

Introdução à Análise de Dados em Arqueologia

João Cascalheira

Table of contents

Relatório de unidade curricular	3
1 Apresentação	4
2 A unidade curricular no quadro da licenciatura em Património Cultural e Arqueologia	7
3 Estrutura da unidade curricular	10
3.1 Objetivos da aprendizagem	10
3.2 Metodologias de ensino e avaliação	10
3.3 Conteúdos programáticos	10
3.4 Bibliografia	10
4 Organização e conteúdos das aulas	11
4.1 Aula 01 - O registo em Arqueologia: do físico ao digital	11
4.2 Aula 02 - Dados e Bases de dados	14
4.2.1 Tipos de bases de dados	14
4.2.2 Tipos de dados	18
Aula 03 - Recolha de dados	19
Aula 04 - Introdução ao R e Manipulação Básica de Dados	23
Aula 05	27
Aula 06	27
Aula 07	27
Aula 08	27
Aula 09	27
Aula 10	27
Aula 11	27
Aula 12	27
Aula 13	27
Bibliografia	28
Anexos	30

Relatório de unidade curricular

Relatório apresentado no âmbito das provas de habilitação para o título de Agregado no ramo de conhecimento de Arqueologia pela Universidade do Algarve, de acordo com o Art.º 8.º do Decreto-Lei n.º 239/2007, de 19 de junho, e com o Art.º 4.º do Despacho n.º 2251/2020, de 17 de fevereiro.

ICArEHB - Centro Interdisciplinar em Arqueologia e Evolução do Comportamento Humano

Faculdade de Ciências Humanas e Sociais

Universidade do Algarve

Faro

2023

1 Apresentação

O termo “dados” (*data* em inglês) é hoje usado em muitos e diferentes contextos e é quase sinónimo da utilização de computadores Lock (n.d.). Veja-se, por exemplo, como o campo interdisciplinar denominado de “Ciência dos Dados” foi pela primeira vez proposto em 1974 por Peter Naur como nome alternativo para a chamada Ciência de Computação. No entanto a definição do que são “dados” é um problema de investigação por si próprio para os cientistas da informação (Leonelli 2016; Borgman 2012).

É comum assumir-se que os arqueólogos recolhem dados e os introduzem e, computadores. Consequentemente, afirmar que tanto a arqueologia como os computadores dependem de dados parece óbvio e carece de profundidade. Para obter uma compreensão mais profunda, temos de explorar o conceito de dados arqueológicos, a sua potencial relação com dados compatíveis com computadores e as suas implicações no registo arqueológico, na cultura material, na teoria, na interpretação, nas metodologias, na análise e no significado. Além disso, devemos examinar a forma como as ideias relacionadas com a computação, como “tecnologia”, “objetividade” e “científico”, se entrelaçam com a arqueologia, particularmente no âmbito da subdisciplina de arqueoinformática.

Em termos gerais, a Arqueologia pré-década de 1960 estava maioritariamente baseada numa descrição empírica da cultural material, o que incluía acreditar que uma grande quantidade de dados falariam sempre por eles próprios. Os padrões emergiriam do estudo descritivo de coleções de dados para que as cerâmicas, ferramentas em pedra ou estruturas arqueológicas fizessem sentido agrupadas de acordo com determinadas características, a que depois eram atribuídos limites espaciais e temporais através do conceito Childeano de Cultura. Devido a este contexto teórico, os dados eram tidos como adquiridos pela maior parte dos arqueólogos e o processo de observação, registo e interpretação precisava de pouca, se não mesmo nenhuma, justificação.

As mudanças com a Nova Arqueologia, a partir dos finais dos anos 1960, mas com maior intensidade a partir da década de 1970, anunciaram a introdução do Método Científico e a rejeição da percepção subjetiva do empirismo arqueológico. No centro desta nova corrente estava o acreditar na objetividade através da observação sistemática, medindo e registando dados através da adopção dos denominados métodos quantitativos. A objetividade era possível separando a teoria da prática para os que os dados existissem e pudessem ser medidos de forma independente por um qualquer sujeito observador. Assim, enquanto a ligação anterior entre dados e teoria era indutiva, i.e. uma recolha imparcial de “todos” os dados iria produzir teoria, o novo paradigma da arqueologia processual tinha como ponto central um raciocínio

hipotético-dedutivo em que o conhecimento é acumulado através do teste de hipóteses explícitas (muitas vezes através do uso de testes estatísticos de significância) sobre dados recolhidos na base do que seria relevante para a análise (Trigger 1998). A grande implicação do método científico para a Arqueologia tornou-se assim na possibilidade de criar uma arqueologia global, unida por métodos standardizados de análise que pudessem ser aplicados a qualquer conjunto de dados para estabelecer generalizações interculturais ou até mesmo “leis” do conhecimento arqueológico. Se avançarmos para os dias de hoje, estas alterações serviram como pedra basilar para algumas das mais recentes revoluções na forma como se pensa e se pratica a Arqueologia. Fenómenos como a chamada “crise da replicação” e as alterações que esta acarretou na maneira de como recolhemos, gerimos e partilhamos os dados arqueológicos, estão a transformar a Arqueologia.

Como noutras disciplinas, quando vários estudos arqueológicos independentes demonstram resultados semelhantes, consideramos que esses resultados são uma aproximação razoável ao comportamento humano do passado. Esta capacidade de reproduzir os resultados de outros estudos é um princípio fundamental do método científico e, quando as reproduções são bem sucedidas, a disciplina avança. Em Arqueologia, existe uma longa tradição de testes empíricos de reproducibilidade quando, por exemplo, se retorna a sítios escavados por gerações anteriores de arqueólogos ou quando se reexaminam coleções de museus com novos métodos. Contudo, pouco progresso foi feito no que diz respeito a testar a replicabilidade dos dados recolhidos e das respetivas análises quantitativas. Este problema advém de duas razões principais. Primeiro, a maior parte das publicações raramente contém informações suficientes para outro arqueólogo reproduzir os resultados (Marwick 2016). Segundo, a formação base dos arqueólogos raramente inclui o desenvolvimento de competências de organização, gestão e análise de dados. Esta falta de integração de uma componente de literacia dos dados (do inglês *data literacy*, definida como a capacidade de ler, trabalhar, analisar e comunicar com dados) na formação em Arqueologia, prejudica a qualidade dos dados arqueológicos, inibe a sua reutilização, e limita a capacidade que os arqueólogos muitas vezes têm em comunicar num contexto interdisciplinar e consequentemente, a contribuição que podemos ter em conversas contemporâneas fora do âmbito da Arqueologia (apesar de estarmos supostamente bem posicionados para o fazer, ver (Kintigh et al. 2014a)).

Neste contexto, a literacia dos dados não se trata principalmente de permitir que os indivíduos dominem uma habilidade específica ou se tornem proficientes numa determinada plataforma de tecnologia. Em vez disso, trata-se de equipar os arqueólogos com ferramentas para entender os princípios subjacentes e os desafios dos dados (Bhargava et al. 2015; Kansa and Kansa 2021). Atualmente não faltam discussões e iniciativas sobre as promessas e os perigos de alavancar dados em vários tamanhos e formas para enfrentar os desafios do mundo como parte da *Data Revolution*, invocada em 2015 pelas Nações Unidas sobre a agenda de desenvolvimento global. Um artigo, amplamente citado, publicado pelo jornal *The Economist* intitulado “The Data Deluge: Businesses, Governments and Society Are Only Starting to Tap Its Vast Potential” (Cukier 2010) teve como um dos primeiros comentários online o seguinte: “Aqui estão os empregos do século XXI (...). Por favor entendam e eduquem a próxima geração em conformidade”.

