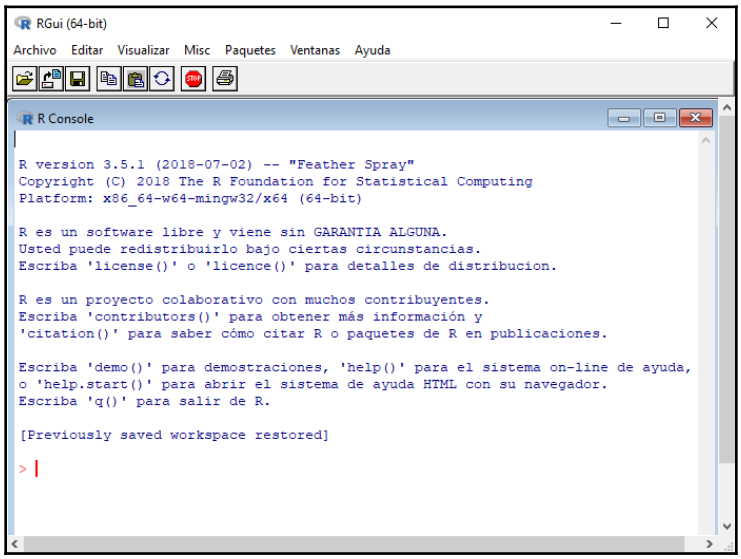
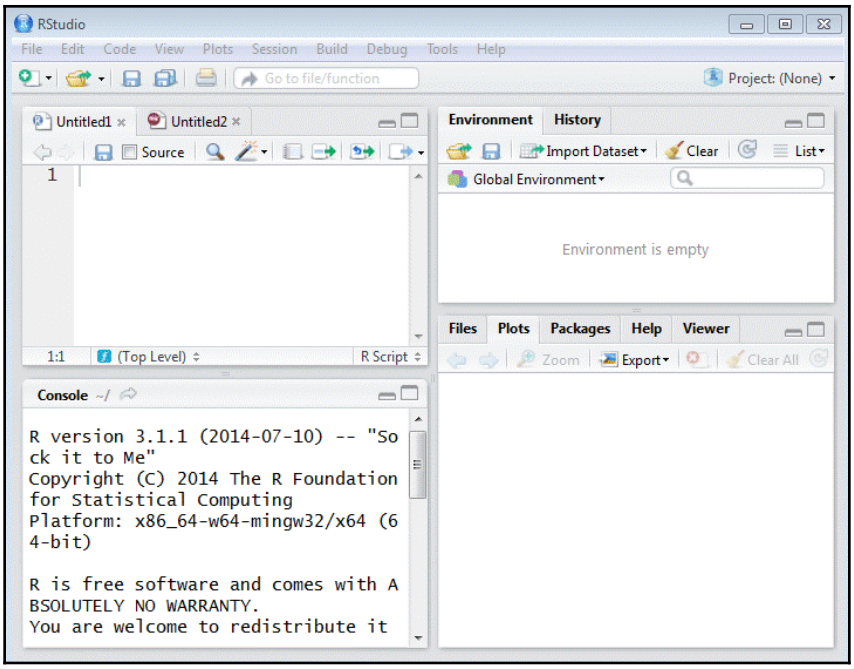
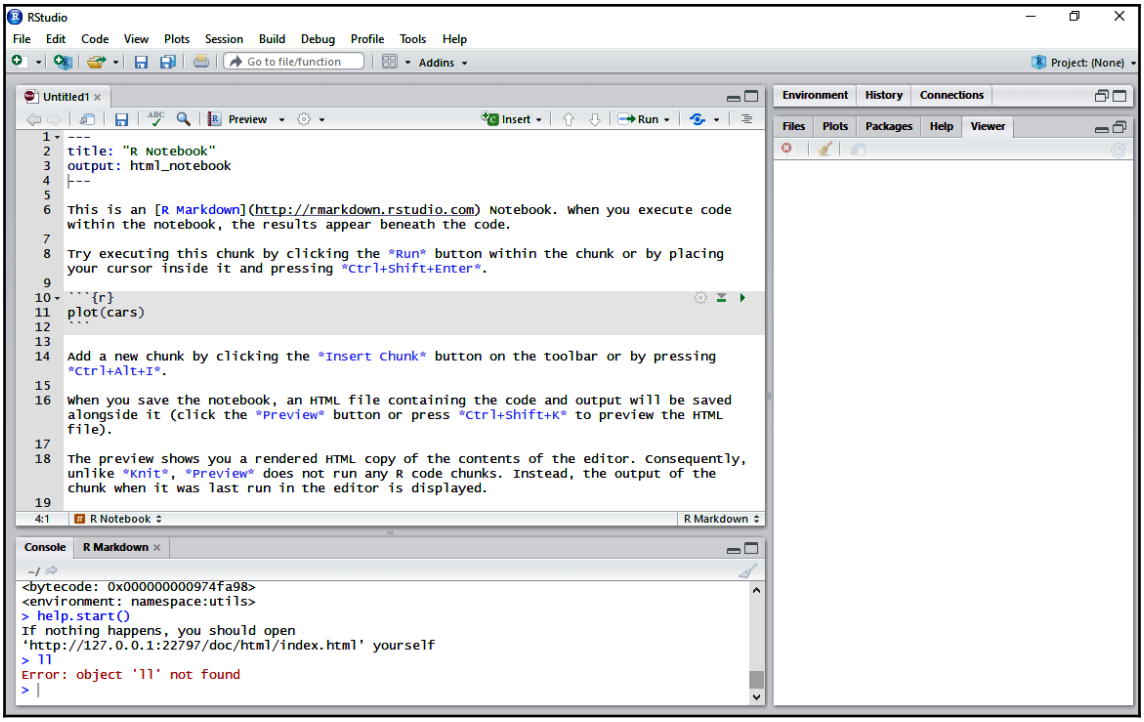
Chapter 01: R Fundamentals for Machine Learning

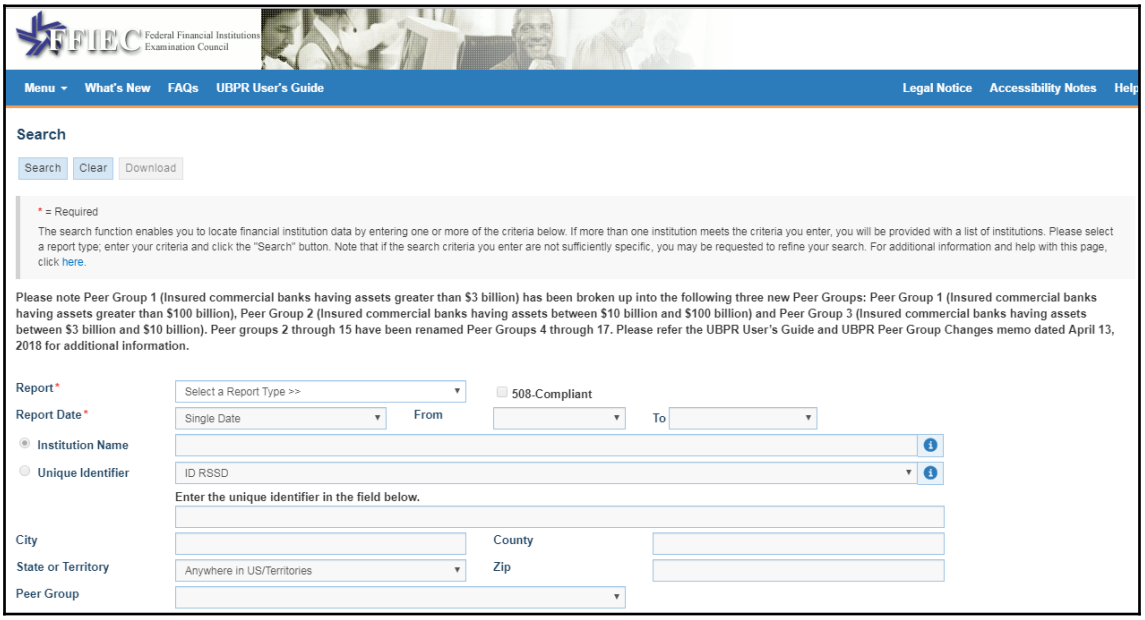


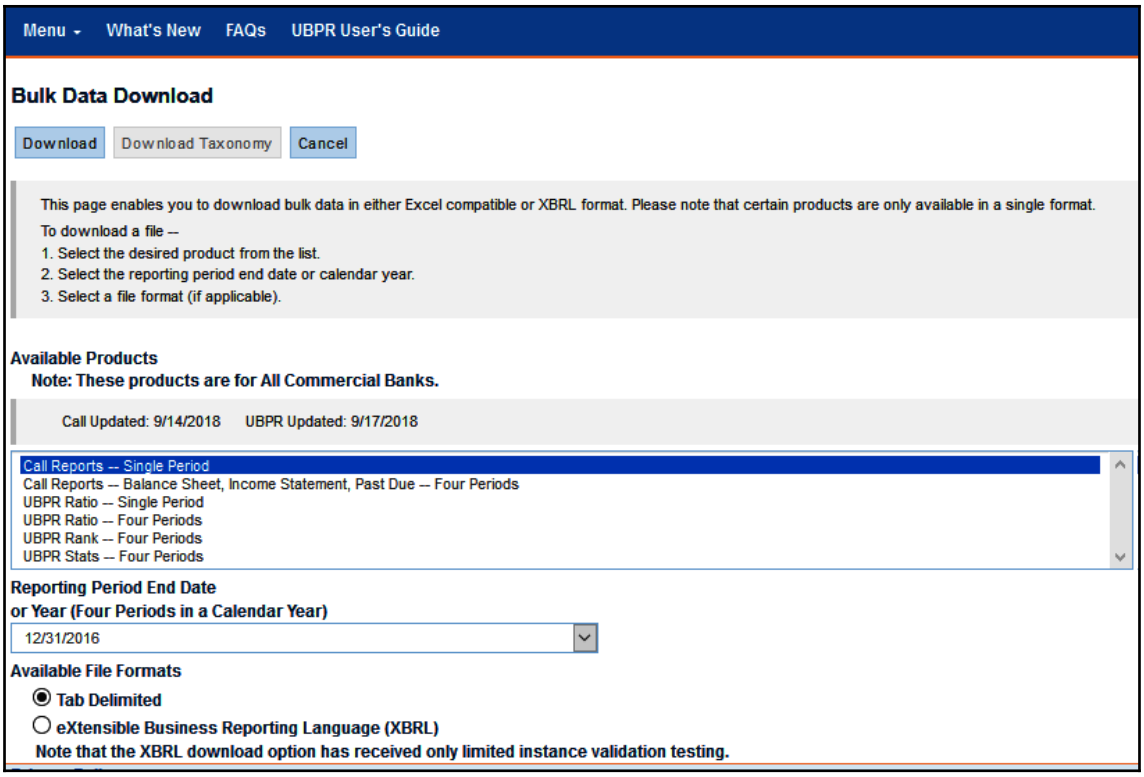

RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
Project: (None)
Untitled1
Untitled2
Source
1
1:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Zoom
Export
Clear All
Console ~/
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Untitled1
Preview
Insert
Run
---
title: "R Notebook"
output: html_notebook

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.
Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.
```{r}
plot(cars)
```
Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.
When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).
The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.
4:1
R Notebook
R Markdown
Console
R Markdown
~/
<bytecode: 0x000000000974fa98>
<environment: namespace:utils>
> help.start()
If nothing happens, you should open
'http://127.0.0.1:22797/doc/html/index.html' yourself
> ll
Error: object 'll' not found
>
Environment
History
Connections
Files
Plots
Packages
Help
Viewer

Chapter 02: Predicting Failures of Banks - Data Collection




Menu
What's New
FAQs
UBPR User's Guide
Bulk Data Download
Download
Download Taxonomy
Cancel
This page enables you to download bulk data in either Excel compatible or XBRL format. Please note that certain products are only available in a single format.
To download a file --
1. Select the desired product from the list.
2. Select the reporting period end date or calendar year.
3. Select a file format (if applicable).
Available Products
Note: These products are for All Commercial Banks.
Call Updated: 9/14/2018
UBPR Updated: 9/17/2018
Call Reports -- Single Period
Call Reports -- Balance Sheet, Income Statement, Past Due -- Four Periods
UBPR Ratio -- Single Period
UBPR Ratio -- Four Periods
UBPR Rank -- Four Periods
UBPR Stats -- Four Periods
Reporting Period End Date
or Year (Four Periods in a Calendar Year)
12/31/2016
Available File Formats
Tab Delimited
eXtensible Business Reporting Language (XBRL)
Note that the XBRL download option has received only limited instance validation testing.

Failed Bank List

The FDIC is often appointed as receiver for failed banks. This page contains useful information for the customers and vendors of these banks. This includes information on the acquiring bank (if applicable), how your accounts and loans are affected, and how vendors can file claims against the receivership. Failed Financial Institution Contact Search displays point of contact information related to failed banks.

This list includes banks which have failed since October 1, 2000. To search for banks that failed prior to those on this page, visit this link: Failures and Assistance Transactions

Click arrows next to headers to sort in Ascending or Descending order.

Download Data

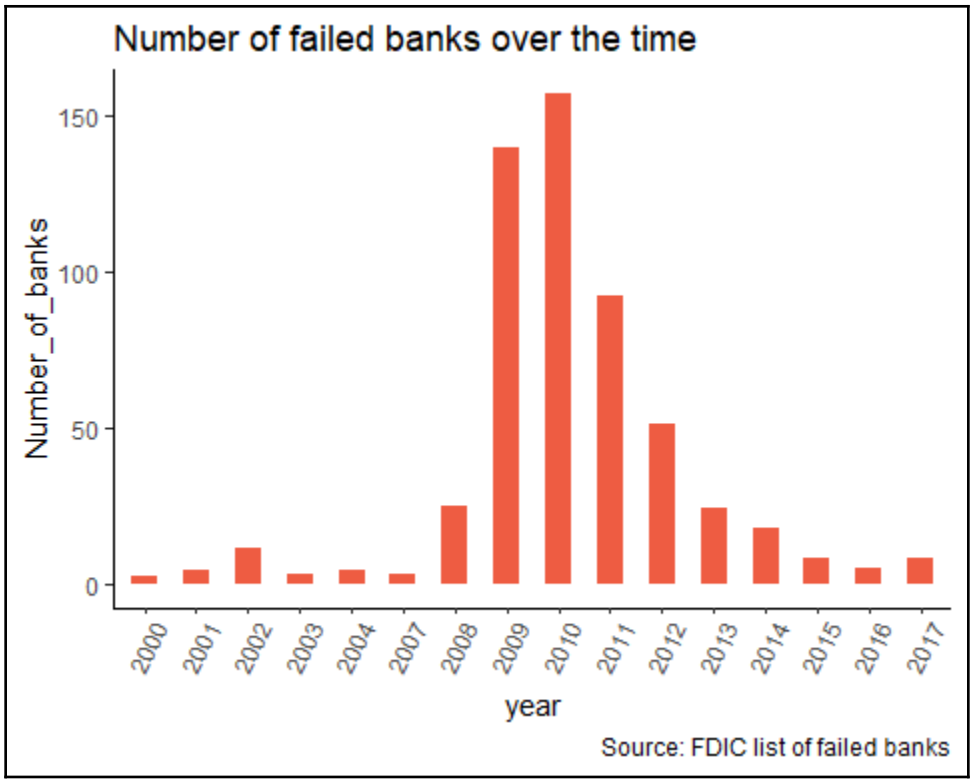
Search:

