**CURSO DE INTRODUÇÃO A MANIPULAÇÃO DE BIG DATA COM R**

OBJETIVOS: Introduzir conceitos, pacotes e funções fundamentais do R para executar as seguintes tarefas: Importar um conjunto de dados de um arquivo; Selecionar, remover e criar colunas; Filtrar linhas, agrupar e sumarizar os dados; Construção de gráficos com os dados; Apresentar conceitos sobre Spark, Parquet e a biblioteca sparklyr do R; Manipulação de dados com sparklyr; Divulgar material e maneiras de procurar ajuda com os erros;

a - f: 3 horas  
g - j: 3 horas

1. INTRODUÇÃO AO R: 6 horas
   1. Instalação do R (5min) ok
   2. Instalação do Rstudio (5min) ok
   3. Primeiros passos no Rstudio (5min) ok
   4. R Base e instalação de pacotes (5min) ok
   5. Pedindo ajuda no/sobre R (função help, Stackoverflow, ...) (5min) ok
   6. R como calculadora (30min) ok
      1. Soma, multiplicação, divisão, srqt, log, …
      2. O objeto atômico (variável) ok
         1. numeric, integer, boolean, string
      3. Calculando comparações boolean ok
   7. Estrutura de dados (Objetos no R) (2hr e 20min, explicação e aplicação)
      1. Vetor (20min)
      2. Misturando objetos (20min)
      3. Matrizes (20min)
      4. Fatores (20min)
      5. Valores especiais (NA, Inf, NULL) (20min)
      6. Listas (20min)
      7. data.frame (20min)
   8. Funções **(40min)**
      1. O que são, pra que servem e como usar funções
      2. Funções do R Base
         1. Mean, median, quantile, ...
      3. Funções de pacotes
      4. Funções customizadas (feitas por você mesmo)
   9. Controles de fluxo **(30min)**
      1. If e Else
      2. For
   10. Fórmulas **(30 min)**
   11. Wrap-up
2. MANIPULAÇÃO DE DADOS COM R: 10 horas
   1. Introdução e contextualização **(10min)**
   2. Importação de dados **(1hr e 30min)**
      1. R Base (20min)
      2. A biblioteca readr (20min)
      3. A biblioteca haven para importação de arquivos .dta .sas .sav (20min)
      4. Manipulando variáveis com R Base (30min)
   3. O operador pipe (%>%) **(20min)**
   4. Trabalhando com tibbles **(10min)**
   5. O pacote dplyr **(3hrs)**
      1. Filter (20min)
      2. Select (20min)
      3. Arrange (20min)
      4. Mutate (20min)
      5. Group\_by (20min)
      6. Summarise (20min)
      7. Merge (20min)
   6. O pacote tidyr **(1hr e 30min)**
      1. Gather (20min)
      2. Spread (20min)
      3. Outras funções do tidyr (10min)
   7. Fazendo descritivas com dplyr e tidyr **(2hrs)**
   8. O pacote ggplot2 **(2hrs)**
      1. O que são gráficos no ggplot2? (10min)
      2. As camadas de um gráfico (20min)
      3. Gráfico de dispersão (20min)
      4. Gráfico de barras (20min)
      5. Gráfico boxplot (20min)
      6. Gráfico histograma (20min)
      7. Onde encontrar exemplos para inspiração e códigos (10min)
   9. Outros pacotes do Tidyverse **(10min)**
   10. Onde encontrar material para estudo **(5min)**
   11. Wrap-up
3. INTRODUÇÃO AO APRENDIZADO DE MÁQUINA (Sujeito a alteração): 4 horas
   1. Definição
   2. Tarefas de Aprendizado
      1. Descritivas
      2. Preditivas
   3. Pré-processamento
   4. Técnicas
   5. Avaliação
   6. Hands-on!
4. MANIPULAÇÃO DE BIG DATA COM SPARKLYR: 6 horas
   1. O que é e pra que serve o Spark? **(20 - 30min)**
   2. O formato de arquivos Parquet **(20 - 30min)**
   3. Instalando o pacote sparklyr **(10min)**
   4. Instalando o Spark através do R **(20min)**
   5. A biblioteca sparklyr **(2hrs e 30min)**
      1. Conectando no Spark (5min)
      2. Importando dados no Spark (10min)
      3. Utilizando o SparkUI para visualizar e gerenciar os jobs (10min)
      4. O objeto tbl\_spark (15min)
      5. Manipulando um data.frame do Spark (wrap-up do que já foi aprendido no tópico 2.e) (30min)
      6. A função collect (20min)
      7. Leitura e escritas de arquivos em Parquet (10min)
      8. Erros do Spark vs erros do R (20min)
      9. Pedindo ajuda sobre Spark e Sparklyr (função help, Stackoverflow, ...) (5min)
      10. Configurando recursos do Spark (10min)
      11. Outras bibliotecas do Spark (15min)
          1. SparkMlib
   6. Aplicação **(1hr)**
   7. Wrap-up
5. TÓPICOS ESPECIAIS: 2 HORAS
   1. Onde achar material para construção de mapas no R
   2. Discutir dúvidas e divulgar materiais