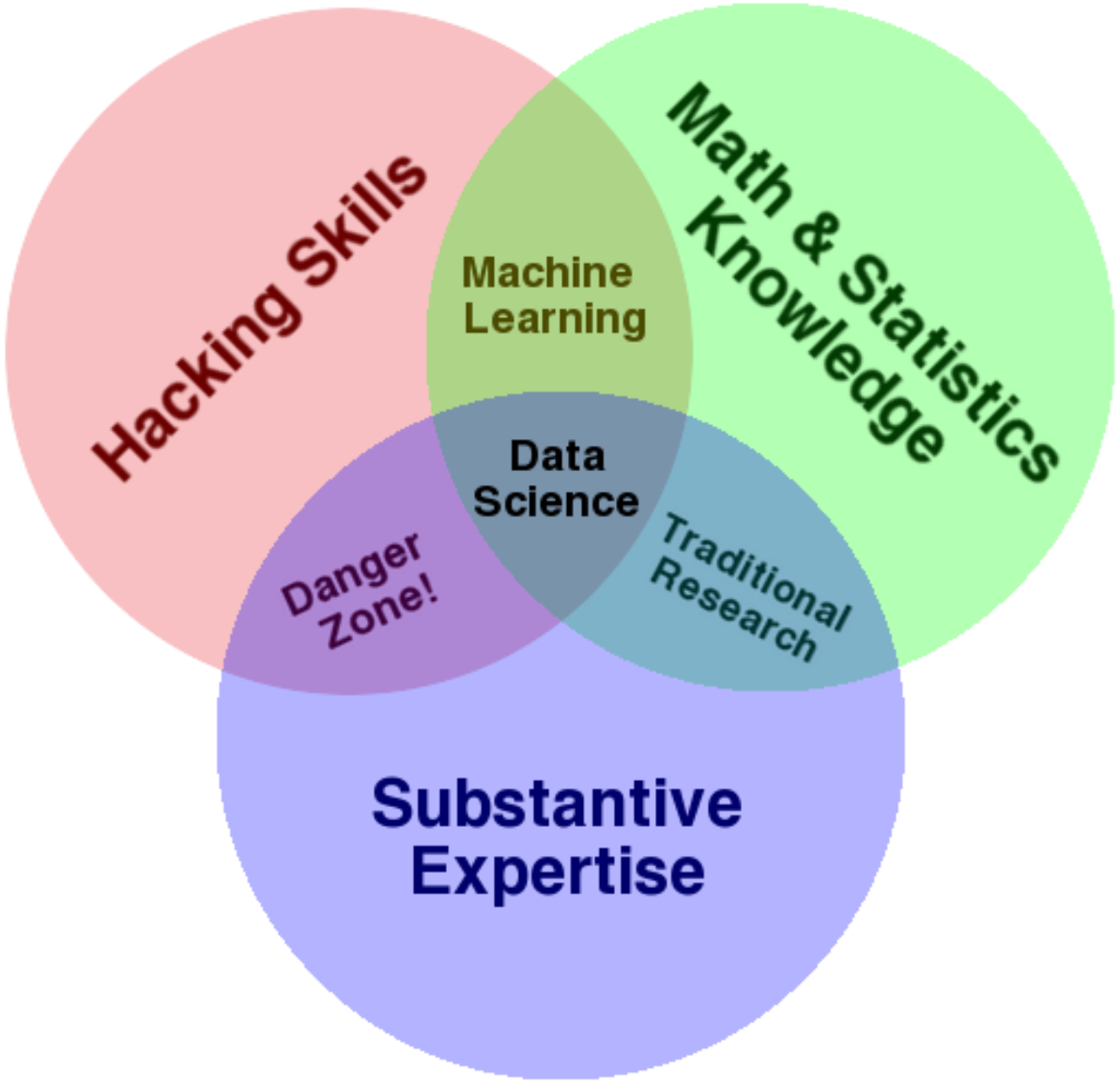
Module 1: Introduction to Informatics

CSCI 1360

Objectives

- Define "Data Science"
- Understand data importance and growth
- Give examples of data science applications
- Know common data types and data representation methods
- Understand data visualization features, types and examples
- Understand data modeling

