

construccion_indice

```
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_15
```

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
##   924  4466 18624   147
```

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
## 1834 2139 13817  834
```

```
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
##   237   211   113   193   291 13817
```

Obra social

Del cuestionario hogar CH08

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

```
##
##      1      2      3      4      9      12      13      23      123
## 13609   523   141  3796    23   517    12     2     1
```

```
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 13609
##      2           0   523
##      3           0   141
##      4    3796      0
```

```
##    9      0      0
##   12      0    517
##   13      0     12
##   23      0      2
##  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

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    0.00   0.50   0.50   0.48   0.50   0.50   3762
```

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##    0.000   0.500   0.500   0.398   0.500   0.500    23
```

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
##    0.00   1.00   1.00   0.92   1.00   1.00   3775
```

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    9
## 3824  207 1357    2
```

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

```
##
##      1      2      3      9
## 1461 1048 2879      2
```

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

```
##
##      1      2      3      9
## 1755  103 3531      1
```

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

```
##
##      6      7      9
##   25 5364      1
```

```
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      9
##  0      0      0 1357      0
##  1      0  207      0      0
##  2 3824      0      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      9
##  0      0      0 2879      0
##  1      0 1048      0      0
##  2 1461      0      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      9
##  0      0      0 3531      0
##  1      0    103      0      0
##  2 1755      0      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      9
##  0    25      0      0
##  3      0 5364      0
```

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

```
table(cuentapropia$estabilidad)
```

```
##
##  0      1      2      3      4      5      6      7      8      9
##  6      1      9 729 226 1616 620 1337 340 504
```

Obra social (del cuestionario hogar)

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

```
##
##      1      2      3      4      9     12     13     23
## 2398  340   81 2426   16  123     2     4
```

```
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 2398
##  2      0 340
##  3      0  81
##  4 2426      0
##  9      0      0
## 12      0  123
## 13      0      2
## 23      0      4
```

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.   NA's
## 0.0000  0.2778  0.3333  0.3251  0.3889  0.5000     2
```

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000  0.0000  0.5000  0.2743  0.5000  0.5000    16
```

Seguridad

Combiación aditiva de las dos

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.0000  0.2778  0.6667  0.5994  0.8889  1.0000    18
```

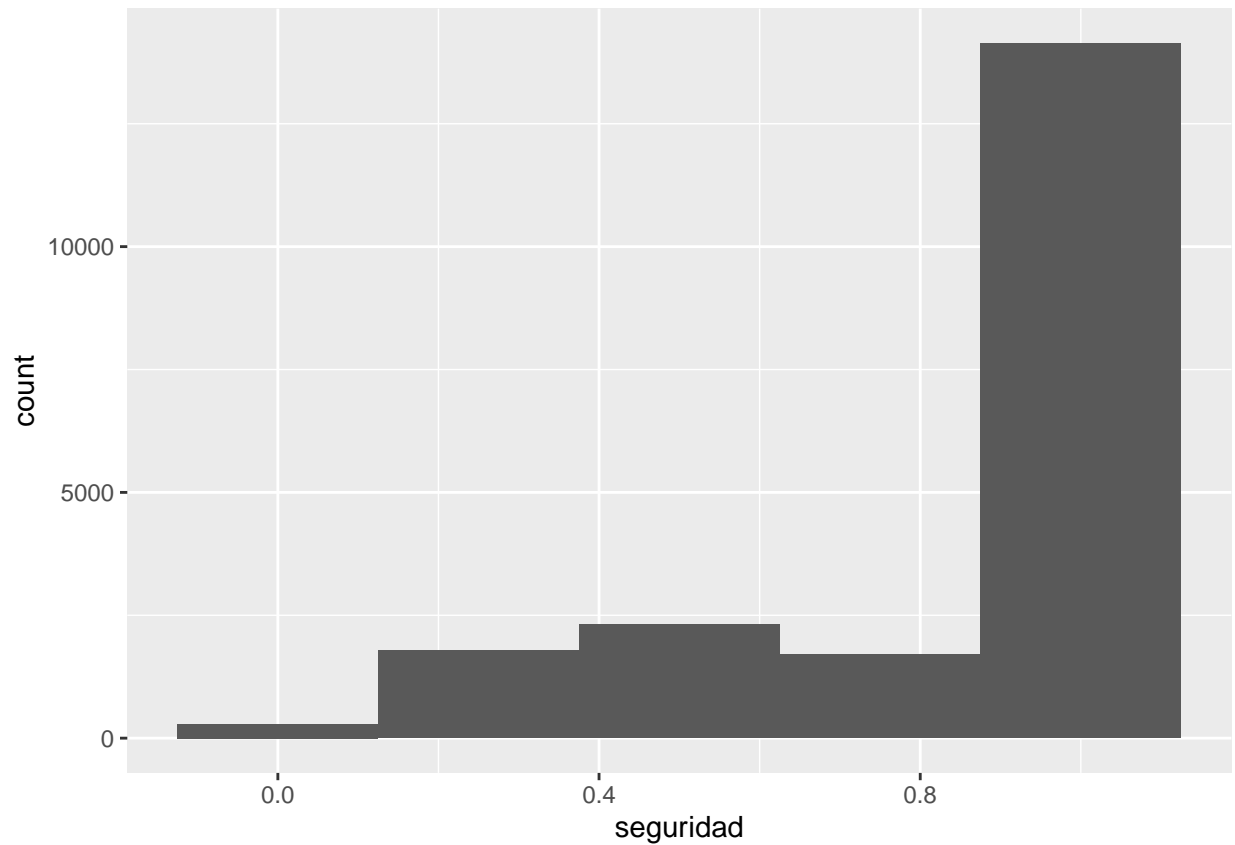
Se eliminan las variables que no están en ambas bases y se unen nuevamente

```
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  0.778  1.000  0.834  1.000  1.000  3793
```

