

---

## MSDS596 PROJECT

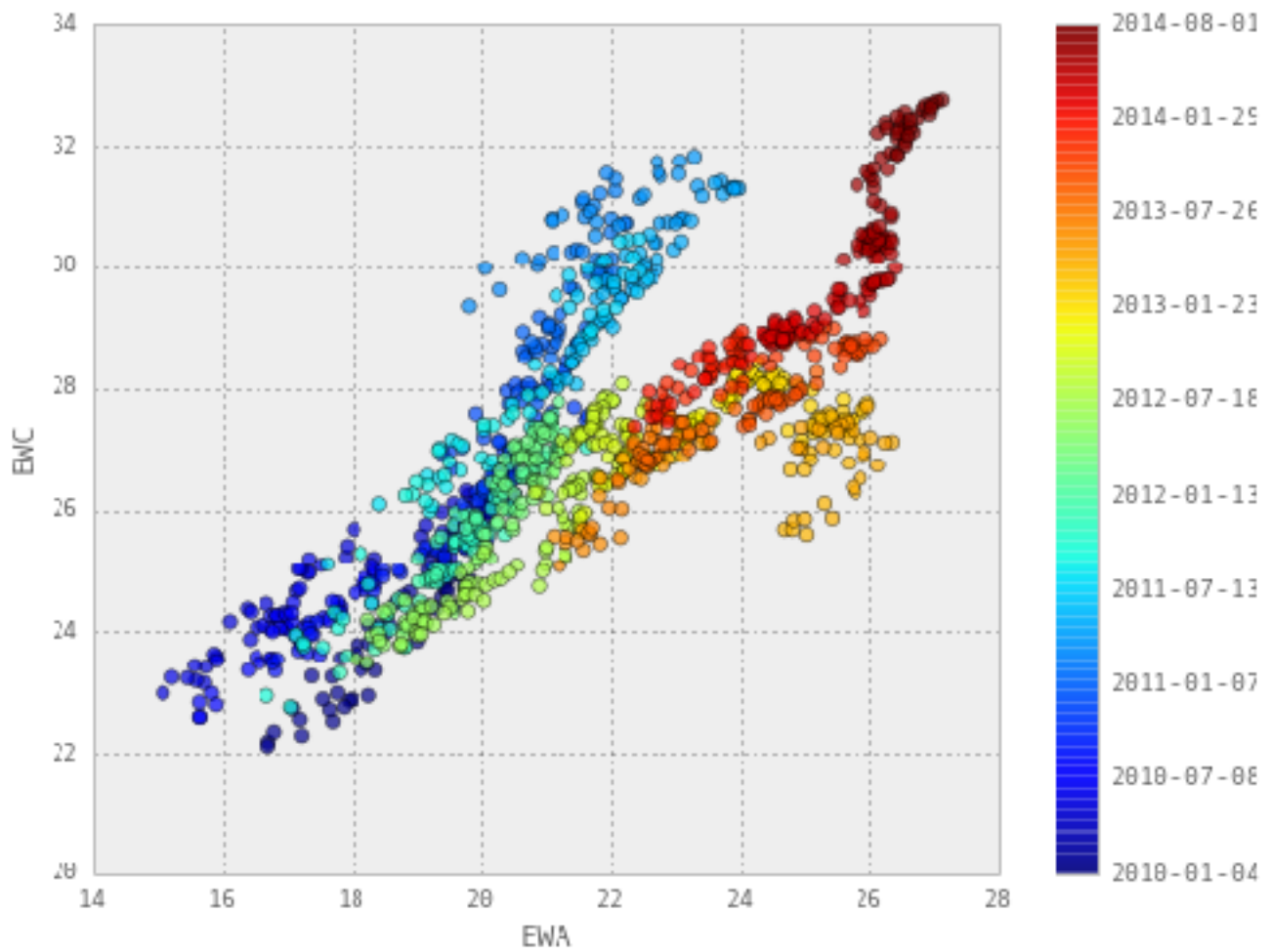


Photo Source: The Algo Engineer ([http://www.thealgoengineer.com/2014/online\\_linear\\_regression\\_kalman\\_filter](http://www.thealgoengineer.com/2014/online_linear_regression_kalman_filter))

---

# Regression Project Proposal

Presented by: Zhenzhen Ge and Linfeng He

October 30, 2016

## Introduction

As we look into demographical aspects of each state, we see widening gap among states in personal income and other factors such as life expectancy, illiteracy and educations, etc. To acquire the knowledges of disparity, we propose to conduct this research examining the relationship between several social-economical factors. Our major dataset will be scrapped and tidied from various online database sources including Bureau of Economic Analysis, U.S. Bureau of Labor Statistics and USA.gov. We anticipated the following factors which will take effects in our studies:

Population: estimate of population

Income: per capita income

Illiteracy: illiteracy

Life Exp: life expectancy in years

Murder: murder and non-negligent manslaughter rate per 100,000 population

HS Grad: percent high-school graduates

Frost: mean number of days with minimum temperature below freezing in capital or large city

Area: land area in square miles

## Research Phase& Expected Outcome

- Implement statistic methodology to testify variable/model and identify statistic confidence intervals;
  - Conduct model diagnostics on constant variance/normality/structure assumptions;
  - Examine outlier existence and adjust the model accordingly;
  - Perform step-wise selection or use information criterion to choose model;
  - Draw inference from the dataset basing on the fit of the model;
  - Change predictors and response to see potential relationship between the variables;
  - Obtain some meaningful conclusion from the rigorous analysis;
  - Demonstrate our findings via various data visualizations.
-