1.csv – Dataset of English words. (<http://crr.ugent.be/programs-data/word-ratings>) It has the mean, st dev and number of ratings for the valence, arousal and domination of each word (In the columns that say Sum). It also has the same values but for specific groups of “raters” (.M = male; .F = female; .O = older; .Y = younger; .H = high education; .L = low education). **13915 words. Crescent Valence and Arousal scores from 1 to 9. Has index column. Column “Word” with the word, “V.Mean.Sum” with the Valence score, “A.Mean.Sum” with the Arousal scores.**

Decision to make: Use just the means for all words, ~~or force some kind of threshold concerning a minimum number of ratings and a maximum allowed SD? Maybe conditional on the mean and st dev of the st dev and number or ratings themselves~~?

2.xlsx – Dataset of Dutch words (<http://crr.ugent.be/programs-data/word-ratings>) (same source as 1, so similar fields). **4299 words (even though the source says 4300). Crescent Valence and Arousal scores from 1 to 9. No index column. Column “Words” with the word, “M V” with the Valence score, “M A” with the Arousal scores.**

3.csv ou 3.xlsx – Dataset of Polish Sentences (https://figshare.com/s/e4b4e339138f07c63153) (with their English translations as well). It has the number of words and letters, the mean valence, arousal, dominance, origin, significance and source. It also has these values segmented by different groups of people and samples. **718 sentences. Crescent Valence and Arousal scores from 1 to 9. Has index Column. Column “Polish sentence” with the sentence, “Valence all” with the Valence, “Arousal All” with the arousal. Has two naming rows, important to IGNORE THE SECOND ROW WHILE IMPORTING.**

4.csv – Dataset of Spanish words (<http://crr.ugent.be/programs-data/word-ratings>) (https://link.springer.com/article/10.3758/s13428-015-0700-2) (same source as 1, so similar fields). **14031 words. Crescent Valence and Arousal scores from 1 to 9. Has index column. Column “Word” with the word, “ValenceMean” with the Valence score, “ArousalMean” with the Arousal scores. IMPORTANT TO REMOVE THE “\*” SIGN FROM THE END OF A LOT OF THE WORDS** (it’s because that word has less than 10 raters for at least 1 on the measures)

5.xlsx – Dataset of Polish Words (https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01081/full) (with their English translations as well). It has the number of letters, the mean valence, arousal, dominance and other measures. It has the POS tag classification. It also has these values segmented by different groups of people and samples. **4905 words. Crescent Valence and Arousal scores from 1 to 9. Has index Column. Column “polish word” with the sentence, “Valence\_M” with the Valence, “arousal\_M” with the arousal.**

5 – Dataset of Polish Words (https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01081/full) (with their English translations as well). It has the number of letters, the mean valence, arousal, dominance and other measures. It has the POS tag classification. It also has these values segmented by different groups of people and samples. **4905**

6.csv ou 6.xlsx – Dataset of English sentences (from Facebook) (https://github.com/wwbp/additional\_data\_sets/tree/master/valence\_arousal). Under consideration because of weird characters and anonimization. **Anonimized URLs, Addresses and Emails were removed, Person and Phone segments were filled with random names and phone numbers. Empty rows were removed (rows that were <URL> now are empty). 2892 sentences.**

7.txt – Same as 6 Problemas graves aqui, nem ftfy resolve. O apóstrofe está hiper desformatado e o caracter utf-8 correspondente é um Õ. Corrigi manualmente mas há imensos artefactos pelo meio. (ISO-8859 text, with CRLF line terminators)

8.xlsx – Dataset of English Sentences, from popular tales (<https://dl.acm.org/doi/abs/10.1007/s10579-011-9140-5>). Same issue as 6, it happens in the ‘ character/symbol. **Uncertain number of sentences. It has multiple sheets with sentences. Will need processing. Has Evaluation (Valence) and Activation (Arousal) scores, crescent, for multiple annotators (will need to be averaged). Problem with symbols TIVE QUE REMOVER PRINCESS AND THE PEA, TINHA COLUNAS A MAIS SEM LABEL. Opá isto é quase tudo perto do 5…. Ficaram 1368 frases.**

9.csv – Dataset of English Sentences, from EmoBank (<https://github.com/JULIELab/EmoBank>). **10062 sentences. Crescent Valence and Arousal scores from 1 to 5. No index Column. Column “text” with the sentence, “V” with the Valence, “A” with the arousal. Has a “split” column with a label suggesting how to split the dataset into train, test and dev, stratified by the text category**

10.xls – Dataset of Portuguese Words (https://link.springer.com/article/10.3758/s13428-011-0131-7#SecESM1). **1034 words. No Index Column. Column “EP-Word” with the Portuguese word. Column “Val-M” with the Valence. Column “Arou-M” with the Arousal.**

11.xslx – Dataset of Portuguese Sentences (https://link.springer.com/article/10.3758%2Fs13428-016-0726-0#Sec13). **192 sentences. Crescent Valence and Arousal scores from 1 to 9. No Index Column. Column “Sentence” with the Portuguese sentence. Column “Valence” with the Valence. Column “Arousal” with the Arousal.**

12.xlsx – Dataset of French Words (https://link.springer.com/article/10.3758%2Fs13428-013-0431-1#Sec14). **1031 words. Crescent Valence and Arousal scores from 1 to 9. No Index Column. Column “French Words” with the French word. Column “valence Mean” with the Valence. Column “arousal Mean” with the Arousal.**

13.csv - Dataset of English Words (https://link.springer.com/article/10.3758/s13428-018-1099-3#Sec13). **5553 words. Crescent Valence and Arousal scores from 1 to 9. No Index Column. Column “Words” with the English word. Column “VAL M” with the Valence. Column “AROU M” with the Arousal. PROBLEM WITH REPEAT WORDS WITH A MEANING IN FRONT BETWEEN ().**

14.txt - Dataset of German Words, nouns (https://link.springer.com/article/10.3758/s13428-018-1099-3#Sec13). **1000 words. Crescent Valence and Arousal scores from 1 to 9. No Index Column. Column “word” with the German word. Column “valence\_mean” with the Valence. Column “arousal\_mean” with the Arousal. Tab as separator (?). (**ISO-8859 text, with CR line terminators”)

15.xlsx - Dataset of Italian Words (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169472#sec015). **1121 words. Crescent Valence and Arousal scores from 1 to 9. No Index Column. Column “Ita\_Word” with the Italian word. Column “M\_Val” with the Valence. Column “M\_Aro” with the Arousal.**