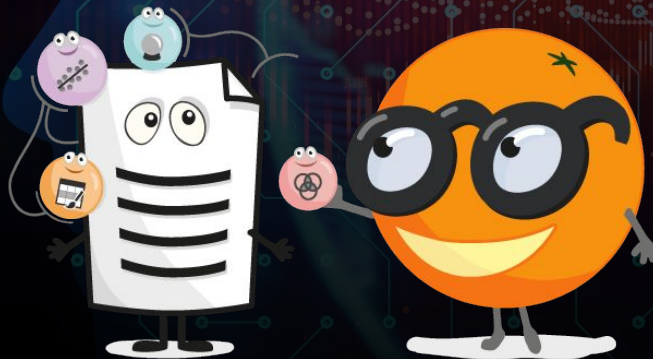


# Introdução à Ferramenta de Mineração de Dados Orange



Helvécio B. Leal Neto

São José dos Campos - Junho de 2023

# Quem sou eu?

- Doutorando em Computação Aplicada pelo Instituto Nacional de Pesquisas Espaciais. (INPE - 2021 ~ 2024).
- Mestre em Computação Aplicada no (INPE - 2021).
- Bacharel em Ciência da Computação Universidade Federal do Oeste do Pará (UFOPA - 2018).
- Tecnólogo em Redes de Computadores pelo Instituto Esperança de Ensino Superior (IESPES - 2013).



Atualmente trabalhando nas áreas de ciência de dados, inteligência artificial aplicada ao monitoramento ambiental e sistemas automáticos para rastreamento e previsão de tempestades em curto prazo (**nowcasting**).

CV - <http://lattes.cnpq.br/7196274083231131>

GitHub: [github.com/helvecioneto/](https://github.com/helvecioneto/)

Kaggle: <https://www.kaggle.com/helvecioneto>

E-mail:

[helvecio.neto@inpe.br](mailto:helvecio.neto@inpe.br)

[helvecioblneto@gmail.com](mailto:helvecioblneto@gmail.com)

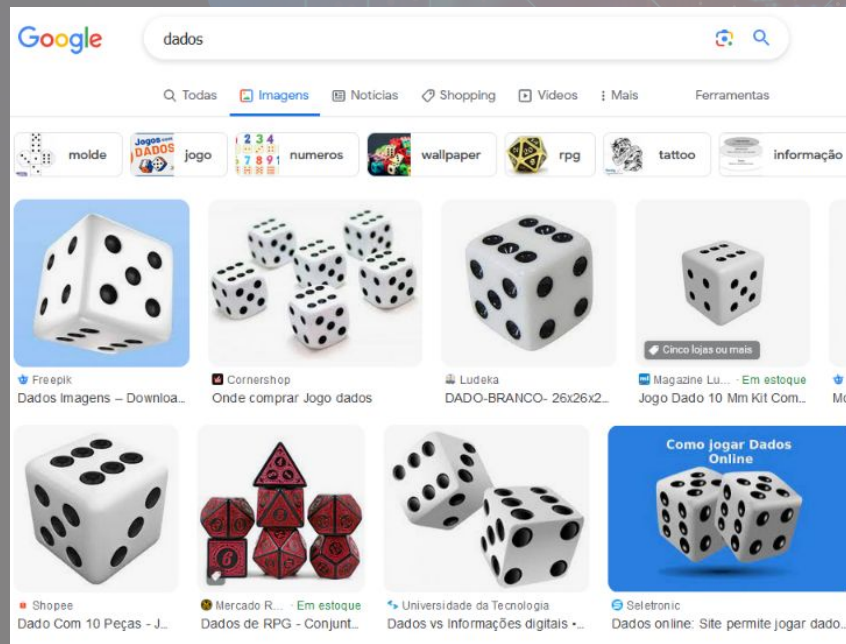
Skype: helvecio.neto

# Introdução

O que são dados ?

# Introdução

O que são dados ?





# Introdução

O que são dados ?



O que são dados ?



Dados são informações coletadas e registradas sobre fatos, eventos, entidades ou medidas.



Eles são elementos brutos e objetivos que, por si só, não possuem significado ou contexto específico. Os dados podem ser representados de várias formas, como números, palavras, imagens, áudios, entre outros formatos.

