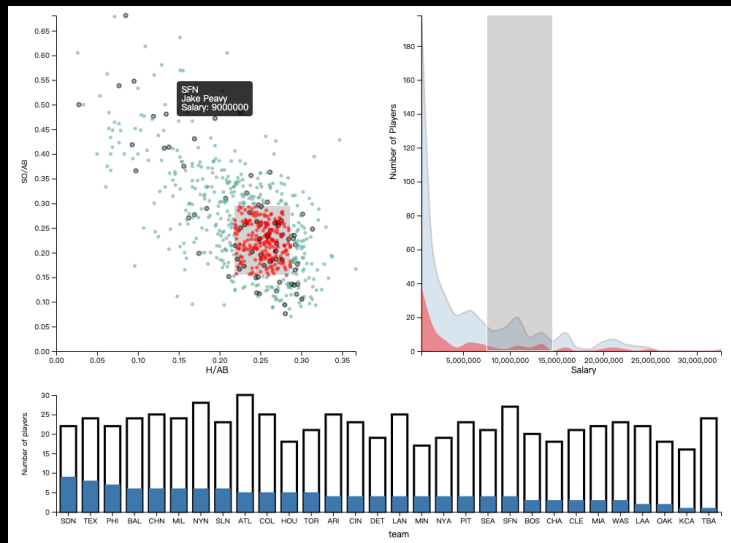# Final Project

Choices of Final Projects

- Web-based **visual analysis application** by D3

- Web-based **interactive data-driven story** by D3 (narrative visualization)

- **Scientific visualization** tool/implementation

Data Analysis Tool

- What you should have in your data analysis tool
  - Multiple views (3 or 4 views)
  - Transition (animation)
  - Link views by interaction
  - Allow users to explore data/discover knowledge/analyze data by interaction
  - Example:
    - our homework3 is an example
    - demo in omni-sci (https://www.omnisci.com/demos)
    - https://www.cs.ubc.ca/~tmm/courses/436V-20/fame/projects/20jan/actors/index.html
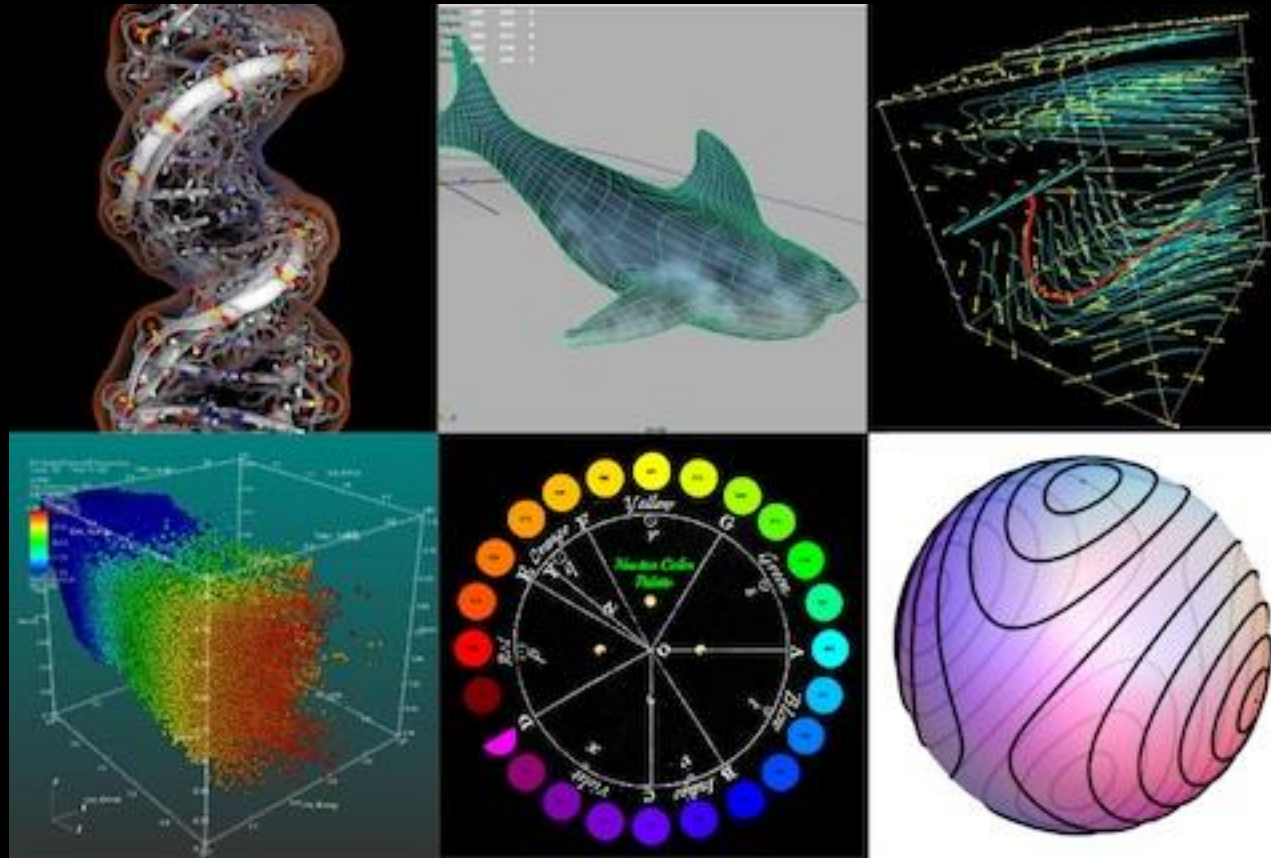
Interactive Data-Driven Story

