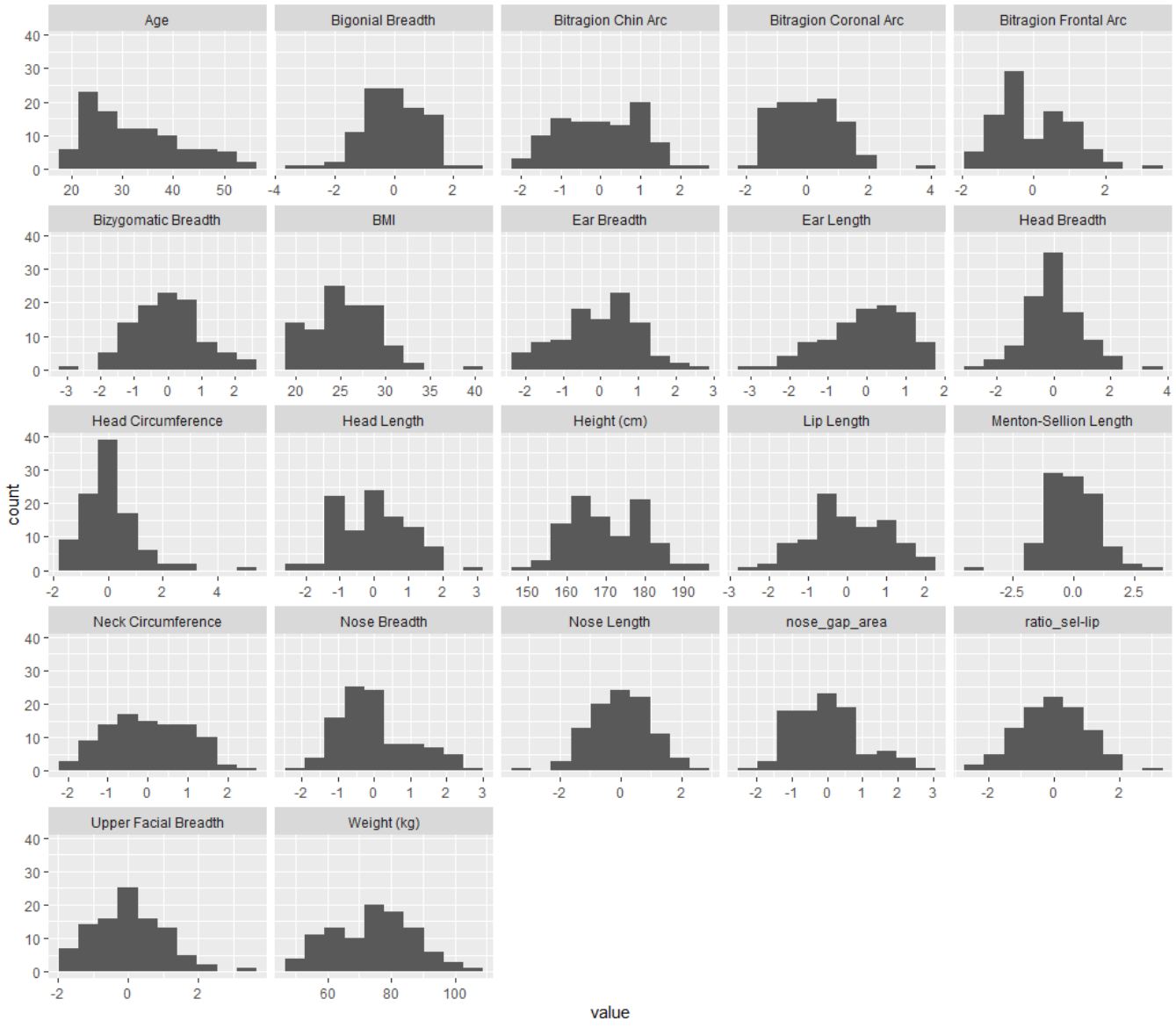
Face Fit Exploratory Analysis and PCA

Variable distribution



Most look normal.

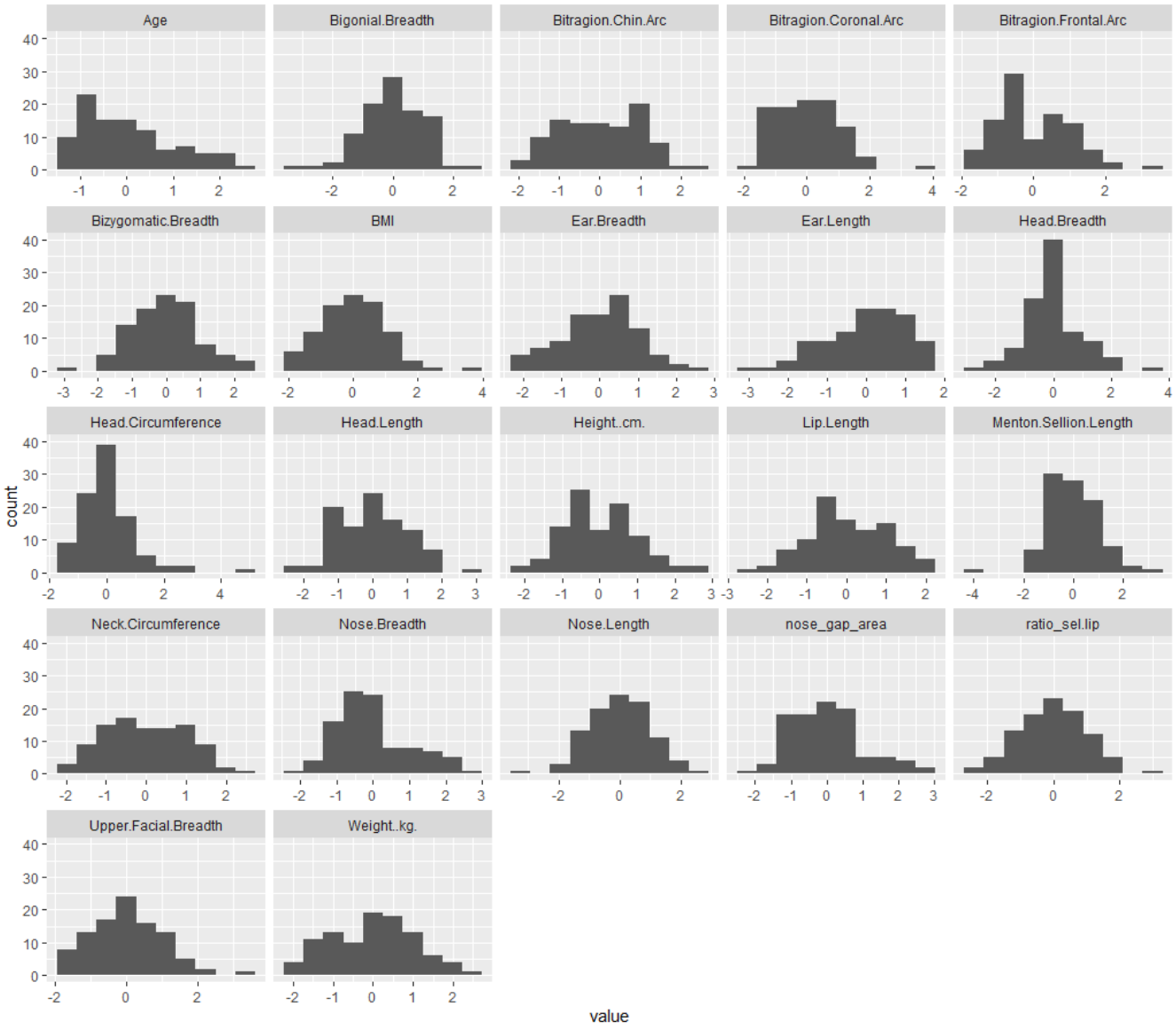
Right-skew for Age, head circumference and nose breadth. May need a transformation

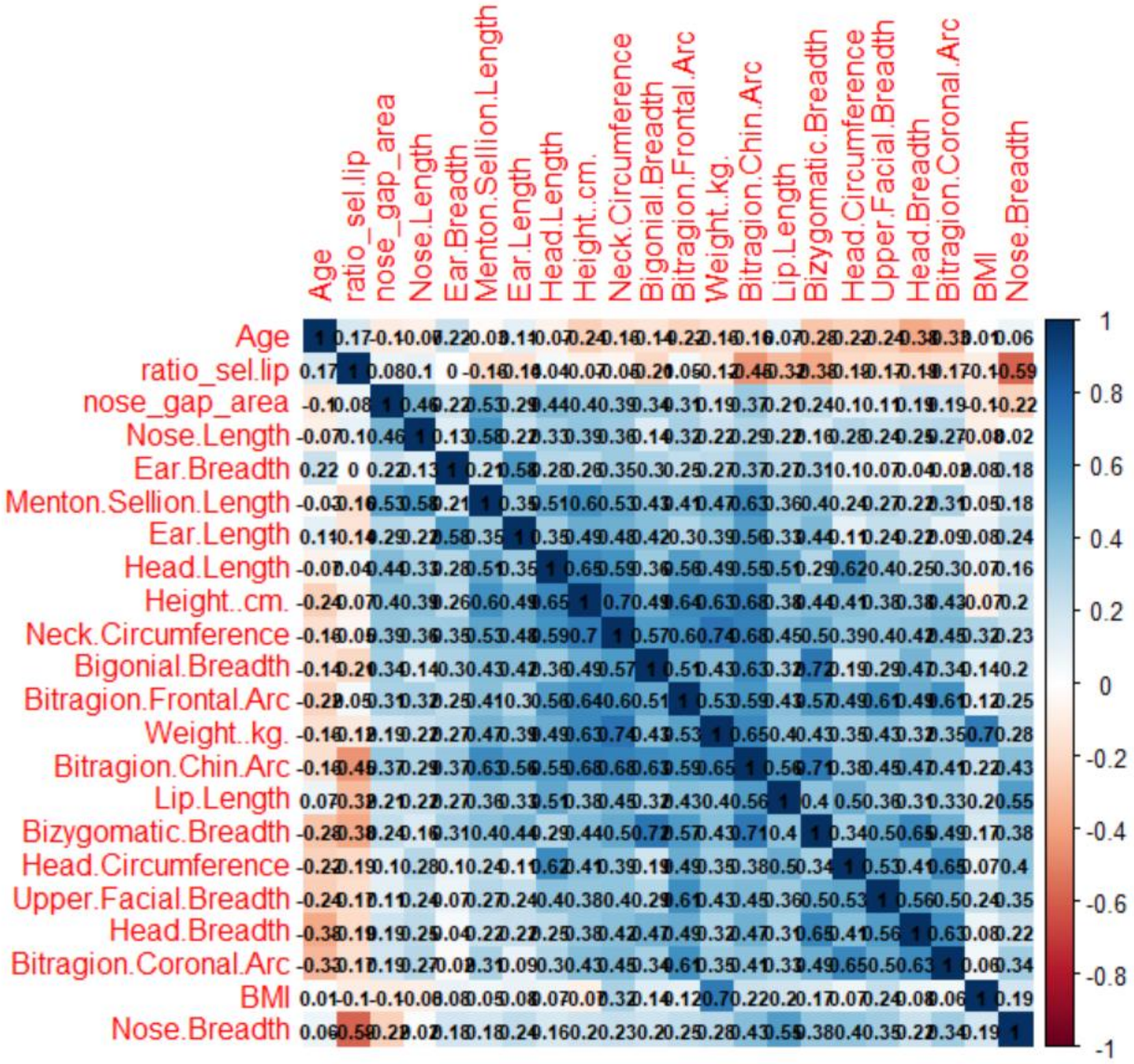
Left-skew for ear length

Uniform distribution for bitragion measurements and possibly neck circumference and nose_gap_area.

