

**Dataset:** IPOcases\_mergepriceindex200110.csv

**Description:** Each row contains the information about a unique firm IPO. The variable **DealNumber** provides a unique identifier for the IPOs. The following variables will be helpful for this task: **DealNumber**, **filingdate**, **long**, **lat**, **GEOID10**, **NumberofEmployees**, **SharesFiledinthisMkt**, **PrincipalAmount\$mil**, **withdrawn\_dummy**<sup>1</sup>. You are welcome to use other variables as well.

**Deliverable:** Please do a brief write-up (2-3 paragraphs) that summarizes your results and findings for both task 1 & 2. Include any graphs or tables you would like to present.

**Task 1:** Plot the locations of the IPOs on the map of U.S. using either Python, R or Stata. Please delete cases located in Alaska and Hawaii. If you need to use a shape file for U.S., you can use the one I included: US\_state\_shapefile/gz\_2010\_us\_040\_00\_500k.shp

The map should also help us visualize the size of the IPOs, as measured by either **NumberofEmployees**, **SharesFiledinthisMkt**, or **PrincipalAmount\$mil**. This type of maps are called the Bubble Maps, and it can be done for example using the Plotly package in Python (<https://plot.ly/python/bubble-maps/>). You can make the size of the dots on the map proportional to the measure of size you choose. If that does not generate enough variation in the size of the bubbles, you can categorize the IPOs into groups, e.g., quartiles or quintiles of IPO size, and plot groups of IPOs on the map. Please include figure notes explaining the size of the bubble or include a legend for your IPO size groups.

For this task, you are welcome to draw maps separately for the successful and withdrawn case, though not necessary.

**Task 2:** In our subsequent analysis of the effect of IPOs on the local housing market, we are especially worried about **more than one big IPOs happening in the same place within a short period of time**. Your job is to help us understand better how often that happens by proposing and producing a few measures on your own (be creative!). For example, for each IPO in the data, does there exist another IPO that happened in the same county within say 1/2/3/4/5 years of its filing date (variable **filingdate**)? With your measures, we can then decide which set of IPOs we should focus on in our analysis by excluding the ones that had multiple IPOs nearby within a short period of time.

To obtain the county that contains each IPO, use the variable **GEOID10**<sup>2</sup>, which is the census tract ([https://transition.fcc.gov/form477/Geo/more\\_about\\_census\\_tracts.pdf](https://transition.fcc.gov/form477/Geo/more_about_census_tracts.pdf)) that contains the IPO. Extract the first 5 digits of the full 11-digit GEOID10 to obtain the 5-digit county fips code

---

<sup>1</sup>withdraw\_dummy == 1 means an IPO has withdrawn after its initial filing. In our econometric analysis, the successful ones are in the treatment group, while the withdrawn ones are in the control/comparison group.

<sup>2</sup> **GEOID10 should be casted as a string variable that has 11 digits. If there are fewer than 11 digits, you need to fill in leading zeros until it is 11 digits.**

([https://en.wikipedia.org/wiki/FIPS\\_county\\_code](https://en.wikipedia.org/wiki/FIPS_county_code)). Note that the first digit of the 5-digit county fips code could be a zero.

For measure of IPO size, you can use **NumberOfEmployees**, **SharesFiledinthisMkt**, and **PrincipalAmount\$mil**. Timing of an IPO is determined by **filingdate**.

For this task, again separating the successful and withdrawn cases potentially adds value to our analysis. For example, you can examine for each IPO firm, how many "successful" and "withdrawn" firms happened in the same county within a short period of time, respectively.