

Categorical Data Analysis

SDS 390CD
Spring 2025

Dr. Kementari Whitcher

MEETS

Monday & Wednesday 10:50 – 12:05
McConnell 403

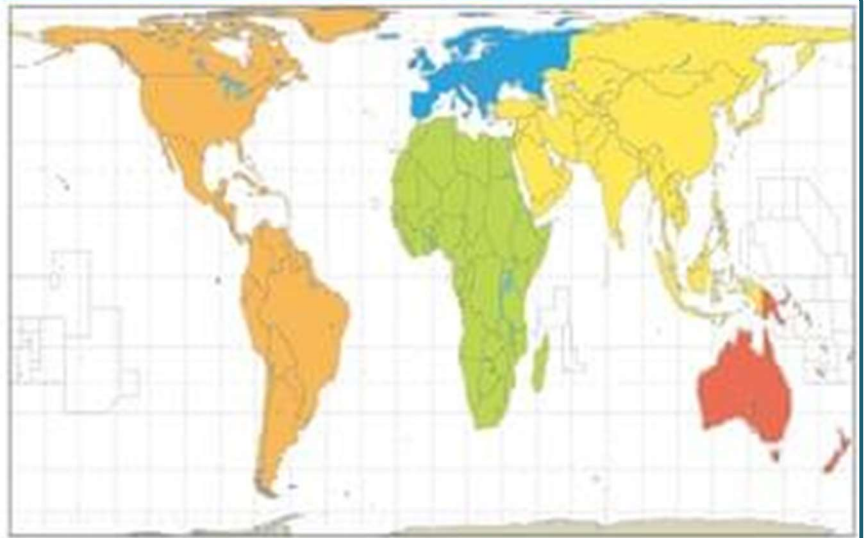


Figure Peters World Map from <https://www.theguardian.com/global/gallery/2009/apr/17/world-maps-mercator-goode-robinson-peters-hammer> "Continent" is a categorical variable.

CONTACT

Office: McConnell 207

Email: kwhitcher@smith.edu

Phone: 413-585-3945

And on Slack

Student Hours

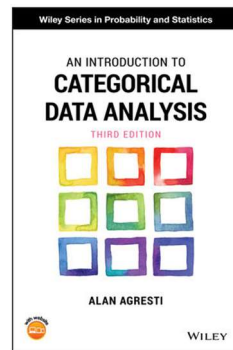
Monday 1:30 to 2:30

Friday 11:00 – noon and 2:00-3:00

Please note I am not on campus on Tuesday and Thursdays but can be available by appointment on zoom.

TEXTBOOK

An Introduction to Categorical Data Analysis by Alan Agresti. 3rd edition. From Wiley publishers. I have ordered copies through the bookstore and there is an e-version available at a reduced cost.



Homework

There is an extensive repository of data sets available with this text. These datasets are available at:

<https://users.stat.ufl.edu/~aa/cat/data/>

or on Github at:

<https://github.com/alanagresti/categorical-data>

or by downloading the zip file in Moodle
Agresti_DataSets



WELCOME TO METHODS IN CATEGORICAL DATA ANALYSIS

This course covers a range of topics to build a strong tool kit for the analysis and interpretation of categorical data, both itself and in combination with continuous data. There is a wealth of data that is either not numerical or does not conform to the level of continuous data. We will rely heavily on reading and practice – your text explains methods thoroughly, gives examples, and provides plenty of practice problems. It relies on R for the analysis of the data, and this course will

PERSPECTIVE

This is an advanced class in methods for analyzing categorical data.

This is a very exciting topic that many undergraduate students do not get the opportunity to explore in any depth. Proficiency with these techniques and concepts will give you a serious tool to add to your analytics tool kit. I am very happy to be undertaking the study of this topic with you!

Emphasis

We will emphasize understanding when to apply various techniques, how to apply them, and what the results mean. It is important that you

engage your skills and previous experiences. Try things out, see what results, and talk about your interpretation. If you get stuck, remember growth only requires that we learn from any bumps along the way. Most important is that you *understand* what we are doing, and why, and when these techniques should be applied.

be able to convey the meaning of your analysis to a broad audience. Many will be unfamiliar with this work.

YOU BELONG HERE – AND YOU HAVE THE POTENTIAL TO SUCCEED!!



OPERATING PRINCIPLES

HOW YOUR LEARNING IS MEASURED

The short answer – you tell me. The longer answer – this class will utilize an *Ungrading*¹ approach known as Labor Based Contract Grading². What you get out of a seminar depends on what you put into it. Much of our work will be in class, but there is also significant outside-class work. You will decide the level of grade you want to contract with yourself – and the class – to receive.

We are all works in progress, and I want to engage your *growth mindset* – the idea that you have great capabilities and the educational journey is about developing them.³ As we journey together, we need feedback about how we're doing – are we getting it? Or not yet? I much prefer assessment to grading, so I will give you feedback that indicates how well I see you demonstrating your understanding. We will talk about how to assess your learning, rather than about your grade. Instead, your final grade will be based on your own description and evaluation of your learning, focused on continuous development. and your ongoing development of knowledge of this topic.

We are operating within a constrained semester system, and at the end of that semester you will need a grade. So how will we get there? We will undertake various activities throughout the semester designed to form and practice your learning, review and demonstrate your learning, and formally assess your learning.

¹ See, for example, *Ungrading : Why Rating Students Undermines Learning (and What to Do Instead)*, edited by Susan D. Blum, West Virginia University Press, 2020. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/smith/detail.action?docID=6370466>.

² With a great deal of appreciation to my colleague Dr. Lindsay Poirier for sharing her Labor Based grading approach and material.

³ See Baldwin, Amy, et al. *Promoting Belonging, Growth Mindset, and Resilience to Foster Student Success*, National Resource Center for The First Year Experience & Students in Transition, 2020. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/smith/detail.action?docID=6132478>.

FORM

You have access to a world of resources to learn about any kind of data analytics. Your text is an excellent resource. We will talk about the concepts and techniques for analyzing this data in class, and I plan to review each chapter with you. Then you will have opportunities to practice your skills through in-class exercises and out-of-class homework. Much of the analysis we will do will require you to use R to analyze a variety of data. Each chapter, I will list in-class problems for practice, as well as homework problems for you to do outside of class. You will get plenty of practice with the material!

DEMONSTRATE

Your first line of assessment comes when you do in-class exercises or homework. Did you understand it? How did it go? It's also important to make sure you can communicate about your analysis with a wide audience, in a way that is understandable and verifiable. You should continually be assessing your own learning. In addition, my role is to give you feedback about your progress – are you on target? Exceeding expectations? Or not quite there yet? (Scale = Satisfactory, Exceptional, Not Yet) The regular practice of thinking about your thinking, your understanding and points of confusion, is called Metacognition. I will provide you with some tools to reflect on your experiences in the class. You will regularly submit these to me and periodically we will have conversations about how it's going.

You will also develop a final project that will allow you to demonstrate your understanding of and competence with the analytic approaches we've learned this semester in a way that works well for you. That will be a data project that includes a poster presentation and discussion of your findings. We can talk about formats - I welcome your ideas!

There will be (at least two, possibly three) instances across the semester where I will ask you to assign yourself a letter grade (either interim or final) and discuss why you assigned the grade you did. We will then have a formal conference about those grade reports. I fully anticipate assigning you the final grade that you assign yourself, unless your self-assessment is unsupportable (you cannot make a reasonable case for it) given the grade contract rubric. You should refer to the rubric frequently to make sure you are on track for your contract grade.

ASSESS

My role is to give you feedback about your progress – are you on target? Exceeding expectations? Or not quite there yet? (Satisfactory, Exceptional, Not Yet or ✓, ✓+, and NY) This scale applies to everything we do, and any time you are “not there yet (NY)” you should come get help and work through any difficulties you are having.

QUESTION – DO YOU WANT EXAM-TYPE ASSESSMENTS? These would likely be a mid-term and a final, and they would also be assessed on the ✓, ✓+, and NY scale.

THE GRADE

In order to assign yourself a grade, you need to consider what each level means. We will discuss that as a class, and come to some understanding, but in general an A to me signifies consistent engagement

with the class, doing the work, achieving a good understanding and being able to use the analytic tools and explain the results, where it is likely you will be able to successfully use this knowledge in the future. You should be able to apply the material beyond the context of the exercises in the book, to new data sets and research questions. A grade of B means doing those things somewhat less consistently – possible lapses in engagement or understanding that you don't get



Figure 3 Image created by Copilot using DALL-E 3. Data Analytics is like playing in a band – it takes a lot of practice, and sometimes practicing with others is a lot of fun!

Group Work

Many of us learn well when we work with others, consult with them about problems, Go for it! Your assessment of your learning should be specific to yourself, but getting there can include work with others.

Deadlines

It's important to complete assignments on time so that you get the opportunity to assess your progress, get feedback, and undertake action to get help in a timely fashion if needed. For things where we need a firm deadline, like a final project or final grade submission, it should be clear to all of us why we have that deadline.

Quantity & Quality

Both Quantity AND Quality are important. In this class, quality is more important – be thoughtful, be thorough, and aim for real understanding and ease. I hope you will have opportunities to apply this knowledge in your career and possibly in competitions, so having a solid comfort with the topic will boost your confidence.

Time Management

This can be a challenge, I know. The expectation at Smith for our course is about 3 times the amount of outside class work time as in class. Plan for that much time and if you don't need it all you can back it off

worked through, perhaps one or two missing areas of competence with the subject. A grade of C means some significant struggles with engagement or understanding that might hinder your ability to successfully use this knowledge in the future. If you are headed for a D or E you should probably take the course over again. Refer to the contract rubric (below) for more details on each level.

DELIVERABLES

Course Deliverables include:

- (1) Class activities – In-class Problems. For each chapter, I will present a summary of the material. I will generally stay close to the chapter but will add some additional information or show you a different way to do the analysis now and then. Then, each chapter has a number of exercises at the end. We will do most of these either in class or as homework. At the end of each summary, I will indicate which problems are for in-class work and which are homework. For in-class problems, we will work on them together (groups allowed) during class and then take some time at the end to present our progress. I anticipate that in many cases we will do this work over 2 class periods.
- (2) Homework – There will be additional problems from the chapter assigned as homework. **I expect these to be done on your own.** If your work is not perfect or you have hit a snag and can't move forward, you should turn in your work anyway, letting me know how far you have gotten. Making a good attempt on every problem is more important than turning in perfect work on only a few problems.

- (3) Short Reflective Summary (one page) – How did you experience this chapter? What particularly got your attention or struck you? What was one new thing you learned? Was anything really surprising? Was this chapter personally interesting to you, or not, and either way – why? How do the concepts/ideas in this chapter relate to your daily life? What questions do you have after this chapter? What one thing would you like to explore more? **I am looking for authenticity in these, not a polished academic essay! Better to write in your own words than to use AI – see the policy below.**
- (4) Ethics Project – Starting in SDS 100, you gained exposure to ethical dilemmas in data science. Those were only a few examples of concerns that arise in the field. Your task here is to find one case of an ethical issue that somehow involves *categorical* data, and analyze how it relates. The source can be a book chapter (out of something like Weapons of Math Destruction) or an article from a reputable peer-reviewed academic journal or from a reputable news source⁴. Guiding questions include – what is the ethical issue? What categorical data is involved in the case? How – if at all – was it analyzed? How could bias or ethical difficulty arise from the analysis? What might you do differently (or additionally) based on your learning in this class? **Similarly, I am looking for authenticity – NOT AI – for these. See the policy below.**
- (5) You will develop a final project of some type that will allow you to demonstrate your understanding in a way that works well for you. That will be a data project that includes a poster presentation and discussion of your findings. General guidelines are to identify data of interest to you, pose your own research/exploration questions, and provide a good-quality analysis and interpretation. You may find it easier to work in a pair or group of three on this project. No more than three people in a group.
- (6) Advanced Work – In order to contract for an A or a B, you must also undertake advanced work where you identify another dataset that was NOT used in that chapter, and demonstrate that chapter's techniques and concepts on the fresh data. You will do a brief presentation to your classmates as well as submitting your results to me. You may do this in a pair or solo. If in a pair, make sure both your names are on the final product.
- (7) Consistent and on-time attendance.

Contract Guidelines

Grade Level	Absent or Late	In-Class Activities Missing or Late	Homework Activities Missing or Late	Chapter Reflection Journal Entry	Ethics Project	Final Project	Advanced Work – New Data Sets
A	<=3	<=3	<=1 missing	All	Required	Complete	At least 4
B	<=6	<=4	<=2 missing	<=1 missing	Required	Complete	At least 3
C	<=9	<=5	<=3 missing	<=2 missing	Required	Complete	At least 2
D	<=12	6 or more	<=4 missing	<=3 missing	Not required	Incomplete	0
E	13 or more	8 or more	5 or more	4 or more missing	Not required	Missing	0

⁴ Refer to the Media Bias Chart from Harvard Library, linked in Moodle, to assess reputation of the source. You must include the assessment in your review.

Civil Interactions

Smith is a community and you are a part of that. A class is like a neighborhood – we get to know our neighbors and we offer support and encouragement to one another. You have a responsibility to be respectful of the class community, including coming to class on time, not leaving early, and not disrupting class (although sometimes there are unusual circumstances). You are also responsible for reading this syllabus and keeping a copy for your reference. Use class time strategically to get answers to things that are not clear from readings, videos, or homework. When you have questions where answers would clarify things for you or for other students, you have a *responsibility* to the class to ask those questions.

We should keep space for others' opinions and ideas, even though we may not agree with them. Our society has gotten rather bad about civil public discourse – let's be the solution, not the problem. In academic settings, specifically, disagreements are resolved by reference to research, analysis, and logical argumentation, with discussions often resulting in compromises. If disagreements cannot be resolved, we agree to disagree but still respect each other. However, this in no way means tolerating harassment or discrimination.

You are students, and also developing citizens and professionals, with the responsibilities that go along with that status. You have a responsibility to your peers and to yourself to be prepared for class and to do your work. You are responsible for recalling and applying your previous knowledge when it is applicable to this class. That includes your knowledge of mathematics, time management, writing, verbal communication, and any other field that is relevant. You don't have to be perfect – just do your best, and ask for help when you need it. That's all we can ask.

Email Communications

Communicating over email or text does not change the rules of polite communication. It actually highlights problems - if you say something insulting to someone, it's even worse over email than in person. If you write poorly, e-mail gives the recipient a permanent record of that fact. Start now to practice good email skills that you can use in the workplace. Start off with a salutation – “Dear Dr. ...” or “Professor ...” – and end with your name. Please use first and last names in case I have more than one student with your first name. Include a descriptive topic of the email, and it is very helpful if you include which class you're asking about. So something like “SDS 390 – Homework CH 5”. Please note that I usually answer email within 24 hours, except for weekends and holidays, and sometimes even then. If you have not heard back in that time during usual business days, please send a follow up in case your message got into spam or otherwise lost.

Some communications are best done in person, not over email. If you have a complex question or problem, need to talk out a serious situation or concern, or are in trouble in some way, it is better to come to office hours or call/email me to make an appointment to meet with me. I also have teleconferencing gear and access to software so we can meet “virtually” if it's time sensitive and one of us is not on campus.

Other Electronic Things (especially cell phones)

Turn them off if you are not actively using them for class (for example, as a calculator). If you forget once in a while, turn them off as soon as you realize your mistake (I tend to forget too). If you have a critical situation and need to be available by phone, tell me about it before class. For non-critical situations – well, our time together is precious, and you don't want to waste it taking social calls. Similarly, unrelated texting during class, emailing, surfing the web, etc. are all reasons we may ask you to leave for the day. Most people need to be paying close attention during class.

ACADEMIC INTEGRITY - PLAGIARISM AND CHEATING

Our goal is to learn⁵. That is our main mission. In order to be a strong, vibrant learning community we must have academic integrity. That means you accurately and honestly represent your work to me and to your classmates. Smith's Honor Code gives you a number of privileges that are unusual elsewhere in education, such as non-proctored exams and self-scheduled final exams. I feel so lucky to be part of a learning community based on this kind of integrity and respect. Please be aware that you are doing a disservice – especially to yourself - if you violate our Code. We uphold this standard as a community, and it is one of our most important obligations as scholars. Any suspected Honor Code infractions—even if minor or unintentional—have to be reported to the Academic Honor Board by any of us (both you and me). While our goal with the Code is to provide for an educational experience, violations can result in serious academic consequences (such as a failing grade on an assignment). We should all agree to abide by the code and protect the learning and privileges it affords us all.

Please remember, if you are struggling, or feel pushed by a deadline, reach out and talk to me before cheating or using resources inappropriately. The College's honor code is articulated here: <https://www.smith.edu/about-smith/offices/student-affairs/student-handbook/academic-honor-code-0>

Cheating is using the work of another student or collaborating on work that is supposed to be done individually or misrepresenting the integrity of your work in some way. **This includes use of material generated by Artificial Intelligence (AI) without attribution.** If you use such material, you should cite it. You are responsible for knowing the Smith policy on academic dishonesty and academic integrity. If you cheat in this ungraded class, you are only cheating yourself. If you cheat, I will pursue actions under the honor code. Cheating includes dishonesty of any kind in an examination situation or paper as well as failure to appropriately cite sources used in assignments. ZERO TOLERANCE POLICY – All cases of suspected cheating in this class will be pursued.

Plagiarism is the representation of another's words, ideas, or works as one's own without adequate attribution (citation). This includes work generated by AI (Artificial Intelligence). This is also a serious offense. ANY time you use another's words or ideas you should cite them. This includes material from books, internet sites, students' contributions on Blackboard, magazine, newspaper, or journal articles, other students' papers, their exams, etc. Obviously, you should also avoid having someone write your papers for you! As college students, I expect you to know when you are using someone else's words or ideas. It is NOT HARD to drop a citation into a paper or presentation. Even if your citation is not 100% correct, make the effort. When in doubt, YES, CITE IT!!! Ignorance of the policy or appropriate citation rules is no defense. If you have any questions, ask me BEFORE you turn a paper in. Be extremely careful about this, I have a zero tolerance policy in this area. Why? Professional ethics, and also because it's personal – I am a published author, and if someone took my words and misrepresented them as their own, that would be disrespectful of the work my colleagues and I put into the research and writing to get published. Hopefully you also will be published someday (if you aren't already). You will want people to respect your work too. Even open source materials ask you to cite the work if you use it.

Generative AI is a powerful tool that you have probably already encountered in your educational journey. I myself use generative AI (such as ChatGPT). It is a very powerful tool. It can also make some very bad errors. You are welcome to use AI in this course to help you with the following types of things:

- Outlining responses (not writing them)
- Summarizing online material for your own understanding (not for your team's summary for class)
- Generating project ideas
- Outlining presentations
- Refining and improving your writing

⁵ With thanks to my colleagues from New Faculty Orientation (2024-25) for suggestions on how to format this section.

You should NOT use AI to generate material that

- Expresses your own understanding of the material in the course
- Substitutes for your own work, including any R code.

Use AI as a tool, not a substitute. And (very importantly) if you do use AI for anything, you **must** cite it (see Figure 2 for an example)! Report (1) which AI tool you used (for example, ChatGPT and version, Bard, etc.), (2) what you used it for, and (3) include the prompt you used/conversation log (it's very easy to share those now). In other words, like any resource you use in academic work, cite it! *Using AI without citation, or using AI generated work and representing it as your own sole product is a violation of academic integrity.*

ACCESSIBILITY & INCLUSIVITY

“Teaching inclusively means embracing student diversity in all forms — race, ethnicity, gender, disability, socioeconomic background, ideology, even personality traits like introversion — as an asset.”⁶ Your experiences in this course, whether in class, office visits, or otherwise should be as accessible as possible. If there is some aspect of your experience, including but not limited to learning differences, that impacts your ability to fully reach your academic goals, please communicate that to me. I also recommend that you work with the Smith College Accessibility Resource Center (ARC). ARC facilitates communication between you and your instructors by generating a letter specific to you that indicates the support you need in the classroom or on assignments. If you already have this letter, please bring it to office hours or email it to me, and let's chat about what you need. I want to hear your ideas about your needs, and I will want to have at least a brief synchronous conversation with you. You are not required to give me details about your learning differences, but please know that I am studying for a graduate certificate in Learning Differences and Neurodiversity so I have some expertise in working with a diverse set of students around their learning needs. It is also important to let you know that I am a proponent of the inclusivity model of learning difference, rather than the medical model. That means that I believe we ALL learn in different ways, and not all of those ways are or need to be considered a disability. I am an advocate of a strengths-based rather than a deficit approach. If you have any concerns about your learning needs please talk with me about them.

DAILY SCHEDULE

Note: I am uncertain of the pace at which we will move through material. Therefore dates and coverage should be considered preliminary. Monitor Moodle for schedule updates.

Week	Monday	Wednesday
1	January 27 First Day of Class – Overview of Syllabus, Introductions Discussion about contracts and assessment practices	29 Classroom Agreements <ul style="list-style-type: none">- Contracts- Assessment practices

⁶ From Viji Sathy and Kelly A. Hogan (2023) “How to Make Your Teaching More Inclusive” *The Chronicle of Higher Education*, August 18, <https://www.chronicle.com/article/how-to-make-your-teaching-more-inclusive/>

		Review from Intro Statistics – what do you remember about? <ul style="list-style-type: none"> - Discrete probability distributions - Especially binomial
2	February 3 Review from Intro Statistics – what do you remember about <ul style="list-style-type: none"> - Proportions hypothesis tests - Chi-squared - Tests of slope coefficients in regression 	5 Chapter 1 – Formal introduction
3	10 Continue chapter 1	12 Chapter 2 – Analyzing Contingency Tables
4	17 Continue chapter 2	19 Chapter 3 – Generalized Linear Models
5	24 Continue chapter 3	26 Chapter 4 – Logistic Regression
6	March 3 Continue chapter 4	5
7	10	12
8	17 SPRING BREAK	19 SPRING BREAK
9	24	26
10	31	April 2
11	7	9
12	14	16
13	21	23
14	28	30

15	STUDY PERIOD – No Classes	5/3 – 5/5
Finals	FINAL EXAM PERIOD	5/6 – 5/9

NOTES

1. I reserve the right to make changes to this syllabus as needed. Changes will be posted to Moodle and you will be notified in class and/or via messaging.
2. **Bad Weather Policy** – I live up near the Vermont border, and the weather is sometime quite different there from the Northampton weather. In the event that classes are in session but I cannot get here, I will hold class via Zoom. Watch for links in our Moodle course site.