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



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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  0.000   0.778   1.000   0.834   1.000   1.000  3793
```

Consistencia

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

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

```
## [1] "factor"
```

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

```
##
##      1      2      3      4      9
## 1943 4124 12709 5196   42
```



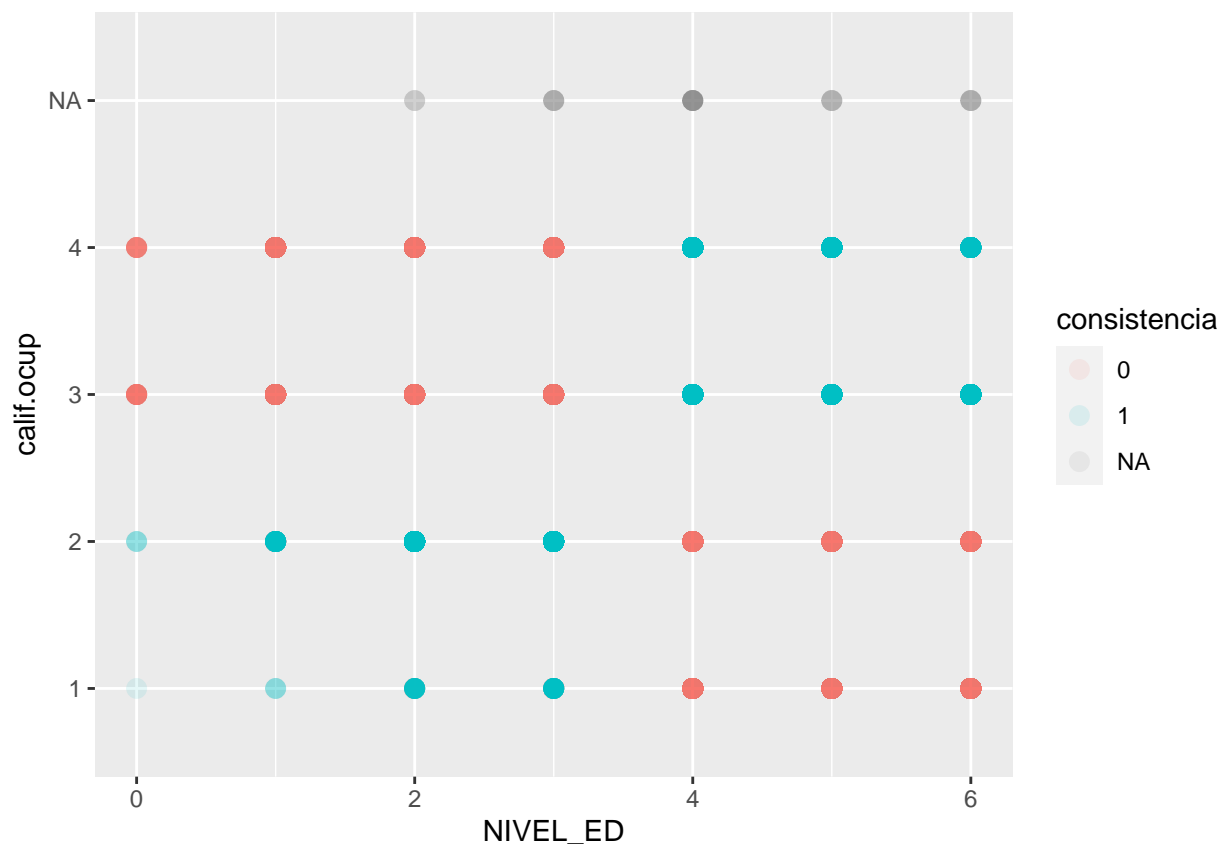
```
ocupades$calif.ocup[ocupades$calif.ocup==7]<-NA
ocupades$calif.ocup[ocupades$calif.ocup==9]<-NA
```

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.
##      -8    3000    5900    6533    8400   300000
```

