HKUST
VISLAB

COMP 4462
Data Visualization Tutorial

Leo Yu Ho, Lo
Ming Yao

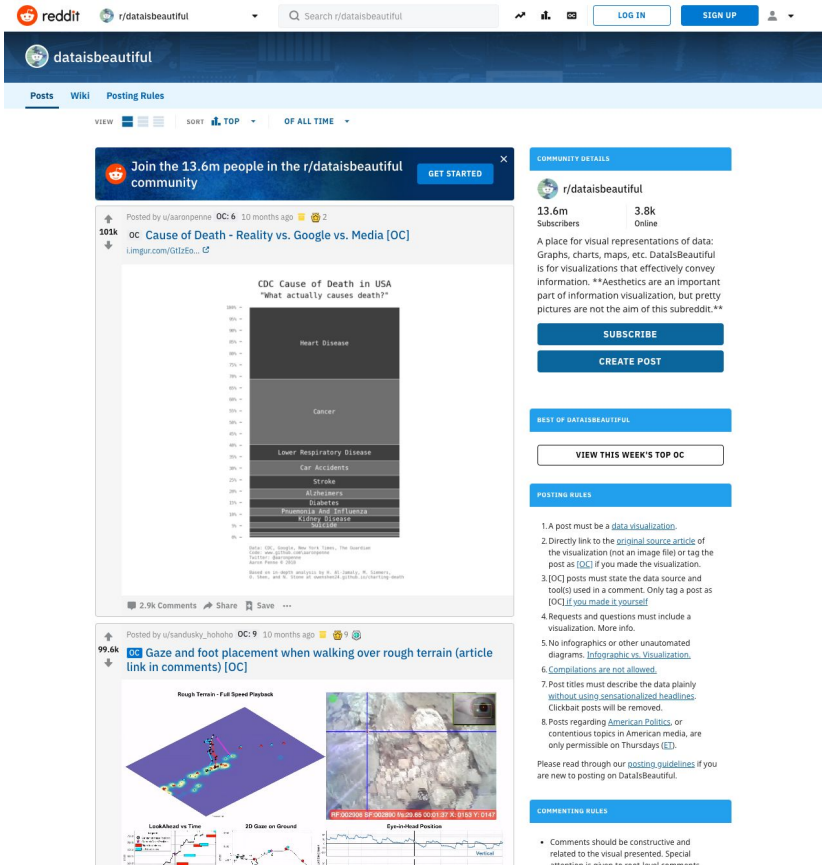
Tuesday 26 February, 2019
https://bit.ly/vis-t03

Course Project & Top-Vis Competition

- Project (50%)
  - Grouping: Sign up sheet (27 Feb)
  - Phase 1: Proposal presentation (27 Mar & 29 Mar)
    - Find a dataset
      - What kind of data? How large is it? Why this dataset?
    - Visualization tasks / data processing / visual encoding
    - Good to have a mock up
  - Phase 2: Project presentation (3 May & 8 May)
    - Make it real! Coding & demo
    - Share stories in the data
- Top-Vis Competition (10%)
  - 2 mins to present 2 visualizations
  - 24 Apr & 26 Apr
  - Write up a short essay
    - Why you chose this visualization?
    - What data are visualized? How are they encoded?

