

construccion_indice_2

```
library(foreign)
library(questionr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(ggthemes)
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##   src, summarize

## The following objects are masked from 'package:questionr':
##
##   describe, wtd.mean, wtd.table, wtd.var

## The following objects are masked from 'package:base':
##
##   format.pval, units
```

lectura bases

```
eph_1_05<-read.dbf("bases_EPH/Ind_t105.DBF")
eph_1_15<-read.dbf("bases_EPH/Ind_t115.DBF")
eph_1_20<-read.table("bases_EPH/usu_individual_T120.txt",
                     header = TRUE, sep= ";")
```

Se elige con que base trabajar

```
base<-eph_1_20
```

Se rotulan las variables y definen niveles

```
base$REGION_rot<-as.factor(base$REGION)
levels(base$REGION_rot)<-c("Gran Buenos Aires", "NOA",
                          "NEA", "Cuyo", "Pampeana", "Patagónica")

base$sexo<-as.factor(base$CH04)
levels(base$sexo)<-c("varones", "mujeres")

base$CH06[base$CH06==1]<-0
base$grupos_edades<-cut(base$CH06, c(0,24,54,max(base$CH06)))

base$NIVEL_ED[base$NIVEL_ED==9]<-NA
base$NIVEL_ED[base$NIVEL_ED==7]<-0
base$educacion<-as.factor(base$NIVEL_ED)
levels(base$educacion)<-c("nunca asistió", "primario incompleto", "primario completo",
                        "secundario incompleto", "secundario completo", "superior incompleto",
                        "superior completo")
```

Se retienen solo personas ocupadas

```
ocupades<-subset(base, base$ESTADO==1)
table(ocupades$CAT_OCUP)
```

```
##
##      1      2      3      4
##  767  4649 15562  161
```

Seguridad = estabilidad + obra social

Asalariados

Estabilidad

Combina “a término” vs “permanente” (PP07C) con duración de los “a término” (PP07D).
Se hace para el subconjunto de asalariados

```
asalariados<-subset(ocupades, ocupades$CAT_OCUP==3)
table(asalariados$PP07C)
```

```
##
##      0      1      2      9
## 1616 2011 11182   753
```

```
asalariados$estabilidad<-ifelse(asalariados$PP07C==2, 6, ifelse(
  asalariados$PP07C==1 & asalariados$PP07D==5, 5, ifelse(
    asalariados$PP07C==1 & asalariados$PP07D==4, 4, ifelse(
      asalariados$PP07C==1 & asalariados$PP07D== 3, 3, ifelse(
        asalariados$PP07C==1 & asalariados$PP07D== 2, 2, ifelse(
          asalariados$PP07C==1 & asalariados$PP07D== 1, 1, NA
        )
      )
    )
  )
))
table(asalariados$estabilidad)
```

```
##
##      1      2      3      4      5      6
##   260   180    93   123   257 11182
```

Obra social

Del cuestionario hogar CH08

```
table(asalariados$CH08)
```

```
##
##      1      2      3      4      9      12      13      23
## 11200   445   146  3425      8   329      6      3
```

```
asalariados$obra_social <-1
asalariados$obra_social[asalariados$CH08==4]<-0
asalariados$obra_social[asalariados$CH08==9]<-NA
table(asalariados$CH08, asalariados$obra_social)
```

```
##
##      0      1
##      1      0 11200
##      2      0   445
##      3      0   146
##      4  3425      0
```

```
##    9      0      0
##   12      0    329
##   13      0      6
##   23      0      3
```

Combinación

Para que tengan igual peso se estandarizan al intervalo [0 - 0.50] por medio de $0.50 * \frac{x-min}{max-min}$ cada una

```
asalariades$estabilidad_st<-.5*(asalariades$estabilidad-1)/5
summary(asalariades$estabilidad_st)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  0.000   0.500   0.500   0.477   0.500   0.500  3467
```

```
asalariades$obra_social_st<-.5*asalariades$obra_social
summary(asalariades$obra_social_st)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000   0.5000   0.5000   0.3899   0.5000   0.5000      8
```

Seguridad

Combiación aditiva de las dos

```
asalariades$seguridad<-asalariades$estabilidad_st+
  asalariades$obra_social_st
summary(asalariades$seguridad)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  0.000   1.000   1.000   0.914   1.000   1.000  3471
```

Cuenta propia

Estabilidad

Combina capital (maquinaria PP05C_1, local PP05C_2, vehículo PP05C_3) y clientes (uno solo o varios PP05F)

Se hace para el subconjunto de cuenta propia

```
cuentapropia<-subset(ocupades,
  ocupades$CAT_OCUP==2 | ocupades$CAT_OCUP ==1)
table(cuentapropia$PP05C_1)
```

```
##
##    1    2    3
## 4120 219 1077
```

```
table(cuentapropia$PP05C_2)
```

```
##
##      1      2      3
## 1413 1059 2944
```

```
table(cuentapropia$PP05C_3)
```

```
##
##      1      2      3
## 1822  136 3458
```

```
table(cuentapropia$PP05F)
```

```
##
##      6      7
##    32 5384
```

```
cuentapropia$maquinas<-0
cuentapropia$maquinas[cuentapropia$PP05C_1==0]<-NA
cuentapropia$maquinas[cuentapropia$PP05C_1==9]<-NA
cuentapropia$maquinas[cuentapropia$PP05C_1==1]<-2
cuentapropia$maquinas[cuentapropia$PP05C_1==2]<-1
table(cuentapropia$maquinas, cuentapropia$PP05C_1)
```

```
##
##      1      2      3
## 0      0      0 1077
## 1      0 219      0
## 2 4120      0      0
```

```
cuentapropia$local<-0
cuentapropia$local[cuentapropia$PP05C_2==0]<-NA
cuentapropia$local[cuentapropia$PP05C_2==9]<-NA
cuentapropia$local[cuentapropia$PP05C_2==1]<-2
cuentapropia$local[cuentapropia$PP05C_2==2]<-1
table(cuentapropia$local, cuentapropia$PP05C_2)
```

```
##
##      1      2      3
## 0      0      0 2944
## 1      0 1059      0
## 2 1413      0      0
```

```
cuentapropia$vehiculo<-0
cuentapropia$vehiculo[cuentapropia$PP05C_3==0]<-NA
cuentapropia$vehiculo[cuentapropia$PP05C_3==9]<-NA
cuentapropia$vehiculo[cuentapropia$PP05C_3==1]<-2
cuentapropia$vehiculo[cuentapropia$PP05C_3==2]<-1
table(cuentapropia$vehiculo, cuentapropia$PP05C_3)
```

```
##
##      1      2      3
##    0      0      0 3458
##    1      0    136      0
##    2 1822      0      0
```

```
cuentapropia$clientes<-0
cuentapropia$clientes[cuentapropia$PP05F==0]<-NA
cuentapropia$clientes[cuentapropia$PP05F==9]<-NA
cuentapropia$clientes[cuentapropia$PP05F==6]<-0
cuentapropia$clientes[cuentapropia$PP05F==7]<-3

table(cuentapropia$clientes, cuentapropia$PP05F)
```

```
##
##      6      7
##    0    32      0
##    3      0 5384
```

```
cuentapropia$estabilidad<-cuentapropia$maquinas+cuentapropia$local+
  cuentapropia$vehiculo+cuentapropia$clientes

table(cuentapropia$estabilidad)
```

```
##
##    0      1      2      3      4      5      6      7      8      9
##    9      3      9    569    189    1715    587    1486    348    501
```

