What’s This Course About?

“Aimed at non-CS undergraduate and graduate students who want to learn the basics of big data tools and techniques and apply that knowledge in their areas of study. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets. This course provides a broad and practical introduction to big data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.”
“Aimed at non-CS undergraduate and graduate students who want to learn the basics of big data tools and techniques and apply that knowledge in their areas of study. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets. This course provides a broad and practical introduction to big data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.”
“Aimed at non-CS undergraduate and graduate students who want to learn the basics of big data tools and techniques and apply that knowledge in their areas of study. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets. This course provides a broad and practical introduction to big data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. **Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application.** Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.”
“Aimed at non-CS undergraduate and graduate students who want to learn the basics of big data tools and techniques and apply that knowledge in their areas of study. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets. This course provides a broad and practical introduction to big data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.”
Who Should Take It?

“Aimed at non-CS undergraduate and graduate students who want to learn the basics of big data tools and techniques and apply that knowledge in their areas of study. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets. This course provides a broad and practical introduction to big data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.”
Who Should Take It?

“Aimed at non-CS undergraduate and graduate students who want to learn the basics of big data tools and techniques and apply that knowledge in their areas of study. Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets. This course provides a broad and practical introduction to big data: data analysis techniques including databases, data mining, machine learning, and data visualization; data analysis tools including spreadsheets, Tableau, relational databases and SQL, Python, and R; introduction to network analysis and unstructured data. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application. Prerequisites: comfort with basic logic and mathematical concepts, along with high school AP computer science, CS106A, or other equivalent programming experience.”
Who Shouldn’t Take It?

Computer Science or MCS students
(except by petition)

If you’re in the wrong place, it’s okay to leave now 😞
Course Staff

Instructor
Jennifer Widom

Course Assistants
Steven Chen (bootcamp)
Alex Haigh (head)
Vera Lin
Leo Mehr
John Solitario
Who’s Taking It - Winter 2019

Undergraduates, Masters, MBA, PhD

American Studies
Art Practice
Bioengineering
Biology
Business Administration
Civil & Environmental Engg
Classics
Comparative Literature
Computer Science
Developmental Biology
East Asian Studies
Economics
Electrical Engineering
English
Environmental Systems Engg
Epidemiology & Clinical Research
French

Human Biology
Iberian & Latin American Cultures
International Policy Studies
International Relations
Law
Management Science & Engg
Materials Science & Engineering
Math & Computational Science
Mechanical Engineering
Philosophy
Physics
Psychology
Product Design
Science, Technology, & Society
Slavic Languages & Literature
Sociology
Undeclared
Who’s Taking It

Programs listed in the bar chart include:
- Undeclared
- Science, Tech, & Soc
- Civil & Environmental Engr
- Mgmt Sci & Engineering
- Business Administration
- Sociology
- Engineering
- Economics
- International Policy
- International Relat
- Envir Systems Engr
- East Asian Studies
- English
- Computer Science
- Mechanical Engineer
- Physics
- Bioengineering
- Developmental Biology
- Classics
- Electrical Engineering
- Comparative Lit
- Human Biology
- Epidemiology
- French
- ILAC
- Law LLM
- Material Sci & Eng
- Math & Comp Science
- Philosophy
- Political Science
- Psychology
- Slavic Lang & Lit
- American Studies

The chart shows the number of students taking the course by program, categorized into Graduate, Professional, and Undergraduate.
Who’s Taking It
Who’s Taking It
Who’s Taking It
# Assigned Work

<table>
<thead>
<tr>
<th>Assignment/Project</th>
<th>Assigned</th>
<th>Due</th>
</tr>
</thead>
</table>
| Assignment #1  
Spreadsheets for Data Analysis and Visualization     | Jan 14   | Jan 21  |
| Project #1  
Personal Data Analysis                           | Jan 14   | Jan 28  
Feb 18        |
| Assignment #2  
Data Visualization Using Tableau, SQL                | Jan 21   | Jan 31  |
| Assignment #3  
Python for Data Analysis and Visualization          | Jan 31   | Feb 11  |
| Assignment #4  
Machine Learning, R Language                         | Feb 18   | Feb 28  |
| Project #2  
Movie-Rating Predictions                           | Feb 18   | March 8 |
| Assignment #5  
Data Mining, Network Analysis, Unstructured Data      | Feb 28   | March 14|
## Exams

<table>
<thead>
<tr>
<th>Exam</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm exam</td>
<td>February 14</td>
</tr>
<tr>
<td>In class</td>
<td>Love symbol</td>
</tr>
<tr>
<td>Final exam</td>
<td>March 18</td>
</tr>
<tr>
<td>12:15-3:15 PM (2 hours)</td>
<td></td>
</tr>
</tbody>
</table>
Logistics

- Units - 4 for undergraduates, 3-4 for graduates
- WAYS requirement - Applied Quantitative Reasoning (WAY-AQR)
- Textbook? No  Readings? Recommended
- Class attendance - Expected
  - Hand-on activities
  - Only cursory notes
  - All class material game for exams
Logistics

- Grade weighting - 1/3 each assignments, projects, exams
- Graded on a curve? Not really
- Late policy - 10%/30% for 24/48 hours late, five free late days
Office Hours

TA office hours
• 20 hours per week
• Times and locations can vary

Prof. Widom office hours
• Wednesdays 4:00-5:00 PM (usually)
• Huang building 2nd floor Dean’s Office #227

Always check the course calendar!
Online

Website - http://cs102.stanford.edu

Piazza
  • Announcements
  • Q&A (private and public)
  • Discussion

Gradescope - Assignment submission & grading
For Thursday’s Class

1) Get set up on Google Drive if you’re not already

2) Download Europe city temperatures data from course website (three files)

3) Copy data files into Google Drive, make sure you can open with Google Sheets

4) Bring laptop to class (or share)
CS 102: Big Data
Tools and Techniques

Questions?