

APLICAÇÕES DE RSTUDIO NA APRENDIZAGEM MATEMÁTICA: ELABORAÇÃO DE ARTEFATOS E CIÊNCIA DE DADOS COMO MATEMÁTICA

RStudio applications in Mathematics Learning: Elaboration of artifact and Data Science as Mathematics

*Isadora Luiz Lemes
Renato P. dos Santos*

Resumo

O objetivo deste trabalho é discutir contribuições para a aprendizagem de Matemática da construção de ‘artefatos’, que possam ser mostrados, discutidos, examinados, sondados e admirados, construídos pelo uso do RStudio que aqui é interpretado como um micromundo de Ciência de Dados. Aqui é trazido o conceito da ‘arte de aprender’, que Papert define como Matemática, visando problematizar, também, o quanto os atuais cursos de formação de professores supervalorizam o ensino enquanto carecem de questões que envolvam o aprender. Demonstraremos, aqui, como é possível utilizar o RStudio mesmo com pouco conhecimento em linguagem de programação, que o campo da Estatística possui alternativas aos seus métodos convencionais de ensino e que estas são capazes de desenvolver aptidões através da construção do conhecimento que parte dos próprios estudantes, visto que eles irão definir, manipular e tratar dados que partam de seus interesses.

Palavras-chave: Ensino; Construcionismo; Aprendizagem Matemática; Ciência de Dados; Matemática; RStudio.

Abstract

The objective of this work is to discuss contributions to mathematics learning from the construction of ‘artefacts’ that can be shown,

discussed, examined, probed and admired, built using the RStudio that is taken here as a Data Science microworld. Here is brought the concept of ‘the art of learning’ that Papert defines as Mathetic, aiming to problematize also how much the current teacher training courses overvalues teaching while they lack issues that involve the learning. We will demonstrate here how it is possible to use RStudio, even with little knowledge of programming language, that the field of Statistics has alternatives to its standard methods, and that these may develop skills through the construction of knowledge that starts from the student themselves, since they will define, manipulate, and treat data that are of their interest.

Keywords: Education; Constructionism; Mathematics learning; Data Science; Mathetics; RStudio.

Introdução

Nestes dias em que denúncias de vigilância e a interceptação do tráfego em telefones, e-mails e redes sociais por agências de segurança inflamam as preocupações com a privacidade (GALLAGHER, 2013), é importante capacitar nossos estudantes, futuros profissionais de Ciências e Matemática, tanto em termos técnicos como em éticos, para os desafios científicos propostos por esta realidade, na qual vão exercer suas profissões.

Autores como Davenport e Patil (2012), Eybers e Hattingh (2016) e Saltz e Heckman (2015) concordam que a profissão Cientista de Dados emerge a cada o momento e, segundo Davenport e Patil (2012) “a escassez de cientistas de dados está se tornando uma séria restrição em alguns setores”. Isso coloca uma demanda por ações urgentes que possam modificar este cenário, exigindo que novas abordagens sejam desenvolvidas em sala de aula, para que os futuros profissionais a serem inseridos no mercado de trabalho estejam continuamente preparados para enfrentar os desafios que são constantemente colocados por nossa sociedade.

Antagonicamente, como veremos adiante, os estudantes ainda se veem diante de formações em que os professores estão focados em ensinar conteúdos a seus alunos, não os ensinando a ‘pescar’ por si sós (PAPERT, 1993, p.139).

Mesmo nos dias de hoje, o foco das formações de professores continua sendo no ensino e muito pouco se fala da aprendizagem e até quando se trata desta, observa-se que ela não aparece descolada do ensino. Contudo, Papert já frisava que, para que alguém aprenda, não é imprescindível que este receba instrução, pois, através de exploração livre, partindo dos objetivos e interesses do sujeito, é possível sua própria construção do conhecimento (PAPERT, 1980, p.184, 204).

Por outro lado, como se verá adiante, a linguagem R e o ambiente RStudio são muito versáteis e permitem intensas atividades oferecendo inúmeras possibilidades aos usuários mesmo que estes não possuam intimidade com linguagem de programação:

[...] além de fornecer operações estatísticas, R tem um propósito geral de linguagem de programação que se pode usar para automatizar análises e criar novas funções que ampliam os recursos de linguagem existentes (MATLOFF, 2011, p.2, tradução nossa).

Acreditamos que a utilização destes recursos impactaria no processo da construção do conhecimento e partiria dos fundamentos da Matemática de Papert (1993, p.84), pois, neste ambiente, os alunos produziram seus próprios artefatos dentro de um micromundo (PAPERT,

1980, p.7–19), aprendendo sem necessidade expressa de um processo transmissivo.

Neste trabalho, procuramos demonstrar os micromundos e artefatos de que estamos falando e, trazemos problemas já anteriormente discutidos pelos autores em outras publicações, com o fito de demonstrar o quanto pode ser praticável a realização destas atividades através de exemplos concretos que podem ‘dar vida’ às tantas abstrações Matemáticas que são vivenciadas em todos os campos desta área, inclusive na Estatística, que podem, assim, ser abordadas com maior interesse, visando atender às necessidades que se mostram cotidianamente em nosso sistema coletivo.