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.
##      0.8333 23.8095 37.5000 46.6275 56.2500 1388.8889
```

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

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

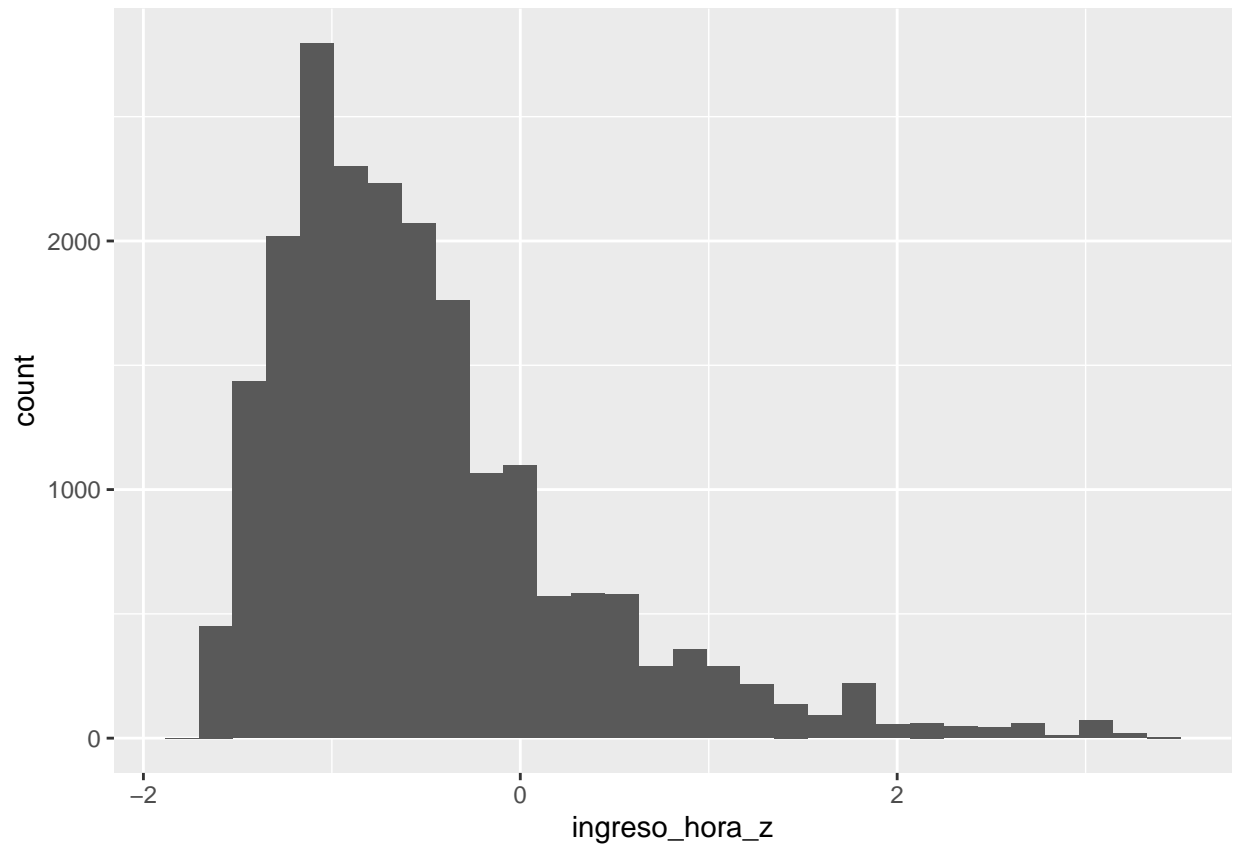
Se estandariza (puntaje z) respecto de las medias y desviaciones (ponderadas) de cada región