A importância da literacia dos dados (bem como a demonimada literacia digital) para um aluno que termina qualquer curso superior nos dias que correm não deve ser desvalorizada. Em Arqueologia, especificamente, é urgente a mudança para perfis profissionais em que a especialização e a profundidade de conhecimentos se conjuguem com uma maior amplitude de conhecimento no uso e tratamento de dados por meios digitais (ver Fig. 1.1). A crescente importância dos métodos digitais de registo, gestão e comunicação dos achados arqueológicos, obrigam necessariamente a que todos os profissionais, tenham um nível básico de literacia dos dados, de preferência através de programas especificamente direcionado para as especificidades dos dados arqueológicos.

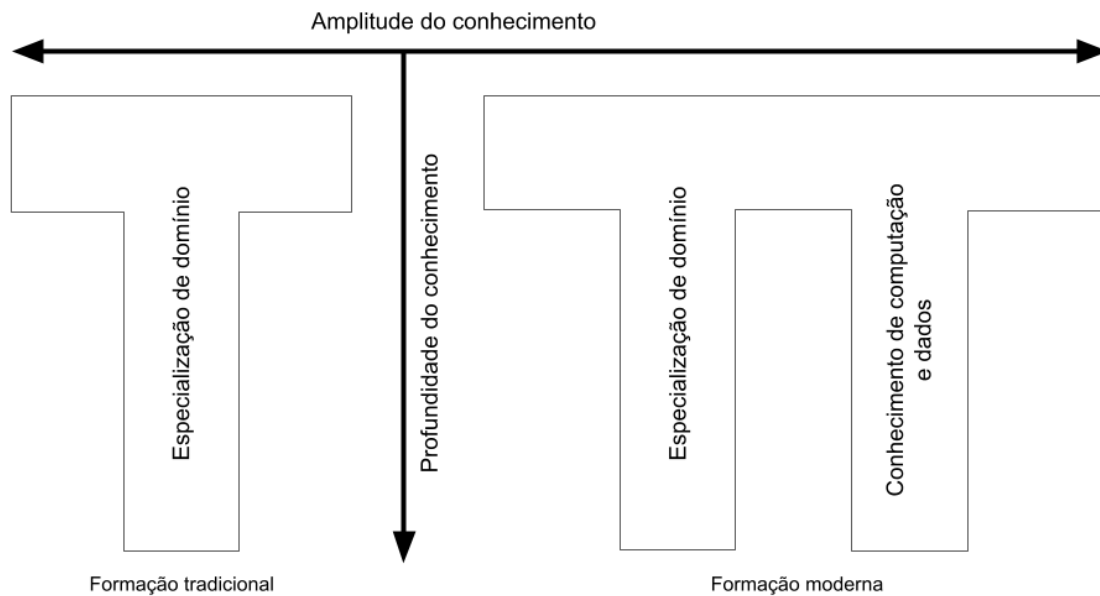


Fig. 1.1: Perfis profissionais em forma de “T” vs forma de “Pi”. Adaptado de (Marwick 2016).

2 A unidade curricular no quadro da licenciatura em Património Cultural e Arqueologia

A Introdução à Análise de Dados em Arqueologia (abreviadamente, IADA) é uma das unidades curriculares optativas do curso de licenciatura em Património Cultural e Arqueologia da Universidade do Algarve, destinando-se, na estrutura do curso, aos alunos que frequentem o 2º e 3º ano.

Esta unidade curricular funcionou pela primeira vez no ano letivo de 2016/2017, repetindo-se em 2017/2018 e 2020/2021. A sua existência resulta de uma necessidade muito concreta: a de preencher uma lacuna na estrutura do curso no que diz respeito ao ensino das técnicas e metodologias computacionais de tratamento, análise e apresentação de dados arqueológicos. Tradicionalmente, no âmbito da licenciatura em Património Cultural e Arqueologia, estes aspectos fundamentais para qualquer profissional de arqueologia, eram lecionados ou no âmbito das disciplinas de técnicas de análise laboratorial (e.g., cerâmicas, líticos, faunas), ou no contexto da realização de trabalhos de seminário de conclusão de curso. Quer num caso, quer noutro, os alunos para além de terem de desenvolver conhecimentos específicos sobre cada uma das disciplinas, bem como desenvolver o trabalho analítico dos materiais em questão, têm ainda que aprender a construir e manter as bases de dados resultantes da análise de conjuntos arqueológicos, escolher as melhores abordagens para as analisar, e finalmente criar e interpretar elementos gráficos com base nos dados recolhidos. Com base na minha experiência anterior na leção das disciplinas de análise de materiais líticos, bem como na orientação de alunos de licenciatura e do mestrado em Arqueologia da UAlg, frequentemente ficou patente a dificuldade que os alunos demonstram neste âmbito, fruto das lacunas relacionadas com a literacia dos dados (ver Chapter 1 para mais informação), nomeadamente a falta de um conhecimento base sobre os diferentes métodos e procedimentos de recolha, análise e apresentação de dados em Arqueologia.

No panorama nacional, IADA não se trata de uma unidade curricular ímpar, sendo que a maior parte dos planos curriculares de licenciaturas em Arqueologia tem atualmente pelo menos uma disciplina dedicada às técnicas informáticas e métodos digitais aplicados à Arqueologia (Tabela 2.1). Ainda assim, torna-se evidente, com base na informação disponibilizada nos respectivos *website*, que os conteúdos programáticos de algumas dessas disciplinas têm como foco dominante o uso de Sistemas de Informação Geográfica ou o desenho assistido por computador. Estes tópicos não fazem parte do programa de IADA, sendo tratados em unidades curriculares

optativas independentes, como por exemplo a de Sistemas de Informação Geográfica Aplicados à Arqueologia. Por outro lado, é certo que os alunos de grande parte destas licenciaturas, incluindo os da UAlg, têm a opção de frequentar outras unidades curriculares optativas de cursos distintos, dedicadas, por exemplo, à ciência de dados ou à estatística introdutória. Nem sempre esta é a melhor opção, tendo em conta, principalmente, as particularidades (inclusivamente éticas) dos dados arqueológicos, e as respectivas limitações e oportunidades na aplicação de determinadas abordagens.

Tabela 2.1: Exemplos de unidades curriculares disponibilizadas noutras instituições com componentes relacionadas com a análise de dados em Arqueologia.

