



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

FACULTAD DE CIENCIAS

CURSO 2022/23

Análisis estadístico para identificar el impacto de variables socioeconómicas en la distribución espacial del COVID-19 en la C.A. de Castilla y León

ANEXOS

Autor

Raúl Eleazar Tizado Núñez

Tutor

José Luis Vicente Villardón

Salamanca, 2023

A

ANEXOS

A.1	Preprocesamiento de bases de datos – código R	2
A.2	Resultados – código R	5
A.3	Mapas e histogramas	6

A.1 Preprocesamiento de bases de datos – código R

```
pob_ref <- function(year) {
  fname <- sprintf("COVID/Poblacion_de_referencia_20%.csv", year)
  d0 <- read.table(fname, header = TRUE, sep=";", stringsAsFactors = TRUE, fileEncoding="latin1")
  colnames(d0) <- c('Periodo', 'Provincia', 'Area', 'Ambito', 'ID', 'Name', 'Edad', 'Sexo')
  zbs <- unique(d0[c('ID', 'Name', 'Area', 'Provincia', 'Ambito')])
  zbs$Name <- substr(zbs$Name, 8, 9999)
  t <- as.data.frame( table(d0$ID, d0$Sexo) )
  pob <- reshape(t, idvar = 'Var1', timevar = 'Var2', direction = 'wide')
  colnames(pob) <- c('ID', 'Hombre', 'Mujer')
  d <- d0[c('ID', 'Edad')]
  d$Clase <- ifelse(d$Edad < 18, 0, ifelse(d$Edad >= 65, 2, 1))
  d$Clase <- factor(d$Clase, levels = 0:2, labels = c('<18', '18-64', '>=65'))
  t <- as.data.frame( table(d$ID, d$Clase) )
  edad <- reshape(t, idvar = 'Var1', timevar = 'Var2', direction = 'wide')
  colnames(edad) <- c('ID', 'E0-17', 'E18-64', 'E65-00')
  zbs <- merge( zbs, pob, by = 'ID')
  zbs <- merge( zbs, edad, by = 'ID')
  zbs
}

pob_20 <- pob_ref('20')
```

JCYL: Poblacion_de_referencia_20__.csv

```
d0 <- read.table("COVID/tasa-mortalidad-covid-por-zonas-basicas-de-salud.csv", header = TRUE, sep=";", stringsAsFactors = TRUE)
d0$FECHA <- as.Date(d0$FECHA)
d <- d0[ c('FECHA', 'CS', 'FALLECIDOS') ]
d$DIA <- factor( format(d$FECHA, format = '%d') )
d$MES <- factor( format(d$FECHA, format = '%m') )
d$ANO <- factor( format(d$FECHA, format = '%y') )

fallecidos <- function(data, year) {
  x <- subset(data, ANO == year)
  x <- aggregate(x['FALLECIDOS'], by=x['CS'], FUN=sum)
  colnames(x) <- c('ID', 'Fallecidos')
  x
}

fal_20 <- fallecidos(d, '20')
```

JCYL: tasa-mortalidad-Covid-19 -por-zonas-basicas-de-salud.csv

```
d0 <- read.table("COVID/tasa-enfermos-acumulados-por-areas-de-salud.csv", header = TRUE, sep=";", stringsAsFactors = TRUE)
d0$FECHA <- as.Date(d0$FECHA)
d <- d0[ c('FECHA', 'CS', 'TOTALENFERMEDAD', 'TOTALENFERMEDAD_7DIAS', 'TOTALENFERMEDAD_14DIAS', 'PCR_REALIZADOS', 'PCR_POSITIVOS') ]
d$DIA <- factor( format(d$FECHA, format = '%d') )
d$MES <- factor( format(d$FECHA, format = '%m') )
d$ANO <- factor( format(d$FECHA, format = '%y') )

enfermos <- function(data, year) {
  x <- subset(data, ANO == year)
  x[ is.na(x) ] <- 0
  x <- aggregate(x[c('TOTALENFERMEDAD', 'TOTALENFERMEDAD_7DIAS', 'TOTALENFERMEDAD_14DIAS', 'PCR_REALIZADOS', 'PCR_POSITIVOS')], by=x['CS'], FUN=sum)
  colnames(x) <- c('ID', 'Enfermos', 'Enfermos7', 'Enfermos14', 'PDIA', 'PDIA+')
  x
}

enf_20 <- enfermos(d, '20')
```

JCYL: tasa-enfermos-acumulados-por-areas-de-salud.csv

```

d0 <- read.table("COVID/prevalencia-coronavirus.csv", header = TRUE, sep=";", stringsAsFactors = TRUE)
d0$FECHA <- as.Date(d0$FECHA)
d <- d0[ c('FECHA','CS','PREVALENCIA') ]
d$DIA <- factor( format(d$FECHA, format = '%d') )
d$MES <- factor( format(d$FECHA, format = '%m') )
d$ANO <- factor( format(d$FECHA, format = '%y') )

prevalencia <- function(data, year) {
  x <- subset(data, ANO == year)
  x[ is.na(x) ] <- 0
  x <- aggregate(x['PREVALENCIA'], by=x['CS'], FUN=mean)
  colnames(x) <- c('ID','Prevalencia')
  x
}

pre_20 <- prevalencia(d,'20')

```

JCYL: prevalencia-coronavirus.csv

```

prov_name <- c('AVILA','BURGOS','LEON','PALENCIA','SALAMANCA','SEGOVIA','SORIA','VALLADOLID','ZAMORA')
prov_ine1 <- c( 30877 , 30934 , 31078 , 31150 , 31186 , 31204 , 31222 , 31267 , 31276 )

d0 <- NULL
for(i in 1:9) {
  fname <- sprintf("INE/%d_%s_ID.csv", prov_ine1[i], prov_name[i])
  d <- read.table(fname, header = TRUE, sep=";", dec = ',',
                 colClasses = c('character','character','character','factor','factor','character') )
  # Hay decimales con , y con . y si no lo coge character falla, ya se pasará luego a número
  d0 <- rbind(d0, d)
}

d0 <- subset(d0, d0$Distrito == '')
d0 <- d0 [ d0$Total != '.', ]
d0$value <- fix_number( d0$Total )
d0$IDm <- as.integer(gsub("\\D", "", d0$Municipios))

demografic <- function(data, year) {
  x <- subset(data, Periodo == year)
  field <- 'Indicadores.demográficos'
  x <- x[ c('IDm', field, 'value') ]
  y <- reshape(x, idvar = 'IDm', timevar = field, v.names = 'value', direction = 'wide')
  colnames(y)[colnames(y) == "value.Edad media de la población"] <- "Edad_media"
  colnames(y)[colnames(y) == "value.Porcentaje de población menor de 18 años"] <- "Menor_18"
  colnames(y)[colnames(y) == "value.Porcentaje de población de 65 y más años"] <- "Mayor_64"
  colnames(y)[colnames(y) == "value.Tamaño medio del hogar"] <- "Hogar_size"
  colnames(y)[colnames(y) == "value.Porcentaje de hogares unipersonales"] <- "Hogar_unip"
  colnames(y)[colnames(y) == "value.Población"] <- "Poblacion"
  colnames(y)[colnames(y) == "value.Porcentaje de población española"] <- "Pob_spain"
  y[ is.na(y$Pob_spain), 'Pob_spain'] <- 100.0
  y$Menor_18 <- round(y$Menor_18 * y$Poblacion / 100.0 ,0)
  y$Mayor_64 <- round(y$Mayor_64 * y$Poblacion / 100.0 ,0)
  y$Entre_18_64 <- y$Poblacion - y$Menor_18 - y$Mayor_64
  y$Pob_spain <- round(y$Pob_spain * y$Poblacion / 100.0 , 0)
  y$Pob_no_sp <- y$Poblacion - y$Pob_spain
  y[ c('IDm','Poblacion','Edad_media','Menor_18','Entre_18_64','Mayor_64','Hogar_size','Hogar_unip','Pob_spain','Pob_no_sp') ]
}

mun_20 <- demografic(d0, '2020')