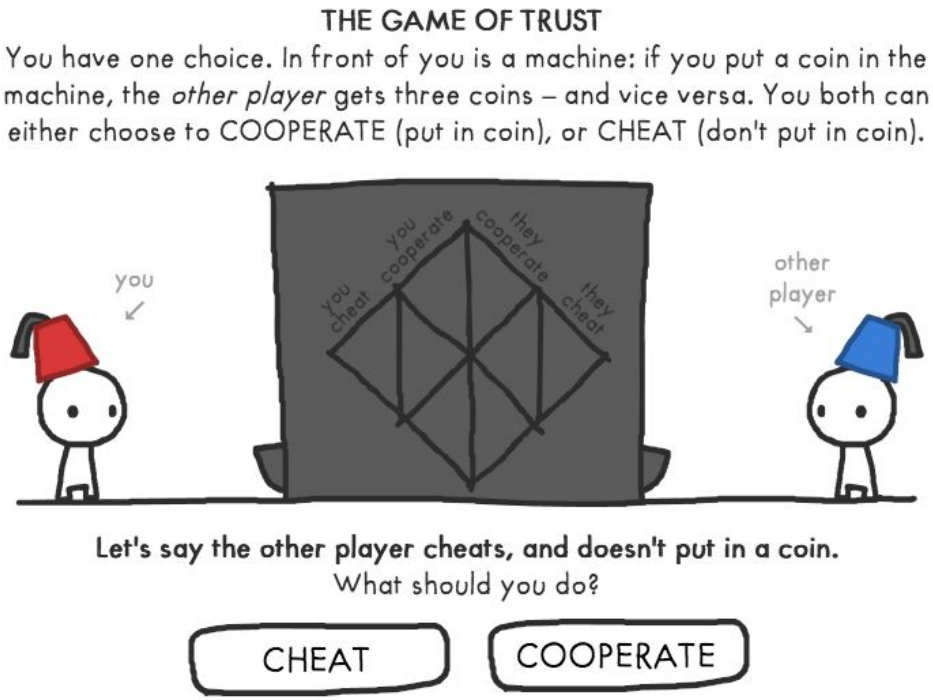
Data is Beautiful

- New visualizations everyday
- Top post of all time
  - Visualization with highest voting of all time
- A lot of remarkable ideas
- Mainstream:
  - Meaning of data > visual effect
  - And some are visually impressive
- Another subreddit: Data is Ugly
  - Lying with charts
  - Deceiving, scam
  - Some are from very authoritative sources
    - Famous news websites
    - Governments
    - Famous companies



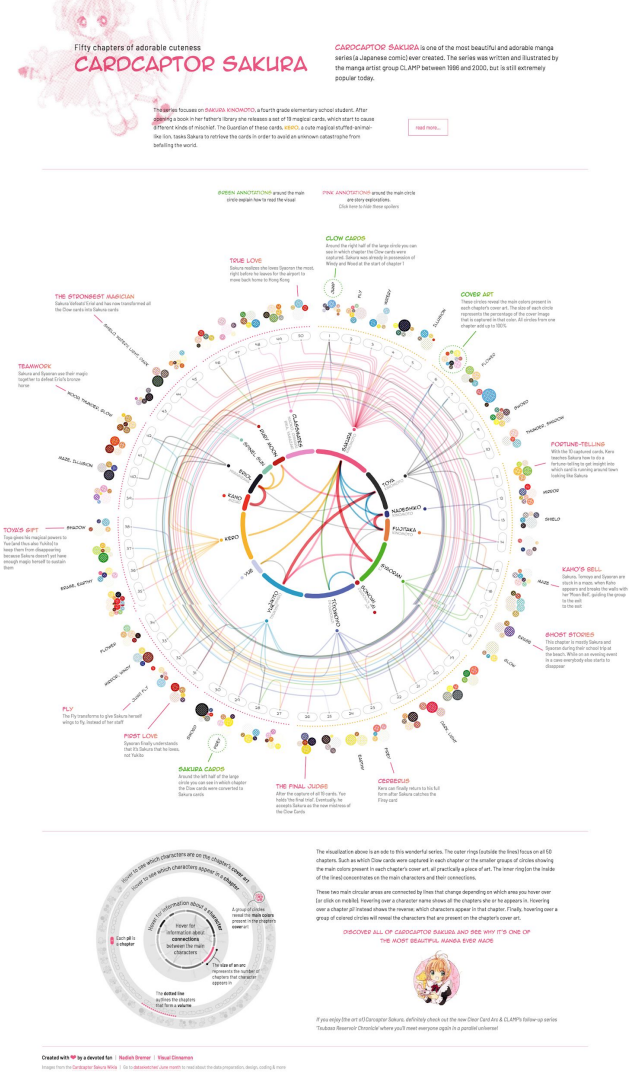
Nick Case

- Narrative visualizations
  - Telling a story with visualizations
- Evolution of Trust
  - Game theory about our society
  - Prisoner dilemma
    - CHEAT?
    - COOPERATE?
  - Interactive
  - Nice graphics and music
  - A sandbox simulator at the end
  - Enjoy!
- More on Nick Case's webpage



