

Minicurso de Introdução à Manipulação de Textos Usando o R

Exercícios Aula 3

1. Baixe o arquivo .txt disponibilizado no Google Classroom e armazene-o em uma variável.
2. Execute os comandos aprendidos na aula e tenha um vetor de caracteres contendo as palavras do arquivo .txt (Lembre-se que cada objeto do vetor precisa ser uma palavra!).
3. De uma olhada no vetor resultante, perceba que como foi dito e mostrado na aula, ele possui algumas palavras estranhas que dificultariam uma futura análise (Ex. “come9ara”). Vamos supor que essas palavras não teriam impacto nenhum em nossa análise e vamos desconsiderá-las, fingindo que são erros de código que apareceram no arquivo durante o download. Crie uma variável chamada “vet_num” e armazene nela todas as palavras que contêm dígitos e uma variável chamada “vet_limpo”, onde você irá armazenar o texto sem as palavras com dígitos.
4. Após futura análise, você identificou outro erro de código. Dessa vez, foram identificados diversos objetos do vetor contendo a escrita “â”. Crie um novo vetor chamado “vet_a” e atribua a ele todas as palavras que possuem “â”. Após isso, atualize a variável “vet_limpo” removendo todas as palavras que possuem “â”.
5. Vamos supor que você decida calcular a proporção de palavras errôneas vindas do arquivo .txt. Para isso, você precisa criar um novo vetor contendo as palavras que foram consideradas erros (Palavras com dígitos e “â”). Após isso, calcule a proporção erro/palavra do arquivo baixado.