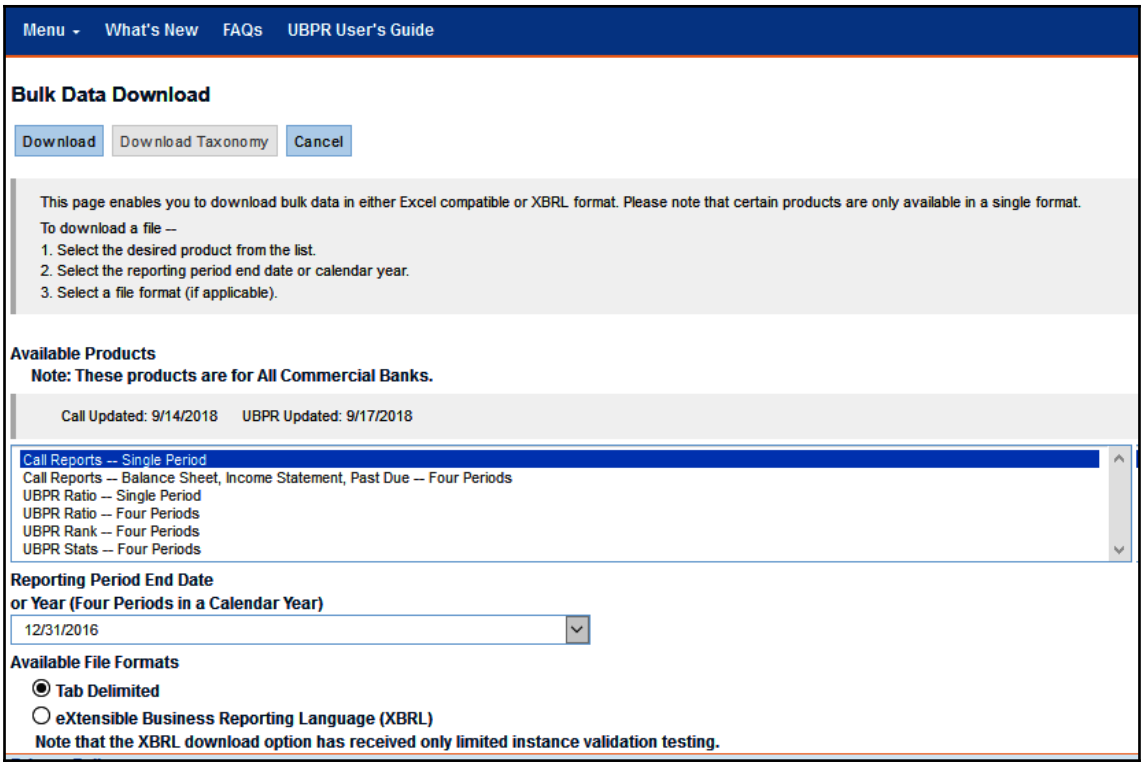
Show 25 **entries**

Showing 1 to 25 of 555 entries

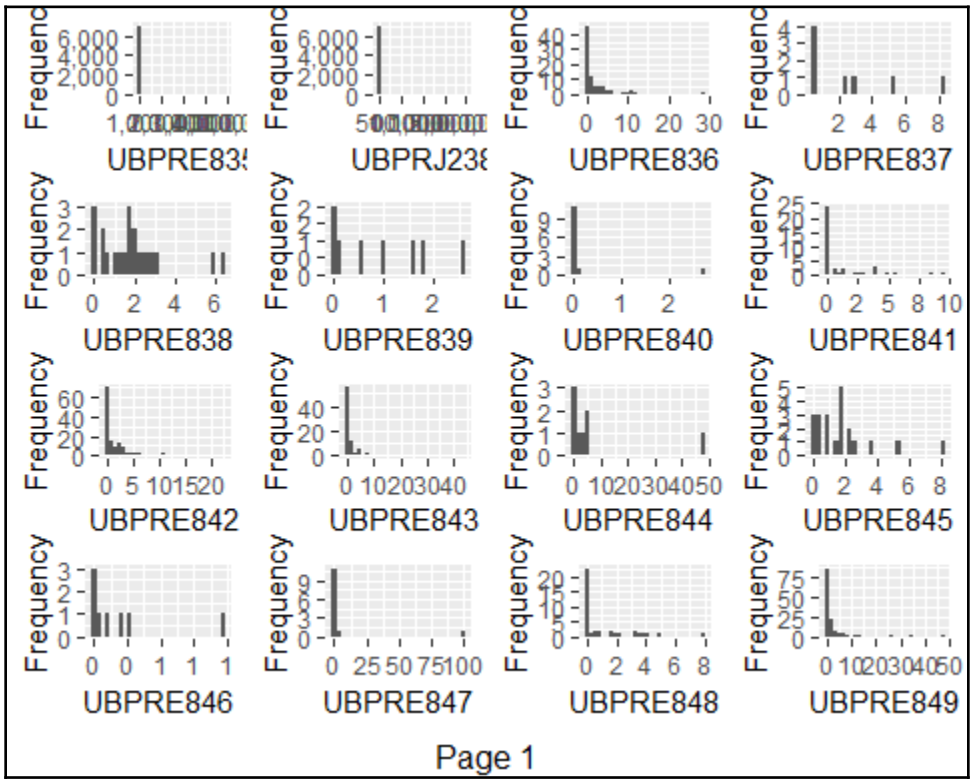
First Previous 1 2 3 4 5 Next Last

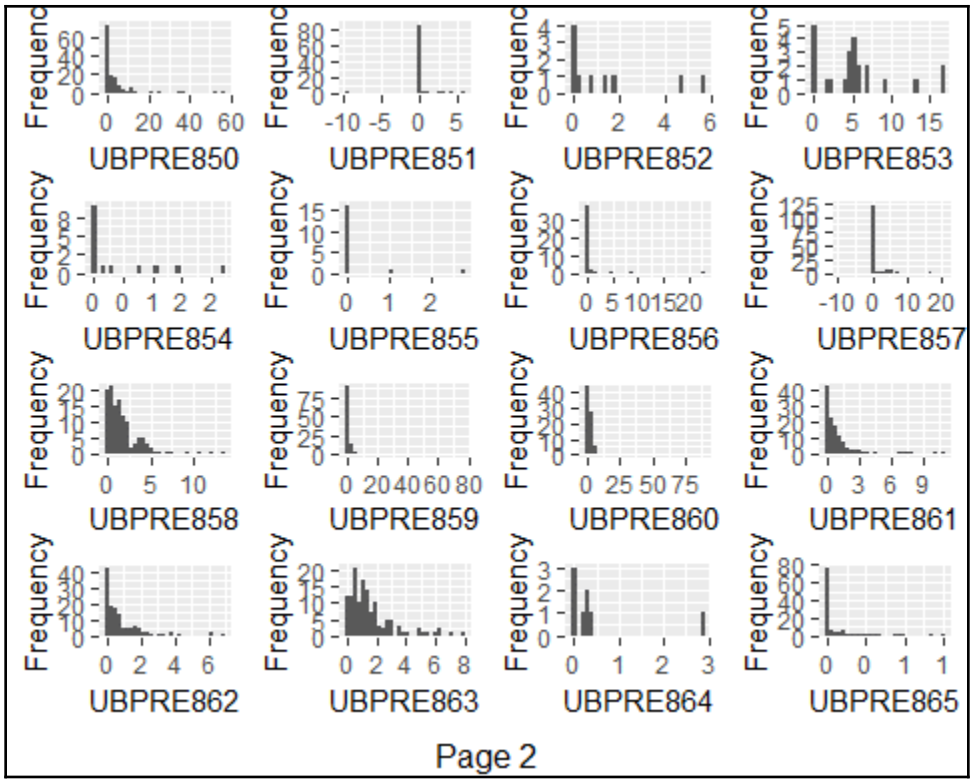
Bank Name	City	ST	CERT	Acquiring Institution	Closing Date	Updated Date
Washington Federal Bank for Savings	Chicago	IL	30570	Royal Savings Bank	December 15, 2017	February 21, 2018
The Farmers and Merchants State Bank of Argonia	Argonia	KS	17719	Conway Bank	October 13, 2017	February 21, 2018
Fayette County Bank	Saint Elmo	IL	1802	United Fidelity Bank, fsb	May 26, 2017	July 26, 2017
Guaranty Bank, (d/b/a BestBank in Georgia & Michigan)	Milwaukee	WI	30003	First-Citizens Bank & Trust Company	May 5, 2017	March 22, 2018
First NBC Bank	New Orleans	LA	58302	Whitney Bank	April 28, 2017	December 5, 2017

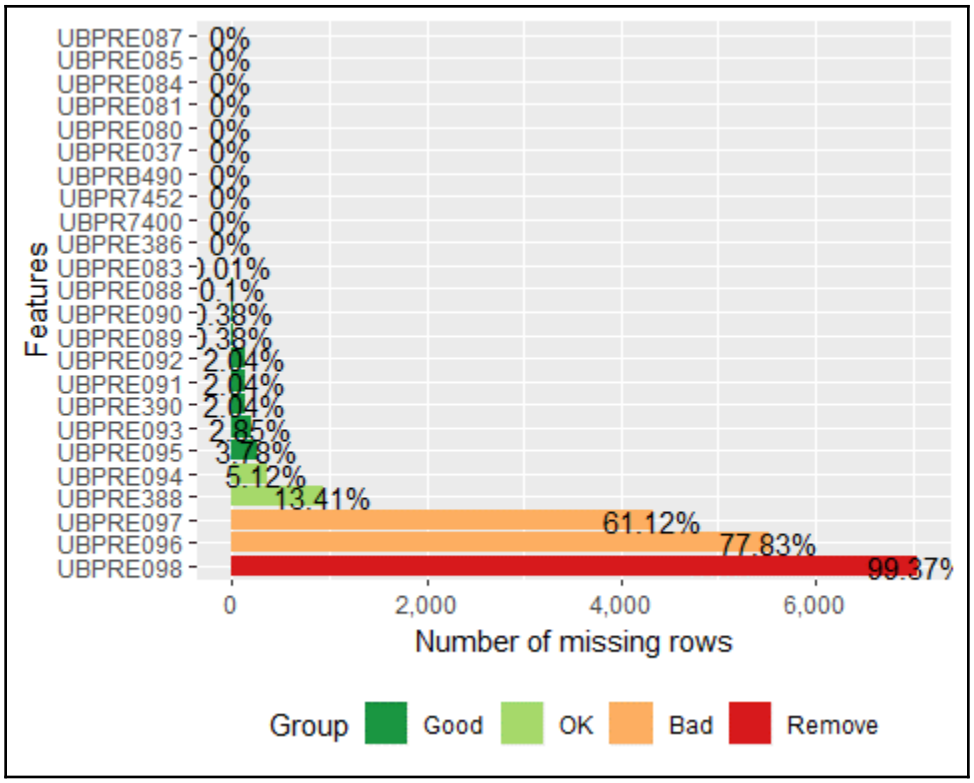

Number of failed banks over the time
Number_of_banks
150
100
50
0
2000
2001
2002
2003
2004
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
year
Source: FDIC list of failed banks

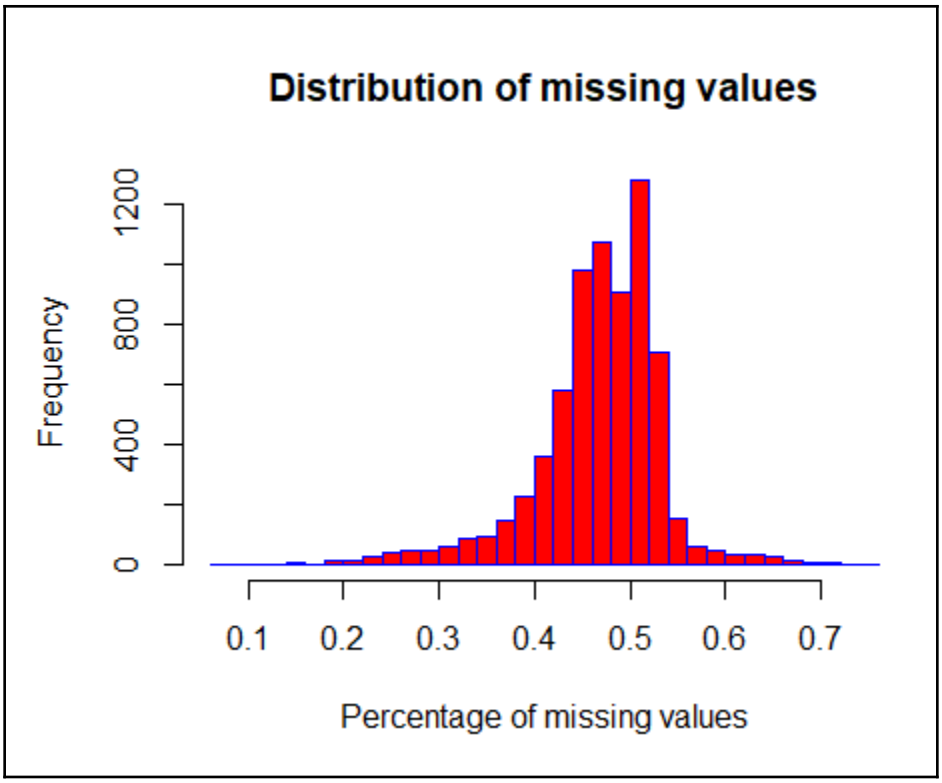

Menu
What's New
FAQs
UBPR User's Guide
Bulk Data Download
Download
Download Taxonomy
Cancel
This page enables you to download bulk data in either Excel compatible or XBRL format. Please note that certain products are only available in a single format.
To download a file --
1. Select the desired product from the list.
2. Select the reporting period end date or calendar year.
3. Select a file format (if applicable).
Available Products
Note: These products are for All Commercial Banks.
Call Updated: 9/14/2018
UBPR Updated: 9/17/2018
Call Reports -- Single Period
Call Reports -- Balance Sheet, Income Statement, Past Due -- Four Periods
UBPR Ratio -- Single Period
UBPR Ratio -- Four Periods
UBPR Rank -- Four Periods
UBPR Stats -- Four Periods
Reporting Period End Date
or Year (Four Periods in a Calendar Year)
12/31/2016
Available File Formats
Tab Delimited
eXtensible Business Reporting Language (XBRL)
Note that the XBRL download option has received only limited instance validation testing.

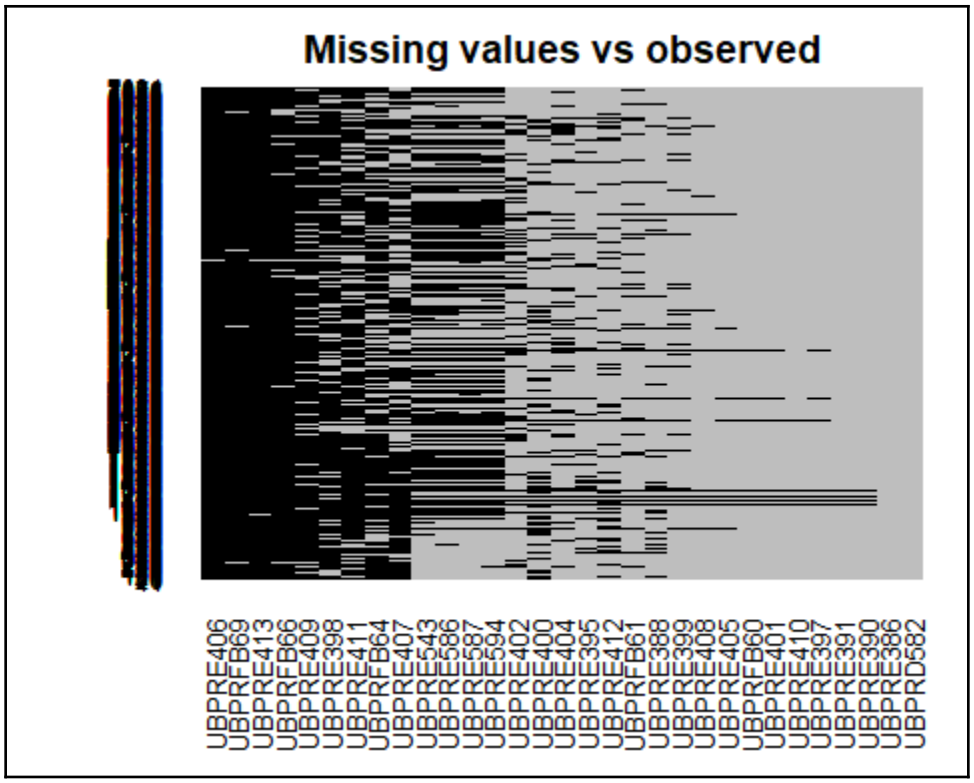
Chapter 03: Predicting Failures of Banks - Descriptive Analysis

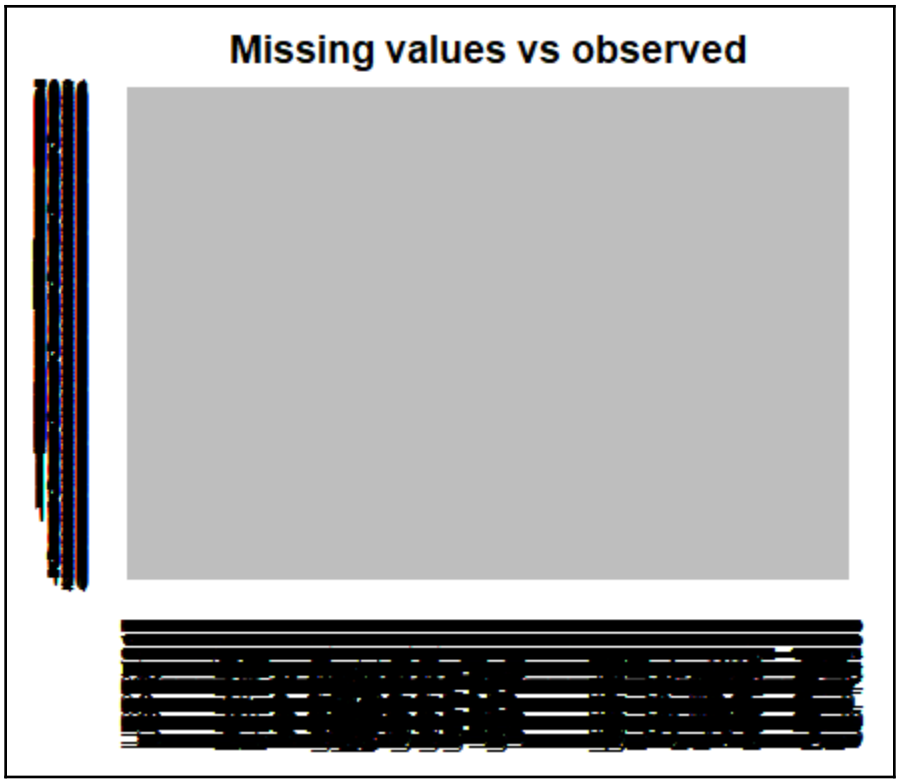


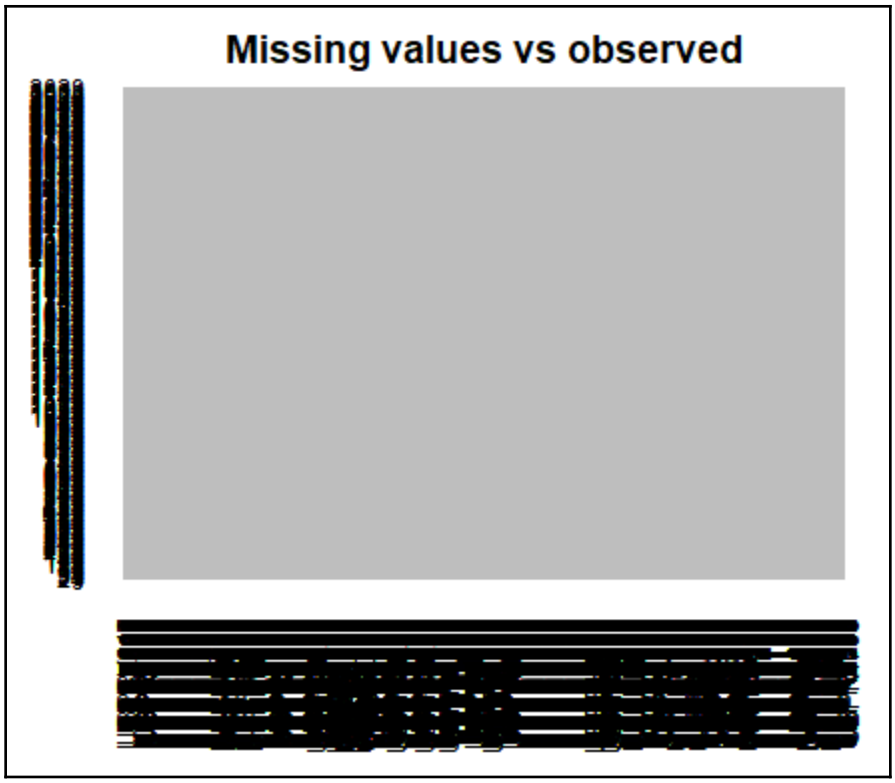

UBPRE850
UBPRE851
UBPRE852
UBPRE853
UBPRE854
UBPRE855
UBPRE856
UBPRE857
UBPRE858
UBPRE859
UBPRE860
UBPRE861
UBPRE862
UBPRE863
UBPRE864
UBPRE865
Frequency
Page 2


UBPRE087 - 0%
UBPRE085 - 0%
UBPRE084 - 0%
UBPRE081 - 0%
UBPRE080 - 0%
UBPRE037 - 0%
UBPRB490 - 0%
UBPR7452 - 0%
UBPR7400 - 0%
UBPRE386 - 0%
UBPRE083 - 0.01%
UBPRE088 - 0.1%
UBPRE090 - 0.38%
UBPRE089 - 0.38%
UBPRE092 - 2.04%
UBPRE091 - 2.04%
UBPRE390 - 2.04%
UBPRE093 - 2.85%
UBPRE095 - 3.78%
UBPRE094 - 5.12%
UBPRE388 - 13.41%
UBPRE097 - 61.12%
UBPRE096 - 77.83%
UBPRE098 - 99.37%
Features
0
2,000
4,000
6,000
Number of missing rows
Group
Good
OK
Bad
Remove


Distribution of missing values
Frequency
0
400
800
1200
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Percentage of missing values