Do Construtivismo à Ciência de Dados

Já em 1993, Papert afirmava que, embora a literatura científica se abrigue sob a denominação “Teorias de Aprendizagem”, as abordagens apresentadas raramente se referem a orientar os professores e pesquisadores sobre ‘como aprender’, e resultam, na prática, em manuais de ‘boa instrução’ (PAPERT, 1993, p.83). Mesmo nos dias de hoje, a literatura corrente em Ensino e Aprendizagem preocupa-se mais em discutir estratégias de ensino do que ajudar o indivíduo a ter melhor aprendizado. De fato, uma rápida pesquisa no Google Acadêmico indica 141.000 títulos contendo a palavra ‘ensino’, excluindo variações da combinação genérica ‘ensino e aprendizagem’ (com cerca de 30 mil entradas), contra 15.800 para apenas ‘aprendizagem’.

Com isso, não é de se espantar que esses em formação busquem ‘novas’ tecnologias, mais com a preocupação de ‘ensinar melhor’ seus alunos do que ajudá-los a aprender, coerentemente à formação recebida, a qual não valorizou a Matemática.

Papert argumentava que os métodos considerados importantes na Educação são voltados a como o ensino deve ser desenvolvido e como tornar-se um professor capacitado, enquanto que não se encontra a preocupação em desenvolver métodos de aprendizagem, para tornar-se um aprendiz capaz, (PAPERT, 1993, p.82). Segundo esse autor, “a Pedagogia como a arte de ensinar foi adotada pelo mundo acadêmico como uma área respeitável e importante”, enquanto que a ‘arte de aprender’ “é uma órfã acadêmica” (1993, p.82).

Ainda segundo Papert, na formação dos professores, é muito valorizado que eles saibam como ensinar e pouco se fala que eles devem também saber aprender (1993, p.82). Para esse autor há um “tratamento desigual das artes de aprender e ensinar”, de que “mesmo os defensores da educação ‘ativa’ ou ‘construtivista’ têm dificuldade em escapar”, o que acaba por reforçar a hierarquização dos papéis de professores e alunos, onde o professor, “aquele que sabe”, ensina e a criança é ensinada, sem considerar que todos devem dominar a arte de aprender e que esta, igualmente, tem de ser aprendida (1993, p.83).

Já em 1638, Comenius, considerado o pai da Educação moderna, prenunciava essa arte de aprender, quando falava que “a proa e a popa da nossa *Didática* será investigar e descobrir o método segundo o qual os professores ensinam menos e os estudantes aprendam mais” (COMENIUS, 1985).

Percebendo a necessidade de um nome para designar essa arte de aprender, Papert propôs o termo *Matética*, inspirado pela origem do termo *Mathēmatikos*, com o significado de ‘disposto a aprender’ (1993, p.84):

O que é Matética aqui é a mudança de foco, desde o pensar sobre se as próprias regras são eficazes na aplicação imediata, para procurar múltiplas explicações sobre como trabalhar com as regras pode contribuir, a longo prazo, para a aprendizagem. Para salientar o argumento de uma forma certamente exagerada, sugiro que qualquer tipo de “brincar com problemas” melhorará as competências subjacentes a sua solução (PAPERT, 1993, p.87).

Papert considerava que buscar conexões fazia parte do ato de aprender intencionalmente e que, a partir de “entidades mentais já existentes, novas entidades mentais” passariam a existir de maneiras sutis e inconscientes (1993, p.105) e que

[...] esse conhecimento cultural é tão interconectado que a aprendizagem propagar-se-á por migração livre para todas as suas áreas. Isso sugere uma estratégia para facilit-

tar a aprendizagem, melhorando a conectividade no ambiente de aprendizagem, por meio de ações sobre culturas e não sobre indivíduos (PAPERT, 1993, p.105).

Por outro lado, o documento da Base Nacional Comum Curricular (BNCC) (Brasil, 2017) traduz a forte importância de efetivar atividades que envolvam a interpretação e leitura de dados, gráficos, tabelas, etc., defendendo

[...] a abordagem de conceitos, fatos e procedimentos presentes em muitas situações-problema da vida cotidiana, das ciências e da tecnologia. [...] todos os cidadãos precisam desenvolver habilidades para coletar, organizar, representar, interpretar e analisar dados em uma variedade de contextos, de maneira a fazer julgamentos bem fundamentados e tomar as decisões adequadas. (BRASIL, 2017, p.230).

Assim, acreditamos que a Ciência de Dados, por ser intensamente abrangente e implicar na necessidade de conhecimentos diversos, em áreas muito variadas, poderia estimular os estudantes a explorar, tratar, analisar e interpretar dados existentes, propiciando que conexões se estabeleçam em sua mente e contribuindo para a formação de um pensamento mais crítico, o que vem muito ao encontro da realidade de nossa sociedade atual.

Do que foi discutido acima, mais do que apenas mais uma ‘nova’ tecnologia educacional, vemos o *aprender-com-Ciência-de-Dados* como uma nova maneira de construir o conhecimento científico, de aprender a fazer Ciência, de aprender a pensar como um cientista.

RStudio como Micromundo de Ciência de Dados

Há, atualmente, muitos ambientes e programas de cálculo estatístico¹, tais como Tableau, GNU Octave, Python/Jupyter Notebook e R/RStudio, dentre outros. Ao contrário dos bem conhecidos SPSS, SAS, MATLAB e Mathemati-

¹ Uma lista não extensiva pode ser encontrada em https://en.wikipedia.org/wiki/List_of_statistical_packages

ca, com licenças bastante dispendiosas, vários destes são abertos e gratuitos ou têm licenças especiais para utilizações educacionais. Embora Python seja frequentemente apontada como uma forte concorrente a R, trata-se de uma linguagem de programação geral, com extensa aplicação no ensino de programação na educação básica, como se pode ver, p.ex., em Marques et al. (2011) e Scaico et al. (2012). Dentre esses ambientes, optamos por trabalhar com a linguagem R e o ambiente RStudio, por nossa familiaridade com esse ambiente.

