



Identify research questions

- Is analyte concentration higher in one catchment than the other?
- Is analyte load higher in one catchment than the other?
- How do flow, ppt, and temp affect analyte load?

List of questions and raw data

Do we need weather data?

Data scrapping

- Analyte concentration data
- Flow data
- Precipitation data

Data gathering/  
scrapping

Preliminary analysis

Descriptive analysis  
Plot visualization

Drop the outliers

Can we answer the questions by looking at the dataset differently?

Do they answer research questions?

Identify the possible reasons and future recommendations

Final report

Develop ML to replace missing data

Fill in missing data

Any issues with the dataset?

Is the outliers valid ?

- Normality test
- T-test/Wilcoxon test
- Correlation test
- Calculate load
- Annual cumulative ppt

Prepare data for analysis

Need normal dataset?

Normal?

Log transform

Normal?

Use original data with caution note

Data analysis

Load calculation

T-test/  
Wilcoxon test

Dimension reduction (PCA)

Analysis outputs