```
#R project
#Dataset : Cities AirQuality and WaterPollution (2020)
#Divyansh Niranjan (20/1683)
#Rajanish Pandey (20/325)

install.packages("dplyr")
library(dplyr)
getwd()
setwd("C:/Users/acer/Documents/math")
as=read.csv("C:/Users/acer/Documents/math/AirWater.csv")
View(as)
# 1. What are the dimension of dataset ?
dim(as)

# 2. Give detail of the structure of datasets.
str(as)

# 3. Write a summary of the dataset along with special remark.
summary(as)

# 4. What are the col umn names of  the datasets ? Also display top 4 row and
bottom 5 rows.
names(as)
slice_head(as,n=4)
slice_tail(as,n=5)

# 5. Is there NA in the data sets ?
library(readxl)
as1 = read_excel("AirWater1.xlsx")
View(as1)

View(is.na(as1))
length(which(is.na(as1)))

# thers is no NAs in this data set.
# 6. If the answer is "Yes" in the above question then remove all NAs .
# there are no NAs in this data set.

# 7. Again find the dimensions of the data sets .
dim(as1)

# 8. Find the measures of central tendencies.
attach(as)
mean(AirQuality)
mean(WaterPollution)
median(AirQuality)
median(WaterPollution)

# 9.Plot atleast two variable of the data sets.
plot(AirQuality,WaterPollution)

# 10. Use which command to fetch certain data.
which("Country"=="United States of America")
d1=data.frame(which(AirQuality>80))
```

```
View(d1)
which(WaterPollution<26)

# Question 1
#Creat a  new data set from the given data set which has air quality less than
50 and water pollution less than 40.
newdata=data.frame(filter(as,AirQuality<50,WaterPollution<40))
View(newdata)

# Question 2
#Find out total number of cities having air quality less than 30 and water
pollution greater than 80 also viwe this data.
dim(filter(as,AirQuality<30,WaterPollution>80))
count(filter(as,AirQuality<30,WaterPollution>80))

# Question 3
# data set of all cities of India .
attach(as1)
View(filter(as,Country ==" India"))

# Question 4
#Find out cities with minimum and maximum water pollution and maximum
airquality and water pollution.
View(slice_max(as1,WaterPollution))
View(slice_min(as1,WaterPollution))
View(filter(as1,AirQuality==100,WaterPollution==100))


# Question 5
#Plot both numeric column with labels.plot histogram for airquality and water
pollution.
boxplot(AirQuality,WaterPollution,names=c('AirQuality','WaterPollution')
        ,col=c('lightblue','lightgreen'),horizontal = TRUE,notch = TRUE,range
= 1)

#6 to display the random numbers of specific column.
fivenum(AirQuality)

#7 To display the specific data of a specific column.
names(as)
as %>% group_by(Country) %>% summarize(mean(AirQuality))

#8 Transpose of the datasets.
transpose.as<-t(as)
View(transpose.as)

#9 how to arrange the data in ascending or descending order?
View(arrange(as,-desc(AirQuality)))

#10 use of summarize() command.
summarize(as,mean(AirQuality))
as %>% summarize((mean(AirQuality)))
```