

Gestão Baseada em Dados no Âmbito da UFRN

Capacitação para Alta Gestão

Marcus Nunes

22 de Novembro de 2021

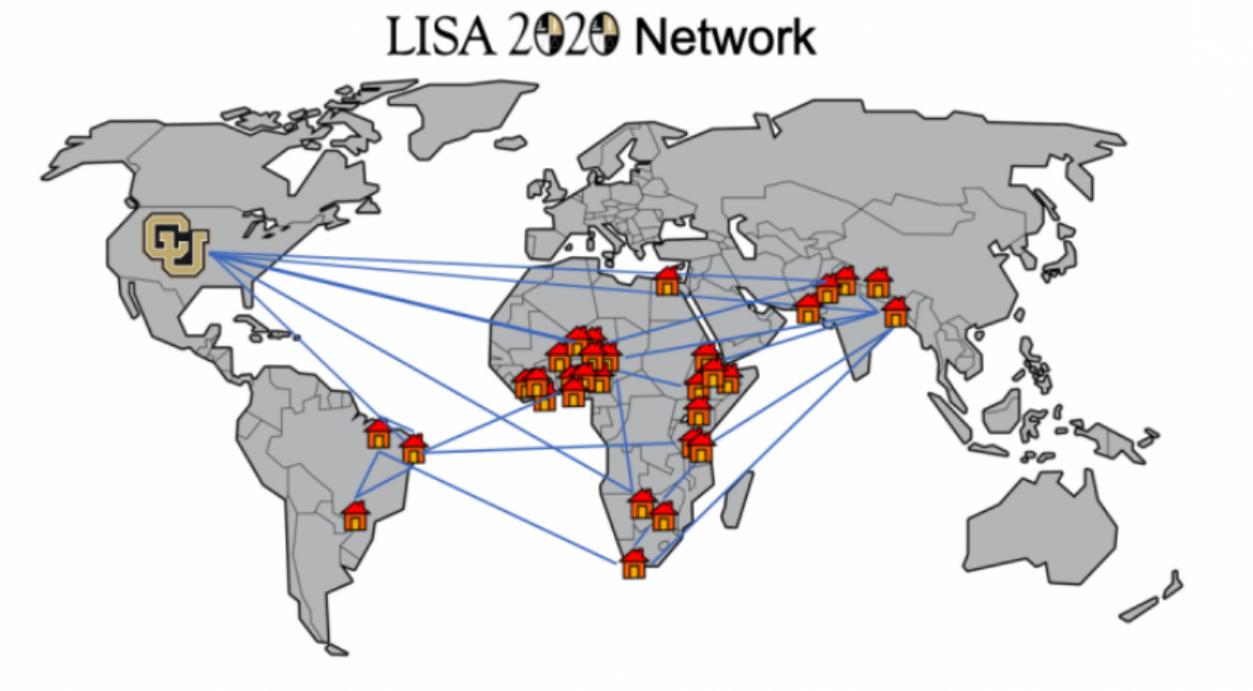
Departamento de Estatística - UFRN

Roteiro

1. Apresentação
2. Motivação
3. Cuidados com a Coleta e Análise de Dados
4. Análise Exploratória, Inferência, Predição
5. Exemplos na UFRN
6. Conclusões

Apresentação

- Marcus Nunes, Professor Adjunto no Departamento de Estatística
- PhD em Estatística pela Pennsylvania State University (2013)
- Interessado em Educação Estatística, Aprendizado de Máquina e Projetos de Colaboração Estatística
- Site pessoal: `marcusnunes.me`
- Email: `marcus@marcusnunes.me`



link

Motivação

- A UFRN registra e produz diversos dados sobre os alunos, funcionários e professores
- Por que não usar estes dados para responder perguntas sobre a universidade?
- Identificar deficiências
- Auxiliar em tomadas de decisão
- Otimizar recursos

Cuidados com a Coleta de Dados

Cuidados com a Coleta de Dados

- A qualidade da análise realizada depende diretamente da qualidade dos dados coletados
- Dados que não foram coletados corretamente não irão gerar boas análises
- É preciso entender bem os dados e saber das suas limitações
- Muitas vezes, a análise já surge sem futuro, pois os dados não foram coletados da maneira devida