Missing values vs observed
UBPRE406
UBPRFB69
UBPRE413
UBPRFB66
UBPRE409
UBPRE398
UBPRE411
UBPRFB64
UBPRE407
UBPRE543
UBPRE586
UBPRE587
UBPRE594
UBPRE402
UBPRE400
UBPRE404
UBPRE395
UBPRE412
UBPRFB61
UBPRE388
UBPRE399
UBPRE408
UBPRE405
UBPRFB60
UBPRE401
UBPRE410
UBPRE397
UBPRE391
UBPRE390
UBPRE386
UBPRD582


Missing values vs observed


Missing values vs observed

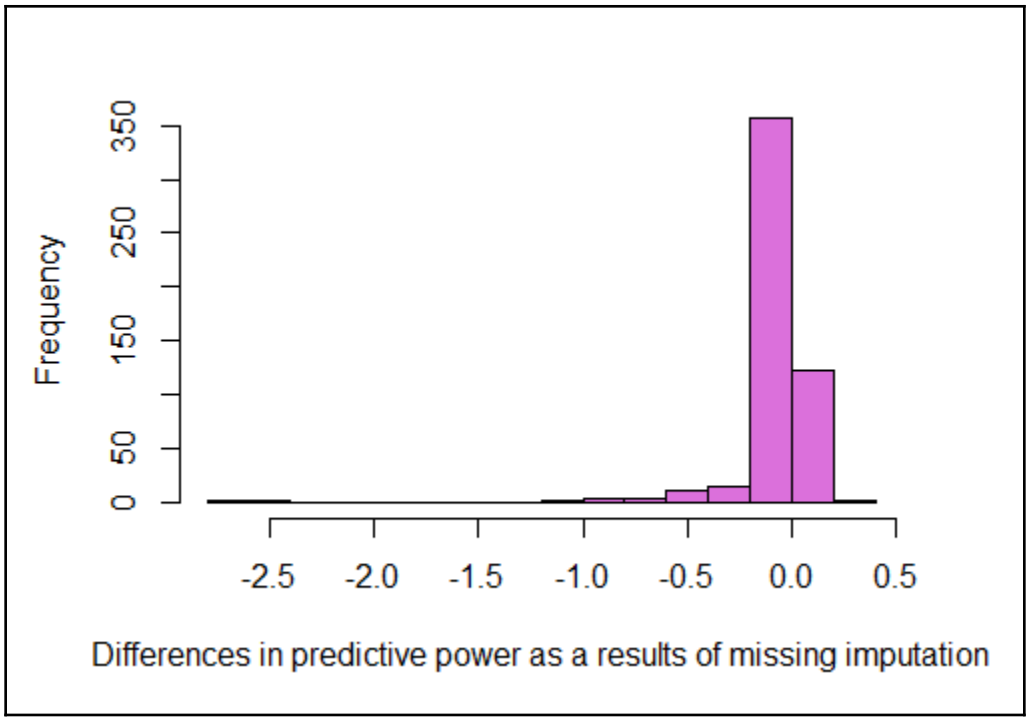
Chapter 04: Predicting Failures of Banks - Univariate Analysis

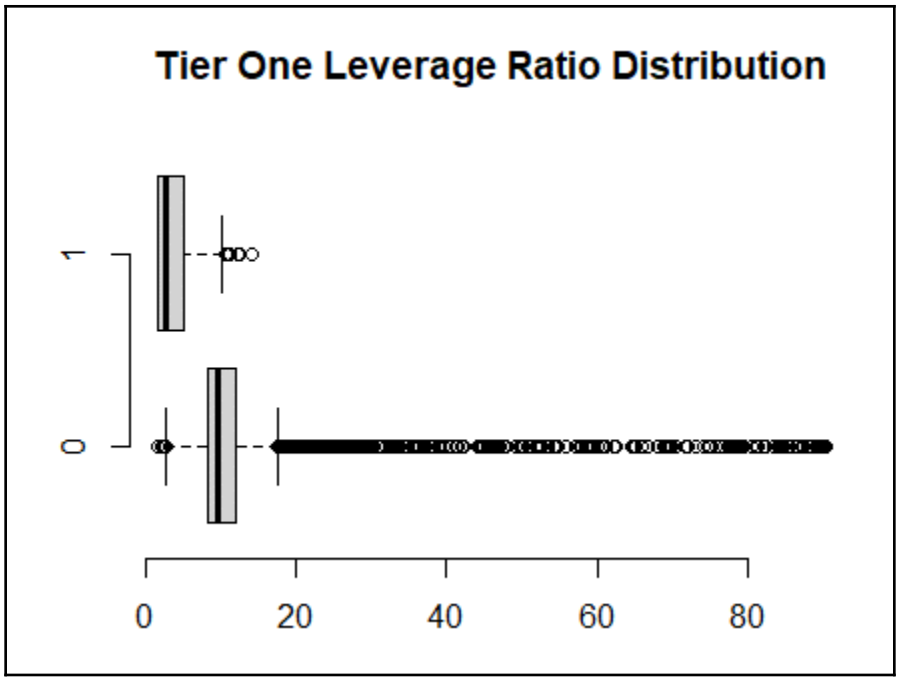
$$IV = \sum_{i-1}^{n} \left[(Distr\ Goods_i - Distr\ Bads_i) \quad * In \left(\frac{Distr\ Goods}{Distr\ Bads} \right) \right]$$

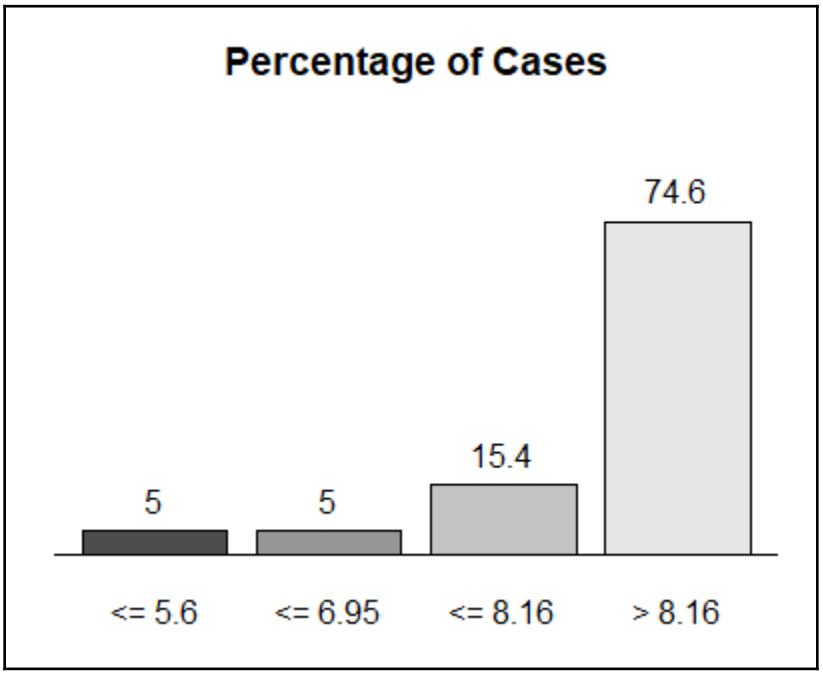
	Cutpoint	CntRec	CntGood	CntBad	CntCumRec	CntCumGood	CntCumBad	PctRec	GoodRate	BadRate	Odds	LnOdds	WoE	IV
1	<= 0.5487	5751	5705	46	5751	5705	46	0.8110	0.9920	0.0080	124.0217	4.8205	1.8101	1.2780
2	<= 0.9175	492	473	19	6243	6178	65	0.0694	0.9614	0.0386	24.8947	3.2147	0.2043	0.0026
3	<= 1.811	397	354	43	6640	6532	108	0.0560	0.8917	0.1083	8.2326	2.1081	-0.9022	0.0692
4	> 1.811	451	226	225	7091	6758	333	0.0636	0.5011	0.4989	1.0044	0.0044	-3.0059	1.9305
5	Missing	0	0	0	7091	6758	333	0.0000	NaN	NaN	NaN	NaN	NaN	NaN
6	Total	7091	6758	333	NA	NA	NA	1.0000	0.9530	0.0470	20.2943	3.0103	0.0000	3.2803

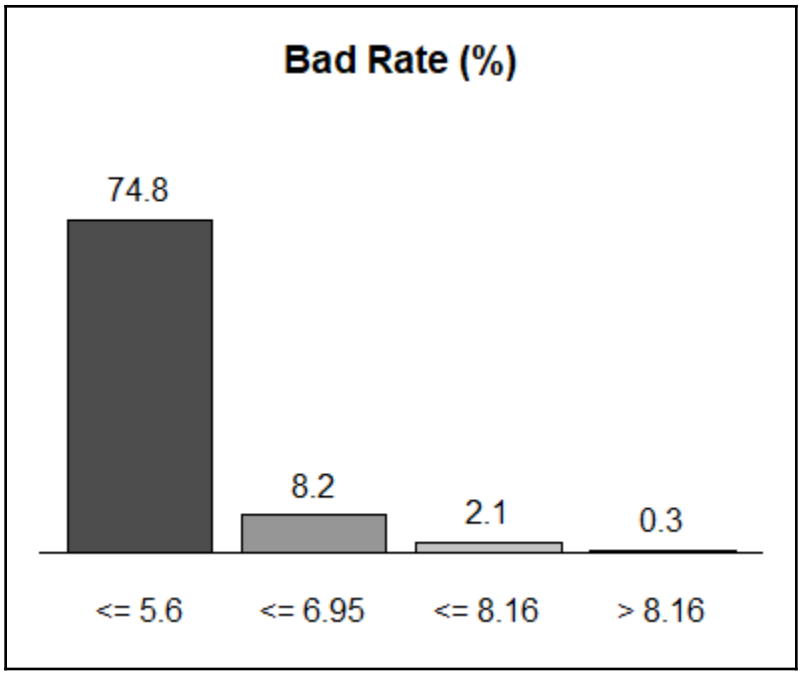
$$IV = (0.9920 - 0.0080) * \left(\frac{0.9920}{0.0080} \right) = 1.2780$$

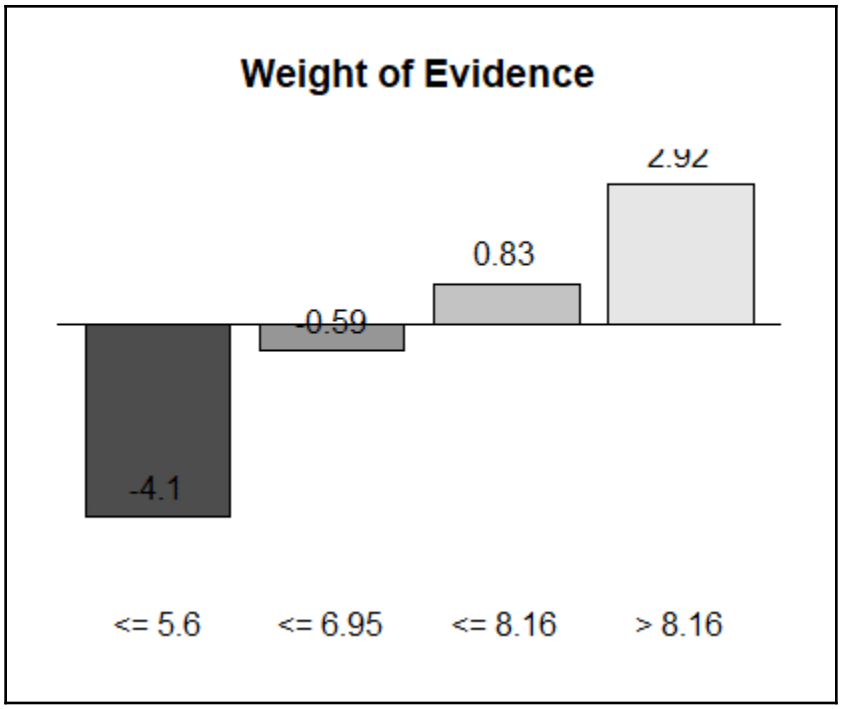
$$WoE = \left[In \left(\frac{Distr\ Goods}{Distr\ Bads} \right) \right] * 100$$

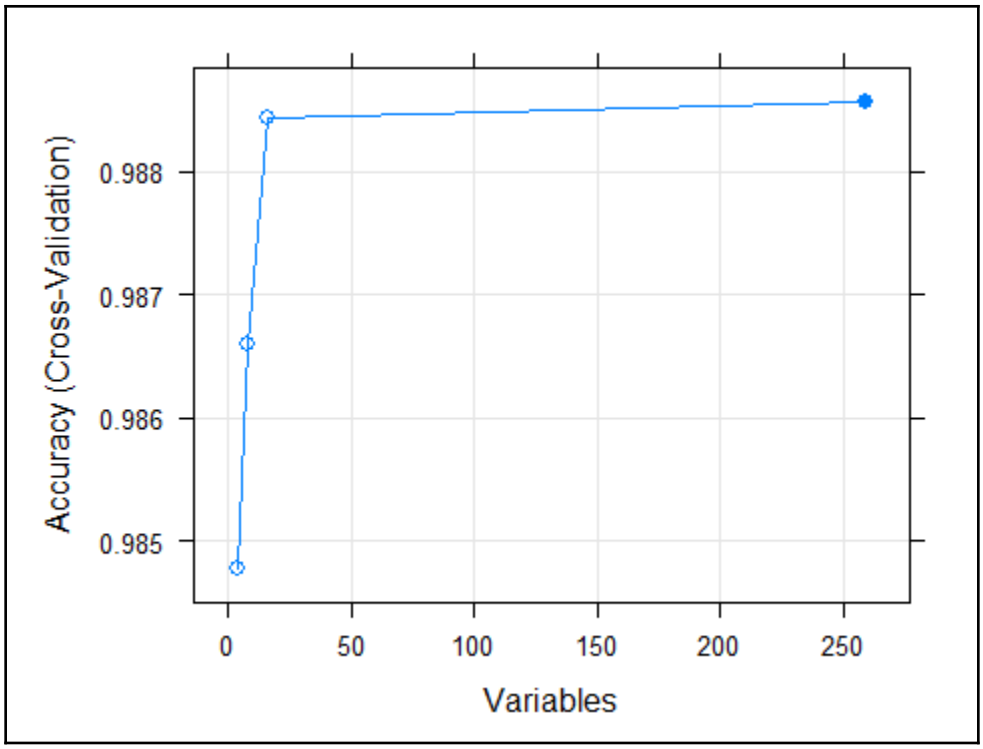

Frequency
0
50
150
250
350
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
Differences in predictive power as a results of missing imputation


Tier One Leverage Ratio Distribution
1
0
0
20
40
60
80


Percentage of Cases
74.6
15.4
5
5
<= 5.6
<= 6.95
<= 8.16
> 8.16


Bad Rate (%)
74.8
8.2
2.1
0.3
<= 5.6
<= 6.95
<= 8.16
> 8.16

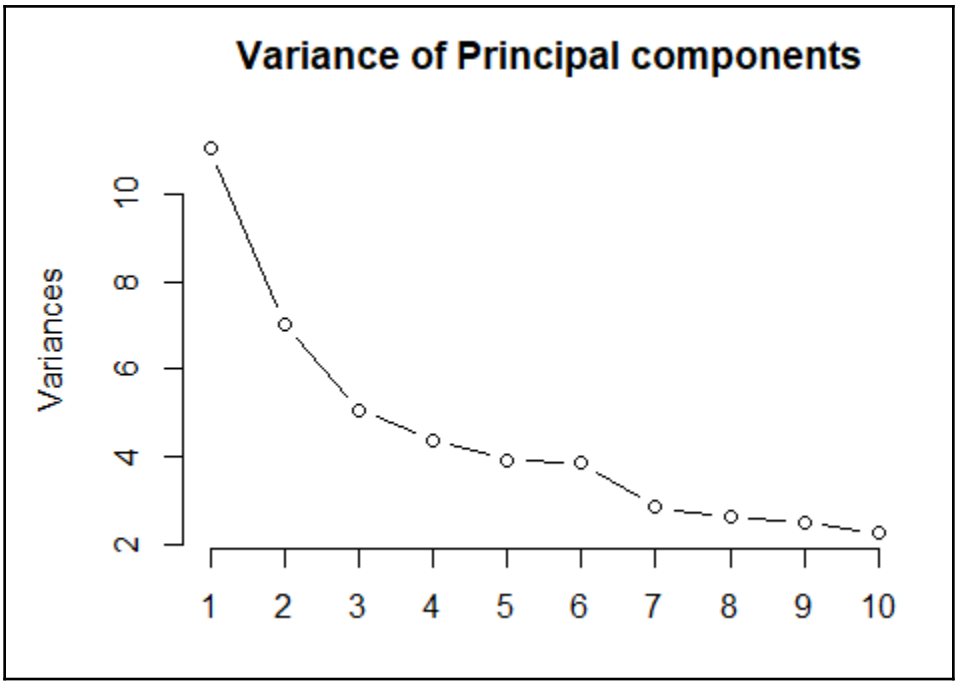

Weight of Evidence
2.92
0.83
-0.59
-4.1
<= 5.6
<= 6.95
<= 8.16
> 8.16

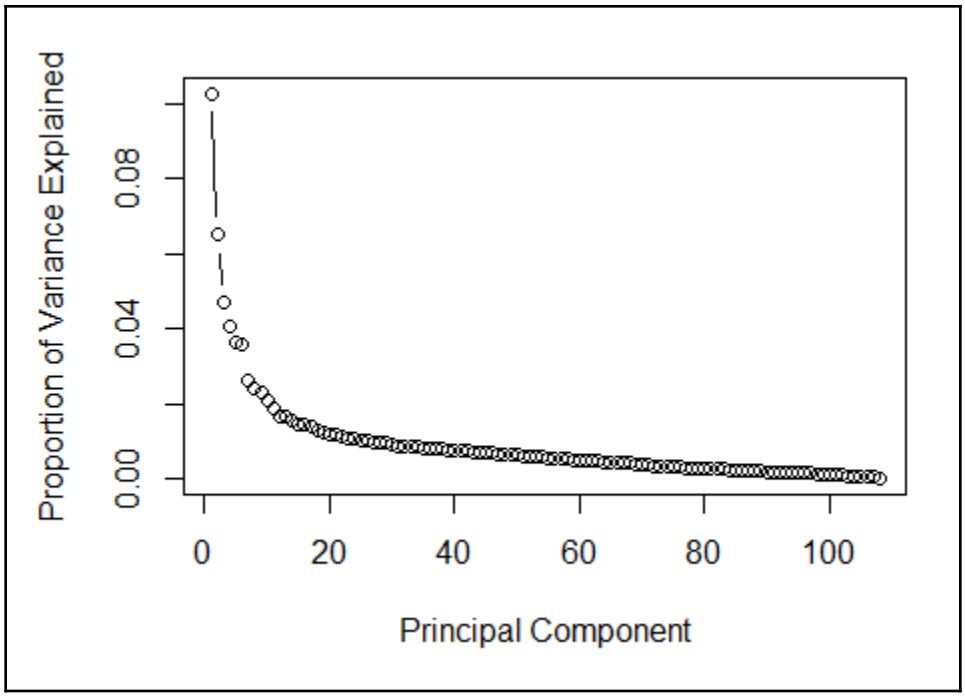


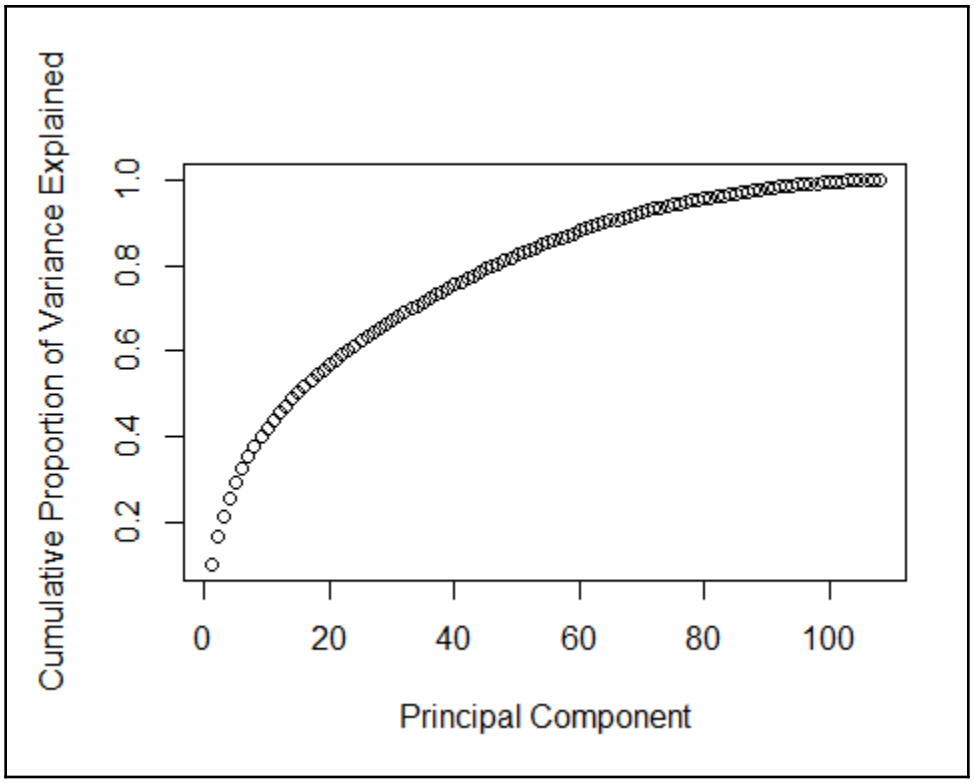
$$argmax_\beta \left(L(\beta|y,x) - \lambda \sum_{k=1}^{K} \beta_k^2 \right)$$

