



**CHANDIGARH
UNIVERSITY**
Discover. Learn. Empower.

University Institute Of Computing

SUBJECT-R PROGRAMING

PROJECT- Household Water Usage Analysis



R Programming

SUBMITTED BY:-

STUDENT NAME:- SAMIKSHA

BRANCH:- BCA(DATA SCIENCE)

SUB CODE:- 24CAP-161

UID:- 24BCD10162

SECTION:-24BCD3

SEMESTER:- SECOND

SUBMITTED To:-

TEACHER NAME:- MS. SHIVANI CHADHA

Household Water Usage Analysis

Project Title: Analysis of Monthly Water Usage Across Households and Regions

Introduction

This report presents an analysis of average monthly water consumption (in liters) across different households in various regions (North, South, East, West). The aim is to identify usage patterns, regional variations, and households with the highest or lowest consumption.

Dataset

The dataset includes 5 households and their average water usage across four regions:

Household	North	South	East	West
Smith	1200	1400	1300	1250
Johnson	1500	1600	1550	1480
Williams	1100	1200	1150	1080
Brown	1300	1350	1320	1290
Jones	1250	1280	1220	1200

Methodology

Using R and the following libraries:

- ggplot2 for visualization
- dplyr and tidyr for data manipulation

We:

- Calculated mean, median, and standard deviation for each household
- Visualized total and regional water usage
- Identified max/min usage regions per household
- Analyzed regional water consumption patterns

Visualizations

1. Total Monthly Water Usage by Household (Bar chart showing overall usage per household)

2. Water Usage Distribution by Region (Histogram showing variation within each region)
3. Water Usage Spread by Household (Boxplot comparing range and outliers per household)
4. Total Water Usage by Region (Line chart showing cumulative usage by region)

Code

```
library(ggplot2)

library(dplyr)

library(tidyr)


# Average monthly household water usage (in liters) across regions
water_usage_data <- data.frame(

  Household = c("Smith", "Johnson", "Williams", "Brown", "Jones"),
  North = c(1200, 1500, 1100, 1300, 1250),
  South = c(1400, 1600, 1200, 1350, 1280),
  East = c(1300, 1550, 1150, 1320, 1220),
  West = c(1250, 1480, 1080, 1290, 1200)
)

print("Monthly Water Usage Data (Liters):")

print(water_usage_data)

# Summary statistics
summary_stats <- water_usage_data %>%

  pivot_longer(-Household, names_to = "Region", values_to = "Water_Usage") %>%

  group_by(Household) %>%

  summarise(

    Mean = mean(Water_Usage),

    Median = median(Water_Usage),

    SD = sd(Water_Usage)
```

```

)

print("Summary Statistics by Household:")

print(summary_stats)

# Convert to long format for plotting

long_water <- pivot_longer(water_usage_data, -Household, names_to = "Region", values_to =
"Water_Usage")

# Total water usage by household

ggplot(long_water %>% group_by(Household) %>% summarise(Total =
sum(Water_Usage)),

  aes(x = reorder(Household, -Total), y = Total, fill = Household)) +

  geom_bar(stat = "identity") +

  theme_minimal() +

  ggtitle("Total Monthly Water Usage by Household")

# Histogram by region

ggplot(long_water, aes(x = Water_Usage, fill = Region)) +

  geom_histogram(binwidth = 50, position = "dodge") +

  facet_wrap(~Region) +

  theme_minimal() +

  ggtitle("Water Usage Distribution by Region")

# Boxplot by household

ggplot(long_water, aes(x = Household, y = Water_Usage, fill = Household)) +

  geom_boxplot() +

  theme_minimal() +

  ggtitle("Water Usage Spread by Household")

# Max and min usage per household

max_usage <- long_water %>%

  group_by(Household) %>%

```

```

filter(Water_Usage == max(Water_Usage))

min_usage <- long_water %>%

  group_by(Household) %>%

  filter(Water_Usage == min(Water_Usage))

print("Highest Usage Region per Household:")

print(max_usage)

print("Lowest Usage Region per Household:")

print(min_usage)

# Total usage by region

region_summary <- long_water %>%

  group_by(Region) %>%

  summarise(Total_Usage = sum(Water_Usage))

ggplot(region_summary, aes(x = Region, y = Total_Usage, group = 1)) +

  geom_line(size = 1.2, color = "blue") +

  geom_point(size = 3) +

  theme_minimal() +

  ggtitle("Total Water Usage by Region")

```

Key Findings

- Johnson household had the highest total water usage.
- Williams consistently showed the lowest usage across all regions.
- South region reported the highest cumulative water usage overall.
- North and West regions had more variability in household usage.

Conclusion

The analysis provides clear insights into household water consumption patterns, highlighting opportunities for conservation efforts in high-usage areas like the South. Future analysis could include time-series trends and comparisons across different months or seasons.



