Selection_de_variables

November 19, 2019

[1]:
```
library(leaps)
library(glmnet)
```

```
Loading required package: Matrix

Loaded glmnet 3.0
```

0.1 IMPORTATION DES DONNEES

[2]:
```
data_tot = read.csv('./Donnees/Plusieurs_pays/Total.csv', header = TRUE)
data_tot = data_tot[-c(47,60)]
```

Création d'un nouveau dataframe sans les variables `Rating` et `Pays`, utilisé après :

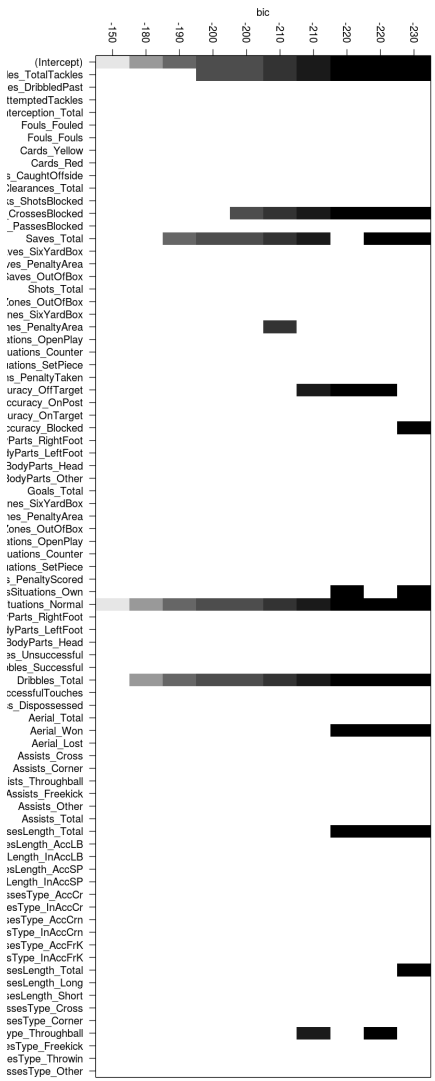
[3]:
```
data = data_tot[,-c(81,82)]
```

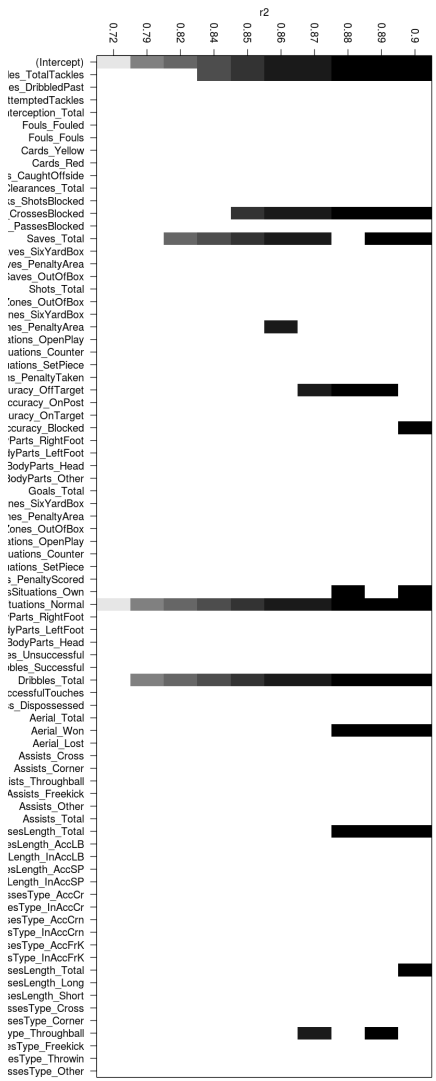
0.2 Sélection de variables

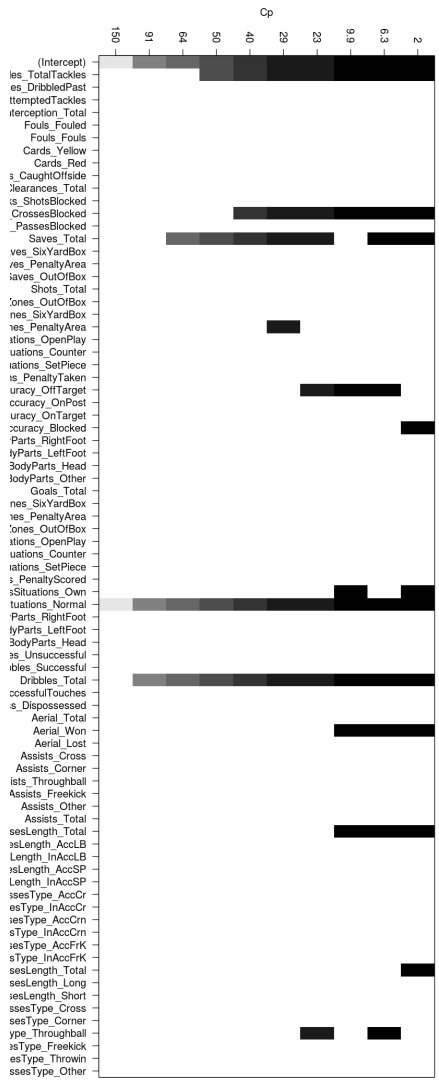
0.2.1 BIC

[39]:
```
choixb <- regsubsets(data_tot$Rating~., data=data,nbest=1, nvmax=10,␣
 ↪method="seqrep")
```

[5]:
```
options(repr.plot.width=18, repr.plot.height=8)
plot(choixb,scale="Cp")
plot(choixb,scale="r2")
plot(choixb,scale="bic")
```


bic
-150
-180
-190
-200
-200
-210
-210
-220
-220
-230
(Intercept)
les_TotalTackles
es_DribbledPast
ttemptedTackles
nterception_Total
Fouls_Fouled
Fouls_Fouls
Cards_Yellow
Cards_Red
s_CaughtOffside
Clearances_Total
is_ShotsBlocked
CrossesBlocked
_PassesBlocked
Saves_Total
ves_SixYardBox
/es_PenaltyArea
Saves_OutOfBox
Shots_Total
'ones_OutOfBox
nes_SixYardBox
nes_PenaltyArea
ations_OpenPlay
uations_Counter
uations_SetPiece
is_PenaltyTaken
uracy_OffTarget
ccuracy_OnPost
uracy_OnTarget
ccuracy_Blocked
Parts_RightFoot
lyParts_LeftFoot
BodyParts_Head
BodyParts_Other
Goals_Total
nes_SixYardBox
nes_PenaltyArea
'ones_OutOfBox
ations_OpenPlay
uations_Counter
uations_SetPiece
s_PenaltyScored
sSituations_Own
tuations_Normal
Parts_RightFoot
lyParts_LeftFoot
BodyParts_Head
es_Unsuccessful
ibles_Successful
Dribbles_Total
ccessfulTouches
s_Dispossessed
Aerial_Total
Aerial_Won
Aerial_Lost
Assists_Cross
Assists_Corner
ists_Throughball
Assists_Freekick
Assists_Other
Assists_Total
sesLength_Total
esLength_AccLB
Length_InAccLB
esLength_AccSP
Length_InAccSP
ssesType_AccCr
esType_InAccCr
esType_AccCrn
sType_InAccCrn
esType_AccFrK
sType_InAccFrK
sesLength_Total
sesLength_Long
esLength_Short
ssesType_Cross
sesType_Corner
ype_Throughball
esType_Freekick
esType_Throwin
ssesType_Other


r2
0.72
0.79
0.82
0.84
0.85
0.86
0.87
0.88
0.89
0.9
(Intercept)
les_TotalTackles
es_DribbledPast
ttemptedTackles
nterception_Total
Fouls_Fouled
Fouls_Fouls
Cards_Yellow
Cards_Red
s_CaughtOffside
Clearances_Total
is_ShotsBlocked
CrossesBlocked
_PassesBlocked
Saves_Total
ves_SixYardBox
/es_PenaltyArea
Saves_OutOfBox
Shots_Total
'ones_OutOfBox
nes_SixYardBox
nes_PenaltyArea
ations_OpenPlay
uations_Counter
uations_SetPiece
is_PenaltyTaken
uracy_OffTarget
ccuracy_OnPost
uracy_OnTarget
ccuracy_Blocked
Parts_RightFoot
lyParts_LeftFoot
BodyParts_Head
BodyParts_Other
Goals_Total
nes_SixYardBox
nes_PenaltyArea
'ones_OutOfBox
ations_OpenPlay
uations_Counter
uations_SetPiece
s_PenaltyScored
sSituations_Own
tuations_Normal
Parts_RightFoot
lyParts_LeftFoot
BodyParts_Head
es_Unsuccessful
ibles_Successful
Dribbles_Total
ccessfulTouches
s_Dispossessed
Aerial_Total
Aerial_Won
Aerial_Lost
Assists_Cross
Assists_Corner
ists_Throughball
Assists_Freekick
Assists_Other
Assists_Total
sesLength_Total
esLength_AccLB
Length_InAccLB
esLength_AccSP
Length_InAccSP
ssesType_AccCr
esType_InAccCr
esType_AccCrn
sType_InAccCrn
esType_AccFrK
sType_InAccFrK
sesLength_Total
sesLength_Long
esLength_Short
ssesType_Cross
sesType_Corner
ype_Throughball
esType_Freekick
esType_Throwin
ssesType_Other


