

## EJEMPLO STATIS

### Datos del ejemplo de Vinos de Abdi et al. (2012)

Advanced Review

# STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling

Hervé Abdi,<sup>1\*</sup> Lynne J. Williams,<sup>2</sup> Domininique Valentin<sup>3</sup>  
and Mohammed Bennani-Dosse<sup>4</sup>



Abdi, H., Williams, L. J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 124–167.  
<https://doi.org/10.1002/wics.198>

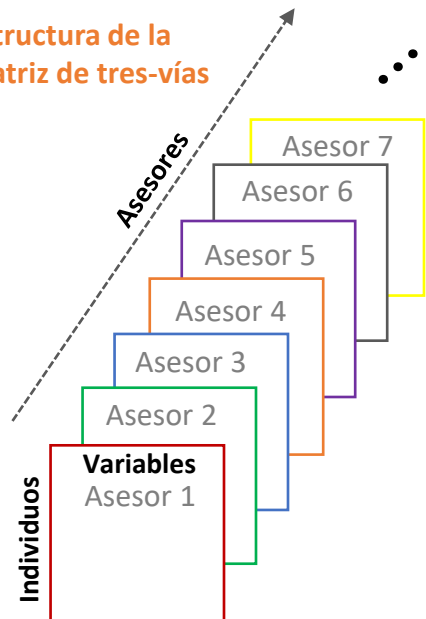
Se seleccionan doce vinos elaborados de uvas *Sauvignon Blanc* procedentes de tres regiones vitivinícolas (cuatro vinos de cada región): Nueva Zelanda, Francia y Canadá. Luego se solicita a 10 asesores expertos evaluar estos vinos. Los evaluadores se les pidieron dos cosas, primero, que evaluaran los vinos con una calificación de 9 puntos, utilizando cuatro variables consideradas como estándar para la evaluación de vinos (Pis de gato, maracuyá, pimienta verde y mineral). Y segundo, si sentían la necesidad los evaluadores tenían la libertad de añadir variables propias (algunos evaluadores no eligen ninguna, algunos eligen una, dos o más variables)

TABLE 1 | Raw Data (Tables  $Y_{[1]}$  through  $Y_{[10]}$ )

		Assessor 1						Assessor 2						Assessor 3						Assessor 4					Assessor 5					
		V1	V2	V3	V4	V5	V6	V1	V2	V3	V4	V7	V8	V1	V2	V3	V4	V9	V10	V1	V2	V3	V4	V8	V1	V2	V3	V4	V11	V12
Vino Nueva Zelanda	NZ <sub>1</sub>	8	6	7	4	1	6	8	6	8	3	7	5	8	6	8	3	7	2	9	5	8	2	6	9	6	9	3	8	2
	NZ <sub>2</sub>	7	5	8	1	2	8	6	5	6	3	7	7	8	7	7	2	8	2	8	7	7	3	5	7	7	7	1	9	2
	NZ <sub>3</sub>	6	5	6	5	3	4	6	6	6	5	8	7	8	7	7	6	9	1	8	8	9	2	7	7	7	7	1	7	2
	NZ <sub>4</sub>	9	6	8	4	3	5	8	6	8	4	6	6	8	2	8	3	9	3	8	8	9	4	7	8	9	7	5	6	1
Vino Francia	FR <sub>1</sub>	2	2	2	8	7	3	2	3	1	7	4	3	3	4	3	6	4	6	4	2	2	4	3	4	4	4	2	4	4
	FR <sub>2</sub>	3	4	4	9	6	1	4	3	4	9	3	5	4	3	4	8	3	9	3	2	2	6	2	4	5	5	6	1	5
	FR <sub>3</sub>	5	3	5	4	8	3	3	3	2	7	4	4	5	4	5	2	3	6	4	4	4	6	4	6	5	7	2	3	1
	FR <sub>4</sub>	5	2	4	8	7	4	4	3	5	5	3	3	6	3	7	7	1	7	5	2	2	9	4	6	6	5	8	4	5
Vino Canadá	CA <sub>1</sub>	8	6	8	4	4	7	8	6	9	5	5	6	8	5	9	1	5	2	7	5	6	3	2	8	6	8	2	5	4
	CA <sub>2</sub>	4	6	2	5	3	4	5	5	5	6	5	8	5	5	4	6	5	1	5	6	6	4	4	6	6	6	4	6	3
	CA <sub>3</sub>	8	4	8	1	3	3	8	4	8	3	7	7	8	3	7	3	5	4	7	3	6	1	6	7	4	8	4	5	1
	CA <sub>4</sub>	5	3	6	4	4	2	5	3	7	4	8	5	5	4	4	5	4	3	5	2	2	6	6	5	5	5	5	6	1

