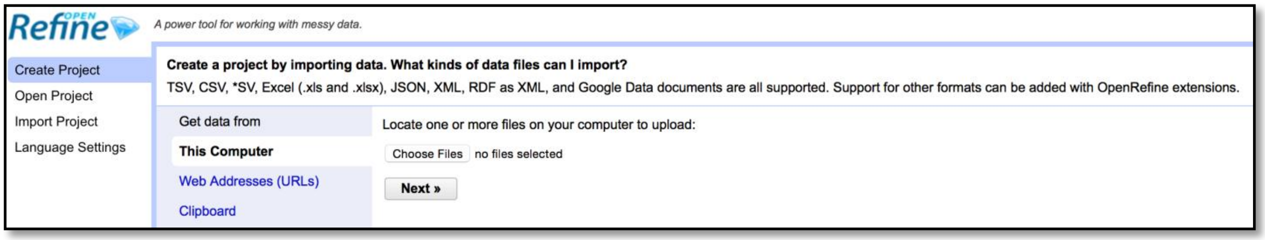
Chapter 1: Preparing the Data



A power tool for working with messy data.

- Create Project
- Open Project
- Import Project
- Language Settings

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.

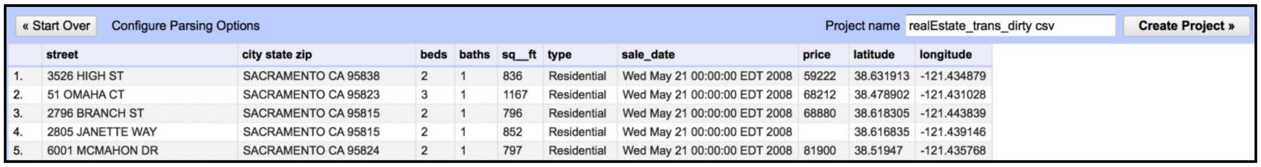
Get data from

- This Computer
- Web Addresses (URLs)
- Clipboard

Locate one or more files on your computer to upload:

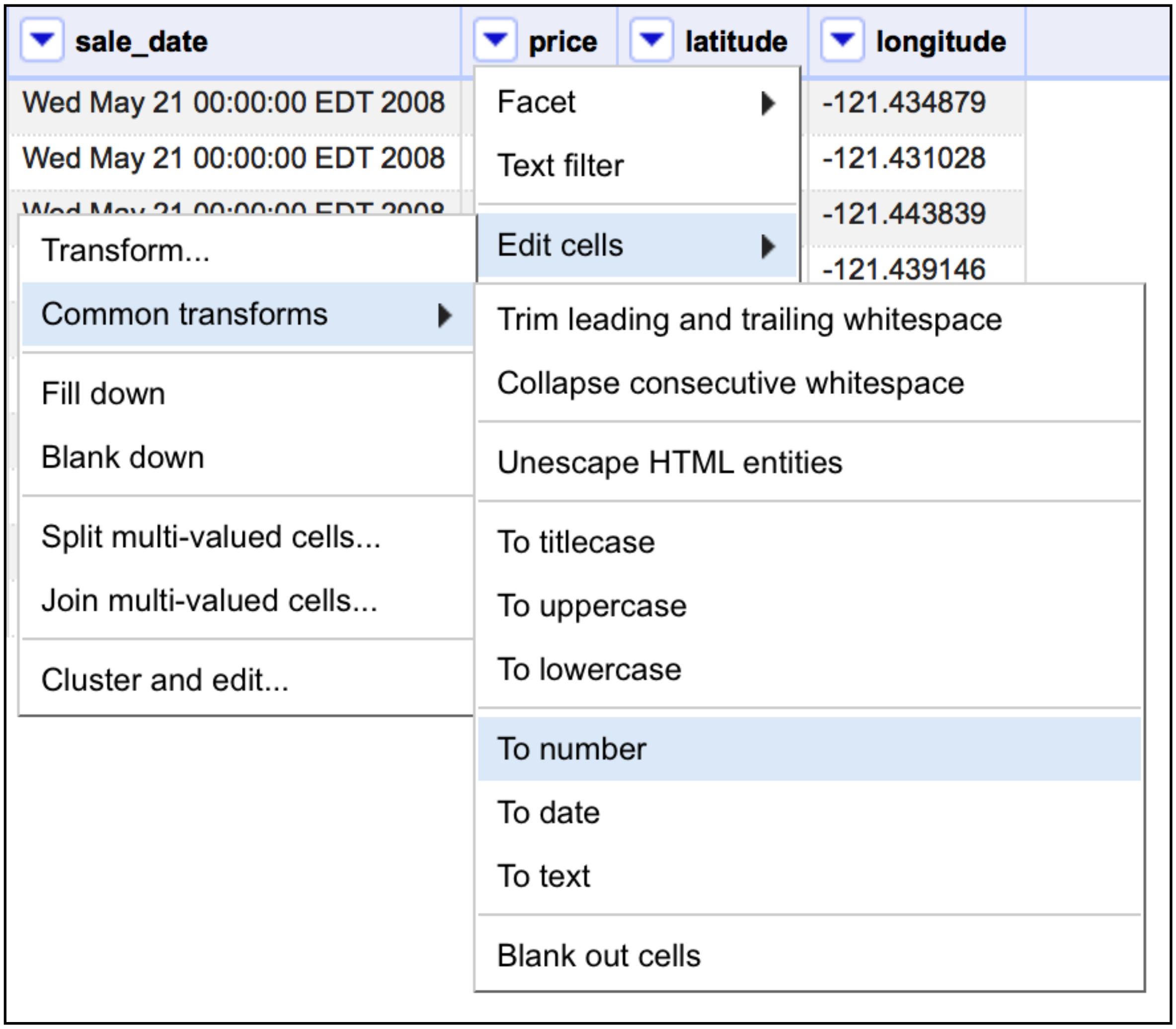
Choose Files no files selected

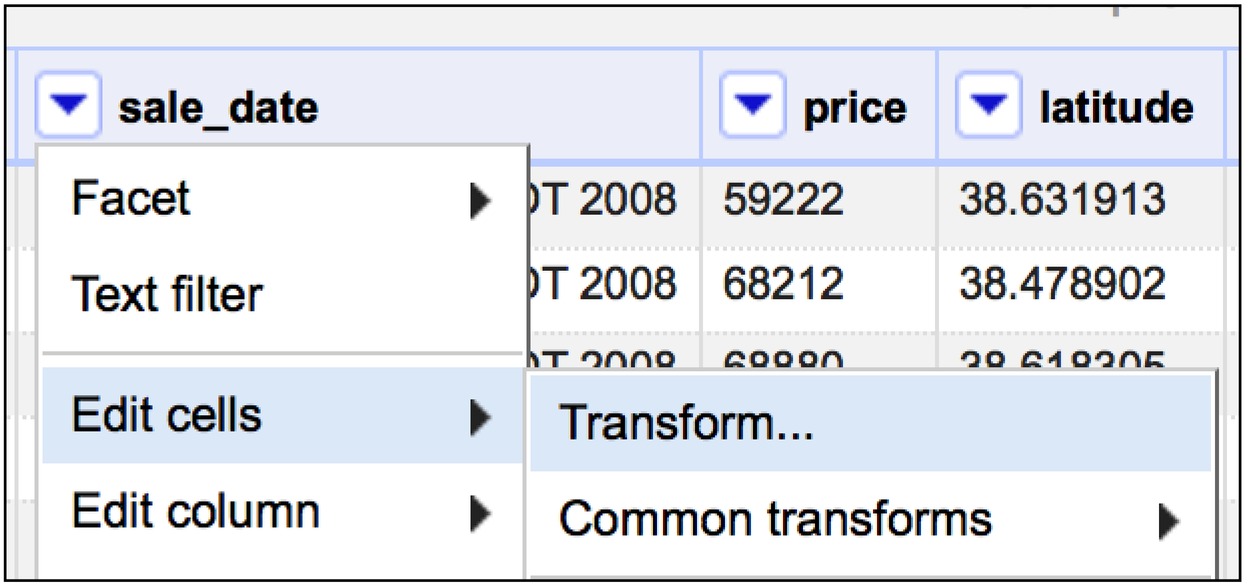
Next »

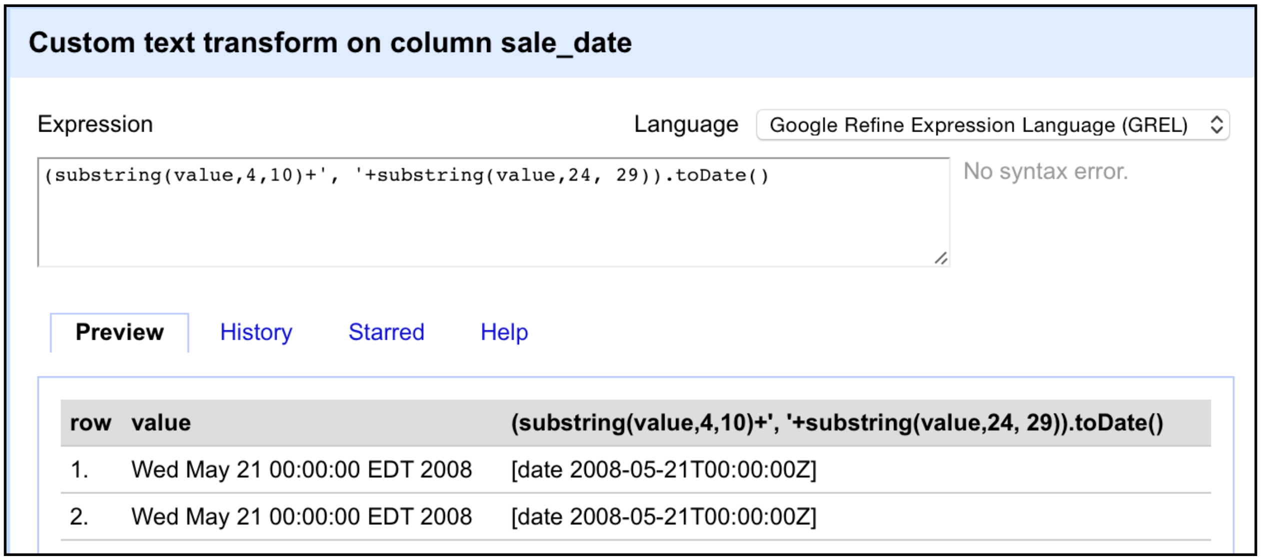


« Start Over Configure Parsing Options Project name realEstate_trans_dirty csv Create Project »

	street	city state zip	beds	baths	sq__ft	type	sale_date	price	latitude	longitude
1.	3526 HIGH ST	SACRAMENTO CA 95838	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.631913	-121.434879
2.	51 OMAHA CT	SACRAMENTO CA 95823	3	1	1167	Residential	Wed May 21 00:00:00 EDT 2008	68212	38.478902	-121.431028
3.	2796 BRANCH ST	SACRAMENTO CA 95815	2	1	796	Residential	Wed May 21 00:00:00 EDT 2008	68880	38.618305	-121.443839
4.	2805 JANETTE WAY	SACRAMENTO CA 95815	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008		38.616835	-121.439146
5.	6001 MCMAHON DR	SACRAMENTO CA 95824	2	1	797	Residential	Wed May 21 00:00:00 EDT 2008	81900	38.51947	-121.435768







Custom text transform on column sale_date

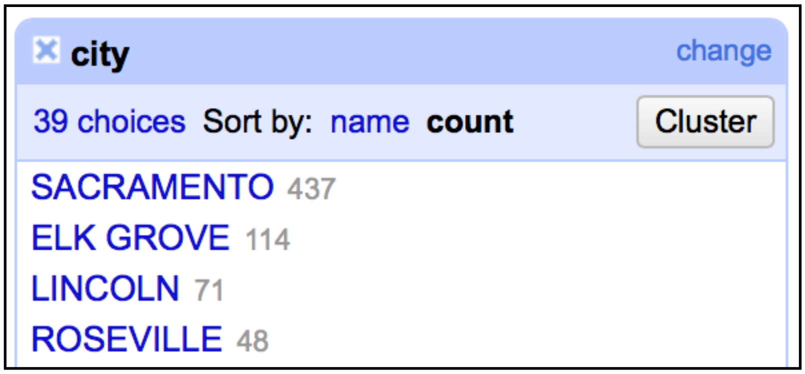
Expression

Language: Google Refine Expression Language (GREL)

```
(substring(value,4,10)+', '+substring(value,24, 29)).toDate()
```

No syntax error.

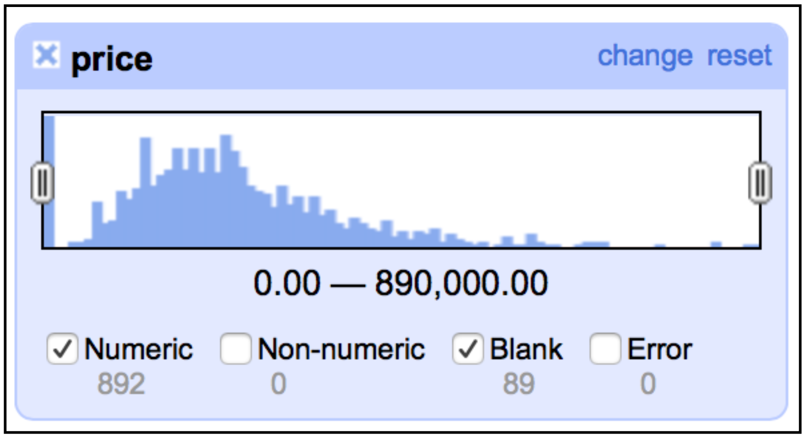
row	value	(substring(value,4,10)+', '+substring(value,24, 29)).toDate()
1.	Wed May 21 00:00:00 EDT 2008	[date 2008-05-21T00:00:00Z]
2.	Wed May 21 00:00:00 EDT 2008	[date 2008-05-21T00:00:00Z]



city

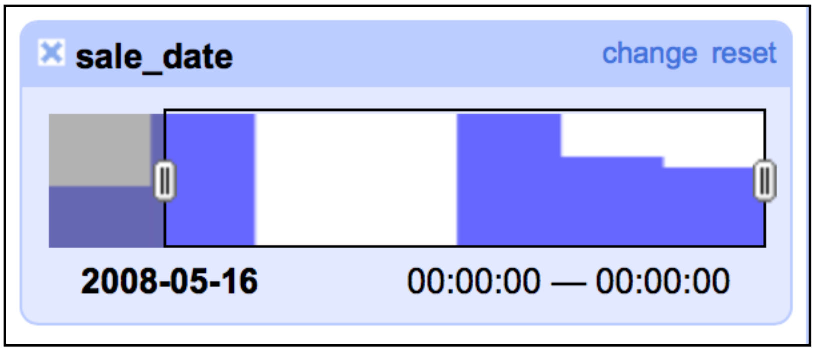
39 choices Sort by: name **count**

SACRAMENTO	437
ELK GROVE	114
LINCOLN	71
ROSEVILLE	48

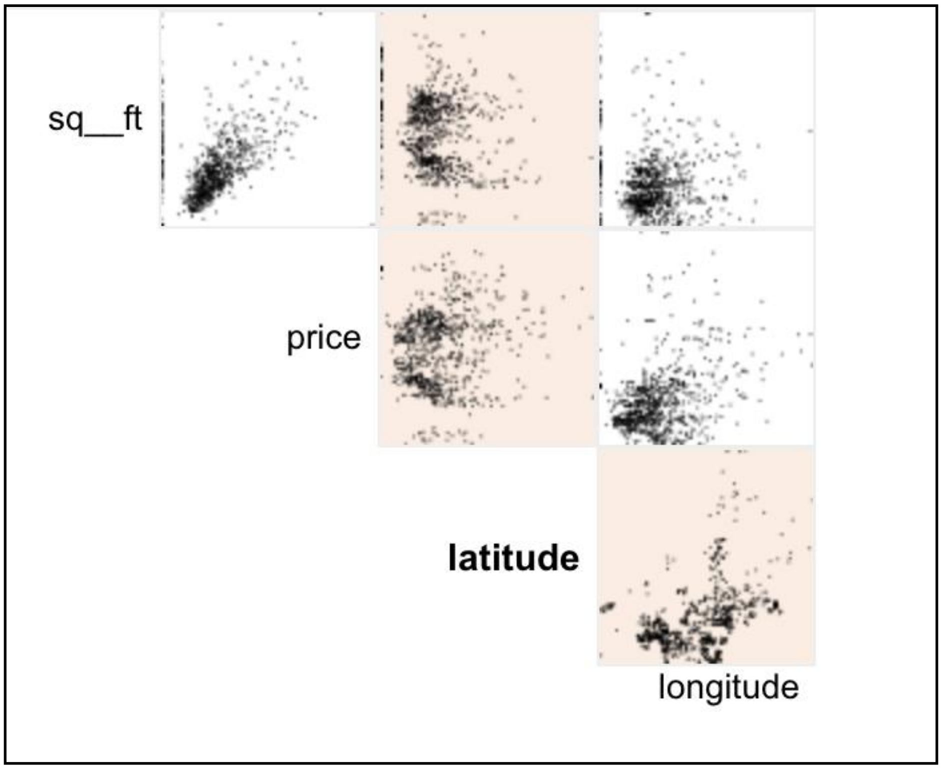


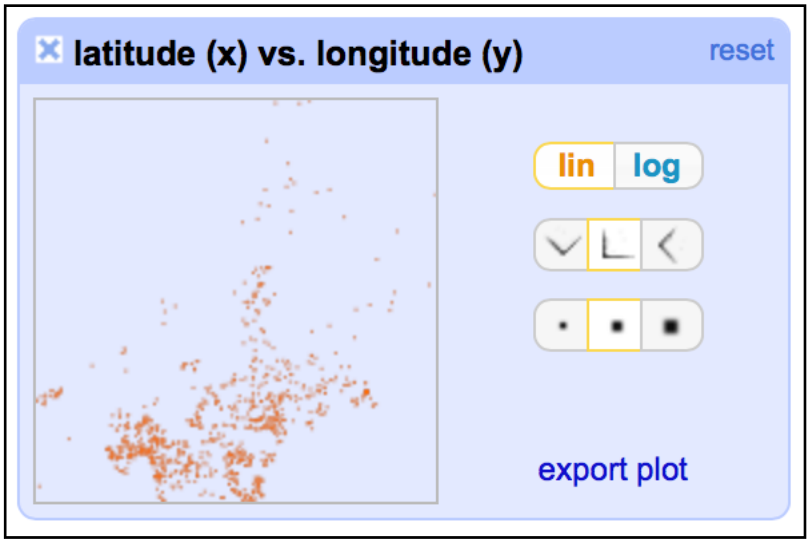
0.00 — 890,000.00

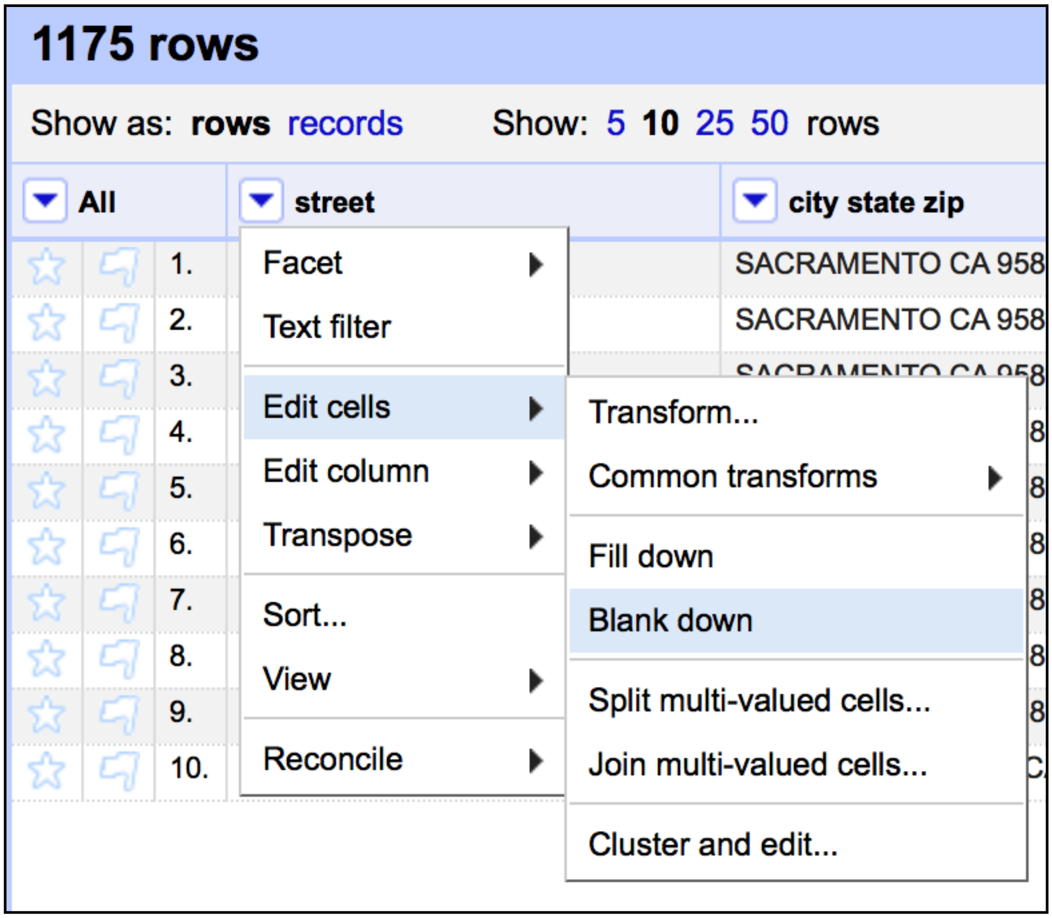
☑ Numeric 892 □ Non-numeric 0 ☑ Blank 89 □ Error 0

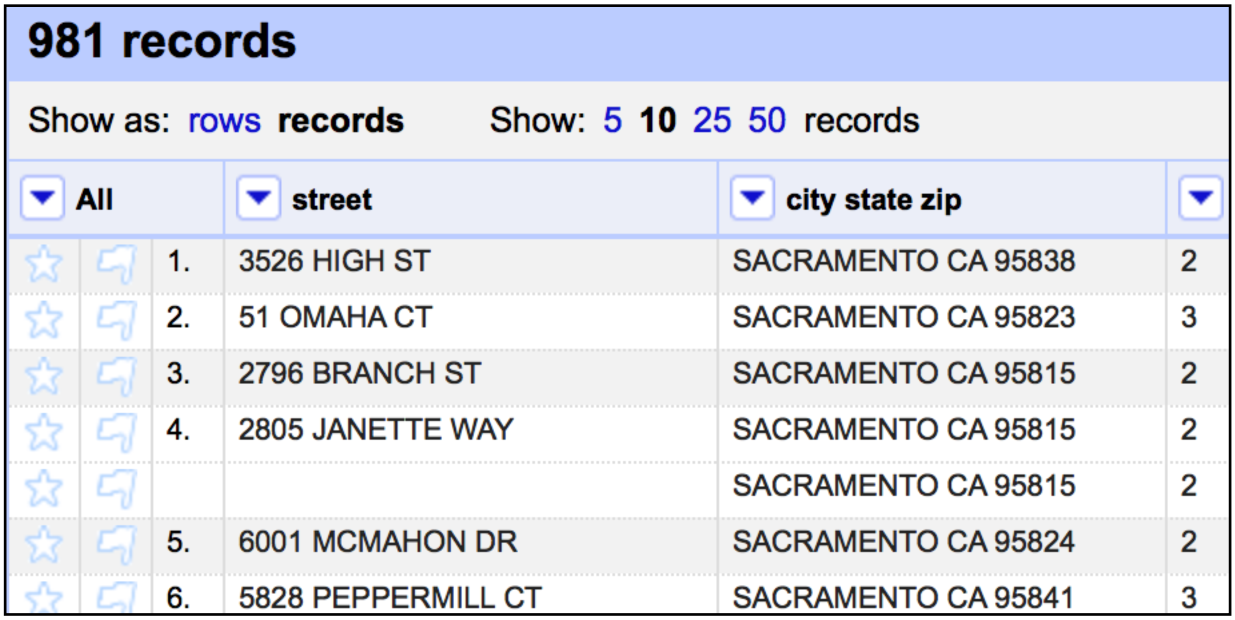


2008-05-16 00:00:00 — 00:00:00




latitude (x) vs. longitude (y)
reset
lin
log
export plot

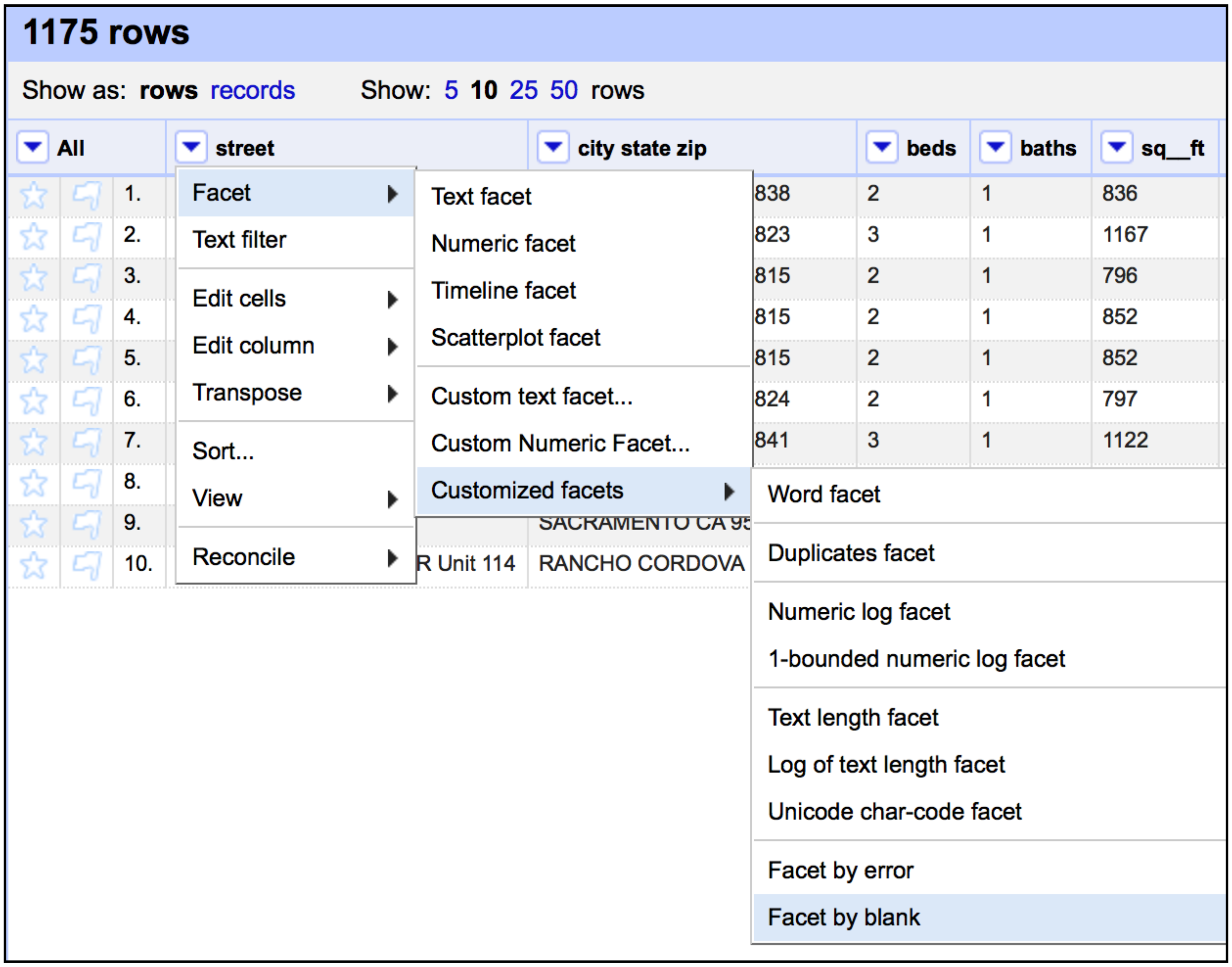

1175 rows
Show as: rows records
Show: 5 10 25 50 rows
All
street
city state zip
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
SACRAMENTO CA 958
SACRAMENTO CA 958
SACRAMENTO CA 958
Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile
Transform...
Common transforms
Fill down
Blank down
Split multi-valued cells...
Join multi-valued cells...
Cluster and edit...



981 records

Show as: rows **records**   Show: 5 **10** 25 50 records

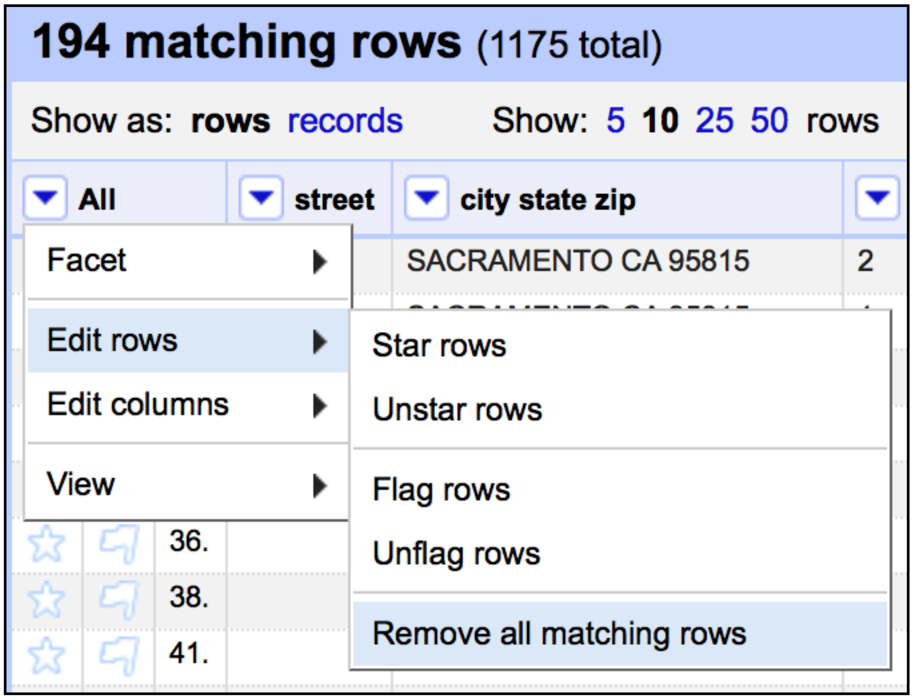
All			street	city state zip	
☆		1.	3526 HIGH ST	SACRAMENTO CA 95838	2
☆		2.	51 OMAHA CT	SACRAMENTO CA 95823	3
☆		3.	2796 BRANCH ST	SACRAMENTO CA 95815	2
☆		4.	2805 JANETTE WAY	SACRAMENTO CA 95815	2
☆				SACRAMENTO CA 95815	2
☆		5.	6001 MCMAHON DR	SACRAMENTO CA 95824	2
☆		6.	5828 PEPPERMILL CT	SACRAMENTO CA 95841	3



1175 rows

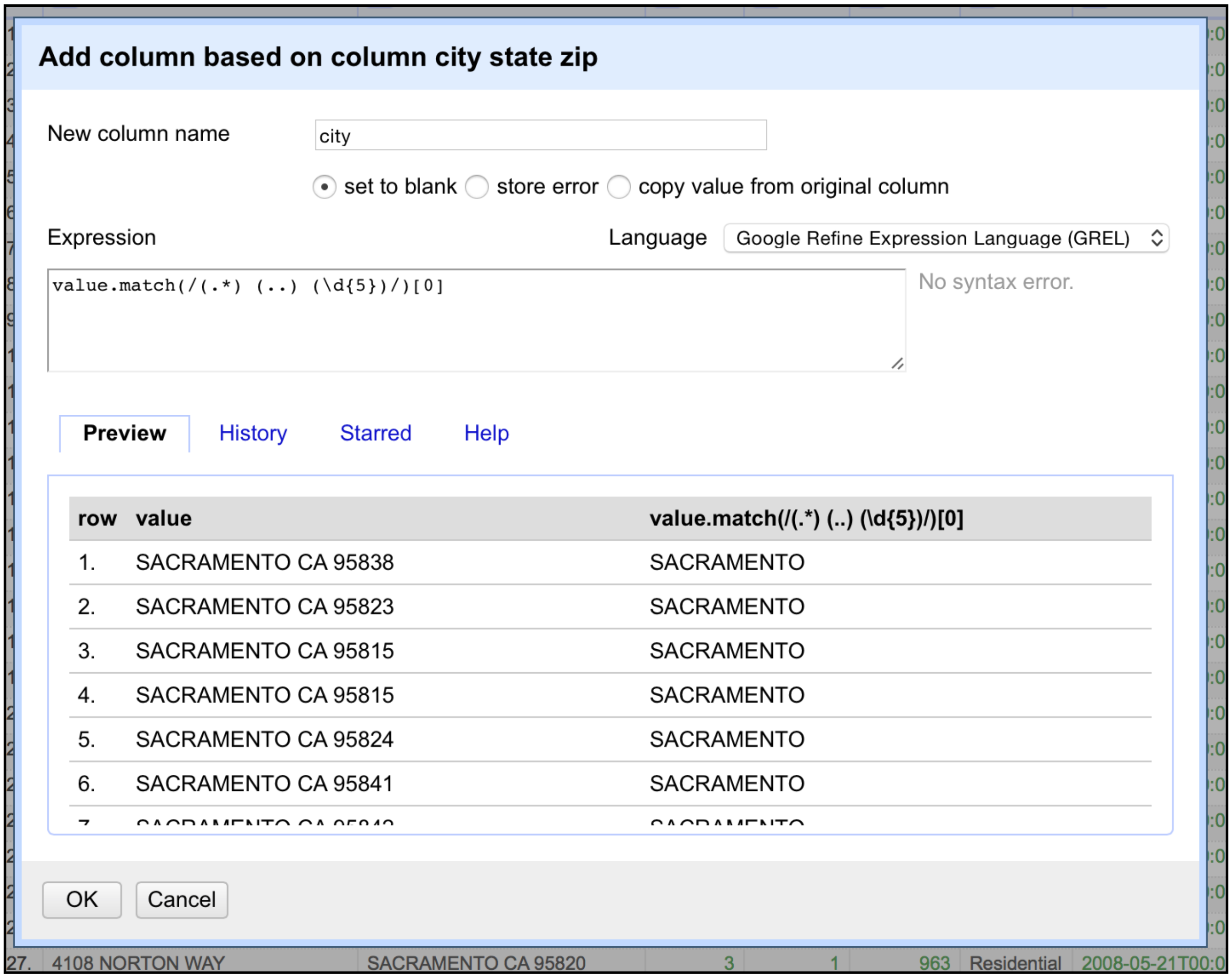
Show as: **rows** records   Show: 5 **10** 25 50 rows

All			street	city state zip	beds	baths	sq__ft
☆		1.		838	2	1	836
☆		2.		823	3	1	1167
☆		3.		815	2	1	796
☆		4.		815	2	1	852
☆		5.		815	2	1	852
☆		6.		824	2	1	797
☆		7.		841	3	1	1122
☆		8.					
☆		9.		SACRAMENTO CA 95			
☆		10.	R Unit 114	RANCHO CORDOVA			



194 matching rows (1175 total)

Show as: **rows** records Show: 5 **10** 25 50 rows



Add column based on column city state zip

New column name: city

On error: ● set to blank ○ store error ○ copy value from original column

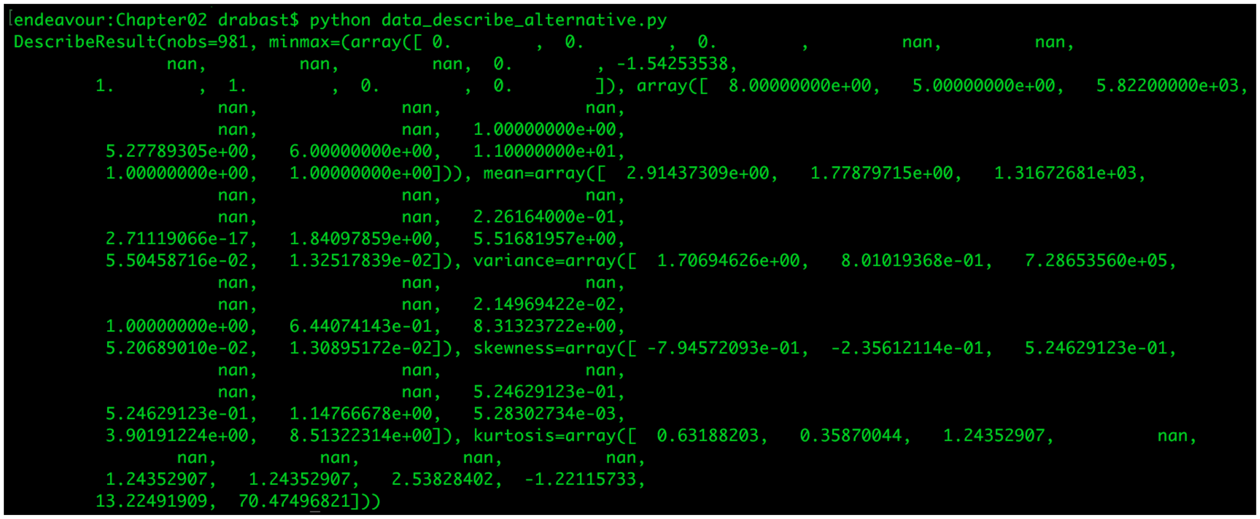
Expression — Language: Google Refine Expression Language (GREL)

```
value.match(/(.*) (..) (\d{5})/)[0]
```

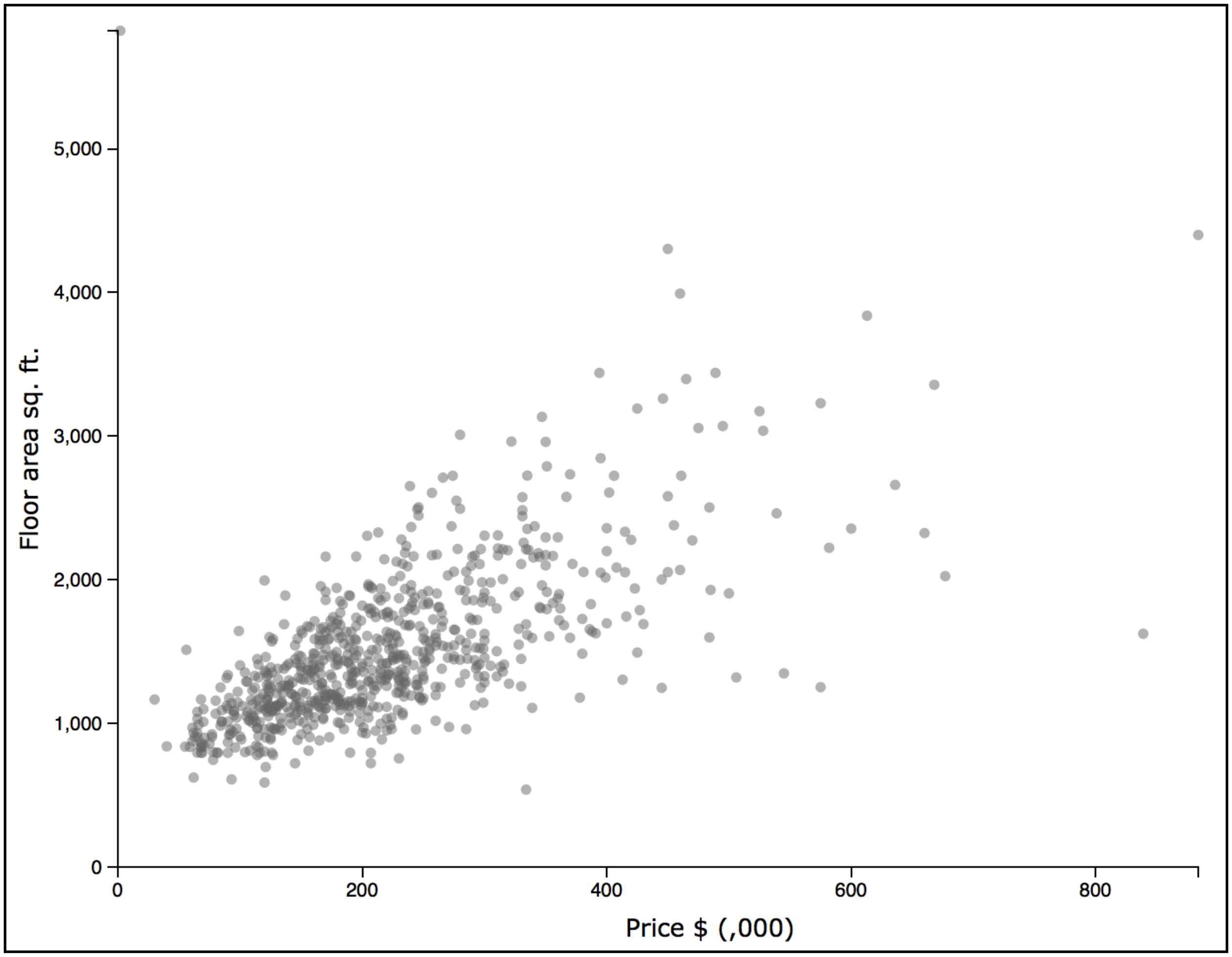
No syntax error.

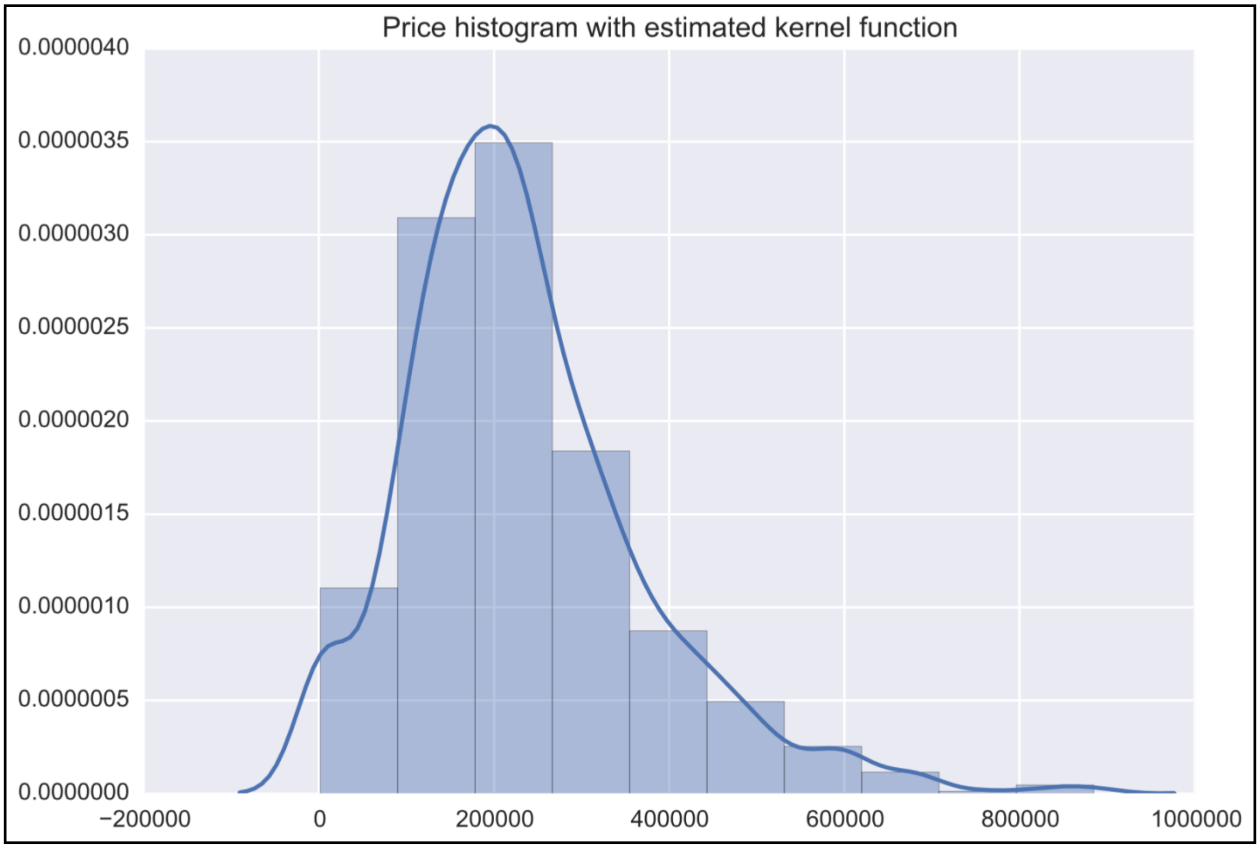
row	value	value.match(/(.*) (..) (\d{5})/)[0]
1.	SACRAMENTO CA 95838	SACRAMENTO
2.	SACRAMENTO CA 95823	SACRAMENTO
3.	SACRAMENTO CA 95815	SACRAMENTO
4.	SACRAMENTO CA 95815	SACRAMENTO
5.	SACRAMENTO CA 95824	SACRAMENTO
6.	SACRAMENTO CA 95841	SACRAMENTO

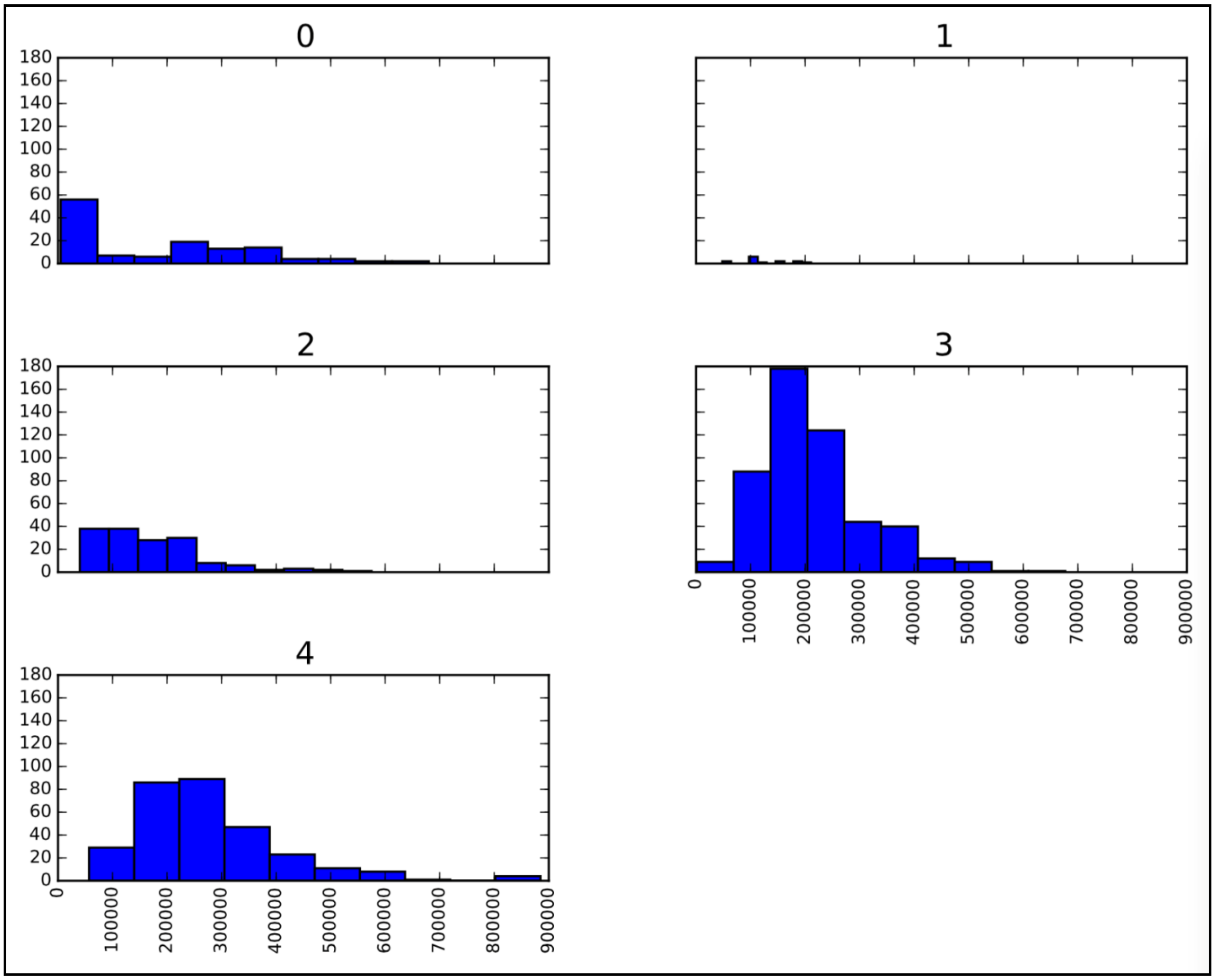
OK
Cancel
27. 4108 NORTON WAY SACRAMENTO CA 95820 3 1 963 Residential 2008-05-21T00:0