- What you should have for interactive data-driven story
  - 3 or 4 views
  - Text along with each view to explain what people should see and understand of each view
  - Transition and interaction in some views
- More information about storytelling, https://medium.com/nightingale/from-storytelling-to-scrollytelling-a-short-introduction-and-beyond-fbda32066964
- Example:
  - http://thewaterweeat.com/
  - https://www.bloomberg.com/graphics/2015-auto-sales/
  - https://www.nytimes.com/interactive/2014/09/14/sports/baseball/jeter-swings.html
  - https://www.cs.ubc.ca/~tmm/courses/436V-20/fame/projects/20jan/disney/index.html
  - More: https://www.cs.ubc.ca/~tmm/courses/436V-20/fame/

Scientific Data Visualization Implementation

- Discuss with me if you want to do scivis and you have ideas

Suggested workflow for D3 Data analysis tool and interactive data-driven story work.

- Do not customize backend and database system.

- This work is just a vis course final project. Focus and spend your time on data understanding/visual design/visualization implementation.

- Do not spend time on system (backend-frontend connection) construction. It may cause implementation problem which does not relate to vis.

- I suggest the following workflow
  - Use python to pre-process and clean up your dataset offline and output a compact csv or json files which only contains information useful for your visualization.
  - Your web-based application just loads csv or json files to create your visualization.

Schedule of Final Project

- Team formation **5%** (due: 5/22 (Sun.))
- Proposal **30%** (due: 6/1)
- ~~Work in progress check (due: ?)~~
- Final project demo and submission **65%** (due: 6/15)
  - Online course: take videos and submit
  - Face-to-Face: presentation in classroom (6/16)

- When you submit your proposal or final project to moodle, just submit them by one team member.
- You do not have grace days for any check points of final project.
- Check points submitted after the due dates be graded with 0 points.

Team Formation

Team formation - 5% (due: 11:59PM, 5/22)

- 1~3 students per team (no more than 3)
- Email TA (60947081s@ntnu.edu.tw) your team members (names, student ID and email) with an email subject, **"DV Project Team – studentname1, studentname2, studentname3"**
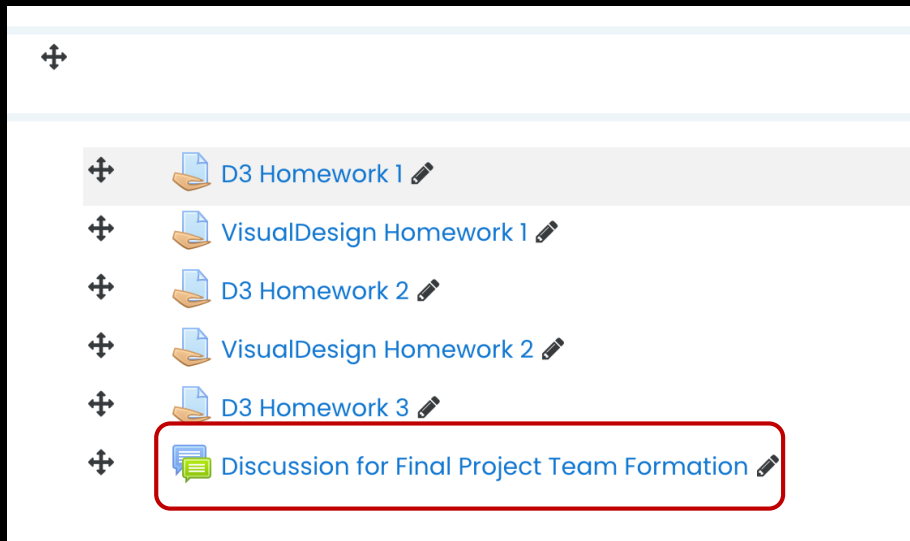
You can try to find your partner here

