

# Workshop: Dados de Comex no Brasil

## Fontes, Metodologias e Aplicações Práticas com R

Alan Leal

Doutor em Teoria Econômica pela Universidade de São Paulo  
Baseado em Leal & Martins (2025) e comexTL (pacote no R)

10 de setembro de 2025

# Bem-vindos ao Workshop!

**Objetivo:** Capacitar no uso de dados de comércio exterior brasileiro

**Duração:** 2 horas

**Formato:** Teoria + Prática (código R)

**Material disponível:** <https://alanleal-econ.com/index.php/teaching/>

## Agradecimentos especiais/Crédito:

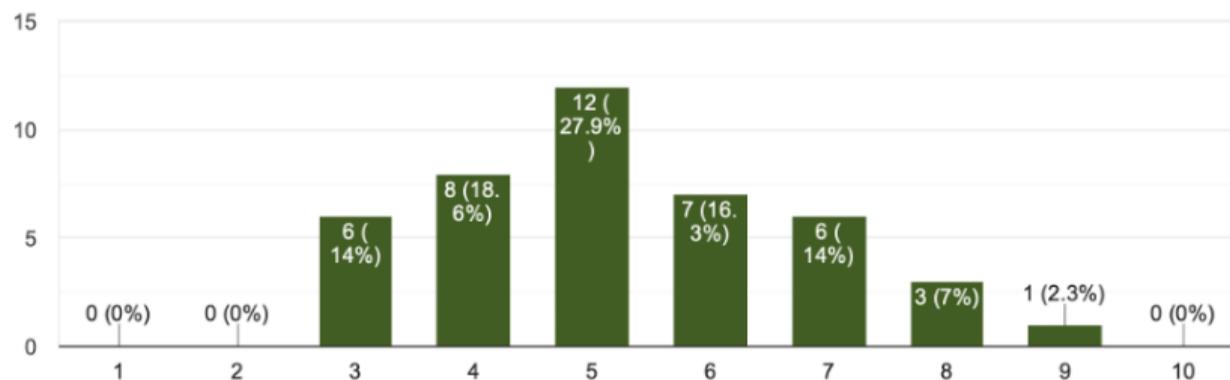
- Profa. Dra. Michelle Martins (UFV)
- MDIC/SECEX - Disponibilização dos dados abertos
- Participantes do workshop

# Conhecimentos Prévios I

## Objetivo: Alinhar expectativas e resultados do Workshop

Qual é seu conhecimento de Economia Internacional? (Mais próximo de 1, menor, mais próximo de 10, maior)

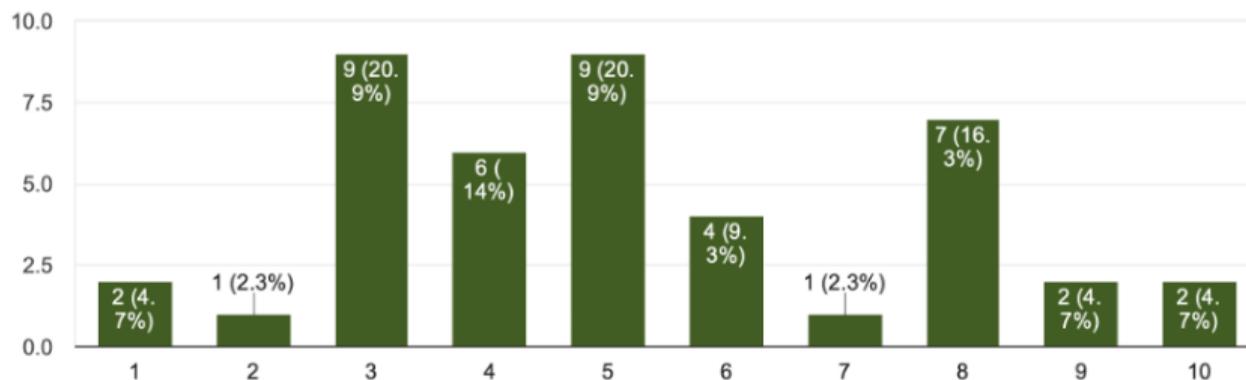
43 respostas



# Conhecimentos Prévios II

Qual é seu conhecimento de Economia Regional? (Mais próximo de 1, menor, mais próximo de 10, maior)

