



Universidad Complutense de Madrid

## **Análisis del comprador de seguros de vida en el periodo 2008-2013: análisis de los seguros de riesgo y ahorro mediante técnicas de data mining**

Madrid 30 de septiembre de 2014



UNIVERSIDAD  
COMPLUTENSE  
MADRID

Memoria del Contrato de Investigación:

***Análisis del comprador de seguros de vida en el periodo 2008-2013: análisis de los seguros de riesgo y ahorro mediante técnicas de data mining***

**Autores**

Lorenzo Escot  
José Andrés Fernández Cornejo  
Javier Portela

Elaborado para *Investigación Cooperativa entre Entidades Aseguradoras y Fondos de Pensiones -ICEA-*.



Madrid a 30 de septiembre de 2014

## **Análisis del comprador de seguros de vida en el periodo 2008-2013: análisis de los seguros de riesgo y ahorro mediante técnicas de data mining**

### **Contenido**

1	Introducción .....	1
2	Estudio comparativo de la evolución temporal de las pólizas en los años de crisis, según modalidad. Relación con segmentos poblacionales de edad y profesión. ....	3
3	Evolución en los diferentes tipos de localidad por tamaño de población .....	15
4	Patrones en las Entidades presentes en la base de datos.....	17
5	Patrones de edad-profesión en las diferentes modalidades.....	19
6	Patrones en los perfiles de las pólizas por prima, duración y capital, segregando por modalidad. ....	21
7	Conclusiones.....	30
	ANEXOS.....	33

# 1 Introducción

El presente estudio tiene como objetivo profundizar en los estudios previos sobre el comprador de seguros a partir de la información suministrada por ICEA, ampliando la categorización de los mismos según el tipo del seguro contratado (riesgo o ahorro) y aplicando las técnicas de minería de datos una vez incorporada la información disponible para el periodo 2013. En particular se analizará la muestra anonimizada de pólizas de seguros de vida para buscar y analizar segmentos de consumidores y tipologías de pólizas de especial interés.

La evolución del número de pólizas en general en el periodo de crisis (2008-2013) sigue una pauta ascendente que alcanza un máximo en 2010 y es descendente a partir de entonces. Sin embargo, una curva agregada de totales puede esconder subsegmentos de población para los cuales la evolución puede ser radicalmente diferente, incluso contradictoria (ascendente en lugar de descendente) con la curva general.

Con la información disponible se pretende en este informe mostrar cómo la forma de la curva de evolución en el periodo de crisis 2008-2013 del número de pólizas depende de ciertas características como son el ***tipo de póliza según modalidad***, de las ***características socioeconómicas como la edad y la profesión del asegurado***, y también en cuanto a la ***terna prima, duración y capital***.

Se realizarán varios análisis por separado para indagar en los ***perfiles de evolución*** y también en las ***características estáticas de contratación de pólizas en cuanto a edad y profesión***. Además de esta introducción el estudio se divide en 5 secciones o capítulos.

En las secciones 2 y 3 se estudiará la evolución del número de pólizas por modalidad antes y durante la crisis de:

- las diferentes modalidades en su conjunto;
- la adscripción a diferentes modalidades de colectivos edad-profesión;
- clusters edad-profesión-modalidad que siguen el mismo tipo de evolución y se pueden agrupar.

En la sección 4 se estudia descriptivamente la participación de entidades en la muestra, intentando estudiar la presencia de la categoría “*No Contesta*” (NC) en el campo Profesión.

En la sección 5 se analizan mediante *árboles de decisión* hasta qué punto la edad y profesión son indicativos de la modalidad de póliza contratada. Se tratan los datos en conjunto, sin tener en cuenta el año.

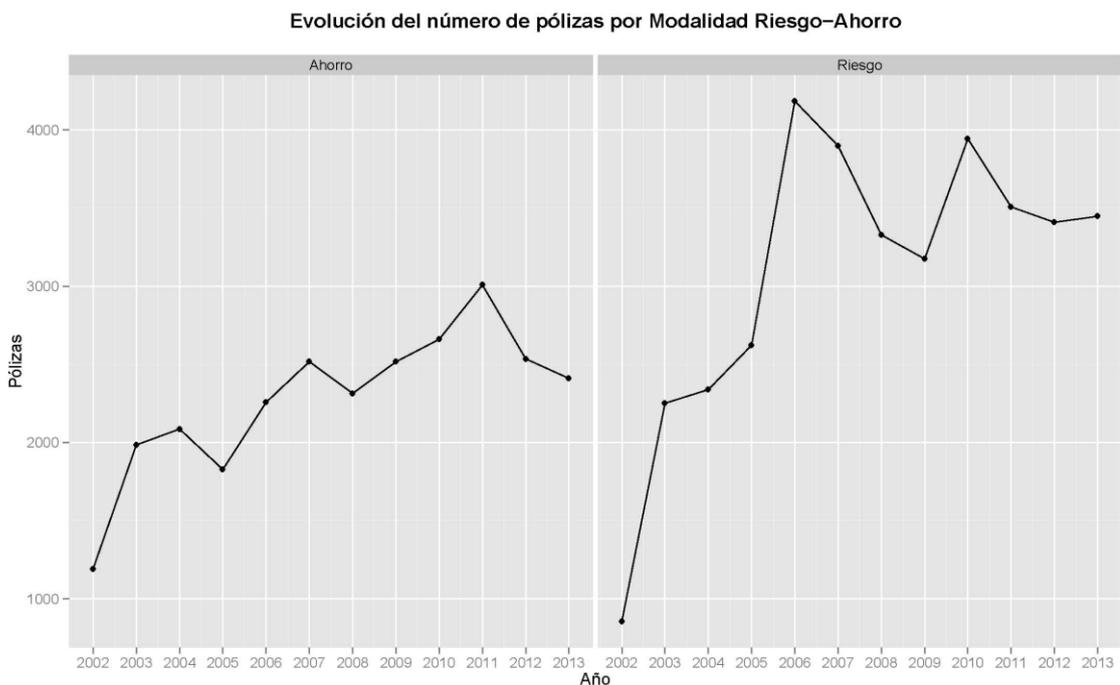
En la sección 6 Se estudian los perfiles prima-capital-duración en las diferentes modalidades. Al existir gran variabilidad de tipos de pólizas, incluso dentro de cada modalidad, se pretenden establecer perfiles de productos para las modalidades más importantes y su presencia en diferentes grupos de profesión.

Finalmente la sección 7 resume las principales conclusiones

## 2 Estudio comparativo de la evolución temporal de las pólizas en los años de crisis, según modalidad. Relación con segmentos poblacionales de edad y profesión.

En primer lugar es necesario un análisis descriptivo para evaluar las técnicas más apropiadas para buscar resultados interesantes a través de técnicas de data mining.

En las gráficas del número de pólizas, se observa de 2008 a 2013, en el total de pólizas, una parte estable y ascendente hasta 2010, y un descenso hasta 2013 regresando a niveles de 2009 (las curvas representan el total de pólizas de vida según modalidad ahorro y riesgo).

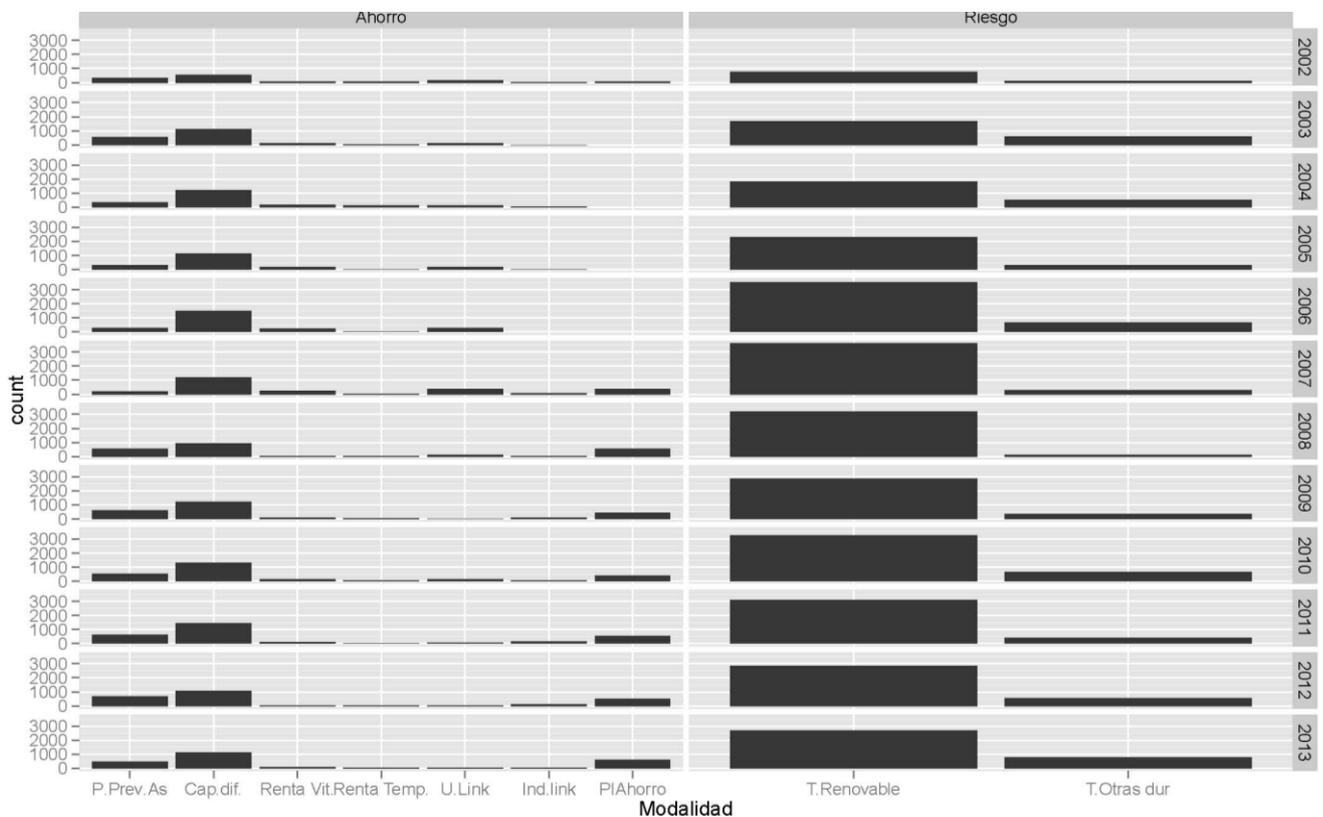
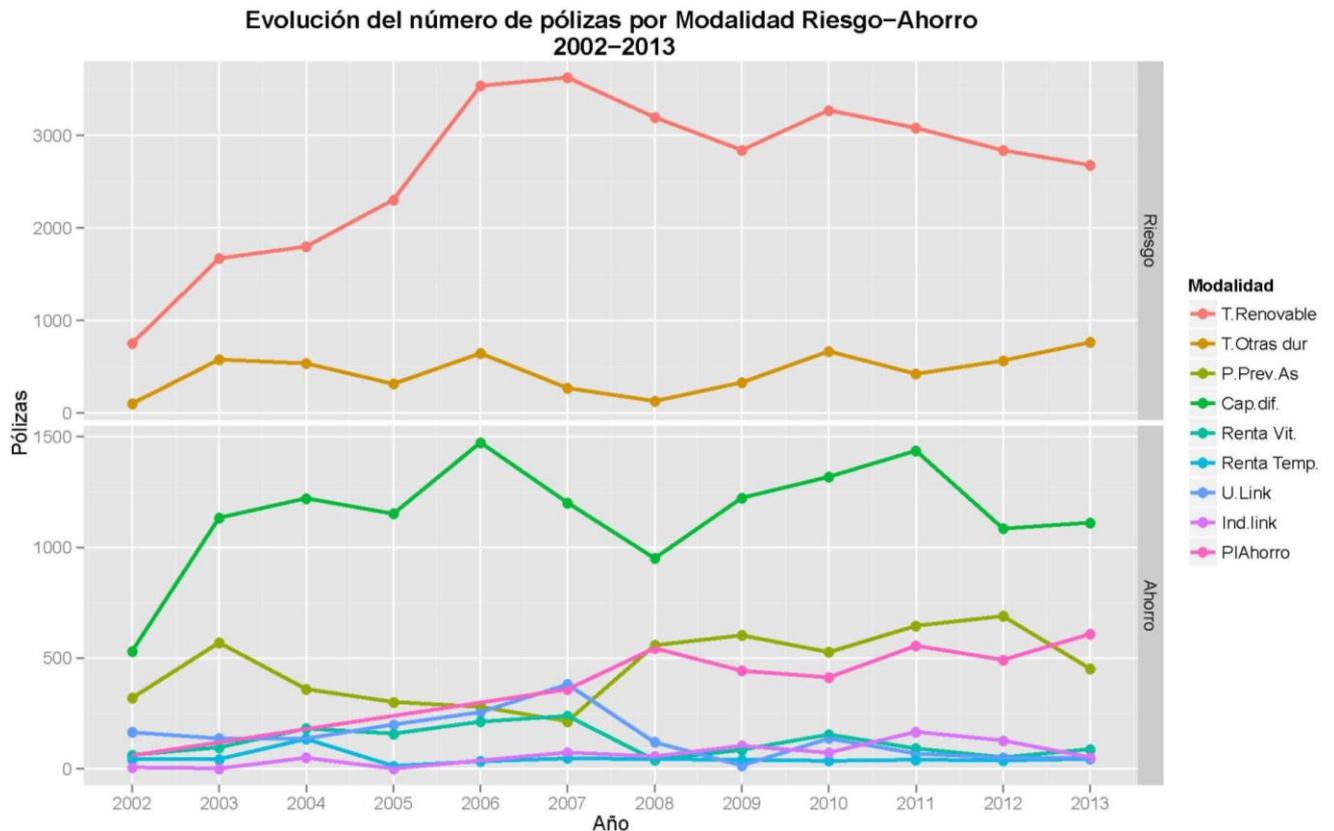


Si se diferencia por modalidad dentro de Ahorro y riesgo, se aprecian diferentes evoluciones:

En Riesgo, un ligero descenso en crisis en la modalidad Temporal renovable y ligero aumento en Temporal otras duraciones, con posible trasvase entre ambas modalidades.

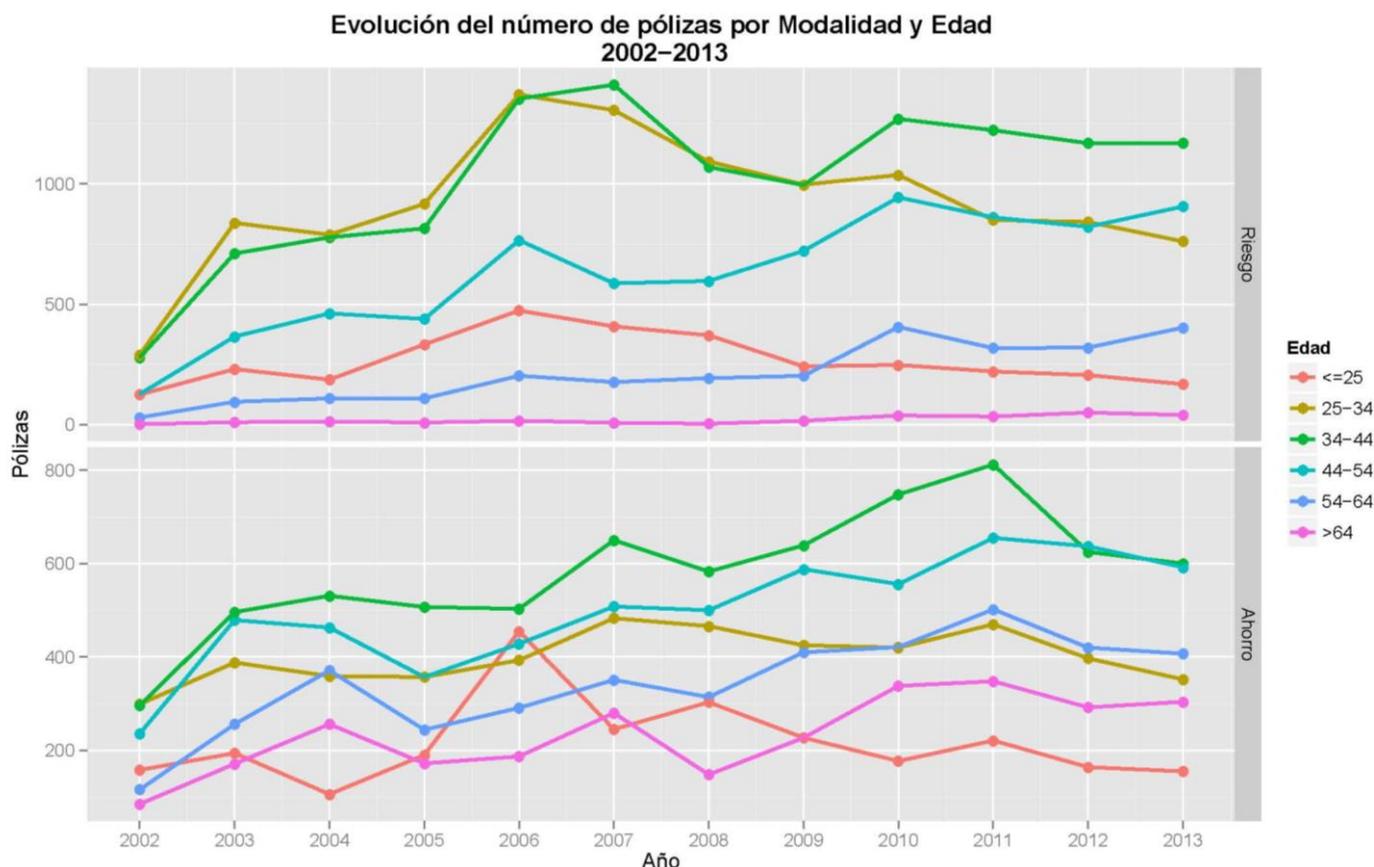
En Ahorro, aumento en Capital diferido hasta 2011 y descenso a continuación, mientras la segunda modalidad más representada, Capital diferido, manifiesta un crecimiento hasta 2012 y descenso en 2013. PIAhorro sigue una curva en ligero ascenso, incluso dentro de la crisis. En otras modalidades menos representadas, Ind. Link manifiesta un ligero descenso y Renta vitalicia, Renta temporal y U. link un ligero ascenso no significativo.

