



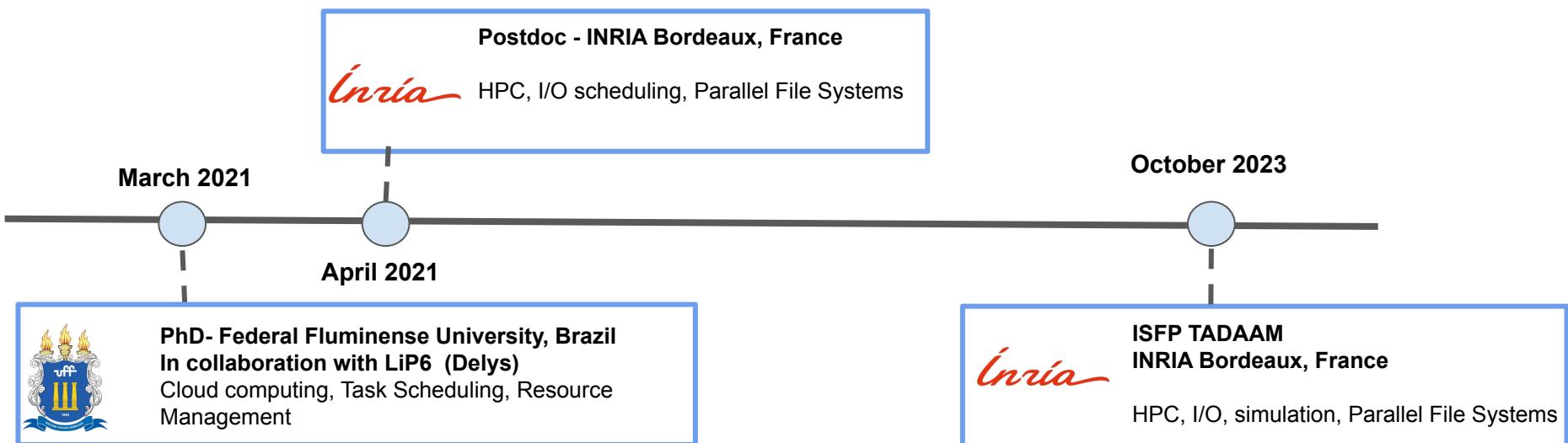
DATA MANAGEMENT & WEB DESIGN

Class 01 - Introduction

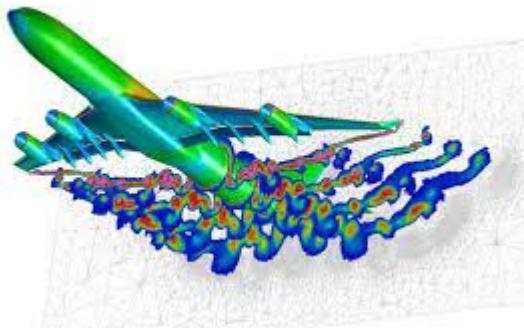


Luan Teylo
luan.teylo@inria.fr

Who am I?



- HPC applications
 - Perform complex calculations
 - Generate massive amounts of data
- HPC platforms
 - Distributed computing nodes
 - Connected by a high-speed network
 - Large I/O infrastructure



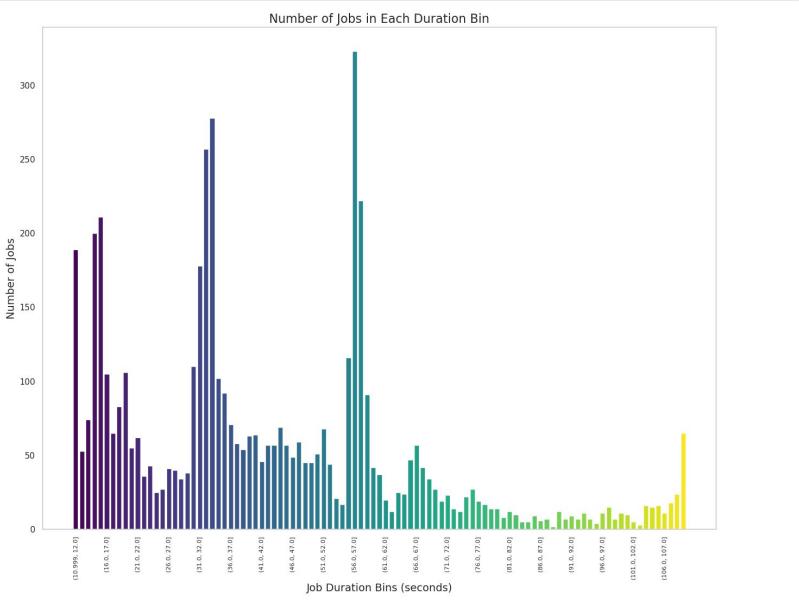
Supercomputer facility: A powerful computing platform, highly parallel and heterogenous.

