CS 594: Empirical Analysis

Deriving Sound Insights from Data

Purpose of this course

Learn to work with data for the purposes of performing and communicating empirical research in Computer Science

Who am I?

Assistant Professor Chris Kanich Main research interest: Socio-technical aspects of security

- Economic issues: how much do attackers profit?
- User-facing issues: what are the most damaging attacks?
- How can we design secure software that meets users needs, but minimizes risk?

Who are you?

Please let us know:

- Your level of study (PhD? MS?)
- Your research interests

The path to a successful project

- Data Aquisition
- Data Analysis
- Data Evaluation
- Data Communication

Data Aquisition

- Using techniques like web scraping
- Other network level measurements
- Detecting and preventing biases in collection

Data Analysis

- Data quality - especially with borrowed datasets!
- Initial data exploration

Data Evaluation and Communication

- Basic shape of the dataset
    - How well does our data fit that distribution?
    - Why would we expect it to fit this distribution?
- Hypothesis testing
    - In exploratory/measurement studies, what more complex, insightful questions can we ask of our data?
    - In system papers (data mining, applied machine learning, computer systems), how can we convince the reader that our system performs as well as we believe it does?

For Wednesday

- Read "Strategies for Sound Internet Measurement" by Vern Paxson
- Get started on the homework assignment (posted via course website)