EDA Project- Degree Days Statistics

Mengyun Li, Shijie Cai, Fei Shi, Xiang Zhao

10/13/2017

1. Introduction

1.1 Project Description

Based on the information from the Climate Prediction Center, our team choose the data from heating degree days for monthly city. We choose in total 36 months (from 2014 - 2016) for 4 cities, Seattle, Boston, Chicago and Detroit (2 cities are on the coast with similar longitude and 2 inland cities close to each other) and the numbers in the raw data represent the summations of negative differences between the mean daily temperature and the 65°F.

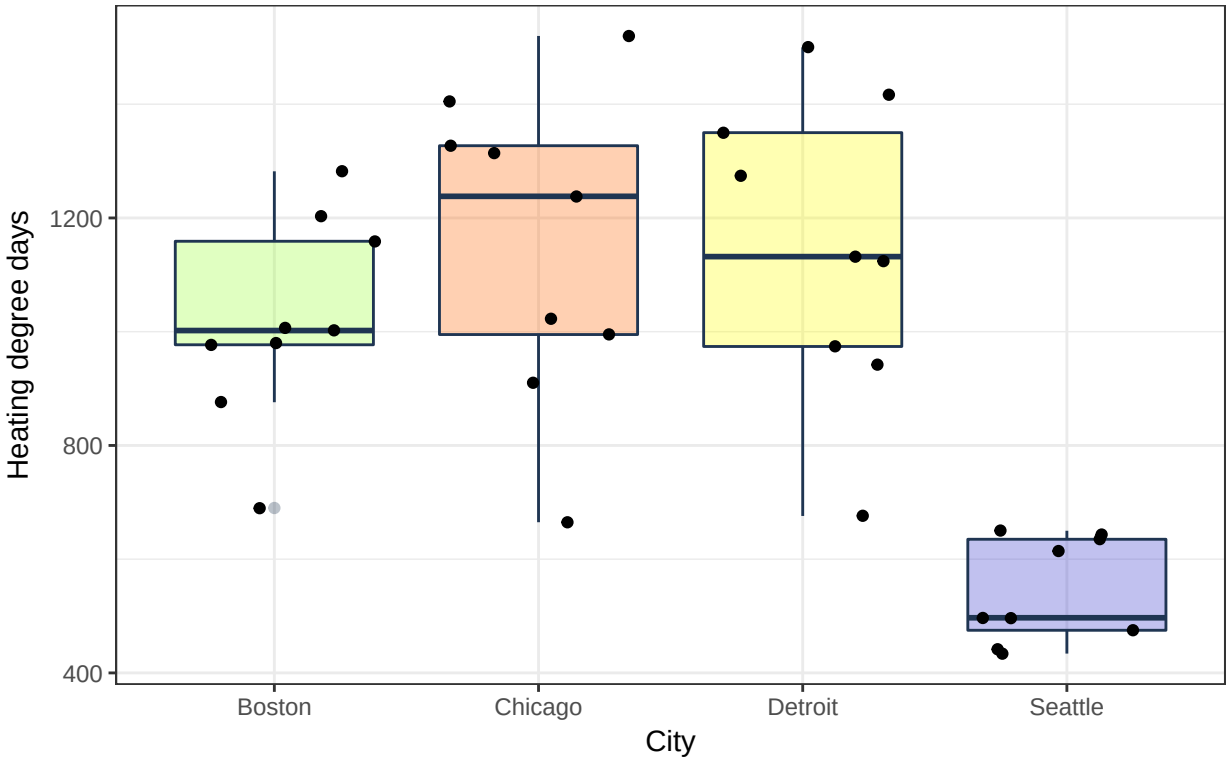
1.2 Research Question

We are interested in how geographic location would influence the usage of heating?

2. Exploratory Data Analysis (EDA)

To show how geographic location would have affected the degree days, we first sort 12 months into 4 seasons, then calculate the mean season indexes for 3 years. Each box, would contain 9 points, stands for 1 season of a single city. Then we would create paired box plots of 2 inland cities and 2 coast cities to see the difference.

