Teaching Data Visualisation and Basic Map-Making Skills at a Liberal Arts College

Michael T. GASTNER

Yale-NUS College, 16 College Avenue West, #01 -220, Singapore 138527 and Information and Communications Technology Cluster, Singapore Institute of Technology, 10 Dover Drive, Singapore 138683
michael.gastner@singaporetech.edu.sg (ORCID: 0000-0002-1097-8833)

Abstract. Communicating data has become vital in an increasingly data-dependent society. Because data are often communicated graphically, many colleges and universities are offering courses that teach the principles of data visualisation. In this article, the author reflects on his experience of teaching such a course at a liberal arts college. The course introduced students to the tidyverse suite of packages for the programming language R, including the data visualisation package ggplot2, which implements Wilkinson's (2005) 'grammar of graphics'. The packages sf and tmap were used to extend the capabilities of the tidyverse packages for analysing geospatial data and producing thematic maps. This article provides insights into effective teaching strategies, including intended learning outcomes, the core syllabus, pedagogy and assessment. Sample tasks and visualisations are also presented to demonstrate how essential cartographic skills were imparted to students.

Keywords: pedagogy, grammar of graphics, visualisation software, tidyverse

1 Introduction

Data visualisation is the graphical representation of information and data (Bikakis 2019). It is a useful tool because 'the human visual system is a pattern seeker of enormous power and subtlety' (Ware 2021, xiv). The importance of data visualisation is recognised in fields such as business and finance (Solis 2019), medicine and healthcare (Park et al. 2022), journalism and media (Engebretsen 2020), public policy (Raineri and Molinari 2021) and environmental planning (Metze 2020). Accordingly, visualisation literacy has been described as 'essential to support cognition and evolve toward a more informed society' (Firat et al. 2022, 99).

Many definitions of visualisation literacy highlight the capability to read, interpret and extract information from visual representations of data (Boy et al. 2014; Börner et al. 2016; Lee et al. 2017). This set of skills is analogous to reading comprehension, which is a requirement but not a synonym for textual literacy. To be considered literate in a language, an individual must also be able to speak fluently and create meaningful written documents. Similarly, visualisation literacy is not solely determined by the ability to read and understand statistical charts but also by the capacity to produce one's own meaningful graphical displays.

In response to the growing need for visualisation literacy in this broader sense, colleges and universities worldwide are offering data visualisation courses as part of their degree programmes or for continuing education of working professionals, such as the course described by Chiera (2017). Yale-NUS College, which was Singapore's first liberal arts college, taught its own course 'Data Analysis and Visualisation'. The liberal arts education aims to cultivate intellectual openness in students (Blaich et al. 2004).

Podučavanje vizualizacije podataka i osnovnih vještina izrade karata na Liberal Arts Collegeu

Michael T. GASTNER

Yale-NUS College, 16 College Avenue West, #01 -220, Singapore 138527 and Information and Communications Technology Cluster, Singapore Institute of Technology, 10 Dover Drive, Singapore 138683
michael.gastner@singaporetech.edu.sg (ORCID: 0000-0002-1097-8833)

Sažetak. U društvu koje sve više ovisi o podatcima komuniciranje podataka postalo je vitalno. Budući da se podatci često prenose grafički, mnogi fakulteti i sveučilišta nude tečajeve koji podučavaju principe vizualizacije podataka. U ovom se članku autor osvrće na svoje iskustvo podučavanja takvog kolegija na Liberal Arts Collegeu. Tečaj je upoznao studente s paketom paketa tidyverse za programski jezik R, uključujući paket za vizualizaciju podataka ggplot2 koji primjenjuje Wilkinsonovu (2005) "gramatiku grafike". Paketi sf i tmap korišteni su za proširenje mogućnosti paketa tidyverse za analizu geoprostornih podataka i izradu tematskih karata. Ovaj članak pruža uvid u učinkovite strategije podučavanja, uključujući predviđene ishode učenja, temeljni nastavni plan i program, odgojno-obrazovni proces i ocjenjivanje. Prikazani su i ogledni zadatci i vizualizacije kako bi se pokazalo kako su osnovne kartografske vještine prenesene studentima.

Ključne riječi: odgojno-obrazovni proces, gramatika grafike, softver za vizualizaciju, tidyverse

1. Uvod

Vizualizacija podataka grafički je prikaz informacija i podataka (Bikakis 2019). To je koristan alat jer je "ljudski vizualni sustav tragač za uzorcima goleme moći i suptilnosti" (Ware 2021, xiv). Važnost vizualizacije podataka prepoznata je u područjima kao što su poslovanje i financije (Solis 2019), medicina i zdravstvo (Park et al. 2022), novinarstvo i mediji (Engebretsen 2020), javna politika (Raineri i Molinari 2021) i planiranje zaštite okoliša (Metze 2020). Sukladno tome, vizualizacijska je pismenost opisana kao "ključna za potporu kogniciji i razvija se prema informiranijem društvu" (Firat et al. 2022, 99).

Mnoge definicije vizualizacijske pismenosti ističu sposobnost čitanja, tumačenja i izdvajanja informacija iz vizualnih prikaza podataka (Boy et al. 2014, Börner et al. 2016, Lee et al. 2017). Taj je skup vještina analogan razumijevanju pročitanog, što je uvjet, ali ne i sinonim za tekstualnu pismenost. Da bi se osoba smatrala pismenom na nekom jeziku, ona također mora moći tečno govoriti i stvarati smislene pisane dokumente. Slično tome, vizualizacijska pismenost nije određena samo sposobnošću čitanja i razumijevanja statističkih dijagrama, već i sposobnošću izrade vlastitih smislenih grafičkih prikaza.

Kao odgovor na sve veću potrebu za vizualizacijskom pismenošću u ovom širem smislu, koledži i sveučilišta diljem svijeta nude tečajeve vizualizacije podataka kao dio svojih studijskih programa ili za kontinuirano obrazovanje stručnjaka, poput tečaja koji opisuje Chiera (2017).

Yale-NUS College, koji je bio prvi singapurski koledž liberalnih vještina, imao je vlastiti kolegij "Analiza podataka i vizualizacija". Liberalno obrazovanje ima za cilj njegovati intelektualnu otvorenost kod studenata (Blaich et al. 2004). Da bi ispunili taj cilj, Yale-NUS College i mnogi drugi koledži za slobodne

To meet this goal, Yale-NUS College and many other liberal arts colleges require students to take courses in humanities, social sciences and natural sciences, instead of narrowly focusing on a single subject early on. Therefore, as the initiator and lead developer of 'Data Analysis and Visualisation', I aimed to create an interdisciplinary course that fostered creative computational thinking, including a basic understanding of visualising geospatial data. Cartographic skills, imparted on students through hands-on map-making tasks, included the ability to choose suitable map projections, selecting appropriate symbology and creating effective legends.

The course was taught five times between 2016 and 2022. However, following the announcement of the closure of Yale-NUS College (National University of Singapore 2021, The Economist 2021), substantial changes were made to the college's original curriculum structure, which rendered 'Data Analysis and Visualisation' defunct. The purpose of this article is to document the insights and experiences gained from incorporating data visualisation into a liberal arts education while Yale-NUS College was still operational.

2 Related Work

While business experts predict that liberal arts students will soon be in demand for data visualisation jobs (Clayton 2016; Hillery 2019), few educators have written about the pedagogical aspects of data visualisation courses specifically tailored for liberal arts colleges.

Owen et al. (2013) discussed the course 'Visualization: Data, Pixels, and Ideas' at Yale University (USA) as an example of a course designed for liberal arts education. Because students came from diverse educational backgrounds, programming was not a prerequisite; however, the course used a variety of software tools, including Excel, Matlab and a Virtual Reality Modeling Language viewer. Applications were mostly based on data sets from the humanities (e.g. politics or history). As a sample application, Owen et al. (2013) showed a digitised historical map of Dublin highlighting geo-referenced events in James Joyce's novel 'Ulysses'.

Myers (2021) described a similar project using map making in the humanities. For a course on ancient travel, geography, and ethnography at Kenyon College (USA), students worked on a project titled 'Mapping Cicero's Letters'. During the project, students identified place names mentioned in Cicero's letters, stored the information (e.g. latitude and longitude) in spreadsheets using Microsoft Excel and created digital visualisations using the web application Carto.

In the natural sciences, Sharma (2017) outlined a model course in scientific computing at Wagner College (USA). The syllabus included elementary statistical charts (e.g. bar charts and histograms) produced by Mathematica and molecular visualisation produced by the open-source software Avogadro. Sharma (2017) argued that similar courses should be routinely offered by liberal arts colleges to improve students' technological competency and career prospects.

To understand the confidence of students and faculty in their own technological competency, Kumar (2020) analysed data from semi-structured interviews at Symbiosis School for Liberal Arts (India). Students complained about a lack of education in data visualisation software (e.g. Tableau), while faculty felt that they lacked expertise to teach in this area. To remedy the situation, Kumar (2020) recommended that Symbiosis School offer more practical training for both students and faculty.

3 Choice of Software

Given the importance of practical training, it is crucial to choose an appropriate software environment when planning a data visualisation course.

Spreadsheet programs (e.g. Excel or Google Sheets) are easy to learn; however, their charts are limited to only a few standard types and are challenging to customise. Applications with graphical user interfaces (e.g. Tableau or Qlik) offer more customisation options for charts, but using them to build charts algorithmically is not straightforward. Current versions of geographic information systems (e.g. ArcGIS and QGIS) allow for algorithmic control using Python scripts; however, these systems are primarily designed for creating maps rather than general statistical charts. An alternative is the direct use of general-purpose programming languages such as Python, R or Julia. These three languages are free of charge, open-source and platform-independent, and each of them has specialised packages for data visualisation (e.g. Python's matplotlib) that can produce professional-quality charts.

