* [Text mining infrastucture in R](http://www.jstatsoft.org/v25/i05/" \t "_blank)
* [CRAN Task View: Natural Language Processing](http://cran.r-project.org/web/views/NaturalLanguageProcessing.html)
* [Videos](https://www.youtube.com/user/OpenCourseOnline/search?query=NLP) and [Slides](https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html) from Stanford Natural Language Processing cour
* [Natural language processing Wikipedia page](https://en.wikipedia.org/wiki/Natural_language_processing)

## Course Tasks

This course will be separated into 8 different tasks that cover the range of activities encountered by a practicing data scientist. They mirror many of the skills you have developed in the data science specialization. The tasks are:

1. Understanding the problem
   * Obtain data from course website
   * Familiarize with background
2. Data acquisition and cleaning
   * Tokenization (punctuation, digits, typos)

 con <- file("en\_US.twitter.txt", "r")

readLines(con, 1) ## Read the first line of text

readLines(con, 1) ## Read the next line of text

readLines(con, 5) ## Read in the next 5 lines of text

close(con) ##  Important to close the connection

* + Profanity cleaning

1. Exploratory analysis
2. Statistical modeling
3. Predictive modeling
4. Creative exploration
5. Creating a data product
6. Creating a short slide deck pitching your product

# Quiz 1 Notes

Twitter:

[1] maxLength 213

[1] numLines 2360148

[1] hasHate 22138

[1] hasLove 90956

[1] hasPhrase 3

News:

[1] maxLength 5760

[1] numLines 77259

[1] hasHate 322

[1] hasLove 1105

Blogs:

[1] maxLength 40835

[1] numLines 899288

[1] hasHate 11098

[1] hasLove 49167