

Prediction of ozone level in Boston

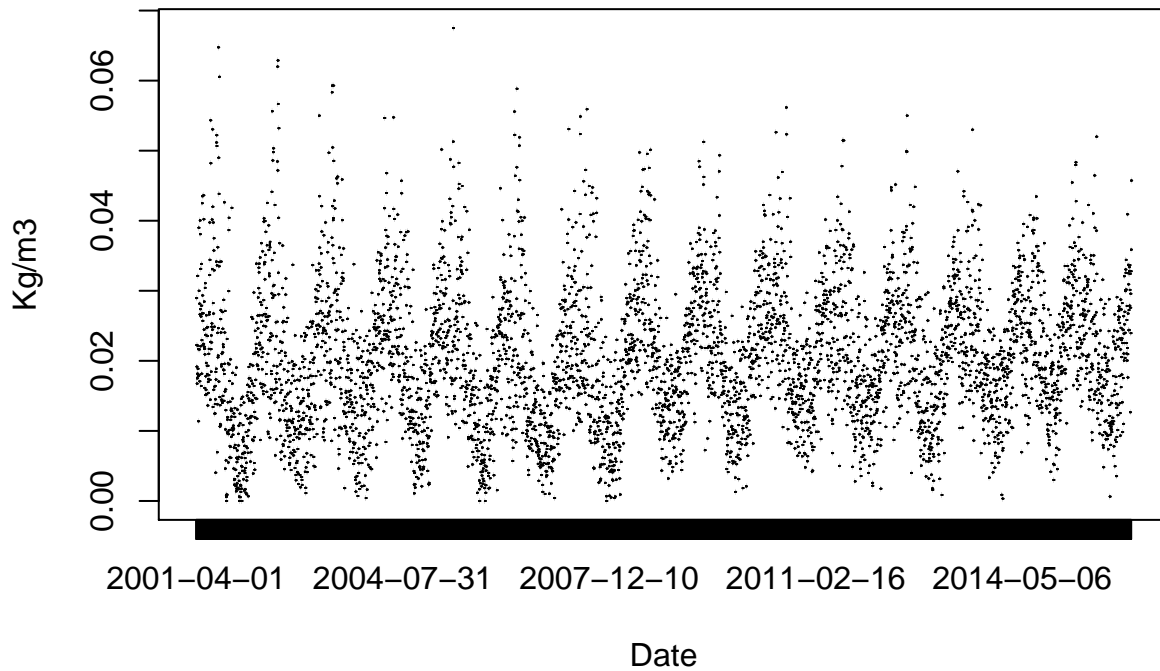
Lucas Emanuel Resck Domingues

Lucas Machado Moscheb*

Predicting O3 in Boston

Load and visualize

Daily average level of O3 in Boston



Data treatment

We noticed that some days do not exist in the dataset, for example, the day August 31, 2001 does not have information in the dataset.

##	X	City	State	Site.Num	Date.Local	O3.Mean
## 148	148	Boston	Massachusetts	42	2001-08-28	0.024583
## 149	149	Boston	Massachusetts	42	2001-08-29	0.015000
## 150	150	Boston	Massachusetts	42	2001-08-30	0.022333
## 151	151	Boston	Massachusetts	42	2001-09-01	0.021958
## 152	152	Boston	Massachusetts	42	2001-09-02	0.018750
## 153	153	Boston	Massachusetts	42	2001-09-03	0.028708

Also, there is duplicated days, as June 9, 2002:

*Escola de Matemática Aplicada

```
##           X   City           State Site.Num Date.Local  O3.Mean
## 412 412 Boston Massachusetts      42 2002-06-08 0.022917
## 413 413 Boston Massachusetts      42 2002-06-09 0.036190
## 414 414 Boston Massachusetts      42 2002-06-09 0.037000
## 415 415 Boston Massachusetts      42 2002-06-10 0.023389
```

The duplicated one is easier to deal, but the nan values are harder. First we calculate the mean value between the duplicated.

The rate of NA values is almost 5% of the dataset.

```
## [1] 0.04453367
```

Models

Now we develop some models using the train data.

The metric to compare is the Mean Absolute Error (MAE) in the predictions:

Decompose

Regression

Holt-Winters

ARMA