

Analyse de données - Rapport

Kilian Blain

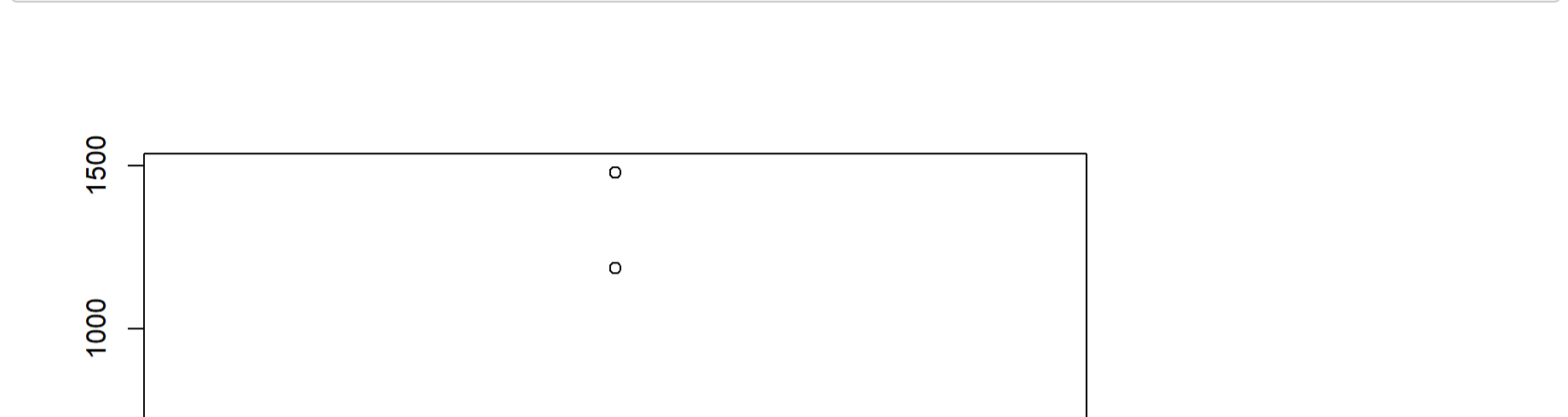
2022-12-14

Analyse en composantes principales

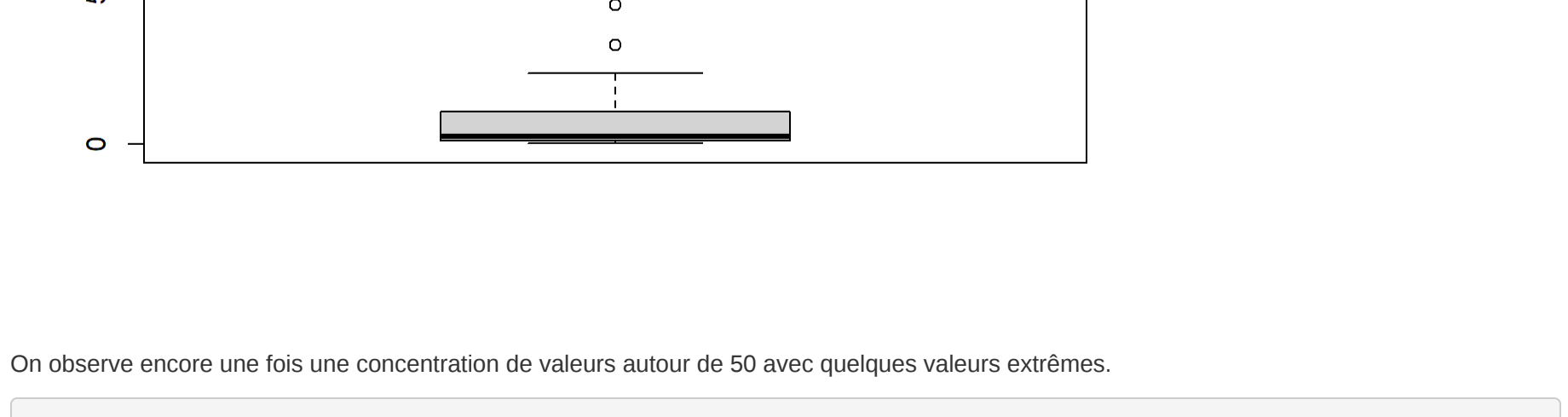
1.