```

INE: Indicadores demográficos

```

prov_name <- c('AVILA', 'BURGOS', 'LEON', 'PALENCIA', 'SALAMANCA', 'SEGOVIA', 'SORIA', 'VALLADOLID', 'ZAMORA')
prov_ine2 <- c( 30870 , 30927 , 31071 , 31143 , 31179 , 31197 , 31215 , 31260 , 31269 )
d0 <- NULL
for(i in 1:9) {
  fname <- sprintf("INE/%s_INGRESO_%d.csv", prov_name[i], prov_ine2[i])
  d <- read.table(fname, header = TRUE, sep=";", dec = ',',
    colClasses = c('character', 'character', 'character', 'factor', 'factor', 'character') )
  d <- read.table(fname, header = TRUE, sep=";", dec = ',')
  d0 <- rbind(d0, d)
}
d0 <- subset(d0, d0$Distrito == '')
d0$Distribución.por.fuente.de.ingresos <- factor(d0$Distribución.por.fuente.de.ingresos)
d0$Periodo <- d0$Periodo
d0 <- d0 [ d0$Total != '.', ]
d0$value <- fix_number( d0$Total )
d0$IDm <- as.integer(gsub("\\D", "", d0$Municipios))
d0[ is.na(d0$value), 'value'] <- 0

ingresos <- function(data, year) {
  x <- subset(data, Periodo == year)
  field <- 'Distribución.por.fuente.de.ingresos'
  x <- x[ c('IDm', field, 'value') ]
  y <- reshape(x, idvar = 'IDm', timevar = field, v.names = 'value', direction = 'wide')
  colnames(y)[colnames(y) == "value.Renta bruta media por persona"] <- "Renta_bruta"
  colnames(y)[colnames(y) == "value.Fuente de ingreso: salario"] <- "Ing_salario"
  colnames(y)[colnames(y) == "value.Fuente de ingreso: pensiones"] <- "Ing_pensiones"
  colnames(y)[colnames(y) == "value.Fuente de ingreso: prestaciones por desempleo"] <- "Ing_desempleo"
  colnames(y)[colnames(y) == "value.Fuente de ingreso: otras prestaciones"] <- "Ing_otras_prestaciones"
  colnames(y)[colnames(y) == "value.Fuente de ingreso: otros ingresos"] <- "Ing_otros"
  y[ c('IDm', 'Renta_bruta', 'Ing_salario', 'Ing_pensiones', 'Ing_desempleo', 'Ing_otras_prestaciones', 'Ing_otros') ]
  y
}

ing_20 <- ingresos(d0, '2020')

ing_20[ is.na(ing_20$Renta_bruta), 'Renta_bruta'] <-
  ing_20[ is.na(ing_20$Renta_bruta), 'Ing_salario'] +
  ing_20[ is.na(ing_20$Renta_bruta), 'Ing_pensiones'] +
  ing_20[ is.na(ing_20$Renta_bruta), 'Ing_desempleo'] +
  ing_20[ is.na(ing_20$Renta_bruta), 'Ing_otras_prestaciones'] +
  ing_20[ is.na(ing_20$Renta_bruta), 'Ing_otros']

```

INE: Distribución por fuente de ingresos

```

prov_name <- c('AVILA', 'BURGOS', 'LEON', 'PALENCIA', 'SALAMANCA', 'SEGOVIA', 'SORIA', 'VALLADOLID', 'ZAMORA')
prov_ine4 <- c( 30869 , 30926 , 31070 , 31142 , 31178 , 31196 , 31214 , 31259 , 31268 )
d0 <- NULL
for(i in 1:9) {
  fname <- sprintf("INE/%d_%s.csv", prov_ine4[i], prov_name[i])
  d <- read.table(fname, header = TRUE, sep=";", dec = ',')
  # Hay decimales con , y con . y si no lo coge character falla, ya se pasará luego a número
  d0 <- rbind(d0, d)
}
d0 <- subset(d0, d0$Distrito == '')
d0$Indicadores.de.renta.media.y.mediana <- factor(d0$Indicadores.de.renta.media.y.mediana)
d0$Periodo <- factor(d0$Periodo)

d0 <- d0 [ d0$Total != '.', ]
d0$value <- fix_number( d0$Total )
d0$IDm <- as.integer(gsub("\\D", "", d0$Municipios))

renta <- function(data, year) {
  x <- subset(data, Periodo == year)
  field <- 'Indicadores.de.renta.media.y.mediana'
  x <- x[ c('IDm', field, 'value') ]
  y <- reshape(x, idvar = 'IDm', timevar = field, v.names = 'value', direction = 'wide')
  colnames(y)[colnames(y) == "value.Renta neta media por persona "] <- "Neta_persona"
  colnames(y)[colnames(y) == "value.Renta neta media por hogar"] <- "Neta_hogar"
  colnames(y)[colnames(y) == "value.Mediana de la renta por unidad de consumo"] <- "Mediana_unidad"
  colnames(y)[colnames(y) == "value.Mediana de la renta por unidad de consumo"] <- "Mediana_unidad"
  colnames(y)[colnames(y) == "value.Renta bruta media por persona"] <- "Bruta_persona"
  colnames(y)[colnames(y) == "value.Renta bruta media por hogar"] <- "Bruta_hogar"
  y[ c('IDm', 'Neta_persona', 'Neta_hogar', 'Mediana_unidad', 'Mediana_unidad', 'Bruta_persona', 'Bruta_hogar') ]
  y
}

renta_20 <- renta(d0, '2020')

```

INE: Indicadores de renta media y mediana


```

prov_name <- c('AVILA', 'BURGOS', 'LEON', 'PALENCIA', 'SALAMANCA', 'SEGOVIA', 'SORIA', 'VALLADOLID', 'ZAMORA')
prov_line3 <- c( 37683 , 37687 , 37703 , 37708 , 37713 , 37715 , 37717 , 37722 , 37723 )
d0 <- NULL
for(i in 1:9) {
  fname <- sprintf("INE/%d_%s_GINI.csv", prov_line3[i], prov_name[i])
  d <- read.table(fname, header = TRUE, sep=";", dec = ',')
  # Hay decimales con , y con . y si no lo coge character falla, ya se pasará luego a número
  d0 <- rbind(d0, d)
}
d0 <- subset(d0, d0$Distrito == '')
d0$Índice.de.Gini.y.Distribución.de.la.renta.P80.P20 <- factor(d0$Índice.de.Gini.y.Distribución.de.la.renta.P80.P20)
d0 <- d0 [ d0$Total != '.', ]
d0$value <- fix_number( d0$Total )
d0$IDm <- as.integer(gsub("\\D", "", d0$Municipios))

gini <- function(data, year) {
  x <- subset(data, Período == year)
  field <- 'Índice.de.Gini.y.Distribución.de.la.renta.P80.P20'
  x <- x[ c('IDm', field, 'value') ]
  y <- reshape(x, idvar = 'IDm', timevar = field, v.names = 'value', direction = 'wide')
  colnames(y)[colnames(y) == "value.Distribución de la renta P80/P20"] <- "P80/P20"
  colnames(y)[colnames(y) == "value.Índice de Gini"] <- "Gini"
  y
}

gini_20 <- gini(d0, '2020')

```

INE: Índice de Gini y Distribución de la renta P80/P20

A.2 Resultados – código R

```

library(corrplot)

data_san <- final_20[ socio_san ]
rownames(data_san) <- final_20$IDz
summary(data_san)

ss_cor <- cor(data_san)
summary(ss_cor)
corrplot(ss_cor, method='square', order='hclust', hclust.method='ward.D', addrect = 6, col.lim = c(0.6, 1), is.corr = FALSE)

socio_san_sel <- c("E0.17", "E18.64", "E65.00", "Fallecidos", "Prevalencia", "Enfermos")

h <- hclust( as.dist(1 - cor(data_san)), method = 'ward.D' )
plot(h, main = NA, xlab = NA, sub = NA, cex=1, lwd=1)

data_eco <- final_20[ socio_eco ]
rownames(data_eco) <- final_20$IDz
summary(data_eco)

ss_cor <- cor(data_eco)
summary(ss_cor)
corrplot(ss_cor, method='square', order='hclust', hclust.method='ward.D', addrect = 5, col.lim = c(-1, 1), is.corr = FALSE)

socio_eco_sel <- c("Edad_media", "Renta_bruta", "Neta_persona", "Gini", "Pob_ine")

h <- hclust( as.dist(1 - cor(data_san)), method = 'ward.D' )
plot(h, main = NA, xlab = NA, sub = NA, cex=1, lwd=1)

data_all <- cbind(data_eco, data_san)
summary(data_all)

ss_cor <- cor(data_all)
summary(ss_cor)
corrplot(ss_cor, method='square', order='hclust', hclust.method='ward.D', addrect = 6, col.lim = c(-1, 1), is.corr = FALSE)

h <- hclust( as.dist(1 - cor(data_san)), method = 'ward.D' )
plot(h, main = NA, xlab = NA, sub = NA, cex=1, lwd=1)

```

Análisis de Correlaciones

```

library(FactoMineR)
library(factoextra)

data <- cbind(data_eco, data_san)
data_san_sel <- data[ socio_san_sel ]
data_eco_sel <- data[ socio_eco_sel ]
data_sel <- cbind(data_eco_sel, data_san_sel)

acp <- PCA(data_sel, scale.unit = TRUE, ncp = 4, graph = FALSE)

fviz_eig(acp, addlabels = TRUE, choice = 'variance', geom = 'line', hjust = -0.2)

var <- get_pca_var(acp)
var$contrib

fviz_contrib(acp, choice = "var", axes = 1, top = 10)
fviz_contrib(acp, choice = "var", axes = 2, top = 10)
fviz_contrib(acp, choice = "var", axes = 3, top = 10)
fviz_contrib(acp, choice = "var", axes = 4, top = 10)

fviz_pca_var(acp, repel = TRUE, axes = c(1,2))
fviz_pca_var(acp, repel = TRUE, axes = c(3,4))

amb <- factor ( final_20$Ambito )

fviz_pca_biplot(acp,
  col.ind = amb,
  geom.ind = "point", pointshape = 21, pointsize = 2,
  palette = "jco", label = "var",
  col.var = "black", repel = TRUE,
  legend.title = "Ambito", axes = c(1,2))

fviz_pca_biplot(acp,
  col.ind = amb,
  geom.ind = "point", pointshape = 21, pointsize = 2,
  palette = "jco", label = "var",
  col.var = "black", repel = TRUE,
  legend.title = "Ambito", axes = c(3,4))

```

Análisis de Componentes Principales

```

library(CCA)

cc1 <- cc(data_san_sel, data_eco_sel)
cc1$cor
cc1$scores

plt.cc(cc1, d1=1, d2=2, var.label = T, int = 0)
plt.cc(cc1, d1=3, d2=4, var.label = T, int = 0)

```

Análisis de Correlación Canónica

A.3 Mapas e histogramas

Mapas de cuartiles de las variables socioeconómicas y sociosanitarias estudiadas y su histograma de distribución de frecuencias.

