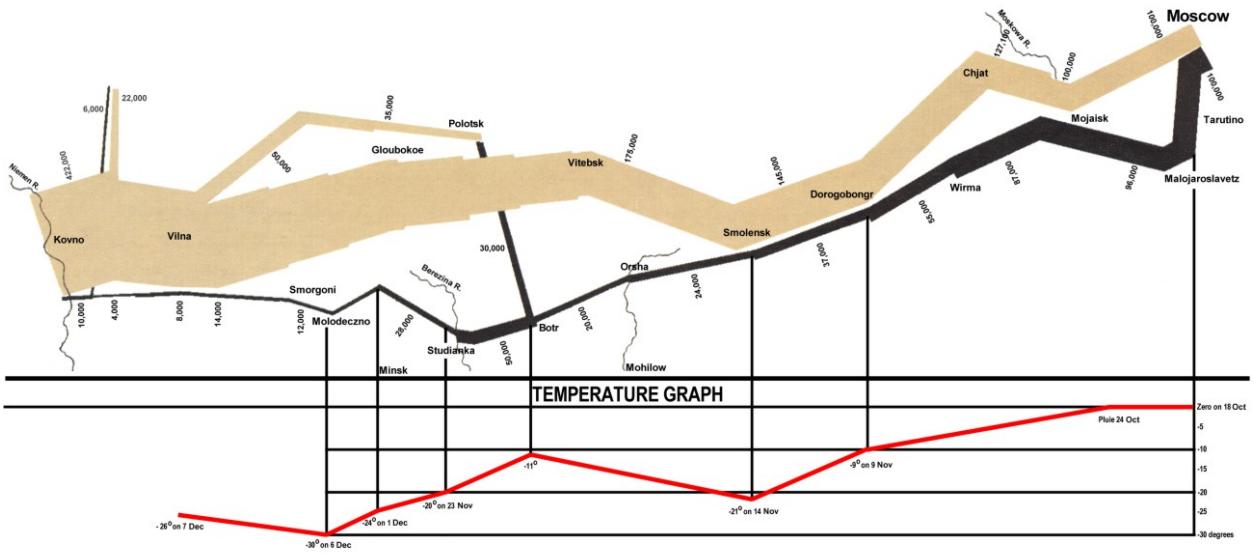
Data Visualization
CSCI 444 (UG)
Fall 2020 Syllabus

> The minimum we should hope for with any display technology is that it should do no harm.
>
> –EDWARD TUFTE



Instructor Details

Name:	Professor Jesse Johnson
Office:	417 Social Science Building
Telephone:	(406) 243-2356
Email:	`jesse.johnson@umontana.edu`
Web:	Faculty Home Page
Office Hours:	MWF 10:00–11:00 , Online Only (via Teams)
	Or, by appointment.

Prerequisites

Students taking this course are expected to have:

- Experience with modern, complex quantitative software packages to the degree that *independently* mastering several new software packages in a semester does not present a problem. *The instructor will not teach students to use the software.*
- Organizational skills and familiarity with computers sufficient to install new software, create a file system for the course, and execute programs and move files from the command line.
- Some programming experience with any language.
- Evidence of mathematical maturity as shown by successful completion of calculus and/or statistics.
- Maturity enough to show up for class, consistently.
- Maturity enough to offer constructive criticism to your peers.

Course Objectives

This course emphasizes the practice of data visualization, compelling students to identify and master tools to produce high quality visualizations. Through the course, students will encounter datasets having different types of relations between members. Hence, a majority of the student's time is spent in the creation of original visuals. To guide the process, lecture time will initially be spent considering visualizations that represent best, and worst practices. Later lectures will consider the way humans perceive, and how limits of human perception drive decisions about the display of quantitative information. Additionally, we will consider photographic principles that are relevant to recording video. Moreover, students acquire a sophisticated framework for assessing the quality of visualizations, the technical skills required to produce visualizations, and an understanding of the mathematical challenges that underlay many data visualization techniques.

Student Outcomes

Upon successful completion of this course, student will be better able to:

1. analyze a problem, and identify and define the computing requirements appropriate to its solution.
2. design, implement, and evaluate a computer-based process to for the visualization of data.
3. quickly adapt to any quantitative visualization programming environment.
4. precondition large data sets to make them readily accessible to the visualization software being used.
5. critique visualizations using a framework focused on the accurate display of quantitative information and the limitations of human perception.
6. communicate effectively with a range of audiences.

