Centre d'Immunologie et des Maladies Infectieuses

Inserm
Institut national de la santé et de la recherche médicale

SORBONNE UNIVERSITÉ

Feature Selector App

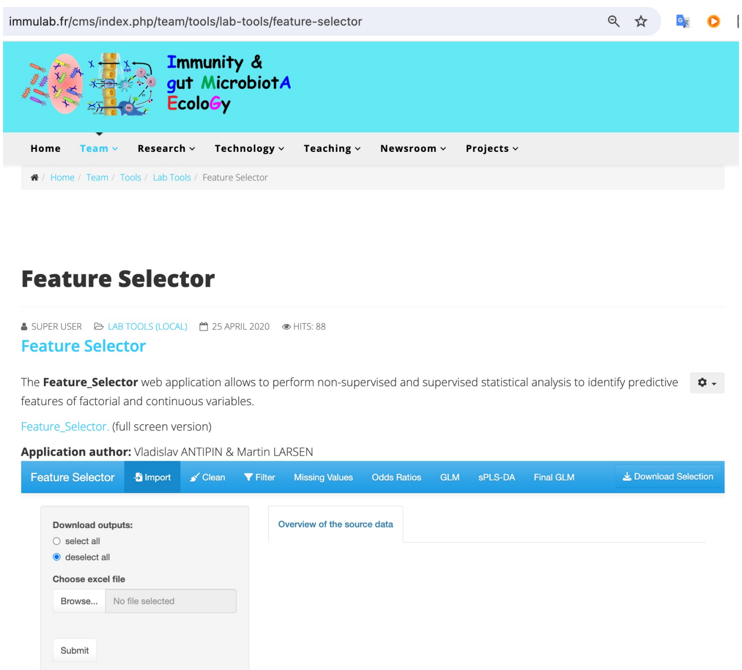
by Vladislav Antipin and Martin Larsen

https://www.immulab.fr/cms/index.php/team/tools/lab-tools/feature-selector

Team: Immunity & microbiota ecology
www.immulab.fr
INSERM UMR-S1135, Sorbonne University, Paris, France

1

1



2

Glossary

- Variable = Parameter = Dimension (= Column)
- Feature = Predictor variable
- Response variable = Outcome Variable = variable to be predicted
- Observation = Sample = Point (= Row)

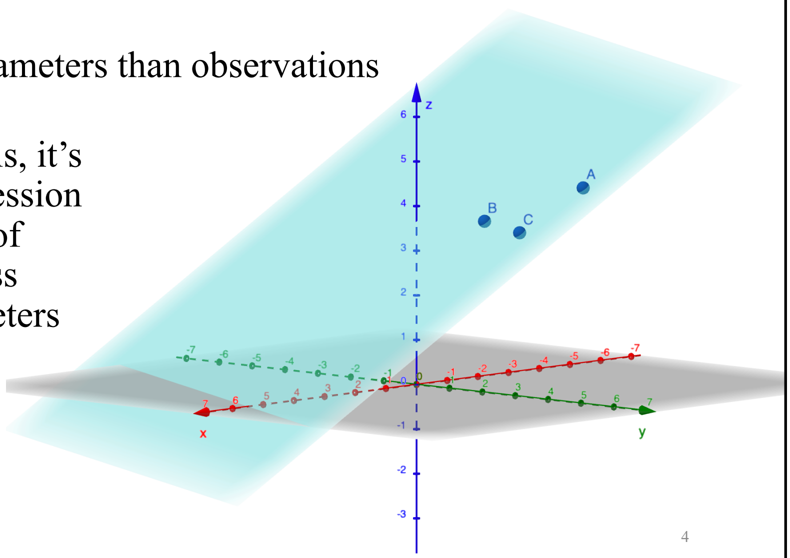
3

3

Aim

p > n problem : more parameters than observations

For most regression models, it's not possible to fit the regression hyperplane if the number of observations (points) is less than the number of parameters (dimensions)



4

4

Aim

Redundancy problem:
we're interested to identify the most important features

●●●

IO	IP	IQ	IR	IS	IT	IU	IV	IW	IX	IY	IZ	JA	JB	JC
nCD4TregT	nCD4Th1 I	nCD4 TH17	nCD4 TH22	nCD4N	nCD4cm	nCD4em	nCD4tm	nCD4M	nCD8N	nCD8 Eff	nCD8cm	nCD8em	nCD8tm	nCD8M

●●●

5

5

Solution: Regularized models

6

6

Regularized models

Cost Function = Loss Function + Penalty

Lasso (L^1 Norm) Ridge (L^2 Norm)

$$\lambda \sum_{j=1}^{p} |\beta_j| \qquad \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2$$

7

7

Regularized models

Cost Function = Loss Function + Penalty

Lasso (L^1 Norm) ElasticNet Ridge (L^2 Norm)

$$\lambda \sum_{j=1}^{p} |\beta_j| \qquad \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + \frac{(1-\alpha)}{2} \sum_{j=1}^{p} \beta_j^2 \right] \qquad \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2$$

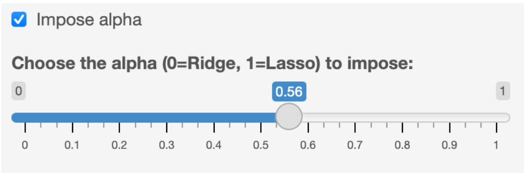
8

8

Regularized models

Cost Function = Loss Function + Penalty

Lasso (L^1 Norm)	ElasticNet	Ridge (L^2 Norm)
$\lambda \sum_{j=1}^{p} \lvert\beta_j\rvert$	$\lambda \left[\alpha \sum_{j=1}^{p} \lvert\beta_j\rvert + \frac{(1-\alpha)}{2} \sum_{j=1}^{p} \beta_j^2 \right]$	$\frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2$

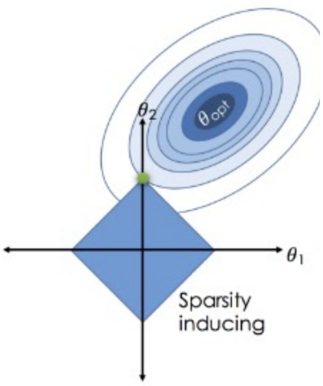


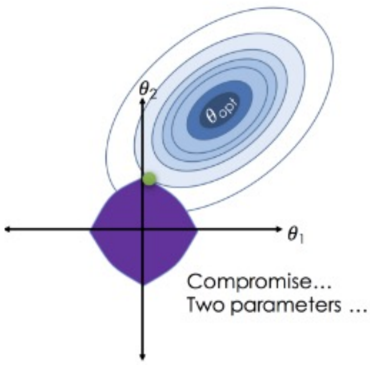
9

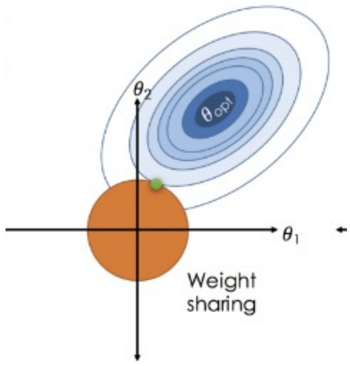
9

Regularized models

“strict” “sparse” — Lasso (L^1 Norm) — ElasticNet — Ridge (L^2 Norm) — “soft” “grouping”







10

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net.

10

Regularized models

Regularized = Penalized (any penalty)
≈ Sparse (only Lasso)

- regularized GLM (Lasso, Elastic-Net, Ridge) you choose alpha
- sparse PLS-DA (only Lasso)
- Multi-Block sparse PLS-DA = DIABLO (only Lasso)

11

11

Suitable methods

Based on response variable:

Categorical

- binary (e.g. dead-alive) — sPLS-DA and GLM
- multinomial (e.g. response - partial response - no response) — only sPLS-DA

Numeric (e.g. CRP level in blood) — only GLM

Survival time

- event + time until event — only GLM

Based on predictors:

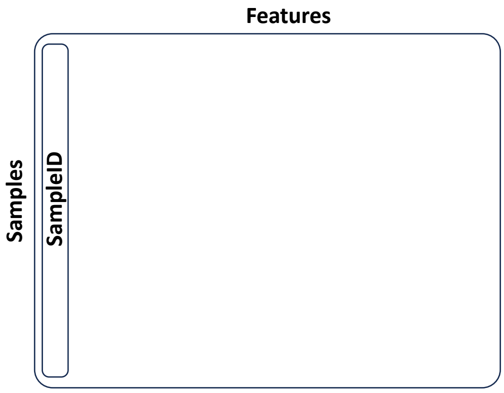
- Numeric + Categorical — GLM
- only Numeric — sPLS-DA (categorical excluded automatically)

12

12

Data structure

Data - Monoblock

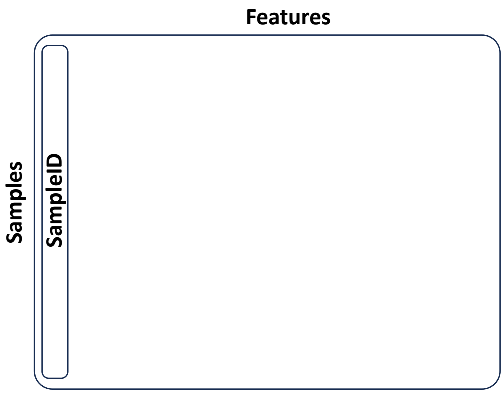


13

13

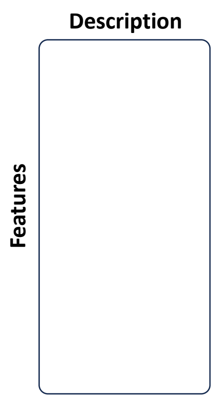
Data structure

Data - Monoblock



ID variable is obligatory and has to UNIQUELY identify each observation after filters are applied!

Feature table



check that there are no duplicated rows or columns!

14

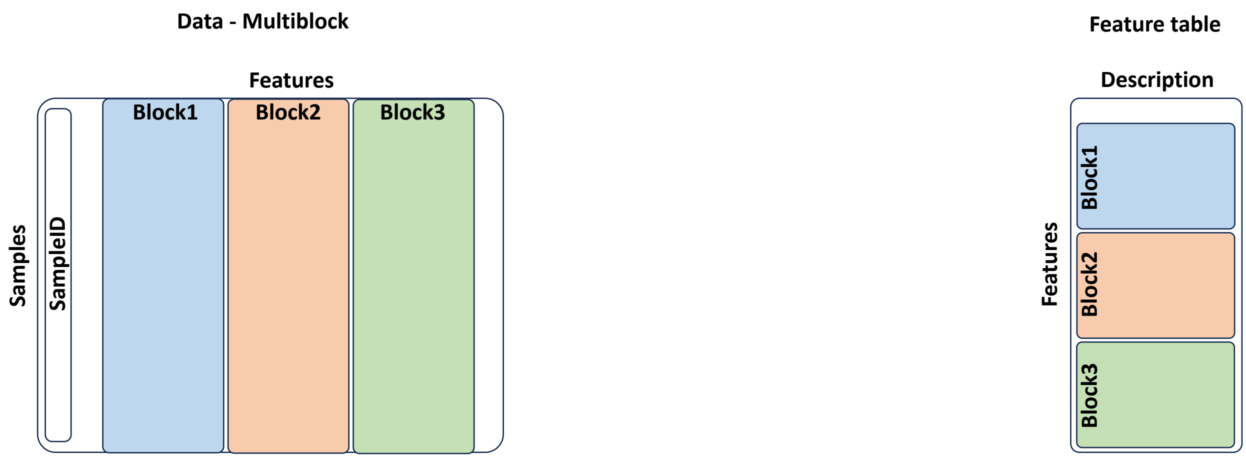
14

Data structure

15

15

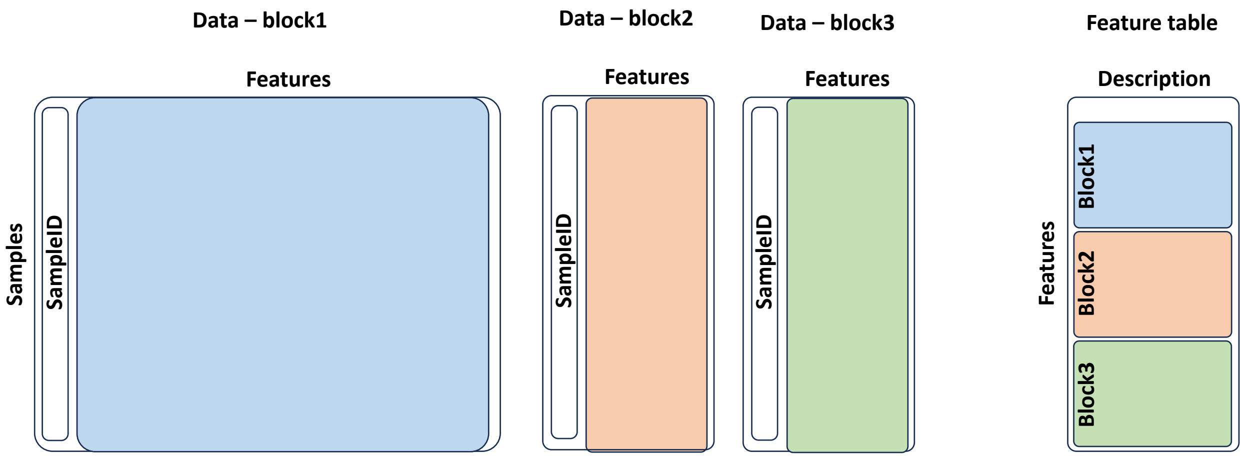
Data structure



16

16

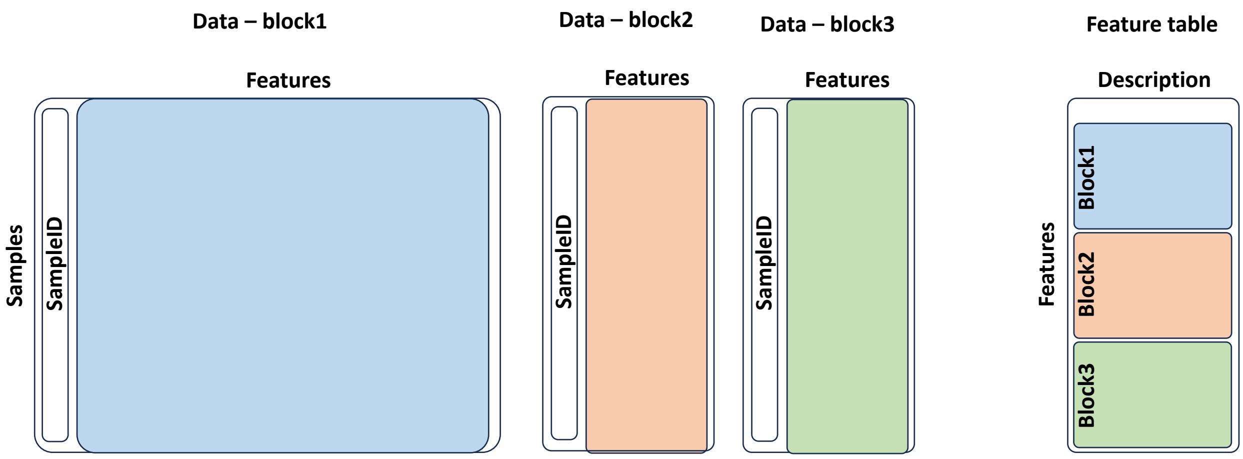
Data structure



17

17

Data structure



key variables, combination of which should UNIQUELY identify each observation in each sheet!

18

18

Wide format of dependent data

e.g. different timepoints for the same variable

FD	FE	FF	FG
ADL M6	IADL M6	ADL6 M12	IADL4 M12
3	0	3	0

19

19

Wi format of depend data

e.g. diffe oints fo variable

FD	FE		FG
ADL M6	IADL	12	IADL4 M12
3			0

20

20

Long format

	A	B	FQ	FR	FS	FT
1	PatientCode	Timepoint	nCD4	%CD4	nCD8	%CD8
2	HIP001	J0	498,64	51,2	311,31	31,96
3	HIP001	J10	666,36	48,59	471,77	34,4
4	HIP001	J13	810,69	50,11	538,75	33,3
5	HIP001	J3	675,57	47,07	530,55	36,96
6	HIP001	J4	419,65	43,48	362,09	37,52
7	HIP001	J7	766,91	47,76	555,47	34,6
8	HIP001	M17	730,35	49,65	482,35	32,79
9	HIP002	J0	251,25	54,8	123,56	26,95
10	HIP002	J1	614,52	62,24	278,86	28,25
11	HIP002	J3	730,64	58,09	348,43	27,7
12	HIP002	J6	516,98	64,84	219,41	27,52
13	HIP002	M15	633,34	60,27	272,96	25,97

21

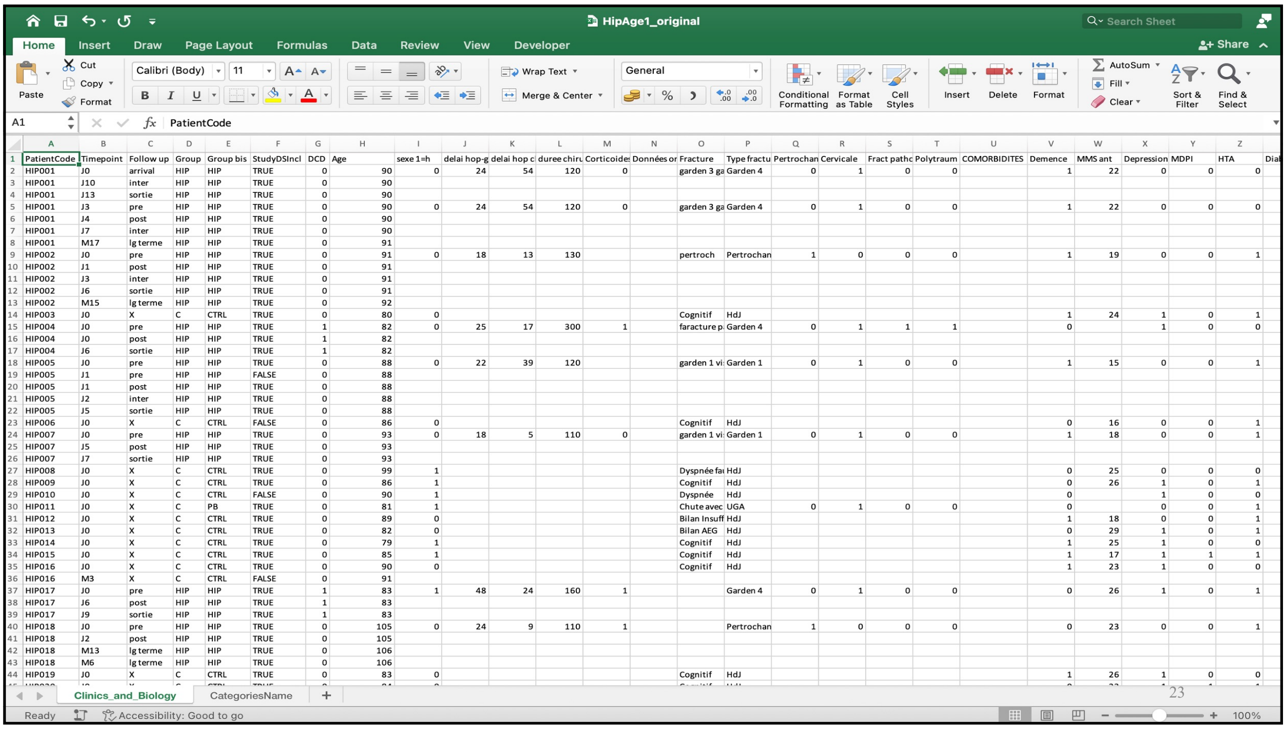
21

Other remarks

- Missing values are marked as ` `,`?`, `ND`, `nd`,`NA`,`na` or `#N/A` (excel-integrated NA notation)
- Check attentively all missing values for response variables (they will be ignored by the model if you don't manage it yourself)
- Models require variability inside each group

22

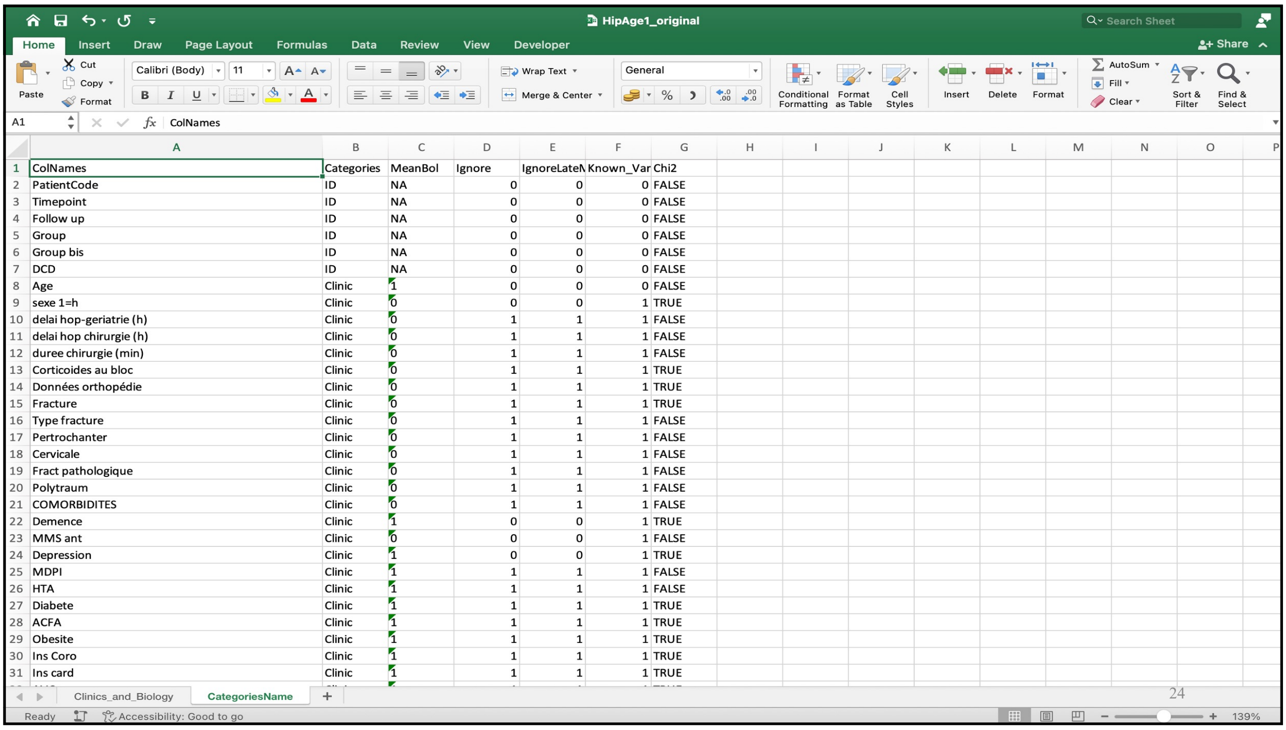
22



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	PatientCode	Timepoint	Follow up	Group	Group bis	StudyDSIncl	DCD	Age	sexe 1=h	delai hop-g	delai hop c	duree chiru	Corticoide:
2	HIP001	J0	arrival	HIP	HIP	TRUE	0	90	0	24	54	120	0
3	HIP001	J10	inter	HIP	HIP	TRUE	0	90					
4	HIP001	J13	sortie	HIP	HIP	TRUE	0	90					
5	HIP001	J3	pre	HIP	HIP	TRUE	0	90	0	24	54	120	0
6	HIP001	J4	post	HIP	HIP	TRUE	0	90					
7	HIP001	J7	inter	HIP	HIP	TRUE	0	90					
8	HIP001	M17	lg terme	HIP	HIP	TRUE	0	91					
9	HIP002	J0	pre	HIP	HIP	TRUE	0	91	0	18	13	130	
10	HIP002	J1	post	HIP	HIP	TRUE	0	91					
11	HIP002	J3	inter	HIP	HIP	TRUE	0	91					
12	HIP002	J6	sortie	HIP	HIP	TRUE	0	91					
13	HIP002	M15	lg terme	HIP	HIP	TRUE	0	92					
14	HIP003	J0	X	C	CTRL	TRUE	0	80	0				
15	HIP004	J0	pre	HIP	HIP	TRUE	1	82	0	25	17	300	1
16	HIP004	J0	post	HIP	HIP	TRUE	1	82					
17	HIP004	J6	sortie	HIP	HIP	TRUE	1	82					
18	HIP005	J0	pre	HIP	HIP	TRUE	0	88	0	22	39	120	
19	HIP005	J1	pre	HIP	HIP	FALSE	0	88					
20	HIP005	J1	post	HIP	HIP	TRUE	0	88					
21	HIP005	J2	inter	HIP	HIP	TRUE	0	88					
22	HIP005	J5	sortie	HIP	HIP	TRUE	0	88					
23	HIP006	J0	X	C	CTRL	FALSE	0	86	0				
24	HIP007	J0	pre	HIP	HIP	TRUE	0	93	0	18	5	110	0
25	HIP007	J5	post	HIP	HIP	TRUE	0	93					
26	HIP007	J7	sortie	HIP	HIP	TRUE	0	93					
27	HIP008	J0	X	C	CTRL	TRUE	0	99	1				
28	HIP009	J0	X	C	CTRL	TRUE	0	86	1				
29	HIP010	J0	X	C	CTRL	FALSE	0	90	1				
30	HIP011	J0	X	C	PB	TRUE	0	81	1				
31	HIP012	J0	X	C	CTRL	TRUE	0	89	0				
32	HIP013	J0	X	C	CTRL	TRUE	0	82	0				
33	HIP014	J0	X	C	CTRL	TRUE	0	79	1				
34	HIP015	J0	X	C	CTRL	TRUE	0	85	1				
35	HIP016	J0	X	C	CTRL	TRUE	0	90	0				
36	HIP016	M3	X	C	CTRL	FALSE	0	91					
37	HIP017	J0	pre	HIP	HIP	TRUE	1	83	1	48	24	160	1
38	HIP017	J6	post	HIP	HIP	TRUE	1	83					
39	HIP017	J9	sortie	HIP	HIP	TRUE	1	83					
40	HIP018	J0	pre	HIP	HIP	TRUE	0	105	0	24	9	110	1
41	HIP018	J2	post	HIP	HIP	TRUE	0	105					
42	HIP018	M13	lg terme	HIP	HIP	TRUE	0	106					
43	HIP018	M6	lg terme	HIP	HIP	TRUE	0	106					
44	HIP019	J0	X	C	CTRL	TRUE	0	83	0				