d <- read.table("Eaux.data", header = TRUE)									
#		Nom	Pays	Bicarbonates	Sulfates	Chlorures	Calcium	Magnesium	
# 1	Aix-les-Bains	F-73	841	27	3	84	23		
# 2	Arvie	F-63	2195	31	387	178	92		
# 3	Badoit	F-42	1360	48	40	198	85		
# 4	Beckersheim	F-68	263	23	9	81	5		
# 5	Cayenne	F-84	287	3	5	44	24		
# 6	Chambon	F-45	268	0	23	96	6		
# 7	Contrex	F-88	463	1187	9	486	84		
# 8	Cristal-Roc	F-72	260	15	8	78	2		
# 9	Saint-Gyr	F-45	250	15	20	71	2		
# 10	Evian	F-74	357	18	4	78	24		
# 11	Ferita	Italie	311	14	18	73	18		
# 12	Hepar	F-88	463	1479	10	555	110		
# 13	Saint-Hippolite	F-37	256	6	23	86	3		
# 14	Laurier	F-77	186	18	16	64	4		
# 15	Ogeu	F-64	183	16	44	48	11		
# 16	Ordine	F-95	398	218	15	157	35		
# 17	Planchet	F-22	160	150	38	141	47		
# 18	Perrier	F-36	448	91	31	148	9		
# 19	Ribes	Espagne	168	24	8	55	5		
# 20	La Salvetat	F-24	327	5	4	83	1		
# 21	Sampellegrino	Italie	245	438	52	174	51		
# 22	Spa	Belgique	110	65	5	4	1		
# 23	Sulfates	F-74	332	2	8	163	14		
# 24	Valvert	F-88	204	18	4	68	2		
# 25	Verl	Espagne	196	18	6	58	6		
# 26	Verniere	F-64	1170	158	18	199	72		
# 27	Vichy-Celestin	F-63	2869	138	235	103	10		
# 28	Vichy-St-Yorre	F-63	4368	174	322	98	11		
# 29	Villedieu	Espagne	19	7	6	16	2		
# 30	Vittel	F-88	402	306	15	282	36		
# 31	Vittel	F-88	65	6	8	18	6		
#	Sodium								
# 1	2								
# 2	650								
# 3	150								
# 4	3								
# 5	23								
# 6	11								
# 7	9								
# 8	4								
# 9	11								
# 10	5								
# 11	13								
# 12	14								
# 13	18								
# 14	9								
# 15	21								
# 16	3								
# 17	17								
# 18	14								
# 19	9								
# 20	7								
# 21	35								
# 22	3								
# 23	5								
# 24	2								
# 25	13								
# 26	154								
# 27	1172								
# 28	1788								
# 29	9								
# 30	3								
# 31	9								

