1. **Download R.**

* First go to <https://www.r-project.org/> and click the “Download R” link under the “Getting Started” Header.

A screenshot of a cell phone

Description automatically generated

* Select a CRAN mirror that is in your country and click the link.
* Click “Download for Mac”, “Download for Linux” or “Download for PC” accordingly.
* Click on the link that has the latest version of R.
* Save and open the downloaded file and follow the installation instructions.

1. **Download RStudio**

* Go to <https://rstudio.com/products/rstudio/download/>.
* Choose the “RStudio Desktop” Option and click “Download”

A screenshot of a cell phone

Description automatically generated

* Click the version recommended for your system and open the .dmg file on your computer.

1. **Get the R Script from Github**

* Go to <https://github.com/jellen44/AutomaticSleepScoringTool>

A screenshot of a cell phone

Description automatically generated

* Click on “SleepScoringTool.R”
* Highlight the text in the file and copy

A screenshot of a social media post

Description automatically generated

* Open RStudio and go to “File” → “New File” → “RScript”
* Paste the cxopied text from Github into this RScript and go to “File” → “Save” to save this new file

1. **Packages**

* There should be a pop-up at the top of the R script explaining that some packages are not installed:



* Click on “Install” and if the bar doesn’t appear, continue because the script will install them automatically.
* The first time you run the script, you will likely to prompted by R to enter a “y” into the command line to approve the installation of certain packages.
* NOTE: Line 124 in the script reads: “install\_tensorflow(restart\_session = FALSE)” – it is not necessary to keep this line in after the first time running the script, so you can delete it and there is a heading labeled “Delete This After First Time Running Script” above it on line 123 so you know to do so.

1. **Sleep Scoring Methodology**

* This model works effectively with small amounts of data, but still needs some amount of data to train.
* You should score fifty different sequences of ten consecutive epochs randomly distributed throughout the file (500 epochs in total).
* Afterwards, you should specifically seek out and score 60 REM epochs to make sure that REM is represented in the dataset.
* Ultimately you should end up with 560 epochs scored in total (2.6% of a 24-hour file with 4-second epochs).

1. **Correctly Formatting Input Data**

* This script works with either one or two EEG signal files and one EMG file (default is two EEGs).
* If you have any questions about the below formatting instructions, please see example files at <https://github.com/jellen44/AutomaticSleepScoringTool>, and click “ExampleData,” and you can view separate example files for the EEG1, EEG2 and EMG files. Or you can send an email to jgellen4@gmail.com.
* This script takes in raw EEG and EMG data so there is no need to process the data at all, but they must be in the correct format.

1. **Requirements for All Files**

* These files must be comma-separated values (.csv) spreadsheets. You can make a spreadsheet a .csv file in Excel by going to “File” → “Save As” and choosing “comma-separated values (.csv)” from the menu.
* These spreadsheets should consist of each row representing every raw datapoint from a given epoch (no matter how long that epoch is). Thus, the number of columns and rows should be the same in all three files.
* The number of rows should equal the number of epochs recorded and the number of columns should equal the number of datapoints recorded within each epoch.
* For example, if you have 4 second epochs with a sampling rate of 250 Hz, there will be 1,000 columns (4 seconds divided by .004 seconds is 1,000).
* There does not need to be any column headers, so the first row can just be the first epoch of raw data.

1. **Formatting the Scoring Results Column**

-The last piece of information is that there needs to be a column of data that with the epochs you scored.

-Here, you can just add that column to the end of the EEG1 csv file as the last column in the dataset.

-You also have the option of either making a separate .csv file that has all of the scoring results in one column which can be entered as an optional parameter in the R script.

-You MUST label a wake epoch a 0, a non-REM epoch a 1 and a REM epoch a 2 within this column.

-You can leave all of the other unscored columns blank.

-Ultimately, each row should contain a 0, a 1 or a 2, or be left blank if it is not scored.

