Project draft

Margaret Perry, Stella Li, Yuqing Geng

11/28/2018

Introduction:

In the past few decades, with the rapid development of social network, governments around the world have adopted it for anti-terrorism and anti-insurgency purposes. The recent Brexit has brought back people's attention to the longstanding issue between northern Ireland and Britain. After decades of conflicts and numerous casualties, many people are still haunted by the aftereffect. We are interested in using the ERGM and Bernoulli Block Model to map out the the ties between members of PIRA, and evaluate what variables contribute to the group-membership within the network based on multivariable logistic regression. We mainly focus on marital status, but for further analysis, we would also take other variables like gender, whether attend university into consideration.

Data:

4. Visualizing

The data that we are using for this experiment was found in the datasets area of the UCINET Software site. According to the site the data was collected by the International Center for the Study of Terrorism, Pennsylvania State University. The ties are composed of 4 types of relationships involvement in a PIRA activity together, friends before joining PIRA movement, blood relatives, and related through marriage. According to the data description the network have "binary and symmetric relations between members". The data covers of the members of the Provisional Irish Republican Army from 1970 to 1998, which the height of the troubles back in Ireland and England. For the members data was collected on information regarding gender, age, marital status, recruiting age, whether or not they attended university, then role and task-related characteristics, lastly brigade membership which is in reference to the cell that the belonged to. The observations are divided in 6 different time periods across the 28 years. However the period 4 and 5 datasets are combined on the website and there is no clear way to distinguish them, but we will not be using this subset in our analysis so it is not our main concern.

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union

## Loading required package: statnet.common

##
## Attaching package: 'statnet.common'

## The following object is masked from 'package:base':
##
##     order
## Loading required package: network

## network: Classes for Relational Data
## Version 1.13.0.1 created on 2015-08-31.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##                     Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Martina Morris, University of Washington
##                     Skye Bender-deMoll, University of Washington
##  For citation information, type citation("network").
##  Type help("network-package") to get started.

##
## Attaching package: 'network'

## The following objects are masked from 'package:igraph':
##
##     %c%, %s%, add.edges, add.vertices, delete.edges,
##     delete.vertices, get.edge.attribute, get.edges,
##     get.vertex.attribute, is.bipartite, is.directed,
##     list.edge.attributes, list.vertex.attributes,
##     set.edge.attribute, set.vertex.attribute

## sna: Tools for Social Network Analysis
## Version 2.4 created on 2016-07-23.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##  For citation information, type citation("sna").
##  Type help(package="sna") to get started.

##
## Attaching package: 'sna'

## The following objects are masked from 'package:igraph':
##
##     betweenness, bonpow, closeness, components, degree,
##     dyad.census, evcent, hierarchy, is.connected, neighborhood,
##     triad.census

##
## ergm: version 3.9.4, created on 2018-08-15
## Copyright (c) 2018, Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Carter T. Butts, University of California -- Irvine
##                     Steven M. Goodreau, University of Washington
##                     Pavel N. Krivitsky, University of Wollongong
##                     Martina Morris, University of Washington
##                     with contributions from
##                     Li Wang
##                     Kirk Li, University of Washington
##                     Skye Bender-deMoll, University of Washington
## Based on "statnet" project software (statnet.org).
## For license and citation information see statnet.org/attribution
## or type citation("ergm").

## NOTE: Versions before 3.6.1 had a bug in the implementation of the
## bd() constriant which distorted the sampled distribution somewhat.
## In addition, Sampson's Monks datasets had mislabeled vertices. See
## the NEWS and the documentation for more details.
```

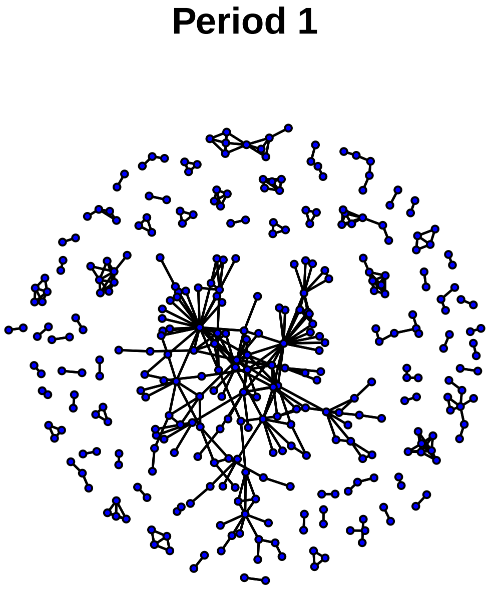
```
##
## Attaching package: 'ergm'

## The following objects are masked from 'package:statnet.common':
##
##     colMeans.mcmc.list, sweep.mcmc.list

## Loading required package: igraphdata

##
## Statistical Analysis of Network Data with R
## Type in C2 (+ENTER) to start with Chapter 2.

## Loading required package: Rcpp

## Loading required package: parallel

## Loading required package: digest

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union

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

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

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:igraph':
##
##     crossing
```

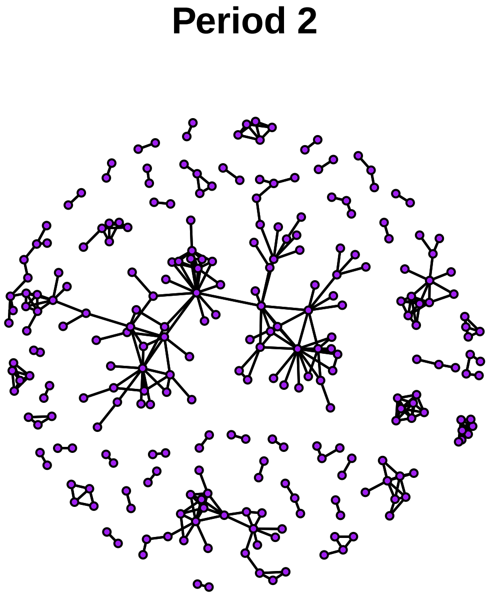
Method

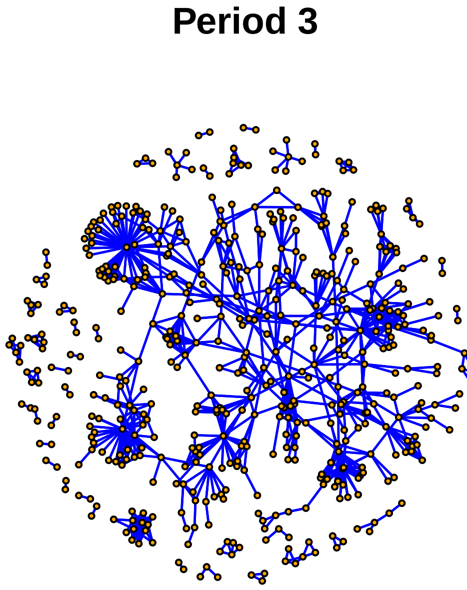
```
# creating networks
net_1ma= network(adj.matrix.1, vertex.attr = data_1ma[,-1], vertex.attrnames = colnames(data1[,-1]), di

net_2ma= network(adj.matrix.2, vertex.attr = data_2ma[,-1], vertex.attrnames = colnames(data2[,-1]), di


net3= network(adj.matrix.3, vertex.attr = data3[,-1], vertex.attrnames = colnames(data3[,-1]), directed


net_6= network(adj.matrix.6, vertex.attr = data_6[,-1], vertex.attrnames = colnames(data6[,-1]), directe
net_6

net_3= network(adj.matrix.3, vertex.attr = data_3[,-1], vertex.attrnames = colnames(data_3[,-1]), direct
net_3
```

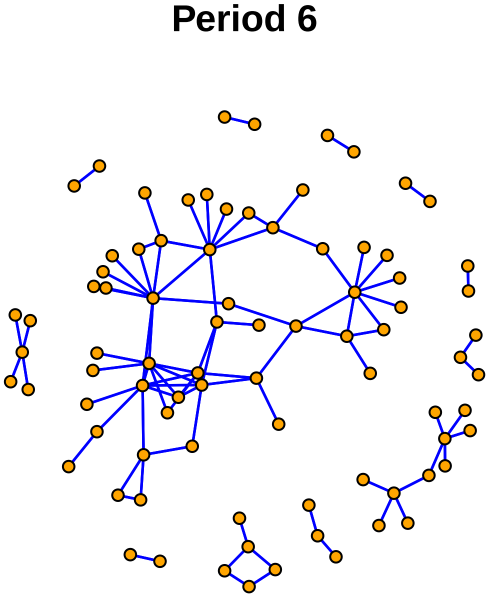
```
net_6ma= network(adj.matrix.6, vertex.attr = data_6ma[,-1], vertex.attrnames = colnames(data6[,-1]), di
net_6

net_3ma= network(adj.matrix.3, vertex.attr = data_3ma[,-1], vertex.attrnames = colnames(data_3ma[,-1]),
net_3
```

Results and Visualization



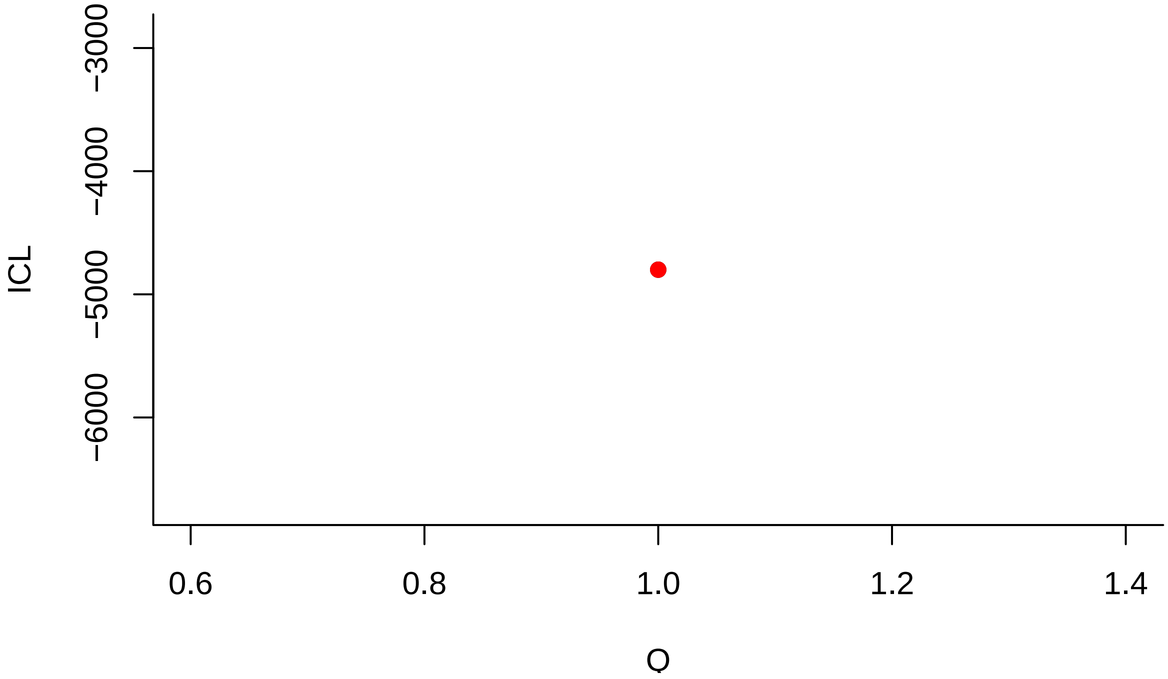


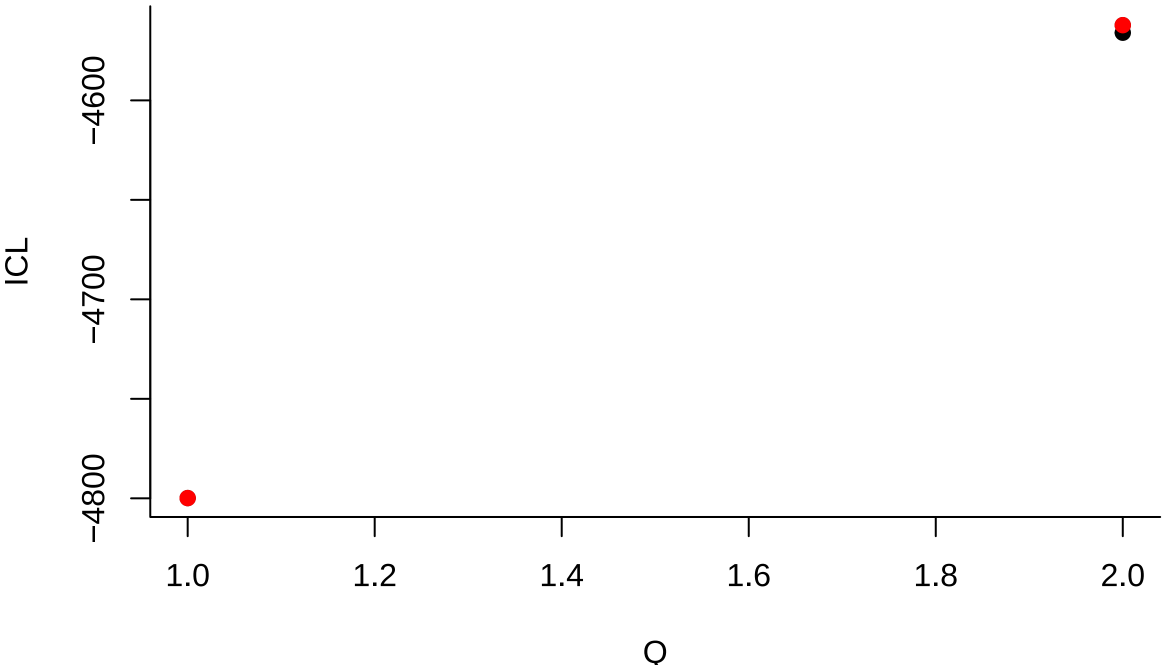


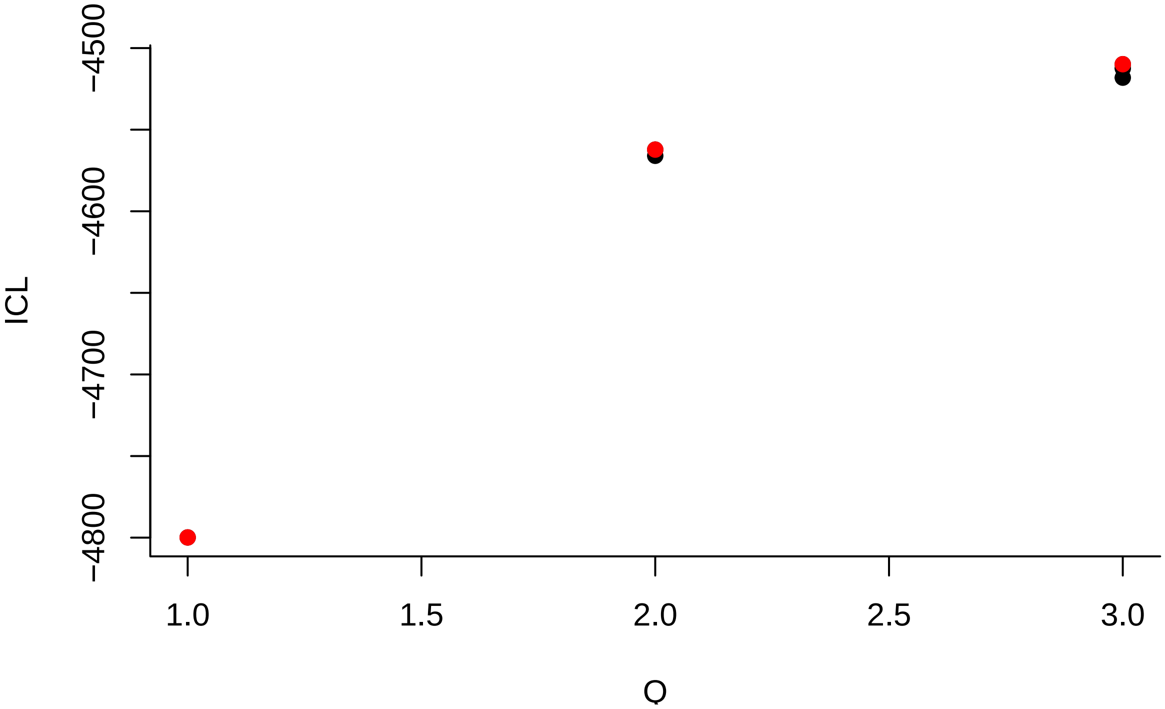


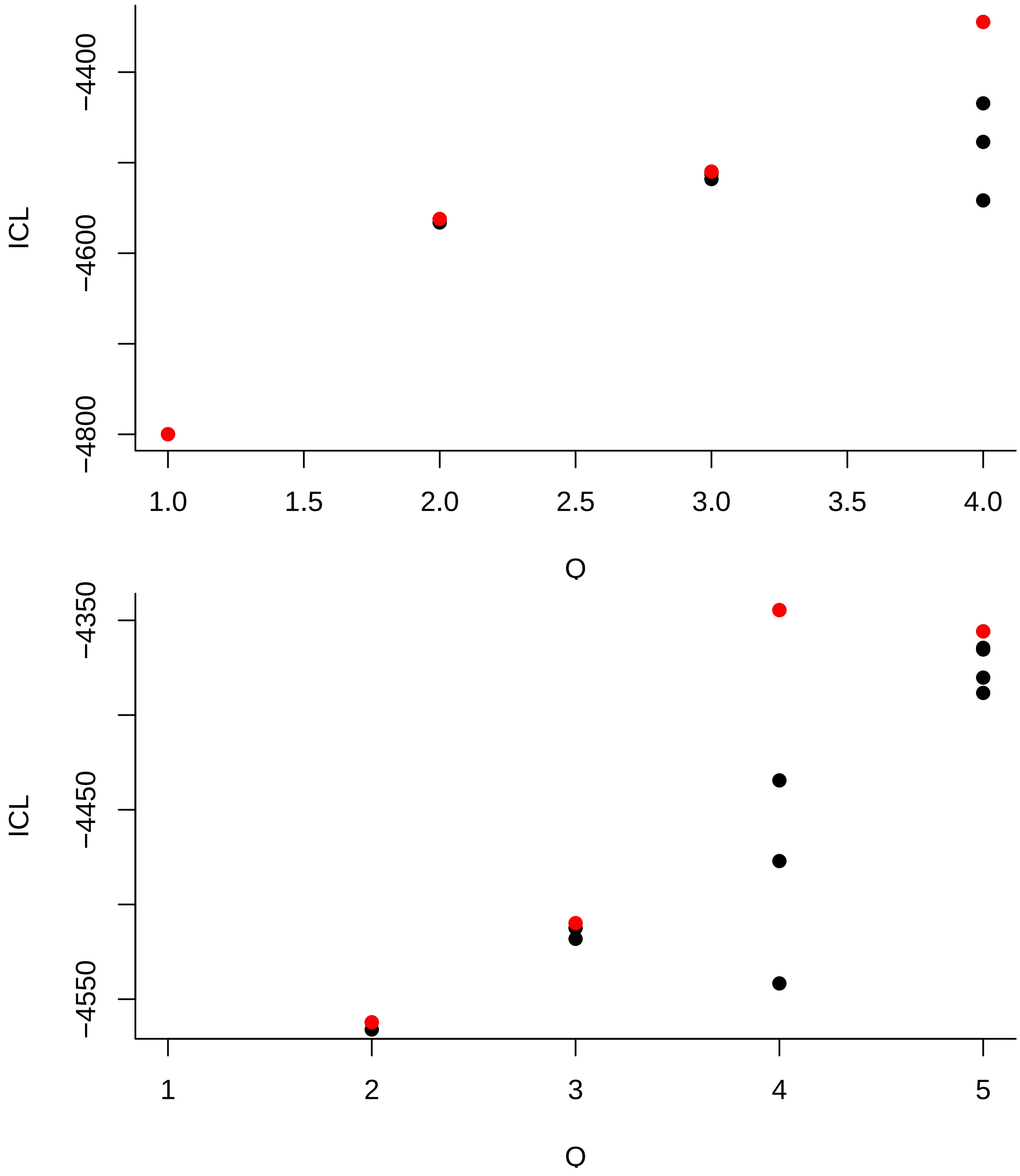
# Bernoulli Block Model

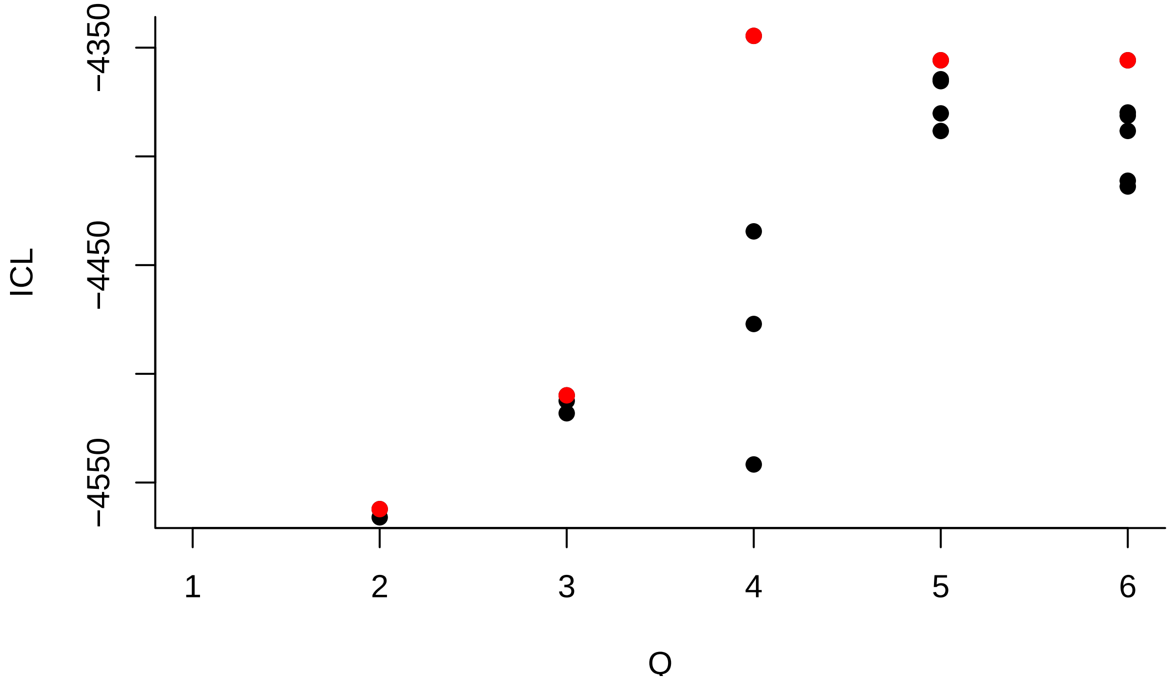
```
my_model<-function(x){
  BM_bernoulli("SBM",x,
verbosity=6,
autosave='',
plotting=character(0),
exploration_factor=1.5,
explore_min=1,
explore_max=7,
ncores=detectCores())}