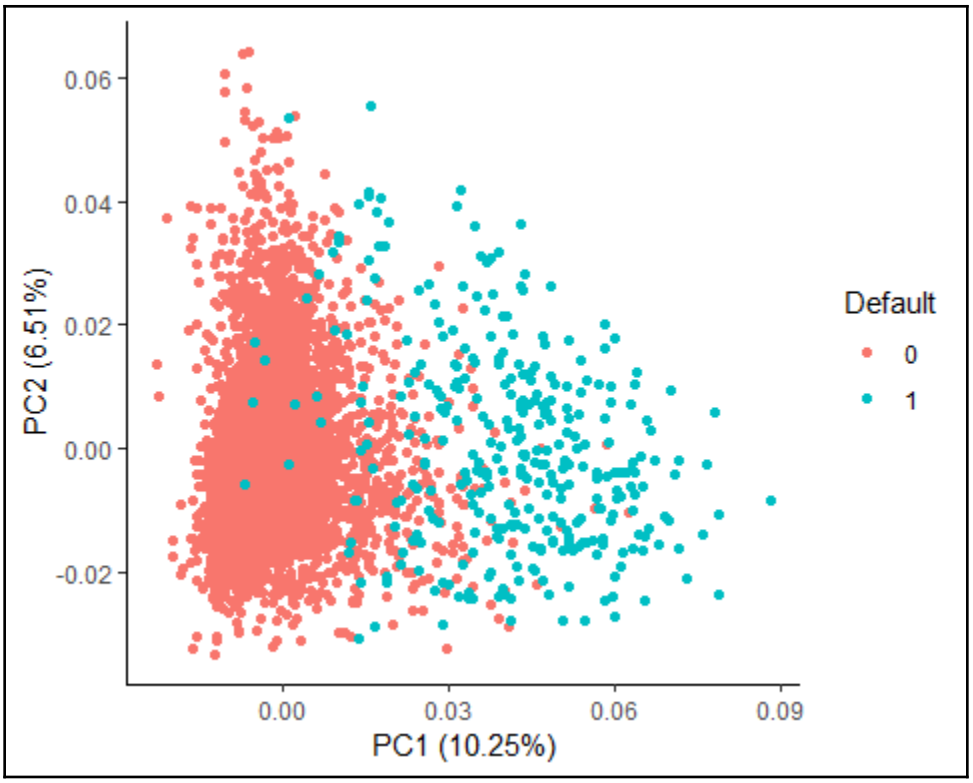
$$argmax_\beta \left(L(\beta|y,x) - \lambda \sum_{k=1}^{K} |\beta_k| \right)$$

$$argmax_\beta \left(L(\beta|y,x) - \lambda_1 \sum_{k=1}^{K} |\beta_k| + \lambda_2 \sum_{k=1}^{K} \beta_k^2 \right)$$

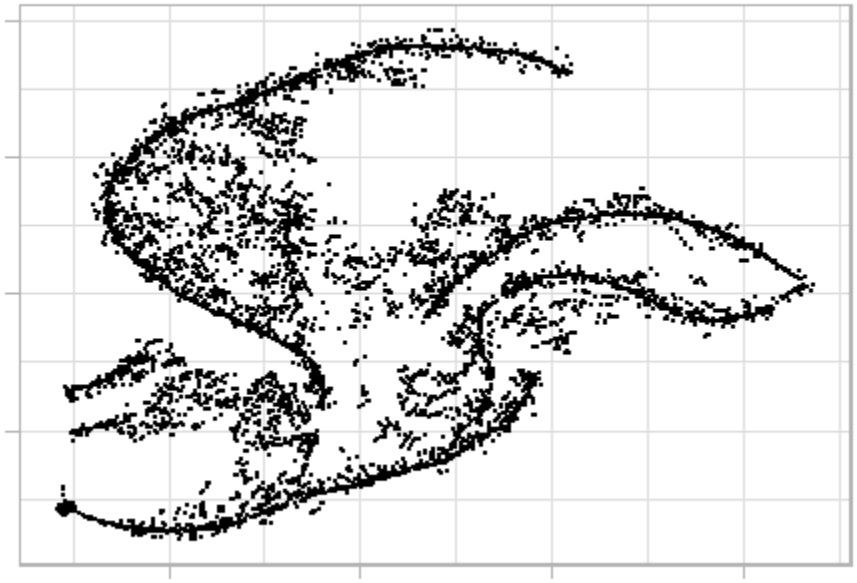



Proportion of Variance Explained
0.00
0.04
0.08
0
20
40
60
80
100
Principal Component


Cumulative Proportion of Variance Explained
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Principal Component

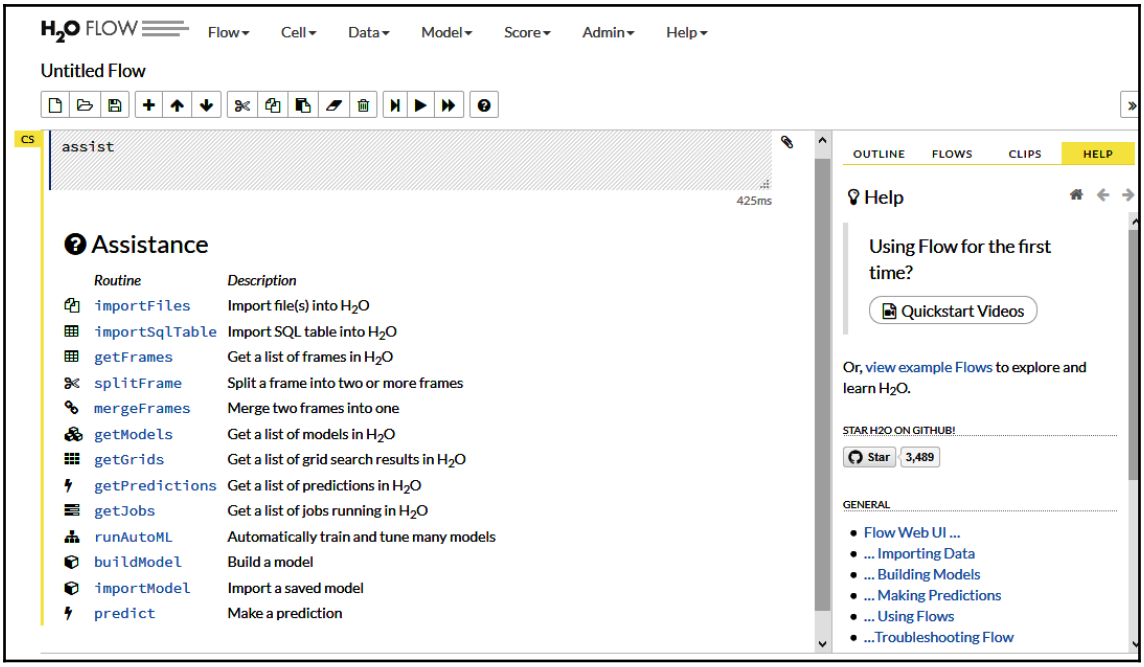

0.06
0.04
0.02
0.00
-0.02
PC2 (6.51%)
0.00
0.03
0.06
0.09
PC1 (10.25%)
Default
0
1

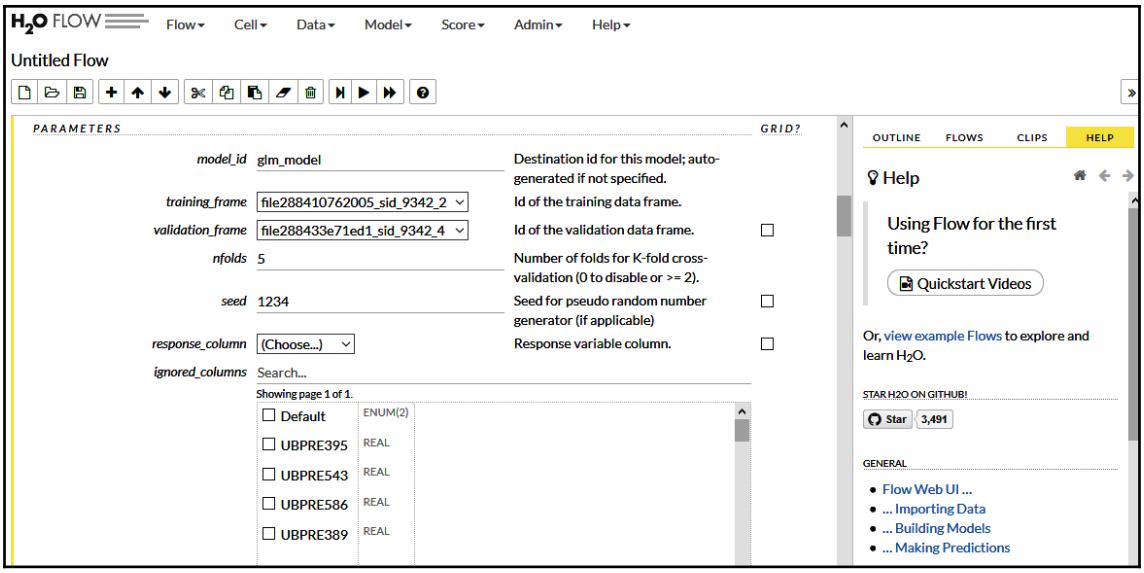
t-SNE

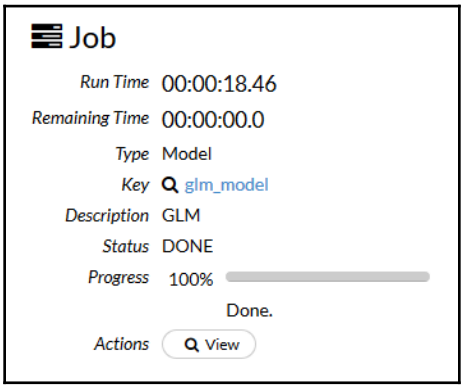


Chapter 05: Predicting Failures of Banks - Multivariate Analysis

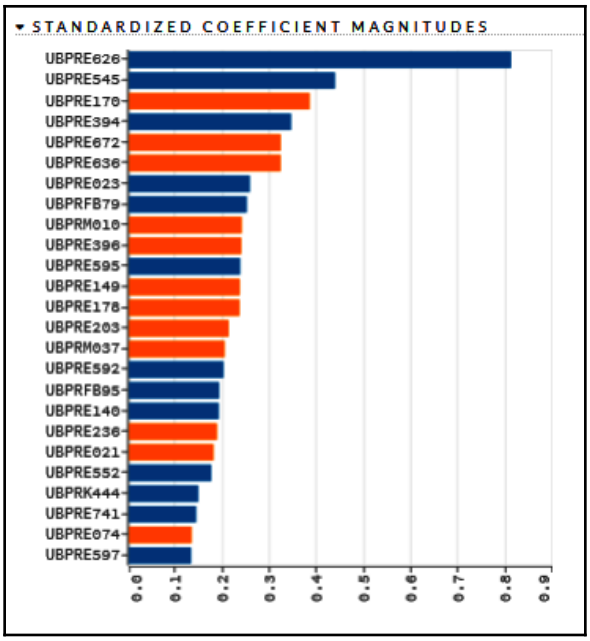
$$L_n\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=\alpha+\sum_{i=1}^{P}\beta_i X_i$$

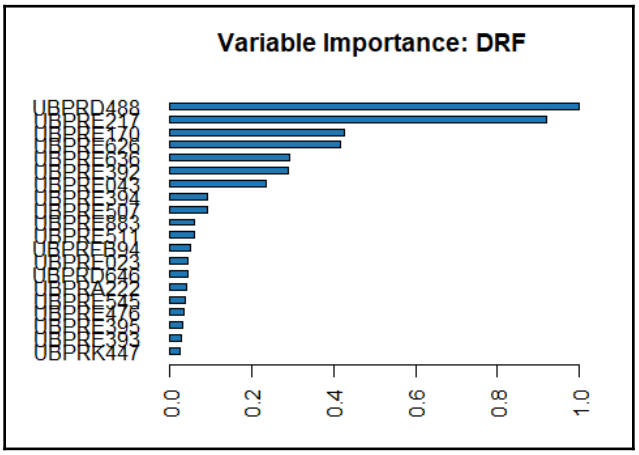


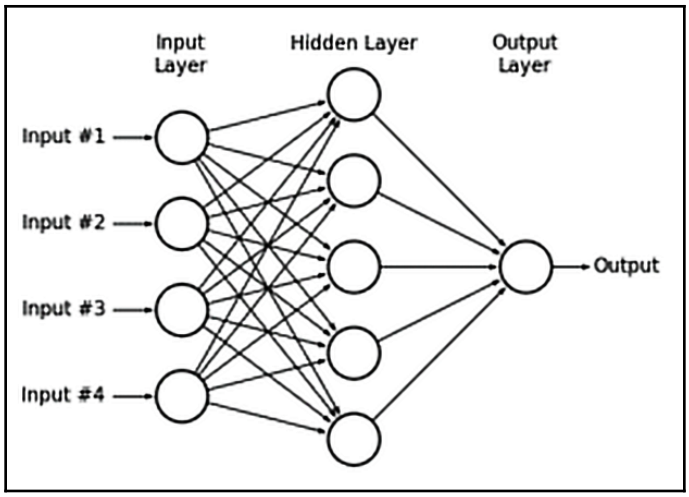


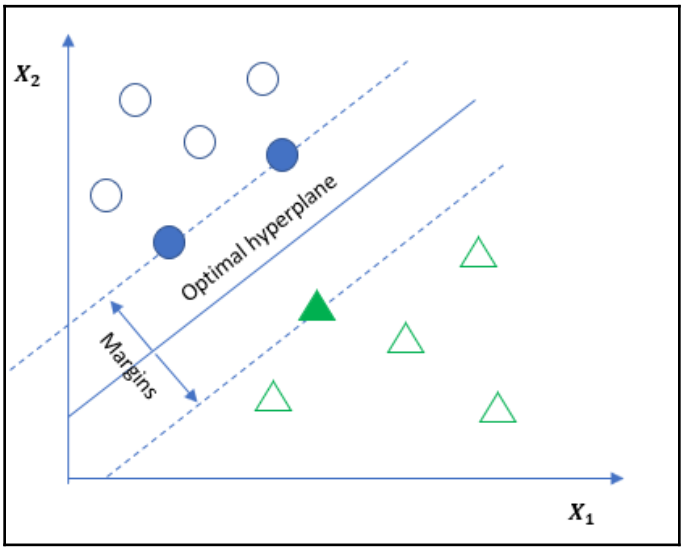


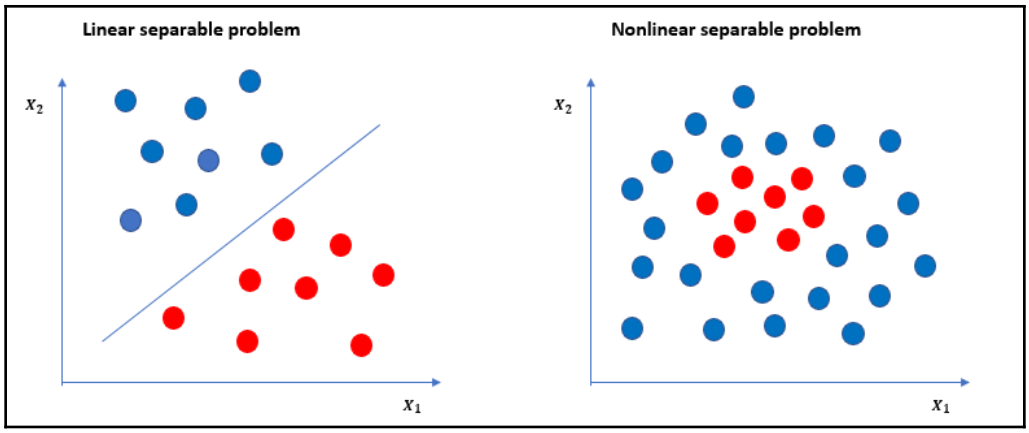
family	binomial
link	logit
regularization	Lasso (lambda = 8.474E-4)
lambda_search	nlambda = 100, lambda.max = 0.1414, lambda.min = 8.474E-4, lambda.1se = 0.003421
number_of_predictors_total	108
number_of_active_predictors	56
number_of_iterations	117
training_frame	train

