

Data Visualization (Syllabus)

csc 9005

王科植 Ko-Chih Wang

- Lab: Visual Data Analysis Lab (ViDA)
 - Visual data analysis (Large-scale data visualization)
 - <https://sites.google.com/view/vida-ntnu>



National Taiwan Normal University Visual Data Analysis Lab.

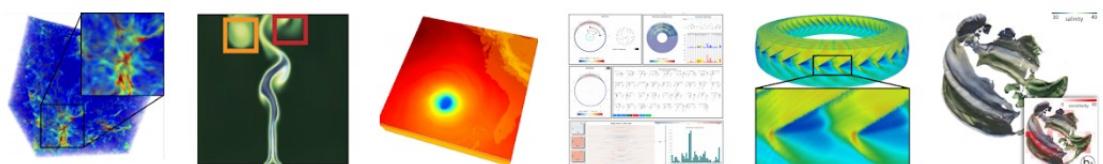
Home People Research Publication

VIDA

Visual Data Analysis Laboratory

News:

- Our laboratory is looking for new talented students. [\[more\]](#) (2019/07/11)

The screenshot shows the homepage of the ViDA website. The header includes the university logo and name, and navigation links for Home, People, Research, and Publication. Below the header is a large banner with the word "VIDA" in white. Underneath is a sub-header "Visual Data Analysis Laboratory". A "News" section contains a single item about recruiting students. The main content area features several small images illustrating research topics like data visualization and machine learning.

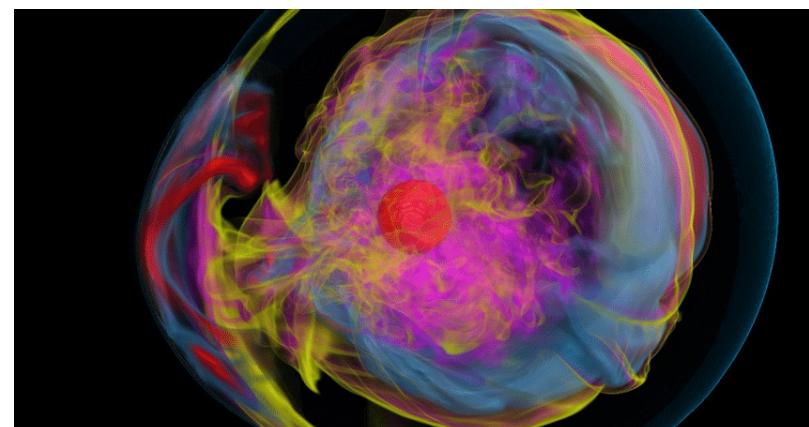
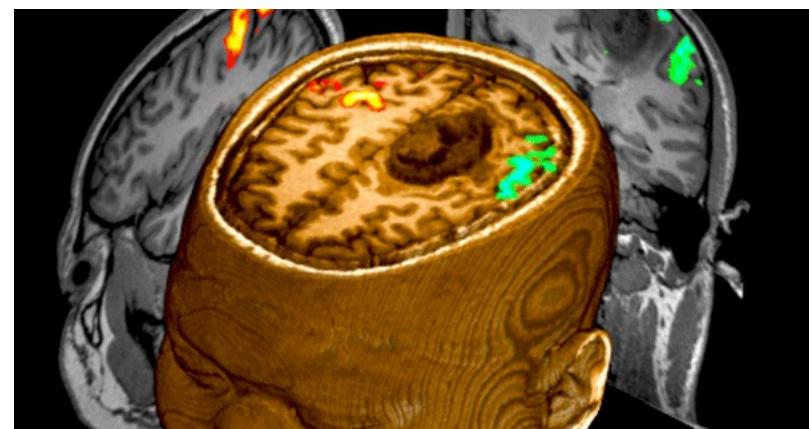
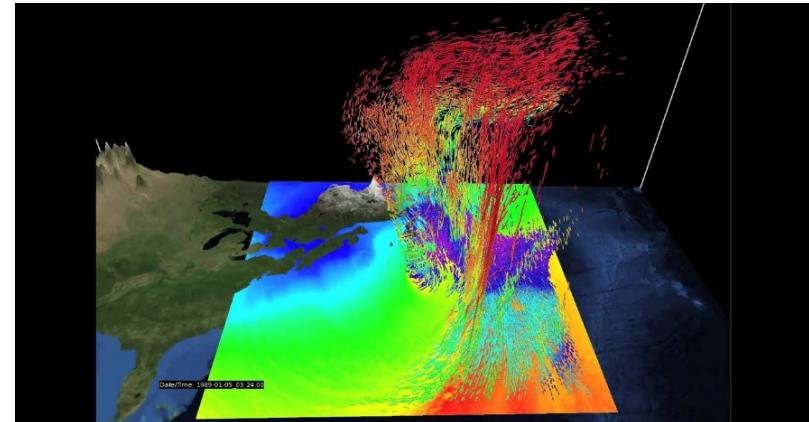
Visual Data Analysis (VIDA) Laboratory is a research group at [Department of Computer Science and Information Engineering, National Taiwan Normal University](#) (NTNU) and led by [Ko-Chih Wang](#) (王科植). The main research directions of ViDA are **large-scale data analysis & visualization**, **high-performance computing**, **computer graphics**, and **machine learning**. We conducts cutting-edge research in data visualization for scientific and information data analysis. Visualization research is at an intersection of data analysis, computer graphics and large-scale data handling, and has been playing increasingly important role in many applications. Our research has two main branches. One is scientific data visualization and the other one is information data visualization. Scientific data visualization focuses on visualizing data from scientific simulation and solving the large-scale data problem in computational science. Information data visualization builds visual tools to facilitate applications such as social network, machine learning, questionnaire, transaction, and traffic data analysis. For more details of ViDA's research, please check the [Research page](#).

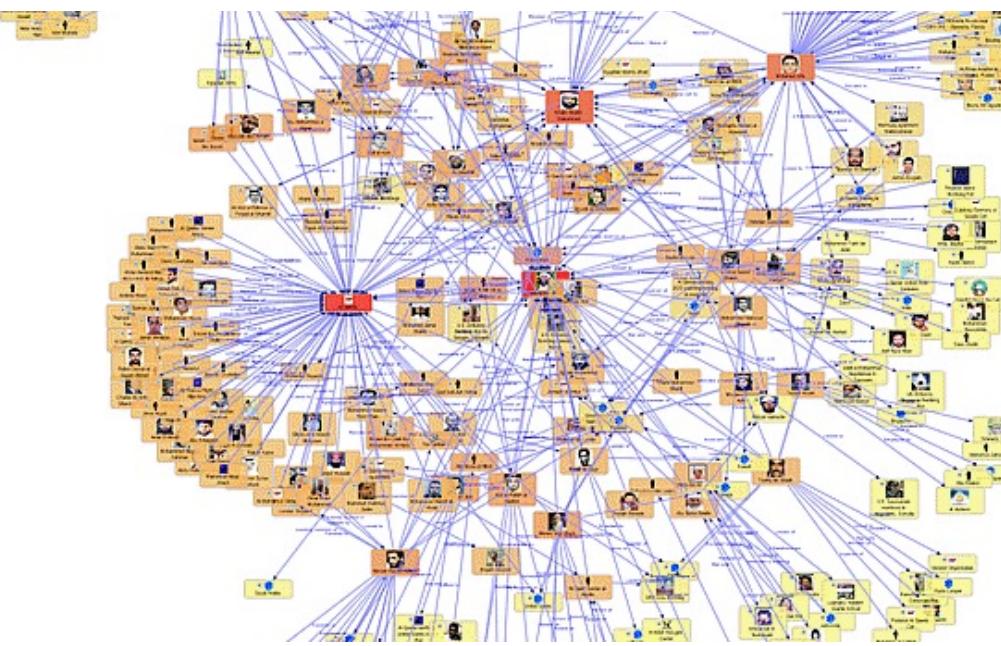
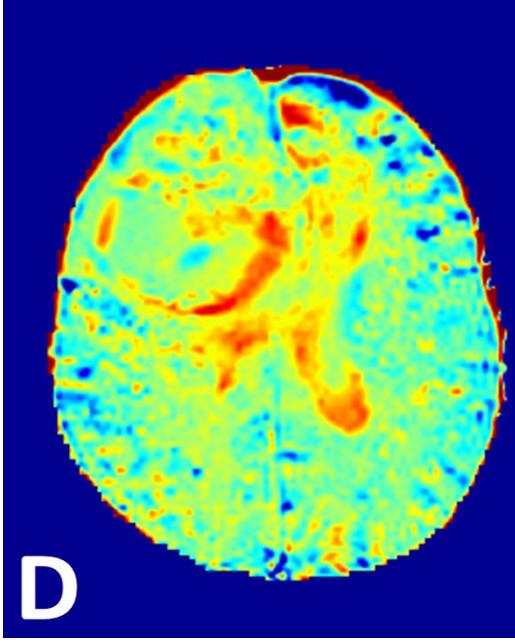
EMI (English Medium Instruction) Course

- All activities in English?
 - Tradeoff between learning our domain knowledge and developing English ability
- English
 - Lecture
 - Turn in your assignments or the project report
 - Project presentation?? (TBD)
- Could be in Mandarin Chinese
 - Students ask questions in class
 - Office hour
 - Talk with me outside of the class/during the break
 - Exam
- I also offer course videos (Mandarin) on youtube and post on moddle.
- If you are not a native mandarin speaker, we can always use English.

Data Visualization

- Visualization emerged as a field of study on about 35 years ago.
- The field was driven by supercomputing for scientific studies at the beginning. Later, data driven problem solving and decision making carried out in all fields of study boosts the development of visualization.
- The field grows mainly in U.S. slowly first, and then explode:
 - IEEE VIS (1990)
 - EuroVis (2005)
 - PacificVis (2008)
 - ChinaVis (2014)
- The growth of Visualization programs in academia continues.
- Commercial and open-source visualization software tools & libraries are widely available





CAPTURE



PROCESS



STORE



ANALYZE



USE

Data Visualization

- A process of transform “data” to “image”
- Purpose of Data visualization
 - Amplify human’s cognition ability
 - Integrate with domain knowledge to analyze the data
 - To explore, analyze or present the data
- Last mile in data analytics pipeline
 - If the data is huge or the result of data analysis is still complicated, how to let human summarizes information and. Make the decision?
 - Visual representation

Data Visualization

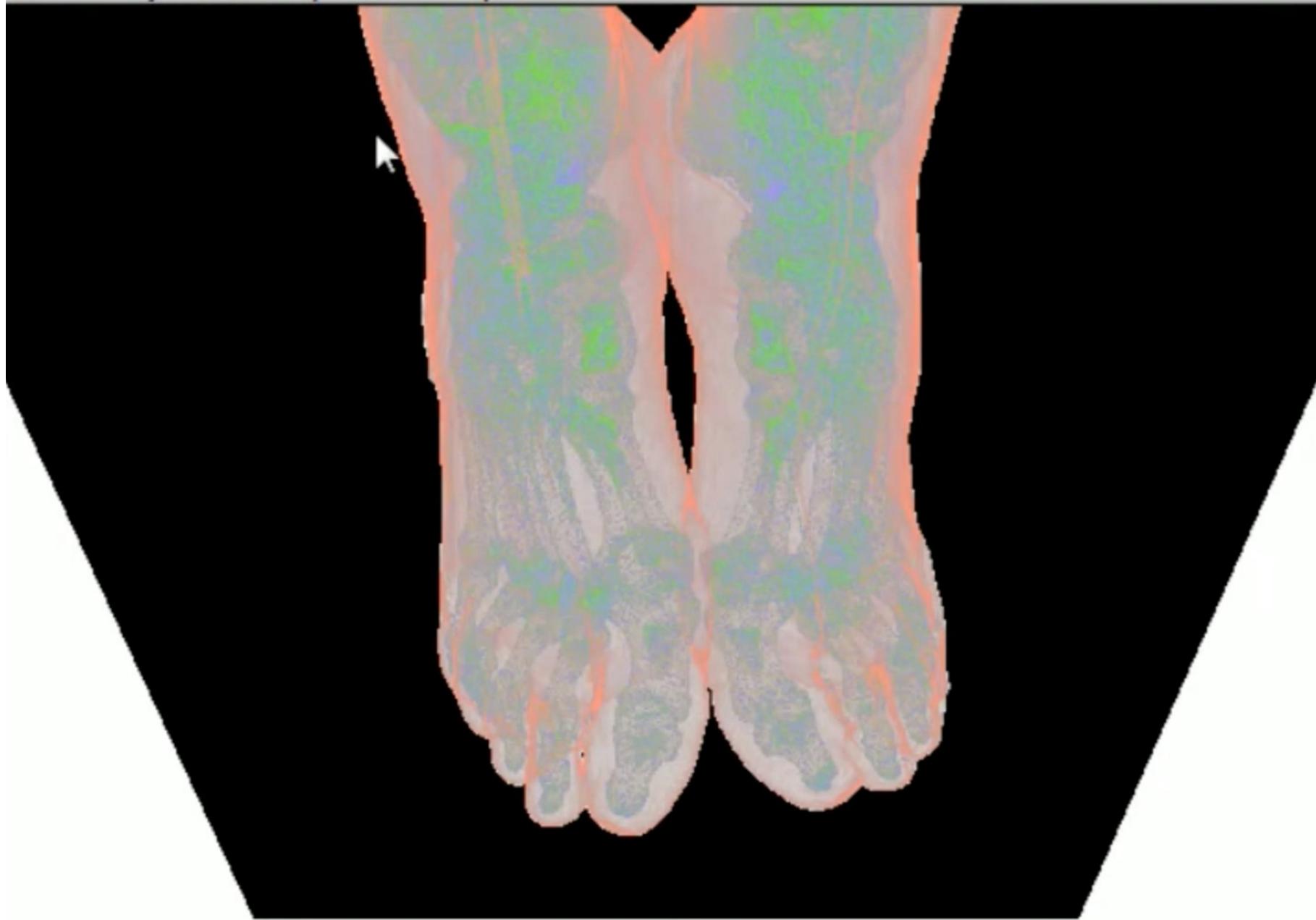
- Scientific Visualization
- Information Visualization, Visual Analytics

