



# MONITORIA

# TADI

Tratamento e Análise de  
Dados e Informações



**EACH**



Escola de Artes, Ciências e Humanidades  
Universidade de São Paulo



# **IMPORTÂNCIA DE SABER MANIPULAR DADOS**

# IMPORTÂNCIA DE SABER MANIPULAR DADOS

- Dados - informação - conhecimento
- Através da manipulação de dados é possível conhecer o perfil e comportamento de indivíduos, o que pode ser fundamental para campanhas de marketing, por exemplo.
- Avaliação de políticas públicas, estudo de conjuntos de dados de biologia, etc...

# IMPORTÂNCIA DE SABER MANIPULAR DADOS

- Saber manipular dados auxilia no processo de tomada de decisão que é muito importante no meio empresarial.
- No meio acadêmico, os dados conferem veracidade a um dado estudo, é necessário realizar a manipulação desses de maneira correta para não chegar a conclusão erradas.

# IMPORTÂNCIA DE SABER MANIPULAR DADOS

- No mundo real, encontramos diversas dificuldades para alcançar nosso objetivo, conjuntos de dados com dados faltantes, ou muitas variáveis, de difícil compreensão, etc...
- O conhecimento de conceitos fundamentais de Estatística, estudados na matéria de TADI, é essencial para entendimento do mundo dos dados.

# DIFERENÇA ENTRE FERRAMENTAS



Planilhas



R



Python

# PLANILHAS

Exemplos: Excel e Google Sheets



## Vantagens

- Bastante abrangente (realiza diversas tarefas)
- Intuitivo e fácil de se realizar tarefas simples
- Utilizado em diversas áreas para atender diversos propósitos

**Bom para poucas tabelas e tarefas simples**

## Desvantagens

- Limitado quando precisamos trabalhar com dados em grande escala
  - Limitação de linhas e formatos de arquivos
  - Difícil visualização e manipulação das tabelas
- Difícil, ou impossível, implementação de alguns algoritmos mais complexos

