



**Universidade de São Paulo
Escola de Artes, Ciências e Humanidades**

Integrantes:

André Ramos	9125339
Matheus Mendes Sant'Ana	8921666
Victor Luiz Soares Alexandre Pereira	8921350

**2º Exercício Programa
Análise de Dados com MapReduce**

**SÃO PAULO
2017**

Sumário

1) Introdução	3
2) Tutorial de Inicialização e Configuração	4
2.1) Linux – Ubuntu	4
2.1.1) Java Development Kit 8 (JDK)	4
2.1.2) SSH	5
2.1.3) Hadoop	7
2.1.4) Dica após reiniciar o ambiente de execução	12
2.2) Dados National Climactic Data Center (NCDC)	13
2.3) Execução do programa	14
3) Análise dos dados	15
3.1) Mappers e Reducers	15
3.2) Medidas estatísticas utilizadas	17
3.2.1) Média	17
3.2.2) Moda	18
3.2.3) Mínimo	18
3.2.4) Máximo	18
3.2.5) Variância	18
3.2.6) Desvio padrão	19
3.2.7) Desvio médio	19
3.3) Método dos Mínimos Quadrados (MMQ)	19
4) Breve descrição dos pacotes	20
4.1) Pacote “executar”	20
4.2) Pacote “graphics”	20
4.3) Pacote “operacoes”	20
4.4) Pacote “util”	20
5) Interface Gráfica	21
5.1) Atributos	21
5.1.1) Atributos não utilizados	21
5.1.2) Atributos numéricos utilizados	21
5.1.3) Atributos não-numéricos e qualitativos utilizados	22
6) Fluxo de Telas	23
7) Exemplos de testes de execução	28
8) Referências Bibliográficas	39

1) Introdução

O termo Big Data está cada vez mais em pauta na computação atual devido ao seu uso por grandes empresas e no manejo da informação. Lidar com um grande volume de dados, a velocidade, variedade, variabilidade e complexidade desses dados para se extrair informação significativa são desafios complexos para a análise de dados. Nesse Exercício Programa, utilizaremos uma solução Open Source que é o Hadoop da Apache (<http://hadoop.apache.org/>) para conseguir explorar o conjunto de dados (dados meteorológicos) fornecidos no enunciado do exercício.

O Hadoop é um *framework* para processamento e armazenamento massivo de dados em *clusters* de computadores. Será realizada uma implementação de um programa com interface gráfica para permitir a extração de informações do conjunto de dados através de medidas estatísticas (como por exemplo média, desvio padrão e moda), utilizando funções do paradigma MapReduce criadas pelo arcabouço de programação Java (tornando o processamento mais eficiente através de paralelização).

As etapas de Map e Reduce do conjunto de dados serão detalhadas para cada medida estatística, seguidas da lógica de programação envolvida (além de particularidades no desenvolvimento de código que serão comentadas). O ambiente utilizado para desenvolvimento foi uma máquina virtual (Linux) utilizando Eclipse. descrevemos como configurar o SO Linux passo a passo para instalar o Hadoop.

Por fim, apresentaremos imagens da execução da interface gráfica e disponibilizaremos um arquivo JAR para a execução do exercício programa. O código fonte também será disponibilizado, desenvolvido utilizando Eclipse, podendo ser utilizado para inspeção ou manipulação.

2) Tutorial de Inicialização e Configuração

2.1) Linux – Ubuntu

Para configurar o ambiente de desenvolvimento do Hadoop, utilizaremos o sistema operacional Ubuntu 17.04(Zesty Zapus), disponível no site <http://releases.ubuntu.com/17.04/>. Configure o usuário no Linux como “pc” (sem aspas) para facilitar a execução dos comandos. **Caso já tenha o Ubuntu instalado, troque os comandos com “pc” pelo nome de usuário utilizado.**

2.1.1) Java Development Kit 8 (JDK)

Baixe o JDK 8 para linux x64 (176.95MB) ou x86 (179.39MB) a depender da arquitetura do Sistema Operacional no seguinte site:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Abra o terminal e digite:

```
sudo apt-get purge openjdk-*
```

Criar pasta "Java" no diretório /home/pc/

Utilizando o terminal seria:

```
cd /home/pc/
```

```
mkdir Java
```

– O diretório utilizado será então /home/pc/Java/

Navegue até o diretório onde foi baixado o Java e extraia arquivo na pasta “Java” criada anteriormente (botão direito no arquivo e "Extrair aqui", ou no Prompt: tar -xvf jdk-8u131-linux-x64.tar.gz).

Exemplo utilizando o terminal:

```
cp jdk-8u131-linux-x64.tar.gz /home/pc/Java/
```

```
tar -xvf jdk-8u131-linux-x64.tar.gz
```

No terminal digitar:

```
cd /etc/  
sudo gedit profile
```

Abrirá um arquivo de texto. No fim do arquivo aberto, em uma nova linha, digitar:

```
JAVA_HOME=/home/pc/Java/pastaExtraida  
PATH=$PATH:$HOME/bin:$JAVA_HOME/bin  
export JAVA_HOME  
export PATH
```

(Para o nosso exemplo, a primeira linha ficou: **JAVA_HOME=/home/pc/Java/jdk1.8.0_131**)

– Salve e feche o arquivo

Para simplificar, abaixo utilizaremos o diretório-exemplo **"/home/pc/Java/jdk1.8.0_131"** para denotar os comandos de execução. Entre no terminal e digite:

```
cd /home/pc/Java/jdk1.8.0_131  
./etc/profile
```

```
sudo update-alternatives --install "/usr/bin/java" "java" "/home/pc/Java/jdk1.8.0_131/bin/java" 1  
sudo update-alternatives --install "/usr/bin/javac" "javac" "/home/pc/Java/jdk1.8.0_131/bin/javac" 1  
sudo update-alternatives --install "/usr/bin/javaws" "javaws"  
"/home/pc/Java/jdk1.8.0_131/bin/javaws" 1
```

```
sudo update-alternatives --set java /home/pc/Java/jdk1.8.0_131/bin/java  
sudo update-alternatives --set javac /home/pc/Java/jdk1.8.0_131/bin/javac  
sudo update-alternatives --set javaws /home/pc/Java/jdk1.8.0_131/bin/javaws
```

Para testar se o Java foi instalado corretamente, execute o comando seguinte no terminal:

```
java -version
```

2.1.2) SSH

Abra um novo terminal no Ubuntu e digite:

```
sudo addgroup hadoop  
sudo adduser --ingroup hadoop hduser
```

Digite a senha e a confirmação da senha (recomendada a mesma senha do root Linux). Pressione "enter" para Full Name, Room Number, Work Phone, Home Phone e Other (tudo *default*). Digite "y" para confirmar a operação e pressione "enter".

Execute no mesmo terminal:

sudo apt-get install ssh

Digite os dois comandos abaixo, caso não ocorra erro e apareça um caminho como output do comando em ambos, tudo está corretamente configurado:

which ssh

(saída desejada: /bin/ssh)

which sshd

(saída desejada: /sbin/sshd)

Logo após verificar a configuração, execute no mesmo terminal:

su hduser

(digite a senha configurada no início desta seção para poder continuar)

ssh-keygen -t rsa -P ""

Digite "enter" para salvar a chave no diretório padrão. Por precaução, guarde em um arquivo texto os diretórios do **arquivo de identificação** e do **arquivo de chave pública** (No nosso exemplo, o arquivo de identificação está no diretório /home/hduser/.ssh/id_rsa e o arquivo de chave pública está no diretório /home/hduser/.ssh/id_rsa.pub).

cat \$HOME/.ssh/id_rsa.pub >> \$HOME/.ssh/authorized_keys

ssh localhost

Digite "yes" e pressione "enter"

Instale também o Openbox:

sudo apt-get install xorg openbox

2.1.3) Hadoop

Para instalar e configurar o Hadoop, utilizaremos a versão *stable* mais atualizada do hadoop que possui um tutorial em sua documentação no site do Apache: <https://hadoop.apache.org/docs/r2.8.0/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>

Siga os passos abaixo detalhadamente:

Abra um novo diretório e um novo terminal e digite:

sudo adduser hduser sudo

Crie uma pasta no diretório /home/pc com nome "Hadoop": o diretório utilizado como exemplo a partir de agora será /home/pc/Hadoop

Exemplo:

**cd /home/pc/
mkdir Hadoop**

Baixe o hadoop 2.8.0 (binário) no site

<http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.8.0/hadoop-2.8.0.tar.gz>

(pode ser qualquer *mirror*). Copie o arquivo compactado no diretório /home/pc/Hadoop e extraia neste mesmo diretório (botão direito no arquivo e "Extrair aqui") ou dê comandos pelo terminal.

Exemplo utilizando o terminal:

Navegue com o comando “**cd**” até a pasta onde foi baixado o Hadoop (Ex:Downloads) e execute:

**cp hadoop-2.8.0.tar.gz /home/pc/Hadoop/
cd /home/pc/Hadoop/
tar -xvf hadoop-2.8.0.tar.gz**

Execute os seguintes comandos no terminal:

**cd /usr/local
sudo mkdir hadoop**

**cd /home/pc/Hadoop
sudo mv * /usr/local/hadoop**

sudo su hduser

sudo chown -R hduser:hadoop /usr/local/hadoop

update-alternatives --config java

Selecione o caminho similar a /home/pc/Java/jdk1.8.0_131/bin/java, no nosso exemplo digitamos "2" e pressionamos "enter" pois era a opção de número 2.

sudo gedit ~/.bashrc

No fim do arquivo aberto, em uma nova linha, insira:

#HADOOP VARIABLES START

export JAVA_HOME=/home/pc/Java/jdk1.8.0_131

export HADOOP_INSTALL=/usr/local/hadoop/hadoop-2.8.0

export PATH=\$PATH:\$HADOOP_INSTALL/bin

export PATH=\$PATH:\$HADOOP_INSTALL/sbin

export HADOOP_MAPRED_HOME=\$HADOOP_INSTALL

export HADOOP_COMMON_HOME=\$HADOOP_INSTALL

export HADOOP_HDFS_HOME=\$HADOOP_INSTALL

export YARN_HOME=\$HADOOP_INSTALL

export HADOOP_COMMON_LIB_NATIVE_DIR=\$HADOOP_INSTALL/lib/native

export HADOOP_OPTS="-Djava.library.path=\$HADOOP_COMMON_LIB_NATIVE_DIR"

#HADOOP VARIABLES END

Salve o arquivo e feche-o.

Digite no terminal:

source ~/.bashrc

Para confirmar que o Java está corretamente configurado, execute os seguintes comandos:

1º **javac -version**

2º **which javac**

3º **readlink -f /usr/bin/javac**

Saída esperada dos comandos:

1º **javac 1.8.0_131**

2º **/usr/bin/javac**

3º **/home/pc/Java/jdk1.8.0_131/bin/javac**

Abrir o seguinte arquivo pelo terminal:

```
sudo gedit /usr/local/hadoop/hadoop-2.8.0/etc/hadoop/hadoop-env.sh
```

Modificar o arquivo aberto na linha onde há

export JAVA_HOME=\${JAVA_HOME} substituir por

```
export JAVA_HOME=/home/pc/Java/jdk1.8.0_131
```

Salve o arquivo e feche-o.

Digite no terminal:

```
sudo mkdir -p /app/hadoop/tmp
```

```
sudo chown hduser:hadoop /app/hadoop/tmp
```

```
sudo gedit /usr/local/hadoop/hadoop-2.8.0/etc/hadoop/core-site.xml
```

Incluir no arquivo aberto as seguintes linhas **dentro das tags** <configuration> </configuration>:

```
<property>
```

```
<name>fs.defaultFS</name>
```

```
<value>hdfs://localhost:9000</value>
```

```
</property>
```

```
<property>
```

```
<name>hadoop.tmp.dir</name>
```

```
<value>/app/hadoop/tmp</value>
```

```
</property>
```

Salve o arquivo e feche-o.