Obra social (del cuestionario hogar)

```
table(cuentapropia$CH08)
```

```
##
##      1      2      3      4      9      12      13      123
## 2343    310    91 2550      6    112      3      1
```

```
cuentapropia$obra_social<-1
cuentapropia$obra_social[cuentapropia$CH08==4]<-0
cuentapropia$obra_social[cuentapropia$CH08==9]<-NA

table(cuentapropia$CH08, cuentapropia$obra_social)
```

```
##
##      0      1
##    1      0 2343
##    2      0    310
##    3      0     91
##    4 2550      0
##    9      0      0
##   12      0    112
##   13      0      3
##  123      0      1
```

Combinación

Para que tengan igual peso se estandarizan al intervalo [0 - 0.50] por medio de $0.50 * \frac{x-min}{max-min}$ cada una y luego se suman

```
cuentapropia$estabilidad_st<-.5*(cuentapropia$estabilidad)/9
summary(cuentapropia$estabilidad_st)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.2778  0.3333  0.3311  0.3889  0.5000
```

```
cuentapropia$sobra_social_st<-.5*cuentapropia$sobra_social
summary(cuentapropia$sobra_social_st)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000  0.0000  0.5000  0.2643  0.5000  0.5000      6
```

Seguridad

Combiación aditiva de las dos

```
cuentapropia$seguridad<-cuentapropia$estabilidad_st+
  cuentapropia$sobra_social_st
summary(cuentapropia$seguridad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000  0.2778  0.6667  0.5954  0.8889  1.0000      6
```

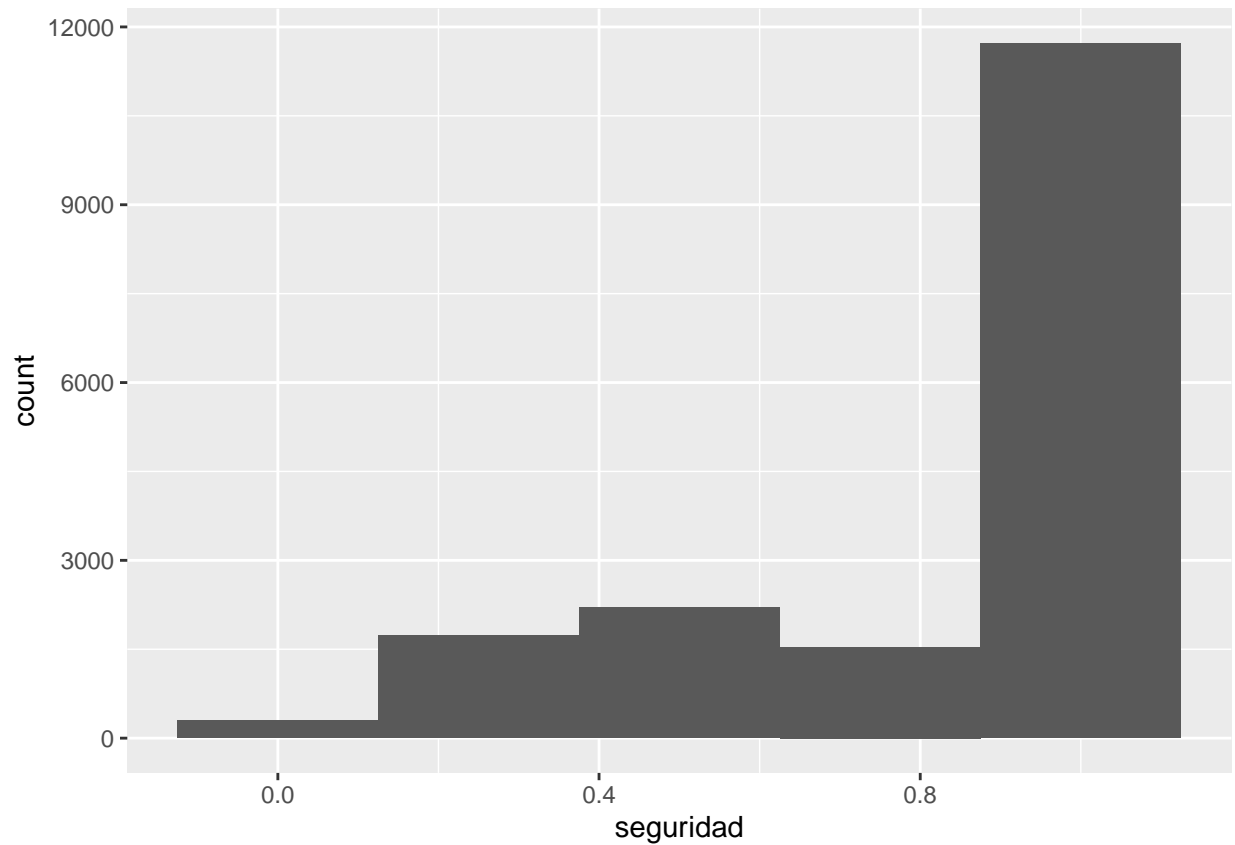
Se eliminan las variables que no están en ambas bases y se unen nuevamente

```
cuentapropia<-
  subset(cuentapropia, select=-c(local, clientes, vehiculo, maquinas))

ocupades<-data.frame(rbind(asalariades, cuentapropia))
summary(ocupades$seguridad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.000  0.667  1.000  0.815  1.000  1.000  3477
```

```
ggplot(ocupades)+geom_histogram(aes(seguridad), bins = 5)
```



```
summary(ocupades$seguridad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  0.000   0.667   1.000   0.815   1.000   1.000  3477
```

