Causas Muerte Republica Domincana, 2011-2012

Pedro Camacho

26 November 2016

# 1.Leyendo la data

La data se encuentra en un repositorio de Github de forma pública, de forma que cualquier persona pueda acceder al mismo. Para incrementar la reproducibilidad de este analisis vamos a descargar los documentos directamente del repositorio de Githup.

url2011 <- "https://github.com/Pedromoisescamacho/Principales-causas-de-muerte-Republica-Domincana/raw/master/Mortalidad%20Por%20Provincia%202011.csv"  
url2012 <- "https://raw.githubusercontent.com/Pedromoisescamacho/Principales-causas-de-muerte-Republica-Domincana/master/Mortalidad%20Por%20Provincia%202012.csv"

# 2. Modificando los datos

la tabla entregada por el Ministerio de Salud Pública se encuentra en formta de excel, para poder trabajar en R la convertimos a CSV desde el mismo excel. Esto se convirtio en dos documentos uno para el 2011 y otro para el 2012. Los documentos presentan la cantidad de muertes por provincia por causa de muerte; donde cada linea es una causa de muerte y cada columna es una provincia de RD. También el documento incluye una linea y una columna sumando los totales. el siguiente codigo lee los dos documentos:

muerte1 <- read.csv(url2011, header = T, nrows = 69) #luego de la linea 70 la tabla contiene informaciones generales que no vienen al caso  
muerte2 <- read.csv(url2012, header = T, nrows = 69) #luego de la linea 70 la tabla contiene informaciones generales que no vienen al caso  
muerte1$periodo <- as.factor("2011")  
muerte2$periodo <- as.factor("2012")  
m <- rbind(muerte1, muerte2)

Este formato no es el apropiado como base de datos dado que limita los analisis, por lo que en primera instancia vamos a modificar la base de datos para que cada linea sea una causa de (observacion) y tengamos una columna con la provincia en la que se dio. Tambien tenemos que sacar las lineas y columnas que tienen totales.

linea <- grep("total", m$Causas,ignore.case = T) #buscamos las lineas que tiene "total"  
columna <- grep("total", names(m), ignore.case = T) #buscamos las columnas con nombre "total"  
m2<- m[c(-linea), -(columna)] #eliminamos de la base de datos las lineas y columnas con "total"  
m2$Causas <- as.factor(as.character(m2$Causas)) #eliminamos el factor total que todavia se encontraba

con el bloque de codigos anterior obtenemos una base de datos donde cada una de las columnas es una variable. Por otro lado debemos de hacer mas leible la columna de Causas. Esta columna inicia con un código de 3 numeros que identifica las causas. Por ejemplo en "1.02" el número antes del punto es el grupo de la causa y los números luego del punto la causa específica. Por lo tanto vamos a crear una variable con el grupo, otra con el numero especifico de la causa y otra con la causa en letras.

m2$grupo <- as.factor(as.character(substr(m2$Causas, 1, 1))) #extrae el primer numero  
m2$Causas <- as.factor(substr(m2$Causas, 6, 100)) #extrae solo el texto de la columna  
levels(m2$grupo) <- c("Mal definidos", "Infección/Parásito", "Tumores/Cáncer", "Sis. Circulatorio", "Perinatal", "Muertes Violentas", "Otras")  
library(reshape2) #cargamos el paquete que utilizaremos para agregar   
m3 <- melt(m2, id = c("Causas", "periodo", "grupo")) #creamos la variable provincia coloca  
names(m3) <- c("causas", "periodo","grupo", "provincia", "muertes")  
m3$provincia <- as.factor(gsub("\\.", " ", m3$provincia)) #quitando los puntos de los nombres de las provincias

Lo numeros en la recien creada variable "grupo" no indican rapidamente a que grupo nos referimos, por lo que ahora vamos a colocar nombres a los grupos.

La tabla final obtenida tiene unas dimensiones de 4488, 5 vs 138, 36 de la tabla inicial, siendo m3 mucho mas facil de trabajar.

# Análisis

Primero vamos a ver cuales grupos de causas tienen mayor participacion.

por\_grupos <-dcast(m3, grupo ~periodo, sum)

## Using muertes as value column: use value.var to override.

por\_grupos <- por\_grupos[order(por\_grupos$`2012`, decreasing = T), ]  
por\_grupos2 <- transform(por\_grupos, "2011" = round(`2011`/sum(`2011`)\*100,1), "2012" = round(`2012`/sum(`2012`)\*100,1))  
names(por\_grupos2) <- c("grupo", "2011", "2012")  
barplot(acast(m3, periodo ~ grupo, sum), legend.text = T, beside = T)

## Using muertes as value column: use value.var to override.

