**CÓDIGO R DE LA PRÁCTICA**

Descarga y lectura del fichero EPA

Leemos el fichero con microdatos de la EPA del trimestre de 2017

**Paquetes utilizados**

readr, plyr,dplyr,highr,rmarkdown,reshape2,stringi,formatR

**Ancho de los campos**

widths\_epa<-c(3,2,2,5,1,2,2,1,1,2,2,2,1,1,2,3,1,3,2,

2,2,3,1,2,1,2,3,2,1,1,1,2,2,1,1,1,2,1,1,1,2,2,2,3,

3,2,3,1,2,4,4,4,1,4,4,2,1,1,1,2,4,1,1,2,2,1,1,1,1,

2,1,1,2,1,1,1,3,1,1,2,1,2,2,2,1,1,1,2,3,1,2,2,7)

**Nombres de los campos**

colnames\_epa<-c('CICLO','CCAA','PROV','NVIVI',

'NIVEL','NPERS','EDAD5', 'RELPP1','SEXO1','NCONY',

'NPADRE', 'NMADRE','RELLMILI','ECIV1','PRONA1',

'REGNA1','NAC1', 'EXREGNA1', 'ANORE1','NFORMA',

'RELLB', 'EDADEST','CURSR','NCURSR', 'CURSNR',

'NCURNR','HCURNR','RELLB','TRAREM', 'AYUDFA',

'AUSENT','RZNOTB','VINCUL','NUEVEM','OCUP1',

'ACT1','SITU','SP', 'DUCON1', 'DUCON2','DUCON3',

'TCONTM','TCONTD','DREN', 'DCOM','PROEST',

'REGEST','PARCO1','PARCO2','HORASP','HORASH',

'HORASE','EXTRA', 'EXTPAG','EXTNPG','RZDIFH',

'TRAPLU', 'OCUPLU1', 'ACTPLU1','SITPLU',

'HORPLU','MASHOR','DISMAS','RZNDISH','HORDES',

'BUSOTR','BUSCA','DESEA','FOBACT','NBUSCA',

'ASALA','EMBUS','ITBU','DISP','RZNDIS','EMPANT',

'DTANT','OCUPA','ACTA','SITUA','OFEMP','SIDI1',

'SIDI2','SIDI3','SIDAC1','SIDAC2','MUN1','PRORE1', 'REPAIRE1','TRAANT','AOI','CSE','FACTOREL')

