```
# This program is written in R programming language version ‘3.1.1’
installed on a Linux server. "R is a free software environment for
statistical computing and graphics" with
# no guarantees. R compiles and runs on a wide variety of UNIX platforms,
Windows and MacOS." To download a free copy of R visit "http://www.r-
project.org/".
# In addition, the following R packages were used in this program:
# package "foreach" version 1.4.0
# package "data.table" version 1.9.4
# package "reshape2" version 1.2.1
# package "XLConnect" version 0.2-10
# package "zoo" version 1.7-7

# This program will downloaed from the internet and install the latest
version of the above packages If they are not installed in your R
environment. It is necessary to
# have internet connection to download these packages.


# If for any reason this program fails to run, please make sure that the
above packages are installed, check the verion of the packages and
# make sure the functions called in this program are still in use and are
compatible with the Operating System you are using.

# A step-by-step description is provided throughout this code.

#########################################################################
#################################################################

# Load Necessary Packages for this analysis

if (!(require(foreach))) install.packages ("foreach")
if (!(require(data.table))) install.packages ("data.table")

# You will need to download Fannie Mae's Single-Family Loan Performance
Data from Fannie Mae's website at
https://loanperformancedata.fanniemae.com/lppub/index.html.
# After downloading the files you will need to unzip the files. Though
read.table function in R can read zipped files,
# we have used the "fread" function from data.table package to read these
files for efficiency and speed. Unfortunately, fread cannot read zipped
files.
# This program will work with any number of pairs of Acquisition and
Performance files. We encourage users to download them all for the
complete data set.
# In order for this code to run properly, the naming of the files should
remain the same after download and unzipping process so that the files
are saved in order.

# You will need the path to the downloaded files, please copy and paste
or type the path below:
fileslocation<-"/<INSERT FILEPATH TO UNZIPPED PAIR(S) OF FILES HERE>/"
```

```
# Check the number of files downloaded (should be even, equal number of
Acquisition and Performance Files)

numberoffiles<-length(list.files(fileslocation, pattern =
glob2rx("*txt"), full.names=TRUE))

# with the help of "foreach" package we contruct a loop so that R can
loop through the downloaded files and perform the analysis
# Number of iteration (files will be processed in pairs, also, could be
used as the number of cores in parallel processing)

numberofcores<-(numberoffiles/2)

# Below, after defining the Acquisition and Performance variables and
their classes, the files are read into R and then data manipulation is
carried out.
# Acquisition and Performance files (from one or many quarters) will be
merged into an R dataframe called "Combined_Data"

# Define Acquisition Variables, variable classes and read the files into
R
Acquisitions <- list.files(fileslocation, pattern =
glob2rx("*Acquisition*txt"), full.names=TRUE)

Acquisitions_Variables = c("LOAN_ID", "ORIG_CHN", "Seller.Name",
"ORIG_RT", "ORIG_AMT", "ORIG_TRM", "ORIG_DTE"
                           ,"FRST_DTE", "OLTV", "OCLTV", "NUM_BO", "DTI",
"CSCORE_B", "FTHB_FLG", "PURPOSE", "PROP_TYP"
                           ,"NUM_UNIT", "OCC_STAT", "STATE", "ZIP_3",
"MI_PCT", "Product.Type", "CSCORE_C", "MI_TYPE")

Acquisition_ColClasses = c("character", "character", "character",
"numeric", "numeric", "integer", "character", "character", "numeric",
                           "numeric", "character", "numeric", "numeric",
"character", "character", "character", "character", "character",
                           "character", "character", "numeric",
"character", "numeric", "numeric")


# Define Performance Variables, variable classes and read the files into
R
Performance <- list.files(fileslocation, pattern =
glob2rx("*Performance*txt"), full.names=TRUE)

Performance_Variables = c("LOAN_ID", "Monthly.Rpt.Prd", "Servicer.Name",
"LAST_RT", "LAST_UPB", "Loan.Age", "Months.To.Legal.Mat"
                          , "Adj.Month.To.Mat", "Maturity.Date", "MSA",
"Delq.Status", "MOD_FLAG", "Zero.Bal.Code",
                          "ZB_DTE", "LPI_DTE", "FCC_DTE","DISP_DT",
"FCC_COST", "PP_COST", "AR_COST", "IE_COST", "TAX_COST", "NS_PROCS",
                          "CE_PROCS", "RMW_PROCS", "O_PROCS",
"NON_INT_UPB", "PRIN_FORG_UPB", "REPCH_FLAG")
```

```
Performance_ColClasses = c("character", "character", "character",
"numeric", "numeric", "numeric", "numeric", "numeric", "character",
                           "character", "character", "character",
"character", "character", "character", "character", "character",
                           "numeric", "numeric", "numeric", "numeric",
"numeric", "numeric", "numeric", "numeric", "numeric", "numeric",
"numeric", "character")

# Save a Copy to disk by executing the following line of code:
save(Performance_Data, file="FANNIEMAE_Performance_Data.Rda")

#Close Connections created as result of Running Foreach
env <- foreach:::.foreachGlobals
rm(list=ls(name=env), pos=env)


Acquisitions_Data <- foreach(k=1:numberofcores, .inorder=FALSE,
.combine=rbind,
                             .packages=c("data.table")) %do% {
                               Data_A<- fread(Acquisitions[k], sep = "|",
colClasses=Acquisition_ColClasses, showProgress=FALSE)
                               setnames(Data_A, Acquisitions_Variables)
                               setkey(Data_A, "LOAN_ID")
                               }

Performance_Data <- foreach(k=1:numberofcores, .inorder=FALSE,
.combine=rbind,
                            .packages=c("data.table")) %do% {
                              Data_P<- fread(Performance[k], sep = "|",
colClasses=Performance_ColClasses, showProgress=FALSE)
                              setnames(Data_P, Performance_Variables)
                              setkey(Data_P, "LOAN_ID")
                            }

# Save a Copy to disk by executing the following line of code:
save(Acquisitions_Data, file="FANNIEMAE_Acquisitions_Data.Rda")

rm(list= ls()[!(ls() %in% c('Acquisitions_Data', 'Performance_Data'))])
```