my_model.period1<-my_model(adj.matrix.1)
#my_model.period2<-my_model(adj.matrix.2)
my_model.period3<-my_model(adj.matrix.3)

estimate_group<-function(x){
  x$estimate()
  which.max(x$ICL)
}
estimate.period1<-estimate_group(my_model.period1)
```

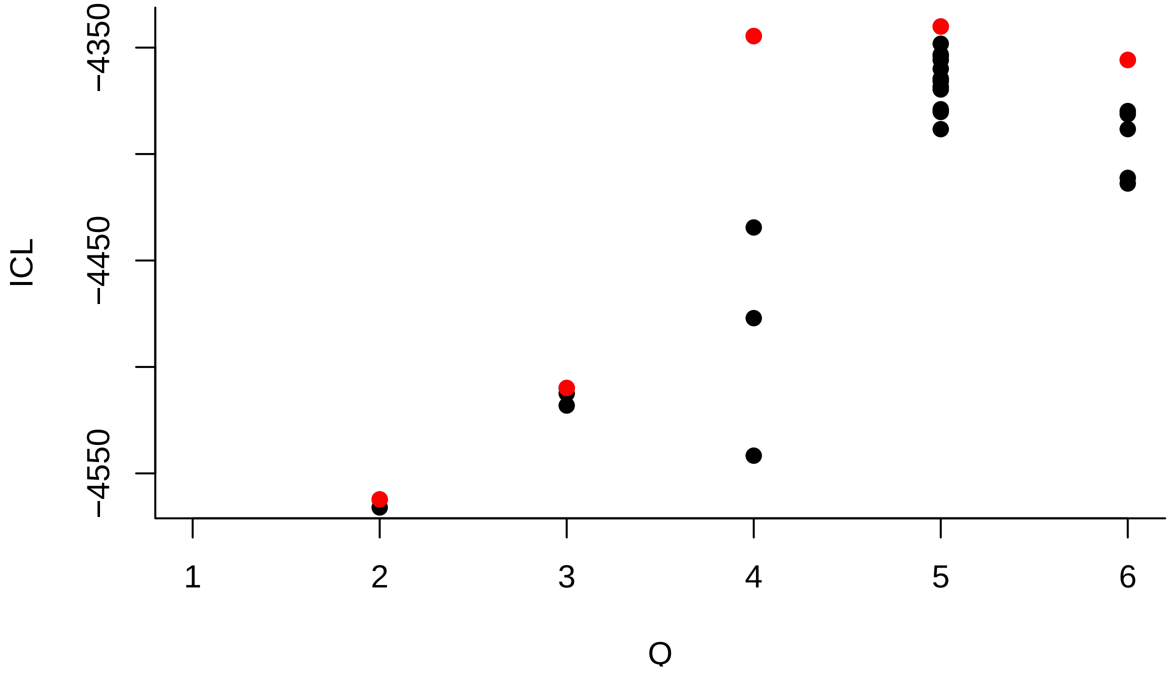


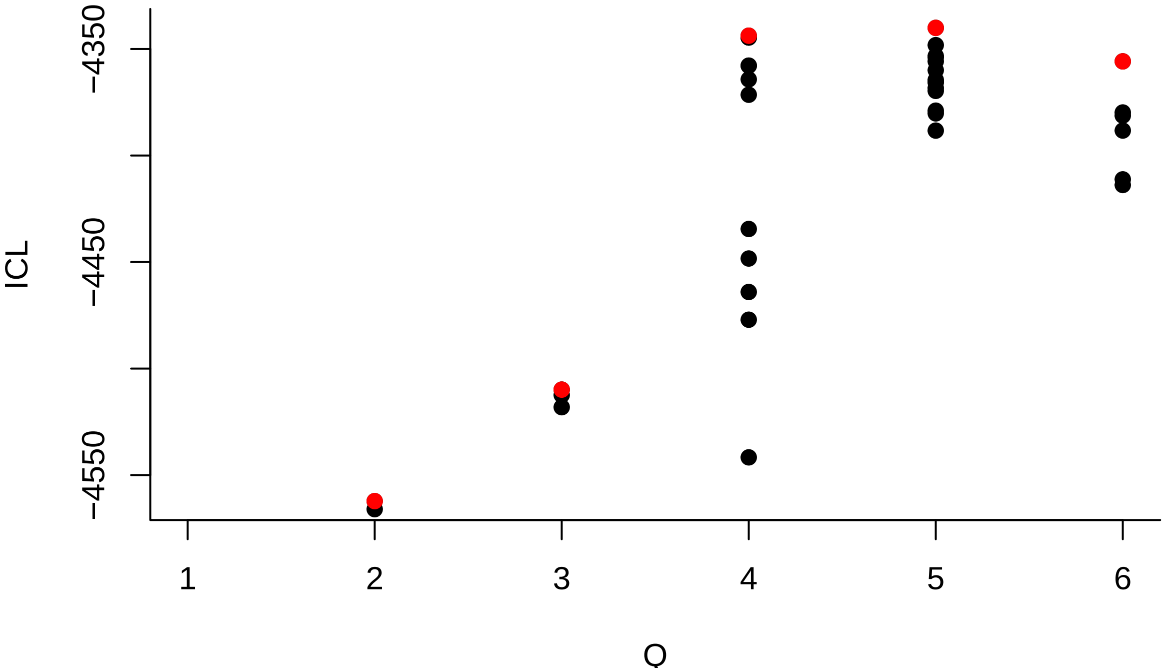

ICL
−4600
−4700
−4800
1.0
1.2
1.4
1.6
1.8
2.0
Q

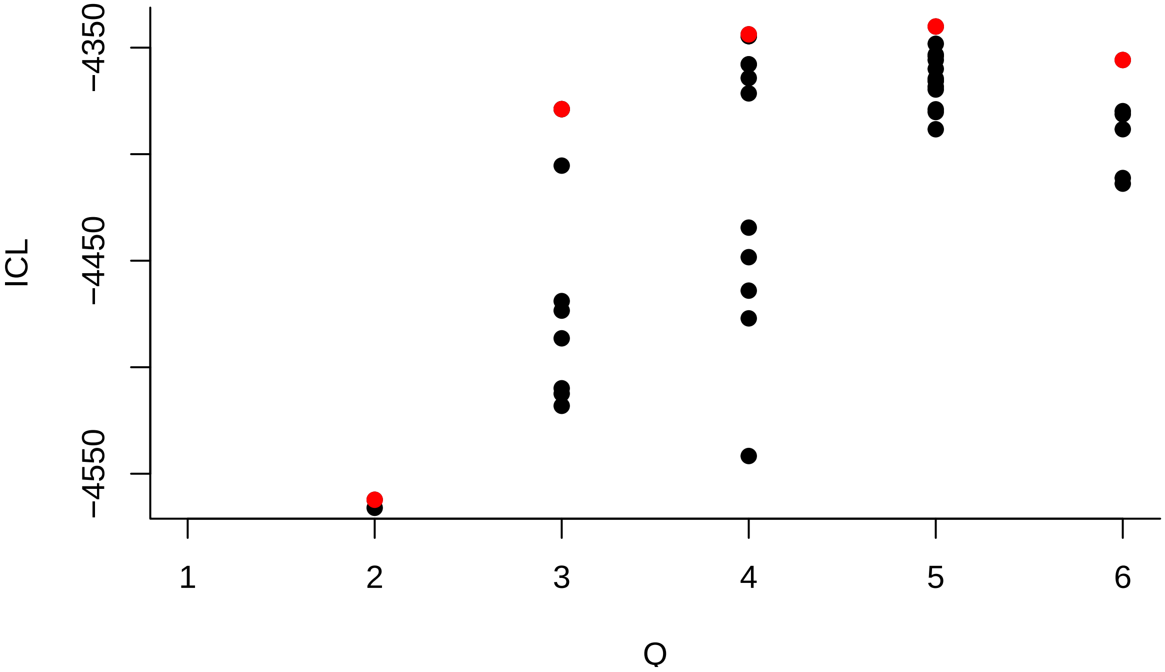

ICL
−4500
−4600
−4700
−4800
1.0
1.5
2.0
2.5
3.0
Q

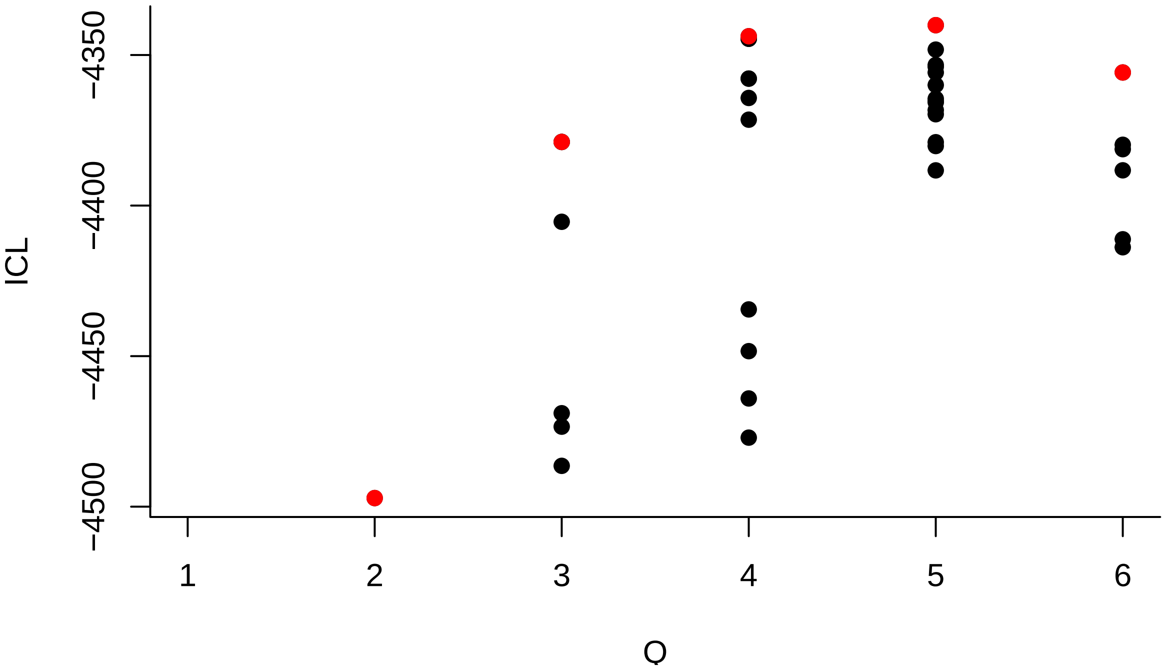

ICL
−4400
−4600
−4800
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Q
ICL
−4350
−4450
−4550
1
2
3
4
5
Q

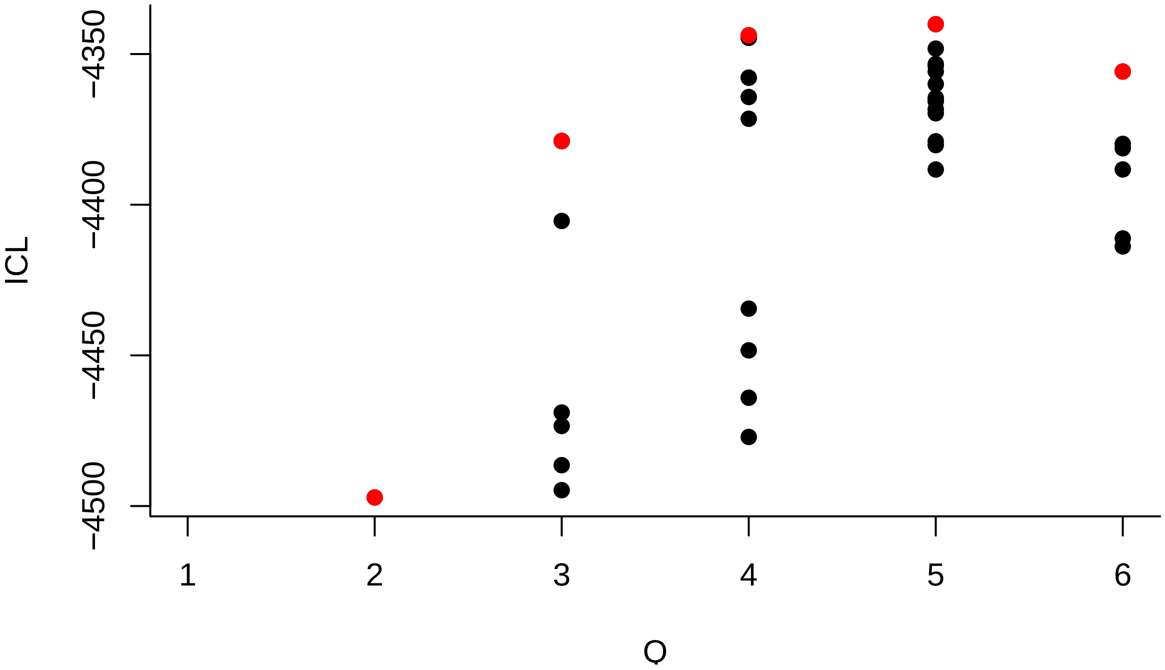

ICL
−4350
−4450
−4550
1
2
3
4
5
6
Q

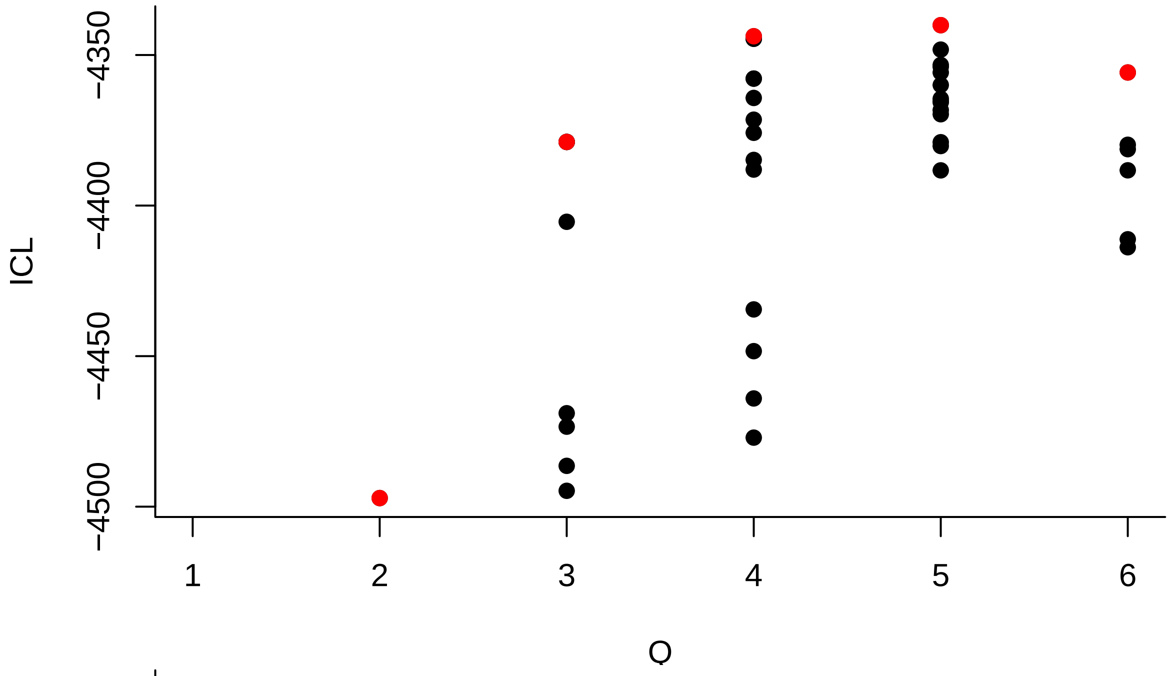

ICL
−4350
−4450
−4550
1
2
3
4
5
6
Q

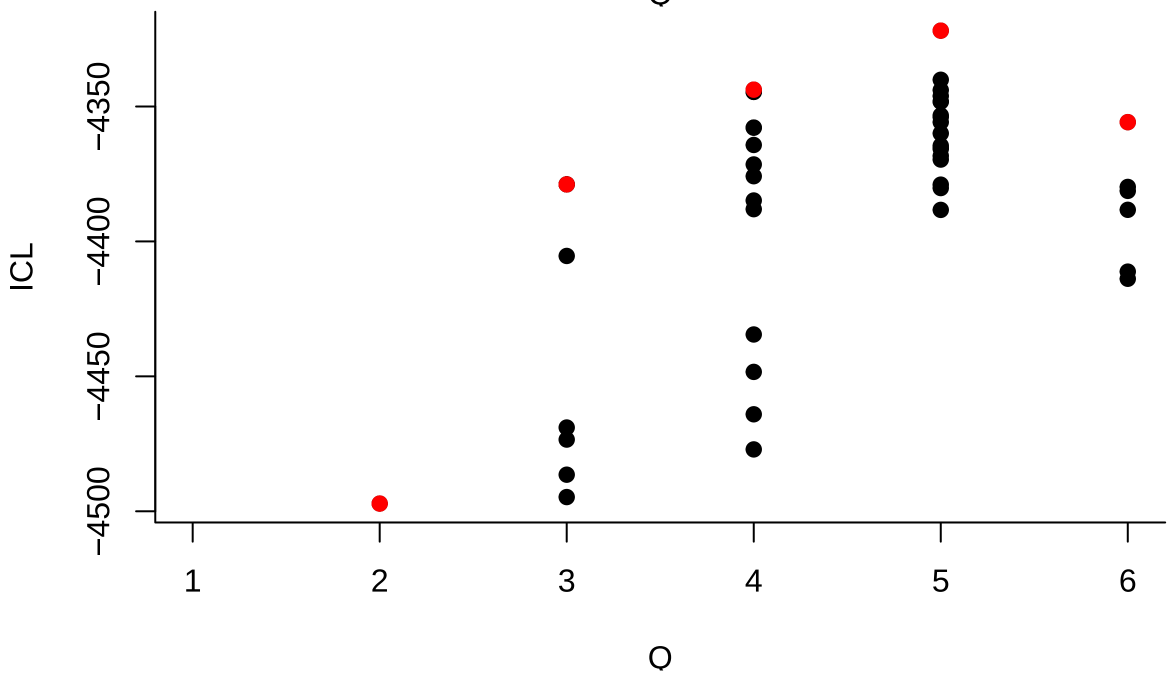

ICL
−4350
−4450
−4550
1
2
3
4
5
6
Q

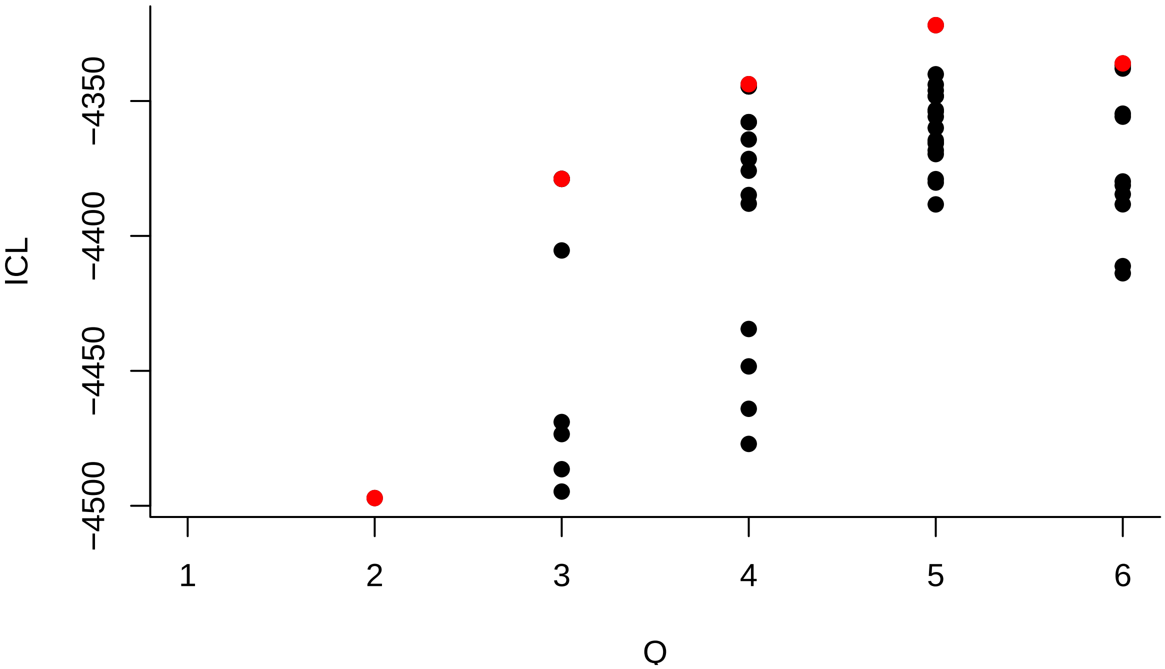

ICL
−4350
−4450
−4550
1
2
3
4
5
6
Q