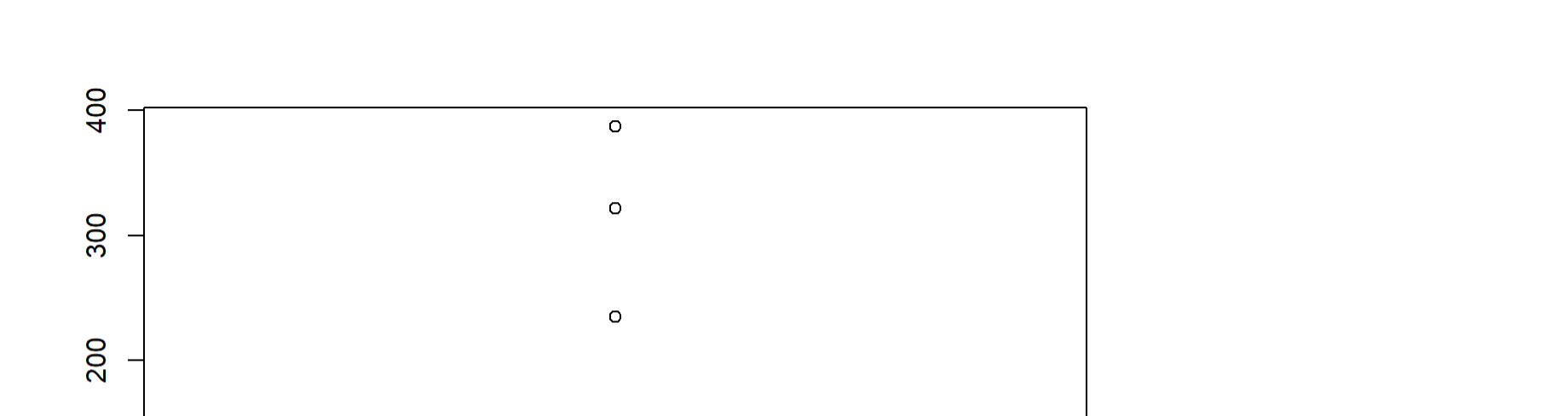
2.



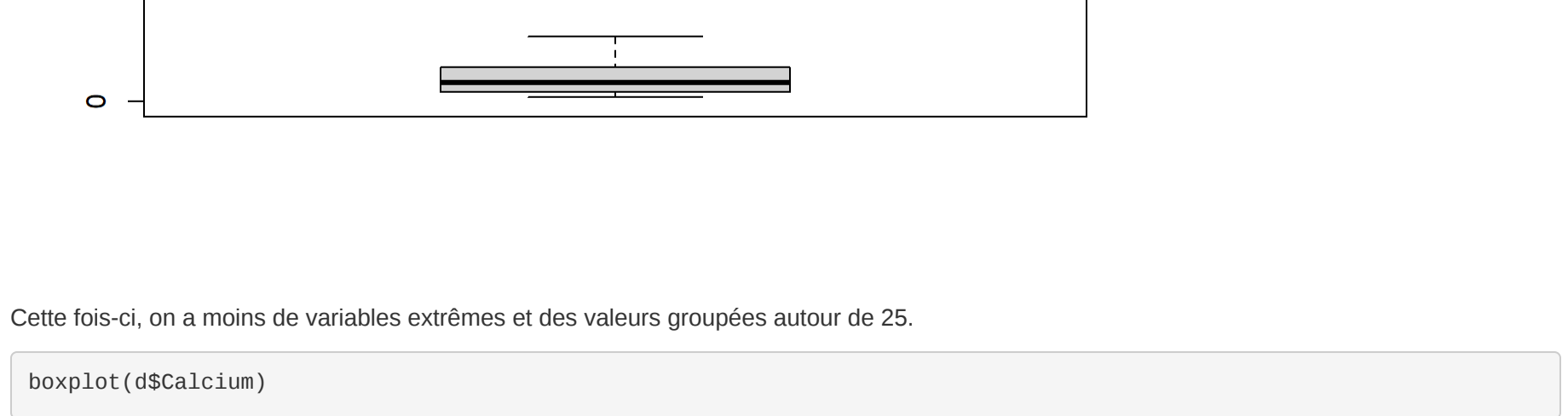
On remarque une concentration autour d'environ 300 mais il y a également quelques valeurs extrêmes allant jusqu'à plus de 4000.