R (R CORE TEAM, 2017) tem a filosofia de fornecer uma linguagem conveniente para o trabalho interativo e, ao mesmo tempo, ser uma linguagem de programação poderosa para o desenvolvimento de novas ferramentas. Isso permitiu a criação de uma comunidade muito ativa e vibrante, com milhares de usuários em todo o mundo, que desenvolvem novos pacotes de recursos, disponíveis gratuitamente nos repositórios CRAN² e GitHub³, e ajudam uns aos outros a usar R através de sites como o Stack Overflow⁴ (PENG, 2016, p.7), um fórum pesqui-

sável contendo dezenas de milhares de perguntas respondidas sobre R, RStudio e seus pacotes.

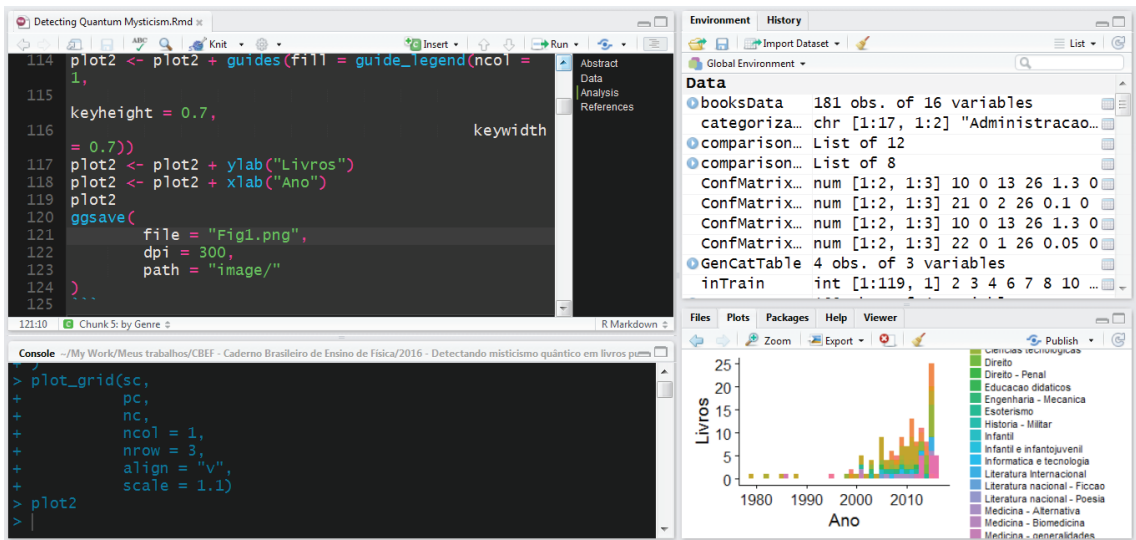
Uma vantagem importante que R tem sobre muitos outros pacotes estatísticos é sua superior capacidade em criar gráficos avançados com “qualidade de publicação”, além de permitir visualizações complexas e sofisticadas de dados em 3D (PENG, 2016, p.7).

Embora R se distinga de muitos outros softwares estatísticos por dispor de um *console*⁵, um meio para emitir comandos de forma interativa, para facilitar esse trabalho, muitas linguagens interativas de programação maduras incluem ambientes de desenvolvimento integrado (IDE), dispendo de interfaces gráficas de usuário (GUI) amigáveis (VERZANI, 2011, p.2), que, no caso de R, é o *RStudio* (RSTUDIO TEAM, 2017).

Basicamente, o RStudio apresenta quatro áreas, visíveis na Figura 1, em sentido anti-horário:

1. O console, onde se digitam comandos de forma interativa e se obtém resultados e mensagens de erro, as quais proporcio-

Figura 1 – Interface Gráfica do RStudio



Fonte: Os autores.

² <http://cran.r-project.org>

³ <http://github.com/>

⁴ <http://stackoverflow.com/questions/tagged/r>

⁵ Por analogia à ‘estação central de controle’, parecida com um teletipo, que ocupava posição de destaque na sala climatizada onde ficam os grandes computadores (*mainframe*) e dava total controle ao operador ao digitar comandos nela como numa máquina de escrever.

nam o necessário e útil feedback para o aprendizado.

2. O código armazenado, a sequência de comandos que, uma vez testada e validada, pode ser armazenada em forma de um ‘programa’, possivelmente contendo ‘funções’, que pode ser reutilizado muitas vezes, com menor chance de erros de digitação.
3. O ‘ambiente’, exibindo as variáveis e bases de dados disponíveis, indicando seus tipos de dados e dimensões.
4. Os gráficos gerados como resultado da análise dos dados realizada pelos comandos.

No entanto, qualquer uma dessas janelas pode ser expandida para tela cheia, caso o usuário deseje trabalhar com mais conforto.

R e RStudio possuem recursos acessíveis, mas poderosos o suficiente para se manipular facilmente dados de variados formatos, obtidos de diferentes fontes, mesmo com volumes da ordem de milhões de dados, o que pode torná-los valiosos para trabalhar a Ciência de Dados com os alunos dos níveis de ensino básico e superior, visando prepará-los para a realidade que os aguarda.