**Tipo de columnas**

coltypes\_epa<-list(CICLO=col\_character(),

CCAA=col\_character(),PROV=col\_character(),

NVIVI=col\_character(),NIVEL=col\_character(),

NPERS=col\_character(),EDAD5=col\_number(),

RELPP1=col\_character(),SEXO1=col\_character(),

NCONY=col\_character(),NPADRE=col\_character(),

NMADRE=col\_character(),RELLMILI=col\_character(),

ECIV1=col\_character(),PRONA1=col\_character(),

REGNA1=col\_character(),NAC1=col\_character(),

EXREGNA1=col\_character(),ANORE1=col\_number(),

NFORMA=col\_character(),RELLB=col\_character(),

EDADEST=col\_number(),CURSR=col\_character(),

NCURSR=col\_character(),CURSNR=col\_character(),

NCURNR=col\_character(),HCURNR=col\_number(),

RELLB=col\_character(),TRAREM=col\_character(),

AYUDFA=col\_character(),AUSENT=col\_character(),

RZNOTB=col\_character(),VINCUL=col\_character(),

NUEVEM=col\_character(),OCUP1=col\_character(),

ACT1=col\_character(),SITU=col\_character(),

SP=col\_character(),DUCON1=col\_character(),

DUCON2=col\_character(),DUCON3=col\_character(),

TCONTM=col\_number(),TCONTD=col\_number(),

DREN=col\_number(),DCOM=col\_number(),

PROEST=col\_character(),REGEST=col\_character(),

PARCO1=col\_character(),PARCO2=col\_character(),

HORASP=col\_number(),HORASH=col\_number(),

HORASE=col\_number(),EXTRA=col\_character(),

EXTPAG=col\_character(),EXTNPG=col\_character(),

RZDIFH=col\_character(),TRAPLU=col\_character(),

OCUPLU1=col\_character(),ACTPLU1=col\_character(),

SITPLU=col\_character(),HORPLU=col\_number(),

MASHOR=col\_character(),DISMAS=col\_character(),

RZNDISH=col\_character(),HORDES=col\_number(),

BUSOTR=col\_character(),BUSCA=col\_character(),

DESEA=col\_character(),FOBACT=col\_character(),

NBUSCA=col\_character(),ASALA=col\_character(),

EMBUS=col\_character(),ITBU=col\_character(),

DISP=col\_character(),RZNDIS=col\_character(),

EMPANT=col\_character(),DTANT=col\_number(),

OCUPA=col\_character(),ACTA=col\_character(),

SITUA=col\_character(),OFEMP=col\_character(),

SIDI1=col\_character(),SIDI2=col\_character(),

SIDI3=col\_character(),SIDAC1=col\_character(),

SIDAC2=col\_character(),MUN1=col\_character(),

PRORE1=col\_character(),REPAIRE1=col\_character(),

TRAANT=col\_character(),AOI=col\_character(),

CSE=col\_character(),FACTOREL=col\_number())

**Lectura del fichero**

epa176<-read\_fwf("EPAWEBT0416",

fwf\_widths(widths\_epa,colnames\_epa),

col\_types =cols(CICLO=col\_character(),

CCAA=col\_character(),PROV=col\_character(),

NVIVI=col\_character(),NIVEL=col\_character(),

NPERS=col\_character(),EDAD5=col\_number(),

RELPP1=col\_character(),SEXO1=col\_character(),

NCONY=col\_character(),NPADRE=col\_character(),

NMADRE=col\_character(),RELLMILI=col\_character(),

ECIV1=col\_character(),PRONA1=col\_character(),

REGNA1=col\_character(),NAC1=col\_character(),

EXREGNA1=col\_character(),ANORE1=col\_number(),

NFORMA=col\_character(),RELLB=col\_character(),

EDADEST=col\_number(),CURSR=col\_character(),

NCURSR=col\_character(),CURSNR=col\_character(),

NCURNR=col\_character(),HCURNR=col\_number(),

RELLB=col\_character(),TRAREM=col\_character(),

AYUDFA=col\_character(),AUSENT=col\_character(),

RZNOTB=col\_character(),VINCUL=col\_character(),

NUEVEM=col\_character(),OCUP1=col\_character(),

ACT1=col\_character(),SITU=col\_character(),

SP=col\_character(),DUCON1=col\_character(),

DUCON2=col\_character(),DUCON3=col\_character(),

TCONTM=col\_number(),TCONTD=col\_number(),

DREN=col\_number(),DCOM=col\_number(),

PROEST=col\_character(),REGEST=col\_character(),

PARCO1=col\_character(),PARCO2=col\_character(),

HORASP=col\_number(),HORASH=col\_number(),

HORASE=col\_number(),EXTRA=col\_character(),

EXTPAG=col\_character(),EXTNPG=col\_character(),

RZDIFH=col\_character(),TRAPLU=col\_character(),

OCUPLU1=col\_character(),ACTPLU1=col\_character(),

SITPLU=col\_character(),HORPLU=col\_number(),

MASHOR=col\_character(),DISMAS=col\_character(),

RZNDISH=col\_character(),HORDES=col\_number(),

BUSOTR=col\_character(),BUSCA=col\_character(),

DESEA=col\_character(),FOBACT=col\_character(),

NBUSCA=col\_character(),ASALA=col\_character(),

EMBUS=col\_character(),ITBU=col\_character(),

DISP=col\_character(),RZNDIS=col\_character(),

EMPANT=col\_character(),DTANT=col\_number(),

OCUPA=col\_character(),ACTA=col\_character(),

SITUA=col\_character(),OFEMP=col\_character(),

SIDI1=col\_character(),SIDI2=col\_character(),

SIDI3=col\_character(),SIDAC1=col\_character(),

SIDAC2=col\_character(),MUN1=col\_character(),

PRORE1=col\_character(),REPAIRE1=col\_character(),

TRAANT=col\_character(),AOI=col\_character(),

CSE=col\_character(),FACTOREL=col\_number()))

