

Análisis Multivariado I

Análisis Factorial

Integrantes:

- Coudet, Lucía CI: 4.545.399-9
- Czarniewicz, Daniel CI: 4.744.781-9
- Talvi, Ramón CI: 4.423.883-5

El presente trabajo busca ser una continuación del primer informe presentado por el equipo. La motivación de este segundo informe fue la aplicación de las técnicas estudiadas en la segunda parte del curso, es decir, las técnicas de Análisis Factorial. Debido a este carácter de continuidad, no es recomendable que se omita la lectura del primer informe, especialmente la introducción del mismo, donde se detalla la estructura de la base de datos con la que se trabajó. Tampoco se incluye aquí un recuento detallado del Análisis de Cluster, y del Análisis Discriminante, realizados en tal oportunidad. Sí se hace uso de los resultados hallados mediante estas técnicas.

Contenido

Análisis de Componentes Principales.....	1
ACP CON LAS 19 VARIABLES.....	1
Localidad	5
Sexo	6
Educación	6
Edad.....	6
Grupos (Clusters).....	7
Sexo - Nivel educativo - Edad	8
Individuos por año.....	8
ACP SOLO CON VARIABLES DE CALIDAD DE LA INSERCIÓN LABORAL.....	8
ACP CON EL RESTO DE LAS VARIABLES	12
Análisis de Correspondencias Simples	14
GRUPOS - LOCALIDAD Y SEXO	15
GRUPOS - NIVEL EDUCATIVO	16
GRUPOS - EDAD	18
Conclusiones	20
Bibliografía y referencias.....	20
Anexos.....	21
ANEXO 1: CÓDIGO R	21

Análisis de Componentes Principales

En virtud de las altas correlaciones observadas en la base de datos, muchas de las cuáles superan el 0.8 en valor absoluto, surge la interrogante de si sería apropiado implementar una técnica de reducción de dimensiones, de manera de poder trabajar en un espacio de dos o a lo sumo tres dimensiones.

Para ello, dado que todas las variables en la base son continuas, se implementó la técnica de componentes principales. Aprovechando la metodología de construcción de los individuos de la base en función de las etiquetas localidad, sexo, edad, y nivel educativo; y teniendo una totalidad de variables socio-económicas referentes a la inserción laboral de los individuos representativos, parece interesante indagar en una posible separación de las observaciones en función de esas variables.

En una primera instancia, se implementó componentes principales con todas las variables de la base, con el objetivo de entender la estructura de los datos de la base sin suprimir del análisis ninguna de las variables.

En una segunda instancia, con el objetivo de analizar la calidad de la inserción laboral de los individuos en la base, de manera de continuar con la pregunta de investigación del primer informe, se implementó componentes principales utilizando solamente las variables *desemp*, *tparcial*, *multiemp*, *subemp*, y *precario*.

Por último, se aplica componentes principales sobre el resto de las variables de la base (es decir, las no incluidas en la segunda implementación), con el objetivo de ver si los grupos formados mediante el análisis cluster según las variables de calidad de inserción laboral se distinguen también en función del resto de las variables.

ACP CON LAS 19 VARIABLES

Considerando todas las variables de la base, como se puede apreciar en la tabla 1.1, el porcentaje de varianza de los datos proyectados retenido por la dimensión 1 (es decir, la primera componente principal) es de un 38.791%. Por su parte, las dimensiones 2 y 3 explican un 21.351% y 11.398% respectivamente. Los porcentajes de varianza acumulada para una totalidad de 2 y 3 dimensiones son de 60.142% y 71.54% respectivamente. A partir de la sexta componente, los aumentos en la variabilidad acumulada comienzan a ser marginales, hasta llegar al 100% para la componente número 19 (lo cual es lo esperado, por construcción del método).

Dado que la tercera dimensión aporta solamente un 10% de la varianza de los datos proyectados, con el fin de simplificar el análisis, parece certero decidir retener solo dos dimensiones.

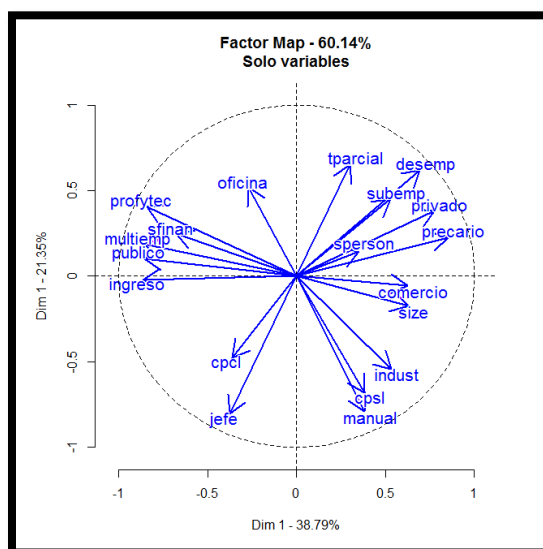
Tabla 1.1

Dim.	Var	% Var	% Var acum
1	7.37	38.79	38.79
2	4.06	21.35	60.14
3	2.17	11.40	71.54
4	1.60	8.43	79.97
5	0.97	5.08	85.05
6	0.79	4.15	89.19
7	0.43	2.26	91.45
8	0.38	2.00	93.46
9	0.26	1.36	94.82
10	0.22	1.17	95.99
11	0.19	1.02	97.00
12	0.14	0.72	97.72
13	0.11	0.56	98.28
14	0.09	0.49	98.77
15	0.09	0.46	99.23
16	0.06	0.30	99.53
17	0.05	0.25	99.79
18	0.03	0.18	99.97
19	0.01	0.04	100.00

En lo que respecta al análisis columna (por variables), el factor map brinda una visión general de cuáles son las variables que más contribuyen a la conformación de uno y otro eje. Teniendo en cuenta que las variables cerca del círculo unidad son las que están mejor representadas, vemos que la dimensión 1 separa entre las variables *ingreso*, *publico*, *multiemp*, y *profytec* por un lado, y *desemp*, *precario*, y *privado* por otro (siendo mencionadas, de nuevo, solamente las que están mejor representadas).

A partir de este primer análisis visual y, complementando con los valores de las coordenadas para las dimensiones uno y dos, es posible sostener que la primera componente está relacionada a la calidad de la inserción laboral de los individuos, y separa entre aquellos con una mejor calidad: profesionales y técnicos, con mayores niveles de ingreso, multiempleados, empleados en el sector público, y aquellos con una mala calidad de inserción laboral: desempleados, individuos en condición de trabajo precario, empleados privados (y la inseguridad que ello conlleva), subempleados.

La segunda componente, por su parte, separa primordialmente a individuos con empleados a tiempo parcial, y empleados de oficina, de los jefes de hogar, y con trabajado manual.



En la tabla 2.2 se muestran las coordenadas, contribuciones, y cosenos cuadrados para el análisis columna. La misma permite apreciar con más detalle cuáles son las variables que contribuyen más a la conformación de los ejes y cuáles son las que están mejor representadas.

Tabla 1.2

Variables	Coordenadas		Contribuciones		Cosenos cuadrados	
	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
jefe	-0.3769	-0.8041	1.9270	15.9379	0.1420	0.6466
desemp	0.6944	0.6188	6.5418	9.4390	0.4821	0.3829
size	0.6264	-0.1753	5.3246	0.7577	0.3924	0.0307
tparcial	0.3015	0.6516	1.2336	10.4647	0.0909	0.4245
multiemp	-0.8559	0.1846	9.9402	0.8404	0.7326	0.0341
privado	0.7745	0.3792	8.1387	3.5448	0.5998	0.1438
publico	-0.8519	0.1038	9.8463	0.2658	0.7257	0.0108
cpsl	0.3836	-0.6802	1.9963	11.4039	0.1471	0.4626
cpcl	-0.3605	-0.4759	1.7629	5.5827	0.1299	0.2265
profytec	-0.8417	0.4067	9.6130	4.0763	0.7085	0.1654
oficina	-0.2700	0.5180	0.9893	6.6137	0.0729	0.2683
manual	0.3816	-0.7908	1.9758	15.4143	0.1456	0.6253
indust	0.5323	-0.5422	3.8449	7.2477	0.2834	0.2940
comercio	0.6233	-0.0555	5.2705	0.0759	0.3885	0.0031
sfinan	-0.6768	0.2490	6.2158	1.5282	0.4581	0.0620
sperson	0.3520	0.1440	1.6810	0.5111	0.1239	0.0207
subemp	0.5263	0.4511	3.7585	5.0152	0.2770	0.2035
precario	0.8509	0.2271	9.8240	1.2709	0.7240	0.0516
ingreso	-0.8635	-0.0199	10.1159	0.0098	0.7456	0.0004

En lo que respecta a la calidad en la representación de las variables, la misma se ve en función de los valores de los cosenos cuadrados. Para la dimensión 1, las variables que están mejor representadas son:

- *multiemp*
- *privado*
- *publico*
- *profytec*
- *precario*
- *ingreso*

Mientras que para la dimensión 2 son:

- *jefe*
- *manual*

Las coordenadas de las variables sobre cada dimensión (es decir, sobre cada dirección) son las correlaciones de cada una de las variables con cada una de las dimensiones. Cada componente debe ser interpretada en función de las variables más correlacionadas con ella. Como muestra la tabla 2.2, resulta semejante interpretar la componente en función de las variables más correlacionadas con la misma o atendiendo a los valores de las contribuciones de cada variable a la conformación del eje.

