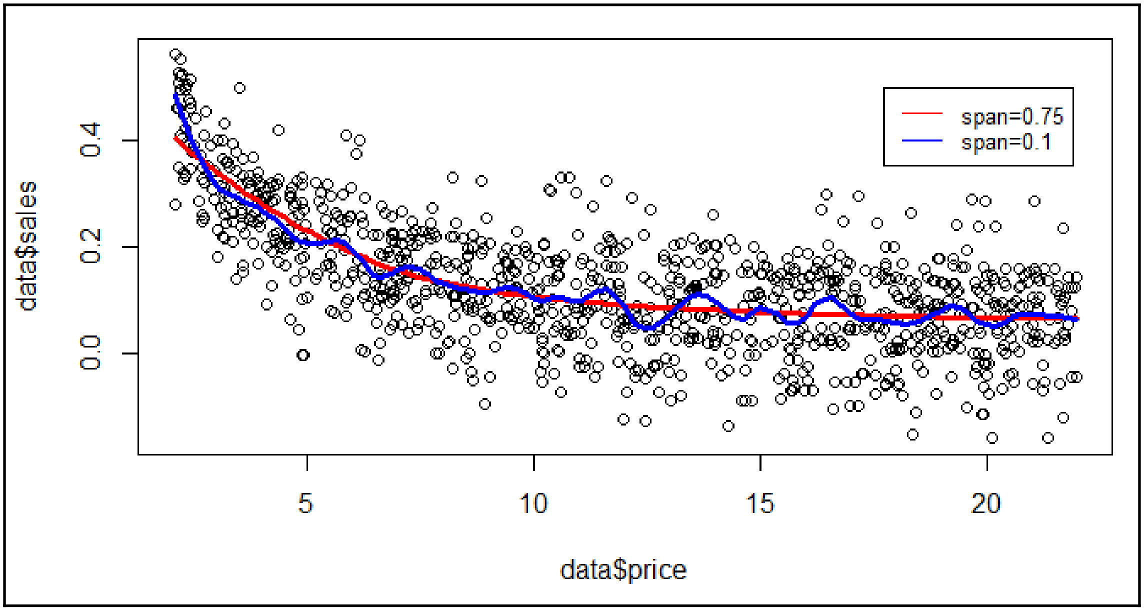
```
> cor.test( ~ x + y, method = "spearman",conf.level = 0.95)

        Spearman's rank correlation rho

data:  x and y
S = 333120, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.9989073

Warning message:
In cor.test.default(x = 1:100, y = c(20, 20, 20, 20, 20, 20, 20,
  Cannot compute exact p-value with ties
>
> cor.test( ~ x + y, method = "pearson",conf.level = 0.95)

        Pearson's product-moment correlation

data:  x and y
t = -18.594, df = 98, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9196625 -0.8302144
sample estimates:
       cor
-0.8826926
```



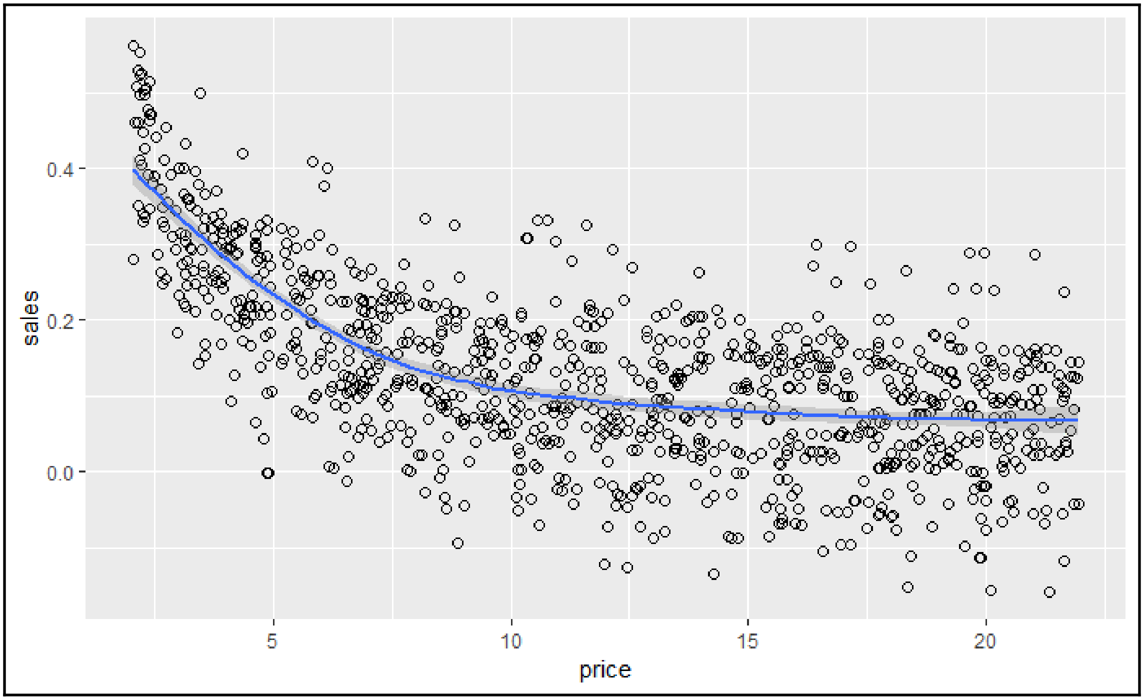
25
20
15
10
5
0
-5
y
0
20
40
60
80
100
x

```
        Spearman's rank correlation rho

data:  x and y
S = 292080, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.7526793

>
> cor.test( ~ x + y, method = "pearson",conf.level = 0.95)

        Pearson's product-moment correlation

data:  x and y
t = -14.616, df = 98, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.8810873 -0.7541889
sample estimates:
       cor
-0.8279566
```

```
        Spearman's rank correlation rho

data:  salary and educ_level
S = 4, p-value = 0.03333
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8857143
```

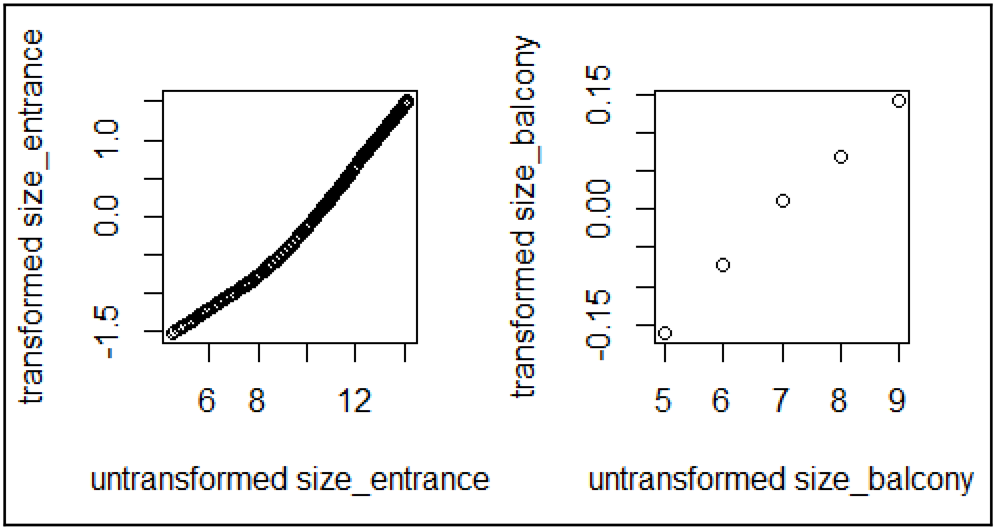


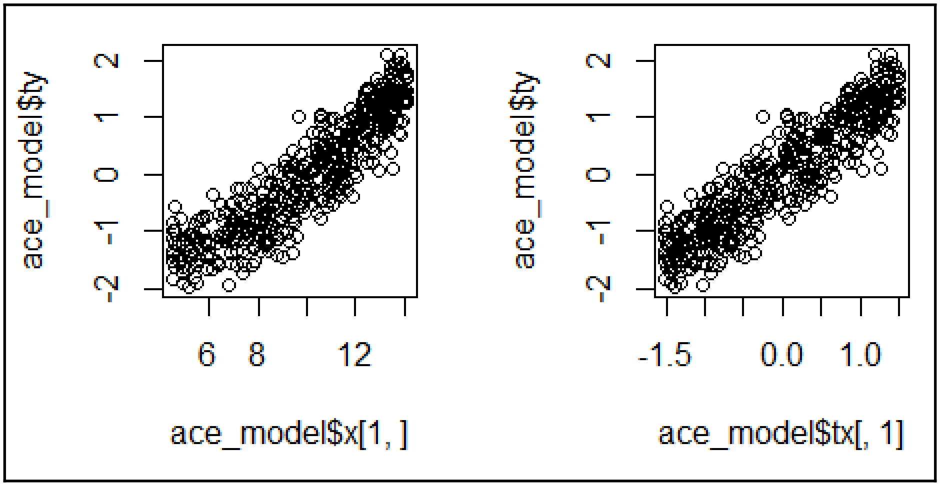

```
> loess1_wrapper(5)   - loess1_wrapper(10)
[1] 0.123627
> loess1_wrapper(10)  - loess1_wrapper(15)
[1] 0.02789907
```


0.4
0.2
0.0
sales
5
10
15
20
price

```
Call:
lm(formula = Property_price ~ size + number.bathrooms + number.bedrooms +
    number.entrances + size_balcony + size_entrance, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-18.4347  -4.6414  -0.0136   4.4449  23.6712

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -16.7063     2.5520  -6.546 1.54e-10 ***
size               5.6270     0.1068  52.666  < 2e-16 ***
number.bathrooms   1.1513     0.1457   7.904 1.91e-14 ***
number.bedrooms    1.4115     0.1758   8.028 7.86e-15 ***
number.entrances   0.6621     0.2079   3.186  0.00154 **
size_balcony       1.3693     0.2137   6.408 3.56e-10 ***
size_entrance      0.4055     0.0906   4.476 9.54e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.491 on 474 degrees of freedom
Multiple R-squared:  0.8612,	Adjusted R-squared:  0.8594
F-statistic: 490.1 on 6 and 474 DF,  p-value: < 2.2e-16
```

```
> ace_model$rsq
[1] 0.89131
```

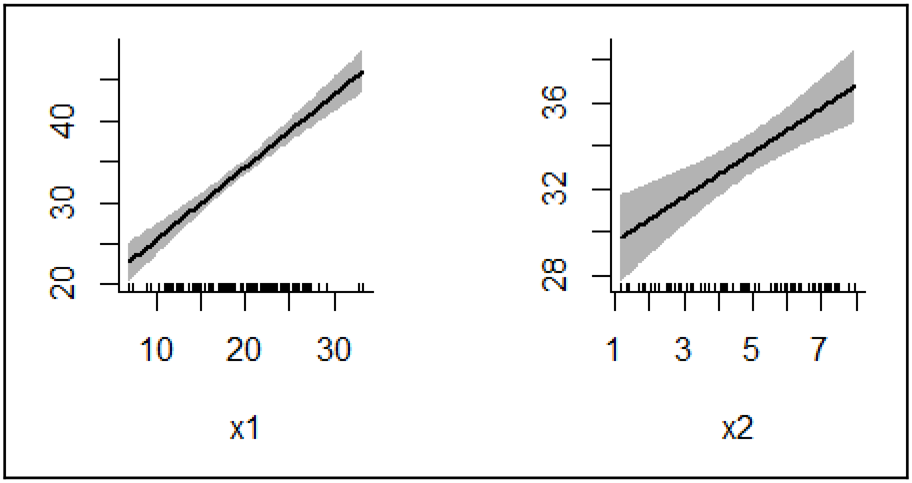




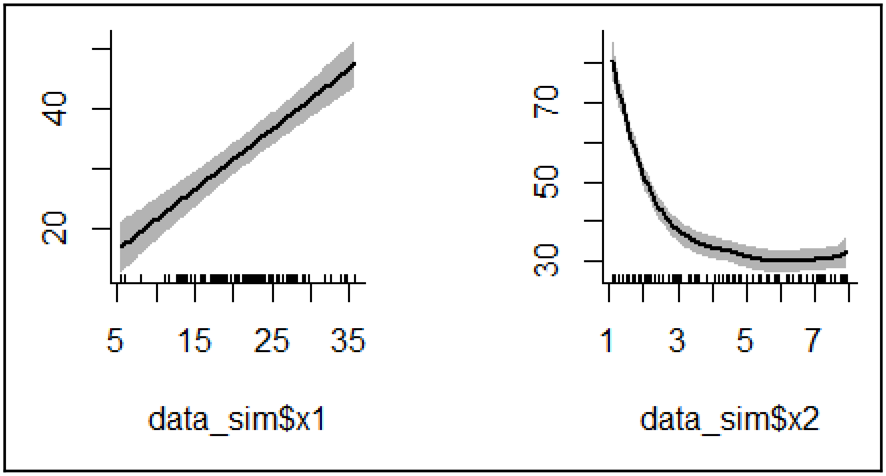

```
$`results`
                                                    Test Statistic   df1      df2 P-value Permutation Test p-value
ANOVA type test p-value                                    178.511 3.826 281.2335       0                        0
McKeon approx. for the Lawley Hotelling Test               316.457 8.000 203.4024       0                        0
Muller approx. for the Bartlett-Nanda-Pillai Test           67.965 8.162 293.8913       0                        0
Wilks Lambda                                               155.763 8.000 288.0000       0                        0

$releffects
           Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa          0.19427     0.75307      0.16667     0.16667
versicolor      0.54767     0.30387      0.50593     0.50653
virginica       0.75807     0.44307      0.82740     0.82680
```

```
The ANOVA type statistic will be used in the following test
The Global Hypothesis is significant, subset algorithm will continue

~Performing the Subset Algorithm based on Factor levels~
The Hypothesis of equality between factor levels  setosa versicolor virginica is rejected
The Hypothesis of equality between factor levels  versicolor virginica is rejected
The Hypothesis of equality between factor levels  setosa virginica is rejected
The Hypothesis of equality between factor levels  setosa versicolor is rejected
All appropriate subsets using factor levels have been checked using a closed multiple testing procedure, which controls the max
imum overall type I error rate at alpha= 0.05

~Performing the Subset Algorithm based on Response Variables~
 The Hypothesis of equality using response variables  Sepal.Length Sepal.Width Petal.Length Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Width Petal.Length Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Length Petal.Length Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Length Sepal.Width Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Length Sepal.Width Petal.Length is rejected
The Hypothesis of equality using response variables  Petal.Length Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Width Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Width Petal.Length is rejected
The Hypothesis of equality using response variables  Sepal.Length Petal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Length Petal.Length is rejected
The Hypothesis of equality using response variables  Sepal.Length Sepal.Width is rejected
The Hypothesis of equality using response variables  Petal.Width is rejected
The Hypothesis of equality using response variables  Petal.Length is rejected
The Hypothesis of equality using response variables  Sepal.Width is rejected
The Hypothesis of equality using response variables  Sepal.Length is rejected
All appropriate subsets using response variables have been checked using a multiple testing procedure, which controls the maxim
um overall type I error rate at alpha= 0.05
```

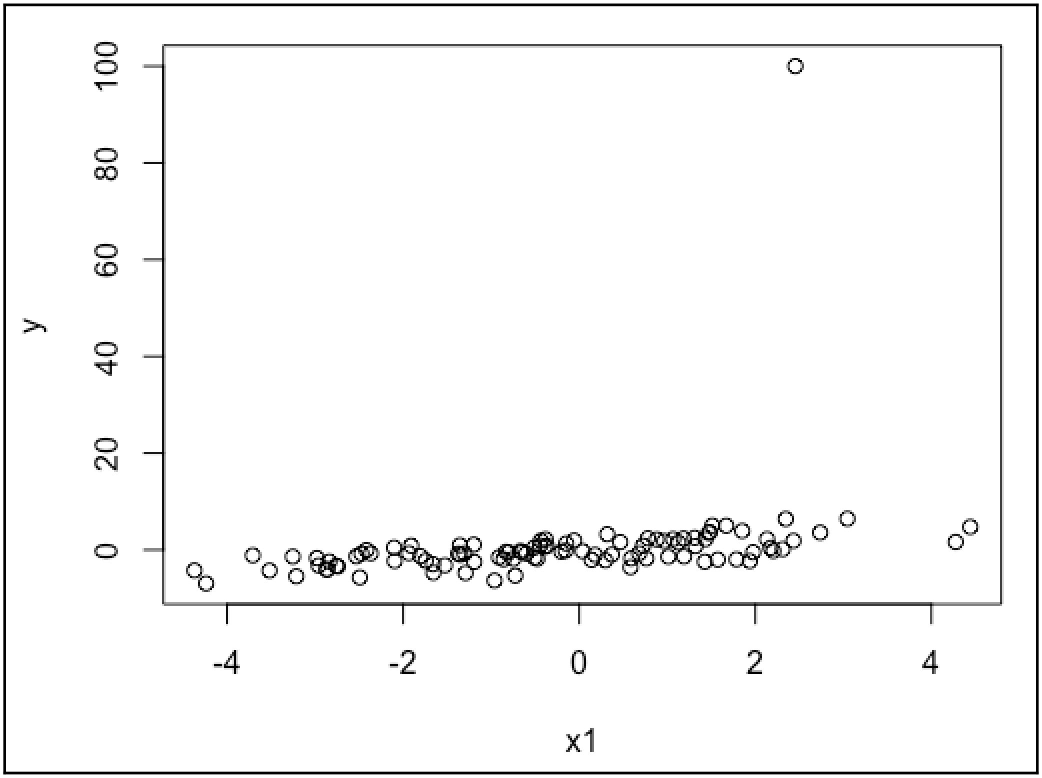
```
Summary for linear components:

             coef      se  ratio p-value
intercept 11.5000 2.31800  4.962       0
x1         0.8857 0.08716 10.160       0
x2         1.0370 0.23420  4.428       0
```



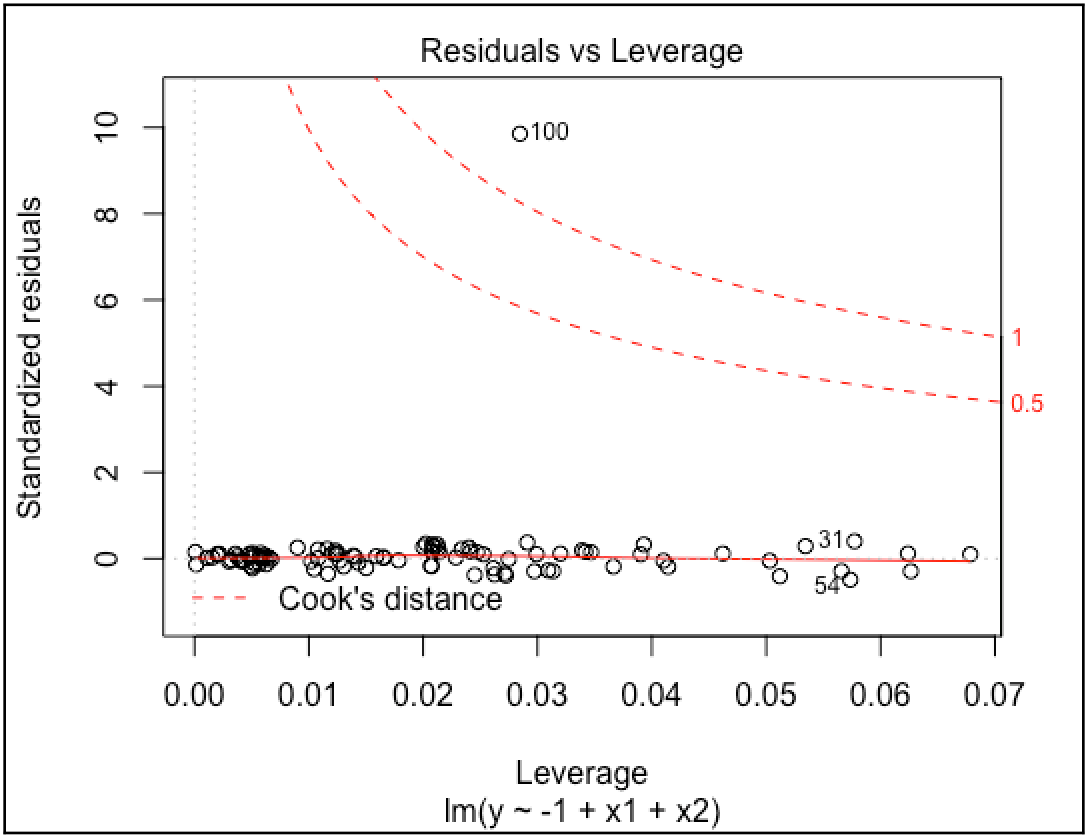

```
Summary for non-linear components:

               df  spar knots
f(data_sim$x1)  1 12870    24
f(data_sim$x2)  1  2592    24
```




```
Summary for linear components:

              coef      se  ratio p-value
intercept   11.220 24.3000 0.4616  0.6454
data_sim$x1  1.012  0.1022 9.8980  0.0000



Summary for non-linear components:

                  df  spar knots
f(data_sim$x2) 5.145 2.209    24
```


100
80
60
40
20
0
y
-4
-2
0
2
4
x1

```
> summary(e)

Call:
lm(formula = y ~ -1 + x1 + x2)

Residuals:
   Min      1Q Median      3Q     Max 
-4.748 -1.088  0.318  1.334 96.795 

Coefficients:
   Estimate Std. Error t value Pr(>|t|)   
x1   1.6489     0.5270   3.129  0.00231 **
x2   0.3825     0.5146   0.743  0.45911   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.966 on 98 degrees of freedom
Multiple R-squared:  0.09393,	Adjusted R-squared:  0.07544 
F-statistic:  5.08 on 2 and 98 DF,  p-value: 0.00796
```


Residuals vs Leverage
Standardized residuals
0
2
4
6
8
10
100
1
0.5
31
54
Cook's distance
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Leverage
lm(y ~ -1 + x1 + x2)

```
> summary(rlm_model)

Call: rlm(formula = y ~ -1 + x1 + x2, psi = psi.huber)
Residuals:
    Min      1Q  Median      3Q     Max
-2.6973 -0.6242  0.0971  0.6817 99.6149

Coefficients:
   Value  Std. Error t value
x1  0.9795  0.0524    18.7067
x2  0.9179  0.0511    17.9545

Residual standard error: 0.9639 on 98 degrees of freedom
```

```
> summary(rlm_model)

Call: rlm(formula = y ~ -1 + x1 + x2, psi = psi.hampel)
Residuals:
     Min       1Q   Median       3Q      Max
-2.64864 -0.58826  0.09075  0.64438 99.64157

Coefficients:
   Value  Std. Error t value
x1  0.9826  0.0509    19.2935
x2  0.9335  0.0497    18.7727

Residual standard error: 0.9359 on 98 degrees of freedom
```

```
> summary(rlm_model)

Call: rlm(formula = y ~ -1 + x1 + x2, psi = psi.bisquare)
Residuals:
    Min      1Q  Median      3Q     Max
-2.6774 -0.6236  0.0923  0.6736 99.6521

Coefficients:
   Value   Std. Error t value
x1  0.9725  0.0526    18.4780
x2  0.9270  0.0514    18.0394

Residual standard error: 0.9702 on 98 degrees of freedom
```

```
> covClassic(d,cor = TRUE)
Call:
covClassic(data = d, corr = TRUE)

Classical Estimate of Correlation:
       V1     V2
V1 1.0000 0.4889
V2 0.4889 1.0000

Classical Estimate of Location:
   V1    V2
5.112 5.071
```

```
> cov.rob(d,cor = TRUE)
$center
[1] 5.119222 5.080008
```

```
$cor
          [,1]      [,2]
[1,] 1.0000000 0.5371111
[2,] 0.5371111 1.0000000
```

```
Call:
covClassic(data = d, corr = TRUE)

Classical Estimate of Correlation:
       V1     V2
V1 1.0000 0.5341
V2 0.5341 1.0000

Classical Estimate of Location:
   V1    V2
5.733 5.811
```

```
$center
[1] 5.095713 5.066022
```

```
$cor
          [,1]      [,2]
[1,] 1.0000000 0.5278468
[2,] 0.5278468 1.0000000
```

```
> covClassic(d,cor = TRUE)
Call:
covClassic(data = d, corr = TRUE)

Classical Estimate of Correlation:
       V1     V2
V1 1.0000 0.6128
V2 0.6128 1.0000

Classical Estimate of Location:
   V1    V2
7.883 7.874
```

```
> cov.rob(d,cor = TRUE)
$center
[1] 5.097614 5.062966
```

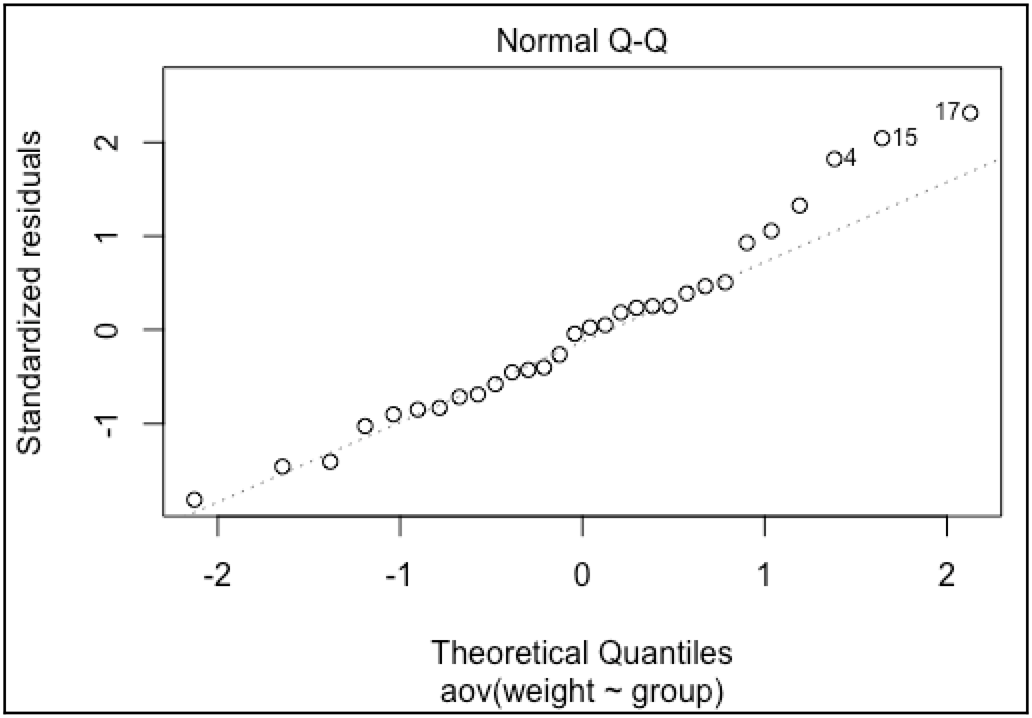
```
$cor
          [,1]      [,2]
[1,] 1.0000000 0.5056711
[2,] 0.5056711 1.0000000
```

```
> glm( y~x1+x2,data=df,family="binomial")

Call:  glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Coefficients:
(Intercept)           x1           x2
      1.767        1.927        4.990

Degrees of Freedom: 999 Total (i.e. Null);  997 Residual
Null Deviance:      1307
Residual Deviance: 440.5        AIC: 446.5
> robust::glmRob(y~x1+x2,data=df,family="binomial")
Call:
robust::glmRob(formula = y ~ x1 + x2, family = "binomial", data = df)

Coefficients:
(Intercept)          x1          x2
      1.764       1.923       4.981

Degrees of Freedom: 1000 Total; 997 Residual
Residual Deviance: 440.5
```

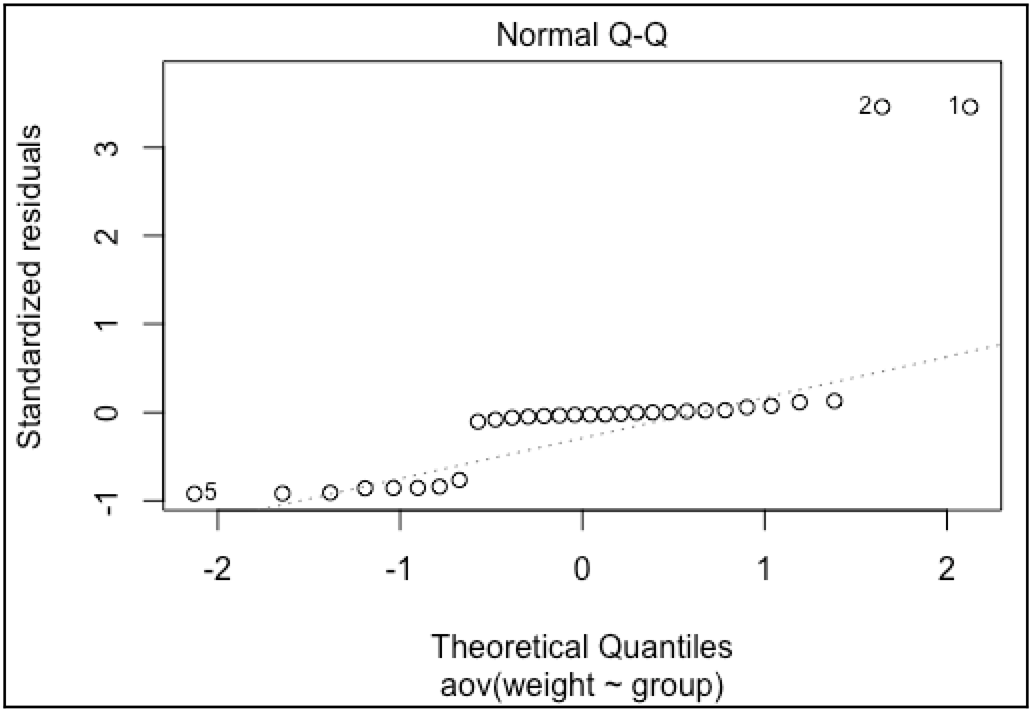
```
> glm( y~x1+x2,data=df,family="binomial")

Call:  glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Coefficients:
(Intercept)           x1           x2
      1.767        1.927        4.990

Degrees of Freedom: 999 Total (i.e. Null);  997 Residual
Null Deviance:      1307
Residual Deviance: 440.5        AIC: 446.5
> robust::glmRob(y~x1+x2,data=df,family="binomial")
Call:
robust::glmRob(formula = y ~ x1 + x2, family = "binomial", data = df)

Coefficients:
(Intercept)          x1          x2
      1.764       1.923       4.981

Degrees of Freedom: 1000 Total; 997 Residual
Residual Deviance: 440.5
```

```
> glm( y~x1+x2,data=df,family="binomial")

Call:  glm(formula = y ~ x1 + x2, family = "binomial", data = df)

Coefficients:
(Intercept)           x1           x2
    1.36236      0.09564      3.46004

Degrees of Freedom: 999 Total (i.e. Null);  997 Residual
Null Deviance:      1320
Residual Deviance: 602.9        AIC: 608.9
> robust::glmRob(y~x1+x2,mthod="cubif",data=df,family="binomial")
Call:
robust::glmRob(formula = y ~ x1 + x2, family = "binomial", data = df,
    mthod = "cubif")

Coefficients:
(Intercept)          x1          x2
     1.3982      0.2321      3.4954

Degrees of Freedom: 1000 Total; 997 Residual
Residual Deviance: 630.7
```

```
> summary(d)
            Df Sum Sq Mean Sq F value Pr(>F)
group        2  3.766  1.8832   4.846 0.0159 *
Residuals   27 10.492  0.3886
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```




```
> robust::anova.lmRob(robanova)

Terms added sequentially (first to last)

            Chisq Df RobustF    Pr(F)
(Intercept)        1
group              2  9.3139 0.001871 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```




```
> summary(d)
            Df Sum Sq Mean Sq F value Pr(>F)
group        2    539   269.7   2.248  0.125
Residuals   27   3239   120.0
```

