**VIT-AP UNIVERSITY, ANDHRA PRADESH**

**CSE2047 – Data Analytics - Lab Sheet : 6**

**Academic year:** 2020-2021  **Branch/ Class:** B.Tech/M.Tech

**Semester:** Fall  **Date:**

**Faculty Name:** Prof. S.Gopikrishnan  **School:** SCOPE

**Student name:** Valiveti Manikanta bhuvanesh **Reg. no.: 19BCD7088**

**LAB 6**

**Questions:**

1. **Create a student result dataset with numeric values.** 
   1. **write a function for calculating the mean.**
   2. **Write a function to compute std.deviation.**

df<-read.csv('Student\_Data\_cleaned.csv')

df

v<-colnames(df)

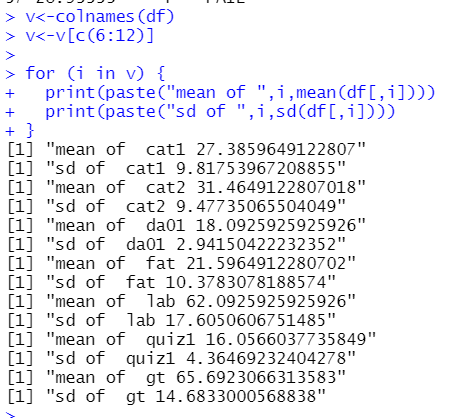
v<-v[c(6:12)]

for (i in v) {

print(paste("mean of ",i,mean(df[,i])))

print(paste("sd of ",i,sd(df[,i])))

}



1. **Use Covid.csv and weather.csv. Do all observations (min, max, mean, variance, SD, range) in both dataframe.**

df<-read.csv('COVID\_country\_wise\_latest.csv')

head(df)

df1<-read.csv('weatherHistory.csv')

head(df1)

v1<-colnames(df)

v2<-colnames(df1)

for(i in v1){

if(typeof(df[,i])=="integer"){

print(paste("mean of ",i,mean(df[,i],na.rm = TRUE)))

print(paste("sd of ",i,sd(df[,i],na.rm = TRUE)))

print(paste("min of ",i,min(df[,i],na.rm = TRUE)))

print(paste("max of ",i,max(df[,i],na.rm = TRUE)))

print(paste("range of ",i,range(df[,i],na.rm = TRUE)[1],range(df[,i],na.rm = TRUE)[2]))

print(paste("var of ",i,var(df[,i],na.rm = TRUE)))

}

}

for(i in v2){

if(typeof(df1[,i])=="double"|typeof(df1[,i])=="integer"){

print(paste("mean of ",i,mean(df1[,i],na.rm = TRUE)))

print(paste("sd of ",i,sd(df1[,i],na.rm = TRUE)))

print(paste("min of ",i,min(df1[,i],na.rm = TRUE)))

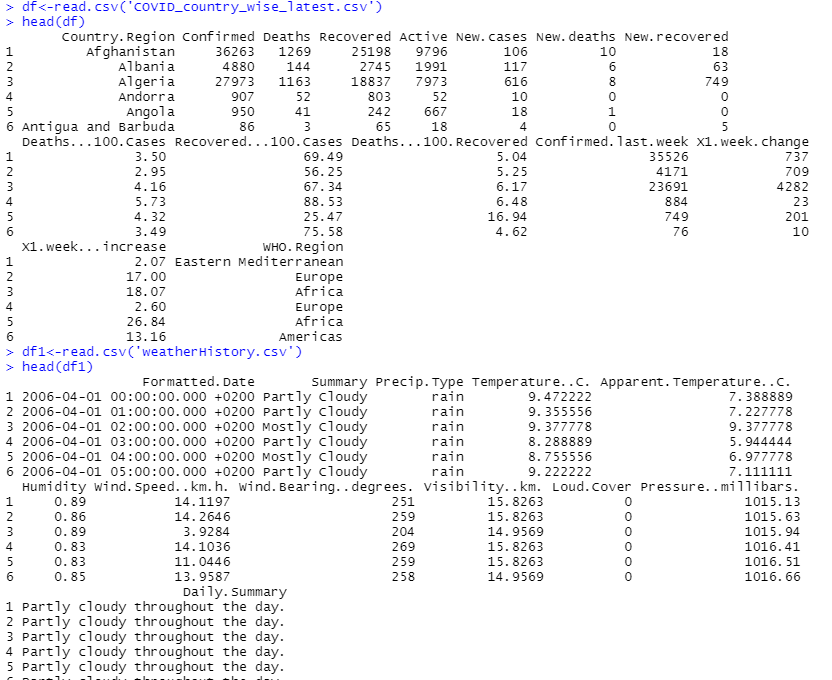
print(paste("max of ",i,max(df1[,i],na.rm = TRUE)))