# Introdução



<https://www.statista.com/chart/25443/estimated-amount-of-data-created-on-the-internet-in-one-minute/> - <https://youtu.be/S8ePpLyTIA>

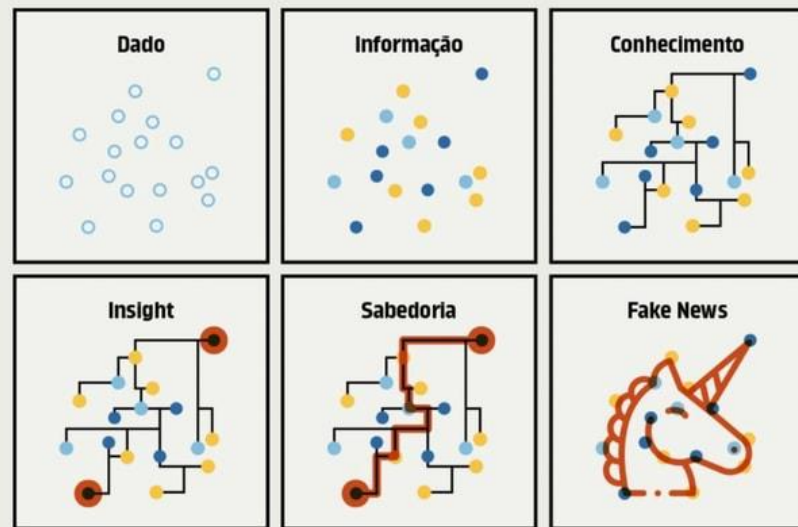
# Introdução - Mineração de dados

Mineração de dados é o processo de **descobrir** padrões, relacionamentos e **informações** úteis a partir de grandes **quantidades de dados**, com o objetivo de **transformar** esses dados em **conhecimento** significativo para tomada de **decisões**. (Fayyad, Piatetsky-Shapiro e Smyth, 1996).



# Introdução - Mineração de dados

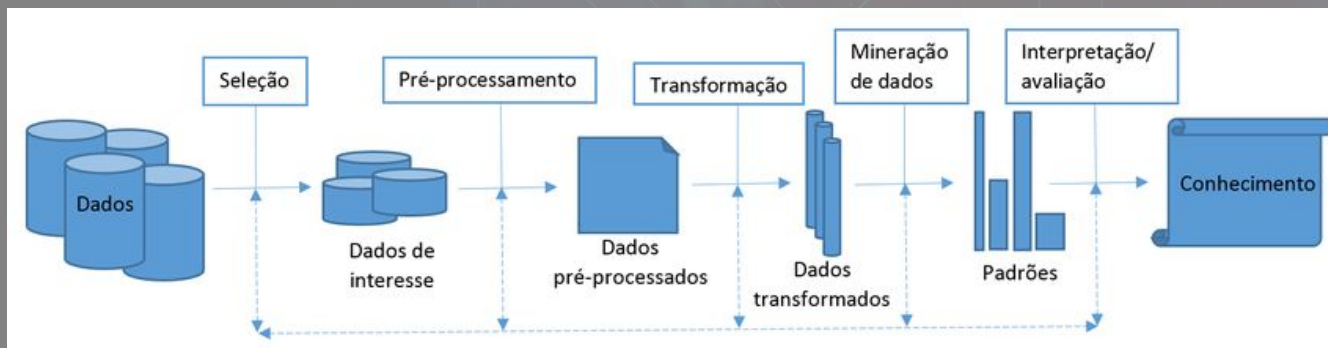
Para diversos autores a mineração de dados é considerada parte do processo de Descoberta de Conhecimento em Base de Dados (**KDD** - *Knowledge Discovery in Databases*) (Fayyad, Piatetsky-Shapiro e Smyth, 1996).





# Introdução - Mineração de dados

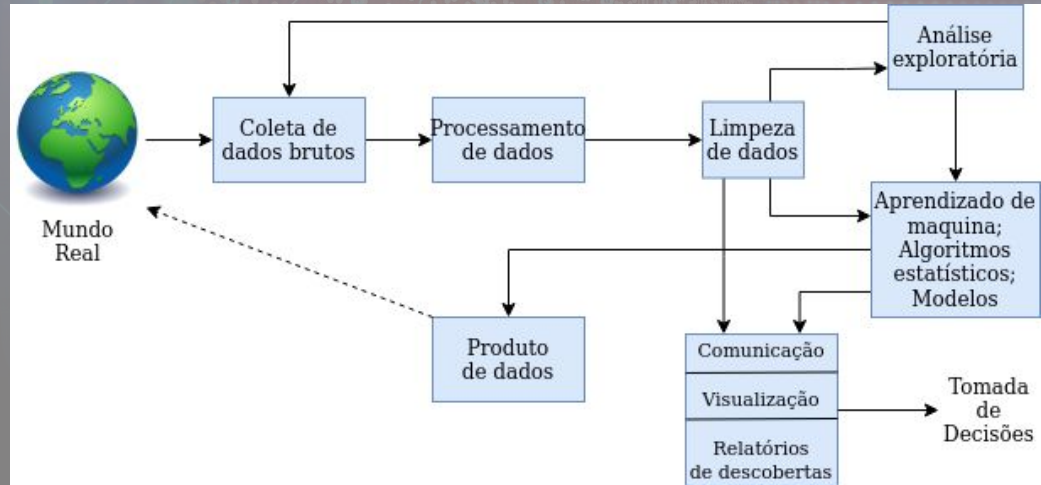
Para diversos autores a mineração de dados é considerada parte do processo de Descoberta de Conhecimento em Base de Dados (**KDD** - *Knowledge Discovery in Databases*) (Fayyad, Piatetsky-Shapiro e Smyth, 1996).



