

# Project 1 Fish McDowell

Data is everywhere. It's power is immeasurable with finding patterns, modeling relationships, and driving decisions. In order to be able to do that, data must be handled appropriately. In this report, we will go through the motions of loading in and preprocessing some data so that it's true power can be used as discussed above.

## Question 1: Selecting Columns

First, we will load in the appropriate data set and select only `area_name`, `STCOU`, and any columns that end with the letter D.

```
sec1 <- read_csv("./data/EDU01a.csv", col_names = TRUE)
```

```
Rows: 3198 Columns: 42
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr (22): Area_name, STCOU, EDU010187N1, EDU010187N2, EDU010188N1, EDU010188...
```

```
dbl (20): EDU010187F, EDU010187D, EDU010188F, EDU010188D, EDU010189F, EDU010...
```

```
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.
```

```
sec1_new <- sec1 |>
  select(area_name = Area_name,
         STCOU,
         ends_with("D"))

head(sec1_new, n = 5)
```

```

# A tibble: 5 x 12
  area_name      STCOU EDU010187D EDU010188D EDU010189D EDU010190D EDU010191D
  <chr>          <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1 UNITED STATES 00000    40024299   39967624   40317775   40737600   41385442
2 ALABAMA      01000     733735    728234     730048     728252     725541
3 Autauga, AL   01001      6829      6900       6920       6847       7008
4 Baldwin, AL  01003     16417     16465      16799      17054      17479
5 Barbour, AL  01005      5071      5098       5068       5156       5173
# i 5 more variables: EDU010192D <dbl>, EDU010193D <dbl>, EDU010194D <dbl>,
#   EDU010195D <dbl>, EDU010196D <dbl>

```

The selected columns look to be what we hoped.