# CS506

# EARI project Final Report (draft)

## Project members :

## Nianyi Zhang,

## Haoran Wei,

## Ruikang Wang

## Introduction

*Project questions and motivation:*

We are mainly concerned about two questions: First, how have economic benefits, specifically defined as the number of energy efficiency jobs, flowed to different census regions in [Massachusetts](https://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.1.4320&q=%E9%A9%AC%E8%90%A8%E8%AF%B8%E5%A1%9E%E5%B7%9E)? How does it vary between advantaged and disadvantaged communities, defined as lower-income, less-educated and enhanced-minority census regions? Second, what dependent variables demonstrate high levels of correlation with the independent variables of higher job numbers and higher wages in those disadvantaged communities?

To address the first question, we want to find out how energy efficiency jobs number would vary as the independent variables changes. Specifically, energy efficiency jobs include highest-paying jobs-architecture and engineering and low-wage jobs like construction and installation trades. We will compare these jobs’ numbers independently between majority-white and majority-non-white communities, between Hispanic and non-Hispanic communities, between educated and less-educated communities, between high-income and low-income communities.

To handle the second question, we will run a simple regression model to find out the variables which satisfy our demands.

*The datasets description:*

We are working on the 2012-2016 ACS 5-year estimates census tract data for race, ethnic, education, and occupation in MA. These data can be retrieved through the Census Bureau by searching the keywords above. At first, not all of these data are useful to us, so what we have to do is to filter first, deleting those data we don't need. And the necessary step for us to do is to give these data proper flags. For instance, we will flag a census region as either educated community or under-educated community depending on what percentage of people in the region has a four-year diploma or higher. We will illustrate all these definitions clearly later in Methodology.

## Methodology

The first question only requires data manipulation skills. At first, we classify these communities by certain definitions. And then we define an ECONOMIC INCLUSION INDEX which is meant to show the comparison between different communities.

**Community Definitions：**

**Income:** Low-income and high-income communities are defined by comparing a region’s median Household Income to the county-specific median Household Income. Median income was used as it better accounts for uneven income distributions, where averages could be skewed to either the high or low end of earnings. A high-income community would be one that has a higher median Household Income compared to the county median Household Income, while a low-income community has a lower median Household Income than county median Household Income.

**Demographics:** There are two comparative groups based on demographics: (a) White communities vs. some other race (African American, American Indian, Asian, Hawaiian, and Pacific Islander, and Other), and (b) Hispanic communities vs. non-Hispanic. The threshold to determine "predominantly" White or "predominantly" Hispanic were based on state-specific averages. A predominantly White community would have a higher proportion of White individuals compared to the statewide average.

A predominantly Hispanic community would have a higher proportion of Hispanic individuals compared to statewide average.

**Education:** Similar to demographics, region-specific baselines were used to identify communities with a higher proportion of individuals with a Bachelor’s degree or higher. An educated community would have a higher proportion of individuals with a Bachelor’s degree or higher.

**Economic Inclusion Index**

The index was generated by comparing resident employed concentration of architecture and engineering occupations, construction and extraction occupations, and installation, maintenance, and repair occupations from the U.S. Census Bureau's American Community Survey. Employment estimates by occupational category were collected for each census tract within the geographies used. The concentration of employment was compared between each set of measures (high income vs. low income, educated vs. under-educated, etc.) for aggregated census tracts. The index measures the comparison of each measure to its opposite category. For example, a 5% employment concentration for architecture and engineering occupations for all employed residents in aggregated high-income census tracts compared to a 2% employment concentration for architecture and engineering occupations for all employed residents in aggregated low-income census tracts would result in a 2.5 index score for aggregated high-income census tracts for that geography.

HIGH INCOME = .05/.02 = 2.5 INDEX – 2.5 times more concentrated in high-income neighborhoods.