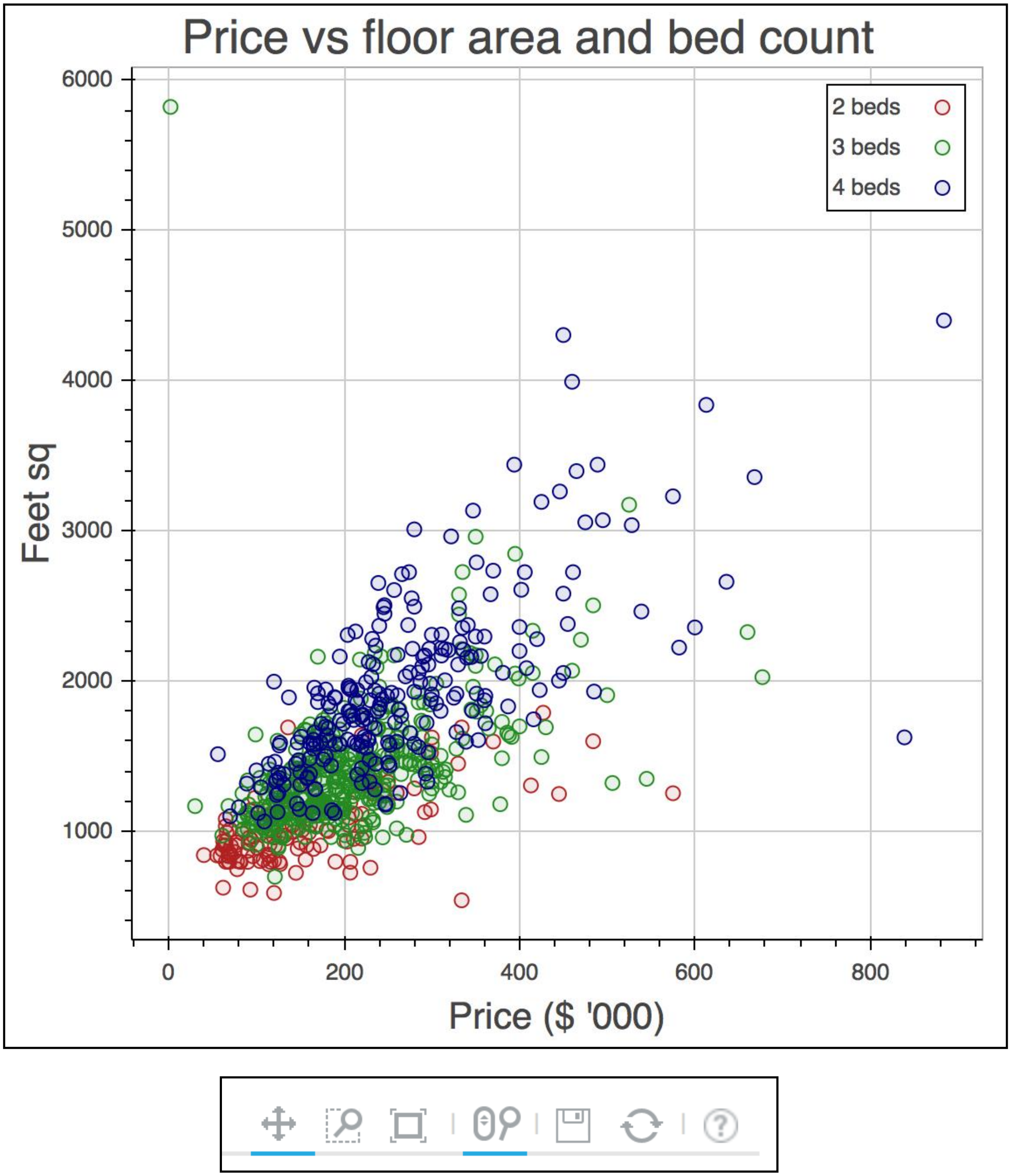


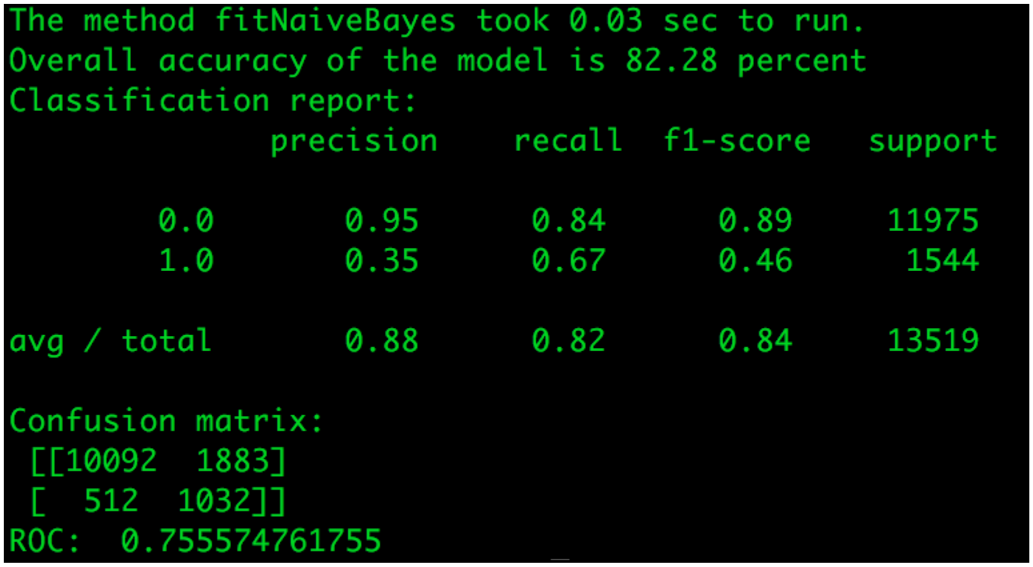
```
[endeavour:Chapter02 drabast$ python data_describe_alternative.py
 DescribeResult(nobs=981, minmax=(array([ 0.         ,  0.         ,  0.         ,          nan,          nan,
              nan,          nan,          nan,  0.         , -1.54253538,
        1.         ,  1.         ,  0.         ,  0.         ]), array([  8.00000000e+00,   5.00000000e+00,   5.82200000e+03,
                   nan,               nan,               nan,
                   nan,               nan,   1.00000000e+00,
         5.27789305e+00,   6.00000000e+00,   1.10000000e+01,
         1.00000000e+00,   1.00000000e+00])), mean=array([  2.91437309e+00,   1.77879715e+00,   1.31672681e+03,
                   nan,               nan,               nan,
                   nan,               nan,   2.26164000e-01,
         2.71119066e-17,   1.84097859e+00,   5.51681957e+00,
         5.50458716e-02,   1.32517839e-02]), variance=array([  1.70694626e+00,   8.01019368e-01,   7.28653560e+05,
                   nan,               nan,               nan,
                   nan,               nan,   2.14969422e-02,
         1.00000000e+00,   6.44074143e-01,   8.31323722e+00,
         5.20689010e-02,   1.30895172e-02]), skewness=array([ -7.94572093e-01,  -2.35612114e-01,   5.24629123e-01,
                   nan,               nan,               nan,
                   nan,               nan,   5.24629123e-01,
         5.24629123e-01,   1.14766678e+00,   5.28302734e-03,
         3.90191224e+00,   8.51322314e+00]), kurtosis=array([  0.63188203,   0.35870044,   1.24352907,          nan,
               nan,          nan,          nan,          nan,
         1.24352907,   1.24352907,   2.53828402,  -1.22115733,
        13.22491909,  70.47496821]))
```



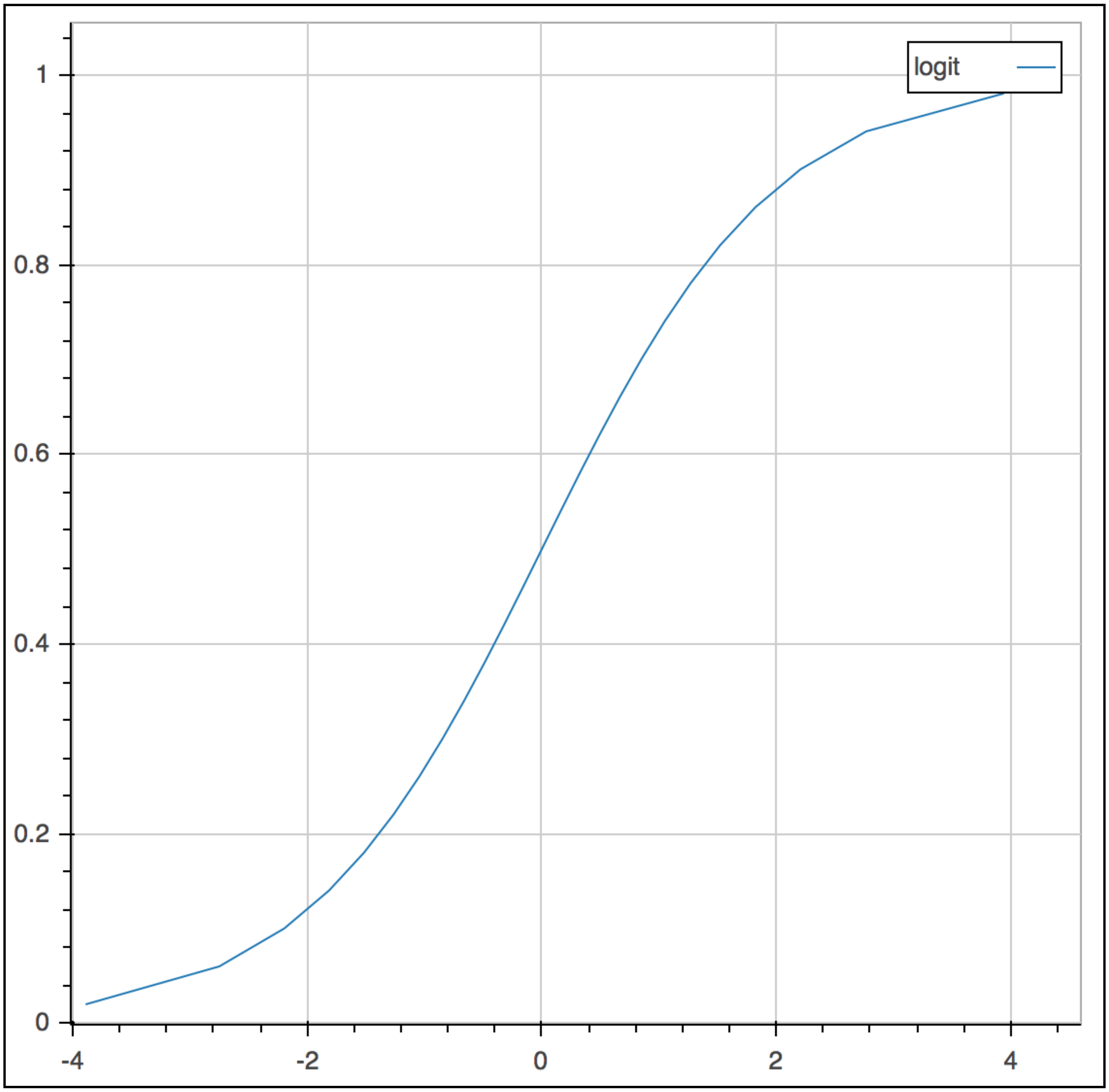

Price histogram with estimated kernel function
0.0000040
0.0000035
0.0000030
0.0000025
0.0000020
0.0000015
0.0000010
0.0000005
0.0000000
−200000
0
200000
400000
600000
800000
1000000

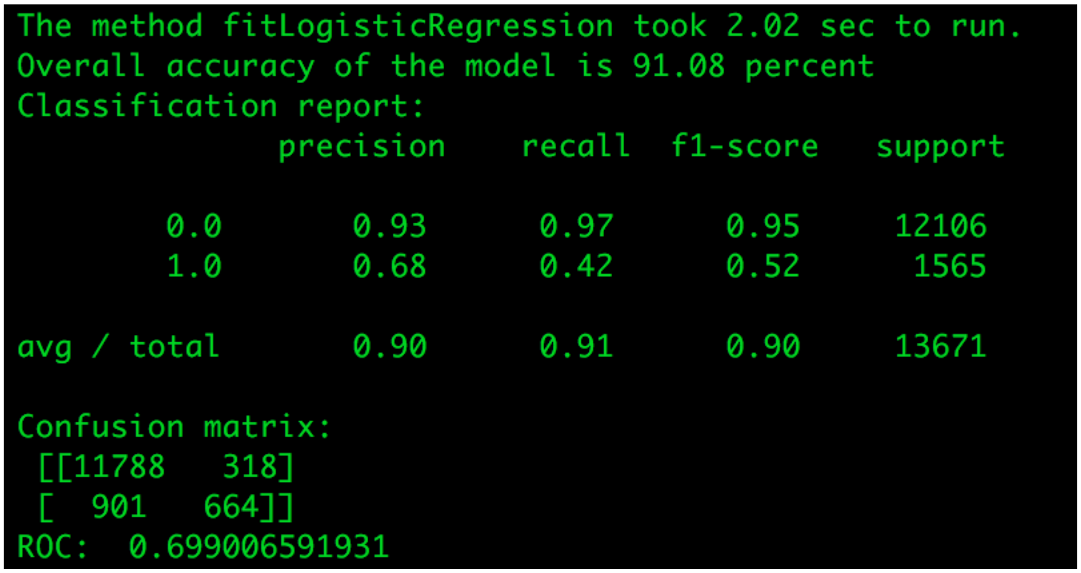

0
1
2
3
4
180
160
140
120
100
80
60
40
20
0
0
100000
200000
300000
400000
500000
600000
700000
800000
900000


Price vs floor area and bed count
Feet sq
6000
5000
4000
3000
2000
1000
0
200
400
600
800
Price ($ '000)
2 beds
3 beds
4 beds

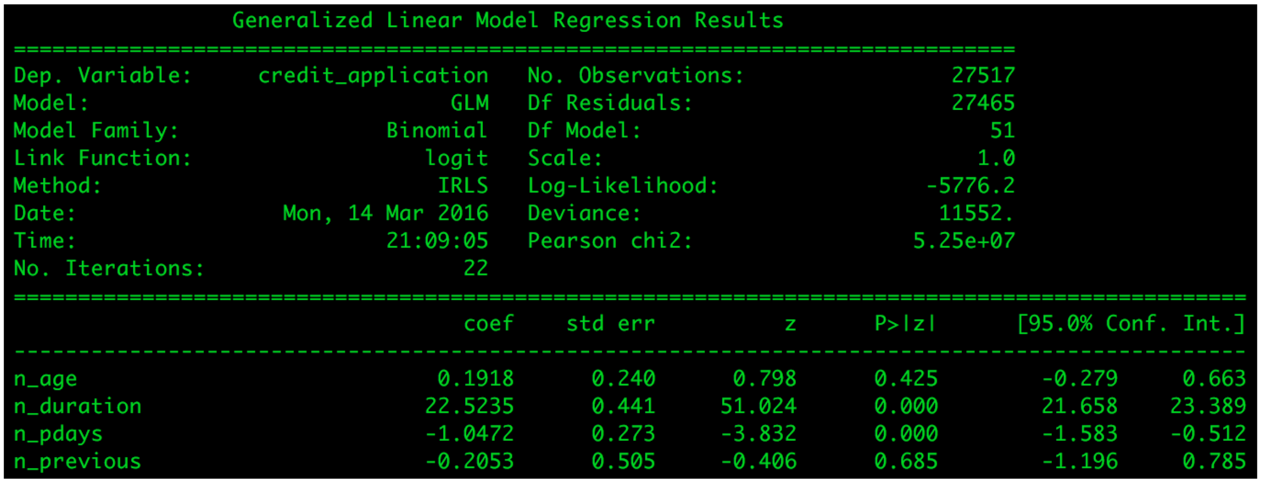


```
The method fitNaiveBayes took 0.03 sec to run.
Overall accuracy of the model is 82.28 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.95      0.84      0.89     11975
        1.0       0.35      0.67      0.46      1544

avg / total       0.88      0.82      0.84     13519

Confusion matrix:
 [[10092  1883]
 [  512  1032]]
ROC:  0.755574761755
```

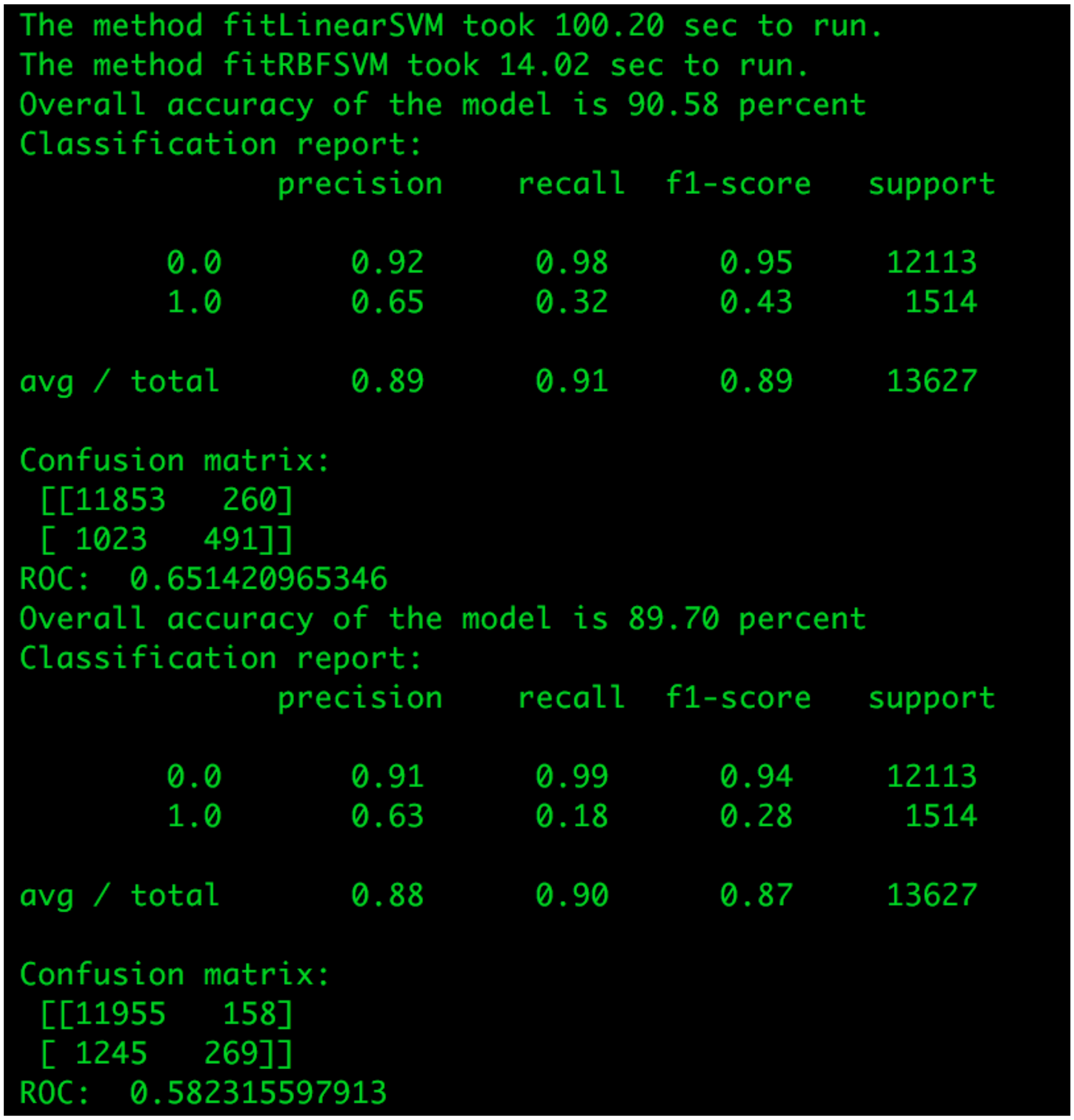

logit
1
0.8
0.6
0.4
0.2
0
-4
-2
0
2
4



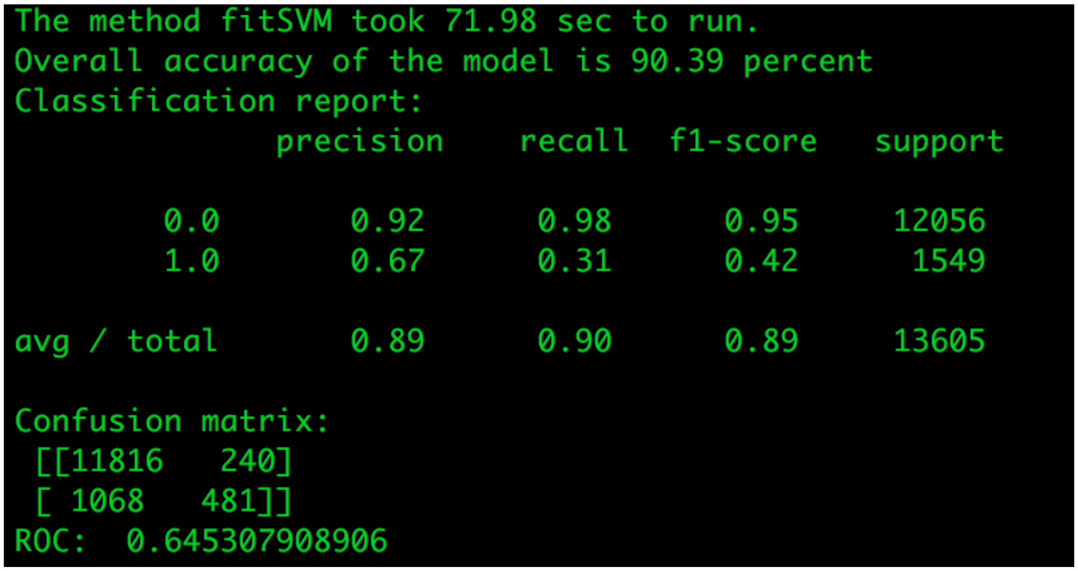
```
The method fitLogisticRegression took 2.02 sec to run.
Overall accuracy of the model is 91.08 percent
Classification report:
             precision    recall  f1-score   support

        0.0       0.93      0.97      0.95     12106
        1.0       0.68      0.42      0.52      1565

avg / total       0.90      0.91      0.90     13671

Confusion matrix:
 [[11788   318]
 [  901   664]]
ROC:  0.699006591931
```



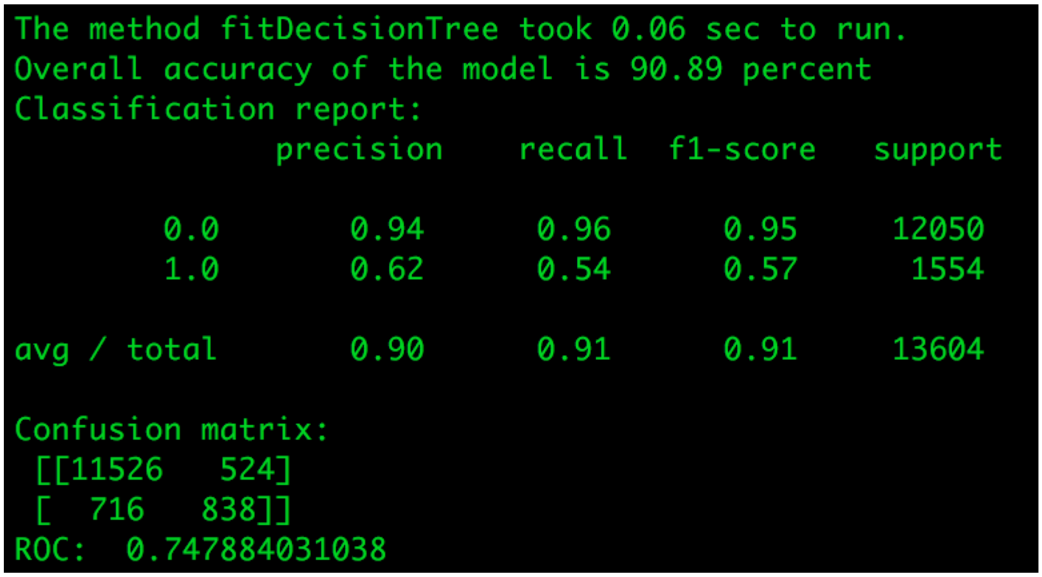
```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:     credit_application   No. Observations:                27517
Model:                            GLM   Df Residuals:                    27465
Model Family:                Binomial   Df Model:                           51
Link Function:                  logit   Scale:                             1.0
Method:                          IRLS   Log-Likelihood:                -5776.2
Date:                Mon, 14 Mar 2016   Deviance:                       11552.
Time:                        21:09:05   Pearson chi2:                 5.25e+07
No. Iterations:                    22
================================================================================================
                                   coef    std err          z      P>|z|      [95.0% Conf. Int.]
------------------------------------------------------------------------------------------------
n_age                            0.1918      0.240      0.798      0.425      -0.279       0.663
n_duration                      22.5235      0.441     51.024      0.000      21.658      23.389
n_pdays                         -1.0472      0.273     -3.832      0.000      -1.583      -0.512
n_previous                      -0.2053      0.505     -0.406      0.685      -1.196       0.785
```



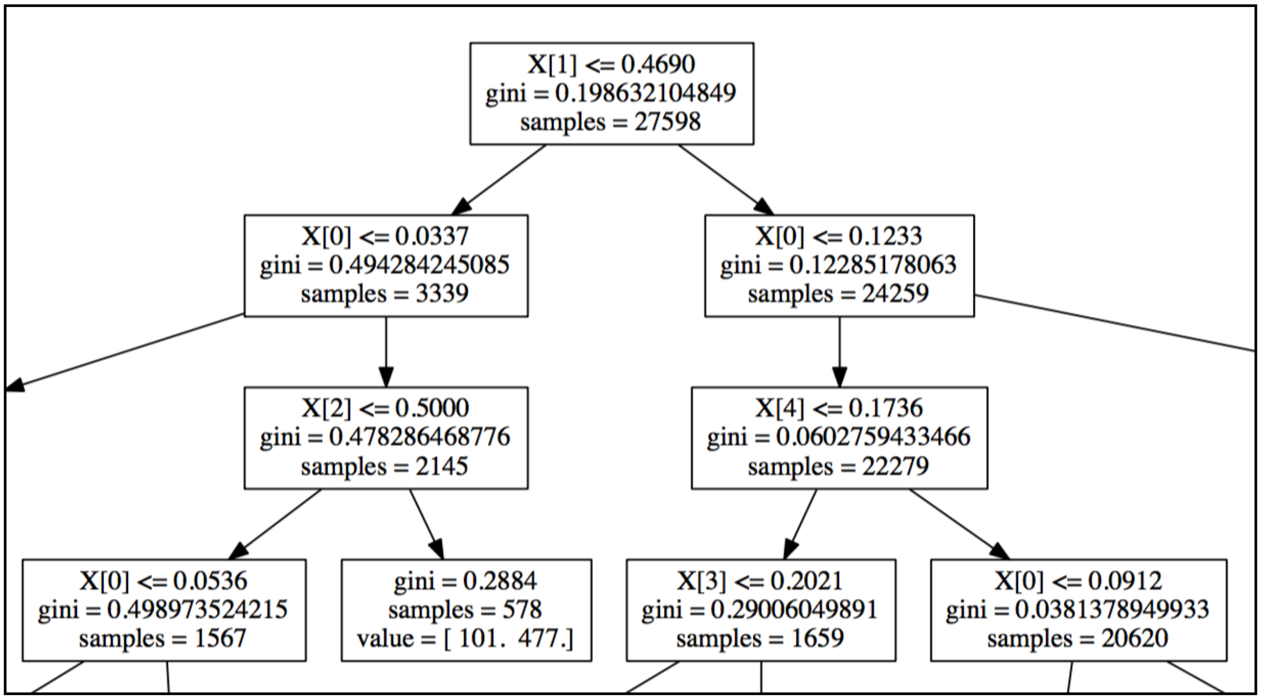
```
The method fitLinearSVM took 100.20 sec to run.
The method fitRBFSVM took 14.02 sec to run.
Overall accuracy of the model is 90.58 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.92      0.98      0.95     12113
        1.0       0.65      0.32      0.43      1514

avg / total       0.89      0.91      0.89     13627

Confusion matrix:
 [[11853   260]
 [ 1023   491]]
ROC:  0.651420965346
Overall accuracy of the model is 89.70 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.91      0.99      0.94     12113
        1.0       0.63      0.18      0.28      1514

avg / total       0.88      0.90      0.87     13627

Confusion matrix:
 [[11955   158]
 [ 1245   269]]
ROC:  0.582315597913
```

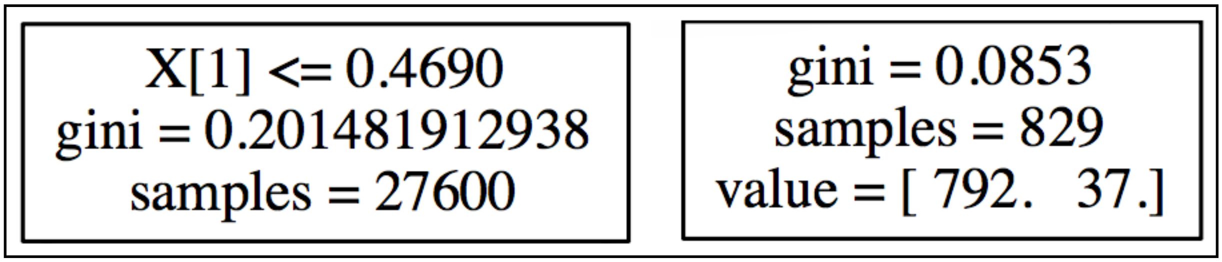


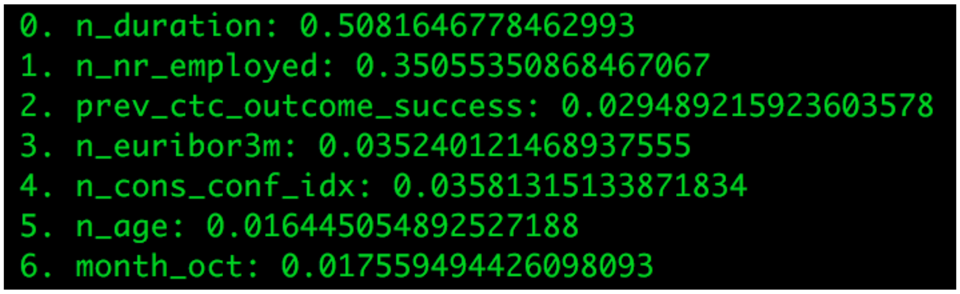
```
The method fitSVM took 71.98 sec to run.
Overall accuracy of the model is 90.39 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.92      0.98      0.95     12056
        1.0       0.67      0.31      0.42      1549

avg / total       0.89      0.90      0.89     13605

Confusion matrix:
 [[11816   240]
 [ 1068   481]]
ROC:  0.645307908906
```



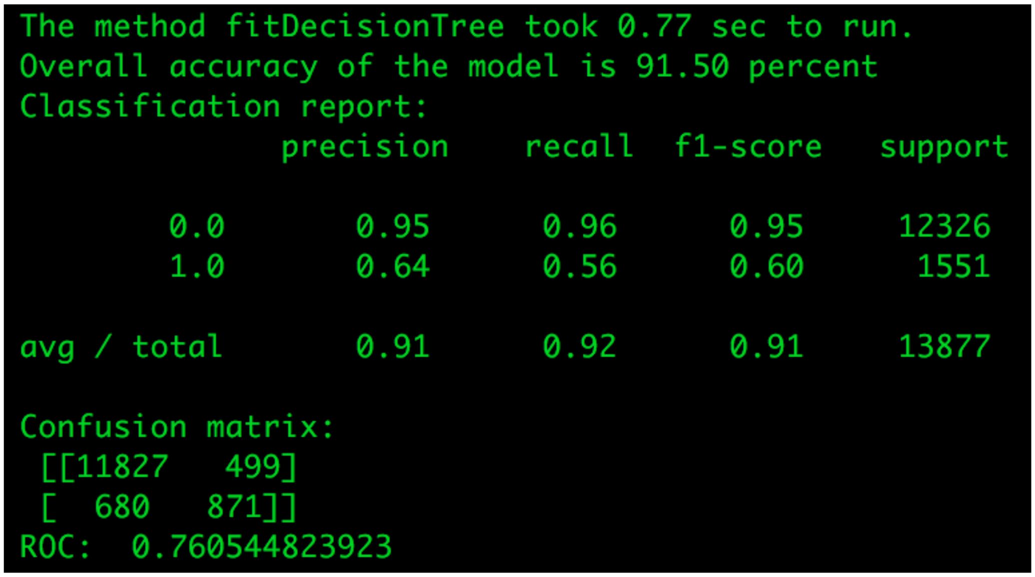
```
The method fitDecisionTree took 0.06 sec to run.
Overall accuracy of the model is 90.89 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.94      0.96      0.95     12050
        1.0       0.62      0.54      0.57      1554

avg / total       0.90      0.91      0.91     13604

Confusion matrix:
 [[11526   524]
 [  716   838]]
ROC:  0.747884031038
```



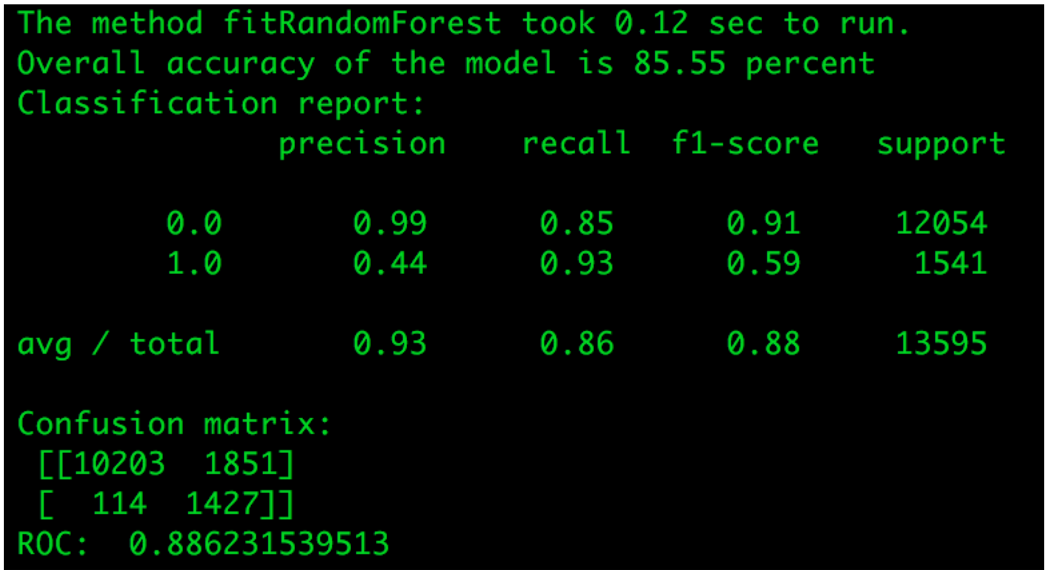




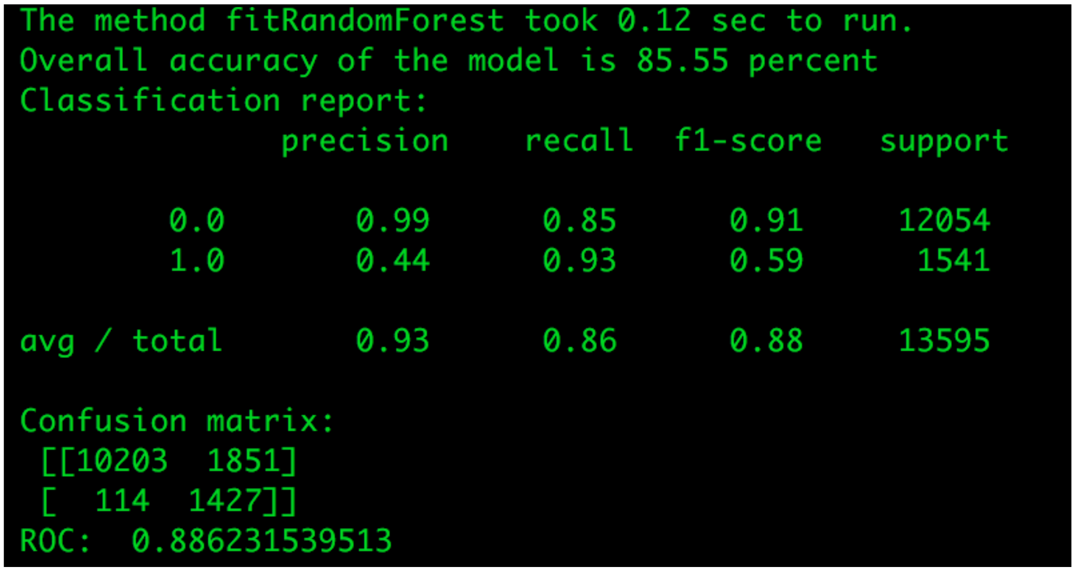
```
0. n_duration: 0.5081646778462993
1. n_nr_employed: 0.35055350868467067
2. prev_ctc_outcome_success: 0.029489215923603578
3. n_euribor3m: 0.035240121468937555
4. n_cons_conf_idx: 0.03581315133871834
5. n_age: 0.016445054892527188
6. month_oct: 0.017559494426098093
```



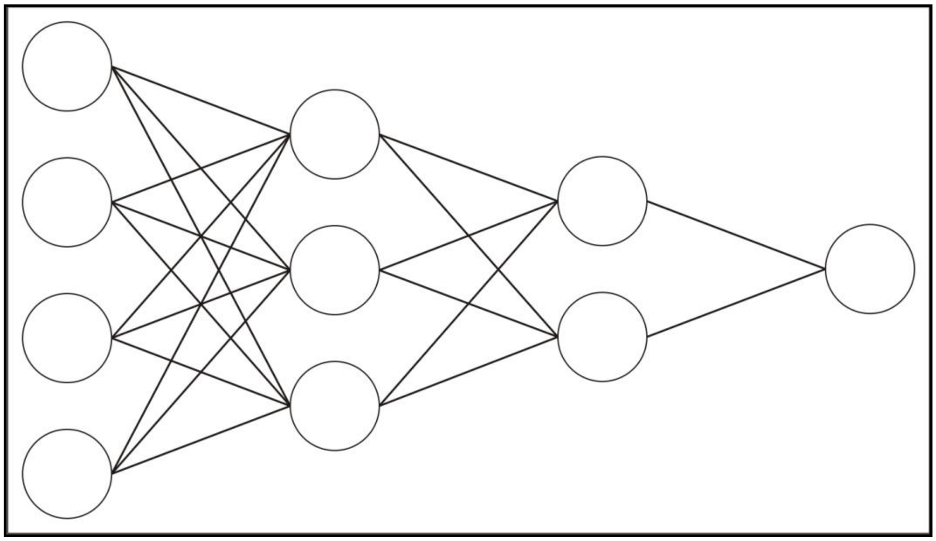
```
The method fitDecisionTree took 0.77 sec to run.
Overall accuracy of the model is 91.50 percent
Classification report:
              precision    recall  f1-score   support


        0.0       0.95      0.96      0.95     12326
        1.0       0.64      0.56      0.60      1551

avg / total       0.91      0.92      0.91     13877

Confusion matrix:
 [[11827   499]
 [  680   871]]
ROC:  0.760544823923
```

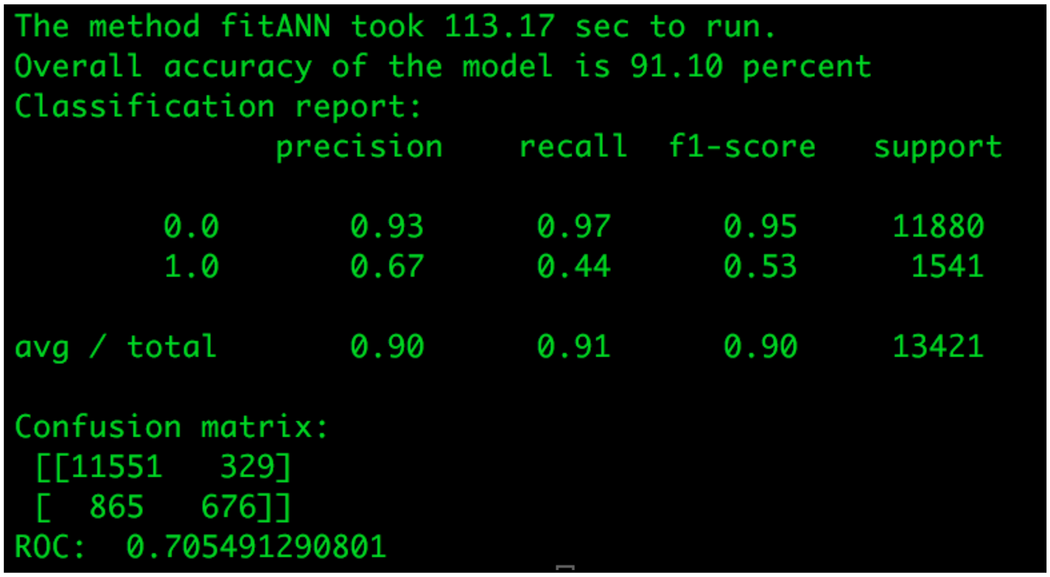


```
The method fitRandomForest took 0.12 sec to run.
Overall accuracy of the model is 85.55 percent
Classification report:
              precision    recall  f1-score   support


        0.0       0.99      0.85      0.91     12054
        1.0       0.44      0.93      0.59      1541

avg / total       0.93      0.86      0.88     13595

Confusion matrix:
 [[10203  1851]
 [  114  1427]]
ROC:  0.886231539513
```

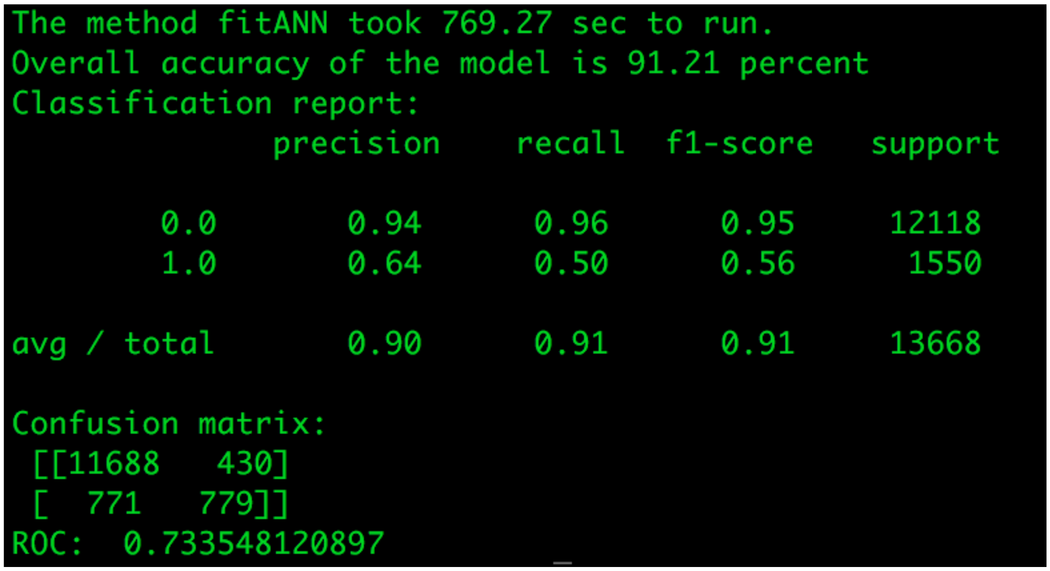


```
The method fitRandomForest took 0.12 sec to run.
Overall accuracy of the model is 85.55 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.99      0.85      0.91     12054
        1.0       0.44      0.93      0.59      1541

avg / total       0.93      0.86      0.88     13595

Confusion matrix:
 [[10203  1851]
 [  114  1427]]
ROC:  0.886231539513
```

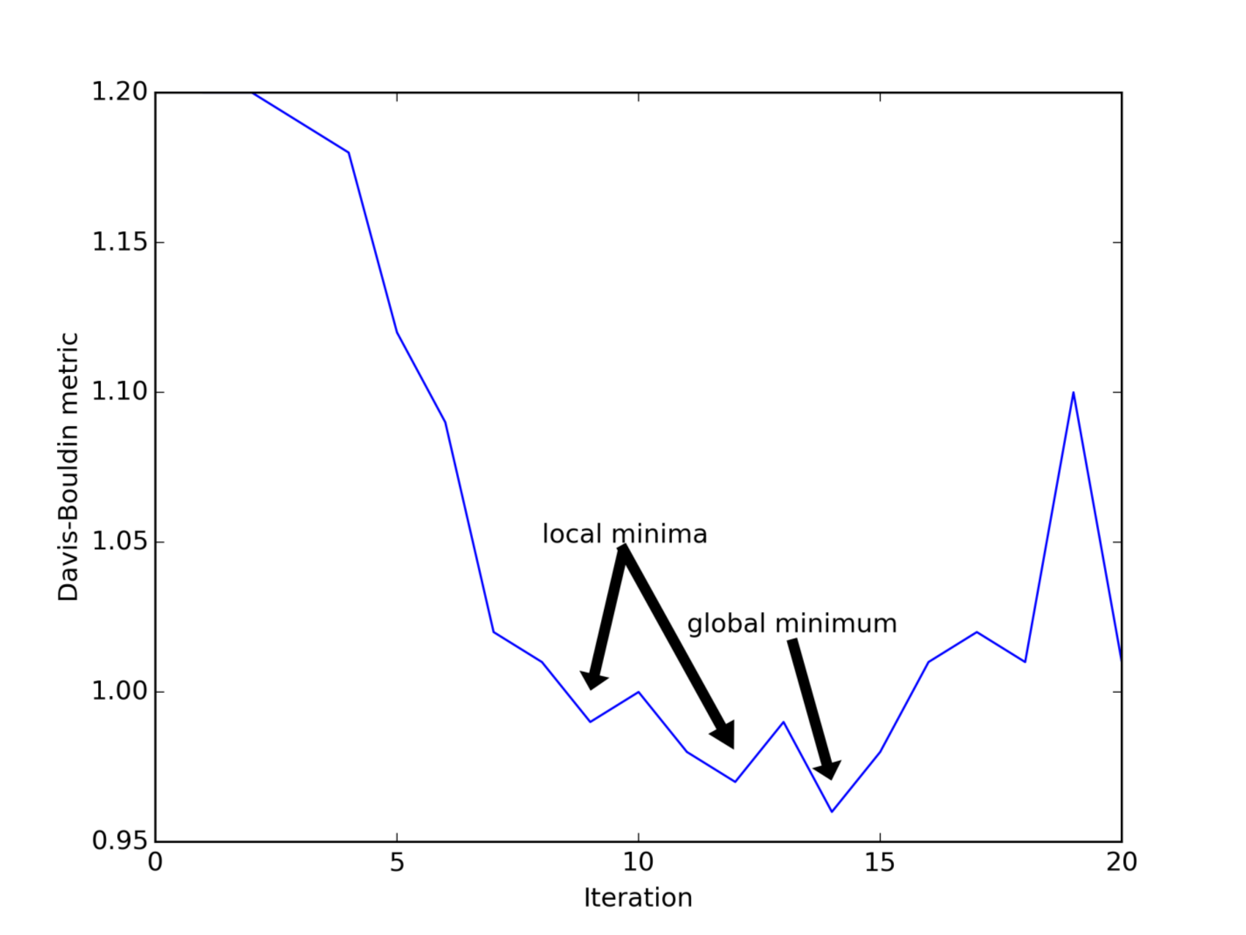


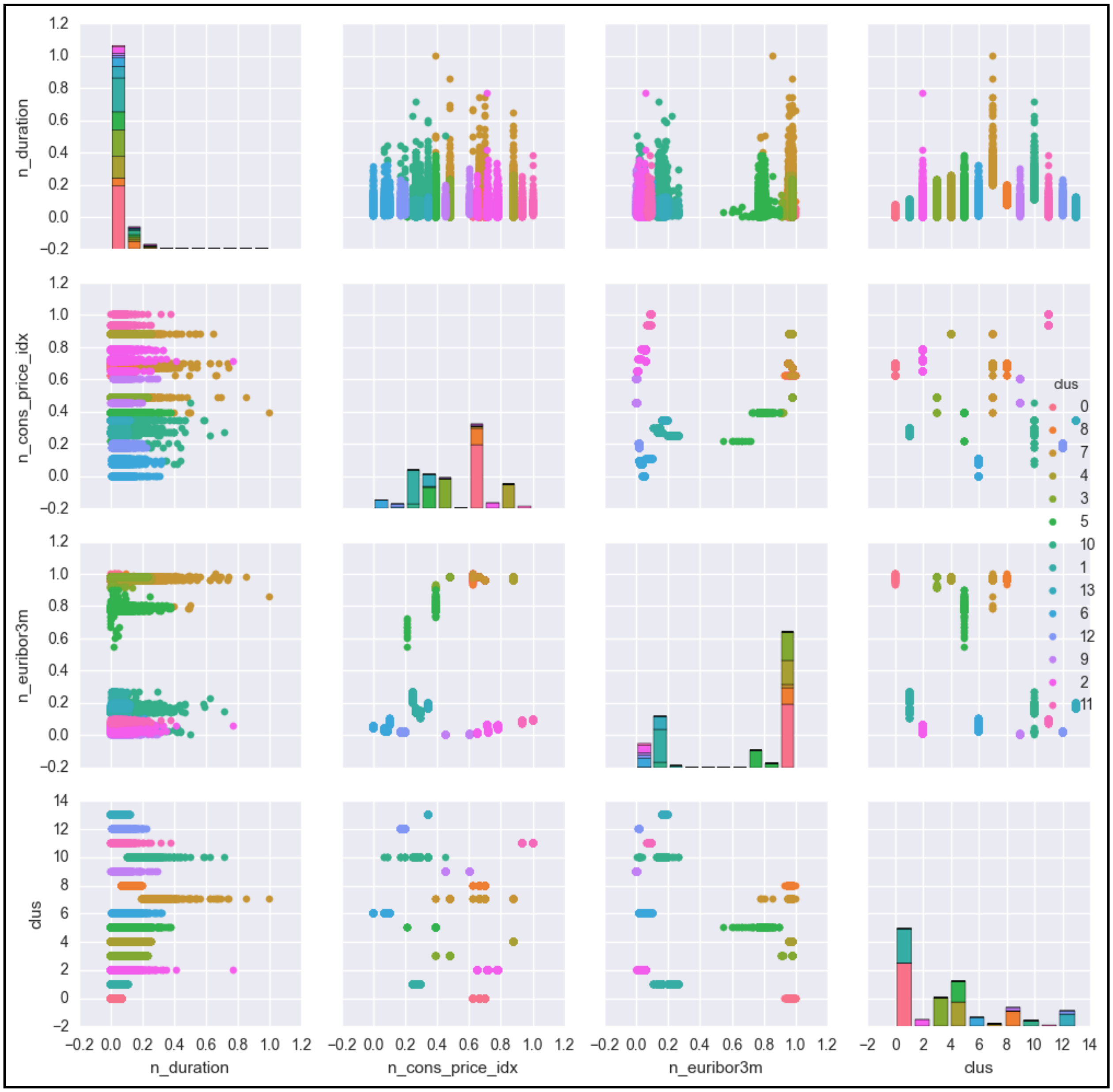


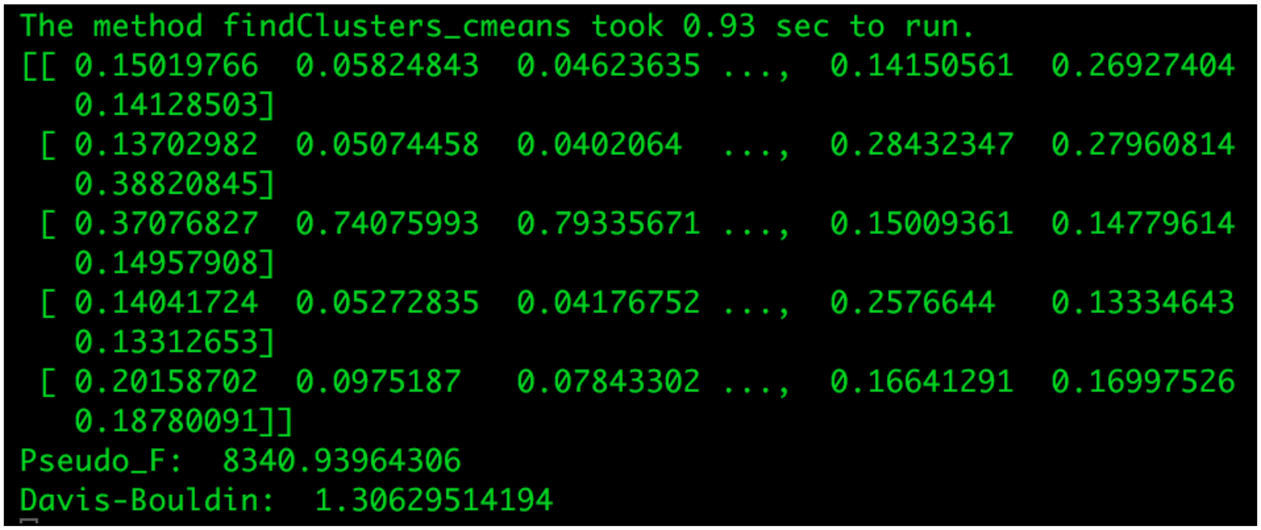
```
The method fitANN took 113.17 sec to run.
Overall accuracy of the model is 91.10 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.93      0.97      0.95     11880
        1.0       0.67      0.44      0.53      1541

avg / total       0.90      0.91      0.90     13421

Confusion matrix:
 [[11551   329]
 [  865   676]]
ROC:  0.705491290801
```