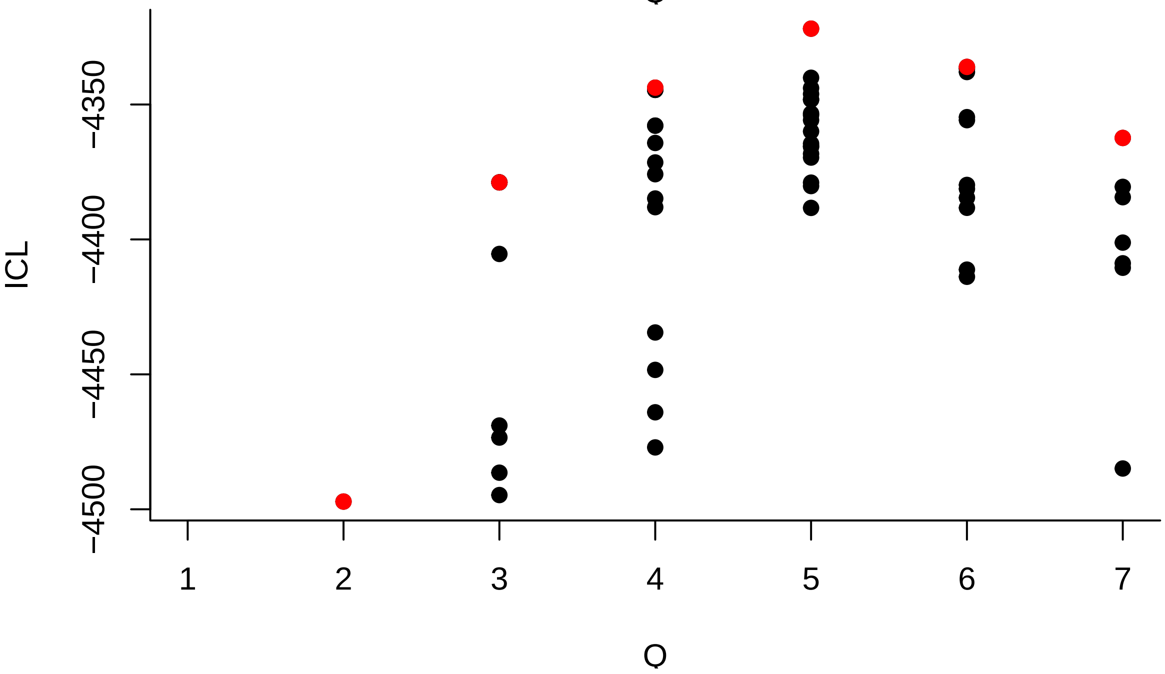

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
Q

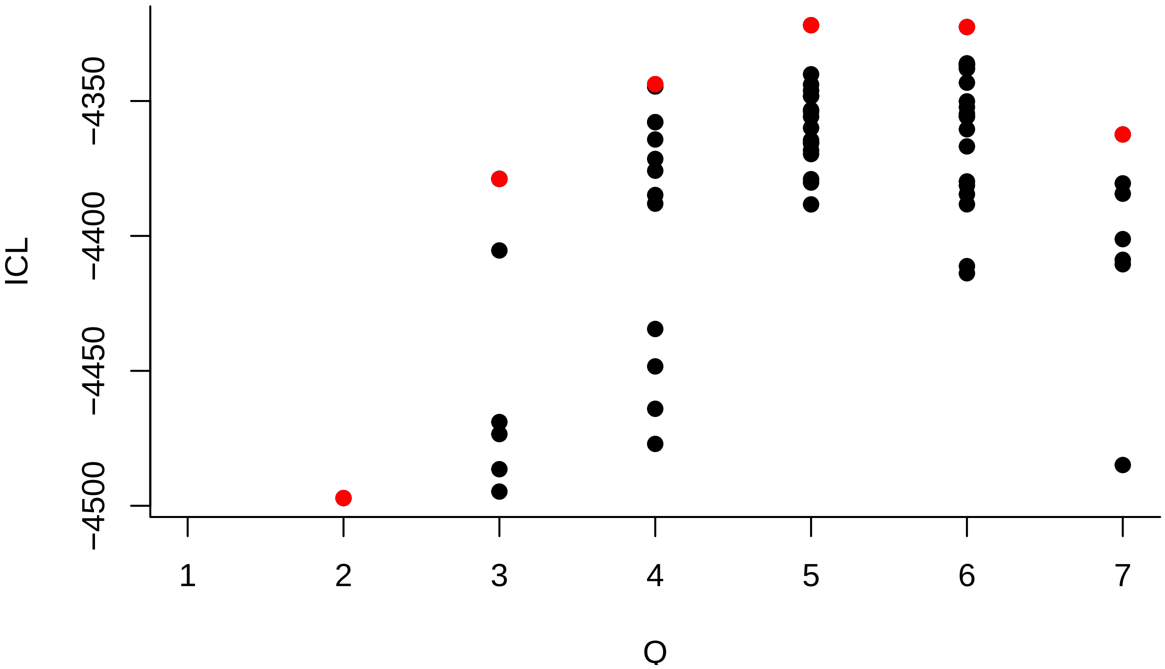

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
Q

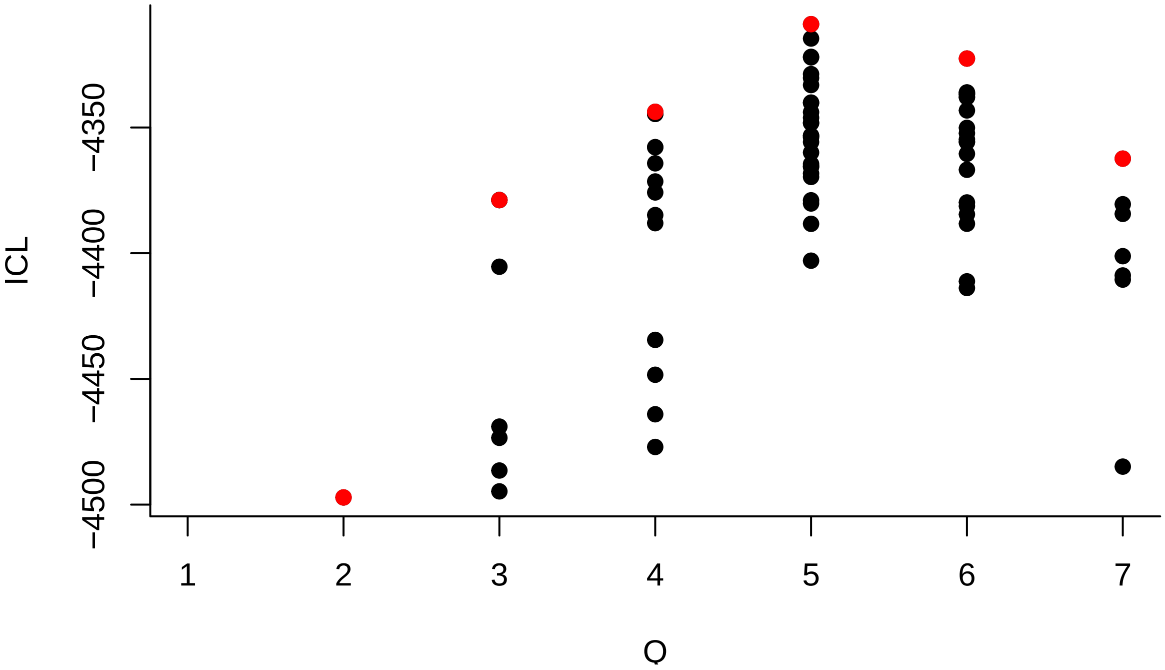

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
Q

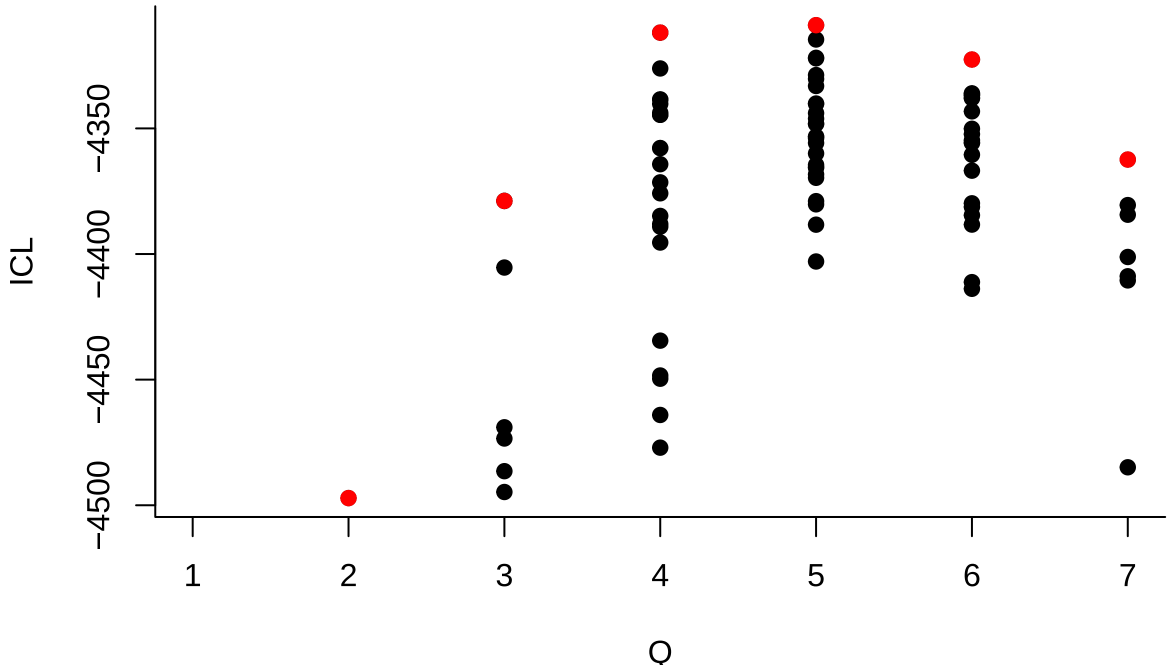

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
Q

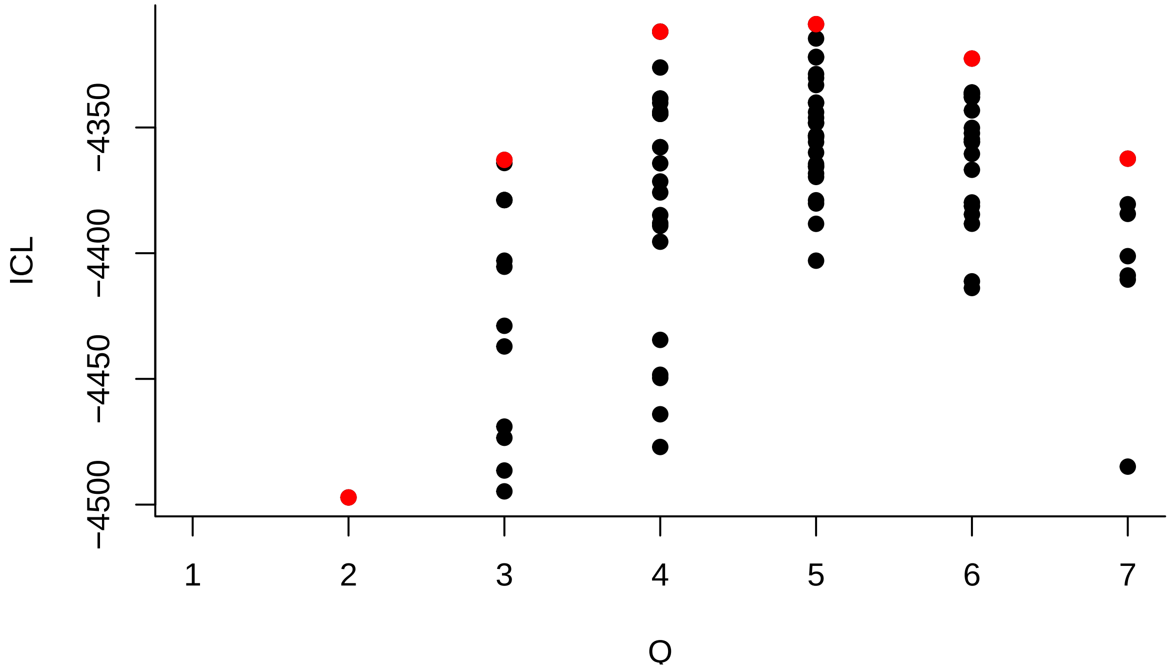

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
Q

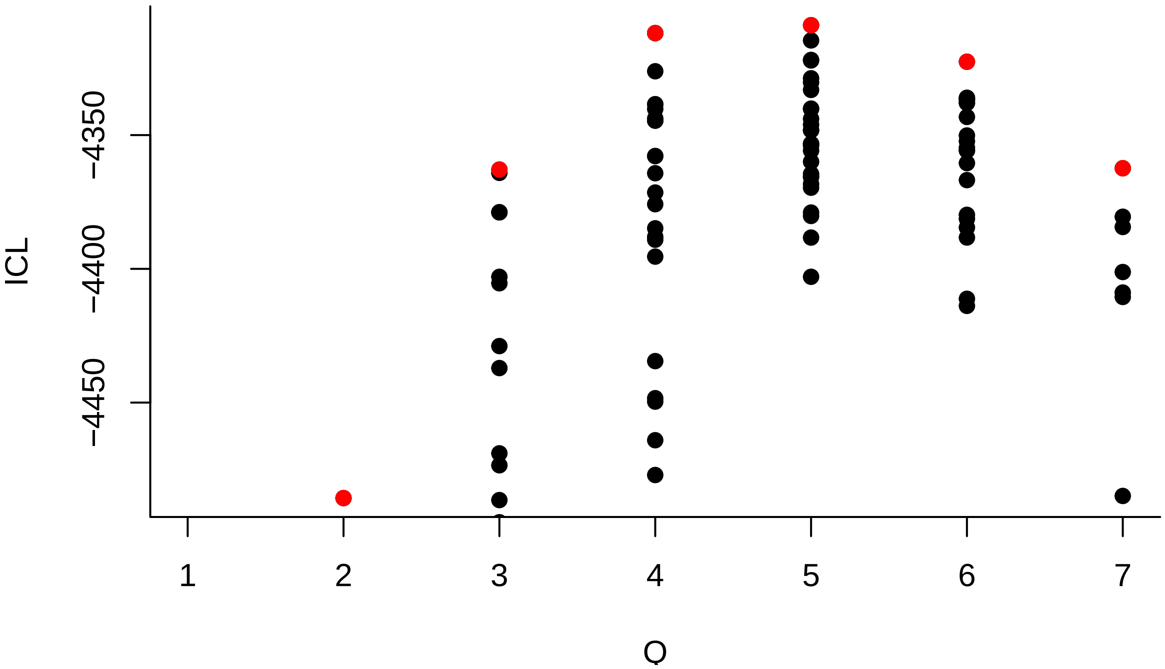

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
7
Q

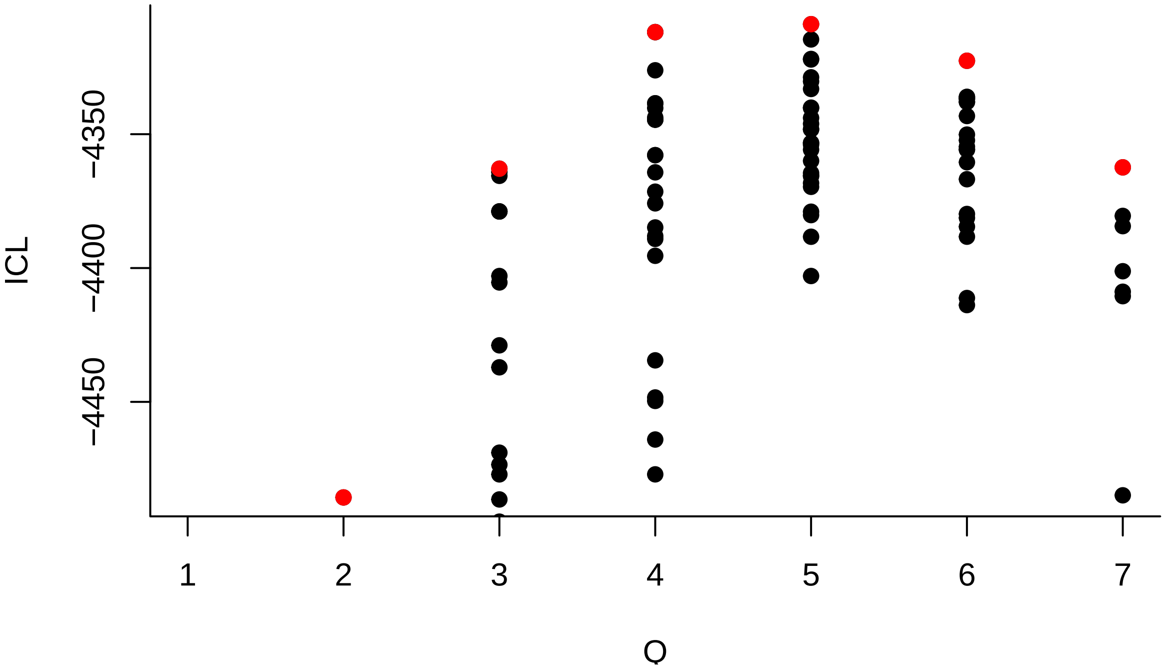

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
7
Q

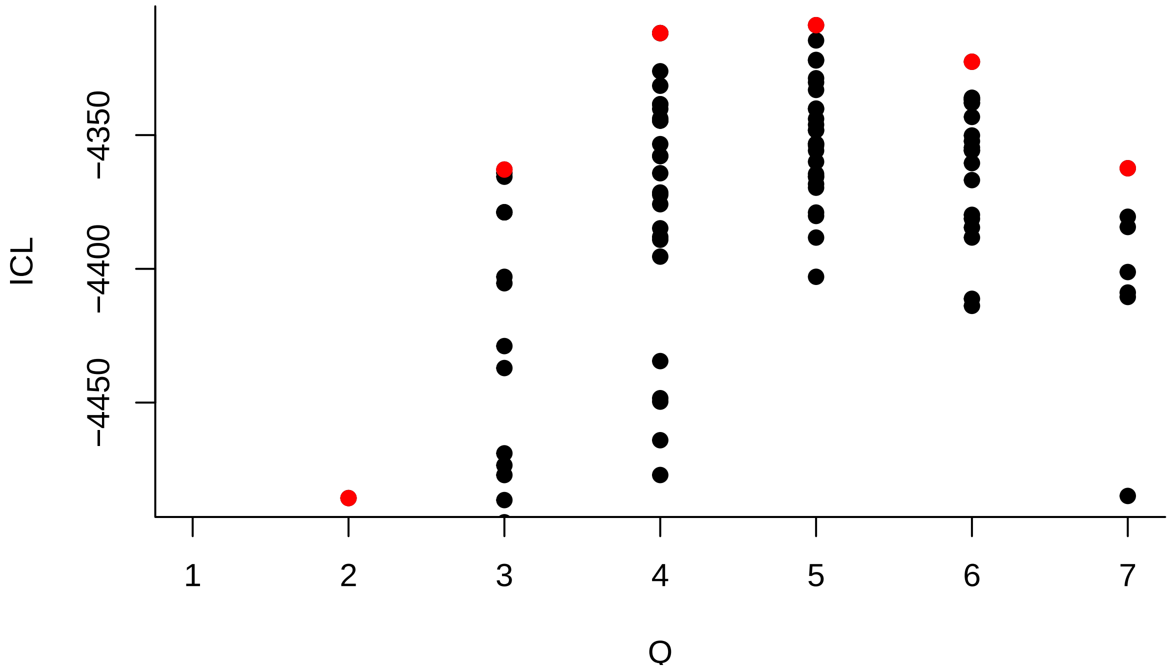

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
7
Q

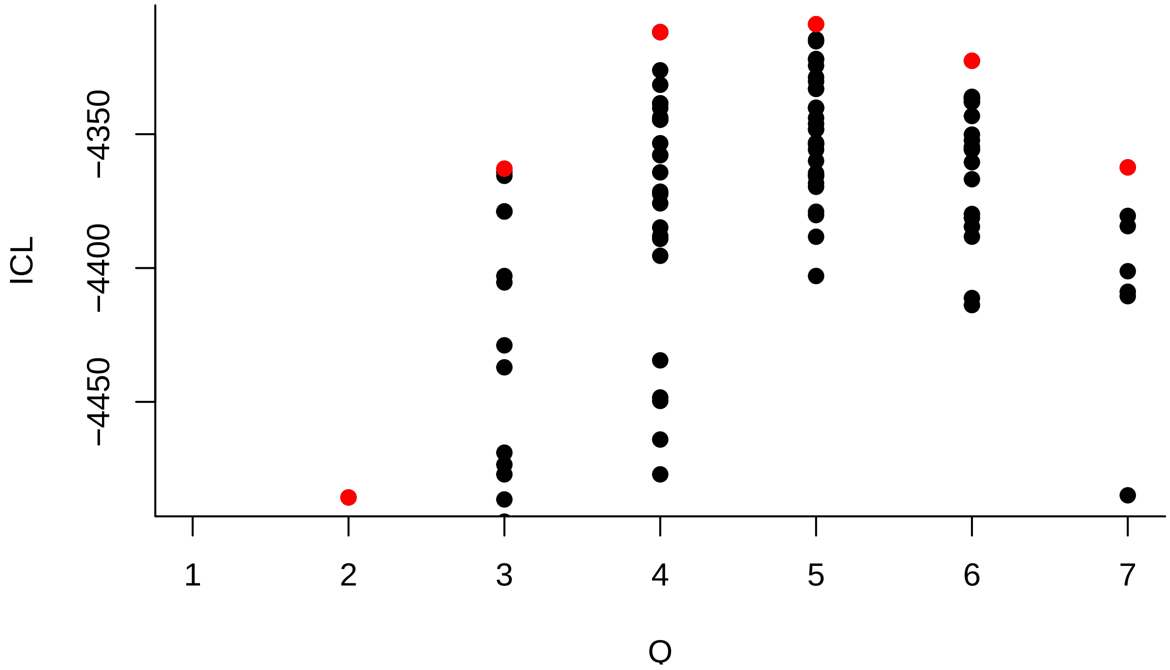

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
7
Q

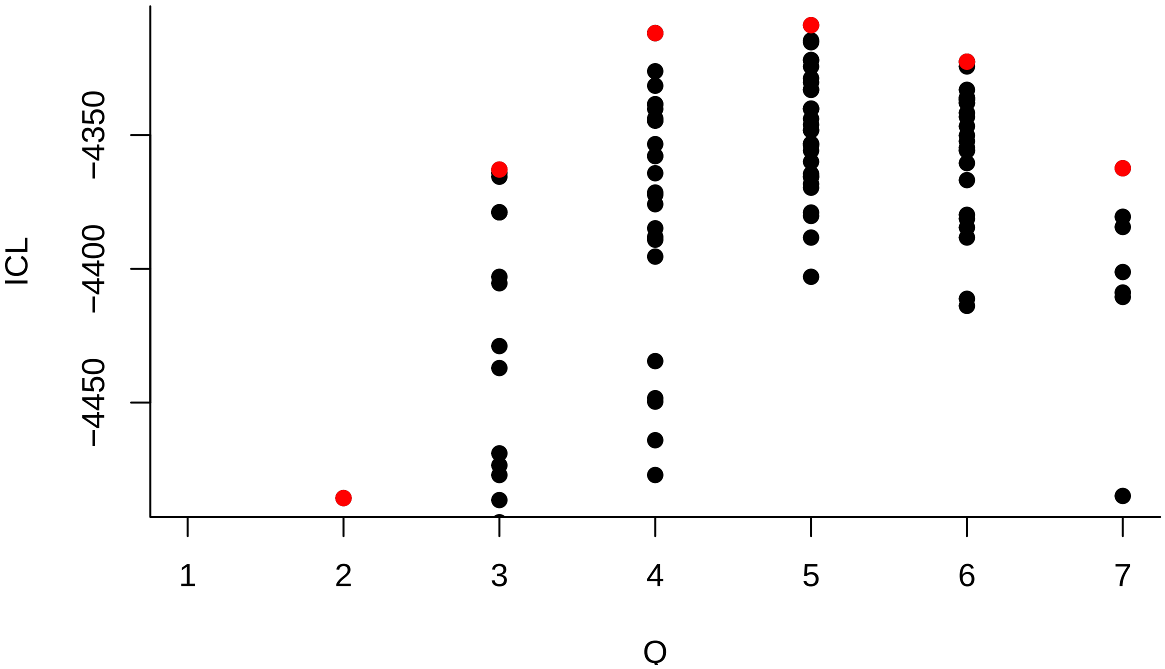