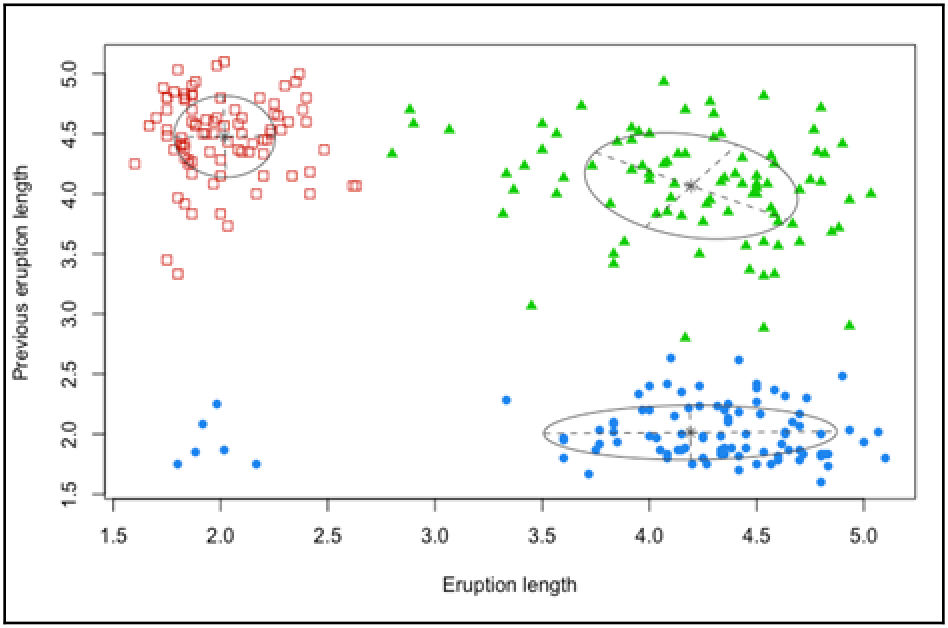
```
> robust::anova.lmRob(robanova)

Terms added sequentially (first to last)

            Chisq Df RobustF    Pr(F)
(Intercept)        1
group              2  7.9748 0.004006 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

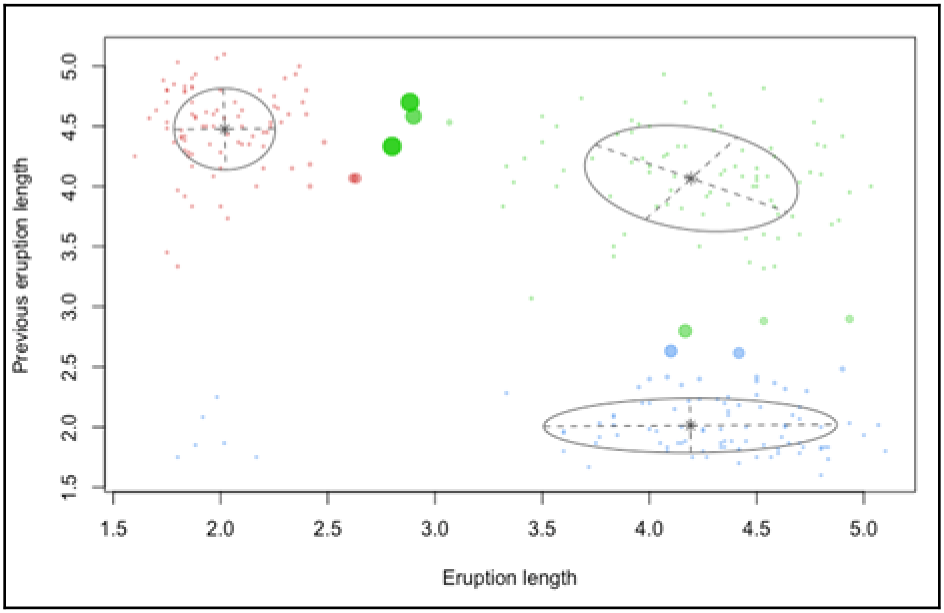
```
Standard deviations (1, .., p=10):
 [1] 1.4164259 1.3425988 1.2985310 1.0337184 0.9984854 0.9763466 0.9583313 0.4550208 0.4264126 0.4229815
```

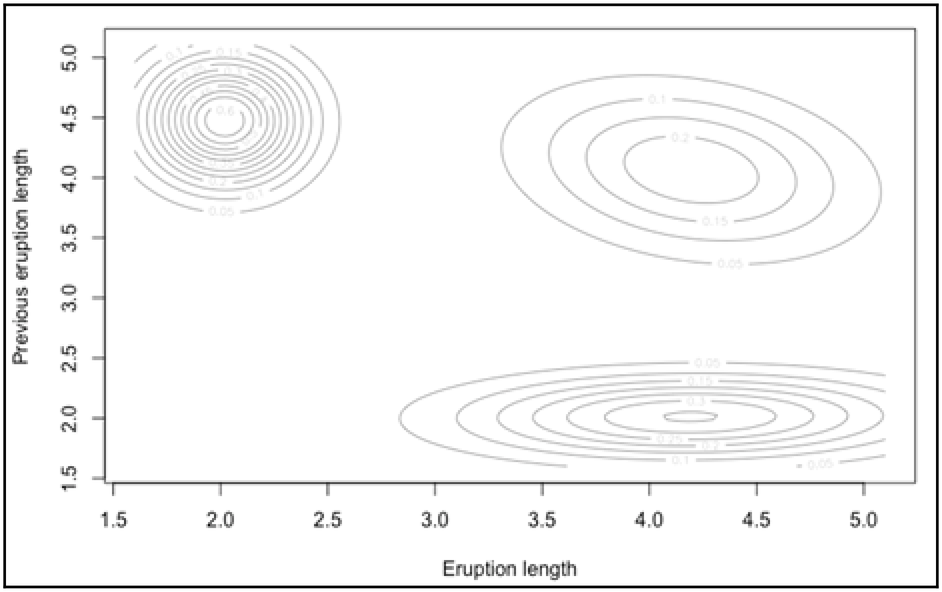
```
> x1$eigenvalues
      PC1       PC2       PC3       PC4       PC5       PC6
2.1628126 1.8364713 1.6630791 1.0449954 0.9669290 0.9306331
```

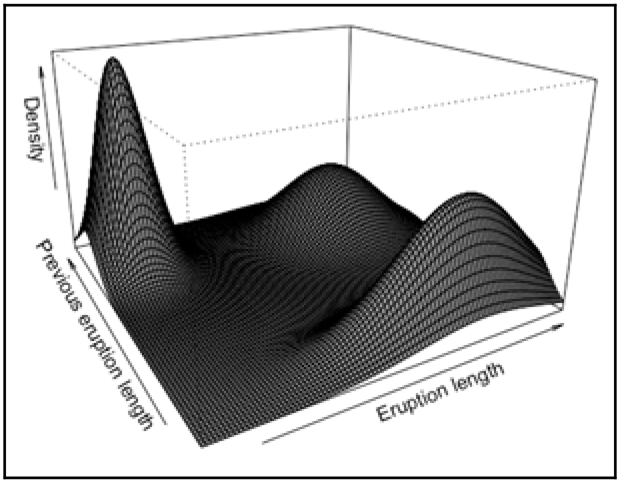
```
Standard deviations (1, .., p=10):
 [1] 3.15375504 0.11429621 0.10610480 0.09134895 0.08006696 0.07545510 0.07356575 0.03584887 0.03512249 0.03358356
```

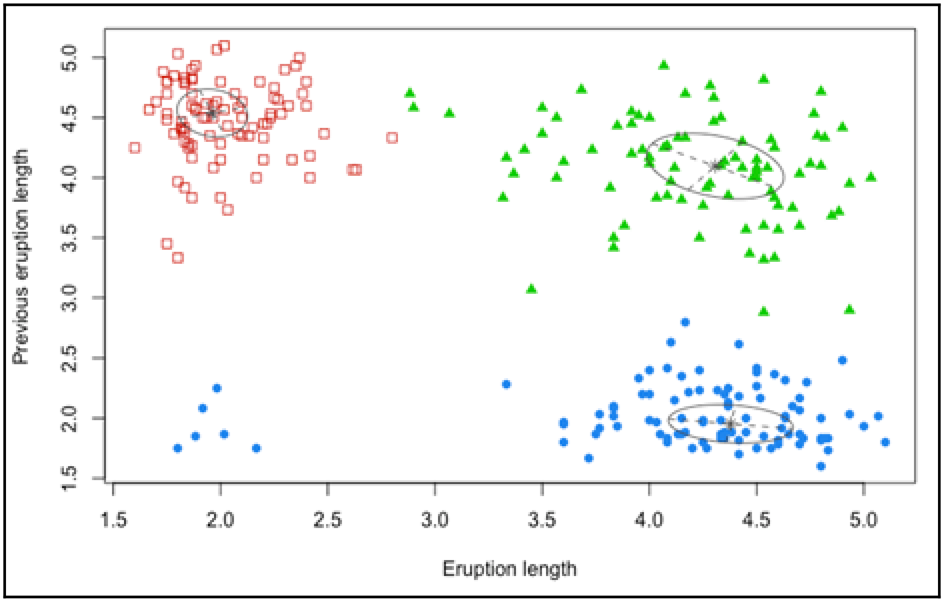
```
> x1$eigenvalues
      PC1       PC2       PC3       PC4       PC5       PC6
2.1675611 1.8456676 1.6573998 1.0494377 0.9724306 0.9306576
```

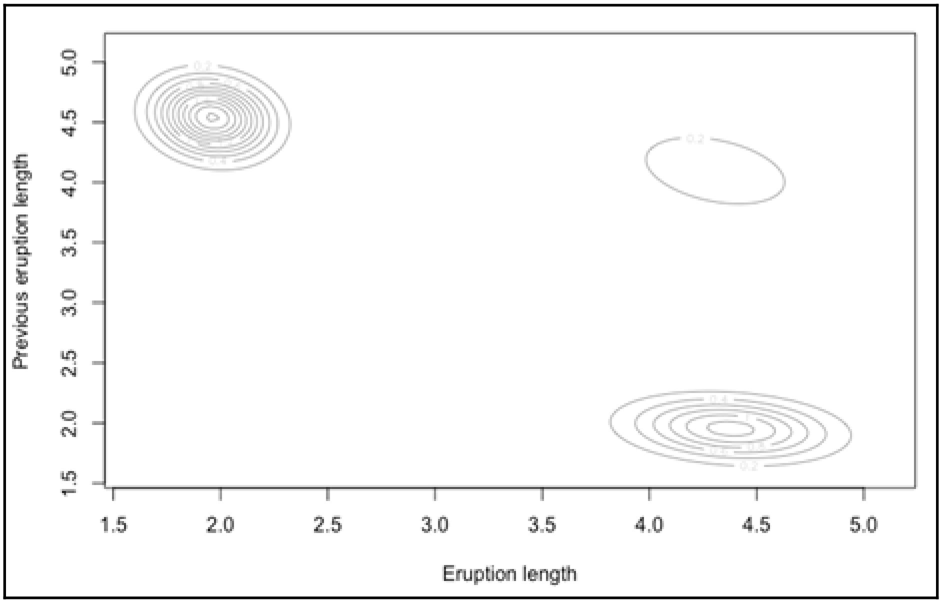

Previous eruption length
Eruption length
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

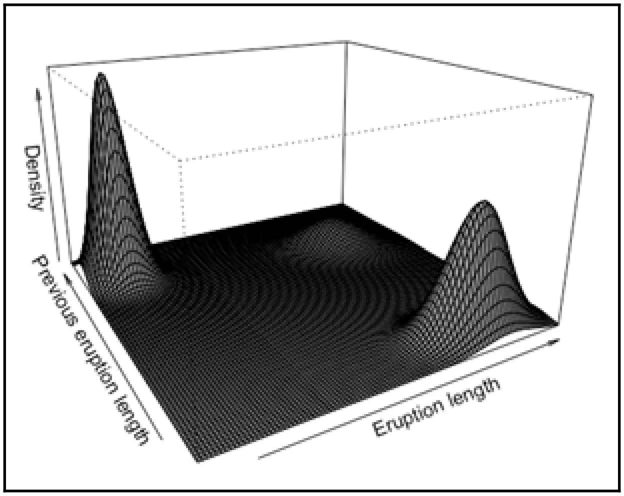

Previous eruption length
Eruption length
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0


Previous eruption length
Eruption length
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

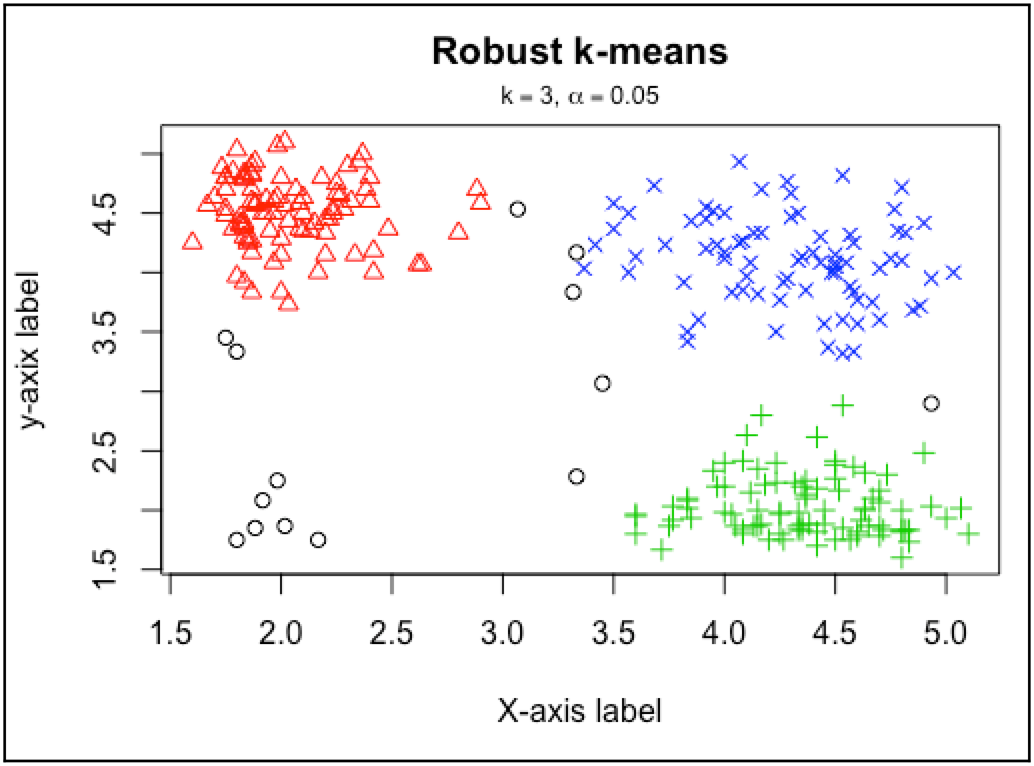

Density
Previous eruption length
Eruption length

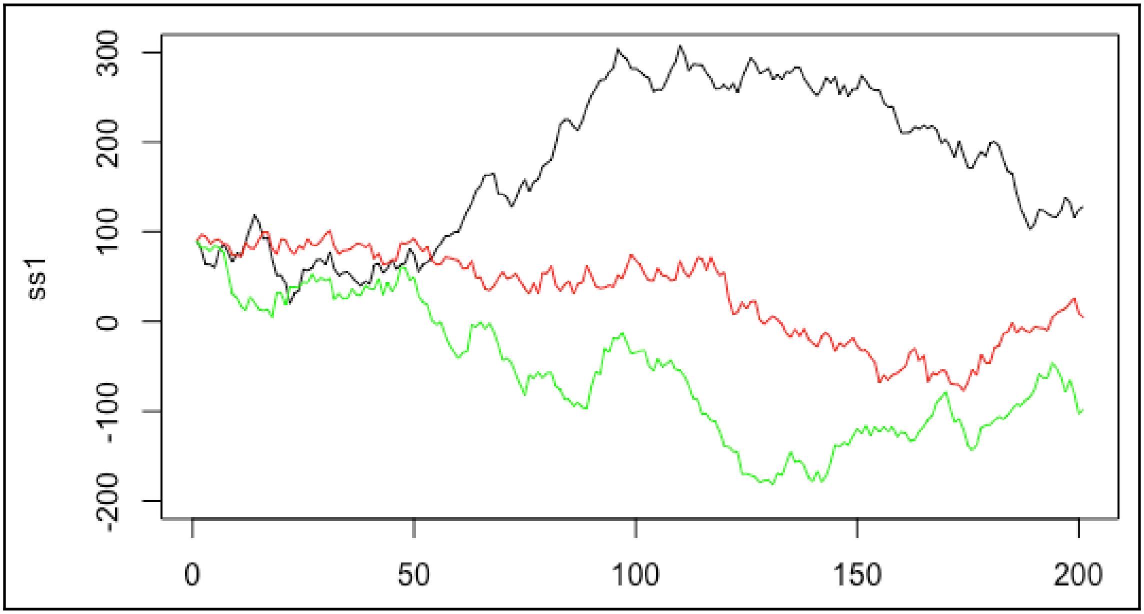

Previous eruption length
Eruption length
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

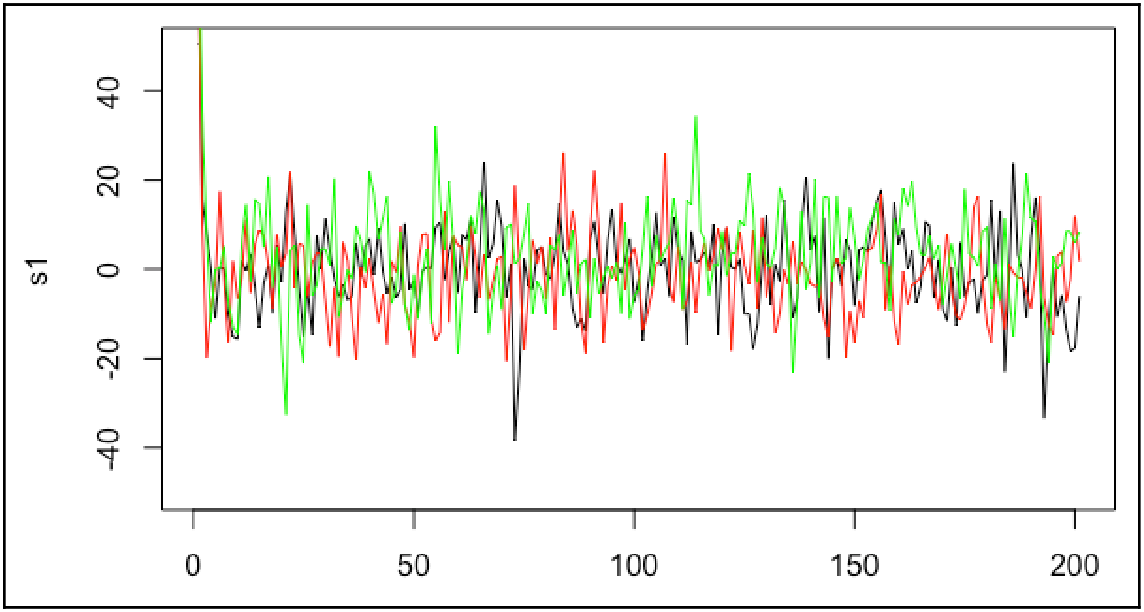

Previous eruption length
Eruption length
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

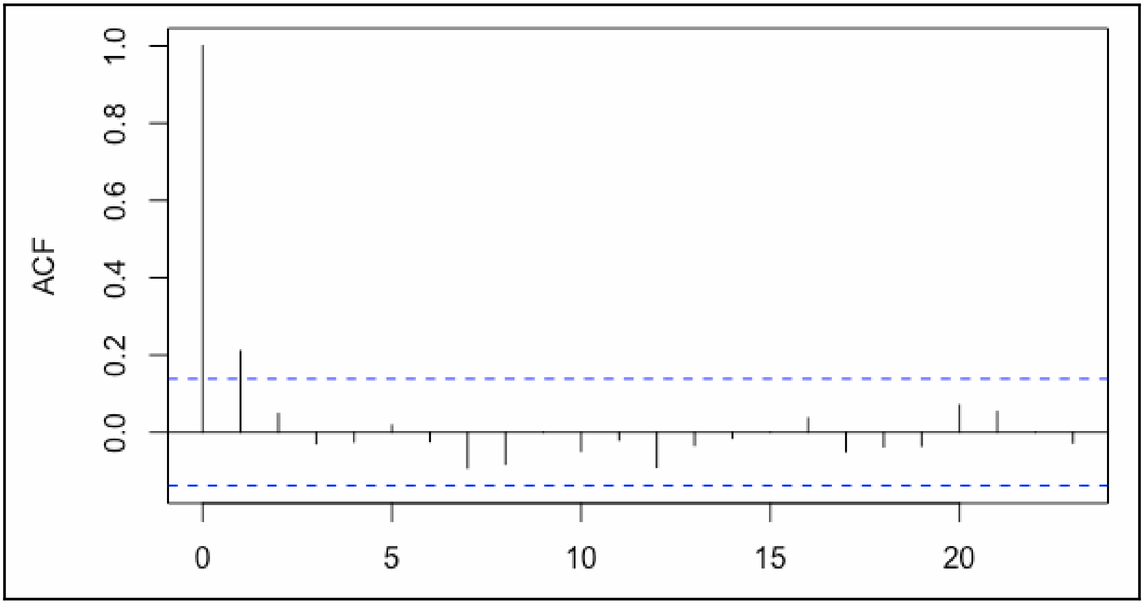

Density
Previous eruption length
Eruption length


Previous eruption length
Eruption length
5
4
3
2
2
3
4
5

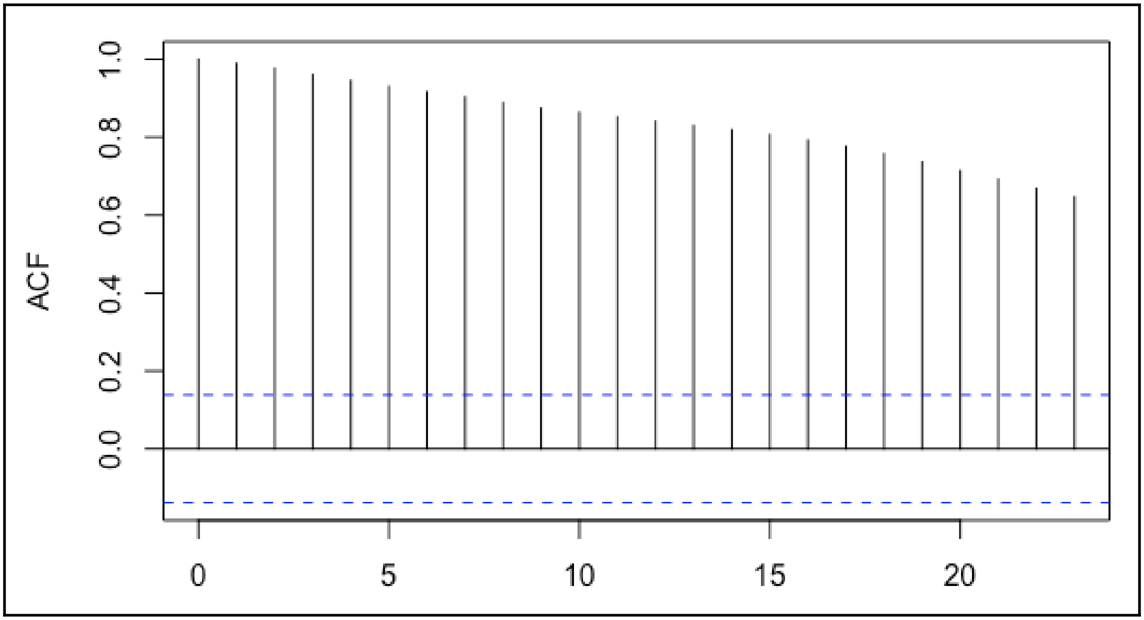

Robust k-means
k = 3, α = 0.05
y-axix label
X-axis label
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
1.5
2.5
3.5
4.5


ss1
300
200
100
0
-100
-200
0
50
100
150
200

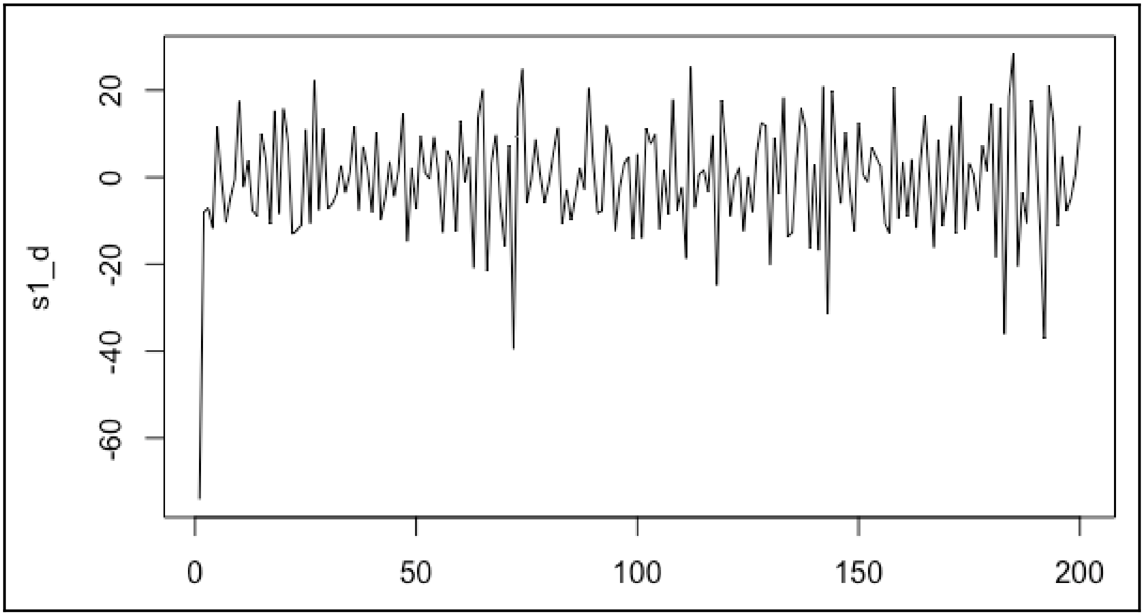

s1
40
20
0
-20
-40
0
50
100
150
200



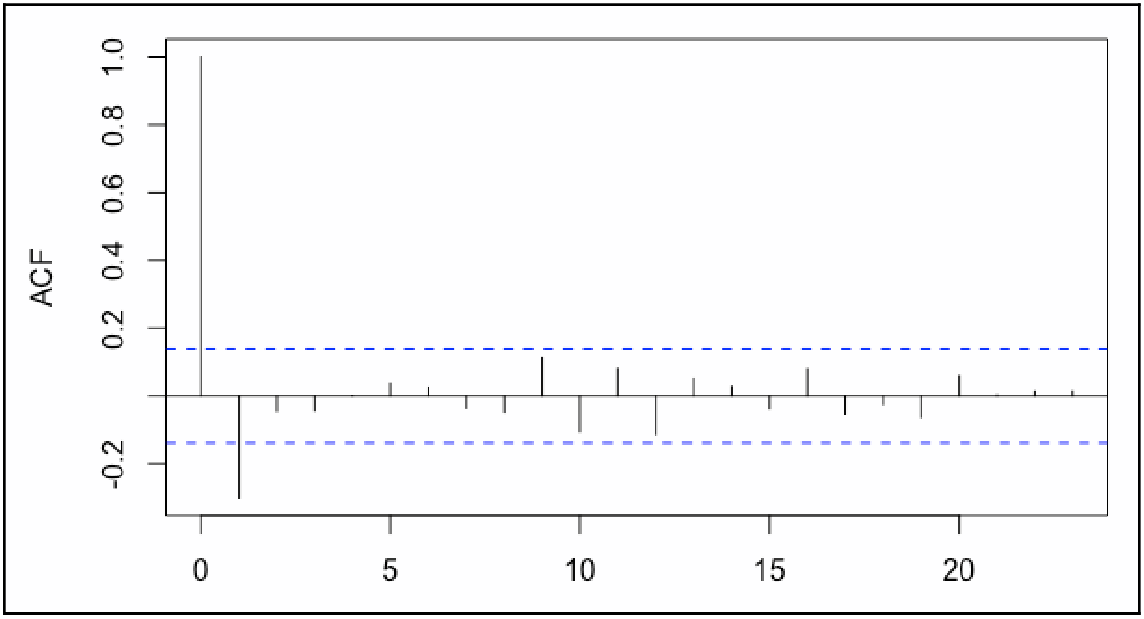

```
Call:
arima(x = s1, order = c(1, 0, 0), include.mean = FALSE)

Coefficients:
         ar1
      0.2986
s.e.  0.0814

sigma^2 estimated as 125.4:  log likelihood = -770.84,  aic = 1545.69
```



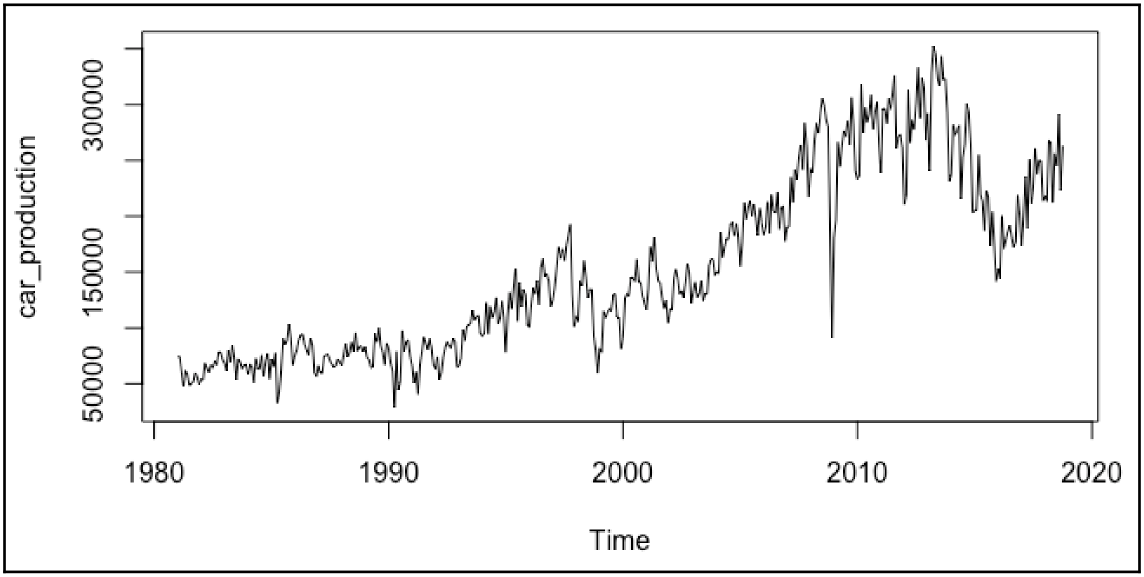

```
Call:
arima(x = ss1, order = c(1, 0, 0), include.mean = FALSE)

