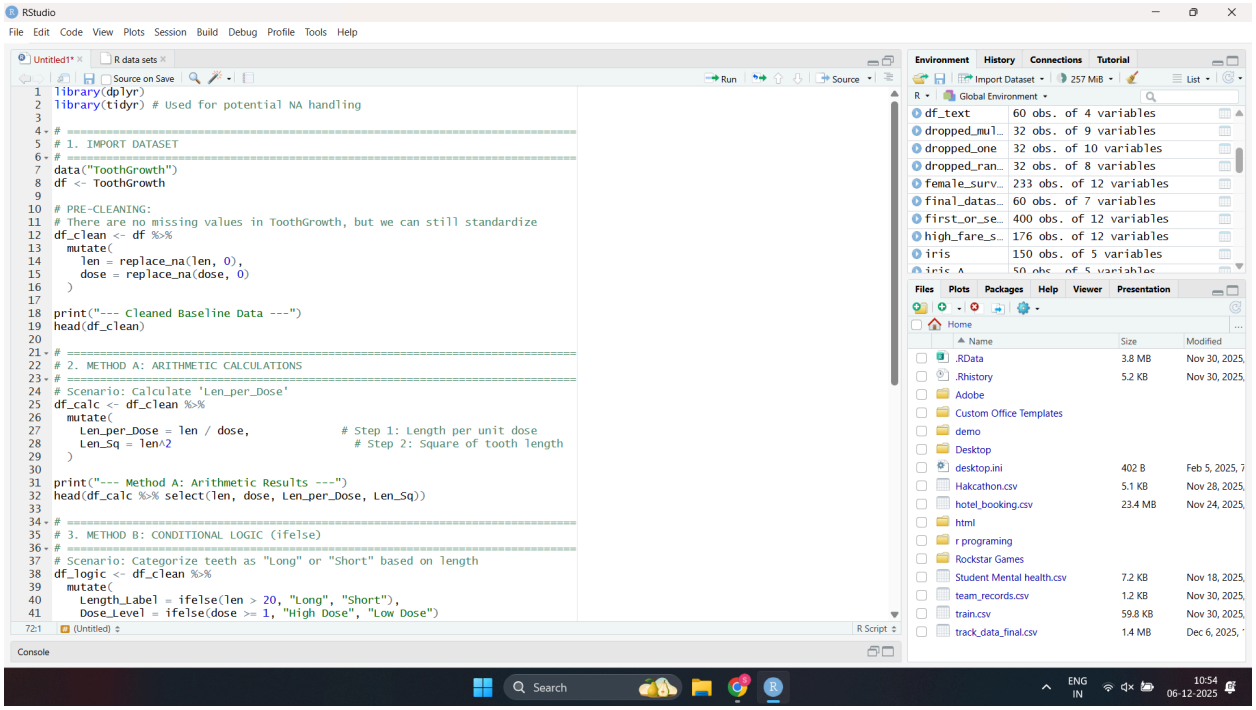
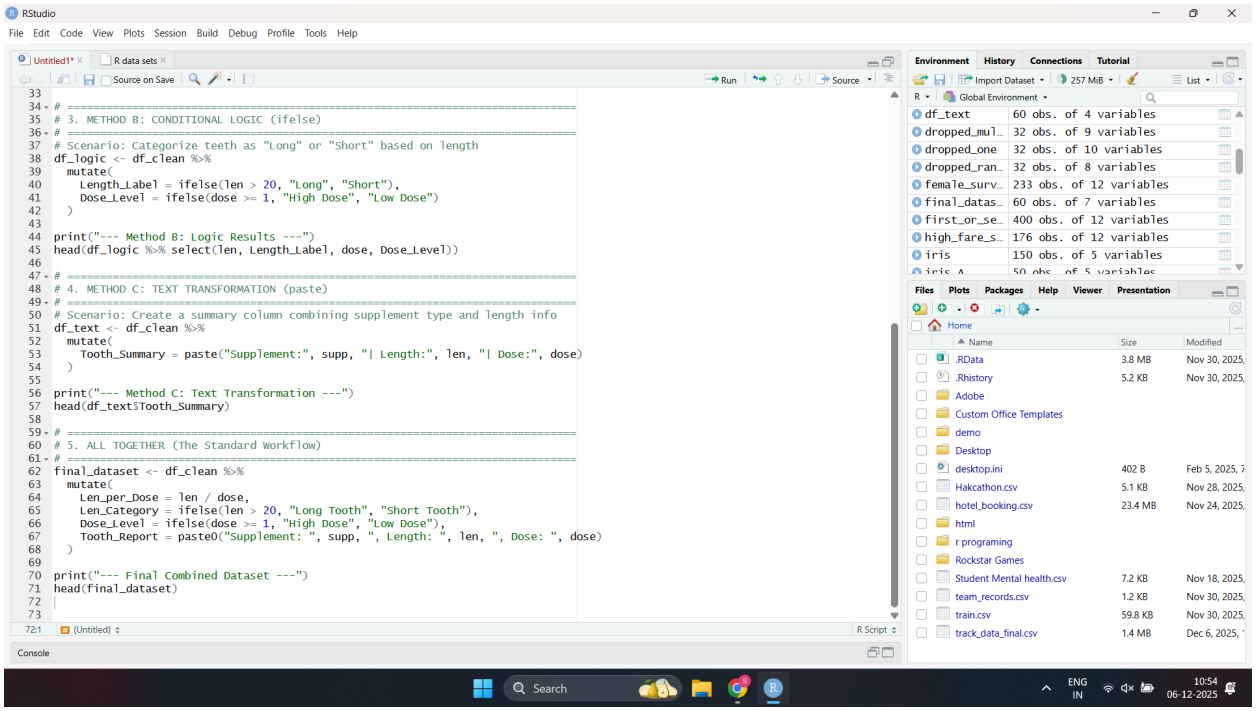
SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis SAS / SPSS / R

