

# 1. Fundamentals

Tags	Done
Módulo	<a href="https://www.coursera.org/learn/data-scientists-tools/home/module/1">https://www.coursera.org/learn/data-scientists-tools/home/module/1</a>

## Objetivos de aprendizagem

- Define data science
- Define data
- Use resources to get help with problems

## Videos automatizados

Por que os vídeos automatizados iguais aos textos do curso?

Por que atualizar aulas gravadas é muito trabalhoso ˘\_(`´)\_/˘

já que o volume de cursos e atualizações (correções, por exemplo) é alto, o uso de textos automatizados ajuda a lidar com os updates necessários

Os videos são então criados com Amazon Polly, utilizando dois pacotes do R que foram criados para isso (se eu entendi direito):Ari

e Detector.

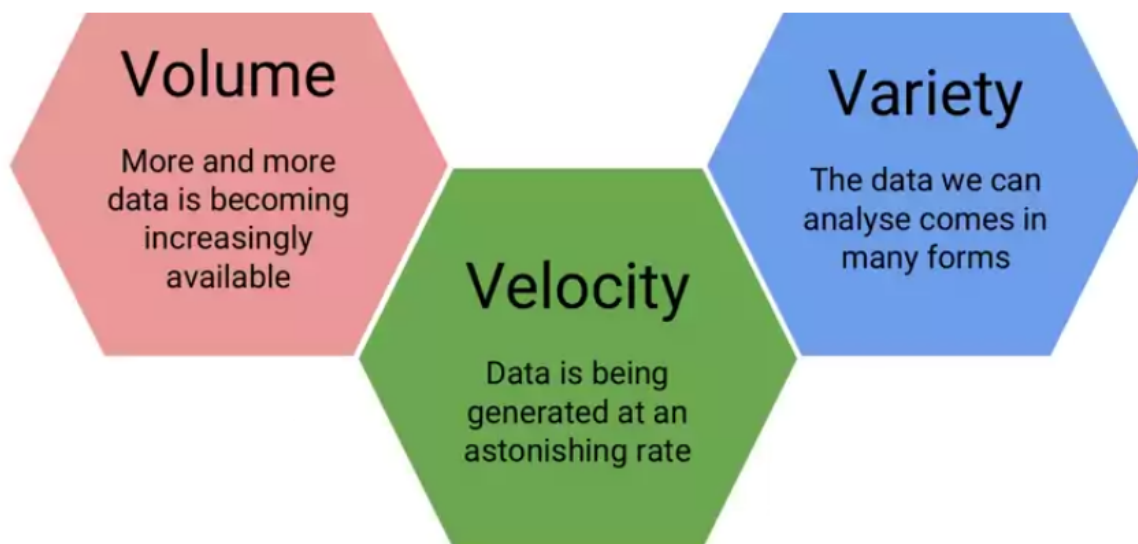
Ao final do curso um dos objetivos é criar coisas legais como o sistema de automatização de criação de vídeos a partir do R markdown, google slides e Amazon Polly (no caso desse exemplo no caso).

---

## o que é ciência de dados?

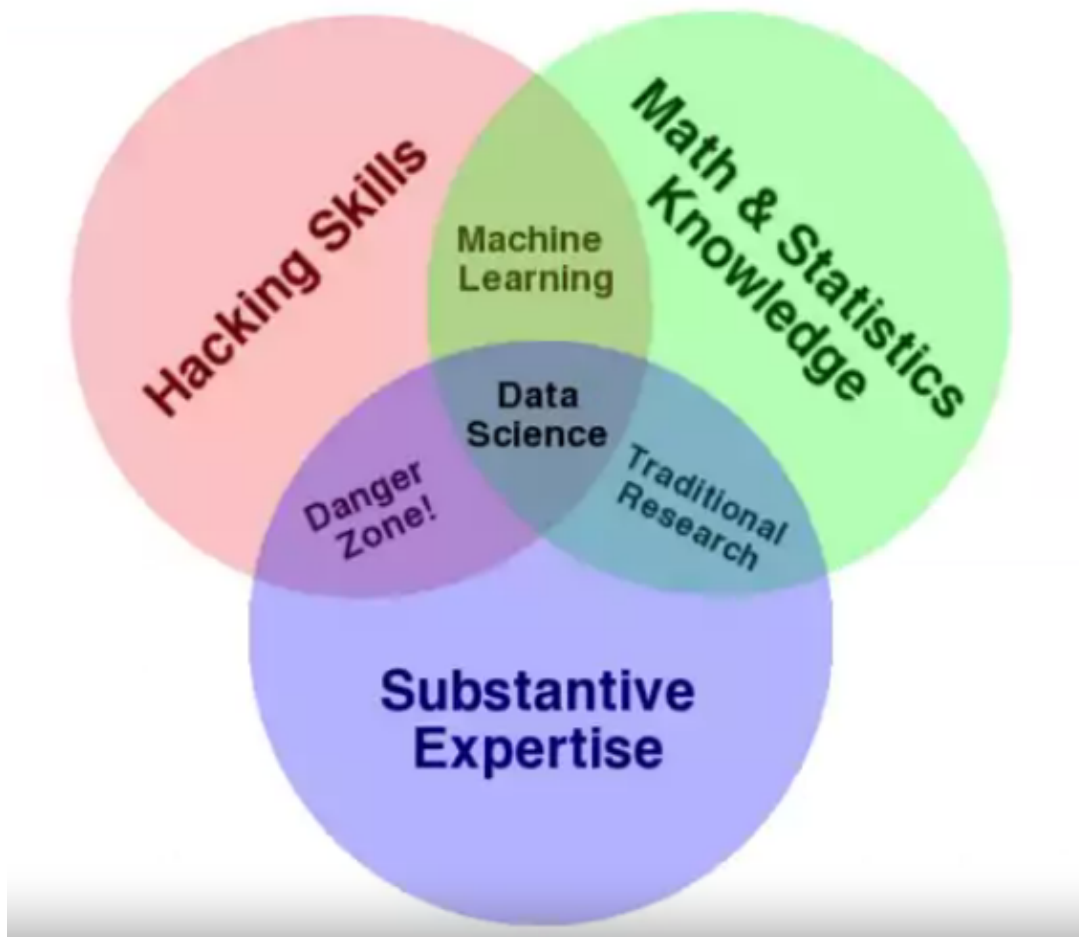
Aparentemente data science é qualquer coisa que usa dados?????

