

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Vamos mostrar nesta abordagem da disciplina, como ela se integra ao projeto de PI, por meio da figura de arquitetura de referência.

A arquitetura de referência visa apenas apresentar um modelo que permite a implementação do conteúdo da disciplina de maneira estritamente aplicada.

Com os datasets implementados ganha-se velocidade na implementação da modelagem computacional, testes e correções podem ser induzidas a partir da experiência no uso direto da matemática.

### **Por que usamos Python?**

Porque o python é um recurso multidisciplinar que pode ser aplicado em vários contextos genéricos e não especializados. Por este motivo não existem ferramentas de implementação exclusiva para python, são necessárias alocações de outros recursos que auxiliam na construção de modelagem matemática. Por exemplo, nativamente gráficos não são gerados pelo python por meio de um pacote ou libs instaladas. É necessário o uso de ferramentas que incorporem essas funcionalidades. O python foi desenvolvido com propósito genérico.

### **Vamos continuar usando python?**

Sim, bastante. Usaremos python para captura de outros dados para formar datasets especializados. Encontramos restrições nas próprias libs dos pacotes, por exemplo, a temperatura e a frequência não são capturadas diretamente da bios, no ambiente windows ou VM com SO Linux, mas em ambiente windows. O Windows 10, à partir dessa versão, incorporou um sistema de proteção contra ataques e infecções de vírus, não permitindo o acesso a mesma de qualquer aplicação que não seja nativa, como uma dll ( biblioteca dinâmica). A implementação de matemática computacional com Python será abordada em conjunto com outra ferramenta.

### **Porque vamos usar R (Rstudio)?**

A linguagem R é uma poderosa ferramenta de modelagem e criação de algoritmos matemáticos. Tem um portfólio colaborativo mundial com respeito a novos algoritmos que se aplicam a tudo. Basta usar os exemplos nos seus conjuntos de dados e aprender se o modelo reflete o comportamento previsto. Além disso as implementações gráficas, que auxiliam na interpretação dos dados são muito bem escritas, gerando inúmeras possibilidades para o data science.

### **Porque usamos o excel e não diretamente o banco de dados?**

O excel é uma ferramenta amplamente utilizada no mundo todo, seu maior potencial é a possibilidade de trabalhar com arquivos csv ( comma separated values) , ou seja, valores separados por vírgula. Este formato de arquivo representa 100% das bases de dados e repositórios mundiais com milhões de bases de dados disponíveis para serem usadas, modeladas e se criar novas oportunidades para empresas e pessoas.

O dados não mentem, o que fazemos aqui em cálculo é aplicar os modelos matemáticos, dos básicos aos avançados, nesses conjuntos de dados. Obviamente é impossível aplicar os métodos

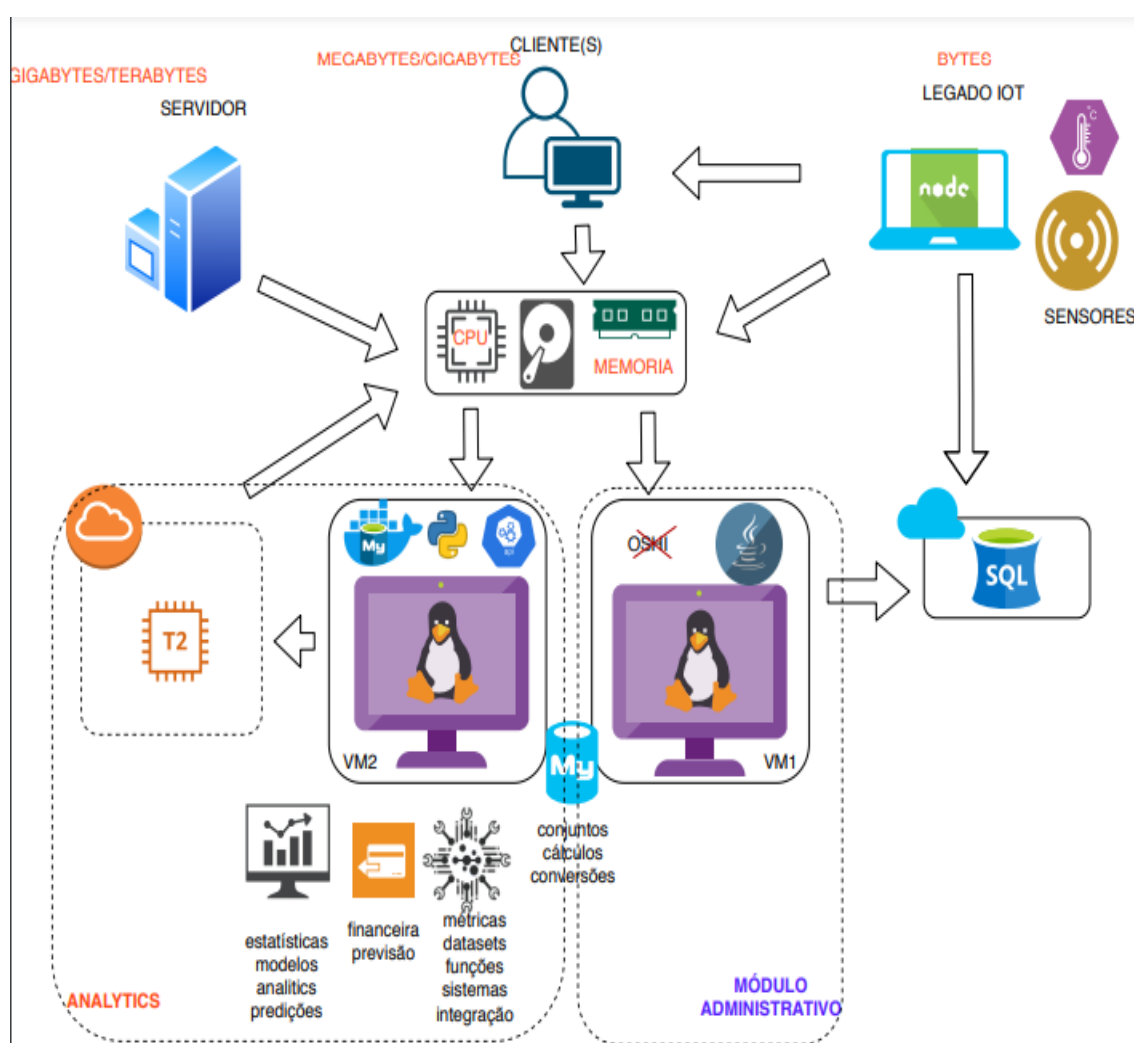
# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

numéricos a todo o conjunto, por isso utilizamos amostras desses conjuntos e vamos testando o comportamento dessas amostras. Deixamos claro que são amostras de diversas partes do conjuntos, se na maioria o comportamento se mostrar o mesmo, então aquele modelo matemático representa os dados de todo esse conjunto.

É necessário esclarecer que os dados são dinâmicos e portanto os modelos matemáticos que os representam devem ser constantemente verificados e validados, quando não mais os representam devem ser experimentados outros modelos que melhor se adequem a essa representação.

