

CSE 5544 Lab 4/4 (15/60 pts):

Finding Noise and Inconsistency | consistencies in human decisions

Assigned 3/7 out, Due 3/11, 3/28 8:59 pm

This homework assignment has two deadlines. The first deadline is for the in-class activities, when we will collaboratively work out a solution. The second part will be the individual's implementation.

The resource folder is online at

https://drive.google.com/drive/folders/1LeJwZ4BX23DtcfBwyuqYzec0Tn9xgNnN?usp=share_link.

(please bear with us on any noise in the assignment. We strive to fix them.)

Learning Goal

In this assignment, we'll begin to provide hands-on experience designing, refining, implementing, and even evaluating a visualization method. In this assignment, we will use pixel-based representation for multivariate data analysis.

The data we collect and visualize can have a big impact on people. One issue with any data analysis is the quality of the input. Modern machine learning algorithms collect data from people. It is not acceptable for similar quality medical imaging scans, to end up with dramatically different cancer diagnoses - say one benign and the other malignant from two different doctors or two hospitals. And yet in real doctors' offices, something like that happens and the frequency is often higher than we expect. The current system relies on down-streaming reviews for patient safety. In this modern machine and human intelligence era, perhaps we can somehow capture this type of errors, communicate to the general public, or create means to fix the system.

To Err is Human.

Our tales involve an empirical study data collected via an image annotation program where 7 expert coders were asked to code 2,000+ images by assigning a visualization type to images. Our goal in this work is to show those inconsistencies and consistencies by offering an overview of the coders' results.

- There are 11 types in Columns (O and P): e.g., surface/volume, bar, point, line, glyph, color, etc. And OTHERS (that we don't have codes)

Many choices we make in visual analytics impact what our audience will be able to see and take away. These choices range from the data we choose to analyze, the variables we choose to show, and the colors we choose to use, to the text we choose to accompany a visual representation.

In this assignment, you will be working in **groups of 2-3** on the first part of the project that will focus on mapping data to some visual forms. We will use the image coding dataset or diagnostic disagreement from the cancer data. You can choose to be persuasive (while being true to the data) and try to make strong cases for your topic.

