

CHRISTOPHER G. PRENER, PH.D.

# SYLLABUS

SOC 5050: QUANTITATIVE ANALYSIS  
FALL, 2016  
SAINT LOUIS UNIVERSITY



# Course Introduction

## Meeting Times and Office Hours

---

### Contact

Office: 1918 Morrissey Hall

Phone: 314-977-4276

Email: [prenercg@slu.edu](mailto:prenercg@slu.edu)

### Course Meeting

Mondays

4:20pm to 6:50pm

3600 Morrissey (GIS Lab)

---

### Office Hours

Wednesdays:

2:30pm to 4:00pm in 3600 Morrissey (GIS Lab)

4:00pm to 5:00pm in 1918 Morrissey (Office)

---

*Figures often beguile me, particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli would often apply with justice and force: 'There are three kinds of lies: lies, damned lies, and statistics'.*  
Mark Twain (1906)

## Course Description

This course provides an introduction to applied statistical analysis for graduate students with an emphasis placed on statistical techniques that are most common in the sociological literature. The statistical techniques introduced include measures of central tendency and dispersion as well as measures of bivariate association. Multivariate statistical analyses are also introduced. While the examples may be specific to the social sciences, the theories and skills that are covered are broadly applicable across academic disciplines.

## Course Objectives

This course has four intertwined objectives:

1. *Fundamentals of inferential statistics*: Describe the use of various statistical tests, their requirements and assumptions, and their interpretation; execute these tests both by hand and programmatically using Stata.
2. *Fundamentals of data analysis*: Perform basic data cleaning and analysis tasks programmatically using Stata in ways that support high quality documentation and replication.
3. *Fundamentals of data visualization*: Create and present publication-quality plots programmatically using Stata.
4. *Quantitative research synthesis*: Plan, implement, and present a research project that uses a large, publicly available sociology dataset known as the General Social Survey.

## Core Documents

There are three core documents for this course. This **Syllabus** sets out core expectations and policies for the course. The **Reading List** contains topics, readings (both required and optional), and assignment due dates for each week. These two documents spell out what is *required* for this course. The **User's Guide** is a companion document that is written with the aim of helping you be *successful* in this course. The idea behind a course User's Guide is to create a reference for many of the intangible, subtle or disparate skills and ideas that contribute to being a successful researcher. All course documents are available in the Core-Documents repository on [GitHub](#).<sup>1</sup>

<sup>1</sup> GitHub is a website designed for collaboratively working on programming and data analysis projects. It is described briefly below (see page 5) and in more detail in the **User's Guide**.

## Readings

There are four books required for this course. Each book has been selected to correspond with one or more of the course objectives. The books are:

1. Acock, Alan. 2016. *A Gentle Introduction to Stata*. 5<sup>th</sup> edition. College Station, TX: Stata Press. ISBN-13: 978-1597181426; List Price: \$79.95; ebook versions available.
2. Freedman, David, Robert Pisani, and Roger Purves. 2007.

*Statistics*. 4<sup>th</sup> edition. New York, NY: W.W. Norton and Company. ISBN-13: 978-0393929720; List Price: \$140.00; ebook versions available.

3. Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press. ISBN-13: 978-1597180474; List Price: \$67.95; ebook versions available.
4. Wheelan, Charles. 2014. *Naked Statistics: Stripping the Dread from the Data*. New York, NY: W.W. Norton and Company. ISBN-13: 978-0393347777; List Price: \$16.95; ebook versions available.

There is one additional book that is optional but highly recommended<sup>2</sup>:

5. Mitchell, Michael N. 2012. *A Visual Guide to Stata Graphics*. 3<sup>rd</sup> edition. College Station, TX: Stata Press. ISBN-13: 978-1597181068; List Price: \$62.00; ebook versions available.

<sup>2</sup> This is particularly the case for students who want to build their skill-set in data visualization and those who foresee themselves undertaking a quantitative thesis or dissertation.

I do not require students to buy physical copies of texts. You are free to select a means for accessing these texts that meets your budget and learning style. If ebook editions (e.g. Kindle, iBooks, pdf, etc) of texts are available, they are acceptable for this course. All texts should be obtained in the edition noted above.

