

# Pronóstico de contrataciones de préstamos bancarios

*Abraham Nieto 51556, Alejandro Hernández 87806 y Omar Reyes 127131*

## Introducción

Es bien sabido que el otorgamiento de créditos es uno de los negocios más rentables para los bancos, es por ello que estos asignan mucho presupuesto a la generación de campañas para la colocación de estos dirigidas a clientes y no clientes de tal modo que el objetivo, en el primer caso, además de acrecentar el negocio también es incrementar la fidelidad del cliente con el banco y en el segundo caso hablamos de atraer nuevos clientes con ofertas crediticias

Por supuesto, se cuenta con información de una muestra de 64mil clientes donde históricamente algunos han contratado un préstamo personal a lo largo de los últimos 2 años en cada uno de los meses estas contrataciones se han hecho debido a que el cliente ha recibido una oferta del banco o bien de modo proactivo el cliente lo ha solicitado.

Uno de los principales objetivos es crear campañas dirigidas donde esto significa una personalización en el modo de comunicarle la oferta y esto es posible ya que se cuenta con datos como saldos, gastos, créditos, productos bancarios, movimientos, etc. del cliente, pero antes de llevar a cabo una campaña, en términos de negocio, es necesario establecer las metas que, como área de colocación de créditos, debemos lograr y para poder establecer estas metas necesitamos conocer la historia de las adquisiciones de créditos personales y además poder conocer el número esperado de estos, en periodos posteriores, de tal modo que este dato se pueda utilizar como referencia de cuales son los posibles escenarios en cuanto a los resultados que se pudiesen obtener desde el punto de vista de los datos con los que se cuentan, entonces utilizando las predicciones del número de créditos que se van a colocar cada mes podemos establecer una meta de colocación de créditos para campañas dirigidas bajo el supuesto de que todos los clientes cumplen las condiciones para dicho otorgamiento.

Una de las hipótesis más comunes es que para la adquisición de un préstamo bancario hay una variable, además de la posible necesidad financiera, que desempeña un papel muy importante está es la etapa de vida de la gente, es decir, por ejemplo una persona que es el sustento de una familia no tiene ni los mismos gastos ni necesidades que una persona joven que está comenzando su vida laboral así como alguien que ya está jubilado es por eso, que es importante detectar a cuales de estos segmentos, desde el punto de vista comercial, es más ‘fácil’ ofrecerles un crédito y que tengan mayor tasa de aceptación.

## Descripción del problema

Se quiere establecer un proceso, mediante un modelo, mensual que pronostique el número de contrataciones de préstamos personales con el fin de establecer la meta mensual de colocaciones de créditos personales a través de campañas, para esto se busca construir un modelo que pronostique el número de préstamos colocados esperado, para ello contamos con 24 meses de historia (Agosto 2016-Agosto 2018) donde contamos con un dataset de 216 registros agrupados por mes y por ciclo de vida de nuestros clientes entonces para cada mes contamos el sumario de nuestros datos por ciclo de vida y contamos con la variable target que representa el número de préstamos contratados en cada uno de los ciclos de vida por mes está será la variable a predecir, además se cuenta con 5 variables adicionales que es necesario probar que tengan o no influencia en el pronóstico: gastos promedio, número de productos bancarios promedio, ratios de endeudamiento.

## Objetivos

- Construir un modelo que pronostique ,puntualmente y por intervalo, el número de préstamos que serán contratados en los siguientes 3 meses y evaluar su desempeño.

- Determinar si existe diferencia significativa en las contrataciones de préstamos por ciclo de vida, es decir, si existe algún o algunos segmentos que tengan mayor influencia en el número de contrataciones totales con el fin de identificar si las estrategias debieran ir dirigidas a ciertos segmentos.

## Análisis exploratorio de datos

Las variables presentes en la base de datos son las siguientes:

- **TARGET:** Número de créditos contratados. Es un entero entre 13 y 150. Esta es nuestra variable respuesta y la renombraremos como  $y$
- **ID:** Número de clientes. Es un entero entre 1939 y 8436.
- **FH\_REF:** Fecha de referencia, inicia el 01-08-2016 y finaliza el 01-07-2018.
- **TP\_SEGMENTO\_FINAL2:** Ciclo de vida de los clientes (ADULTO EN PLENITUD, ADULTOS INDEPENDIENTES, DIVORCIADO, HOGARES CON HIJOS, JOVEN PROFESIONAL, JOVEN TRABAJADOR, PAREJA JOVEN, PAREJAS ADULTAS y PAREJAS SENIOR). La renombraremos como  $x_1$ .
- **TO\_PROM\_TO\_CARGOS\_3M:** Número promedio de cargos en los últimos 3 meses. Es un número real entre 2.75 y 12.96. La renombraremos como  $x_2$ .
- **NU\_VINC\_COGNODATA:** Número promedio de productos bancarios. Es un número real entre 1.57 y 4.20. La renombraremos como  $x_3$ .
- **TO\_NECESIDAD\_FINAN\_CAP\_3M:** Cociente del gasto promedio y el saldo. Es un número real entre -1367.4 y 843804.73. La renombraremos como  $x_4$ .
- **IM\_PROM\_GASTOS\_3M:** Importe promedio de gastos (compras+cargos recurrentes) en los últimos 3 meses. Es un número real entre 24057.97 y 834859.95. La renombraremos como  $x_5$ .
- **IM\_SUM\_SDO\_CORTE\_1M:** Importe promedio del saldo de cuenta de cheques. Es un número real entre 828293.4 y 3364972.9. La renombraremos como  $x_6$ .

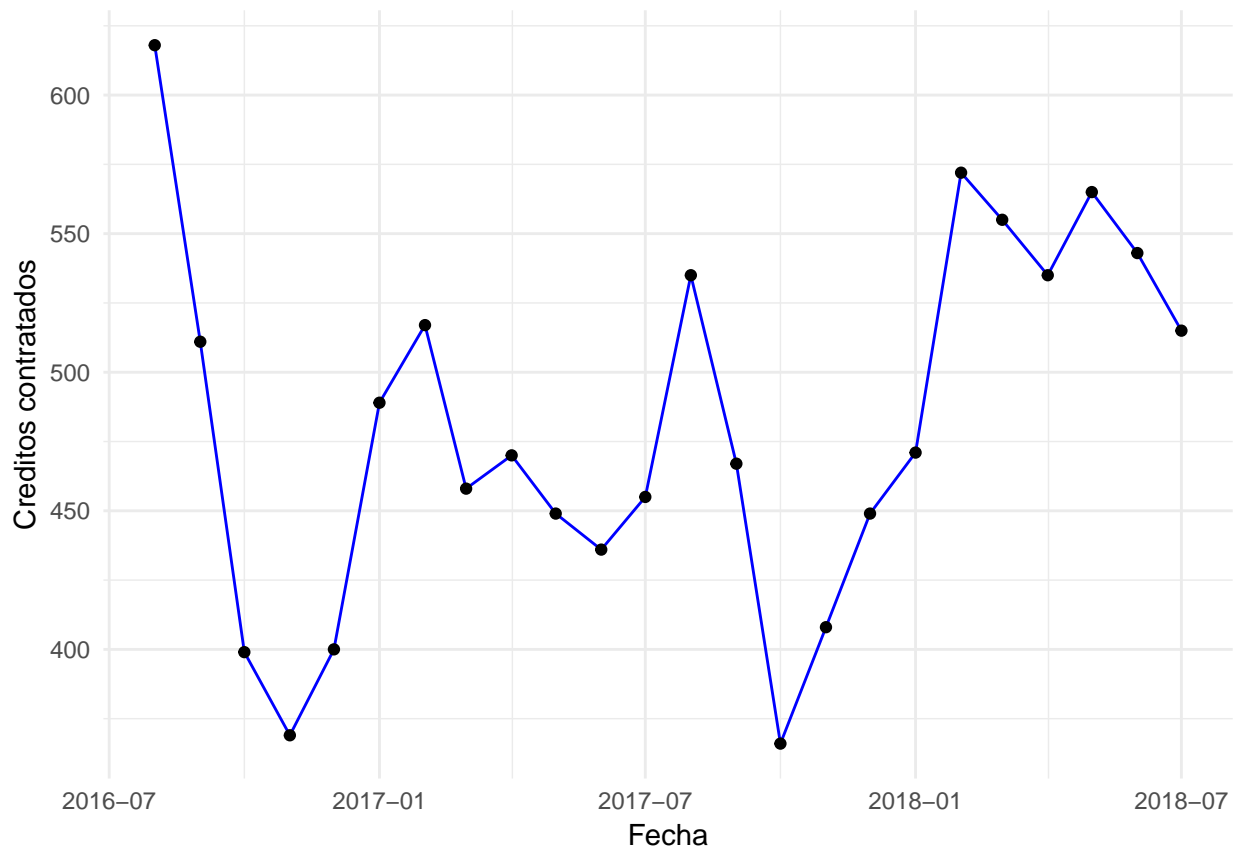
Es importante destacar que como la variable  $x_1$  es una variable categórica tenemos que incorporar la restricción para los sus coeficientes asociados, lo cual se detallará en cada uno de los modelos.

Por otro lado las variables  $x_2, \dots, x_6$  están en escalas con distintos órdenes de magnitud, por lo que las escalamos.

A continuación se muestra un análisis exploratorio de las variables en el dataset. El objetivo es obtener una perspectiva de que tipo de relaciones hay entre ellas. El primer análisis es de las ventas contra el tiempo, para identificar tendencia o estacionalidad en los datos para los 9 segmentos diferentes que son objeto de análisis.

