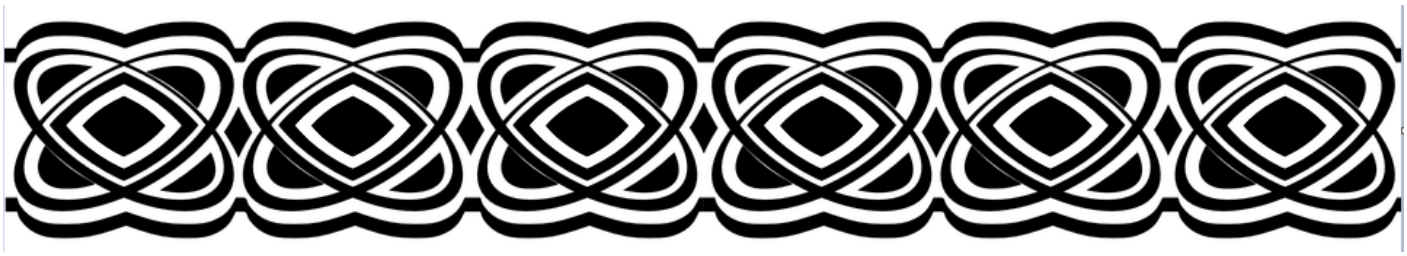
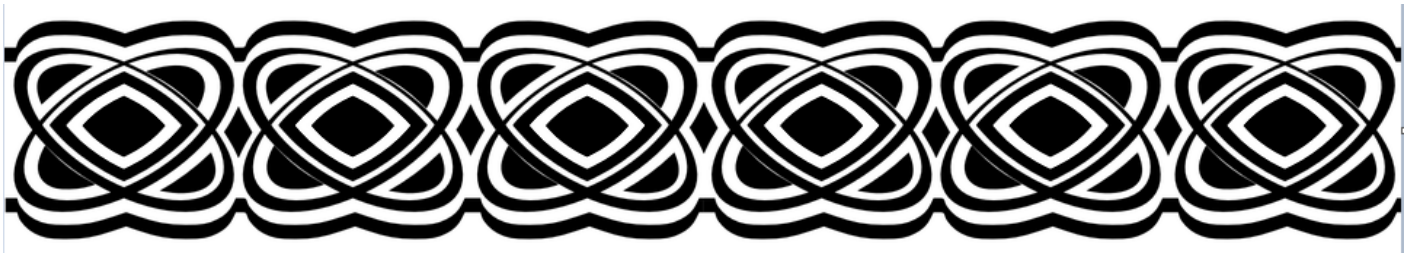
Table of Contents

1 My data and computer science courses Spring 23



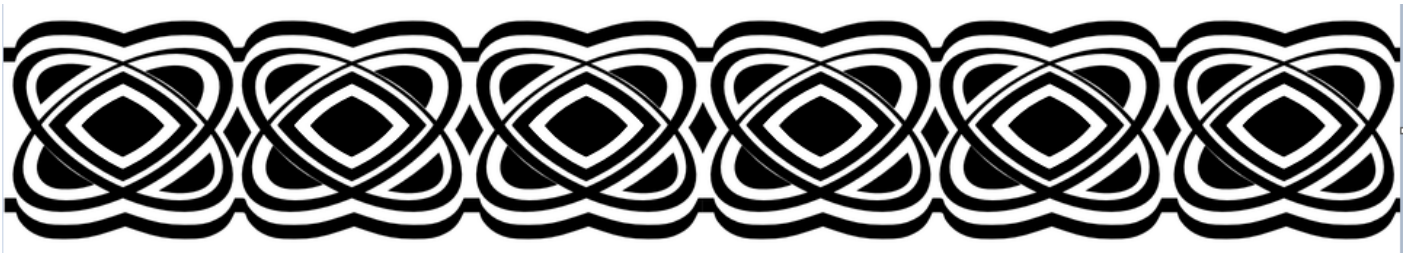
2 Introduction to Programming in C/C++ (CSC 100)

This course introduces you to programming using C. We cover C++ as an extension. C is a system programming language of pure power: it enables you to converse with the computer at a level unknown to users other high level languages many of which come from C. You also learn about: compilers, working on the command line, text editors Emacs, using C for Internet of Things (IoT) devices, cybersecurity, using pseudocode and process modeling. You get a foundation in critical thinking in concert with one of the three currently most popular languages (the other two, according to the TIOBE index, are Python and Java). The course is for anyone who is interested in learning computational thinking and basic programming skills. You get: short lectures and practice sessions with interactive notebooks, weekly quizzes and program assignments. You'll learn some great tools: we use the "hacker's editor" Emacs, the world's foremost software engineering platform, GitHub, and SAP Signavio for process modeling. This course also prepares you for "Data structures with C++" (CSC 240) and "Algorithms" (CSC 265). Prerequisites: none. Offered again in: summer 2023, spring 2024.



