

- DANL 310: Data Visualization and Presentation -

# 1. Introduction

Byeong-Hak Choe

SUNY Geneseo

January 27, 2022

# Instructor

## Current Appointment & Education

- Name: Byeong-Hak Choe.
- Lecturer in Data Analytics at 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.

# Instructor

## Data Science and Climate Change

- Choe, B.H., 2021. “Social Media Campaigns, Lobbying and Legislation: Evidence from #climatechange/#globalwarming 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.).

# Syllabus

## Email, Class & Office Hours

- Email: [bchoe@geneseo.edu](mailto:bchoe@geneseo.edu)
- Classroom: Newton Lecture Hall, Room 205.
- Class Hours:  
Tuesdays and Thursdays 01:00 PM – 02:15 PM.
- Class Homepage:  
<https://canvas.geneseo.edu/courses/23754>
- Office: South Hall 117B.
- Office Hours:  
Tuesdays and Wednesdays 03:45 PM – 05:45 PM.

# Syllabus

## Course Prerequisites

### (1) Business & Economics Statistics or Equivalent:

- Economics 205, Geography 278, Mathematics 242, Mathematics 262, Political Science 251, Psychology 250, or Sociology 211.

### (2) Programming for Data Analytics:

- Data Analytics 100.

# Syllabus

## Required Materials

- **“Data Visualization: A Practical Introduction”** by Kieran Healy
  - A free online version of this book is available at <https://socviz.co/> .
  - ISBN-13: 978-0691181622; ISBN-10: 0691181624
- **“Tableau eLearning”** and *“LinkedIn Learning”*
  - I will briefly summarize contents in basic Tableau courses from Tableau eLearning (<https://elearning.tableau.com>) and LinkedIn Learning (<https://www.linkedin.com/learning>).
  - I will require you to obtain certificates for some Tableau courses from the LinkedIn Learning.

# Syllabus

## Course Description

- This course covers tools and methodologies that visually represent data using well-presented and visually appealing graphics in order to understand data better and perform useful data analytics tasks.
- Topics covered in this course include
  - (1) visualizing many forms of graphs such as line graphs, scatter plots, bar charts, and more;
  - (2) tidying data from various sources for data visualization;
  - (3) customizing graphics using various formats and styles including colors, fonts, lines, and more;
  - (4) visualizing geographical data such as maps;
  - (5) showing an overview and telling a story using dashboard.
- During the course, you will work hands-on with R, R Markdown, Git, GitHub, and/or Tableau.

# Syllabus

## Course Requirements

- **Laptop:** You should bring your own laptop (Mac or Windows) to the classroom. It is recommended to have 2+ core CPU, 4+ GB RAM, and 500+ GB disk storage in your laptop for this course.
- **Homework:** There will be six homework assignments.
- **Exam:** There will be one midterm exam.
- **Team project:** There will be one team project.



# Syllabus

## Website

- You will create their own website using R Markdown and GitHub.
- You will publish R codes, dashboards and your team project on their website.

# Syllabus

## Team Project

- Each team consists of two students.
- Each team must select data for the team project.
- The data for the team project must be approved by Byeong-Hak Choe.
- The write-up for the team project must include machine learning methods (e.g., linear regression, clustering) and several visualizations.
- The document for the team project must be published in each member's website.
- The document for the team project must be identical for both team members.

# Syllabus

## Team Project

- Team projects will be evaluated based on
  - (1) the quality of exploratory data analysis using ggplot and dashboards;
  - (2) the quality of applying machine learning methods;
  - (3) the quality of the 10-minute midterm presentation and the 18-minute final presentation;
  - (4) peer evaluation on other teams' project presentations and web-pages;
  - (5) your evaluation on your member's contribution to the team project;
  - (6) self-evaluation on your contribution to the team project.
- Changing a team member must be approved by Byeong-Hak Choe.

# Syllabus

## Team Project

- Successful team projects ...
  - pose an interesting and open-ended question/problem that can be addressed using data science for which data is available or obtainable.
  - pose a multi-faceted question/problem, containing enough dimensions to be addressed in a multitude of ways. are sufficiently deep, such that a useful data product can be made in two months that makes a solid advance on the problem.
  - can be split up into milestones, such that concrete progress can be made in two months.
  - draw on various tools and topics you have learned during your undergraduate years.