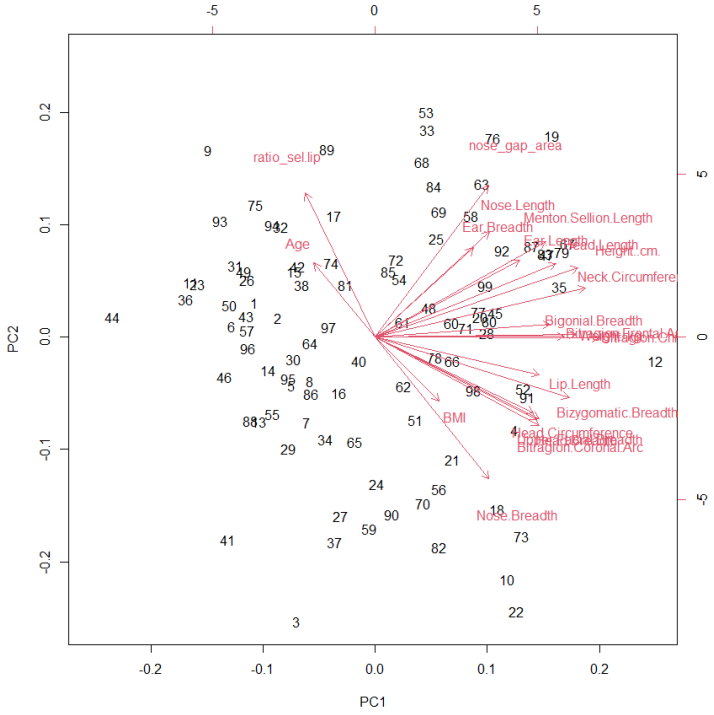
Weight, Age, BMI, and height need to be scaled and centered.

With scaling for all variables, include outliers

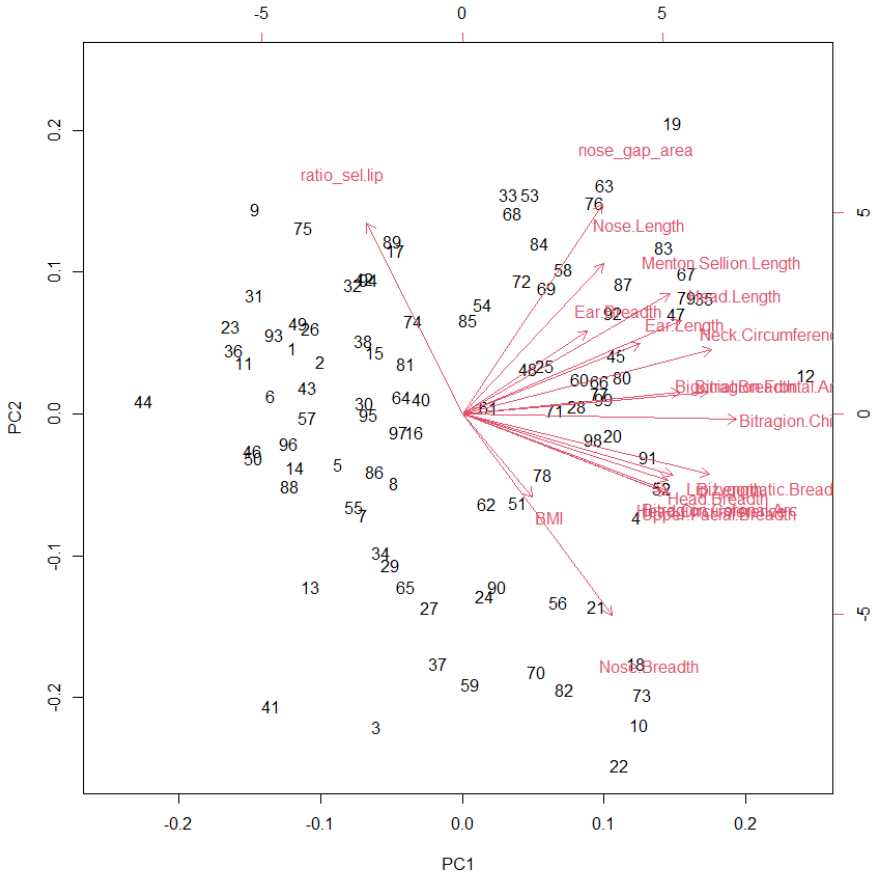




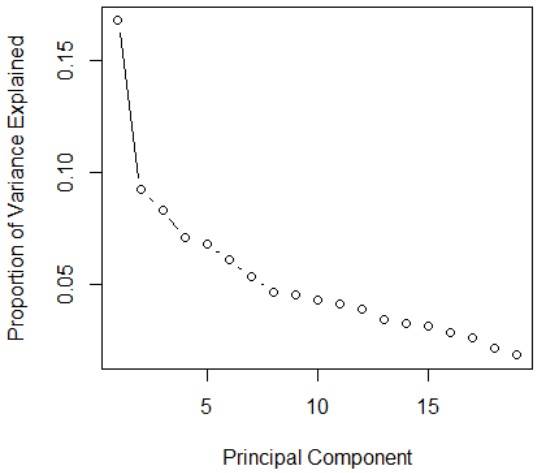
PCA (98 observations)

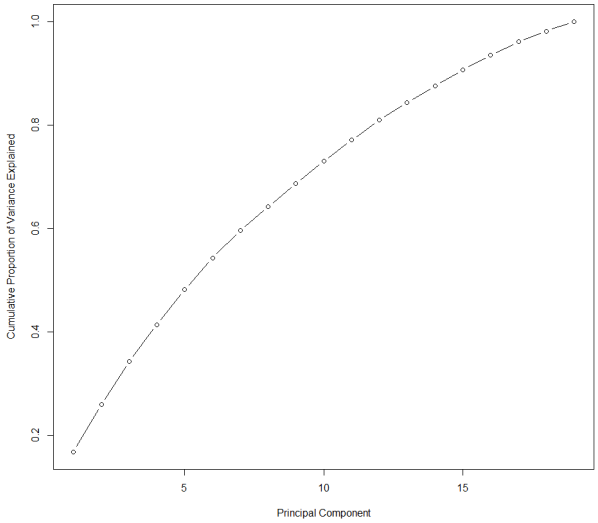


Remove some variables



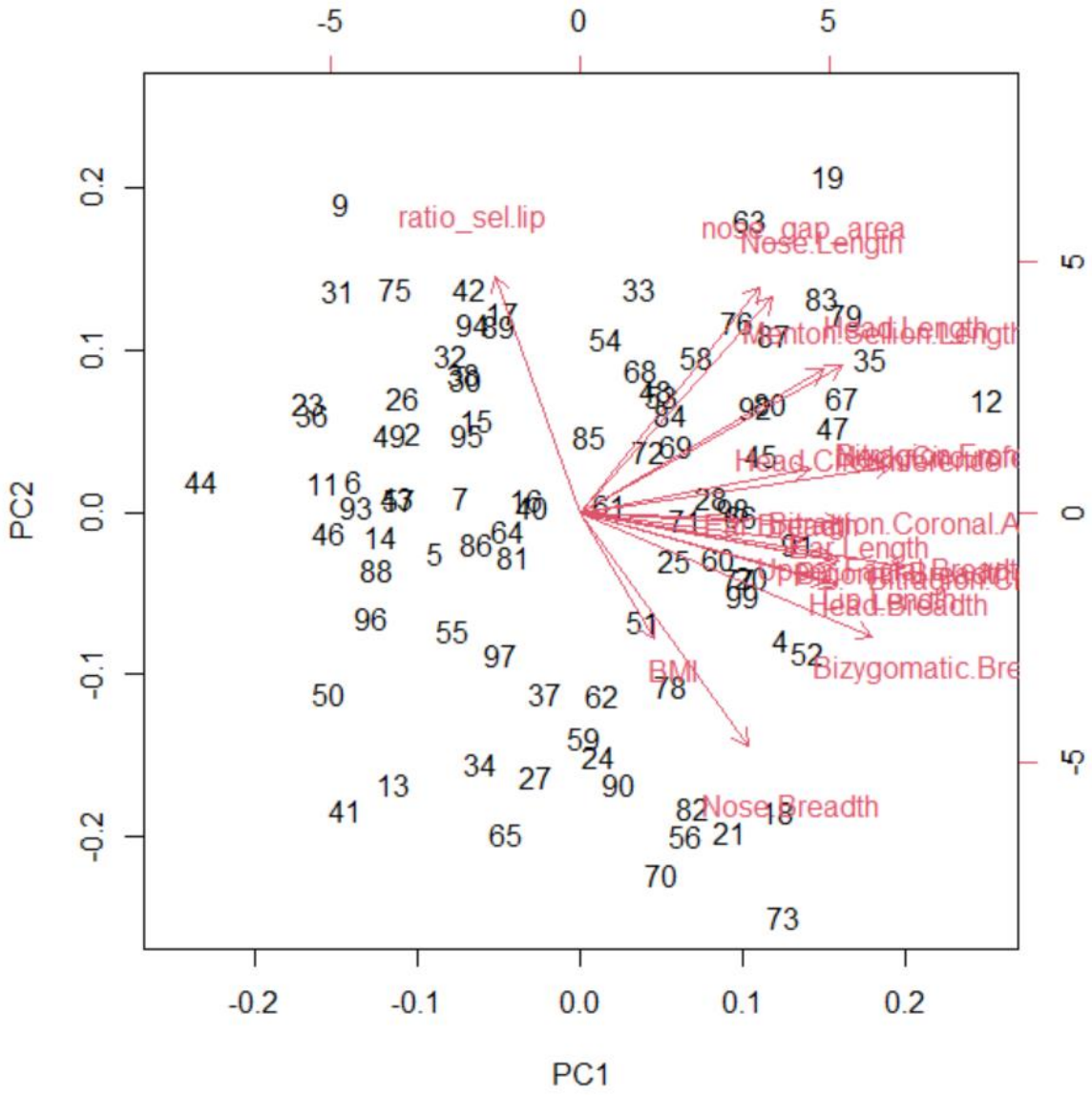

```
Standard deviations (1, .., p=19):
 [1] 2.6932432 1.4773317 1.3353824 1.1324011 1.0858758 0.9761661 0.8525111 0.7423209 0.7230402 0.6837451 0.6575282 0.6218437 0.54
09994 0.5210889 0.4969525
[16] 0.4486753 0.4150951 0.3362548 0.2888591

Rotation (n x k) = (19 x 19):
                              PC1         PC2         PC3         PC4         PC5         PC6         PC7         PC8
PC9         PC10
Bitragion.Chin.Arc      0.32070065 -0.01209426  0.20367880  0.02857164 -0.109673680 -0.1140396731  0.107085725 -0.068492573  0.126
36844 -0.09362322
Bitragion.Coronal.Arc   0.24070221 -0.16379799 -0.37828353  0.10812519 -0.061400099  0.0778080209 -0.050896428  0.415919562 -0.166
23219 -0.24849398
Bitragion.Frontal.Arc   0.28755832  0.04640290 -0.20280657  0.09530582  0.214942082  0.1276212115  0.043863405  0.121336837  0.383
09681  0.27623893
Head.Circumference      0.23307361 -0.16332657 -0.32781465 -0.33674934  0.072436050  0.1980310630  0.121221895  0.004853124 -0.169
89931 -0.37415242
Neck.Circumference      0.29152181  0.13679441  0.04786664  0.01438199  0.228787017 -0.1731270170  0.057643335  0.259165243 -0.026
15754 -0.22311236
Ear.Breadth             0.14605198  0.17624732  0.41776170 -0.09600803  0.234849680  0.4223344605 -0.226930809  0.156519895 -0.160
25348 -0.04926398
Ear.Length              0.20813452  0.14969971  0.39622454  0.01078915  0.065080261  0.2648975853 -0.287783038 -0.220898698 -0.039
28683 -0.23497536
Lip.Length              0.24585847 -0.12968356  0.08525766 -0.37665990 -0.001755909 -0.0011999272  0.175173161  0.005193864 -0.333
44997  0.65132903
Upper.Facial.Breadth    0.24105019 -0.17097861 -0.23280860  0.04619004  0.175325716  0.0008780183 -0.279908013 -0.609081004  0.290
92395  0.08252396
Menton.Sellion.Length   0.24245347  0.25613342  0.03565874 -0.11857341 -0.307056277 -0.3423442335 -0.059093486  0.107956299  0.346
95131 -0.06777356
Nose.Breadth            0.17513284 -0.42902735  0.17093677 -0.29858998 -0.127002585  0.0701167481 -0.160108758  0.232332942  0.168
89753  0.15722007
Nose.Length             0.16517690  0.32050611 -0.20174556 -0.18238061 -0.272073562 -0.1517955618 -0.584556747  0.132692848 -0.055
41927  0.11627262
Bigonial.Breadth        0.25506724  0.04607166  0.17488017  0.39666576 -0.019093162 -0.0130831588  0.318004722  0.202113032  0.113
11750  0.10875454
Bizygomatic.Breadth     0.28876841 -0.12911106  0.11685662  0.37717690 -0.079283406  0.0681880705  0.009012348 -0.020911588  0.024
61804  0.09016002
Head.Breadth            0.23992159 -0.14170021 -0.23099608  0.38905638 -0.058141758  0.0975136002 -0.156523724 -0.145597740 -0.396
54693  0.07494365
Head.Length             0.25561045  0.19868069 -0.10111356 -0.34383847  0.166911550  0.0641191107  0.395342748 -0.234675195  0.111
71900 -0.15632450
nose_gap_area           0.16354945  0.44623822 -0.03526863  0.04609129 -0.220503904 -0.1225460310  0.197783444 -0.238231827 -0.403
25489  0.07267182
ratio_sel.lip          -0.11162262  0.40625201 -0.26127058  0.06046757  0.473785318  0.1510126055 -0.085723285  0.196277450  0.082
40620  0.26943820
BMI                     0.08197889 -0.17621115  0.15173939 -0.00271261  0.543007419 -0.6674608459 -0.152793626  0.005254696 -0.229
18130 -0.06331756
```





