

Social Science Inquiry II

Week 1: Course introduction

Molly Offer-Westort

Department of Political Science,
University of Chicago

Winter 2023

Overview

This course:

- ▶ Part of a sequence
 - ▶ Fall: research design
 - ▶ Winter: methods
 - ▶ Spring: practical applications

Did you take SOSC 13100 in fall quarter of 2022?

Overview

This course:

- ▶ Part of a sequence
 - ▶ Fall: research design
 - ▶ Winter: methods
 - ▶ Spring: practical applications

Did you take SOSC 13100 in fall quarter of 2022?

Overview

This course:

- ▶ Part of a sequence
 - ▶ Fall: research design
 - ▶ Winter: methods
 - ▶ Spring: practical applications

Did you take SOSC 13100 in fall quarter of 2022?

Overview

This course:

- ▶ Part of a sequence
 - ▶ Fall: research design
 - ▶ Winter: methods
 - ▶ Spring: practical applications

Did you take SOSC 13100 in fall quarter of 2022?

Overview

This course:

- ▶ Part of a sequence
 - ▶ Fall: research design
 - ▶ Winter: methods
 - ▶ Spring: practical applications

Did you take SOSC 13100 in fall quarter of 2022?

Overview

This course:

- ▶ Part of a sequence
 - ▶ Fall: research design
 - ▶ Winter: methods
 - ▶ Spring: practical applications

Did you take SOSC 13100 in fall quarter of 2022?

Main take-aways from fall quarter?

Course objectives

- ▶ Students will approach data thinking like a social scientist; i.e., thinking about what data inputs are, how they're measured, relationships among variables, thinking about data generating processes when making inferences
- ▶ Build on foundations of causal inference introduced in fall SSI I, with respect to both experimental and observational data
- ▶ Have facility with basic statistical tests and methods
- ▶ Have facility with basic statistical software programming and data visualization

Course objectives

- ▶ Students will approach data thinking like a social scientist; i.e., thinking about what data inputs are, how they're measured, relationships among variables, thinking about data generating processes when making inferences
- ▶ Build on foundations of causal inference introduced in fall SSI I, with respect to both experimental and observational data
- ▶ Have facility with basic statistical tests and methods
- ▶ Have facility with basic statistical software programming and data visualization

Course objectives

- ▶ Students will approach data thinking like a social scientist; i.e., thinking about what data inputs are, how they're measured, relationships among variables, thinking about data generating processes when making inferences
- ▶ Build on foundations of causal inference introduced in fall SSI I, with respect to both experimental and observational data
- ▶ Have facility with basic statistical tests and methods
- ▶ Have facility with basic statistical software programming and data visualization

Course objectives

- ▶ Students will approach data thinking like a social scientist; i.e., thinking about what data inputs are, how they're measured, relationships among variables, thinking about data generating processes when making inferences
- ▶ Build on foundations of causal inference introduced in fall SSI I, with respect to both experimental and observational data
- ▶ Have facility with basic statistical tests and methods
- ▶ Have facility with basic statistical software programming and data visualization

Course objectives

- ▶ Students will approach data thinking like a social scientist; i.e., thinking about what data inputs are, how they're measured, relationships among variables, thinking about data generating processes when making inferences
- ▶ Build on foundations of causal inference introduced in fall SSI I, with respect to both experimental and observational data
- ▶ Have facility with basic statistical tests and methods
- ▶ Have facility with basic statistical software programming and data visualization

Course objectives

By the end of the quarter, you should be able to ...

- ▶ Formulate a falsifiable research “hypothesis”
- ▶ Determine what evidence (data, statistical tests) you would need to falsify that hypothesis
- ▶ Interpret results from statistical analysis
- ▶ Run basic statistical analysis in R

Course objectives

By the end of the quarter, you should be able to ...

- ▶ Formulate a falsifiable research “hypothesis”
- ▶ Determine what evidence (data, statistical tests) you would need to falsify that hypothesis
- ▶ Interpret results from statistical analysis
- ▶ Run basic statistical analysis in R

Course objectives

By the end of the quarter, you should be able to . . .

- ▶ Formulate a falsifiable research “hypothesis”
- ▶ Determine what evidence (data, statistical tests) you would need to falsify that hypothesis
- ▶ Interpret results from statistical analysis
- ▶ Run basic statistical analysis in R

Course objectives

By the end of the quarter, you should be able to . . .

- ▶ Formulate a falsifiable research “hypothesis”
- ▶ Determine what evidence (data, statistical tests) you would need to falsify that hypothesis
- ▶ Interpret results from statistical analysis
- ▶ Run basic statistical analysis in R

Course objectives

By the end of the quarter, you should be able to . . .

- ▶ Formulate a falsifiable research “hypothesis”
- ▶ Determine what evidence (data, statistical tests) you would need to falsify that hypothesis
- ▶ Interpret results from statistical analysis
- ▶ Run basic statistical analysis in R

Overview

What about qualitative research methods?

- ▶ Data can be both quantitative and qualitative.
- ▶ Principles of research design can be shared across methodological approaches.

Overview

What about qualitative research methods?

- ▶ Data can be both quantitative and qualitative.
- ▶ Principles of research design can be shared across methodological approaches.

Overview

What about qualitative research methods?

- ▶ Data can be both quantitative and qualitative.
- ▶ Principles of research design can be shared across methodological approaches.

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
 - ▶ Formalizing our uncertainty about that model, using statistics
 - ▶ What it requires to infer causal relationships
 - ▶ Linear regression (and inference with)
 - ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

Overview

- ▶ High level: how to answer questions with data
- ▶ Summarizing univariate and multivariate data numerically and visually, data exploration
- ▶ Probability as a model of the world
- ▶ Formalizing our uncertainty about that model, using statistics
- ▶ What it requires to infer causal relationships
- ▶ Linear regression (and inference with)
- ▶ Beyond linear regression: other statistical methods

What we do with data

We'll get started on working with data, and we'll also get started on thinking critically about how you use data to answer questions.

What data would you need to make the argument in the article below?

A screenshot of the top portion of a news article from The Guardian. The header is dark blue with the Guardian logo and 'For 200 years' in white. Below the header, the title 'More officer diversity won't cut racial disparity in US police shootings - study' is in large black font. A sub-headline in bold black font reads: 'Research found as percentage of black officers who fired in fatal shootings increased, the citizen shot was more likely to be black'. Below this, the author 'Miranda Bryant in New York' is written in red. At the bottom, the date and time 'Mon 22 Jul 2019 17:56 EDT' are shown in a small grey font.

Sign in **The Guardian**
For 200 years

More officer diversity won't cut racial disparity in US police shootings - study

Research found as percentage of black officers who fired in fatal shootings increased, the citizen shot was more likely to be black

Miranda Bryant *in New York*

Mon 22 Jul 2019 17:56 EDT

What we do with data

We'll get started on working with data, and we'll also get started on thinking critically about how you use data to answer questions.

What data would you need to make the argument in the article below?

A screenshot of the top portion of a news article from The Guardian. The header is dark blue with the Guardian logo and 'For 200 years' in white. Below the header, the title 'More officer diversity won't cut racial disparity in US police shootings - study' is in large black font. Underneath the title is a bolded summary: 'Research found as percentage of black officers who fired in fatal shootings increased, the citizen shot was more likely to be black'. Below this is the author's name 'Miranda Bryant' in red, followed by 'in New York' in red. At the bottom of the header section is the date and time 'Mon 22 Jul 2019 17:56 EDT' in a small grey font.

Sign in **The Guardian**
For 200 years

More officer diversity won't cut racial disparity in US police shootings - study

Research found as percentage of black officers who fired in fatal shootings increased, the citizen shot was more likely to be black

Miranda Bryant *in New York*

Mon 22 Jul 2019 17:56 EDT

Inferential questions

- ▶ What can the data you *do* have tell you about data you *don't* have?
- ▶ What data would you need to answer questions about *what would have happened*?
- ▶ What can we say about our *uncertainty* about estimates or predictions?

Inferential questions

- ▶ What can the data you *do* have tell you about data you *don't* have?
- ▶ What data would you need to answer questions about *what would have happened*?
- ▶ What can we say about our *uncertainty* about estimates or predictions?

Inferential questions

- ▶ What can the data you *do* have tell you about data you *don't* have?
- ▶ What data would you need to answer questions about *what would have happened*?
- ▶ What can we say about our *uncertainty* about estimates or predictions?

Statistical software

Programming in R.

- ▶ Assuming you're starting from scratch.
- ▶ But if you do have some programming skills, with statistical software, python, or other languages, there can be a lot of carryover.
- ▶ If you don't have any baseline skills, focus on the rstudio primers <https://rstudio.cloud/learn/primers>

Statistical software

Programming in R.

- ▶ Assuming you're starting from scratch.
- ▶ But if you do have some programming skills, with statistical software, python, or other languages, there can be a lot of carryover.
- ▶ If you don't have any baseline skills, focus on the rstudio primers <https://rstudio.cloud/learn/primers>

Statistical software

Programming in R.

- ▶ Assuming you're starting from scratch.
- ▶ But if you do have some programming skills, with statistical software, python, or other languages, there can be a lot of carryover.
- ▶ If you don't have any baseline skills, focus on the rstudio primers
<https://rstudio.cloud/learn/primers>

Statistical software

Programming in R.

- ▶ Assuming you're starting from scratch.
- ▶ But if you do have some programming skills, with statistical software, python, or other languages, there can be a lot of carryover.
- ▶ If you don't have any baseline skills, focus on the rstudio primers <https://rstudio.cloud/learn/primers>

Reference texts

[See syllabus]

Asking and answering questions

- ▶ Google!
- ▶ Ask on our course StackOverflow team:
<https://stackoverflowteams.com/c/sosc13200-w23-2> [I'll send invites shortly]
- ▶ Use general StackOverflow (<https://stackoverflow.com/>) for R questions or CrossValidated (<https://stats.stackexchange.com/>) for stats.

Class participation includes asking and *answering* questions on the class StackOverflow.

Asking and answering questions

- ▶ Google!
- ▶ Ask on our course StackOverflow team:
<https://stackoverflowteams.com/c/sosc13200-w23-2> [I'll send invites shortly]
- ▶ Use general StackOverflow (<https://stackoverflow.com/>) for R questions or CrossValidated (<https://stats.stackexchange.com/>) for stats.

Class participation includes asking and *answering* questions on the class StackOverflow.

Asking and answering questions

- ▶ Google!
- ▶ Ask on our course StackOverflow team:
<https://stackoverflowteams.com/c/sosc13200-w23-2> [I'll send invites shortly]
- ▶ Use general StackOverflow (<https://stackoverflow.com/>) for R questions or CrossValidated (<https://stats.stackexchange.com/>) for stats.

Class participation includes asking and *answering* questions on the class StackOverflow.

Asking and answering questions

- ▶ Google!
- ▶ Ask on our course StackOverflow team:
<https://stackoverflowteams.com/c/sosc13200-w23-2> [I'll send invites shortly]
- ▶ Use general StackOverflow (<https://stackoverflow.com/>) for R questions or CrossValidated (<https://stats.stackexchange.com/>) for stats.

Class participation includes asking and *answering* questions on the class StackOverflow.

Asking and answering questions

- ▶ Google!
- ▶ Ask on our course StackOverflow team:
<https://stackoverflowteams.com/c/sosc13200-w23-2> [I'll send invites shortly]
- ▶ Use general StackOverflow (<https://stackoverflow.com/>) for R questions or CrossValidated (<https://stats.stackexchange.com/>) for stats.

Class participation includes asking and *answering* questions on the class StackOverflow.

Assessment

- ▶ Homework: 60%
- ▶ Final project: 30%
- ▶ Participation: 10%

Homework

- ▶ Solution sets will be posted *at the same time* as problem sets.
- ▶ Do as much as you can on the problem set before checking the solutions.
- ▶ Check your work, and then fill out a form on how you did, what you understood and didn't.
- ▶ You get marked both on completion of the problem set, **AND** filling out the form.
- ▶ (If you find errors in the solution set, post them on the class StackOverflow and you will get extra credit)
- ▶ For homework assignments, always submit *both* your .R file showing your work, and a compiled .pdf file on Canvas.

Homework

- ▶ Solution sets will be posted *at the same time* as problem sets.
- ▶ Do as much as you can on the problem set before checking the solutions.
- ▶ Check your work, and then fill out a form on how you did, what you understood and didn't.
- ▶ You get marked both on completion of the problem set, **AND** filling out the form.
- ▶ (If you find errors in the solution set, post them on the class StackOverflow and you will get extra credit)
- ▶ For homework assignments, always submit *both* your .R file showing your work, and a compiled .pdf file on Canvas.

Homework

- ▶ Solution sets will be posted *at the same time* as problem sets.
- ▶ Do as much as you can on the problem set before checking the solutions.
- ▶ Check your work, and then fill out a form on how you did, what you understood and didn't.
- ▶ You get marked both on completion of the problem set, **AND** filling out the form.
- ▶ (If you find errors in the solution set, post them on the class StackOverflow and you will get extra credit)
- ▶ For homework assignments, always submit *both* your .R file showing your work, and a compiled .pdf file on Canvas.

Homework

- ▶ Solution sets will be posted *at the same time* as problem sets.
- ▶ Do as much as you can on the problem set before checking the solutions.
- ▶ Check your work, and then fill out a form on how you did, what you understood and didn't.
- ▶ You get marked both on completion of the problem set, **AND** filling out the form.
- ▶ (If you find errors in the solution set, post them on the class StackOverflow and you will get extra credit)
- ▶ For homework assignments, always submit *both* your .R file showing your work, and a compiled .pdf file on Canvas.

Homework

- ▶ Solution sets will be posted *at the same time* as problem sets.
- ▶ Do as much as you can on the problem set before checking the solutions.
- ▶ Check your work, and then fill out a form on how you did, what you understood and didn't.
- ▶ You get marked both on completion of the problem set, **AND** filling out the form.
- ▶ (If you find errors in the solution set, post them on the class StackOverflow and you will get extra credit)
- ▶ For homework assignments, always submit *both* your .R file showing your work, and a compiled .pdf file on Canvas.

Homework

- ▶ Solution sets will be posted *at the same time* as problem sets.
- ▶ Do as much as you can on the problem set before checking the solutions.
- ▶ Check your work, and then fill out a form on how you did, what you understood and didn't.
- ▶ You get marked both on completion of the problem set, **AND** filling out the form.
- ▶ (If you find errors in the solution set, post them on the class StackOverflow and you will get extra credit)
- ▶ For homework assignments, always submit *both* your .R file showing your work, and and a compiled .pdf file on Canvas.

Homework grading

check(+/-)

- ▶ Check: You fully completed the assignment, and submitted all components. (A)
- ▶ Check plus: You went above and beyond, your solutions were clear and detailed. (A+)
- ▶ Check minus: You made an attempt, but it wasn't complete. Maybe you didn't submit all components, or didn't fully answer some of the questions. (B or C)
- ▶ Unmarked: You did not submit enough of an assignment for credit.

Homework grading

check(+/-)

- ▶ Check: You fully completed the assignment, and submitted all components. (A)
- ▶ Check plus: You went above and beyond, your solutions were clear and detailed. (A+)
- ▶ Check minus: You made an attempt, but it wasn't complete. Maybe you didn't submit all components, or didn't fully answer some of the questions. (B or C)
- ▶ Unmarked: You did not submit enough of an assignment for credit.

Homework grading

check(+/-)

- ▶ Check: You fully completed the assignment, and submitted all components. (A)
- ▶ Check plus: You went above and beyond, your solutions were clear and detailed. (A+)
- ▶ Check minus: You made an attempt, but it wasn't complete. Maybe you didn't submit all components, or didn't fully answer some of the questions. (B or C)
- ▶ Unmarked: You did not submit enough of an assignment for credit.

Homework grading

check(+/-)

- ▶ Check: You fully completed the assignment, and submitted all components. (A)
- ▶ Check plus: You went above and beyond, your solutions were clear and detailed. (A+)
- ▶ Check minus: You made an attempt, but it wasn't complete. Maybe you didn't submit all components, or didn't fully answer some of the questions. (B or C)
- ▶ Unmarked: You did not submit enough of an assignment for credit.

Homework grading

check(+/-)

- ▶ Check: You fully completed the assignment, and submitted all components. (A)
- ▶ Check plus: You went above and beyond, your solutions were clear and detailed. (A+)
- ▶ Check minus: You made an attempt, but it wasn't complete. Maybe you didn't submit all components, or didn't fully answer some of the questions. (B or C)
- ▶ Unmarked: You did not submit enough of an assignment for credit.

Assignment for Friday

- ▶ Download software on your own computer.
- ▶ Set up your working directory.
- ▶ Compile and submit a report.
- ▶ (Fill out the feedback form as well, even though there aren't solutions.)

Assignment for Friday

- ▶ Download software on your own computer.
- ▶ Set up your working directory.
- ▶ Compile and submit a report.
- ▶ (Fill out the feedback form as well, even though there aren't solutions.)

Assignment for Friday

- ▶ Download software on your own computer.
- ▶ Set up your working directory.
- ▶ Compile and submit a report.
- ▶ (Fill out the feedback form as well, even though there aren't solutions.)

Assignment for Friday

- ▶ Download software on your own computer.
- ▶ Set up your working directory.
- ▶ Compile and submit a report.
- ▶ (Fill out the feedback form as well, even though there aren't solutions.)

Final Project

- ▶ More in-depth, open-ended problem set.
- ▶ Make sure you understand the homework solutions, the final will not have a solution set.
- ▶ You pick from approved data sets—or propose your own.
- ▶ Run through a series of coding exercises and statistical analyses.

Final Project

- ▶ More in-depth, open-ended problem set.
- ▶ Make sure you understand the homework solutions, the final will not have a solution set.
- ▶ You pick from approved data sets—or propose your own.
- ▶ Run through a series of coding exercises and statistical analyses.

Final Project

- ▶ More in-depth, open-ended problem set.
- ▶ Make sure you understand the homework solutions, the final will not have a solution set.
- ▶ You pick from approved data sets—or propose your own.
- ▶ Run through a series of coding exercises and statistical analyses.

Final Project

- ▶ More in-depth, open-ended problem set.
- ▶ Make sure you understand the homework solutions, the final will not have a solution set.
- ▶ You pick from approved data sets—or propose your own.
- ▶ Run through a series of coding exercises and statistical analyses.

Final Project

- ▶ More in-depth, open-ended problem set.
- ▶ Make sure you understand the homework solutions, the final will not have a solution set.
- ▶ You pick from approved data sets—or propose your own.
- ▶ Run through a series of coding exercises and statistical analyses.

Github

- ▶ <https://github.com/UChicago-pol-methods/SOSC13200-W23>
- ▶ Serves as a repository for class documents.
- ▶ You don't need a GitHub account, or need to know how to use git.
- ▶ But if you *do* find mistakes in slides or homeworks, you can get extra extra credit for submitting a pull request on GitHub with revisions.

Github

- ▶ <https://github.com/UChicago-pol-methods/SOSC13200-W23>
- ▶ Serves as a repository for class documents.
- ▶ You don't need a GitHub account, or need to know how to use git.
- ▶ But if you *do* find mistakes in slides or homeworks, you can get extra extra credit for submitting a pull request on GitHub with revisions.

Github

- ▶ <https://github.com/UChicago-pol-methods/SOSC13200-W23>
- ▶ Serves as a repository for class documents.
- ▶ You don't need a GitHub account, or need to know how to use git.
- ▶ But if you *do* find mistakes in slides or homeworks, you can get extra extra credit for submitting a pull request on GitHub with revisions.

Github

- ▶ <https://github.com/UChicago-pol-methods/SOSC13200-W23>
- ▶ Serves as a repository for class documents.
- ▶ You don't need a GitHub account, or need to know how to use git.
- ▶ But if you *do* find mistakes in slides or homeworks, you can get extra extra credit for submitting a pull request on GitHub with revisions.

Other questions/concerns about class policies?

Getting started in R

[R script]