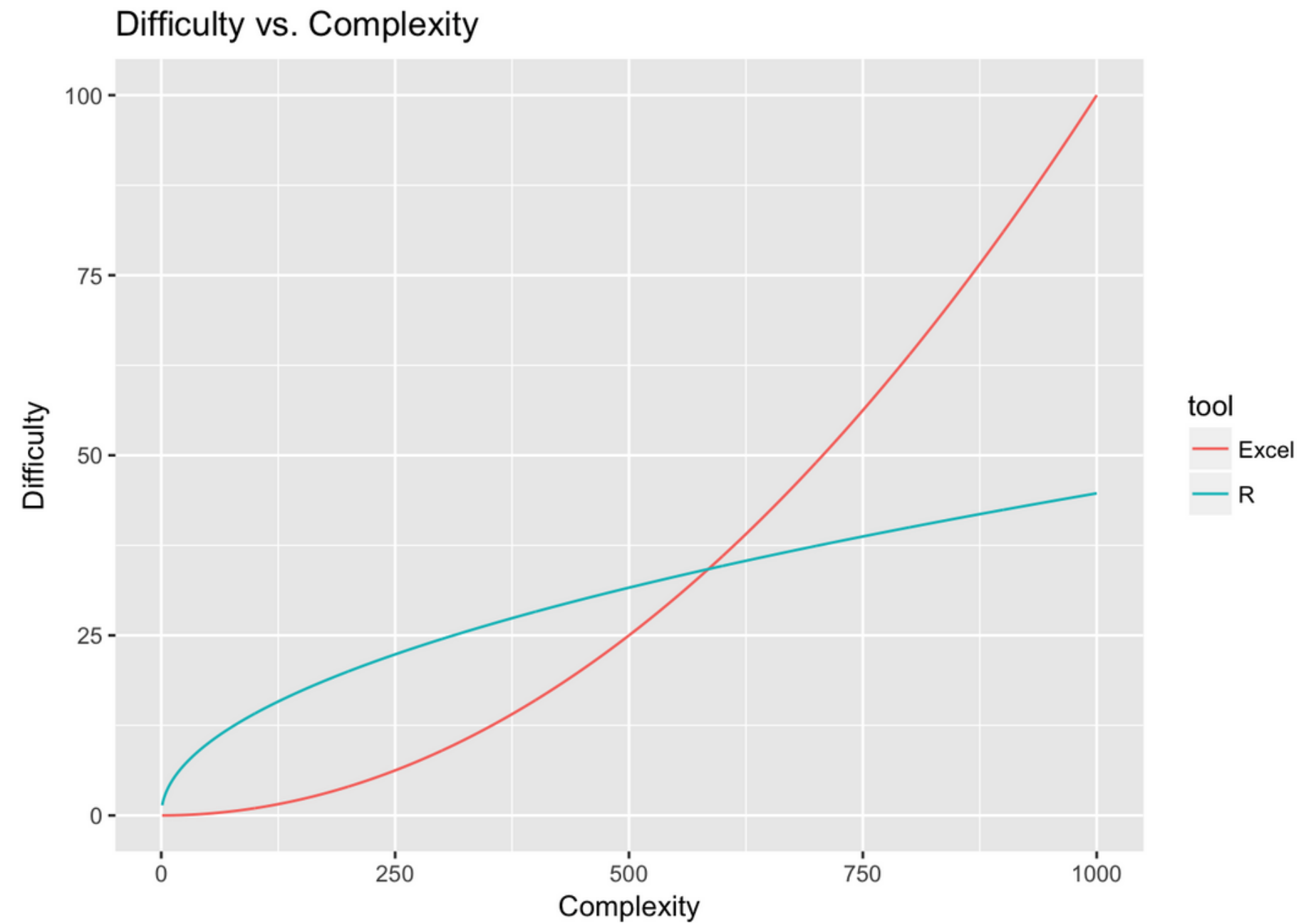


Ferramenta que vamos usar



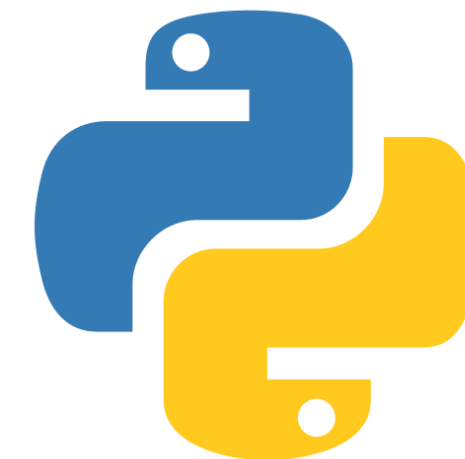
## Vantagens

- Trabalha muito bem com base de dados grandes
- Fácil automação
- Curva de complexidade x dificuldade (imagem)





# PORQUE NÃO PYTHON?



## Vantagens

- Uma das linguagens mais famosas e mais usadas atualmente
- Realiza as mesmas tarefas que o R dentro do que vamos aprender
- Generalista - serve para outras tarefas como desenvolvimento web

## Desvantagens

- Menos amigável para pessoas que não sabem programar
- Generalista - não é uma linguagem para estatística especificamente

# **CONCEITOS IMPORTANTES ANTES DE IR PARA O R**

# SOBRE O R



## O que é?

- R - linguagem de programação
- RStudio - ambiente de desenvolvimento

## Como usar?

### Baixando (recomendado):

- Baixar R e RStudio: [link download](#)

### Online/nuvem:

- RStudio Cloud: [link download](#)



# TIPOS DE ARQUIVOS



## .R

- Só código :
  - não dá para personalizar
  - não dá para adicionar outros arquivos (imagens, vídeo, etc)
- Não fica tão legal para organizar as análises
- Bom para rodar aplicações ou caso você goste

## .Rmd (R markdown)

- Código e texto
  - altamente personalizável
  - suporta vários tipos de arquivos
- Muito fácil de organizar as análises e bom para compartilhar seus resultados (gerar relatórios)

**Vamos preferir usar esse tipo de arquivo**

# FORMATO PARA ARMAZERNAR OS CONJUNTOS DE DADOS

- **CSV: comma-separated values**
- **TSV: tab-separated values)**
- **XLS e XLSX: arquivos de planilhas Excel**
- **outros (bd, json, etc)**

# DICA: ONDE ACHAR DATASETS

kaggle



**VAMOS PARA O  
RSTUDIO**



# FONTES

- <https://www.aquare.la/saindo-do-limitado-excel-para-r-ou-melhor-python/>
-