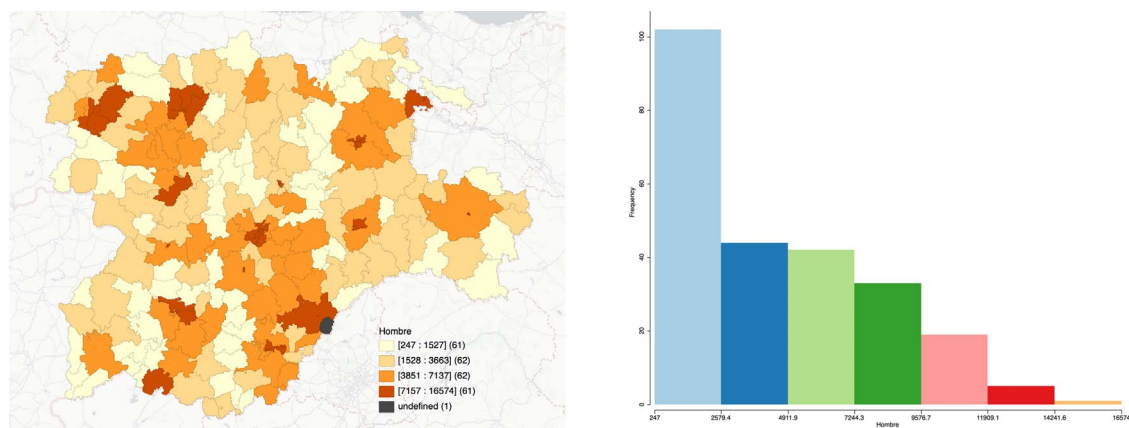


Figura A.1: Distribución espacial e histograma: Hombres

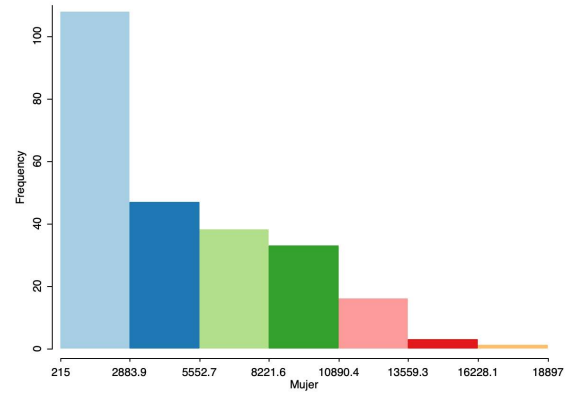
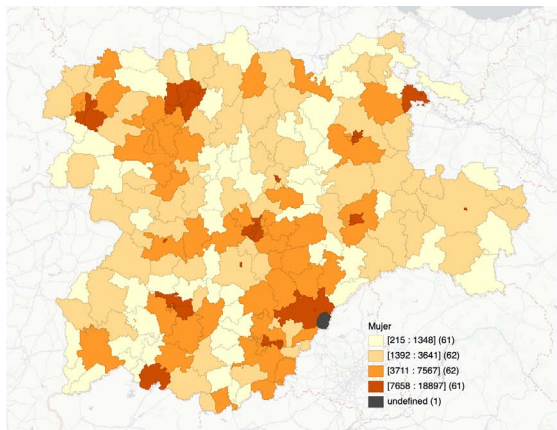


Figura A.2: Distribución espacial e histograma: Mujeres

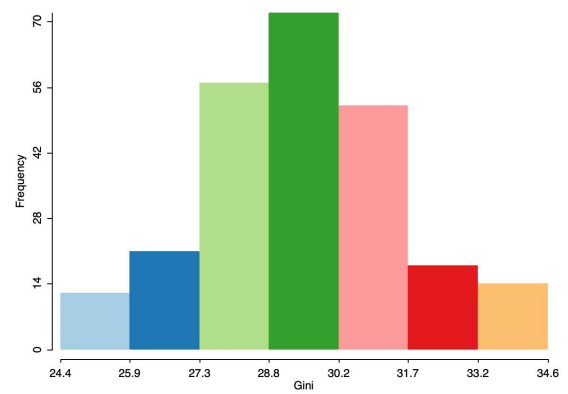
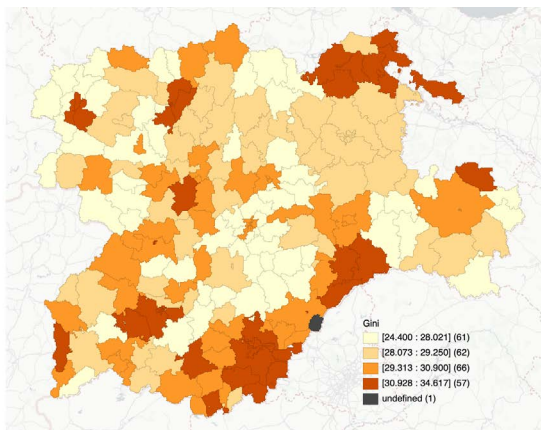


Figura A.3: Distribución espacial e histograma: Índice de Gini

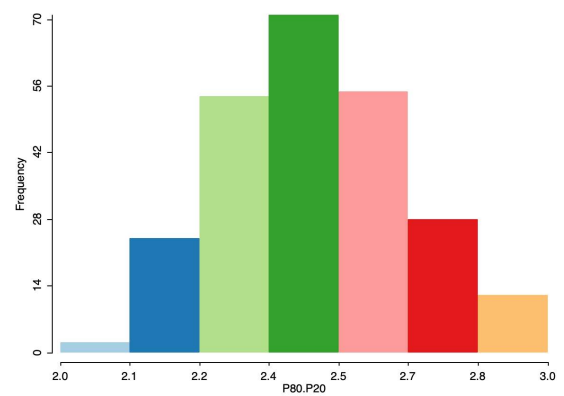
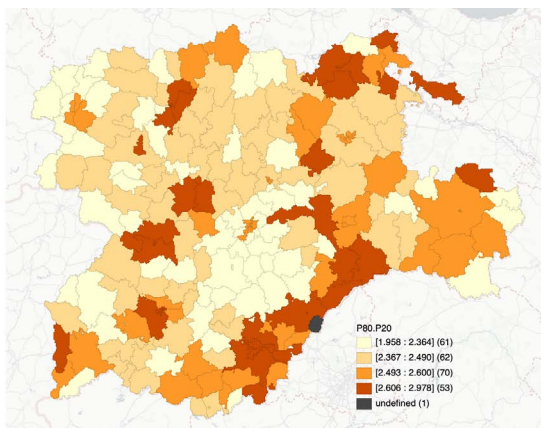


Figura A.4: Distribución espacial e histograma: Ratio entre percentil 80 y percentil 20

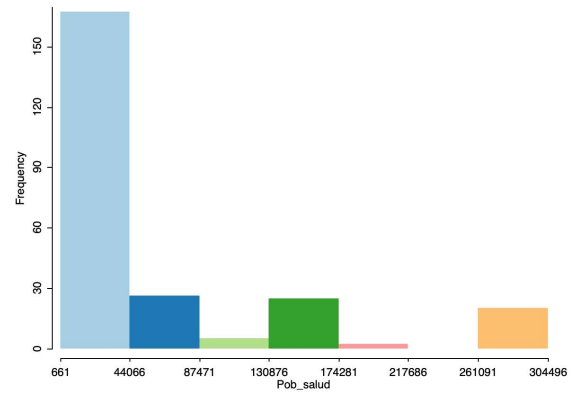
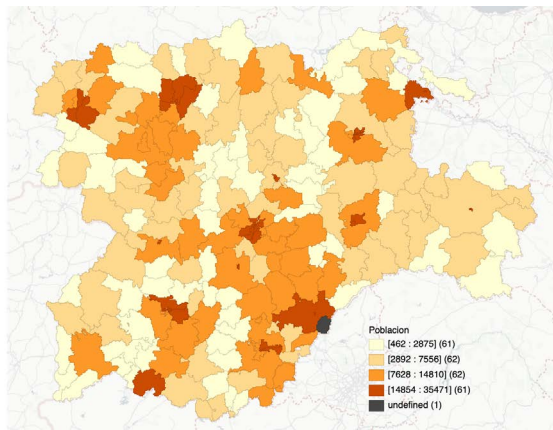


