EJPracticoentiendetutabladedatos.R

apalmad

2021-01-18

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
# 1.1 INSTALAR PAQUETES DE FUNCIONES  
#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
# Lista de paquetes de funciones a instalar  
.packages = c("ggplot2", "plotly", "xlsx","scales","readr","dplyr","psych","readxl")  
  
  
# Instala los paquetes sinó los tienes instalados  
.inst <- .packages %in% installed.packages()  
if(length(.packages[!.inst]) > 0) install.packages(.packages[!.inst])  
  
  
#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
# 1.2 CARGAR PAQUETES O CREAR FUNCIONES  
#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
# Carga los paquetes sinó los tienes cargados  
lapply(.packages, require, character.only=TRUE)

## Loading required package: ggplot2

## Loading required package: plotly

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

## Loading required package: xlsx

## Error: package or namespace load failed for 'xlsx':  
## .onLoad failed in loadNamespace() for 'rJava', details:  
## call: fun(libname, pkgname)  
## error: JAVA\_HOME cannot be determined from the Registry

## Loading required package: scales

## Loading required package: readr

##   
## Attaching package: 'readr'

## The following object is masked from 'package:scales':  
##   
## col\_factor

## Loading required package: 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

## Loading required package: psych

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:scales':  
##   
## alpha, rescale

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

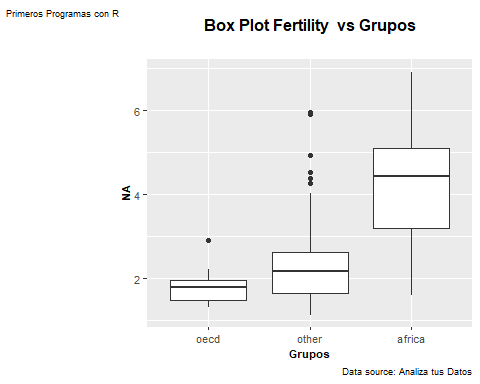
## Loading required package: readxl

## [[1]]  
## [1] TRUE  
##   
## [[2]]  
## [1] TRUE  
##   
## [[3]]  
## [1] FALSE  
##   
## [[4]]  
## [1] TRUE  
##   
## [[5]]  
## [1] TRUE  
##   
## [[6]]  
## [1] TRUE  
##   
## [[7]]  
## [1] TRUE  
##   
## [[8]]  
## [1] TRUE

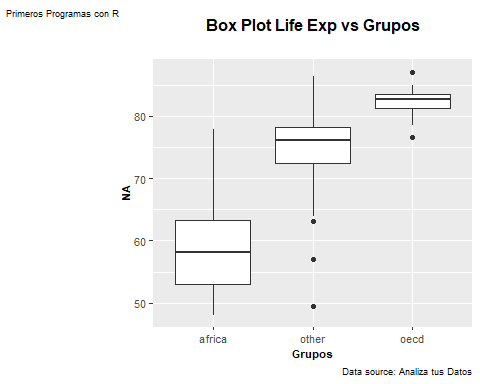
#####################################  
#  
#Setear el working directory  
#####################################  
setwd("C:/Users/apalmad/Desktop/Analiza tus Datos/Proyectos R")  
  
  
####################################  
# Leeo la tabla de datos  
####################################  
  
Datos <- read\_xlsx("Caso Practico/Entiende tu Tabla de Datos/Datos.xlsx")  
  
#################################  
# Describo las variables  
###############################  
str(Datos)

## tibble [193 x 8] (S3: tbl\_df/tbl/data.frame)  
## $ pais : chr [1:193] "Afghanistan" "Albania" "Algeria" "Angola" ...  
## $ region : chr [1:193] "Asia" "Europe" "Africa" "Africa" ...  
## $ group : chr [1:193] "other" "other" "africa" "africa" ...  
## $ fertility : num [1:193] 5.97 1.52 2.14 5.13 2.17 ...  
## $ ppgdp : num [1:193] 499 3677 4473 4322 9162 ...  
## $ lifeExpF : num [1:193] 49.5 80.4 75 53.2 79.9 ...  
## $ pctUrban : num [1:193] 23 53 67 59 93 64 47 89 68 52 ...  
## $ infantMortality: num [1:193] 124.5 16.6 21.5 96.2 12.3 ...

