

# Son Bias in the US: Evidence from Business Names\*

Walter Guillioli and Gaurav Sood

02 February, 2020

Are American businesses with the word “son(s)” more common than the word “daughter(s)”?

 To answer the question, we assemble data on business names. In the US, businesses are registered with the state. And to allow businesses to check whether a particular name is taken, etc., each state provides a way to search for business names. But most states do not allow for sophisticated searches or return all the relevant search results. Because of these reasons, we can only estimate a very conservative lower bound. In all, based on data from 40 states, across states, the median ratio of businesses with the word ‘son(s)’ in their name to the word ‘daughter(s)’ is 12:1.

When we split by region, we find that the son to daughter ratio is generally higher in the Northeast and the Midwest compared to the South and the West. Analysis by political preferences leaves that democratic majority states have higher ratios than republican majority states. Lastly, counter-intuitively, correlation with state GDP is weak.

## Data and Methods

In the United States, businesses have to register with a state. Each state provides a website to search for business names. But the functionality of these websites varies a lot. Some states do not return all the search results. Some only allow for particular regex searches—returning all the names that include the search term. We describe the concerns in further detail below, including how we address the challenges:

### 1. Search results for son(s) are inflated.

- There are three reasons for that:
- Son is part of many English words, from names such as Jason and Robinson to ordinary English words like mason (which can also be a name).
  - Son is a Korean name.
  - Some businesses use the word “son” playfully. For instance, “son” is a homonym of sun, and some people use that to create names like “son of a beach.”

We address the first issue by cleaning the data using regular expressions, only looking for exact matches of son and sons. We don’t address the other two concerns because when we carefully browsed the search results, we found that such cases were not that uncommon.

### 2. Limited number of results returned.

Some states return all the relevant search results, but others limit the number of returned results. For example, Alabama only displays the “first” 1,000 businesses. The truncated returned set means we only know that there were more than 1,000 results. We do, however, know how many—it could be 1,001 or 500,000. This limitation only impacted search results of “son(s).” And it means that we can only provide a conservative estimate of the lower bound.

Outside of the two concerns we note above, there is one more concern. Having “son” in the name doesn’t preclude the existence of the word daughter. And vice versa. When we browsed the results, we found that this was fairly infrequent. So we chose to ignore this.

In all, we got data for 36 states—the remaining 14 were left because getting the data proved onerous. The 36 states for which we got the data represent 69.9% of the US population, 71.2% of the GDP and 71% of the registered businesses.

---

\*The data and scripts used are posted here: [https://github.com/soodoku/sonny\\_side](https://github.com/soodoku/sonny_side).

We enriched the dataset with some state level data. We added the census region (Northeast, West, Midwest, and South), population of the state, GDP, major political party of the state, and number of businesses in each state. The sources for each are in the references. Table 1 describes some of the key features in the data.