Figura A.5: Distribución espacial e histograma: Población sanitaria total

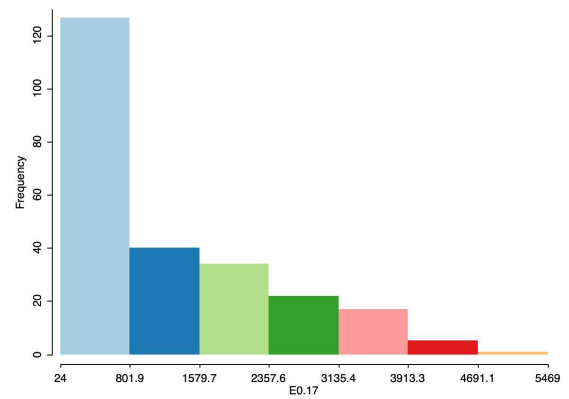
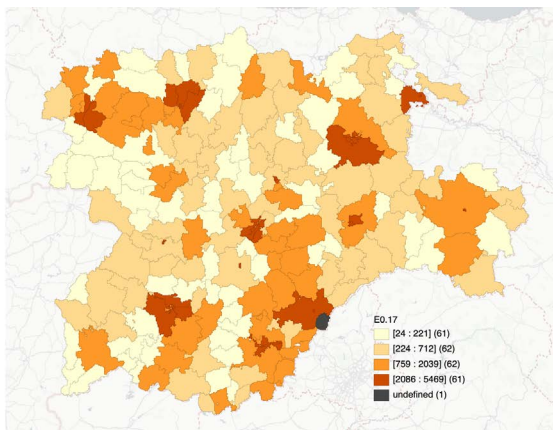


Figura A.6: Distribución espacial e histograma: Población sanitaria menor de 18 años

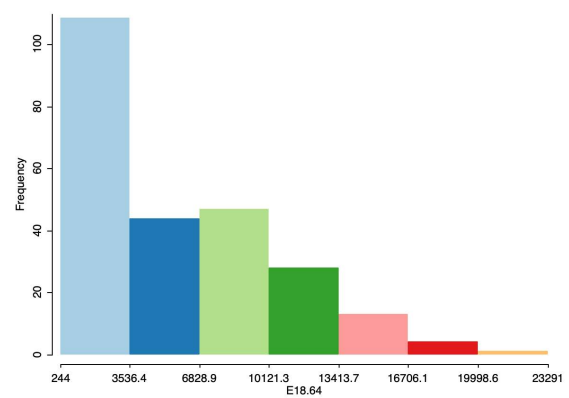
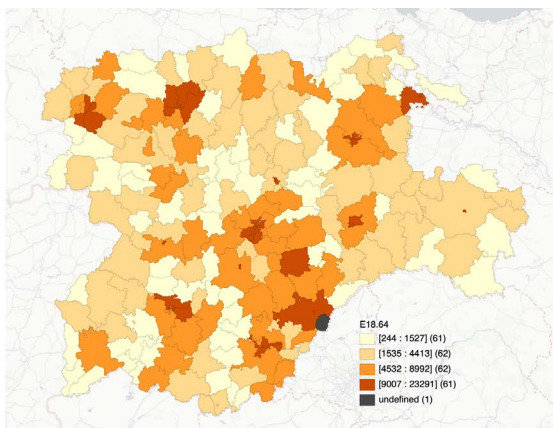


Figura A.7: Distribución espacial e histograma: Población sanitaria entre de 18 y 64 años

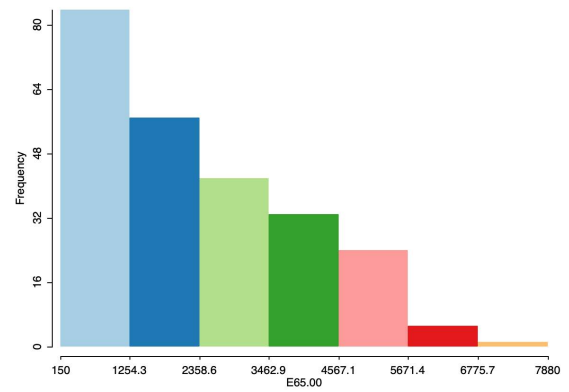
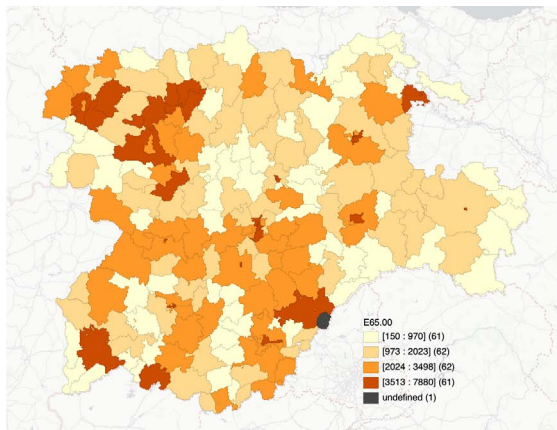


Figura A.8: Distribución espacial e histograma: Población sanitaria mayor de 64 años

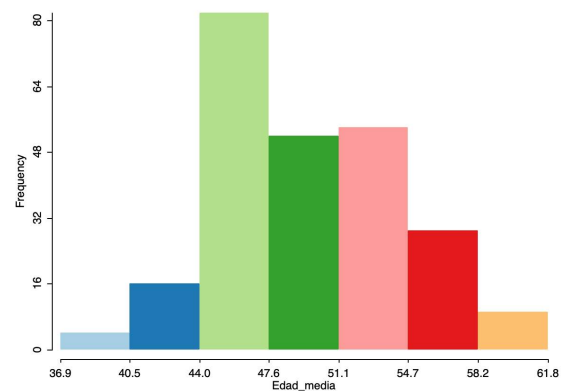
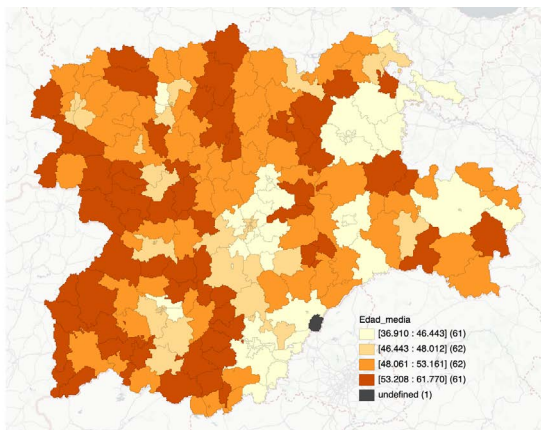


Figura A.9: Distribución espacial e histograma: Edad media

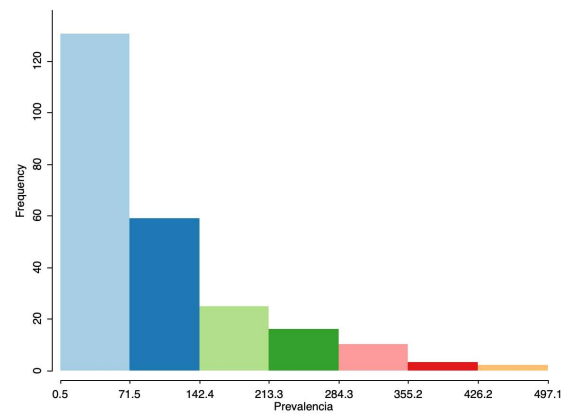
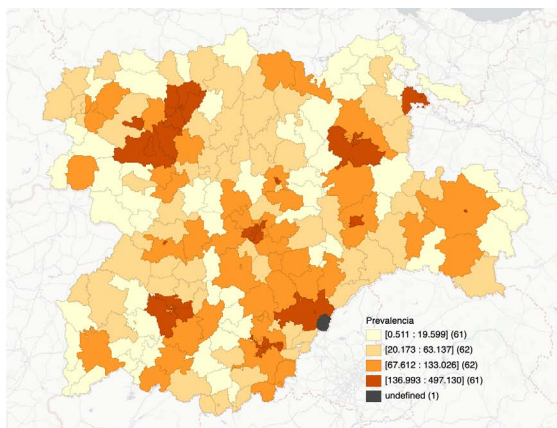


Figura A.10: Distribución espacial e histograma: Prevalencia del COVID-19

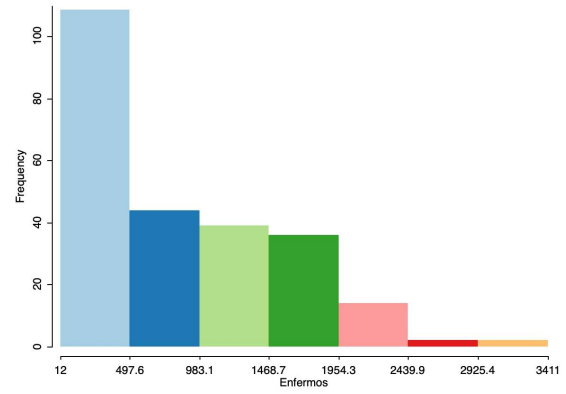
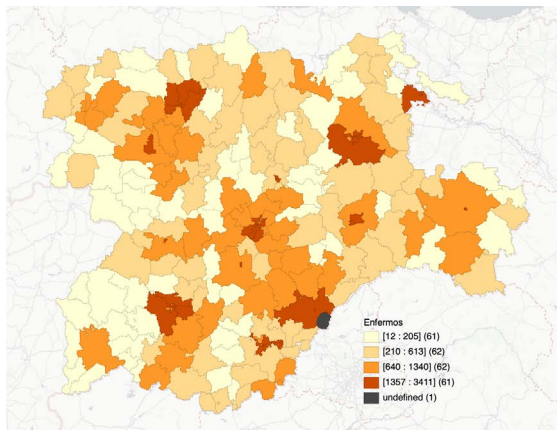