- V1 Cat Pee
- V2 Passion Fruit
- V3 Green Pepper
- V4 Mineral
- V5 Smoky
- V6 Citrus
- V7 Tropical
- V8 Leafy
- V9 Grassy
- V10 Flinty
- V11 Vegetal
- V12 Hay
- V13 Melon
- V14 Grass
- V15 Peach

Estructura de la matriz de tres-vías



K= 10 Asesores

J= diferentes variables en cada k tabla

I= mismos individuos 12 vinos (3 regiones, 4 por cada región)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Asesor	Var	NZ1	NZ2	NZ3	NZ4	FR1	FR2	FR3	FR4	CA1	CA2	CA3	CA4
2	AS1	V1_Cat Pe	8	7	6	9	2	3	5	5	8	4	8	5
3	AS1	V2_Passio	6	5	5	6	2	4	3	2	6	6	4	3
4	AS1	V3_Green	7	8	6	8	2	4	5	4	8	2	8	6
5	AS1	V4_Miner	4	1	5	4	8	9	4	8	4	5	1	4
6	AS1	V5_Smoky	1	2	3	3	7	6	8	7	4	3	3	4
7	AS1	V6_Citrus	6	8	4	5	3	1	3	4	7	4	3	2
8	AS2	V1_Cat Pe	8	6	6	8	2	4	3	4	8	5	8	5
9	AS2	V2_Passio	6	5	6	6	3	3	3	3	6	5	4	3
10	AS2	V3_Green	8	6	6	8	1	4	2	5	9	5	8	7
11	AS2	V4_Miner	3	3	5	4	7	9	7	5	5	6	3	4
12	AS2	V7_Tropic	7	7	8	6	4	3	4	3	5	5	7	8
13	AS2	V8_Leafy	5	7	7	6	3	5	4	3	6	8	7	5
14	AS3	V1_Cat Pe	8	8	8	8	3	4	5	6	8	5	8	5
15	AS3	V2_Passio	6	7	7	2	4	3	4	3	5	5	3	4
16	AS3	V3_Green	8	7	7	8	3	4	5	7	9	4	7	4
17	AS3	V4_Miner	3	2	6	3	6	8	2	7	1	6	3	5
18	AS3	V9_Grassy	7	8	9	9	4	3	3	1	5	5	5	4
19	AS3	V10_Flnty	2	2	1	3	6	9	6	7	2	1	4	3
20	AS4	V1_Cat Pe	9	8	8	8	4	3	4	5	7	5	7	5

Para los análisis colocamos en las columnas los individuos (mismos para todas las k-tablas)

Ordenamos los datos según la variable que forma la tercera dimensión-vía, en este caso es Asesores

Trabajaremos con la librería **ade4**.

Para importar los datos, utilizamos el siguiente código en R:

```
wine<- read_excel(here::here("Data/wine.xlsx"))  
  
asesores<- factor(wine$Asesor,levels = c("AS1","AS2","AS3",  
                                         "AS4", "AS5", "AS6",  
                                         "AS7", "AS8", "AS9", "AS10"))
```

Para estandarizar y formar las k-tablas se utiliza el siguiente código:

```
#---- Aplica componentes principales, normaliza las variables.  
wpca = withinpca(wine[,c(3:14)], asesores,scaling = "total", scannf = F)  
#---- Abre los objetos en K tablas para poder hacer el STATIS.  
ktab1 = ktab.within(wpca,colnames = wine$Var)
```

Estandarizamos los datos:

1. Aplicamos un PCA normalizado a las columnas de la 3 a la 14 (las que tienen los datos)
2. Utilizamos la variable ASESORES para identificar las K-tablas.
3. Aplicar escalamiento TOTAL o PARTIAL de Bouroche (1975)

### Procesos de Estandarización de Bouroche (1975):

**Parcial:** Primero realiza una estandarización por variables (centra y normaliza). Segundo a esa matriz normalizada, le aplica una estandarización (centra y normaliza) cada k-tabla, de tal forma que cada tabla tendrá media 0 y varianza 1.