**Dividimos por 100 la variable FACTOREL para reducirla a unidades y calculamos el total** **poblacional.**

epa176$FACTOREL<-epa176$FACTOREL/100

Poblacion<-sum(epa176$FACTOREL)

summary(epa176$FACTOREL)

Poblacion

**Filtramos el fichero por las variables que nos interesan les damos nuevos nombres: CCAA, EDAD5, SEXO1, NAC1, NFORMA, AOI, FACTOREL.**

library(plyr); library(dplyr)

epa\_sel<-select(epa176,CCAA,EDAD5,SEXO1,NAC1,NFORMA,AOI,FACTOREL)

glimpse(epa\_sel)

**Creamos una nueva variable "ACTIVIDAD" que identifica la situación respecto del empleo**

epa\_sel<-epa\_sel%>%mutate(ACTIVIDAD=cut(as.numeric(AOI),breaks=c(0,4,6,9),labels=c("Ocupados","Parados","Inactivos")))

**Creamos una nueva variable "ESTUDIOS" que agrupa e identifica los niveles de estudio**

epa\_sel$NFORMA <- mapvalues(epa\_sel$NFORMA, from = c('AN','P1','P2','S1','SG','SP','SU'), to = c('01','02','03','04','05','06','07'))

epa\_sel<-epa\_sel%>%mutate(ESTUDIOS=cut(as.numeric(NFORMA),breaks=c(0,1,3,6,7),labels=c("Analfabetos","E. Primarios","E. Secundarios","E. Universitarios")))

**Creamos una nueva variable "GRUPO\_EDAD" que reduce y explicita los grupos de edad**

epa\_sel<-epa\_sel%>%mutate(GRUPO\_EDAD=cut(as.numeric(EDAD5),breaks=c(-1,10,16,25,35,45,55,65),labels=c("Menores de 16 años","De 16 a 20 años","De 20 a 29 años","De 30 a 39 años","De 40 a 49 años","De 50 a 59 años","Mayores de 60 años")))

muestra\_edad<-as.data.frame(table(epa\_sel$GRUPO\_EDAD))

library(knitr)

kable(muestra\_edad, align=c('l','r'),caption = "Distribución de la muestra por edades",format.args = list(decimal.mark = ",", big.mark = "."))

**Eliminamos variables redundantes**

epa\_sel<-epa\_sel%>%select(-EDAD5,-NFORMA,-AOI)

glimpse(epa\_sel)

**Convertimos en factores las variables CCAA, SEXO1 y NAC1, y renombramos estas dos** **últimas**

colnames(epa\_sel)[2:3]<-c("SEXO", "NACIONALIDAD")

epa\_sel$CCAA<-as.factor(epa\_sel$CCAA)

epa\_sel$SEXO<-as.factor(epa\_sel$SEXO)

epa\_sel$NACIONALIDAD<-as.factor(epa\_sel$NACIONALIDAD)

glimpse(epa\_sel)

**Agrupamos las observaciones por todas las variables de nuestro interés (CCAA, SEXO,** **ACTIVIDAD, ESTUDIOS,GRUPO\_EDAD)y obtenemos los totales poblacionales para cada combinación posible**

tablas\_EPA<-epa\_sel%>%group\_by(CCAA,SEXO,NACIONALIDAD,ACTIVIDAD,ESTUDIOS,GRUPO\_EDAD)%>%summarise(Total=sum(FACTOREL))

#Reordenamos las columnas

tablas\_EPA<-select(tablas\_EPA,CCAA,SEXO,GRUPO\_EDAD,ESTUDIOS,NACIONALIDAD,ACTIVIDAD,TOTAL=Total)

glimpse(tablas\_EPA)