Figura A.11: Distribución espacial e histograma: Enfermos de COVID-19

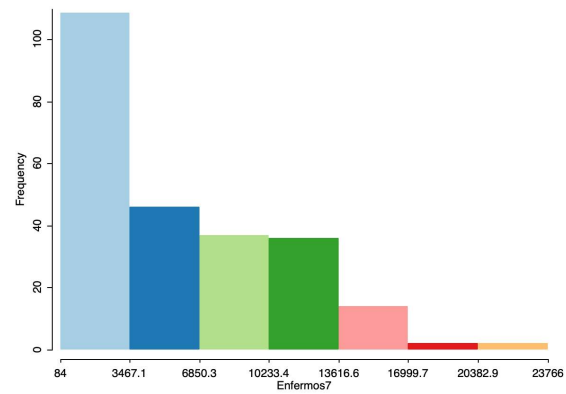
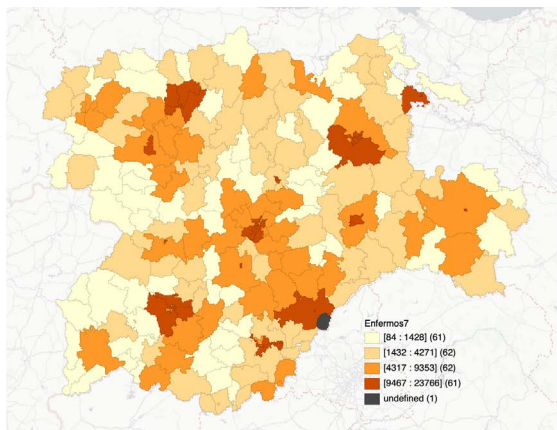


Figura A.12: Distribución espacial e histograma: Enfermos de COVID-19 acumulados en los últimos 7 días

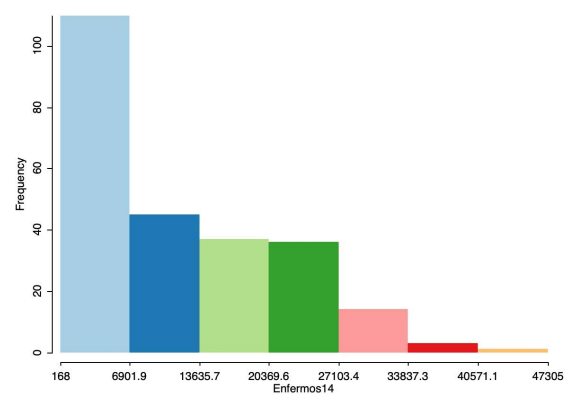
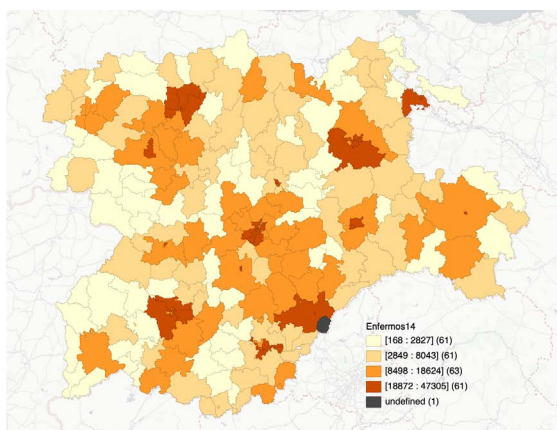


Figura A.13: Distribución espacial e histograma: Enfermos de COVID-19 acumulados en los últimos 14 días

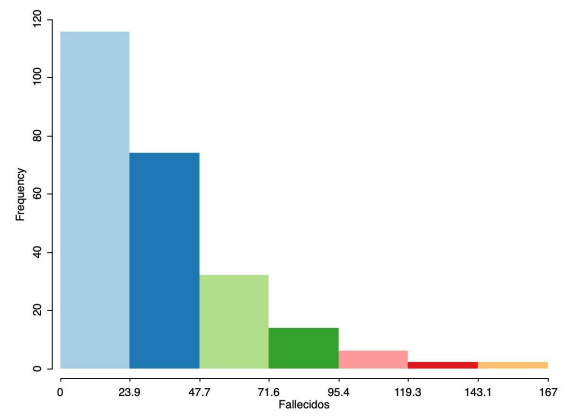
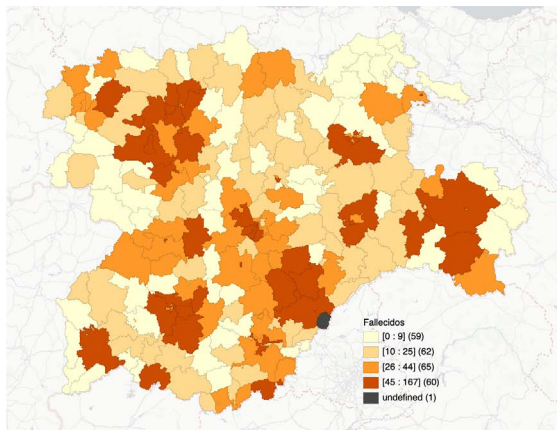


Figura A.14: Distribución espacial e histograma: Número de fallecimientos

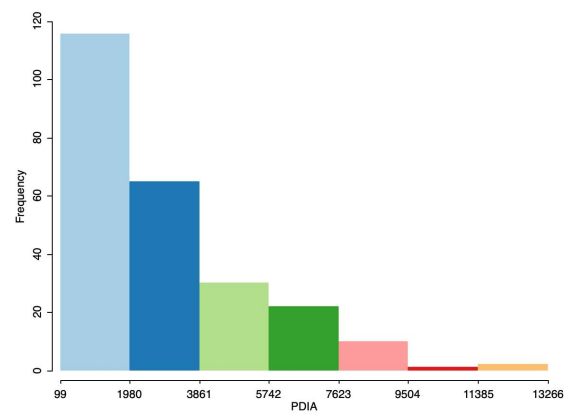
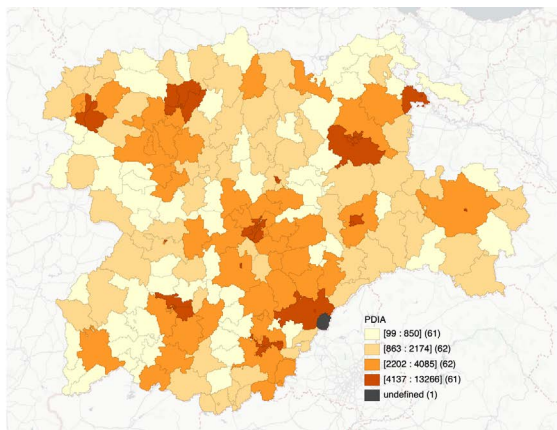


Figura A.15: Distribución espacial e histograma: PDIA

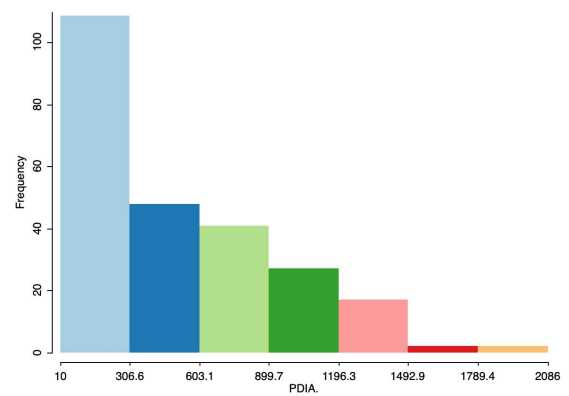
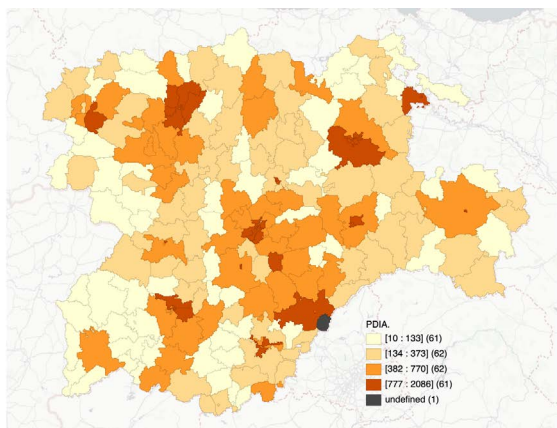