Data Science

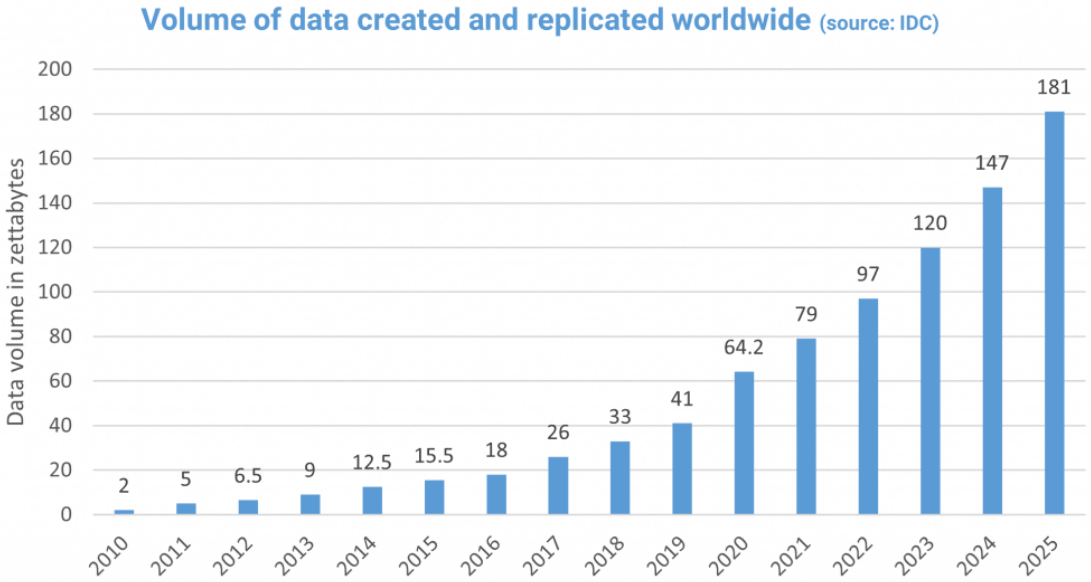


Definition

- Data Science as a proper field of study is the confluence of three major aspects:
  - Hacking skills: the ability to code, and knowledge of available tools.
  - Math and statistics: strong quantitative skills that can be implemented in code.
  - Substantive expertise: some specialized area of emphasis.