Table 1: Summary of the Data

| State          | Ratio | US Region | Population | GDP   | Establishments | Political Party |
|----------------|-------|-----------|------------|-------|----------------|-----------------|
| Alabama        | 7     | South     | 0.015      | 0.011 | 0.013          | Republican      |
| Alaska         | 11    | West      | 0.002      | 0.003 | 0.003          | Republican      |
| Arkansas       | 17    | South     | 0.009      | 0.006 | 0.008          | Republican      |
| California     | 24    | West      | 0.120      | 0.145 | 0.119          | Democratic      |
| Connecticut    | 20    | Northeast | 0.011      | 0.013 | 0.012          | Democratic      |
| Florida        | 4     | South     | 0.065      | 0.051 | 0.070          | Republican      |
| Georgia        | 12    | South     | 0.032      | 0.029 | 0.029          | Republican      |
| Hawaii         | 17    | West      | 0.004      | 0.004 | 0.004          | Democratic      |
| Idaho          | 2     | West      | 0.005      | 0.004 | 0.006          | Republican      |
| Illinois       | 48    | Midwest   | 0.039      | 0.042 | 0.041          | Democratic      |
| Indiana        | 25    | Midwest   | 0.021      | 0.018 | 0.019          | Republican      |
| Kansas         | 5     | Midwest   | 0.009      | 0.008 | 0.010          | Republican      |
| Kentucky       | 4     | South     | 0.014      | 0.010 | 0.012          | Republican      |
| Maryland       | 2     | South     | 0.018      | 0.020 | 0.018          | Democratic      |
| Massachusetts  | 41    | Northeast | 0.021      | 0.028 | 0.023          | Democratic      |
| Michigan       | 24    | Midwest   | 0.030      | 0.026 | 0.028          | Republican      |
| Minnesota      | 2     | Midwest   | 0.017      | 0.018 | 0.019          | Democratic      |
| Mississippi    | 12    | South     | 0.009      | 0.006 | 0.008          | Republican      |
| Montana        | 4     | West      | 0.003      | 0.002 | 0.005          | Republican      |
| Nevada         | 72    | West      | 0.009      | 0.008 | 0.008          | Democratic      |
| New Hampshire  | 27    | Northeast | 0.004      | 0.004 | 0.005          | Democratic      |
| New Jersey     | 2     | Northeast | 0.027      | 0.030 | 0.030          | Democratic      |
| New York       | 2     | Northeast | 0.059      | 0.082 | 0.070          | Democratic      |
| North Dakota   | 26    | Midwest   | 0.002      | 0.003 | 0.003          | Republican      |
| Ohio           | 26    | Midwest   | 0.036      | 0.033 | 0.033          | Republican      |
| Oregon         | 4     | West      | 0.013      | 0.012 | 0.015          | Democratic      |
| Rhode Island   | 17    | Northeast | 0.003      | 0.003 | 0.004          | Democratic      |
| South Carolina | 69    | South     | 0.016      | 0.011 | 0.014          | Republican      |
| South Dakota   | 12    | Midwest   | 0.003      | 0.003 | 0.003          | Republican      |
| Tennessee      | 2     | South     | 0.021      | 0.018 | 0.017          | Republican      |
| Utah           | 5     | West      | 0.010      | 0.009 | 0.010          | Republican      |
| Vermont        | 22    | Northeast | 0.002      | 0.002 | 0.003          | Democratic      |
| Washington     | 15    | West      | 0.023      | 0.028 | 0.024          | Democratic      |
| West Virginia  | 2     | South     | 0.005      | 0.004 | 0.005          | Republican      |
| Wisconsin      | 20    | Midwest   | 0.018      | 0.016 | 0.018          | Republican      |
| Wyoming        | 11    | West      | 0.002      | 0.002 | 0.003          | Republican      |

## Results

In all, as shown in Figure 1 we find that a conservative estimate of son to daughter ratio is between 2:1 to 72:1 across the 36 states where we have data with a median of 12:1.

