ReadMe for "The Impact of Smoking Bans in Bars on Alcohol Consumption and Smoking" Anne M. Burton

In the zip file, I've included all Stata do-files to replicate all results, tables, and figures in the paper. I've also included all publicly available data.

Code

- big do.do runs all the code
 - You will need to change anne_base (and nielsen_base, if you have the Nielsen data) to the correct directory
 - If you have access to all datasets, these are the only two lines of code you will need to change to get everything to run on your machine
 - FYI, this code takes 5-6 days to run start to finish with 4-core Stata. It will run faster if you have more cores
 - o Note that if you do not have access to the Nielsen data, you will need to comment out any line of code with Nielsen in it for the code to run
- config.do creates filepaths and installs necessary packages
 - Note that if you do not have access to the Nielsen data, you will need to comment out the line that begins "global nielsen_base"
 - o If you do not have access to the Nielsen data, once you make the above changes to big_do.do and config.do you will be able to run big_do.do from start to finish
- build brfss individual data.do imports and cleans the raw BRFSS data
- build_smoking_bans.do imports the raw smoking bans and population data, cleans the data, and creates the treatment variable
- build controls do imports and cleans the raw alcohol and tobacco policy data
- build nielsen data.do imports and cleans the raw Nielsen data
- analysis_sd_brfss.do merges the BRFSS data with the treatment and control variables and creates additional variables to make the final analysis dataset, and runs other do-files that generate summary statistics and most of the regression analysis
 - o analysis_sd_brfss_sumstats.do creates the summary statistics tables for the BRFSS data and the map of smoking bans
 - o analysis_sd_brfss_individual_twfe.do runs the main BRFSS regressions and creates most of the estimates for the main tables
 - o analysis sd brfss event studies.do runs the BRFSS event studies
 - o analysis_sd_brfss_heterogeneous_twfe.do runs the BRFSS robustness checks and creates estimates for the heterogeneous effects and robustness checks tables
- analysis_sd_nielsen.do merges the Nielsen data with the treatment and control variables and creates additional variables to make the final analysis dataset, and runs other do-files that generate summary statistics and most of the regression analysis
 - o analysis_sd_nielsen_sumstats.do creates the summary statistics table for the Nielsen data
 - o analysis_sd_nielsen_house_twfe.do runs the main Nielsen regressions and creates most of the estimates for the main tables
 - o analysis_sd_nielsen_household_robust.do runs the Nielsen robustness checks and creates estimates for the heterogeneous effects and robustness checks tables
 - o analysis sd nielsen event studies.do runs the Nielsen event studies

- analysis_sd_brfss_individual_newdid.do formats the BRFSS data for the stacked and DiD imputation estimators and runs the regressions/event studies for those alternative estimators
- analysis_sd_nielsen_household_newdid.do formats the Nielsen data for the stacked and DiD imputation estimators and runs the regressions for those alternative estimators

Raw Data: description of raw data files with links where available

- brfss_raw_yyyy.XPT are the raw SAS files for each BRFSS wave downloaded from the CDC's BRFSS website
 - o link: https://www.cdc.gov/brfss/annual data/annual data.htm
- Not included: Nielsen Consumer Panel data 2004-2012.
 - To request access to the data, see
 https://www.chicagobooth.edu/research/kilts/research-data/nielseniq
- ansi_county_codes_2010.txt is a text file of U.S. county and state names and corresponding FIPS codes, which is used in the creation of the smoking bans data
 - o link: https://www2.census.gov/geo/docs/reference/codes/files/national-county.txt
- ansi_place-to-county_2010.txt is a text file of U.S. place names and place types linked to counties, which is used in the creation of the smoking bans data to match cities to counties
 - o link: https://www2.census.gov/geo/docs/reference/codes/files/national_places.txt
- census_incorporated_place_pop_est_2000_2009.csv is a CSV file of city population estimates for 2000-2009, which is used in the creation of the smoking bans data to generate the fraction of the county population subject to city-level smoking bans
 - o link: https://www2.census.gov/programs-surveys/popest/datasets/2000-2009/cities/totals/sub-est2009-ip.csv
- census_incorporated_place_pop_est_2010_2018.csv is a CSV file of city population estimates for 2010-2018, which is used in the creation of the smoking bans data to generate the fraction of the county population subject to city-level smoking bans
 - o link: https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html
- census_subcounty_pop_est_2000_2010.csv is a CSV file of subcounty population estimates for 2000-2010, which is used in the creation of the smoking bans data for population estimates for cities that span multiple counties
 - o link: https://www2.census.gov/programs-surveys/popest/datasets/2000-2010/intercensal/cities/
- census_subcounty_pop_est_2010_2018.csv is a CSV file of subcounty population estimates for 2010-2018, which is used in the creation of the smoking bans data for population estimates for cities that span multiple counties
 - o link: https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html
- co-est00int-tot.csv is a CSV file of county population estimates for 2000-2010, which is used in the creation of the smoking bans data to generate the fraction of the county population subject to smoking bans
 - o link: https://www2.census.gov/programs-surveys/popest/datasets/2000-2010/intercensal/county/

- co-est2017-alldata.csv is a CSV file of county population estimates for 2010-2017, which is used in the creation of the smoking bans data to generate the fraction of the county population subject to smoking bans
 - o link: https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/counties/asrh/
- export-data-xlsx--adult-operators-of-noncommercial-motor-vehicles_changes.xlsx is an Excel spreadsheet of alcohol policy changes over time, which is used for the BAC data
 - o link: https://alcoholpolicy.niaaa.nih.gov/apis-policy-topics/adult-operators-of-noncommercial-motor-vehicles/12/changes-over-time#page-content
- export-data-xlsx--adult-operators-of-noncommercial-motor-vehicles_2021.xlsx is an Excel spreadsheet of alcohol policies in 2021, which is used for the BAC data
 - o link: https://alcoholpolicy.niaaa.nih.gov/apis-policy-topics/adult-operators-of-noncommercial-motor-vehicles/12#page-content
- smoke_free_laws_2018_07_01.pdf is the PDF file of effective dates of smoking bans in bars and restaurants downloaded from the American Nonsmokers' Rights Foundation (I use the July 1, 2018 version but it has been updated since then)
 - o link: https://no-smoke.org/wp-content/uploads/pdf/EffectivePopulationList.pdf
- smoke_free_laws_1jul2018_copied.xlsx is an Excel spreadsheet of effective dates of smoking bans that I copied over from the pdf file to import into Stata
- state_fips_codes.xlsx is an Excel spreadsheet of state names, postal abbreviations, and FIPS codes used to help merge different datasets for the control variables
- tbot_vot51_1970_2016.xlsx is an Excel spreadsheet of the Tax Burden on Tobacco (TBOT) data, which is used for the cigarette tax data