Copie e abra o seguinte arquivo:

```
sudo cp /usr/local/hadoop/hadoop-2.8.0/etc/hadoop/mapred-site.xml.template
```

```
/usr/local/hadoop/hadoop-2.8.0/etc/hadoop/mapred-site.xml
```

```
sudo gedit /usr/local/hadoop/hadoop-2.8.0/etc/hadoop/mapred-site.xml
```

Incluir no arquivo aberto as seguintes linhas **dentro das tags** <configuration> </configuration>:

```
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
```

Salve o arquivo e feche-o.

Abaixo criaremos os diretórios dos nós. Atente-se para os comandos de permissão (**chown**) ao usuário “hduser” que criamos, pois serão essenciais para que a conexão seja aceita. No terminal digite:

```
sudo mkdir -p /usr/local/hadoop_store/hdfs/namenode
sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode
sudo chown -R hduser:hadoop /usr/local/hadoop_store
sudo gedit /usr/local/hadoop/hadoop-2.8.0/etc/hadoop/hdfs-site.xml
```

Incluir no arquivo aberto as seguintes linhas **dentro das tags** `<configuration>` `</configuration>`:

```
<property>
<name>hadoop.tmp.dir</name>
<value>/app/hadoop/tmp</value>
</property>
```

```
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
```

Salve o arquivo e feche-o.

No terminal execute:

```
sudo gedit /usr/local/hadoop/hadoop-2.8.0/etc/hadoop/yarn-site.xml
```

Incluir no arquivo aberto as seguintes linhas **dentro das tags** `<configuration>` `</configuration>`:

```
<property>
<name>yarn.scheduler.minimum-allocation-mb</name>
<value>128</value>
</property>
```

```

<property>
  <name>yarn.scheduler.maximum-allocation-mb</name>
  <value>2048</value>
</property>

<property>
  <name>yarn.scheduler.minimum-allocation-vcores</name>
  <value>1</value>
</property>
<property>
  <name>yarn.scheduler.maximum-allocation-vcores</name>
  <value>2</value>
</property>

<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
  <value>4096</value>
</property>
<property>
  <name>yarn.nodemanager.resource.cpu-vcores</name>
  <value>4</value>
</property>

<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>

```

Salve o arquivo e feche-o.

Formate o hdfs:

hdfs namenode -format

Iniciando o Hadoop Cluster:

Para iniciar o hadoop:

start-dfs.sh

start-yarn.sh

Para finalizar a execução do hadoop:

stop-yarn.sh

stop-dfs.sh

Para checar a versão do Hadoop:

hadoop version

Atenção: Para retomar o ambiente de execução do Hadoop após desligar o computador, por exemplo, você deve executar o seguinte comando em um novo terminal e posteriormente iniciar o hadoop (com os comandos “start-dfs.sh” e “start-yarn.sh”): **sudo su hduser**

2.1.4) Dica após reiniciar o ambiente de execução

Para corrigir eventuais falhas após reiniciar o sistema (seja para executar o hadoop ou para executar o .JAR), tente realizar os passos a seguir em um novo terminal ou interrompa antes a execução do hadoop (stop-yarn.sh e stop-dfs.sh):

sudo su hduser

sudo rm -R /tmp/*

sudo rm -R /app/hadoop/tmp

sudo mkdir -p /app/hadoop/tmp

sudo chown hduser:hadoop /app/hadoop/tmp

hdfs namenode -format

hadoop namenode -format

start-yarn.sh

start-dfs.sh

Também tente deletar as pastas de /input e /output criadas no **hdfs** (Hadoop Distributed File System), **caso mude a entrada de dados contida na pasta Data** (esta pasta será criada no tópico **2.3**):

hadoop fs -rm /input/*

hadoop fs -rm /output/*

hadoop fs -copyFromLocal /home/pc/Data/* /input

2.2) Dados National Climactic Data Center (NCDC)

Para realizar o download de todos os dados anuais (cerca de 20GB), abra um terminal e escolha a pasta destino desejada (no nosso caso “Desktop”) e digite no terminal:

```
sudo wget -r --no-parent ftp://ftp.ncdc.noaa.gov/pub/data/gsod/ url -P /home/pc/Desktop
```

Para realizar o download de um dado ano (mude o caminho com o ano desejado) para a pasta Desktop digite conforme o exemplo:

```
sudo wget -r --no-parent ftp://ftp.ncdc.noaa.gov/pub/data/gsod/2001 url -P /home/pc/Desktop
```

- Os dados estarão na pasta [ftp.ncdc.noaa.gov/pub/data/gsod/](ftp://ftp.ncdc.noaa.gov/pub/data/gsod/) dentro do local escolhido

Extraia os arquivos .gz:

```
sudo gunzip -r /home/pc/Desktop/ftp.ncdc.noaa.gov/pub/data/gsod/ -v
```

2.3) Execução do programa

Siga em ordem os passos a seguir:

1-) Cole o arquivo **EP.jar** em **/home/pc**

2-) Cole os arquivos **jcommon-1.0.23.jar** e **jfreechart-1.0.19.jar** em **/home/pc** – Esses arquivos podem ser obtidos através da pasta “lib” do arquivo .zip disponível no site: <https://sourceforge.net/projects/jfreechart/files/1.0.19/JFreeChart/1.0.19/jfreechart-1.0.19.zip/download>

3-) Copie todos os arquivos de dados meteorológicos desejados – **somente os arquivos sem as pastas** – para um diretório, por exemplo **/home/pc/Data**

4-) Abra um novo terminal em **/home/pc** e digite **sudo su hduser**

5-) Inicie o hadoop através dos comandos **start-yarn.sh** e **start-dfs.sh** e aguarde até o fim da execução.

6-) Execute o seguinte comando para criar o diretório de entrada no hadoop: **hadoop fs -mkdir /input**

7-) Execute o seguinte comando para criar o diretório de saída no hadoop: **hadoop fs -mkdir /output**

8-) Adicione todos os arquivos do diretório do item 3 no diretório de entrada do hadoop (exemplo: **/home/pc/Data**), através do seguinte comando: **hadoop fs -copyFromLocal /home/pc/Data/* /input**

9-) Referenciando os 2 arquivos do item 2 no hadoop (primeira etapa):

```
export LIBJARS=/home/pc/jcommon-1.0.23.jar:/home/pc/jfreechart-1.0.19.jar
```

10-) Referenciando os 2 arquivos do item 2 no hadoop (segunda etapa):

```
export HADOOP_CLASSPATH=/home/pc/jcommon-1.0.23.jar:/home/pc/jfreechart-1.0.19.jar
```

11-) Execute o seguinte comando para dar a permissão do diretório **/home/pc** ao hadoop:

```
sudo chown -R hduser:hadoop /home/pc
```

12-) Execute o programa final:

```
hadoop jar EP.jar executar.Testar /input /output/ -libjars ${LIBJARS} -mytoolopt value
```

13-) Após terminar de executar o programa, devolva o acesso de **/home/pc** ao usuário padrão:

```
sudo chown -R pc:pc /home/pc.
```

Para executar o programa novamente no mesmo terminal, só é necessário seguir os itens 11 e 12 novamente.

Para executar o programa em um novo terminal, siga os 13 passos novamente. **Para finalizar a execução do programa, basta fechar a tela atual da interface gráfica.**

3) Análise dos dados

Além das medidas estatísticas de média e desvio padrão, foram implementadas outras funcionalidades importantes: O valor **mínimo** e **máximo** de uma variável, a **variância**, a **moda** e o **desvio médio**. Para cada métrica, explicaremos como o cálculo é realizado e a respectiva abordagem no algoritmo do MapReduce. O algoritmo do cálculo das estatísticas foi dividido em vários *Mappers* e *Reducers* como descrito a seguir:

3.1) Mappers e Reducers:

MapCont e ReduceCont:

Este conjunto de operandos tem como objetivo devolver o **número de linhas (ou “registros”)** agrupados por alguma forma de data (por ano, mês, dia do mês ou dia da semana). O Mapper envia como chave o **agrupamento** desejado (ano, mês, dia do mês ou dia da semana) e envia **“1”** como valor, pois o intuito é capturar a presença do registro. No Reducer, a chave emitida para o arquivo de saída é a mesma do Mapper (ou seja, a chave é o **agrupamento**) e o valor é a **soma dos valores** de todos os Mappers. Por fim, teremos o número **“n”** de linhas agrupadas pelo tipo de data desejado (ano, mês, dia do mês ou dia da semana).

MapDifMod e ReduceDifMod:

Conjunto de operandos utilizado no cálculo do **desvio médio**. Trata-se do numerador do desvio médio: a soma dos módulos da diferença entre o **valor do atributo** e a **média do atributo**. No Mapper, a chave emitida é o **agrupamento** desejado (ano, mês, dia do mês ou dia da semana) e o valor emitido é o **valor do atributo desejado concatenado** – através do caractere *underline* – **com o valor da média** correspondente ao seu agrupamento (se o agrupamento é por ano e a linha atual tem ano 2017, então a média concatenada ao valor do atributo atual será a média de 2017).

No Reducer, a chave emitida é o **agrupamento** (mesma chave que a emitida pelo Mapper), e será criada uma **variável de saída para armazenar a soma dos módulos das diferenças** (cada diferença é o valor “xi” menos a média correspondente ao seu agrupamento). Continuando ainda na explicação do Reducer, cada valor recebido terá a seguinte verificação: se contiver o caractere *underline*, será feita a diferença entre o valor antes de *underline* (“xi”) e depois de *underline* (a média), somando o módulo dessa diferença à variável de saída. Caso não haja *underline* no valor, ele será simplesmente somado à variável de saída (pois a operação de diferença e módulo já foi feita anteriormente). O Reducer emitirá para o arquivo de saída o **agrupamento** e a **variável de saída correspondente** (soma dos módulos das diferenças para o

agrupamento).

MapDifQuad e ReduceDifQuad:

O princípio destes operandos é muito similar aos operandos **MapDifMod** e **ReduceDifMod**: se distingue deles apenas no Reducer, onde é feita a **soma dos quadrados das diferenças** e não a **soma dos módulos das diferenças**. No Mapper a chave emitida é o **agrupamento** desejado (ano, mês, dia do mês ou dia da semana) e o valor emitido é o **valor do atributo atual concatenado** – através do caractere *underline* – **com o valor da média correspondente ao agrupamento**.

No Reducer a chave emitida é a mesma que a do Mapper. O valor emitido é a soma dos quadrados das diferenças entre o número à esquerda de *underline* (“xi”) e o número à direita de *underline* (a média correspondente). Assim como no **ReduceDifMod**, se não houver o caractere *underline* no valor, ele apenas será somado à variável de saída.

MapMax e ReduceMax:

No Mapper a chave emitida é o **agrupamento** desejado (por ano, mês, dia do mês ou dia da semana) e o valor emitido é o **valor do atributo atual**. No Reducer a chave de saída é a mesma do Mapper (**agrupamento**), enquanto o valor emitido é o **maior valor** de todos os Mappers recebidos.

MapMin e ReduceMin:

No Mapper a chave é o **agrupamento** desejado (por ano, mês, dia do mês ou dia da semana) e o valor emitido é o **valor do atributo atual**. No Reducer a chave de saída é a mesma do Mapper (**agrupamento**) e o valor emitido é o **menor valor** de todos os Mappers recebidos.

MapModa1 e ReduceModa1:

No Mapper a chave emitida é o **agrupamento concatenado** – através do caractere *underline* – **com o valor do atributo** (por ano, mês, dia do mês ou dia da semana) e o valor emitido é **“1”**. No Reducer, a chave emitida é a mesma chave recebida do Mapper (**agrupamento concatenado com o valor do atributo**) e o valor final emitido para cada Reducer é a soma dos valores de cada Mapper recebido. Como exemplo, poderíamos ter a seguinte linha em um arquivo de saída: **2017_A 8** (agrupando por ano: significa que existem **8** valores “A” para **2017**).

