Mapping SBA Loans

Methodology and Data Sources

Project History

On July 6, 2020, the Small Business Administration (SBA) released first round loan data from the Paycheck Protection Program (PPP). This project maps the spatial distribution of all these loan recipients. The team that worked on this project is a part of the IDA B. WELLS Just Data Lab, run by Professor Ruha Benjamin of Princeton University.

The final product can be accessed [here](). Datasets, license and codebook are available on [GitHub]().

Data Sources and Methodology:

SBA Loan Data

1. **Source:**

The SBA loan data were obtained directly from the [SBA website](). According to the website, 4,885,388 loans were approved, totalling $521.48 billions dollars. We compiled 58 files, consisting of 57 files recording loans below $150,000 that went to the US states and territories, and one national file recording loans above $150,000.

2. **Loan estimation:**

For loans higher than $150,000, the SBA did not release exact amounts but in ranges (ie. $150,000-$350,000, $350,000-$1 million, $1 mil-$2mil, $2 mil-$5 mil, $5mil- $10mil). This prevents accurate data analysis because even though these big loans constitute about [14.3% of the total number of loans, they account for 73.3% of the total loan amount](). To tackle this problem, we assumed that each loan is equal to the midpoint of its range (eg. a loan in the $1-2 mil range was estimated to be $1.5 millions). The [breakdown of total loan amount per range]() shows that with this method, we consistently overestimated the loan amount across all loan ranges. The sum of all the estimated loans in our final dataset is $544,828,131,868 - a 4% difference from the SBA's declared $521,483,817,756.

3. **Geocoding:**

To obtain the loan recipients' counties, we geocoded the addresses provided by loan recipients above $150,000 and ZIP codes provided by loan recipients below $150,000. When one ZIP

code corresponded with more than one county, we recorded the county where the ZIP code's centroid fell. This process was largely done on ArcGIS by the Princeton map librarian.

In our datasets, 509 businesses were not accounted for since they either did not fill out ZIP codes or filled out non-existent ones. These businesses in total account for 15 million dollars.

Population Estimate:

Our population estimates were taken from the 2019 County Metro Population Estimates. The dataset name is Population, Population Change, and Estimated Components of Population Change: April 1, 2010 to July 1, 2019

2020 US Average Household Income

This layer was obtained from and is the property of Esri.