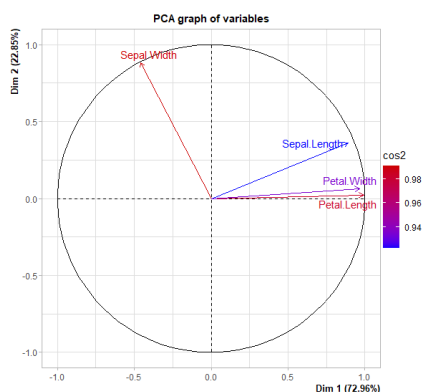


TP 2 Analyse de données

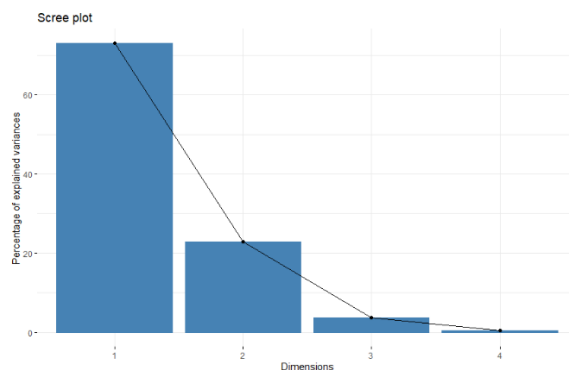
Arthur Crochemore et Alaâ Chakori Semmane

Partie 1 :

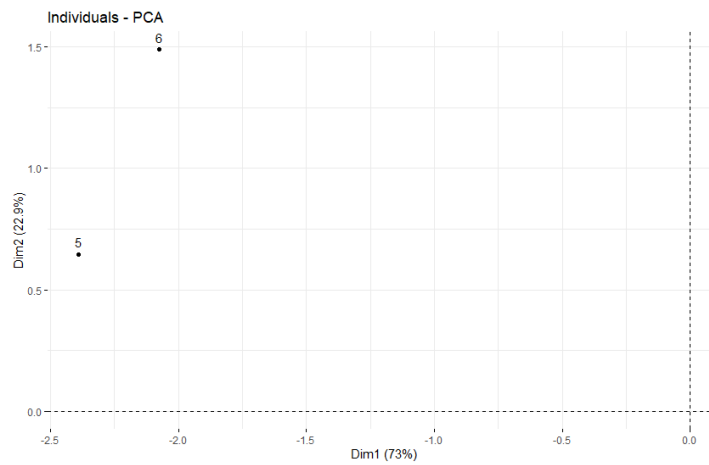
- On Sauvegarde des données iris dans un fichier **.xlsx**
- Matrice de corrélation, qui sont stocke dans un fichier **.xlsx**
- Matrice de covariance variance, qui sont stocké dans un fichier **.xlsx**
- Valeur moyenne de des variables quantitatives dans un fichier **.xlsx**
- Pour chaque variable quantitative on calcule sa variance qui est stocké dans un fichier **.xlsx**
- On normalise les données et celles-ci sont stocké dans un fichier **.xlsx**
- On fait la moyenne des valeurs centré des valeurs, qui sont stocké dans un fichier **.xlsx**
(RQ : on constate que celles-ci sont nulles ce qui est cohérent avec les cours)
- On fait la variance des valeurs centrée et réduite
(On obtient des valeurs différentes qui sont stocké dans un fichier **.xlsx**)
- On réalise un ACP et on réalise la projection du nuage des variables sur les deux axes factoriels.
(RQ : plus la couleur tend vers le rouge plus sa déformation sera faible)



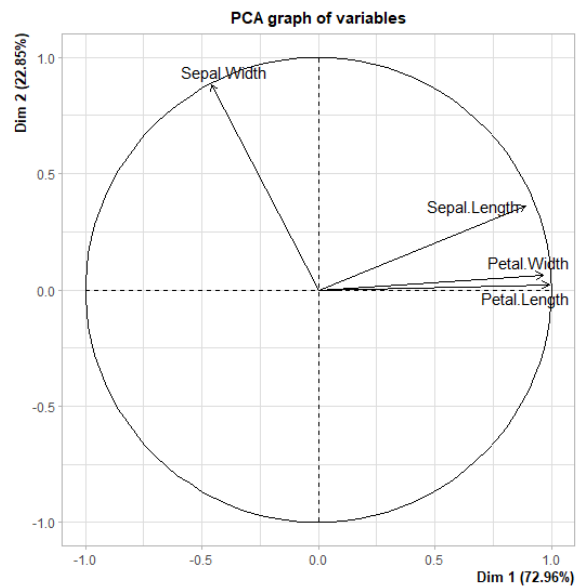
- On réalise une visualisation de la valeur propre de chaque axe généré par l'ACP



- On réalise une projection du Nuages des Individus sur les deux premières composantes principales.



- On réalise un ACP et on réalise la projection du nuage des variables sur les deux axes factoriels.
(RQ : plus la couleur tend vers le rouge plus son inertie ??? la déformation est faible)



Partie 2

Question 1 : Donner et exécuter la commande qui permet de retrouver l'ensemble des valeurs fournies par l'ACP.

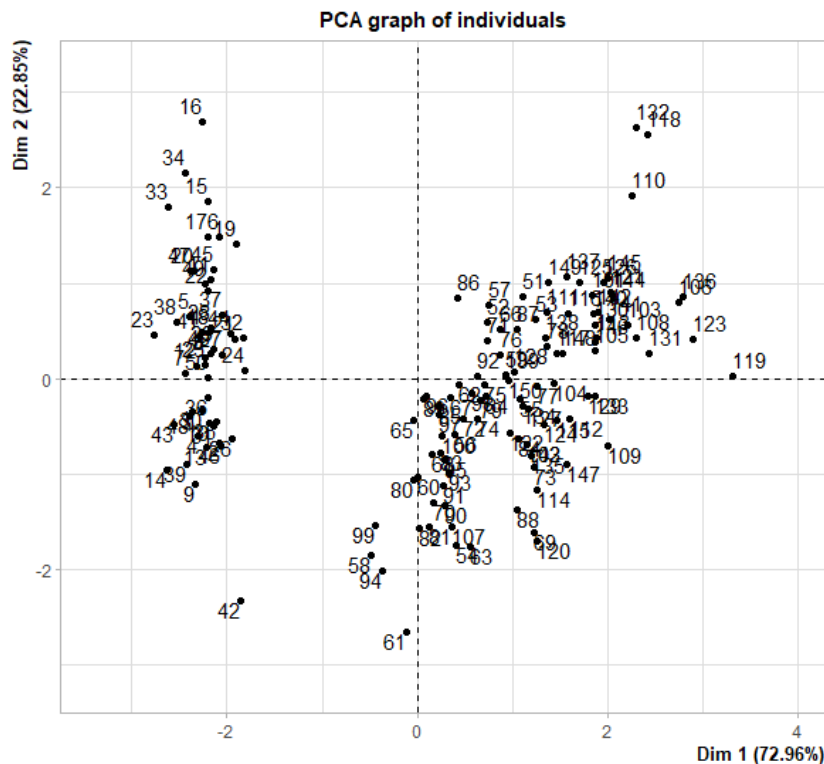
```
res.PCA = PCA(iris[,1:4], scale.unit = TRUE, ncp=4, graph=F)
```

Question 2 : Examiner la matrice des corrélations entre les variables. Faites un commentaire.