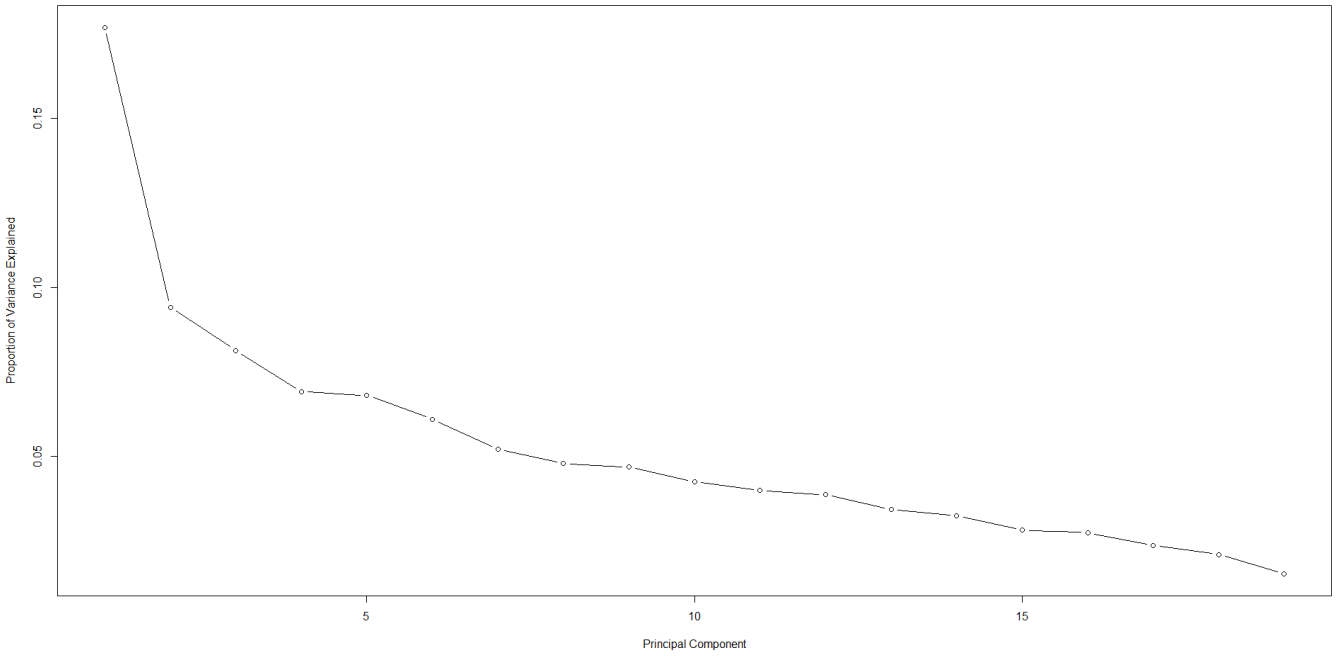

```
> cumsum(pve)
 [1] 0.1680199 0.2601842 0.3434930 0.4141386 0.4818818 0.5427806 0.595
9651 0.6422753 0.6873826 0.7300385 0.7710589 0.8098531 0.8436037 0.876
1122 0.9071149
[16] 0.9351058 0.9610018 0.9819793 1.0000000
```

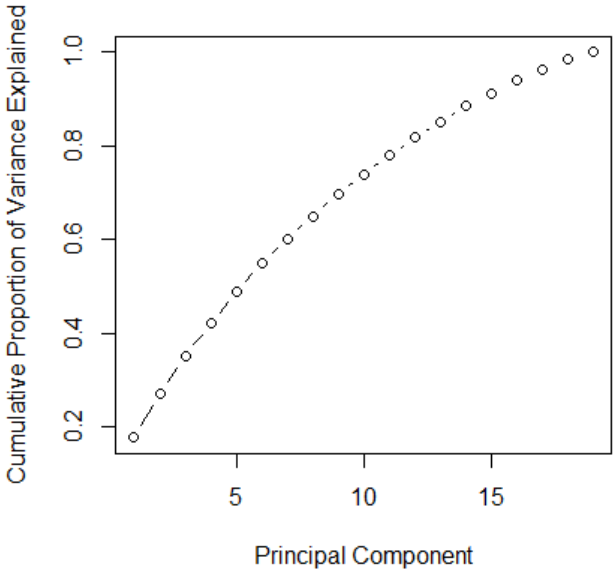
PCA with outliers removed (91)




```
Standard deviations (1, ..., p=19):
 [1] 2.7244968 1.4489630 1.2528949 1.0641833 1.0471490 0.9390362 0.8015775 0.7373125 0.7213920 0.6533415 0.6144736 0.5949551 0.52
73393 0.4998791 0.4332635
[16] 0.4215612 0.3637535 0.3221182 0.2343555

Rotation (n x k) = (19 x 19):
                          PC1         PC2          PC3         PC4          PC5          PC6           PC7          PC8
PC9         PC10
Bitragion.Chin.Arc     0.31298344 -0.09814171  0.163822697 -0.13376304  0.002499892 -0.122417142  7.354774e-02 -0.137403199  0.14
089917 -0.096140953
Bitragion.Coronal.Arc  0.25132128 -0.01721176 -0.343842168 -0.03552463 -0.009032435  0.063761370  5.779092e-02  0.215729947 -0.21
834478  0.119693015
Bitragion.Frontal.Arc  0.29963269  0.08150416 -0.155855304  0.15240708 -0.146082882  0.223174780 -1.252609e-01  0.176005479  0.28
813849  0.323223718
Head.Circumference     0.22185036  0.07814445 -0.127505551  0.26480324  0.160997939  0.170956111  1.202745e-01 -0.042093005 -0.12
692537 -0.152464221
Neck.Circumference     0.29914197  0.08337936  0.125790109  0.12499923 -0.212931175 -0.225228198  1.640220e-01  0.112746147  0.05
211530 -0.242695554
Ear.Breadth            0.15540211 -0.01691726  0.498247176 -0.01458845 -0.236258850  0.341399889 -2.370727e-01  0.200012750 -0.33
619706  0.247041509
Ear.Length             0.21438831 -0.05335268  0.409425027 -0.15744998 -0.100846989  0.130862205 -2.849413e-01 -0.335939315 -0.15
611770 -0.350734897
Lip.Length             0.23932516 -0.12648587  0.132688821  0.32938490  0.366224639  0.085575203  2.105844e-01  0.159313700 -0.25
879719 -0.016126410
Upper.Facial.Breadth   0.24374687 -0.08384045 -0.287004230  0.17805184 -0.007852332  0.008378773 -4.833396e-01 -0.488829670  0.16
012311  0.245166957
Menton.Sellion.Length  0.23488541  0.25824145  0.117932894 -0.16604879  0.121415840 -0.299013377 -1.445071e-01  0.071441906  0.29
180525  0.003981032
Nose.Breadth           0.16195952 -0.42250319  0.143837209  0.08477476  0.399697717  0.020606819 -1.109800e-01  0.279405687  0.15
863208  0.212300146
Nose.Length            0.18583154  0.38943790 -0.035017875 -0.17136667  0.287865852 -0.326486241 -3.930305e-01  0.348260286 -0.11
180883 -0.135467215
```

```
Bigonial.Breadth       0.24899605 -0.08918309  0.006284172 -0.25560361 -0.319030321  0.017669679  3.662975e-01  0.197347343  0.29
761745 -0.016262919
Bizygomatic.Breadth    0.28126158 -0.22498141 -0.120138955 -0.27203608 -0.223217535  0.096432085 -5.587082e-05  0.016847747  0.04
221360  0.038168034
Head.Breadth           0.24684334 -0.13104936 -0.456439742 -0.17382455 -0.010960759  0.099478735 -2.678847e-02 -0.025811680 -0.37
788569 -0.366507251
Head.Length            0.25250936  0.26748857  0.117740975  0.35852124  0.144874670  0.155685793  2.500804e-01 -0.274402274  0.24
028322 -0.178985099
nose_gap_area          0.17296253  0.40678658  0.039242246 -0.21240039  0.043135070 -0.115586794  3.223231e-01 -0.292541342 -0.36
075726  0.541321179
ratio_sel.lip         -0.08225021  0.42736952 -0.088348112  0.29283476 -0.345251054  0.266344946 -1.683478e-01  0.258162636  0.02
095823 -0.117900999
BMI                    0.07196002 -0.22789456  0.029397350  0.46215894 -0.394968942 -0.621001935 -3.868785e-02  0.006791015 -0.23
730589  0.083347097
```



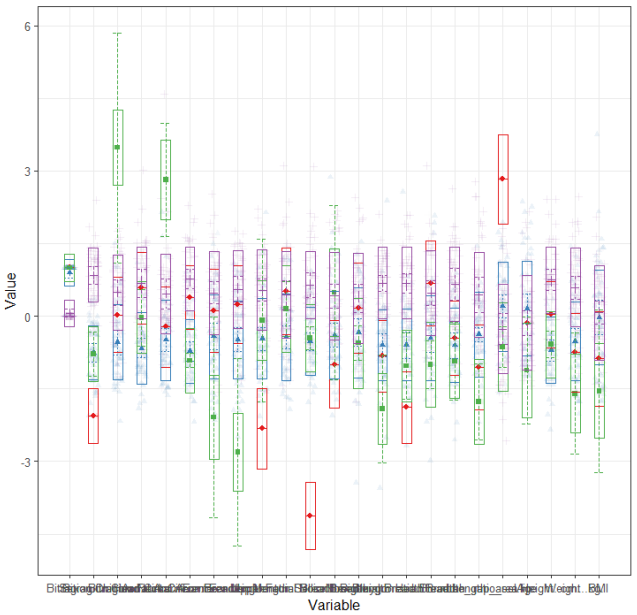


```
> cumsum(pve)
 [1] 0.1768919 0.2709679 0.3523139 0.4214075 0.4893952 0.5503635 0.602
4070 0.6502781 0.6971155 0.7395347 0.7794303 0.8180586 0.8522968 0.884
7522 0.9128825
[16] 0.9402529 0.9638701 0.9847841 1.0000000
```

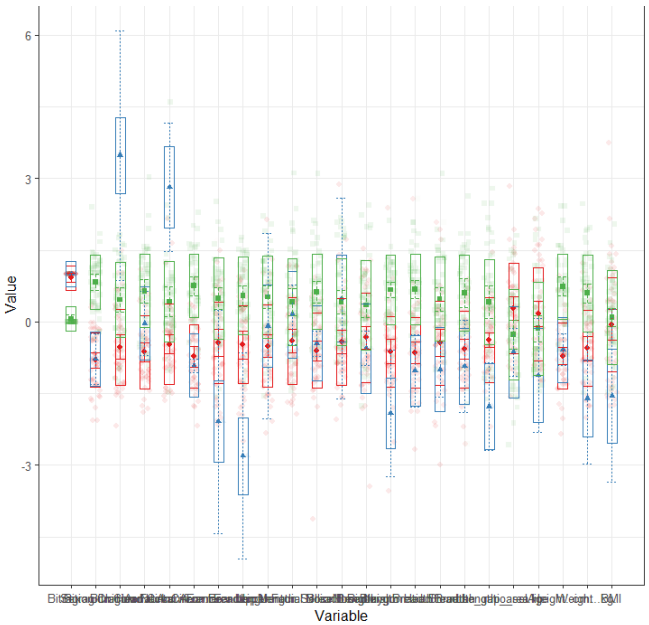
Factor Analysis, LCA, Bayesian profile regression, hierarchical clustering, k-means?