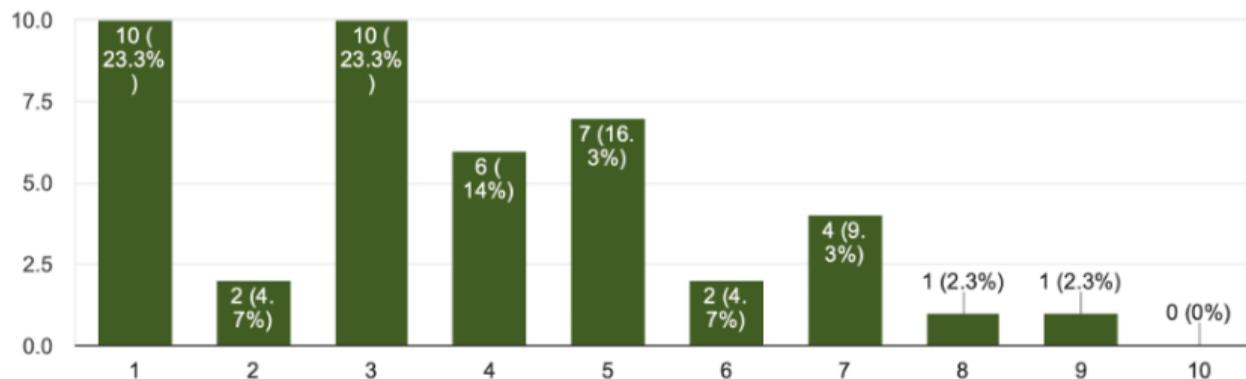
43 responses



# Conhecimentos Prévios III

Qual é seu conhecimento de R (programming language)? (Mais próximo de 1, menor, mais próximo de 10, maior)

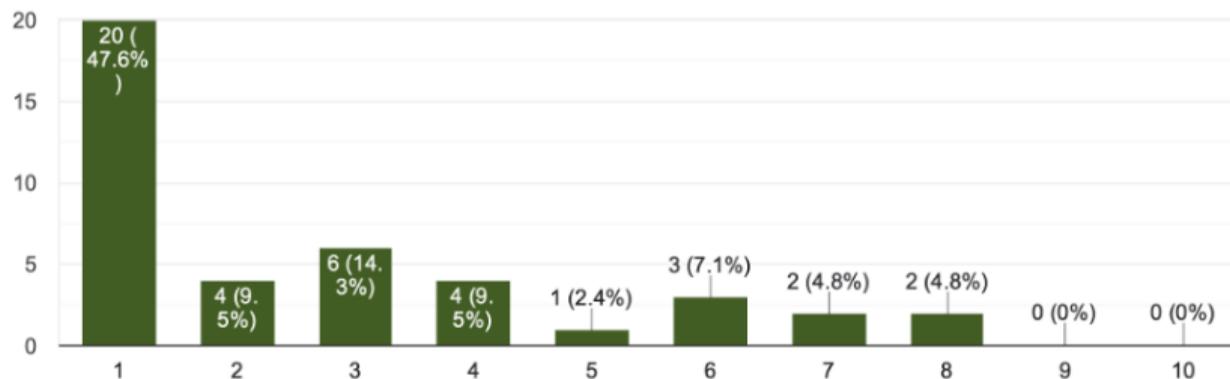
43 responses



# Conhecimentos Prévios IV

Qual é seu conhecimento de dados de comércio exterior (Comex Stat, UN COMTRADE, WITS, USA Trade Online, etc.)? (Mais próximo de 1, menor, mais próximo de 10, maior)

42 respostas



# Sistema Harmonizado (HS) I

## O que é o Sistema Harmonizado?

Sistema internacional de classificação de mercadorias desenvolvido pela Organização Mundial das Aduanas (WCO)

### Níveis de agregação:

- **HS2**: 2 dígitos (21 seções)
- **HS4**: 4 dígitos ( 1.200 categorias)
- **HS6**: 6 dígitos ( 5.000 produtos)
- **NCM**: 8 dígitos (Brasil/Mercosul)

### Exemplo - Soja:

- HS2: 12 - Sementes oleaginosas
- HS4: 1201 - Soja
- HS6: 120190 - Soja exceto para semeadura
- NCM: 12019000 (Brasil)