La largeur des sépales, n'as pas de corrélation forte avec les 3 autres variables (-0.118, -0.428 et -0.366).

Tandis que les variables des pétales sont fortement corrélées entre elles (0.963), et que chacune d'elles est très corrélée avec la longueur des sépales (respectivement 0.872 et 0.818).

Question 3 : Examiner le nuage de points des individus :



Donner l'expression et la valeur de l'inertie du premier plan factoriel ?

```
result$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.91849782	72.9624454	72.96245
comp 2	0.91403047	22.8507618	95.81321
comp 3	0.14675688	3.6689219	99.48213
comp 4	0.02071484	0.5178709	100.00000

Comme le premier plan factoriel est composé de l'axe 1 et 2, on peut facilement déterminer son

inertie en faisant la somme des deux premières composantes principales :

$$I_{\text{inertie}} = \sum (\lambda_i) = 2.91 + 0.91 = 3.82$$

Donner l'expression de l'inertie d'un point.

L'inertie d'un point c'est la distance du point par rapport au point de reference (0,0).Ce qui represente la somme des coordonnees du point au carée

```
Inertie_individus <- sort(rowSums(res.PCA$ind$coord^2))
```

Quels sont ceux dont l'inertie est la plus faible ?

On calcule les longueurs euclidiennes de chaque point à l'origine pour le nuage des 2 premières composantes principales après ACP, et on trie le tableau ainsi obtenu :

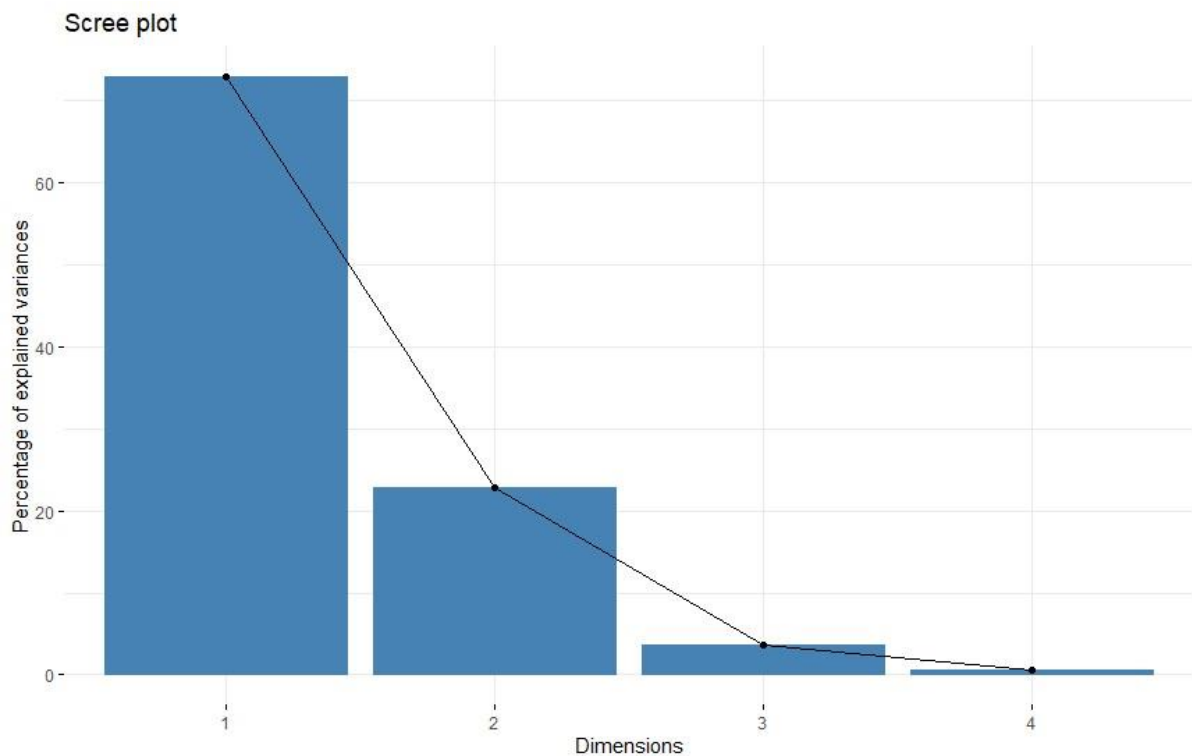
```
longueursApresProjection = (sqrt(res.PCA$ind$coord[,1]^2 + res.PCA$ind$coord[,2]^2))
sort(longueursApresProjection)
```

```
> longueursApresProjection = (sqrt(res.PCA$ind$coord[,1]^2 + res.PCA$ind$coord[,2]^2))
> sort(longueursApresProjection)
96      83      85      67      65      62      97      98      92      72      100      79      75      56      64      74
0.2029220 0.2302699 0.3646643 0.4015153 0.4402685 0.4452157 0.4472342 0.5968327 0.6250669 0.6332501 0.6515628 0.7021417 0.7055167 0.7092696 0.7432691 0.7575541
68      83      71      95      76      139      59      52      86      150      66      128      60      93      80      57
0.8078591 0.8139215 0.8375230 0.9030816 0.9095338 0.9249853 0.9284611 0.9425609 0.9482456 0.9609641 1.0126618 1.0229763 1.0340814 1.0441241 1.0594837 1.0746467
55      122      134      91      87      127      84      77      70      102      143      90      53      78      51      138
1.0954840 1.1328551 1.1520108 1.1542739 1.1714814 1.2169528 1.2359931 1.2588819 1.3130075 1.3524455 1.3524455 1.3592142 1.3855874 1.3982246 1.3995153 1.4110103
124      104      135      117      111      115      148      73      81      82      107      99      112      149      114      116
1.4176213 1.4409170 1.4505773 1.4934701 1.5299955 1.5328361 1.5447841 1.5473502 1.5677374 1.5726505 1.6040553 1.6073891 1.6572297 1.7050460 1.7142752 1.7279042
88      129      54      147      24      63      133      32      105      137      146      58      113      130      21      140
1.7331196 1.7981378 1.8011041 1.8033190 1.8206809 1.8520847 1.8712411 1.8799619 1.8909717 1.9043236 1.9101113 1.9154574 1.9299876 1.9466261 1.9574973 1.9716035
125      44      142      69      101      26      94      27      141      120      109      37      35      29      40      2
1.9800337 2.0204324 2.0229427 2.0328624 2.0396244 2.0496592 2.0514625 2.0653971 2.1062565 2.1236963 2.1290692 2.1494842 2.1595786 2.1623684 2.1863075 2.1874312
46      31      126      50      144      36      25      28      121      10      8      18      145      103      30      1
2.1884783 2.1987915 2.1988653 2.2038527 2.2173114 2.2174085 2.2313886 2.2317288 2.2318689 2.2341135 2.2439700 2.2442357 2.2568445 2.2757542 2.2902996 2.3150171
41      12      13      108      19      4      3      22      11      45      48      49      7      131      5      6
2.3287510 2.3299345 2.3350575 2.3404449 2.3619173 2.3757205 2.3888240 2.3926776 2.4046184 2.4231609 2.4253298 2.4426125 2.4444932 2.4497141 2.4758315 2.5545828
9      39      38      43      20      47      61      17      14      23      106      15      136      123      110      42
2.5873736 2.5924228 2.5958076 2.5981929 2.6006581 2.6348273 2.6563595 2.6598008 2.8031615 2.8119508 2.8659138 2.8799779 2.9269827 2.9270263 2.9659526 2.9859887
33      34      119      132      16      118
3.1709410 3.2572096 3.3107433 3.4943192 3.5119423 3.5247112
```

En rouge les points avec la plus grande inertie, en bleu ce avec la plus petite.

Question 4 : Donner le profil des valeurs propres et proposer le meilleur plan d'observation à retenir pour l'analyse. Justifier le nombre de composantes principales trouvées ?

