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 Linux” 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 copied text from Github into this RScript and go to “File” → “Save” to save this new file

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 thirty-one different sequences of ten consecutive epochs randomly distributed throughout the file (310 epochs in total).
* Afterwards, you should specifically seek out and score 20 REM epochs to make sure that REM is represented in the dataset.
* Ultimately you should end up with 330 epochs scored in total (1.5% 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.
* 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 or normalize the data at all, but they must be in the correct format.

1. **Requirements for All Files**

* These files must be a 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 dropdown 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 datapoints recorded every 4 milliseconds, 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 have the option of either making a separate .csv file that has all of the scoring results in one column, or you can just add that column to the end of the EEG1 csv file as the last column in the dataset.

-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, so even if you don’t score the first 100 epochs, for example, make sure that they are still the first 100 epochs in your column and do not change the order.

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

**-** **A screenshot of a social media post

Description automatically generated**

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.

**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.

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

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

-*vectorinput* – if you would like to input your raw EEG and EMG data as single vectors (all in one single column), then put “Yes”

*-numberofdatapointsperepoch* – input the number of samples per epoch (can be found by dividing the seconds per epoch by the seconds per datapoint)

-*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.

-*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. **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 “Model Predictions,” 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 means REM.

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