Consistencia

Es la relación entre la calificación de la tarea y la educación del trabajador

```
class(ocupades$PP04D_COD)
```

```
## [1] "integer"
```

```
ocupades$calif.ocup= substr(ocupades$PP04D_COD, 5,5)
table(ocupades$calif.ocup)
```

```
##
##      1      2      3      4      6      7      8      9
## 1013 1311 3245 10668 4558      1     87      4     91
```



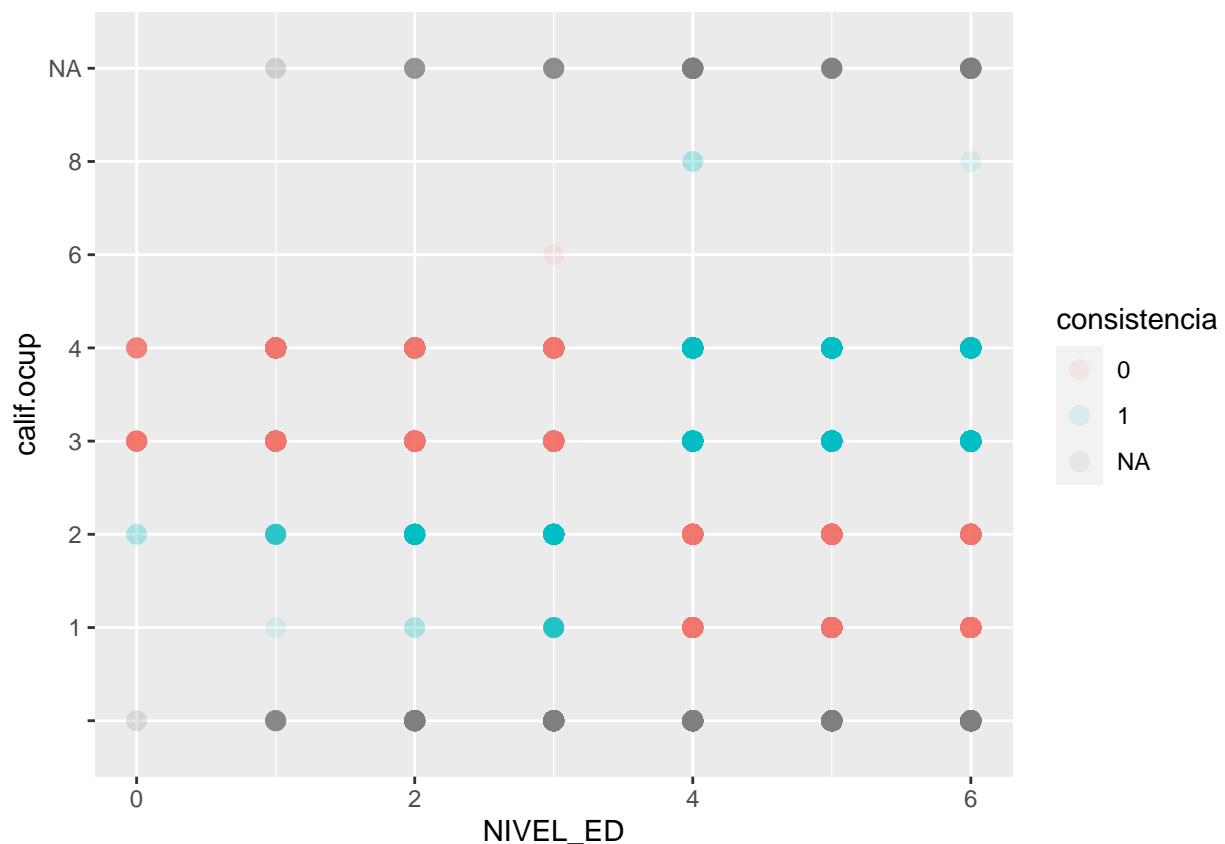
```
ocupades$calif.ocup[ocupades$calif.ocup==7]<-NA
ocupades$calif.ocup[ocupades$calif.ocup==9]<-NA
```

Se considera consistente (1) si educación y calificación están ambas por encima o ambas por debajo de sus medianas e inconsistente (0) en caso contrario

```
ocupades$educa_num<-as.numeric(ocupades$educacion)
ocupades$calif_num<-as.numeric(ocupades$calif.ocup)

ocupades$consistencia<-as.factor(
  ifelse(
    (ocupades$educa_num >= median(ocupades$educa_num, na.rm = TRUE)) &
    (ocupades$calif_num >= median(ocupades$calif_num, na.rm = TRUE)) |
    (ocupades$educa_num < median(ocupades$educa_num, na.rm = TRUE)) &
    (ocupades$calif_num < median(ocupades$calif_num, na.rm = TRUE)),1,0))

ggplot(ocupades)+
  geom_point(aes(NIVEL_ED, calif.ocup, col=consistencia), alpha=.1, size=3)
```



Ingresos

Se construye el ingreso laboral, combinando ingresos salariales con ingresos de cuentapropistas con o sin socios (en EPH son tres variables)

```

ocupades$ingreso_laboral<-ifelse(
  (ocupades$CAT_OCUP==1 | ocupades$CAT_OCUP ==2) &
  ocupades$PP06A==1,ocupades$PP06D, ifelse(
    (ocupades$CAT_OCUP==1 | ocupades$CAT_OCUP ==2) &
    ocupades$PP06A==2,ocupades$PP06C, ifelse(
      ocupades$CAT_OCUP==3, ocupades$PP08D1, 0
    )
  )
)
summary(ocupades$ingreso_laboral)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      -9    6000   19000   22556   30000   500000