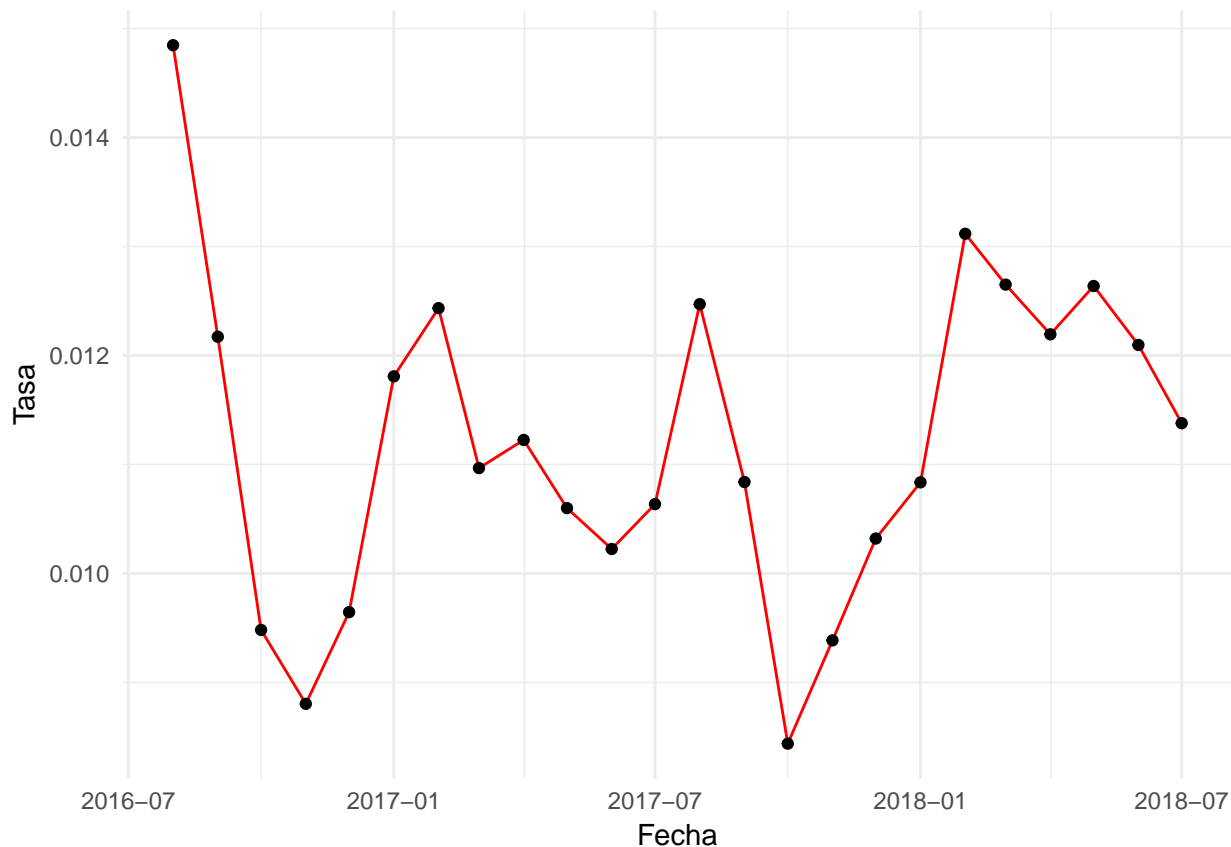
## Totales

En la siguiente gráfica se observa el número de contrataciones totales por mes, podemos ver que en agosto de 2016 tenemos el mayor número histórico de contrataciones (más de 600), debido a algunas estrategias comerciales en ese año, por otro lado aunque la estacionalidad no es del todo marcada exactamente por mes si se puede observar por periodos, donde tenemos caídas muy claras en las contrataciones de los últimos meses del año de octubre a diciembre, muchas veces en este inter la gente tiene más liquidez, al comienzo de año hay un repunte importante sobre todo en Enero ya que normalmente los clientes están gastados y su necesidad financiera crece, del mismo modo en los agostos siempre hay un número importante de contrataciones debido a que en muchos casos hay regresos a clases.



Ahora observemos del mismo modo las tasas contratación, donde de los últimos meses del año tienen una tasa de contratación menor al 1% y los meses con mayor número de contrataciones superan el 1.2% de tasa de contratación.

```
ggplot(tot,aes(x=as.Date(FH_REF),y=tasa)) + geom_line(col="red") +geom_point()+
  xlab("Fecha") +
  ylab("Tasa")+theme_minimal()
```



## Segmentos

Los distintos segmentos con los que contamos y su descripción son los siguientes:

**Joven Trabajador** Clientes entre 18 a 34 años, Obreros calificados, Con ingresos de \$3,300 a \$11,000 mensuales.

**Joven Profesionalista** Clientes entre 18 a 34 años, Joven cuya profesión requiere una calificación superior para ser desempeñada, Con ingresos superiores a los \$3,200 mensuales.

**Pareja Joven** Clientes entre 18 a 34 años, Sin hijos y conviven en pareja, Con ingresos superiores a los \$5,000 mensuales.

**Hogares con hijos** Clientes entre 18 a 65 años, Conviven con sus hijos en el hogar, Con ingresos promedio de \$25,000 mensuales.

**Pareja Adulta** Clientes entre 35 a 44 años, Sin hijos en el hogar, que conviven en pareja, Con ingresos superiores a los \$7,000 mensuales.

**Pareja Senior** Clientes entre 45 a 65 años, Sin hijos en el hogar, que conviven en pareja, Con ingresos superiores a los \$7,000 mensuales.

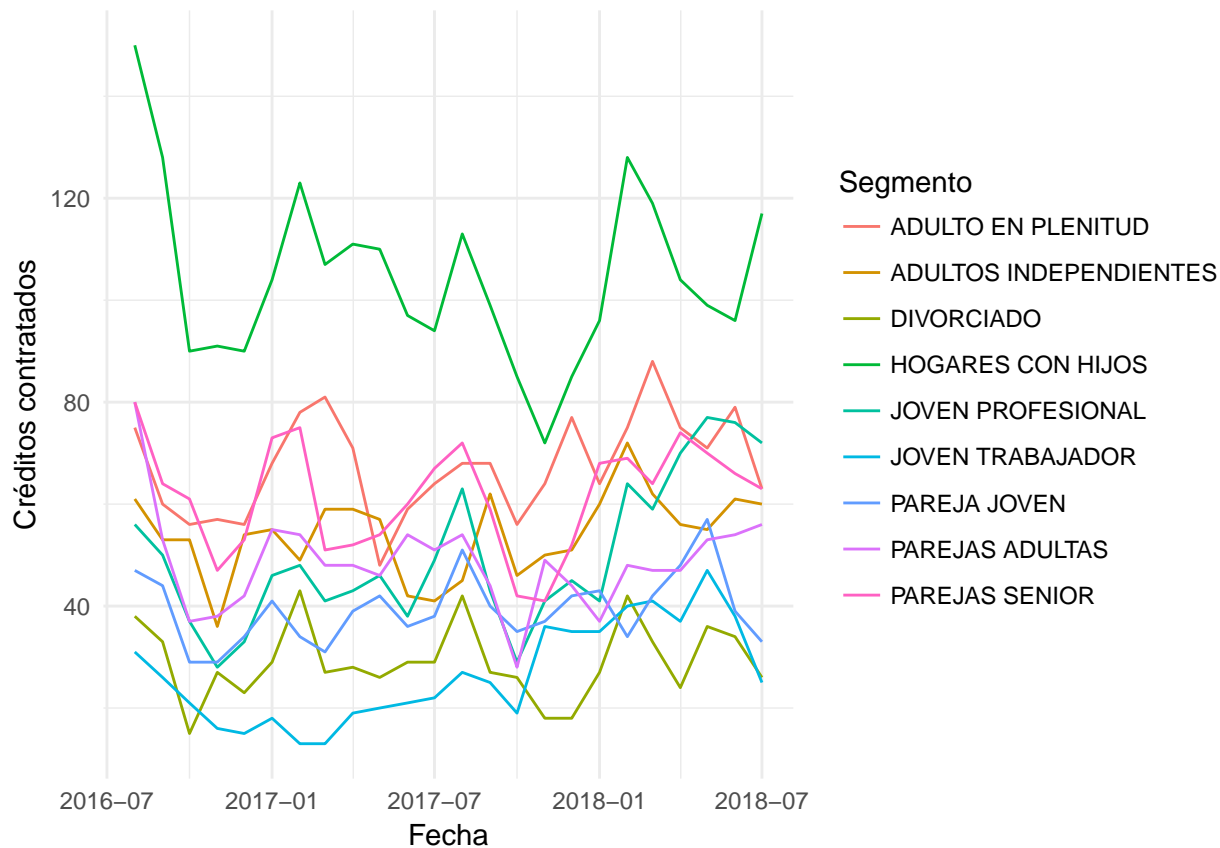
**Divorciado** Clientes entre 35 a 65 años, Divorciados.

**Adulto Independiente** Clientes entre 35 a 65 años, Sin hijos en el hogar, no conviven en pareja, Con ingresos superiores a \$6,000 mensuales.

