Data Visualization Distributions

Md Easin Hasan

3/27/2021

data<-read.csv(file="owid-covid-data.csv",header=TRUE)  
dim(data)

## [1] 73384 59

#colnames(data)  
data1 <- data[data$iso\_code == "USA", ]   
data2 <- data[data$iso\_code == "GBR", ]   
data3<-rbind(data1,data2)  
#dim(data1)  
#dim(data2)  
country <- ifelse(data3$iso\_code == "USA",0,1)  
data4 <- cbind(country,data3)  
data4$country<-as.factor(data4$country)  
data4$country<-factor(data4$country,levels=c(0,1),labels=c("USA", "UK"))  
data5 <- data4[,-c(2)]  
dat2 <-data5[,c(1,6,8,11,18,20)]  
colnames(dat2)

## [1] "country" "new\_cases"   
## [3] "total\_deaths" "total\_cases\_per\_million"  
## [5] "icu\_patients" "hosp\_patients"

miss.info<- function(dat, filename=NULL){   
   
 vnames <- colnames(dat); vnames   
   
 n <- nrow(dat)   
   
 out <- NULL   
   
 for (j in 1: ncol(dat)){   
   
 vname <- colnames(dat)[j]   
   
 x <- as.vector(dat[,j])   
   
 n1 <- sum(is.na(x), na.rm=T)   
   
 n2 <- sum(x=="NA", na.rm=T)   
   
 n3 <- sum(x=="", na.rm=T)   
   
 nmiss <- n1 + n2 + n3   
   
 ncomplete <- n-nmiss   
   
 out <- rbind(out, c(col.number=j, vname=vname, mode=mode(x), n.levels=length(unique(x)), ncomplete=ncomplete, miss.perc=nmiss/n)) }   
   
 out <- as.data.frame(out)   
   
 row.names(out) <- NULL   
   
 return(out)   
   
}   
  
miss.info(dat2)

## col.number vname mode n.levels ncomplete  
## 1 1 country character 2 813  
## 2 2 new\_cases numeric 741 812  
## 3 3 total\_deaths numeric 734 740  
## 4 4 total\_cases\_per\_million numeric 776 813  
## 5 5 icu\_patients numeric 520 563  
## 6 6 hosp\_patients numeric 559 568  
## miss.perc  
## 1 0  
## 2 0.001230012300123  
## 3 0.0897908979089791  
## 4 0  
## 5 0.30750307503075  
## 6 0.301353013530135

library(mice)

## Warning: package 'mice' was built under R version 4.0.4

##   
## Attaching package: 'mice'

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

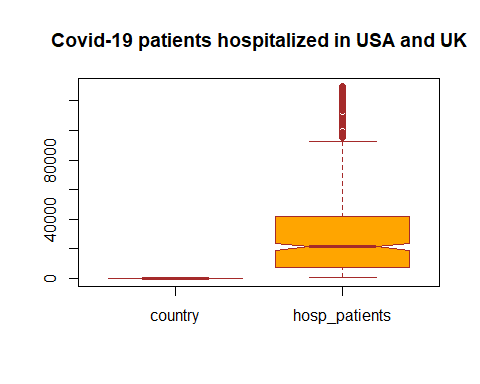
## The following objects are masked from 'package:base':  
##   
## cbind, rbind

fit.mice.dat\_f <- mice(dat2, m = 1, maxit = 50, method = 'pmm', seed = 500)

##   
## iter imp variable  
## 1 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 2 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 3 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 4 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 5 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 6 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 7 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 8 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 9 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 10 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 11 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 12 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 13 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 14 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 15 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 16 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 17 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 18 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 19 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 20 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 21 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 22 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 23 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 24 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 25 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 26 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 27 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 28 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 29 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 30 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 31 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 32 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 33 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 34 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 35 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 36 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 37 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 38 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 39 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 40 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 41 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 42 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 43 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 44 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 45 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 46 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 47 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 48 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 49 1 new\_cases total\_deaths icu\_patients hosp\_patients  
## 50 1 new\_cases total\_deaths icu\_patients hosp\_patients

dat\_f.imputed <- complete(fit.mice.dat\_f,1)  
datf <- as.data.frame(dat\_f.imputed)

dat3 <- dat2[,-c(2,3,4,5)]  
  
boxplot(dat3, main = "Covid-19 patients hospitalized in USA and UK",col="orange", border="brown",  
 notch = TRUE)



## Data visualization distribution improvement

library(ggplot2)  
ggplot(dat2, aes(x = country, y = hosp\_patients, fill = country)) +  
 geom\_boxplot(alpha=0.3) +  
 theme(legend.position="right")+  
 ggtitle("COVID-19 patients hospitalized in USA & UK")