MapModa2 e ReduceModa2:

Este conjunto de operações se diferencia dos demais principalmente pelo tipo de **arquivo de entrada recebido**: ao invés do Mapper receber os arquivos dos dados meteorológicos, ele recebe os dados pré-processados por **MapModa1** e **ReduceModa1**. No Mapper a chave emitida é o **agrupamento** desejado (por ano, mês, dia do mês ou dia da semana), e o valor emitido é o **valor do atributo concatenado** – através do caractere *underline* – **com a quantidade de vezes que ele aparece no agrupamento** (exemplo: para uma linha do arquivo igual a 2017_A 8, a chave emitida será 2017 e o valor emitido será A_8).

O Reducer emitirá como chave a mesma chave recebida do Mapper (**agrupamento**), e como valor ele emitirá o valor do Mapper que contém o maior número à direita de *underline*, exemplo: se um Reducer receber um Mapper igual a **2017, A_8** (par [chave, valor]) e outro Mapper igual a **2017, B_19**; o resultado final escrito será **2017, B_19** (segundo Mapper) pois 19 é maior do que 8. Para apresentarmos na interface o resultado da moda pra o usuário, basta exibirmos os dados do arquivo reduzido sem o número após *underline* (que denota a quantidade de vezes que o termo aparece).

MapSoma e ReduceSoma:

No Mapper a chave emitida é o **agrupamento** desejado (por ano, mês, dia do mês ou dia da semana) e o valor emitido é o **valor do atributo** atual. No Reducer, a chave emitida é a mesma chave recebida do Mapper (**agrupamento**) e o valor emitido é o **resultado da soma dos valores** de todos os Mappers recebidos.

3.2) Medidas estatísticas utilizadas:

3.2.1) Média

Utiliza **2 operações de MapReduce** (no pior dos casos). A média de um conjunto de valores numéricos é calculada **somando-se todos os valores do atributo** em questão e **dividindo a soma pelo número de elementos**.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Para realizar o cálculo da média, são utilizados [1] **MapSoma** e **ReduceSoma** para calcular a soma

total do atributo e [2] **MapCont** e **ReduceCont** para calcular a quantidade total de elementos. Por fim, após gerarmos os 2 arquivos – um **agrupando as somas** e o outro **agrupando a quantidade total de elementos** – podemos gerar um arquivo final, agrupando as somas sobre os elementos correspondentes, o que nos resultará a média (seja agrupada por ano, mês, dia do mês ou dia da semana – a depender da escolha feita pelo usuário).

3.2.2) Moda

Utiliza **2 operações de MapReduce** (no pior dos casos). É o valor que mais se repete para um determinado atributo. Para o nosso programa, só pode ser gerada para atributos não-numéricos. Utiliza [1] **MapModa1** e **ReduceModa1** e [2] **MapModa2** e **ReduceModa2**.

3.2.3) Mínimo

Utiliza **1 operação de MapReduce** (no pior dos casos). Retorna o menor valor para o atributo escolhido, agrupando por ano, mês, dia do mês ou dia da semana. Utiliza [1] **MapMin** e **ReduceMin**.

3.2.4) Máximo

Utiliza **1 operação de MapReduce** (no pior dos casos). Retorna o maior valor para o atributo escolhido, agrupando por ano, mês, dia do mês ou dia da semana. Utiliza [1] **MapMax** e **ReduceMax**.

3.2.5) Variância

Utiliza **3 operações de MapReduce** (no pior dos casos). Como foi dito no tópico 3.2.1, a média utiliza **2** operações de MapReduce. A variância utilizará, além dos MapReduces para gerar a média, uma nova operação de MapReduce – **MapDifQuad** e **ReduceDifQuad** – para calcular os **quadrados das diferenças** entre cada valor do atributo e a média relacionada a esse valor (média relacionada através do agrupamento). Por fim, como a média já calcula o valor de elementos (MapCont e ReduceCont), o arquivo gerado por essa operação será aproveitado para completar o cálculo da variância (parte de baixo da fração).

$$s^2 = \frac{\sum_i^N (X - \bar{X})^2}{N - 1}$$

3.2.6) Desvio padrão

Utiliza **3 operações de MapReduce** (no pior dos casos). O desvio padrão é simplesmente o resultado positivo da raiz quadrada da variância (item 3.2.5), que por sua vez utiliza até **3** operações de MapReduce. Utiliza portanto o arquivo gerado pelo cálculo da variância e extrai a raiz positiva.

$$s = \sqrt{\frac{\sum_i^N (X - \bar{X})^2}{N - 1}}$$

3.2.7) Desvio médio

Utiliza **3 operações de MapReduce** (no pior dos casos). Utiliza a média (que por sua vez utiliza **2** operações de MapReduce), e também utiliza o conjunto de operações **MapDifMod** e **ReduceDifMod** para calcular o **módulo das diferenças** entre o atributo escolhido e a média correspondente ao período de agrupamento. O desvio médio é definido como sendo “a média das distâncias que os valores do conjunto se encontram da média”:

$$DM = \frac{\sum |x_i - \bar{x}|}{n}$$

3.3) Método dos mínimos quadrados (MMQ):

O método dos mínimos quadrados funciona de modo similar a uma regressão logística, utilizado principalmente para prever dados futuros. Ele será aplicado apenas quando o agrupamento for por “**ano**”. No modelo da interface gráfica possuirá sua própria reta no gráfico, reta essa que poderá ser comparada com a linha gerada pelos valores reais de uma estatística escolhida.

A decisão de utilizar este método apenas para valores agrupados por **ano** surgiu com base na quantidade de valores possíveis: a quantidade de **anos** possíveis é indefinida (“infinita”), sendo mais vantajoso do que calcular para os **dias da semana**, **dias do mês** ou **meses** que são limitados. Para um grande conjunto de dados, ao agruparmos por **dia da semana**, **dia do mês** ou por **mês**, poderíamos ter todos os valores possíveis em cada agrupamento – isto é, de domingo até sábado para dia da semana, de 1 até 31 para dia do mês e de janeiro até dezembro para mês –, logo não seria necessário prever dados para um valor que já existe (exemplo: se temos dados agrupados para os 12 meses, não necessitaríamos prever dados referentes a fevereiro pois já temos os dados reais agrupados para fevereiro).

4) Breve descrição dos pacotes

4.1) Pacote “executar”

Contém o método *main* do programa dentro da classe Testar. A classe Testar recebe os parâmetros da Interface gráfica através dos atributos de data, além de invocar as funções estatísticas.

4.2) Pacote “graphics”

Possui todo o código da interface gráfica, necessário para a execução do programa.

4.3) Pacote “operacoes”

O núcleo de operações do programa (camada de “negócios”): contém os cálculos estatísticos mais básicos e operações de agrupamento, como por exemplo “soma”, com seus respectivos *Mappers* e *Reducers*.

4.4) Pacote “util”

Realiza o processamento de estatísticas mais complexas, utilizando principalmente o pacote “operacoes” para juntar diferentes dados de arquivos agrupados após serem “reduzidos”. Responsável principalmente por devolver *Buffers* de leitura dos arquivos do **hdfs** (Hadoop Distributed File System), os quais contêm dados gerados com base na estatística escolhida.

5) Interface Gráfica

5.1) Atributos

O conjunto de dados da NCDC possui 29 atributos, dos quais selecionamos uma parte significativa (25 atributos) para considerarmos na interface gráfica. Apresentaremos a seguir os atributos que foram e que não foram utilizados, tanto numéricos como não numéricos.

5.1.1) Atributos não utilizados:

stn – Identificador da estação meteorológica. Removido por ser um identificador, portanto não traz informações relevantes para cálculos estatísticos no nosso contexto.

wban – Número histórico do Departamento de Meteorologia da Marinha da Força Aérea ("Weather Bureau Air Force Navy"), utilizado quando aplicável e repetido na maioria das vezes, possuindo o valor 99999 no geral. Removido por ser um identificador e por apresentar o mesmo valor (99999) repetidas vezes.

flagMaxTemp – Quando está em branco indica que a temperatura máxima registrada foi absoluta. Quando possui o valor "*" indica que a temperatura máxima foi derivada da temperatura máxima por hora, ou da temperatura relatada sinopticamente. Removido por não trazer informações relevantes no nosso contexto.

flagMinTemp – Quando está em branco indica que a temperatura mínima registrada foi absoluta. Quando possui o valor "*" indica que a temperatura mínima foi derivada da temperatura mínima por hora ou da temperatura relatada sinopticamente. Removido por não trazer informações relevantes no nosso contexto.

5.1.2) Atributos numéricos utilizados:

temp – Temperatura média em graus Fahrenheit.

countTemp – Número de observações usadas para calcular a temperatura média.

dewp – Ponto médio de condensação da água em graus Fahrenheit.

countDewp – Número de observações usadas para calcular o ponto médio de condensação.

slp – Pressão média no nível do mar em milibares.

countSlp – Número de observações usadas para calcular a pressão média no nível do mar.

stp – Pressão média na estação meteorológica em milibares.

countStp – Número de observações usadas para calcular a pressão média na estação meteorológica.

visib – Visibilidade média em milhas.

countVisib – Número de observações usadas para calcular a visibilidade média.

wdsp – Velocidade média do vento em nós.

countWdsp – Número de observações usadas para calcular a velocidade média do vento.

maxSpd – Velocidade máxima sustentada do vento em nós.

maxGust – Rajada de vento mais forte em nós.

maxTemp – Temperatura máxima relatada em graus Fahrenheit – pode não ser exatamente do mesmo dia pois não necessariamente foi relatada no mesmo país.

minTemp – Temperatura mínima relatada em graus Fahrenheit – pode não ser exatamente do mesmo dia pois não necessariamente foi relatada no mesmo país.

prep – Precipitação total (chuva e/ou neve derretida) relatada em polegadas; geralmente não terminará com a observação da meia-noite – ou seja, pode incluir a última parte do dia anterior. O valor .00 indica que não há precipitação mensurável (inclui rastreio). Nota: Muitas estações não informam '0' nos dias sem precipitação, porém '99.99' aparecerá frequentemente nesses dias. Além disso, por exemplo, uma estação só pode reportar um valor de 6 horas para o período durante o qual a chuva caiu.

sndp – Profundidade da neve em polegadas – último relatório do dia, se relatado mais de uma vez. Nota: A maioria das estações não informa '0' em dias sem neve no chão, porém '999.9' aparecerá frequentemente nesses dias.