```

Se retienen los casos con ingreso no nulo y horas semanales trabajadas no nulas y menores a 999

```

ocupades<-subset(ocupades, ocupades$ingreso_laboral>0 &
  ocupades$PP3E_TOT>0 & ocupades$PP3E_TOT < 999
)

```

Ingreso - hora

```

ocupades$ingreso_hora_bruto<-ocupades$ingreso_laboral/(4*ocupades$PP3E_TOT)
summary(ocupades$ingreso_hora)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.786  97.222  153.058  195.574  237.500  5000.000

```

Se retienen los que tienen ingreso hora menor al P_{99}

```

ocupades<-subset(ocupades, ocupades$ingreso_hora_bruto<
  quantile(ocupades$ingreso_hora_bruto,.99))

```

Se estandariza (puntaje z) respecto de las medias y desviaciones (ponderadas) de cada aglomerado

```

aglos<-as.numeric(as.character(levels(as.factor(ocupades$AGLOMERADO))))

for (i in aglos) {ocupades$ingreso_hora_z=(
  ocupades$ingreso_hora_bruto-wtd.mean(
    ocupades[ocupades$AGLOMERADO==i,]$ingreso_hora_bruto,
    weights = ocupades[ocupades$AGLOMERADO==i,]$PONDERA)/sqrt(wtd.var(
      ocupades[ocupades$AGLOMERADO==i,]$ingreso_hora_bruto,
      weights=ocupades[ocupades$AGLOMERADO==i,]$PONDERA))
}

summary(ocupades$ingreso_hora_bruto)

```

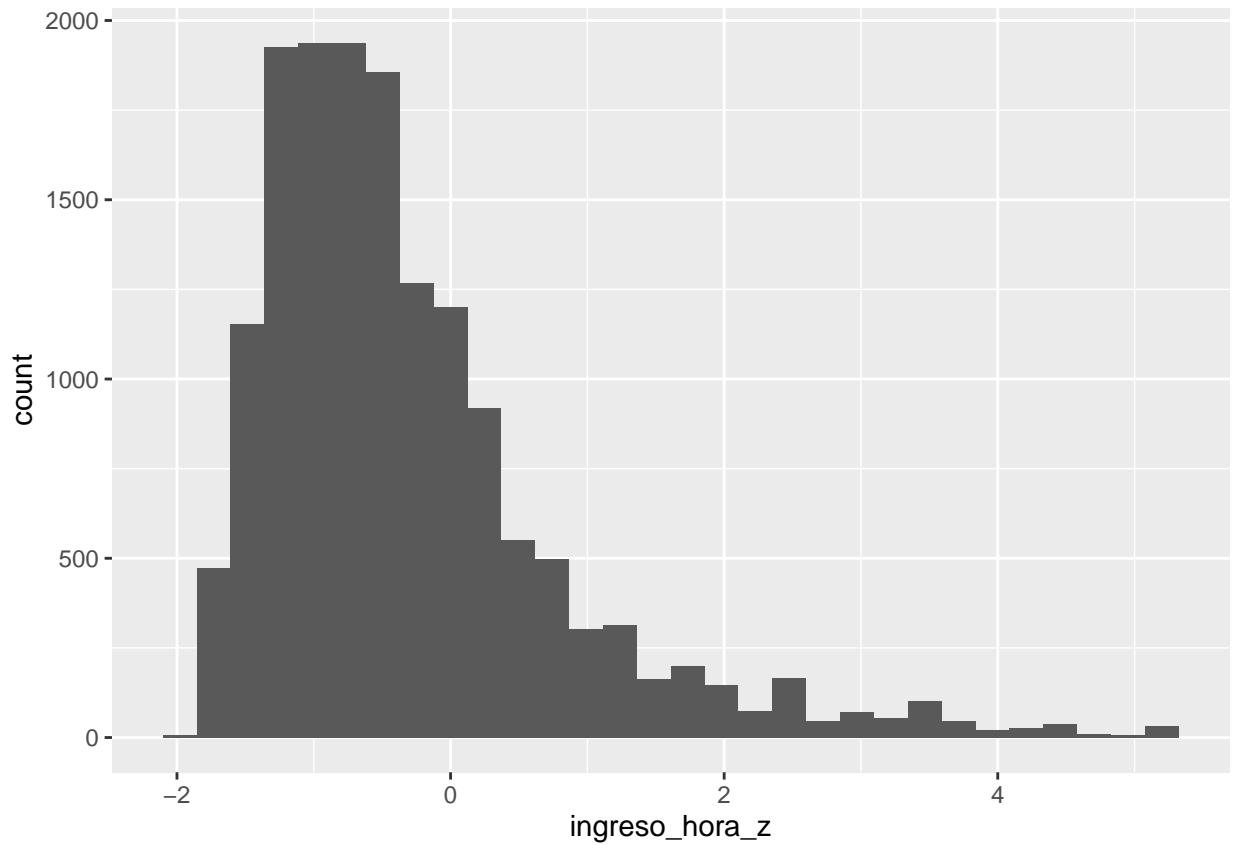
```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.786  95.238  150.000  183.331  233.333  839.286