No entanto, para Papert, a Matemática, a ‘arte de aprender’, que corresponde ao estudante aprender a pescar, na conhecida metáfora, não é suficiente. Para esse autor, é necessário também ter “boas linhas de pesca”, as ferramentas – que no contexto da época, ele praticamente restringiu aos computadores – e “águas ricas”, ambientes dispostos de grande variedade de atividades matematicamente ricas, ou seja, seus ‘micromundos’. (PAPERT, 1993, p.139).

Os micromundos⁶ diferem dos objetos de aprendizagem, *a partir dos quais* o aluno aprende, e das simulações, que são apenas interativas, por serem ambientes intelectuais cuja ênfase está no processo de livre exploração e manipulação dos recursos e objetos de um universo regrado que ‘incorpora’ um pequeno subdomínio da Matemática ou das Ciências e que é imediatamente compreensível e também intrinsecamente motivador para ele (PAPERT, 1980, p.184, 204).

Note-se, porém, que, com a ampliação dos domínios de conhecimento sobre os quais os micromundos foram projetados, essa restrição a computadores e linguagens de programação no Construcionismo desapareceu. (HEALY; KYNI-GOS, 2009). Hoje, se aceita como micromundo qualquer ambiente que incorpore as características mencionadas de se constituir num universo regrado que

- incorpore um pequeno subdomínio da Matemática ou das Ciências,
- permita a livre exploração e manipulação dos recursos e objetos existentes,
- seja imediatamente compreensível e intrinsecamente motivador para seus usuários e
- disponha de grande variedade de atividades matematicamente ricas.

Por outro lado, Papert, em seu Construcionismo, acreditava que a aprendizagem natural, piagetiana (1980, p.7), de “construção de estruturas do conhecimento”, que ocorre ‘na cabeça’, ocorreria com maior proveito num ambiente, alicerçado em computadores (1993, p.142), que dispusesse de *objetos-de-pensar* e *objetos-de-pensar-com* (1980, p.11), objetos de que as crianças pudessem “se apropriar à sua própria maneira” (1980, p.19) e, a partir deles, engajarem-se conscientemente na construção ‘no mundo’ do que Papert chama de *entidades públicas, artefatos que possam ser mostrados, discutidos, examinados, sondados e admirados*, sejam elas “um castelo de areia, um bolo, uma casa de Lego, uma corporação, um programa de computador, um poema ou uma teoria do universo”. (PAPERT, 1993, p.142, grifo nosso).

Como veremos adiante, esses recursos poderosos, mas acessíveis, de R de manipulação de dados e gráficos permitem aos estudantes a construção de artefatos, na forma de gráficos sofisticados e visualizações complexas, que podem ser exibidos com orgulho aos colegas, discutidos e analisados.

Desta forma, entendemos que o aplicativo *RStudio* constitui-se num micromundo de Ciência de Dados, por ser um ambiente amigável, dispor de inúmeros recursos poderosos e acessíveis e tornar acessíveis a livre exploração, manipulação e visualização de conjuntos de dados, permitindo ao mesmo tempo, tornar

⁶ Para uma revisão atualizada, ver Noss e Hoyles (2017) (NOSS E HOYLES, 2017).

mais concreto e personalizado o conhecimento “antes restrito a processos formais e abstratos da Matemática e da Estatística”, conforme já discutido por estes autores (LEMES; DOS SANTOS, 2017).

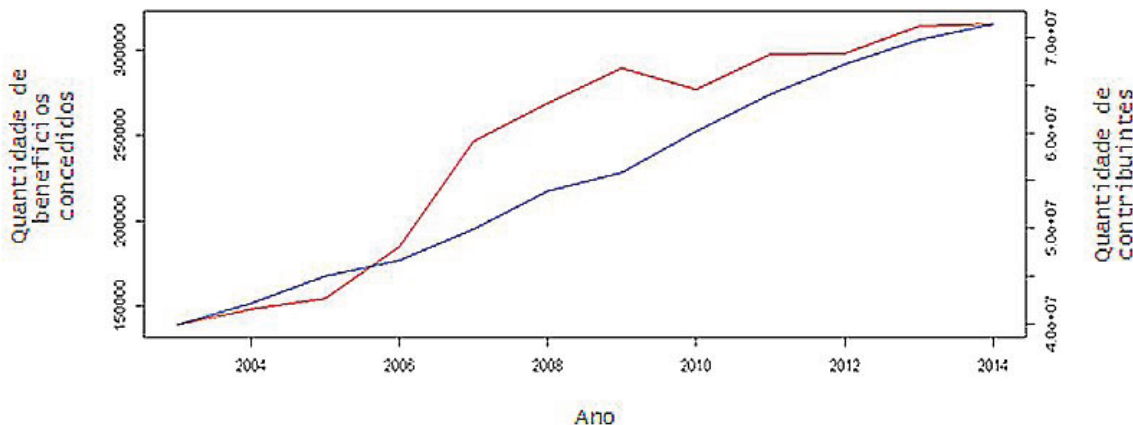
Discussão

Em outro trabalho, estes autores (2017) já discutiram algumas implicações da Ciência de Dados e da interpretação de gráficos matemáticos, a partir de dados extraídos de bases de dados diversas, explorados pelos próprios alunos, para a aprendizagem de Matemática, propiciando a concretização e a personalização do conhecimento antes restrito a processos formais e abstratos. Aqui, traremos alguns exemplos de aplicação do RStudio como um micromundo de Ciência de Dados na Educação Matemática, explorando dados públicos abertos. Esta questão de dados públicos abertos será discutida em outro trabalho (LEMES; DOS SANTOS, 2018). Também não serão apresentadas aqui as necessárias etapas iniciais de preparação das bases de dados no RStudio, já discutidas por estes autores em outro trabalho (2017).