Coefficients:
         ar1
      0.9975
s.e.  0.0026

sigma^2 estimated as 104.6:  log likelihood = -755.24,  aic = 1514.48
```

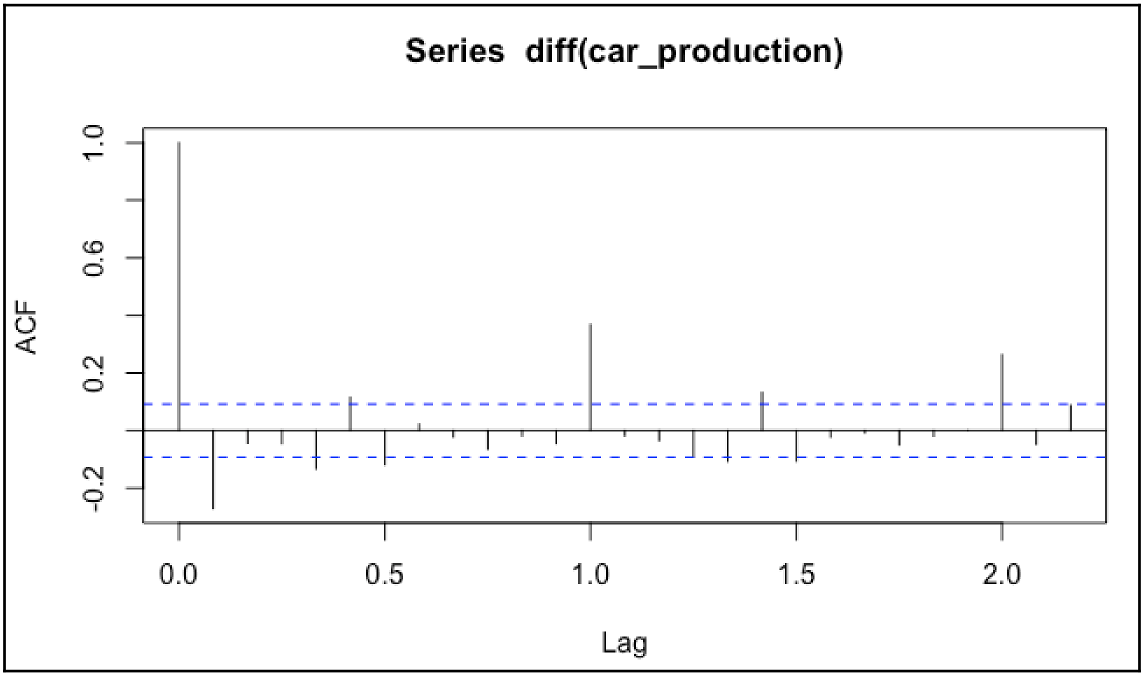


s1_d
20
0
-20
-40
-60
0
50
100
150
200

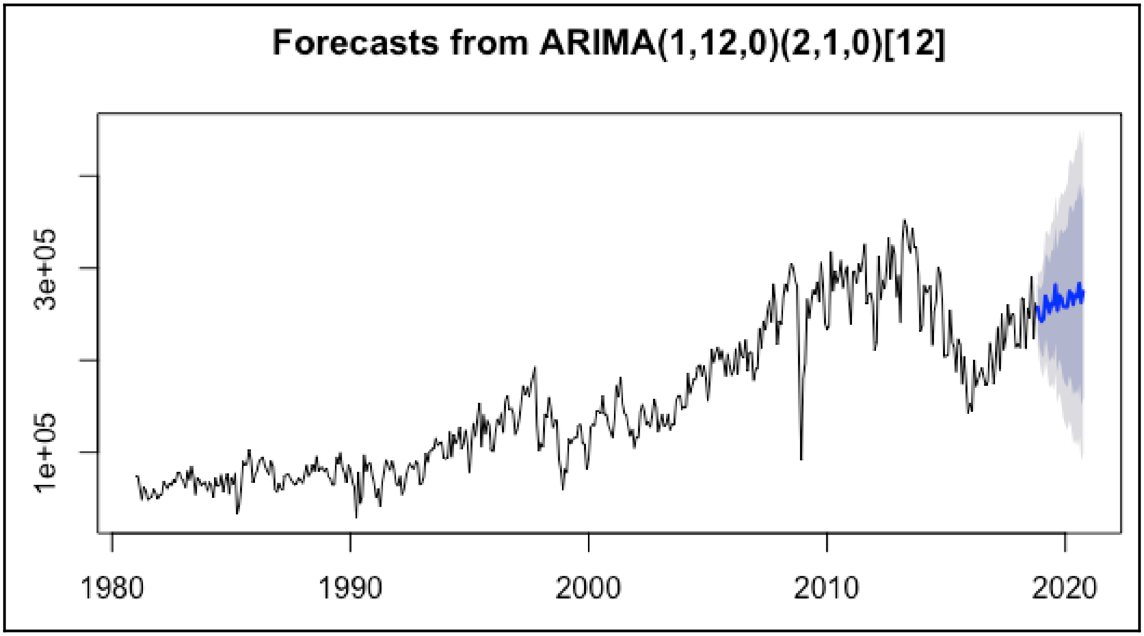


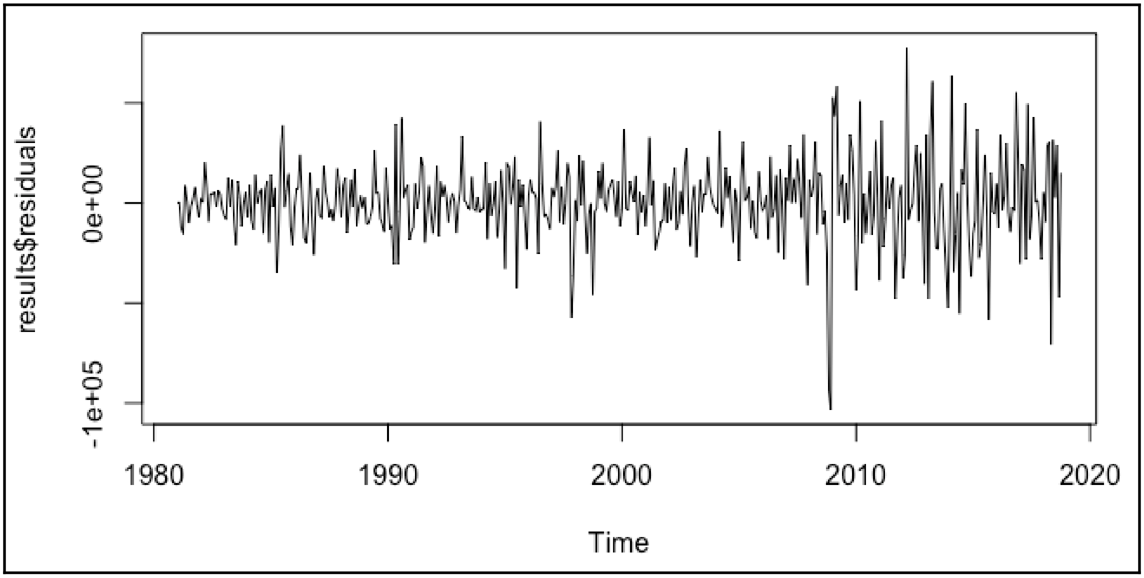
ACF
1.0
0.8
0.6
0.4
0.2
-0.2
0
5
10
15
20

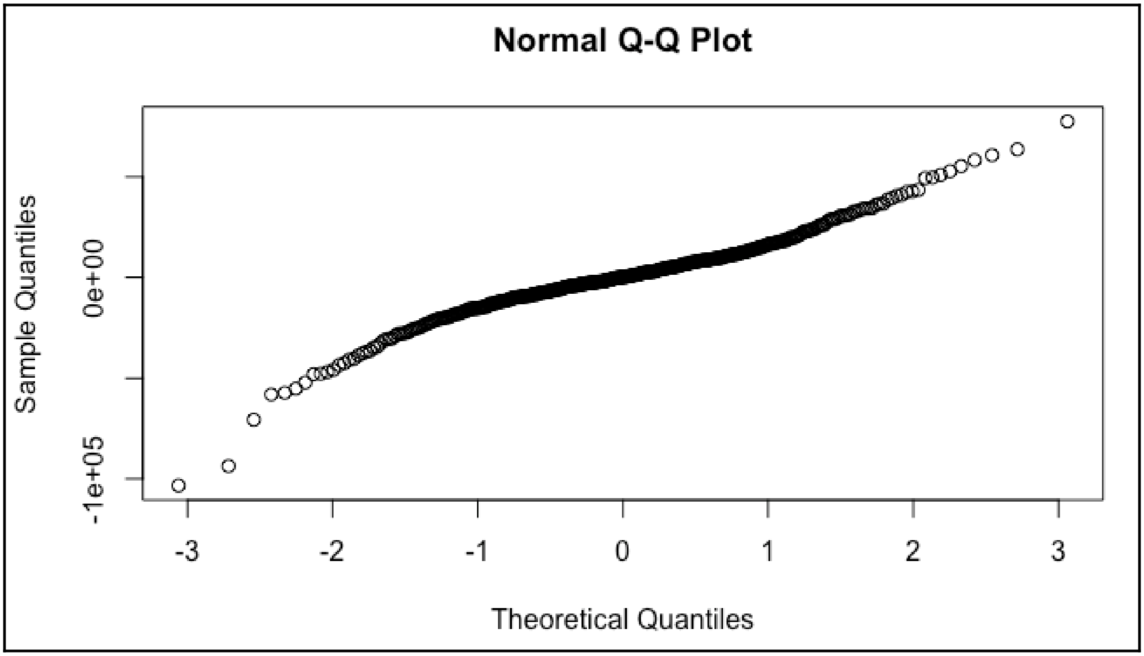

car_production
50000
150000
300000
1980
1990
2000
2010
2020
Time

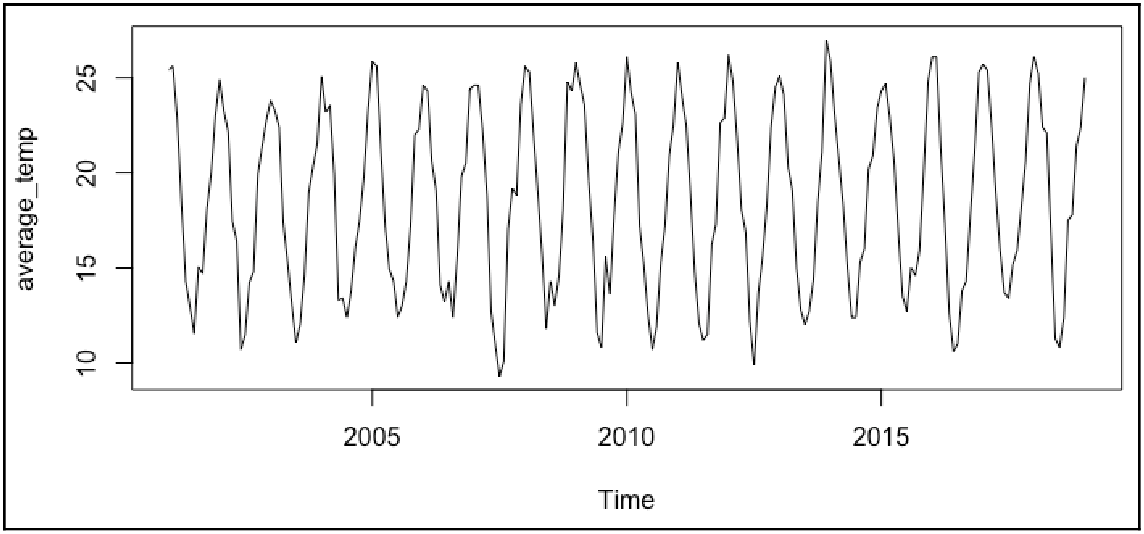

Series diff(car_production)
ACF
-0.2
0.2
0.6
1.0
0.0
0.5
1.0
1.5
2.0
Lag

```
Call:
arima(x = car_production, order = c(1, 1, 0), seasonal = c(2, 0, 0, 12))

Coefficients:
          ar1    sar1    sar2
      -0.3343  0.3648  0.1494
s.e.   0.0447  0.0477  0.0482

sigma^2 estimated as 408522768:  log likelihood = -5135.38,  aic = 10278.77
```

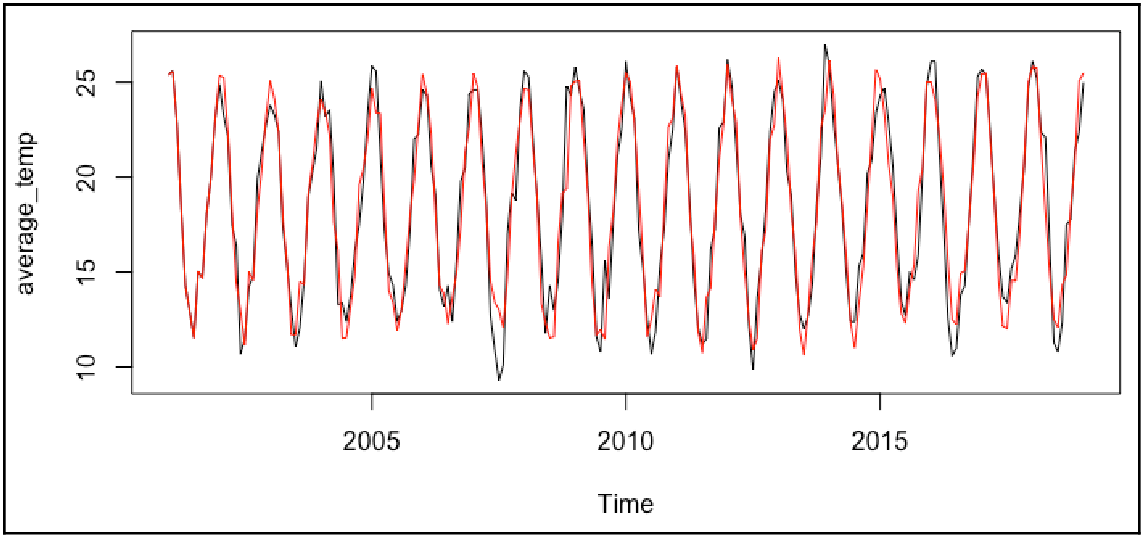


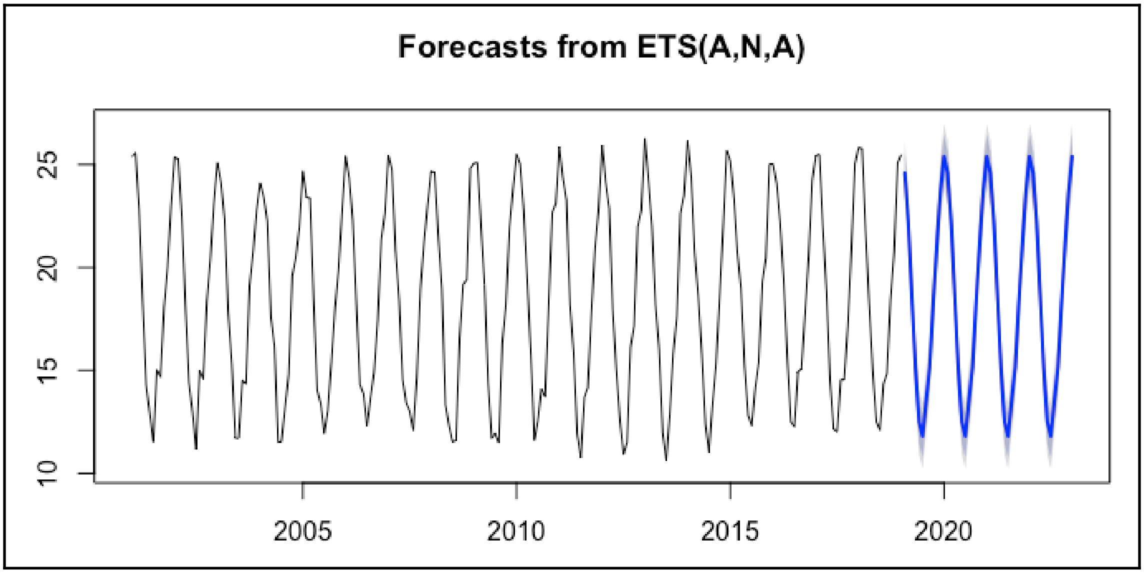

results$residuals
0e+00
-1e+05
1980
1990
2000
2010
2020
Time


Normal Q-Q Plot
Sample Quantiles
0e+00
-1e+05
-3
-2
-1
0
1
2
3
Theoretical Quantiles

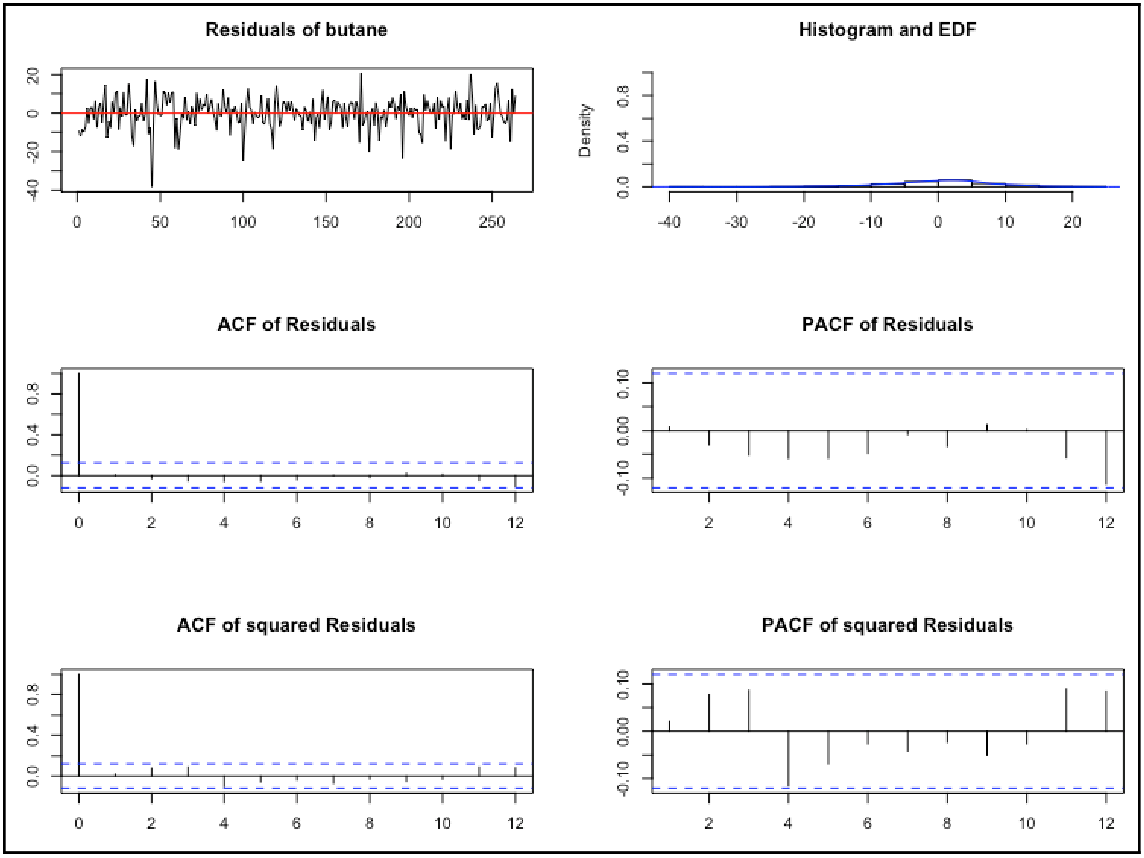



```
Series: average_temp
ARIMA(0,0,1)(1,1,0)[12]

Coefficients:
         ma1     sar1
      0.1569  -0.4626
s.e.  0.0711   0.0635

sigma^2 estimated as 2.126:  log likelihood=-368.64
AIC=743.27   AICc=743.39   BIC=753.24
```






