

# 1 Basic Info

This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por). In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful (see paper source for more details).

Attribute Information:

- 1 school - student's school
- 2 sex - student's sex
- 3 age - student's age
- 4 address - student's home address type
- 5 famsize - family size
- 6 Pstatus - parent's cohabitation status
- 7 Medu - mother's education
- 8 Fedu - father's education
- 9 Mjob - mother's job
- 10 Fjob - father's job
- 11 reason - reason to choose this school
- 12 guardian - student's guardian
- 13 traveltime - home to school travel time
- 14 studytime - weekly study time
- 15 failures - number of past class failures
- 16 schoolsup - extra educational support
- 17 famsup - family educational support
- 18 paid - extra paid classes within the course subject (Math or Portuguese)
- 19 activities - extra-curricular activities
- 20 nursery - attended nursery school
- 21 higher - wants to take higher education
- 22 internet - Internet access at home
- 23 romantic - with a romantic relationship
- 24 famrel - quality of family relationships
- 25 freetime - free time after school
- 26 goout - going out with friends
- 27 Dalc - workday alcohol consumption
- 28 Walc - weekend alcohol consumption
- 29 health - current health status
- 30 absences - number of school absences

## 2 Background and Motivation

This data set is using student's information to predict his/her grades in Math and Portuguese. We choose this data set because from ourselves experiences, learning a foreign language and learning the math both needs some "talents". But the talent needed for language and math seems different. We want to unveil the relation between students' info and their grades in math and Portuguese and also to verify our assumption that the "talents" for math and Portuguese are different.

## 3 Project Objectives

### 3.1 Build informative and vivid plots that helps understand the data

#### 3.1.1 Discover the trends and relationships

Any trend among the relationships? Any features are strongly correlated? good students always good?

### **3.1.2 Discover the outliers**

Any student stands out? or departure from the main trend? For example, for certain kind of student's background, he/she should get some good or bad grade, but his/her performance is far away from our expectation. Why this outliers happen? We expect our visualization will offer some evidence or hints to discover and help explain outliers.

### **3.1.3 Discover the feature Importance**

We could split the data according to some

## **4 Data**

We obtain our data from the UCI machine learning data repo:  
<http://archive.ics.uci.edu/ml/datasets/Student+Performance>

## **5 Data Processing**

The data contains many ordinary features, we need to translate these ordinary features to dummy variables.

## **6 Visualization Design**

### **6.1 Trends discovery**

1. parallel plot to show the trend relationships between features. 2. stream plot to show the trend of grads for different class of students.

### **6.2 Outlier discovery**

3. Apply multiple projection method to discover the best projection method and according to the projection method to discovery the

### **6.3 Feature Importance Discovery**

## **7 Must-Have Features**

1. Allow zoom-in and zoom-out for the trends' and outliers' detail. 2. Allow customized manipulation of data. 3. Apply distinguishable color map to represent the categories.

## **8 Optional Features**

## **9 Project Schedule**

## **10 Reference**

UCI data repo