DA 6233: Data Analytics Visualization and Communication

Ashwin Malshe

Fall 2023

Day Class Time and Classroom	1:00 pm - 3:45 pm Tuesday | Room 250
Day Office Hours	11:45 am to 12:45 pm Tuesday | TBD
Evening Class Time and Classroom	6:00 pm - 8:45 pm Tuesday | Room 200A
Evening Office Hours	4:45 pm - 5:45 pm Tuesday | TBD
Email	ashwin.malshe@utsa.edu
Phone	530-762-1669
Office	Main Campus BB 4.06.20
Website	www.dataviz.school

This version of the syllabus is created on 2023-08-10. This will supersede any previous versions of the syllabus.

Software

The following free software are required

- R 4.3.1 or later
- RStudio Desktop 2023.06.1+524 or later

Follow these instructions for installing and registering Tableau only after you have the license key.

1. Download Tableau Desktop and Tableau Prep here
2. Select each product download link to get started. When prompted, enter your school email address for Business E-mail and enter the name of your school for Organization.

3. Activate with your product key, which will be posted on Canvas by 08/31.

This license is valid only for fall 2023 semester. You can continue using Tableau after the semester is over by individually requesting your own one-year license through the Tableau for Students program.

Books (Not required but HIGHLY recommended)

All these books are freely available in the public domain. I strongly encourage buying them for future use.

- Fundamentals of Data Visualization (Free Online) and (Amazon)
- R for Data Science (2e) (Free Online) and (Amazon)
- R Markdown: The Definitive Guide (Free Online) and (Amazon)
- ggplot2: Elegant Graphics for Data Analysis (Free Online) and (Amazon)

Note: Also available in PDF format from the UTSA library.

- Practical Tableau (Amazon)

Note: Also available in HTML format from the UTSA library.

Additional Reading:

- How Charts Lie by Alberto Cairo: http://albertocairo.com
- `ggplot2` book website: http://ggplot2.org/book/
- `Htmlwidgets` for R: http://www.htmlwidgets.org/
- Flowing Data: https://flowingdata.com/

Online Tutoring

UTSA has partnered with TutorMe to provide online tutoring up to four hours a week per student. You can access this service from Canvas by clicking on "TutorMe" link in the left menu bar.

To use TutorMe effectively, I strongly recommend watching this video and reading this document.

Learning Objectives

1. Get proficient in R coding, specifically in data wrangling
    - Primarily using `tidyverse` packages
2. Learn fundamentals of data visualization and communication
    - Learn about data storytelling
3. Visualization for cross-sectional, time-series, and spatial data
    - Using R packages such as `ggplot2` for static visualizations
    - Using `JavaScript` libraries in R for interactive visualizations
    - Using Tableau for creating dashboards

Course Description

This course is designed to instruct students in the fundamental principles of data visualization and acquaint them with tools that facilitate more effective visual communication. As data analytics methods continue to gain prominence, conveying the results of such analyses to decision-makers remains a considerable challenge. Visualization tools, including graphs, dashboards, and websites, can greatly aid in this communication process. However, without a comprehensive understanding of best practices in visualization, it is easy to inadvertently mislead audiences through inaccurate object sizes, inappropriate color shading, excessive clutter, or arbitrarily truncated scales. To emphasize this point, the syllabus concludes with three images sourced from the subreddit r/dataisugly, which clearly illustrate these pitfalls.

Throughout the course, students will utilize R and Tableau to create visualizations. R offers several specialized libraries for plotting, with a particular focus on `ggplot2` and its dependent libraries. These libraries share a similar syntax for their functions, which simplifies the learning process. Tableau, a Salesforce-owned company, is widely regarded as one of the most popular visualization software tools among industry professionals. It competes with various dashboard software options, including MicroStrategy, Alteryx, and Microsoft's Power BI.

Schedule

Week 01, 08/22: Introduction and Basics of R, RStudio, and Github
Week 02, 08/29: Data wrangling with R
Week 03, 09/05: Data wrangling with R
Week 04, 09/12: *What we can easily see and Data storytelling*
Week 05, 09/19: The grammar of graphics with `ggplot2`
Week 06, 09/26: The grammar of graphics with `ggplot2`
Week 07, 10/03: The grammar of graphics with `ggplot2`
Week 08, 10/10: Final project topics discussion
Week 09, 10/17: Introduction to interactive visualizations
Week 10, 10/24: Introduction to interactive visualizations
Week 11, 10/31: Tableau
Week 12, 11/07: Tableau
Week 13, 11/14: Tableau
Week 14, 11/21: Tableau
Week 15, 11/28: Final presentations

Grading

Grading is based on group and individual assignments and attendance.