Cuidados com a Coleta de Dados

- **População:** é o conjunto de todos os elementos ou resultados sob investigação
- **Amostra:** qualquer subconjunto da população de interesse

Cuidados com a Coleta de Dados

- As observações contidas em uma amostra são mais ou menos informativas quanto mais conhecermos a população que nos interessa
- Por exemplo, ao analisarmos a glicose de um paciente a partir de uma amostra de sangue retirada do seu dedo, teremos uma boa ideia a respeito da quantidade de glicose no corpo todo
- Por outro lado, se estivermos interessados na opinião da comunidade acadêmica a respeito de um projeto de infraestrutura, escolher uma amostra de 200 alunos e servidores do departamento ou centro beneficiado pelo projeto vai gerar um viés de seleção

- **Estudo Experimental:** grupos específicos para os sujeitos do estudo são designados pelo pesquisador
- **Estudo Observacional:** não é possível definir grupos específicos para os sujeitos do estudo

- Utilizado quando é possível dividir aleatoriamente os sujeitos do estudo em grupos
- Permite controlar (quase) todos os aspectos que se queiram do estudo
- Permite estabelecer relações de causa e efeito

- Utilizados quando não é possível, por quaisquer razões, dividir os objetos de estudo em grupos
- Não permite controlar todos os aspectos do estudo
- Não permite estabelecer relações de causa e efeito

Cuidados com a Coleta de Dados

- Decidir se o estudo é observacional ou experimental
- Tomar uma amostra representativa
- Tomar uma amostra grande o suficiente

Análise Exploratória, Inferência, Predição

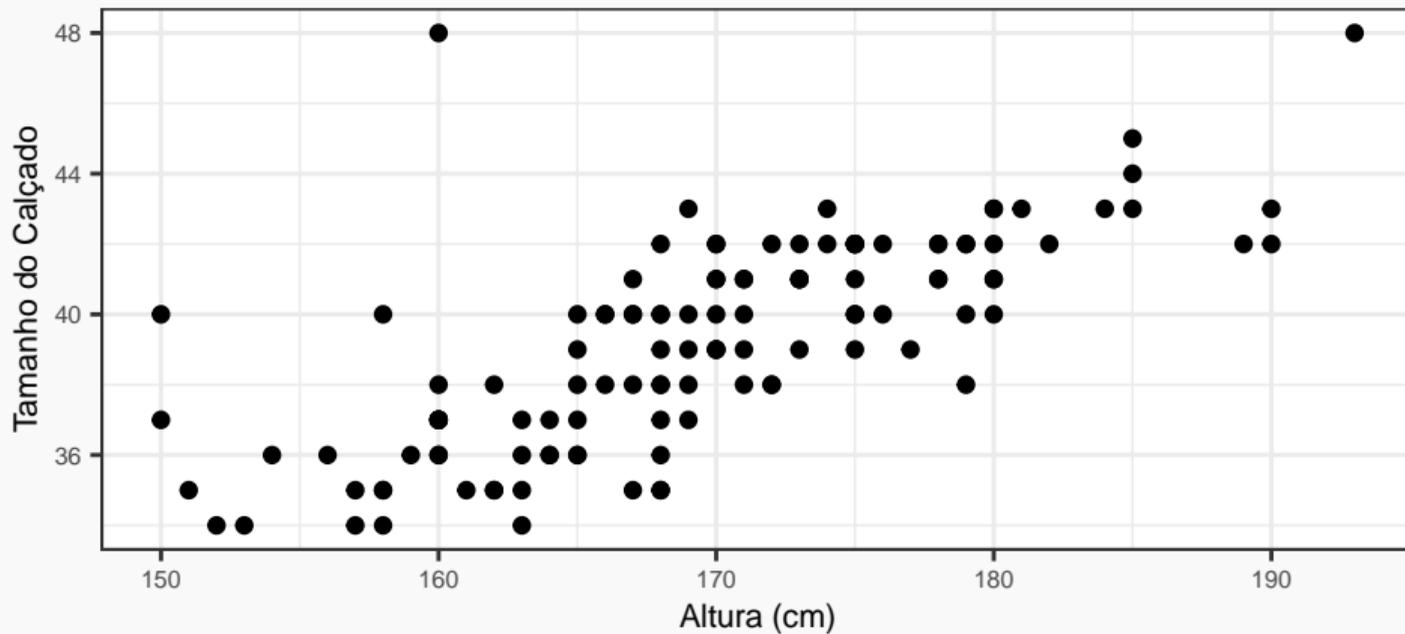
Análise Exploratória, Inferência, Predição

