Problem Set 4

Leah Pomerantz

February 22, 2022

1 Data Sources

I think I'd be interested in scraping data from Twitter. I really enjoyed the Twitter scraping done in Dr. Wang's class, and it played nicely with R. I would be interested in looking at tweets using the hashtag "antisemitism." I think there are a lot of other topics I would be interested in, as the general concept of using Twitter data and text analysis fascinates me.

2 Questions from Problem 6

(a) 5.d
mydf$dates is a "character" type

(b) 6.7
"class(df1)" gives "tbl_df" "tbl" "data.frame"
"class(df)" gives "tbl_spark" "tbl_sql" "tbl_lazy" "tbl"

(c) 6.8
The names of the columns are different by deliminator. For "df", there is a period between the two words and for "df1", there is a "_" I think it might be because of how tibbles and Sparks may be read differently by the computer.