Cp
150
91
64
50
40
29
23
9.9
6.3
2
(Intercept)
les_TotalTackles
es_DribbledPast
ttemptedTackles
nterception_Total
Fouls_Fouled
Fouls_Fouls
Cards_Yellow
Cards_Red
s_CaughtOffside
Clearances_Total
is_ShotsBlocked
CrossesBlocked
_PassesBlocked
Saves_Total
ves_SixYardBox
/es_PenaltyArea
Saves_OutOfBox
Shots_Total
'ones_OutOfBox
nes_SixYardBox
nes_PenaltyArea
ations_OpenPlay
uations_Counter
uations_SetPiece
is_PenaltyTaken
uracy_OffTarget
ccuracy_OnPost
uracy_OnTarget
ccuracy_Blocked
Parts_RightFoot
lyParts_LeftFoot
BodyParts_Head
BodyParts_Other
Goals_Total
nes_SixYardBox
nes_PenaltyArea
'ones_OutOfBox
ations_OpenPlay
uations_Counter
uations_SetPiece
s_PenaltyScored
sSituations_Own
tuations_Normal
Parts_RightFoot
lyParts_LeftFoot
BodyParts_Head
es_Unsuccessful
ibles_Successful
Dribbles_Total
ccessfulTouches
s_Dispossessed
Aerial_Total
Aerial_Won
Aerial_Lost
Assists_Cross
Assists_Corner
ists_Throughball
Assists_Freekick
Assists_Other
Assists_Total
sesLength_Total
esLength_AccLB
Length_InAccLB
esLength_AccSP
Length_InAccSP
ssesType_AccCr
esType_InAccCr
esType_AccCrn
sType_InAccCrn
esType_AccFrK
sType_InAccFrK
sesLength_Total
sesLength_Long
esLength_Short
ssesType_Cross
sesType_Corner
ype_Throughball
esType_Freekick
esType_Throwin
ssesType_Other

Les trois différents critères utilisés ci-dessus pour la sélection de modèle (Cp de Mallows, R2 et BIC) semblent donner les mêmes résultats.
Suivant le critère choisi, il faut soit le maximiser (R2), soit le minimiser (Cp et BIC). Dans les deux cas, il s'agit de trouver les variables mises en noir sur la ligne du haut.

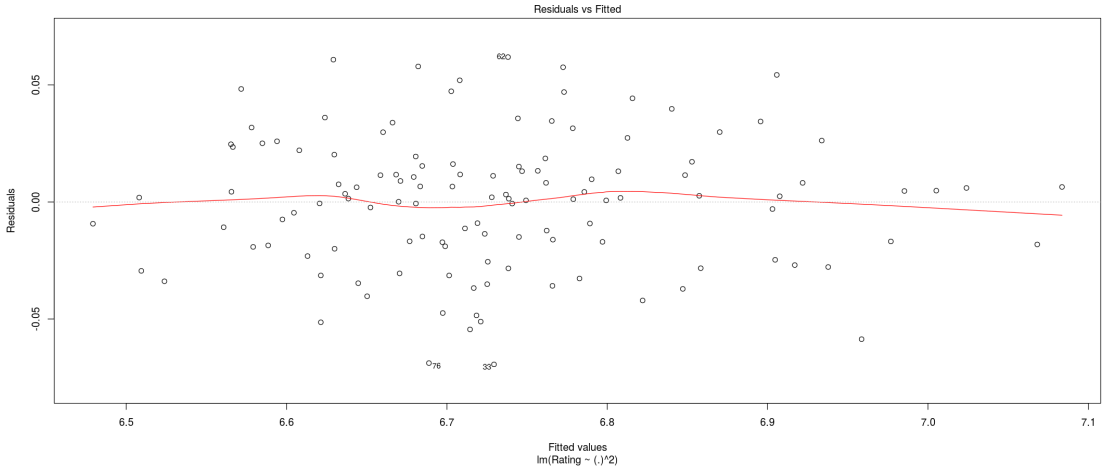
[6]:
```
nb_min = which.min(summary(choixb)$bic)
coef(choixb, nb_min)
```

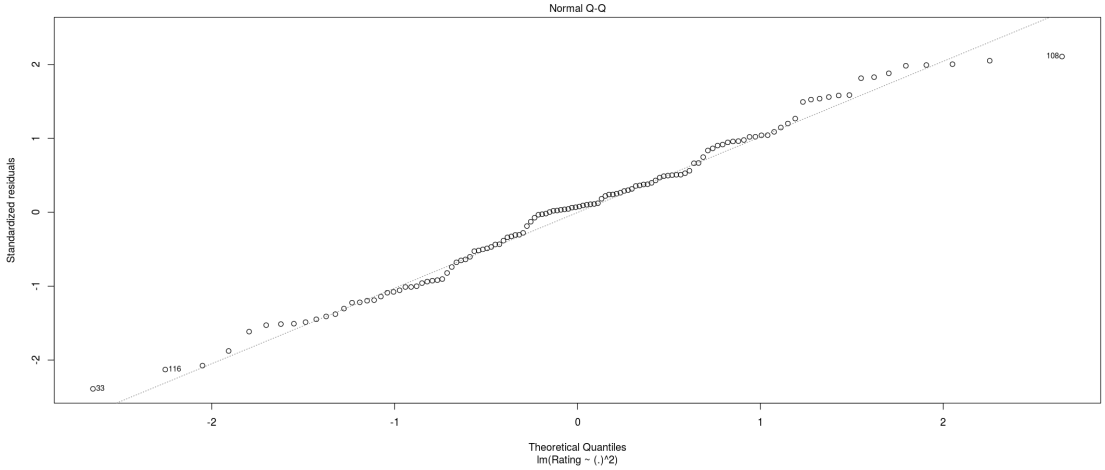
(Intercept) 5.75215912848794 **Tackles_TotalTackles** 0.0116722138782329
Blocks_CrossesBlocked 0.0446894205980564 **Saves_Total** -0.0197312817417468
ShotsAccuracy_Blocked -0.0298931882729577 **GoalsSituations_Own** 0.221852796031417
GoalsSituations_Normal 0.194642946229757 **Dribbles_Total** 0.00852515056606557
Aerial_Won 0.00705457233332821 **PassesLength_Total** 0.000369394176703441
KeyPassesLength_Total 0.0189994394720078

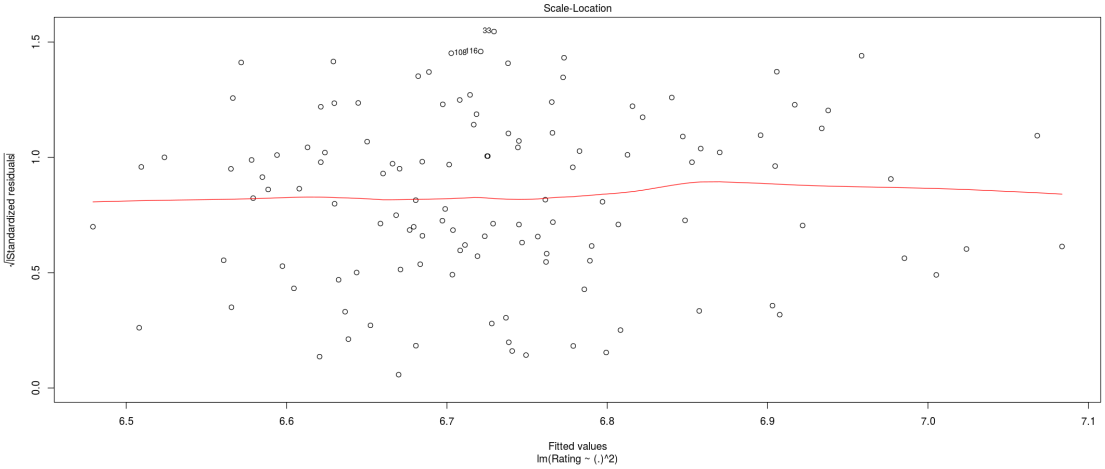
[7]:
```
data_bic = data.frame(data_tot$Rating, data_tot$Tackles_TotalTackles,␣
↪data_tot$Blocks_CrossesBlocked, data_tot$Saves_Total,
                      data_tot$ShotsAccuracy_Blocked,␣
↪data_tot$GoalsSituations_Own, data_tot$GoalsSituations_Normal,
                      data_tot$Dribbles_Total, data_tot$Aerial_Won,␣
↪data_tot$PassesLength_Total, data_tot$KeyPassesLength_Total)

names(data_bic) <- c("Rating", "Tackles_TotalTackles", "Blocks_CrossesBlocked",␣
↪"Saves_Total", "ShotsAccuracy_Blocked",
                     "GoalsSituations_Own", "GoalsSituations_Normal",␣
↪"Dribbles_Total", "Aerial_Won",
                     "PassesLength_Total", "KeyPassesLength_Total")
```