```

```
ggplot(ocupades)+geom_histogram(aes(ingreso_hora_z))
```

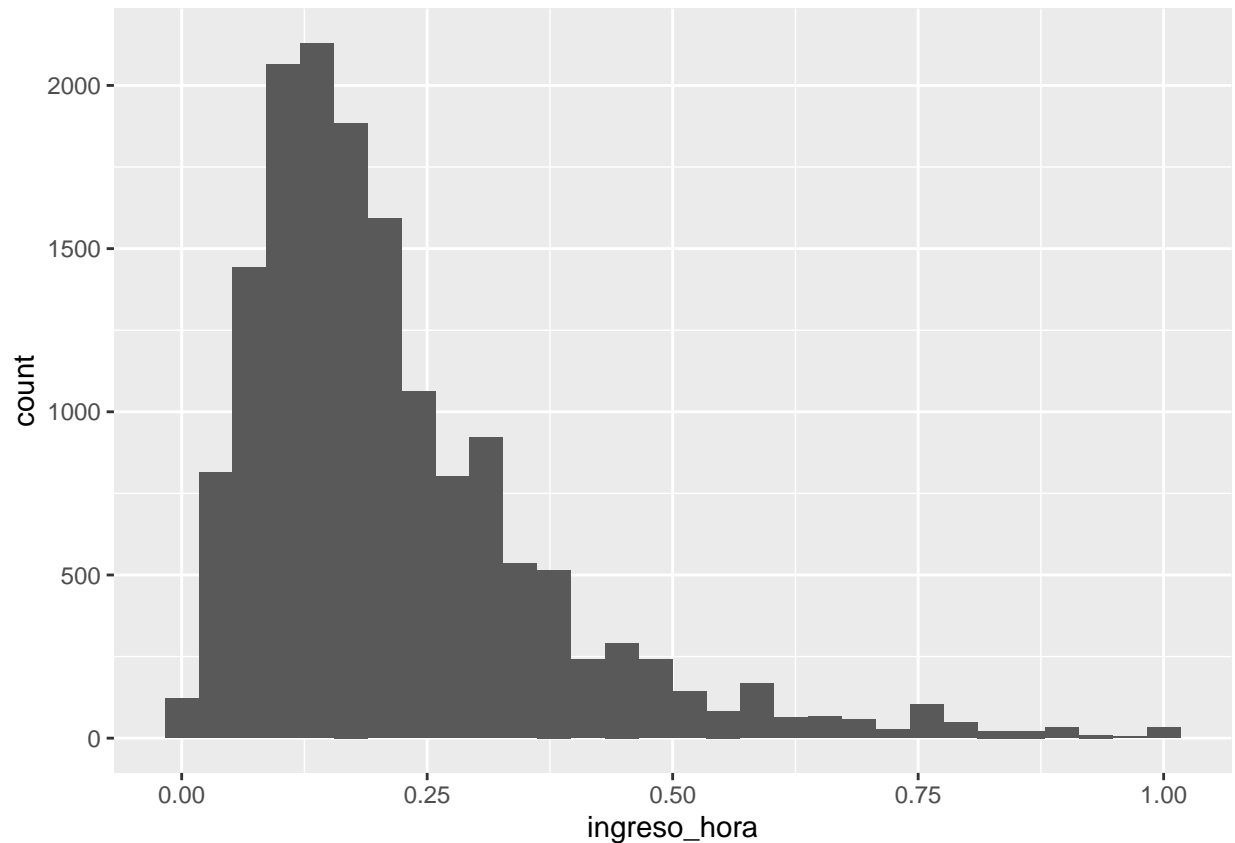
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Y se estandariza al [0 1]

```
ocupades$ingreso_hora<-(ocupades$ingreso_hora_z-min(ocupades$ingreso_hora_z))/  
  (max(ocupades$ingreso_hora_z)-min(ocupades$ingreso_hora_z))  
  
ggplot(ocupades)+geom_histogram(aes(ingreso_hora))
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Combinación de los indicadores

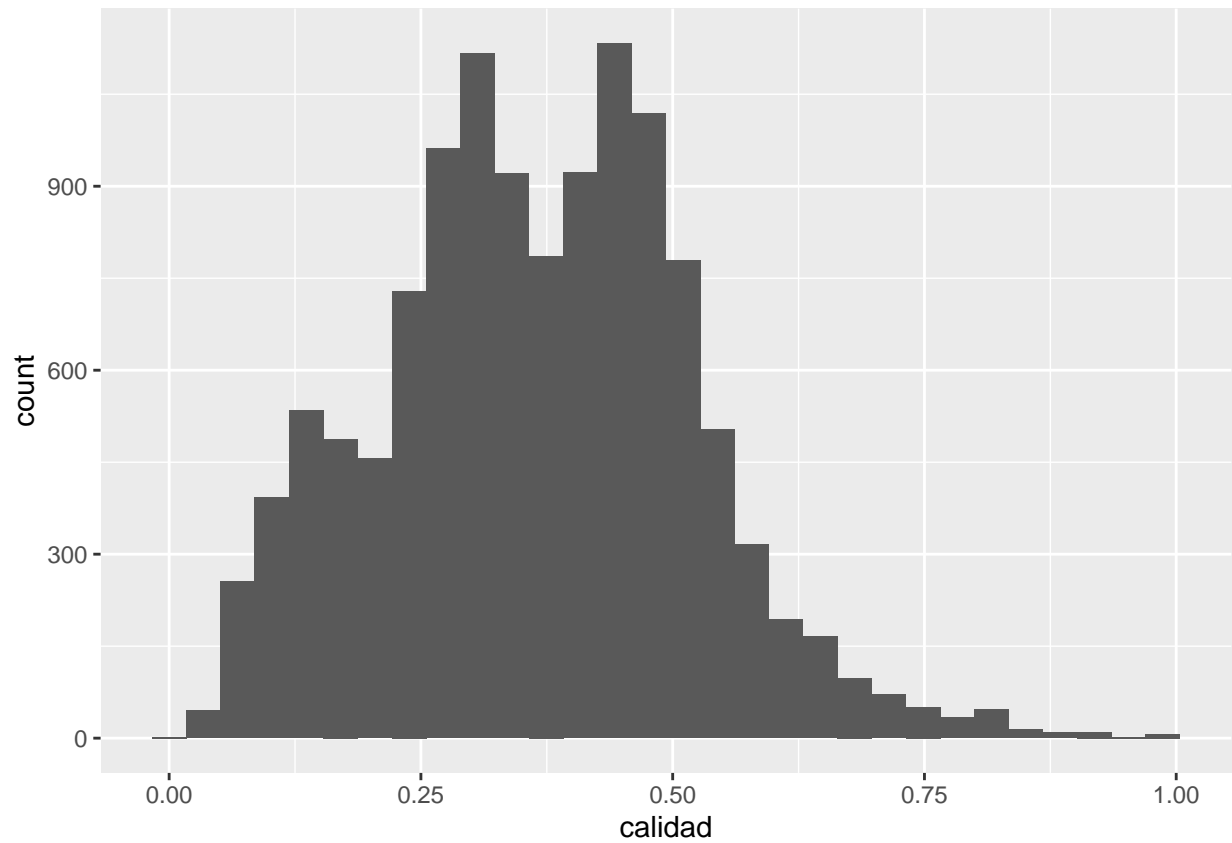
¿será por PCA? por ahora aditivo nomás

