PFAS Dashboard Project

# Semester 1: Foundational Analysis & Static Visualization

**Aims:** By the end of the semester, main goals of the project will be to have a solid understanding and familiarity with the PFAS data from the [Michigan PFAS Action Response Team](https://www.michigan.gov/pfasresponse) (MPART), several effective data visualizations, several well-organized summary tables showcasing key metrics or takeaways, a draft dashboard created using Quarto containing some of these visualizations and summary tables, and a well-organized and well-documented GitHub project repository prepared for organizing the project and deploying the dashboard to showcase and facilitate exploration of the Michigan PFAS data related to drinking water.

## Month 1: Orientation & Exploration (Weeks 1-4, August 27th - September 24th)

* **Student Focus/Objectives:** Understand the project's scientific and public health context, establish a reproducible workflow by creating a GitHub repo (Michigan-PFAS) for the project adding Andrew (@dilernia) as a collaborator, and become familiar with the data.
* **Key Tasks:**
  + **Context & Literature Review:** Review the [Environmental Protection Agency (EPA) PFAS regulations](https://www.epa.gov/sdwa/and-polyfluoroalkyl-substances-pfas) and [Michigan PFAS regulations](https://www.michigan.gov/pfasresponse/drinking-water/mcl) for drinking water. Explore the [MPART website](https://www.michigan.gov/pfasresponse) and related [literature](https://drive.google.com/drive/folders/1IvjtLwF91auyxnWrILYKXimr_rIdz_Z2?usp=drive_link) regarding PFAS pollution.
  + **Technical Setup:**
    - Create a GitHub account if needed, and then create a new repository on [GitHub.com](http://github.com) for the project called PFAS or Michigan-PFAS.
    - Create a README file including motivation for the project, sources of the data, and a data dictionary by first creating a Quarto file for the README (README.qmd), and then using *quarto render README.qmd --to markdown* in a Terminal application to create the corresponding [README.md](http://readme.md) to be uploaded to the repo on GitHub.
    - Below is a proposed structure for the README:

### Michigan PFAS Pollution Data Dashboard

#### Introduction

In this section, we will provide important background information and the motivation for the dashboard (PFAS health implications, PFAS pollution in Michigan, and the importance of making this data easy to access and explore). We will describe the source of the data here as well, but not significant details.

#### Dashboard

**View the Live Dashboard Here (this will be a clickable link at some point)**

*(We will replace the link above with our actual GitHub Pages URL once available)*

#### Key Features

A brief description (possibly a bulleted list) of the visualizations, tables, and other features available in the dashboard.

*(We will update this section once the dashboard is more finalized)*

#### Data Source

A description of all data sources with formal citations and clickable URLs that link directly to where the raw data can be obtained. This section or the Introduction section would likely be a good place for the data dictionaries.

#### Technologies Used

* **Frontend**: *(e.g., Leaflet, Plotly, but we will update this section once the dashboard is more finalized)*
* **Backend**: R and Quarto
* **Hosting**: GitHub Pages

#### Attribution & Licensing

* **Data Providers:** Briefly indicate the sources of data (e.g., MPART), and indicate that the dashboard is an independent project and is not affiliated with or endorsed by MPART or the state of Michigan.
* **Project License:** The code for this project is licensed under the [MIT License](https://www.google.com/search?q=LICENSE).
* **Acknowledgments:** This project was created by Nancy Odhiambo and funded as part of a Master's Graduate Research Assistant position in the Department of Statistics at Grand Valley State University. Dr. Andrew DiLernia served as the faculty mentor for this project.
  + **Exploratory Data Analysis (EDA):** Use Quarto to import and conduct exploratory data analysis of the PFAS data related to drinking water (01-exploration.qmd).
    - Use packages from the tidyverse for data loading and manipulation.
    - Use the skimr package to generate detailed summary statistics.
    - Use the naniar package to visualize and quantify patterns of missingness in the data to understand its integrity.
    - Create and maintain a data dictionary in the project's README.md file based on the data dictionary created in the [PFAS-data.qmd](https://drive.google.com/file/d/1PYI6GZNqEiCoVSlVddH8501lRzg1XAnX/view?usp=sharing) file.
* **Potential Deliverables:**
  + A GitHub repository with a README.md file outlining project goals and the data dictionary.
  + A Quarto document (01-exploration.qmd) containing the initial EDA with summaries of missingness.

## Month 2: Deeper Analysis & Key Metric Definition (Weeks 5-8, September 24th - October 15th)

* **Student Focus/Objectives:** Begin conducting more targeted analyses, defining metrics based on [federal](https://www.epa.gov/sdwa/and-polyfluoroalkyl-substances-pfas) and [state](https://www.michigan.gov/pfasresponse/drinking-water/mcl) PFAS regulations for drinking water that will constitute key aspects of the dashboard.
* **Key Tasks:**
  + **Focused Analysis:** Using Quarto (02-analysis.qmd), analyze distributions of key analytes and explore relationships between the contamination levels of different analytes.
  + **Define Key Numerical Summaries:** Define and calculate summaries based on EPA and state-level limits:
    - Summarize the amount of samples exceeding the PFOA and PFOS limits as specified in the state and federal regulations.
    - Identification of sites with the highest concentrations and greatest amount of samples exceeding specified limits.
  + **Static Visualizations:**
    - Use ggplot2 to create a core set of static plots.
    - Create a histogram of PFOA/PFOS concentrations with vertical lines (geom\_vline()) at the regulation limits.
    - Develop a bar chart showing the number of sites per county that exceed the regulation limits.
    - Possibly create a map of locations of interest using leaflet. This [demo](https://rpubs.com/dilernia/maps) could be used to aid in creating maps
* **Potential Deliverables:**
  + A Quarto document (02-analysis.qmd) complete with static visualizations, summary tables, and interpretations of the results with respect to the federal and state regulations. The dashboard should link to MPART or EPA resources directly with formal references.

## Month 3 & 4: Dashboard Structuring & Publishing (Weeks 9-15, October 15th - December 3rd)

* **Student Focus/Objectives:** Build the dashboard in Quarto and publish it to GitHub Pages.
* **Key Tasks:**
  + **Quarto Dashboard Layout:**
    - Create a Quarto dashboard project, creating a primary .qmd file for the dashboard.
    - Brainstorm the design and implement the dashboard's layout, planning the placement of key summary tables and data visualizations.
    - Include accompanying text with explanations and interpretations of visualizations and summary tables, defining key metrics, providing key takeaways, and including formal citations when possible to the EPA and state of Michigan’s websites and formal documentation regarding regulation limits and data sources.
    - Investigate and properly include licensing agreements regarding data in the dashboard, adhering to requirements from the MPART website if there are any.
  + **Publishing to GitHub Pages:**
    - Configure the project to render the Quarto dashboard to a docs folder.
    - Upload project files and organize them using the GitHub repository.
    - Enable GitHub Pages in the repository settings to serve the site.
* **Potential Deliverables:**
  + A Quarto dashboard containing key metrics showcased via summary tables and visualizations.
  + A public link to the dashboard hosted on GitHub Pages.