

R Notebook

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*. Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

Hide

```
#install.packages("titanic")
library(titanic)
data(titanic_train)
View(titanic_train)
```

Hide

```
#remove and transform data
titanic_train$PassengerId<-NULL
titanic_train$Name<-NULL
titanic_train$Ticket<-NULL
titanic_train$Cabin<-NULL
titanic_train$Fare<-log(titanic_train$Fare)
```

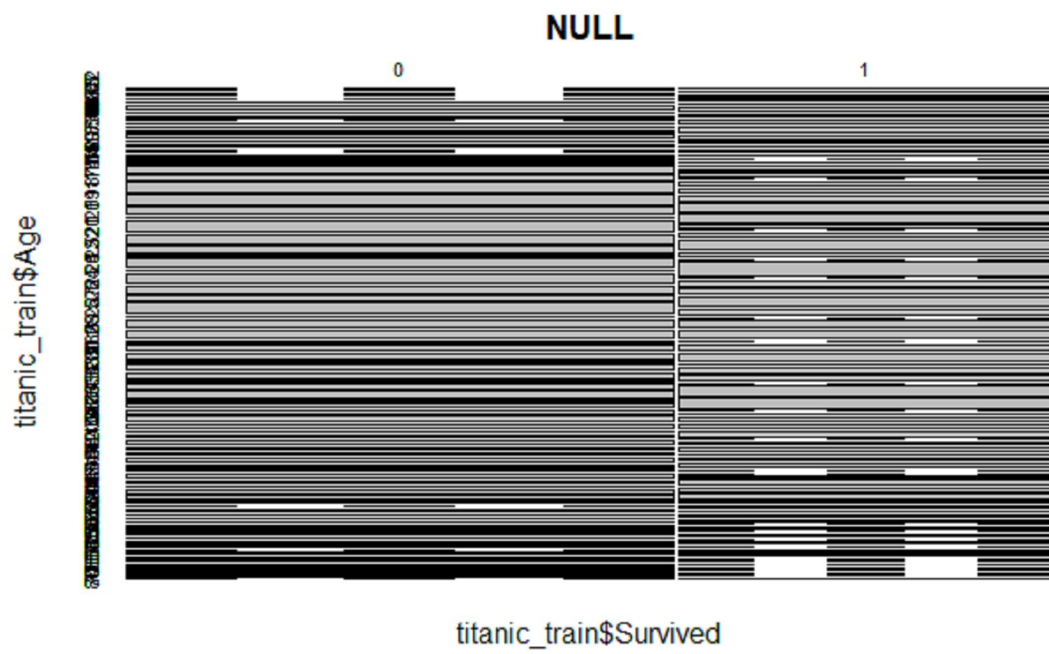
NaNs produced

Hide

```
View(titanic_train)
```

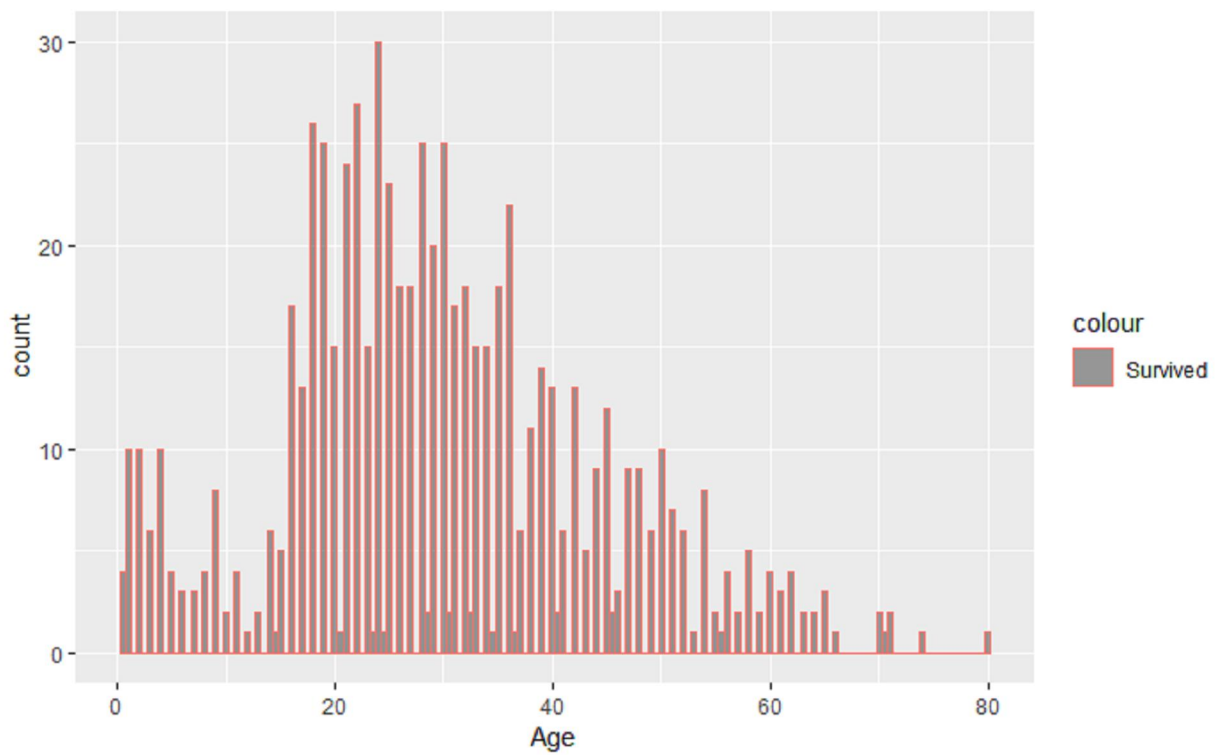
Hide

```
#plots for 3 pairs of variables
titanic1<- as.data.frame(titanic_train)
library(GGally)
#Age Survived
mosaicplot(titanic_train$Survived~titanic_train$Age)
```



