

Disciplina: Tópicos Especiais em Botânica: Importação, manejo e visualização de bases de dados

Professor responsável: Gerhard Ernst Overbeck

Professoras: Bianca Ott Andrade e Luciana da Silva Menezes.

Carga Horária: 30 horas, ou 2 créditos (1 teórico, 1 prático).

Modalidade: Semi concentrada, ensino remoto emergencial.

Súmula

Com o aumento das colaborações científicas, o interesse em se armazenar e organizar estes dados a longo prazo para análises em âmbito regional até global tornam-se indispensáveis. Entretanto, a grande maioria desses dados não estão prontamente disponíveis para uso. Conhecimentos de manipulação, limpeza de e visualização de dados são fundamentais. Estudos mostram que essas tarefas ocupam 50-80% do tempo de um profissional que trabalha com análise de dados. Portanto, esse curso tem por objetivo apresentar ferramentas e estratégias para tornar o processo de importação e manejo de dados mais rápido e eficiente. As aulas consistirão de uma curta exposição teórica (20%) e os restantes 80% de aula dedicados para atividades práticas. Ao longo do curso trataremos de fundamentos em organização e gerenciamento de base de dados. Serão utilizados diferentes softwares para montagem, organização e gestão de base de dados. Vantagens e desvantagens do uso de cada software serão discutidos. Além disso, vamos aprender como acessar informações de repositórios globais. Espera-se que ao final do curso os alunos sejam capazes de buscar informação, montar, consultar e manejar bases de dados, bem como desenvolvam habilidade de resolução de problemas no R.

Objetivos de aprendizagem

- Ser capaz de gerar e manejar bases de dados para armazenamento de longo prazo, bem como gerar metadados claro e autoexplicativo;
- Utilizar os diferentes softwares para importação, limpeza, manipulação e visualização dos dados (Excel e R);
- Identificar e distinguir vantagens e desvantagens do uso de diferentes softwares de manejo e análise de dados;
- Conhecer as bases de dados de maior relevância para estudos nas áreas de Ecologia/Botânica/Zoologia;
- Importar matriz utilizando repositórios e bases de dados globais (TRY, GBIF, Worldclim, SoilGrids);
- Encontrar padrões nos dados utilizando ferramentas básicas de visualização de dados;
- Solucionar problemas no R e aprimorar seus conhecimentos de forma independente.

Método de trabalho

Todas as aulas serão síncronas, compostas por uma breve exposição/explicação, seguida de atividade prática realizada em aula. Para as aulas síncronas utilizaremos o Google Meet. As atividades práticas serão realizadas utilizando os softwares Microsoft Excel, R e RStudio. Não é necessário possuir o R e o RStudio instalados no computador, faremos isso nas primeiras aulas. Os participantes da disciplina também devem estar familiarizados com a plataforma Google Drive, que será utilizada para compartilhamento dos arquivos utilizados em aula e construção do portfólio avaliativo.

Avaliação

- Presença e participação nas aulas síncronas 10% do conceito final;
- Método do portfólio: ao longo do curso serão realizadas atividades que os alunos deverão solucionar e depositar a resolução na pasta compartilhada do Google Drive, compondo seu portfólio. O portfólio dos alunos será avaliado ao final da disciplina pelos critérios de: entrega das atividades estipuladas no prazo (pontualidade), organização, apresentação e reprodutibilidade dos resultados (25% do conceito final);
- Avaliação entre pares: os alunos também deverão avaliar o trabalho dos colegas em 4 momentos distintos durante a disciplina, fornecendo *feedbacks* construtivos (40% do conceito final);
- Realizaremos um exercício prático, individual, de caráter avaliativo ao final da disciplina (25% do conceito final);
- Os conceitos finais serão aplicados seguindo a correspondência com as seguintes notas: $\geq 9,0$ = A; $8,99 - 7,5$ = B; $7,49 - 6,0$ = C; $< 6,0$ = D; participação $< 75\%$ = FF. O conceito final mínimo para aprovação é C.

Conteúdo programático e cronograma 2022/1

As aulas ocorrerão sempre às **segundas, quartas e sextas feiras**, no horário das **10:00 até 11:30**, conforme descrito na tabela abaixo:

Semana	Aulas	Data	Conteúdo
Semana 1	Aula 1	4 de abril	Introdução, construção de metadados.
Semana 1	Aula 2	6 de abril	Manejo de base de dados em Excel.
Semana 1	Aula 3	8 de abril	Planilha dinâmica em Excel, criar e exportar gráficos.
Semana 2	Aula 4	11 de abril	Introdução ao R.
Semana 2	Aula 5	13 de abril	Usando o R: importar e manipular dados.
Semana 2	-	15 de abril	FERIADO
Semana 3	Aula 6	18 de abril	Faxina de dados (pacotes do tidyverse), parte 1.
Semana 3	Aula 7	20 de abril	Faxina de dados (pacotes do tidyverse), parte 2.
Semana 3	-	22 de abril	Não haverá aula.
Semana 4	Aula 8	25 de abril	Faxina de dados (pacotes do tidyverse), parte 3.
Semana 4	Aula 9	27 de abril	Visualização dos dados (base e ggplot2), parte 1.
Semana 4	Aula 10	29 de abril	Visualização dos dados (base e ggplot2), parte 2.
Semana 5	Aula 11	2 de maio	Visualização dos dados (base e ggplot2), parte 3.

Semana 5	Aula 12	4 de maio	Base de dados Gbif, parte 1.
Semana 5	Aula 13	6 de maio	Base de dados Gbif, parte 2.
Semana 6	Aula 14	9 de maio	Bases de dados TRY e sPlot, parte 1.
Semana 6	Aula 15	11 de maio	Bases de dados TRY e sPlot, parte 2.
Semana 6	Aula 16	13 de maio	Bases de dados Worldclim e SoilGrids, parte 1.
Semana 7	Aula 17	16 de maio	Bases de dados Worldclim e SoilGrids, parte 2.
Semana 7	Aula 18	18 de maio	Aula sob demanda do grupo.
Semana 7	Aula 19	20 de maio	Exercício prático avaliativo.
Semana 8	Aula 20	23 de maio	Aula de SQL.

Referências

Boehmke, BC, 2016. Data Wrangling with R. Ed. Springer, 257 p.

Grolemund, G, Wickham, H, 2019. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Ed. Alta Books, 528 p.

Mello, MP, Peternello, LA, 2013. Conhecendo o R: uma visão mais que estatística. Ed. UFV., 222 p.

Wickham, H, 2014. Tidy Data. Journal of Statistical Software 59: 1-23.

Zuur A, Ieno EN, Meesters E. 2009. A Beginner's Guide to R. Springer Science & Business Media, 218 p.

Sites

<https://r-project.org/>

<https://rstudio.com/>

<https://gbif.org/>

<https://try-db.org/>

<https://idiv.de/en/splot.html>

<https://worldclim.org/>

<https://soilgrids.org/>