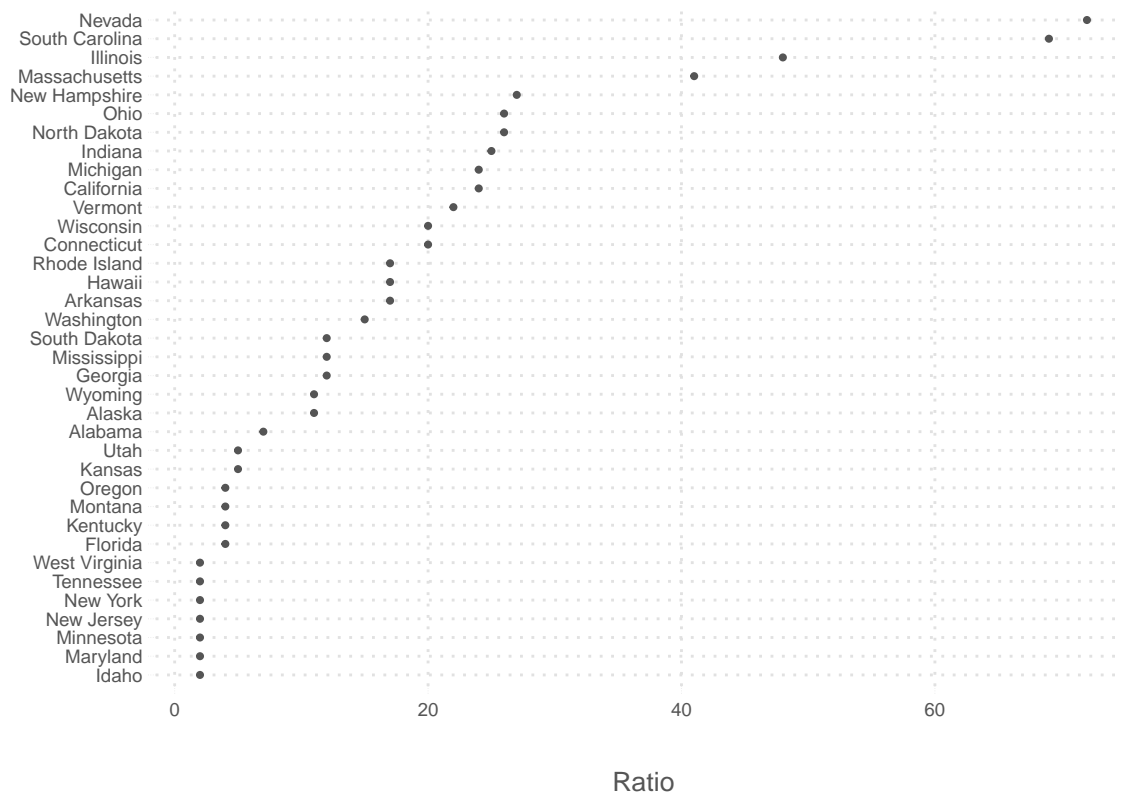


Figure 1: Ratio of business names with the word ‘son’ to ‘daughter’

### Covariates of Son Bias

We begin our investigation with whether the son to daughter ratio varies by population of the state or number of businesses in the state. We do this as a way to assess the generalizability of our result. The upshot is that the ratio doesn’t varies in any systematic way by state population (see Figure 2) or by number of establishments—a proxy of total number of businesses in a state (see Figure 3).