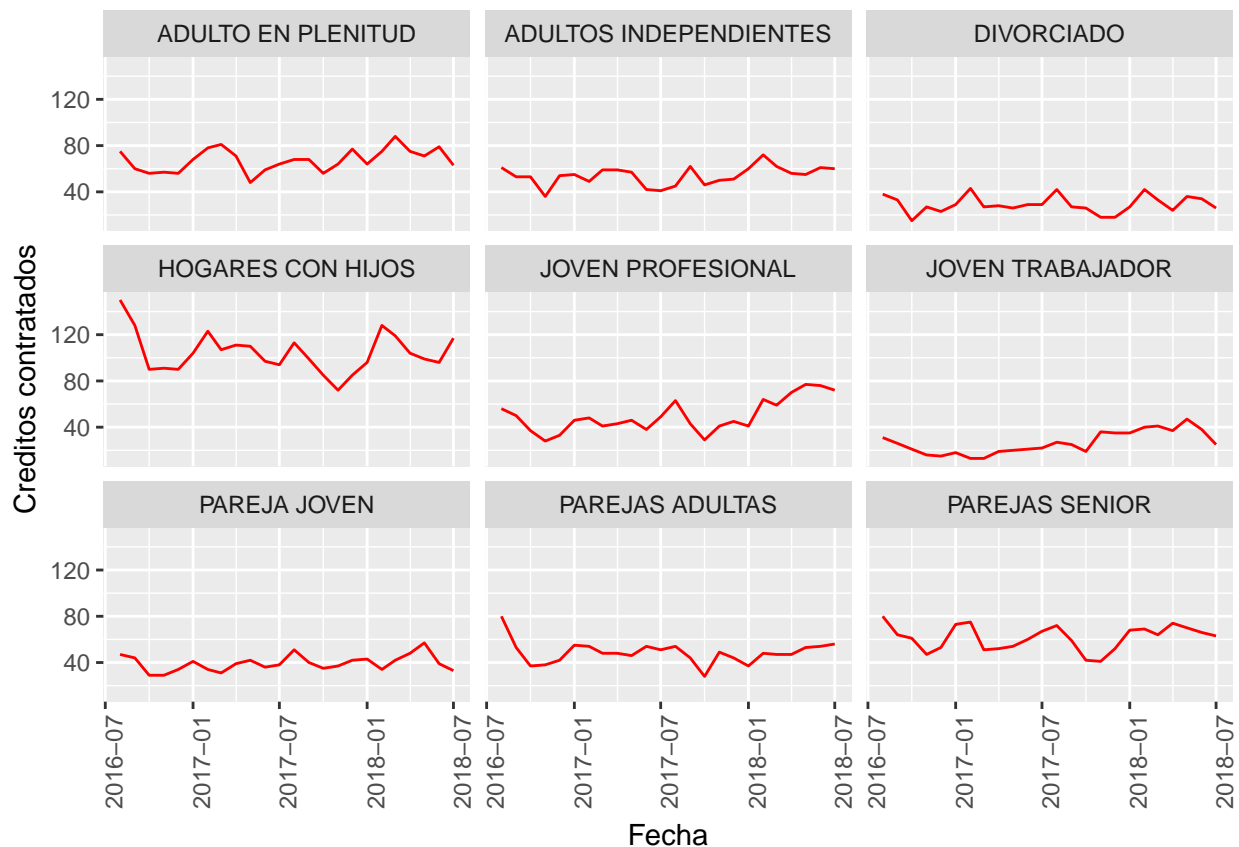
**Adulto en Plenitud** Clientes mayores a 65 años.

De forma similar, en la siguiente gráfica se puede observar la serie de número de créditos contratados por segmento.

En esta gráfica sobresalen los hogares con hijos, aunque en contraste de la tendencia general cabe resaltar que los segmentos de joven profesional y trabajador tuvieron una tendencia creciente constante a partir del 2018 lo que pudiera significar que más jóvenes cumplieron los requisitos para adquirir préstamos.



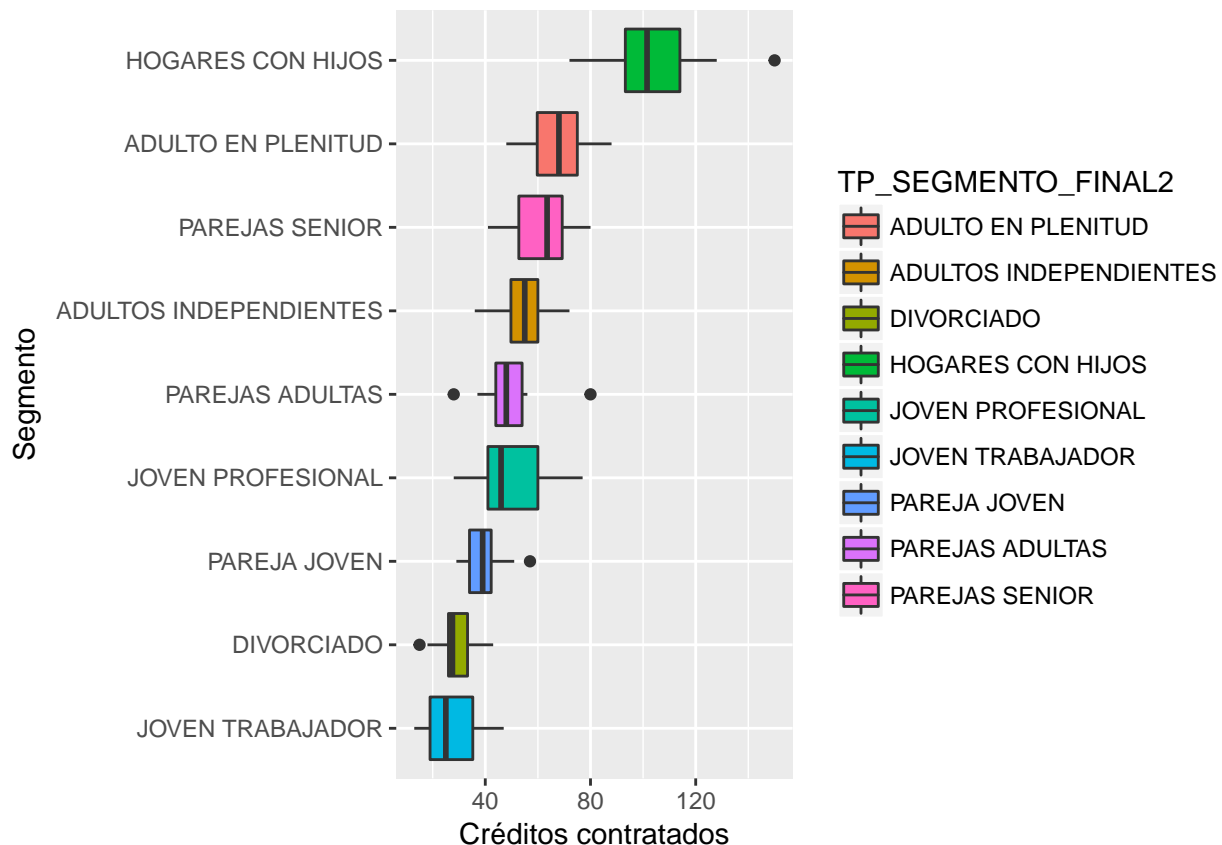
Y si separamos las tendencias:



En la gráfica anterior se muestra que el segmento de hogares con hijos es el que tiene una tendencia más parecida a la total es decir con más variaciones, entonces podemos pensar que este segmento tiene una influencia importante en las contrataciones con respecto al total por mes, en contraste las parejas jóvenes y los divorciados muestran una tendencia más constante.

Los segmentos con mayor número de contrataciones de forma mensual son Hogares con Hijos y Adultos en plenitud, y los que menos son Divorciados y jóvenes trabajadores.

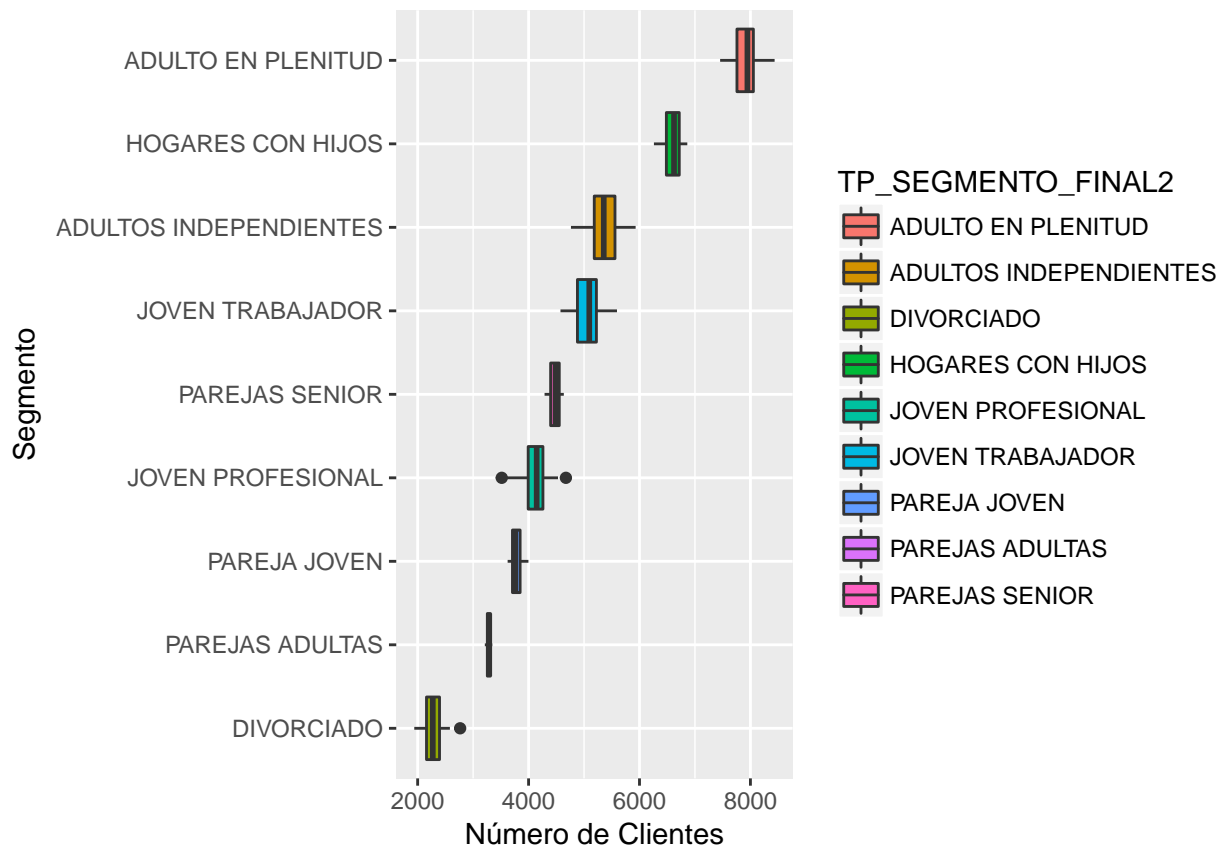
Se presenta una gráfica de caja y brazos donde se notan diferencias entre los distintos segmentos, como se había mencionado el segmento de Hogares con hijos es un grupo totalmente separado y el que tiene mayor volumen de contrataciones, al comparar su rango intercuartílico con respecto a los demás, por otro lado se puede observar que desde el segmento de adultos en plenitud hasta jóvenes profesionales al menos el 75% de los meses han sobrepasado las 40 contrataciones y existe un tercer grupo donde el volumen y la distribución de contrataciones mensuales es el más bajo, pareja joven, divorciado y joven trabajador.