En virtud de lo anterior, es posible establecer que las variables

- *multiemp*
- *privado*
- *publico*
- *profytec*
- *precario*
- *ingreso*

son las que más contribuyen a la conformación del eje 1. Mientras que las variables:

- *jefe*
- *desemp*
- *tparcial*
- *cpsl*
- *manual*

son las que más contribuyen a la conformación del eje 2.

A modo de enriquecer la comprensión de la estructura subyacente en la base de datos y, abogando por un análisis dual, resulta interesante proyectar a los individuos de la base sobre el factor map, en función de las distintas etiquetas: *localidad*, *sexo*, *edad*, *nivel educativo*, *clusters*, y *tiempo*. Se realiza la implementación, en todos los casos, haciendo un paralelismo con los resultados obtenidos en el análisis Cluster. A continuación se presentan los resultados:

Localidad

El factor map nos permite observar es que no existe una agrupación de los individuos en función de las variables que discrimine según el individuo resida en el interior o en Montevideo. Es importante destacar que el resultado anterior es consistente con uno de los hallazgos del primer informe al aplicar Cluster. De los 6 grupos seleccionados, solo el grupos 5 discernía parcialmente en función de la localidad de los

individuos. Para los restantes 5 grupos, resultó que cerca de la mitad de los individuos agrupados pertenecía al interior, y cerca de la otra mitad a Montevideo.

Sexo

Aquí se ve que efectivamente existe cierta agrupación de las observaciones en función del sexo. Uno de los puntos a destacar es que entre los profesionales y técnicos las mujeres son las que predominan, así como también lo hacen en variables como desempleo, precariedad y subempleo. También es posible observar que existe una mayor cercanía entre las mujeres y el empleo público.

Los hombres, por su parte, son los que tienen mayores ingresos y son en su mayoría cuenta propistas con y sin local, jefes de hogar, trabajadores manuales, trabajadores en la industria.

Educación

En esta oportunidad, se demuestra claramente cómo las observaciones toman valores altos en distintas variables según el nivel educativo. Como era de esperarse, y en concordancia con los resultados en Cluster, los profesionales y técnicos, los oficinistas, los empleados públicos y con multiempleo son en su mayoría individuos con nivel educativo terciario. Asimismo, los mismos son los que retienen más ingreso. La división entre los individuos con nivel educativo primario y secundario no parece ser tan clara pero vemos que la variable desempleo está más presente entre estos últimos mientras que la situación de precariedad está más asociada a los individuos que cuentan solamente con nivel educativo primario.

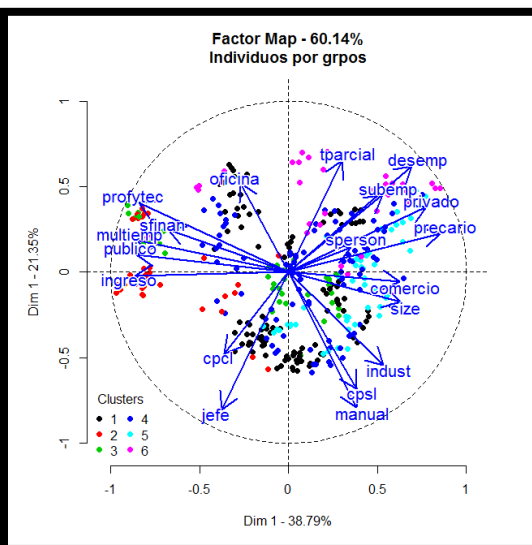
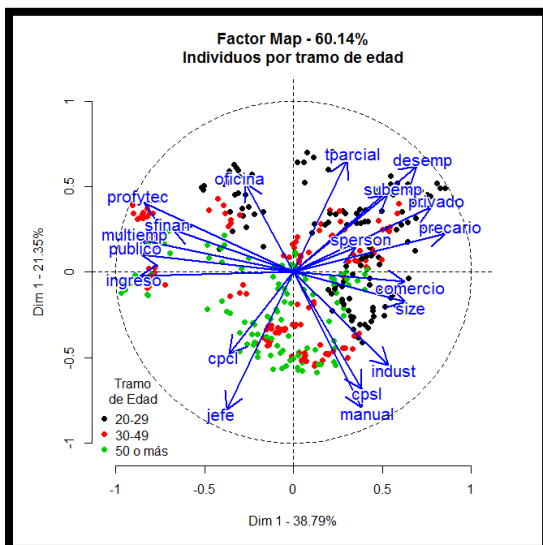
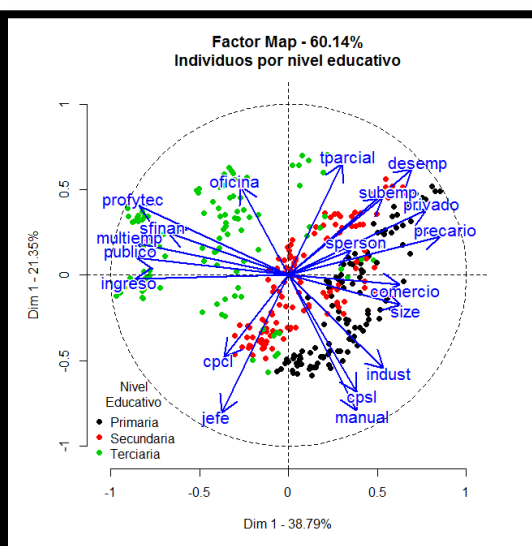
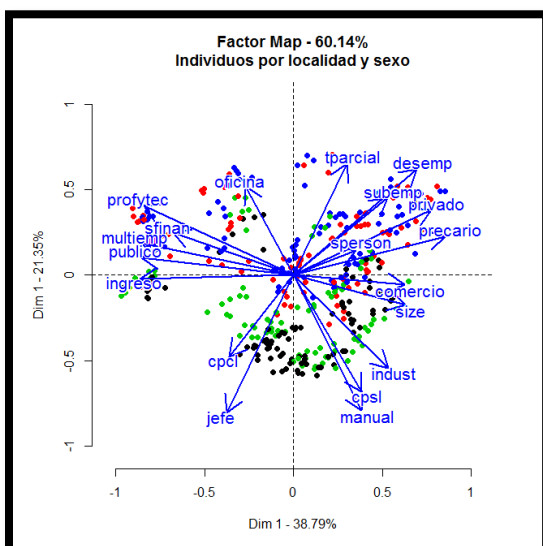
Edad

En lo que respecta a la edad, si bien ya no es tan clara la separación de los individuos como para el caso del nivel educativo, el plano factorial sugiere que los profesionales y técnicos, personas con multiempleo, y empleados públicos, son en su mayoría individuos de más de 30 años. Para la variable ingreso, son también los individuos mayores de 30 años los que alcanzan niveles más altos.

Se detecta también que la precariedad, el subempleo, desempleo, y tiempo parcial abunda entre los individuos de entre 20 y 29 años.

Grupos (Clusters)

Proyectando los individuos según el cluster al que pertenecen vemos claramente que los profesionales y técnicos, multiempleados, y empleados públicos son en su mayoría individuos pertenecientes a los grupos 2 y 3, los cuáles son ambos formados por individuos de nivel educativo terciario y mayores de 30 años. Por su parte, el grupo 2, en el que predominan universitarios de mayores de 29 años, es el que recibe mayores ingresos. También podemos extraer del factor map que el desempleo abunda en el grupo 6, conformado en su mayoría por mujeres jóvenes (de entre 20 y 29 años) con nivel educativo terciario.

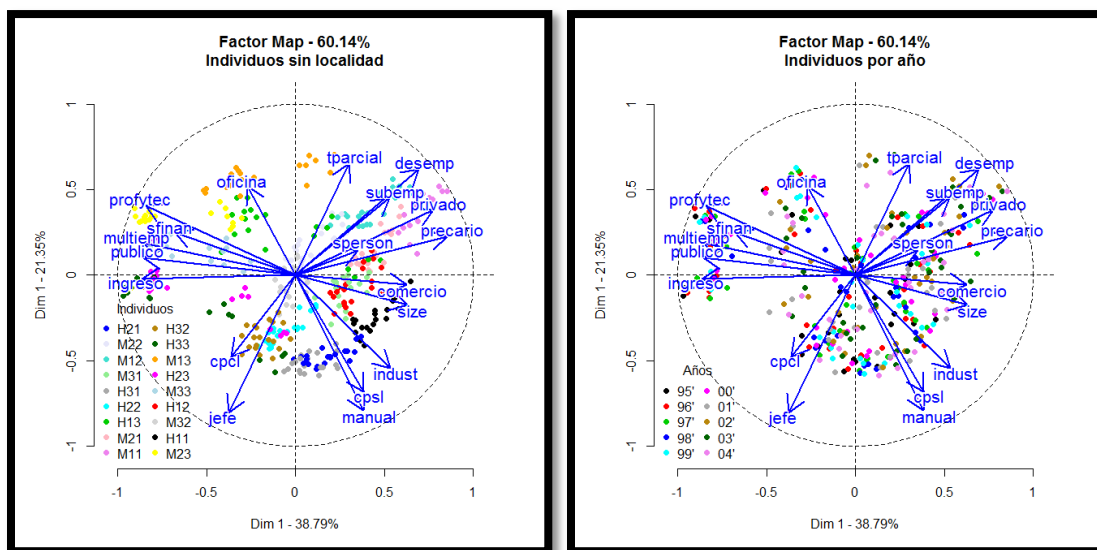


Sexo - Nivel educativo - Edad

A modo de combinar los análisis anteriores respecto a separación de los individuos en las variables según sexo, edad, y nivel educativo, y con el fin de tener una aproximación multidimensional (en el sentido de combinar etiquetas), el plano factorial sugiere que, como habíamos encontrado, los profesionales y técnicos son en su mayoría mujeres de entre 30 y 49 años con nivel educativo terciario. En lo que respecta al ingreso, los hombres de más de 30 años y nivel educativo terciario son los que alcanzan los más altos valores. Los desempleados y subempleados son en su mayoría hombres de entre 20 y 49 años y nivel educativo secundario. La precariedad y el empleo privado abundan entre las mujeres de entre 20 y 49 años con nivel educativo primario.

