

UIUC CS 498 Data Visualization Narrative

Visualization

Wonhee Jung (wonheej2@illinois.edu) - University of Illinois at Urbana-Champaign

This repository is for the University of Illinois at Urbana-Champaign MCS-DS CS 498 Data Visualization summer 2019 assignment - Narrative Visualization and published to Github Page per instructor's request.

Github Page for demo: https://freesoft.github.io/UIUC-CS498_DV_Narrative_Visualization/
Git Repo: https://github.com/freesoft/UIUC-CS498_DV_Narrative_Visualization/

This doc contains a several sections, including `Review Criteria` by instructors, to explain the result of narrative visualization.

Data Source

Instead of using static and embedded CSV data source for data visualization, I've decided to use APIs so that the data source of this chart doesn't need to be updated later on. To do that, I've decided to use World Bank APIs, not downloadable bulk CSVs from World Bank website.

World Bank APIs :

- Developer Information: <https://datahelpdesk.worldbank.org/knowledgebase/topics/125589-developer-information>
- API Basic Call Structures: <https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structures>
- Country API: <https://datahelpdesk.worldbank.org/knowledgebase/articles/898590-country-api-queries>
- Aggregate API query: <https://datahelpdesk.worldbank.org/knowledgebase/articles/898614-aggregate-api-queries>
- Metadata API query: <https://datahelpdesk.worldbank.org/knowledgebase/articles/1886695-metadata-api-queries> etc

World Bank API provides APIs that users can call with specific indicators, and I used `SL.EMP.WORK.ZS` to get % of employment for average wage and salary workers. The API also provides the way to obtain either world information(with country code `WLD`) or each country's unique three-digit country code. (e.g., USA for the United States of America, KOR for South Korea, etc.)

How to use World Bank APIs

Following HTTP Get request for the API is to get average wage and salary worker's % of employment(`SL.EMP.WORK.ZS`), worldwide(`WLD`), and between 2000 and 2017.(`date=2000:2017`).

```
http://api.worldbank.org/v2/country/WLD/indicator/SL.EMP.WORK.ZS?format=json&date=2000:2017
```

And you'll get the response as below;

```
[
  {
    "page": 1,
    "pages": 1,
    "per_page": 50,
    "total": 18,
    "sourceid": "2",
    "lastupdated": "2019-07-10"
  },
  [
    {
      "indicator": {
        "id": "SL.EMP.WORK.ZS",
        "value": "Wage and salaried workers, total (% of total employment) (modeled ILO
      },
      "country": {
        "id": "1W",
        "value": "World"
      },
      "countryiso3code": "",
      "date": "2017",
      "value": 51.8887433801719,
      "unit": "",
      "obs_status": "",
      "decimal": 1
    },
    // more data, skipped
  ]
]
```

Which means that wage and salary workers' percentage in 2017 worldwide was 51.88%.

Review Criteria (copied from the instruction in Coursera)

An essay will be required and will be submitted along with the URL of the narrative visualization. This essay is an important piece of the assignment as it is used for you to communicate your understanding of the concepts of narrative visualization and how they apply to the one you created.

The essay should contain the following sections.

- **Messaging.** What is the message you are trying to communicate with the narrative visualization?

This visualization assignment asks why the average rate of waged & salaried worker is so different between countries. I've found this oddness while I was working on a previous Tableau Dashboard assignment, and decided to make a story to question myself and also for the viewers.

For example, regardless of male or female, the worldwide rate for wage&salary worker has been increasing for years, and now it reaches over 50%. It is known that as a country being developed well, this rate should go up, means self-employment rate should have decreased. However, while this wage&salaried worker's rate stays around 50%, some countries like the US has more than 90%. And it was similar for many developed western countries.

Thus, it was natural to think that the rate of highly developed country's WDI indicator will be higher than developing country's one. But this hypothesis wasn't correct when I checked China and Russia. I expected a 50% rate for both countries, and if there is one country higher than the other, I thought it's China over Russia. But well, Russia's plotting showed pretty much the same as the US, while China's number is just over 50%.

So I wanted to share what I found previously, and make it a narrative story so that viewers can have time to think, and hope to see other evidence and data to answer these differences. Also, I gave a feature to investigate other country's WDI data for employment rate.

- **Narrative Structure.** Which structure was your narrative visualization designed to follow (martini glass, interactive slide show or drop-down story)? How does your narrative visualization follow that structure? (All of these structures can include the opportunity to "drill-down" and explore. The difference is where that opportunity happens in the structure.)