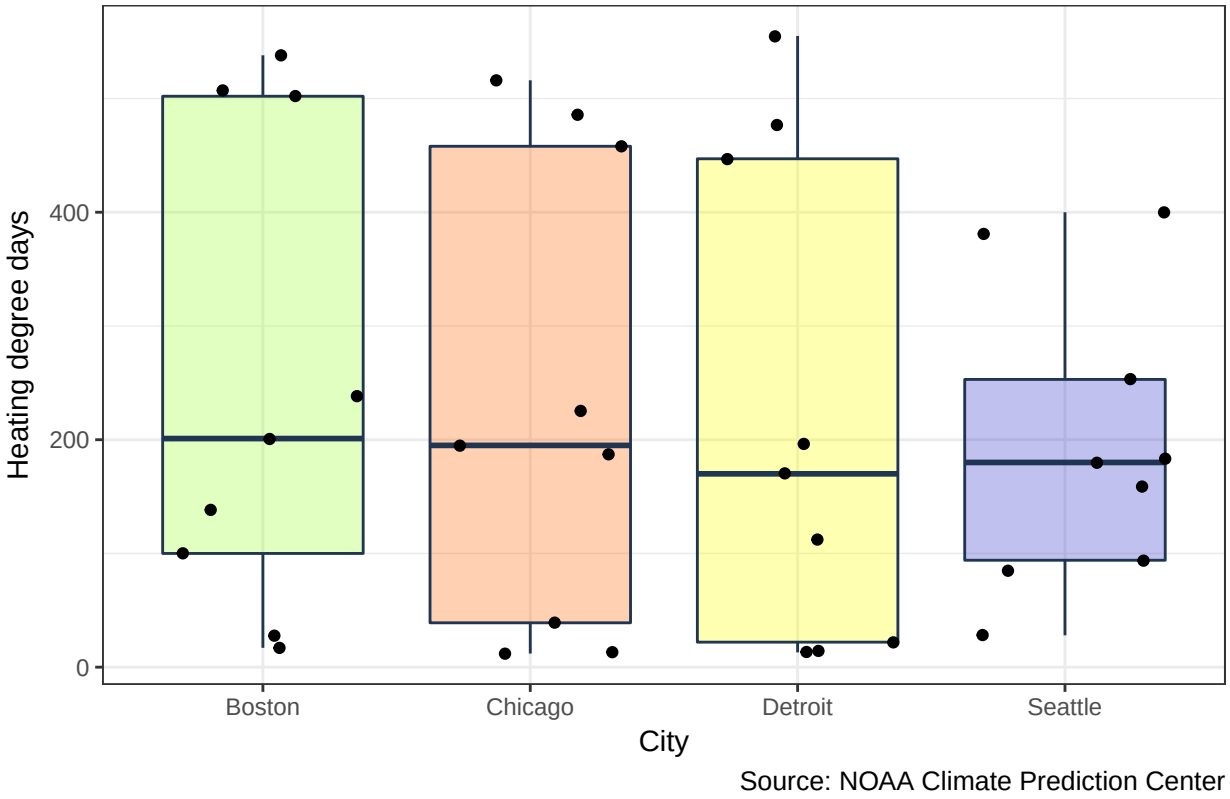
Figure2.1: 2014−2016 heating degree days of four cities in spring