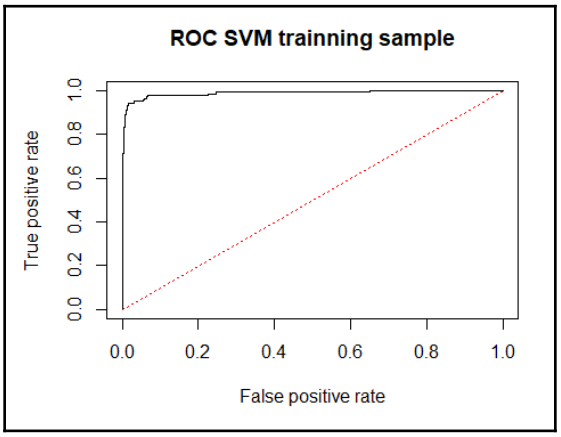

STANDARDIZED COEFFICIENT MAGNITUDES
UBPRE626
UBPRE545
UBPRE170
UBPRE394
UBPRE672
UBPRE636
UBPRE023
UBPRFB79
UBPRM010
UBPRE396
UBPRE595
UBPRE149
UBPRE178
UBPRE203
UBPRM037
UBPRE592
UBPRFB95
UBPRE140
UBPRE236
UBPRE021
UBPRE552
UBPRK444
UBPRE741
UBPRE074
UBPRE597
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


Variable Importance: DRF
UBPRD488
UBPRE217
UBPRE170
UBPRE626
UBPRE636
UBPRE396
UBPRE043
UBPRE394
UBPRE507
UBPRE883
UBPRE511
UBPRFB94
UBPRE023
UBPRD646
UBPRA222
UBPRE545
UBPRE476
UBPRE395
UBPRE393
UBPRK447
0.0
0.2
0.4
0.6
0.8
1.0


Input Layer
Hidden Layer
Output Layer
Input #1
Input #2
Input #3
Input #4
Output


X_2
Optimal hyperplane
Margins
X_1


Linear separable problem
x_2
x_1
Nonlinear separable problem
x_2
x_1

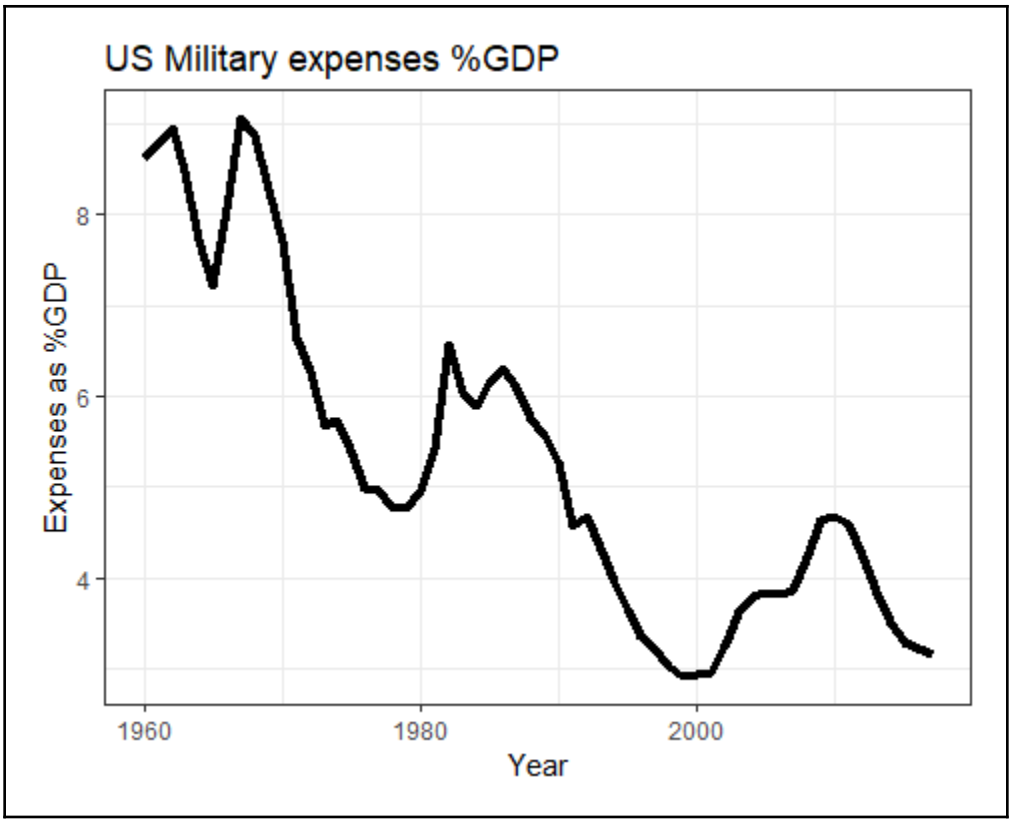

ROC SVM trainning sample
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

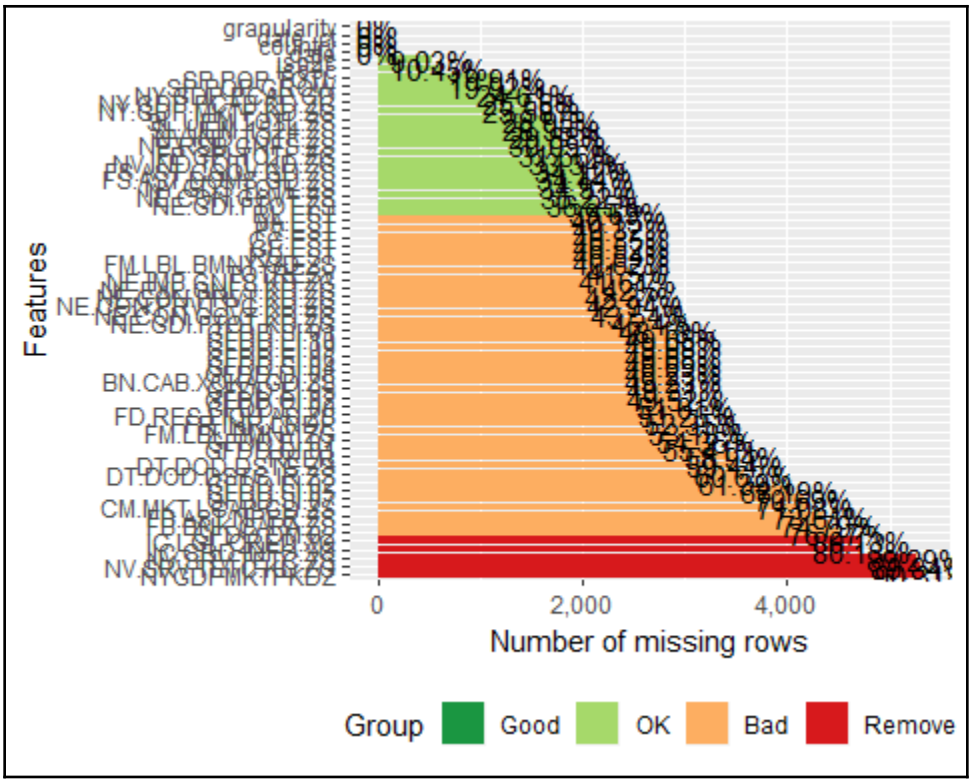
Chapter 06: Visualizing Economic Problems in the European Union

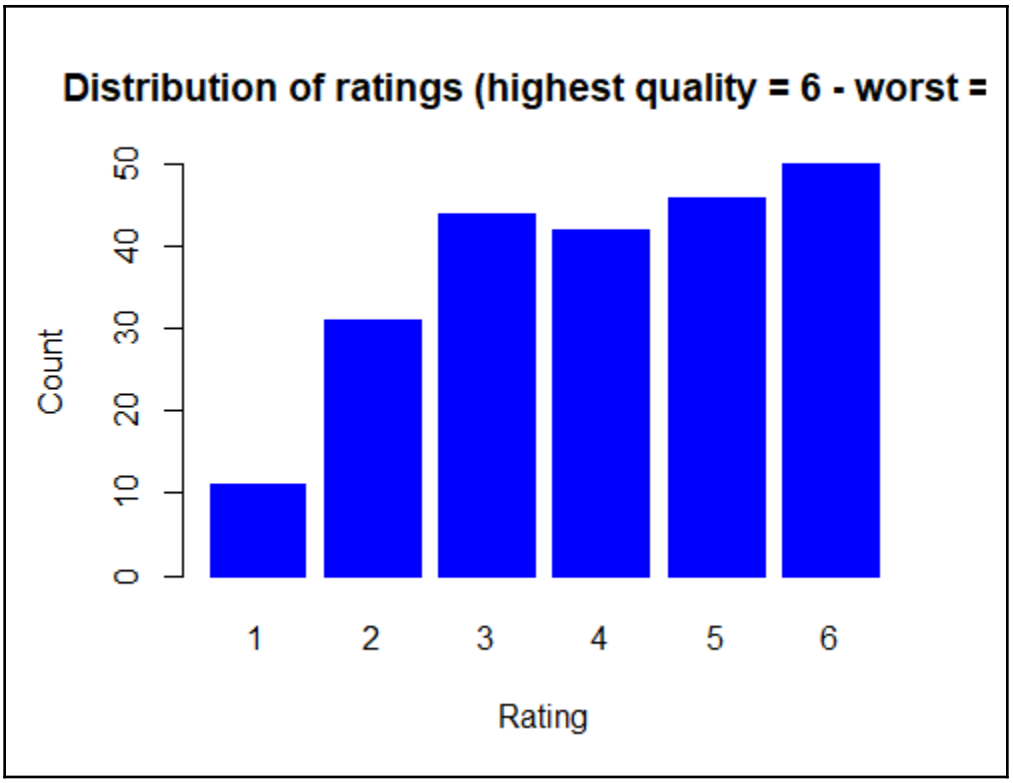
General summary of the opinions reflected by our ratings

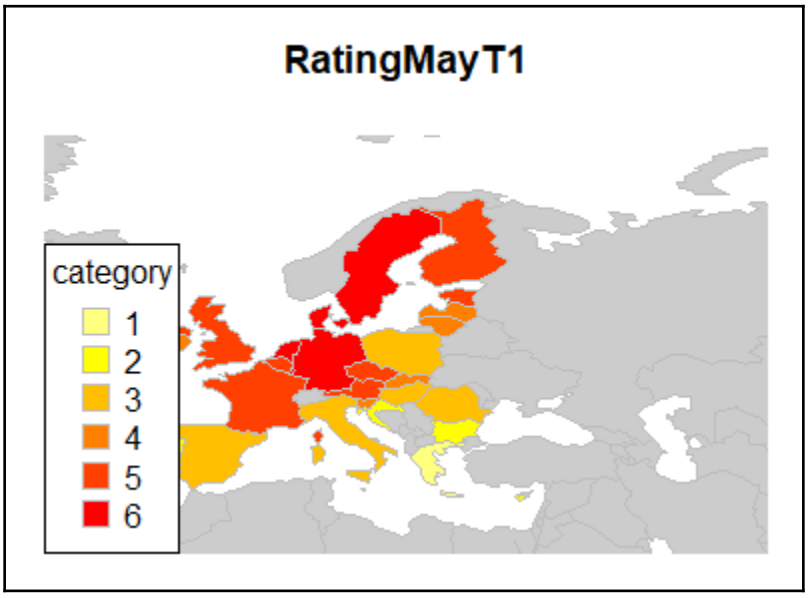
Investment Grade	**AAA**	Extremely strong capacity to meet financial commitments. Highest rating
	AA	Very strong capacity to meet financial commitments
	A	Strong capacity to meet financial commitments, but somewhat susceptible to adverse economic conditions and changes in circumstances
	BBB	Adequate capacity to meet financial commitments, but more subject to adverse economic conditions
	BBB-	Considered lowest investment-grade by market participants
Speculative Grade	**BB+**	Considered highest speculative-grade by market participants
	BB	Less vulnerable in the near-term but faces major ongoing uncertainties to adverse business, financial and economic conditions
	B	More vulnerable to adverse business, financial and economic conditions but currently has the capacity to meet financial commitments
	CCC	Currently vulnerable and dependent on favorable business, financial and economic conditions to meet financial commitments
	CC	Highly vulnerable; default has not yet occurred, but is expected to be a virtual certainty
	C	Currently highly vulnerable to non-payment, and ultimate recovery is expected to be lower than that of higher rated obligations
	D	Payment default on a financial commitment or breach of an imputed promise; also used when a bankruptcy petition has been filed or similar action taken

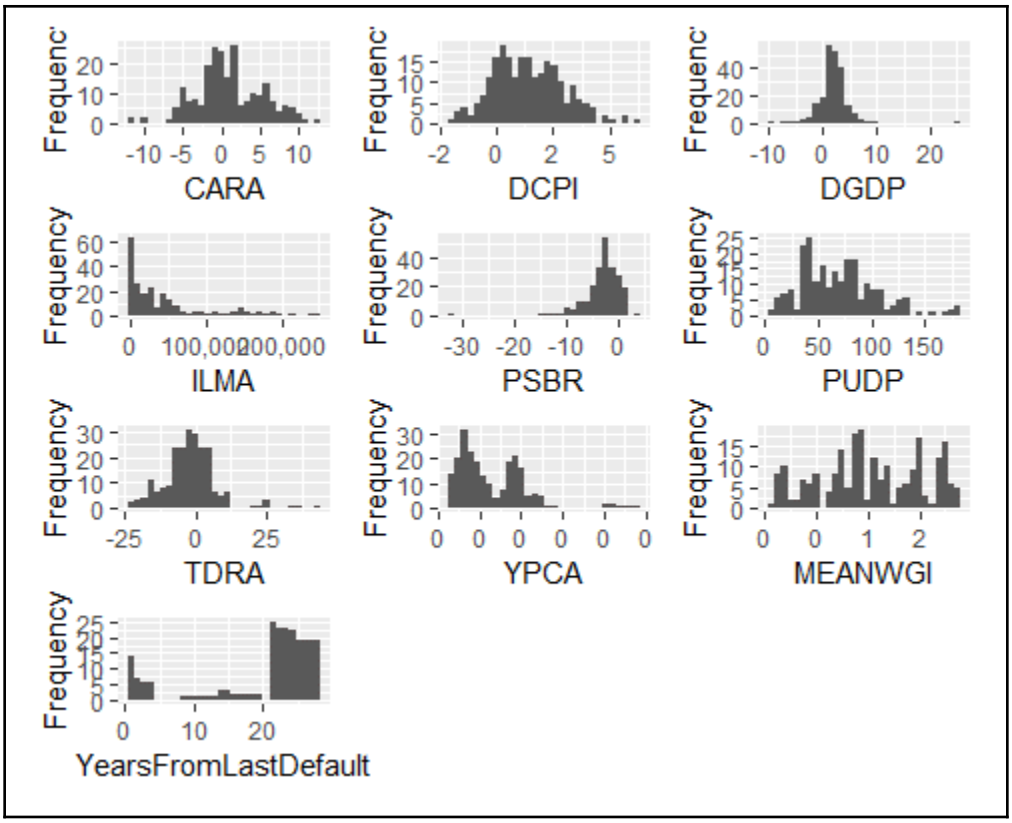
Ratings from 'AA' to 'CCC' may be modified by the addition of a plus (+) or minus (-) sign to show relative standing within the major rating categories.

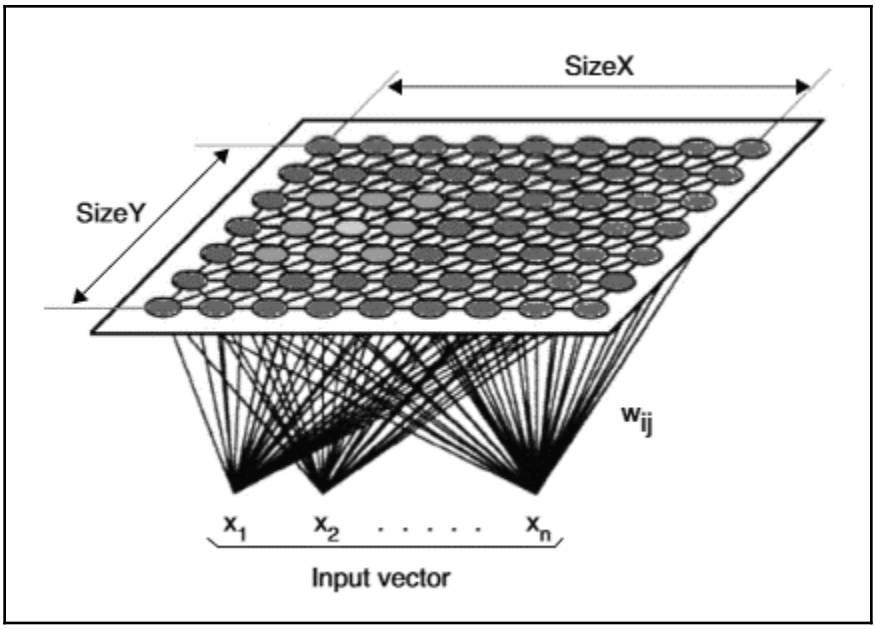

US Military expenses %GDP
Expenses as %GDP
8
6
4
1960
1980
2000
Year

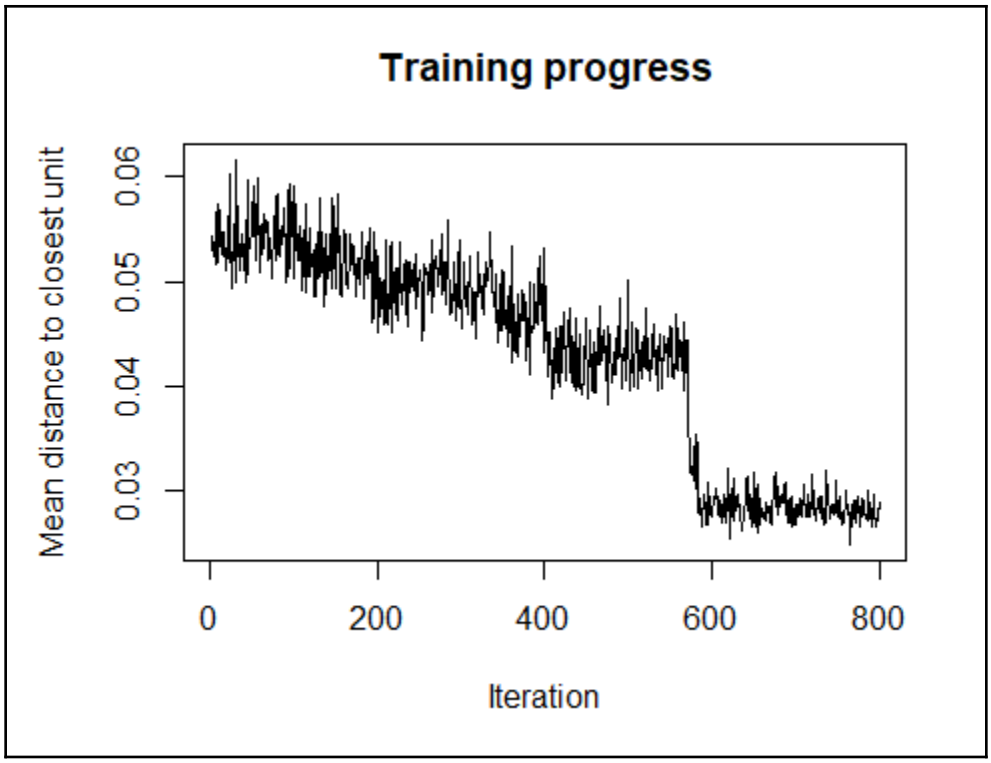

granularity
Features
0
2,000
4,000
Number of missing rows
Group
Good
OK
Bad
Remove