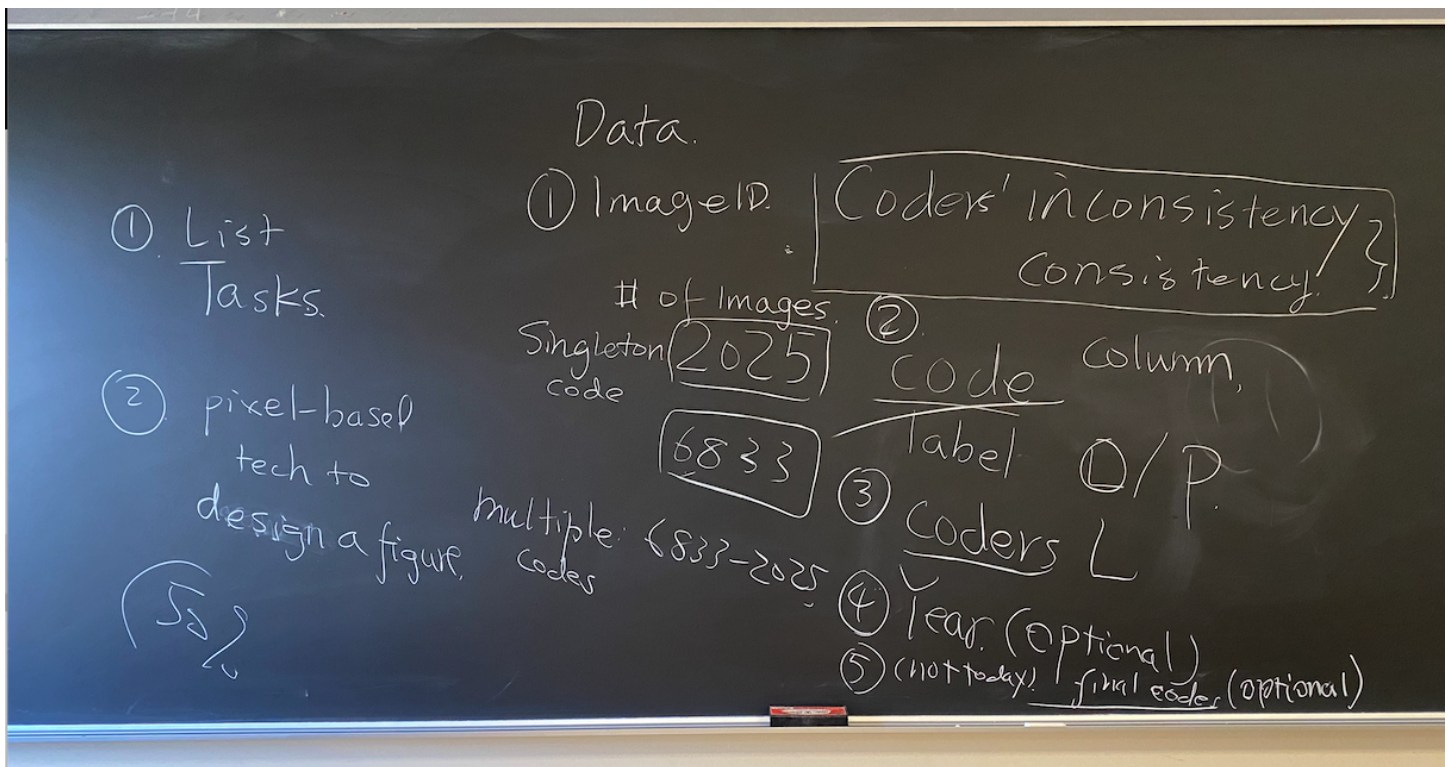
Prior to starting your project, it is helpful to gain a sense of what goes into formulating a successful visualization project and to beware of common pitfalls. We encourage you to read [A Nested Model for Visualization Design and Validation](#) by Tamara Munzner. You will also present your team's design online and hope to receive peer's consent for being the best and receive 10% extra credits for your work.

Step 1: Data Analysis and initial design (individual, in-class exercise, 20%)

In this assignment, you can keep working on the image coding dataset - or the cancer diagnosis dataset. We have collected coders' results. See the blackboard discussion below.

In class, we have assigned two work:

- List the set of tasks your visualizations will address;
- Illustrate the solution space using the [pixel-based bar chart](#) to show the multivariate dataset.
 - [See alternatives from Microsoft Research.](#)



Now after the March 7th class, you shall have your own design.

Please read the instruction here before you cast your vote.

<https://docs.google.com/document/d/16V5eGCacSbcNy0CwWfdqA0leUZ3IIspiKs0NZSKS3Kw/edit>.

Case your vote (for 1 extra credits):

<https://docs.google.com/forms/d/1kx7FgrS01oIV5ENrR58hHYli0S8SH-NxIXGLK40wdSE/edit>

The index is the same with the index in section 2 (21 design solutions by you)

In the March 9th class, we will form groups to discuss the design. This will be the 1st iteration for your team to pick and choose a final solution for your group. by following Step 2.

Step 2: Task Refinement and Visual Refinement (team, 20%)

Now you feel like you have a good grasp of how to convert data to visual representations. Now, turn around and compare and contrast your task list with those of your peers. Try to arrive a consensus by listening to others and try to write a half to one-pager report about,

- [Description of how someone did a data visualization project - not too much unlike the one you will work on](#) but someone else's from the following resources.

Then your team is supposed to come up with THREE tasks, and ONE visualization that will provide solutions for someone else to fulfill these tasks.

We will spend the 1st half of the lecture on Thursday. Please turn it on by ~~Friday~~ **Saturday** evening.