```
fviz_eig(res.PCA)
```



Le meilleur plan factoriel est donc de projeter sur les Composantes principales 1 et 2 car les 3 et 4 apportent très peu d'informations.

On a un tableau de 150 individus et 4 variables, on part donc d'une dimension 4, après l'ACP on obtient donc 4 Composantes Principales.

Question 5 : Donner l'expression et la valeur de l'inertie de ce nuage de points.

Le nuage utilise les 2 premières composantes principales donc son inertie est celle du premier plan factoriel (3.82 comme calculé à la question 3).

Question 6 : Donner l'expression et la valeur de l'indice ponctuel de qualité pour le point le mieux placé pour ce critère

On a vu à la question 3 qu'il s'agissait du point 118, il suffit donc de comparer son abscisse avant projection dans le nuage centré réduit à son abscisse après projection :

```
coordPoint118AvantProjection = scale(iris[,1:4])[118,]
longueurAvantProjection <- sqrt(sum(coordPoint118AvantProjection^2))
longueurApresProjection <- longueurApresProjection[118]
longueurAvantProjection
longueurApresProjection
longueurApresProjection / longueurAvantProjection
```

```
> coordPoint118AvantProjection = scale(iris[,1:4])[118,]
> coordPoint118AvantProjection
Sepal.Length Sepal.Width Petal.Length Petal.Width
2.242172 1.703886 1.666574 1.312801
> longueurAvantProjection <- sqrt(sum(coordPoint118AvantProjection^2))
> longueurApresProjection <- longueursApresProjection[118]
> longueurAvantProjection
[1] 3.52583
> longueurApresProjection
118
3.524711
> longueurApresProjection / longueurAvantProjection
118
0.9996826
```

La valeur de l'indice ponctuel de qualité du point 118 est donc 0.9996828

Question 7 : Donner l'expression qui mesure la contribution d'un point sur un axe factoriel. Quels sont les individus qui contribuent le plus fortement à la formation du premier axe principal ?

Comme on a fait une ACP centrée réduite, il suffit de regarder la plus grande absces après projection

```
sort(abs(res.PCA$ind$coord[,1]))
```

```
> sort(abs(res.PCA$ind$coord[,1]))
60 82 65 80 89 96 61 81 68 70 85 97 83 100
0.01142619 0.02345269 0.03335470 0.04025861 0.06958832 0.09136066 0.11019628 0.13079518 0.15881005 0.16491790 0.22397877 0.22771687 0.24153827 0.25673059
91 90 95 93 67 94 107 56 54 86 62 99 72 58
0.27907778 0.28347724 0.28858624 0.33653037 0.35025167 0.36218338 0.36701769 0.38868734 0.40748306 0.42913912 0.44069345 0.44766702 0.47628719 0.48732274
63 98 92 74 79 75 64 52 71 57 76 66 139 59
0.56210831 0.57638829 0.62456979 0.63285820 0.66480037 0.70266118 0.71956189 0.73133743 0.73768265 0.74652974 0.87427365 0.87540719 0.92482492 0.92790164
150 122 128 88 87 84 55 51 134 102 143 127 135 69
0.96065603 0.97798073 1.02095055 1.04453138 1.04872805 1.06109461 1.07547470 1.10178118 1.11414774 1.15788161 1.15788161 1.17510363 1.20247330 1.22509363
73 53 77 114 120 124 138 78 111 149 104 115 117 148
1.23417810 1.24097932 1.25650912 1.26011510 1.26376667 1.33323218 1.34629210 1.35840512 1.36417549 1.37278779 1.44015066 1.46764520 1.47143146 1.52117050
147 137 116 112 125 129 24 32 101 140 42 133 130 105
1.56458048 1.57625591 1.59007732 1.60267867 1.70073390 1.78834992 1.81866953 1.83159477 1.84456887 1.85204505 1.85812246 1.86270322 1.86364755 1.86781222
146 113 19 142 21 26 126 44 145 109 141 121 144 37
1.87050329 1.88390070 1.89857200 1.90178409 1.91432300 1.95184633 1.95432671 1.96444768 1.99814710 2.00668647 2.01481043 2.03771630 2.04055823 2.04514621
27 46 6 2 35 45 29 31 11 28 40 10 18 15
2.05115137 2.06974430 2.07563095 2.08096115 2.10997488 2.13705901 2.13956345 2.14012214 2.16631010 2.16857717 2.16971071 2.18432817 2.19034951 2.19874060
50 103 22 17 36 13 25 49 8 110 16 1 30 41
2.20383344 2.20526679 2.20701284 2.20758770 2.20780890 2.21845090 2.22716331 2.22944655 2.23284716 2.25977735 2.26221453 2.26470281 2.26526149 2.28647514
4 108 132 12 9 20 3 47 5 48 118 39 131 7
2.29938422 2.30243944 2.30492772 2.32613087 2.33464048 2.34336905 2.36422905 2.38473317 2.38984217 2.39437631 2.42632899 2.42963258 2.43595373 2.44402884
34 38 43 33 14 106 23 136 123 119
2.44617739 2.52733191 2.55363840 2.61494794 2.63310070 2.75187334 2.77434470 2.79877045 2.89765149 3.31069558
```

Le point 119 est le point qui contribue le plus avec une contribution de 3.31069558, puis le point 123, puis 136, puis 23, ...

Question 8 : Donner deux individus qui sont bien représentés par leur première composante principale et deux les plus mal représentés par cette composante.

On se réfère aux indices de qualités des absces sur l'axes 1 :