Attendance - 10%

Homework assignments – 60%

Final project presentation – 30%

Score	Grade
97.5-100	A+
95-97.5	A
90-95	A-
87.5-90	B+
85-87.5	B
80-85	B-
77.5-80	C+
75-77.5	C+
< 75	F

Attendance

Although the expected attendance for everyone is 100%, a minimum of 80% attendance is required to pass the course. This translates to 12 classes out of 15. ***Attendance less than***

80% will result in an F. However, if your attendance is more than or equal to 60% and less than 80% (i.e., you attended 9-11 classes out of 15) and you have supporting documents to justify your absence, you will get 5% for the attendance.

Supporting documents when attendance >= 60% but < 80%

If you miss any classes due to illness, please submit a physician's note to me. The students in the evening cohort may also miss classes due to work-related travel. In those cases, they will have to submit a letter from their immediate supervisor. There are other possible reasons such as family emergencies and weather-related absences. In such cases I will make a decision on a case-by-case basis.

The supporting documents are necessary only if the attendance is falling below 80%. The supporting documents don't earn you a "presence"! For example, if you miss 1 class out of 15 due to illness and you have a doctor's note, it doesn't grant you 100% attendance. Your attendance is still 14/15. **Thus, documents help you out only when you are failing.**

Table 1: Attendance

Attendance	Points
15/15 (100%)	10%
14/15	9%
13/15	8.5%
12/15 (80%)	8%

Note:

Attendance less than 80% translates to F

Homework Assignments (Individual)

There will be four individual homework assignments, each with a one week deadline. The homework will involve creating visualizations. You will submit your code and output combined as an Rmarkdown file along with the output html file. There will be a homework on Tableau as well. In that case you will submit your Tableau Workbook. Each homework assignment is worth 15 points.

Homework Deadlines

Table 2: Homework Submission

Homework	Maximum Points	Topic	Made Available On	Due Date
1	15%	R Coding	09/06	09/13
2	15%	ggplot2	10/04	10/11
3	15%	Interactive Viz	10/25	11/01
4	15%	Tableau	11/08	11/15

Note:
The hard deadline is 18:00 CST on the submission date

Final Project (Group work)

The final project will require you to create a static or interactive visualization that helps educate organizations (profit or non-profit), policymakers, consumers, etc., using publicly available data sets. The final project will be entirely graded on the dashboard. I will grade the content of the presentation according to the relevance of the topic to practitioners (purely academic visualizations for the sake of it will get low grades), attention to details, and data storytelling. Pretty visualizations that do not tell us about anything interesting are meaningless.

I will post a list of project topics and related datasets on Github and make them available to you in the first week of classes. Some of the examples are the Covid-19 dashboard, 2020 US Presidential Elections dashboard, etc.

> You have the option of forming your own group or letting me assign you to a group. I will share a Google spreadsheet on the first day where you can indicate your preference for group formation method. If you prefer to form your own group, make sure that the group has exactly four students. I will make groups of four students each based on the UTSA IDs. The groups will be finalized by the end of the first week because we may have late admissions.

Late Submissions

No late submissions are allowed. The submission links on Canvas will disappear past the deadline. If you anticipate late submission, email me *in advance* so that I can give you extra time if needed.

Artificial Intelligence (AI) Use Guidance

This is not a campus-wide policy and only applies to this course. This policy may change, with prior notice, to reflect developments in the technology. Generative AI tools such as ChatGPT may be not used in this course. Note that generative AI can produce questionable results, including misinformation, nonexistent references, and bias. More importantly, whereas generative AI can make you a better programmer once you know enough programming, relying on it as a beginner will likely to harm more than help you. (Adopted from Dr. Max Kilger's syllabus)

Policy on Cheating

Students are expected to be above reproach in academic activities. Students who violate University rules on academic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and dismissal from the University. "Scholastic dishonesty includes, but is not limited to, cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an exam for another person, any act designed to give an unfair advantage to a student or the attempt to commit such acts." From The University of Texas System Rules and Regulations of the Board of Regents, Rule: 50101. (www.utsystem.edu/BOR/rules.htm).