A close up of a screen

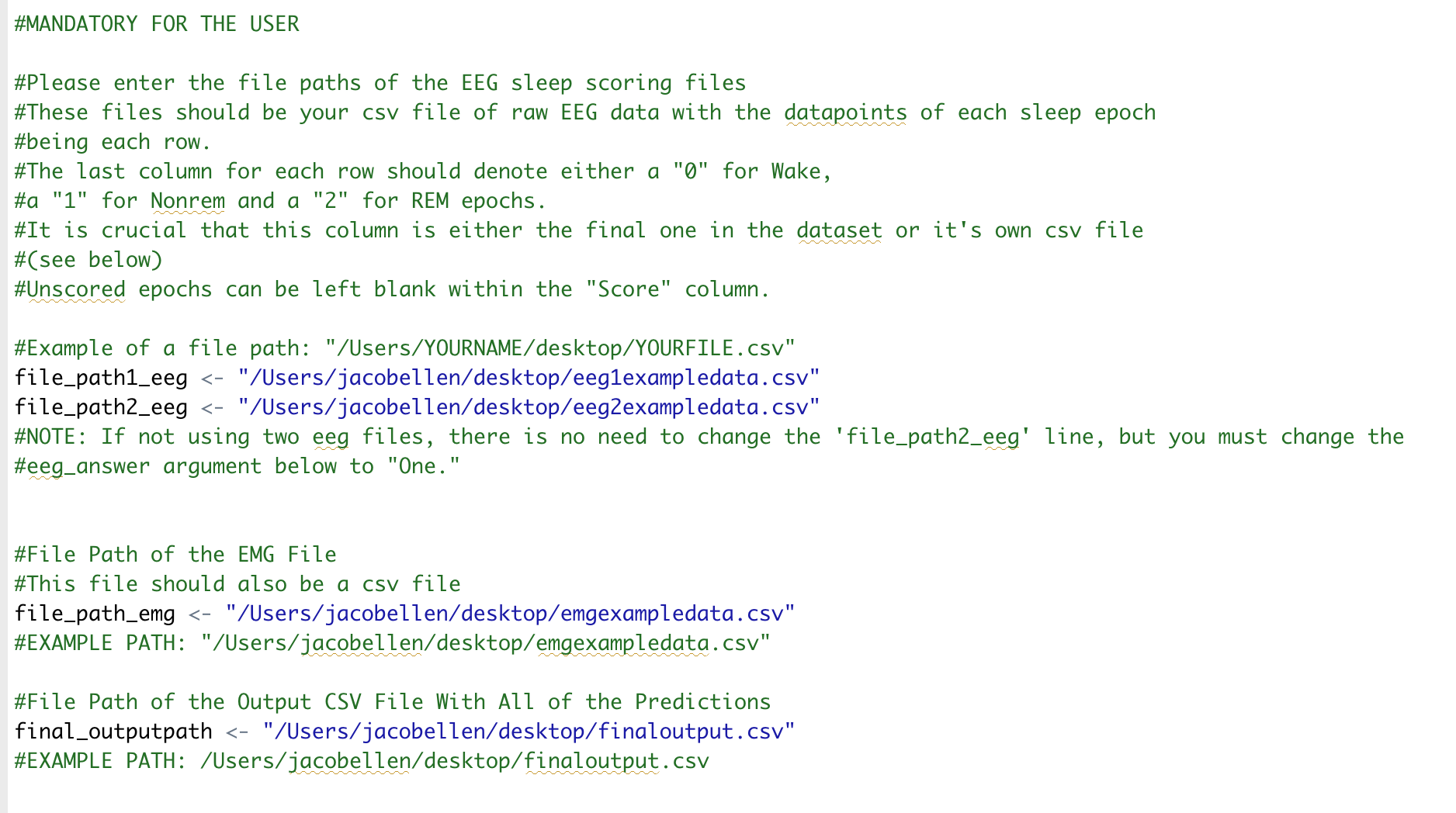
Description automatically generated

* This column would represent a series of wake epochs, followed by unscored ones
* It is crucial that this Scoring column above is matched by row to the epoch it is scoring.

1. **Using the Script**

**-** If you have one file with the EEG1 data, a file with the EMG data (the second EEG is optional), then you can start using the script.

**-** Read the “MANDATORY FOR THE USER” section within the script as you must change some of the answers.

**-** ****

Note: if you do not have experience writing file paths, refer to (<http://www.mactips.info/2011/11/how-to-read-and-write-a-filepath>) for Macs and (<https://www.pcworld.com/article/251406/windows-tips-copy-a-file-path-show-or-hide-extensions.html>) for Windows.

**Mandatory Arguments:**

­-*file\_path1\_eeg* – this is the file path to your first eeg .csv file (in quotes)

-*file\_path2\_eeg* – this is the file path to your second eeg file (if you don’t have one just put empty quotations in that section “”)

-*file\_path\_emg* – this is the file path to your emg .csv file

-*final\_outputpath* – this is the path of where you would like the final output file to go on your computer.

*-samplingrate* – input the number of samples taken per second of the EEG (in Hertz). Note that this script will accept any sampling rate greater than or equal to 150 Hz (see warning section below) and 250 Hz is the default.

**Optional Arguments:** you don’t have to change these unless you want to

A. Number of EEGs

-*eeg\_answer* – the number of EEG’s you would like to use for scoring. If one, put “One” (default is two).