- A escolha do tipo de análise estatística a ser utilizada vai depender dos objetivos que desejamos alcançar com ela
- Há três objetivos principais quando realizamos uma análise de dados:
 - Análise Exploratória
 - Inferência
 - Predição

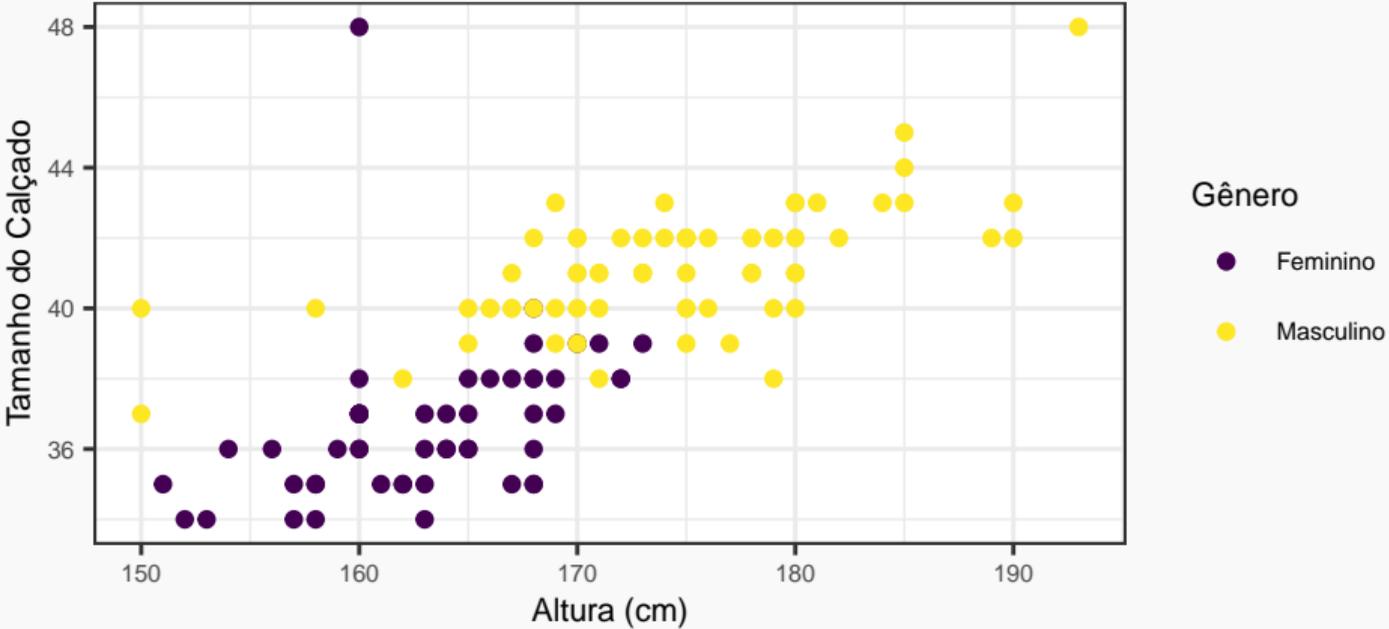
- Nos próximos slides veremos como estes três conceitos se relacionam na prática
- Vou utilizar um conjunto de dados coletado por mim, durante quatro anos, com os alunos ingressantes no Bacharelado em Estatística da UFRN
- Queremos estudar a relação entre tamanho do calçado, altura e gênero dos alunos

- Análise Exploratória é o primeiro passo em qualquer análise de dados
- Resume os dados através de gráficos e tabelas
- Ajudar a verificar o comportamento geral dos dados, criar perguntas e levantar hipóteses

Alunos Ingressantes no Bacharelado em Estatística



Alunos Ingressantes no Bacharelado em Estatística



- Embora nos dê uma ideia do que se passa com os dados, a Análise Exploratória é limitada
- Ela não permite que conclusões a respeito dos dados sejam tiradas
- Desta forma, outras ferramentas devem ser adicionadas ao arsenal do cientista de dados