ICL
−4350
−4400
−4450
−4500
1
2
3
4
5
6
7
Q

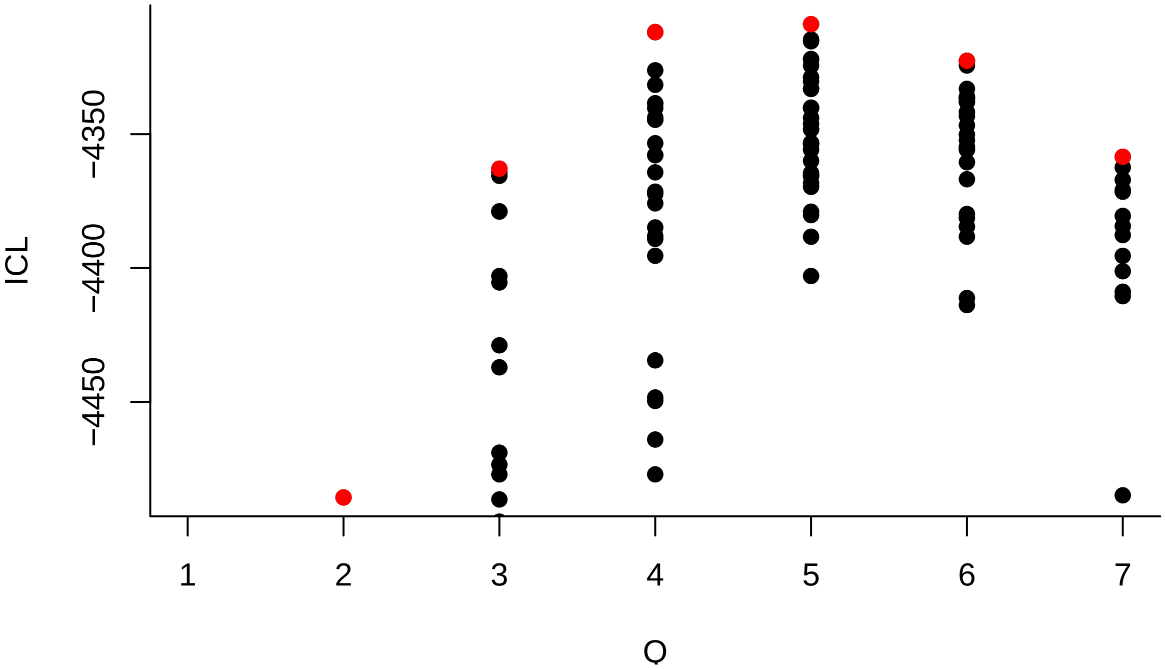

ICL
−4350
−4400
−4450
1
2
3
4
5
6
7
Q

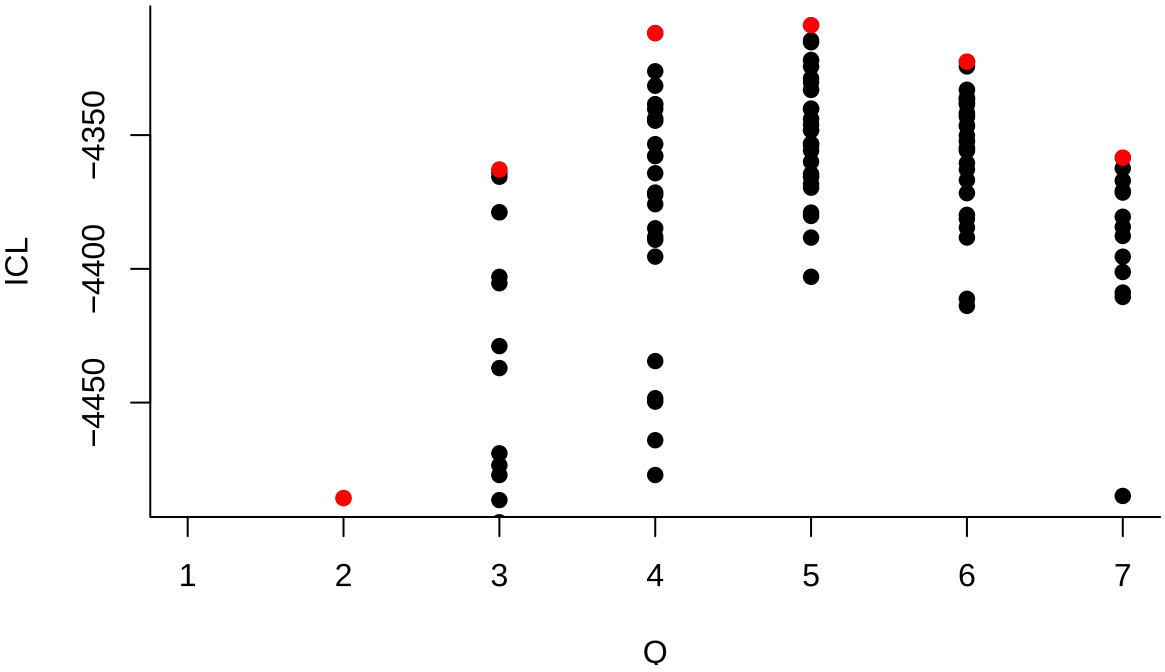

ICL
−4350
−4400
−4450
1
2
3
4
5
6
7
Q


ICL
−4350
−4400
−4450
1
2
3
4
5
6
7
Q

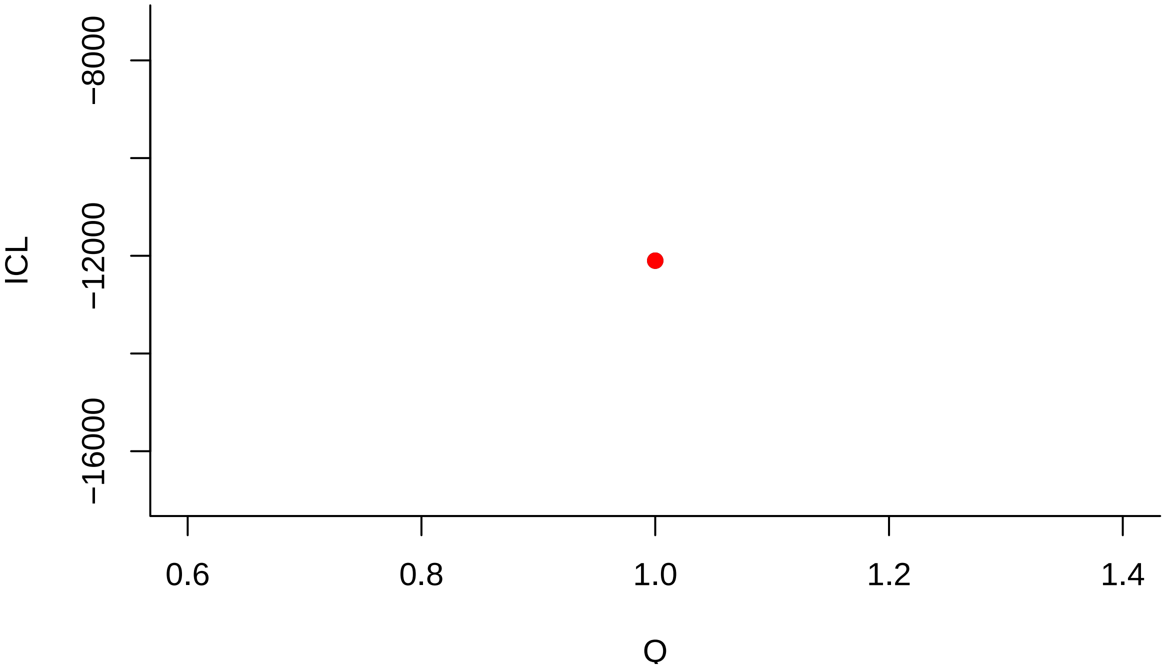

ICL
−4350
−4400
−4450
1
2
3
4
5
6
7
Q

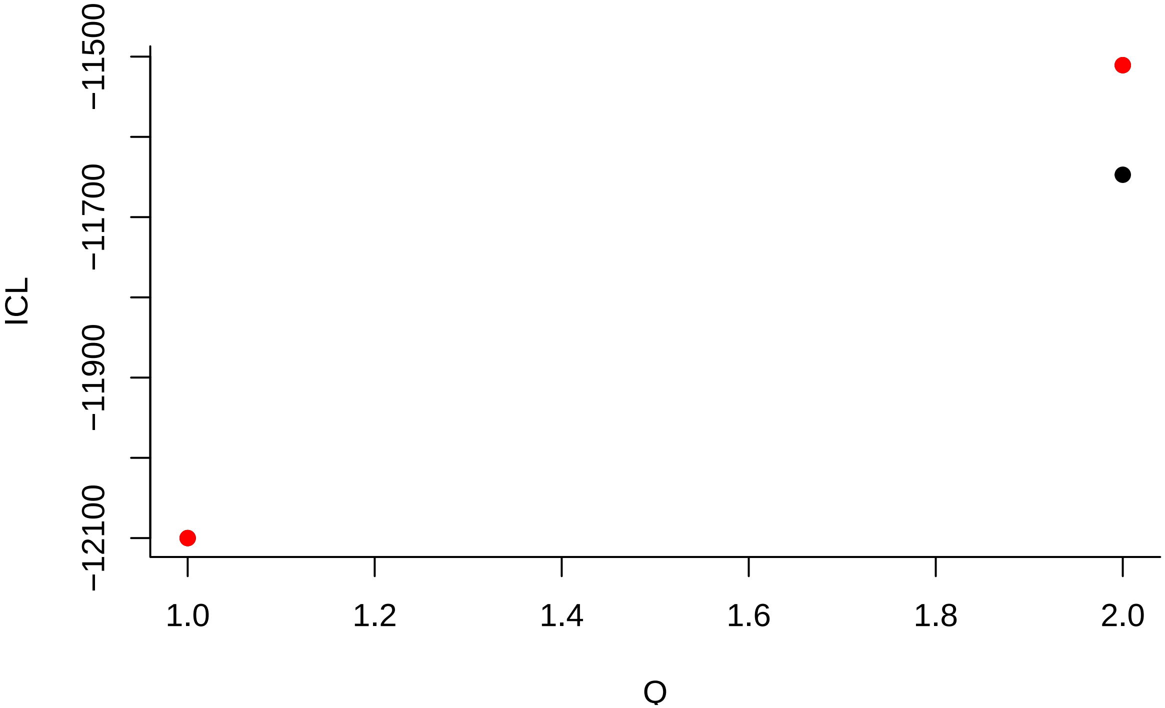

ICL
−4350
−4400
−4450
1
2
3
4
5
6
7
Q

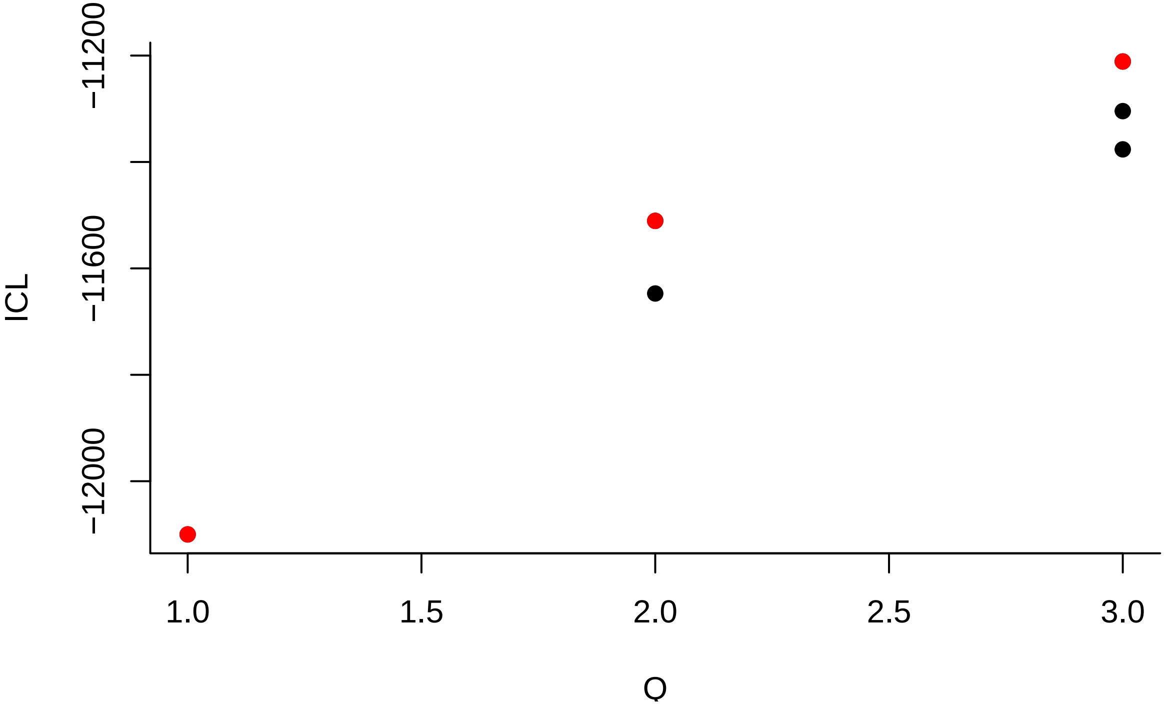

ICL
−4350
−4400
−4450
1
2
3
4
5
6
7
Q

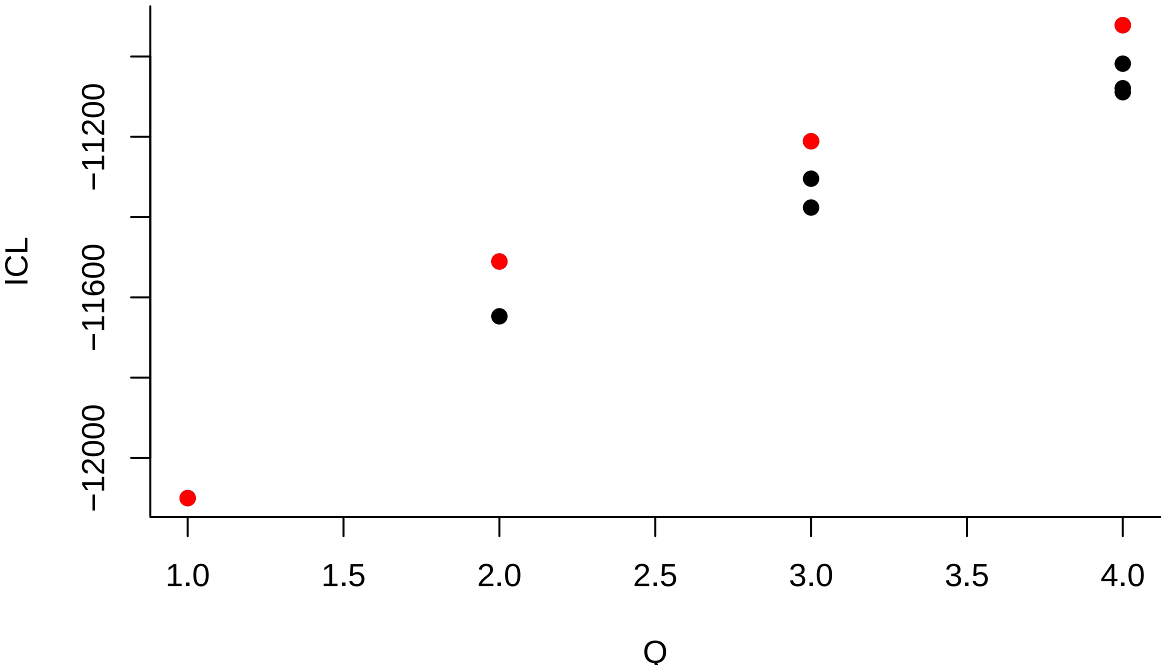


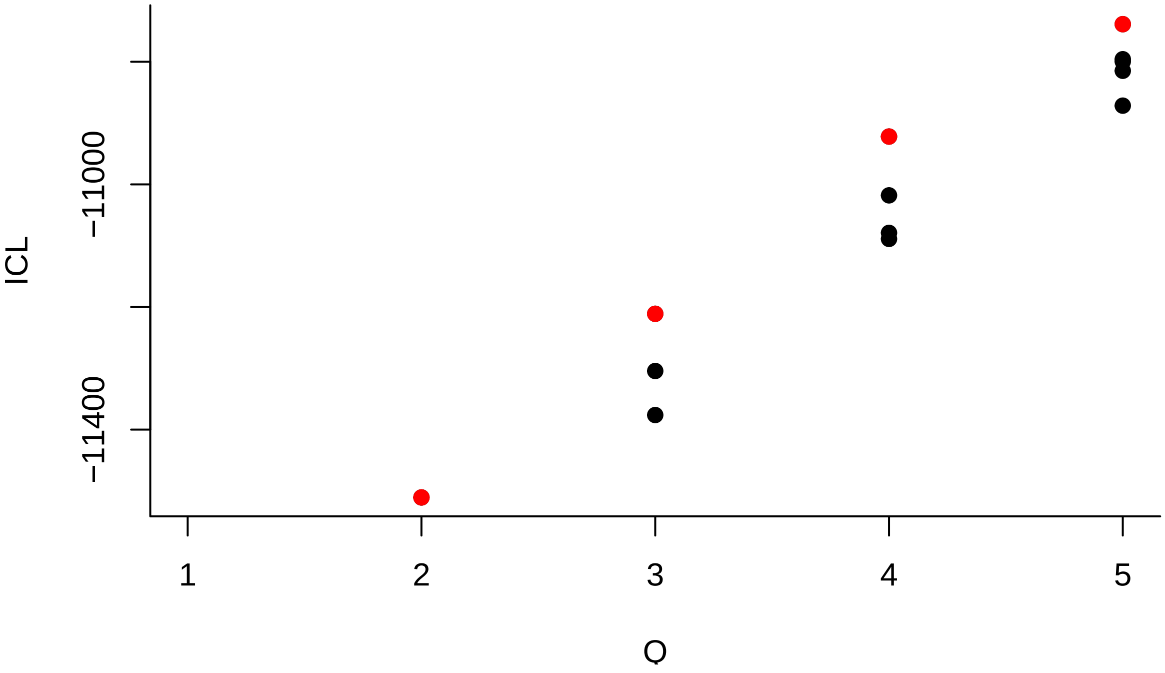
```
#estimate.period2<-estimate_group(my_model.period2)
estimate.period3<-estimate_group(my_model.period3)
```

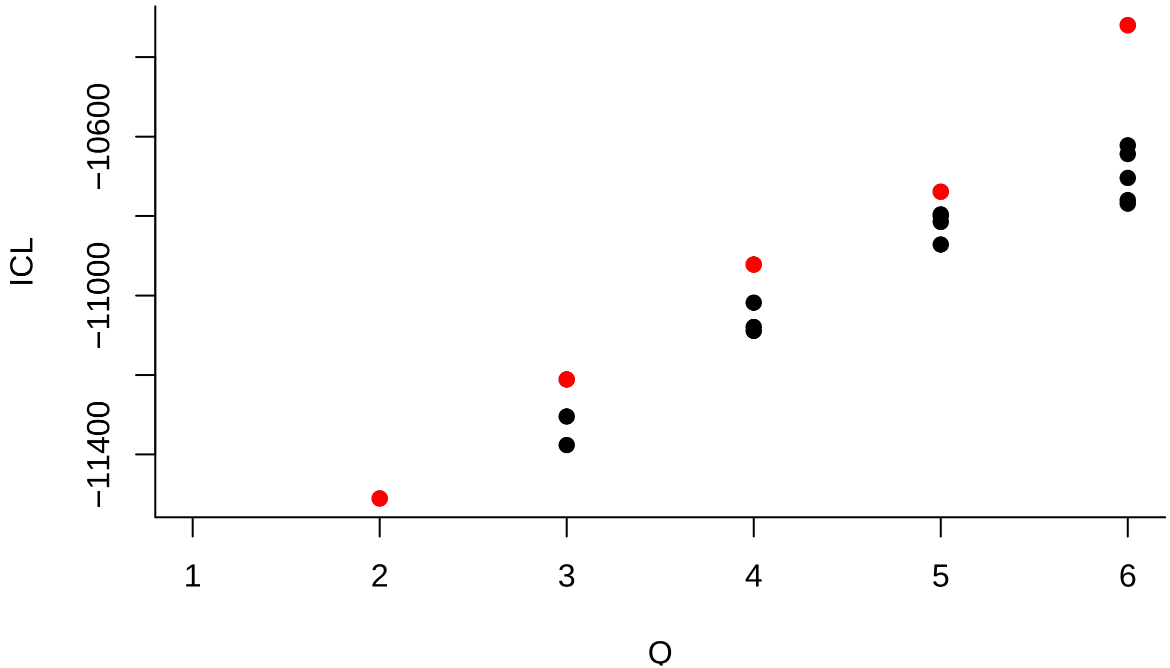


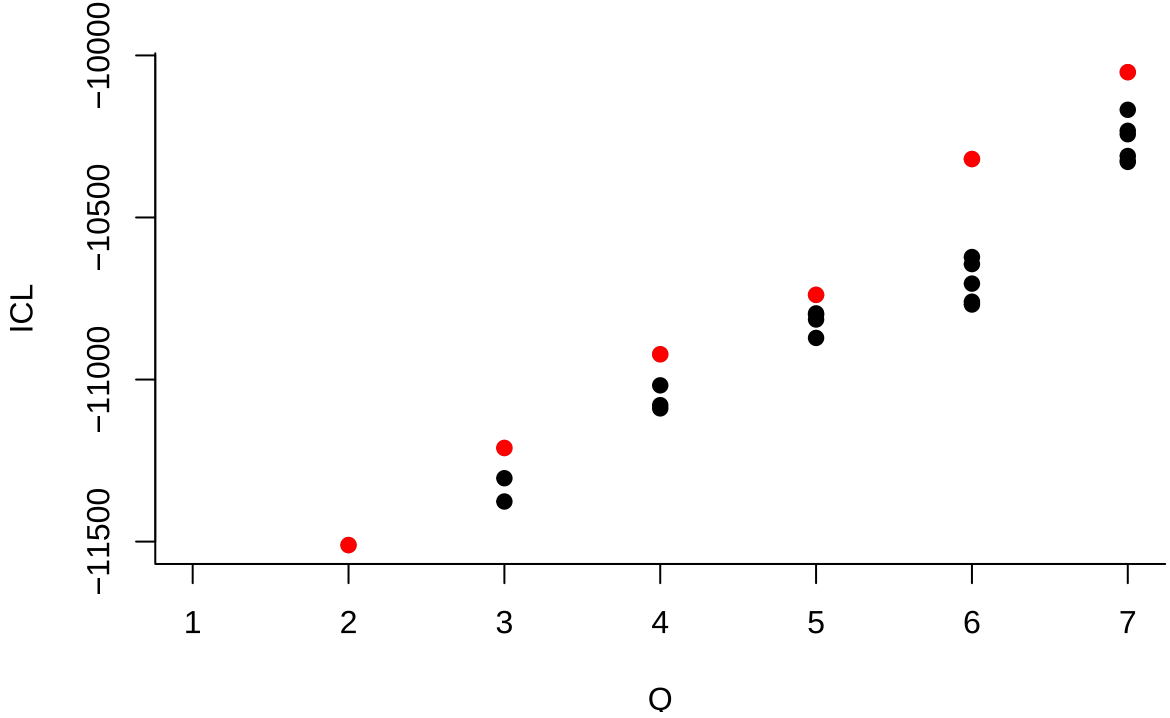

−11500
−11700
−11900
−12100
ICL
1.0
1.2
1.4
1.6
1.8
2.0
Q

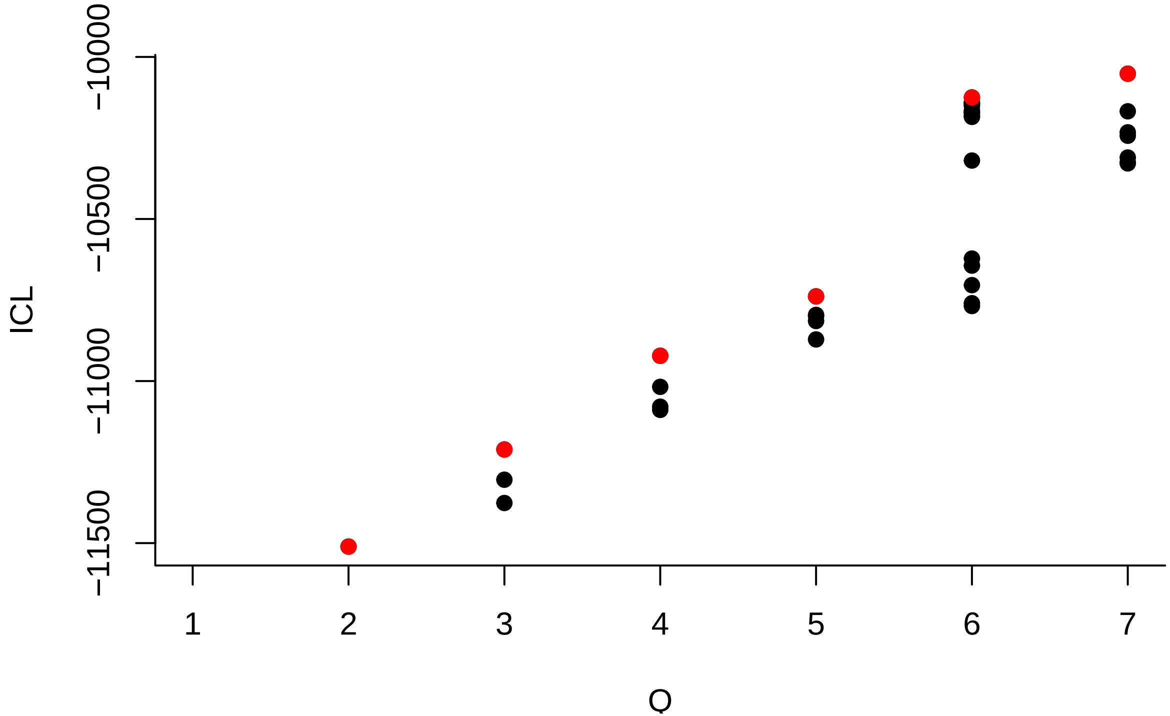

−11200
−11600
−12000
ICL
1.0
1.5
2.0
2.5
3.0
Q

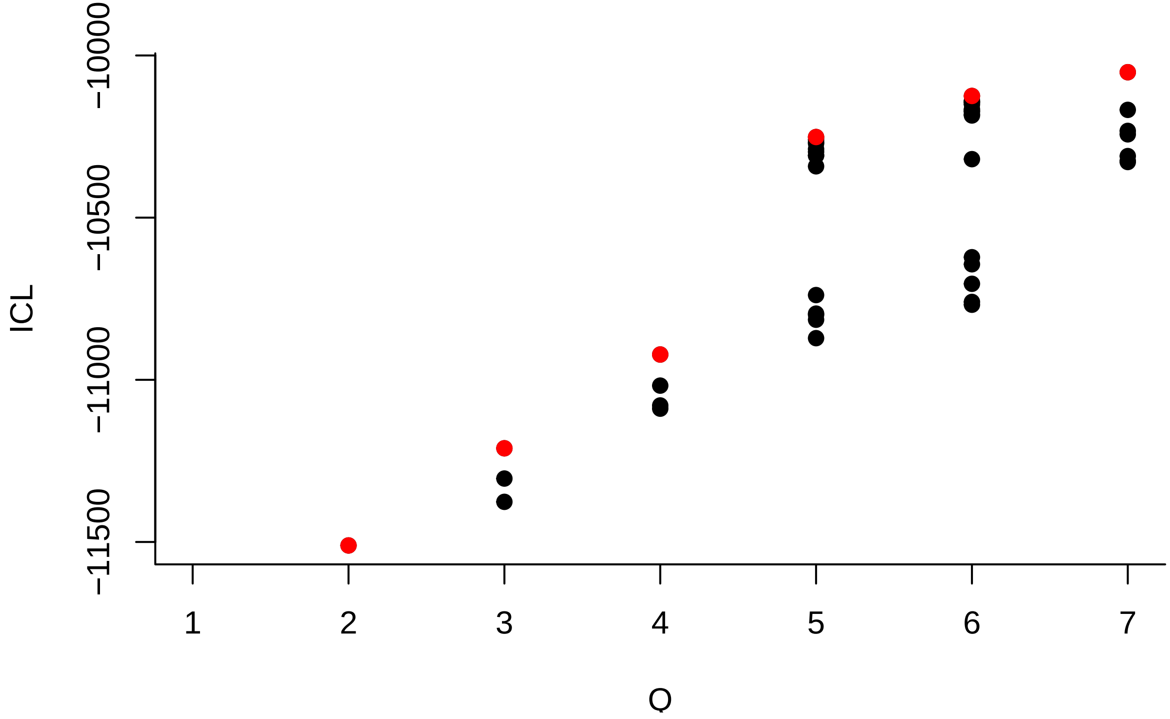