LCA

Example output with 4 classes with all variables



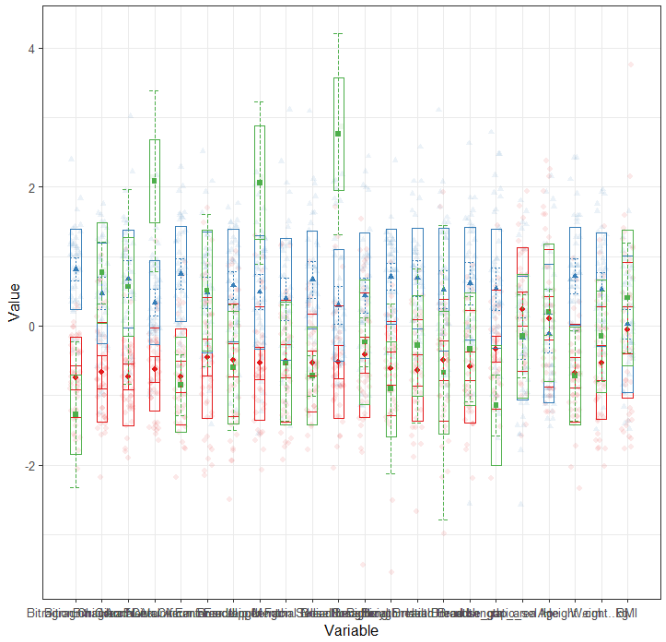
Class	n
<dbl>	<int>
1	1
2	2
3	3
4	4



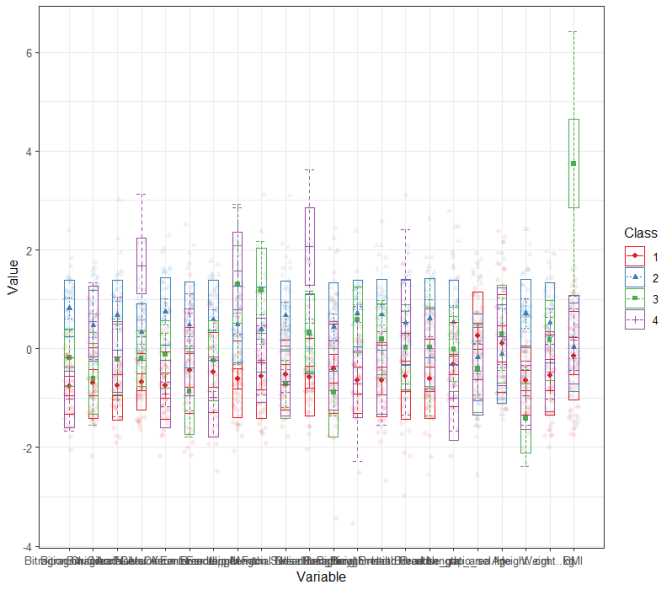
3 classes with all variables

Class	n
<dbl>	<int>
1	1
2	2
3	3

Without 7 outliers and sex



Class	n
<dbl>	<int>
1	46
2	46
3	1



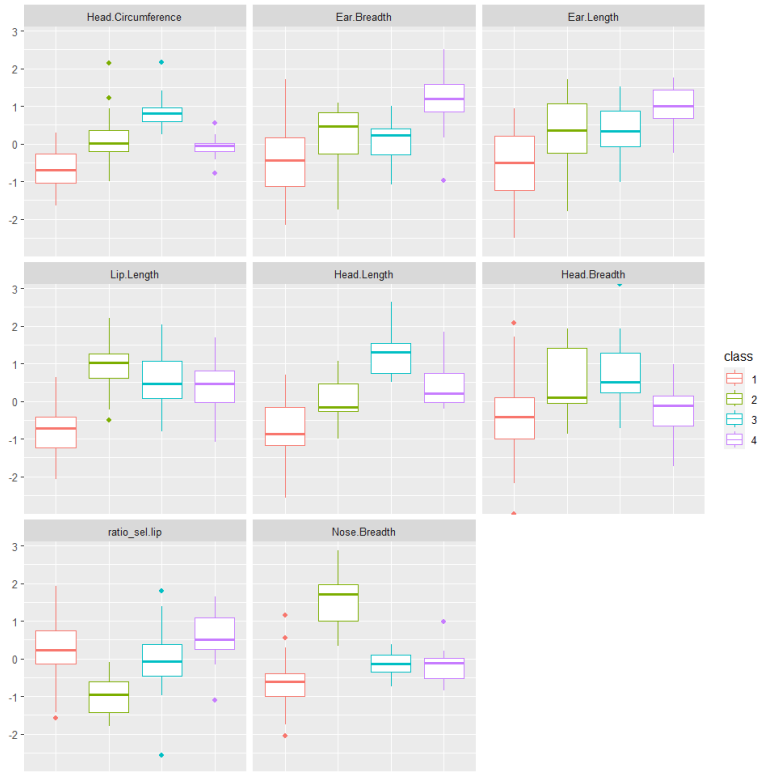
Class	n
<dbl>	<int>
1	21
2	32
3	38
4	2

Compare tidyLPA solutions:

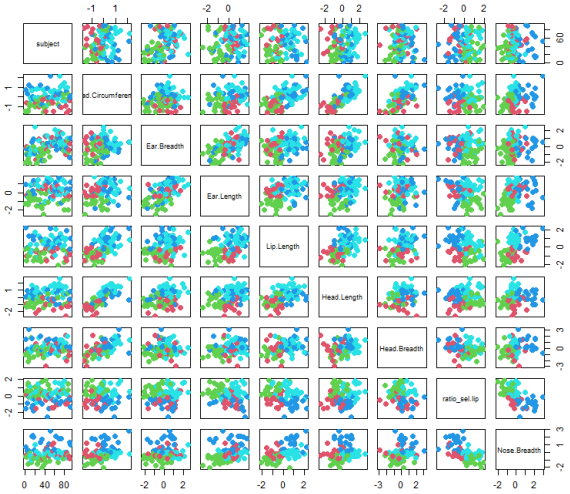
Model	Classes	AIC	BIC
1	2	5128.994	5298.678
1	3	5031.592	5259.526
1	4	5011.852	5298.036
1	5	5047.539	5391.973

Best model according to AIC is Model 1 with 4 classes.
Best model according to BIC is Model 1 with 3 classes.

With smaller subset of variables



Kmeans clustering



K = 4, nstart = 20

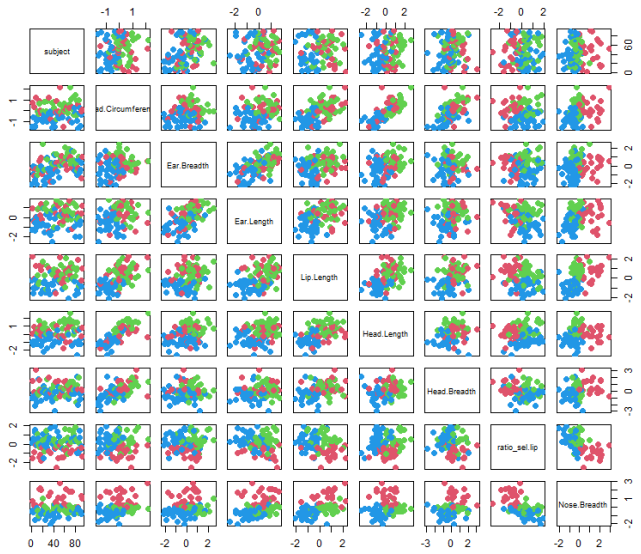
km.out$tot.withinss = 332.5704 (within-cluster sum of squares)

Within cluster sum of squares by cluster:
[1] 62.21006 79.89359 89.98450 100.48230
(between_SS / total_SS = 46.2 %)
(sum of squared distances of centroid mean of groups)/(sum of square distances of each data point to the global mean)

means clustering with 4 clusters of sizes 16, 26, 21, 29

Cluster means:

	Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
1	0.29172761	0.2648979	-0.7057184	-1.0131655	-0.1919569	-0.5646799	-0.1227434
2	-0.78499742	-0.9796451	-0.8448185	-0.5940373	-0.5632204	0.6882322	-0.9121198
3	-0.01488748	0.2722193	0.9067729	0.2267173	0.6520617	-0.9362417	1.2002521
4	0.64241935	0.6500054	0.4518716	0.9615764	0.1473800	0.4598564	-0.2573451



K=3, nstart = 20

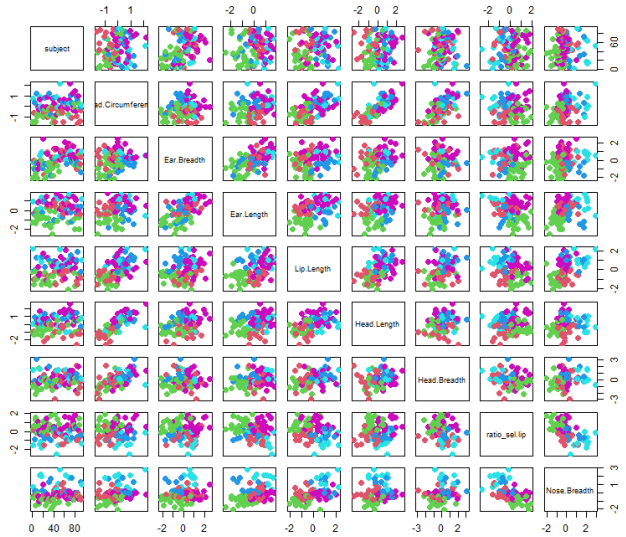
km.out$tot.withinss = 372.8572

Within cluster sum of squares by cluster:
[1] 107.9379 104.9004 160.0188
(between_SS / total_SS = 39.7 %)

K-means clustering with 3 clusters of sizes 24, 30, 38

Cluster means:

	Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
1	0.02711945	0.2649385	0.6818385	0.1172631	0.6193224	-0.9545004	1.1617096
2	0.65819053	0.6539821	0.4319330	0.9286597	0.1701395	0.4074541	-0.2656344
3	-0.46898289	-0.5958852	-0.8008522	-0.7811301	-0.5188328	0.3478502	-0.7328612



K= 5 , nstart = 20

km.out$tot.withinss = 306.874

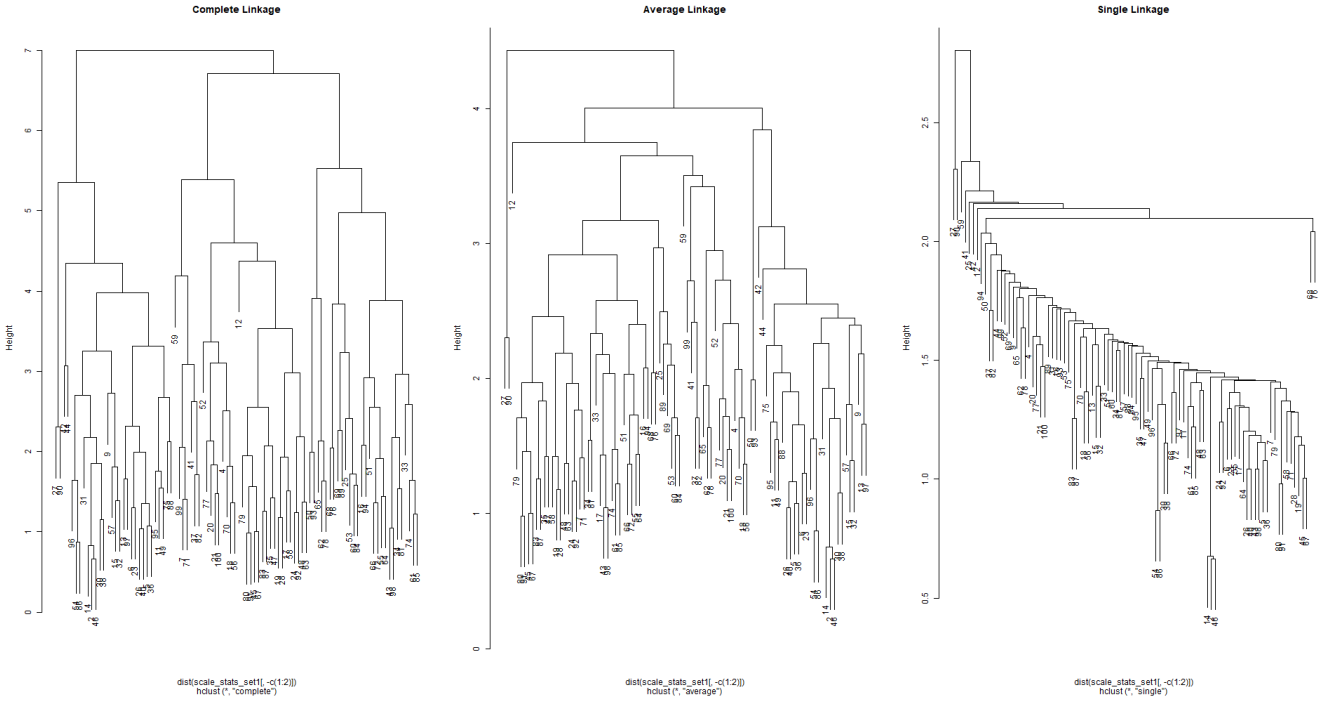
Within cluster sum of squares by cluster:
[1] 65.41579 79.89359 42.59014 35.53327 83.44119
(between_SS / total_SS = 50.4 %)