For 'Data Analysis and Visualisation', the final choice of programming language was R for the following reasons:
- R is widely used by professional data scientists. Thus, learning R can lead to attractive employment opportunities for students after graduation.

vještine zahtijevaju od studenata da pohađaju kolegije iz humanističkih, društvenih i prirodnih znanosti umjesto da se rano usredotoče na jedan predmet. Stoga sam, kao inicijator i voditelj razvoja "Analize podataka i vizualizacije", imao za cilj stvoriti interdisciplinarni tečaj koji potiče kreativno računalno razmišljanje uključujući osnovno razumijevanje vizualizacije geoprostornih podataka. Kartografske vještine, prenesene studentima kroz praktične zadatke izrade karte, sadržavale su sposobnost odabira prikladnih kartografskih projekcija, odabira odgovarajućih kartografskih znakova i stvaranja učinkovitih tumača znakova. Predmet se predavao pet puta između 2016. i 2022. godine. Međutim, nakon najave zatvaranja Yale-NUS College (Nacionalno sveučilište u Singapuru 2021, The Economist 2021) napravljene su značajne promjene u izvornoj strukturi kurikuluma koledža, što je rezultiralo ukidanjem predmeta "Analiza podataka i vizualizacija". Svrha je ovoga članka dokumentirati uvide i iskustva stečena uključivanjem vizualizacije podataka u obrazovanje slobodnih vještina dok je Yale-NUS College još bio operativan.

2. Što su drugi pisali o toj temi

Dok poslovni stručnjaci predviđaju da će studenti liberalnih vještina uskoro biti traženi za poslove vizualizacije podataka (Clayton 2016, Hillery 2019), malo je nastavnika pisalo o pedagoškim aspektima tečajeva vizualizacije podataka posebno skrojenih za koledže liberalnih vještina.

Owen i sur. (2013) raspravljali su o tečaju "Vizualizacija: podatci, pikseli i ideje" na Sveučilištu Yale (SAD) kao primjeru tečaja osmišljenog za obrazovanje slobodnih vještina. Budući da su studenti dolazili iz različitih obrazovnih sredina, programiranje nije bilo preduvjet. Međutim, tečaj je koristio razne softverske alate, uključujući Excel, Matlab i preglednik Virtual Reality Modeling Language. Prijave su se uglavnom temeljile na skupovima podataka iz humanističkih znanosti (npr. politika ili povijest). Kao primjer prijave, Owen et al. (2013) pokazali su digitaliziranu povijesnu kartu Dublina ističući georeferencirane događaje u romanu Jamesa Joycea "Uliks".

Myers je (2021) opisao sličan projekt koristeći izradu karata u humanističkim znanostima. Za kolegij o drevnim putovanjima, geografiji i etnografiji na Kenyon Collegeu (SAD) studenti su radili na projektu pod nazivom "Kartiranje Ciceronovih pisama". Tijekom projekta studenti su identificirali imena mjesta koja se spominju u Ciceronovim pismima, pohranili podatke (npr. geografsku širinu i dužinu) u

proračunske tablice Microsoft Excela i s pomoću web aplikacije Carto izradili digitalne vizualizacije.

U prirodnim je znanostima Sharma (2017) opisao model tečaja znanstvenog računarstva na Wagner Collegeu (SAD). Nastavni plan i program sadržavao je osnovne statističke grafikone (npr. stupčaste dijagrame i histograme) koje je izradila Mathematica i molekularnu vizualizaciju koju je izradio softver otvorenog koda Avogadro. Sharma (2017) tvrdi da bi koledži liberalnih vještina trebali rutinski nuditi slične tečajeve kako bi se poboljšala tehnološka kompetencija studenata i izgledi za karijeru.

Da bi razumio povjerenje studenata i nastavnika u vlastitu tehnološku kompetenciju, Kumar je (2020) analizirao podatke iz polustrukturiranih intervjua u Symbiosis School for Liberal Arts (Indija). Studenti su se žalili na nedostatak obrazovanja o softveru za vizualizaciju podataka (npr. Tableau), dok su nastavnici smatrali da im nedostaje stručnost za podučavanje u tom području. Da bi se situacija poboljšala, Kumar je (2020) preporučio da Symbiosis School za studente i nastavno osoblje ponudi više praktične obuke.

3. Izbor softvera

S obzirom na važnost praktične obuke, prilikom planiranja tečaja vizualizacije podataka ključno je odabrati odgovarajuće softversko okruženje.

Programe za proračunske tablice (npr. Excel ili Google Sheets) lako je naučiti. Međutim, njihove su karte ograničene na samo nekoliko standardnih tipova i teško ih je prilagoditi. Aplikacije s grafičkim korisničkim sučeljima (npr. Tableau ili Qlik) nude više mogućnosti prilagodbe za grafikone, ali njihovo korištenje za algoritamsku izradu grafikona nije jednostavno. Trenutačne verzije geoinformacijskih sustava (npr. ArcGIS i QGIS) omogućuju algoritamsku kontrolu upotrebom skripti Pythona. Međutim, ti su sustavi prvenstveno dizajnirani za izradu karata, a ne općih statističkih grafikona. Alternativa je izravna uporaba programskih jezika opće namjene kao što su Python, R ili Julia. Ta su tri jezika besplatna, otvorenoga koda i neovisna o platformi, a svaki od njih ima specijalizirane pakete za vizualizaciju podataka (npr. Pythonov matplotlib) koji mogu proizvesti grafikone profesionalne kvalitete.

Za "Analizu podataka i vizualizaciju" R je bio konačni izbor programskog jezika iz sljedećih razloga:
- R naširoko koriste znanstvenici koji se bave podatcima. Stoga učenje R može dovesti do atraktivnih mogućnosti zapošljavanja studenata nakon diplome.
- Kolekcija R paketa "tidyverse" ima dosljednu sintaksu jednostavnu za upotrebu i širok raspon paketa za

Table 1 Learning objectives and categorisation according to Bloom's revised taxonomy (Anderson and Krathwohl 2001).
Tablica 1. Ciljevi učenja i kategorizacija prema Bloomovoj revidiranoj taksonomiji (Anderson i Krathwohl 2001).

Objective / Cilj	Category / Kategorija
Acquire the skill of importing data from various file formats (e.g. CSV, XLSX or JSON) / Steći vještinu uvoza podataka iz različitih formata datoteka (npr. CSV, XLSX ili JSON)	Remember / Zapamtiti
Devise algorithmic strategies for tidying, analysing and visualising statistical data / Osmisliti algoritamske strategije za pospremanje, analizu i vizualizaciju statističkih podataka	Understand / Razumjeti
Implement algorithmic strategies through stylistically refined R code using tidyverse syntax / Primijeniti algoritamske strategije s pomoću stilski profinjenog R koda upotrebom sintakse tidyverse	Apply / Primijeniti
Interpret R's documentation, find additional help online and incorporate suggested solutions into your own code / Protumačiti dokumentaciju R-a, pronaći dodatnu pomoć na internetu i uključiti predložena rješenja u svoj kod	Analyse / Analizirati
Produce visually appealing graphics that effectively convey information / Izraditi vizualno privlačnu grafiku koja učinkovito prenosi informacije	Create / Kreirati
Document code and outputs with accompanying text using R Markdown / Dokumentirati kod i izlaze s popratnim tekstom upotrebom R Markdowna	Understand / Razumjeti
Review code written by other students and give constructive written feedback / Pregledati kôd koji su napisali drugi studenti i dati konstruktivne pisane povratne informacije	Evaluate / Procijeniti

- The 'tidyverse' collection of R packages features a consistent, user-friendly syntax and a broad range of packages for various data science tasks, from data import and cleaning to visualisation and modelling.
- The core visualisation package in the tidyverse, ggplot2 (Wickham 2016), is an implementation of Wilkinson's (2005) 'grammar of graphics'. The grammar provides a consistent and structured approach to data visualization that emphasizes the underlying principles, rather than memorizing specific chart types, such as scatter plots or box plots.
- The add-on package sf (Pebesma 2018) enables advanced geocomputation using R (Lovelace et al. 2019). Another add-on package, tmap (Tennekes 2018), simplifies the creation of thematic maps by extending Wilkinson's (2005) grammar.
- The prerequisite course 'Quantitative Reasoning' taught basic elements of R using the integrated development environment RStudio, which provides a user-friendly interface for writing, executing and documenting R code. RStudio also supports R Markdown, a document format that combines text, R code and code outputs (e.g. plots) into a single document. As part of the course assignments, students were required to submit R Markdown code, along with a rendered PDF or HTML document.

Instructors who are more familiar with Python than R may consider using the Python package plotnine (Kibirige et al. 2022), which emulates the syntax of ggplot2. Similarly, the Julia package Gadfly.jl (Jones et al. 2021), which is heavily influenced by ggplot2, allows changing the programming language from R to Julia. However, Gadfly.jl is still new; thus, it does not yet offer the same functionality as ggplot2. Furthermore, since the last offering of 'Data Analysis and Visualisation', R Markdown has been superseded by Quarto, which supports code written in R, Python and Julia. Thus, any of these three programming languages could now be used as the basis for the course.

4 Learning Objectives

'Data Analysis and Visualisation' was designed for second-year students at Yale-NUS College. Students had not previously encountered ggplot2 and the tidyverse syntax; thus, 'Data Analysis and Visualisation' needed to teach these topics in a self-contained manner. The aim was to build upon the students' prior R knowledge and provide them with the tools to present real-world data from various disciplines in informative, elegant and structured ways. To prepare students for their careers after graduation and encourage critical thinking about their own code and plots, they were expected to search for help online and critique the work submitted by other students.

različite zadatke iz znanosti o podatcima, od uvoza i čišćenja podataka do vizualizacije i modeliranja.

- Temeljni vizualizacijski paket u tidyverseu, ggplot2 (Wickham 2016), primjena je Wilkinsonove (2005) "gramatike grafike". Gramatika pruža dosljedan i strukturiran pristup vizualizaciji podataka koji naglašava temeljna načela umjesto pamćenja određenih vrsta grafikona, kao što su raspršeni dijagrami ili dijagrami pravokutnika (box plots).
- Dodatni paket sf (Pebesma 2018) omogućuje napredno georačunanje s pomoću R-a (Lovelace et al. 2019). Još jedan dodatni paket, tmap (Tennekes 2018), pojednostavljuje stvaranje tematskih karata proširujući Wilkinsonovu (2005) gramatiku.
- Preduvjetni tečaj "Kvantitativno rasuđivanje" podučavao je osnovne elemente R-a upotrebom integriranog razvojnog okruženja RStudio koje pruža korisničko sučelje za pisanje, izvođenje i dokumentiranje R koda. RStudio također podržava R Markdown, format dokumenta koji kombinira tekst, R kod i izlaze koda (npr. crteže) u jedan dokument. Kao dio zadaća na kolegiju studenti su morali predati R Markdown kod, zajedno s renderiranim dokumentom u PDF-u ili HTML-u.

