Webscraping U.S. Fast Food Chains Data

Shayan Golshan

Contents

Libraries that we will be using

```r
library(rvest)
library(tidyverse)
library(stringr)
library(tibble)
```

1. Function Definition

```r
stateabb <- function(state) {
    if("District of Columbia" %in% state){
      return ("DC")
    }

    if("Virginia" %in% state){
      return ("VA")
    }

    if("West VA" %in% state){
      return("WV")
    }

    loc <- grep(state, state.name)
    stateA <- state.abb[loc]
    return(stateA)
  }


readURL <- function (data){
      URL <-
        data

      Link <-
        read_html(URL)

      Html <-
        html_nodes(Link, css=".list-unstyled-links")

      html2 <-
        html_nodes(Link, css="h1")

      Data <-
        html_text2(Html)

      companyName <-
        html_text2(html2)

      split <-
        str_split(Data, "\n")

      splitDF <-
        as.data.frame(split)
```

```
    states <-
      str_extract(splitDF[,1], ".*(?=  or| McDonald's| Starbucks| Peet's| Dunkin'|
       ↪  Panera| Caribou| Au Bon| The Coffee| Tim)")
      print(states)

    urlDF <-
      splitDF %>%
      mutate(
             State = map_chr(states, ~{
                r <- stateabb(.x)
                if (length(r) == 0) NA_character_ else paste(r, collapse = ", ")
               }),
             Count = parse_number(splitDF[,1]),
             CompanyName = str_extract(companyName, ".*(?= Loc)")
      )

    newDF <-
      urlDF[-1]

    DF <-
      newDF %>%
      drop_na()

    newDF <-
      DF %>% filter(State !=
    "character(0)")

  return(newDF)

  }
```

2. Scraping the Data from the Websites

```
starbs <-
  readURL(paste0("https://web.archive.org/web/20240331211512/",
                 "https://www.menuism.com/restaurant-locations/",
                 "starbucks-coffee-39564"))
```

```
##  [1] NA                      "Alaska"                "Alabama"
##  [4] "Arkansas"              "Arizona"               "California"
##  [7] "Colorado"              "Connecticut"           "District of Columbia"
## [10] "Delaware"              "Florida"               "Georgia"
## [13] "Hawaii"                "Iowa"                  "Idaho"
## [16] "Illinois"              "Indiana"               "Kansas"
## [19] "Kentucky"              "Louisiana"             "Massachusetts"
## [22] "Maryland"              "Maine"                 "Michigan"
## [25] "Minnesota"             "Missouri"              "Mississippi"
## [28] "Montana"               "North Carolina"        "North Dakota"
## [31] "Nebraska"              "New Hampshire"         "New Jersey"
## [34] "New Mexico"            "Nevada"                "New York"
```

```
## [37] "Ohio"                  "Oklahoma"             "Oregon"
## [40] "Pennsylvania"          "Rhode Island"         "South Carolina"
## [43] "South Dakota"          "Tennessee"            "Texas"
## [46] "Utah"                  "Virginia"             "Vermont"
## [49] "Washington"            "Wisconsin"            "West Virginia"
## [52] "Wyoming"               NA                     "County Dublin"
## [55] NA                      "England"              "Greater London"
## [58] "Surrey"                NA                     "Madrid"
## [61] NA                      "British Columbia"     "Ontario"
```

```
dd <-
  readURL(paste0("https://web.archive.org/web/20211025021928/",
                 "http://www.menuism.com/restaurant-locations/",
                 "dunkin-donuts-181624"))
```

```
##  [1] NA                      "Alabama"              "Arkansas"
##  [4] "Arizona"               "California"           "Colorado"
##  [7] "Connecticut"           "District of Columbia" "Delaware"
## [10] "Florida"               "Georgia"              "Hawaii"
## [13] "Iowa"                  "Illinois"             "Indiana"
## [16] "Kansas"                "Kentucky"             "Louisiana"
## [19] "Massachusetts"         "Maryland"             "Maine"
## [22] "Michigan"              "Minnesota"            "Missouri"
## [25] "Mississippi"           "North Carolina"       "Nebraska"
## [28] "New Hampshire"         "New Jersey"           "New Mexico"
## [31] "Nevada"                "New York"             "Ohio"
## [34] "Oklahoma"              "Oregon"               "Pennsylvania"
## [37] "Rhode Island"          "South Carolina"       "Tennessee"
## [40] "Texas"                 "Utah"                 "Virginia"
## [43] "Vermont"               "Washington"           "Wisconsin"
## [46] "West Virginia"         NA                     "Prince Edward Island"
## [49] "Quebec"
```

```
peets <-
  readURL(paste0("https://web.archive.org/web/20240618034417/",
                 "https://www.menuism.com/restaurant-locations/",
                 "peets-coffee-tea-84051"))
```

```
##  [1] NA              "California"   "Colorado"     "Hawaii"       "Illinois"
##  [6] "Nevada"        "Oregon"       "Pennsylvania" "Texas"        "Washington"
## [11] NA              "MA"
```

```
panera <-
  readURL(paste0("https://web.archive.org/web/20240617000440/",
                 "http://www.menuism.com/restaurant-locations/",
                 "panera-bread-4258"))
```

```
##  [1] NA                      "Alabama"              "Arkansas"
##  [4] "Arizona"               "California"           "Colorado"
##  [7] "Connecticut"           "District of Columbia" "Delaware"
## [10] "Florida"               "Georgia"              "Iowa"
## [13] "Idaho"                 "Illinois"             "Indiana"
```

```
## [16] "Kansas"                 "Kentucky"             "Louisiana"
## [19] "Massachusetts"          "Maryland"             "Maine"
## [22] "Michigan"               "Minnesota"            "Missouri"
## [25] "Mississippi"            "North Carolina"       "North Dakota"
## [28] "Nebraska"               "New Hampshire"        "New Jersey"
## [31] "New Mexico"             "Nevada"               "New York"
## [34] "Ohio"                   "Oklahoma"             "Oregon"
## [37] "Pennsylvania"           "Rhode Island"         "South Carolina"
## [40] "South Dakota"           "Tennessee"            "Texas"
## [43] "Virginia"               "Vermont"              "Washington"
## [46] "Wisconsin"              "West Virginia"        NA
## [49] "Ontario"
```

```
caribou <-
  readURL(paste0("https://web.archive.org/web/20220814224211/",
                 "http://www.menuism.com/restaurant-locations/",
                 "caribou-coffee-164861"))
```

```
##  [1] "Colorado"               "District of Columbia" "Georgia"
##  [4] "Iowa"                   "Illinois"             "Indiana"
##  [7] "Kansas"                 "Maryland"             "Michigan"
## [10] "Minnesota"              "Missouri"             "North Carolina"
## [13] "North Dakota"           "Nebraska"             "Ohio"
## [16] "Oregon"                 "Pennsylvania"         "South Dakota"
## [19] "Virginia"               "Wisconsin"
```

```
pain <-
  readURL(paste0("https://web.archive.org/web/20231004193712/",
                 "http://www.menuism.com/restaurant-locations/",
                 "au-bon-pain-69342"))
```

```
##  [1] "Connecticut"            "District of Columbia" "Florida"
##  [4] "Georgia"                "Illinois"             "Indiana"
##  [7] "Kentucky"               "Massachusetts"        "Maryland"
## [10] "Maine"                  "Michigan"             "Minnesota"
## [13] "Missouri"               "New Hampshire"        "New Jersey"
## [16] "Nevada"                 "New York"             "Ohio"
## [19] "Pennsylvania"           "Rhode Island"         "Texas"
## [22] "Virginia"
```

```
beanleaf <-
  readURL(paste0("https://web.archive.org/web/20220628010828/",
                 "http://www.menuism.com/restaurant-locations/",
                 "the-coffee-bean-tea-leaf-165988"))
```

```
## [1] "Arizona"          "California"     "Florida"          "Georgia"
## [5] "Hawaii"           "North Carolina" "Nevada"           "Texas"
```

```
mcD <-
  readURL(paste0("https://web.archive.org/web/20240224131056/",
                 "http://www.menuism.com/restaurant-locations/",
                 "mcdonalds-21019"))
```

```
##   [1] NA                          "Alaska"
##   [3] "Alabama"                   "Arkansas"
##   [5] "Arizona"                   "California"
##   [7] "Colorado"                  "Connecticut"
##   [9] "District of Columbia"      "Delaware"
##  [11] "Florida"                   "Georgia"
##  [13] "Hawaii"                    "Iowa"
##  [15] "Idaho"                     "Illinois"
##  [17] "Indiana"                   "Kansas"
##  [19] "Kentucky"                  "Louisiana"
##  [21] "Massachusetts"             "Maryland"
##  [23] "Maine"                     "Michigan"
##  [25] "Minnesota"                 "Missouri"
##  [27] "Mississippi"               "Montana"
##  [29] "North Carolina"            "North Dakota"
##  [31] "Nebraska"                  "New Hampshire"
##  [33] "New Jersey"                "New Mexico"
##  [35] "Nevada"                    "New York"
##  [37] "Ohio"                      "Oklahoma"
##  [39] "Oregon"                    "Pennsylvania"
##  [41] "Rhode Island"              "South Carolina"
##  [43] "South Dakota"              "Tennessee"
##  [45] "Texas"                     "Utah"
##  [47] "Virginia"                  "Vermont"
##  [49] "Washington"                "Wisconsin"
##  [51] "West Virginia"             "Wyoming"
##  [53] NA                          "Gävleborg"
##  [55] "Östergötland"              NA
##  [57] "Friuli-Venezia Giulia"     NA
##  [59] NA                          "D"
##  [61] NA                          "England"
##  [63] "Greater London"            "Northern Ireland"
##  [65] "Scotland"                  NA
##  [67] "Bretagne"                  "Centre"
##  [69] "Franche-Comte"             "Ile-de-France"
##  [71] "Lorraine"                  "Lower-Normandy"
##  [73] "Midi-Pyrénées"             "Occitanie"
##  [75] "Rhône-Alpes"               "Upper-Normandy"
##  [77] NA                          "CN"
##  [79] "Madrid"                    NA
##  [81] "Hovedstaden"               NA
##  [83] "Bavaria"                   "Berlin"
##  [85] "NDS"                       "NRW"
##  [87] "Saar"                      "Saxony-Anhalt"
##  [89] NA                          "British Columbia"
##  [91] "Manitoba"                  "Newfoundland and Labrador"
##  [93] "Nova Scotia"               "Ontario"
##  [95] "Prince Edward Island"      "Quebec"
##  [97] "Saskatchewan"              NA
##  [99] "New South Wales"           "Victoria"
## [101] NA                          "Steiermark"
## [103] "Tirol"                     NA
## [105] "Cordoba"
```

```r
tim <-
  readURL(paste0("https://web.archive.org/web/20220809051154/",
                 "https://www.menuism.com/restaurant-locations/",
                 "tim-hortons-190025"))
```

```
##  [1] NA                        "Connecticut"             "Delaware"
##  [4] "Indiana"                 "Kentucky"                "Massachusetts"
##  [7] "Maine"                   "Michigan"                "Minnesota"
## [10] "Missouri"                "New Jersey"              "New York"
## [13] "Ohio"                    "Pennsylvania"            "Rhode Island"
## [16] "Virginia"                "West Virginia"           NA
## [19] "British Columbia"        "Manitoba"                "Northwest Territories"
## [22] "Ontario"                 "Prince Edward Island"    "Quebec"
## [25] "Saskatchewan"
```