```
> any(roots(m)>0.9999)
[1] FALSE
```


Residuals of butane
Histogram and EDF
Density
ACF of Residuals
PACF of Residuals
ACF of squared Residuals
PACF of squared Residuals

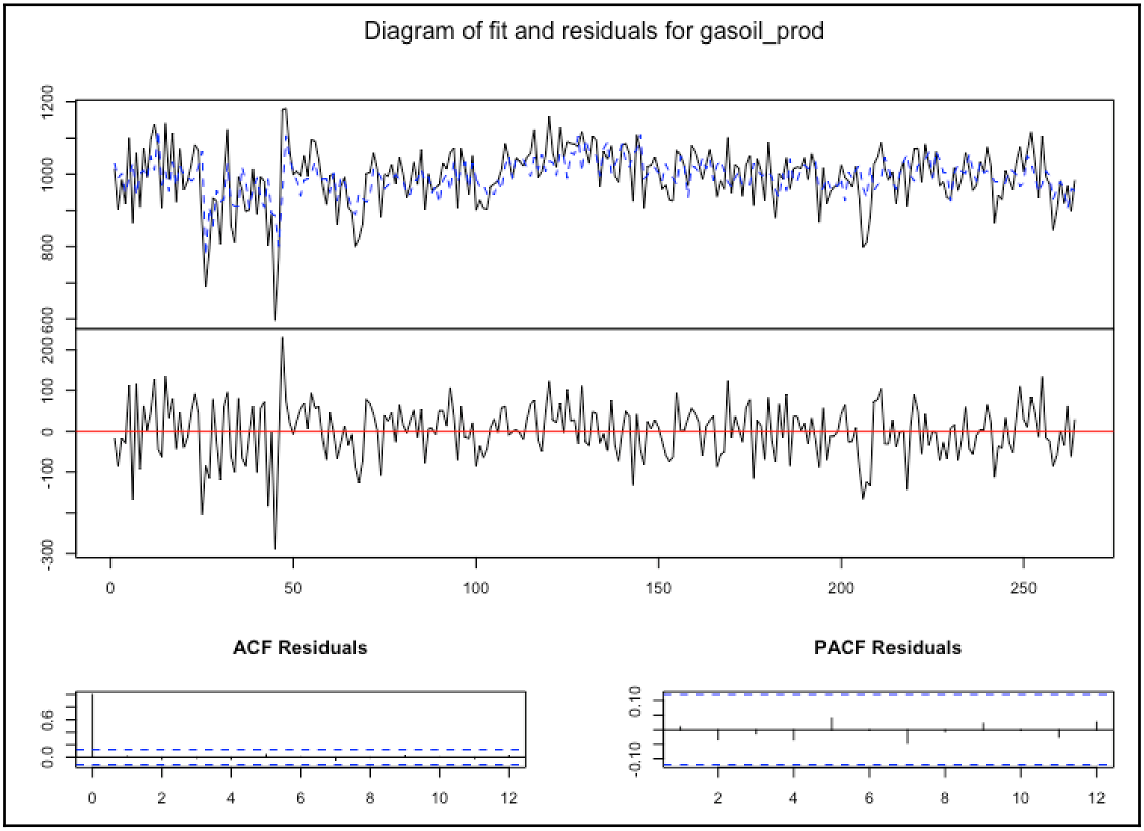
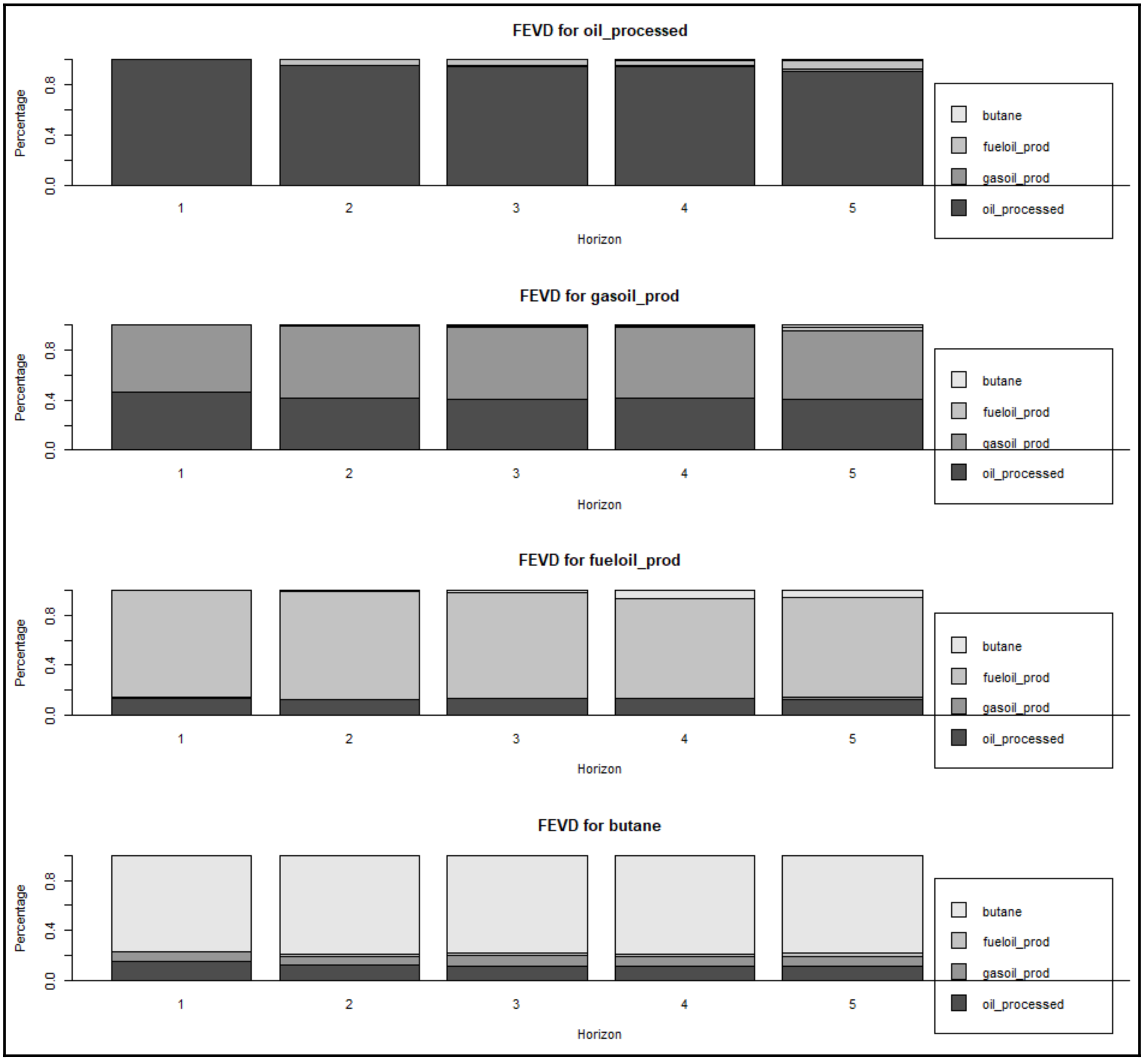
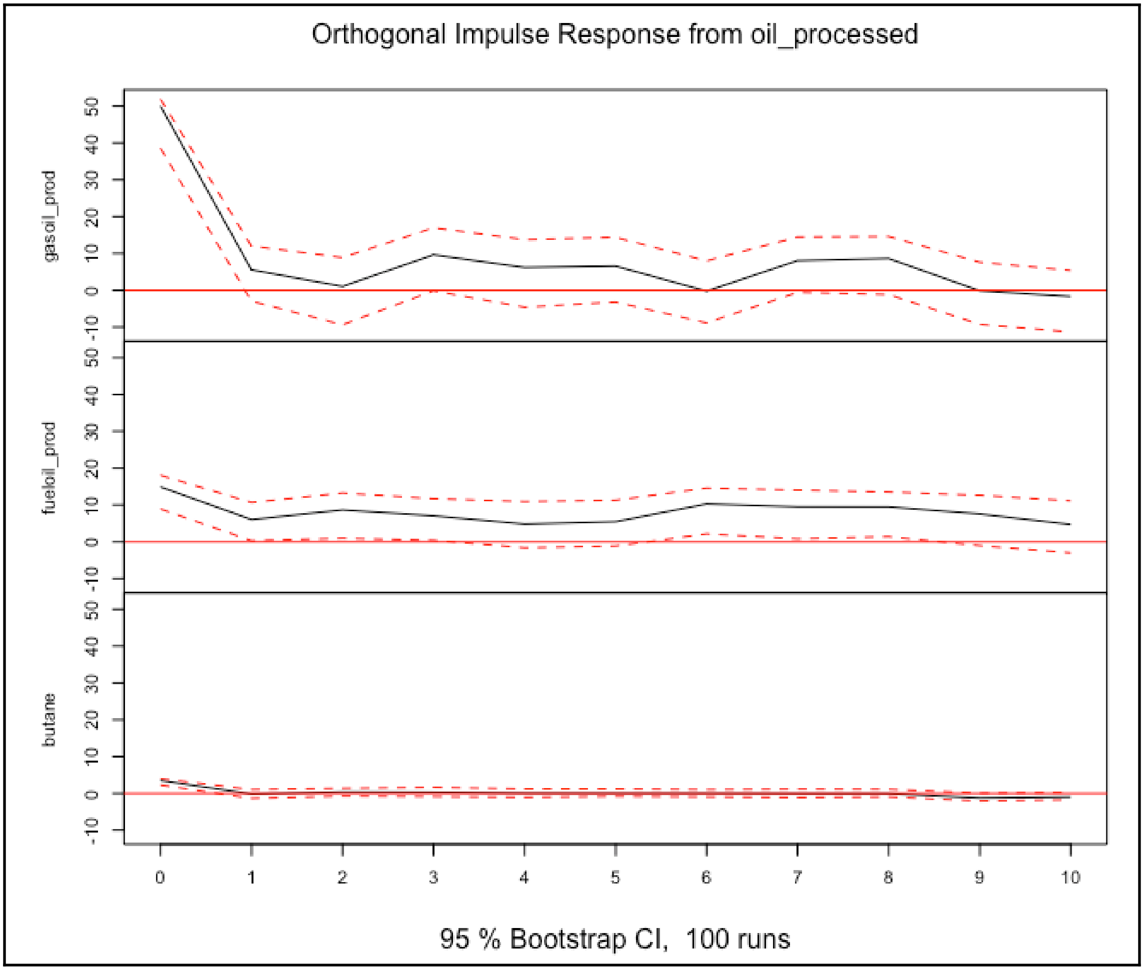
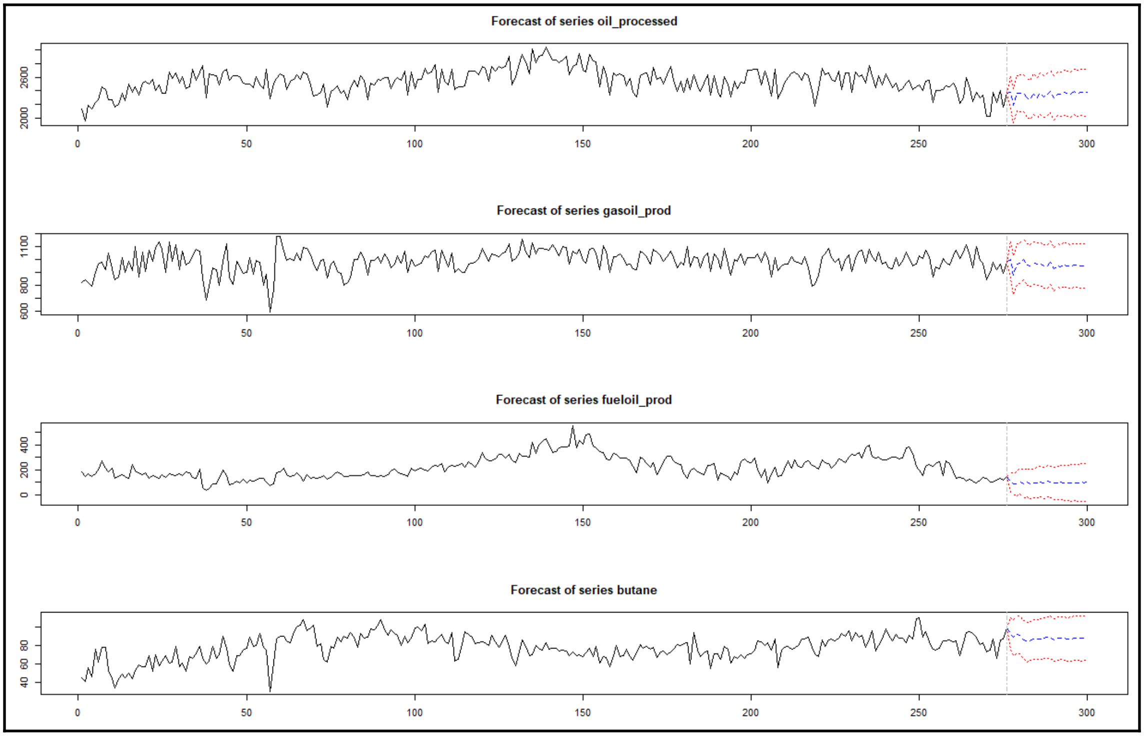


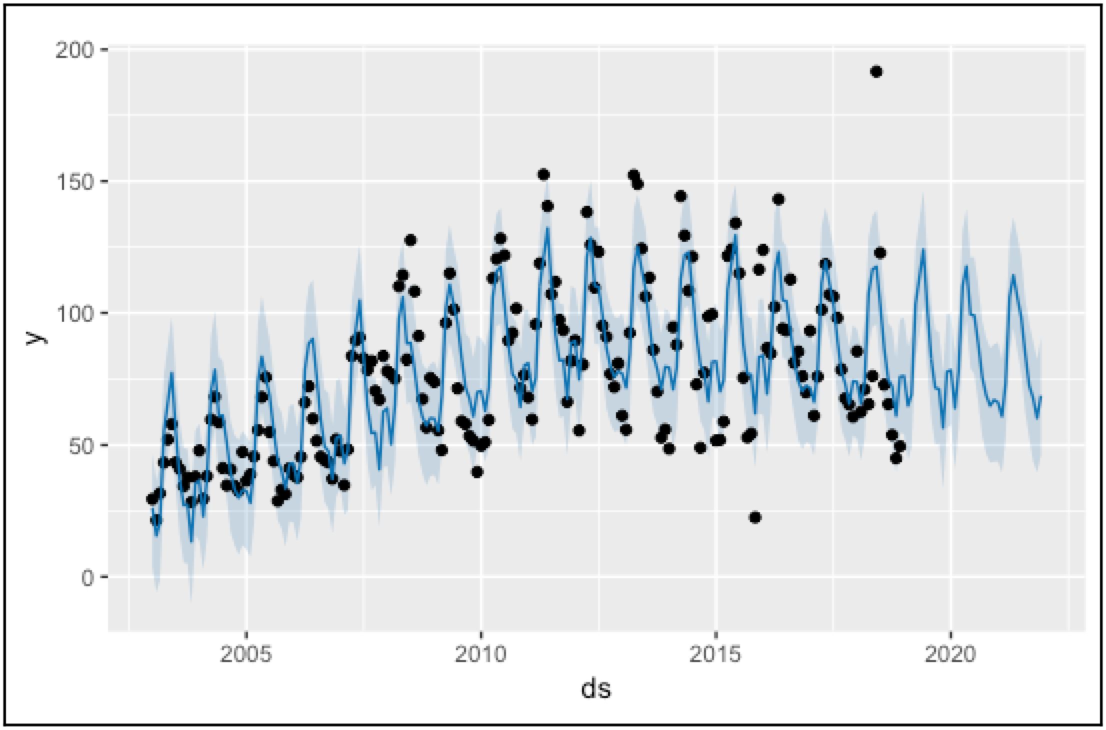
Diagram of fit and residuals for gasoil_prod
1200
1000
800
600
200
100
0
-100
-300
0
50
100
150
200
250
ACF Residuals
0.6
0.0
0
2
4
6
8
10
12
PACF Residuals
0.10
-0.10
2
4
6
8
10
12

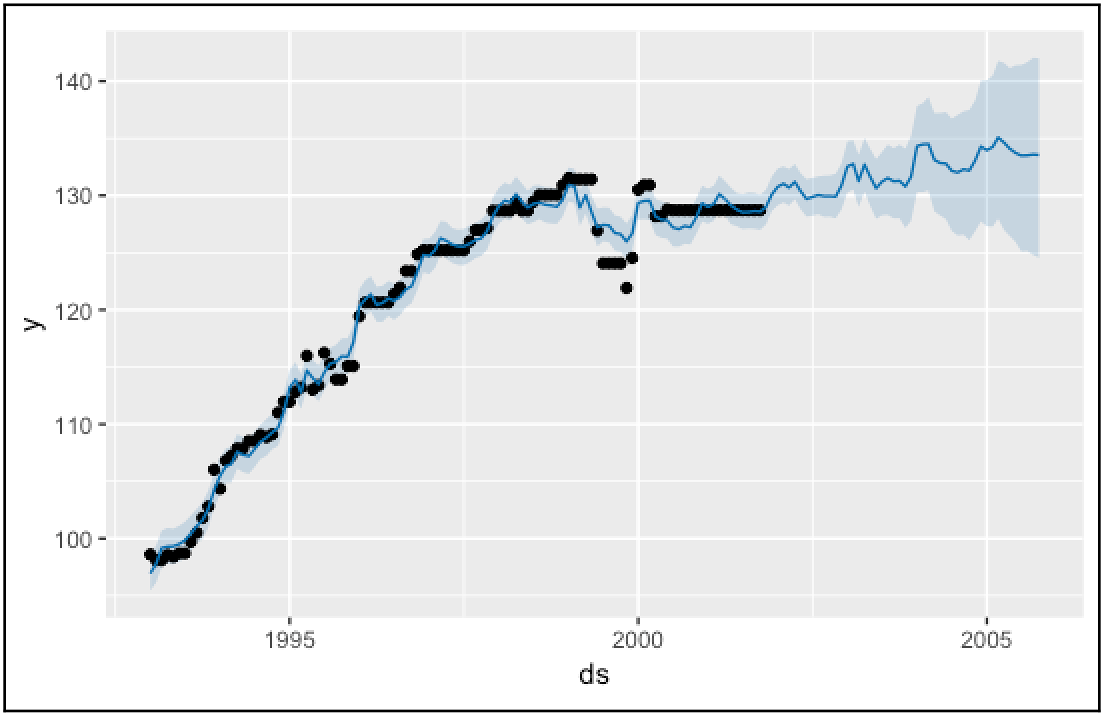

FEVD for oil_processed
Percentage
0.0
0.4
0.8
1
2
3
4
5
Horizon
butane
fueloil_prod
gasoil_prod
oil_processed
FEVD for gasoil_prod
Percentage
0.0
0.4
0.8
1
2
3
4
5
Horizon
butane
fueloil_prod
gasoil_prod
oil_processed
FEVD for fueloil_prod
Percentage
0.0
0.4
0.8
1
2
3
4
5
Horizon
butane
fueloil_prod
gasoil_prod
oil_processed
FEVD for butane
Percentage
0.0
0.4
0.8
1
2
3
4
5
Horizon
butane
fueloil_prod
gasoil_prod
oil_processed

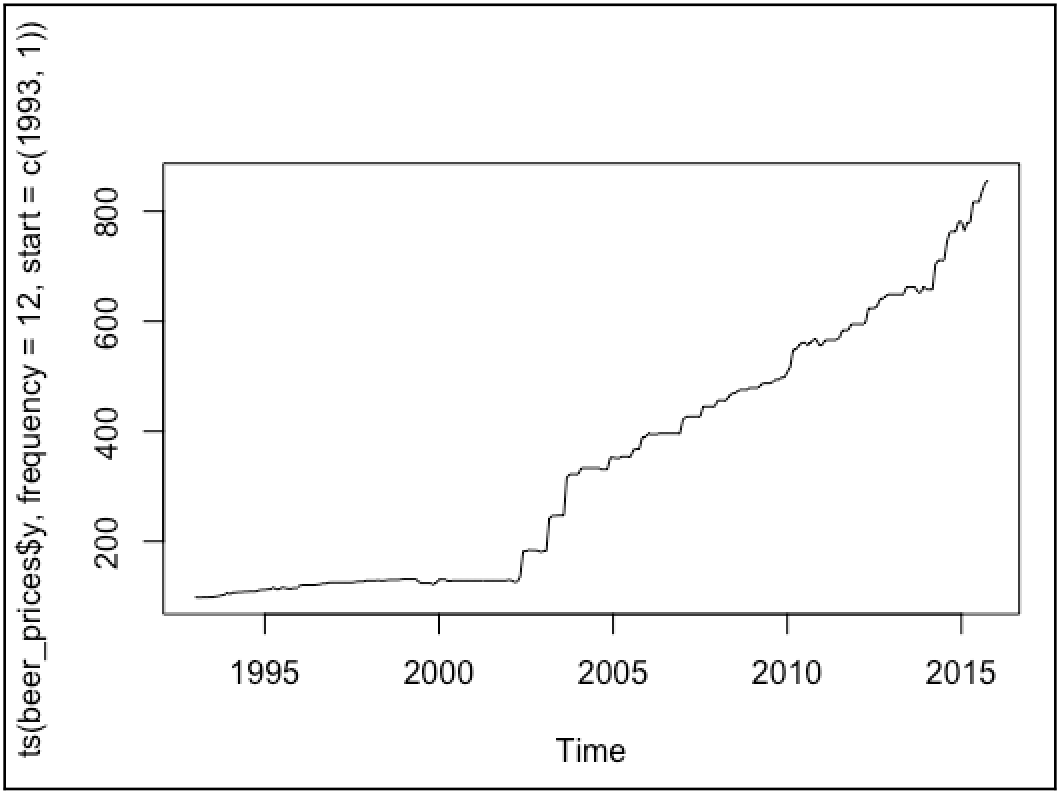


Orthogonal Impulse Response from oil_processed
gasoil_prod
fueloil_prod
butane
-10
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8
9
10
95 % Bootstrap CI, 100 runs

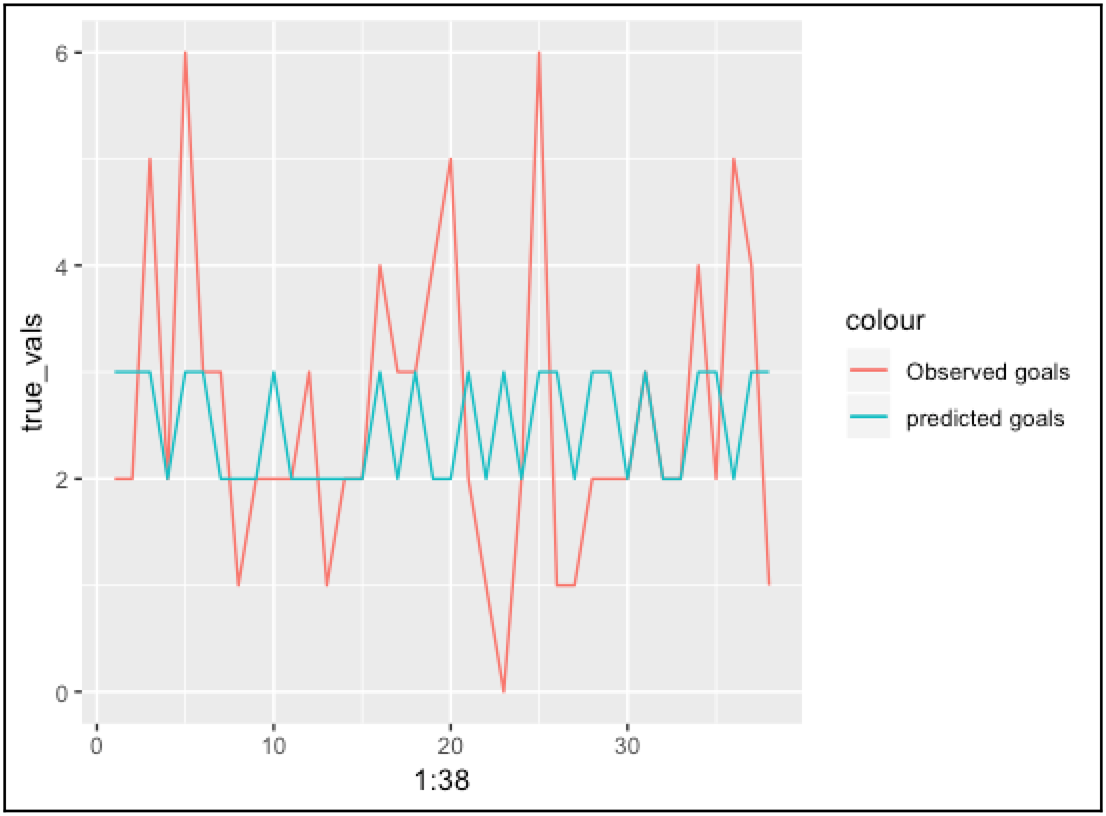

Forecast of series oil_processed
2000
2600
0
50
100
150
200
250
300
Forecast of series gasoil_prod
600
800
1100
0
50
100
150
200
250
300
Forecast of series fueloil_prod
0
200
400
0
50
100
150
200
250
300
Forecast of series butane
40
60
80
0
50
100
150
200
250
300


200
150
100
50
0
y
2005
2010
2015
2020
ds

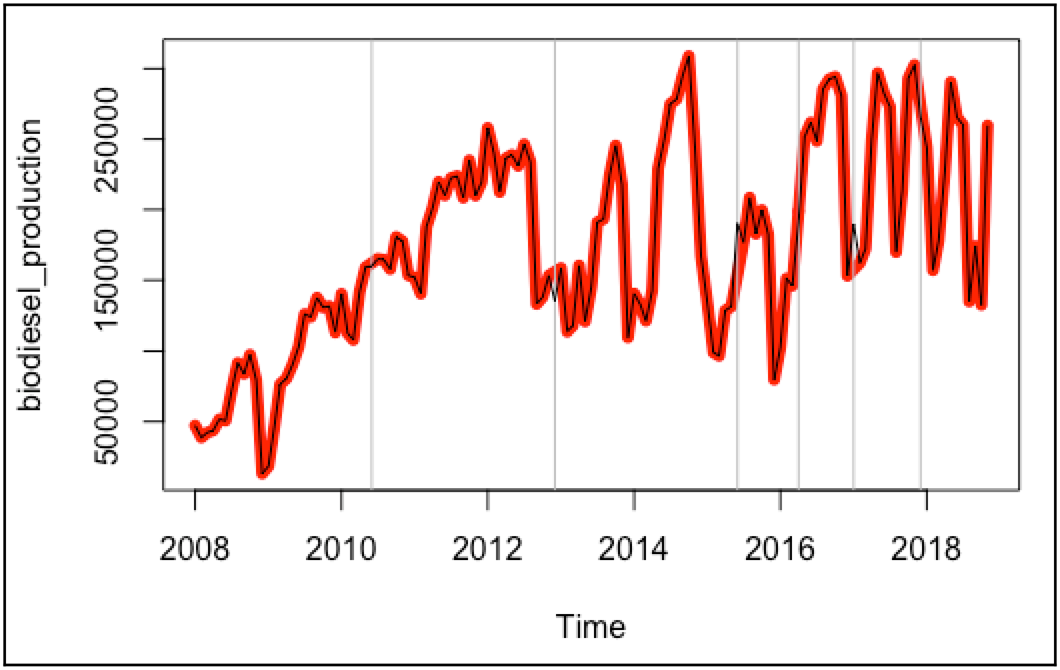

140
130
120
110
100
y
1995
2000
2005
ds


ts(beer_prices$y, frequency = 12, start = c(1993, 1))
200
400
600
800
1995
2000
2005
2010
2015
Time

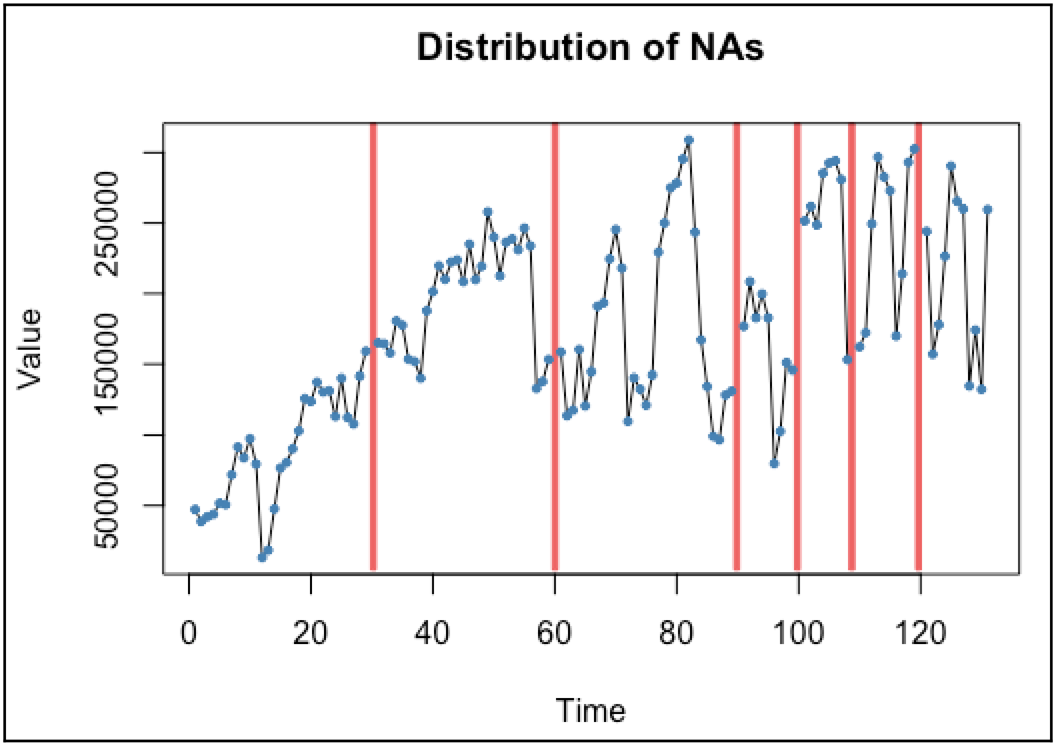
```
Call:
tsglm(ts = data$Goals, model = list(past_obs = c(1:3)), xreg = cbind(data$home_away,
    data$pos, data$diff_days, data$champions_next_days_after,
    data$champions_next_days_before), link = "log", distr = "poisson")

Coefficients:
             Estimate  Std.Error  CI(lower)  CI(upper)
(Intercept)   0.74618     0.5976    -0.4251     1.9175
beta_1        0.24035     0.2854    -0.3190     0.7997
beta_2       -0.27568     0.2969    -0.8576     0.3063
beta_3       -0.03185     0.2717    -0.5644     0.5007
eta_1         0.26207     0.2425    -0.2133     0.7374
eta_2         0.01427     0.0195    -0.0239     0.0524
eta_3         0.00574     0.0330    -0.0589     0.0704
eta_4        -0.02630     0.2783    -0.5718     0.5192
eta_5        -0.18015     0.2641    -0.6978     0.3375
Standard errors and confidence intervals (level =  95 %) obtained
by normal approximation.

Link function: log
Distribution family: poisson
Number of coefficients: 9
Log-likelihood: -64.84624
AIC: 147.6925
BIC: 162.4308
QIC: 147.6925
```

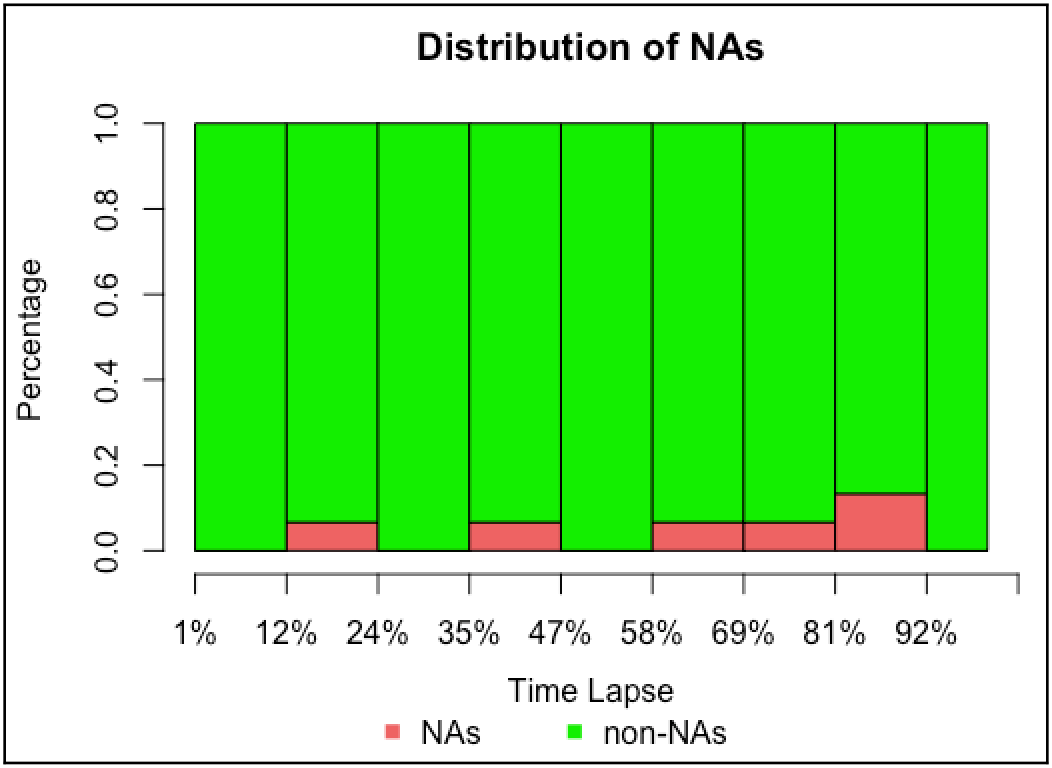


true_vals
6
4
2
0
0
10
20
30
1:38
colour
Observed goals
predicted goals

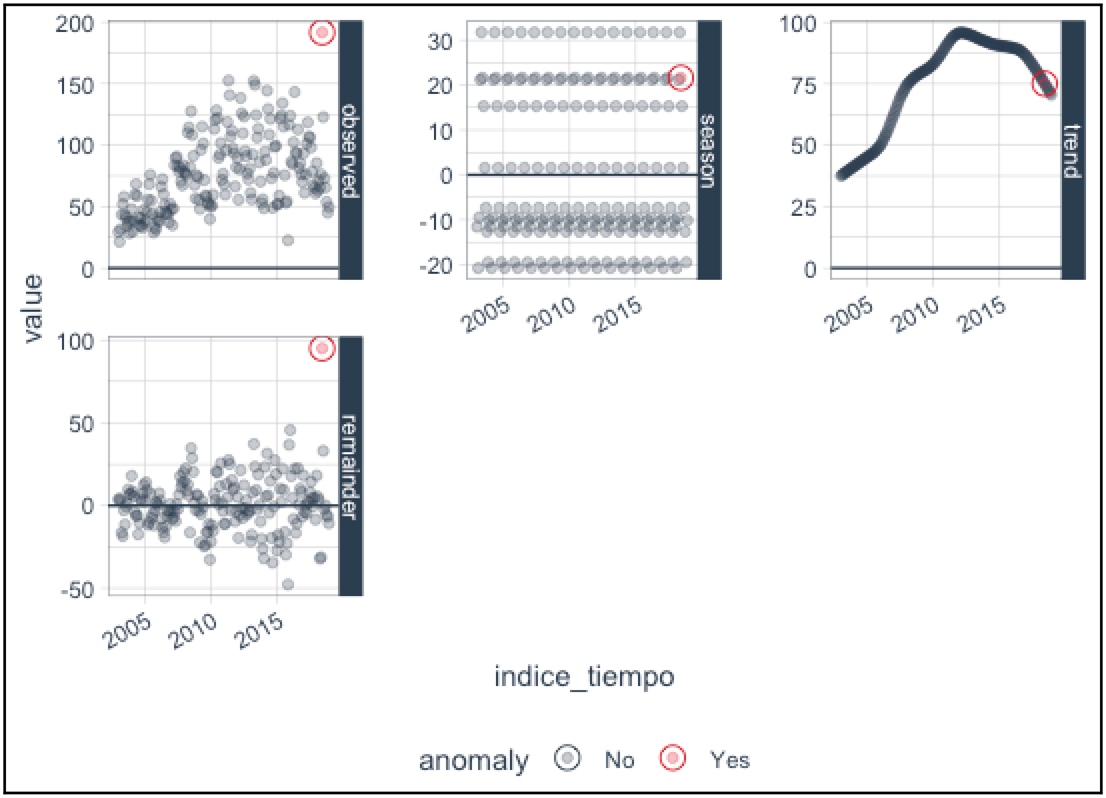


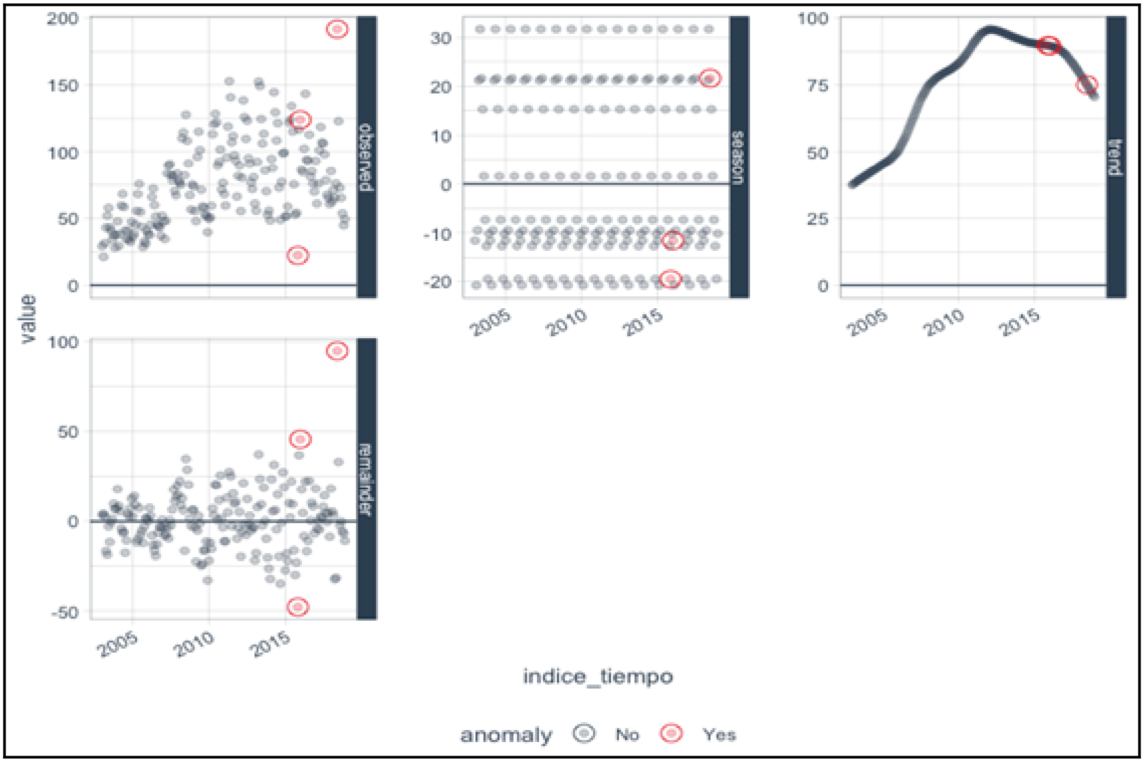
biodiesel_production
50000
150000
250000
2008
2010
2012
2014
2016
2018
Time


Distribution of NAs
Value
50000
150000
250000
0
20
40
60
80
100
120
Time

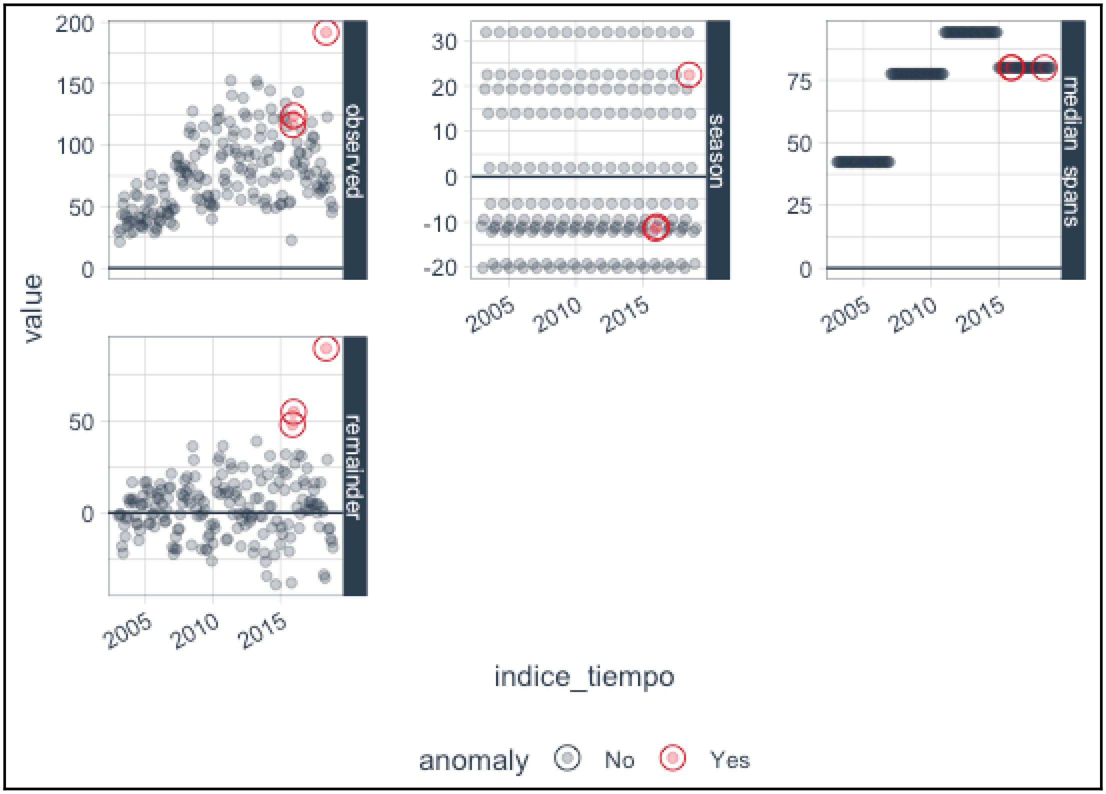

Distribution of NAs
Percentage
0.0
0.2
0.4
0.6
0.8
1.0
1%
12%
24%
35%
47%
58%
69%
81%
92%
Time Lapse
NAs
non-NAs