# Introdução - Mineração de dados

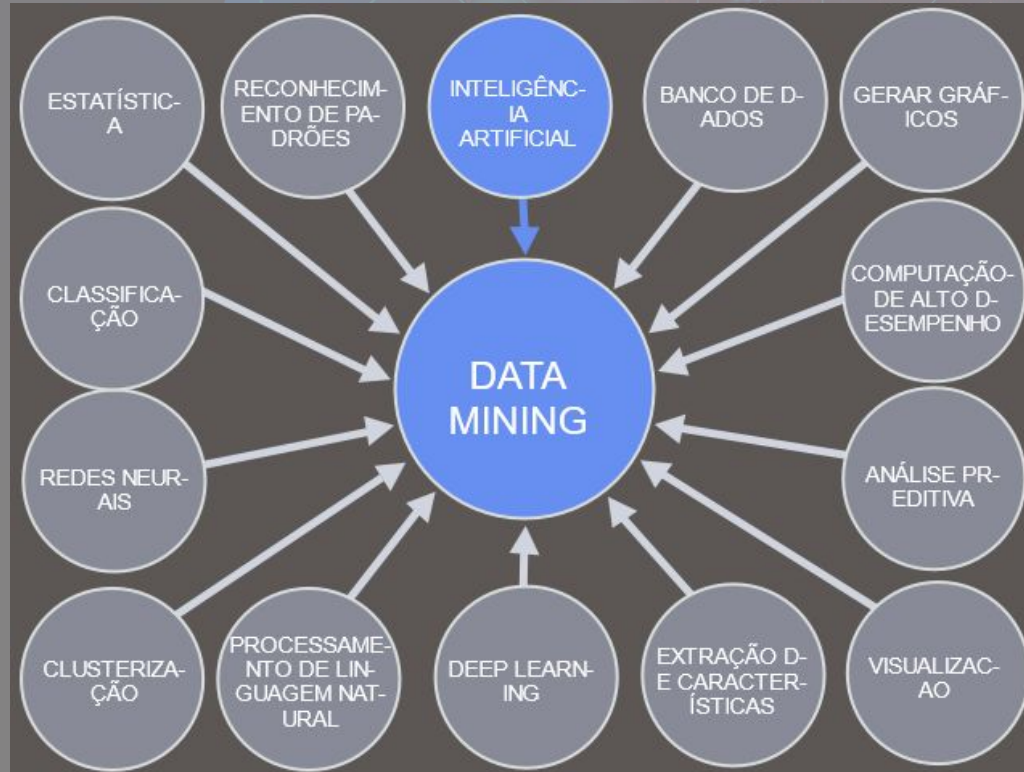
Quais são as etapas do processo de Mineração de dados?

- Planejamento estratégico
- Seleção de dados
- Modelagem de dados
- Avaliação dos resultados



O'Neil e Schutt, (2013).

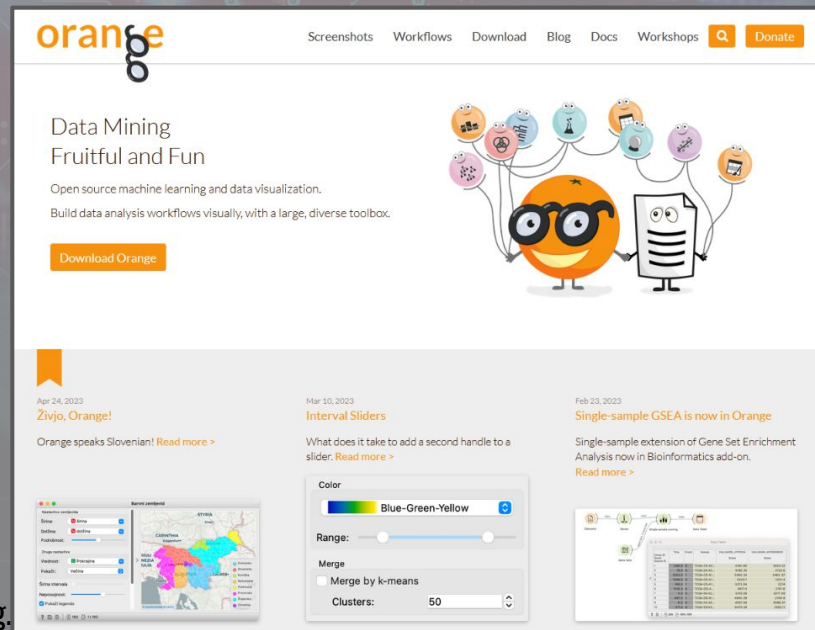
# Introdução - Mineração de dados



# Introdução - Orange

O Orange é uma ferramenta de código aberto que oferece uma interface gráfica intuitiva para facilitar a análise de dados e a criação de modelos preditivos.

