

Lecture 1

Syllabus and Course Outline

Byeong-Hak Choe

bchoe@geneseo.edu

SUNY Geneseo

January 21, 2026

Instructor

Current Appointment & Education

- Name: Byeong-Hak Choe.
- Assistant Professor of Data Analytics and Economics, School of Business at SUNY Geneseo.
- Ph.D. in Economics from University of Wyoming.
- M.S. in Economics from Arizona State University.
- M.A. in Economics from SUNY Stony Brook.
- B.A. in Economics & B.S. in Applied Mathematics from Hanyang University at Ansan, South Korea.
 - Minor in Business Administration.
 - Concentration in Finance.

Research 1: Social Media vs. Fossil Fuel Lobbying on Climate Legislation

- Choe, B.-H., 2021. *“Social Media Campaigns, Lobbying and Legislation: Evidence from #climatechange and Energy Lobbies.”*
- **Question:** To what extent do social media campaigns compete with fossil fuel lobbying on climate change legislation?
- Data include:
 - 5.0 million tweets with #climatechange/#globalwarming around the globe;
 - 12.0 million retweets/likes to those tweets;
 - 0.8 million Twitter users who wrote those tweets;
 - 1.4 million Twitter users who retweeted or liked those tweets;
 - 0.3 million US Twitter users with their location at a city level;
 - Firm-level lobbying data (expenses, targeted bills, etc.).

Research 2: Measuring the Value of Statistical Life Using Big Data

- Choe, B.-H. and Newbold, Steve, *“Estimating the Value of Statistical Life (VSL) through Big Data”*
- VSL is the monetary value associated with reducing the risk of death.
 - How much value would that be? How can we measure it?
 - How do government agencies use the VSL to decide which policies are worth the cost when they reduce the risk of death?

Research 3: Governance Barriers to Effective Climate Finance

- Choe, B.-H. and Ore-Monago, T., 2024. “***Governance and Climate Finance in the Developing World***”
- **Climate finance** refers to the financial resources allocated for mitigating and adapting to climate change, including support for initiatives that reduce greenhouse gas emissions and enhance resilience to climate impacts.
 - We focus on *transnational financing* that rich countries provide poor countries with financial resources, in order to help them adapt to climate change and mitigate greenhouse gas (GHG) emissions.
 - Since the GHG emissions in developing countries are rapidly growing, it is crucial to assess the effectiveness of climate finance.
 - Poor governance (e.g., legal system, rule of law, and accountability) can be significant barriers to emissions reductions.

Research 4: Climate Finance with Renegotiations

- Choe, B.-H., “*Climate Finance under Limited Commitment and Renegotiations: A Dynamic Contract Approach*” (2026)
 - Using climate funds (e.g., the **Green Climate Fund**) as the main mechanism for financing developing countries, this paper studies a long-term funding relationship between a rich and a poor country.
 - The parties disagree over (1) the level of funding and (2) how to allocate it between adaptation and mitigation. Commitment is limited, and contracts may be renegotiated each period with some probability.
 - **Main results:** **Less frequent renegotiation** improves contract efficiency and reduces inequality between the countries.
 - These findings suggest design principles for climate funds: strengthen commitment and limit renegotiation, especially as climate risks intensify.

Let's Introduce Ourselves

- Now that you've heard a bit about me and my research, I'd like to learn about **you**.
- Please briefly introduce yourself:
 - Your **name** and **year** (first-year, sophomore, junior, senior)
 - Your **major/minor** (or what you are considering)
 - Why you are taking **DANL 210**
 - One topic you are interested in exploring with data this semester (sports, finance, social issues, business, etc.)

Syllabus

Course & Meeting Info

- **Course:** DANL 210-01 — *Data Preparation and Management* (**3 credits**)
- **Semester:** Spring 2026
- **Instructor:** Byeong-Hak Choe
- **Email (preferred):** bchoe@geneseo.edu
- **Class time:** MWF 10:30 A.M. – 11:20 A.M.
- **Class location:** Newton 205
- **Office:** South 227B
- **Office Hours:** MWF 9:15 A.M. – 10:15 A.M. (or by appointment)

Course Website & Brightspace

Course Website (GitHub Pages)

(<https://bcdanl.github.io/210>)

- Lecture slides
- Classwork
- Homework
- Project resources

Brightspace

- Announcements
- Grades
- Assignment submissions

I will post a **Brightspace announcement** whenever new course materials (e.g., homework) are added to the course website.

What This Course Is About

- This course provides hands-on practice in **collecting, cleaning, and transforming real-world datasets** for analysis and decision-making.
- We will use **Jupyter Notebooks (.ipynb)** in **Google Colab** for Python fundamentals and data transformation, and **Python scripts (.py)** in **Spyder** for data collection workflows.
- We will focus on the core Python data workflow (e.g., **pandas**, **selenium**, **requests**, and related tools).

Key skills you'll build:

- Collect data from real-world sources
- Work with common Python data objects (and understand their structure and attributes)
- Perform essential data-wrangling tasks (filtering, grouping, pivoting, joining)
- Clean messy data (missing values, duplicates, inconsistent types)
- Create clear, effective visualizations to communicate results
- Explain how Python tools are used strategically to solve data analytics problems

Prerequisites

- **Statistics:** (ECON 205 or GEOG 278 or MATH 242 or MATH 262 or PLSC 251 or PSYC 250 or SOCL 211)
- **Introduction to Programming:** DANL 201

Communication Guidelines

- I check email **daily (Mon–Fri)**
- Expect responses within **12–72 hours**

Course Contents & Schedule

- **Weeks 1–3:** Python Basics
 - HW 1: Feb 4 (before class time)
- **Weeks 4–8:** Data Collection (Web scraping with `selenium` + application programming interfaces (APIs))
 - HW 2: Feb 25 (before class time)
 - HW 3: Mar 9 (before class time)
 - **Midterm Exam: March 13** (during class time)
- **Week 9:** Spring Break (March 14–21)

Course Contents & Schedule

- **Weeks 10–16:** Data Transformation + Visualization ([pandas](#) + plotting)
 - HW 4: Apr 13 (before class time)
 - **GREAT Day: April 22** (no class)
 - HW 5: Apr 22 (before class time)
 - HW 6: May 4 (before class time)
- **Week 17:** Final Exam + Data Storytelling Report
 - **Final Exam: May 14, 8:30 A.M. – 10:30 A.M.**
 - **Data Storytelling Report: May 14, 11:59 P.M.**

Grading

Grade components

- Attendance — 5%
- Participation — 5%
- Group Project — 20%
- Homework (total) — 20%
- Exams (total) — 50%

Important grading details

- **Lowest homework score is dropped**
- **Total exam grade** is the **maximum** of:
 - Simple average of midterm and final, **or**
 - Weighted average (**33% midterm / 67% final**)
- I will post detailed grading feedback for each assignment on Brightspace once grading is completed.
 - The final grade displayed in Brightspace will NOT reflect your official course grade because we will follow the grading scheme described above.

Grading Data Storytelling Report

- Data collection (30%)
- Descriptive statistics (5%)
- Data transformation (30%)
- Data visualization (10%)
- Data storytelling (15%)
- Code (10%)

Course Policies

- **Late work:** Homework is accepted up to **3 days late** with a **30% penalty**.
- Make-up exams are allowed **only with documented excuses** or university-excused absences.
- **Attendance** is taken with a **sign-in sheet**
 - You must email me **before** missing class for illness or other valid reasons
 - You have **7 unexcused absences** available in the semester
 - ▶ Beyond that, your total percentage grade may be reduced

Academic integrity & AI use

- Cheating and plagiarism violate college policy and may result in serious consequences.
- You **may use AI tools** (e.g., ChatGPT, Gemini) to support your learning, **but** you must:
 - **Disclose** AI use when it meaningfully contributes to your work
 - **Cite/attribute** AI-generated material appropriately
 - Using AI **without disclosure/citation** may be treated as academic dishonesty
- There is **no penalty** for using AI on Homework and the Data Storytelling Project.
 - However, you are still responsible for understanding, verifying, and being able to explain any AI-assisted work.

Accessibility

SUNY Geneseo is committed to equitable access. If you have approved accommodations, please coordinate through OAS and communicate with me early.

Course Tools

Useful Keyboard Shortcuts for Lecture Slides

- **CTRL + Shift + F**: Search
- **M**: Toggle menu
- **E**: PDF export/printing mode

Let's Practice Markdown!

- **Markdown is a lightweight writing format** that lets you create well-structured documents using plain text (headings, lists, links, quotes, images, and code blocks).
- **Why Markdown is widely used on the internet:**
 - **Works everywhere:** supported by many platforms, including **Jupyter Notebook, GitHub, Reddit, Slack/Discord/Obsidian**, and documentation websites.
 - **Great for publishing and future-proof:** Markdown can be converted into professional webpages/PDFs, and because it is plain text, and it is easy to store and reuse.
- **In this course, Markdown is essential** because you will write narrative explanations in Jupyter Notebook (*.ipynb) when creating your **Data Storytelling Report**.
- Start here: [Classwork 1](#)

What is Python?

- **Python** is a versatile programming language known for its simplicity and readability.
- Python has become a dominant tool in various fields including data analysis, machine learning, and web development.
 - It is widely used among developers, data scientists, and researchers for building applications and performing data-driven tasks.
 - Python is open source and has a vast ecosystem of libraries and frameworks.

What is Jupyter?

- **Jupyter** is an open-source integrated development environment (IDE) primarily for **Python**, though it supports many other languages.
 - An IDE is a software application that provides comprehensive facilities (e.g., text code editor, graphical user interface (GUI)) to users for a programming-related project.
 - Jupyter provides a notebook interface that allows users to write and execute code in a more interactive and visual format.


What is Jupyter Notebook?

- Jupyter Notebook (*.ipynb) is a user-friendly environment that enhances coding, data analysis, and visualization.
 - It offers a web-based interface that combines live code, equations, visualizations, and narrative text.
 - Jupyter Notebook is widely used for data science, machine learning, and research, enabling easy sharing and collaboration.
- For Python basics and data transformation topics, we will use **Google Colab**, Google's free cloud-based version of Jupyter Notebook.

Getting Started with Google Colab

- Step 1: Accessing **Google Colab**
 - Open your web browser and go to Google Colab.
 - Sign in with your Google account if you haven't already.
- Step 2: Creating a New Notebook
 - Once on the Colab homepage, click on the **+ New notebook** button.

Turn off AI Assistance in Google Colab

- On Google Colab:
 1. From the top-right corner, click 
 2. Click “**AI Assistance**” from the side menu.
 3. Disable the first two and enable the last:
 - ☐ **Show AI-powered inline completions**
 - ☐ **Consented to use generative AI features**
 - ☒ **Hide generative AI features**