1 Instructions

This README file contains instructions for how to run the programs in the folder "Prepare Data". The file "1_masterprep.R" will run them all. Below we describe what each individual file does.

1. **1_mergeSEDA_pollution.R**. This file takes as inputs

   - "Raw Data\1. Pollution Data\PM25_2008_2018_District.csv": cleaned pollution data
   - "Raw Data\6. SEDA DATA\seda_geodist_long_cs_4.1.csv": raw SEDA test score data
   - "Raw Data\6. SEDA DATA\seda_cov_geodist_long_4.1.csv": raw SEDA demographics data

   The file merges the pollution and school data. The output file is "Derived Data\df_avg_noinst.csv" and "Derived Data\df_bwg.csv".

2. **2_prepforbartik.R**. This file takes as inputs

   - "Raw Data\5. Electricity Data\plant_by_year_withloc.dta"
   - "Raw Data\5. Electricity Data\eia_retirementlist_2019.csv"
   - "Raw Data\5. Electricity Data\PlantY07.csv"

   The file combines all electricity plant information (location and production) into one file. The file also creates Figure A1 in the appendix. The program creates five output files in the derived folder that are used for later analysis:

   - "Derived Data\finalfuelagg.csv": total production in U.S., by year and fuel type
   - "Derived Data\dfcoal.csv" : coordinates of each plant in the U.S.
   - "Derived Data\dfcoalprod.csv" and "Derived Data/aggprod.csv" : production information for all plants in the U.S. with location coordinates
   - "Derived Data\distanalysisleaid.csv" production for each district across years

3. **3a_createbartik2005.R**. This file takes as inputs the files created from "2_prepforbartik.R" and calculates the distance between school districts and power plants. It also creates a dataset for the Bartik instrument with 2005 shares. The output files are:

   - "dfbartik2005_40km.csv"

- "dfbartik2005_60km.csv",

4. **3b_createbartik2001.R** This file takes as inputs the files created from "2_prepforbartik.R" and calculates the distance between school districts and power plants. It also calculates creates a dataset to prepare for the bartik instrument with 2001 shares. The output files are:

   - "dfbartik2001_40km.csv"

5. **4_preparebartikanalysis.do** This do-file takes the files created by **3a_createbartik2005.R** and **3b_createbartik2001.R** and computes the bartik instrument. It outputs three files for analysis:

   - dfbartik2005_40kmanalysis.csv
   - dfbartik2005_60kmanalysis.csv
   - dfbartik2001_40kmanalysis.csv