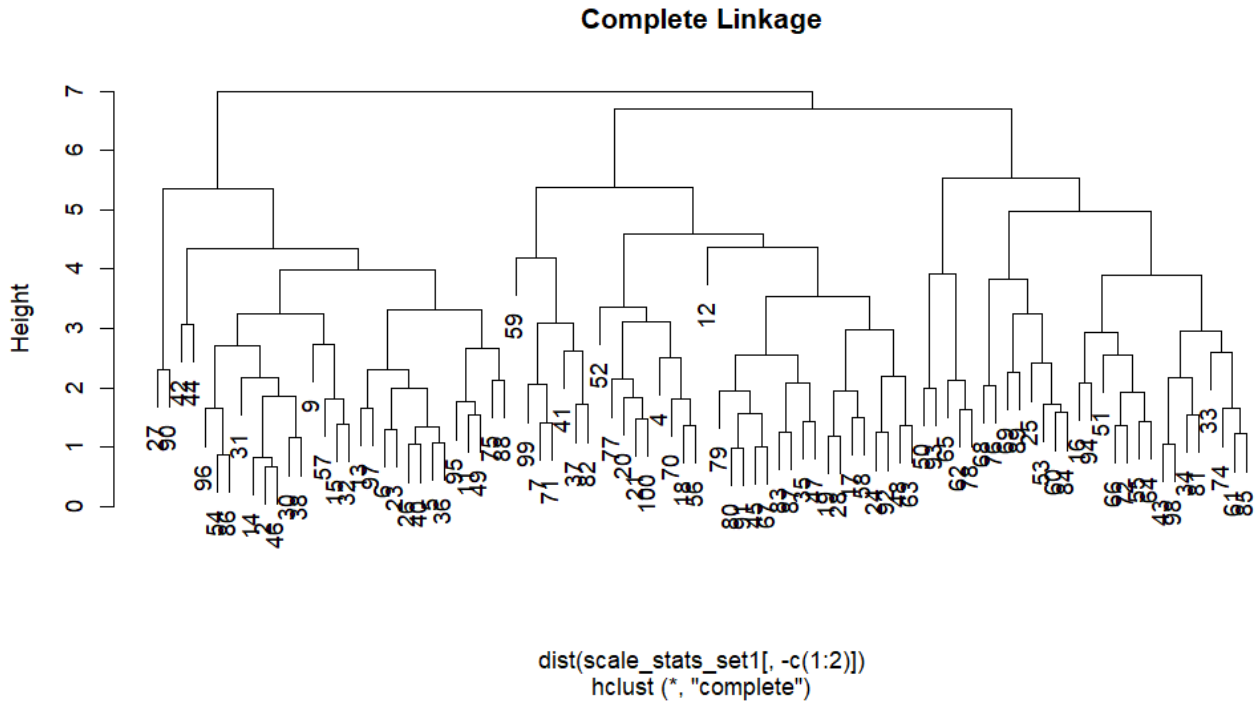
K-means clustering with 5 clusters of sizes 17, 26, 14, 11, 24

Cluster means:

	Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
1	0.3017580	0.2695526	-0.6660923	-0.9692783	-0.10594893	-0.5666258	-0.09506646
2	-0.7849974	-0.9796451	-0.8448185	-0.5940373	-0.56322036	0.6882322	-0.91211979
3	-0.5130530	-0.3820276	0.7992453	0.6803242	0.58396397	-0.4475035	0.46103420
4	0.5717324	0.8774907	1.0607625	0.1549441	0.57015394	-1.1934750	1.63569140
5	0.7812068	0.8299467	0.3883651	0.9035390	0.09374783	0.5694082	-0.29385679

Hierarchical Clustering





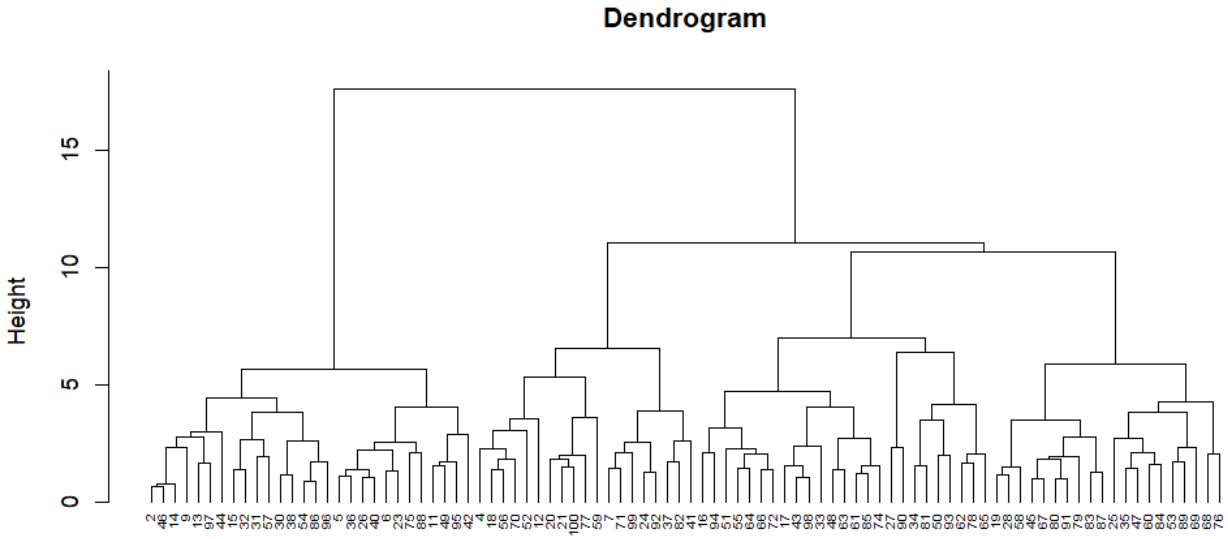
Cut at 6 clusters?

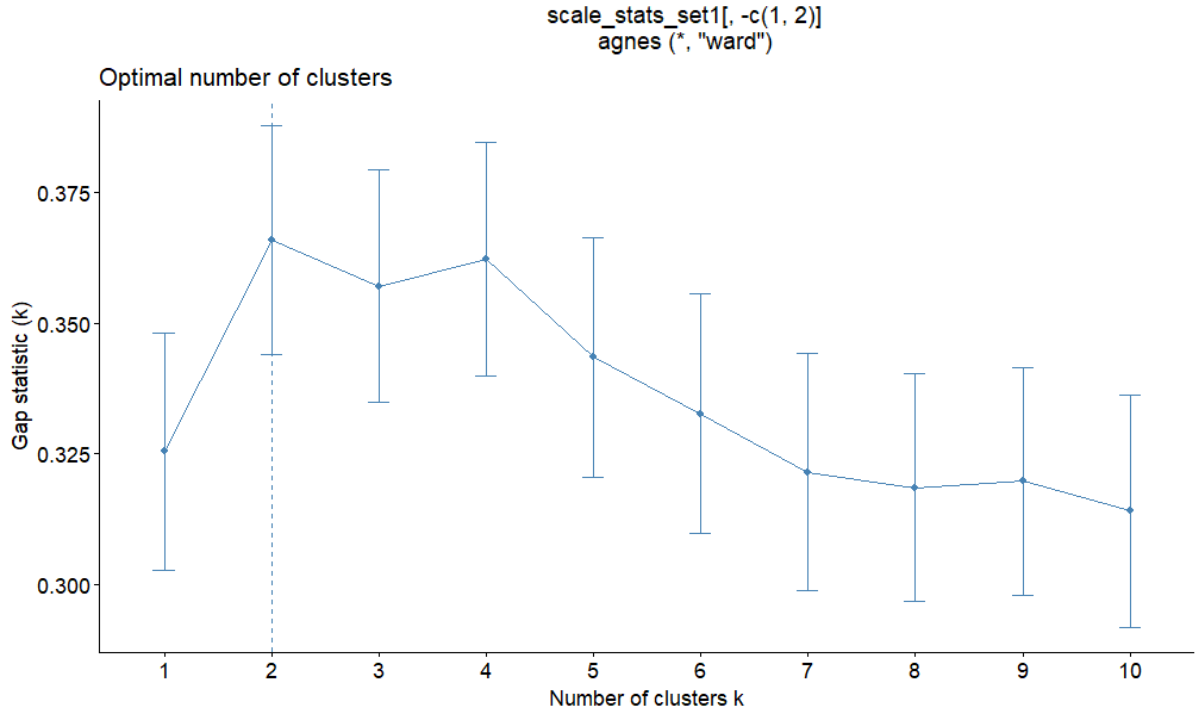
Another method

average single complete ward

0.6319527 0.4735829 0.7583956 0.9063407

Since Ward's method is closest to one it does the best job of putting the data in strong clusters





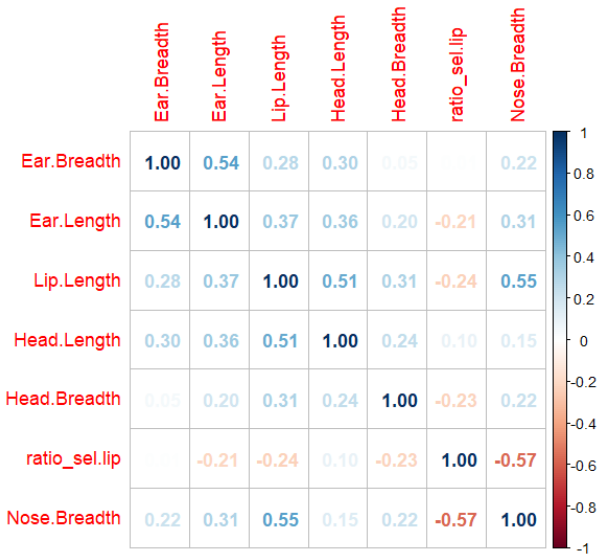
Two clusters is the optimal, but 4 is also strong.

groups
1 2 3 4
28 19 25 20

cluster	Head.Circumference	Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
1	-0.6848510	-0.76737710	-0.8861073	-0.8068998	-0.6008282	-0.64441377	0.6383809	-0.8463593
2	0.4122909	-0.09568445	0.2807804	0.9867346	0.3167739	0.66805222	-0.9530946	1.2119184
3	-0.4116520	0.45355312	0.2231551	-0.4336097	-0.4961828	0.07920286	-0.2374679	-0.1519401
4	0.5334525	0.72704842	0.8615826	0.6787605	1.2100108	0.18113988	0.4352375	-0.1733304

Compare Methods
Are clusters/classes/groups meaningful?

Factor Analysis



— Kaiser-Meyer-Olkin criterion (KMO)

! The overall KMO value for your data is mediocre.
These data are probably suitable for factor analysis.

