**Análisis de Componentes Principales (ACP)**

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.

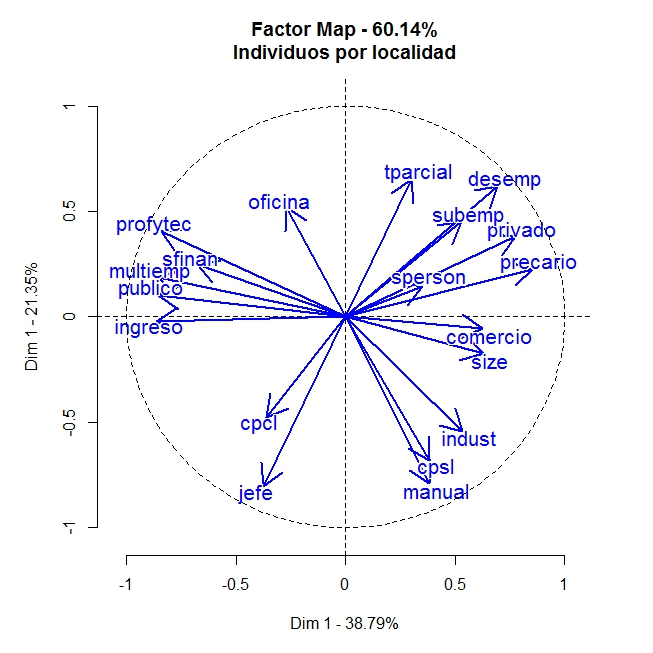
**Tabla 1.1**



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 2.2**



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. Es decir, para la variable *jefe*, la correlación con la dimensión 1 es 1.93, mientras que la correlación con la dimensión dos es 15.94. Razonando de la misma manera con el resto de las variables, es posible concluir que las variables:

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

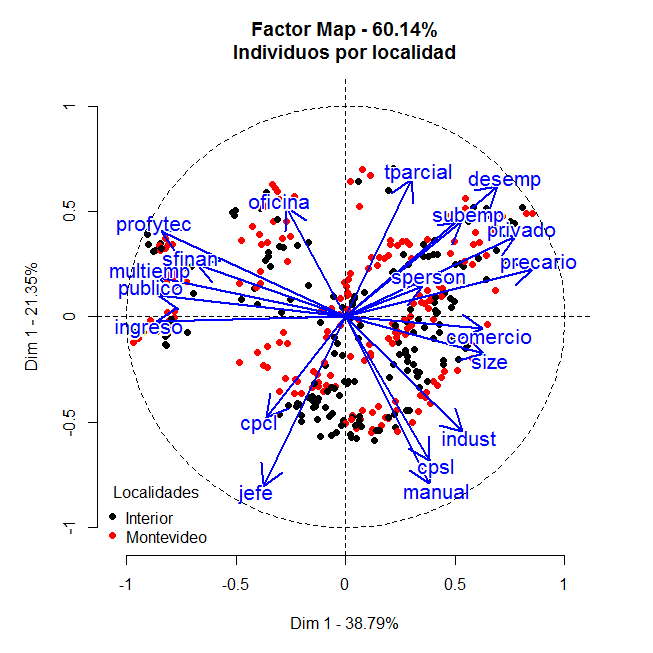
son las que más contribuyen a la conformación del eje 1. Mientras que las variables:

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

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

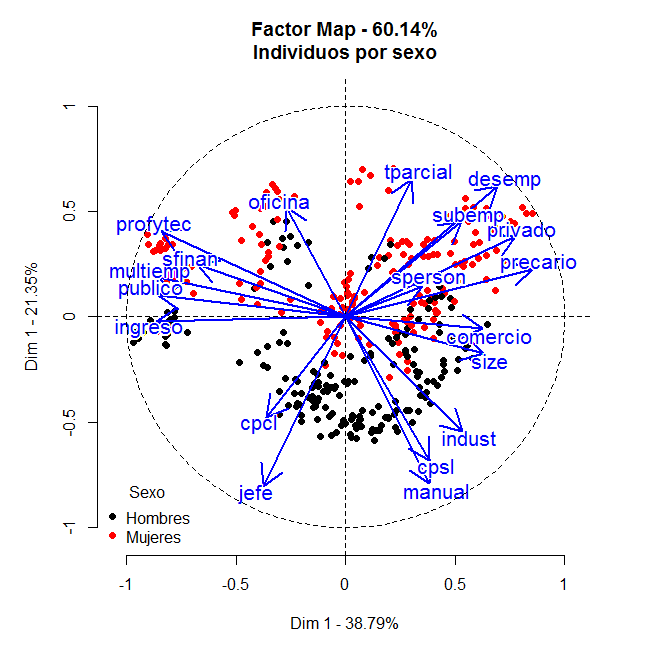
A modo de enriquecer la comprensión de la estructura subyacente en la base de datos y, abogando por un análisis dual, resulta interesante proyectar a los individuos de la base sobre el factor map, en función de las distintas etiquetas: *localidad*, *sexo*, *edad*, *nivel educativo*, *clusters*, y *tiempo*. 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 resultad 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.