ICL
−12000
−11600
−11200
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Q

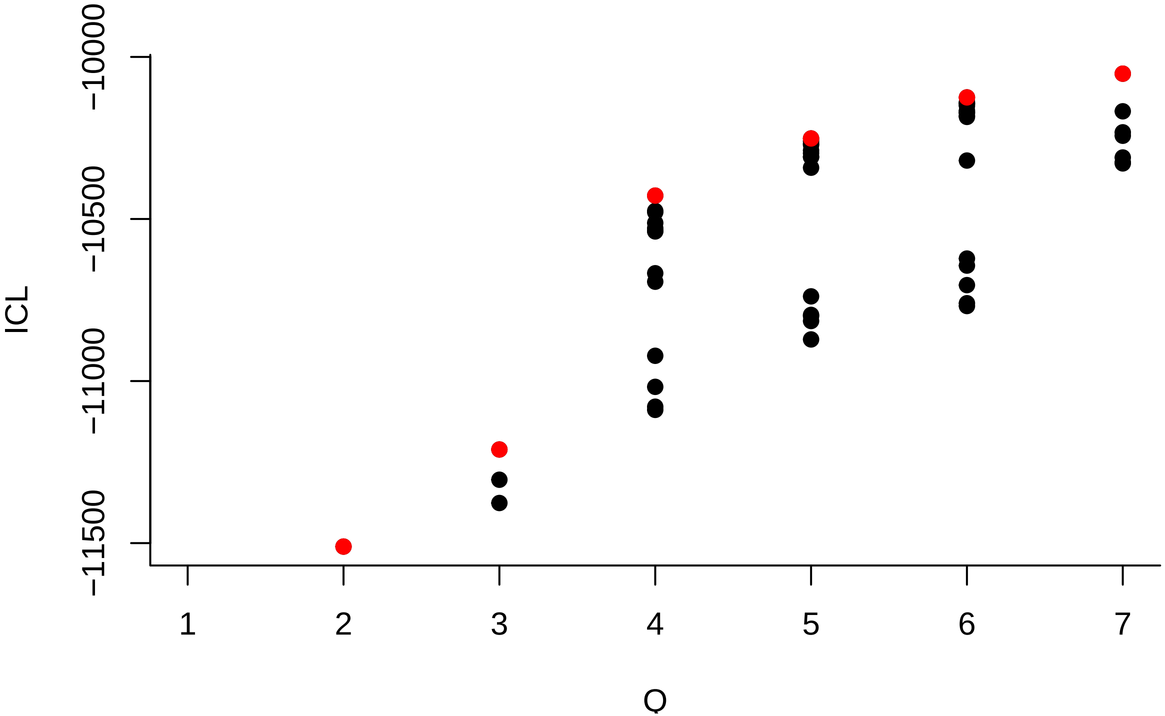

ICL
−11400
−11000
1
2
3
4
5
Q

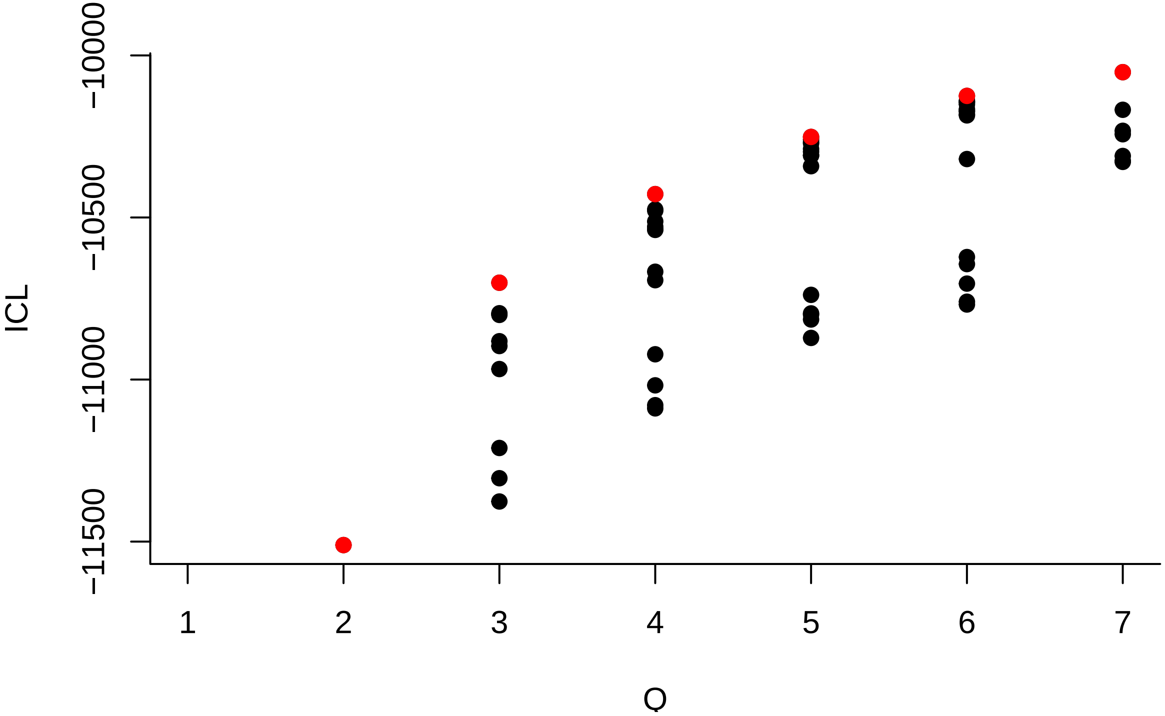

ICL
−10600
−11000
−11400
1
2
3
4
5
6
Q

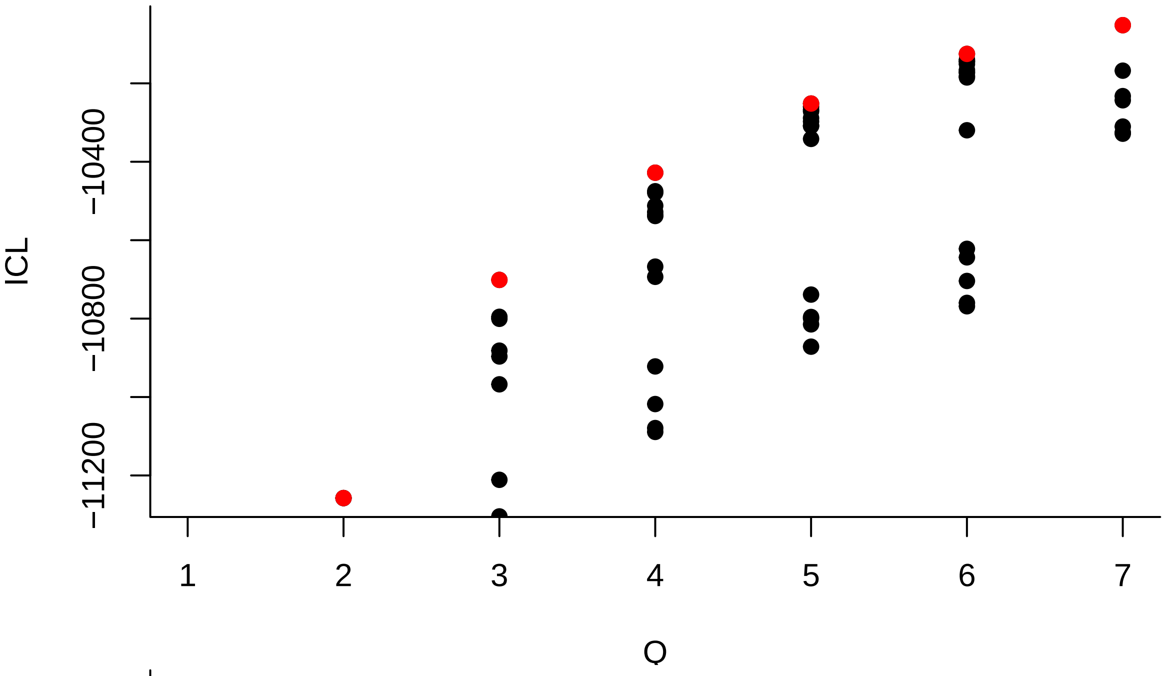

ICL
−10000
−10500
−11000
−11500
1
2
3
4
5
6
7
Q

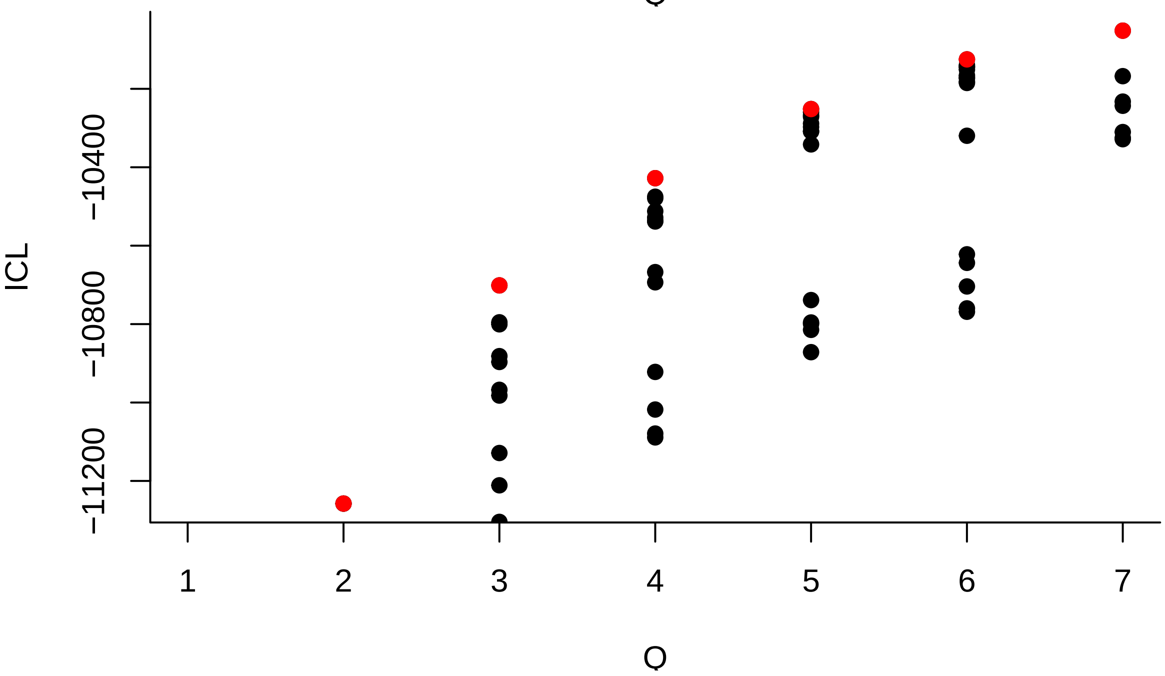

ICL
−10000
−10500
−11000
−11500
1
2
3
4
5
6
7
Q

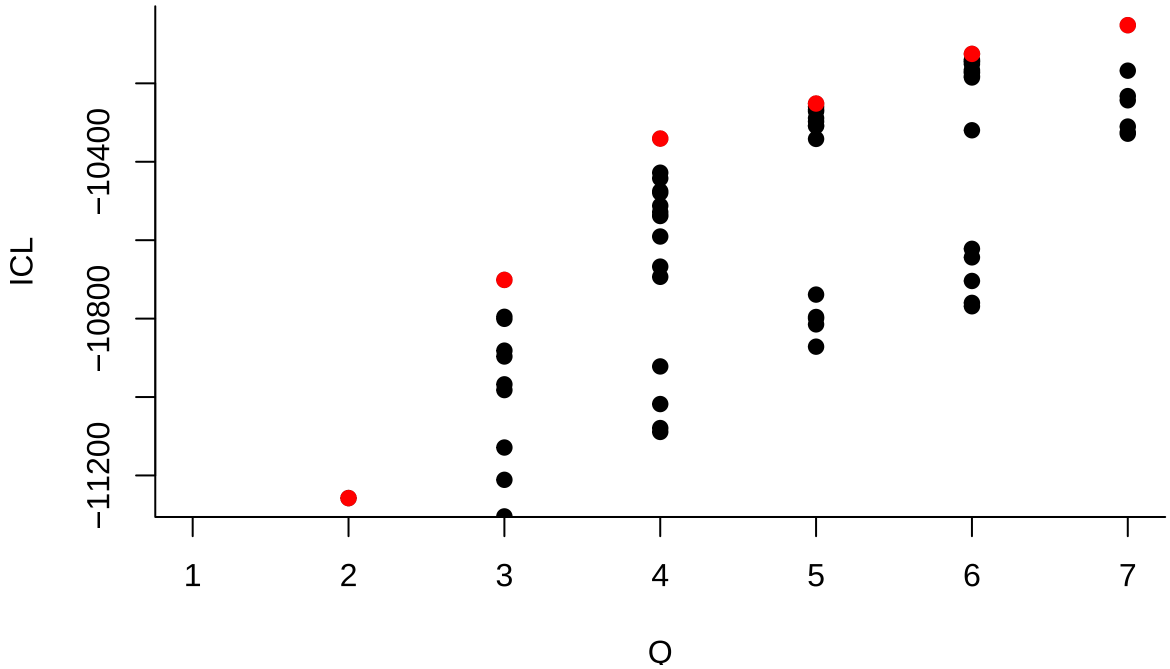

ICL
−10000
−10500
−11000
−11500
1
2
3
4
5
6
7
Q

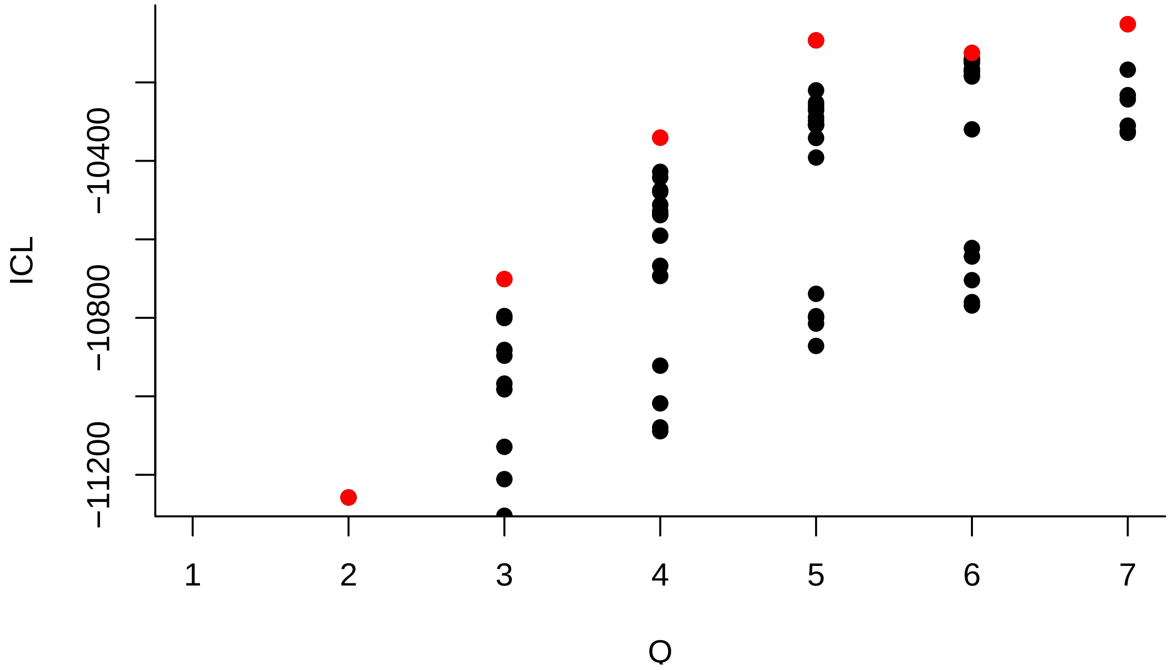

ICL
−10000
−10500
−11000
−11500
1
2
3
4
5
6
7
Q

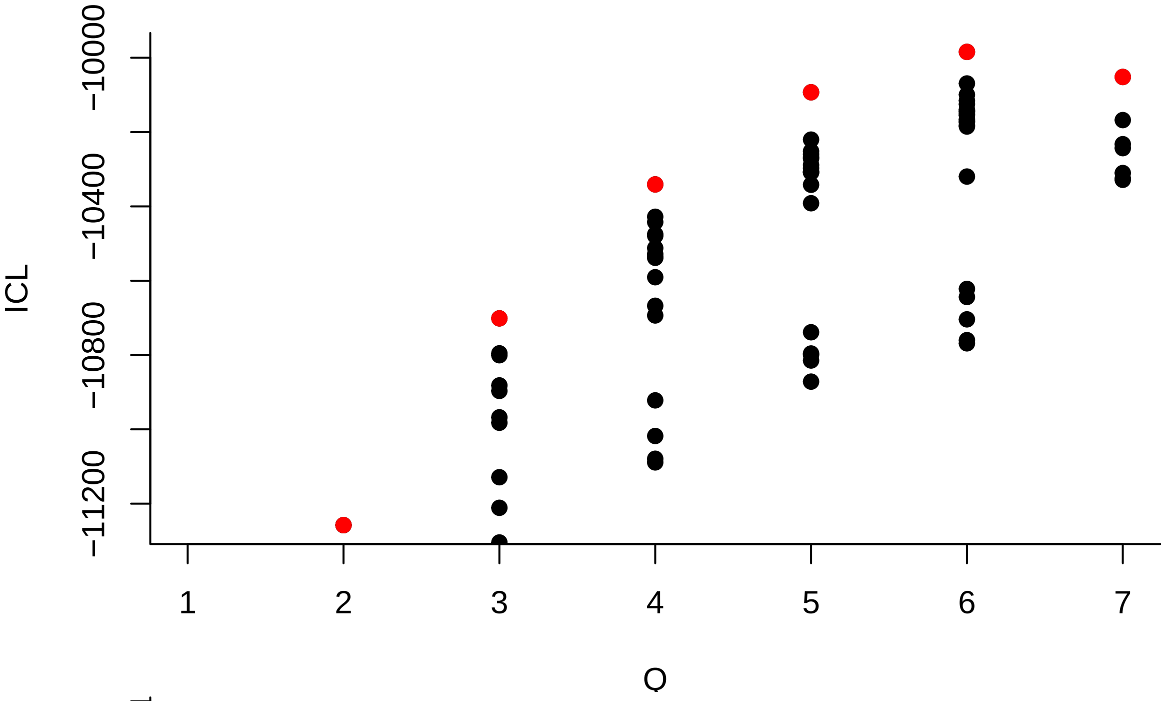

ICL
−10000
−10500
−11000
−11500
1
2
3
4
5
6
7
Q


ICL
−10400
−10800
−11200
1
2
3
4
5
6
7
Q

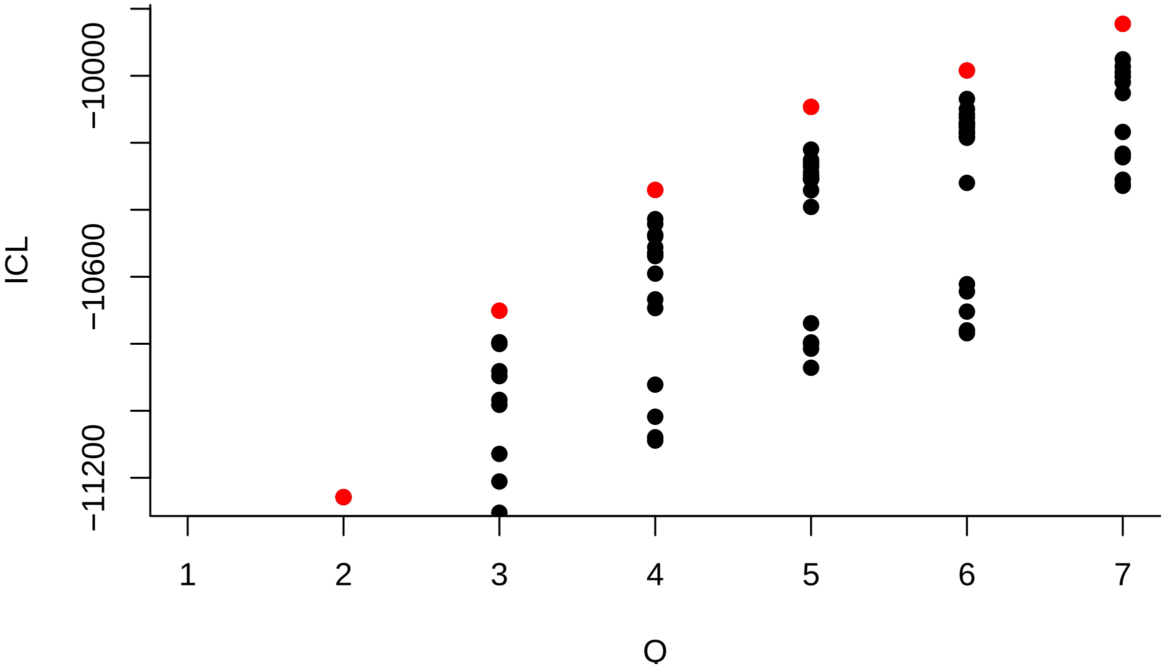

ICL
−10400
−10800
−11200
1
2
3
4
5
6
7
Q


ICL
−10400
−10800
−11200
1
2
3
4
5
6
7
Q

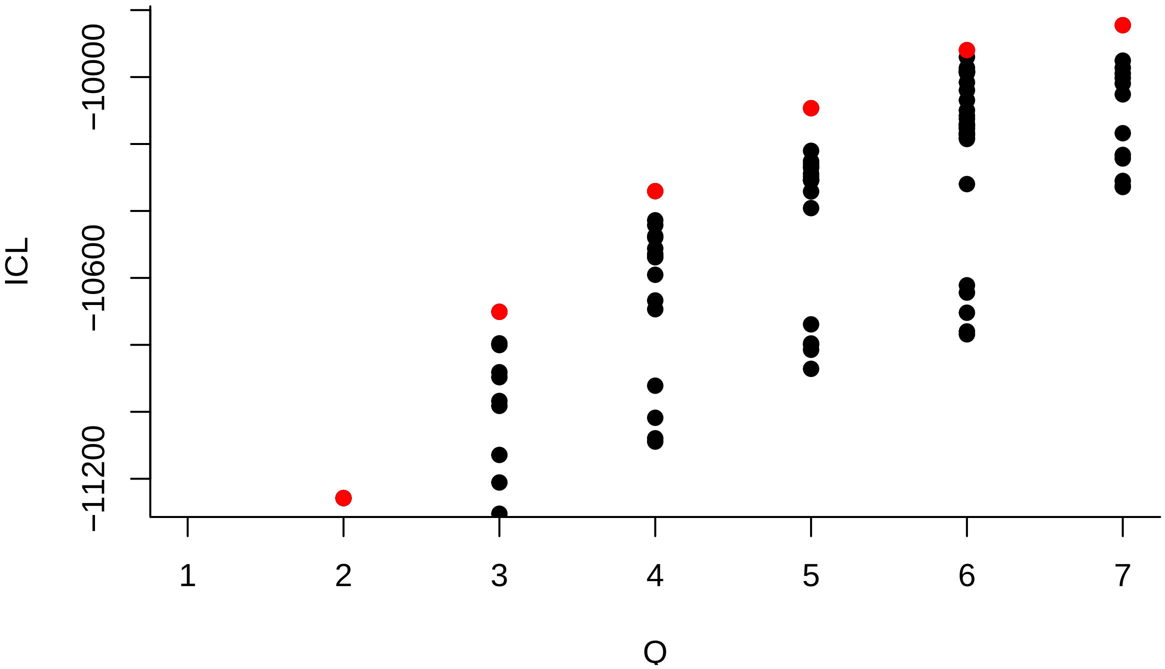

ICL
−10400
−10800
−11200