Scientific Visualization

Scientific Discovery

- https://www.youtube.com/watch?v=95z0qRNFFxs&ab_channel=LosAlamosNationalLaboratory

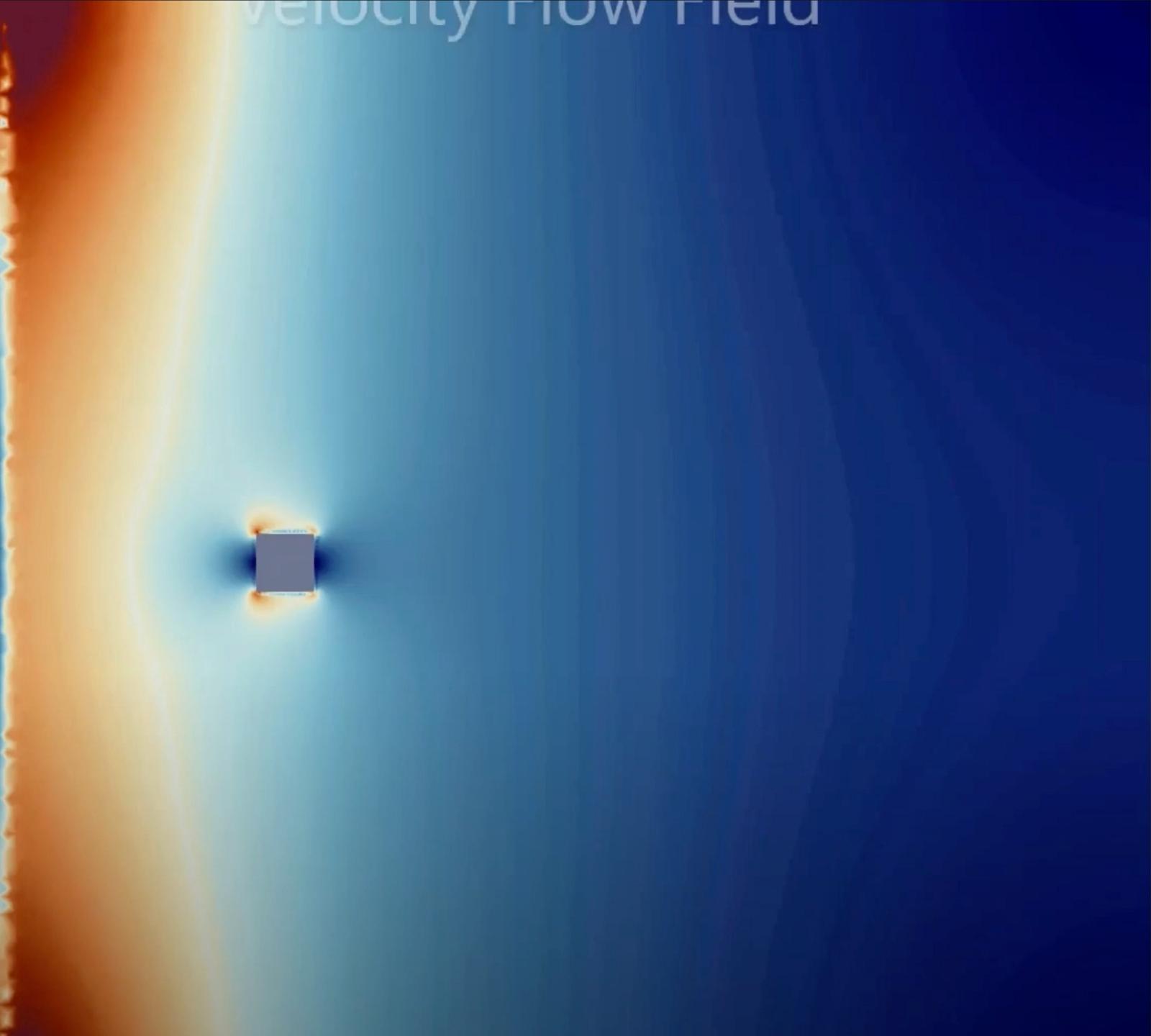
File Ray Trace Spline Sharpen View



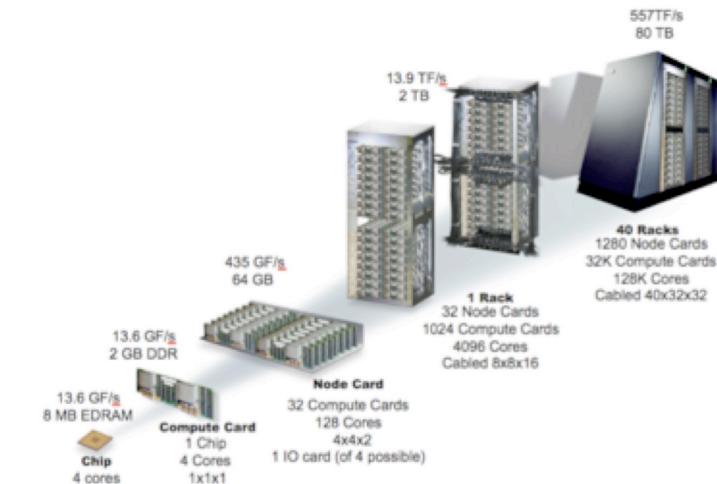
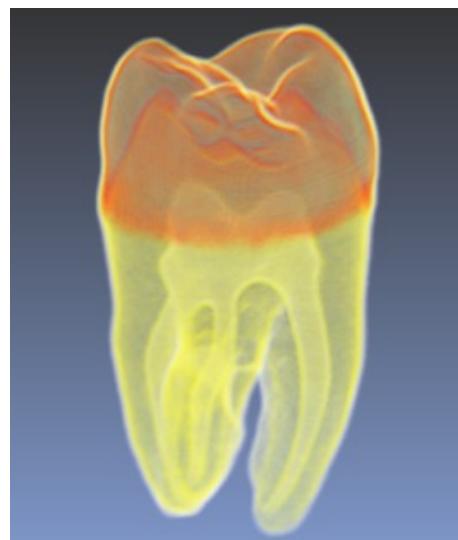
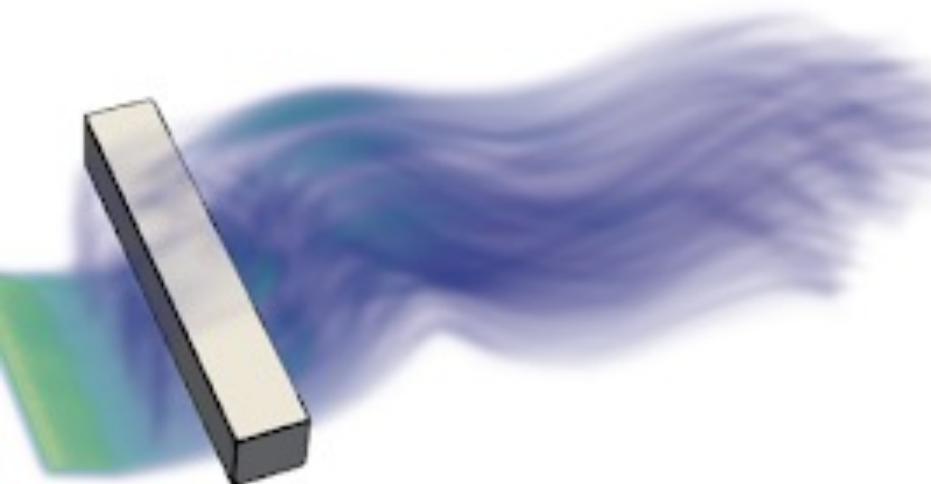
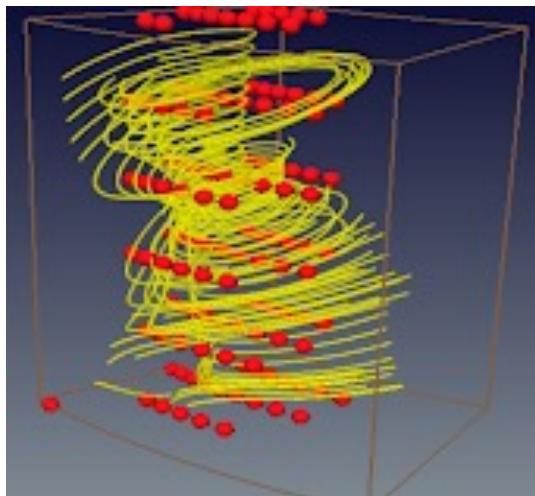
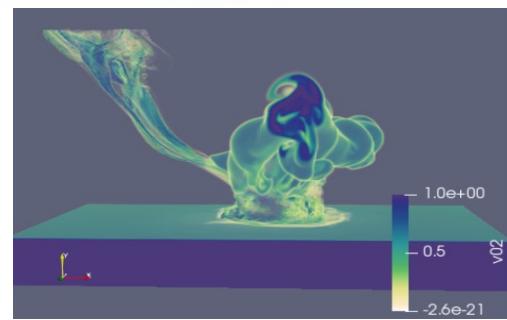
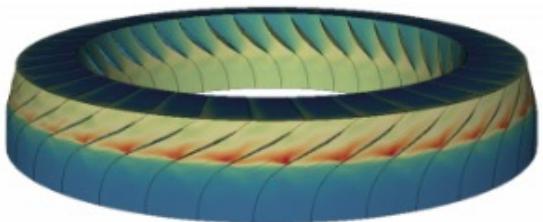


Yx
停 (k)
Z

Velocity Flow Field

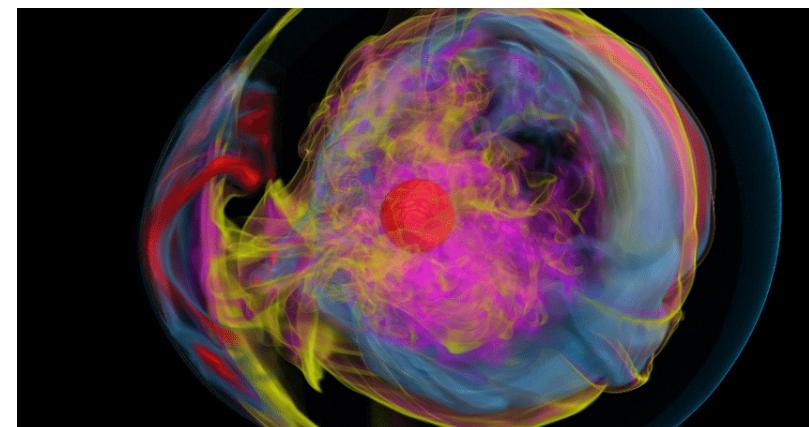
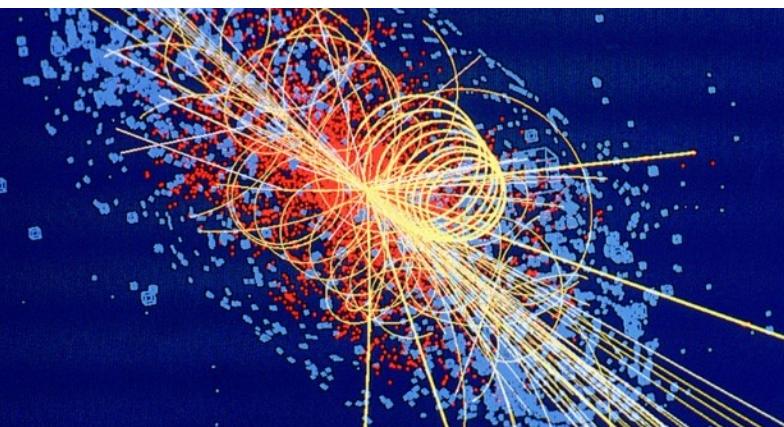
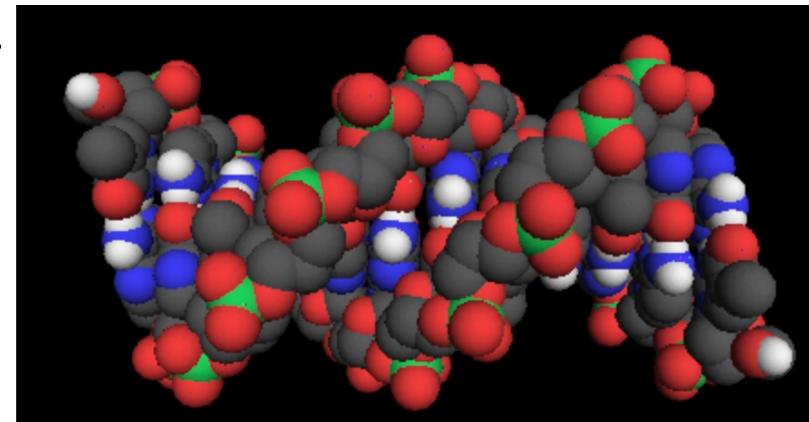
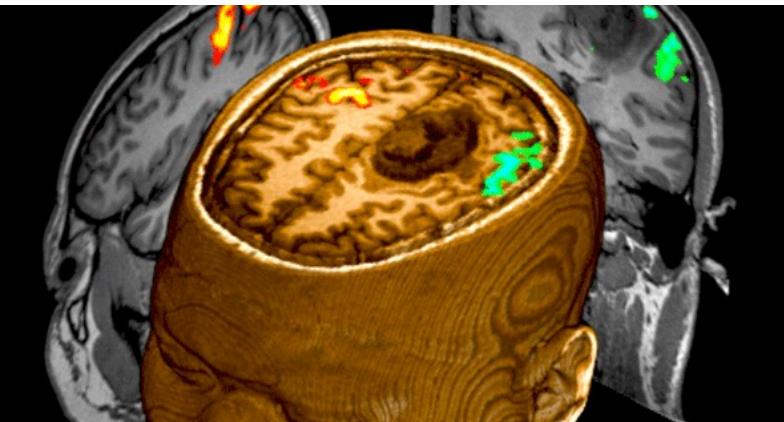
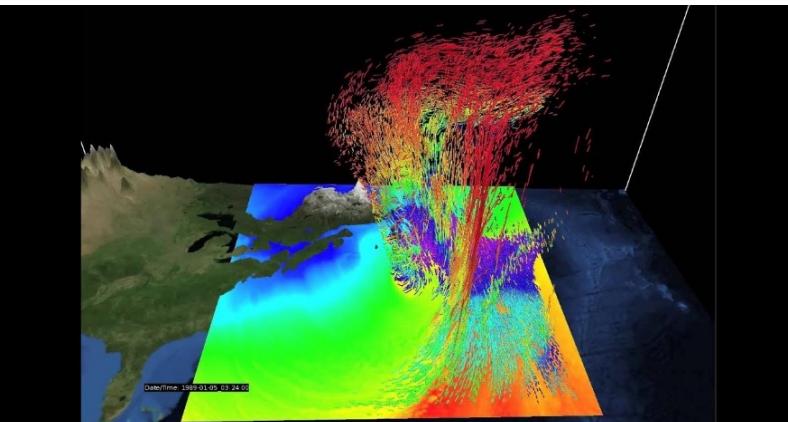