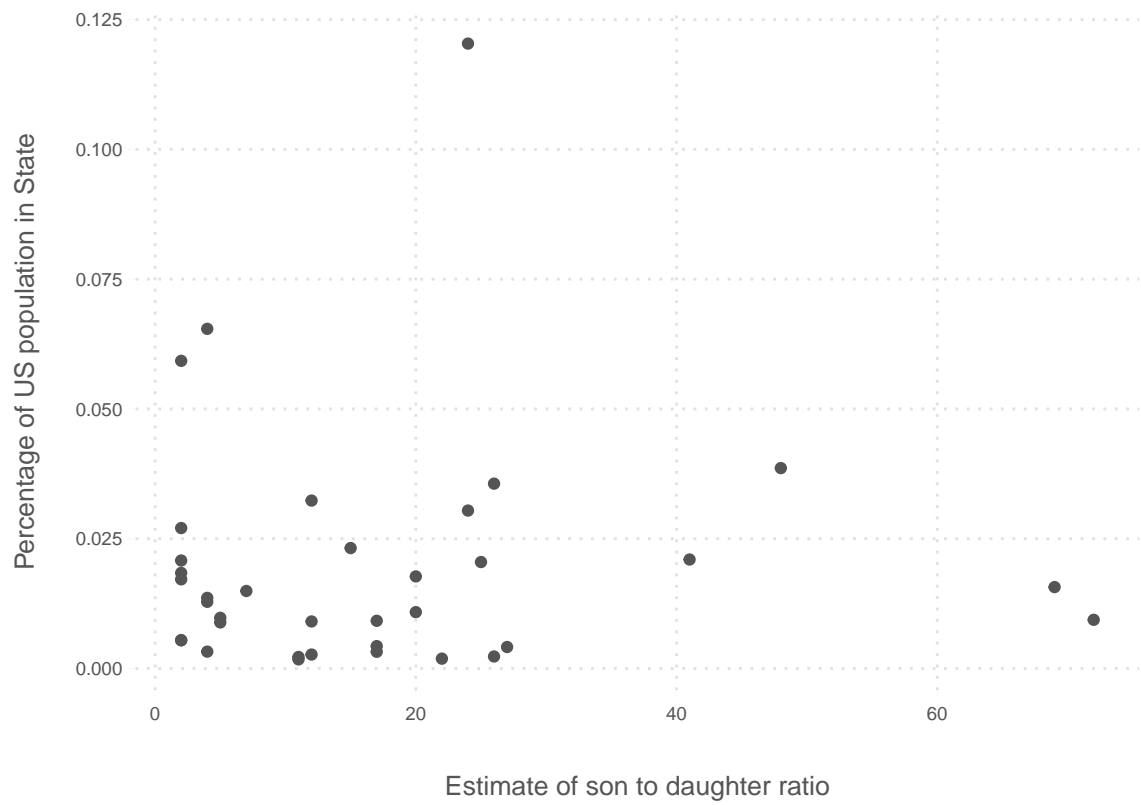


Figure 2: Estimate of son to daughter ratio vs % of US population per state

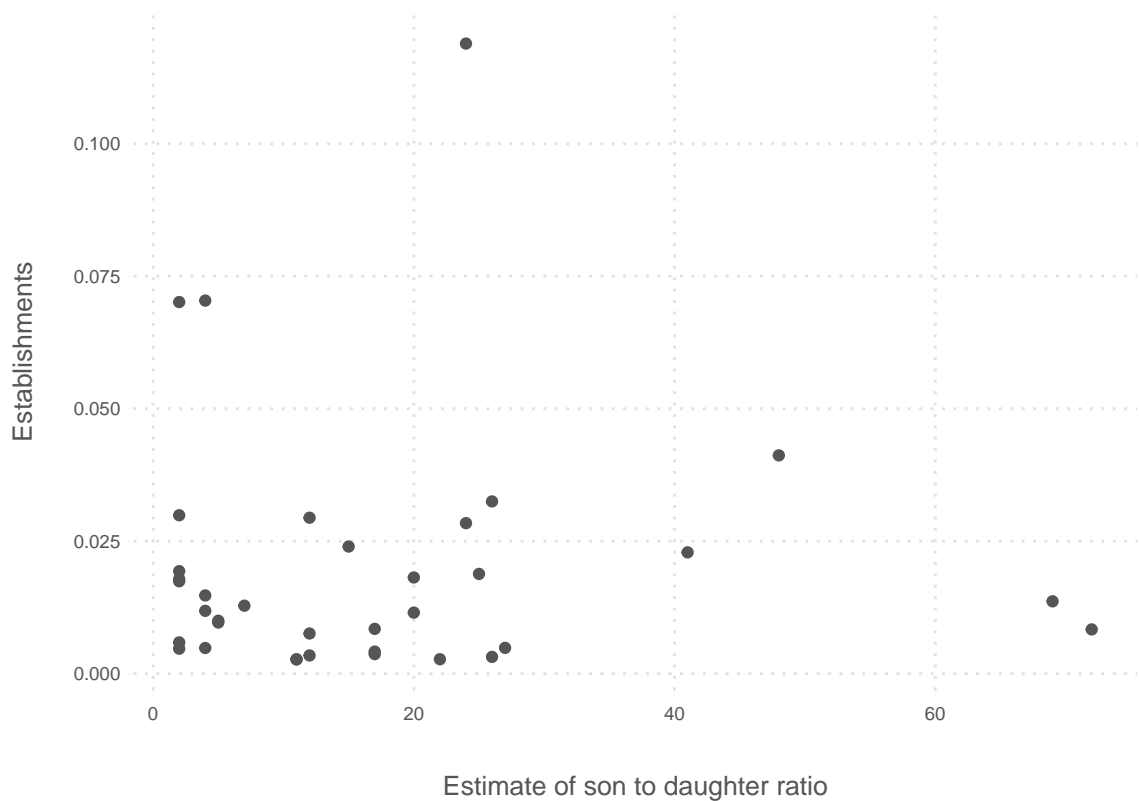


Figure 3: Estimate of son to daughter ratio vs # of Establishments

Next, we assess whether the ratio of son to daughter in business names varies by politics of the state. We get to the question in two different ways.

- a) **Region:** When we look at estimates of son to daughter ratio by region, states in the Midwest and Northeast generally have higher ratios than in the West and the South. Figure 4 shows the biggest gap is Midwest vs. South with a median ratio of 24:1 vs 5:1 respectively.