En los gráficos de barras se observa también como el porcentaje relativo de pólizas en las diferentes modalidades se conserva aproximadamente a lo largo del tiempo, siendo en riesgo la modalidad más frecuente Temporal renovable y en ahorro Capital diferido, plan de previsión asegurado y PIAhorro.



## Evolución de los grupos de Edad en las diferentes modalidades

Se presenta en el siguiente gráfico la evolución general en Riesgo y Ahorro, para a continuación diferenciar según las modalidades internas.



### **Riesgo**

En la modalidad Temporal Renovable se aprecia un descenso de pólizas del tramo de edad 25-34 desde el 2007. El tramo 34-44 refleja un descenso de 2007 a 2009 para recuperarse y mantenerse aproximadamente en 2010-2013. Otros tramos de edad se mantienen sin apenas cambios.

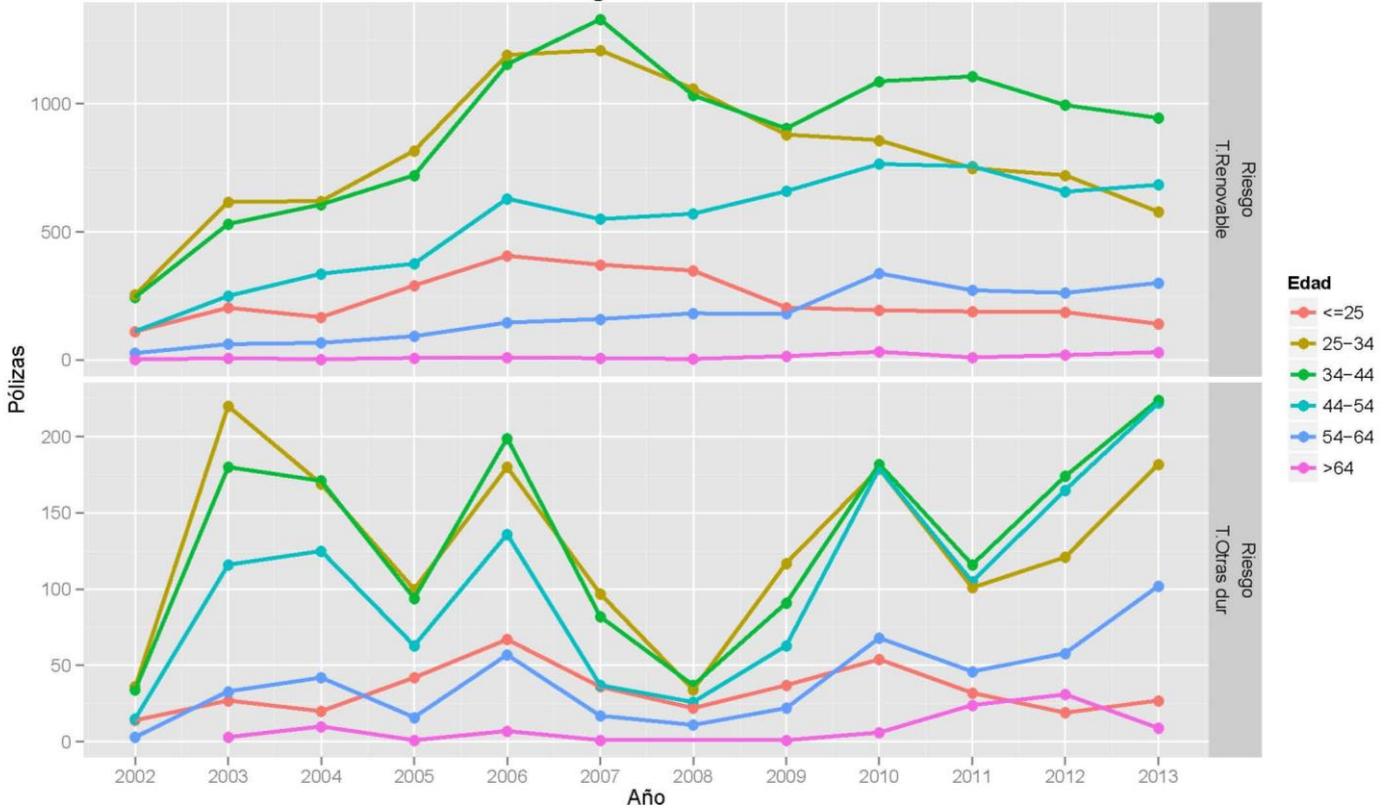
En otras duraciones, hay un comportamiento homogéneo en casi todos los grupos de edad, salvo en  $\geq 64$ , donde se aprecia un descenso en 2013.

### **Ahorro**

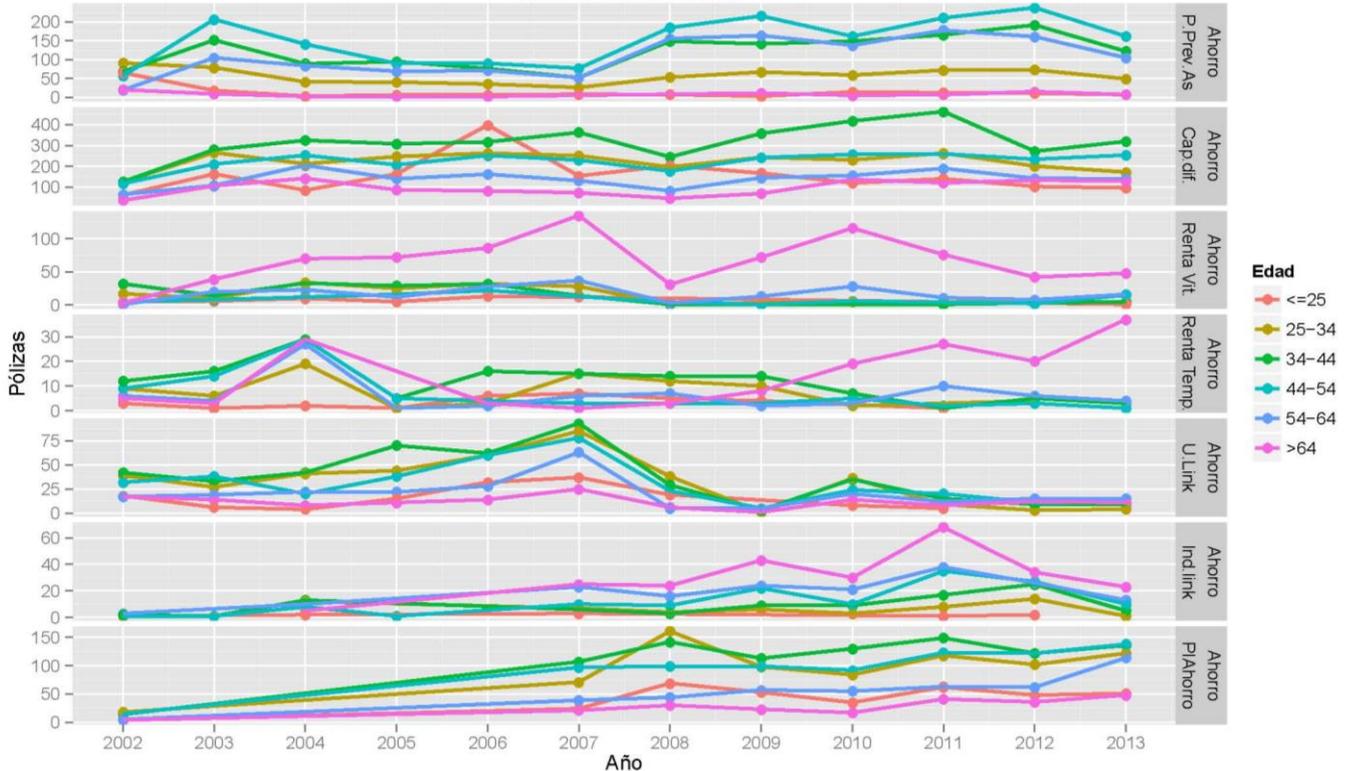
En general no hay grandes diferencias en la evolución de los diferentes tramos de edad, salvo en Renta Temporal, en la cual el rango de edad  $\geq 64$  tiene una evolución creciente diferente de los tramos de edad de la misma modalidad.

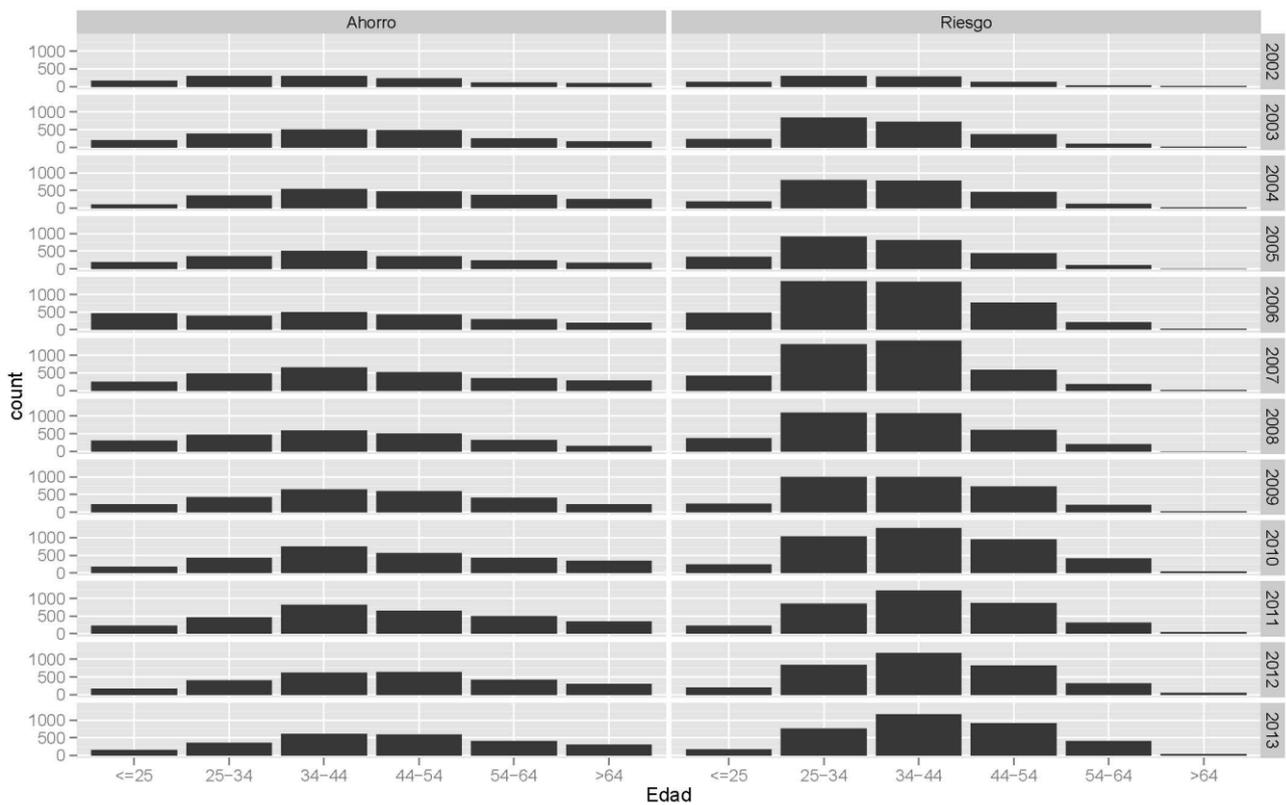
En el diagrama de barras se observa cómo en Riesgo, la proporción relativa de jóvenes (25-34) se reduce en los últimos años en beneficio de una mayor proporción de mayores (34-54)

**Evolución del número de pólizas por Modalidad y Edad  
Riesgo. 2002-2013**



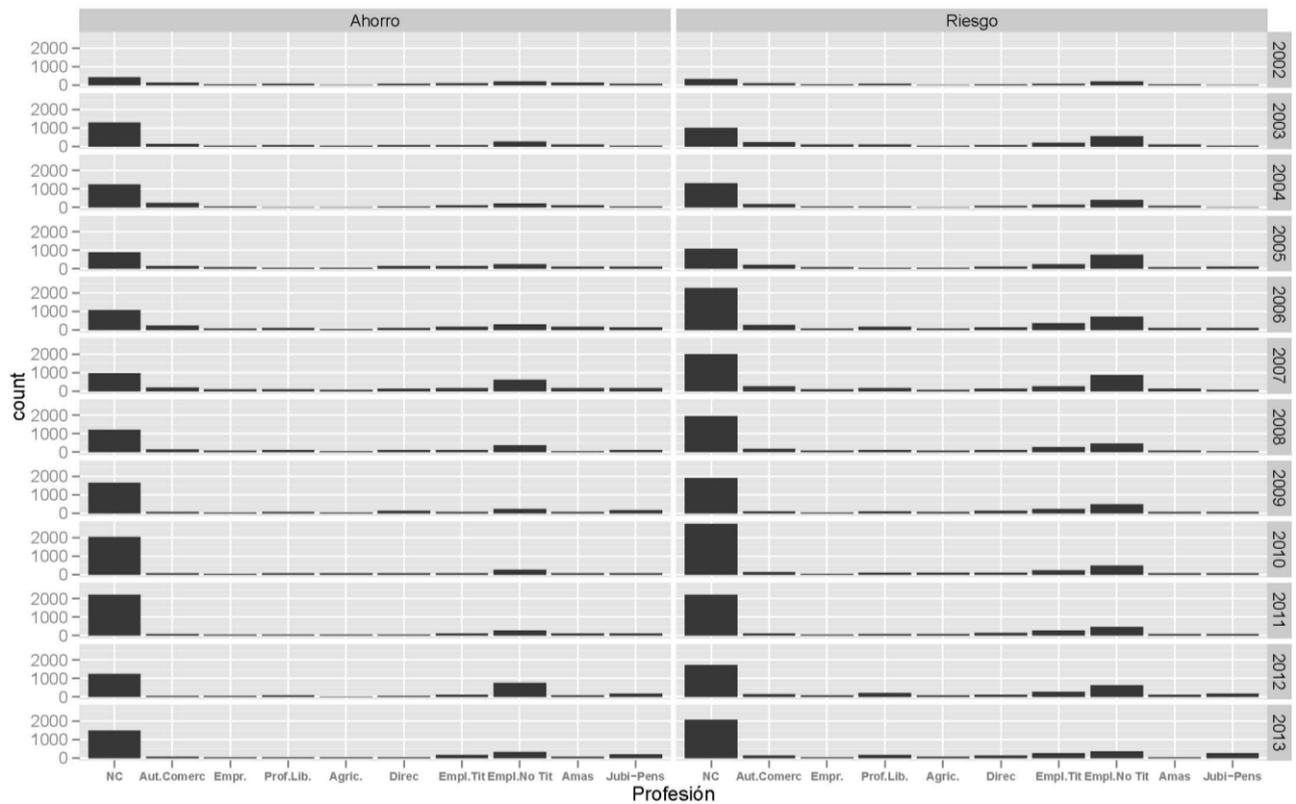
**Evolución del número de pólizas por Modalidad y Edad  
Ahorro. 2002-2013**