Scientific Visualization



Scientific Visualization

- Visualization for scientific data
- Scientific data: phenomenon of our real world



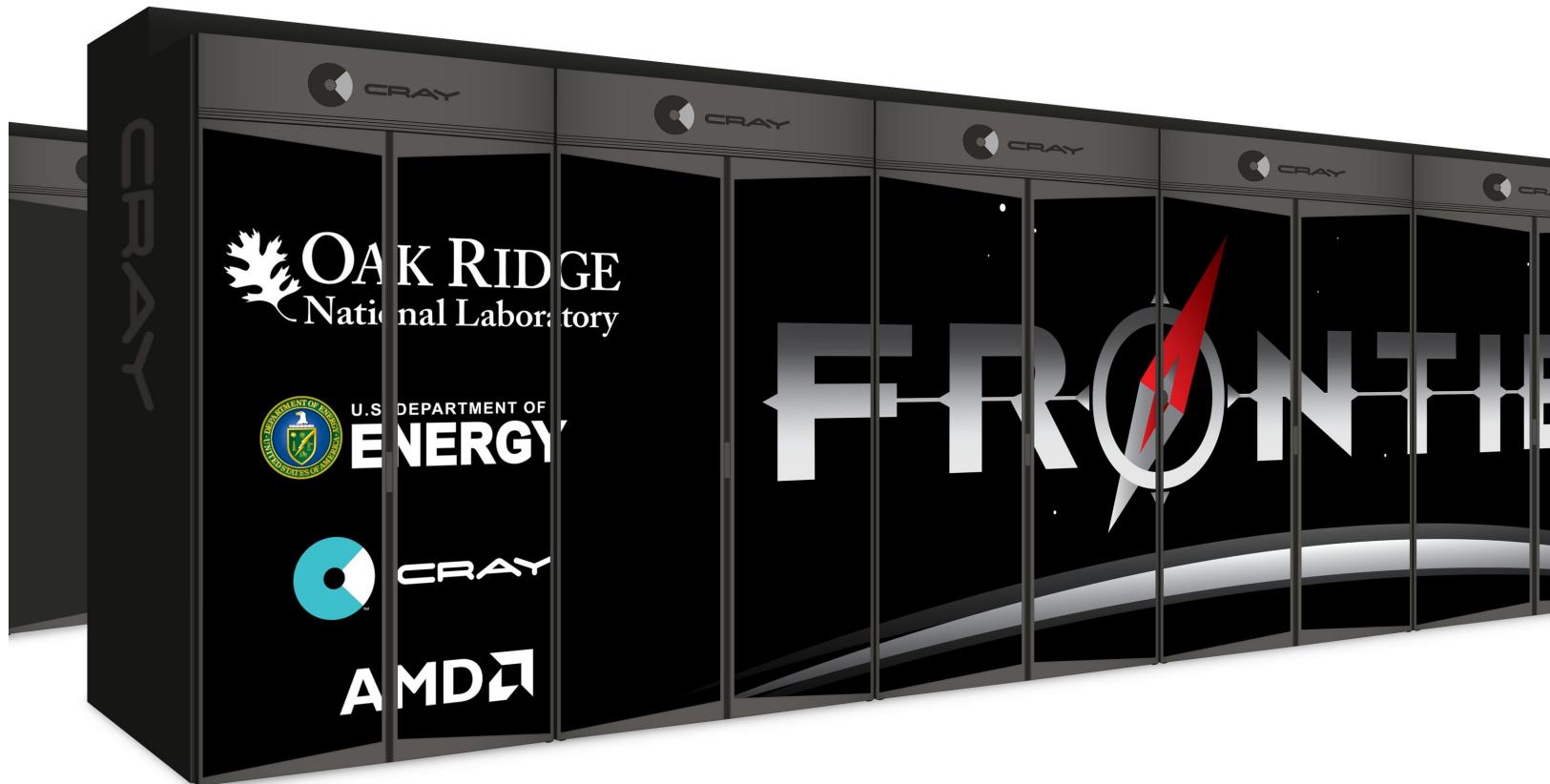
Scientific Studies using Simulations

- <https://avl.ncsa.illinois.edu/>

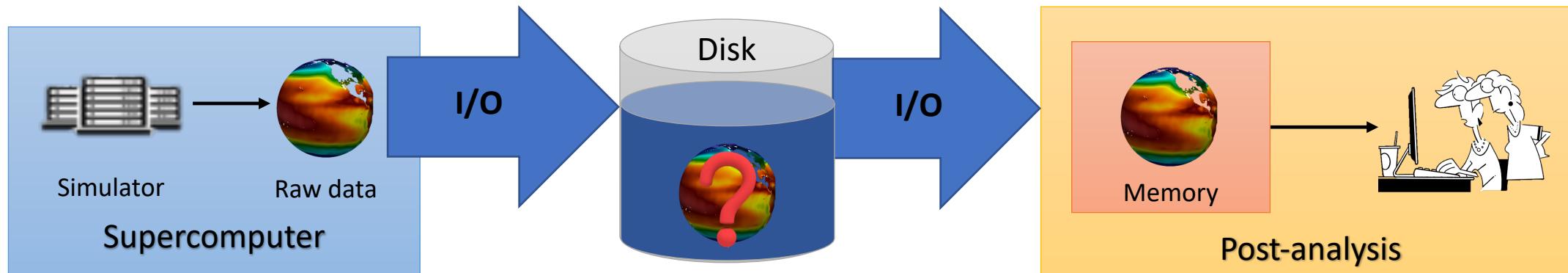
Interactive 3D Visualization

- https://www.youtube.com/watch?v=tl_sJuA2LWg&ab_channel=AssociationforComputingMachinery%28ACM%29

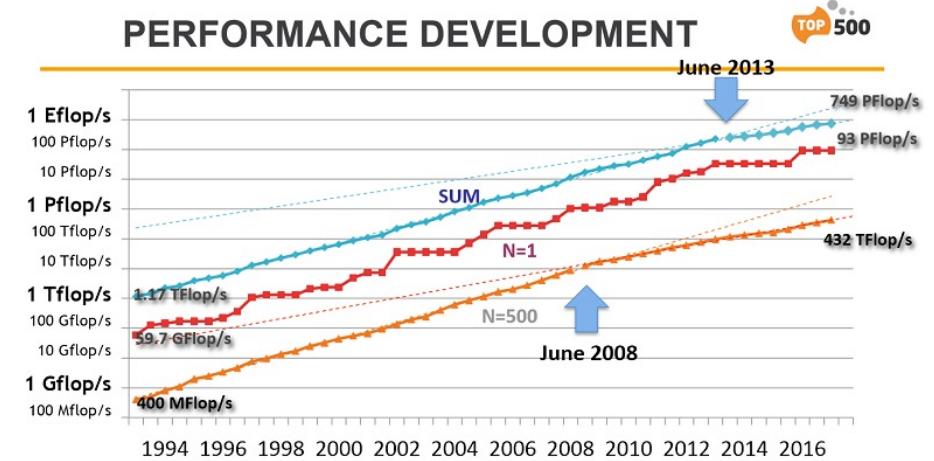
Scientific Studies using Simulations



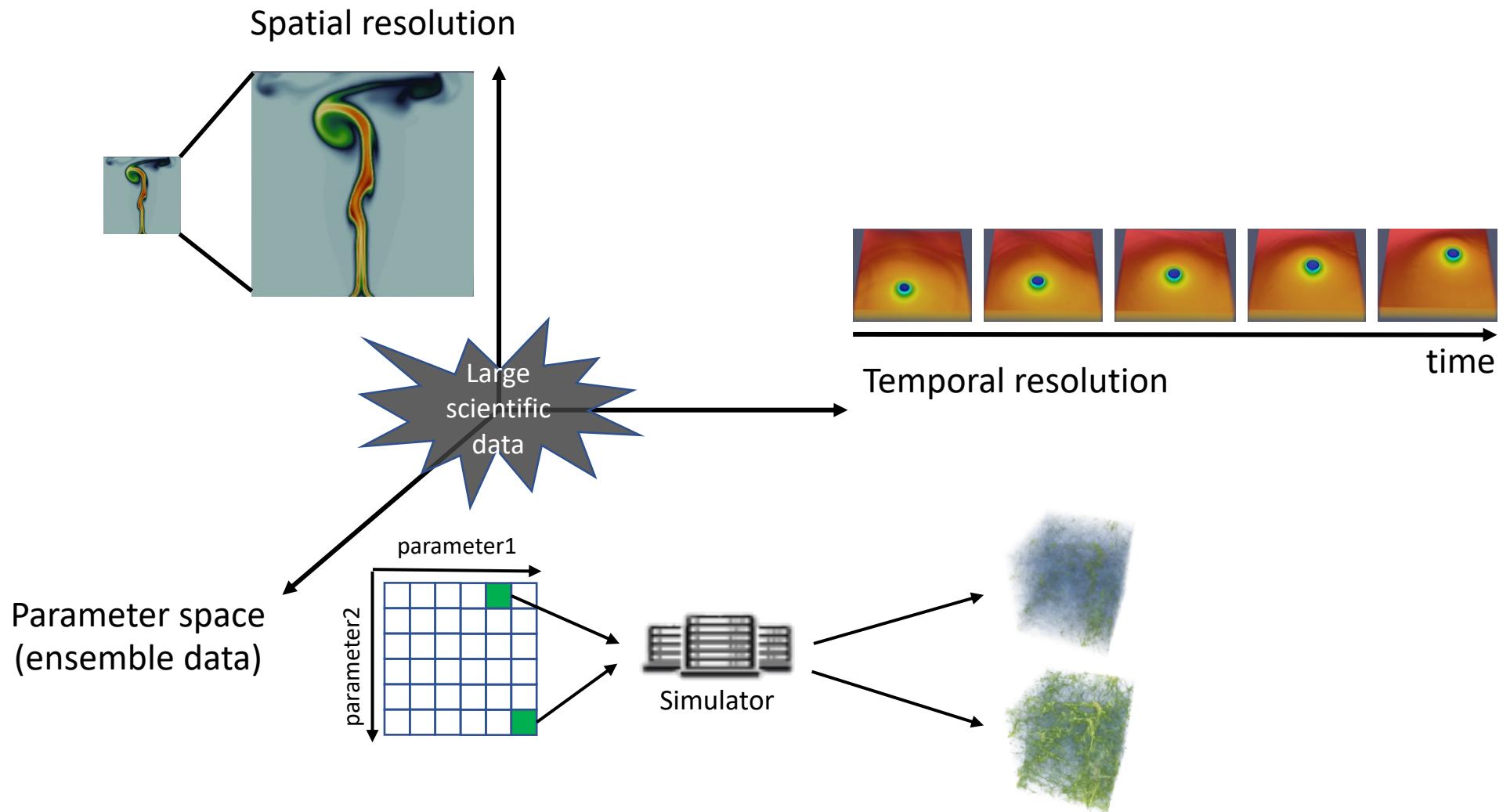
Scientific Studies using Simulations



- Data size is often large
- Long I/O time
- Hard to keep all datasets for long time
- Usually, batch processing
- Difficult for Interactive exploration



Large Scientific Datasets

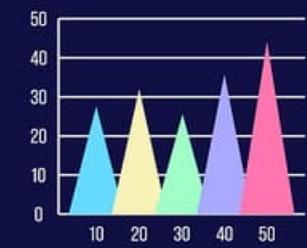
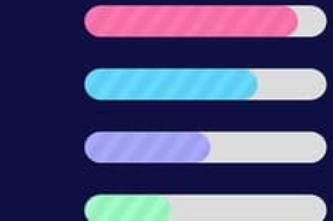
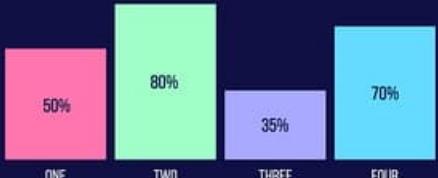
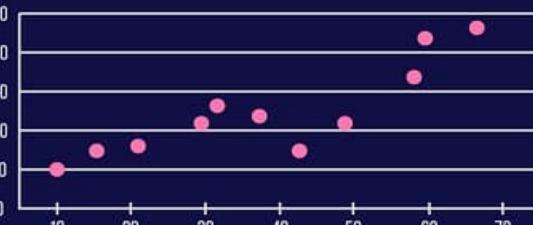
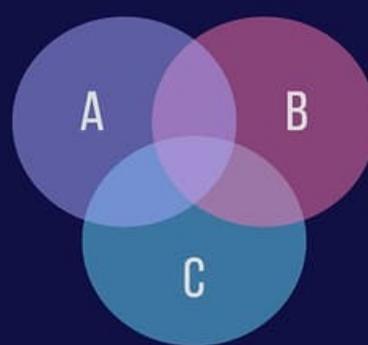