Textbook

This semester I'll be using the following text. You need to purchase a copy.

The Visual Display of Quantitative Information
Edward Tufte
Graphics Press
2001

Course Logistics

Hypothesis Driven Visualization

The course is project driven. All but one of the projects share the same approach, but differ in data sets used. This semester, expect to do 6-7 of these projects. Given a data set, your approach will always be as follows:

1. Form a hypothesis from the data. State the hypothesis as clearly as possible.
2. Produce a visual from the data to support your hypothesis. Your visual may have multiple panels, but should be no more than one 8.5 by 11 inch page.
3. Produce a second visual that has arrows and labels that indicate how and where your visual is consistent with best practices discussed in class. It may be the best way to do this is to create your visual, print it out, mark it up with a pen, photograph or scan it with your phone, and then submit that.

Rubrics for grading projects will be provided to students.

Required Software

I've made the class accessible to more students by allowing assignments to be done with software packages instead of programming languages. Upper division and graduate level MIS, Math, Data Science, Economics, Physics, Chemistry, Business Analytics, and other students are welcome here. You are free to complete the assignments using the suggested or other software. All I ask is that you do not use Microsoft Excel or Google Sheets, just because you've probably used those elsewhere.

If you are a student that can program, you should program. It'll help you think about data in a more structured way. However, if you are a computer science student, you must program. The requirements follow:

- **Undergraduate computer science** students have to complete 3 of the assignments using either Vega or Vega-lite
- **Graduate computer science** students have to complete 3 of the assignments using d3js.

These are web-based tools for preparing visualizations. For your assignments, turn in a link to the web page for me to review, as well as a printout highlighting features of the visualization. I'd like you to use your github account to create the web page.

Online Resources

As in previous years, there is a Moodle supplement to the course.

In addition, I've created a Microsoft Teams supplement. I recommend we use that to discuss assignments and software. I am also considering its use for recording and storing classroom sessions, but am still inquiring about the disk space available.

COVID-19 Information

You will no doubt see these expectations in other courses, but I've been asked to make them clear to you, so here they are:

- Mask use is required within the classroom
- Each student is provided with a cleaning kit. The expectation is that students will clean their personal work space when they arrive for class, and before they leave the classroom
- Students are discouraged from congregating outside the classroom before and after class
- Specific seating arrangements will be used to ensure social distancing and support contact tracing efforts
- Class attendance will be recorded to support contact tracing efforts
- Drinking liquids and eating food is discouraged within the classroom (which requires mask removal)
- Hand sanitizer and cleaning supplies can be requested in the main office, in Social Science 401.

- The classroom sessions will be recorded and streamed, as well as stored for asynchronous review.
- Stay home if you feel sick and/or if exhibiting COVID-19 symptoms, please contact the Curry Health Center at (406) 243-4330
- Up-to-Date COVID-19 Information from the University of Montana UM Coronavirus Website: UM COVID-19 Fall 2020 website
- You are strongly encouraged to remain vigilant outside the classroom in mitigating the spread of COVID-19

Assignments

Much of the learning in this class achieved by doing. I've been developing assignments over the years and they break down about as follows. I'll continue to refine these through the semester, so the final form may be slightly different from what I say here.

Throughout, data are required. There are no constraints on the data you use for your assignments, provided the *data relation* is upheld. One excellent clearing house for data is awesome-public-datasets.

1. **The scatter plot**

   Data Relation: One to one relations, or ordered-pairs

   English Language Description: The most common type of data visualization for a reason, plots of ordered-pairs reveal trends in data that can only be seen in a graphical construction. Fundamentally, the relation is that of independent and dependent variables.

   Example Data Sources: The bureau of labor statistics provides a lot to think about: Bureau of Labor Statistics. If work and money isn't your thing, try disease Center for Disease Control.

   Example Hypotheses: Montana's wages are higher for workers without college degrees than wages for similar workers in Washington, Oregon, and Idaho. Summer is the best season to contract an STD, except when the carnival comes to town. The period having the highest incidence of tuberculosis in the United States corresponded to the period of highest troop deployments overseas.