All readings are listed on the **Reading List** and should be completed before the course meeting on the week in which they are assigned. Full text versions of most readings not found in the books assigned for the course can be obtained using the library's [Electronic Reserves](#) system. The password for the Electric Reserves will be emailed out to students at the beginning of the semester.

## *GitHub*

The majority of course content (sample code, documentation, assignments, and some readings) for this course will be made available using [GitHub](#).<sup>3</sup> GitHub is a website used by programmers, data analysts, and researchers to share computer code and projects. GitHub will also be used for assignment submission and feedback. In addition to providing us with platform for hosting course content, using GitHub will give you experience in some of the techniques that researchers use to conduct both open-source and collaborative research. Finally, we will use GitHub's [wiki tool](#) to host a weekly jotter with course notes, links, and updates. The jotter is located on the Core-Documents repository's [wiki page](#).

<sup>3</sup> You will need to [sign-up](#) for a free GitHub account and then become a member of the [SOC 5050 course organization](#) at the beginning of the semester.

## Software

This course uses the statistical software program **Stata**, which is available for Windows and OS X. Stata is available in 3600 Morrissey (GIS Lab) and thus purchasing it is not required for this course. However, I recommend purchasing Stata if you foresee yourself completing quantitative research or a quantitative dissertation.<sup>4</sup>

I use and recommend Stata not because it is easy but because it is powerful. For most uses, I believe that it is simply the best statistical package available. It is an open-platform, meaning that user-written packages are available that extend and expand Stata's base capabilities. Since Stata is designed to operate based on coded commands, it is both powerful and flexible. Learning to program for Stata will position you well not only to conduct reproducible research<sup>5</sup> but also to make the jump to R.<sup>6</sup>

You will also need a free application called **Atom** ([website](#)). Atom is a text editor for both Windows and OS X developed by the same team that operates GitHub. It is available in 3600 Morrissey (GIS Lab). We will use Atom for editing Stata code and for writing in a markup language called **Markdown**.<sup>7</sup> Markdown is the "official" language for writing documentation on GitHub, and you will use it for formatting your lab assignments, problem sets, and research files. Writing in simple, open-source documents helps to ensure your work is reproducible and can be opened without issue on (almost) any computer in the future. Once again, we do this not necessarily because it is the easiest approach but because it has significant advantages over the long-term.<sup>8</sup>

Finally, you will need another free application called **GitHub Desktop** ([website](#)). This program allows you to easily copy data from GitHub onto your computer. It also makes it easy to upload files like labs and problem sets to GitHub. It is available for both Windows and OS X, and is available in 3600 Morrissey (GIS Lab).<sup>9</sup>

## Data

I will make all of the required data available on Dropbox<sup>10</sup> in a single .zip file. These data should be downloaded and extracted as soon as possible after the start of the semester. The .zip file will include data for all textbooks as well as the data for the labs, problem sets, and final project. Data for labs and problem sets will be drawn from the 2011 Current Population Survey (CPS) and the 2013 National Health Interview Survey (NHIS). Data for the final project will be drawn from the 2012 General Social Survey (GSS).

<sup>4</sup> Stata can be purchased by students at a reduced rate through the Stata Corporation's [website](#). If you buy Stata for yourself, please note that Small Stata will not be sufficient for completing the course; you must use a minimum of Stata/IC to complete coursework.

<sup>5</sup> One of the mantras of the course; see the **User's Guide** for more details.

<sup>6</sup> R is the wave of the future. Since it is available for free and exceptionally powerful, many organizations and academics are moving their workflows to R.

<sup>7</sup> Markup languages are used to specify how outputted text will look, as opposed to "what-you-see-is-what-you-get" editors like Microsoft Word that allow you to format your text as you write.

<sup>8</sup> Details on setting up Atom and writing in Markdown and are included in the **User's Guide**.

<sup>9</sup> Details on setting up GitHub Desktop are included in the **User's Guide**.

<sup>10</sup> Dropbox is a cloud storage service. You do not need a Dropbox account to download these data. However, I *highly* recommend obtaining a free account and using Dropbox to store files instead of keeping them stored locally on your computer. This should be one element of your strategy for protecting your work, something which is discussed further in the **User's Guide**.