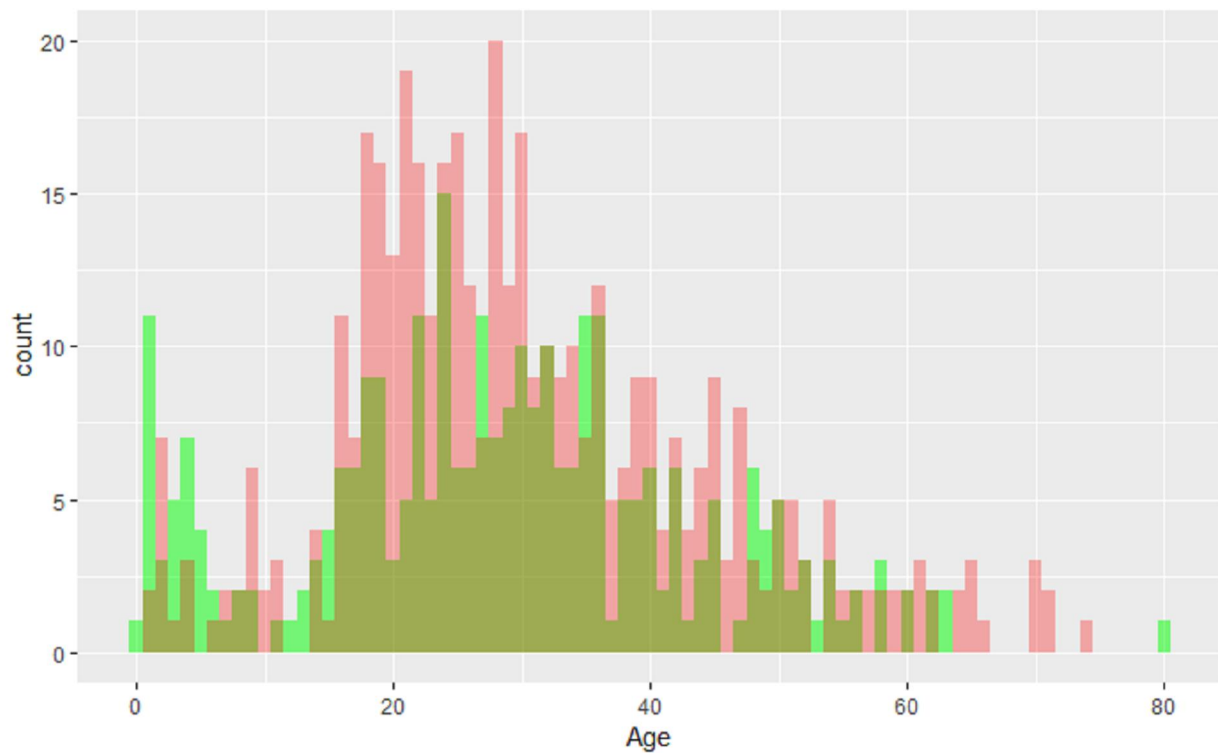
Hide

```
ggplot(titanic_train, aes(x=Age,color='Survived')) +  
  geom_histogram(binwidth=.5, alpha=0.6, position="identity")
```