```
[1] "Length of time series:"
[1] 131
[1] "-------------------------"
[1] "Number of Missing Values:"
[1] 6
[1] "-------------------------"
[1] "Percentage of Missing Values:"
[1] "4.58%"
[1] "-------------------------"
[1] "Stats for Bins"
[1] "  Bin 1 (33 values from 1 to 33) :      1 NAs (3.03%)"
[1] "  Bin 2 (33 values from 34 to 66) :      1 NAs (3.03%)"
[1] "  Bin 3 (33 values from 67 to 99) :      1 NAs (3.03%)"
[1] "  Bin 4 (32 values from 100 to 131) :      3 NAs (9.38%)"
[1] "-------------------------"
[1] "Longest NA gap (series of consecutive NAs)"
[1] "1 in a row"
[1] "-------------------------"
[1] "Most frequent gap size (series of consecutive NA series)"
[1] "1 NA in a row (occuring 6 times)"
[1] "-------------------------"
[1] "Gap size accounting for most NAs"
[1] "1 NA in a row (occuring 6 times, making up for overall 6 NAs)"
[1] "-------------------------"
[1] "Overview NA series"
[1] "  1 NA in a row: 6 times"
```

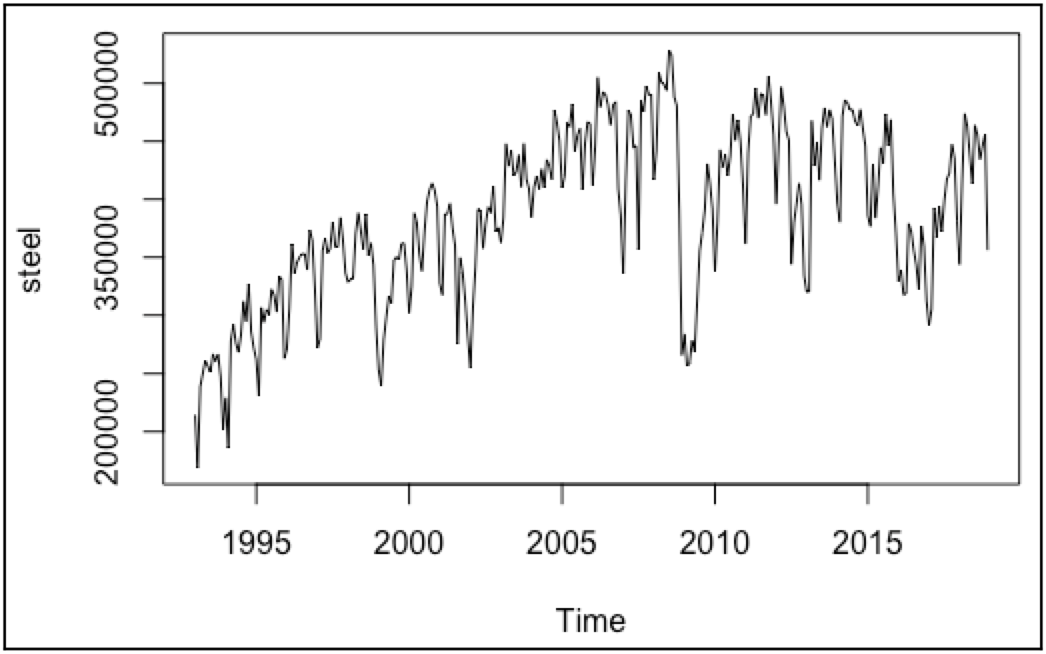

value
observed
season
trend
remainder
indice_tiempo
anomaly
No
Yes
2005
2010
2015

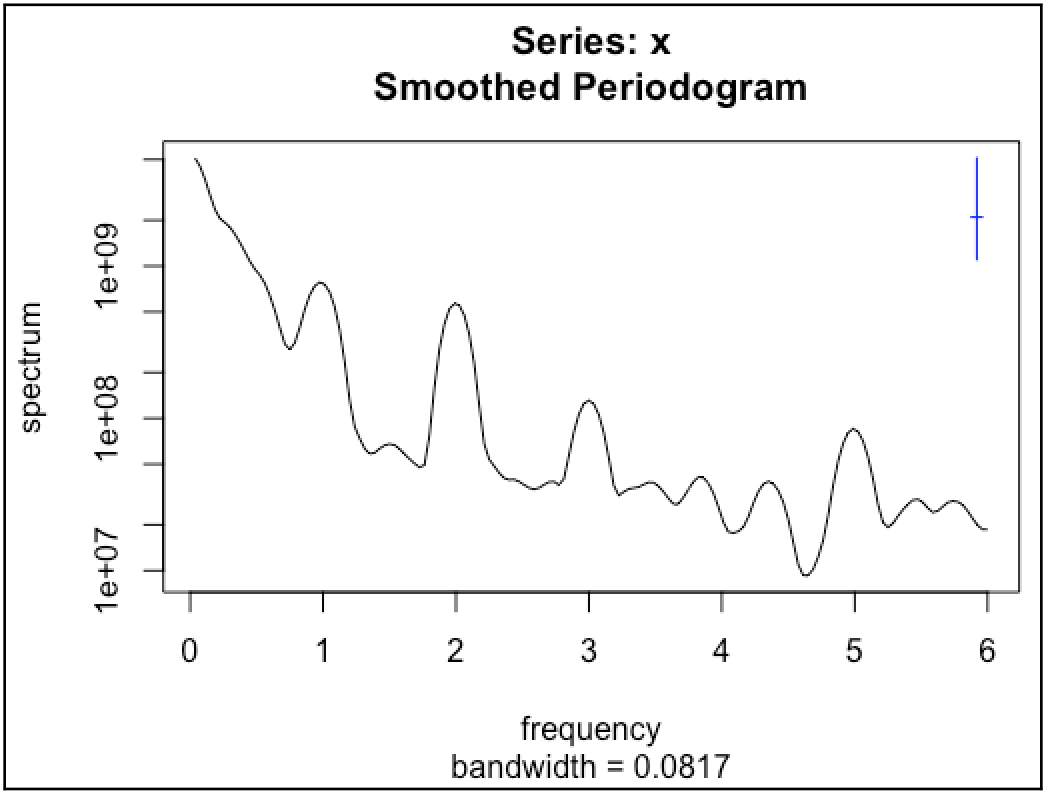


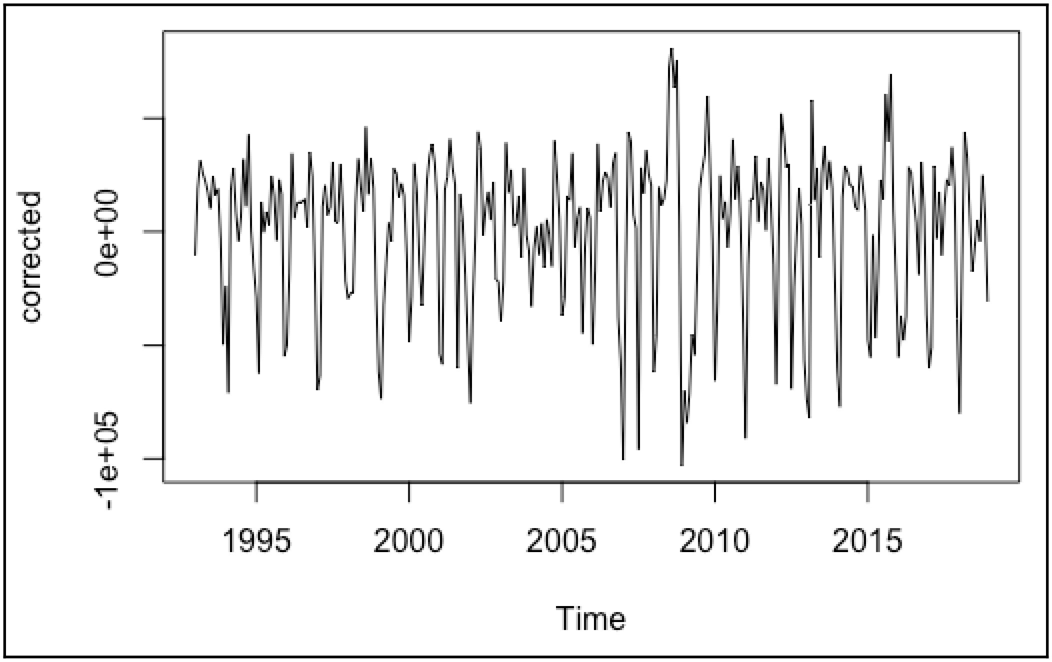
observed
season
trend
remainder
value
indice_tiempo
anomaly
No
Yes

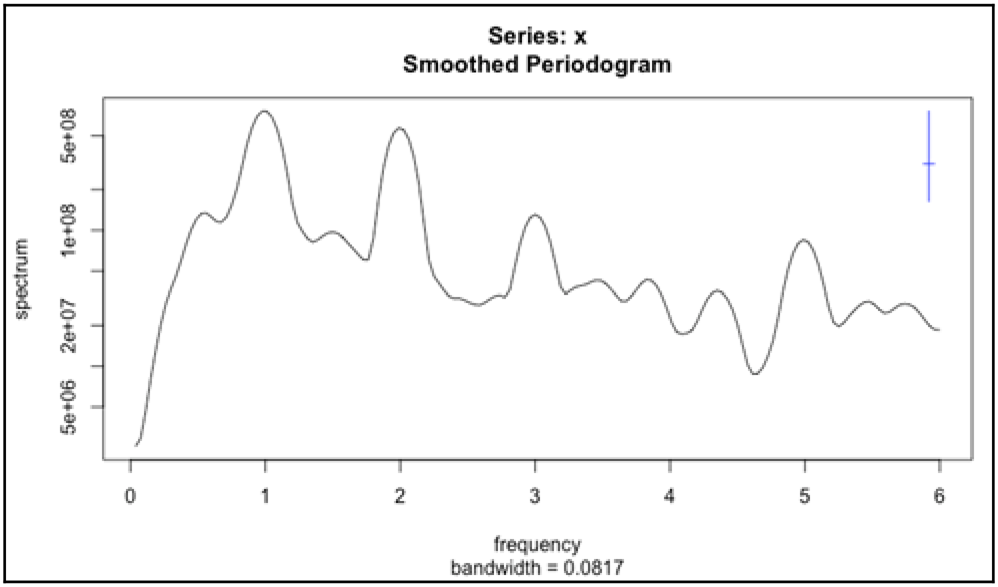


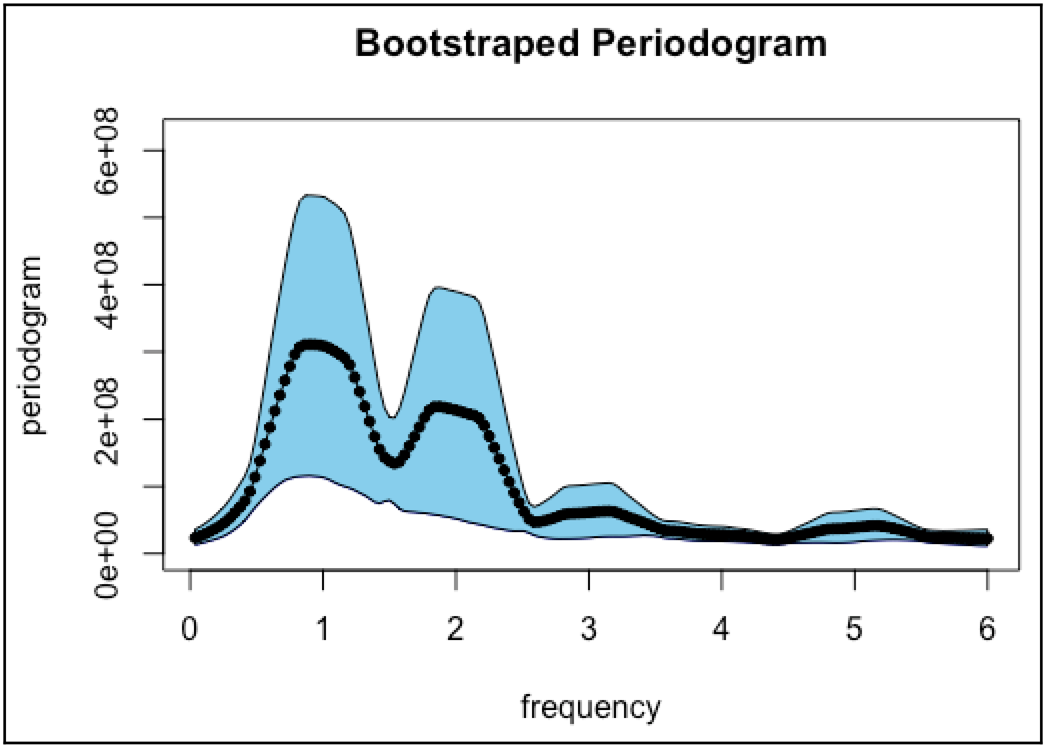
value
observed
season
median_spans
remainder
200
150
100
50
0
30
20
10
0
-10
-20
75
50
25
0
50
0
2005
2010
2015
indice_tiempo
anomaly
No
Yes


steel
200000
350000
500000
1995
2000
2005
2010
2015
Time


Series: x
Smoothed Periodogram
spectrum
1e+07
1e+08
1e+09
0
1
2
3
4
5
6
frequency
bandwidth = 0.0817


corrected
0e+00
-1e+05
1995
2000
2005
2010
2015
Time


Series: x
Smoothed Periodogram
spectrum
5e+08
1e+08
2e+07
5e+06
0
1
2
3
4
5
6
frequency
bandwidth = 0.0817



Bootstraped Periodogram
periodogram
0e+00
2e+08
4e+08
6e+08
0
1
2
3
4
5
6
frequency

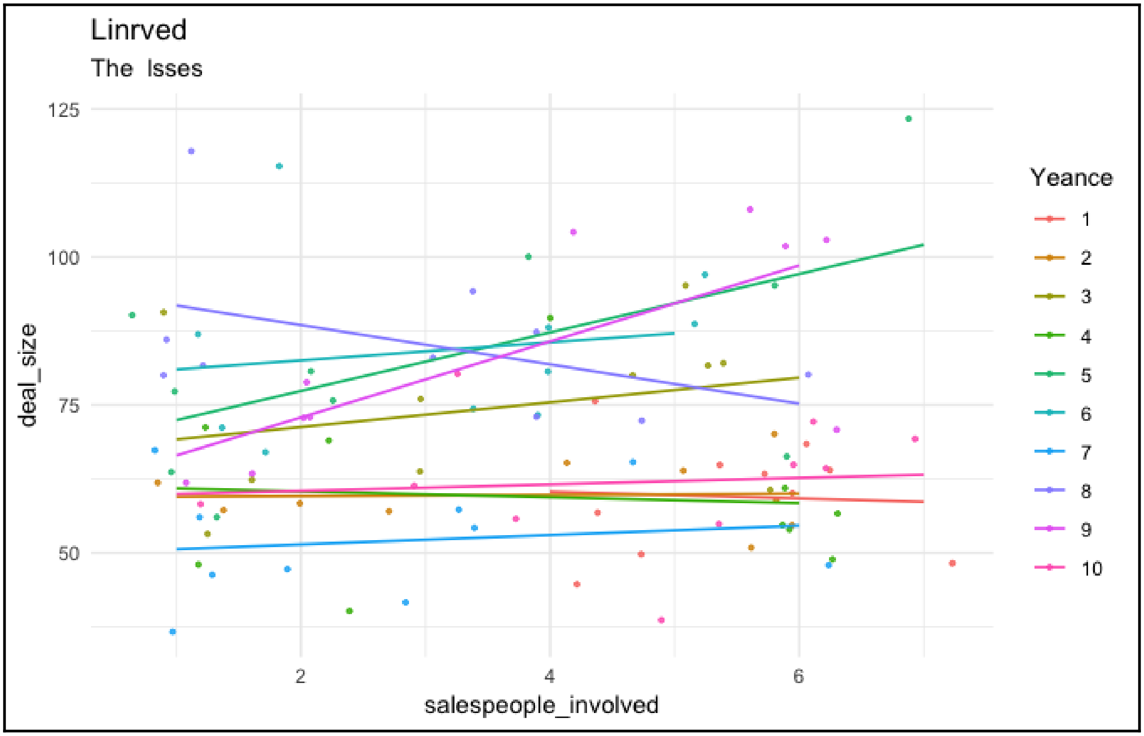
Chapter 8: Mixed Effects Models

```
Linear mixed model fit by REML ['lmerMod']
Formula: deal_size ~ salespeople_involved + time_spent_deal + (-1 + salespeople_involved |
    clientid) + (-1 + time_spent_deal | clientid)
   Data: data
REML criterion at convergence: 774.8291
Random effects:
 Groups     Name                 Std.Dev.
 clientid   salespeople_involved  2.297
 clientid.1 time_spent_deal       1.266
 Residual                        10.244
Number of obs: 100, groups:  clientid, 10
Fixed Effects:
         (Intercept)  salespeople_involved       time_spent_deal
              48.833                 1.956                 2.759
```

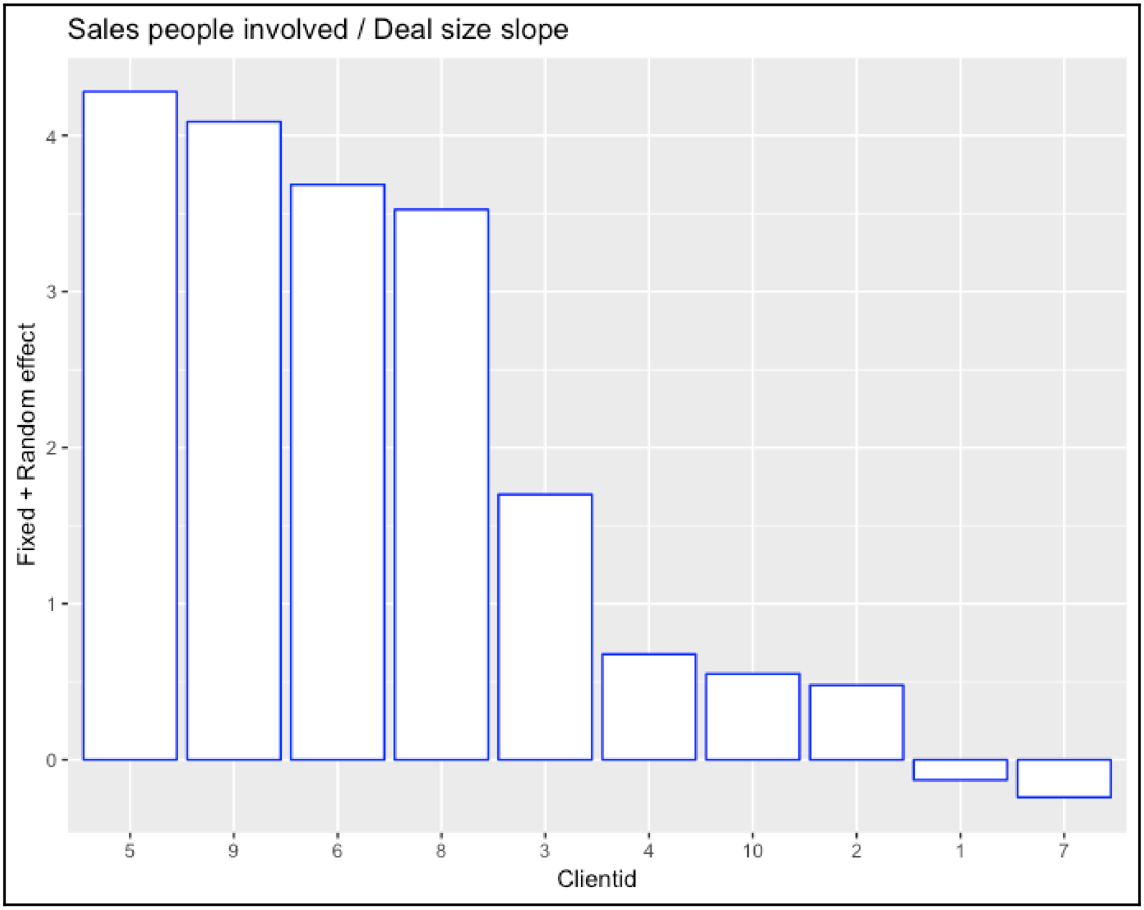
```
Linear mixed model fit by REML ['lmerMod']
Formula: deal_size ~ salespeople_involved + time_spent_deal + (1 + salespeople_involved |      clientid)
   Data: data
REML criterion at convergence: 773.9846
Random effects:
 Groups   Name                 Std.Dev. Corr
 clientid (Intercept)           7.816
          salespeople_involved  1.980   0.49
 Residual                      10.350
Number of obs: 100, groups:  clientid, 10
Fixed Effects:
         (Intercept)  salespeople_involved       time_spent_deal
              50.295                 1.627                 2.739
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: deal_size ~ salespeople_involved + time_spent_deal + (1 + time_spent_deal +      salespeople_involved | clientid)
   Data: data
REML criterion at convergence: 768.192
Random effects:
 Groups   Name                 Std.Dev. Corr
 clientid (Intercept)          6.624
          time_spent_deal      1.046    -0.24
          salespeople_involved 2.057    -0.31  1.00
 Residual                      9.959
Number of obs: 100, groups:  clientid, 10
Fixed Effects:
         (Intercept)  salespeople_involved       time_spent_deal
               49.50                  1.86                  2.71
convergence code 0; 1 optimizer warnings; 0 lme4 warnings
```

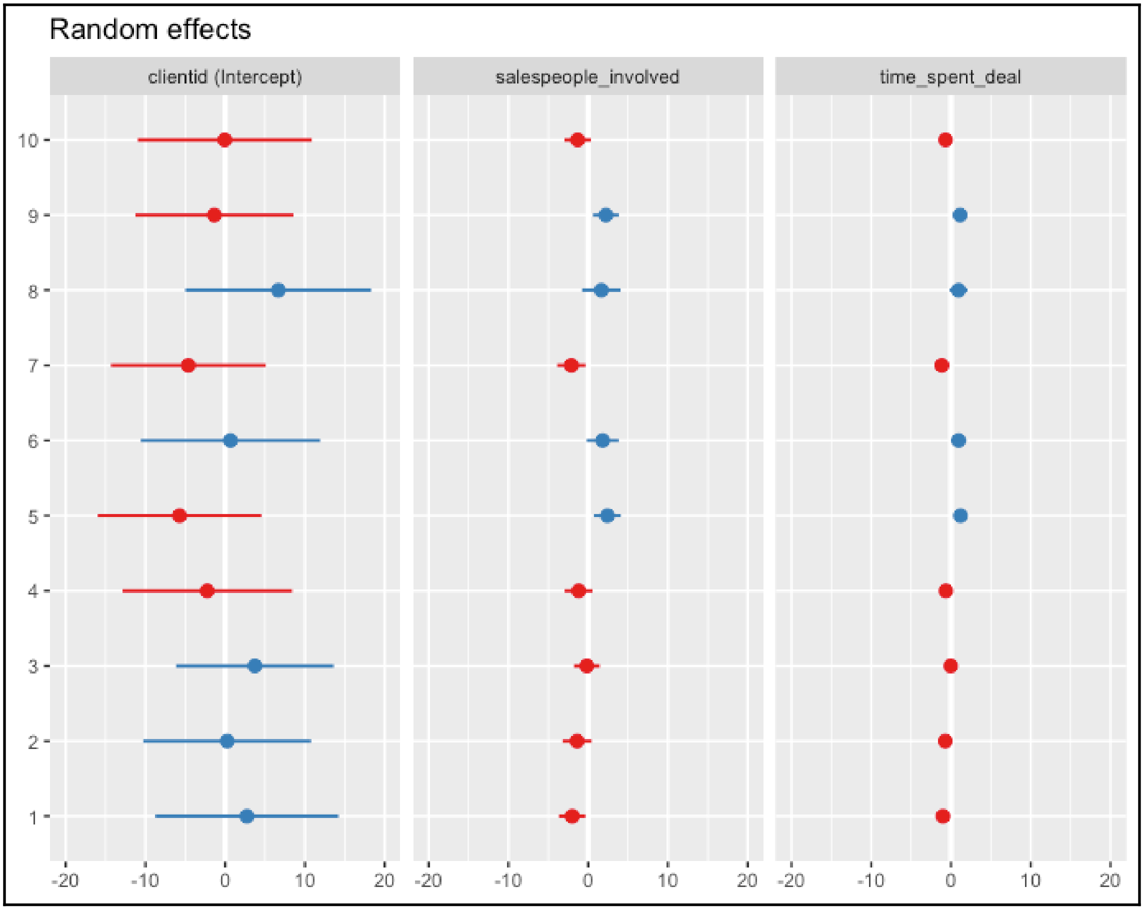
```
$clientid
   salespeople_involved time_spent_deal
1            -1.2168381      -1.2586696
2            -0.9524152      -1.0154345
3            -0.5808371       0.7857888
4            -1.9134247      -0.4041685
5             2.2273657       0.4613553
6             1.5211412       1.1708320
7            -2.8446889      -1.3699988
8             2.6419682       1.5660341
9             2.5395813       0.6396795
10           -1.4218524      -0.5754181

with conditional variances for "clientid"
```



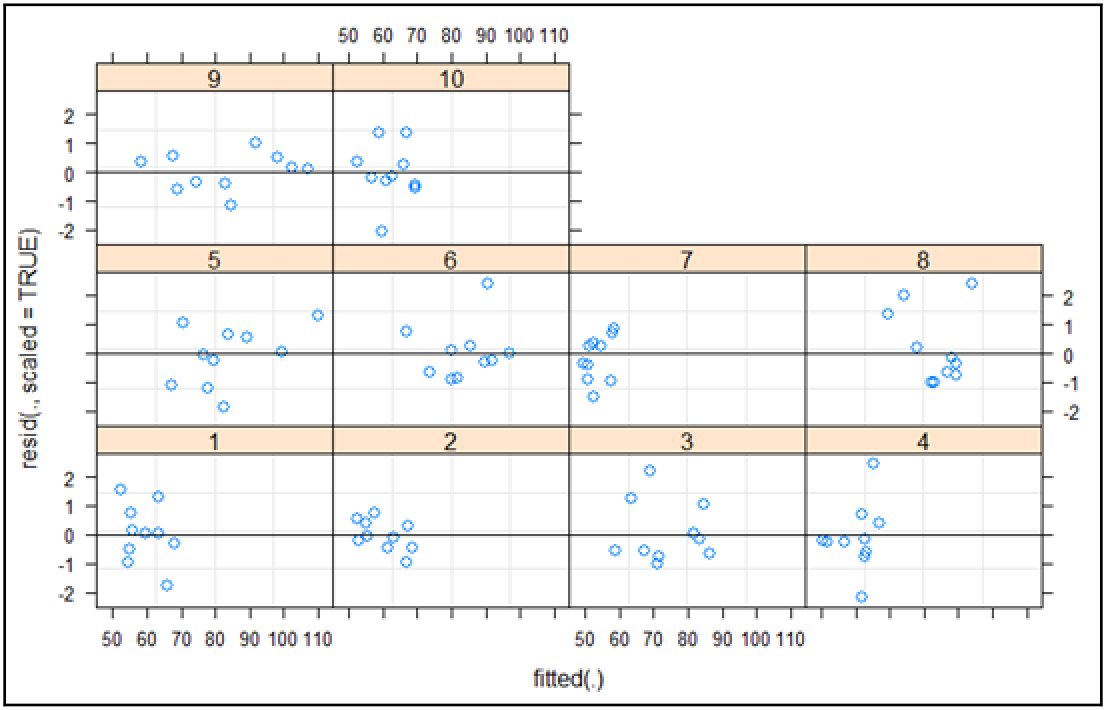
	(Intercept)	salespeople_involved	time_spent_deal	clientid
1	52.20568	-0.1303600	1.708985	1
2	49.72813	0.4777053	1.995823	2
3	53.21839	1.6994957	2.669033	3
4	47.22548	0.6760296	2.069997	4
5	43.80163	4.2804007	3.899074	5
6	50.17681	3.6832969	3.662500	6
7	44.84697	-0.2417718	1.567070	7
8	56.18871	3.5225308	3.647934	8
9	48.13624	4.0871472	3.848485	9
10	49.42407	0.5496581	2.029642	10

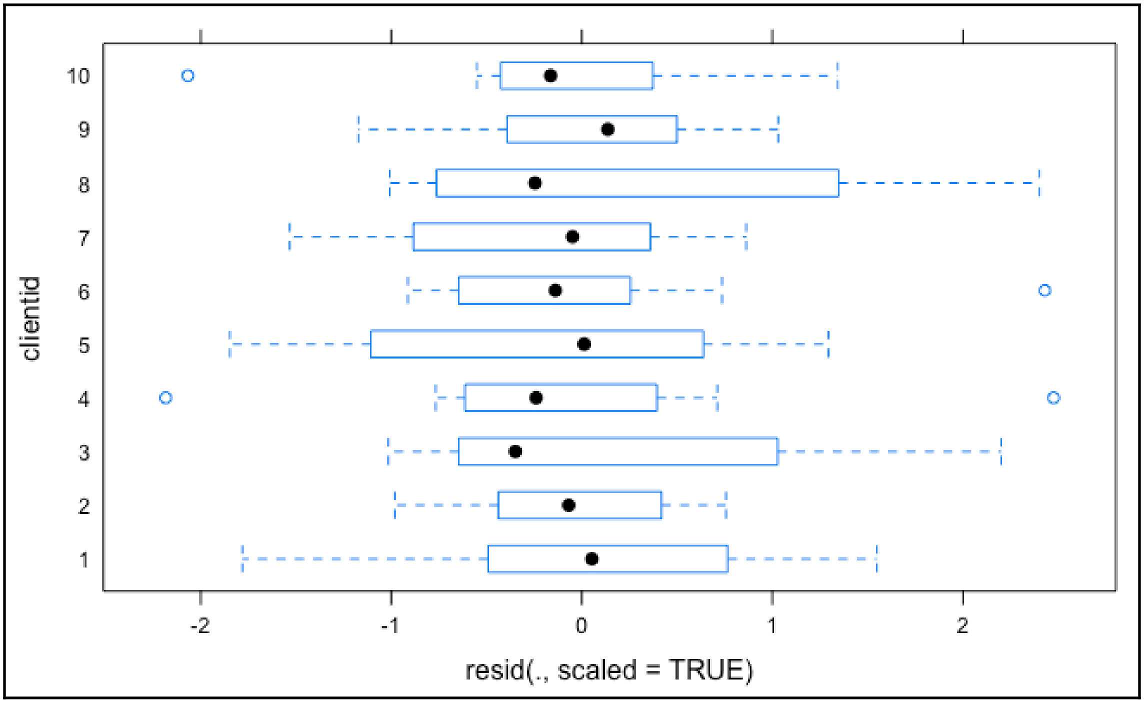


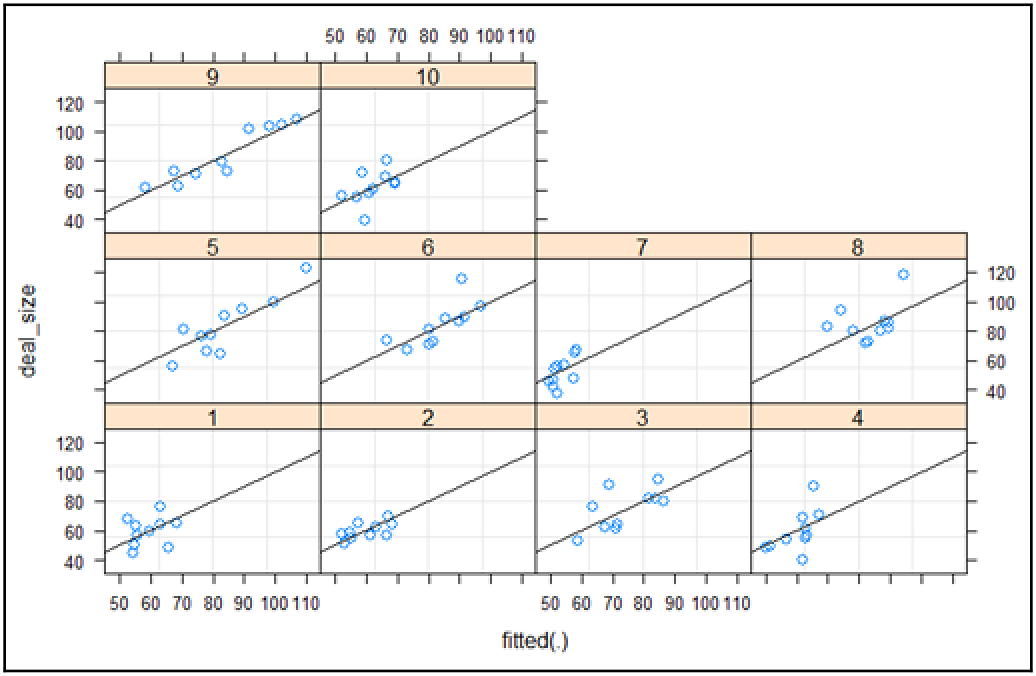
Sales people involved / Deal size slope
Fixed + Random effect
0
1
2
3
4
5
9
6
8
3
4
10
2
1
7
Clientid

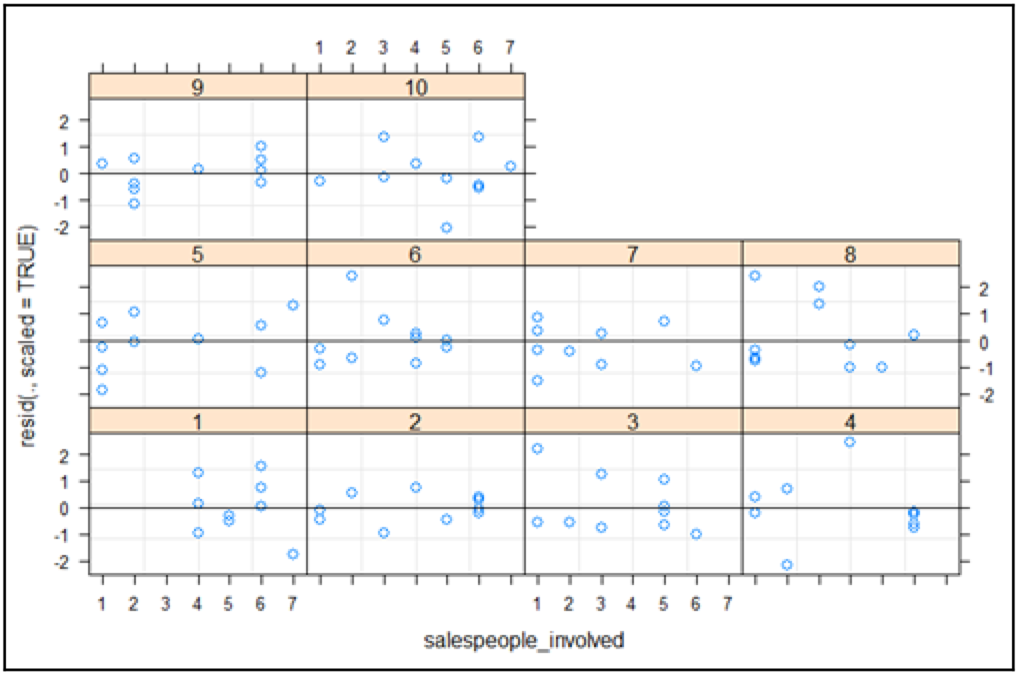


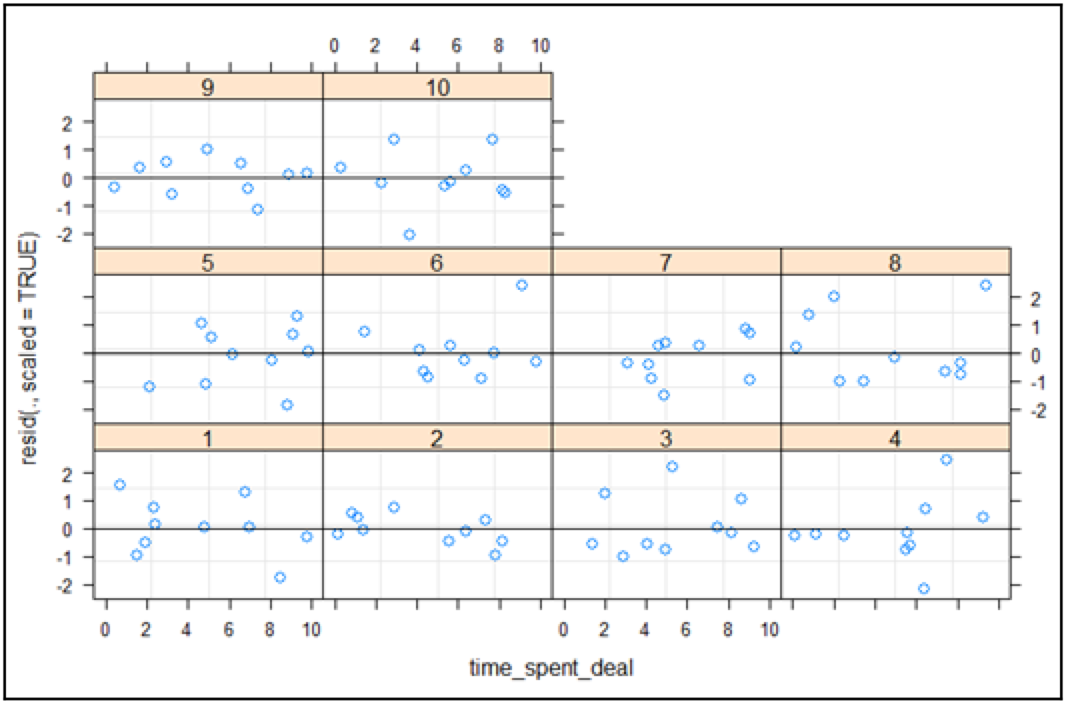
Random effects
clientid (Intercept)
salespeople_involved
time_spent_deal
10
9
8
7
6
5
4
3
2
1
-20
-10
0
10
20


9
10
5
6
7
8
1
2
3
4
resid(., scaled = TRUE)
fitted(.)
50 60 70 80 90 100 110
2
1
0
-1
-2


10
9
8
7
6
5
4
3
2
1
clientid
-2
-1
0
1
2
resid(., scaled = TRUE)

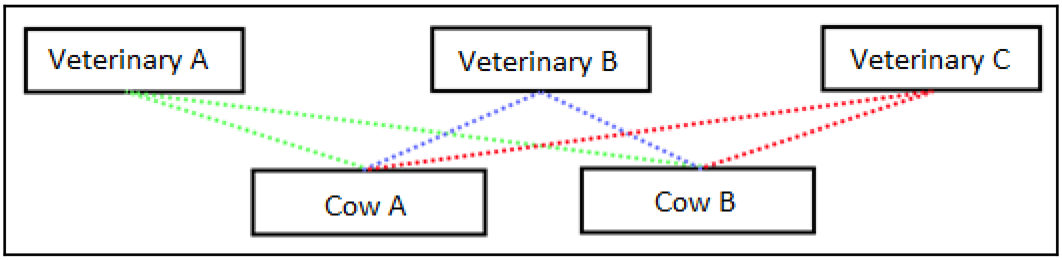

9
10
5
6
7
8
1
2
3
4
deal_size
fitted(.)

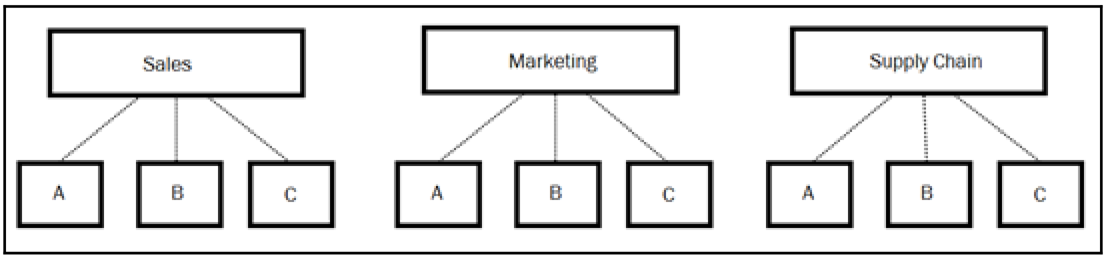

9
10
5
6
7
8
1
2
3
4
resid(., scaled = TRUE)
salespeople_involved

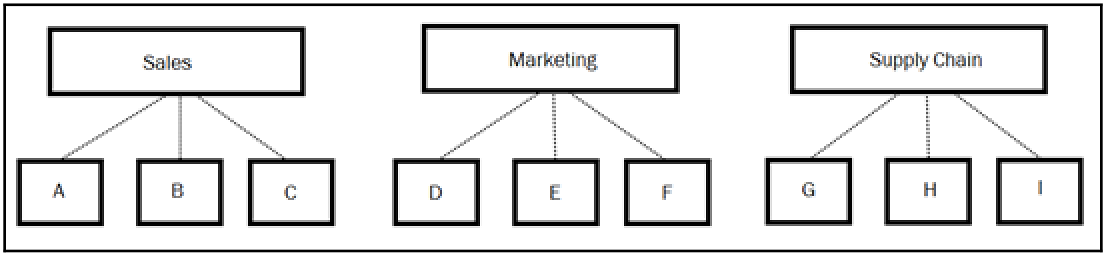


9
10
5
6
7
8
1
2
3
4
resid(., scaled = TRUE)
time_spent_deal

