Dataset I: Unemployment and Household Income in the United States

United States Department of Agriculture Economic Research Service. (2018). Unemployment and median household income for the U.S., States, and counties, 2007-17 [comma separated values file]. Retrieved from <https://www.ers.usda.gov/webdocs/DataFiles/48747/Unemployment.xls?v=0>. Last accessed 10 September 2018.

License: [CC0 1.0 Universal (CC0 1.0) Public Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/)

The owner of this work has dedicated it to the public domain. The owner waives copyright so any member of the public can “copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission” (“CC0,” 2018). Under this license, the person who has dedicated this work to the public is not held liable for its uses. Users should not “imply endorsement by the author or the affirmer” when citing or referring to the work (“CC0,” 2018).

U.S Department of Agriculture Economic Research Service (ERS) compiles data in their efforts to study rural America (Economic Research Service, 2018). This dataset is robust because it contains granular, county-level data over several years (2009-2017), making it possible to observe trends over time and over space. This dataset is sourced from the Bureau of Labor Statistics and ERS has added data that describes how urban or rural the area is based on their classification system (United States, 2018). This classification of how urbanized a county is allows the user to make observations about employment and median household income in urban and rural parts of the country. This dataset is interesting on a few levels. It could reveal trends in economic health in rural compared with urban areas. On a more local level, it can be interesting for a resident to learn that their county of residence has a much higher or much lower median household income than the state or national average. When I showed this to my parents, they were floored by the income disparity within our home state of Maryland. Even without additional statistical analysis, the numbers seem to tell a story.

Potential data users for this dataset are government workers and/or academics that study economic health in the United States. The dataset starts at the end of the Great Recession and may provide insight to those looking to understand the economic status of American counties in the post-Recession era. This dataset could also help policymakers decide where to allocate social welfare benefits, based on need. A county with a high median household income may still have pockets of poverty, but this data could be used to reevaluate the block grants given to each state. Similarly, non-profit or charity groups could use this data to decide where to focus their efforts. This dataset does not include the needed geospatial mapping data, but it may be possible to map this data since the county name is provided. An anti-poverty group could create a map to illustrate the areas in critical need of economic assistance. An aspiring state politician could use a similar map to prove to potential supporters that the incumbent candidate has not succeeded in fulfilling their economic promises, or the incumbent could use such a map to demonstrate their success.

1. Is there geospatial wealth disparity in the United States? If so, where is this most extreme?
2. What is the status of employment and wealth accumulation in the United States on a county level? This question could be more fully answered if combined with Bureau of Labor Statistics reports on which industries are employing how many people over the same time period (2009-2017).
3. What is the correlation between urbanization and employment rates in the United States over the past nine years?

In order to use this dataset, it is necessary to understand the classification codes ERS assigns each county. Two of these codes are linked below:

[Rural-Urban Continuum Code](https://www.ers.usda.gov/webdocs/DataFiles/53251/ruralurbancodes2013.xls?v=0)

[Urban Influence Code](https://www.ers.usda.gov/webdocs/DataFiles/53797/UrbanInfluenceCodes2013.xls?v=0)

Dataset II: Microfinance Loan Profiles of Kiva-Facilitated Loans

Kiva. (2018). Kiva\_loans.csv (Version 5) [comma separated values file]. Retrieved from <https://www.kaggle.com/kiva/data-science-for-good-kiva-crowdfunding/downloads/kiva_loans.csv/5>. Last accessed 11 September 2018.

License: [CC0 1.0 Universal (CC0 1.0) Public Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/)

The owner of this work has dedicated it to the public domain. The owner waives copyright so any member of the public can “copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission” (“CC0,” 2018). Under this license, the person who has dedicated this work to the public is not held liable for its uses. Users should not “imply endorsement by the author or the affirmer” when citing or referring to the work (“CC0,” 2018).

This data from the non-profit organization Kiva (www.kiva.org) is a detailed summary of all the loans given through their organization in the past few years (2014-2017). Kiva’s main activity is to connect lenders to people in need of small loans across the world. Anyone can be a lender as long as they have Internet access and money. The lenders can find projects or people seeking small loans (about $25-$100) through Kiva’s website (Kiva.org, 2018). The project description provides details about the potential recipient, what they will use the money for, and how it may help their family or community. For example, one could lend $25 to a woman in Guatemala seeking to buy materials for a new tailoring business. When the loan is repaid, the lender can choose to invest in another project, donate the money, or withdraw the funds completely (Kiva.org, 2018). This dataset contains several parameters of data for every loan in the last three years, such as amount, date of disbursal, categorical data about the project theme or topic (agriculture, education, transportation, etc.), and country of recipient, among others (Kiva, 2018). This dataset is interesting because it is a fairly comprehensive snapshot of the types of projects lenders like to fund, where these projects are in the world, how much (or how little) is needed to accomplish projects, and may provide a sense of who the average Kiva loan recipient is.

Kiva opened up this data as part of a data challenge hosted by the website Kaggle (“Data,” 2018). This dataset is posted along with several others. Their stated goal is to “build more localized models to estimate the poverty levels of residents in the regions where Kiva has active loans” (“Data,” 2018). With this in mind, of course the main user of this dataset would be Kiva and the data scientists who choose to enter the data challenge. This data could also be used by researchers interested in evaluating efficacy of microloans, the types of projects funded by microloans, and the types of recipients of microloans. This dataset could potentially be used by lenders who may wish to loan money to recipients in countries that are funded less frequently, or sectors that are funded less frequently. However, it does seem unlikely that a casual lender would take the time to download this dataset to make their decision. Most likely this trend and country profile information would be used by Kiva itself to develop strategies to accomplish their operational goals.