Overall: 0.656

For each variable:

Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
0.633	0.702	0.693	0.658	0.787	0.547	0.632

Call:
factanal(x = scale_stats_set1[, -c(1, 2)], factors = 3, n.obs = 7776, scores = "Bartlett")

Uniquenesses:

Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
0.556	0.287	0.005	0.623	0.872	0.005	0.484

Loadings:

	Factor1	Factor2	Factor3
Ear.Breadth		0.171	0.644
Ear.Length	0.221	0.181	0.795
Lip.Length	0.274	0.942	0.179
Head.Length		0.496	0.352
Head.Breadth	0.237	0.244	0.111
ratio_sel.lip	-0.997		
Nose.Breadth	0.591	0.379	0.151

	Factor1	Factor2	Factor3
SS loadings	1.530	1.401	1.237
Proportion Var	0.219	0.200	0.177
Cumulative Var	0.219	0.419	0.595

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 4.1 on 3 degrees of freedom.
The p-value is 0.251

3 Factors

Uniquenesses:

Head.Circumference	Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
0.005	0.498	0.492	0.460	0.285	0.591	0.439	0.302

Loadings:

	Factor1	Factor2	Factor3
Head.Circumference	0.977		0.195
Ear.Breadth			0.703
Ear.Length	0.147	0.201	0.668
Lip.Length	0.507	0.377	0.375
Head.Length	0.722	-0.144	0.415
Head.Breadth	0.591	0.235	
ratio_sel.lip		-0.749	
Nose.Breadth	0.223	0.767	0.245

	Factor1	Factor2	Factor3
SS loadings	2.161	1.410	1.356
Proportion Var	0.270	0.176	0.170
Cumulative Var	0.270	0.446	0.616

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 25.56 on 7 degrees of freedom.
The p-value is 0.000604 (significant p-value indicates more factors are needed)

4 Factors

Uniquenesses:

Head.Circumference	Ear.Breadth	Ear.Length	Lip.Length	Head.Length	Head.Breadth	ratio_sel.lip	Nose.Breadth
0.066	0.669	0.005	0.464	0.198	0.155	0.571	0.005

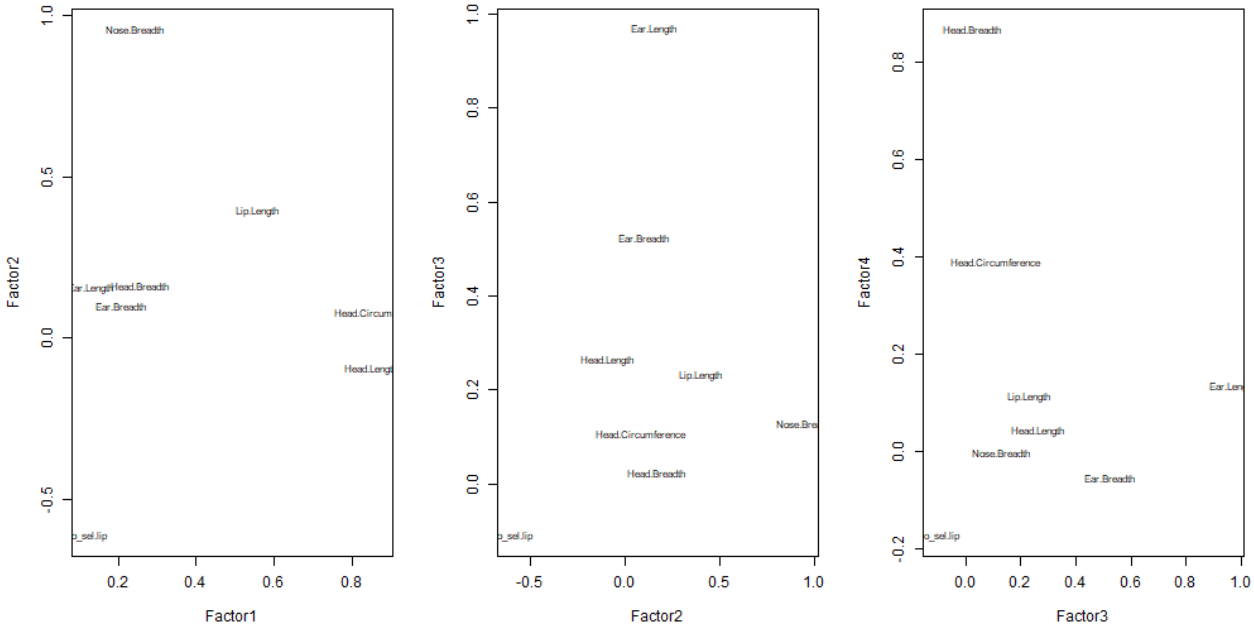
Loadings:

	Factor1	Factor2	Factor3	Factor4
Head.Circumference	0.873		0.110	0.392
Ear.Breadth	0.207	0.101	0.525	
Ear.Length	0.130	0.155	0.968	0.133
Lip.Length	0.558	0.398	0.232	0.114
Head.Length	0.850		0.263	
Head.Breadth	0.254	0.165		0.868
ratio_sel.lip	0.110	-0.612	-0.110	-0.172
Nose.Breadth	0.242	0.959	0.130	

Negative numbers indicate a contrast between the variables

	Factor1	Factor2	Factor3	Factor4
SS loadings	1.990	1.530	1.376	0.971
Proportion Var	0.249	0.191	0.172	0.121
Cumulative Var	0.249	0.440	0.612	0.733

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 2.56 on 2 degrees of freedom.
The p-value is 0.279



Factor loadings close to 1 or -1 indicate that the factor has a strong influence on the variable.
Try different rotation methods to interpret factor loadings better. (varimax, equimax, quartimax)

Compare Groupings

Hierarchical/k-means

> table(scale_stats_cats$cluster,scale_stats_cats$k_clusters)

	1	2	3	4	5
1	0	1	25	2	0
2	9	0	0	0	10
3	2	6	1	15	1
4	3	17	0	0	0

Hierarchical/LCA

> table(scale_stats_cats$cluster,scale_stats_cats$class)

	1	2	3	4
1	27	0	0	1
2	0	14	5	0
3	14	3	3	5
4	0	0	12	8

LCA/k-means

> table(scale_stats_cats$class,scale_stats_cats$k_clusters)

	1	2	3	4	5
1	0	1	26	14	0
2	5	0	0	2	10
3	9	10	0	0	1
4	0	13	0	1	0

Outcome Prediction
(cross validation,lowest RMSE)

Also try supervised learning (xgboost, random forest, neural network, LDA, penalized regression to find important variables and better prediction)

Hierarchical clustering

```
glm.fit<-glm(KN95~cluster, data = complete_stats)
cv.err<-cv.glm(complete_stats, glm.fit)
cv.err$delta          #146.8020 146.7339

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-24.1474   -9.0115    0.3363    8.2062   24.8240

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   64.418      2.241  28.741  < 2e-16 ***
cluster2       8.430      3.525   2.391  0.01892 *
cluster3       2.558      3.263   0.784  0.43522
cluster4      11.362      3.472   3.272  0.00153 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 140.6621)

    Null deviance: 14260  on 91  degrees of freedom
Residual deviance: 12378  on 88  degrees of freedom
AIC: 722.06

Number of Fisher Scoring iterations: 2
```

K-means Clustering

```
cv.err$delta         #140.4521 140.3740

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-22.525   -8.152   -1.009    8.708   23.607

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  67.6929     3.0910  21.900  < 2e-16 ***
k_clusters2   4.6321     3.8895   1.191  0.23691
k_clusters3  -4.1775     3.8340  -1.090  0.27890
k_clusters4   0.1777     4.1741   0.043  0.96613
k_clusters5  13.1344     4.6599   2.819  0.00597 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 133.7633)

    Null deviance: 14260  on 91  degrees of freedom
Residual deviance: 11637  on 87  degrees of freedom
AIC: 718.38
```

Factor Analysis

```
cv.err$delta          #150.7898 150.6986

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-22.7096   -7.7240   -0.9772    8.9471   25.1268
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  69.3239     1.2441  55.724   <2e-16 ***
Factor1       2.4885     1.1970   2.079   0.0406 *
Factor2       3.1403     1.2440   2.524   0.0134 *
Factor3       1.9509     1.2458   1.566   0.1210
Factor4       0.7048     1.1400   0.618   0.5380
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 142.3876)

    Null deviance: 14260  on 91  degrees of freedom
Residual deviance: 12388  on 87  degrees of freedom
AIC: 724.13
```

Latent Class Analysis

```
cv.err$delta        #151.1825 151.1096

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-24.594  -7.605   0.000   9.218  22.000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   64.800      1.877  34.530   <2e-16 ***
class2         8.494      3.466   2.450   0.0162 *
class3         7.130      3.277   2.175   0.0323 *
class4         9.229      3.720   2.481   0.0150 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 144.395)

    Null deviance: 14260  on 91  degrees of freedom
Residual deviance: 12707  on 88  degrees of freedom
AIC: 724.47
```

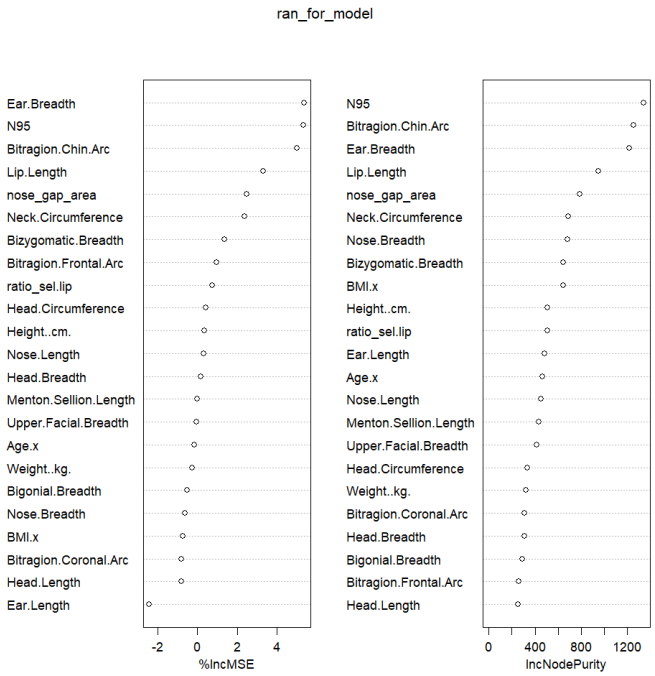
Supervised Learning

Random Forest

```
        Mean of squared residuals: 142.4891
              % Var explained: 8.07

                Number of trees: 500
No. of variables tried at each split: 5

        Mean of squared residuals: 139.0716
              % Var explained: 10.27
```

Importance indicates the total amount that the amount that the error is decreased due to splits in that variable. Larger % shows more importance because it is reducing the error (residual sum squares (RSS) or mean square error (MSE))

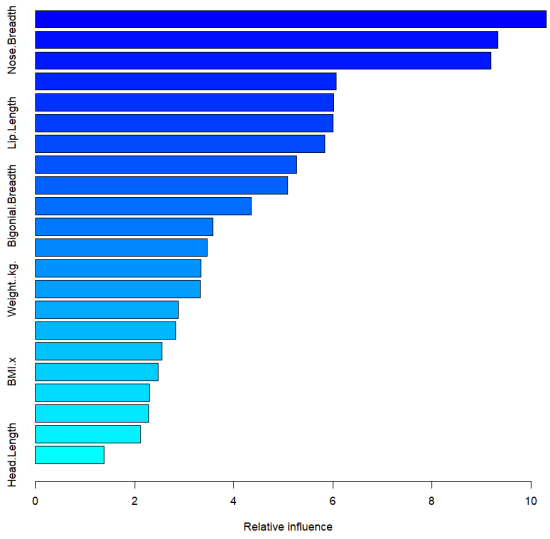
	%IncMSE	IncNodePurity
Bitragion.Chin.Arc	5.024717760	1256.9559
Bitragion.Coronal.Arc	-0.779033746	310.6126
Bitragion.Frontal.Arc	0.988387545	257.4253
Head.Circumference	0.417898491	332.1163
Neck.Circumference	2.362752430	687.5447
Ear.Breadth	5.385025178	1214.7591
Ear.Length	-2.407957275	480.8372
Lip.Length	3.304436818	951.2186
Upper.Facial.Breadth	-0.049421776	414.0601
Menton.Sellion.Length	0.005477826	433.7530
Nose.Breadth	-0.597461227	684.7406
Nose.Length	0.329356475	452.4049
Bigonial.Breadth	-0.491994626	293.6838
Bizygomatic.Breadth	1.380379078	646.9542
Head.Breadth	0.186922114	310.0315
Head.Length	-0.794316736	250.8811
nose_gap_area	2.496648011	789.2567
ratio_sel.lip	0.754019448	505.5186
Age.x	-0.135430154	462.7062
Height..cm.	0.372286170	507.9880
Weight..kg.	-0.263289913	318.6144
BMI.x	-0.736760054	642.6659
N95	5.319027997	1343.7329

