Process Book

Initial Questions (Answer based on Proposal):

- **Overview and Motivation:** Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.
    - **Overview:**
        - We want to figure out which factors are the greatest predictors of river discharge. These factors include precipitation or river discharge of previous weeks. Based on these predictions, we want to visualize which rivers are at the greatest risk of flood and also allow for users to explore each river's past history of precipitation and discharge and also allow the user insights into our prediction model such as feature importance and also the efficacy of our model for a given prediction.
    - **Motivation:**
        - We wanted to explore the applications of machine learning on environmental problems and we already had some data about river discharge. More specifically, discharge prediction can help predict extreme conditions which is useful for forecasting floods and drought-like conditions.
- **Related Work:** Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.
    - Charles' advisor (Dr. Randhir) suggested an ML model to predict discharge, using discharge from 1 week ago, 2 weeks ago, and precipitation today, 1 week ago and 2 weeks ago as inputs. We were curious to see the extent to which an ML model could predict discharge simply based on past observations and also learning how to visualize elements such as flood risk spatially with a map.
- **Questions:** What questions are you trying to answer with your visualization(s)?
    - Which factors are greatest predictors of river discharge?
    - What are the forecasted discharges of rivers in New England, given their predictors (such as precipitation from previous weeks and discharge from previous weeks) in the dataset? (The Outcome of the Model)
    - Which rivers in New England are at risk of severe conditions (drought or flood)?
    - How can we display each New England river's past precipitation and discharge?
    - How can we evaluate our model and figure out which rivers it is the most effective at predicting discharge for?
- **Data:** Include information about the source, how you collected it (e.g., web scraping), cleaning methods, etc.
    - Sources:
        - U.S. Geological survey for historic gauge discharge data
        - PRISM for precipitation and temperature historical data
    - Links:

    - https://waterdata.usgs.gov/nwis/rt
    - https://prism.oregonstate.edu/recent/

Questions to ask periodically:

- **Questions:** What questions are you trying to answer with your visualization(s)? How did these questions evolve/change over the course of the project? What new questions did you consider?
- **Exploratory Data Analysis:** What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
- **Design Evolution:** What are the different visualizations you considered? Did you deviate from your proposal, if so, why?
- **Focus on the important decisions you made, and be concise in what you choose to include in your process book.**
- **We strongly advise you to include many figures in your process book, including photos of your sketches of potential designs, screenshots from different visualization tools you explored, inspirations of visualizations you found online, etc.**

Meeting Notes Template

Tasks Completed from the previous meeting:

-

Questions Answered:

-

Work Log:

-

Tasks for next time:

- Milestone goals (Have nearly done by next meeting):
    -

Task completion summary:

- Charles
- John
- Nathaniel

Meeting 1: 3/28/25

Questions Answered:

- Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
    - We looked through the dataset using a pandas dataframe in a python notebook.

    - We learned that our dataset is too large to host on Supabase so we considered alternatives such as simply using python in our backend to store the data in a pandas dataframe and also preprocessing the dataset to remove unnecessary columns.
        - Removed elevation, site name, latitude and longitude and moved to a different table as these are a single value that corresponds to a gauge point and does not change over time.

Work Log:

```
df_no_name_2
[21] ✓ 0.0s
```

	Date	Mean Discharge (cubic ft/sec)	ppt (inches)	tmean (degrees F)	Site Number
0	2010-01-01	901.0	0.09	10.5	1010000
1	2010-01-02	893.0	0.09	19.2	1010000
2	2010-01-03	885.0	0.56	23.3	1010000
3	2010-01-04	872.0	0.06	31.3	1010000
4	2010-01-05	858.0	0.10	30.2	1010000
...	...	...	...	...	...
2173020	2023-12-27	NaN	0.00	41.0	413413071270400
2173021	2023-12-28	NaN	0.53	45.4	413413071270400
2173022	2023-12-29	NaN	0.66	44.2	413413071270400
2173023	2023-12-30	NaN	0.03	43.4	413413071270400
2173024	2023-12-31	NaN	0.00	40.7	413413071270400

2173025 rows × 5 columns

-
- Table with unnecessary columns removed, cleaned up for demonstrative purposes using pandas library in python.
- Organized data into two different tables, one for static data such as gauge point latitude and longitude and one for date-based data such as mean discharge.
- Decided to have a simple dropdown instead of an interactive map.
- Decided how to filter search based on dataset (Supabase).