# Syllabus

## Course Schedule

There will be tentatively twenty-eight class sessions:

- (1) twenty-four class sessions for the twenty-four lectures (one class session for one lecture)
- (2) one class session for the midterm exam
- (3) one class session for the midterm presentation for the team project
- (4) two class sessions for the final presentation for the team project.

# Syllabus

## Course Contents

Class	Lecture	Contents	HW
<b>Data Visualization: A Practical Introduction</b>			
1	1	Ch 1. Look at Data	1
2–3	2–3	Ch 2. Get Started	1
4–5	4–5	Ch 3. Make a Plot	1
6	6	Ch 4. Show the Right Numbers	1
7–8	7–8	Ch 5. Graph Tables, Add Labels, Make Notes	1
9–10	9–10	Ch 6. Work with Models	2
11–12	11–12	Ch 7. Draw Maps	2
13		<b>Midterm Exam</b>	
14–15	13–14	Ch 8. Refine Your Plots	3–6
16		<b>Midterm Presentation for Team Projects</b>	
17–26	15–24	<b>Tableau, R Markdown, GitHub, and Dashboard</b>	3–6
27–28		<b>Final Presentation for Team Projects</b>	

# Syllabus

## Important Dates

- You must form the team and let Byeong-Hak Choe ([bchoe@geneseo.edu](mailto:bchoe@geneseo.edu)) know each member's name by midnight, February 24, 2022 via email.
- The midterm exam will be tentatively scheduled on March 10, 2022.
- The midterm presentation for the team projects will be tentatively scheduled on March 29, 2022. You must submit your evaluation on the other teams' presentations at the end of the class period.
- There will be no class on April 21, 2022.
- The final presentation for the team projects will be tentatively scheduled on May 10, 2022 and May 12, 2022. You must submit your evaluation on the other teams' presentations at the end of the class period.

# Syllabus

## Important Dates

- The deadline for the team project's web-page is midnight, May 12, 2022. After this deadline, you are not allowed to modify the team project's web-page.
- You must start evaluating other team projects' websites from May 13, 2022.
- You must finish evaluating other team projects' websites by 2:30 P.M. on May 19, 2022.
- You must submit your peer evaluation to Byeong-Hak Choe via email by 2:30 P.M. on May 19, 2022.
- You must also submit (1) your evaluation on your member's contribution to the team project and (2) self-evaluation on your contribution to the team project to Byeong-Hak Choe via email by 2:30 P.M. on May 19, 2022.



# Syllabus

## Grading

- Homework assignments account for 40% of the total percentage grade.
- The midterm exam and the team project account for 60% of the total percentage grade.

$$\begin{aligned} & \text{(Total Percentage Grade)} \\ &= 0.4 \times (\text{Total Homework Score}) \\ &+ 0.6 \times (\text{Total Exam and Project Score}). \end{aligned}$$

# Syllabus

## Grading

- Tentatively, one of the homework assignments is to obtain certificates for Tableau courses from LinkedIn Learning.
- The lowest homework score will be dropped when calculating the total homework score.
- When selecting the homework assignment with the lowest score, the homework assignment for Tableau is not considered.
- Each of the five homework accounts for 20% of the total homework score.

# Syllabus

## Grading

The total exam and project score is the maximum between

- (1) the simple average of the midterm exam score and the team project score and
- (2) the weighted average of them with one-third weight on the midterm exam score and two-third weight on the team project score.

(Total Exam and Project Score)

$$= \max \{ 0.5 \times (\text{Midterm Exam Score}) + 0.5 \times (\text{Team Project Score}) , \\ 0.33 \times (\text{Midterm Exam Score}) + 0.67 \times (\text{Team Project Score}) \} .$$

# Syllabus

## Grading

Letter grades will be determined by the total percentage grade:

$$\begin{aligned} 100 &\geq A \geq 93 > A- \geq 90; \\ 90 &> B+ \geq 87 > B \geq 83 > B- \geq 80; \\ 80 &> C+ \geq 77 > C \geq 73 > C- \geq 70; \\ 70 &> D \geq 60 > E. \end{aligned}$$

# Syllabus

## Make-up exams

- Make-up exams will not be given unless you have either a medically verified excuse or an absence excused by the University.
- If you cannot take exams because of religious obligations, notify me by email at least two weeks in advance so that an alternative exam time may be set.
- A missed exam without an excused absence earns a grade of zero.