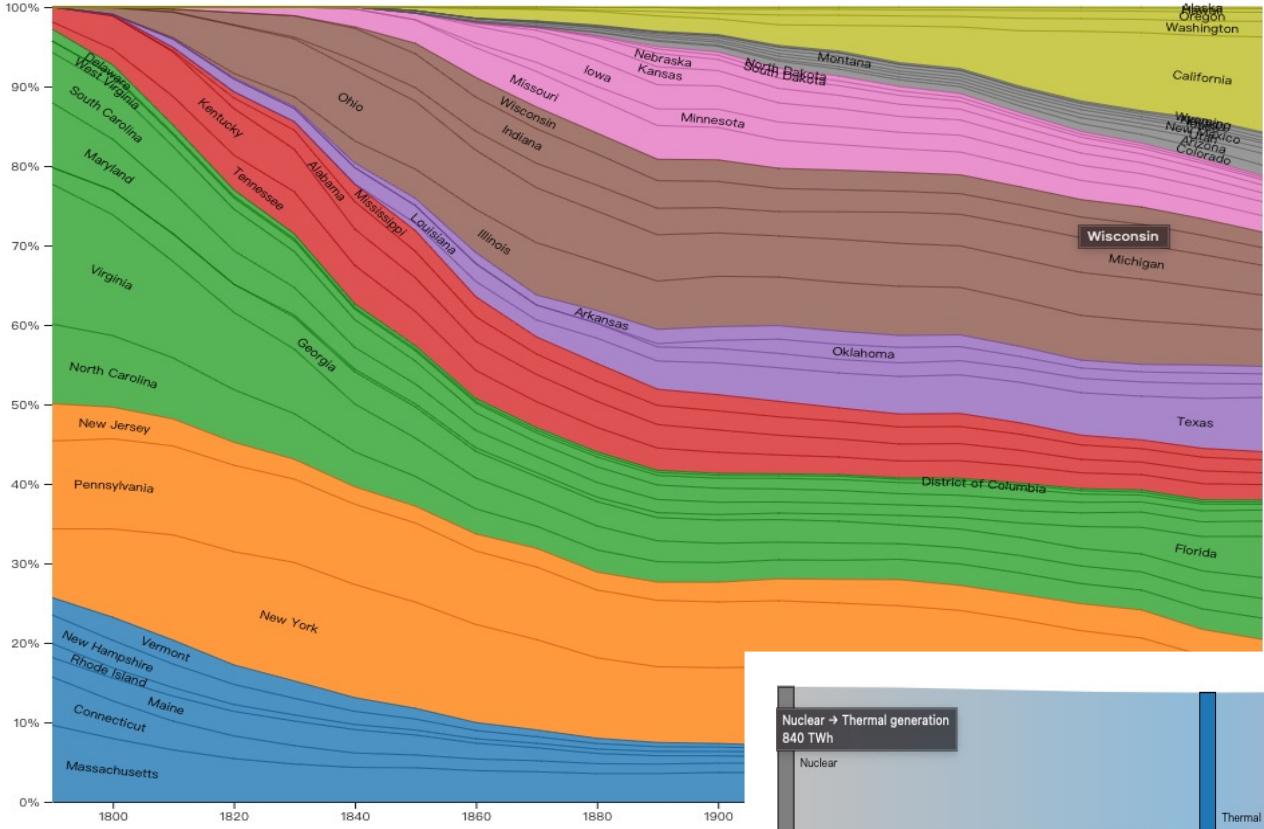
Information Visualization & Visual Analytics

Hans Rosling

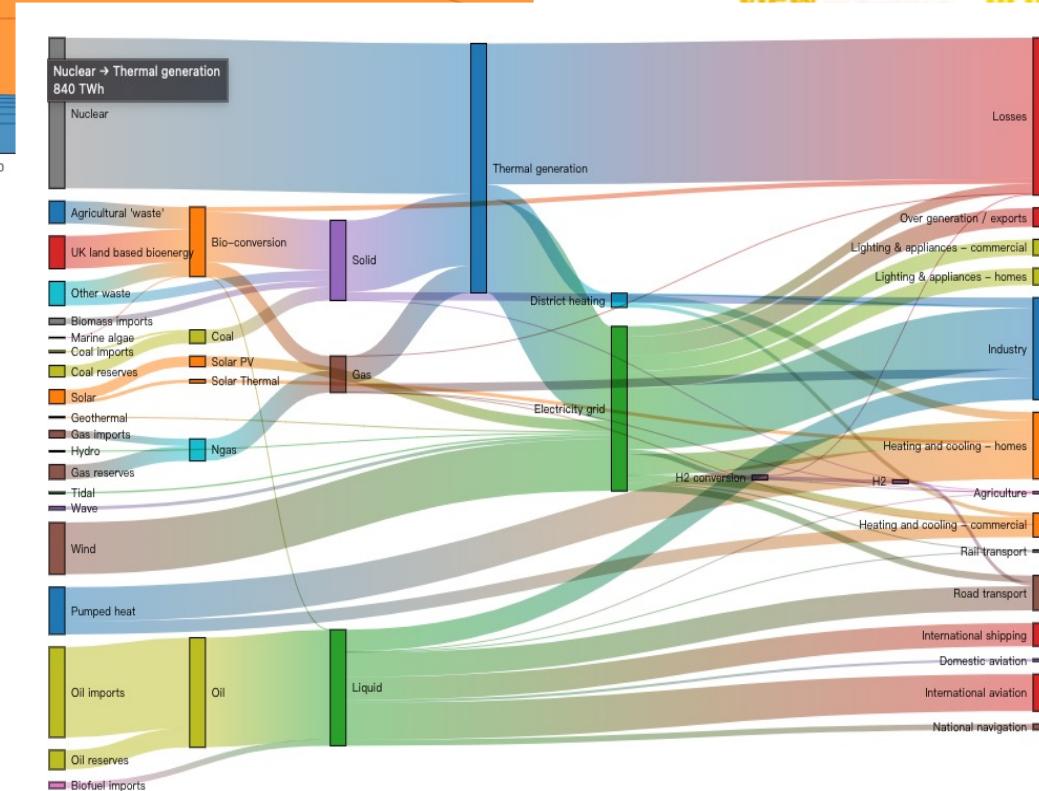
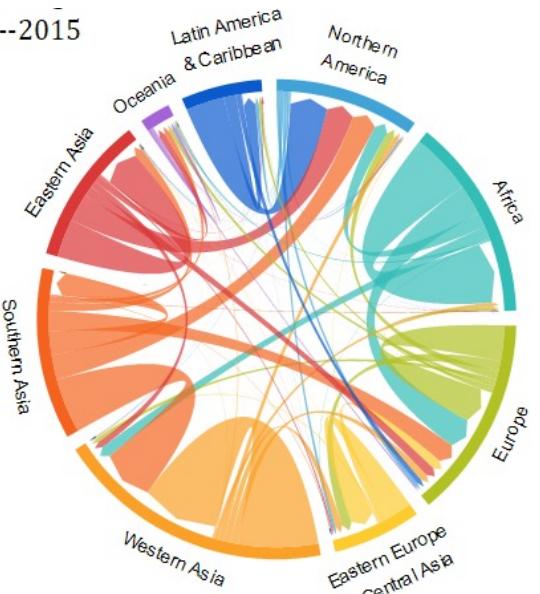
- <https://www.youtube.com/watch?v=jbkSRLYSoj0>







2010-2015



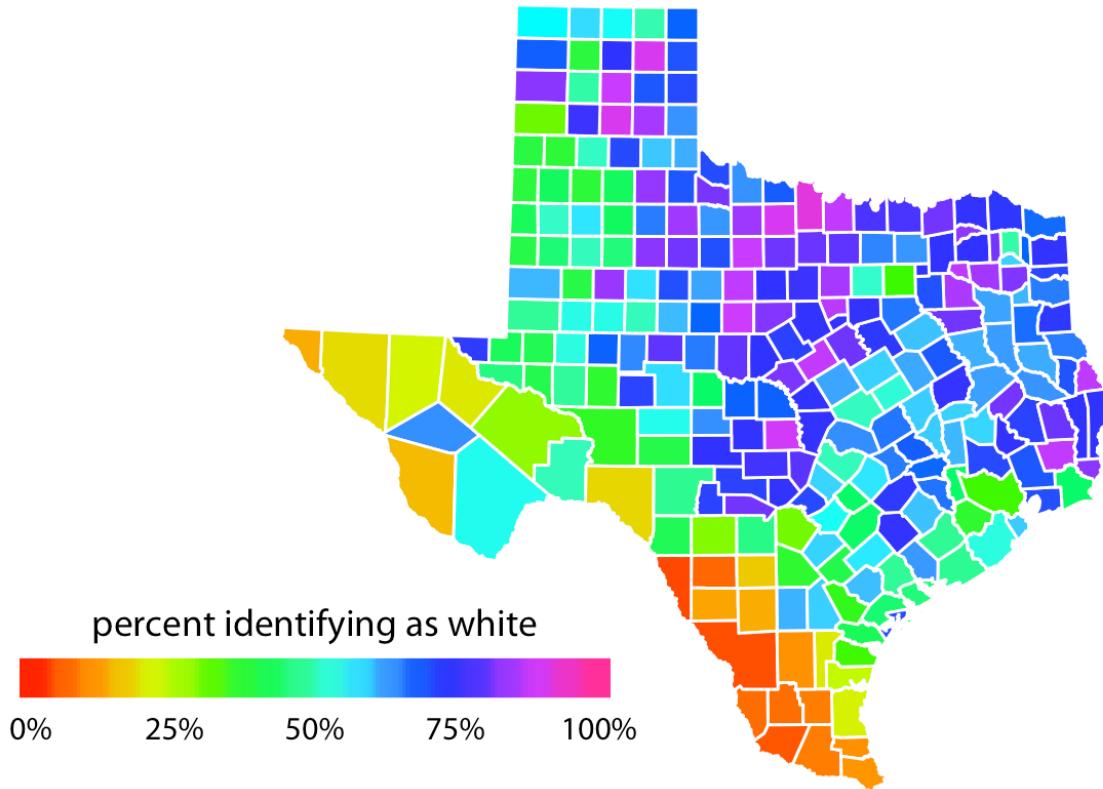
tie coat childish
auction interest home count
countess grabern andrew letter
hand smile leave
maneuver account
holiday face perplexity
intrigue connection
garter ending surroundings
moment possession
erision princess
rears unconquerable
beginning proposal
trouser spring consideration
constraint knowledge property
knee breast steward
view brother intercourse
promise waistcoat
society
husband nataša's serf
mother nicholas estate nski room
sha foot

embarrassment

iS Car Sales 2017



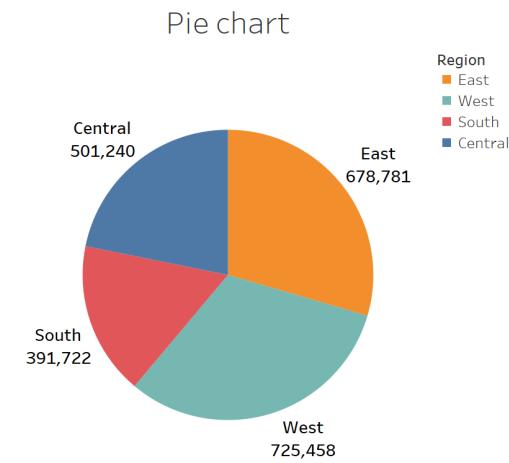
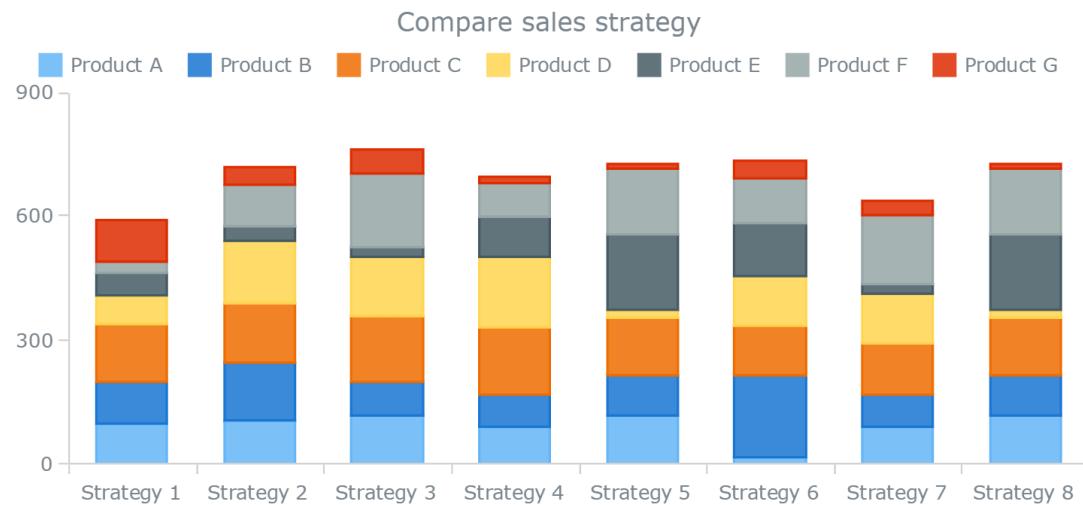
Color Choice