Figura A.16: Distribución espacial e histograma: PDIA+

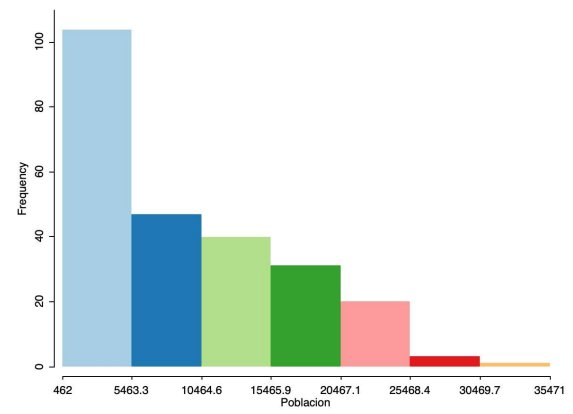
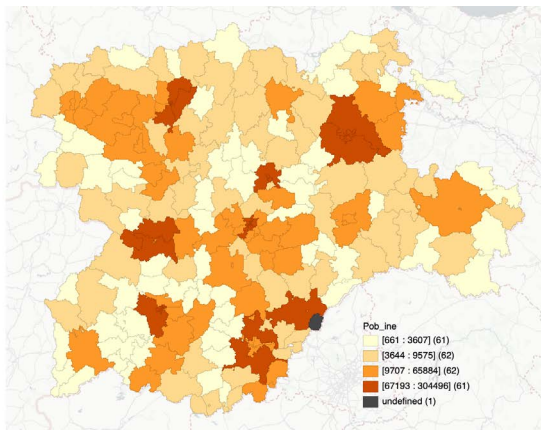


Figura A.17: Distribución espacial e histograma: Población total INE

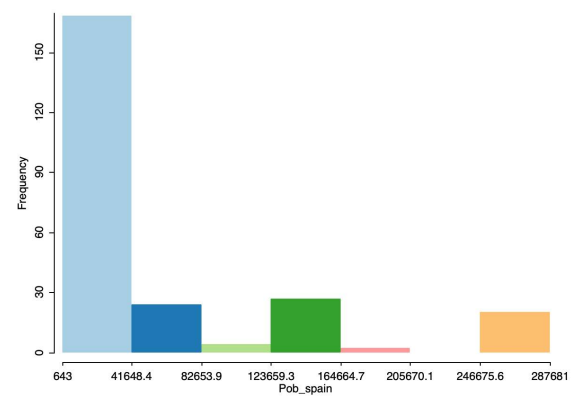
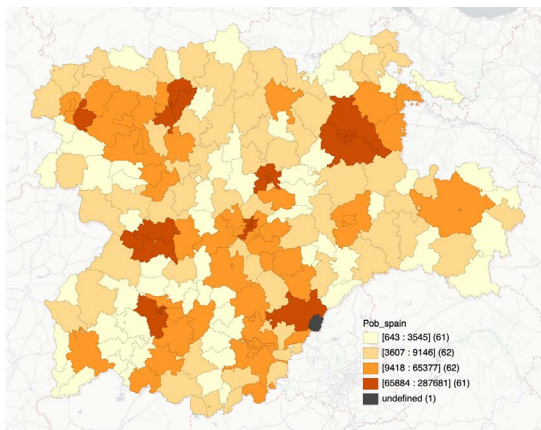


Figura A.18: Distribución espacial e histograma: Población de nacionalidad española

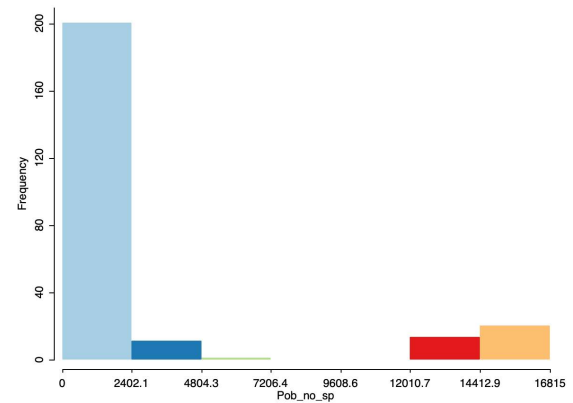
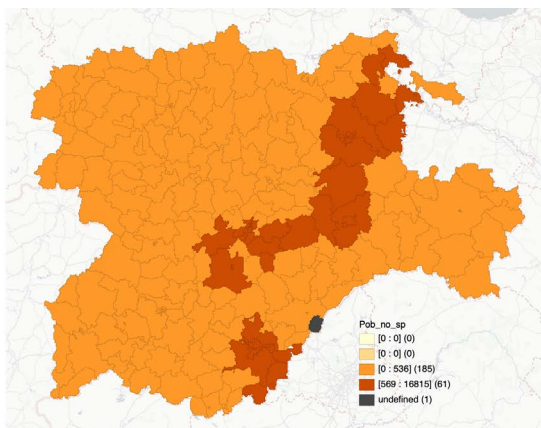


Figura A.19: Distribución espacial e histograma: Población de nacionalidad no española

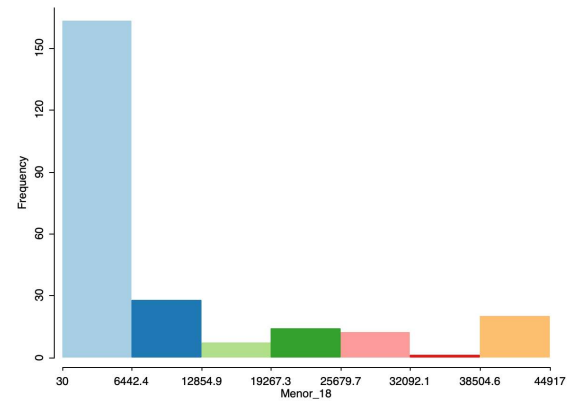
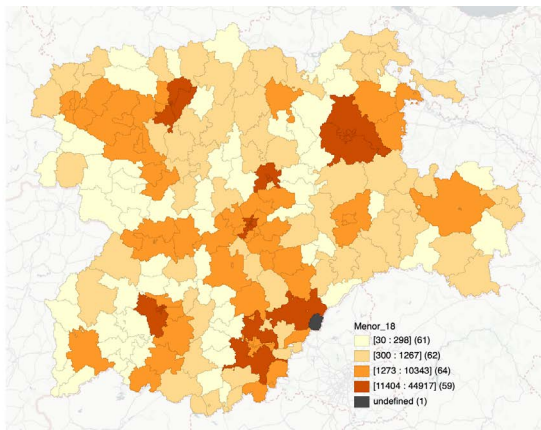


Figura A.20: Distribución espacial e histograma: Población menor de 18 años

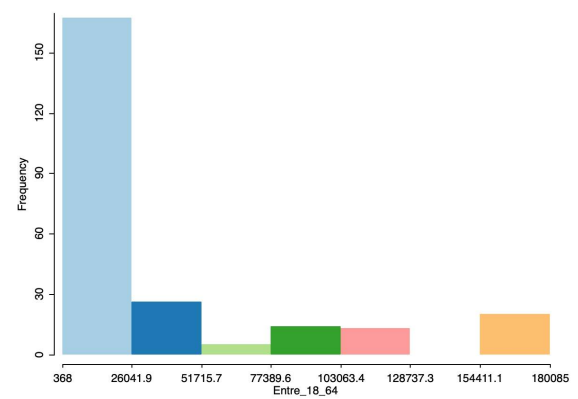
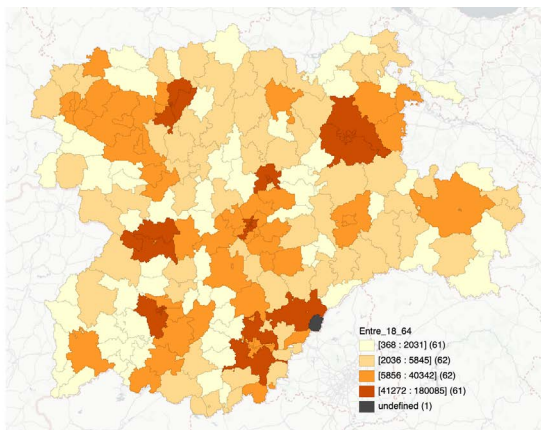


Figura A.21: Distribución espacial e histograma: Población entre de 18 y 64 años

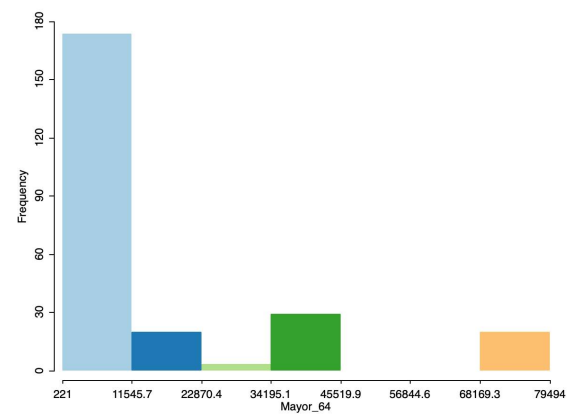
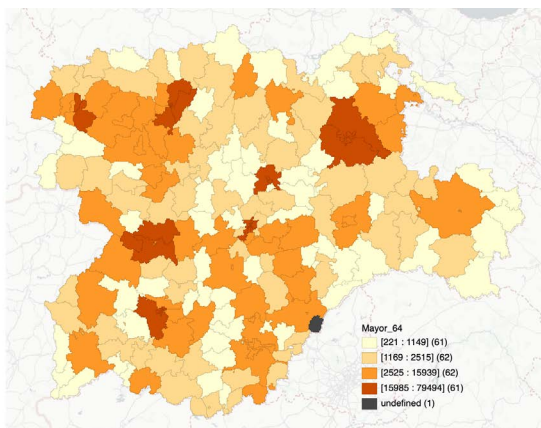