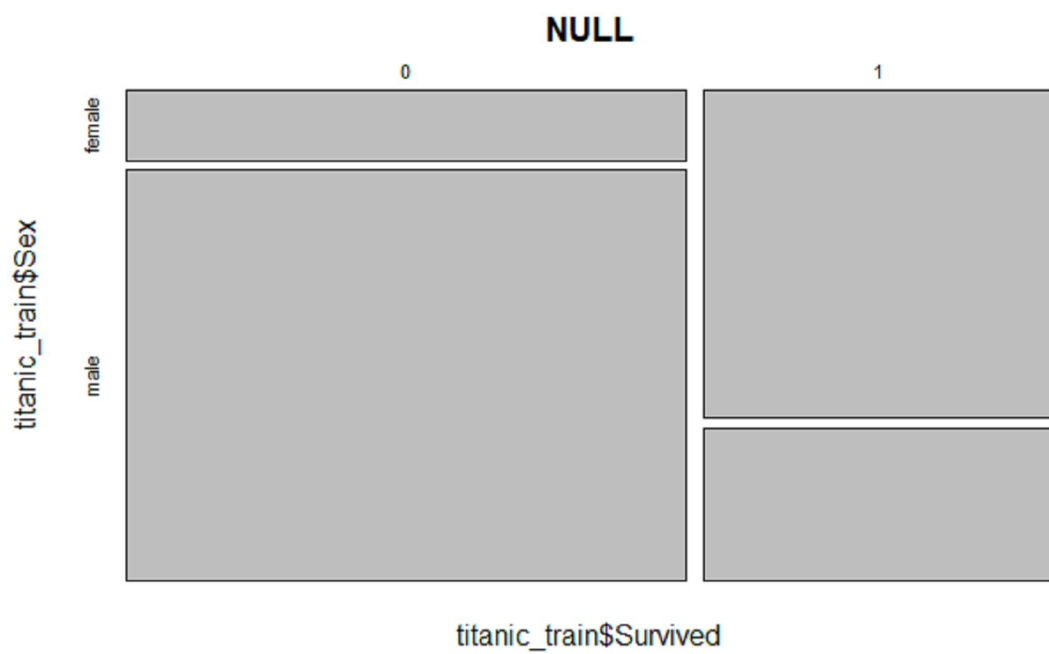
Hide

```
# Change histogram plot fill colors by if survived
cond <- titanic_train$Survived == 1
ggplot(titanic_train, aes(x=Age)) +
  geom_histogram(data=subset(titanic_train,cond==TRUE),binwidth=1,fill="green",alpha=.5) +
  geom_histogram(data=subset(titanic_train,cond==FALSE),binwidth=1,fill="red",alpha=.3)
```



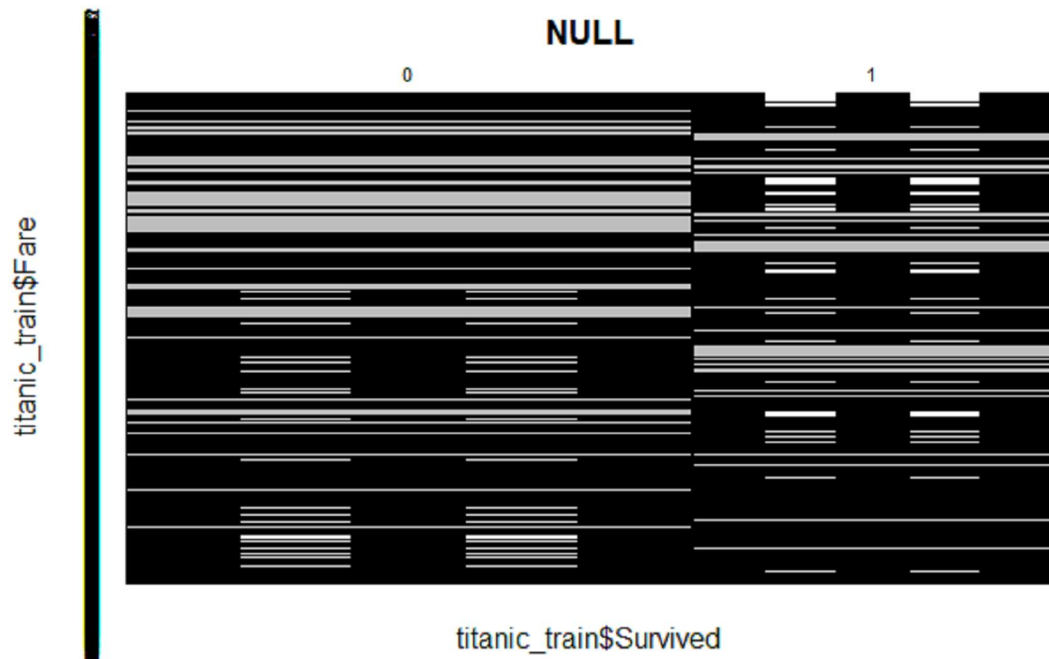
Hide

```
#Sex Survived
mosaicplot(titanic_train$Survived~titanic_train$Sex)
```



