**Case-study: 33\_US\_WaterQuality\_BottledWaterSales\_2006\_2019Data**

* US\_Water**Violations**\_Quality\_BottledWaterSales\_1982\_2015RegressionData.csv
* US\_**WaterQuality\_BottledWaterSales**\_2006\_2019RegressionData

**Summary**

Dataset allows examining the quality of drinking-water and sales of bottled water in US. Globally, in developed countries, 10-40 million people have been affected by water quality violations (1985-2019).

Weekly data of sales and violation records in 2,151 US counties (2006-2015) show that violations posing an immediate health risk (tier-1) are associated with a 14% increase in bottled water sales.

Greater averting action is taken against water contaminants that might pose a greater perceived health risk and that require more immediate public notification.

Rural, low-income communities do not take significant averting action for elevated levels of nitrate, but experience a higher prevalence of nitrate violations.

This study can inform improvements in public notification and targeting of technical assistance from state regulators and public health agencies in order to reduce community exposure to contaminants.

**Provenance**

* **Paper**: Allaire, M., Mackay, T., Zheng, S., Lall, U. (2019). Community Response to Impaired Water: Understanding Averting Behavior with Bottled Water Sales. Proceedings of the National Academy of Sciences. (2019-08-30): <https://doi.org/10.1073/pnas.1905385116>
* **Database:** Allaire, M., Mackay, T., Zheng, S., Lall, U. (2019). Community Response to Impaired Water: Understanding Averting Behavior with Bottled Water Sales. Proceedings of the National Academy of Sciences: <https://dataverse.harvard.edu/dataverse/allaire>

**Meta-Data**

See paper appendix for details:

<https://www.pnas.org/content/pnas/suppl/2019/09/24/1905385116.DCSupplemental/pnas.1905385116.sapp.pdf>.

* **Violations:**
	+ Water violation data comes from the EPA Safe Drinking Water Information System (SDWIS), 2006 to 2015. These data include the contaminant that triggered the violation, start and end date of the violation, and community water system (CWS) characteristics, including the number of people and counties served. For CWS that serve multiple counties, we split the population served between counties based on the share of combined population. For example, if a CWS serves two counties, where County A has twice the population of County B, then two-thirds of the service population is designated as County A. Few water systems (n=102) serve multiple counties.
	+ Tier 1 violation notices for noncompliance with total coliform and turbidity state that elevated levels have been found in the water supply and *recommend boiled or bottled water* *use*. In addition, public notices for a confirmed presence of E. coli will state that this indicates possible contamination from human or animal wastes, which can cause illness. Tier 1 notices for nitrate state that infants under six months old should not drink tap water.
	+ Tier 2 violation notices, *possible health effects* are only mentioned for some Pathogen violations. Notification for these violations state that there is an increased chance that disease-causing organisms are in the water supply and not all in-home filters can effectively remove parasites. Tier 2 violations of ‘Other’ chemical types mention increased health risks due to drinking tap water with contaminants in excess of the MCL over many years.

|  |
| --- |
| **Violations Features** |
| utility\_id |
| location\_state\_abbrev |
| year |
| All\_Violation\_Categories |
| Lag\_All\_Violation\_Categories |
| TotalColiform\_Violation |
| Lag\_TotalColiform\_Violation |
| private\_ownership |
| size\_medium |
| size\_large |
| water\_source\_purchased |
| water\_source\_surfacewater |
| ln\_median\_household\_income |
| nonwhite |
| ln\_housing\_density |
| hhi |

* **Bottled Water Sales**:
	+ Bottled water consumption uses weekly sales (Sunday to Saturday) from the Nielsen Retail Scanner dataset for weeks from January 1, 2006 to December 31, 2015. The Nielsen dataset includes grocery stores, drugstores, convenience stores, and mass merchandisers (e.g., club, Walmart, military, dollar stores). These stores are affiliated with about 90 participating retail chains and are estimated by Nielsen to represent over half of total sales volume of U.S. grocery and drug stores as well as more than 30% of mass merchandiser sales volume. It should be noted that conclusions drawn from these data are our own and Nielsen had no role in the analysis. Our assembled dataset includes sales in dollars for over 5,000 Universal Product Codes (UPCs) of bottled water, which include flavored and unflavored water, as well as different container sizes of the same product. Since the Nielsen dataset does not consistently provide information about volume of water sold, we define our dependent variable as the log of aggregate sales in dollars (constant $2015) for all UPCs by county and week. We exclude stores that report sales during fewer than 469 weeks, which represents 90% of the weeks in our study period.
* **Weather Data**
	+ Temperature and precipitation data were obtained from the NOAA National Centers for Environmental Prediction (NCEP). Daily average air temperature is from the NCEP North American Regional Reanalysis (NARR) dataset, which has a 32km resolution. Daily average precipitation was from the U.S. Climate Prediction Center (CPC) Gauge-Based Analysis of Global Daily Precipitation. This is a global dataset available at 0.5-degree resolution. Gridded climate data was converted to county-level information by calculating average values of cells within county boundaries. Number of rainy days in a given week and county was calculated based on the number of days that average precipitation was 1mm or greater.
* **Demographic Data**
	+ Annual, county-level Census characteristics include median household income, housing density, percent nonwhite population, and median year housing stock built. These demographic data were obtained from the decennial census in 2010 and the American Community Survey data in years of 2006, 2008, and 2012. Values in inter-census years were estimated using monotone piecewise cubic interpolation.

|  |
| --- |
| **Water Quality & Sales** |
| fips |
| year |
| week |
| nweeks |
| ln\_water\_sales |
| tier1\_violation |
| tier1\_BPV\_violation |
| tier1\_NN\_violation |
| tier2\_violation |
| tier2\_BPV\_violation |
| tier2\_other\_violation |
| temp\_mean |
| precip\_rainy\_days |
| ln\_deflated\_median\_household\_nc |
| ln\_housing\_density |
| nonwhite |
| ln\_median\_year\_structure\_built |
| deflated\_median\_household\_income |