

## **Blockwell Consulting, LLC**

Megan A. Gall, PhD, GISP  
www.blockwellconsulting.com

# Montana Prisoner Reallocation Final Project Report April 22, 2022

## **SUMMARY**

As Principal at Blockwell Consulting, LLC and at the request of the Montana Legislative Services Division (LSD) on behalf of the Montana Districting and Apportionment Commission, I conducted a data cleaning and reallocation project between April 1 to April 22, 2022. The commission seeks to adjust the U.S. Census Bureau 2020 Public Law 94-171 redistricting data (P.L. 94-171 data) to assign inmate addresses to their last known residence whenever possible. Data reallocation must be accomplished prior to completing redistricting map recommendations for the Montana Legislature in January 2023. Aside from product support (as needed), work contracted to Blockwell Consulting, LLC for the Prisoner Reallocation Project is complete.

## **DATA CLEANING**

I performed address cleaning and standardization of residential inmate address data provided to me for geocoding, including correction of initially unusable addresses.

Prisoner address data was delivered to me in four datasets. 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 P.O. Boxes and not included for geocoding. (P.O. Boxes are generally located inside United States Postal Service facilities and do not represent residential locations. Additionally, most reference datasets for geocoding do not include P.O. Boxes.) 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 similar facilities, 20 records noted transience, and 12 addresses were P.O. 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 records with partial address elements.

Data cleaning took seven hours to accomplish<sup>1</sup>.

## GEOCODING

After cleaning the data, 1,335 addresses were appropriate for geocoding. There are many ways to geocode addresses, including the ability to use addresses to return census block FIPS (Federal Information Processing Series) codes or latitude/longitude coordinates. I used an R package called *tidygeocoder* to assign each prisoner address to a Census block<sup>2</sup>. The *tidygeocoder* package uses the U.S. Census application programming interface (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.

Next, I attempted to correct addresses which could not initially be geocoded. 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 QGIS<sup>3</sup> 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. In total, 1,286 (96.3%) addresses successfully geocoded and 49 (3.7%) did not<sup>4</sup>. A geocoding rate of 96.3% is a successful geocode in most situations but is particularly true for this project given the often-transient nature of this population. Of the addresses that successfully geocoded, 908 were geocoded with *tidygeocoder* and 378 were geocoded with *gmapsdistance*.

The 49 addresses which could not be geocoded were additionally excluded from the reallocation process. These records are provided in a Microsoft Excel file called *MT\_PrisonerFailedGeocodes.xls*.

---

<sup>1</sup> The original scope of work for this project included approximately 1,132 prisoners. After the initial contract was negotiated, Montana requested the addition of approximately 1,708 individual addresses, adding to the project scope with a new total of 2,840. Additional time was required because cleaning addresses for geocoding is a more manual data process than most.

<sup>2</sup> 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.

<sup>3</sup> [QGIS](https://qgis.org/en/site/) is the predominate open-source geographic information system (GIS): <https://qgis.org/en/site/>

<sup>4</sup> This is an improvement from the numbers reported in the mid-project report. I was able to allocate additional, partially invalid addresses to the center of relevant jurisdictions.

---

Geocoding took 13 hours to accomplish.

## REALLOCATION

I began the reallocation of residential addresses to 2020 Census blocks using the cleaned and geocoded data. First, I created crosstabs including race against Census block and Hispanic against Census block. I converted the tables from long format to wide format and then joined the crosstab tables to the Public Law 94-171 (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.

I repeated the process for prisons. First, using the full dataset of prisoners, I created a set of crosstabs for race against facility and Hispanic against facility. 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 race and total population, away from the prison facility blocks.

There were some malalignments between the Census demographic data and the prisoner demographic data. First, the Census used separate categories to tabulate 'Asian' and 'Native Hawaiian and Other Pacific Islander' (NHPI) respondents. The prisoner demographic data combined them into a single category called 'Asian or Pacific Islander'. To accommodate this issue, I added the total Asian and NHPI population together for the state of Montana. Asian respondents represent 91% of that combined population. Seven total prisoners were logged as 'Asian or Pacific Islander' and because 91% of 7 people is 6.37, I rounded up and allocated all seven 'Asian or Pacific Islander' prisoners as 'Asian' for reallocation into the PL 94-171 data.