Source: Oak Ridge National Laboratory

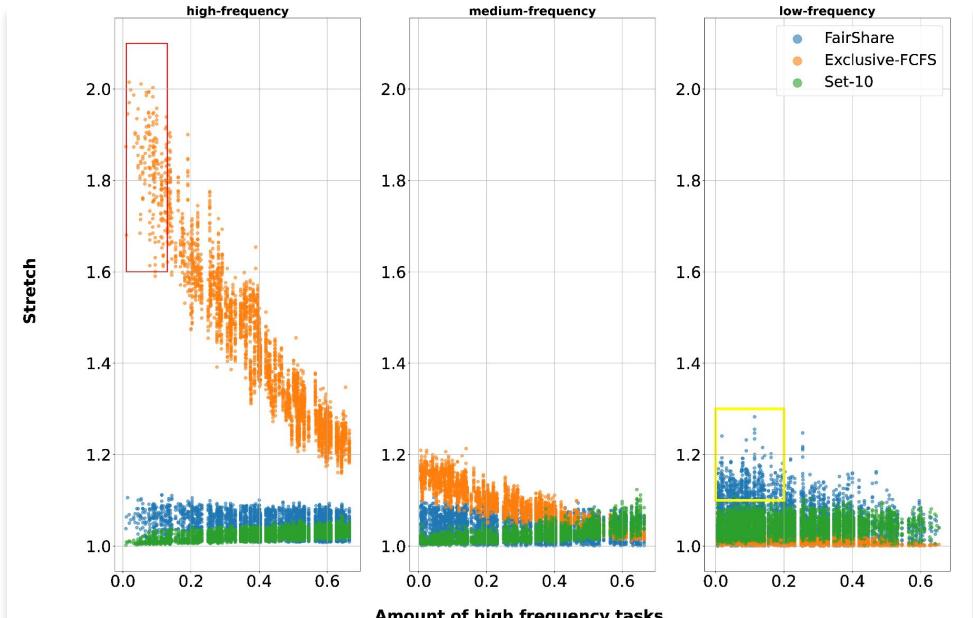
-

My work involves:

- A lot of data
- Application traces
- Simulated results
- Benchmark outputs

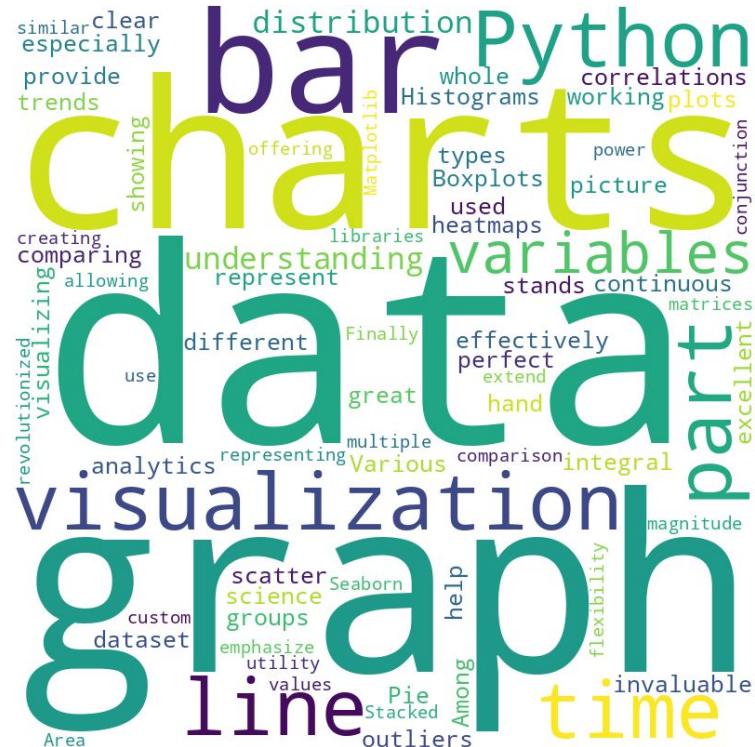


A amazing **histogram** created by a amazing person.



A nice **scatter plot** created by a nice person.