```
sort(res.PCA$ind$cos2[,1])
```

60	82	80	61	65	81	70	89	94
9.595978e-05	2.171750e-04	1.318942e-03	1.720338e-03	4.564686e-03	6.897068e-03	1.548382e-02	3.032379e-02	3.108397e-02
68	107	90	54	91	58	99	96	85
3.208691e-02	3.778803e-02	4.330787e-02	5.110830e-02	5.506021e-02	6.362155e-02	7.534247e-02	7.540386e-02	7.832704e-02
63	83	95	93	100	86	97	56	67
7.862358e-02	8.505812e-02	9.864922e-02	1.002241e-01	1.510282e-01	1.652746e-01	2.142720e-01	2.620665e-01	2.795423e-01
88	69	120	42	16	132	72	118	57
3.120681e-01	3.398538e-01	3.482637e-01	3.818483e-01	4.148109e-01	4.250633e-01	4.685974e-01	4.704046e-01	4.710788e-01
114	122	149	51	71	74	137	34	15
4.833376e-01	4.855240e-01	4.986771e-01	4.998618e-01	4.999383e-01	5.454390e-01	5.494001e-01	5.635205e-01	5.650488e-01
110	66	52	73	135	115	19	102	143
5.695564e-01	5.940507e-01	5.960682e-01	6.020253e-01	6.098699e-01	6.273744e-01	6.298932e-01	6.348502e-01	6.348502e-01
101	6	33	17	53	59	75	84	87
6.576882e-01	6.600989e-01	6.763790e-01	6.854138e-01	6.915685e-01	7.066943e-01	7.067637e-01	7.159556e-01	7.186556e-01
76	125	145	116	147	77	150	126	111
7.204579e-01	7.207085e-01	7.208004e-01	7.371151e-01	7.415355e-01	7.464447e-01	7.508462e-01	7.549599e-01	7.647242e-01
45	98	130	11	62	20	9	47	121
7.667622e-01	7.718989e-01	7.993811e-01	8.015666e-01	8.029139e-01	8.096466e-01	8.115435e-01	8.160634e-01	8.198686e-01
144	49	139	142	55	22	37	109	141
8.273394e-01	8.284719e-01	8.308982e-01	8.353172e-01	8.444751e-01	8.465230e-01	8.547102e-01	8.575580e-01	8.672096e-01
39	140	136	138	124	14	79	106	64
8.734875e-01	8.737961e-01	8.752553e-01	8.755744e-01	8.759355e-01	8.786492e-01	8.809180e-01	8.843819e-01	8.860991e-01
26	134	108	46	2	13	128	123	146
8.867166e-01	8.895068e-01	8.896069e-01	8.900194e-01	8.927725e-01	8.938769e-01	8.980987e-01	9.010088e-01	9.050786e-01
131	78	127	21	32	44	103	5	112
9.085510e-01	9.101054e-01	9.123146e-01	9.140097e-01	9.153582e-01	9.167790e-01	9.309818e-01	9.315095e-01	9.339097e-01
4	105	28	10	92	38	31	113	35
9.346682e-01	9.346725e-01	9.361470e-01	9.434549e-01	9.446604e-01	9.452702e-01	9.457919e-01	9.471838e-01	9.485212e-01
18	148	43	1	133	119	117	23	104
9.505502e-01	9.510633e-01	9.521845e-01	9.539975e-01	9.539992e-01	9.566234e-01	9.597199e-01	9.599880e-01	9.608218e-01
41	29	129	48	30	36	40	3	25
9.617761e-01	9.624243e-01	9.669461e-01	9.710330e-01	9.752045e-01	9.756149e-01	9.785321e-01	9.790410e-01	9.791281e-01
7	27	8	24	12	50			
9.809239e-01	9.834773e-01	9.884478e-01	9.906097e-01	9.918759e-01	9.946874e-01			

Les deux mieux représentés sont les points 12 et 50 et les deux moins bien représentés sont les points 60 et 82.

Question 9 : Analyser de la même façon le deuxième puis le troisième axe principal (reprendre les questions 6,7 et 8)

2eme axe principal :

`sort(abs(res.PCAindcoord[,2]))`

> sort(abs(res_PCA\$indscord[,2]))												
50	139	119	150	92	59	104	7	62	75	128	77	24
0.009216358	0.017223100	0.017780949	0.024331668	0.024923029	0.032226078	0.046987588	0.047644198	0.063295188	0.063411820	0.064346029	0.077256020	0.085558526
12	25	98	133	96	64	129	67	36	55	89	8	79
0.133078335	0.137254455	0.154873597	0.178549495	0.181192126	0.186224606	0.187361215	0.196311735	0.206107398	0.208421046	0.219503335	0.223148073	0.225927855
27	76	117	131	40	148	85	134	105	29	127	78	30
0.242163553	0.250793393	0.255621824	0.259284433	0.268878961	0.269069144	0.287773512	0.292922623	0.295044824	0.313217810	0.316394472	0.331311682	0.337731904
3	97	48	146	71	21	123	74	72	113	108	112	138
0.341908024	0.384920081	0.386246873	0.386966082	0.396571562	0.408855708	0.413641060	0.416387721	0.417320281	0.419249651	0.420065580	0.421700450	0.422430611
32	65	41	115	23	35	10	44	43	1	124	18	31
0.423695068	0.439003210	0.441715388	0.442271587	0.458343668	0.460201841	0.469013561	0.472326668	0.479100690	0.480026597	0.481811219	0.488838316	0.504540690
66	87	28	103	130	122	38	56	52	4	100	141	53
0.509063957	0.522051797	0.527149525	0.562010477	0.562290726	0.571764325	0.592292774	0.593283636	0.594614726	0.597394508	0.598851796	0.613885637	0.616297654
26	84	5	37	2	140	116	142	111	102	143	46	109
0.625618588	0.633843245	0.646835383	0.661558111	0.674133557	0.676128174	0.676244806	0.689575494	0.692756454	0.698869862	0.698869862	0.711052725	0.711438654
13	57	83	68	106	135	86	95	136	51	144	101	147
0.728676165	0.773019312	0.777256383	0.792095742	0.800409201	0.811315271	0.845582241	0.855730320	0.856803329	0.862972418	0.867520601	0.870421312	0.896686809
39	121	22	73	14	93	49	126	149	125	60	11	145
0.904180040	0.910467410	0.924121427	0.933325729	0.961506729	0.988404018	0.997959764	1.007777596	1.011254419	1.013921867	1.034018275	1.043690653	1.049168747
80	137	9	91	47	20	45	114	70	90	88	19	17
1.058718547	1.068581107	1.115327675	1.120028524	1.120429702	1.127849382	1.142229262	1.162260421	1.302609230	1.329324639	1.382988719	1.405018794	1.483609363
6	99	107	81	82	69	120	54	63	33	58	15	110
1.489177523	1.543792034	1.561502891	1.562271834	1.572475594	1.622243803	1.706745380	1.754403989	1.764724381	1.793575856	1.852429087	1.860057113	1.921010376
94	34	42	118	132	61	16						
2.019237873	2.150727877	2.337415158	2.556661251	2.626323468	2.654072819	2.686284485						

Le point 16 est le point qui contribue le plus avec une contribution de 2.686284485 , puis le point 61, puis 132, puis 118, ...

On se réfère aux indices de qualités des absces sur l'axes 1 :

sort(res.PCA\$ind\$cos2[,2])

	50	119	139	7	150	59	104	92	24
1.739593e-05	2.759385e-05	2.881718e-04	3.727715e-04	4.816803e-04	8.523983e-04	1.022806e-03	1.504234e-03	2.192409e-03	
77	12	128	25	75	36	133	8	131	
2.821831e-03	3.246413e-03	3.567452e-03	3.718675e-03	5.756026e-03	8.502422e-03	8.765525e-03	9.872384e-03	1.029352e-02	
129	27	40	62	123	3	29	30	105	
1.061345e-02	1.370839e-02	1.502744e-02	1.656293e-02	1.836048e-02	2.047578e-02	2.062572e-02	2.167720e-02	2.332217e-02	
48	23	117	108	148	55	43	41	146	
2.526848e-02	2.620158e-02	2.896412e-02	2.961119e-02	2.975647e-02	3.171536e-02	3.351629e-02	3.589435e-02	3.873603e-02	
21	1	10	35	113	18	32	38	31	
4.169275e-02	4.286032e-02	4.349669e-02	4.512209e-02	4.690981e-02	4.734550e-02	4.898229e-02	5.191636e-02	5.256672e-02	
44	78	28	98	115	76	64	103	134	
5.299912e-02	5.413854e-02	5.531742e-02	5.572946e-02	5.697216e-02	5.928516e-02	5.934992e-02	6.046551e-02	6.148505e-02	
4	112	127	5	130	106	141	136	138	
6.308947e-02	6.465768e-02	6.613785e-02	6.823959e-02	7.276930e-02	7.481817e-02	8.050645e-02	8.202807e-02	8.620357e-02	
67	37	26	2	13	79	46	109	142	
8.781721e-02	8.943472e-02	9.109885e-02	9.369248e-02	9.643777e-02	1.017402e-01	1.050436e-01	1.077902e-01	1.098228e-01	
124	140	14	39	85	116	71	101	22	
1.143969e-01	1.164569e-01	1.171616e-01	1.209721e-01	1.293003e-01	1.333235e-01	1.444842e-01	1.464500e-01	1.484182e-01	
144	121	122	49	53	87	47	9	11	