```
> optim(c(1,1,0.03),get__loglikelihood,method="BFGS",control=list(trace=1,REPORT=1))
initial  value -778.940438
iter   2 value -1150.292479
iter   3 value -1161.257702
iter   4 value -1175.141029
iter   5 value -1175.588396
iter   5 value -1168.403484
final  value -1175.588396
converged
$par
[1] 1.01401561 0.96580905 0.05792872

$value
[1] -1175.588

$counts
function gradient
      68        5

$convergence
[1] 0

$message
NULL
```








```
> xtabs(~ Group + Person, data)
           Person
Group       A B C D
  Marketing 5 5 5 5
  Sales     5 5 5 5
```

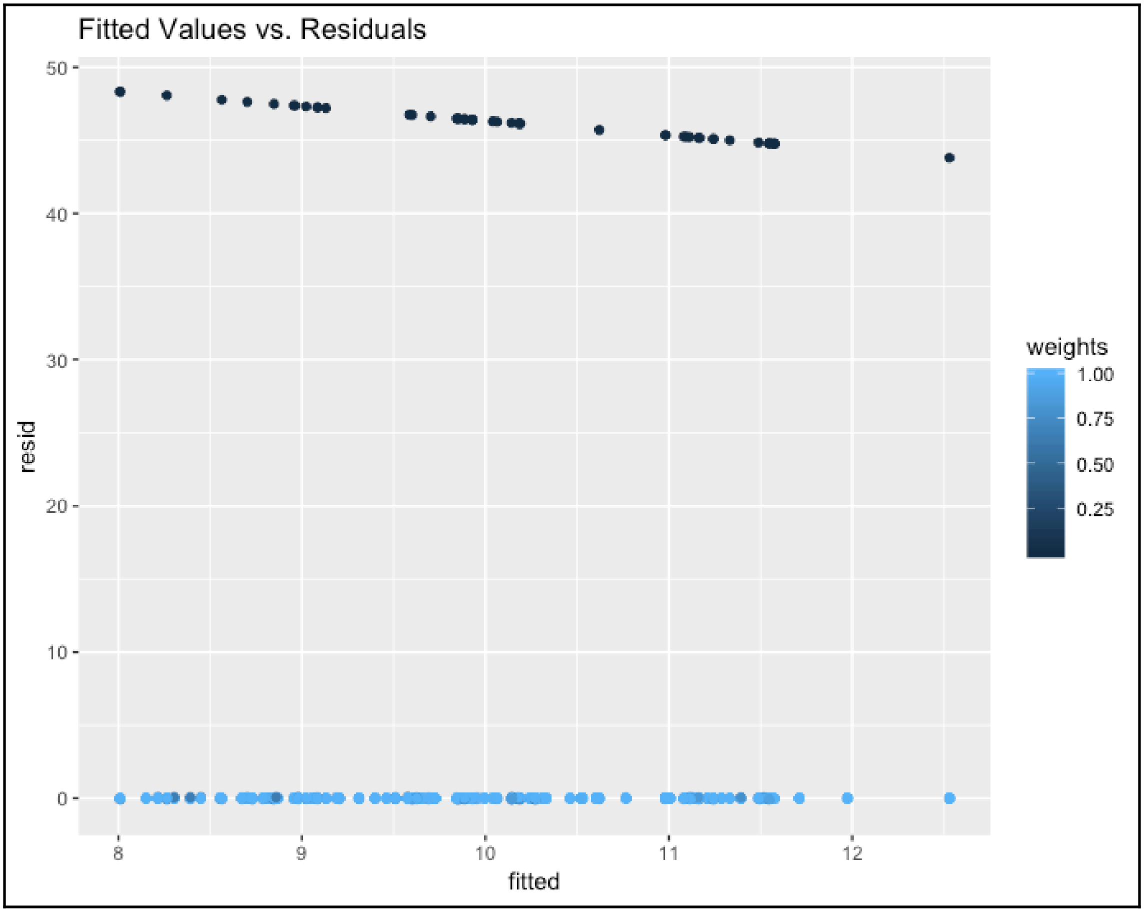
```
> lmer(Rating ~ -1 + (1 | Group/Person)        , data = data)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Rating ~ -1 + (1 | Group/Person)
   Data: data
      AIC       BIC    logLik  deviance  df.resid
 223.0909  228.1576 -108.5455  217.0909        37
Random effects:
 Groups       Name        Std.Dev.
 Person:Group (Intercept)  3.001
 Group        (Intercept) 26.755
 Residual                  2.584
Number of obs: 40, groups:  Person:Group, 8; Group, 2
No fixed effect coefficients
>
> lmer(Rating ~ -1 + (1 | Group) + (1 | Person), data = data)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Rating ~ -1 + (1 | Group) + (1 | Person)
   Data: data
      AIC       BIC    logLik  deviance  df.resid
 217.4469  222.5135 -105.7234  211.4469        37
Random effects:
 Groups   Name        Std.Dev.
 Person   (Intercept)  3.061
 Group    (Intercept) 26.688
 Residual              2.530
Number of obs: 40, groups:  Person, 4; Group, 2
No fixed effect coefficients
```

```
> xtabs(~ Group + Person, data2)
           Person
Group       A AA B BB C CC D DD
  Marketing 5  0 5  0 5  0 5  0
  Sales     0  5 0  5 0  5 0  5
```

```
> lmer(Rating ~ -1 + (1 | Group/Person)          , data = data2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Rating ~ -1 + (1 | Group/Person)
   Data: data2
      AIC       BIC    logLik  deviance  df.resid
 223.0909  228.1576 -108.5455  217.0909        37
Random effects:
 Groups       Name        Std.Dev.
 Person:Group (Intercept)  3.001
 Group        (Intercept) 26.755
 Residual                  2.584
Number of obs: 40, groups:  Person:Group, 8; Group, 2
No fixed effect coefficients
>
> lmer(Rating ~ -1 + (1 | Group) + (1 | Person), data = data2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Rating ~ -1 + (1 | Group) + (1 | Person)
   Data: data2
      AIC       BIC    logLik  deviance  df.resid
 223.0909  228.1576 -108.5455  217.0909        37
Random effects:
 Groups   Name        Std.Dev.
 Person   (Intercept)  3.001
 Group    (Intercept) 26.755
 Residual              2.584
Number of obs: 40, groups:  Person, 8; Group, 2
No fixed effect coefficients
```

```
> lmer(data=XYG, Y ~ 1 + (1|G))
Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ 1 + (1 | G)
   Data: XYG
REML criterion at convergence: -3197.028
Random effects:
 Groups   Name        Std.Dev.
 G        (Intercept) 1.02950
 Residual             0.03066
Number of obs: 1000, groups:  G, 100
Fixed Effects:
(Intercept)
      9.988
> rlmer(data=XYG, Y ~ 1 + (1|G))
Robust linear mixed model fit by DAStau
Formula: Y ~ 1 + (1 | G)
   Data: XYG
Random effects:
 Groups   Name        Std.Dev.
 G        (Intercept) 1.14192
 Residual             0.03073
Number of obs: 1000, groups: G, 100
Fixed Effects:
(Intercept)
      9.984
```

```
> lmer(data=XYG, Y ~ 1 + (1|G))
Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ 1 + (1 | G)
   Data: XYG
REML criterion at convergence: 7468.825
Random effects:
 Groups   Name        Std.Dev.
 G        (Intercept) 1.929
 Residual             9.974
Number of obs: 1000, groups:  G, 100
Fixed Effects:
(Intercept)
       12.3
> rlmer(data=XYG, Y ~ 1 + (1|G))
Robust linear mixed model fit by DAStau
Formula: Y ~ 1 + (1 | G)
   Data: XYG
Random effects:
 Groups   Name        Std.Dev.
 G        (Intercept) 1.14287
 Residual             0.03437
Number of obs: 1000, groups: G, 100
Fixed Effects:
(Intercept)
      9.987
```



Fitted Values vs. Residuals
50
40
30
20
10
0
resid
8
9
10
11
12
fitted
weights
1.00
0.75
0.50
0.25

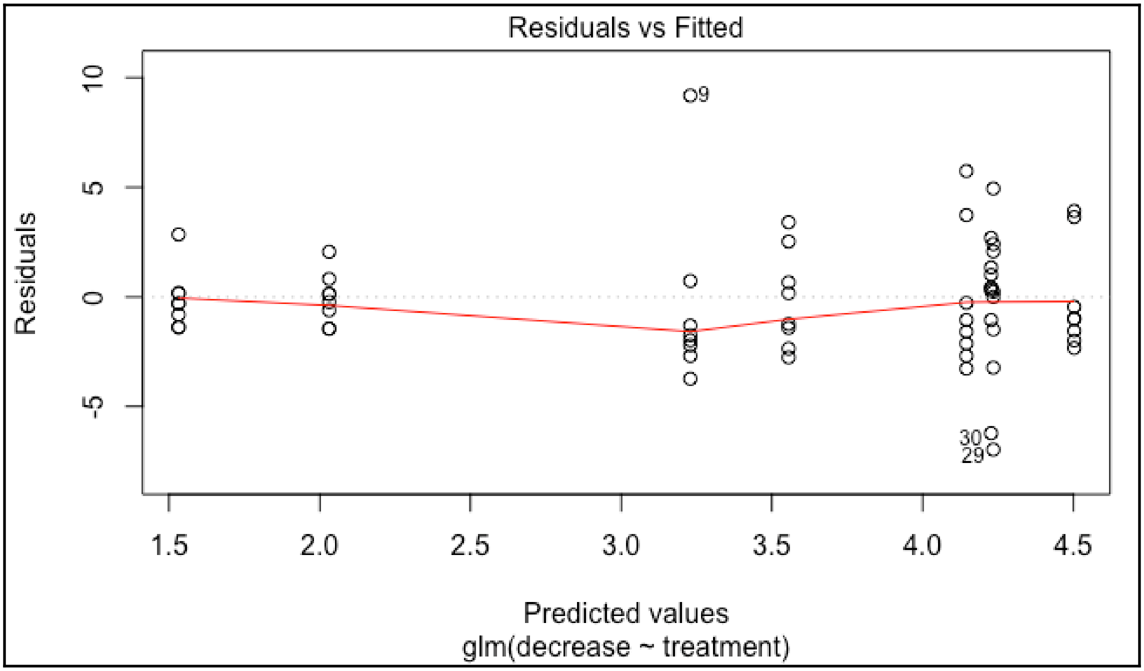
```
> cAIC(m1)$caic
[1] 769.6015
>
>
> cAIC(m2)$caic
[1] 771.1832
>
>
> cAIC(m3)$caic
[1] 765.4679
>
>
> cAIC(m4)$caic
[1] 775.5881
>
>
> cAIC(m5)$caic
[1] 773.8559
>
>
> cAIC(m6)$caic
singular fit
[1] 780.0531
>
>
> cAIC(m7)$caic
[1] 769.7357
```

```
Call:
glm(formula = decrease ~ treatment, family = poisson(), data = OrchardSprays)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.9617  -1.5516  -0.3743   0.7517   9.1888

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.5315     0.1644   9.316   <2e-16 ***
treatmentB    0.5000     0.2084   2.399   0.0164 *
treatmentC    1.6973     0.1788   9.492   <2e-16 ***
treatmentD    2.0239     0.1749  11.570   <2e-16 ***
treatmentE    2.6136     0.1703  15.346   <2e-16 ***
treatmentF    2.7026     0.1698  15.915   <2e-16 ***
treatmentG    2.6954     0.1699  15.868   <2e-16 ***
treatmentH    2.9711     0.1686  17.627   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1904.73  on 63  degrees of freedom
Residual deviance:  438.95  on 56  degrees of freedom
AIC: 786.47

Number of Fisher Scoring iterations: 5
```

```
$`emmeans of treatment`
 treatment  rate    SE  df asymp.LCL asymp.UCL
 A          4.62 0.760 Inf      3.35      6.38
 B          7.62 0.976 Inf      5.93      9.80
 C         25.25 1.777 Inf     22.00     28.98
 D         35.00 2.092 Inf     31.13     39.35
 E         63.12 2.809 Inf     57.85     68.88
 F         69.00 2.937 Inf     63.48     75.00
 G         68.50 2.926 Inf     63.00     74.48
 H         90.25 3.359 Inf     83.90     97.08
```

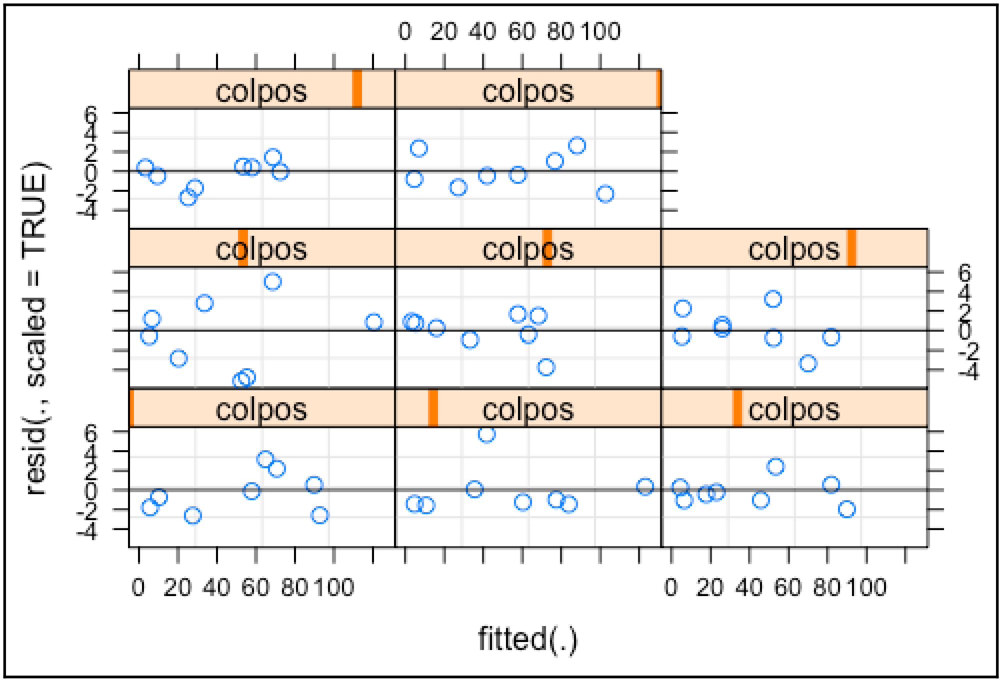
```
Confidence level used: 0.95
Intervals are back-transformed from the log scale

$`pairwise differences of treatment`
 contrast  ratio      SE  df z.ratio p.value
 A / B    0.6066 0.12639 Inf  -2.399 0.2415
 A / C    0.1832 0.03275 Inf  -9.492 <.0001
 A / D    0.1321 0.02311 Inf -11.570 <.0001
 A / E    0.0733 0.01248 Inf -15.346 <.0001
 A / F    0.0670 0.01138 Inf -15.915 <.0001
 A / G    0.0675 0.01147 Inf -15.868 <.0001
 A / H    0.0512 0.00864 Inf -17.627 <.0001
 B / C    0.3020 0.04412 Inf  -8.196 <.0001
 B / D    0.2179 0.03078 Inf -10.785 <.0001
 B / E    0.1208 0.01637 Inf -15.593 <.0001
 B / F    0.1105 0.01491 Inf -16.325 <.0001
 B / G    0.1113 0.01502 Inf -16.265 <.0001
 B / H    0.0845 0.01127 Inf -18.533 <.0001
 C / D    0.7214 0.06660 Inf  -3.537 0.0096
 C / E    0.4000 0.03330 Inf -11.006 <.0001
 C / F    0.3659 0.03009 Inf -12.225 <.0001
 C / G    0.3686 0.03034 Inf -12.125 <.0001
 C / H    0.2798 0.02227 Inf -16.003 <.0001
 D / E    0.5545 0.04131 Inf  -7.915 <.0001
 D / F    0.5072 0.03722 Inf  -9.251 <.0001
 D / G    0.5109 0.03753 Inf  -9.141 <.0001
 D / H    0.3878 0.02730 Inf -13.455 <.0001
 E / F    0.9149 0.05633 Inf  -1.445 0.8361
 E / G    0.9215 0.05684 Inf  -1.325 0.8897
 E / H    0.6994 0.04058 Inf  -6.162 <.0001
 F / G    1.0073 0.06074 Inf   0.121 1.0000
 F / H    0.7645 0.04323 Inf  -4.749 0.0001
 G / H    0.7590 0.04300 Inf  -4.867 <.0001

P value adjustment: tukey method for comparing a family of 8 estimates
Tests are performed on the log scale
```

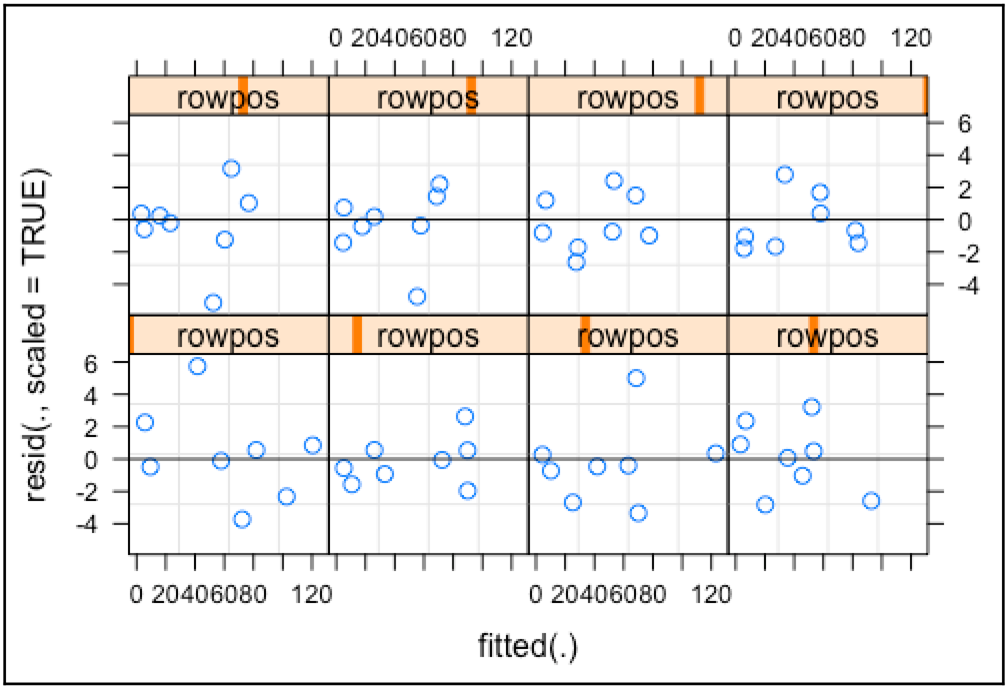
```
> predict(fixed_std_model,data.frame(treatment="D"),type="response")
 1
35
```



```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: poisson  ( log )
Formula: decrease ~ treatment + (1 | colpos) + (1 | rowpos)
   Data: OrchardSprays

     AIC      BIC   logLik deviance df.resid
   667.6    689.2   -323.8    647.6       54

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4858 -1.2184 -0.2889  0.8911  6.5399

Random effects:
 Groups Name        Variance Std.Dev.
 colpos (Intercept) 0.02496  0.1580
 rowpos (Intercept) 0.03792  0.1947
Number of obs: 64, groups:  colpos, 8; rowpos, 8

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.5079     0.1867   8.076 6.68e-16 ***
treatmentB    0.4850     0.2083   2.329   0.0199 *
treatmentC    1.6776     0.1788   9.385  < 2e-16 ***
treatmentD    2.0067     0.1749  11.475  < 2e-16 ***
treatmentE    2.6155     0.1702  15.364  < 2e-16 ***
treatmentF    2.7070     0.1697  15.952  < 2e-16 ***
treatmentG    2.6908     0.1698  15.846  < 2e-16 ***
treatmentH    2.9691     0.1685  17.623  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
           (Intr) trtmnB trtmnC trtmnD trtmnE trtmnF trtmnG
treatmentB -0.694
treatmentC -0.808  0.725
treatmentD -0.826  0.741  0.863
treatmentE -0.849  0.761  0.886  0.907
treatmentF -0.852  0.763  0.889  0.909  0.934
treatmentG -0.851  0.763  0.889  0.909  0.934  0.936
treatmentH -0.858  0.769  0.896  0.916  0.941  0.944  0.943
```

```
> ranef(model_2)
$colpos
  (Intercept)
1  0.20893166
2  0.21171940
3 -0.14365310
4 -0.01940481
5 -0.17732000
6 -0.08980566
7 -0.08152766
8  0.09836171

$rowpos
  (Intercept)
1  0.33434211
2  0.16578727
3  0.12408467
4 -0.15557697
5 -0.23460180
6 -0.16512827
7 -0.07662162
8  0.01880953

with conditional variances for "colpos" "rowpos"
> fixef(model_2)
(Intercept)  treatmentB  treatmentC  treatmentD  treatmentE  treatmentF  treatmentG  treatmentH
  1.5078872   0.4850256   1.6776380   2.0066968   2.6154742   2.7070001   2.6908219   2.9691472
> VarCorr(model_2)
 Groups Name        Std.Dev.
 colpos (Intercept) 0.15798
 rowpos (Intercept) 0.19474
```


colpos
resid(., scaled = TRUE)
fitted(.)


rowpos
resid(., scaled = TRUE)
fitted(.)

```
Computing profile confidence intervals ...
                 2.5 %    97.5 %
.sig01      0.09442293 0.2976190
.sig02      0.12077033 0.3597700
(Intercept) 1.12576452 1.8632398
treatmentB  0.08171115 0.9019814
treatmentC  1.34000298 2.0435608
treatmentD  1.67758110 2.3658719
treatmentE  2.29653348 2.9665200
treatmentF  2.38922095 3.0571158
treatmentG  2.37281217 3.0411236
treatmentH  2.65411339 3.3171222
```

```
$`pairwise differences of treatment`
 contrast  ratio      SE  df z.ratio p.value
 A / B    0.6157 0.12824 Inf  -2.329 0.2778
 A / C    0.1868 0.03339 Inf  -9.385 <.0001
 A / D    0.1344 0.02351 Inf -11.475 <.0001
 A / E    0.0731 0.01245 Inf -15.364 <.0001
 A / F    0.0667 0.01133 Inf -15.952 <.0001
 A / G    0.0678 0.01152 Inf -15.846 <.0001
 A / H    0.0513 0.00865 Inf -17.623 <.0001
 B / C    0.3034 0.04432 Inf  -8.166 <.0001
 B / D    0.2183 0.03085 Inf -10.771 <.0001
 B / E    0.1188 0.01610 Inf -15.718 <.0001
 B / F    0.1084 0.01463 Inf -16.463 <.0001
 B / G    0.1102 0.01487 Inf -16.339 <.0001
 B / H    0.0834 0.01112 Inf -18.632 <.0001
 C / D    0.7196 0.06654 Inf  -3.559 0.0089
 C / E    0.3915 0.03272 Inf -11.221 <.0001
 C / F    0.3572 0.02945 Inf -12.485 <.0001
 C / G    0.3631 0.02993 Inf -12.290 <.0001
 C / H    0.2749 0.02194 Inf -16.179 <.0001
 D / E    0.5440 0.04065 Inf  -8.148 <.0001
 D / F    0.4964 0.03659 Inf  -9.501 <.0001
 D / G    0.5045 0.03717 Inf  -9.287 <.0001
 D / H    0.3820 0.02699 Inf -13.620 <.0001
 E / F    0.9125 0.05644 Inf  -1.480 0.8184
 E / G    0.9274 0.05749 Inf  -1.215 0.9277
 E / H    0.7021 0.04093 Inf  -6.067 <.0001
 F / G    1.0163 0.06153 Inf   0.267 1.0000
 F / H    0.7694 0.04374 Inf  -4.611 0.0001
 G / H    0.7571 0.04318 Inf  -4.880 <.0001

P value adjustment: tukey method for comparing a family of 8 estimates
Tests are performed on the log scale
```

```
Call:
glm.nb(formula = decrease ~ treatment, data = OrchardSprays,
    init.theta = 6.700856316, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4728  -0.7266  -0.2311   0.3033   3.5589

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.5315     0.2137   7.165 7.76e-13 ***
treatmentB    0.5000     0.2841   1.760   0.0785 .
treatmentC    1.6973     0.2632   6.448 1.13e-10 ***
treatmentD    2.0239     0.2606   7.766 8.07e-15 ***
treatmentE    2.6136     0.2575  10.149  < 2e-16 ***
treatmentF    2.7026     0.2572  10.508  < 2e-16 ***
treatmentG    2.6954     0.2572  10.479  < 2e-16 ***
treatmentH    2.9711     0.2564  11.590  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(6.7009) family taken to be 1)

    Null deviance: 318.938  on 63  degrees of freedom
Residual deviance:  63.893  on 56  degrees of freedom
AIC: 527.16

Number of Fisher Scoring iterations: 1


              Theta:  6.70
          Std. Err.:  1.46

 2 x log-likelihood:  -509.16
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Negative Binomial(8.6028)  ( log )
Formula: decrease ~ treatment + (1 | colpos) + (1 | rowpos)
   Data: OrchardSprays

     AIC      BIC   logLik deviance df.resid
   526.5    550.2   -252.2    504.5       53

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7969 -0.5954 -0.0862  0.2879  4.0502

Random effects:
 Groups Name        Variance Std.Dev.
 colpos (Intercept) 0.003765 0.06136
 rowpos (Intercept) 0.031372 0.17712
Number of obs: 64, groups:  colpos, 8; rowpos, 8

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.4986     0.2155   6.954 3.55e-12 ***
treatmentB    0.5262     0.2706   1.944   0.0519 .
treatmentC    1.6457     0.2498   6.587 4.47e-11 ***
treatmentD    2.0298     0.2456   8.263  < 2e-16 ***
treatmentE    2.6410     0.2429  10.874  < 2e-16 ***
treatmentF    2.7176     0.2426  11.202  < 2e-16 ***
treatmentG    2.7262     0.2422  11.254  < 2e-16 ***
treatmentH    2.9892     0.2410  12.405  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr) trtmnB trtmnC trtmnD trtmnE trtmnF trtmnG
treatmentB -0.721
treatmentC -0.775  0.619
treatmentD -0.794  0.633  0.685
treatmentE -0.805  0.641  0.689  0.707
treatmentF -0.805  0.643  0.695  0.707  0.717
treatmentG -0.806  0.642  0.691  0.707  0.717  0.716
treatmentH -0.810  0.645  0.696  0.710  0.721  0.720  0.721
```

```
> print(paste("p of edible in data=",round(total_proportion,3),
+             "/p of edible in train=",round(train_proportion,3),
+             "/p of edible in test=",round(test_proportion,3)))
[1] "p of edible in data= 0.518 /p of edible in train= 0.518 /p of edible in test= 0.518"
```

```
> print(paste("proportion of edible in fold1=",r1,
+             "/proportion of edible in fold2=",r2,
+             "/proportion of edible in fold3=",r3,
+             "/proportion of edible in fold4=",r4))
[1] "proportion of edible in fold1= 0.517971442639094 /proportion of edible in fold2= 0.517971442639094 /proportion of edible in fold3
= 0.517971442639094 /proportion of edible in fold4= 0.517971442639094"
```