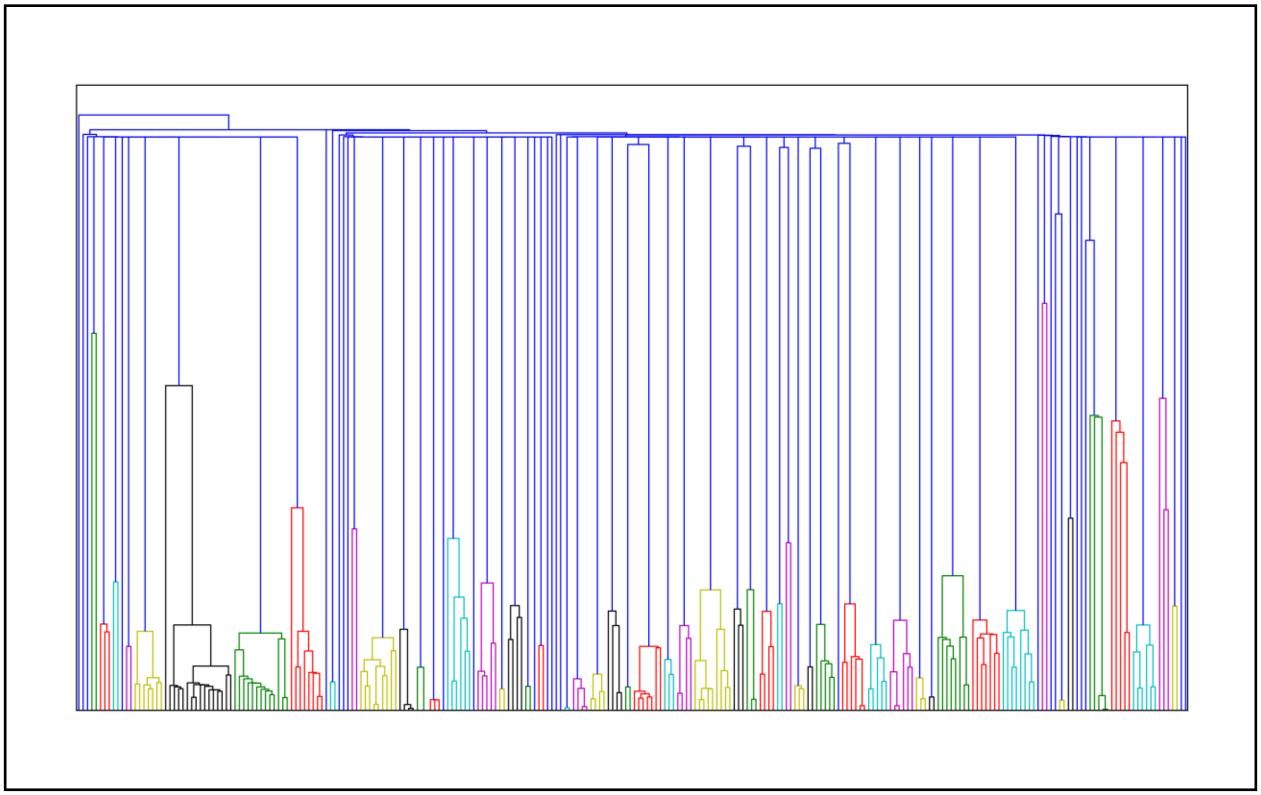
```
The method fitANN took 769.27 sec to run.
Overall accuracy of the model is 91.21 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.94      0.96      0.95     12118
        1.0       0.64      0.50      0.56      1550

avg / total       0.90      0.91      0.91     13668

Confusion matrix:
 [[11688   430]
 [  771   779]]
ROC:  0.733548120897
```

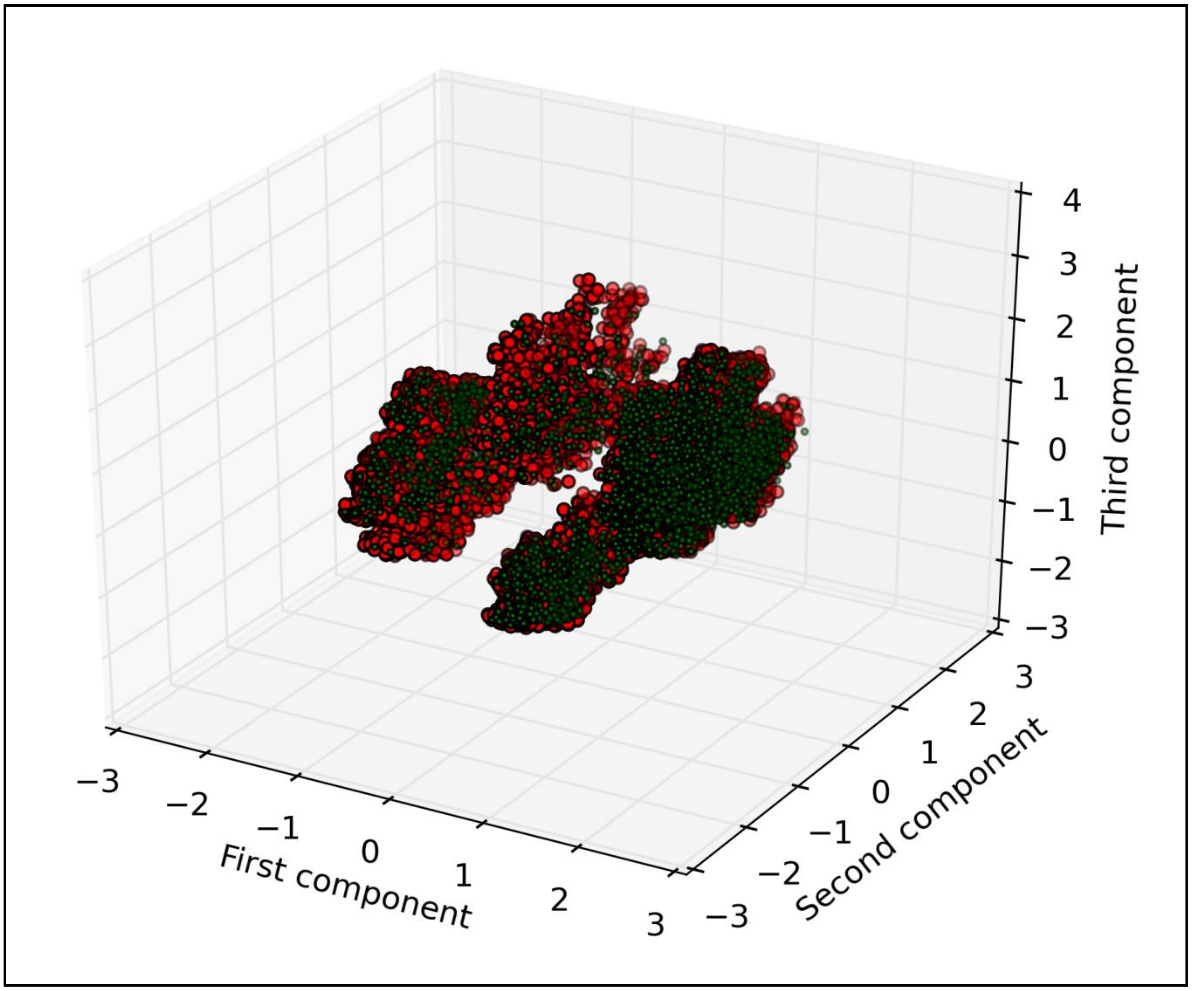

1.20
1.15
1.10
1.05
1.00
0.95
Davis-Bouldin metric
0
5
10
15
20
Iteration
local minima
global minimum

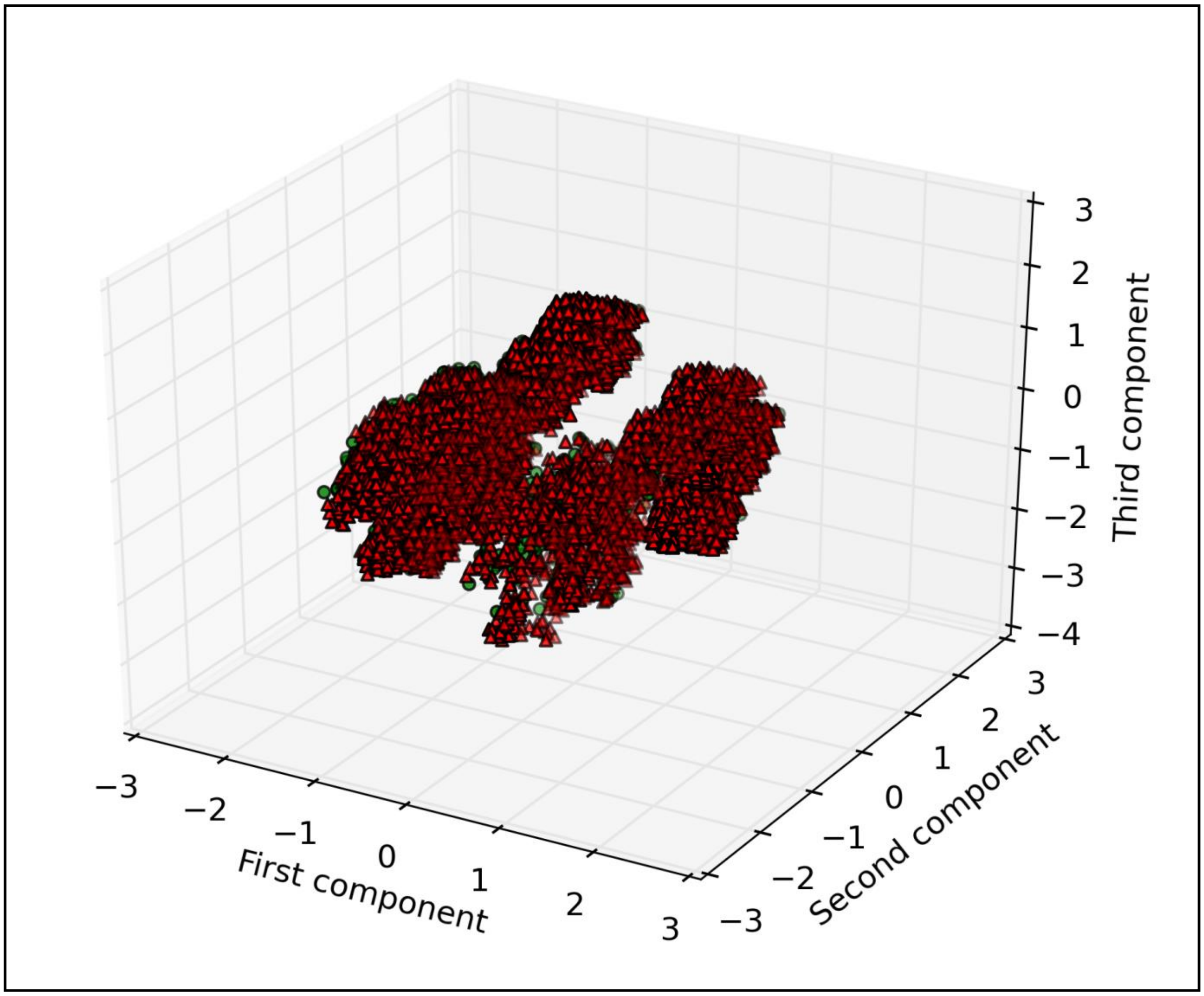

n_duration
n_cons_price_idx
n_euribor3m
clus
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
14
12
10
8
6
4
2
0
−2
−0.2 0.0 0.2 0.4 0.6 0.8 1.0 1.2
−2 0 2 4 6 8 10 12 14
clus
0
8
7
4
3
5
10
1
13
6
12
9
2
11

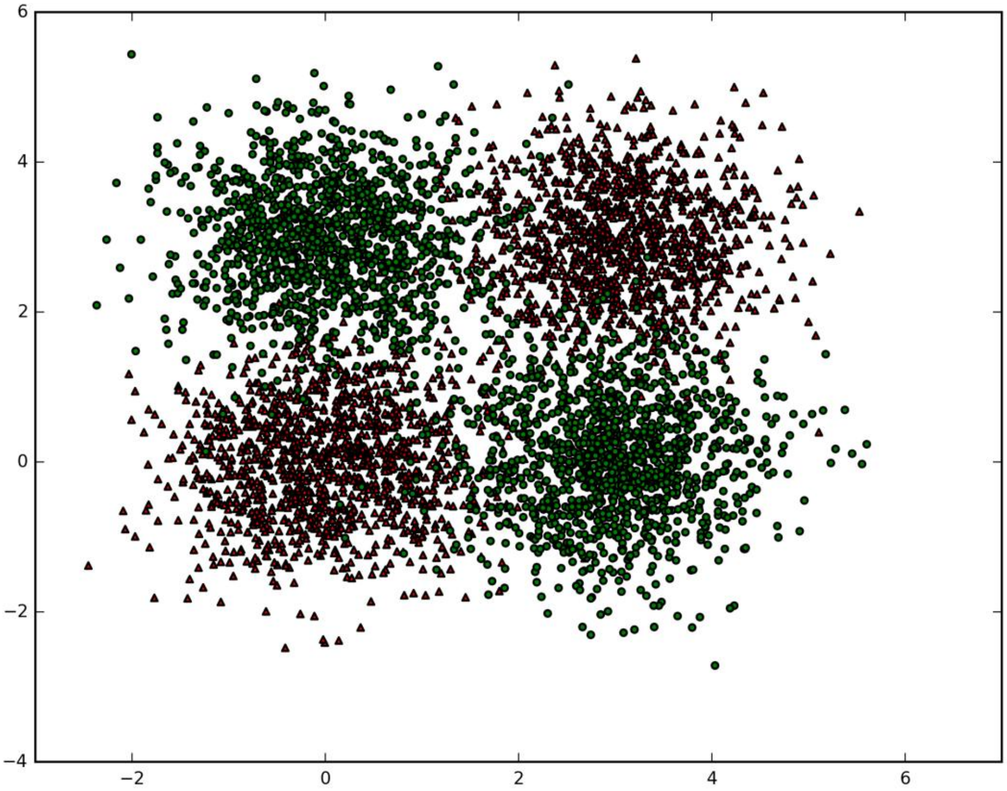


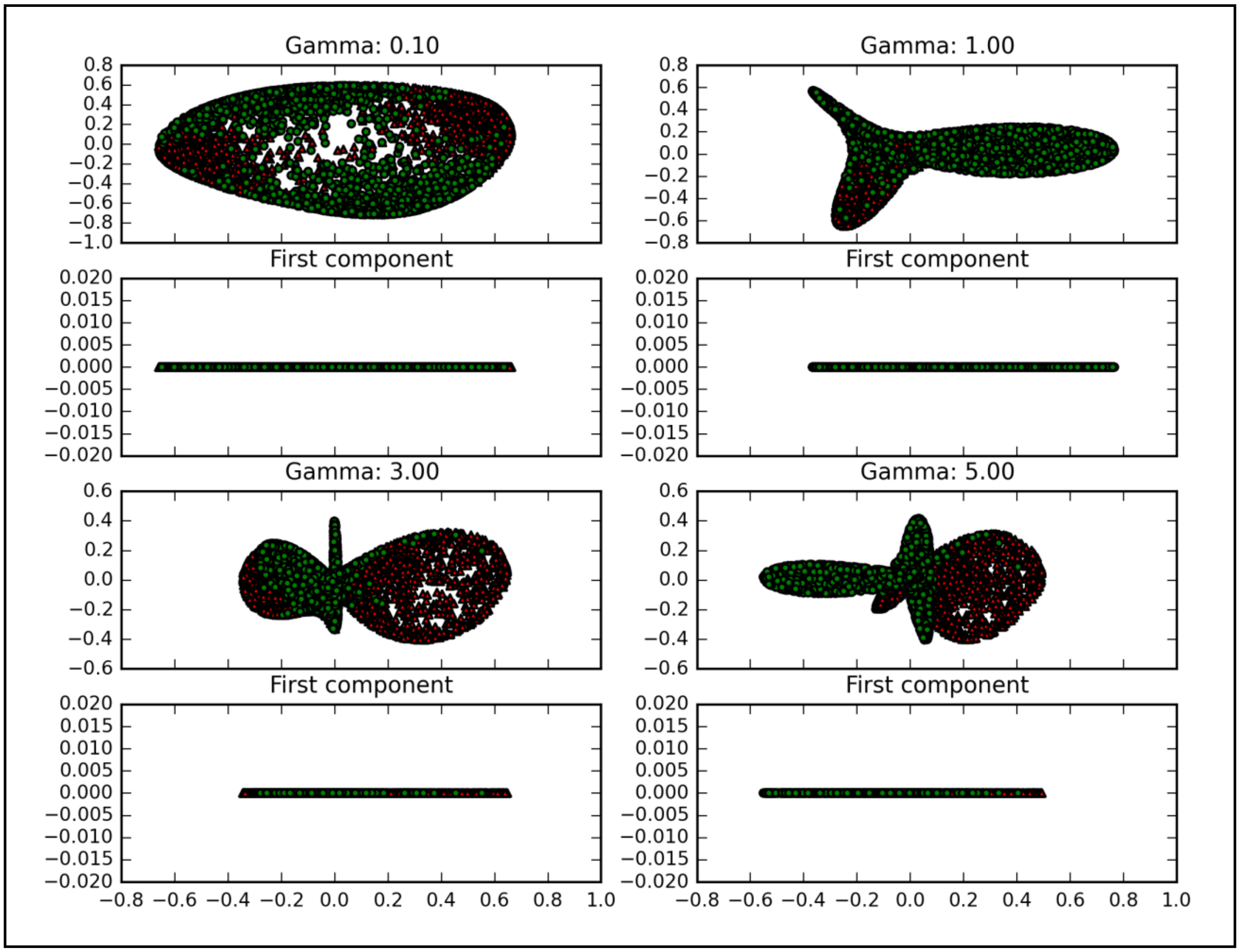
```
The method findClusters_cmeans took 0.93 sec to run.
[[ 0.15019766  0.05824843  0.04623635 ...,  0.14150561  0.26927404
   0.14128503]
 [ 0.13702982  0.05074458  0.0402064  ...,  0.28432347  0.27960814
   0.38820845]
 [ 0.37076827  0.74075993  0.79335671 ...,  0.15009361  0.14779614
   0.14957908]
 [ 0.14041724  0.05272835  0.04176752 ...,  0.2576644   0.13334643
   0.13312653]
 [ 0.20158702  0.0975187   0.07843302 ...,  0.16641291  0.16997526
   0.18780091]]
Pseudo_F:  8340.93964306
Davis-Bouldin:  1.30629514194
```

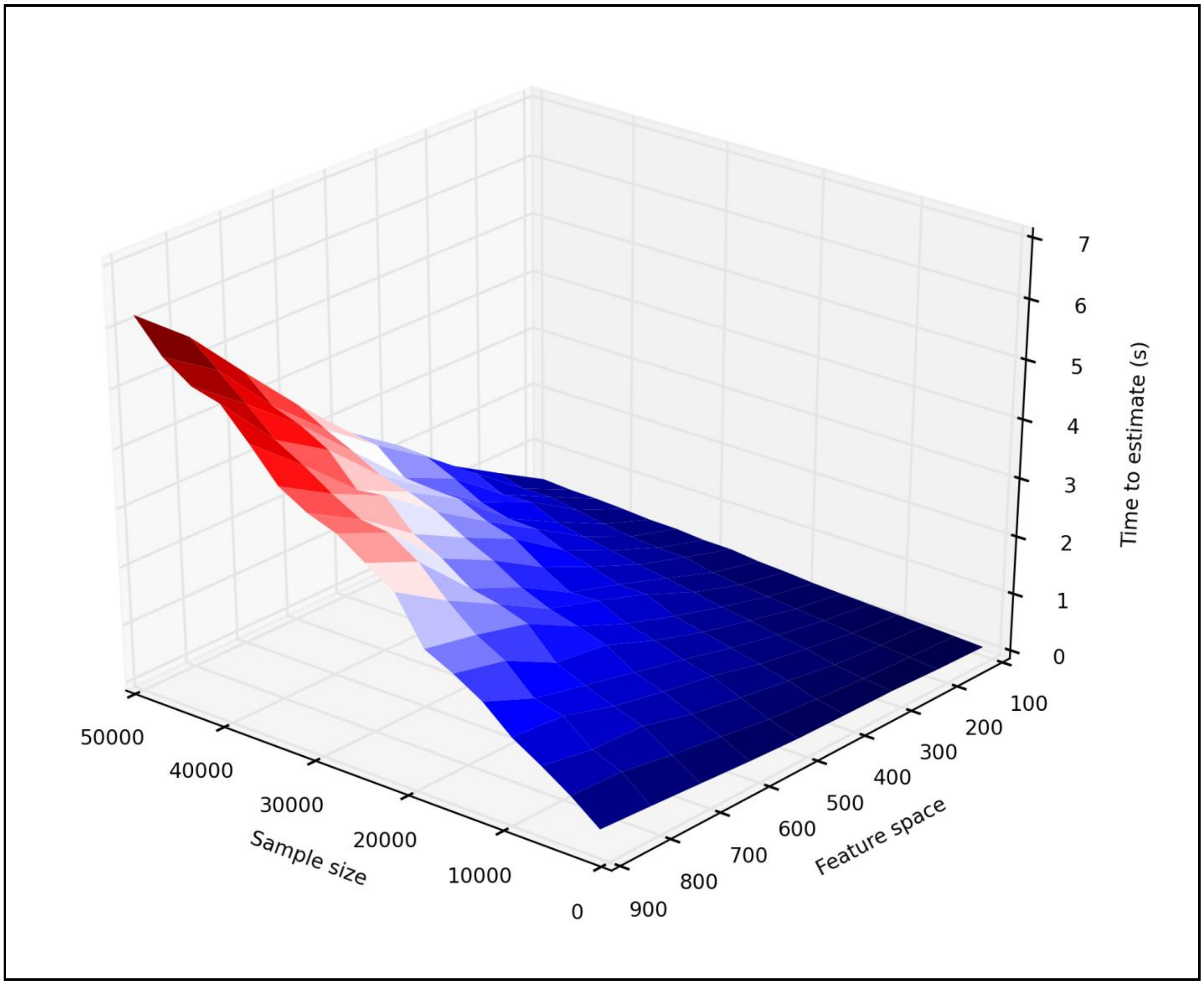


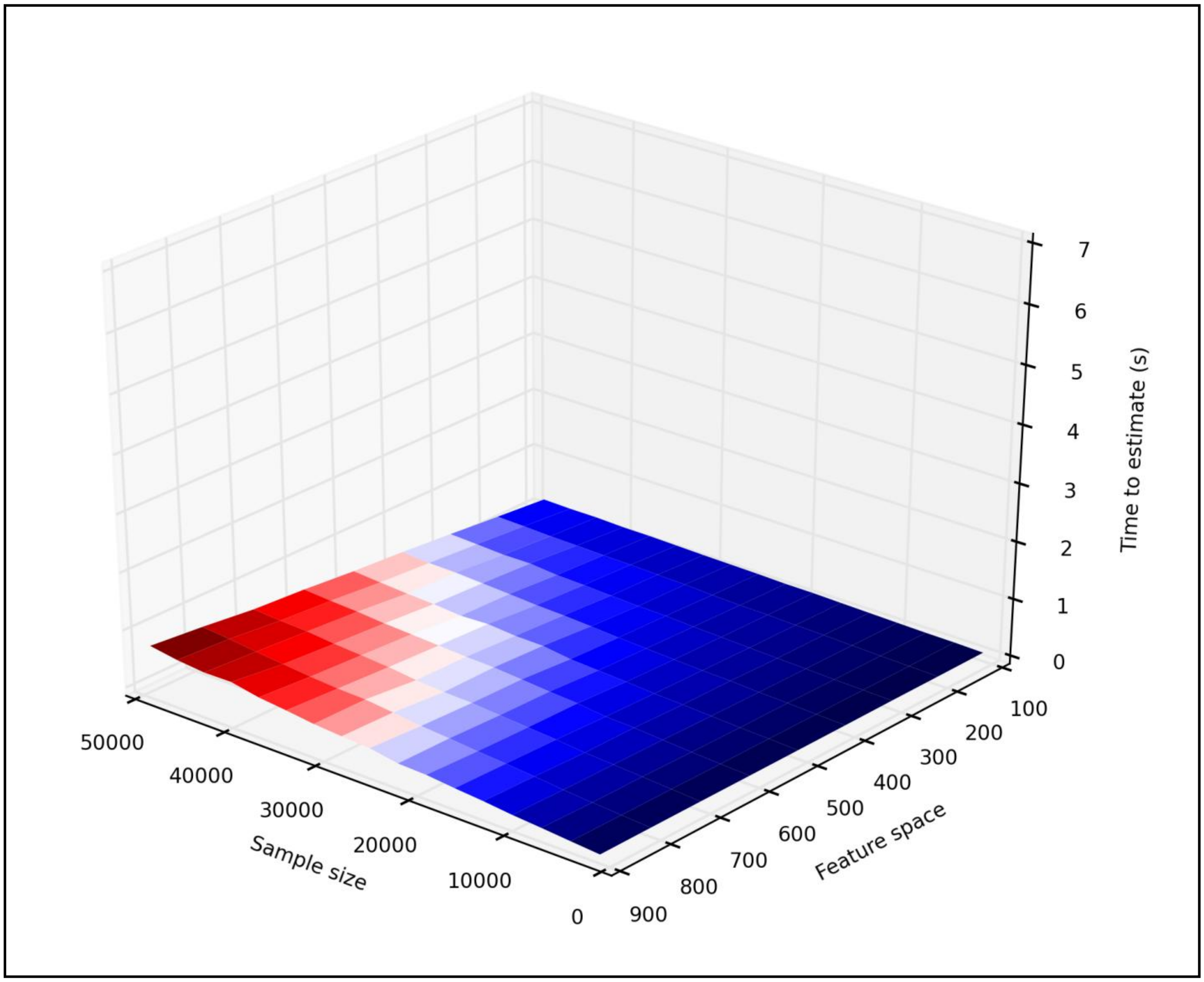

First component
−3
−2
−1
0
1
2
3
Second component
−3
−2
−1
0
1
2
3
Third component
−3
−2
−1
0
1
2
3
4

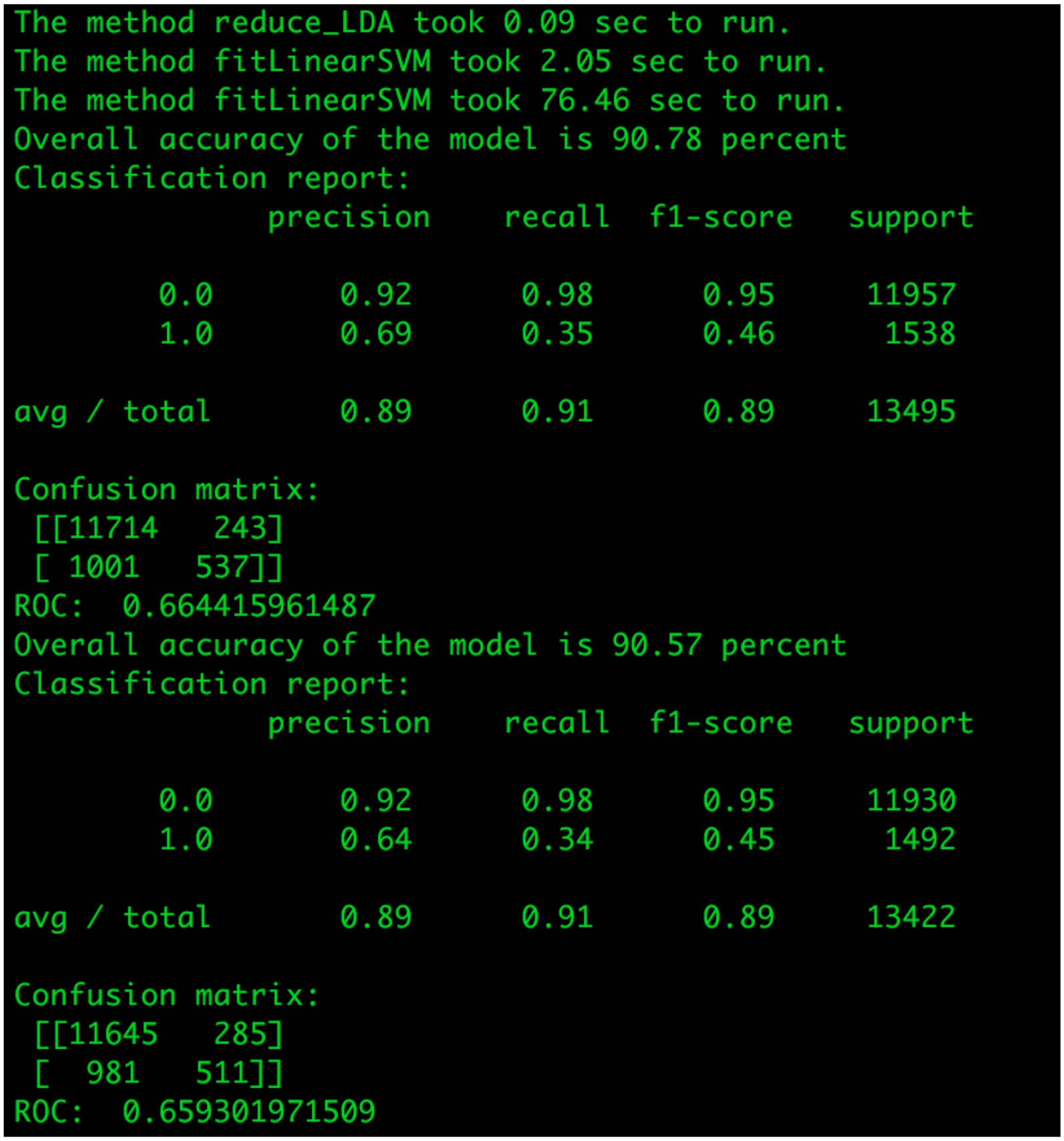

First component
−3
−2
−1
0
1
2
3
Second component
−3
−2
−1
0
1
2
3
Third component
3
2
1
0
−1
−2
−3
−4


6
4
2
0
−2
−4
−2
0
2
4
6

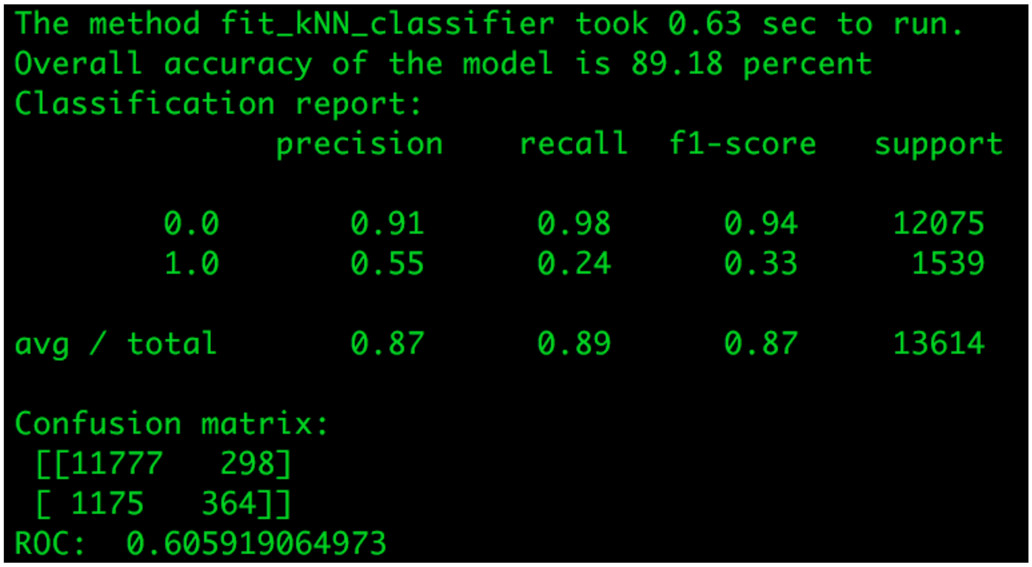

Gamma: 0.10
First component
Gamma: 1.00
First component
Gamma: 3.00
First component
Gamma: 5.00
First component
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
0.020
0.015
0.010
0.005
0.000
−0.005
−0.010
−0.015
−0.020
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0


Time to estimate (s)
0
1
2
3
4
5
6
7
Sample size
50000
40000
30000
20000
10000
0
Feature space
100
200
300
400
500
600
700
800
900

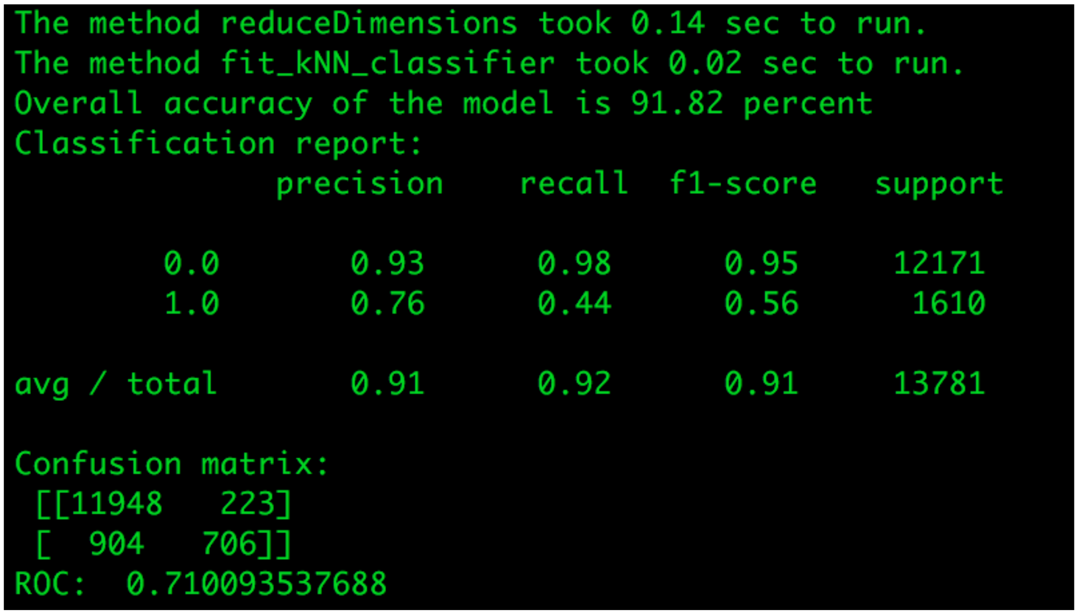

Time to estimate (s)
0
1
2
3
4
5
6
7
Sample size
50000
40000
30000
20000
10000
0
Feature space
100
200
300
400
500
600
700
800
900



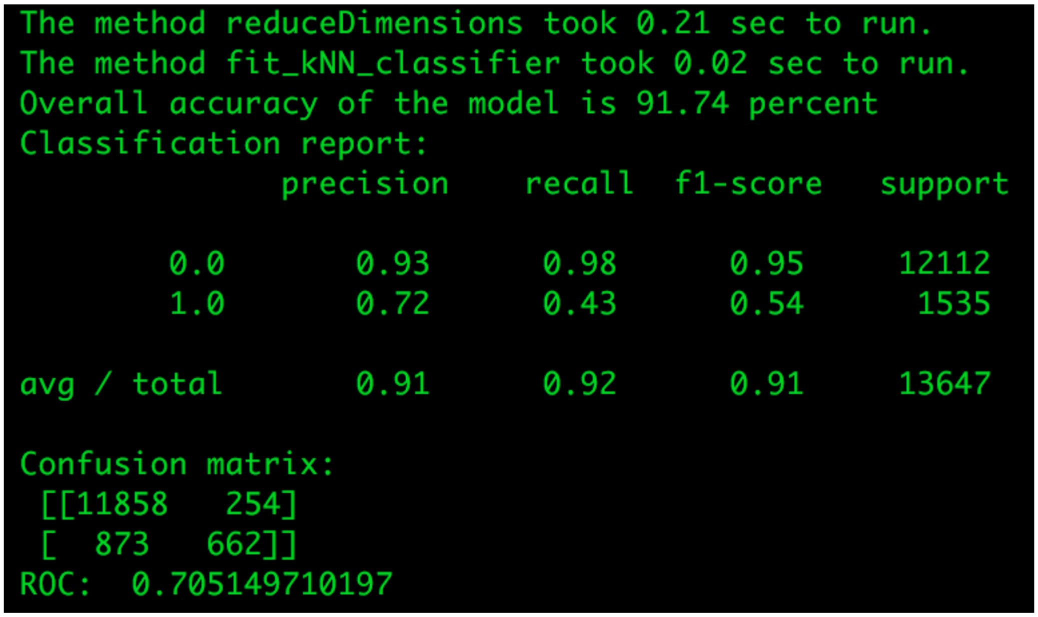
```
The method reduce_LDA took 0.09 sec to run.
The method fitLinearSVM took 2.05 sec to run.
The method fitLinearSVM took 76.46 sec to run.
Overall accuracy of the model is 90.78 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.92      0.98      0.95     11957
        1.0       0.69      0.35      0.46      1538

avg / total       0.89      0.91      0.89     13495

Confusion matrix:
 [[11714   243]
 [ 1001   537]]
ROC:  0.664415961487
Overall accuracy of the model is 90.57 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.92      0.98      0.95     11930
        1.0       0.64      0.34      0.45      1492

avg / total       0.89      0.91      0.89     13422

Confusion matrix:
 [[11645   285]
 [  981   511]]
ROC:  0.659301971509
```



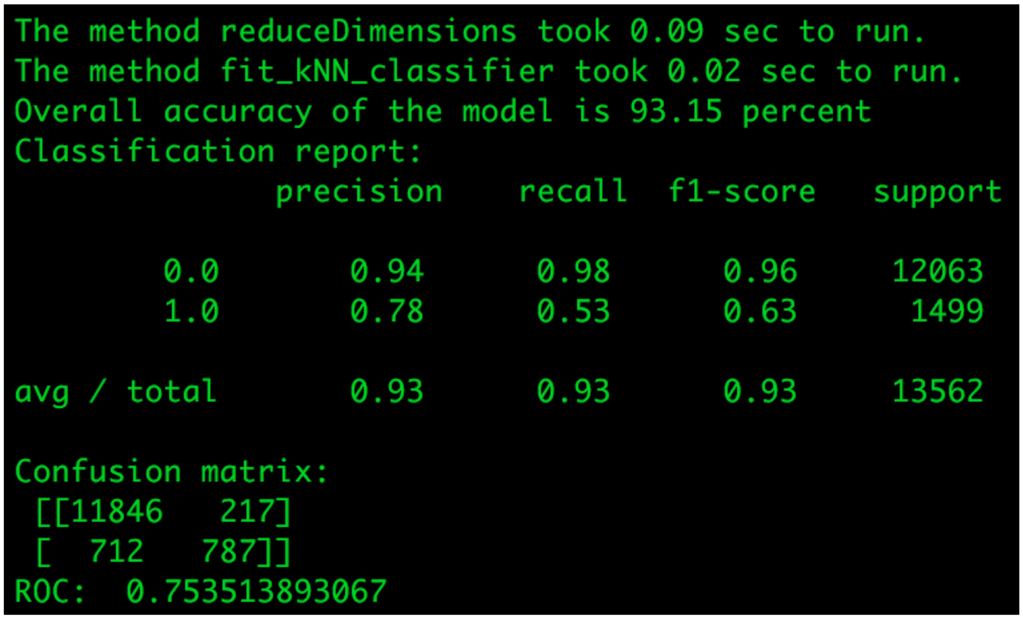
```
The method fit_kNN_classifier took 0.63 sec to run.
Overall accuracy of the model is 89.18 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.91      0.98      0.94     12075
        1.0       0.55      0.24      0.33      1539

avg / total       0.87      0.89      0.87     13614

Confusion matrix:
 [[11777   298]
 [ 1175   364]]
ROC:  0.605919064973
```



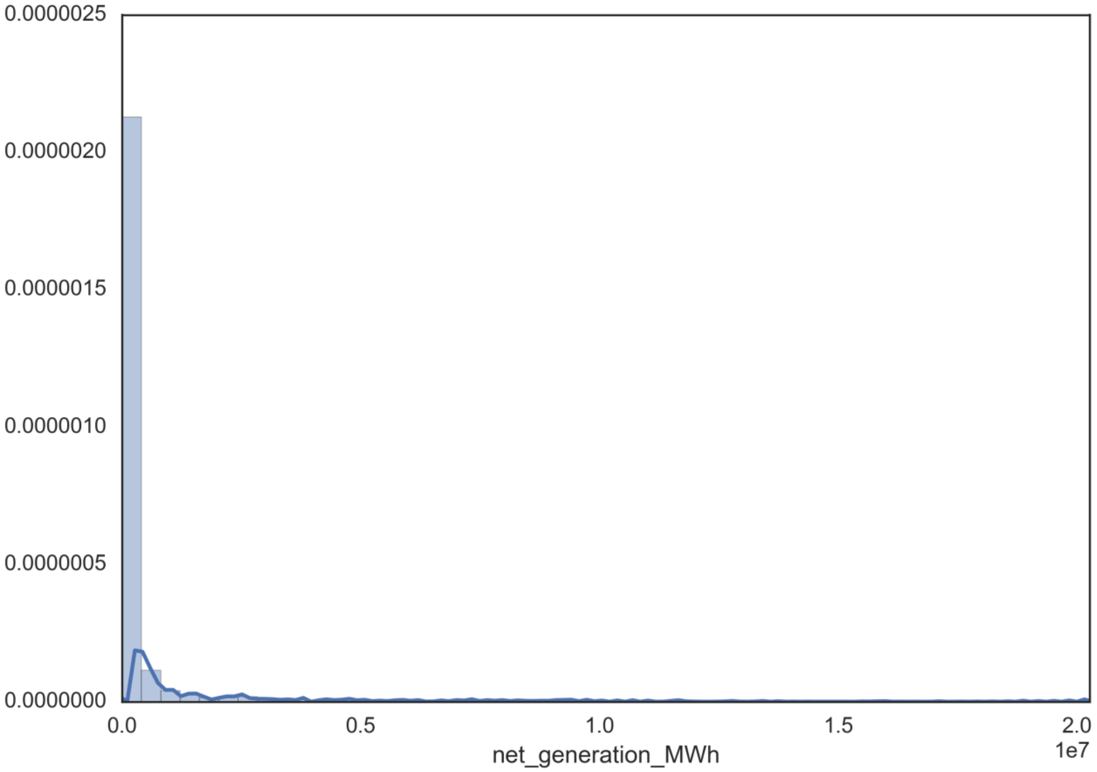
```
The method reduceDimensions took 0.14 sec to run.
The method fit_kNN_classifier took 0.02 sec to run.
Overall accuracy of the model is 91.82 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.93      0.98      0.95     12171
        1.0       0.76      0.44      0.56      1610

avg / total       0.91      0.92      0.91     13781

Confusion matrix:
 [[11948   223]
 [  904   706]]
ROC:  0.710093537688
```