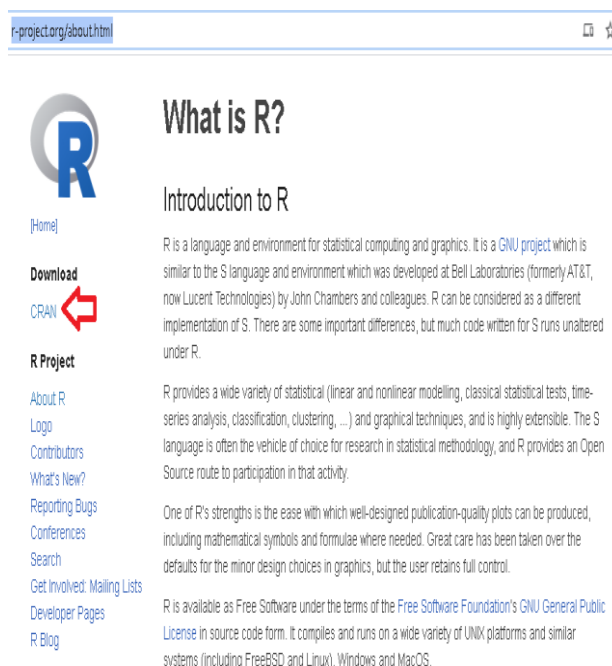
### Arquitetura de referência:



# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Baixe o R em sua máquina ( <https://www.r-project.org/about.html>)



Aqui você pode baixar qualquer mirror de sua preferência. O projeto R é mundial

Brazil

<https://nbcgib.uesc.br/mirrors/cran/>

<https://cran-r.c3sl.ufpr.br/>

<https://cran.fiocruz.br/>

<https://vps.fmvz.usp.br/CRAN/>

<https://brieger.esalq.usp.br/CRAN/>



Computational Biology Center at Universidade Estadual de Santa Cruz

Universidade Federal do Parana

Oswaldo Cruz Foundation, Rio de Janeiro

University of Sao Paulo, Sao Paulo

University of Sao Paulo, Piracicaba

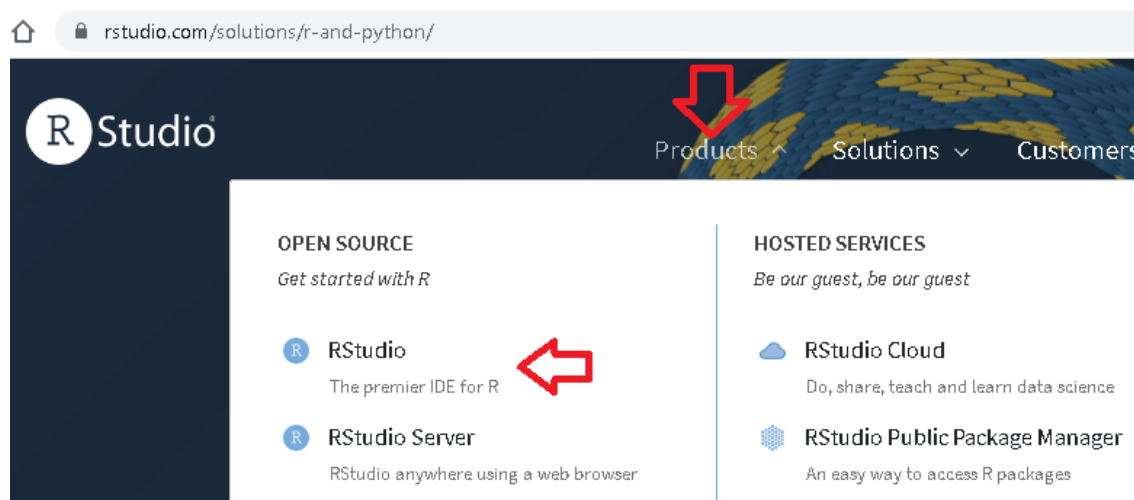
Agora vamos colocar uma “capinha bonita” na tool R usando o Rstudio, que vai se ajustar como uma automação de dados, importação e exportação, etc.

Baixe o RStudio em sua máquina (<https://rstudio.com/>)

Vá em produtos e clique opção RStudio

# Importação e Exportação dos datasets

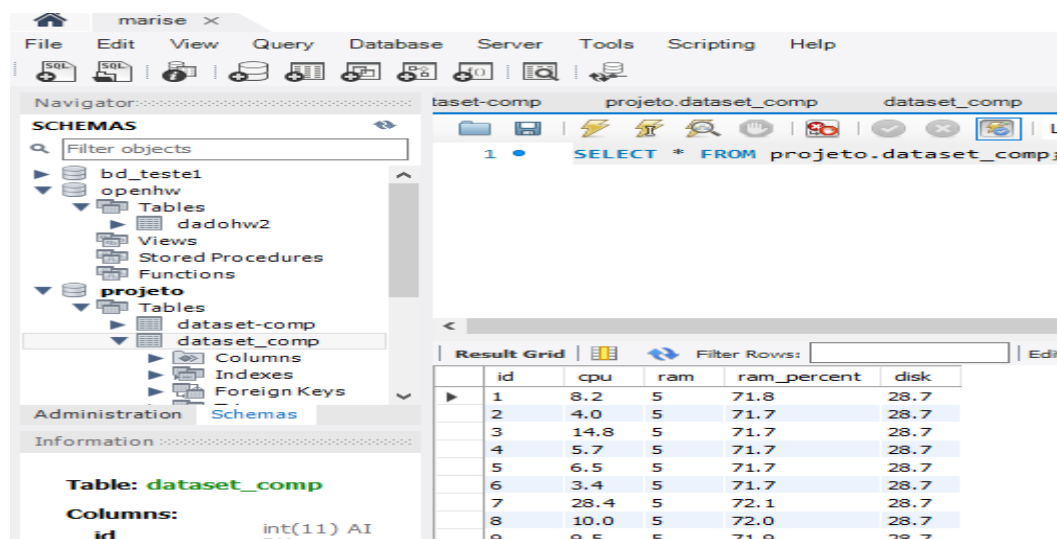
## Aplicando modelagem de cálculo computacional com R aos data sets



### Exportação e Importação de dados do banco de dados.

A API Pythonics realiza inserts no banco de dados Mysql. Precisamos desses dados para realizar nossas análises por meio da matemática computacional.

O ambiente que estamos usando do R não tem conector para mysql, então usaremos a exportação da tabela dataset\_comp em formato csv, que está armazenada no banco de dados Mysql local.