   Suggested Software: Tableau

2. **The raster plot**

   Data Relation: A matrix of values, each value corresponding to some location.

   English Language Description: A two-dimensional plane is broken into cells. In each cell, there is a value. Each cell is associated with a location and has a size. Such data are used to represent everything from satellite data to the output of a mathematical function.

Example Data Sources: This assignment requires students use geoTIFF formatted data, there are alternatives to that format, but geoTIFF is the most common. As such, you'll find many potential sources, but a few that are interesting include National Snow and Ice Datacenter and Socioeconomic Data and Applications Center

Example Hypotheses: The extent of glaciated areas has decreased more in the 90s that the period 2000-2010. The fastest growing populations are in equatorial regions of the world.

Suggested Software: Tableau, QGIS

3. The geographic vector data

Data Relation: A set of points, distributed on the surface of Earth, or other planet. Each point registers the location of a feature of interest. In some cases, there is a relationship between the points, as they define a road, or a political boundary. Other times they are independent, like cities.

English Language Description: Here, ordered pairs are longitude and latitude. Special care must be taken to assure that the projection of points on the surface of a sphere (lat./long. pair) to a plane (such as a Mercator projection)is done correctly and consistently.

Example Data Sources: Focus on ERSI Shapefile format, which is supported by most software. A create hub for free shapefile data is GISGeography

Example Hypotheses: Coral Reefs are more common in the tropics. The ice shelves of Antarctica occur only in regions that would be massive bays.

Suggested Software: Tableau, QGIS

4. The network

Data Relation: A graph, or numerous one-to-one and one-to-many relations within the same data set.

English Language Description: Networks are graphs of complex relations between nodes. A node might be a person, and edge would denote a friendship. Evaluation of these relations is an open problem and useful for a addressing questions ranging from marketing to radicalization of terrorists.

Example Data Sources: Sometimes smaller is easier in this assignment, try this site.

Example Hypotheses: Networks often have a highly connected node that, if removed, would result in several disjoint networks.

Suggested Software: Gephi

5. The simulation model output

Data Relation: Time dependent relations across a polygonal mesh of nodal points.

English Language Description: This will be the output of a simulation model, which solves a set of equations, providing the value on the nodal points, and then interpolating between nodal points.

Example Data Sources: Paraview tutorial data

Example Hypotheses: Very dependant on the data set used.

Suggested Software: Paraview

6. **The movie**

Data Relation: Multiple raster frames, collected with a lens and an image sensor.

English Language Description: In this assignment you will take a video that demonstrates some physical, chemical, or biological phenomena.

Example Data Sources: I got excited about this assignment by reviewing this classic set of films: National Committee for Fluid Mechanics Films and thinking about my own interest in photography. I will draw some basic lectures on photography from the Mark Levoy's excellent resource.

Example Hypotheses: Not applicable, here you will focus on the best way to reveal what you are looking at.

Suggested Software: For this assignment you will need a cell phone with a 'pro' mode that can be accessed for video. Slow motion and time lapse modes could be useful as well. It is also acceptable if you have a camera with a movie mode. If you don't have either of these, see me. We'll figure something out.

7. **Human Acuities**

Data Relation: Not applicable

English Language Description: This assignment is to simply identify the best possible display, but not to over-buy. Many display technologies provide specifications that exceed human limits of perception in some ways, but have a marketing appeal. Others lag human perception. For this assignment consider at least resolution, dynamic range, and color gamut to identify the best possible display for a person that will view the display from a distance of 2 meters.

Example Data Sources: Look to articles like this to inform your search. There is a lot out there on this topic.

Example Hypotheses: Not applicable in this assignment.

Suggested Software: Not applicable.

Graduate Increment

This is a **"UG"** course, meaning it can be taken for graduate credit.

If you are a graduate student taking the course, then you'll be required to do one additional assignment. You can do use any dataset you like for this assignment. You will present the results of this assignment to the class during the final examination meeting.

Meeting Times/Place

Times: Monday, Wednesday, Friday 15:00–15:50
Place: Social Science 362 and online
Online: *Microsoft Teams Classroom*

Final Exam Time and Place

Time: 13:10-15:10, Monday, November 23
Place: Social Science 362 and online
Online: *Microsoft Teams Classroom*

Grading Policy

Grading scale

A	94-100
A-	90-93
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	0-59

Students achieving the numerical scores above are guaranteed the associated letter grade. However, if average performance is low, I may decide to assign a higher letter grade for a lower score; e.g. a B+ for a numerical score of 84.

