Problem Set Four

Devin Williams

February 18, 2025

- Issues with running script in R, Java continually is defaulting to a older version. Seems to be running now, and will run outside of OSCER. Tried to implement change mentioned in email but have not been able to fix it fully.
- Data Sources
  - ◇ I am very interested in looking at data scraping for multiple different sources. Including...
    - ◇ Professional soccer player values in leagues around the world.
    - ◇ Transit data including things like rider patterns or bus schedules.
    - ◇ NFL performance statistics year over year.
    - ◇ Professional athletes interview scripts.
    - ◇ Boeing flight data, looking at how true issues with their planes are.
- Question One: Check what type of object mydf is. What type of an object is mydf$date?
  - ◇ It is a character.
- Question Two: Verify that the two dataframe are different types: type class(df1) and class(df). What is the class of each?
  - ◇ There are two different classes: df1 is tbl_df and df2 is tbl_spark
- Question Three: Are the column names any different across the two objects? If so, why might that be?
  - ◇ For the dataframe labeled "df" has the column names: Sepal_Length, Sepal_Width, Petal_Length, Petal_Width, Species. While the dataframe labeled "df1" has the column names: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, Species
  - ◇ Spark typically uses an underscore as a separator rather than a dot like R. This is why df1 has dots and df has underscores in this example.