
Blockwell Consulting, LLC

Megan A. Gall, PhD, GISP

www.blockwellconsulting.com

Montana Prisoner Reallocation Project

Mid-Project Report

April 14, 2022

SUMMARY

The Prisoner Reallocation Project has seven specific milestones. This is a mid-project report detailing the preliminary findings and project status for Milestones 1 through 4, representing work accomplished from April 1 to April 14.

DATA CLEANING

Prisoner address data came to me in 4 files. First, I verified variables across datasets and then merged them into a single dataset. Second, I cleaned the address variable for geocoding. I standardized addresses by doing things like stripping all punctuation and turning all addresses to lowercase. I then parsed the address field into four separate fields for street address, city, state, and zip code.

Not all addresses were appropriate to geocode. Many addresses noted transience or locations of prior correctional facilities. Several addresses were for PO Boxes, which cannot be geocoded. Some addresses were in states outside of Montana. I created a new 'status' variable to log how many address records were in these categories. Out of 2,840 total records, 1,505 were not appropriate for reallocation for the following reasons: 1,389 records had no address, 56 records were outside of Montana, 28 records were addresses for jails or other similar facilities, 20 records noted transience, and 12 addresses were PO boxes.

In total, there were 1,335 records with addresses appropriate for geocoding. All records with addresses were complete in that a street address, city, state, and zip code was recorded for each. As found in the geocoding, some addresses are invalid, but there were no cases in which we had only partial address elements.

Data cleaning took 7 hours to accomplish.

GEOCODING

After cleaning the data, 1,335 addresses were appropriate for geocoding. I used an R package called *tidygeocoder* to accomplish the first round of geocoding¹. There are many ways to geocode addresses, including the ability to use addresses to return census block FIPS codes or latitude/longitude coordinates. The *tidygeocoder* package uses the US Census API to look up street addresses and return the block level FIPS code for the corresponding census block. Of the 1,335 addresses, 908 (68%) geocoded to the corresponding Census block with *tidygeocoder* and 427 (32%) did not.

I examined the 427 records that did not geocode and could not ascertain certain patterns or specific reasons for geocode failure. I investigated several individual addresses picked at random to verify authenticity through Google maps. Then I geocoded the remaining 427 addresses with the R package *gmapsdistance*, which relies on the Google API to return the latitude/longitude coordinates. Batch processing on latitude/longitude coordinates is not possible in *tidygeocoder* or the Census API, so an extra step was required to assign the block. I used QGIS² software to create a shapefile out of latitude/longitude coordinates. I then created a new variable for the shapefile to 'tag' with the block FIPS code.

This process resulted in two final tables. One table contains all records that had an address inside of Montana. There is a flag for the geocode status detailing whether the record was geocoded by *tidygeocoder* or *gmapsdistance* or did not geocode. In total, 1,256 (93.7%) of addresses successfully geocoded and 79 (6.3%) did not. A geocoding rate of 93.7% is a successful threshold given the highly transient nature of this population. Of the addresses that successfully geocoded, 908 were geocoded with *tidygeocoder* and 348 were geocoded with *gmapsdistance*.

Geocoding took 10 hours to accomplish.

REALLOCATION

I began the reallocation with the final geocoded data. First, I created sets of crosstabs including race/block and Hispanic/block. I converted the tables from long format to wide format and then joined the crosstab tables to the PL 94-171 tables. I then reallocated prisoners, by demographic group and total population, back to the home Census block by adding the relevant columns together. In the final deliverable file, original PL 94-171 data are preserved and new reallocation fields are appended.

I repeated the process for prisons. First, using the full dataset of prisoners, I created a set of crosstabs including race/facility and hispanic/facility. I converted the tables from long format to wide format and

¹ My original proposal stated I would use the *censusxy* package for this work. However, *censusxy* did not function as designed. The *tidygeocoder* package has the same capabilities as *censusxy* but is more developed and supported. In the end, the different R packages accomplish the same goal.

² [QGIS](https://qgis.org/en/site/) is the predominate open-source geographic information system (GIS): <https://qgis.org/en/site/>

then joined the crosstab tables to the PL 94-171 tables. I then reallocated prisoners, by race and total population, away from the prison facility blocks.

Although the reallocation is done, not all the diagnostic checks and balances on the final dataset are complete. The final report will include more specific reporting on reallocation once the final checks are complete.

Reallocation took a total of 3 hours to accomplish.

MID PROJECT REPORT

Milestone 4 is complete with the delivery of this mid-project report.

The mid-project report took a total of 4 hours to accomplish.

SUMMARY

This mid-project report details methods, complications, results to date, and time spent on each task. Data cleaning, geocoding, and reallocation are complete except for some additional data checks and accuracy protocols.

Overall, the project has required more time than estimated due to unforeseen issues with *censusxy* and additional geocoding steps. However, the project is still estimated to be completed on time.