Second, data from the prisons showed two of the seven prisons<sup>5</sup> tallied a population higher than the population counted in the Census. More significantly than the total count, there is a malalignment between the race/ethnicity details reported by the prisons and those reported by the Census. Reallocation by race/ethnicity group away from Census blocks with prisons led to small negative population counts for those few blocks. There are several examples. For the associated block, the Montana State Prison tallied 462 white respondents while the Census tallied 355 white respondents. The Riverside Special Needs unit tallied 21 white respondents while the Census tallied 17. Finally, Montana Women's Prison tallied 85 American Indian or Alaskan Native respondents while the Census only tallied 80.

These errors are a result of malalignment between the prisoner demographics and Census demographics. It is not possible to ascertain the origin of these errors with the information currently available, but there are many possibilities. The Census relies on self-identification. I am not familiar with intake procedures at the Montana prisons under consideration, but demographic collection is likely a combination of self-identification and identification by personnel. Respondents may self-report race and ethnicity

---

<sup>5</sup> These are the Riverside Special Needs Unit and the Montana State Prison.

---

differently based on the circumstances, timing, or method of questioning. It is also possible that prisons used different addresses for respondents quartered in different areas. Finally, the Census data were collected on a precise date within the Census intake season and that may differ slightly from the precise date for which prisoner data were pulled. All these highly plausible scenarios may account for the data malalignment.

To solve this issue, I created an additional set of variables. New variable names retain the original Census conventions. Additional variable names are preceded with either an 'R' (e.g., RP0010001) or an 'S' (e.g., SP0010001). Variables preceded with 'R' represent those variables reallocated according to the original and maligned prisoner data. Variables preceded with 'S' represent those variables reallocated according to the original inmate and with the additional step of suppressing negative numbers to zero. I recommend the Montana Commission use the 'R' variable data for compliance with equal population redistricting principles and the 'S' variable data for compliance with the Voting Rights Act (VRA). Technically, this method undercounts inmates but is a functional solution around an intractable problem of conflict between primary data sources.

In the final deliverable file, original PL 94-171 data are preserved and new reallocation fields are appended.

Reallocation took a total of 10 hours to accomplish. Additional hours were incurred due to malignment between the primary datasets.

## **MID PROJECT REPORT**

The mid-project report was submitted on April 14, 2022.

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

## **CREATE MAPTITUDE FILES**

Once the PL 94-171 file was adjusted, I joined the tabular data to the spatial data to create a shapefile with geometry and demographic data. I then imported the shapefile into Maptitude and converted it into a Caliper Compact Geographic Database file, appropriate for redistricting activities in Maptitude for Redistricting software. The final Maptitude files are delivered in a zipped file, all named *Census2020\_PL94171\_Reallocation*.

Maptitude file creation took two hours to accomplish.

---

## DELIVER FINAL PRODUCT

I am delivering all final products with this report, including the adjusted PL 94-171 Maptitude geographic file, R script with project code, and a list of addresses that failed to geocode. Final files and names are:

- Maptitude Reallocation PL 94-171 files:
  - *Census2020\_PL94171\_Reallocation.zip*
    - *Census2020\_PL94171\_Reallocation.BIN*
    - *Census2020\_PL94171\_Reallocation.cdf*
    - *Census2020\_PL94171\_Reallocation.DCB*
    - *Census2020\_PL94171\_Reallocation.sty*
- File of addresses that did not geocode:
  - *MT\_PrisonerFailedGeocode.csv*
- R script for data processing:
  - *MT\_PrisonerReallocation2020.R*
- Final project report:
  - *MT\_PrisonerReallocationProject\_BlockwellConsulting\_FinalReport.pdf*

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

## PRODUCT SUPPORT

I am available for product support as needed between 4/23/2022 and 5/6/2022.

## DISCUSSION OF RESULTS

Of the 2,840 total inmates in Montana facilities during the 2020 census, 1,505 records did not have enough address information to geocode and reallocate. In total, there were 1,335 records with addresses appropriate for geocoding. Of those, 1,286 (96.3%) addresses successfully geocoded and 49 (3.7%) did not. After all adjustments were made, 1,076 blocks had prisoners reallocated to home addresses and seven blocks had prisoners reallocated away from prisons.

Reallocation and adjustment of the Public Law 94-171 data were successful, despite additions to the prisoner data sets, challenges cleaning and preparing data, and a reduced timeline.