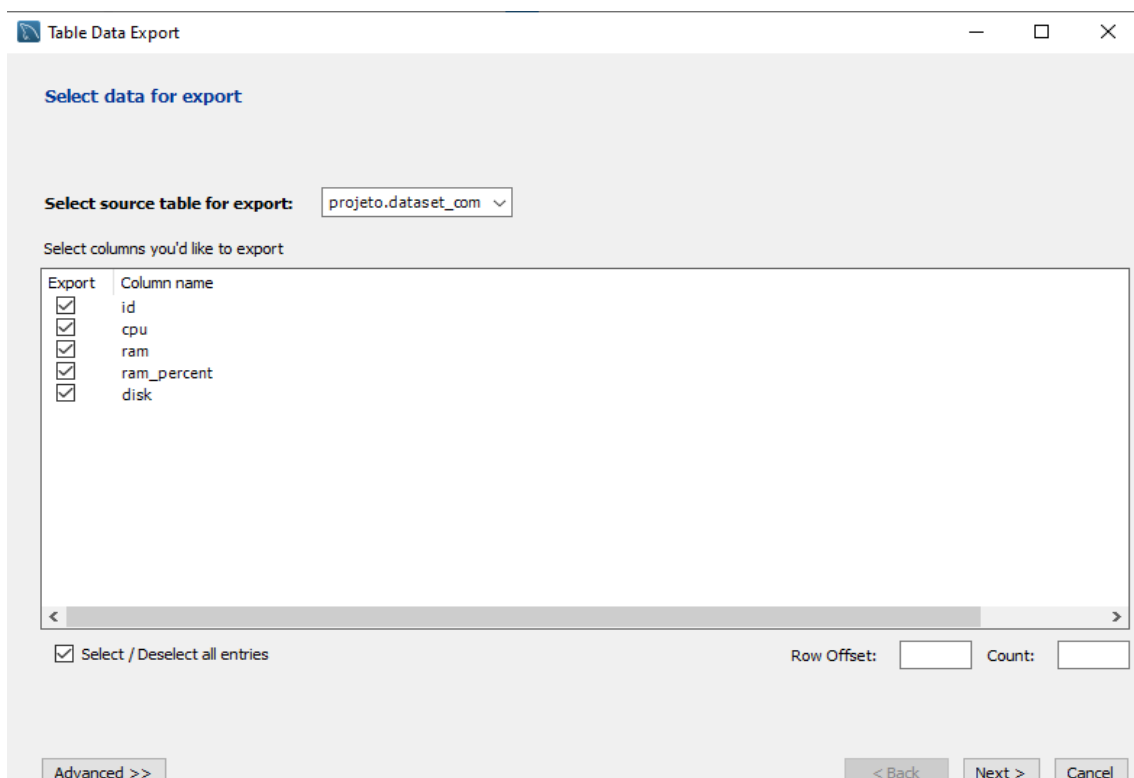
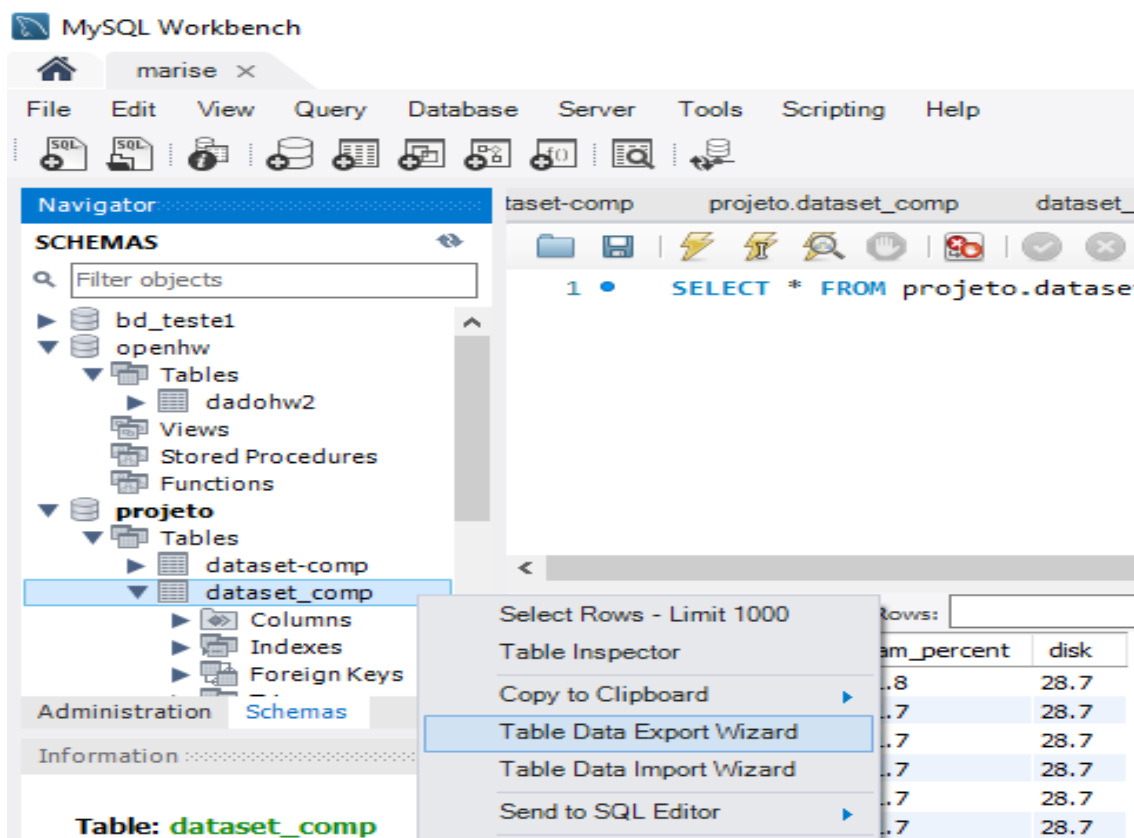


Clique com o botão direito sobre o nome da tabela, na janela dos schemas.

Selecione Table Data Export Wizard

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets



Deixe todas as colunas selecionadas e clique em next.

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Pode-se exportar tabelas como formato csv ou json.

Deixar os campos conforme padrão da ferramenta wizard.

Clique em browse para local onde será salvo a tabela. Criar um nome para que a tabela seja salva com o nome conhecido. Recomenda-se salvar dentro de uma pasta.

Clicar em next.

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

**Export Data**

The following tasks will now be performed. Please monitor the execution.

☐ Prepare Export  
☐ Export data to file

Click [Next >] to execute.

Show Logs

< Back   Next >   Cancel

Clique em next.

**Export Data**

The following tasks will now be performed. Please monitor the execution.

☒ Prepare Export  
☒ Export data to file

Finished performing tasks. Click [Next >] to continue.

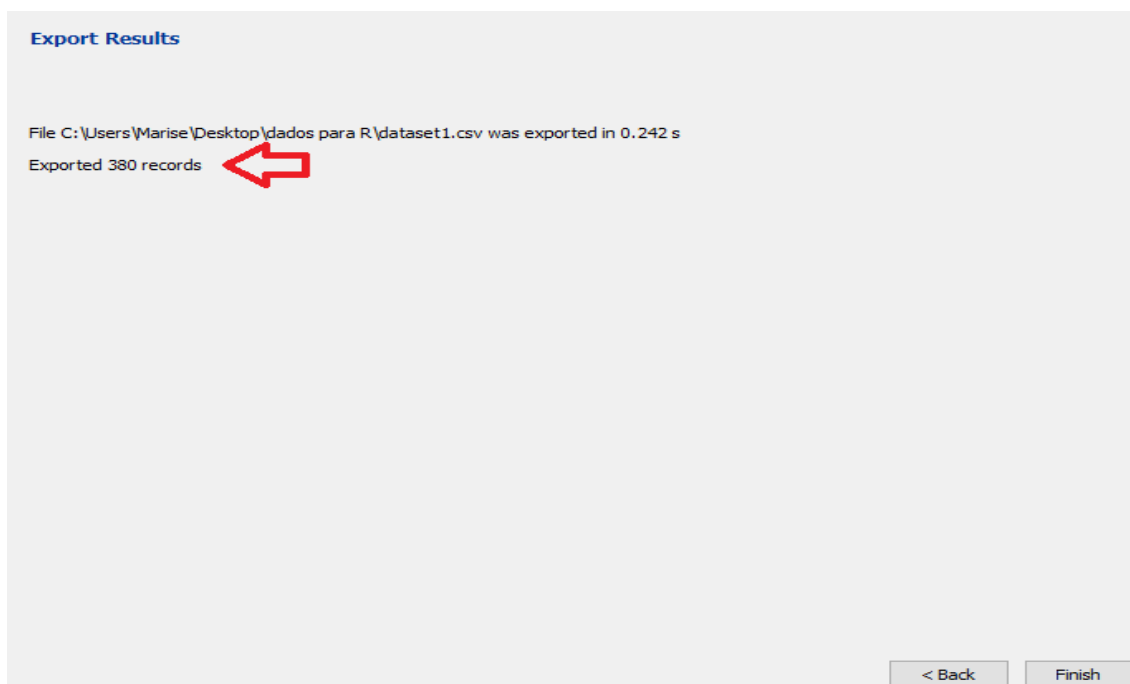
Show Logs

< Back   Next >

Verifique que o processo de exportação foi finalizado, clique em next.

