

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 :

```
1 library(dplyr)
2 library(tidyrr) # Used for potential NA handling
3
4 # =====
5 # 1. IMPORT DATASET
6 # =====
7 data("ToothGrowth")
8 df <- ToothGrowth
9
10 # PRE-CLEANING:
11 # There are no missing values in ToothGrowth, but we can still standardize
12 df_clean <- df %>%
13   mutate(
14     len = replace_na(len, 0),
15     dose = replace_na(dose, 0)
16   )
17
18 print("--- Cleaned Baseline Data ---")
19 head(df_clean)
20
21 # =====
22 # 2. METHOD A: ARITHMETIC CALCULATIONS
23 # =====
24 # Scenario: Calculate 'Len_per_Dose'
25 df_calc <- df_clean %>%
26   mutate(
27     Len_per_Dose = len / dose,          # Step 1: Length per unit dose
28     Len_Sq = len^2                     # Step 2: Square of tooth length
29   )
30
31 print("--- Method A: Arithmetic Results ---")
32 head(df_calc %>% select(len, dose, Len_per_Dose, Len_Sq))
33
34 # =====
35 # 3. METHOD B: CONDITIONAL LOGIC (ifelse)
36 # =====
37 # Scenario: Categorize teeth as "Long" or "Short" based on length
38 df_logic <- df_clean %>%
39   mutate(
40     Length_Label = ifelse(len > 20, "Long", "Short"),
41     Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose")
42   )
43
44 print("--- Method B: Logic Results ---")
45 head(df_logic %>% select(len, Length_Label, dose, Dose_Level))
46
47 # =====
48 # 4. METHOD C: TEXT TRANSFORMATION (paste)
49 # =====
50 # Scenario: Create a summary column combining supplement type and length info
51 df_text <- df_clean %>%
52   mutate(
53     Tooth_Summary = paste("Supplement:", supp, "| Length:", len, "| Dose:", dose)
54   )
55
56 print("--- Method C: Text Transformation ---")
57 head(df_text$Tooth_Summary)
58
59 # =====
60 # 5. ALL TOGETHER (The Standard Workflow)
61 # =====
62 final_dataset <- df_clean %>%
63   mutate(
64     Len_per_Dose = len / dose,
65     Len_Category = ifelse(len > 20, "Long Tooth", "Short Tooth"),
66     Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose"),
67     Tooth_Report = paste0("Supplement: ", supp, ", Length: ", len, ", Dose: ", dose)
68   )
69
70 print("--- Final Combined Dataset ---")
71 head(final_dataset)
72
73
```

```
33 # =====
34 # 3. METHOD B: CONDITIONAL LOGIC (ifelse)
35 # =====
36 # Scenario: Categorize teeth as "Long" or "Short" based on length
37 df_logic <- df_clean %>%
38   mutate(
39     Length_Label = ifelse(len > 20, "Long", "Short"),
40     Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose")
41   )
42
43 print("--- Method B: Logic Results ---")
44 head(df_logic %>% select(len, Length_Label, dose, Dose_Level))
45
46 # =====
47 # 4. METHOD C: TEXT TRANSFORMATION (paste)
48 # =====
49 # Scenario: Create a summary column combining supplement type and length info
50 df_text <- df_clean %>%
51   mutate(
52     Tooth_Summary = paste("Supplement:", supp, "| Length:", len, "| Dose:", dose)
53   )
54
55 print("--- Method C: Text Transformation ---")
56 head(df_text$Tooth_Summary)
57
58 # =====
59 # 5. ALL TOGETHER (The Standard Workflow)
60 # =====
61 final_dataset <- df_clean %>%
62   mutate(
63     Len_per_Dose = len / dose,
64     Len_Category = ifelse(len > 20, "Long Tooth", "Short Tooth"),
65     Dose_Level = ifelse(dose >= 1, "High Dose", "Low Dose"),
66     Tooth_Report = paste0("Supplement: ", supp, ", Length: ", len, ", Dose: ", dose)
67   )
68
69 print("--- Final Combined Dataset ---")
70 head(final_dataset)
71
72
73
```

Riya rane  
S107 sycs

# SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE

## SUBJECT - Data Analysis SAS / SPSS / R

Output :