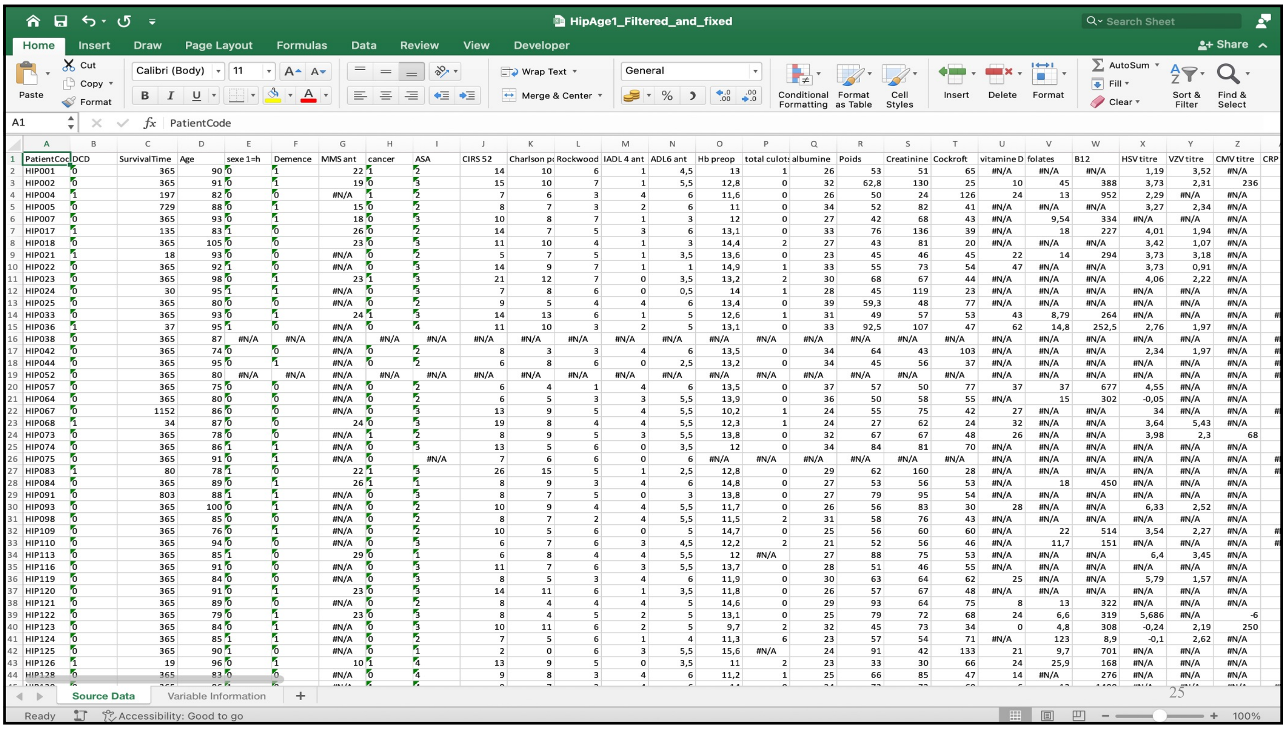
	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	Données or	Fracture	Type fractu	Pertrochan	Cervicale	Fract pathc	Polytraum	COMORBIDITES	Demence	MMS ant	Depression	MDPI	HTA	Dia
2		garden 3 ga	Garden 4	0	1	0	0		1	22	0	0	0	
3														
4														
5		garden 3 ga	Garden 4	0	1	0	0		1	22	0	0	0	
6														
7														
8														
9		pertroch	Pertrochan	1	0	0	0		1	19	0	0	1	
10														
11														
12														
13														
14		Cognitif	HdJ						1	24	1	0	1	
15		faracture p:	Garden 4	0	1	1	1		0		1	0	0	
16														
17														
18		garden 1 vi:	Garden 1	0	1	0	0		1	15	0	0	1	
19														
20														
21														
22														
23		Cognitif	HdJ						0	16	0	0	1	
24		garden 1 vi:	Garden 1	0	1	0	0		1	18	0	0	1	
25														
26														
27		Dyspnée fa	HdJ						0	25	0	0	0	
28		Cognitif	HdJ						0	26	1	0	1	
29		Dyspnée	HdJ						0		1	0	0	
30		Chute avec	UGA	0	1	0	0		0		0	0	1	
31		Bilan Insuff	HdJ						1	18	0	0	1	
32		Bilan AEG	HdJ						0	29	1	0	1	
33		Cognitif	HdJ						1	25	1	0	0	
34		Cognitif	HdJ						1	17	1	1	1	
35		Cognitif	HdJ						1	23	1	0	0	
36														
37			Garden 4	0	1	0	0		0	26	1	0	1	
38														
39														
40			Pertrochan	1	0	0	0		0	23	0	0	1	
41														
42														
43														
44		Cognitif	HdJ						1	26	1	0	0	

23

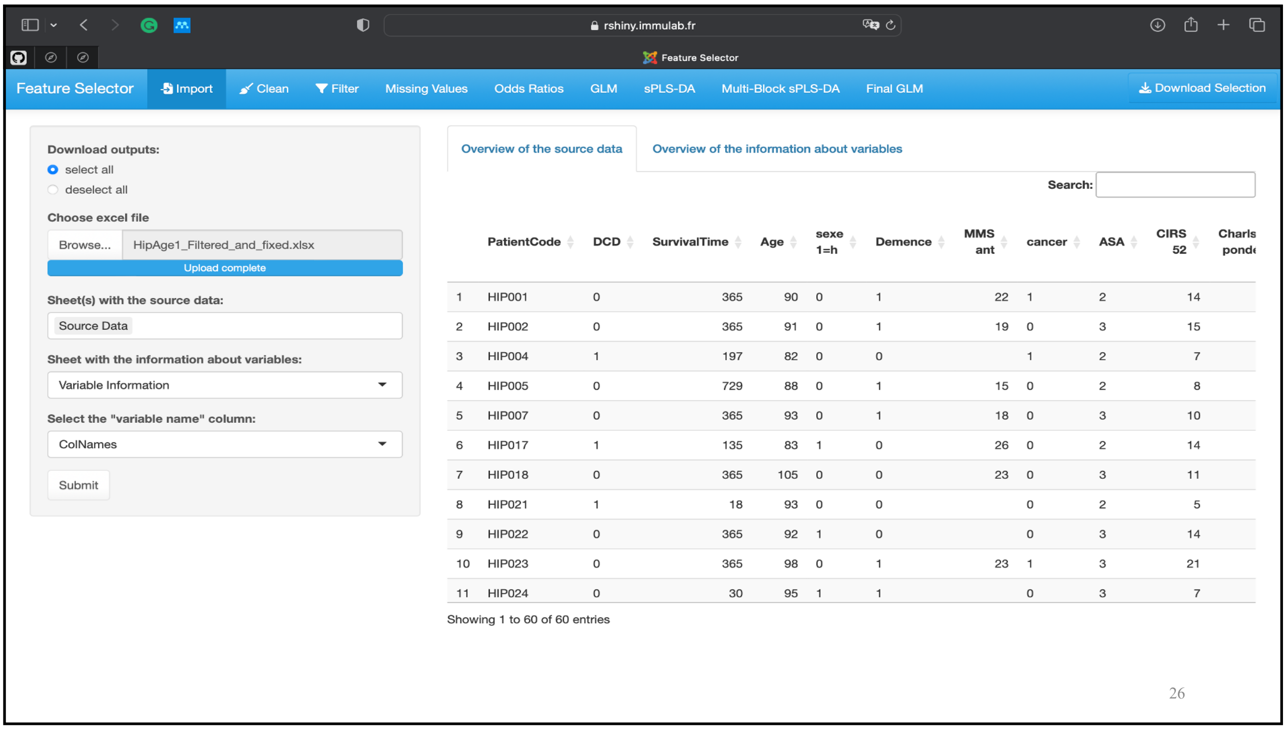


	A	B	C	D	E	F	G
1	ColNames	Categories	MeanBol	Ignore	IgnoreLateN	Known_Var	Chi2
2	PatientCode	ID	NA	0	0	0	FALSE
3	Timepoint	ID	NA	0	0	0	FALSE
4	Follow up	ID	NA	0	0	0	FALSE
5	Group	ID	NA	0	0	0	FALSE
6	Group bis	ID	NA	0	0	0	FALSE
7	DCD	ID	NA	0	0	0	FALSE
8	Age	Clinic	1	0	0	0	FALSE
9	sexe 1=h	Clinic	0	0	0	1	TRUE
10	delai hop-geriatrie (h)	Clinic	0	1	1	1	FALSE
11	delai hop chirurgie (h)	Clinic	0	1	1	1	FALSE
12	duree chirurgie (min)	Clinic	0	1	1	1	FALSE
13	Corticoides au bloc	Clinic	0	1	1	1	TRUE
14	Données orthopédie	Clinic	0	1	1	1	TRUE
15	Fracture	Clinic	0	1	1	1	TRUE
16	Type fracture	Clinic	0	1	1	1	FALSE
17	Pertrochanter	Clinic	0	1	1	1	FALSE
18	Cervicale	Clinic	0	1	1	1	FALSE
19	Fract pathologique	Clinic	0	1	1	1	FALSE
20	Polytraum	Clinic	0	1	1	1	FALSE
21	COMORBIDITES	Clinic	0	1	1	1	FALSE
22	Demence	Clinic	1	0	0	1	TRUE
23	MMS ant	Clinic	0	0	0	1	FALSE
24	Depression	Clinic	1	0	0	1	TRUE
25	MDPI	Clinic	1	1	1	1	FALSE
26	HTA	Clinic	1	1	1	1	FALSE
27	Diabete	Clinic	1	1	1	1	TRUE
28	ACFA	Clinic	1	1	1	1	TRUE
29	Obesite	Clinic	1	1	1	1	TRUE
30	Ins Coro	Clinic	1	1	1	1	TRUE
31	Ins card	Clinic	1	1	1	1	TRUE

24

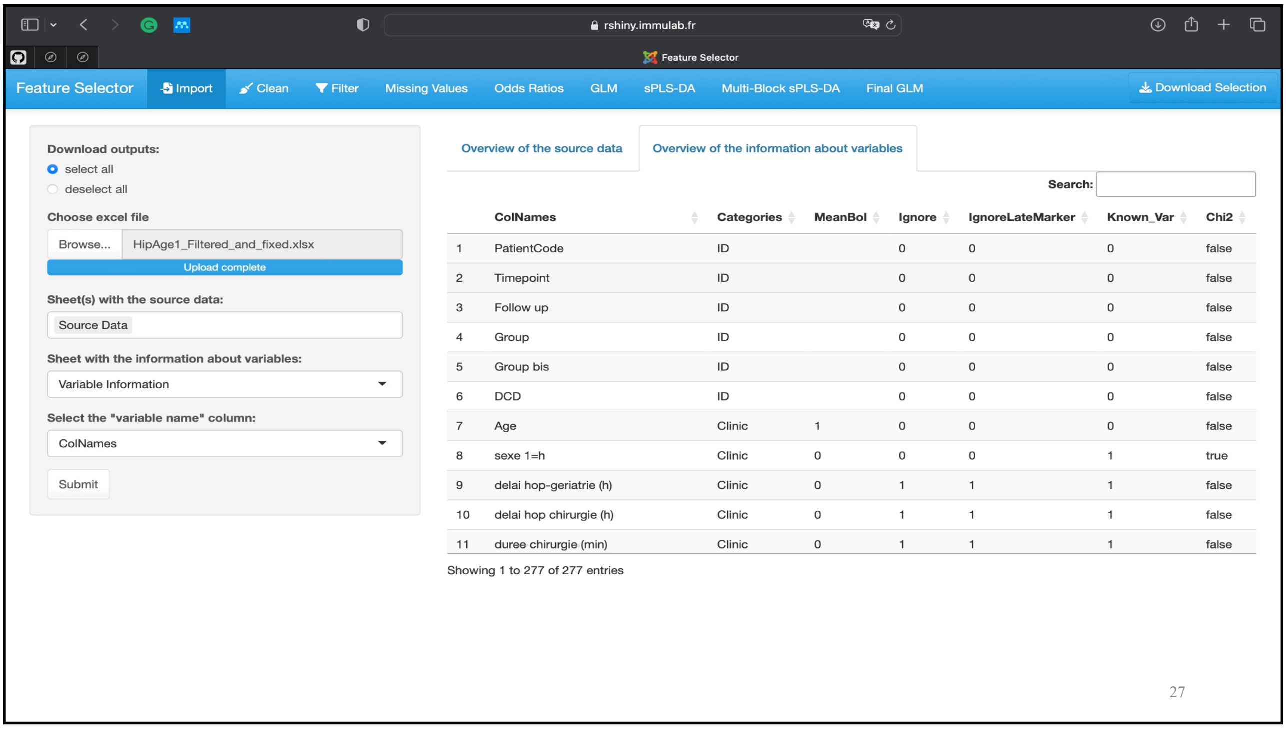


25



	PatientCode	DCD	SurvivalTime	Age	sexe 1=h	Demence	MMS ant	cancer	ASA	CIRS 52	Charls ponde
1	HIP001	0	365	90	0	1	22	1	2	14	
2	HIP002	0	365	91	0	1	19	0	3	15	
3	HIP004	1	197	82	0	0		1	2	7	
4	HIP005	0	729	88	0	1	15	0	2	8	
5	HIP007	0	365	93	0	1	18	0	3	10	
6	HIP017	1	135	83	1	0	26	0	2	14	
7	HIP018	0	365	105	0	0	23	0	3	11	
8	HIP021	1	18	93	0	0		0	2	5	
9	HIP022	0	365	92	1	0		0	3	14	
10	HIP023	0	365	98	0	1	23	1	3	21	
11	HIP024	0	30	95	1	1		0	3	7	

26



27

Clean the data

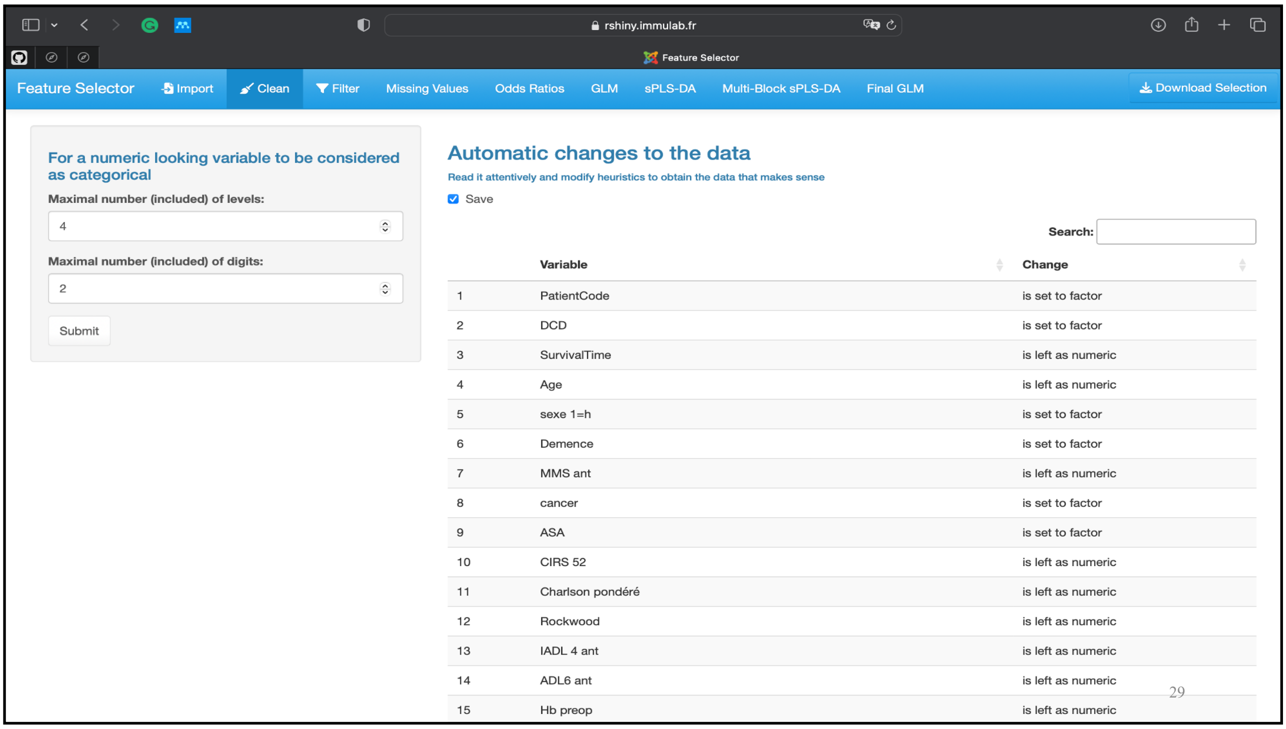
- Removes variables with no variation
- Removes dates
- Sets appropriate types to variables (categorical / numeric / logical)
- Problem – if looks numeric, but is to be better considered categorical

e.g. medical score going from 1 to 3 is either approximated as a number or as the group variable (giving 3 groups of patients). For small numbers of groups and in the absence of numerical relations between the score values.

- The user has to choose two heuristics – maximal nb of categories and digits for such variables to be considered as categorical

28

28



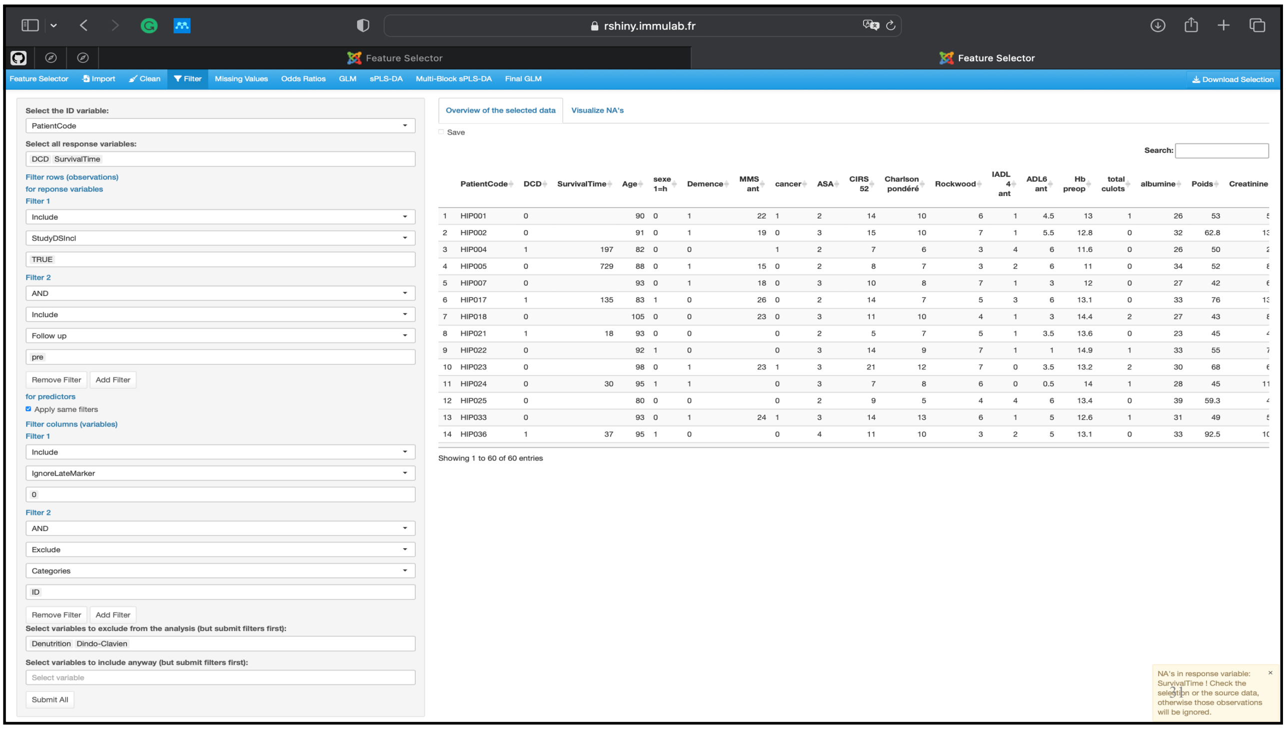
29

Filter

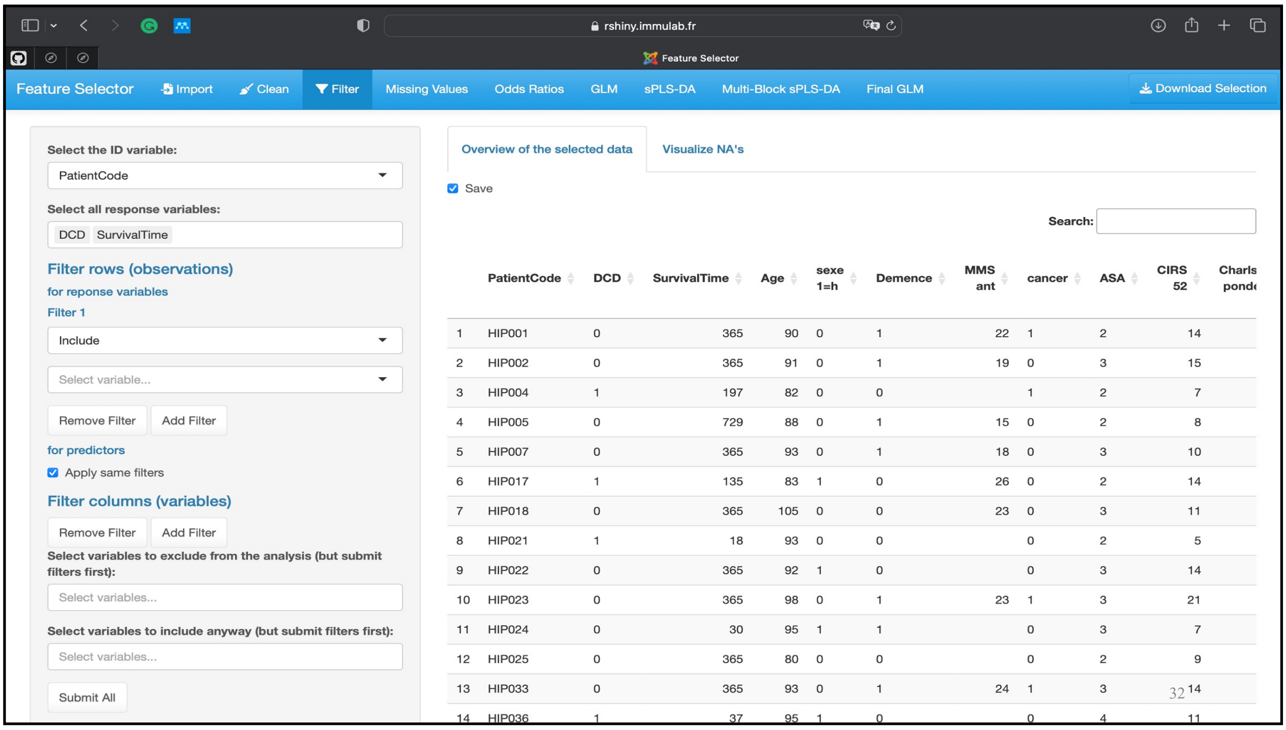
- Select the ID variable (uniquely identifying each filtered observation)
- Select the response variable(s)
- Set filters for response variable(s) and predictors (same or separate)
- If separate filters are set, pay attention to whether you consider the response variable as the potential predictor or not ! (e.g. we might be interested to predict post-surgery CRP level with pre-surgery CRP level, but not with post-surgery CRP)

30

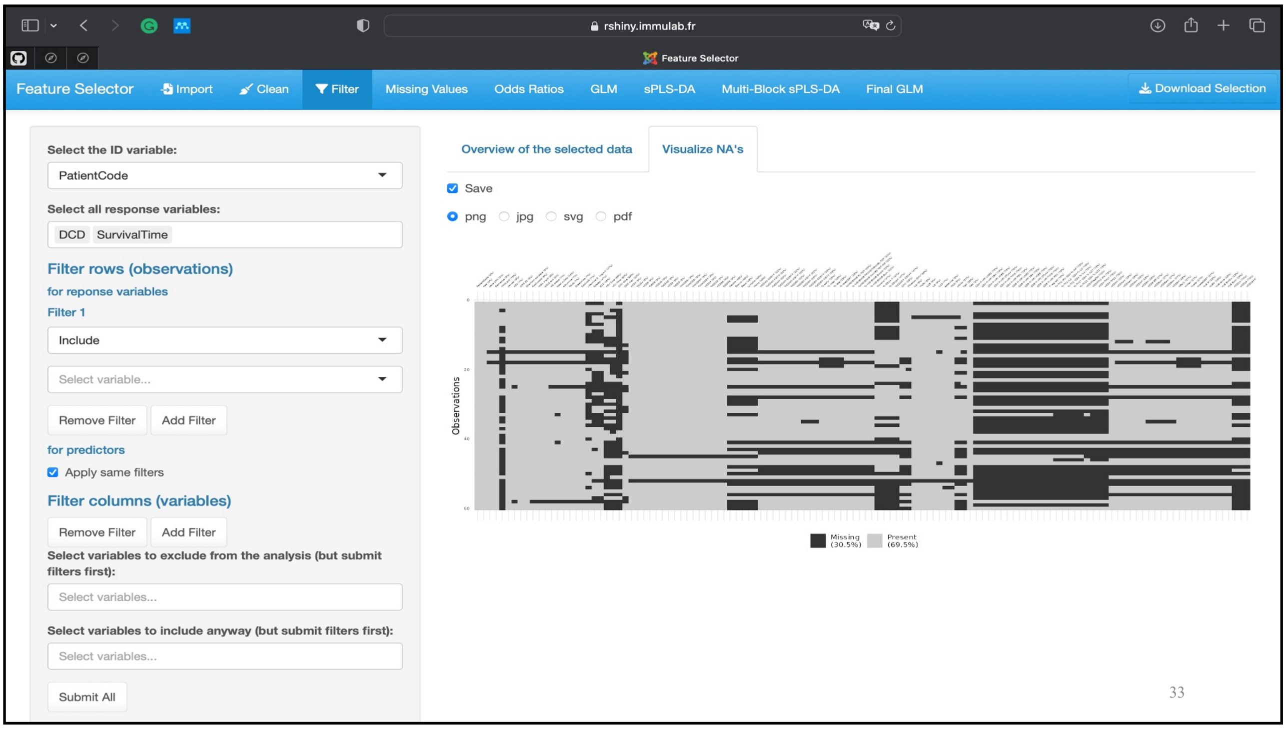
30


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the ID variable:
PatientCode
Select all response variables:
DCD SurvivalTime
Filter rows (observations)
for reponse variables
Filter 1
Include
StudyDSIncl
TRUE
Filter 2
AND
Include
Follow up
pre
Remove Filter
Add Filter
for predictors
Apply same filters
Filter columns (variables)
Filter 1
Include
IgnoreLateMarker
0
Filter 2
AND
Exclude
Categories
ID
Remove Filter
Add Filter
Select variables to exclude from the analysis (but submit filters first):
Denutrition Dindo-Clavien
Select variables to include anyway (but submit filters first):
Select variable
Submit All
Overview of the selected data
Visualize NA's
Save
Search:
Showing 1 to 60 of 60 entries
NA's in response variable: SurvivalTime ! Check the selection or the source data, otherwise those observations will be ignored.