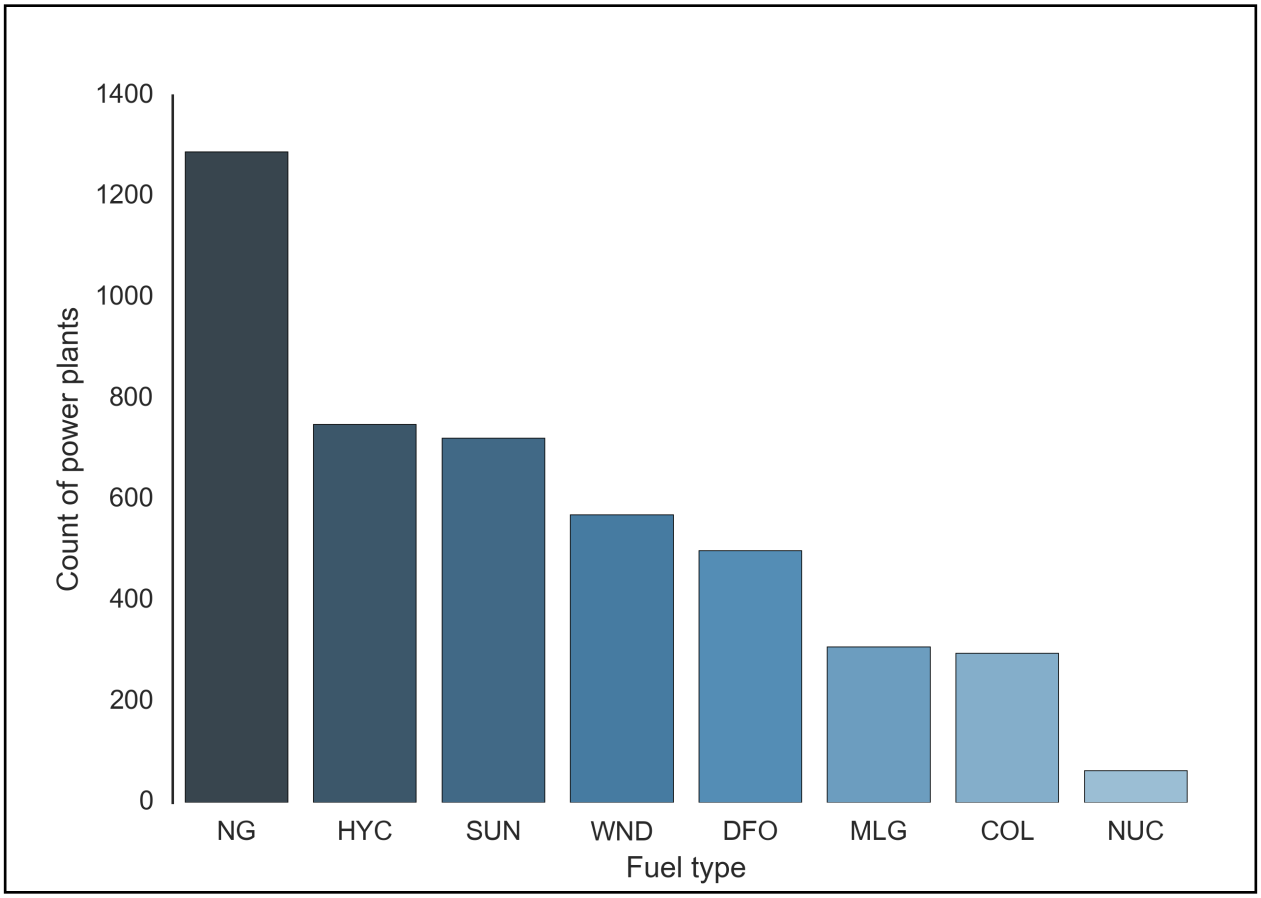


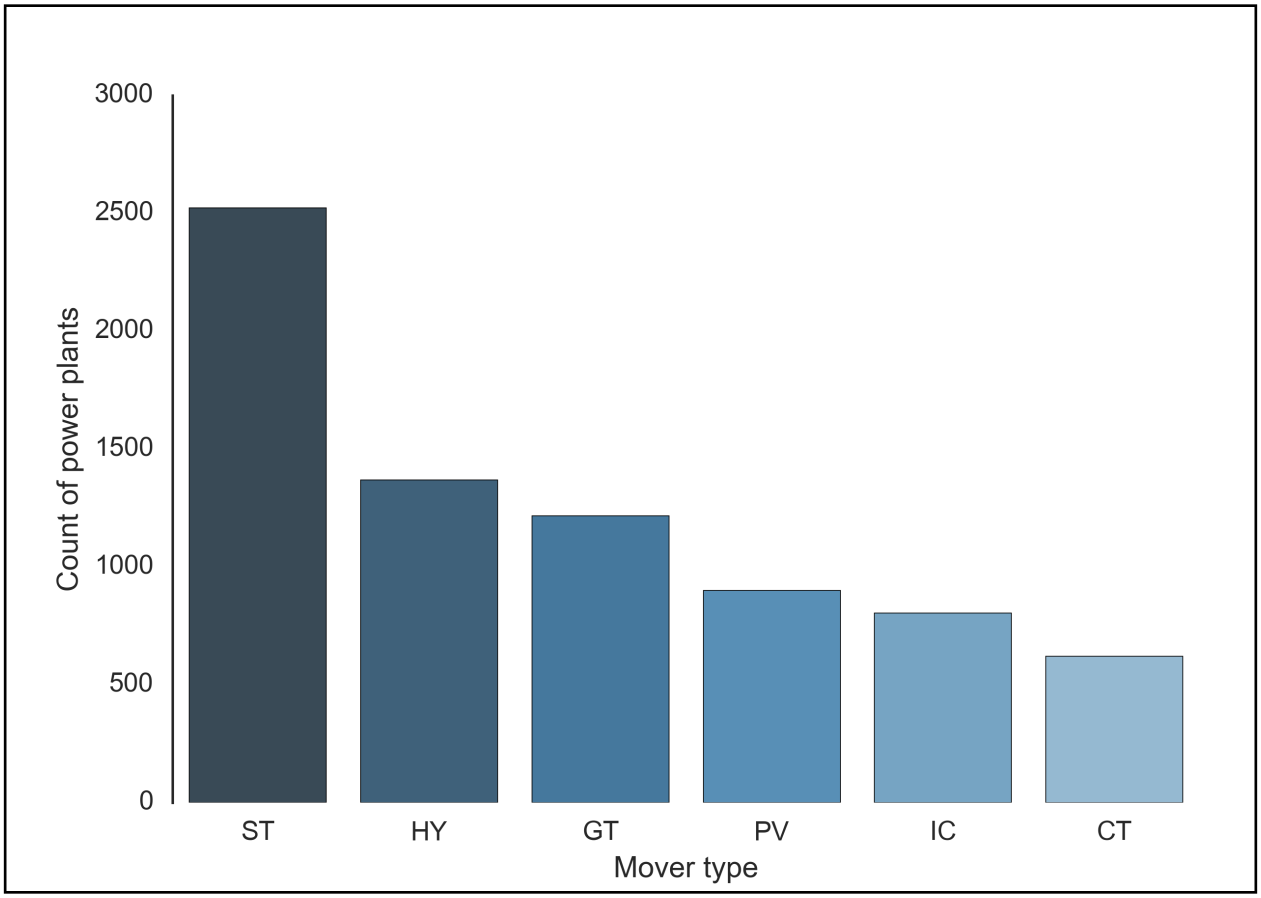
```
The method reduceDimensions took 0.21 sec to run.
The method fit_kNN_classifier took 0.02 sec to run.
Overall accuracy of the model is 91.74 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.93      0.98      0.95     12112
        1.0       0.72      0.43      0.54      1535

avg / total       0.91      0.92      0.91     13647

Confusion matrix:
 [[11858   254]
 [  873   662]]
ROC:  0.705149710197
```

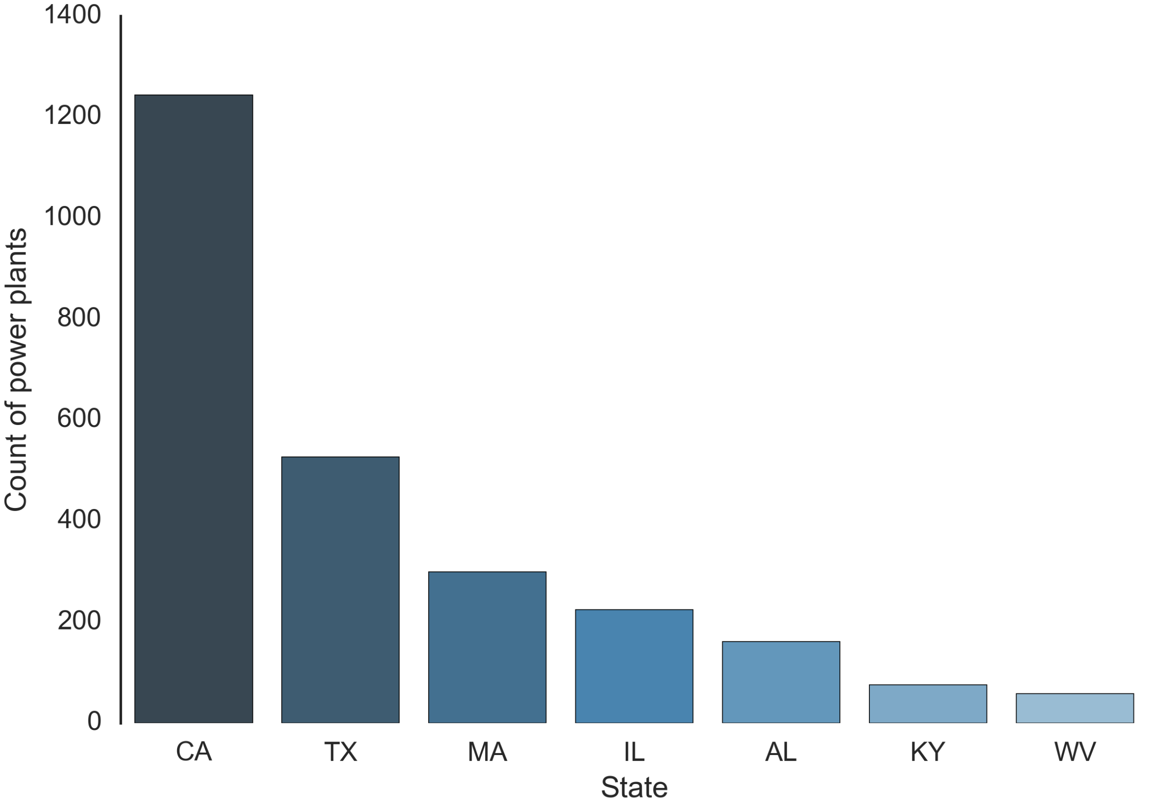


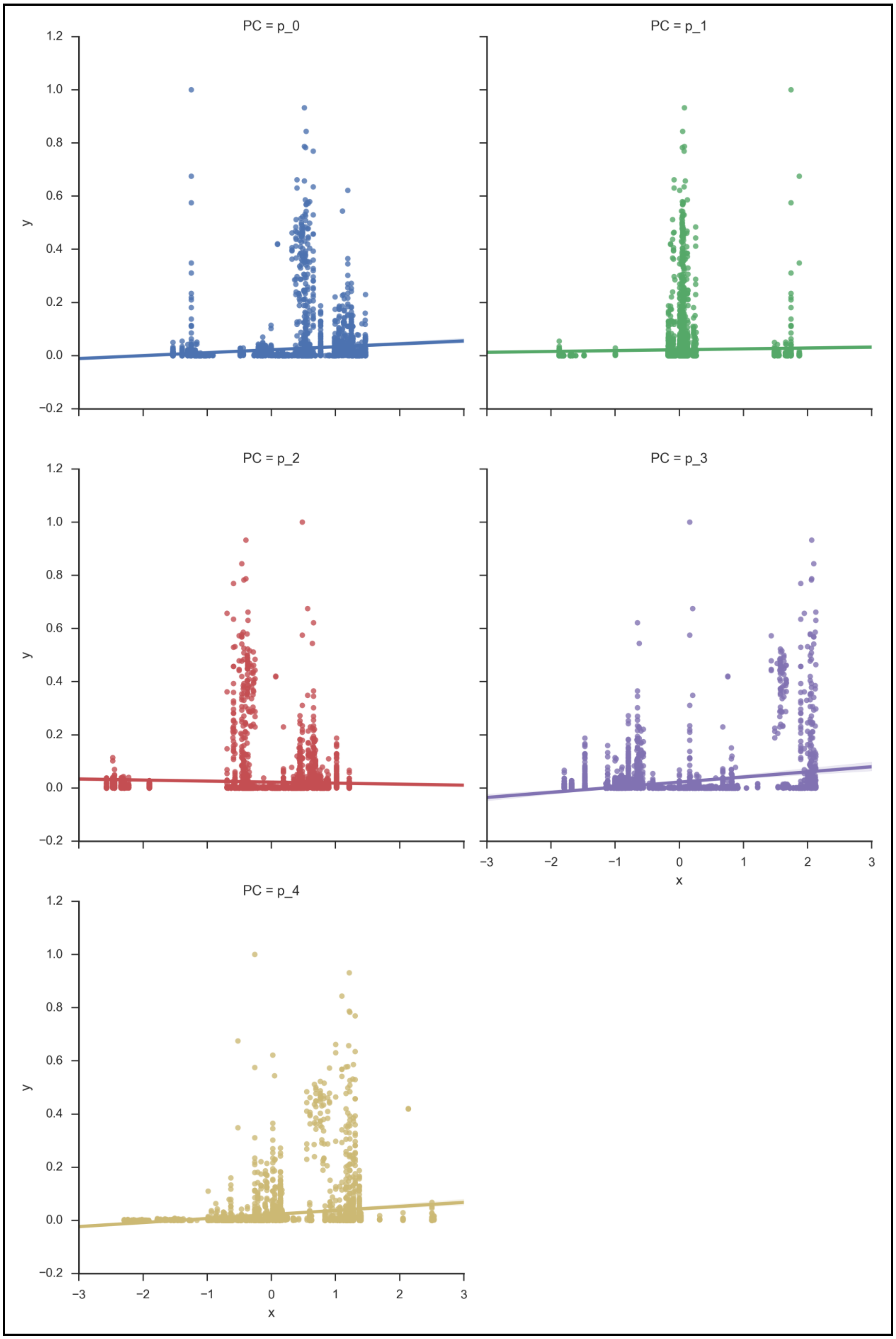
```
The method reduceDimensions took 0.09 sec to run.
The method fit_kNN_classifier took 0.02 sec to run.
Overall accuracy of the model is 93.15 percent
Classification report:
              precision    recall  f1-score   support

        0.0       0.94      0.98      0.96     12063
        1.0       0.78      0.53      0.63      1499

avg / total       0.93      0.93      0.93     13562

Confusion matrix:
 [[11846   217]
 [  712   787]]
ROC:  0.753513893067
```

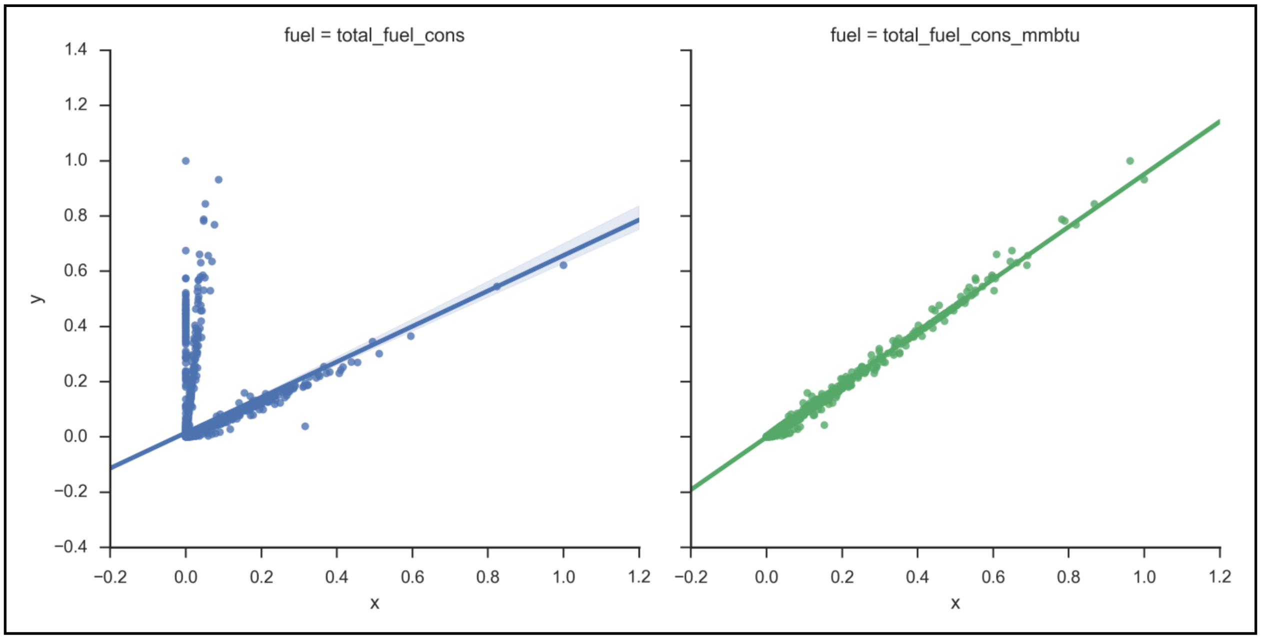

0.0000025
0.0000020
0.0000015
0.0000010
0.0000005
0.0000000
0.0
0.5
1.0
1.5
2.0
1e7
net_generation_MWh

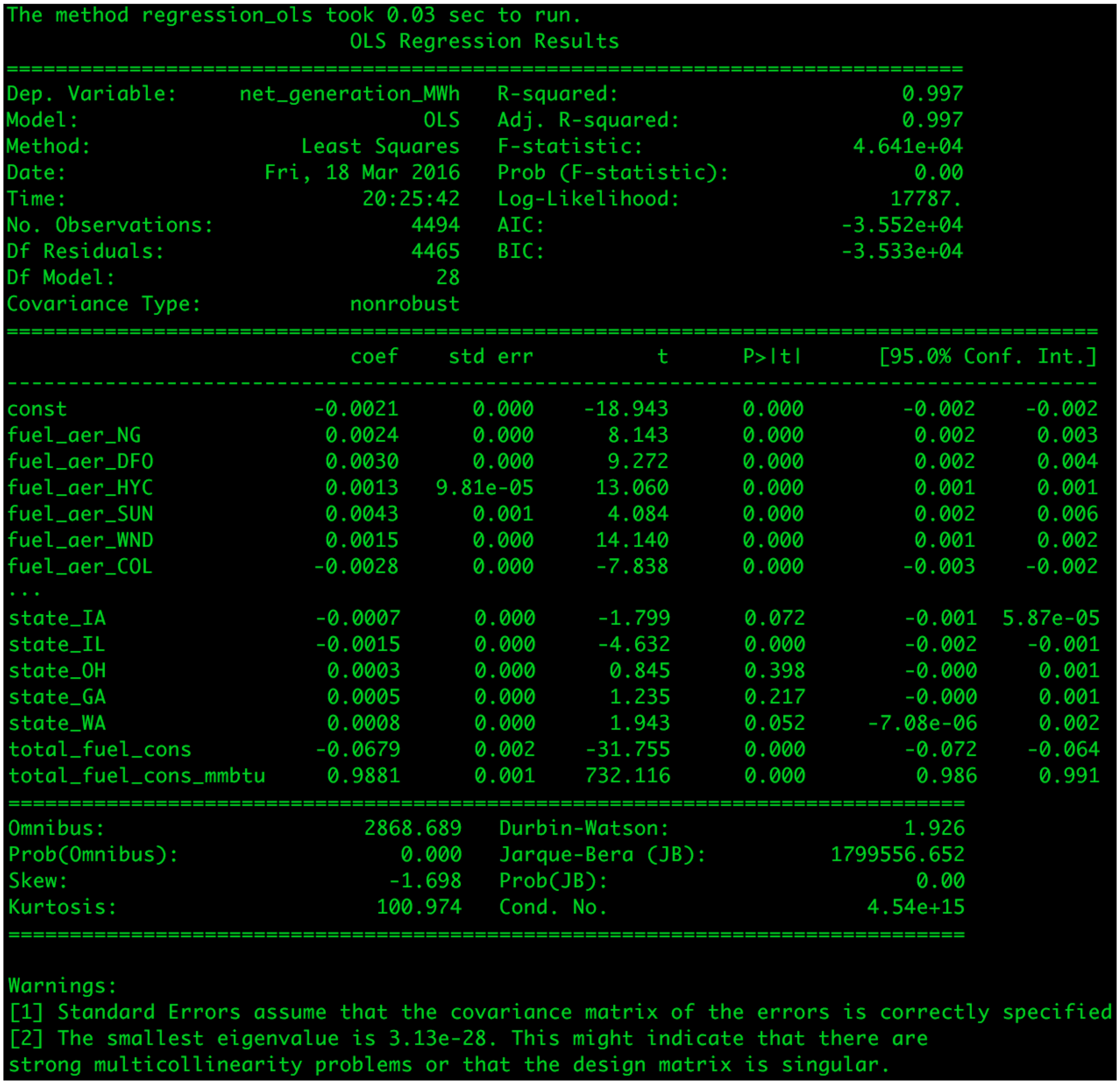

1400
1200
1000
800
600
400
200
0
Count of power plants
NG
HYC
SUN
WND
DFO
MLG
COL
NUC
Fuel type


3000
2500
2000
1500
1000
500
0
Count of power plants
ST
HY
GT
PV
IC
CT
Mover type

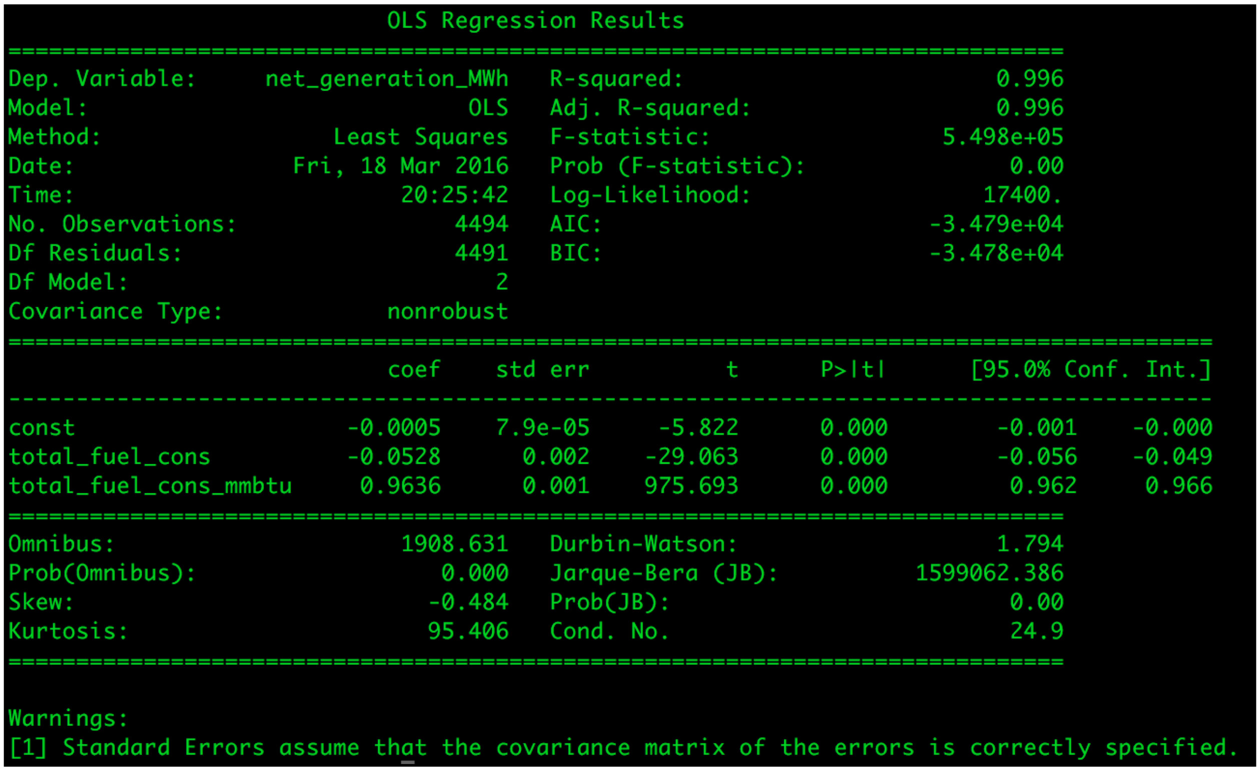

1400
1200
1000
800
600
400
200
0
Count of power plants
CA
TX
MA
IL
AL
KY
WV
State


PC = p_0
PC = p_1
PC = p_2
PC = p_3
PC = p_4
y
x
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−3
−2
−1
0
1
2
3

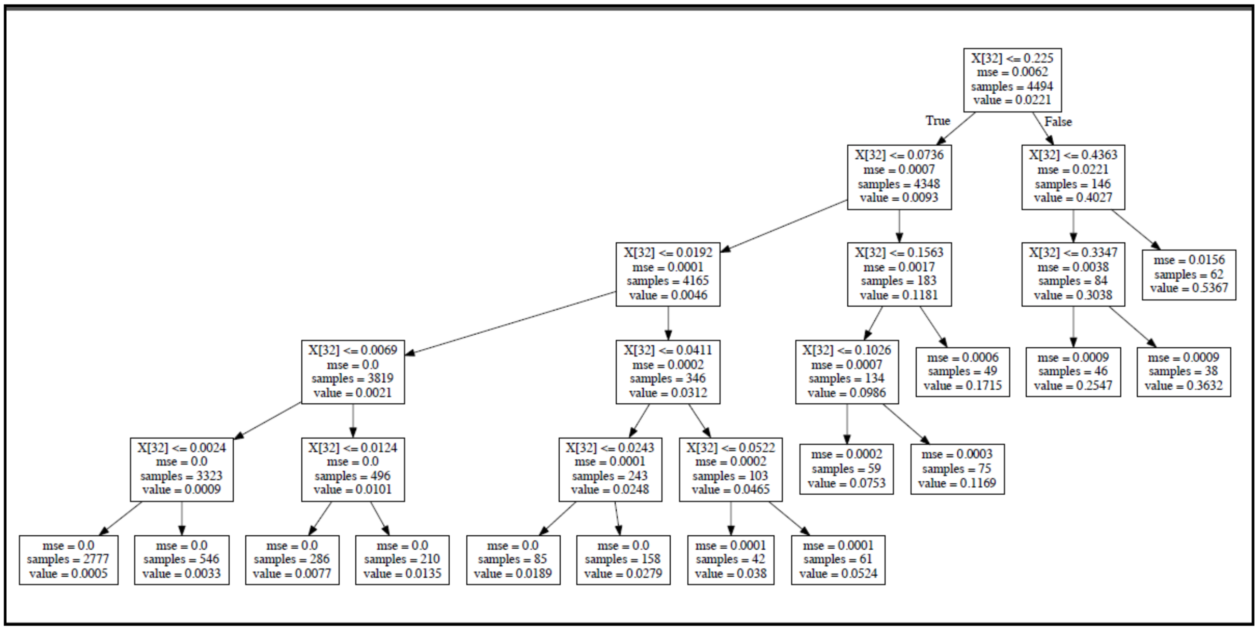

fuel = total_fuel_cons
fuel = total_fuel_cons_mmbtu
y
x
x
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2

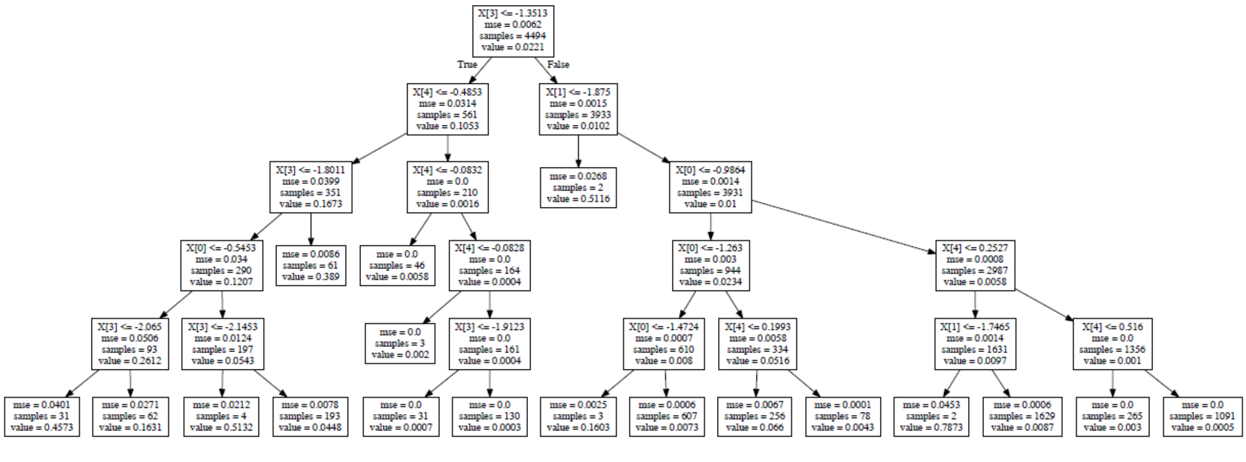


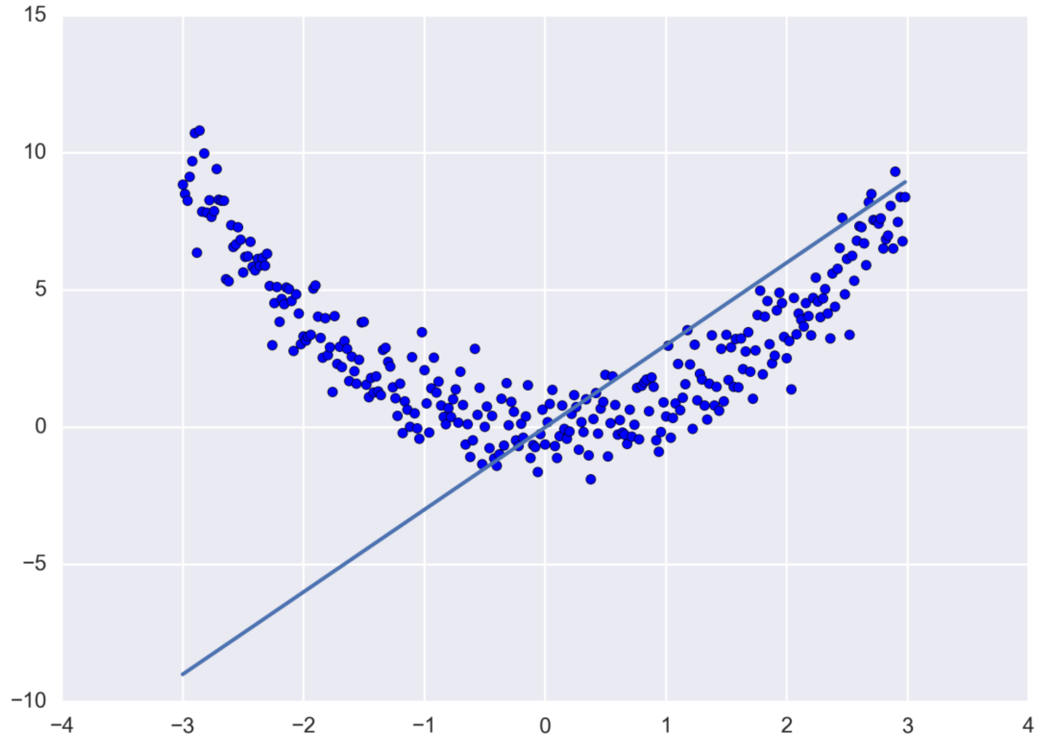
```
The method regression_ols took 0.03 sec to run.
                            OLS Regression Results
==============================================================================
Dep. Variable:     net_generation_MWh   R-squared:                       0.997
Model:                            OLS   Adj. R-squared:                  0.997
Method:                 Least Squares   F-statistic:                 4.641e+04
Date:                Fri, 18 Mar 2016   Prob (F-statistic):               0.00
Time:                        20:25:42   Log-Likelihood:                 17787.
No. Observations:                4494   AIC:                        -3.552e+04
Df Residuals:                    4465   BIC:                        -3.533e+04
Df Model:                          28
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [95.0% Conf. Int.]
-----------------------------------------------------------------------------------------
const                    -0.0021      0.000    -18.943      0.000        -0.002    -0.002
fuel_aer_NG               0.0024      0.000      8.143      0.000         0.002     0.003
fuel_aer_DFO              0.0030      0.000      9.272      0.000         0.002     0.004
fuel_aer_HYC              0.0013   9.81e-05     13.060      0.000         0.001     0.001
fuel_aer_SUN              0.0043      0.001      4.084      0.000         0.002     0.006
fuel_aer_WND              0.0015      0.000     14.140      0.000         0.001     0.002
fuel_aer_COL             -0.0028      0.000     -7.838      0.000        -0.003    -0.002
...
state_IA                 -0.0007      0.000     -1.799      0.072        -0.001  5.87e-05
state_IL                 -0.0015      0.000     -4.632      0.000        -0.002    -0.001
state_OH                  0.0003      0.000      0.845      0.398        -0.000     0.001
state_GA                  0.0005      0.000      1.235      0.217        -0.000     0.001
state_WA                  0.0008      0.000      1.943      0.052     -7.08e-06     0.002
total_fuel_cons          -0.0679      0.002    -31.755      0.000        -0.072    -0.064
total_fuel_cons_mmbtu     0.9881      0.001    732.116      0.000         0.986     0.991
==============================================================================
Omnibus:                     2868.689   Durbin-Watson:                   1.926
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1799556.652
Skew:                          -1.698   Prob(JB):                         0.00
Kurtosis:                     100.974   Cond. No.                     4.54e+15
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified
[2] The smallest eigenvalue is 3.13e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

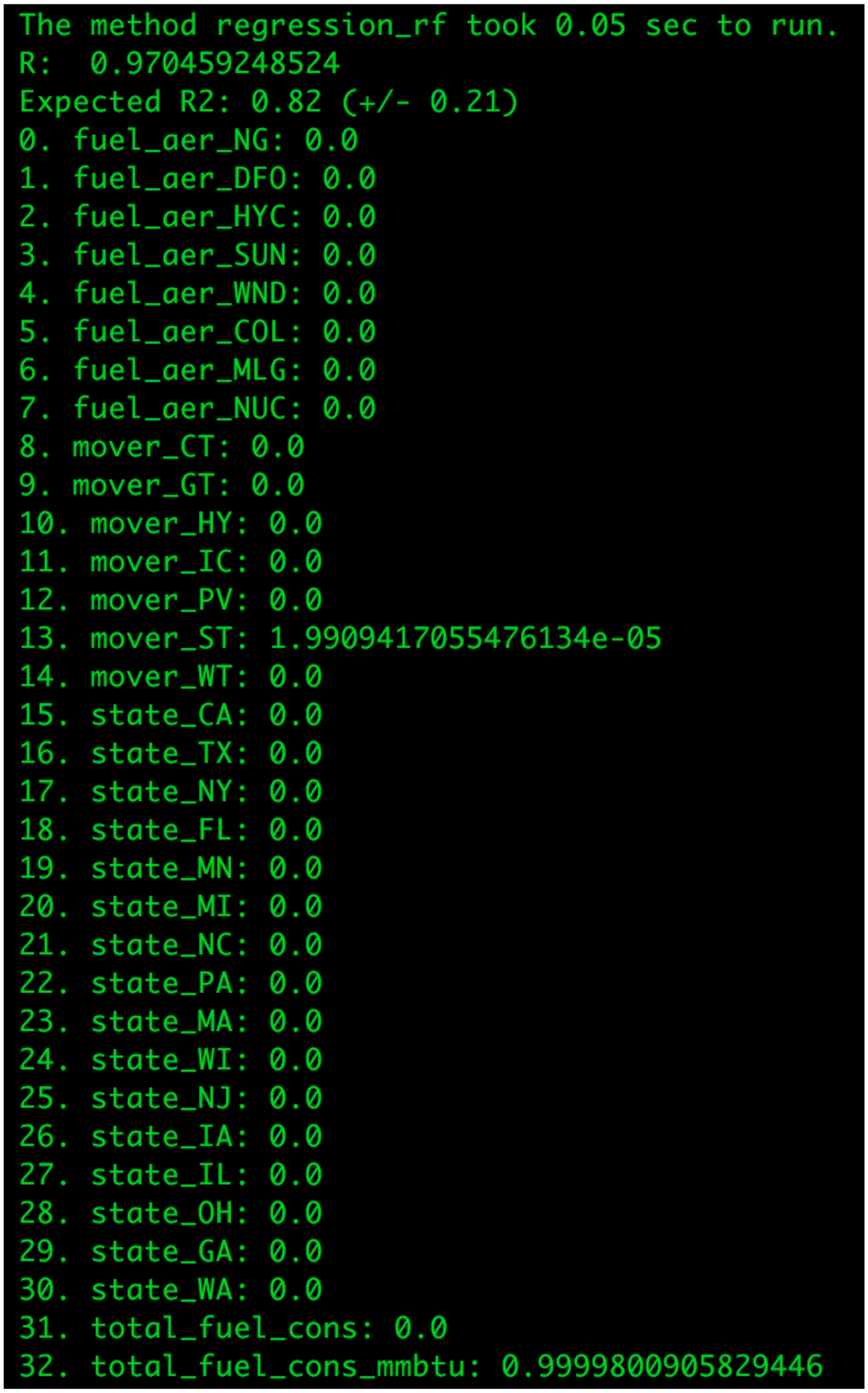


```
                            OLS Regression Results
==============================================================================
Dep. Variable:     net_generation_MWh   R-squared:                       0.996
Model:                            OLS   Adj. R-squared:                  0.996
Method:                 Least Squares   F-statistic:                 5.498e+05
Date:                Fri, 18 Mar 2016   Prob (F-statistic):               0.00
Time:                        20:25:42   Log-Likelihood:                 17400.
No. Observations:                4494   AIC:                        -3.479e+04
Df Residuals:                    4491   BIC:                        -3.478e+04
Df Model:                           2
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [95.0% Conf. Int.]
-----------------------------------------------------------------------------------------
const                    -0.0005    7.9e-05     -5.822      0.000        -0.001    -0.000
total_fuel_cons          -0.0528      0.002    -29.063      0.000        -0.056    -0.049
total_fuel_cons_mmbtu     0.9636      0.001    975.693      0.000         0.962     0.966
==============================================================================
Omnibus:                     1908.631   Durbin-Watson:                   1.794
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1599062.386
Skew:                          -0.484   Prob(JB):                         0.00
Kurtosis:                      95.406   Cond. No.                         24.9
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

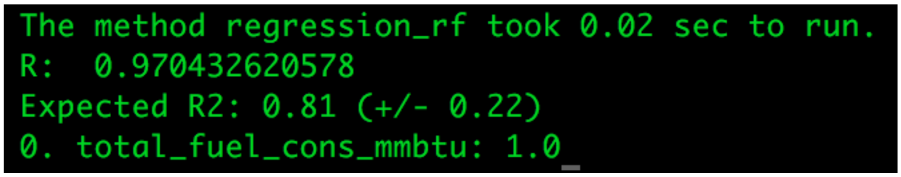



X[3] <= -1.3513 mse = 0.0062 samples = 4494 value = 0.0221
True
False
X[4] <= -0.4853 mse = 0.0314 samples = 561 value = 0.1053
X[1] <= -1.875 mse = 0.0015 samples = 3933 value = 0.0102
X[3] <= -1.8011 mse = 0.0399 samples = 351 value = 0.1673
X[4] <= -0.0832 mse = 0.0 samples = 210 value = 0.0016
mse = 0.0268 samples = 2 value = 0.5116
X[0] <= -0.9864 mse = 0.0014 samples = 3931 value = 0.01
X[0] <= -0.5453 mse = 0.034 samples = 290 value = 0.1207
mse = 0.0086 samples = 61 value = 0.389
mse = 0.0 samples = 46 value = 0.0058
X[4] <= -0.0828 mse = 0.0 samples = 164 value = 0.0004
X[0] <= -1.263 mse = 0.003 samples = 944 value = 0.0234
X[4] <= 0.2527 mse = 0.0008 samples = 2987 value = 0.0058
X[3] <= -2.065 mse = 0.0506 samples = 93 value = 0.2612
X[3] <= -2.1453 mse = 0.0124 samples = 197 value = 0.0543
mse = 0.0 samples = 3 value = 0.002
X[3] <= -1.9123 mse = 0.0 samples = 161 value = 0.0004
X[0] <= -1.4724 mse = 0.0007 samples = 610 value = 0.008
X[4] <= 0.1993 mse = 0.0058 samples = 334 value = 0.0516
X[1] <= -1.7465 mse = 0.0014 samples = 1631 value = 0.0097
X[4] <= 0.516 mse = 0.0 samples = 1356 value = 0.001
mse = 0.0401 samples = 31 value = 0.4573
mse = 0.0271 samples = 62 value = 0.1631
mse = 0.0212 samples = 4 value = 0.5132
mse = 0.0078 samples = 193 value = 0.0448
mse = 0.0 samples = 31 value = 0.0007
mse = 0.0 samples = 130 value = 0.0003
mse = 0.0025 samples = 3 value = 0.1603
mse = 0.0006 samples = 607 value = 0.0073
mse = 0.0067 samples = 256 value = 0.066
mse = 0.0001 samples = 78 value = 0.0043
mse = 0.0453 samples = 2 value = 0.7873
mse = 0.0006 samples = 1629 value = 0.0087
mse = 0.0 samples = 265 value = 0.003
mse = 0.0 samples = 1091 value = 0.0005

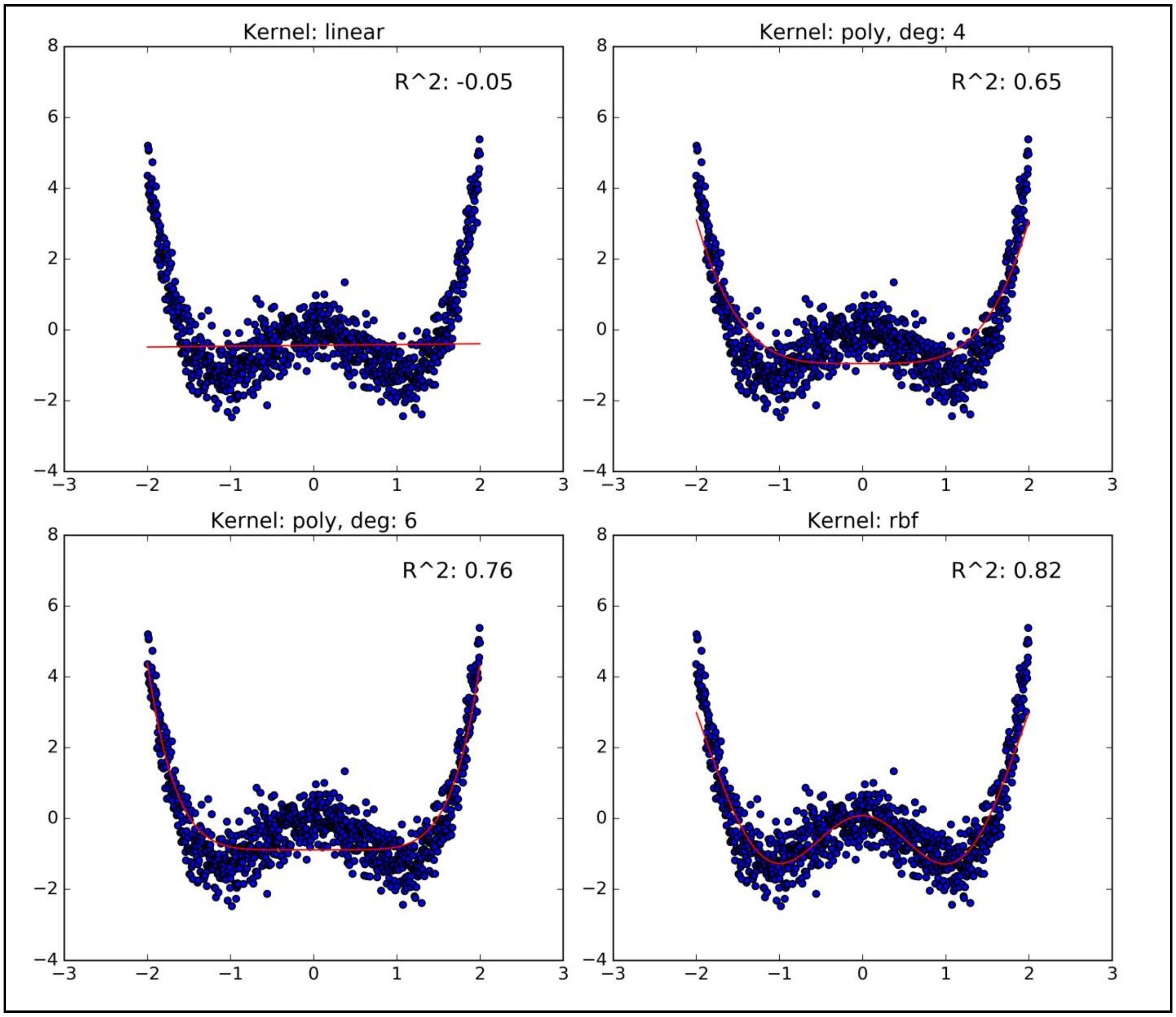

15
10
5
0
−5
−10
−4
−3
−2
−1
0
1
2
3
4

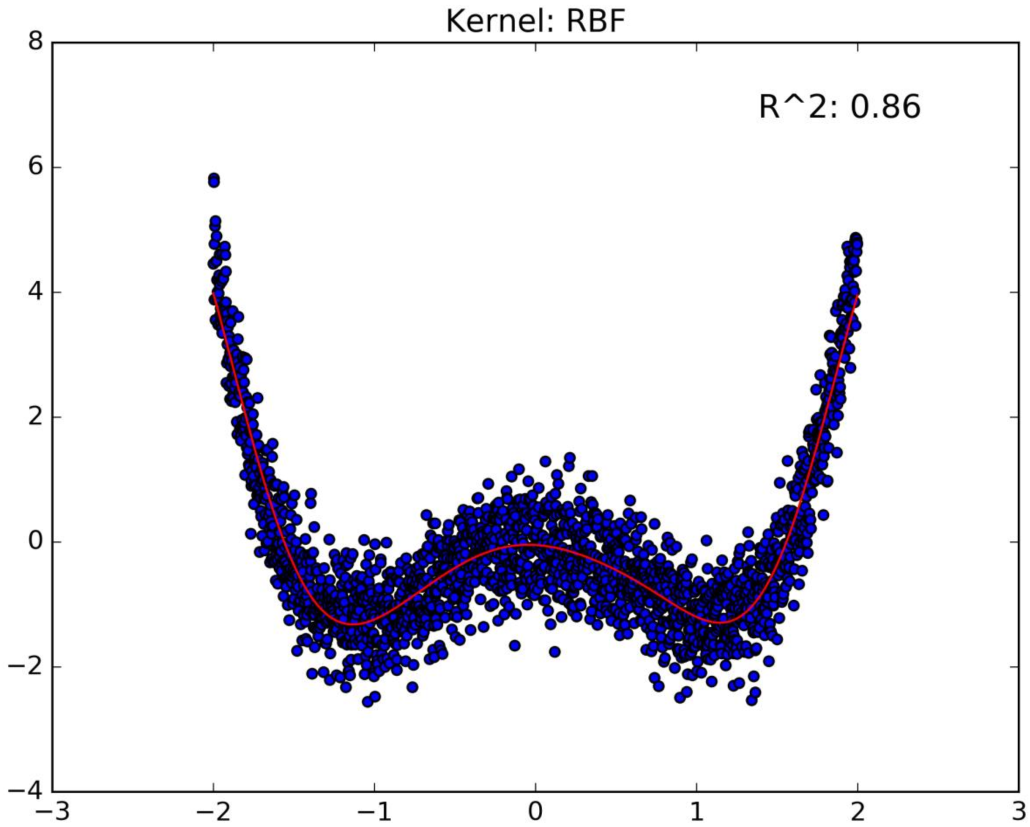


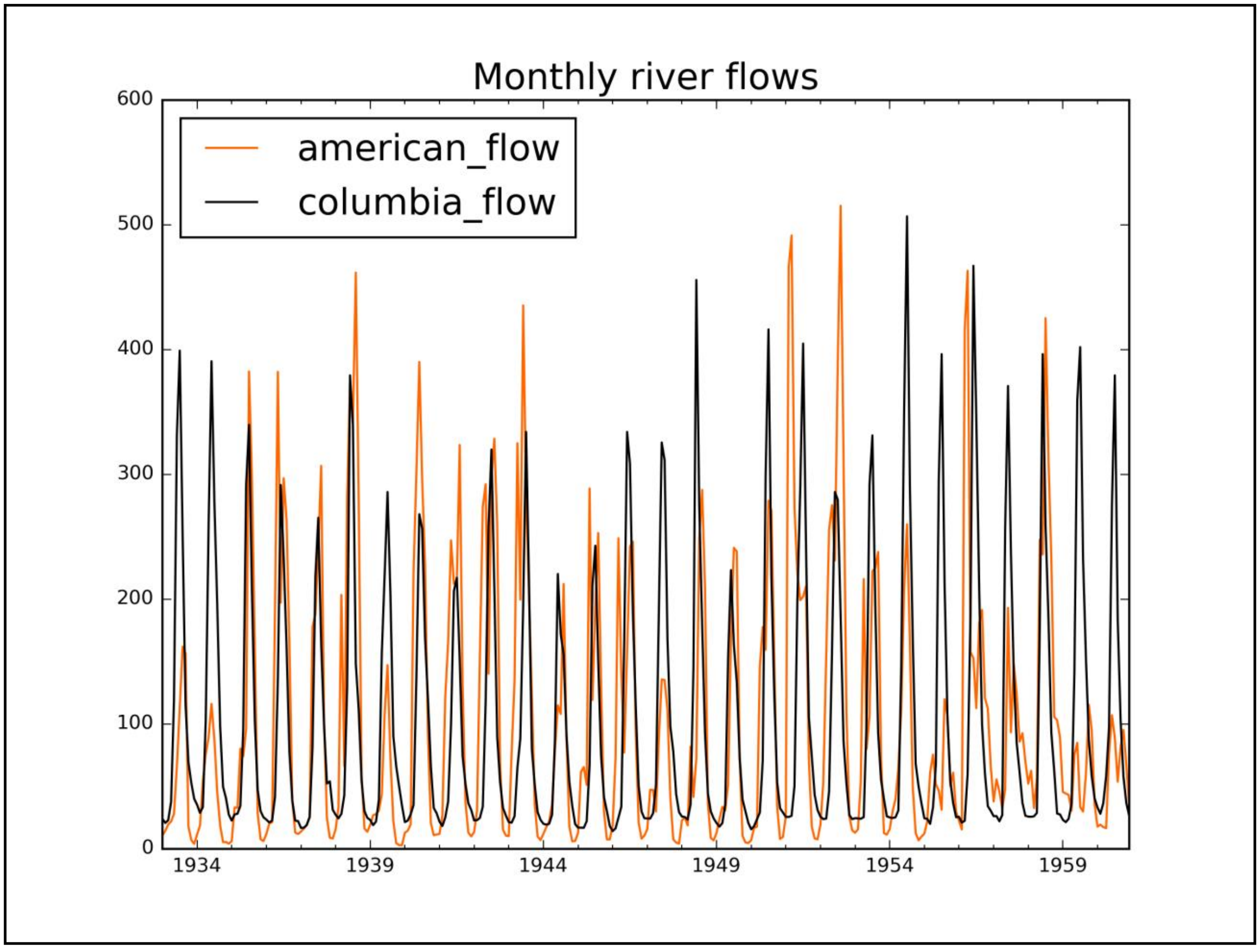
```
The method regression_rf took 0.05 sec to run.
R:  0.970459248524
Expected R2: 0.82 (+/- 0.21)
0. fuel_aer_NG: 0.0
1. fuel_aer_DFO: 0.0
2. fuel_aer_HYC: 0.0
3. fuel_aer_SUN: 0.0
4. fuel_aer_WND: 0.0
5. fuel_aer_COL: 0.0
6. fuel_aer_MLG: 0.0
7. fuel_aer_NUC: 0.0
8. mover_CT: 0.0
9. mover_GT: 0.0
10. mover_HY: 0.0
11. mover_IC: 0.0
12. mover_PV: 0.0
13. mover_ST: 1.9909417055476134e-05
14. mover_WT: 0.0
15. state_CA: 0.0
16. state_TX: 0.0
17. state_NY: 0.0
18. state_FL: 0.0
19. state_MN: 0.0
20. state_MI: 0.0
21. state_NC: 0.0
22. state_PA: 0.0
23. state_MA: 0.0
24. state_WI: 0.0
25. state_NJ: 0.0
26. state_IA: 0.0
27. state_IL: 0.0
28. state_OH: 0.0
29. state_GA: 0.0
30. state_WA: 0.0
31. total_fuel_cons: 0.0
32. total_fuel_cons_mmbtu: 0.9999800905829446
```