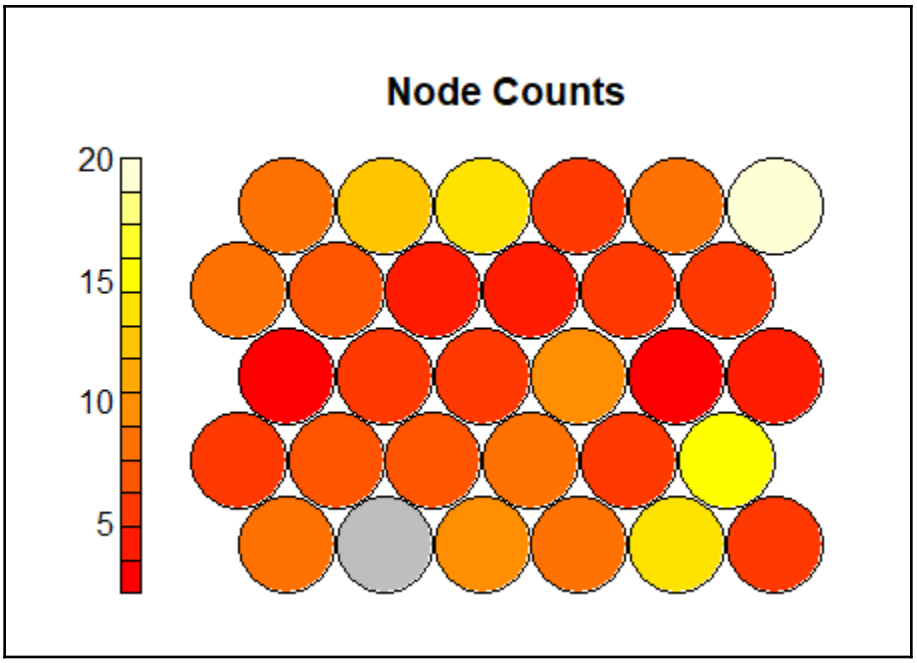

Distribution of ratings (highest quality = 6 - worst =
Count
0
10
20
30
40
50
1
2
3
4
5
6
Rating

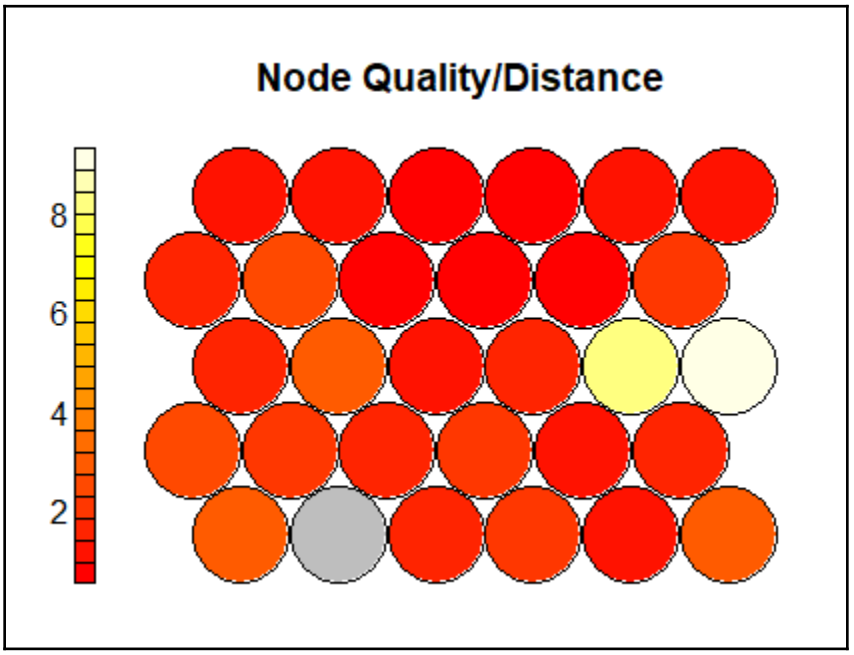

RatingMayT1
category
1
2
3
4
5
6

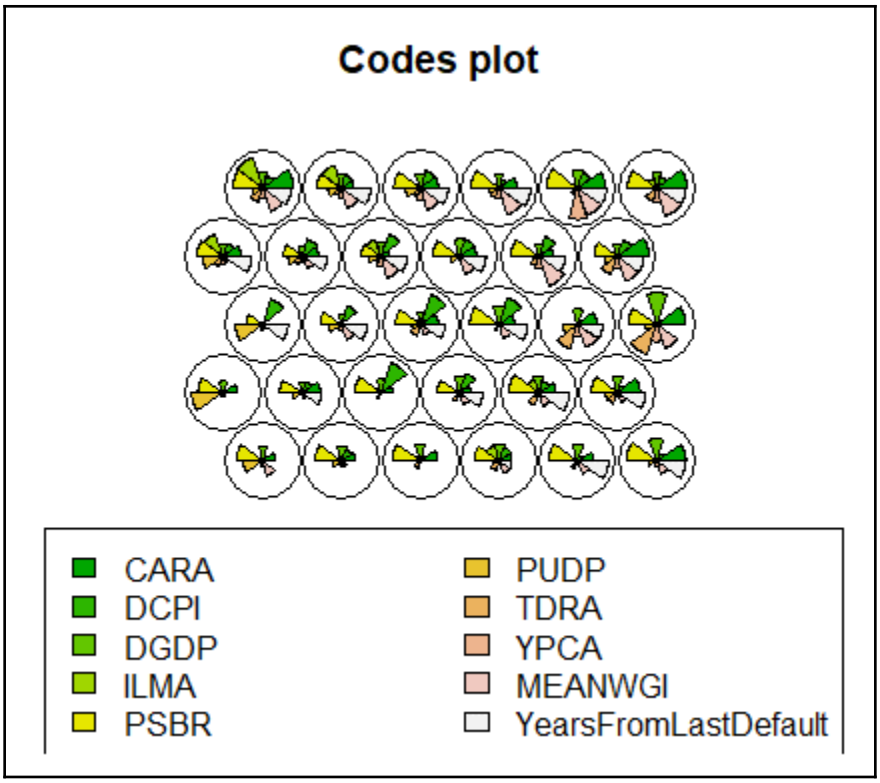

Frequency
CARA
DCPI
DGDP
ILMA
PSBR
PUDP
TDRA
YPCA
MEANWGI
YearsFromLastDefault

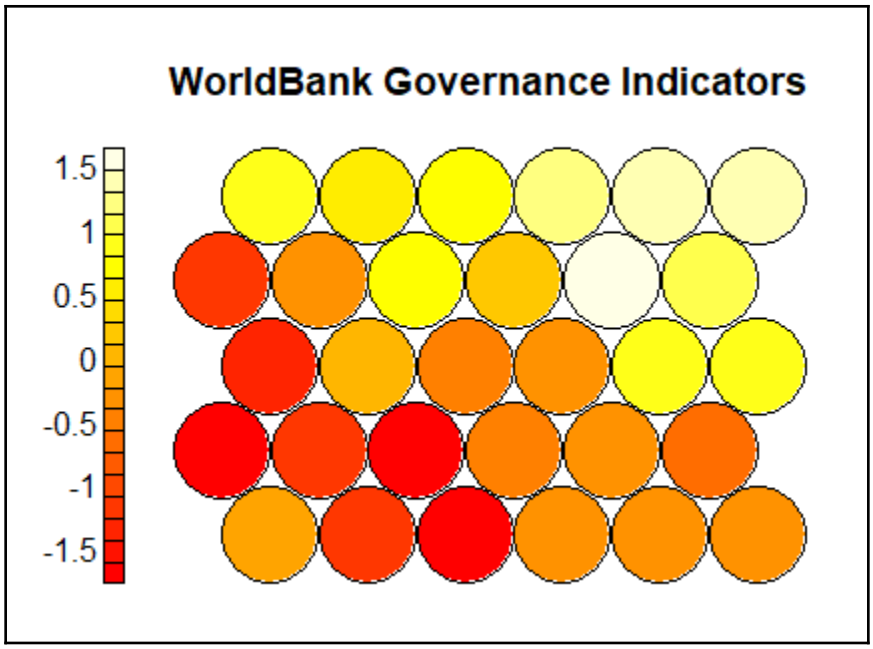

SizeX
SizeY
w_{ij}
x_1 x_2 . . . . . x_n
Input vector

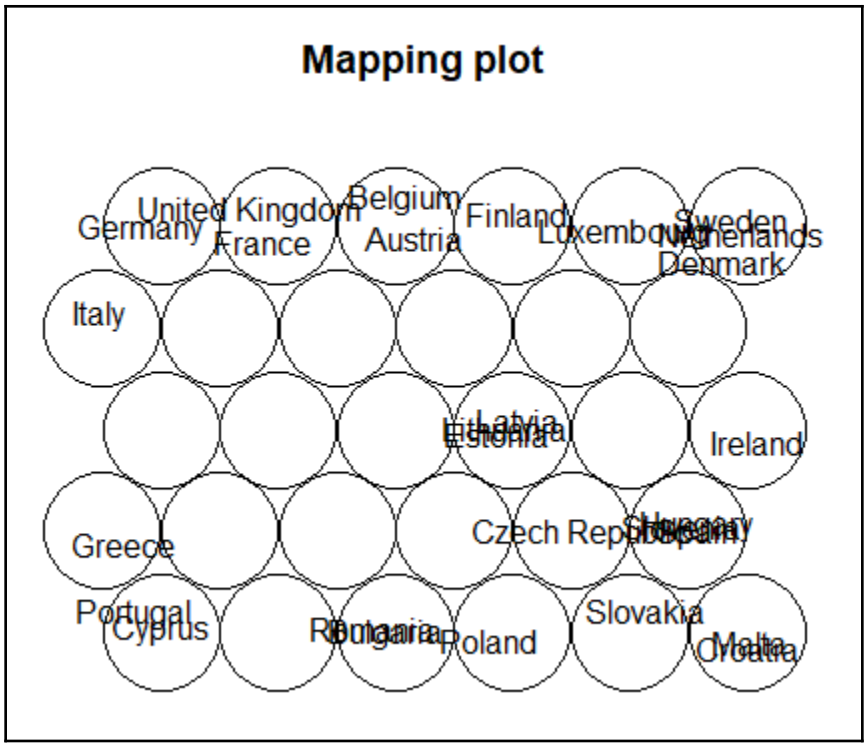

Training progress
Mean distance to closest unit
0.03
0.04
0.05
0.06
0
200
400
600
800
Iteration