Resources

- [A list of additional articles, podcasts, papers, book on Vis4Good](#)
- [Article on how to make a real-world impact with data visualization](#)

Examples

- [Data Science for Social Good projects at the eScience Institute](#)
- [Perceiving Yemen](#)

What to turn in

March 9 after class: Finish reading the materials above to fine-tune your design. We will do in-class activities to expand your design solution space via team discussion. Then each team finalizes an initial set of **five** questions (your team will have 3 final questions out of all questions being discussed). We'd like to investigate key topics of interest.

- **Team submission (1 per team. Please write down everyone's name)**
 - **This is the agreement each team arrives for your best design**
- **Your individual submission (describe the mapping tables of your two teammates' original design)**

March 10: Turn in your team's solution by 8:59pm. Each team only needs to submit one entry and please clearly show the team information (who and your OSU dot id in the file names).

The best way to come up with a good idea is to have lots of them, and narrow them down to the one your team agrees. In this phase, you and your teammates will brainstorm designs that help to clearly communicate the inconsistency and consistency of data or help other people to explore the data according to the list of questions. Your task in this assignment is to brainstorm many designs individuals conceive, and then use some basic criteria to filter and select the best ideas. You will then hand in a polished version of the final designs that your team chooses.

The goal of this exercise is to show you that if you have an open mind, your first idea is (unlikely) to be your best idea. Instead, the process of brainstorming, discussing, and affinity diagramming can help you find good ideas that you had not considered before.

Important:

- do not limit yourself to things you know how to build yet - purely think about the best design you can come up with.

Your Team Exercise

In this exercise, go through the following steps:

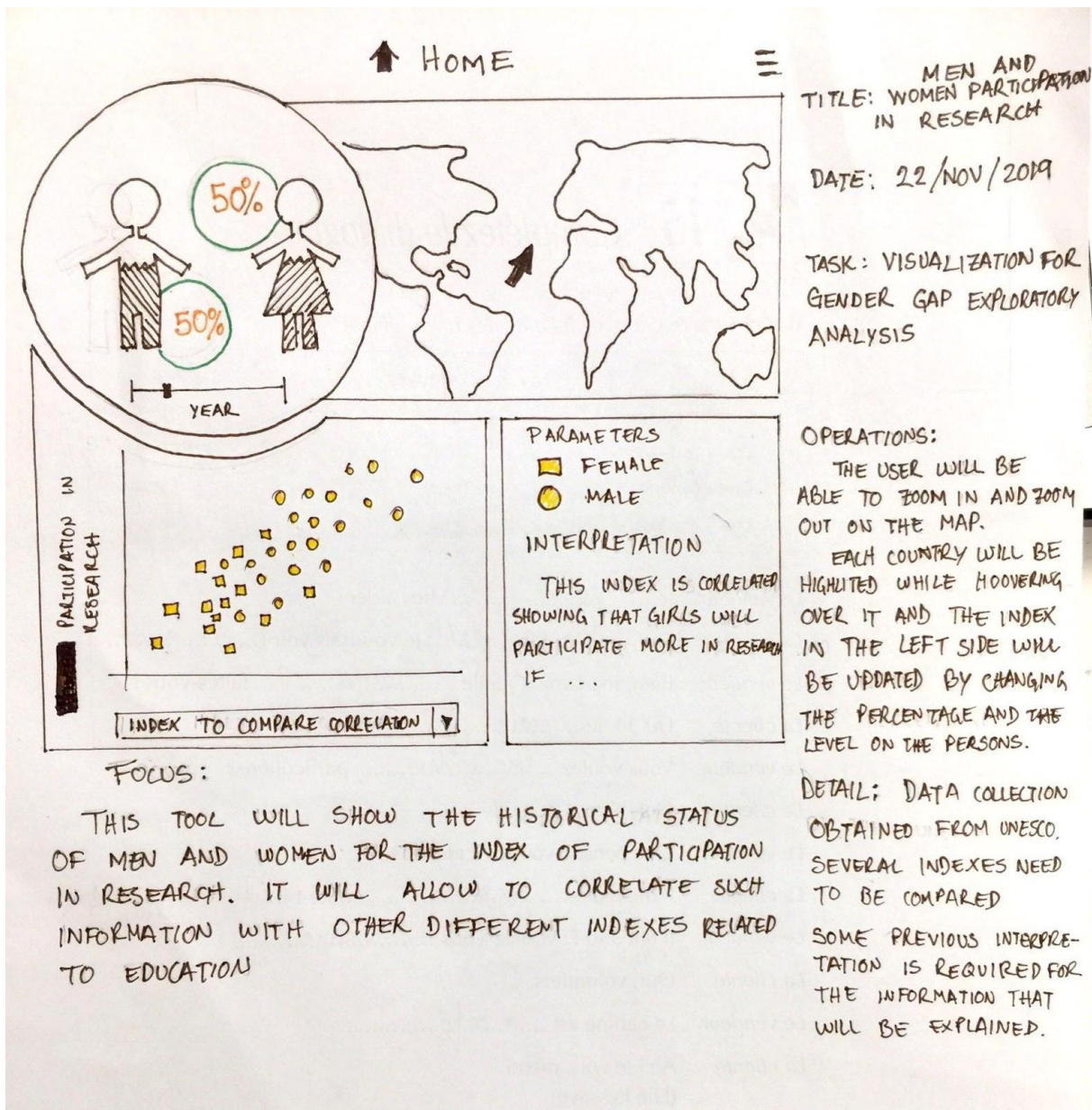
- **Brainstorm session.** With your group, schedule a mutual time that you can get together and work for at least an hour together. Brainstorm and show your own ideas, and explain via your sketch -- on a single sheet of paper. The goal here is to share each other's distinct ideas (hopefully) -- your team should aim to address the questions of interest. You/your team are aiming here for variance: the ideas should be different from one another. You are allowed to build off of one another's ideas, but make sure that they're different. If you end up with a bunch of different ways to solve the same problem and essentially variations on the exact same idea, try again when you feel your team is doing it right.
- **Discussion.** This can be part of your brainstorming session, or a different one altogether. As a group, go through each of the sketches one by one, discussing the main idea of the sketch. Group your sketches or the ideas extracted from the sketches. At the end of this, you will have several different groups of ideas. Discuss each of these groups in relation to your project, their weaknesses, strengths, feasibility and originality.
- **Select and polish ideas.** From your discussion session, select the most promising ideas (at least one per team member present), discuss them
- **Re-sketching your most promising ideas:** Re-sketch the 1-3 most promising sketches neatly on a piece of paper, each student needs to sketch one!
- Your sketch needs to be understandable by others, so add legible annotations and/or provide descriptions where appropriate.
- Don't put any names on the sketch. Name the digital file to assign a project name and ask the team to choose the best options.
- Each team needs to submit one sketch by March 10, 2023.

Sketch

For your final sketch you can follow the following structure:

Layout	Title: _____ of your tool
	Date: _____ when sketch created
a version of what the final visualization will look like	Task: _____ what project are you working on
	Operations
imagine a sketched screenshot of the main interface	add a brief description or sketch of how the user interacts with the tool
Focus	Detail
focus in on something you think is particularly novel or interesting about your tool	any specific detail required for your tool
	e.g. what data do you need, what algorithms you need to write, what metrics you need to define,...

Here is an example design:

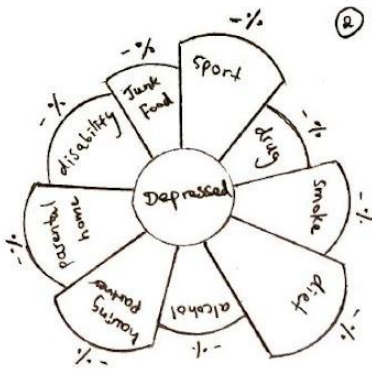


And another one:

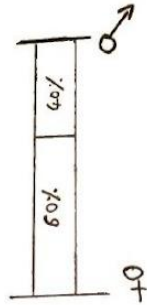
Layout

Lifestyle of depression

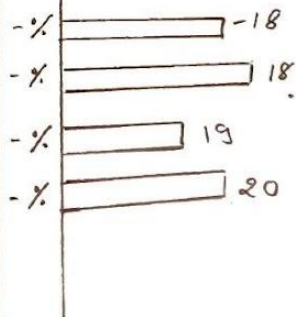
①



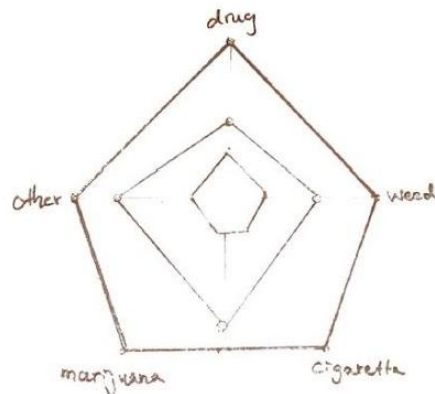
② Depressed students by gender



③ depressed students by age



④ Spider chart of drug usage



Title : Major Lifestyle of depressed students

Date : 26 - 11 - 2019

Task : Depressive Students

Operations

- ① user can hover on each activity to see more detail on (regularly, occasionally) percentage and see increasing size animation.
- ② user can hover on the bar to view the university year by percentage from those group age. so that user can know for that group age, how many percentage of student that studying in (year1, year2, year3).
- ③ user can hover on the dot point to see the percentage they also can click on it to see more details on (occasionally, regularly).

Focus

In our previous visualization assignment, we show both Life style of depressed and non-depressed, also other mental illness comparison which not seem to answer our question clearly. Thus, to answer to our question regarding to Lifestyle of depressed students, we only show the Lifestyle of depression student. And we focus on every students from the dataset which not just some specific major as previous work.

Detail

- need to convert the number of each activities into percentage and combine them together. then filter out those who has no depression. for ①
- for ②, need to get percentage of the student in each years regarding to each group age.
- for ③, need to get the percentage of each drug type usage. also filter out who does not use it.

- Your own task list and sketches, as well as a brief description of your teammates' design.
- Your team's task list and digitized final sketch by your team (you can photograph them, if you don't have a scanner at hand) as .jpg/.png images of best quality. The text and details of your sketch need to be clearly visible. **Name each image <your last name>.dotid <project team> design.jpg | png.**
 - Write down how the mapping works (think about lab 1)
- **Bonus (due on the same day, 5% credits):** [Description of how someone did a data visualization project - not too much unlike the one you will work on](#), but someone else's from the resources.

Part 2: Implementation

Your TA will post answers and ask for your votes by Sat. The winning one will be assigned and implemented (to be added based on your design.)

- CS students: D3
- Others: python

Hints: for D3, drawing something like a heatmap: <https://gist.github.com/mbostock/3202354>.

Rubric

Component	Excellent (extra 20% pts)	Satisfactory (full points)	Poor
Depth of Exploration	Several follow-up questions are asked and yielded insights that helped to more deeply explore how others do things	A good description of questions were asked by others. But may not take the analysis much deeper than the initial questions.	No description of what others do and the design choices they made.
Visualization design	Visualizations were produced, using a variety of marks and encodings explored. All design decisions were both expressive and effective.	One visualization was produced. The visual encodings chosen were largely effective and expressive, but some ineffective ones.	Several ineffective or inexpressive design choices are made.
Task list and sketches	Description of how someone did a data visualization project - not too much unlike the one you will work on, but someone else's from the resources	Submit your own task list and sketches, as well as a brief description of your teammates' design. One team submit their team's task list, mapping tablet and digitized final sketch	Lack some parts of task list and sketch
Piazza post		Post:Task; mapping table; illustration	Post part of the task, mapping table and illustration

Programming		Finish all of the design in the sketch	Finish part of design
--------------------	--	--	-----------------------