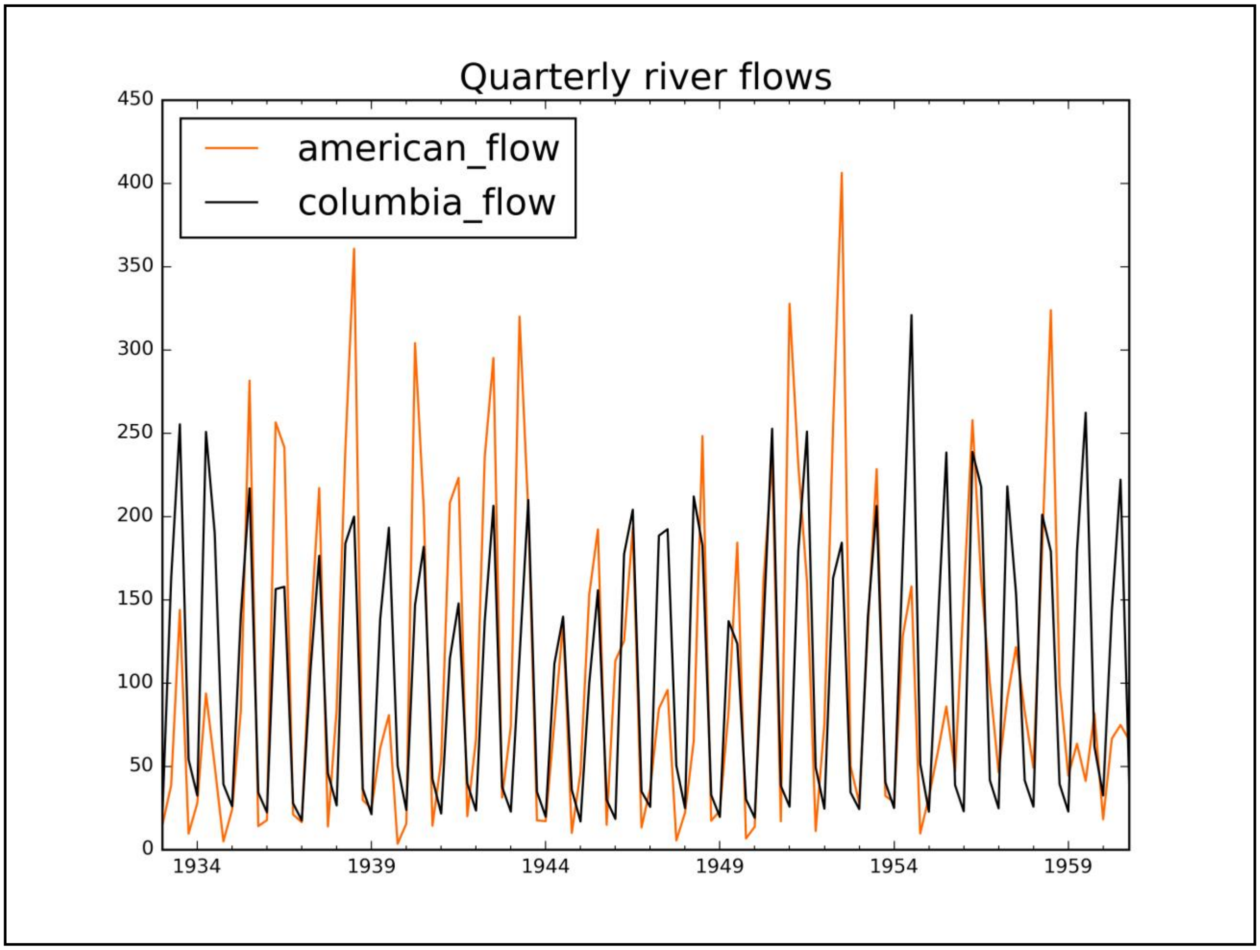


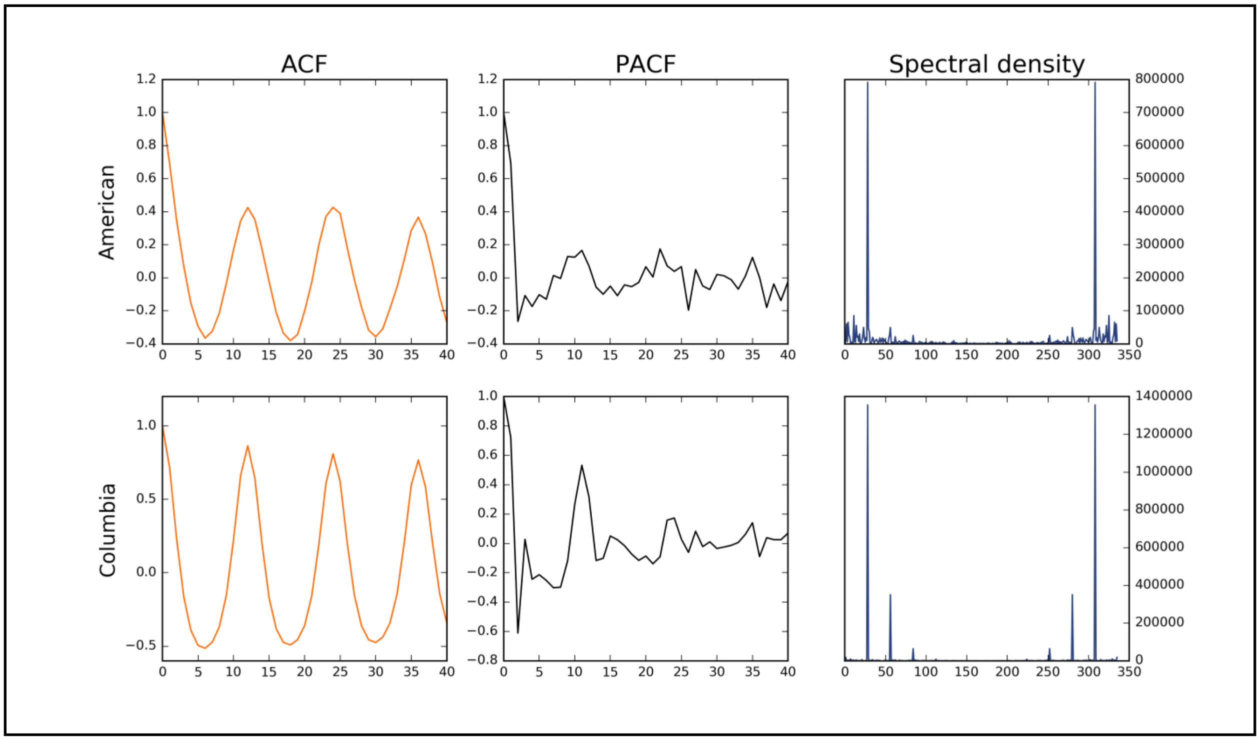
```
The method regression_rf took 0.02 sec to run.
R:  0.970432620578
Expected R2: 0.81 (+/- 0.22)
0. total_fuel_cons_mmbtu: 1.0
```

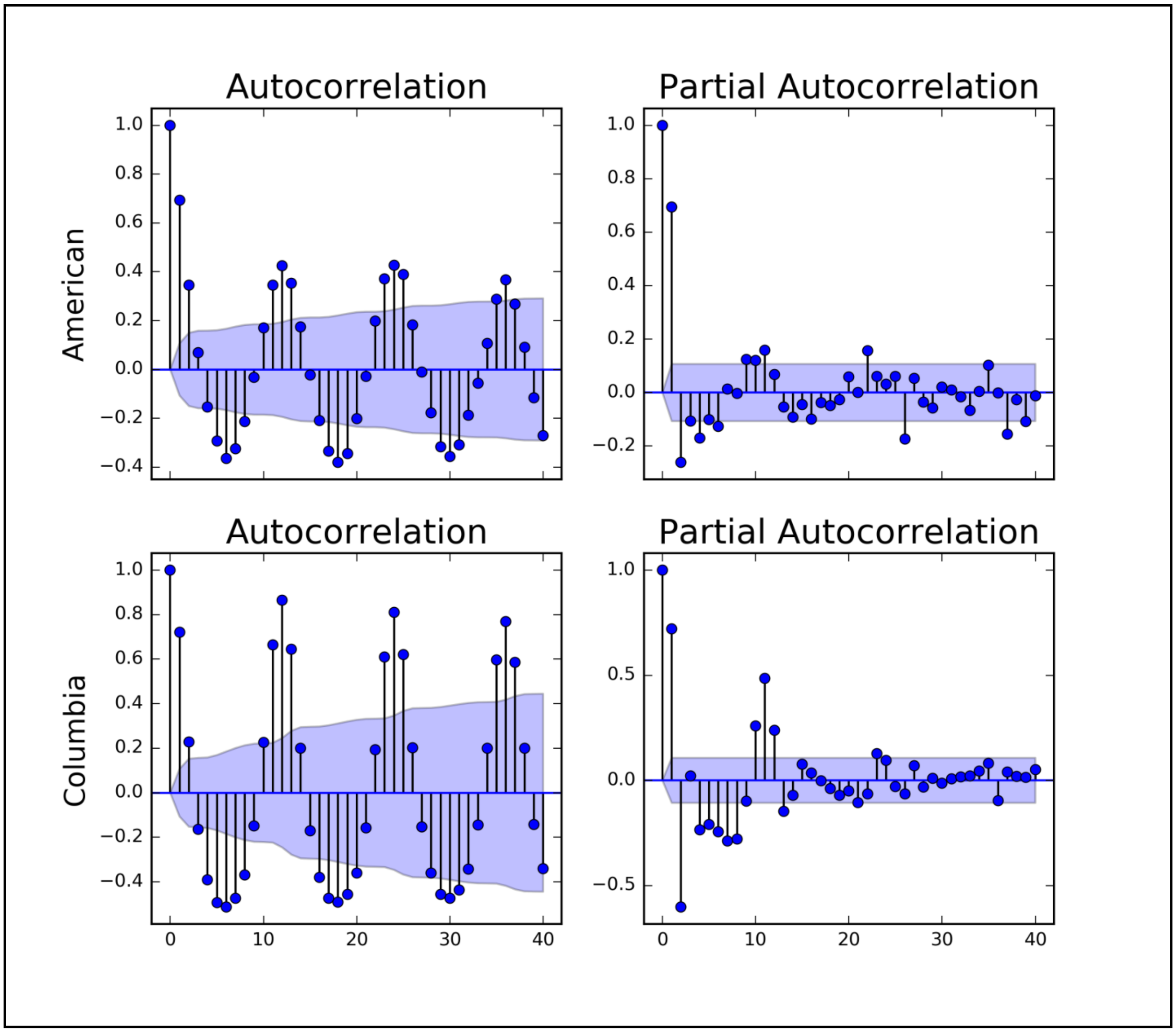

Kernel: linear
R^2: -0.05
Kernel: poly, deg: 4
R^2: 0.65
Kernel: poly, deg: 6
R^2: 0.76
Kernel: rbf
R^2: 0.82
8
6
4
2
0
−2
−4
−3
−2
−1
0
1
2
3

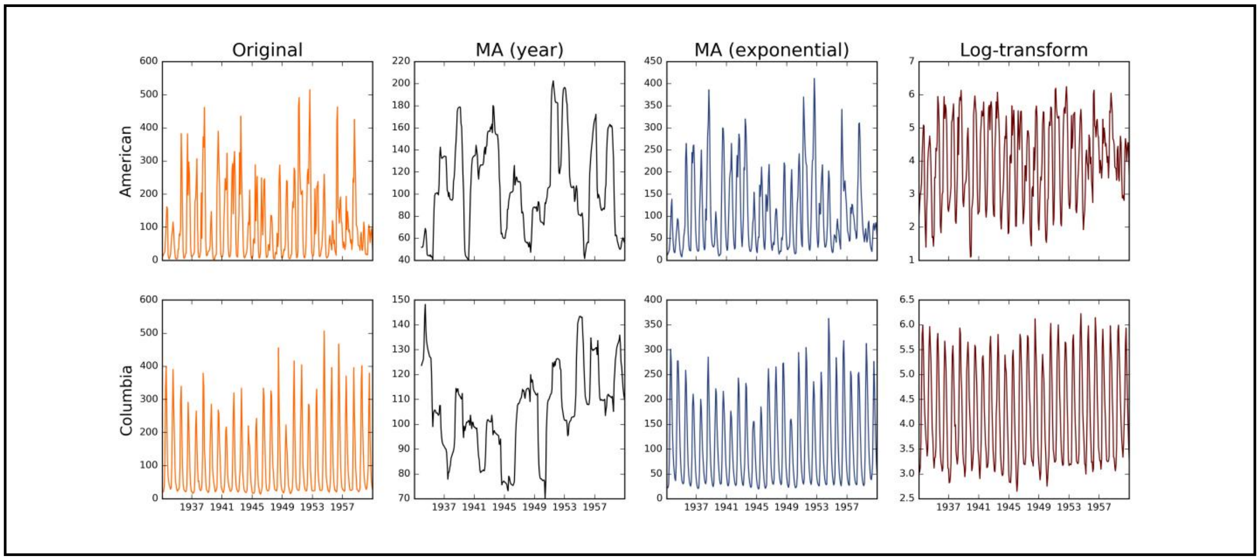

Kernel: RBF
R^2: 0.86
8
6
4
2
0
−2
−4
−3
−2
−1
0
1
2
3

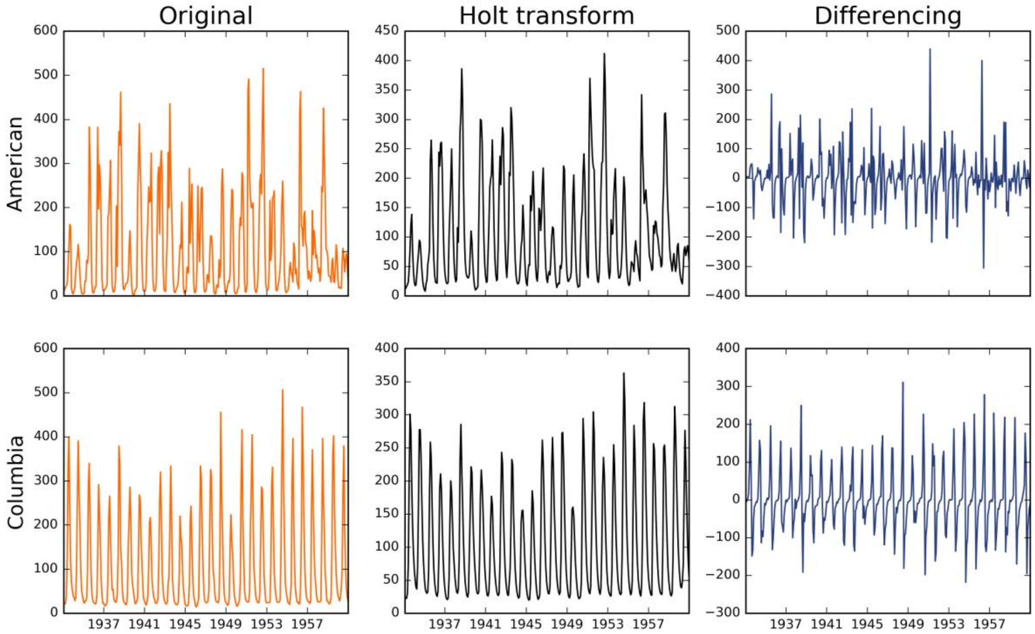

Monthly river flows
american_flow
columbia_flow
600
500
400
300
200
100
0
1934
1939
1944
1949
1954
1959

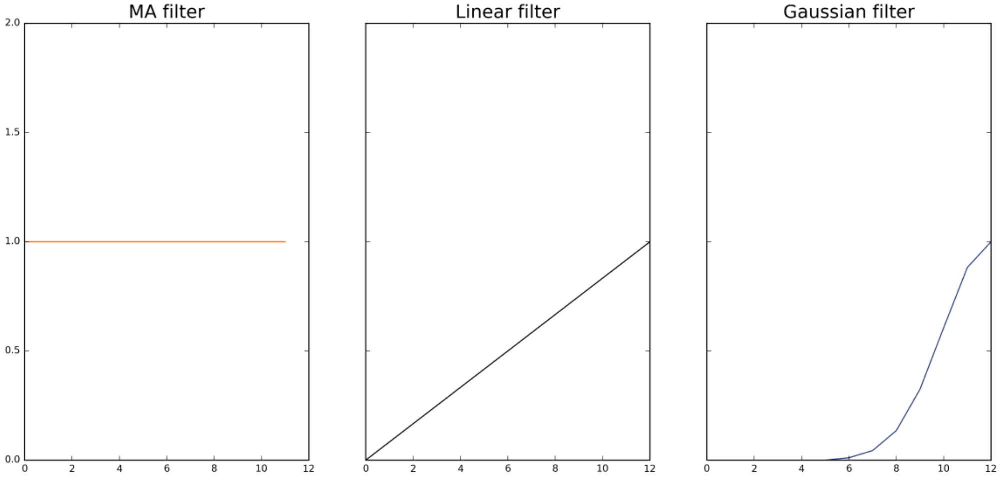

Quarterly river flows
american_flow
columbia_flow
450
400
350
300
250
200
150
100
50
0
1934
1939
1944
1949
1954
1959

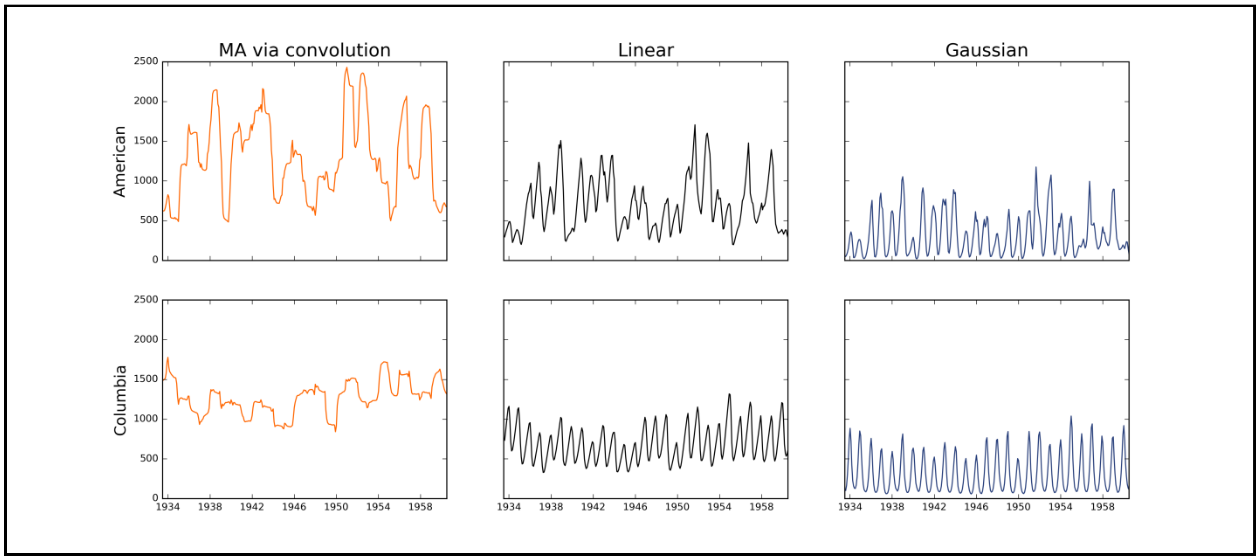

ACF
PACF
Spectral density
American
Columbia
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
5
10
15
20
25
30
35
40
800000
700000
600000
500000
400000
300000
200000
100000
0
50
100
150
200
250
300
350
1.0
0.5
0.0
−0.5
−0.6
−0.8
1400000
1200000
1000000
800000
600000
400000
200000

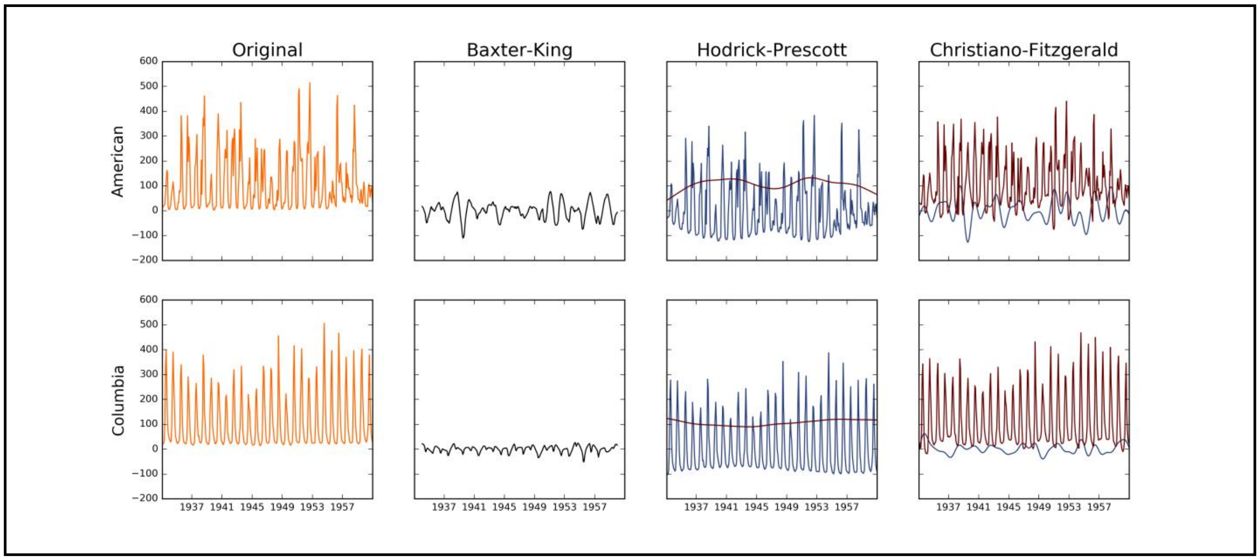

Autocorrelation
Partial Autocorrelation
American
Autocorrelation
Partial Autocorrelation
Columbia
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
1.0
0.5
0.0
−0.5
0
10
20
30
40

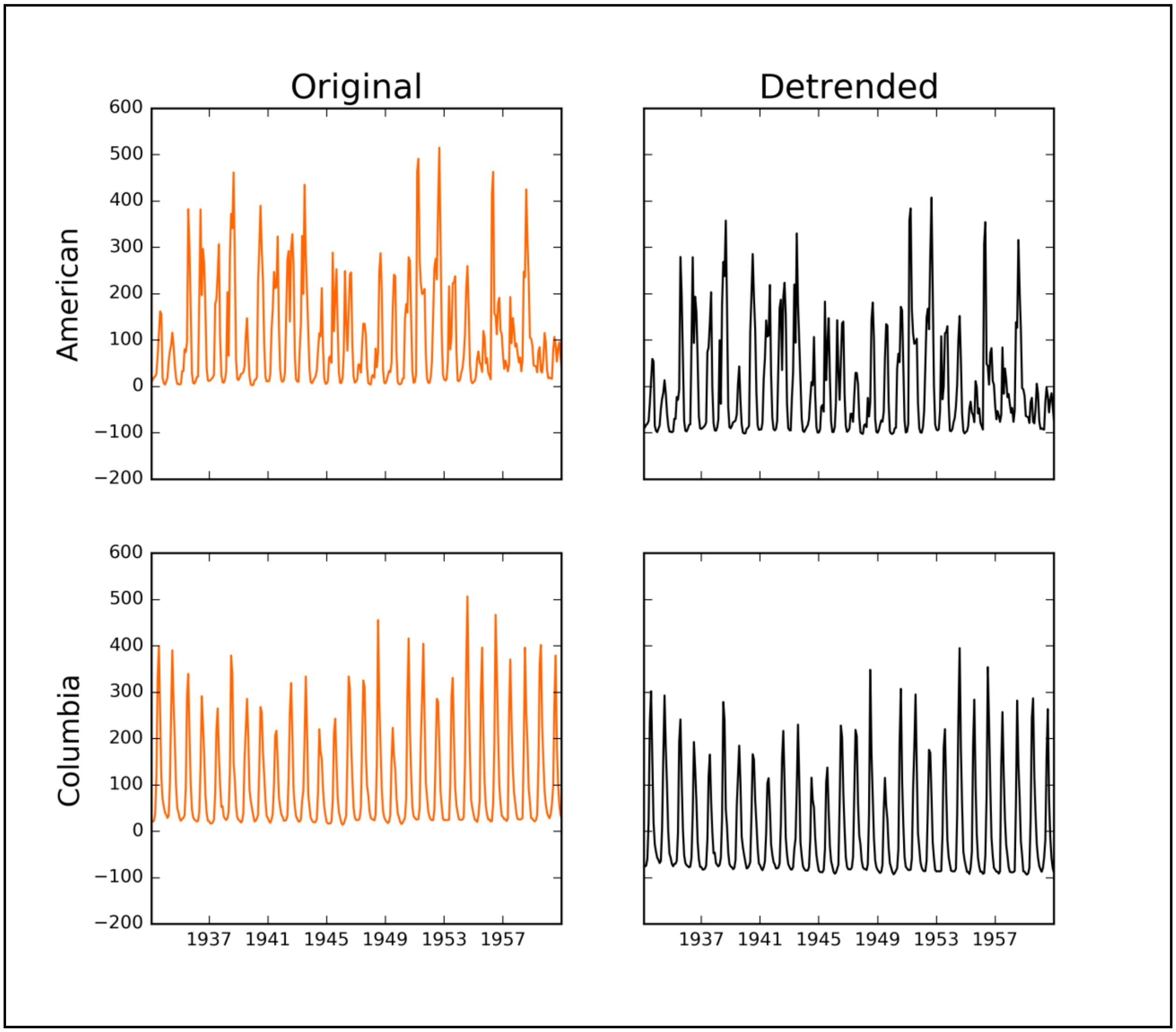

Original
MA (year)
MA (exponential)
Log-transform
American
Columbia
600
500
400
300
200
100
0
220
200
180
160
140
120
100
80
60
40
450
400
350
300
250
200
150
100
50
0
7
6
5
4
3
2
1
150
140
130
120
110
100
90
80
70
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
1937
1941
1945
1949
1953
1957

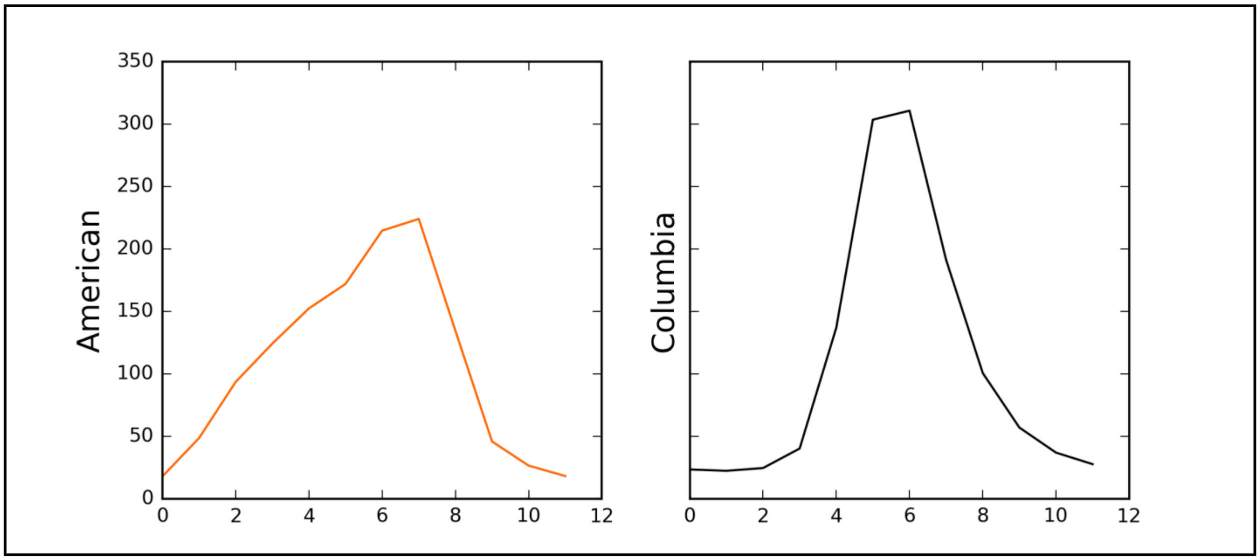

Original
Holt transform
Differencing
American
Columbia
0
100
200
300
400
500
600
450
350
250
150
50
−100
−200
−300
−400
1937
1941
1945
1949
1953
1957

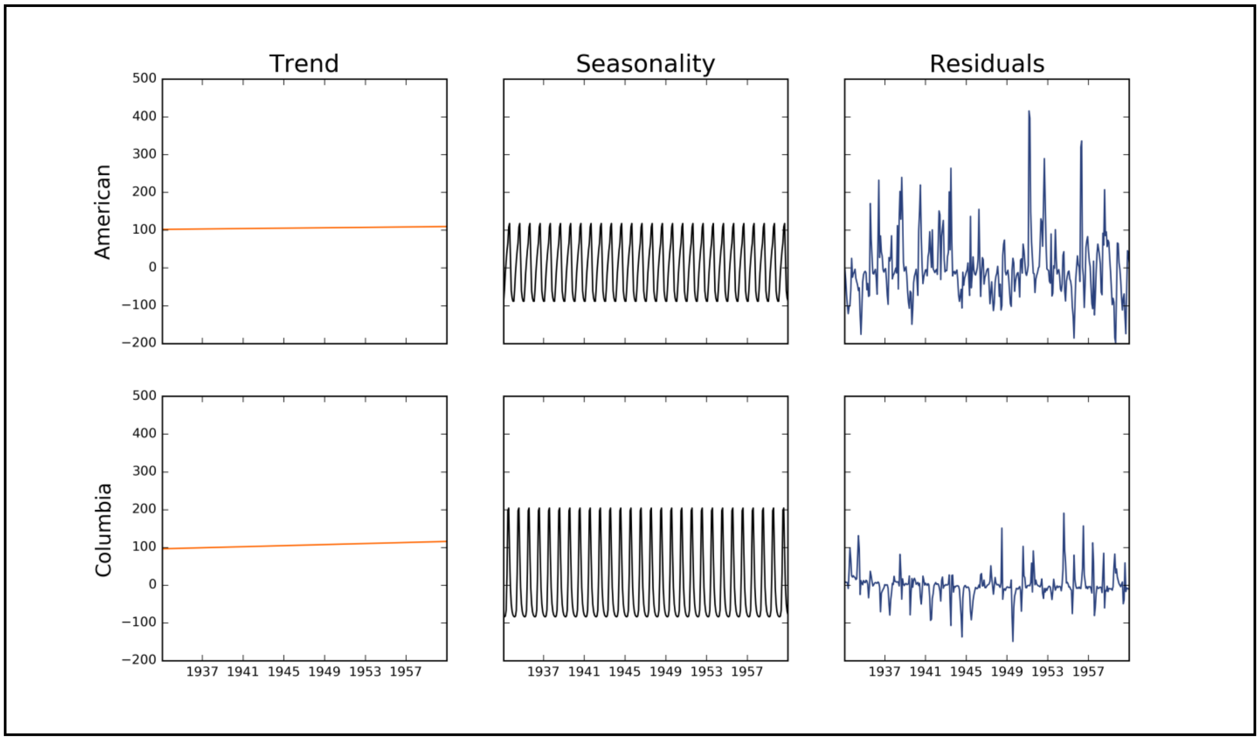

MA filter
Linear filter
Gaussian filter
0.0
0.5
1.0
1.5
2.0
0
2
4
6
8
10
12

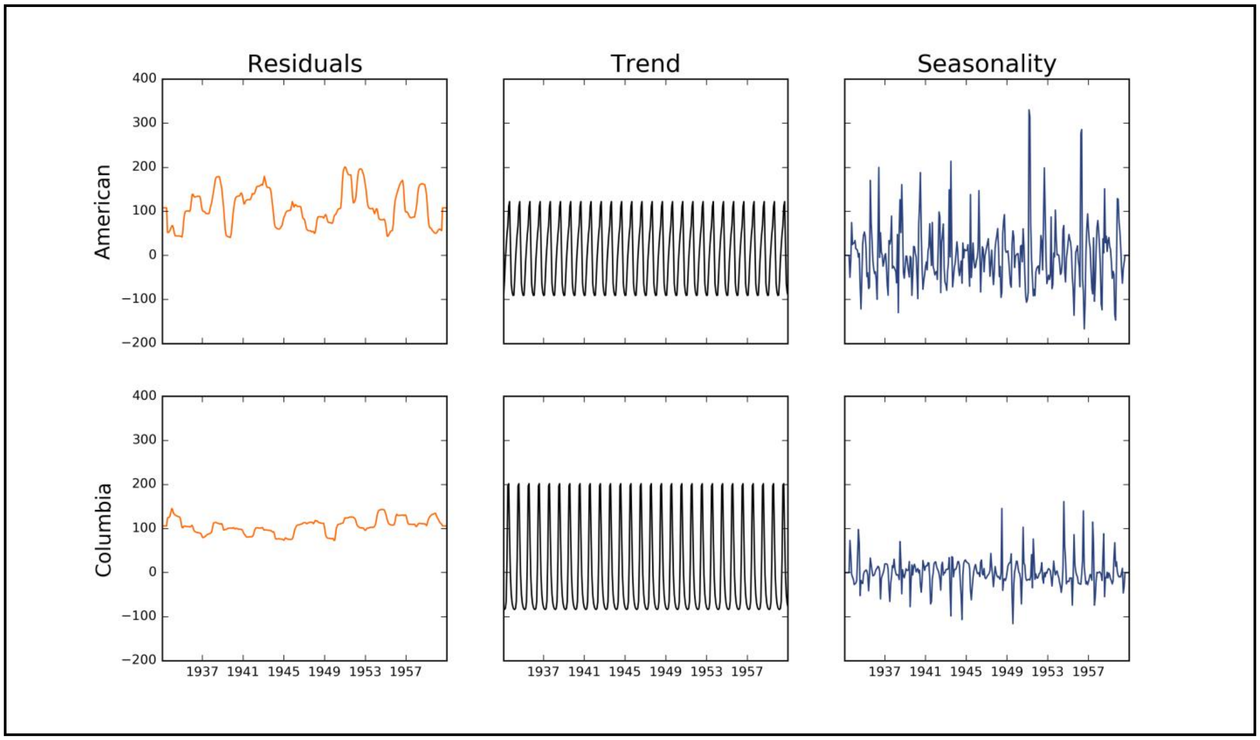

MA via convolution
Linear
Gaussian
American
Columbia
2500
2000
1500
1000
500
0
1934
1938
1942
1946
1950
1954
1958

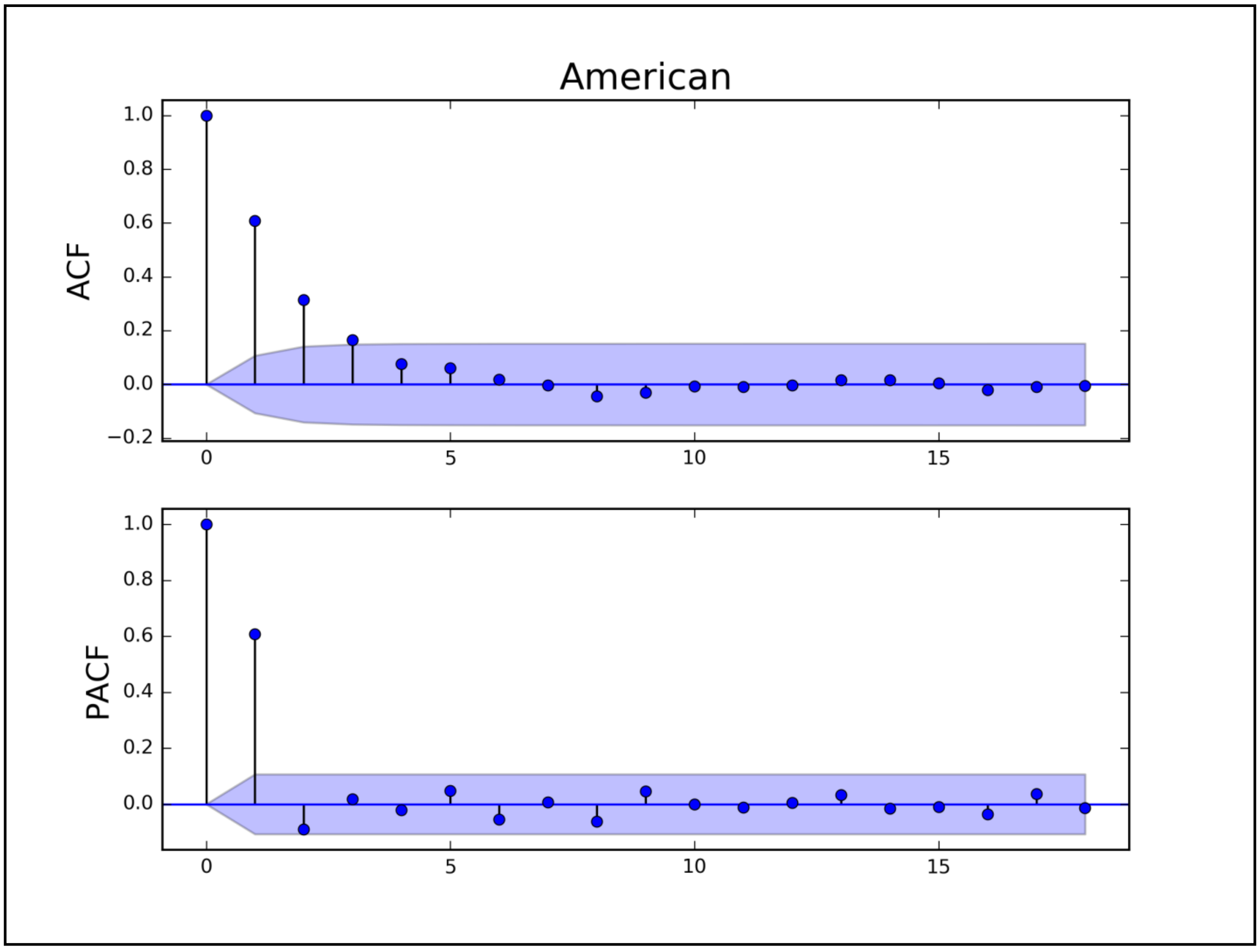

Original
Baxter-King
Hodrick-Prescott
Christiano-Fitzgerald
American
Columbia
600
500
400
300
200
100
0
−100
−200
1937
1941
1945
1949
1953
1957

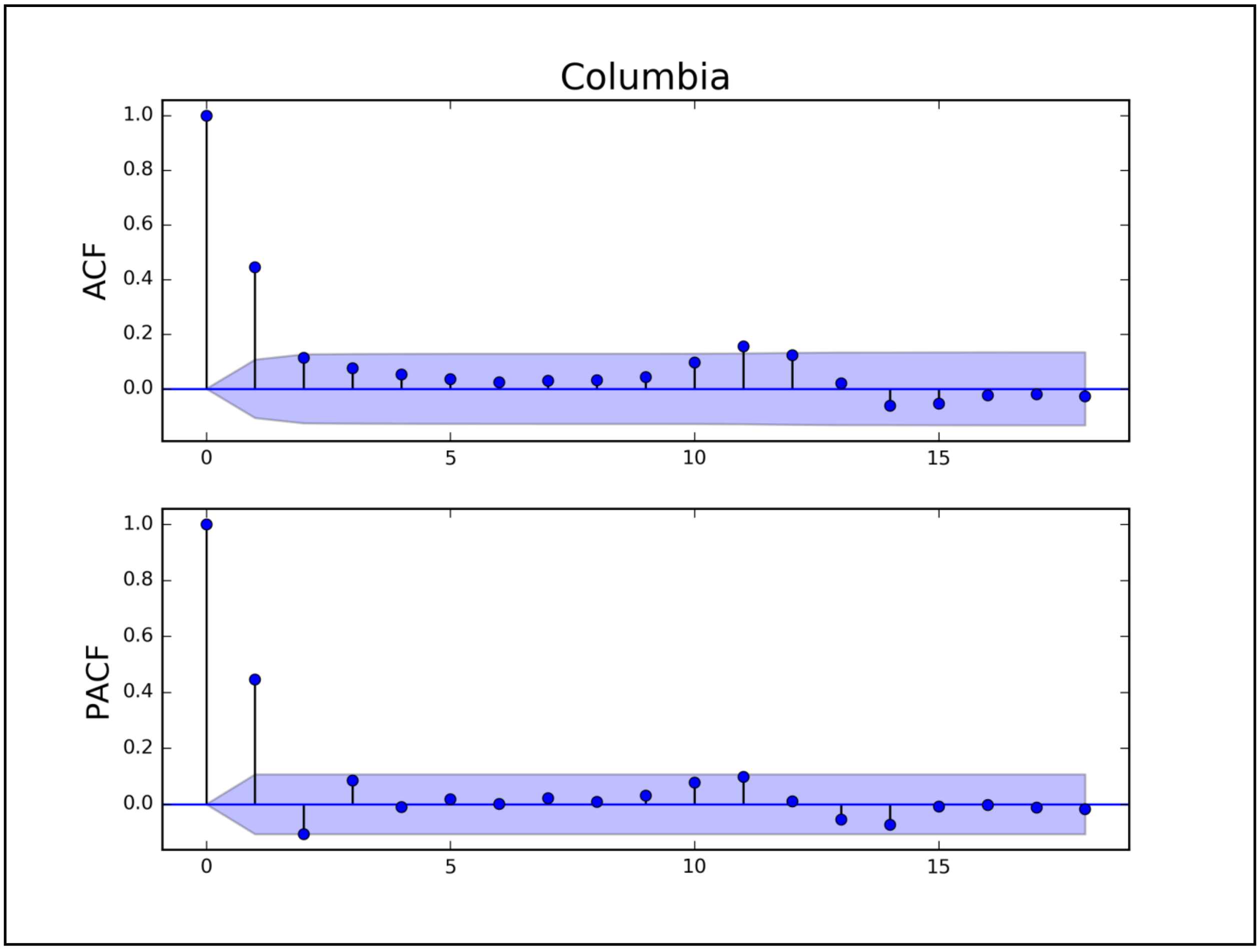

Original
Detrended
American
Columbia
600
500
400
300
200
100
0
−100
−200
1937
1941
1945
1949
1953
1957

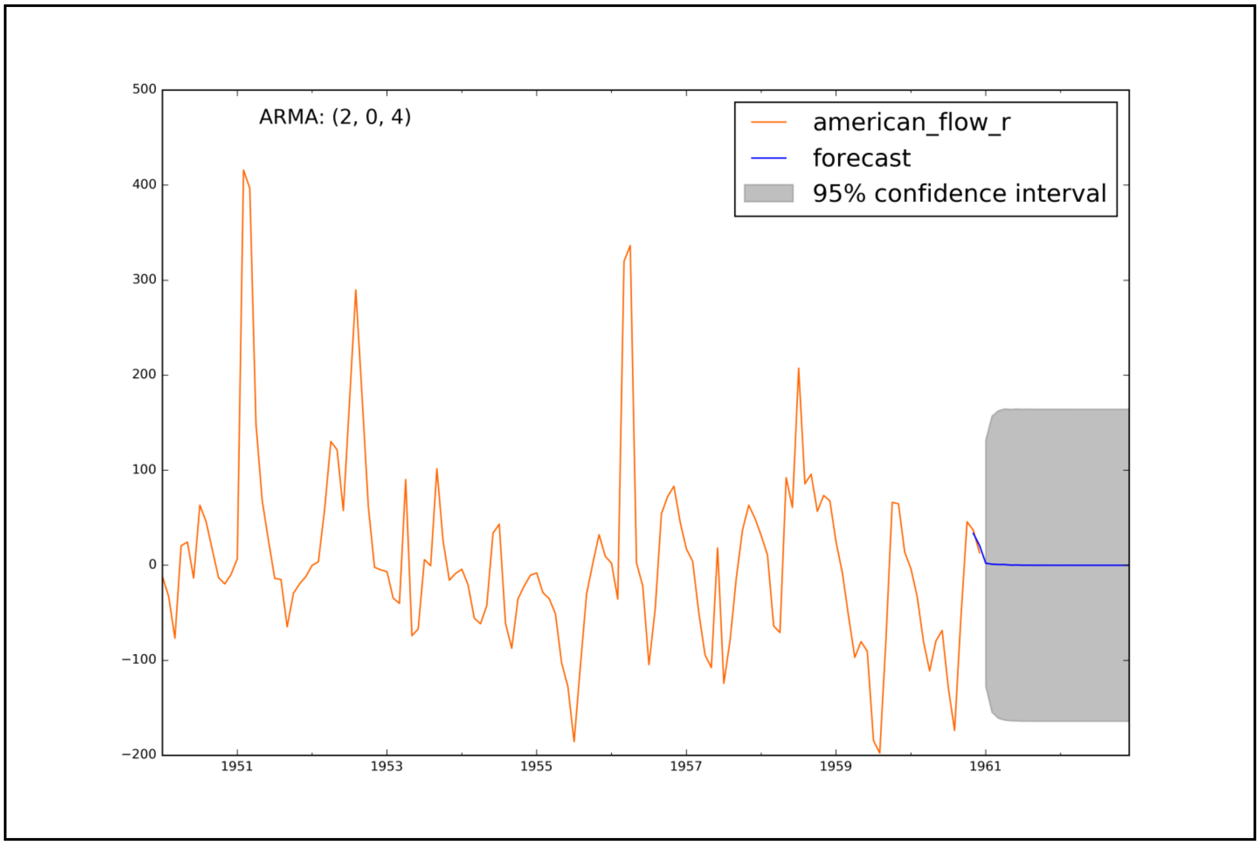

American
Columbia
350
300
250
200
150
100
50
0
0
2
4
6
8
10
12

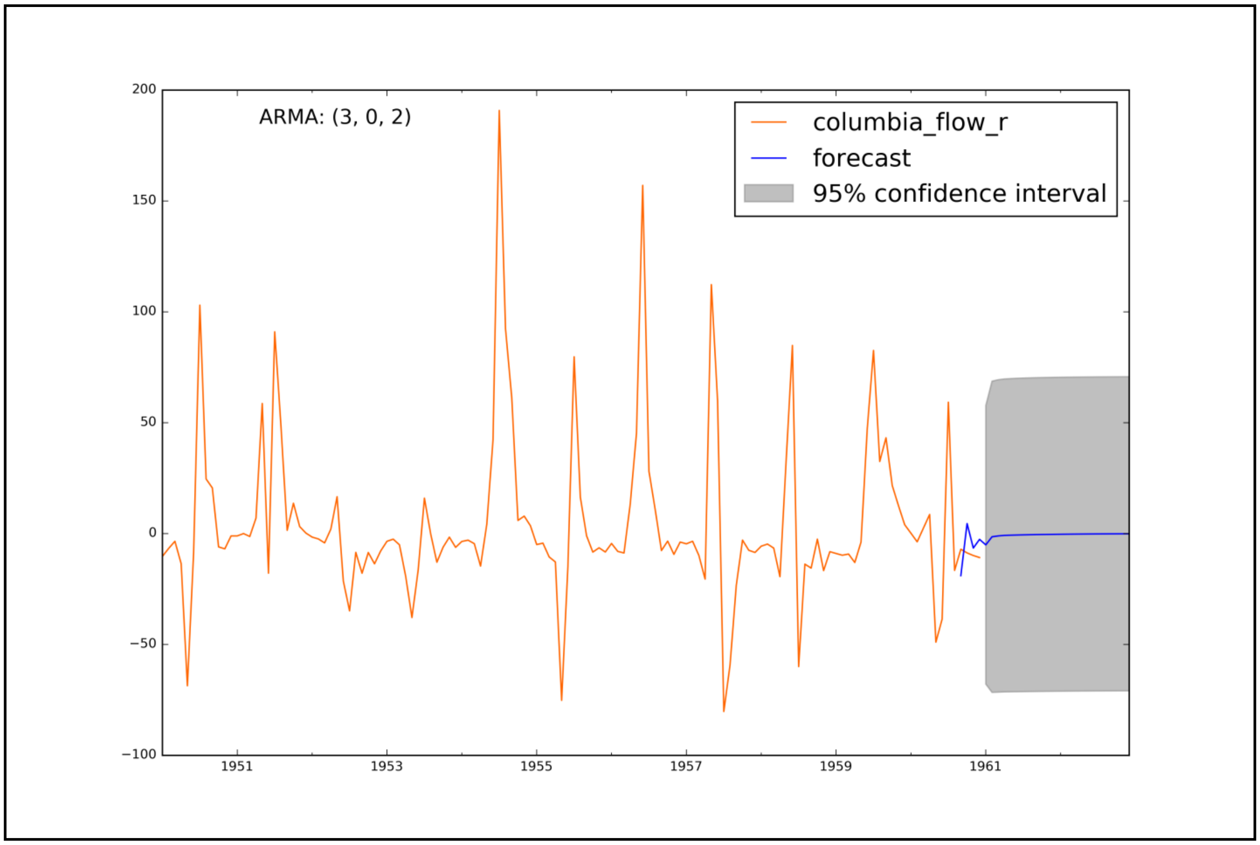

Trend
Seasonality
Residuals
American
Columbia
500
400
300
200
100
0
−100
−200
1937
1941
1945
1949
1953
1957