Aim - Creating new variables using transformations and calculations in R. import dataset.

Input :



```
library(dplyr)
library(tidyr) # Used for potential NA handling

# ==========================================================================
# 1. IMPORT DATASET
# ==========================================================================
data("ToothGrowth")
df <- ToothGrowth

# PRE-CLEANING:
# There are no missing values in ToothGrowth, but we can still standardize
df_clean <- df %>%
  mutate(
    len = replace_na(len, 0),
    dose = replace_na(dose, 0)
  )

print("--- Cleaned Baseline Data ---")
head(df_clean)

# ==========================================================================
# 2. METHOD A: ARITHMETIC CALCULATIONS
# ==========================================================================
# Scenario: Calculate 'Len_per_Dose'
df_calc <- df_clean %>%
  mutate(
    Len_per_Dose = len / dose,              # Step 1: Length per unit dose
    Len_Sq = len^2                          # Step 2: Square of tooth length
  )

print("--- Method A: Arithmetic Results ---")
head(df_calc %>% select(len, dose, Len_per_Dose, Len_Sq))

# ==========================================================================
# 3. METHOD B: CONDITIONAL LOGIC (ifelse)
# ==========================================================================
# Scenario: Categorize teeth as "Long" or "Short" based on length
df_logic <- df_clean %>%
  mutate(
    Length_Label = ifelse(len > 20, "Long", "Short"),
    Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose")
```

