IS677 HW13

Mike Lehman

November 28, 2016

#1  
airline\_08 <- read.big.matrix("2008.csv", type = "integer", header = TRUE, backingfile = "2008.bin", descriptorfile = "2008.desc")

#2  
options(scipen = 999, digits = 0)  
  
start\_time <- Sys.time()  
  
summary(airline\_08)

## min max mean NAs  
## Year 2008 2008 2008 0  
## Month 1 12 6 0  
## DayofMonth 1 31 16 0  
## DayOfWeek 1 7 4 0  
## DepTime 1 2400 1334 136246  
## CRSDepTime 0 2359 1326 0  
## ArrTime 1 2400 1481 151649  
## CRSArrTime 0 2400 1495 0  
## UniqueCarrier 9 9 9 6747520  
## FlightNum 1 9743 2224 0  
## TailNum 9149 96079 86552 6754265  
## ActualElapsedTime 12 1379 127 154699  
## CRSElapsedTime -141 1435 129 844  
## AirTime 0 1350 104 154699  
## ArrDelay -519 2461 8 154699  
## DepDelay -534 2467 10 136246  
## Origin 7009728  
## Dest 7009728  
## Distance 11 4962 726 0  
## TaxiIn 0 308 7 151649  
## TaxiOut 0 429 16 137058  
## Cancelled 0 1 0 0  
## CancellationCode 7009728  
## Diverted 0 1 0 0  
## CarrierDelay 0 2436 16 5484993  
## WeatherDelay 0 1352 3 5484993  
## NASDelay 0 1357 17 5484993  
## SecurityDelay 0 392 0 5484993  
## LateAircraftDelay 0 1316 21 5484993

stop\_time <- Sys.time() - start\_time  
stop\_time

## Time difference of 1 secs

Machine Details: Surface Pro 4 Windows 10 16.0 GB RAM 2 physical cores, 4 threads

#3  
late\_flights\_ind <- mwhich(airline\_08, "ArrDelay", 0, "gt")  
nrow(airline\_08[late\_flights\_ind,])

## [1] 2979504

#4  
early\_flights\_ind <- mwhich(airline\_08, "ArrDelay", 0, "lt")  
nrow(airline\_08[early\_flights\_ind,])

## [1] 3845305

#5  
ontime\_flights\_ind <- mwhich(airline\_08, "ArrDelay", 0, "eq")  
nrow(airline\_08[ontime\_flights\_ind,])

## [1] 184919

#6  
distances <- airline\_08[,"Distance"]  
mean(distances)

## [1] 726

#7  
on\_time\_distances <- airline\_08[ontime\_flights\_ind,"Distance"]  
mean(on\_time\_distances)

## [1] 662

#8  
early\_distances <- airline\_08[early\_flights\_ind,"Distance"]  
mean(early\_distances)

## [1] 708

#9  
late\_distances <- airline\_08[late\_flights\_ind,"Distance"]  
mean(late\_distances)

## [1] 754

# 10

This question cannot be answered because the read.big.matrix command can only accept one data type for all the values it imports from an external file. We used the integer data type as most of the data points were integers/numeric in nature. Subsequently, the bigmemory package imported the Dest feature with all NAs as the original value was of data type string since they are the corresponding IATA airport code.