

Big Data Analytics

Projeto de bloco



Agenda

- Projeto de bloco;
- Repositório de dados;
- Primeira tarefa.



Projeto de Bloco



Projeto de bloco

- O projeto de bloco de nossa disciplina será gerado ao longo das aulas, conforme formos avançando no conhecimento da linguagem R e de técnicas de análise de dados.



Projeto de bloco

- O projeto de bloco de nossa disciplina será gerado ao longo das aulas, conforme formos avançando no conhecimento da linguagem R e de técnicas de análise de dados.



Repositórios de dados



Repositórios de dados

- Para executarmos o projeto, precisaremos de dados.
- Porém dados são muito valiosos, portanto contaremos com repositórios de dados públicos ou de interesse acadêmico.



Repositórios de dados

- [Data Repositories;](#)
- [Recommended Data Repositories;](#)
- [20 Big Data Repositories You Should Check Out;](#)
- [Digging into Data Challenge;](#)
- [UCI Machine Learning Repository;](#)
- [Portal Brasileiro de Dados Abertos.](#)



Primeira tarefa



Primeira tarefa

- Analisar as sugestões de repositórios e escolher um conjunto de dados para estudo.
- Pode ser usada uma base de dados escolhida pelos alunos, que não esteja nas sugestões de repositórios, porém cuidado com direitos autorais e/ou confidencialidade dos dados.
- **A base escolhida deverá sempre ser aprovada pelo professor (sugestão: na segunda aula do curso)!**



Primeira tarefa

- Fazer um estudo inicial da base escolhida e documentar suas características básicas.
- Esta primeira análise pode ser feita em Excel, se preferirem.
- A base escolhida, a princípio, deverá ser utilizada durante todo o módulo. Apenas uma alteração será permitida!
- Grupos de até 3 alunos!



Primeira tarefa

- Nosso objetivo é executar uma análise de dados completa até o final do módulo.
- Neste primeiro momento, escolham a base de dados e pensem em uma ou duas perguntas iniciais que deverão ser respondidas durante a análise.

