**STUFF FOR THE READ ME**

**Data Sets**

Crime Data Set: <https://catalog.data.gov/dataset/crime-data-from-2010-to-2019>

Dispensary Data Set: <https://cannabis.lacity.org/personal-activity/find-licensed-retailers>

Dispensary Data Set: [https://search.cannabis.ca.gov](https://search.cannabis.ca.gov/)

**ETL for Crime Data**

* Imported as CSV
* Stripped out all rows with no location coordinates
* Extracted a variable to just report year of occurrence
* Deleted variables that were not needed for analysis
* Deleted rows with crime descriptions that seemed to be domestic in nature, related to employee theft, moving violations, or clearly non-violent
* Grouped weapon used data into smaller categories
* Grouped premise type data into smaller categories
* Grouped crime type data into smaller categories
* Exported data into JSON files

**ETL for Dispensary Data**

* Deleted all non-commercial retail facilities such as delivery-only services and cultivation businesses if known
* Stripped out all rows with no location coordinates
* Manually examined data for locations with matching addresses or coordinates
* Could not find a solid single source. All sources had SOME info but not all, and all had contradictions.
* Contacted CA Bureau of Cannabis Control who sent a website that has the same issues – not everything checks out, etc.
* Resigned to do the best with what we have (this would be great for “future plans”).