**Total:** Primero realiza una estandarización por variables (centra y normaliza). Segundo a esa matriz normalizada, le aplica un centrado a cada k-tabla, luego a esta nueva matriz centrada la normaliza (divide para la desviación estándar columna de la tabla total), de tal forma que cada k-tabla tendrá media 0 y la varianza total será igual a 1.

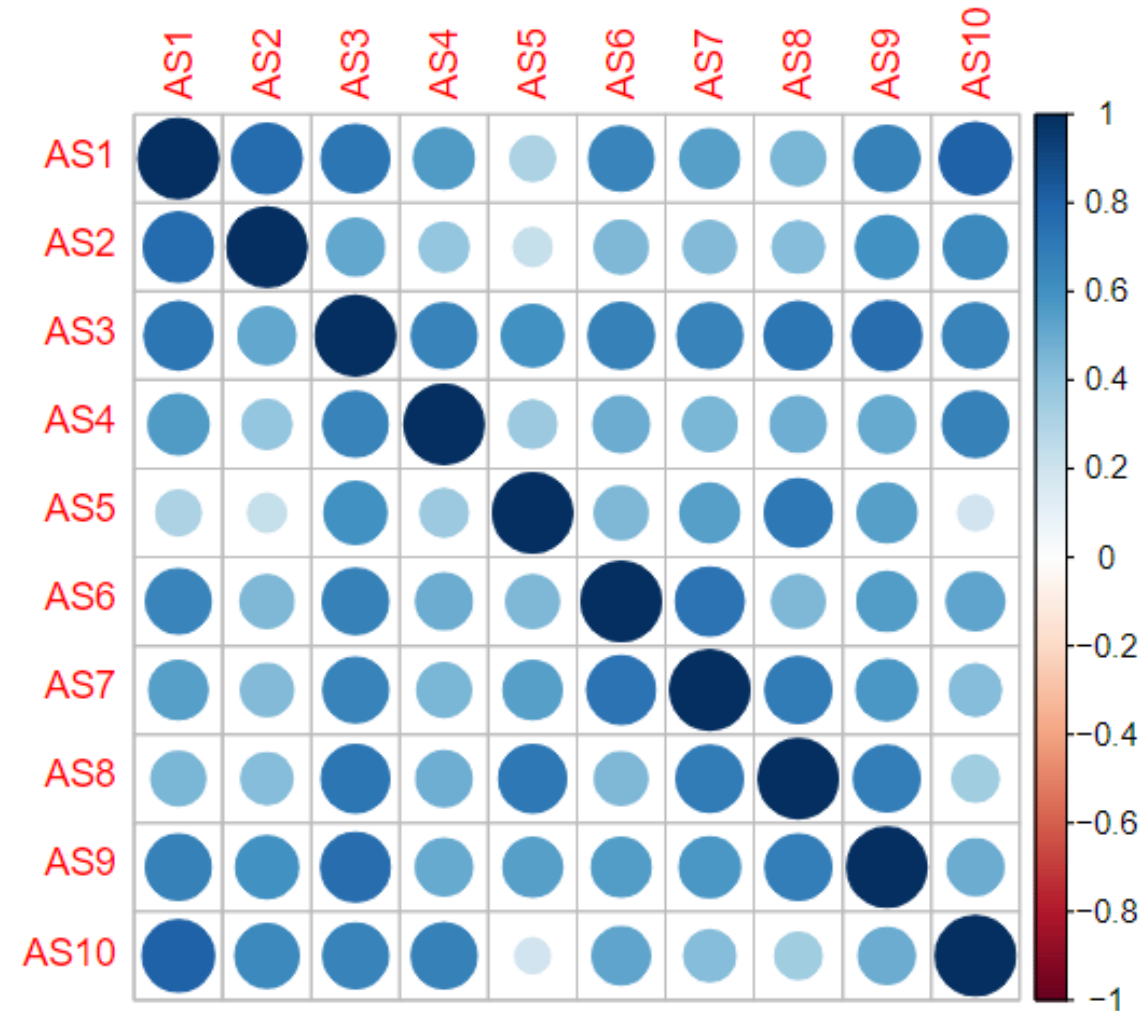
```
statis1 = statis(ktab1, scannf = F) #Aplicar el statis
```

No especificamos la  
cantidad de ejes a  
retener...es a su elección



```
corrplot(statis1$RV)
```