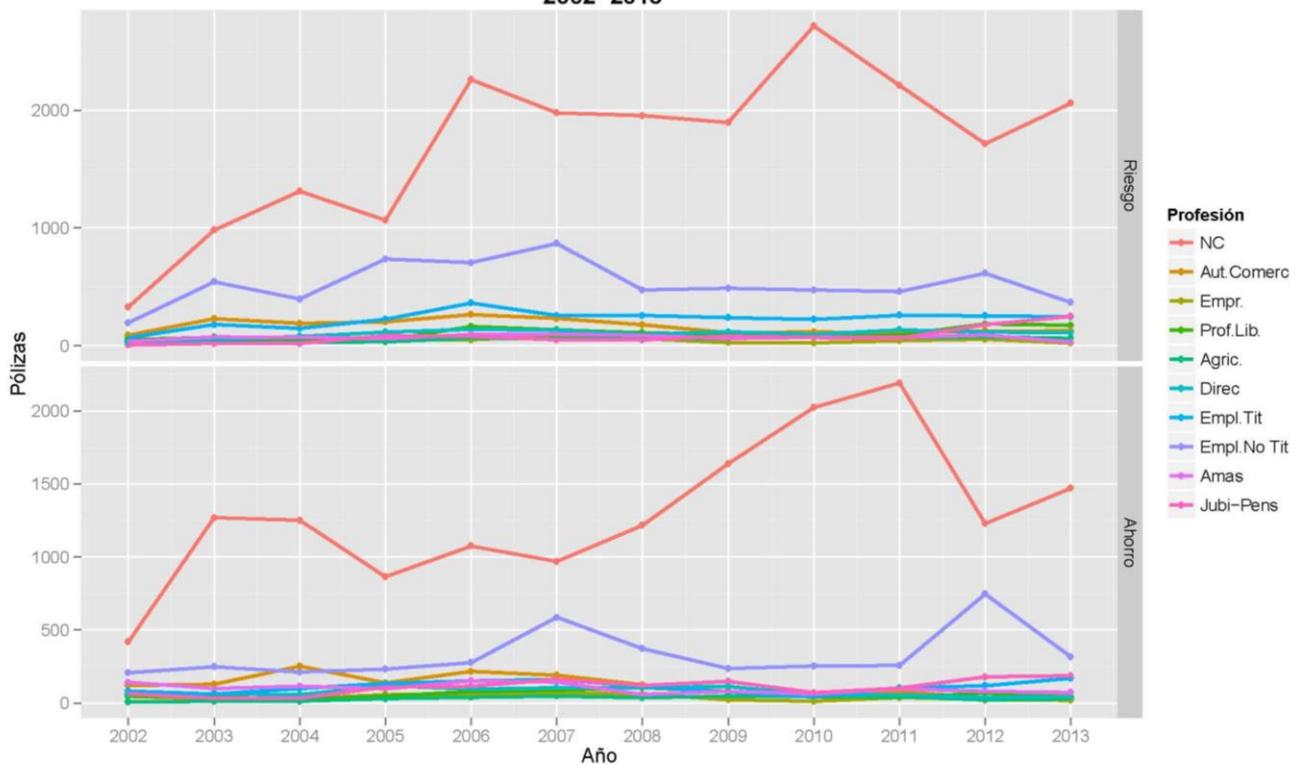
### Evolución del número de pólizas por modalidad respecto a Profesión

La presencia de pólizas con profesión desconocida NC en la base de datos marca los resultados. En esta categoría entrarían tanto los parados, como gente que no quiere declarar su profesión o simple ausencia del campo en la información recogida por las entidades. Después de esta categoría, la presencia más frecuente recae sobre Empleados No Titulados, y después Jubilados y pensionistas y Autónomops y Comerciantes, tanto en Ahorro como en Riesgo. A partir de 2005 en Riesgo se observa una mayor presencia relativa de la categoría NC, debida posiblemente al aumento del paro. En Ahorro este aumento se observa a partir de 2008.



En la gráfica de evolución se observa el efecto de la crisis, en Riesgo a partir de 2010 y en Ahorro a partir de 2011. Puede existir un trasvase de pólizas de Empleados No Titulados a NC (posiblemente debido a formar parte de la población de parados), en 2012-2013.

Evolución del número de pólizas por Modalidad y Profesión 2002-2013

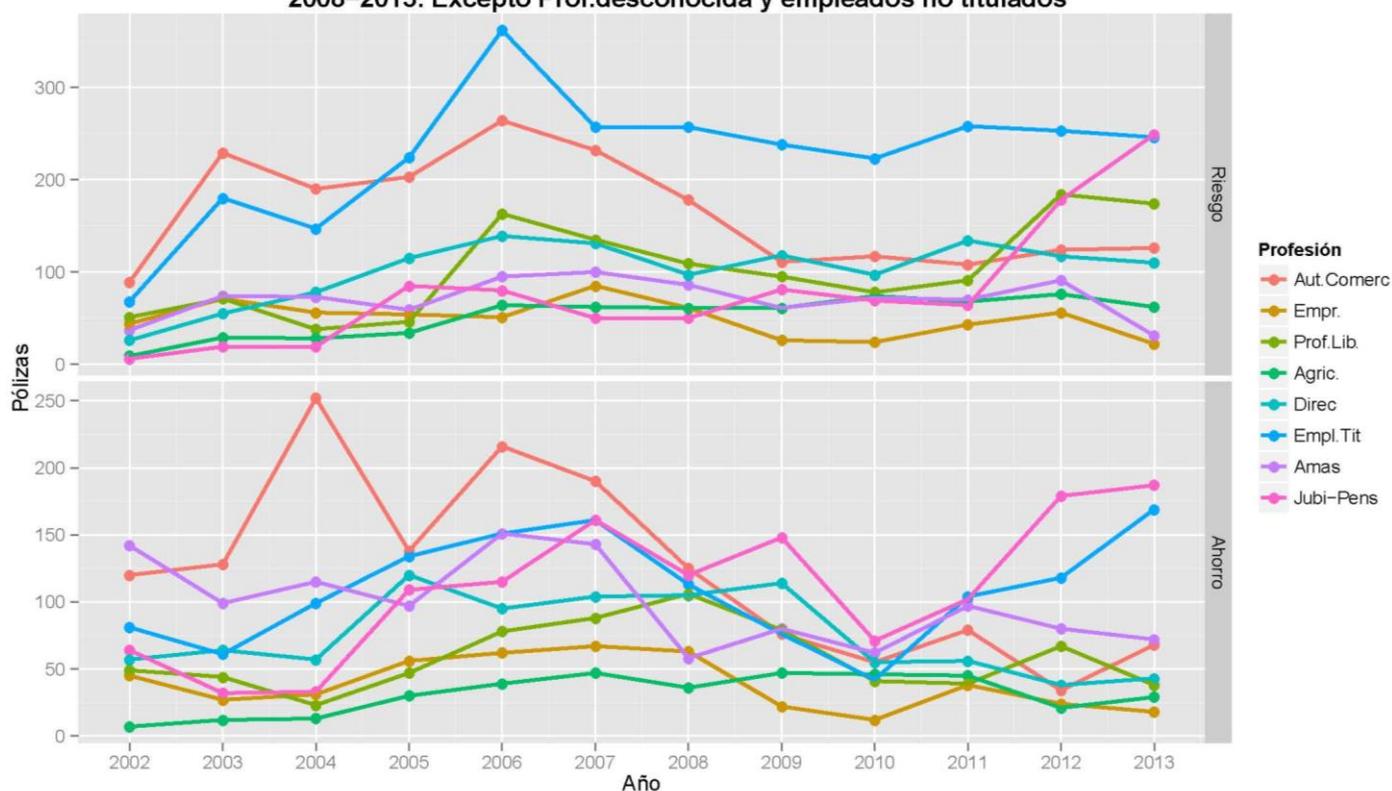


Para observar mejor las evoluciones durante el periodo de crisis no se tienen en cuenta en el siguiente gráfico las categorías NC y Empleados no Titulados. Tanto en Ahorro como en Riesgo se aprecia que la categoría Jubilados-pensionistas manifiesta una evolución creciente, frente a otras categorías con evoluciones estables o con ligeros decrecimientos o aumentos.

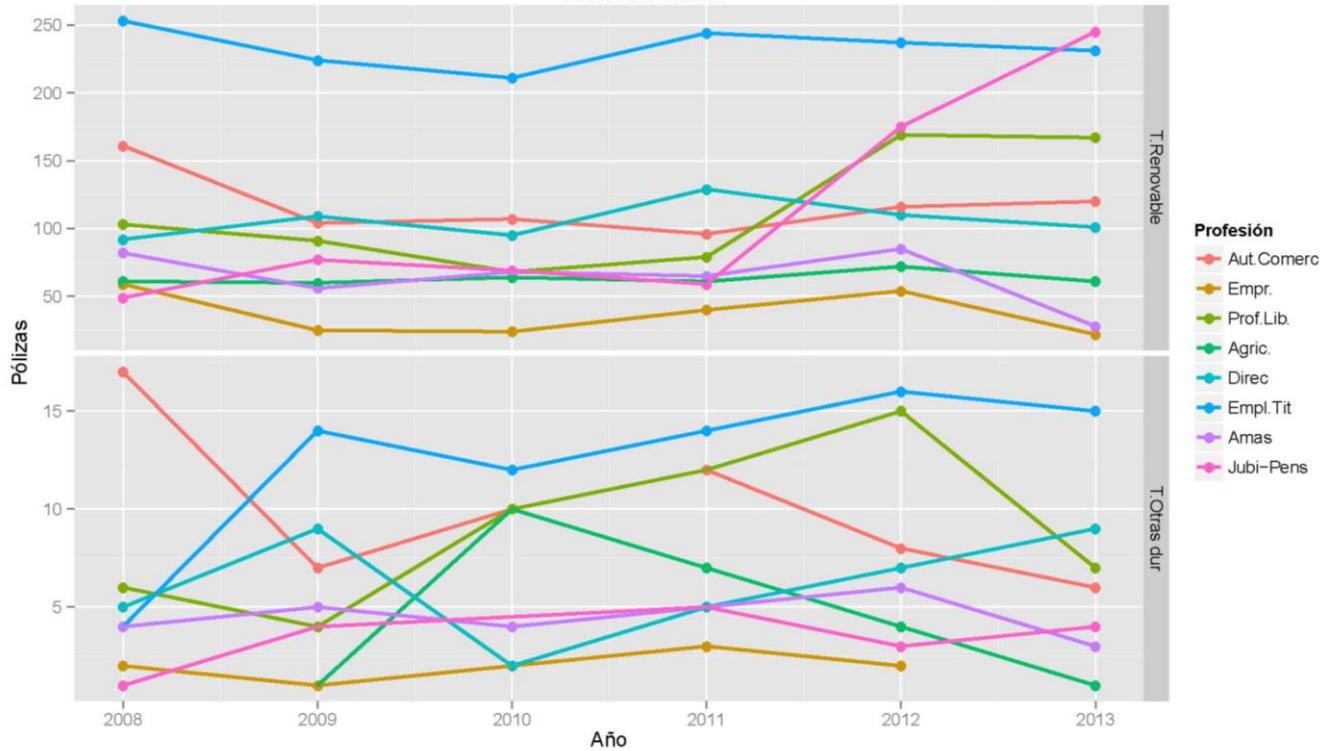
Estos patrones se observan mejor descendiendo a nivel modalidad, dentro de Ahorro y Riesgo. En Riesgo, la evolución creciente de la categoría Jubilados-pensionistas es debida a las pólizas de modalidad Temporal Renovable (en Otras duraciones no existe este crecimiento). En Ahorro, existe un crecimiento general de esta categoría a partir de 2010 en casi todas las modalidades, con probable descenso en 2012-2013.

### Evolución del número de pólizas por Modalidad y Profesión

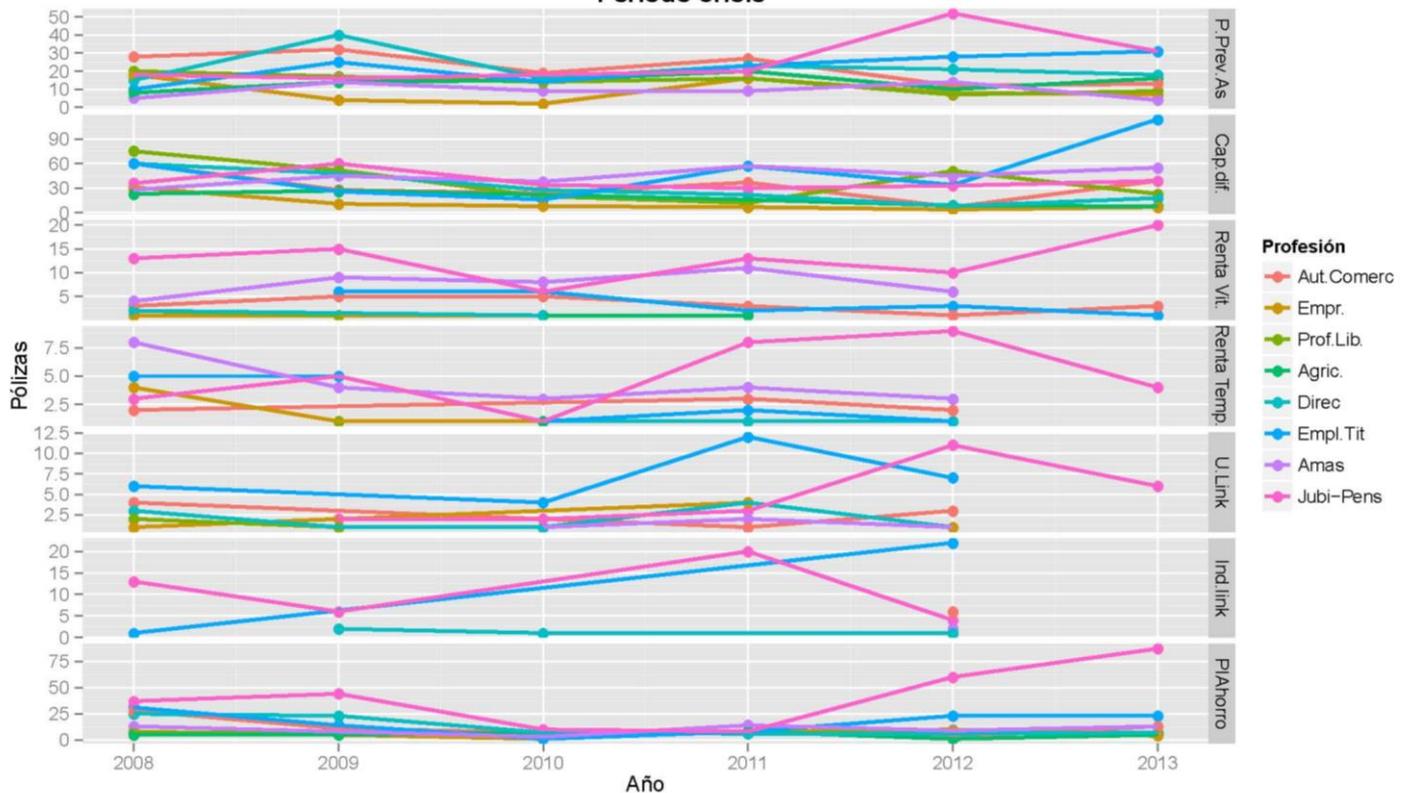
2008–2013. Excepto Prof.desconocida y empleados no titulados



**Evolución del número de pólizas por Modalidad y Profesión  
Riesgo.Salvo NC y Empleados no titulados.  
Periodo crisis**



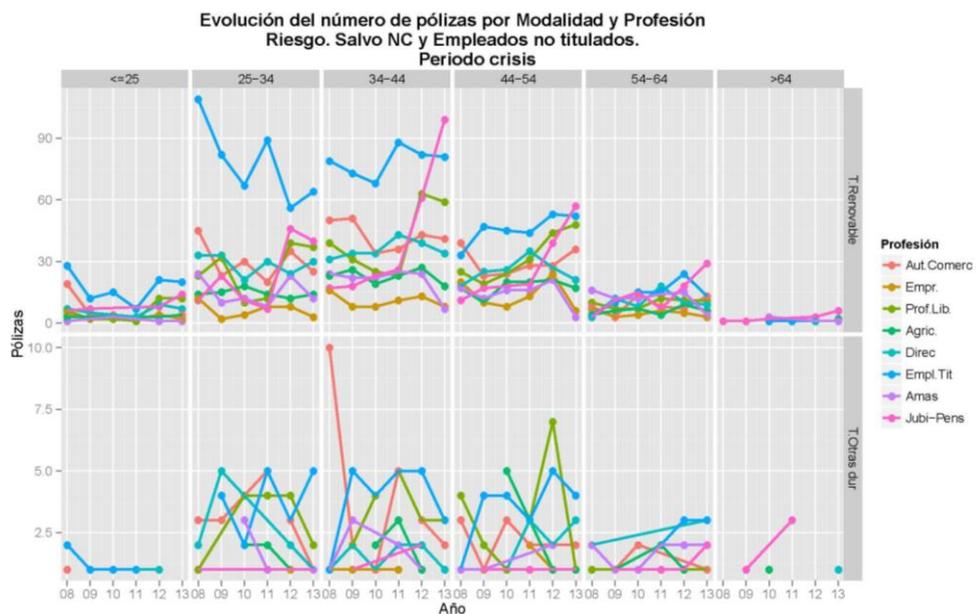
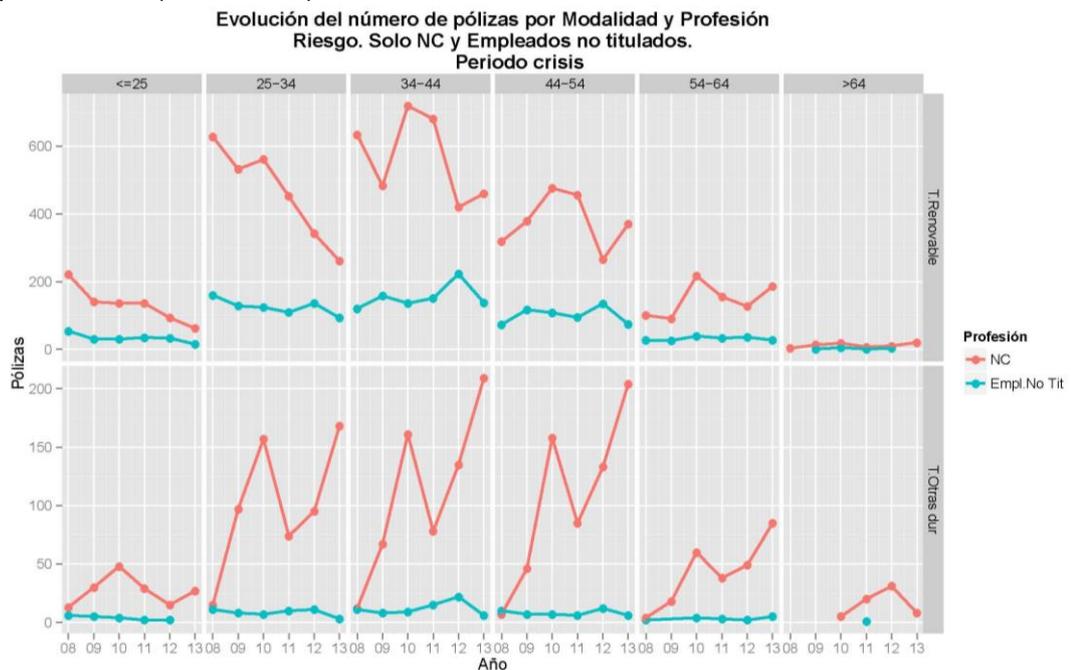
**Evolución del número de pólizas por Modalidad y Profesión  
Ahorro.Salvo NC y Empleados no titulados.  
Periodo crisis**



