**Prerequisites**

This will be an interactive workshop introducing a natural language processing / machine learning workflow using R. We will specifically use the tidymodels framework for the analyses. We will be analyzing text data from medical examiners across the United States. We will develop a model to classify overdose deaths by type of substance involved.

If you want to follow along with the code, please do the following before the workshop (will speed things up the day of):

1. Download and install R: https://cran.rstudio.com/
2. Downlaod and install RStudio (free version): https://www.rstudio.com/products/rstudio/download/
3. Clone this repository / or download the .rmd file and the /Data/ folder
4. Open RStudio
5. Open the .rmd file
6. Run the first chunk to install the required packages or following code:

install.packages("pacman")

library(pacman)

p\_load(tidyverse, tidylog, purrr, forcats, colorspace, gtsummary, flextable, tictoc) #Basic

p\_load(textrecipes, tidytext, wordcloud, stopwords, textdata)

p\_load(tidymodels, parsnip, discrim, naivebayes, ranger, xgboost, kknn, keras, workflowsets, themis, stacks, vip) #Models

p\_load(parallel, doMC)

If this feels to onerous, you can also follow me along as (I struggle) we go through the code the day of the workshop.

Thanks for taking the time to check out the repository and being present at the workshop.

David