```
> createTimeSlices(r,4,horizon=2)
$train
$train$Training4
[1] 1 2 3 4

$train$Training5
[1] 2 3 4 5

$train$Training6
[1] 3 4 5 6

$train$Training7
[1] 4 5 6 7

$train$Training8
[1] 5 6 7 8


$test
$test$Testing4
[1] 5 6

$test$Testing5
[1] 6 7

$test$Testing6
[1] 7 8

$test$Testing7
[1] 8 9

$test$Testing8
[1]  9 10
```

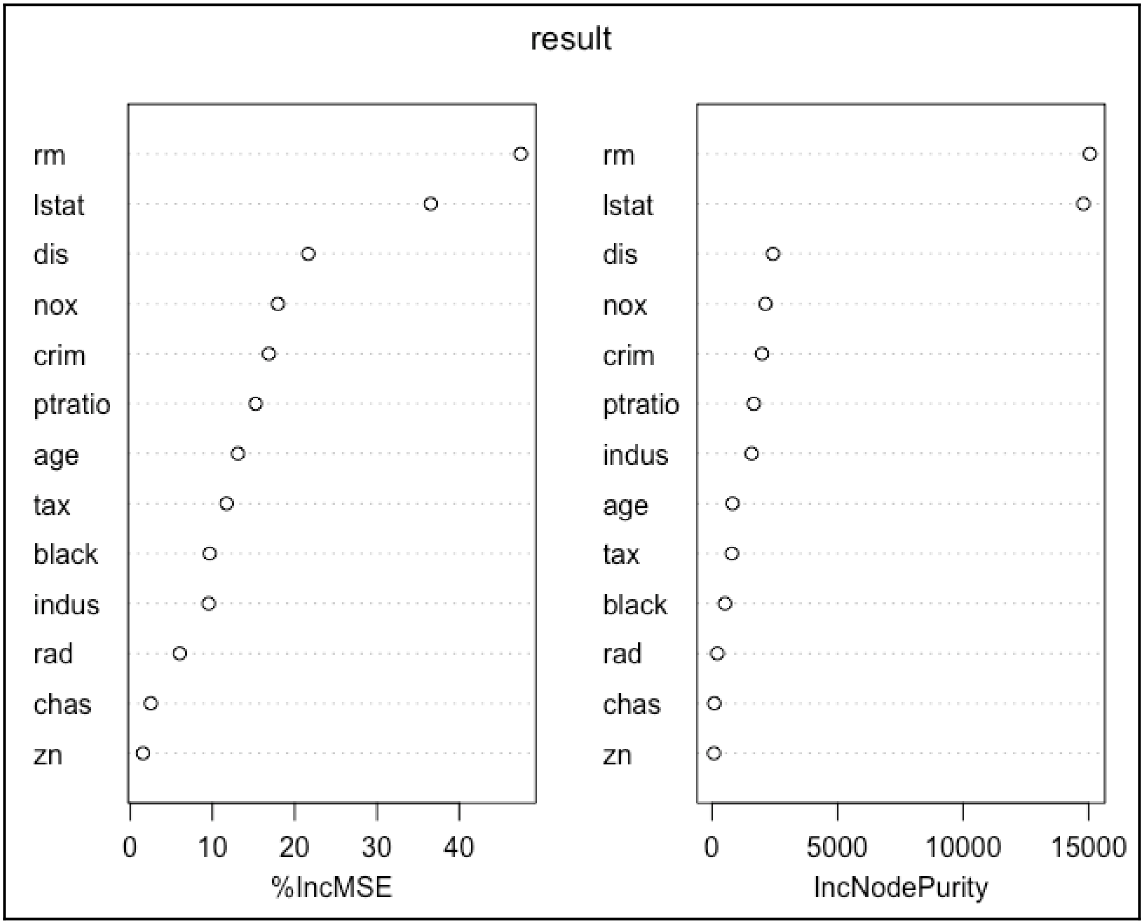
```
> postResample(pred = y_test_pred, obs = testdata$m)
    RMSE Rsquared      MAE
3.019687 0.444644 2.419719
```

```
> postResample(pred = y_test_pred, obs = testdata$m)
    RMSE Rsquared      MAE
3.010622 0.450088 2.393453
```

```
> m3
Support Vector Machines with Linear Kernel

752 samples
  8 predictor

Pre-processing: median imputation (8), scaled (8), centered (8)
Resampling: Cross-Validated (4 fold, repeated 1 times)
Summary of sample sizes: 564, 564, 564, 564
Resampling results across tuning parameters:

  C     RMSE      Rsquared   MAE
  0.01  3.167553  0.4535466  2.501964
  0.10  3.151757  0.4619144  2.492518
  0.20  3.152360  0.4617206  2.491767
  0.30  3.149831  0.4622470  2.489312
  0.40  3.150668  0.4617750  2.490378
  0.50  3.151037  0.4617684  2.491050

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was C = 0.3.
```

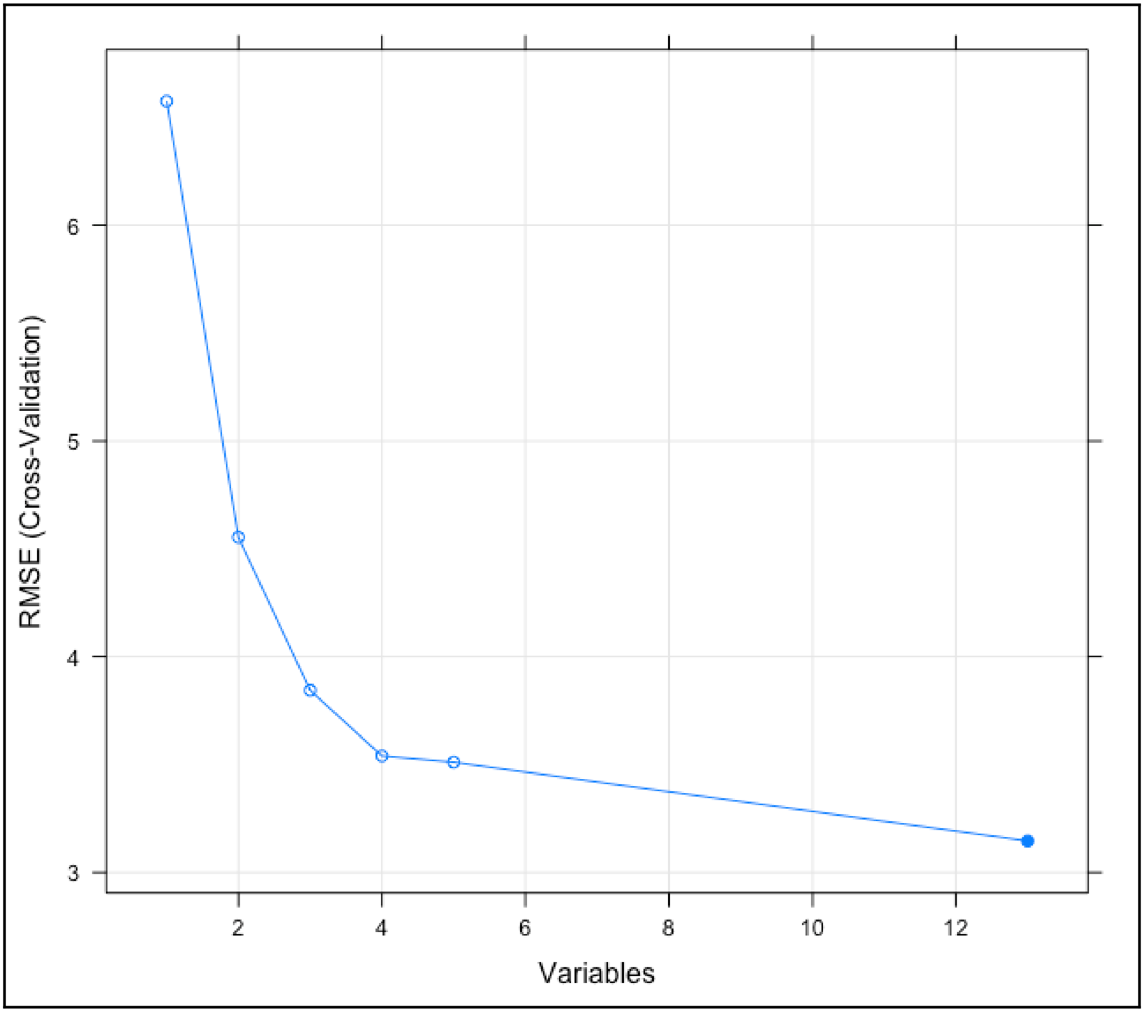
```
Call:
 randomForest(x = x, y = y, mtry = param$mtry, importance = TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 7

          Mean of squared residuals: 9.941527
                    % Var explained: 88.22
```

```
> importance(result)
          %IncMSE IncNodePurity
crim    16.861785    1986.58738
zn       1.558330      74.09081
indus    9.546889    1568.62781
chas     2.505455      87.21040
nox     17.955073    2124.22826
rm      47.477796   15049.08565
age     13.089950     810.19076
dis     21.635264    2425.25057
rad      6.008844     207.38032
tax     11.713522     785.05155
ptratio 15.244567    1654.38507
black    9.659428     512.22050
lstat   36.523813   14796.48140
```




```
Call:
 randomForest(x = x, y = y, mtry = param$mtry)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 1

          Mean of squared residuals: 21.18228
                    % Var explained: 74.91
```

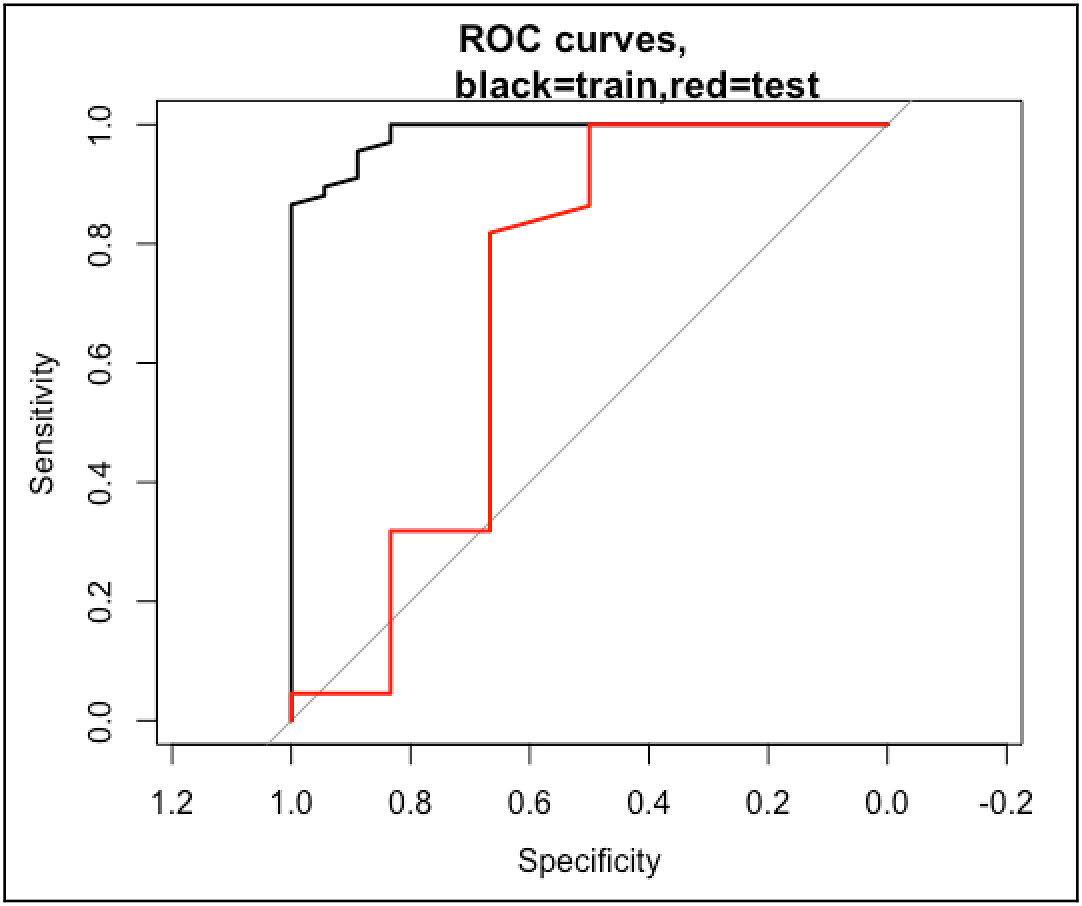
```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold)

Resampling performance over subset size:

 Variables  RMSE Rsquared   MAE RMSESD RsquaredSD  MAESD Selected
         1 6.574   0.5145 4.706 0.9108    0.11433 0.5566
         2 4.553   0.7570 3.122 0.5942    0.05851 0.3117
         3 3.844   0.8261 2.577 0.7870    0.07370 0.3623
         4 3.540   0.8584 2.394 0.5727    0.04900 0.2479
         5 3.511   0.8656 2.387 0.6134    0.05770 0.3051
        13 3.146   0.8916 2.141 0.4805    0.04053 0.2546        *

The top 5 variables (out of 13):
   rm, lstat, nox, dis, crim
```

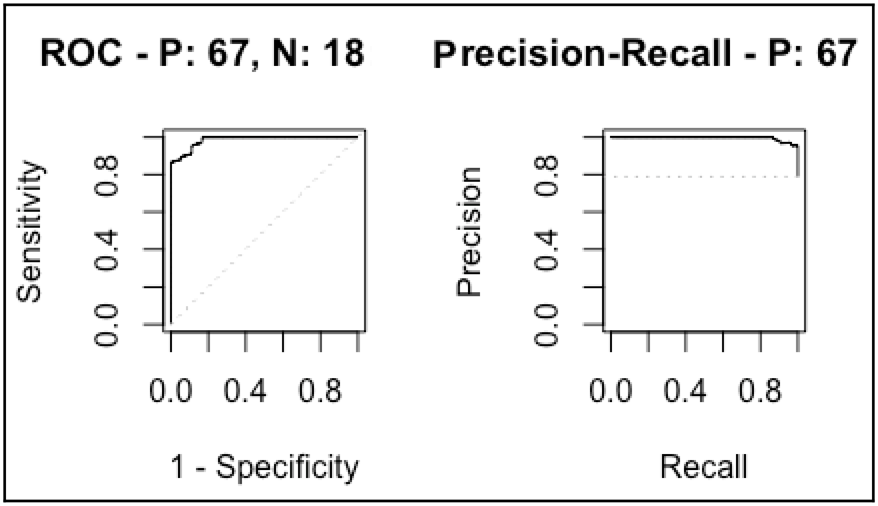

RMSE (Cross-Validation)
6
5
4
3
2
4
6
8
10
12
Variables

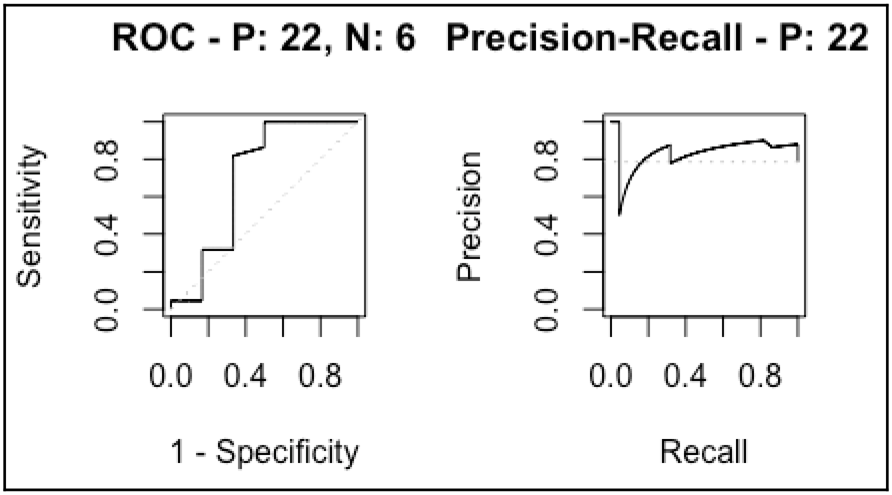
```
Random Forest

8124 samples
 116 predictor
   2 classes: 'e', 'p'

No pre-processing
Resampling: Cross-Validated (4 fold, repeated 1 times)
Summary of sample sizes: 6093, 6093, 6093, 6093
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
   2    0.9767356  0.9533301
   5    1.0000000  1.0000000
   7    1.0000000  1.0000000
  10    1.0000000  1.0000000

Accuracy was used to select the optimal model using  the largest value.
The final value used for the model was mtry = 5.
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved           13            5
  not_approved        0           67

               Accuracy : 0.9412
                 95% CI : (0.868, 0.9806)
    No Information Rate : 0.8471
    P-Value [Acc > NIR] : 0.006832

                  Kappa : 0.8039
 Mcnemar's Test P-Value : 0.073638

            Sensitivity : 1.0000
            Specificity : 0.9306
         Pos Pred Value : 0.7222
         Neg Pred Value : 1.0000
             Prevalence : 0.1529
         Detection Rate : 0.1529
   Detection Prevalence : 0.2118
      Balanced Accuracy : 0.9653

       'Positive' Class : approved
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            3            3
  not_approved        1           21

               Accuracy : 0.8571
                 95% CI : (0.6733, 0.9597)
    No Information Rate : 0.8571
    P-Value [Acc > NIR] : 0.6292

                  Kappa : 0.5172
 Mcnemar's Test P-Value : 0.6171

            Sensitivity : 0.7500
            Specificity : 0.8750
         Pos Pred Value : 0.5000
         Neg Pred Value : 0.9545
             Prevalence : 0.1429
         Detection Rate : 0.1071
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.8125

       'Positive' Class : approved
```


ROC curves,
black=train,red=test
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
Specificity


ROC - P: 67, N: 18
Precision-Recall - P: 67
Sensitivity
0.0 0.4 0.8
1 - Specificity
Precision
0.0 0.4 0.8
Recall


ROC - P: 22, N: 6
Precision-Recall - P: 22
Sensitivity
0.0 0.4 0.8
1 - Specificity
Precision
0.0 0.4 0.8
Recall

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            3            3
  not_approved        1           21

               Accuracy : 0.8571
                 95% CI : (0.6733, 0.9597)
    No Information Rate : 0.8571
    P-Value [Acc > NIR] : 0.6292

                  Kappa : 0.5172
 Mcnemar's Test P-Value : 0.6171

            Sensitivity : 0.7500
            Specificity : 0.8750
         Pos Pred Value : 0.5000
         Neg Pred Value : 0.9545
             Prevalence : 0.1429
         Detection Rate : 0.1071
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.8125

       'Positive' Class : approved
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            4            2
  not_approved        4           18

               Accuracy : 0.7857
                 95% CI : (0.5905, 0.917)
    No Information Rate : 0.7143
    P-Value [Acc > NIR] : 0.2718

                  Kappa : 0.4324
 Mcnemar's Test P-Value : 0.6831

            Sensitivity : 0.5000
            Specificity : 0.9000
         Pos Pred Value : 0.6667
         Neg Pred Value : 0.8182
             Prevalence : 0.2857
         Detection Rate : 0.1429
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.7000

       'Positive' Class : approved
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            3            3
  not_approved        1           21

               Accuracy : 0.8571
                 95% CI : (0.6733, 0.9597)
    No Information Rate : 0.8571
    P-Value [Acc > NIR] : 0.6292

                  Kappa : 0.5172
 Mcnemar's Test P-Value : 0.6171

            Sensitivity : 0.7500
            Specificity : 0.8750
         Pos Pred Value : 0.5000
         Neg Pred Value : 0.9545
             Prevalence : 0.1429
         Detection Rate : 0.1071
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.8125

       'Positive' Class : approved
```

```
> roc(testdata$Approved_,predict(baseline,testdata,type="prob")$approved)

Call:
roc.default(response = testdata$Approved_, predictor = predict(baseline,     testdata, type = "prob")$approved)

Data: predict(baseline, testdata, type = "prob")$approved in 6 controls (testdata$Approved_ approved) > 22 cases (testdata$Approved_ not_approved).
Area under the curve: 0.7008
> roc(testdata$Approved_,predict(up,testdata,type="prob")$approved)

Call:
roc.default(response = testdata$Approved_, predictor = predict(up,     testdata, type = "prob")$approved)

Data: predict(up, testdata, type = "prob")$approved in 6 controls (testdata$Approved_ approved) > 22 cases (testdata$Approved_ not_approved).
Area under the curve: 0.7083
> roc(testdata$Approved_,predict(smote,testdata,type="prob")$approved)

Call:
roc.default(response = testdata$Approved_, predictor = predict(smote,     testdata, type = "prob")$approved)

Data: predict(smote, testdata, type = "prob")$approved in 6 controls (testdata$Approved_ approved) > 22 cases (testdata$Approved_ not_approved).
Area under the curve: 0.7197
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            3            3
  not_approved        1           21

               Accuracy : 0.8571
                 95% CI : (0.6733, 0.9597)
    No Information Rate : 0.8571
    P-Value [Acc > NIR] : 0.6292

                  Kappa : 0.5172
 Mcnemar's Test P-Value : 0.6171

              Precision : 0.5000
                 Recall : 0.7500
                     F1 : 0.6000
             Prevalence : 0.1429
         Detection Rate : 0.1071
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.8125

       'Positive' Class : approved
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            4            2
  not_approved        4           18

               Accuracy : 0.7857
                 95% CI : (0.5905, 0.917)
    No Information Rate : 0.7143
    P-Value [Acc > NIR] : 0.2718

                  Kappa : 0.4324
 Mcnemar's Test P-Value : 0.6831

              Precision : 0.6667
                 Recall : 0.5000
                     F1 : 0.5714
             Prevalence : 0.2857
         Detection Rate : 0.1429
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.7000

       'Positive' Class : approved
```

```
Confusion Matrix and Statistics

              Reference
Prediction     approved not_approved
  approved            4            2
  not_approved        3           19

               Accuracy : 0.8214
                 95% CI : (0.6311, 0.9394)
    No Information Rate : 0.75
    P-Value [Acc > NIR] : 0.2638

                  Kappa : 0.5
 Mcnemar's Test P-Value : 1.0000

              Precision : 0.6667
                 Recall : 0.5714
                     F1 : 0.6154
             Prevalence : 0.2500
         Detection Rate : 0.1429
   Detection Prevalence : 0.2143
      Balanced Accuracy : 0.7381

       'Positive' Class : approved
```

```
Linear Regression

16 samples
 6 predictor

Pre-processing: centered (6), scaled (6)
Resampling: Cross-Validated (7 fold)
Summary of sample sizes: 14, 13, 13, 14, 14, 14, ...
Resampling results:

  RMSE       Rsquared
  0.3958705  0.999043

Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
The lasso

16 samples
 6 predictor

Pre-processing: centered (6), scaled (6)
Resampling: Cross-Validated (7 fold)
Summary of sample sizes: 14, 14, 13, 14, 13, 14, ...
Resampling results across tuning parameters:

  fraction   RMSE       Rsquared
  0.1000000  2.0163297  0.9983535
  0.1888889  0.8871784  0.9983535
  0.2777778  0.5064000  0.9996527
  0.3666667  0.4786030  0.9997923
  0.4555556  0.3950334  0.9993535
  0.5444444  0.3691649  0.9992959
  0.6333333  0.3942011  0.9991461
  0.7222222  0.3963325  0.9990120
  0.8111111  0.4022890  0.9987205
  0.9000000  0.4044215  0.9985525

RMSE was used to select the optimal model using  the smallest value.
The final value used for the model was fraction = 0.5444444.
```

```
Ridge Regression

16 samples
 6 predictor

Pre-processing: centered (6), scaled (6)
Resampling: Cross-Validated (7 fold)
Summary of sample sizes: 14, 14, 14, 13, 14, 14, ...
Resampling results across tuning parameters:

  lambda        RMSE       Rsquared
  0.0000000000  0.4161288  0.9974995
  0.0001000000  0.3888210  0.9974647
  0.0002371374  0.3725052  0.9973616
  0.0005623413  0.3625536  0.9970605
  0.0013335214  0.3722161  0.9964052
  0.0031622777  0.4134904  0.9953960
  0.0074989421  0.4447313  0.9943925
  0.0177827941  0.4578839  0.9937277
  0.0421696503  0.4658095  0.9930094
  0.1000000000  0.5286827  0.9906528

RMSE was used to select the optimal model using  the smallest value.
The final value used for the model was lambda = 0.0005623413.
```

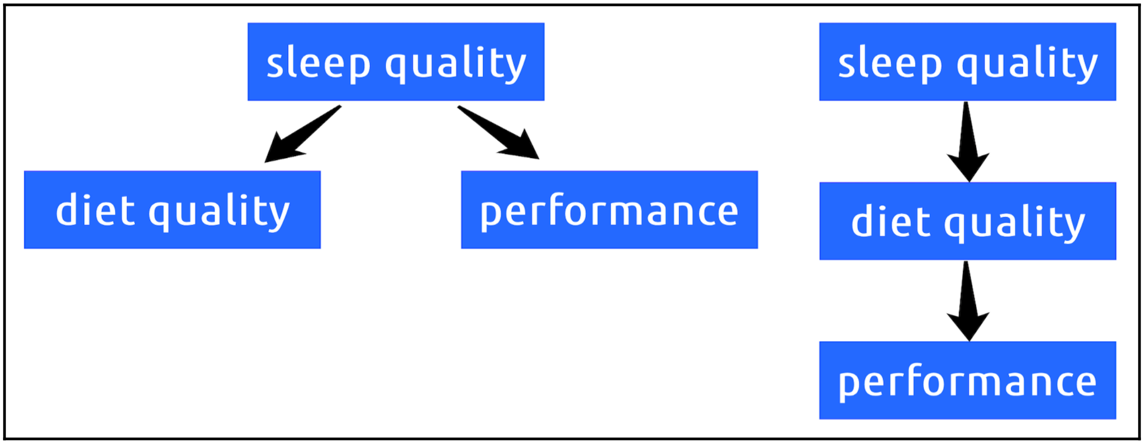
```
  0.8     0.542568144  0.8250467  0.9975710
  0.8     1.253402784  1.3779181  0.9974856
  0.8     2.895522997  2.7206532  0.9974589
  0.9     0.001545254  0.3502197  0.9972639
  0.9     0.003569736  0.3408876  0.9971775
  0.9     0.008246553  0.3346229  0.9967991
  0.9     0.019050608  0.3713379  0.9963457
  0.9     0.044009375  0.4677234  0.9956117
  0.9     0.101667364  0.5640010  0.9964296
  0.9     0.234864797  0.6740046  0.9978305
  0.9     0.542568144  0.8336168  0.9976902
  0.9     1.253402784  1.4628706  0.9975585
  0.9     2.895522997  2.9890655  0.9975556
  1.0     0.001545254  0.3398837  0.9973507
  1.0     0.003569736  0.3289765  0.9972444
  1.0     0.008246553  0.3289986  0.9967737
  1.0     0.019050608  0.3640645  0.9963612
  1.0     0.044009375  0.4913479  0.9957447
  1.0     0.101667364  0.5567210  0.9969338
  1.0     0.234864797  0.6563954  0.9978855
  1.0     0.542568144  0.8211473  0.9978855
  1.0     1.253402784  1.5201484  0.9978855
  1.0     2.895522997  3.2526324  0.9978855

RMSE was used to select the optimal model using  the smallest value.
The final values used for the model were alpha = 1 and lambda = 0.003569736.
```

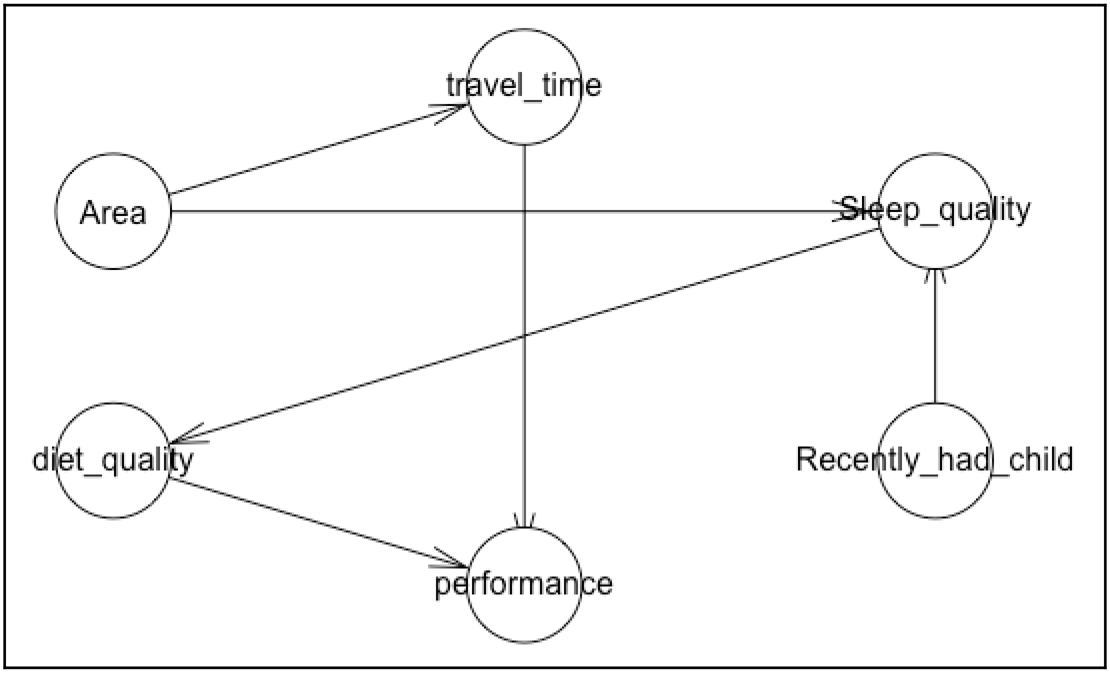
```
> varImp((elasticnet_))
glmnet variable importance

              Overall
Year           100.00
Unemployed      26.34
Population      11.51
Armed.Forces    10.80
GNP              0.00
GNP.deflator     0.00
```

```
Logic Regression

24 samples
13 predictors
 2 classes: '0', '1'

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 20, 19, 19, 19, 19
Resampling results across tuning parameters:

  ntrees  treesize  Accuracy  Kappa
  2        4        0.40      -0.26545455
  2        8        0.36      -0.33818182
  2       16        0.40      -0.26545455
  2       32        0.32      -0.39878788
  3        4        0.48      -0.05030303
  3        8        0.40      -0.23212121
  3       16        0.40      -0.26545455
  3       32        0.32      -0.39878788
  4        4        0.28      -0.47151515
  4        8        0.32      -0.39878788
  4       16        0.32      -0.36545455
  4       32        0.32      -0.36895105

Accuracy was used to select the optimal model using the largest value.
The final values used for the model were treesize = 4 and ntrees = 3.
```

