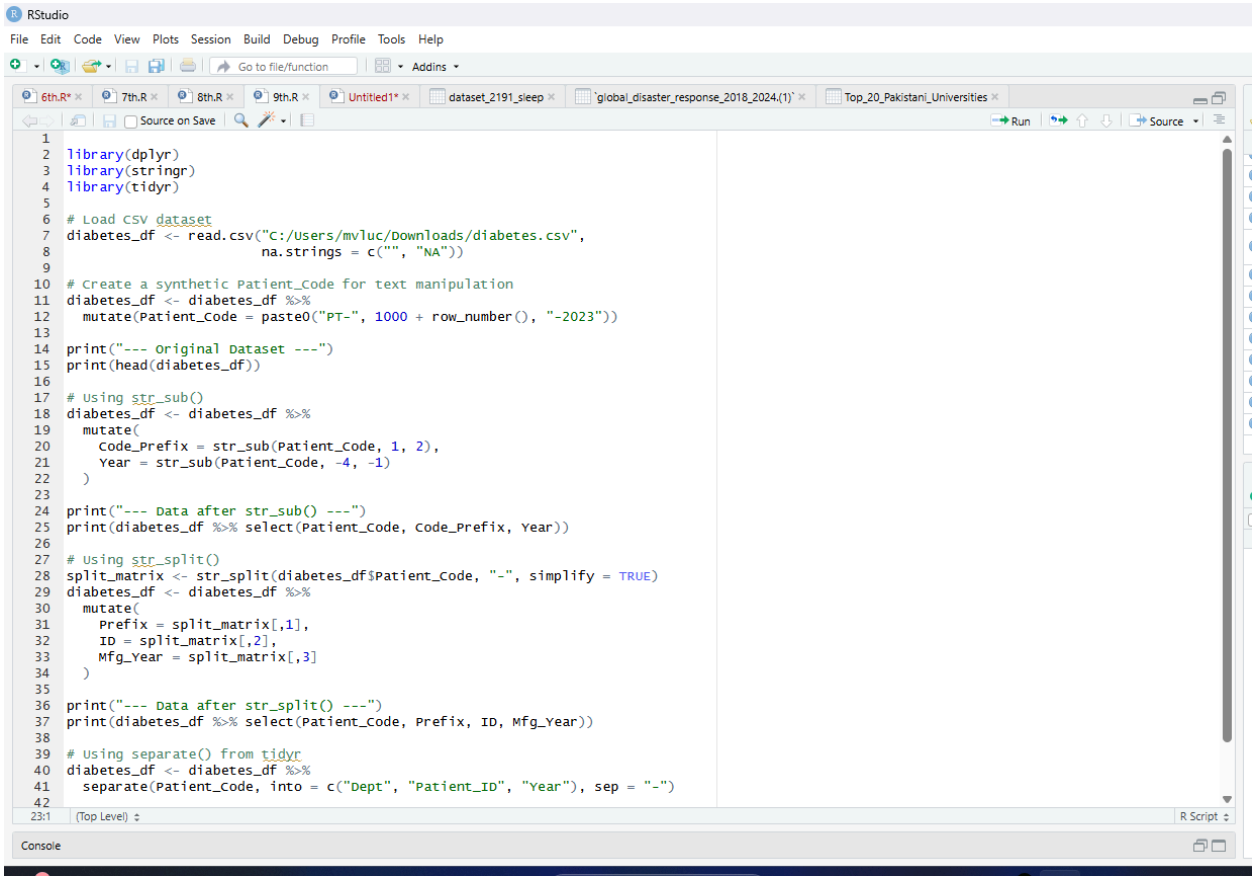
Sheth l.u.j. And sir m.v. college of arts science and commerce

Practical 10 R

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



Name: Simran S113

Sheth l.u.j. And sir m.v. college of arts science and commerce

```
diabetes_df <- read.csv("C:/Users/mvluc/Downloads/diabetes.csv",
```

```
> library(dplyr)
> library(tidyr)
>
> df <- read.csv("C:/Users/mvluc/Downloads/Employee.csv", na.strings = c("", "NA"))
>
> df_clean <- df %>%
+   mutate(
+     PaymentTier = replace_na(PaymentTier, 0),
+     Age = replace_na(Age, 0)
+   )
>
> print("--- Cleaned Baseline Data ---")
[1] "--- Cleaned Baseline Data ---"
> print(head(df_clean))
  Education JoiningYear      City PaymentTier Age Gender EverBenched ExperienceInCurrentDomain LeaveOrNot
1 Bachelors        2017 Bangalore           3  34   Male          No                         0          0
2 Bachelors        2013      Pune           1  28 Female          No                         3          1
3 Bachelors        2014 New Delhi           3  38 Female          No                         2          0
4   Masters        2016 Bangalore           3  27   Male          No                         5          1
5   Masters        2017      Pune           3  24   Male         Yes                         2          1
6 Bachelors        2016 Bangalore           3  22   Male          No                         0          0
>
> df_calc <- df_clean %>%
+   mutate(
+     Bonus = PaymentTier * 0.1,
+     Total_Compensation = PaymentTier + Bonus
+   )
>
> print("--- Method A: Arithmetic Results (Total Compensation) ---")
[1] "--- Method A: Arithmetic Results (Total Compensation) ---"
> print(df_calc %>% select(PaymentTier, Bonus, Total_Compensation))
    PaymentTier Bonus Total_Compensation
1             3   0.3                3.3
2             1   0.1                1.1
```

```
library(tidyr)

# Load CSV dataset
diabetes_df <- read.csv("C:/Users/mvluc/Downloads/diabetes.csv",
```

```
332           2   0.2               2.2
333           3   0.3               3.3
 [ reached 'max' / getOption("max.print") -- omitted 4320 rows ]
>
> df_logic <- df_clean %>%
+   mutate(
+     Age_Group = ifelse(Age < 30, "Young", ifelse(Age <= 50, "Mid-Age", "Senior")),
+     Tier_Label = ifelse(PaymentTier > 5000, "Premium", "Standard")
+   )
>
> print("--- Method B: Logic Results ---")
[1] "--- Method B: Logic Results ---"
> print(df_logic %>% select(Age, Age_Group, PaymentTier, Tier_Label))
    Age Age_Group PaymentTier Tier_Label
1    34   Mid-Age           3   Standard
2    28     Young           1   Standard
3    38   Mid-Age           3   Standard
4    27     Young           3   Standard
5    24     Young           3   Standard
6    22     Young           3   Standard
7    38   Mid-Age           3   Standard
8    34   Mid-Age           3   Standard
9    23     Young           3   Standard
10   37   Mid-Age           2   Standard
11   27     Young           3   Standard
12   34   Mid-Age           3   Standard
13   32   Mid-Age           3   Standard
14   39   Mid-Age           3   Standard
15   37   Mid-Age           3   Standard
16   29     Young           1   Standard
17   34   Mid-Age           3   Standard
18   34   Mid-Age           3   Standard
19   30   Mid-Age           2   Standard
```

