Get Parkinson's Kaggle data into R dataframes (on Mac)

Make sure that you have 30+ GB free storage on PC before using this help file.

1. Manually unzip Parkinson.zip.
    # remove zip file afterward to free space.
2. Open Mac terminal.

step 3-4 are in mac terminal.
3. Set working directory to Parkinson folder.
    cd PathTo/Parkinson
    # for example, on my mac, PathTo is ~/Dropbox/Courses/PH244/Project_III/data.

4. Uncompress files.
    # uncompress mjff_binary_files and get MJFF-binary-files folder of tar.bz2 files.
    unzip mjff_binary_files.zip

    # remove zip to free space
    rm mjff_binary_files.zip

    # go to MJFF-binary-files folder.
    cd ./MJFF-binary-files/

    # compress all .tar.bz2 files and delete afterward.
    # this operation is going to take a while.
    # after the step, MJFF-binary-files folder consists of subfolders of bin and log files.
    *for file in *.tar.bz2; do tar -jxf "$file"; rm "$file" ; done*

step 5 is in Rstudio
5. Open Rstudio, and create .csv.bz2 files using HumDyn.R.
    # I slightly changed HymDyn.R. Please use the version offered by class, not by Kaggle.
    # modify input in line 12 into the path of MJFF-binary-files.
    # for example, in my case, paths = list.dirs("~/Dropbox/Courses/PH244/project_III/data/MJFF-binary-files/")
    # this step is going to take a while
    # after HumDyn.R, in each subfolder you will see .csv.bz2 files

Step 6 done in mac terminal.
6. Uncompress .csv.bz2 files
    # make sure that you are in the folder MJFF-binary-files
    # this step is going to take a while
    find . -name '*.csv.bz2' -execdir bzip2 -d '{}' \;

Step 7 is done in Rstudio
7. Process the .csv files into R data frames using Process.R
    # set mypath variable in line 5 as the path of MJFF-binary-files
    # for example, in my case, mypath="~/Dropbox/Courses/PH244/project_III/data/MJFF-binary-files/"

this script creates 7 data frames, one for each data type, with an extra indicator column for whether the row of data was from a person with Parkinson's (PK=1) or not.
data frames are saved in mjff.Rdata in MJFF-binary-files folder. You may delete the rest subfolders therein to free space.