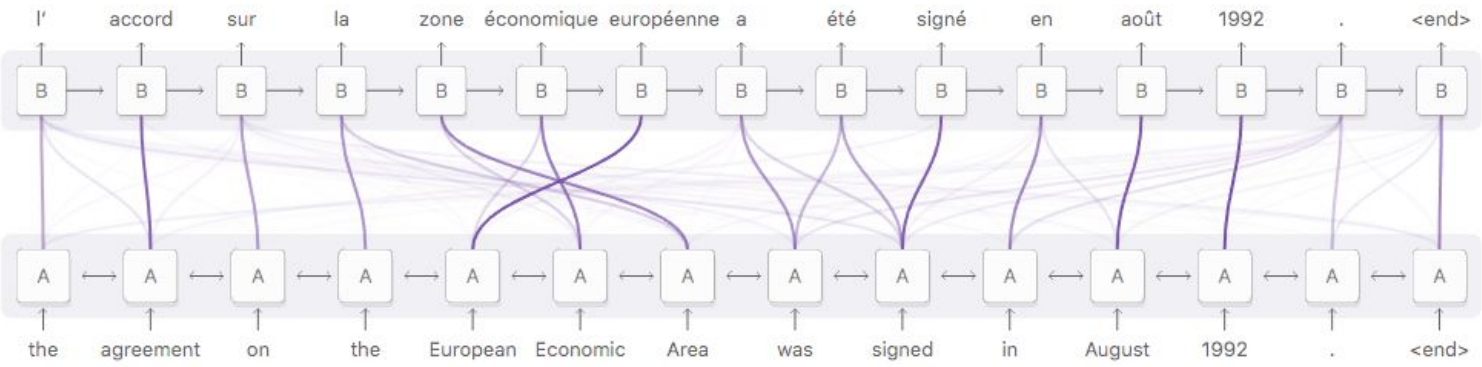
Data Sketches

- Beautiful! Eye pleasing! Fun datasets!
- And they have 24 of them!
- By:
  - Nadieh Bremer
  - Susie Lu
- Cardcaptor Sakura
  - Visualizing 50 chapters of the manga
    - Appeared characters
    - Magic spells
    - Annotations
- Another one on Dragon Ball Z
- With explanations!
  - They have journaled the process in details!



