

Data Sharing Plan

No original data will be created in the course of the proposed work. Rather, the tangible products of the proposed work include:

- Computer source code
- Summaries, samples, and transformation of existing data from the Census Bureau
- Summaries, samples, and transformation of existing data from other sources that are publicly available
- Articles to be submitted for publication in refereed journals
- Tutorials, descriptions of workflows, and other documentation to be provided to the Census Bureau

We present our plans for the dissemination of each of these products below.

Computer source code

The PIs are committed to research reproducibility. Co-PI Gagnon-Bartsch in particular has recently given presentations and tutorials on tools for reproducibility and reproducible workflows. The PIs are also committed to making their statistical analyses publicly available, and easily accessible so as to allow others to inspect the underlying code as well as the robustness of the findings. As an example, a recently published article jointly authored by all three PIs (Conrad, et al., 2019), <https://doi.org/10.1177/0894439319875692> explored the impact of varying analytic decisions in analyses of Twitter posts. To accompany the manuscript, the PIs made publicly available an interactive web application that allows users to reproduce the analyses in the article as well as to explore in even greater detail the impact of alternative analytic decisions. The web application is available at:

<https://robynferg.shinyapps.io/Twitter-ICS/>

All computer codes created in the course of the proposed work will be hosted and made publicly available on Github. Examples of previous work of the PIs are available at:

https://github.com/robynferg/Twitter_ICS

https://github.com/robynferg/Tracking_Presidential_Approval_with_Twitter

<https://github.com/robynferg/Population-Auxiliary-Data-and-Sample-Data>

In addition, to further ensure reproducibility and archival permanence of the analyses, Docker images of the analyses (which include not only the code but also the computational environment needed to run the code) will be made available on Docker Hub. Examples of previous work of the PIs are available at:

<https://hub.docker.com/r/johanngb/twitter>

<https://hub.docker.com/r/johanngb/twitter-presidential>

Note that the first example reproduces the analyses in Conrad, et al. (2019), and in particular allows one to run the interactive web application locally on one's own computer, thereby serving as a long-term archive of this interactive analysis.

Products derived from Census Bureau data

The proposed activities include analysis of Census Bureau data, including multiple tracking surveys. This may involve new data products that aggregate or summarize the original data in new ways. If the original Census Bureau data is already publicly available, we will also make any products we derive from the original data publicly available as well, via Github and/or Docker Hub as described above. If the original Census Bureau data is not publicly available (e.g. survey microdata or Sprinklr data), we will consult with our Census Bureau colleagues to determine if and when it is appropriate to release any derived products.

Products derived from publicly available data

The proposed activities include analysis of publicly available data, including data from social media such as Twitter and Reddit. Our analyses may involve new data products that aggregate or summarize this publicly available data, e.g., computing the number of tweets per day referencing the Census. We will make any products we derive from this social media data publicly available, via Github and/or Docker Hub as described above, but only to the extent that it does not unduly violate the privacy of social media users. In particular, although tweets are publicly available, we will not redistribute datasets that include the twitter handles of individual tweets. This is consistent with our privacy policy in previous work. See, for example, <https://hub.docker.com/r/johanngb/twitter-presidential>

Journal articles

In order to rapidly disseminate significant findings, draft manuscripts will be made available on arXiv (<https://arxiv.org/>) or other preprint servers, and / or the PIs' personal web sites. The manuscripts will then be promptly submitted for publication in peer-reviewed journals.

Tutorials and other documentation

The proposed activities include developing tutorials and example work flows for our Census Bureau colleagues. These documents will be shared with the Census Bureau in a mutually agreed upon format, and will only be made publicly available if requested by the Census Bureau.