Individuos por año

El gráfico no nos permite extraer conclusiones relevantes en función de la variable tiempo.



ACP SOLO CON VARIABLES DE CALIDAD DE LA INSERCIÓN LABORAL

Como fue mencionado anteriormente, resulta interesante analizar cómo es la calidad de la inserción laboral de los individuos. En esta oportunidad, las variables seleccionadas son: *desemp*, *subemp*, *precario*, *tparcial*, y *multiemp*.

Es importante destacar que las variables *desemp*, *subemp*, y *precario*, están asociadas a una situación del individuo en la cual no tiene posibilidad de elección. Es decir, representan situaciones no controladas por los individuos, en el sentido que el

desempleado desea ser empleado, y el subempleado lo es porque desea trabajar más horas.

Las variables *tparcial* y *multiemp*, por su parte, representan decisiones propias de los individuos. El empleo a tiempo parcial está asociada a estudiantes en nivel terciario (eligen trabajar menos de 8 horas), mientras que, como revelaron las estadísticas descriptivas y los análisis anteriores, el multiempleo se concentran predominantemente en torno a profesionales y técnicos, que a su vez son quienes perciben mayores ingresos y quienes tienen una mejor calidad de inserción laboral.

Como muestra la tabla 2.1, en esta oportunidad, quedarse con dos dimensiones implica retener un 83.15% de la varianza de los datos proyectados, lo cual es un porcentaje significativamente alto. Esto quiere decir que, reducir de 5 a 2 dimensiones, implica perder solamente un 16.85% de la varianza.

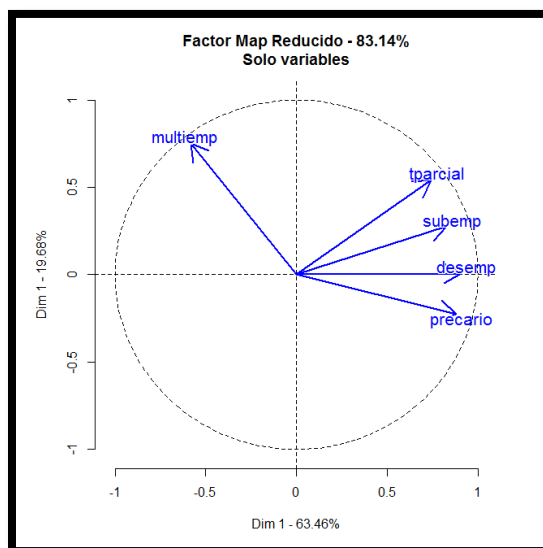
Tabla 2.1

Dim.	Var	% Var	% Var acum
1	3.17	63.47	63.47
2	0.98	19.68	83.15
3	0.35	7.06	90.21
4	0.30	5.92	96.13
5	0.19	3.87	100.00

A partir de la tabla 2.2 y el análisis visual del factor map, es posible constatar que las variables que reflejan decisiones personales de los individuos en cuanto a su inserción laboral, fueron las que más contribuyeron a la formación del segundo eje, mientras que aquellas que están fuera del ámbito de las decisiones personales hicieron lo propio para el primer eje.

Tabla 2.2

Variables	Coordenadas		Contribuciones		Cosenos cuadrados	
	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
desemp	0.9070	0.0076	25.9229	0.0059	0.8226	0.0001
tparcial	0.7430	0.5425	17.3955	29.9071	0.5520	0.2943
multiemp	-0.5844	0.7506	10.7645	57.2579	0.3416	0.5634
subemp	0.8268	0.2746	21.5436	7.6633	0.6836	0.0754
precario	0.8794	-0.2255	24.3736	5.1659	0.7734	0.0508

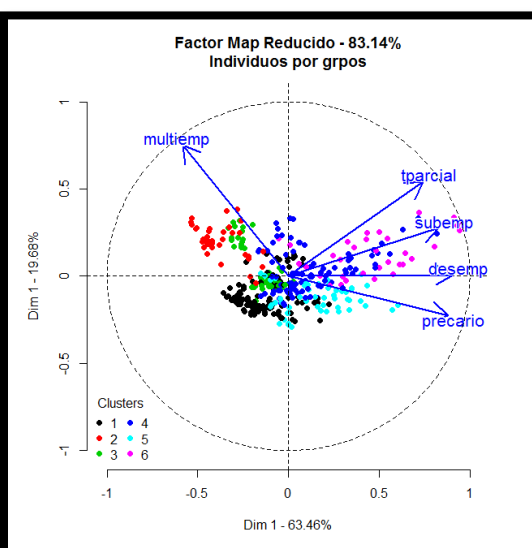
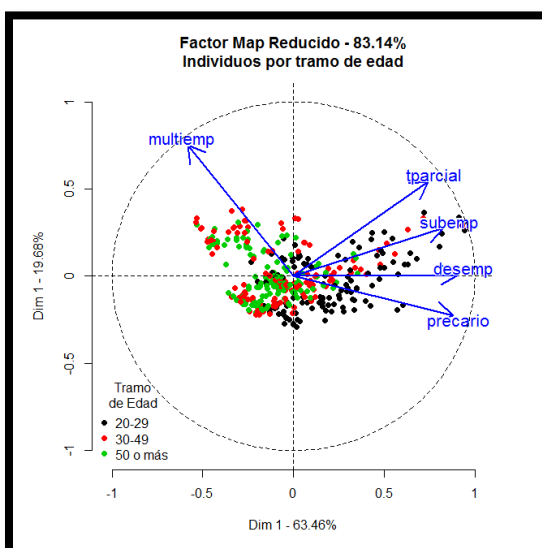
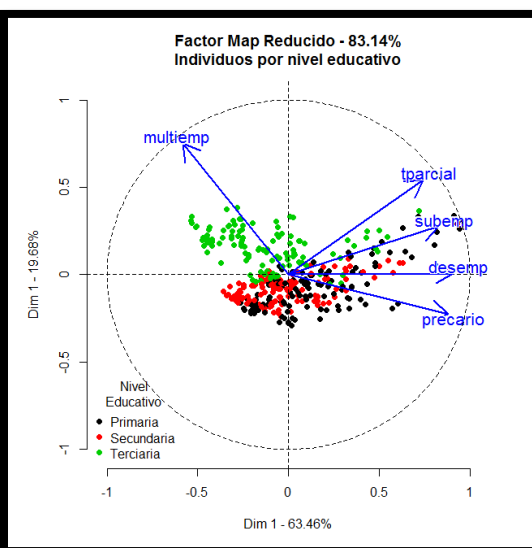
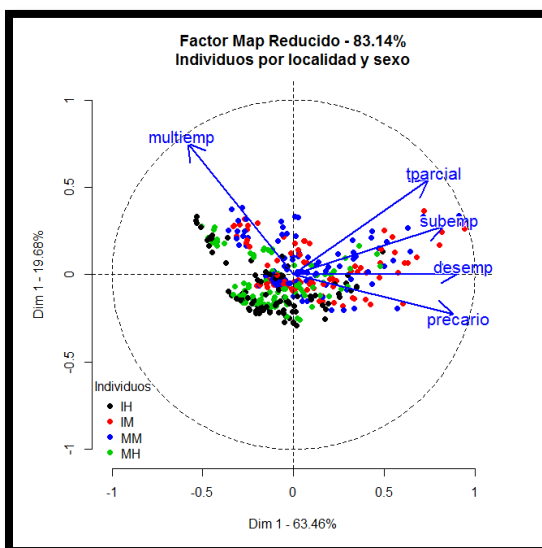


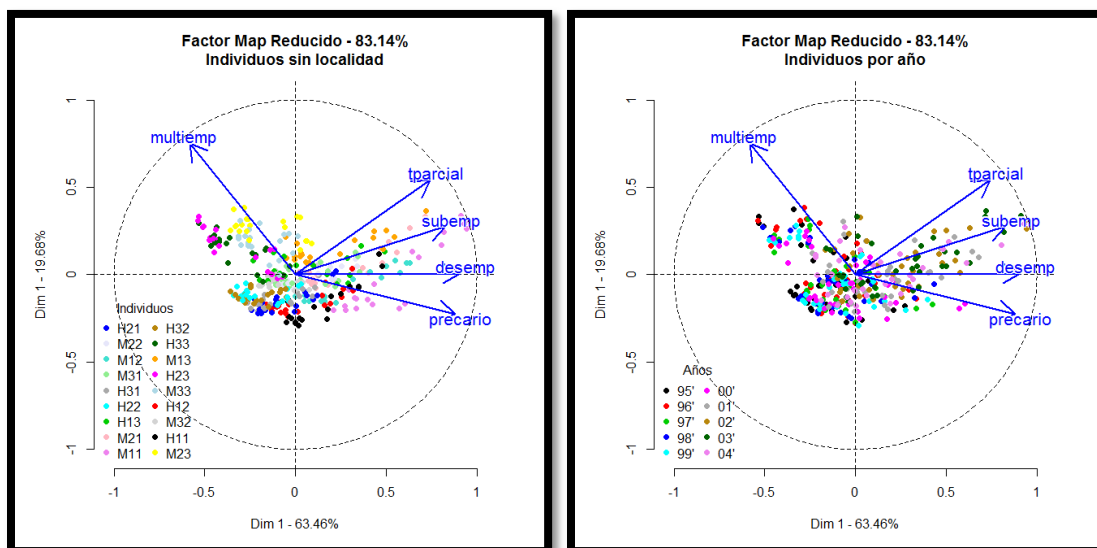
Recurriendo al análisis fila de manera similar a lo realizado en el análisis para todas las variables de la base, se proyectan los individuos sobre el factor map y se diferencia según las etiquetas: *localidad, sexo, nivel educativo, edad, grupos en Cluster*, por año, y considerando *sexo, nivel educativo, y edad* en forma conjunta.