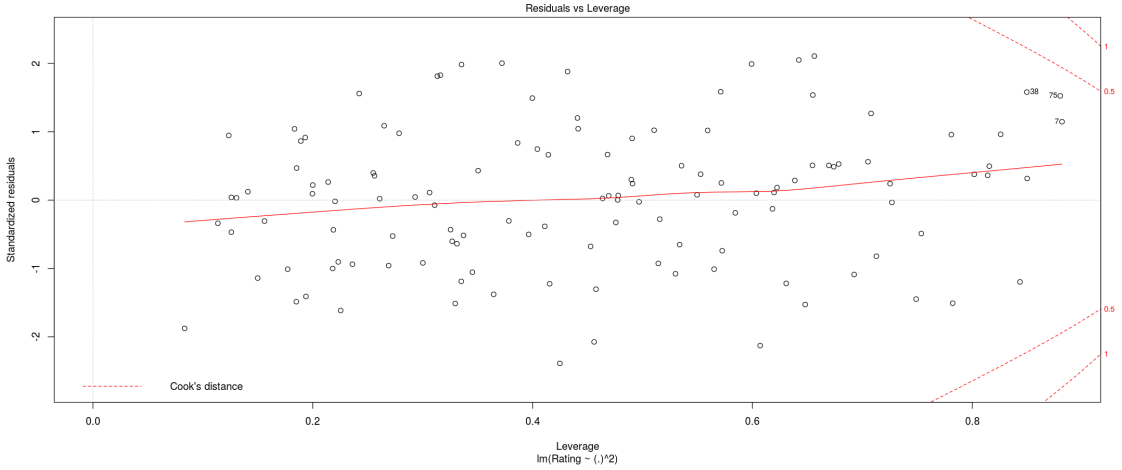
[8]:
```
reg_bic = lm(Rating~(.)^2, data = data_bic)
```

[9]:
```
plot(reg_bic)
```




Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Rating ~ (.)^2)
108
116
33


Scale-Location
√|Standardized residuals|
Fitted values
lm(Rating ~ (.)^2)
33
108
116


Residuals vs Leverage
Standardized residuals
Leverage
lm(Rating ~ (.)^2)
Cook's distance
38
75
7

On voit que le graphe des résidus ne présente pas de forme particulière. De plus, le graphe quantile-quantile est plus ou moins aligné (quelques soucis sur les petits et grands quantiles).

On va refaire de la sélection de variables, mais pour chaque pays, et ainsi observer les variables vraiment influentes.

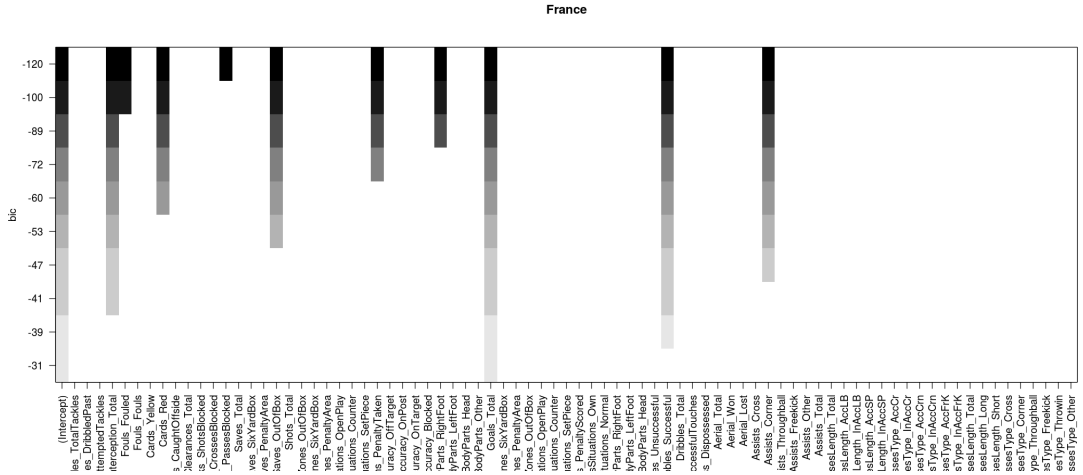
[10]:
```
France = data[data_tot$Pays == "France",]
Allemagne = data[data_tot$Pays == "Allemagne",]
Italie = data[data_tot$Pays == "Italie",]
Espagne = data[data_tot$Pays == "Espagne",]
Argentine = data[data_tot$Pays == "Argentine",]
Angleterre = data[data_tot$Pays == "Angleterre",]
```

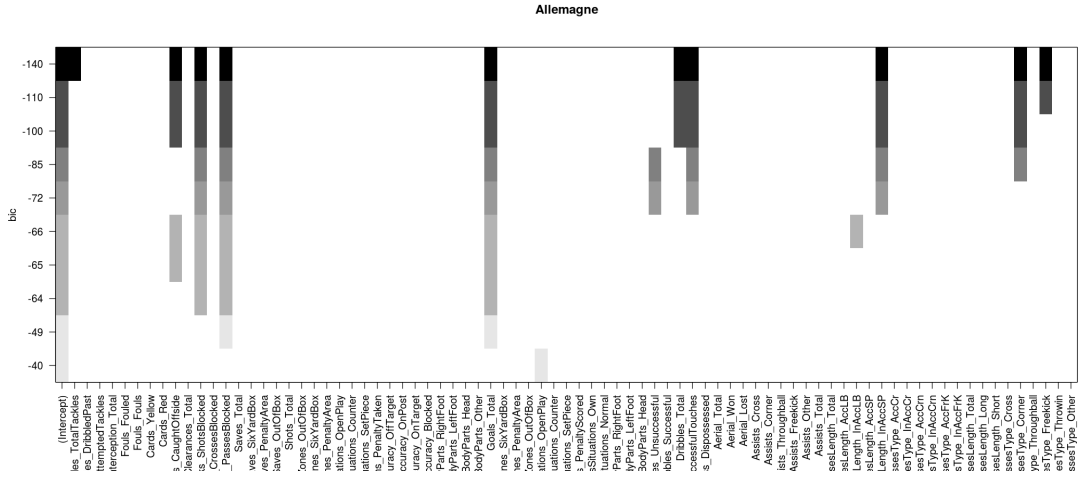
[11]:
```
choix_France <- regsubsets(data_tot[data_tot$Pays=="France",]$Rating~.,␣
↪data=France,nbest=1, nvmax=10, method="seqrep")
choix_Allemagne <- regsubsets(data_tot[data_tot$Pays=="Allemagne",]$Rating~.,␣
↪data=Allemagne,nbest=1, nvmax=10, method="seqrep")
choix_Italie <- regsubsets(data_tot[data_tot$Pays=="Italie",]$Rating~.,␣
↪data=Italie,nbest=1, nvmax=10, method="seqrep")
choix_Espagne <- regsubsets(data_tot[data_tot$Pays=="Espagne",]$Rating~.,␣
↪data=Espagne,nbest=1, nvmax=10, method="seqrep")
choix_Argentine <- regsubsets(data_tot[data_tot$Pays=="Argentine",]$Rating~.,␣
↪data=Argentine, nbest=1, nvmax=10, method="seqrep")
choix_Angleterre <- regsubsets(data_tot[data_tot$Pays=="Angleterre",]$Rating~.,␣
↪data=Angleterre, nbest=1, nvmax=10, method="seqrep")
```

