# YALEI DU
*MASTER OF STATISTICS(IN READING)*

Web : badbye.github.io
Email: yaleidu@163.com

## Profile

I'm a graduate student, currently study Statistics in University of Shanghai for Science and Technology. Interested in **Data Mining** and **Machine Learning**, as well as **Data Visualization**.

## Skills

### Data Processing
Familiar with many types of data: csv, json, xml. Able to handle unstructured data like text characters.

### Data Analysis
Perform basic descriptive and advanced inferential statistics(machine learning algorithms and statistical models).

### Data Presentation
Compose analysis report using html5 slide, including dynamic and interactive graphics.

## Technical

| | | |
|---|---|---|
| **R** | Excel | Git |
| SQL | Matlab/Octaves | Linux |

## Projects

### Calendar-heatmap of Zhengzhou

Collect Zhengzhou's temperture data in 2012 via API, both xml and json format; reshape this dataset to dataframe in R(just like a table in excel); Then visualize it as a heatmap using ggplot2.

### Counting word frequency in GCT's Enlish test papers

GCT refers to Graduate Candidate Test. Deal with English test papers in 2003 to 2013, the first part is counting each word's frequence of occurence, representing them in a Word-Cloud graph. The second part, treat these frequence as features building a SVM model to classify every reading materials(this part isn't showed in the link).

### More

- Computing Principle Component in R
- Non-parameter tests in R

## Education

**2009 - 2013:**
- Bachelor of Economics, School of Business, Zhengzhou University.

**2013 - Present:**
- Master of Statistics, College of Science, University of Shanghai for Science and Technology.