This is a first attempt at establishing some form of electric cooking potential index. I have taken four datasets to identify large urban populations that are (1) using primarily biomass fuels (2) have access to electricity and (3) have sufficient income to purchase and use an induction stove, potentially with government or NGO assistance.

The four datasets I am using are:

1. The United Nations World Urbanization Prospects 2018 dataset on all cities with over 300000 residents (<https://population.un.org/wup/Download/>)
2. The WHO Household Energy database for proportions of residents using dirty fuels (by country) (<https://www.who.int/data/gho/data/themes/air-pollution/who-household-energy-db>)
3. The World Bank World Development Indicators for proportion of urban residents with access to electricity (by country) (<https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators>)
4. Data from Our World in Data's poverty explorer (<https://ourworldindata.org/poverty>) originally compiled from World Bank Income Data (<https://pip.worldbank.org/home>) to get the proportion of residents living in extreme poverty.

I combine these datasets into one dataframe containing:

1. Urban areas over 300000 residents, for which all below data is available
2. Population of the urban area
3. Proportion of the parent country's urban residents using dirty fuels (biomass, charcoal, kerosene, or coal)
4. Proportion of the parent country's urban residents with electricity access
5. Proportion of the parent country's residents living in extreme poverty (less than $1 per day)

To get a sense of which urban areas might benefit from an electric cook stove intervention, I computed an “estimated addressable population” column by multiplying the dirty fuel, electricity, and not-extreme poverty proportions by the population of each urban area:

Estimated addressable population = population \* (% dirty fuel use) \* (% electricity access) \* (% above extreme poverty).

# Major Issues

I can see several major issues with this first attempt.

First, the proportions I am using are by country, not by urban area. This will definitely lead to weird results where countries can have highly varied development (e.g. Shanghai is listed as one of the largest populations with dirty fuels, which seems unlikely. But maybe not—I’ve never been to Shanghai). Ajay pointed out [this dataset](https://www.pnas.org/doi/10.1073/pnas.2113658119#supplementary-materials) which might help, at least for income.

Second, there is no information on the quality of electricity access. The robustness of a city's infrastructure will play a huge role in whether induction stoves are a scalable solution in the near term, so we have to find a way to incorporate this.

Third, I'm not really sure how to define a cutoff point for "able to afford an induction stove." This depends on the price of stoves, electricity prices, the size of potential subsidies, and income. Related is how willing people would actually be to use such stoves. If LPG is cheaper, then people will just use that instead (Ecuador is a good example).

Fourth, we don't have any information on potential for government cooperation (e.g. corruption indexes).

Fifth, just wondering about other factors I may be missing. For example, environmental factors seem important. For example, electric stoves wouldn't work in Mongolia because the stoves are used for heating homes, in addition to cooking.

Sixth, there is probably high overlap between people using dirty fuels, without electricity access, and with low income. So this first analysis probably overestimates the size of the addressable population.

# Other Thoughts

Seems like we’re attempting to measure (1) the potential health impacts of flipping a population to electricity and (2) the actual uptake scales and probabilities.

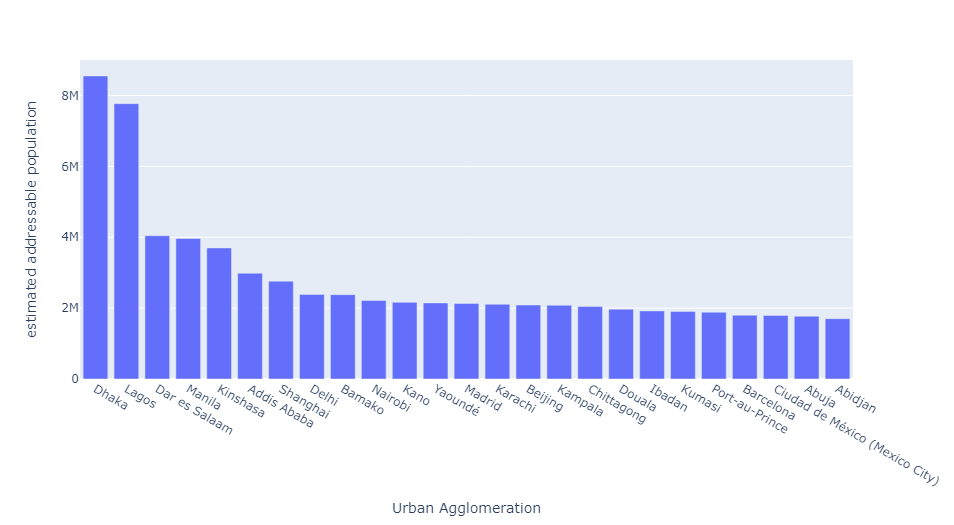
Point (1) seems much easier to measure (just the total population using dirty fuels). However there is still room to add nuance: urban density, extent of fuel use, and size of clean fuel use population would also matter in terms of total health effects.

Point (2) Is difficult. It requires understanding electricity infrastructure/stability, government stability, financing, and political support, income and economics of the population, geography/environment, and a whole lot else. Phew.

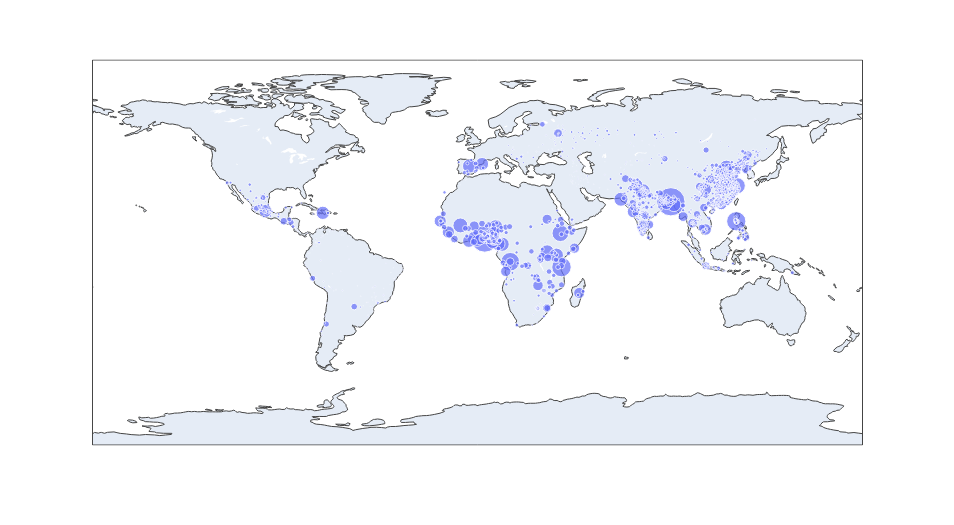
Other observations:

1. We're missing out on small island countries, because they don’t have large urban areas. Might be reasonable to include these anyway though. Issues of justice/bias here, plus such countries could be interesting proving grounds for scaling up interventions.
2. Almost half the cities in this list are in India or China
3. Income data isn't consistent across years (some countries have more recent data than others). More of a reason to use the dataset Ajay suggested.

# Plots



The estimated addressable population of the top 25 urban areas.



Map showing the location and addressable population size of all ~1500 urban areas