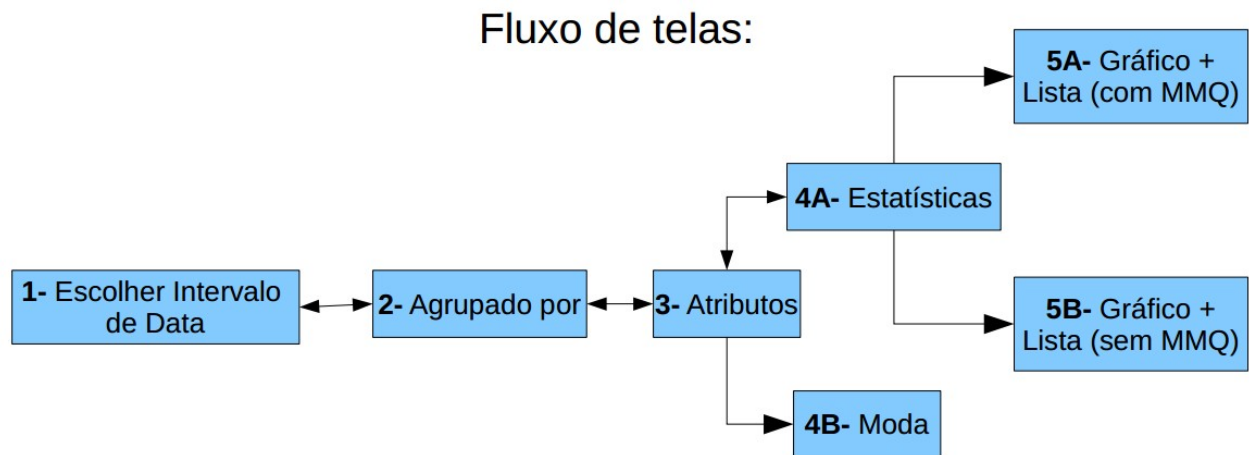
5.1.3) Atributos não-numéricos e qualitativos utilizados:

flagPrp – A = 1 relato de 6 horas de precipitação
B = 2 relatos de 6 horas de precipitação
C = 3 relatos de 6 horas de precipitação
D = 4 relatos de 6 horas de precipitação
E = 1 relato de 12 horas de precipitação
F = 2 relatos de 12 horas de precipitação
G = 1 relato de 24 horas de precipitação
H = A estação informou '0' como a quantidade do dia (por exemplo, a partir de relatos de 6 horas), mas também relatou pelo menos uma ocorrência de precipitação em observações por hora – isso poderia indicar que um rastreio ocorreu, mas deve ser considerado como dados incompletos para o dia.
I = A estação não relatou nenhum dado sobre precipitação no dia e não relatou nenhuma ocorrência de precipitação em suas observações por hora – é possível que tenha ocorrido precipitação mas não foi reportada.

fog – névoa	(1 para sim e 0 para não)
rain – chuva ou garoa	(1 para sim e 0 para não)
snow – neve	(1 para sim e 0 para não)
hail – granizo	(1 para sim e 0 para não)
thunder – trovão	(1 para sim e 0 para não)
tornado – tornado ou nuvem funil	(1 para sim e 0 para não)

6) Fluxo de Telas

Fluxo de interação do usuário e o sistema:



1-) Escolher Intervalo de Data: Após a instalação e configuração descritas anteriormente, quando o usuário iniciar o sistema no terminal, ele deverá escolher o período dos dados a serem analisados. Depois disso, ele deve clicar no botão Próximo (Retiramos o acento da palavra “próximo” para evitar erros de caracteres *unicode*):

A interface do sistema, intitulada 'Defina o periodo dos dados a serem analisados', permite ao usuário definir o intervalo de tempo para a análise. Ela é dividida em duas seções principais: 'Data inicial' e 'Data final'.

Data inicial: O usuário seleciona o mês 'Janeiro' e o ano '1901'.

Data final: O usuário seleciona o mês 'Junho' e o ano '2017'.

Abaixo de cada seleção, há uma grade de dias da semana (D, S, T, Q, Q, S, S) com os dias numerados. Na grade da data final, o dia 28 de Junho está destacado em vermelho.

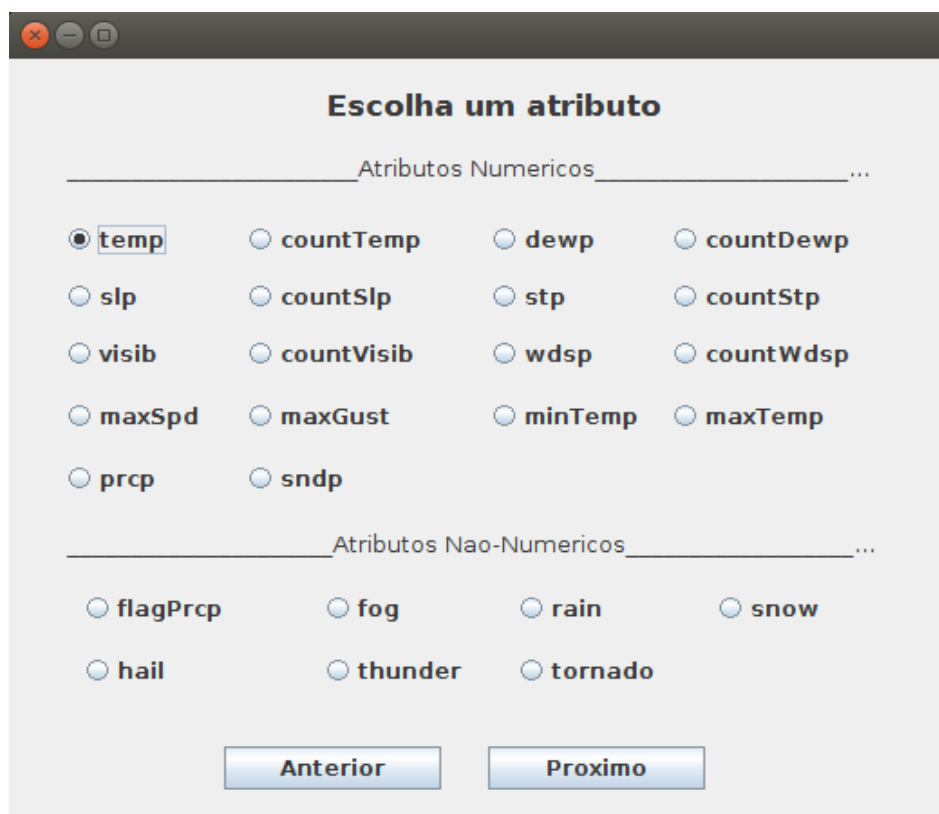
Na parte inferior da interface, há um botão 'Proximo' para avançar para a próxima etapa.

2-) Agrupado por: Na tela seguinte será solicitado que o usuário determine o agrupamento dos dados, como na imagem a seguir. Assim que o fizer, deve clicar no botão Próximo (para prosseguir) ou no botão Anterior (caso queira retornar à página anterior):



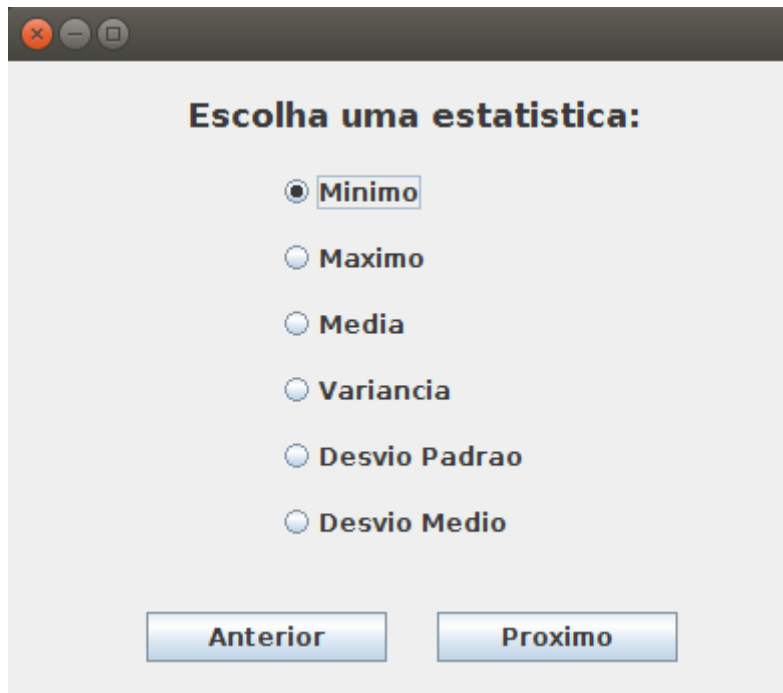
A screenshot of a software window titled "Agrupar valores por:". It contains four radio button options: "Ano" (selected), "Mes", "Dia do mes", and "Dia da semana". At the bottom, there are two buttons: "Anterior" and "Proximo".

3-) Atributos: Neste momento será solicitado que o usuário escolha um atributo, **numérico ou não-numérico**, dentre as opções disponíveis. Segue o exemplo na imagem abaixo. Novamente, o usuário terá a opção de seguir para a próxima tela, clicando em Próximo, ou retornar à tela anterior, clicando Anterior:



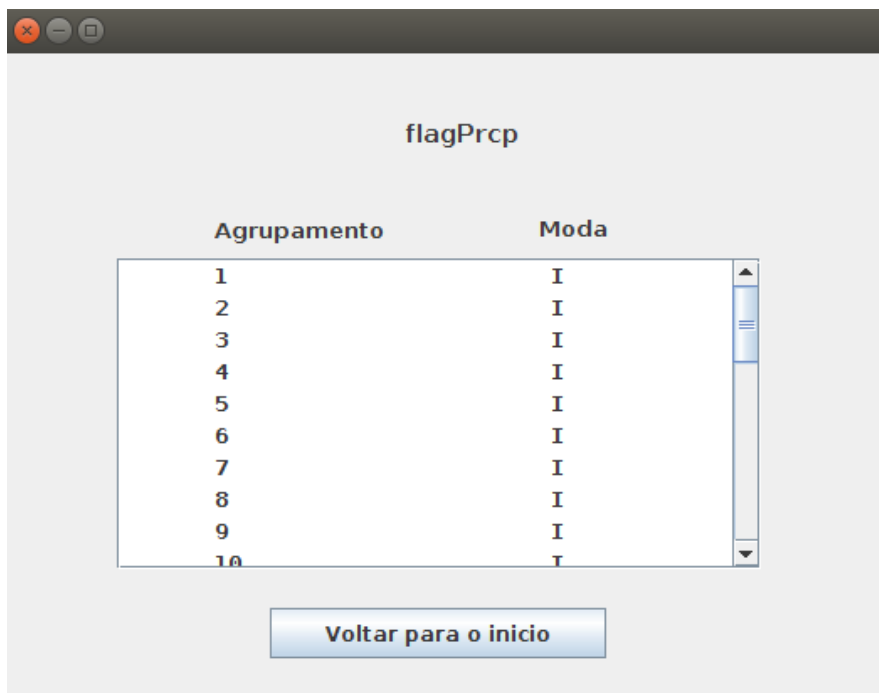
A screenshot of a software window titled "Escolha um atributo". It is divided into two sections: "Atributos Numericos" and "Atributos Nao-Numericos". The "Atributos Numericos" section contains a list of attributes with radio buttons: "temp" (selected), "countTemp", "dewp", "countDewp", "slp", "countSlp", "stp", "countStp", "visib", "countVisib", "wdsp", "countWdsp", "maxSpd", "maxGust", "minTemp", "maxTemp", "prcp", and "sndp". The "Atributos Nao-Numericos" section contains a list of attributes with radio buttons: "flagPrcp", "fog", "rain", "snow", "hail", "thunder", and "tornado". At the bottom, there are two buttons: "Anterior" and "Proximo".

4A-) Estatísticas: Se o usuário escolheu um **atributo numérico** ele chegará nesta página, na qual deverá escolher uma estatística para o cálculo dos dados. Ele poderá seguir para a próxima página, clicando em Próximo, ou retornar à página anterior, clicando em Anterior:



A screenshot of a software window titled "Escolha uma estatística:". The window has a standard OS title bar with close, maximize, and minimize buttons. Inside, there are six radio buttons for selecting a statistical measure: "Minimo" (selected), "Maximo", "Media", "Variancia", "Desvio Padrao", and "Desvio Medio". At the bottom, there are two buttons: "Anterior" and "Proximo".

