Programming for Data Science
Lab Assignment – 4

Experiment Title: Optimizing Urban Flood Control using R Programming

Total Marks: 100

Due Date: 06-08-2025

Submission Mode: Upload a single lab report (PDF or DOCX) to the LMS

Objective

To simulate and analyze flood control strategies in an urban environment using datasets created with vectors and lists, and to apply built-in and user-defined functions to assess risks and propose effective mitigation strategies.

Problem Statement

Urban flooding is becoming increasingly common due to factors such as heavy rainfall, insufficient drainage systems, and a lack of green cover. As a data analyst, you are tasked with simulating flood control data for a city consisting of at least 20 urban zones.

You must build a dataset using vectors and lists only, and perform various analysis tasks using built-in and user-defined functions to evaluate flood risk levels and suggest solutions.

Tasks to be implemented (in R)

Append the last four digits to all variables and user-defined functions you are using in your implementation

1. Dataset Creation (20 Marks)

Create a dataset using vectors representing each of the following attributes, and store all vectors inside a list named city_flood_data. Each attribute must have at least 20 entries.

Attributes:
- zone_name: Names of the zones (character vector)
- rainfall_mm: Rainfall received (numeric vector)
- drainage_capacity: Drainage capacity (numeric vector)
- population: Total population in the zone (numeric vector)
- water_logging_cm: Simulated water logging levels (numeric vector)

- green_cover_percent: Percentage of green cover (numeric vector)

 You must define and use a user-defined function named display_data_structure() that performs the following:
- Prints all attributes in the dataset
- Displays the structure of city_flood_data using str()

2. Built-in Functions Usage (15 Marks)

Use appropriate built-in R functions to:
- Identify the zone with the highest rainfall
- Compute the average water logging level
- Identify zones with below-average green cover
- Sort zones in descending order of population

3. User-defined Functions (20 Marks)

Define and apply the following two functions for all 20 zones:
1. assess_flood_risk(rainfall, drainage) → Returns "High", "Moderate", or "Low" based on the difference between rainfall and drainage capacity.
2. suggest_mitigation(green_cover, population) → Returns a relevant mitigation suggestion such as "Install Rain Garden", "Create Green Roofs", "Increase Tree Plantation", etc.

Store the function outputs into two new list items:
- risk_level
- mitigation_suggestion

4. Vector Operations (15 Marks)

- Reduce water_logging_cm by 20% for zones where green_cover_percent > 30%
- Create a new vector called severity_index using the formula:
  severity_index = (rainfall_mm - drainage_capacity + adjusted_water_logging_cm) / population

Store both adjusted_water_logging_cm and severity_index inside city_flood_data.

5. Filtering and Summary (10 Marks)

- Identify and list all zones where severity_index > 0.05
- Create and display a summary for these zones that includes:
  - Zone name
  - Severity index
  - Suggested mitigation strategy

Lab Report Structure (Submit as a Single PDF or DOCX on LMS)

File Name Format: Experiment_4_YourName_YourRegNo.pdf or .docx

Your report should include the following sections:

1. Title Page – Student Name, Reg. No., Experiment Title, Date
2. Objective
3. Rephrased Problem Statement
4. Dataset Description – Describe your data and how you generated it
5. Code Implementation – With detailed comments for clarity
6. Output Summary – Results from built-in and user-defined functions
7. Complete Dataset and Structure Output (via display_data_structure())
8. Filtered High-Risk Zone Summary
9. Conclusion and Recommendations
10. Appendix – Include the full R code and any additional information about your R program at the end of the document.

Evaluation Rubric (Total: 100 Marks)

Component	Marks
Dataset creation with at least 20 entries per attribute	20
Use of built-in functions	15
User-defined functions (assess_flood_risk, suggest_mitigation)	15
Dataset display function (display_data_structure)	5
Vector operations and severity index calculation	15
Filtering and zone summary	10
Code quality, structure, and readability	10
Lab report formatting and explanation	10

Instructions

- Use only vectors and lists. Do not use data frames or external files.
- Simulate at least 20 entries for each attribute.
- Submit a single lab report file to the LMS in PDF or DOCX format.
- Ensure your code is well-commented and understandable.
- Plagiarism or AI-generated content submitted without understanding will receive zero marks.

Expected OUTPUT: (Output can be enhanced based upon your Analysis)

```
----- City Flood Dataset -----
$zone_name
 [1] "Zone 1"  "Zone 2"  "Zone 3"  "Zone 4"  "Zone 5"  "Zone 6"  "Zone 7"
 [8] "Zone 8"  "Zone 9"  "Zone 10" "Zone 11" "Zone 12" "Zone 13" "Zone 1
4"
[15] "Zone 15" "Zone 16" "Zone 17" "Zone 18" "Zone 19" "Zone 20"

$rainfall_mm
 [1]  63 362  51 440 384 126 316 400 190 320 481  53 162  86 404 300   9
[18] 140 268 422

$drainage_capacity
 [1] 463 234 194  75 174  95  63 107 447  27  33 320 331 308  60 290  60
[18] 195 167 265

$population
 [1] 3445 1251 8716 2157 1553 7269 4485 3078 1166 7968 3312 7908 8858
 1017
[15] 4292 4098 1737 9312 5188 7765

$water_logging_cm
 [1] 127  63 149 194  94  59  72 133  50 167 145  97 101 193 142 186 173
[18]  66 180 194

$green_cover_percent
 [1] 32 17 21 41 31 27 13 39 22 16 13 12 31 39 20 24 30 27 35 14


----- Structure of the Dataset -----
List of 6
 $ zone_name          : chr [1:20] "Zone 1" "Zone 2" "Zone 3" "Zone 4" ...
 $ rainfall_mm        : int [1:20] 63 362 51 440 384 126 316 400 190 320 ...
 $ drainage_capacity  : int [1:20] 463 234 194 75 174 95 63 107 447 27 ...
 $ population         : int [1:20] 3445 1251 8716 2157 1553 7269 4485 3078 1
166 7968 ...
 $ water_logging_cm   : int [1:20] 127 63 149 194 94 59 72 133 50 167 ...
 $ green_cover_percent: int [1:20] 32 17 21 41 31 27 13 39 22 16 ...
```

Zone with Highest Rainfall: Zone 11
Average Water Logging (cm): 129.25
Zones with Below-Average Green Cover:
 [1] "Zone 2" "Zone 3" "Zone 7" "Zone 9" "Zone 10" "Zone 11" "Zone 12
"
 [8] "Zone 15" "Zone 16" "Zone 20"
Zones Sorted by Population (Descending):
 [1] "Zone 18" "Zone 13" "Zone 3" "Zone 10" "Zone 12" "Zone 20" "Zone 6
"
 [8] "Zone 19" "Zone 7" "Zone 15" "Zone 16" "Zone 1" "Zone 11" "Zone 8
"
[15] "Zone 4" "Zone 17" "Zone 5" "Zone 2" "Zone 9" "Zone 14"

--- High Risk Zone Summary (Severity Index > 0.05) ---
    Zone Severity_Index   Mitigation
1 Zone 2   0.1527 Increase Tree Plantation
2 Zone 4   0.2412   Install Rain Gardens
3 Zone 5   0.1836   Install Rain Gardens
4 Zone 7   0.0725 Increase Tree Plantation
5 Zone 8   0.1298   Install Rain Gardens
6 Zone 10   0.0577 Increase Tree Plantation
7 Zone 11   0.1790 Increase Tree Plantation
8 Zone 15   0.1132 Increase Tree Plantation
9 Zone 17   0.0702   Install Rain Garden