### Limpando e organizando seus dados

Faculdade de Ciências Econômicas Programa de Pós-Graduação em Economia PPGE-FACECON-ICSA-UFPA

Sérgio Rivero

PPGE-FACECON-ICSA

11 de maio de 2019





#### Sumário

- Objetivos da Aula O que é uma boa base de dados e que tipos de bases existem? Algumas das funcionalidade do R são[VS11]:
- Dúvidas
- Avaliação
- Para Levar para Casa
- 6 Agradecimentos
- Referências





### Objetivos da Aula

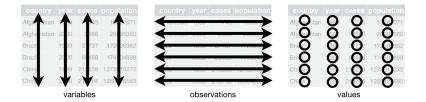
Discutir algumas funções de correção e limpeza de bases de dados em R.





# O que é uma boa base de dados e que tipos de bases existem?

Boas bases de dados para trabalhar, sobretudo com tidy, são bases que possuem variáveis nas colunas e observações nas linhas. Podemos identificar tal estrutura na figura 1 abaixo:



Organização ideal para base de dados





### Tipos de Bases de Dados

Há três tipos bases de dados principais que comportam a estrutura de variáveis nas colunas e observações nas linhas que são:

- Séries Temporais
- Cortes Transversais
- Dados em Painel





## Série Temporal

Séries temporais(ou longitudinais) são identificadas como observações ordenadas ao longo do "tempo". Exemplos:

- PIB distribuído anualmente;
- Arrecadação mensal de um ou mais Estados:
- Taxa de câmbio diária.





#### A base do BCB

Um bom exemplo de fonte de bases de dados de series temporais é o SGS - Banco Central. Segue o link abaixo:

	Parâmetros informados	
Séries selecionadas		
20783 - Spread médio das operações	de crédito - Total	
Período	Função	
01/03/2018 a 27/03/2019	Linear	
Registros encontrados por série: 12		
Lista de valores (Formato numérico: Europeu - 123.456.789,00)		
Data		
mês/AAAA	p.p.	
mar/2018	19,93	
abr/2018	19,56	
mai/2018	18,46	
jun/2018	17,68	
jul/2018	17,69	
ago/2018	17,58	
set/2018	17,26	
out/2018	17,88	
nov/2018	18,08	
dez/2018	16,94	
jan/2019	18,58	
fev/2019	19,04	
Fonte	BCB-DSTAT	

Exemplo de série temporal

https://www3.bcb.gov.br/sgspub/localizarseries/ localizarSeries.do?method=prepararTelaLocalizarSeries

Dados transversais(ou cross-section) são identificados quando não há importância de ordenação das observações, além de representar dados em um único ponto do "tempo". Exemplos:

- Arrecadação de ICSM para todos Estados do Brasil para o ano de 2018:
- Dados demográficos para todos os municípios do Pará para o ano de 2010.

Um bom exemplo de fonte de bases de dados de cortes transversais é o Sidra - IBGE. Segue o link abaixo:

https://sidra.ibge.gov.br/home

Tabela 3939 - Efetivo dos rebanhos, por tipo de rebanho Variável - Efetivo dos rebanhos (Cabeças) Ano - 2017					
			Tipo de rebanho - Bovino		
			Município		
Abaetetuba (PA)	2.800				
Abel Figueiredo (PA)	46.364				
Acará (PA)	10.740				
A4++ (DA)	1 571				









## Avaliação

Avaliação





## Para levar para casa

Fechamento da Aula





### Agradecimentos

Ao Monstro do Espaguete Voador À UFPA Etc...





### Referências I



W N Venables and D M Smith.

An Introduction to R, volume 2.

CRAN - Comprehensive R Archive Network, 2011.



