Austin Datasets

Contents

[Notes 1](#_Toc499482675)

[Merge 1](#_Toc499482676)

[Socioeconomic Data 1](#_Toc499482677)

[Old 2](#_Toc499482678)

# Notes

11/26, Mika

## Merge

* SQL script: Merge.sql
* Database name: Austin.sqlite
  + 2 tables
    - AAnn2015 – the annual crime data from Austin\_Annual\_Crime\_Dataset\_2015.csv
    - AUF2015 – the Use of Force report data from Austin\_UF\_R2R\_2015.csv
* Merged dataset on council district
  + Dataset too big to post to git. Will need to run it yourself to generate.
  + If want to view what it will look like ahead of time without waiting two million years, look at “testAustinMerge\_CD.csv” or run that part of the Merge.sql script.
* Merging UF on *PrimaryKey*…Problem: the merge is keeping rows only where we have a match between the AAnn2015.*GOPrimaryKey* and the AUF2015.*PrimaryKey* (UoF incidents accounted for in overall crime dataset. Also, there’s 300-sum entries - quite a bit more than the 192 I got from Pandas...).
  + So, what we need to figure out is how to get all of the incidents to show up - for this query to not exclude them – while not creating duplicate rows for UF incidents that have a match in the annual crime set. This might be tangential to our new questions, though, which as I understand it are looking just at council district.

## Socioeconomic Data

* Located in <https://github.com/MikaArmenta/IPPP_Abbott_Lovins_Armenta/tree/master/Final/Austin> > Districts10\_Socioeconomics.csv

# Old

* + Primary Key
* 11% of UoF incidents are accounted for in the annual crime stats.
* 192 duplicate keys (Crime IDs) between Austin Annual Crime 2015 and Austin UoF 2015
  + Total Austin Annual Crime: 38,573
  + Total Austin UoF 2015: 1,681
* *DupKey* = dictionary of duplicate keys
  + Cleaning
* Removed all white spaces from indexes in both datasets
  + e.g. “Primary Key” now = “PrimaryKey”, “GP Primary Key” = “GOPrimaryKey”
* NaNs
  + *AAnn\_nan\_rows* and *AUF2015\_nan\_rows* = dataframe of rows with NaNs in them
* Drop rows with NaNs for *CouncilDistrict*