31


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the ID variable:
PatientCode
Select all response variables:
DCD SurvivalTime
Filter rows (observations)
for reponse variables
Filter 1
Include
Select variable...
Remove Filter
Add Filter
for predictors
Apply same filters
Filter columns (variables)
Remove Filter
Add Filter
Select variables to exclude from the analysis (but submit filters first):
Select variables...
Select variables to include anyway (but submit filters first):
Select variables...
Submit All
Overview of the selected data
Visualize NA's
Save
Search:
PatientCode DCD SurvivalTime Age sexe 1=h Demence MMS ant cancer ASA CIRS 52
1 HIP001 0 365 90 0 1 22 1 2 14
2 HIP002 0 365 91 0 1 19 0 3 15
3 HIP004 1 197 82 0 0 1 2 7
4 HIP005 0 729 88 0 1 15 0 2 8
5 HIP007 0 365 93 0 1 18 0 3 10
6 HIP017 1 135 83 1 0 26 0 2 14
7 HIP018 0 365 105 0 0 23 0 3 11
8 HIP021 1 18 93 0 0 0 2 5
9 HIP022 0 365 92 1 0 0 3 14
10 HIP023 0 365 98 0 1 23 1 3 21
11 HIP024 0 30 95 1 1 0 3 7
12 HIP025 0 365 80 0 0 0 2 9
13 HIP033 0 365 93 0 1 24 1 3 14

32



33

Missing Values

Stringent = exclude rows / columns having too many NA's

Impute (= fill in NA's in each variable) with :

- central tendency – the mean, median or mode
- k Nearest Neighbors – the mean of k most resembling samples
- other methods are to come…

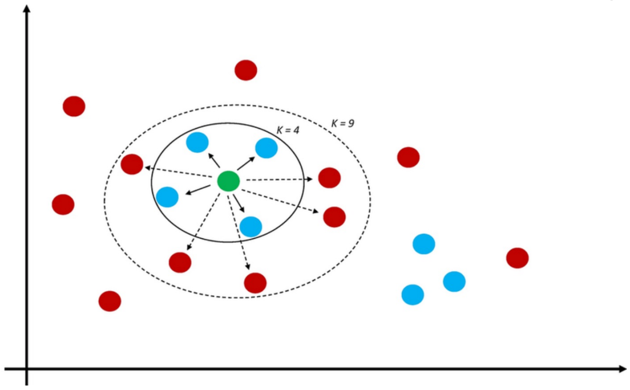
Musolf, Anthony & Holzinger, Emily & Malley, James & Bailey-Wilson, Joan. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Human Genetics. 141. 10.1007/s00439-021-02402-z.

34

34

Missing Values

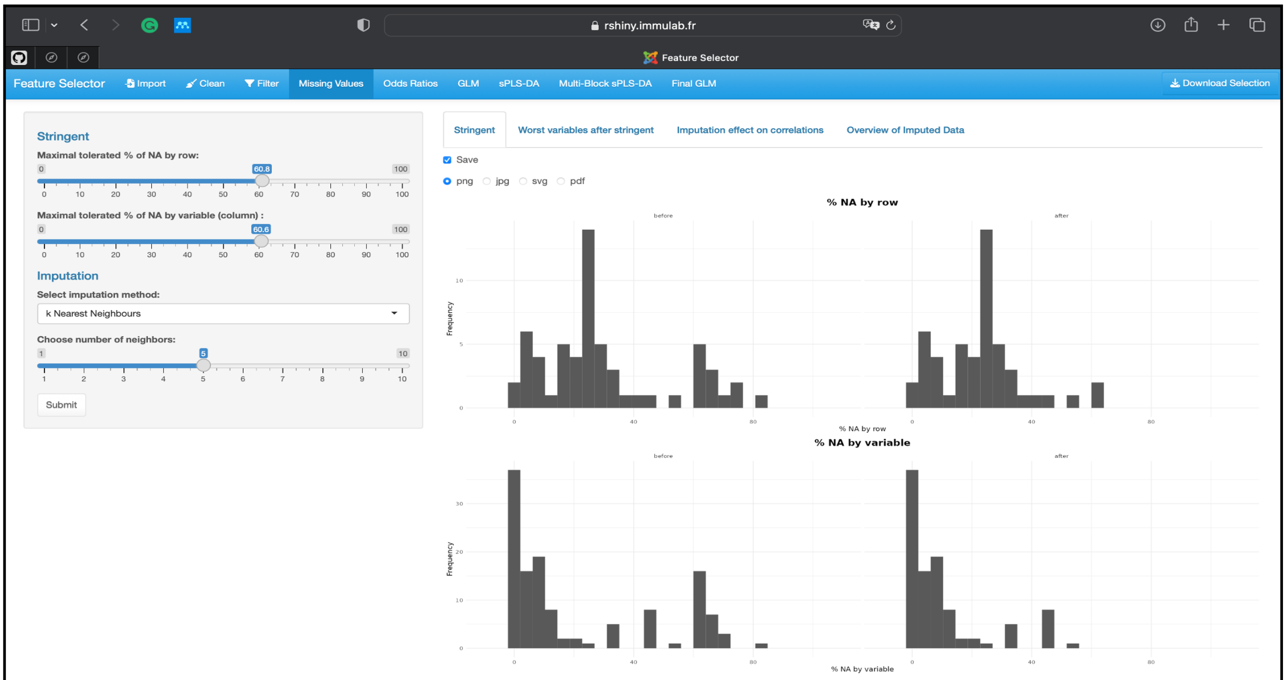
kNN – replace NA's with the mean of k most resembling samples



Musolf, Anthony & Holzinger, Emily & Malley, James & Bailey-Wilson, Joan. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Human Genetics.

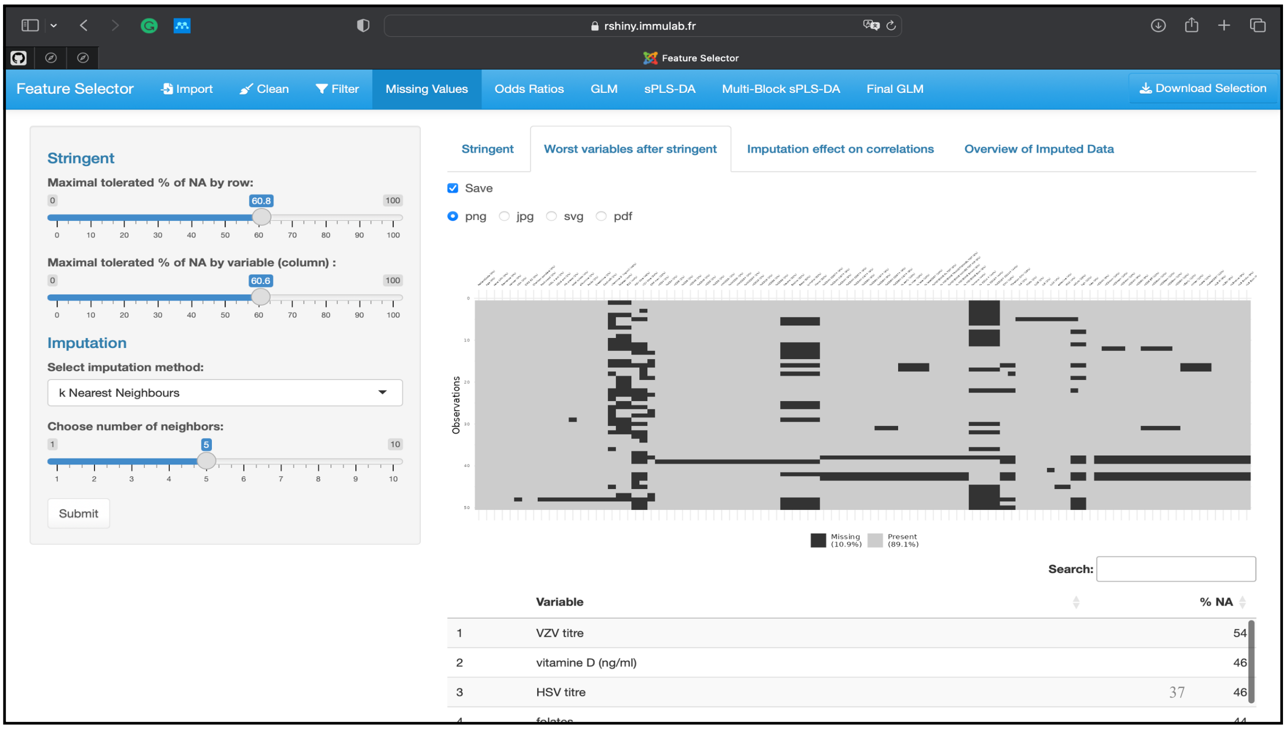
35

35

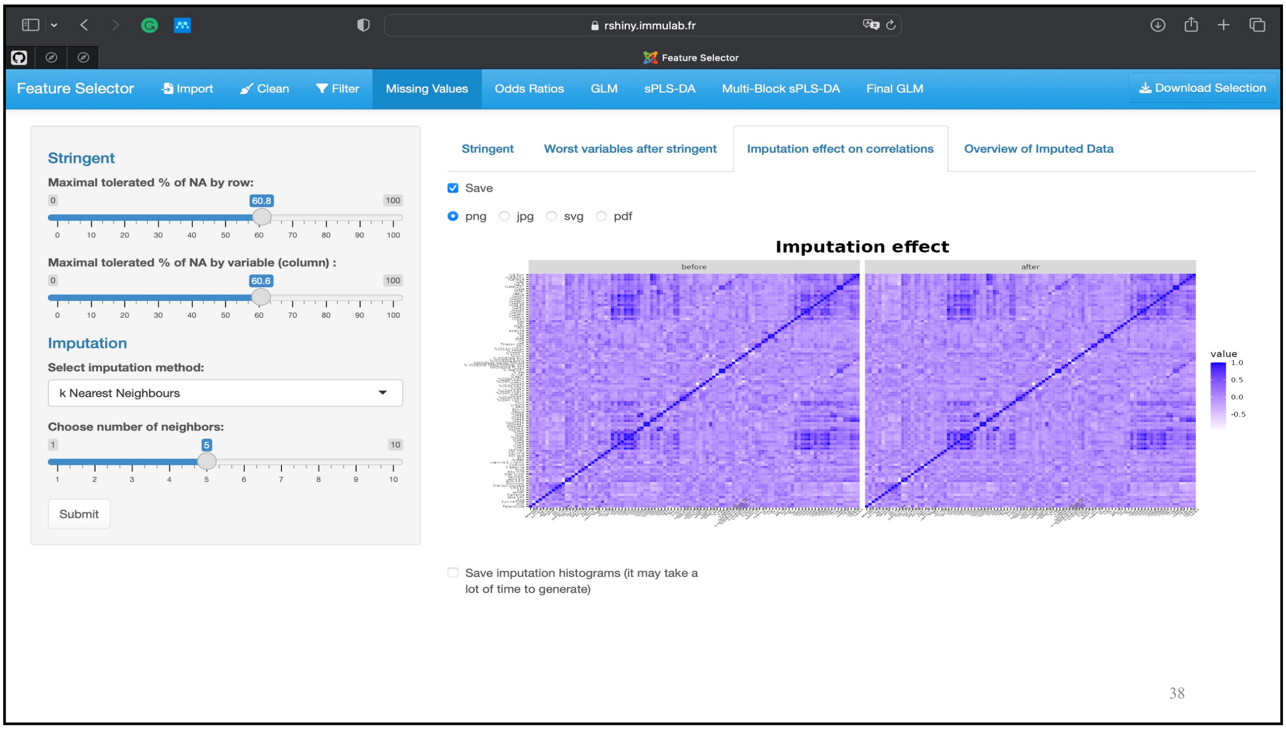


36

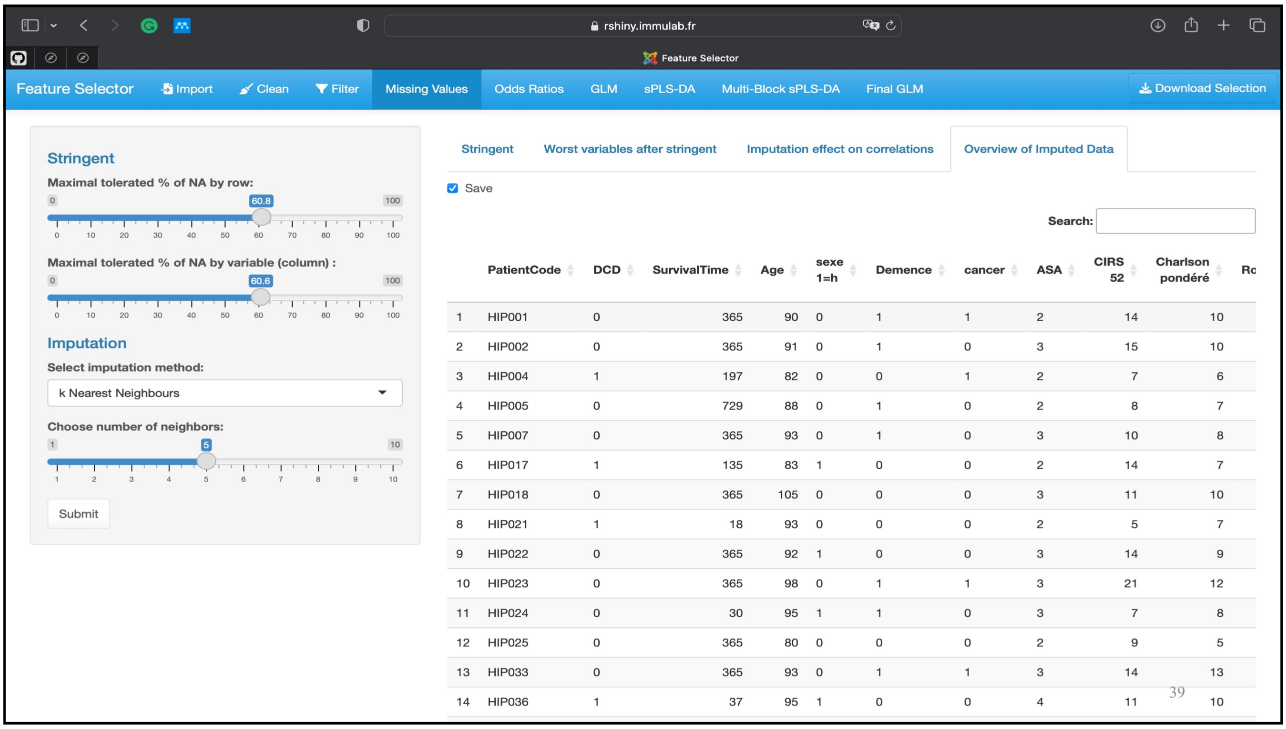
36


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Stringent
Maximal tolerated % of NA by row:
60.8
Maximal tolerated % of NA by variable (column) :
60.6
Imputation
Select imputation method:
k Nearest Neighbours
Choose number of neighbors:
5
Submit
Stringent
Worst variables after stringent
Imputation effect on correlations
Overview of Imputed Data
Save
png
jpg
svg
pdf
Observations
Missing (10.9%)
Present (89.1%)
Search:
Variable
% NA
1 VZV titre 54
2 vitamine D (ng/ml) 46
3 HSV titre 46
37

37


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Stringent
Maximal tolerated % of NA by row:
60.8
Maximal tolerated % of NA by variable (column) :
60.6
Imputation
Select imputation method:
k Nearest Neighbours
Choose number of neighbors:
5
Submit
Stringent
Worst variables after stringent
Imputation effect on correlations
Overview of Imputed Data
Save
png
jpg
svg
pdf
Imputation effect
before
after
value
1.0
0.5
0.0
-0.5
Save imputation histograms (it may take a lot of time to generate)
38

38



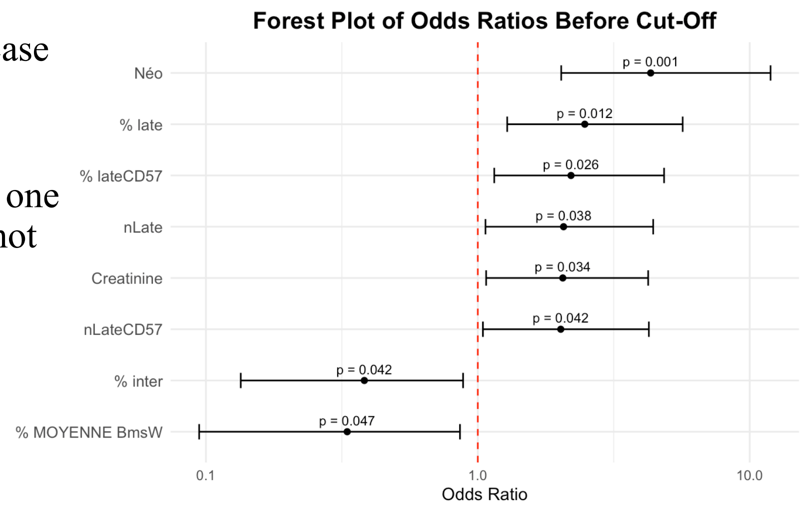
	PatientCode	DCD	SurvivalTime	Age	sexe 1=h	Demence	cancer	ASA	CIRS 52	Charlson pondéré	Ro
1	HIP001	0	365	90	0	1	1	2	14	10	
2	HIP002	0	365	91	0	1	0	3	15	10	
3	HIP004	1	197	82	0	0	1	2	7	6	
4	HIP005	0	729	88	0	1	0	2	8	7	
5	HIP007	0	365	93	0	1	0	3	10	8	
6	HIP017	1	135	83	1	0	0	2	14	7	
7	HIP018	0	365	105	0	0	0	3	11	10	
8	HIP021	1	18	93	0	0	0	2	5	7	
9	HIP022	0	365	92	1	0	0	3	14	9	
10	HIP023	0	365	98	0	1	1	3	21	12	
11	HIP024	0	30	95	1	1	0	3	7	8	
12	HIP025	0	365	80	0	0	0	2	9	5	
13	HIP033	0	365	93	0	1	1	3	14	13	
14	HIP036	1	37	95	1	0	0	4	11	10	

39

Odds Ratios

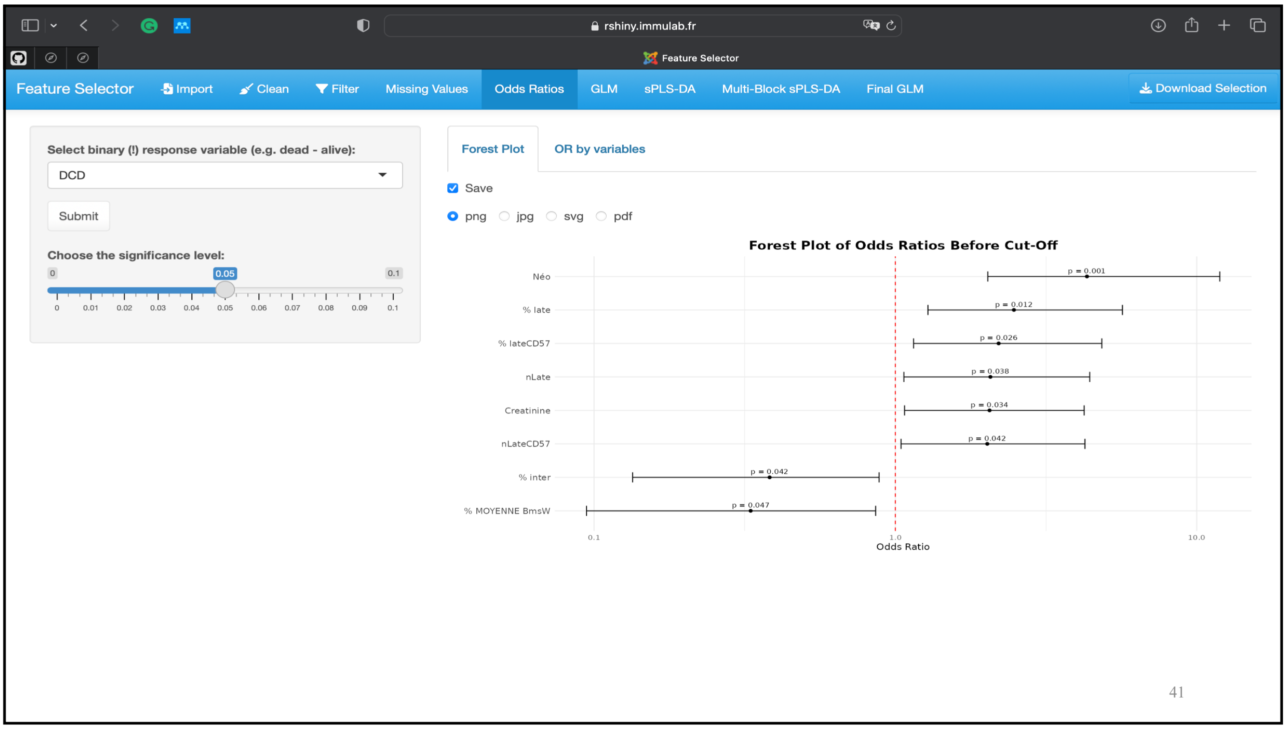
Preliminary analysis in case of binary response

Variables are considered one by one, interactions are not taken into account

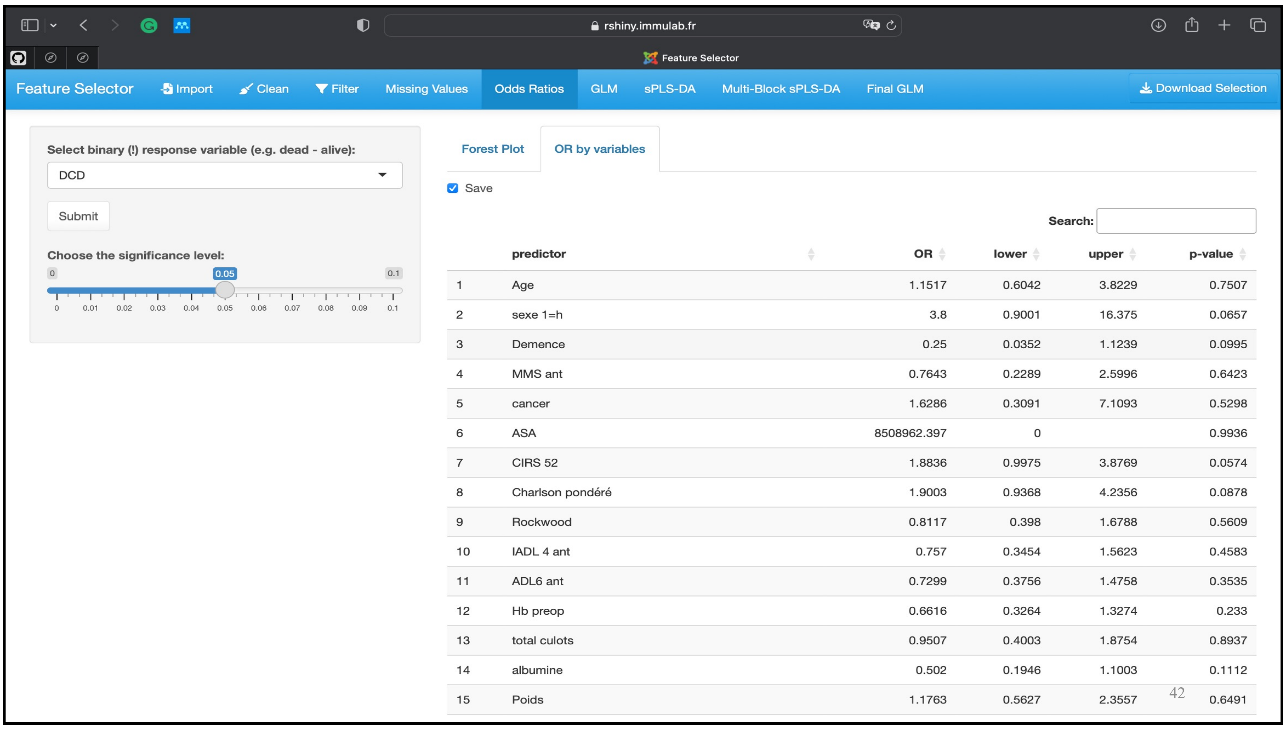


attention to the x axis!

