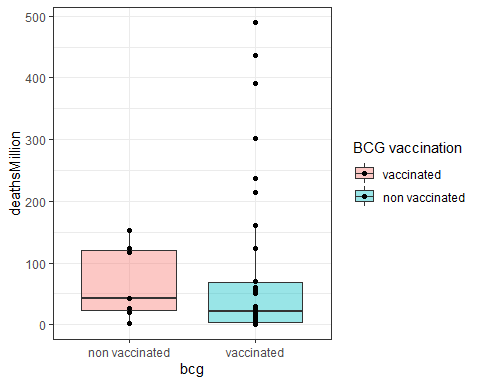
BCG\_covid19 correlation

Oussama Kadded

21/04/2020

The original data was devided into countries with current BCG vaccination programs and countries with no BCG programs. I think it makes no sense to devide the data in this manner since the age groups that are most vulnerable to the infection have already had BCG vaccination before a lot of countries stopped the programs and since most countries don’t give booster shots, there would be no difference in coverage for the elderly population. Therefore it makes more sense to devide the data into countries that never had a BCG program vs countries that still have or used to have a universal BCG vaccination program.

#loading the data   
bcg=read.csv2("C:/RepTemplates/BCG\_covid19/BCG vaccination 2(modified).csv",header = T)  
bcg$bcg=factor(bcg$bcg,levels=c(0,1),labels=c("non vaccinated","vaccinated"))  
ggplot(bcg,aes(bcg,deathsMillion,fill=bcg)) +   
 geom\_boxplot(alpha=0.4) +   
 geom\_point() +  
 scale\_fill\_discrete(name="BCG vaccination",labels=c("vaccinated","non vaccinated")) +  
 theme\_bw()



#run anova test  
anova=aov(deathsMillion~bcg, bcg)  
summary(anova)

## Df Sum Sq Mean Sq F value Pr(>F)  
## bcg 1 1445 1445 0.091 0.764  
## Residuals 39 619000 15872

H0 : the number of deaths/million in countries with a vaccinated high risk groups = the number of deaths/million in countries with non vaccinated high risk groups.

The anova test performed on the modified data resulted in a p value = 0.76 showing no significant difference in covid-19 deaths between countries with vaccinated high risk population and countries with no vaccination.

**The null hypothesis is retained.**