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets



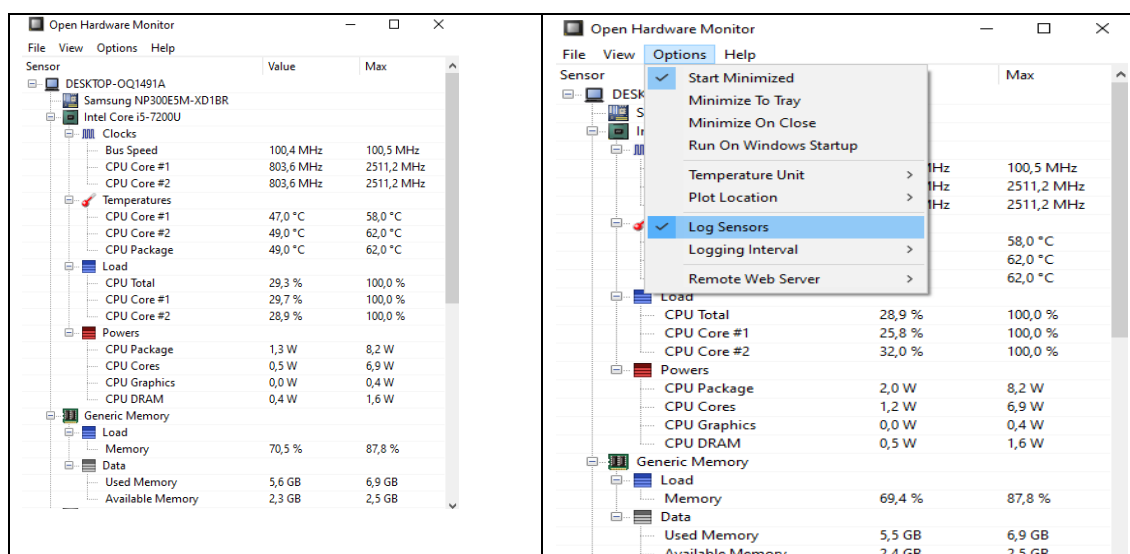
Foram exportados 380 registros, clicar em finish.

Os registros são salvos na pasta indicada para exportação.

Nome	Data de modificação	Tipo	Tamanho
dataset1.csv	24/09/2020 21:43	Arquivo de Valores Separados por Vírgulas ...	11 KB

Agora será realizada a importação para o banco de dados Mysql, vamos importar os arquivos gerados pelo log sensor do OHM.

(OHM → OpenHardwareMonitor → <https://openhardwaremonitor.org/downloads/>)





# Importação e Exportação dos datasets

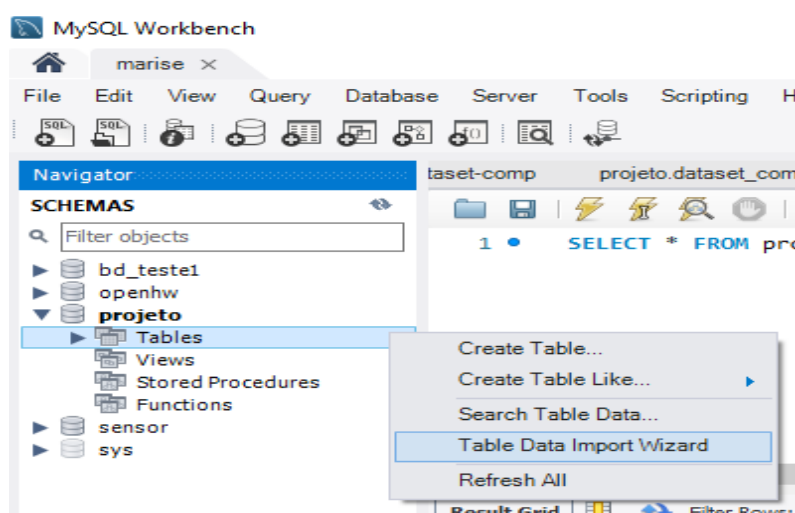
## Aplicando modelagem de cálculo computacional com R aos data sets

Habilitar o OpenHardwareMonitor.exe e em “options” setar “Log Sensors”. Esta ação gerará um arquivo no formato csv com os dados capturados por esse executável.

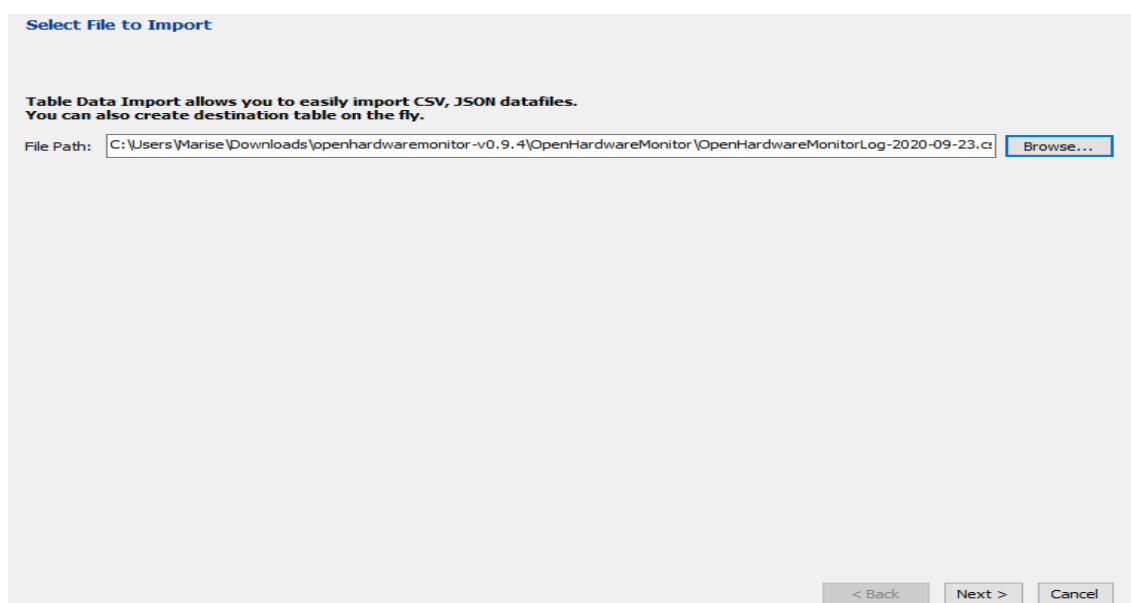
Lembrando que este executável é uma dll (biblioteca dinâmica) e portanto, para SO Windows acessa a BIOS da máquina.

##### Atenção: cuidando com códigos disponibilizados na internet que utilizam o módulo ou pacote WMI (Windows Management Instrumentation), este pacote permite coletar a temperatura ThermalZone que é utilizada erroneamente como sendo da CPU, mas na verdade é do sensor da placa mãe. #####

No Mysql Workbench, em tabela, clicar com o botão direito e depois Table Data Import Wizard.



Depois procure a tabela que deseja importar, se preferir mude o nome.



# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Mude o nome desejado, da nova tabela que será importada, na seta vermelha a seguir. Clique em next. Veja as opções para usar uma tabela existente ou dropar a tabela existente.

**Select Destination**

Select destination table and additional options.

☐ Use existing table: projeto..datasetmaq1

☒ Create new table: projeto . OpenHardwareMonitorLog1

☐ Drop table if exists

< Back Next > Cancel

Esta é uma etapa importante, pois podemos incluir algum tratamento na tabela, uma evz que essas manipulações de dados acarretam alterações nos tipos de dados a serem reconhecidos pelas diversas ferramentas utilizadas.

**Configure Import Settings**

Detected file format: csv

Encoding: utf-8

Columns:

Column	Type
/intelcpu/0/load/0	text
/intelcpu/0/temperature/0	text
/intelcpu/0/temperature/1	text
/intelcpu/0/temperature/2	text
/intelcpu/0/clock/1	int
/intelcpu/0/clock/2	int

MyUnknown... /intelcpu/0... /intelcpu/0... /intelcpu/0... /intelcpu/0... /intelcpu/0... /intelcpu/0... /intelcpu/0... /intelcpu/0... /intelcpu/0...

Time	CPU Core #1	CPU Core #2	CPU Package	CPU Core #1	CPU Core #2	CPU Package	CPU Core #1	CPU Core #2	CPU Package
09/23/202...	0	0	31	3113.77734	3113.77734	10.5545	3113.77734	3113.77734	10.5545
09/23/202...	76.62338	76.62338	31	3113.77734	3113.77734	7.93385	3113.77734	3113.77734	7.93385
09/23/202...	100	100	47	3113.77563	3113.77563	13.7205	3113.77563	3113.77563	13.7205

< Back Next > Cancel

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Veja que podemos ter uma visualização prévia do que estamos importando e realizando o tratamento do tipo dos dados que podemos armazenar para cada coluna.

Observe que o wizard detectou o formato do arquivo, o padrão de econdig deve ser respeitado em geral é UTF-8.

Os dados podem ser ajustados, como text, int, double, etc.

É possível ter a visualização dos dados dispostos previamente na tabela importada.

O ID será automaticamente assumido pela primeira coluna.

**Table Data Import**

**Configure Import Settings**

Detected file format: csv

Encoding: utf-8

Columns:

Source Column	Field Type
MyUnknownColumn	int
/intelcpu/0/load/1	int
/intelcpu/0/load/2	int
/intelcpu/0/load/0	int
/intelcpu/0/temperature/0	int
/intelcpu/0/temperature/1	int

MyUnknown...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...	/intelcpu/0...
Time	CPU Core #1	CPU Core #2	CPU Total	CPU Core #1	CPU Core #2	CPU Package	CPU Core #1	CPU Core #2	CPU Pa
09/23/202...	0	0	0	30	30	31	3113.77734	3113.77734	10.5545
09/23/202...	76.62338	76.62338	76.62338	40	42	42	3113.77734	3113.77734	7.93385
09/23/202...	100	100	100	42	47	47	3113.77563	3113.77563	13.7205

Não se preocupe em trabalhar com muitos dados neste momento. O foco agora é a aplicação da matemática computacional a um conjunto de dados que queremos estudar.

Clique em next.

Veja que as seguintes etapas são mostradas

Preparar a importação

Importar dados de arquivo

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

**Import Data**

The following tasks will now be performed. Please monitor the execution.

☐ Prepare Import  
☐ Import data file

Click [Next >] to execute.

Show Logs

< Back

Next >

Cancel

As cadeias de caracteres podem ser declaradas como codificadas em "latin1" ou "UTF-8" ou como "bytes". Essas declarações podem ser lidas por Encoding, que retornará um vetor de caracteres de valores "latin1", "UTF-8", "bytes" ou "unknown". Escolha sempre UTF-8.

### Table Data Import

**Import Data**

The following tasks will now be performed. Please monitor the execution.

☒ Prepare Import  
☒ Import data file

Row import failed with error: ("Data truncated for column 'MyUnknownColumn' at row 1", 1265)

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Uma barra de progresso mostra os dados sendo carregado para o banco.

### Aplicando Modelagem de dados aos datasets

Vamos supor que você queira somar dois números em R. Depois veremos como fica estes conceitos em Python.

Abra o RStudio e vamos trabalhar na console. Primeiro dê um CTRL + I

Limpará a console.

Clique em Tools e Check for Packages update.

```
a=2
```

```
b=3
```

```
c=a+b
```

```
c
```

```
5
```

**EX1) Faça diretamente na console as outras 3 operações diretas como a e b, subtração, multiplicação e divisão.**

## ENTRADA DE DADOS EM R

Funções de entrada de dados:

- `c()`
- `rep()`
- `seq()` OU :

Objetos de classes mais específicas como entrada de dados:

- `matrix()`
- `list()`
- `data.frame()`

Entrada de dados via teclado:

- `scan()`

Entrada de dados no formato texto:

- `readlines()`

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

**EX2) Vamos calcular a IMC agora, é bem simples trabalhar com a ferramenta.**

Temos os valores do peso e altura, a seguir, como poderia ser feito para calcular a IMC, sabendo que,

$$\text{IMC} = \text{Peso} / (\text{altura} * \text{altura})$$

Peso =75

Altura =1.80      # atenção use ponto em não vírgula para decimais

Digite no R:

```
> peso=75  
  
> altura = 1.80  
  
> IMC = 75/(1.80 *1.8)  
  
> IMC  
  
[1] 23.14815  
  
> round(IMC, 1)  
  
[1] 23.1
```

**EX3) Calculando o IMC como “ Projeto a ser aplicado a conjunto de dados”, integração de funções**

Vamos criar um dataset com pelo menos 5 colunas numéricas e 3 colunas categóricas, através do uso de **funções de criação de distribuições randômicas (aleatórias).**

Tarefas:

1. Calcular somas e médias através do uso das funções apply, lapply, sapply, etc.
2. Usar a função split
3. Criar gráficos simples

A especificação detalhada de como deve ser o dataset não foi especificada, usaremos a simulação para definir variáveis que visem a demonstração do modelo.

**Resolução:**

### **Passo 1 Conhecer o cálculo especializado de IMC**

Vamos usar um modelo de questão **do ENEM2010 - Matemática:**

**Questão 2)** Embora o índice de Massa Corporal (IMC) seja amplamente utilizado, existem ainda inúmeras restrições teóricas ao uso e às faixas de normalidade preconizadas. O Recíproco do Índice Ponderal (RIP), de acordo com o modelo alométrico(\*), possui uma melhor fundamentação matemática, já que a massa é uma variável de dimensões cúbicas e altura, uma variável de dimensões lineares. As fórmulas que determinam esses índices são:

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

$$IMC = \frac{massa(kg)}{[altura(m)]^2} \quad RIP = \frac{altura(cm)}{\sqrt[3]{massa(kg)}}$$

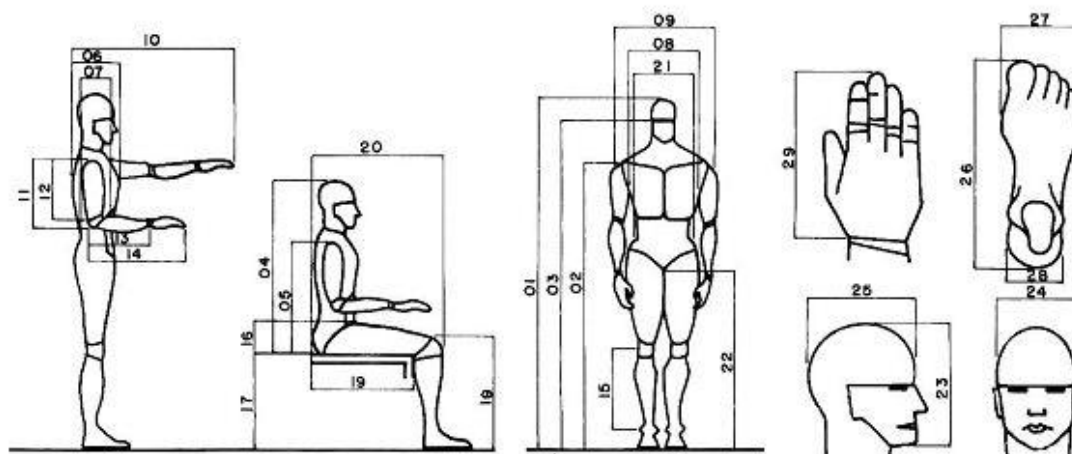
Se uma menina, com 64 kg de massa, apresenta IMC igual a 25 kg/m<sup>2</sup>, então ela possui RIP igual a:

- a) 0,4 cm/kg<sup>1/3</sup>
- b) 2,5 cm/kg<sup>1/3</sup>
- c) 8 cm/kg<sup>1/3</sup>
- d) 20 cm/kg<sup>1/3</sup>
- e) 40 cm/kg<sup>1/3</sup>

### Passo 2: Por que isso e para que serve?

(\*) **alometria** significa "crescimento ou desenvolvimento anormal ou desproporcional"

**Antropometria** é a ciência que estuda as medidas do corpo humano, a fim de estabelecer diferenças entre indivíduos, sexo, idade, raças, status socioeconômico, entre outros. As medidas **antropométricas** são usadas para avaliar o tamanho, forma e composição do corpo humano.



### Passo 3 Resolução da questão do Enem:

Para começar a resolução desse exercício, devemos fazer a **análise dimensional** das duas grandezas, o IMC e o RIP:

Para começar a resolução desse exercício, devemos fazer a análise dimensional das duas grandezas, o IMC e o RIP:

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

$$IMC = \frac{massa(kg)}{[altura(m)]^2} \quad RIP = \frac{altura(cm)}{\sqrt[3]{massa(kg)}}$$

$$[IMC] = \frac{[M]}{[L^2]} \quad [RIP] = \frac{[L]}{[\sqrt[3]{M}]}$$

$$[IMC] = [M].[L]^{-2} \quad [RIP] = [L].[M]^{-\frac{1}{3}}$$

Como sabemos o IMC e a massa da menina, fica fácil encontrarmos a altura dela. Depois disso, basta aplicarmos na fórmula do RIP esses valores, transformando a altura da menina em centímetros, de modo a calculá-lo.

$$[IMC] = \frac{[M]}{[L^2]}$$

$$25 = \frac{64}{L^2} \rightarrow L = \sqrt{\frac{64}{25}} = \frac{8}{5} = 1,6m \text{ (160 cm)}$$

$$[RIP] = [L].[M]^{-\frac{1}{3}} \rightarrow [RIP] = 160.64^{-\frac{1}{3}} = \frac{160}{4} = 40 \text{ cm/kg}^{\frac{1}{3}}$$

**Passo 4** A partir da compreensão anterior, criar o projeto e implementar:

### Criando um dataset simulado

**Contexto simulado:** estudo sobre o índice de massa corpórea em adultos jovens na qual foram inserido dados simulados antropométricos, demográficos e de alguns fatores de risco.

As variáveis coletadas estão detalhadas na tabela abaixo (incluindo o tipo da variável, sua representação em R e outras informações importantes):



# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

Variável	Observações	Tipo	Representação no R
idade	Em anos completos	Dimensional de razão, discreta	Numeric
altura	Em metros (m)	Dimensional de razão, contínua	Numeric
peso	Em quilo (Kg)	Dimensional de razão, contínua	Numeric
imc	peso/altura2	Índice dimensional de razão, contínuo	Numeric
sexo	1 = Masculino		
2 = Feminino	Nominal	Unordered Factor	
escolaridade	0 = Analfabeto		
1 = 1º grau completo			
2 = 2º grau completo			
3 = 3º grau completo			
4 = mestrado			
5 = doutorado			
6 = pós-doutorado	Ordinal	Ordered Factor	
profissao	1 = Humanas		
2 = Exatas			
3 = Biológicas	Nominal	Unordered Factor	
fumante	0 = Não		
1 = Sim	Binária	Ordered Factor	
salario	Em reais (R\$)	Dimensional de razão, contínua	Numeric
carros	Número de carros	Dimensional de razão, discreta	Numeric
filhos	Número de filhos	Dimensional de razão, discreta	Numeric

### Simulação das variáveis do dataset

Podemos propor, como teste, que as informações simuladas advém de 20.000 observações, retiradas aleatoriamente por algum processo de amostragem a partir de uma população de 200.000 indivíduos. Sendo  $n$ , número de observações e  $p$  o universo amostral.

$n <- 20000$

$p <- 200000$

Variáveis dimensionais:

A1) Idade

A variável idade (em anos completos) foi simulada a partir de uma distribuição normal, com o uso da função `rnorm` ajustada para uma média de idade de 37 anos com desvio padrão de 7 anos. Para evitar quaisquer números negativos foi utilizada a função `abs` e para manter a idade em anos completos o resultado foi arredondado para zero casas decimais com a função `round`. A função `set.seed` foi utilizada para tornar os resultados reprodutíveis.

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

```
> n <- 20000
```

Observe que já atribuímos o valor para n anteriormente

```
> set.seed(1234)
```

```
> idade <- abs(round(rnorm(n, 35, 7),0))
```

```
> summary(idade)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
6.00 30.00 35.00 34.99 40.00 61.00
```

Agora vamos representar essas idades do novo conjunto de dados

```
> idade
```

```
[1] 27 37 43 19 38 39 31 31 31 29 32 28 30 35 42 34 31 29 29 52 36 32 32 38 30 25 39 28 35 28 43 32 30
[34] 31 24 27 20 26 33 32 45 28 29 33 28 28 27 26 31 32 22 31 27 28 34 39 47 30 46 27 40 53 35 30 35 47
[67] 27 45 44 37 35 32 32 40 49 34 25 30 37 33 34 34 25 34 41 40 39 32 34 27 35 37 47 42 32 37 27 41 42
[100] 50 38 32 35 31 29 36 29 36 37 35 34 30 27 41 35 41 26 36 40 35 34 30 49 40 48 36 31 24 31 37 42 37
[133] 27 40 23 32 33 21 41 31 33 45 39 34 39 38 47 37 39 37 32 36 46 29 36 45 33 28 29 32 29 33 32 34 38
[166] 39 47 35 33 45 47 35 33 22 45 29 27 56 37 35 16 34 42 38 41 49 43 31 40 34 31 15 29 38 50 39 39 28
[199] 36 20 38 40 36 40 37 40 48 43 35 27 38 32 45 24 32 38 34 31 33 39 41 40 51 43 37 30 55 40 30 36 33
[232] 33 28 36 37 27 12 33 35 39 35 38 27 40 40 37 45 38 37 37 38 42 38 40 27 32 45 26 37 38 42 33 36 23
[265] 37 30 33 27 38 40 33 48 40 42 49 30 41 42 35 37 28 38 30 41 29 37 29 33 24 36 42 36 23 31 47 48 27
[298] 32 37 37 31 28 34 42 35 30 31 46 32 40 45 38 32 35 32 44 34 46 31 30 36 35 31 24 35 30 43 28 43 30
[331] 37 41 31 39 34 27 33 46 34 29 35 37 42 43 36 43 30 19 32 31 19 39 30 33 24 39 48 29 34 49 42 48 35
[364] 34 31 45 26 37 35 51 31 24 37 35 29 36 33 30 39 33 22 16 31 45 41 44 42 37 24 36 46 11 30 43 39 41
[397] 38 39 35 29 26 35 32 29 38 36 45 27 31 34 25 33 30 20 33 31 43 31 39 36 26 38 34 36 40 36 46 36 26
[430] 40 26 22 29 30 31 46 46 40 26 35 38 34 36 28 26 45 37 32 37 49 38 45 45 32 40 30 25 27 34 48 36 29
[463] 37 40 26 28 21 26 29 33 45 41 37 31 35 39 31 42 43 27 30 43 23 38 37 57 16 29 40 47 41 25 51 23 27
[496] 35 51 39 40 47 42 26 40 34 47 33 24 38 37 35 28 29 35 42 32 29 16 38 38 44 45 34 44 35 30 29 33 37
[529] 35 36 36 40 33 46 38 27 33 39 32 37 38 44 30 35 31 27 22 36 39 39 25 27 27 30 36 39 29 15 42 39 30
[562] 27 33 36 40 45 29 37 44 29 33 23 37 31 29 30 38 20 32 28 28 38 37 34 39 31 21 43 33 32 36 40 38 41
[595] 35 31 33 41 29 28 37 39 23 40 35 34 43 45 27 42 31 36 38 36 25 31 46 36 29 30 42 37 29 35 39 47 29
```

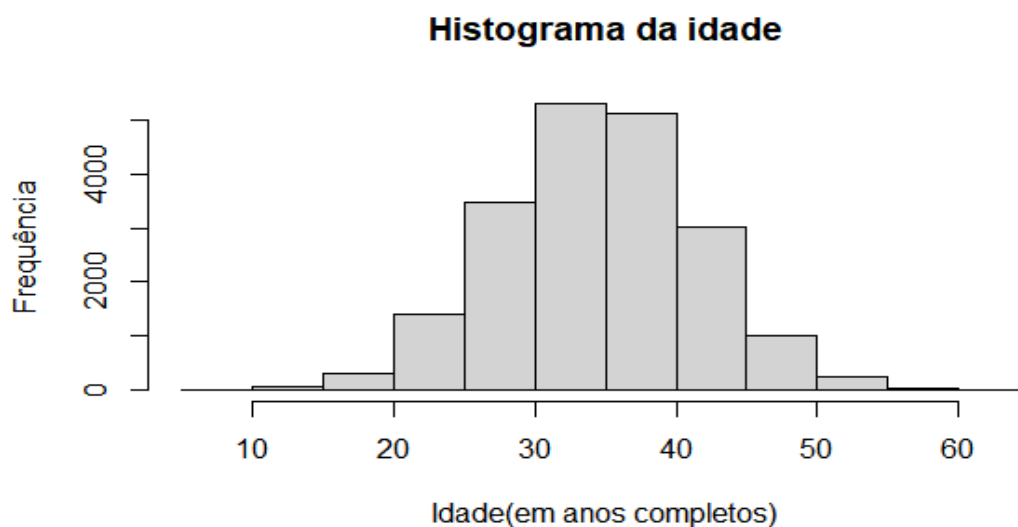
# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

```
[628] 32 42 32 45 42 23 38 25 25 35 35 35 44 37 30 36 31 35 27 34 36 30 35 30 30 34 38 34 36 42 33 34 39
[661] 26 39 31 27 23 30 27 27 46 27 42 39 29 31 32 36 33 41 33 20 34 22 42 30 27 31 39 22 37 28 35 40 39
[694] 46 36 30 50 33 41 42 38 45 34 26 32 31 37 28 40 27 47 45 32 36 32 27 16 35 35 46 21 30 41 37 36 31
[727] 36 33 36 31 43 42 44 32 33 33 45 33 36 38 40 33 51 29 25 35 47 35 36 41 34 36 43 36 30 38 27 25 34
[760] 27 39 34 30 46 34 45 34 33 38 42 54 28 34 43 37 34 26 33 34 34 46 32 24 38 34 34 39 15 31 34 39 33
[793] 28 38 35 36 43 31 26 26 28 25 35 48 34 40 27 33 39 33 40 40 31 30 52 34 35 35 31 29 27 41 28 40 45
[826] 43 34 31 27 35 33 36 31 39 40 43 51 36 31 34 24 36 32 38 33 25 35 26 41 22 34 38 24 45 29 31 36 33
[859] 28 38 31 39 31 28 34 40 37 35 21 45 35 29 34 37 24 40 37 43 42 31 32 28 32 42 48 41 22 41 42 39 29
[892] 42 40 40 32 41 36 35 45 40 15 32 25 39 32 41 38 35 34 32 36 27 40 32 30 33 48 23 33 36 31 32 38 37
[925] 30 39 36 40 26 30 34 42 31 28 40 38 26 49 45 25 50 23 36 27 32 33 37 27 16 34 30 36 29 41 35 39 28
[958] 34 37 34 47 21 40 35 42 34 16 38 33 28 30 35 52 36 28 48 32 29 26 30 49 41 40 36 37 39 37 36 42 34
[991] 37 29 46 45 22 33 39 38 40 32

[ reached getOption("max.print") -- omitted 19000 entries ]
```

```
> hist. (idade,
        main = "Histograma da idade",
        ylab = "Frequência",
        xlab= "Idade (em anos completos)")
```



Observe que pelo gráfico conseguimos compreender o conjunto de dados.

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

A simulação acima atingiu o objetivo de manter a média em 37 anos, mas o range de dados foi um pouco maior do que eu gostaria, de 11 a 60 anos, mas não comprometerá a simulação.

### A2) Altura

A altura (em metros) seguiu a mesma lógica da simulação da idade, utilizando-se uma distribuição normal com média 1,50 m e desvio padrão de 0,2 m. Entretanto, como a altura é uma variável dimensional de razão e contínua, foram utilizadas duas casas decimais na simulação.

```
> altura <- abs(round(rnorm(n, 1.50, 0.1), 2))
```

```
> summary(altura)
```

```
Min. 1st Qu. Median Mean
```

```
1.09 1.43 1.50 1.50
```

```
3rd Qu. Max.
```

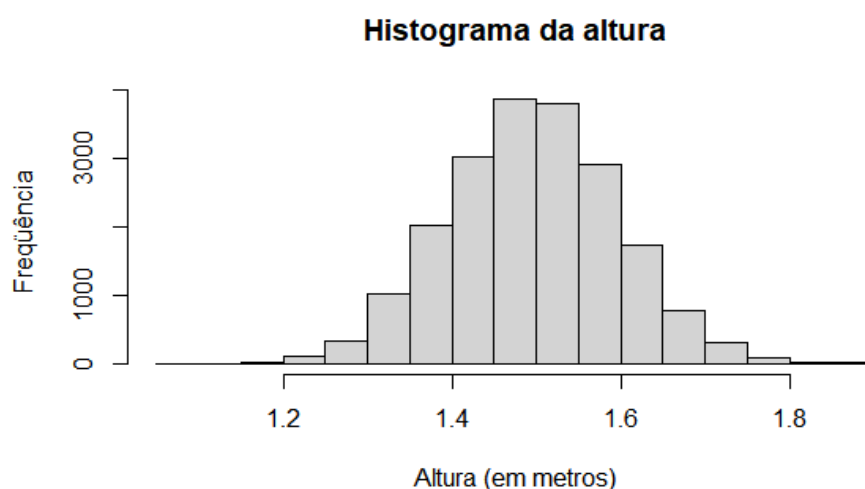
```
1.57 1.87
```

```
> hist(altura,
```

```
+ main = "Histograma da altura",
```

```
+ ylab = "Frequência",
```

```
+ xlab = "Altura (em metros)")
```



# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

### A3) Peso

A variável peso (em Kg) seguiu a mesma lógica de simulação da altura, mas com média de 55 Kg e desvio padrão de 4 Kg (com duas casas decimais).

```
> peso <- abs(round(rnorm(n, 55, 4), 2))
```

```
> summary(peso)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
------	---------	--------	------	---------	------

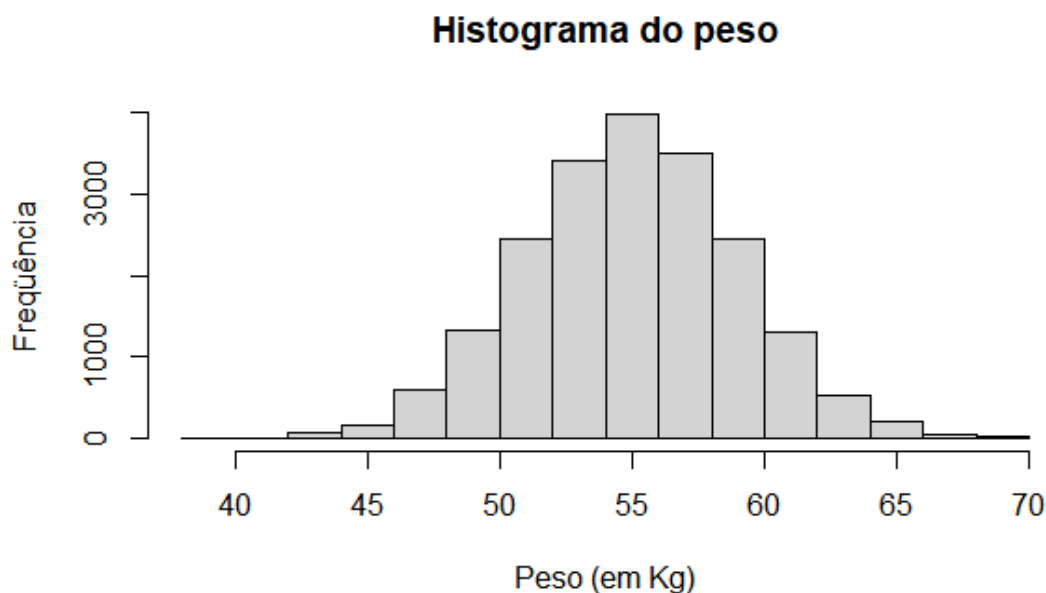
38.49	52.27	55.01	55.00	57.69	69.91
-------	-------	-------	-------	-------	-------

```
> hist(peso,
```

```
+   main = "Histograma do peso",
```

```
+   ylab = "Frequência",
```

```
+   xlab = "Peso (em Kg)"))
```



A4) índice de massa corpórea

Faça você mesmo

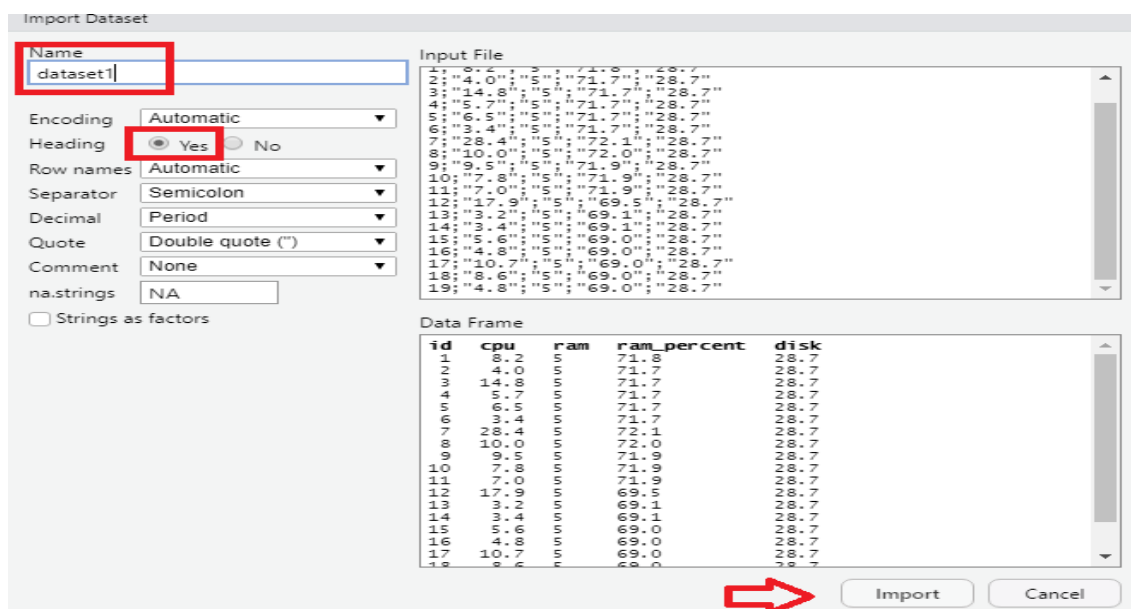
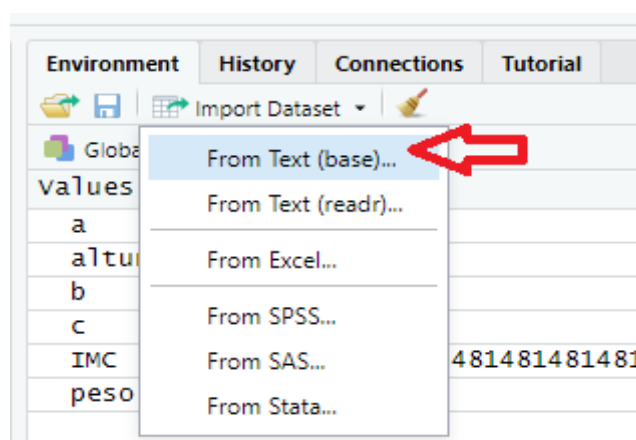
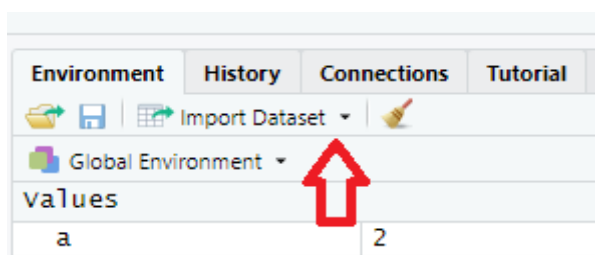
**Entrada de dados por importação de tabela**

Padrão deve ser csv

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets

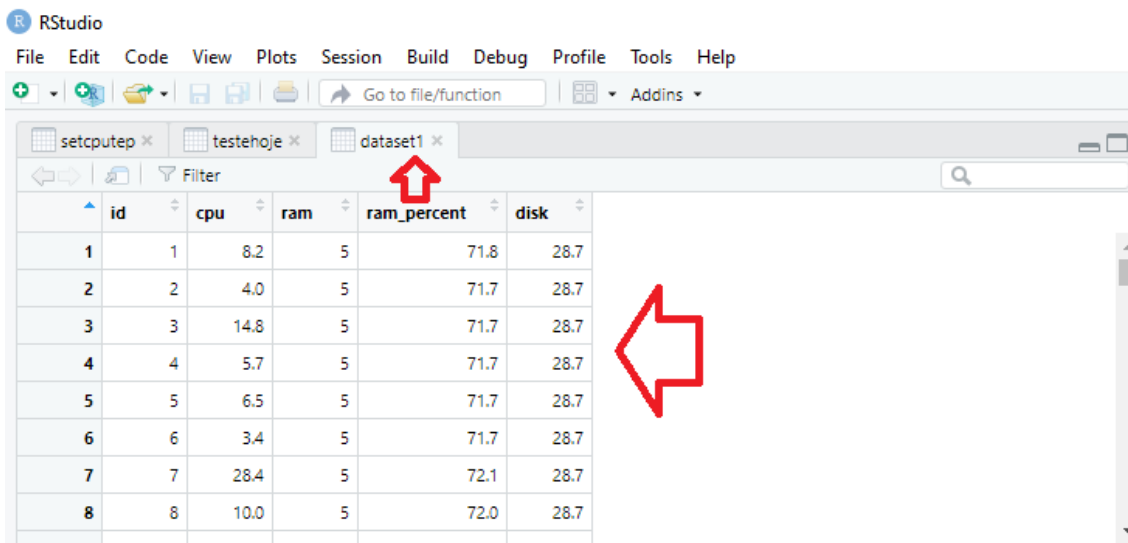
Utilizar o Wizard de importação de tabelas do RStudio



Clique em importar

# Importação e Exportação dos datasets

## Aplicando modelagem de cálculo computacional com R aos data sets



RStudio interface showing a dataset named 'dataset1' with 8 rows and 6 columns: id, cpu, ram, ram\_percent, and disk. A red arrow points to the 'ram\_percent' column header, and a large red arrow points to the right side of the table.

	id	cpu	ram	ram_percent	disk
1	1	8.2	5	71.8	28.7
2	2	4.0	5	71.7	28.7
3	3	14.8	5	71.7	28.7
4	4	5.7	5	71.7	28.7
5	5	6.5	5	71.7	28.7
6	6	3.4	5	71.7	28.7
7	7	28.4	5	72.1	28.7
8	8	10.0	5	72.0	28.7

Veja na console o resultado dessa automação:

```
dataset1 <- read.csv("C:/Users/Marise/Desktop/dados para R/dataset1.csv", sep=";")
```

Agora que temos os dados de máquina no ambiente R, vamos pensar em como utilizar estes dados e com qual objetivo.