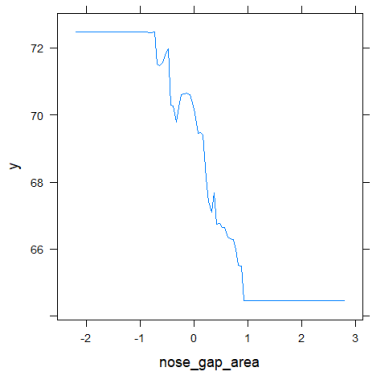


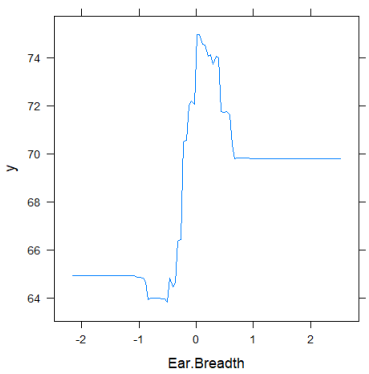
Boosting

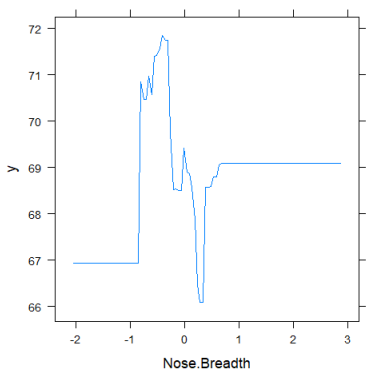
KN95

	var	rel.inf
nose_gap_area	nose_gap_area	10.305552
Nose.Breadth	Nose.Breadth	9.331326
Ear.Breadth	Ear.Breadth	9.196040
Ear.Length	Ear.Length	6.062168
Height..cm.	Height..cm.	6.015255
Lip.Length	Lip.Length	6.009020
Upper.Facial.Breadth	Upper.Facial.Breadth	5.845039
ratio_sel.lip	ratio_sel.lip	5.263364
Neck.Circumference	Neck.Circumference	5.092802
Bigonial.Breadth	Bigonial.Breadth	4.358403
Bitragion.Chin.Arc	Bitragion.Chin.Arc	3.574952
Nose.Length	Nose.Length	3.462573
Menton.Sellion.Length	Menton.Sellion.Length	3.342496
Weight..kg.	Weight..kg.	3.322764
Bizygomatic.Breadth	Bizygomatic.Breadth	2.882803
Bitragion.Coronal.Arc	Bitragion.Coronal.Arc	2.832282
Age.x	Age.x	2.545495
BMI.x	BMI.x	2.470395
Head.Breadth	Head.Breadth	2.300711
Bitragion.Frontal.Arc	Bitragion.Frontal.Arc	2.286070
Head.Circumference	Head.Circumference	2.117332
Head.Length	Head.Length	1.383157



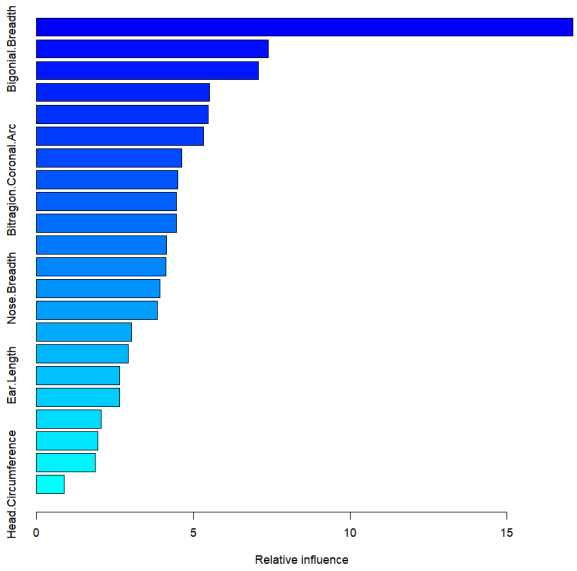






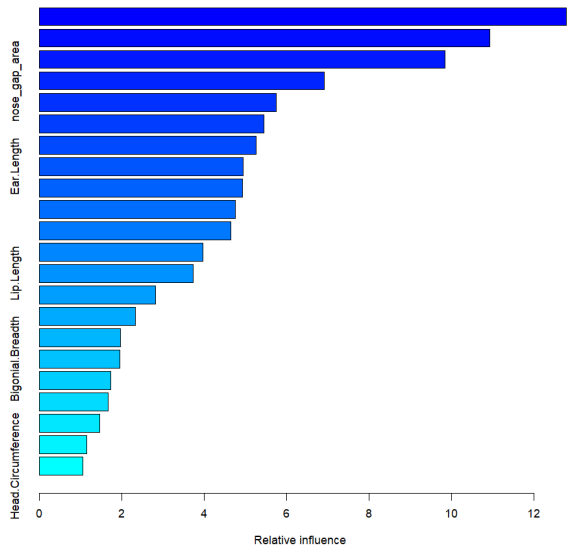
KF94

Bitragion.Chin.Arc	Bitragion.Chin.Arc 17.0976217
Bigonial.Breadth	Bigonial.Breadth 7.3841740
Menton.Sellion.Length	Menton.Sellion.Length 7.0697379
Bizygomatic.Breadth	Bizygomatic.Breadth 5.5107890
Age.x	Age.x 5.4666848
Head.Length	Head.Length 5.3352446
ratio_sel.lip	ratio_sel.lip 4.6285604
Bitragion.Coronal.Arc	Bitragion.Coronal.Arc 4.5087110
Neck.Circumference	Neck.Circumference 4.4669091
nose_gap_area	nose_gap_area 4.4632674
Lip.Length	Lip.Length 4.1521970
Ear.Breadth	Ear.Breadth 4.1325515
Nose.Breadth	Nose.Breadth 3.9256349
Weight..kg.	Weight..kg. 3.8500504
Nose.Length	Nose.Length 3.0236146
BMI.x	BMI.x 2.9240620
Ear.Length	Ear.Length 2.6549362
Bitragion.Frontal.Arc	Bitragion.Frontal.Arc 2.6419446
Height..cm.	Height..cm. 2.0567032
Upper.Facial.Breadth	Upper.Facial.Breadth 1.9532858
Head.Breadth	Head.Breadth 1.8676215
Head.Circumference	Head.Circumference 0.8856985



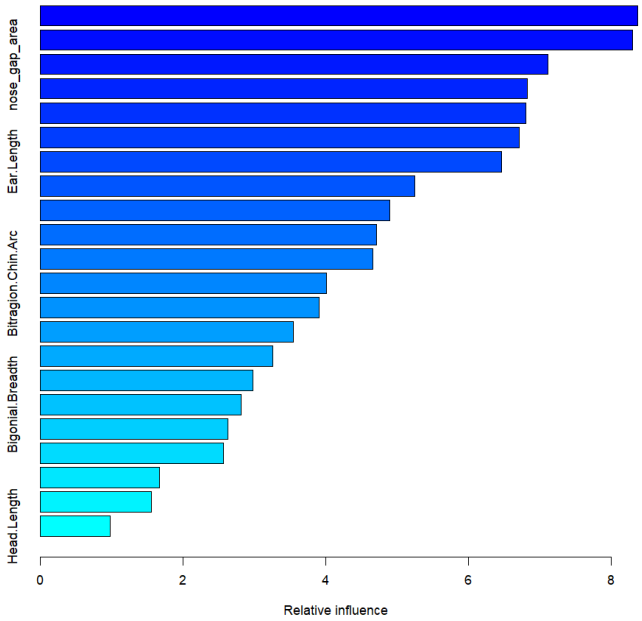
Surgical

	var rel.inf
Bitragion.Chin.Arc	Bitragion.Chin.Arc 12.780778
Bizygomatic.Breadth	Bizygomatic.Breadth 10.926142
Nose.Breadth	Nose.Breadth 9.841078
nose_gap_area	nose_gap_area 6.911546
ratio_sel.lip	ratio_sel.lip 5.750443
Upper.Facial.Breadth	Upper.Facial.Breadth 5.451953
BMI.x	BMI.x 5.255903
Ear.Length	Ear.Length 4.936017
Menton.Sellion.Length	Menton.Sellion.Length 4.927084
Nose.Length	Nose.Length 4.757492
Ear.Breadth	Ear.Breadth 4.644290
Age.x	Age.x 3.971384
Lip.Length	Lip.Length 3.727854
Bitragion.Coronal.Arc	Bitragion.Coronal.Arc 2.816576
Neck.Circumference	Neck.Circumference 2.322960
Weight..kg.	Weight..kg. 1.962327
Bigonial.Breadth	Bigonial.Breadth 1.942871
Head.Breadth	Head.Breadth 1.733376
Head.Length	Head.Length 1.669605
Height..cm.	Height..cm. 1.464615
Bitragion.Frontal.Arc	Bitragion.Frontal.Arc 1.149922
Head.Circumference	Head.Circumference 1.05578



MKF94

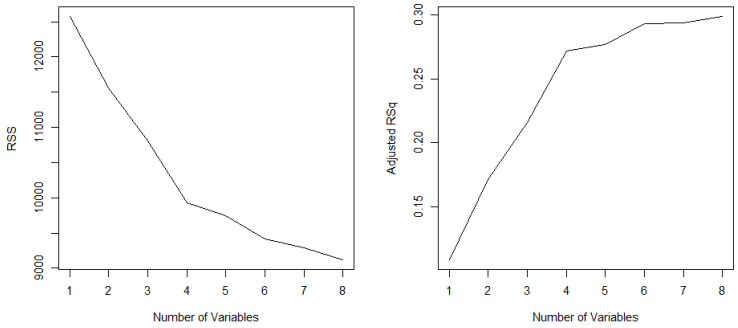
var	rel.inf
Age.x	Age.x 8.3669942
Nose.Breadth	Nose.Breadth 8.2995780
nose_gap_area	nose_gap_area 7.1143605
Lip.Length	Lip.Length 6.8263855
BMI.x	BMI.x 6.7981118
Bizygomatic.Breadth	Bizygomatic.Breadth 6.7069722
Ear.Length	Ear.Length 6.4638645
Ear.Breadth	Ear.Breadth 5.2478557
Upper.Facial.Breadth	Upper.Facial.Breadth 4.9003659
ratio_sel.lip	ratio_sel.lip 4.7104439
Neck.Circumference	Neck.Circumference 4.6611234
Bitragion.Chin.Arc	Bitragion.Chin.Arc 4.0056550
Bitragion.Frontal.Arc	Bitragion.Frontal.Arc 3.9094330
Weight..kg.	Weight..kg. 3.5446295
Menton.Sellion.Length	Menton.Sellion.Length 3.2581455
Head.Breadth	Head.Breadth 2.9781748
Bigonial.Breadth	Bigonial.Breadth 2.8127678
Bitragion.Coronal.Arc	Bitragion.Coronal.Arc 2.6267851
Nose.Length	Nose.Length 2.5643238
Height..cm.	Height..cm. 1.6718965
Head.Circumference	Head.Circumference 1.5531240
Head.Length	Head.Length 0.9790094

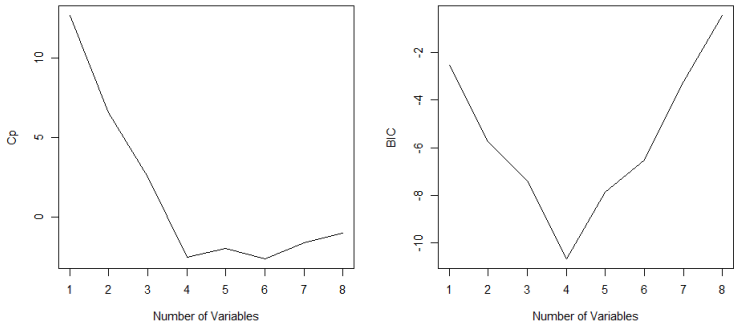


KN95	Surgical	KF94	MKF94
nose_gap_area 10.305552	Bitragion.Chin.Arc 12.780778	Bitragion.Chin.Arc 17.0976217	Age.x 8.3669942
Nose.Breadth 9.331326	Bizygomatic.Breadth 10.926142	Bigonial.Breadth 7.3841740	Nose.Breadth 8.2995780
Ear.Breadth 9.196040	Nose.Breadth 9.841078	Menton.Sellion.Length 7.0697379	nose_gap_area 7.1143605
Ear.Length 6.062168	nose_gap_area 6.911546	Bizygomatic.Breadth 5.5107890	Lip.Length 6.8263855
Height..cm. 6.015255	ratio_sel.lip 5.750443	Age.x 5.4666848	BMI.x 6.7981118
Lip.Length 6.009020	Upper.Facial.Breadth 5.451953	Head.Length 5.3352446	Bizygomatic.Breadth 6.7069722
Upper.Facial.Breadth 5.845039	BMI.x 5.255903		Ear.Length 6.4638645
ratio_sel.lip 5.263364			Ear.Breadth 5.2478557
Neck.Circumference 5.092802			