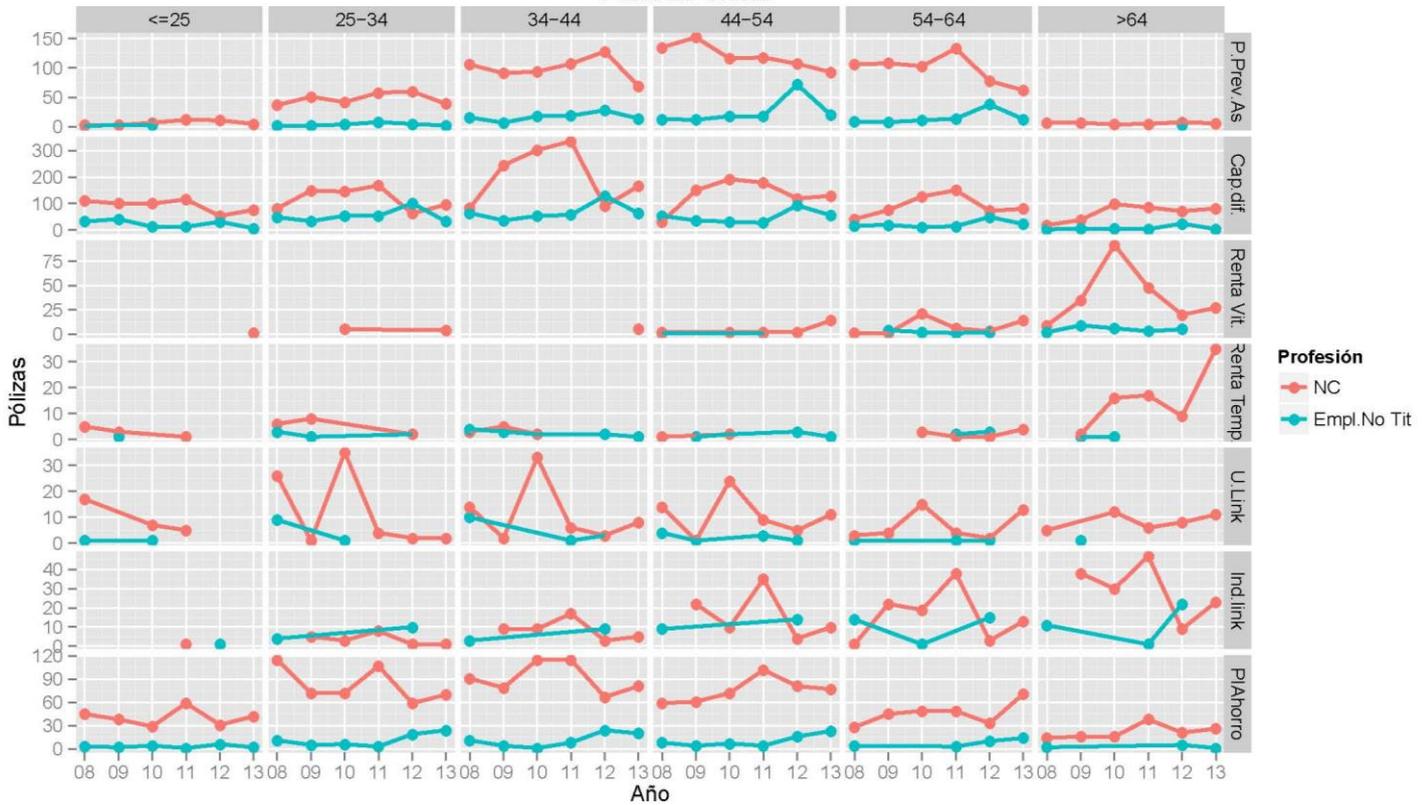
## Evolución del número de pólizas por modalidad respecto a Profesión y Edad

Es necesario estudiar la interacción Profesión-edad, pues la combinación de ambos factores puede originar diferentes decisiones para la contratación de pólizas y evoluciones en crisis también diferentes. Como se ve en el siguiente gráfico, la combinación NC, <=34, en modalidad Temporal Renovable, tiene una evolución diferente que NC, 34-64 en Temporal Renovable y NC, <=64 en modalidad Temporal Otras duraciones.

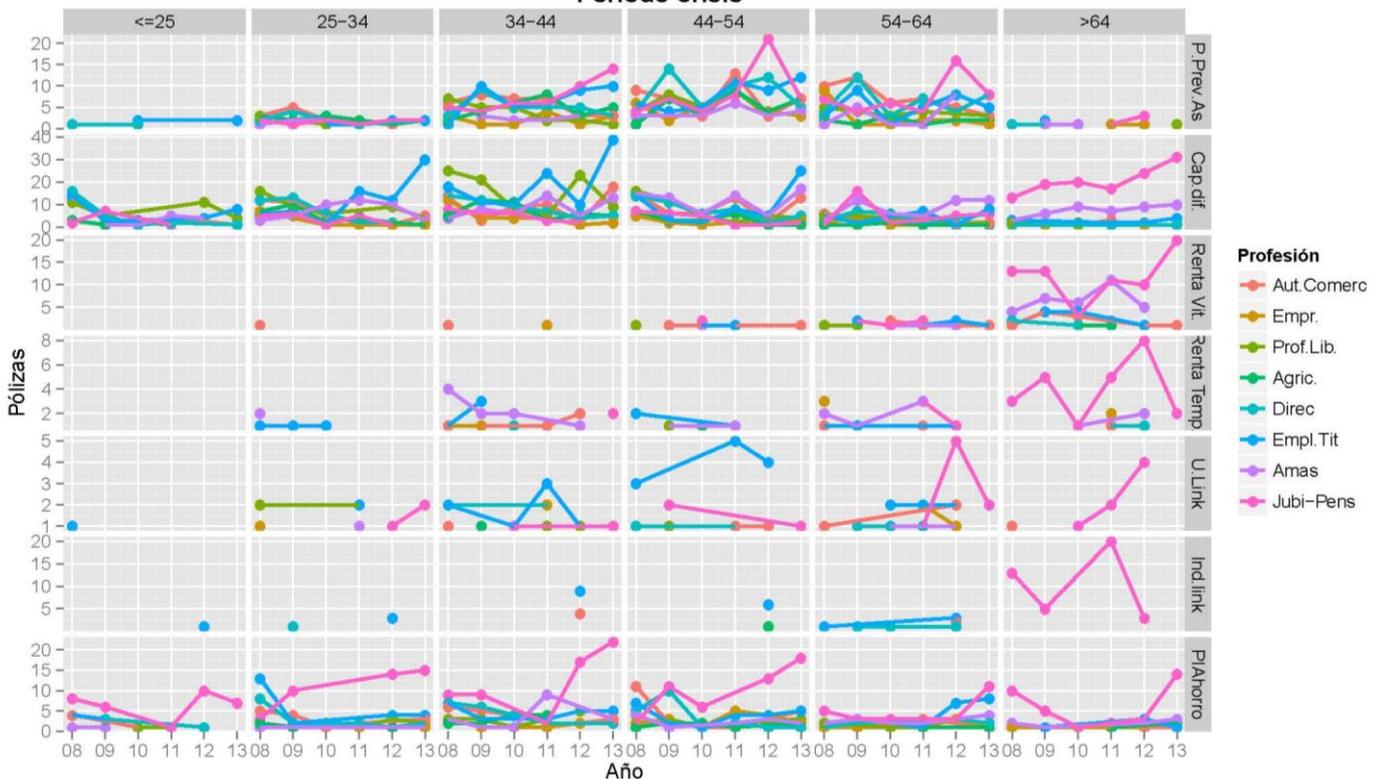
Si en Riesgo se realiza el mismo gráfico exceptuando las profesiones NC y empleados no titulados, también se aprecian diferentes evoluciones, en muchos casos poco distinguibles al disponer de poca información (pocas pólizas de determinadas combinaciones de profesión-edad). En Ahorro también es posible detectar patrones de evolución que se pueden agrupar (por ejemplo NC, PIAhorro, edades <=54) y otros patrones específicos de alguna modalidad para alguna profesión independientemente del grupo de edad (NC, U.Link).



### Evolución del número de pólizas por Modalidad y Profesión Ahorro. Solo NC y Empleados no titulados. Periodo crisis



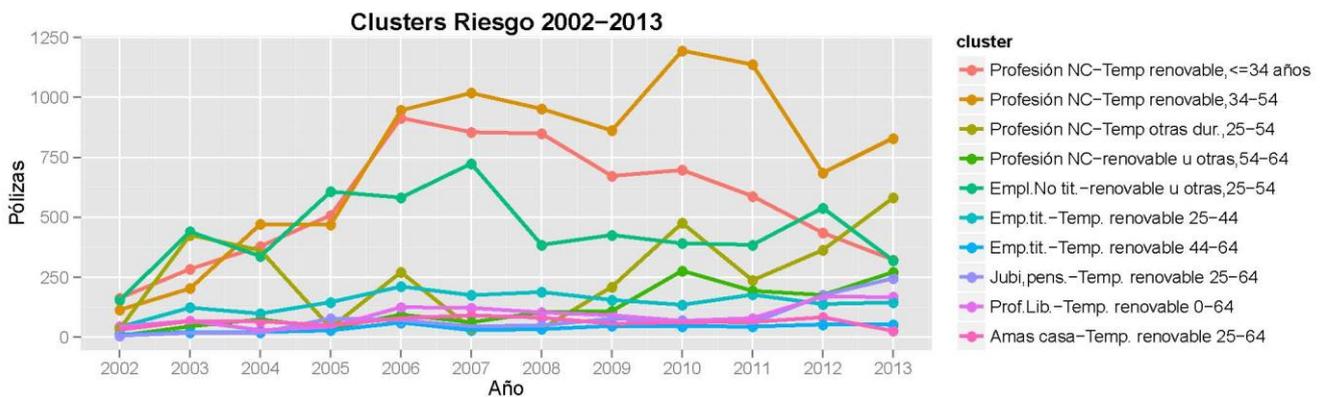
### Evolución del número de pólizas por Modalidad y Profesión Ahorro. Salvo NC y Empleados no titulados. Periodo crisis



Debido al número de combinaciones existentes, es interesante utilizar algún procedimiento de Data Mining para agrupar combinaciones de edad-profesión-modalidad que tengan una evolución similar en tiempo de crisis. En este caso se utilizarán algoritmos de clustering (distancia euclídea, método de Ward), utilizando como elementos las combinaciones edad-modalidad-profesión y planteando la distancia entre elementos como la distancia entre el vector promedio del número de pólizas en los años 2007-2013. Se han dejado fuera clusters poco claros, restringiéndonos en la presentación a los más fáciles de comprender e identificar.

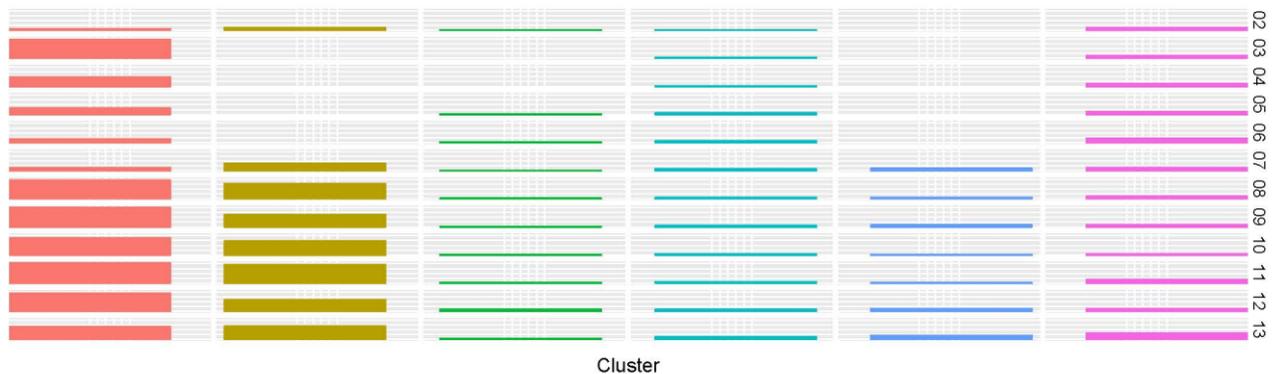
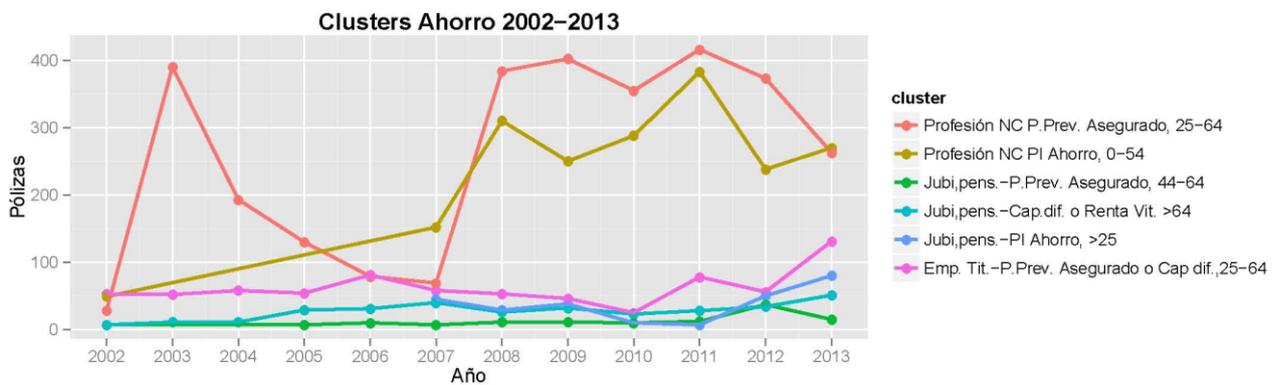
### Clusters en Riesgo

Se han construido 10 clusters diferenciados en cuanto a su evolución en crisis. Los más numerosos son los 1 y 2, repectivamente NC,T.renovable, <=34 y NC,T.renovable, 34-54, con una evolución conjunta contrapuesta en crisis, claramente descendiente en el primero desde 2006. El tercer cluster de interés engloba a Profesión NC, Temp. otras duraciones, 25-54, con trayectoria creciente en crisis. Un cluster mencionado intuitivamente es el de Jubilados-pensionistas, temp. renovable, 25-64 con trayectoria creciente en crisis, como el de Profesiones liberales, temp. renovable, 0-64.



## Clusters en Ahorro

En Ahorro se han construido 6 clusters diferenciados. Los clusters más numerosos son los formados por NC, P.P. Asegurado, 25-64, con evolución creciente hasta 2011 y decreciente después, y NC, PIAhorro, 0-54 con un patrón evolución de fuerte decrecimiento en 2011-2012 para recuperarse en 2012-2013. Con evolución creciente en crisis están los Empleados titulados 25-64, con P.P. Asegurado o Cap.diferido. También Jubilados-pensionistas >64, Cap. diferido o Renta Vitalicia, o bien Jubilados-pensionistas >25, con PIAhorro.

