# DSSL Reading Group: Differential Privacy

## Dwork 2009:

1. What is differential privacy?
    a. Definition is inherently symmetric, so bound is above and "below"
    b. What are the attack vectors this is protecting against?
    c. Function sensitivity is critical here:
        - Key: everything is a histogram
    d. Alex's crazy idea: just tell people you're adding noise, but don't do anything. How does this help us to understand the benefits / drawbacks of DP?
        - It obviously fails if K is public, but what if K could remain private? If people think there is noise that there isn't, does that still guarantee privacy?
2. What does or does not DP guarantee?
    a. How does the guarantee in Definition 1 relate to intuitive notions of privacy?
    b. Remark 2 on page 3?
3. Laplacian noise example/theorem
4. Examples from Dwork 2009
    a. Boolean vectors
        i. Complicated process just to get a conditional probability
        ii. "The number t is rather large": There's something odd here. With DP we limit the amount of information released in a query, and this forces us to make many queries to get useful information. So are our aims actually compatible?
        iii. Can we use a query that directly asks the conditional probability question? (or is it difficult to determine the sensitivity?).
            - E.g. two "count" queries: count (A and B) and count(A)
    b. Contingency table release
        i. Add noise to the "full" contingency table
            1. Non-integer
            2. Possibly negative
            3. Consistency in projections is immediate
            4. Very low signal to noise ratio
        ii. Add noise to each sub-table independently
            1. Possible inconsistencies between tables
            2. Type mismatch (for method i as well) with downstream applications
        iii. Fourier Domain
            1. Fourier transform of 2^k binary vectors
            2. Fewer noise terms needed in some cases
        iv. Linear Programming & Rounding:
            1. Create a whole new dataset with marginals pretty close to the true marginals.
    c. Learning Halfspaces

## Papers

Dwork 2009
Ruggles 2019: DP and the Census (interesting arguments both for and against the use of DP in the Census, non-technical)
Dong 2020: Gaussian DP