PSTAT 234: Statistical Data Science

Sang-Yun Oh

What the course is about

- **Practical case studies**: Movie recommender system, portfolio allocation, basketball shot pattern analysis, data scraping, etc.

- **Real datasets**: [Movielense](), stock prices, NBA, EPA air quality index, 2016 election results

- **Quantitative methods**: Latent factor analysis, non-negative matrix factorization, numerical optimization, linear algebra

- **Modern computing tools**: python, Jupyter notebook, virtual machine, GitHub

- High-level use of data science tools and concepts

Class Demographic

PSTAT majors	Non-PSTAT
18	27

- PSTAT majors: ACTSC, STATS, STATSAP
- Other majors: CHEM, COMM, ECON, EDUC, ESM, GEOG, MARSC, MATRL, ME, PBS, POLS

- Zoom breakouts!

Class Information

- Syllabus
- GauchoSpace: central portal
- Slack: instant messaging
- Google Calendar: zoom links, lab hours
- GitHub repository: course files (lecture notebooks, assignment templates, etc.)
- Gradescope: submit Jupyter notebook assignments
- DataCamp: computing tool exercises

First Week

- First lecture video: GauchoCast video
- First assignment: Introduction to Shell (DataCamp)
- First lab: get Jupyter lab running on Virtualbox

Final Projects

- Interdisciplinary team projects

- Groups of 4

- e-poster session

- Recorded videos