De nuevo, localidad no aporta a separar a los individuos en las variables. En lo que respecta a sexo, los subempleados son predominantemente mujeres. El nivel educativo es, de nuevo, en lo que más se diferencian los individuos. Vemos que los individuos con multiempleo son los de nivel educativo terciario, mientras que entre precarios, subempleados, y desempleados, predominan individuos de nivel educativo primario y secundario. Según tramo de edad, vemos que los precarios, subempleados, y desempleados son en su mayoría individuos de entre 20 y 29 años y, en menor medida, de 30 a 49 años.

En lo que respecta a los grupos obtenidos en cluster, se constata que los individuos del grupo 2, donde predominan universitarios mayores de 29 años, y del grupo 3, compuesto en su totalidad por individuos de nivel educativo terciario, con predominancia de mayores de 50 años, son en su mayoría multiempleados. El grupo 6, compuesto en su mayoría por mujeres de nivel educativo terciario y rango de edad entre 20 y 29 años, son en su mayoría subempleadas.

Si vemos a los individuos según el año, de nuevo no podemos extraer ninguna conclusión clara y, analizando según sexo, edad, y nivel educativo de forma conjunta, tampoco podemos concluir nada. Esto está probablemente relacionado al resultado en cluster de que algunos de los grupos formados resultaron ser muy grandes (en cantidad de individuos), y bastante heterogéneos.





ACP CON EL RESTO DE LAS VARIABLES

Al igual que cuando se aplicó el método sobre la totalidad de las variables de la base, aquí también se obtienen resultados que deben tomarse con precaución debido a la baja inercia capturada por las dos primeras dimensiones del análisis (en este caso, es de un 60.99%).

Tabla 3.1

Dim	Var	% Var	% Var acum
1	5.46	38.98	38.98
2	3.08	22.01	60.99
3	1.82	13.00	73.99
4	1.14	8.11	82.10
5	0.83	5.91	88.00
6	0.59	4.18	92.18
7	0.29	2.10	94.28
8	0.27	1.90	96.18
9	0.18	1.30	97.48
10	0.14	0.98	98.46
11	0.11	0.77	99.22
12	0.06	0.39	99.62
13	0.05	0.33	99.95
14	0.01	0.05	100.00

En la tabla 3.2 es posible apreciar que son ahora las variables *profytec* y *público*; y *jefe* y *privado*, las que más contribuyen a la conformación de los ejes 1 y 2

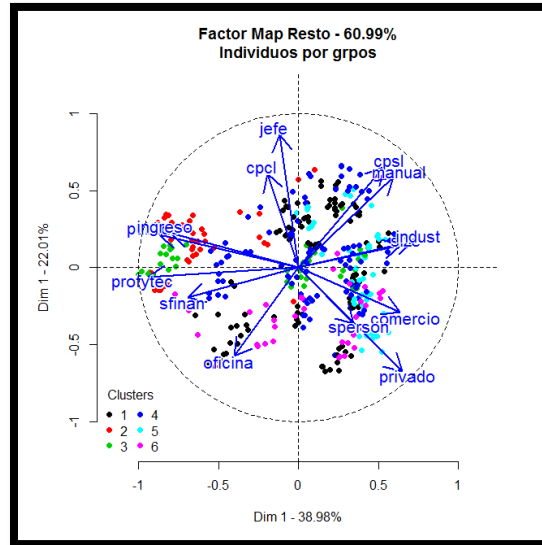
respectivamente. Si hacemos un paralelismo con la implementación del método para las 5 variables de calidad de inserción laboral, en esta oportunidad también observamos que la primera componente separa entre aquellos con una mejor inserción laboral y aquellos con una más endeble. La variable *profytec* tiene una correlación alta con las variables multiempleo, e ingreso, por lo que parecería que la información revelada es de alguna manera similar.

Tabla 3.2

Variables	Coordenadas		Contribuciones		Cosenos cuadrados	
	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
jefe	-0.1194	0.8653	0.2613	24.3044	0.0143	0.7488
size	0.6395	0.1448	7.4941	0.6806	0.4090	0.0210
privado	0.6464	-0.6729	7.6573	14.6945	0.4179	0.4527
publico	-0.8693	0.2151	13.8468	1.5018	0.7557	0.0463
cpsl	0.5301	0.6543	5.1492	13.8939	0.2810	0.4281
cpcl	-0.1938	0.6085	0.6884	12.0181	0.0376	0.3703
profytec	-0.9383	-0.0587	16.1321	0.1118	0.8804	0.0034
oficina	-0.4013	-0.5752	2.9516	10.7376	0.1611	0.3308
manual	0.5930	0.5894	6.4436	11.2744	0.3516	0.3474
indust	0.7167	0.1723	9.4124	0.9639	0.5137	0.0297
comercio	0.6331	-0.2886	7.3450	2.7037	0.4008	0.0833
sfinan	-0.6907	-0.1916	8.7413	1.1915	0.4770	0.0367
sperson	0.3459	-0.3638	2.1927	4.2945	0.1197	0.1323
ingreso	-0.7985	0.2240	11.6842	1.6292	0.6376	0.0502

Proyectando los individuos sobre el grupo al que pertenecen según el análisis de Cluster vemos que en los grupos 2 y 3 predominan individuos empleados en el sector público, con ingresos altos. Asimismo, los profesionales y técnicos son en su mayoría individuos pertenecientes a estos dos grupos. Recordando la conformación de los grupos, lo común a estos dos es que en ellos existe amplia predominancia de individuos con nivel educativo terciario y de edad avanzada.

El resto de los grupos se encuentran dispersos por el resto de las variables, por lo cual no es posible obtener conclusiones claras.



Análisis de Correspondencias Simples

Con el objetivo de determinar si existe alguna asociación entre los grupos resultantes del análisis Cluster y las etiquetas con las que fueron contruidos los individuos representativos de la base de datos, se implementó la técnica de análisis de correspondencias simples.

El presente análisis sólo tiene sentido si existe alguna dependencia entre las variables utilizadas en la tabla de contingencia. Para testear lo anterior, se utilizó el test Chi-cuadrado de Pearson. Para todos los casos (los grupos de cluster contra cada una de las etiquetas de los individuos representativos), se rechazó la hipótesis nula de independencia, lo cual habilita la implementación del ACS.

Es importante destacar que, a priori, es posible establecer que los pesos para las modalidades localidad, sexo, nivel educativo, y edad van a ser los mismos (es decir, el valor que nos devuelve R como *colmass*). Esto es así puesto que, por construcción de la base, se tiene la misma cantidad de individuos en cada modalidad, por lo que, los pesos van a ser los mismos para cada uno de ellos. No se hará referencia a este hecho nuevamente en lo que resta del análisis.

Otro punto importante a subrayar es el hecho de que, también a priori, se sabe que los grupos en Cluster son de diferentes tamaños (en cantidad de individuos), siendo el grupo 1 y 4 particularmente altos. Esto significa que, los grupos que van a tener mayor peso van a ser siempre primero el 1 y luego el 4. Tampoco se volverá a mencionar este hallazgo en lo que sigue.

GRUPOS - LOCALIDAD Y SEXO

En una primera instancia, se implementó ACS con los grupos de cluster y una variable cualitativa cuyas modalidades son en función del sexo y la localidad donde reside el individuo. De esta manera, se tienen 4 modalidades: interior-hombre (IH), interior-mujer (IM), Montevideo-hombre (MH), Montevideo-mujer (MM). La tabla de contingencia se presenta a continuación:

Tabla 4.1.1

		Localidad - Sexo			
		IH	IM	MH	MM
Grupos	1	38	7	41	18
	2	14	3	17	6
	3	0	24	1	18
	4	11	26	21	30
	5	24	15	6	6
	6	3	15	4	12

Tabla 4.1.2

Dim	Value	% Var	% Var acum
1	0.28	83.10	83.10
2	0.06	16.70	99.80
3	0.00	0.20	100.00

La tabla 4.1.2 muestra que con 2 dimensiones se logra captar casi el 100% de la inercia de la nube de puntos proyectada, lo cual quiere decir que reducir dimensiones prácticamente no implica pérdida de información.

