

This project is to be an INDIVIDUAL project! Anyone caught collaborating with anyone - student or non-student - will receive a zero for both the project and the presentation. You can accomplish this project using methods you have learned in class, especially in activities and homework. Think of this project as a larger activity - except you must work alone!


You must submit the following to Canvas:

- an R Notebook (nb.html) document

Also, make sure you have completed the accompanying

The final project analysis should be an HTML produced by R Notebook with embedded Rmd.

Additional stop words

- As stated in the Data Wrangling section of the rubric there are some other 'stop words' you will need to filter out. This is the [file](#)  of those eleven (11) words

Some suggestions

- When you go to add the lyrics from genius to the songs this may take some time - there are 200 songs that need to be "analyzed". Two options are to attach lyrics per decade (i.e. scrape 1980s table, attach lyrics, repeat) then combine these four tables, or combine the four tables then attach the lyrics.
- For Graph 3, Top Ten Words by Decade, you will need to install the gridExtra package (pay attention to spelling and capital 'E') to use the grid.arrange function to combine the four decade graphs. To see how to do this Google 'grid.arrange to combine several plots into one' and open link to cran.r-project.org
- For Graph 6, to add a single regression line when you are using a color to create point (scatterplot) graph, Google 'fit single regression line to ggplot with a color' and open link to stackoverflow.com
- I would encourage you that after you create the data frame of all the records and lyrics that you write this file to your working directory. This way you can import this data file to continue with project without having to continually scrape and organize. HOWEVER, your project file MUST include the scraping statements and how you manipulated the data. In your project file do NOT begin with reading in the data set! You can write the file to your working directory as a text file by the following (this code my data frame is lyrics and I am writing to my working directory as a text file also name lyrics):
`write.table(lyrics, "lyrics.txt", sep = ' ', col.names = T, row.names = FALSE)`

See grading rubric for additional details in Submit Final Project

Please submit just a link to your Zoom presentation. This is done (1) as a precaution in the event the link to your Zoom presentation in your Project notebook file does not work, and (2) to allow me to assign a separate grade for your presentation. Guidelines for the presentation are these:

- The presentation should be a minimum of 3 minutes and maximum of 4 minutes.
- At beginning of presentation, you **MUST** include yourself in the video holding (visible for me to see) your PSU ID with photo and clearly state your name and major or intended major. You are welcome to hide our 9-digit number (e.g. put small piece of paper over it). The purpose of this step is to verify the person doing the presentation. If there is a noticeable change in voice (i.e. suggesting someone else is now providing the presentation) you will receive a zero for **BOTH** the project and the presentation.
- The presentation itself should just provide a screen shot of each of the project graphs and your interpretation of them.
- The last part of video (e.g. the last slide if using PowerPoint) will include your thoughts on the project – what you found most challenging and at least two “things” you learned by doing this project. This can be anything related to R coding, statistics, or about yourself.