

Trabalho da Disciplina MATA60 T01 - Banco de dados

## **Mineração de dados da base aberta do ENEM 2012**

### **1ª Etapa - Análise Estatística da Base**

**Mateus Cordeiro, Ramon Dias, Roberto Sales**

Departamento de Ciência da Computação (DCC)

Universidade Federal da Bahia

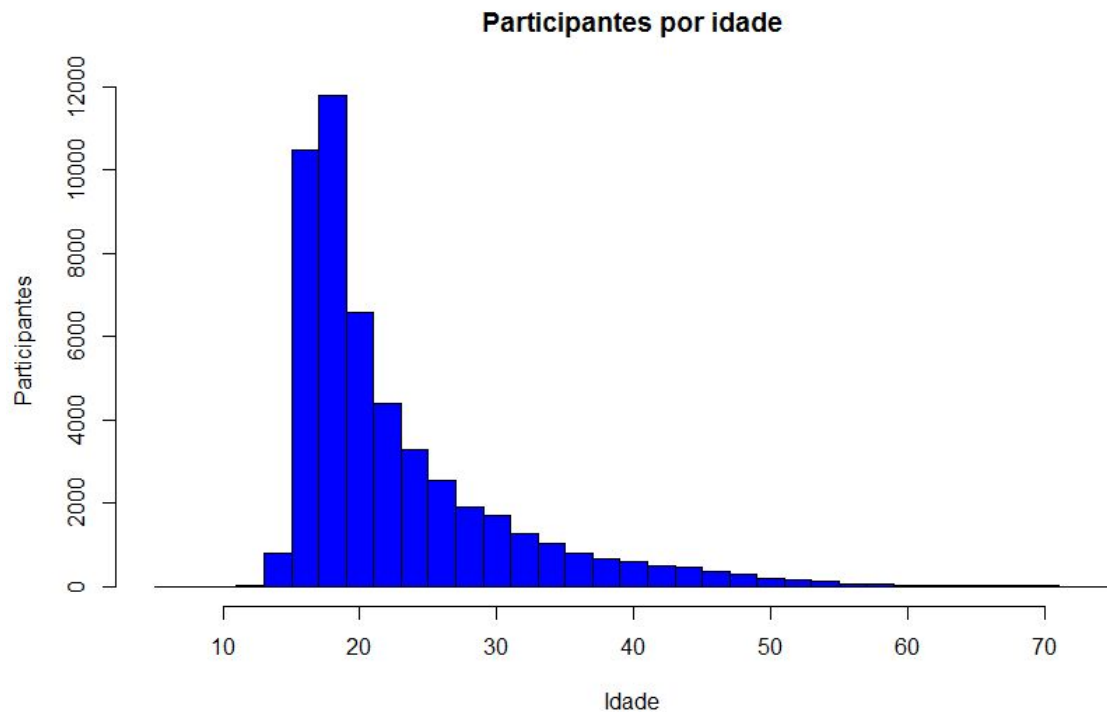
#### **Introdução**

A mineração de dados é um processo muito importante na sociedade atual, especialmente devido aos avanços tecnológicos que nos permitem gerar e armazenar uma grande quantidade de dados. Pois nesse contexto, as instituições investem para serem capazes de extrair informações valiosas a partir desses dados brutos, que irão guiar as suas ações e estratégias.

Nesse trabalho temos o objetivo de desenvolver e utilizar técnicas de mineração de dados sobre uma amostra de 50mil registros da base de dados aberta da ENEM 2012. Na primeira etapa do trabalho foram gerados diversos gráficos com o objetivo de prover um conhecimento mais aprofundado da base.

## Análise estatística

### Distribuição dos participantes por idade



O histograma acima descreve a distribuição dos participantes por faixas de idade de 2 anos. Percebe-se a predominância de participantes entre 16 e 20 anos e uma curva em declive bastante regular com o aumento da idade.

## Distribuição dos participantes por sexo



O gráfico acima mostra o número de participantes dos sexos masculino e feminino, existe uma superioridade considerável do número de mulheres.