I've used Martini Glass structure. I added four scenes to show the different country's line chart, including world average, and then attached the last scene in the visualization with a few controllable options that viewer can navigate different country's data, including data type, e.g., male, female, or total ratio.

- **Visual Structure.** What visual structure is used for each scene? How does it ensure the viewer can understand the data and navigate the scene? How does it highlight to urge the viewer to focus on the important parts of the data in each scene? How does it help the viewer transition to other scenes, to understand how the data connects to the data in other scenes?

Simplified the scene navigation for the Martini Glass narration model, so view can only navigate to the next scene until it goes to the last page, which makes viewers less interrupted by lots of navigation buttons and menus.

Five scenes in the visualization use the same template. The top area of the template tells what data viewer see, followed by a detailed message that I asked myself and also wanted to show other viewers.

Plotting section (SVG) comes in the middle of the page, followed by line chart color information to explain the meaning of each different color in the line chart.

At the bottom, I added WDI indicator explanation section in 5 scenes so when a viewer can double-check the meaning of current WDI indicator without moving to the previous page or clicking the help button when they needed.

- **Scenes.** What are the scenes of your narrative visualization? How are the scenes ordered, and why

Those five scenes are re-ordered as the way I wanted to tell the story. First, it shows world average data so that viewers can have a measure for the following scene. 2nd scene is for the US, which is a highly developed country and has a high rate of wage/salaried employment rate. 3rd and 4th one are impressive because it's both the country ruled by a communist party.

China is one of the most powerful countries in the 21st century, not just for the military but economically as well whereas Russia is still a powerful country for military perspective but not as much as China when it goes to the economy. Thus, my visualization shows two countries, China and Russia, one by one, but the first one has world average level of wage/salary employment rate (China) whereas second country(Russian) has a surprisingly high percentage, almost same as the US.

I wanted to make the viewer think, "Hmmm, that's interesting. Yeah, now I want to know why?". If you feel the way as I hoped, I can say my narrative visualization completed the mission.

- **Annotations.** What template was followed for the annotations, and why that template? How are the annotations used to support the messaging? Do the annotations change within a single scene, and if so, how and why

I added each year's rate value as d3-tooltip on each datapoint's circle, so when a viewer performs mouseover event, tooltip annotation pops up and shows what the rate for that specific year was. d3-tooltip also provides 3-digit country code in the last scene and help the viewer to distinguish between multiple line charts.

I've tried to add general annotation on the chart, but there weren't significant changes I found for the wage/salary worker's average ratio. If zoomed in, the rate might drop 2~3% on a specific event; however, the overall percentage doesn't get impacted by individual activities. That said, I've decided not to add redundant and useless annotation on the chart.

- **Parameters.** What are the parameters of the narrative visualization? What are the states of the narrative visualization? How are the parameters used to define the state and each scene?

Current visualization impementation has a few parameters internally.

1) Plotting works differently based on given 3-digit country code.

2) Gender type for data navigation. D3 code calls different APIs to get separate dataset, and also draw a

chart with different color based on this selection. (orange: total, blue: male, red: female)

3) Based on current scene location, visualization page shows different option. For page 1~4, it doesn't show any option for viewer to navigate; However the page shows gender radio box and country select box in scene 5.

- **Triggers.** What are the triggers that connect user actions to changes of state in the narrative visualization? What affordances are provided to the user to communicate to them what options are available to them in the narrative visualization?

1) Page navigation link: D3 code internally keep track of current scene location and call different WDI APIs and different function with parameter to draw different chart. In WorldBank WDI API side, there are three different indicator name depends on "total", "male", and "female", and JSON query in the code also need to be changed.

2) Radio box: gender radio box in scene 5 directly decide which parameter to use for data query.

3) country select list: Depends on which country viewer choose in scene 5, it also impacts the behavior of WDI API call, and D3 code either calls world data or each country's data.

Why my visualization is different than others?

My visualization is based on realtime WorldBank API calling and parse the JSON result. This behavior caused lots of data parsing and data sync issue but also taught me a lot more than I learned from the classes. e.g., how to pass extra function parameter to the callback function which usually takes only `d` or `d, i`. Also, the visualization page will work well with updated data on WDI Worldbank data change as my code don't use locally loaded CSV.

Future work

- There is still some display issue on the last page during overlapping the line chart.
- Also, internal chart drawing functions are flexible enough to take a query period(year) as a parameter, but I didn't have enough time to add it as query parameter, so the date is hardcoded from 2000 to 2017. It would be better with some range/scroller to allow viewers to control of it.