1-Black
2-Green
3-purple
Shared with Elastic Net Regression Variables

Best Subset Selection
KN95





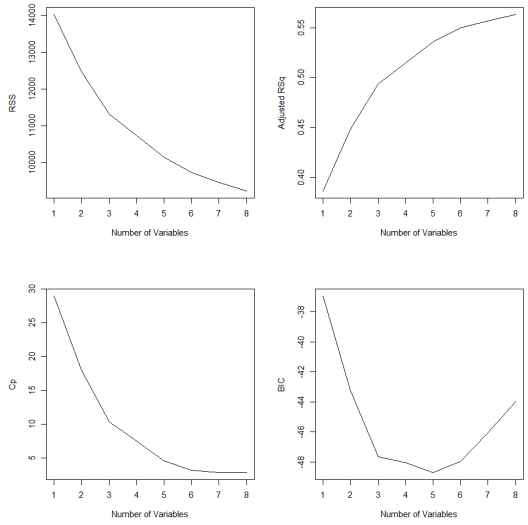
4 Variables

(Intercept)	Ear.Breadth	Lip.Length	Menton.Sellion.Length	nose_gap_area
69.477436	3.354625	3.369148	4.244381	-5.218057

6 Variables

(Intercept)	Ear.Breadth	Lip.Length	Upper.Facial.Breadth	Menton.Sellion.Length	Bizygomatic.Breadth
69.246286	2.702040	3.495693	-2.393592	4.279744	2.674251

nose_gap_area
-5.292972



KF94

(Intercept)	Neck.Circumference	Menton.Sellion.Length
5.150786	5.546481	5.344939
Bizygomatic.Breadth	**nose_gap_area**	**Age.x**
5.507095	-3.357483	4.726084

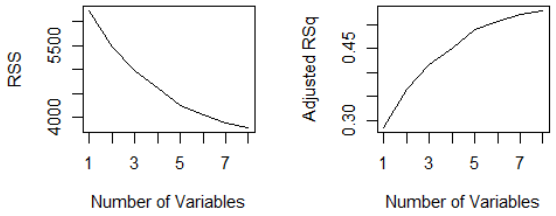
Surgical

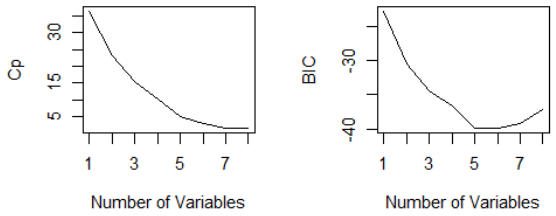
5 variables

(Intercept)	Bitragion.Chin.Arc	Ear.Breadth	Menton.Sellion.Length	Nose.Length	nose_gap_area
57.197332	4.751789	2.278044	4.246167	-3.941203	-3.222140

6 variables

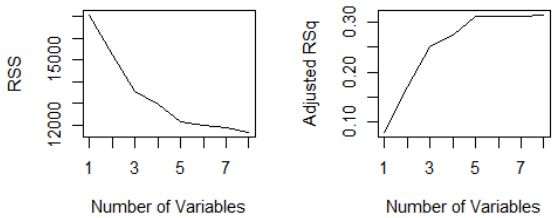
(Intercept)	Bitragion.Chin.Arc	Ear.Breadth	Menton.Sellion.Length	Nose.Breadth	Nose.Length	nose_gap_area
57.329008	3.786049	2.139793	4.153065	1.890864	-3.956852	-2.469667

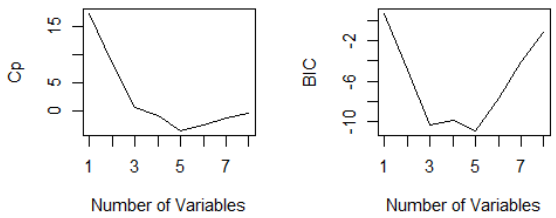




MKF94

(Intercept)	Bitragion.Chin.Arc	Neck.Circumference	Ear.Length	nose_gap_area	Age.x
66.819995	6.345920	4.740477	-6.230462	-3.426972	6.030945





Elastic Net Regression (nonzero coefficients)

KN95	s1
(Intercept)	69.427926563
Bitragion.Chin.Arc	0.868310870
Head.Circumference	0.250662861
Neck.Circumference	1.100548082
Ear.Breadth	1.847755691
Lip.Length	1.694117054
Upper.Facial.Breadth	-0.785949160
Menton.Sellion.Length	1.015436809
Nose.Breadth	0.562156594
Nose.Length	0.673638269
Bigonial.Breadth	0.430640694
Bizygomatic.Breadth	0.934639906
Head.Breadth	0.003943001
nose_gap_area	-3.011890165
Age.x	0.404811635
BMI.x	-0.598810227

KF94

	s1
(Intercept)	55.187267238
Bitragion.Chin.Arc	2.286371160
Bitragion.Coronal.Arc	0.974059176
Bitragion.Frontal.Arc	0.881650356
Head.Circumference	0.156788285
Neck.Circumference	1.985462690
Ear.Breadth	1.355486323
Upper.Facial.Breadth	-1.003675766
Menton.Sellion.Length	2.801212790
Nose.Breadth	0.512539091
Nose.Length	-0.004712357
Bigonial.Breadth	2.119372556
Bizygomatic.Breadth	1.566417250
Head.Length	0.482498608
nose_gap_area	-1.808991830
ratio_sel.lip	-0.562179661
Age.x	3.048380048
Weight..kg.	1.431855205

Surgical

	s1
(Intercept)	57.16858500
Bitragion.Chin.Arc	2.52950592
Bitragion.Coronal.Arc	-0.54630017
Head.Circumference	-0.98116649
Neck.Circumference	0.86199879
Ear.Breadth	1.83003310
Ear.Length	-0.57717693
Upper.Facial.Breadth	-0.09360404
Menton.Sellion.Length	2.95144412
Nose.Breadth	2.02968682
Nose.Length	-2.86874968
Bizygomatic.Breadth	2.15101967
nose_gap_area	-1.75495990
Age.x	0.19971203

MKF94

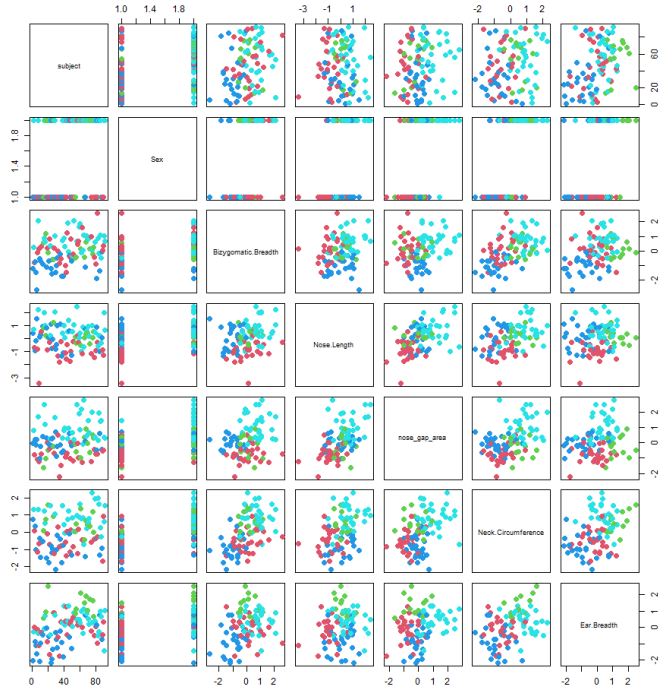
	s1
(Intercept)	66.6839526
Bitragion.Chin.Arc	2.5236911
Head.Circumference	-0.1976336
Neck.Circumference	2.6639496

Ear.Length -3.4388664
Lip.Length 1.2209772
Menton.Sellion.Length 0.5617950
Nose.Breadth 0.5647308
Bigonial.Breadth 0.2579257
Bizygomatic.Breadth 0.8553541
nose_gap_area -1.8470165
Age.x 4.2235796
BMI.x 0.8374927

Clustering with selected variables

(Bizygomatic Breadth, Nose Length, Nose Gap Area, Neck Circumference, Ear Breadth)

4 k-means clusters



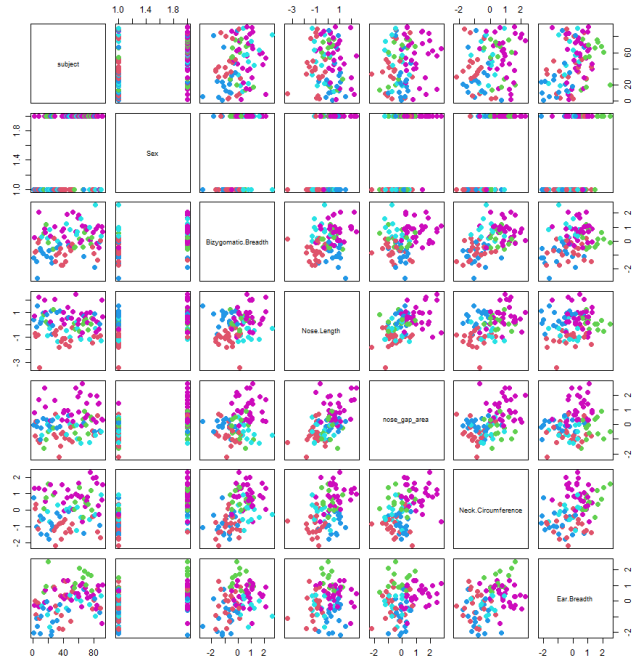
K-means clustering with 4 clusters of sizes 25, 13, 24, 30

Cluster means:

	Bizygomatic.Breadth	Nose.Length	nose_gap_area	Neck.Circumference	Ear.Breadth
1	0.1482384	-0.92412932	-0.75632430	-0.5393143	-0.1610692
2	0.1076945	0.19422189	-0.28171767	0.5775855	1.3355823
3	-1.0482425	0.06596069	-0.06428665	-0.8869017	-0.8833717
4	0.7493256	0.67971436	1.08701650	0.9289105	0.3480105

Within cluster sum of squares by cluster:
[1] 62.71906 21.03027 56.50618 78.93027
(between_SS / total_SS = 51.4 %)

5 Clusters



K-means clustering with 5 clusters of sizes 20, 13, 17, 12, 30

Cluster means:

	Bizygomatic.Breadth	Nose.Length	nose_gap_area	Neck.Circumference	Ear.Breadth
1	-0.7411918	-1.1274833	-0.67937915	-0.9549055	-0.2691778
2	0.1076945	0.1942219	-0.28171767	0.5775855	1.3355823
3	-0.9776698	0.4995841	-0.04710818	-0.7806054	-0.9439581
4	0.8326970	-0.6219534	-0.50521375	-0.2000081	-0.3164007
5	0.7493256	0.6797144	1.08701650	0.9289105	0.3480105

Within cluster sum of squares by cluster:
[1] 45.07786 21.03027 32.80232 20.57489 78.93027
(between_SS / total_SS = 56.0 %)

Relative Influence

Add up to 100 or have the most important have 100. Most important is the variable that reduces the error the most when it is used for splitting. It has nothing to do with how it describes the variance. In gbm() the relative influence is averaged across all trees generated by the boosting algorithm. (empirical improvements by splitting on x at that point, best at training the model). Sex has low importance because it is the only binary variable. It can only be used once in a tree, while other variables can appear multiple times. Frequency gives it low importance even if it does a good job splitting the data and generating a prediction.