

# PS4

Becky Crouse

February 21, 2023

## Question 7a

I would be interested in scraping data from SEC filings, specifically forms 10k, to gather data from the reconciliation schedules for unrecognized tax benefits. Compustat data is known for being incomplete for these data points, so I see a lot of value in having a robust, reliable data set.

## Question 7b

- The class for df1 is "tbl\_df" "tbl" "data.frame"
- The class for df is "tbl\_spark" "tbl\_sql" "tbl\_lazy" "tbl"
- Yes, the "." in the original column names is replaced with a "\_" in the Sparklyr table. It's possible that this package utilizes the "." differently, so columns that contain a "." need to be changed in order for the package to function correctly.