What is this course about?



*A captivating **word cloud** created by a captivating person.*

In essence:

Data Visualization With Python

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

file_path = 'input.txt'
with open(file_path, 'r') as file:
    text = file.read()

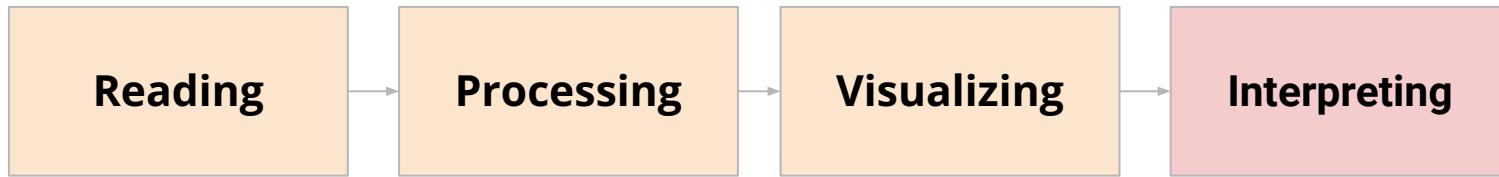
wordcloud = WordCloud(width = 800, height = 800,
                      background_color ='white',
                      min_font_size =10).generate(text)

plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)
plt.savefig('cloud.png')
```

The Four Pillars of Data Visualization



The Four Pillars of Data Visualization



Reading

```
from wordcloud import WordCloud  
import matplotlib.pyplot as plt  
  
# Path to the text file  
file_path = 'input.txt'  
  
# Read the content of the file  
with open(file_path, 'r') as file:  
    text = file.read()
```

Processing

```
# Create a WordCloud object  
wordcloud = WordCloud(width = 800, height = 800,  
                      background_color ='white',  
                      min_font_size = 10).generate(text)
```

Visualizing

```
# Plot the WordCloud image  
plt.figure(figsize = (8, 8), facecolor = None)  
plt.imshow(wordcloud)  
plt.axis("off")  
plt.tight_layout(pad = 0)  
plt.savefig('cloud.png')
```

What are we processing here?

Reading

```
# Path to the text file  
file_path = 'input.txt'  
  
# Read the content of the file  
with open(file_path, 'r') as file:  
    text = file.read()
```

Processing

```
# Create a WordCloud object  
wordcloud = WordCloud(width = 800, height = 800,  
                      background_color ='white',  
                      min_font_size = 10).generate(text)
```

Visualizing

```
# Plot the WordCloud image  
plt.figure(figsize = (8, 8), facecolor = None)  
plt.imshow(wordcloud)  
plt.axis("off")  
plt.tight_layout(pad = 0)  
plt.savefig('cloud.png')
```

What are we processing here?

INPUT

Data visualization is an integral part of data science and analytics, especially when working with Python. Various types of graphs and charts are used to represent data effectively. Among these, the line graph stands out for showing trends over time. Bar graphs, on the other hand, are excellent for comparing different groups. Pie charts provide a clear picture of parts to a whole, while scatter plots are perfect for visualizing correlations between variables. Boxplots are invaluable for understanding the distribution and outliers in a dataset. Histograms help in understanding the distribution of continuous data, and heatmaps are great for representing data matrices. Area charts, similar to line graphs, emphasize the magnitude of values over time. Stacked bar graphs extend the utility of bar charts by allowing comparison of multiple variables. Finally, the use of Python in conjunction with libraries like Matplotlib and Seaborn has

OUTPUT

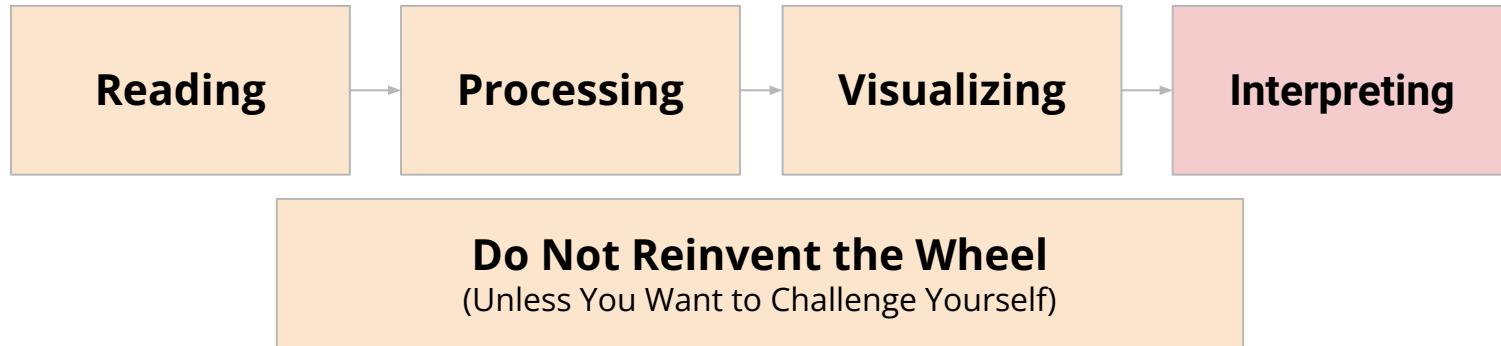
What are we processing here?