Referência: <https://www.wcoomd.org/>

# Nomenclatura no Brasil

## NCM - Nomenclatura Comum do Mercosul

- Baseada no Sistema Harmonizado (6 dígitos)
- Adiciona 2 dígitos específicos do Mercosul (8 dígitos total)
- Utilizada para classificação fiscal e estatística

## Hierarquia Importante

- HS6 = internacional, comparável entre países
- NCM (HS8) = nacional, específico por país
- Dados municipais brasileiros: limitados ao HS4
- Dados estaduais/nacionais: disponíveis em NCM (HS8)

Fonte: <https://www.gov.br/produtividade-e-comercio-externo/>

# WITS - World Integrated Trade Solution I

## Características

- Plataforma do Banco Mundial
- Dados de 190+ países
- Período: 1962-presente
- Níveis: HS2, HS4, HS6
- Gratuito com registro

## Vantagens:

- Interface amigável
- Dados padronizados
- Ferramentas de análise integradas
- Comparabilidade internacional

## Limitações:

- Dados agregados por país
- Sem informações sub-nacionais
- Pode ter defasagem temporal
- Limitado ao HS6

Acesso: <https://wits.worldbank.org/>

# UN Comtrade I

## Base de Dados da ONU

- Repositório oficial da ONU
- 170+ países reportando
- Dados desde 1962
- API disponível
- Mais de 3 bilhões de registros

## Vantagens:

- Fonte primária oficial
- Dados muito detalhados
- API para automação
- Metadados extensivos

## Limitações:

- Interface menos intuitiva
- Necessita conhecimento técnico
- Limitações na API gratuita
- Podem existir discrepâncias

Acesso: <https://comtrade.un.org/>

# BACI CEPII I

## Base Harmonizada (CEPII)

- Dados do UN Comtrade reconciliados
- Metodologia de limpeza avançada
- Elimina inconsistências mirror
- Formato padronizado
- Atualização anual

## Vantagens:

- Dados mais confiáveis
- Menos discrepâncias
- Metodologia transparente
- Documentação excelente

## Limitações:

- Defasagem de 1-2 anos
- Apenas comércio bilateral
- Não inclui serviços

Acesso: [http://www.cepii.fr/CEPII/en/bdd\\_modele/bdd\\_modele\\_item.asp?id=37](http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37)

# Disponibilização Gratuita: Casos Internacionais I

## Brasil - ComexStat (MDIC)

- Portal oficial do MDIC
- Dados municipais (HS4) e estaduais/nacionais (NCM)
- Série histórica desde 1997
- Atualizações mensais

## Outros Países com Dados Sub-nacionais

- **Estados Unidos:** USA Trade Online - dados por estado e distritos
- **China:** China Customs Statistics - dados provinciais
- **Espanha:** ICEX DataComex - dados regionais
- **Canadá:** Statistics Canada - dados provinciais
- **União Europeia:** Eurostat Comext - dados regionais NUTS

Brasil é um dos poucos países com dados municipais gratuitos!

# O Grande Problema: Cobertura vs. Produção I

## Limitação Crítica dos Dados Brasileiros

Os dados municipais do ComexStat mostram onde a operação de exportação foi **registrada burocraticamente**, não onde o produto foi **produzido** para exportação.

### Evidências do Problema (Leal & Martins, 2025)

- 62% dos municípios de Mato Grosso não exportaram soja oficialmente em 2023
- Mas MT é o maior produtor de soja do Brasil!
- Concentração artificial em grandes centros urbanos
- Sub-representação de municípios produtores rurais

### Consequências:

- Análises enviesadas de competitividade regional
- Políticas públicas mal direcionadas
- Subestimação do potencial exportador local

# Outras Peculiaridades I

## Nomenclatura:

- Municipal: apenas HS4
- Estadual/Nacional: NCM (HS8)
- Mudanças de classificação ao longo do tempo
- Correspondência entre HS e NCM nem sempre perfeita

## Temporalidade:

- Dados desde 1997
- Atualizações mensais
- Revisões retrospectivas
- Defasagem de 30-45 dias

## Principais Dificuldades:

- Quebras estruturais nas séries
- Mudanças metodológicas do MDIC
- Qualidade heterogênea por período
- Códigos municipais alterados (IBGE)
- Empresas que mudam de domicílio
- Terceirização de operações de comex