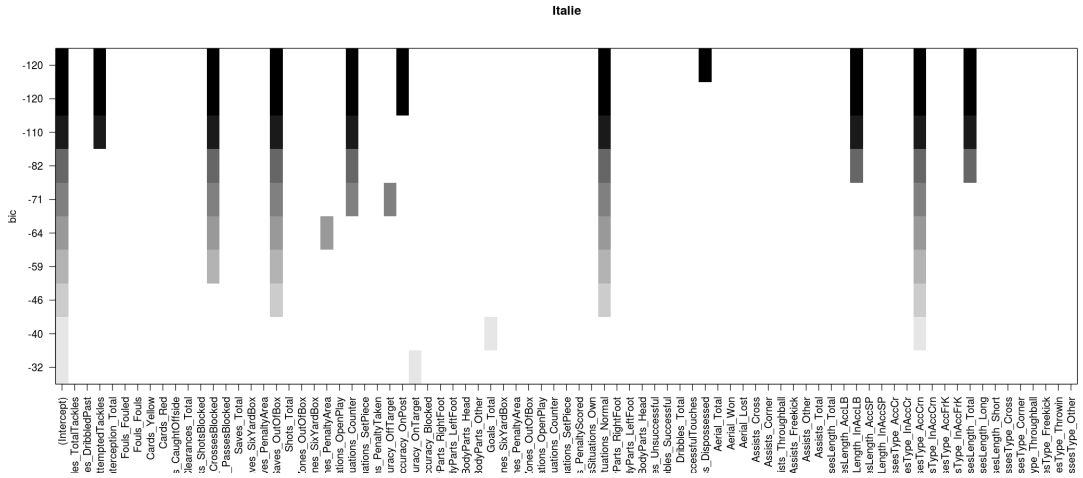
```
Warning message in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
force.in = force.in, :
61  linear dependencies found
Warning message in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
force.in = force.in, :
63  linear dependencies found
Warning message in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
force.in = force.in, :
61  linear dependencies found
Warning message in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
force.in = force.in, :
61  linear dependencies found
Warning message in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
force.in = force.in, :
55  linear dependencies found
Warning message in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
force.in = force.in, :
61  linear dependencies found
```

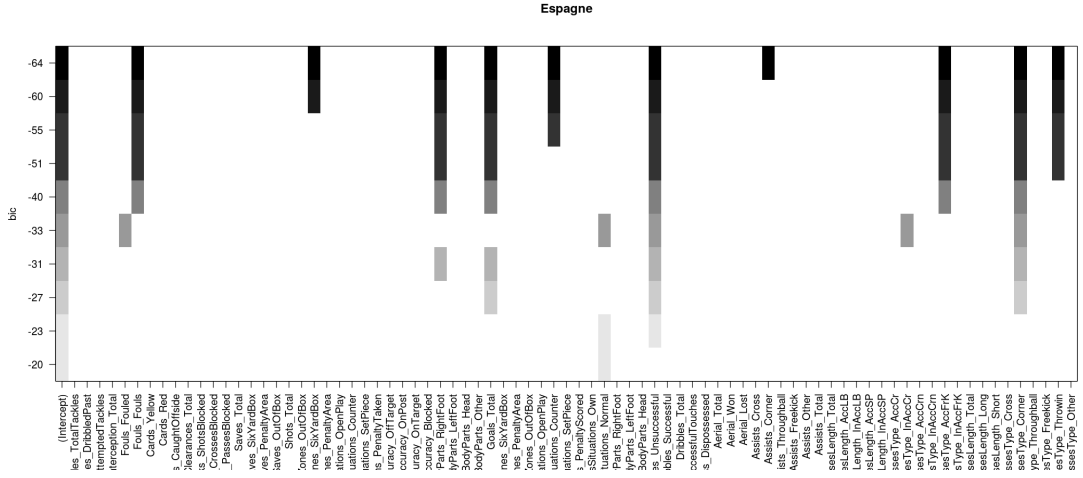
[12]:
```
plot(choix_France,scale="bic", main = "France")
plot(choix_Allemagne,scale="bic", main = "Allemagne")
```

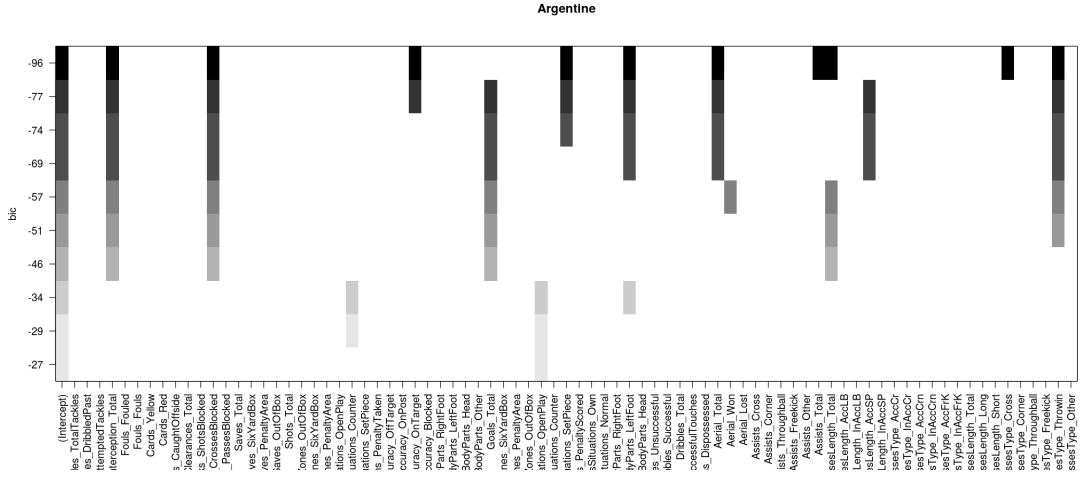
```
plot(choix_Italie,scale="bic", main = "Italie")
plot(choix_Espagne,scale="bic", main = "Espagne")
plot(choix_Argentine,scale="bic", main = "Argentine")
plot(choix_Angleterre,scale="bic", main = "Angleterre")
```

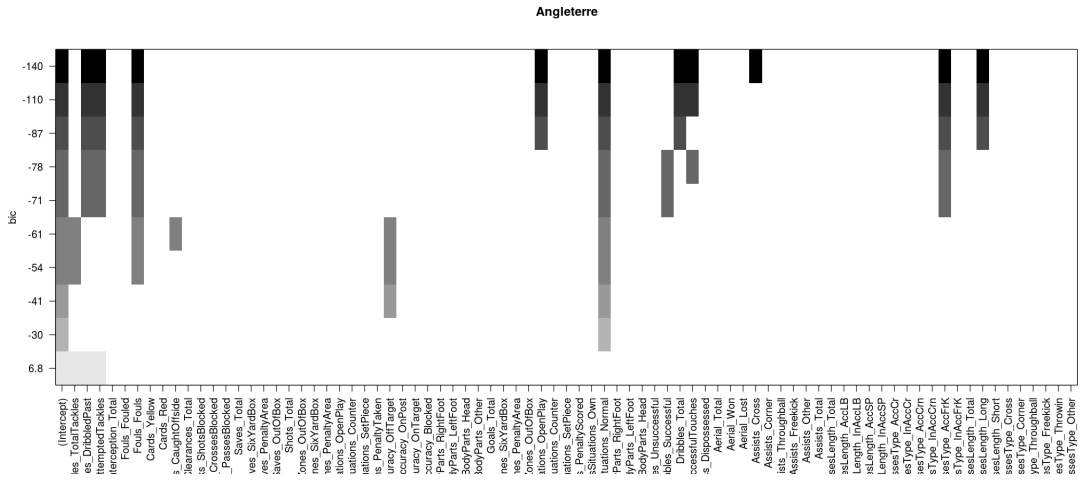





Italie
bic
-120
-120
-110
-82
-71
-64
-59
-46
-40
-32
(Intercept)
Tackles_TotalTackles
Tackles_DribbledPast
Challenges_AttemptedTackles
Challenges_Interception_Total
Fouls_Fouled
Fouls_Fouls
Cards_Yellow
Cards_Red
Challenges_CaughtOffside
Clearances_Total
Shots_ShotsBlocked
Crosses_CrossesBlocked
Passes_PassesBlocked
Saves_Total
Saves_SixYardBox
Saves_PenaltyArea
Saves_OutOfBox
Shots_Total
Zones_OutOfBox
Passes_SixYardBox
Passes_PenaltyArea
Situations_OpenPlay
Situations_Counter
Situations_SetPiece
Accuracy_PenaltyTaken
Accuracy_OffTarget
Accuracy_OnPost
Accuracy_OnTarget
Accuracy_Blocked
BodyParts_RightFoot
BodyParts_LeftFoot
BodyParts_Head
BodyParts_Other
Goals_Total
Passes_SixYardBox
Passes_PenaltyArea
Zones_OutOfBox
Situations_OpenPlay
Situations_Counter
Goals_PenaltyScored
Situations_SetPiece
Situations_Normal
BodyParts_RightFoot
BodyParts_LeftFoot
BodyParts_Head
Passes_Unsuccessful
Tackles_Total
Dribbles_Total
Dribbles_Successful
Touches_SuccessfulTouches
Touches_Dispossessed
Aerial_Total
Aerial_Won
Aerial_Lost
Assists_Cross
Assists_Corner
Assists_Throughball
Assists_Freekick
Assists_Other
Assists_Total
PassesLength_Total
PassesLength_AccLB
PassesLength_InAccLB
PassesLength_AccSP
PassesLength_InAccSP
PassesType_AccCr
PassesType_InAccCr
PassesType_AccCrn
PassesType_InAccCrn
PassesType_AccFrK
PassesType_InAccFrK
PassesLength_Total
PassesLength_Long
PassesLength_Short
PassesType_Cross
PassesType_Corner
PassesType_Throughball
PassesType_Freekick
PassesType_Thrown
PassesType_Other