Instituição - Curso	Unidade Curricular	Links
Universidade Nova de Lisboa - Licenciatura em Arqueologia	Métodos digitais em Arqueologia	https://guia.unl.pt/pt/2022/fcsh/program/9006/course/01
Universidade de Coimbra - Licenciatura em Arqueologia	Informática Aplicada à Arqueologia	https://apps.uc.pt/courses/PT/unit/80479/22541/2023-2024?type=ram&id=5382
Universidade de Évora - Licenciatura em História e Arqueologia	Introdução às Humanidades Digitais	https://www.uevora.pt/estudar/cursos/licenciaturas?cod=8estudos&uc=HIS12024L
Universidade do Minho - Licenciatura em Arqueologia	Tratamento Digital da Informação	https://www.ics.uminho.pt/pt/Ensino/Licenciaturas/Arqueologia

O número de alunos inscritos em IADA, nas três edições do curso em que esta foi leccionada, foi de 4 alunos em 2016/2017, e 7 alunos em 2017/2018 e em 2020/2021. Tratando-se de uma unidade curricular optativa destinada apenas aos alunos que optem pelo ramo de Arqueologia no 2º ano do curso, a consolidação numérica das duas últimas edições transpõe uma relativamente boa aceitação dos alunos. Por outro lado, tendo em conta a grande componente prática da unidade curricular, estes números permitiram cumprir os pressupostos do processo de Bolonha com um processo de ensino/aprendizagem mais centrado no aluno.

Do ponto de vista da avaliação, a média geral da avaliação de todos os alunos de todas as edições de IADA situa-se em 14.8 valores. Deve salientar-se ainda, a este respeito, que as avaliações individuais registadas acima destas médias se distribuem entre 15 e 19 valores, e abaixo entre os 14 e os 4 valores. Neste último caso, as duas reprovações estão claramente relacionadas com a fraca assiduidade dos alunos, o que revela algum peso para a componente letiva presencial no âmbito desta unidade curricular.

Apesar do sucesso relativamente elevado dos alunos de IADA, são de salientar dois casos específicos, por terem não só cumprido os objetivos da unidade curricular, mas porque puderam,

posteriormente, revelar a importância dos conteúdos aprendidos para o sucesso do seu percurso enquanto alunos e investigadores. Os dois casos são os de David Nora e Joana Belmiro, cujos seminários finais de licenciatura, bem como as teses de mestrado defendidas na UAlg, incluem ampla utilização das ferramentas e conceitos de recolha, gestão, análise, e partilha de dados introduzidas e exploradas nas versões anteriores de IADA. No caso da Joana Belmiro, a publicação com revisão por pares que teve origem na sua tese de mestrado (Belmiro et al. 2019), utilizou os princípios discutidos em IADA sobre a importância da transparência e da partilha dos dados em Arqueologia.

3 Estrutura da unidade curricular

No Anexo 2 apresenta-se a proposta de Ficha de Unidade Curricular de IADA para os próximos anos letivos em que está previsto o funcionamento da disciplina. Nesta secção esclarecem-se algumas das opções tomadas nos campos mais relevantes para a estruturação da unidade curricular, nomeadamente algumas das alterações efetuadas com base na experiência tida nos anos anteriores.

3.1 Objetivos da aprendizagem

3.2 Metodologias de ensino e avaliação

A avaliação é distribuída pelos três seguintes elementos: 1) elaboração de exercícios práticos - 50%; 2) Trabalho final - 40%; 3) participação e assiduidade - 10%.

Os exercícios práticos são individuais e feitos ao longo das várias aulas de cariz teórico-prático. Na sua essência, estes exercícios correspondem à aplicação prática de cada um dos principais blocos letivos da unidade curricular. Na maior parte dos casos, estes exercícios iniciam-se e terminam-se em contexto de aula, sendo que nalguns casos, em que, por vários motivos, a conclusão dos mesmos não é alcançada em aula, é pedido aos alunos que terminem até à próxima aula, onde serão reservados 30 minutos para revisão das dúvidas e dos resultados finais. O objetivo destes exercícios

3.3 Conteúdos programáticos

A unidade curricular de IADA tem um total de 30 horas de horas teórico-práticas e 5 horas de orientação tutorial; portanto, 140 horas totais de trabalho, equivalentes a 5 ECTS. Tendo em conta que os horários semanais-tipo estabelecem 3 horas de aulas por semana, as aulas distribuem-se ao longo de 13 semanas do respetivo semestre.

3.4 Bibliografia

4 Organização e conteúdos das aulas

4.1 Aula 01 - O registo em Arqueologia: do físico ao digital

A primeira aula da unidade curricular é dedicada à importância do registo em Arqueologia. Foca-se sobretudo como a transição para a recolha, processamento, e análise de dados em ambiente totalmente digital agiliza o fluxo de trabalho e permite uma melhor planificação do trabalho arqueológico e reduz o potencial de erros e a perda de informação irrecuperável do registo arqueológico.

Embora a opção preferencial para o registo arqueológico seja a sua preservação *in situ*, a maior parte das realidades com que o arqueólogo se depara não permitem esta conservação, sendo muitos dos sítios arqueológicos destruídos, tapados ou transformados depois da intervenção. A escavação de um sítio arqueológico é, desta forma, uma experiência que não se pode voltar a repetir e, portanto, a importância do registo exato de todas as suas particularidades deve ser a principal preocupação do arqueólogo.

As abordagens atuais ao processo de registo durante a escavação arqueológica prevêm sobretudo, e de uma forma o mais lógica possível, ligar todos os elementos díspares de um sítio arqueológico no âmbito do registo espacial tridimensional. Naturalmente, a escavação arqueológica não é uma prática precisa, uma vez que existe muito espaço para ambiguidade. O processo de escavar e registar um sítio arqueológico é uma mistura curiosa entre intuição, interpretação e rigor científico Lock (n.d.). O resultado desse processo é um arquivo composto por dados escritos, desenhados e fotografados dos objetos escavados, e da sua relação espacial. Esta preservação por registo tem como principal vantagem o facto de que esse arquivo irá estar acessível para análise, interpretação, e futuras reinterpretações. Coletivamente, esse registo fornece o contexto dos achados feitos, ou seja, os padrões específicos de associação física que os relacionam entre si e com os sedimentos em que foram encontrados (Mitchell 2018).

No entanto, para garantir sucesso neste processo é essencial que os dados estejam estruturados e guardados de forma lógica e sem ambiguidades. O uso crescente de computadores, e particularmente de *software* de Sistemas de Bases de Dados tem sido fundamental para o desenvolvimento dos sistemas de registo de escavação nas últimas décadas.

Devido às particularidades dos sítios arqueológicos e as diferentes ideias e objetivos de equipas de investigação e responsáveis da investigação, conhecemos hoje um conjunto muito alargado de sistemas de registo. Ainda assim é possível reconhecer alguns conceitos-chave e requisitos-chave que integram, actualmente, a maior parte dos trabalhos de escavação. Um destes conceitos

é, por exemplo, o da [Matriz de Harris](#) (Harris 1997) nomeadamente o da conceptualização do registo de contextos únicos, em que um contexto é qualquer evento natural ou antrópico que pode ser distinguido no registo arqueológico.

As primeiras abordagens ao registo de escavação consistiam em texto descritivo, notas curtas e esquemas anotados nos cadernos de campo, produzindo um tipo de registo que é difícil de armazenar e, principalmente, de analisar de forma objetiva, por ser pouco lógico na sua estrutura e não ter sentido explícito. O início da busca por cumprir com estes dois últimos requisitos foi aquando da introdução das folhas *proforma* de registo de campo. Ainda hoje, em várias escavações, é comum ter vários formulários de registo que representam várias classes diferentes de informação, *e.g.* formulários de contexto, formulários de contextos especiais, formulários individuais de achados, etc.