1.495359e-01	1.636762e-01	1.659527e-01	1.660006e-01	1.705639e-01	1.780829e-01	1.801414e-01	1.852153e-01	1.860553e-01	
20	111	145	126	66	45	102	143	74	
1.875497e-01	1.972084e-01	1.987242e-01	2.007516e-01	2.008852e-01	2.190454e-01	2.312787e-01	2.312787e-01	2.361181e-01	
147	137	84	125	149	135	96	89	51	
2.435666e-01	2.524935e-01	2.554710e-01	2.561509e-01	2.706036e-01	2.776295e-01	2.965881e-01	3.017123e-01	3.066571e-01	
17	33	6	73	19	72	52	15	114	
3.095680e-01	3.182025e-01	3.397830e-01	3.442909e-01	3.449666e-01	3.597502e-01	3.940322e-01	4.043808e-01	4.111848e-01	
110	34	57	118	88	132	16	69	42	
4.115901e-01	4.356170e-01	5.051030e-01	5.222982e-01	5.470709e-01	5.518683e-01	5.849066e-01	5.959170e-01	6.042461e-01	
56	97	120	86	107	63	60	65	68	
6.105701e-01	6.122323e-01	6.352021e-01	6.416844e-01	6.840153e-01	7.749373e-01	7.858559e-01	7.907372e-01	7.982287e-01	
100	93	95	83	91	99	80	58	54	
8.217536e-01	8.645554e-01	8.673937e-01	8.807901e-01	8.868405e-01	8.959985e-01	9.121560e-01	9.192932e-01	9.473960e-01	
90	70	94	82	81	61				
9.523411e-01	9.659862e-01	9.661717e-01	9.763206e-01	9.839981e-01	9.979427e-01				

Les deux mieux représentés sont les points 81 et 61 et les deux moins bien représentés sont les points 50 et 119.

3eme axe principal :

sort(abs(res.PCA\$ind\$coord[,3]))

	17	124	140	5	38	92	54	112	113	147	6	16	24
0.005344094	0.005428364	0.012672115	0.015738196	0.019435812	0.020481147	0.023101768	0.023186408	0.026338410	0.026371352	0.026968294	0.030526609	0.034488596	
41	117	3	18	61	33	47	46	30	64	31	34	79	
0.034894909	0.037431260	0.044201485	0.044215316	0.046790444	0.047228419	0.057217858	0.063929826	0.068435776	0.068658945	0.075008442	0.082668045	0.085863889	
27	8	90	4	100	12	52	91	84	94	25	56	1	
0.086364011	0.088695498	0.089410023	0.091290106	0.091879161	0.093759244	0.094121716	0.094487601	0.105311387	0.105821048	0.117993536	0.124191550	0.127706022	
118	96	127	142	95	20	48	9	57	81	83	50	97	
0.127881459	0.128978343	0.129937757	0.130075005	0.130889685	0.132630467	0.139467905	0.145076864	0.148969403	0.149983478	0.151211957	0.153029490	0.156213154	
22	104	135	35	70	40	148	14	138	49	134	99	39	
0.159865277	0.163630107	0.164723757	0.170274861	0.172837808	0.175883821	0.180178380	0.180796084	0.180875478	0.181492780	0.183488392	0.190400930	0.193254662	
65	93	103	42	62	28	36	13	2	121	82	45	58	
0.194932893	0.199389755	0.202914170	0.204234223	0.205073815	0.206816248	0.225441580	0.230911237	0.234608854	0.234799484	0.241552281	0.248433561	0.249265266	
10	146	78	72	120	11	32	129	29	98	111	89	74	
0.253765567	0.256273852	0.260826577	0.264952227	0.267536893	0.268681102	0.270467377	0.270658006	0.271150240	0.271650362	0.284612074	0.291579274	0.291896252	
125	68	26	43	44	80	23	7	128	144	133	73	19	
0.298450613	0.302037174	0.305640982	0.305766453	0.309601318	0.319573330	0.332179098	0.335470401	0.337715967	0.338144000	0.354148712	0.368412272	0.374343275	
109	105	87	110	55	139	126	21	141	75	86	76	15	
0.393990571	0.395628375	0.395786384	0.397551897	0.398255523	0.416826193	0.419984722	0.421292594	0.428332842	0.446027008	0.450634071	0.472578954	0.472900998	
69	37	67	132	66	150	102	143	60	136	53	114	106	
0.482304024	0.484537410	0.490873075	0.493473808	0.503511382	0.528248807	0.530160149	0.530160149	0.538899390	0.542906499	0.554006835	0.580249290	0.582309103	
59	71	145	116	108	85	51	88	119	130	77	131	63	
0.596169361	0.616526306	0.632413436	0.638428708	0.651706439	0.665249720	0.684586163	0.688295960	0.703305304	0.715634119	0.727155002	0.727816146	0.765771394	
122	123	149	137	107	115	101							
0.828127201	0.857421825	0.933395241	0.945853819	0.986893267	1.003869574	1.005401018							

Le point 101 est le point qui contribue le plus avec une contribution de 1.005401018, puis le point 115, puis 107, puis 137, ...

sort(res.PCA\$ind\$cos2[,3])

17	124	5	140	38	16	6	54	113
4.016660e-06	1.452106e-05	4.039790e-05	4.090771e-05	5.590322e-05	7.553351e-05	1.114335e-04	1.642717e-04	1.851388e-04
112	147	33	41	61	3	24	18	47
1.954695e-04	2.106691e-04	2.206332e-04	2.240088e-04	3.101660e-04	3.422122e-04	3.562427e-04	3.873412e-04	4.697940e-04
117	34	46	30	92	31	118	4	8
6.210597e-04	6.435898e-04	8.491279e-04	8.900726e-04	1.015833e-03	1.161820e-03	1.306734e-03	1.473268e-03	1.559692e-03
12	27	20	94	25	1	9	48	142
1.611452e-03	1.743550e-03	2.593586e-03	2.653527e-03	2.748222e-03	3.033525e-03	3.133777e-03	3.294562e-03	3.907659e-03
14	90	22	42	50	49	39	35	91
4.142463e-03	4.308269e-03	4.441578e-03	4.613167e-03	4.796002e-03	5.490379e-03	5.526313e-03	6.177227e-03	6.311564e-03
40	84	103	64	28	81	13	52	36
6.430181e-03	7.052264e-03	7.882131e-03	8.067524e-03	8.514587e-03	9.069177e-03	9.684299e-03	9.872793e-03	1.017240e-02
45	121	127	2	135	11	104	10	148
1.036211e-02	1.088557e-02	1.115483e-02	1.134754e-02	1.144455e-02	1.233027e-02	1.240376e-02	1.273359e-02	1.334317e-02
99	43	23	79	29	120	138	58	146
1.362912e-02	1.365155e-02	1.376225e-02	1.469515e-02	1.545739e-02	1.560780e-02	1.580425e-02	1.664541e-02	1.698940e-02
70	110	7	57	100	132	32	95	26
1.700670e-02	1.762760e-02	1.848124e-02	1.875830e-02	1.934358e-02	1.948351e-02	1.996007e-02	2.029333e-02	2.174287e-02
129	125	144	44	82	134	19	15	56
2.214823e-02	2.219381e-02	2.271900e-02	2.277138e-02	2.303811e-02	2.412575e-02	2.448797e-02	2.613836e-02	2.675438e-02
136	109	111	83	78	133	126	93	141
3.293444e-02	3.305796e-02	3.328672e-02	3.33620e-02	3.355342e-02	3.448514e-02	3.486555e-02	3.518273e-02	3.919390e-02
106	105	119	21	37	69	73	108	145
3.959954e-02	4.193416e-02	4.317082e-02	4.426781e-02	4.797608e-02	5.267388e-02	5.364475e-02	7.127311e-02	7.220417e-02
123	131	80	128	97	87	114	55	74
7.889070e-02	8.110630e-02	8.310921e-02	9.826930e-02	1.008348e-01	1.023568e-01	1.024848e-01	1.158004e-01	1.160354e-01
68	130	116	102	143	88	53	72	63
1.160627e-01	1.178714e-01	1.188293e-01	1.330936e-01	1.330936e-01	1.355055e-01	1.378277e-01	1.450098e-01	1.459187e-01
96	65	139	98	62	86	51	101	66
1.502826e-01	1.559075e-01	1.687869e-01	1.714552e-01	1.738666e-01	1.822459e-01	1.929814e-01	1.953930e-01	1.965268e-01
137	76	60	150	149	77	107	75	59
1.978260e-01	2.105051e-01	2.134528e-01	2.270347e-01	2.305387e-01	2.499887e-01	2.732251e-01	2.847768e-01	2.917205e-01
115	122	71	89	67	85			
2.935210e-01	3.481322e-01	3.492051e-01	5.323828e-01	5.490673e-01	6.909830e-01			

Les deux mieux représentés sont les points 67 et 85 et les deux moins bien représentés sont les points 17 et 124.

Question 10 : Quelles sont les variables les plus fortement corrélées avec la première composante principale. Interpréter cet axe à l'aide de ces variables.

Comme on a fait une ACP centrée réduite, un point avec une forte contribution sur un axe a forcément une faible contribution sur les autres axes. Il faut donc regarder les plus grandes absises positives et les plus grandes absises négatives.