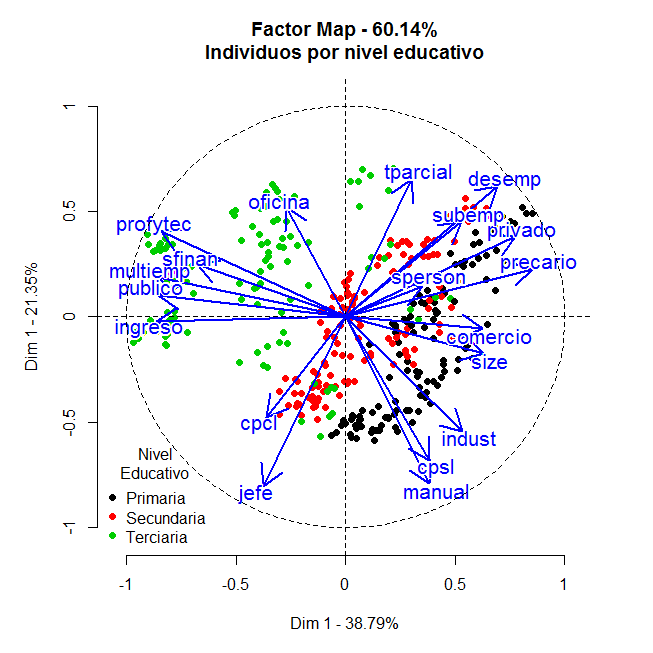
**Sexo**

****

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.

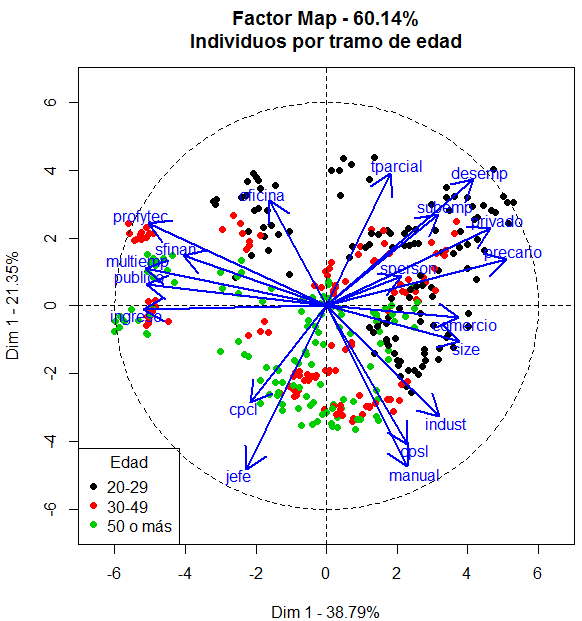
Los hombres, por su parte, son los que tienen mayores ingresos y son en su mayoría cuenta propistas con y sin local, jefes de hogar, trabajadores manuales, trabajadores en la industria.

**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.

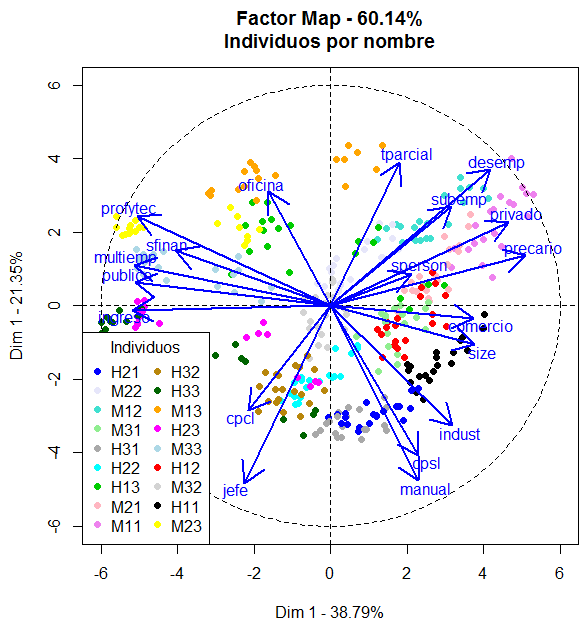
**Edad**



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 lo que alcanzan niveles más altos.