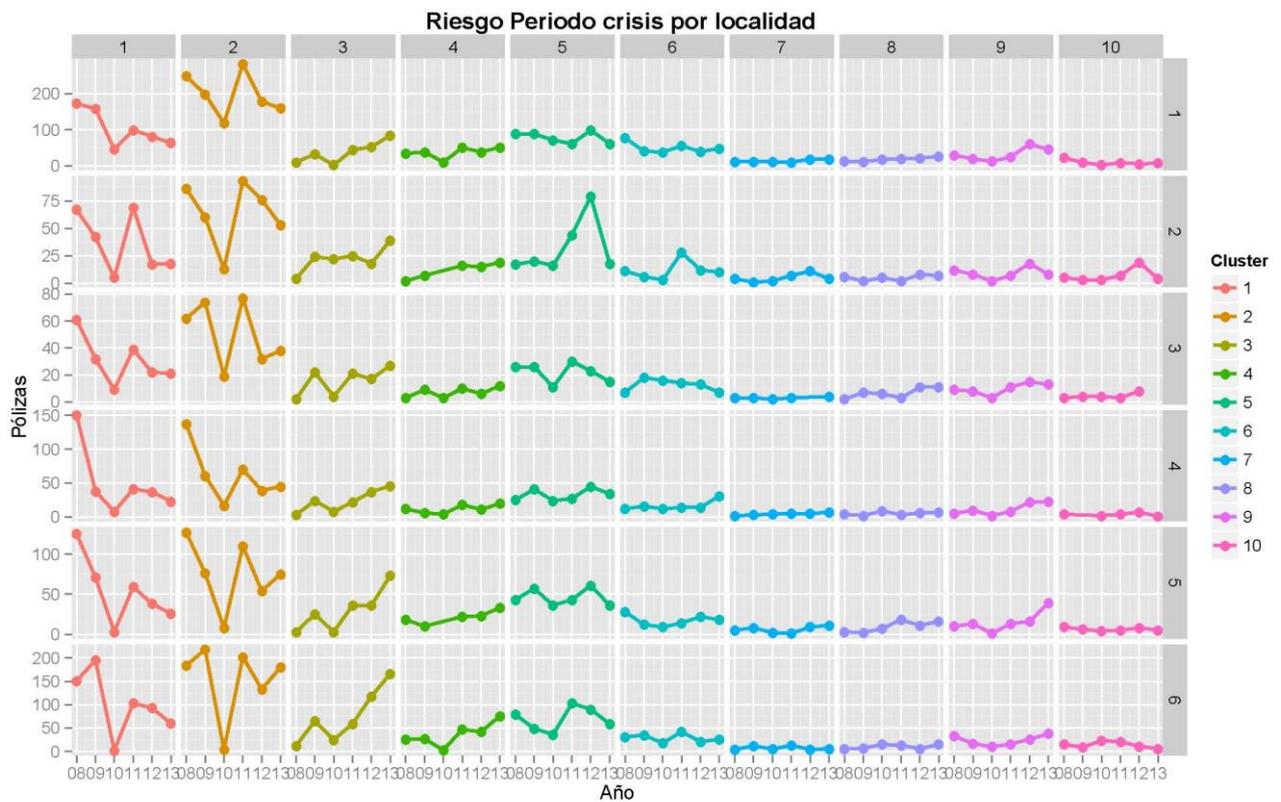


### 3 Evolución en los diferentes tipos de localidad por tamaño de población

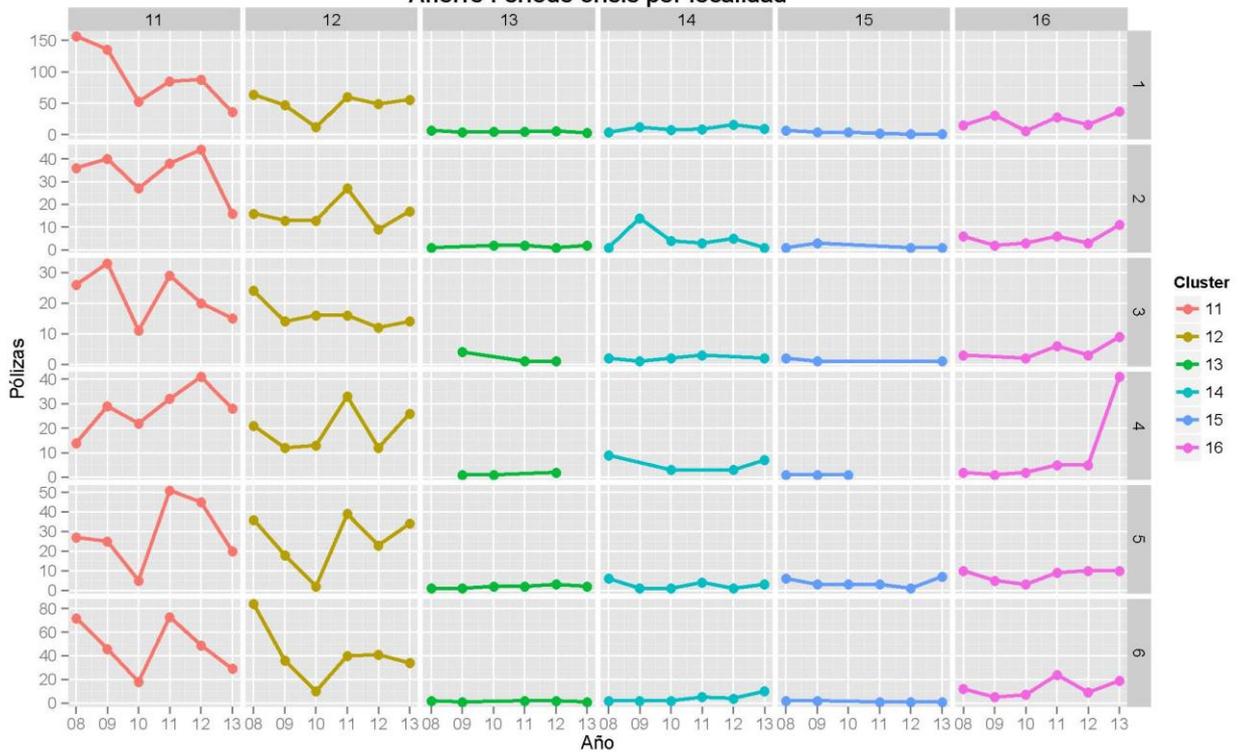
Cabe preguntarse si la pertenencia a diferentes tipos de localidades origina diferentes patrones de evolución. En general no se han encontrado diferencias. Como ejemplo, si se presenta la evolución de los clusters construidos en los diferentes tipos de localidades se observa que la evolución es en general similar.

En la base de datos, consta el tipo de localidad, según la siguiente especificación:

- 1 Capital de Provincia
- 2 Más de 100.000 hab.
- 3 De 50.000 a 100.000 hab.
- 4 De 25.000 a 50.000 hab.
- 5 De 10.000 a 25.000 hab.
- 6 Menos de 10.000 hab.



### Ahorro Periodo crisis por localidad



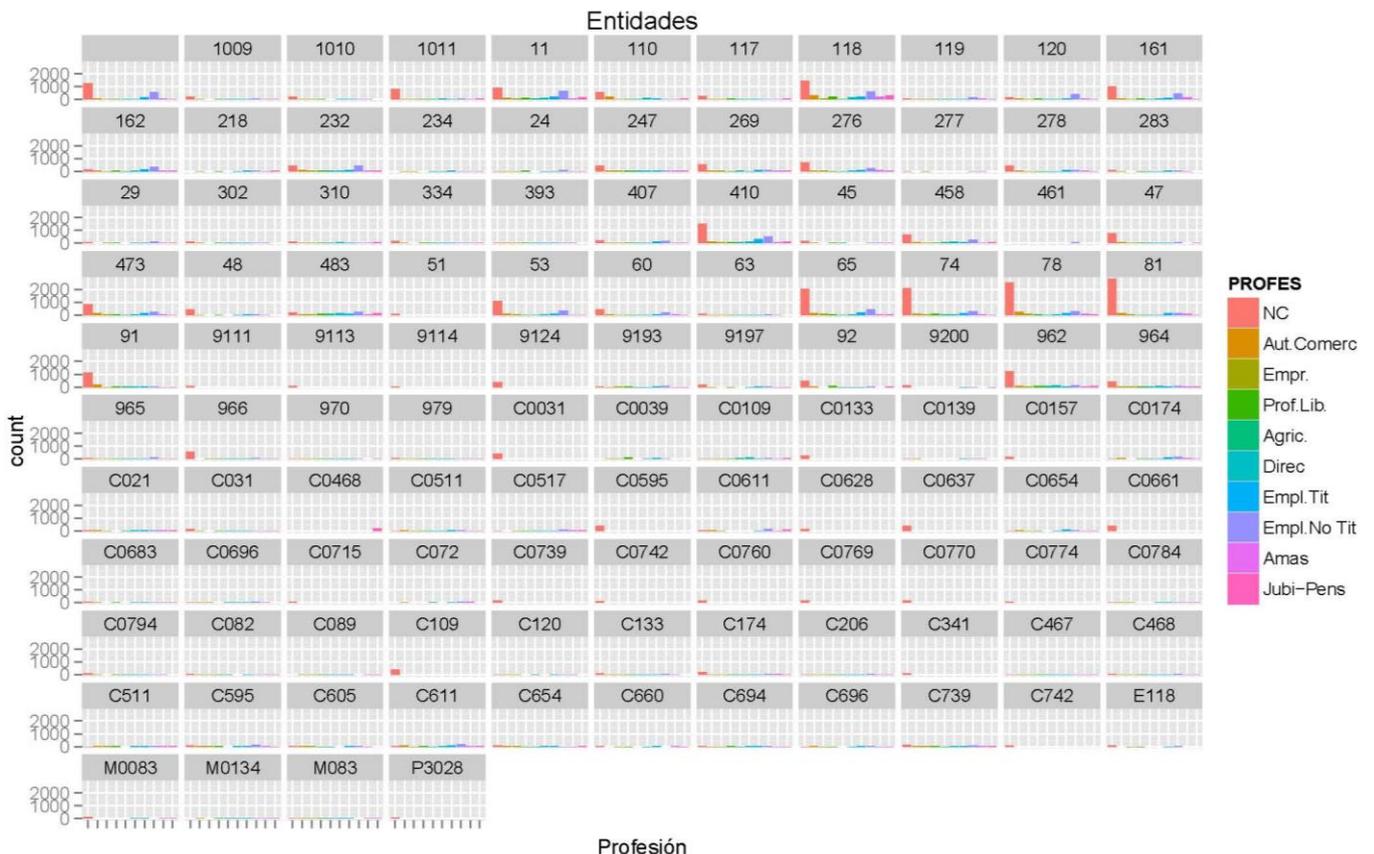
## 4 Patrones en las Entidades presentes en la base de datos

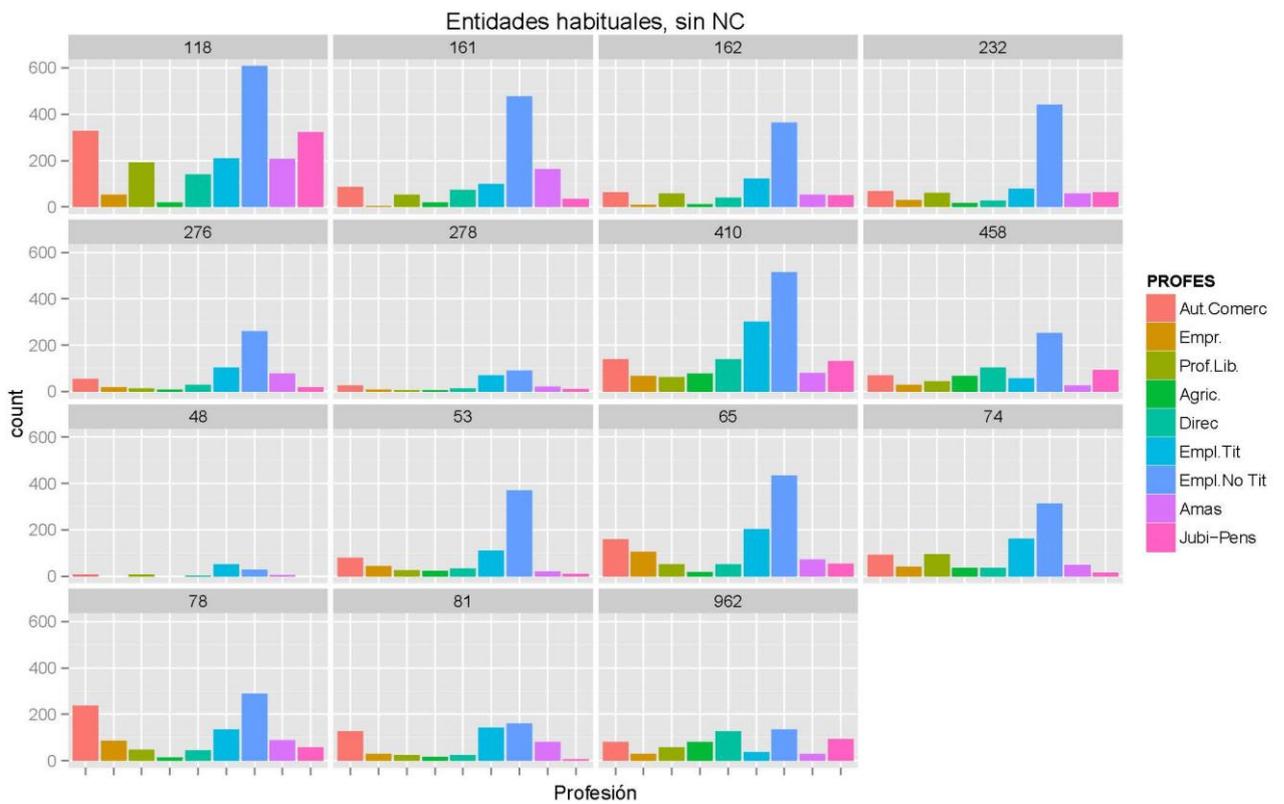
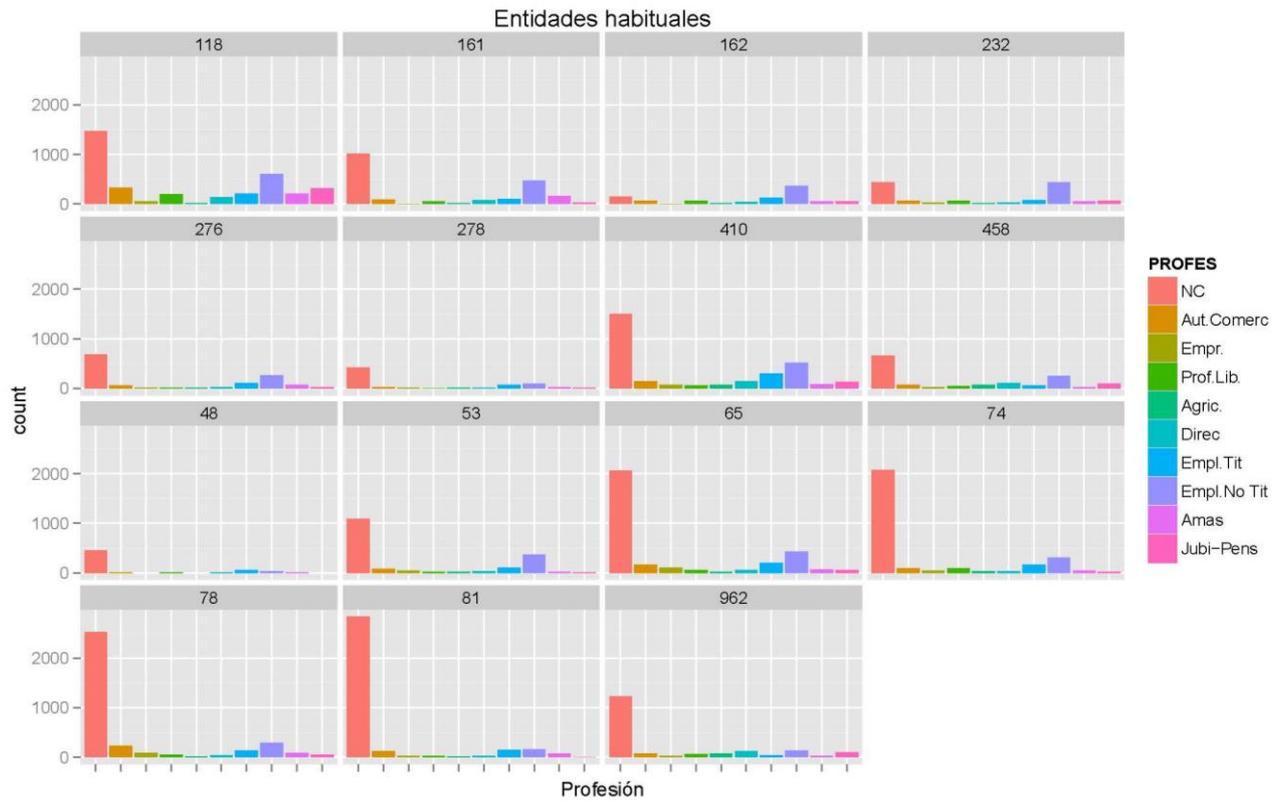
Un problema que surge a la hora de interpretar los datos es la categoría NC en el campo Profesión. Cabe preguntarse si se puede restringir la base de datos a entidades con información sólida y continuada respecto al campo profesión. La respuesta es negativa.

Como resumen (en el anexo está la información más detallada), se dispone de los siguientes datos:

- El número de entidades diferentes en la base de datos es 114.
- Las entidades que no presentan individuos con profesión \_NC (No Sabe, No Contesta) son en total 7, pero cada una de ellas sólo está presente en un año determinado.
- El número de entidades que solamente reportan NC en el campo Profesión asciende a 23 entidades. Cada una de ellas sólo está presente en un año determinado. 15 entidades de las 23 se han añadido recientemente, son del 2013.
- Buscando entidades habituales, el número de entidades presentes en 7 o más años del periodo 2002-2013 son 15 entidades. Si se trata de de entidades presentes en 8 o más años, son 11.