Espagne
bic
-64
-60
-55
-51
-40
-33
-31
-27
-23
-20
(Intercept)
Tackles_TotalTackles
Tackles_DribbledPast
Challenges_AttemptedTackles
Challenges_Interception_Total
Fouls_Fouled
Fouls_Fouls
Cards_Yellow
Cards_Red
Challenges_CaughtOffside
Clearances_Total
Shots_ShotsBlocked
Crosses_CrossesBlocked
Passes_PassesBlocked
Saves_Total
Saves_SixYardBox
Saves_PenaltyArea
Saves_OutOfBox
Shots_Total
Zones_OutOfBox
Passes_SixYardBox
Passes_PenaltyArea
Situations_OpenPlay
Situations_Counter
Situations_SetPiece
Accuracy_PenaltyTaken
Accuracy_OffTarget
Accuracy_OnPost
Accuracy_OnTarget
Accuracy_Blocked
BodyParts_RightFoot
BodyParts_LeftFoot
BodyParts_Head
BodyParts_Other
Goals_Total
Passes_SixYardBox
Passes_PenaltyArea
Zones_OutOfBox
Situations_OpenPlay
Situations_Counter
Goals_PenaltyScored
Situations_SetPiece
Situations_Normal
BodyParts_RightFoot
BodyParts_LeftFoot
BodyParts_Head
Passes_Unsuccessful
Tackles_Total
Dribbles_Total
Dribbles_Successful
Touches_SuccessfulTouches
Touches_Dispossessed
Aerial_Total
Aerial_Won
Aerial_Lost
Assists_Cross
Assists_Corner
Assists_Throughball
Assists_Freekick
Assists_Other
Assists_Total
PassesLength_Total
PassesLength_AccLB
PassesLength_InAccLB
PassesLength_AccSP
PassesLength_InAccSP
PassesType_AccCr
PassesType_InAccCr
PassesType_AccCrn
PassesType_InAccCrn
PassesType_AccFrK
PassesType_InAccFrK
PassesLength_Total
PassesLength_Long
PassesLength_Short
PassesType_Cross
PassesType_Corner
PassesType_Throughball
PassesType_Freekick
PassesType_Thrown
PassesType_Other


Argentine
bic
-96
-77
-74
-69
-57
-51
-46
-34
-29
-27
(Intercept)
Tackles_TotalTackles
Tackles_DribbledPast
Challenges_AttemptedTackles
Challenges_Interception_Total
Fouls_Fouled
Fouls_Fouls
Cards_Yellow
Cards_Red
Challenges_CaughtOffside
Clearances_Total
Shots_ShotsBlocked
Crosses_CrossesBlocked
Passes_PassesBlocked
Saves_Total
Saves_SixYardBox
Saves_PenaltyArea
Saves_OutOfBox
Shots_Total
Zones_OutOfBox
Passes_SixYardBox
Passes_PenaltyArea
Situations_OpenPlay
Situations_Counter
Situations_SetPiece
Accuracy_PenaltyTaken
Accuracy_OffTarget
Accuracy_OnPost
Accuracy_OnTarget
Accuracy_Blocked
BodyParts_RightFoot
BodyParts_LeftFoot
BodyParts_Head
BodyParts_Other
Goals_Total
Passes_SixYardBox
Passes_PenaltyArea
Zones_OutOfBox
Situations_OpenPlay
Situations_Counter
Goals_PenaltyScored
Situations_SetPiece
Situations_Normal
BodyParts_RightFoot
BodyParts_LeftFoot
BodyParts_Head
Passes_Unsuccessful
Tackles_Total
Dribbles_Total
Dribbles_Successful
Touches_SuccessfulTouches
Touches_Dispossessed
Aerial_Total
Aerial_Won
Aerial_Lost
Assists_Cross
Assists_Corner
Assists_Throughball
Assists_Freekick
Assists_Other
Assists_Total
PassesLength_Total
PassesLength_AccLB
PassesLength_InAccLB
PassesLength_AccSP
PassesLength_InAccSP
PassesType_AccCr
PassesType_InAccCr
PassesType_AccCrn
PassesType_InAccCrn
PassesType_AccFrK
PassesType_InAccFrK
PassesLength_Total
PassesLength_Long
PassesLength_Short
PassesType_Cross
PassesType_Corner
PassesType_Throughball
PassesType_Freekick
PassesType_Thrown
PassesType_Other



On voit que les variables retenues ne sont pas les mêmes d'un pays à l'autre. On va les afficher.

[13]:
```
nb_min = which.min(summary(choix_France)$bic)
coef(choix_France, nb_min)
```

(Intercept) 6.13982138539476 **Interception_Total** 0.0237925361751378 **Fouls_Fouled** 0.00780851950097679 **Cards_Red** -0.599989959343333 **Blocks_PassesBlocked** 0.00953063969054786 **Saves_OutOfBox** -0.158696058583241 **ShotsSituations_PenaltyTaken** -0.299210436279684 **ShotsBodyParts_RightFoot** 0.01058761355573 **Goals_Total** 0.248842297964619 **Dribbles_Successful** 0.0144764285972766 **Assists_Corner** -0.685200840055928

[14]:
```
nb_min = which.min(summary(choix_Allemagne)$bic)
coef(choix_Allemagne, nb_min)
```

(Intercept) 6.3990913593768 **Tackles_TotalTackles** -0.00307838302127169 **OffSides_CaughtOffside** 0.019588686984888 **Blocks_ShotsBlocked** -0.100248995583956 **Blocks_PassesBlocked** 0.0312030351619673 **Goals_Total** 0.254083488003104 **Dribbles_Total** -0.0108021870270855 **PossesionLoss_UnsuccessfulTouches** 0.0329907893164837 **PassesLength_InAccSP** -0.00653498957293113 **KeyPassesType_Corner** 0.0420783100445682 **KeyPassesType_Freekick** 0.0317966473768425

[15]:
```
nb_min = which.min(summary(choix_Italie)$bic)
coef(choix_Italie, nb_min)
```

(Intercept) 6.1219354788306 **Tackles_TotalAttemptedTackles** -0.00631437235721584 **Blocks_CrossesBlocked** 0.0829437516623944 **Saves_OutOfBox** -0.107609112290232 **shotsSituations_Counter** -0.162027230276744 **ShotsAccuracy_OnPost** 0.0470456703888898 **GoalsSituations_Normal** 0.22052772131892 **PossesionLoss_Dispossessed** 0.00214167732977552 **PassesLength_InAccLB** 0.00409971082734128 **PassesType_AccCrn** 0.0741779917647036 **KeyPassesLength_Total** 0.0154845158237651

[16]:
```
nb_min = which.min(summary(choix_Espagne)$bic)
coef(choix_Espagne, nb_min)
```

(Intercept) 5.30226068030442 **Fouls_Fouls** 0.0603328099672517 **ShotsZones_SixYardBox** -0.0494355377391404 **ShotsBodyParts_RightFoot** -0.0951975313204938 **Goals_Total** 0.646177372605525 **GoalsSituations_Counter** 0.311904628461455 **Dribbles_Unsuccessful** -0.0773271313750928 **Assists_Corner** 0.162092345825822 **PassesType_AccFrK** 0.0438940115077385 **KeyPassesType_Corner** 0.445351698903317 **KeyPassesType_Throwin** 0.512586963670849

[17]:
```
nb_min = which.min(summary(choix_Argentine)$bic)
coef(choix_Argentine, nb_min)
```

(Intercept) 5.16725857068163 **Interception_Total** 0.0155505147131027 **Blocks_CrossesBlocked** 0.118954098176166 **ShotsAccuracy_OnTarget** 0.0553495653421166 **GoalsSituations_SetPiece** 0.158874580769146 **GoalsBodyParts_LeftFoot** 0.238693148373298 **Aerial_Total** 0.00628894471029627 **Assists_Total** 0.118270593028633 **PassesLength_Total** 0.000998176423661522 **KeyPassesType_Cross** 0.0280459864635727 **KeyPassesType_Throwin** 0.330646866627707

[18]:
```
nb_min = which.min(summary(choix_Angleterre)$bic)
coef(choix_Angleterre, nb_min)
```

(Intercept) 6.49408039603212 **Tackles_DribbledPast** -0.0586785202081505 **Tackles_TotalAttemptedTackles** 0.0274205807376198 **Fouls_Fouls** -0.0285663008528862 **GoalsSituations_OpenPlay** 0.102413063843859 **GoalsSituations_Normal** 0.235515646057419 **Dribbles_Total** 0.0115469947556096 **PossesionLoss_UnsuccessfulTouches** -0.0103487550216349 **Assists_Cross** 0.0600858505639187 **PassesType_AccFrK** -0.016677853678321 **KeyPassesLength_Long** 0.036631274846773