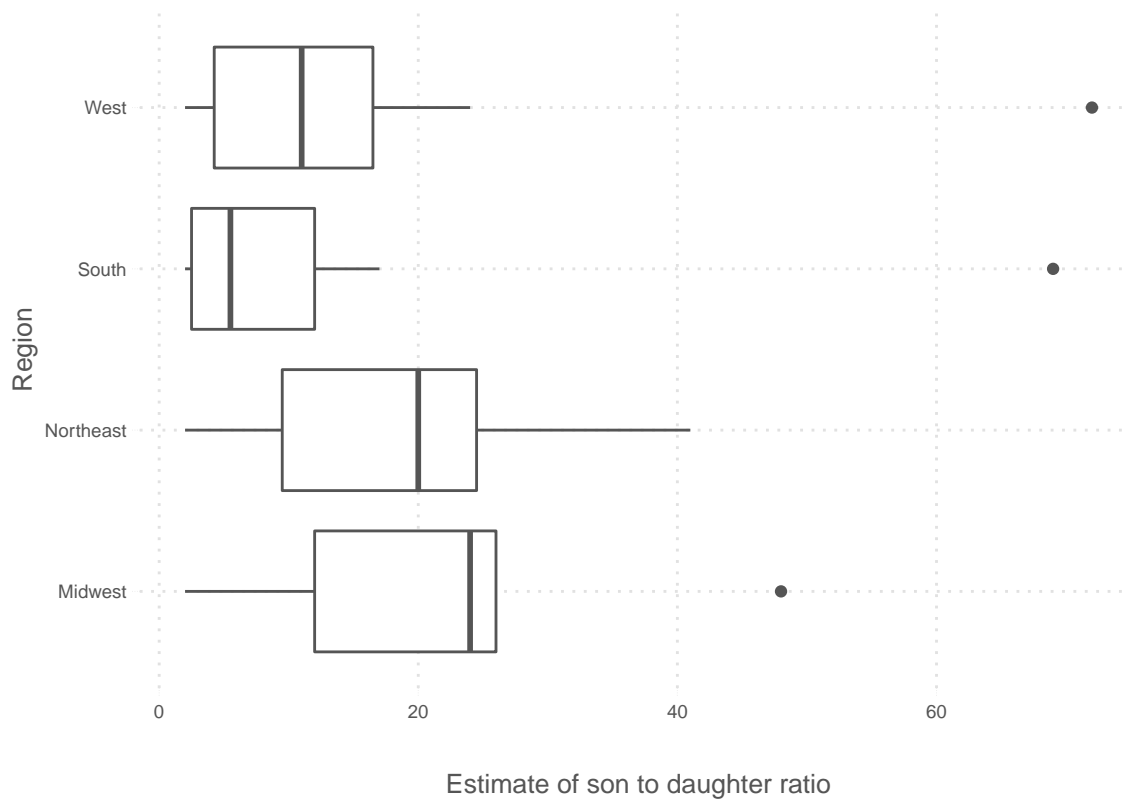


Figure 4: Estimate of son to daughter ratio by US Region

- b) **Political Inclination of the State:** We expect liberal states to have smaller ratios than more conservative states. To do that, we see how ratios compare between states where the Republicans won the presidential vote in 2016 than where the Democrats won. Counterintuitively, the ratio tends to be higher in Democrat states, with a median of 17 vs. a median of 11 for Republican states (see Figure 5).

