Project: Rainfall-Runoff Relationship

Project Overview

In hydrology, understanding how **rainfall translates to runoff** is essential for flood control, drainage design, and watershed planning. In this project, you'll analyze synthetic rainfall and runoff data across multiple watershed zones, each with different land cover, slope, and soil type. Your goal is to uncover patterns that explain why some areas generate more runoff than others under similar rainfall.

Objective

You'll work with a **synthetic dataset** of rainfall and runoff across different sub-watersheds. Each sub-area includes descriptors like **land cover**, **average slope**, and **soil type**. Your analysis should explore how runoff varies with storm intensity and watershed characteristics.

Major Questions for Your Proposal

1. **Runoff Drivers**
   Which variables (e.g., impervious surface %, soil type, slope) do you think will have the strongest effect on runoff? Why?

2. **Analysis Approach**

   - Will you group by rainfall event? Compare across land types? Explore seasonal or geographic differences?
   - How will you handle missing values?
   - Will you model the runoff with a regression model? Will you try to predict runoff based on rainfall?

3. **Visualization Plan**

   - What figures will you use to summarize and/or explore the data?
   - How will you show relationships between rainfall and runoff? Will you use scatter plots, grouped bars, or time series visuals?

Project Definitions

Input Variables

- **zone_id**: Sub-watershed identifier
- **date**: Date of rainfall event
- **rainfall_mm**: Total precipitation during the event
- **impervious_pct**: % of the zone covered by roads, roofs, pavement
- **slope_pct**: Average land slope in the watershed (%)
- **soil_type**: Hydrologic soil group (A, B, C, D – with A being high infiltration, D being poor infiltration)
- **land_use_type**: Residential, Commercial, Forest, Agriculture
- **season**: Season when the rainfall occurred

Output Variables

- **runoff_mm**: Observed or estimated surface runoff

Data Source

You'll be provided a **synthetic dataset** representing runoff responses across zones during a range of storm events. The dataset will include built-in relationships (e.g., impervious zones = higher runoff), allowing you to test hypotheses and surface trends.

Project Requirements

1. **Data Cleaning & Grouping**

   - Group events by various input variables (e.g., zone, season, land use)
   - Consider filtering out unrealistic values and/or aggregating by storm size

2. **Exploratory Analysis**

   - Identify how runoff varies by input variables
   - Consider relationships between input variables - are there correlations?
   - Derive summary metrics based on input variables (e.g., average rainfall/runoff per zone, season, etc.)
   - Consider deriving additional metrics that might be useful for analysis such as runoff ratios or runoff coefficients.

3. **Visualization**

   - Visualize relationships between input variables and across input variables and runoff to help you understand the data and form hypotheses to explore further
   - Visualize any summary and/or derived metrics (e.g., runoff ratios)

4. **Discussion**

   - What were some of the biggest contributors to runoff? Why do you think they are important? Are there ways to control for them?
   - Do your findings align with what you expected in your proposal?
   - What might a civil engineer do with this information?