1
2
3
4
5
6
7
Q

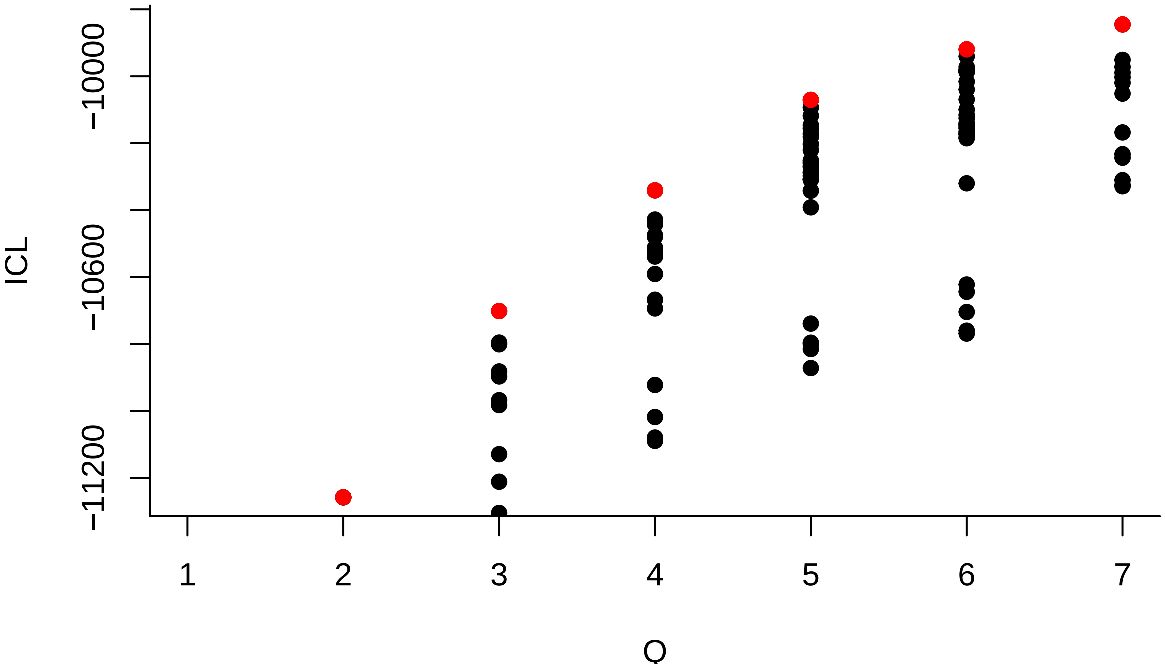


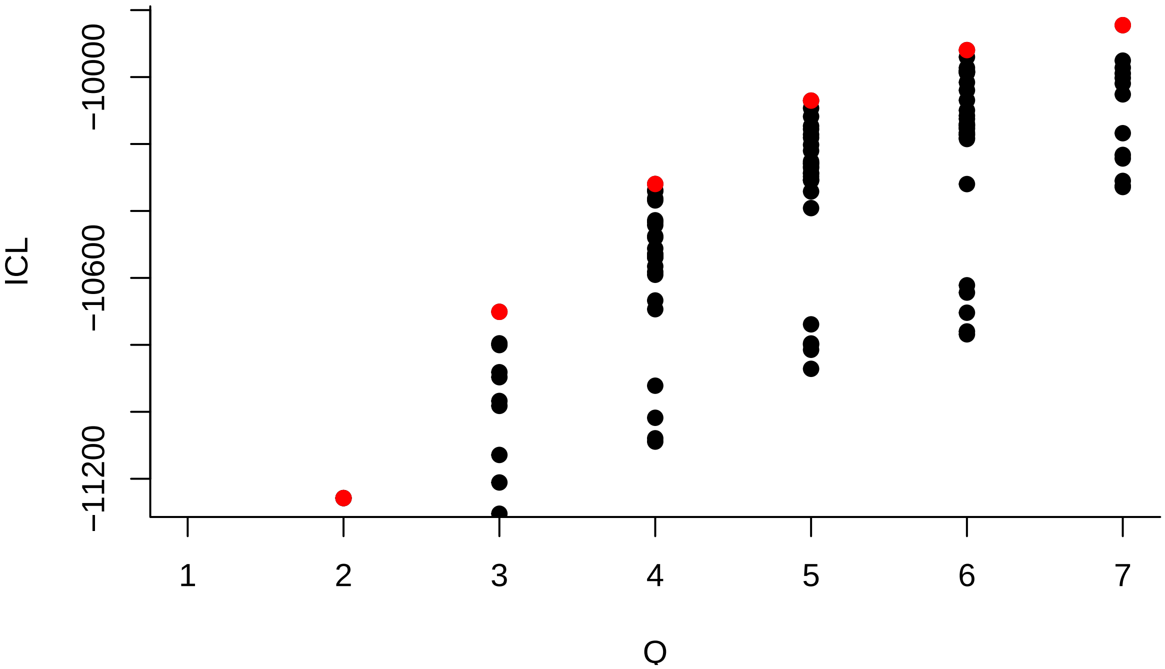
ICL
−10000
−10400
−10800
−11200
1
2
3
4
5
6
7
Q

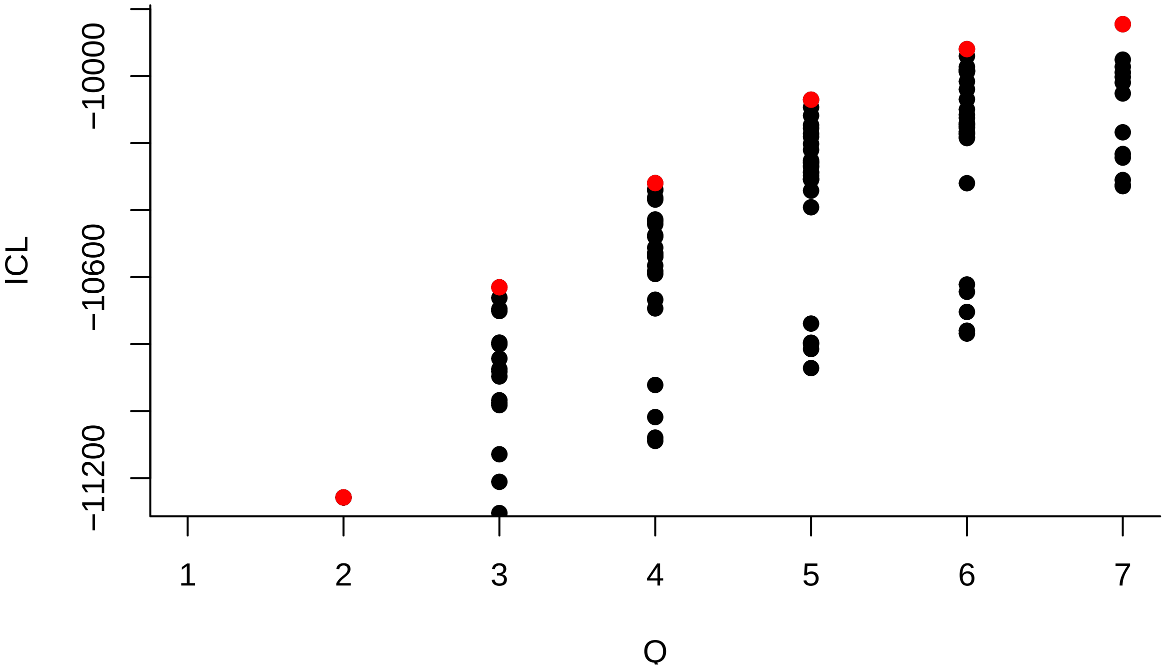


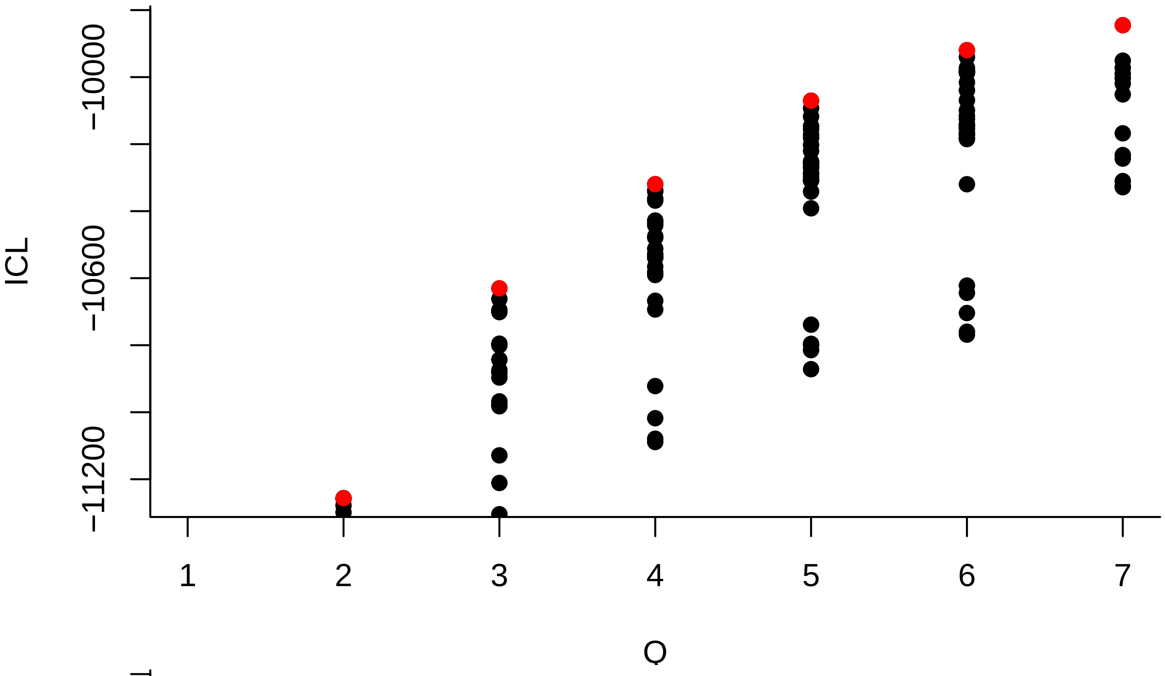
ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q

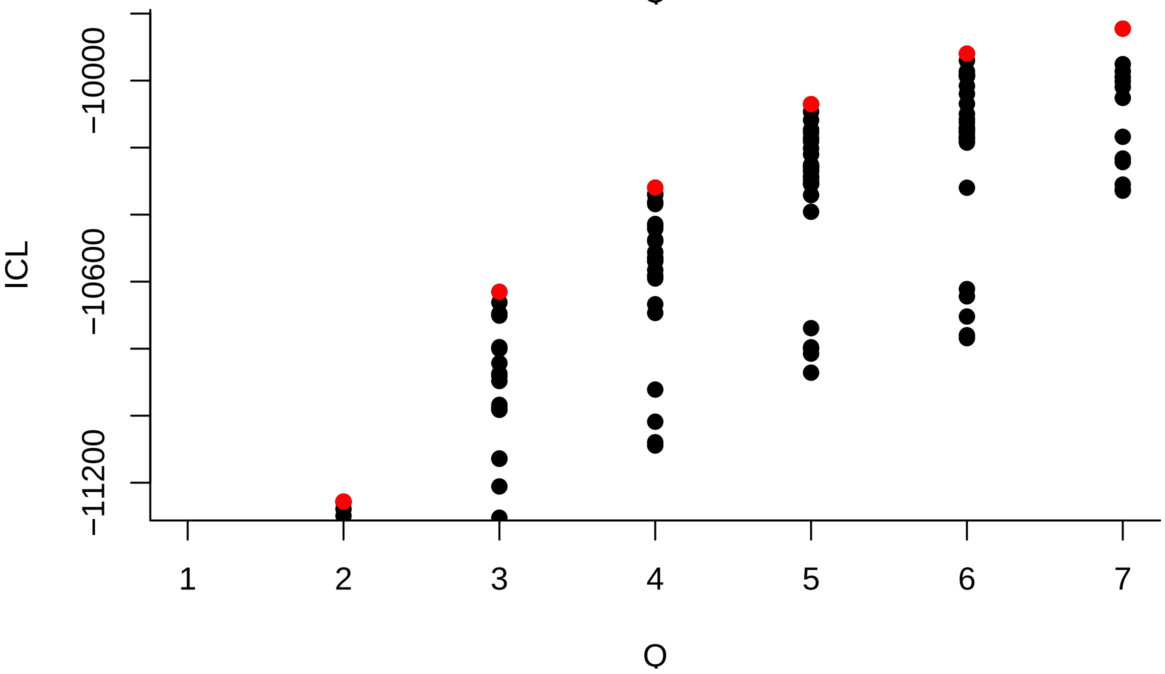

ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q

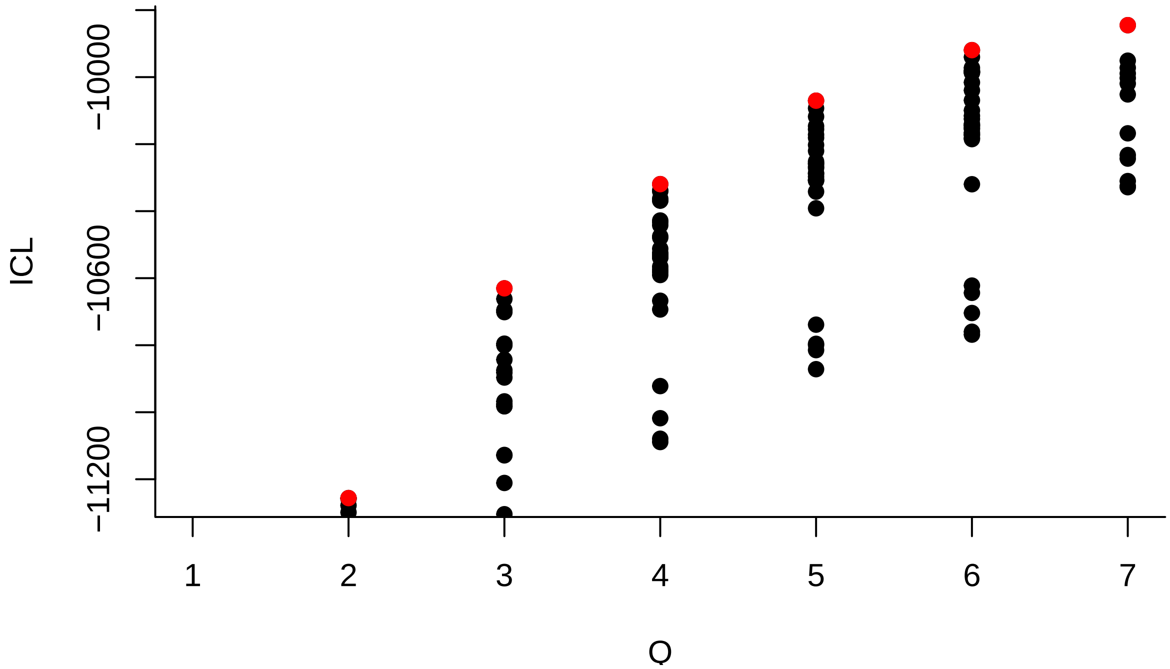

ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q

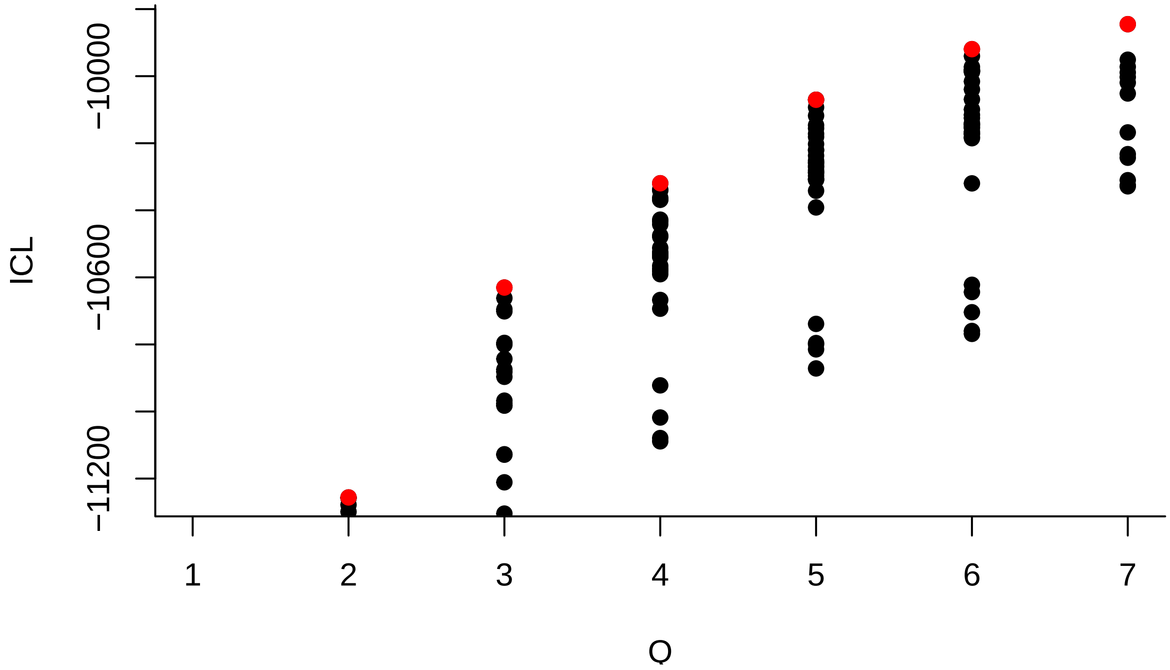

ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q

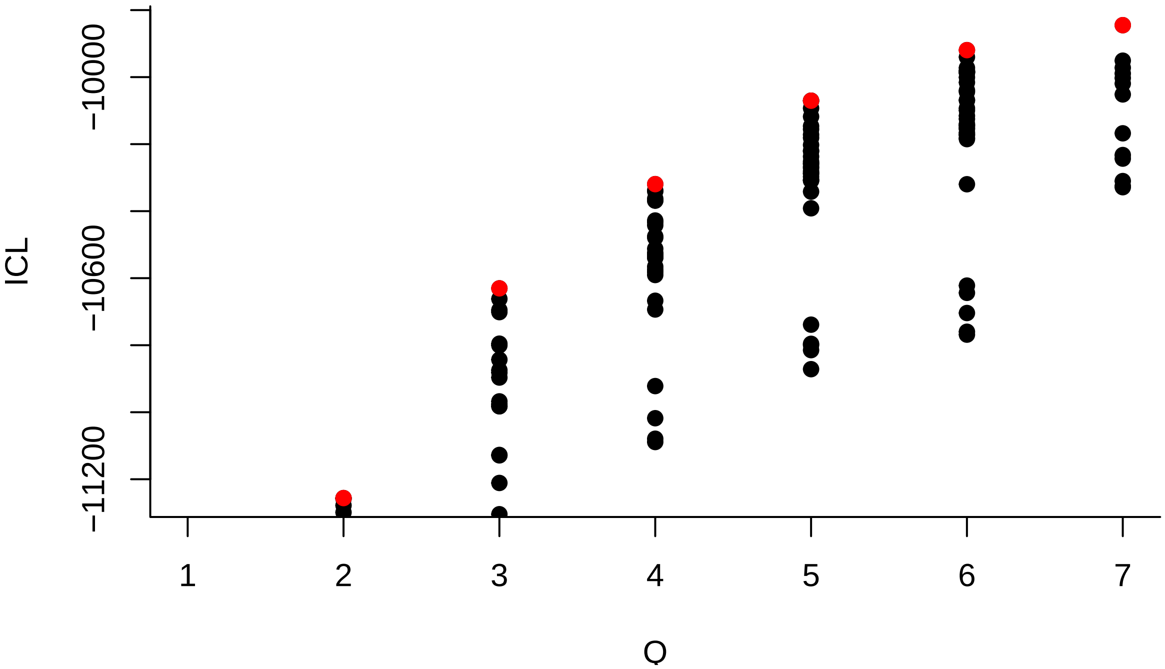

ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q

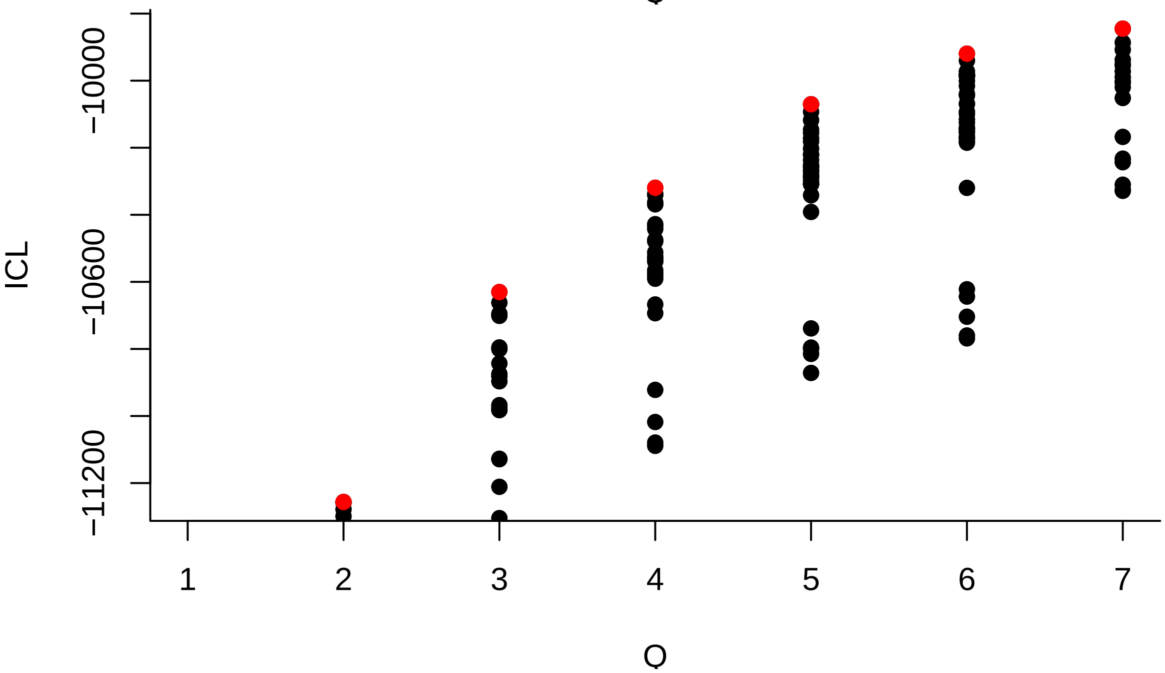

ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q


ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q


ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q


ICL
−10000
−10600
−11200
1
2
3
4
5
6
7
Q