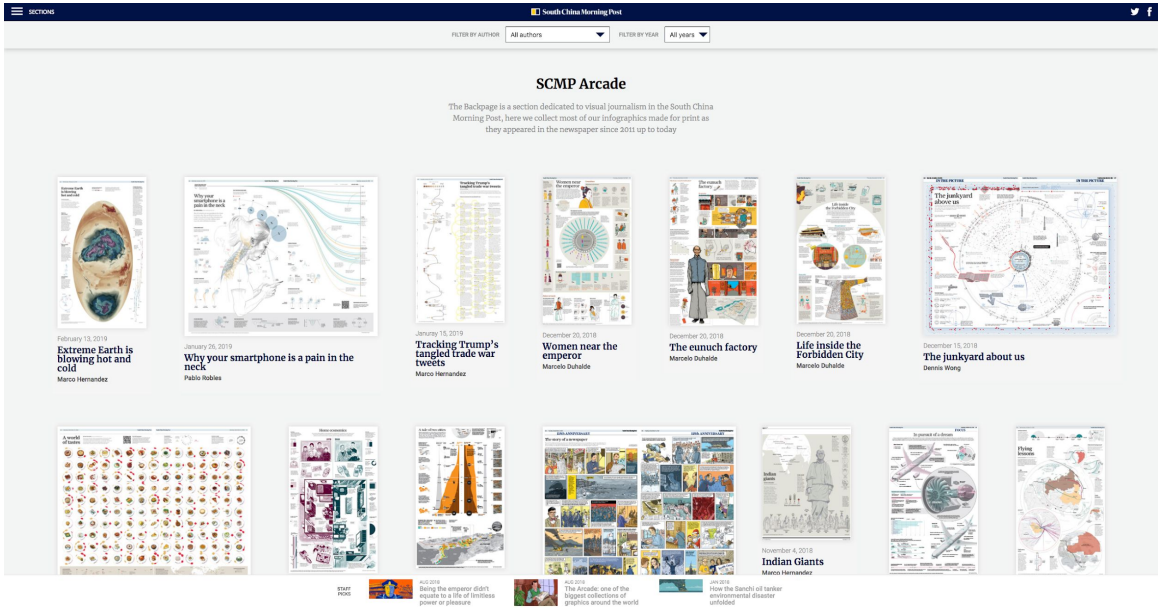
Distill

- Visual Explanation of Machine Learning Algorithms
- Attention and Augmented Recurrent Neural Networks
  - Visualizing a neural translation model
  - Which word in a French sentence <=> which word in English?



The list of 2018 visualization lists

- 33 lists, each has 10+ visualizations!
- 2018 in visuals: South China Morning Post's infographic highlights
- SCMP Print Arcade
  - 217 visualizations from 2011 to 2019

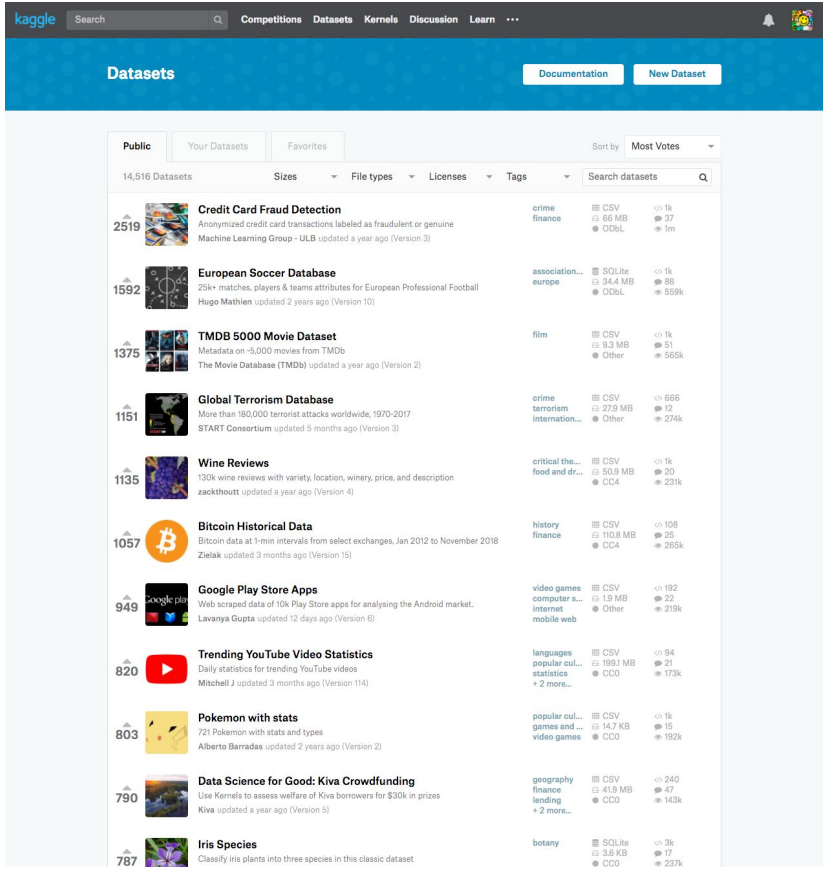


Junk Charts

- A collection of bad visualizations
  - How to lie with visualizations
  - Like Data is Ugly subreddit
  - With explanations
  - Update frequently