En lo que respecta a contribución a la inercia, la misma está dada primordialmente por el grupo 3. La distancia chi-cuadrado es la distancia de cada grupo al centroide. Cuando la misma es cercana a uno, la modalidad está mejor representada. Este último punto es de suma importancia ya que para interpretar los puntos, los mismos deben estar bien proyectados. El grupo 3 es totalmente dominante en lo que respecta a calidad de representación en el factor map.

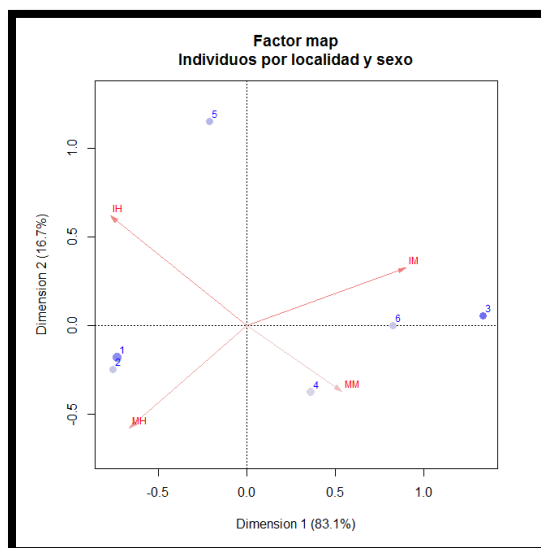
Tabla 4.1.3

Grupos	Mass	ChiDist	Inertia	Coordenadas	
				Dim. 1	Dim. 2
1	0.289	0.542	0.085	-1.002	-0.366
2	0.111	0.570	0.036	-1.038	-0.504
3	0.119	0.974	0.113	1.825	0.120
4	0.244	0.323	0.026	0.497	-0.765
5	0.142	0.585	0.049	-0.292	2.359
6	0.094	0.603	0.034	1.129	0.003

Tabla 4.1.4

Loc/Sexo	Mass	ChiDist	Inertia	Coordenadas	
				Dim. 1	Dim. 2
IH	0.250	0.640	0.102	-1.056	1.266
IM	0.250	0.677	0.114	1.231	0.666
MH	0.250	0.562	0.079	-0.910	-1.178
MM	0.250	0.433	0.047	0.735	-0.754

Para las modalidades de localidad-sexo, si bien no existen grandes diferencias, vemos que Interior-Hombre e Interior-Mujer son las que están mejor representadas, al mismo tiempo que son las que más contribuyen a la inercia. A continuación se presenta el factor map asociado:



No parece haber ningún punto que esté especialmente alejado de la nube de puntos, por lo cual no es preciso quitar ningún punto e incluirlo luego como suplementario. El factor map sugiere que, mientras que el primer eje separa por localidad, el segundo eje separa por sexo.

GRUPOS - NIVEL EDUCATIVO

En una segunda instancia nos proponemos investigar si existe alguna asociación entre los grupos formados, y el nivel educativo de los individuos. A continuación se presentan la tabla de contingencia, y la tabla de valores propios y varianzas:

Tabla 4.2.1

		Nivel educativo		
		Primaria	Secundaria	Terciario
G r u p o s	1	29	55	20
	2	0	1	39
	3	13	13	17
	4	37	30	21
	5	34	13	4
	6	7	8	19

Tabla 4.2.2

Dim	Value	% Var	% Var acum
1	0.31	83.00	83.00
2	0.06	17.00	100.00

La tabla 4.2.2 muestra que la primera dimensión capta el 83% de la varianza de los datos proyectados, mientras que la segunda capta el 17%. Entre las dos, se acumula el 100% de la varianza.

Tabla 4.2.3

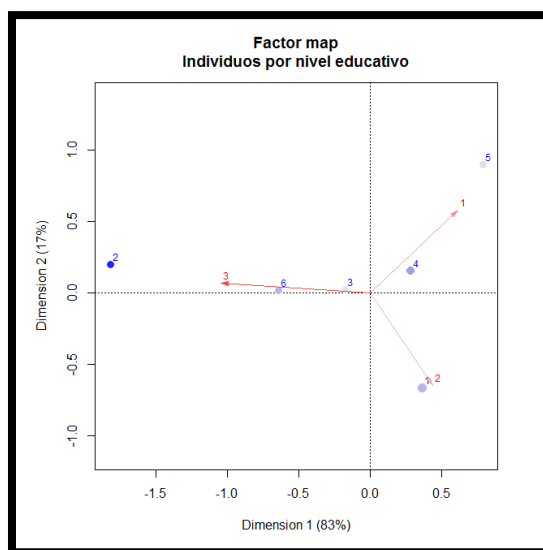
Grupos	Mass	ChiDist	Inertia	Coordenadas	
				Dim. 1	Dim. 2
1	0.289	0.428	0.053	0.483	-1.323
2	0.111	1.362	0.206	-2.443	0.392
3	0.119	0.132	0.002	-0.236	0.050
4	0.244	0.223	0.012	0.376	0.309
5	0.142	0.739	0.077	1.055	1.789
6	0.094	0.480	0.022	-0.863	0.038

Tabla 4.2.4

Nivel Educativo	Mass	ChiDist	Inertia	Coordenadas	
				Dim. 1	Dim. 2
Primaria	0.333	0.540	0.097	0.820	1.152
Secundaria	0.333	0.460	0.071	0.588	-1.286
Terciario	0.333	0.783	0.204	-1.408	0.134

En esta oportunidad, la modalidad que está mejor representada es el grupo 2, el cuál es totalmente dominante, ya que es el que toma el valor más alto en la distancia chi-cuadrado al centroide. En lo que respecta a inercia, el grupo 2 es también el que más aporta. Para las modalidades de la variable nivel educativo, es nivel terciario la que queda mejor representada a la vez que es la que contribuye más a la inercia.

El factor map se presenta a continuación:



A partir del análisis visual se ve que el nivel educativo terciario es el que más contribuye a la conformación del eje 1 mientras que, para el eje2, no resulta tan claro si es la modalidad 1 o 2 de la variable nivel educativo la que más contribuye a su conformación.

En esta oportunidad, el factor map sugiere que hay una clara asociación entre el grupo 1 y el nivel educativo 2 (secundaria). A la vez que los grupos 3, 6, y 2 se asocian más claramente con el nivel educativo 3 (terciario). Es importante destacar la concordancia de lo que revela el factor map con la caracterización de los grupos realizada mediante el análisis de clusters.

GRUPOS - EDAD

Por último, se implementó correspondencia simple con los grupos obtenidos mediante el análisis de cluster, y los tramos de edad. Las tablas de contingencia y de varianzas son presentadas a continuación:

Tabla 4.3.1

		Edad		
		20-29	30-49	50 y más
G r u p o s	1	49	30	25
	2	0	23	17
	3	0	6	37
	4	8	40	40
	5	29	21	1
	6	34	0	0

Tabla 4.3.2

Dim	Var	% Var	% Var acum
1	0.46	82.00	82.00
2	0.10	18.00	100.00

La dimensión 1 capta el 82% de la varianza de los datos proyectados mientras que la dimensión 2 el 18%. Entre las dos, se acumula el total de 100%.

Tabla 4.3.3

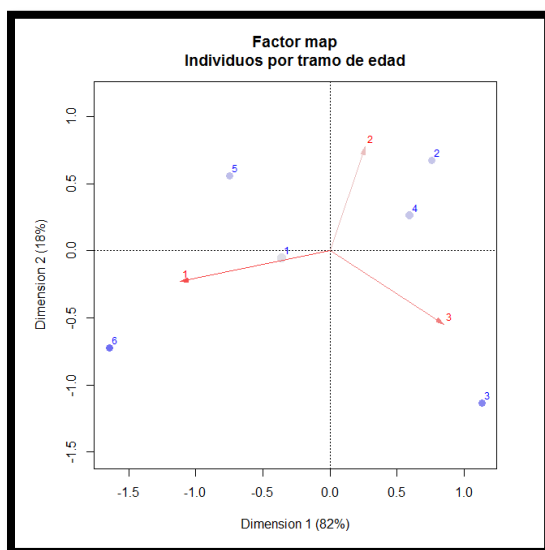
Grupos	Mass	ChiDist	Inertia	Coordenadas	
				Dim. 1	Dim. 2
1	0.289	0.298	0.026	-0.438	-0.089
2	0.111	0.731	0.059	0.921	1.199
3	0.119	1.131	0.153	1.376	-2.020
4	0.244	0.514	0.065	0.727	0.469
5	0.142	0.693	0.068	-0.910	0.995
6	0.094	1.414	0.189	-1.999	-1.288

Tabla 4.3.4

Edad	Mass	ChiDist	Inertia	Coordenadas	
				Dim. 1	Dim. 2
1	0.333	0.926	0.286	-1.354	-0.409
2	0.333	0.488	0.080	0.323	1.377
3	0.333	0.763	0.194	1.031	-0.968

La tabla 4.3.3 evidencia que los grupos 6 y 3 son los que están mejor representados, puesto que son los más próximos al círculo unidad. Asimismo, son los que contribuyen más a la inercia. Por su parte, para los tramos de edad, el tramo de edad 1 es la modalidad que queda mejor representada a la vez que es la que más contribuye a la inercia.

