Proposal Submission Tracking

Md Ismail Hossain

2025-02-03

Data Processing and Merging Overview

This script processes and merges **proposal data, award data, and college name data** to create a comprehensive dataset for research funding analysis. The process begins by **loading and cleaning the proposal data**, where cost-related columns are renamed for clarity, and missing statuses are filtered out. Next, the **award data is loaded and processed similarly**, with cost columns renamed and an additional `"Proposal #"` column generated from `"Award #"` to facilitate merging.

After that, the proposal and award datasets are **merged using a left join on "Proposal #"**, ensuring that all proposal data is retained while incorporating matching award information. Following this, **college name data is merged** using `"PI Unit"` to associate each proposal with its respective college or department.

Finally, the `"Created Date"` column is converted into a **Date-Time format**, and the dataset is **filtered to include only records from the year 2024**. The resulting dataset provides a **consolidated view of research proposals, awarded funding, and associated institutions**, making it ready for further analysis or reporting.

Step 1: Load Required Libraries

```r
# Load necessary libraries
library(readr)      # For reading CSV files
library(dplyr)      # For data manipulation
library(readxl)     # For reading Excel files
library(lubridate)  # For handling date-time conversions
```

Step 2: Load and Process Proposal Data

```r
# Read the main Proposal Data file
p_data <- read_csv("C:/Users/mhossa11/OneDrive - University of Wyoming/Projects/Proposal health Project

# Read the Proposal Cost Data file
Proposal_cost_data_all <- read_csv("C:/Users/mhossa11/OneDrive - University of Wyoming/Projects/Proposa
  select(-`Project Title`)  # Remove the "Project Title" column

# Rename columns (except "Proposal #") by appending "in Proposal"
colnames(Proposal_cost_data_all)[-1] <- paste0(colnames(Proposal_cost_data_all)[-1], " in Proposal")

# Merge Proposal Data with Proposal Cost Data on "Proposal #"
p_data <- left_join(p_data, Proposal_cost_data_all, by = "Proposal #")

# Select relevant columns for further processing
p_data_selected_columns <- p_data %>%
  select(`Proposal #`, `Project Title`, `Created Date`, PI, `PI Unit`,
         `History Action`, `History Action Date`, `Status`,
         `Total Sponsor Costs in Proposal`, `Total Direct Costs in Proposal`,
         `Total Indirect Costs in Proposal`, `Total Project Cost in Proposal`) %>%
  filter(!is.na(Status))  # Remove rows with missing Status
```

Step 3: Load and Process Award Data

```r
# Read the main Award Data file
a_data <- read_csv("C:/Users/mhossa11/OneDrive - University of Wyoming/Projects/Proposal health Project

# Read the Award Cost Data file
Award_cost_data_all <- read_csv("C:/Users/mhossa11/OneDrive - University of Wyoming/Projects/Proposal he
  select(-`Project Title`)  # Remove "Project Title" column

# Rename columns (except "Award #") by appending "in Award Data"
colnames(Award_cost_data_all)[-1] <- paste0(colnames(Award_cost_data_all)[-1], " in Award Data")

# Merge Award Data with Award Cost Data on "Award #"
a_data <- left_join(a_data, Award_cost_data_all, by = "Award #")

# Select relevant columns for further processing
a_data_selected_columns <- a_data %>%
  select(`Award #`, `Project Title`, PI, `Admin Unit`, `Status`,
         `Obligated Amount in Award Data`, `Total Expected Amount in Award Data`,
         `Direct Costs in Award Data`, `expected direct in Award Data`,
         `expected indirect in Award Data`, `Indirect Costs in Award Data`,
         `Total Project Cost in Award Data`) %>%
  filter(!is.na(Status))  # Remove rows with missing Status

# Rename "Status" column to "Status_after_award"
colnames(a_data_selected_columns)[5] <- "Status_after_award"

# Generate a "Proposal #" column from "Award #" by replacing "A" with "P"
a_data_selected_columns <- a_data_selected_columns %>%
  mutate(`Proposal #` = gsub("A", "P", `Award #`))

# Perform a left join to merge proposal and award data on "Proposal #"
award_non_award <- p_data_selected_columns %>%
  left_join(a_data_selected_columns, by = "Proposal #")
```

Step 4: Load and Merge College Name Data

```r
# Read the College Name data from an Excel file
college_name <- read_excel("C:/Users/mhossa11/OneDrive - University of Wyoming/Projects/Proposal health

# Select relevant columns and remove missing values
college_name_only <- college_name %>%
  select(`College (Subdivision)`, `Department (Organization)`) %>%
  na.omit()

# Rename "Department (Organization)" column to "PI Unit" for merging
colnames(college_name_only)[2] <- "PI Unit"

# Merge college name data with the award/non-award dataset on "PI Unit"
```

```
award_non_award_with_college_name <- left_join(college_name_only, award_non_award, by = "PI Unit") %>%
  distinct()  # Remove duplicate rows
```

Step 5: Convert "Created Date" to Date-Time and Filter for 2024

```
# Convert "Created Date" to Date-Time format
award_non_award_with_college_name <- award_non_award_with_college_name %>%
  mutate(`Created Date` = as.POSIXct(`Created Date`, format = "%m/%d/%Y %H:%M", tz = "UTC"))

# Define the date range for filtering (1st Jan 2024 - 31st Dec 2024)
start_date <- as.POSIXct("2024-01-01 00:00:00", tz = "UTC")
end_date <- as.POSIXct("2024-12-31 23:59:59", tz = "UTC")

# Filter data within the specified date range
filtered_data <- award_non_award_with_college_name %>%
  filter(`Created Date` >= start_date & `Created Date` <= end_date)

# View the filtered data
write.csv(filtered_data,"C:/Users/mhossa11/OneDrive - University of Wyoming/Projects/Proposal health Pr
```