# *Course Expectations and Policies*

## *General Expectations for Students*

My priority is that class periods are productive learning experiences for all students. In order to foster this type of productive environment, I ask students to follow a few policies:

1. Arrive to class on time and stay for the entire class period.<sup>11</sup>
2. Silence *all* electronic devices before entering the classroom.
3. Do not engage in side conversations. This is disrespectful to the speaker (whether me or a classmate), and can affect the ability of others in the class to learn.
4. Be respectful of your fellow classmates. Do not interrupt when someone is speaking, monopolize the conversation, or belittle the ideas or opinions of others.
5. Complete the assigned readings for each class in advance, and come prepared with discussion points and questions.

<sup>11</sup> If you drive to campus, please get an on-campus parking pass for the semester or use a smartphone app to top off your meter. Leaving class to feed your meter is disruptive for both you and your classmates.

## *Attendance and Participation*

Attendance and participation are important components of this course since we only meet once a week. Students are expected to attend all class sessions. Making up missed classes are your responsibility. Please email me if you will be missing class so that we can touch base about any assignments, and make sure to obtain notes from a classmate. I do post slides to GitHub, but my slides are intended only to serve as references. Please note that lectures and discussions cannot be recorded by any means (e.g. audio or video recordings, or photographs) without my permission.

## *Communication*

Email is my preferred method of communication. I dedicate time to email responses each workday, meaning that my response time

is typically within 24 hours during the workweek. If you have not received a response from me after 48 hours (or by end of business on Monday if you emailed me over the weekend), please follow-up to ensure that your message did not get lost in the shuffle.

Please use your SLU email account when emailing me. All messages regarding course updates, assignments, and changes to the class schedule including cancellations will be sent to your SLU email account. It is imperative that you check your SLU email account regularly.

Please also ensure that all concerns or questions about your standing in the course are directed to me immediately. Inquires from parents, SLU staff members, and others will not be honored.

### *Electronic Devices*

During class periods, students are asked to refrain from using electronic devices (including cell phones) for activities not directly related to the course. For this class, I expect students to limit their use of electronic devices to accessing course software, readings, and notes.

There is evidence that using electronic devices during lectures results in decreased retention of course content<sup>12</sup> and lower overall course performance.<sup>13</sup> Students who are not using a laptop but are in direct view of another student's laptop also have decreased performance in courses.<sup>14</sup> Conversely, students who take notes the "old fashioned way" have better performance on tests compared to students who take notes on laptops.<sup>15</sup>

I therefore ask students to be conscious of how they are using their devices, the ways such use impacts their own learning, and the effect that it may have on others around them. I reserve the right to alter this policy if electronic device use becomes problematic during the semester.

### *Student Support*

If you meet the eligibility requirements for **academic accommodations** through the Disability Services office (located within the Student Success Center), you should arrange to discuss your needs with me after the first class. All discussions of this nature are treated confidentially, and I will make every effort to work with you to come up with a plan for successfully completing the course requirements. Please note that I will not provide accommodations to students who are not working with Disability Services.<sup>16</sup>

<sup>12</sup> Hembrooke, Helene and Geri Gay. 2003. "The Laptop and the Lecture: The Effects of Multitasking in Learning Environments". *Journal of Computing in Higher Education* 5(1): 46-64.

<sup>13</sup> Fried, Carrie. 2008. "In-class laptop use and its effects on student learning". *Computers & Education* 50(3): 906-914.

<sup>14</sup> Sana, Faria et al. 2013. "Laptop Multitasking Hinders Classroom Learning for Both Users and Nearby Peers". *Computers & Education* 62: 24-31.

<sup>15</sup> Mueller, Pam and Daniel Oppenheimer. 2014. "The Pen Is Mightier Than the Keyboard Advantages of Longhand Over Laptop Note Taking". *Psychological Science* 25(6): 1159-1168.

<sup>16</sup> Additional details can be found on the Disability Services [website](#). You can contact them at [disability\\_services@slu.edu](mailto:disability_services@slu.edu) or 314-977-3484 to schedule an appointment.



