Teaching Assistants

Section BA: Yuan “Monica” Gao Rollinson
Email: monicagr@uw.edu
Zoom link:
Lab times (Zoom): Friday 2:30-3:20pm
Office hours (Zoom): Tuesday 1pm-2pm & Friday 11:30am-12:30pm, or by appointment

Section BB: İhsan Kahveci
Email: ikahveci@uw.edu
Zoom link:
Lab times (Zoom): Friday 1:30-2:20pm
Office hours (Zoom): Monday 10am-11am & Wednesday 10am-11am, or by appointment

Objectives

- Supplement the lecture and reading portions of class
- Understand the basics of a data science workflow/pipeline
- Execute exploratory data analysis with real-world datasets
- Practice data importing, manipulation, and visualization
- Learn how to use data to answer social science questions
- Reflect on the ethics of the work, often by getting our hands dirty
- Gain familiarity with R Studio and RMarkdown

Access to RStudio

- The Server for RStudio can be found here: http://128.95.72.79:8787
- Every student will have an account and password generated for you. You should not change the password and this will allow your TAs to log in and help if necessary.
Assessments

Lab assignments (7)  30%
Lab homework (3)  30%
Lab final project  40%

Reminder that these lab-specific assessments represent 40% of your final grade while the assessments outlined in the Lecture Syllabus represent the other 60% of your final grade.

Lab Assignments

We will be working through some in-class lab assignments during lab sections. Each assignment will be posted on Canvas on Thursday and due by the following Friday noon. You will submit your work on Canvas to receive participation credit. You will be graded on your completion (not accuracy) of these assignments. There are eight assignments in total. Bonus questions will be given to some assignments.

<table>
<thead>
<tr>
<th>Date (Friday)</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2 (April 10)</td>
<td>Reading data</td>
</tr>
<tr>
<td>Week 3 (April 17)</td>
<td>Data transformation</td>
</tr>
<tr>
<td>Week 4 (April 24)</td>
<td>Data visualization and mapping</td>
</tr>
<tr>
<td>Week 5 (May 1)</td>
<td>Categorical and text data</td>
</tr>
<tr>
<td>Week 6 (May 8)</td>
<td>Exploratory data analysis</td>
</tr>
<tr>
<td>Week 7 (May 15)</td>
<td>Collecting data using the Internet</td>
</tr>
<tr>
<td>Week 8 (May 22)</td>
<td>Political ads &amp; merging data</td>
</tr>
<tr>
<td>Week 9 (May 29)</td>
<td>Office help time for final project</td>
</tr>
<tr>
<td>Week 10 (June 5)</td>
<td>Office help time for final project</td>
</tr>
</tbody>
</table>
**Lab Homework**

There are three lab homeworks. They will be due on Canvas. The due dates will depend on how we progress through the lab materials and will be announced on Canvas. To encourage you to learn, homeworks will be graded on both completion and accuracy. For example, you would receive half of the full points for every question you complete, and you will receive full points for every question you answer correctly.

**Lab Final Project**

The project is meant to show your ability to use RStudio to explore a data set and produce relevant output and visualizations. Think about this project as turning data into information. You will work with groups of five people. The format will be a 5-min length video presentation. You can do it synchronously or asynchronously, just make sure each participant takes an equal role in the presentation.

In order to receive credit for the final project, you will need to submit four items on Canvas:

1. Link of your group video presentation (YouTube link);
2. 5-page presentation slides (in pdf format);
3. R code used to create the presentation (in html or RMarkdown format);
4. Lastly, you need to individually comment on other videos via Canvas.

**Final Project Rubric:**

<table>
<thead>
<tr>
<th>Research Question</th>
<th>15 pts</th>
<th>10 pts</th>
<th>5 pts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present a research question that:</td>
<td>Fulfills all expectations</td>
<td>Attempts all requirements with / some errors execution</td>
<td>Missing at least one requirement superficial</td>
</tr>
<tr>
<td>· Applies data to a social issue (it doesn’t have to be a novel problem!)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· Has reasonable scope for a lab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· Identifies focal independent and dependent variables</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>15 pts</th>
<th>10 pts</th>
<th>5 pts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect data from an appropriate source (fits your question, reliable)</td>
<td>Fulfills all expectations</td>
<td>Attempts all requirements with some errors</td>
<td>Data source unreliable or summary statistics inaccurate</td>
</tr>
<tr>
<td>Provide summary statistics for relevant variables</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Preparation

Provide code that demonstrates at least four (4) of:
- Cleaning NAs
- Cleaning other missing values
- Cleaning numeric data
- Recoding a variable
- Creating a numeric variable
- Creating a dummy variable
- Creating a character variable
- Creating a categorical variable

BONUS: Merging two sources of data
BONUS: Cleaning text data

Visual Analysis

Produce two visualizations one that shows variation expectations and one that shows covariation that each:
- Show the relationship between key variables (your focal dependent and independent variables must be present in at least one)
- Have clear labels
- Have clear and appropriate scales

BONUS: High Data/Ink ratio
BONUS: Show uncertainty

Discussion

Provide a preliminary answer to your research question that offers:
- A discussion of ethical issues (privacy, anonymity, etc.)
- A clear connection to your data
- A clear explanation of your visualizations
- A discussion of confidence
- A discussion of limitations
Lab History

This lab is based on SOC 225 lab developed by Connor Gliroy and Ian Kennedy, Department of Sociology, University of Washington.