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal
R - R 4.5.2 - - / -
> library(dplyr)
> library(tidyverse) # 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 ---")
[1] "--- Cleaned Baseline Data ---"
> head(df_clean)
  len supp dose
1  4.2   VC  0.5
2 11.5   VC  0.5
3  7.3   VC  0.5
4  5.8   VC  0.5
5  6.4   VC  0.5
6 10.0   VC  0.5
> # 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 ---")
[1] "--- Method A: Arithmetic Results ---"
> head(df_calc %>% select(len, dose, Len_per_Dose, Len_Sq))
  len dose Len_per_Dose Len_Sq
1  4.2  0.5         8.4  17.64
2 11.5  0.5        23.0 132.25
3  7.3  0.5        14.6  53.29
4  5.8  0.5        11.6  33.64
5  6.4  0.5        12.8  40.96

Environment History Connections Tutorial
R - Global Environment
df_text 60 obs. of 4 variables
dropped_mul_ 32 obs. of 9 variables
dropped_one 32 obs. of 10 variables
dropped_ran_ 32 obs. of 8 variables
female_surv_ 233 obs. of 12 variables
final_data_ 60 obs. of 7 variables
first_or_se_ 400 obs. of 12 variables
high_fare_s_ 176 obs. of 12 variables
iris 150 obs. of 5 variables
iris_A 50 obs. of 5 variables
Files Plots Packages Help Viewer Presentation
Home
Name Size Modified
.RData 3.8 MB Nov 30, 2025
.Rhistory 5.2 KB Nov 30, 2025
Adobe
Custom Office Templates
demo
Desktop
desktop.ini 402 B Feb 5, 2025, 7
Hakathon.csv 5.1 KB Nov 28, 2025
hotel_booking.csv 23.4 MB Nov 24, 2025
html
r programming
Rockstar Games
Student Mental health.csv 7.2 KB Nov 18, 2025
team_records.csv 1.2 KB Nov 30, 2025
train.csv 59.8 KB Nov 30, 2025
track_data_final.csv 1.4 MB Dec 6, 2025, 10:55
06-12-2025
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal
R - R 4.5.2 - - / -
> # 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 ---")
[1] "--- Method B: Logic Results ---"
> head(df_logic %>% select(len, Length_Label, dose, Dose_Level))
  len Length_Label dose Dose_Level
1  4.2      Short  0.5   Low Dose
2 11.5      Short  0.5   Low Dose
3  7.3      Short  0.5   Low Dose
4  5.8      Short  0.5   Low Dose
5  6.4      Short  0.5   Low Dose
6 10.0      Short  0.5   Low Dose
> # 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 ---")
[1] "--- Method C: Text Transformation ---"
> head(df_text$Tooth_Summary)
[1] "Supplement: VC | Length: 4.2 | Dose: 0.5" "Supplement: VC | Length: 11.5 | Dose: 0.5"
[3] "Supplement: VC | Length: 7.3 | Dose: 0.5" "Supplement: VC | Length: 5.8 | Dose: 0.5"
[5] "Supplement: VC | Length: 6.4 | Dose: 0.5" "Supplement: VC | Length: 10 | Dose: 0.5"
> # 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 = paste("Supplement: ", supp, ", Length: ", len, ", Dose: ", dose)
+   )
> print("--- Method D: Final Dataset ---")
[1] "--- Method D: Final Dataset ---"
> head(final_dataset)
  len supp dose Len_per_Dose Len_Category Dose_Level Tooth_Report
1  4.2   VC  0.5         8.4   Short Tooth   Low Dose "Supplement: VC, Length: 4.2, Dose: 0.5"
2 11.5   VC  0.5        23.0   Short Tooth   Low Dose "Supplement: VC, Length: 11.5, Dose: 0.5"
3  7.3   VC  0.5        14.6   Short Tooth   Low Dose "Supplement: VC, Length: 7.3, Dose: 0.5"
4  5.8   VC  0.5        11.6   Short Tooth   Low Dose "Supplement: VC, Length: 5.8, Dose: 0.5"
5  6.4   VC  0.5        12.8   Short Tooth   Low Dose "Supplement: VC, Length: 6.4, Dose: 0.5"
6 10.0   VC  0.5        12.8   Short Tooth   Low Dose "Supplement: VC, Length: 10, Dose: 0.5"