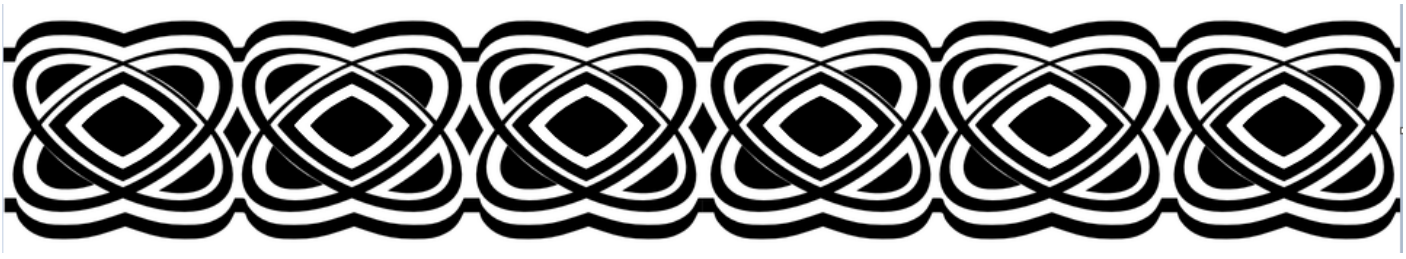
3 Digital humanities - working with text data (CSC 105)

WHY? Have you ever wondered how DALL-E creates realistic images from text input, how bots like Alexa and Siri work, or how AI writes automatic essays? The basis of these technologies, which are widely applied in the humanities, is Natural Language Processing (NLP), and its foundation is text mining with computers. **WHAT?** This course serves as an introduction to text mining using R. R is a statistical programming language used mostly by non-programmers to manipulate, analyse and visualize data. We cover: fundamentals of R, basic quantitative text data analysis, metadata and markup languages like XML, and many examples including: the novels of Jane Austen, sentiment analysis from Twitter messages, and NASA datasets. **WHO?** The course is for anyone who is interested in learning computational thinking and understanding digital technology through a data science lense. **HOW?** Short lectures and practice sessions with interactive notebooks, weekly ungraded quizzes and home assignments. You'll learn some great tools of lasting importance: the "hacker's editor" GNU Emacs, the world's top software development platform, GitHub, and command line "UNIX" tools for text manipulation. You work on your own project throughout the term using the agile project management methodology Scrum. Offered again in: spring 2025. **Prerequisites:** none. Introductory knowledge of R or of another programming language is useful but not obligatory[1].



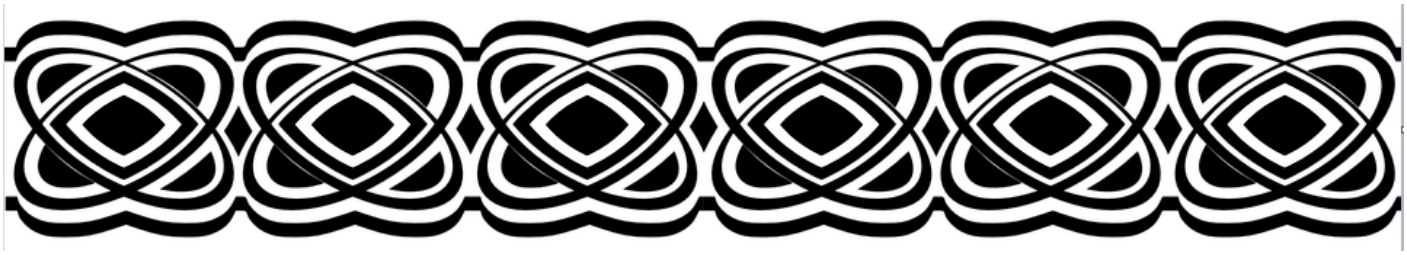
4 Introduction to advanced data science (DSC 205)

This course continues the journey into data science using the functional, object-oriented statistical programming language R. We will also explore data science using command line UNIX tools. In contrast to the basic introduction (DSC 105), we will explore some "Tidyverse" packages like `dplyr` (for data frame manipulation) and `ggplot2` (for plotting), as well as the RStudio IDE and R Markdown for data science publishing. The course is accompanied by DataCamp assignments from the "Intermediate R" course and includes a certificate. This course prepares you well for 300-level data science courses like data visualization (302) and machine learning (305) or special topics courses like "Extending R" (482). You'll also learn some great tools of lasting importance: the "hacker's editor" GNU Emacs, the world's top software development platform, GitHub. You work on your own project throughout the term using the agile project management methodology Scrum. Offered again in: spring 2024. **Prerequisites:** introductory knowledge of R as taught in DSC 105 or obtained independently by completing the DataCamp online course "Introduction to R" or "fasteR: Fast Lane to Learning R!" (chapters 1-15 only, freely available on GitHub), or Davies, The Book of R (NoStarch, 2016, Part I only). Basic R concepts are repeated and practiced at the start of the term.