**Sustituimos los valores de CCAA, SEXO y NACIONALIDAD por sus literales**

tablas\_EPA$CCAA <- mapvalues(tablas\_EPA$CCAA, from = c('01','02','03','04','05','06','07','08','09','10', '11','12','13','14','15','16','17','51','52'), to = c('01 Andalucía', '02 Aragón', '03 Asturias', '04 Baleares', '05 Canarias', '06 Cantabria', '07 Castilla-León', '08 Castilla-La Mancha', '09 Cataluña', '10 Comunidad Valenciana', '11 Extremadura', '12 Galicia', '13 Madrid', '14 Murcia', '15 Navarra', '16 País Vasco', '17 Rioja, La','51 Ceuta','52 Melilla '))

tablas\_EPA$SEXO <- mapvalues(tablas\_EPA$SEXO, from = c('1','6'), to = c('Varones', 'Mujeres'))

tablas\_EPA$NACIONALIDAD <-mapvalues(tablas\_EPA$NACIONALIDAD, from = c('1','2','3'), to = c('Española','Española y doble nacionalidad','Extranjera'))

kable(head(tablas\_EPA), caption = "Distribución de la población por todas las variables de interés", format.args = list(decimal.mark = ",", big.mark = "."))

**Ordenamos las filas de datos por los valores de las sucesivas columnas**

tablas\_EPA<-arrange(tablas\_EPA,CCAA,SEXO,GRUPO\_EDAD,ESTUDIOS,NACIONALIDAD,ACTIVIDAD)

kable(head(tablas\_EPA), caption = "Distribución de la población por todas las variables de interés", format.args = list(decimal.mark = ",", big.mark = "."))

**Eliminamos los menores de 16 años**

tablas\_EPA<-tablas\_EPA%>%filter(GRUPO\_EDAD!="Menores de 16 años")

kable(head(tablas\_EPA), caption = "Distribución de los mayores de 16 años por todas las variables de interés", format.args = list(decimal.mark = ",", big.mark = "."))

library(reshape2)

Valores<-dcast(tablas\_EPA,CCAA+SEXO+GRUPO\_EDAD+ESTUDIOS+NACIONALIDAD ~ ACTIVIDAD,value.var = "TOTAL")

**Calculamos valores absolutos y relativos de empleo , paro e inactividad para el Conjunto** **Nacional**

TotalNac\_EPA<-Valores%>%summarise(

Total\_Ocupados=round(sum(Ocupados,na.rm = TRUE),1),

Total\_Parados=round(sum(Parados,na.rm = TRUE),1),

Total\_Activos=Total\_Ocupados+Total\_Parados,

Tasa\_desempleo=round(100\*(Total\_Parados/Total\_Activos),2),

Total\_Inactivos=sum(Inactivos,na.rm = TRUE),

Tasa\_actividad=round(100\*Total\_Activos/(Total\_Activos+Total\_Inactivos),2))

kable(TotalNac\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad. Total Nacional", format.args = list(decimal.mark = ",", big.mark = "."))

**Calculamos los valores absolutos y relativos de empleo, paro e inactividad por CCAA**

TablaCCAA\_EPA<-Valores%>%group\_by(CCAA)%>%summarise(

Total\_Ocupados=round(sum(Ocupados,na.rm = TRUE),1),

Total\_Parados=round(sum(Parados,na.rm = TRUE),1),

Total\_Activos=Total\_Ocupados+Total\_Parados,

Tasa\_desempleo=round(100\*(Total\_Parados/Total\_Activos),2),

Total\_Inactivos=sum(Inactivos,na.rm = TRUE),

Tasa\_actividad=round(100\*Total\_Activos/(Total\_Activos+Total\_Inactivos),2))

kable(TablaCCAA\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad. Comunidades Autónomas", format.args = list(decimal.mark = ",", big.mark = "."))