Gráfico de correlaciones



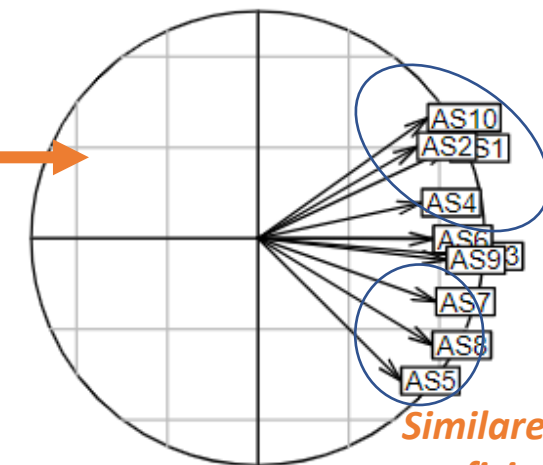


14 plot(statis1)

AVANTE

Interestructura, indica las k-tablas que presentan estructuras similares (opiniones similares)

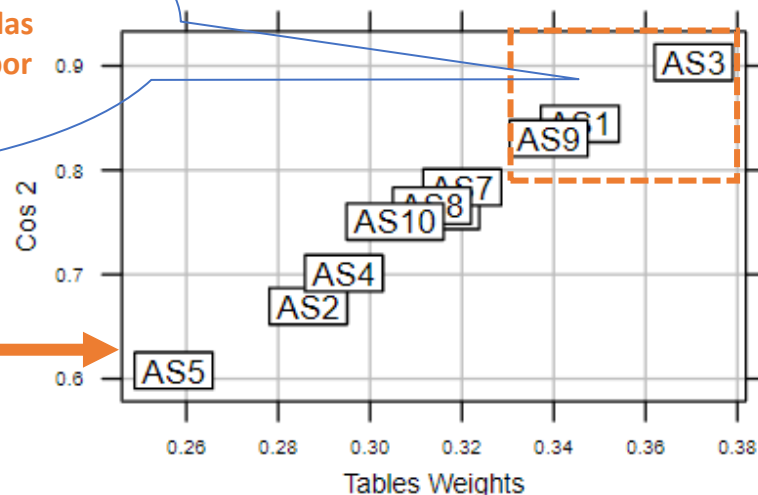
Interestructura



Similares - altos coeficientes RV

Typological Value

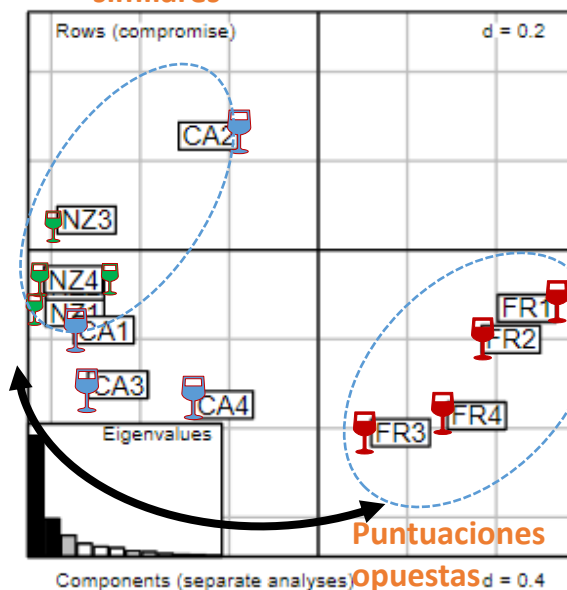
K-tablas que más aportan con información para construir la matriz consenso (pesos altos) y a su vez son las tablas que son mejor representadas por la matriz consenso (cos2 altos)



Se grafican los pesos y el  $\cos^2$  de cada k-tabla

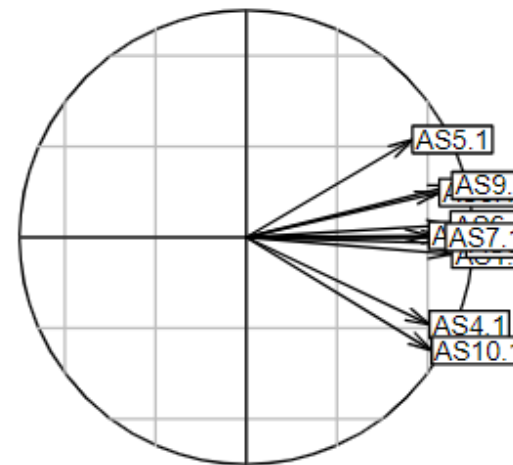
STATIS  
Similares - altos coeficientes RV  
 $d = 0.4$