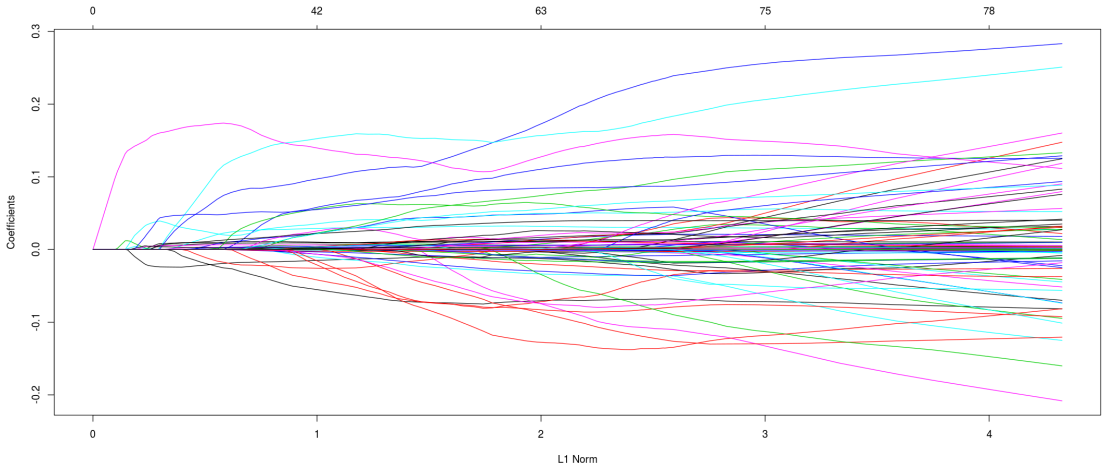
On remarque que c'est seulement en France, Allemagne et Espagne que la variable `Goals_Total` est sélectionnée.

1 Regression LASSO

alpha=1 is the lasso penalty, and alpha=0 the ridge penalty

[19]:
```
m_lasso = glmnet(as.matrix(data), data_tot$Rating, alpha = 1, nlambda = 100)
```

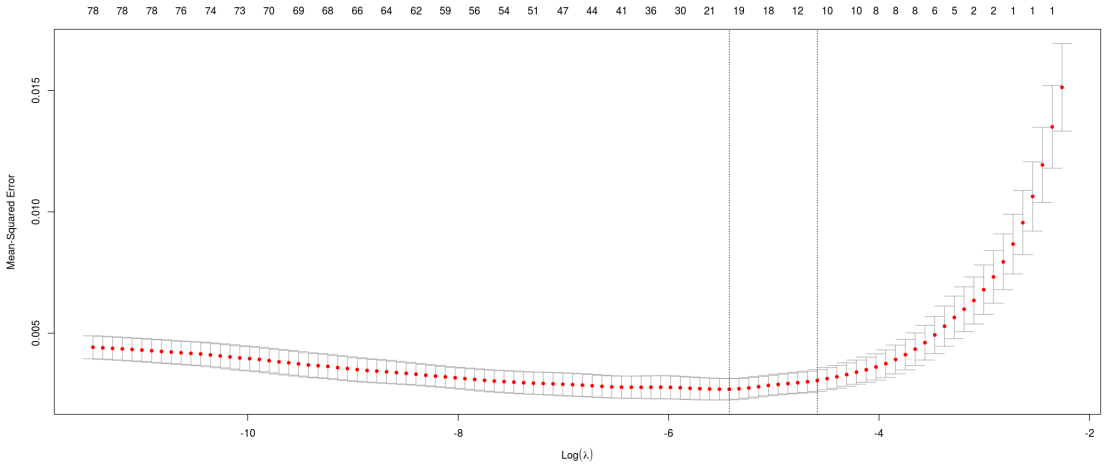
[20]:
```
plot(m_lasso)
```



Cross validation

[22]:
```
cv.out <- cv.glmnet(as.matrix(data), data_tot$Rating, alpha = 1)
```

[23]:
```
plot(cv.out)
```



[24]:
```
bestlam <- cv.out$lambda.min
```

[25]:
```
predict(m_lasso, type = "coefficients", s = bestlam)
```

```
81 x 1 sparse Matrix of class "dgCMatrix"
                                       1
(Intercept)                6.0663284835
Tackles_TotalTackles       0.0078627329
```

Tackles_DribbledPast	.
Tackles_TotalAttemptedTackles	.
Interception_Total	0.0002451712
Fouls_Fouled	.
Fouls_Fouls	.
Cards_Yellow	.
Cards_Red	-0.0192991092
OffSides_CaughtOffside	.
Clearances_Total	.
Blocks_ShotsBlocked	.
Blocks_CrossesBlocked	0.0160727193
Blocks_PassesBlocked	0.0030481778
Saves_Total	-0.0213412187
Saves_SixYardBox	-0.0113095109
Saves_PenaltyArea	.
Saves_OutOfBox	.
Shots_Total	.
ShotsZones_OutOfBox	.
ShotsZones_SixYardBox	.
ShotsZones_PenaltyArea	.
ShotsSituations_OpenPlay	.
shotsSituations_Counter	0.0045381507
ShotsSituations_SetPiece	.
ShotsSituations_PenaltyTaken	.
ShotsAccuracy_OffTarget	0.0095423106
ShotsAccuracy_OnPost	.
ShotsAccuracy_OnTarget	.
ShotsAccuracy_Blocked	.
ShotsBodyParts_RightFoot	.
ShotsBodyParts_LeftFoot	.
ShotsBodyParts_Head	.
ShotsBodyParts_Other	.
Goals_Total	.
GoalsZones_SixYardBox	.
GoalsZones_PenaltyArea	0.0245991643
GoalsZones_OutOfBox	.
GoalsSituations_OpenPlay	.
GoalsSituations_Counter	.
GoalsSituations_SetPiece	.
GoalsSituations_PenaltyScored	0.0579291028
GoalsSituations_Own	0.0869207019
GoalsSituations_Normal	0.1717135160
GoalsBodyParts_RightFoot	.
GoalsBodyParts_LeftFoot	.
GoalsBodyParts_Head	.
Dribbles_Unsuccessful	.
Dribbles_Successful	.
Dribbles_Total	0.0062294301

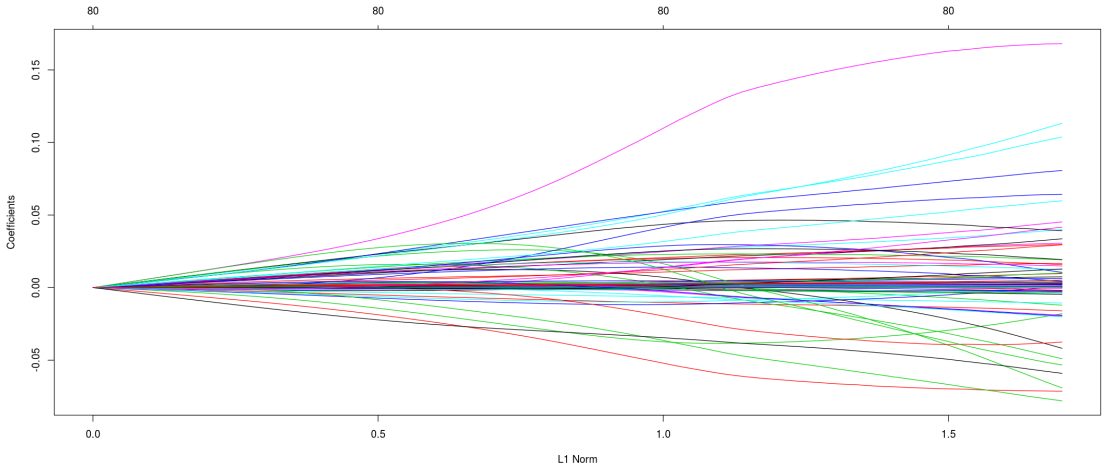
```
PossesionLoss_UnsuccessfulTouches  .
PossesionLoss_Dispossessed         .
Aerial_Total                       .
Aerial_Won                         0.0019604181
Aerial_Lost                        .
Assists_Cross                      .
Assists_Corner                     .
Assists_Throughball                .
Assists_Freekick                   .
Assists_Other                      .
Assists_Total                      .
PassesLength_Total                 0.0001116956
PassesLength_AccLB                 0.0003080580
PassesLength_InAccLB               .
PassesLength_AccSP                 .
PassesLength_InAccSP               .
PassesType_AccCr                   .
PassesType_InAccCr                 .
PassesType_AccCrn                  0.0107611979
PassesType_InAccCrn                .
PassesType_AccFrK                  .
PassesType_InAccFrK                .
KeyPassesLength_Total              .
KeyPassesLength_Long               .
KeyPassesLength_Short              0.0026221787
KeyPassesType_Cross                .
KeyPassesType_Corner               .
KeyPassesType_Throughball          0.0482067835
KeyPassesType_Freekick             .
KeyPassesType_Throwin              .
KeyPassesType_Other                .
```