```
ocupades$consistencia<-as.numeric(as.character(ocupades$consistencia))
ocupades$calidad<-(ocupades$seguridad+ocupades$consistencia+4*ocupades$ingreso_hora)/6
summary(ocupades$calidad)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  0.009  0.260   0.364   0.364  0.468   0.995   3481
```

```
ggplot(ocupades)+geom_histogram(aes(calidad))
```

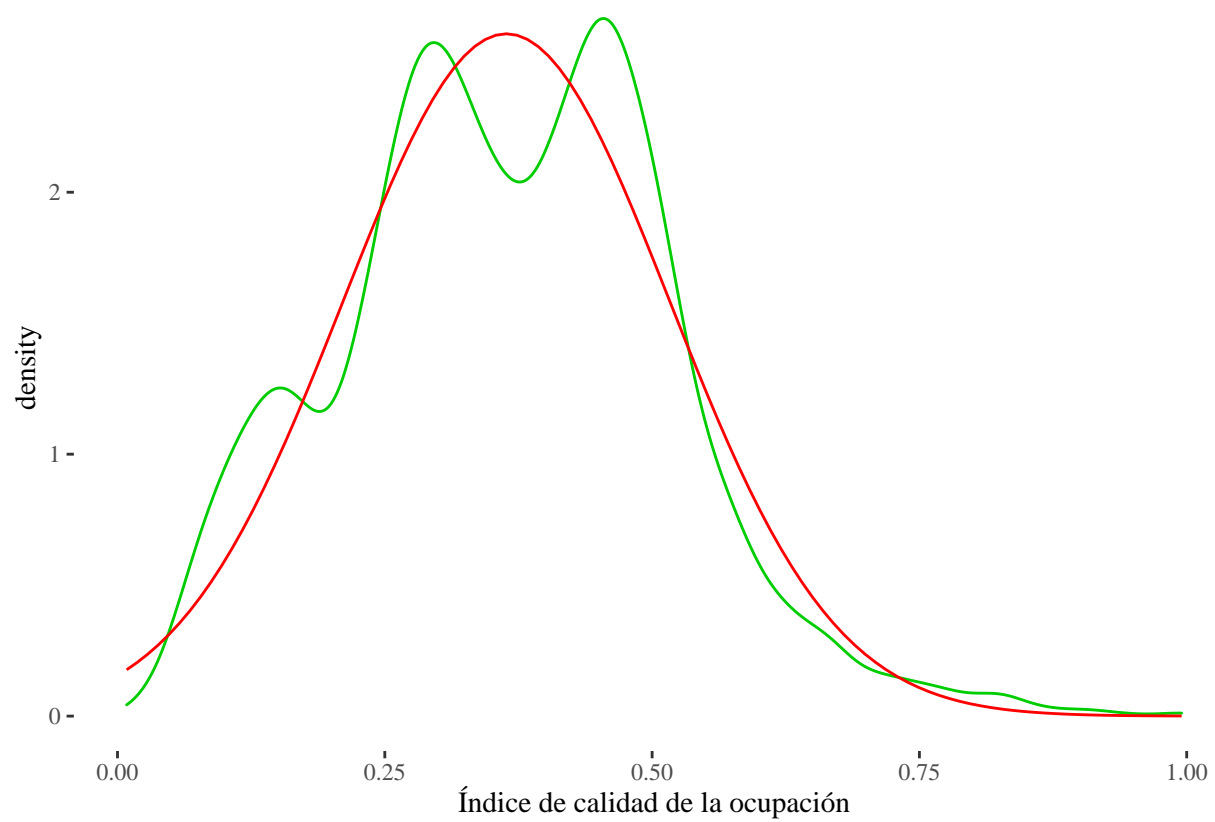
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



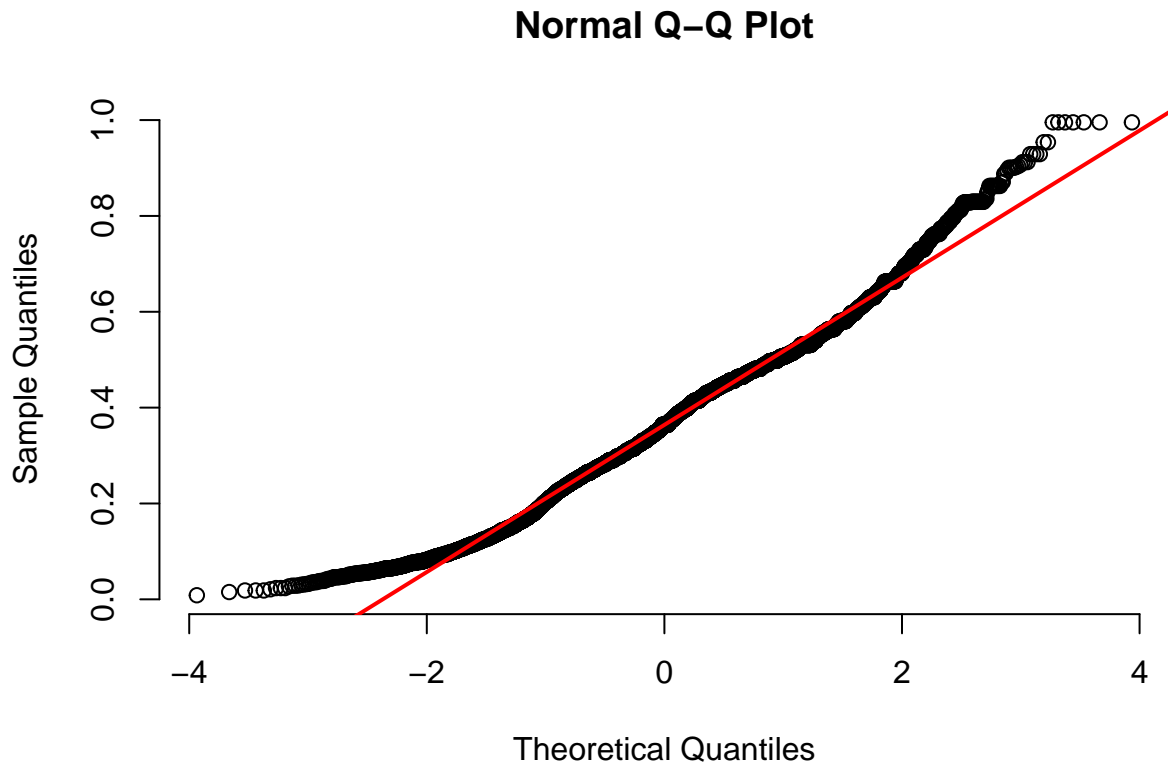
Normalidad

Gráficos

