

Boston University
Department of Mathematics and Statistics

MA575 – Fall 2018
Linear Models
Lab Session 3: Data Visualization

**Project Report Due: Week of the 24th of September 2018 at
beginning of lab**

In this course you will learn to build regression models. It is used to estimate the relationships among observed variables. Regression analysis helps us understand the changes in one noisy variable given that the other variables are changed. Regression is commonly used for the purposes of prediction, inference and to determine statistical relationships between the variables (statistical significance).

The purpose of this lab session is to process and visualize datasets.

Pre-lab activities

Download the *Lab3* packet, including the R files and the AirQualityData.csv data.
Run the Lab3.R R script and examine the output.

Lab activities

1. Introduction to Air Quality data set.
2. Run the Lab3.R file in R studio step by step and analyze output.
3. Now, visualize and explore the datasets **Bike Sharing**, **Facebook** and **Forest Fires**. Read the DataSetDescription.txt file for each of the data sets. Use R to see the relationships between the variables. Discuss with your team.

Post-lab activities

You have a project deliverable, "Description of Potential Topic Areas" due at the beginning of lab in one week. At most one page, it should summarize your tentative topic (bike rentals, forest fires or social media metrics) and list interesting questions that can be approached by regression analysis, prediction and inference. **Please put at least 3 questions of interest and the rational behind them.**