NOT IDEAL

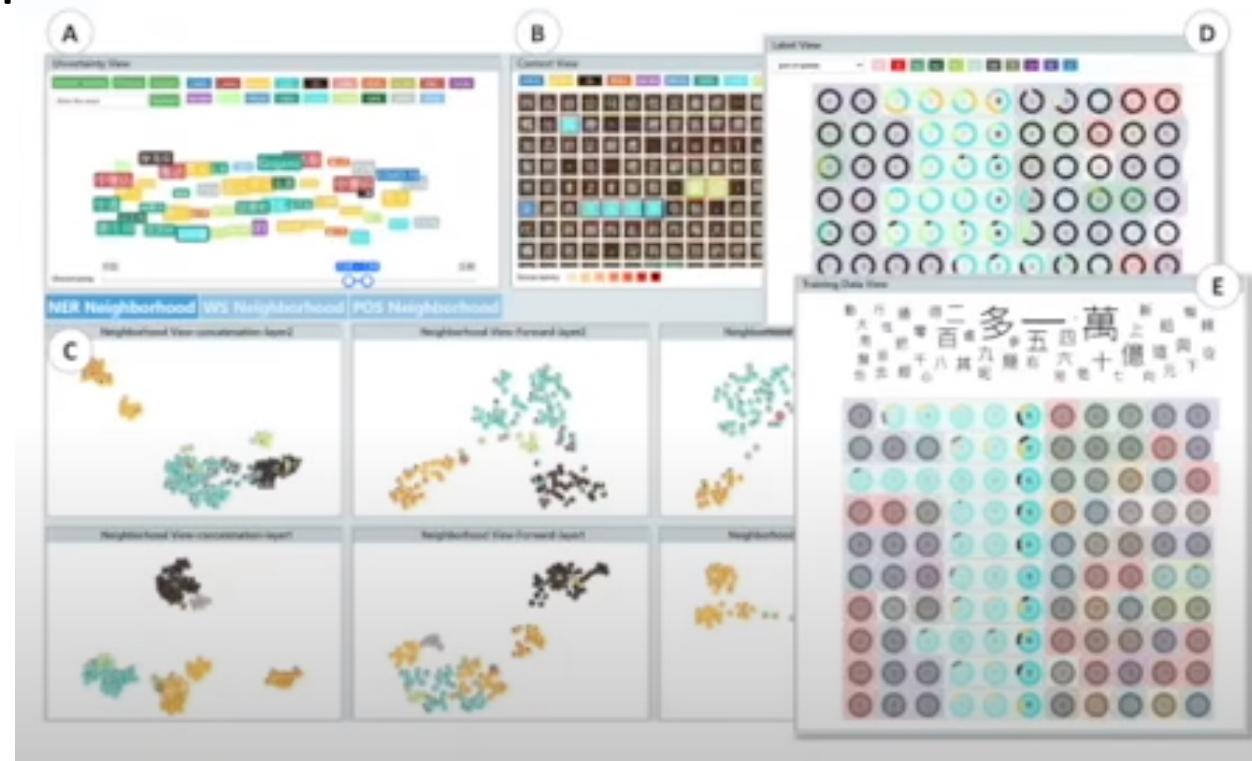
Visualization Choice

- Bar chart or pie chart



Visual Analytics Tool

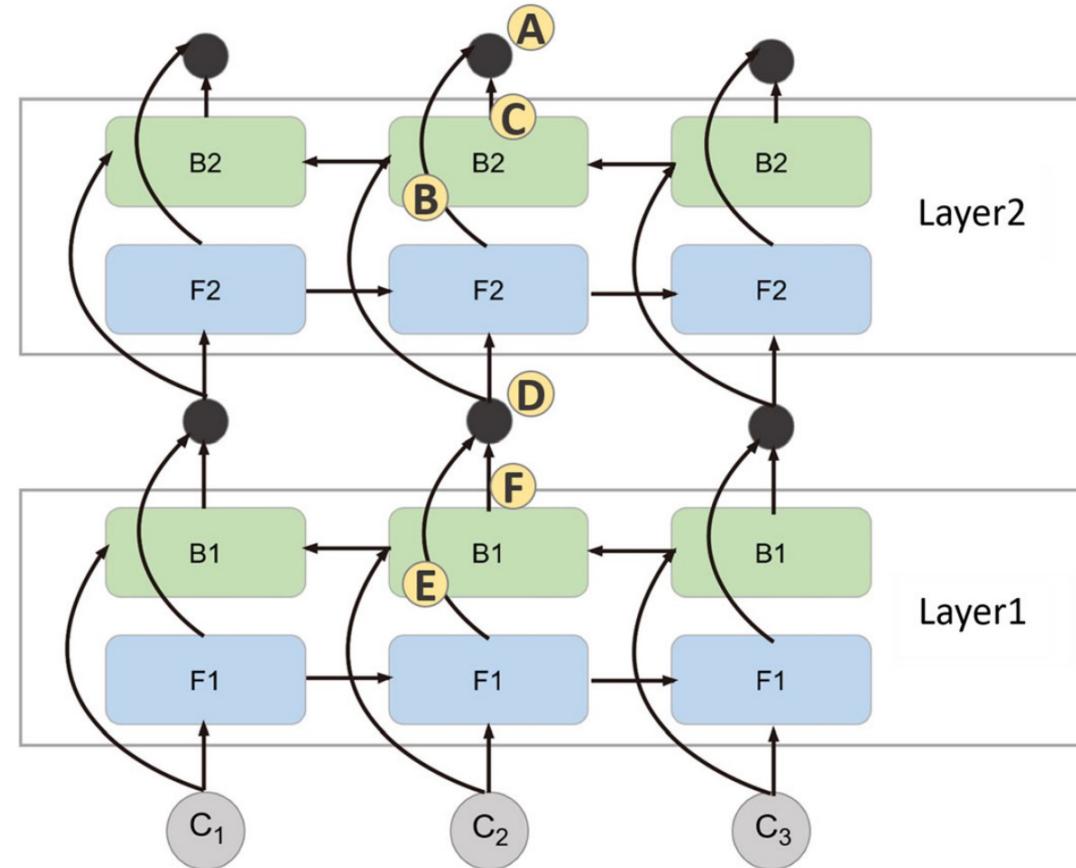
- <https://youtu.be/WKVZqHfyP0k>
- CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition



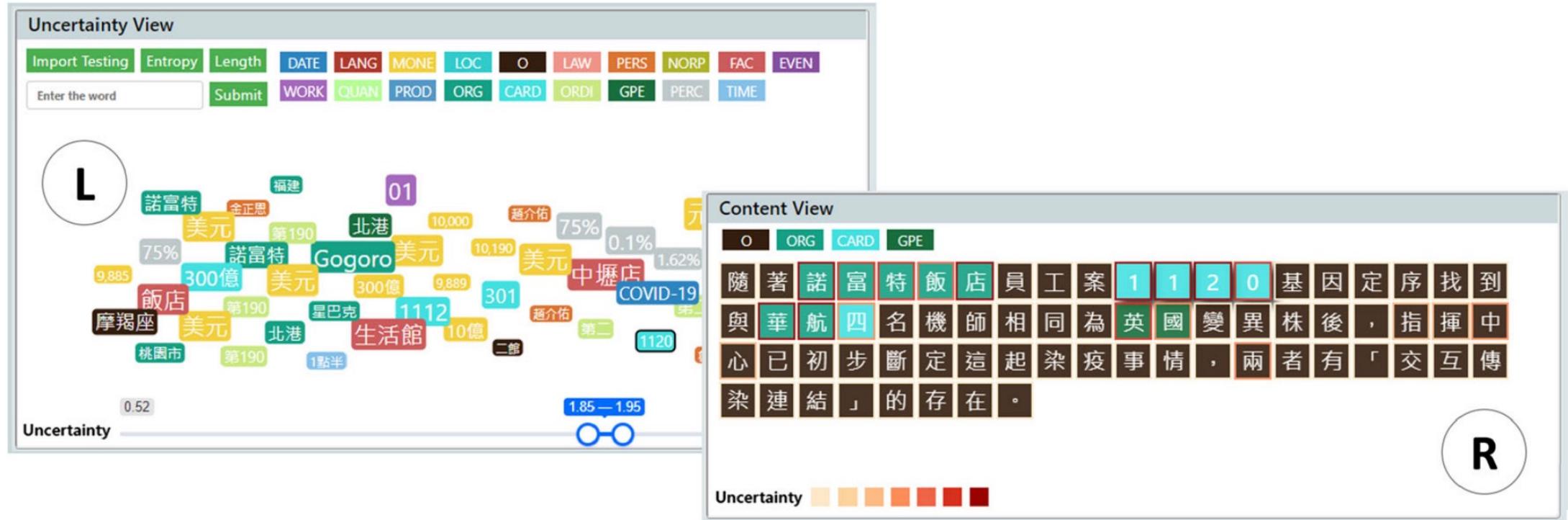
CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition

- NER: name entity
 - 布希(Bush) -> name
 - 聯合國(United Nations) -> organization
- POS: part-of-speech
 - 布希(Bush) -> noun
 - 跑步(run) -> verb
- WS: word segmentation
 - 這是一隻狗 -> 這 是 一隻 狗
 - This is a dog

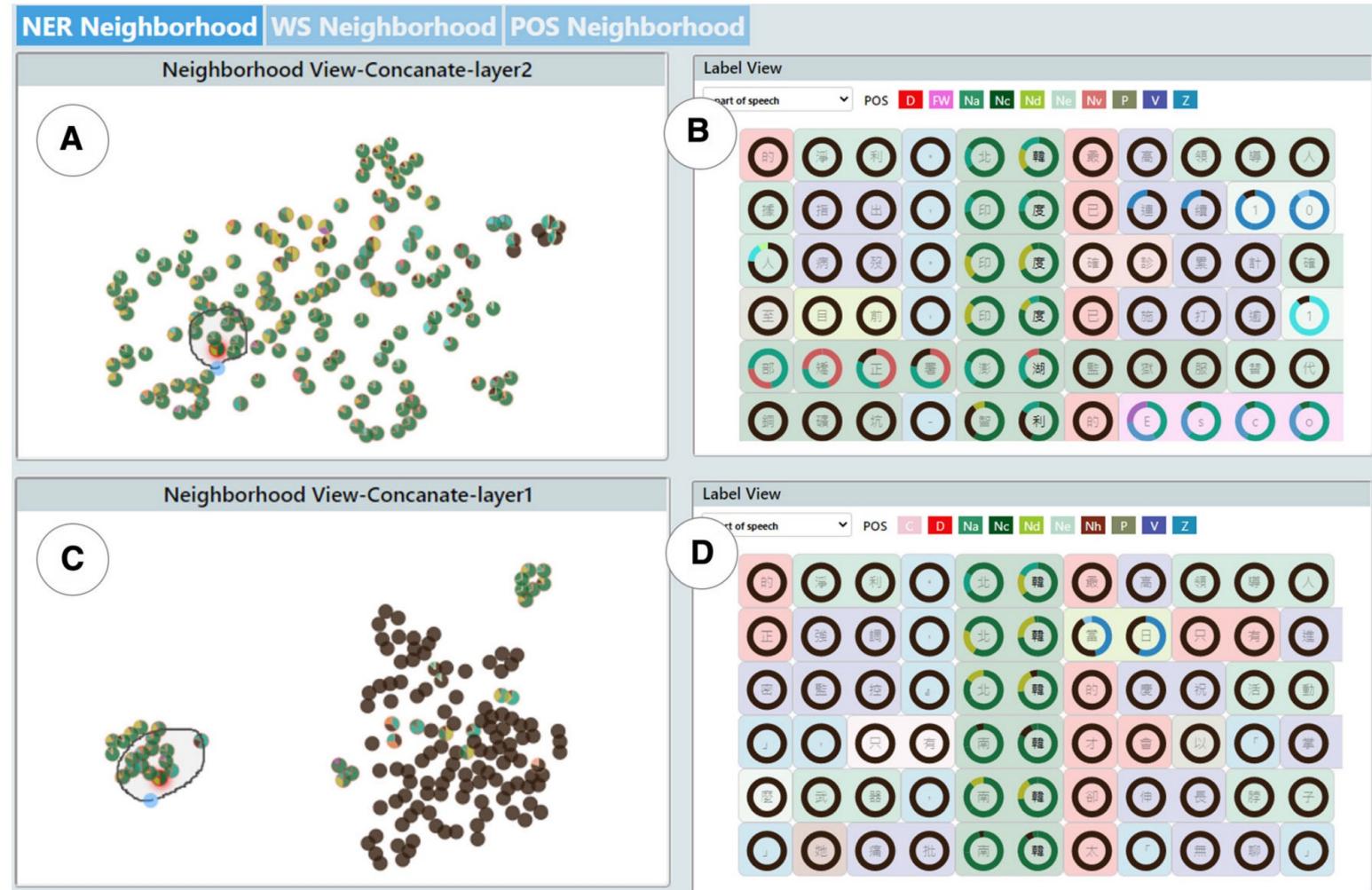
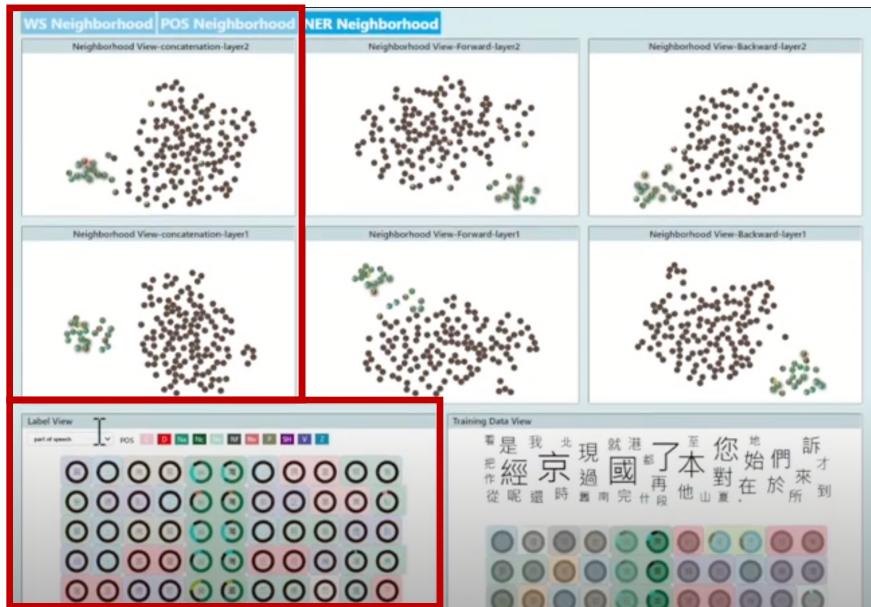
CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition



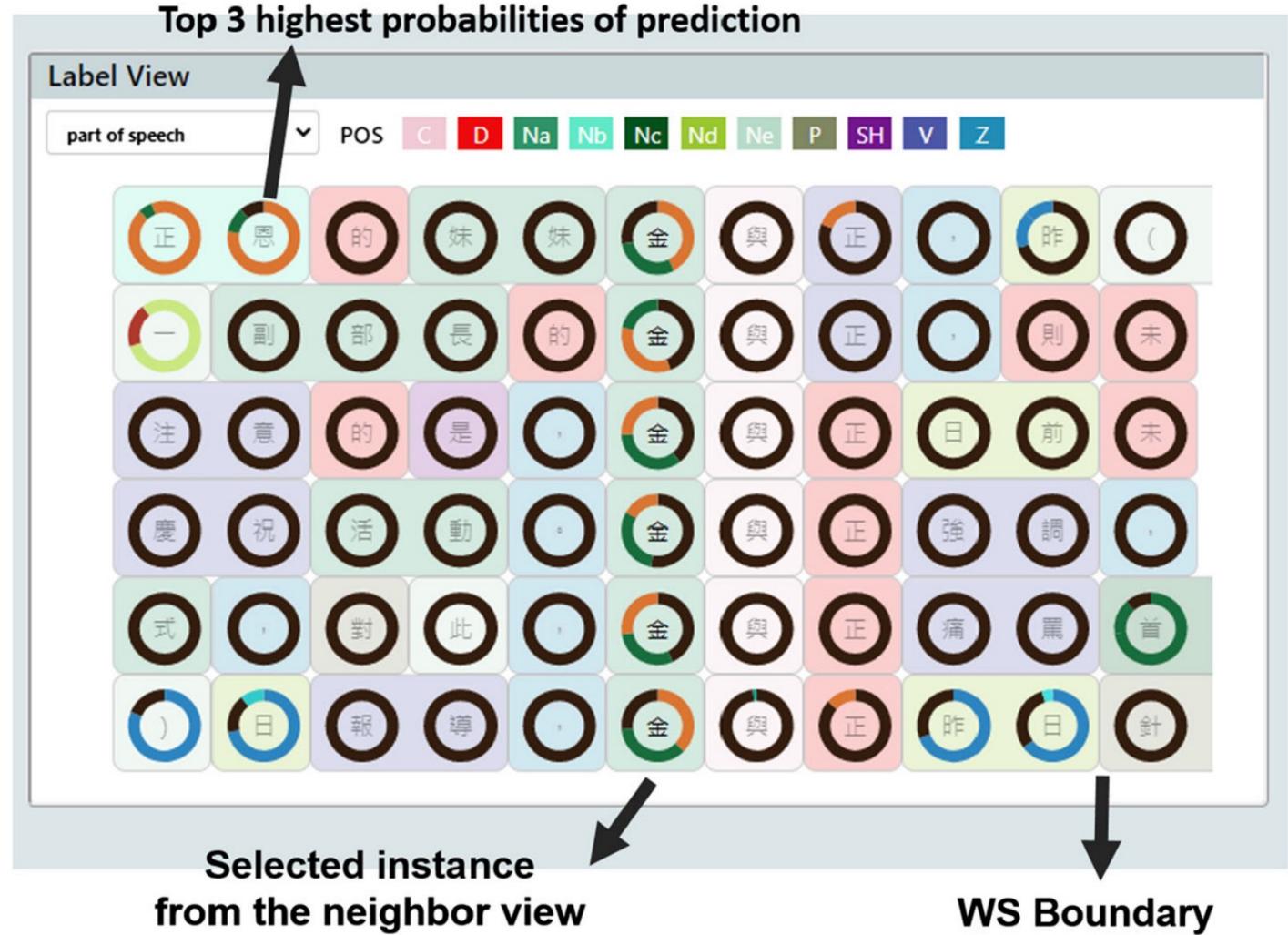
CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition



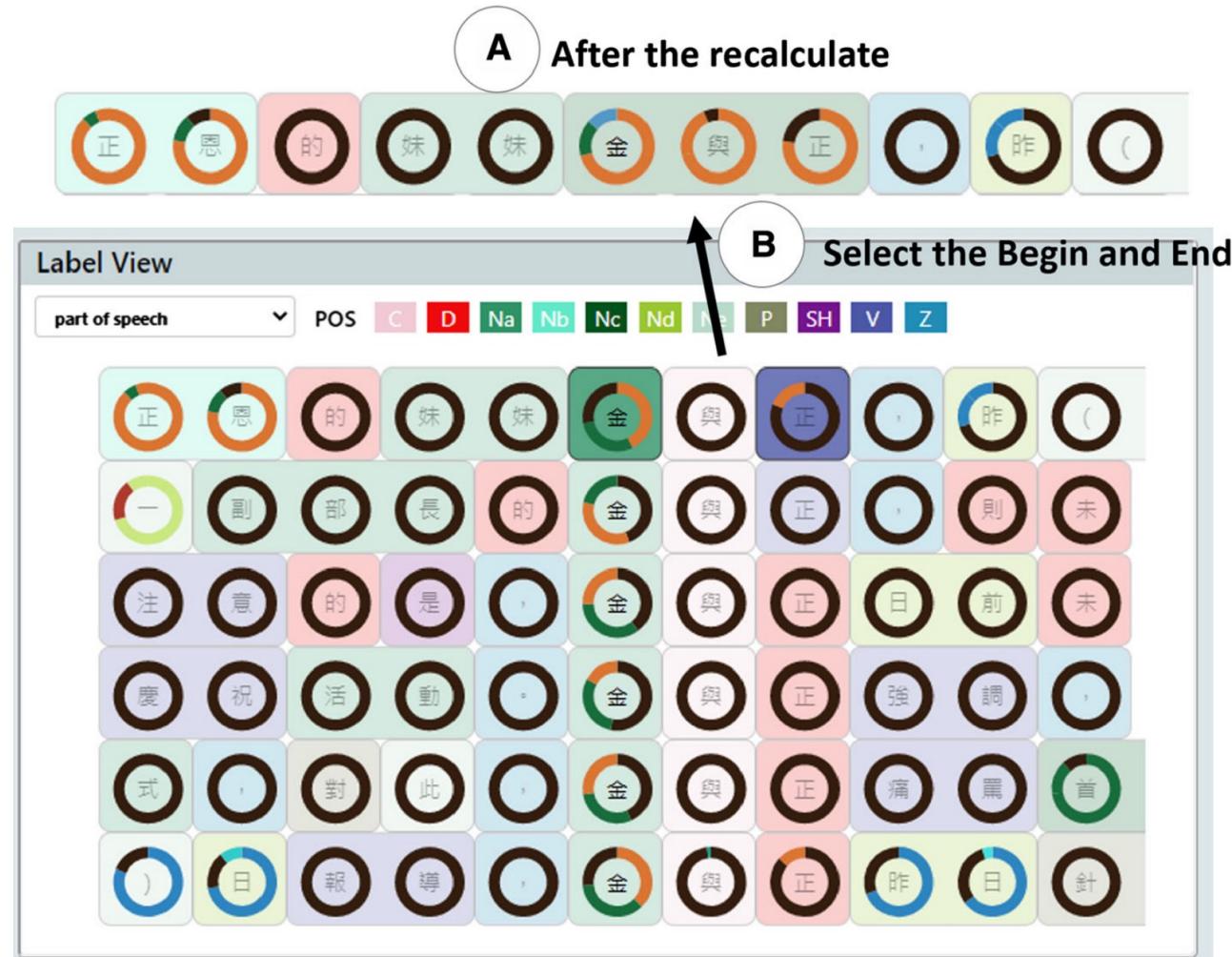
CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition



CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition

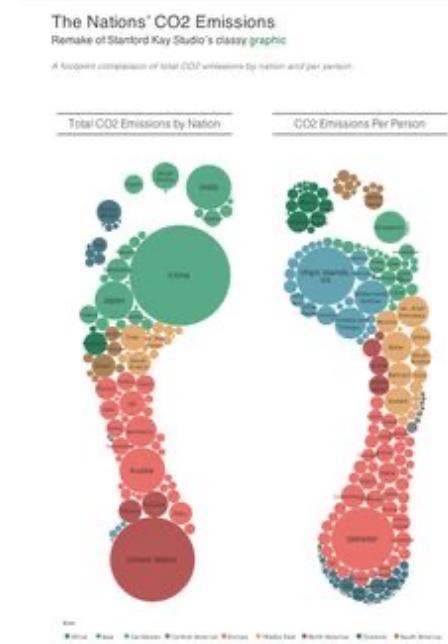
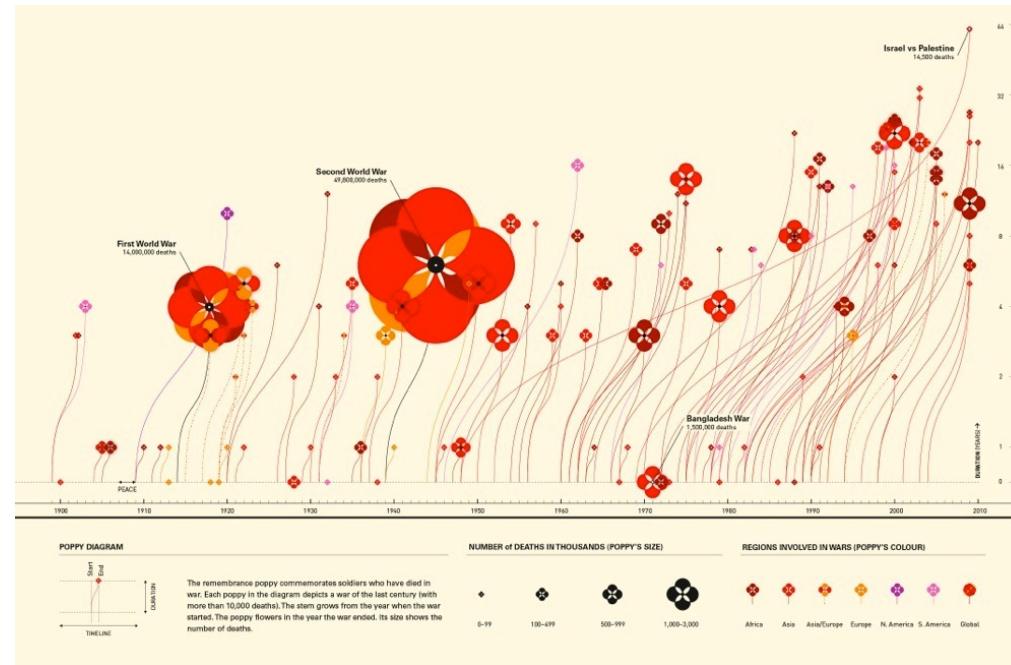


CNERVis: A Visual Diagnosis Tool for Chinese Named Entity Recognition



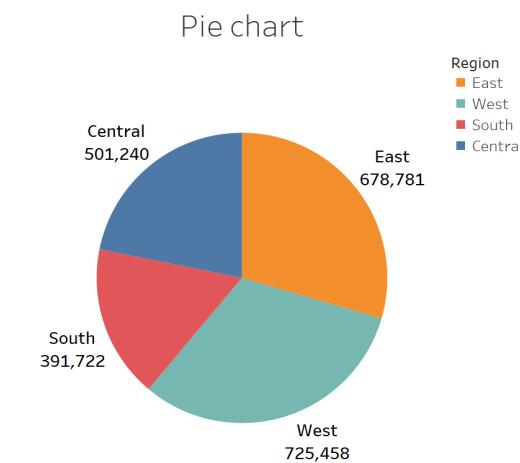
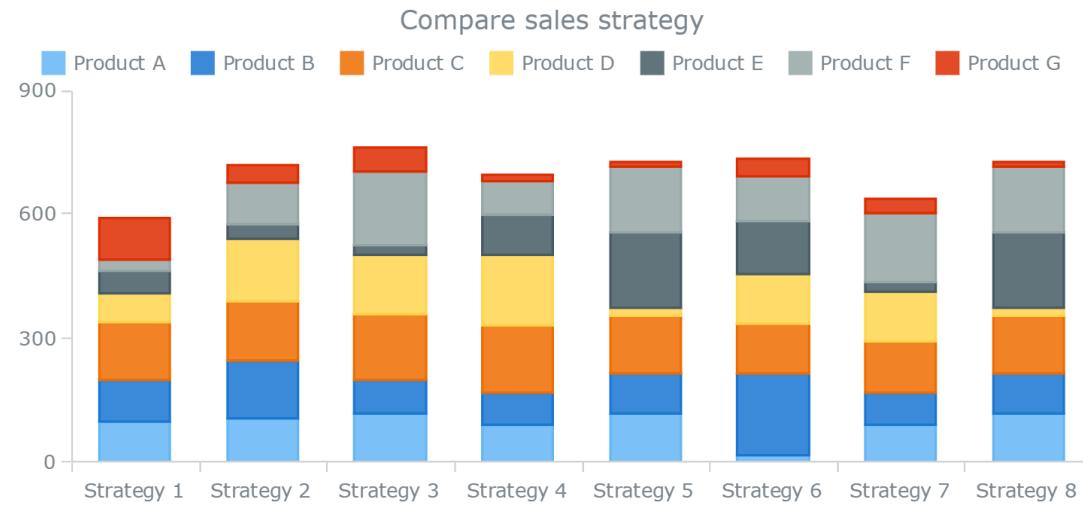
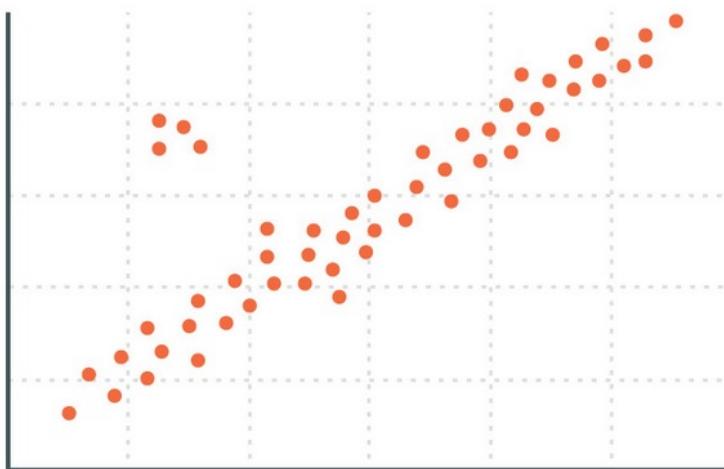
This course

◎ Have a nice-looking plot?