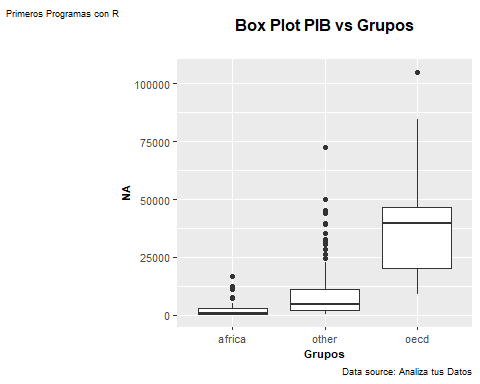
creaboxplot <- function(Datos, var\_x,var\_y,Titulo,xlabel,ylabel="NA"){  
 ggplot(data = Datos) +  
 geom\_boxplot(mapping = aes(  
 x = reorder(get(var\_x),get(var\_y),FUN = median),  
 y = get(var\_y)  
 )) +  
 labs(title = Titulo,  
 subtitle = "",  
 caption = "Data source: Analiza tus Datos",  
 x = xlabel,  
 y = ylabel,  
 tag = "Primeros Programas con R") +  
 theme(plot.title = element\_text(size =12, face = "bold", hjust = 0.5 ),  
 plot.subtitle = element\_text(size = 9, hjust = 0.5),  
 plot.caption = element\_text(size = 7),  
 plot.tag = element\_text(size = 7),  
 axis.text = element\_text(size=8),  
 axis.title = element\_text(size=8,face="bold"))  
   
   
}  
  
  
creascatplot <- function(Datos, var\_x,var\_y,Titulo,xlabel,ylabel="NA",categoria="group"){  
 ggplot(data = Datos) +  
 geom\_point(mapping = aes(x = get(var\_x), y = get(var\_y),color=get(categoria))) +  
   
 labs(title = Titulo,  
 subtitle = "",  
 caption = "Data source: Analiza tus Datos",  
 x = xlabel,  
 y = ylabel,  
 tag = "Primeros Programas con R") +  
 theme(plot.title = element\_text(size =12, face = "bold", hjust = 0.5 ),  
 plot.subtitle = element\_text(size = 9, hjust = 0.5),  
 plot.caption = element\_text(size = 7),  
 plot.tag = element\_text(size = 7),  
 axis.text = element\_text(size=8),  
 axis.title = element\_text(size=8,face="bold")) +  
 geom\_smooth(mapping = aes(x = get(var\_x), y = get(var\_y)), orientation = "y")+  
 facet\_grid(. ~ get(categoria))   
   
   
}  
  
#############################  
#  
# Primeros Analisis  
#  
#############################  
##  
##Box Plot para comparar grupos (comparacion de medias entre grupos)  
##  
creaboxplot(Datos,"group","fertility","Box Plot Fertility vs Grupos","Grupos")



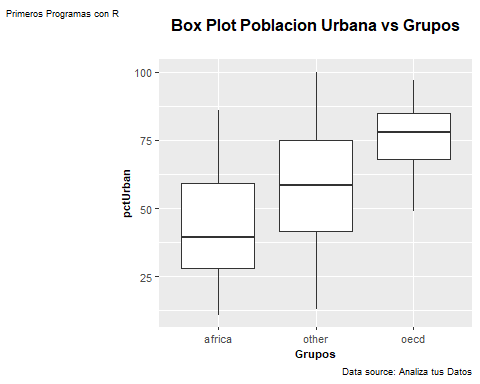
creaboxplot(Datos,"group","lifeExpF","Box Plot Life Exp vs Grupos","Grupos")



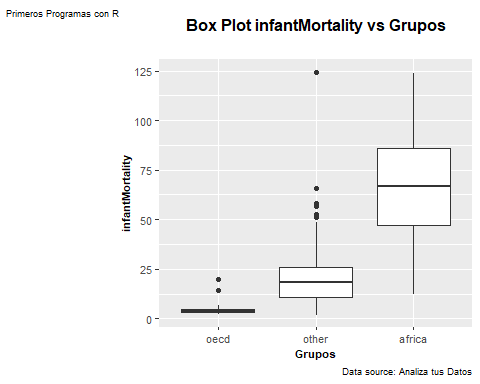
creaboxplot(Datos,"group","ppgdp","Box Plot PIB vs Grupos","Grupos")



creaboxplot(Datos,"group","pctUrban","Box Plot Poblacion Urbana vs Grupos","Grupos","pctUrban")