El factor map se presenta a continuación:



Atendiendo al análisis visual del factor map, la modalidad 1 de la variable edad (20-29 años) es la que más contribuye a la conformación del eje 1, y la modalidad 2 (30-49 años) es la que más lo hace a la conformación del eje 2. El efecto de la modalidad 3 en los ejes no parece demasiado claro.

Conclusiones

Del análisis de correspondencias simples puede concluirse que:

- Si bien el grupo 4 fue caracterizado en Cluster como un grupo de observaciones heterogéneas, el presente análisis sugiere que está asociado a la modalidad Montevideo-Mujeres.
- El grupo 1, que había sido caracterizado como de observaciones heterogéneas, parecería guardar una relación con la modalidad nivel educativo 2, mientras que, los grupos 3, 6, y 2 guardan relación con el nivel educativo 3.
- En lo que respecta a las edades, no es posible exponer ninguna conclusión clara en relación a su asociación con los clusters formados.
- En definitiva, los grupos parecen diferenciarse más que nada en función del nivel educativo.

Combinando los resultados encontrados en la aplicación de las dos metodologías (componentes principales y correspondencias simples), puede esbozarse una importante conclusión sobre la estructura del mercado laboral uruguayo, la cual hace referencia a la movilidad que proporciona la educación en el ámbito laboral.

Parecería ser que la calidad de la inserción laboral de los individuos está estrechamente relacionada en primera instancia con el nivel educativo alcanzado en sus vidas y, en menor medida, su edad (entre los individuos con niveles educativos altos, son los jóvenes los que parecen tener una inserción laboral un poco más endeble).

Bibliografía y referencias

- Blanco (2006) - Introducción al análisis multivariado - 1era edición
- Dueñas Rodríguez - Modelos de respuesta multinomial con R y aplicación con datos reales (Tesis de maestría)
- Everitt & Hothorn (2014) - A handbook of statistical analyses using R - 3rd edition
- Peña (2002) - Análisis de datos multivariantes
- Rencher (2002) - Methods of multivariate analysis - 2nd edition

Anexos

ANEXO 1: CÓDIGO R

```
#####  
##### SEGUNDO INFORME ANÁLISIS MULTIVARIADO I 2016 #####  
#####  
  
rm(list=ls())  
setwd("C:/Users/dacza_000/Dropbox/1er informe")  
  
#### Cargo librerías y funciones ####  
  
library(cluster); library(lattice); library(mclust)  
library(fpc); library(MVN); library(MASS)  
library(dplyr); library(nnet); library(Metrics)  
library(bitops); library(ggplot2); library(RCurl)  
library(caret); library(boot); library(StatMatch)  
library(car); library(ca); library(FactoMineR)  
library(stats); library(mitools); library(WritexLS)  
library(Rcmdr)  
  
source("panelcor.R")  
source("panelhist.R")  
source("indicadores.R")  
source("shadowtext.R")  
  
#$#####  
#### PREPARACIÓN DE LA BASE ####  
#####  
  
#### Exploración inicial ####  
  
ech = read.table("ech95-04.txt", header = TRUE, sep = "")  
  
# La base es un pseudo panel de la ech para el período 1995 a  
2004.  
# Para cada año se muestra la variable resumen (nombre) para  
cada uno de los grupos construidos  
  
# No existen NA en la base  
table(complete.cases(ech))  
  
# Primera vista de las variables  
str(ech)
```

```
# Cambio el nombre de las variables: AÑO -> year, TAMAÑO ->
size, y PÚBLICO -> public
ech = rename(ech, year = AÑO, size = TAMAÑO, publico =
PÚBLICO)

# Cambio los nombres de las variables en uppercase por
lowercase
names(ech) <- tolower(names(ech))

#### Creación de dummies ####

# Creo variables binarias para cada uno de los atributos de
la variable "nombre"

ech[, "educ"] <- substr(ech$nombre, 4, 4)
ech <- select(ech, educ, everything())
ech$educ <- as.factor(gsub(",", "", as.character(ech$educ)))

ech[, "edad"] <- substr(ech$nombre, 3, 3)
ech <- select(ech, edad, everything())
ech$edad <- as.factor(gsub(",", "", as.character(ech$edad)))

ech[, "sexo"] <- substr(ech$nombre, 2, 2)
ech <- select(ech, sexo, everything())
ech$sexo <- as.factor(gsub(",", "", as.character(ech$sexo)))

ech[, "loc"] <- substr(ech$nombre, 1, 1)
ech <- select(ech, loc, everything())
ech$loc <- as.factor(gsub(",", "", as.character(ech$loc)))

# Muevo la variable year a la primera posición
ech <- select(ech, year, everything())

# Muevo la variable udad_anio a la primera posición
ech <- select(ech, udad_anio, everything())

# Muevo la variable unidad a la primera posición
ech <- select(ech, unidad, everything())

# Creo la variable nombre2 (=nombre sin localidad)

ech[, "nombre2"] <- substr(ech$nombre, 2, 4)
ech$nombre2 <- as.factor(gsub(",", "", as.character(ech$nombre2)))

# Creo la variable locsex
```

```
ech[, "locsex"] <- substr(ech$nombre, 1, 2)
ech$locsex <- as.factor(gsub(",", "",
as.character(ech$locsex)))

#### Generación de submatrices e inversas ####

# Genero la base ech2, la cual contiene la variable nombre +
vars explicativas

ech2 <- select(ech, (nombre:ingreso))

# Genero la base ech3, la cual contiene únicamente las vars
explicativas

ech3 <- select(ech2, -nombre)

# Genero la base ech_redu con las variables de situación de
empleo

ech_redu <- select(ech3, desemp, tparcial, multiemp, subemp,
precario)

# Genero la base ech_res con el resto de las variables

ech_res <- select(ech, -(1:8),-10,-12,-13,-25,-26, -28, -29)

# Inversas de las matrices

echinv <- t.data.frame(ech)
ech2inv <- t.data.frame(ech2)
ech3inv <- t.data.frame(ech3)

#### Paletas de colores ####

colores6=c("#7fc97f", "#beaed4", "#fdc086", "#ffff99",
"#386cb0", "#f0027f")
sexcolor <- c("lightskyblue", "lightskyblue", "lightskyblue",
"lightskyblue", "lightskyblue", "lightskyblue",
"lightskyblue", "lightskyblue", "lightskyblue", "palegreen3",
"palegreen3", "palegreen3", "palegreen3", "palegreen3",
"palegreen3", "palegreen3", "palegreen3", "palegreen3",
"lightskyblue", "lightskyblue", "lightskyblue",
"lightskyblue", "lightskyblue", "lightskyblue",
"lightskyblue", "lightskyblue", "lightskyblue", "palegreen3",
"palegreen3", "palegreen3", "palegreen3", "palegreen3",
"palegreen3", "palegreen3", "palegreen3", "palegreen3")
```

```
#####  
#### Clusters (dist: Mahalanobis) ####  
#####  
  
#### Formación de Clusters ####  
  
dist_mah <- mahalanobis.dist(ech_redu)  
agnes_mah_redu_wa <- agnes(dist_mah, diss=TRUE,  
method="ward")  
  
clusters <- factor(cutree(agnes_mah_redu_wa, k=6))  
table(clusters)  
  
ech <- cbind(ech, clusters)  
write.csv(ech, "tablas/ech con grupos.csv")  
  
#####  
#### Análisis Discriminante ####  
#####  
  
#### Construcción del modelo ####  
  
rmlog <- multinom(ech$clusters ~ ., data=ech_redu,  
maxit=1000, Hess=TRUE, model=TRUE, wald=TRUE)  
  
# Prueba stepwise de significación global (criterios BIC y  
AIC)  
  
stepwise(rmlog, criterion="AIC")  
stepwise(rmlog, criterion="BIC")  
  
# Devuelve los coeficientes estimados por ecuación, los  
errores estandard, y  
# los valores del estadístico de ward para cada regresor en  
cada ecuación  
w1 <- summary(rmlog, wald=TRUE)  
  
# Armo una matrix con el valor absoluto de los estadísticos  
de ward para cada  
# regresor (por filas), en cada ecuación (por columnas)  
w1_abs <- abs(w1$wald.ratios)  
  
# calculo p valores para la prueba de significación  
individual a dos colas  
  
p1 <- round(pnorm(w1_abs, lower.tail=FALSE), 3)
```

```
write.csv(p1, "tablas/pvalores modelo 1.csv")

# Las variables no significativas para los grupos indican que
# esa variable no
# contribuye a explicar la formación de ese grupo

# Devuelve los valores ajustados para la muestra
fitted(rmlog)

# cálculo de los valores predichos
probspredict <- predict(rmlog, newdata=ech_redu, "probs")
grupospredict <- predict(rmlog, newdata=ech_redu, "class")

# Predicciones correctas

postResample(obs=ech$clusters, pred=grupospredict)

x1 <- table(ech$clusters, grupospredict)

write.csv(x1, "tablas/grupospredict.csv")

# Agrego los grupos a la base

ech <- cbind(ech, grupospredict)

#####
##### Discriminante Multinomial (vars explicativas) #####
#####

##### Construcción del modelo #####

ech_wa <- select(ech, grupospredict, jefe, size, privado,
publico, cpsl, cpcl, profytec, oficina, manual, indust,
comercio, sfinan, sperson, ingreso)

# Estimamos el modelo multi logístico

rmlog2 <- multinom(grupospredict ~ ., data=ech_wa,
maxit=1000)

# Inferencia

stepwise(rmlog2, criterion="AIC") # Según criterio AIC se
debe estimar rmlog3
```

```
rmlog3 <- multinom(formula = grupospredict ~ jefe + size +
privado + cpsl + cpc1 + profytec + oficina + manual +
comercio + sfinan + sperson + ingreso, data = ech_wa, maxit =
1000)

stepwise(rmlog3, criterion="AIC")

w3 <- summary(rmlog3, wald=TRUE)
w3_abs <- abs(w3$wald.ratios)
p3 <- round(pnorm(w3_abs, lower.tail=FALSE), 3)

p3<0.05

# Devuelve los coeficientes estimados del modelo tomando como
referencia al grupo IH11
coef(rmlog3)

# Devuelve los valores ajustados para la muestra
fitted(rmlog3)

# Cálculo de los valores predichos
predict3 <- predict(rmlog3, newdata=ech3, "probs")
predict4 <- predict(rmlog3, newdata=ech3, "class")

a <- NULL

for(i in 1:360){a[i] <- max(predict3[i,])}

print(a)

# Predicciones correctas

postResample(obs=ech_wa$grupospredict, pred=predict4)

table(ech_wa$grupospredict, predict4)

ech_predict4 <- cbind(ech_wa, predict4)

#####
#### Componentes principales ####
#####

#### ACP ####

for(i in 1:360){row.names(ech)[i] <-
as.character(ech$udad_anio[[i]])}
```

```
# x11(15,15)
# par(mfrow=c(1,2), oma=c(0,1,0,1))
a <- PCA(ech3, graph=FALSE, ncp=2)

summary(a)

#### Gráficos con individuos y grupos ####

x1 <- a$ind$coord[,1]; x2 <- a$var$coord[,1]
y1 <- a$ind$coord[,2]; y2 <- a$var$coord[,2]

normas <- NULL
for(i in 1:length(x1)){normas[i]=sqrt((x1[i])^2+(y1[i])^2)}

wx <- NULL; wy <- NULL; t=0.2; z=max(normas); theta <- seq(0,
2*pi, length=200)

# Factor map solo variables
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg="white", ylim=c(-
6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map - ",
round(a$eig[2,3],2), "%", "\nSolo variables"), axes=FALSE)
axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
abline(h=0, lty=2)
abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)

# Factor map según localidad
x11(15,15)
par(oma=c(0,1,0,1))
```



```

plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$loc),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map - ",
round(a$eig[2,3],2), "%", "\nIndividuos por localidad"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Localidades",
c("Interior","Montevideo"), pch=c(21,21), col=c(1,2),
pt.bg=c(1,2), pt.lwd=0, bty="n")