Proposal

Proposal – **30%** (due: 11:59PM, 6/1)

- Submit a **report** includes multiple sections

- Section 1: Basic information
  - Project title
  - A list of all team members and studentIDs and emails

- Section 2: Overview
  - A few sentences that describe what problem your visualization is tackling.
  - Example:
    - Missed medical appointments cost the healthcare system a lot of money and affects the quality of care. If we could understand what factors lead to missed appointments it may be possible to reduce their frequency. To address this challenge, I propose building a data visualization that allows health care administrators to visually explore a dataset of missed appointments. My app will use show the distribution of factors contributing to appointment show/no show and allow users to explore different aspects of this data by filtering and re-ordering on different variables in order to compare factors that contribute to absence.

Proposal – 30% (due: 11:59PM, 6/1)

- Section3: Data and Data Processing
  - You are free to choose any dataset. Here are some resources
    - https://www.kaggle.com/datasets
    - https://datasetsearch.research.google.com/
    - https://docs.google.com/spreadsheets/d/1wZhPLMCHKJvwOkP4juclhjFgqIY8fQFMemwKL2c64vk/edit#gid=0
  - Describe the dataset and the variables that you will visualize.
  - Provide a link to your data source or describe where and how to get the dataset.
  - Example:
    - I will be visualizing a dataset of approximately 300,000 missed patient appointments. Each appointment has 15 associated variables that describe the patient who made the appointment (PatientID, Gender, Age), the health status of the patient (Hypertension, Diabetes, Alcohol intake, physical disabilities), information about the appointment itself (appointment ID, appointment date), whether the patient showed up (status), and if a text message was sent to the patient about the appointment (SMSsent). Using this data I will also derive a new variable, which is the predicted probability that a patient will show up for their appointment (ProbShow).
  - If you have to preprocess or clean your data before visualize it, briefly describe what you will do and your plan.
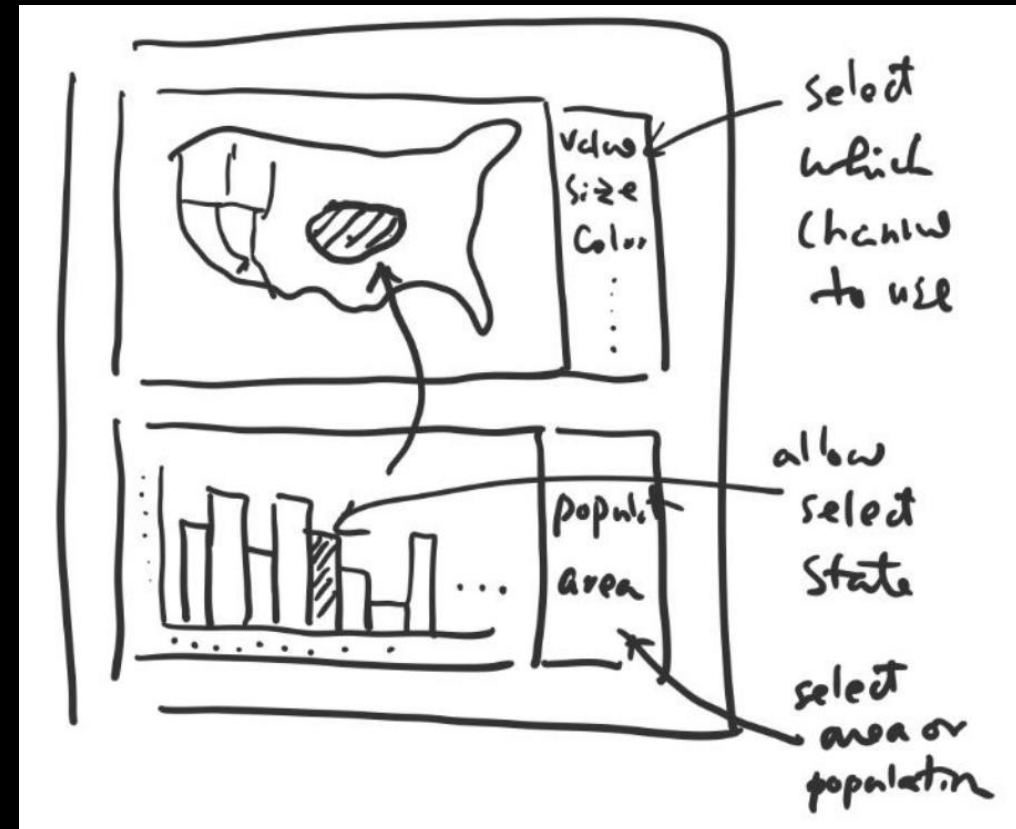
Proposal – **30%** (due: 11:59PM, 6/1)

