Notes from Discussion regarding Enron Corporation email corpus.
-Monte Carlo Pi solution with Map-Reduce.
-Map-Reduce to collect stop word frequencies.
-So group email messages by sender (sorting).
-Use a mail parser (some are built-in).

Interesting Areas of Research:

- Tone changes by recipient
- Timestamps; was the person in the office when it was sent, or are emails different after a certain time?
- Can determine gender?
    - Simple comparisons against the names, then run the stats see how they match up
- Languages if they are different than the base language tend to really jump out.
- How often there is ghost writing, do the outliers match someone else?

Partition the corpus to run in parallel. We should be able to parse it in parallel and then run the statistics in parallel.

Independent Component Analysis, can we parallelize it. Prof Adali.