Tasks for next time:

- Milestone goals (Have nearly done by next meeting):
    - Make a graph showing precipitation and discharge over time. (John)
    - Make a state column (Charles).
    - Making a dropdown for selecting by state then displays all possible gauge points. (Charles / Nathaniel)
    - Embed interactive map (Charles)
    - Figuring out data structures (Nathaniel)

Task completion summary:

- Charles:
    - Dataset preprocessing
    - Planned out database tables
- Nathaniel:
    - Build Supabase
    - Dataset preprocessing
    - Planned out database tables
    - Tried new database website (Railway)
- John:
    - Fixed Project proposal according to comments.
    - Set up a process book.

Meeting 2: 4/2/25

Tasks Completed from the previous meeting:

- Charles
    - Made a state column for the data
    - Embedded an interactive map
- Nathaniel
    - Finished registering for database

Questions Answered:

- **Exploratory Data Analysis:** What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
    - We realized the embedded map from ArcGIS Online didn't integrate well into an interactive html page. The elements within the map cannot be easily modified by changes on the html page. Further research into an alternative is necessary.

Work Log:

- Charles - uploaded data to group members
- Nathaniel - uploaded data to new database CloudFlare
- John - Created a drop down on the website

Tasks for next time:

- Milestone goals (Have nearly done by next meeting):
    - Gauge selector basic implementation on dummy data (John)
    - Finishing Wrangler setup (Nathaniel)
    - Adding new column for river name (Charles)

Task completion summary:

- Charles
    - Created a new column for the state a gauge point is located in and added an interactive map to the website.
- John
    - Began line graph based on dummy data to represent temperature mean over time for a given gauge point.
- Nathaniel
    - Found a platform (Wrangler) for hosting our database for free.
    - Started uploading data to the database.

Meeting 3: 4/9/25

Tasks Completed from the previous meeting:

- John: Began gauge filter, added selection for state using drop down.
- Nathaniel: Finished wrangler setup

Questions Answered:

-

Work Log:

- Nathaniel: Created API to call from database using URL
- John: Worked on creating table based on dummy data

- Charles: Creating line graph

Tasks for next time:

- Milestone goals (Have nearly done by next meeting):
    - John: Finish table
    - Charles: Finish line graph

Meeting 4: 4/11/25

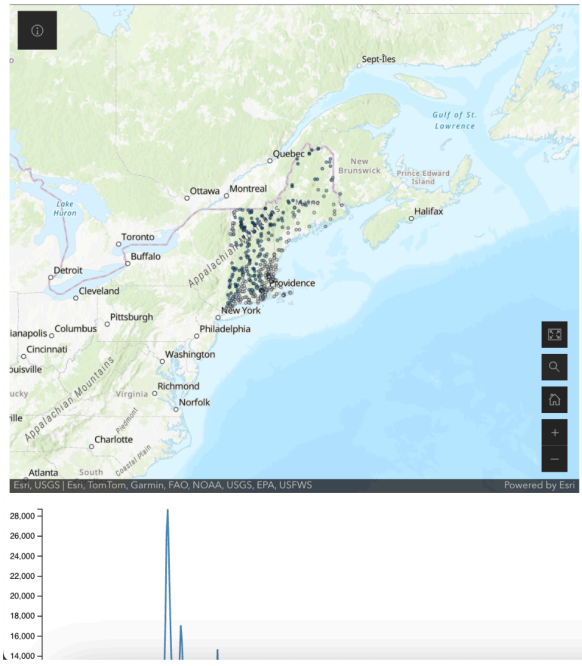
Tasks Completed from the previous meeting:

- John: Finished table
- Charles: Finished example line graph
- Nathaniel: Added additional queries to the database

Questions Answered:

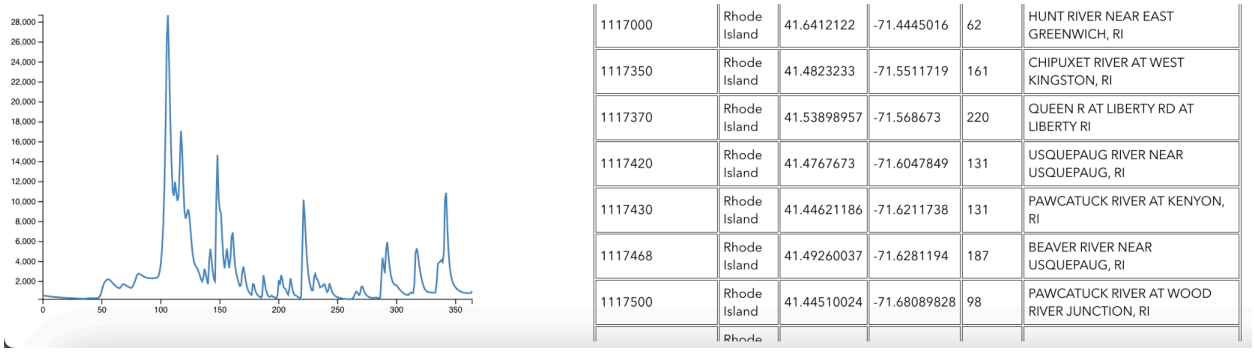
- **Design Evolution:** What are the different visualizations you considered? Did you deviate from your proposal, if so, why?
    - We settled on a layout of map in the top left, graph in the bottom left and table on the right
- **Questions:** What questions are you trying to answer with your visualization(s)? How did these questions evolve/change over the course of the project? What new questions did you consider?
    - We might shift our project to have more of an emphasis on river data visualization rather than analyzing the quality of a river model.

Work Log:

- 

State: Rhode Island

site_number	state	latitude	longitude	elevation	site_name
1109403	Rhode Island	41.83093377	-71.3503315	56	TEN MILE R., PAWTUCKET AVE. AT E. PROVIDENCE, RI
1111300	Rhode Island	41.9812093	-71.68590049	476	NIPMUC RIVER NEAR HARRISVILLE, RI
1111500	Rhode Island	41.99648716	-71.56200758	262	BRANCH RIVER AT FORESTDALE, RI
1112500	Rhode Island	42.0064871	-71.5025613	210	BLACKSTONE RIVER AT WOONSOCKET, RI
1113895	Rhode Island	41.8884331	-71.38144467	79	BLACKSTONE R AT ROOSEVELT ST AT PAWTUCKET RI
1114000	Rhode Island	41.833989	-71.41061208	69	MOSHASSUCK RIVER AT PROVIDENCE, RI
1114500	Rhode Island	41.8589884	-71.4872823	243	WOONASQUATUCKET RIVER AT CENTERDALE, RI
1115098	Rhode Island	41.8525992	-71.6061754	436	PEEPTOAD BROOK AT ELMDALE RD NR NORTH SCITUATE, RI
1115110	Rhode Island	41.8456548	-71.6095088	436	HUNTINGHOUSE BK AT ELMDALE RD AT N SCITUATE, RI
1115114	Rhode Island	41.83759929	-71.6120089	436	RUSH BROOK NEAR ELMDALE RD NEAR NORTH SCITUATE, RI
1115170	Rhode Island	41.84093269	-71.584508	354	MOSWANSICUT STREAM NR NORTH SCITUATE, RI
1115183	Rhode Island	41.79759966	-71.5847857	374	QUONOPAUG BK AT RT 116 NR NORTH SCITUATE, RI
1115185	Rhode Island	41.83621005	-71.72256829	469	WINSOR BROOK AT WINSOR RD NEAR SOUTH FOSTER, RI