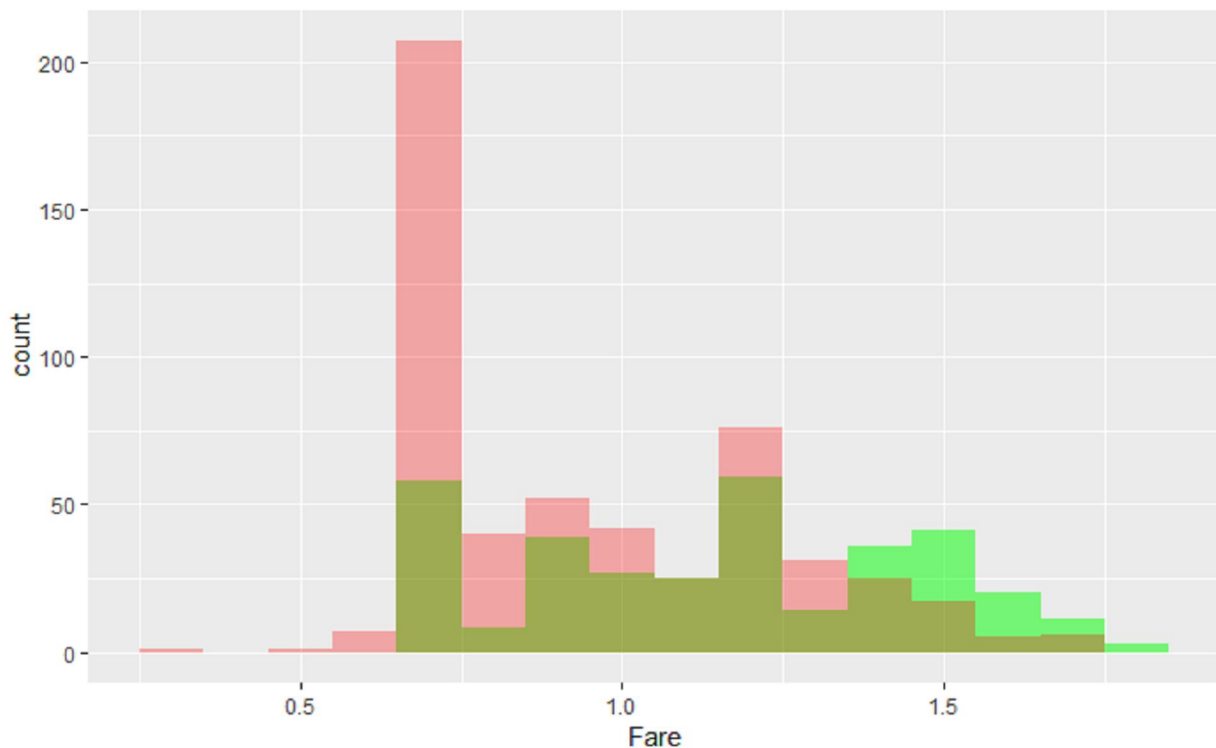
Hide

```
#Fare Survived
mosaicplot(titanic_train$Survived~titanic_train$Fare)
```



Hide

```
# Change histogram plot fill colors by if survived
cond <- titanic_train$Survived == 1
ggplot(titanic_train, aes(x=Fare)) +
  geom_histogram(data=subset(titanic_train,cond==TRUE),binwidth=0.1,fill="green",alpha
=.5) +
  geom_histogram(data=subset(titanic_train,cond==FALSE),binwidth=0.1,fill="red",alph
a=.3)
```



Hide

```
#create a model matrix
#matrix only contains numbers, convert a data frame into numeric
#remove Survived variable
titanic2 <- model.matrix(~Pclass+Age+SibSp+Parch+Fare, data=titanic1)
#delete rows with inf values
titanic2[is.infinite(titanic2)] <- NA
titanic2 <- na.omit(titanic2)
#turn into list titanic2<- split(titanic2, rep(1:ncol(titanic2), each = nrow(titanic
2)))
```