40



41



	predictor	OR	lower	upper	p-value
1	Age	1.1517	0.6042	3.8229	0.7507
2	sexe 1=h	3.8	0.9001	16.375	0.0657
3	Demence	0.25	0.0352	1.1239	0.0995
4	MMS ant	0.7643	0.2289	2.5996	0.6423
5	cancer	1.6286	0.3091	7.1093	0.5298
6	ASA	8508962.397	0		0.9936
7	CIRS 52	1.8836	0.9975	3.8769	0.0574
8	Charlson pondéré	1.9003	0.9368	4.2356	0.0878
9	Rockwood	0.8117	0.398	1.6788	0.5609
10	IADL 4 ant	0.757	0.3454	1.5623	0.4583
11	ADL6 ant	0.7299	0.3756	1.4758	0.3535
12	Hb preop	0.6616	0.3264	1.3274	0.233
13	total culots	0.9507	0.4003	1.8754	0.8937
14	albumine	0.502	0.1946	1.1003	0.1112
15	Poids	1.1763	0.5627	2.3557	0.6491

42

GLM

Multiple Regression models can predict :

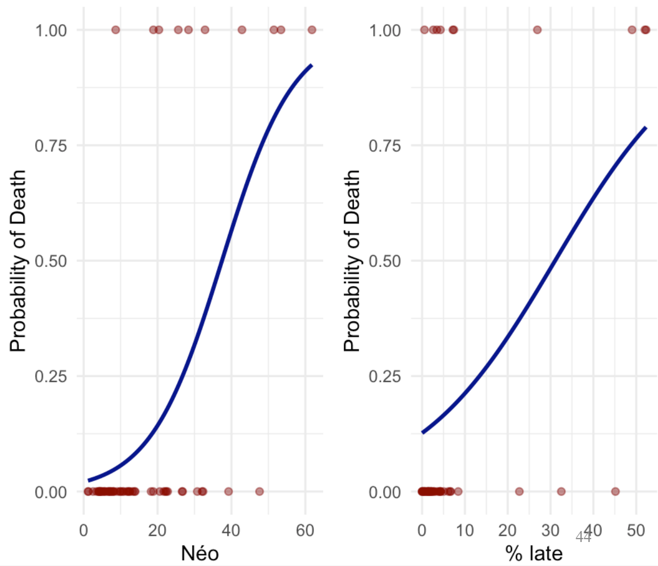
- Linear – numeric outcome
- Logistic – binary outcome
- Cox Proportional Hazards model (Survival analysis) – hazard function

43

43

GLM

- Logistic – binary outcome

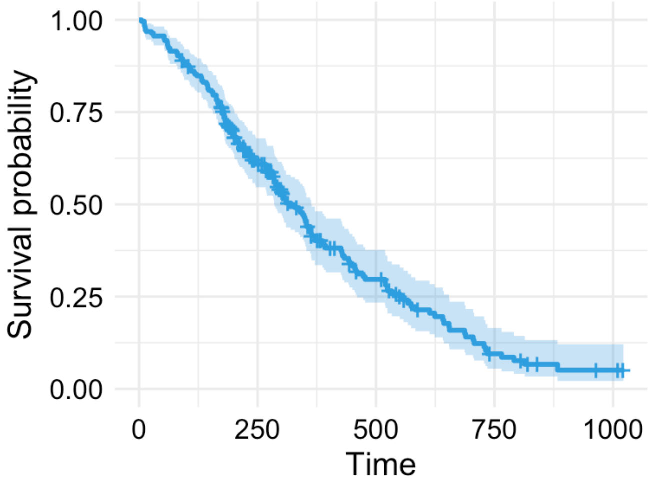


44

44

GLM

- Cox Proportional Hazards model (Survival analysis) – hazard function



45

http://www.sthda.com/english/wiki/cox-proportional-hazards-model

45

GLM Assumptions

Linear regression:

- **linearity** of relationship
- independence of observations
- no perfect multicollinearity
- (no endogeneity)

Regularization is more robust to violations of these assumptions:

- homoscedasticity (constant variance)
- residual normality

46

46

GLM Assumptions

Logistic regression:

- **linearity** of log-odds
- independence of observations
- no perfect multicollinearity

- large sample size

47

47

GLM Assumptions

Cox PH model:

- **linearity** of log-hazard
- independence of observations
- no perfect multicollinearity

- proportional hazards (hazard ratios are constant over time)
- non-informative censoring

48

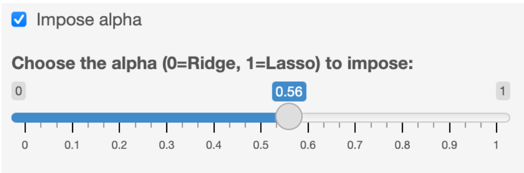
48

Regularized GLM

Cost Function = Loss Function + Penalty

Lasso (L^1 Norm) | ElasticNet | Ridge (L^2 Norm)

$$\lambda \sum_{j=1}^{p} |\beta_j| \qquad \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + \frac{(1-\alpha)}{2} \sum_{j=1}^{p} \beta_j^2 \right] \qquad \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2$$



49

49

Regularized GLM Assumptions

- **the same scale** – variables must be standardized, but no clear consensus on whether dummy variables (indicators of levels of categorical variables) have to be standardized or not – try both ways

50

50

Model Robustness

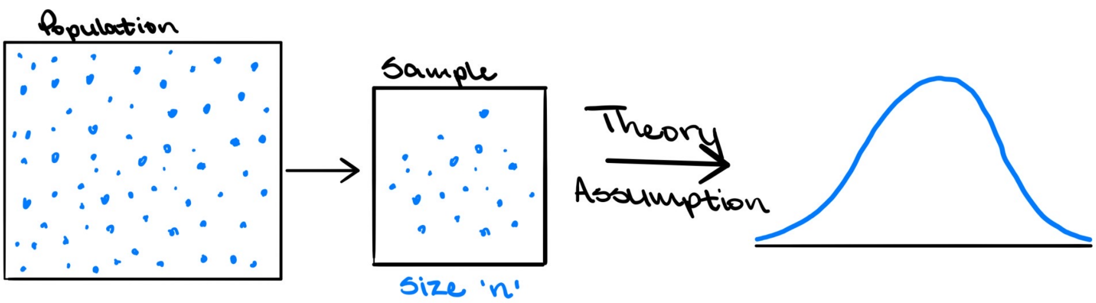
If I repeat the experiment, will I get the same result?

51

51

Model Robustness

Theoretical approach:

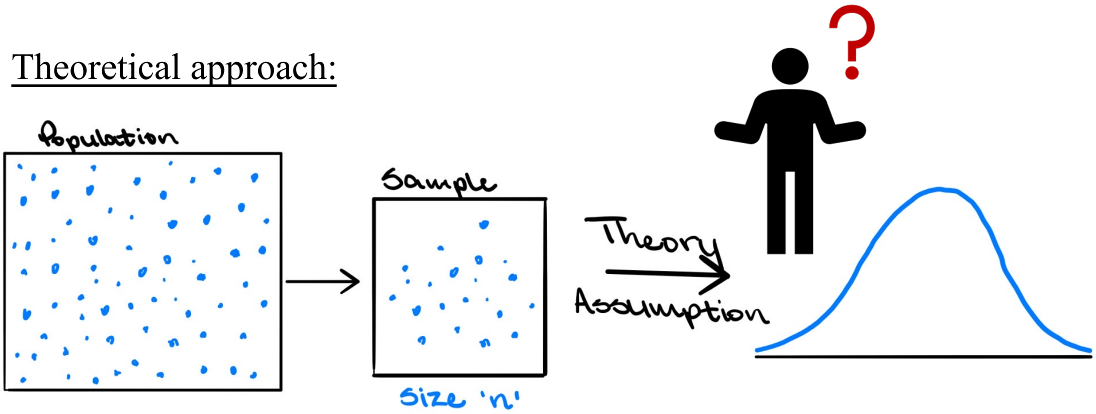


52
https://www.linkedin.com/pulse/bootstrapping-statistics-what-why-its-used-trist-n-joseph

52

Model Robustness

Theoretical approach:



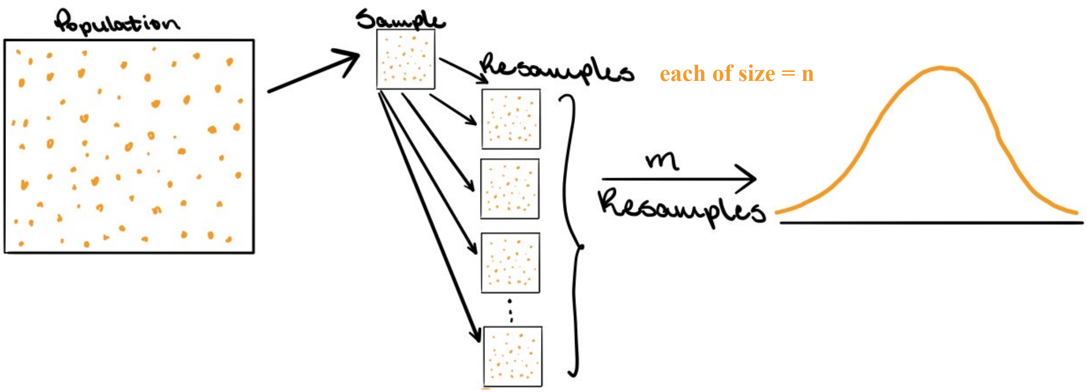
https://www.linkedin.com/pulse/bootstrapping-statistics-what-why-its-used-trist-n-joseph

53

53

Model Robustness

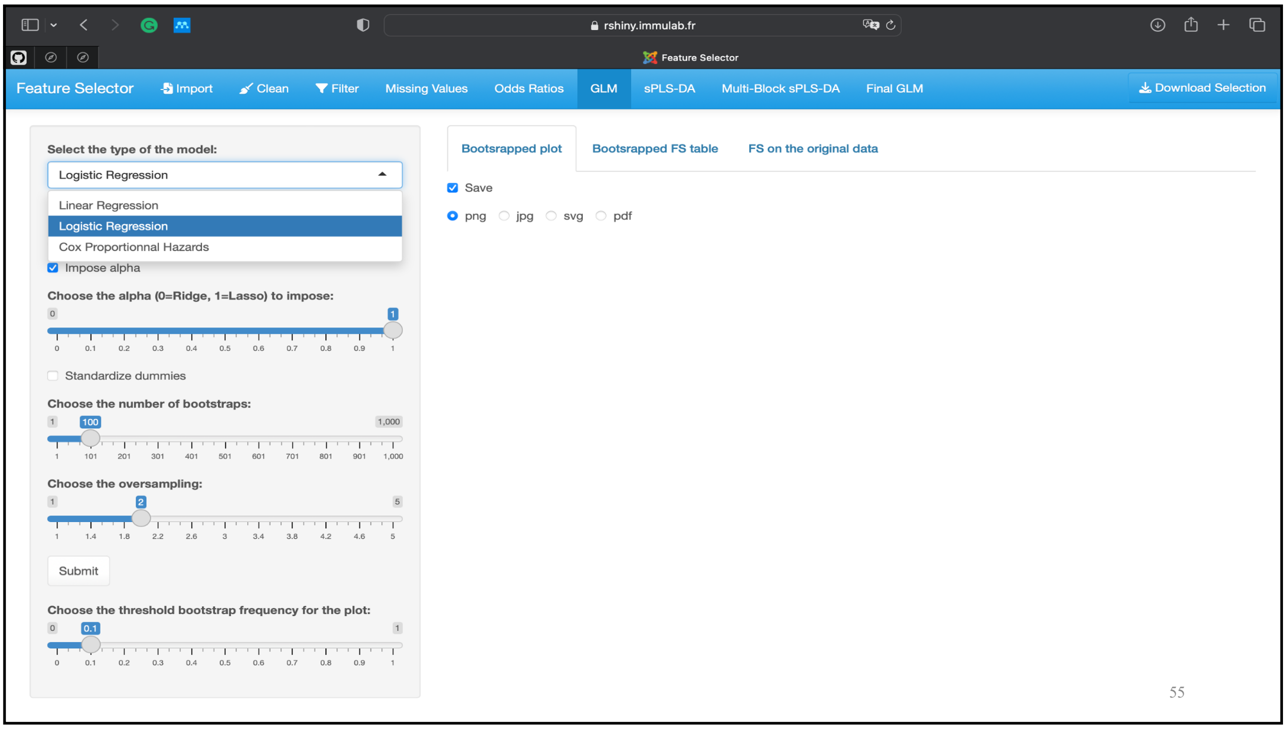
Brute force simulation approach: **Bootstrapping** or Cross-Validation



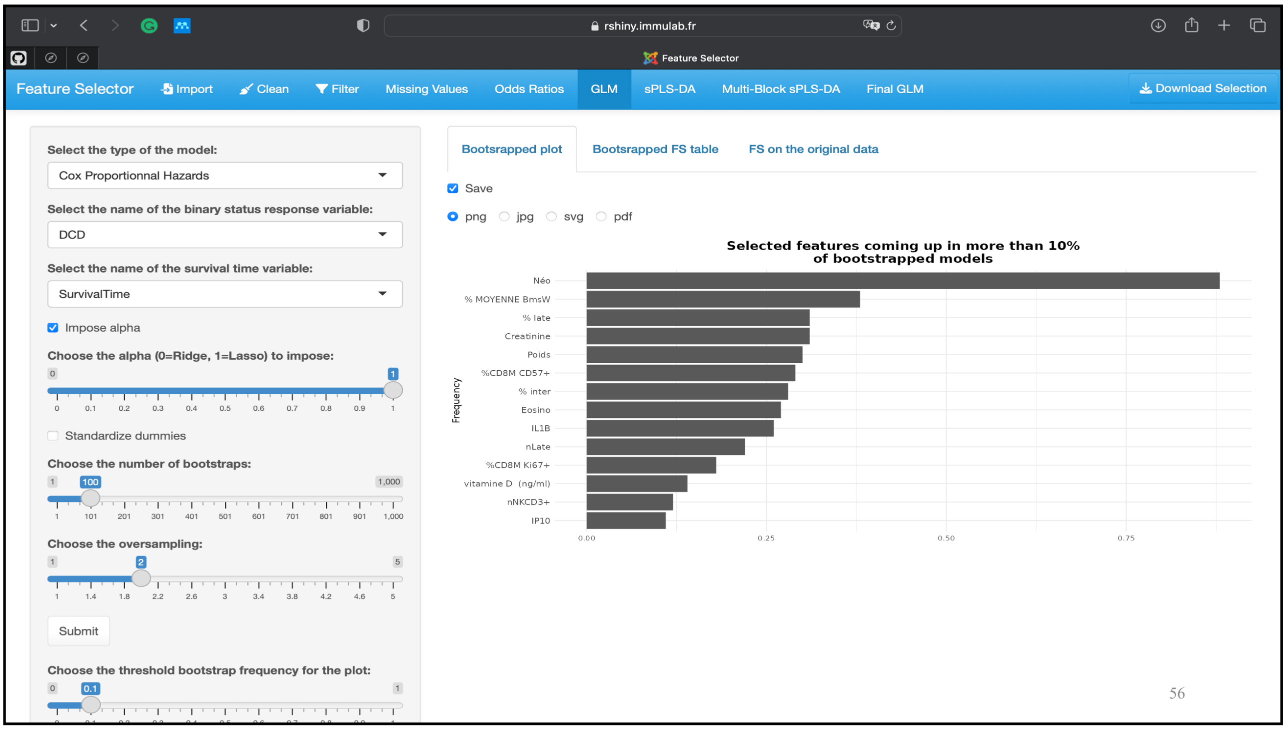
https://www.linkedin.com/pulse/bootstrapping-statistics-what-why-its-used-trist-n-joseph

54

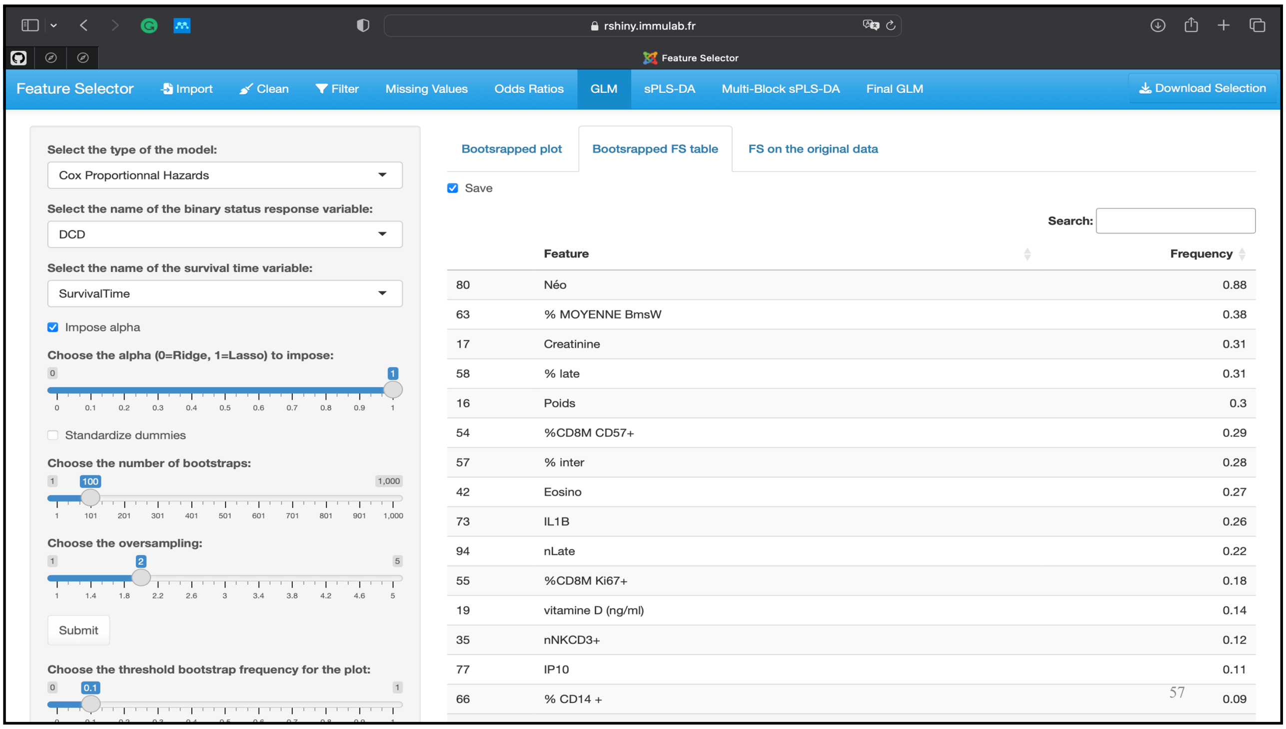
54


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Logistic Regression
Linear Regression
Logistic Regression
Cox Proportionnal Hazards
Impose alpha
Choose the alpha (0=Ridge, 1=Lasso) to impose:
Standardize dummies
Choose the number of bootstraps:
Choose the oversampling:
Submit
Choose the threshold bootstrap frequency for the plot:
Bootstrapped plot
Bootstrapped FS table
FS on the original data
Save
png
jpg
svg
pdf
55

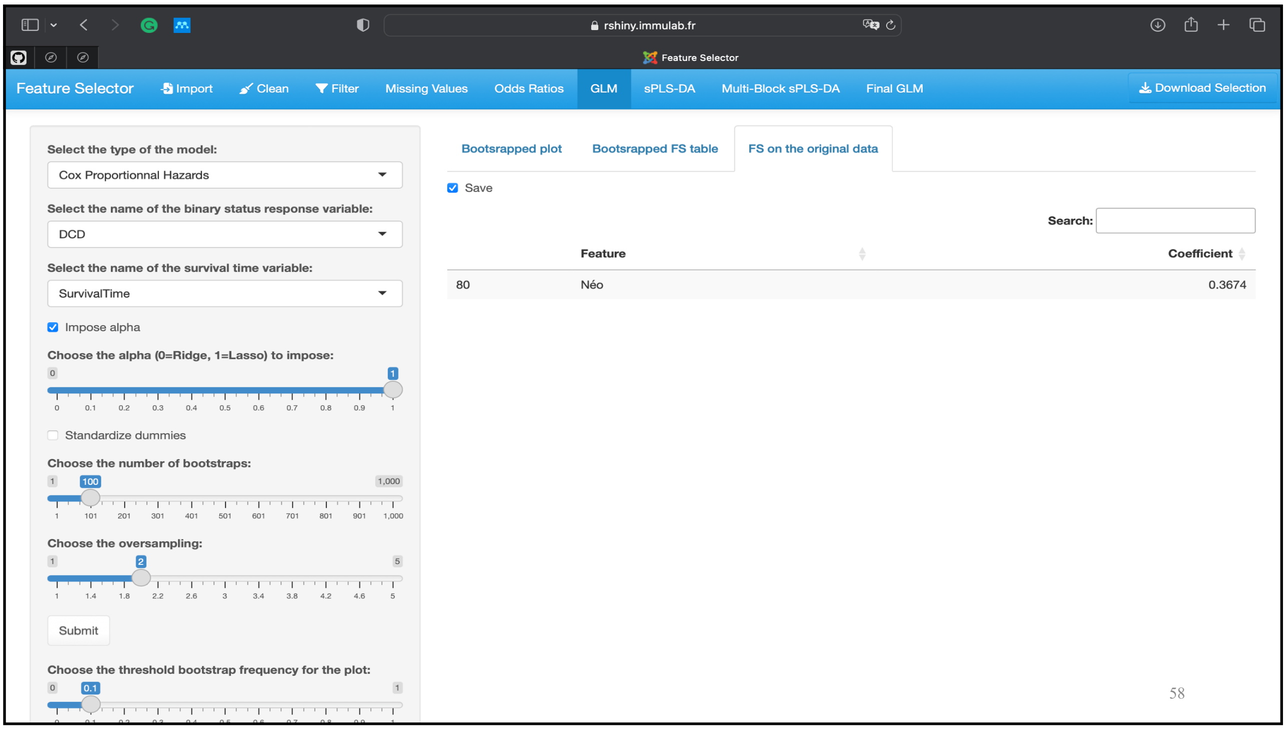
55


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Cox Proportionnal Hazards
Select the name of the binary status response variable:
DCD
Select the name of the survival time variable:
SurvivalTime
Impose alpha
Choose the alpha (0=Ridge, 1=Lasso) to impose:
Standardize dummies
Choose the number of bootstraps:
Choose the oversampling:
Submit
Choose the threshold bootstrap frequency for the plot:
Bootstrapped plot
Bootstrapped FS table
FS on the original data
Save
png
jpg
svg
pdf
Selected features coming up in more than 10%
of bootstrapped models
Néo
% MOYENNE BmsW
% late
Creatinine
Poids
%CD8M CD57+
% inter
Eosino
IL1B
nLate
%CD8M Ki67+
vitamine D (ng/ml)
nNKCD3+
IP10
Frequency
0.00
0.25
0.50
0.75
56

56


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Cox Proportionnal Hazards
Select the name of the binary status response variable:
DCD
Select the name of the survival time variable:
SurvivalTime
Impose alpha
Choose the alpha (0=Ridge, 1=Lasso) to impose:
Standardize dummies
Choose the number of bootstraps:
Choose the oversampling:
Submit
Choose the threshold bootstrap frequency for the plot:
Bootsrapped plot
Bootsrapped FS table
FS on the original data
Save
Search:
Feature
Frequency
80 Néo 0.88
63 % MOYENNE BmsW 0.38
17 Creatinine 0.31
58 % late 0.31
16 Poids 0.3
54 %CD8M CD57+ 0.29
57 % inter 0.28
42 Eosino 0.27
73 IL1B 0.26
94 nLate 0.22
55 %CD8M Ki67+ 0.18
19 vitamine D (ng/ml) 0.14
35 nNKCD3+ 0.12
77 IP10 0.11
66 % CD14 + 0.09
57

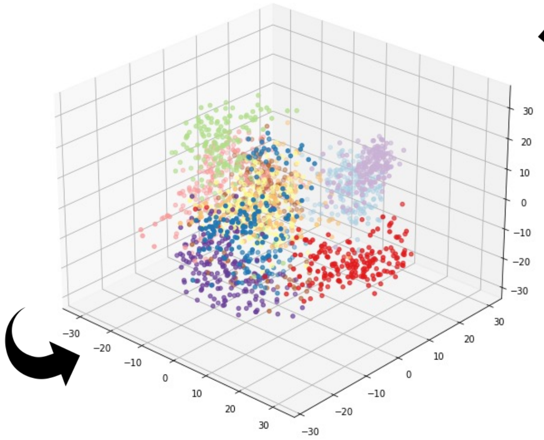
57


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Cox Proportionnal Hazards
Select the name of the binary status response variable:
DCD
Select the name of the survival time variable:
SurvivalTime
Impose alpha
Choose the alpha (0=Ridge, 1=Lasso) to impose:
Standardize dummies
Choose the number of bootstraps:
Choose the oversampling:
Submit
Choose the threshold bootstrap frequency for the plot:
Bootsrapped plot
Bootsrapped FS table
FS on the original data
Save
Search:
Feature
Coefficient
80 Néo 0.3674
58

58

Dimensionality reduction: Intuition

Different goals, but the same principle – find the "good perspective"



<u>PCA</u> – increase variation

<u>PLS-DA</u> – increase class separation

"New perspective's" coordinate axes
= **Components**
= linear combinations of features

https://www.atmosera.com/blog/principal-component-analysis/

59

59

sPLS-DA

Classification – prediction of categorical variable (binary or multinomial)

Assumptions:
- **linearity** – classes can be separated by linear combinations of variables
- no perfect multicollinearity
- independence of observations