print(paste("range of ",i,range(df1[,i],na.rm = TRUE)[1],range(df1[,i],na.rm = TRUE)[2]))

print(paste("var of ",i,var(df1[,i],na.rm = TRUE)))

}

}







1. **Write a function that has three vector arguments for merging the into an existing dataframe.**

merger<-function(a,b,c){

df$dummy1=a

df$dummy2=b

df$dummy3=c

head(df)

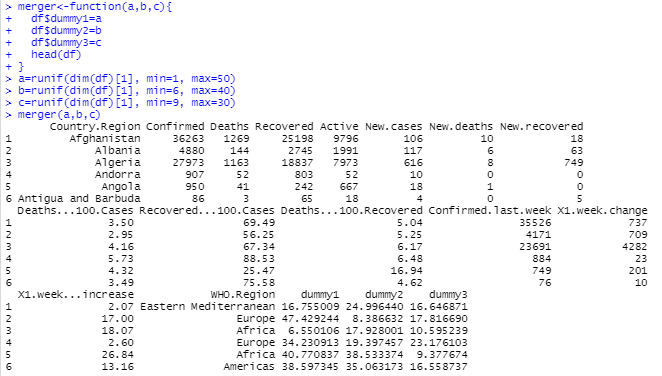
}

a=runif(dim(df)[1], min=1, max=50)

b=runif(dim(df)[1], min=6, max=40)

c=runif(dim(df)[1], min=9, max=30)

merger(a,b,c)



1. **After merging create a function compute to find out min, max and avg of all numeric columns.**

v3<-colnames(df)

for(i in v3){

if(typeof(df[,i])=="integer"){

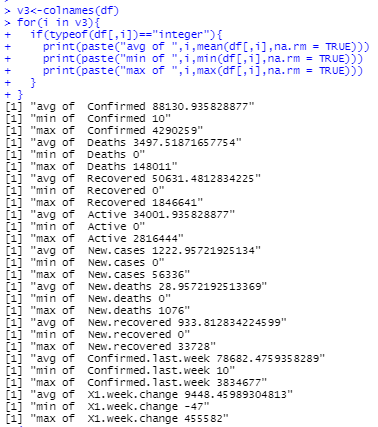
print(paste("avg of ",i,mean(df[,i],na.rm = TRUE)))

print(paste("min of ",i,min(df[,i],na.rm = TRUE)))

print(paste("max of ",i,max(df[,i],na.rm = TRUE)))

}

}



1. **The summary values should be in a single data frame with the following columns: variable name, mean, sd, minimum, and maximum.**

j=0

summary<-data.frame()

for(i in v3){

if(typeof(df[,i])=="integer"){

j=j+1

summary[j,1]=i

summary[j,2]=mean(df[,i],na.rm = TRUE)

summary[j,3]=sd(df[,i],na.rm = TRUE)

summary[j,4]=min(df[,i],na.rm = TRUE)

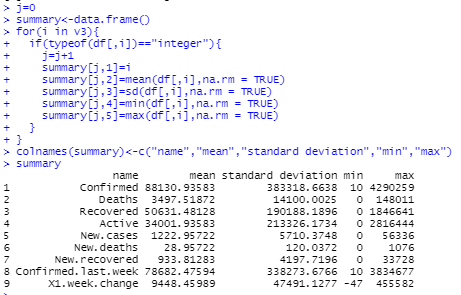
summary[j,5]=max(df[,i],na.rm = TRUE)

}

}

colnames(summary)<-c("name","mean","standard deviation","min","max")

summary



1. **Write a function so that the summary of the dataframe should be written to a csv file and to R.**

writer<-function(df){

write.csv(df,'summary.csv')

}

writer(summary)

df<-read.csv("summary.csv")

df