If you are a **student-athlete** who is in-season, you should discuss your game schedule with me after the first class and share your travel letter with me as soon as you have a copy. You are reminded that games and tournaments are not excuses for failing to complete assignments, and that NCAA rules prohibit student-athletes from missing classes for practice. Low grades that jeopardize eligibility must be addressed immediately by you, not by a coach or academic coordinator.<sup>17</sup>

I also encourage you to take advantage of the **University Writing Services (UWS) program**. Getting feedback benefits writers at all skill levels and the quality of your writing will be reflected in assignment grades. The UWS has trained writing consultants who can help you improve the quality of your written work. UWS's consultants are available to address everything from brainstorming and developing ideas to crafting strong sentences and documenting sources.<sup>18</sup>

<sup>17</sup> More information about resources and academic support for student-athletes can be found at the Student-Athlete Academic Support Services [website](#).

<sup>18</sup> More information on the UWS program can be found on their [website](#). The UWS program has a number of on-campus locations.

### *Academic Honesty*

All students should familiarize themselves with [Saint Louis University's policies](#) concerning cheating, plagiarism, and other academically dishonest practices.<sup>19</sup>

<sup>19</sup> This course is also governed by the College of Arts and Sciences' academic honesty policies, which are available on their [website](#).

*Academic integrity is honest, truthful and responsible conduct in all academic endeavors. The mission of Saint Louis University is "the pursuit of truth for the greater glory of God and for the service of humanity." Accordingly, all acts of falsehood demean and compromise the corporate endeavors of teaching, research, health care, and community service via which SLU embodies its mission. The University strives to prepare students for lives of personal and professional integrity, and therefore regards all breaches of academic integrity as matters of serious concern.*

Any work that is taken from another student, copied from printed material, or copied the internet without proper citation is expressly prohibited. Note that this includes all computer code written for class assignments - each student is expected to author and de-bug their own code.<sup>20</sup> All assignments should include in-text citations and references formatted using the American Sociological Association [style guidelines](#). Any student who is found to have been academically dishonest in their work risks failing both the assignment and this course.

<sup>20</sup> See the **User's Guide** for details on collaboration with your colleagues.

## *Title IX*

All students should familiarize themselves with Saint Louis University's policies on bias, discrimination, and harassment:

*Saint Louis University and its faculty are committed to supporting our students and seeking an environment that is free of bias, discrimination, and harassment. If you have encountered any form of sexual misconduct (e.g. sexual assault, sexual harassment, stalking, domestic or dating violence), we encourage you to report this to the University. If you speak with a faculty member about an incident of misconduct, that faculty member must notify SLU's Title IX coordinator, Anna R. Kratky (DuBourg Hall, Room 36; [akratky@slu.edu](mailto:akratky@slu.edu)); 314-977-3886) and share the basic facts of your experience with her. The Title IX coordinator will then be available to assist you in understanding all of your options and in connecting you with all possible resources on and off campus.*

*If you wish to speak with a confidential source, you may contact the counselors at the University Counseling Center at 314-977-TALK.*

# Course Assignments and Grading

## *Assignments*

*Attendance and Participation (10%)* As discussed above, both attendance and participation are important aspects of this class. The class participation grade will be based on (a) attendance, (b) level of engagement during lectures and labs, and (c) occasional class “entry” and “exit tickets”.

*Lab Exercises (10%)* Each course meeting will include time dedicated to practicing the techniques and applying the theories described during the day’s lecture. There will be a total of sixteen lab exercises over the course of the semester, each of which is worth approximately 0.63% of your final grade. These exercises will give you an opportunity to practice skills that correspond with the first three course objectives. Lab exercises will build upon each other, meaning that they will increase in complexity each week.

It is imperative that you keep up with these exercises and seek out help from myself or your colleagues if you have questions about the course materials. Lab exercises are graded simply for completion. Exercise documents will detail the deliverables to be submitted to demonstrate completion of the assignment. Replication files will be posted on GitHub that illustrate my approach to each exercise.

*Problem Sets (30%)* There will be a total of ten problem sets over the course of the semester, each of which is worth 3% of your final grade. Problem sets will require students to draw on a variety of skills, including cleaning data in Stata, visualizing distributions and relationships, and performing statistical analyses as well as calculating test statistics by hand. They are designed to assess your progress with the first three course objectives.

Like lab exercises, problem sets will build upon each other. As with labs, students must ensure that they are keeping up with these assignments. Problem set documents will detail the deliverables to be submitted to demonstrate completion of the assignment. Replication

files will be posted on GitHub that illustrate my approach to each exercise.

*Final Paper and Presentation (35%)* The data analysis paper (which corresponds to the fourth course objective) will be completed using the 2012 wave of the General Social Survey. Each student will be responsible for selecting a research question, setting out hypotheses based on a brief literature review, and performing analyses to these these hypotheses. The final paper will include that brief literature review as well as a description of the analyses used, and an interpretation of the results. Students will be expected to include formatted tables, graphs as needed, and all associated Stata code and output.<sup>21</sup>

Students will be required to submit a brief memo (worth 5%) identifying your topic, citations for relevant articles, and key variables in class on **September 12<sup>th</sup>**. The literature review, hypotheses, and data analysis plan (worth 5%) will be due on **October 10<sup>th</sup>**. Drafts of the presentation and paper, which will only be graded for completion, will be due in class on **November 14<sup>th</sup>**. Students will be expected to deliver a research presentation (much like a conference presentation; worth 10%) describing their findings during our course's final exam period on **December 12<sup>th</sup>**. The final paper and associated deliverables (data, code, and documentation; worth 15%) will be due via email and hardcopy by **4:20pm on December 12<sup>th</sup>**.

<sup>21</sup> Detailed instructions will be provided on Blackboard under 'Assignments' and we will talk about this project in class as the semester progresses.

*Quizzes (15%)* Three short quizzes covering statistical theories and concepts will be given throughout the semester. Like the problem sets, these are designed to assess your progress with the first three course objectives. Each will constitute 5% of your final grade. Quizzes will be given on the following schedule: Quiz 1 on **October 3<sup>rd</sup>**, Quiz 2 on **November 7<sup>th</sup>**, and Quiz 3 on **December 5<sup>th</sup>**.

### *Submission and Late Work*

Copies of all assignment materials (written responses, code, and visualizations) should be uploaded to your private repository on GitHub before class on the day that the assignments are due. Written responses should be uploaded as properly formatted Markdown (.md) files.

Once the class begins, any assignments submitted will be treated as late. Assignments handed in within 24-hours of the beginning of class will have 15% deducted from the grade. I will deduct 15% per day for the next two 24-hour periods that assignments are late; after 72-hours, I will not accept late work.<sup>22</sup> If you cannot attend

<sup>22</sup> These periods begin at 4:20pm each day. All assignments must be turned in by 4:20pm on the Thursday after they are due to receive partial credit.

class because of personal illness, a family issue, jury duty, an athletic match, or a religious observance, you must contact me beforehand to discuss alternate submission of work.

These same policies cover missed quizzes - you may only take a quiz at an alternate time for one of the aforementioned reasons and you must speak with me before the quiz date to schedule a makeup quiz. Failure to do so will result in a "0". Please note that non-emergent travel plans are not considered valid reasons for missing an quiz or an assignment submission.

### *Extra Credit*

From time to time I may offer extra credit to be applied to your final grade. I will only offer extra credit if it is open to the entire class (typically for something like attending a lecture or event on-campus). If I offer extra credit, I will typically require you to submit a short written summary of the activity within a week of the event to obtain the credit. When offered, extra credit opportunities cannot be made-up or substituted if you are unable to attend the event.

### *Grading*

Grades will be included with assignment feedback and will also be made available through Blackboard's grade center.<sup>23</sup> I use a point system for calculating grades. The following table gives the weighting and final point totals for all assignments for this course:

<sup>23</sup> This is the only official use of Blackboard for this course.

Assignment	Grading Point System			
	Weight	Points	Qty.	Total
Attendance	10%	100 pts	x1	100 pts
Labs	10%	6.25 pts	x16	100 pts
Problem Sets	30%	30 pts	x10	300 pts
Final Project	35%	350 pts	x1	350 pts
Quizzes	15%	50 pts	x3	150 pts
<i>Total</i>	100%			1,000 pts

I will return assignments with grades that represent number of points earned. If you want to know your percentage on a particular assignment, divide the number of points earned by the number of points possible and then multiple it by 100.