- no strong outliers
- no extremely imbalanced classes

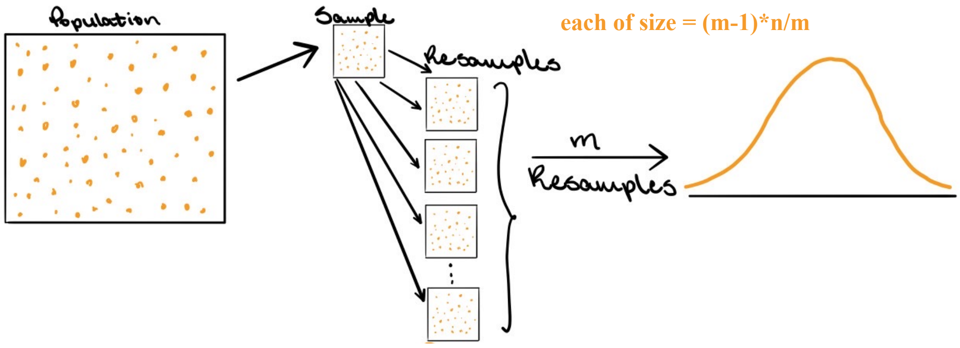
- only numeric predictors accepted

60

60

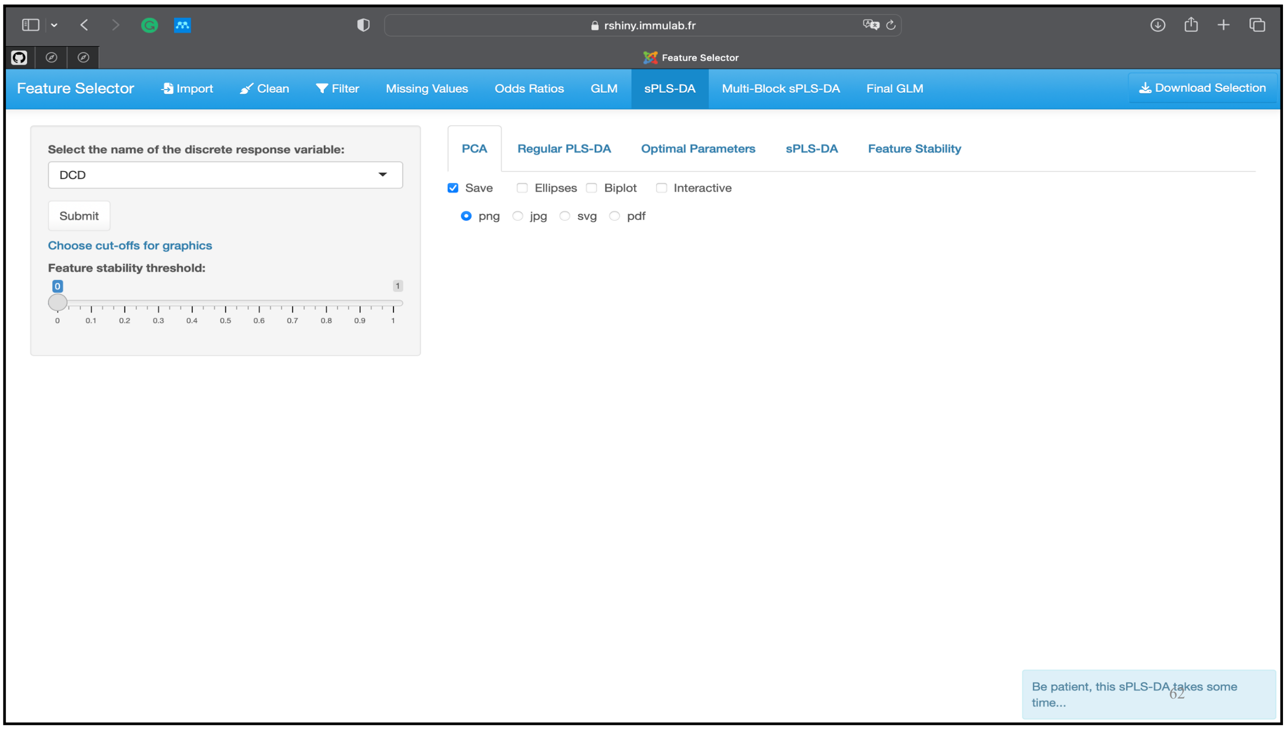
Model Robustness

Brute force simulation approach: Bootstrapping or **Cross-Validation**

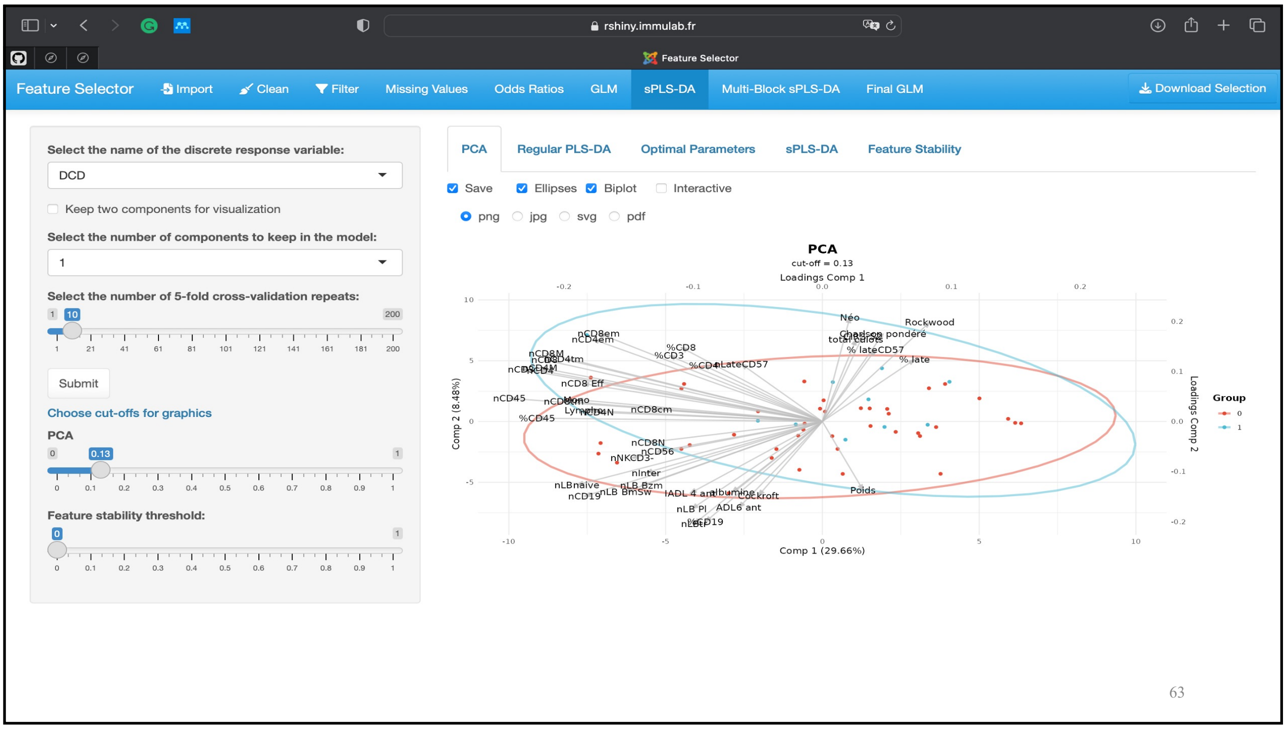


https://www.linkedin.com/pulse/bootstrapping-statistics-what-why-its-used-trist-n-joseph

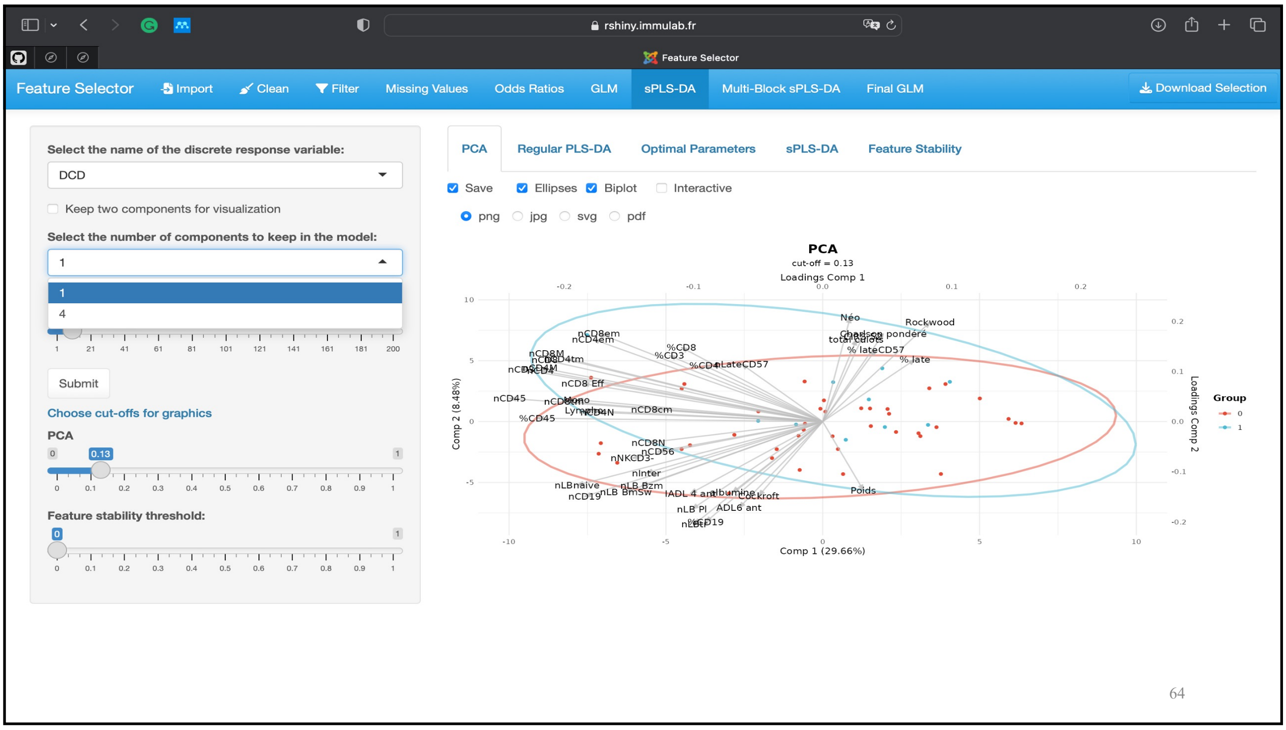
61



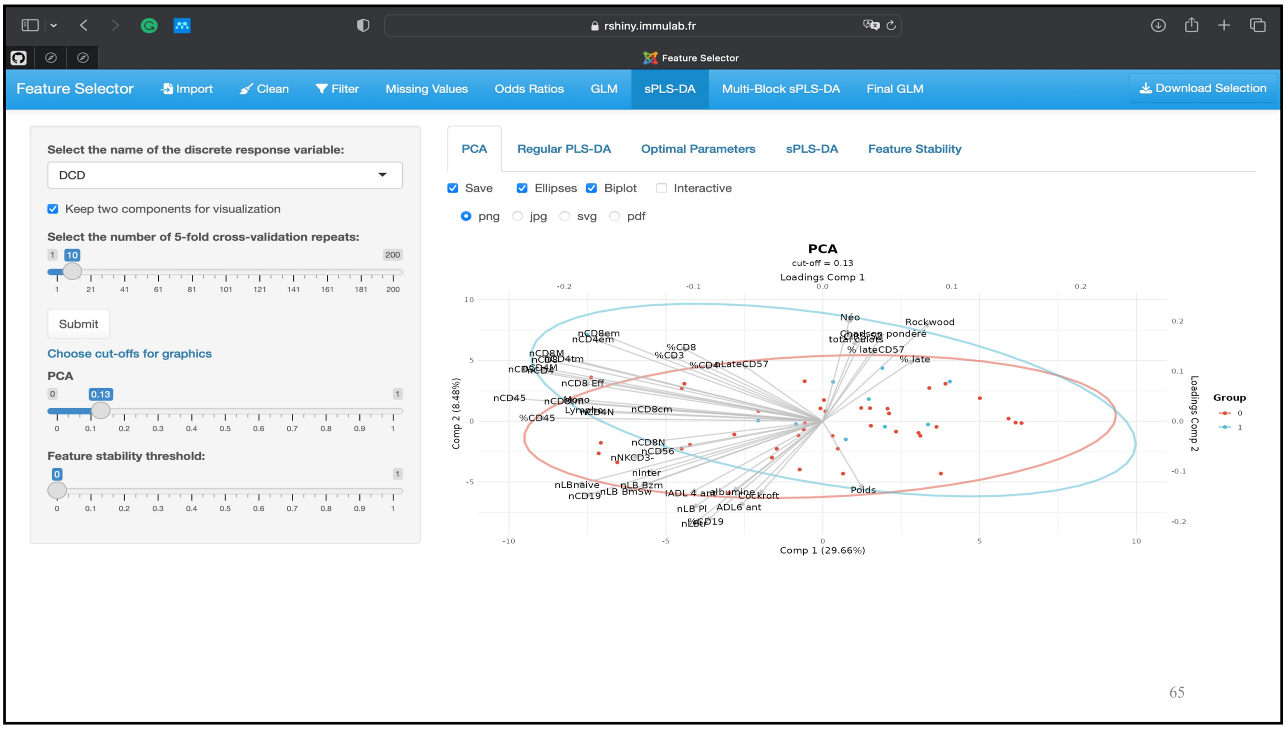
62


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of components to keep in the model:
1
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Feature stability threshold:
PCA
Regular PLS-DA
Optimal Parameters
sPLS-DA
Feature Stability
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
PCA
cut-off = 0.13
Loadings Comp 1
Comp 2 (8.48%)
Comp 1 (29.66%)
Loadings Comp 2
Group
63

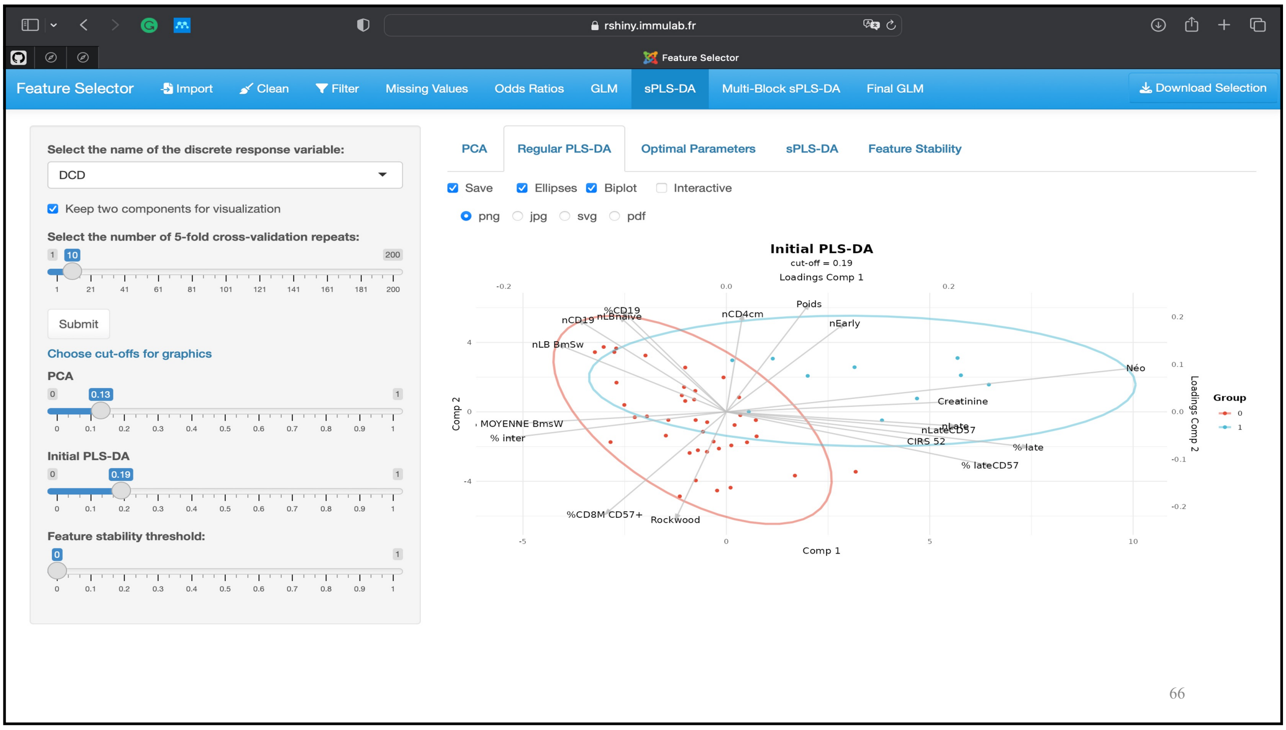
63


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of components to keep in the model:
1
1
4
Submit
Choose cut-offs for graphics
PCA
Feature stability threshold:
PCA
Regular PLS-DA
Optimal Parameters
sPLS-DA
Feature Stability
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
PCA
cut-off = 0.13
Loadings Comp 1
Comp 2 (8.48%)
Comp 1 (29.66%)
Loadings Comp 2
Group
64

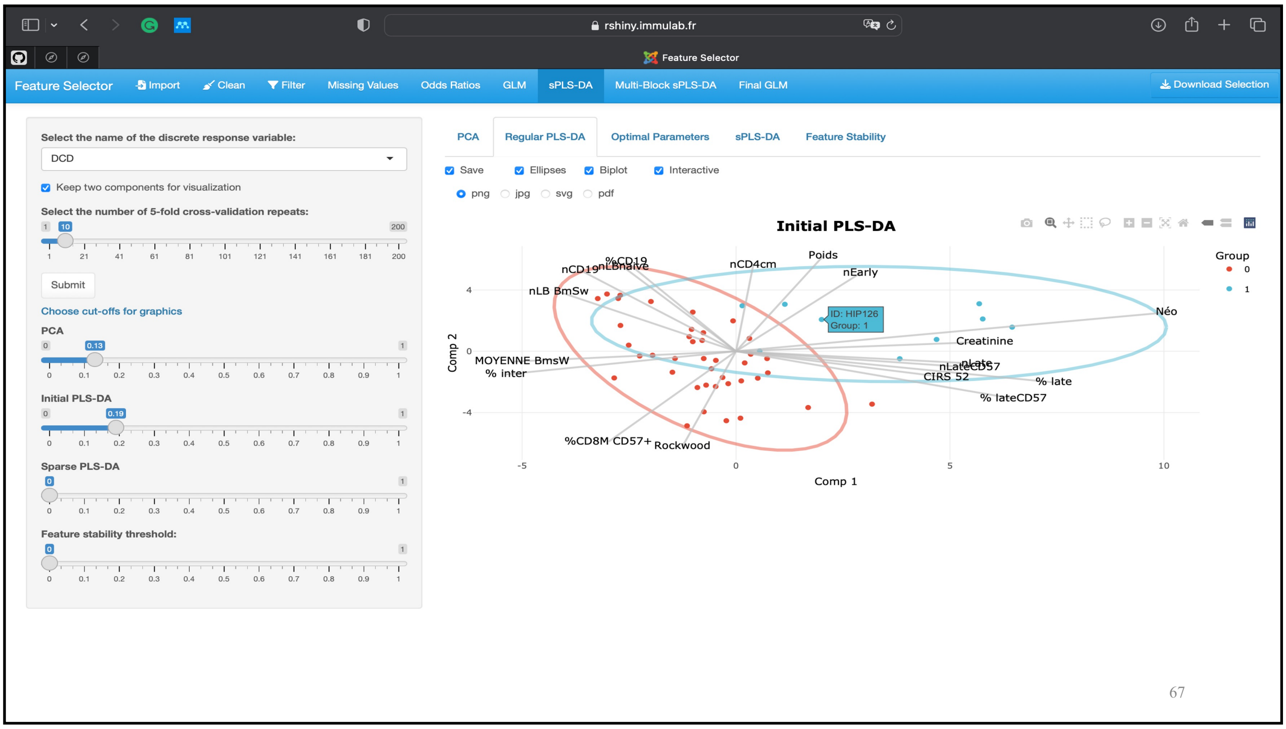
64


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Feature stability threshold:
PCA
Regular PLS-DA
Optimal Parameters
sPLS-DA
Feature Stability
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
PCA
cut-off = 0.13
Loadings Comp 1
Comp 1 (29.66%)
Comp 2 (8.48%)
Loadings Comp 2
Group
65

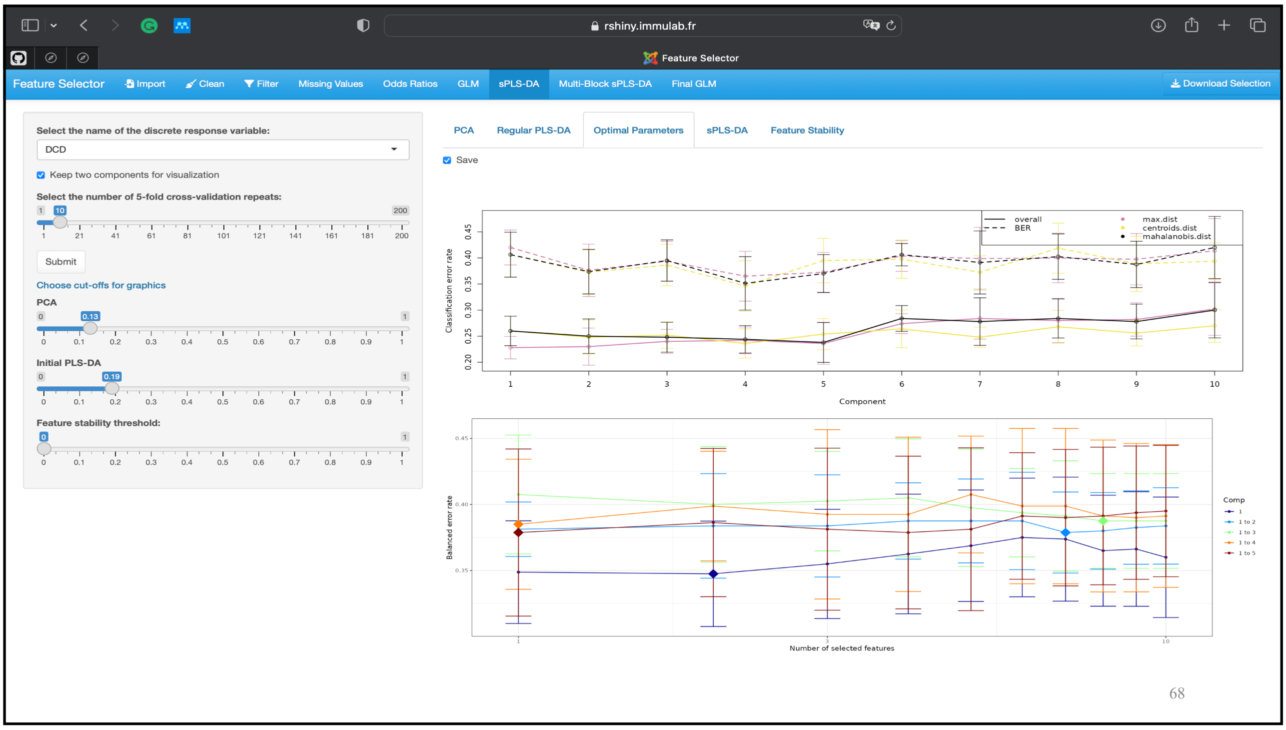
65


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Initial PLS-DA
Feature stability threshold:
PCA
Regular PLS-DA
Optimal Parameters
sPLS-DA
Feature Stability
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
Initial PLS-DA
cut-off = 0.19
Loadings Comp 1
Comp 1
Comp 2
Loadings Comp 2
Group
66

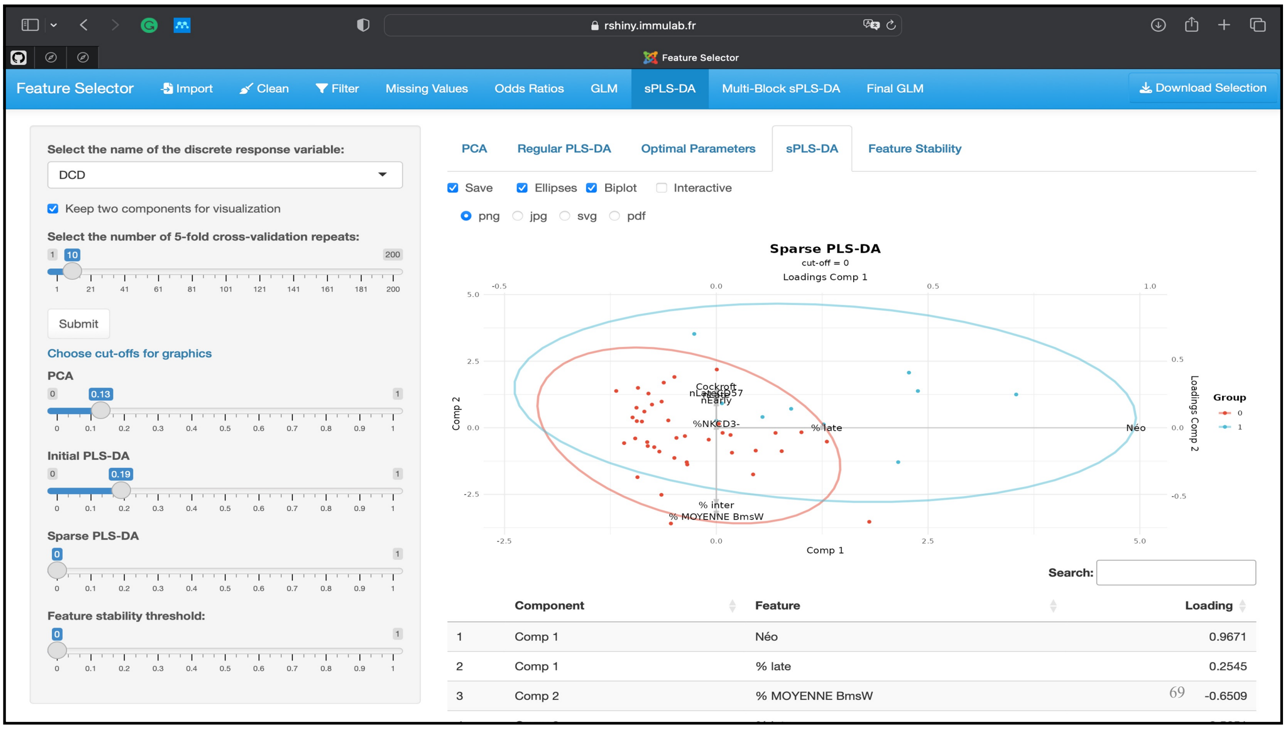
66


rshiny.immulab.fr
Feature Selector
Feature Selector Import Clean Filter Missing Values Odds Ratios GLM sPLS-DA Multi-Block sPLS-DA Final GLM Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Initial PLS-DA
Sparse PLS-DA
Feature stability threshold:
PCA Regular PLS-DA Optimal Parameters sPLS-DA Feature Stability
Save Ellipses Biplot Interactive
png jpg svg pdf
Initial PLS-DA
Group
Comp 1
Comp 2
67