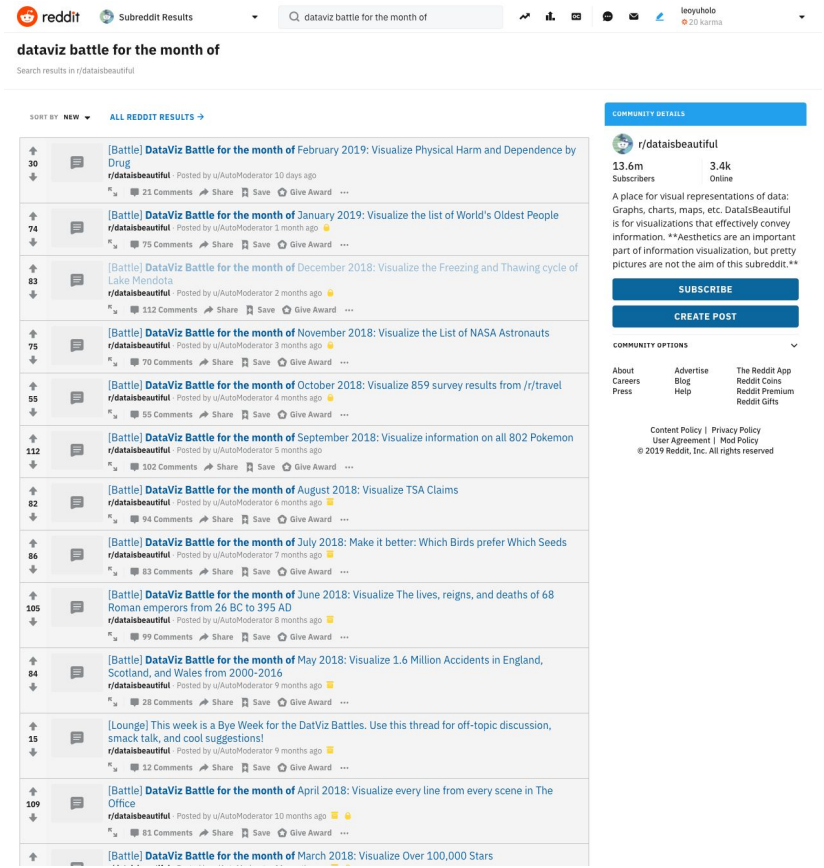
Kaggle Datasets

- No.1 source of datasets
- A lot of datasets
- Data are clean (relatively)
- A lot of kernels (jupyter notebooks)
  - See what the others do with the datasets
- Can seek help very easily
  - Can also raise questions to the authors



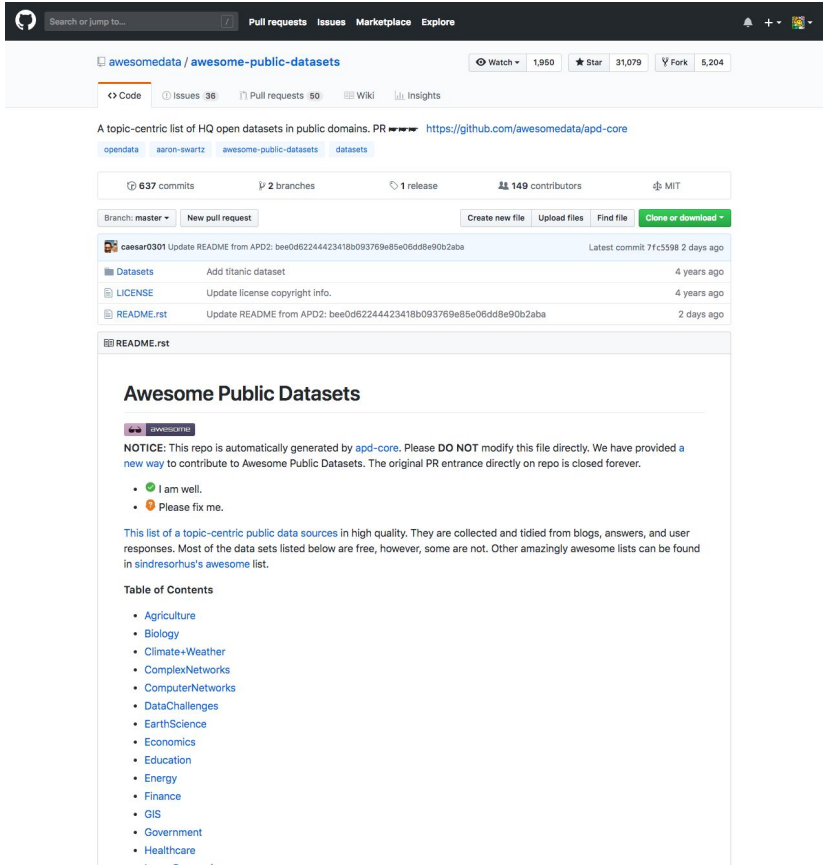
Dataviz Battle on r/dataisbeautiful

- Monthly competition on r/dataisbeautiful
- A lot of submissions for references
- September 2018: Visualize information on all 802 Pokemon
  - Winners are announced in the Dataviz Battle thread of next month
  - For example, October 2018 announced the winners of visualizing Pokemon