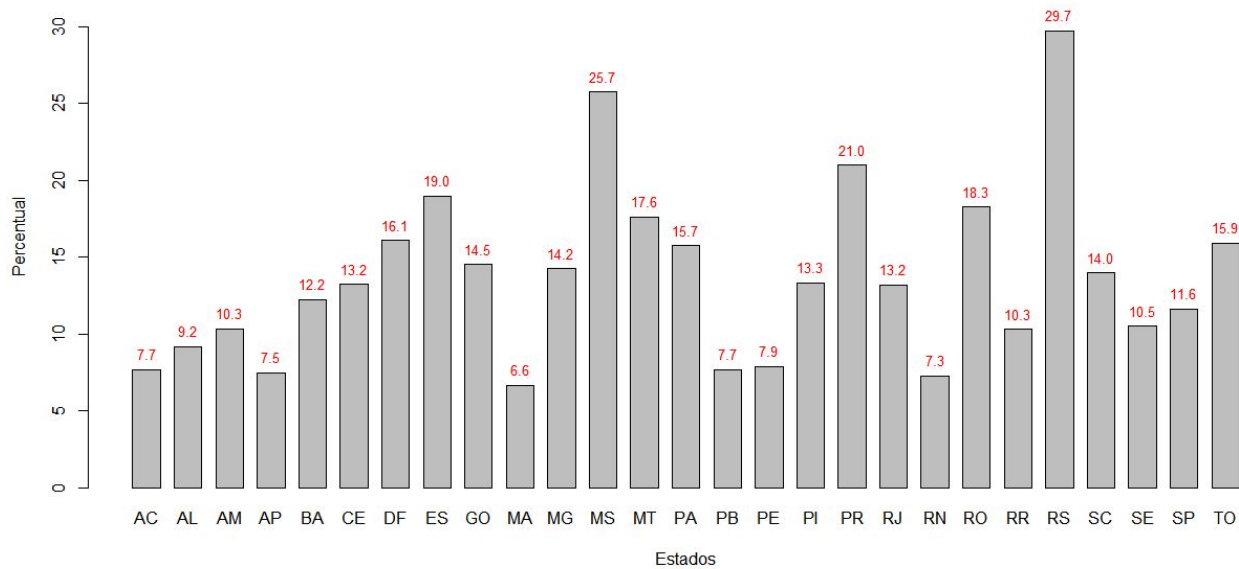
## Distribuição dos participantes por estado



O gráfico acima registra o número de participantes em cada estado mais o Distrito Federal. Como esperado o número de participantes é proporcional a população dos estados.

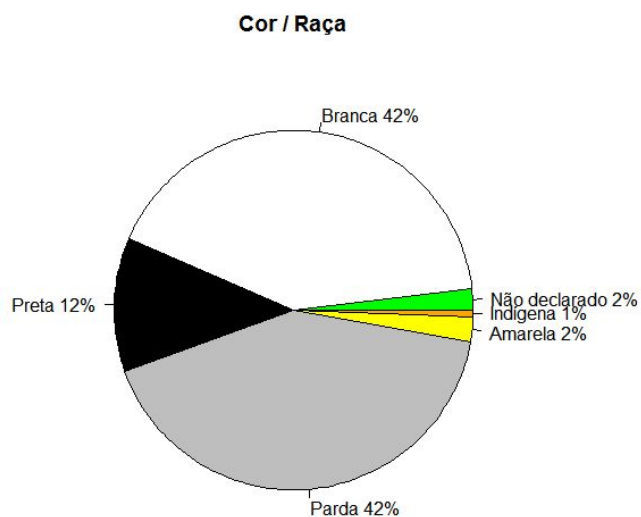
## Análise de solicitação de certificação no ensino médio

Percentual de participantes maiores de idade que solicitaram certificação de ensino médio