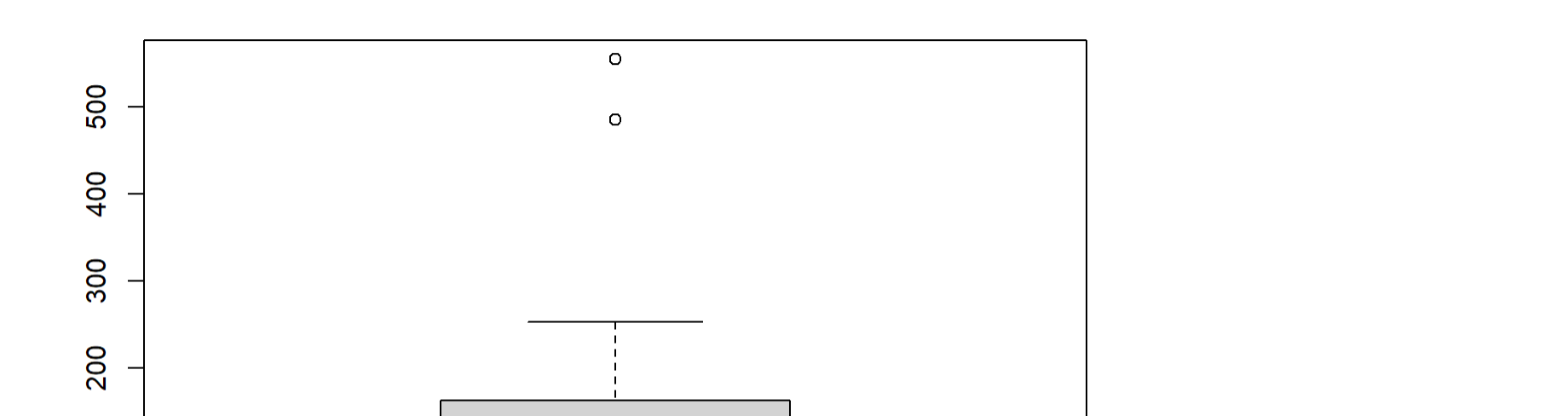
On observe encore une fois une concentration de valeurs autour de 50 avec quelques valeurs extrêmes.



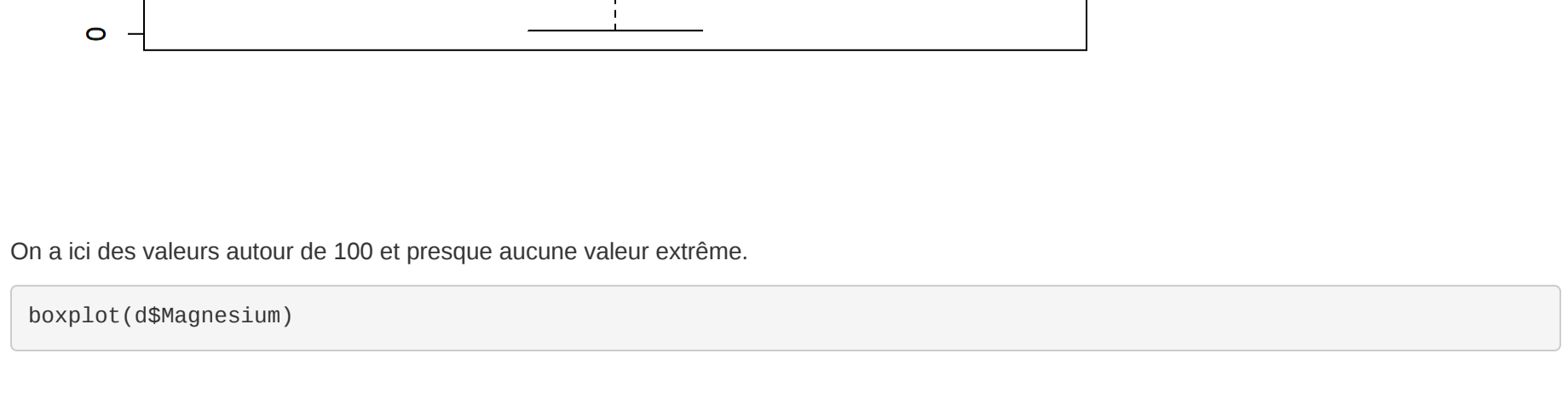
Cette fois-ci, on a moins de variables extrêmes et des valeurs groupées autour de 25.



On a ici des valeurs autour de 100 et presque aucune valeur extrême.



On a des valeurs centrées autour de 20 et quelques valeurs extrêmes.