1. Inputting Raw EEG/EMG Data as One Long Vector

-*vectorinput* – if you would like to input your raw EEG and EMG data as single vectors (.csv’s in one single column), then put “Yes” and you must change the “epochlength” arguments below.

-*epochlength* – the number of seconds for each epoch for scoring.

-*scoringcolumn\_filepath* – If you select *vectorinput* as “Yes,” then you need to provide an external scored file (because the scored file can no longer be in EEG1 spreadsheet). This external file should be a .csv with one column that is the scored epochs (and many more unscored epochs). Here, you provide the path to that file.

C. Copy of Features Extracted

-*frequency\_answer* – if you would like a spreadsheet of the data after it has been analyzed into features, such as delta band power and theta band power for example, put “Yes.”

*-file\_output\_path* – if you answered “Yes” to frequency\_answer, this is the output path of that file.

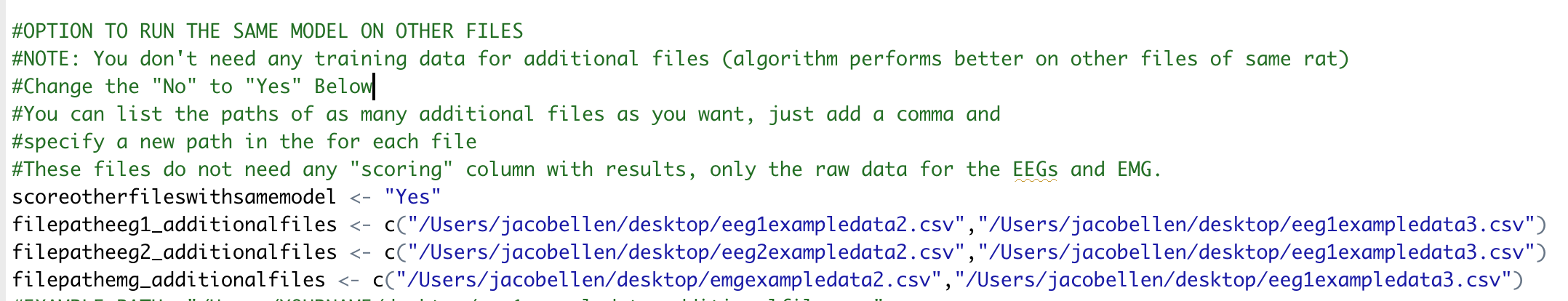
Example of what the “frequency\_answer” dataset would look like:

A picture containing window, sitting, large, table

Description automatically generated

1. **Scoring Additional Files**

* One last optional paramater is the ability to score multiple other files based on the training of a single file (see paper: model performs well on different files of the same rat, but not as well on other rats).
* *scoreotherfileswithsamemodel* – put “Yes” if you would like to do this (default is “No”)
* *Filepatheeg1\_additionalfiles* – path to the additional first EEG file in the same format as explained above
* *Filepatheeg2\_additionalfiles* – path to the additional second EEG file
* *Filepathemg\_additionalfiles* – path to the additional EMG file
* Note: there is no need to provide any scoring data or additional column to these extra files.
* You can add as many additional files as you want here by putting commas between them as shown below.



1. **Model Output**

-This model outputs a .csv file to the “final\_outputpath” specified.

-This file is the EEG1 data, but with the final column named “ModelPredictions,” which contains the prediction for each epoch in the whole dataset (training epochs are kept to the same label that the user originally gave it).

-Again, these predictions will be in the form where a 0 means wake, a 1 means non-REM and a 2 is REM.

1. **Rescoring Option**

-There is one additional output column in the output file called “ModelCertainty,” where each epoch will be labelled as “Uncertain” if the average certainty of the five models run is under 90% (left blank if not).

-This will usually correspond to a small percentage of the dataset and can be used to rescore these “uncertain” epochs to achieve more accurate results (see paper).

**Explaining Potential Warnings**

1. *Insufficient REM Warning*

* If the training data consists of less than 10% REM, a warning will be sent and your REM scoring will likely not be very effective.

1. *Poor EMG Quality Warning*

* If the EMG coefficient of variation is worse than 1.67, meaning the signal quality of your EMG is poor, a warning will be sent and your REM scoring will likely not be very effective (see paper for details).

1. *Sampling Rate Warning*

* If the sampling rate is lower than 150 Hz (the minimum), the script will stop and an error will be produced.
* Any sampling rate >=150 Hz and above will work, although the “high gamma” feature will be removed if the sampling rate is under 250 Hz because not all “high gamma” data will be present.

Thank you and please direct any questions or concerns to jgellen4@gmail.com