



INTEGRAÇÃO POSTGRESQL E RSTUDIO PARA ANÁLISE EXPLORATÓRIA DE EMISSÕES EM CO₂EQ PROVENIENTES DE METANO NA AGRICULTURA

Rafael Salazar Stavale¹, Bruno Francisco Teixeira Simões²,

1. Aluno do Bacharelado em Sistemas de Informação - Universidade Federal do Estado do Rio de Janeiro

2. Departamento de Métodos Quantitativos - Universidade Federal do Estado do Rio de Janeiro

Introdução

A questão ambiental quanto ao aquecimento global é uma grande preocupação científica a várias décadas. Entender as suas causas e consequências necessita de uso de várias áreas do conhecimento humano para chegar a soluções mais assertivas e eficientes. Os órgãos engajados com essa preocupação ambiental disponibilizam suas bases de dados com o intuito de gerar estudos colaborativos, compondo uma grande quantidade de dados (Big Data) que exerce um papel vital em estudos de meio ambiente. Esse trabalho se propõe a mostrar um modelo de criação de um sistema integrado entre o banco de dados PostgreSQL e o Software R tornando viável a realização de análises do grande volume dos dados abertos da ONU, mais especificamente em relação às Emissões em CO₂eq de todos os animais na pecuária no mundo. A pecuária é responsável por 8-10.8% das emissões de gases de efeito estufa, de acordo com O'MARA (2011).

Objetivos

O trabalho consiste em criar um ambiente integrado do banco de dados PostgreSQL e o Software R, proporcionando acesso direto às informações armazenadas no banco e realizar uma análise exploratória sobre Emissões de CO₂eq proveniente de animais na pecuária no mundo no período de 1961 até 2016.

Material e Método

Os dados obtidos são secundários e de acesso aberto. Um ambiente de banco de dados foi criado com a solução gratuita PostgreSQL, utilizado padrão ANSI e o número total de linhas acessadas na base de dados utilizadas para a carga inicial do banco foi de 9.193.883.



No R foram utilizados dois pacotes de conexão direta com o banco, o *DBI* e o *RPostgres*. Dessa forma foi possível realizar consultas SQL diretamente dos scripts R e criar dataframes com o resultado, tornando-se possível importar toda a base de dados. Após as operações de importação foi criada uma função de transposição de dataframes que foi usada sistematicamente para todos os resultados da consulta no banco, resultando nos dataframes em formato 'tidy' desejados. Os dataframes foram agrupados em um único objeto com 1152 variáveis. Foi obtido um valor mediano com objetivo de se utilizar um estimador que seja robusto para as variáveis se surgirem valores arbitrariamente grandes. O projeto foi versionado em um repositório GitHub e todas as mudanças foram feitas pela integração do RStudio na própria interface sem a necessidade de uso de linha de comando.

Resultados e Discussão

A qualidade da integração entre o banco de dados e a ferramenta R foi considerada satisfatória quanto à facilidade de uso quando pronta. O PostgreSQL otimiza o tamanho da base de dados em relação aos dados, com a carga inicial foi obtido um tamanho de base útil de 1.3GB. Uma desvantagem dessa arquitetura é o esforço de produzir scripts para o banco de dados, para o versionamento no GitHub e para as conexões no Rstudio.

Na Figura 1 verifica-se a distribuição da variável de emissões totais em CO₂eq e a representação da magnitude (em bolhas) do PIB de cada um dos 246 países do estudo. Uma hipótese que pode ser aqui feita é a existência de uma relação entre o PIB do país e o comportamento dessa variável. Foi realizado o teste de Shapiro-Wilk e verificado que a variável de emissões totais em CO₂eq não segue distribuição normal (p-valor<2.2e-16), ao nível de significância de 0,01. O valor da correlação de Spearman de 0,6824 confirma esta hipótese.