Hide

titanicpca

Standard deviations (1, ..., p=5):

```
[1] 1.4030982 1.3011731 0.8197669 0.7486194 0.3252923
```

Rotation (n x k) = (5 x 5):

	PC1	PC2	PC3	PC4	PC5
Pclass	-0.5875243	0.34802286	-0.2622727	0.2095212	0.64885217
Age	0.1786743	-0.56484574	-0.5560489	0.5805830	0.05251396
SibSp	0.2542585	0.56541730	0.2373679	0.7180822	-0.20897428
Parch	0.2974435	0.48947403	-0.7366459	-0.3033919	-0.19299911
Fare	0.6853981	0.02340574	0.1517626	-0.1064680	0.70378565

Age and EmbarkedQ are chosen.

Hide

```
#NMF rank 2
#install.packages("NMF")
library(NMF)
#W (u x k) and H (k x v)
#A ≈ WH
#matrix titanic2[, -1] with desired dimension k=2
#perform multiple runs of one algorithm (default is to keep only best fit)
titanicnmf <- nmf(titanic2[, -1], rank = 2, method = "snmf/r", nrun = 10)
#consensusmap(titanicnmf)
#Sys.setlocale('LC_ALL', 'C') Chinese character not error
w <- basis(titanicnmf) # W user feature matrix matrix
h <- coef(titanicnmf)
A <- w %*% h #is matrix multiplication
dim(w)
```

```
[1] 707 2
```

Hide

```
df <- as.data.frame(w)
head(df, 10)
```

	V1 <dbl>	V2 <dbl>
1	0.026214891	0.010867491
2	0.045236002	0.001064432
3	0.030988776	0.008200979
4	0.041665688	0.001493020
5	0.041695126	0.006722347
7	0.064255733	0.000000000
8	0.002395084	0.020167110

	V1 <dbl>	V2 <dbl>
9	0.032164461	0.011182300
10	0.016695727	0.008753946
11	0.004801688	0.015668638
1-10 of 10 rows		

[Hide](#)

```
#View(titanic2[, -1])
scaledifference <- scale(A -titanic2[, -1], scale=FALSE) ##lookingt at the sum of suqa
red difference compared to the original data and delete the smallest, why scale= FA
LSE?
squarescale <- scaledifference^2
colSums(squarescale) #since Age and EmbarkedQ has the lowest square difference, the
se two are chosen
```

```
Pclass      Age      SibSp      Parch      Fare
292.90644  15.49182 402.29678 410.24345 106.96629
```

[Hide](#)

```
summary(squarescale)
```

```

Pclass      Age      SibSp      Parch
Fare
Min.   :0.000001  Min.   :3.200e-07  Min.   : 0.000006  Min.   : 0.000002  Mi
n.     :0.000007
1st Qu.:0.073509  1st Qu.:1.934e-03  1st Qu.: 0.039515  1st Qu.: 0.050505  1st
Qu.:0.0176337
Median :0.308048  Median :9.391e-03  Median : 0.266784  Median : 0.196584  Medi
an :0.0666765
Mean   :0.414295  Mean   :2.191e-02  Mean   : 0.569019  Mean   : 0.580259  Mea
n    :0.1512960
3rd Qu.:0.573993  3rd Qu.:2.608e-02  3rd Qu.: 0.480116  3rd Qu.: 0.316182  3rd
Qu.:0.1861113
Max.   :4.185267  Max.   :3.063e-01  Max.   :10.547229  Max.   :24.936083  Ma
x.    :1.3877824
```

[Hide](#)

```
testpca
```

	PC1	PC2
1	1.5270306	-0.355528761
2	1.6968861	-0.243201948
3	3.0540012	-1.821837573
4	1.6663424	-0.056647983
5	3.0244464	1.342217786
6	1.6507094	0.452632057
7	1.4126120	-0.182962015
8	5.7400805	0.838087437
9	1.1421538	0.278399519
10	4.4829990	1.468301201
12	5.8182243	-1.539979583
13	8.4432835	0.048374367
14	5.5970709	-1.177478378
15	8.0602290	-0.905901490
16	5.2650343	0.340624550
17	3.2780717	-0.755341521
18	1.1776421	0.161962052
19	1.7353312	0.542439987
20	1.4721829	-0.769174520
21	8.0910578	-1.218580341
22	-0.5061484	1.134732919
24	7.8235500	0.068585242
25	12.4449864	0.883637813
26	3.3994861	-0.302709021
27	7.8580741	0.030547756
28	1.1960509	0.103766016
29	6.1220017	-1.333523585
31	5.4375279	-0.673112735
32	5.8299012	0.956646321
33	4.6859022	1.527247434
35	7.7198060	-0.250846927
36	1.1482901	0.259000840
38	1.5927072	0.176136161
39	1.8528992	0.028155672
41	3.1616584	0.083491684
43	1.6128710	-0.607504333
44	3.3340380	-0.557348055
45	7.6882405	-0.840171661
46	1.4382610	0.013996143
47	6.1102933	-1.490789233
49	8.4523660	-1.999056315
50	3.8605964	0.784682544
51	7.7335985	-0.015077205
52	3.6296236	-0.429604708
53	5.4087444	1.658803387
54	12.4027190	2.300314226
56	5.6688222	3.693314518
57	1.5525427	-0.374265772
58	1.3574771	0.011237447
60	10.9832622	-0.971433558
61	1.3316371	0.324086657