1. What sectors or topics receive the most loans?
2. What is the average time of repayment by country? By topic? By sex of recipient? Depending on the answer, it may be important to contextualize the findings through socioeconomic and historical information about the county.
3. Which topics or sectors are popular in which regions of the world?

Analyzing this dataset requires understanding the column ID meanings, which are defined here: <https://www.kaggle.com/kiva/data-science-for-good-kiva-crowdfunding#kiva_loans.csv>

Dataset III: Comprehensive United States Health Indicator Data

Centers of Disease Control and Prevention Division of Population Health. (2018). U.S. Chronic Disease Indicators (CDI) [comma separated values file]. Retrieved from <https://chronicdata.cdc.gov/api/views/g4ieh725/rows.csv?accessType=DOWNLOAD&bom=true&format=true>. Last accessed 10 September 2018.

License: [ODC Public Domain Dedication and License (PDDL)](https://opendatacommons.org/licenses/pddl/summary/)

Under the terms of this license, the user can use the data in any way they wish, such as copying, distributing, and modifying, with no requirements for accreditation (“ODC,” 2018).

The Centers for Disease Control and Prevention has a set of indicators to define different types of chronic disease and human health outcomes. These 124 indicators are designed for local health departments to survey and classify the health of their municipalities using standards that are consistent across the country (“U.S. Chronic,” 2018). The data set lists the levels of each indicator on a state-level from 2001 to 2016. Not every state reported on each indictor every year. The data is also broken down by indicator level by sex and by race in each state for every year that is reported (Centers, 2018). Some indicators are age-specific, providing further demographic details. This data is interesting because it includes a wide range of health topics, such as alcohol use, prevalence of diabetes, sexual health, and levels of physical activity.

Local public health officials can use this data to understand how their municipality compares with others, and with national averages. Public health decision-makers could use this data to determine health topics that need to be addressed and also determine areas where their policies may have improved health outcomes or health literacy. Health insurance companies may use this data to determine what types of coverage they should offer in certain areas, or at certain times. Health literacy groups may use this data to determine what types of campaigns they should offer. For example, one indicator is median fruit consumption among adults over age 18 (Centers, 2018). A health literacy group may determine certain racial groups, or certain states overall, have low fruit consumption. They could then develop culturally appropriate nutrition curricula and target those populations specifically. Due to the number of indicators and topic areas, dozens if not hundreds of questions can be asked of this data.

1. Is there a racial difference in incidences of cancer, diabetes and other non-communicable diseases within a certain state? Across the country? If there is, additional datasets on lifestyle factors or poverty in different communities might help contextualize why these differences exist.
2. Within the topic of alcohol, which indicators have improved in the state of Maryland between 20001 and 2016? Additional information about educational campaigns and alcohol regulations can bring additional meaning to these findings. It could be useful to confirm any findings with other types of data, such as alcohol-related car accidents or crimes. This broader view can help illustrate impacts on society of improved health indicators, in addition to just looking at impacts on the user.
3. Which states have the highest rates of premature mortality? Additional information about concentration of industrial factories, clean water status, and other environmental health data from Environmental Protection Agency or local health departments could provide more context about why certain states or certain racial groups may have greater incidence of premature mortality.

A drawback of this dataset is the number of column ID abbreviations that need to be defined for a user to understand what the data means. For example, the value for median fruit consumption seems to range from 1 to 1.4. Does that number refer to how many fruits are being eaten? Is it a ratio? All of the indicators and values are defined in another CDC document titled [Indicators for Chronic Disease Surveillance – 2013](https://www.cdc.gov/mmwr/pdf/rr/rr6401.pdf). It is 252 pages long, so it may be burdensome to an information seeker to properly define all the terms he or she is analyzing.

Word Count Excluding References: 1, 866

References

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. (2018). Retrieved from <https://creativecommons.org/publicdomain/zero/1.0/>. Last accessed 11 September 2018.

Centers of Disease Control and Prevention Division of Population Health. (2018). U.S. Chronic Disease Indicators (CDI) [comma separated values file]. Retrieved from <https://chronicdata.cdc.gov/api/views/g4ieh725/rows.csv?accessType=DOWNLOAD&bom=true&format=true>. Last accessed 10 September 2018.

Centers for Disease Control and Prevention. “U.S. Chronic Disease Indicators (CDI).” 9 July 2018. Retrieved from: <https://chronicdata.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-CDI-/g4ie-h725>. Last accessed 13 September 2018.

Data science for good: Kiva crowdfunding. N.d. Retrieved from <https://www.kaggle.com/kiva/data-science-for-good-kiva-crowdfunding/home>. Last accessed 10 September 2018.

Economic Research Service. “Overview.” 7 June 2018. Retrieved from <https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications.aspx>. Last accessed 11 September 2018.

Kiva. (2018). Kiva\_loans.csv (Version 5) [comma separated values file]. Retrieved from <https://www.kaggle.com/kiva/data-science-for-good-kiva-crowdfunding/downloads/kiva_loans.csv/5>. Last accessed 11 September 2018.

Kiva.org. “The journey of a Kiva loan.” N.d. Retrieved from <https://www.kiva.org/about/how>. Last accessed 11 September 2018.

ODC Public Domain Dedication and License Summary. N.d. Retrieved from: [https://opendatacommons.org/licenses/pddl/summary/](https://opendatacommons.org/licenses/odbl/summary/). Last accessed 11 September 2018.

United States Department of Agriculture Economic Research Service. (2018). Unemployment and median household income for the U.S., States, and counties, 2007-17 [comma separated values file]. Retrieved from <https://www.ers.usda.gov/webdocs/DataFiles/48747/Unemployment.xls?v=0>. Last accessed 10 September 2018.