Figura A.22: Distribución espacial e histograma: Población mayor de 64 años

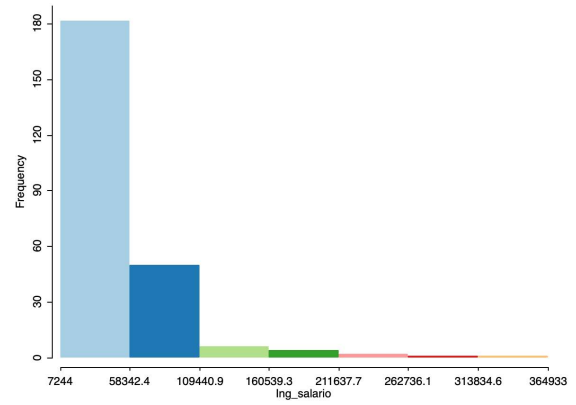
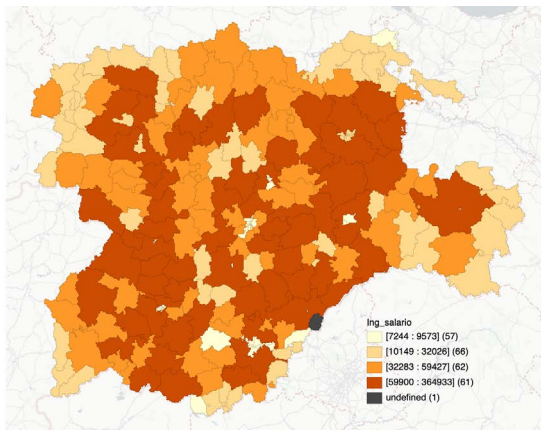


Figura A.23: Distribución espacial e histograma: Ingresos por salario

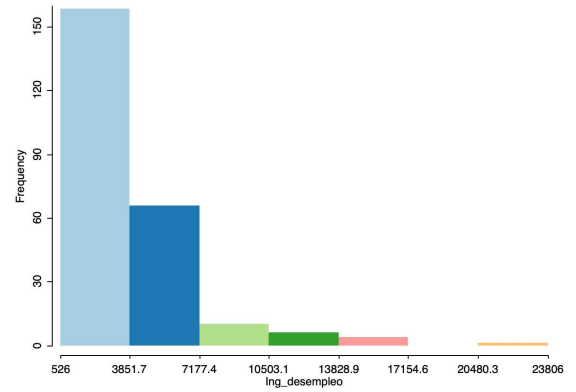
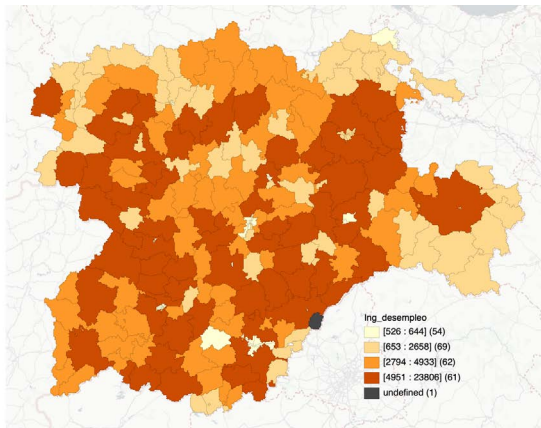


Figura A.24: Distribución espacial e histograma: Ingresos por prestaciones de desempleo

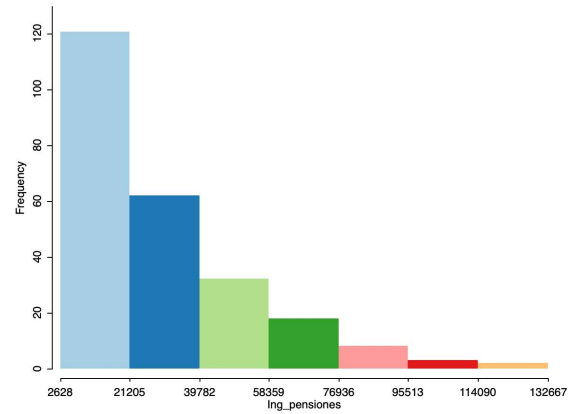
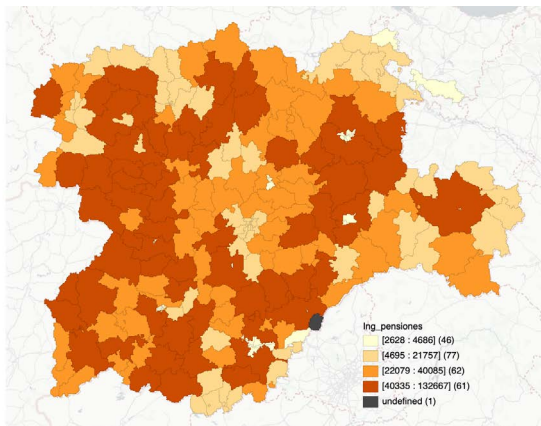


Figura A.25: Distribución espacial e histograma: Ingresos por pensiones

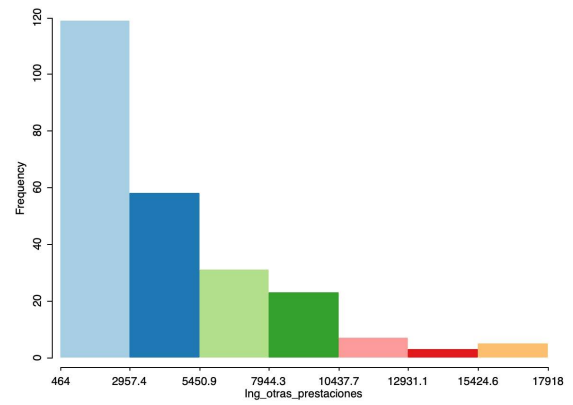
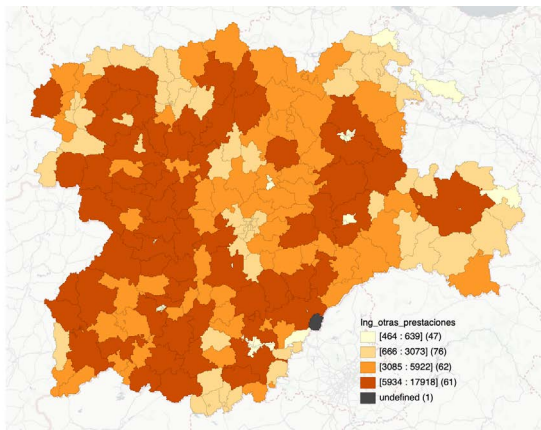


Figura A.26: Distribución espacial e histograma: Ingresos por otras prestaciones

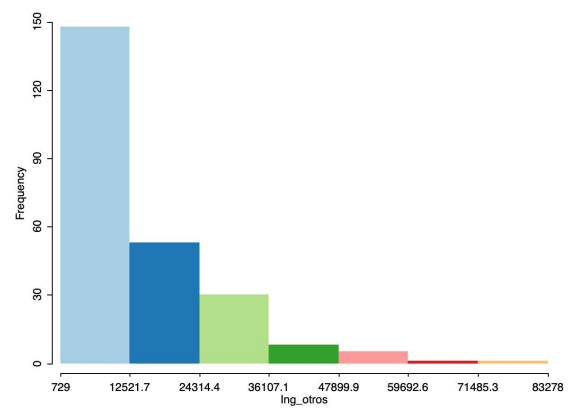
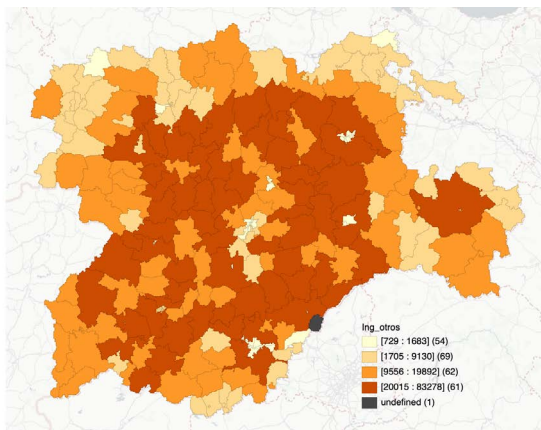