Teniendo en cuenta el apartado b) no es posible prescindir del término NC en el campo Profesión en el estudio. Se presentan algunos gráficos con la presencia de distintas profesiones en las diferentes entidades.





## 5 Patrones de edad-profesión en las diferentes modalidades

En este apartado se pretende estudiar mediante árboles de decisión hasta que punto la edad y profesión son indicativos de la modalidad contratada. Se tratan los datos sin tener en cuenta el año.

Los árboles se construyen con el campo modalidad como variable dependiente, y las variables edad2 y profesión como independientes. El contenido formal está en el anexo.

Como resultados relevantes, se tiene:

### Riesgo

- En todas las profesiones salvo NC, y todas las edades, el porcentaje de Temporal renovable es 90.3 y de temporal otras duraciones, 9.7.
- Para el rango de edad  $\leq 44$ , profesión NC, el porcentaje de Otras duraciones sube a 16.3.
- Para el rango de edad  $> 44$ , profesión NC, el porcentaje de Otras duraciones sube a 23.

### Ahorro

En general se distingue una presencia mayoritaria de Capital diferido, P.Prev. As. y PIAhorro, poca Renta Vitalicia para edades menores de 64. Para Autónomos y comerciantes  $> 64$ , una clara predisposición hacia Renta Vitalicia, y para Amas de casa, Empleados titulados o Directivos mayores de 64, predisposición hacia Capital diferido o Renta Vitalicia.

En concreto:

Para todas las **edades  $\leq 64$** , el porcentaje de pertenencia a las diferentes modalidades es:

U.link	:	6.4%
Cap.dif	:	51.8%
P.Prev.As.	:	22.2%
PIAhorro	:	13.3%
Renta Vit.	:	2.8%
Renta Temp.	:	1.7%
Ind.Link:		1.9%

Para **edad  $> 64$** :

Si la Profesión es Autónomos y Comerciantes, los porcentajes quedan:

U.link	:	4.5%
Cap.dif	:	18.9%
P.Prev.As.	:	1.8%
PIAhorro	:	2.7%
Renta Vit.	:	71.2%
Renta Temp.	:	0.9%
Ind.Link:		0.0%

Si la Profesión es Amas de casa, Empleados titulados o Directivos, los porcentajes son:

U.link	:	4.2%
Cap.dif	:	42.3%
P.Prev.As.	:	5.3%
PIAhorro	:	4.9%
Renta Vit.	:	39.1%
Renta Temp.	:	4.2%
Ind.Link:		0.0%

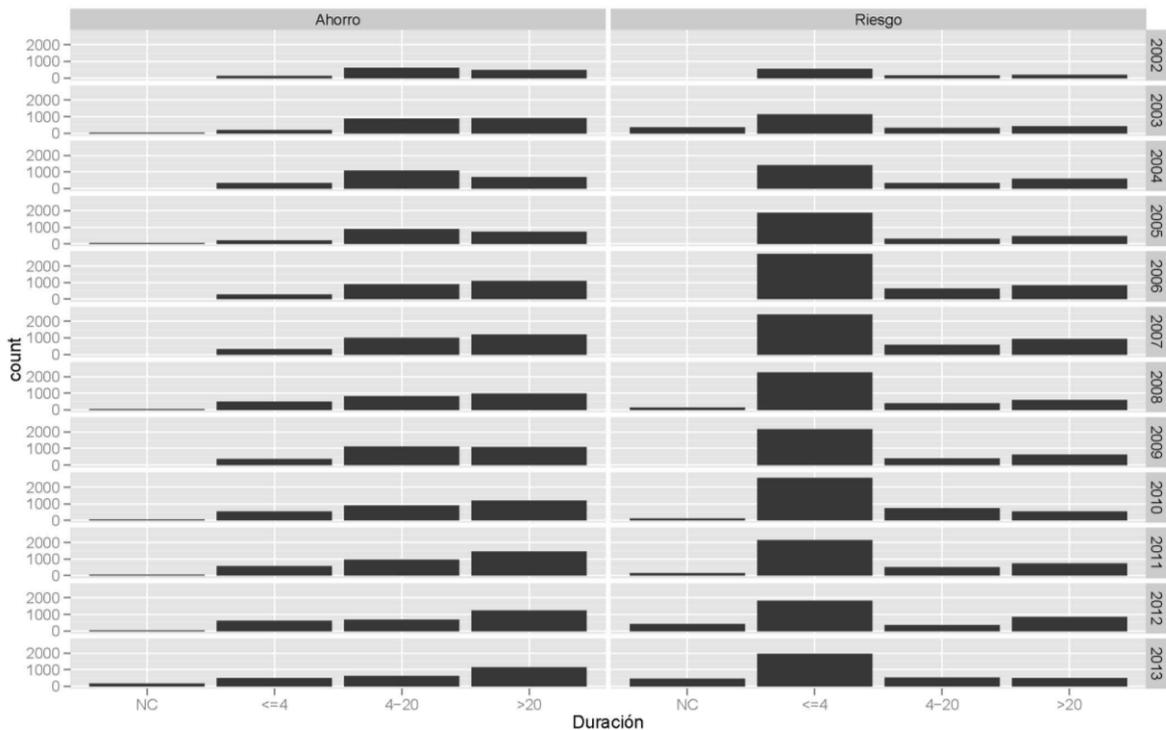
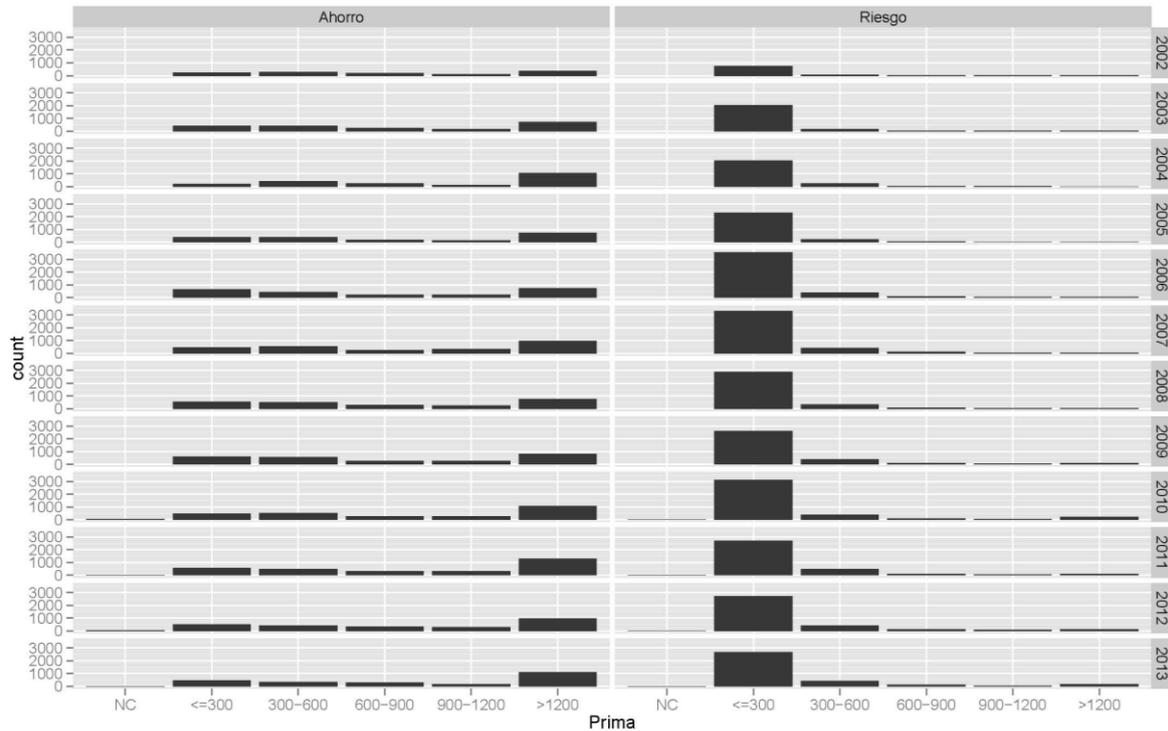
Para el resto de Profesiones (siempre>64), es

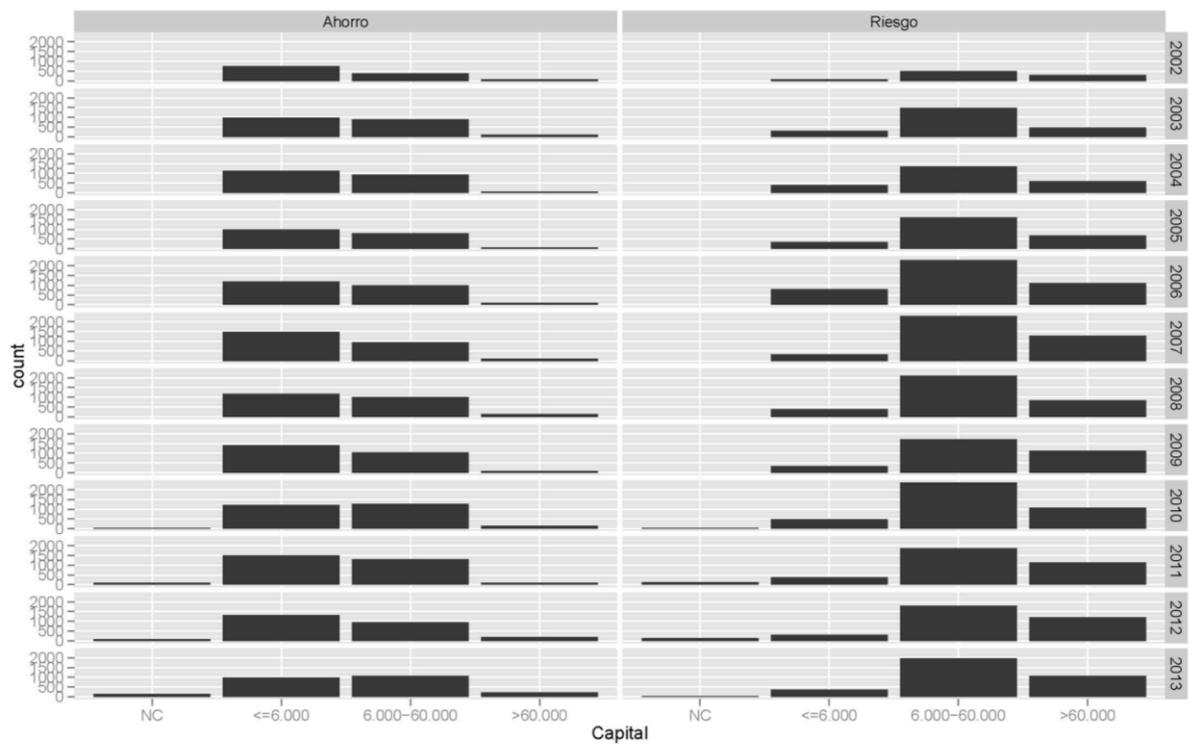
U.link	:	5.1%
Cap.dif	:	42.3%
P.Prev.As.	:	2.9%
PIAhorro	:	8.4%
Renta Vit.	:	24.9%
Renta Temp.	:	5.9%
Ind.Link:		2.4%

## 6 Patrones en los perfiles de las pólizas por prima, duración y capital, segregando por modalidad.

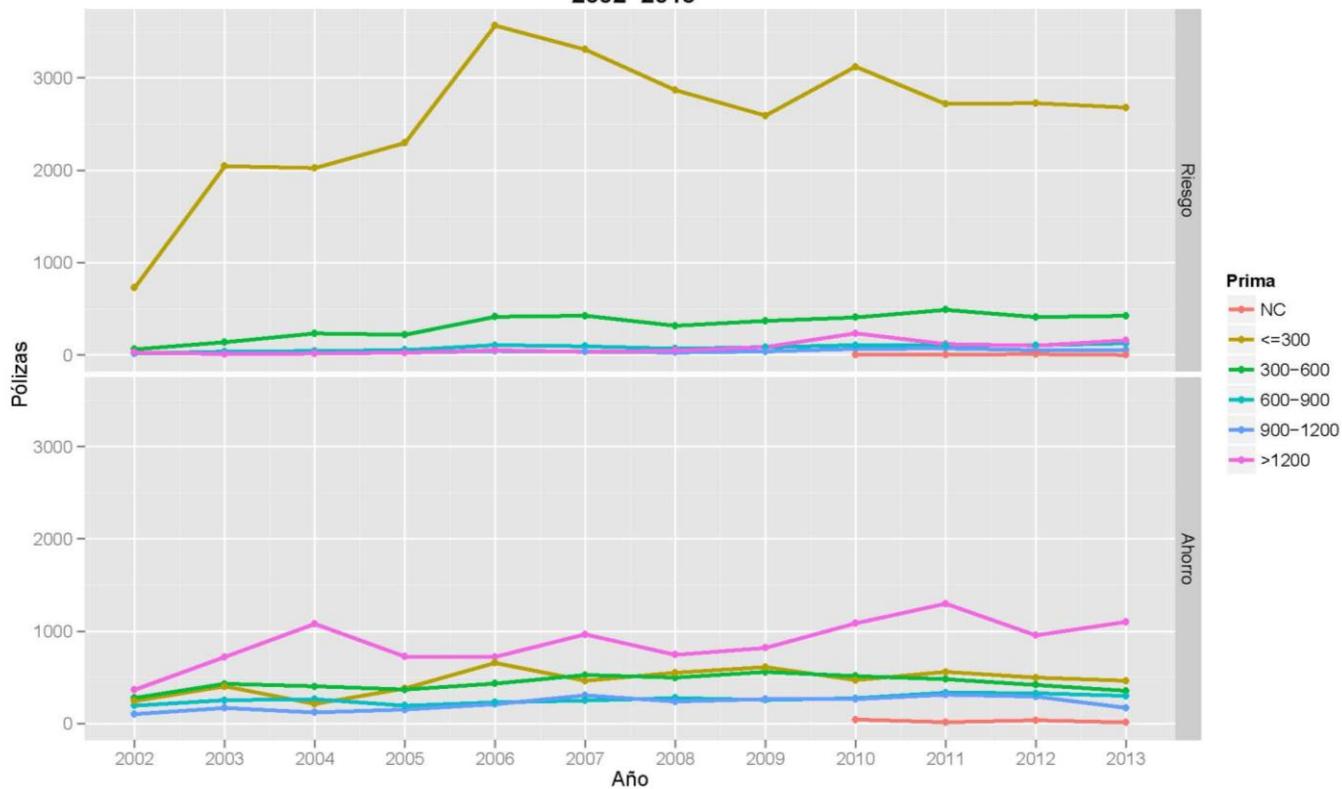
En este apartado se estudian los perfiles prima capital duración en las diferentes modalidades.

En primer lugar se observa cierta estabilidad en el porcentaje relativo de las categorías de prima, duración y capital a lo largo de los años. Es difícil establecer patrones de evolución en estas variables por separado, debido a su interrelación.