Puntuaciones similares



Puntuaciones similares

Puntuaciones similares  
opuestas  
 $d = 0.4$



Compromiso, indica la estructura promedio de la puntuación de los 10 asesores

statis1\$RV

Coeficientes de correlación altos, indican estructura similar entre la opinión de los Asesores, esto se ve reflejado en el gráfico de la Inter-estructura

	AS1	AS2	AS3	AS4	AS5	AS6	AS7	AS8	AS9	AS10
AS1	1.0000000	0.7621837	0.7233690	0.5698505	0.3196213	0.6566314	0.5451215	0.4587821	0.6727076	0.8063799
AS2	0.7621837	1.0000000	0.5150242	0.3853480	0.2312591	0.4428257	0.4321389	0.4226073	0.6080369	0.6390418
AS3	0.7233690	0.5150242	1.0000000	0.6636754	0.6048818	0.6798981	0.6646965	0.7218333	0.7513215	0.6671891
AS4	0.5698505	0.3853480	0.6636754	1.0000000	0.3612642	0.4920976	0.4587143	0.4862200	0.5066833	0.6726889
AS5	0.3196213	0.2312591	0.6048818	0.3612642	1.0000000	0.4431497	0.5444689	0.7184164	0.5483896	0.1951651
AS6	0.6566314	0.4428257	0.6798981	0.4920976	0.4431497	1.0000000	0.7310576	0.4400904	0.5520014	0.5230607
AS7	0.5451215	0.4321389	0.6646965	0.4587143	0.5444689	0.7310576	1.0000000	0.6963468	0.5782340	0.4215883
AS8	0.4587821	0.4226073	0.7218333	0.4862200	0.7184164	0.4400904	0.6963468	1.0000000	0.6859440	0.3442326
AS9	0.6727076	0.6080369	0.7513215	0.5066833	0.5483896	0.5520014	0.5782340	0.6859440	1.0000000	0.4983218
AS10	0.8063799	0.6390418	0.6671891	0.6726889	0.1951651	0.5230607	0.4215883	0.3442326	0.4983218	1.0000000

## Coseno cuadrado

`statis1$cos2`

AS	cos2
----	------

AS1 0.8444117

AS2 0.6695396

AS3 0.9034202

AS4 0.7010952

AS5 0.6078350

AS6 0.7609272

AS7 0.7836120

AS8 0.7658789

AS9 0.8304622

AS10 0.7506780

Valores altos de Cos2, indican que el compromiso representa de mejor manera la opinión de esos Asesores (k-tablas)