```
sort(res.PCA$ind$coord[,1])
```

23	14	33	43	38	34	7	39	48	5	47	3	20
-2.77434470	-2.63310070	-2.61494794	-2.55363840	-2.52733191	-2.44617739	-2.44402884	-2.42963258	-2.39437631	-2.38984217	-2.38473317	-2.36422905	-2.34336905
9	12	4	41	30	1	16	8	49	25	13	36	17
-2.33464048	-2.32613087	-2.29938422	-2.28647514	-2.26526149	-2.26470281	-2.26221453	-2.23284716	-2.22944655	-2.22716331	-2.21845090	-2.20780890	-2.20758770
22	50	15	18	10	40	28	11	31	29	45	35	2
-2.20701284	-2.20383344	-2.19874060	-2.19034951	-2.18432817	-2.16971071	-2.16857717	-2.16631010	-2.14012214	-2.13956345	-2.13705901	-2.10997488	-2.08096115
6	46	27	37	44	26	21	19	42	32	24	58	99
-2.07563095	-2.06974430	-2.05115137	-2.04514621	-1.96444768	-1.95184633	-1.91432300	-1.89857200	-1.85812246	-1.83159477	-1.81866953	-0.48732274	-0.44766702
94	61	80	65	60	82	89	96	81	68	70	85	97
-0.36218338	-0.11019628	-0.04025861	-0.03335470	0.01142619	0.02345269	0.06958832	0.09136066	0.13079518	0.15881005	0.16491790	0.22397877	0.22771687
83	100	91	90	95	93	67	107	56	54	86	62	72
0.24153827	0.25673059	0.27907778	0.28347724	0.28858624	0.33653037	0.35025167	0.36701769	0.38868734	0.40748306	0.42913912	0.44069345	0.47628719
63	98	92	74	79	75	64	52	71	57	76	66	139
0.56210831	0.57638829	0.62456979	0.63285820	0.66480037	0.70266118	0.71956189	0.73133743	0.73768265	0.74652974	0.87427365	0.87540719	0.92482492
59	150	122	128	88	87	84	55	51	134	102	143	127
0.92790164	0.96065603	0.97798073	1.02095055	1.04453138	1.04872805	1.06109461	1.07547470	1.10178118	1.11414774	1.15788161	1.15788161	1.17510363
135	69	73	53	77	114	120	124	138	78	111	149	104
1.20247330	1.22509363	1.23417810	1.24097932	1.25650912	1.26011510	1.26376667	1.33323218	1.34629210	1.35840512	1.36417549	1.37278779	1.44015066
115	117	148	147	137	116	112	125	129	101	140	133	130
1.46764520	1.47143146	1.52117050	1.56458048	1.57625591	1.59007732	1.60267867	1.70073390	1.78834992	1.84456887	1.85204505	1.86270322	1.86364755
105	146	113	142	126	145	109	141	121	144	103	110	108
1.86781222	1.87050329	1.88390070	1.90178409	1.95432671	1.99814710	2.00668647	2.01481043	2.03771630	2.04055823	2.20526679	2.25977735	2.30243944
132	118	131	106	136	123	119						
2.30492772	2.42632899	2.43595373	2.75187334	2.79877045	2.89765149	3.31069558						

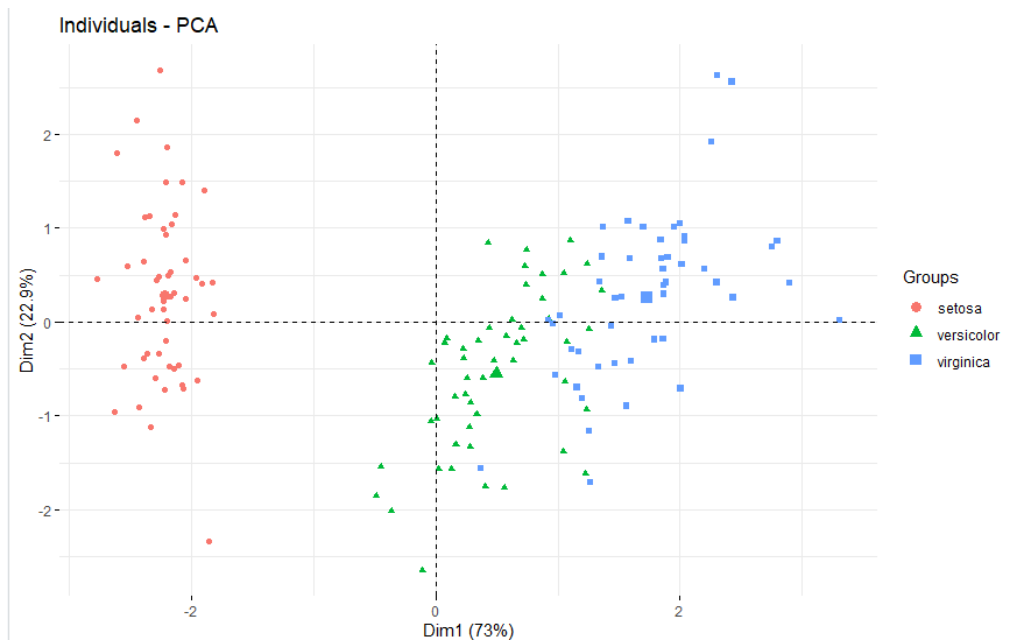
Il oppose les individus 119, 123, 136 aux individus 23, 14 et 33

Cet axe semble opposer les verginica (101 à 150) aux setosas (1 à 50)

Aussi, à partir du tableaux individus/variables, on peut voir que cet axe oppose les faibles longueurs de Sepals, largeurs de Pétales et longueurs de Pétales aux fortes longueurs de Sepals, largeurs de Pétales et longueurs.

	Sepal.Len	Sepal.Wid	Petal.Len	Petal.Wid	Species
119	7,7	2,6	6,9	2,3	virginica
123	7,7	2,8	6,7	2	virginica
136	7,7	3	6,1	2,3	virginica
14	4,3	3	1,1	0,1	setosa
23	4,6	3,6	1	0,2	setosa
33	5,2	4,1	1,5	0,1	setosa

De plus, au vu du nuage de point, on peut observer que cet axe semble être discriminant car les versicolores se trouvent proche de l'origine :



Question 11 : Faisant suite à la question précédente, donner une interprétation concernant le deuxième et troisième axe factoriel.

2eme axe principal :

sort(res.PCA\$ind\$coord[,2])