This course

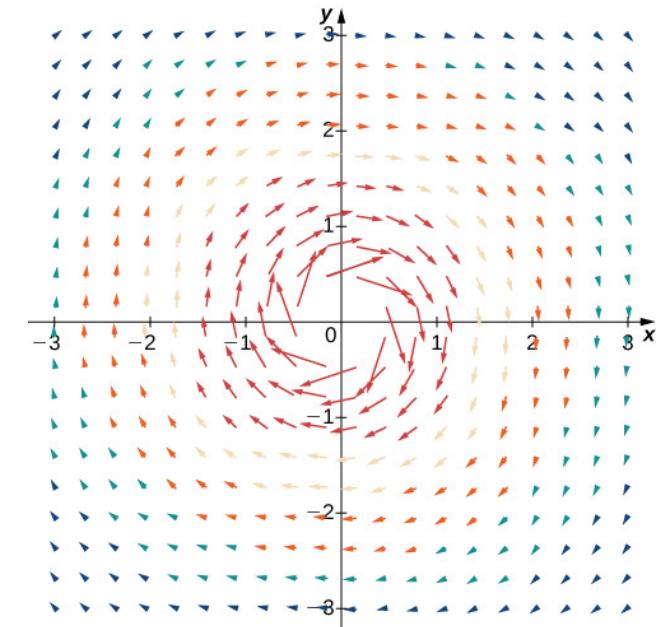
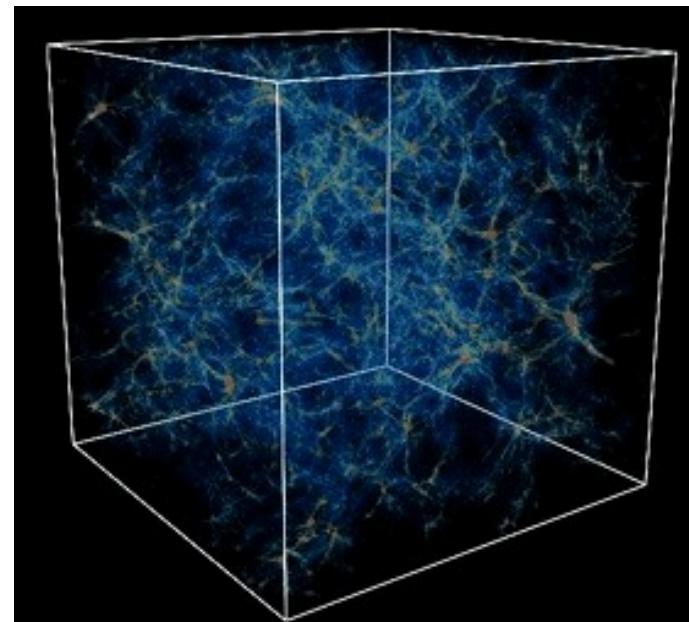
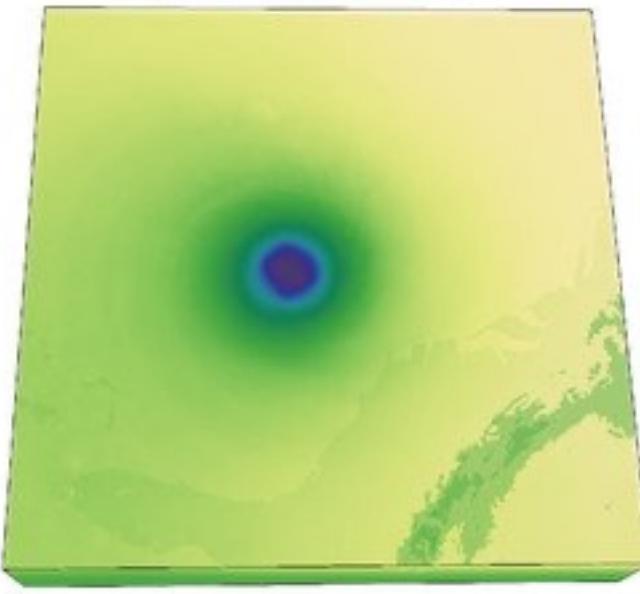
- ~~Have a nice looking plot?~~
- Effectively and correctly convey the information



Data

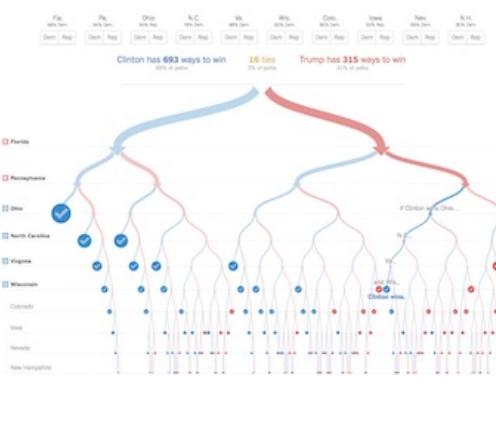
Data for Scientific Visualization

- Graphical representation of data
 - **Spatial data:** scalar/vector /tensor field defined on structured/unstructured meshes; particles
 - Non-spatial data: graph, tree, texts, table, etc



Data for Information Visualization

- Graphical representation of data
 - Spatial data: scalar/vector /tensor field defined on structured/unstructured meshes; particles
 - **Non-spatial data:** graph, tree, texts, table. etc



Jackson Fortson Campbell Mercer Miller Gatling Jackson Nash Howie
Hughes Sura Jamison Francis Croshere O'Neal Finley McInnis Nowi
Stackhouse Blaylock Olajuwon Best Miller Piatkowski Miles Mag
Robinson Mills Brandon Taylor Mobley Martin Rose Marbury Oc
Davis Allen Garnett Billups Gill Rice John
Harrington Hunter Szczerbiak Ellis Camby Hous
Caron Harris Van Horn Barry Pollard Sprewell Anderson Tho
Jones Stoudamire Peeler Martin Jackson Anderson Robir Dun
Hill Mutombo Kidd Cassell Williams Smith Turkoglu Chris
McLeod Delk Wells Strickland Long Laettner Funderburke Dai
y Ratiiff Sabonis Hardaway Pippen Dickerson Long Laettner Funderburke Dai
Robinson Wallace Divac Hamilton

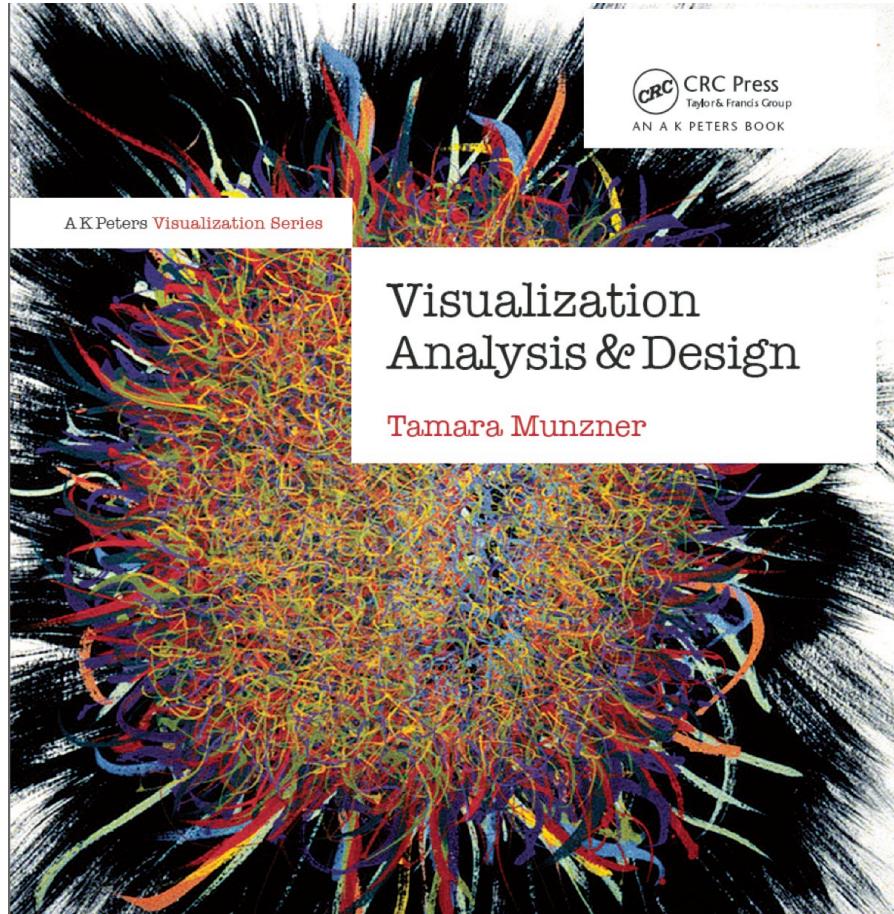


A	B	C	D	E
1 Last Name	Sales	Country	Quarter	
2 Smith	\$16,753.00	UK	Qtr 3	
3 Johnson	\$14,808.00	USA	Qtr 4	
4 Williams	\$10,644.00	UK	Qtr 2	
5 Jones	\$1,390.00	USA	Qtr 3	
6 Brown	\$4,865.00	USA	Qtr 4	
7 Williams	\$12,438.00	UK	Qtr 1	
8 Johnson	\$9,339.00	UK	Qtr 2	
9 Smith	\$18,919.00	USA	Qtr 3	
10 Jones	\$9,213.00	USA	Qtr 4	
11 Jones	\$7,433.00	UK	Qtr 1	
12 Brown	\$3,255.00	USA	Qtr 2	
13 Williams	\$14,867.00	USA	Qtr 3	
14 Williams	\$19,302.00	UK	Qtr 4	
15 Smith	\$9,698.00	USA	Qtr 1	
16				

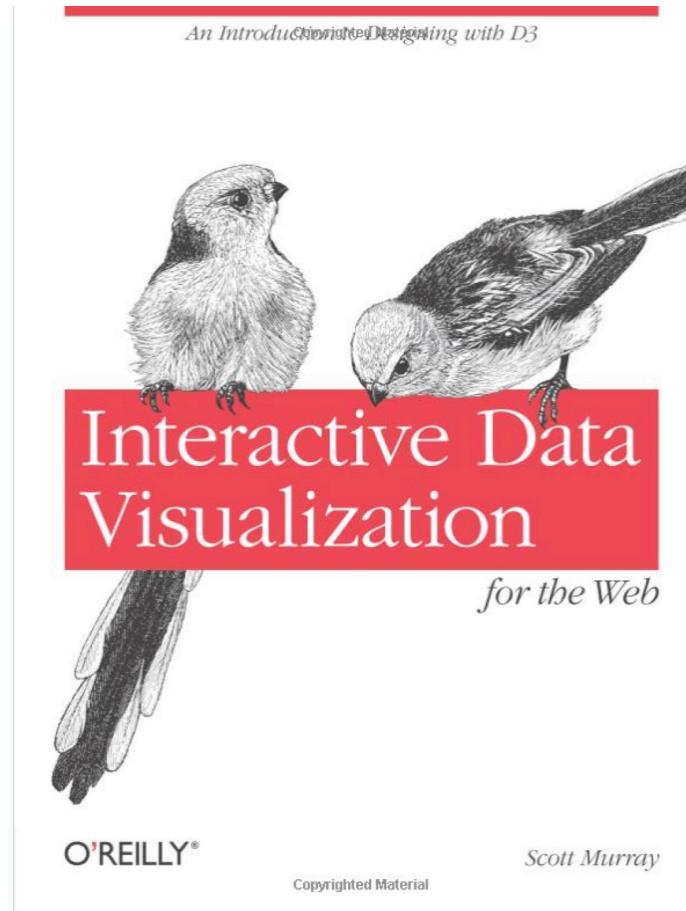
Course Focus

- Analyzing existing visualization techniques/idioms as a springboard for designing new ones
- Understand how we think visually
 - How to organize space
 - Which color and shape will stand out
 - When we should use images instead of words
- Practical Examples of Visual Design for Envisioning information
 - Theories of visual design for data analysis
 - Design strategies
- Information and Scientific Visualization
- Data visualization via Python/Matplotlib; Javascript/D3; Paraview/VTK; Tableau

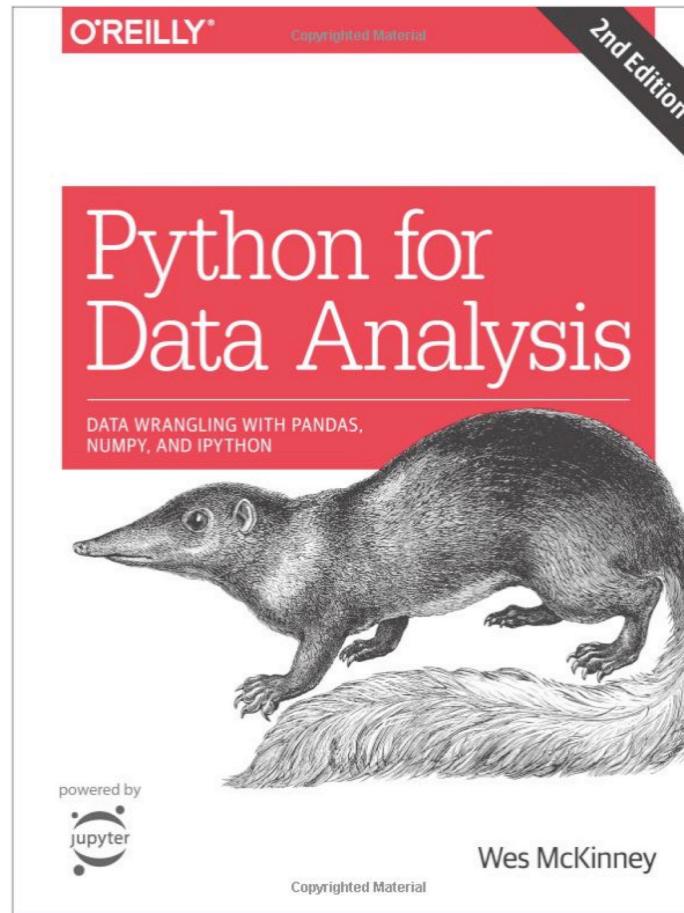
Textbook



D3 Programming



Python and Data



Data Visualization

Information Visualization

Vis concept & strategies



What?