Les variables qu'il semble intéressant de retenir, d'après le modèle de régression Lasso, pour l'ensemble des données, sont :
Tackles_TotalTackles ; Interception_Total ; Cards_Red ; Blocks_CrossesBlocked ; Blocks_PassesBlocked ; Saves_Total ; Saves_SixYardBox ; shotsSituations_Counter ; ShotsAccuracy_OffTarget ; GoalsZones_PenaltyArea ; GoalsSituations_PenaltyScored ; GoalsSituations_Own ; GoalsSituations_Normal ; Dribbles_Total ; PossesionLoss_UnsuccessfulTouches ; Aerial_Won ; PassesLength_Total ; PassesLength_AccLB ; PassesType_AccCrn ; KeyPassesLength_Short ; KeyPassesType_Throughball

1.1 Régression RIDGE

[26]:
```
m_ridge = glmnet(as.matrix(data), data_tot$Rating, alpha = 0, nlambda = 100)
```

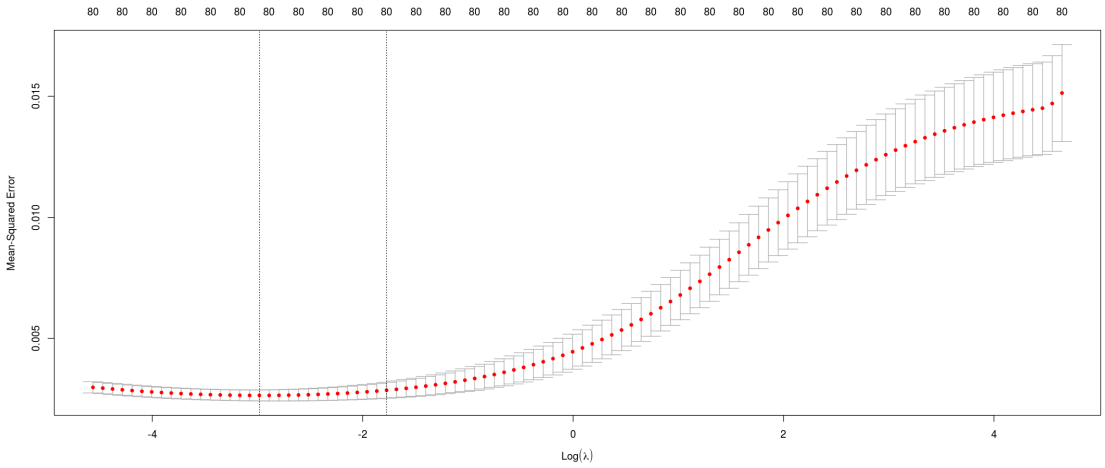
[27]:
```
plot(m_ridge)
```



Cross validation

[28]:
```
ridge.out <- cv.glmnet(as.matrix(data), data_tot$Rating, alpha = 0)
```

[29]:
```
plot(ridge.out)
```



[30]:
```
bestlam_ridge <- ridge.out$lambda.min
```

[31]:
```
p = predict(m_ridge, type = "coefficients", s = bestlam_ridge)
```

[32]:
```
print(p)
```

```
81 x 1 sparse Matrix of class "dgCMatrix"
                                      1
(Intercept)                5.933385e+00
```

Tackles_TotalTackles	5.029562e-03
Tackles_DribbledPast	-1.050493e-03
Tackles_TotalAttemptedTackles	1.930825e-03
Interception_Total	4.113718e-03
Fouls_Fouled	1.577533e-03
Fouls_Fouls	-2.254448e-03
Cards_Yellow	1.383592e-03
Cards_Red	-6.233630e-02
OffSides_CaughtOffside	-7.043090e-04
Clearances_Total	1.302633e-03
Blocks_ShotsBlocked	-7.378592e-03
Blocks_CrossesBlocked	2.057568e-02
Blocks_PassesBlocked	2.638850e-03
Saves_Total	-1.150756e-02
Saves_SixYardBox	-3.801867e-02
Saves_PenaltyArea	-1.020541e-02
Saves_OutOfBox	-8.376784e-03
Shots_Total	9.779845e-04
ShotsZones_OutOfBox	-1.632885e-03
ShotsZones_SixYardBox	1.275089e-02
ShotsZones_PenaltyArea	2.791921e-03
ShotsSituations_OpenPlay	1.726147e-03
shotsSituations_Counter	2.834539e-02
ShotsSituations_SetPiece	-7.443996e-03
ShotsSituations_PenaltyTaken	-1.412563e-03
ShotsAccuracy_OffTarget	5.380717e-03
ShotsAccuracy_OnPost	-1.023178e-02
ShotsAccuracy_OnTarget	4.390700e-03
ShotsAccuracy_Blocked	-9.128814e-03
ShotsBodyParts_RightFoot	1.041340e-03
ShotsBodyParts_LeftFoot	1.114678e-03
ShotsBodyParts_Head	5.744813e-04
ShotsBodyParts_Other	-4.878783e-02
Goals_Total	1.339259e-02
GoalsZones_SixYardBox	6.535650e-02
GoalsZones_PenaltyArea	2.934959e-02
GoalsZones_OutOfBox	4.638599e-02
GoalsSituations_OpenPlay	2.071820e-02
GoalsSituations_Counter	-1.613327e-03
GoalsSituations_SetPiece	6.073203e-02
GoalsSituations_PenaltyScored	6.462555e-02
GoalsSituations_Own	1.381570e-01
GoalsSituations_Normal	2.226449e-02
GoalsBodyParts_RightFoot	2.316906e-02
GoalsBodyParts_LeftFoot	2.313799e-02
GoalsBodyParts_Head	2.928428e-02
Dribbles_Unsuccessful	3.883473e-03
Dribbles_Successful	3.194385e-03

```
Dribbles_Total                      2.300626e-03
PossesionLoss_UnsuccessfulTouches   2.807157e-03
PossesionLoss_Dispossessed         -2.276545e-03
Aerial_Total                        4.737453e-04
Aerial_Won                          1.708597e-03
Aerial_Lost                         8.702412e-05
Assists_Cross                       2.681161e-02
Assists_Corner                     -3.031979e-02
Assists_Throughball                -6.927754e-03
Assists_Freekick                    5.183321e-02
Assists_Other                       1.724586e-02
Assists_Total                       1.762455e-02
PassesLength_Total                  7.758378e-05
PassesLength_AccLB                  9.972125e-04
PassesLength_InAccLB               -6.788826e-05
PassesLength_AccSP                  6.728336e-05
PassesLength_InAccSP                4.571695e-04
PassesType_AccCr                    9.478577e-04
PassesType_InAccCr                 -1.727322e-03
PassesType_AccCrn                   1.661938e-02
PassesType_InAccCrn                 3.615475e-03
PassesType_AccFrK                   8.949087e-06
PassesType_InAccFrK                -1.598352e-03
KeyPassesLength_Total               2.588134e-03
KeyPassesLength_Long                3.320002e-03
KeyPassesLength_Short               2.748708e-03
KeyPassesType_Cross                 4.819064e-04
KeyPassesType_Corner               -7.117289e-03
KeyPassesType_Throughball           3.987658e-02
KeyPassesType_Freekick              2.053771e-02
KeyPassesType_Throwin              -3.904237e-02
KeyPassesType_Other                 3.006856e-03
```

Ici c'est beaucoup moins évident de faire de la sélection de variables : les coefficients ne s'annulent pas. Certains sont cependant très petits ($1e-4$).
Si on ne souhaite garder que celles dont le coefficient est au moins de l'ordre de 10^{-2}, on peut citer :