Right to Privacy

Except under specific exceptions provided in the Family Education Rights and Privacy Act of 1974, I will not give information concerning your grades, academic progress, attendance, address, phone, or e-mail to anyone outside the UTSA system unless you give your prior written permission. In addition, I will not provide or discuss grade information over the phone or by e-mail.

Special Needs

If you feel that you are eligible for or maybe helped by accommodations in the class due to a disability or special need, contact the Office of Disability Services (ODS). Students with disabilities must be registered with the ODS located in MS 2.03.18 (458 4157 – voice; 458 4981 – TRY) or UTSA Downtown in FS 1.526 (458-2816) in order to receive support services. To see if you are eligible for these services and privileges, visit the website below: http://www.utsa.edu/disability/studeligibility.htm

Three Examples of Bad Visualizations

Here I show 3 examples of bad visualizations. I will discuss these briefly in the class.

The fast browser that broke the scale (Source)

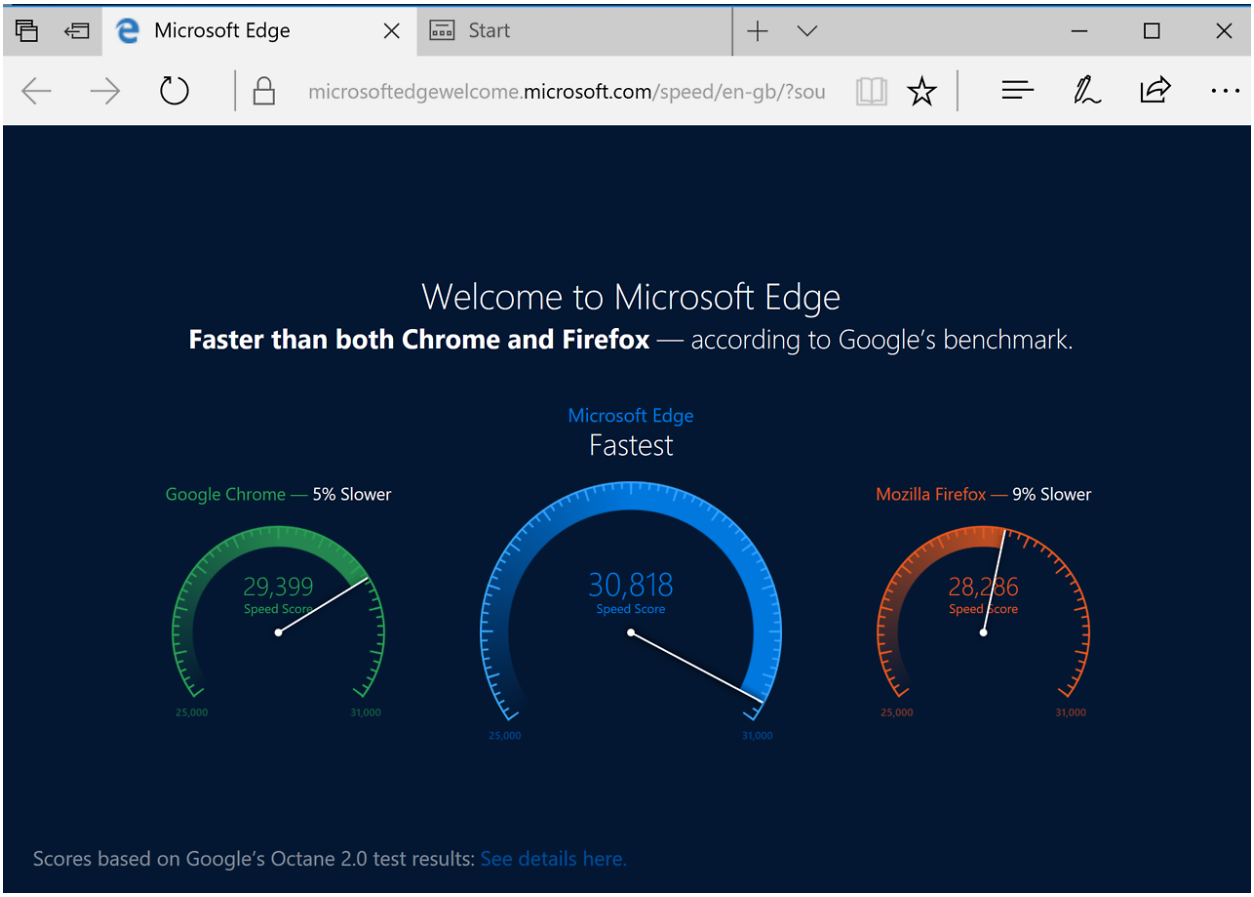

Figure 1: Microsoft Edge

The pie chart (pizza chart?) that doesn't add up to 100%(Source)

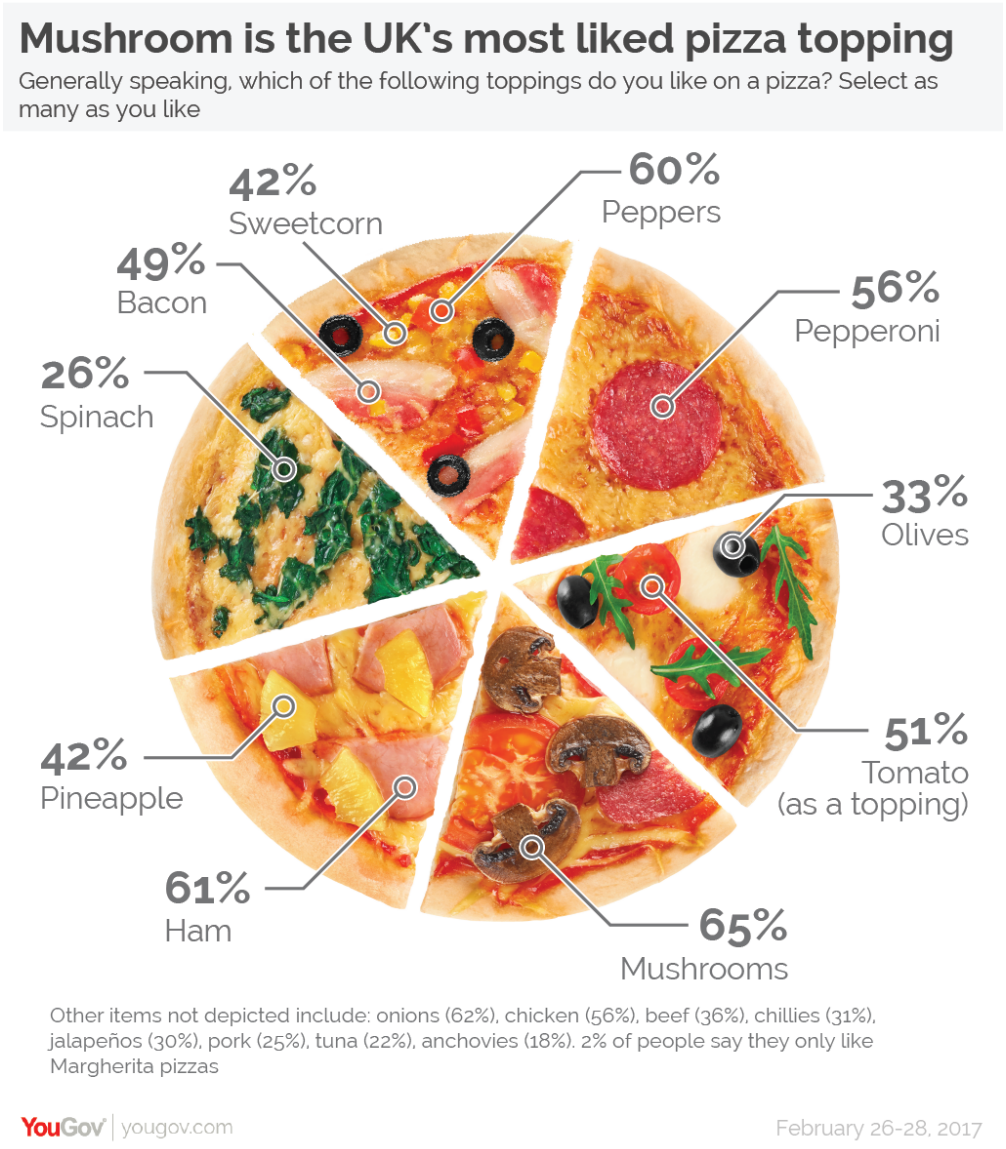

Figure 2: Pizza Chart

Stand your ground flips the image (Source)