```

61 42 94 58 63 54 120 69 82 81 107 99
-2.654072819 -2.337415158 -2.019237873 -1.852429087 -1.764724381 -1.754403989 -1.706745380 -1.622243803 -1.572475594 -1.562271834 -1.561502891 -1.543792034
88 90 70 114 91 9 80 60 93 14 73 39
-1.382988719 -1.329324639 -1.302609230 -1.162260421 -1.120028524 -1.115327675 -1.058718547 -1.034018275 -0.988404018 -0.961506729 -0.933325729 -0.904180040
147 95 135 68 83 13 109 46 102 143 2 84
-0.896686809 -0.855730320 -0.811315271 -0.792095742 -0.777256383 -0.728676165 -0.711438654 -0.711052725 -0.698869862 -0.698869862 -0.674133557 -0.633843245
26 100 4 56 122 31 124 43 10 35 115 65
-0.625618588 -0.598851796 -0.597394508 -0.593283636 -0.571764325 -0.504540690 -0.481811219 -0.479100690 -0.469013561 -0.460201841 -0.442271587 -0.439003210
112 72 74 48 97 3 30 127 134 85 79 89
-0.421700450 -0.417320281 -0.416387721 -0.386246873 -0.384920081 -0.341908024 -0.337731904 -0.316394472 -0.292922623 -0.287773512 -0.225927855 -0.219503335
55 36 67 129 64 96 133 98 77 75 62 104
-0.208421046 -0.206107398 -0.196311735 -0.187361215 -0.186224606 -0.181192126 -0.178549495 -0.154873597 -0.077256020 -0.063411820 -0.063295188 -0.046987588
150 50 139 119 92 59 7 128 24 12 25 8
-0.024331668 0.009216358 0.017223100 0.017780949 0.024923029 0.032226078 0.047644198 0.064346029 0.085558526 0.133078335 0.137254455 0.223148073
27 76 117 131 40 148 105 29 78 146 71 21
0.242163553 0.250793393 0.255621824 0.259284433 0.268878961 0.269069144 0.295044824 0.313217810 0.331311682 0.386966082 0.396571562 0.408855708
123 113 108 138 32 41 23 44 1 18 66 87
0.413641060 0.419249651 0.420065580 0.422430611 0.423695068 0.441715388 0.458343668 0.472326668 0.480026597 0.488838316 0.509063957 0.522051797
28 103 130 38 52 141 5 37 140 116 142
0.527149525 0.562010477 0.562290726 0.592292774 0.594614726 0.613885637 0.616297654 0.646835383 0.661558111 0.676128174 0.676244806 0.689575494
111 57 106 86 136 51 144 101 121 22 49 126
0.692756454 0.773019312 0.800409201 0.845582241 0.856803329 0.862972418 0.867520601 0.870421312 0.910467410 0.924121427 0.997959764 1.007777596
149 125 11 145 137 47 20 45 19 17 6 33
1.011254419 1.013921867 1.043690653 1.049168747 1.068581107 1.120429702 1.127849382 1.142229262 1.405018794 1.483609363 1.489177523 1.793575856
15 110 34 118 132 16
1.860057113 1.921010376 2.150727877 2.556661251 2.626323468 2.686284485

```

Il oppose les individus 16, 132, 118 et 34 aux individus 61, 42 et 94. Cela ne semble pas opposer des classes, on va donc regarder les coordonnées de cet individu dans le nuage initial.

	Sepal. Len	Sepal. Wic	Petal. Leng	Petal. Widi	Species
16	5,7	4,4	1,5	0,4	setosa
132	7,9	3,8	6,4	2	virginica
118	7,7	3,8	6,7	2,2	virginica
34	5,5	4,2	1,4	0,2	setosa
61	5	2	3,5	1	versicolor
42	4,5	2,3	1,3	0,3	setosa
94	5	2,3	3,3	1	versicolor

La différence semble être au niveau de la largeur des Sepal. En triant les largeurs de Sepal on voit que les largeurs minimales sont bien entre 2 et 2,3 et les max entre 4.4 et 3.8.