## Summary and Results interpretation

**In general**, it was found that energy efficiency-related employment is found across all types of communities, advantaged and disadvantaged alike. However, the highest-paying jobs, which are typically found in architecture and engineering are most highly concentrated in advantaged—high-income, educated, White and non-Hispanic communities, while lower-wage energy efficiency jobs in construction and extraction or installation, maintenance, and repair are most often found in disadvantaged communities with lower education and a higher prevalence of ethnic and racial minorities residents. It should be noted that while these are the general trends seen across the state, there is some variation by geography.

**The highest-paying energy efficiency jobs—architecture and engineering—are more likely to be concentrated in high-income, non-Hispanic, educated neighborhoods.** In most of the 14 counties of Massachusetts, energy efficiency-related architecture and engineering jobs are at least one and a half times more likely to belong to residents who live in high-income neighborhoods, except a few exceptions. The overall contrast is most striking in Berkshire County, where high-wage energy efficiency jobs are eight to thirty times more likely to be concentrated in “advantaged” communities (Figures 1 through 4).

**At the same time, low-wage energy efficiency jobs across construction and installation trades are most likely found in disadvantaged communities.** For example, energy efficiency workers living in low-income neighborhoods are most likely to work in construction and installation, particularly in Nantucket, Norfolk and Suffolk (Figures 5 and 6). These construction and installation jobs are more concentrated in the non-white majority, undereducated and Hispanic communities (Figures 7 through 9).