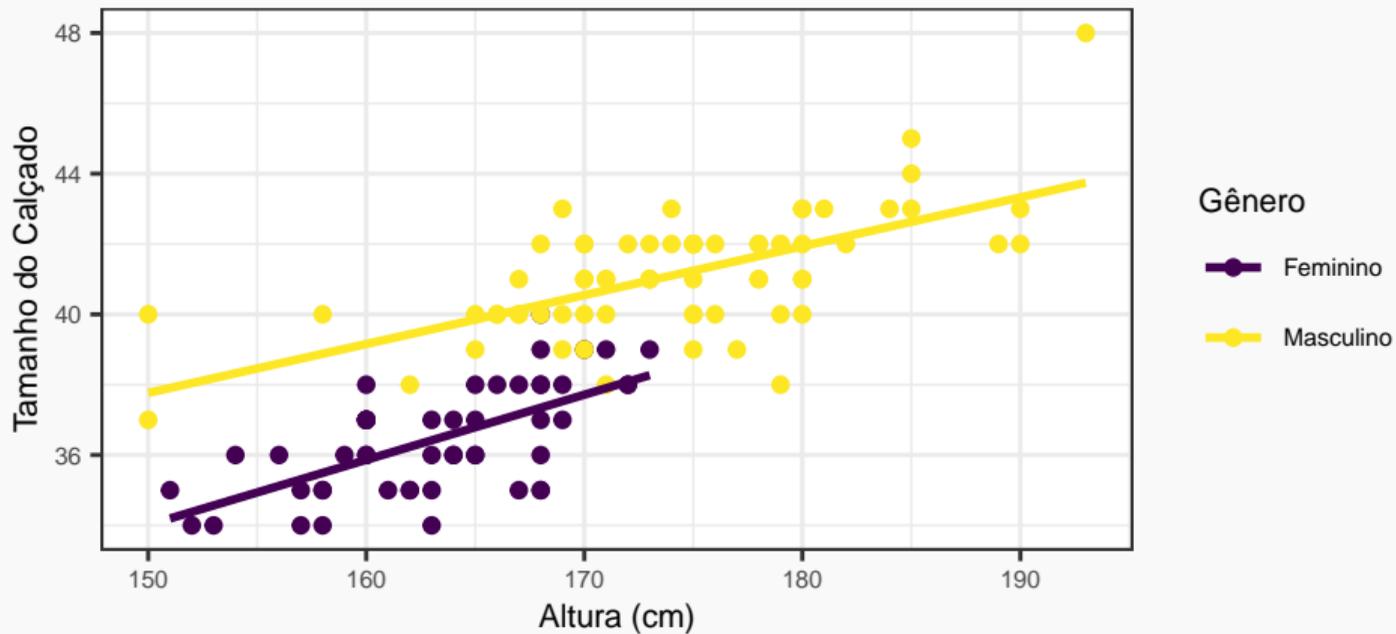
- Inferência é o conjunto de técnicas estatísticas que nos permite tirar conclusões sobre a população de interesse
- Pode ser aplicada na população como um todo ou em uma amostra
- Se aplicada em uma amostra, vem acompanhada de medidas de incerteza

- Uma pergunta que podemos fazer em relação a este conjunto de dados é
Existe diferença na média do tamanho do calçado entre os gêneros feminino e masculino?
- E esta pergunta pode ser tornada mais específica:
Caso exista diferença na média do tamanho do calçado entre os gêneros feminino e masculino, qual é a magnitude dessa diferença?

- Podemos responder estas perguntas ajustando um modelo a estes dados
- No caso específico deste problema, queremos ajustar um modelo de acordo com a equação abaixo

$$\text{calcado} = \alpha + \beta_1(\text{altura}) + \beta_2(\text{genero}_{\text{Masculino}}) + \epsilon \quad (1)$$

Alunos Ingressantes no Bacharelado em Estatística



Gênero	Média	EP	gl	Lim Inf	Lim Sup
Feminino	37.4	0.2	117	37.0	37.8
Masculino	40.4	0.2	117	40.1	40.8

- Com um modelo ajustado aos dados é possível realizar predições
- Fazer uma predição é estimar características de interesse para um sujeito que não estava presente no conjunto de dados original
- Neste contexto que estamos vendo, o objetivo é estimar o tamanho do calçado quando sabemos apenas a altura e o gênero do indivíduo