3. Checking the Contents of the Websites

```r
head(starbs, 10)
```

```
##    State Count      CompanyName
## 1     AK    24 Starbucks Coffee
## 2     AL    73 Starbucks Coffee
## 3     AR    33 Starbucks Coffee
## 4     AZ   279 Starbucks Coffee
## 5     CA  2362 Starbucks Coffee
## 6     CO   371 Starbucks Coffee
## 7     CT   107 Starbucks Coffee
## 8     DC    72 Starbucks Coffee
## 9     DE    20 Starbucks Coffee
## 10    FL   616 Starbucks Coffee
```

```r
head(dd, 10)
```

```
##    State Count    CompanyName
## 1     AL     1 Dunkin' Donuts
## 2     AR    11 Dunkin' Donuts
## 3     AZ    74 Dunkin' Donuts
## 4     CA    46 Dunkin' Donuts
## 5     CO     5 Dunkin' Donuts
## 6     CT   406 Dunkin' Donuts
## 7     DC    15 Dunkin' Donuts
## 8     DE    57 Dunkin' Donuts
## 9     FL   654 Dunkin' Donuts
## 10    GA   147 Dunkin' Donuts
```

```r
head(peets, 10)
```

```
##   State Count         CompanyName
```

```
## 1    CA   163 Peet's Coffee & Tea
## 2    CO     3 Peet's Coffee & Tea
## 3    HI     1 Peet's Coffee & Tea
## 4    IL     3 Peet's Coffee & Tea
## 5    NV     1 Peet's Coffee & Tea
## 6    OR     8 Peet's Coffee & Tea
## 7    PA     1 Peet's Coffee & Tea
## 8    TX     3 Peet's Coffee & Tea
## 9    WA    14 Peet's Coffee & Tea
```

```
head(panera, 10)
```

```
##    State Count  CompanyName
## 1     AL    20 Panera Bread
## 2     AR    13 Panera Bread
## 3     AZ    29 Panera Bread
## 4     CA   216 Panera Bread
## 5     CO    32 Panera Bread
## 6     CT    41 Panera Bread
## 7     DC     4 Panera Bread
## 8     DE     9 Panera Bread
## 9     FL   227 Panera Bread
## 10    GA    71 Panera Bread
```

```
head(caribou, 10)
```

```
##    State Count    CompanyName
## 1     CO     9 Caribou Coffee
## 2     DC     8 Caribou Coffee
## 3     GA    25 Caribou Coffee
## 4     IA    11 Caribou Coffee
## 5     IL    81 Caribou Coffee
## 6     IN     1 Caribou Coffee
## 7     KS     5 Caribou Coffee
## 8     MD     9 Caribou Coffee
## 9     MI    30 Caribou Coffee
## 10    MN   312 Caribou Coffee
```

```
head(pain, 10)
```

```
##    State Count CompanyName
## 1     CT     8 Au Bon Pain
## 2     DC    21 Au Bon Pain
## 3     FL    23 Au Bon Pain
## 4     GA     3 Au Bon Pain
## 5     IL    32 Au Bon Pain
## 6     IN     8 Au Bon Pain
## 7     KY     1 Au Bon Pain
## 8     MA    67 Au Bon Pain
## 9     MD     9 Au Bon Pain
## 10    ME     1 Au Bon Pain
```

```r
head(beanleaf,10)
```

```
##   State Count                CompanyName
## 1    AZ    19 The Coffee Bean & Tea Leaf
## 2    CA   175 The Coffee Bean & Tea Leaf
## 3    FL     1 The Coffee Bean & Tea Leaf
## 4    GA     2 The Coffee Bean & Tea Leaf
## 5    HI    22 The Coffee Bean & Tea Leaf
## 6    NC     1 The Coffee Bean & Tea Leaf
## 7    NV    23 The Coffee Bean & Tea Leaf
## 8    TX     5 The Coffee Bean & Tea Leaf
```

```r
head(mcD, 10)
```

```
##    State Count CompanyName
## 1     AK    33  McDonald's
## 2     AL   279  McDonald's
## 3     AR   190  McDonald's
## 4     AZ   326  McDonald's
## 5     CA  1623  McDonald's
## 6     CO   237  McDonald's
## 7     CT   173  McDonald's
## 8     DC    37  McDonald's
## 9     DE    46  McDonald's
## 10    FL  1142  McDonald's
```

```r
# Some non-United States locations were being aggregated since McDonald's is very
↪  popular.  We will focus only on the United States Fast Food Locations
mcD <- mcD[-c(52, 53), ]

head(tim, 10)
```

```
##    State Count CompanyName
## 1     CT    10 Tim Hortons
## 2     DE     1 Tim Hortons
## 3     IN     5 Tim Hortons
## 4     KY     3 Tim Hortons
## 5     MA     5 Tim Hortons
## 6     ME    27 Tim Hortons
## 7     MI   191 Tim Hortons
## 8     MN     7 Tim Hortons
## 9     MO     1 Tim Hortons
## 10    NJ     2 Tim Hortons
```