Em vez de apresentar exemplos conjecturais de aplicação desta estratégia, traremos aqui problemas concretos, já tratados pelos autores, com o intuito de demonstrar de forma categórica a viabilidade de sua execução. De fato, graças à enorme quantidade de pacotes já disponíveis, nada há nestes exemplos que não pudesse ter sido realizado em sala de aula com alunos minimamente proficientes em R.

Como primeiro exemplo, extraído de Lemes e dos Santos (2017), analisaram-se dados obtidos do Portal Brasileiro de Dados Abertos⁷ que conta com mais de 3.000 bases de dados, referentes a temáticas variadas, incluindo Turismo, Qualidade de vida, Educação, entre outras. As bases de dados escolhidas foram referentes a *Estatísticas de Contribuição de Pessoas Físicas por Unidade de Federação e Aposentadorias Concedidas por anos de serviço*. Após a escolha das informações de interesse e agregação das bases pela variável comum ‘ano’, apresenta-se, na Figura 2, o resultado em forma de gráfico de linhas, em que se observa uma certa correlação entre as variáveis *Quantidade de benefícios concedidos* e *Quantidade de contribuintes*, o que pode propiciar interessantes discussões em sala de aula.

Figura 2 – Gráfico de linha relacionando as variáveis Quantidade de Benefícios (Aposentadorias) Concedidos e Quantidade de contribuintes



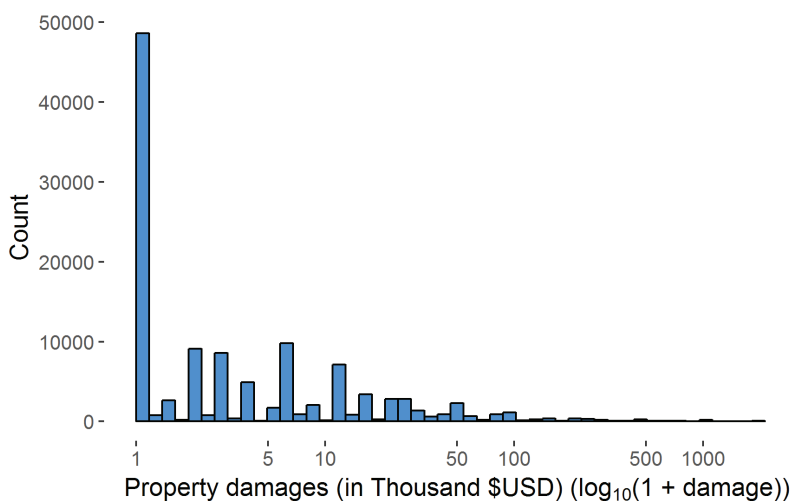
Fonte: Lemes e dos Santos (2017).

⁷ <http://dados.gov.br>

Em seguida, apresenta-se um gráfico de barras (Figura 3), referente a um estudo de dos Santos (2016) sobre a base de dados de grandes tempestades e eventos climáticos (*Storm Events Database*) da NOAA (*U.S. National Oceanic and Atmospheric Administration*), exibindo contagem de eventos por total de danos causados por tempestades. Uma vez que foi usada uma escala logarítmica, os valores de dano tiveram que ser aumentados em US \$ 1,00,

seguindo a prática usual, já que o logaritmo de zero é definido matematicamente apenas como um valor assintótico. Tendo isso em conta, observa-se que os valores reportados de propriedade e danos à cultura se concentram em US \$ 0,00, o que a análise de dos Santos (2016) identificou como muito mais devida a um persistente problema de extensa omissão de dados na base da NOAA do que a danos efetivamente nulos.

Figura 3 – Gráfico de barras exibindo contagem de eventos por total de danos causados por tempestades, em escala logarítmica.



Fonte: dos Santos (2016).

Na sequência, mostra-se um mapa temático coroplético bivariado (*multicolour bivariate choropleth map*)⁸, extraído de dos Santos e Lemes (2017), em que se analisa a correlação entre resultados médios por município no Ideb⁹ e o valor, calculado para cada um dos 5.564 municípios brasileiros, do Índice de Atendimento de Serviços Coletivos Urbanos (D4), parte do Índice de Bem-Estar Urbano (IBEU) desenvolvido

pelo Observatório das Metrôpoles, que avalia o adequado fornecimento de água e eletricidade, atendimento de esgoto e coleta de lixo. Observa-se que, na maioria dos municípios, ocorre a situação esperada de que o atendimento de serviços coletivos é deficiente e simultaneamente o Ideb também é baixo (verde), ou atendimento é adequado e o Ideb também é alto (púrpura). No entanto, surpreendentemente, em muitos municípios, especialmente das regiões Norte e Centro-Oeste, ocorre de o Ideb ser alto com fornecimento deficiente de eletricidade (azul), enquanto que, em alguns poucos municípios, o Ideb é baixo, mesmo com fornecimento adequado de eletricidade (vermelho). Desta forma, embora a relação visual observada sugira uma correlação positiva entre o Ideb e o atendimento de serviços coletivos, parece haver algum fator regional que não foi levado em conta na construção do IBEU.

