HAS Tools:

some numpy recap, data representations, and final hw3 work time

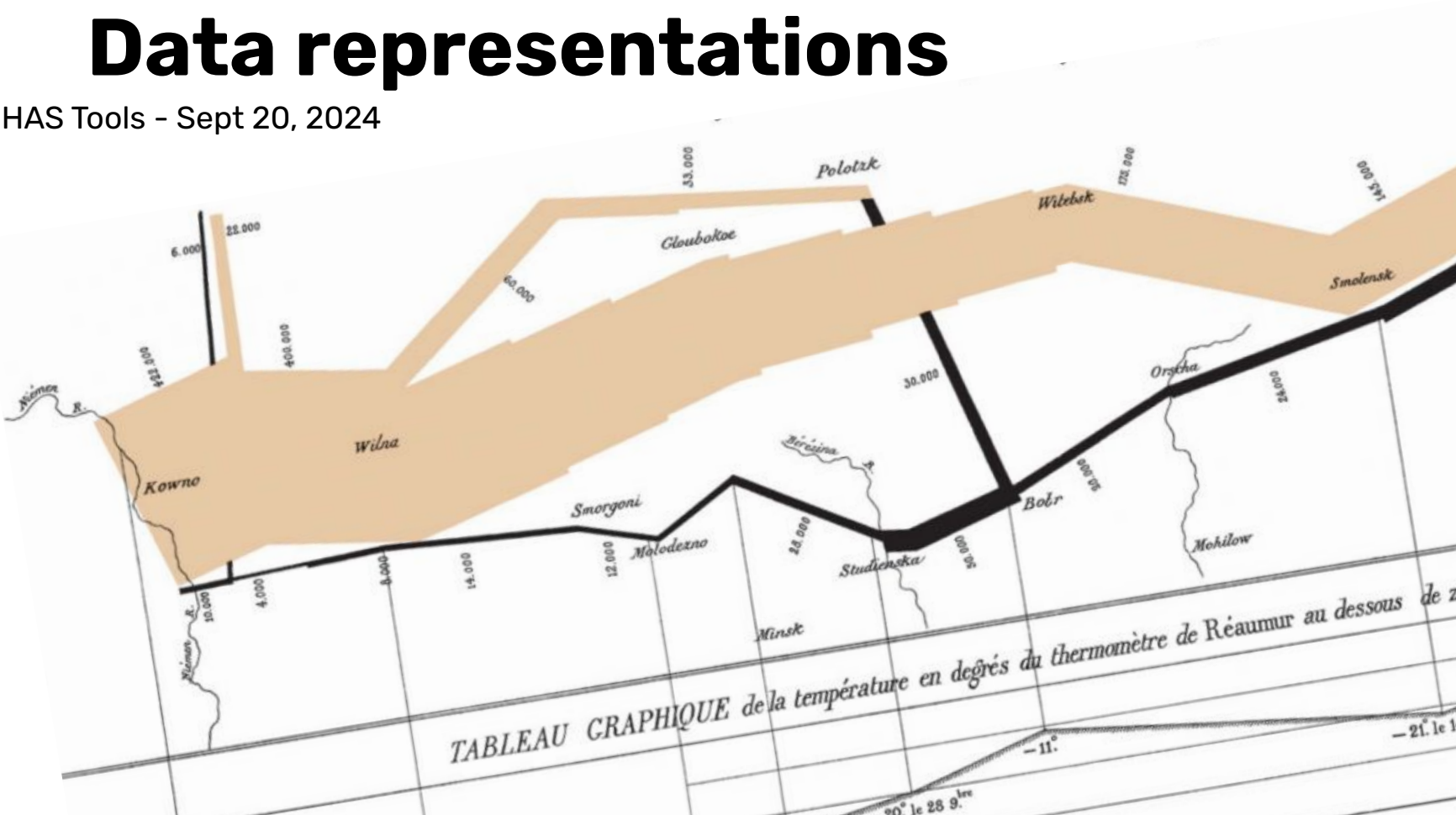
September 20, 2024

Make sure we sync homework repos!