4. Getting Population Data for the United States

```r
populationURL <-
    "https://simple.wikipedia.org/wiki/List_of_U.S._states_by_population"
```

```r
populationLink <-
    read_html(populationURL)

populationHtml <-
    html_nodes(populationLink, css="table")

populationData <-
    html_table(populationHtml[[1]])

populationDataDF <-
  populationData[-c(30, 53:56, 57:60),]

populationDataDF$State <-
  str_replace_all(
    populationDataDF$State,
    setNames(
      vapply(populationDataDF$State, function(s) {
        r <- stateabb(s)
        if (length(r) == 0) s else r[1]
      }, FUN.VALUE = character(1)),
      populationDataDF$State
    )
  )

populationDataDF
```

```
## # A tibble: 51 x 11
##    Rank in states & territ~1 Rank in states & ter~2 State Census population, A~3
##    <chr>                     <chr>                  <chr> <chr>
##  1 1                         1                      CA    39,538,223
##  2 2                         2                      TX    30,145,505
##  3 3                         4                      FL    21,538,187
##  4 4                         3                      NY    20,201,249
##  5 5                         6                      PA    13,002,700
##  6 6                         5                      IL    12,812,508
##  7 7                         7                      OH    11,799,448
##  8 8                         9                      GA    10,711,908
##  9 9                         10                     NC    10,439,388
## 10 10                        8                      MI    10,077,331
## # i 41 more rows
## # i abbreviated names: 1: `Rank in states & territories, 2020`,
## #   2: `Rank in states & territories, 2010`,
## #   3: `Census population, April 1, 2020[1][2]`
## # i 7 more variables: `Census population, April 1, 2010[1][2]` <chr>,
## #   `Percent change, 2010-2020[note 1]` <chr>,
## #   `Absolute change, 2010-2020` <chr>, ...
```

```r
populationDataDF$State
```

```
##  [1] "CA"      "TX"      "FL"      "NY"      "PA"      "IL"      "OH"
##  [8] "GA"      "NC"      "MI"      "NJ"      "VA"      "WA"      "AZ"
## [15] "MA"      "TN"      "IN"      "MD"      "MO"      "WI"      "CO"
```

```
## [22] "MN"       "SC"       "AL"       "LA"       "KY"       "OR"       "OK"
## [29] "CT"       "UT"       "IA"       "NV"       "AR"       "MS"       "KS"
## [36] "NM"       "NE"       "ID"       "West VA"  "HI"       "NH"       "ME"
## [43] "RI"       "MT"       "DE"       "SD"       "ND"       "AK"       "DC"
## [50] "VT"       "WY"
```

Cleaning up the States inside the scraped data to ensure a clean join.

```
cleanPopulation <- function(x) {
  x <- trimws(x)
  out <- rep(NA_character_, length(x))

  lut <- setNames(
    c(state.abb, "DC", "WV"),
    c(state.name, "District of Columbia", "West VA")
  )

  idx <- match(x, names(lut))
  out[!is.na(idx)] <- lut[idx[!is.na(idx)]]

  valid_abbs <- c(state.abb, "DC")
  is_valid <- toupper(x) %in% valid_abbs
  out[is_valid] <- toupper(x[is_valid])

  drop_pr <- toupper(x) %in% c("Puerto Rico", "PR")
  out[drop_pr] <- NA_character_

  out
}
```

```
populationDataDF$State <- cleanPopulation(populationDataDF$State)
populationDataDF <- subset(populationDataDF, !is.na(State))
print(populationDataDF$State)
```

```
##  [1] "CA" "TX" "FL" "NY" "PA" "IL" "OH" "GA" "NC" "MI" "NJ" "VA" "WA" "AZ" "MA"
## [16] "TN" "IN" "MD" "MO" "WI" "CO" "MN" "SC" "AL" "LA" "KY" "OR" "OK" "CT" "UT"
## [31] "IA" "NV" "AR" "MS" "KS" "NM" "NE" "ID" "WV" "HI" "NH" "ME" "RI" "MT" "DE"
## [46] "SD" "ND" "AK" "DC" "VT" "WY"
```

5. Joining all of the scraped dataframes with eachother

```
bigDataset <- rbind(mcD, starbs, beanleaf, pain, panera, peets, tim, caribou, dd)

bigDataset <-
  bigDataset %>%
  mutate(
    State = as.character(State)
  )

joined <-
```

```r
  full_join(bigDataset, populationDataDF, by = c("State" = "State"))

head(joined, 10)
```

```
##    State Count CompanyName Rank in states & territories, 2020
## 1     AK    33  McDonald's                                 49
## 2     AL   279  McDonald's                                 24
## 3     AR   190  McDonald's                                 34
## 4     AZ   326  McDonald's                                 14
## 5     CA  1623  McDonald's                                  1
## 6     CO   237  McDonald's                                 21
## 7     CT   173  McDonald's                                 29
## 8     DC    37  McDonald's                                 50
## 9     DE    46  McDonald's                                 46
## 10    FL  1142  McDonald's                                  3
##    Rank in states & territories, 2010 Census population, April 1, 2020[1][2]
## 1                                  48                                733,391
## 2                                  23                              5,024,279
## 3                                  33                              3,011,524
## 4                                  16                              7,151,502
## 5                                   1                             39,538,223
## 6                                  22                              5,773,714
## 7                                  30                              3,605,944
## 8                                  51                                689,545
## 9                                  46                                989,948
## 10                                  4                             21,538,187
##    Census population, April 1, 2010[1][2] Percent change, 2010-2020[note 1]
## 1                                 710,231                              3.3%
## 2                               4,779,736                              5.1%
## 3                               2,915,918                              3.3%
## 4                               6,392,017                             11.9%
## 5                              37,253,956                              6.1%
## 6                               5,029,196                             14.8%
## 7                               3,574,097                              0.9%
## 8                                 601,723                             14.6%
## 9                                 897,934                             10.2%
## 10                             18,801,310                             14.6%
##    Absolute change, 2010-2020
## 1                     +23,160
## 2                    +244,543
## 3                     +95,606
## 4                    +759,485
## 5                  +2,284,267
## 6                    +744,518
## 7                     +31,847
## 8                     +87,822
## 9                     +92,014
## 10                 +2,736,877
##    Total seats in the U.S. House of Representatives, 2023-2033
## 1                                                            1
## 2                                                            7
## 3                                                            4
## 4                                                            9
## 5                                                           52
```

```
## 6                                                                  8
## 7                                                                  5
## 8                                                     1 (non-voting)
## 9                                                                  1
## 10                                                                28
##    Census population per electoral vote[note 2]
## 1                                       244,464
## 2                                       558,253
## 3                                       501,921
## 4                                       650,137
## 5                                       732,189
## 6                                       577,371
## 7                                       515,135
## 8                                       229,848
## 9                                       329,983
## 10                                      717,940
##    Census population per House seat
## 1                           733,391
## 2                           717,754
## 3                           752,881
## 4                           794,611
## 5                           760,350
## 6                           721,714
## 7                           721,189
## 8                                 -
## 9                           989,948
## 10                          769,221
##    Percent of the total U.S. population, 2020[note 3]
## 1                                              0.22%
## 2                                              1.50%
## 3                                              0.90%
## 4                                              2.13%
## 5                                             11.80%
## 6                                              1.72%
## 7                                              1.08%
## 8                                              0.21%
## 9                                              0.30%
## 10                                             6.43%
```

6. Getting the Financial Data

Now, let's use the yfR library to get the stock prices of the companies that we can.

- Unfortunately, Dunkin' Donuts, Panera Bread, Caribou Coffee, and Au Bon Pain are all owned by private brands now, so the stock data returns as null values.

```
table(joined$CompanyName)
```

```
##
##                Au Bon Pain              Caribou Coffee
##                         22                          20
```

```
##              Dunkin' Donuts                   McDonald's
##                          45                           51
##                Panera Bread         Peet's Coffee & Tea
##                          46                            9
##            Starbucks Coffee The Coffee Bean & Tea Leaf
##                          51                            8
##                 Tim Hortons
##                          16
```

```
library(yfR)

tickers <- c("MCD","JDEP.AS","SBUX","JBFCF","QSR")
raw <- yf_get(tickers = tickers, first_date = Sys.Date()-7, last_date = Sys.Date())
latest <- raw %>%
  slice_max(ref_date, n = 1) %>%
  transmute(Ticker = ticker, Price = price_adjusted)
latest
```

```
## # A tibble: 5 x 2
##   Ticker   Price
##   <chr>    <dbl>
## 1 JBFCF     4.10
## 2 JDEP.AS  26.9
## 3 MCD     313.
## 4 QSR      63.8
## 5 SBUX     89.5
```

```
companyName_lookup <- tribble(
  ~CompanyName,                                ~Ticker,
  "Au Bon Pain",                  NA_character_,    # private
  "Caribou Coffee",               NA_character_,    # private
  "Dunkin' Donuts",               NA_character_,    # private
  "McDonald's",                   "MCD",
  "Panera Bread",                 NA_character_,    # private
  "Peet's Coffee & Tea",          "JDEP.AS",        # JDE Peet's (Euronext); US OTC alt:
  ↪ JDEPF
  "Starbucks Coffee",             "SBUX",
  "The Coffee Bean & Tea Leaf",   "JBFCF",          # Jollibee (OTC US); primary: JFC.PS
  "Tim Hortons",                  "QSR"             # Restaurant Brands International
)
```

```
stock_data <- full_join(companyName_lookup, latest, by = c("Ticker" = "Ticker"))
head(stock_data, 10)
```

```
## # A tibble: 9 x 3
##   CompanyName                 Ticker   Price
##   <chr>                       <chr>    <dbl>
## 1 Au Bon Pain                 <NA>      NA
## 2 Caribou Coffee              <NA>      NA
## 3 Dunkin' Donuts              <NA>      NA
## 4 McDonald's                  MCD      313.
## 5 Panera Bread                <NA>      NA
```

```
## 6 Peet's Coffee & Tea        JDEP.AS  26.9
## 7 Starbucks Coffee           SBUX     89.5
## 8 The Coffee Bean & Tea Leaf JBFCF     4.10
## 9 Tim Hortons                QSR      63.8
```

Now, let's join the stock data to the population data.

```
populationStocks_fullDF <- full_join(stock_data, joined, by = c("CompanyName" =
↪  "CompanyName"))
# removing the nulls removes the privately owned corporations
populationStocks_noNullsDF <- full_join(stock_data, joined, by = c("CompanyName" =
↪  "CompanyName")) %>% drop_na()
```

7. Regional and Population Analysis

```
northeast <-
  c("CT", "ME", "MA", "NH", "RI", "VT", "NJ", "NY", "PA")
northeastTF <-
  populationStocks_fullDF$State %in% northeast

midwest <-
  c("IL", "IN", "MI", "OH", "WI", "IA", "KS", "MN", "MO", "NE", "ND", "SD")
midwestTF <-
  populationStocks_fullDF$State %in% midwest

south <-
  c("DE", "FL", "GA", "MD", "NC", "SC", "VA", "DC", "WV", "AL", "KY", "MS", "TN", "AR",
   ↪  "LA", "OK", "TX")
southTF <-
  populationStocks_fullDF$State %in% south

west <-
  c("AZ", "CO", "ID", "MT", "NV", "NM", "UT", "WY", "AK", "CA", "HI", "OR", "WA")
westTF <-
  populationStocks_fullDF$State %in% west


populationStocks_fullDF$Region <-
  as.factor(ifelse(northeastTF == TRUE, "Northeast",
          ifelse(midwestTF == TRUE, "Midwest",
          ifelse(southTF == TRUE, "South",
          ifelse(westTF == TRUE, "West", "NULL")
         )
        )
       )
      )

head(populationStocks_fullDF, 10)
```

```
## # A tibble: 10 x 16
##    CompanyName Ticker Price State Count `Rank in states & territories, 2020`
```

```
##    <chr>       <chr>  <dbl> <chr> <dbl> <chr>
##  1 Au Bon Pain <NA>      NA CT        8 29
##  2 Au Bon Pain <NA>      NA DC       21 50
##  3 Au Bon Pain <NA>      NA FL       23 3
##  4 Au Bon Pain <NA>      NA GA        3 8
##  5 Au Bon Pain <NA>      NA IL       32 6
##  6 Au Bon Pain <NA>      NA IN        8 17
##  7 Au Bon Pain <NA>      NA KY        1 26
##  8 Au Bon Pain <NA>      NA MA       67 15
##  9 Au Bon Pain <NA>      NA MD        9 18
## 10 Au Bon Pain <NA>      NA ME        1 43
## # i 10 more variables: `Rank in states & territories, 2010` <chr>,
## #   `Census population, April 1, 2020[1][2]` <chr>,
## #   `Census population, April 1, 2010[1][2]` <chr>,
## #   `Percent change, 2010-2020[note 1]` <chr>,
## #   `Absolute change, 2010-2020` <chr>,
## #   `Total seats in the U.S. House of Representatives, 2023-2033` <chr>,
## #   `Census population per electoral vote[note 2]` <chr>, ...
```

Q: What are the top chains by region?

```r
#Northeast
populationStocks_fullDF %>%
  filter(Region=="Northeast") %>%
  group_by(Region, CompanyName, State) %>%
  summarize(Count) %>%
  arrange(desc(Count))
```

```
## # A tibble: 53 x 4
## # Groups:   Region, CompanyName [8]
##    Region    CompanyName      State Count
##    <fct>     <chr>            <chr> <dbl>
##  1 Northeast Dunkin' Donuts   MA     1101
##  2 Northeast Dunkin' Donuts   NY     1022
##  3 Northeast McDonald's       NY      811
##  4 Northeast McDonald's       PA      603
##  5 Northeast Starbucks Coffee NY      492
##  6 Northeast Dunkin' Donuts   NJ      477
##  7 Northeast Dunkin' Donuts   CT      406
##  8 Northeast Dunkin' Donuts   PA      402
##  9 Northeast McDonald's       NJ      335
## 10 Northeast McDonald's       MA      306
## # i 43 more rows
```

```r
#Midwest
populationStocks_fullDF %>%
  filter(Region=="Midwest") %>%
  group_by(Region, CompanyName, State) %>%
  summarize(Count) %>%
  arrange(desc(Count))
```

```
## # A tibble: 70 x 4
```

```
## # Groups:   Region, CompanyName [8]
##    Region  CompanyName      State Count
##    <fct>   <chr>            <chr> <dbl>
##  1 Midwest McDonald's       OH      843
##  2 Midwest McDonald's       IL      791
##  3 Midwest McDonald's       MI      662
##  4 Midwest Dunkin' Donuts   IL      579
##  5 Midwest Starbucks Coffee IL      455
##  6 Midwest McDonald's       IN      406
##  7 Midwest McDonald's       MO      396
##  8 Midwest McDonald's       WI      353
##  9 Midwest Caribou Coffee   MN      312
## 10 Midwest McDonald's       MN      305
## # i 60 more rows
```

```
#South
populationStocks_fullDF %>%
  filter(Region=="South") %>%
  group_by(Region, CompanyName, State) %>%
  summarize(Count) %>%
  arrange(desc(Count))
```

```
## # A tibble: 89 x 4
## # Groups:   Region, CompanyName [9]
##    Region CompanyName      State Count
##    <fct>  <chr>            <chr> <dbl>
##  1 South  McDonald's       TX     1303
##  2 South  McDonald's       FL     1142
##  3 South  Starbucks Coffee TX      720
##  4 South  Dunkin' Donuts   FL      654
##  5 South  Starbucks Coffee FL      616
##  6 South  McDonald's       GA      563
##  7 South  McDonald's       NC      475
##  8 South  McDonald's       VA      473
##  9 South  McDonald's       TN      409
## 10 South  McDonald's       MD      402
## # i 79 more rows
```

```
#West
populationStocks_fullDF %>%
  filter(Region=="West") %>%
  group_by(Region, CompanyName, State) %>%
  summarize(Count) %>%
  arrange(desc(Count))
```

```
## # A tibble: 56 x 4
## # Groups:   Region, CompanyName [8]
##    Region CompanyName      State Count
##    <fct>  <chr>            <chr> <dbl>
##  1 West   Starbucks Coffee CA     2362
##  2 West   McDonald's       CA     1623
##  3 West   Starbucks Coffee WA      634
##  4 West   Starbucks Coffee CO      371
```

```
## 5 West   McDonald's      AZ      326
## 6 West   McDonald's      WA      326
## 7 West   Starbucks Coffee AZ     279
## 8 West   Starbucks Coffee OR     279
## 9 West   McDonald's      CO      237
## 10 West  Panera Bread    CA      216
## # i 46 more rows
```

Q: Are some of these chains more prevalent in certain states than others? Possibly despite having less stores overall? Same questions for regions instead of states.

```r
top_states <- populationStocks_fullDF %>%
  group_by(CompanyName, State) %>%
  summarise(stores = sum(Count), .groups = "drop_last") %>%
  mutate(share_of_chain = stores / sum(stores)) %>%    # within-chain share
  group_by(CompanyName) %>%
  slice_max(share_of_chain, n = 5, with_ties = FALSE) %>%
  arrange(CompanyName, desc(share_of_chain))

top_states
```

```
## # A tibble: 45 x 4
## # Groups:   CompanyName [9]
##    CompanyName    State stores share_of_chain
##    <chr>          <chr>  <dbl>          <dbl>
##  1 Au Bon Pain    MA        67         0.202
##  2 Au Bon Pain    NY        58         0.175
##  3 Au Bon Pain    IL        32         0.0964
##  4 Au Bon Pain    PA        28         0.0843
##  5 Au Bon Pain    FL        23         0.0693
##  6 Caribou Coffee MN       312         0.501
##  7 Caribou Coffee IL        81         0.130
##  8 Caribou Coffee OH        46         0.0738
##  9 Caribou Coffee MI        30         0.0482
## 10 Caribou Coffee NC        26         0.0417
## # i 35 more rows
```

```r
most_prevalent_chains_states <- top_states %>%
  group_by(CompanyName) %>% group_split()

most_prevalent_chains_states
```

```
## <list_of<
##   tbl_df<
##     CompanyName   : character
##     State         : character
##     stores        : double
##     share_of_chain: double
##   >
## >[9]>
```

```
## [[1]]
## # A tibble: 5 x 4
##   CompanyName State stores share_of_chain
##   <chr>       <chr>  <dbl>          <dbl>
## 1 Au Bon Pain MA        67         0.202
## 2 Au Bon Pain NY        58         0.175
## 3 Au Bon Pain IL        32         0.0964
## 4 Au Bon Pain PA        28         0.0843
## 5 Au Bon Pain FL        23         0.0693
##
## [[2]]
## # A tibble: 5 x 4
##   CompanyName    State stores share_of_chain
##   <chr>          <chr>  <dbl>          <dbl>
## 1 Caribou Coffee MN       312         0.501
## 2 Caribou Coffee IL        81         0.130
## 3 Caribou Coffee OH        46         0.0738
## 4 Caribou Coffee MI        30         0.0482
## 5 Caribou Coffee NC        26         0.0417
##
## [[3]]
## # A tibble: 5 x 4
##   CompanyName    State stores share_of_chain
##   <chr>          <chr>  <dbl>          <dbl>
## 1 Dunkin' Donuts MA      1101         0.170
## 2 Dunkin' Donuts NY      1022         0.158
## 3 Dunkin' Donuts FL       654         0.101
## 4 Dunkin' Donuts IL       579         0.0893
## 5 Dunkin' Donuts NJ       477         0.0736
##
## [[4]]
## # A tibble: 5 x 4
##   CompanyName State stores share_of_chain
##   <chr>       <chr>  <dbl>          <dbl>
## 1 McDonald's  CA      1623         0.0967
## 2 McDonald's  TX      1303         0.0776
## 3 McDonald's  FL      1142         0.0681
## 4 McDonald's  OH       843         0.0502
## 5 McDonald's  NY       811         0.0483
##
## [[5]]
## # A tibble: 5 x 4
##   CompanyName  State stores share_of_chain
##   <chr>        <chr>  <dbl>          <dbl>
## 1 Panera Bread FL       227         0.0955
## 2 Panera Bread CA       216         0.0908
## 3 Panera Bread OH       170         0.0715
## 4 Panera Bread IL       154         0.0648
## 5 Panera Bread NY       138         0.0580
##
## [[6]]
## # A tibble: 5 x 4
##   CompanyName         State stores share_of_chain
##   <chr>               <chr>  <dbl>          <dbl>
## 1 Peet's Coffee & Tea CA       163         0.827
```

```
## 2 Peet's Coffee & Tea WA        14         0.0711
## 3 Peet's Coffee & Tea OR         8         0.0406
## 4 Peet's Coffee & Tea CO         3         0.0152
## 5 Peet's Coffee & Tea IL         3         0.0152
##
## [[7]]
## # A tibble: 5 x 4
##   CompanyName      State stores share_of_chain
##   <chr>            <chr>  <dbl>          <dbl>
## 1 Starbucks Coffee CA      2362         0.229
## 2 Starbucks Coffee TX       720         0.0699
## 3 Starbucks Coffee WA       634         0.0616
## 4 Starbucks Coffee FL       616         0.0598
## 5 Starbucks Coffee NY       492         0.0478
##
## [[8]]
## # A tibble: 5 x 4
##   CompanyName                State stores share_of_chain
##   <chr>                      <chr>  <dbl>          <dbl>
## 1 The Coffee Bean & Tea Leaf CA       175         0.706
## 2 The Coffee Bean & Tea Leaf NV        23         0.0927
## 3 The Coffee Bean & Tea Leaf HI        22         0.0887
## 4 The Coffee Bean & Tea Leaf AZ        19         0.0766
## 5 The Coffee Bean & Tea Leaf TX         5         0.0202
##
## [[9]]
## # A tibble: 5 x 4
##   CompanyName State stores share_of_chain
##   <chr>       <chr>  <dbl>          <dbl>
## 1 Tim Hortons MI       191          0.382
## 2 Tim Hortons OH       105          0.21
## 3 Tim Hortons NY       100          0.2
## 4 Tim Hortons ME        27          0.054
## 5 Tim Hortons RI        26          0.052
```

Here is each company's top 5 states by share of each company's total U.S. stores

- Au Bon Pain shows strong Northeast concentration with Massachusetts (20.2%) and New York (17.5%)
- Caribou Coffee demonstrates extreme regional focus with Minnesota alone accounting for 50.1% of all locations
- Dunkin' Donuts maintains Northeast dominance across Massachusetts (17.0%) and New York (15.8%)
- McDonald's shows more geographic distribution with California (9.7%) and Texas (7.8%) leading
- Panera Bread has relatively even distribution with Florida (9.6%) and California (9.1%) at the top
- Peet's Coffee & Tea exhibits extreme California concentration at 82.7%
- Starbucks shows California preference (22.9%) but more geographic spread
- The Coffee Bean & Tea Leaf is heavily California-focused (70.6%)
- Tim Hortons concentrates in Michigan (38.2%) and Ohio (21.0%), reflecting its Canadian heritage in border states

```r
top_regions <- populationStocks_fullDF %>%
  group_by(CompanyName, Region) %>%
  summarise(stores = sum(Count), .groups = "drop_last") %>%
  mutate(share_of_chain = stores / sum(stores)) %>%
  group_by(CompanyName) %>%
  slice_max(share_of_chain, n = 3, with_ties = FALSE) %>%
  arrange(CompanyName, desc(share_of_chain))

top_regions
```

```
## # A tibble: 26 x 4
## # Groups:   CompanyName [9]
##    CompanyName    Region    stores share_of_chain
##    <chr>          <fct>      <dbl>          <dbl>
##  1 Au Bon Pain    Northeast    184         0.554
##  2 Au Bon Pain    South         83         0.25
##  3 Au Bon Pain    Midwest       64         0.193
##  4 Caribou Coffee Midwest      520         0.835
##  5 Caribou Coffee South         87         0.140
##  6 Caribou Coffee West          10         0.0161
##  7 Dunkin' Donuts Northeast   3871         0.597
##  8 Dunkin' Donuts South       1512         0.233
##  9 Dunkin' Donuts Midwest      929         0.143
## 10 McDonald's     South       6643         0.396
## # i 16 more rows
```

```r
most_prevalent_chains_regions <- top_regions %>%
  group_by(CompanyName) %>% group_split()

most_prevalent_chains_regions
```

```
## <list_of<
##   tbl_df<
##     CompanyName   : character
##     Region        : factor<06deb>
##     stores        : double
##     share_of_chain: double
##   >
## >[9]>
## [[1]]
## # A tibble: 3 x 4
##   CompanyName Region    stores share_of_chain
##   <chr>       <fct>      <dbl>          <dbl>
## 1 Au Bon Pain Northeast    184          0.554
## 2 Au Bon Pain South         83          0.25
## 3 Au Bon Pain Midwest       64          0.193
##
## [[2]]
## # A tibble: 3 x 4
##   CompanyName    Region  stores share_of_chain
##   <chr>          <fct>    <dbl>          <dbl>
## 1 Caribou Coffee Midwest    520          0.835
## 2 Caribou Coffee South       87          0.140
```

```
## 3 Caribou Coffee West        10         0.0161
##
## [[3]]
## # A tibble: 3 x 4
##   CompanyName    Region    stores share_of_chain
##   <chr>          <fct>      <dbl>          <dbl>
## 1 Dunkin' Donuts Northeast   3871          0.597
## 2 Dunkin' Donuts South       1512          0.233
## 3 Dunkin' Donuts Midwest      929          0.143
##
## [[4]]
## # A tibble: 3 x 4
##   CompanyName Region  stores share_of_chain
##   <chr>       <fct>    <dbl>          <dbl>
## 1 McDonald's  South     6643          0.396
## 2 McDonald's  Midwest   4280          0.255
## 3 McDonald's  West      3411          0.203
##
## [[5]]
## # A tibble: 3 x 4
##   CompanyName  Region    stores share_of_chain
##   <chr>        <fct>      <dbl>          <dbl>
## 1 Panera Bread South        809          0.340
## 2 Panera Bread Midwest      748          0.315
## 3 Panera Bread Northeast    498          0.209
##
## [[6]]
## # A tibble: 3 x 4
##   CompanyName         Region  stores share_of_chain
##   <chr>               <fct>    <dbl>          <dbl>
## 1 Peet's Coffee & Tea West       190         0.964
## 2 Peet's Coffee & Tea Midwest      3         0.0152
## 3 Peet's Coffee & Tea South        3         0.0152
##
## [[7]]
## # A tibble: 3 x 4
##   CompanyName      Region  stores share_of_chain
##   <chr>            <fct>    <dbl>          <dbl>
## 1 Starbucks Coffee West      4414          0.429
## 2 Starbucks Coffee South     2909          0.283
## 3 Starbucks Coffee Midwest   1660          0.161
##
## [[8]]
## # A tibble: 2 x 4
##   CompanyName                Region stores share_of_chain
##   <chr>                      <fct>   <dbl>          <dbl>
## 1 The Coffee Bean & Tea Leaf West      239         0.964
## 2 The Coffee Bean & Tea Leaf South       9         0.0363
##
## [[9]]
## # A tibble: 3 x 4
##   CompanyName Region    stores share_of_chain
##   <chr>       <fct>      <dbl>          <dbl>
## 1 Tim Hortons Midwest      309          0.618
## 2 Tim Hortons Northeast    179          0.358
```

```
## 3 Tim Hortons South          12           0.024
```

Now, when we looking at the top regions, many of these chains appear to be very concentrated in their area

- Au Bon Pain is Northeast-focused (55.4%)
- Caribou Coffee heavily concentrates in the Midwest (83.5%)
- Dunkin' Donuts dominates the Northeast (59.7%)
- McDonald's shows the most even distribution with slight South preference (39.6%)
- Panera Bread favors the South (34.0%) and Midwest (31.5%)
- Peet's Coffee & Tea is almost exclusively Western (96.4%)
- Starbucks leads in the West (42.9%) but maintains significant presence across regions
- The Coffee Bean & Tea Leaf is nearly exclusively Western (96.4%)
- Tim Hortons concentrates in the Midwest (61.8%) and Northeast (35.8%)

```r
state_density <- populationStocks_fullDF %>%
  mutate(
    pop_2020 = as.numeric(gsub("[^0-9.]", "", `Census population, April 1, 2020[1][2]`)),
    Count    = as.numeric(Count)
  ) %>%
  group_by(CompanyName, State) %>%
  summarise(
    stores   = sum(Count, na.rm = TRUE),
    pop_2020 = first(pop_2020),
    .groups  = "drop"
  ) %>%
  mutate(stores_per_100K = stores / pop_2020 * 1e5) %>%
  group_by(CompanyName) %>%
  slice_max(stores_per_100K, n = 5, with_ties = FALSE) %>%
  arrange(CompanyName, desc(stores_per_100K))

state_density
```

```
## # A tibble: 45 x 5
## # Groups:   CompanyName [9]
##    CompanyName    State stores pop_2020 stores_per_100K
##    <chr>          <chr>  <dbl>    <dbl>           <dbl>
##  1 Au Bon Pain    DC        21   689545           3.05
##  2 Au Bon Pain    MA        67  7029917           0.953
##  3 Au Bon Pain    RI         6  1097379           0.547
##  4 Au Bon Pain    NH         6  1377529           0.436
##  5 Au Bon Pain    NY        58 20201249           0.287
##  6 Caribou Coffee MN       312  5706494           5.47
##  7 Caribou Coffee DC         8   689545           1.16
##  8 Caribou Coffee ND         7   779094           0.898
##  9 Caribou Coffee IL        81 12812508           0.632
## 10 Caribou Coffee SD         4   886667           0.451
## # i 35 more rows
```

```r
pop_density_by_company <- state_density %>%
  group_by(CompanyName) %>% group_split()

pop_density_by_company
```

```
## <list_of<
##   tbl_df<
##     CompanyName    : character
##     State          : character
##     stores         : double
##     pop_2020       : double
##     stores_per_100K: double
##   >
## >[9]>
## [[1]]
## # A tibble: 5 x 5
##   CompanyName State stores pop_2020 stores_per_100K
##   <chr>       <chr>  <dbl>    <dbl>           <dbl>
## 1 Au Bon Pain DC        21   689545           3.05
## 2 Au Bon Pain MA        67  7029917           0.953
## 3 Au Bon Pain RI         6  1097379           0.547
## 4 Au Bon Pain NH         6  1377529           0.436
## 5 Au Bon Pain NY        58 20201249           0.287
##
## [[2]]
## # A tibble: 5 x 5
##   CompanyName    State stores pop_2020 stores_per_100K
##   <chr>          <chr>  <dbl>    <dbl>           <dbl>
## 1 Caribou Coffee MN       312  5706494           5.47
## 2 Caribou Coffee DC         8   689545           1.16
## 3 Caribou Coffee ND         7   779094           0.898
## 4 Caribou Coffee IL        81 12812508           0.632
## 5 Caribou Coffee SD         4   886667           0.451
##
## [[3]]
## # A tibble: 5 x 5
##   CompanyName    State stores pop_2020 stores_per_100K
##   <chr>          <chr>  <dbl>    <dbl>           <dbl>
## 1 Dunkin' Donuts MA      1101  7029917           15.7
## 2 Dunkin' Donuts NH       185  1377529           13.4
## 3 Dunkin' Donuts RI       142  1097379           12.9
## 4 Dunkin' Donuts CT       406  3605944           11.3
## 5 Dunkin' Donuts ME       102  1362359            7.49
##
## [[4]]
## # A tibble: 5 x 5
##   CompanyName State stores pop_2020 stores_per_100K
##   <chr>       <chr>  <dbl>    <dbl>           <dbl>
## 1 McDonald's  OH       843 11799448            7.14
## 2 McDonald's  MI       662 10077331            6.57
## 3 McDonald's  MD       402  6177224            6.51
## 4 McDonald's  MO       396  6154913            6.43
## 5 McDonald's  KS       188  2937880            6.40
##
```

```
## [[5]]
## # A tibble: 5 x 5
##   CompanyName  State stores pop_2020 stores_per_100K
##   <chr>        <chr>  <dbl>    <dbl>           <dbl>
## 1 Panera Bread MO        91  6154913            1.48
## 2 Panera Bread OH       170 11799448            1.44
## 3 Panera Bread IL       154 12812508            1.20
## 4 Panera Bread CT        41  3605944            1.14
## 5 Panera Bread VA        98  8631393            1.14
##
## [[6]]
## # A tibble: 5 x 5
##   CompanyName         State stores pop_2020 stores_per_100K
##   <chr>               <chr>  <dbl>    <dbl>           <dbl>
## 1 Peet's Coffee & Tea CA       163 39538223          0.412
## 2 Peet's Coffee & Tea OR         8  4237256          0.189
## 3 Peet's Coffee & Tea WA        14  7705281          0.182
## 4 Peet's Coffee & Tea HI         1  1455271          0.0687
## 5 Peet's Coffee & Tea CO         3  5773714          0.0520
##
## [[7]]
## # A tibble: 5 x 5
##   CompanyName      State stores pop_2020 stores_per_100K
##   <chr>            <chr>  <dbl>    <dbl>           <dbl>
## 1 Starbucks Coffee DC        72   689545           10.4
## 2 Starbucks Coffee WA       634  7705281            8.23
## 3 Starbucks Coffee OR       279  4237256            6.58
## 4 Starbucks Coffee CO       371  5773714            6.43
## 5 Starbucks Coffee NV       188  3104614            6.06
##
## [[8]]
## # A tibble: 5 x 5
##   CompanyName                State stores pop_2020 stores_per_100K
##   <chr>                      <chr>  <dbl>    <dbl>           <dbl>
## 1 The Coffee Bean & Tea Leaf HI        22  1455271          1.51
## 2 The Coffee Bean & Tea Leaf NV        23  3104614          0.741
## 3 The Coffee Bean & Tea Leaf CA       175 39538223          0.443
## 4 The Coffee Bean & Tea Leaf AZ        19  7151502          0.266
## 5 The Coffee Bean & Tea Leaf GA         2 10711908          0.0187
##
## [[9]]
## # A tibble: 5 x 5
##   CompanyName State stores pop_2020 stores_per_100K
##   <chr>       <chr>  <dbl>    <dbl>           <dbl>
## 1 Tim Hortons RI        26  1097379           2.37
## 2 Tim Hortons ME        27  1362359           1.98
## 3 Tim Hortons MI       191 10077331           1.90
## 4 Tim Hortons OH       105 11799448           0.890
## 5 Tim Hortons NY       100 20201249           0.495
```

Here, I calculated stores per 100,000 population to identify where chains achieve highest market exposure relative to local population.

- Au Bon Pain achieves remarkable density in Washington DC (3.05 stores per 100K)
- Caribou Coffee shows exceptional penetration in Minnesota (5.47 per 100K)

- Dunkin' Donuts demonstrates extraordinary Northeast density with Massachusetts leading at 15.7 per 100K
- McDonald's shows highest density in Ohio (7.14 per 100K), Panera Bread peaks in Missouri (1.48 per 100K)
- Peet's Coffee & Tea maintains modest California density (0.41 per 100K)
- Starbucks achieves highest density in Washington DC (10.4 per 100K)
- The Coffee Bean & Tea Leaf shows strongest penetration in Hawaii (1.51 per 100K)
- Tim Hortons achieves highest density in Rhode Island (2.37 per 100K)

Several brands are regional specialists (Peet's and Coffee Bean in the West, Caribou in the Upper Midwest, Au Bon Pain and Dunkin' in the Northeast), while McDonald's and Starbucks are broadly national. When ranking while adjusted for population, small states and D.C. draw different results—e.g., Dunkin' in MA, Caribou in MN, and Starbucks/Au Bon Pain in DC—proving deep local penetration can matter more than raw store counts. Overall, the data highlights home-market advantages and urban concentration (notably DC), confirming that some chains are far more prevalent in specific states and regions even if they don't have the most locations overall.

8. Financial Analysis

Q: Do the financial data match what you'd expect based on the number and locations (footprint) of the stores? Why or why not?

```
financial_check <- populationStocks_fullDF %>%
  mutate(Price = suppressWarnings(as.numeric(Price))) %>%
  group_by(CompanyName) %>%
  summarise(
    total_stores   = sum(as.numeric(Count), na.rm = TRUE),
    stock_price    = median(Price, na.rm = TRUE),    # robust per company
    states_present = n_distinct(State[State %in% state.abb]),
    .groups = "drop"
  ) %>%
  filter(!is.na(stock_price)) %>%                        # public only
  mutate(
    footprint  = total_stores * log1p(states_present), # use logs to respond to skewness
     ↪  of states_present, attempt to "dampen" the spread
    store_rank     = dense_rank(desc(total_stores)),
    footprint_rank = dense_rank(desc(footprint)),
    price_rank     = dense_rank(desc(stock_price))
  )

# How aligned are price and footprint among the public names?
spearman_r <- suppressWarnings(cor(financial_check$stock_price,
                                   financial_check$footprint,
                                   method = "spearman",
                                   use = "complete.obs"))


spearman_r
```

```
## [1] 0.9
```

```
financial_check
```

```
## # A tibble: 5 x 8
##   CompanyName       total_stores stock_price states_present footprint store_rank
##   <chr>                    <dbl>       <dbl>          <int>     <dbl>      <int>
## 1 McDonald's               16781      313.               50    65980.          1
## 2 Peet's Coffee & ~          197       26.9               9      454.          5
## 3 Starbucks Coffee         10294       89.5              50    40474.          2
## 4 The Coffee Bean ~          248        4.10              8      545.          4
## 5 Tim Hortons                500       63.8              16     1417.          3
## # i 2 more variables: footprint_rank <int>, price_rank <int>
```

It looks like the financial data does match what I expect for the most part based on the number and locations (footprint), with some caveats.

- Price broadly matches footprint for the big three — McDonald's (Ranks 1/1/1), Starbucks (2/2/2), Tim Hortons (3/3/3).
- Peet's looks slightly rich (price 4 vs. footprint 5); Coffee Bean slightly cheap (5 vs. 4)
- Some caveats are that per-share price isn't comparable across firms; tickers reflect global, multi-brand parents, while counts are U.S. store brand-level
- I think that store scale helps explain valuation, but unit economics, franchise mix, and growth drive the true differences.

Next steps (better test): Scrape data related to the USD market cap/EV to a global footprint or revenue/margin proxy and incorporate into analysis