# Syllabus

## Academic Integrity and Plagiarism

- All homework assignments and exams must be the original work by you.
- Examples of academic dishonesty include:
  - *representing the work, thoughts, and ideas of another person as your own;*
  - *allowing others to represent your work, thoughts, or ideas as theirs; and*
  - *being complicit in academic dishonesty by suspecting or knowing of it and not taking action.*
- Geneseo's Library offers frequent workshops to help you understand how to **paraphrase**, **quote**, and **cite** outside sources properly. See <https://www.geneseo.edu/library/library-workshops>.

# Syllabus

## Class Attendance

- If you are experiencing symptoms associated with COVID on a day that class meets in-person, do not attend.
- If you anticipate an absence due to religious observations, please contact me as soon as possible in advance to discuss your needs and arrange make up plans.
- If you are called to active military duty and need to miss classes, please let me know and consult as soon as possible with the Dean of Students.

# Syllabus

## Accessibility

- The Office of Accessibility will coordinate reasonable accommodations for persons with physical, emotional, or cognitive disabilities to ensure equal access to academic programs, activities, and services at Geneseo.
- Please contact me and the Office of Accessibility Services for questions related to access and accommodations.



# Syllabus

## Well-being

- You are strongly encouraged to communicate your needs to faculty and staff and seek support if you are experiencing unmanageable stress or are having difficulties with daily functioning.
- Liz Felski, the School of Business Student Advocate ([felski@geneseo.edu](mailto:felski@geneseo.edu), South Hall 303), or the Dean of Students (585-245-5706) can assist and provide direction to appropriate campus resources.
- For more information, see [https://www.geneseo.edu/dean\\_students](https://www.geneseo.edu/dean_students).

# Career Development

## Quick Chats (Virtual Drop-In)

- To get information about career development, you can visit the Career Development Events Calendar ([https://www.geneseo.edu/career\\_development/events/calendar](https://www.geneseo.edu/career_development/events/calendar)) or register via Handshake (<https://app.joinhandshake.com/login>).

# Look at Data

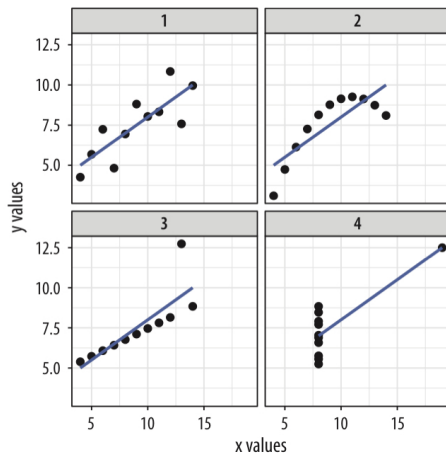


Figure 1.1 Plots of Anscombe's quartet

# Look at Data

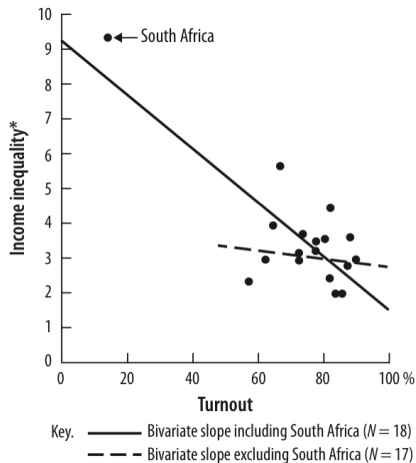


Figure 1.2 Seeing the effect of an outlier on a regression line.

# Look at Data

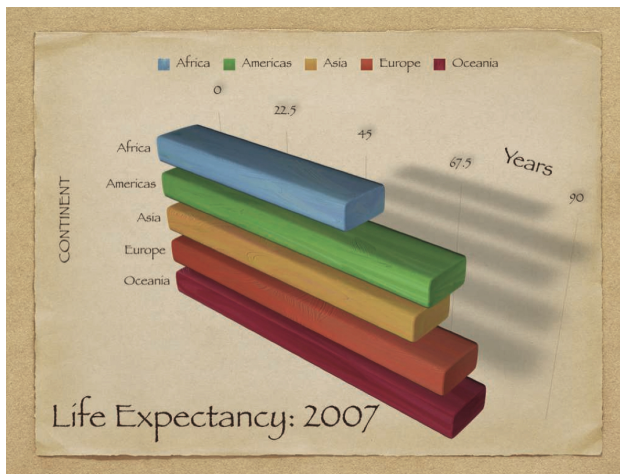


Figure 1.4 A chart with a considerable amount of junk in it.

# Look at Data

Percentage of people who say it is “essential” to live in a democracy

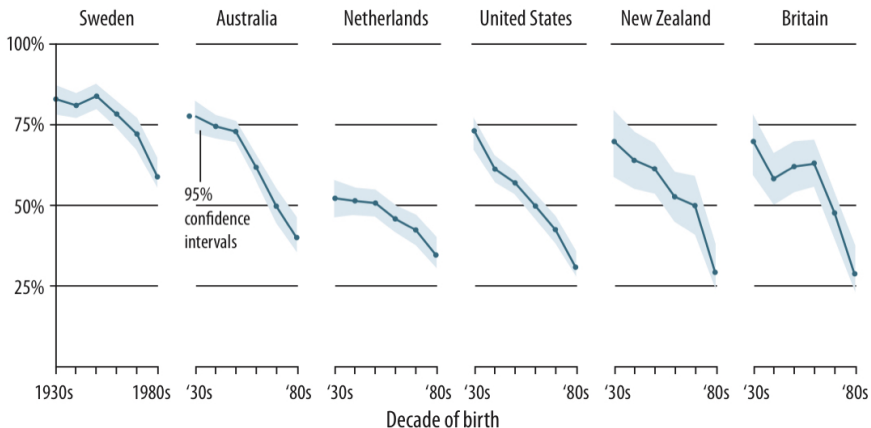
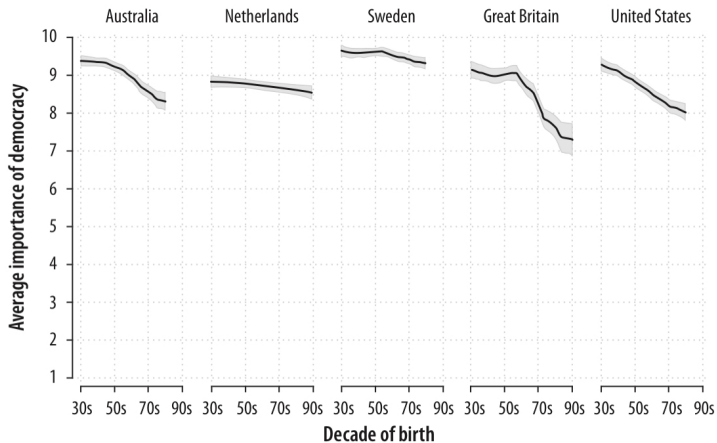


Figure 1.8 A crisis of faith in democracy?

# Look at Data



Graph by Erik Voeten, based on WVS 5

**Figure 1.9** Perhaps the crisis has been overblown.

# Look at Data

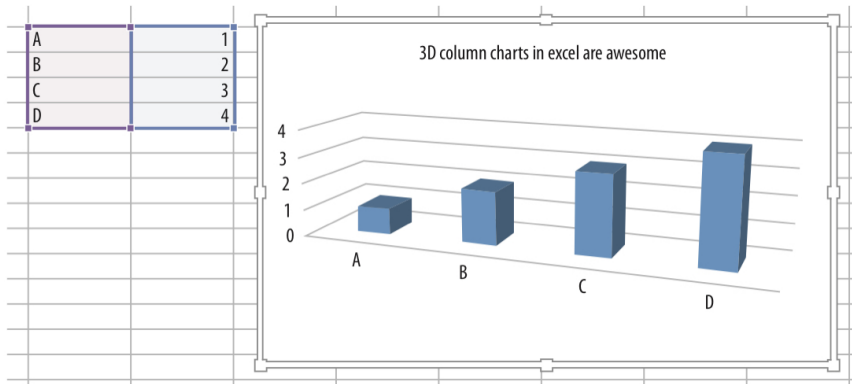


Figure 1.10 A 3-D column chart created in Microsoft Excel.



# Look at Data

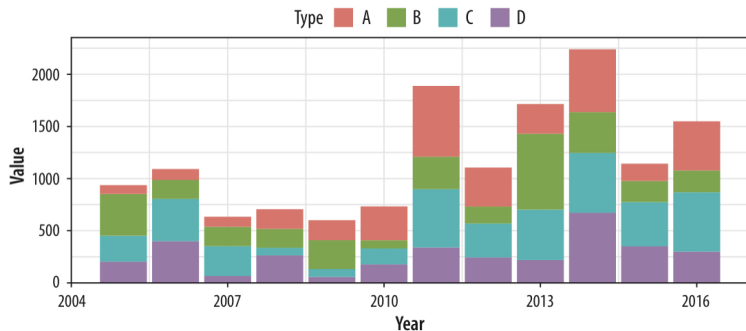
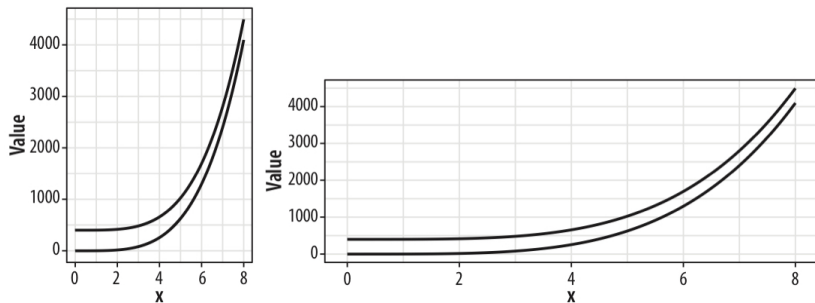


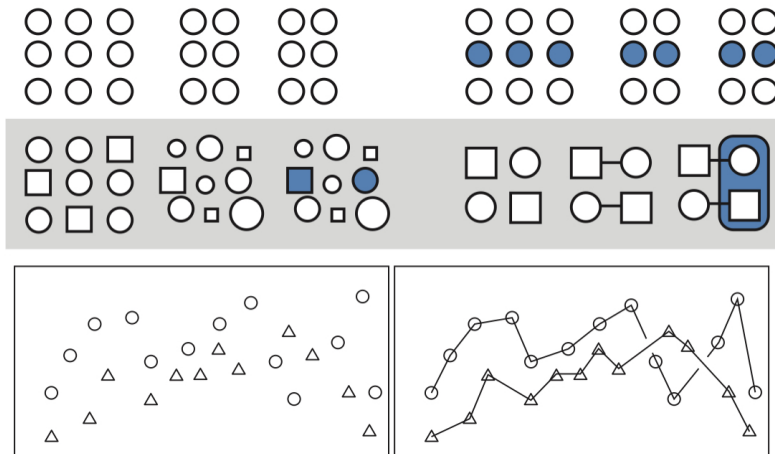
Figure 1.11 A junk-free plot that remains hard to interpret.

# Look at Data



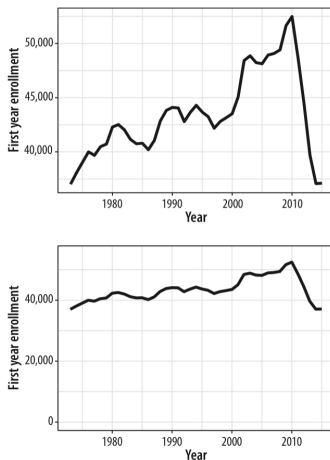
**Figure 1.12** Aspect ratios affect our perception of rates of change.

## Look at Data



**Figure 1.21** The strong inferences we make about relationships between visual elements from relatively sparse visual information are called “gestalt rules.”

## Look at Data



**Figure 1.27** Two views of the rapid decline in law school enrollments in the mid-2010s.