Environment History Connections Tutorial
R - Global Environment
df_text 60 obs. of 4 variables
dropped_mul_ 32 obs. of 9 variables
dropped_one 32 obs. of 10 variables
dropped_ran_ 32 obs. of 8 variables
female_surv_ 233 obs. of 12 variables
final_data_ 60 obs. of 7 variables
first_or_se_ 400 obs. of 12 variables
high_fare_s_ 176 obs. of 12 variables
iris 150 obs. of 5 variables
iris_A 50 obs. of 5 variables
Files Plots Packages Help Viewer Presentation
Home
Name Size Modified
.RData 3.8 MB Nov 30, 2025
.Rhistory 5.2 KB Nov 30, 2025
Adobe
Custom Office Templates
demo
Desktop
desktop.ini 402 B Feb 5, 2025, 7
Hakathon.csv 5.1 KB Nov 28, 2025
hotel_booking.csv 23.4 MB Nov 24, 2025
html
r programming
Rockstar Games
Student Mental health.csv 7.2 KB Nov 18, 2025
team_records.csv 1.2 KB Nov 30, 2025
train.csv 59.8 KB Nov 30, 2025
track_data_final.csv 1.4 MB Dec 6, 2025, 10:55
06-12-2025
```

Riya rane  
S107 sycs

# SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE

## SUBJECT - Data Analysis SAS / SPSS / R

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal
R - R4.5.2 - /
1 4.2 VC 0.5 8.4 Short Tooth Low Dose
4 5.8 Short 0.5 Low Dose
5 6.4 Short 0.5 Low Dose
6 10.0 Short 0.5 Low Dose
> # =====
> # 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 ---")
[1] "--- Method C: Text Transformation ---"
> head(df_text$Tooth_Summary)
[1] "Supplement: VC | Length: 4.2 | Dose: 0.5" "Supplement: VC | Length: 11.5 | Dose: 0.5"
[3] "Supplement: VC | Length: 7.3 | Dose: 0.5" "Supplement: VC | Length: 5.8 | Dose: 0.5"
[5] "Supplement: VC | Length: 6.4 | Dose: 0.5" "Supplement: VC | Length: 10 | Dose: 0.5"
> # =====
> # 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 ---")
[1] "--- Final Combined Dataset ---"
> head(final_dataset)
  len supp dose len_per_dose len_category dose_level Tooth_Report
1  4.2  VC  0.5      8.4 Short Tooth Low Dose Supplement: VC, Length: 4.2, Dose: 0.5
2 11.5  VC  0.5     23.0 Short Tooth Low Dose Supplement: VC, Length: 11.5, Dose: 0.5
3  7.3  VC  0.5     14.6 Short Tooth Low Dose Supplement: VC, Length: 7.3, Dose: 0.5
4  5.8  VC  0.5     11.6 Short Tooth Low Dose Supplement: VC, Length: 5.8, Dose: 0.5
5  6.4  VC  0.5     12.8 Short Tooth Low Dose Supplement: VC, Length: 6.4, Dose: 0.5
6 10.0  VC  0.5     20.0 Short Tooth Low Dose Supplement: VC, Length: 10, Dose: 0.5
>
>
> |

Environment History Connections Tutorial
R - Global Environment
df_text 60 obs. of 4 variables
dropped_mu1... 32 obs. of 9 variables
dropped_one 32 obs. of 10 variables
dropped_ran... 32 obs. of 8 variables
female_surv... 233 obs. of 12 variables
final_datas... 60 obs. of 7 variables
first_or_se... 400 obs. of 12 variables
high_fare_s... 176 obs. of 12 variables
iris 150 obs. of 5 variables
iris_A 50 obs. of 5 variables
Files Plots Packages Help Viewer Presentation
Home
Name Size Modified
.RData 3.8 MB Nov 30, 2025
Rhistory 5.2 KB Nov 30, 2025
Adobe
Custom Office Templates
demo
Desktop
desktop.ini 402 B Feb 5, 2025, 7
Hakathon.csv 5.1 KB Nov 28, 2025
hotel_booking.csv 23.4 MB Nov 24, 2025
html
r programming
Rockstar Games
Student Mental health.csv 7.2 KB Nov 18, 2025
team_records.csv 1.2 KB Nov 30, 2025
train.csv 59.8 KB Nov 30, 2025
track_data_final.csv 1.4 MB Dec 6, 2025, 1
```

Riya rane  
S107 sycs