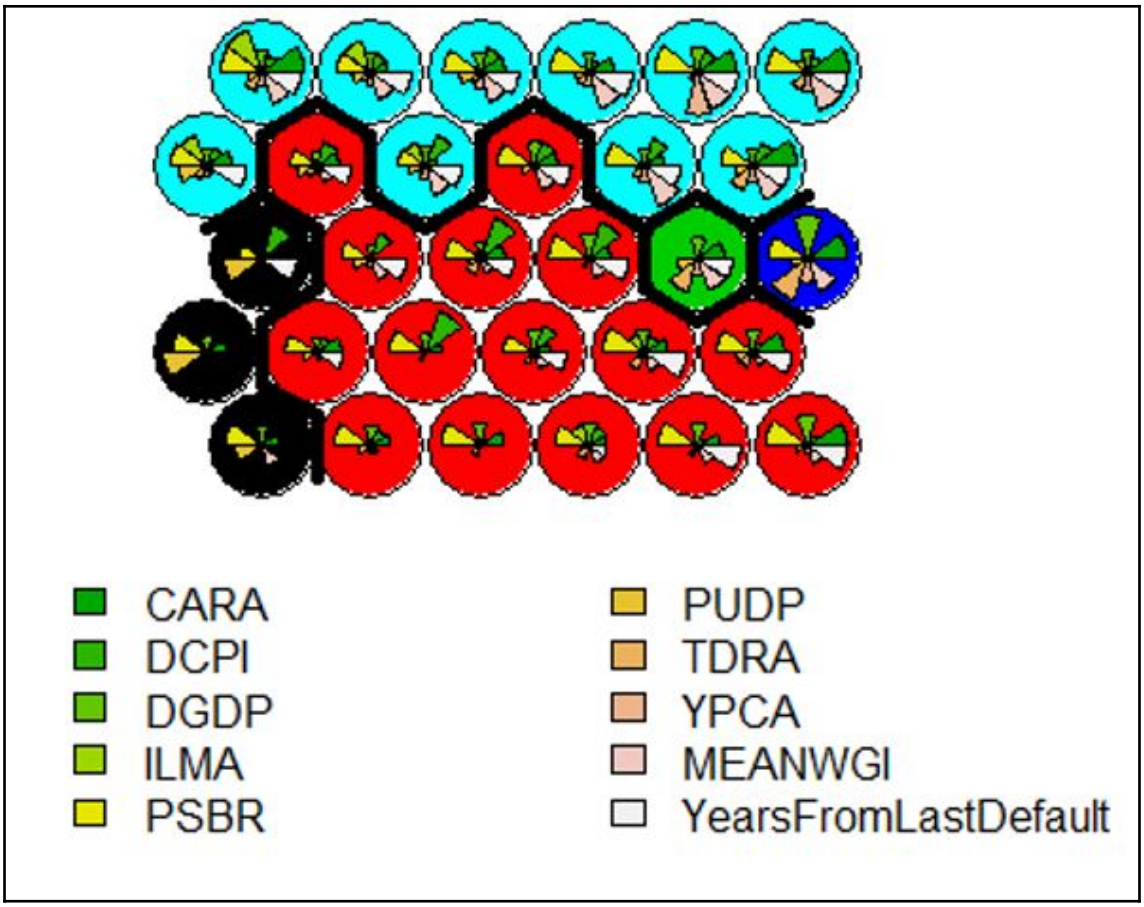

Node Counts
20
15
10
5


Node Quality/Distance
8
6
4
2

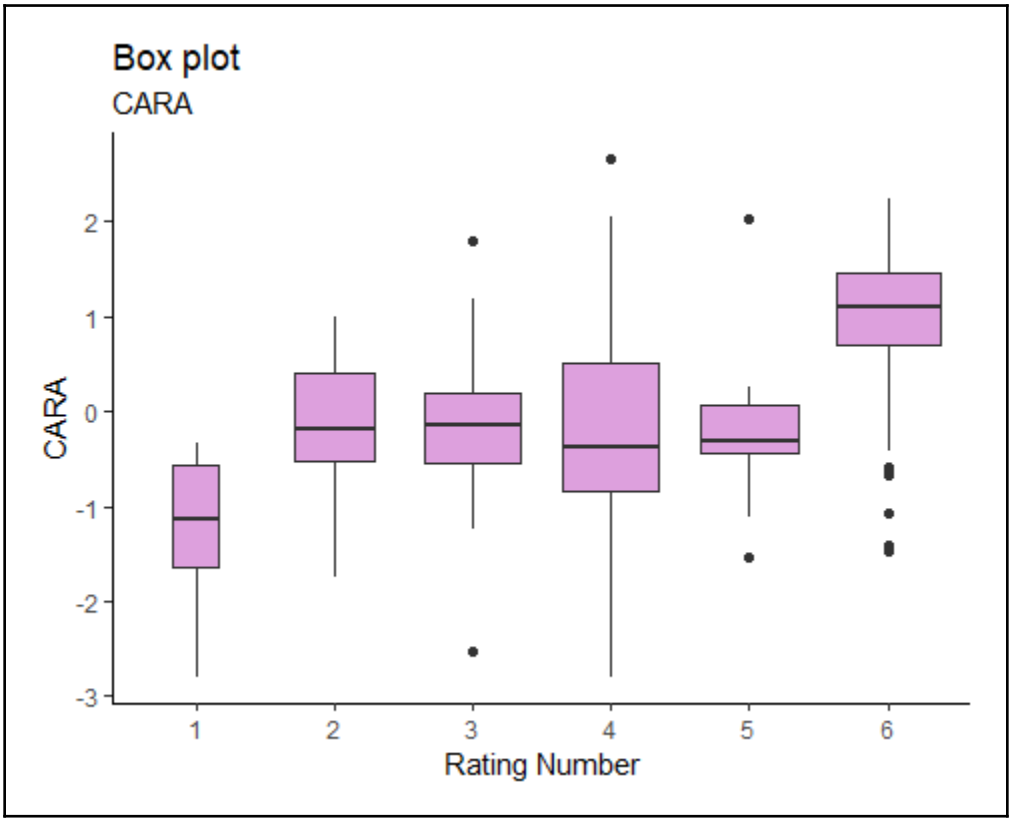

Codes plot
CARA
DCPI
DGDP
ILMA
PSBR
PUDP
TDRA
YPCA
MEANWGI
YearsFromLastDefault

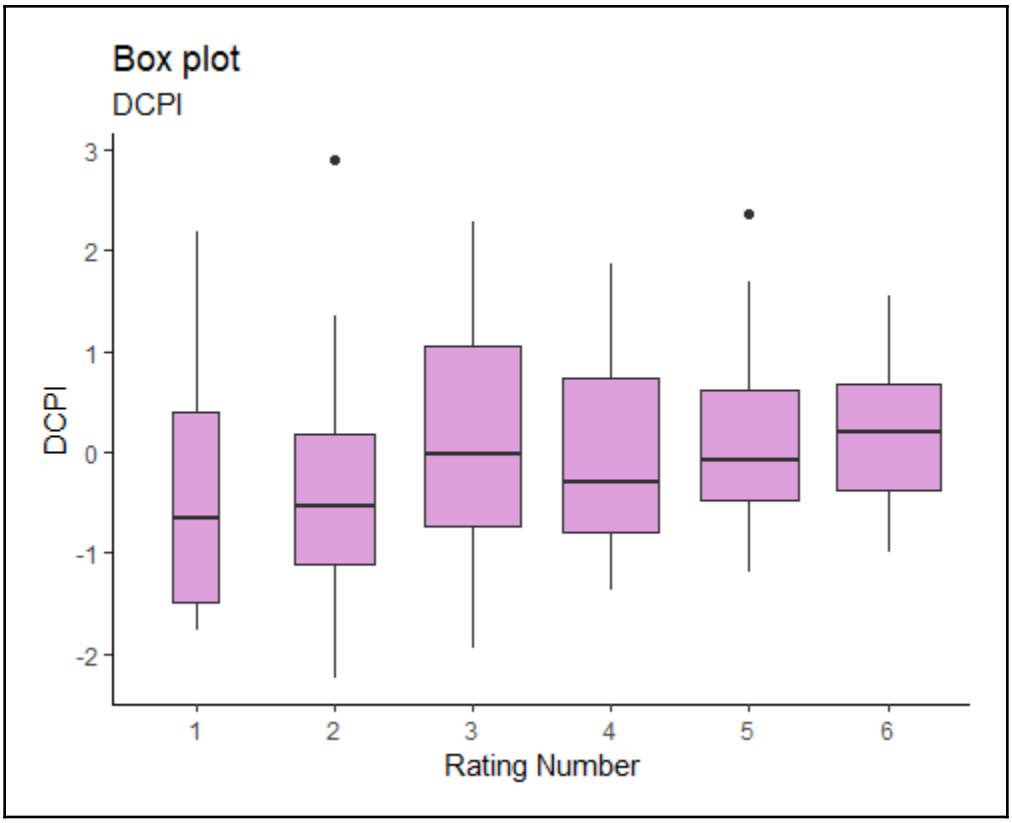

WorldBank Governance Indicators
1.5
1
0.5
0
-0.5
-1
-1.5

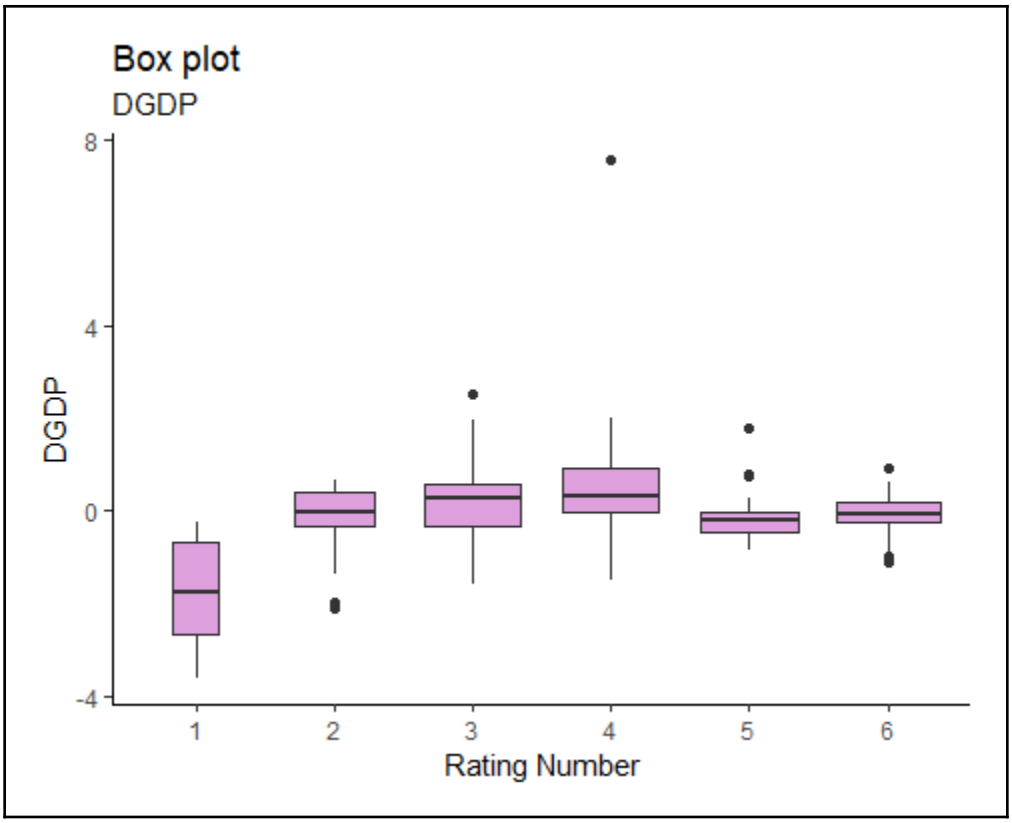

Mapping plot
United Kingdom
Belgium
Germany
France
Austria
Finland
Luxembourg
Sweden
Netherlands
Denmark
Italy
Latvia
Ireland
Greece
Czech Republic
Portugal
Cyprus
Slovakia
Poland
Malta
Croatia

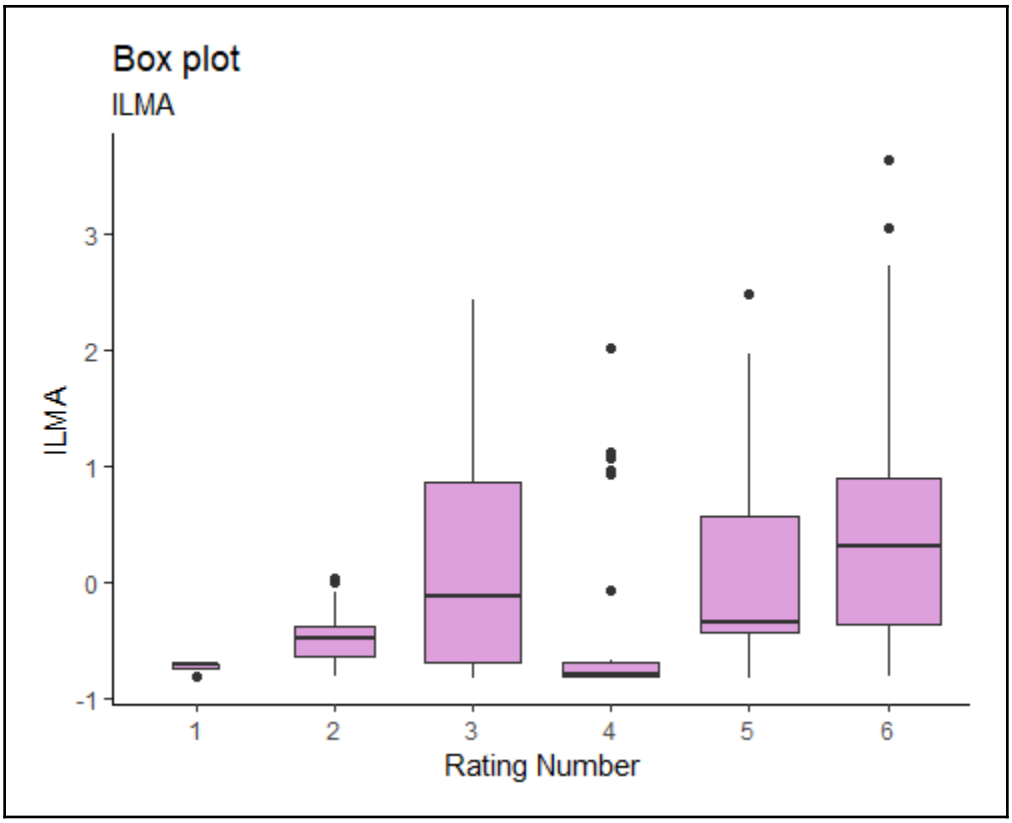

CARA
DCPI
DGDP
ILMA
PSBR
PUDP
TDRA
YPCA
MEANWGI
YearsFromLastDefault

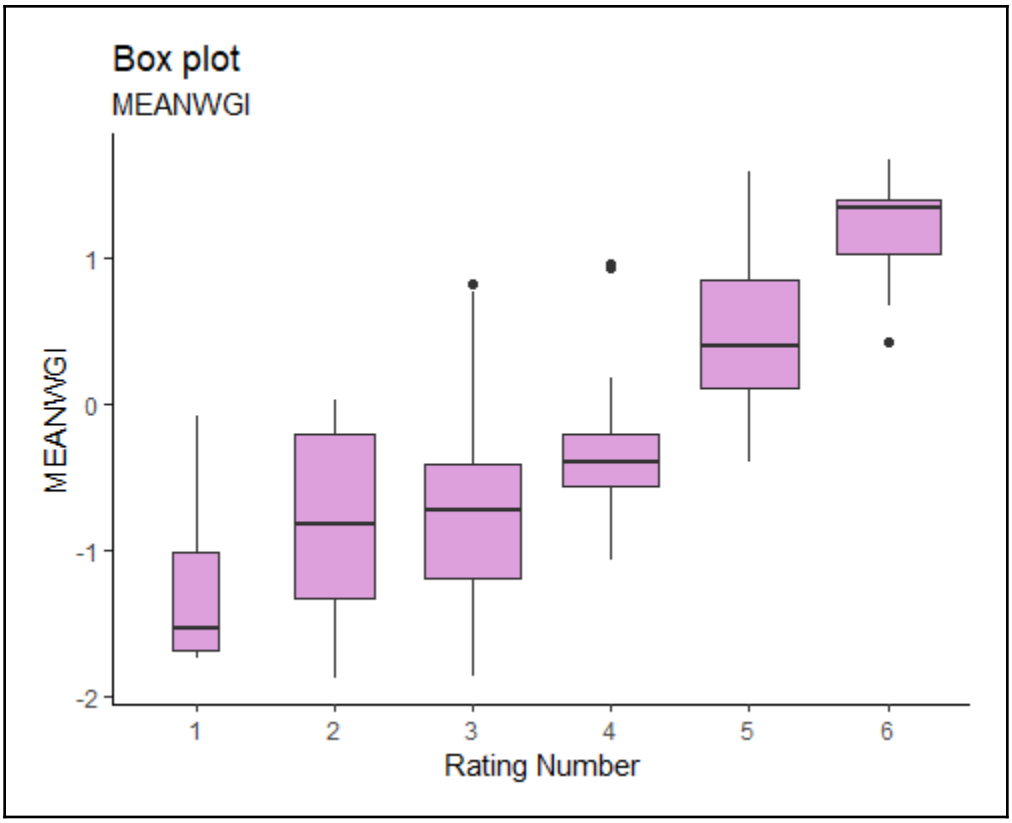
Chapter 07: Sovereign Crisis - NLP and Topic Modeling