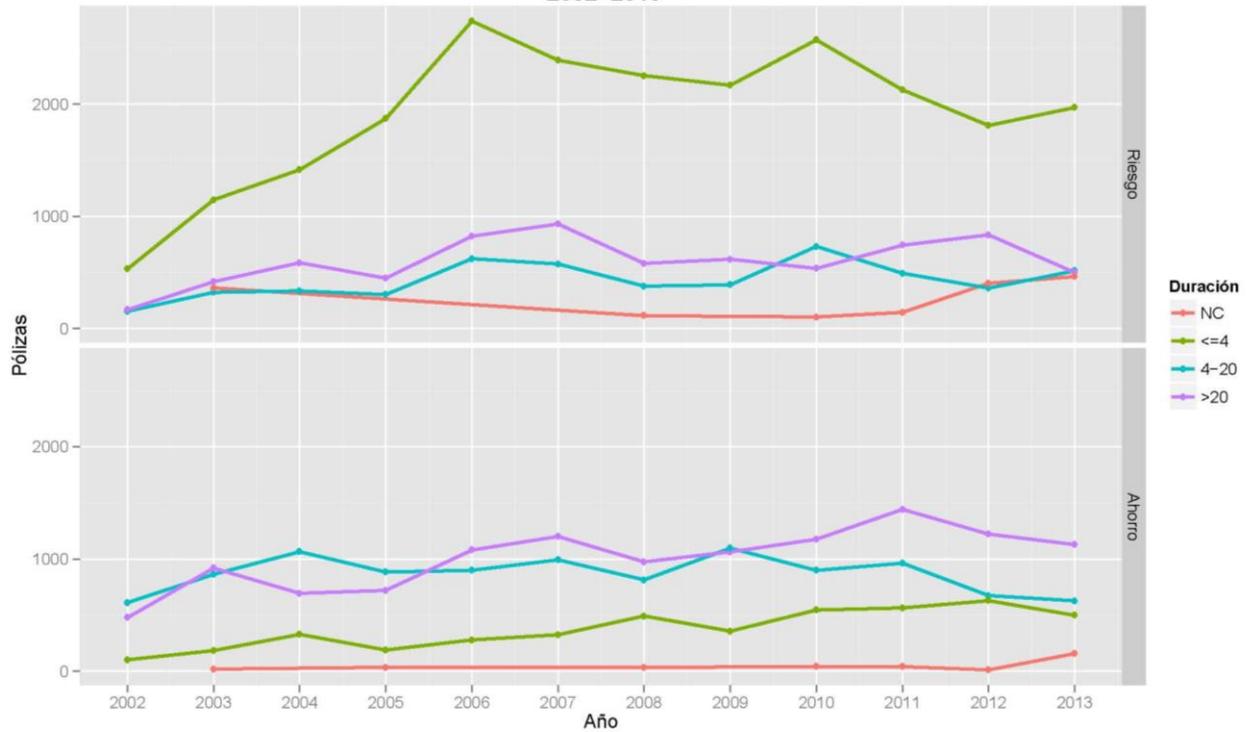




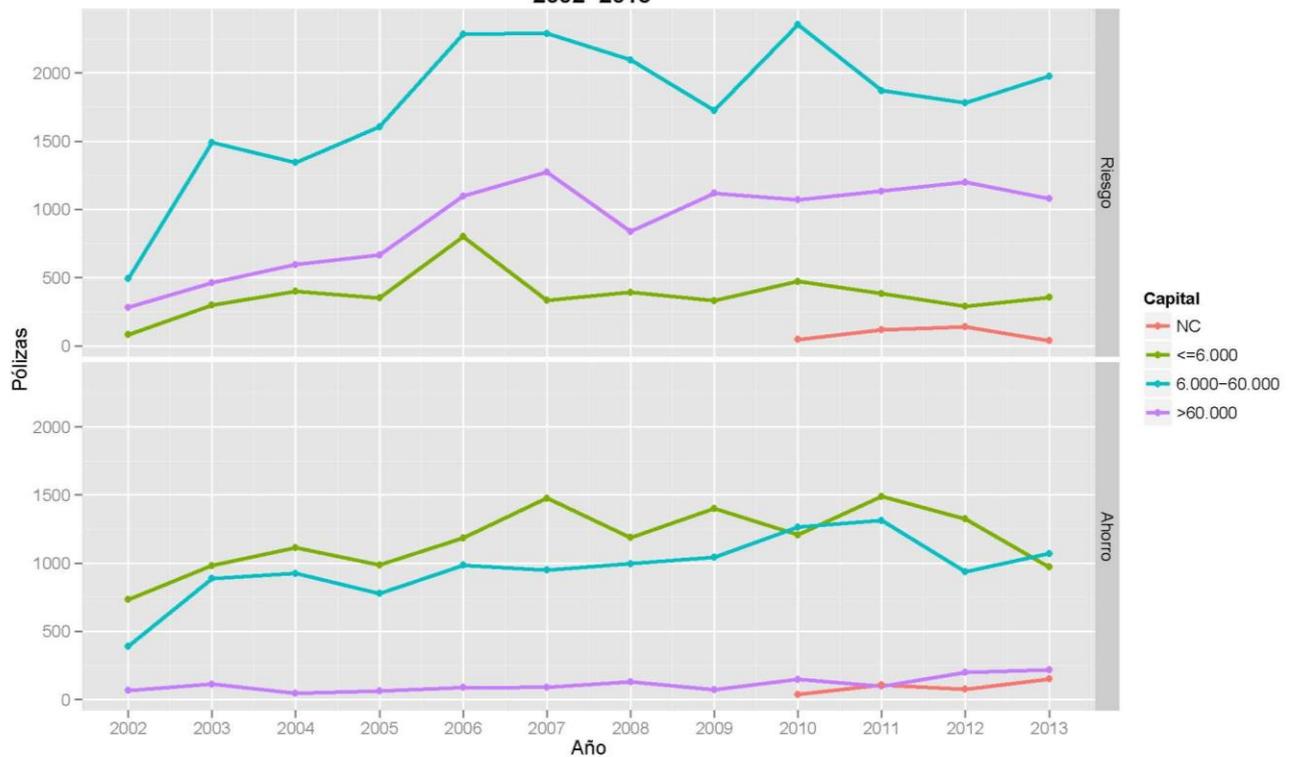
**Evolución del número de pólizas por Prima y Modalidad 2002-2013**



**Evolución del número de pólizas por Duración y Modalidad  
2002-2013**

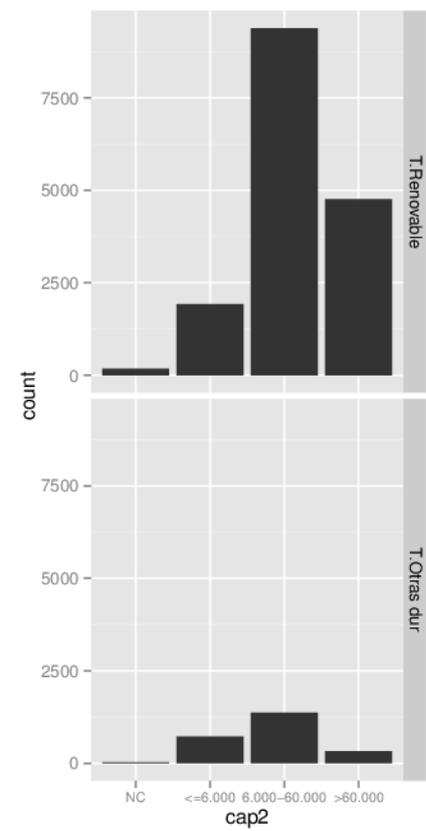
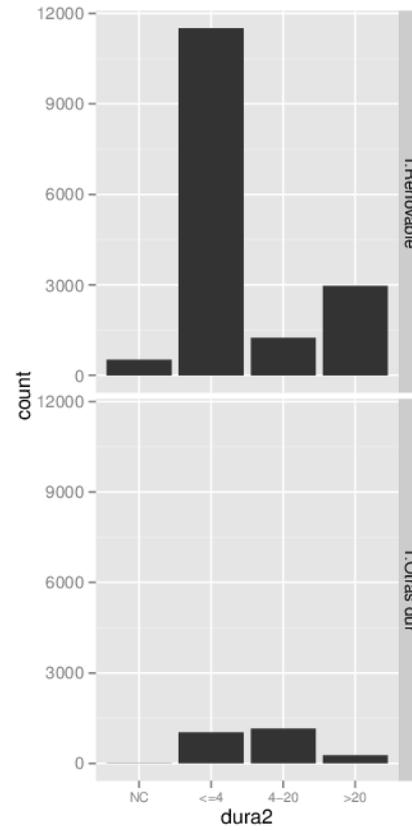
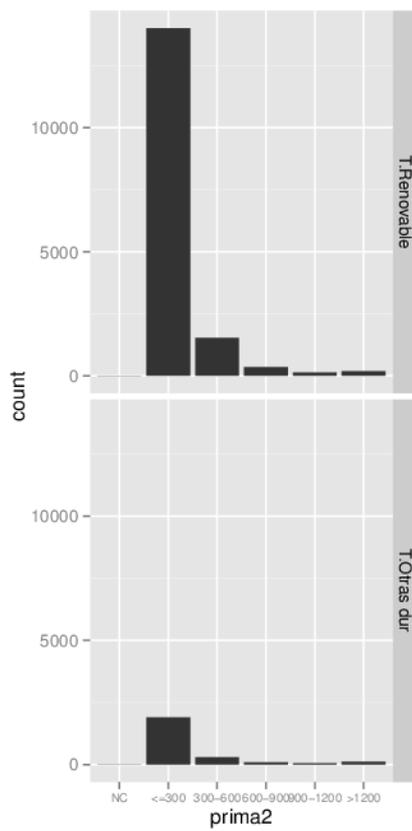
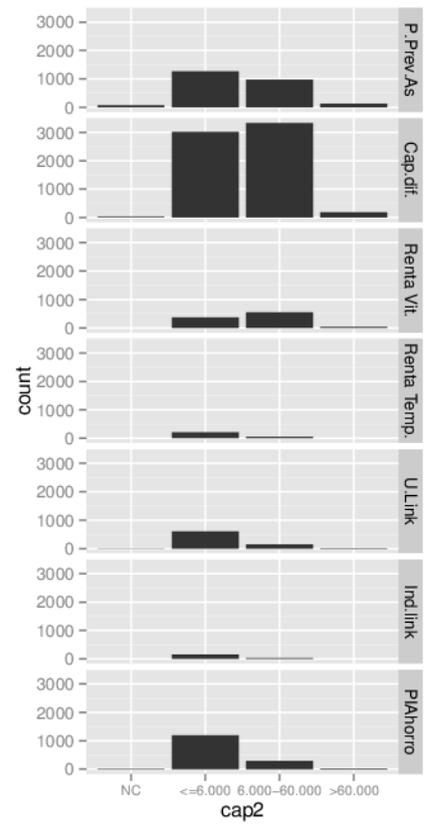
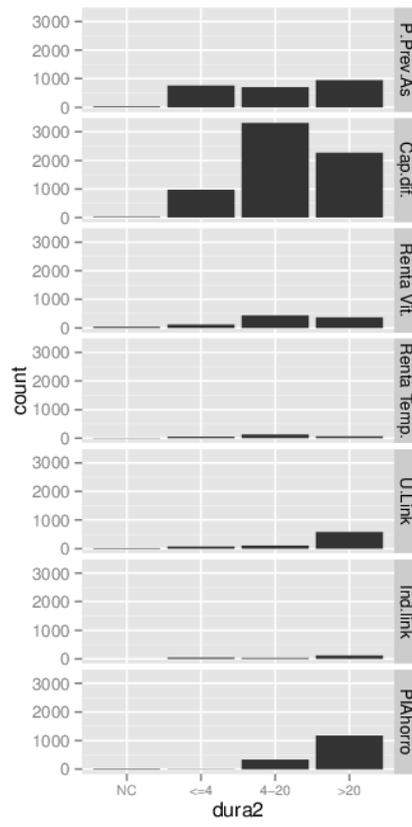
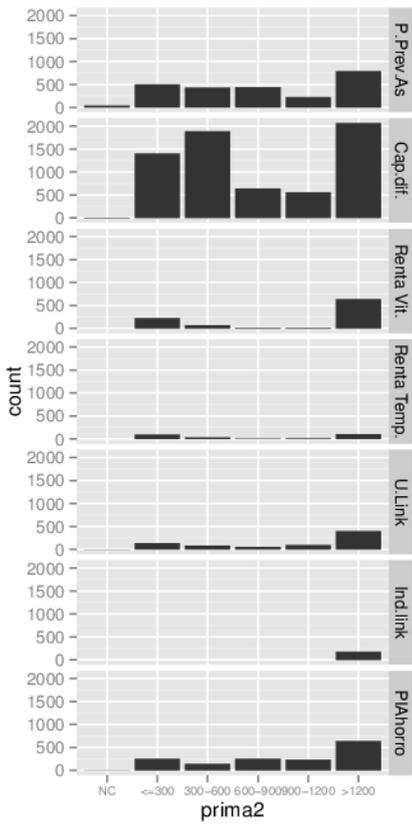


**Evolución del número de pólizas por Capital y Modalidad  
2002-2013**



Mucho más relevante es segregar por modalidad esta información, ya sin tener en cuenta los años. Se observan las diferentes frecuencias relativas de prima, duración y capital por separado según cada modalidad.

# Ahorro



## Combinaciones de prima, duración y capital

Aún dentro de cada modalidad, existe cierta dispersión en los valores de las tres variables. Cada combinación de las tres da lugar a un producto diferente. Es necesario estudiar la distribución de las diferentes combinaciones. En las siguientes gráficas se ha creado la variable código que corresponde a la combinación de las categorías de prima, duración y capital, para estudiar si en cada modalidad existen productos “estrella” o al menos cierta preponderancia de unos productos sobre otros. Dada la siguiente categorización:

prima2	1	<=300
	2	300-600
	3	600-900
	4	900-1200
	5	>1200
dura2	1	<=4
	2	4-20
	3	>20
cap2	1	<=6.000
	2	6.000-60.000
	3	>60.000

El código consiste en 3 números que indican respectivamente la categoría en prima2, dura2 y cap2. Por ejemplo, en Renta Vitalicia, el producto más frecuente tiene el **código 532**, que consiste en prima2=5, dura2=3, cap2=2, o sea,

prima>1200  
duración>20 años  
capital entre 6.000 y 60.000

Como resumen, se tiene:

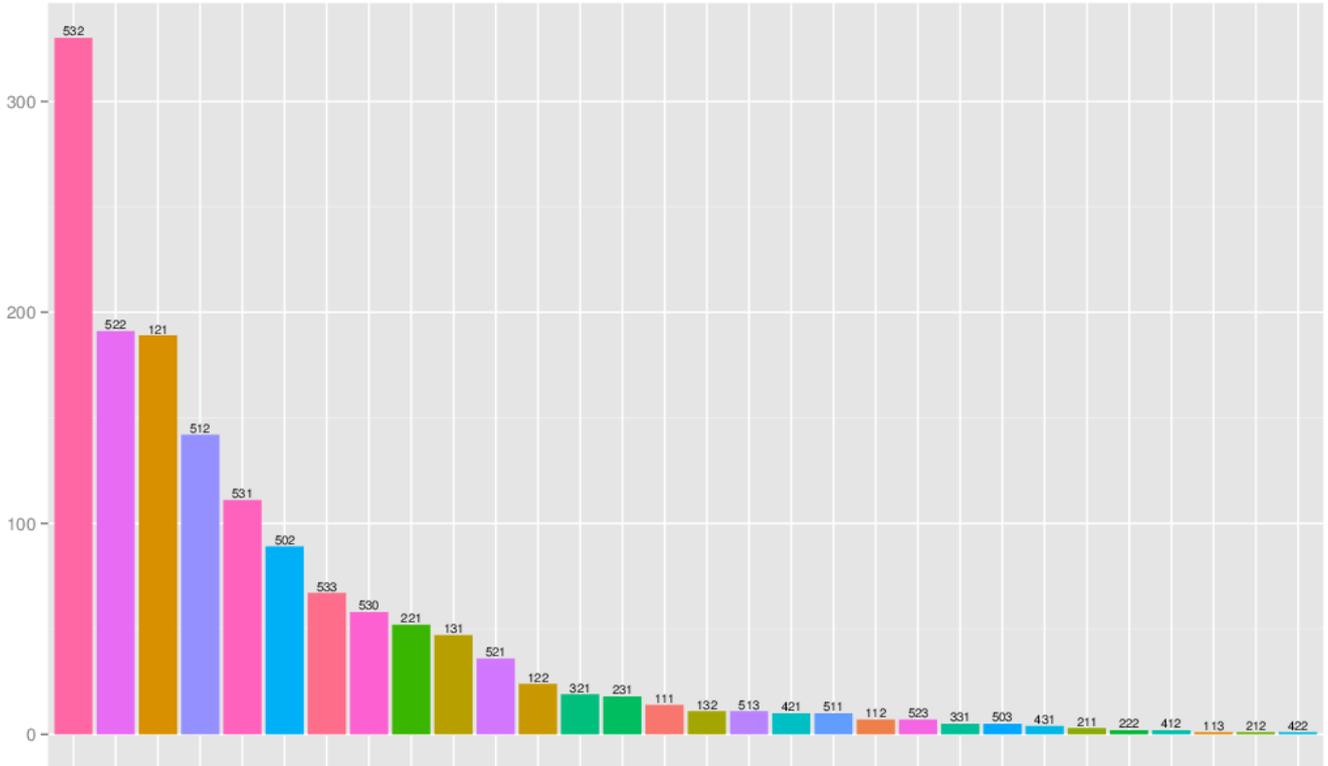
### Ahorro

- En capital diferido hay cierta dispersión de productos. Con poca diferencia entre ellos, las pólizas más frecuentes son tipo 232 (prima media-baja, duración alta, capital medio) y 522 (prima alta y duración y capital medios).
- En Ind. link, U. Link y PIAhorro el tipo de póliza más frecuente con cierta diferencia, es 531 (prima alta, duración alta, capital bajo).
- En Renta Vitalicia, las pólizas más frecuentes con cierta diferencia son 532 (prima alta, duración alta, capital medio).
- En Renta Temporal, las pólizas más frecuentes con cierta diferencia son 521 (prima alta, duración media, capital bajo).