Students taking the course pass/no pass are required to earn a grade of D or better in order to pass.

Assessments and weights

The following assessments will be used and weighted according to the values in the table to determine final grades.

Component	Description	Weight (UG/G)
Projects	A total of 6-7 projects that use various sets of data and software tools.	80/60%
Final project	A more comprehensive final project based on data students identify.	0/20%
Group/classroom	Some grades will be developed from participation in classroom and online activities	20/20%

Tentative schedule:

MONDAY	WEDNESDAY	FRIDAY
Aug 19th **1** Graphical Excellence I	21st **2** Graphical Excellence II	23rd **3** Graphical Integrity I
26th **4** Graphical Integrity II	28th **5** Theory of Data Graphics I	30th **6** Theory of Data Graphics II
Sep 2nd **7** Assignment 1 Due: The Scatter Plot	4th **8** Data-Ink and Graphical Redesign I	6th **9** Data-Ink and Graphical Redesign II
9th **10** Chartjunk: Vibrations, Grids, and Ducks I	11th **11** Chartjunk: Vibrations, Grids, and Ducks II	13th **12** Data-Ink Maximization and Graphical Design I
16th **13** Data-Ink Maximization and Graphical Design II	18th **14** Assignment 2 Due: The Raster Plot	20th **15** Sparklines: Intense, Simple, Word-Sized Graphics
23rd **16** Multi-functioning Graphical Elements I	25th **17** Multi-functioning Graphical Elements II	27th **18** Data Density and Small Multiples I
30th **19** Data Density and Small Multiples II	Oct 2nd **20** Aesthetics and Technique in Graphical Design	4th **21** Assignment 3 Due:The Geographic Data Vector
7th **22** Design for the Display of Information	9th **23** Mapped Pictures: Images as Evidence and Explanation	11th **24** Links and Causal Arrows: Ambiguity in Action
14th **25** The Fundamental Principles of Analytical Design	16th **26** Corruption in Evidence	18th **27** The Cognitive Style of PowerPoint: Pitching Out Corrupts within I
21st **28** Assignment 4 Due: The Network	23rd **29** The Cognitive Style of PowerPoint: Pitching Out Corrupts within II	25th **30** Elements of photography I
28th **31** Elements of photography II	30th **32** Elements of photography III	Nov 1st **33** Assignment 5 Due: The Simulation Model Output

MONDAY	WEDNESDAY	FRIDAY
4th **34** Elements of photography IV	6th **35** Sculptural Pedestals: Meaning, Practice, and Depedestalization I	8th **36** Sculptural Pedestals: Meaning, Practice, and Depedestalization II
11th *Veterans Day*	13th **37** Assignment 6 Due: The Movie	15th **38** Visual acuities I
18th **39** Visual acuities II	20th **40** Assignment 7 Due: Visual Acuities	22nd **41** Wrap up/Course evaluation
25th **42** Finals Week	27th **43** Final Exam	29th **44** Finals Week

Attendance Policy

Attendance will not be taken. Students absent when called up to critique visuals will be given a grade of 0%. Students informing the instructor of a valid reason for missing class *in advance*, via email, will not be called to the board. Valid reasons include family emergencies and illness. I may ask for documentation of absence (doctors note, death certificate, etc.). Cell phone photos are useful for this - a selfie in the doctors office, or next to a car that won't start...

Online presence will be counted the same as physical presence for the purpose of any required presentations.

Academic Integrity

All students must practice academic honesty. Academic misconduct is subject to an academic penalty by the course instructor and/or a disciplinary sanction by the University. All students need to be familiar with the Student Conduct Code. I will follow the guidelines given there. In cases of academic dishonesty, I will seek out the maximum allowable penalty. If you have questions about which behaviors are acceptable, especially regarding use of code found on the internet or shared by your peers, please ask me.

Disabilities

Students with disabilities may request reasonable modifications by contacting me. The University of Montana assures equal access to instruction through collaboration between students with disabilities, instructors, and Disability Services for Students. "Reasonable" means the University permits no fundamental alterations of academic standards or retroactive modifications.