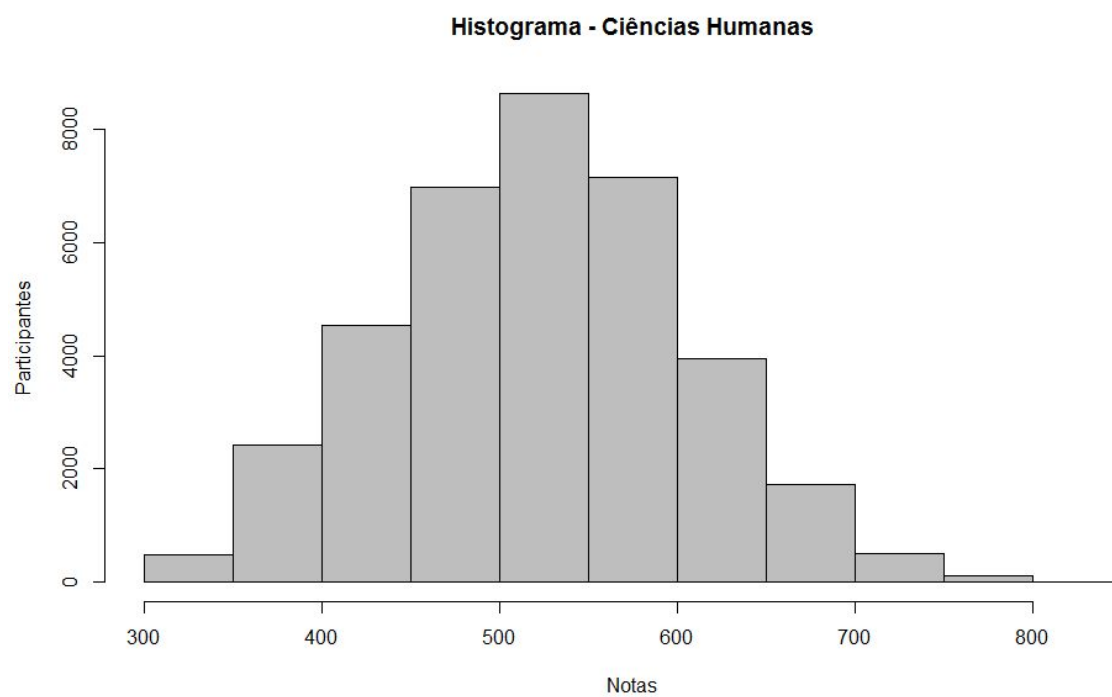
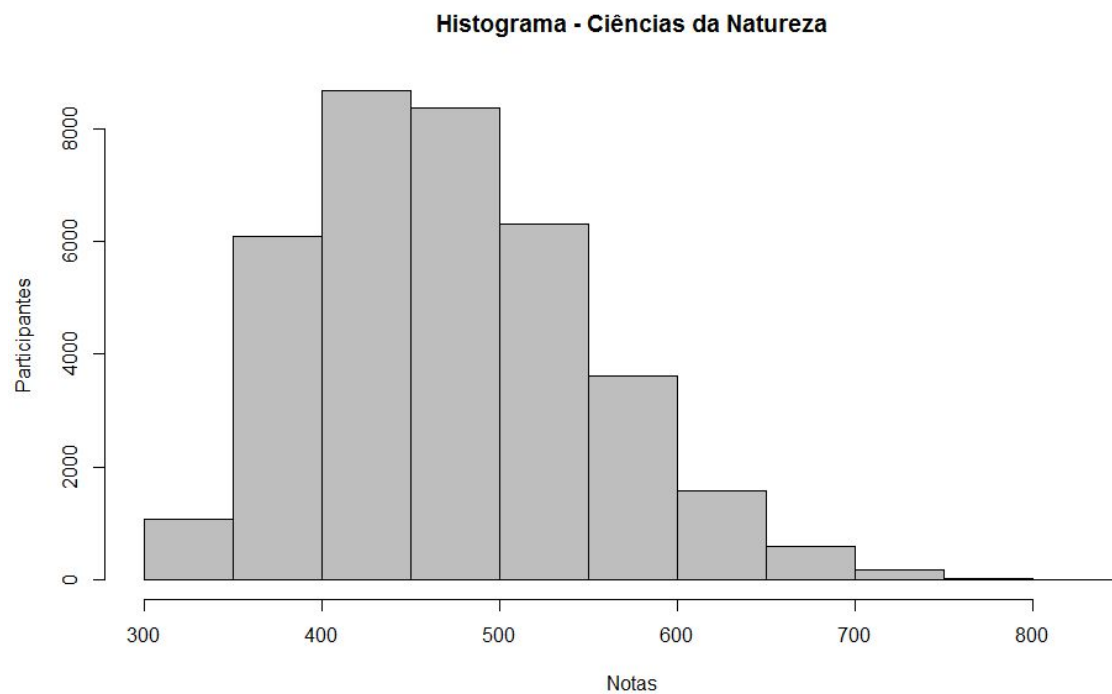
O gráfico mostra o percentual de participantes maiores de idade que não concluíram o ensino médio e que solicitaram essa certificação através do ENEM.

## Distribuição dos participantes por cor/raça

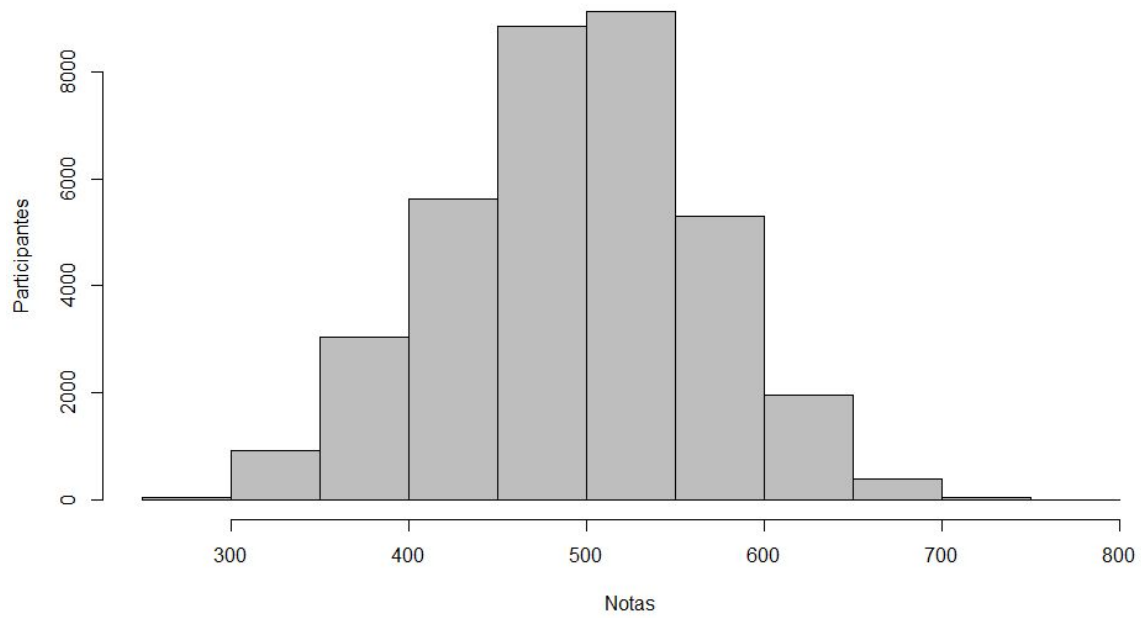


## Análise de notas por prova

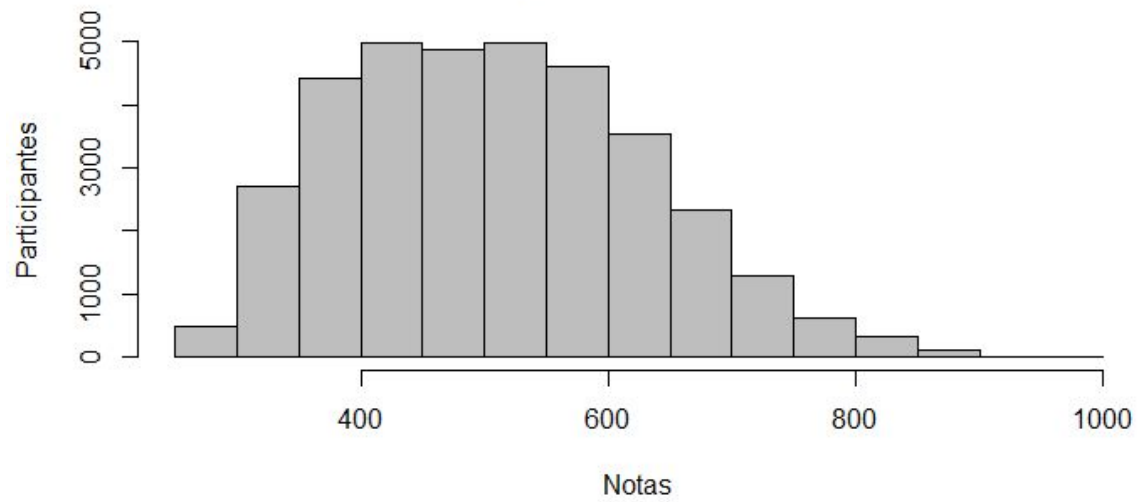
Os gráficos a seguir mostram a distribuição dos participantes por faixas de notas para cada prova do ENEM.

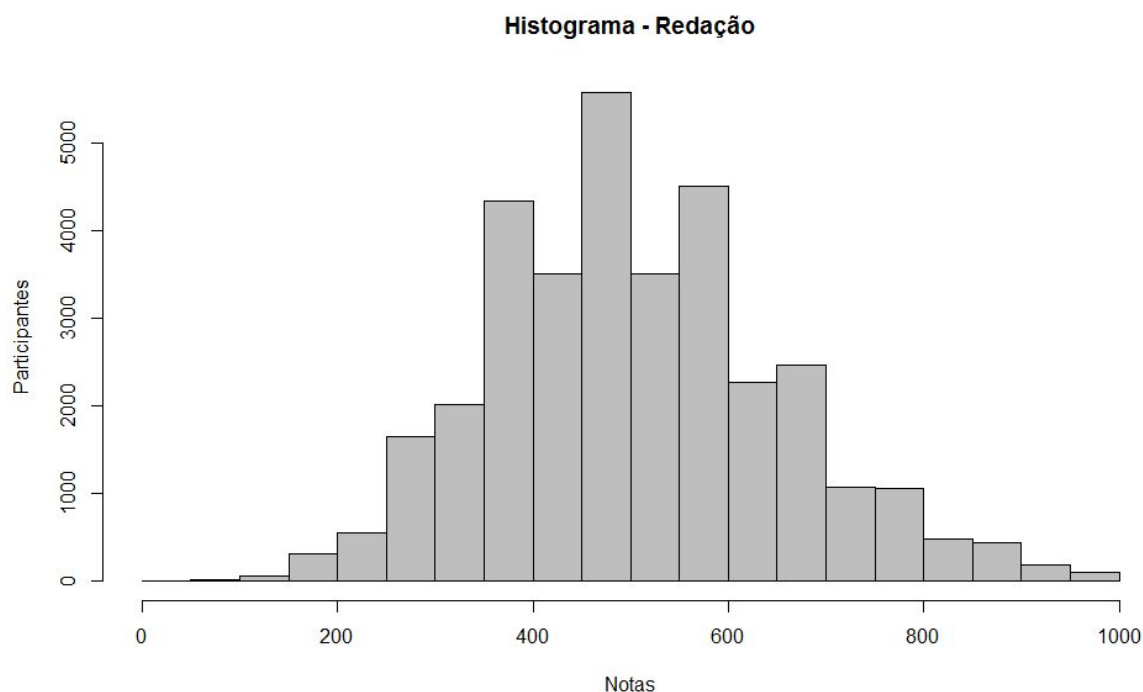


**Histograma - Linguagens e Códigos**



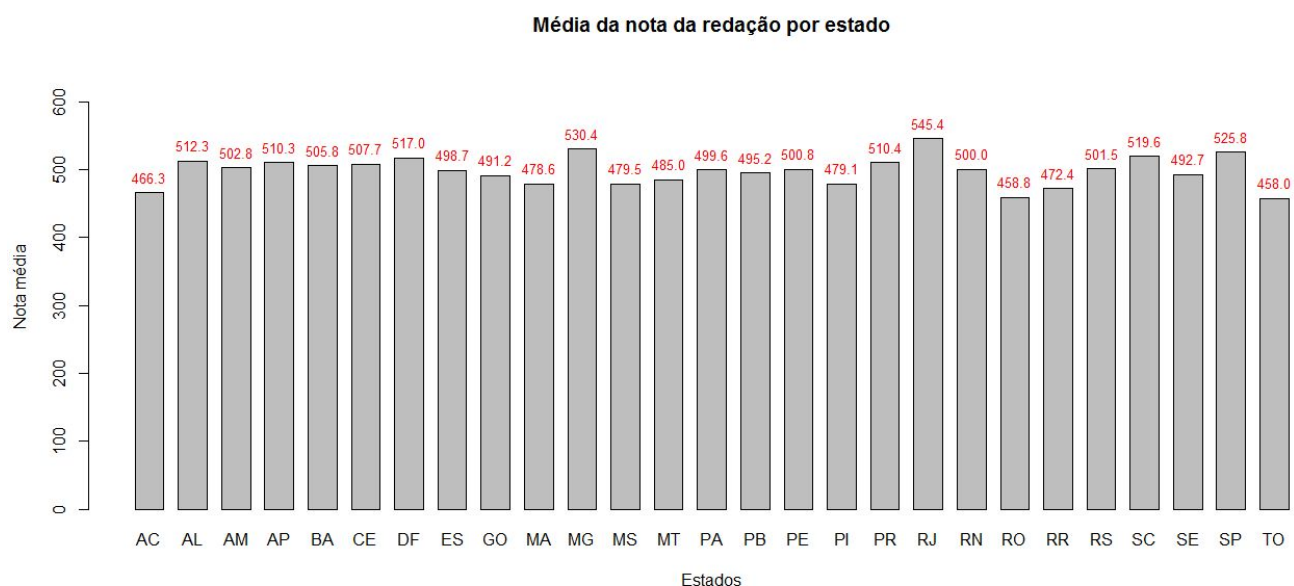
**Histograma - Matemática**





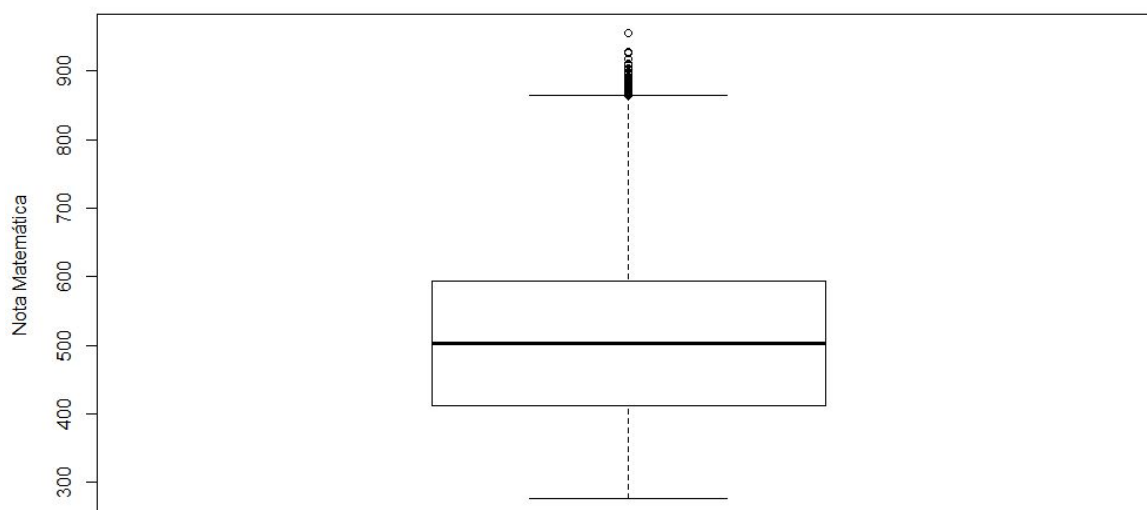
Os gráficos acima mostram que entre as provas objetivas, a de ciências da natureza é a qual maioria dos participantes tem o pior desempenho e que a de matemática tem a maior distribuição entre as faixas de notas.

## Nota média da redação por estado



Os três estados com maiores médias na redação são da região Sudeste

### Análise das notas da prova de matemática



Esse gráfico mostra que a mediana da nota de matemática está em torno de 500 pontos e que a maioria das notas se concentram entre 400 e 600. Pontuações acima de 850 são consideradas outliers.