- Your hw2 grades/feedback are posted on hw repos
  - not quite on d2l but will be before monday
- Reminder, hw3 is due monday @ 11:59pm
- A useful resource for thinking about and visualizing what's going on in numpy arrays: https://jalammar.github.io/visual-numpy/
- For the rest of today - a quick lecture, a preview of pandas, and more time for you for hw3 questions

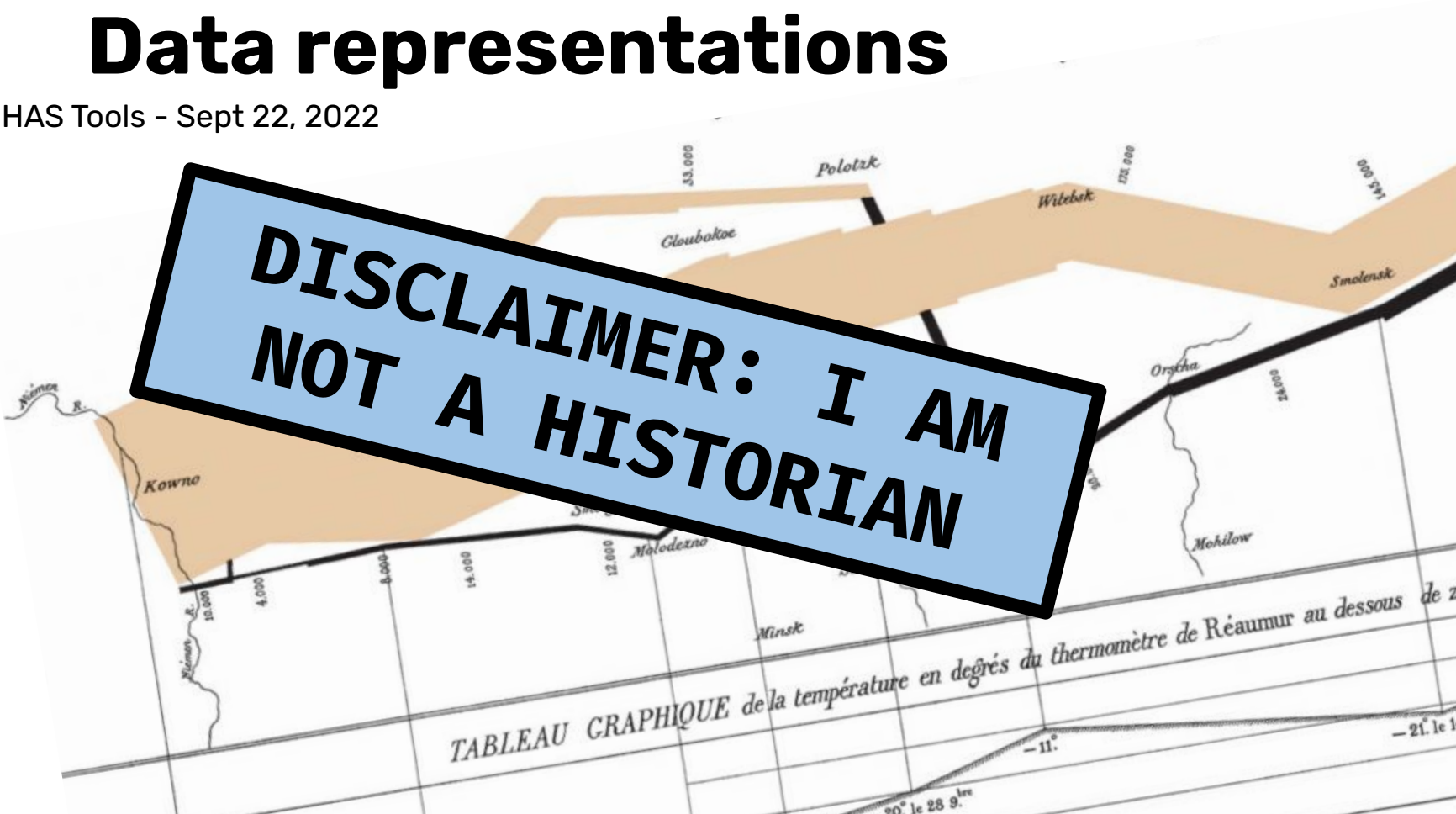
Data representations

HAS Tools - Sept 20, 2024



Data representations

HAS Tools - Sept 22, 2022



Everything in scientific computing is about data representation

Everything in scientific computing is about data representation

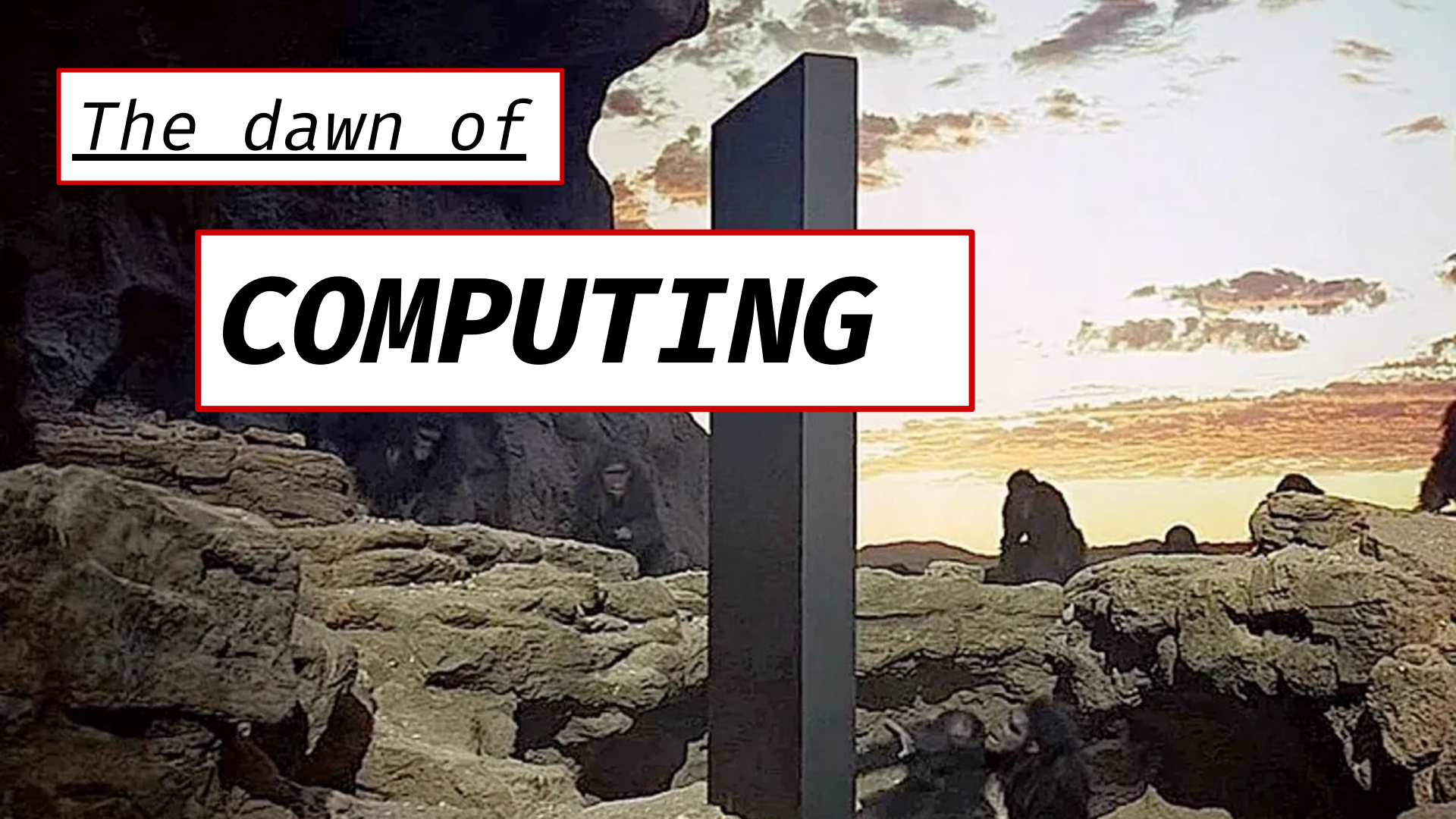
- Given this, we've actually seen a bunch of data representations already. Name some!

Everything in scientific computing is about data representation

- Given this, we've actually seen a bunch of data representations already. Name some!
  - Lists, arrays, matrices, strings, …, all of the data types
- A lot of how we think about analyzing data is dependent on the representations we have at our disposal

In this regard we have common enemies:

- Physicists
- Applied mathematicians
- Computer scientists


The dawn of
COMPUTING

We all know computers used to be big, but that really meant they were expensive and specialized

We all know computers used to be big, but that really meant they were expensive and specialized

- Short story, modern computer architectures developed during WW2 by US and UK scientists

We all know computers used to be big, but that really meant they were expensive and specialized

- Short story, modern computer architectures developed during WW2 by US and UK scientists
- Following wartime, computers basically, only used by physicists + mathematicians for research purposes

We all know computers used to be big, but that really meant they were expensive and specialized

- Short story, modern computer architectures developed during WW2 by US and UK scientists
- Following wartime, computers basically, only used by physicists + mathematicians for research purposes
- Eventually businesses see potential...