62	3.4449108	-0.631994754
63	1.3012765	0.283833417
64	1.3429759	0.128391604
65	11.9406634	2.276050174
67	1.3390955	0.285124904
68	6.9491604	-1.540575565
69	5.8471459	-0.950745136
70	12.9450087	0.989789078
71	1.3749117	0.051049274
72	1.3807272	0.168897229
73	1.4873511	-0.141193286
74	5.7500472	-0.856019835
75	10.4779383	-0.949473472
76	10.4472570	-0.852480079
78	6.4471808	-0.677985226
79	3.3340380	-0.557348055
80	1.3749117	0.051049274
81	3.3215302	1.979826191
82	11.2517390	-1.581243305
83	5.8550419	-1.656371655
87	1.4495483	-0.064051311
88	1.3881507	0.286800096
90	4.9153136	1.751117257
91	2.9591895	0.780315933
93	8.1377044	1.000420066
95	5.5605011	-0.725235082
96	1.4006623	0.012712182
97	8.9966979	-2.011199053
98	1.4873511	-0.141193286
99	1.3563713	0.207281949
100	1.5722387	-0.295160262
101	7.7860423	-0.758398909
102	5.1552596	0.219226480
104	1.4068236	-0.026293868
105	2.1123518	1.546369280
106	3.8645183	-0.020798667
107	1.3588958	0.168151707
110	3.1929039	-0.111178447
111	3.8033402	-0.972702785
113	6.1474089	-1.136573684
114	1.1653443	0.259583226
115	11.2026489	-1.426053876
116	2.9995285	0.938559288
118	3.4685615	2.180929448
119	8.1261249	-1.069002278
120	5.1798047	0.141631766
121	3.5520664	0.155993584
123	7.7811687	-0.444833713
124	1.2714511	-0.109349627
126	3.3087017	0.951275058
127	1.3638447	0.129104254
129	3.4813084	-1.022916341
130	1.4617859	0.054015953