Predição

Gênero	Altura	Calçado	Lim Inf	Lim Sup
Feminino	155	35	33	38
Masculino	155	38	36	41
Feminino	160	36	33	39
Masculino	160	39	36	42
Feminino	165	37	34	39
Masculino	165	40	37	42
Feminino	170	37	35	40
Masculino	170	40	38	43

Exemplos na UFRN

- Mesmo sem um projeto institucional oficial, a UFRN possui algumas iniciativas para análise de dados
- Veremos algumas destas iniciativas a partir de agora

- Portal de Dados Abertos da UFRN - <https://dados.ufrn.br/>
- Disponibiliza dados e informações públicas da Instituição
- 84 conjuntos de dados em 12 grupos (Biblioteca, Despesas e Orçamentos, Ensino, Pesquisa, Extensão etc.)



PESQUISAR DADOS

Ex.: cursos



Etiquetas mais comuns

requisição

requisições

servidores

Pesquisar conjuntos de dados... 

84 conjuntos de dados encontrados Ordenar por: Relevância 

PESQUISADORES
Relação de pesquisadores da UFRN
 

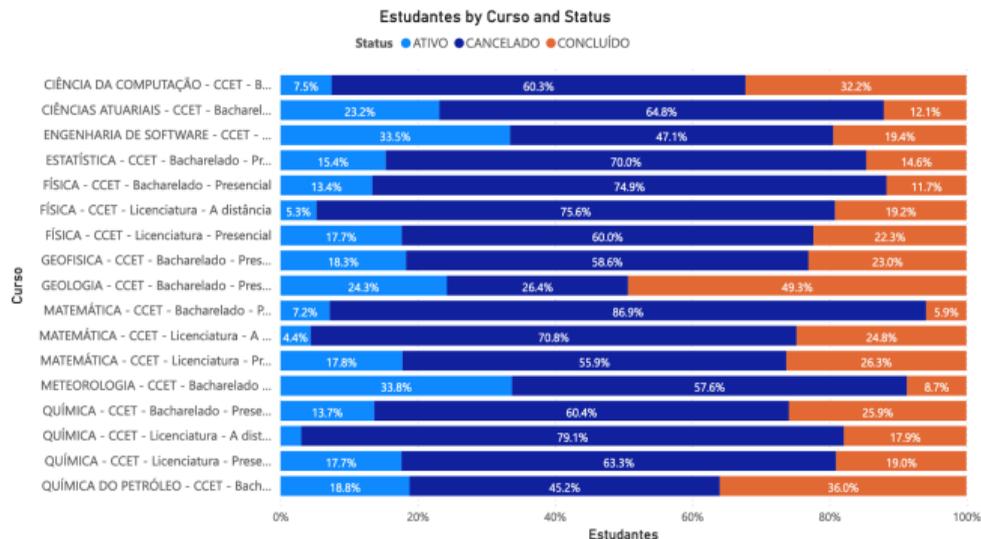
BOLSISTAS DE INICIAÇÃO CIENTÍFICA
Relação de bolsistas de iniciação científica de projetos de pesquisa da UFRN
 

REQUISIÇÃO DE DIÁRIAS
Relação das requisições de diárias da UFRN
  

AUSÊNCIAS DE SERVIDORES
Relação de ausências de servidores da instituição.
 

- Estes dados podem ser usados no desenvolvimento de aplicativos e ações
- Avaliação de professores, livros mais retirados na biblioteca, áreas de pesquisa com mais bolsistas
- Podem ser cruzados com outras fontes de informação

- Programa de Apoio à Melhoria da Qualidade do Ensino de Graduação
- Coordenado pelo Professor Fabiano Gomes, do Instituto de Química e Procurador Educacional Institucional da UFRN



link

Scylax: ferramenta de análise de produção acadêmica

- Scylax: ferramenta de análise de produção acadêmica - <https://scylax.ufrn.br/>
- Ferramenta de exploração da atuação de pesquisadores ou organizações
- Relatórios gráficos e intuitivos
- Produto desenvolvido dentro do Projeto Plataforma para Avaliação de Dados das Atividades de Pesquisa, coordenado pelo Professor Gibeon Soares de Aquino Júnior, do DIMAp

