

# Introdução à análise de dados

**Felipe Carvalho**     *National Institute for Space Research (INPE), Brazil*

**Felipe Menino**     *National Institute for Space Research (INPE), Brazil*

---

A aplicação dos conceitos vistos no minicurso de introdução à análise de dados é extremamente importante para a fixação dos conceitos. Desta forma, esta lista de exercícios busca criar situações que ajudem o aluno na aplicação e fixação de tais conceitos.

---

## 1. Contexto

A Universidade Federal do Oeste da Bahia (UFOB), disponibiliza dados nominais de alunos que trancaram suas matrículas na universidade. Os dados disponibilizados referentes aos anos de 2014 à 2017 estão no [Portal Brasileiro de Dados Abertos](#)

## 2. Atividade

Como apresentado ao longo das atividades do curso, a análise de dados é basicamente uma atividade utilizada para responder perguntas através de diferentes métodos de tratamento e manipulação dos dados.

Com base nisto, faça a análise do conjunto de dados **MIAD002**, que está no repositório do minicurso, dentro do diretório **atividades**.

Com este conjunto de dados, responda as seguintes questões:

1. Qual o curso com maior quantidade de matrículas trancadas no ano de 2015 ?
2. Faça um **plot** com a variação entre os anos do número de matrículas trancadas por curso.
3. Qual semestre (de todos os anos) contém a maior taxa de matrículas trancadas ?

### ***Dicas***

Abaixo uma tabela com a relação de funções que podem ser úteis para a solução dos exercícios propostos

Descrição	R	Python
Leitura de dados	<code>readr::read_csv()</code>	<code>pd.read_csv()</code>
Agrupamento	<code>dplyr::group_by()</code>	<code>pd.DataFrame.groupby()</code>
Filtro	<code>dplyr::filter()</code>	<code>pd.DataFrame[Condição]</code>
Desvio padrão	<code>stats::sd()</code>	<code>pd.DataFrame.std()</code>
Média	<code>mean()</code>	<code>pd.DataFrame.mean()</code>
Mínimo	<code>min()</code>	<code>pd.DataFrame.min()</code>
Máximo	<code>max()</code>	<code>pd.DataFrame.max()</code>