Data Importance

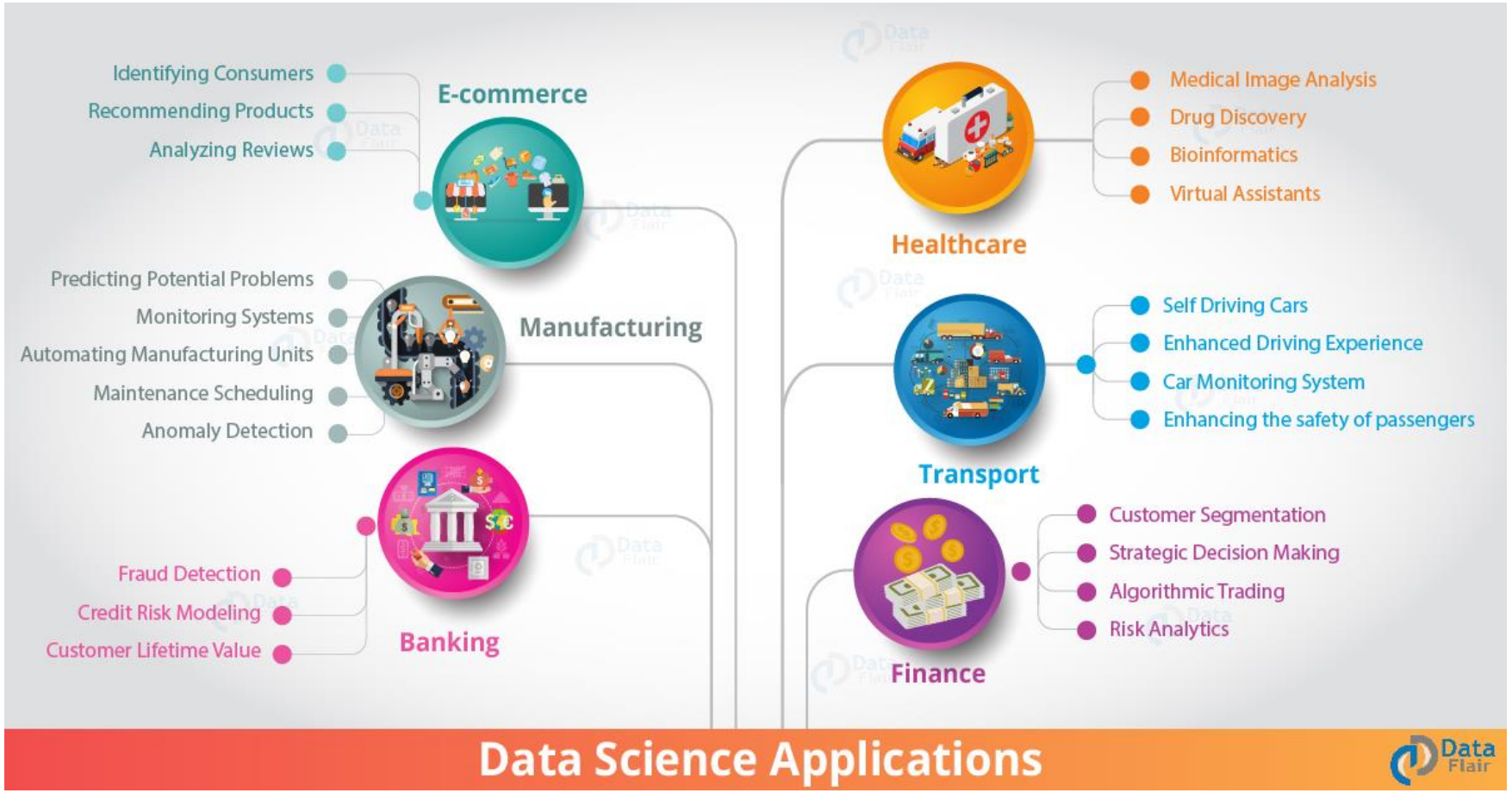
- Data Growth
- Zettabyte = 1 Billion Terabyte



Reasons Behind Data Growth

- Storage Technologies and Prices
  - Cloud and On premise
- Computation Technologies and Prices
  - CPUs and GPUs capabilities
- Data Collection Technologies
  - Sensors and Cameras