Name: Simran S113

Sheth l.u.j. And sir m.v. college of arts science and commerce

```
 [ reached 'max' / getOption("max.print") -- omitted 4403 rows ]
>
> df_text <- df_clean %>%
+   mutate(
+     Employee_Summary = paste(Education, "employee from", City, "with Tier", PaymentTier)
+   )
>
> print("--- Method C: Text Transformation ---")
[1] "--- Method C: Text Transformation ---"
> print(head(df_text$Employee_Summary))
[1] "Bachelors employee from Bangalore with Tier 3" "Bachelors employee from Pune with Tier 1"
[3] "Bachelors employee from New Delhi with Tier 3" "Masters employee from Bangalore with Tier 3"
[5] "Masters employee from Pune with Tier 3"        "Bachelors employee from Bangalore with Tier 3"
>
> final_dataset <- df_clean %>%
+   mutate(
+     Total_Compensation = PaymentTier + (PaymentTier * 0.1),
+     High_Earner = ifelse(Total_Compensation > 7000, TRUE, FALSE),
+     Summary_Report = paste0(Education, ", Age: ", Age, ", Tier: ", PaymentTier)
+   )
>
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset))
  Education JoiningYear      City PaymentTier Age Gender EverBenched ExperienceInCurrentDomain LeaveOrNot Total_Compensation High_Earner
1 Bachelors        2017 Bangalore           3  34   Male          No                          0          0               3.3       FALSE
2 Bachelors        2013      Pune           1  28 Female          No                          3          1               1.1       FALSE
3 Bachelors        2014 New Delhi           3  38 Female          No                          2          0               3.3       FALSE
4   Masters        2016 Bangalore           3  27   Male          No                          5          1               3.3       FALSE
5   Masters        2017      Pune           3  24   Male         Yes                          2          1               3.3       FALSE
6 Bachelors        2016 Bangalore           3  22   Male          No                          0          0               3.3       FALSE
              Summary_Report
1 Bachelors, Age: 34, Tier: 3
```

```
Console  Background Jobs
R 4.5.2 · ~/
>
> print("--- Method C: Text Transformation ---")
[1] "--- Method C: Text Transformation ---"
> print(head(df_text$Employee_Summary))
[1] "Bachelors employee from Bangalore with Tier 3" "Bachelors employee from Pune with Tier 1"
[3] "Bachelors employee from New Delhi with Tier 3" "Masters employee from Bangalore with Tier 3"
[5] "Masters employee from Pune with Tier 3"        "Bachelors employee from Bangalore with Tier 3"
>
> final_dataset <- df_clean %>%
+   mutate(
+     Total_Compensation = PaymentTier + (PaymentTier * 0.1),
+     High_Earner = ifelse(Total_Compensation > 7000, TRUE, FALSE),
+     Summary_Report = paste0(Education, ", Age: ", Age, ", Tier: ", PaymentTier)
+   )
>
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset))
  Education JoiningYear      City PaymentTier Age Gender EverBenched ExperienceInCurrentDomain LeaveOrNot Total_Compensation High_Earner
1 Bachelors        2017 Bangalore           3  34   Male          No                          0          0               3.3       FALSE
2 Bachelors        2013      Pune           1  28 Female          No                          3          1               1.1       FALSE
3 Bachelors        2014 New Delhi           3  38 Female          No                          2          0               3.3       FALSE
4   Masters        2016 Bangalore           3  27   Male          No                          5          1               3.3       FALSE
5   Masters        2017      Pune           3  24   Male         Yes                          2          1               3.3       FALSE
6 Bachelors        2016 Bangalore           3  22   Male          No                          0          0               3.3       FALSE
              Summary_Report
1 Bachelors, Age: 34, Tier: 3
2 Bachelors, Age: 28, Tier: 1
3 Bachelors, Age: 38, Tier: 3
4   Masters, Age: 27, Tier: 3
5   Masters, Age: 24, Tier: 3
6 Bachelors, Age: 22, Tier: 3
>
```

Name: Simran S113