- Importação de conjuntos de dados.
- Pré-processamento:
  - a. limpeza
  - b. transformação
  - c. redução de dimensionalidade.
- Visualização: gráficos e representações interativas
- Aprendizado de máquina.
  - a. Divisão de conjuntos de dados em treino e teste.
  - b. Construção de modelos de classificação, regressão e clustering.



<https://orangedatamining.com/>



# Introdução - Orange (Download e Instalação)



The screenshot shows the Orange Data Mining website. At the top, the 'orange' logo is on the left, and navigation links for 'Screenshots', 'Workflows', 'Download', 'Blog', 'Docs', 'Workshops', a search icon, and a 'Donate' button are on the right. The main content area features the text 'Data Mining Fruitful and Fun', a description of the software as open source machine learning and data visualization tools, and a 'Download Orange' button. A large cartoon illustration of an orange character with glasses and a notepad is positioned to the right. Below this, the URL <https://orangedatamining.com/> is displayed in large, bold, black text. The bottom section contains three news items: 'Zivjo, Orange!' dated April 24, 2023, 'Interval Sliders' dated March 10, 2023, and 'Single-sample GSEA is now in Orange' dated February 23, 2023. Each item includes a brief description and a 'Read more >' link. Below the news items are three small preview images: a map of Slovenia, a 'Color' widget interface showing a 'Blue-Green-Yellow' gradient and a 'Range' slider, and a 'Merge by k-means' widget interface showing a 'Clusters' value of 50.

orange

Screenshots Workflows Download Blog Docs Workshops [Donate](#)

Data Mining  
Fruitful and Fun

Open source machine learning and data visualization.  
Build data analysis workflows visually, with a large, diverse toolbox.

[Download Orange](#)

<https://orangedatamining.com/>

Apr 24, 2023  
**Zivjo, Orange!**  
Orange speaks Slovenian! [Read more >](#)

Mar 10, 2023  
**Interval Sliders**  
What does it take to add a second handle to a slider. [Read more >](#)

Feb 23, 2023  
**Single-sample GSEA is now in Orange**  
Single-sample extension of Gene Set Enrichment Analysis now in Bioinformatics add-on. [Read more >](#)







# Introdução - Orange (Download e Instalação)

**orange**

Screenshots Workflows Download Blog Docs Workshops

Data Mining  
Fruitful and Fun

Open source machine learning and data visualization.  
Build data analysis workflows visually, with a large, diverse toolbox.

<https://orangedatamining.com/>

**News**

**Apr 24, 2023**  
**Zivjo, Orange!**  
Orange speaks Slovenian! [Read more >](#)

**Mar 10, 2023**  
**Interval Sliders**  
What does it take to add a second handle to a slider. [Read more >](#)

**Feb 23, 2023**  
**Single-sample GSEA is now in Orange**  
Single-sample extension of Gene Set Enrichment Analysis now in Bioinformatics add-on. [Read more >](#)