ou melhor, são as ferramentas que usamos para extrair a pepita de ouro de uma montanha de dados...



também tem a questão que o volumes de dados aumenta exponencialmente...

Ta, mas e aí, o que é um cientista de dados??



Nesse curso o foco vai ser o conhecimento de matemática e estatística e as habilidades de programação, deixando a expertise nas mãos do aluno.

...e da emprego 🤔

fluxograma do conteúdo do curso:

```
R > Rstudio > Version Control > Applying the tools to answer questions
```

JÁ MANJAMO DOIS

## O que são dados?

Data possui algumas definições, algumas focam em um conjunto de valores, outras acoplam o fato de que dados são coletados, analisados e levam à conclusões

## Definitions of “data”

First, let's look at what a few trusted sources consider data to be.

First up, we'll look at the [Cambridge English Dictionary](#), which states that data is:

Information, especially facts or numbers, collected to be examined and considered and used to help decision-making.

Second, we'll look at the definition provided by [Wikipedia](#), which is:

A set of values of qualitative or quantitative variables.

There are clearly different definitions and there are different components of what data is. Both agree that data is values – numbers or

“A set of values of qualitative or quantitative variables”

**Set:** In statistics, the population you are trying to discover something about

**Variable:** Measurements or characteristics of an item

**Qualitative variable:** Measurements or information about qualities

**Quantitative variable:** Measurements or information about quantities or numerical items

Nos exemplos, os conjuntos de dados são coisas simples... já na realidade é geralmente uma bagunça 🧑



Importante!!

O ideal é definir qual é a pergunta antes, e então definir o método de coleta de dados...

Infelizmente, os dados disponíveis geralmente irão limitar as perguntas feitas.

Resolver problemas envolve *pedir ajuda!*

Eric Raymond - How to Ask questions the smart way

Roger Peng - Getting Help

[https://www.youtube.com/watch?v=ZFawxxzouCY&ab\\_channel=RogerPeng](https://www.youtube.com/watch?v=ZFawxxzouCY&ab_channel=RogerPeng)

Essa etapa do curso é focada em ensinar técnicas de resolução de problemas quando não há uma figura de educador para te ajudar.



Read the manual > go to google read forums

códigos geralmente apresentam dois tipos de erro:

- Your command produces no data and spits out an error message
  - mais fácil de lidar, checar por erros no código ou então analisar a mensagem de erro para identifica-lo e resolve-lo
- Your command produces no data and spits out an error message

- Neste caso o código funciona perfeitamente, mas provavelmente você não soube pedir direito o que queria
- 



Basicamente, explique seu código para o pato 🦆

finalmente, quando precisar perguntar no forum, siga algumas recomendações:

## Asking questions on forums - details to include:

- The question you are trying to answer
- How you approached the problem, what steps you took to answer the question
- What steps will reproduce the problem (including sample data for troubleshooters to work from!)
- What was the expected output
- What you saw instead (including any error messages you received!)
- What troubleshooting steps you have already tried
- Details about your set-up, eg: what operating system you are using, what version of the product you have installed (eg: R, Rpackages)

Enfim, quando finalmente precisar perguntar em fóruns: seja específico, seja educado, quando resolver o problema dê a devolutiva, e só pergunte quando for genuinamente uma dúvida, e não use os fóruns uma forma dos outros fazerem seu trabalho

---

## 0 Processo de ciência de dados

O processo da ciência de dados inicia com uma pergunta, é importante definir bem a pergunta, como por exemplo:

*como o fluxo de diatomáceas influencia no fluxo vertical de carbono no ambiente marinho???11*


com a pergunta estabelecida os dados são obtidos, ou coletados, e então o primeiro passo é explorar os dados, e geralmente modelar o modelo. Finalmente, o projeto deve ser comunicado para outros!

Projetos de ciência de dados quase sempre envolve a comunicação dos resultados obtidos! PÃ

O módulo segue o projeto:

## Hilary: the most poisoned baby name in US history

I've always had a special fondness for my name, which – according to Ryan Gosling in “Lars and the Real Girl” – is a scientific fact for most people (Ryan Gosling

 <https://hilaryparker.com/2013/01/30/hilary-the-most-poisoned-baby-name-in-us-history/>