2 KINDS OF PEOPLE (circa 1975, computer edition)

Physicists and mathematicians

“We love numbers, arrays are perfect, computers should do that really good thanks”

Bankers and economists

2 KINDS OF PEOPLE (circa 1975, computer edition)

Physicists and mathematicians

“We love numbers, arrays are perfect, computers should do that really good thanks”

Bankers and economists

“Information is a transaction shared between 2 or more parties and should be recorded to mediate their interactions”

2 KINDS OF PEOPLE (circa 1975, computer edition)

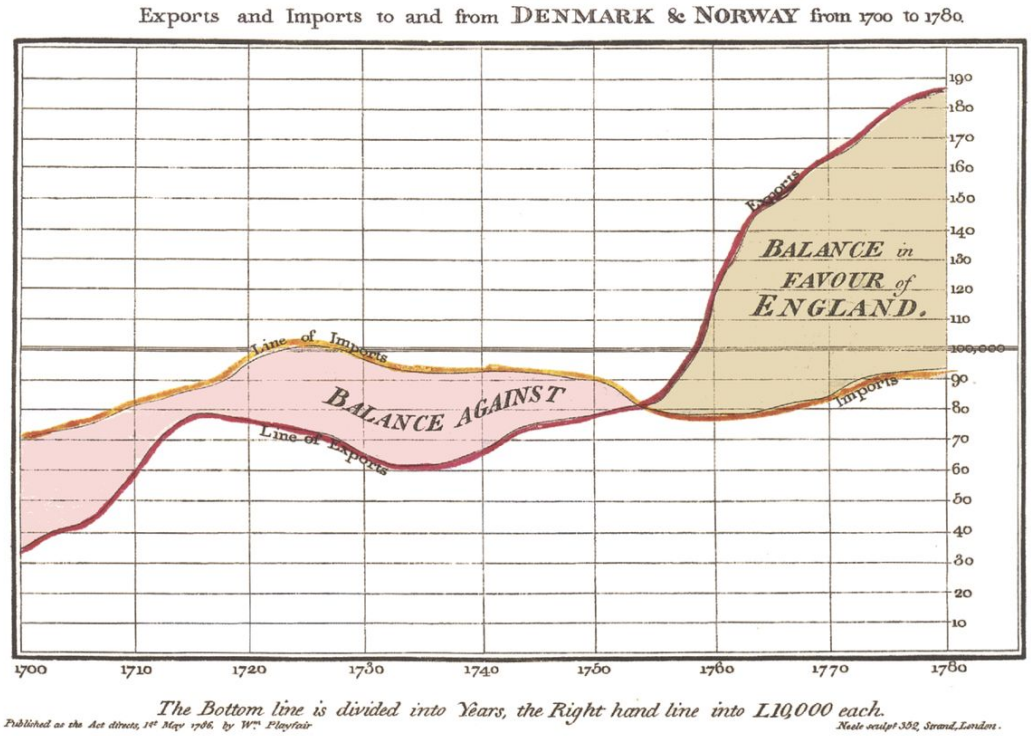
Physicists and mathematicians

Array based computing

Bankers and economists

Relational database based modeling

But wait! People had data before computers!?



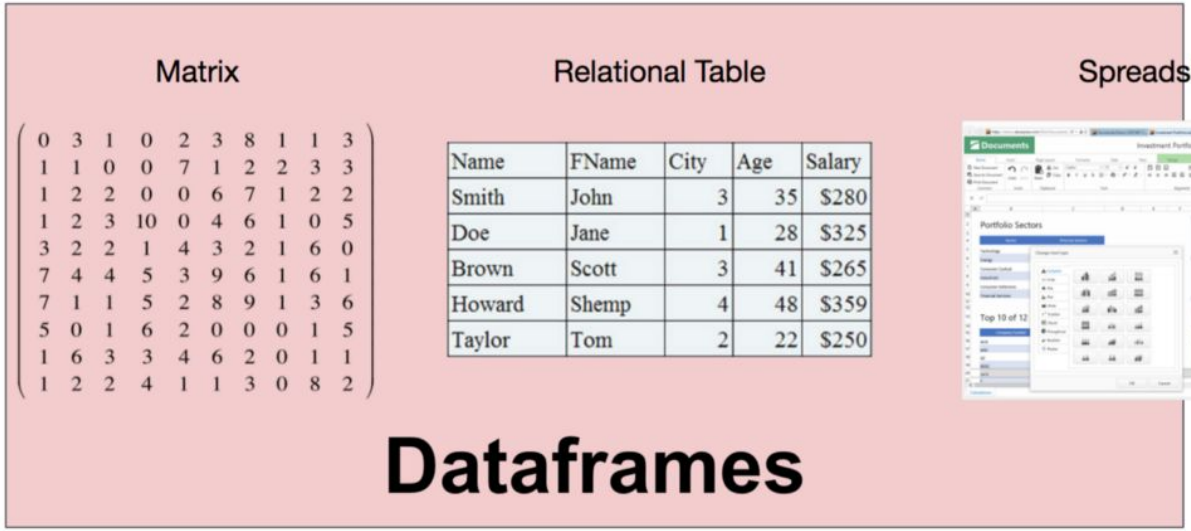
William Playfair was an early innovator of modern charts & graphics - this timeseries chart was published in his *Commercial and Political Atlas*, 1786