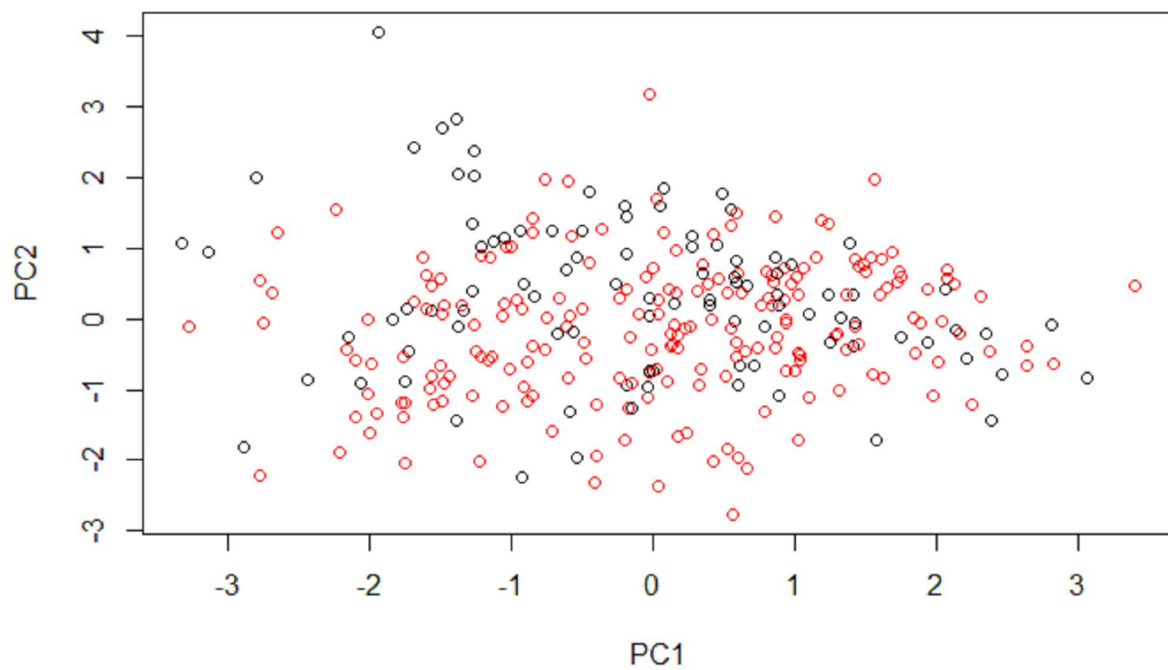
131 1.5599662 -0.256362905
132 6.1141338 -1.804389707
135 1.6507229 -0.684644629
136 1.4054614 0.052092520
137 1.2451410 -0.051423413
138 3.2849479 -0.402158626
139 1.4495134 0.092813310
140 8.0457414 3.606161348
141 7.3784280 4.908101500
142 9.6909769 -0.897914602
143 12.6045294 0.379272170
144 4.8950069 -0.425609436
145 5.8170171 -1.383154996
146 4.2059499 1.663407749
148 1.4372409 0.131610667
150 5.5393867 0.674368109
151 8.5427253 0.014711537
154 3.2515659 0.763884689
155 6.2240771 4.154300010
156 1.3151066 0.049006984
157 10.5128575 -0.712982291
158 1.3931889 0.090889878
159 5.8170171 -1.383154996
160 3.4097101 1.161449711
162 3.3338027 1.941028833
163 3.3712756 -0.399210611
165 3.4690359 -0.984118984
166 4.4084610 1.232614850
167 7.3611397 -1.565723291
168 5.9483606 2.755338140
170 1.6828567 0.139998231
172 1.2512773 -0.070822091
173 2.9714620 0.741518576
175 6.7797885 3.003256789
176 6.3575493 1.253495761
177 5.5883670 -0.088200174
178 7.9210402 -1.185169838
179 6.9625820 1.010284960
180 9.3932087 -1.004446408
181 3.3340380 -0.557348055
182 8.9870660 0.077587095
183 7.3805154 0.208164040
185 11.3593334 0.550762468
186 3.9317204 -0.929102263
187 4.6678765 0.400822820
188 1.9209286 1.537674571
190 3.4567634 -0.945321627
191 5.2411674 -0.052355020
193 3.2743030 1.762522930
194 3.5971576 -1.764072808
195 5.8545968 1.510835551
196 1.5159142 -0.297083694
197 9.7812316 1.283358528

198	1.3086434	0.284084990
199	2.7595981	-0.302449499
202	3.1961732	2.160843973
203	11.0647954	-0.803298196
204	5.2693910	1.527909967
205	2.7841431	-0.380044213
207	1.5099096	-0.375721655
208	2.7718706	-0.341246856
209	5.8052737	-1.030607942
210	1.4298173	0.013707800
211	3.9136085	-0.175988096
213	7.1370634	0.082335802
214	5.5602532	-1.061086307
215	3.3388139	3.075359481
216	6.8931862	-1.346404767
218	10.7980022	-0.644900097
219	11.2818610	-0.353861571
221	3.7528089	0.053684666
222	1.4249683	0.170408024
223	2.7473255	-0.263652142
224	1.3515722	0.167901611
225	6.0279191	-1.807333865
227	1.3761172	0.090306896
229	3.1031175	-0.537005504
230	3.4076732	-0.790132198
231	6.8191270	0.189127914
232	5.5592938	-0.568410495
233	1.2068062	0.759690043
235	8.3118479	-0.583577300
236	1.3563713	0.207281949
237	8.7425334	-1.549281104
238	1.1653696	0.200759409
239	3.8066023	1.085792490
240	9.3390466	-0.901447451
241	6.0752694	-1.884149800
242	6.1255898	0.069080948
243	10.1850761	-0.195233543
246	7.6088444	-0.686017156
247	4.3328395	-0.209508237
248	6.1970787	0.226230267
249	5.1798047	0.141631766
251	5.6788115	2.379217191
252	1.3768983	0.207982929
253	9.6552981	-0.067102220
254	1.7980971	0.065500684
255	1.9449432	-0.262824507
258	1.5108761	-0.101173476
259	3.1990401	-0.130577126
260	1.3454610	0.167692918
261	3.6508516	0.235297010
262	1.3686438	0.168484592
263	5.3214564	0.669083754
264	2.7472428	2.156297050

