Springboard Deliverable Data Wrangle

Taylor Hughes

4/11/2018

# How to Obtain Necessary Data

In order to get our data ready to run a regression, we must change it to be useful and able to be interpreted. In order to do this, we must make our Census data uniform with the format of the 2017 Global Coworking Survey results. You can see the different demographics we will be anayzing below, and for a detailed write-up please refer to my [comprehensive paper](https://github.com/taylorhughes291/Coworking/blob/master/Data/Coworking_Data%20Wrangling.rmd).

## What data do we need?

In summary, we need to be on the lookout for the following in our PUMS Data:

* Does the person Drive, Bike, Walk, or Transit as their commute?
* How long would that person’s mode of transportation take to get to our coworking space?
* What is the age of the person?
* What is the person’s job type?
* What field does the person work in?
* What is the person’s gender?
* What is the person’s marital status?
* How many children does the person have?
* What type of health insurance does the person have, if any?
* What is the person’s current workspace?
* What is that person’s highest level of education?
* What is this person’s relative income?

## How to create uniform data formats

Most of the fields were relatively easy to convert. Every field had to be converted to a factor, even integer values such as age, since the Global Coworking Survey formatted things in bins such as “Age 18-29”. However, it was very difficult to interpret commute time in such a way that was relevant and uniform with the results of the survey. In order to accomplish this, the following steps were taken:

* Local Market Area Census Data
  + Assign Census Observations into Census Tracts by Census Tract Population and observation closeness to Census Tract Average Income
  + Determine latitude and longitude of centroid point of each Census Tract and compute commute times from these points to the proposed Coworking Space location using Google Maps API.
  + Use this information to determine if each Census observation can commute to the Coworking Location within 20 minutes.
* Training Data Census Data - LA County
  + Assign LA Census Observations into LA Census Tracts by Census Tract Population and observation closeness to Census Tract Average Income
  + Determine latitude and longitude of centroid point of each Census Tract and latitude and longitude of every Coworking Space in Los Angeles County
  + Compute for each Census Tract the geographically closest Coworking Space in terms of Euclidean Distance
  + Compute the commute time between each Census Tract centroid point and its closest geographical Coworking Space.
  + Use this information to determine if each LA County Census observation can commute to any Coworking Location within 20 minutes.

## Completed Market Area Census Data

Please see the .txt file for the [Market Area Census Data](https://github.com/taylorhughes291/Coworking/blob/master/Data/Springboard%20Deliverables/Market%20Census%20Final%20Data.txt).