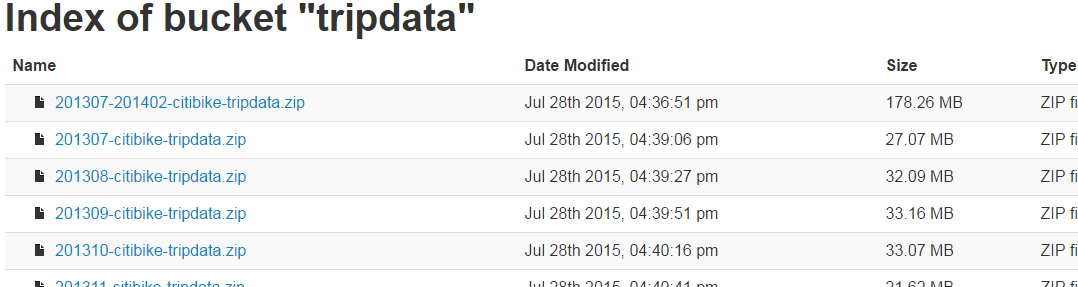
R: Direct File Download

Example #1

Consider the following website: <https://s3.amazonaws.com/tripdata/index.html>



|  |  |
| --- | --- |
| Citi Bike Website | HTML Code does not contain links |

The following code produces no output.

library(rvest)

url<-"https://s3.amazonaws.com/tripdata/index.html"

page <- read\_html(url)

temp<-html\_text(html\_nodes(page, "a"))

df<-as.data.frame(temp)  
View(df)



Direct File Download

The download.file() function will allow one to directly download a file. The file being downloaded here is a compressed file (\*.zip) and thus needs to be unzipped before using the read.csv().

#Using download.file to get file

#create a temporary file

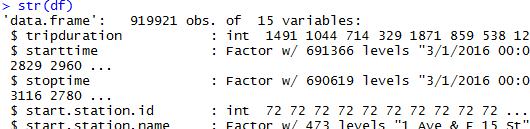
temp <- tempfile()

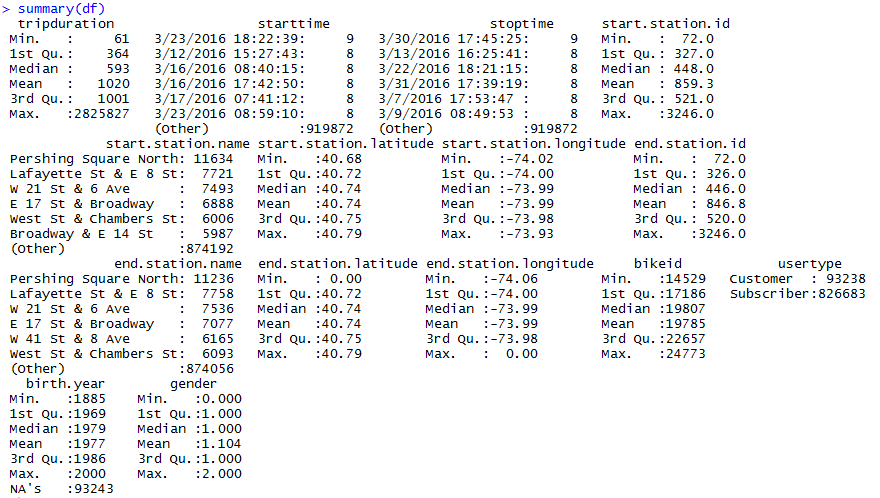
download.file("https://s3.amazonaws.com/tripdata/201603-citibike-tripdata.zip",temp,mode="wb")

temp1 <- unzip(temp, "201603-citibike-tripdata.csv")

df <- read.csv(temp1, header=T, sep=",")

unlink(temp)

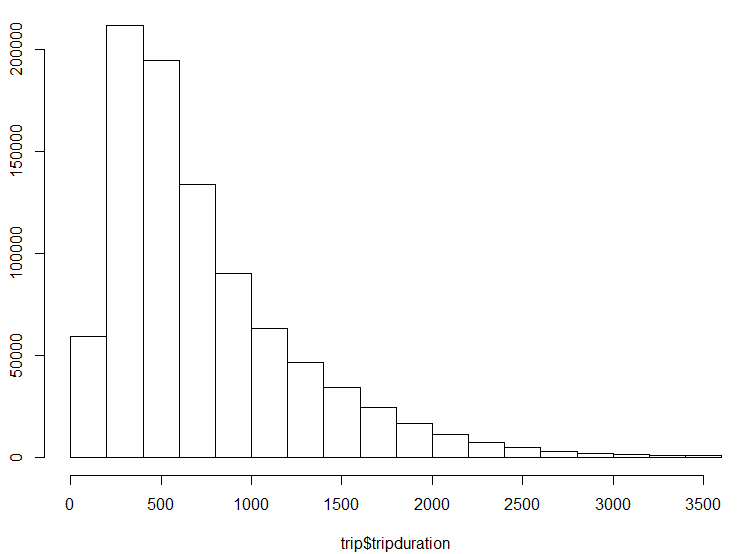




The download.file() function

trip<-subset(df,tripduration<3600,select=tripduration)

hist(trip$tripduration)



Example #2

Consider the following website: https://www.ssa.gov/oact/babynames/limits.html

#Reading in zip file that contains several files  
zipdir <- tempfile()

dir.create(zipdir)

download.file("https://www.ssa.gov/oact/babynames/state/namesbystate.zip", temp, mode="wb")

unzip(temp, exdir=zipdir)

files <- list.files(zipdir, pattern="\\.TXT$")

df<-data.frame()

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

filepath <- file.path(zipdir,files[i])

temp <- read.csv(filepath,header=F)

df<-rbind(df, temp)

}

unlink(zipdir)