```

```
# Factor map según sexo
```

```

x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$sexo),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map - ",
round(a$eig[2,3],2), "%", "\nIndividuos por sexo"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}

```

```

for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Sexo", c("Hombres","Mujeres"),
pch=c(21,21), col=c(1,2), pt.bg=c(1,2), pt.lwd=0, bty="n")

# Factor map según educ
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$educ),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map - ",
round(a$eig[2,3],2), "%", "\nIndividuos por nivel
educativo"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Nivel\nEducativo",
c("Primaria","Secundaria", "Terciaria"), pch=c(21,21, 21),
col=c(1,2,3), pt.bg=c(1,2,3), pt.lwd=0, bty="n")

# Factor map según edad
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$edad),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map - ",
round(a$eig[2,3],2), "%", "\nIndividuos por tramo de edad"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))

```

```

      axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      abline(h=0, lty=2)
      abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Tramo\nde Edad", c("20-29","30-
49", "50 o más"), pch=c(21,21, 21), col=c(1,2,3),
pt.bg=c(1,2,3), pt.lwd=0, bty="n")

# Factor map según localidad y sexo
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$locsex), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map - ", round(a$eig[2,3],2), "%",
"\nIndividuos por localidad y sexo"), axes=FALSE)
      axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      abline(h=0, lty=2)
      abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Individuos", c("IH","IM", "MM",
"MH"), pch=c(21,21,21,21), col=c(1,2,4,3), pt.bg=c(1,2,4,3),
pt.lwd=0, bty="n")

```

```
# Factor map según grupos
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$grupospredict), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map - ", round(a$eig[2,3],2), "%",
"\nIndividuos por grpos"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Clusters",
c("1","2","3","4","5","6"), pch=c(21,21,21,21,21,21),
col=c(1,2,3,4,5,6), pt.bg=c(1,2,3,4,5,6), pt.lwd=0, ncol=2,
bty="n")
```

```
# Factor map según año
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=(as.numeric(ech$year)-
1994), ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1
- ", round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map - ",
round(a$eig[2,3],2), "%", "\nIndividuos por año"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
```

```

for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Años",
c("95'", "96'", "97'", "98'", "99'", "00'", "01'", "02'", "03'", "04'"
), pch=c(21,21,21,21,21,21,21,21,21,21),
col=c(1,2,3,4,5,6,7,8,9,10), pt.bg=c(1,2,3,4,5,6,7,8,9,10),
ncol=2, pt.lwd=0, bty="n")

```

```

# Factor map según nombre2
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$nombre2), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map - ", round(a$eig[2,3],2), "%",
"\nIndividuos sin localidad"), axes=FALSE)
axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
abline(h=0, lty=2)
abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
nombres=unique(as.character(ech$nombre2))
legend("bottomleft", title="Individuos", nombres,
col=as.numeric(unique(ech$nombre2)), pch=21,
pt.bg=as.numeric(unique(ech$nombre2)), ncol=2, pt.lwd=0,
bg="white", bty="n")

```

```

#$#####
#### ACP reducido ####

```

```
#####
```

```
#### ACP reducido ####
```

```
for(i in 1:360){row.names(ech_redu)[i] <-
as.character(ech$udad_anio[[i]])}
```

```
# x11(15,15)
# par(mfrow=c(1,2), oma=c(0,1,0,1))
a <- PCA(ech_redu, graph=FALSE, ncp=2)
```

```
summary(a)
```

```
#### Gráficos con individuos y grupos ####
```

```
x1 <- a$ind$coord[,1]; x2 <- a$var$coord[,1]
y1 <- a$ind$coord[,2]; y2 <- a$var$coord[,2]
```

```
normas <- NULL
for(i in 1:length(x1)){normas[i]=sqrt((x1[i])^2+(y1[i])^2)}
```

```
wx <- NULL; wy <- NULL; t=0.2; z=max(normas); theta <- seq(0,
2*pi, length=200)
```

```
# Factor map solo variables
```

```
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg="white", ylim=c(-
6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Reducido -
", round(a$eig[2,3],2), "%", "\nSolo variables"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
```

```

for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
# Factor map reducido según localidad
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$loc),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Reducido -
", round(a$eig[2,3],2), "%", "\nIndividuos por localidad"),
axes=FALSE)
axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
abline(h=0, lty=2)
abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Localidades",
c("Interior","Montevideo"), pch=c(21,21), col=c(1,2),
pt.bg=c(1,2), pt.lwd=0, bty="n")

```

```

# Factor map reducido según sexo
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$sexo),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Reducido -
", round(a$eig[2,3],2), "%", "\nIndividuos por sexo"),
axes=FALSE)
axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
abline(h=0, lty=2)

```

```

      abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Sexo", c("Hombres","Mujeres"),
pch=c(21,21), col=c(1,2), pt.bg=c(1,2), pt.lwd=0, bty="n")

# Factor map reducido según educ
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$educ),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Reducido -
", round(a$eig[2,3],2), "%", "\nIndividuos por nivel
educativo"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Nivel\nEducativo",
c("Primaria","Secundaria", "Terciaria"), pch=c(21,21, 21),
col=c(1,2,3), pt.bg=c(1,2,3), pt.lwd=0, bty="n")

# Factor map reducido según edad
x11(15,15)
par(oma=c(0,1,0,1))

```



```

plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$edad),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Reducido -
", round(a$eig[2,3],2), "%", "\nIndividuos por tramo de
edad"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Tramo\nde Edad", c("20-29","30-
49", "50 o más"), pch=c(21,21, 21), col=c(1,2,3),
pt.bg=c(1,2,3), pt.lwd=0, bty="n")

```

Factor map reducido según localidad y sexo

```

x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$locsex), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map Reducido - ", round(a$eig[2,3],2),
"%", "\nIndividuos por localidad y sexo"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}

```

```

for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Individuos", c("IH","IM", "MM",
"MH"), pch=c(21,21,21,21), col=c(1,2,4,3), pt.bg=c(1,2,4,3),
pt.lwd=0, bty="n")

```

```

# Factor map reducido según grupos
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$grupospredict), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map Reducido - ", round(a$eig[2,3],2),
"%", "\nIndividuos por grpos"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Clusters",
c("1","2","3","4","5","6"), pch=c(21,21,21,21,21,21),
col=c(1,2,3,4,5,6), pt.bg=c(1,2,3,4,5,6), pt.lwd=0, ncol=2,
bty="n")

```

```

# Factor map reducido según año
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=(as.numeric(ech$year)-
1994), ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1
- ", round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Reducido -
", round(a$eig[2,3],2), "%", "\nIndividuos por año"),
axes=FALSE)