Instruktori koji su bolje upoznati s Pythonom nego s R-om mogu razmotriti upotrebu Pythonova paketa plotnine (Kibirige et al. 2022) koji emulira sintaksu ggplot2. Slično tome, Julia paket Gadfly.jl (Jones et al. 2021), koji je pod velikim utjecajem ggplot2, omogućuje promjenu programskog jezika s R na Julia. Međutim, Gadfly.jl je još uvijek nov, stoga još ne nudi istu funkcionalnost kao ggplot2. Nadalje, od zadnje je ponude "Analize podataka i vizualizacije" R Markdown zamijenjen Quartom koji podržava kod napisan u jezicima R, Python i Julia. Stoga se bilo koji od tih triju programskih jezika sada može upotrebljavati kao osnova za tečaj.

4. Ciljevi učenja

"Analiza podataka i vizualizacija" osmišljena je za studente druge godine Yale-NUS Collegea. Studenti se prije nisu susreli s ggplot2 i sintaksom tidyverse, stoga je "Analiza podataka i vizualizacija" trebala poučavati ove teme na samostalan način. Cilj je bio nadgraditi prethodno znanje studenata o R-u i pružiti im alate za prikazivanje podataka iz stvarnog svijeta iz različitih disciplina na informativan, elegantan i strukturiran način. Kako bi pripremili studente za njihove karijere nakon diplome i potaknuli kritičko razmišljanje o vlastitom kodu i zapletima, od njih se očekivalo da potraže pomoć na internetu i

kritiziraju radove drugih studenata. Predviđeni su ishodi učenja sadržavali ciljeve učenja navedene u tablici 1, a koji pokrivaju svih šest kategorija Bloomove revidirane taksonomije (Anderson i Krathwohl 2001): zapamtiti, razumjeti, primijeniti, analizirati, procijeniti i stvoriti.

5. Nastavni plan i program

Ciljevi učenja u tablici 1 prevedeni su u temeljni nastavni plan i program sa sedam modula:
I. Motivacija
II. Upoznavanje sa softverskim alatima
III. Strukture podataka
IV. Vizualizacija podataka pomoću ggplot2
V. Pročišćavanje neobrađenih podataka s pomoću sintakse tidyverse
VI. Mrežni podatci
VII. Geoprostorni podatci.

Svaki je od tih modula obuhvaćao od jedne do četiri lekcije, kao što je naznačeno u tablici 2. Modul I sadržavao je raspravu o svrsi vizualizacije podataka i najboljim praksama koristeći primjere iz medija. Nakon toga, svaka je lekcija započela kratkom raspravom o primjeru vizualizacije podataka koji su odabrali studenti. Počevši od uvoda R Markdowna u Lekciji 3 Modula II, studenti su mogli kopirati i zalijepiti ggplot2 kod u vlastite dokumente. Međutim, lekcije o tome kako izraditi dijagrame pomoću ggplot2 započele su tek u Modulu IV nakon što su studenti stekli solidne temelje u R-u i R Markdownu.

Modul V vratio se tehničkom sadržaju o jeziku R omogućujući studentima da razviju projekte vizualizacije temeljene na neurednim podatcima kakvi se obično susreću u skupovima podataka iz stvarnog svijeta. Na primjer, jedan zadatak tijekom modula V zahtijevao je od studenata da preuzmu podatke iz Odjela za statistiku Singapura (2022) i kreiraju niz populacijskih piramida kao što je prikazano na slici 1. Uz fasetiranje, koje je već obrađeno u Modulu IV, studenti su morali primijeniti nekoliko operacija dplyr (npr. pivotiranje) kako bi transformirali neobrađene podatke u konačni dijagram. Koraci su bili navedeni u zadatku, ali su studenti morali napisati vlastiti kod.

Moduli VI i VII bavili su se podatcima koji zahtijevaju alate za vizualizaciju izvan ggplot2: paketi tidygraph i ggraph za mrežne podatke, te sf, stars i tmap za geoprostorne podatke. Iako su mreže i karte ključni za analitiku podataka, moduli VI i VII mogu se izostaviti ako je vrijeme ograničeno.

U semestru od 13 tjedana temeljni je nastavni plan ostavio dovoljno vremena za dodavanje više

Table 2 Core syllabus. Each lesson lasted for 110 minutes.
Tablica 2. Temeljni nastavni plan i program. Svaki je sat trajao 110 minuta.

Module Modul	Title Naslov	Lesson Lekcija	Topic Tema
I I	Motivation Motivacija	1 1	Discussing best practices in data visualisation. Demonstration of R, RStudio and R Markdown. Rasprava o najboljim praksama u vizualizaciji podataka. Demonstracija R, RStudio i R Markdown.
II	Getting familiar with the software tools	1	R vectors. Numeric, integer, character and logical data types. Data type conversion.
II	Upoznavanje sa softverskim alatima	1	R vektori. Numerički, cjelobrojni, znakovni i logički tipovi podataka. Pretvorba tipa podataka.
		2	Getting help from the R console and the world-wide web. Setting arguments of built-in R functions. Writing R scripts and functions.
		2	Dobivanje pomoći od R konzole i weba. Postavljanje argumenata ugrađenih R funkcija. Pisanje R skripti i funkcija.
		3	Installing and loading packages. Creating documents with R Markdown.
		3	Instalacija i učitavanje paketa. Izrada dokumenata s R Markdownom.
		4	Logical operators. Conditional statements (`if_else`, `case_when` and `case_match`). Missing and undefined values.
		4	Logički operatori. Uvjetne naredbe (`if_else`, `case_when` and `case_match`). Vrijednosti koje nedostaju ili nisu definirane.
III III	Data structures Strukture podataka	1 1	Factors. Faktori.
		2	Data frames and tibbles. Importing spreadsheet data as a tibble.
		2	Podatkovni okviri i tablice (frames and tibbles) znakovi. Uvoz podataka proračunske tablice kao tibble.
IV	Data visualisation using ggplot2	1	Exploring ggplot2 using real-world examples. Terminology of ggplot2: geoms, scales and c ordinates. Setting titles for plots and legends.
IV	Vizualizacija podataka upotrebom ggplot2	1	Istraživanje ggplot2 s pomoću primjera iz stvarnoga svijeta. Terminologija ggplot2: geomi, mjerila i koordinate. Postavljanje naslova za crteže i tumače znakova.
		2	Individual versus collective geoms. Visualising statistical summaries. Annotations. Colour scales
		2	Individualni naspram kolektivnih geoma. Vizualizacija statističkih sažetaka. Bilješke. Ljestvice boja.
		3	Facets. Themes in ggplot2.
		3	Fasetiranje. Teme u ggplot2.
V	Refining raw data using tidyverse syntax	1	Data transformation using the dplyr package.
V	Pročišćavanje neobrađenih podataka s pomoću sintakse tidyverse	1	Transformacija podataka s pomoću paketa dplyr.
		2	Pivoting to produce 'tidy' data.
		2	Pivotiranje za dobivanje "urednih" podataka.
		3	Joining data from different tibbles.
		3	Spajanje podataka iz različitih tablica.
		4	Lists. Functional programming tools in the purrr package.
		4	Popisi. Alati za funkcionalno programiranje u paketu purrr.
VI VI	Network data Mrežni podatci	1 1	Importing network data. Network analysis using the tidygraph package. Uvoz mrežnih podataka. Analiza mreže s pomoću paketa tidygraph.
		2	Network visualisation using the ggraph package.
		2	Vizualizacija mreže upotrebom paketa ggraph.
VII VII	Geospatial data Geoprostorni podatci	1 1	Geocomputation using the sf package. Choropleth maps using the tmap package. Georačunanje s pomoću paketa sf. Koropletne karte upotrebom paketa tmap.
		2	Working with raster data using the stars package. Raster maps using the tmap package.
		2	Rad s rasterskim podatcima s pomoću paketa stars. Rasterske karte s pomoću paketa tmap.
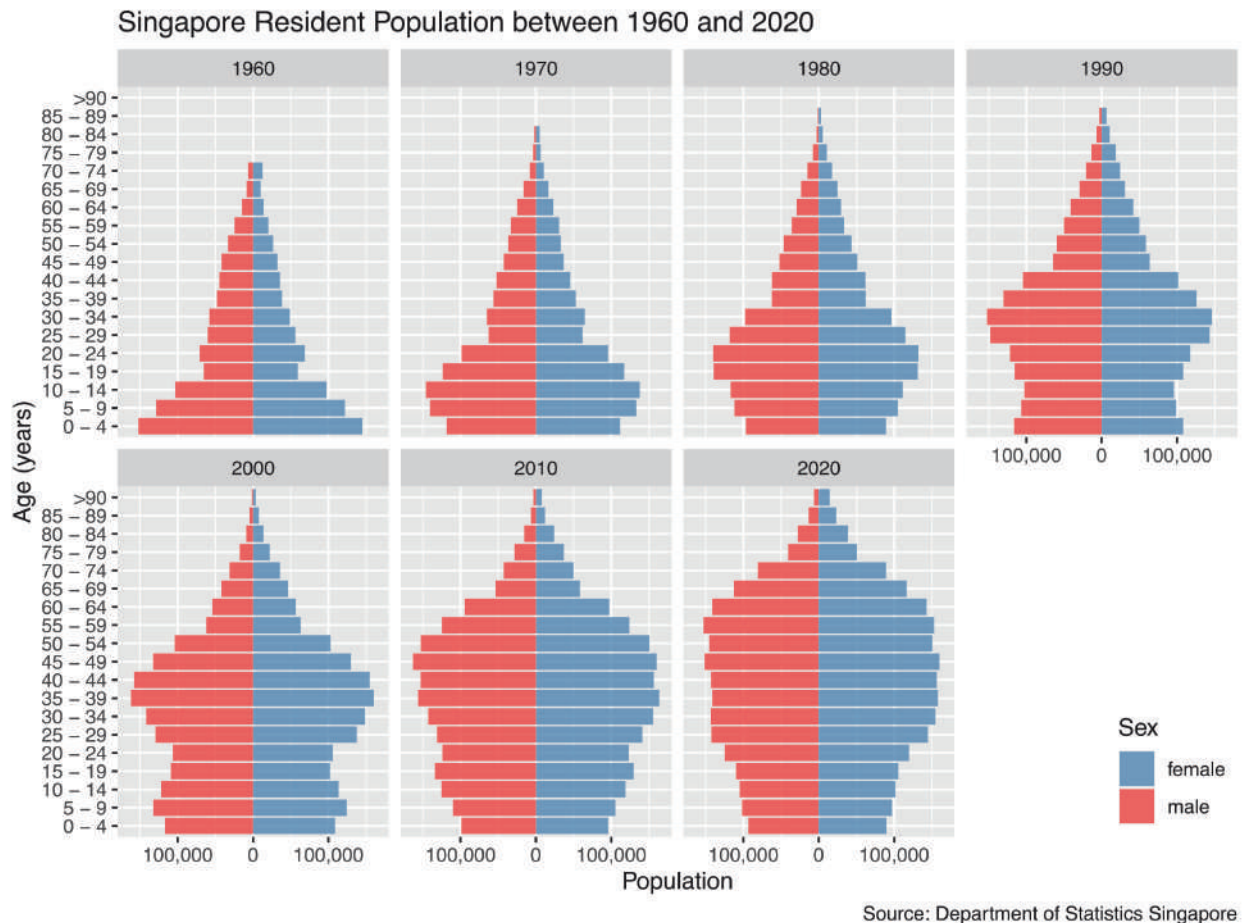
Slika 1. Uzorak vizualizacije koji će izraditi studenti tijekom Modula V: niz piramidalnih dijagrama temeljen na podatcima popisa stanovništva u Singapuru.
Fig. 1 Sample visualisation to be created by students during Module V: a sequence of pyramid plots based on Singapore census data.

tema na kraju. Tijekom godina bavio sam se sljedećim temama:

- Trokutasti grafikoni
- Trodimenzionalni crteži
- Integracija grafike (npr. sparkline) u tablice
- Analiza teksta i oblaci riječi
- Interaktivni crteži sa zumiranjem, pomicanjem i informativnim savjetima
- Animacije
- Web aplikacije koje koriste paket Shiny.

