

Notes for Data Visualization Lecture Week 3

-Warm-up

- What is the visualization trying to show?
Rainfall affect the output of refinery
- What are its methods?
Geographical data, weather data, refinery database
- Weakness:
Color is hard to tell
Not sure what the refinery out of the area of the rainfall is about
Not clear about the rain is for one time or many times
- Outright error: One side of the map has two datasets, the other side has only one.

-Data formats

- Text:
ASCII: human readable text e.g. csv file
JSON: popular online infrastructure, very flexible
- Binary: not human readable
- Query-based: Web version of data format
SQL

-Organization:

- Row: take every filed for one entry, and line them up
- Column: go through all the heights, and then weights
- CSV: can have a header of metadata
- JSON: It's like XML for data storage, can store lots of different data types
ipynb has the same formatting as JSON
- HDF5: HDF5 readers has the ability to jump to the next field without reading anything before, don't have to wait for reading to happen. It's binary, no need to transform it.

-Operational Palette

- Filter: Take big data set and put into subset, the majority
Relationships
Subsampling