```

[1] 2.0 2.2 2.2 2.2 2.3 2.3 2.3 2.3 2.4 2.4 2.4 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.6 2.6 2.6 2.6 2.6 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.8 2.8 2.8 2.8 2.8 2.8
[40] 2.8 2.8 2.8 2.8 2.8 2.8 2.8 2.8 2.9 2.9 2.9 2.9 2.9 2.9 2.9 2.9 2.9 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0
[79] 3.0 3.0 3.0 3.0 3.0 3.1 3.1 3.1 3.1 3.1 3.1 3.1 3.1 3.1 3.1 3.1 3.1 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2 3.2
[118] 3.4 3.4 3.4 3.4 3.4 3.4 3.4 3.4 3.5 3.5 3.5 3.5 3.5 3.5 3.5 3.6 3.6 3.6 3.6 3.7 3.7 3.7 3.8 3.8 3.8 3.8 3.8 3.8 3.8 3.8 3.9 3.9 4.0 4.1 4.2 4.4

```

3eme axe principal :

```
sort(res.PCA$ind$coord[,3])
```

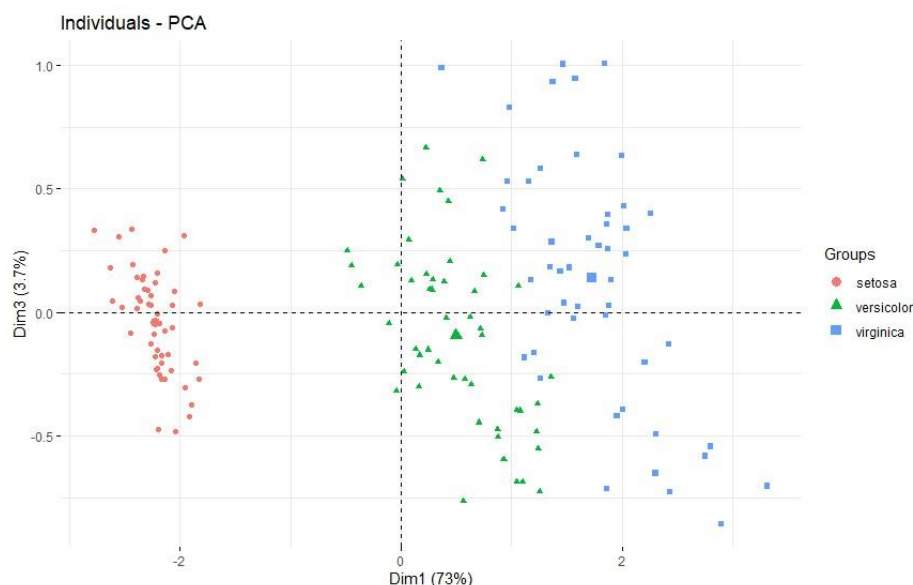
```
> sort(res.PCA$ind$coord[,3])
123      63      131      77      130      119      88      51      108      59      106      53
-0.857421825 -0.765771394 -0.727816146 -0.727155002 -0.715634119 -0.703305304 -0.688295960 -0.684586163 -0.651706439 -0.596169361 -0.582309103 -0.554006835
136      66      132      37      69      15      76      75      21      126      55      87
-0.542906499 -0.503511382 -0.493473808 -0.484537410 -0.482304024 -0.472900998 -0.472578954 -0.446027008 -0.421292594 -0.419984722 -0.398255523 -0.395786384
109      19      73      80      26      68      74      98      29      32      11      120
-0.393990571 -0.374343275 -0.368412272 -0.319573330 -0.305640982 -0.302037174 -0.291896252 -0.271650362 -0.271150240 -0.270467377 -0.268681102 -0.267536893
72      78      10      82      2      13      36      28      42      103      93      134
-0.264952227 -0.260826577 -0.253765567 -0.241552281 -0.234608854 -0.230911237 -0.225441580 -0.206816248 -0.204234223 -0.202914170 -0.199389755 -0.183488392
49      40      70      35      135      50      83      81      118      1      52      8
-0.181492780 -0.175883821 -0.172837808 -0.170274861 -0.164723757 -0.153029490 -0.151211957 -0.149983478 -0.127881459 -0.127706022 -0.094121716 -0.088695498
34      31      64      46      61      18      147      54      92      140      124      17
-0.082668045 -0.075008442 -0.068658945 -0.063929826 -0.046790444 -0.044215316 -0.026371352 -0.023101768 -0.020481147 -0.012672115 -0.005428364 -0.005344094
5      38      112      113      6      16      24      41      117      3      33      47
0.015738196 0.019435812 0.023186408 0.026338410 0.026968294 0.030526609 0.034488596 0.034894909 0.037431260 0.044201485 0.047228419 0.057217858
30      79      27      90      4      100      12      91      84      94      25      56
0.068435776 0.085863889 0.086364011 0.089410023 0.091290106 0.091879161 0.093759244 0.094487601 0.105311387 0.105821048 0.117993536 0.124191550
96      127      142      95      20      48      9      57      97      22      104      148
0.128978343 0.129937757 0.130075005 0.130889685 0.132630467 0.139467905 0.145076864 0.148969403 0.156213154 0.159865277 0.163630107 0.180178380
14      138      99      39      65      62      121      45      58      146      129      111
0.180796084 0.180875478 0.190400930 0.193254662 0.194932893 0.205073135 0.234799484 0.248433561 0.249265266 0.256273852 0.270658006 0.284612074
89      125      43      44      23      7      128      144      133      105      110      139
0.291579274 0.298450613 0.305766453 0.309601318 0.332179098 0.335470401 0.337715967 0.338144000 0.354148712 0.395628375 0.397551897 0.416826193
141      86      67      150      102      143      60      114      71      145      116      85
0.428332842 0.450634071 0.490873075 0.528248807 0.530160149 0.530160149 0.538899390 0.580249290 0.616526306 0.632413436 0.638428708 0.665249720
122      149      137      107      115      101
0.828127201 0.933395241 0.945853819 0.986893267 1.003869574 1.005401018
```

Cet axe oppose donc les individus 107, 115 et 101 aux individus 63, 131 et 77. On regarde sur Excel comme précédemment :

	Sepal.Len	Sepal.Wic	Petal.Len	Petal.Wid	Species
107	4,9	2,5	4,5	1,7	virginica
101	6,3	3,3	6	2,5	virginica
115	5,8	2,8	5,1	2,4	virginica
63	6	2,2	4	1	versicolor
131	7,4	2,8	6,1	1,9	virginica
77	6,8	2,8	4,8	1,4	versicolor

Aucune observation ne peut être faites, nous allons donc essayer de représenter le nuage :

```
fviz_pca_ind(res.PCA, axes = c(1, 3), label="none", habillage=iris$Species)
```



L'axe ne semble définitivement n'avoir aucune signification particulière

Question 12 : Proposez une/des interprétation(s) possible(s) suite à l'étude des graphiques obtenus dans R.

Les setosas sont très différents des versicolors et des virginica, du fait de la faible longueur de leurs sépals, de leurs pétales et de la faible largeur de leurs pétales.

Les virginica et les versicolors sont plus difficiles à différencier, bien que les versicolors semblent avoir de des sépales de plus petites longueurs et des pétales de plus petite largeur et longueur que ceux des virginica, bien que la différence soit bien plus faible.

Le critère de la largeur des sépales ne semble quant à lui pas être utile pour différencier ces différentes variétés d'Iris