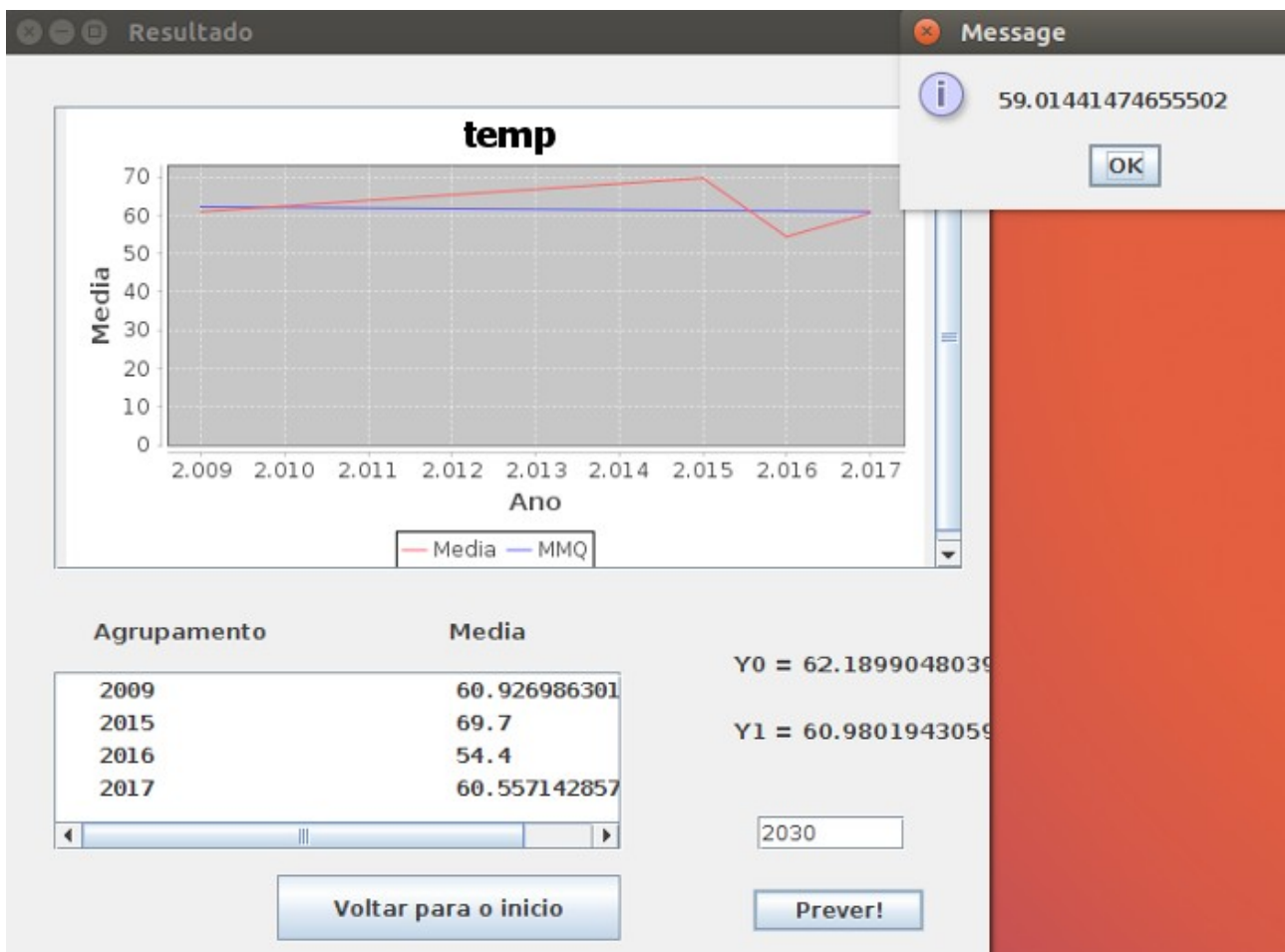
4B-) Moda: Caso o usuário tenha escolhido um **atributo não-numérico**, ele chegará nesta página, na qual será gerada a moda do atributo escolhido anteriormente. Caso queira, o usuário poderá ir para a primeira tela, clicando em Voltar para o início:



A screenshot of a software window titled "flagPrcp". It displays a table with two columns: "Agrupamento" and "Moda". The table has 10 rows, with "Agrupamento" values from 1 to 10 and "Moda" values all being "I". To the right of the table is a vertical scrollbar. Below the table is a button labeled "Voltar para o início".

Agrupamento	Moda
1	I
2	I
3	I
4	I
5	I
6	I
7	I
8	I
9	I
10	I

5A-) Gráfico + Lista (com MMQ): Esta tela aparecerá caso o usuário tenha selecionado o **agrupamento por ano** e um **atributo numérico**. Será gerado um gráfico, no qual a **abcissa (x)** possuirá os valores do **agrupamento** escolhido – somente ano nesta tela –, enquanto a **ordenada (y)** possuirá os valores da **estatística escolhida**. No gráfico estarão presentes duas retas: uma reta **MMQ**, que denota a reta criada com base no método dos mínimos quadrados; e uma reta da **estatística escolhida**. O usuário pode inserir no campo texto um ano para o qual pretende prever os dados, com base no método dos mínimos quadrados, clicando então no botão Prever!: será retornada uma mensagem contendo o dado previsto em uma pequena tela. Por fim, o usuário pode também voltar à tela inicial, clicando em Voltar para o início. Abaixo temos um exemplo desta tela, juntamente com um exemplo hipotético de dados previstos para 2030 utilizando o MMQ. Perceba também a presença dos dados Y0 e Y1 neste exemplo, relacionados ao cálculo do MMQ para 2009 e 2017 respectivamente.



5B-) Gráfico + Lista (sem MMQ): Aparecerá caso o usuário tenha escolhido um **atributo numérico**, porém com **agrupamento diferente de ano**. Esta tela é semelhante à tela 5A, porém sem a opção de prever um valor e sem a reta gerada pelo MMQ (esta exclusão está explicada na seção 3.2). O usuário então pode visualizar os dados e, caso queira, retornar ao início clicando em [Voltar para o início](#).

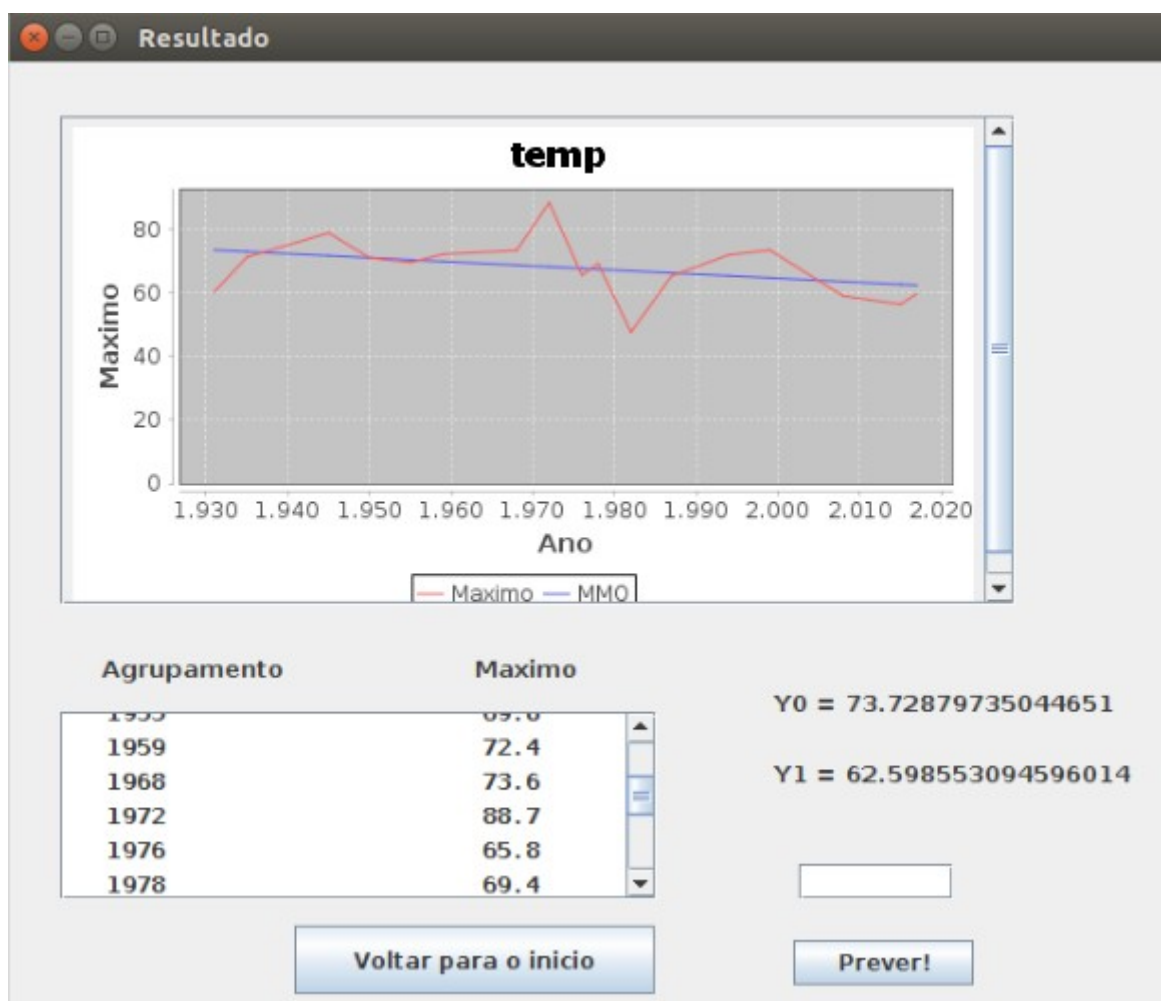


7) Exemplos de testes de execução

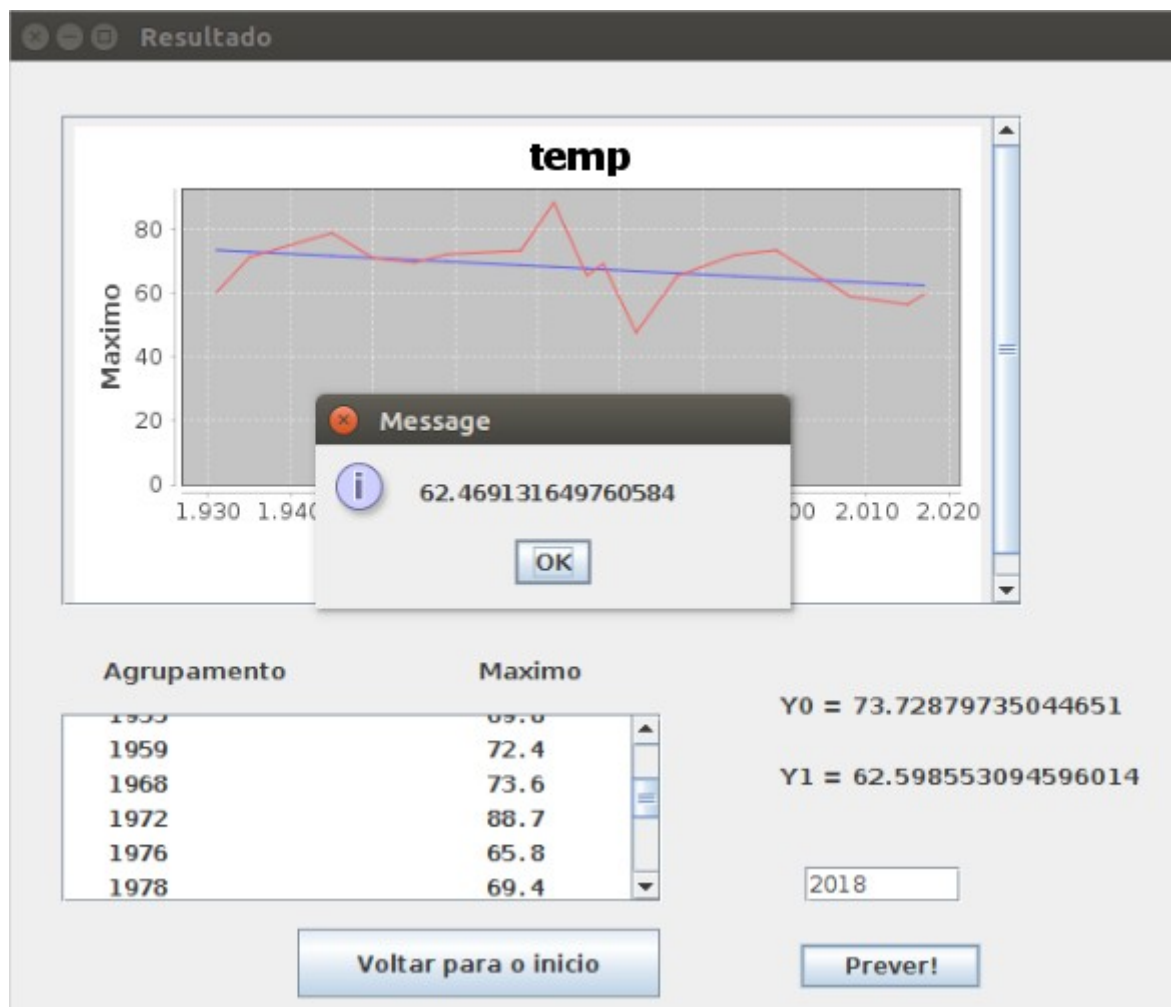
Os exemplos de execução de vários testes utilizam uma amostra de cerca de 30 arquivos de diferentes anos. Foi necessário reduzir o conjunto de dados para teste pois a implementação *single-node cluster* do Hadoop apresenta erros quando lida com uma grande base de dados: localmente não consegue tratar um número grande de arquivos como argumento (aproximadamente 500 mil). Além disso, levaria um tempo muito maior para executar cada tarefa.