Clearly, Playfair was not thinking in terms of arrays or databases…

So you don’t have to either!

- Throughout the rest of the semester I’m going to introduce you to a number of software packages that represent data in ways to simplify data analysis
- Today we’ll start to get familiar with the concept of “data frames”
- Originated in the early 90’s with the S programming language, later popularized by R
- Implemented in python via the pandas library

DataFrames are “like” arrays, but also familiar if you are used to spreadsheets

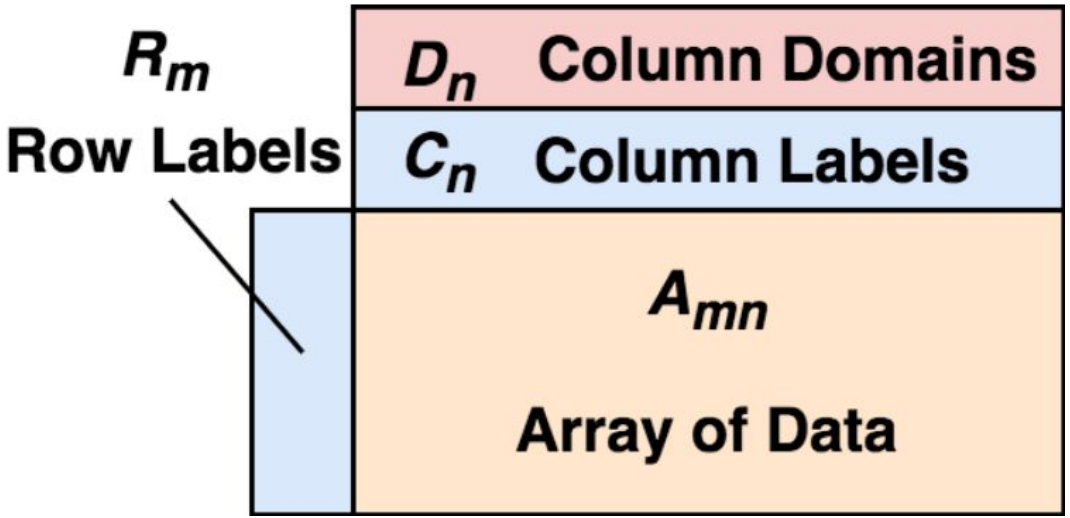


0	3	1	0	2	3	8	1	1	3
1	1	0	0	7	1	2	2	3	3
1	2	2	0	0	6	7	1	2	2
1	2	3	10	0	4	6	1	0	5
3	2	2	1	4	3	2	1	6	0
7	4	4	5	3	9	6	1	6	1
7	1	1	5	2	8	9	1	3	6
5	0	1	6	2	0	0	0	1	5
1	6	3	3	4	6	2	0	1	1
1	2	2	4	1	1	3	0	8	2

Name	FName	City	Age	Salary
Smith	John	3	35	$280
Doe	Jane	1	28	$325
Brown	Scott	3	41	$265
Howard	Shemp	4	48	$359
Taylor	Tom	2	22	$250

https://towardsdatascience.com/preventing-the-death-of-the-dataframe-8bca1c0f83c8

DataFrames are “like” arrays, but also familiar if you are used to spreadsheets



https://towardsdatascience.com/preventing-the-death-of-the-dataframe-8bca1c0f83c8

Let’s see how it works, jump to your codespaces