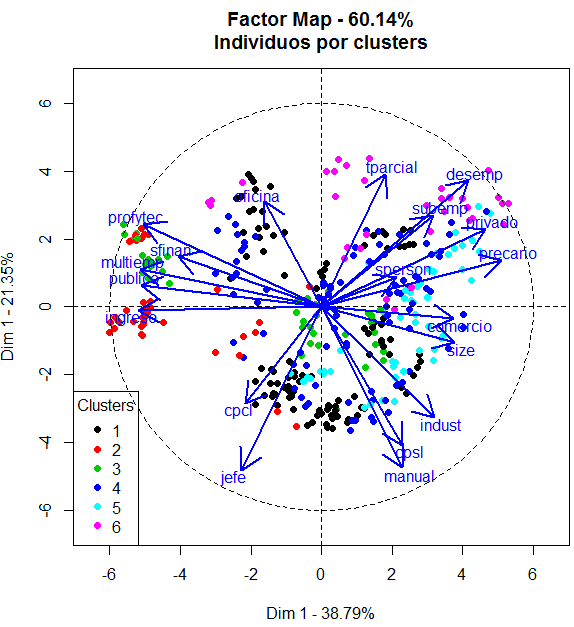
Se detecta también que la precariedad, el subempleo, desempleo, y tiempo parcial abunda entre los individuos de entre 20 y 29 años.

**Sexo – Nivel educativo – Edad**



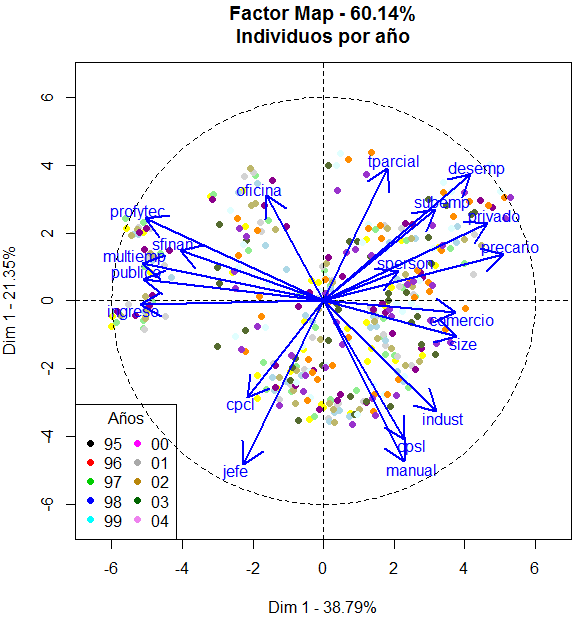
A modo de combinar los análisis anteriores respecto a separación de los individuos en las variables según sexo, edad, y nivel educativo, y con el fin de tener una aproximación multidimensional (en el sentido de combinar etiquetas), el plano factorial sugiere que, como habíamos encontrado, los profesionales 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 abunda entre las mujeres de entre 20 y 49 años con nivel educativo primario.

**Grupos (Cluster)**



Proyectando los individuos según el grupo al que pertenecen en cluster vemos claramente que los profesionales y técnicos, multiempleados, y empleados públicos son en su mayoría individuos pertenencientes 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 entre el grupo 6, conformado en su mayoría por mujeres jóvenes (de entre 20 y 29 años) con nivel educativo terciario.

**Individuos por año**



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

**Componentes principales sólo con variables 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 deseadas para 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. Tparcial está asociada a estudiantes en nivel terciario (eligen trabajar menos de 8 horas), mientras que, como revelaron las estadísticas descriptivas y los análisis anteriores, en la variable multiemp se concentran predominantemente profesionales y técnicos que a su vez son los que perciben mayores ingresos y que 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**