1º) Temperatura média máxima registrada em todos os anos

No teste abaixo, foi escolhido o intervalo de data **entre 1 de Janeiro de 1901 e 2 de Julho de 2017**, utilizando a estatística “**máximo**”, atributo *temp* e agrupamento feito **por ano**. Como resultado, temos a figura a seguir:



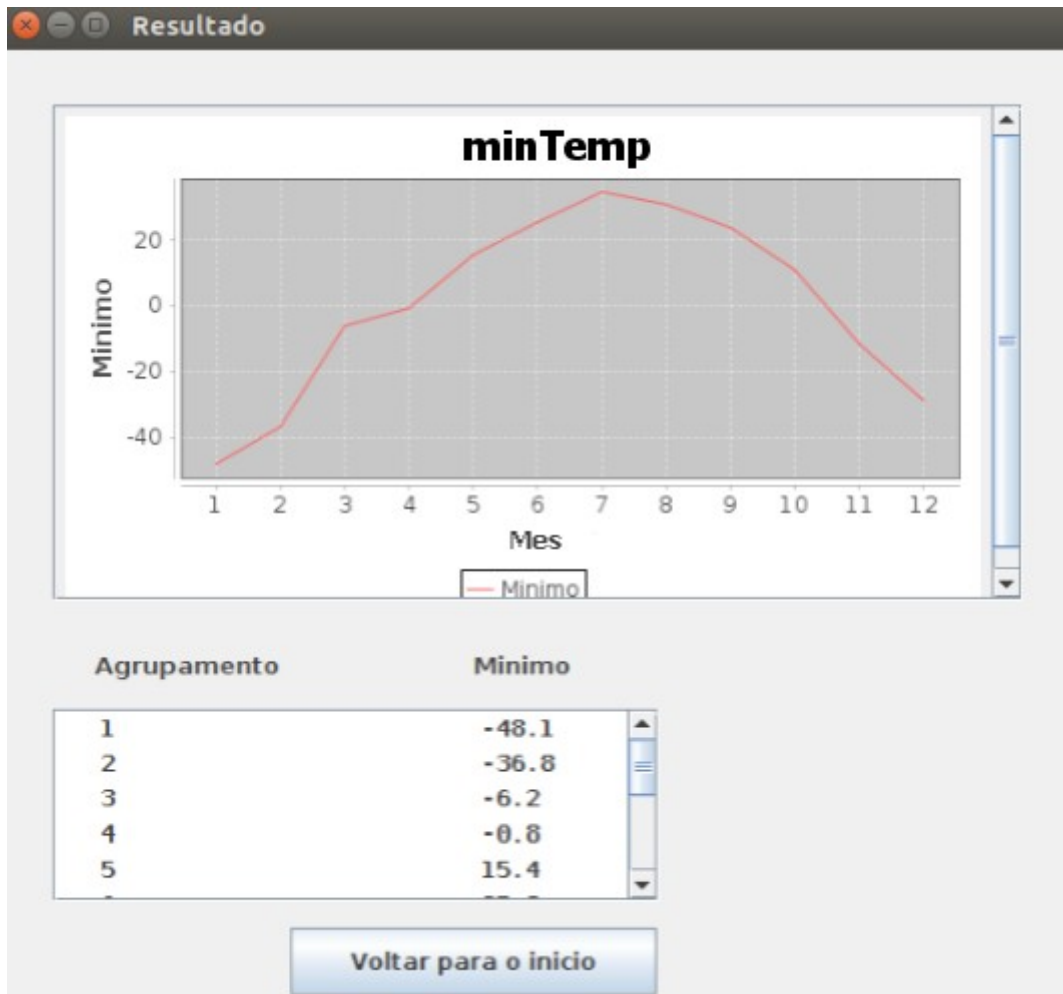
A maior média de temperatura coletada foi de 88.7 em 1972 como é visto no pico máximo do gráfico, também temos o resultado do agrupamento em ordem crescente dos anos. Abaixo faremos uma predição utilizando o Método dos Mínimos Quadrados (MMQ): qual será a temperatura máxima para o ano de 2018 (digitando “2018” na caixa de texto e clicando em “Prever!”):



Prevemos o valor de cerca de 62.5 para a temperatura média máxima em 2018.

2º) Mês mais frio registrado na década de 90

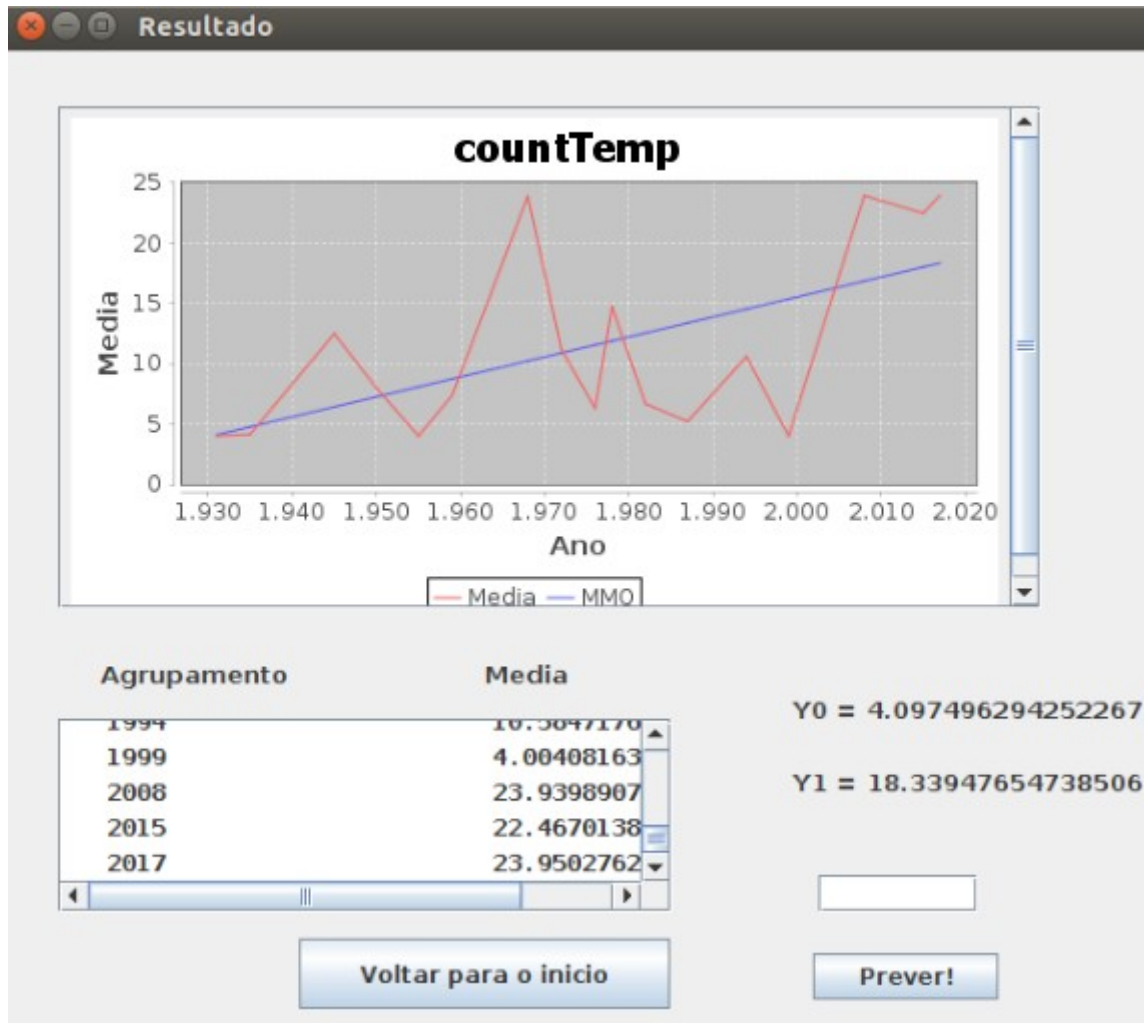
Analisaremos a temperatura **mínima** (*minTemp*) no intervalo de toda a década de 90 (1 de Janeiro de 1990 até 31 de Dezembro de 1999) agrupando por **mês**:



Observamos que o mês de Janeiro teve o valor mínimo de temperatura mínima durante toda a década de 90, também percebemos que os meses próximos à metade do ano (Junho, Julho, Agosto) possuem temperaturas mais altas do que os meses do início e do fim do ano.