Contudo, aspectos como a ilegibilidade, folhas de dados incompletas e erros entre observadores são algumas das inúmeras dificuldades que os arqueólogos enfrentam quando recolhem dados no terreno utilizando abordagens puramente analógicas. Por outro lado, os arqueólogos lamentam muitas vezes as horas passadas a preencher formulários, que consomem tempo precioso que poderiam dedicar a outras tarefas.

Por estas razões, atualmente, muitos destes formulários foram substituídos por abordagens totalmente digitais ao trabalho de campo arqueológico. De facto, a completa digitalização de dados está a tornar-se um requisito prático para todo o trabalho de campo e de laboratório (Austin 2014; Kintigh et al. 2014b; Ogburn 2010). Várias entidades de financiamento e tutelares de Património e Arqueologia, obrigam os profissionais a demonstrarem uma estratégia para o arquivo e divulgação a longo prazo de todos os dados criados durante a investigação. Em Portugal, por exemplo, a Direção Geral do Património Cultural (DGPC) exige a submissão dos dados recolhidos em todos os trabalhos de campo em suporte físico e digital, para efeitos de preservação dos dados a longo prazo.

A digitalização é frequentemente vista como proporcionadora de uma maior flexibilidade através da sua separação entre função e forma, entre conteúdo e meio, na forma como pode quebrar as fronteiras entre os dados, encoraja e apoia uma utilização dinâmica e colaborativa e proporciona mais oportunidades para a recombinação de dados e a criação de novos conjuntos de dados. Kaufmann and Jeandesboz (2017) sugerem uma série de possibilidades digitais, muitas das quais diretamente relacionadas com a nossa utilização e relação com os dados. Estas incluem a maleabilidade e flexibilidade dos dispositivos digitais, as suas capacidades de armazenamento, a sua capacidade de pesquisa, a sua conectividade, a sua computabilidade, a sua natureza interactiva e a sua criação e organização de dados. A combinação de todos estes factores, e não só, cria um ambiente indiscutivelmente atrativo para a produção, manipulação, consumo e criação de conhecimento a partir dos dados.

No trabalho arqueológico, outra das inevitabilidades para a recolha e gestão de dados de forma totalmente digital relaciona-se com os próprios instrumentos de medição que se utilizam atualmente no campo e em laboratório. A adoção generalizada da Estação Total em trabalhos de escavação e prospeção Bernatchez and Marean (2011); McPherron and Dibble (2002), por

Ficha de Sítio/Trabalho Arqueológico
(para acompanhar o relatório)

Sítio Arqueológico

Designação

Distrito

Concelho

Freguesia

Lugar

C.M.P. 1:25.000 folha n.º

Altitude (m)

Coordenada X

Coordenada Y

Tipo de sítio *

Período cronológico *

Descrição do sítio (15 linhas)

Bibliografia

Proprietários

Classificação *

Decreto

Estado de conservação *

Uso do solo *

Ameaças *

Proteção/Vigilância *

* Preencher de acordo com a lista do *Thesaurus* do ENDOVÉLICO. Essa lista poderá ser consultada em: www.patrimoniocultural.gov.pt

Fig. 4.1: Ficha de sítio da Direção Geral do Património Cultural. Disponível em:
<https://www.patrimoniocultural.gov.pt/static/data/recursos/formularios/fichasitiodgpc_distributed>

exemplo, teve como consequência o armazenamento totalmente digital dos dados espaciais, seja na memória interna do equipamento, seja em computadores conectados por cabo ou por *bluetooth* ao equipamento topográfico (muitas vezes dando lugar ao desenvolvimento de *software* próprio para estabelecer essa ligação e cumprir com necessidades muito específicas em Arqueologia - ver, por exemplo, as soluções disponibilizadas em www.oldstoneage.com). Em laboratório, por outro lado, quase todos os equipamentos utilizados para leitura, registo e análise de materiais arqueológicos são, hoje em dia, completamente digitais. Até mesmo um instrumento amplamente usado para obter medições lineares de objetos - o paquímetro - tem, há vários anos, versões digitais que permitem a sua conexão a um computador, e através de um simples pressionar de botão as medidas são enviadas de forma imediata para o campo da base de dados utilizada.

4.2 Aula 02 - Dados e Bases de dados

Note

Nesta aula são apresentados os princípios básicos de uma base de dados, os tipos de bases de dados mais utilizadas, bem como os tipos de variáveis e escalas de medida. O objetivo é, para além da familiarização com os vários conceitos, demonstrar quão importantes são as decisões tomadas antes do início da recolha de dados, e como essas decisões vão influenciar os tipos de análises que poderão ser posteriormente feitas.

4.2.1 Tipos de bases de dados

Uma base de dados é um simples repositório de informação relacionado com determinado assunto ou finalidade, ou seja, é uma coleção de dados ou itens de informação estruturados de determinada maneira, que permite a sua consulta, atualização, e outros tipos de operações processadas por meios informáticos. Serve para gerir vastos conjuntos de informação de modo a facilitar a organização, manutenção e pesquisa de dados. Existem vários tipos de bases de dados digitais, incluindo bases de dados relacionais, NoSQL, orientadas a objetos, multimodais, espaciais, etc. Em Arqueologia, são sobretudo utilizados três tipos principais de bases de dados: folhas de cálculo, bases de dados relacionais, e bases de dados espaciais.

4.2.1.1 Folhas de cálculo

Uma folha de cálculo é um tipo de ficheiro de armazenamento de dados no qual os dados são armazenados numa estrutura semelhante a uma tabela. É, em certo sentido, muito semelhante ao sistema de quadrícula arqueológica utilizada para identificar unidades de escavação, com linhas horizontais (normalmente identificadas com números) e colunas verticais (normalmente

identificadas com letras) McPherron and Dibble (2003). A intersecção de uma linha específica com uma coluna específica é denominada de célula. Numa folha de cálculo utilizada como base de dados, cada linha representa um registo, cada coluna representa um campo e cada célula contém valores.

Em Arqueologia, as folhas de cálculo são amplamente utilizadas para recolha e análise de dados. A principal vantagem é a sua simplicidade e o facto de poderem ser lidas pela maior parte dos software mais utilizados para análises quantitativas de dados (como MS Excel, Google Sheets, R, Python, SPSS, entre outros), sistemas de gestão de bases de dados e linguagens de programação. Muitas ferramentas de manipulação de dados e bibliotecas de programação oferecem suporte integrado para a leitura e escrita destes ficheiros tornando-os uma escolha popular para o intercâmbio de dados em vários domínios.

Os formatos mais comuns e os mais aconselháveis para este tipo de base de dados são os CSV (comma-separated values) e TXT (ficheiro de texto simples). Ao contrário de outros formatos, este são completamente abertos e, por essa razão são facilmente partilhados entre utilizadores que usem diferentes programas ou sistemas operativos, e tendo em conta que não precisam de *software* específico podem ser abertos e modificados até mesmo com um simples editor de texto.

Contudo, à medida que os requisitos de armazenamento de dados crescem em tamanho e complexidade, as folhas de cálculo podem tornar-se menos eficientes e escaláveis em comparação com bases de dados relacionais, que oferecem capacidades mais avançadas de consulta, indexação, e gestão de dados.

4.2.1.2 Bases de dados relacionais

As bases de dados relacionais são construídas com base nos princípios do modelo relacional, que define os dados como conjuntos de tabelas. De uma forma simplista, uma base de dados relacional agrega várias folhas de cálculo ou ficheiros planos, que contém informação distinta sobre o mesmo conjunto de dados.