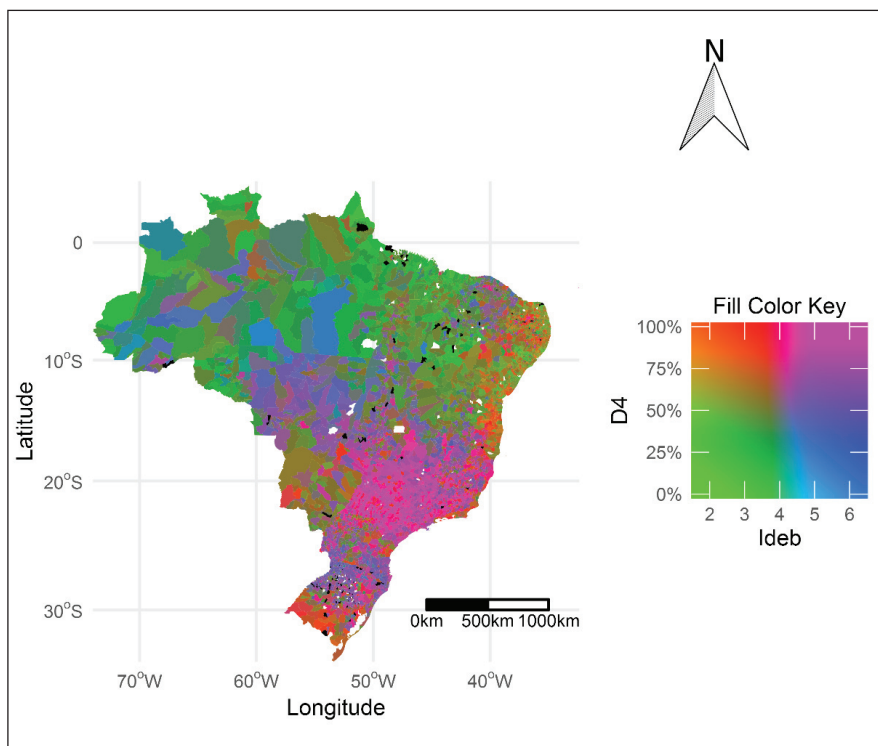
⁸ Mapas coropléticos bivariados multicolores (MURPHY, 2016) superpõem duas quantidades de interesse, de acordo com uma escala específica transcrita visualmente por meio padrões de cores, às respectivas áreas geográficas de ocorrência. As correlações positivas entre elas são indicadas pelas cores que variam de verde (as duas variáveis têm valores baixos) ao violeta (ambos altos), enquanto as cores vermelho e azul indicam correlações negativas (uma é alta e a outra baixa, ou vice-versa).

⁹ Disponíveis no site do Inep: <http://portal.inep.gov.br/web/portal-ideb/planalhas-para-download>

Além disso, ela também parece confirmar o “efeito vizinhança” apontado por de Queiroz Ribeiro (2005), no sentido de que a distribuição desigual das condições urbanas, devido ao crescimento

descontrolado da população e das cidades no Brasil, reduz as oportunidades educacionais e engendra desigualdades sociais incompatíveis com uma sociedade sustentável.

Figura 4 – Mapa temático do tipo coroplético correlacionando resultados no Ideb e Índice de Atendimento de Serviços Coletivos Urbanos (D4) por município.



Fonte: dos Santos e Lemes (2017).

Em outro exemplo, foram utilizados os ‘Microdados do Censo da Educação Básica 2015’¹⁰ e os ‘Microdados do ENEM 2015’¹¹. A Figura 5 apresenta um exemplo de gráfico de violino¹², relacionando as proficiências médias por escola em Ciências Humanas no ENEM e o Indicador de Nível Socioeconômico de seus

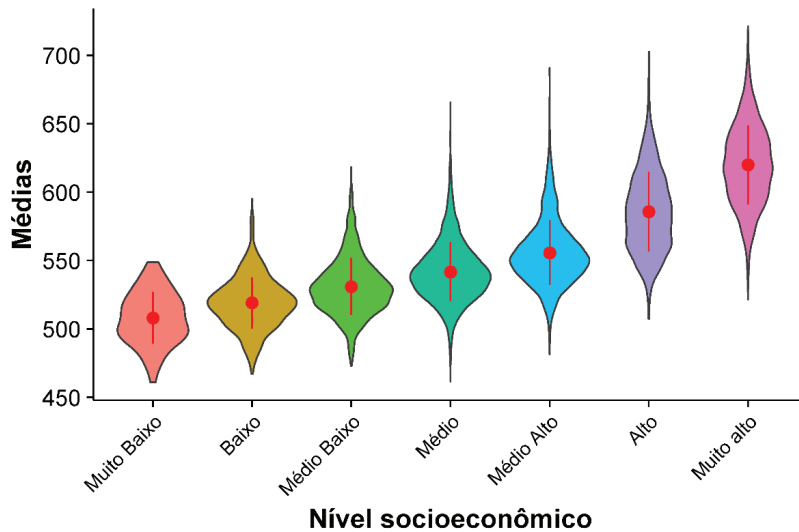
alunos matriculados no 3º ano do Ensino Médio regular, extraído de dos Santos e Lemes ([S.d.]). A análise dos dados mostrou que o fator com que as proficiências nas áreas de conhecimento abrangidas pelo ENEM 2015 melhor se correlacionam foi o Indicador de Nível Socioeconômico dos alunos, com correlações variando de 0.67 a 0.80, resultados estatisticamente significativos ($p=0,00$), considerando o tamanho da amostra ($n=14.976$). Os resultados da análise de dos Santos e Lemes ([S.d.]) indicaram, não surpreendentemente, que as melhores performances corresponderam a alunos com nível socioeconômico muito alto, em escolas com administração federal ou privada, de menor porte e com professores mais bem formados.