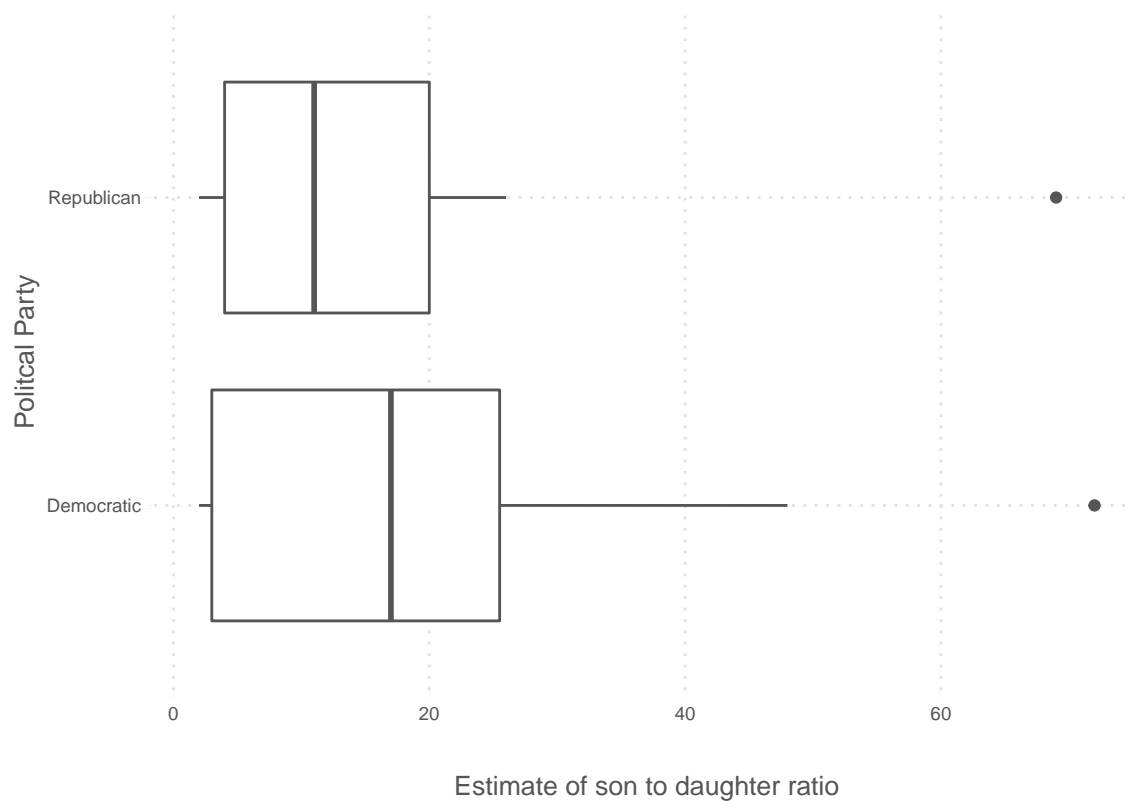


Figure 5: Estimate of son to daughter ratio by majority Political party (2016 elections)

Lastly, we investigate whether the ratio varies by income. We expect rich states to have smaller ratios than poorer states with the hypothesis that wealth is correlated with liberal attitudes. As Figure 6 shows, the relationship is weak.

