Ang Li

MA681

2017/10/13

Personal Assessment

In this EDA project, the main goal of our group is to find the distribution of cooling degree-day data in Florida in years of 2008, 2011 and 2014. To do this, we basically divide works into three parts: Data cleaning and manipulation, Data visualization and Model estimation. In data cleaning part, we using loops, R packages such as “dplyr” and “tidyverse” to read data in three years into R and arranged headers, rows and columns; In visualization part, we plotted and compared plots such as histograms, boxplots and ecdf and analyzed between graphs and years; In the end, we plotted the distribution graph for our data and calculated the estimated values for peaks and variances.

We did this project collaborate as a team; thus, I participated in every part in this project.

First, at the data cleaning part, since our data files were named in a complicated way that we cannot read into R, I changed all the names of data files into numbers according to their dates, which can be read methodically and correctly. After reading the data by years and loaded them into three list, I typed codes in R and divided them into three data frames so that we can use it efficiently later.

Second, at data visualization part, I plotted and analyzed every plot, such as density plot, boxplot, ecdf and histogram, individually in three years, in order to find the similarity and difference between years. In model estimation part,

Third, at model estimation model, I wrote codes to calculate the estimation of x values at its peak and variances for each year, which was a more specific way compare to visualization to analyze models ‘s differences between years.