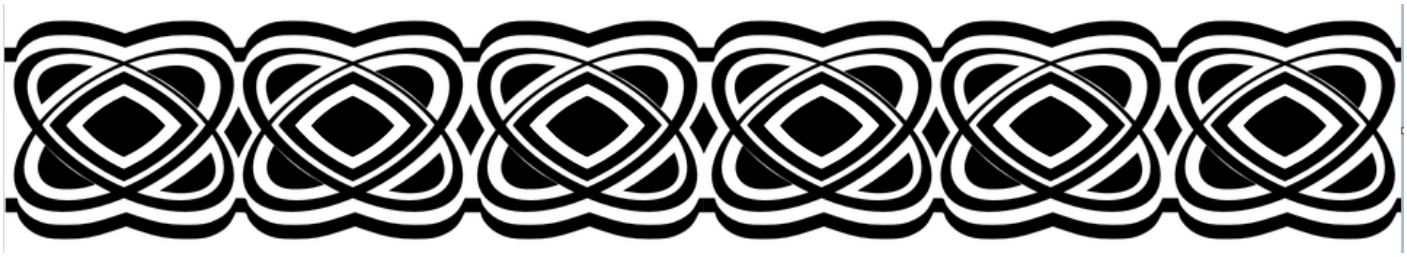
5 Machine learning (DSC 305)

Why? Machine learning (ML), a subfield of Artificial Intelligence (AI) is everywhere! It supports any job that requires repetition or looking at data, and gathering conclusions to make decisions. The course covers different methods to make predictions with ML: linear and logistic regression, decision trees, and deep learning (neural network) models. You'll learn to work with open-source libraries like `Keras` and `TensorFlow` by Google. Application examples include: predicting house prices, image recognition, and transformers, the technology behind language processing AI. Some assignments will come from DataCamp, an online learning platform for data scientists (with certificate). You'll also learn some great tools of lasting importance: the "hacker's editor" GNU Emacs, the world's top software development platform, GitHub. You work on your own project throughout the term using the agile project management methodology Scrum. Offered again in: spring 2025. **Prerequisites:** none. ML requires imagination, creativity, and a visual mind - if you enjoy finding patterns and spotting correlations, then you can do ML. Some proficiency in R or Python or another programming language is useful. If you completed the advanced introduction to data science (DSC 205) and/or data visualization (DSC 302), this course is definitely for you. If you're just curious but not sure if you can manage, talk to me and/or learn the basics of R on your own[1]!



6 Internship supervision (CSC 301/401)

Why? An internship can be an important "rite of passage", allowing you to test your skills in a real work environment. It can help you to find out who you are, what you want (or don't want) to do with your life and your education, it tests your assertiveness and networking skills, professional attitude,punctuality, stamina leadership issues, and much more. This course was designed to support you during your internship and help you make the most of the experience. **What?** We will meet weekly for one hour of discussion, you will write a weekly short (150 words) blog on GitHub, and briefly present your internship. **Prerequisites: none.** Offered again in: fall 2023. **Who?** This course serves any student who wants to have a successful career right out of the academic gates. Some topics may also be of topical interest to you as a student of psychology, or other disciplines related to success in the workplace.



7 Textbooks

These books are likely to play a role in my courses next term:

- CSC 100: C Programming by King (2008), Modern C (Gustedt, 2021), Tiny C projec (Gookin, 2021) and others.
- DSC 205: Based mostly on Davies, Book of R (2016), Part II, DataCamp, Intermediate R (2020), and Matloff, The Art of R Programming (2011).
- DSC 305: Grokking Machine Learning (Serrano, 2020), The Art of Machine Learning (Matloff, 2023), and Deep Learning with R/2e (Chollet, 2022). Machine Learning with R/3e (Lantz, 2019), Hands-on Machine Learning with R (Boehmke/Brandon, 2023).
- CSC 105: Pearson, Exploratory Data Analysis (2016), chapter 8, and selected texts, like Robinson/Silge, Text mining with R (2017).

Footnotes:

[1] as taught in DSC 105 or obtained independently by completing the free DataCamp online course "Introduction to R", or "fasteR: Fast Lane to Learning R!" (chapter 1-15 only, freely available on GitHub), or Davies, The Book of R (NoStarch, 2016, Part I only).

Created: 2022-10-25 Tue 22:07