```
regiones<-c(1,40,41,42,43,44)
for (i in regiones) {ocupades$ingreso_hora_z=(
  ocupades$ingreso_hora_bruto-wtd.mean(
    ocupades[ocupades$REGION==i,]$ingreso_hora_bruto, weights = ocupades[ocupades$REGION==i,]$PONDERA)),
  ocupades[ocupades$REGION==i,]$ingreso_hora_bruto, weights=ocupades[ocupades$REGION==i,]$PONDERA))
}
summary(ocupades$ingreso_hora_bruto)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.8333 23.4375 37.5000 43.7163 55.5556 185.1852
```

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

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

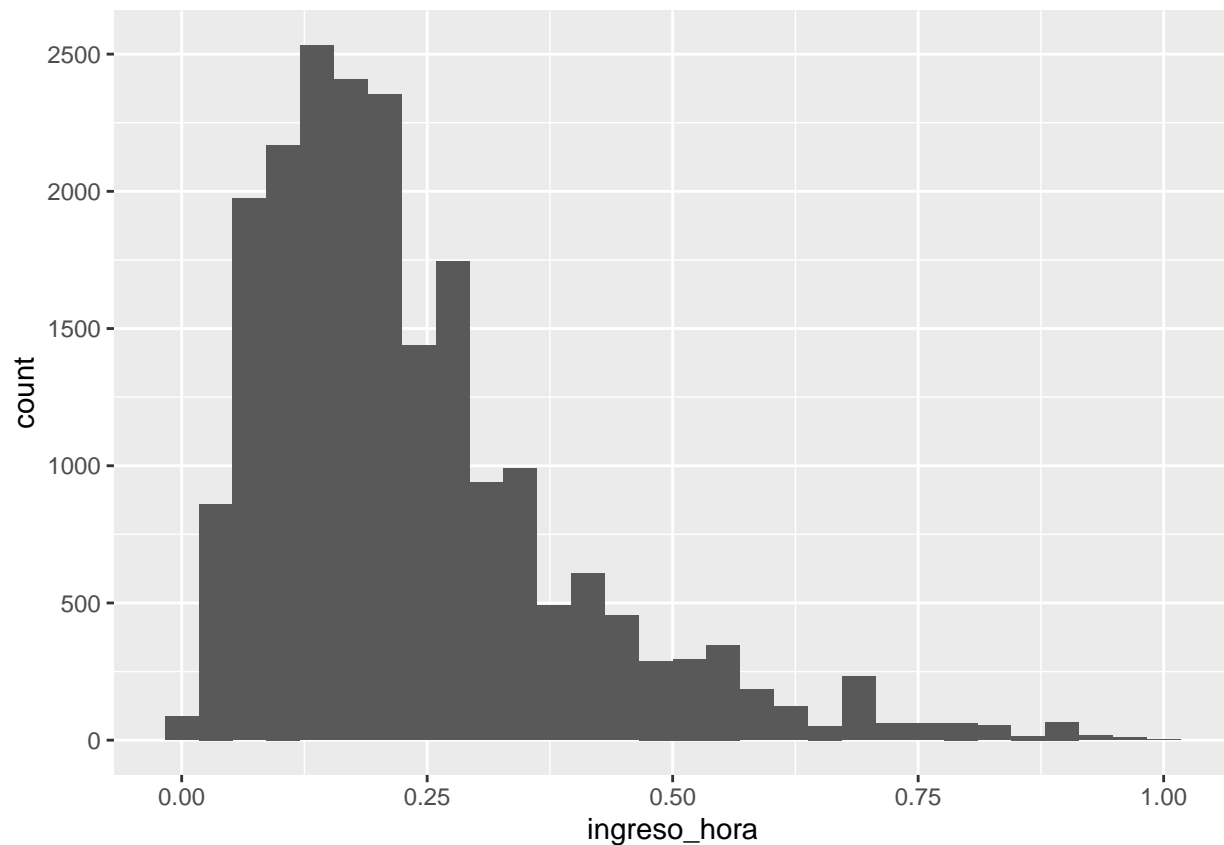


Y se estandariza al [0 1]