Le graphe est dur à lire, on a l'air d'avoir majoritairement des petites valeurs (en dessous de 50) mais plusieurs valeurs très extrêmes sont présentes (plus de 1000).

3.

		Bicarbonates	Sulfates	Chlorures	Calcium	Magnesium
#	Bicarbonates	881187.632	4857.470	75563.2341	8964.32473	5539.8591
#	Sulfates	4857.470	111192.824	-1292.56989	36862.92581	7065.9559
#	Chlorures	75563.2341	-1292.570	16594.80982	-40.17527	1541.2075
#	Calcium	8964.325	36862.926	48.17527	14784.66452	2954.3398
#	Magnesium	5539.859	7065.956	1541.20753	2954.33978	956.1896
#	Sodium	342297.995	-1396.804	30225.03871	-1293.97634	561.1957
#	Sodium					
#	Bicarbonates	342297.9946				
#	Sulfates	-1396.8039				
#	Chlorures	30225.0387				
#	Calcium	-1293.9763				
#	Magnesium	561.1957				
#	Sodium	140584.8865				

Sans grande surprise, les valeurs sur lesquelles on a remarqué des valeurs extrêmes possèdent des variances élevées.

4.

		Bicarbonates	Sulfates	Chlorures	Calcium	Magnesium
#	Bicarbonates	1.08009800	0.81295791	0.86742370	0.89376788	0.1907861
#	Sulfates	0.81295791	1.00989809	0.841787152	0.903991893	0.6951841
#	Chlorures	0.8674237	0.84178715	1.008006080	0.894271157	0.235485
#	Calcium	0.8937677	0.90399188	0.894271157	1.08009800	0.785786
#	Magnesium	0.1907861	0.69518406	0.235485026	0.78578603	1.0000000
#	Sodium	0.97275533	0.81276784	0.869257384	-0.028346747	0.0484173
#	Sodium					
#	Bicarbonates	0.97275533				
#	Sulfates	-0.01276784				
#	Chlorures	0.86925738				
#	Calcium	-0.02834676				
#	Magnesium	0.04841710				
#	Sodium	1.00000000				

Il semble exister une corrélation entre : Bicarbonates et sodium ; Bicarbonates et chlorures ; Sulfates et calcium ; Chlorures et sodium ; Calcium et magnésium ; et peut-être magnésium et sulfates

5.

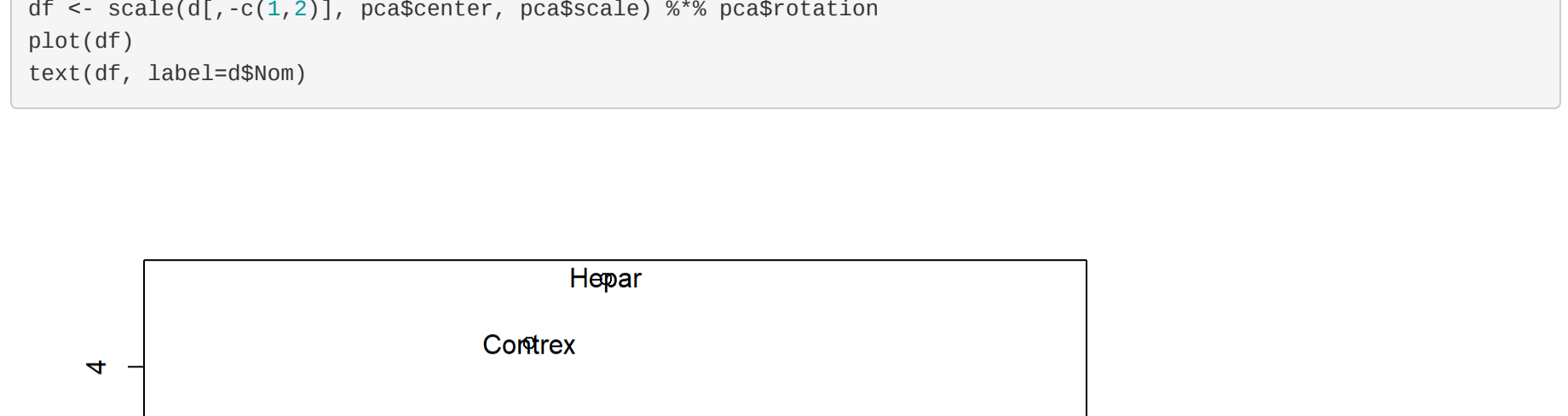
pr <- prcomp(d[,c(1,2)], center = TRUE, scale = TRUE)									
#	Standard deviations (1...p=6):								
#	[1] 1.78958342 1.58644233 0.68574244 0.36312270 0.28612339 0.02506336								
#	Rotation (n x k) = (6 x 6):								
#		PC1	PC2	PC3	PC4	PC5	PC6		
#	Bicarbonates	0.533709	-0.231609	-0.440916	0.531045	-0.0189122	0.6317069		
#	Sulfates	0.216108	0.541609	-0.469316	0.469636	0.41895399	0.2708923		
#	Chlorures	0.507845	0.2427741	0.303885	-0.691786	-0.3242345	0.6899175		
#	Calcium	0.248046	0.5015843	0.1874200	0.207060	-0.71847216	-0.2284214		
#	Magnesium	0.3055515	0.4562973	0.7185506	0.194964	0.36307454	-0.11171519		
#	Sodium	0.507838	-0.287328	-0.3147315	0.1109149	0.2884232	-0.68217221		

