Anthony Rentsch and I are still planning on working together (as of this submission, we have not had any sort of falling out yet). We have pivoted our final project idea after some discussions and conversations with Jayshree and Dr. Honaker. We would like to explore topics related to differentially-private algorithm design.

Given that the 2020 Census will be releasing differentially-private stats, one of our interests is investigating how this will affect downstream uses of Census data. Many institutions use Census data as a benchmark for the calibration of their survey data for national/state estimates [1]. In particular, we are interested in multilevel regression and poststratification [2] and survey weighting [3]. We would be interested in performing experiments to see how these non-private algorithms would behave when we add differential privacy into historical Census statistics and then how they would behave after we apply our own private algorithm. One potential dataset we could use for this is the Cooperative Congressional Election Study [4].

Our other idea is to do another type of differentially private algorithm design not necessarily related to Census data but still appropriate for use in the social/life sciences. We would like to develop a ML model (possibly a linear regression model since we have discussed this in class) that would provide a high level of privacy and utility. As evident from this paper by Wang [5], the implementation of differentially private regression, or other ML models, is still being experimented with. We would like to apply this to some category in the Census data, but are open to suggestions.

We are having trouble gauging how to scope this - we aren't sure how ambitious designing a differentially private algorithm is. We would appreciate some guidance in developing our scope. Additionally, we think these ideas may be interesting to pursue but we would definitely be open pursuing a different direction in algorithm design, especially if it has a ML component. If there are ideas/directions that either of you have that might make for a solid project, we'd be open to that!

[1] https://clintonwhitehouse3.archives.gov/WH/EOP/CEA/html/censusreview.html
[2] http://www.princeton.edu/~jkastell/MRP_primer/mrp_primer.pdf
[3] http://www.pewresearch.org/methods/2018/01/26/how-different-weighting-methods-work/
[4] https://cces.gov.harvard.edu/
[5] https://arxiv.org/abs/1803.02596