```
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

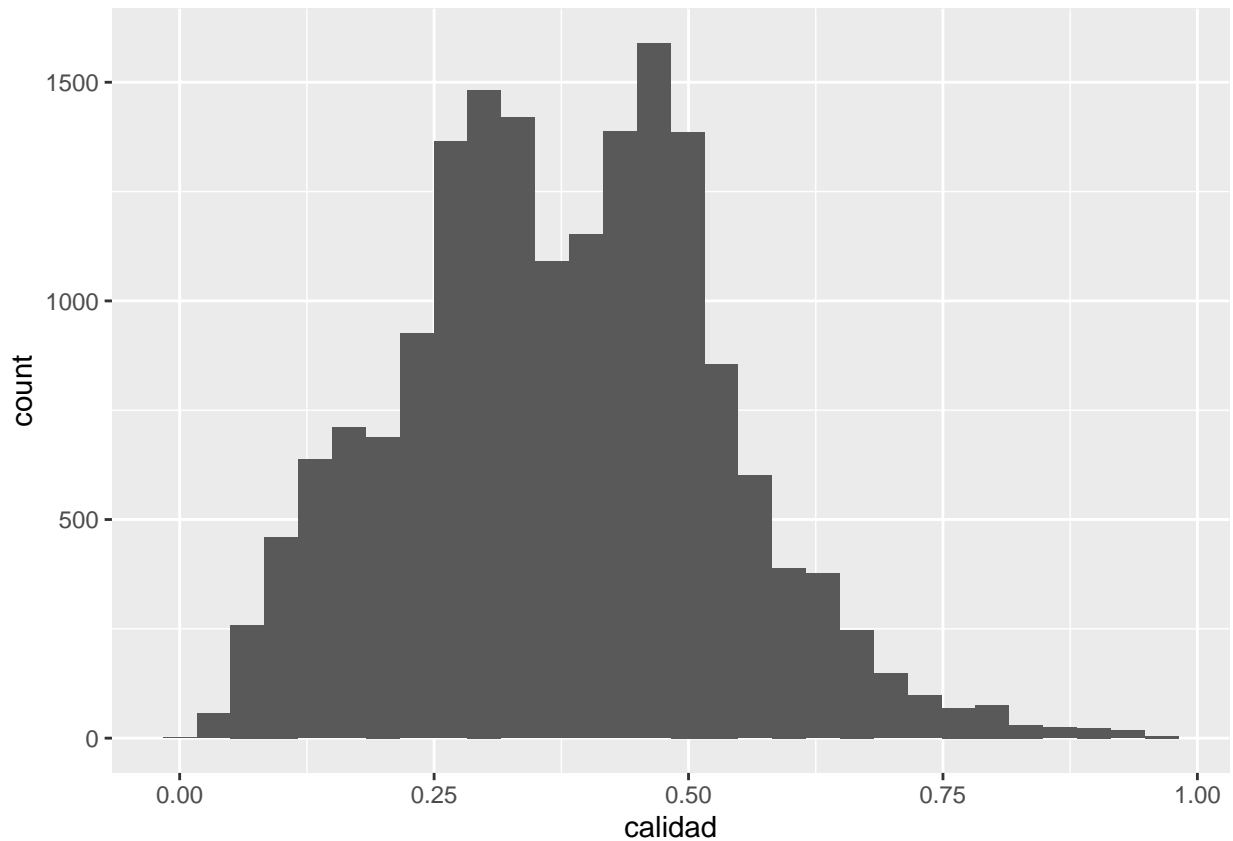
¿será por PCA? por ahora aditivo nomás

```
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
##  0.016  0.265   0.372   0.376  0.481   0.981   3347
```