A estrutura de uma base de dados relacional baseia-se no conceito de chaves ou identificadores únicos (em inglês *Globally Unique Identifier*, ou GUID), que desempenham um papel vital na manutenção da integridade dos dados e no estabelecimento de relações entre tabelas. A seleção de uma chave primária é essencial, pois influencia o desempenho e a eficiência das consultas e junções na base de dados. Eis algumas das principais características das chaves primárias:

Unicidade: Cada valor na coluna da chave primária deve ser único. Isto garante que não há dois registos na tabela com o mesmo identificador.

Não-nulidade: Uma chave primária não pode conter valores NULL. Cada registo na tabela tem de ter um valor de chave primária válido e não nulo.

Vale Boi - Suportes

File Edit View Insert Format Data Tools Extensions Help

Q Menu

100%

123

Default

10

B

Z

A

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

10

Fig. 4.2: Exemplo de folha de cálculo do Google Sheets com dados provenientes da análise de ferramentas em pedra do sítio arqueológico de Vale Boi. Cada linha representa um artefacto, e cada coluna uma das variáveis analisadas. Os dados presentes nas células representam os valores para cada uma das variáveis ou campos.

Imutabilidade: Idealmente, os valores da chave primária não devem mudar depois de serem atribuídos a um registo. Isto assegura a consistência e evita potenciais problemas com a referenciação de dados.

Coluna única ou chave composta: Uma chave primária pode ser composta por uma única coluna ou por uma combinação de várias colunas. Neste último caso, é designada por chave composta.

Um dos exemplos práticos de uma base de dados relacional é a utilizada pelo software Newplot, um programa open source desenvolvido para gerir dados provenientes de trabalhos de escavação arqueológica. O Newplot funciona sobre uma base de dados relacional em MS Access. Esta é composta por várias tabelas, das quais as mais importantes são as *Context* e *XYZ*. A primeira contém informação sobre um determinado artefacto ou conjunto de artefactos e o contexto em que estes foram encontrados (e.g., unidade estratigráfica, tipo de material). A segunda contém informação espacial da proveniência desses artefactos, nomeadamente as coordenadas tridimensionais x, y, e z. As duas tabelas encontram-se relacionadas através de uma chave primária composta, representada pelas variáveis *UNIT* e *ID*, que correspondem, respetivamente, ao quadrado, área de escavação ou sítio arqueológico (e.g., A1, L15, Abrigo, Vale Boi), e a um número sequencial de identificação. No seu conjunto, a combinação de *UNIT* e *ID* tem de ser única. Por exemplo, em toda a base de dados, apenas um objeto pode ter como chave primária “A1-100” para que se possa relacionar as suas coordenadas na tabela *XYZ* com a respetiva informação contextual na tabela *Context*. Nalguns casos, contudo, é necessário registar várias coordenadas tridimensionais para um único objeto, que permitam recolher informação relativa à sua orientação ou ao seu tamanho e forma geral. Note-se, no entanto, que com esta abordagem estaremos a criar mais do que um registo na tabela da base de dados para cada objeto registado. A base de dados associada ao program Newplot está preparada para que um objeto tenha várias entradas na tabela *XYZ*, utilizando o mesmo conjunto de UNIT-ID mas adicionando uma outra variável, denominada *SUFFIX*. Neste caso, a relação entre as tabelas *Context* e *XYZ* é denominada de um-para-múltiplos (1:M), em que várias entradas em *XYZ* correspondem apenas a uma entrada de *Context*.

4.2.1.3 Bases de dados espaciais

As bases de dados espaciais são um tipo especializado de bases de dados concebido para armazenar, gerir e processar dados espaciais ou geográficos. Desempenham um papel vital em várias aplicações que envolvem informações baseadas na localização, como os Sistemas de Informação Geográfica (SIG), a monitorização ambiental, o planeamento urbano, a logística, serviços baseados na localização, ou em Arqueologia. Estas bases de dados estão equipadas com técnicas de indexação espacial e capacidades de consulta espacial que permitem a recuperação e análise eficientes de dados espaciais. Permitem a representação de objectos geométricos simples, como pontos, linhas e polígonos, e, em alguns casos, permitem tratar de estruturas mais complexas, como objectos 3D, e redes irregulares trianguladas (em inglês Triangulated Irregular Network - TIN).

Quase todo o trabalho de campo arqueológico produz grandes quantidades de informação baseada na localização, seja de sítios arqueológicos, ou de objetos dentro de um sítio arqueológico. A proveniência de cada fóssil, artefacto e amostra é fundamental para compreender a sua idade, tafonomia, contexto ecológico, contexto cultural e significado comportamental. Desta forma, a proveniência espacial de um item é muitas vezes tão importante como o próprio item, sendo a utilização de bases de dados espaciais em Arqueologia particularmente útil. Principalmente quando estão incluídos dados geográficos com coordenadas absolutas relativas a um datum e sistema de projecção específicos.

Um dos exemplos do uso de uma base de dados espacial para consulta online do inventário de sítios arqueológicos de Portugal Continental, é a do [Geoportal da Direção Geral do Património Cultural](#). Esta está baseada no formato Geodatabase da ESRI. Este é um formato proprietário utilizado pelo conjunto de software ArcGIS da ESRI. Fornece um modelo de dados espaciais abrangente com vários tipos de dados, incluindo pontos, linhas, polígonos e dados raster.

Outro dos exemplos disponíveis para consulta é o do projeto [Paleo Core](#), uma plataforma online para gestão de dados arqueológicos, paleontológicos, e geológicos, com particular ênfase nos períodos Pliocénico e Pleistocénico. Toda a base de dados deste projeto é suportado por programas de acesso aberto, incluindo a extensão PostGIS do popular sistema de bases de dados relacionais PostgreSQL. O PostGIS adiciona suporte para gerir e analisar de dados espaciais, transformando o PostgreSQL num banco de dados espacial completo, com tipos específicos de dados espaciais, índices espaciais e funções espaciais.

4.2.2 Tipos de dados

Todos os formatos de bases de dados aceitam a definição de vários tipos de dados, que devemos escolher de acordo com a informação que será colocada dentro desse campo. Destes tipos de dados devemos escolher aqueles que melhor se adaptem à informação que queremos armazenar. O tipo de dados utilizado determina a natureza dos dados que podem ser armazenados, o intervalo de valores que pode assumir e as operações que podem ser efectuadas sobre ele. Por vezes, são confundidos os tipos de dados possíveis de escolher numa base de dados com os tipos de variáveis estatísticas. Apesar de relacionados, no caso das bases de dados são tidas em consideração sobretudo as operações de armazenamento e manipulação e por isso existem mais tipos que no caso das variáveis estatísticas. Em baixo são apresentados os principais exemplos dos tipos de dados.

Tabela 4.1: Principais tipos de dados que definem os valores que uma coluna de uma base de dados pode conter.