2.1 Analysis of Heating Degree Days of Four Cities in Spring

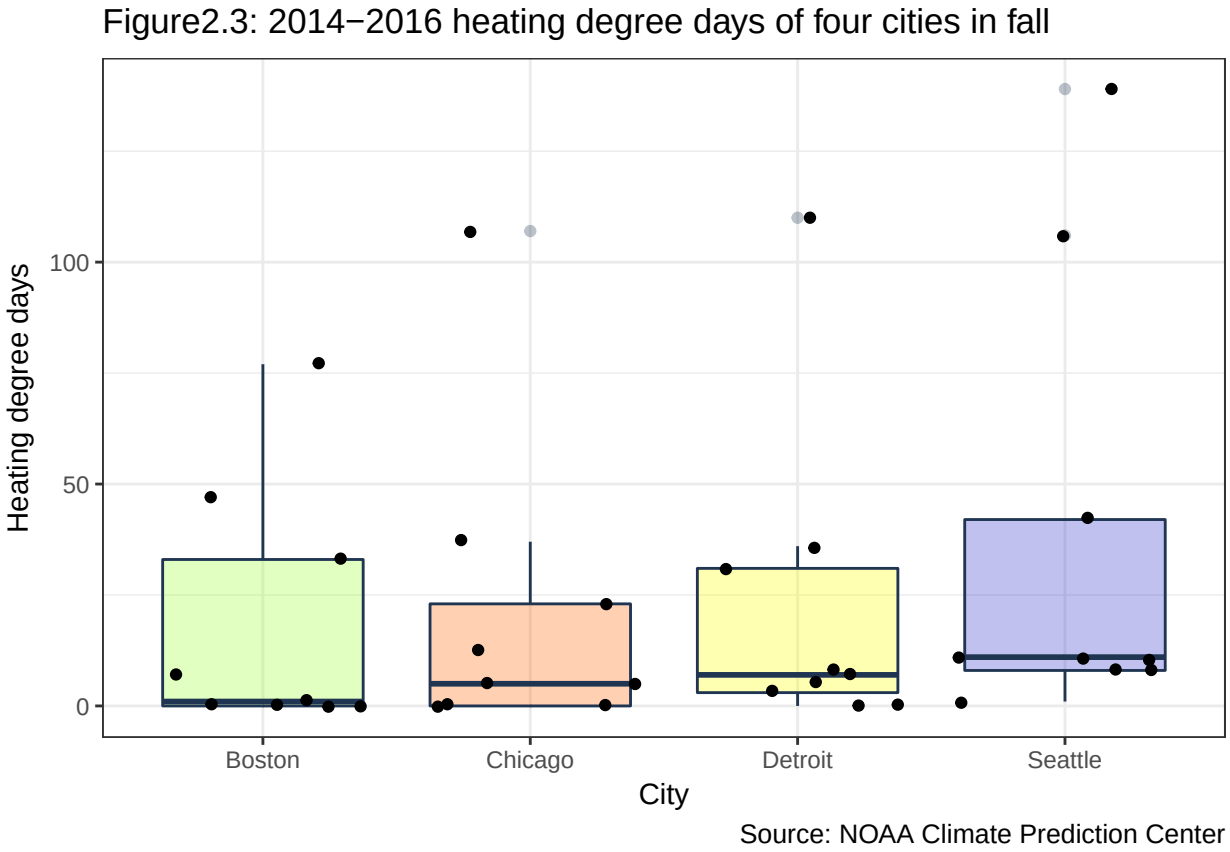
Figure 2.1 shows the heating degree days in the season Spring for four cities. From the box plot we observe that the heating degrees days lies in similar interval for inland cities but have different means. Instead of inland cities, although the distributions of heating degree days in Boston and Seattle looks almost same , the heating degree days in Seattle is extremely smaller than Boston. The variance for coast cities are obviously smaller than inland cities.