¹⁰ Obtidos do site do Inep <http://portal.inep.gov.br/microdados>

¹¹ Também obtidos do site do Inep: http://download.inep.gov.br/microdados/microdados_enem2015.zip

¹² Gráficos de violino combinam dois gráficos rotacionados a 90°, espelhados, da densidade de probabilidade da variável com um marcador para a média e uma indicação do desvio padrão dessa mesma variável (HINTZE; NELSON, 1998) (HINTZE; NELSON, 1998).

Figura 5 – Gráfico de violino relacionando desempenhos médios por escola em Ciências Humanas no ENEM e o Indicador de Nível Socioeconômico dos alunos

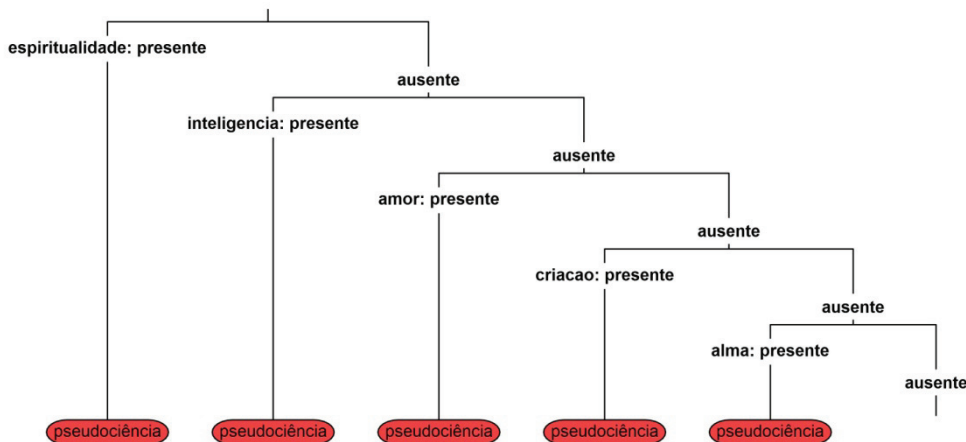


Fonte: dos Santos e Lemes ([S.d.]).

No próximo exemplo, Moura e dos Santos (2017) extraíram títulos de livros contendo as palavras ‘quântica’ ou ‘quântico’ dos catálogos on-line das quatro maiores livrarias do Brasil e os analisaram com recursos de mineração de dados textuais e aprendizagem de máquina (*machine learning*). A Figura 6 traz parte de uma árvore de regressão do modelo que identificou 22 termos,

com uma acurácia de 94% de discriminação entre as categorias. Com isto, um visitante de uma livraria, tendo em mãos um livro que supostamente se refira a ‘quântica’, antes mesmo de lê-lo, caso identifique qualquer uma dessas palavras (espiritualidade, inteligência, amor, criação, alma, etc.) em seu título, poderá assumir, com grande segurança, que ele trata de ‘misticismo quântico’ (pseudociência) e não de ciência.

Figura 6 – Parte da árvore de regressão de classificação de livros através de palavras em seus títulos



Fonte: Moura e dos Santos (2017).

Conclusão

Acreditamos ter mostrado como o RStudio pode ser interpretado como um micromundo de aprendizagem de Matemática e Estatística. Os estudantes podem apropriar-se dos recursos desse ambiente e utilizá-los para fazer alguma coisa nova com eles, brincar com eles e construir com eles (PAPERT, 1980, p.120).

Acreditamos, também, ter mostrado como, através do uso do RStudio, os estudantes podem se pôr em contato com a realidade complexa dos interesses e necessidades reais e atuais das pessoas e, ao mesmo tempo, o professor brasileiro de Educação Matemática pode identificar temas de interesse para o mundo real de seus estudantes. Certamente, haverá outras formas de se utilizar este aplicativo no ensino, cabendo ao professor criativo identificá-las e desenvolvê-las.

Mais ainda, acreditamos que a Ciência de Dados tem grandes potencialidades na aprendizagem de Ciências e Matemática, numa abordagem Construcionista, em que o aluno tem a oportunidade de construção ‘no mundo’ de seus artefatos, em forma dos resultados de suas pesquisas, de gráficos e explicações, que podem ser mostrados aos colegas, discutidos, examinados, sondados e até admirados por seus pares, como interessantes e relevantes, de acordo com a visão acima de Papert (PAPERT, 1993, p.142–143), contribuindo para a construção ‘na mente’ de seu conhecimento.

Neste sentido, redimensionar as formações dos futuros professores poderia representar uma solução para que eles tenham condições de abordar o aprender um pouco desconectado do ensinar e, além disso, compreender que é necessário que também saibam dominar a arte de aprender.

Trata-se não apenas de desenvolver ‘novas’ abordagens para as aulas de Matemática, mas propiciar que os alunos vivenciem essa arte de aprender autonomamente, por exploração própria, partindo de assuntos que sejam de seu genuíno interesse, em maior consonância com as necessidades atuais.

Agradecimentos e Apoios

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/CAPES pela bolsa de mestrado.

Referências

COMENIUS, Ioannes Amos. *Didáctica Magna (orig. publ. em 1649)*. Tradução Joaquim Ferreira Gomes. 3. ed. Lisboa: Fundação Calouste Gulbenkian, 1985.

DAVENPORT, Thomas H.; PATIL, D.J. Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, out. 2012. Disponível em: <<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>>. Acesso em: 19 jun. 2017.