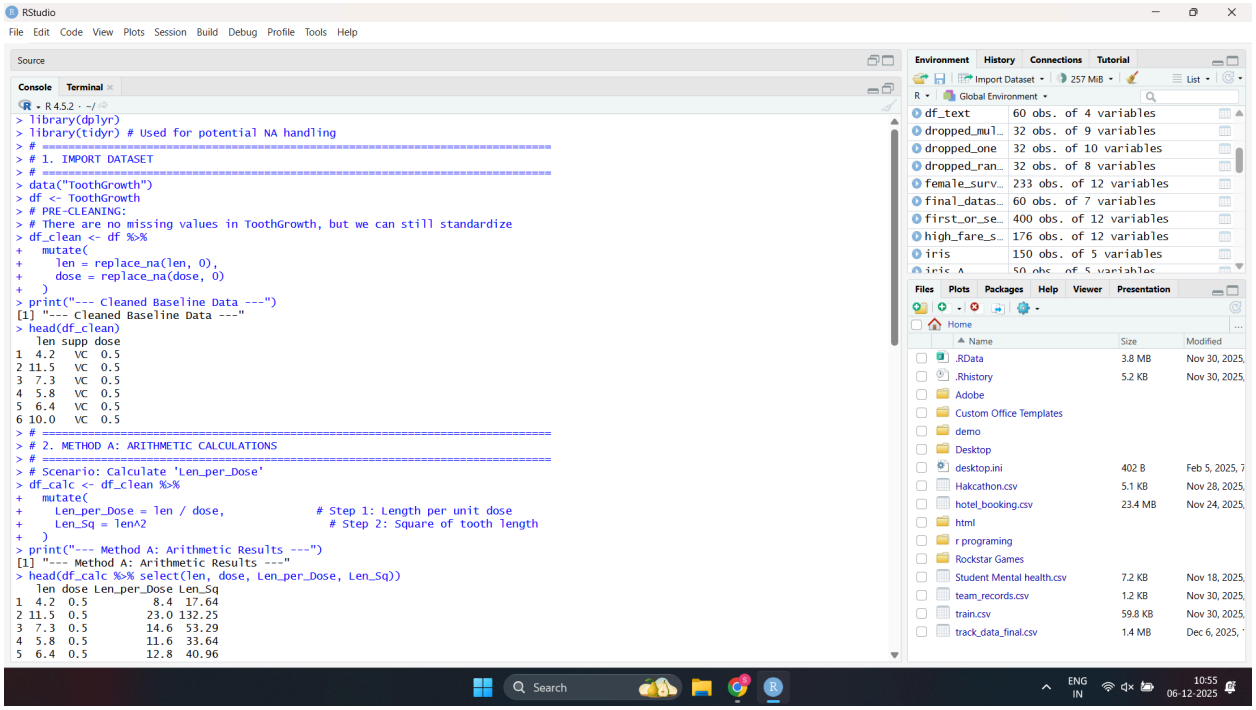


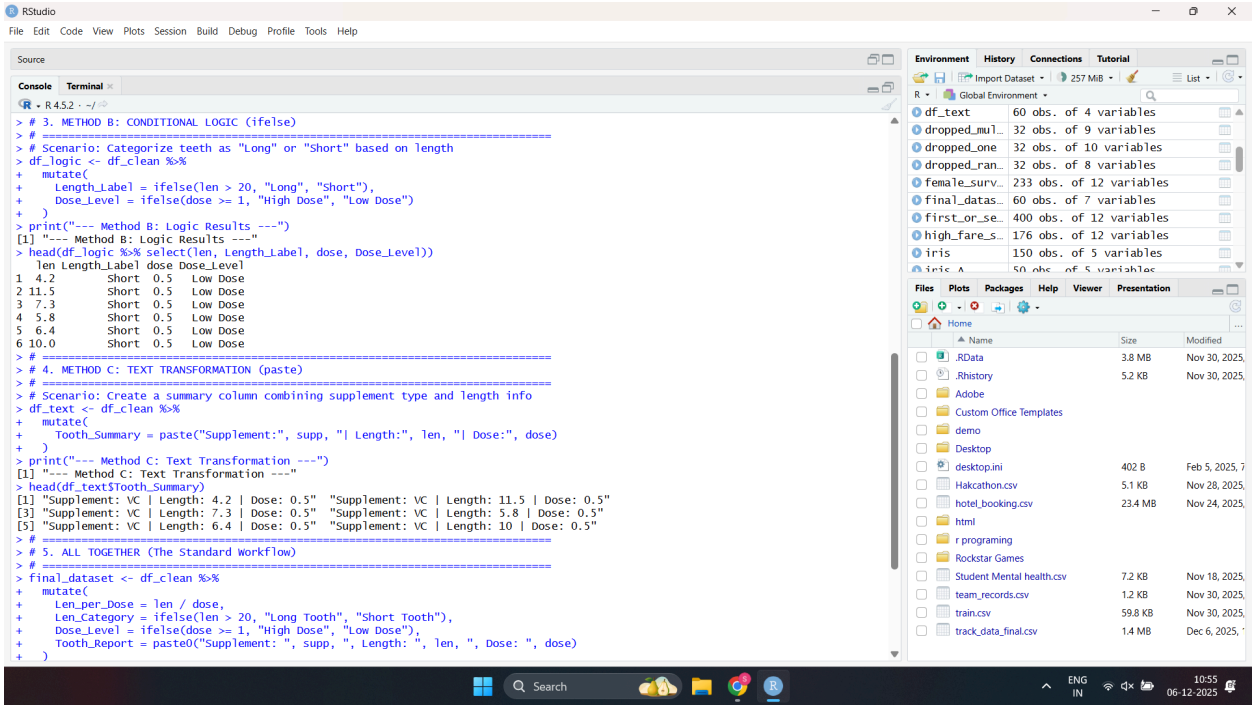
```
# ==========================================================================
# 3. METHOD B: CONDITIONAL LOGIC (ifelse)
# ==========================================================================
# Scenario: Categorize teeth as "Long" or "Short" based on length
df_logic <- df_clean %>%
  mutate(
    Length_Label = ifelse(len > 20, "Long", "Short"),
    Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose")
  )

print("--- Method B: Logic Results ---")
head(df_logic %>% select(len, Length_Label, dose, Dose_Level))

# ==========================================================================
# 4. METHOD C: TEXT TRANSFORMATION (paste)
# ==========================================================================
# Scenario: Create a summary column combining supplement type and length info
df_text <- df_clean %>%
  mutate(
    Tooth_Summary = paste("Supplement:", supp, "| Length:", len, "| Dose:", dose)
  )

print("--- Method C: Text Transformation ---")
head(df_text$Tooth_Summary)

# ==========================================================================
# 5. ALL TOGETHER (The Standard Workflow)
# ==========================================================================
final_dataset <- df_clean %>%
  mutate(
    Len_per_Dose = len / dose,
    Len_Category = ifelse(len > 20, "Long Tooth", "Short Tooth"),
    Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose"),
    Tooth_Report = paste0("Supplement: ", supp, ", Length: ", len, ", Dose: ", dose)
  )

print("--- Final Combined Dataset ---")
head(final_dataset)
```