Quanto às análises, é possível identificar a Índia como um outlier em relação aos outros países e isso é corroborado em PATRA et.al (2012). A Índia tem uma das maiores produções pecuárias do mundo e comparado a outros grandes produtores não executa muito significativamente mitigação de metano mostrados em GERBER et.al (2013).

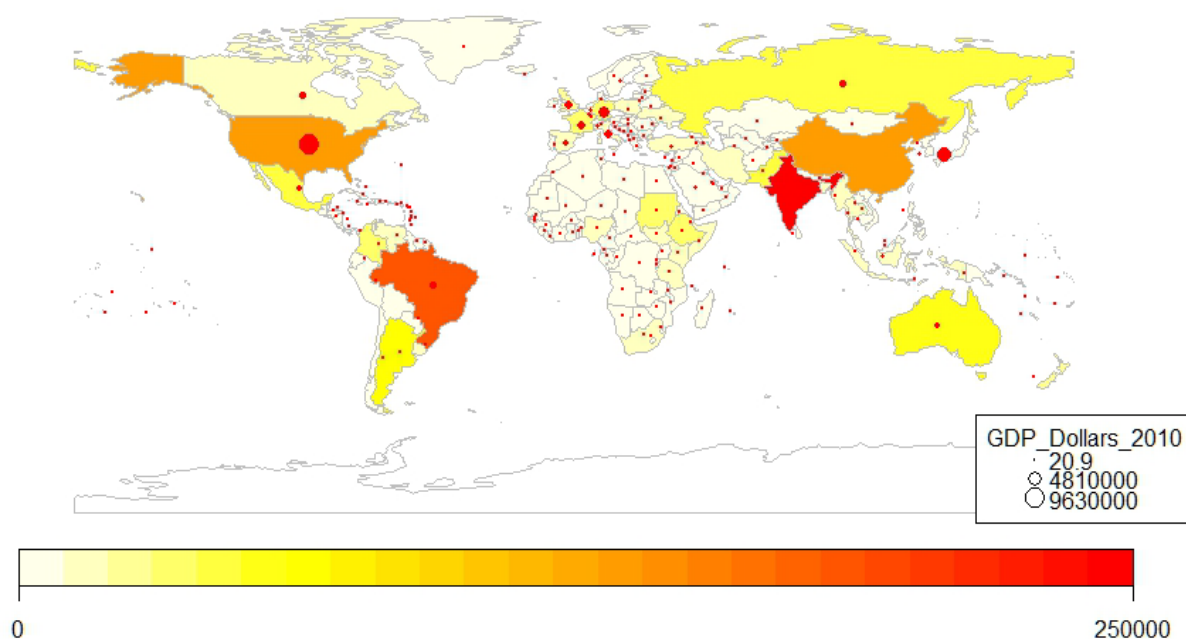


Figura 1 – Emissões em CO₂eq totais de animais na pecuária com bolhas de GDP

Conclusão

A arquitetura provou-se eficiente ao lidar com o volume de dados e a vantagem é que foi totalmente construída com ferramentas gratuitas e pacotes disponibilizados no CRAN do R. Uma hipótese da relação de ações de mitigação de gás metano com o PIB dos países foi confirmada, com uma correlação de Spearman de 0.68. Como sugestão para trabalho futuro, poderão ser identificadas mais variáveis que se relacionam com as emissões totais em CO₂eq e a construção de um modelo para a previsão desta variável.

Referências

GERBER, P. J. et al. Technical options for the mitigation of direct methane and nitrous oxide emissions from livestock: a review. *Animal*, v. 7, n. s2, p. 220-234, 2013.

O'MARA, Frank P. The significance of livestock as a contributor to global greenhouse gas emissions today and in the near future. *Animal Feed Science and Technology*, v. 166, p. 7-15, 2011.

PATRA, Amlan Kumar. Enteric methane mitigation technologies for ruminant livestock: a synthesis of current research and future directions. *Environmental Monitoring and Assessment*, v. 184, n. 4, p. 1929-1952, 2012.