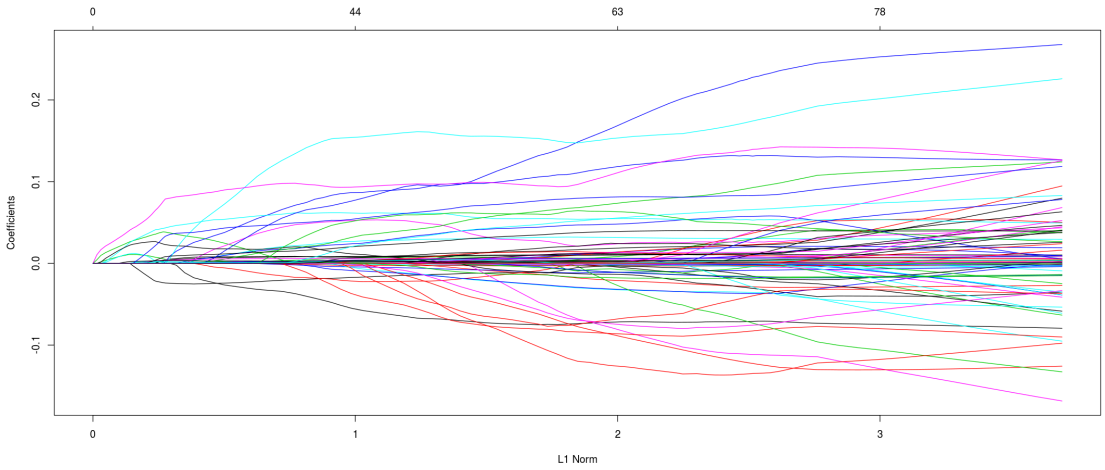
(ancienne version erreur)
Cards_Red ; OffSides_CaughtOffside ; Blocks_ShotsBlocked ; Blocks_CrossesBlocked ; Saves_Total ; Saves_SixYardBox ; ShotsZones_SixYardBox ; shotsSituations_Counter ; ShotsSituations_SetPiece ; ShotsSituations_PenaltyTaken ; ShotsAccuracy_OnPost ; ShotsAccuracy_Blocked ; ShotsBodyParts_Other ; GoalsZones_SixYardBox ; GoalsZones_PenaltyArea ; GoalsZones_OutOfBox ; GoalsSituations_OpenPlay ; GoalsSituations_Counter ; GoalsSituations_SetPiece ; GoalsSituations_PenaltyScored ; GoalsSituations_Own ; GoalsSituations_Normal ; GoalsBodyParts_RightFoot ; GoalsBodyParts_LeftFoot ; GoalsBodyParts_Head ; Assists_Cross ; Assists_Corner ; Assists_Throughball ; Assists_Freekick ; Assists_Other ; Assists_Total ; PassesType_AccCrn ; PassesType_InAccCrn ; KeyPassesLength_Long ; KeyPassesType_Corner ; KeyPassesType_Throughball ; KeyPassesType_Freekick ; KeyPassesType_Throwin

correction :
Cards_Red ; ShotsBodyParts_Other ; Saves_SixYardBox ; KeyPassesType_Throwin ; Assists_Corner ; Saves_Total ; Saves_PenaltyArea ; ShotsZones_SixYardBox ; Goals_Total ; PassesType_AccCrn ; Assists_Other ; Assists_Total ; KeyPassesType_Freekick ; Blocks_CrossesBlocked ; GoalsSituations_OpenPlay ; GoalsSituations_Normal ; GoalsBodyParts_RightFoot ; GoalsBodyParts_LeftFoot ; Assists_Cross ; shotsSituations_Counter ; GoalsZones_PenaltyArea ; GoalsBodyParts_Head ; KeyPassesType_Throughball ; GoalsZones_OutOfBox ; Assists_Freekick ; GoalsSituations_SetPiece ; GoalsSituations_PenaltyScored ; GoalsZones_SixYardBox ; GoalsSituations_Own

1.2 Régression Elastic Net

[33]:
```
m_enet = glmnet(as.matrix(data), data_tot$Rating, alpha = 0.5, nlambda = 100)
```

[34]:
```
plot(m_enet)
```

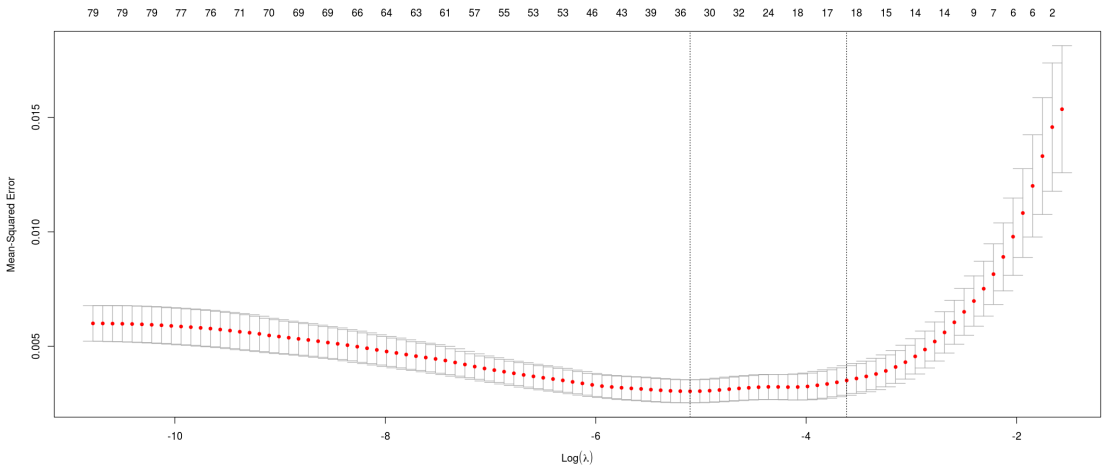


Cross validation

[35]:
```
enet.out <- cv.glmnet(as.matrix(data), data_tot$Rating, alpha = 0.5)
```

[36]:
```
bestlam_enet <- enet.out$lambda.min
```

[37]:
```
plot(enet.out)
```



[38]:
```
predict(m_enet, type = "coefficients", s = bestlam_enet)
```

```
81 x 1 sparse Matrix of class "dgCMatrix"
                                       1
(Intercept)                 5.960152e+00
Tackles_TotalTackles        7.884472e-03
Tackles_DribbledPast        .
Tackles_TotalAttemptedTackles .
Interception_Total          3.387571e-03
Fouls_Fouled                .
Fouls_Fouls                -3.226268e-04
Cards_Yellow                .
Cards_Red                  -3.414001e-02
OffSides_CaughtOffside      .
Clearances_Total            .
Blocks_ShotsBlocked        -1.336716e-03
Blocks_CrossesBlocked       2.042328e-02
Blocks_PassesBlocked        2.586785e-03
Saves_Total                -2.065990e-02
Saves_SixYardBox           -1.518781e-02
Saves_PenaltyArea           .
Saves_OutOfBox              .
Shots_Total                 .
ShotsZones_OutOfBox         .
ShotsZones_SixYardBox       .
ShotsZones_PenaltyArea      .
ShotsSituations_OpenPlay    .
shotsSituations_Counter     1.723389e-02
ShotsSituations_SetPiece    .
ShotsSituations_PenaltyTaken .
ShotsAccuracy_OffTarget     8.531610e-03
```

ShotsAccuracy_OnPost	.
ShotsAccuracy_OnTarget	7.341222e-04
ShotsAccuracy_Blocked	.
ShotsBodyParts_RightFoot	.
ShotsBodyParts_LeftFoot	.
ShotsBodyParts_Head	.
ShotsBodyParts_Other	-3.363974e-04
Goals_Total	4.232759e-03
GoalsZones_SixYardBox	6.576370e-02
GoalsZones_PenaltyArea	6.021550e-02
GoalsZones_OutOfBox	4.416339e-02
GoalsSituations_OpenPlay	1.916481e-02
GoalsSituations_Counter	.
GoalsSituations_SetPiece	1.666822e-02
GoalsSituations_PenaltyScored	5.519323e-02
GoalsSituations_Own	1.257398e-01
GoalsSituations_Normal	9.781903e-02
GoalsBodyParts_RightFoot	.
GoalsBodyParts_LeftFoot	.
GoalsBodyParts_Head	.
Dribbles_Unsuccessful	.
Dribbles_Successful	.
Dribbles_Total	5.999990e-03
PossesionLoss_UnsuccessfulTouches	1.624991e-03
PossesionLoss_Dispossessed	-7.128936e-04
Aerial_Total	.
Aerial_Won	2.635428e-03
Aerial_Lost	.
Assists_Cross	.
Assists_Corner	.
Assists_Throughball	.
Assists_Freekick	1.884938e-02
Assists_Other	.
Assists_Total	.
PassesLength_Total	1.662660e-04
PassesLength_AccLB	5.011314e-04
PassesLength_InAccLB	.
PassesLength_AccSP	.
PassesLength_InAccSP	.
PassesType_AccCr	.
PassesType_InAccCr	.
PassesType_AccCrn	1.466273e-02
PassesType_InAccCrn	.
PassesType_AccFrK	.
PassesType_InAccFrK	.
KeyPassesLength_Total	1.150391e-03
KeyPassesLength_Long	.
KeyPassesLength_Short	2.638233e-03

```
KeyPassesType_Cross                .
KeyPassesType_Corner               .
KeyPassesType_Throughball          4.837109e-02
KeyPassesType_Freekick             .
KeyPassesType_Throwin              .
KeyPassesType_Other                2.021758e-05
```

[ ]: