



Instituto Infnet

# AT

Data: 05/04/2022

Nome: Eduardo Marcello Duarte

Disciplina: Fundamentos de Descrição e Análise de Dados

Professor: Luís Estevão

# Etapa 1

Leia o texto abaixo, extraído do Big Data Business em abril de 2016:

“A cada segundo, 75 sanduíches são vendidos na cadeia de restaurantes do McDonald 's em todo o mundo. Achou muito? Imagine então que existem mais de 34 mil restaurantes da rede, servindo mais de 62 milhões de pessoas diariamente em 118 países.

Todas essas informações, assim como milhões de outros dados obtidos por meio da coleta de dados, ajudam o McDonald 'sa compreender melhor seus clientes e, assim, maximizar sua presença e todo seu lucro...

## Otimização do tráfego de drive-thru

Para oferecer uma experiência mais agradável aos clientes, a cadeia de restaurantes McDonald 's conduziu um estudo com base nos dados coletados por câmeras e sistemas automatizados dos restaurantes. O objetivo era identificar os motivos que faziam com que a fila do drive-thru fosse longa e demorada.

Entre outras análises realizadas, foram avaliados três pontos principais:

1. O primeiro era referente ao design e estrutura do drive-thru, ou seja, consistia em determinar se o caminho percorrido pelo cliente desde o momento do pedido até a retirada da compra era realmente o mais adequado. Caso não fosse, o restaurante deveria passar por uma reforma que alterasse e melhorasse a situação.
2. O segundo ponto consistia em avaliar a qualidade da informação oferecida ao consumidor sobre o menu. Se o cardápio não fosse objetivo o suficiente para que os clientes conseguissem se decidir pelo pedido somente com as informações apresentadas, ele precisaria ser revisto.
3. Já o terceiro ponto avaliou quem eram os clientes que tinham preferência pelo drive-thru: por exemplo, tentaram entender os perfis que compravam, tentando identificar se eram famílias que

vinham em grandes carros para passear ou mães atarefadas que precisavam apenas pegar o lanche do filho. Dessa forma, foram criados padrões de demanda e, nos dias de hoje, já é possível prever quem são os clientes dos restaurantes e os do drive-thru.

Hoje a rede de restaurantes tem mais de 34 mil restaurantes ao redor do mundo. É uma verdadeira avalanche de dados!"

*Fonte:*

*<http://www.bigdatabusiness.com.br/como-o-mcdonalds-cria-estrategias-a-partir-de-big-data/>*

Para o caso acima, responda as perguntas abaixo:

## Questão 1

Defina com suas próprias palavras o conceito de análise de dados para o caso acima.

R: A empresa do McDonald's viu que seu tráfego do drive-thru estava muito desorganizado, então utilizaram suas câmeras e sistemas automatizados para obter os dados necessários para uma análise de dados. A análise de dados é um processo que pega os dados existentes para poder relacioná-los, de forma ordenada e estruturada, assim gerando as informações. Podemos dizer que é a porta de entrada para o conhecimento interno, ou externo. Com isso, a organização conseguiu obter informações necessárias para reformas dentro de seus restaurantes em relação aos tipos de clientes que os frequentam.

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275758>

## Questão 2

Você considera o conceito de Big Data aplicado ao caso acima? Se sim, como você define “Análise de Dados com Big Data” (diferenciando sua resposta do item anterior)?

R: Sim, pois vemos que a McDonald 's SA procurou obter dados através dos seus sistemas, o que gerou uma enorme quantidade de dados extraídos para serem processados. Big Data é a extração de uma grande coleta de dados, assim como sua leitura e interpretação para poderem ser transformados em informações. Através destes dados, semi-estruturados, a organização os organizou com base nos tipos de clientes e separando os de carro com os que consumiam dentro do local.

Fonte:

[https://novavidati.com.br/2020/12/o-que-e-big-data-conceito-e-aplicacoes/?gclid=Cj0KCQjw5-WRBhCKARIsAAId9FkXDHbIPMb9nB1ol1w1Ap3xbCvrk5F7-U3boyNgcFcXT08CD9aLdLAaAtE6EALw\\_wcB](https://novavidati.com.br/2020/12/o-que-e-big-data-conceito-e-aplicacoes/?gclid=Cj0KCQjw5-WRBhCKARIsAAId9FkXDHbIPMb9nB1ol1w1Ap3xbCvrk5F7-U3boyNgcFcXT08CD9aLdLAaAtE6EALw_wcB)

<https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275758>

## Questão 3

Você acredita que foram aplicadas regras e medidas estatísticas no caso acima? Se sim, quais delas você destacaria e por que (para que)?

R: Sim, as regras usadas foram:

- **População:** que são os elementos dentro de um estudo, no caso do McDonald 's foram os clientes.
- **Elementos:** A empresa verificou 3 exemplos grandes para serem analisados, que poderiam ser as causas para as longas filas e demoras, e eles são:
  - A estrutura do drive-thru.
  - O cardápio do restaurante.
  - A preferência pelo drive-thru.

- **Amostra:** Como o foco do problema era o drive-thru então a empresa deve ter criado um foco maior neste momento, então a investigação foi mais forte nesta área.

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275758>

## Questão 4

Para o caso acima foi aplicado alguma análise descritiva, diagnóstica, preditiva e/ou prescritiva? Justifique.

R: Sim, a análise diagnóstica. Dentro do caso mostra claramente que a empresa queria entender o motivo do seu tráfego do drive-thru ser longa e demorada, ou seja casos que aconteceram e estavam acontecendo e para isso eles usaram seus sistemas automatizados e câmeras para entender, o que bate com uma técnica chamada **Data discovery and visualization** que visa entender o motivo das filas serem longas, e o uso de outra técnica **OLAP cubes** que tenta entender os padrões de consumo.

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275764>

## Etapa 2

Considere o arquivo [vendasExemploBR.xlsx](#) para as questões abaixo:

### Questão 1

Caracterize cada coluna quanto ao tipo de dados.

R:

- Ordem: Ordinal
- Nome: Nominal
- Estado: Nominal
- Vendas: Intervalar de Razão
- Desconto: Intervalar de Razão
- Produto: Nominal
- Data: Ordinal
- Segmento: Nominal

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275771>

## Questão 2

Caracterize cada coluna quanto aos tipos de escala (quando aplicável).

R:

- Ordem: escala nominal
- Nome: escala nominal
- Estado: escala nominal
- Vendas: escala de intervalo
- Desconto: escala de intervalo
- Produto: escala nominal
- Data: escala ordinal
- Segmento: escala nominal

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275771>

## Questão 3

Caracterize cada coluna quanto a dados discretos e contínuos.

R:

- Ordem: Contínuo
- Nome: Discreto
- Estado: Discreto
- Vendas: Contínuo
- Desconto: Contínuo
- Produto: Discreto
- Data: Contínuo
- Segmento: Discreto

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275771>

## Questão 4

Caracterize cada coluna quanto a dados qualitativos e quantitativos.

R:

- Ordem: Qualitativos.
- Nome: Qualitativos.

- Estado: Qualitativos.
- Vendas: Quantitativos.
- Desconto: Quantitativos.
- Produto: Qualitativos.
- Data: Quantitativos.
- Segmento: Qualitativos.

Fonte: <https://lms.infnet.edu.br/moodle/mod/page/view.php?id=275771>

## Questão 5

Cite 2 tipos de problemas caso ocorra falta de qualidade de dados. Justifique.

R:

1. **Data errada:** Caso uma data estivesse errada os cálculos das vendas sairiam da linha esperada do mês, o que poderia causar uma perda de tempo na busca pelo erro.
2. **Desconto errado:** Caso o desconto não tenha sido aplicado da forma correta, fará com que os cálculos finais deem diferença, acarretando em problemas no planejamento de vendas.

## Questão 6

Para o caso anterior, o que você faria para corrigir o erro do dado para evitar que ocorram erros nas análises dos dados? Justifique.

R: Sempre ao final de cada dia, ou semana, gerar uma contagem das ordens, verificando cada detalhe da mesma. Pois se um dado não é mais confiável é preciso filtrar novamente cada dado para encontrar a falha, caso dê respostas positivas poderá seguir com o planejado, ou terá que tomar novas providências.

## Questão 7

Cite uma possível variável calculada e qual o seu propósito.

R: Vendas por desconto, pois seu propósito é mostrar o valor final das vendas feitas por um indivíduo cujo o nome se encontra na coluna "NOME" e deve ser usada como uma possível comparação para um planejamento de alcance de metas.

## Etapa 3

Considere o arquivo vendasExemploBR.xlsx e o RStudio para as questões abaixo:

### Questão 1

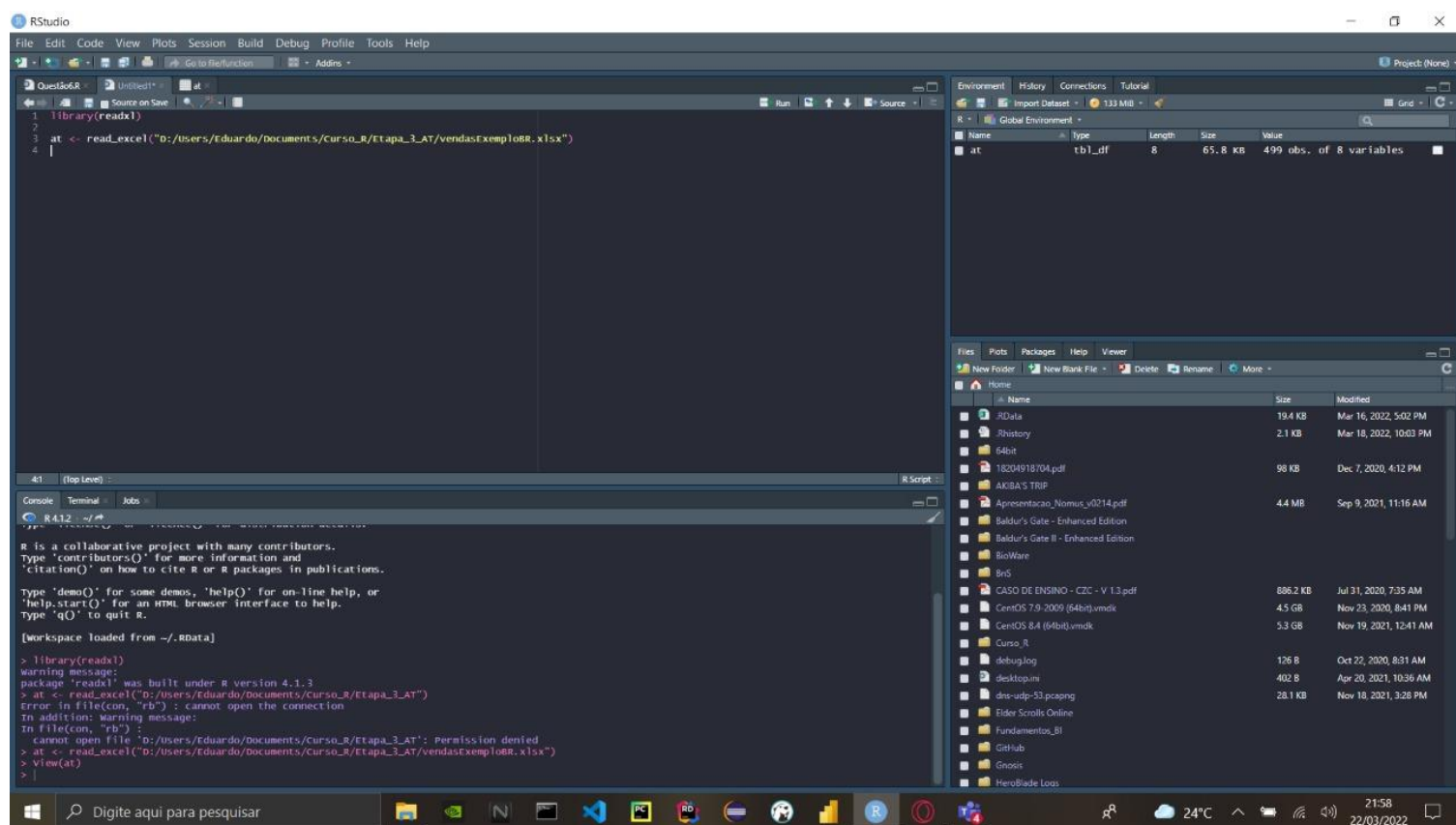
Diferenciam, conceitualmente, o motor R e RStudio.

R: O Motor R é a linguagem de programação sendo muito utilizada para uma análise de dados iterativa é a ferramenta mais usada pelos estatísticos, cientistas e cientistas de dados. O RStudio é um ambiente de desenvolvimento próprio para a linguagem de programação R que permite utilizar as funcionalidades só R de forma prática, simples e visual.

### Questão 2

Carregue o arquivo xls no RStudio.

R:





RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Project: (None)

Environment History Connections Tutorial

R Global Environment

Name	Type	Length	Size	Value
at	tbl_df	8	65.8 KB	499 obs. of 8 variables

Files Plots Packages Help Viewer

New Folder New Blank File Delete Rename More

Home

Name	Size	Modified
RData	19.4 KB	Mar 16, 2022, 5:02 PM
Rhistory	2.1 KB	Mar 18, 2022, 10:03 PM
64bit		
18204918704.pdf	98 KB	Dec 7, 2020, 4:12 PM
AKIBA'S TRIP		
Apresentacao_Nomus_v0214.pdf	4.4 MB	Sep 9, 2021, 11:16 AM
Baldur's Gate - Enhanced Edition		
Baldur's Gate II - Enhanced Edition		
BioWare		
BnS		
CASO DE ENSINO - CZC - V.1.3.pdf	886.2 KB	Jul 31, 2020, 7:35 AM
CentOS 7.9-2009 (64bit).vmdk	4.5 GB	Nov 23, 2020, 8:41 PM
CentOS 8.4 (64bit).vmdk	5.3 GB	Nov 19, 2021, 12:41 AM
Curso_R		
debuglog	126 B	Oct 22, 2020, 8:31 AM
desktop.ini	402 B	Apr 20, 2021, 10:36 AM
dns-udp-53.pcapng	28.1 KB	Nov 18, 2021, 3:26 PM
Elder Scrolls Online		
Fundamentos_BI		
GitHub		
Gnosis		
HeroBlade Loas		

Showing 1 to 20 of 499 entries, 8 total columns

Console Terminal Jobs

```
R 4.1.2 ~\#
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[workspace loaded from ~\.RData]

> library(readxl)
warning message:
package 'readxl' was built under R version 4.1.3
> at <- read_excel("D:/Users/Eduardo/Documents/Curso_R/Etapa_3_AT")
Error in file(con, "rb") : cannot open the connection
In addition: warning message:
In file(con, "rb") :
cannot open file 'D:/Users/Eduardo/Documents/Curso_R/Etapa_3_AT': Permission denied
> at <- read_excel("D:/Users/Eduardo/Documents/Curso_R/Etapa_3_AT/VendasExemploR.xlsx")
> view(at)
> |
```

Digite aqui para pesquisar

24°C 21:58 22/03/2022

### Questão 3

Aplique o valor real da venda para cada linha da tabela, a saber coluna VENDA - percentual do DESCONTO aplicado.

R:

The screenshot displays the RStudio interface. The main window shows a data frame with the following columns: ORDEM, INOME, ESTADO, VENDAS, DESCONTO, VENDAS\_POR\_DESCONTO, PRODUTO, DATA, and SEGMENTO. The data is sorted by ORDEM. A tooltip for the 'VENDAS\_POR\_DESCONTO' column indicates it is numeric with a range of 0 to 24000. The console shows the following R code:

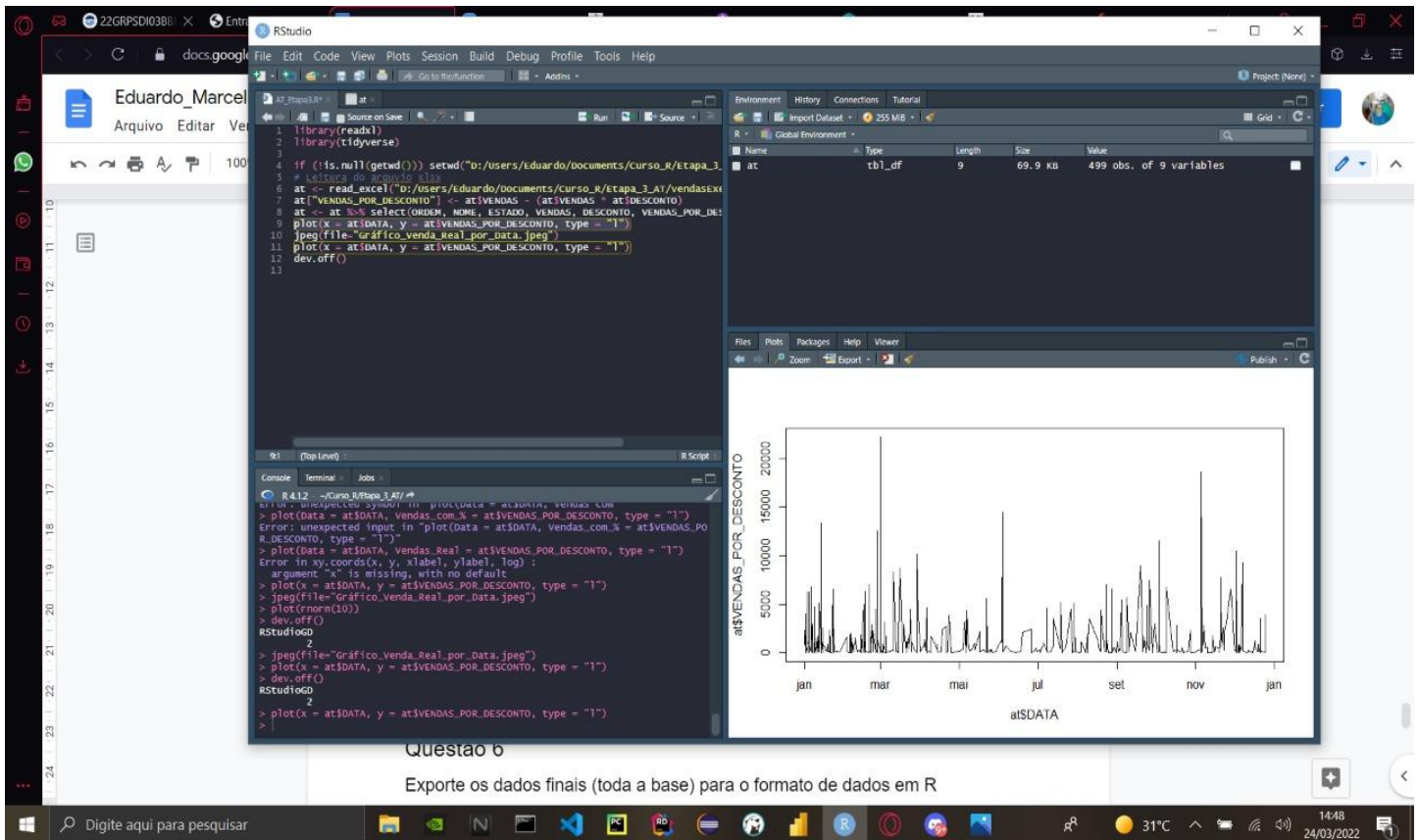
```
R 4.1.2 - ~/Curso R/Etapa 3 AT/ #
> plot(data = at$VENDAS_POR_DESCONTO, type = "l")
Error: unexpected input in "plot(data = at$VENDAS_POR_DESCONTO, type = "l")"
> plot(data = at$VENDAS_POR_DESCONTO, type = "l")
Error in xy.coords(x, y, xlab, ylab, log) :
  argument "x" is missing, with no default
> plot(x = at$VENDAS_POR_DESCONTO, type = "l")
> jpeg(file="Gráfico_Venda_Real_por_Data.jpeg")
> plot(rnorm(10))
> dev.off()
RStudio
2
> jpeg(file="Gráfico_Venda_Real_por_Data.jpeg")
> plot(x = at$VENDAS_POR_DESCONTO, type = "l")
> dev.off()
RStudio
2
> plot(x = at$VENDAS_POR_DESCONTO, type = "l")
```

Obs: A resposta se encontra na coluna “VENDAS\_POR\_DESCONTO”.

## Questão 4

Plote o valor de venda real (item anterior) por dia em um gráfico de linhas.

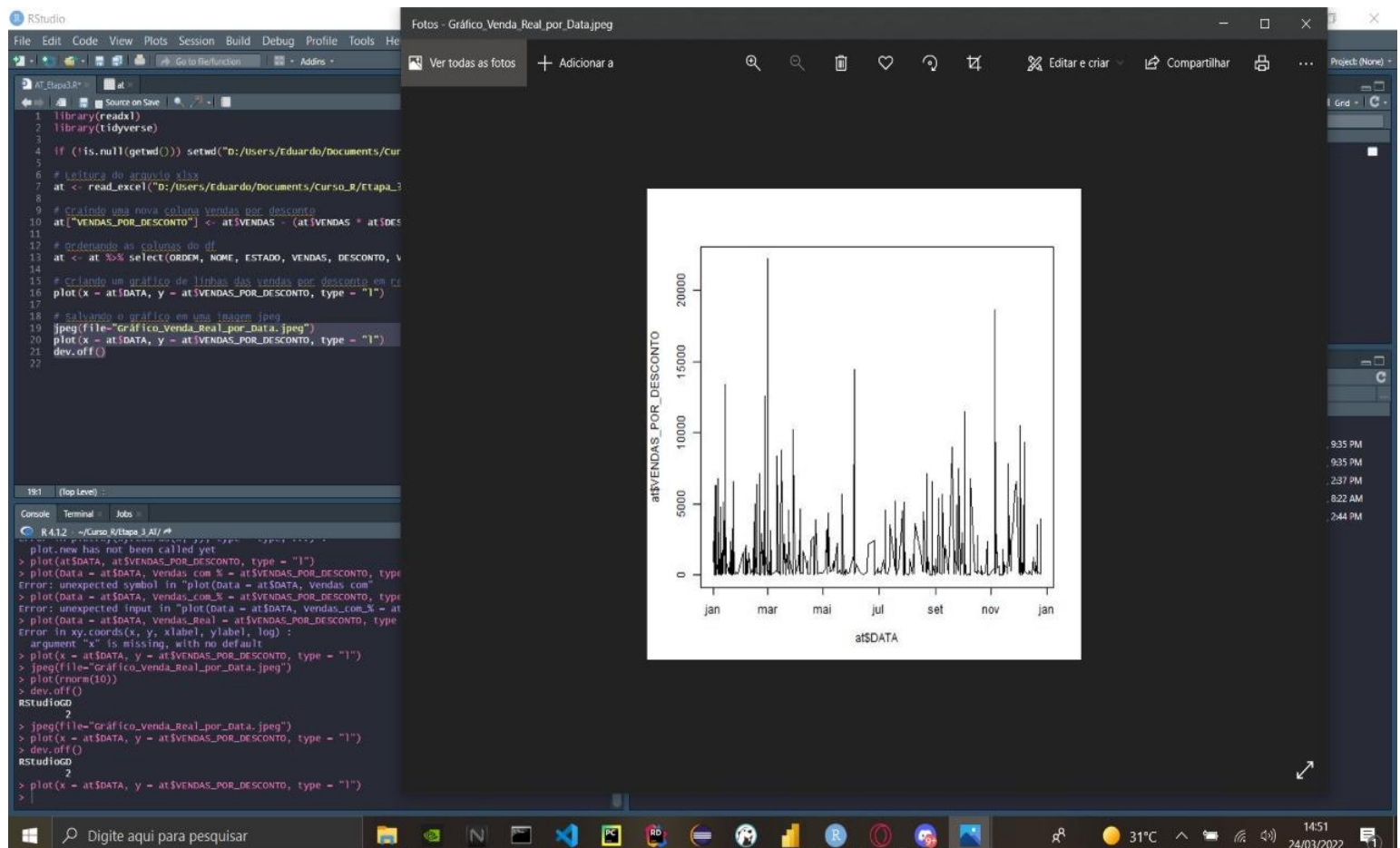
R:



## Questão 5

Exporte a imagem.

R:



## Questão 6

Exporte os dados finais (toda a base) para o formato de dados em R

R:

