US Bank Integration 2.0

This is the restarting of the US bank integration project with better quality data compiled from WRDS.

Standard Industrial Classification Code

We include a short description of SIC codes and what they stand for.

```
url_SIC <- "https://en.wikipedia.org/wiki/Standard_Industrial_Classification"
webpage <- xml2::read_html(url_SIC)
SIC_tables <- rvest::html_nodes(webpage, "table")
table_1 <- rvest::html_table(SIC_tables[2], fill = F)
knitr::kable(table_1[[1]])
```

Range of SIC Codes	Division
0100-0999	Agriculture, Forestry and Fishing
1000-1499	Mining
1500-1799	Construction
1800-1999	not used
2000-3999	Manufacturing
4000-4999	Transportation, Communications, Electric, Gas and Sanitary service
5000-5199	Wholesale Trade
5200-5999	Retail Trade
6000-6799	Finance, Insurance and Real Estate
7000-8999	Services
9100-9729	Public Administration
9900-9999	Nonclassifiable

Directory Management

First we need to be able to navigate the directory where data are stored.

```
data_folder_path <- "../Data_Bank_Int/"
file_path <- paste0(data_folder_path, "SICCD_6000_6999_20171031.csv")
```

Tidy Reading

After navigating to the relevant folder, one needs to be able to read the data file, stored in this case, with a `.csv` extension.

```
data <- readr::read_csv(file_path,
                        na = c("", "NA", "NaN", "C", "B"),
                        col_types =
```

```
                              cols(`Cumulative Factor to Adjust Shares/Vol` = col_double(),
                                   `Cumulative Factor to Adjust Prices` = col_double(),
                                   `Factor to Adjust Price` = col_double(),
                                   `Factor to Adjust Shares` = col_double(),
                                   `CUSIP Header` = col_character(),
                                   CUSIP = col_character()
                                   )
                    )

bank_name <- data %>% dplyr::distinct(`Company Name`)
```