Data Science Applications



https://data-flair.training/blogs/data-science-applications/

Data Types

Some forms of data that data scientists handle

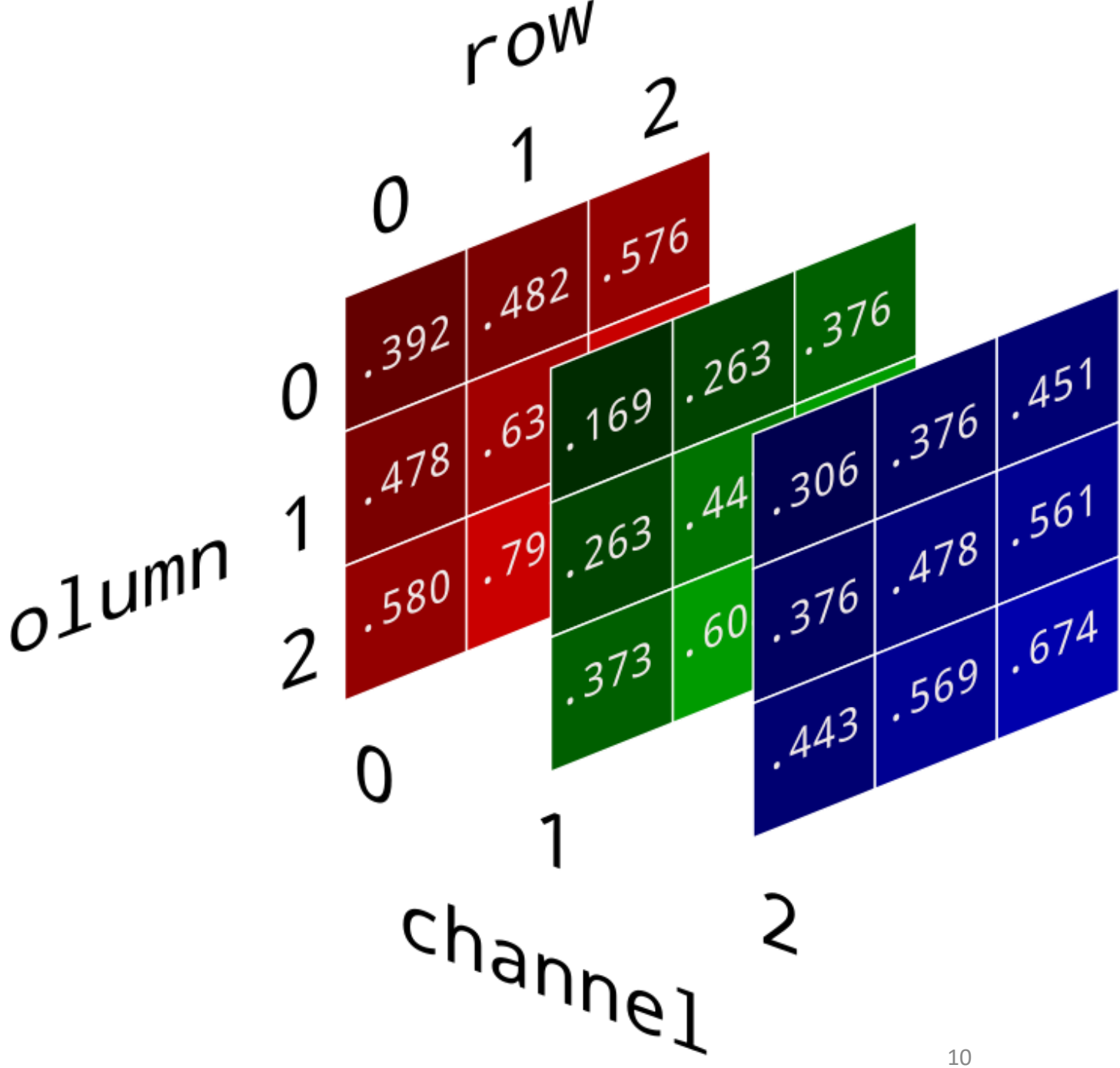
- Numbers
  - Bank transactions
  - Weather sensor readings
- Text
  - Movie reviews
  - News articles
- Image
  - X-Ray images
  - Road signs

Data Representation

- Computer can process data as numbers
- Image representation
- Text representation

Image Encoding

- R, G, B
- (256, 256, 256)
- 16.7 million colors



Text Representation – Bag of Words

```
"John","likes","to","watch","movies","Mary","likes","movies","too"

"Mary","also","likes","to","watch","football","games"
```

Representing each bag-of-words as a JSON object, and attributing to the respective JavaScript variable:

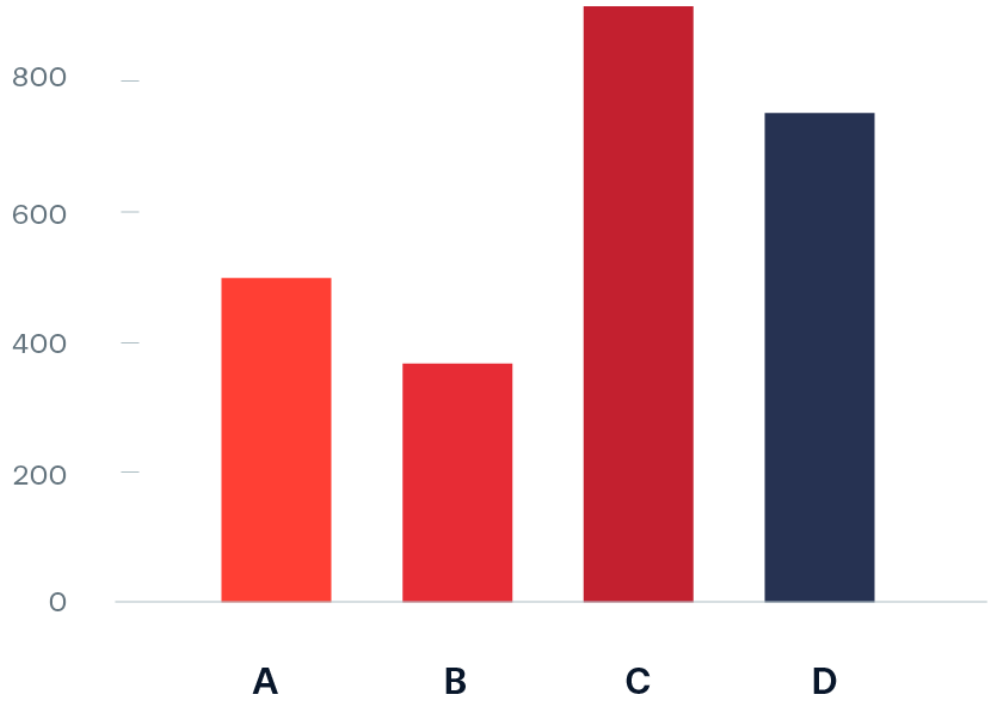
```
BoW1 = {"John":1,"likes":2,"to":1,"watch":1,"movies":2,"Mary":1,"too":1};
BoW2 = {"Mary":1,"also":1,"likes":1,"to":1,"watch":1,"football":1,"games":1};
```

https://en.wikipedia.org/wiki/Bag-of-words_model

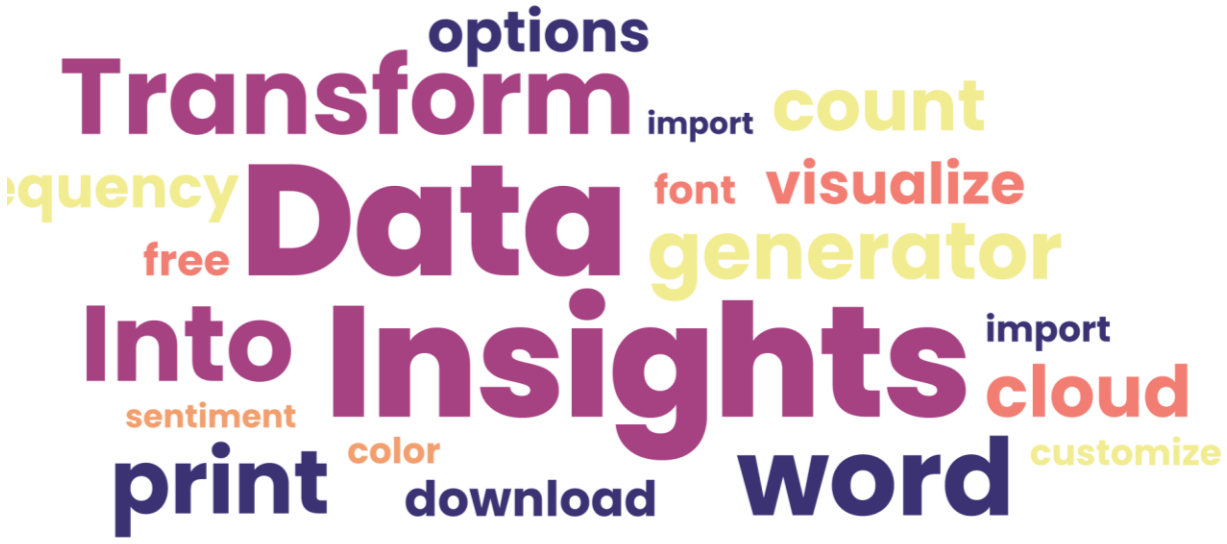
Data Visualization

- Why visualization is important
  - To understand data
  - To share data easily
  - To visualize patterns

Some Types of Visualizations



Some Types of Visualizations



Data Modeling

- Discriminative modeling
  - Learn the boundaries between data samples
- Generative modeling (beyond the scope of the course)