

EPPS6356 Data Visualization

Fall 2023 Wednesday: 7:00 pm – 9:45 pm GR3.402

Instructor: Dr. Karl Ho; 972-883-2017; GR3.203; kyho@utdallas.edu
Office Hours: Wednesday, Thursday: 2 pm – 4 pm, by appointment.
Teaching Assistant: Kelvin Chan; GR.3.318; KelvinChunMan.Chan@utdallas.edu

Overview: This course is a comprehensive training on data visualization from data literacy to advanced data presentation designs. Members will learn to communicate data patterns, deliver data messages and insights via visual representations. This course starts with data theory, principles and concepts of visualizing data, followed by different approaches and best practices in data visualization. This is a programming-based course and students will write codes with programming tools to generate charts, animated graphs and interactive contents such as dashboards. Topics on new developments and tools of data visualization will also be covered.

Learning Objectives:

At completion of this course, students will be able to:

1. Understand data theory
2. Be familiar with principles behind effective data visualization
3. Read complex data through educated eyes
4. Communicate message in data effectively using advanced visualization techniques

Required Reading:

Wilkinson, Leland. 2005. *The Grammar of Graphics. Second edition.* Springer (‡) (LW)
Murrell, Paul. 2016. *R graphics.* CRC Press.(‡) (PM)
Wickham, Hadley. *ggplot2: elegant graphics for data analysis.* springer, 2016. (*) (HW)
Cairo, Alberto. 2012. *The Functional Art: An introduction to information graphics and visualization.* New Riders. (‡) (AC1)
Cleveland, William S. and McGill, Robert. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387), pp.531-554. (CM)
Ware, Colin. 2012. *Information visualization: perception for design.* Elsevier. (‡) (CW)
Yau, Nathan. 2011. *Visualize This: The Flowingdata Guide To Design, Visualization, And Statistics.* John Wiley & Sons. (‡) (NY)
Unwin, Antony. 2015. *Graphical data analysis with R.* Vol. 27. CRC Press, (*) (AU)

(*) Online version available

(‡) Available electronically at UTD Library

Online reading:

Seaborn documentation: <https://seaborn.pydata.org/>
Matplotlib documentation: <https://matplotlib.org/>
Shiny tutorial: <http://shiny.rstudio.com>
Quarto: <https://quarto.org>
Nathan Yau Flowing Data: <https://flowingdata.com/>
Alberto Cairo Functional Art: <http://www.thefunctionalart.com>

Recommended Reading:

Healy, Kieran. 2019. *Data Visualization: A Practical Introduction.* Princeton University Press (KH)
Beeley, C., 2013. *Web application development with R using Shiny.* Packt Publishing Ltd. (CB)
Cairo, Alberto, 2016. *The truthful art: Data, charts, and maps for communication.* New Riders. (AC2)
Chang, Winston. 2012. *R graphics cookbook: practical recipes for visualizing data.* O'Reilly Media, Inc. (‡) (WC)
Chen, Chun-houh, Wolfgang Karl Härdle, and Antony Unwin, eds. 2007. *Handbook of data visualization.* Springer Science & Business Media.

- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Interpretation and Presentation." *American Journal of Political Science*. 44(2) 341-355.
- Lanzetta, Vitor Bianchi. 2017. *R Data Visualization Recipes : a Cook Book with 65+ Data Visualization Recipes for Smarter Decision-Making*. 1st edition. Birmingham, England: Packt. (‡)
- Andrew Gelman, Cristian Pasarica, Rahul Dodhia. 2002. "Let's practice what we preach: Turning tables into graphs." *The American Statistician*. 56:2(May): 121-130.
- Jake Bowers and Katherine W. Drake. 2005. "EDA for HLM: Visualization when Probabilistic Inference Fails." *Political Analysis*. 13(4): 301-326.
- Murray, Scott. 2017. *Interactive data visualization for the web: an introduction to designing with D3*. O'Reilly Media, Inc.
- Tremmel, Lothar. 1995. "The Visual Separability of Plotting Symbols in Scatterplots." *Journal of Computational and Graphical Statistics*. 4(2): 101-112.
- Ben Fry. 2007. *Visualizing Data*. O'Reilly. (‡)
- Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. Graphics Press. 2nd ed.
- Cleveland, William S., 1985. *The elements of graphing data* (pp. 135-143). Monterey, CA: Wadsworth Advanced Books and Software.
- Wong, Donna. 2014. *Wall Street Journal Guide to Information Graphics*, Norton.
- Wills, Graham. 2011. *Visualizing time: Designing graphical representations for statistical data*. Springer Science & Business Media. (GW)
- Lamigueiro, Oscar Perpinan. 2014. *Displaying Time Series, Spatial, and Space-Time Data with R*. Chapman & Hall/CRC.
- Campeato, Oswald. 2016. *D3*. Dulles, Virginia: Mercury Learning and Information.