DE QUEIROZ RIBEIRO, Luiz César. Segregação residencial e segmentação social: o “efeito vizinhança” na reprodução da pobreza nas metrópoles brasileiras. *Cadernos Metrópole*, n.13, p.47–70, 2005.

DOS SANTOS, Renato P. Some comments on the reliability of NOAA’s Storm Events Database. *arXiv*, n.1606.06973, 22 jun. 2016. Disponível em: <<http://arxiv.org/abs/1606.06973>>. Acesso em: 23 jun. 2016.

DOS SANTOS, Renato P.; LEMES, Isadora Luiz. Impacto do ensino de Filosofia e Sociologia no desempenho escolar: uma análise a partir dos microdados do censo escolar do Inep. *em preparação*.

DOS SANTOS, Renato P.; LEMES, Isadora Luiz. Urban Welfare: Going beyond Socioeconomic Status in School Inequalities. In: 8. DEUTSCH-BRASILIANISCHES SYMPOSIUM FÜR NACHHALTIGE ENTWICKLUNG, 2.-7. Oktober 2017. *Proceedings...*Porto Alegre: PUCRS, 2017.

EYBERS, Sunet; HATTINGH, Marie. Teaching Data Science to Post Graduate Students: A Preliminary Study Using a “F-L-I-P” Class Room Approach. 6 dez. 2016, [S.l.]: International Association for the Development of the Information Society, 6 dez. 2016. p.189–196.

GALLAGHER, Sean. What the NSA can do with “big data”. *Ars Technica*, 11 jun. 2013.

HEALY, Lulu; KYNIGOS, Chronis. Charting the microworld territory over time: design and construction in mathematics education. *ZDM Mathematics Education*, v.42, n.1, p.63–76, 2009.

HINTZE, Jerry L.; NELSON, Ray D. Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, v.52, n.2, p.181–184, maio 1998.

LEMES, Isadora Luiz; DOS SANTOS, Renato P. Ciência de Dados mediada pelo software RStudio como Tecnologia na Aprendizagem de Ciências e Matemática. *em preparação*, 2018.

LEMES, Isadora Luiz; DOS SANTOS, Renato P. Utilização do software RStudio como um micromundo de Ciência de Dados no ensino de Matemática e Estatística. In: VII CIEM – CONGRESSO

INTERNACIONAL DE ENSINO DE MATEMÁTICA, Canoas, 4-7/10/2017. Canoas: ULBRA – Universidade Luterana do Brasil, 2017.

MARQUES, Diego Lopes et al. Atraindo Alunos do Ensino Médio para a Computação: Uma Experiência Prática de Introdução a Programação utilizando Jogos e Python. In: XXII SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, Aracaju, 2011. *Anais* p.1138–1147. Aracaju: SBC – Sociedade Brasileira de Computação, 2011. Disponível em: <<http://www.br-ie.org/pub/index.php/wie/article/view/1954>>. Acesso em: 21 out. 2017.

MATLOFF, Norman. *The Art of R Programming: A Tour of Statistical Software Design*. San Francisco: No Starch Press, 2011.

MOURA, Mairus Disconzi De; DOS SANTOS, Renato P. Detectando misticismo quântico em livros publicados no Brasil com Ciência de Dados. *Caderno Brasileiro de Ensino de Física*, v.34, n.3, dez. 2017.

MURPHY, William. *colorplaner: A “ggplot2” Extension to Visualize Two Variables per Color Aesthetic Through Color Space Projections*. . [S.l: s.n.]. Disponível em: <<https://cran.r-project.org/package=colorplaner>>. Acesso em: 8 out. 2016, 2016

NOSS, Richard; HOYLES, Celia. Constructionism and Microworlds. In: DUVAL, ERIK; SHARPLES, MIKE; SUTHERLAND, ROSAMUND (Org.). *Technology Enhanced Learning: Research Themes*. p.29–35. Cham: Springer International Publishing, 2017.

PAPERT, Seymour A. *Mindstorms – Children, Computers and Powerful Ideas*. New York: Basic Books, 1980.

PAPERT, Seymour A. *The Children’s Machine: Bringing the Computer Revolution to Our Schools*. New York: Basic Books, 1993.

PENG, Roger D. *R Programming for Data Science*. Victoria, CA-BC: Leanpub, 2016.

R CORE TEAM. *R: A language and environment for statistical computing*. Disponível em: <<https://www.r-project.org/>>. Acesso em: 21 abr. 2017.

RSTUDIO TEAM. *RStudio: Integrated Development for R*. Disponível em: <<http://www.rstudio.com/>>. Acesso em: 19 abr. 2017.

SALTZ, Jeffrey; HECKMAN, Robert. Big Data science education: A case study of a project-focused introductory course. *Themes in Science and Technology Education*, v.8, n.2 – Special Issue on Big Data in Education, p.85–94, dez. 2015.

SCAICO, Pasqueline et al. Um jogo para o ensino de programação em Python baseado na taxonomia de Bloom. In: XXIII SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, Curitiba, 2012. *Anais*.... Curitiba: SBC – Sociedade Brasileira de Computação, 2012. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wei/2012/0028.pdf>>. Acesso em: 21 out. 2017.

VERZANI, John. *Getting Started with RStudio*. Sebastopol, CA: O’Reilly Media, 2011. (Nutchell Handbook).

Isadora Luiz Lemes – Acadêmica do Mestrado em Ensino de Ciências e Matemática – PPGEICM/ULBRA.

Renato P. dos Santos – Docente e Pesquisador do Programa de Pós-Graduação em Ensino de Ciências e Matemática (PPGEICM) da ULBRA Canoas.