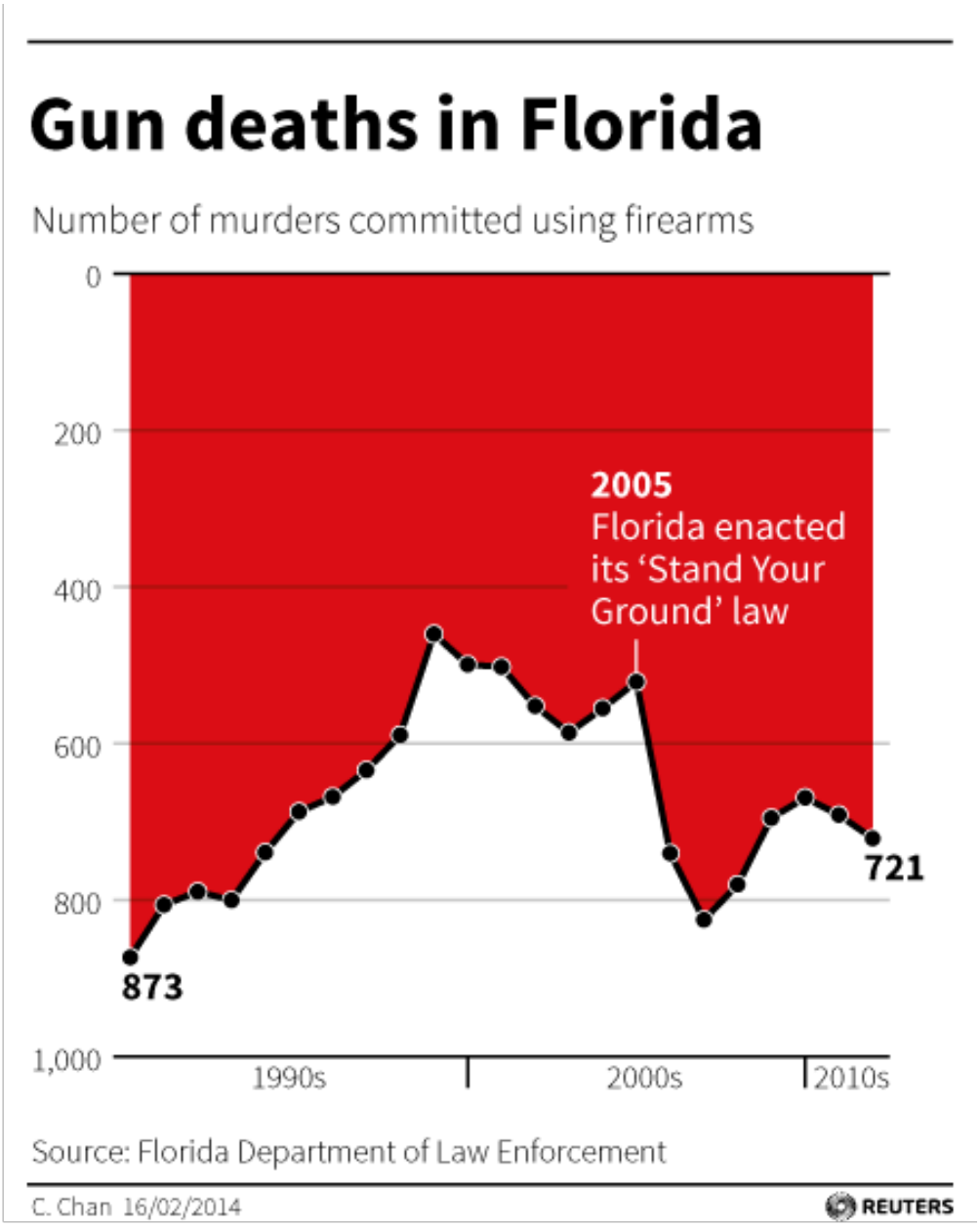

Figure 3: Gun Deaths