summary(pr)									
#	Importance of components:								
#		PC1	PC2	PC3	PC4	PC5	PC6		
#	Standard deviation	1.78958	1.5864	0.68574	0.36312	0.28612	0.02506		
#	Proportion of Variance	0.4837	0.4195	0.06115	0.02189	0.01364	0.00018		
#	Cumulative Proportion	0.4837	0.9031	0.96427	0.98609	0.99998	1.00000		

Il semble qu'il y ait deux composantes principales (il nous permettrait de récupérer 90% de l'information. Le paramètre center permet de décaler les valeurs afin que leurs moyennes soient 0. Le paramètre scale permet de diviser toutes les valeurs par l'écart-type afin d'avoir des valeurs sur une plus petite échelle.

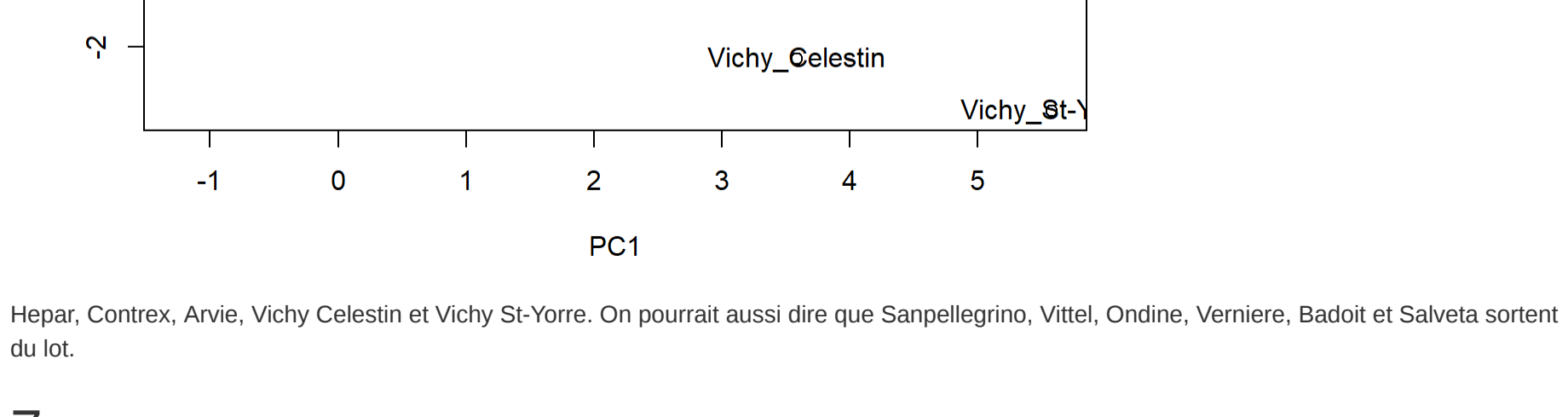
6.