En la figura anterior podemos constatar que los hogares con hijos es el segmento con mayor número de contrataciones promedio mensual y también en la mediana de los meses, después los adultos en plenitud y parejas senior tienen estadísticos similares y al final los Divorciados y jóvenes trabajadores son los segmentos con menor número de contrataciones de manera mensual.

Podemos observar es que estas distribuciones no necesariamente tienen una relación directamente proporcional al número de clientes que tenemos en cada segmento por mes, pero si nos muestra que hay segmentos a los que de forma más recurrente se les otorgan los préstamos, es el caso de los adultos en plenitud que quizás son económicamente más solventes y los hogares con hijos.

```
ggplot(camp, aes(x=fct_reorder(TP_SEGMENTO_FINAL2, ID), y=ID, fill=TP_SEGMENTO_FINAL2)) +
  geom_boxplot() + coord_flip() + xlab('Segmento') + ylab('Número de Clientes')
```



En la tabla siguiente se observan los estadísticos puntuales de cada segmento, en esta tabla se puede confirmar lo explicado en la figuras anteriores de manera puntual, los hogares con hijos es el segmento con mayor número de contrataciones promedio mensual 104, también en la mediana de los meses 101 y tiene la mayor tasa promedio de contrataciones por mes 1.58%, después los adultos en plenitud y parejas senior tienen estadísticos similares entre sí, aunque la tasa de contratación promedio de los primeros es la segunda más baja en promedio y al final los Divorciados y jóvenes trabajadores son los segmentos con menor número de contrataciones de manera mensual y estos últimos también tienen la tasa de contratación promedio mensual más baja .5%.

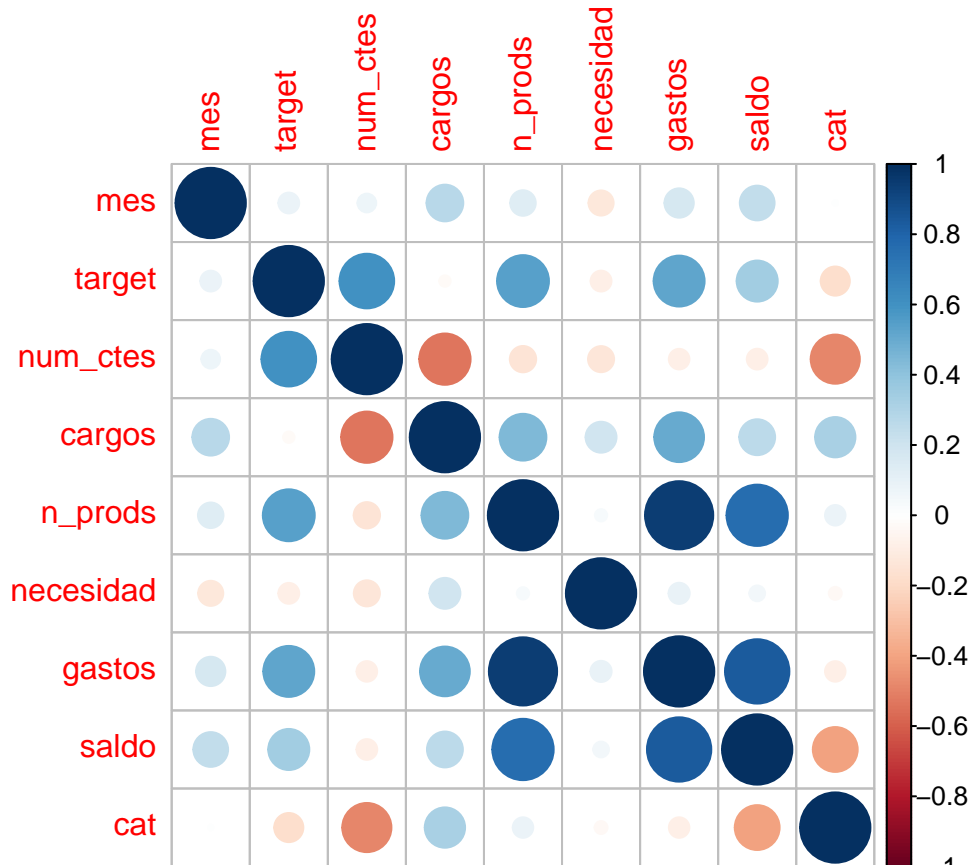
```
## # A tibble: 9 x 4
##   TP_SEGMENTO_FINAL2 mediana promedio tasa_promedio
##   <fctr>      <dbl>      <dbl>      <dbl>
## 1 HOGARES CON HIJOS    101.5  104.50000    1.5814570
## 2 ADULTO EN PLENITUD    68.0   67.54167    0.8517157
## 3 PAREJAS SENIOR       63.5   61.54167    1.3746084
## 4 ADULTOS INDEPENDIENTES 55.0   54.12500    1.0100024
## 5 JOVEN PROFESIONAL    46.0   49.79167    1.2058444
## 6 PAREJAS ADULTAS      48.0   48.62500    1.4831303
## 7 PAREJA JOVEN         39.0   39.37500    1.0410249
## 8 DIVORCIADO           27.5   29.16667    1.2739482
## 9 JOVEN TRABAJADOR     25.0   26.66667    0.5227769
```

## Variables Numéricas

Para realizar una primera exploración a la interacción de las variables, graficamos un correlograma de nuestros datos. Para hacer esto se transformó la variable de fecha a numérica:



```
library(corrplot)
campx<-camp
campx$FH_REF<-as.numeric(campx$FH_REF)
names(campx) <- c("mes","seg","target","num_ctes","cargos","n_prods","necesidad","gastos","saldo","cat")
nums <- unlist(lapply(campx, is.numeric))
corrplot(cor(campx[,nums]))
```