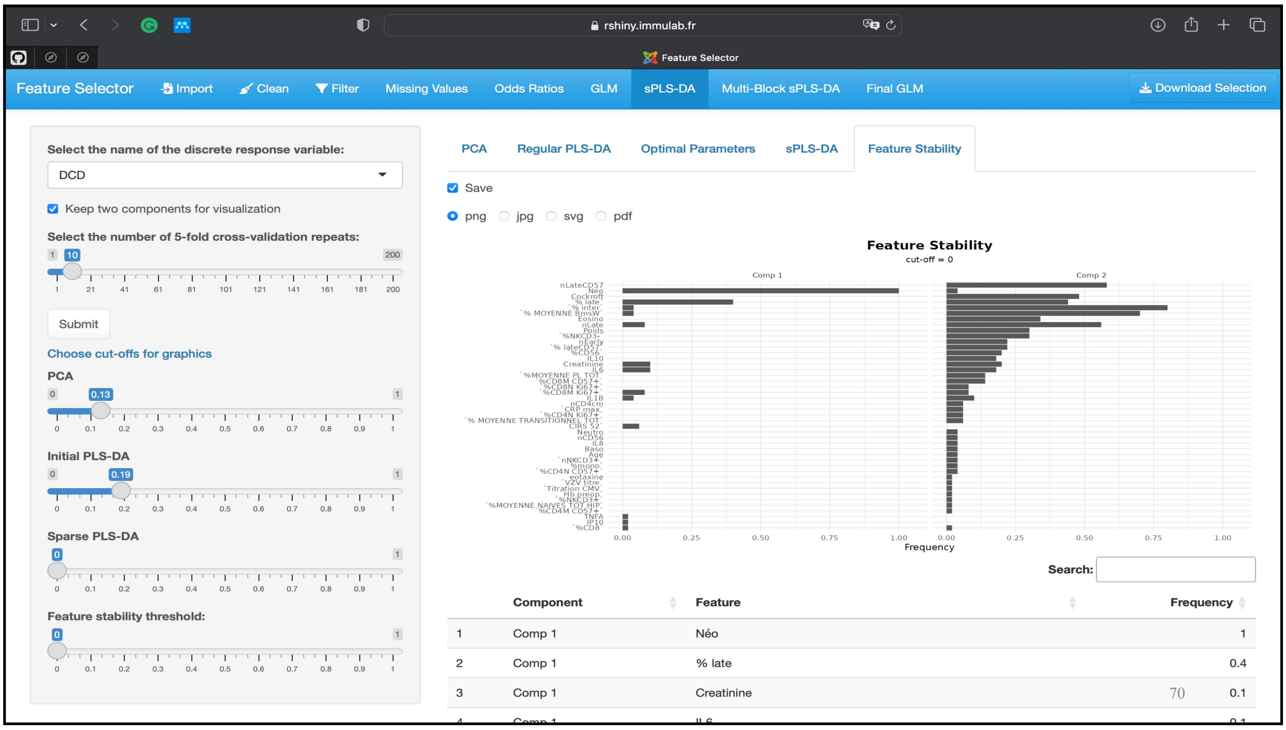
67


rshiny.immulab.fr
Feature Selector
Feature Selector Import Clean Filter Missing Values Odds Ratios GLM sPLS-DA Multi-Block sPLS-DA Final GLM Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Initial PLS-DA
Feature stability threshold:
PCA Regular PLS-DA Optimal Parameters sPLS-DA Feature Stability
Save
Classification error rate
Component
Balanced error rate
Number of selected features
Comp
68

68


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Initial PLS-DA
Sparse PLS-DA
Feature stability threshold:
PCA
Regular PLS-DA
Optimal Parameters
sPLS-DA
Feature Stability
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
Sparse PLS-DA
cut-off = 0
Loadings Comp 1
Comp 2
Comp 1
Loadings Comp 2
Group
Néo
% late
% inter
% MOYENNE BmsW
Search:
Component
Feature
Loading
1 Comp 1 Néo 0.9671
2 Comp 1 % late 0.2545
3 Comp 2 % MOYENNE BmsW -0.6509
69

69


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Keep two components for visualization
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PCA
Initial PLS-DA
Sparse PLS-DA
Feature stability threshold:
PCA
Regular PLS-DA
Optimal Parameters
sPLS-DA
Feature Stability
Save
png
jpg
svg
pdf
Feature Stability
cut-off = 0
Comp 1
Comp 2
Frequency
Search:
Component
Feature
Frequency
1 Comp 1 Néo 1
2 Comp 1 % late 0.4
3 Comp 1 Creatinine 0.1
70

70

Multi-Block sPLS-DA (DIABLO)

- classification + integration of information from different blocks, takes into account inter-block interactions

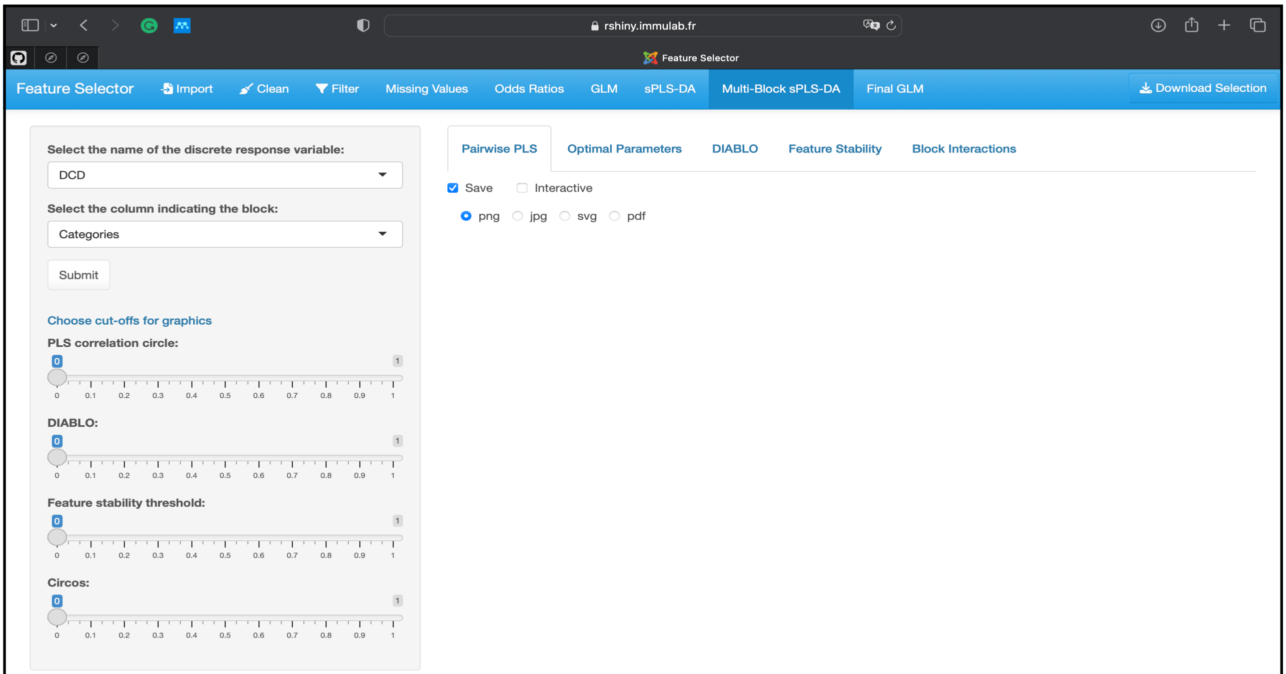
Assumptions: in addition to sPLS-DA

- linearity– classes can be separated by linear combinations of variables inside and across blocks

- Contribution of variables from different blocks to the same component might mean the relation to the same phenomenon (e.g. pathological process, metabolic pathway)

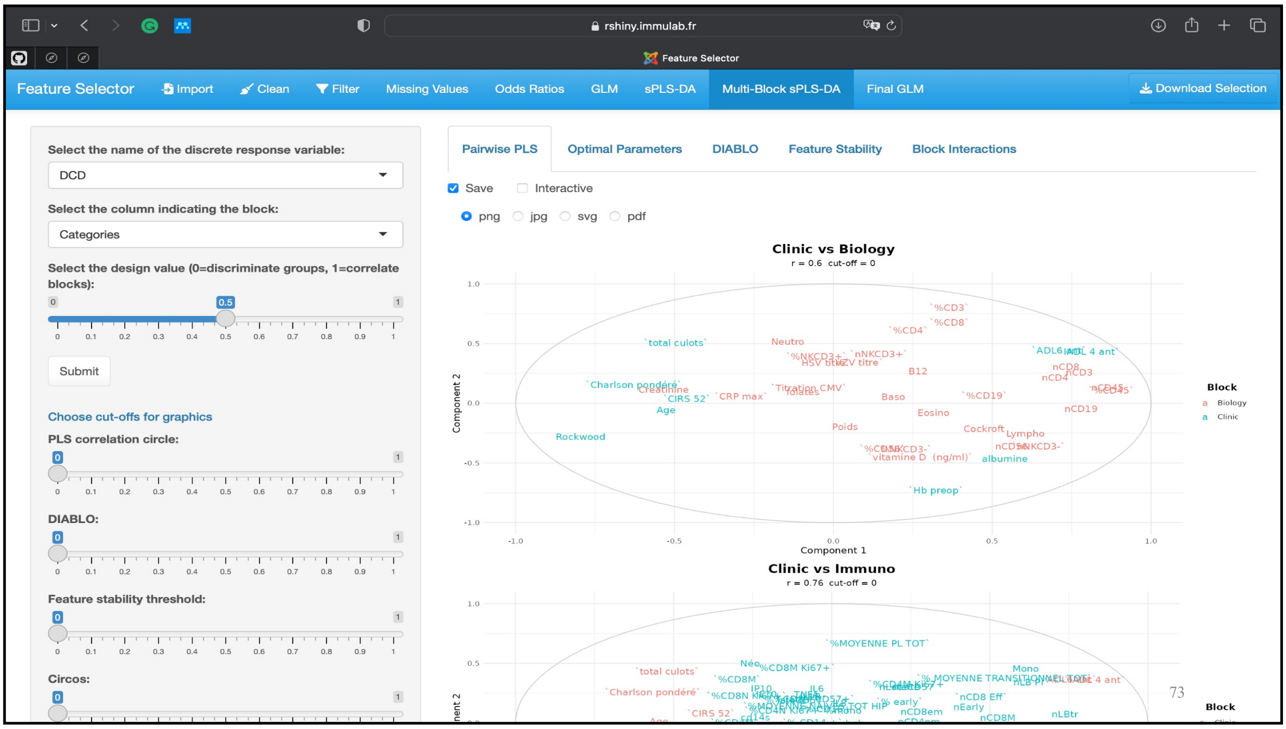
71

71

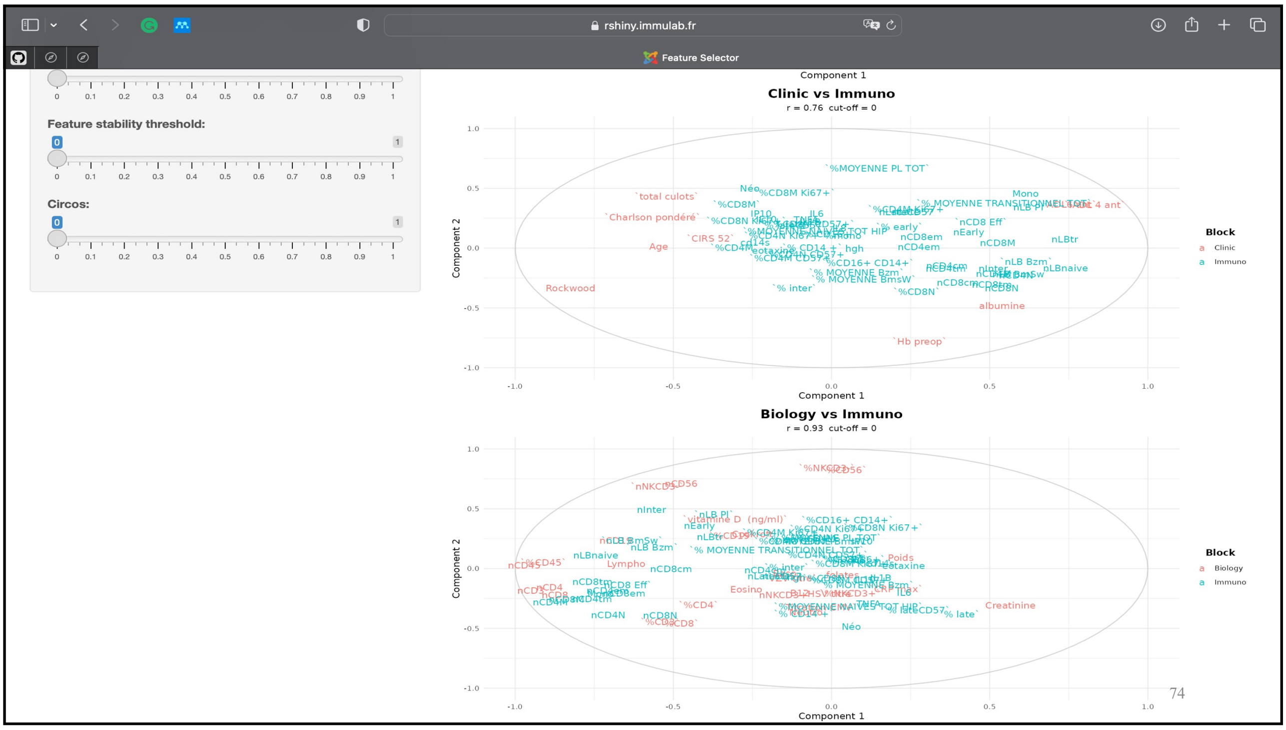


72

72


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Select the column indicating the block:
Categories
Select the design value (0=discriminate groups, 1=correlate blocks):
Submit
Choose cut-offs for graphics
PLS correlation circle:
DIABLO:
Feature stability threshold:
Circos:
Pairwise PLS
Optimal Parameters
DIABLO
Feature Stability
Block Interactions
Save
Interactive
png
jpg
svg
pdf
Clinic vs Biology
r = 0.6 cut-off = 0
Component 1
Component 2
Block
Biology
Clinic
Clinic vs Immuno
r = 0.76 cut-off = 0
73

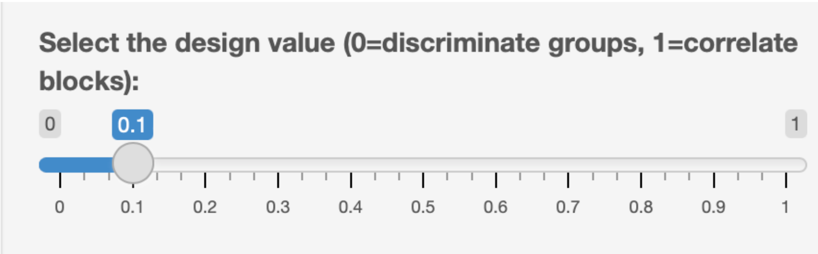
73


rshiny.immulab.fr
Feature Selector
Feature stability threshold:
Circos:
Component 1
Clinic vs Immuno
r = 0.76 cut-off = 0
Component 2
Block
Clinic
Immuno
Biology vs Immuno
r = 0.93 cut-off = 0
Block
Biology
Immuno
74

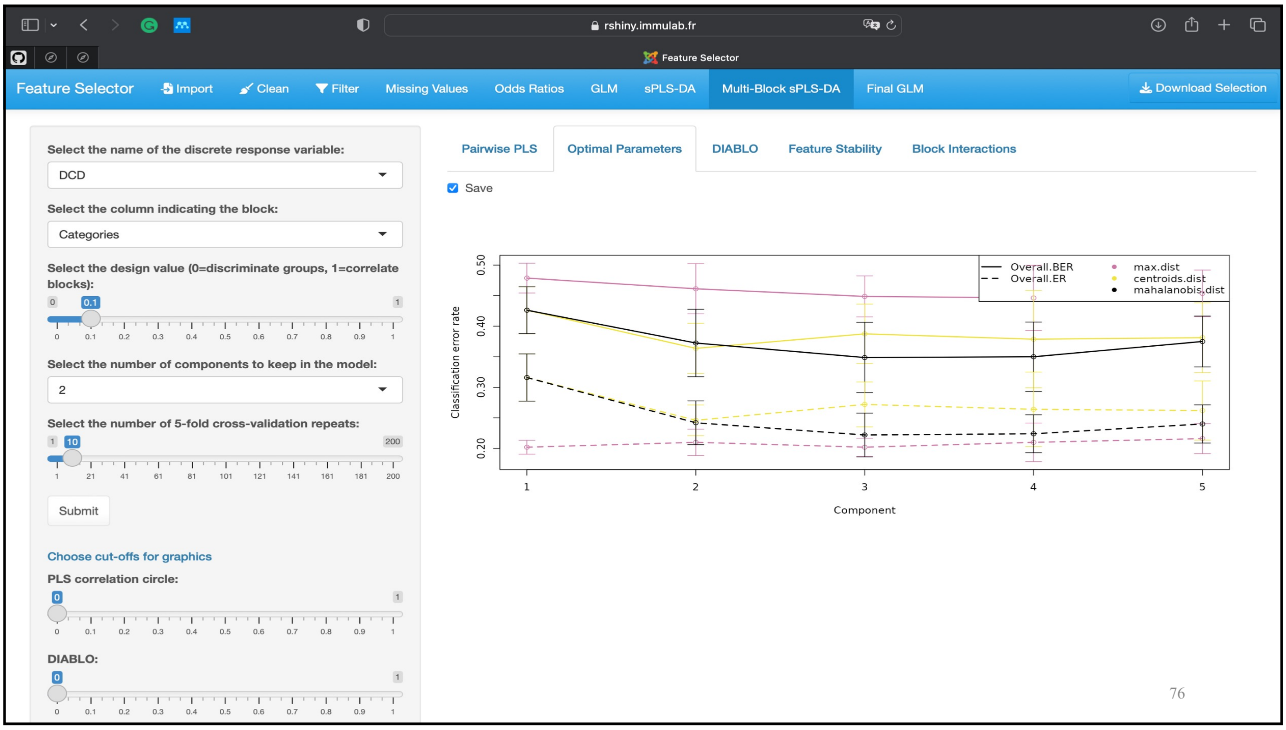
74

Choose the design value

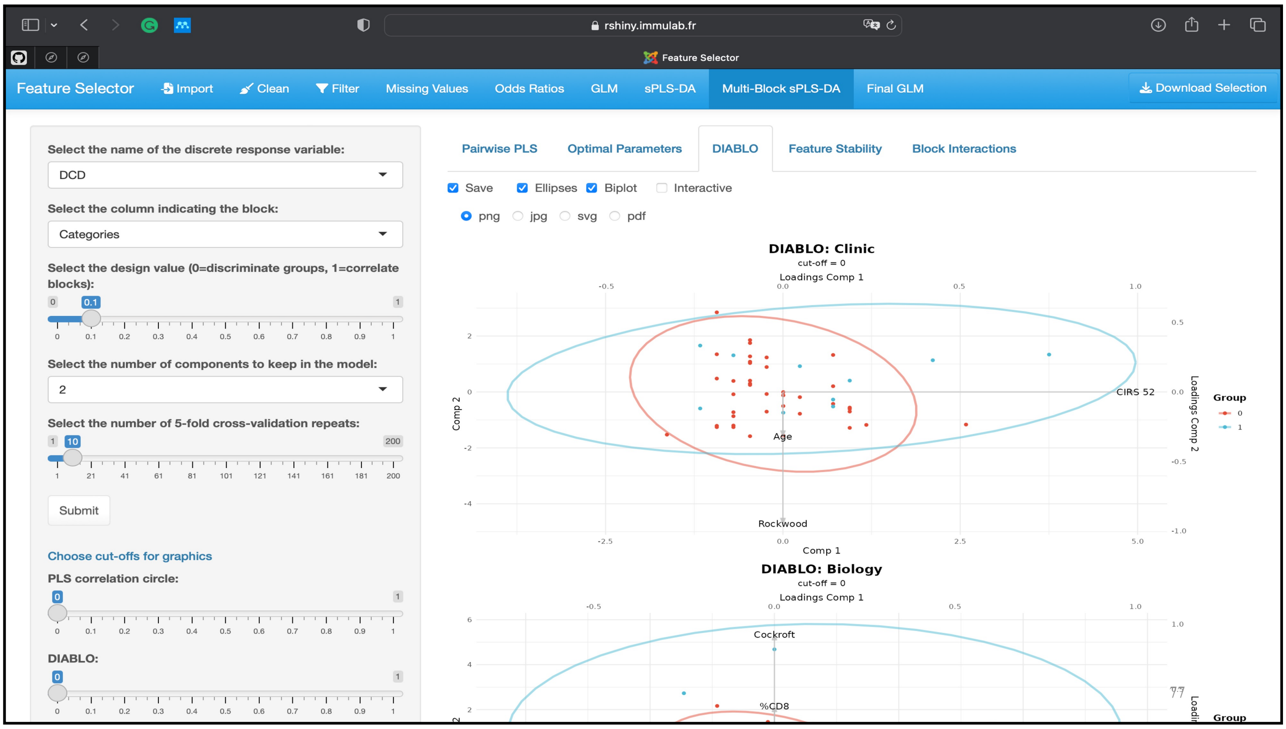
- Based on PLS results and the aim of the study
- From 0 to 1
- 0 - maximize discriminative ability
- 1 - maximize correlation between datasets



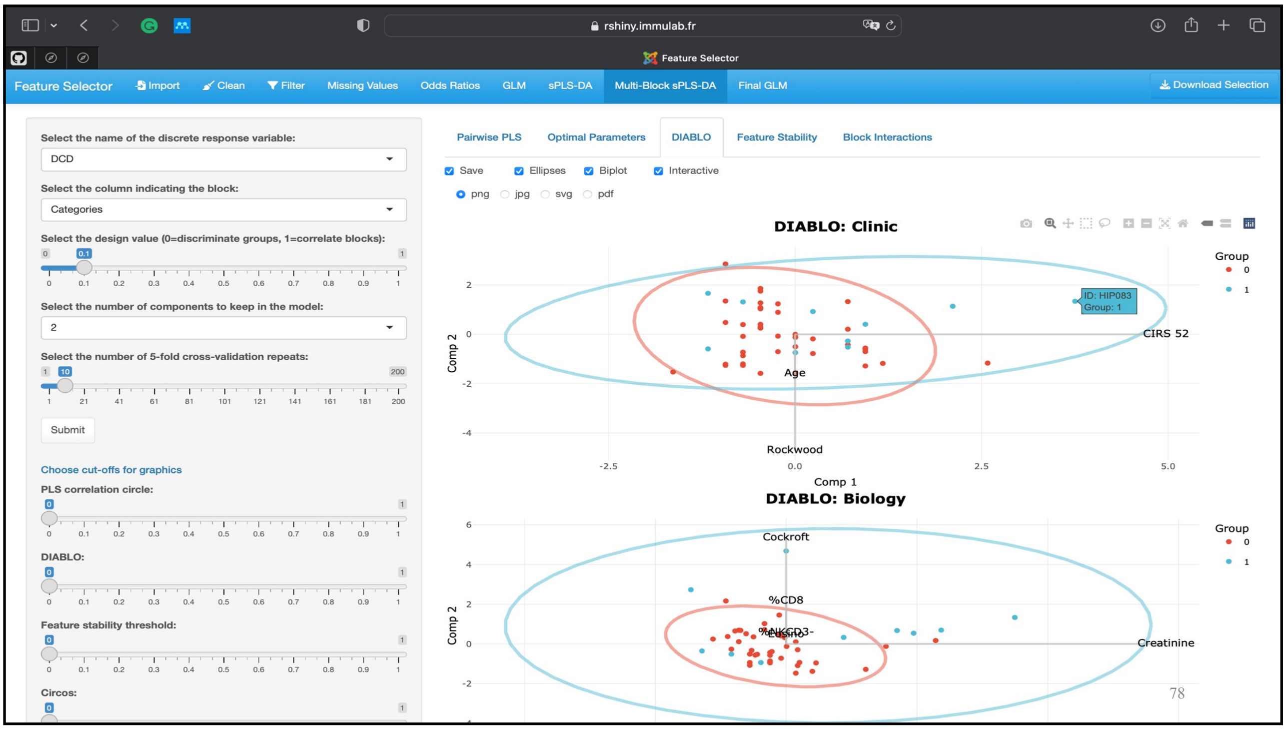
75



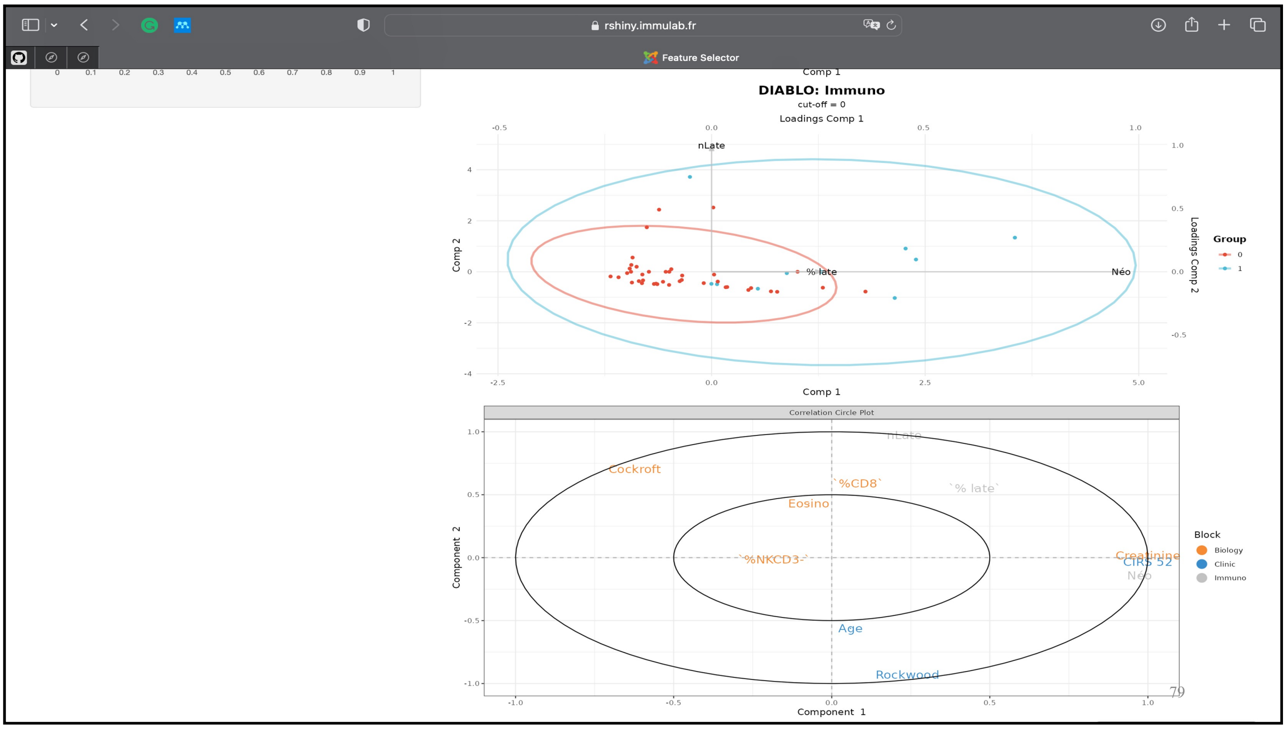
76


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Select the column indicating the block:
Categories
Select the design value (0=discriminate groups, 1=correlate blocks):
Select the number of components to keep in the model:
2
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PLS correlation circle:
DIABLO:
Pairwise PLS
Optimal Parameters
DIABLO
Feature Stability
Block Interactions
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
DIABLO: Clinic
cut-off = 0
Loadings Comp 1
CIRS 52
Age
Rockwood
Comp 1
Comp 2
Loadings Comp 2
Group
DIABLO: Biology
cut-off = 0
Loadings Comp 1
Cockroft
%CD8