Final grades will be calculated by taking the sum of all points earned and dividing it by the total number of points possible (1,000). This will be multiplied by 100 and then converted to a letter grade

using the following table:

Final Grading Scale					
Grade Point	Letter	Percentage	Grade Point	Letter	Percentage
4.0	A	93.0-100	2.3	C+	77.0-79.9
3.7	A-	90.0-92.9	2.0	C	73.0-76.9
3.3	B+	87.0-89.9	1.7	C-	70.0-72.9
3.0	B	83.0-86.9	1.0	D	63.0-69.9
2.7	B-	80.0-82.9	0.0	F	< 63.0

Borderline grades (e.g. 89.9%) will not be rounded up, so please do not ask. No chances will be given for revisions of poor grades. Incomplete grades will be given upon request only if you have a "C" average and have completed at least two-thirds of the assignments. You should note that incomplete grades must be rectified by the specified deadline or they convert to an "F".

# Course Schedule

## Course Overview

Week	Dates	Topics
1	August 22 <sup>nd</sup>	Course Introduction
2	August 29 <sup>th</sup>	Describing Distributions
3	September 5 <sup>th</sup>	Working with Data (Part 1)
4	September 12 <sup>th</sup>	Probability and Bayes' Theorem
5	September 19 <sup>th</sup>	Distribution of Random Variables
6	September 26 <sup>th</sup>	Foundations for Inference
7	October 3 <sup>rd</sup>	Difference of Means (Part 1)
8	October 10 <sup>th</sup>	Difference of Means (Part 2)
9	October 17 <sup>th</sup>	Working with Data (Part 2)
10	October 24 <sup>th</sup>	Correlations (Part 1)
11	October 31 <sup>st</sup>	Correlations (Part 2)
12	November 7 <sup>th</sup>	ANOVA
13	November 14 <sup>th</sup>	Bivariate Regression
14	November 21 <sup>st</sup>	Multivariate Regression (Part 1)
15	November 28 <sup>th</sup>	Multivariate Regression (Part 2)
16	December 5 <sup>th</sup>	Analyzing Categorical Data
17	December 12 <sup>th</sup>	Finals Week - Presentations

## Planned Tegrity Lectures

This semester, we have two classes that fall on official university holidays: **Labor Day** (Week 3, September 5<sup>th</sup>) and **Fall Break** (Week 9, October 17<sup>th</sup>). These two weeks will both have materials assigned for them, which will include a Tegrity<sup>24</sup> lecture. These two lectures will be shorter than typical in-class lectures. Students should view these lectures during that week and complete the associated readings and lab exercises. A link to the video for each lecture will be sent out ahead of time.

<sup>24</sup> Tegrity is a lecture capture system that allows instructors to capture video and audio of a lecture as well as images of the associated lecture slides.

## *Reading List*

Please consult the stand-alone **Reading List** document for details on readings and assignments for each week.<sup>25</sup>

<sup>25</sup> All course documents including the **Reading List** are available in the Core-Documents repository on [GitHub](#).

## *Class Progression*

Each week will be broken down roughly the same way. Class will begin with any relevant “follow-up” from the previous weeks and relevant announcements. We will then segue into a short lecture on a data management topic before tackling the main statistical topic for the week. For many weeks, this will include a discussion of a chapter from Wheelan (2014). Depending on the week, we will take an in-depth look either at calculating a particular test statistic by hand or using Stata to estimate the same statistic. Most classes will end with time dedicated to working through the lab exercise.

## *Scheduling Notes*

The weekly schedule may change as it depends on the progress of the class. However, you must keep up with the reading assignments. In the event of a cancellation due to weather or another disruption, I may alter the weekly schedule.

Since this course only meets once per week, cancellations are particularly disruptive. I will make every effort to schedule make-up classes at a time that works for at least a portion of the class. These class sessions will be recorded and made immediately available using Tegrity for students who are unable to attend the make-up class. All students will be responsible for either attending the make-up class or watching the lecture as well as completing all readings, lab assignments, and problem sets for make-up classes.