Figure2.2: 2014−2016 heating degree days of four cities in summer



2.2 Analysis of Heating Degree Days of Four Cities in Summer

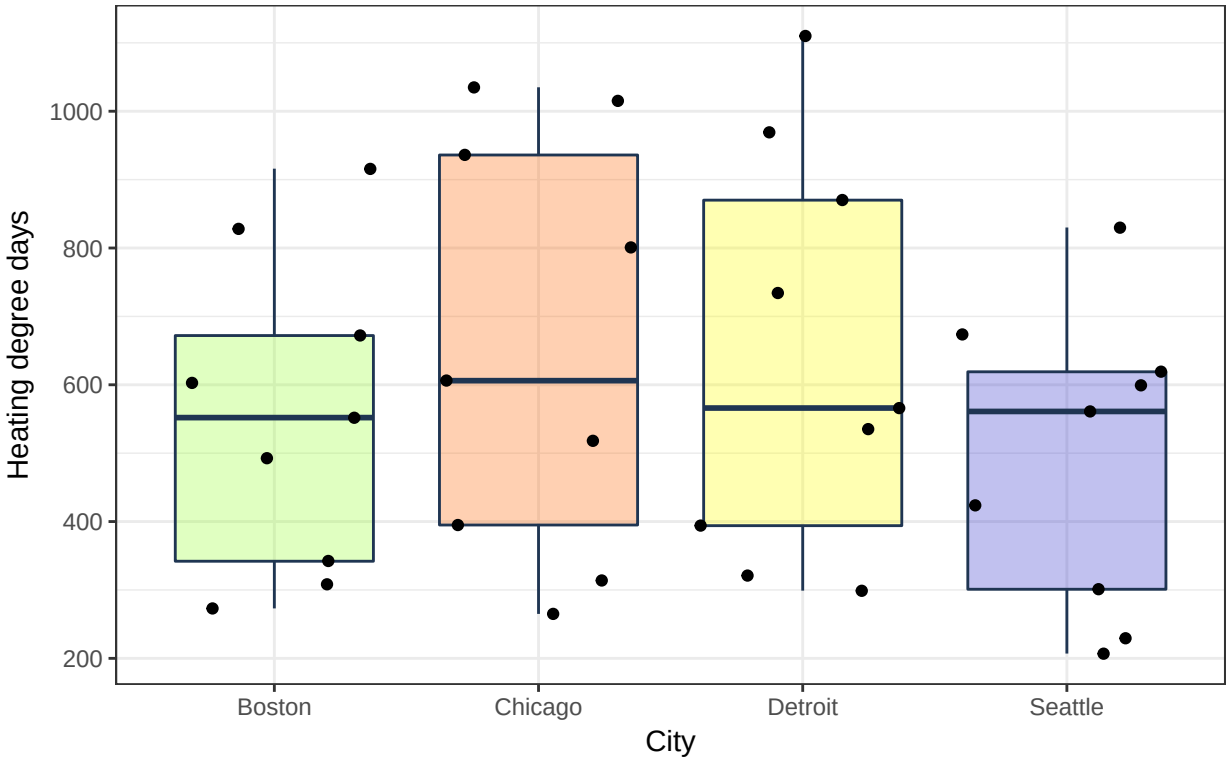
Figure 2.2 is the heating degree days in the season Summer. Mean value is generally the same and the ranges for Boston, Chicago and Detroit are similar. However the range for seattle is much smaller than the rest three. Our thought is, the temperature in Seattle in Summer is much warmer than the other three. However, Seattle has much difference of the 25th percentile than Boston which is on the other side of the coast but with pretty much the same longitude. So geographic location has some sort influence but whether the influence is significant or not requires further analysis.



2.3 Analysis of Heating Degree Days of Four Cities in Fall

Figure 2.3 is the heating degree days in the season fall. Mean value is generally in the same level for the four cities. There are several outliers in Chicago, Detroit, Seattle. The mean of Boston is in the same level as the 75th percentile but the variance is large. Compare to the figure 2.1 and figure 2.2, the heating degree days in season fall is obviously lower than spring and summer.

Figure2.4: 2014−2016 heating degree days of four cities in winter



2.4 Analysis of Heating Degree Days of Four Cities in Winter

Figure 2.4 is the heating degree days for winter. We can see inland cities have big difference with coast cities for the range, especially at the 25 percentile. Interestingly, the mean of these four cities are about the same, maybe because what we create is the average of three months. If we can take a look at one single month, there will be difference in the mean as well.

3. Conclusion

From four boxplot above we observe that the heating degree days in inland cities is significantly larger than coast cities and they have a larger variance. Also the heating degree days is smallest in fall and largest in spring. The variance of heating degree days is largest in summer. This results shows that the inland cities may have larger electricity demand than the coast cities. And it may have larger fluctuation in electricity demand in summer.