```

```

      axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      abline(h=0, lty=2)
      abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Años",
c("95'", "96'", "97'", "98'", "99'", "00'", "01'", "02'", "03'", "04'"),
pch=c(21,21,21,21,21,21,21,21,21,21),
col=c(1,2,3,4,5,6,7,8,9,10), pt.bg=c(1,2,3,4,5,6,7,8,9,10),
ncol=2, pt.lwd=0, bty="n")

# Factor map reducido según nombre2
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$nombre2), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map Reducido - ", round(a$eig[2,3],2),
"%", "\nIndividuos sin localidad"), axes=FALSE)
      axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
      abline(h=0, lty=2)
      abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)

```

```

nombres=unique(as.character(ech$nombre2))
legend("bottomleft", title="Individuos", nombres,
col=as.numeric(unique(ech$nombre2)), pch=21,
pt.bg=as.numeric(unique(ech$nombre2)), ncol=2, pt.lwd=0,
bg="white", bty="n")

####$#####
#### ACP resto de las variables ####
#####

### ACP resto de las variables ###

for(i in 1:360){row.names(ech_res)[i] <-
as.character(ech$udad_anio[[i]])}

# x11(15,15)
# par(mfrow=c(1,2), oma=c(0,1,0,1))
a <- PCA(ech_res, graph=FALSE, ncp=2)

summary(a)

#### Gráficos con individuos y grupos ####

x1 <- a$ind$coord[,1]; x2 <- a$var$coord[,1]
y1 <- a$ind$coord[,2]; y2 <- a$var$coord[,2]

normas <- NULL
for(i in 1:length(x1)){normas[i]=sqrt((x1[i])^2+(y1[i])^2)}

wx <- NULL; wy <- NULL; t=0.2; z=max(normas); theta <- seq(0,
2*pi, length=200)

# Factor map Resto solo variables
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg="white", ylim=c(-
6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Resto - ",
round(a$eig[2,3],2), "%", "\nSolo variables"), axes=FALSE)
axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
abline(h=0, lty=2)
abline(v=0, lty=2)

```

```

for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)

# Factor map Resto según localidad
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$loc),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Resto - ",
round(a$eig[2,3],2), "%", "\nIndividuos por localidad"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Localidades",
c("Interior","Montevideo"), pch=c(21,21), col=c(1,2),
pt.bg=c(1,2), pt.lwd=0, bty="n")

# Factor map resto según sexo
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$sexo),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Resto - ",

```

```

round(a$eig[2,3],2), "%", "\nIndividuos por sexo"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-1", "-0.5", "0", "0.5", "1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-1", "-0.5", "0", "0.5", "1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Sexo", c("Hombres", "Mujeres"),
pch=c(21,21), col=c(1,2), pt.bg=c(1,2), pt.lwd=0, bty="n")

# Factor map resto según educ
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$educ),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Resto - ",
round(a$eig[2,3],2), "%", "\nIndividuos por nivel
educativo"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-1", "-0.5", "0", "0.5", "1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-1", "-0.5", "0", "0.5", "1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)

```

```

legend("bottomleft", title="Nivel\nEducativo",
c("Primaria","Secundaria", "Terciaria"), pch=c(21,21, 21),
col=c(1,2,3), pt.bg=c(1,2,3), pt.lwd=0, bty="n")

# Factor map resto según edad
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=as.numeric(ech$edad),
ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Resto - ",
round(a$eig[2,3],2), "%", "\nIndividuos por tramo de edad"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Tramo\nde Edad", c("20-29","30-
49", "50 o más"), pch=c(21,21, 21), col=c(1,2,3),
pt.bg=c(1,2,3), pt.lwd=0, bty="n")

# Factor map resto según localidad y sexo
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$locsex), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map Resto - ", round(a$eig[2,3],2), "%",
"\nIndividuos por localidad y sexo"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)

```

```

    abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Individuos", c("IH","IM", "MM",
"MH"), pch=c(21,21,21,21), col=c(1,2,4,3), pt.bg=c(1,2,4,3),
pt.lwd=0, bty="n")

```

```

# Factor map resto según grupos
x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$grupospredict), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map Resto - ", round(a$eig[2,3],2), "%",
"\nIndividuos por grpos"), axes=FALSE)
    axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
    axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
    abline(h=0, lty=2)
    abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Clusters",
c("1","2","3","4","5","6"), pch=c(21,21,21,21,21,21),
col=c(1,2,3,4,5,6), pt.bg=c(1,2,3,4,5,6), pt.lwd=0, ncol=2,
bty="n")

```

```

# Factor map resto según año
x11(15,15)

```



```

par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL, bg=(as.numeric(ech$year)-
1994), ylim=c(-6.5,6.5), xlim=c(-6.5,6.5), xlab=paste0("Dim 1 - ",
round(a$eig[1,2],2), "%"), ylab=paste0("Dim 1 - ",
round(a$eig[2,2],2), "%"), main=paste0("Factor Map Resto - ",
round(a$eig[2,3],2), "%", "\nIndividuos por año"),
axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
legend("bottomleft", title="Años",
c("95'", "96'", "97'", "98'", "99'", "00'", "01'", "02'", "03'", "04'"),
pch=c(21,21,21,21,21,21,21,21,21,21),
col=c(1,2,3,4,5,6,7,8,9,10), pt.bg=c(1,2,3,4,5,6,7,8,9,10),
ncol=2, pt.lwd=0, bty="n")

```

```
# Factor map resto según nombre2
```

```

x11(15,15)
par(oma=c(0,1,0,1))
plot(a$ind$coord, pch=21, col=NULL,
bg=as.numeric(ech$nombre2), ylim=c(-6.5,6.5), xlim=c(-
6.5,6.5), xlab=paste0("Dim 1 - ", round(a$eig[1,2],2), "%"),
ylab=paste0("Dim 1 - ", round(a$eig[2,2],2), "%"),
main=paste0("Factor Map Resto - ", round(a$eig[2,3],2), "%",
"\nIndividuos sin localidad"), axes=FALSE)
  axis(side=1, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  axis(side=2, at=c(-z,-(z/2),0,(z/2),z), labels=c("-
1","-0.5","0","0.5","1"))
  abline(h=0, lty=2)
  abline(v=0, lty=2)
for(h in 1:length(x2)){arrows(x0=0, y0=0, x1=x2[h]*z,
y1=y2[h]*z, lwd=2, col="blue")}

```

```
for(i in 1:length(x2)){if(x2[i]<0){wx[i]=x2[i]*z-t}
else{wx[i]=x2[i]*z+t}}
for(j in 1:length(y2)){if(y2[j]<0){wy[j]=y2[j]*z-t}
else{wy[j]=y2[j]*z+t}}
for(k in 1:length(wx)){shadowtext(x=wx[k], y=wy[k],
labels=row.names(a$var$coord)[k], col="blue", bg="white",
cex=1.3, r=0.1)}
lines(x=z*cos(theta), y=z*sin(theta), lty=2)
nombres=unique(as.character(ech$nombre2))
legend("bottomleft", title="Individuos", nombres,
col=as.numeric(unique(ech$nombre2)), pch=21,
pt.bg=as.numeric(unique(ech$nombre2)), ncol=2, pt.lwd=0,
bg="white", bty="n")
```

```
#####
#### Correspondencias Simples ####
#####
```

```
#### Independencia ####
```

```
chisq.test(ech$grupospredict, ech$loc) # rechazo
independencia
chisq.test(ech$grupospredict, ech$sexo) # rechazo
independencia
chisq.test(ech$grupospredict, ech$educ) # rechazo
independencia
chisq.test(ech$grupospredict, ech$edad) # rechazo
independencia
chisq.test(ech$grupospredict, ech$locsex) # rechazo
independencia
```

```
#### Tablas de contingencia ####
```

```
table(ech$grupospredict, ech$locsex)
table(ech$grupospredict, ech$educ)
table(ech$grupospredict, ech$edad)
```

```
#### ACS por localidad y sexo ####
```

```
conti1 <- with(ech, table(ech$grupospredict,ech$locsex))
```

```
round(prop.table(conti1, 1),4)
round(prop.table(conti1, 2),4)
```

```
acs1 <- ca(conti1)
acs1
summary(acs1)
```

```
x11(15,15)
par(oma=c(0,1,0,1))
plot(acs1, mass=TRUE, map="sympbiplot", contrib="absolute",
arrows=c(FALSE, TRUE), main="Factor map\nIndividuos por
localidad y sexo")

#### ACS por nivel educativo ####

conti2 <- with(ech, table(ech$grupospredict,ech$educ))

round(prop.table(conti2, 1),4)
round(prop.table(conti2, 2),4)

acs2 <- ca(conti2)
acs2
summary(acs2)

x11(15,15)
par(oma=c(0,1,0,1))
plot(acs2, mass=TRUE, map="sympbiplot", contrib="absolute",
arrows=c(FALSE, TRUE), main="Factor map\nIndividuos por nivel
educativo")

#### ACS por tramo de edad ####

conti3 <- with(ech, table(ech$grupospredict,ech$edad))

round(prop.table(conti3, 1),4)
round(prop.table(conti3, 2),4)

acs3 <- ca(conti3)
acs3
summary(acs3)

x11(15,15)
par(oma=c(0,1,0,1))
plot(acs3, mass=TRUE, map="sympbiplot", contrib="absolute",
arrows=c(FALSE, TRUE), main="Factor map\nIndividuos por tramo
de edad")

#$#####
### FIN DE LA PROGRAMACIÓN ###
#####
```