

Final Project Cleaning

Fall 2025

importing each set of data

```
circulation <- drop_na(read_csv("../data/library_circulation.csv"))
```

```
## Rows: 9832 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (3): FSCS Key, Library Name, Locale
## dbl (1): Percentage of Children's Material Circulation
## num (5): Service Area Population, Total Circulation, Electronic Material Cir...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
demographics <- drop_na(read_csv("../data/library_demographics.csv"))
```

```
## Rows: 9248 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (3): FSCS Key, Library Name, Locale
## dbl (3): Central Library, Branch Library, Bookmobiles
## num (1): Service Area Population
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
electronic <- read_csv("../data/library_electronic.csv")%>%
  filter(is.na(`Suppression Note`)) %>% # This only exists if data is missing, so we can dr
  select(-(`Suppression Note`)) %>%
  drop_na()
```

```
## Rows: 11090 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (4): FSCS Key, Library Name, Locale, Suppression Note
## num (4): Service Area Population, Internet Computers, Computer Uses, Wireles...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
programs <- drop_na(read_csv("../data/library_programs.csv"))
```

```
## Rows: 9343 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (3): FSCS Key, Library Name, Locale
## num (10): Service Area Population, Total Library Programs, Children's Progra...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
revenue <- read_csv("../data/library_revenue.csv") %>%
  filter(is.na(`Suppression Note`)) %>% # This only exists if data is missing, so we can dr
  select(-(`Suppression Note`)) %>%
  drop_na()
```

```
## Rows: 9248 Columns: 10
## -- Column specification -----
## Delimiter: ","
## chr (4): FSCS Key, Library Name, Locale, Suppression Note
## num (6): Service Area Population, Local Revenue ($), State Revenue ($), Fede...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
services <- drop_na(read_csv("../data/library_services.csv"))
```

```
## Rows: 9626 Columns: 10
## -- Column specification -----
## Delimiter: ","
## chr (3): FSCS Key, Library Name, Locale
## num (7): Service Area Population, Hours/Year, Physical Visits, Reference Tra...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

joining data together (some weird quirks in the data, but this shouldn't change anything just make column names cleaner)

```
libraries <- inner_join(circulation, demographics, by = c("FSCS Key", "Library Name", "Service Area Popu
libraries <- inner_join(libraries, electronic, by = c("FSCS Key", "Library Name", "Service Area Populat
libraries <- inner_join(libraries, programs, by = c("FSCS Key", "Library Name", "Service Area Population
libraries <- inner_join(libraries, revenue, by = c("FSCS Key", "Library Name", "Service Area Population
libraries <- inner_join(libraries, services, by = c("FSCS Key", "Library Name", "Service Area Population
```

removing all negative numbers which mean missing values and saving

```
libraries_filtered <- libraries %>%
  filter(if_all(where(is.numeric), ~ . >= 0))
write_csv(libraries_filtered, "../data/libraries.csv", row.names = FALSE)
```

Show the final file:

```
show(libraries_filtered)
```

```
## # A tibble: 6,832 x 35
##   'FSCS Key' 'Library Name'  Locale Service Area Populat~1 'Total Circulation'
##   <chr>      <chr>          <chr>          <dbl>          <dbl>
## 1 AK0001    ANCHOR POINT PU~ Rural          2105          8684
## 2 AK0007    BIG LAKE PUBLIC~ Rural        10066         28701
## 3 AK0008    CANTWELL COMMUN~ Rural          196           775
## 4 AK0011    CHINIAK PUBLIC ~ Rural           61          2303
## 5 AK0014    COLD BAY PUBLIC~ Rural           50          1722
## 6 AK0015    COOPER LANDING ~ Rural          344           349
## 7 AK0016    CORDOVA PUBLIC ~ Rural        2609         13162
## 8 AK0017    CRAIG PUBLIC LI~ Rural        1036         17254
## 9 AK0019    DELTA COMMUNITY~ Rural          918         40001
## 10 AK0020   DILLINGHAM PUBL~ Rural        2249          6601
## # i 6,822 more rows
## # i abbreviated name: 1: 'Service Area Population'
## # i 30 more variables: 'Percentage of Children's Material Circulation' <dbl>,
## #   'Electronic Material Circulation' <dbl>,
## #   'Print Material Circulation' <dbl>,
## #   'Other Physical Item Circulation' <dbl>, 'Central Library' <dbl>,
## #   'Branch Library' <dbl>, Bookmobiles <dbl>, 'Internet Computers' <dbl>, ...
```