

Advanced Mongo Assignment Mongo

The 2002 Enron fraud case uncovered financial deception in the world's largest energy trading company and at the time, triggered the largest US bankruptcy and its most massive audit failure. This data was originally made public, and posted to the web, by the Federal Energy Regulatory Commission during its investigation. As such, it has become a standard for data science and text mining.

enron.email2 contains all the emails from CEO Kenneth Lay (including those sent by secretary Rosalie Fleming), many of the emails from Chief of Staff Steven Kean, and all the emails from Senior Specialist Phillip Platter. Your job is to analyze this dataset using advanced Mongo querying techniques and the R language with ggplot.

Read the documents describing the Mongo dataset:

- [Background](#) on Enron and the data
- [Information](#) about the dataset.
- Its use in [data science](#).

Using the email2 document store, **formulate some kind of theory about the data**. Use advanced Mongo query techniques to do the analysis. Your mongo queries should be contained within a single javascript file that can be run inside the Mongo Shell. Your queries need to include the following:

- A print statement to print your theory and what queries you are using to support the theory
- Use of Mongo regular expressions
- Mongo aggregation
- Map-Reduce (When you map-reduce, you'll build new collections in Mongo. Please name them as <username>_<title>, such as myersjac_layStatistics).

Once the queries work the way you expect, move all of them (except the Map-Reduce ones) to R. Make at least two ggplots (a bar graph and a scatterplot with a trendline). At least one scatterplot should show different groups.