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

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

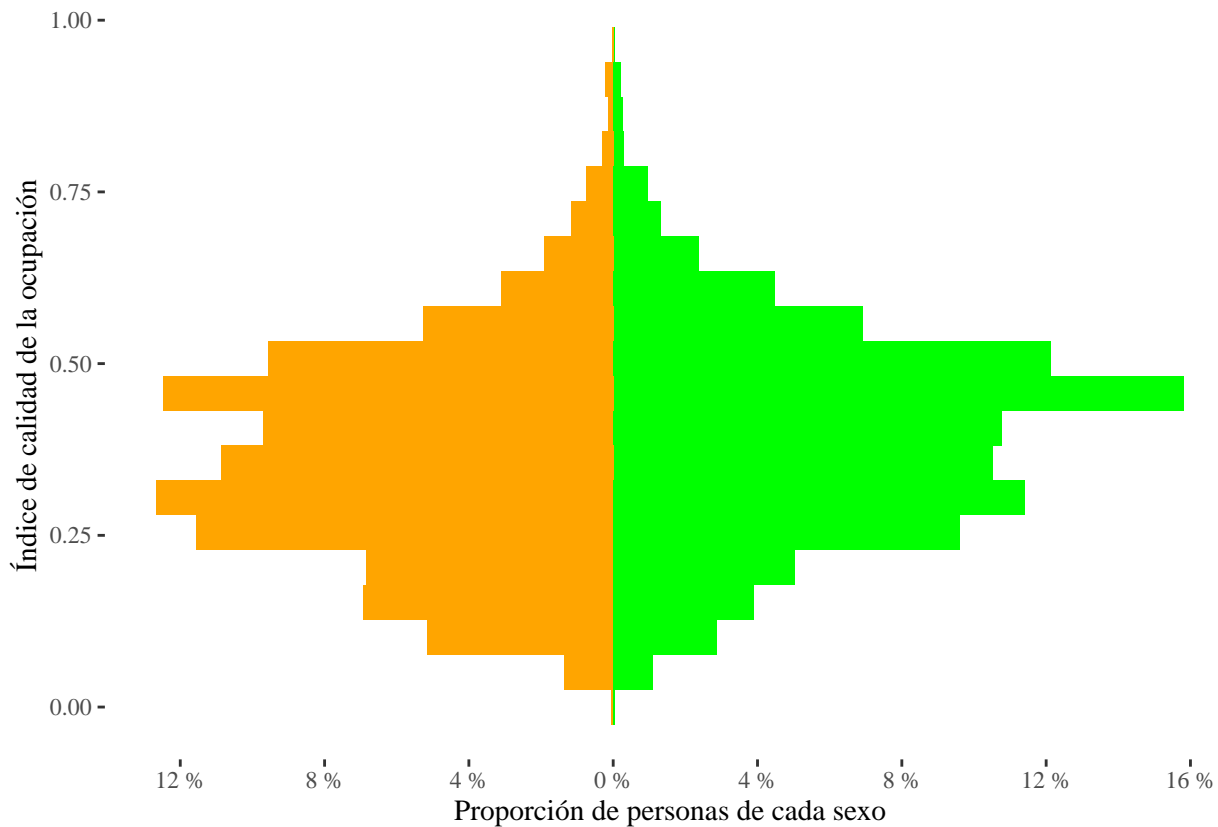


Comparación por sexos

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

```
##
##  Welch Two Sample t-test
##
## data:  ocupades$calidad by ocupades$sexo
## t = -14.9, df = 13145, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.04076028 -0.03128267
## sample estimates:
## mean in group varones mean in group mujeres
##           0.3630771           0.3990986
```

```
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 = -16.597, df = 15171, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.07466073 -0.05888876
## sample estimates:
## mean in group varones mean in group mujeres
## 0.8048610 0.8716357
```

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

```
##
## Welch Two Sample t-test
##
## data: ocupades$consistencia by ocupades$sexo
## t = -7.5277, df = 17718, p-value = 5.411e-14
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.06642077 -0.03897685
## sample estimates:
## mean in group varones mean in group mujeres
##          0.4079094          0.4606082
```