Figura A.27: Distribución espacial e histograma: Otros ingresos

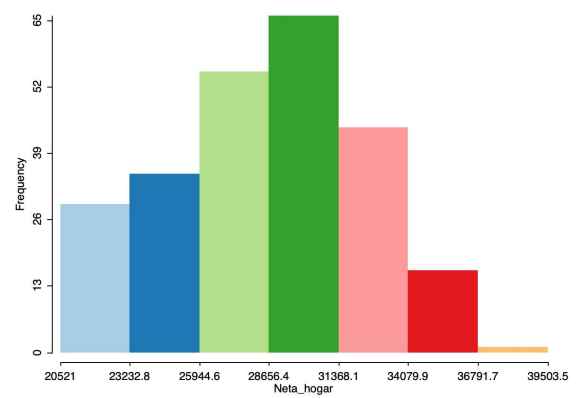
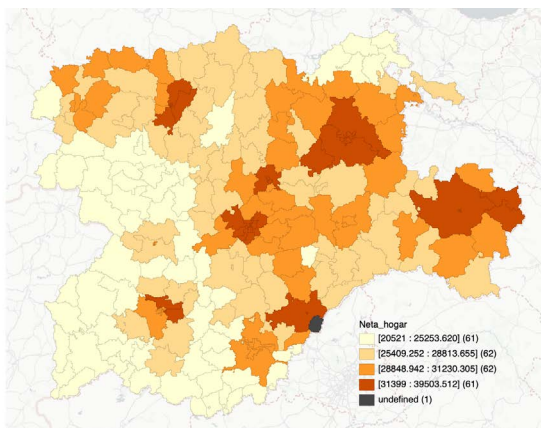


Figura A.28: Distribución espacial e histograma: Renta neta por hogar

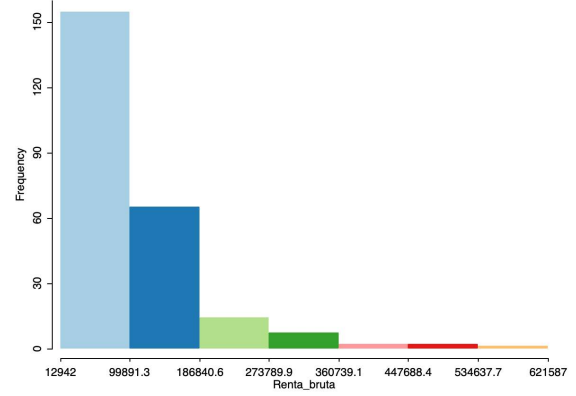
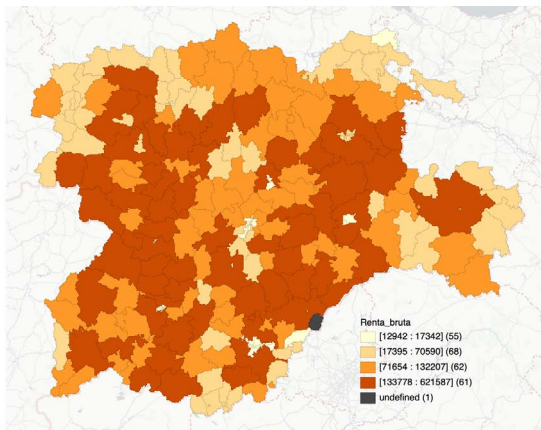


Figura A.29: Distribución espacial e histograma: Renta bruta

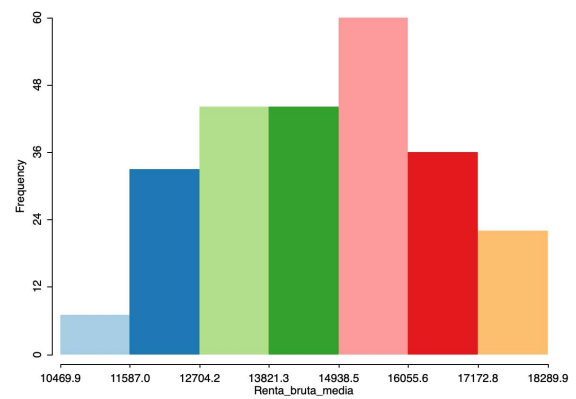
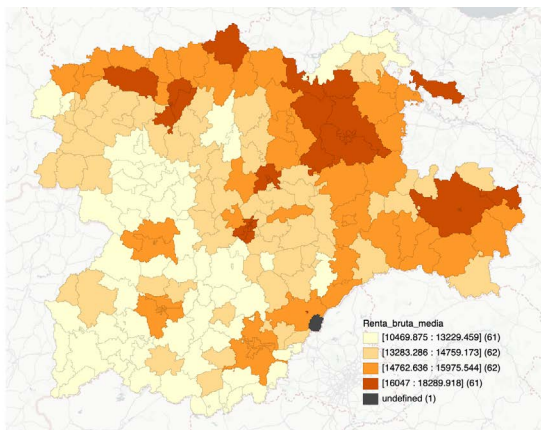


Figura A.30: Distribución espacial e histograma: Renta bruta media

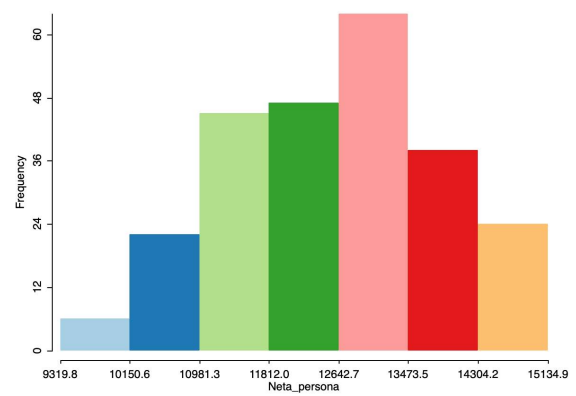
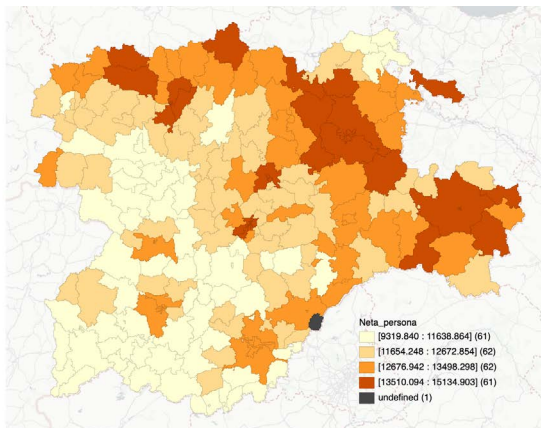


Figura A.31: Distribución espacial e histograma: Renta neta por persona

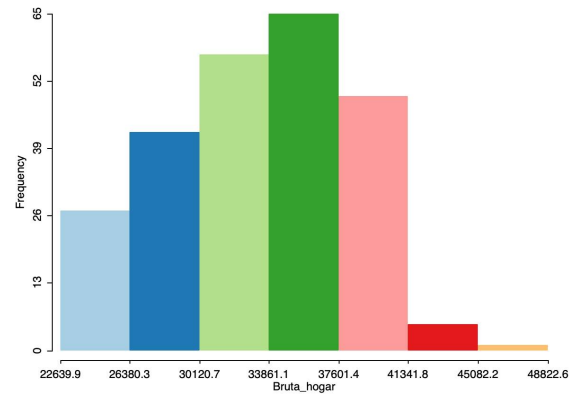
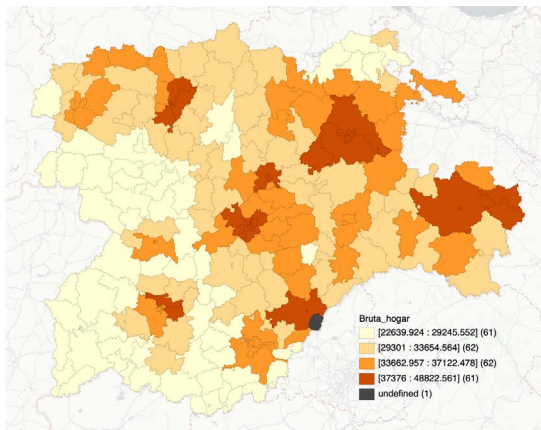


Figura A.32: Distribución espacial e histograma: Renta bruta por hogar

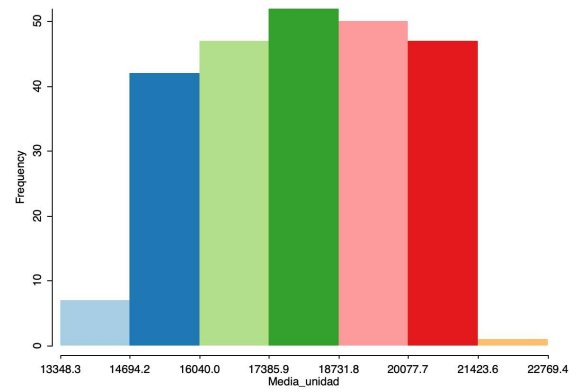
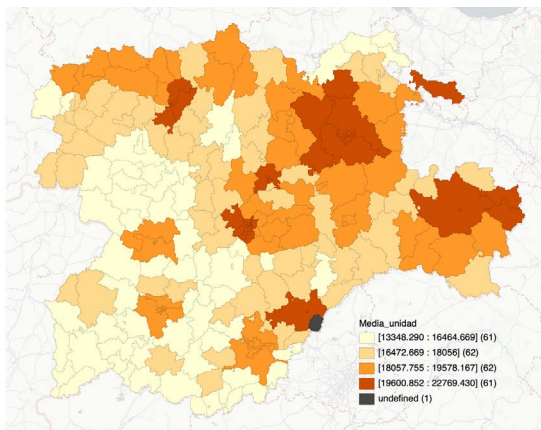


Figura A.33: Distribución espacial e histograma: Renta media por unidad familiar