```
ggplot(ocupades)+geom_density(aes(calidad), col="green3")+  
  xlab("Índice de calidad de la ocupación")+stat_function(  
    fun = dnorm, args = list(  
      mean(ocupades$calidad, na.rm = TRUE),  
      sd(ocupades$calidad, na.rm = TRUE)), col="red")+ theme_tufte()
```



```
qqnorm(ocupades$calidad, pch = 1, frame = FALSE)
qqline(ocupades$calidad, col = "red", lwd = 2)
```



Prueba de bondad de ajuste

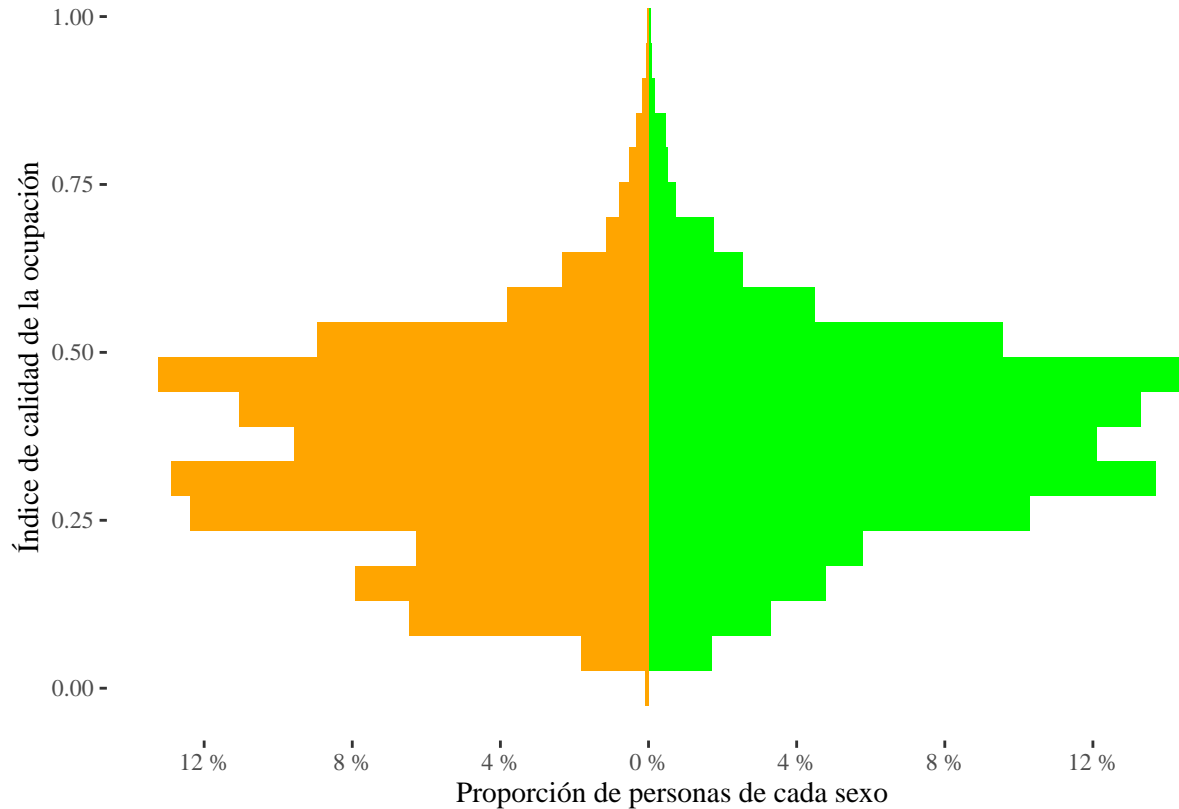
Comparación por sexos

```
t.test(ocupades$calidad~ocupades$sexo)
```

```
##
##  Welch Two Sample t-test
##
## data:  ocupades$calidad by ocupades$sexo
## t = -9.219, df = 10281, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03149661 -0.02045120
## sample estimates:
## mean in group varones mean in group mujeres
##           0.3537666           0.3797405
```

```
ggplot(ocupades) + geom_histogram(data=subset(ocupades, ocupades$sexo=="mujeres"),
                                aes(calidad, y=..count../sum(..count..)),fill="green", bins = 20)+
  geom_histogram(data=subset(ocupades, ocupades$sexo=="varones"),
                aes(calidad, y=(-1)*..count../sum(..count..)),fill="orange", bins = 20)+
  coord_flip()+ xlab("Índice de calidad de la ocupación")+
```

```
ylab("Proporción de personas de cada sexo")+
scale_y_continuous(breaks=seq(-.16,.16,.04),
                   labels=paste(100*abs(seq(-.16,.16,.04)),"%")) + theme_tufte()
```



Da mejor para mujeres, pero cuando se miran las componentes

```
t.test(ocupades$seguridad~ocupades$sexo)
```

```
##
## Welch Two Sample t-test
##
## data: ocupades$seguridad by ocupades$sexo
## t = -11.821, df = 11570, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.06921780 -0.04952682
## sample estimates:
## mean in group varones mean in group mujeres
## 0.7809560 0.8403283
```