## Pesos de las tablas

`statis1$RV.tabw`

Asesor	Peso
--------	------

Asesor 1 0.3457081

Asesor 2 0.2865465

Asesor 3 0.3704914

Asesor 4 0.2943041

Asesor 5 0.2572661

Asesor 6 0.3153251

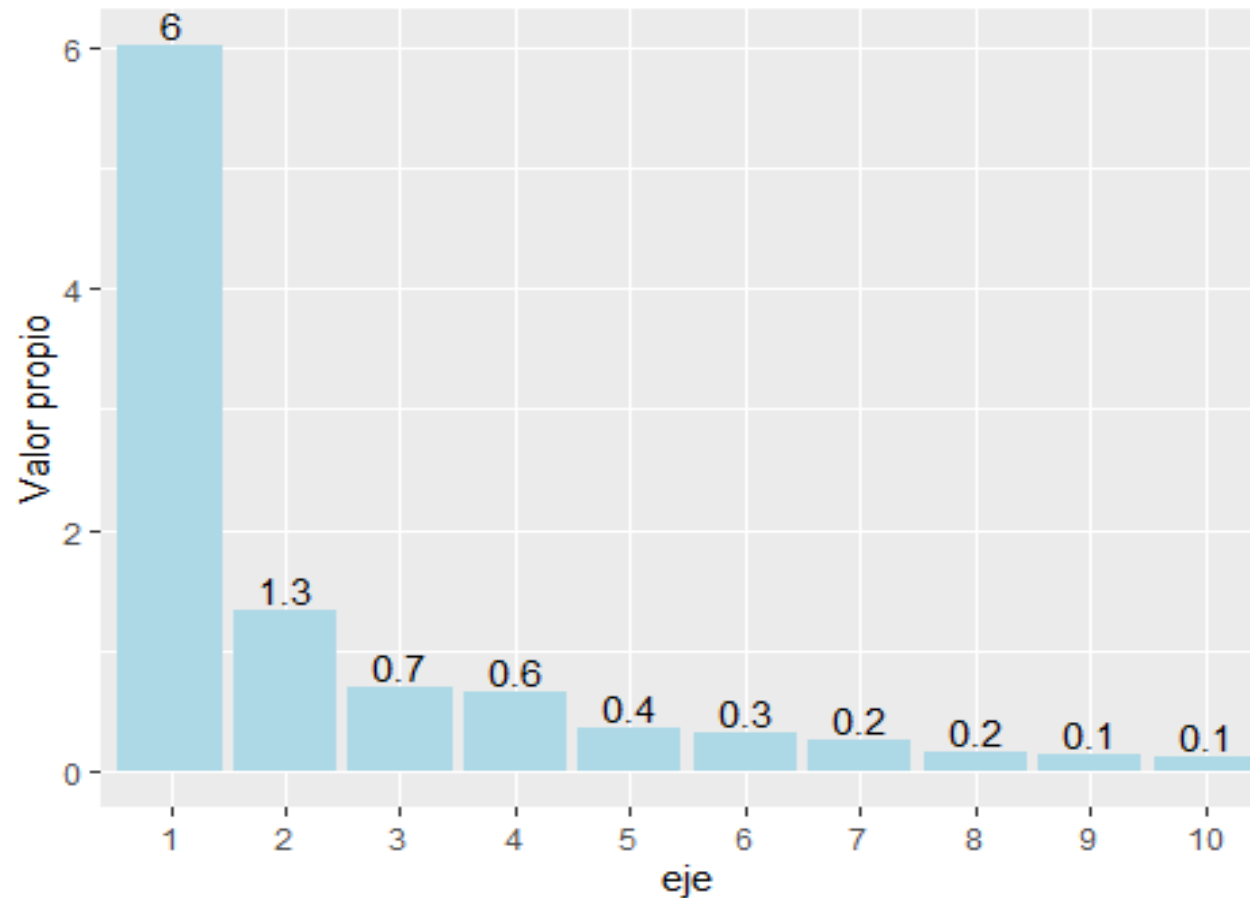
Asesor 7 0.3200585

Asesor 8 0.3135103

Asesor 9 0.3388892

Asesor 10 0.3054677

Valores altos de pesos, indican que esas k-tablas aportan mayor información para contruir el compromiso



la primera componente principal explica el 60.2%, indicando una estructura similar entre las  $K$ -tablas (opiniones de los diferentes asesores).

Valores propios para la inter-estructura

`statis1$RV.eig`

RVEig	Porc.Var	Porc.Var.Acum
6.0215557	0.6021556	0.6021556
1.3365891	0.1336589	0.7358145
0.6874537	0.0687454	0.8045599
0.6469985	0.0646999	0.8692597
0.3571411	0.0357141	0.9049738
0.3134076	0.0313408	0.9363146
0.2483262	0.0248326	0.9611472
0.1525745	0.0152574	0.9764046
0.1272438	0.0127244	0.9891290
0.1087098	0.0108710	1.0000000

*En la inter-estructura los 2 primeros componentes explican un 73,58% de la varianza*