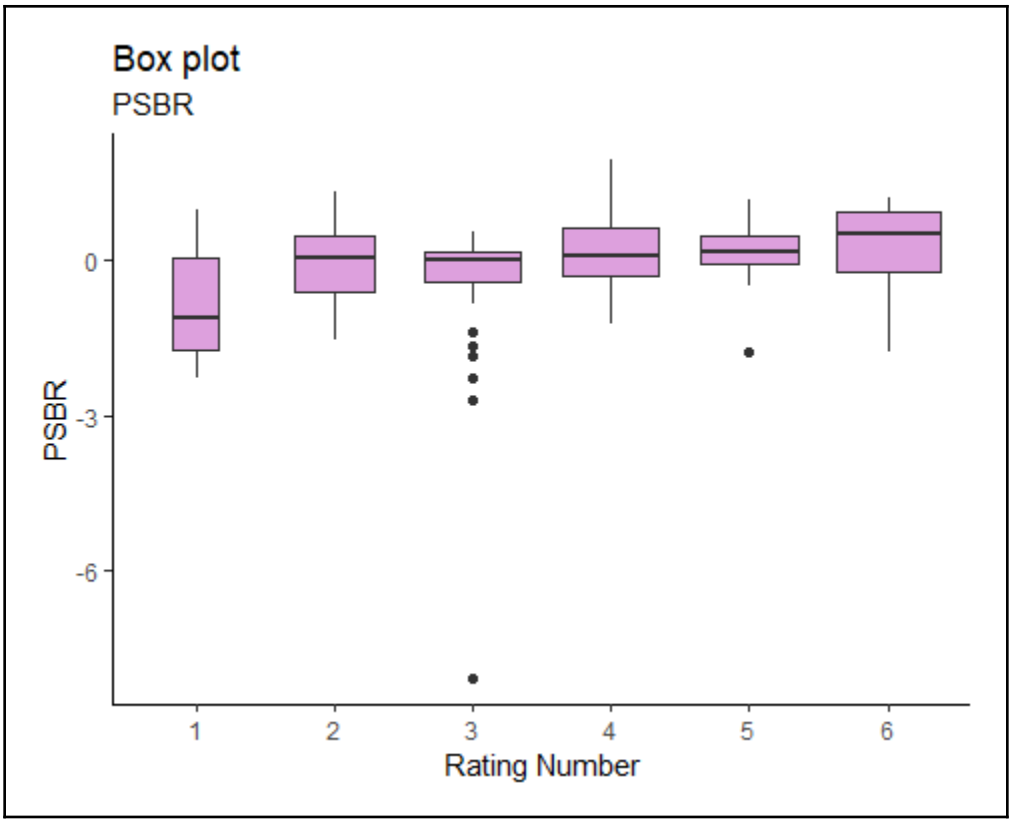


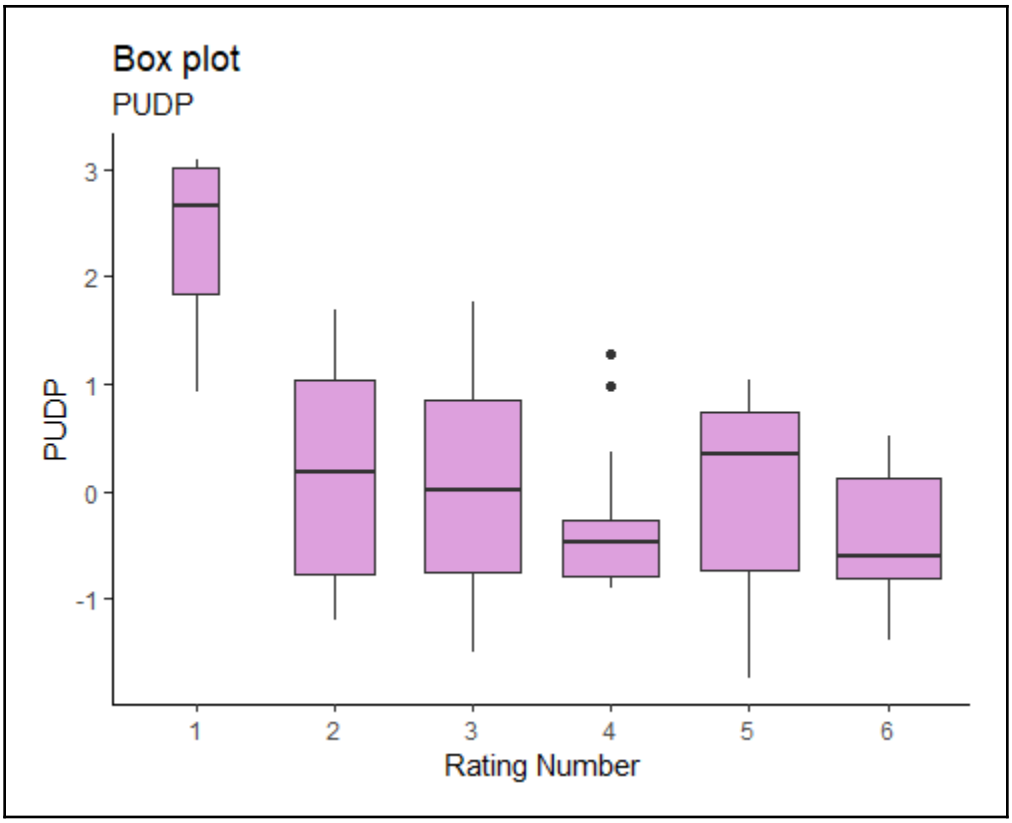

Box plot
DCPI
3
2
1
0
-1
-2
DCPI
1
2
3
4
5
6
Rating Number

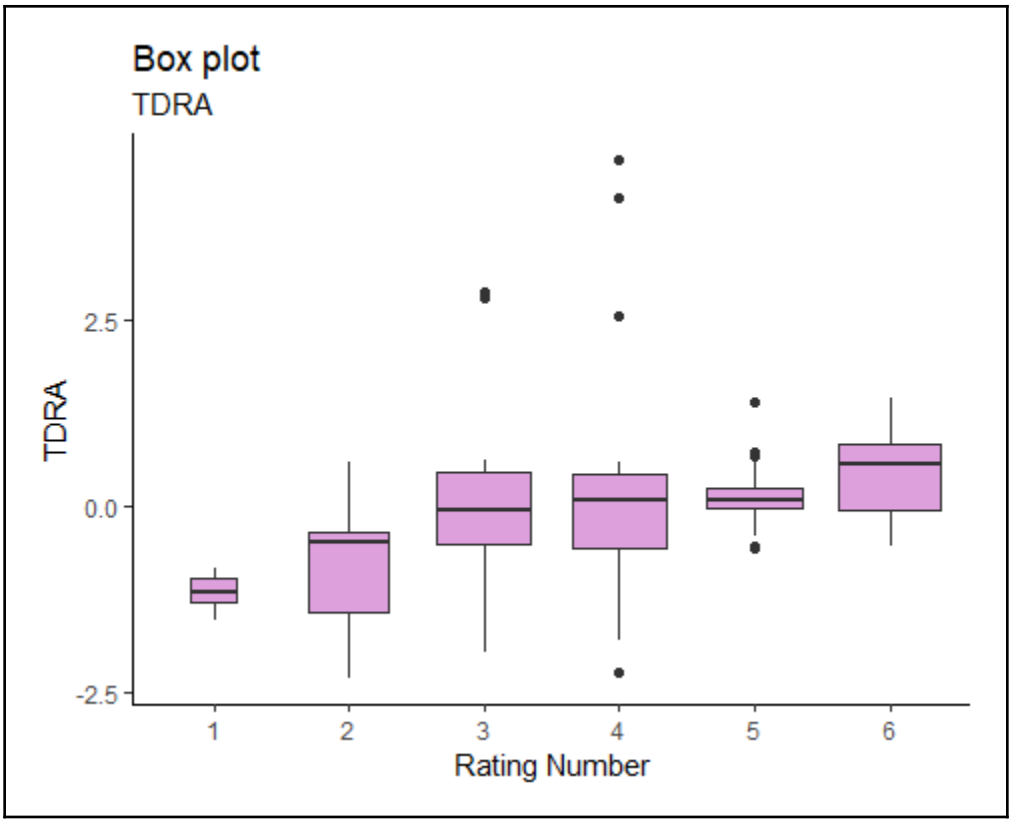

Box plot
DGDP
8
4
0
-4
DGDP
1
2
3
4
5
6
Rating Number

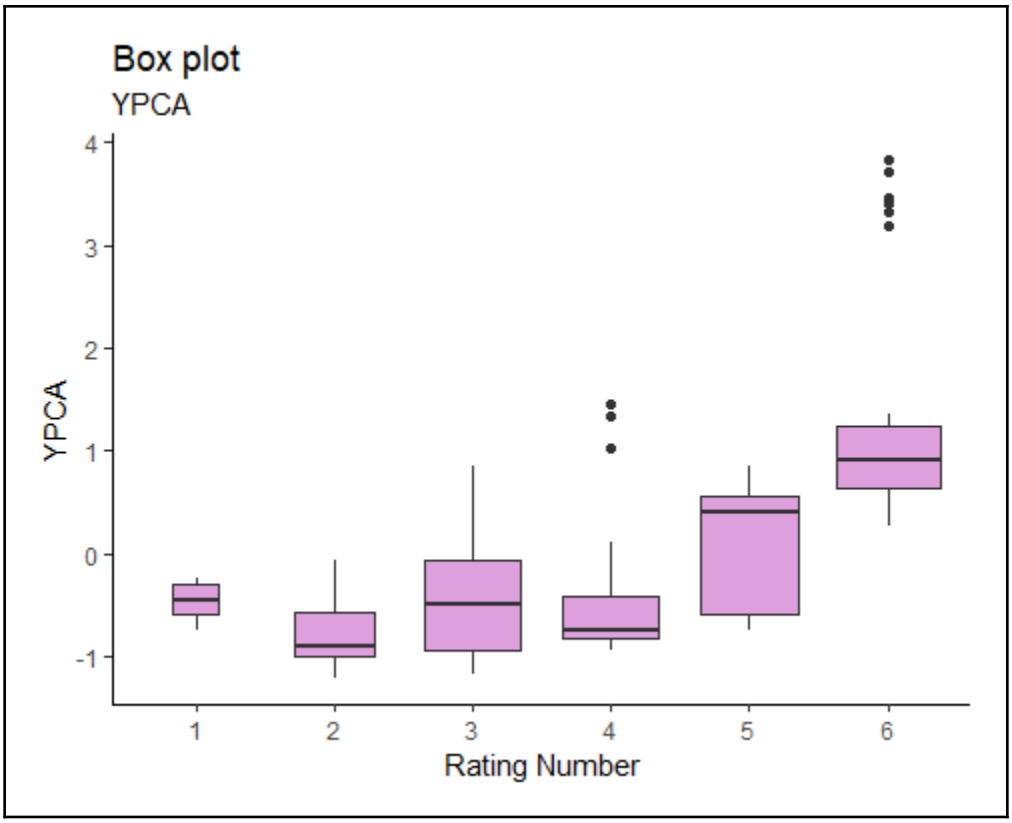

Box plot
ILMA
ILMA
3
2
1
0
-1
1
2
3
4
5
6
Rating Number

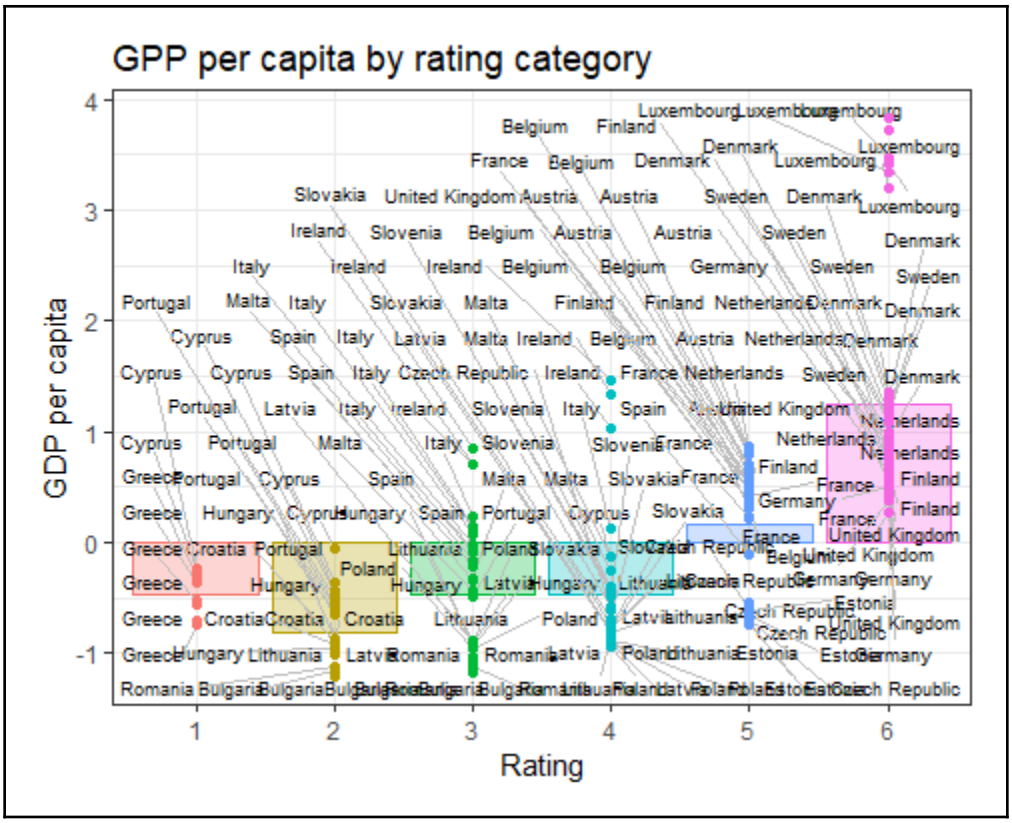

Box plot
MEANWGI
MEANWGI
1
0
-1
-2
1
2
3
4
5
6
Rating Number

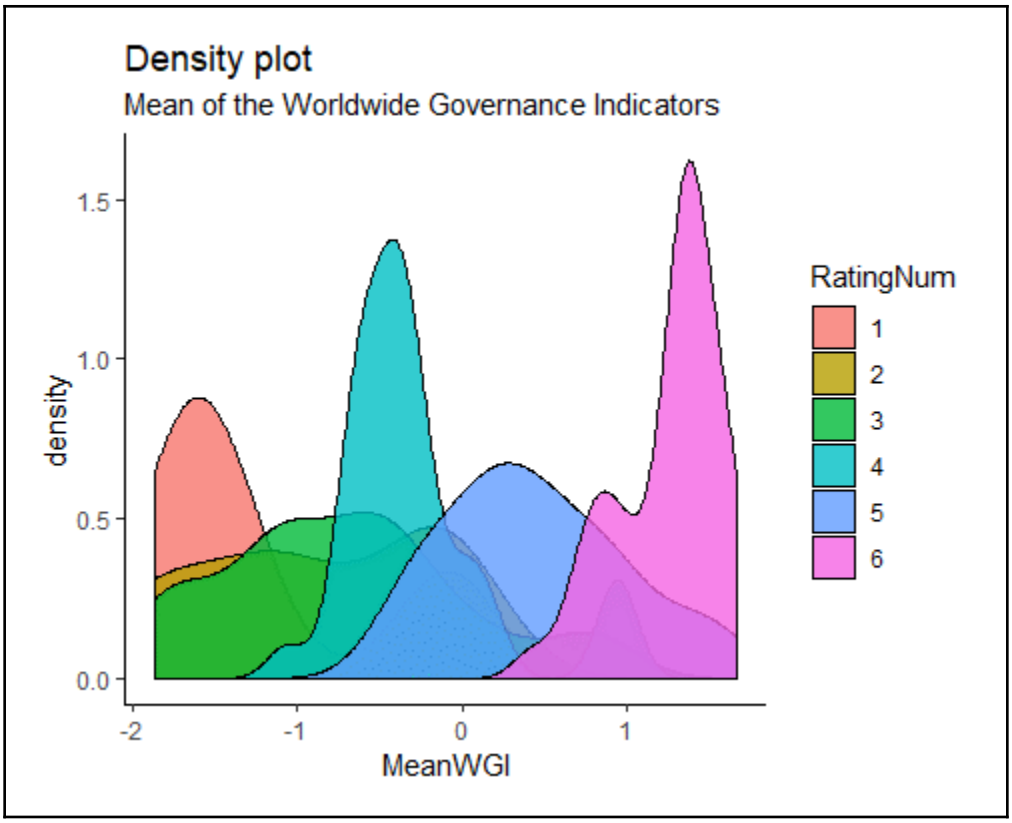

Box plot
PSBR
0
-3
-6
PSBR
1
2
3
4
5
6
Rating Number

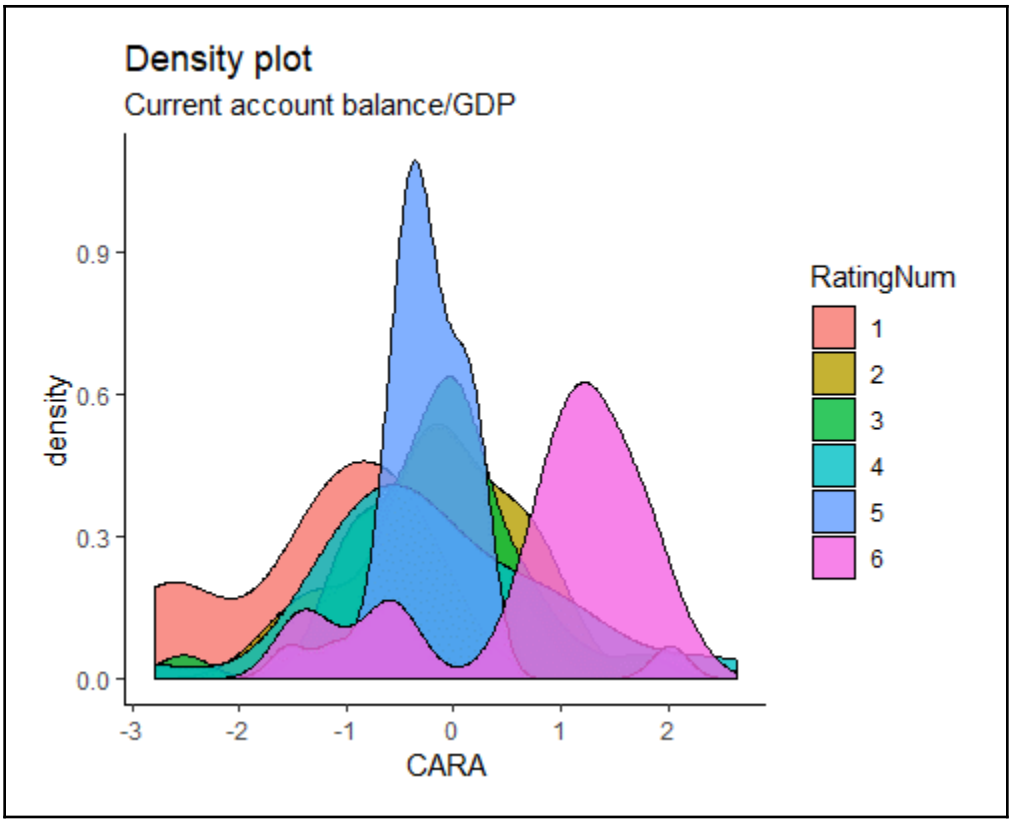

Box plot
PUDP
PUDP
3
2
1
0
-1
1
2
3
4
5
6
Rating Number

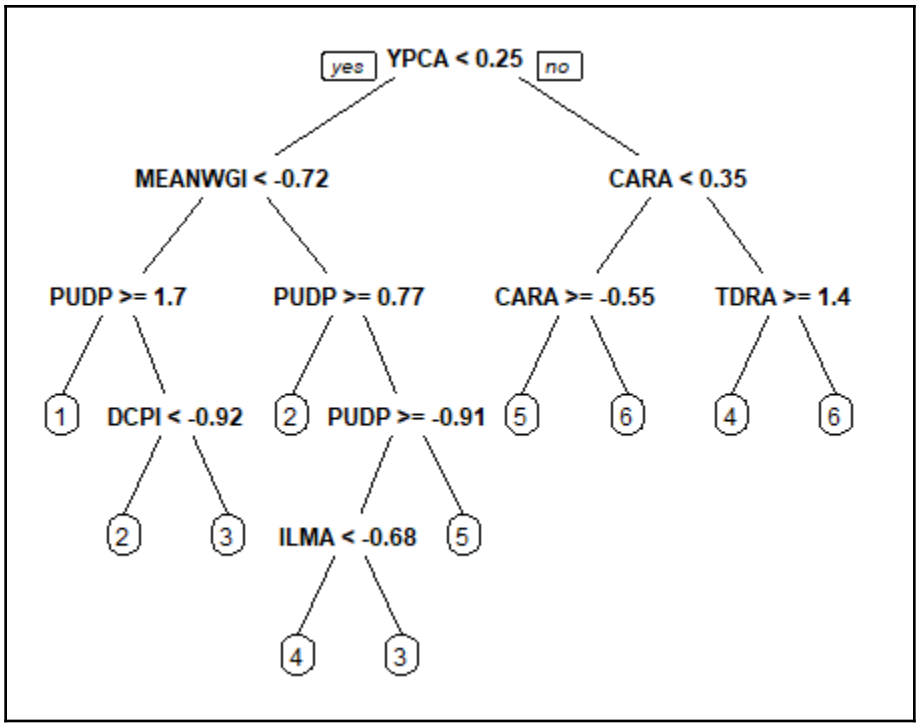

Box plot
TDRA
TDRA
2.5
0.0
-2.5
1
2
3
4
5
6
Rating Number

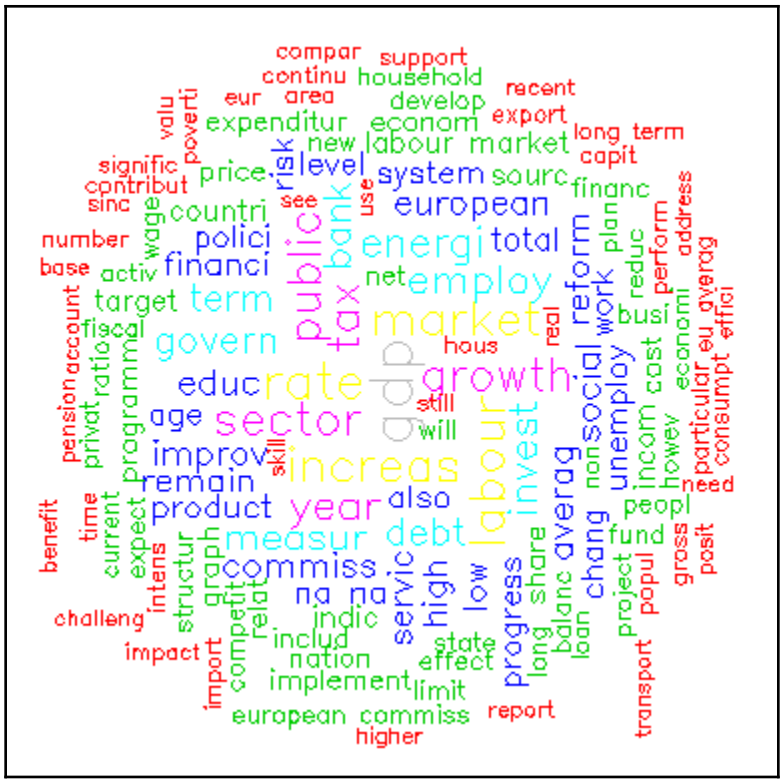

Box plot
YPCA
YPCA
4
3
2
1
0
-1
1
2
3
4
5
6
Rating Number

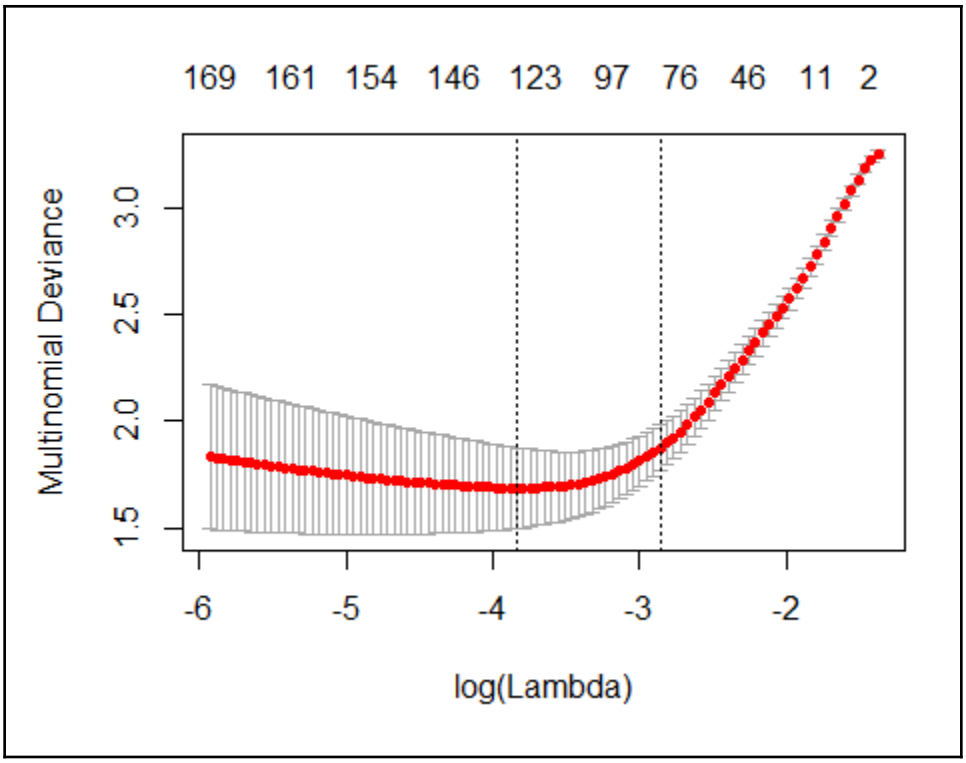

GPP per capita by rating category
GDP per capita
Rating
4
3
2
1
0
-1
1
2
3
4
5
6


Density plot
Mean of the Worldwide Governance Indicators
density
1.5
1.0
0.5
0.0
-2
-1
0
1
MeanWGI
RatingNum
1
2
3
4
5
6


Density plot
Current account balance/GDP
density
0.0
0.3
0.6
0.9
-3
-2
-1
0
1
2
CARA
RatingNum
1
2
3
4
5
6


yes
YPCA < 0.25
no
MEANWGI < -0.72
CARA < 0.35
PUDP >= 1.7
PUDP >= 0.77
CARA >= -0.55
TDRA >= 1.4
1
DCPI < -0.92
2
PUDP >= -0.91
5
6
4
6
2
3
ILMA < -0.68
5
4
3


compar
support
continu
household
recent
eur
area
develop
valu
poverti
expenditur
econom
export
long
term
new
labour
market
capit
signific
price
risk
level
system
sourc
contribut
financ
sinc
see
use
european
address
countri
wage
number
polici
public
bank
energi
total
plan
reform
perform
reduc
base
activ
financi
net
employ
work
average
target
term
tax
market
busi
eu
effici
account
fiscal
govern
real
hous
social
cast
economi
ratio
educ
rate
gdp
growth
unemploy
particular
consumpt
pension
privat
programm
age
sector
still
will
labour
invest
incom
howev
improv
skill
increas
non
need
remain
also
average
peopl
benefit
time
current
expect
product
year
debt
fund
gross
posit
intens
structur
graph
measur
commiss
share
chang
project
popul
challeng
competit
relat
na
na
servic
high
low
progress
balanc
loan
indic
long
impact
import
includ
state
nation
effect
implement
limit
transport
european
commiss
report
higher


169 161 154 146 123 97 76 46 11 2
Multinomial Deviance
1.5
2.0
2.5
3.0
-6
-5
-4
-3
-2
log(Lambda)

.