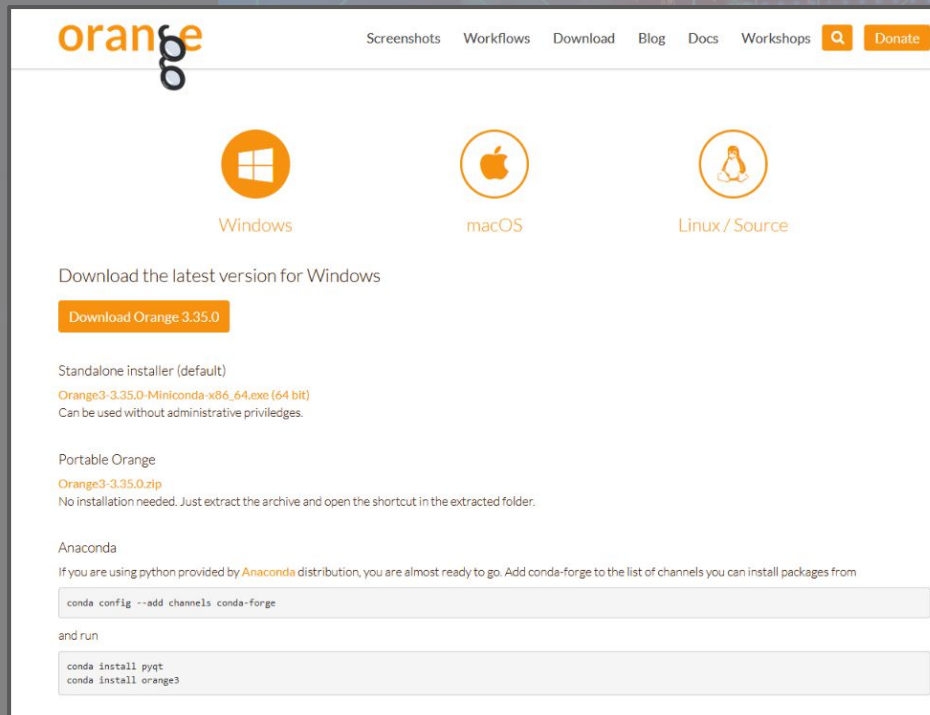
**Workflow**

**Color**  
Blue-Green-Yellow  
Range:   
Merge  
☐ Merge by k-means  
Clusters: 50

**Workflow**

| Time | Class | Score    | Score (normalized) | Score (log-odds) |
|------|-------|----------|--------------------|------------------|
| 1    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 2    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 3    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 4    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 5    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 6    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 7    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 8    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 9    | 0     | 0.000000 | 0.000000           | 0.000000         |
| 10   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 11   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 12   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 13   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 14   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 15   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 16   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 17   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 18   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 19   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 20   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 21   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 22   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 23   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 24   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 25   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 26   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 27   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 28   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 29   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 30   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 31   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 32   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 33   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 34   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 35   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 36   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 37   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 38   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 39   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 40   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 41   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 42   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 43   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 44   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 45   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 46   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 47   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 48   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 49   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 50   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 51   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 52   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 53   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 54   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 55   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 56   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 57   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 58   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 59   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 60   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 61   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 62   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 63   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 64   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 65   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 66   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 67   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 68   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 69   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 70   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 71   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 72   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 73   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 74   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 75   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 76   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 77   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 78   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 79   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 80   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 81   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 82   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 83   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 84   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 85   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 86   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 87   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 88   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 89   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 90   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 91   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 92   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 93   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 94   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 95   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 96   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 97   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 98   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 99   | 0     | 0.000000 | 0.000000           | 0.000000         |
| 100  | 0     | 0.000000 | 0.000000           | 0.000000         |

# Introdução - Orange (Download e Instalação)



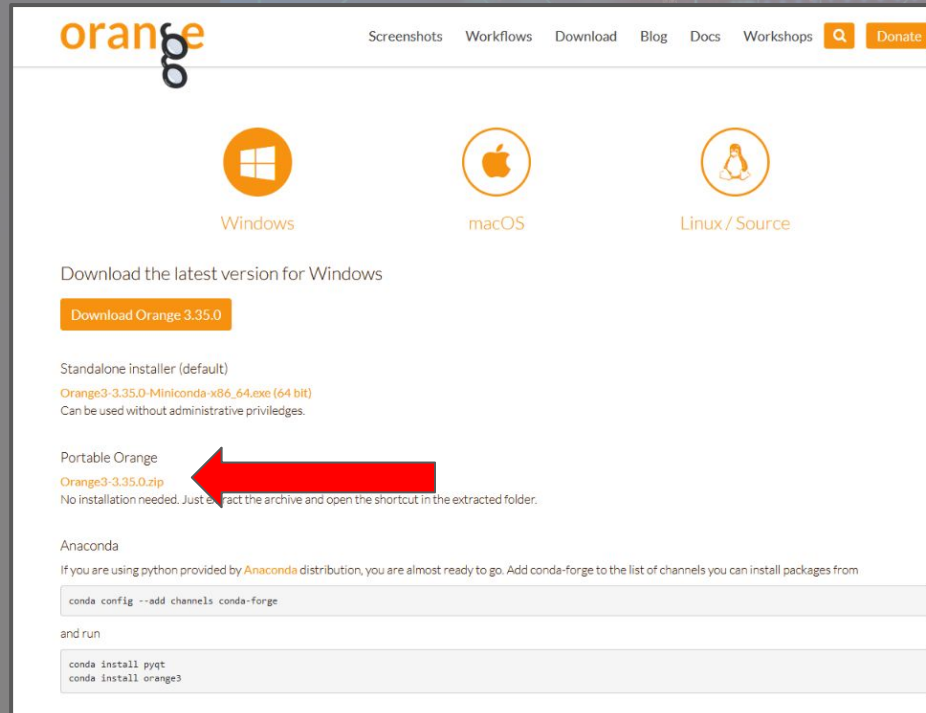
The screenshot shows the Orange3 website with the following content:

- Navigation Bar:** Screenshots, Workflows, Download, Blog, Docs, Workshops, a search icon, and a Donate button.
- Platform Icons:** Three large orange icons representing Windows, macOS (Apple logo), and Linux (Tux penguin).
- Windows Section:**
  - Text: "Download the latest version for Windows"
  - Button: "Download Orange 3.35.0"
  - Text: "Standalone installer (default)"
  - Text: "Orange3-3.35.0-Miniconda-x86\_64.exe (64 bit)"
  - Text: "Can be used without administrative privileges."
- Portable Orange Section:**
  - Text: "Portable Orange"
  - Text: "Orange3-3.35.0.zip"
  - Text: "No installation needed. Just extract the archive and open the shortcut in the extracted folder."
- Anaconda Section:**
  - Text: "Anaconda"
  - Text: "If you are using python provided by **Anaconda** distribution, you are almost ready to go. Add conda-forge to the list of channels you can install packages from"
  - Code Block:

```
conda config --add channels conda-forge
```
  - Text: "and run"
  - Code Block:

```
conda install pyqt
conda install orange3
```

# Introdução - Orange (Download e Instalação)



The screenshot shows the Orange3 website with the following content:

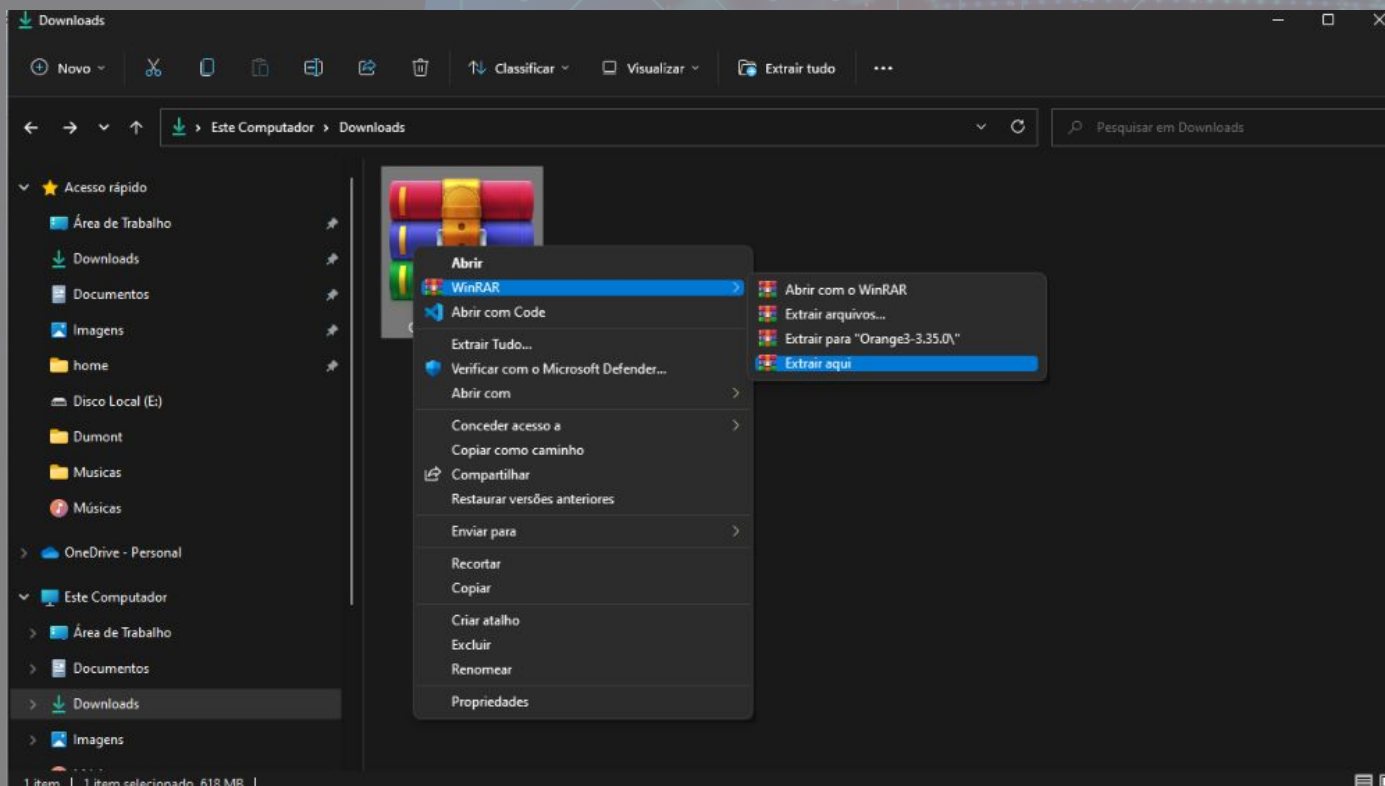
- Navigation bar: Screenshots, Workflows, Download, Blog, Docs, Workshops, Search, Donate
- Operating system icons: Windows, macOS, Linux / Source
- Text: Download the latest version for Windows
- Button: Download Orange 3.35.0
- Section: Standalone installer (default)  
Orange3-3.35.0-Miniconda-x86\_64.exe (64 bit)  
Can be used without administrative privileges.
- Section: Portable Orange  
Orange3-3.35.0.zip  
No installation needed. Just extract the archive and open the shortcut in the extracted folder. (A red arrow points to this section)
- Section: Anaconda  
If you are using python provided by Anaconda distribution, you are almost ready to go. Add conda-forge to the list of channels you can install packages from
- Code block:

```
conda config --add channels conda-forge
```
- Text: and run
- Code block:

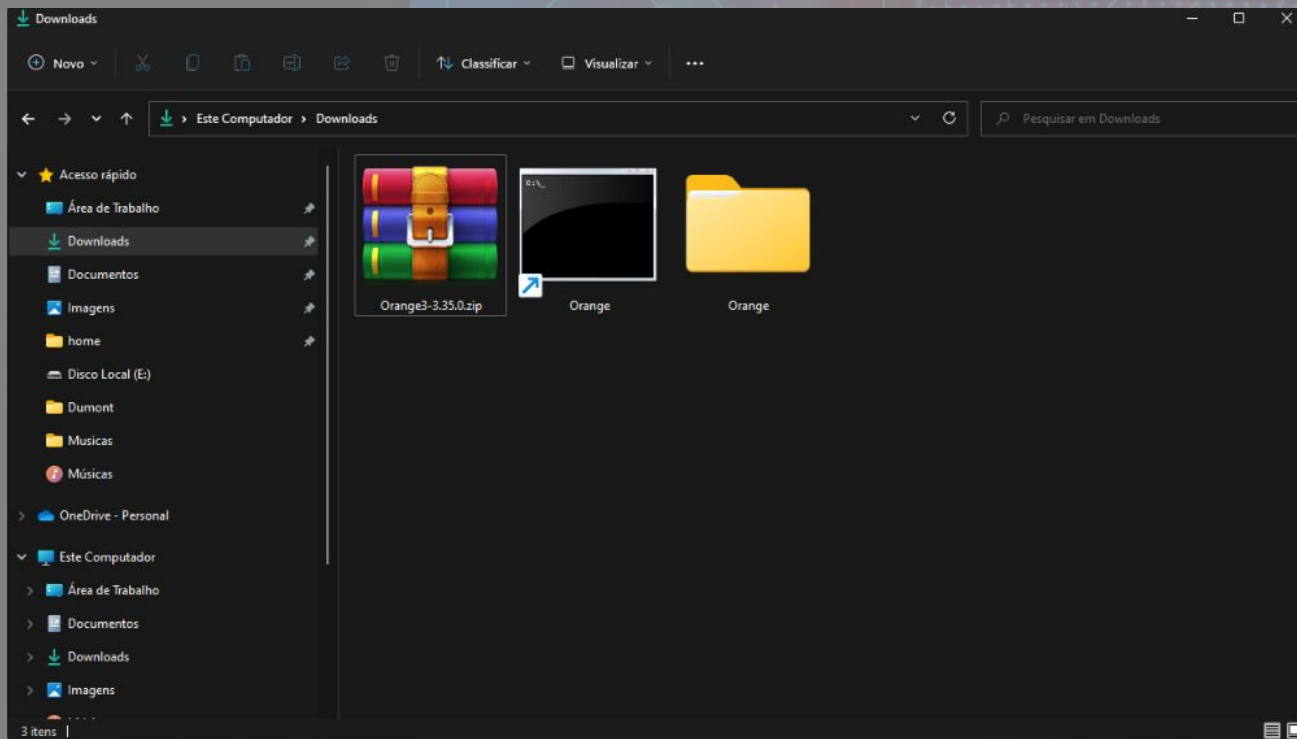
```
conda install pyqt  
conda install orange3
```



