

# Data Science 101

Mark Fredrickson

# DS101 Basics

- “Introduction to Data Science”
- Aimed at wide variety of students: majors, minors, distribution requirement (e.g. econ, psych, CS)
- Typical enrollment: 240 - 300/semester, Fall and Winter
- Alternative to STATS 250, our “classical” introductory course
- Topics:
  - Python programming
  - Descriptive statistics
  - Inferential statistics
  - Regression modeling
  - Machine learning
  - Research design

# Python

- Jupyter notebooks
  - Weekly labs
  - Weekly homework
  - Final project
- Online implementations
  - Great Lakes
  - Google Colab
  - GitHub Codespaces
  - Back to Google Colab
- Packages
  - Pandas (and numpy)
  - Seaborn
  - Statsmodels, scipy, scikit-learn

# Other technical stuff

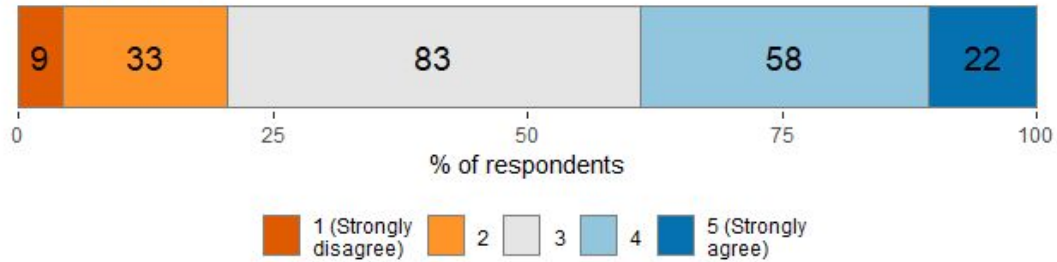
- Grading in Gradescope
  - Convenient upload of .ipynb, though frequent file corruption issues
  - Autograding possible, but not used
- Lecture slides use Pythontex
- Labs and HW use Quarto
  - Plain text file format (easy to manage in git)
  - Render to .ipynb (some manual edits to break up cells)
  - Moving over lectures, render to PDF

# Future Directions

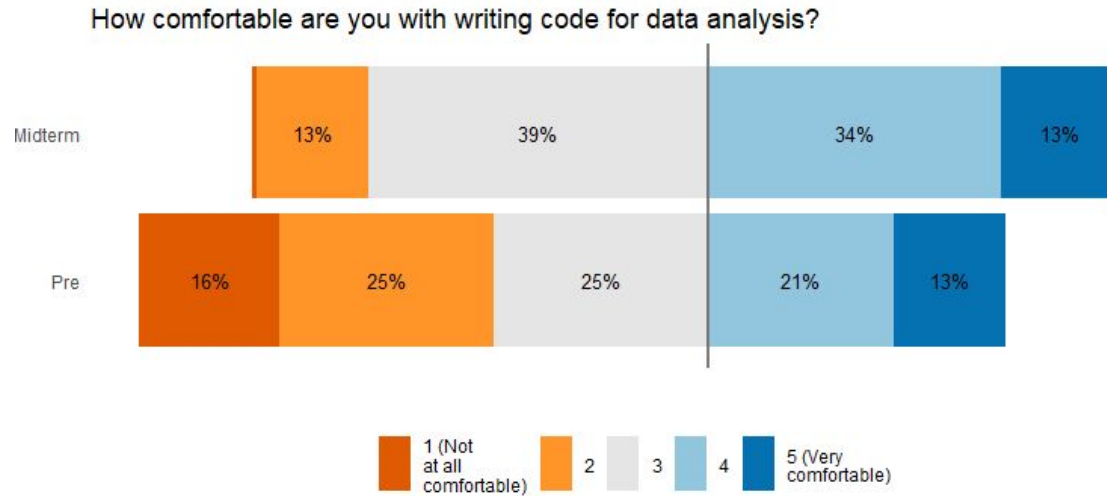
- Use of AI
  - Copilot is best, integrates with Visual Studio and Codespaces
  - Gemini is.... ok
  - Drawing some material from [“Learn AI Assisted Programming with Python”](#)
- Foundational Course Initiative
  - Started summer 2024, three year grant
  - First year goals were/are improving student and GSI support, gather student feedback
  - Second year: flip the classroom
- More “cool factor”
  - Image processing
  - Text/language models
  - More data sets, community linked project

# This course was fun!

This course was fun!

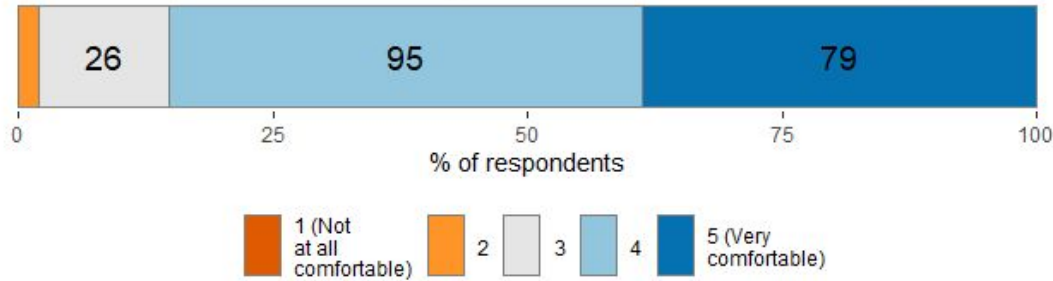


# How comfortable are you with writing code for data analysis?



# How comfortable are you with writing code WITH GenAI for data analysis?

How comfortable are you with writing code WITH GenAI for data analysis?





# How comfortable are you with writing code WITHOUT GenAI for data analysis?

How comfortable are you with writing code WITHOUT GenAI for data analysis?