**Unimos ambas tablas**

TablaNACIONAL\_EPA<-TotalNac\_EPA%>%mutate(CCAA="Conjunto Nacional")

glimpse(TablaCCAA\_EPA)

TablaNACIONAL\_EPA<-TablaNACIONAL\_EPA%>%select(CCAA,Total\_Ocupados,Total\_Parados,Total\_Activos,Tasa\_desempleo,Total\_Inactivos,Tasa\_actividad)

TablaCCAA\_EPA<-rbind(TablaNACIONAL\_EPA,TablaCCAA\_EPA)

kable(TablaCCAA\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad por Comunidades Autónomas y Conjunto Nacional", format.args = list(decimal.mark = ",", big.mark = "."))

**Calculamos los valores absolutos y relativos de empleo , paro e inactividad por SEXO**

TablaSEXO\_EPA<-Valores%>%group\_by(SEXO)%>%summarise(

Total\_Ocupados=round(sum(Ocupados,na.rm = TRUE),1),

Total\_Parados=round(sum(Parados,na.rm = TRUE),1),

Total\_Activos=Total\_Ocupados+Total\_Parados,

Tasa\_desempleo=round(100\*(Total\_Parados/Total\_Activos),2),

Total\_Inactivos=sum(Inactivos,na.rm = TRUE),

Tasa\_actividad=round(100\*Total\_Activos/(Total\_Activos+Total\_Inactivos),2))

kable(TablaSEXO\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad. Distribución por sexos", format.args = list(decimal.mark = ",", big.mark = "."))

**Calculamos valores absolutos y relativos de empleo , paro e inactividad por GRUPOS DE** **EDAD**

TablaGRUPO\_EDAD\_EPA<-Valores%>%group\_by(GRUPO\_EDAD)%>%summarise(

Total\_Ocupados=round(sum(Ocupados,na.rm = TRUE),1),

Total\_Parados=round(sum(Parados,na.rm = TRUE),1),

Total\_Activos=Total\_Ocupados+Total\_Parados,

Tasa\_desempleo=round(100\*(Total\_Parados/Total\_Activos),2),

Total\_Inactivos=sum(Inactivos,na.rm = TRUE),

Tasa\_actividad=round(100\*Total\_Activos/(Total\_Activos+Total\_Inactivos),2))

kable(TablaGRUPO\_EDAD\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad. Distribución por GRUPOS DE EDAD", format.args = list(decimal.mark = ",", big.mark = "."))

**Calculamos valores absolutos y relativos de empleo , paro e inactividad por** **NACIONALIDAD**

TablaNACIONALIDAD\_EPA<-Valores%>%group\_by(NACIONALIDAD)%>%summarise(

Total\_Ocupados=round(sum(Ocupados,na.rm = TRUE),1),

Total\_Parados=round(sum(Parados,na.rm = TRUE),1),

Total\_Activos=Total\_Ocupados+Total\_Parados,

Tasa\_desempleo=round(100\*(Total\_Parados/Total\_Activos),2),

Total\_Inactivos=sum(Inactivos,na.rm = TRUE),

Tasa\_actividad=round(100\*Total\_Activos/(Total\_Activos+Total\_Inactivos),2))

kable(TablaNACIONALIDAD\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad. Distribución por NACIONALIDAD", format.args = list(decimal.mark = ",", big.mark = "."))

**Calculamos valores absolutos y relativos de empleo , paro e inactividad por ESTUDIOS**

TablaESTUDIOS\_EPA<-Valores%>%group\_by(ESTUDIOS)%>%summarise(

Total\_Ocupados=round(sum(Ocupados,na.rm = TRUE),1),

Total\_Parados=round(sum(Parados,na.rm = TRUE),1),

Total\_Activos=Total\_Ocupados+Total\_Parados,

Tasa\_desempleo=round(100\*(Total\_Parados/Total\_Activos),2),

Total\_Inactivos=sum(Inactivos,na.rm = TRUE),

Tasa\_actividad=round(100\*Total\_Activos/(Total\_Activos+Total\_Inactivos),2))

kable(TablaESTUDIOS\_EPA, caption = "Población mayor de 16 años según su relación con la Actividad. Distribución por NIVEL DE ESTUDIOS", format.args = list(decimal.mark = ",", big.mark = "."))