Tipo de dados	Descrição
Texto	Valores curtos e alfanuméricos, tais como um apelido ou um endereço
Texto Breve	
Número, Número Grande	Valores numéricos, tais como distâncias.
Porcentagem	Porcentagens.
Científico	Aceita numeração científica.
Moeda	Valores monetários.
Sim/Não	Também conhecido como lógico ou booleano. Os valores Sim e Não e os campos que contêm apenas um de dois valores.
Data/Hora	Data/Hora: valores de data e hora dos anos 100 a 9999.
Data/Hora Prolongada	Data/Hora Prolongada: os valores de data e hora dos anos 1 a 9999.
Rich Text	Texto ou combinações de texto e números que podem ser formatados ao utilizar controlos de cor e de tipo de letra.
Campo de Cálculo	Resultados de um cálculo. O cálculo tem de fazer referência a outros campos da mesma tabela.
Anexo	Imagens, ficheiros de folha de cálculo, documentos, gráficos e outros tipos de ficheiros suportados anexados aos registos na base de dados, semelhante à forma de anexar ficheiros a mensagens de e-mail.
Hiperligação	Texto ou combinações de texto e números armazenados como texto e utilizados como um endereço de hiperligação.
Memo Texto Longo	Longos blocos de texto. Uma utilização típica de um campo Memo seria uma descrição detalhada do produto.
Pesquisa	Lista de valores que é obtida a partir de uma tabela ou consulta ou apresenta um conjunto de valores que especificou quando criou o campo.

Aula 03 - Recolha de dados

Em geral, por vários motivos, na prática arqueológica os principais objectivos são recolher o máximo de dados necessários, o mais rapidamente possível, com o mínimo de erros possível. Nesta aula são apresentadas e postas em prática técnicas para atingir estes objectivos,

nomeadamente no que diz respeito ao trabalho de laboratório e à recolha de dados a partir de artefactos arqueológicos. Algumas destas técnicas envolvem hardware especializado. Outras, tais como a criação de menus e ficheiros de ajuda que contêm listas de respostas correctas ou válidas para as variáveis seleccionadas, envolvem software que pode ser aplicado à maioria dos programas de gestão de bases de dados.

Decidir que variáveis incluir em determinada base de dados para análise de materiais arqueológicos é o primeiro passo a dar antes de iniciar a análise em si. Contudo, a escolha dessas variáveis vai depender muito do objetivo do estudo e do tipo de objetos em análise e, portanto, não nos preocuparemos com esse passo. Preocupamo-nos sim, em perceber como é que vamos traduzir essas observações num conjunto de variáveis numa base de dados que (1) guarde a informação de forma eficiente, (2) torne a recuperação e análise dos dados simples, (3) integre facilmente com as outras análises que eventualmente irão ser feitas no contexto do projeto arqueológico em causa.

Neste contexto, uma das primeiras preocupações será a da organização e decisão da unidade de análise. Frequentemente, cada objeto é definido como a unidade de análise, e os campos da base de dados representam as várias observações que são feitas individualmente. Num exemplo muito simples, no âmbito da análise de artefactos líticos, cada peça é considerada uma unidade de análise, registando-se variáveis como o Comprimento, a Largura, a Espessura, o Tipo de Talão, etc., para cada um dos artefactos, individualmente. Nestes casos, a introdução dos dados não apresenta grandes desafios, tendo em conta que se pode fazer a análise utilizando um único ficheiro plano ou folha de cálculo.

Noutros casos, como por exemplo na análise de marcas de corte em ossos de animais, a situação é muito mais desafiante. A principal razão é porque se passa a ter duas unidades principais de análise: o osso em si (do que se podem tirar medidas, identificar a espécie, etc.), e cada uma das marcas de corte identificadas (que podem ser individualmente caracterizadas segundo as suas dimensões, perfil, etc.). Uma abordagem possível é adicionar um conjunto de variáveis para cada marca de corte na base de dados, mas isto levanta dois problemas: por um lado, nem todos os ossos terão marcas de corte e isso implica que existirão muitas células em branco; por outro lado, é impossível adivinhar o número de marcas de corte antes de se começar a análise e isso implicaria adicionar novas variáveis durante o estudo, o que deve ser sempre evitado.

Assim, nestes casos a melhor estratégia será sempre separar a análise em pelo menos duas tabelas distintas, de acordo com as duas unidades de análise definidas -uma para os ossos, outra para as marcas de corte. Naturalmente, estas duas tabelas deverão estar sempre relacionadas através de um identificador único (i.e., uma chave primária) que permite ligar os dois registos. Mais uma vez, os identificadores únicos são muito importantes, uma vez que sem essa referência não se consegue associar os dados registados com o(s) objeto(s) analisados. Assim, é também extremamente importante manter junto do objeto analisado (através de uma etiqueta ou, quando possível, marcação direta na peça) informação de qual o identificador único utilizado. Mesmo no caso de se tratarem de objetos de um coleção de museu, por exemplo, em que não exista identificador único associado, deve ser criado um sistema de

identificação que permita a qualquer momento durante a análise associar os dados recolhidos com o objeto analisado.

Uma vez decida a(s) unidade(s) de análise, pode ser criada a base dados atribuindo a cada variável que irá ser medida um nome e especificando que tipo de dados irão ser introduzidos nesse campo. Apesar de parecer frugal, a nomeação das variáveis é muito importante quando se desenha uma bases de dados de qualquer tipo. Os nomes devem não apenas ser únicos, mas devem também ser descritivos e facilmente interpretáveis. Nomes como VAR1, VAR2, VAR3, etc., não permitirão identificar facilmente que informação foi registada naquele campo. Espaços em branco, acentos, e simbolos especiais (e.g., %, \$) devem também ser evitados, uma vez que alguns software têm problemas na conversão desses caracteres. Um caracter que é frequentemente usado, contudo, é o chamado *underscore*, sobretudo para substituir o espaço entre duas ou mais palavras (e.g., Tipo_Talao).

Como anteriormente referido, um dos mais utilizados suportes informáticos para introdução de dados em Arqueologia são as chamadas Folhas de Cálculo. Com os princípios definidos acima, é relativamente fácil abrir um documento do Goodle Sheets, inserir os nomes das variáveis na primeira linha e iniciar a análise inserindo informação sobre cada objeto em cada uma das linhas. Esta é a maneira que muitos arqueólogos usam, em todo mundo, para analisar as suas coleções. A principal razão é, claro, a simplicidade. No entanto, alguns problemas são comuns quando se usa esta opção. Um dos melhores exemplos é que quando um conjunto de variáveis não são relevantes para aquele objeto específico, o utilizador tem de olhar para os nomes das variáveis, saltar com o cursor todas as células que não interessam e reiniciar a análise na próxima variável que é possível registar. Não só este processo demora alguns preciosos segundos, mas principalmente pode facilmente levar a que se preencha células que não era suposto.

Existem hoje opções mais práticas que permitem a introdução mais rápida e com menos erros que a tradicional Folha de Cálculo. Uma das soluções desenvolvidas especificamente para projetos arqueológicos é o software E5, que faz parte do ecossistema de programas criados por Shannon McPherron e Harold Dibble. É totalmente livre e pode ser descarregado em: <https://github.com/surf3s/E5>

O E5 é um programa generalizado de introdução de dados, que funciona com um ficheiro de configuração onde são definidos os campos de introdução de dados. A principal vantagem do E5 é que permite condicionar o preenchimento de variáveis de acordo com valores preenchidos em variáveis anteriores.