Valores propios para el compromiso

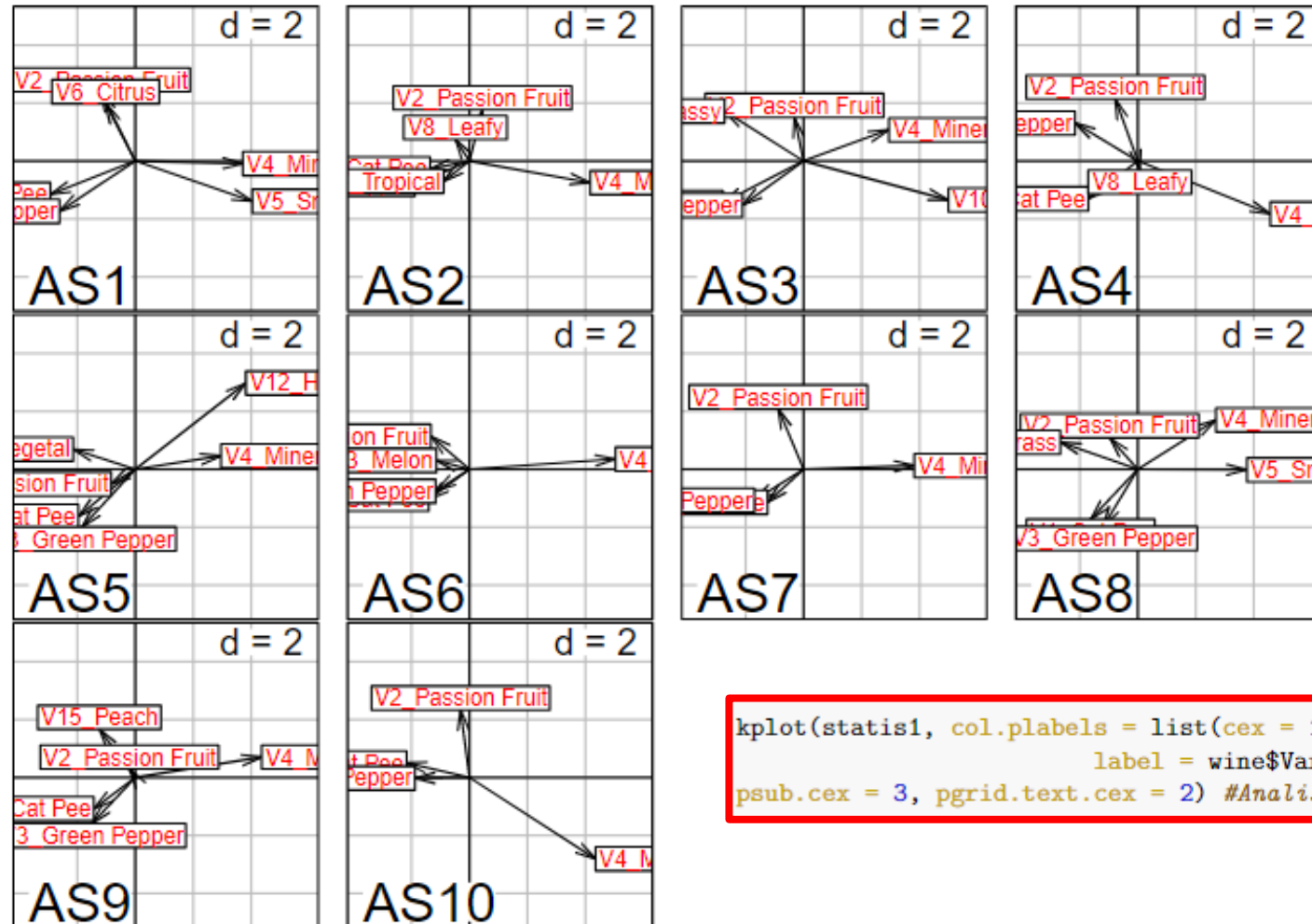
`statis1$C.eig`

CEig	Porc.Var	Porc.Var.Acum
2.2811198	0.5241446	0.5241446
0.7097760	0.1630889	0.6872334
0.4032558	0.0926581	0.7798916
0.2500076	0.0574455	0.8373371
0.1864172	0.0428340	0.8801711
0.1670185	0.0383767	0.9185478
0.1042215	0.0239475	0.9424953
0.0923323	0.0212157	0.9637110
0.0629073	0.0144545	0.9781655
0.0387673	0.0089078	0.9870733
0.0340663	0.0078276	0.9949008
0.0221920	0.0050992	1.0000000

*En el compromiso los 2 primeros componentes explican un 68,72% de varianza*



Intra-estructura, proyecta la información de cada k-tala sobre la información del compromiso



```
kplot(statis1, col.plabels = list(cex = 1.3, col = "red",
                                label = wine$Var),
      psub.cex = 3, pgrid.text.cex = 2) #Análisis separados sobre el consenso
```

*Para la interpretación, nos basamos en el compromiso*

**Asesor 1:** El vino Francés obtuvo puntuaciones por encima del valor promedio en Mineral y Smoky y por debajo del valor promedio en Passion Fruit y Citrus.

Los vinos de Nueva Zelanda y los 3 de Canadá (1, 3, y 4) obtuvieron puntuaciones por encima del valor promedio en Cat Pee y Green Pepper

El vino Canadá 2 obtuvo puntuaciones alrededor del valor promedio.

**Este análisis se debe realizar para cada Asesor (k-tabla)**