APSA style manual (2018 edition):

<https://connect.apsanet.org/stylemanual/>

Course GitHub:

<https://github.com/datageneration/datavisualization>

Participation: Full attendance of all classes is required and imperative. All class members have to actively participate in class preparation and discussion. Participation entails full preparation for class including research of class materials, completion of assignments and full involvement in the team project. Only medical emergency (with documentation) is considered excused absence. If you have a special medical condition, please contact me in first week and special arrangements can be made. Missing one class is equivalent to missing a chance of moving one letter grade up.

Assignments:

Exercises will be assigned. Students are expected to finish all assignments and publish products to own website. apply techniques to own work. At beginning of each class, students will be selected to present and discuss assignment solutions.

Class Q&A:

Ask class related questions on Piazza (<https://piazza.com/utdallas/fall2023/epps6356/home>). Frequently Asked Questions (Q&A) will be posted there. In other words, seek answers to class, assignment, coding questions on that site.

Grading and Requirements:

Team Project:

Collaborative research is important part of this class. Each student must join a team composed of four members to work on a team project and final executive presentation. The final project must be between 25-30 pages in length. Originality is the first quality the instructor demands. Please read carefully the University's policy about cheating and plagiarism (see below for a brief version). Due dates of the proposal and progress reports are listed on the schedule below. No late submission will be accepted. Proposal of the project must be consulted with the instructor in advance and subject to the instructor's approval. It constitutes 20% of the final grade. The group project reports account for 30% of final grade. The final team project presentation is responsible for 30 percent of the final grade.

Each team must design and implement a data visualization project using the ***systems and techniques covered in class***. Identify one original topic to investigate. Originality is the first quality the instructor demands. Original means the topic or data or method(s) must be originally designed and has never been published. Replication with new data and/or method does

count but warrants strong justification (e.g. current project, dissertation research). NOTE: no Kaggle project/data replications.

Participation	10%
Assignment (posted on website)	10%
Project Proposal	20%
Final project presentation:	30%
Final project	30%

Document guideline:

All documents in this class must adhere to the following general guidelines:

- Submit electronic document to instructor’s email in PDF format (NO Word or GoogleDoc):
 - With name and content on file name:
 - E.g. smith_proposal.pdf, doe_finalproject.pdf
 - On letter size paper
 - One inch on all margins
 - 12-point font
- NO submission using link to cloud or online drive (e.g. Google drive, box.com)
- Following the APSA style and format is **mandatory**
- Suggested image format: SVG or PNG

Schedule (Tentative)

Week	Topic	Suggested readings	Due
1	Introduction: Approaches of Data Visualization		
2	Data Theory, Grammar of Graphic	LW, KH: 1	Team building
3	Messages in data	NY: 1, 2	
4	Cognitive Science in Data Visualization	CW: 1, 4, 5	
5	Functionalism in Data Visualization	AC1; CM	
6	R Graphics	NY: 3, PM: 1, 3, 6, 7	
7	R ggplot2	KH: 2, 3, 4	
8	R Shiny	http://shiny.rstudio.com	Proposal & presentation 10/11
9	JavaScript: D3	SM	
10	Building Dashboard	https://quarto.org	
11	Data topic: Time Series Data	NY: 4; GW: 3, 4	
12	Data topic: Spatial Data	NY: 8; OPL: 7, 8, 9	
13	Data topic: Relationships and Models	NY: 6,7	
14	<i>Fall break</i>		
15	Special Topic		
16	Final Project Due		12/6/2023

Comet Creed

*This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:
“As a Comet, I pledge honesty, integrity, and service in all that I do.”*

Academic Support Resources

The information contained in the following link lists the University’s academic support resources for all students. Please see <http://go.utdallas.edu/academic-support-resources>.

UT Dallas Syllabus Policies and Procedures

The information contained in the following link constitutes the University’s policies and procedures segment of the course syllabus. Please review the catalog sections regarding the [credit/no credit](#) or [pass/fail](#) grading option and withdrawal from class.

Please go to <http://go.utdallas.edu/syllabus-policies> for these policies.

The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.