O elemento-chave do E5 é o ficheiro de configuração onde são definidos os campos de introdução de dados. Os ficheiros de configuração (que terminam com a extensão CFG) podem parecer um pouco complicados à primeira vista, e devem ser escritos num programa separado (um editor de texto como o [NotePad](#), [NotePad++](#), [Atom](#) ou [Sublime text](#)). No entanto, o esforço de pensar num ficheiro de configuração significa pensar na estrutura dos seus dados antes de começar a recolhê-los (ao contrário, por exemplo, de usar uma folha de cálculo), e este esforço normalmente compensa mais tarde, durante a análise dos dados.

Vários exemplos de ficheiros CFG estão incluídos na plataforma do E5, e aqui está um exemplo de ficheiro de configuração para ilustrar algumas características:

```
[E5]
TABLE=lithics

[ID]
TYPE=TEXT
PROMPT=Enter the artifact ID
UNIQUE=True

[ARTIFACTTYPE]
TYPE=MENU
PROMPT=Select the artifact type
MENU=Tool,Flake,Core

[TOOLTYPE]
TYPE=MENU
PROMPT=Select the tool type
MENU=Scraper,Notch,Point,Other
CONDITION1=ArtifactType Tool

[PLATFORMTYPE]
TYPE=MENU
PROMPT=What is the platform
MENU=Plain,Cortical,Missing,Other
CONDITION1=ArtifactType Tool,Flake

[PLATFORMWIDTH]
TYPE=NUMERIC
PROMPT=Measure the platform width
CONDITION1=ArtifactType Tool,Flake
CONDITION2=PlatformType not Missing

[WEIGHT]
TYPE=NUMERIC
PROMPT=WEIGHT
```

Aula 04 - Introdução ao R e Manipulação Básica de Dados

4.2.0.1 Começar a utilizar o R

O R é uma poderosa linguagem de programação amplamente utilizada para computação estatística e análise de dados em vários domínios científicos, incluindo a Arqueologia. De facto, na última década, a linguagem R é a que mais tem sido utilizada, salientando alguns autores que a sua crescente adopção tem provocado uma verdadeira revolução na disciplina Schmidt and Marwick (2020). Algumas das principais razões pelas quais o R é utilizado em arqueologia:

- Análise e visualização de dados: Os arqueólogos lidam frequentemente com grandes conjuntos de dados provenientes de escavações ou trabalho laboratorial. O R fornece inúmeros pacotes para manipulação, análise e visualização de dados, facilitando a exploração e interpretação de dados arqueológicos. Desde estatísticas sumárias simples a métodos estatísticos avançados, o R pode lidar com uma vasta gama de tarefas analíticas.
- Ferramentas estatísticas: O R oferece um conjunto abrangente de ferramentas estatísticas que são essenciais, nomeadamente uma grande variedade de modelos de regressão, testes de hipóteses, análises multivariadas e estatísticas espaciais.
- SIG e análise espacial: Os Sistemas de Informação Geográfica (SIG) desempenham um papel crucial na Arqueologia, uma vez que a distribuição espacial de sítios e elementos arqueológicos é frequentemente significativa. O R tem pacotes como ‘sp,’ ‘raster,’ e ‘sf’ que permitem efetuar análises espaciais sofisticadas, criar mapas e sobrepor dados arqueológicos em camadas geográficas.
- Reprodutibilidade e documentação: O R promove um fluxo de trabalho transparente e reproduzível, o que é crucial para a investigação científica, incluindo a Arqueologia. Ao utilizar scripts R, podem-se documentar as análises passo a passo, facilitando a compreensão e a reprodução dos resultados por outras pessoas.
- Integração com outras ferramentas: O R pode ser facilmente integrado noutras linguagens e ferramentas de programação, permitindo utilizar *software* e bibliotecas especializadas para tarefas específicas. Por exemplo, o R pode interagir com programas como o QGIS, GRASS GIS e bases de dados como o PostgreSQL, melhorando as capacidades de gestão e análise de dados.
- Comunidade e recursos: O R tem uma comunidade de utilizadores grande e ativa. Esta comunidade cria e mantém numerosos pacotes adaptados a necessidades arqueológicas específicas. A disponibilidade destes pacotes expande consideravelmente a gama de análises que podem ser efectuadas com o R.
- Flexibilidade e personalização: O R é uma linguagem de código aberto, o que significa que se podem personalizar e alargar as funcionalidades para se adequarem às suas necessidades específicas de investigação. Podem-se criar funções, pacotes e visualizações, tornando o R uma ferramenta flexível e adaptável para necessidades diversas.

- Económica: O R é um software gratuito e de código aberto, tornando-o acessível a todos. Elimina a necessidade de adquirir licenças dispendiosas de software estatístico, tornando-o uma opção atractiva para investigadores e instituições académicas.

4.2.0.1.1 R Instalação e configuração

O R pode ser descarregado a partir do [Comprehensive R Archive Network \(CRAN\)](#). Depois de instalar o R, deve instalar o [RStudio](#), um ambiente de desenvolvimento integrado que fornece uma interface de fácil utilização para trabalhar com o R. O RStudio oferece funcionalidades como realce de sintaxe, conclusão de código e visualização de gráficos, tornando a experiência de programação mais fácil. Depois de instalados os dois programas apenas é necessário abrir o RStudio, uma vez que o R será automaticamente iniciado.

4.2.0.2 Noções básicas de R e estruturas de dados

4.2.0.2.1 A interface do RStudio

O RStudio é um ambiente de desenvolvimento integrado (IDE na sigla inglesa) concebido especificamente para a linguagem de programação R. Fornece uma interface de fácil utilização que simplifica o processo de escrita, execução e gestão de código R. A interface está dividida em vários componentes-chave, cada um servindo um objetivo único para aumentar a produtividade do utilizador.

- *Source*: Este painel é onde se escreve/visualiza os scripts R (identificados com a extensão .R). Alguns resultados (tais como visualizar um conjunto de dados utilizando `View()`) aparecerão como um separador aqui.
- *Console/Terminal/Jobs*: É aqui que se vê a execução dos comandos. É o mesmo ecrã que se veria se estivessemos a utilizar o R na linha de comandos sem o RStudio. Pode-se trabalhar interactivamente (ou seja, introduzir comandos R aqui), mas na maior parte das vezes executa-se um *script* (ou linhas num *script*) no painel *Source* e vê-se a sua execução e resultados aqui. O separador *Terminal* dá acesso ao terminal BASH (o sistema operativo Linux, não relacionado com o R).
- *Environment/History*: Aqui, o RStudio mostra quais os conjuntos de dados e objectos que criou e quais estão definidos na memória. Também pode ver algumas propriedades de objectos/conjuntos de dados, como o seu tipo e dimensões. O separador *History* contém um histórico dos comandos R executados durante a sessão.
- *Files/Plots/Packages/Help/Viewer*: Este painel polivalente mostra-lhe o conteúdo dos directórios no seu computador. Também pode usar o separador *Files* para navegar e definir o directório de trabalho. O separador *Plots* mostra a saída de quaisquer gráficos gerados. Em *Packages* verá os pacotes que estão ativamente carregados ou pode anexar pacotes instalados. “Help” mostra os ficheiros de ajuda para as funções e pacotes do R.

O separador *Viewer* permite-lhe ver o conteúdo local da Web (por exemplo, resultados HTML).

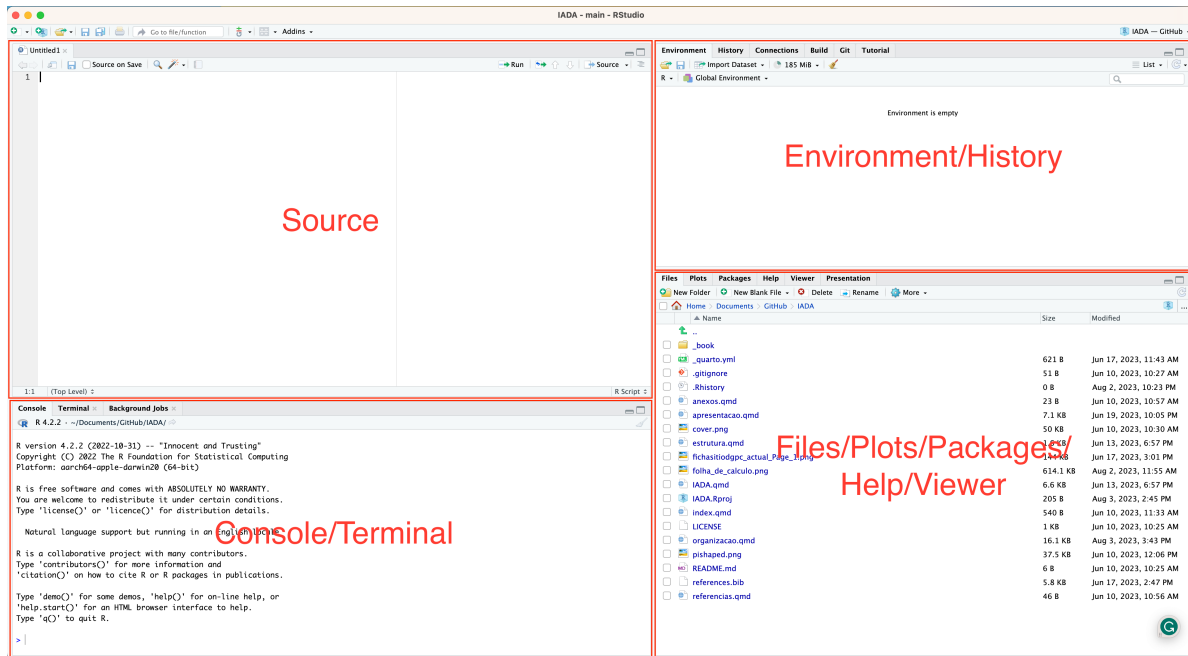


Fig. 4.3: A interface do RStudio com a indentificação das principais componentes.

4.2.0.3 O R como calculadora

O R pode ser utilizado como uma calculadora simples para operações aritméticas básicas. Pode-se efetuar cálculos como adição, subtração, multiplicação e divisão diretamente na *Console* do R.

```
# Adição  
5 + 3
```

```
[1] 8
```

```
# Subtração  
10 - 2
```

```
[1] 8
```

```
# Multiplicação  
4 * 6
```

```
[1] 24
```

```
# Divisão  
16 / 4
```

```
[1] 4
```

4.2.0.3.1 Estruturas de dados em R

Compreender as estruturas de dados em R é essencial, pois determina como os dados são armazenados e manipulados. As principais estruturas de dados incluem:

- *Vectors*: Uma estrutura de dados básica em R, que pode conter elementos do mesmo tipo de dados, como valores numéricos, de caracteres ou lógicos.
- *Matrices*: Uma matriz bidimensional com linhas e colunas, contendo elementos do mesmo tipo de dados.
- *Arrays*: Generalização multidimensional de vetores e matrizes.
- *Dataframes*: Uma estrutura bidimensional semelhante a uma tabela em que as colunas podem ser de diferentes tipos de dados. As estruturas de dados são amplamente utilizadas no tratamento e análise de dados do mundo real.

Aula 05

Aula 06

Aula 07

Aula 08

Aula 09

Aula 10

Aula 11

Aula 12

Aula 13

Bibliografia

- Austin, Anne. 2014. "Mobilizing Archaeologists: Increasing the Quantity and Quality of Data Collected in the Field with Mobile Technology." *Advances in Archaeological Practice* 2: 13–23.
- Belmiro, Joana, João Cascalheira, Nuno Bicho, and Jonathan Haws. 2019. "The Gravettian-Solutrean Transition in Western Iberia: New Data from the Sites of Vale Boi and Lapa Do Picareiro (Portugal)."
- Bernatchez, J., and Curtis W. Marean. 2011. "Total Station Archaeology and the Use of Digital Photography." *SAA Archaeol. Rec* 11.
- Bhargava, Rahul, Erica Deahl, Emmanuel Letouzé, Amanda Noonan, David Sangokoya, and Natalie Shoup. 2015. "Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data."
- Borgman, Christine L. 2012. "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology* 63 (6): 1059–78.
- Cukier, K. 2010. "The Data Deluge: Businesses, Governments and Society Are Only Starting to Tap Its Vast Potential." *The Economist* 23.
- Harris, Edward C. 1997. *Principles of Archaeological Stratigraphy*. 2. ed., 3. print. London: Acad. Pr.
- Kansa, Eric, and Sarah Whitcher Kansa. 2021. "Digital Data and Data Literacy in Archaeology Now and in the New Decade." *Advances in Archaeological Practice* 9 (1): 81–85. <https://doi.org/10.1017/aap.2020.55>.
- Kaufmann, Mareile, and Julien Jeandesboz. 2017. *Politics and 'the Digital' from Singularity to Specificity*. Vol. 20. SAGE Publications Sage UK: London, England.
- Kintigh, Keith W, Jeffrey H Altschul, Mary C Beaudry, Robert D Drennan, Ann P Kinzig, Timothy A Kohler, W Fredrick Limp, et al. 2014a. "GRAND CHALLENGES FOR ARCHAEOLOGY." *AMERICAN ANTIQUITY* 79 (1): 20.
- , et al. 2014b. "GRAND CHALLENGES FOR ARCHAEOLOGY." *AMERICAN ANTIQUITY* 79 (1): 20.
- Leonelli, Sabina. 2016. "The Philosophy of Data." In, 207–18. Routledge.
- Lock, Gary. n.d. "Using Computers in Archaeology: Towards Virtual Pasts," 317.
- Marwick, Ben. 2016. "Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation." *Journal of Archaeological Method and Theory* 24 (2): 424–50. <https://doi.org/10.1007/s10816-015-9272-9>.
- McPherron, Shannon P., and Harold L. Dibble. 2002. *Using Computers in Archaeology: A Practical Guide*. McGraw-Hill.
- . 2003. *Using Computers in Archaeology: A Practical Guide*. Mayfield: McGraw-Hill.

- Mitchell, Peter. 2018. "Introduction to Archaeological Methods and Sources." In. <https://doi.org/10.1093/acrefore/9780190277734.013.367>.
- Ogburn, Joyce L. 2010. "The Imperative for Data Curation." *Portal: Libraries and the Academy* 10 (2): 241–46.
- Schmidt, Sophie C., and Ben Marwick. 2020. "Tool-Driven Revolutions in Archaeological Science." *Journal of Computer Applications in Archaeology* 3 (1): 1832. <https://doi.org/10.5334/jcaa.29>.
- Trigger, Bruce G. 1998. "Archaeology and Epistemology: Dialoguing Across the Darwinian Chasm." *American Journal of Archaeology* 102 (1): 1–34.

Anexos