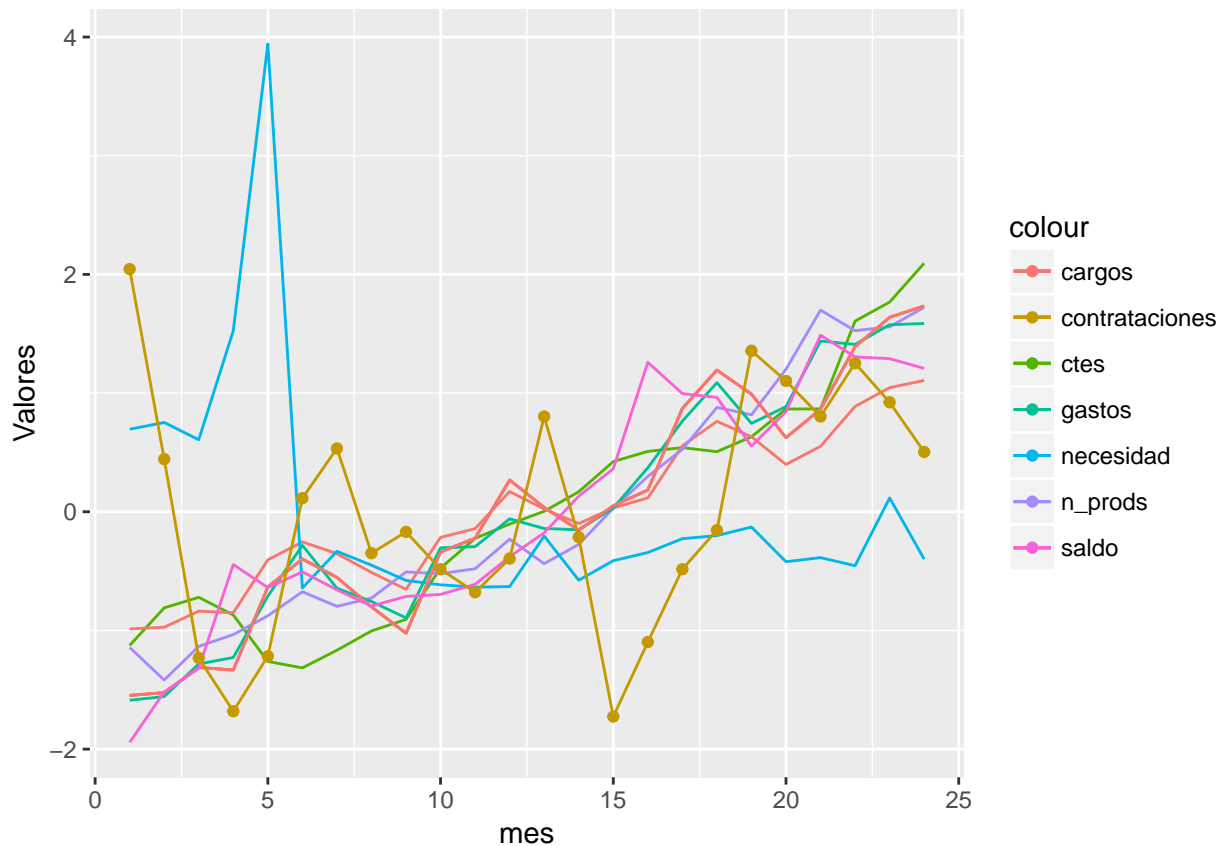


Podemos observar que las variables más relacionadas con nuestro target o número de contrataciones de créditos son el número de clientes, el número de productos y los gastos, otras correlaciones positivas que podemos ver son entre los cargos y el número de productos de forma positiva así como con los gastos, también el número de productos tiene una relación positiva con los saldos de cuenta de cheques, por otro lado vemos una relación negativa entre “cat” que es el segmento numérico que está construido con un valor numérico de forma descendente con respecto a la etapa de vida, es por ello que los saldos y esta variable tienen relación negativa (tienes más liquidez mientras más grande eres). El tiempo tiene mayor relación positiva con los cargos y los saldos lo cual suena coherente en el sentido de que los gastos son distintos en el año y por tanto la liquidez también.

Ahora podemos ver las distintas variables en el tiempo...

```
totp<-campx %>% group_by(mes)%>%summarise(total=sum(target),ctes=sum(num_ctes),cargos=mean(cargos),n_pr
#ggplot(tot,aes(x=as.Date(FH_REF))) + geom_line(y=total,col="blue")
ggplot(totp, aes(x=mes)) + geom_line(aes(y = scale(ctes), colour = "ctes")) +
  geom_line(aes(y = scale(cargos,scale = FALSE), colour = "cargos"))+
  geom_line(aes(y = scale(n_prods), colour = "n_prods"))+
  geom_line(aes(y = scale(cargos), colour = "cargos"))+
  geom_line(aes(y = scale(necesidad), colour = "necesidad"))+
  geom_line(aes(y = scale(cargos), colour = "cargos"))+
```

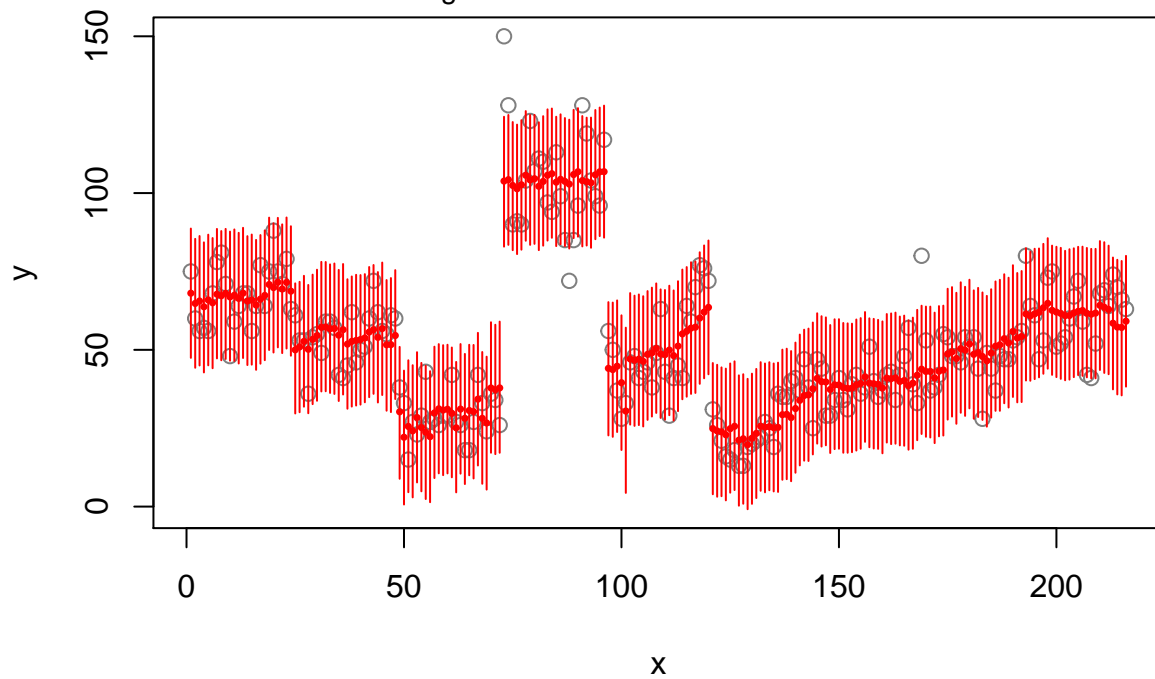
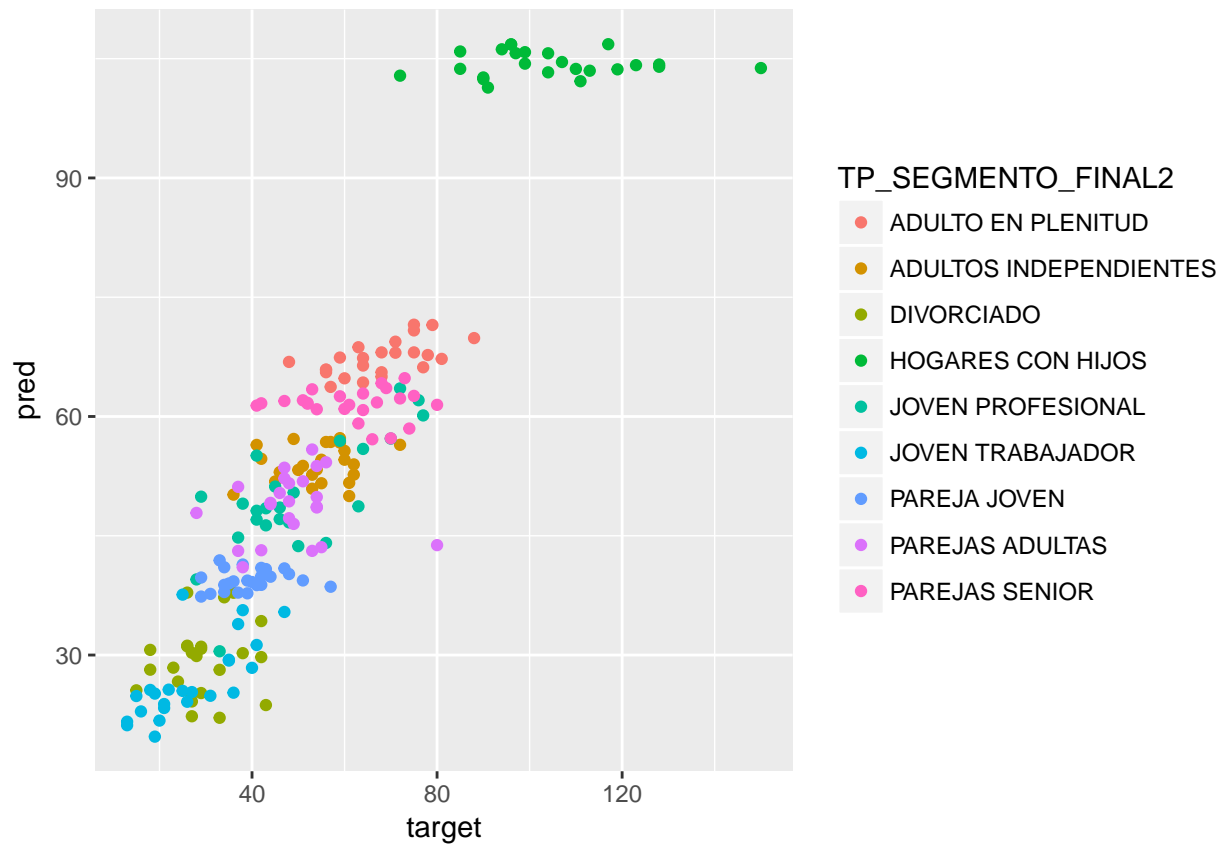
```
geom_line(aes(y = scale(gastos), colour = "gastos"))+
geom_line(aes(y = scale(cargos), colour = "cargos"))+
geom_line(aes(y = scale(saldo), colour = "saldo"))+
geom_point(aes(y = scale(total), colour = "contrataciones"))+
geom_line(aes(y = scale(total), colour = "contrataciones"))+
ylab('Valores')
```