# Outras Peculiaridades II

Fonte: MDIC/SECEX - <http://comexstat.mdic.gov.br/>

# Portal ComexStat: Capacidades e Limitações I

## Funcionalidades do Portal

- Interface web para consultas interativas
- Download de dados em CSV/Excel
- Filtros por período, produto, país, UF, município
- Visualizações básicas integradas

## Limitações Críticas

- **Consulta e Downloads limitados a 150.000 registros por consulta**
- Interface não permite consultas complexas programáticas
- A API oficial disponível é relativamente confusa, pois segue uma filosofia de consulta semelhante ao Comex Stat, que não se traduz bem para situações de código.
- Dados municipais limitados ao HS4
- Processo manual é demorado para grandes volumes

# Portal ComexStat: Capacidades e Limitações II

**Solução:** Pacote comexTL automatiza e supera essas limitações!

Acesso: <https://alanleal-econ.com/index.php/codes/comextl/>

**comexTL:** Acesso programático aos dados hospedados em Cloudflare R2, sem limitações de volume.

# Pacote comexTL: Visão Geral I

## O que é o comexTL?

Pacote R desenvolvido por Alan Leal que oferece acesso programático completo aos dados de comércio exterior brasileiro, buscando contornar as limitações do portal ComexStat.

## Principais Características:

- **Sem limitações de volume:** Download de datasets completos
- **Sistema de cache inteligente:** Trabalho offline após download inicial
- **Quatro módulos especializados:** Tabelas de referência, exportações, importações, dados municipais
- **Tecnologia Apache Arrow:** Processamento eficiente de arquivos Parquet
- **Interface unificada:** Funções consistentes para diferentes tipos de análise

GitHub: <https://github.com/alanleal-econ/comexTL>

# Arquitetura do comexTL I

## Quatro Módulos Principais

- 1 **Tabelas de Referência:** NCM, países, UFs, municípios
- 2 **Dados Gerais de Exportação:** Análise por NCM/SH completo
- 3 **Dados Gerais de Importação:** Análise por NCM/SH com métricas específicas (frete, seguro)
- 4 **Dados Municipais:** Análise por município da empresa (limitado a SH4)

## Infraestrutura:

- **Armazenamento:** Cloudflare R2 (alta disponibilidade)
- **Formato:** Arquivos Parquet (dados) + CSV (referência)
- **Cache:** Sistema multicamadas para performance
- **Atualizações:** Seguem cronograma oficial do MDIC, sendo disponibilizado geralmente nos dez primeiros dias do mês seguinte à operação.

# O Problema Identificado por Leal e Martins (2025) I

**Evidência Empírica** " Em 2023, cerca de 62% dos municípios de Mato Grosso não exportaram oficialmente soja segundo o ComexStat, apesar de serem grandes produtores do grão."

## **Problemas Decorrentes:**

- Análises econométricas enviesadas
- Políticas públicas mal direcionadas
- Subestimação do potencial exportador regional
- Concentração artificial em poucos municípios
- Dificuldade em medir competitividade real

**Necessidade:** Desenvolver método para estimar a verdadeira produção municipal destinada à exportação, baseada na localização da produção, não do registro burocrático.

# A Solução: Algoritmo de Imputação Espacial I

## Insights Fundamentais

- 1 Existe mais conhecimento sobre distribuição espacial da produção do que das exportações burocráticas
- 2 Dados brasileiros têm "âncoras espaciais" imperfeitas mas utilizáveis
- 3 Podemos combinar dados de comércio com dados de produção/emprego

**Método Proposto:** Combinar informações de:

- Dados oficiais de exportação (origem burocrática) - ComexStat
- Dados de emprego formal por setor (origem econômica) - RAIS
- Relações espaciais entre municípios - matrizes de contiguidade
- Coeficientes de proporção estadual (HS6/HS4)

**Resultado:** Estimativa mais realista da produção municipal destinada ao mercado externo.

# Metodologia: Equação Principal I

O algoritmo estima exportações municipais em nível HS6 usando:

$$E_{i,HS6} = \underbrace{\alpha WE_{i,HS4}}_{\text{Componente Espacial}} + \underbrace{rE_{i,HS6}^*}_{\text{Componente de Ajuste}} \quad (1)$$

Onde:

- $E_{i,HS6}$ : exportações estimadas do município  $i$  (HS6)
- $\alpha$ : coeficiente de proporção estadual =  $\frac{E_{estado,HS6}}{E_{estado,HS4}}$
- $W$ : matriz de pesos espaciais normalizada ( $n \times n$ )
- $E_{i,HS4}$ : exportações observadas do município  $i$  (HS4)
- $r$ : vetor de relevância municipal (baseado em emprego formal)
- $E_{i,HS6}^*$ :  $\sum_{i=1}^I \alpha (E_{i,HS4} - WE_{i,HS4})$  (componente de ajuste)

# Componentes do Algoritmo I

## 1. Matriz de Pesos Espaciais (W):

- **Queen (Contiguidade):**  $w_{ij} = 1$  se municípios são vizinhos, 0 caso contrário
- **Distância Inversa:**  $w_{ij} = \frac{1}{d_{ij}}$  onde  $d_{ij}$  é a distância entre centróides
- **Normalização:**  $\sum_j w_{ij} = 1$  para cada linha  $i$

## 2. Vetor de Relevância (r):

- Baseado em dados de emprego formal (RAIS)
- $r_i = \frac{\text{empregos formais setor } i \text{ no município}}{\text{total empregos formais setor } i \text{ no estado}}$
- Propriedade:  $\sum_i r_i = 1$  (normalizado)
- Captura importância relativa do setor em cada município

## 3. Coeficiente :

- Proporção estadual entre HS6 específico e HS4 geral
- Permite desagregação de HS4 municipal para HS6 municipal

# Propriedades Matemáticas I

## 1. Consistência Dimensional

A operação matricial é bem definida:  $(n \times n) \cdot (n \times 1) = (n \times 1)$

## 2. Propriedade de Agregação

A soma dos componentes espacial e de ajuste preserva o total estadual:

$$\sum_{i=1}^n E_{i,HS6} = E_{estado,HS6}$$

## 3. Tratamento de Zeros

Municípios com exportações zero são naturalmente acomodados pela estrutura matricial, recebendo valores estimados baseados na proximidade espacial e relevância setorial.

**Resultado:** Método matematicamente consistente e empiricamente fundamentado que preserva agregados conhecidos.

# Por que a Madeira no Mato Grosso em 2022? I

## Justificativa para o caso de estudo:

- O comércio de madeira é um importante driver de desmatamento
- Produto com forte concentração geográfica conhecida
- Dados de desmatamento são usados para calibrar o vetor  $r$ , de relevância municipal
- Mato Grosso é um estado de transição de três biomas, no qual a madeira tem um papel importante para diversos municípios.

## Dados utilizados:

- **Exportações:** ComexStat - HS4: 4409 (Madeira)
- **Desmatamento:** PRODES 2022 - desmatamento no Mato Grosso em 2022
- **Geometria:** Malha municipal de MT (IBGE)

**Expectativa:** Algoritmo deve identificar municípios das regiões tradicionais (Cerrado e Amazônia).

# Recursos Adicionais I

## Documentação e Tutoriais:

- **GitHub:** <https://github.com/alanleal-econ/comexTL>
- **Documentação completa:** Manual de 23 páginas disponível
- **Exemplos práticos:** Códigos completos no repositório
- **Issues:** Reporte bugs e sugira melhorias

## Outras Ferramentas Úteis:

- **Pacote geobr:** Dados geográficos brasileiros
- **Pacote sidrar:** Dados do IBGE via API
- **Pacote ipeadatar:** Dados econômicos do IPEA
- **Portal ComexStat:** <http://comexstat.mdic.gov.br/>
- **IBGE PAM:** Produção Agrícola Municipal para validação

## Recursos Adicionais II

### Artigos de Referência:

- Leal, A. & Martins, M. M. V. (2025). Brazilian exports imputation: A new algorithm for estimating municipal production directed at the foreign market. *Revista Brasileira de Estudos Regionais e Urbanos*, 19(2), 266-288.

# Próximos Passos

## Para continuar aprendendo:

- 1 **Prática:** Execute os exemplos do manual do comexTL
- 2 **Leitura:** Artigo completo de Leal & Martins (2025)
- 3 **Aplicação:** Use o algoritmo para seu setor/região de interesse
- 4 **Contribuição:** Reporte bugs, sugira melhorias
- 5 **Rede:** Conecte-se com outros usuários nas discussões

**Obrigado pela participação!**

*Dúvidas e sugestões são sempre bem-vindas*

*E-mail: [prof@alanleal-econ.com](mailto:prof@alanleal-econ.com)*