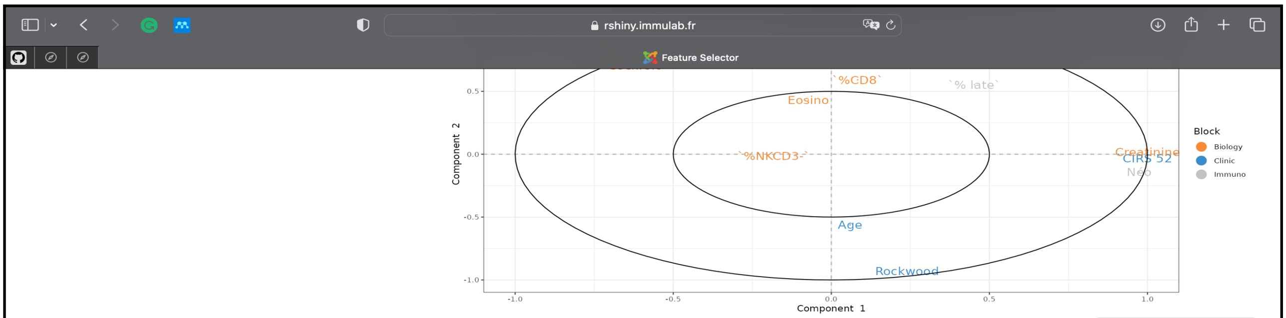
77


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Select the column indicating the block:
Categories
Select the design value (0=discriminate groups, 1=correlate blocks):
Select the number of components to keep in the model:
2
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PLS correlation circle:
DIABLO:
Feature stability threshold:
Circos:
Pairwise PLS
Optimal Parameters
DIABLO
Feature Stability
Block Interactions
Save
Ellipses
Biplot
Interactive
png
jpg
svg
pdf
DIABLO: Clinic
Group
ID: HIP083
Group: 1
CIRS 52
Age
Rockwood
Comp 1
Comp 2
DIABLO: Biology
Cockroft
%CD8
Creatinine

78



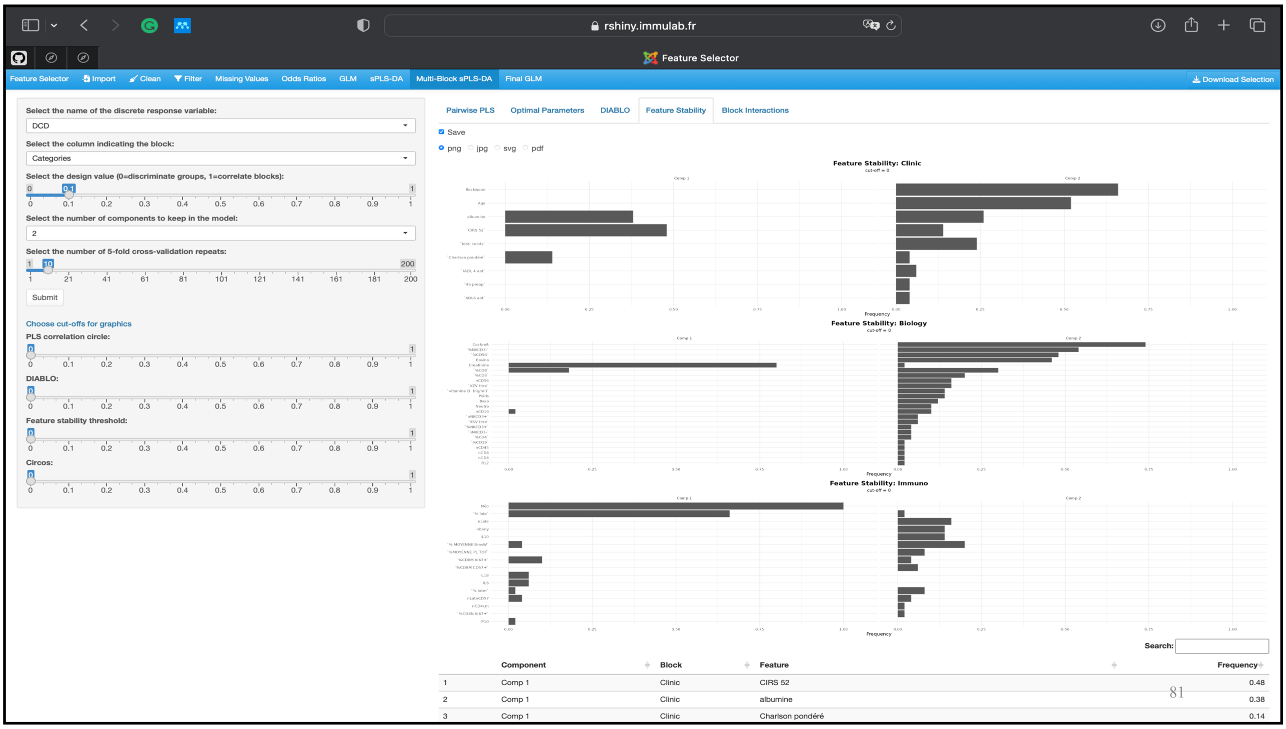
79



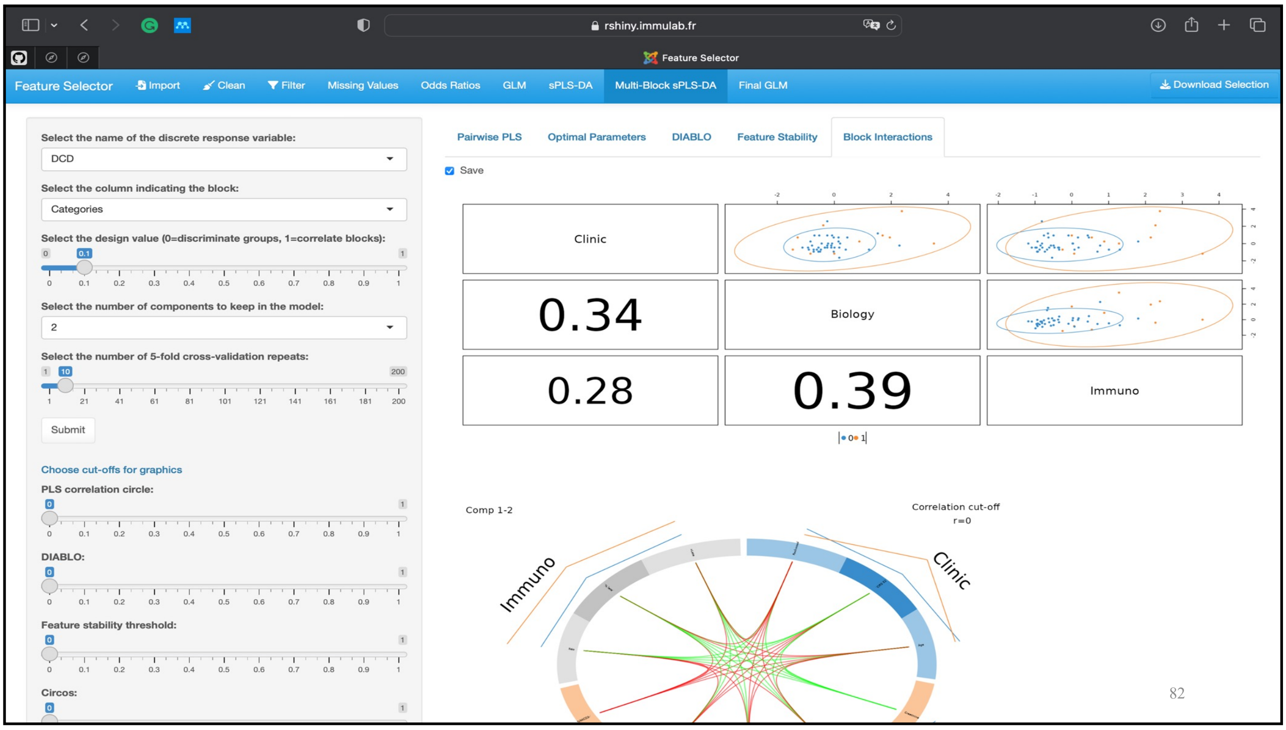
Search:

	Component	Block	Feature	Loading
1	Comp 1	Clinic	CIRS 52	1
2	Comp 1	Biology	Creatinine	1
3	Comp 1	Immuno	Néo	0.9658
4	Comp 1	Immuno	% late	0.2594
5	Comp 2	Clinic	Rockwood	-0.9478
6	Comp 2	Clinic	Age	-0.3188
7	Comp 2	Biology	Cockroft	0.9201
8	Comp 2	Biology	%CD8	0.371
9	Comp 2	Biology	%NKCD3-	0.0964
10	Comp 2	Biology	Eosino	0.0802
11	Comp 2	Immuno	nLate	1

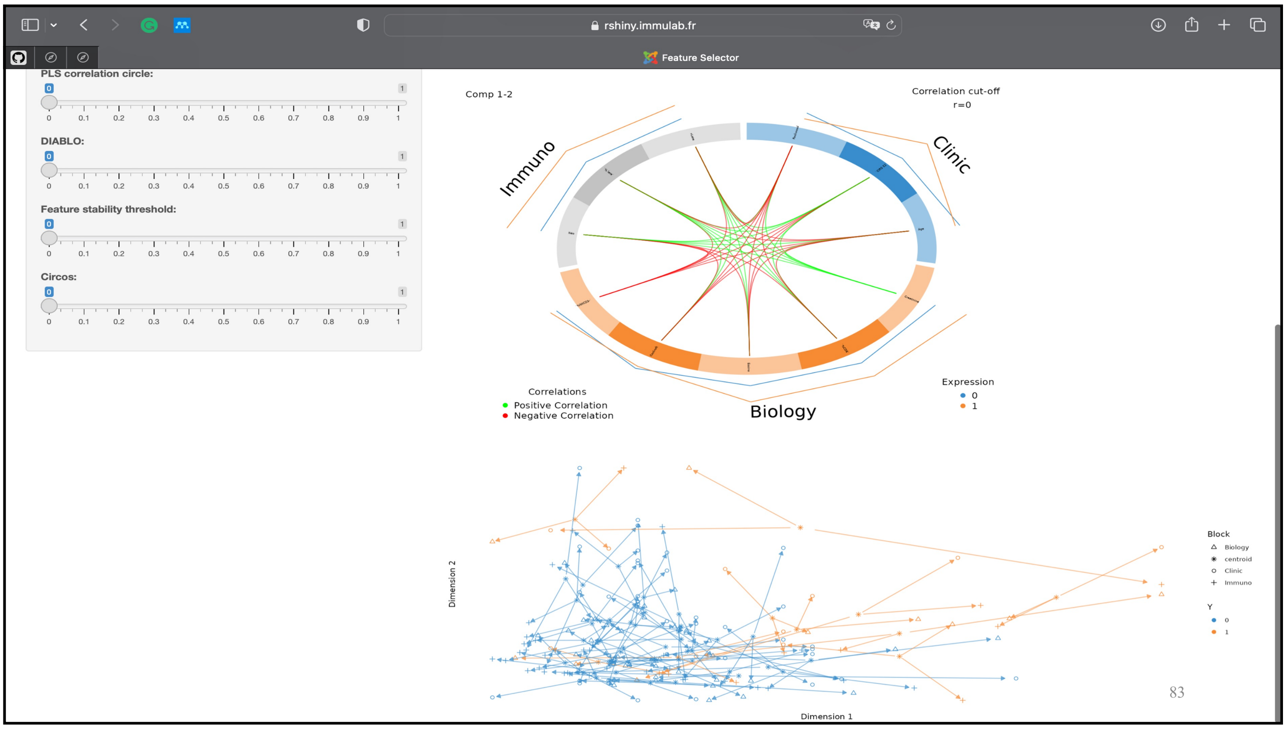
80


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Select the column indicating the block:
Categories
Select the design value (0=discriminate groups, 1=correlate blocks):
Select the number of components to keep in the model:
2
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PLS correlation circle:
DIABLO:
Feature stability threshold:
Circos:
Pairwise PLS
Optimal Parameters
DIABLO
Feature Stability
Block Interactions
Save
png
jpg
svg
pdf
Feature Stability: Clinic
Feature Stability: Biology
Feature Stability: Immuno
Search:
Component
Block
Feature
Frequency
1
Comp 1
Clinic
CIRS 52
0.48
2
Comp 1
Clinic
albumine
0.38
3
Comp 1
Clinic
Charlson pondéré
0.14
81

81


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the name of the discrete response variable:
DCD
Select the column indicating the block:
Categories
Select the design value (0=discriminate groups, 1=correlate blocks):
Select the number of components to keep in the model:
2
Select the number of 5-fold cross-validation repeats:
Submit
Choose cut-offs for graphics
PLS correlation circle:
DIABLO:
Feature stability threshold:
Circos:
Pairwise PLS
Optimal Parameters
DIABLO
Feature Stability
Block Interactions
Save
Clinic
0.34
Biology
0.28
0.39
Immuno
Comp 1-2
Correlation cut-off
r=0
Immuno
Clinic
82

82

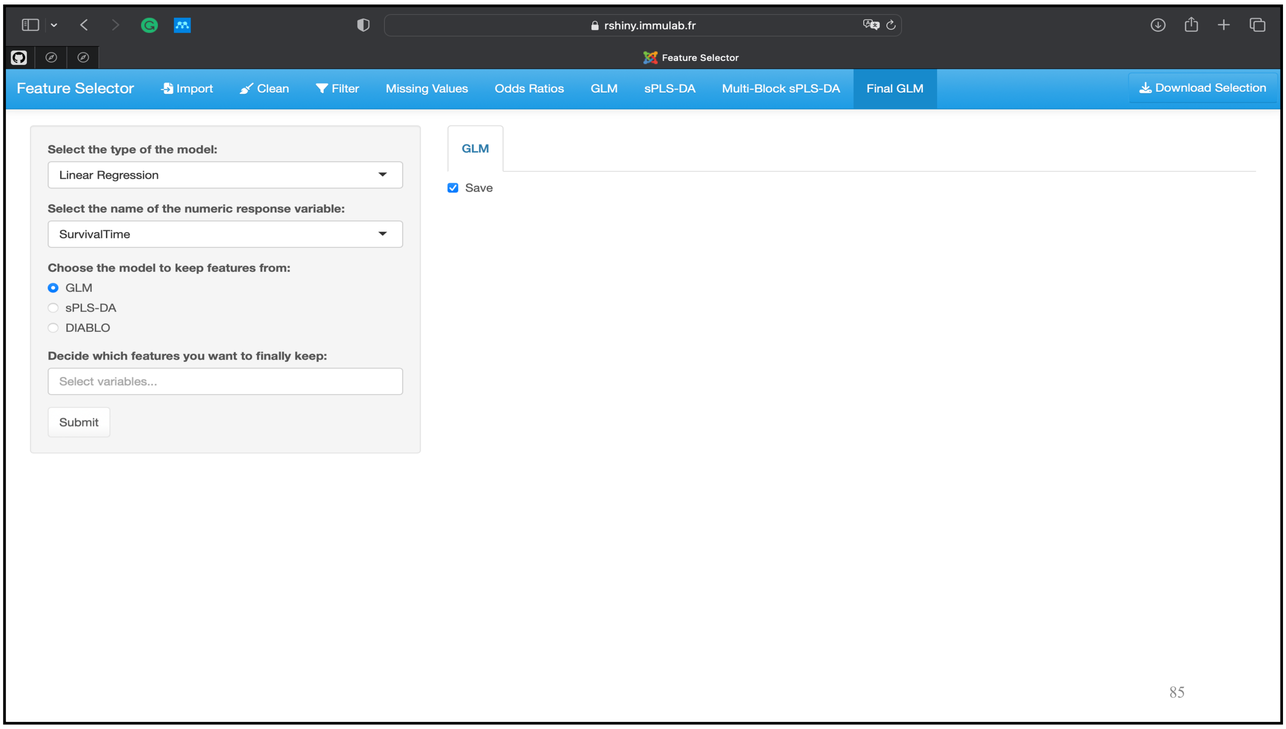


83

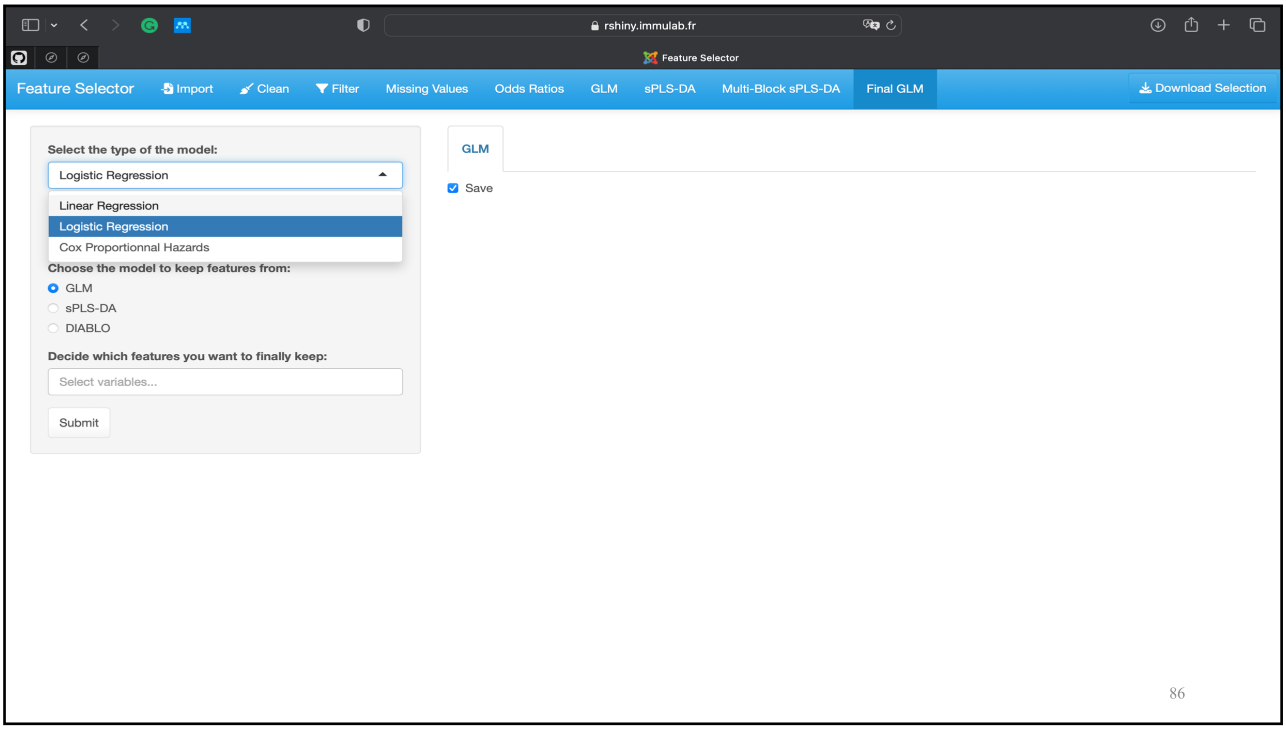
Final GLM

84

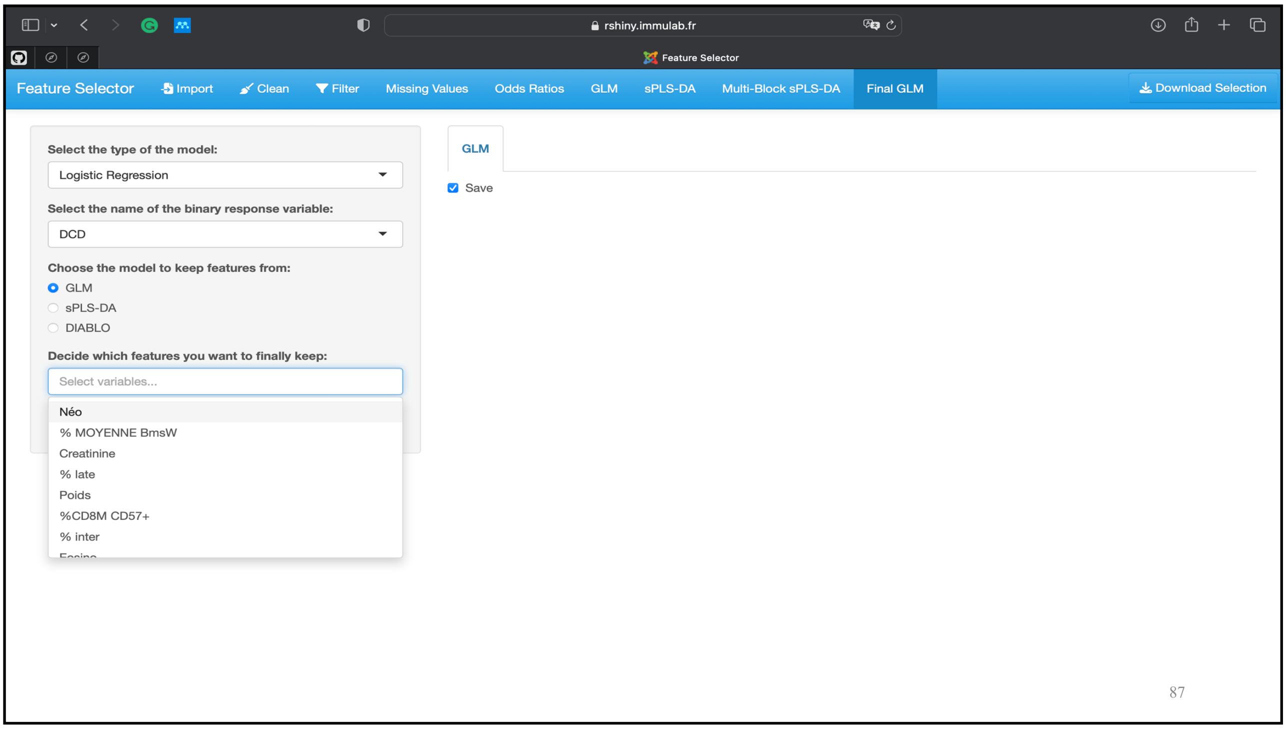
84


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Linear Regression
Select the name of the numeric response variable:
SurvivalTime
Choose the model to keep features from:
GLM
sPLS-DA
DIABLO
Decide which features you want to finally keep:
Select variables...
Submit
GLM
Save
85

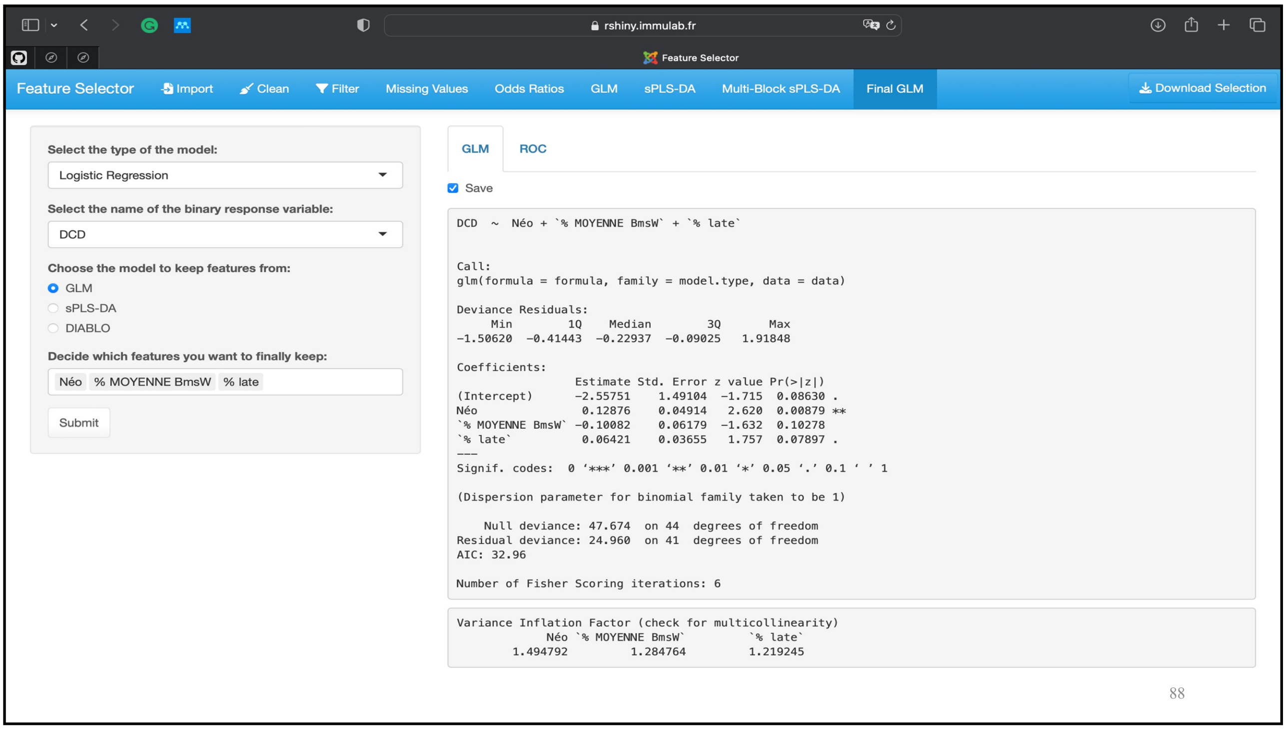
85


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Logistic Regression
Linear Regression
Logistic Regression
Cox Proportionnal Hazards
Choose the model to keep features from:
GLM
sPLS-DA
DIABLO
Decide which features you want to finally keep:
Select variables...
Submit
GLM
Save
86

86


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Logistic Regression
Select the name of the binary response variable:
DCD
Choose the model to keep features from:
GLM
sPLS-DA
DIABLO
Decide which features you want to finally keep:
Select variables...
Néo
% MOYENNE BmsW
Creatinine
% late
Poids
%CD8M CD57+
% inter
GLM
Save
87

87


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Logistic Regression
Select the name of the binary response variable:
DCD
Choose the model to keep features from:
GLM
sPLS-DA
DIABLO
Decide which features you want to finally keep:
Néo
% MOYENNE BmsW
% late
Submit
GLM
ROC
Save
DCD ~ Néo + `% MOYENNE BmsW` + `% late`
Call:
glm(formula = formula, family = model.type, data = data)
Deviance Residuals:
Min 1Q Median 3Q Max
-1.50620 -0.41443 -0.22937 -0.09025 1.91848
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.55751 1.49104 -1.715 0.08630 .
Néo 0.12876 0.04914 2.620 0.00879 **
`% MOYENNE BmsW` -0.10082 0.06179 -1.632 0.10278
`% late` 0.06421 0.03655 1.757 0.07897 .