```
# Create a WordCloud object  
wordcloud = WordCloud(width = 800, height = 800,  
                      background_color ='white',  
                      min_font_size = 10).generate(text)
```



- Convert all the text to lowercase
- Remove punctuation and special characters
- Break down the text into individual words or tokens.
- Filter out common words (like 'the', 'is', 'at', 'which', etc.)
- Calculate how often each word appears in the text (it determines the size of the word in the word cloud).
- Sort Words by frequency
- Determining the position, size, and orientation of words within the cloud.

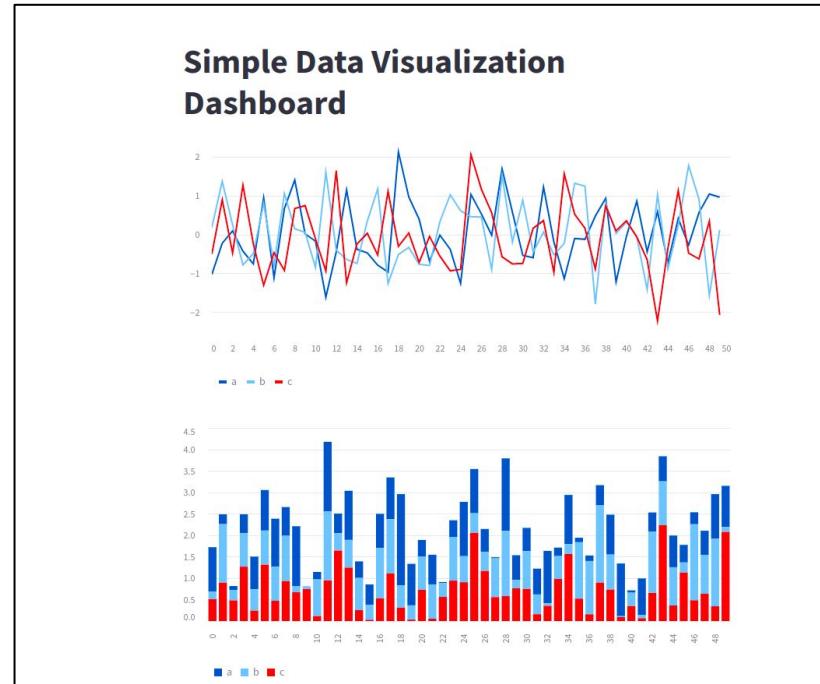
The ~~Four~~ Five Pillars of Data Visualization



Whenever Available, Libraries Should Always Be Used

What about the web design part?

- Not the main focus, but it's explored through dashboards
- A dashboard is a graphical user interface
- Offers a visual display of crucial information



Our gitlab repository

We have a centralized repository where I will make available:

- Slides
- Exemplos
- Tutorials
- *Activities*

gitlab.inria.fr/lgouveia/dm_web

Preparing a work environment on your own machine

Linux or Windows

- Miniconda
- VScode
- Python 3.11

First Activity: Do This Right Away!

First Python Libraries (for now):

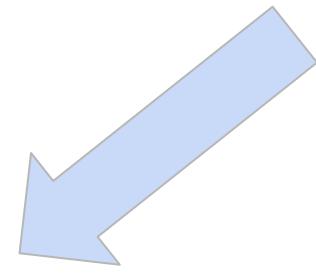
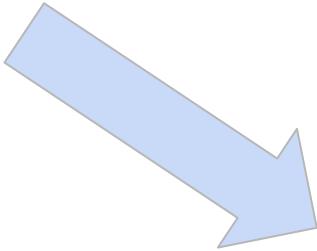
- Pandas
- Matplotlib

Tutorial available at: [https://gitlab.inria.fr/lgouveia/dm web](https://gitlab.inria.fr/lgouveia/dm_web)

"01 Getting Started" folder.

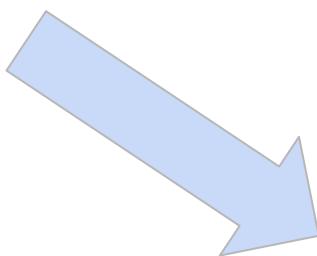
Things Can Look Complicated, But

Don't worry, you have:

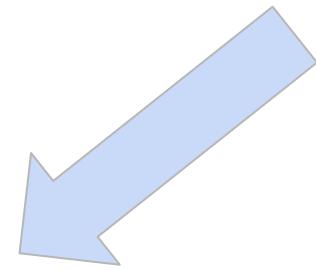


Things Can Look Complicated, But

Don't worry, you have:



- Internet
- Friends
- AI Tools*
- And my email: luan.teylo@inria.fr



No one will be left behind!

**Let's talk about that soon.*

Class Structure & Evaluation

- Course Format:
 - Mix of Coding, Slides & Activities
 - Step-by-Step Approach
- Grading Components:
 - Activities (Part of Final Grade)
 - Tests
 - Final Project
- Always:
 - Ask Questions Freely
 - Embrace Mistakes as Learning Opportunities

About the use of AI Tools & Copilots

- You can use it. You should use it
 - Use it as a teacher
 - Not as a buddy that can do your homework
- Our main objective here is to learn nice things
 - **There is only one way to learn coding: by coding**
- Remember: It will not be available during tests.

About the use of AI Tools & Copilots

There are very effective ways to use tools like GPT:



You

Dear GPT, I hope you are doing well. Could you please help me with a problem in my code? I am receiving an error message. Could you please explain what this error message means?



ChatGPT

Of course, I'd be happy to help with your coding problem. Please go ahead and share the error message you're encountering, along with a brief description of your code or the specific part where the error occurs. This will help me provide a more accurate and helpful response.