Scylax: ferramenta de análise de produção acadêmica

Adicionar Organizações Educacionais ✕

IES **CENTRO/INSTITUTO** DEPARTAMENTO PROGRAMA GRUPO DE PESQUISA

Universidade

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE ✕

Centro/Instituto

Digite para buscar

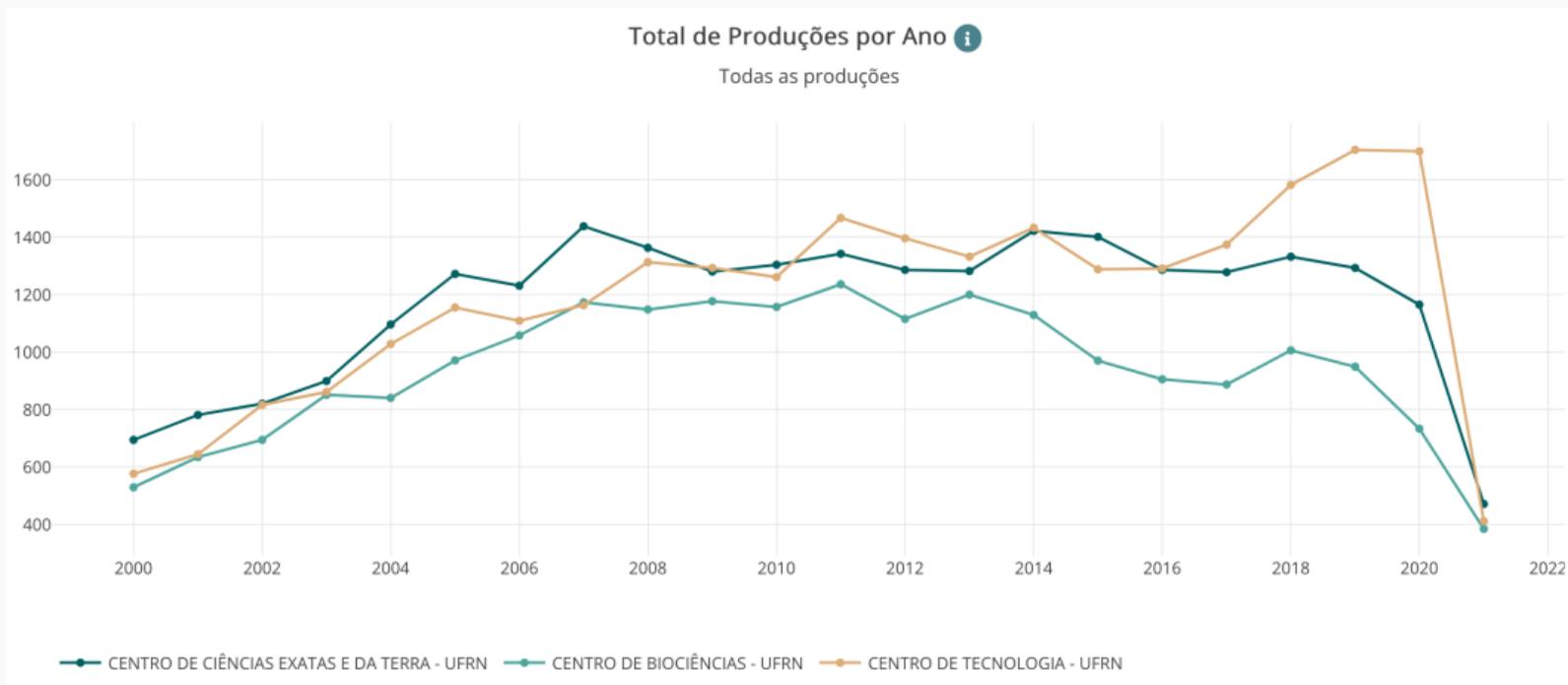
Você ainda pode adicionar 17 organização(ões)

CENTRO DE CIÊNCIAS EXATAS E DA TERRA - UFRN ✕ CENTRO DE BIOCÊNCIAS - UFRN ✕

CENTRO DE TECNOLOGIA - UFRN ✕

CANCELAR ADICIONAR

Scylax: ferramenta de análise de produção acadêmica



link

Scylax: ferramenta de análise de produção acadêmica

- Permite verificar o histórico de pesquisadores e organizações
- Facilita a coleta de dados quando comparada com o lattes
- Democratiza o acesso da informação