```
t.test(ocupades$consistencia~ocupades$sexo)
```

```
##
## Welch Two Sample t-test
```



```
##
## data: ocupades$consistencia by ocupades$sexo
## t = -5.5063, df = 13754, p-value = 3.729e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.06176228 -0.02933399
## sample estimates:
## mean in group varones mean in group mujeres
## 0.4411237 0.4866719

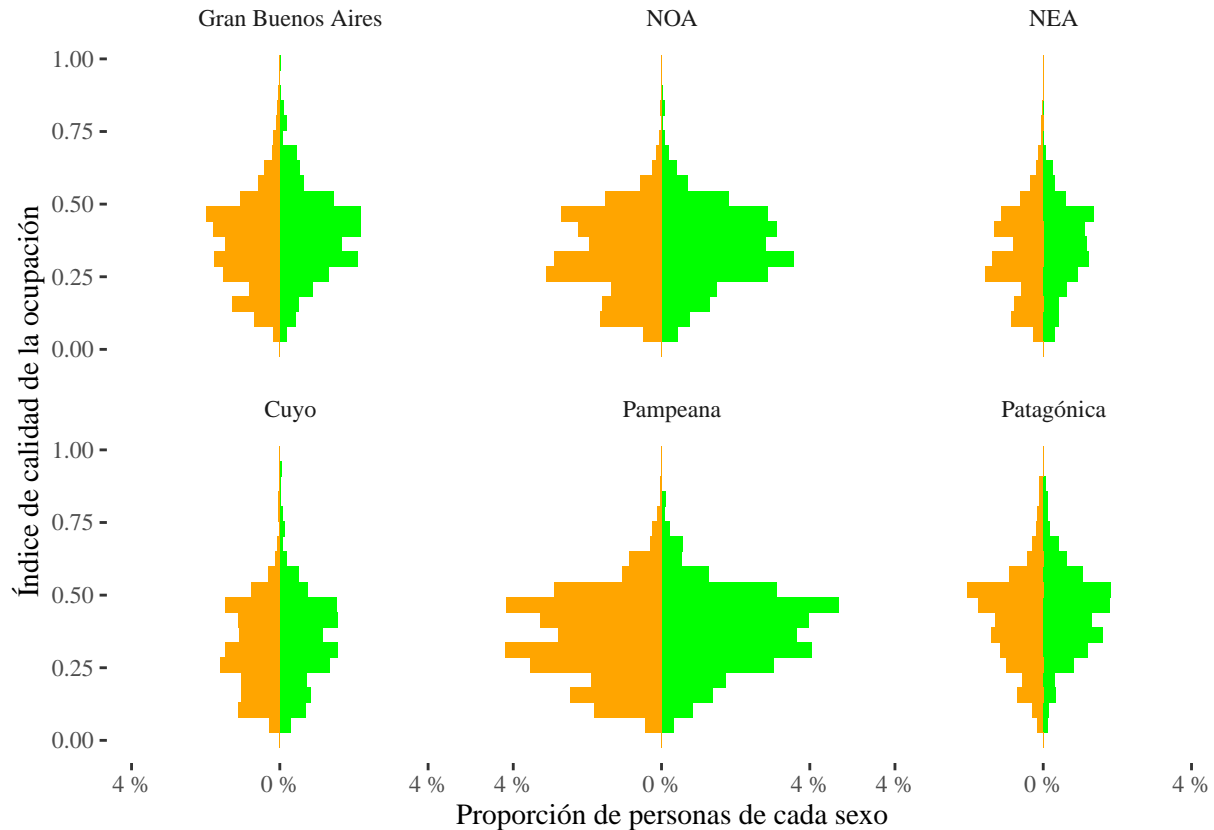
t.test(ocupades$ingreso_hora~ocupades$sexo)
```

```
##
## Welch Two Sample t-test
##
## data: ocupades$ingreso_hora by ocupades$sexo
## t = 1.5704, df = 14187, p-value = 0.1163
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.0009842578 0.0089166418
## sample estimates:
## mean in group varones mean in group mujeres
## 0.2184682 0.2145020
```

Las que aportan positivamente en las mujeres son seguridad y consistencia, mientras que en el ingreso la diferencia es a favor de los varones

Comparación por sexos y regiones

```
ggplot(ocupades) + geom_histogram(data=subset(ocupades, ocupades$sexo=="mujeres"),
  aes(calidad, y=..count../sum(..count..)),fill="green", bins = 20)+
  geom_histogram(data=subset(ocupades, ocupades$sexo=="varones"),
  aes(calidad, y=(-1)*..count../sum(..count..)),fill="orange", bins = 20)+
  coord_flip()+ xlab("Índice de calidad de la ocupación")+
  ylab("Proporción de personas de cada sexo")+
  scale_y_continuous(breaks=seq(-.16,.16,.04),
    labels=paste(100*abs(seq(-.16,.16,.04)),"%")) +
  facet_wrap(~REGION_rot)+theme_tufte()
```



```
u<-lm(calidad~sexo+REGION_rot+CH06+CH06^2, data = ocupades)
summary(u)
```

```
##
## Call:
## lm(formula = calidad ~ sexo + REGION_rot + CH06 + CH06^2, data = ocupades)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39829 -0.10176 -0.00303  0.10482  0.65229
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3360068   0.0058300  57.634 < 2e-16 ***
## sexomujeres     0.0263155   0.0028032   9.388 < 2e-16 ***
## REGION_rotNOA   -0.0375524   0.0046271  -8.116 5.30e-16 ***
## REGION_rotNEA   -0.0373533   0.0056540  -6.607 4.10e-11 ***
## REGION_rotCuyo  -0.0470302   0.0053635  -8.769 < 2e-16 ***
## REGION_rotPampeana -0.0116580  0.0043525  -2.678  0.00741 **
## REGION_rotPatagónica 0.0477962  0.0052672   9.074 < 2e-16 ***
## CH06            0.0007949  0.0001088   7.303 2.98e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1497 on 12054 degrees of freedom
```

```
## (3481 observations deleted due to missingness)
## Multiple R-squared:  0.04505,    Adjusted R-squared:  0.0445
## F-statistic: 81.24 on 7 and 12054 DF,  p-value: < 2.2e-16
```