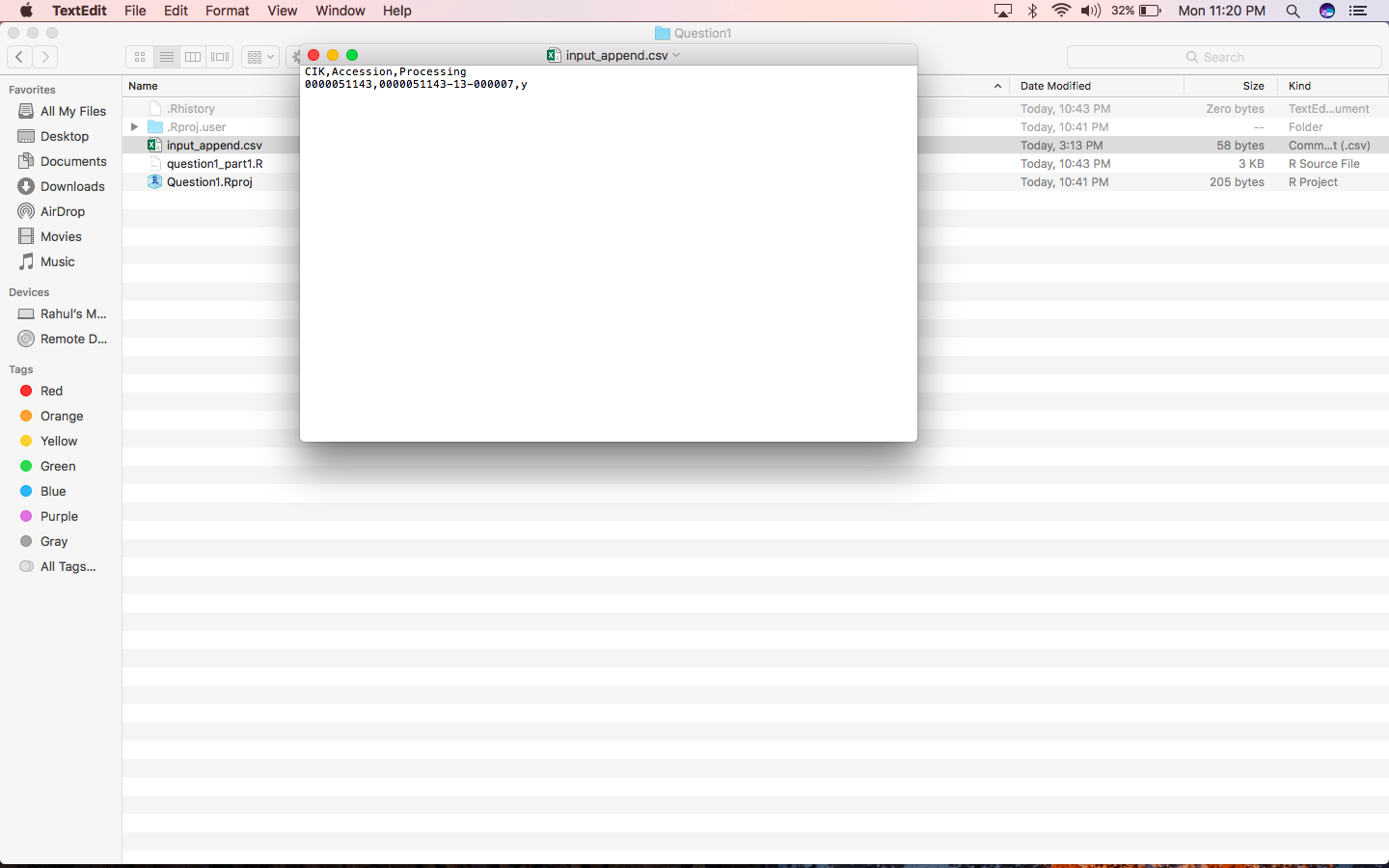
QUESTION 1 :

The motive of the question was to detect all the tables and extract them to them separately. This is process is done many steps.

The first step is to get the url of the Company’s 10-Q and 10-K files.

For this a config file is made where there are 3 columns :

1. CIK
2. Accession
3. Processing



The values in the file are CIK of the company and Accession is the accession no of the company that is provided as input. The processing column has values ‘y’ or ‘n’ to specify the cik to be executed. Note that only one value in the file should have ‘y’ value.

The code first takes the input as the values of CIK and accession number and creates the links for the html page. Attached is the code snippet:

filename <- gsub(" ", "\_", name[1])

field1 = substr(cik,regexpr("[^0]",cik),nchar(cik))

field2 = unlist(strsplit(acc,"-"))

print(paste(Sys.time(),"Generating files for CIK : ",field1," and name : ",name[1],sep=" "))

create\_url\_1 <- paste("https://www.sec.gov/Archives/edgar/data/",field1,"/",field2[1],field2[2],field2[3],"/",field2[1],"-",field2[2],"-",field2[3],"-index.html",sep = "")

url\_accession <- read\_html(create\_url)

url\_accession %>% html\_nodes(xpath = './/\*[@id="formDiv"]/div/table') %>% .[[1]] %>% html\_table(fill = TRUE) -> testTable

for (i in 1:length(testTable)){

if (testTable$Description[i] == "10-Q"){

field3 <- testTable$Document[i]

}

if(testTable$Description[i] == "FORM 10-K"){

field3 <- testTable$Document[i]

}

}

print(paste(Sys.time(),"Fetching from ",field1," and name : ",name[1],sep=" "))

create\_url\_2 <- paste("https://www.sec.gov/Archives/edgar/data/",field1,"/",field2[1],field2[2],field2[3],"/",field3,sep = "")

After fetching the values form the table it is cleaned and all the unwanted values are removed such as spaces and NA values.

The corrected tables are then send to folder with name as that of it’s company’s name.

Then this folder is zipped and pushed to S3 bucket.

A log file is generated and it is also pushed to S3 bucket.