Ako vrijeme dopušta, druga je mogućnost proširiti Modul VII kako bi sadržavao više vrsta tematskih karata (npr. dasimetrijske karte ili kartogrami), karte geoprostornih mreža (npr. ceste ili javni prijevoz) ili karte temeljene na webu (npr. pomoću ggmapa ili letka). U kasnijim verzijama tečaja studenti su također imali priliku raditi na temi koju su sami odabrali tijekom svog timskog projekta; među njihovim su izborima bili financijski grafikoni koji koriste

quantmod i crteži spremni za objavljivanje s pomoću ggpubra. Alternativno, kolegij bi mogao pokriti literaturu o psihofizici vizualizacije podataka (npr. Brewer 1994) i imati timove studenata koji će provoditi manje eksperimente (npr. za usporedbu učinkovitosti različitih paleta boja) koristeći kolege kao sudionike. U tom bi slučaju kolegij o eksperimentalnom dizajnu mogao biti neophodan preduvjet.

Materijal za lektiru za kolegij bile su knjige Wickhama (2016) za modul IV i Wickhama i Grolemunda (2017) za modul V. Za sve sam ostale module studentima osigurao materijale za čitanje koje su sami napisali. Trenutačno prikupljam svoje materijale za objavljivanje udžbenika.

6. Odgojno-obrazovni proces

U skladu s naglaskom na interakciju student - student i student - nastavnik u obrazovanju

The intended learning outcomes involved the learning objectives listed in Table 1, which cover all six categories of Bloom's revised taxonomy (Anderson and Krathwohl 2001): remember, understand, apply, analyse, evaluate, and create.

5 Syllabus

The learning objectives in Table 1 were translated into a core syllabus with seven modules:

I. Motivation
II. Getting familiar with the software tools
III. Data structures
IV. Visualising data using ggplot2
V. Refining raw data using tidyverse syntax
VI. Network data
VII. Geospatial data

Each of these modules was covered over one to four lessons, as indicated in Table 2. Module I involved a class discussion on the purpose of data visualisation and best practices, using examples from the media. Afterwards, each lesson began with a brief class discussion on a data visualisation example selected by the students. Starting from the introduction of R Markdown in Lesson 3 of Module II, students were able to copy and paste ggplot2 code into their own documents. However, lessons on how to create plots with ggplot2 only commenced in Module IV after students had acquired solid foundations in R and R Markdown.

Module V returned to technical content about the R language, enabling students to develop visualisation projects based on untidy data, as typically encountered in real-world data sets. For instance, one assignment during Module V required students to download data from the Department of Statistics Singapore (2022) and create a sequence of population pyramids as shown in Figure 1. In addition to faceting, which was already covered in Module IV, students had to apply several dplyr operations (e.g. pivoting) to transform raw data into the final plot. The steps were outlined in the assignment, but students had to write their own code.

Modules VI and VII dealt with data that required visualisation tools beyond ggplot2: the tidygraph and ggraph packages for network data, and sf, stars and tmap for geospatial data. Although networks and maps are essential for data analytics, Modules VI and VII can be omitted if time is limited.

In a 13-week semester, the core syllabus left enough time to add more topics at the end. Over the years, I have covered the following topics:

- Ternary plots
- Three-dimensional plots
- Integration of graphics (e.g. sparklines) into tables
- Text analysis and word clouds
- Interactive plots with zooming, panning and infotips
- Animations
- Web applications using the Shiny package

If time permits, another option is to expand Module VII to include more types of thematic maps (e.g. dasymetric maps or cartograms), maps of geospatial networks (e.g. roads or public transport) or web-based maps (e.g. using ggmap or leaflet). In later versions of the course, students also had the opportunity to work on a self-selected topic during their team project; among their choices were financial charts using quantmod and publication-ready plots using ggpubr. Alternatively, the course could cover literature on the psychophysics of data visualisation (e.g. Brewer 1994) and have teams of students conduct small-scale experiments (e.g. to compare the effectiveness of different colour palettes) using classmates as participants. In that case, a course on experimental design may be necessary as a prerequisite.

Reading material for the course were the books by Wickham (2016) for Module IV and by Wickham and Grolemund (2017) for Module V. For all other modules, I provided self-written reading materials to the students. I am currently collating my materials for publication as a textbook.

6 Pedagogy

In line with the emphasis on student-student and student-faculty interaction in liberal arts education (Blaich et al. 2004), in-class exercises were integrated into the design of 'Data Analysis and Visualisation' to encourage collaboration and engagement. During the early stages of course development, lectures provided support and structure for the in-class exercises. However, based on student feedback, subsequent iterations gradually shifted away from lecturing towards more hands-on activities and a 'flipped classroom' pedagogy (Mok 2014). This approach involved students reading or watching preparatory materials and taking an online multiple-choice quiz before attending class, freeing up nearly the entire 110-minute class period for discussions and activities. The students responded positively to this teaching style, rating it 4.6 or higher out of 5 on all criteria tested in the course evaluation (Table 3).

Before the COVID-19 pandemic, the pedagogy also incorporated elements of team-based learning

slobodnih vještina (Blaich et al. 2004), vježbe na nastavi integrirane su u dizajn "Analize podataka i vizualizacije" kako bi se potaknula suradnja i angažman. Tijekom ranih faza razvoja kolegija predavanja su pružala podršku i strukturu za vježbe u nastavi. Međutim, na temelju povratnih informacija studenata, naknadne su se iteracije postupno udaljile od predavanja prema više praktičnih aktivnosti i pedagogiji "okrenute učionice (flipped classroom)" (Mok 2014).

Pri tom su pristupu studenti čitali ili gledali pripremne materijale i prije pohađanja nastave rješavali online kviz s višestrukim odgovorima omogućujući tako da gotovo cijeli sat od 110 minuta bude slobodan za rasprave i aktivnosti. Studenti su pozitivno odgovorili na taj stil poučavanja ocijenivši ga ocjenom 4,6 ili više od maksimalno 5 prema svim kriterijima testiranima u evaluaciji kolegija (tablica 3).

Prije pandemije COVID-19 odgojno-obrazovni proces je također sadržavao elemente timskog učenja (Michaelsen i Sweet 2008) sa studentima koji su formirali fiksne timove od 4 do 6 osoba za cijeli semestar. Svaki se tim sastojao od mješavine singapurskih i stranih studenata, muškaraca i žena, te studenata s različitim razinama iskustva. Zajedno su radili na vježbama primjene, kvizovima u nastavi i projektu na kraju semestra. Kako bi se pravedno procijenio pojedinačni doprinos, na kraju semestra provedeno je anonimno ocjenjivanje suigrača. Ocjene unutar svakog tima, koje je dao svaki član ekipe, standardizirane su kako bi se destimuliralo davanje jednakih bodova svima ili pretjerano visoke ili niske ocjene.

Prve povratne informacije studenata pokazale su pozitivan stav prema primijenjenim elementima timskog učenja. Razredno je okruženje percipirano kao suradničko, pri čemu su slabiji studenti dobivali smjernice od boljih, a bolji su studenti produbljivali svoje razumijevanje objašnjavajući svoje metode slabijim studentima.

Mjere sigurnog distanciranja zbog pandemije COVID-19 bile su izazov za timski rad. Nadalje, neki strani studenti nisu mogli osobno prisustvovati zbog ograničenja putovanja, što je otežavalo suradnju na kodiranju. Iako je okruženje za suradničko uređivanje u stvarnom vremenu temeljeno na oblaku (RStudio Workbench) bilo dostupno, bilo je sporo kada ga je istovremeno koristilo više studenata. Dodatno, studenti su iskusili tjeskobu uzrokovanu pandemijom i predosjećajem zatvaranja koledža. U tim su uvjetima lako bili frustrirani slabim rezultatima na "testu osiguranja individualne spremnosti" kojim je započinjao svaki sat slijedeći preporučene timske prakse učenja (Michaelsen i Sweet 2008). Kako bi se

riješio taj problem, posljednja je iteracija kolegija pretvorila testove osiguranja spremnosti u pojedinačne online kvizove s višestrukim pokušajima odgovora smanjujući anksioznost studenata i oslobađajući vrijeme nastave za rasprave i praktične aktivnosti.

Stručno ocjenjivanje unutar timova bilo je dodatni izvor tjeskobe za studente koji su izostajali zbog bolesti ili ograničenja putovanja jer su se bojali da bi njihovi timski kolege mogli nepravedno ocijeniti njihov doprinos. Kao odgovor na to, recenziranje je eliminirano u konačnoj verziji kolegija. Međutim, ta je izmjena dovela do pritužbi o nejednakom doprinosu timskim zadacima. Na temelju ovog iskustva, slažem se s Cestoneom i sur. (2008) da je kolegijalno ocjenjivanje važno za predmete koji uključuju značajan timski rad, ali studenti moraju biti spremni na davanje i primanje povratnih informacija.