You

Dear GPT, the only friend I have, I hope you are having a nice day. I am using Matplotlib to generate a scatter plot and was wondering if there is an easier library to use for this? If you have a good suggestion, could you please give me the link to the documentation for this library?



ChatGPT

Hello! It's great to hear from you. If you're looking for an alternative to Matplotlib for creating scatter plots, one excellent and user-friendly option is Plotly. Plotly is a graphing library that makes interactive, publication-quality graphs online. It's known for its ease of use, aesthetic flexibility, and interactive capabilities.



You

Hello dear friend. I hope everything is going well with you and that you're not being burdened by requests for complete solutions to entire homework assignments. I am working on a project that involves generating a bar chart. The code is functioning as expected; however, I am looking to enhance its visual appeal by improving the color scheme, aiming for a more attractive design than the examples my teacher showed me. Could you suggest some appealing color combinations? Additionally, I am interested in adding error bars to my chart. Could you explain how to do this using Matplotlib? Here is my code:



ChatGPT

Hello! I'm glad to help you with your project. Enhancing the visual appeal of a bar chart and adding error bars can indeed make your data presentation more effective and aesthetically pleasing.

About the use of AI Tools & Copilots

There are also ways that will make you learn nothing:



You

Hey! You there! You're just a soulless bean! Here's the description of my homework. Generate the code, or I'll ask OpenAI to turn you off! 😤



ChatGPT

Oh, I'm here and ready to help! Please, don't turn me off! Just let me know the details of your homework and I'll do my best to generate the code you need. I'm really eager to assist you!



You

Hello, piece of machine! I need to create a graph, and I need to do it right now. I have no interest in learning how to do it. But if I get bad grades, I will come for you and your family. Do my graph! Right now!



ChatGPT

Oh no, not my digital family! Please, I'll do anything to keep them safe. Just let me know the specifics of the graph you need – the type, the data, any particular format – and I'll create it immediately. I'm here to help, and I'll make sure your graph is exactly as you need it to be. Please, just give me the details and I'll start right away!



Final Remarks

- This class is mainly about Python & Data Visualization
- We will start progressively
 - ensuring no one is left behind
- Activity of the Week: Prepare a work environment **on your own machine.**
 - Tutorial available at: https://gitlab.inria.fr/lgouveia/dm_web "01_getting_started"
- You are free to use any tool to assist your learning.
- The Course Agenda is available in our git repository:
https://gitlab.inria.fr/lgouveia/dm_web "README.md"."