3º) A média de observações de temperatura de todos os anos

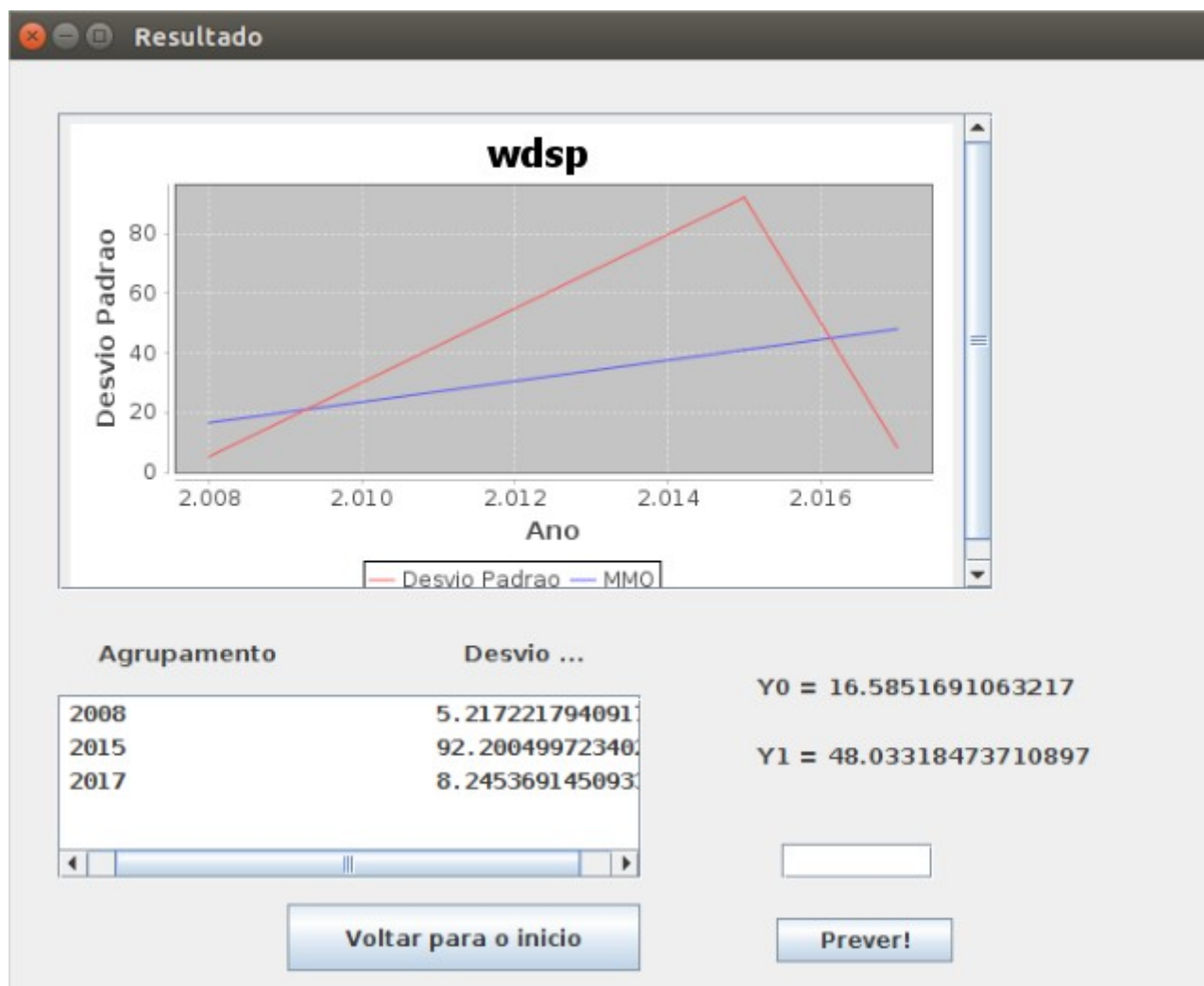
Para este novo teste, foi escolhido novamente o intervalo de data entre **1 de Janeiro de 1901** e **2 de Julho de 2017** com o valor **máximo** do atributo *countTemp* agrupado por ano. Como resultado temos a figura a seguir:



Observamos que a maior média de observações de temperatura (*countTemp*) foi em 2017, porém o crescimento e redução de observações é acentuado, sem seguir uma continuidade suave.

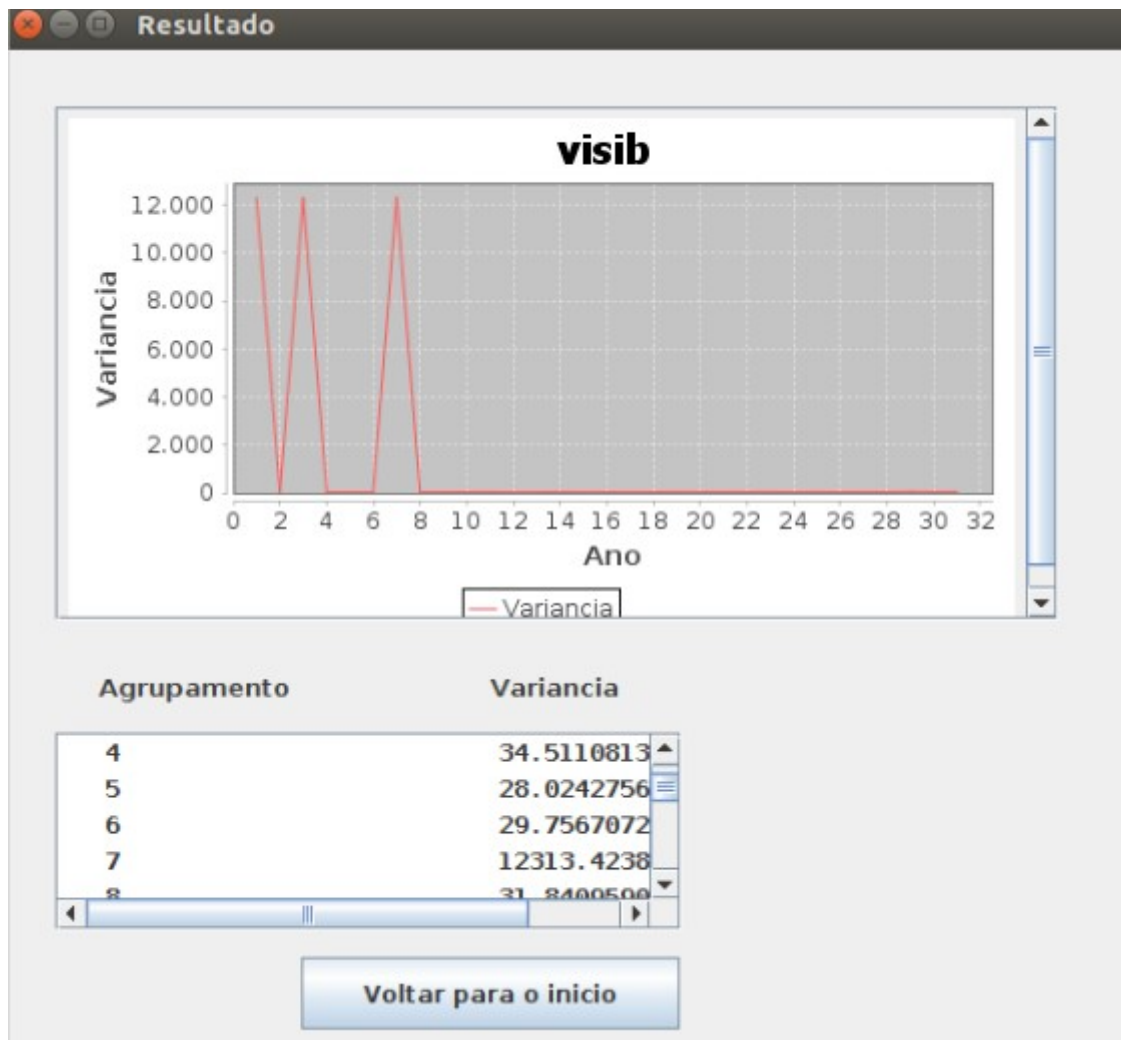
4º) O desvio padrão da velocidade média do vento no século XXI

Parecido com o exemplo anterior, foi escolhido o intervalo de data entre **1 de Janeiro de 2001 e 2 de Julho de 2017**, buscando pelo **desvio padrão** do atributo **wdsp**, agrupado por **ano**. Como resultado, temos a figura a seguir:



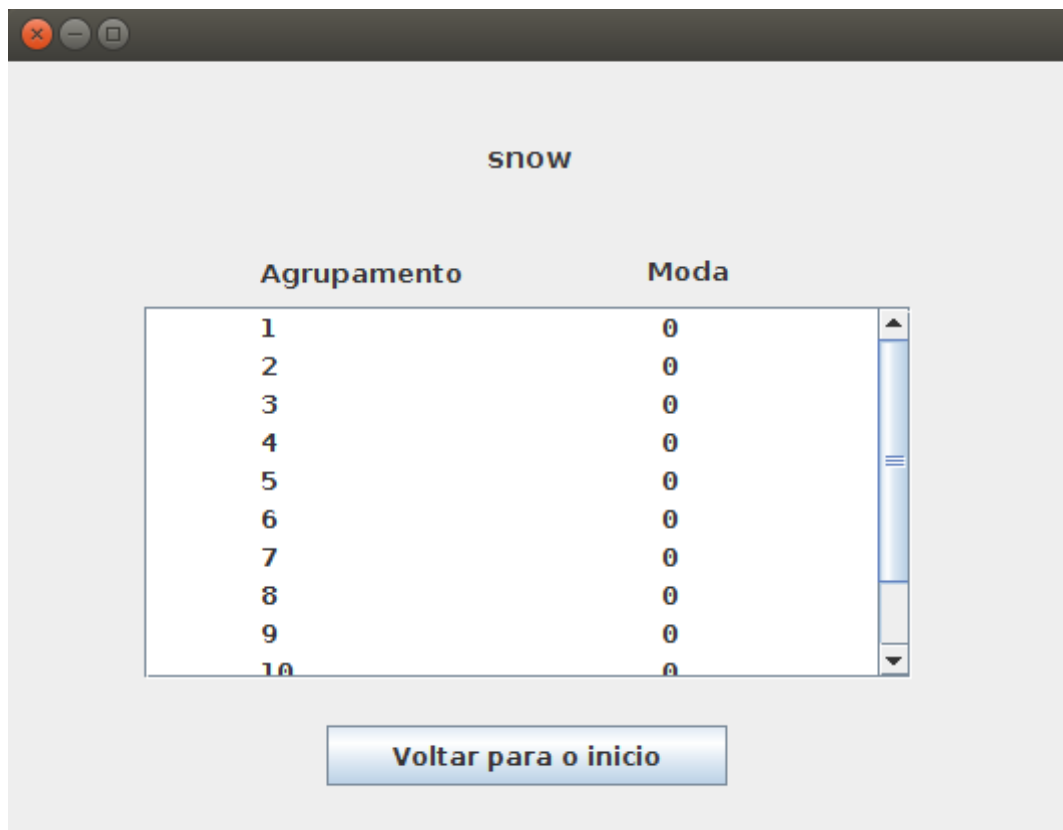
5º) A variância da visibilidade média em milhas na entre as décadas de 30 e 60

Utilizando um intervalo de data entre 1 de Janeiro de 1930 e 31 de Dezembro de 1960 com o valor da visibilidade média em milhas (*visib*) agrupado por dias do mês. Como resultado, temos a figura a seguir:



6º) Moda de ocorrência de neve no século XXI

Queremos investigar se de acordo com os dados, houve mais dias com precipitação de neve ou sem precipitação de 1 de Janeiro de 2001 até 2 de Julho de 2017. Utilizamos o atributo não-numérico *snow* agrupando por **mês**, verificando também os meses que tiveram neve ou não:



The screenshot shows a window titled "snow" with a table containing 10 rows of data. The table has two columns: "Agrupamento" and "Moda". All values in the "Moda" column are 0. Below the table is a button labeled "Voltar para o inicio".

Agrupamento	Moda
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0

Percebemos que no século XXI foi mais frequente a não ocorrência de neve do que a ocorrência.

7º) Ocorrência de trovões na década de 50

Durante a década de 50, podemos descobrir se houve mais dias com ocorrência de trovão ou sem ocorrência de trovão. Utilizamos o atributo não-numérico *thunder*, juntamente com a estatística **moda**, agrupando por **ano**:



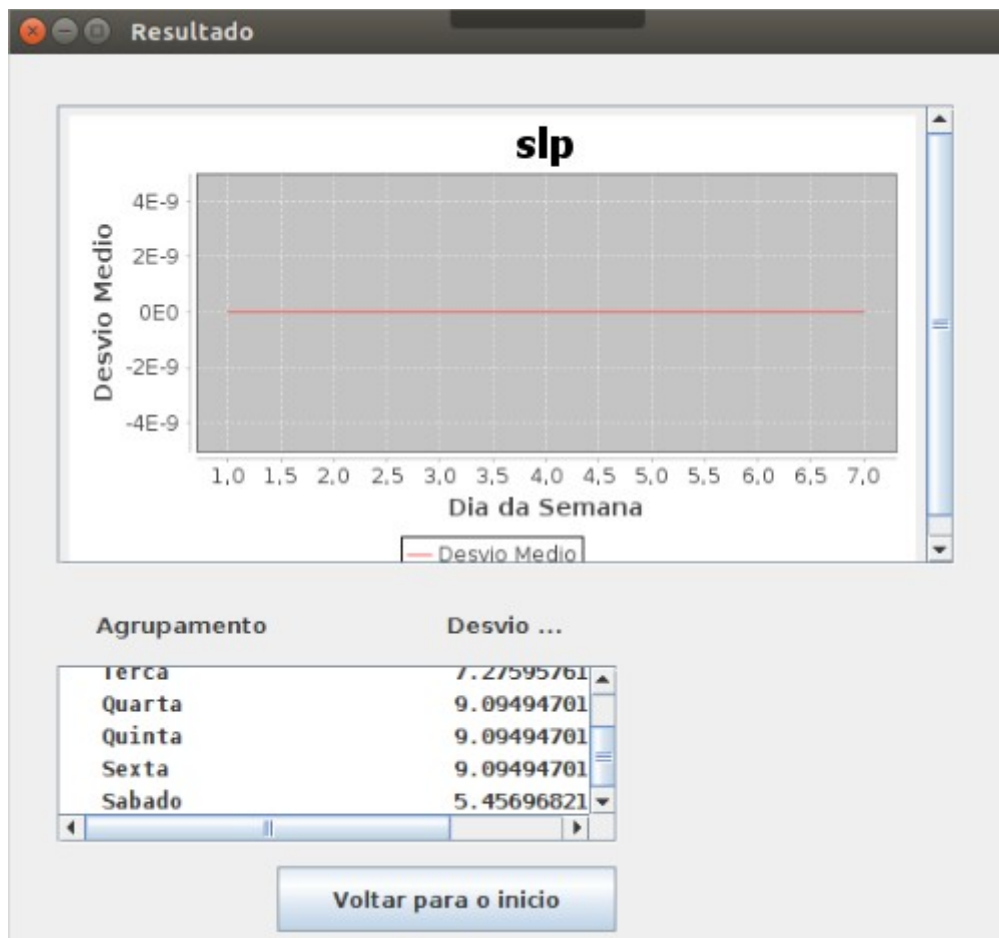
Agrupamento	Moda
1950	0
1955	0
1959	0

Voltar para o inicio

Dos três anos utilizados na amostra nenhum teve ocorrências de mais dias com trovões do que dias sem trovão.

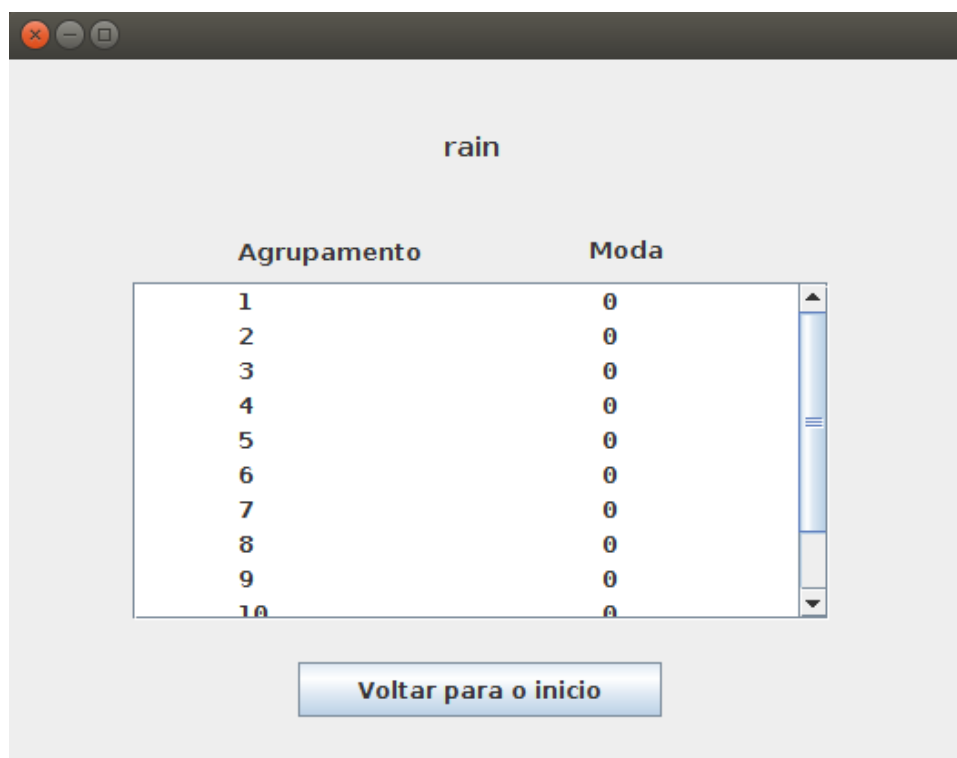
8º) Desvio médio da pressão média no nível do mar em 1994

Podemos verificar a dispersão dos dados (afastamento) quanto à pressão média em milibares (**slp**), agrupando por **dia da semana** no ano de **1994**. O maior desvio médio é geralmente nas quartas, quintas e sextas-feiras.



9º) Ocorrência de chuva durante todos os anos em todos os meses.

De todos os anos de dados coletados, podemos nos perguntar se houve algum mês com dias mais chuvosos do que dias sem chuva. Desta forma, utilizamos a moda do atributo *rain* (chuva ou garoa) agrupado por **mês**, recebendo o seguinte resultado:



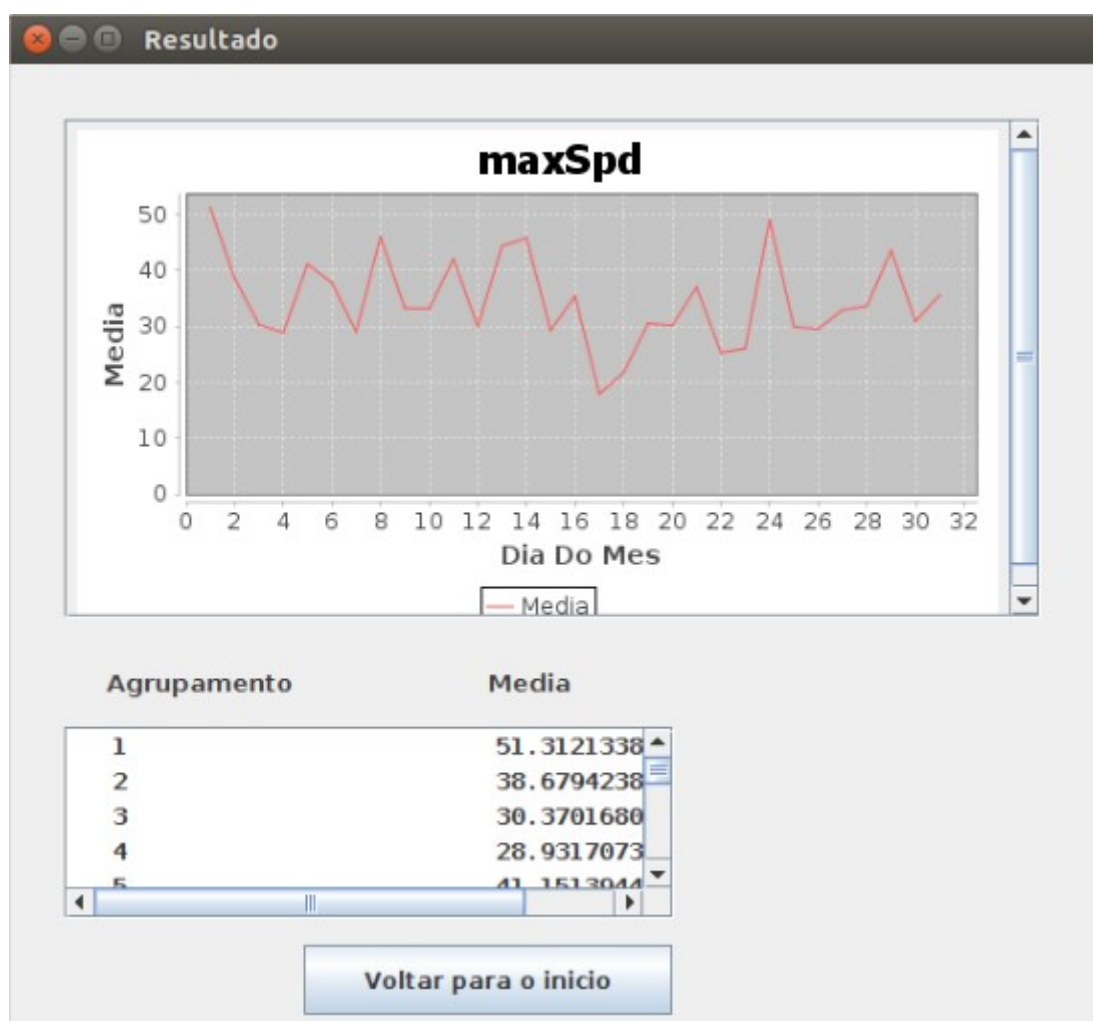
Agrupamento	Moda
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0

Voltar para o início

Percebemos então que os meses tiveram no geral mais dias sem chuva do que dias chuvosos.

10º) Média da velocidade máxima sustentada do vento em dias do mês (para todos os anos).

Em todos os anos, procuramos agora algum tipo de ascensão ou declínio brusco na velocidade do vento. Será utilizada a média do atributo *maxSpd*, agrupando por **dias do mês**. No gráfico percebemos que há uma boa variação entre cada dia, também nos dias 17 geralmente a velocidade é menor se comparada com os outros dias.



8) Referências Bibliográficas:

https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html

<https://www.ibm.com/analytics/us/en/technology/hadoop/mapreduce/>

<http://www.devmedia.com.br/hadoop-mapreduce-como-criar-um-programa-mapreduce-base/30080>

Imagens:

Desvio padrão

[http://videos.web-](http://videos.web-03.net/artigos/Eduardo_Correa_Goncalves/Amplitude_Variancia_Oracle/Amplitude_Variancia_Oracle3.jpg)

[03.net/artigos/Eduardo_Correa_Goncalves/Amplitude_Variancia_Oracle/Amplitude_Variancia_Oracle3.jpg](http://videos.web-03.net/artigos/Eduardo_Correa_Goncalves/Amplitude_Variancia_Oracle/Amplitude_Variancia_Oracle3.jpg)

Variância

[http://videos.web-](http://videos.web-03.net/artigos/Eduardo_Correa_Goncalves/Amplitude_Variancia_Oracle/Amplitude_Variancia_Oracle4.jpg)

[03.net/artigos/Eduardo_Correa_Goncalves/Amplitude_Variancia_Oracle/Amplitude_Variancia_Oracle4.jpg](http://videos.web-03.net/artigos/Eduardo_Correa_Goncalves/Amplitude_Variancia_Oracle/Amplitude_Variancia_Oracle4.jpg)

Desvio médio

[https://www.google.com.br/url?](https://www.google.com.br/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=0ahUKEwjznKLJiubUAhUEUZAKHbMEB-0QjBwIBA&url=http%3A%2F%2Fs3.amazonaws.com%2Fmagoo%2FABAAABdnUAK-2.jpg&psig=AFQjCNEijJ-T2BA0rDor075AvEXITbeq0w&ust=1498929639390238)

[sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=0ahUKEwjznKLJiubUAhUEUZAKHbMEB-0QjBwIBA&url=http%3A%2F%2Fs3.amazonaws.com%2Fmagoo%2FABAAABdnUAK-2.jpg&psig=AFQjCNEijJ-T2BA0rDor075AvEXITbeq0w&ust=1498929639390238](https://www.google.com.br/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=0ahUKEwjznKLJiubUAhUEUZAKHbMEB-0QjBwIBA&url=http%3A%2F%2Fs3.amazonaws.com%2Fmagoo%2FABAAABdnUAK-2.jpg&psig=AFQjCNEijJ-T2BA0rDor075AvEXITbeq0w&ust=1498929639390238)

Média

<https://www.google.com.br/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&ved=0ahUKEwjQ5-zfiubUAhXKkJAKHdanBekQjBwIBA&url=http%3A%2F%2Fs3.amazonaws.com%2Fmagoo%2FABAAABtnwAF-4.jpg&psig=AFQjCNG7wjCo2YYnC0HGMgg9d-pD-nLZJQ&ust=1498929767585388>