Links

- https://research.neustar.biz/2014/09/08/differential-privacy-the-basics/
- https://www.microsoft.com/en-us/research/publication/a-firm-foundation-for-private-data-analysis
- http://www.win-vector.com/blog/2015/10/a-simpler-explanation-of-differential-privacy/

General Notes on DP

- Laplace mechanism for non-categorical data, exponential mechanism for categorical data.
  - Laplace(μ, b) where μ is the centre point, and b is the scale
  - Laplace distribution is like two Exp distributions stuck back-to-back at centre μ
  - VARIATE: A Laplace$(0, b)$ variate can also be generated as the difference of two i.i.d. Exponential$(1/b)$ random variables.
- Why can't you do a bunch of queries and take the average? Total privacy budget ε_{total}
- "63% of people in the US can be uniquely identified by just their birth date, zip code and gender."

Experiment Notes

- Need to find suitable data. Needs to be accurate, not pre-randomized. Needs to be large (so statistical analysis is useful). (Census data?)
- Implement the following SQL statements/keywords:
  - `SELECT` allowing selection of the following. (Not sure if all of these are necessary or feasible. Prioritised by importance.)
    - `SUM(col)`
    - `AVG(col)`
    - `MEDIAN(col)`
    - `COUNT[(col)]`
    - `MAX(col)`
    - `MIN(col)`
  - `FROM`
  - `WHERE`

Report Ideas

- Have cool graphs like this: http://content.research.neustar.biz/blog/differential-privacy/DensityWidget.html

Proposed Outline

- Introduction to DP
  - previous attempts to anonymize data and problems with them
  - DP's contributions
- DP Methods
  - Laplace
  - Exponential
- Our Experiment
  - Describe Dataset
  - Describe Database and statistical data

  - Describe targeted queries
- Our DP methods
  - Laplace
  - ???
- Results
  - Show that targeted queries are more difficult when DP is used
  - Show that DP does not adversely affect the statistical accuracy of the queries
  - Graphs
- Conclusion