

---

# **Doctoral Dissertation Research: Voting and the behavioral economics of housing in an affordability crisis**

*A Data management plan created using the DMPTool*

Creator(s): Michael Hankinson

Affiliation: Harvard University

Last modified: August 17, 2015

Copyright information: The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creators as the source of the language used, but using any of their plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.



# Doctoral Dissertation Research: Voting and the behavioral economics of housing in an affordability crisis

---

## **Roles and responsibilities**

Co-PI Michael Hankinson will be responsible for all data management, monitoring, and custodianship in perpetuity once the exit poll surveys are collected.

## **Expected data**

Existing data used for this project will consist of publicly available US Census data, the San Francisco voter file, and the San Francisco tax assessor's database. Each document is freely available through the City of San Francisco government. Data from the voter file and tax assessor's database will be merged with the new data collected through the exit poll surveys at the individual level. This merged csv file will then have unique identifiers removed before being publicly shared.

## **Period of data retention**

Our team will hold access to the data for a two month period, allowing for the data to be merged, cleaned, and de-identified prior to sharing. Afterwards, the data, metadata, and relevant R code will be freely available. There are no expected embargo periods for political/commercial/patent or publisher reasons.

## **Data format and dissemination**

The exit surveys will be coded and merged into a csv file for ease of analysis. Metadata of analytical steps and question wording will be shared in pdf format for ease of presentation. R code will be annotated for those seeking to replicate the analysis. All files will be made publicly available within 2 months by posting to the Harvard Dataverse Network. No permission restrictions will be placed on the data. Data will be deidentified to preserve anonymity. There are no foreseeable ethical or privacy issues. The final data management process will be vetted by the Harvard IRB.

## **Data storage and preservation of access**

Metadata will include the full research plan and exit survey wording for comprehension of responses. The metadata will also include a walkthrough of the annotated R code to facilitate replication of the analysis.

## **Additional possible data management requirements**