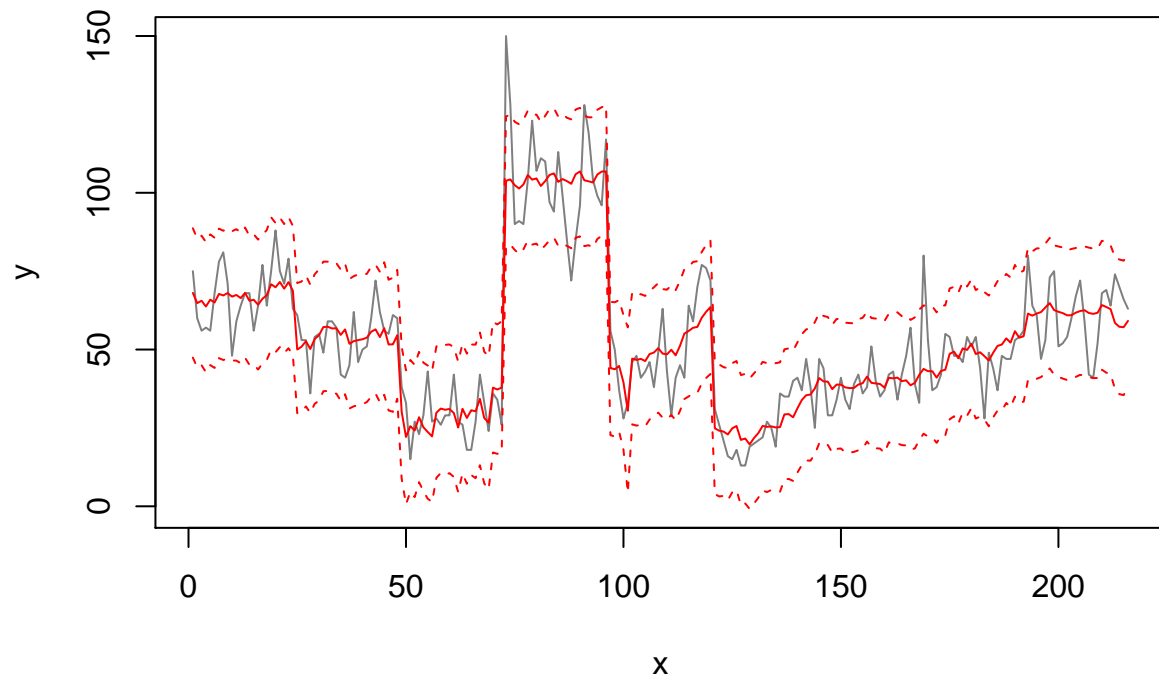


En la gráfica anterior se presentan las variables escaladas con el fin de poder observar mejor el efecto y las contrataciones que se ven en la línea dorada mostrando los puntos a lo largo del tiempo. Podemos observar que para estos clientes la tendencia de las variables es creciente, excepto la de necesidad financiera lo cual tiene sentido por la forma en que se construyó, sabemos que las variables con mayor correlación con las contrataciones son el importe promedio de gastos de los últimos 3 meses(gastos) y el número de productos bancarios promedio que poseen los clientes, en ninguno de los casos sus series se ven como las contrataciones.

## Modelos Estáticos

### Modelo lineal normal





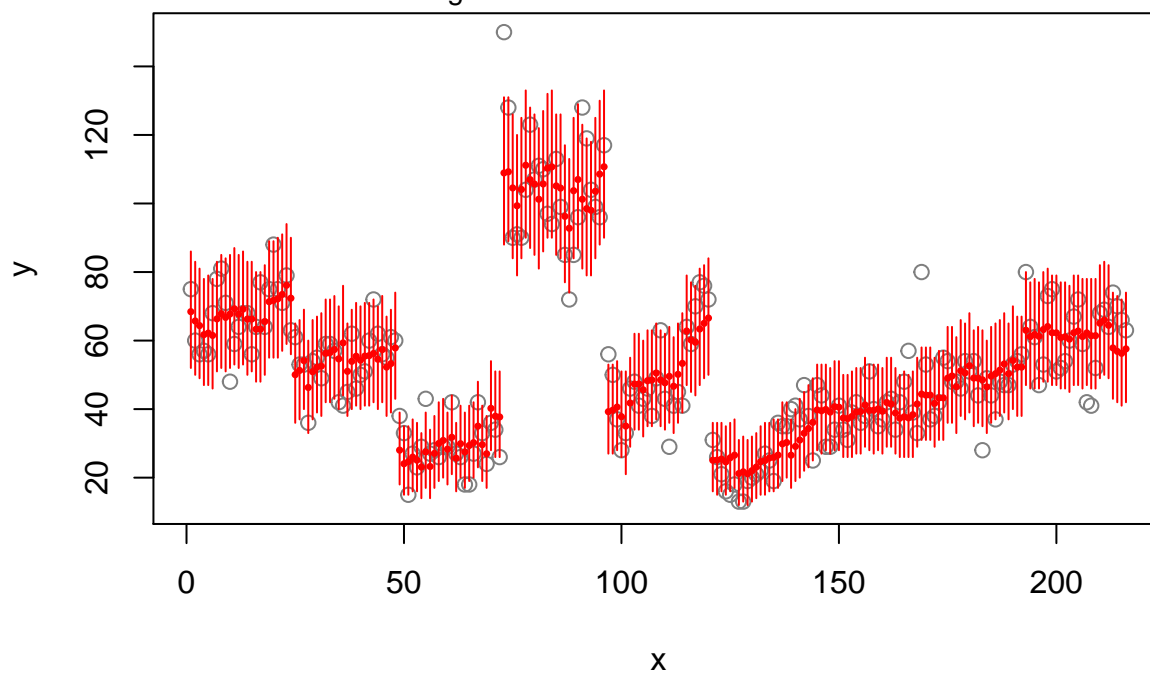
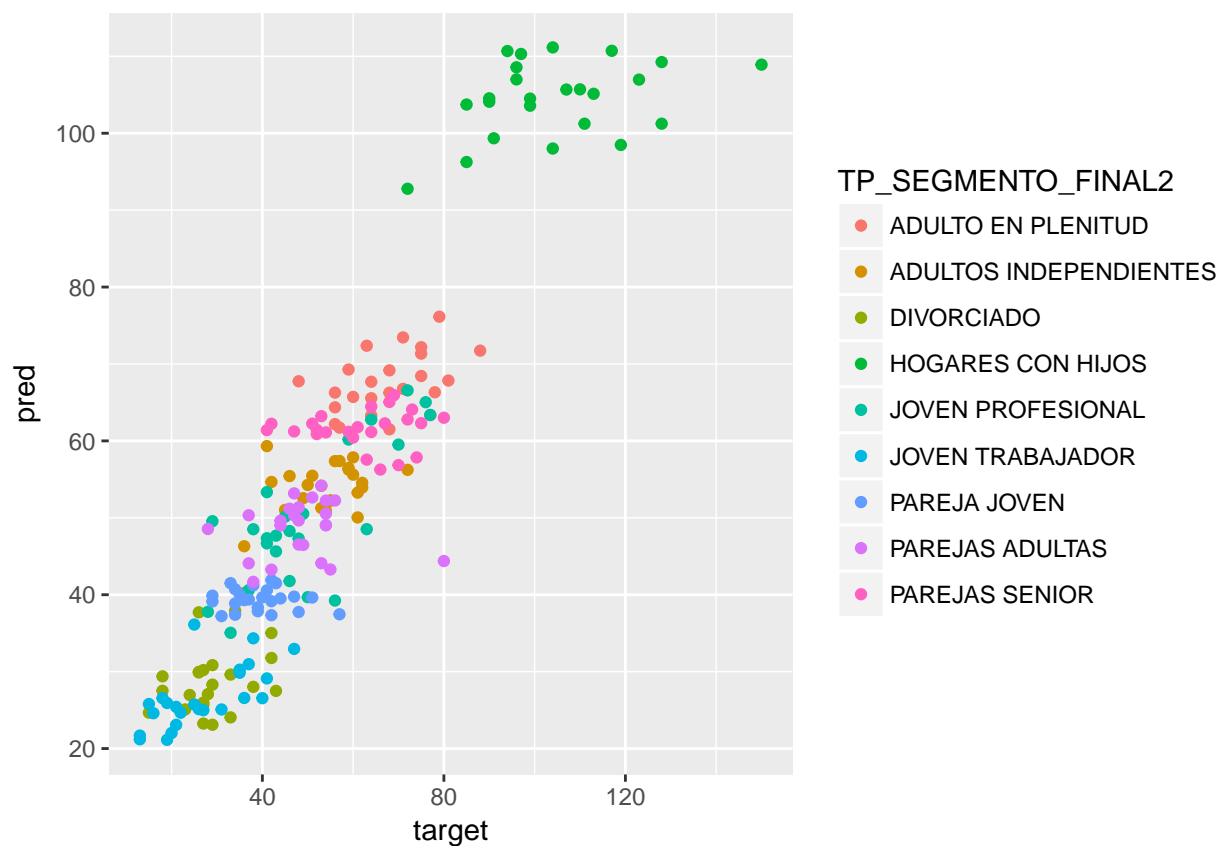
DIC:

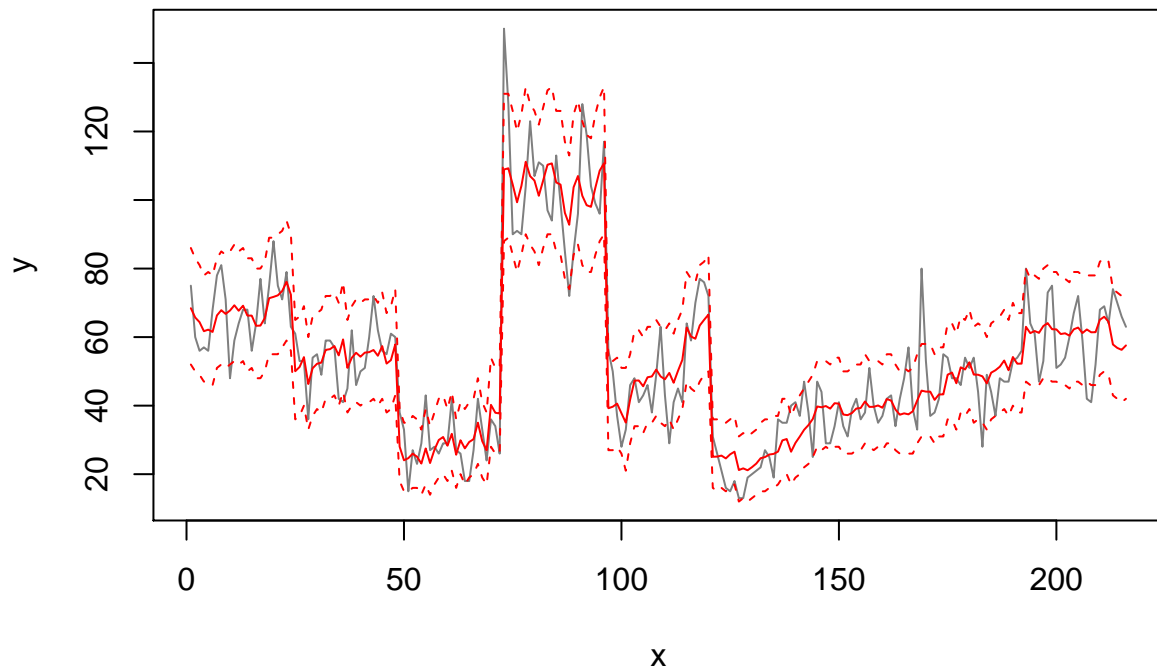
## [1] 1634.666

Pseudo- $R^2$ :

## [1] 0.8372262

# Modelo lineal generalizado poisson con liga log





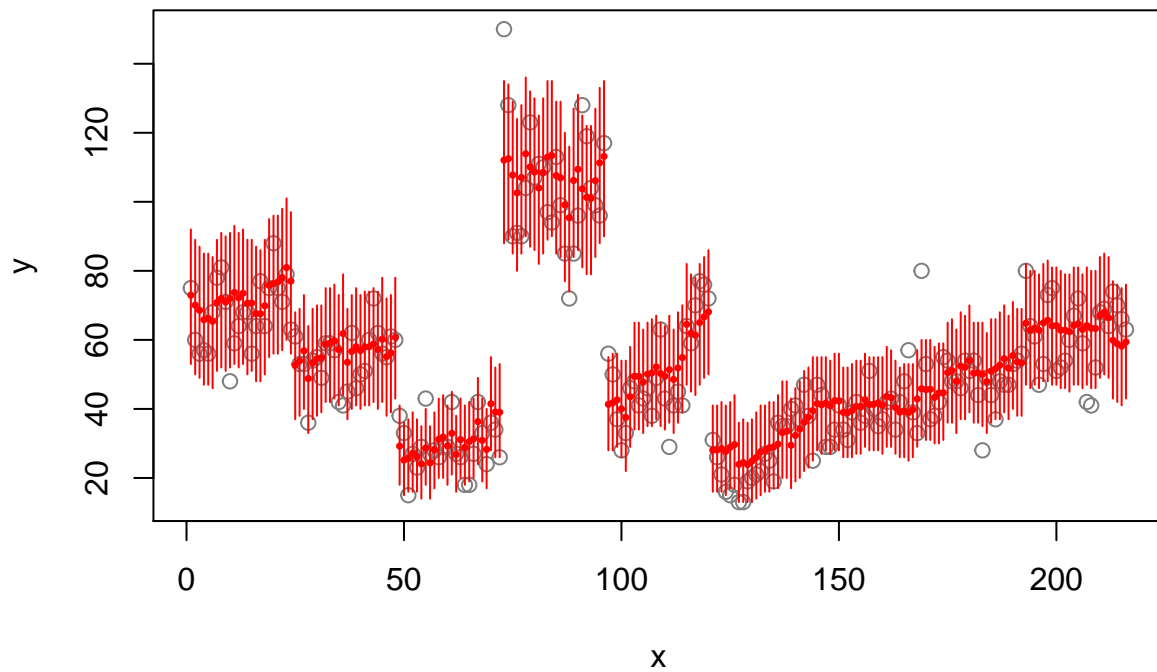
DIC:

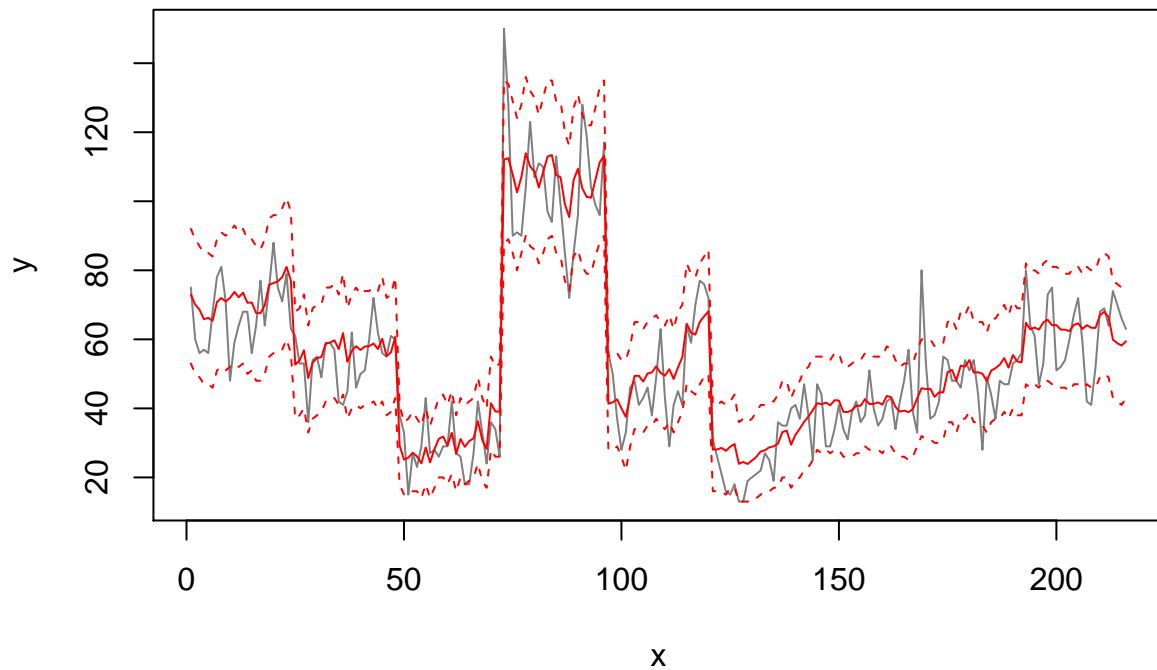
## [1] 1626.342

Pseudo- $R^2$ :

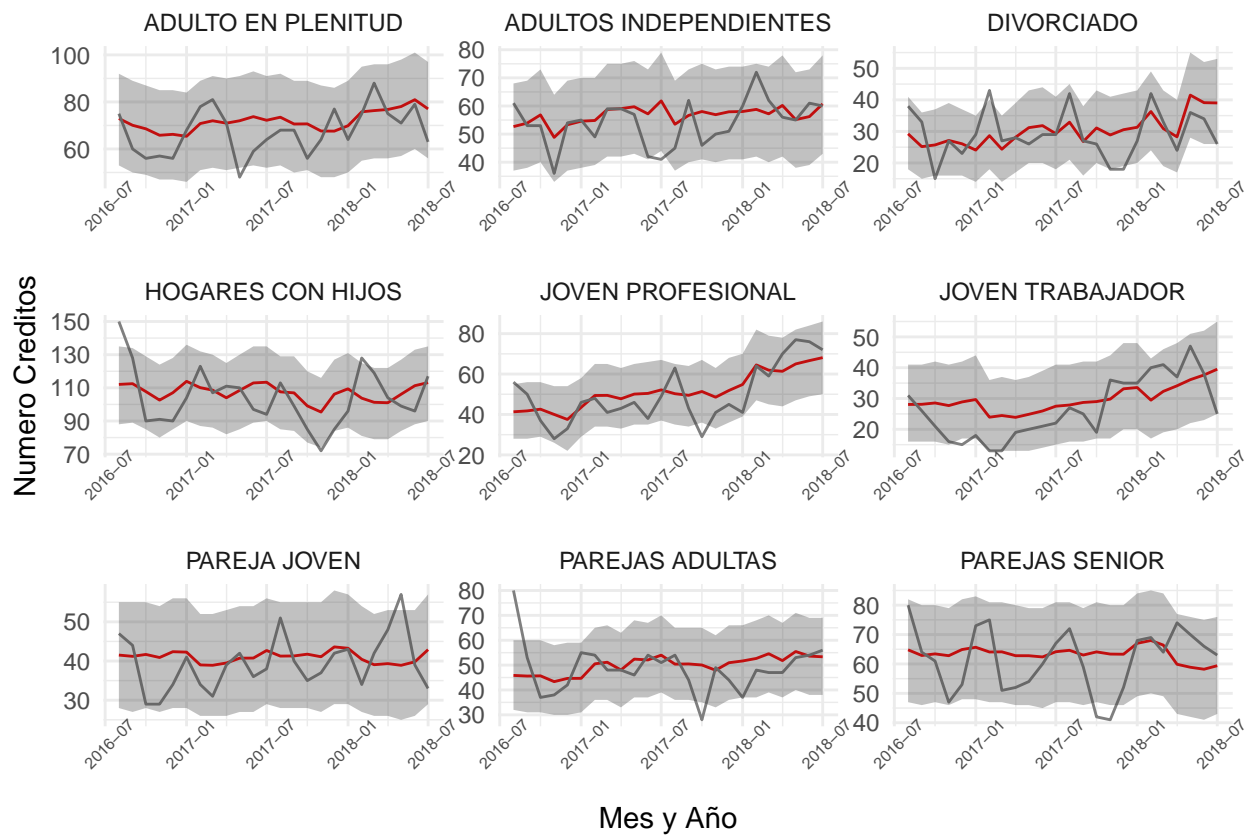
## [1] 0.8474719

**Modelo lineal generalizado binomial con liga logística**





```
ggplot(base_graf) +
  geom_line(aes(x=as.Date(FH_REF),y = mean),color = 'red') +
  geom_line(aes(x=as.Date(FH_REF),y = target), color = 'gray50') +geom_ribbon(aes(x=as.Date(FH_REF),ymin=
  ylab("Numero Credits") +
  xlab("Mes y Año")+theme_minimal()+theme(axis.text.x = element_text(size=6,angle=45))
```