- Angelini, Nunes, Nascimento e Carvalho (202?)
- Quais fatores influenciam a produção acadêmica dos professores da UFRN?
- Analisamos dados entre 2014 e 2017

Média	Desvio Padrão
6.92	10.23

Produção Total

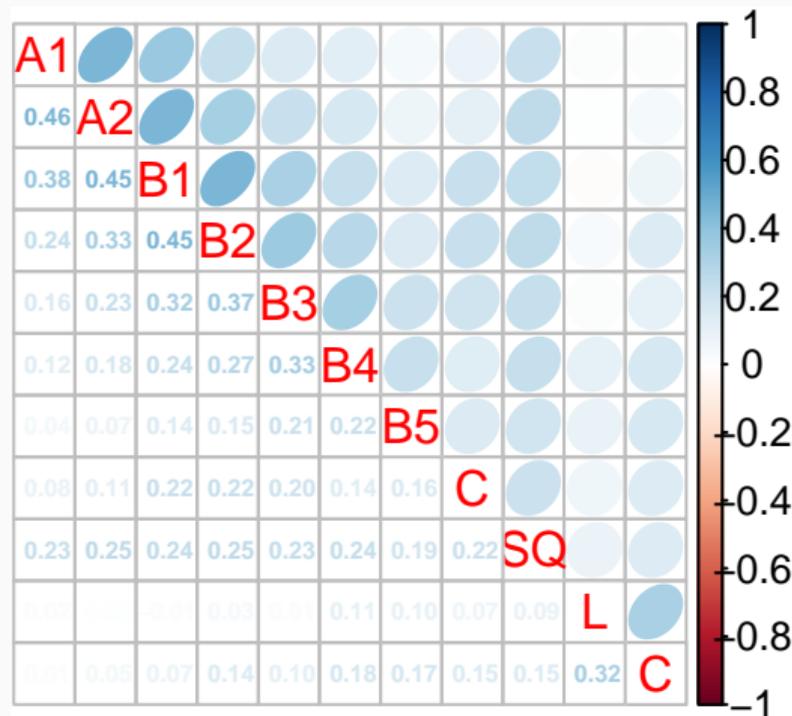
Número de Artigos	Qtde de Professores	Percentual
0	327	18.31
1	227	12.71
2	187	10.47
3	168	9.41
4	126	7.05
5	92	5.15
6	81	4.54
7	59	3.30
8	64	3.58
9	38	2.13
10+	417	23.35