```
#for period 1:
my_model.period1$memberships[[estimate.period1]]$Z#gives probability of being in each group
test1=my_model.period1$memberships[[estimate.period1]]$Z>0.5
test1

# for period 3:
my_model.period3$memberships[[estimate.period3]]$Z#gives probability of being in each group
test3=my_model.period3$memberships[[estimate.period3]]$Z>0.5
test3
```

In order to understand how people are connected, we applied the Bernoulli blockmodel to the edge connection among the three time periods.By doing so, we can see the estimate number of groups within each time period has. We will also be able to see the probability each individual belong to each of these groups.

By assigning people to the group they have more than 50% chance to be in, we created new matrices– "test1"" for period 1 and "test3"" for period 3. Those two matrices can tell us whether or not an individual is in certain group. In order to use the group-membership information alone with other variables, we mutated a new column to our original datasets which include information about node characteristics. This new column, called "group", use number to indicate which group each individual belongs to. We can then use this information to understand what makes an individual in one group but not another.

Note that we tried to run the estimate group function for all three time periods. However, it only works for the first and the third period. We changed the min and max number of group inside the Bernoulli model and re-examine the adjacency matrix for period 2, but it still does not work. So in the following part we will just focus on period 1 and 3.

Based on the Bernoulli Blockmodel, we know that period one has estimately 5 groups, and period 3 has estimately 7 groups. The above code are used to mutate the group-membership information to the original datasets.

Multinomial Logistic Regression