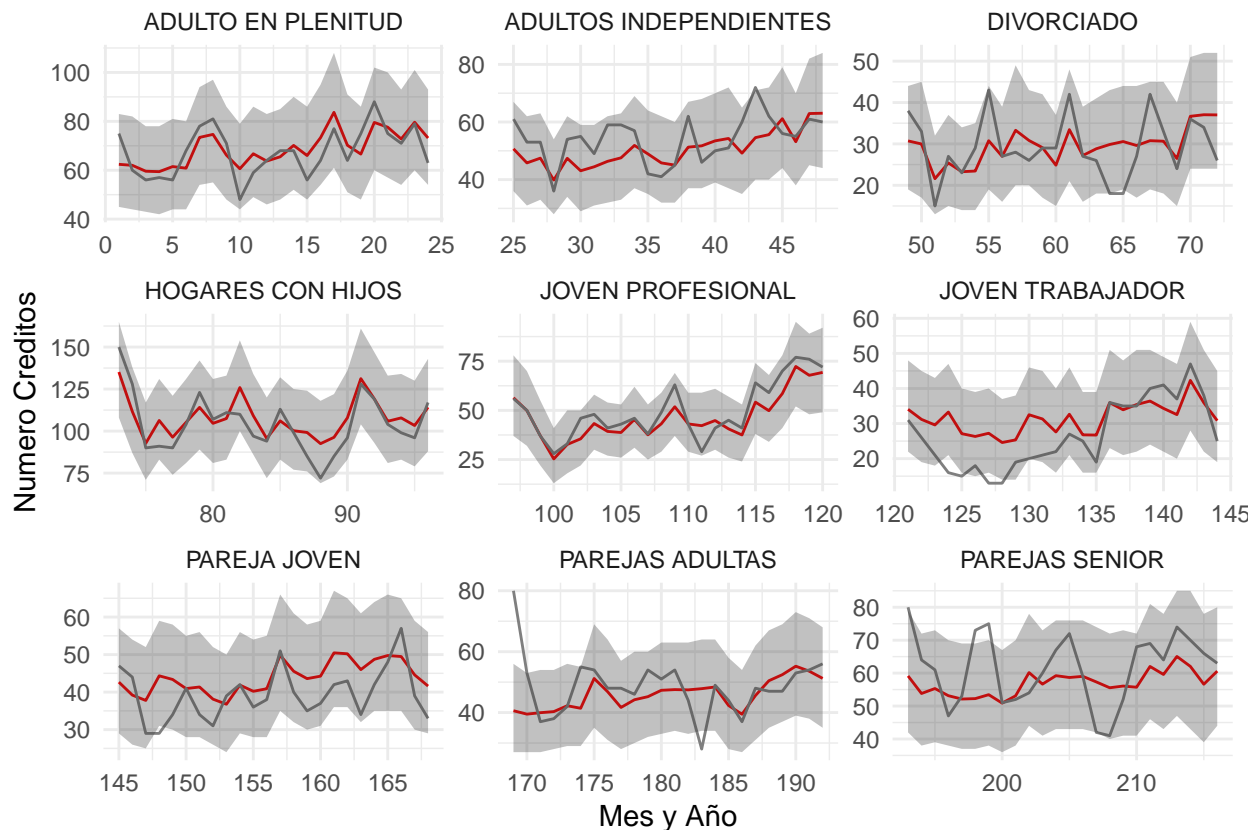
DIC:

```
## [1] 130331954
```

Pseudo- $R^2$ :

```
## [1] 0.8467358
```

```
out<-poisson_dinamico2$BUGSoutput$sims.list
out.sum<-poisson_dinamico2$BUGSoutput$summary
out.yf<-out.sum[grep("yf1",rownames(out.sum)),]
ymin<-min(y,out.yf[,c(1,3,7)])
ymax<-max(y,out.yf[,c(1,3,7)])
base_graf<-as.data.frame(cbind(camp,out.yf[,c(1,3,7)])) %>% mutate(fecha=format(as.Date(FH_REF),"%Y-%m"),
ggplot(base_graf) +
  geom_line(aes(x=x,y = mean),color = 'red') +
  geom_line(aes(x=x,y = target), color = 'gray50') +geom_ribbon(aes(x=x,ymin = `2.5%`, ymax = `97.5%`),
  ylab("Numero Creditos") +
  xlab("Mes y Año")+theme_minimal()
```



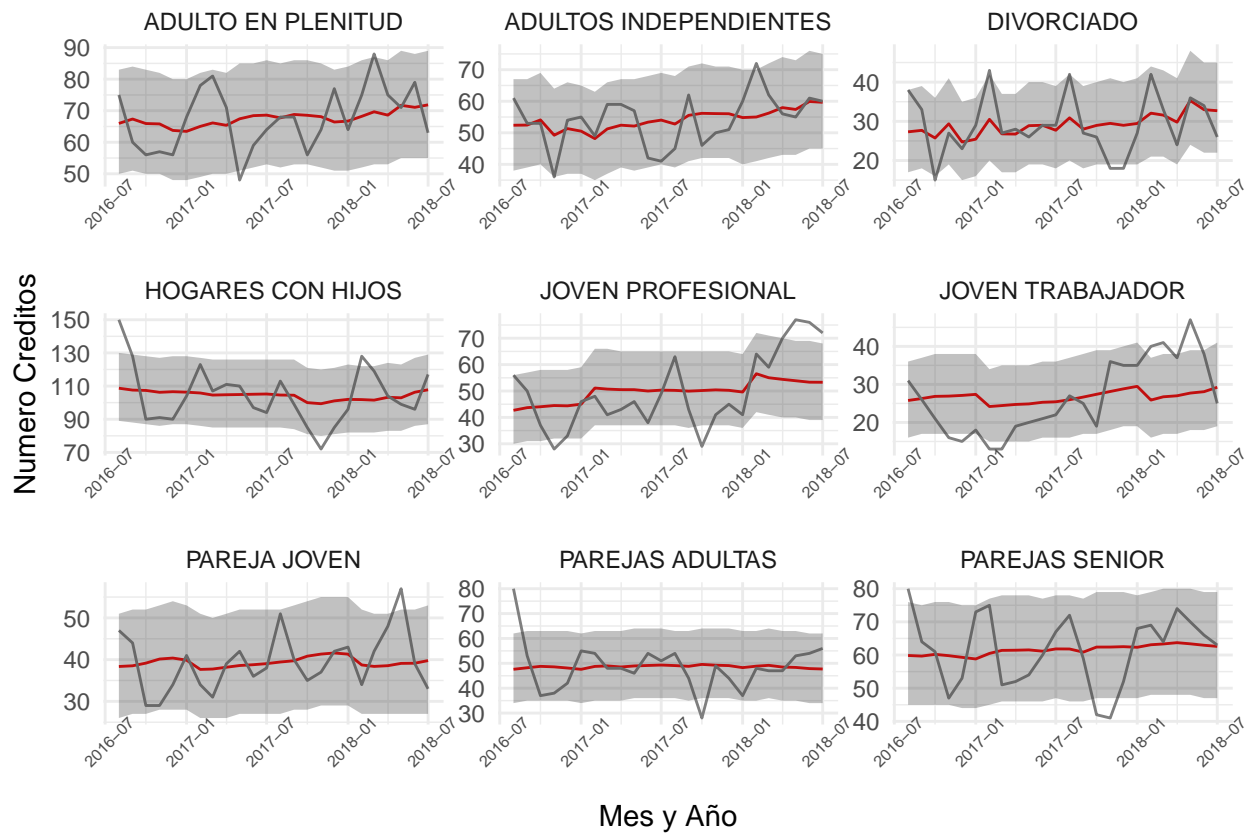
```
r2_dinamico2<-cor(y,out.yf[,1])^2
r2_dinamico2
```

```
## [1] 0.8853767
```

```
out<-poisson_cat$BUGSoutput$sims.list
out.sum<-poisson_cat$BUGSoutput$summary
out.yf<-out.sum[grep("yf1",rownames(out.sum)),]
ymin<-min(y,out.yf[,c(1,3,7)])
ymax<-max(y,out.yf[,c(1,3,7)])
```



```
base_graf<-as.data.frame(cbind(camp,out.yf[,c(1,3,7)])) %>% mutate(fecha=format(as.Date(FH_REF), "%Y-%m"))
ggplot(base_graf) +
  geom_line(aes(x=as.Date(FH_REF),y = mean),color = 'red') +
  geom_line(aes(x=as.Date(FH_REF),y = target), color = 'gray50') +geom_ribbon(aes(x=as.Date(FH_REF),ymin
  ylab("Numero Creditos") +
  xlab("Mes y Año")+theme_minimal()+theme(axis.text.x = element_text(size=6,angle=45))
```



```
r2_cat<-cor(y,out.yf[,1])^2
r2_cat
```

```
## [1] 0.8303296
```