Média	Desvio Padrão
0.7	1.9

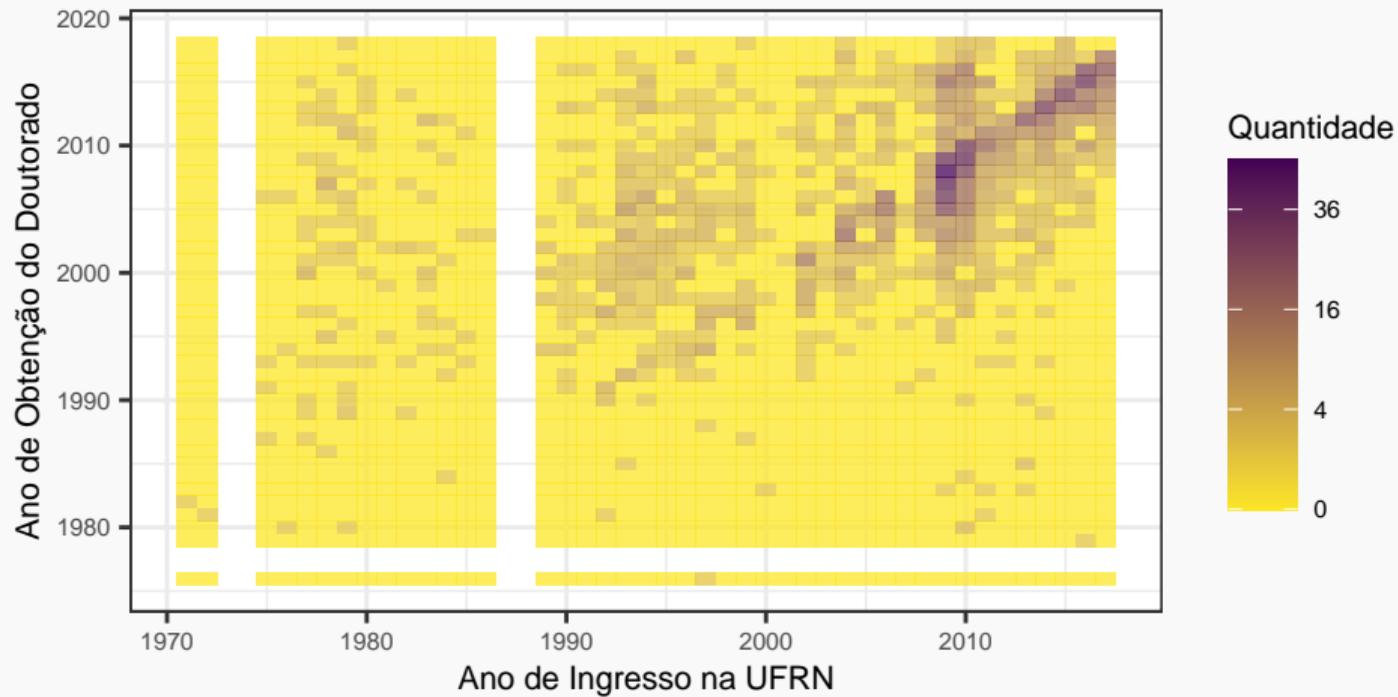
Produção A1

Número de Artigos	Qtde de Professores	Percentual
0	1311	73.40
1	239	13.38
2	84	4.70
3	53	2.97
4	25	1.40
5	28	1.57
6	13	0.73
7	7	0.39
8	5	0.28
9	4	0.22
10+	17	0.95

Correlações entre os Tipos de Produção



Heatmap



Variáveis Utilizadas

- UNG: horas de ensino anuais na graduação
- GRA: horas de ensino anuais na pós-graduação
- SDOC: orientações de doutorado concluídas
- SMAS: orientações de mestrado concluídas
- SIC: orientações de iniciação científica concluídas
- SMON: número de TCCs orientados
- PFU: número de projetos financiados
- PNF: horas dedicadas à pesquisa
- OUT: número de eventos
- BEF: número de artigos publicados antes de obter o doutorado
- YDOC: ano de obtenção do doutorado
- YUFRN: ano de ingresso na UFRN
- GEN: gênero

- Dados discretos, provenientes de contagens
- O natural seria escolher distribuições como Poisson ou Binomial Negativa para o ajuste do modelo
- Entretanto, a análise exploratória nos sugere um excesso de zeros
- Isto nos levou a considerar o ajuste de um modelo hurdle

Modelagem - Produção Total Não-Nula

Preditores	Coefficientes	Média	Desvio.Padrão	p.valor
(Intercept)	5.0415	NA	NA	0.0000
UNG	1.0013	13.9999	6.3341	0.9600
GRA	1.0763	3.4784	4.5456	0.0207
SDOC	1.3162	0.6988	1.7804	0.0000
SMAS	1.2484	2.2447	3.1184	0.0000
SIC	1.1551	2.0666	3.6978	0.0000
SMON	1.0558	3.3645	5.9741	0.0186
PFU	1.1478	3.6534	5.6673	0.0000
PNF	1.0776	21.7013	52.0439	0.0110
OUT	1.0114	5.8018	8.7833	0.6360
BEF	1.4226	3.2156	5.7848	0.0000
YDOC	0.9108	2007.2374	6.7722	0.0083
YUFRN	1.1558	2005.8970	9.7173	0.0000
GENMasculino	1.0695	NA	NA	0.1806

Conclusões

- Os dados da universidade estão aí, mas é necessário fazer perguntas interessantes para analisá-los
- Esta é uma tarefa multidisciplinar: não é possível para uma pessoa sozinha assumir todas as responsabilidades e tarefas envolvidas nela
- A questão não é apenas o tempo e o conhecimento que iniciativas assim demandam, mas também é importante conhecer as demandas das outras áreas



Contato

- Marcus Nunes
- email: `marcus.nunes@ufrn.br`
- site: `https://marcusnunes.me`