**Dataset:** \_\_\_\_\_\_[Housing Code Violation](https://data.montgomerycountymd.gov/Consumer-Housing/Housing-Code-Violations/k9nj-z35d)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Comment on the quality of the data.

* very large dataset ~0.5M rows
* represents code violations found by Moco Dept of Housing & Community Affairs between 2013 and current date
* some rows appear doubtful: extremely quick resolution and inappropriate: (ie. air conditioning in January

1. What are the primary columns of interest?

* all date fields: filed, assigned, closed. Interesting to see inspector workload.
* locationDescription: part of property/dwelling that is in violation
* item: name of structure/appliance/utility in violation
* action: remedial work ordered
* disposition: ultimate action by HCA

1. What are the filters of relevance?

* eliminate NA’s or impute values in columns of interest
* parse and convert dates
* identify questionable rows, isolate or discard

1. Give potential visualization options.

* Sankey diagrams illustrating process flows and throughput.
* Plot workload (cases in process/count of inspectors) vs date
* Histograms of permutations and filters of (locationDesc.,item, action, disposition

1. Discuss any limitations of the dataset.

* Data appears to include non-violations
* Geolocation column contains many zeroed entries, needs verification.
* Rows in years before 2013 questionable.

1. Give ideas for how the dataset can be improved/expanded/deepened.

* Include de-identified inspectors assigned to cases
* Link to census tract.
* Map county code/section to concise description ie. 26(b)4a to exposed wiring.
* Link to MD State Parcel dataset for unit county, property owner, recent sales