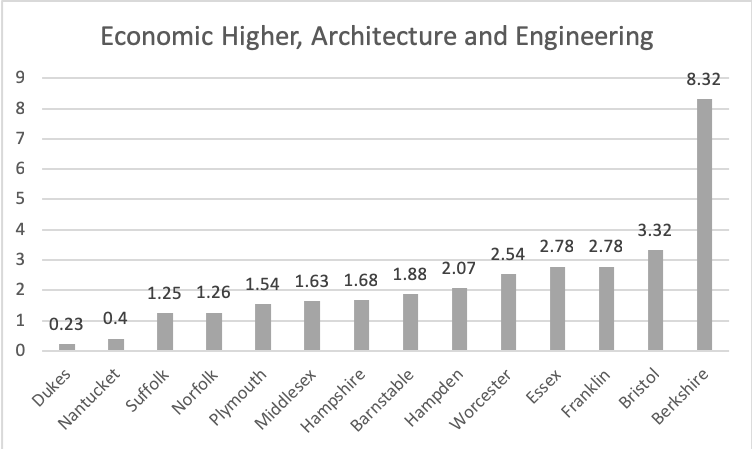


Figure 1

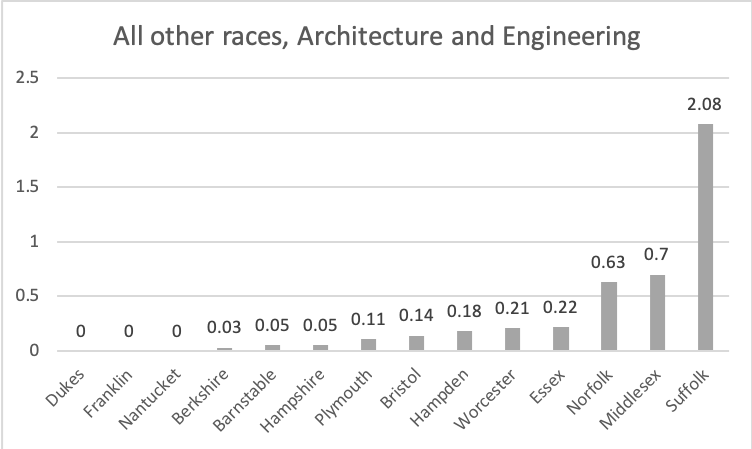


Figure 2

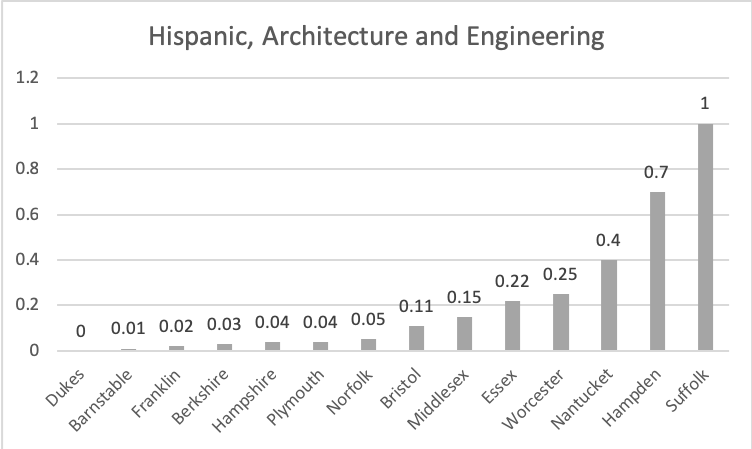


Figure 3

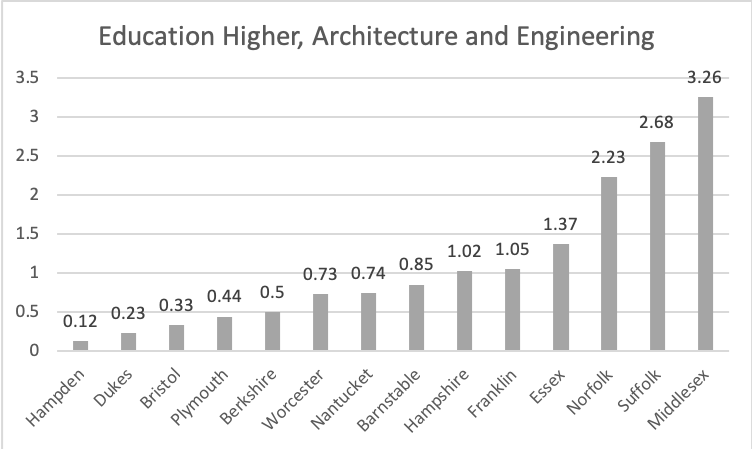


Figure 4

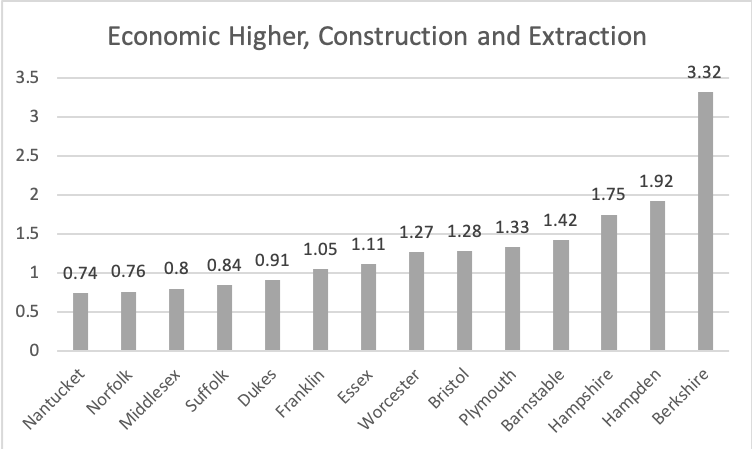


Figure 5

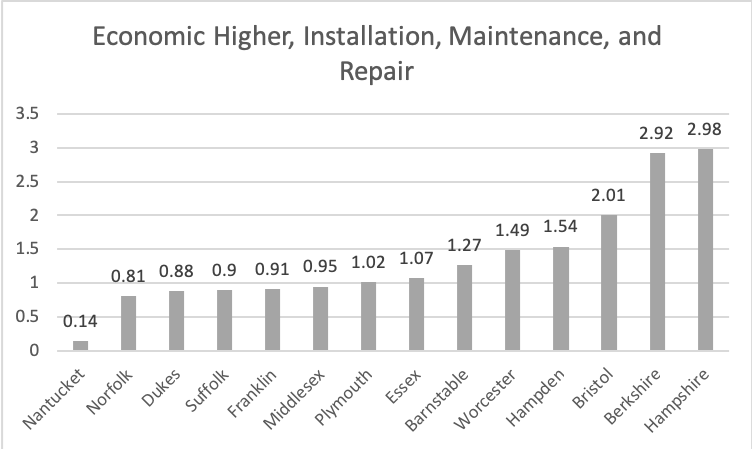


Figure 6

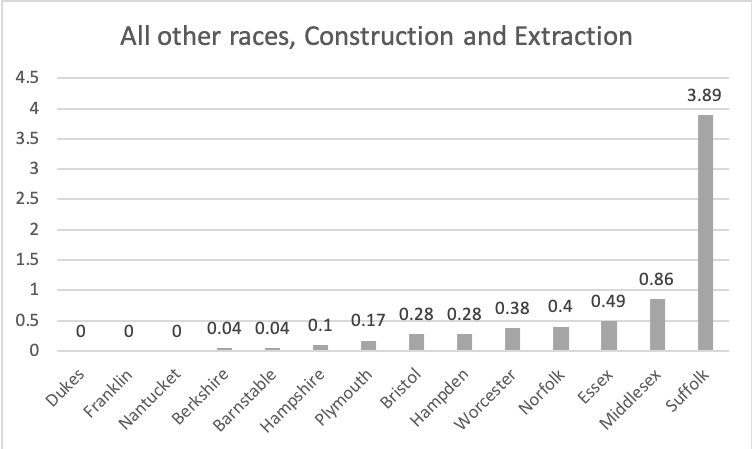


Figure 7

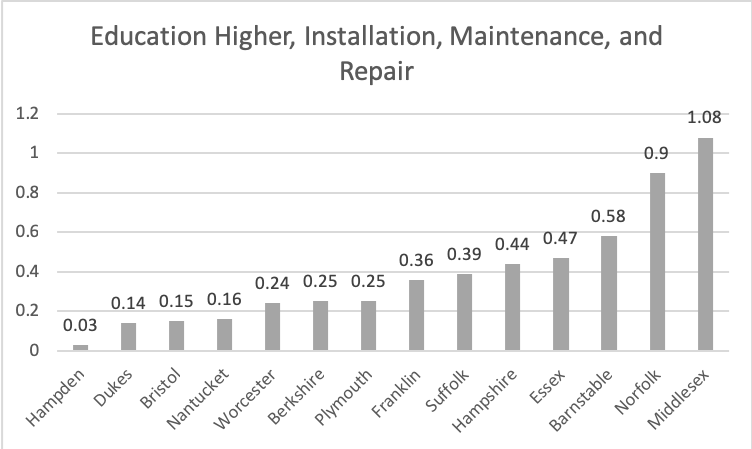


Figure 8

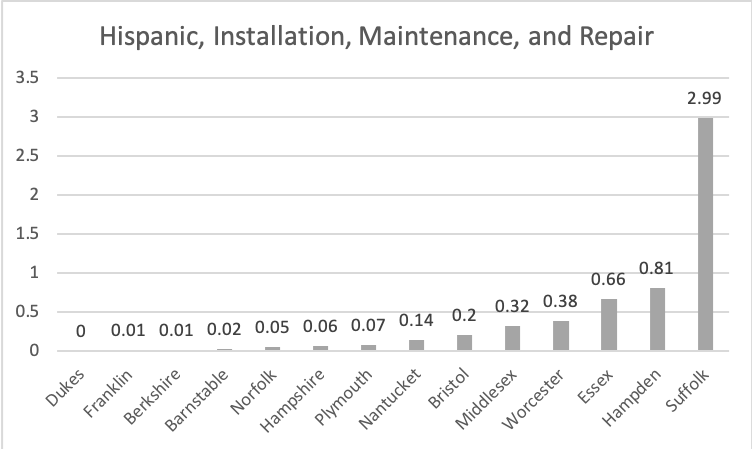


Figure 9

## Extensive thoughts

However, all these steps can be done within the Excel software only with some function specifying parameter to manipulating all the data we need. Because of limited time, we cannot fully go through the step of learning advanced skills of Excel.

From the figures we retrieved above, we could find that some county like Suffolk is pretty different from the normal case where it has more Hispanic and other races people but it could beat other county in some areas. We could step further to understand that occasion.