Welcome to...

DATA 311!

Instructor: Scott Wehrwein (he/him/his)

call me this!

Plan for today:

**Logistics**
- Course webpage/syllabus
- Assessment
- Policies
- Textbook
- Inclusive learning environment
- Flexibility and feedback

**Introduction**
- What is data science?
- What is data?
- Course overview
Quarters Taught

<table>
<thead>
<tr>
<th></th>
<th>F2F</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

DATA 311 Taught

<table>
<thead>
<tr>
<th></th>
<th>F2F</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Family Size

Stats
- Organize, find patterns
- Applying info from data
- Prediction
- Analyze results

Solve problems
- Ethics of data collection + usage
- Presenting results (to non-DS stats)
- Biases in data
- Lurking variables
What is data science?

The process and/or study of extracting insights from data.

Premise: “data ≠ information ≠ understanding ≠ wisdom.”
- Clifford Stoll

The tools:

- CS
- Stats
- Domain Expertise
- Questions
- Interpretation
- Impact

The Practitioners:

- CS
- Stats
- Business

- Solve problems
- Find patterns
- Predict things

- Optimize
- Get data
- Filter, sample, clean
- Should we?
What do data scientists do?

**Answer questions; test hypotheses**

- How much do daily temperatures vary between Dublin and Seattle?
- Are movie budgets correlated with critic ratings?

**Explore datasets to find trends, patterns, and insights**

- What impacts did COVID have on _____?

**Make predictions**

- Given academic transcripts, can we identify students at risk of dropping out ahead of time so we can provide support when they need it?