1117000	Rhode Island	41.6412122	-71.4445016	62	HUNT RIVER NEAR EAST GREENWICH, RI
1117350	Rhode Island	41.4823233	-71.5511719	161	CHIPUXET RIVER AT WEST KINGSTON, RI
1117370	Rhode Island	41.53898957	-71.568673	220	QUEEN R AT LIBERTY RD AT LIBERTY RI
1117420	Rhode Island	41.4767673	-71.6047849	131	USQUEPAUG RIVER NEAR USQUEPAUG, RI
1117430	Rhode Island	41.44621186	-71.6211738	131	PAWCATUCK RIVER AT KENYON, RI
1117468	Rhode Island	41.49260037	-71.6281194	187	BEAVER RIVER NEAR USQUEPAUG, RI
1117500	Rhode Island	41.44510024	-71.68089828	98	PAWCATUCK RIVER AT WOOD RIVER JUNCTION, RI

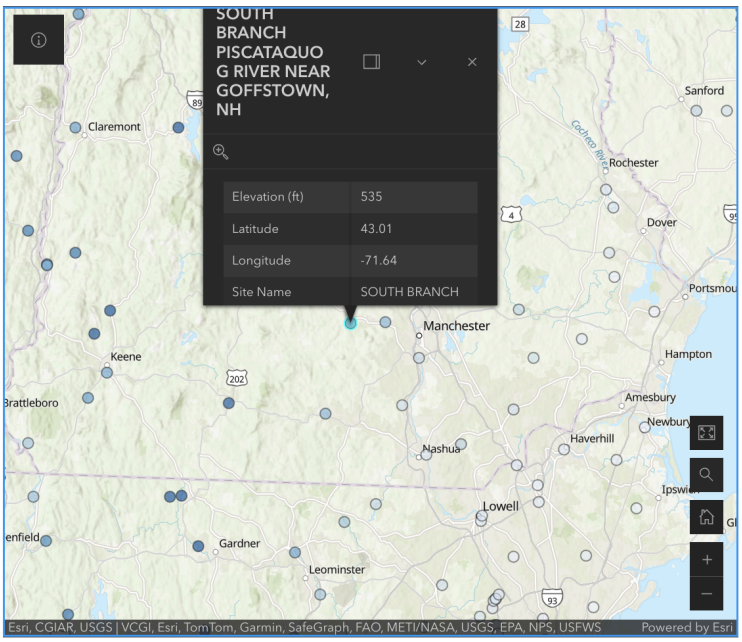
- The above screenshots show our visualizations in their current state. In the top-left, we have an interactive map which allows users to see the locations of gauge points spatially and gather basic information, such as elevation, longitude and longitude, by clicking on certain gauge points.
- On the right we have a filter selector, which allows users to see all gauge points in a given state.
- On the bottom left we have a sample graph visualization precipitation at a predetermined gauge point at a specified date range. We will link this graph to the table to allow for the user to select a specific gauge point and we will also add a date range selector component onto the website.
- Looked into alternative web maps.
- Table now responds to dropbox filtering by state and uses our actual data.

Tasks for next time:

- Milestone goals (Have nearly done by next meeting):
    - Give line graph ability to update based on select gauge point.
    - Add date range selector and link to the graph.
    - Use a leaflet map for visualization (instead of ArcGIS online), have the map update as selection changes

Milestone Retrospective:

- **Implementation:** Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

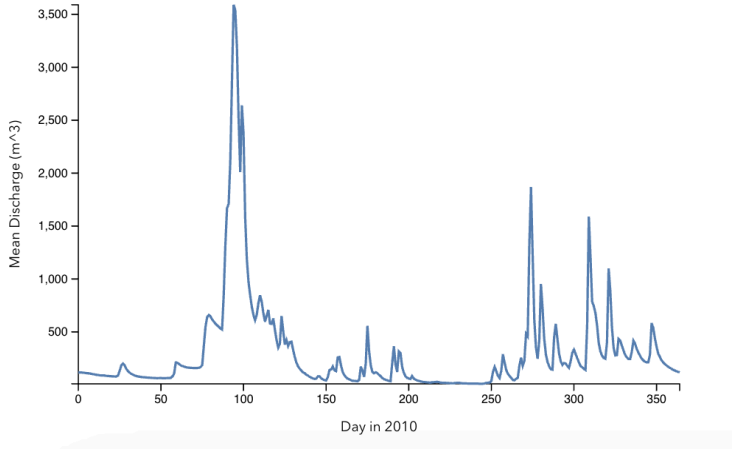
- 

- This is our interactive map, which allows for users to explore the locations of gauge points spatially and click on them to gather basic information such as elevation, latitude, longitude, etc. Users can also search for specific locations and enter full screen mode for easier viewing.
- State: Massachusetts

site_number	state	latitude	longitude	elevation	site_name
1094400	Massachusetts	42.57620079	-71.7881285	518	NORTH NASHUA RIVER AT FITCHBURG, MA
1094500	Massachusetts	42.49506389	-71.7219333	348	NORTH NASHUA RIVER NEAR LEOMINSTER, MA
1095220	Massachusetts	42.41092507	-71.7911829	489	STILLWATER RIVER NEAR STERLING, MA
1095375	Massachusetts	42.37286967	-71.8281279	620	QUINAPOXET RIVER AT CANADA MILLS NEAR HOLDEN, MA
1095434	Massachusetts	42.36453675	-71.775349	509	GATES BROOK NEAR WEST BOYLSTON, MA
1095503	Massachusetts	42.41944444	-71.6661111	364	NASHUA RIVER, WATER STREET BRIDGE, AT CLINTON, MA
1096000	Massachusetts	42.63425619	-71.6578479	351	SQUANNACOOK RIVER NEAR WEST GROTON, MA
1096500	Massachusetts	42.66758945	-71.57506809	243	NASHUA RIVER AT EAST PEPPERELL, MA
1097000	Massachusetts	42.432038	-71.4497848	207	ASSABET RIVER AT MAYNARD, MA
1097300	Massachusetts	42.51259289	-71.40422829	207	NASHOBA BROOK NEAR ACTON, MA
1098500	Massachusetts	42.31514167	-71.3838083	167	COCHITUATE BK BL LAKE COCHITUATE AT FRAMINGHAM, MA
1098530	Massachusetts	42.3253732	-71.3975605	187	SUDBURY RIVER AT SAXONVILLE, MA

- Our table allows users to search for all gauge points within the selected state, and view basic information such as site number, latitude, longitude, etc. This is

meant to allow for a quicker but less intuitive method to gather information about a specific gauge point.



- Our line graph depicts the amount of mean discharge for dummy data over the course of 2010. This will later be linked with the table such that a gauge point can be selected and subsequently a date range can be selected instead of being hard-coded.
- This visualization is meant for analyzing rivers' patterns of discharge and will later be revamped to visualize other patterns such as precipitation over time.

- **Evaluation:** What did you learn about the data by using your visualizations? How did you answer your research questions? How well does your visualization work, and how could you further improve it? (Be honest here. Limitations are a part of any project, and they will be noticeable during the grading process. Acknowledging them in this section indicates thoughtfulness in your design process, and, as such, will only help your grade.)
  - Since our visualizations are still in development, not much can be analyzed about the rivers in our website's current state.
  - Our visualizations are effective for basic functions, however we still need to build upon them more by adding linking and adding more interactive components such as allowing users to define a date range and a river to link the graph to.

Final Answers (Answer after project is complete):

- **Questions:** What questions are you trying to answer with your visualization(s)? How did these questions evolve/change over the course of the project? What new questions did you consider as the project progressed?
- **Design Evolution:** What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?
- **Implementation:** Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.
- **Evaluation:** What did you learn about the data by using your visualizations? How did you answer your research questions? How well does your visualization work, and how

could you further improve it? (Be honest here. Limitations are a part of any project, and they will be noticeable during the grading process. Acknowledging them in this section indicates thoughtfulness in your design process, and, as such, will only help your grade.)