```
## Loading required package: nnet

## Loading required package: ggplot2

## Loading required package: reshape2

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths

## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival

## # weights:  25 (16 variable)
## initial  value 267.166693
## iter  10 value 187.816675
## iter  20 value 186.163361
## iter  30 value 186.093250
## final  value 186.092991
## converged

##
## z test of coefficients:
##
##                     Estimate  Std. Error     z value  Pr(>|z|)
## 2:(Intercept)     4.1015e+00  1.0084e+00  4.0675e+00 4.753e-05 ***
## 2:Gender          1.0372e+01  1.7877e+02  5.8000e-02   0.95373
## 2:Marital.Status -2.5090e+00  1.0954e+00 -2.2904e+00   0.02200 *
## 2:University     -5.4564e-01  1.2173e+00 -4.4820e-01   0.65398
## 3:(Intercept)     2.1168e+00  1.0610e+00  1.9952e+00   0.04602 *
## 3:Gender         -4.5563e+00  6.5996e+00 -6.9040e-01   0.48995
## 3:Marital.Status -1.7181e+00  1.1787e+00 -1.4576e+00   0.14494
## 3:University     -1.4518e+01  7.5769e-05 -1.9161e+05 < 2.2e-16 ***
## 4:(Intercept)     2.4922e+00  1.0412e+00  2.3936e+00   0.01669 *
## 4:Gender         -4.9278e+00  5.2840e+00 -9.3260e-01   0.35103
## 4:Marital.Status -2.6125e+00  1.1868e+00 -2.2013e+00   0.02772 *
## 4:University     -2.7951e-01  1.5253e+00 -1.8320e-01   0.85461
## 5:(Intercept)     2.4525e+00  1.0422e+00  2.3532e+00   0.01861 *
## 5:Gender          9.5996e+00  1.7877e+02  5.3700e-02   0.95718
## 5:Marital.Status -2.1830e+00  1.1626e+00 -1.8776e+00   0.06043 .
## 5:University     -6.4246e-01  1.5127e+00 -4.2470e-01   0.67105
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## # weights:  35 (24 variable)
## initial  value 482.585717
## iter  10 value 330.799021
## iter  20 value 327.342119
## iter  30 value 327.228823
## iter  40 value 327.227907
## final  value 327.227902
## converged

##
## z test of coefficients:
##
##                     Estimate  Std. Error     z value  Pr(>|z|)
## 2:(Intercept)    -4.3910e+00  1.0621e+00 -4.1341e+00 3.563e-05 ***
## 2:Gender          1.1007e+00  1.0738e+00  1.0251e+00 0.3053392
## 2:Marital.Status  2.5679e+00  1.1774e+00  2.1809e+00 0.0291876 *
## 2:University      2.5898e+00  1.5936e+00  1.6251e+00 0.1041319
## 3:(Intercept)    -2.4642e-02  1.9364e-01 -1.2730e-01 0.8987369
## 3:Gender         -6.5903e-01  5.8085e-01 -1.1346e+00 0.2565420
## 3:Marital.Status  1.3104e+00  3.5451e-01  3.6964e+00 0.0002186 ***
## 3:University      1.5396e+00  1.1133e+00  1.3829e+00 0.1667009
## 4:(Intercept)    -4.0602e+00  1.0166e+00 -3.9939e+00 6.501e-05 ***
## 4:Gender          3.5183e-01  1.2311e+00  2.8580e-01 0.7750512
## 4:Marital.Status  2.6838e+00  1.1638e+00  2.3060e+00 0.0211089 *
## 4:University     -1.0899e+01  8.4114e+02 -1.3000e-02 0.9896618
## 5:(Intercept)    -1.8717e+00  3.7931e-01 -4.9345e+00 8.034e-07 ***
## 5:Gender         -2.1588e+01  1.2617e-08 -1.7110e+09 < 2.2e-16 ***
## 5:Marital.Status -6.3729e-01  1.1029e+00 -5.7780e-01 0.5633732
## 5:University     -1.1809e+01  1.3270e-03 -8.8995e+03 < 2.2e-16 ***
## 6:(Intercept)    -4.0762e+00  1.0131e+00 -4.0237e+00 5.729e-05 ***
## 6:Gender         -3.4471e-01  9.3034e-01 -3.7050e-01 0.7109997
## 6:Marital.Status  4.0028e+00  1.0784e+00  3.7119e+00 0.0002057 ***
## 6:University      1.6918e+00  1.5237e+00  1.1104e+00 0.2668421
## 7:(Intercept)    -1.3976e+00  3.0213e-01 -4.6259e+00 3.729e-06 ***
## 7:Gender         -1.0933e+00  8.8519e-01 -1.2351e+00 0.2168043
## 7:Marital.Status  1.6441e+00  4.6167e-01  3.5611e+00 0.0003693 ***
## 7:University      2.4103e+00  1.1742e+00  2.0528e+00 0.0400918 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the filtered data, we then ran the multinomial regression model on period 1 and period 3 data. The goal is to understand what variable(s) determine which group a certain belongs to. We choose three major variable: whether or not went to university, marital status, and gender. The hypothesis is that people are more likely to be with people with similar background (married ppl are more likely to befriend with married ppl, etc.).

In period one, we saw that marital status among the five groups does not have statistically significant influence (p-value >0.05 for group 2-4, indicating that marital status variable does not have significant effect on distinguishing them from group 1). The other two variable, "University" and "Gender", however, has great influence. We can conclude that in period one, gender and education level contribute significantly on who PIRA members befriend with.

In Period two, we saw a different trend. Marriage status has significant influence on the chance of being assigned into certain for most of the time; only in group 2 this variable does not has a significant p-value. The other two variables work efficiently for all seven groups. Thus we can conclude that in most cases at period 2, PIRA members' friend choices is heavily influenced by their marriage status, gender, and education background.

ERGM Model

```
#Period 1:
MaritalStatus_1=ergm(net_1ma~edges+nodematch("Marital.Status"))
```

```
## Starting maximum pseudolikelihood estimation (MPLE):

## Evaluating the predictor and response matrix.

## Maximizing the pseudolikelihood.

## Finished MPLE.

## Stopping at the initial estimate.

## Evaluating log-likelihood at the estimate.
```

```
summary(MaritalStatus_1)
```

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   net_1ma ~ edges + nodematch("Marital.Status")
```

```
##
## Iterations:  7 out of 20
##
## Monte Carlo MLE Results:
##                            Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                      -4.90527    0.07133      0 -68.772   <1e-04 ***
## nodematch.Marital.Status -0.01882    0.09976      0  -0.189     0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 77093  on 55611  degrees of freedom
##  Residual Deviance:  4794  on 55609  degrees of freedom
##
## AIC: 4798    BIC: 4816    (Smaller is better.)
```

```
#Period 1:
MaritalStatus_2=ergm(net_2ma~edges+nodematch("Marital.Status"))
```

```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```

```
summary(MaritalStatus_2)
```

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   net_2ma ~ edges + nodematch("Marital.Status")
##
## Iterations:  7 out of 20
##
## Monte Carlo MLE Results:
##                            Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                      -4.51446    0.07433      0 -60.739   <1e-04 ***
## nodematch.Marital.Status -0.14748    0.10933      0  -1.349    0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 46677  on 33670  degrees of freedom
##  Residual Deviance:  3800  on 33668  degrees of freedom
##
## AIC: 3804    BIC: 3820    (Smaller is better.)
```

```
#Period 3
MaritalStatus_3=ergm(net_3ma~edges+nodematch("Marital.Status"))
```

```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```

```r
summary(MaritalStatus_3)
```

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   net_3ma ~ edges + nodematch("Marital.Status")
##
## Iterations:  7 out of 20
##
## Monte Carlo MLE Results:
##                            Estimate Std. Error MCMC %  z value Pr(>|z|)
## edges                      -4.96878    0.04628      0 -107.355   <1e-04 ***
## nodematch.Marital.Status    0.13953    0.06291      0    2.218   0.0266 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 191413  on 138075  degrees of freedom
##  Residual Deviance:  12089  on 138073  degrees of freedom
##
## AIC: 12093    BIC: 12112    (Smaller is better.)
```

```r
#Period 6
MaritalStatus_6=ergm(net_6ma~edges+nodematch("Marital.Status"))
```

```
## Starting maximum pseudolikelihood estimation (MPLE):
## Evaluating the predictor and response matrix.
## Maximizing the pseudolikelihood.
## Finished MPLE.
## Stopping at the initial estimate.
## Evaluating log-likelihood at the estimate.
```

```r
summary(MaritalStatus_6)
```

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   net_6ma ~ edges + nodematch("Marital.Status")
##
## Iterations:  6 out of 20
##
## Monte Carlo MLE Results:
##                            Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                       -3.3769     0.1551      0 -21.776   <1e-04 ***
## nodematch.Marital.Status    -0.3579     0.2120      0  -1.688   0.0914 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 4718  on 3403  degrees of freedom
##  Residual Deviance:  843  on 3401  degrees of freedom
##
```

```
## AIC: 847    BIC: 859.3    (Smaller is better.)
```

For all three periods when node matching on marital status we see that we obtain a p-value of 0.0266, which is statistically significant. Meaning that there is a tendency within the network for members of the PIRA with similar marital status have some form of relationship to each other. This is interesting because it is an under considered version of homophily in that we are often most concerned with the criminal actives of such organizations that there may be very mundane ways to analyze the network. Knowing this it may be worth looking to how marital status relates to the other nodal characteristics. If we had access to the information it could be avantages to look at the types of relationship best described by marital status save for related for marriage, which would give us a valuable perspective of the inner workings of the network. (This is sort of discussion but I'm not sure how to dive it up).

Discussion:

Ethics: Due to the data having already been collected and published the ethical concerns of the data have hypothetically been addressed. However little is available on how this data was collected, and so we are unsure when in relation to the network the data was collected and the status of the individuals. There is a lot of concern with the vaility and the ethics of collecting data from prisoners, this may have been the case of these individuals. It could have been collected retroactively from events and individuals, but this is likely not the case. When data is collected from inmates it is unlike that the information is truly voluntary. They may also be concerned with confidentiality and refuse to disclose some, more sensitive, information about relatives or significant others in hopes of keeping them safe.

We have some missing data in all of our data sets, due to lack of information, we do not know whether our data is missing at random or not, but we would like to see in the further analysis, whether there will be more data added in or changed, meaning new relationship being formed or old relationship being changed.