7. Ocjenjivanje

Ocjenjivanje tečaja vagalo je pojedinačne i timske doprinose otprilike u omjeru 2 : 1. Sljedeći je popis reprezentativan za shemu ocjenjivanja u kasnijim ponudama tečaja:
- Individualna ocjena (65% od ukupnoga):
  - 10%: Individualni test osiguranja spremnosti
  - 10%: Procjena kolega iz tima
  - 5%: Sudjelovanje I: Dolazak na satove kolegija
  - 5%: Participation II: Pisane primjedbe za poboljšanje nastavnih materijala
  - 15%: Međuispit
  - 20%: Konačni ispit
- Timska ocjena (35% od ukupnoga):
  - 15%: Rješenja vježbi primjene
  - 5%: Povratne informacije o rješenjima drugih timova
  - 15%: Konačni projekt

7.1. Testovi osiguranja spremnosti

Da bi se pripremili za svaku lekciju, studenti su morali pročitati zadano i, povremeno, pogledati videolekciju. Da bi procijenili svoje razumijevanje, pristupali su 10-minutnom individualnom testu osiguranja spremnosti (individual readiness assurance test – IRAT) u obliku kviza od pet pitanja višestrukog izbora. U prvoj su ponudi kolegija odgovori prikupljani u razredu. U sljedećim trima ponudama umjesto toga je korišten softver za timsko učenje InteDashboard, a nakon IRAT-a, prema preporuci Michaelsena i Sweeta (2008), odmah je uslijedio test osiguranja spremnosti tima (team readiness assurance test – TRAT).

(Michaelsen and Sweet 2008), with students forming fixed teams of 4 to 6 for the entire semester. Each team consisted of a mix of Singaporean and foreign students, males and females, and students with varying levels of experience. They worked together on application exercises, in-class quizzes and an end-of-semester project. To fairly assess individual contributions, anonymous peer evaluations of teammates were conducted at the end of the semester. Scores within each team and given by each team member were standardised to disincentivise giving equal scores to everybody or excessively high or low scores.

Early student feedback indicated a positive attitude towards the implemented team-based learning elements. The class environment was perceived as cooperative, with weaker students receiving guidance from stronger ones and stronger students deepening their understanding by explaining their methods to weaker students.

Safe distancing measures due to the COVID-19 pandemic presented challenges for in-class teamwork. Furthermore, some foreign students were unable to attend in person because of travel restrictions, which made collaboration on coding difficult. Although a cloud-based real-time collaborative editing environment (RStudio Workbench) was available, it was slow when used by multiple students simultaneously. Additionally, students experienced anxiety caused by the pandemic and the premonition of the college's closure. Under these conditions, they were easily frustrated by low scores on the 'individual readiness assurance test', which started each lesson, following recommended team-based learning practices (Michaelsen and Sweet 2008). To address this issue, the last iteration of the course converted the readiness assurance tests to individual online quizzes with multiple answer attempts, reducing student anxiety and freeing up class time for discussions and hands-on activities.

Peer review within teams was an additional source of anxiety for students who were absent due to illness or travel restrictions because they feared that their teammates might rate their contributions unfairly. In response, peer review was eliminated in the final version of the course. However, this modification led to complaints about unequal contributions to team assignments. Based on this experience, I concur with Cestone et al. (2008) that peer review is important for courses involving substantial teamwork, but students must be prepared for giving and receiving feedback.

7 Assessment

The course assessment weighed individual and team contributions approximately in a 2:1 ratio. The following list is representative of the grading scheme in later offerings of the course:

- Individual assessment (65% of total):
  - 10%: Individual readiness assurance test
  - 10%: Peer assessment by teammates
  - 5%: Participation I: Visit to office hour
  - 5%: Participation II: Written suggestions for improvements of reading materials
  - 15%: Midterm exam
  - 20%: Final exam
- Team assessment (35% of total):
  - 15%: Solutions to application exercises
  - 5%: Feedback on solutions by other teams
  - 15%: Final project

7.1 Readiness Assurance Tests

To prepare for each lesson, students had to complete a reading assignment and, occasionally, a video tutorial. To assess their understanding, they took a closed-book, 10-minute 'individual readiness assurance test' (IRAT) in the form of a multiple-choice quiz with five questions. In the first offering of the course, responses were collected in class using clickers. In the next three offerings, the team-based learning software InteDashboard was used instead, and the IRAT was immediately followed by a 'team readiness assurance test' (TRAT), as recommended by Michaelsen and Sweet (2008). Teams had to agree on a single answer and, then, received feedback on whether they had the correct answer. If not, teams were permitted to retake the question for a reduced score.

The TRATs promoted lively discussions among team members. However, when questions were complex, teams often resorted to guessing, leading to additional class time spent on explanations of edge cases or obscure features of R. As a result, the final offering of the course made changes to the assessment process. All IRATs were moved to before class, TRATs were dropped, and the questions were simplified. The time gained at the start of each lesson was instead used for a class discussion about a student-selected example of data visualisation in the media.

7.2 Application Exercises

After the TRAT or class discussion, teams were given exercises to apply what they had read before

Tablica 3. Završne ocjene kolegija "Analiza podataka i vizualizacija" upotrebom Likertove ljestvice od 5 stupnjeva (1 = uopće se ne slažem, 5 = u potpunosti se slažem).
Table 3 Final course evaluations of 'Data Analysis and Visualisation' using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

Statement / Tvrdnja	Agreement / Slaganje
The course was effectively organized. Kolegij je bio učinkovito organiziran.	4.69
The course clearly articulated objectives and expectations. Kolegij je jasno artikulirao ciljeve i očekivanja.	4.85
The course assignments helped me learn. Zadaci na kolegiju pomogli su mi u učenju.	4.77
The instructor helped me understand course materials and concepts. Instruktor mi je pomogao razumjeti materijale i koncepte tečaja.	4.62
The instructor encouraged student participation and/or engagement. Instruktor je poticao sudjelovanje i/ili angažman studenata.	4.62
The instructor provided constructive feedback on my work. Instruktor je dao konstruktivne povratne informacije o mojem radu.	4.77
The instructor was helpful outside of class. Instruktor je pomagao i izvan nastave.	4.69

Timovi su se morali složiti oko pojedinog odgovora, a zatim su dobili povratnu informaciju o tome je li odgovor bio točan. Ako nije, timovima je bilo dopušteno ponovno rješavati pitanje za smanjeni rezultat.

TRAT-ovi su poticali živahne rasprave među članovima tima. Međutim, kada su pitanja bila složena, timovi su često pribjegavali pogađanju, što je dovelo do dodatnog vremena provedenog u razredu na objašnjenja rubnih slučajeva ili opskurnih značajki R-a. Kao rezultat toga, konačna je ponuda kolegija unijela promjene u proces ocjenjivanja. Svi IRAT-ovi su premješteni prije nastave, TRAT-ovi su izbačeni, a pitanja su pojednostavljena. Vrijeme dobiveno na početku svake lekcije umjesto toga iskorišteno je za raspravu o primjeru vizualizacije podataka u medijima koji su odabrali studenti.

7.2. Vježbe primjene

Nakon TRAT-a ili rasprave, timovi su dobili vježbe za primjenu onoga što su pročitali prije nastave. Kao što je uobičajeno na koledžima slobodnih vještina, prosječan je broj studenata po sekciji bio oko 20, što mi je omogućilo da posjetim svaki tim i odgovorim na pitanja dok rade na problemima. Kako je kolegij sazrijevao, svaka je vježba za primjenu bila strukturirana oko koherentnog narativa, ali podijeljena u 3 do 4 odvojena materijala (handouts). Prvi je materijal predstavljao temu i uključivao prvih nekoliko zadataka; drugi materijal je pokazao isti tekst kao prvi, ali je dodao više zadataka itd. Prije nego što sam podijelio sljedeći materijal svim timovima, nakratko sam zastao kako bih im dao priliku da zamole mene ili druge timove za pomoć. Kompletan je skup problema objavljen pred kraj lekcije, a svaki je tim trebao predati pisana rješenja unutar jednog tjedna. U konačnoj sam ponudi kolegija dao sumirane povratne informacije o rješenjima nakon lekcija s neparnim brojevima, a svaki je tim dao formativne povratne informacije drugom nasumično dodijeljenom timu nakon lekcija s parnim brojevima. Timovi su dobili rubrike za formativne povratne informacije i dobili su ocjene za dovršetak pregleda.

7.3. Završni timski projekt

Za završni je projekt svaki tim odabrao vizualizacijski paket R-a koji nije uključen u temeljni nastavni program. Cilj je bio izraditi kratki vodič s objašnjenjem značajki paketa i popratnu vježbu s primjenom te to dati drugom nasumično dodijeljenom timu. Radi jasnoće, autorski ćemo tim nazvati Tim A, a tim
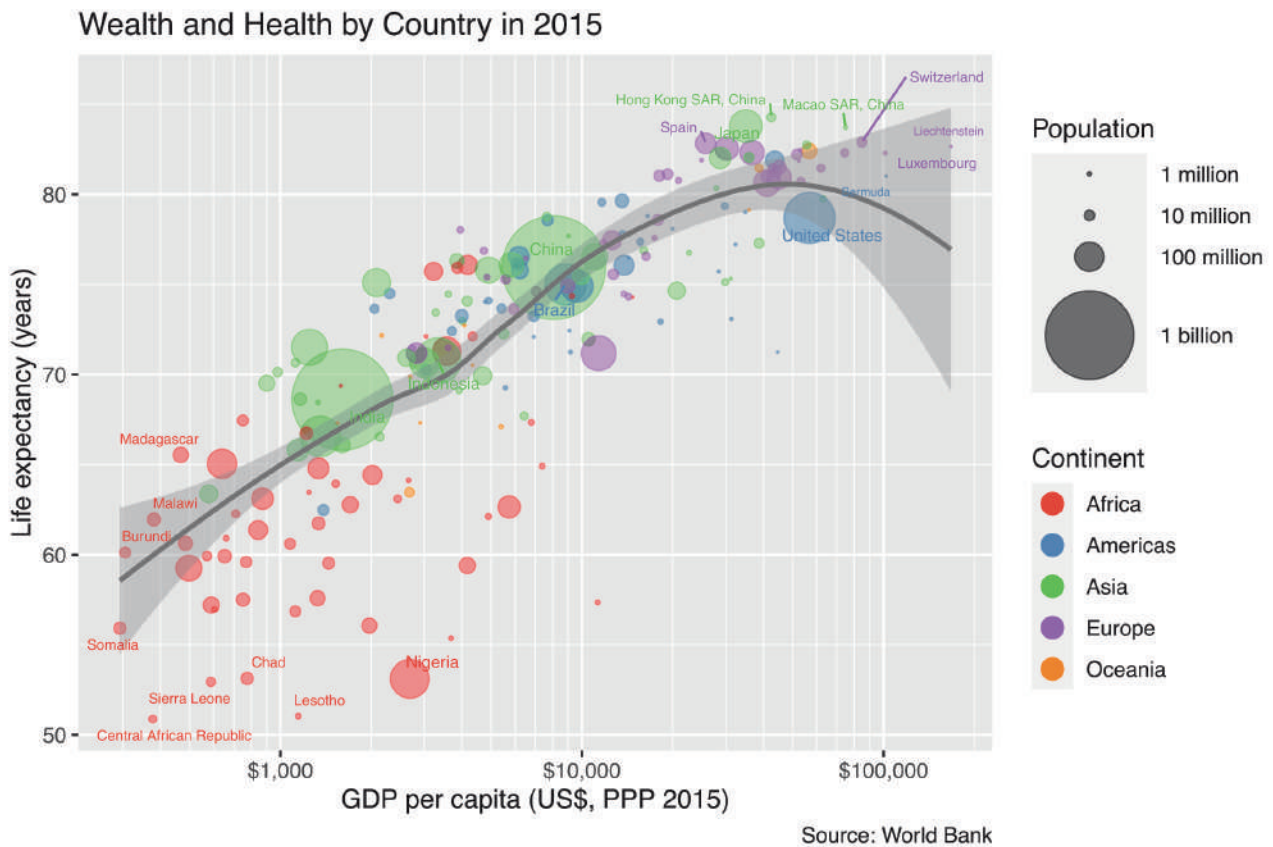
Slika 2. Primjer vizualizacije koju će izraditi studenti u modulu IV: grafikon s mjehurićima očekivanog životnog vijeka u od-
nosu na bruto domaći proizvod po zemlji.
Fig. 2 Sample visualisation to be created by students in Module IV: a bubble chart of life expectancy versus gross domestic
product by country.

class. As common in liberal arts colleges, the average number of students per section was only around 20, which allowed me to visit each team and answer questions while they were working on the problems.

As the course matured, each application exercise was structured around a coherent narrative but divided into 3 to 4 separate handouts. The first handout introduced the topic and included the first few tasks; the second handout showed the same text as the first one but added more tasks, and so on. Before distributing the next handout to all teams, I briefly paused to give them the opportunity to ask me or other teams for help. The complete problem set was released near the end of the lesson, and each team was required to submit written solutions within one week. In the final offering of the course, I gave summative feedback on solutions after odd-numbered lessons, and each team provided formative feedback to another randomly assigned team after even-numbered lessons. Teams were given rubrics for the formative feedback and received marks for completing the review.

7.3 Final Team Project

For the final project, each team chose an R visualisation package not included in the core syllabus. The aim was to explain the features of the package in a short self-written tutorial and accompanying application exercise, given to another randomly assigned team. For clarity, the authoring team will be referred to as 'Team A' and the recipient team as 'Team B'. All teams had to play both roles in parallel. Team A had to send the tutorial to Team B several days before the last class to allow sufficient time for reading. The exercise was then received by Team B at the start of the class, with 45 minutes allotted for completion. Before the class ended, Team A had to provide a sample solution and feedback on Team B's solution attempt. Afterwards, all students in Team B had to submit individual responses to a questionnaire evaluating the tutorial, exercise, sample solution and feedback provided by Team A. Scores and comments were communicated to Team A in an anonymised and aggregated form.

primatelj Tim B. Svi su timovi morali igrati obje uloge paralelno. Tim A je morao poslati tutorial timu B nekoliko dana prije zadnjeg sata kako bi bilo dovoljno vremena za čitanje. Vježbu je zatim primio tim B na početku sata, s 45 minuta dodijeljenih za završetak. Prije završetka sata tim A je morao dati primjer rješenja i povratnu informaciju o pokušaju rješenja tima B. Nakon toga, svi su studenti u timu B morali predati pojedinačne odgovore na upitnik ocjenjujući tutorial, vježbu, primjer rješenja i povratne informacije koje je dao tim A. Rezultati i komentari priopćeni su timu A u anonimnom i zbirnom obliku.

Predviđeni ishodi učenja završnog projekta bili su sljedeći:

▪ Razumijevanje paketa R-a koji nije obuhvaćen drugdje u tečaju
▪ Kontekstualizacija i prezentacija paketa
▪ Usporedba i procjena različitih grafičkih i softverskih alata za vizualizaciju
▪ Dokumentacija i objašnjenje R koda s pomoću R Markdowna
▪ Pružanje stručnih smjernica za rješavanje problema.

Završni su projekti ocjenjivani prema kriterijima u tablici 4, uz razmatranje povratnih informacija studenata iz upitnika. Ekstremnim je odgovorima dana manja težina tijekom moderiranja kako bi se osigurala uravnotežena procjena.

Završni je rad trebao biti kulminacija semestra gdje bi studenti mogli pokazati ovladavanje sadržajem kolegija preuzimajući ulogu instruktora na jednom satu. Nastavnik promatrač primijetio je o završnom projektu u sumarnoj procjeni kolegija: "Studenti su bili aktivni i angažirani tijekom cijele sesije. Štoviše, činilo se da je timski aspekt nastave potpuno integriran u širi odgojno-obrazovni rad, umjesto da se o njemu razmišlja naknadno" (Wertz 2022).

8. Primjeri vizualizacija na vježbama

Slika 2 prikazuje vizualizaciju koju su studenti trebali izraditi kao rješenje timske vježbe u modulu IV. Cilj je bio "uvježbati svoje vještine ggplot2a stvaranjem grafikona s mjehurićima koji se može usporediti s grafikonom očekivanog životnog vijeka u odnosu na bruto domaći proizvod (BDP) Gapmindera (2016)." Podatci su preuzeti s web stranice Svjetske banke (2022).

U modulu IV studenti su dobili prethodno obrađene podatke jer još nisu bili upoznati s paketom dplyr. U zadatku s modulom V dobili su hiperveze na skupove podataka Svjetske banke i morali su dovesti podatke u format koji im je dan u modulu IV koristeći dplyr.

U modulu VII cilj je bio vizualizirati iste podatke u dijagramu na dvama crtežima, od kojih svaki sadrži koropletnu kartu kao što je prikazano na slici 3. Taj je zadatak bio složen od deset podzadataka koji su vodili studente prema rezultatima vježbi u modulima IV i V do konačnog crteža. Među zadatcima je bio odabir prikladnih paleta boja iz Colorbrewer2 (Harrower i Brewer 2003) i primjena ekvivalentne projekcije (u ovom primjeru: Eckert IV).

Za posljednja su dva zadatka studenti trebali odgovoriti na sljedeća pitanja:
(9) Što čitatelj može zaključiti iz karata?
(10) Koje su prednosti i nedostaci ove vizualizacije u usporedbi sa slikom 2?

Kao odgovor na zadatak (9) od timova se očekivalo da komentiraju geografske trendove (npr. većina zemalja koje su siromašne i imaju kratak prosječni životni vijek nalazi se u središnjoj Africi). Većina je timova također istaknula pozitivnu korelaciju između BDP-a po glavi stanovnika i prosječnog životnog vijeka.

U odgovoru na zadatak (10) primijetili su da je otkrivanje korelacije iz dijagrama s mjehurićima na slici 2 lakše nego iz karte na dvama crtežima na slici 3, a koja zahtijeva od gledatelja da prebacuju pogled s jednog crteža na drugi. Timovi su također komentirali da karte na slici 3 imaju prednost jer pokazuju točno gdje se zemlje nalaze, dok boje na slici 2 označavaju samo kontinente i, prema tome, samo grube lokacije. Kao nedostatak slike 3 timovi su istaknuli da se točne vrijednosti za BDP po glavi stanovnika i očekivani životni vijek ne mogu dohvatiti iz diskretizirane palete boja, za razliku od kontinuiranih prostornih koordinata na slici 2. Općenito, odgovori timova pokazali su da su studenti postali promišljeni potrošači i proizvođači vizualizacije podataka u ovoj fazi kolegija.

9. Zaključci

Liberalno obrazovanje vještina ima za cilj prenijeti široko znanje i razvijati kritičko razmišljanje kroz aktivno učenje. Studenti bi trebali moći donijeti zaključak o tome je li vizualizacija prikladna metoda priopćavanja određenog skupa podataka. Ako da, trebali bi razmotriti koja će vizualizacija biti najučinkovitija. Također bi trebali poznavati „najbolje prakse" i razloge za njih. Ipak, studenti bi se trebali osjećati ugodno odstupajući od ustaljenih praksi kad ih se može zamijeniti boljom alternativom. Mnogi široko rasprostranjeni softverski alati za vizualizaciju podataka guše kreativnost nudeći samo ograničen
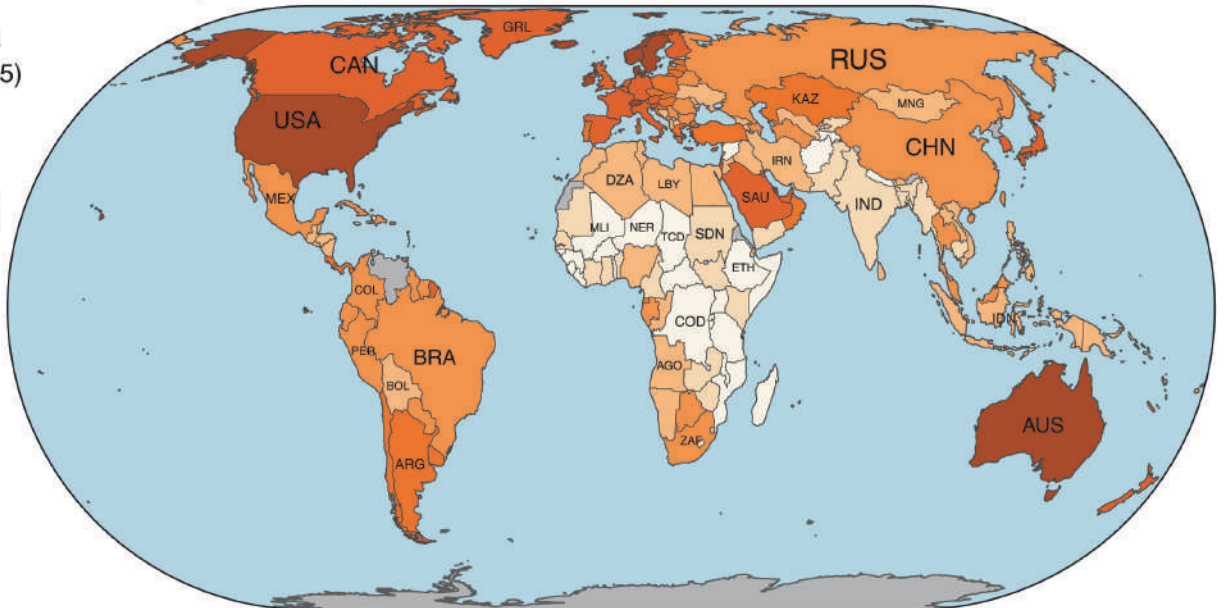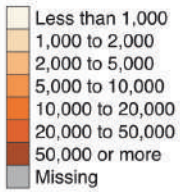
Table 4 Rubrics for final team project.
Tablica 4. Rubrike za završni timski projekt.

Rubric / Rubrika	Weight / Težina	Criteria for full score / Kriterij za puni uspjeh
Writing	20%	The tutorial is complete, well-structured and free of spelling and grammatical errors. It is properly formatted with R Markdown. Language and graphics are polished.
Pisanje	20%	Tutorial je potpun, dobro strukturiran i bez pravopisnih i gramatičkih pogrešaka. Ispravno je formatiran s R Markdownom. Jezik i grafika su dotjerani.
Contextualisation	10%	The tutorial explains why the topic is important and interesting. Sufficient information is provided to ensure that the content is comprehensible, even for readers without prior familiarity with the topic.
Kontekstualizacija	10%	Tutorial objašnjava zašto je tema važna i zanimljiva. Navedeno je dovoljno informacija kako bi se osiguralo da je sadržaj razumljiv, čak i za čitatelje koji prethodno nisu upoznati s temom.
Quality of code in tutorial	10%	The code is free of mistakes (i.e. contains no errors or warnings except for demonstration purposes) and adheres to good coding design and style.
Kvaliteta koda u tutorialu	10%	Kod je bez pogrešaka (tj. ne sadrži pogreške ili upozorenja, osim u svrhu demonstracije) i pridržava se dobrog dizajna i stila kodiranja.
Evaluation of R package	5%	The tutorial summarises the results. It either compares different approaches and recommends a 'best practice', or it evaluates the contribution of a package to the R community.
Procjena paketa R-a	5%	Tutorial sažima rezultate. Ili uspoređuje različite pristupe i preporučuje "najbolju praksu" ili ocjenjuje doprinos paketa zajednici.
Conciseness	5%	The tutorial does not exceed a length of 20 pages.
Konciznost	5%	Tutorial ne premašuje 20 stranica.
Engaging application exercise	10%	The authors provide a compelling narrative for engaging in the exercise.
Zanimljivost vježbe primjene	10%	Autori daju uvjerljiv narativ za sudjelovanje u vježbi.
Clarity of tasks	10%	The exercise describes the tasks precisely.
Jasnoća zadatka	10%	Vježba precizno opisuje zadatak.
Consistency of application exercise and tutorial	10%	The exercise aligns with the previously distributed written material without assuming additional knowledge.
Dosljednost vježbe primjene i tutoriala	10%	Vježba je usklađena s prethodno podijeljenim pisanim materijalom bez pretpostavke dodatnog znanja.
Duration	5%	The exercise can be completed in 45 minutes.
Trajanje	5%	Vježba se može završtiti za 45 minuta.
Quality of sample solution	5%	The code is error-free, concise and follows best coding practices. Adequate explanations are provided.
Kvaliteta rješenja	5%	Kod je bez pogrešaka, sažet je i slijedi najbolje prakse kodiranja. Daju se odgovarajuća objašnjenja.
Guidance	5%	The authors of the exercise provided constructive feedback in class to those attempting to solve the tasks, offering suggestions for improvements where appropriate.
Vođenje	5%	Autori vježbe dali su konstruktivne povratne informacije onima koji su pokušavali riješiti zadatke na satu nudeći prijedloge za poboljšanja gdje je bilo potrebno.
Responses to individual survey on tutorial and exercise	5%	The score is reduced in proportion to the number of teammates who fail to fill out the questionnaire on time.
Odgovori na pojedinačnu anketu o nastavi i vježbama	5%	Rezultat se smanjuje proporcionalno broju suigrača koji ne uspiju na vrijeme ispuniti upitnik.
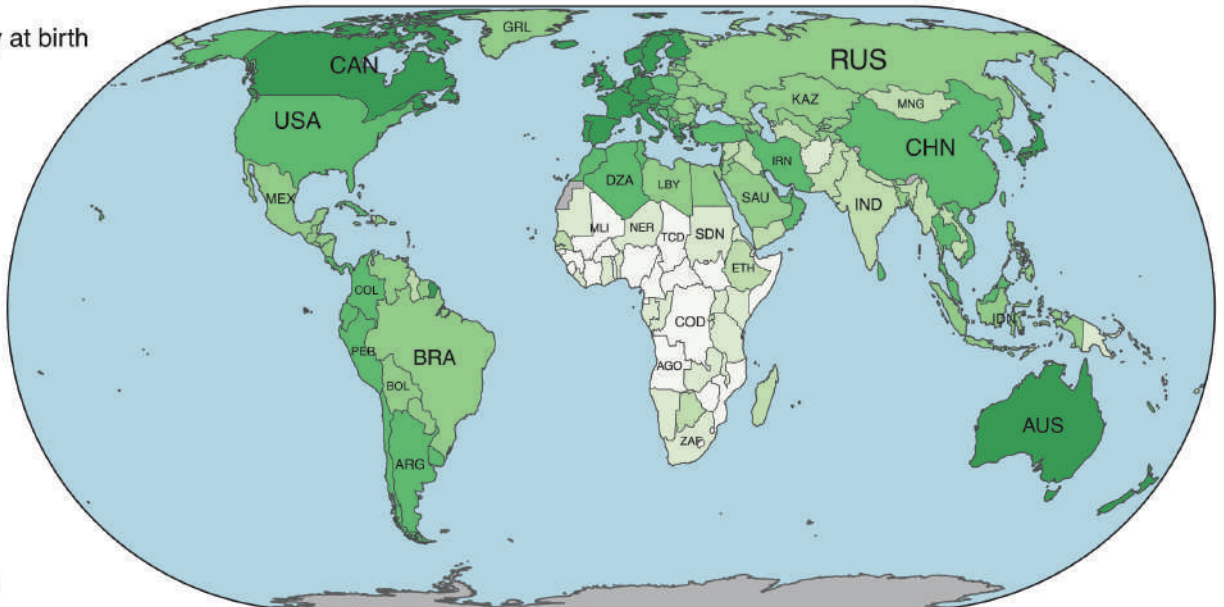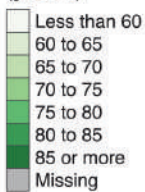
Slika 3. Primjer vizualizacije koju će izraditi studenti u modulu VII: dijagram na dvama crtežima od kojih svaki sadrži koropletnu kartu.
Fig. 3 Sample visualisation to be created by students in Module VII: a two-panel plot, each containing a choropleth map.

skup mogućih grafikona, šaljući pogrešnu poruku da je vizualizacija podataka jednaka odabiru iz malog kataloga vrsta grafikona. Nasuprot tome, Wilkinsonova (2005) gramatika grafike i njezina primjena s pomoću paketa ggplot2 omogućuje studentima stvaranje složenih i kreativnih vizualizacija. Stoga ovaj članak zagovara da bi značajan dio vremena u nastavi trebao biti posvećen praktičnom učenju paketa ggplot2 ili sličnih paketa u drugim programskim okruženjima. R ima prednost pružanja dodatnih softverskih paketa (npr. tmap i sf) koji proširuju gramatiku grafike i

sintaksu tidyverse na geoprostorne podatke, ali Python i Julia brzo uklanjaju prazninu.

U ovom su članku opisani zadatci razvijeni za relativno malu grupu od oko 20 studenata. Ako su grupe veće, instruktori se moraju više oslanjati na studente za pregledavanje koda kako bi nadoknadili smanjenu osobnu interakciju sa studentima. U tom bi slučaju također trebali biti angažirani pomoćnici u nastavi za interakciju sa studentima, kao što predlažu Ruder i Stanford (2018). Unatoč tome, temeljni nastavni plan i program u tablici 2 primjenjiv je i na

The intended learning outcomes of the final project were as follows:

- Understanding of an R package not covered elsewhere in the course
- Contextualisation and presentation of the package
- Comparison and evaluation of different graphics and visualisation software tools
- Documentation and explanation of R code using R Markdown
- Provision of professional guidance for problem-solving

Final projects were assessed using the criteria in Table 4, with the students' feedback from the questionnaire considered. Extreme responses were given less weight during moderation to ensure a balanced evaluation.

The final project was intended to be the culmination of the semester, where students could display mastery of the course content by taking on the role of the instructor for one lesson. A faculty observer noted about the final project in a summative assessment of the course: 'The students were active and engaged throughout the session. Moreover, the team aspect of the class felt fully integrated into the larger pedagogy, rather than tacked on as an afterthought' (Wertz 2022).

8 Samples of Visualisations Used in Exercises

Figure 2 shows a visualisation that students had to produce as a solution to a team exercise in Module IV. The objective was to 'Practise your ggplot2 skills by creating a bubble plot comparable to the plot of life expectancy versus gross domestic product (GDP) by Gapminder (2016).' The data was sourced from the World Bank's (2022) website.

In Module IV, students received preprocessed data because they had not yet been introduced to the dplyr package. In an assignment during Module V, they were given hyperlinks to the World Bank's data sets and had to bring the data into the format given to them in Module IV using dplyr.

In Module VII, the objective was to visualise the same data in a two-panel plot, each containing a choropleth map, as shown in Figure 3. That assignment was scaffolded as ten tasks that guided students from the results of the exercises in Modules IV and V to the final plot. Among the tasks was the selection of suitable colour palettes from Colorbrewer2 (Harrower and Brewer 2003) and applying an equal-area map projection (in this example: Eckert IV).

For the last two tasks, students had to answer the following questions:

(9) What can a reader conclude from the maps?
(10) What are the advantages and disadvantages of this visualisation in comparison to Figure 2?

In response to Task (9), teams were expected to comment on geographic trends (e.g. most countries that are poor and have a short average life expectancy are in central Africa). Most teams also pointed out the positive correlation between per-capita GDP and average life expectancy. In response to Task (10), they noted that detecting the correlation from the bubble plot in Figure 2 is easier than from the two-panel map in Figure 3, which requires viewers to shift their gaze from one panel to the other. Teams also commented that the maps in Figure 3 have the advantage of showing precisely where countries are located, whereas the colours in Figure 2 only indicate continents and, thus, coarse-grained locations. As a disadvantage of Figure 3, teams pointed out that exact values for per-capita GDP and life expectancy cannot be retrieved from the discretised colour palette, in contrast to the continuous spatial coordinates in Figure 2. In general, team responses indicated that students had become thoughtful consumers and producers of data visualisation at this stage of the course.

9 Conclusions

Liberal arts education aims at imparting broad knowledge and critical thinking through active learning. Students should be able to reflect on whether visualisation is an appropriate method of communicating a given data set. If yes, they should consider which visualisation will be most effective. They should also know the established 'best practices' and the reasons for them. Still, students should feel comfortable deviating from established practices when they can be replaced by a better alternative. Many widespread data visualisation software tools stifle creativity by offering only a limited set of possible charts, sending the wrong message that data visualisation is tantamount to choosing from a small catalogue of chart types. By contrast, Wilkinson's (2005) grammar of graphics and its implementation by the R package ggplot2 enable students to produce complex and creative visualisations. Therefore, this article advocates that a substantial part of class time should be devoted to learning ggplot2 or similar packages in other programming environments hands-on. R has the advantage of providing additional software packages (e.g. tmap and sf) that extend the grammar of graphics and tidyverse syntax to geospatial

veće grupe studenata. Nadam se, stoga, da će prijedlozi u ovom članku pridonijeti novom aspektu obrazovanja o vizualizaciji podataka čak i izvan okruženja koledža slobodnih vještina.

Zahvale

References / Literatura

Anderson LW, Krathwohl DR (eds) (2001) A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman, New York, NY

Bikakis N (2019) Big Data Visualization Tools. In: Sakr S, Zomaya AY (eds) Encyclopedia of Big Data Technologies, Springer, Cham, p 336–340. doi: 10.1007/978-3-319-77525-8_109

Blaich C, Bost A, Chan E, Lynch R (2004) Defining Liberal Arts Education. Wabash College. https://web.archive.org/web/20121021110926/http://www.liberalarts.wabash.edu/storage/Defining_Liberal_Education.pdf (accessed on 14 Feb 2023)

Börner K, Maltese A, Balliet RN, Heimlich J (2016) Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. Information Visualization, vol. 15, no. 3, 198–213. doi: 10.1177/1473871615594652

Boy J, Rensink RA, Bertini E, Fekete J-D (2014) A principled way of assessing visualization literacy. IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, 1963–1972. doi: 10.1109/TVCG.2014.2346984

Brewer CA (1994) Color Use Guidelines for Mapping and Visualization. In: MacEachren AM, Fraser Taylor DR (eds) Visualization in Modern Cartography, Academic Press, Cambridge, p 123–147. doi: 10.1016/B978-0-08-042415-6.50014-4

Cestone CM, Levine RM, Lane DR (2008) Peer assessment and evaluation in team-based learning. New Directions for Teaching and Learning, vol. 116, p 69–78. doi: 10.1002/tl.334

Chiera BA (2017) Teaching Visualisation in the Age of Big Data: Adopting Old Approaches to Address New Challenges. In: Prodromou T (ed) Data Visualization and Statistical Literacy for Open and Big Data, IGI Global, Hershey, PA, p 103–125

Clayton R (2016) Wanted: Data Scientists with Liberal Arts Training. The Economic Times, 27 Aug 2016. https://economictimes.indiatimes.com/small-biz/hr-leadership/people/wanted-data-scientists-with-liberal-arts-training/articleshow/53884306.cms (accessed on 14 Feb 2023)

Department of Statistics Singapore (2022) DOS | SingStat Website - Population and Population Structure - Latest Data. https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data (accessed on 14 Feb 2023)

Engebretsen M (2020) From decoding a graph to processing a multimodal message: interacting with data visualisation in the news media. Nordicom Review, vol. 41, no. 1, 33–50. doi: 10.2478/nor-2020-0004

Firat EE, Joshi A, Laramee RS (2022) VisLitE: visualization literacy and evaluation. IEEE Computer Graphics and Applications vol. 42, no. 3, 99–107. doi: 10.1109/MCG.2022.3161767

Gapminder (2016) Updated Gapminder World Poster 2015! https://www.gapminder.org/downloads/updated-gapminder-world-poster-2015/ (accessed on 14 Feb 2023)

Harrower MA, Brewer CA (2003) ColorBrewer.org: an online tool for selecting color schemes for maps. The Cartographic Journal, vol. 40, no. 1, 27-37. doi: 10.1179/000870403235002042

Hillery A (2019) Data literacy is the key to the future of data visualization. Nightingale, October 2019. https://medium.com/nightingale/why-data-literacy-is-the-key-to-the-future-of-data-visualization-57639532377f (accessed on 14 Feb 2023)

Jones DC, Arthur B, Nagy T et al. (2021) GiovineItalia/Gadfly.jl, version v1.3.4, Zenodo. doi: 10.5281/zenodo.5559613

Kibirige H, Lamp G, Katins J et al. (2022) has2k1/plotnine, version v0.9.0, Zenodo. doi: 10.5281/zenodo.7124918

Kumar SP (2020) Understanding the significance of digital literacy for students of Symbiosis School for Liberal Arts. Confluence: Journal of Interdisciplinary Studies, vol. 4, commentary 1. https://cjids.in/volume-iv-2020/understanding-the-significance-of-digital-literacy-for-students-of-symbiosis-school-for-liberal-arts/ (accessed on 14 Feb 2023)

Lee S, Kim S-H, Kwon BC (2017) VLAT: Development of a visualization literacy assessment test. IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 1, 551–560. doi: 10.1109/TVCG.2016.2598920

data, but Python and Julia are quickly closing the gap.

The assignments described here were developed for a relatively small class size of around 20 students. If classes are larger, instructors must rely more strongly on code review by peers to compensate for reduced personal interaction with students. In that case, teaching assistants should also be employed to interact with students in class, as suggested by Ruder and Stanford (2018). Nevertheless, the core syllabus in Table 2 is also applicable to larger classrooms. Therefore, I hope that the suggestions in this article will contribute a new facet to the education in data visualisation even outside the setting of a liberal arts college.

Acknowledgements

Lovelace R, Nowosad J, Muenchow J (2019) Geocomputation with R. CRC Press, Boca Raton, FL

Metze T (2020) Visualization in environmental policy and planning: a systematic review and research agenda. Journal of Environmental Policy & Planning, vol. 22, no. 5, 745–760. doi: 10.1080/1523908X.2020.1798751

Michaelsen LK, Sweet M (2008) The essential elements of team-based learning. New Directions for Teaching and Learning, vol. 2008, no. 116, 7–27. doi: 10.1002/tl.330

Mok HN (2014) Teaching tip: the flipped classroom. Journal of Information Systems Education, vol. 25, no. 1, 7–12. https://jise.org/Volume25/n1/JISEv25n1p7.html (accessed on 14 Feb 2023)

Myers MY (2021) Mapping Cicero's Letters: Digital Visualisations in the Liberal Arts Classroom. In Feuser S, Merten S, Wesselmann K et al. (eds) Teaching Classics in the Digital Age, Universitätsverlag Kiel, Kiel, p 93–103. doi: 10.38072/2703-0784/p24

National University of Singapore (2021) NUS launches two new colleges to deliver flexible, interdisciplinary education more accessibly, and at greater scale. https://news.nus.edu.sg/nus-launches-two-new-colleges-to-deliver-flexible-interdisciplinary-education-more-accessibly-and-at-greater-scale/ (accessed on 23 Feb 2023)

Owen GS, Domik G, Ebert DS, Kohlhammer J et al. (2013) How visualization courses have changed over the past 10 years. IEEE Computer Graphics and Applications, vol. 33, no. 4, 14–19. doi: 10.1109/MCG.2013.57

Park S, Bekemeier B, Flaxman A, Schultz M (2022) Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. Informatics for Health and Social Care, vol. 47, no. 2, 175–193. doi: 10.1080/17538157.2021.1982949

Pebesma E (2018) Simple features for R: standardized support for spatial vector data. The R Journal, vol. 10, no. 1, 439–446. doi: 10.32614/RJ-2018-009

Raineri P, Molinari F (2021) Innovation in Data Visualisation for Public Policy Making. In Grazia C, Pucci P, Raes L, Mareels G (eds) The Data Shake: Opportunities and Obstacles for Urban Policy Making, Springer, Cham, p 47–59. doi: 10.1007/978-3-030-63693-7_4

Ruder SM, Stanford C (2018) Strategies for training undergraduate teaching assistants to facilitate large active-learning classrooms. Journal of Chemical Education, vol. 95, no. 12, 2126–2133. doi: 10.1021/acs.jchemed.8b00167

Sharma AK (2017) A model scientific computing course for freshman students at liberal arts colleges. The Journal of Computational Science Education, vol. 8, no. 2, 2–9. doi: 10.22369/issn.2153-4136/8/2/1

Solis J (2019) Data visualization is king. The Journal of Private Equity, vol. 22, no. 3, 102–107. doi: 10.3905/jpe.2019.1.079

Tennekes M (2018). tmap: thematic maps in R. Journal of Statistical Software, vol. 84, no. 6, 1–39. doi: 10.18637/jss.v084.i06.

The Economist (2021) "Yale-NUS, Singapore's First Liberal-Arts College, Closes Its Doors," 4 Sep 2021. https://www.economist.com/asia/2021/09/04/yale-nus-singapores-first-liberal-arts-college-closes-its-doors (accessed on 14 Feb 2023)

Ware C (2021) Information Visualization, 4th ed. Morgan Kaufmann, Burlington, MA. doi: 10.1016/B978-0-12-812875-6.05001-5

Wertz T (2022) Senior Lecturer, personal communication, 13 Apr 2022

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer, New York.

Wickham H, Grolemund G (2017) R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 1st ed. O'Reilly Media, Farnham, UK

Wilkinson L (2005) The Grammar of Graphics, 2nd ed. Springer, New York, NY. doi: 10.1007/0-387-28695-0

World Bank (2022) World Bank Open Data. https://data.worldbank.org (accessed on 14 Feb 2023)