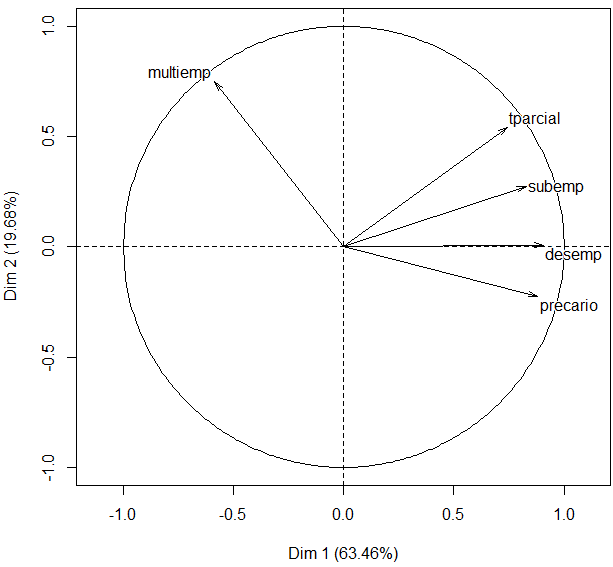


A partir de la tabla 2.2 y el análisis visual del factor map es posible constatar que lo que surgió fue que las variables que más contribuyen a la conformación del eje 1 son aquellas variables que representan una situación del individuo en la cual no tiene opción de decidir en lo que respecta a su inserción laboral, es decir: desempleo, precario, y subempleo. Las mismas, como se mencionó anteriormente, representan situaciones que no son deseables y que no son elegidas por los individuos. Para la dimensión dos, aparecen las variables en las cuales el individuo si tiene opción de decidir: tparcial y multiemp.

**Tabla 2.2**



**Factor map**



Recurriendo al análisis fila de manera similar a la realizada 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 mas se diferencias los individuos. Vemos que los individuos con multiempleo son los de nivel educativo terciario, mientras que los precarios, subempleados, y desempleados, son en su predominancia invididuos de nivel educativo primaria y secundaria. 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 del grupo 2, donde predominan universitarios mayores de 29 años, y el grupo 3, compuesto en su totalidad por individuos de nivel educativo terciario, con predominancia de mas 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 individos según el año, de nuevo no podemos extraer ninguna conclusión clara y, analizando según sexo, edad, y nivel educativo en su conjunto, 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) al igual que bastante heterogéneos.

**Componentes principales con el resto de las variables**

Al igual que cuando se aplicó el método sobre la totalidad de las variables de la base, la inercia acumulada por las primeras dos dimensiones es solamente de un 60.99%.

**Tabla 3.1**

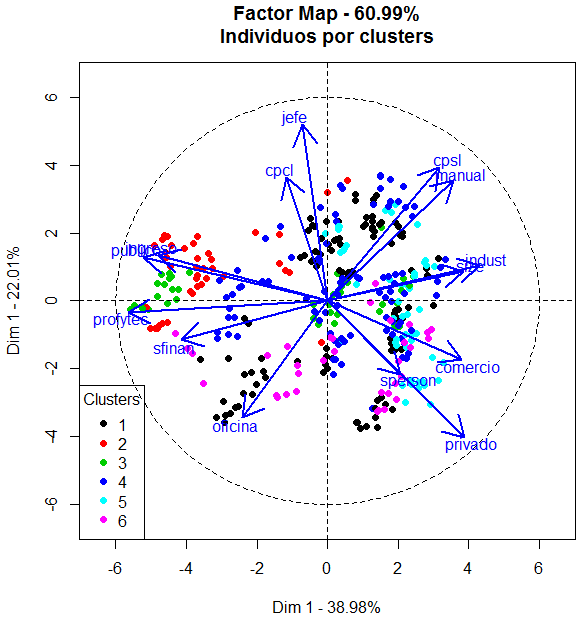


En la tabla 3.2 es posible apreciar que son ahora las variables profytec y público; y jefe y privado, las que mas contribuyen a la conformación de los ejes 1 y 2 respectivamente . Si hacemos un pararelismo con la implementación del método para las 5 variables de calidad de inserción laboral, en esta oportunidad tambié observamos que la primera componente separa entre aquellos con una mejor insersión laboral y aquellos con una mas endeble.

La variable profytec tiene una correlación alta con la variable multiempleo e ingreso, por lo que parecería que la información revelada es de alguna manera similar.

**Tabla 3.2**





Proyectando los individuos sobre el grupo al que pertenecen en Cluster vemos que los grupos 2 y 3 son en su predominancia individuos empleados en el sector público, con ingresos altos. Asimismo,los profesionales y técnicos son en su mayoría individuos pertenencientes 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 e las variables, por lo cuál no es posible sacar conclusiones claras.

**Factor Maps de ACP con las variables de calidad de inserción laboral (ANEXO)**