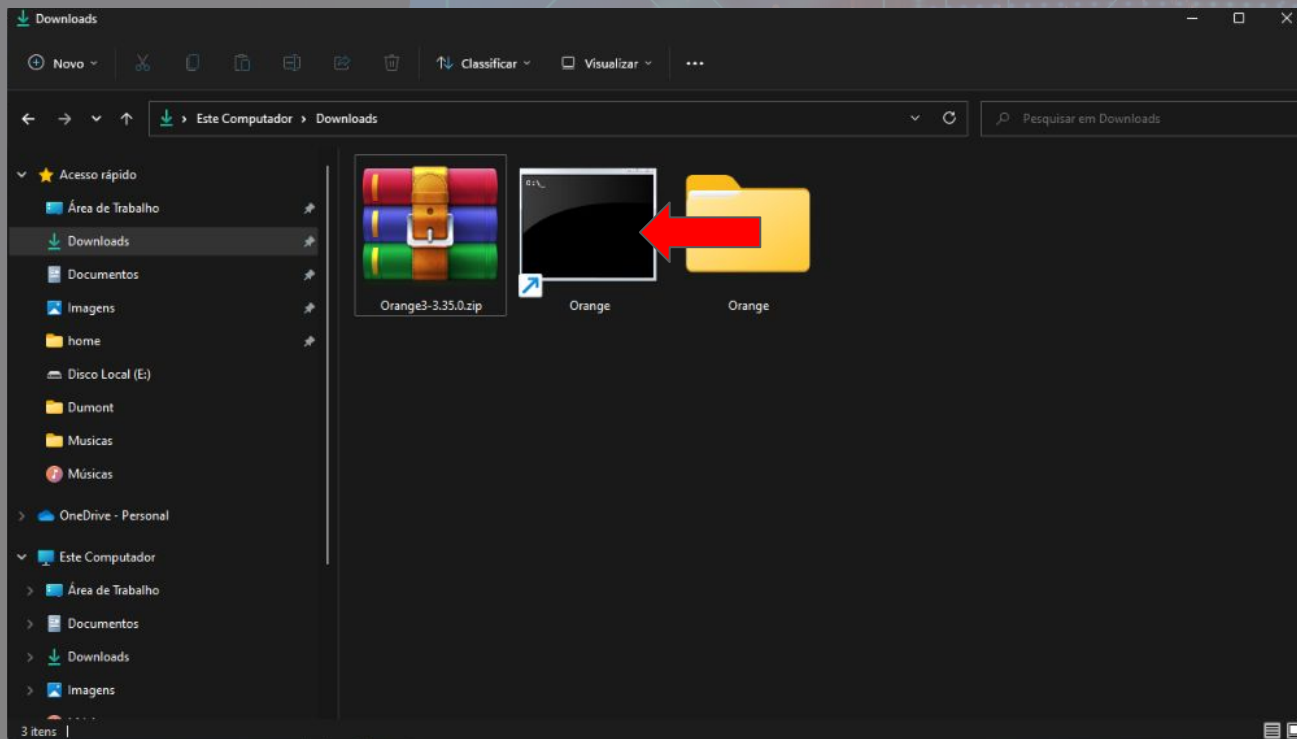
# Introdução - Orange (Download e Instalação)



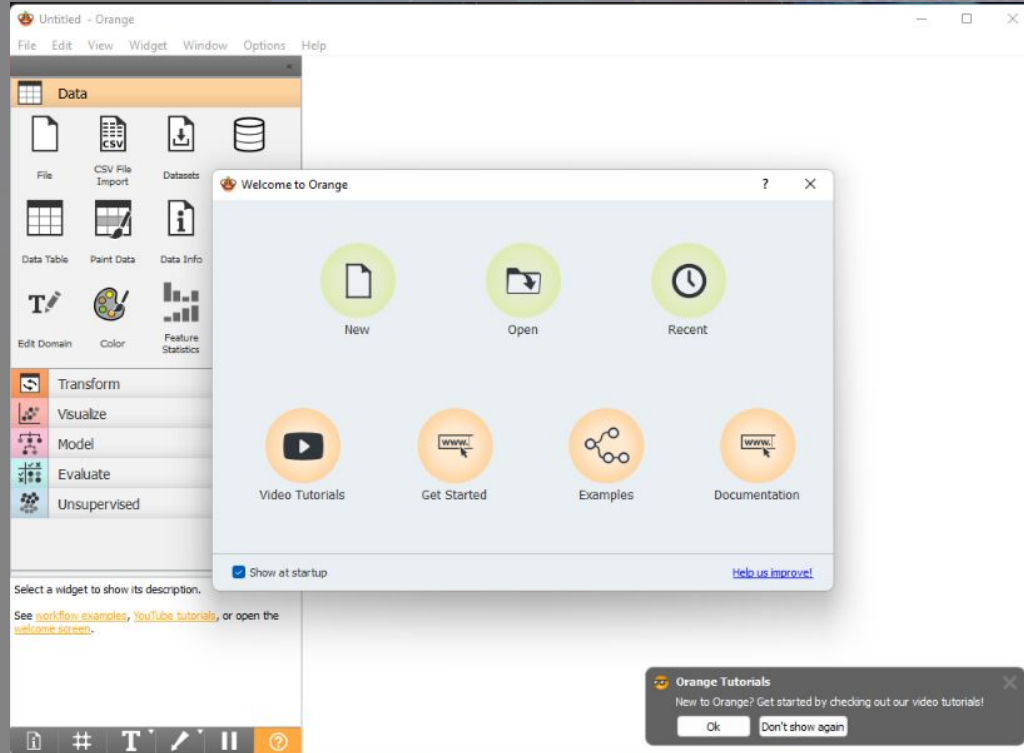
# Introdução - Orange (Download e Instalação)



# Introdução - Orange (Download e Instalação)



# Introdução - Orange





# Introdução - Orange

**Hands ON!**