

DSC 365/465 Syllabus

Spring 2021

Basic Info:

Terriell (Doug) Scrimager, MSPA

Office: Online Only

Office Hours:

Wednesdays from 7:00-9:00pm – appointments only – BlueStar – email me as well

Mondays after class (after recorded class)

Open Hours Thursdays from 7:00-9:00pm or until everyone has left:

show up promptly or email me if you are going to be late. If no one is on by 7:15 I will leave.

e-mail: tscrimag@depaul.edu

will respond by night of next business day

include DSC 365/465 in the subject!

Location:

Monday/Seminar: Zoom

Link: <https://depaul.zoom.us/j/99437556034?pwd=MDJ6VFJRWTNxOGI3NkR2blkwNE43Zz09>

Password: dsc465

Thursday Open Hours: Zoom

Link: <https://depaul.zoom.us/j/92326120592?pwd=TkRZQUwrdFVySmQ2ekxZaUhKVURxQT09>

Password: dsc465

Key Dates:

Midterm: May 1st – May 9th

Final Presentations: May 31st

Final Project Reports Due: May 6th

Course Summary:

This course will be an introduction to data visualization techniques for exploration and analysis of large data sets with a wide variety of data types, such as might be discovered from a wide range of fields including commercial, financial, medical, scientific and engineering. Topics will include visual encoding of numeric data, human visual perception, effective visualization design, graphical integrity, visualizing distributions, color theory, basic network graph visualization, geospatial visualization and some additional topics.

Prerequisites: PREREQUISITE(S): IT 403 and (CSC 401 or IT 411)

This course assumes that you have had a basic course in statistics along with an introductory programming course (e.g., Python).

Learning Goals:

By the end of the quarter, students will be able to use visualization to discover insights in data and craft visualizations to communicate such insights. You will be able to choose appropriate visualizations to perform exploratory analysis and investigate data, and you will be able to create explanatory visualizations to communicate your findings. To do this you will leverage your knowledge of a collection of different visualization techniques along with understanding of how to appropriately apply them and optimize their appearance for clutter, clarity, graphical integrity and human perception. Specific goals for each module are included to help you track and reinforce your learning.

Expectations:

Student/Instructor responsibilities:

First and foremost, this is a graduate course, so you are expected to be a part of a learning community and to take primary responsibility for your overall experience in this class. This class provides a structured learning experience, guidance, resources as well as the instructor which are a significant part of your success in learning so please take advantage of all parts of the course. As an instructor I am here to facilitate learning and am dedicated to the success of each individual in the course. The online tools are there to help you manage your course-load, due dates, homework and class project activities. All work is expected to be completely reproducible with all needed code, data, citations and documentation included for any submission whether homework, online activities or final project. I should be able to follow your process and achieve/find the same results. For every quarter-hour expect to spend a similar amount outside of class each week so ***often at least 4 hours per week over and above the lecture***. Final projects usually require some extra effort, but the latter half of the course is designed to leave you more time for project work.

Course News Items

I provide most of the critical information during the quarter through news items. They can get long and involved since I make them comprehensive of many items that come up during each week (like completion of grading). Please read these thoroughly. In addition, I publish a (mostly) weekly 'checklist for success' as a news item – please read this ASAP as this will point to the priorities of the upcoming week. I strive to publish this Fridays for the next week.

Contacting Me

Please get in touch if you have questions or would like to schedule a meeting outside office hours. **Email is the best option.** NOTE: I respond quickly when possible, but my policy is that you should receive a response by the night of the next business day (i.e., an email Tuesday gets a response by Wednesday night). I have extremely limited availability on the weekends as a last resort and Sundays are preferred for this.

Additionally, please do not email me questions to which the answer is right here on the syllabus (e.g. when is the midterm?) or posted in a note/news item on the D2L page. I may not answer such emails, so if you do not receive a response, double-check that the answer is not easy to find. When you email me, **include DSC 465 online** in the subject.

I will help debug code after you have made a significant effort (including google and StackOverflow searching) to get it working on your own. Do not send me a snippet of code and ask 'what is wrong?'. If you want help debugging, you must include a complete example of what you are trying to accomplish. See Expectations/Coding for what you should do.

Finally - come to my weekly individual office hours or open office hours (office hours are not recorded). I encourage you to ask for help if you need it – a large part of my role in this class is to be sure you have all the opportunity and resources to succeed.

Workload

Visualization is a wide-ranging topic and we have a lot to cover, so the course ***is a significant amount of work***. Except for full-time graduate students, most students are not able to take this class concurrently with another graduate level course. Part of the work is learning to create visualizations, and like any skill this takes practice. You are expected to put in the time making mistakes and figuring things out, and in return you can expect not to be wasting your time with unhelpful exercises. Much of that practice will be on the homework assignments and the project. Once we get to the second half of the class, there is less new material and less individual homework because we shift focus to group project work. The dividing line between halves of the class is the midterm in week 6, which will test your core visualization knowledge. Expect it to be fair but not easy – see midterm details below.

Finally, there are additional weekly exercises that will aid your learning and build community, including **quizzes** to reinforce key points, **tutorials** to give you practice with R implementation, and **discussions** to develop the class' higher level skills like critique. These are individually small parts of your grade, but I expect you to complete all of them because they will make a big difference in your outcomes for the course. The grades section below explains the components of the course in more detail.

Coding

We will generally use Tableau and R in this course, it is important to use both types of tools to build these skills for visualization (and marketability). Though you may use Tableau for most assignments, you will likely find yourselves drawn toward coding in R because you get more control. You are required to use R for some of the homework questions. If this sounds unlikely to you, at least give it a shot though the tutorials. The prerequisites require some familiarity with coding, so I will not be covering the very basics, e.g., concepts of writing instructions to the computer, variables and loops. Learning R and its data toolkits is well worth the time, but you should be able to get through by adapting examples from the ggplot book (see below) and elsewhere, including my demonstrations. This will require experimenting.

I will help debug code if you demonstrate that you made significant efforts to make progress. You cannot learn coding without practicing and making mistakes (banging your head against the wall, frankly). Work with each other, use the recommended books and provided tutorials. Start by making an established example run and then start adding and changing. Come to office hours (or screenshare) and always start by telling me what you have already tried (this is always a good idea when asking anyone a technical question, including a boss or IT support).

Statement on tools and data preparation

This class is about data visualization. It also has a basic programming prerequisite so any participant should be enabled to do their own data preparation. However, many are not proficient in 'R' prior to this class. R is NOT required for data prep but I highly recommend you learn the basics and perform as much as possible in R as you can. However, data preparation is not intended as a roadblock so it should not be something that raises your level of frustration. Use whatever tools you want to perform your data preparation -even Excel if you must, but you MUST document the steps in sufficient detail in your homework and include any artifacts (like an excel workbook) that you used during any homework.

Tableau is used since it is a leading visualization tool. However, it can take significant effort to learn as well. Tableau has published many hours of great tutorials on their site as resources. Use these to supplement your learning. It's meant to be easy to use - yet I personally find it's unintuitive and idiosyncratic in many ways - be prepared to view some videos (our tutorials and theirs) and play with the tool. Sometimes starting a visualization over from scratch is best as you learn. I have discontinued use of Tableau professionally so I will help where I can, but you will also be expected to use google and other resources as much as possible. I have some great examples I will share.

Additionally, this quarter I am introducing PowerBI from Microsoft. This has been requested by many students since it is starting to rival Tableau in popularity. You are NOT required to use PowerBI but if you want to use it, I will be able to grade these submissions. I will offer basic help and a familiarization demo but for more complex questions we will often consult google together. Be ready to try and find the answers online or through community chat first.

If you want to use an alternate language/tool such as Python/PowerBI please let me know ASAP. I can grade those submissions since we are looking at the outcome. You would submit the homework but note in your submission that you used a different tool for the question. You cannot use a non-programming tool to answer a question which specifies using R – you CAN use another programming language such as python. The same is true for questions that specify using Tableau. You can use a non-programming tool like PowerBI to answer Tableau questions.

Many successful students complete each homework question in both a programming tool (R/Python) and a visual tool (like Tableau/PowerBI) in order to best understand what works given the situation. This also helps you understand which might be better for you personally. In a professional setting there is often the expectation that you can do both. Students who make on-time complete submissions may get extra credit if they have applied multiple tools to a question. So, if you do a question correctly and well in BOTH R and Tableau you will usually receive extra points.

Statement on class as an online experience.

Online Materials are designed by Dr. McDonald, Dr. Brown with a small splash of material from me, and then produced by Dr. Brown. So, who you see in the recordings is not me. You are essentially getting the benefit of three professor's knowledge along with updated, custom material that I produce every quarter. You are expected to view/read all the videos/baseline materials each week. To distinguish this from a purely online experience, I supplement this during seminar hours, open hours, email and individual office hours. I am dedicated to the success in learning for each person. These supplemental

contact hours ensure that students have access to help with course materials and to bring contemporary materials into your learning experience. I am here to help. I am currently full-time employed so please be aware that I may not be generally available during business days and much of my course work, prep, forum work, etc. is performed on the weekends.

The weekly seminars, held at class time and recorded, include practical and pragmatic applications of data visualization. These will not last 3 hours since some of the contact time is spread over other nights. These are completely customized each quarter and evolve as do the course materials from student feedback. **Seminar attendance is highly recommended** and is also recorded for later viewing. Since seminar is scheduled at the given class time, I hope that you will attend live as much as possible since that provides the most benefit to you and the community. If you are a distance student doing async learning please review these weekly seminar records as soon as possible so that you can use the personal or open office hours for any questions you might have. These seminars will be applicable to the final project as well as being drawn from examples that I use in a professional setting along with covering questions around class material and live demos. Since I actively do data visualization and data science in my professional role, I will expose you to techniques that will be useful to you in a professional data science/visualization role. I will be specific about what materials are in the midterm – as you will see on the midterm study guide.

Finally, the weekly open-hour group zooms are to have open discussions, ask questions and interact with peers. It is not required to attend, but I have found this enriches the experience for students who take advantage of this. **To provide a safe learning environment the group zoom is not recorded**, so our discussions can be open as long as you follow DePaul's policy 'Guiding Principles on Speech and Expression' http://policies.depaul.edu/documents/other/Guiding_Principles_20170530.pdf). We tend to stick to course topics, or those closely related, and the discussion usually goes as long as needed.

Discussions Boards and Homework Discussion

The discussion boards are for you. I generally browse them to look for issues from time to time. Specifically, what I must check for and cannot tolerate on the boards is academic integrity violations. You may not post answers or answer code to the forums. It is okay to discuss homework, give suggestions for plan of attack or discuss strategies, but everyone must do and submit their own work. The forums are also used for forming project groups and weekly discussion posts. In all cases I expect you all to be courteous to each other and respectful of others' time and effort.

Tutorials, Quizzes

Like the Discussion boards, Tutorials and Quizzes are there *for you*. Quizzes help you remember key points of information and Tutorials give you examples to work from for your assignments.

Grades

See the Grades section for details on the proportional components of your final grade and policies, e.g., for late assignments.

Textbooks and Software:

Books

The following books are **necessary** (they are 'required', but they are also free to read online, and 'required' makes it sound like you must buy them, so I use a different word. It is worth noting that it's not expensive, as textbooks go, to get a hardcopy.

- Covers ggplot with examples that will be critical for doing the homework:
R Graphics Cookbook: Practical Recipes for Visualizing Data, 2nd Edition. Winston Chang. O'Reilly Media, November 2018. <https://r-graphics.org/>

The following books are **strongly** recommended for the course. You need them to get the most out of the course.

- Covers the world of R libraries for data manipulation and more on ggplot:
R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Hadley Wickham and Garrett Grolemund. O'Reilly Media, January 2017. <https://r4ds.had.co.nz/>
- Nearly a textbook for the course, provides more detail on most course topics:
Fundamentals of Data Visualization. Claus O. Wilke. O'Reilly Media, April 2019.

The following books are **recommended**, with explanations of what they cover:

- Excellent coverage of the graphing parts of the course, including clutter, clarity, graphical integrity and guidelines. This is a very influential book on visualization design and is beautifully written with excellent explanations and illustrations (some used for the course).
The Visual Display of Quantitative Information, 2nd Edition, Edward Tufte. ISBN-13 978-0961392147
- An actual visualization textbook and written by one of the top scientists and most influential people in the field. It is a little too geared for future visualization researchers to be our textbook, e.g., including how to perform design studies to make interactive visualization software, but I definitely recommend it for its overall coverage of the science and practice of visualization.
Visualization Analysis and Design. Tamara Munzner. A K Peters/CRC Press. December 2014.
- Much more detail on visual perception for a visualization audience.
Visual Thinking for Design. Colin Ware. ISBN-13: 978-0123708960
- Introduction to visualization from someone who has built a career out of doing it in industry, starting at Google and now consulting. Implementation info is available in Excel.
Storytelling with Data. Cole Nussbaumer Knaflic. ISBN-13: 978-1119002253.
- More design-oriented view of visualization, beautiful book including case studies.
The Functional Art: An Introduction to Information Graphics and Visualization. Alberto Cairo. ISBN-13: 978-0321834737
- Covering D3, a JavaScript-based library for building interactive tools that we discuss at the end of the course.
- *Interactive Data Visualization for the Web*. Scott Murray. <https://alignedleft.com/work/d3-book-2e>
- Using MATLAB for visualization
MATLAB Data Analysis and Visualization. Antonio Siciliano. ISBN-13 978-9812837516

Software

- **Tableau** kindly provides student licenses!
 - You can go to their website and sign up as a student and they will verify your student address at DePaul: <https://www.tableau.com/academic/students>
 - You have to renew every year but it is free while you are a student
- **R Studio** is a convenient and free way to use R (the language). You can download it here but note in the installation instructions that if you don't have R itself, you'll need to install that too by following their link: <https://rstudio.com/products/rstudio/download/> (Get the *Desktop* version. Server is for running your own server for interactive R visualizations.)

Grading:

Work in this course will be evaluated principally through homework assignments (4), a midterm exam and a final project. The breakdown for the grading will be:

- Quizzes: 10%
- Discussions and Participation: 10%
- Homework: 30%
- Midterm: 20%
- Final Project 30%

Quizzes

Weekly quizzes will be administered through D2L and will test the material in the lectures. These quizzes are open book/notes/friends. I ask that you consider the questions yourself in order to test yourself on the material as this 'recall practice' helps you remember the material. I will review each quiz the following week, but you can still retake to improve your score.

Discussions and Participation

The discussion assignments will help you engage with the material and the community by finding discussing examples and questions. You must make your own post and then respond to others. This is worth a portion of your grade. Additional credit can be granted for strong participation in the forums, e.g., helping fellow students.

Attendance and Participation:

You are expected to participate in online discussions with your peers - this is the critical part of your participation score. One of the most important parts of a class is building a learning community. I will be expecting you to be available to help each other with constructive suggestions and to complete the online activities as discussed previously. **Keep up with discussions - waiting until the last week to comment on a bunch of forums is not participation - it's these ongoing thoughtful interactions that are the key to the value of discussion groups.** The discussion assignments will help you engage with the material and the community by finding and discussing examples. You generally must make your own post and then respond to others in the formal discussion topics.

Please ask questions (in class or online forum or both) if something is not clear - if you have a question, chances are that someone else has the same question.

Homework

Homework assignments will be submitted online on the D2L website. **No email submissions will be accepted for homework.** Submissions by the due date will receive feedback and if complete (all questions answered, and all graphs/code included) will be given the option to resubmit for full points. Do NOT submit partial homework before the due date. Homework resubmitted only for extra credit is not allowed – complete any desired homework extra credit with your first submission. It is rare that homework will be done perfectly the first time. Please take advantage of this resubmission policy but follow this exactly. **Homework after the due date receives the points given the first time with no resubmission grading and no homework extra credit.** So, be on time and you can fix anything wrong for full points. The same type of discussion with peers that is allowed on the discussion board is allowed with respect to homework assignments, however, each student must turn in original work, i.e. your own write-up.

All homework submissions must be made up of two files:

- A PDF with all questions answered and all graphs included. Indicate which homework, your name and indicate which question and part are being answered by which written answer and/or graph. Put your responses and graphs in the same order as they are in the homework document. Do not send separate files – all work must be in ONE file easy to grade and to provide feedback. If there is an issue or question, then I may look at the zip file. You will have to cut-n-paste from various tools to do this.
- A Zip file with all other elements (your R files, Tableau workbooks, anything else you used to complete the homework – do not include the data file). I should be able to follow exactly what you did and reproduce it.

If any of the above is missing or incomplete, then your submission is incomplete. You are only given the chance to revise any submission if it is comprehensive, complete, on time and follows the guidelines. Include any homework extra credit in your initial, on-time, complete submission to be considered.

This is a safe learning environment. There is little worry about ‘wrong answers’ as long as you submit full work by the deadline, incorporate my feedback in resubmissions and complete the work. I have found over the years that this is a key to learning. This is a unique element, and opportunity for you, in how I teach this course that you will rarely find elsewhere. So, outside a classroom setting, you will get feedback and have time for refinement of your output. In addition, data visualization often needs a few iterations to ‘get it right’. The midterm is a one-time test, but you have sufficient time and materials to complete this for a great grade.

Midterm

The midterm is held on the sixth week of class. Makeups for the midterm will only be given in extreme circumstances and requests for a makeup must be made as soon as possible. Documentation must be supplied of the relevant circumstances. **You must take the midterm to pass the course.** Midterms will be separately curved for undergraduate students. There is no seminar the week of the midterm. Group hours and individual office hours will still be scheduled during the midterm week. You are expected to take the midterm free of materials and outside assistance – this is to test your learning. Please expect to devote the full 3 three hours provided to complete the midterm. So, plan accordingly. The final midterm grade is not calculated until I manually grade the written section – you will see a lower grade at first for

only the portion that is auto-graded. Please do not email about these scores until I announce that the midterm grading is complete.

This midterm is much harder than most students expect. It is comprehensive of the first 5 weeks of material. At the end of each recorded lecture and your study periods, review the midterm study guide to make sure you have internalized the materials. **The midterm materials are taken directly from the recorded lectures and associated documents.** Many questions require you to apply the principles from the course. It is a mix of T/F, matching and written (short answer) – composed of approximately 65 questions depending upon how many test questions I have included. Poorly performing questions (we have metrics to understand how well a question is predictive of an overall midterm grade) can be curved out at my discretion. Some questions are test questions so will not count against you if incorrect but help me refine the testing contents of the course. All correct answers contribute to your grade even if they are a test question or are curved out otherwise – credit where credit is due.

Final project

The final project in the course will be a group effort to build an original collection of visualizations for a dataset of your choice. Most of the work in the latter half of the course is focused on the project. The deliverables consist of check-ins, a presentation with infographic and a written final detailed report. You will present your final project infographic on the 10th week of class, and then you will have until the final exam date (one week later) to submit your final report (see dates above).

- Groups MUST have someone available live to present for Week #10 in case there are questions or comments that need to go back to the group.
- We will work to use recordings, voiceovers, Zoom, whatever - each person should plan to present something from the project in week 10- but you do not necessarily have to be live to present your contribution.

The project description document is your friend. There are ideas for data sources, an effective outline for the report, and the grading rubric. Most importantly, refer to that document for project milestones, group composition rules, and policies, e.g., for removing a group member if absolutely necessary. I will discuss the project description early on.

Extra Credit

There are experiments going on with human computer interaction (HCI) and psychology researchers at DePaul. They need participants. You can get extra credit in this class for participating. Usually, they will start to be available around the fifth week of the quarter. Go to this link to check it out:

<http://www.cdm.depaul.edu/academics/research/Pages/Instructions-for-Participants.aspx>

There are also plenty of extra credit assignments available – however ALL regular coursework MUST be completed (excluding the final project) before any extra credit is counted. All regular course work means all discussions, quizzes and homework. Some homework will include extra credit, and these will be graded only if the homework is a complete submission the first time by the due date. No resubmission of homework is allowed if only for extra credit.

Final Grading Note

I aim to make grading fair and focused on feedback. Email me if you think this has not been the case.

Feedback for Me

Online Course Evaluations

Evaluations are a way for students to provide valuable feedback regarding their instructor and the course. Detailed feedback will enable the instructor to continuously tailor teaching methods and course content to meet the learning goals of the course and the academic needs of the students. They are a requirement of the course and are key to continue to provide you with the highest quality of teaching. The evaluations are anonymous; the instructor and administration do not track who entered what responses. A program is used to check if the student completed the evaluations, but the evaluation is completely separate from the student's identity. Since 100% participation is our goal, students are sent periodic reminders over three weeks. Students do not receive reminders once they complete the evaluation. Students complete the evaluation online in [CampusConnect](#). I personally read these after each quarter, so while Administration will review them for my job performance, I will be looking carefully for useful constructive criticism that I can use to improve the course. Please help me out.

Other Feedback

Please feel free to email me feedback about the course, during or after. I work hard to make this course effective, interesting and to contribute additional material based upon my professional and academic experience. If you let me know what worked for you and what didn't, I can improve the course. Even if I cannot do that in time to help you, future students will appreciate your time. I will too. One extra credit question/discussion is to help me reword one specific question that I provide and additionally contribute what you think might be a good question for the midterm question bank – this is your chance to directly add to the course.

Policies

Changes to Syllabus

This syllabus is subject to change as necessary during the quarter. If a change occurs, it will be thoroughly addressed during class, posted under Announcements in D2L.

Academic Integrity and Plagiarism

This course will be subject to the university's academic integrity policy. More information can be found at <http://academicintegrity.depaul.edu/>. If you have any questions be sure to consult with your professor.

Academic Policies

All students are required to manage their class schedules each term in accordance with the deadlines for enrolling and withdrawing as indicated in the [University Academic Calendar](#). Information on enrollment, withdrawal, grading and incompletes can be found at: cdm.depaul.edu/enrollment.

Preferred Name & Gender Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the quarter so that I may make appropriate changes to my records. Please also note that students may choose to identify within the University community with a preferred first name that differs from their legal name and may also update their gender. The preferred first name will appear in University related systems and documents except where the use of the legal name is necessitated or required by University business or legal need. For more information and instructions on how to do so, please see the Student Preferred Name and Gender Policy at <http://policies.depaul.edu/policy/policy.aspx?pid=332>

Students with Disabilities

Students who feel they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss their specific needs. All discussions will remain confidential. To ensure that you receive the most appropriate accommodation based on your needs, contact the instructor as early as possible in the quarter (preferably within the first week of class), and make sure that you have contacted the Center for Students with Disabilities (CSD) at: csd@depaul.edu.

Lewis Center 1420, 25 East Jackson Blvd.

Phone number: (312)362-8002

Fax: (312)362-6544

TTY: (773)325.7296