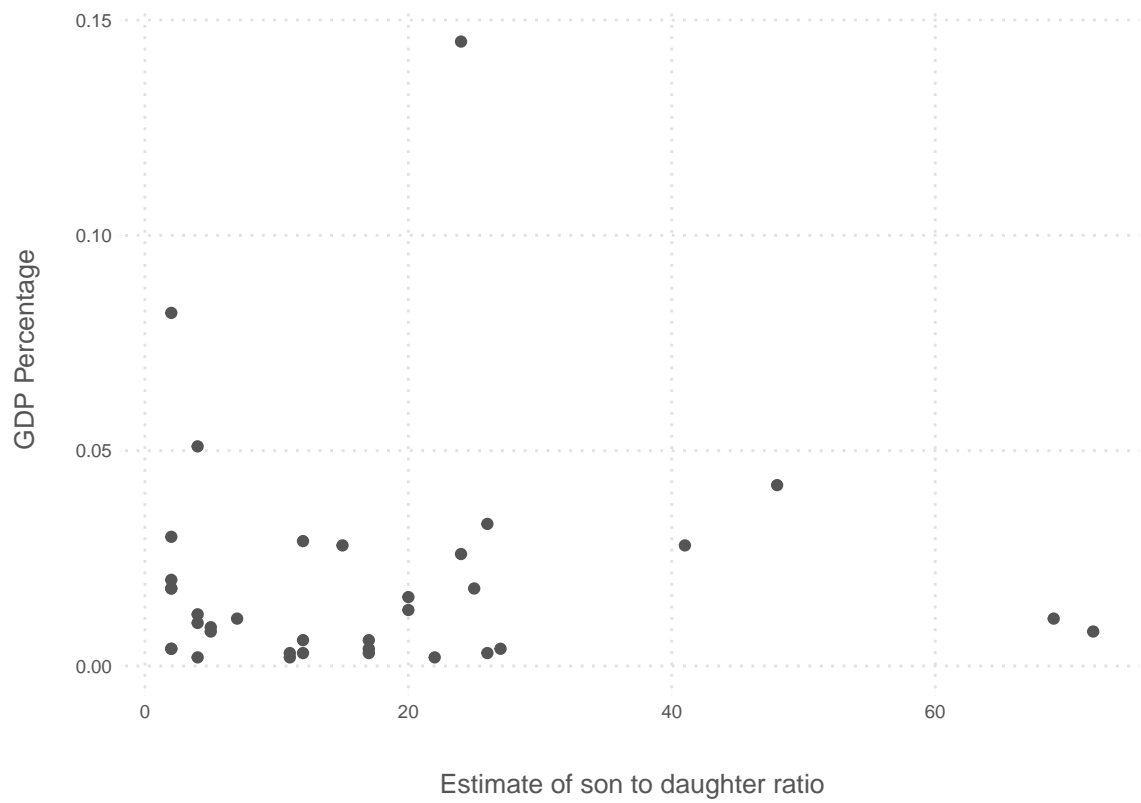


Figure 6: Estimate of son to daughter ratio vs GDP % of US



## References

1. Halpert, Chris. US census bureau regions and divisions.  
<https://github.com/cphalpert/census-regions/>
2. Kaushik, Saurav. Beginner's Guide on Web Scraping in R (using rvest) with hands-on example.  
<https://bit.ly/2Gj0sF6>
3. Kingl, Arvid. Web Scraping in R: rvest Tutorial.  
<https://www.datacamp.com/community/tutorials/r-web-scraping-rvest>
4. United States Census Bureau. State Population Totals and Components of Change: 2010-2019.  
<https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>
5. United States Census Bureau. SUSB Historical Data.  
<https://www.census.gov/data/tables/time-series/econ/susb/susb-historical.html>
6. Wikipedia contributors. (2020, January 13). Political party strength in U.S. states. In Wikipedia, The Free Encyclopedia. Retrieved 03:11, January 20, 2020.  
<https://bit.ly/37o6AYB>