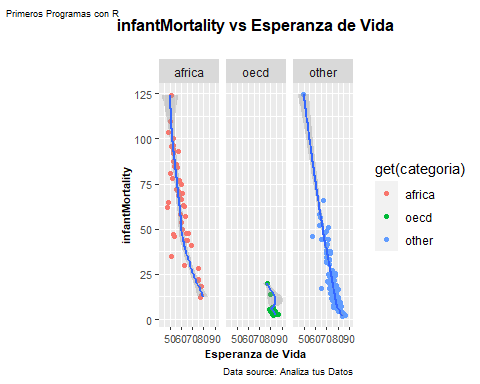


creaboxplot(Datos,"group","infantMortality","Box Plot infantMortality vs Grupos","Grupos","infantMortality")



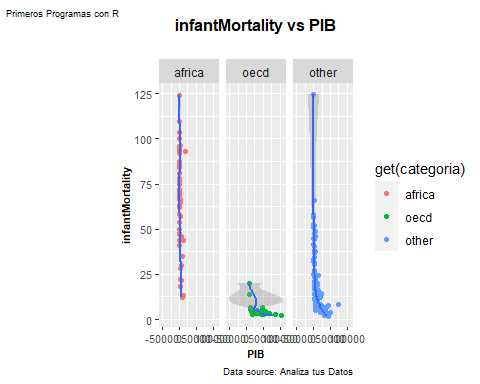
##  
## Scatterplot para buscar asociaciones entre variables  
##  
creascatplot(Datos,"lifeExpF","infantMortality","infantMortality vs Esperanza de Vida","Esperanza de Vida","infantMortality")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



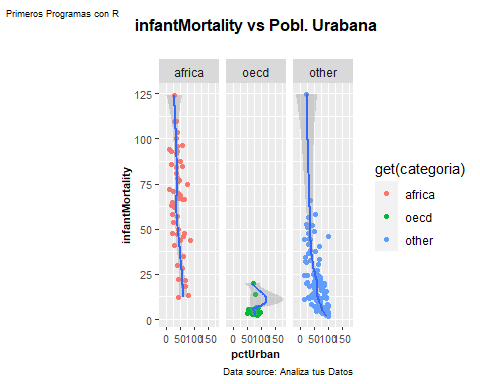
creascatplot(Datos,"ppgdp","infantMortality","infantMortality vs PIB","PIB","infantMortality")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



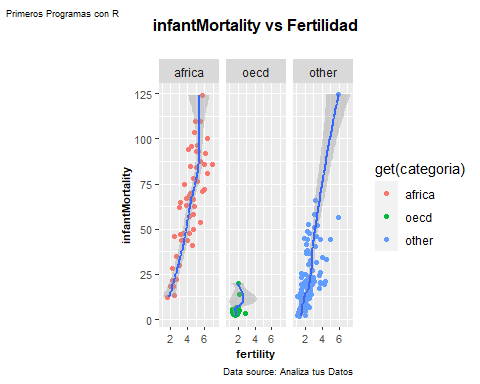
creascatplot(Datos,"pctUrban","infantMortality","infantMortality vs Pobl. Urabana","pctUrban","infantMortality")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



creascatplot(Datos,"fertility","infantMortality","infantMortality vs Fertilidad","fertility","infantMortality")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



####  
# Coef. de Correlacion InfanMortality vs variables  
###  
print(paste("r de InfantMortality vs Fertilidad ", cor(Datos$infantMortality,Datos$fertility)))

## [1] "r de InfantMortality vs Fertilidad 0.857909640351146"

print(paste("r de InfantMortality vs Poblacion Urbana ", cor(Datos$infantMortality,Datos$pctUrban)))

## [1] "r de InfantMortality vs Poblacion Urbana -0.599161713229932"

print(paste("r de InfantMortality vs PIB ", cor(Datos$infantMortality,Datos$ppgdp)))

## [1] "r de InfantMortality vs PIB -0.515384695319901"

print(paste("r de InfantMortality vs Esperzana de Vida ", cor(Datos$infantMortality,Datos$lifeExpF)))

## [1] "r de InfantMortality vs Esperzana de Vida -0.933168259138956"