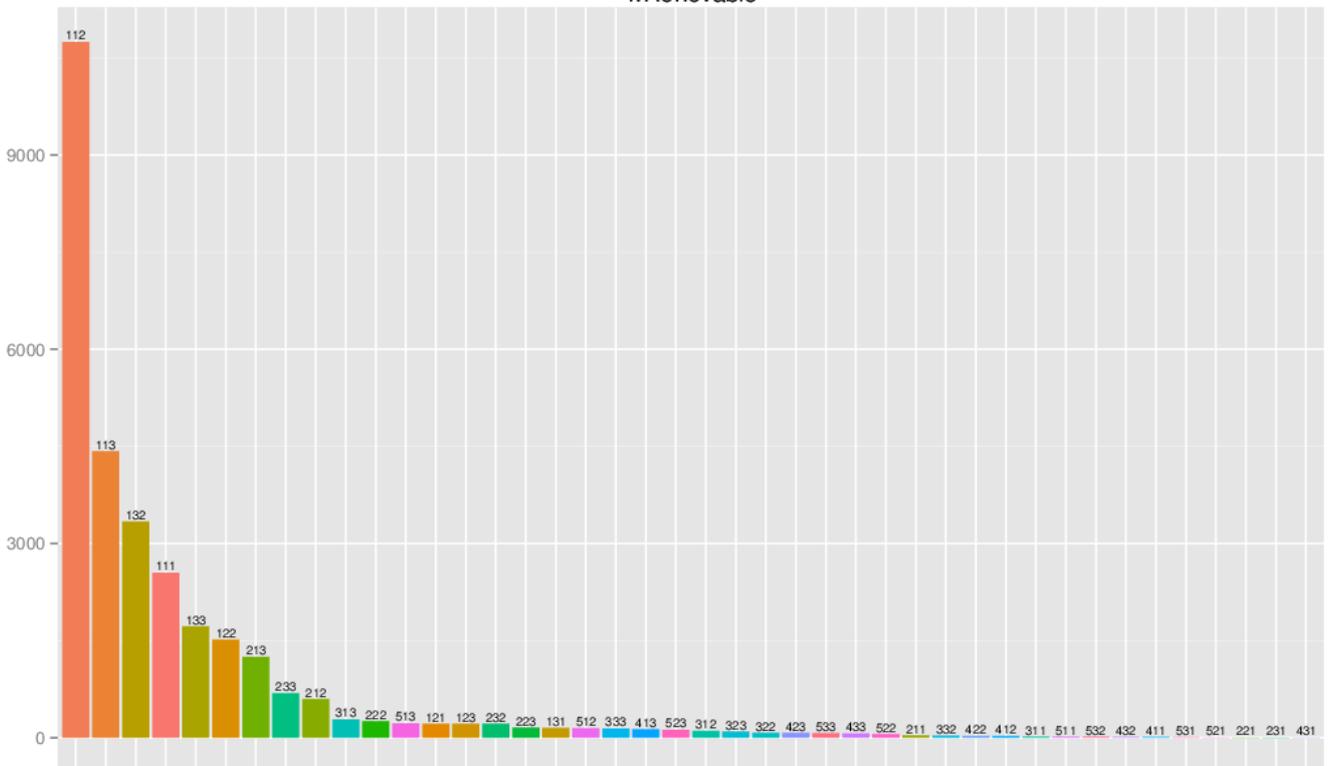
### Riesgo

- En Temporal Renovable, la combinación más frecuente con cierta diferencia es 112 (prima baja, duración baja, capital medio).
- En Temporal otras duraciones, las diferencias entre tipos de pólizas son menores, siendo la más frecuente 122 (prima baja, duración media, capital medio).
- Por fluidez de lectura, solo se presentan aquí las modalidades Renta Vital. de Ahorro y Temporal Renovable de Riesgo. El resto se pueden consultar en el Anexo.

Renta Vit.



T.Renovable

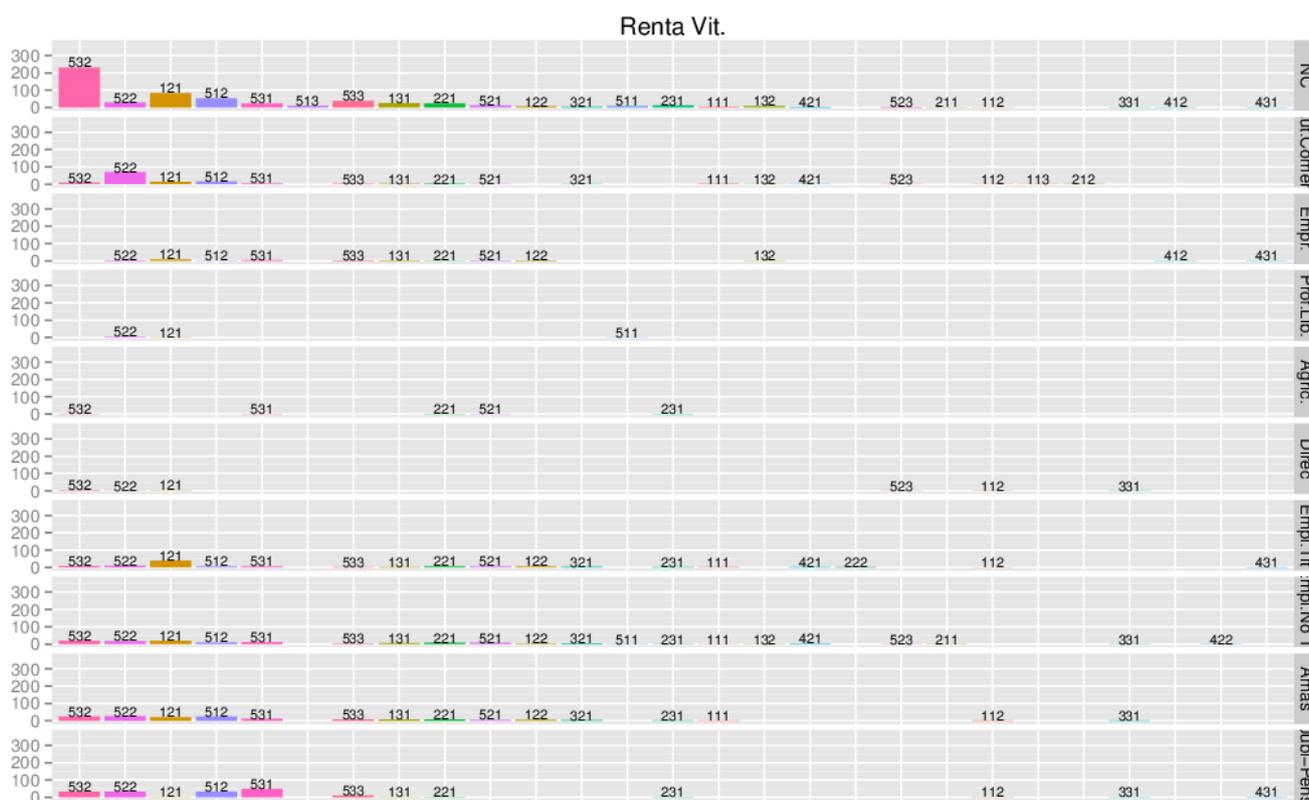


## Perfiles de póliza más contratadas por profesión.

Dada la gran cantidad de combinaciones-tipo de póliza, incluso dentro de cada modalidad, es interesante reducir la información estudiando qué profesiones contratan más o menos qué tipos de pólizas.

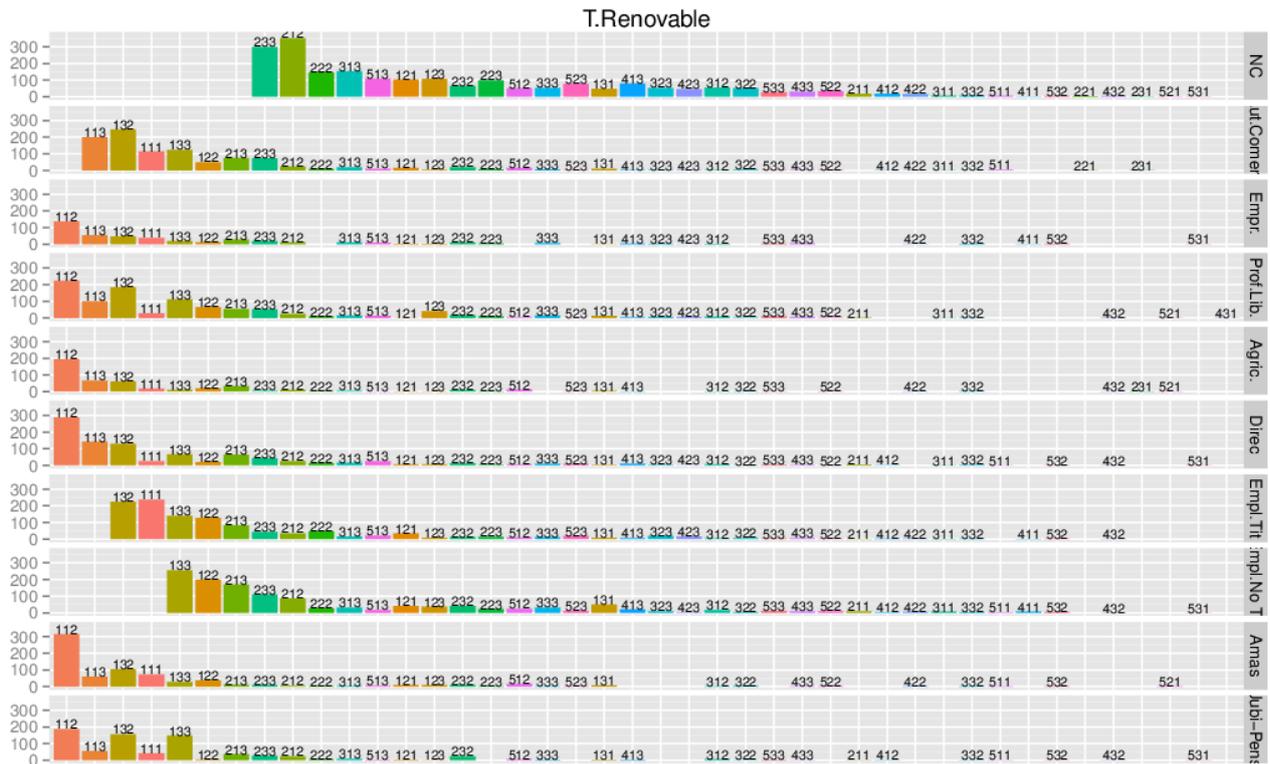
Para estudiar para cada profesión qué productos son los más contratados, se realizan diagramas de barras que dan resultados muy interesantes (solo se presentan algunos por fluidez en el documento; el resto se puede consultar en el Anexo).

El gráfico siguiente muestra la distribución de la contratación de productos en la modalidad Renta Vital. de Ahorro, por profesiones. Se observa que pólizas 532 (prima alta, duración alta, capital medio) son las más contratadas en profesión desconocida NC, pero Autónomos y Comerciantes suelen contratar más pólizas 522 (prima alta, duración media, capital medio) y Empleados Titulados pólizas con código 121 (prima baja, duración media, capital bajo).



Como ejemplo en Riesgo, en la categoría Temporal Renovable se observa cómo mientras en profesión desconocida los productos más frecuentes son 212 y 233, ambos de prima media, en las otras profesiones el producto más contratado es 121 (prima baja, duración media, capital bajo).

Esta diferencia de preferencias nos indica que quizá la categoría de profesión NC, que ya vimos que tenía una evolución diferenciada respecto de otras categorías, representa un segmento de población en sí misma y posiblemente pueda no contener demasiadas pólizas de otras profesiones que han dejado el campo missing, sino que NC indique parados, economía sumergida u otras profesiones no contempladas.



En segundo lugar se plantea la cuestión de cómo los diferentes productos se reparten entre las profesiones. Para ello se utilizan técnicas de segmentación (árboles) donde se plantea como variable dependiente la profesión y como variables independientes tanto código como prima2, dura2 y cap2. El resultado es interesante en cuanto a la detección de ciertos patrones de comportamiento por profesiones.

Como habitualmente, se diferencia el análisis por modalidad.

## Ahorro

### P. Prev. as.

Si la duración es baja, prima bajo, capital bajo (código 111):

alto porcentaje relativo de NC (44%), empleados no titulados (23%) y Empresarios (7.5%).

Con duración baja, exceptuando el código 111, hay cerca de un 98% de profesión NC.

Con duración media alta, los % se reparten de otra manera.

## Renta Vitalicia

En esta modalidad hay un desigual reparto en cuanto a las profesiones y a los tipos de pólizas. Por ejemplo, con prima alta y los códigos 511 512 531, hay una alta participación relativa de la profesión jubilados-pensionistas, mientras que con los códigos 113 212 522 hay una alta participación relativa de la profesión autónomos y comerciantes.

## En Renta Temporal

En esta modalidad hay tres tipologías de pólizas muy diferenciadas:

- Capital medio, con alta participación de profesión NC
- Prima 4 (media-alta) y capital bajo o alto, con participación relativa de Autónomos (20%), Directivos (10%), Empleados titulados (25%), amas de casa(12%).
- Resto de pólizas, con participación de NC, empleados no titulados, Amas de casa y jubilados-pensionistas.

## Riesgo

### Temporal Renovable

Si la duración es alta, se reparten las pólizas mayoritariamente entre NC(45%), empleados no titulados (18%), Autónomos-comerciantes(7.5),y empleados titulados(7%).

En caso de duración baja o media, la mayoría es profesión NC y empleados no titulados.

En otras modalidades de Ahorro y Riesgo se dan datos interesantes a investigar (ver anexo), con diferentes conjuntos de tipologías de pólizas accedidos por diferentes porcentajes de profesión.

## 7 Conclusiones

La evolución del número de pólizas en el periodo de crisis sigue en general una pauta ascendente que alcanza un máximo en 2010 y es descendente a partir de entonces. Sin embargo, se observa que la evolución del total de pólizas en este periodo puede llegar a ser muy diferente según la modalidad de póliza suscrita, y el segmento poblacional de edad y profesión. Además, se observa que el comportamiento en crisis y antes de ella no es constante en homogéneo, con lo cual análisis estadísticos que agreguen datos en dos periodos “antes de la crisis” y “durante toda la crisis” no estarían bien planteados, hay que tener en cuenta la serie en su totalidad.

La proporción de pólizas suscritas por modalidad, sin tener en cuenta otras variables como edad o profesión, parece respetarse a lo largo del periodo estudiado, siendo en Riesgo la modalidad Temporal Renovable la más solicitada, y en Ahorro Capital Diferido, Plan de Previsión Asegurado y PIAhorro, aunque esta última modalidad sólo está disponible a partir de 2007.

Si tenemos en cuenta el grupo de edad dentro de cada modalidad, existen diferencias en cuanto a la evolución del número de pólizas, sobre todo en lo que respecta a los grupos de edad de menos de 25 años y alternativamente, de más de 64 años, comportándose estos grupos a contracorriente de la tónica general en algunas modalidades.

Si se tiene en cuenta el campo profesión, se advierte una gran presencia de la categoría NC (profesión desconocida) respecto a otras profesiones. La evolución de esta categoría es característica, en general creciente hasta 2010, decreciente después y creciente en 2012 en lo que parece un trasvase de pólizas desde la profesión Empleados No titulados.

Este campo es difícil de interpretar; podría deberse a simple falta de información (es decir, valores ausentes en el campo profesión correspondientes a pólizas de personas que realmente trabajan en alguna de las categorías de profesión), pero parece más una categoría propia en sí (por ejemplo, parados y/o profesionales pertenecientes a ciertas áreas propias de economía sumergida). Las razones para considerarlo así son varias:

- Evolución diferenciada de otros grupos de profesión
- En el análisis de las entidades participantes en la encuesta, se observa que prácticamente todas tienen una gran porcentaje de NC, incluso las entidades más habituales.
- Al estudiar los productos contratados por cada grupo de profesión, se observa que los productos contratados por el grupo NC son a menudo muy diferentes de los productos contratados por otras profesiones o por la media de éstas.

Se hace necesario estudiar si existen grupos de edad-profesión con una evolución general similar en tiempo de crisis. A través de un análisis cluster se diferencian 10

clusters o grupos de interés en Riesgo y 6 en Ahorro . Los clusters creados no son exhaustivos, es decir, no cubren todas las pólizas, son solamente aquellos grupos que tienen un comportamiento homogéneo en su evolución en crisis y diferenciado de otros grupos.

En estos clusters se observan a veces grandes diferencias de evolución; por ejemplo, el cluster de Riesgo Profesión NC-Temporal Renovable, <=34 años tiene una evolución claramente decreciente, mientras que el cluster Profesión NC-Temporal Otras duraciones, 25-54 años tiene una evolución creciente. Dado que no en la base de datos no hay seguimiento de pólizas individuales, se necesitaría esta información para corroborar, por ejemplo, si existen trasvases de unos tipos de modalidad contratada a otros, por edad y profesión.

Por otro lado, a través de árboles de decisión se pueden estudiar las preferencias por parte de grupos de edad y profesión, por ciertas modalidades. Se descubre por ejemplo, que en Riesgo, sin ser profesión NC, casi todos los grupos de edad eligen Temporal Renovable en un 90%. En Ahorro se descubre un comportamiento respecto a la selección de de modalidad característico de Autónomos y Comerciantes >64 años, que escogen Renta vitalicia en un 70% y Capital Diferido en un 18%. Otras relaciones se pueden ver más concretamente en el Anexo.

El tamaño de la localidad no parece ser indicativo de diferentes evoluciones o preferencias según nuestros estudios.

Otro aspecto de interés es estudiar, dentro de cada modalidad, cómo son los patrones de selección de pólizas en cuanto a prima, duración y capital (estas tres variables se han codificado respectivamente en 5,3 y 3 categorías). Se observa que, mientras para algunas modalidades (Capital diferido) hay gran variedad de productos diferentes contratados, en algunas modalidades dominan algunos productos. Por ejemplo, en U. link se dan más frecuentemente pólizas con prima y duración alta, y capital bajo.

De gran interés es investigar las preferencias de cada profesión por diferentes productos dentro de cada modalidad. Por ejemplo, en Temporal Renovable el producto más contratado en la profesión NC con diferencia es con prima y capital medios y duración baja, mientras que en otras profesiones suelen ser más frecuentes prima y duración bajos, capital medio o bien prima baja, duración alta y capital medio. También se descubren mediante árboles, grupos de productos que son seleccionados mayoritariamente por ciertos grupos de profesión.