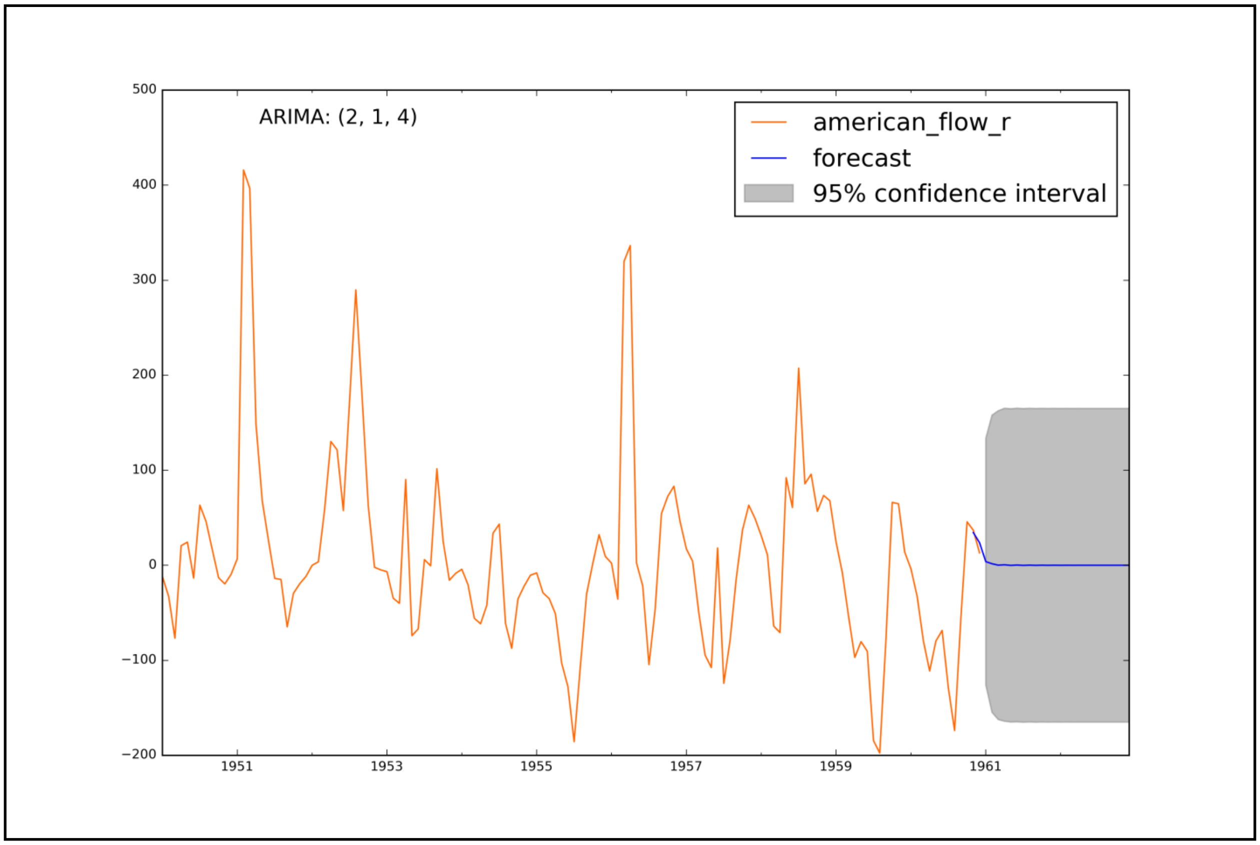

Residuals
Trend
Seasonality
American
Columbia
400
300
200
100
0
−100
−200
1937
1941
1945
1949
1953
1957

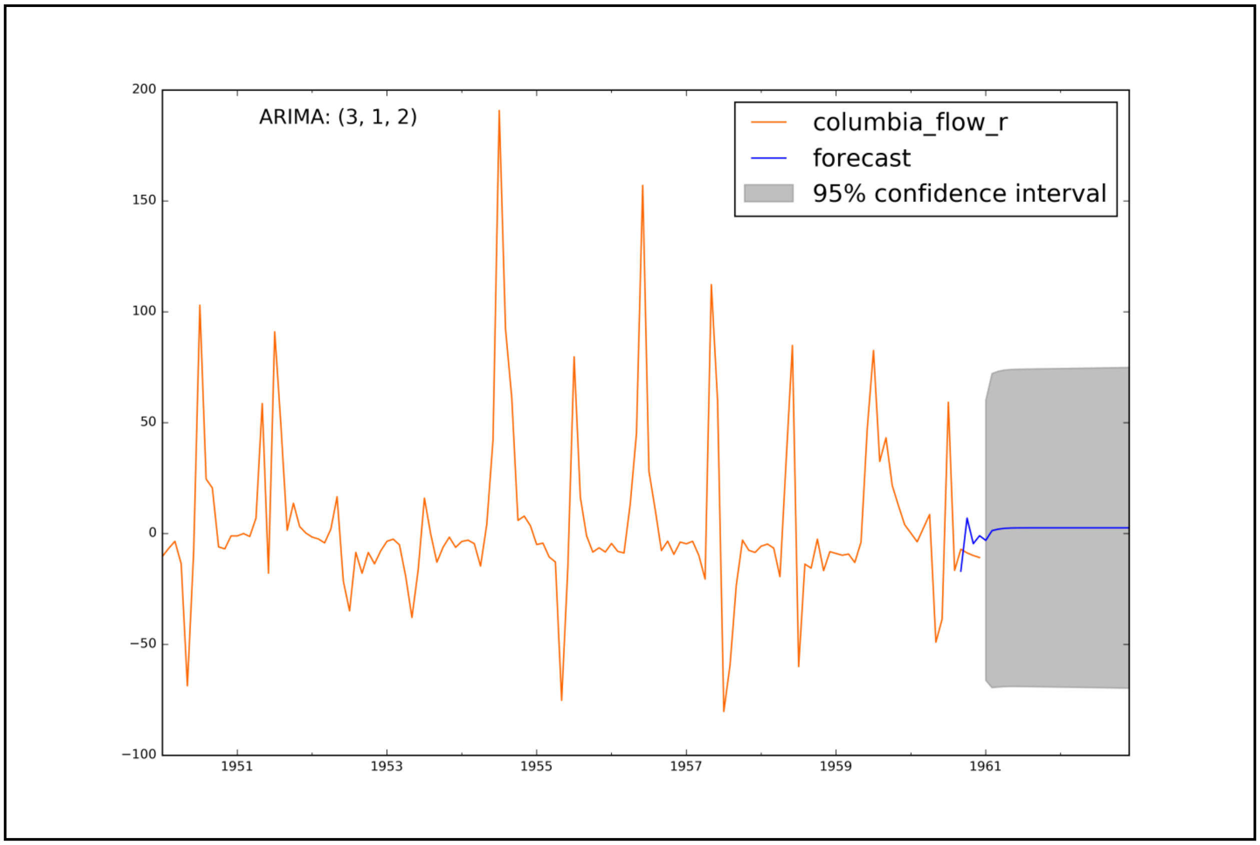

American
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
PACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15

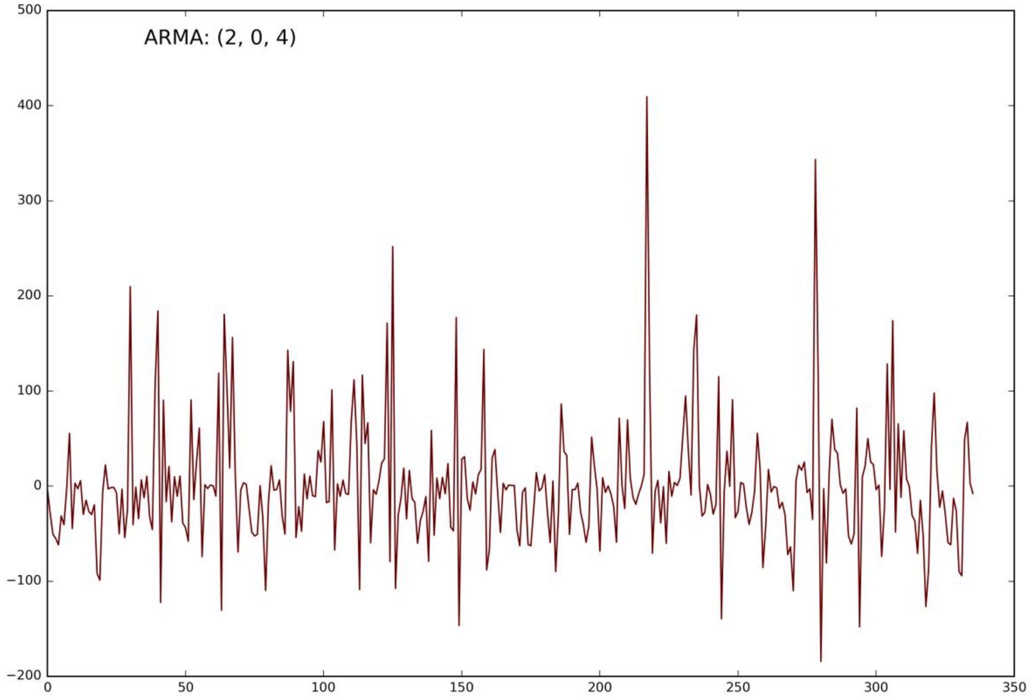

Columbia
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
PACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15

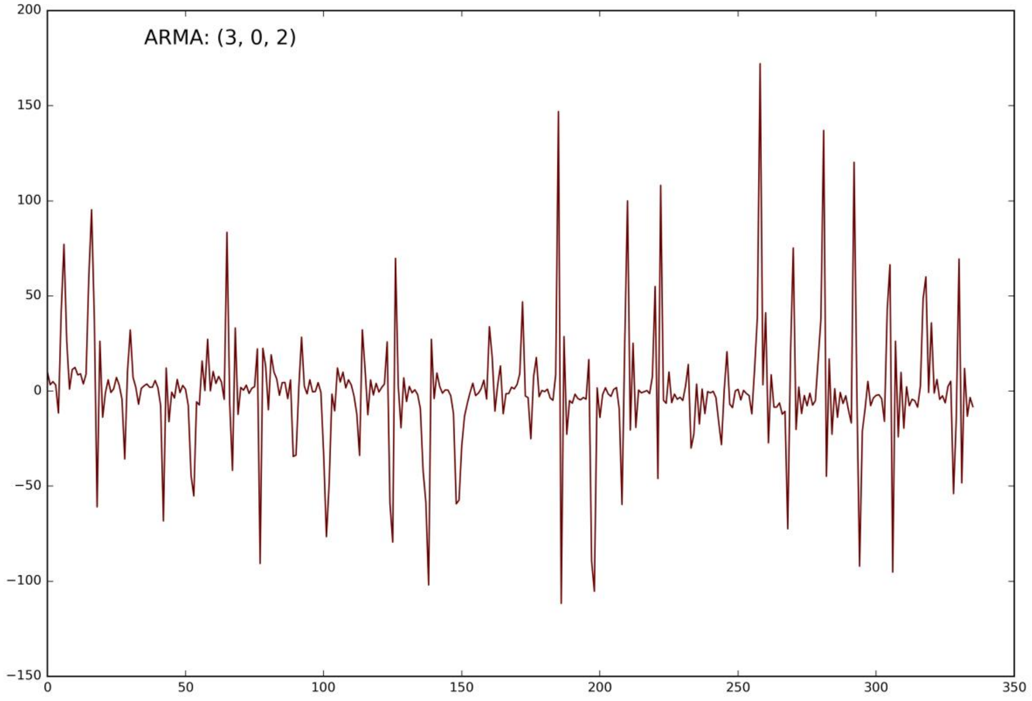

ARMA: (2, 0, 4)
american_flow_r
forecast
95% confidence interval
500
400
300
200
100
0
−100
−200
1951
1953
1955
1957
1959
1961

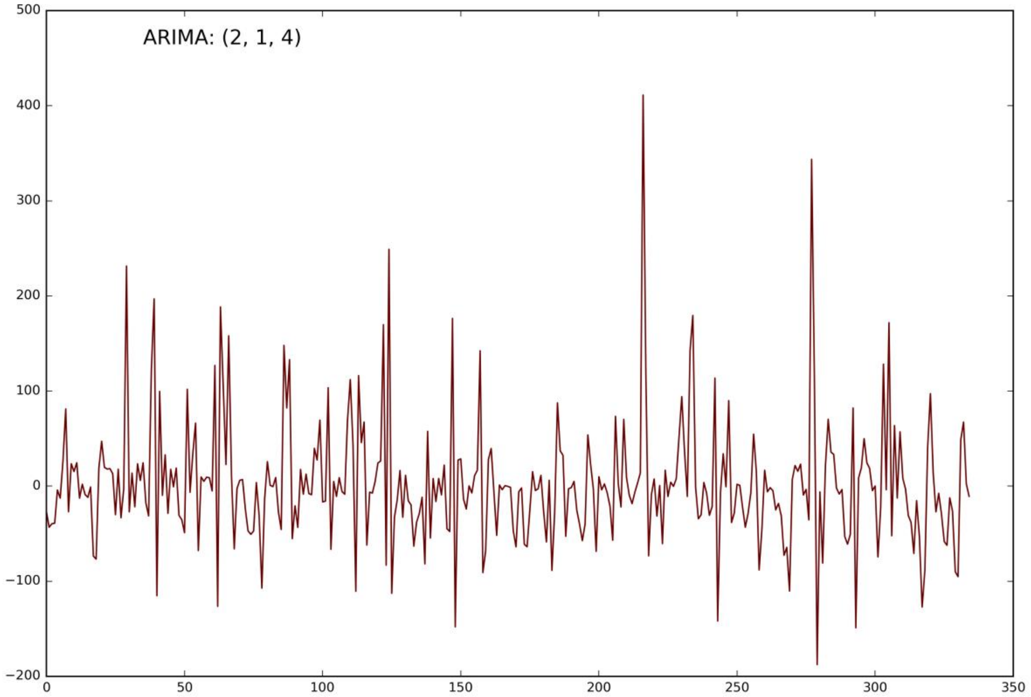

ARMA: (3, 0, 2)
columbia_flow_r
forecast
95% confidence interval
200
150
100
50
0
−50
−100
1951
1953
1955
1957
1959
1961

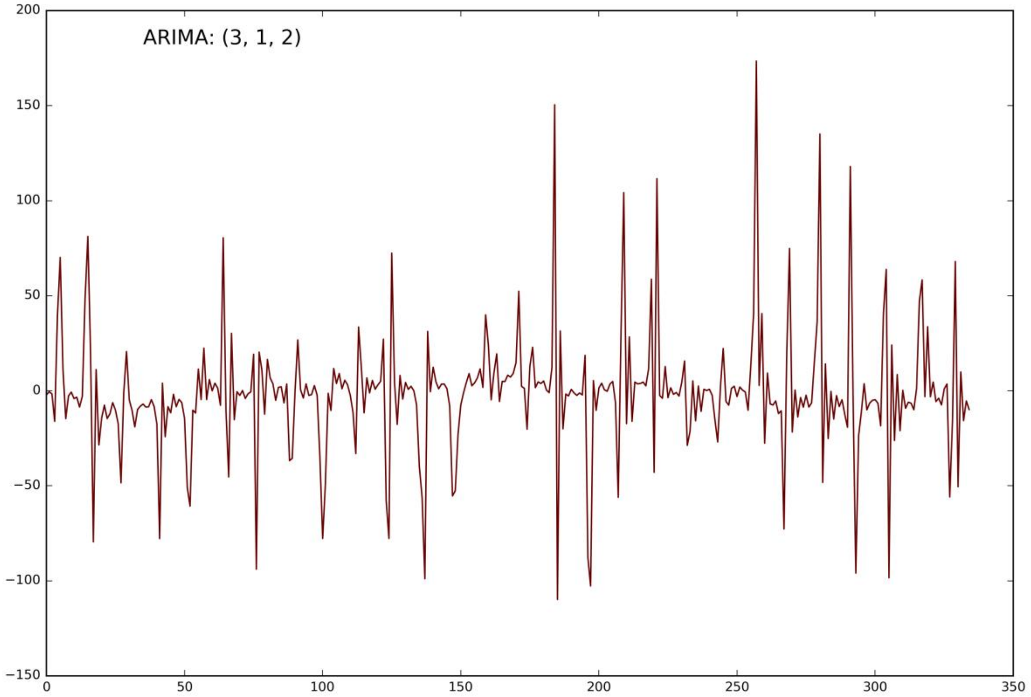

ARIMA: (2, 1, 4)
american_flow_r
forecast
95% confidence interval
500
400
300
200
100
0
−100
−200
1951
1953
1955
1957
1959
1961


ARIMA: (3, 1, 2)
columbia_flow_r
forecast
95% confidence interval
200
150
100
50
0
−50
−100
1951
1953
1955
1957
1959
1961

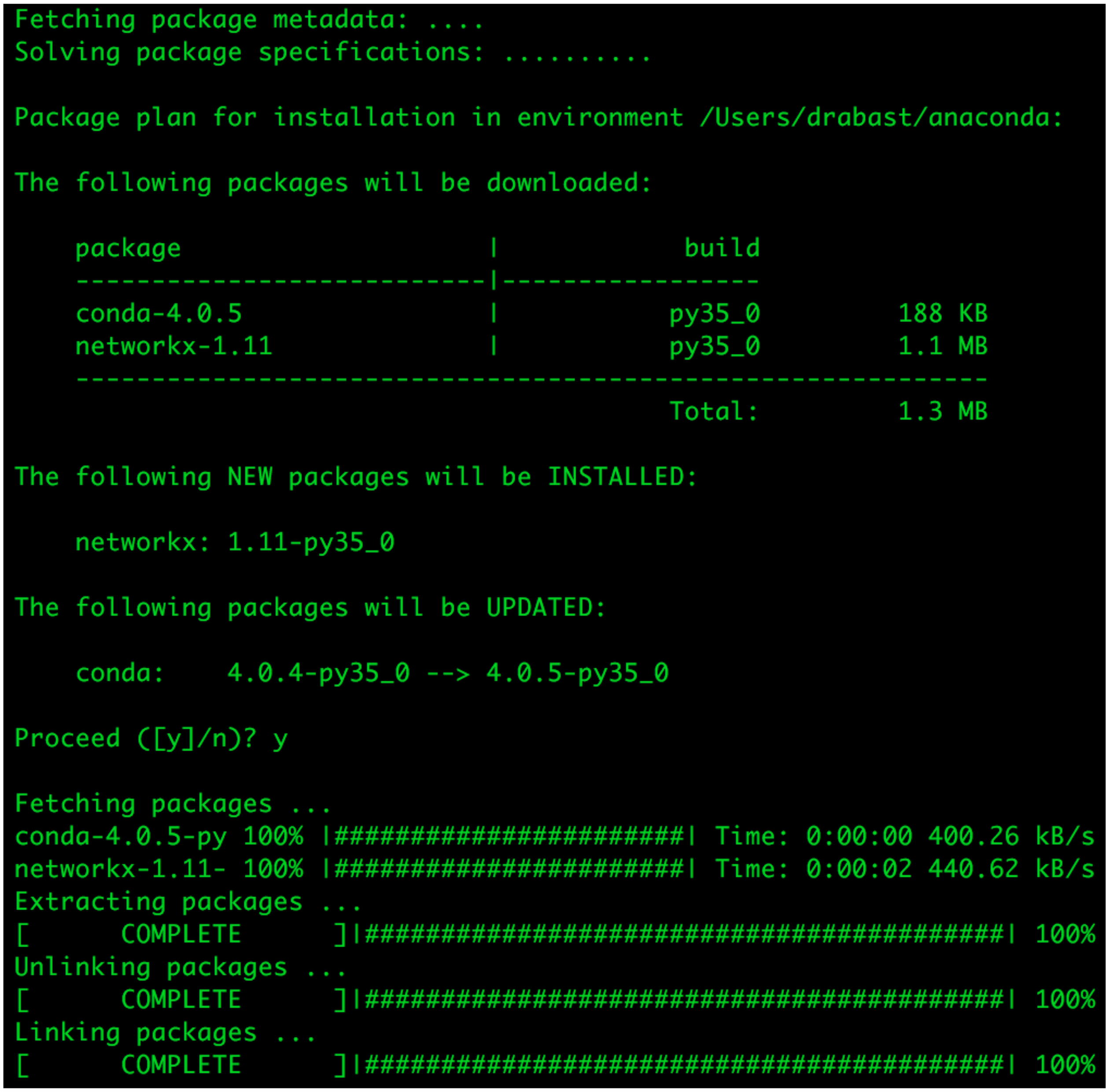

ARMA: (2, 0, 4)
500
400
300
200
100
0
−100
−200
0
50
100
150
200
250
300
350


ARMA: (3, 0, 2)
200
150
100
50
0
−50
−100
−150
0
50
100
150
200
250
300
350

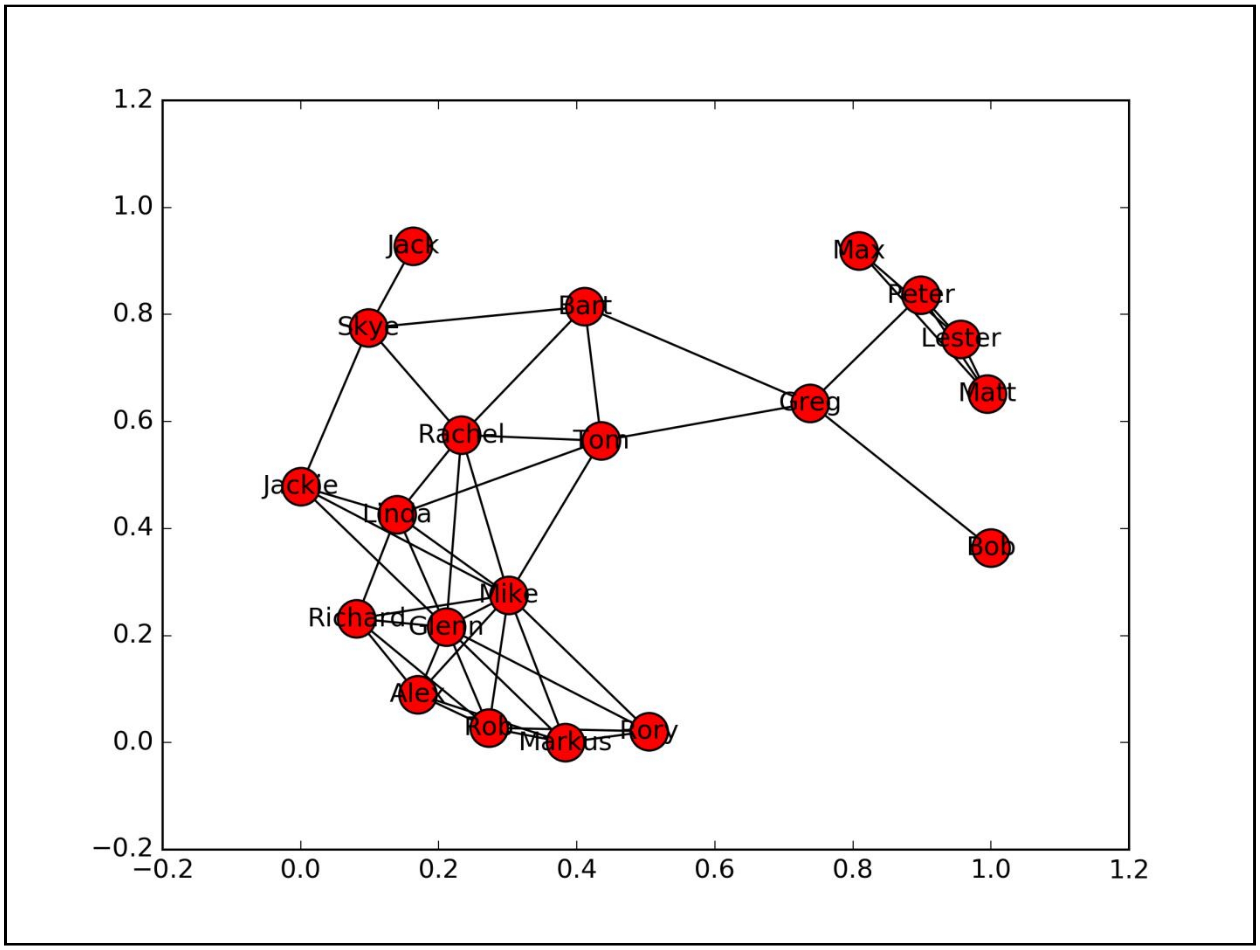

ARIMA: (2, 1, 4)
500
400
300
200
100
0
−100
−200
0
50
100
150
200
250
300
350

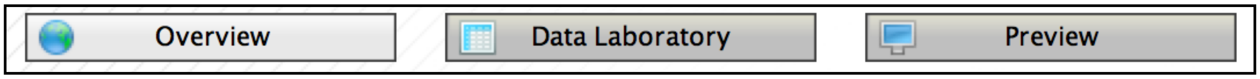

ARIMA: (3, 1, 2)
200
150
100
50
0
−50
−100
−150
0
50
100
150
200
250
300
350

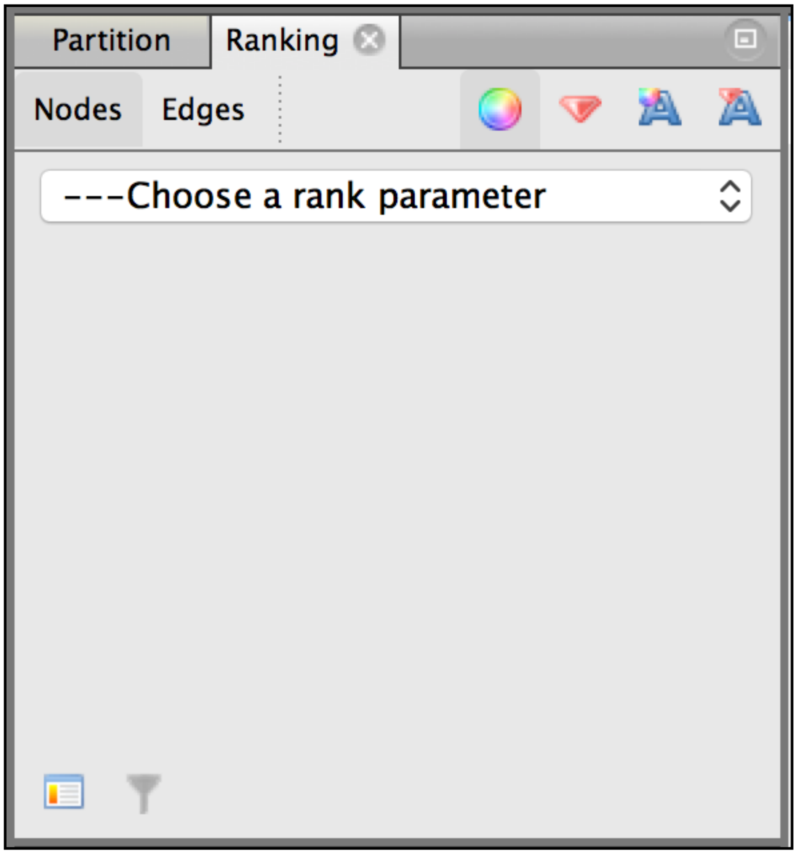
Chapter 8: Graphs

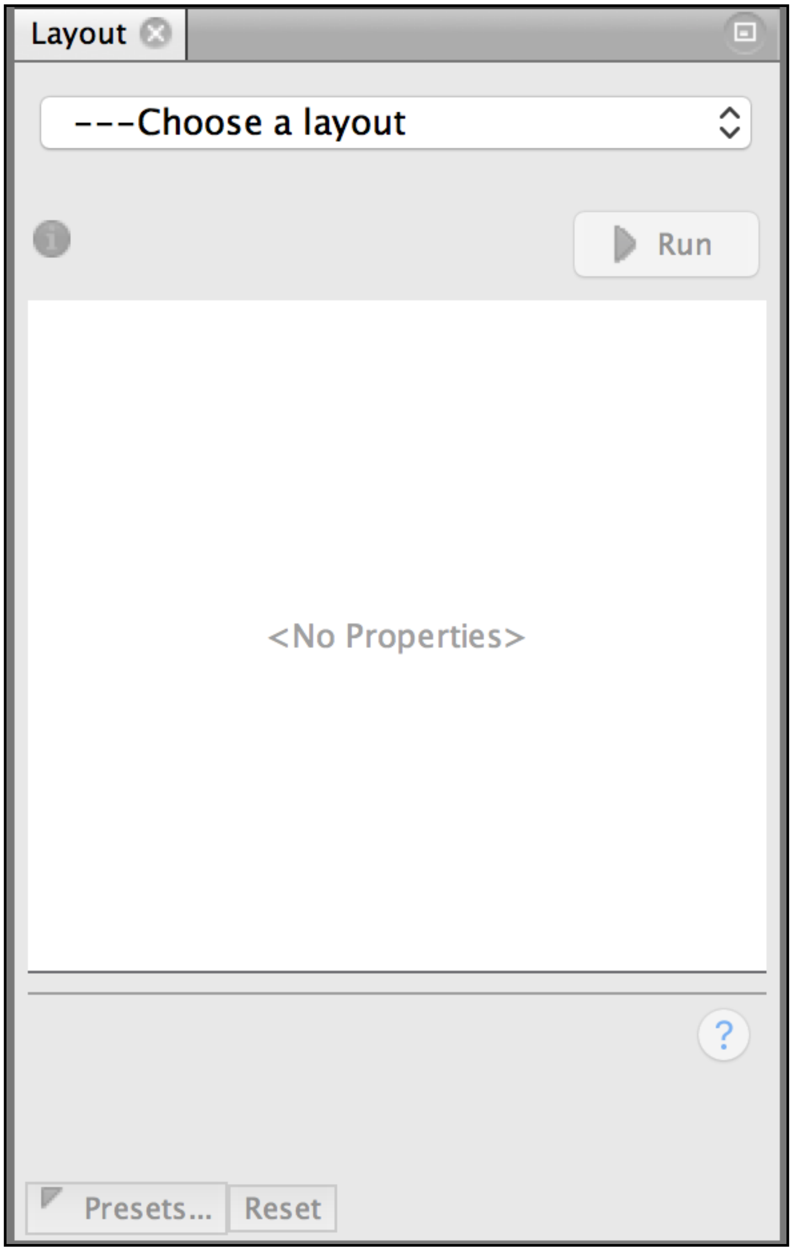


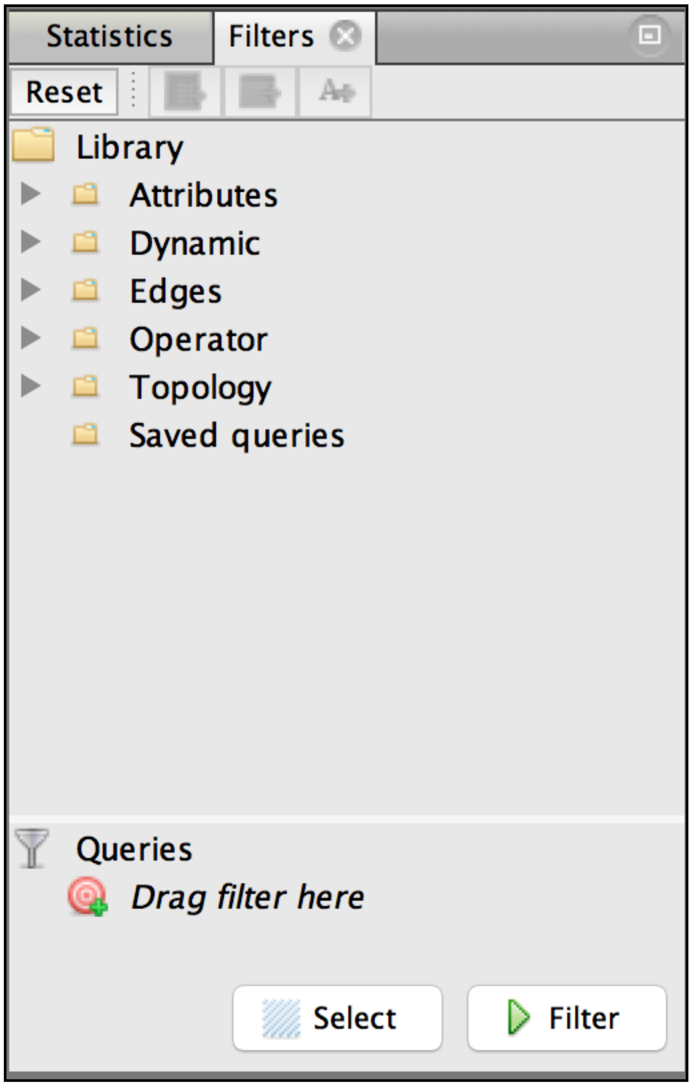
```
Fetching package metadata: ....
Solving package specifications: ..........

Package plan for installation in environment /Users/drabast/anaconda:

The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    conda-4.0.5                |           py35_0         188 KB
    networkx-1.11              |           py35_0         1.1 MB
    ------------------------------------------------------------
                                           Total:         1.3 MB

The following NEW packages will be INSTALLED:

    networkx: 1.11-py35_0

The following packages will be UPDATED:

    conda:    4.0.4-py35_0 --> 4.0.5-py35_0

Proceed ([y]/n)? y

Fetching packages ...
conda-4.0.5-py 100% |#######################| Time: 0:00:00 400.26 kB/s
networkx-1.11- 100% |#######################| Time: 0:00:02 440.62 kB/s
Extracting packages ...
[      COMPLETE      ]|##########################################| 100%
Unlinking packages ...
[      COMPLETE      ]|##########################################| 100%
Linking packages ...
[      COMPLETE      ]|##########################################| 100%
```

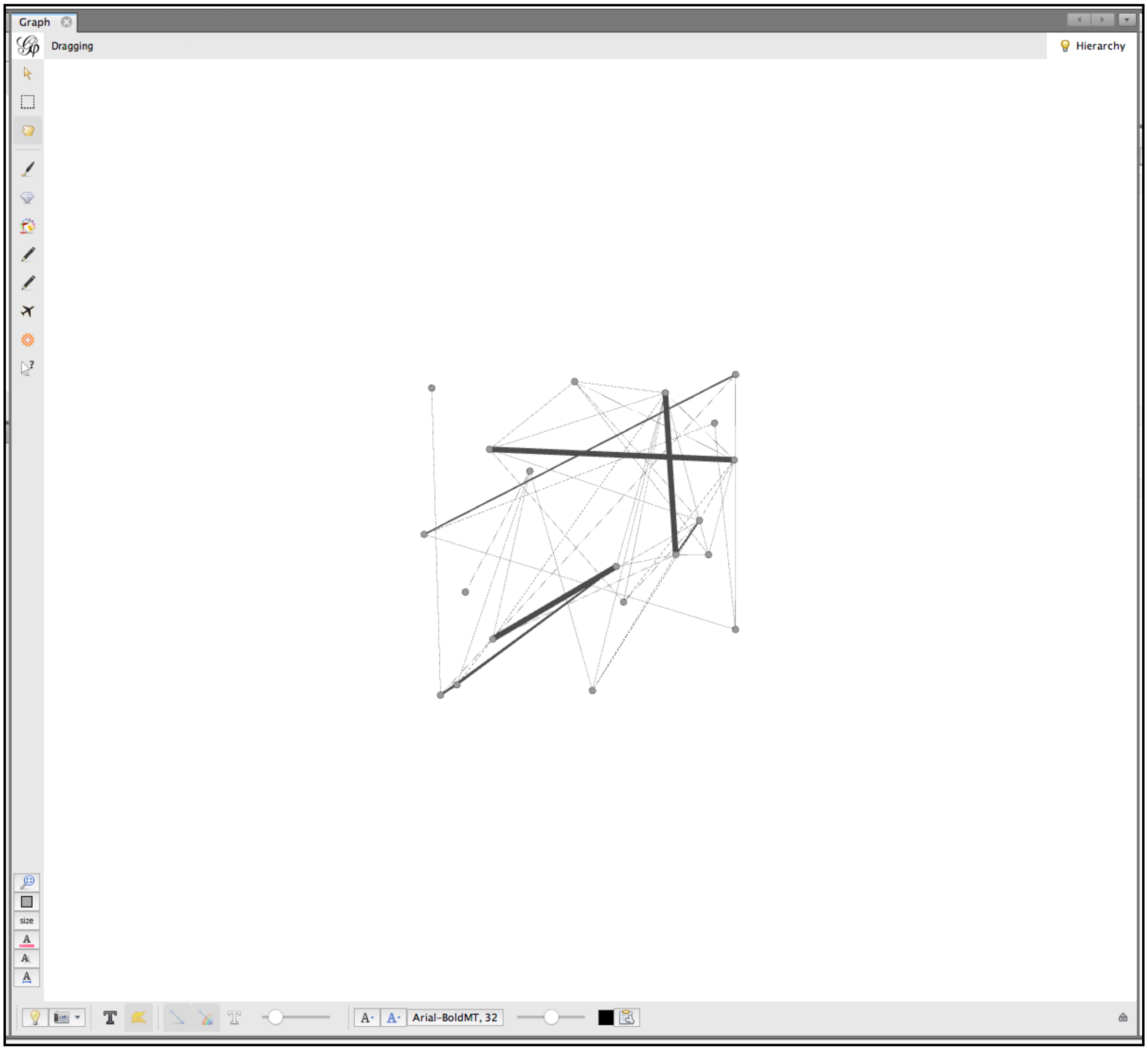

1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Jack
Max
Skye
Bart
Peter
Lester
Matt
Greg
Rachel
Tom
Jackie
Linda
Bob
Mike
Richard
Glenn
Alex
Rob
Markus
Rory

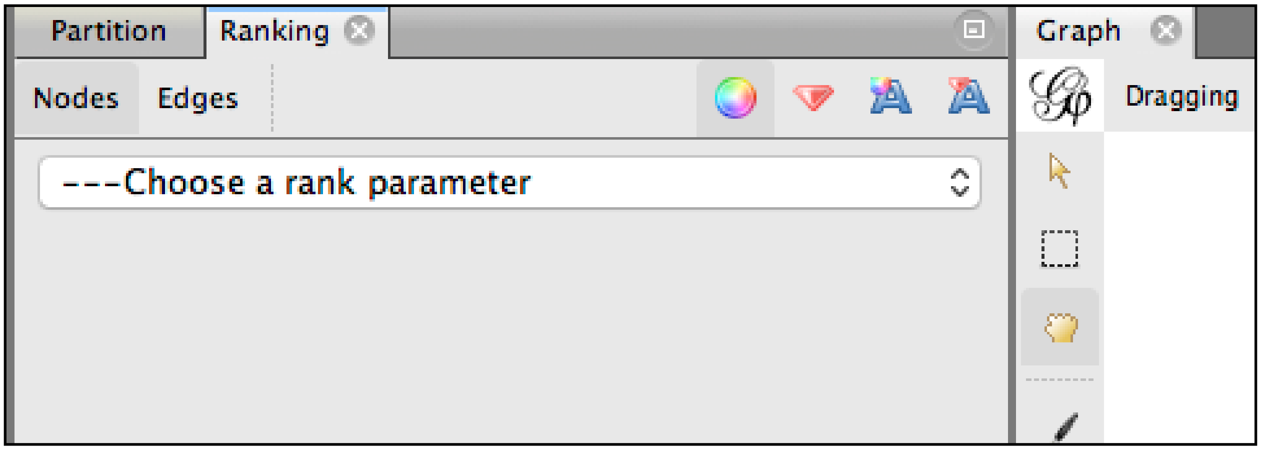

Overview
Data Laboratory
Preview

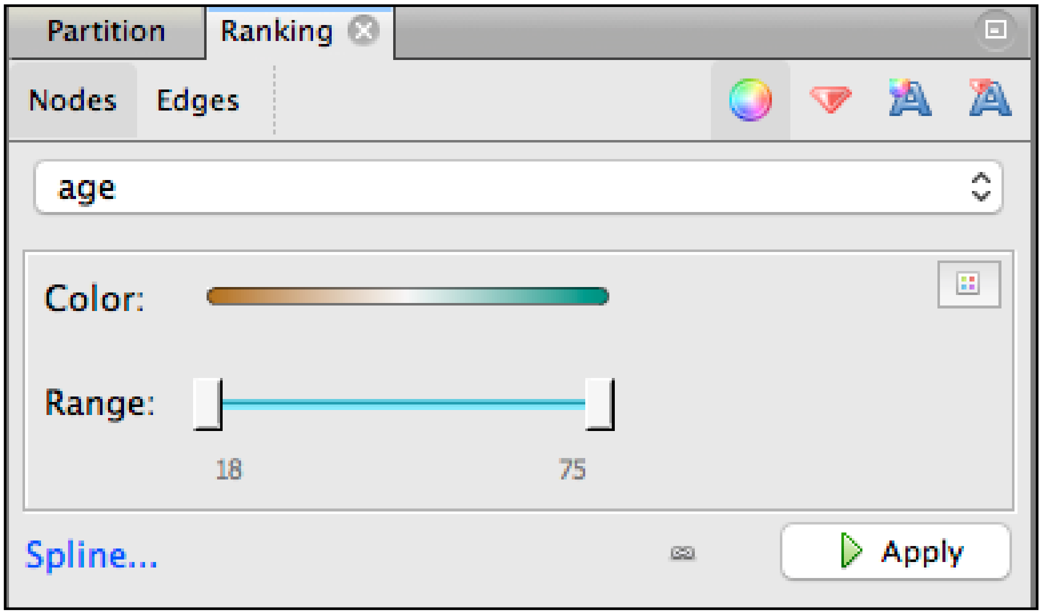

Partition
Ranking
Nodes
Edges
---Choose a rank parameter


Layout
---Choose a layout
Run
<No Properties>
?
Presets...
Reset


Statistics
Filters
Reset
Library
Attributes
Dynamic
Edges
Operator
Topology
Saved queries
Queries
Drag filter here
Select
Filter


Graph
Dragging
Hierarchy
size
Arial-BoldMT, 32

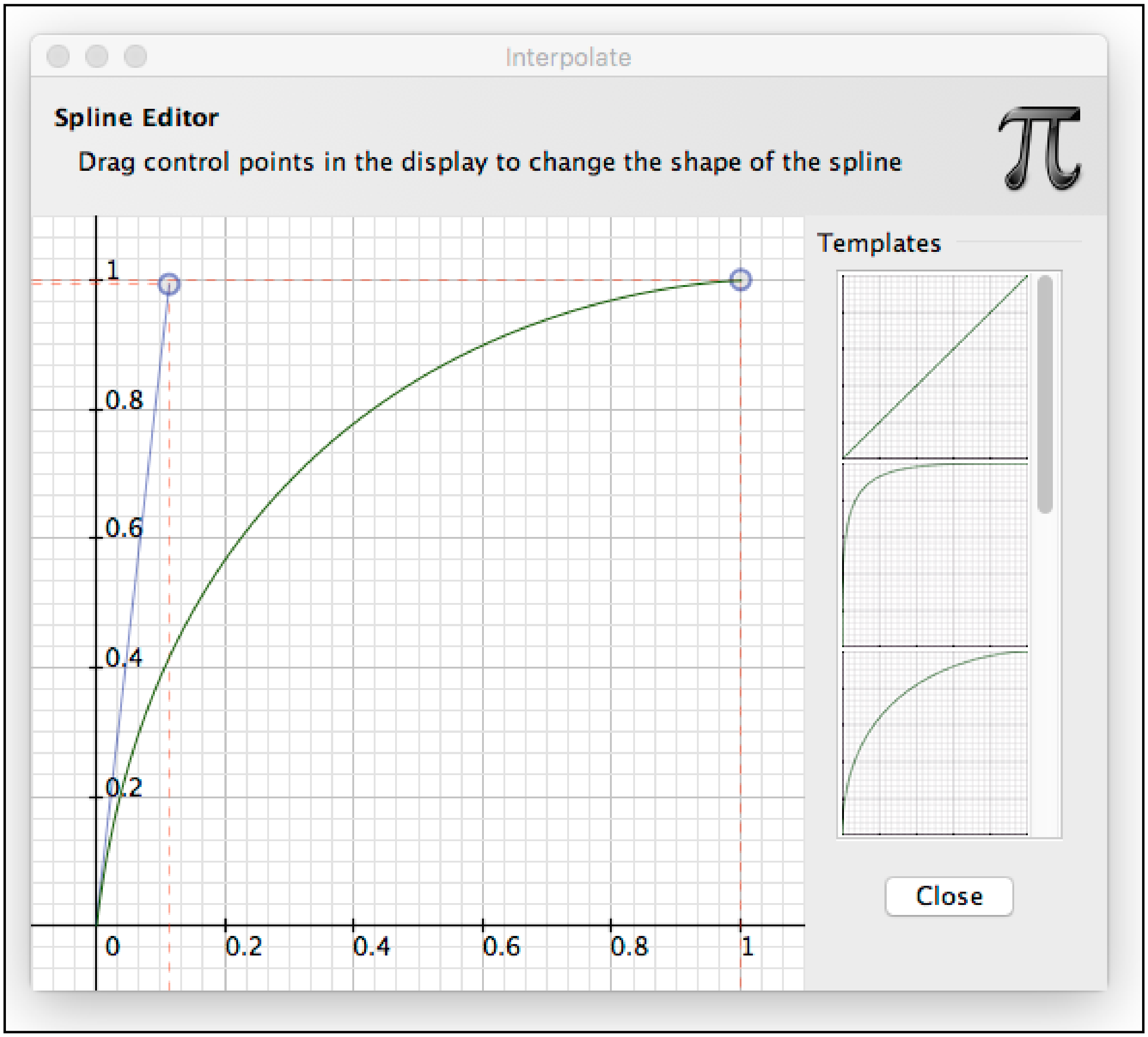

Partition
Ranking
Nodes
Edges
---Choose a rank parameter
Graph
Dragging



age

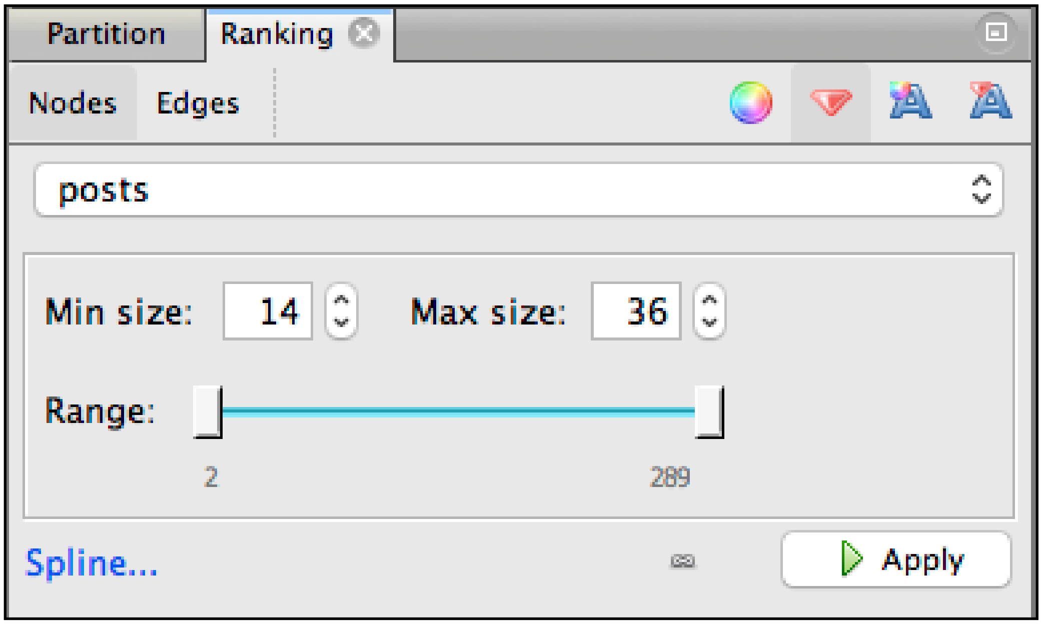
Color:

Range: 18 75



Spline Editor

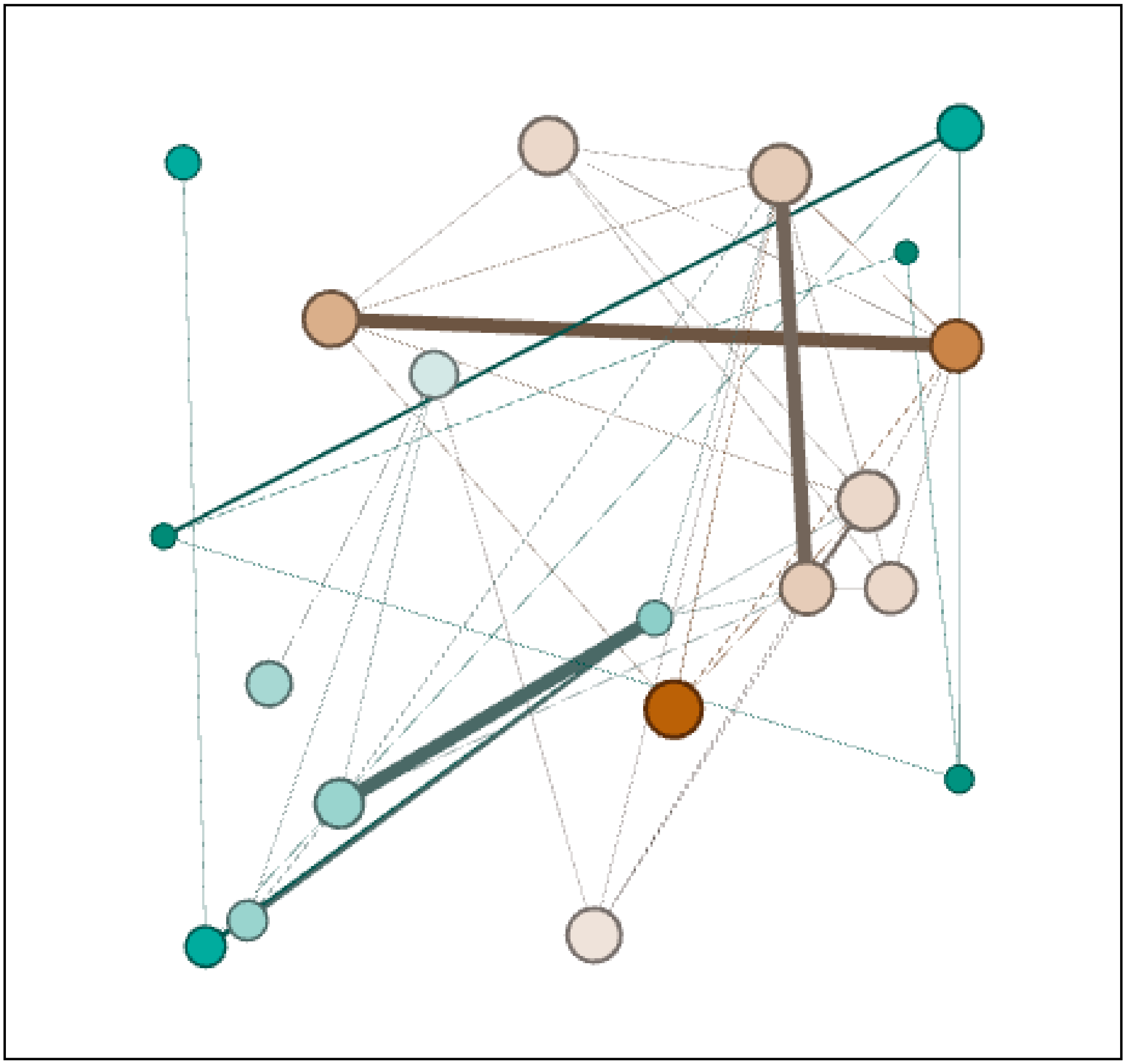
Drag control points in the display to change the shape of the spline

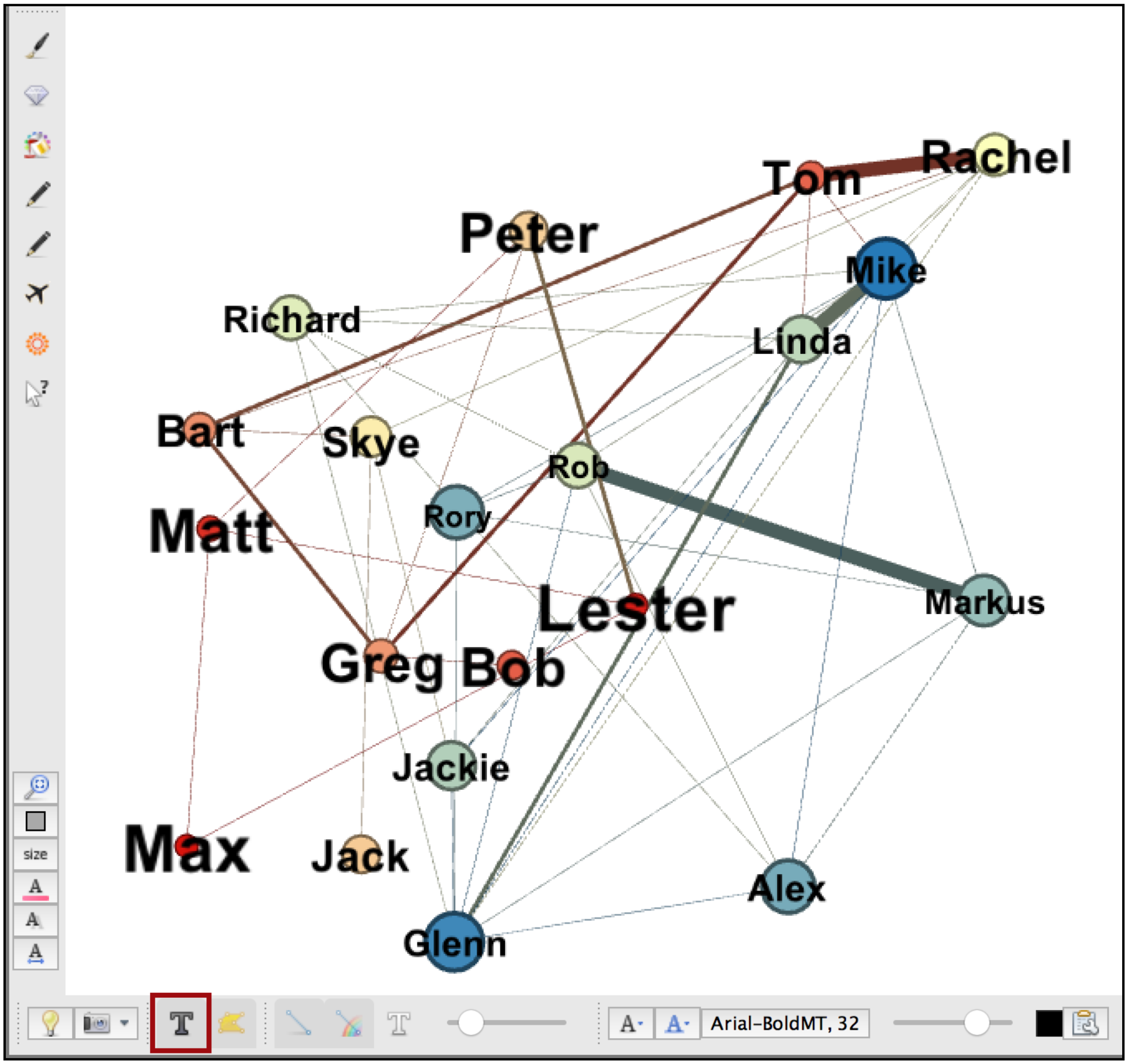


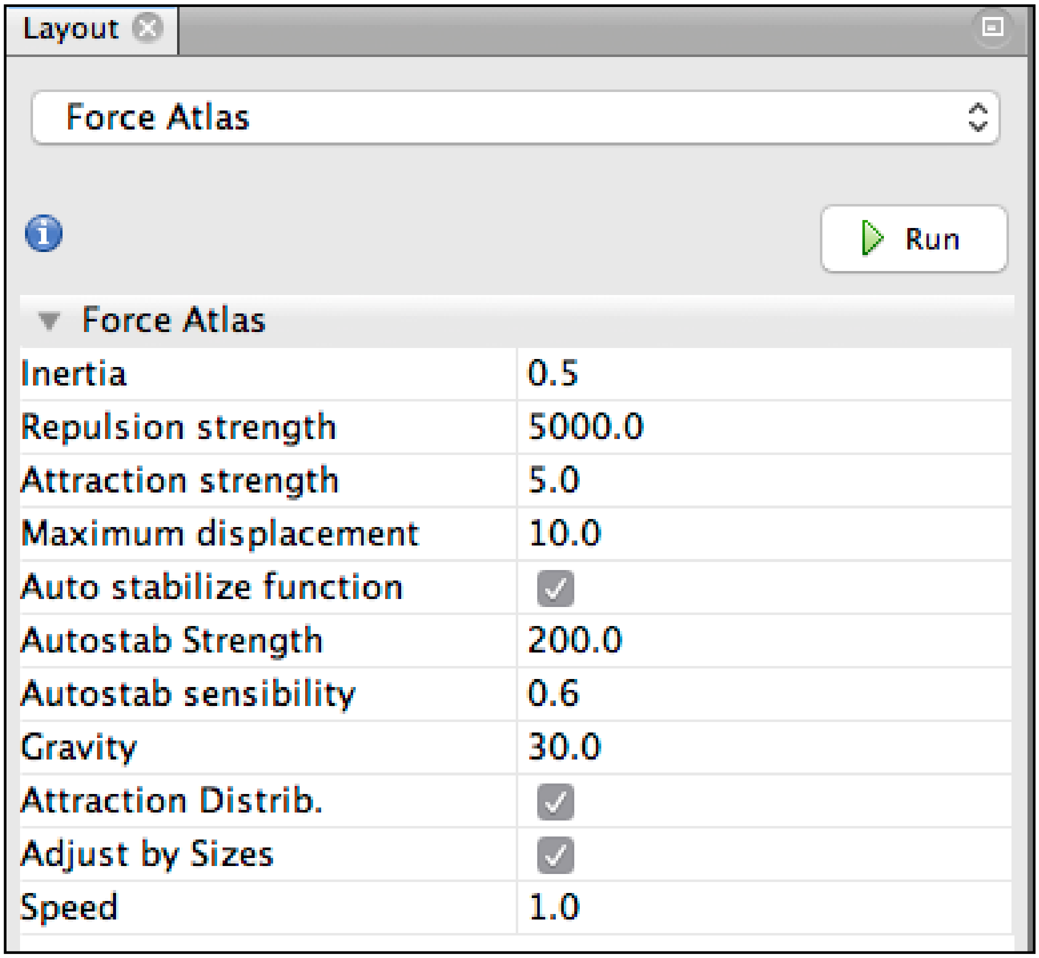
posts

Min size: 14 Max size: 36

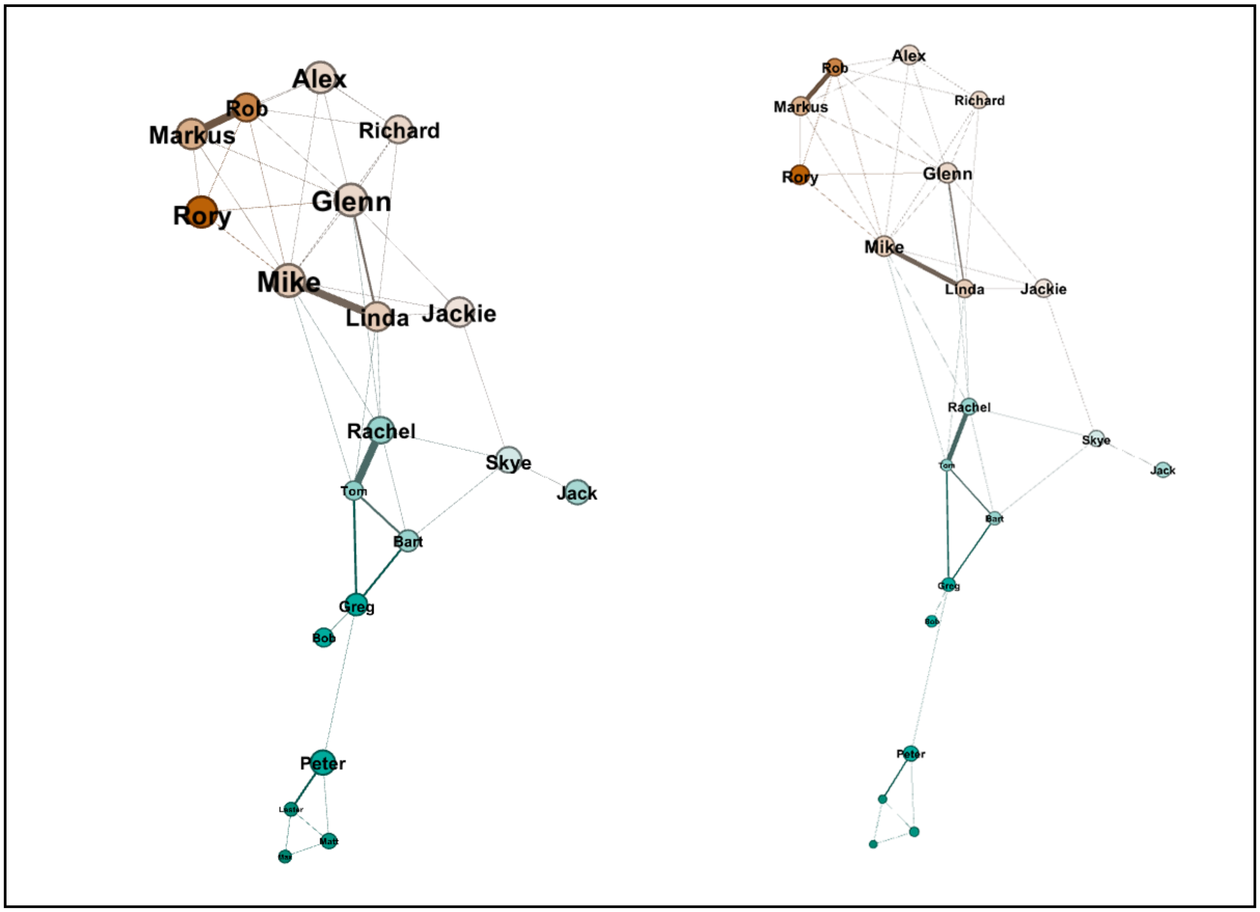
Range: 2 289

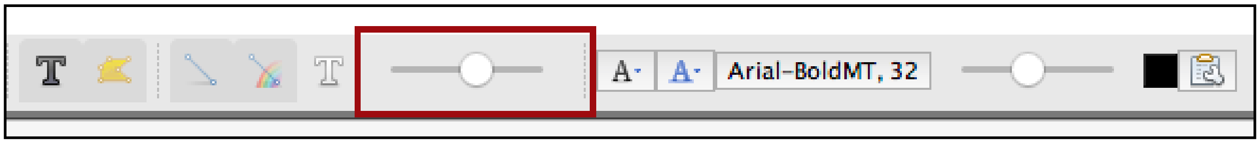



Tom
Rachel
Peter
Mike
Richard
Linda
Bart
Skye
Rob
Rory
Matt
Markus
Lester
Greg
Bob
Jackie
Max
Jack
Alex
Glenn
size
Arial-BoldMT, 32

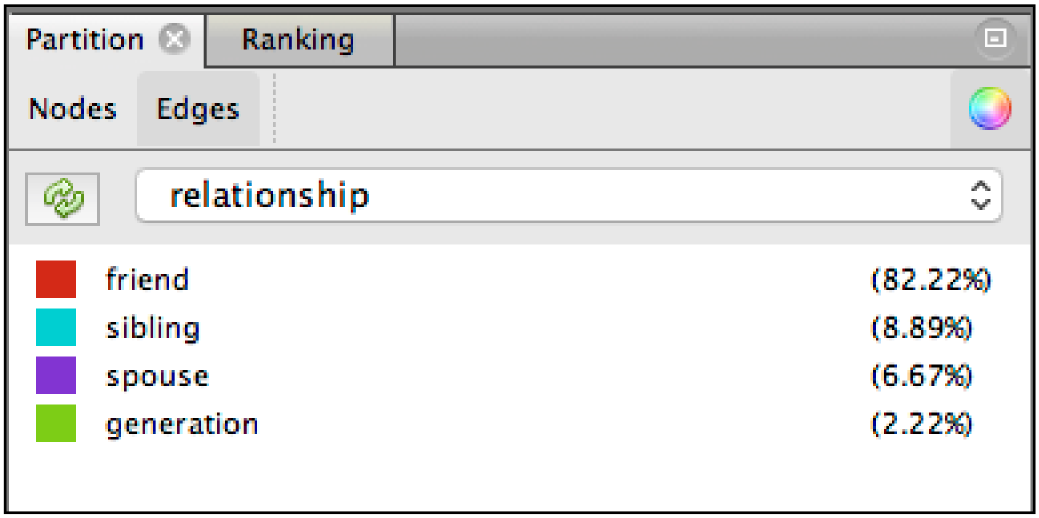


Inertia	0.5
Repulsion strength	5000.0
Attraction strength	5.0
Maximum displacement	10.0
Auto stabilize function	☑
Autostab Strength	200.0
Autostab sensibility	0.6
Gravity	30.0
Attraction Distrib.	☑
Adjust by Sizes	☑
Speed	1.0



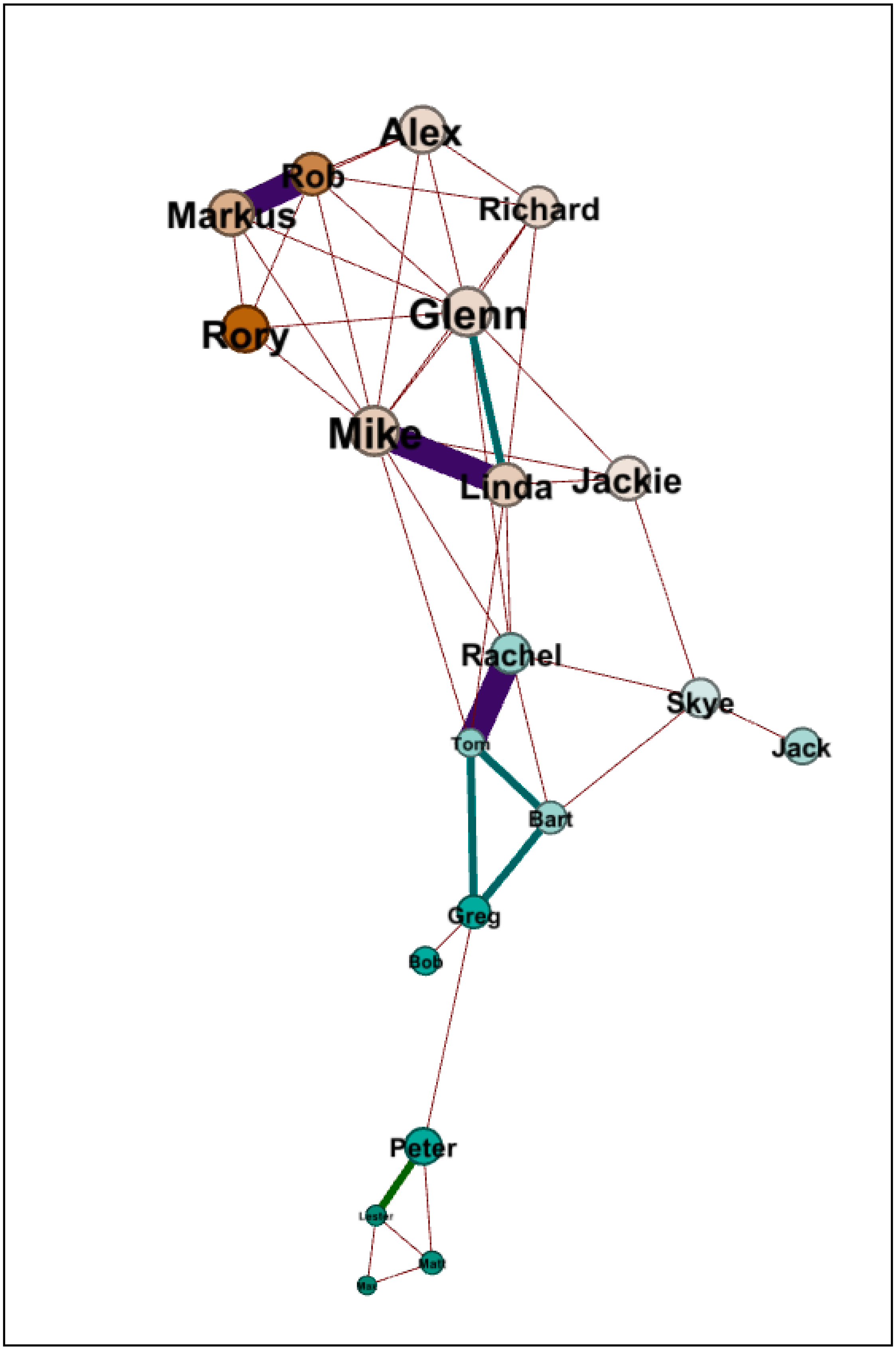


Arial-BoldMT, 32

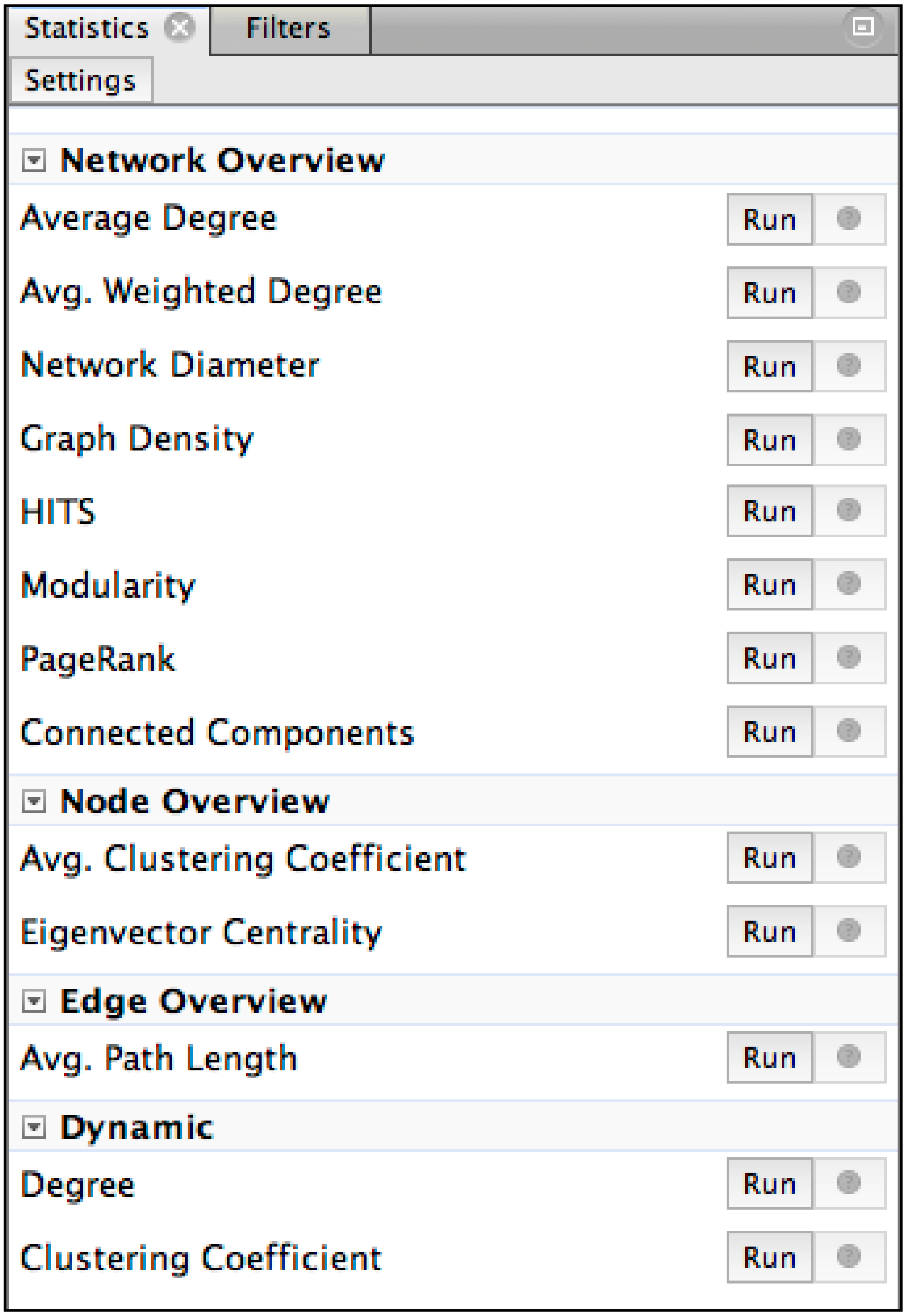


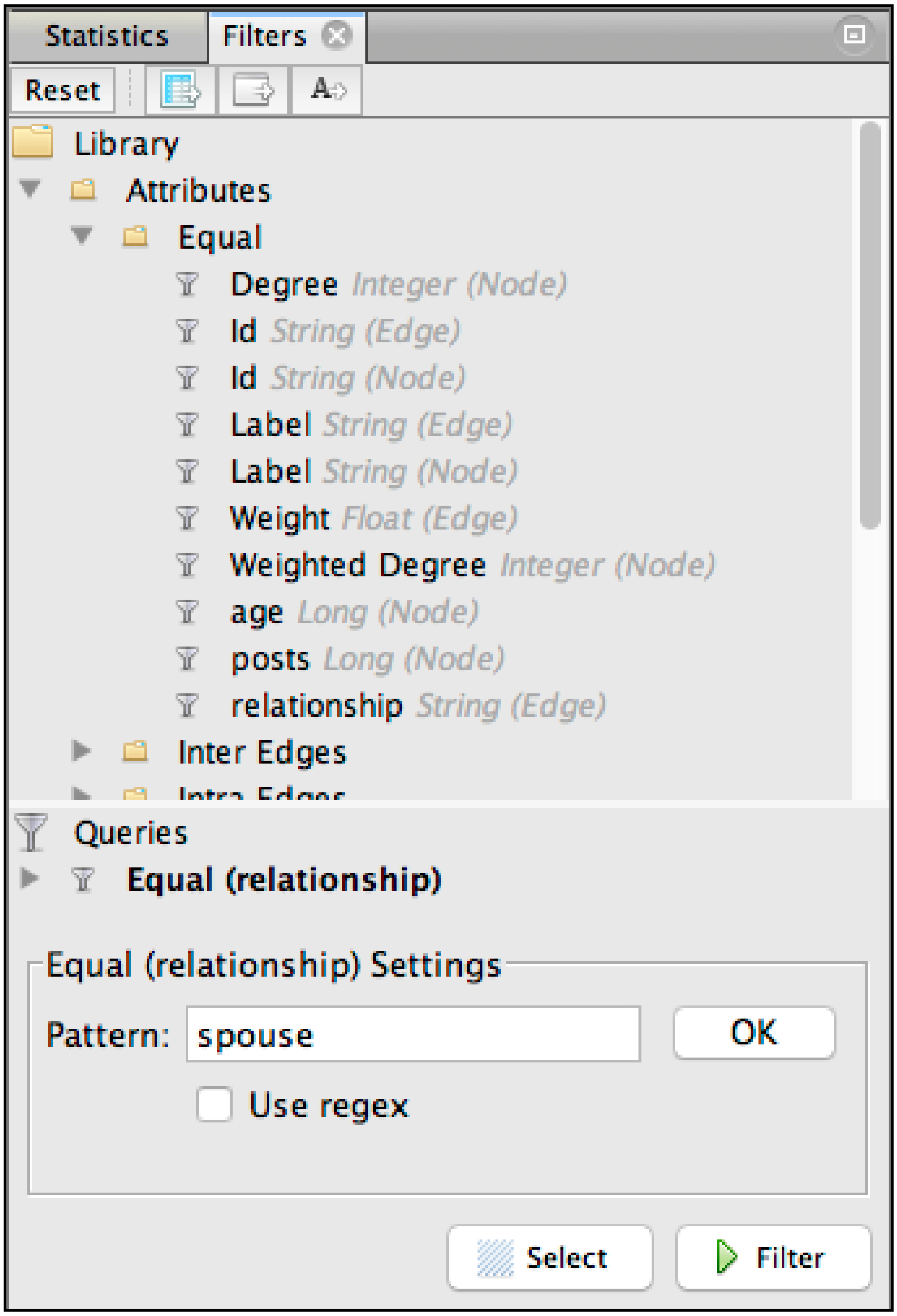
relationship

🟥	friend	(82.22%)
🟦	sibling	(8.89%)
🟪	spouse	(6.67%)
🟩	generation	(2.22%)



Alex
Rob
Markus
Richard
Glenn
Rory
Mike
Linda
Jackie
Rachel
Skye
Jack
Tom
Bart
Greg
Bob
Peter
Lester
Matt
Max

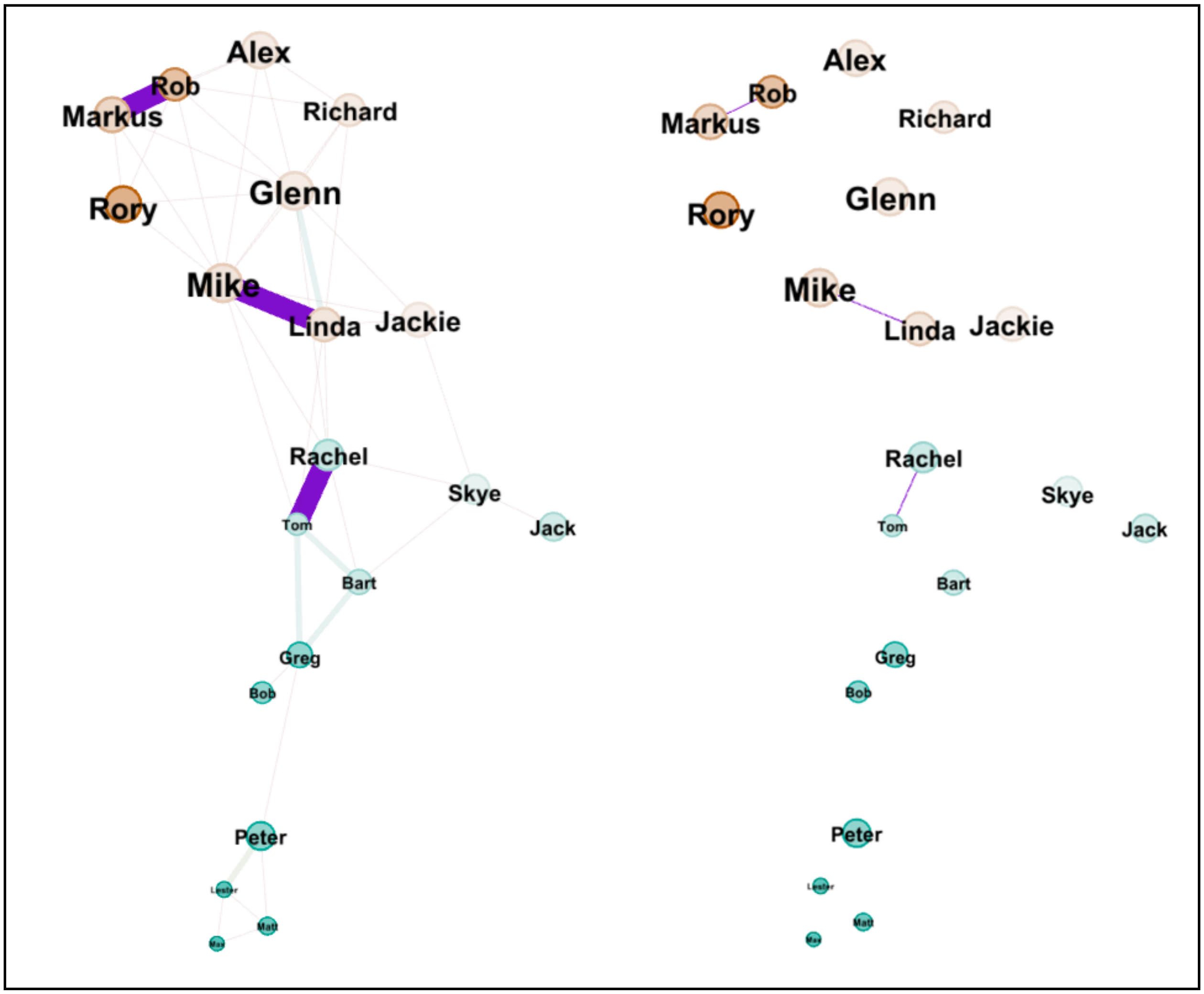

Statistics
Filters
Settings
Network Overview
Average Degree Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
HITS Run
Modularity Run
PageRank Run
Connected Components Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Degree Run
Clustering Coefficient Run

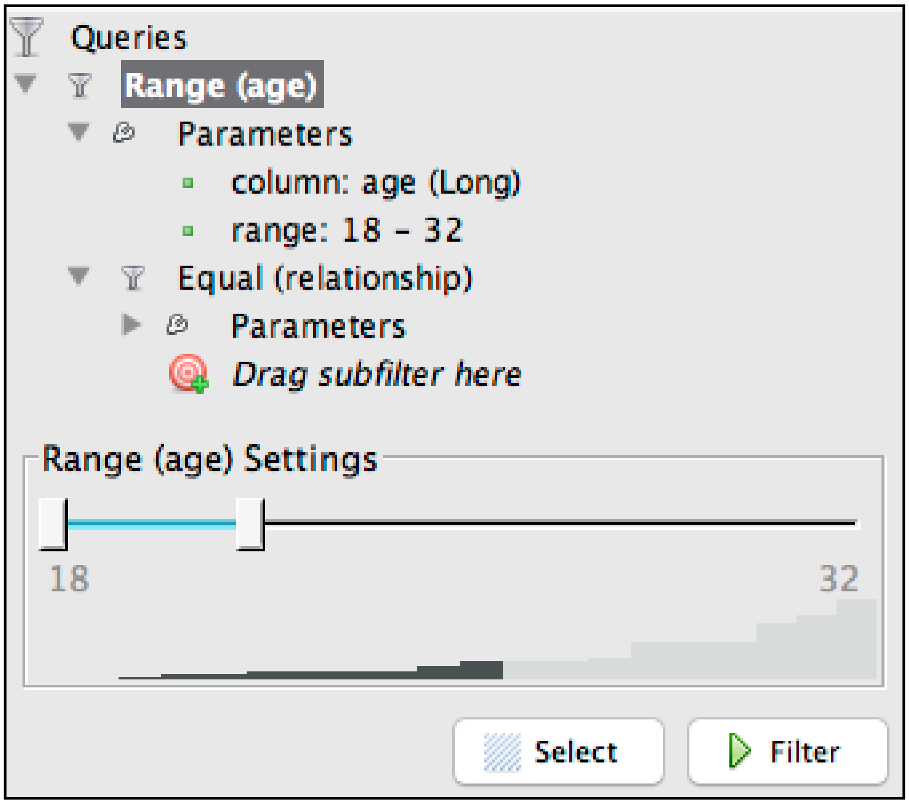


Equal (relationship) Settings

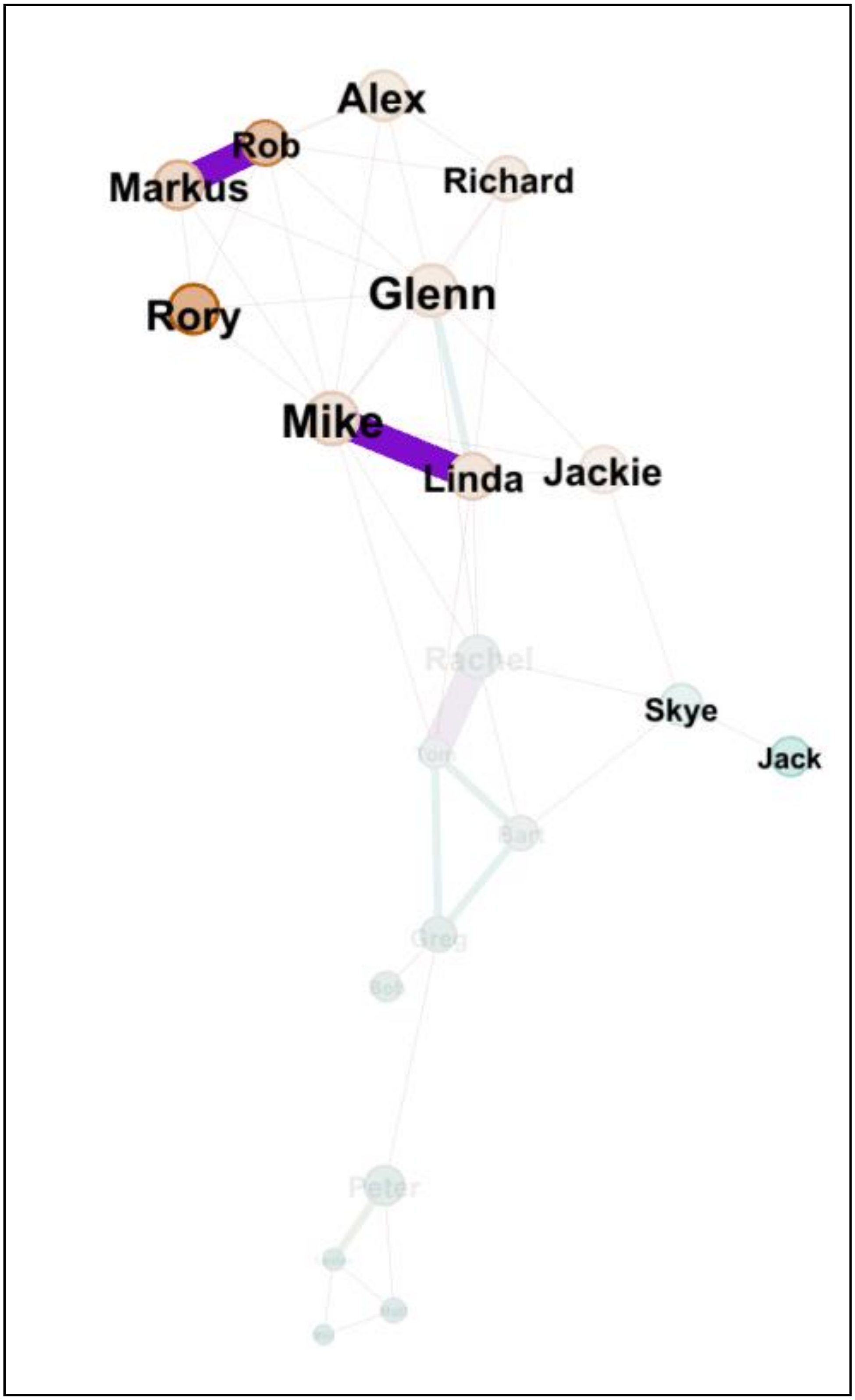
Pattern: spouse

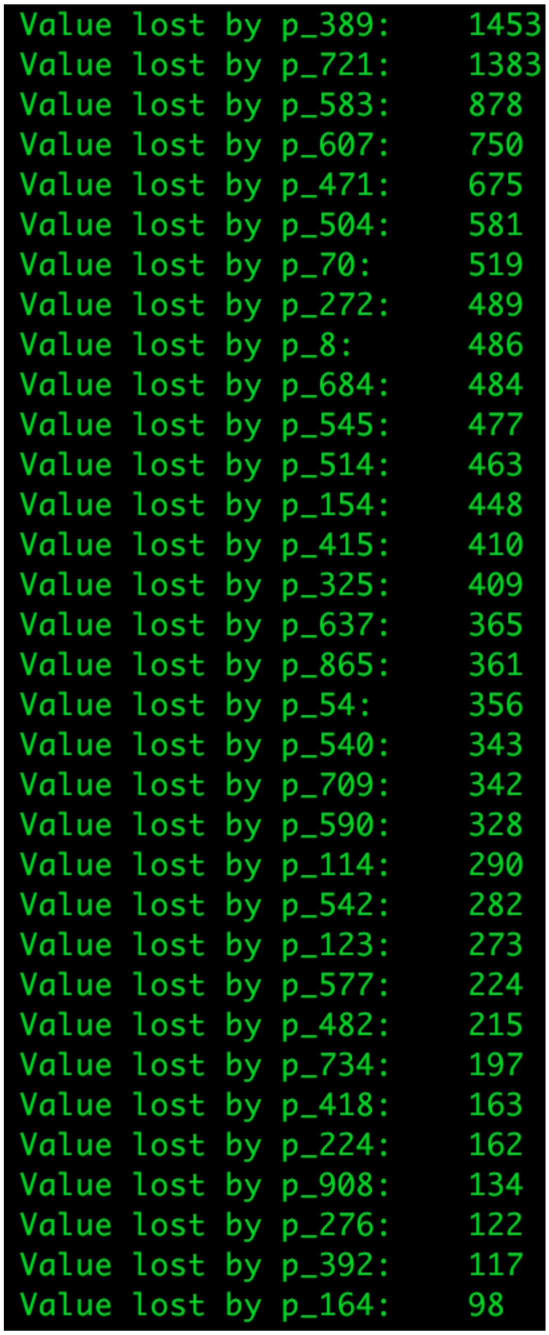
□ Use regex


Alex
Rob
Markus
Richard
Rory
Glenn
Mike
Linda
Jackie
Rachel
Skye
Tom
Jack
Bart
Greg
Bob
Peter
Lester
Matt
Max
Alex
Rob
Markus
Richard
Rory
Glenn
Mike
Linda
Jackie
Rachel
Skye
Tom
Jack
Bart
Greg
Bob
Peter
Lester
Matt
Max

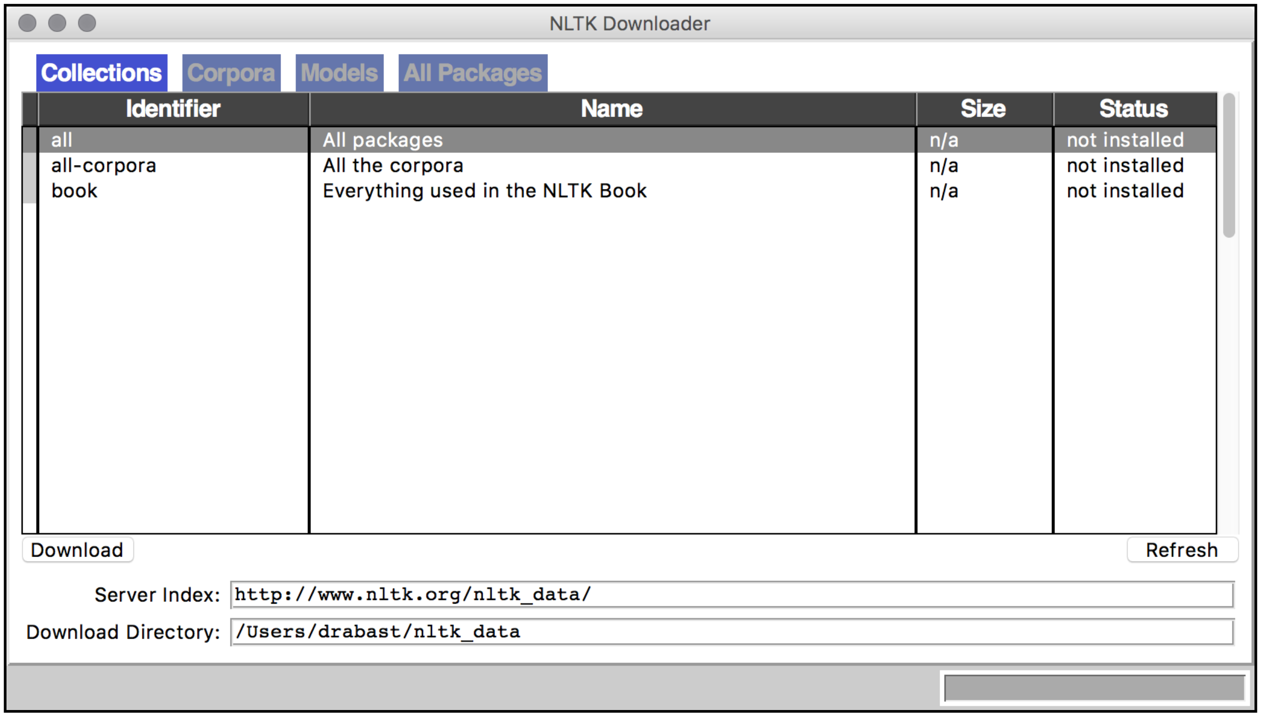


Queries
Range (age)
Parameters
column: age (Long)
range: 18 - 32
Equal (relationship)
Parameters
Drag subfilter here


Alex
Rob
Markus
Richard
Glenn
Rory
Mike
Linda
Jackie
Rachel
Skye
Jack
Tom
Bart
Greg
Bob
Peter



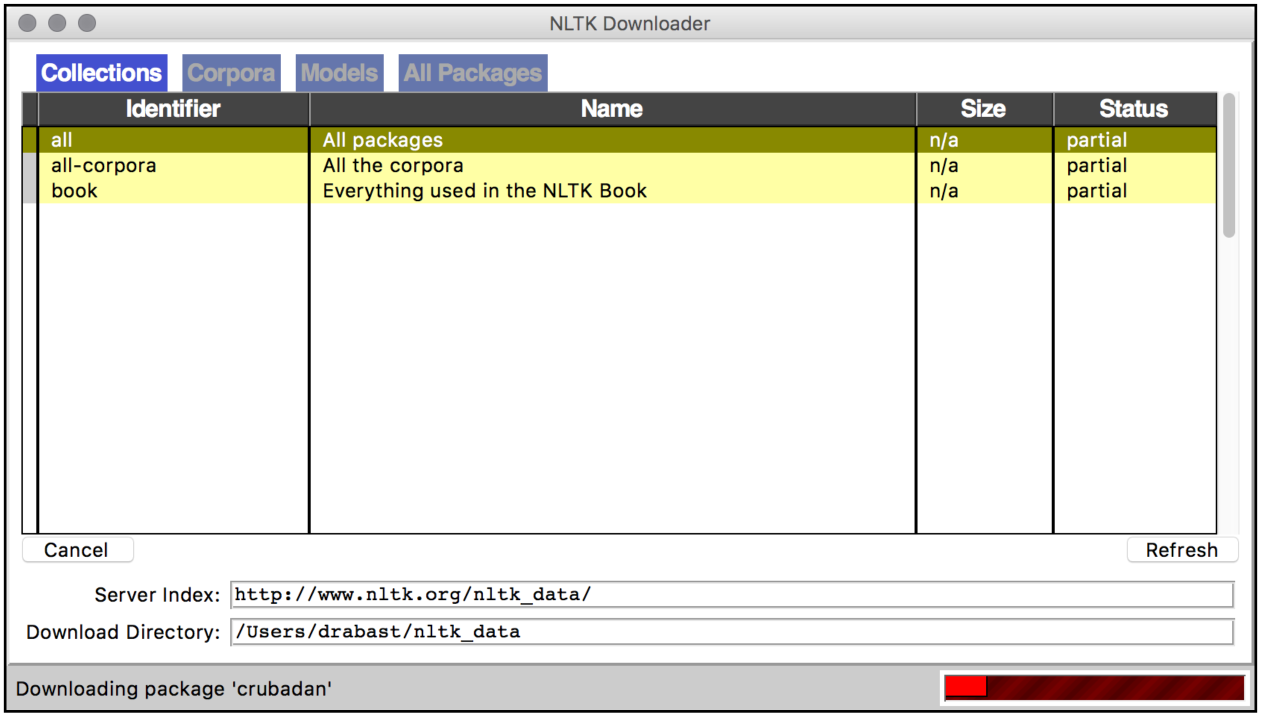
```
Value lost by p_389:    1453
Value lost by p_721:    1383
Value lost by p_583:    878
Value lost by p_607:    750
Value lost by p_471:    675
Value lost by p_504:    581
Value lost by p_70:     519
Value lost by p_272:    489
Value lost by p_8:      486
Value lost by p_684:    484
Value lost by p_545:    477
Value lost by p_514:    463
Value lost by p_154:    448
Value lost by p_415:    410
Value lost by p_325:    409
Value lost by p_637:    365
Value lost by p_865:    361
Value lost by p_54:     356
Value lost by p_540:    343
Value lost by p_709:    342
Value lost by p_590:    328
Value lost by p_114:    290
Value lost by p_542:    282
Value lost by p_123:    273
Value lost by p_577:    224
Value lost by p_482:    215
Value lost by p_734:    197
Value lost by p_418:    163
Value lost by p_224:    162
Value lost by p_908:    134
Value lost by p_276:    122
Value lost by p_392:    117
Value lost by p_164:    98
```