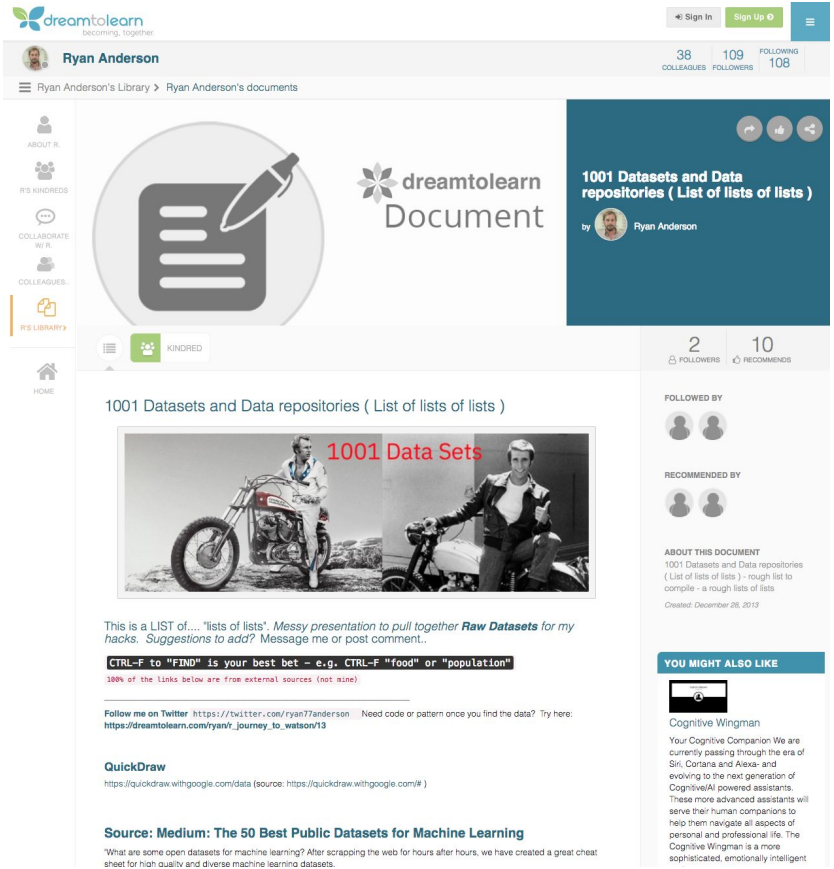
awesome-public-datasets

- A very thorough list
- With active update
- Search Engine subsection
  - Websites that have “search for datasets”
- Data Challenge subsection
  - More Kaggle like websites
- Complementary Collection subsection
  - More dataset lists



1001 Datasets and data repositories

- Another list of lists
- A long long list
- Compiled from many sources
  - GitHub
  - Government
  - Blogs
  - Big Companies
  - Quora
  - Reddit
- In the blog.visual.ly subsection
  - Specifically suitable for visualization



Tasks

- Get the whole list of "Where to find visualizations and datasets" on GitHub
- If you don't have a group yet:
  a. Talk to your classmates about the visualizations you like
  b. Form a group if interest matches
     i. Signup on Google Docs
- Go to your group mates
  a. Find some visualizations
  b. Talk to your group mates
  c. Find a dataset to work on
  d. Talk about what interesting insight can be found in the dataset
- Make amazing visualizations!

Next tutorial

Python, Jupyter and Pandas

- Prepare your Google account beforehand
  - For using Google Colab
  - Jupyter notebook environment
  - Free!
  - No setup
- Alternatively, you can use jupyter notebook on your computer, but that is cumbersome