- Section 4: Usage scenarios & tasks
  - You have to think about how someone else will use your visualization and think about their need to write this section (and design your visualization)
  - This section may include "who will use your visualization", "what data exploration or analysis tasks will be done by them", "a hypothetical walkthrough of how they accomplish the tasks with your visualization"
  - Example:
    - Mary is a policy maker with the Canadian Ministry of Health and she wants to understand what factors lead to missed appointments in order to devise an intervention that improves attendance numbers. She wants to be able to [explore] a dataset in order to [compare] the effect of different variables on absenteeism and [identify] the most relevant variables around which to frame her intervention policy. When Mary logs on to the "Missed Appointments app", she will see an overview of all the available variables in her dataset, according to the number of people that did or did not show up to their medical appointment. She can filter out variables for head-to-head comparisons, and/or rank order patients according to their predicted probability of missing an appointment. When she does so, Mary may notice that "physical disability" appears to be a strong predictor missing appointments, and in fact patients with a physical disability also have the largest number of missed appointments. She hypothesizes that patients with a physical disability could be having a hard time finding transportation to their appointments, and decides she needs to conduct a follow-on study since transportation information is not captured in her current dataset.

Proposal – **30%** (due: 11:59PM, 6/1)

- Section 5: Visualization Design & Sketch
  - From your scenario and tasks, design your visualization and interface.
  - Include and emphasize the visual designs and interaction features you need to accomplish the tasks
  - Your sketch can be hand-drawn (take picture and paste on the report), or mocked up by powerpoint, graphics editor…….
  - Your sketch could be multiple images to describe your design and plan
  - Note: this basic sketch is going to help you to think what you are going to do and your first prototype implementation. It is by no means about what you should submit in the end.

Example of one sketch image:

Proposal – **30%** (due: 11:59PM, 6/1)

- Section 6: Work breakdown and schedule
  - Subdivide your project into many small features (milestones)
    - Each feature should associate with an expected working hour
  - Describe the expected date to accomplish each feature (6/2, 6/9, 6/16…)

Final Project Demo and Submission

Final project demo and submission - **65%** (due: 6/15)

- Time: in class on 6/15 or submit the demo video on 6/15
- If in-class demonstration,
  - I will announce a google spreadsheet (later) for your team to sign up for order of demonstration
    - Sign up before 6/13
- Each team has 5-10 minutes to introduce and demonstrate your project
- Presentation
  - Prepare slides to introduce (3 ~4 minutes)
    - Your project title and team members
    - Introduce the dataset you use with some details
    - Introduce your usage scenario and tasks
    - Introduce what you have done for your project (overview of your visualization/visual design/data processing/ algorithm you use….)
  - Live demonstration (4 ~ 5 minutes)
  - Q & A (in-class)

Final project demo and submission - 65% (due: 6/15)

- Submit your final project on moddle before 11:59PM 6/15. Otherwise, your will receive 0 of this part. What you submit should include
  - All of your code
  - Dataset
  - A txt file about the instruction to run your code
  - Slides
  - **A short video to demonstrate your project (ex: screen record video with explanation of your work)**
  - **A text file to describe what each team member did (in terms of project features)**
  - **If we have an on-line class,**
    - **Submit a video for your presentation**

- If your final project is too large to upload, please contact with TA right after our demonstration.

Grading Rubric

- Visual design:
  - effective of visualization and interaction
  - It links back to your scenarios and tasks well or not.
- Level of technical difficulty
- Quality of implementation
- Quality of your presentation/demonstration
- Time control of your presentation/demonstration
- Teamwork