Identifier	Name	Size	Status
all	All packages	n/a	not installed
all-corpora	All the corpora	n/a	not installed
book	Everything used in the NLTK Book	n/a	not installed

Server Index: `http://www.nltk.org/nltk_data/`

Download Directory: `/Users/drabast/nltk_data`

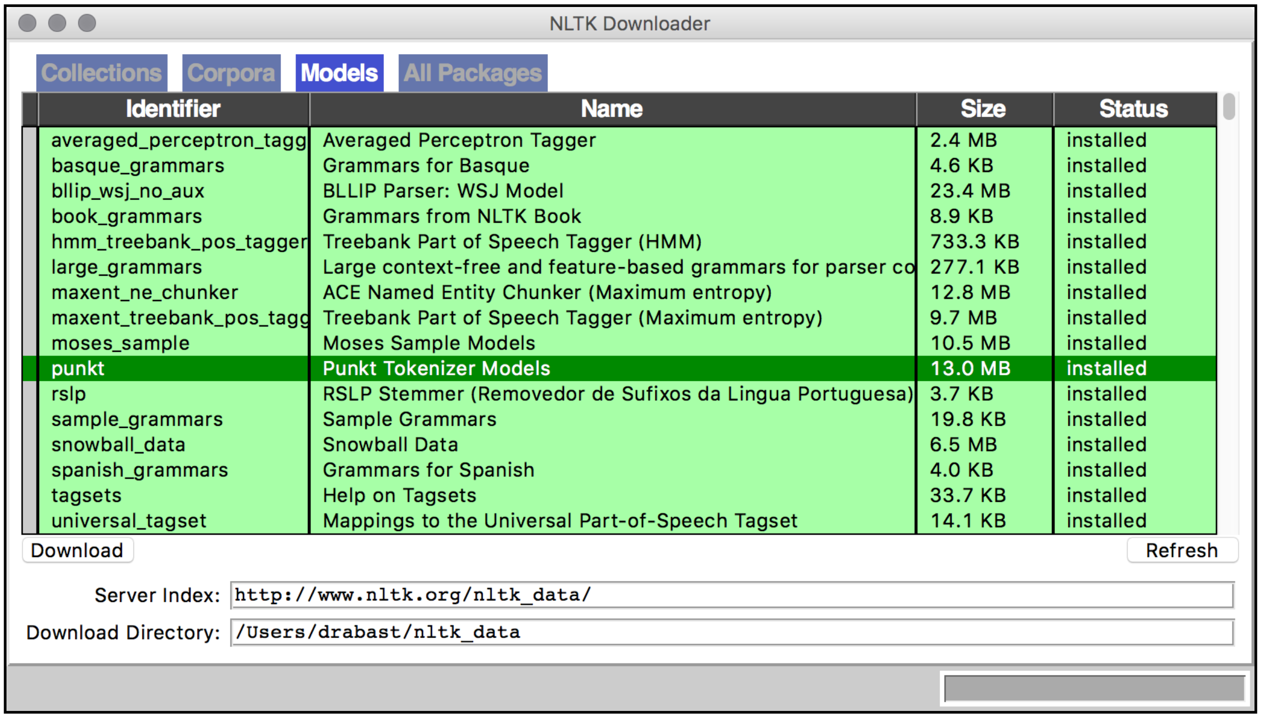


Identifier	Name	Size	Status
all	All packages	n/a	partial
all-corpora	All the corpora	n/a	partial
book	Everything used in the NLTK Book	n/a	partial

Server Index: `http://www.nltk.org/nltk_data/`

Download Directory: `/Users/drabast/nltk_data`

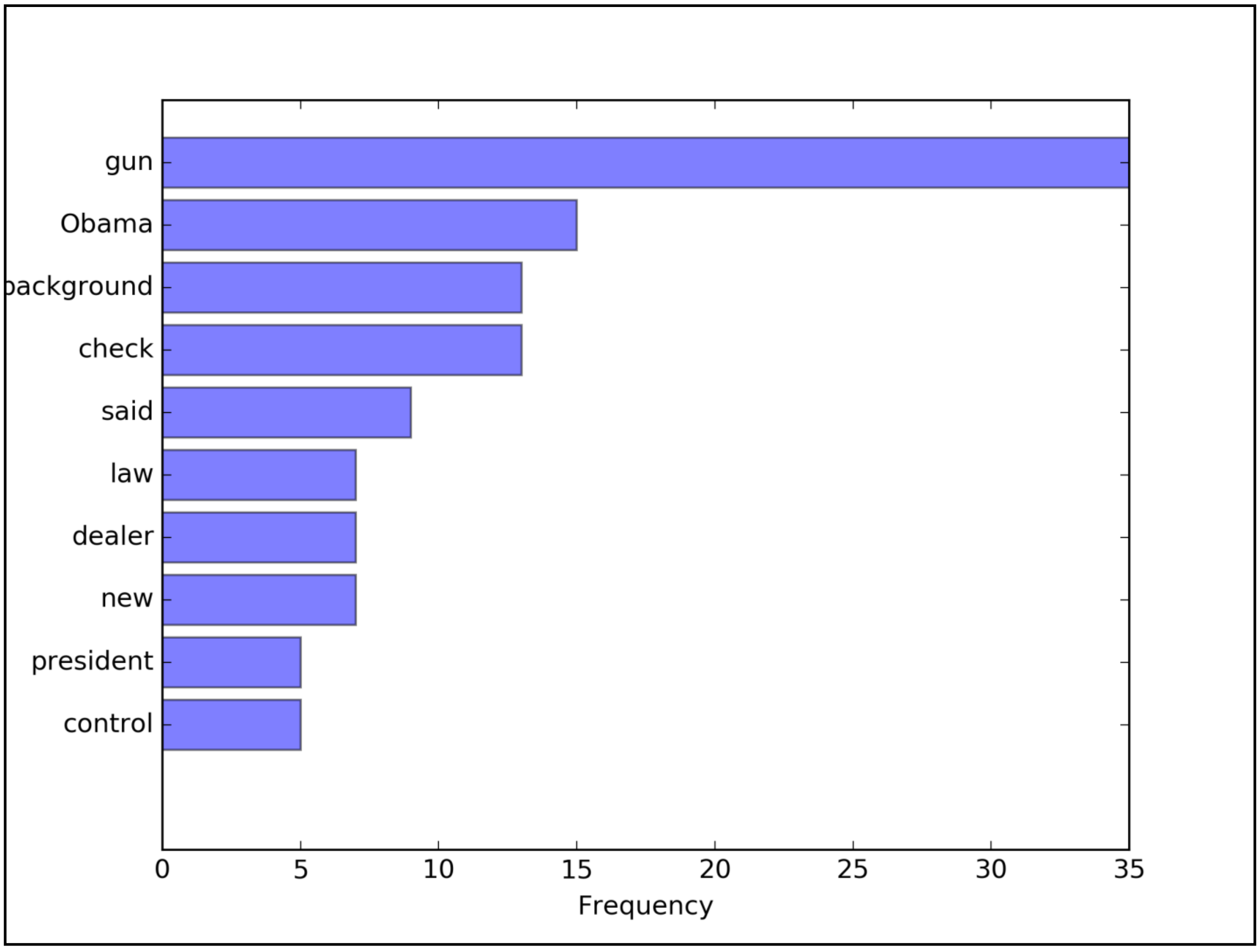
Downloading package 'crubadan'

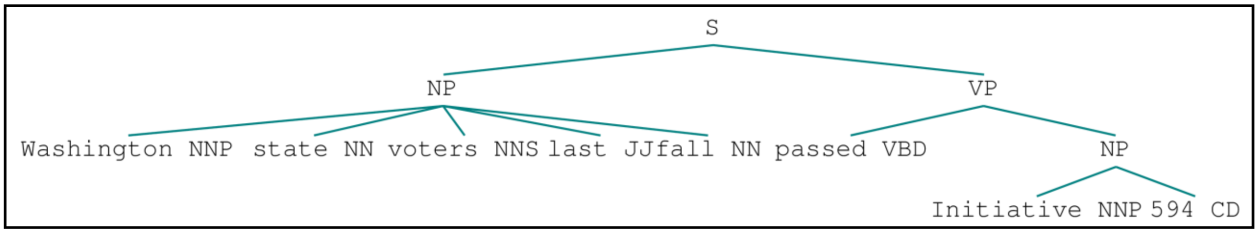


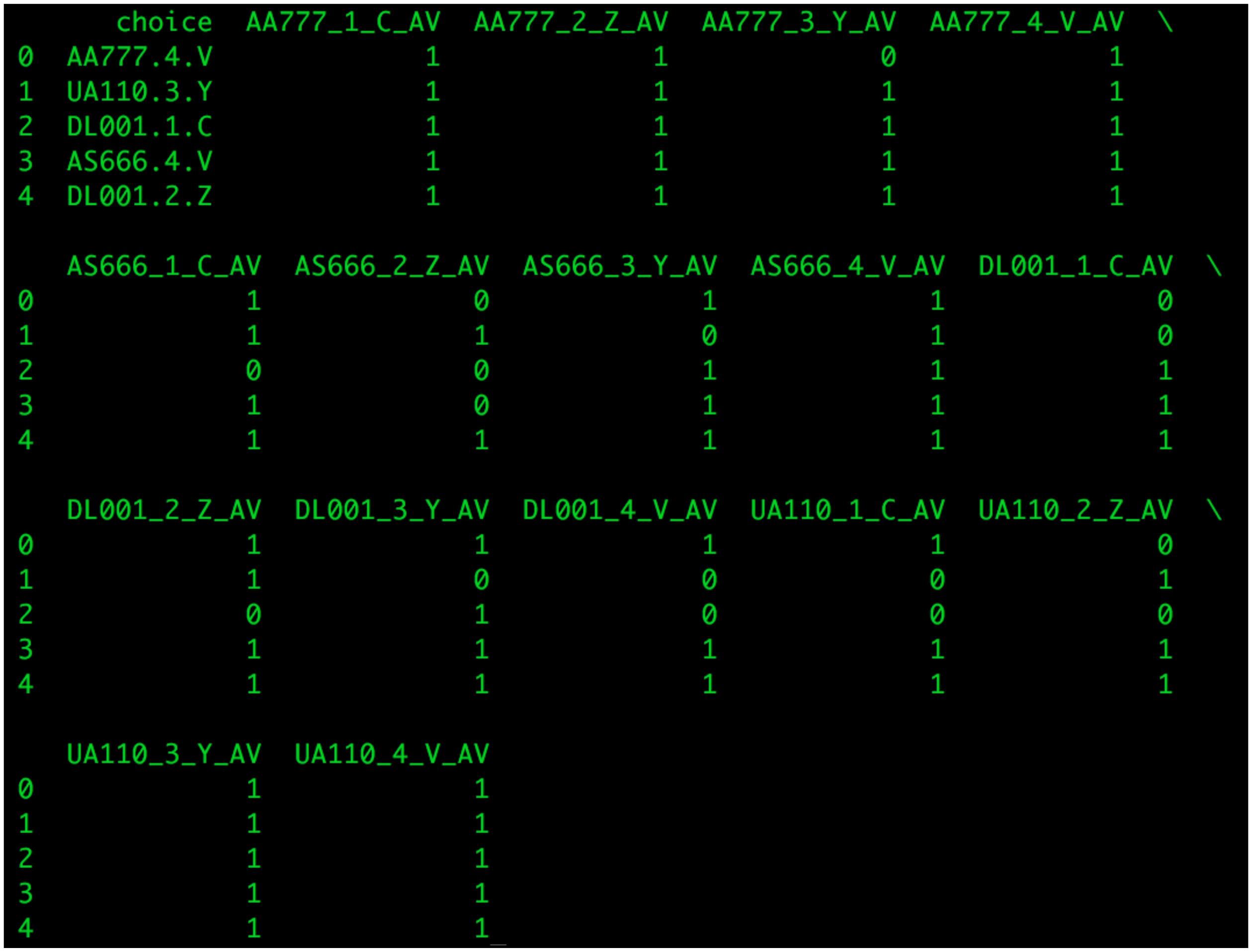
Identifier	Name	Size	Status
averaged_perceptron_tagg	Averaged Perceptron Tagger	2.4 MB	installed
basque_grammars	Grammars for Basque	4.6 KB	installed
bllip_wsj_no_aux	BLLIP Parser: WSJ Model	23.4 MB	installed
book_grammars	Grammars from NLTK Book	8.9 KB	installed
hmm_treebank_pos_tagger	Treebank Part of Speech Tagger (HMM)	733.3 KB	installed
large_grammars	Large context-free and feature-based grammars for parser co	277.1 KB	installed
maxent_ne_chunker	ACE Named Entity Chunker (Maximum entropy)	12.8 MB	installed
maxent_treebank_pos_tagg	Treebank Part of Speech Tagger (Maximum entropy)	9.7 MB	installed
moses_sample	Moses Sample Models	10.5 MB	installed
punkt	Punkt Tokenizer Models	13.0 MB	installed
rslp	RSLP Stemmer (Removedor de Sufixos da Lingua Portuguesa)	3.7 KB	installed
sample_grammars	Sample Grammars	19.8 KB	installed
snowball_data	Snowball Data	6.5 MB	installed
spanish_grammars	Grammars for Spanish	4.0 KB	installed
tagsets	Help on Tagsets	33.7 KB	installed
universal_tagset	Mappings to the Universal Part-of-Speech Tagset	14.1 KB	installed

Server Index: `http://www.nltk.org/nltk_data/`

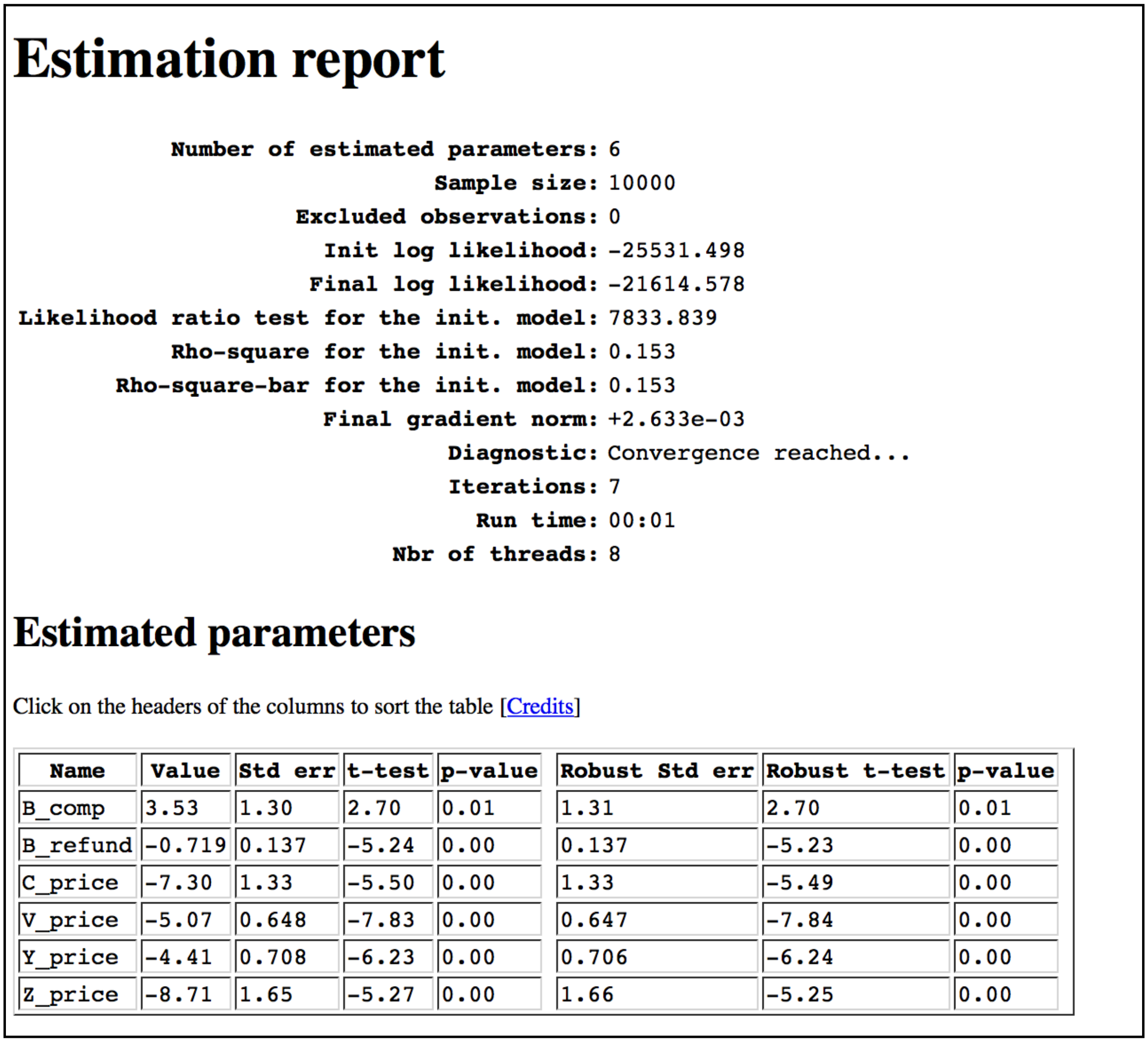
Download Directory: `/Users/drabast/nltk_data`




S
NP
VP
Washington NNP
state NN
voters NNS
last JJ
fall NN
passed VBD
NP
Initiative NNP
594 CD



```
      choice  AA777_1_C_AV  AA777_2_Z_AV  AA777_3_Y_AV  AA777_4_V_AV  \
0  AA777.4.V             1             1             0             1
1  UA110.3.Y             1             1             1             1
2  DL001.1.C             1             1             1             1
3  AS666.4.V             1             1             1             1
4  DL001.2.Z             1             1             1             1

   AS666_1_C_AV  AS666_2_Z_AV  AS666_3_Y_AV  AS666_4_V_AV  DL001_1_C_AV  \
0             1             0             1             1             0
1             1             1             0             1             0
2             0             0             1             1             1
3             1             0             1             1             1
4             1             1             1             1             1

   DL001_2_Z_AV  DL001_3_Y_AV  DL001_4_V_AV  UA110_1_C_AV  UA110_2_Z_AV  \
0             1             1             1             1             0
1             1             0             0             0             1
2             0             1             0             0             0
3             1             1             1             1             1
4             1             1             1             1             1

   UA110_3_Y_AV  UA110_4_V_AV
0             1             1
1             1             1
2             1             1
3             1             1
4             1             1
```



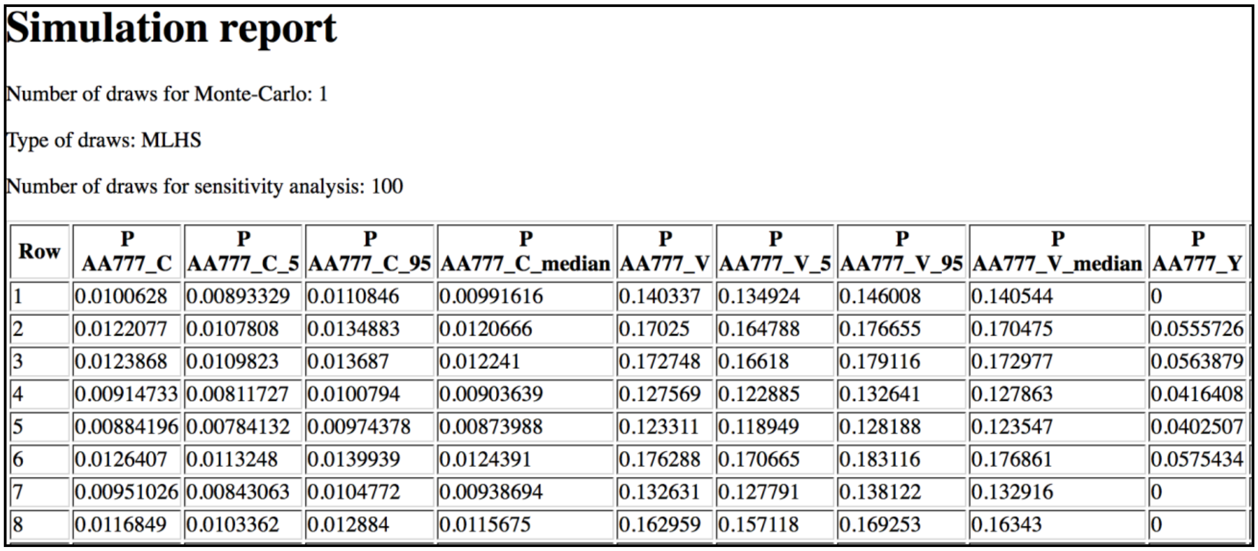
Estimation report

```
      Number of estimated parameters: 6
                         Sample size: 10000
              Excluded observations: 0
                Init log likelihood: -25531.498
               Final log likelihood: -21614.578
Likelihood ratio test for the init. model: 7833.839
      Rho-square for the init. model: 0.153
  Rho-square-bar for the init. model: 0.153
                Final gradient norm: +2.633e-03
                         Diagnostic: Convergence reached...
                         Iterations: 7
                           Run time: 00:01
                     Nbr of threads: 8
```

Estimated parameters

Click on the headers of the columns to sort the table [Credits]

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
B_comp	3.53	1.30	2.70	0.01	1.31	2.70	0.01
B_refund	-0.719	0.137	-5.24	0.00	0.137	-5.23	0.00
C_price	-7.30	1.33	-5.50	0.00	1.33	-5.49	0.00
V_price	-5.07	0.648	-7.83	0.00	0.647	-7.84	0.00
Y_price	-4.41	0.708	-6.23	0.00	0.706	-6.24	0.00
Z_price	-8.71	1.65	-5.27	0.00	1.66	-5.25	0.00



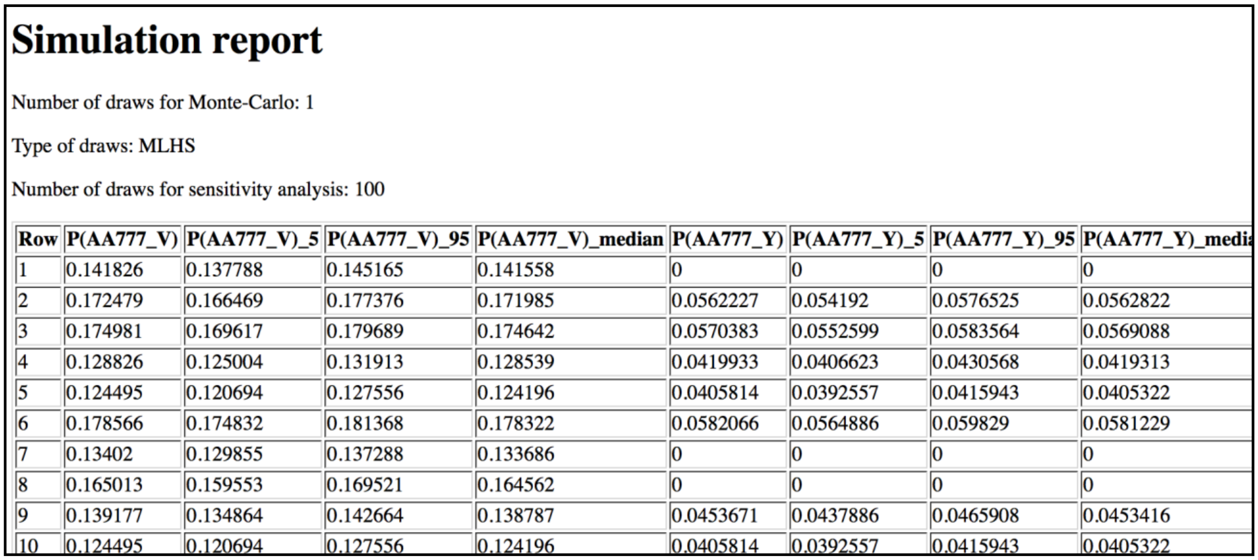
Simulation report

Number of draws for Monte-Carlo: 1

Type of draws: MLHS

Number of draws for sensitivity analysis: 100

Row	P AA777_C	P AA777_C_5	P AA777_C_95	P AA777_C_median	P AA777_V	P AA777_V_5	P AA777_V_95	P AA777_V_median	P AA777_Y
1	0.0100628	0.00893329	0.0110846	0.00991616	0.140337	0.134924	0.146008	0.140544	0
2	0.0122077	0.0107808	0.0134883	0.0120666	0.17025	0.164788	0.176655	0.170475	0.0555726
3	0.0123868	0.0109823	0.013687	0.012241	0.172748	0.16618	0.179116	0.172977	0.0563879
4	0.00914733	0.00811727	0.0100794	0.00903639	0.127569	0.122885	0.132641	0.127863	0.0416408
5	0.00884196	0.00784132	0.00974378	0.00873988	0.123311	0.118949	0.128188	0.123547	0.0402507
6	0.0126407	0.0113248	0.0139939	0.0124391	0.176288	0.170665	0.183116	0.176861	0.0575434
7	0.00951026	0.00843063	0.0104772	0.00938694	0.132631	0.127791	0.138122	0.132916	0
8	0.0116849	0.0103362	0.012884	0.0115675	0.162959	0.157118	0.169253	0.16343	0



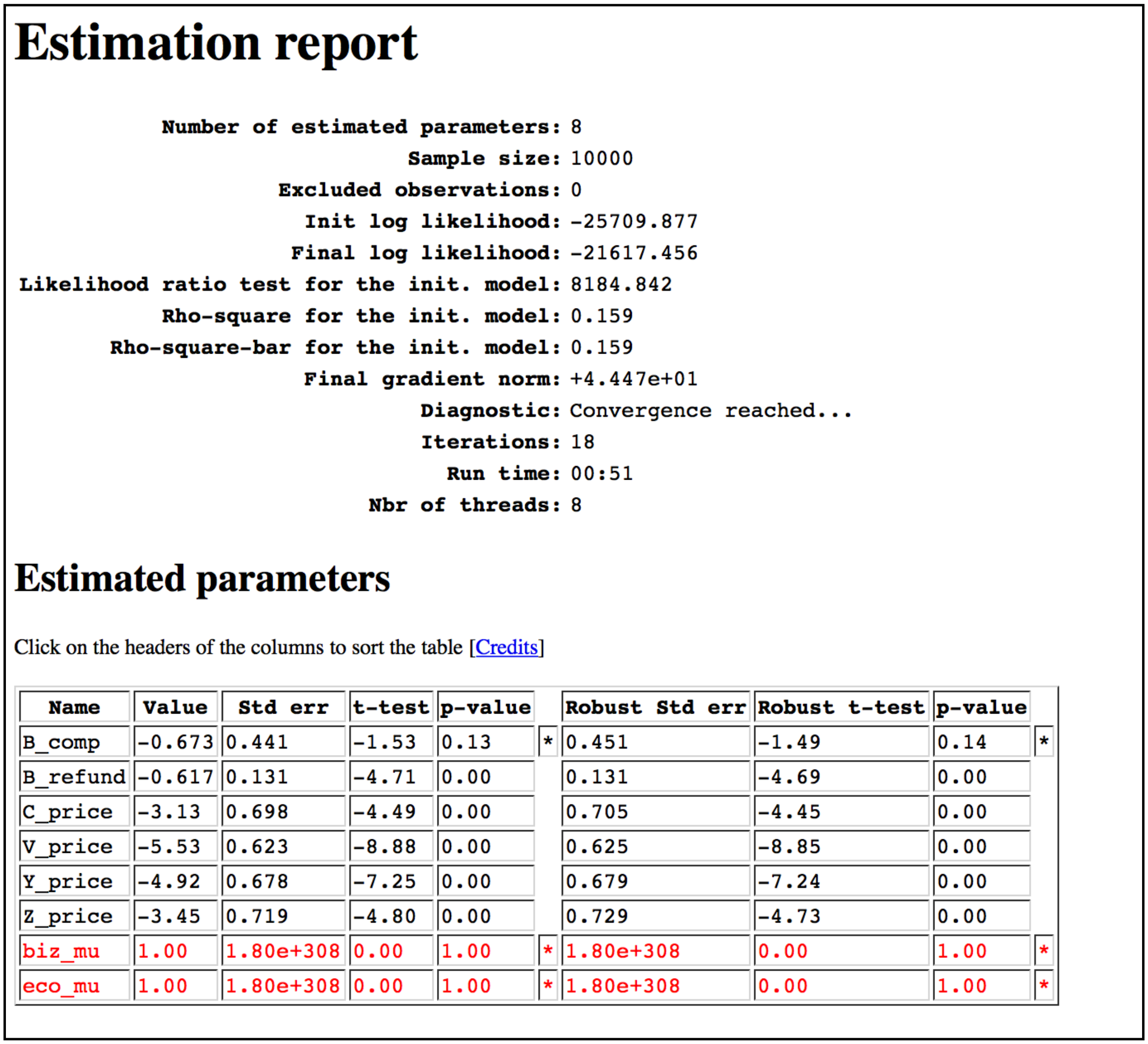
Simulation report

Number of draws for Monte-Carlo: 1

Type of draws: MLHS

Number of draws for sensitivity analysis: 100

Row	P(AA777_V)	P(AA777_V)_5	P(AA777_V)_95	P(AA777_V)_median	P(AA777_Y)	P(AA777_Y)_5	P(AA777_Y)_95	P(AA777_Y)_medi
1	0.141826	0.137788	0.145165	0.141558	0	0	0	0
2	0.172479	0.166469	0.177376	0.171985	0.0562227	0.054192	0.0576525	0.0562822
3	0.174981	0.169617	0.179689	0.174642	0.0570383	0.0552599	0.0583564	0.0569088
4	0.128826	0.125004	0.131913	0.128539	0.0419933	0.0406623	0.0430568	0.0419313
5	0.124495	0.120694	0.127556	0.124196	0.0405814	0.0392557	0.0415943	0.0405322
6	0.178566	0.174832	0.181368	0.178322	0.0582066	0.0564886	0.059829	0.0581229
7	0.13402	0.129855	0.137288	0.133686	0	0	0	0
8	0.165013	0.159553	0.169521	0.164562	0	0	0	0
9	0.139177	0.134864	0.142664	0.138787	0.0453671	0.0437886	0.0465908	0.0453416
10	0.124495	0.120694	0.127556	0.124196	0.0405814	0.0392557	0.0415943	0.0405322



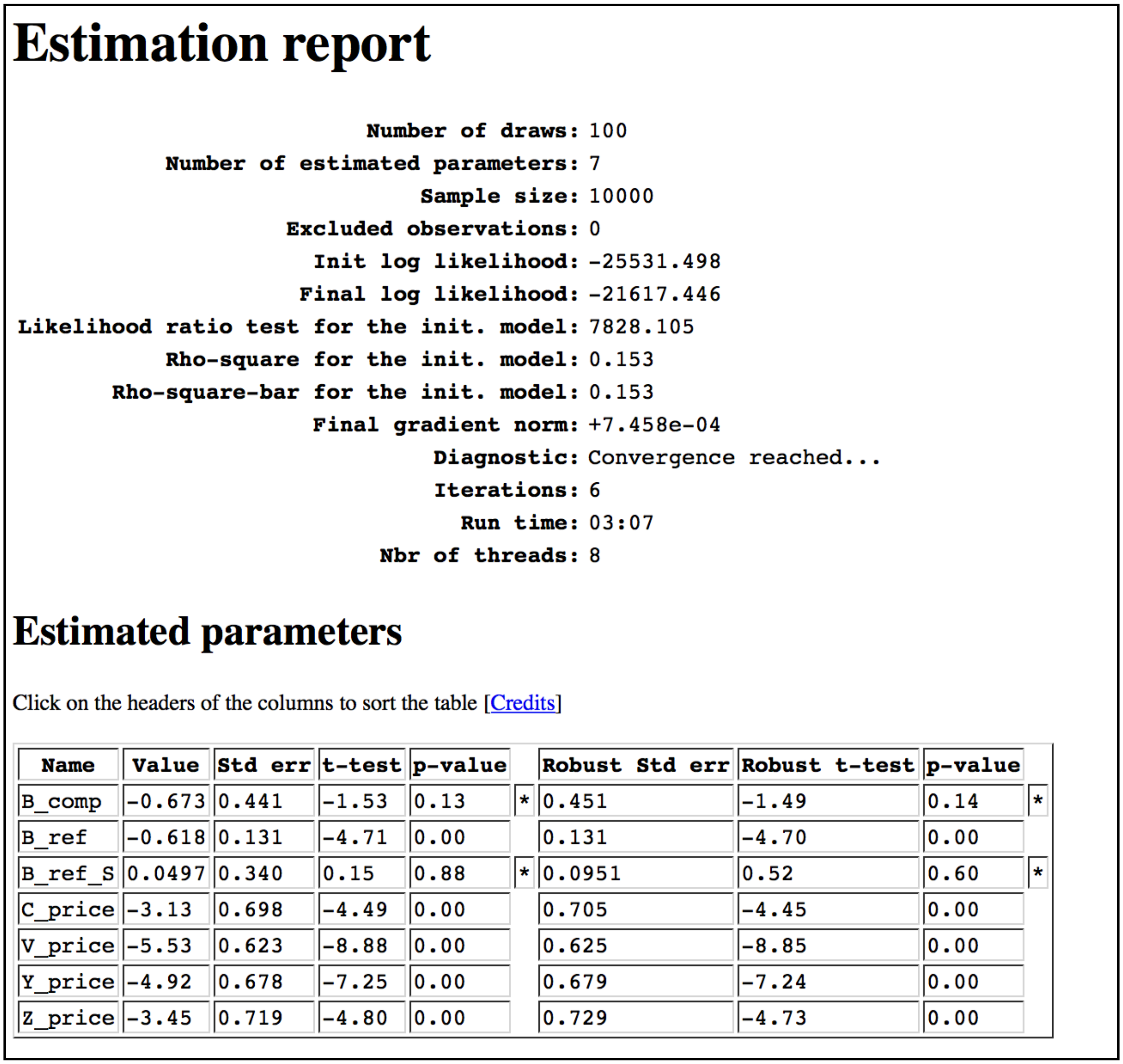
Estimation report

```
            Number of estimated parameters: 8
                               Sample size: 10000
                     Excluded observations: 0
                       Init log likelihood: -25709.877
                      Final log likelihood: -21617.456
Likelihood ratio test for the init. model: 8184.842
            Rho-square for the init. model: 0.159
        Rho-square-bar for the init. model: 0.159
                       Final gradient norm: +4.447e+01
                                Diagnostic: Convergence reached...
                                Iterations: 18
                                  Run time: 00:51
                            Nbr of threads: 8
```

Estimated parameters

Click on the headers of the columns to sort the table [Credits]

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_comp	-0.673	0.441	-1.53	0.13	*	0.451	-1.49	0.14	*
B_refund	-0.617	0.131	-4.71	0.00		0.131	-4.69	0.00	
C_price	-3.13	0.698	-4.49	0.00		0.705	-4.45	0.00	
V_price	-5.53	0.623	-8.88	0.00		0.625	-8.85	0.00	
Y_price	-4.92	0.678	-7.25	0.00		0.679	-7.24	0.00	
Z_price	-3.45	0.719	-4.80	0.00		0.729	-4.73	0.00	
biz_mu	1.00	1.80e+308	0.00	1.00	*	1.80e+308	0.00	1.00	*
eco_mu	1.00	1.80e+308	0.00	1.00	*	1.80e+308	0.00	1.00	*



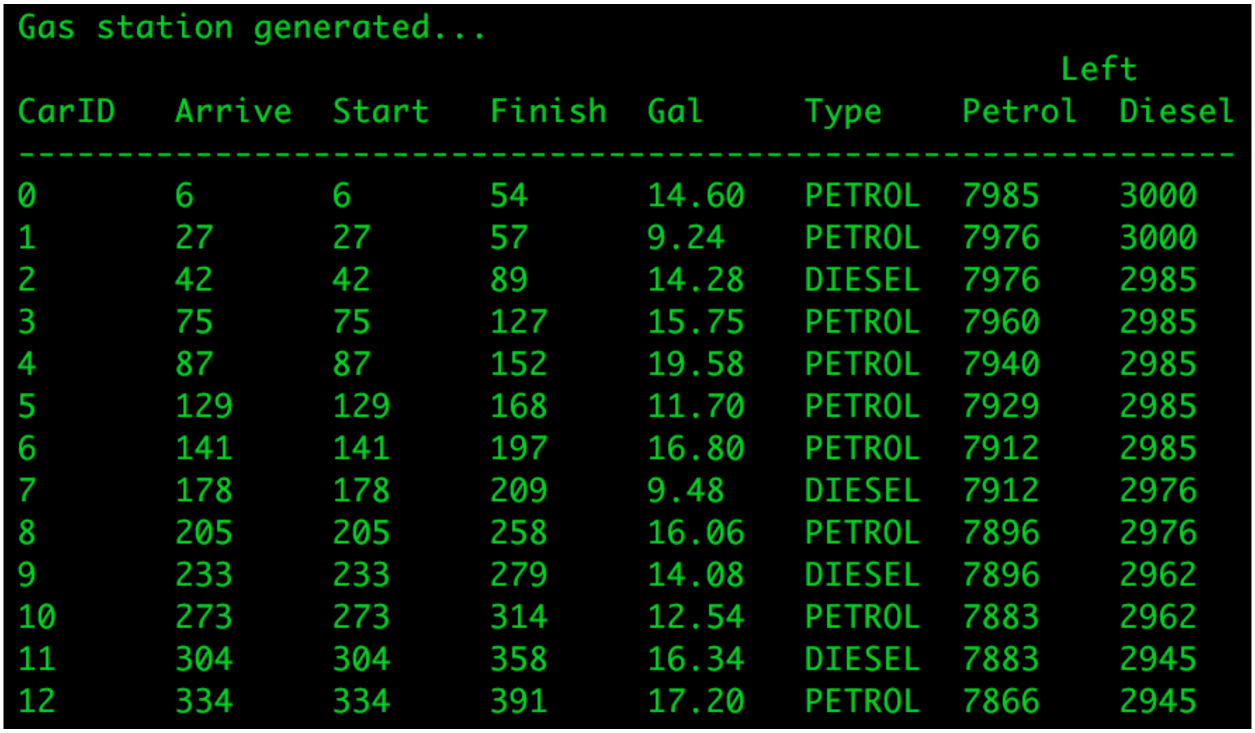
Estimation report

```
                    Number of draws: 100
       Number of estimated parameters: 7
                         Sample size: 10000
                Excluded observations: 0
                  Init log likelihood: -25531.498
                 Final log likelihood: -21617.446
Likelihood ratio test for the init. model: 7828.105
       Rho-square for the init. model: 0.153
   Rho-square-bar for the init. model: 0.153
                  Final gradient norm: +7.458e-04
                           Diagnostic: Convergence reached...
                           Iterations: 6
                             Run time: 03:07
                       Nbr of threads: 8
```

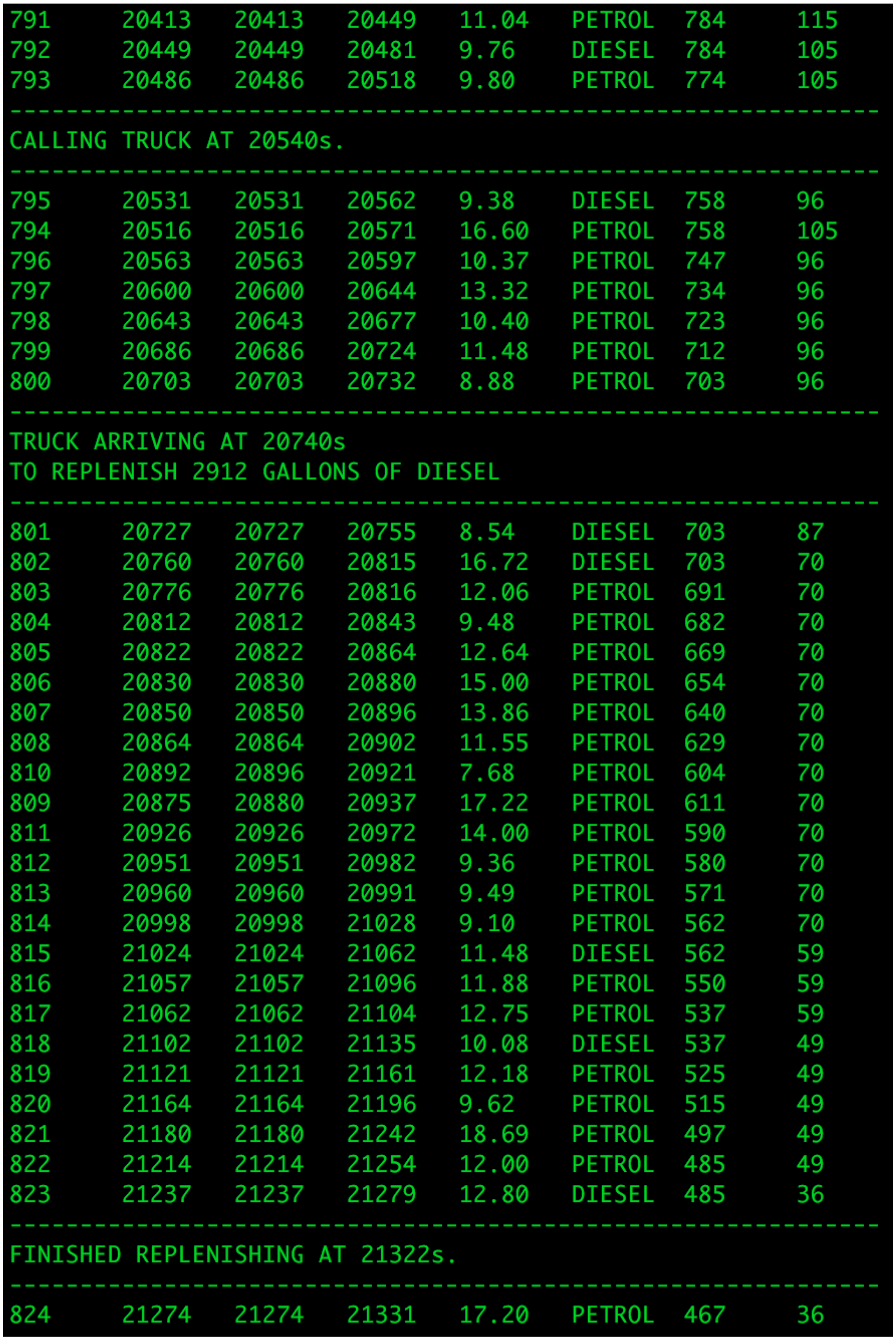
Estimated parameters

Click on the headers of the columns to sort the table [Credits]

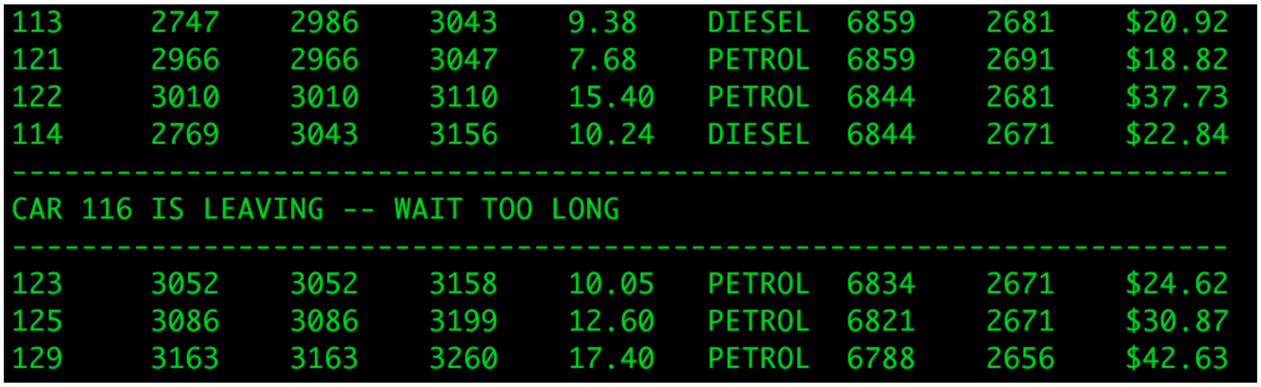
Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_comp	-0.673	0.441	-1.53	0.13	*	0.451	-1.49	0.14	*
B_ref	-0.618	0.131	-4.71	0.00		0.131	-4.70	0.00	
B_ref_S	0.0497	0.340	0.15	0.88	*	0.0951	0.52	0.60	*
C_price	-3.13	0.698	-4.49	0.00		0.705	-4.45	0.00	
V_price	-5.53	0.623	-8.88	0.00		0.625	-8.85	0.00	
Y_price	-4.92	0.678	-7.25	0.00		0.679	-7.24	0.00	
Z_price	-3.45	0.719	-4.80	0.00		0.729	-4.73	0.00	



```
Gas station generated...
                                                     Left
CarID   Arrive  Start   Finish  Gal     Type    Petrol  Diesel
--------------------------------------------------------------
0       6       6       54      14.60   PETROL  7985    3000
1       27      27      57      9.24    PETROL  7976    3000
2       42      42      89      14.28   DIESEL  7976    2985
3       75      75      127     15.75   PETROL  7960    2985
4       87      87      152     19.58   PETROL  7940    2985
5       129     129     168     11.70   PETROL  7929    2985
6       141     141     197     16.80   PETROL  7912    2985
7       178     178     209     9.48    DIESEL  7912    2976
8       205     205     258     16.06   PETROL  7896    2976
9       233     233     279     14.08   DIESEL  7896    2962
10      273     273     314     12.54   PETROL  7883    2962
11      304     304     358     16.34   DIESEL  7883    2945
12      334     334     391     17.20   PETROL  7866    2945
```



```
791     20413   20413   20449   11.04   PETROL  784     115
792     20449   20449   20481   9.76    DIESEL  784     105
793     20486   20486   20518   9.80    PETROL  774     105
---------------------------------------------------------------
CALLING TRUCK AT 20540s.
---------------------------------------------------------------
795     20531   20531   20562   9.38    DIESEL  758     96
794     20516   20516   20571   16.60   PETROL  758     105
796     20563   20563   20597   10.37   PETROL  747     96
797     20600   20600   20644   13.32   PETROL  734     96
798     20643   20643   20677   10.40   PETROL  723     96
799     20686   20686   20724   11.48   PETROL  712     96
800     20703   20703   20732   8.88    PETROL  703     96
---------------------------------------------------------------
TRUCK ARRIVING AT 20740s
TO REPLENISH 2912 GALLONS OF DIESEL
---------------------------------------------------------------
801     20727   20727   20755   8.54    DIESEL  703     87
802     20760   20760   20815   16.72   DIESEL  703     70
803     20776   20776   20816   12.06   PETROL  691     70
804     20812   20812   20843   9.48    PETROL  682     70
805     20822   20822   20864   12.64   PETROL  669     70
806     20830   20830   20880   15.00   PETROL  654     70
807     20850   20850   20896   13.86   PETROL  640     70
808     20864   20864   20902   11.55   PETROL  629     70
810     20892   20896   20921   7.68    PETROL  604     70
809     20875   20880   20937   17.22   PETROL  611     70
811     20926   20926   20972   14.00   PETROL  590     70
812     20951   20951   20982   9.36    PETROL  580     70
813     20960   20960   20991   9.49    PETROL  571     70
814     20998   20998   21028   9.10    PETROL  562     70
815     21024   21024   21062   11.48   DIESEL  562     59
816     21057   21057   21096   11.88   PETROL  550     59
817     21062   21062   21104   12.75   PETROL  537     59
818     21102   21102   21135   10.08   DIESEL  537     49
819     21121   21121   21161   12.18   PETROL  525     49
820     21164   21164   21196   9.62    PETROL  515     49
821     21180   21180   21242   18.69   PETROL  497     49
822     21214   21214   21254   12.00   PETROL  485     49
823     21237   21237   21279   12.80   DIESEL  485     36
---------------------------------------------------------------
FINISHED REPLENISHING AT 21322s.
---------------------------------------------------------------
824     21274   21274   21331   17.20   PETROL  467     36
```



```
113     2747    2986    3043    9.38    DIESEL  6859    2681    $20.92
121     2966    2966    3047    7.68    PETROL  6859    2691    $18.82
122     3010    3010    3110    15.40   PETROL  6844    2681    $37.73
114     2769    3043    3156    10.24   DIESEL  6844    2671    $22.84
----------------------------------------------------------------------
CAR 116 IS LEAVING -- WAIT TOO LONG
----------------------------------------------------------------------
123     3052    3052    3158    10.05   PETROL  6834    2671    $24.62
125     3086    3086    3199    12.60   PETROL  6821    2671    $30.87
129     3163    3163    3260    17.40   PETROL  6788    2656    $42.63
```