Why?

How?

Actions

Analyze

→ Consume

→ Discover

→ Present

→ Enjoy

→ Produce

→ Annotate

→ Record

→ Derive

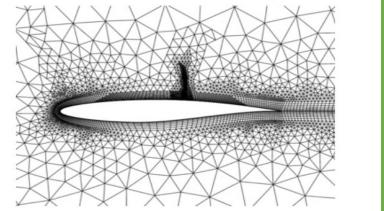
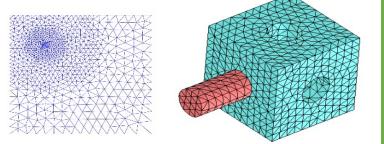
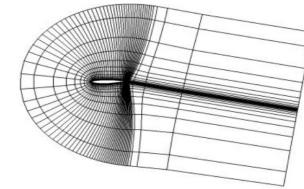
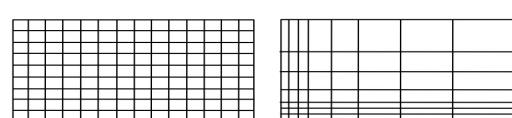
Tool



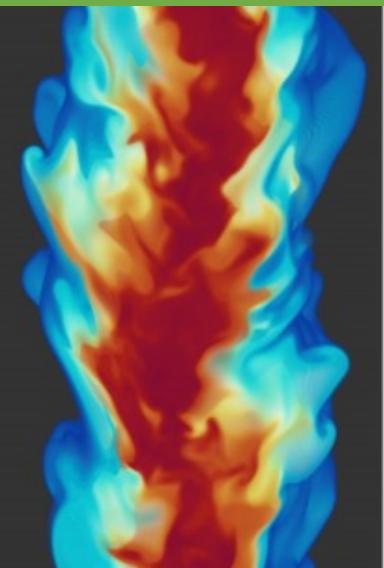
Data-Driven Documents



Scientific Visualization



 **ParaView**



Prerequisites

- No, but...
- Self-learning ability for programming languages
 - Python – Pandas, Matplotlib
 - HTML, CSS,...
 - Javascript
 - D3 (TA)

Data Visualization Course

- **Information Vis:** visual design concept for visual analytics tool
- **Scientific Visualization:** visualization techniques for scientific data
- **D3 (for information visualization):** techniques to implement an interactive tool (TA)
- **Final Project Demonstration:** week 16

Course Topics

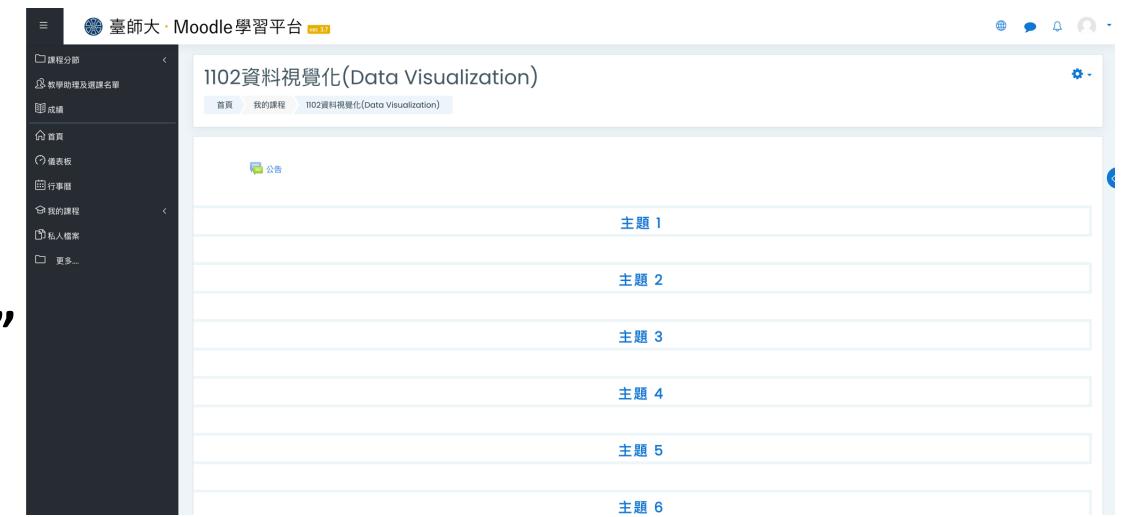
- Foundation
 - Visual Perception
 - Information and Data Analysis
- Visualization Components
 - Data Abstraction
 - Task Abstraction
 - Retina Attributes (Marks and Channels)

Course Topics

- Basic Visualization Techniques
 - Tables
 - Diagrams
 - Networks
 - Maps
- Scientific Visualization
 - Paraview
 - Data Models
 - Scalar Visualization (isosurface, volume rendering)
 - Vector Visualization

Material, Homeworks, Project, Announcement...

- All course materials are on Moodle.
- Assignments and the project
 - I will announce them on Moodle.
 - Submit your assignments and project on “Moodle”
 - I will grade and give the score on “Moodle”
- **Please check the new announcement on Moodle frequently.** Check the announcements on Moodle is your responsibility.



Student Assessments

- Homeworks: 50%
- Midterm: 20%
- Attendance: 5%
- Final Projects: 25% (teamwork)
 - The detail will be announced later

Late Assignment/Project Policy

- Assignment
 - 80% credit for one day (within 24 hours) late
 - 60% credit for two days late
 - 40% credit for three days late
 - 20% credit for four days late
 - **No credit given after four days late**
- Project
 - No credit given with any delay (**No delay!!**)

Academic Misconduct Policy

With regards to your homework and project, students are encouraged to discuss with other classmates. However, students should not copy any code from other people. If students copy any code from websites, you must indicate the code segment that you copied, including the source web address of the code. You are required to contribute most of the code in your assignments and the projects in principle; do not turn in anything mostly written by other people. If you are not sure whether you can use a code, please talk to me before submitting the code.

I reserve the right to give you a 0% on an assignment or fail (0 points) of this course if I believe it is clear that you copied most or all of the answers.

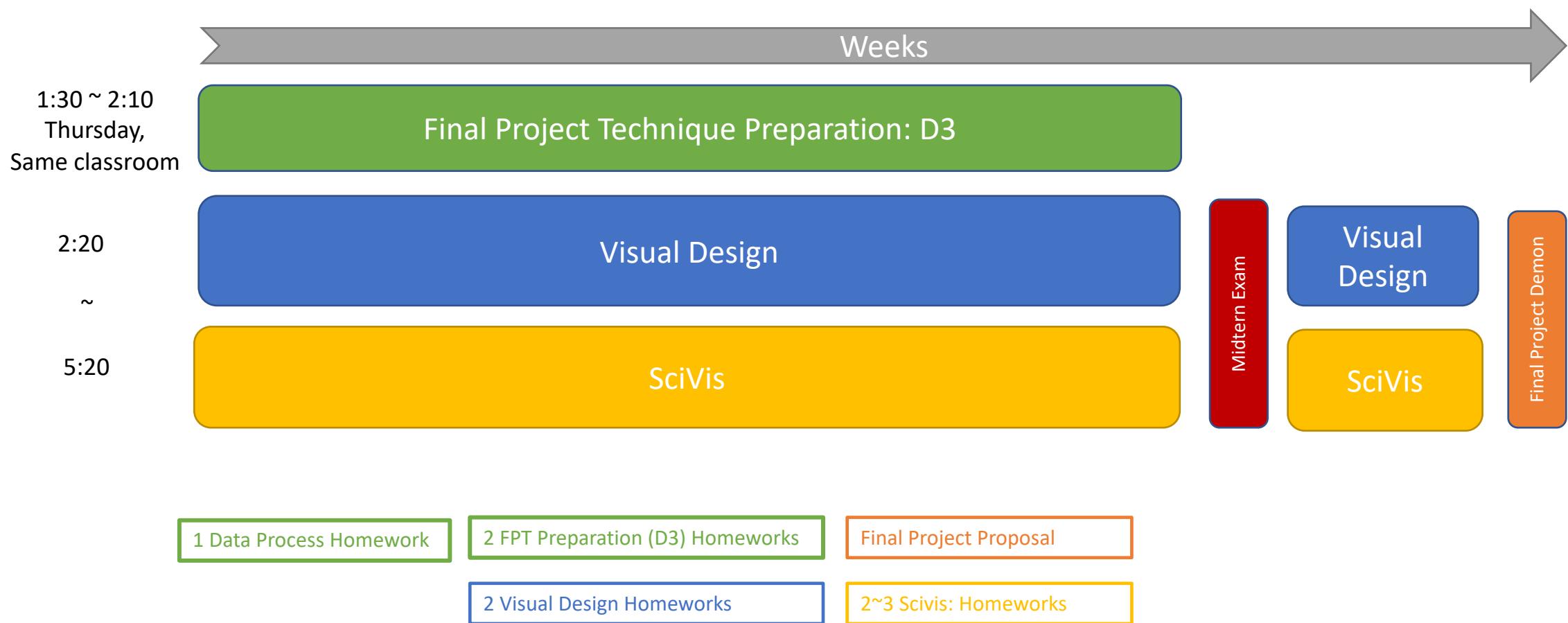
Final Project

- Group project (2 students)
- You will select a dataset and create a visual analytics tool by D3
- You have to find your partner around Week 10 and select a dataset
- You have to turn in a proposal around Week 12 (I will review and accept it or ask you to revise it)
- Week16 final project demonstration

Final Project Technique Preparation: D3

- Not mandatory to attend this Final Project Technique Preparation hour
- You need D3 to implement your visual design and create your visual analytics tool for the final project
- TA (my four master's students) will teach D3 for you to prepare your final project
- TA D3 tutorial will start at Week2
 - Time: 1:20PM every Thursday (may last 8 ~ 10 weeks)
 - Location: same classroom (S601)
 - Language: English (or Mandarin)
 - TA will post the topic on Wednesday
- Alternative: I will put my D3 tutorial videos on youtube
- I will announce the first D3 homework early of next week
 - Deadline could be in week 4 or week5
 - This homework help you to judge whether you need the TA D3 tutorial

Timeline



SciVis, VisualDesign

- You do not need the knowledge from VisualDesign to learn Scivis, and visa versa
- You can think of our 3 hours lecture as you step into another classroom and listen to another professor to teach scivis when you finish the VisualDesign part here

Office Hour

- Instructor: 王科植
- Email: kcwang@ntnu.edu.tw
 - I suggest that you also copy the text and send it by personal message on Moodle to me if you email me anything. Students' emails are sometimes classified as junk mail and I may never see it.
- Location: 應用科學大樓506
- Time: Email me to schedule (Monday and Tuesday are better).

- TA: 王昱琳
- Email: yuling.wang6@gmail.com
- Location: 應用科學大樓109
- Time: TBD (or email)

The Very First Homework

- Check Moddle
- Deadline: 8/21 (Wed) 11:59PM
- Data Process, analysis and basic visualization by Python (Pandas, Matplotlib)
- Install anaconda, use Jupyter notebook to answer and submit
 - You need Numpy, Pandas, and Matplotlib package



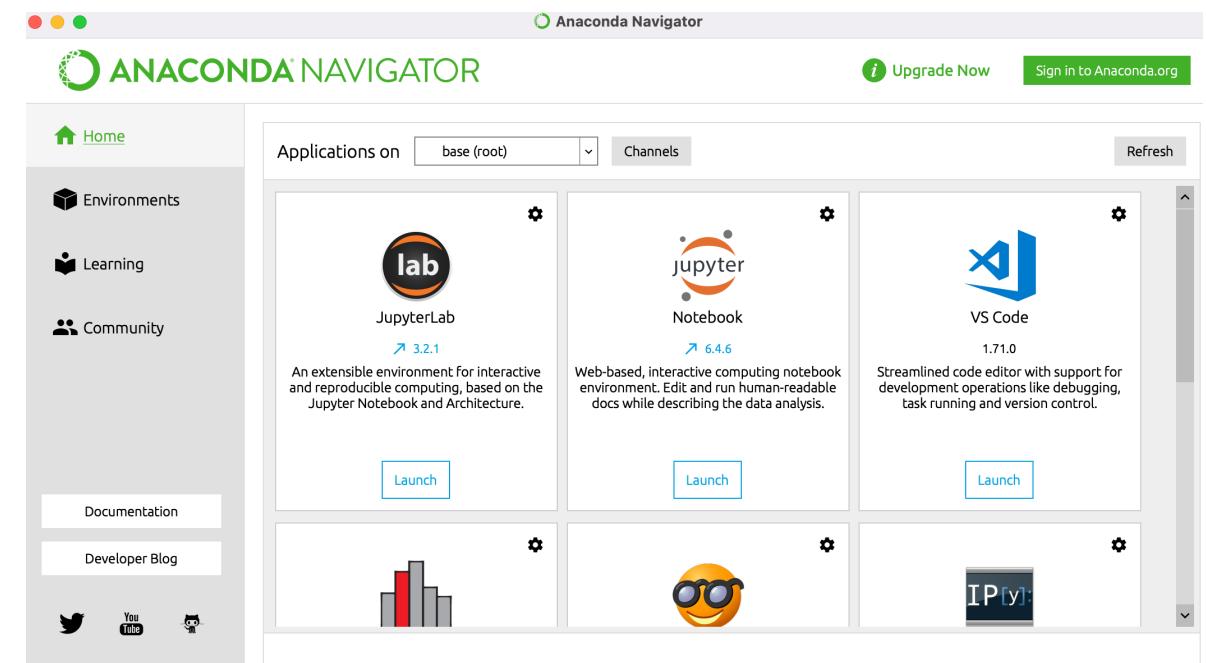
Data science technology for
a better world.

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of open-source packages and libraries today.

Download

For MacOS

Python 3.9 • 64-Bit Graphical Installer • 591 MB
Get Additional Installers



Summary

- The very first homework (deadline: 9/21)
- TA D3 tutorial right before our main lecture start every Thursday (alternative: youtube)
- Main lecture: visual design, scientific visualization
- Make sure you have good self-learning ability for programming languages
- You will have around 10 deadlines (includes all homework, project, proposal, midterm)