265 3.2873772 -0.558941482
270 1.5436171 0.331325589
271 8.2488503 -1.456975850
273 9.6430255 -0.028304863
276 5.0693519 0.490807981
277 2.8209607 -0.496436285
278 5.3148026 -0.285139163
279 4.7035450 0.086151465
280 2.7473255 -0.263652142
281 1.6172523 0.098541446
282 3.0250443 2.175587839
284 3.3583478 1.863434119
285 3.9174841 2.157043305
286 1.3696314 -0.419728485
288 8.4555561 0.009577010
292 1.1993305 -0.190245388
294 9.1473853 -0.544401395
295 1.9878970 -0.398615257
296 1.4420899 -0.025089557
297 6.6047413 2.407699561
299 6.9019363 -0.875508520
300 1.4668241 -0.141894265
301 1.4804588 -0.259078011
303 4.9378723 -0.452719038
304 1.6295248 0.059744089
306 6.7068476 -1.059124595
307 10.6213036 0.967583428
308 1.8671763 1.536178288
309 9.4760917 -0.611609269
310 3.2754817 -0.110861393
311 1.5558896 0.292528232
312 1.1899146 0.123164695
314 1.5344546 -0.453316369
315 9.7071598 -1.760123948
316 1.2840696 0.322462272
317 10.1808469 -1.225648794
318 2.7105079 -0.147260070
319 1.4422790 -0.064299551
320 5.8053562 1.034241035
321 1.4068236 -0.026293868
322 1.2280616 0.006818018
323 3.2849479 -0.402158626
324 5.7065643 -1.033978781
325 10.5252703 -1.104722940
326 1.1461007 0.082452033
327 6.5184726 2.010431251
328 8.3661276 -1.452970934
329 5.1798047 0.141631766
330 3.2235852 -0.208171840
331 6.6266170 -0.031458776
332 6.0366580 -1.258005090
334 3.5542914 1.477960885
335 1.4543624 -0.063886914

336 5.6218638 -0.919221868
337 3.3585831 -0.634942769
338 2.0944961 0.606722417
339 5.2971234 -0.294228154
341 2.6982354 -0.108462713
342 1.4221165 -0.261070349
344 13.1312875 -1.201168979
346 1.2470243 0.360413661
347 3.2849479 -0.402158626
348 1.3876045 -0.497547625
349 3.3467305 -0.321615896
350 4.4432923 -0.558684451
351 8.1906817 -0.860072417
352 2.7841431 -0.380044213
353 7.1493359 0.043538445
354 8.2158099 0.580326371
355 4.2829949 2.800964670
356 5.8673144 -1.695169012
357 8.0853068 -0.778910481
360 3.3139530 0.478699167
361 9.1525355 6.582407854
362 6.2730594 0.934721105
363 4.4432923 -0.558684451
364 1.6663424 -0.056647983
365 7.5651367 -0.060046480
368 5.0864979 0.256229578
369 8.0431168 -0.865111628
370 3.4680112 -0.513556536
371 3.2156682 0.388289866
372 9.3934259 -0.829642801
374 3.5058535 -1.100511055
375 9.1596579 -0.583198752
376 11.0937151 -1.320609773
377 2.1500302 1.349415921
378 2.9431430 -0.217748693
379 7.4244771 -1.838075546
380 6.1258968 4.464678868
382 1.4372758 -0.025253954
384 3.2584625 0.908185202
386 7.9089966 1.550260300
387 1.3822786 0.051300847
388 3.6653964 -1.604876699
389 1.3380941 0.167441346
390 4.6072071 3.217195051
391 8.4635359 -0.547666098
392 7.1777039 -1.129963436
393 4.1315062 1.695912304
394 3.0541389 -1.233586071
395 4.9903290 2.328299897
396 7.6599633 0.217706938
397 1.2223610 0.045839801
398 9.0105074 -0.352993390
399 1.3577335 0.128895561

```
400 1.4558851 -0.220700729
401 9.8468091 -0.774943712
402 4.8017253 -0.224227393
403 7.7608485 0.027227587
404 6.8214479 -0.368443234
405 6.2005242 -0.812542955
406 3.3582515 -0.164356651
407 3.0321233 0.303589060
408 11.2818610 -0.353861571
410 3.0526575 2.088293785
412 8.8206077 -0.487770694
413 1.4313687 -0.103888582
415 9.0086551 -1.156514012
416 1.4003127 -0.516721878
```



Hide

```
logpca_cv = cv.lpca(data1, ks = 2, ms = 1:10)
```