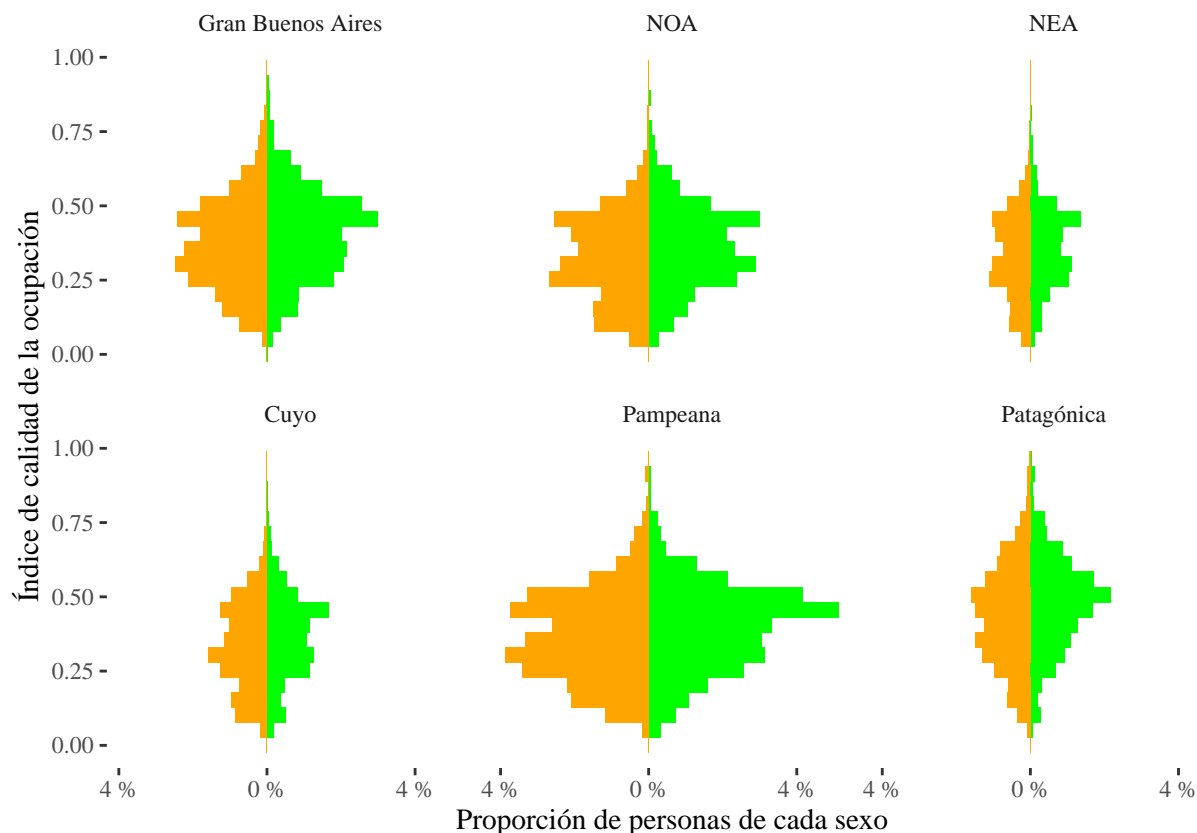
```
t.test(ocupades$ingreso_hora~ocupades$sexo)
```

```
##
## Welch Two Sample t-test
##
## data: ocupades$ingreso_hora by ocupades$sexo
## t = -0.82635, df = 17289, p-value = 0.4086
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.006275952 0.002553567
## sample estimates:
## mean in group varones mean in group mujeres
##          0.2318708          0.2337320
```

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, data = ocupades)
summary(u)
```

```
##
## Call:
## lm(formula = calidad ~ sexo + REGION_rot, data = ocupades)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45290 -0.10636 -0.00319  0.10530  0.56846
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.370589   0.002750  134.774 <2e-16 ***
## sexomujeres     0.035930   0.002380   15.095 <2e-16 ***
## REGION_rotNOA   -0.043688   0.003683  -11.862 <2e-16 ***
## REGION_rotNEA   -0.042330   0.004800   -8.818 <2e-16 ***
## REGION_rotCuyo  -0.036188   0.004379   -8.264 <2e-16 ***
## REGION_rotPampeana -0.003956   0.003345   -1.183  0.237
## REGION_rotPatagónica 0.067471   0.004032  16.732 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1508 on 17562 degrees of freedom
## (3347 observations deleted due to missingness)
```



```
## Multiple R-squared:  0.06255,    Adjusted R-squared:  0.06223  
## F-statistic: 195.3 on 6 and 17562 DF,  p-value: < 2.2e-16
```