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 47.674 on 44 degrees of freedom
Residual deviance: 24.960 on 41 degrees of freedom
AIC: 32.96
Number of Fisher Scoring iterations: 6
Variance Inflation Factor (check for multicollinearity)
Néo `% MOYENNE BmsW` `% late`
1.494792 1.284764 1.219245
88

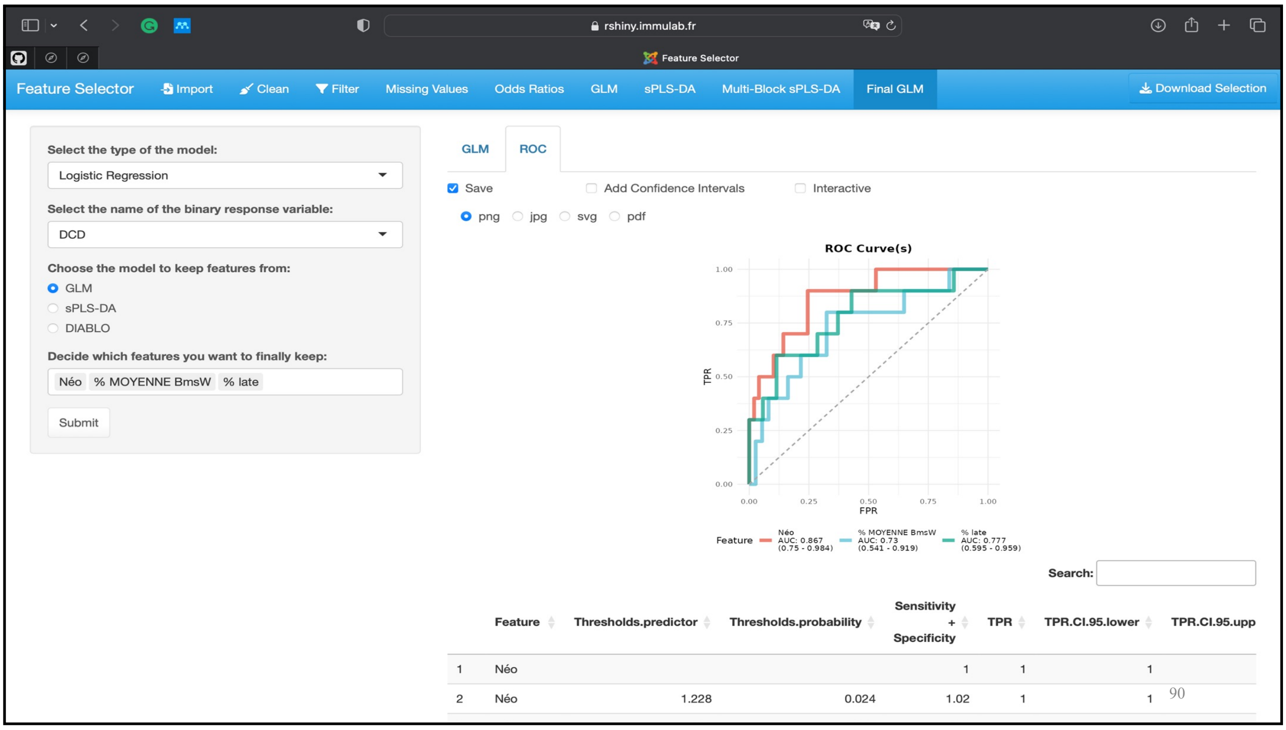
88

ROC curves

* in case of binary response

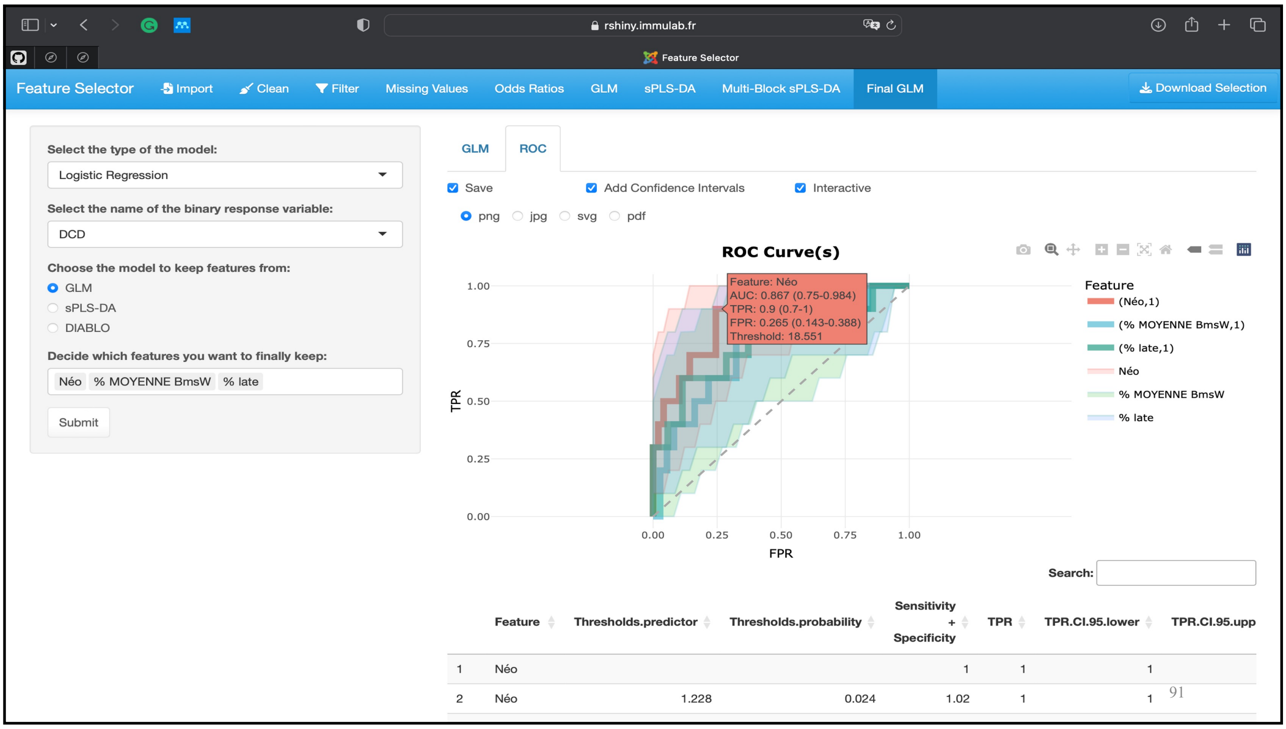
89

89



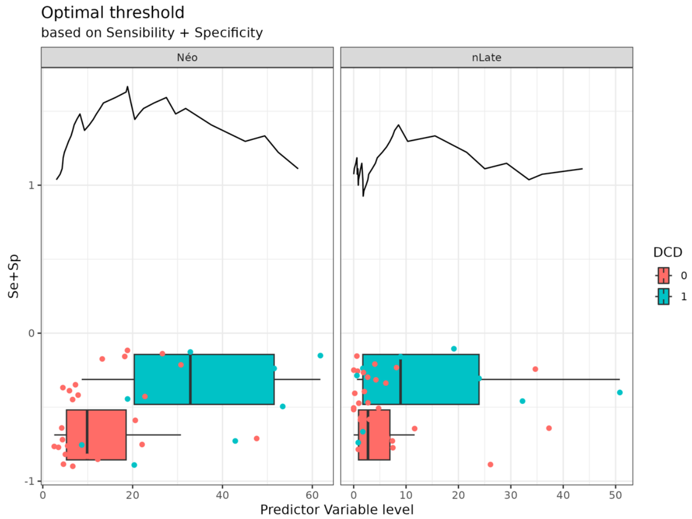
90

90



91

Threshold visualisation



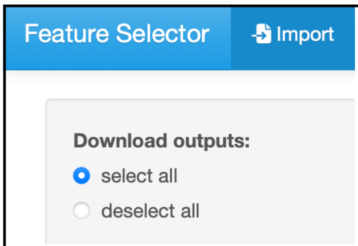
92

92

Download Selection

93

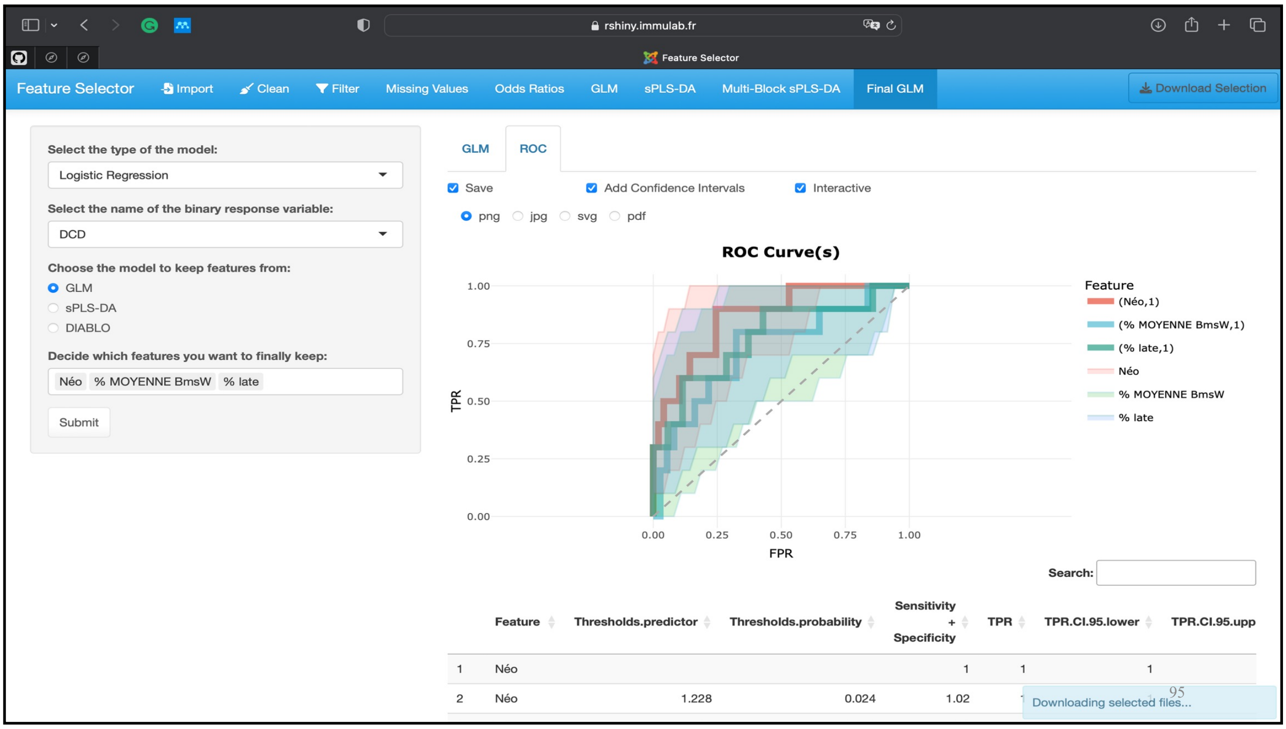
93



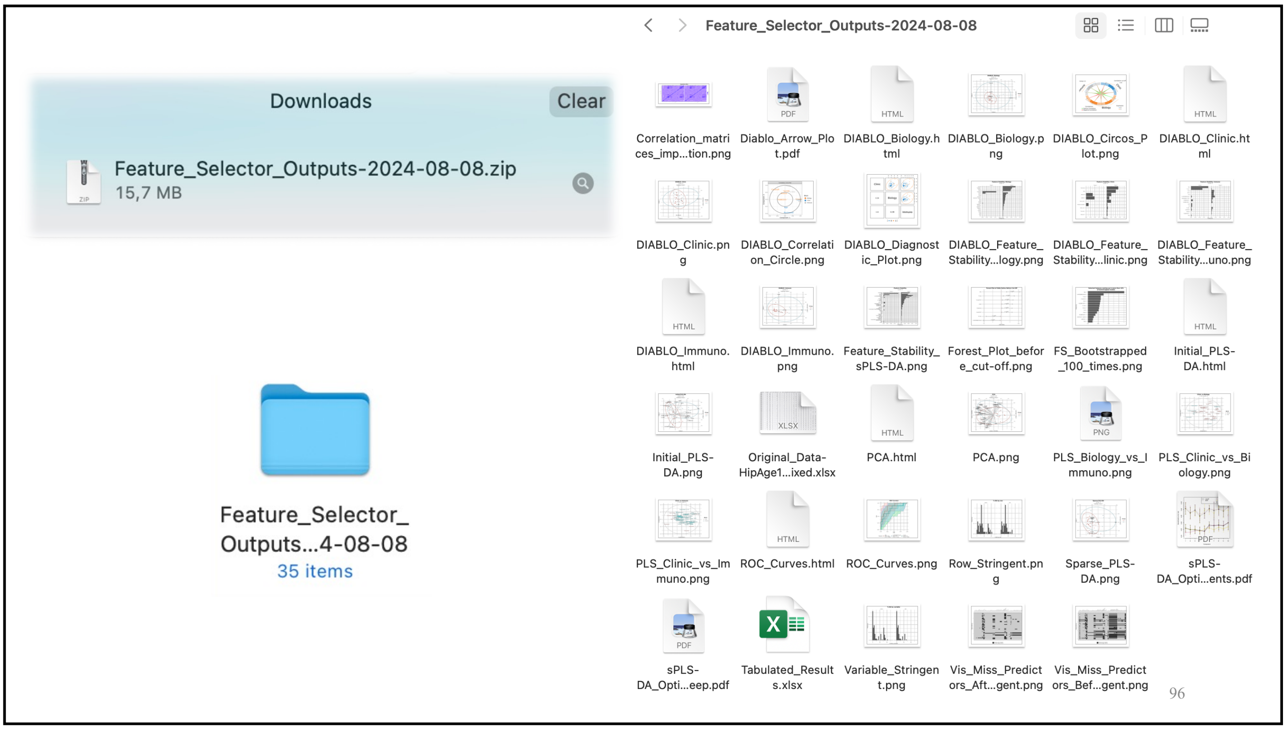
Download Selection

94

94


rshiny.immulab.fr
Feature Selector
Feature Selector
Import
Clean
Filter
Missing Values
Odds Ratios
GLM
sPLS-DA
Multi-Block sPLS-DA
Final GLM
Download Selection
Select the type of the model:
Logistic Regression
Select the name of the binary response variable:
DCD
Choose the model to keep features from:
GLM
sPLS-DA
DIABLO
Decide which features you want to finally keep:
Néo % MOYENNE BmsW % late
Submit
GLM
ROC
Save
Add Confidence Intervals
Interactive
png jpg svg pdf
ROC Curve(s)
TPR
FPR
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
Feature
(Néo,1)
(% MOYENNE BmsW,1)
(% late,1)
Néo
% MOYENNE BmsW
% late
Search:
Feature
Thresholds.predictor
Thresholds.probability
Sensitivity + Specificity
TPR
TPR.CI.95.lower
TPR.CI.95.upp
1 Néo 1 1 1
2 Néo 1.228 0.024 1.02
Downloading selected files...
95

95

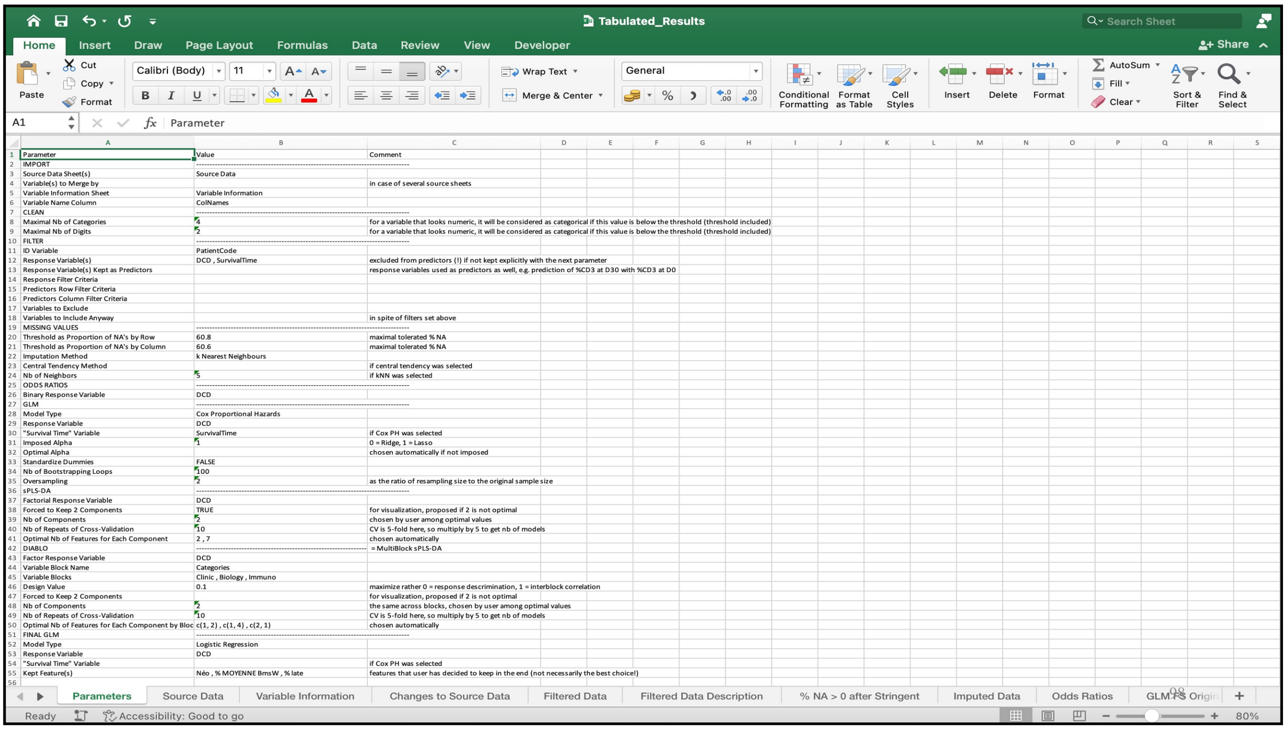

Feature_Selector_Outputs-2024-08-08
Downloads
Clear
Feature_Selector_Outputs-2024-08-08.zip
15,7 MB
Feature_Selector_Outputs...4-08-08
35 items
Correlation_matrices_imp...tion.png
Diablo_Arrow_Plot.pdf
DIABLO_Biology.html
DIABLO_Biology.png
DIABLO_Circos_Plot.png
DIABLO_Clinic.html
DIABLO_Clinic.png
DIABLO_Correlation_Circle.png
DIABLO_Diagnostic_Plot.png
DIABLO_Feature_Stability...logy.png
DIABLO_Feature_Stability...linic.png
DIABLO_Feature_Stability...uno.png
DIABLO_Immuno.html
DIABLO_Immuno.png
Feature_Stability_sPLS-DA.png
Forest_Plot_before_cut-off.png
FS_Bootstrapped_100_times.png
Initial_PLS-DA.html
Initial_PLS-DA.png
Original_Data-HipAge1...ixed.xlsx
PCA.html
PCA.png
PLS_Biology_vs_Immuno.png
PLS_Clinic_vs_Biology.png
PLS_Clinic_vs_Immuno.png
ROC_Curves.html
ROC_Curves.png
Row_Stringent.png
Sparse_PLS-DA.png
sPLS-DA_Opti...ents.pdf
sPLS-DA_Opti...eep.pdf
Tabulated_Results.xlsx
Variable_Stringent.png
Vis_Miss_Predictors_Aft...gent.png
Vis_Miss_Predictors_Bef...gent.png
96

96

Reproducibility

97

97



	Parameter	Value	Comment
1	Parameter	Value	Comment
2	IMPORT	--	
3	Source Data Sheet(s)	Source Data	
4	Variable(s) to Merge by		in case of several source sheets
5	Variable Information Sheet	Variable Information	
6	Variable Name Column	ColNames	
7	CLEAN	--	
8	Maximal Nb of Categories	4	for a variable that looks numeric, it will be considered as categorical if this value is below the threshold (threshold included)
9	Maximal Nb of Digits	2	for a variable that looks numeric, it will be considered as categorical if this value is below the threshold (threshold included)
10	FILTER	--	
11	ID Variable	PatientCode	
12	Response Variable(s)	DCD , SurvivalTime	excluded from predictors (!) if not kept explicitly with the next parameter
13	Response Variable(s) Kept as Predictors		response variables used as predictors as well, e.g. prediction of %CD3 at D30 with %CD3 at D0
14	Response Filter Criteria		
15	Predictors Row Filter Criteria		
16	Predictors Column Filter Criteria		
17	Variables to Exclude		
18	Variables to Include Anyway		in spite of filters set above
19	MISSING VALUES	--	
20	Threshold as Proportion of NA's by Row	60.8	maximal tolerated % NA
21	Threshold as Proportion of NA's by Column	60.6	maximal tolerated % NA
22	Imputation Method	k Nearest Neighbours	
23	Central Tendency Method		if central tendency was selected
24	Nb of Neighbors	5	if kNN was selected
25	ODDS RATIOS	--	
26	Binary Response Variable	DCD	
27	GLM	--	
28	Model Type	Cox Proportional Hazards	
29	Response Variable	DCD	
30	"Survival Time" Variable	SurvivalTime	if Cox PH was selected
31	Imposed Alpha	1	0 = Ridge, 1 = Lasso
32	Optimal Alpha		chosen automatically if not imposed
33	Standardize Dummies	FALSE	
34	Nb of Bootstrapping Loops	100	
35	Oversampling	2	as the ratio of resampling size to the original sample size
36	sPLS-DA	--	
37	Factorial Response Variable	DCD	
38	Forced to Keep 2 Components	TRUE	for visualization, proposed if 2 is not optimal
39	Nb of Components	2	chosen by user among optimal values
40	Nb of Repeats of Cross-Validation	10	CV is 5-fold here, so multiply by 5 to get nb of models
41	Optimal Nb of Features for Each Component	2 , 7	chosen automatically
42	DIABLO	--	= MultiBlock sPLS-DA
43	Factor Response Variable	DCD	
44	Variable Block Name	Categories	
45	Variable Blocks	Clinic , Biology , Immuno	
46	Design Value	0.1	maximize rather 0 = response descrimination, 1 = interblock correlation
47	Forced to Keep 2 Components		for visualization, proposed if 2 is not optimal
48	Nb of Components	2	the same across blocks, chosen by user among optimal values
49	Nb of Repeats of Cross-Validation	10	CV is 5-fold here, so multiply by 5 to get nb of models
50	Optimal Nb of Features for Each Component by Bloc	c(1, 2) , c(1, 4) , c(2, 1)	chosen automatically
51	FINAL GLM	--	
52	Model Type	Logistic Regression	
53	Response Variable	DCD	
54	"Survival Time" Variable		if Cox PH was selected
55	Kept Feature(s)	Néo , % MOYENNE BmsW , % late	features that user has decided to keep in the end (not necessarily the best choice!)
56			

98

Perspectives

- Other imputation methods
- Basic visualization (e.g. boxplots)
- Interactive KM / Cox curves
- Set seed to reproduce randomness

- User-friendly tooltips and help windows

99

99

Let’s try the app!

https://www.immulab.fr/cms/index.php/team/tools/lab-tools/feature-selector

100

100