pca <- prcomp(d[,c(1,2)], center = TRUE, scale = TRUE)									
#	df <- scale(d[,c(1,2)], pca\$center, pca\$scale, pca\$rotation)								
#	plot(df, label="don")								



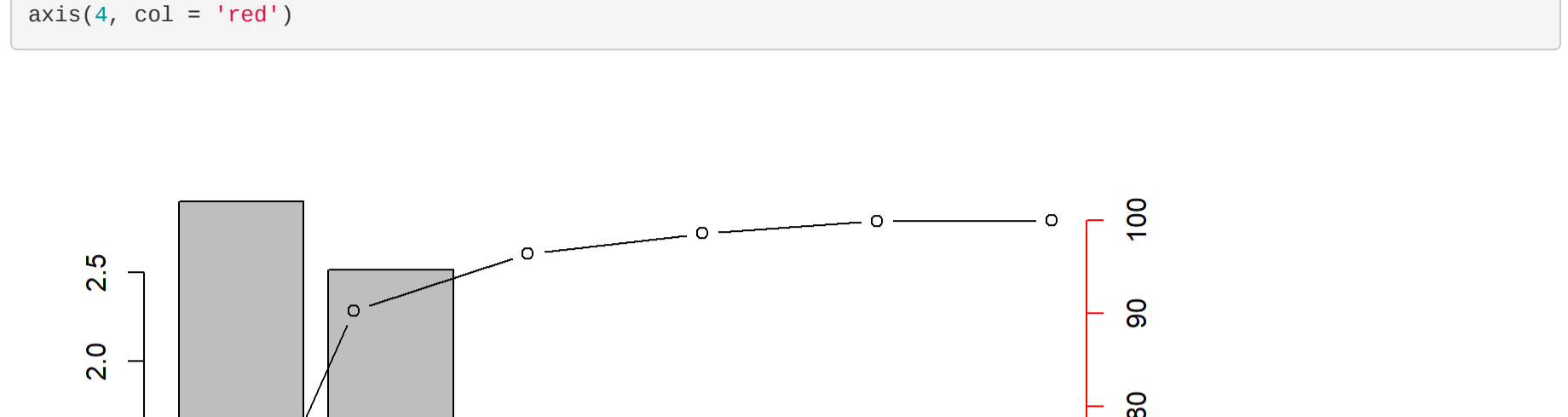
Hepar, Contrex, Arvie, Vichy Celestin et Vichy St-Yorre. On pourrait aussi dire que Sannepellegrino, Vittel, Ordine, Badoit et Salvetat sortent du lot.

7.



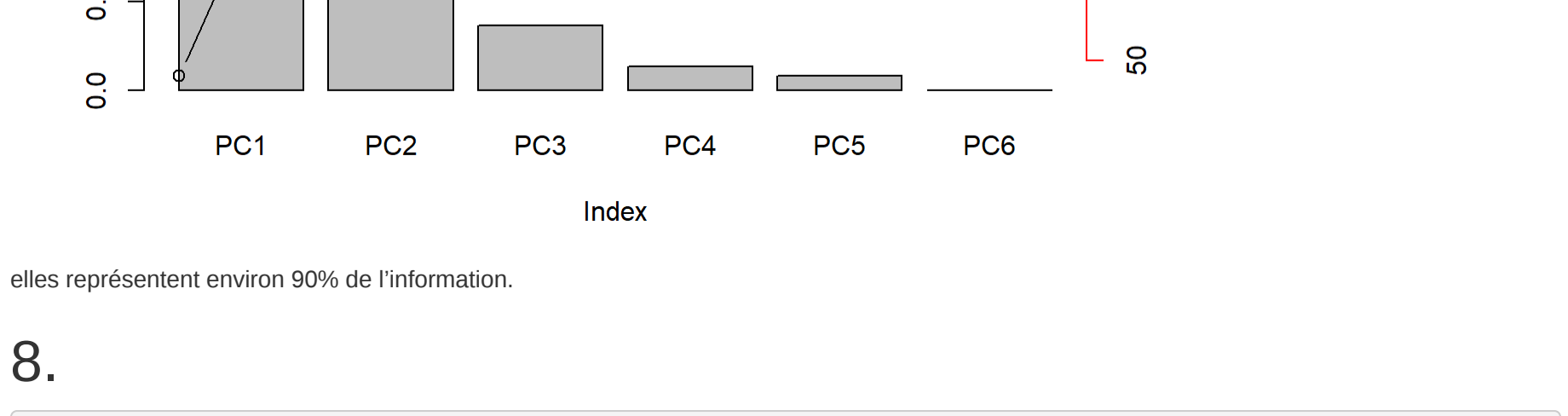
elles représentent environ 90% de l'information.