$$g(E[Y]) = \beta_0 + \sum_{j=1}^{t} \beta_j L_j$$

$$L_j = (X_1 \vee X_2)$$

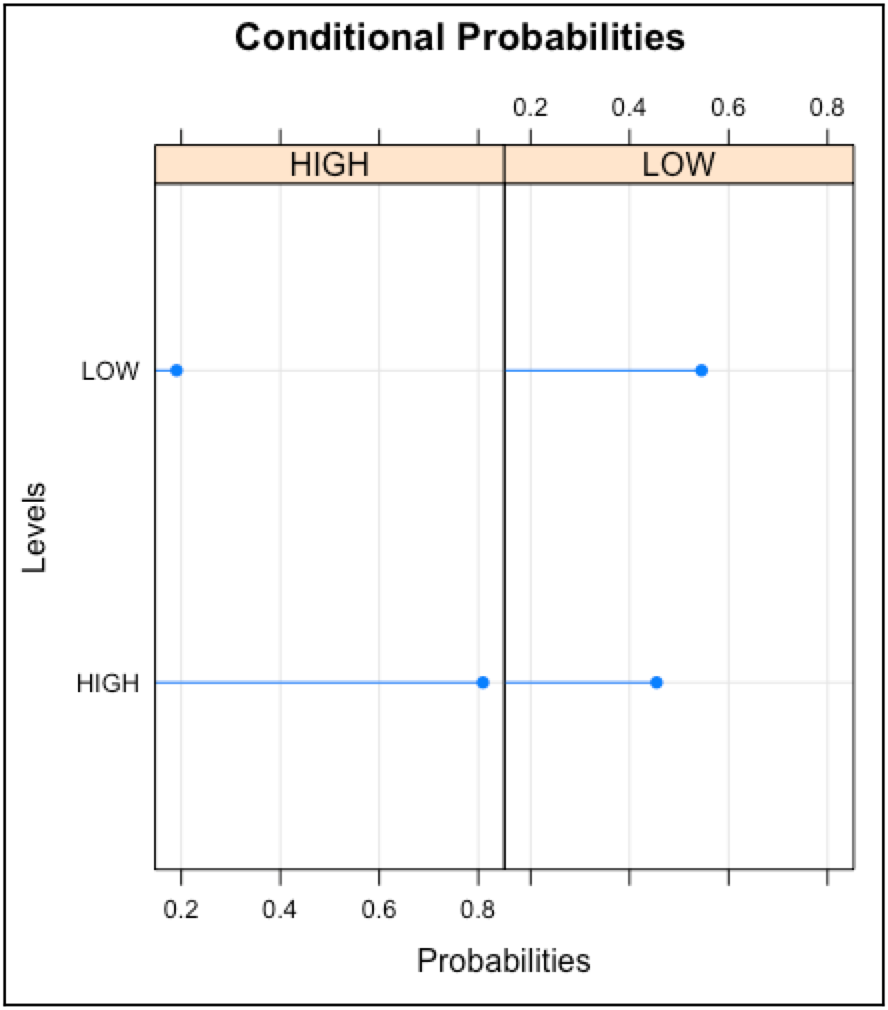

sleep quality
diet quality
performance
sleep quality
diet quality
performance

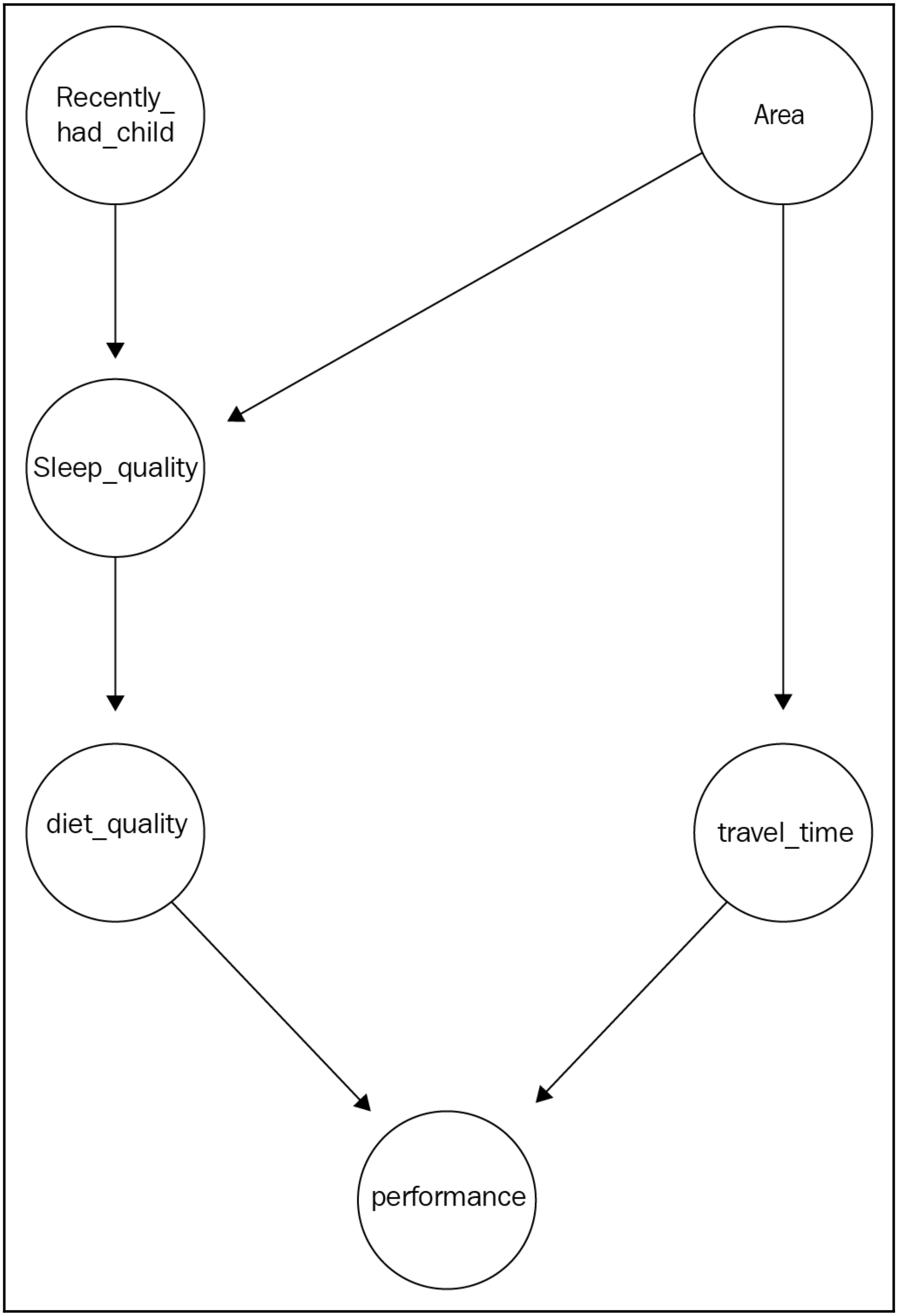

travel_time
Area
Sleep_quality
diet_quality
Recently_had_child
performance

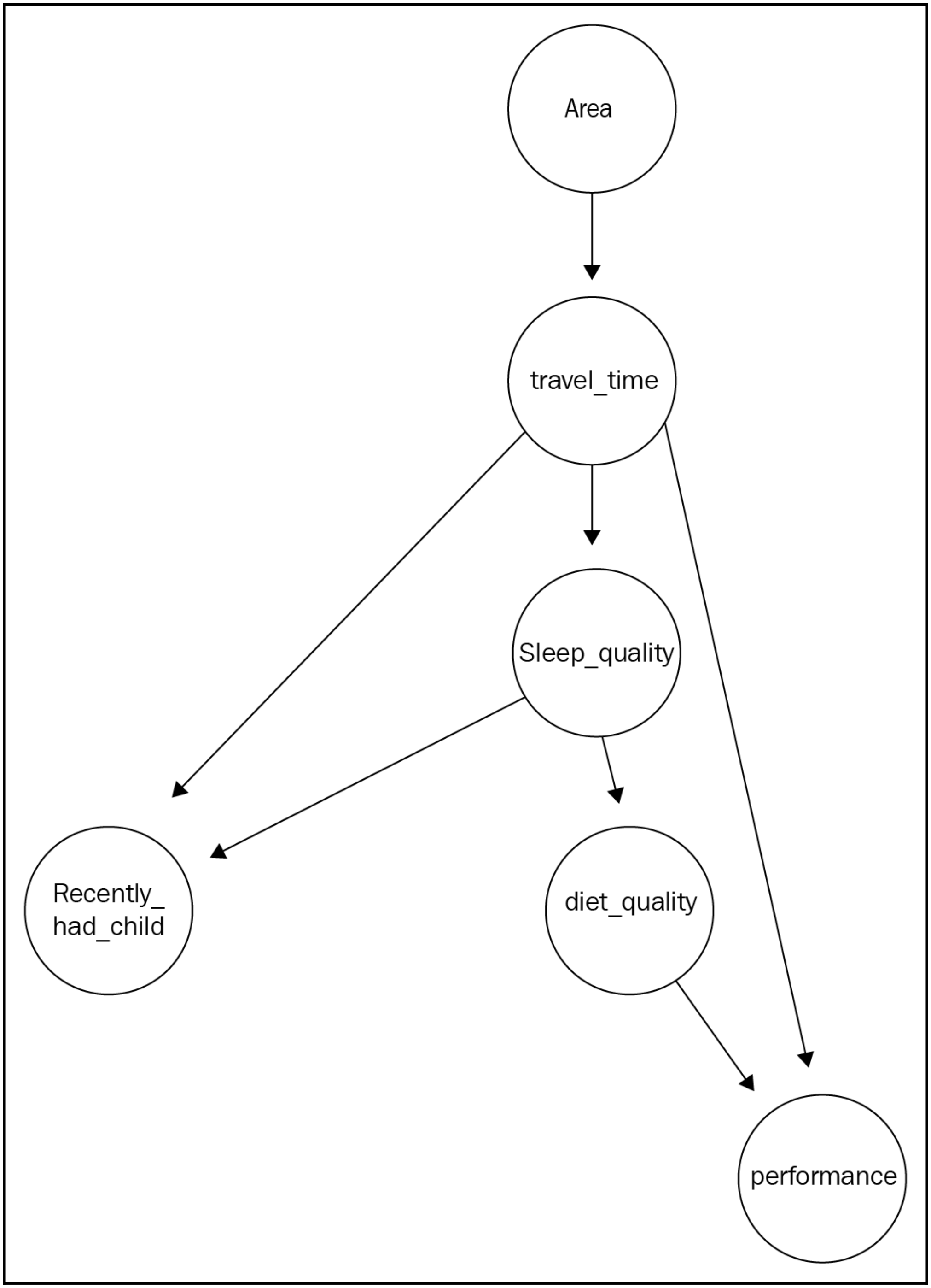
```
> cpquery(fitted, (performance=="HIGH"), (Area=="URBAN"))
[1] 0.7796
> cpquery(fitted, (performance=="HIGH"), (Area=="SUBURBAN"))
[1] 0.6857835
```

```
> cpquery(fitted, (performance=="HIGH"), (travel_time=="HIGH" & Sleep_quality=="HIGH"))
[1] 0.6323752
> cpquery(fitted, (performance=="HIGH"), (travel_time=="HIGH" & Sleep_quality=="LOW"))
[1] 0.5307557
```

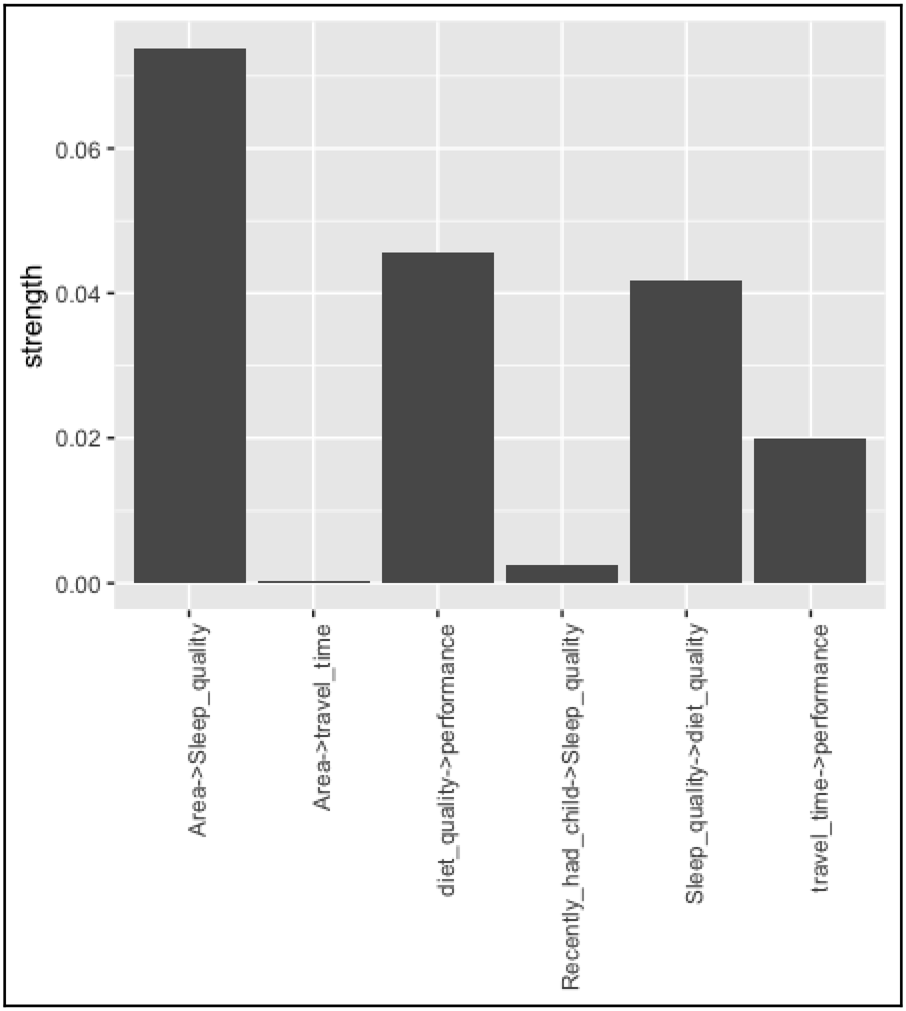
```
> cpquery(fitted, (Sleep_quality=="HIGH"), (performance=="HIGH"))
[1] 0.5980094
> cpquery(fitted, (Sleep_quality=="LOW") , (performance=="HIGH"))
[1] 0.3983425
```

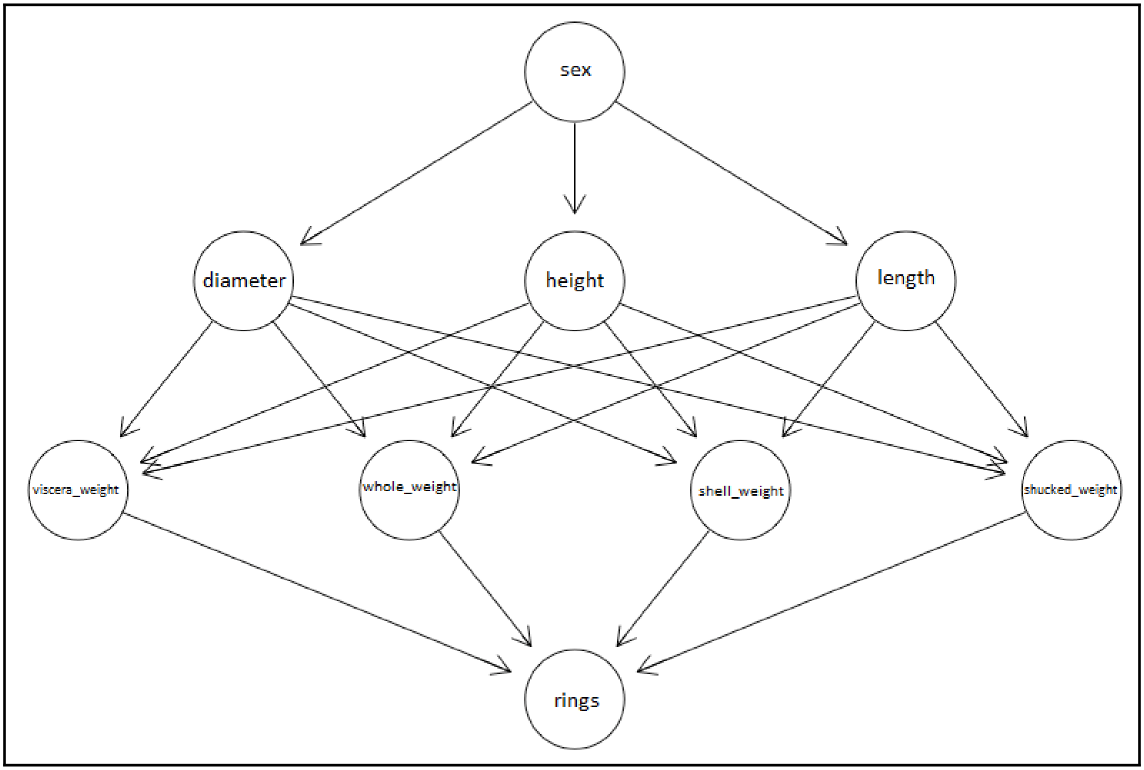

Conditional Probabilities
HIGH
LOW
0.2
0.4
0.6
0.8
LOW
HIGH
Levels
Probabilities


Recently_
had_child
Area
Sleep_quality
diet_quality
travel_time
performance


Area
travel_time
Sleep_quality
Recently_
had_child
diet_quality
performance

```
> score(dag, data, type = "aic")
[1] -115.7599
> score(dag2, data, type = "aic")
[1] -110.9179
```






```
> cpquery(fitted, event = (sex == "M"),evidence = list(diameter=0.65,whole_weight=.8), method = "lw")
[1] 0.7029717
> cpquery(fitted, event = (sex == "M"),evidence = list(diameter=0.15,whole_weight=.8), method = "lw")
[1] 0.1890514
```

```
> cpdist(fitted, nodes = c("height"),evidence = (viscera_weight > 0.4))
      height
1  0.1422997
2  0.2006702
3  0.1655565
4  0.2107666
5  0.1884822
6  0.1556774
7  0.1931824
8  0.1035670
9  0.1453130
10 0.1531892
```

```
> head(predict(fitted,"rings",data))
[1] 1.997482 2.257931 2.134850 1.979325 2.087611 2.521133
> head(predict(fitted,"whole_weight",data))
[1] 0.1706331 0.8539106 0.5759641 0.0993662 0.3762936 0.8709074
```

Abalone network

User defined network

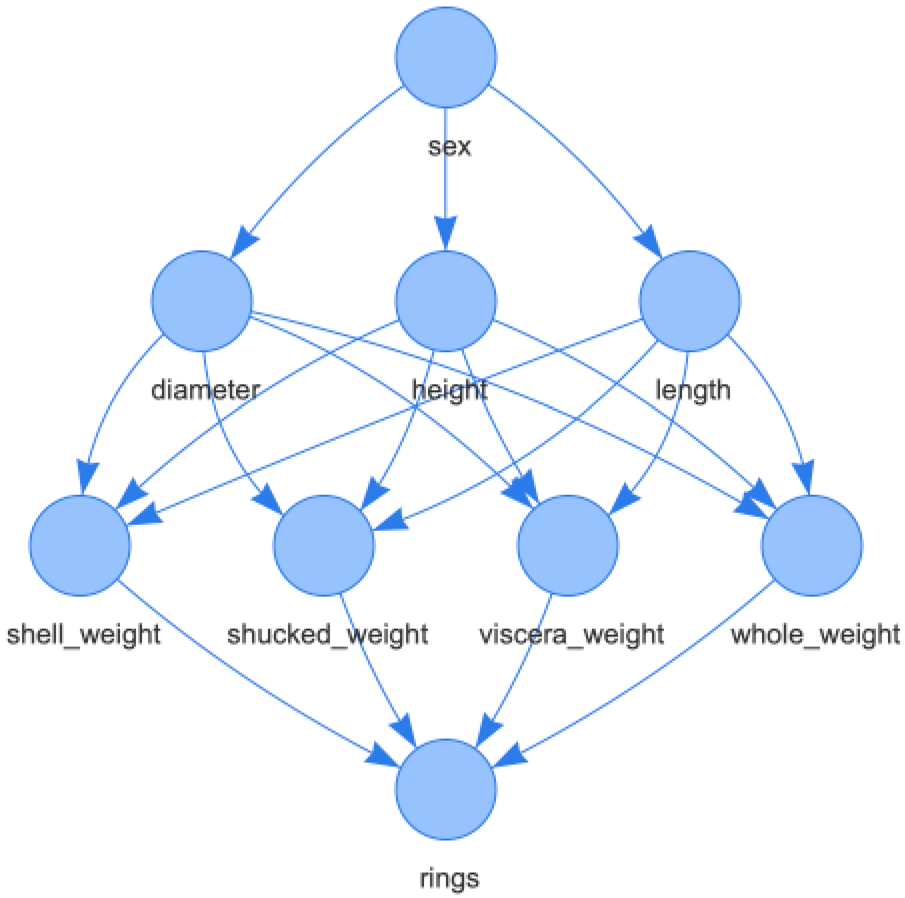


Fig. 1 - live diagram

Abalone network

User defined network

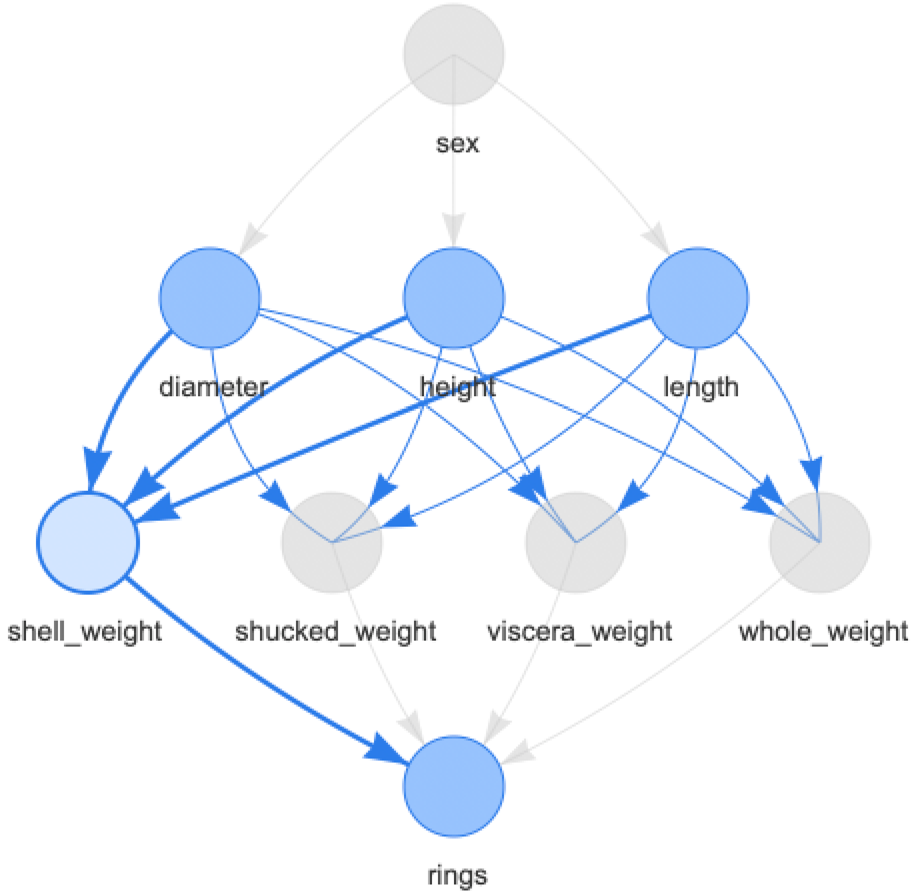


Fig. 1 - live diagram

Abalone network

User defined network

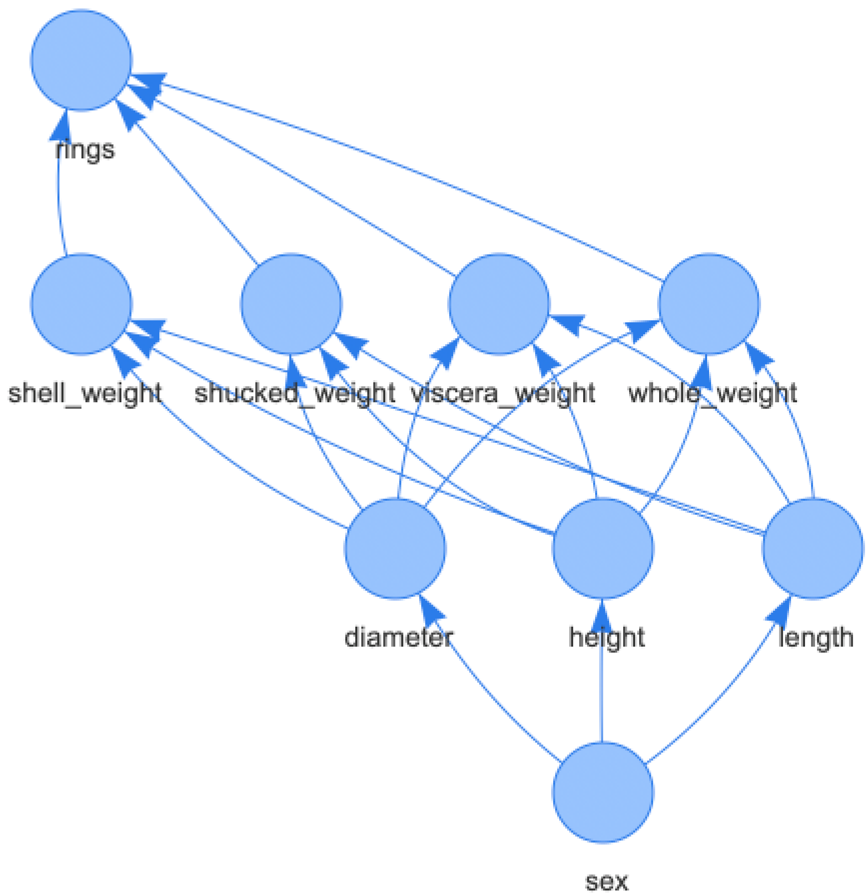
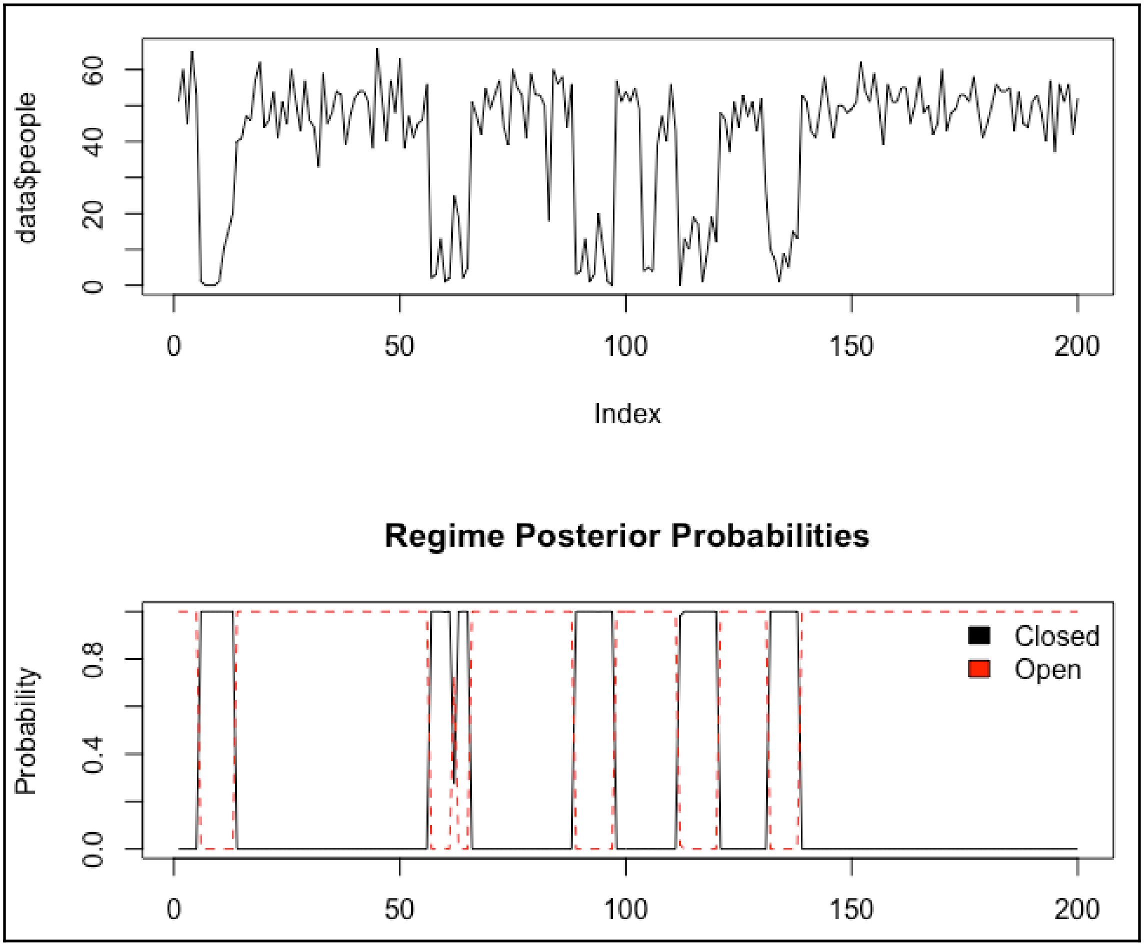
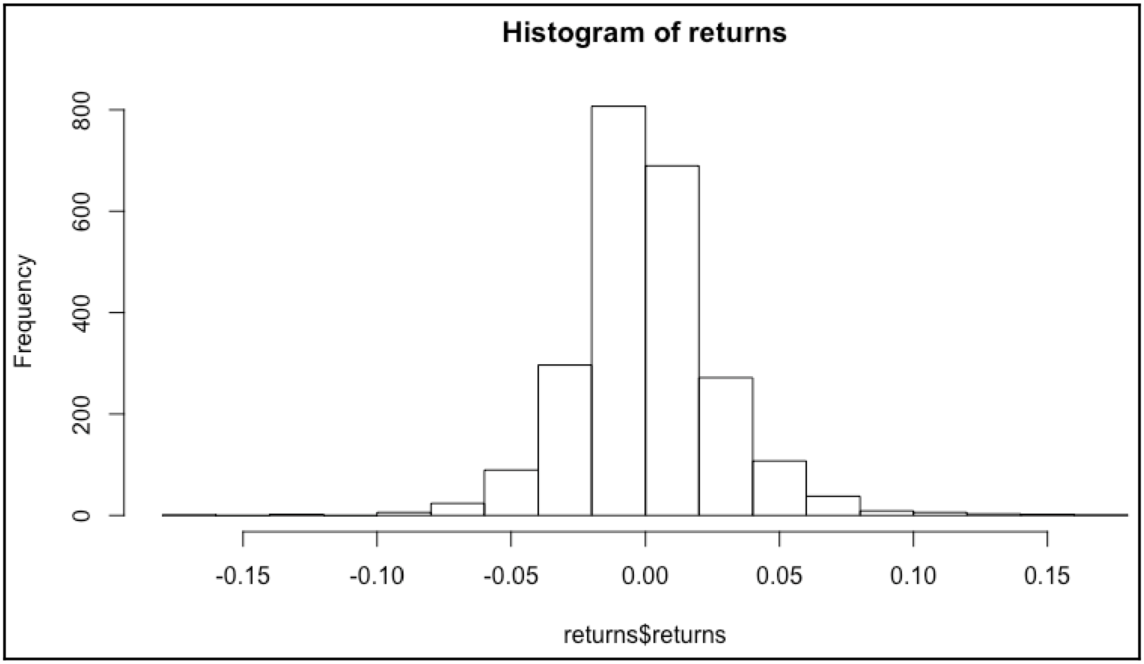


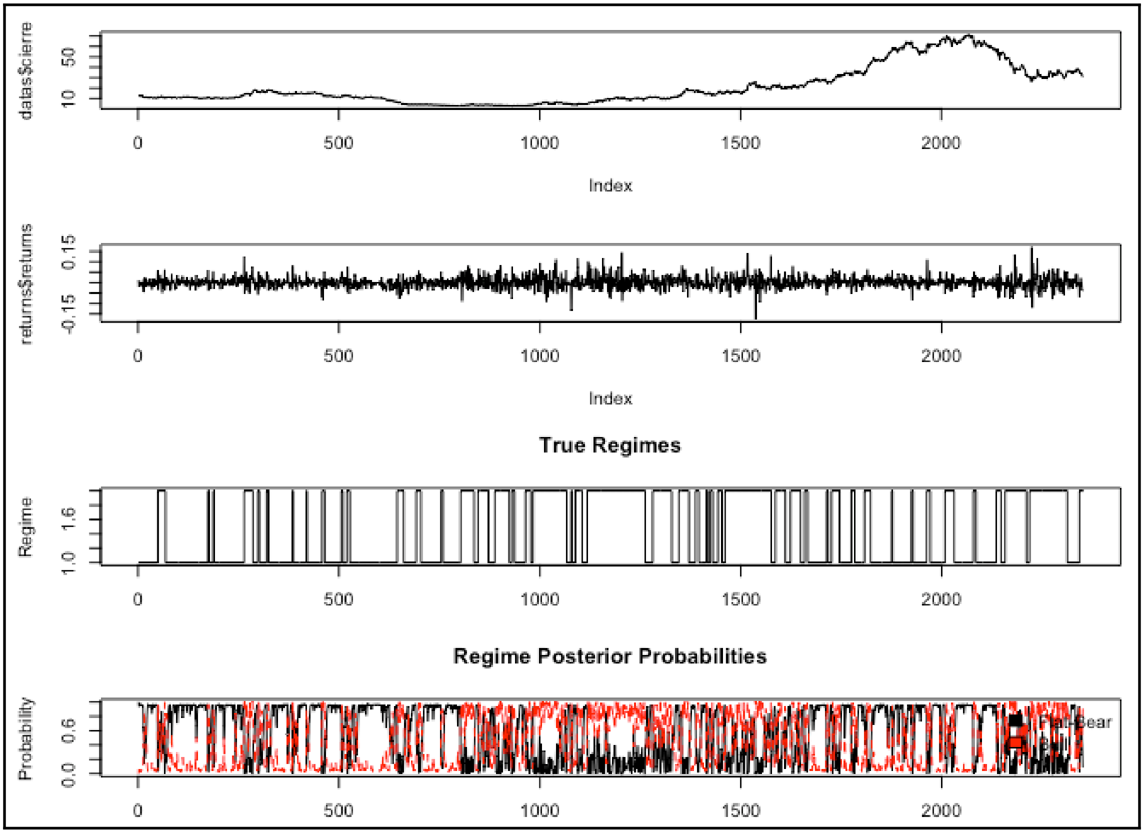
Fig. 1 - live diagram




```
> table(data$state,data$state_pred)

           1   2
  closed  41   1
  open     1 157
```




```
> hmmfit <- fit(hmm, verbose = FALSE)
converged at iteration 65 with logLik: 5336.259
```


datas$cierre
Index
returns$returns
Index
True Regimes
Regime
Regime Posterior Probabilities
Probability
Flat-Bear