Riya rane
S107 sycs

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis SAS / SPSS / R

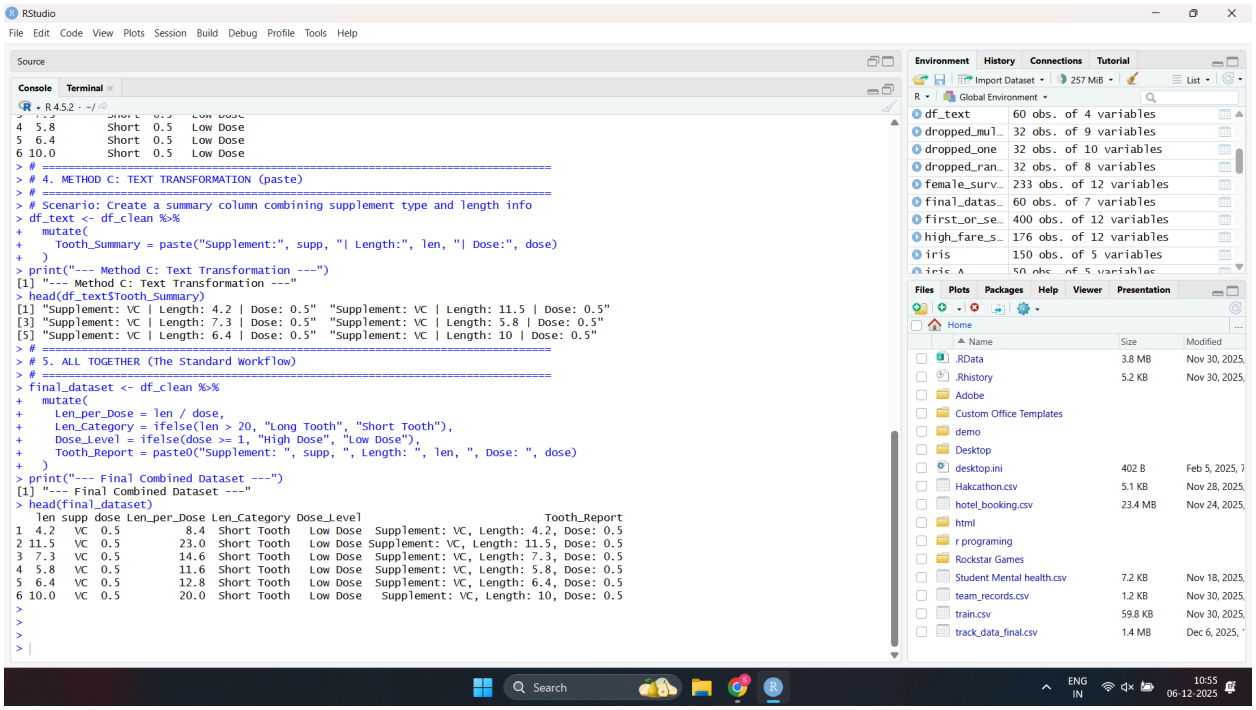
Output :





Riya rane
S107 sycs

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis SAS / SPSS / R



Riya rane
S107 sycs