8.

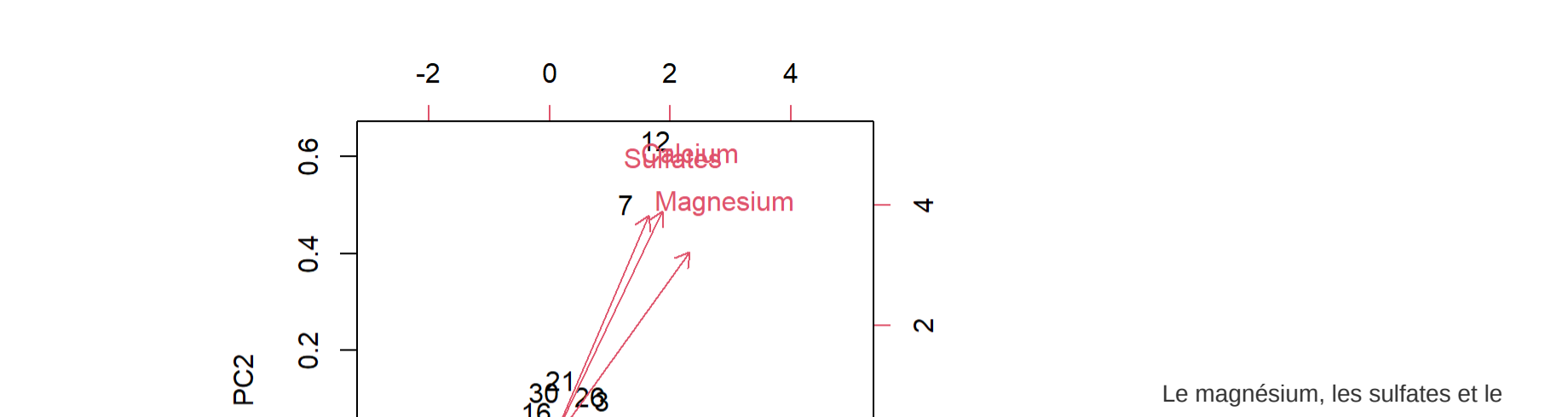


calcium semblent être corrélés entre eux, tout comme le bicarbonate, le sodium et le chlorure.

9.



choisissant les composantes principales 2 et 3.



10.

square <- pr\$rotation ^ 2									
rowSums(square)									
#	Bicarbonates	Sulfates	Chlorures	Calcium	Magnesium	Sodium			
#	1	1	1	1	1	1			

quality <- rowSums(square[,1:2])									
quality									
#	Bicarbonates	Sulfates	Chlorures	Calcium	Magnesium	Sodium			
#	8	0.338922	0.340192	0.3160365	0.354299	0.3014568			0.3400264

Aucune variable n'a l'air mieux représentée qu'une autre sur le premier plan factoriel. On peut couper à la hauteur 21 pour former quatre groupes différents.

#Partitionnement

1.

choisissant les composantes principales 2 et 3.

```
biplot(pr, choices = c(1,3))
```