

# Week 9: Choosing an Engine

Spring 2017  
Matthew Turk

Broadcasting

[go.ischool.illinois.edu/meet2](https://go.ischool.illinois.edu/meet2)

# Evaluating Visualization Systems

- Costs
- Functionality
- Aesthetics

# Choices

- Can I get ahold of this software?
- Do I install it, or do I use it on a server?
- What's the user interface like?
- Is it declarative or is it procedural?

# License: Software

- What can you do with the software?
- Can you study the software?
- Who can you share it with?
- Who can you give your derivative works to?

# License: Software

- Copyleft: share and share-alike
- Non-copyleft: share, but don't necessarily need to share-alike
- <https://choosealicense.com/>

# License: Data

- What can you do with the data?
- How do you credit that data?
- Can the data be redistributed, remixed, modified?
- <http://opendefinition.org/guide/data/>
- <https://theodi.org/guides/publishers-guide-open-data-licensing>

# Accessibility

- Is the software installed locally on your machine?
- Is it hosted at a local or remote instance?
- Who owns the visualizations, and how is access to them controlled?



# Interface

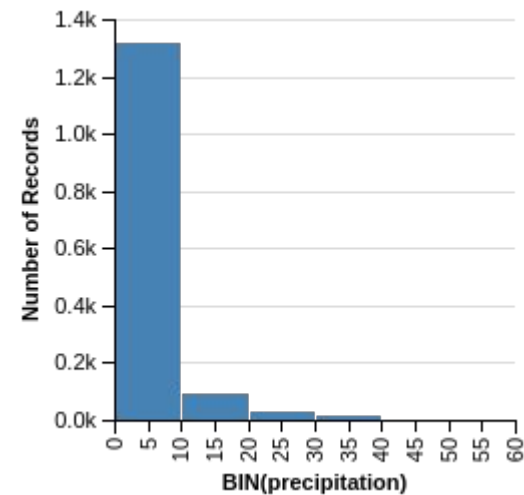
- How do you interact with the software?

# Declarative vs Procedural

- Declarative: how do you want the plot to look?
- Procedural: what are the steps to make the plot look that way?

# Declarative vs Procedural: Example Declarative

```
Chart(df).mark_bar().encode(  
    X('precipitation', bin=True),  
    Y('count(*) : Q')  
)
```



Altair documentation: <https://altair-viz.github.io/tutorials/exploring-weather.html>

# Evaluation: Costs

- Costs
- Functionality
- Aesthetics

# Evaluation: Aesthetics

- Costs
- Functionality
- Aesthetics

# Visualization Engines

- Matplotlib
- Bokeh
- Plotly
- D3 / Vega / Lyra
- Altair

# Visualization Engines: Matplotlib

- License: non-copyleft open source
- Construction of a *plot* is the focus
- Interactivity is possible
- Procedural
- Pros?
- Cons?

# Visualization Engines: Bokeh

- License: non-copyleft open source
- Construction of a set of visualizations is focus
- Interactivity is fundamental
- Mix of declarative and procedural
- Pros?
- Cons?



# Visualization Engines: Plotly

- License: non-copyleft open source & commercial
- Construction of a set of visualizations or dashboards is focus
- Interactivity is possible
- Web UI promotes interactivity
- Pros?
- Cons?

# Visualization Engines: D3 / Vega / Lyra

- License: non-copyleft open source
- Construction of a set of visualizations or dashboards is focus
- Interactivity is possible
- Web UI promotes interactivity
- Pros?
- Cons?

# Visualization Engines: Altair

- License: non-copyleft open source
- Construction of charts and visualizations
- Interactivity is coming
- Domain-specific language for reductions and transformations
- Pros?
- Cons?

# Project

- Groups of 3 or 4 (please choose people you don't know that well)
- Four visualizations or visualization systems
  - a. Procedural, bespoke visualization
  - b. Interactive, bespoke visualization
  - c. System for receiving new data
  - d. Narrative report around a dataset
- Projects due April 28, including report, full source code, and example outputs

# Project: Datasets

- Procedural, bespoke visualization
  - CUMTD dataset
  - Describe the travel times for different routes and distances
  - GTFS format: <https://developers.google.com/transit/gtfs/reference/>
  - Available at: <https://developer.cumtd.com/>
- Generate a single PNG or PDF

# Project: Datasets

- Interactive, bespoke visualization
  - IMLS grant data ([data.imls.gov](http://data.imls.gov))
  - Build out data explorer for grants
    - Amount
    - Institution (and location)
    - Purpose
- Bonus:
  - Fuse with similar data from NSF, NIH, NEA, NEH

# Project: System

- System for receiving new data
  - Data will be provided, and of fixed format
  - At least eight columns:
    - Name
    - Date
    - Latitude, longitude
    - Categorical label
    - 3 quantitative columns
  - Your system should accept a dataframe object
  - Your system must produce a set of plots describing the data
  - Bonus: interactivity
- Three sample datasets will be provided

# Project: Report

- Build a “report” for a set of data
  - Break down the data into different aspects
  - Describe each visualization
  - Use the CUMTD dataset to tell the story of delays
  - More guidance forthcoming
- You will provide a PDF of the report and the source code for the visualizations
- This should be stylized like an infographic, for qualitative rather than quantitative understanding