

## MAT 128 Lab 2: Columns operations

(Idea and some text from Katherine St. John: <https://stjohn.github.io/teaching/cmp108/s17/lab6.html>)

**Motivating question:** How has the shelter population in New York City changed over time?

**Data:** We will use the Department of Homeless Services (DHS) Daily Report dataset, which stores the number of individuals and families staying in the shelter system each day:

<https://data.cityofnewyork.us/Social-Services/DHS-Daily-Report/k46n-sa2m>

Click on the "View Data" button. To keep the data set from being very large (and avoid some missing values in 2014), we are going filter the data to be all counts after January 1, 2015.

To do this:

- Click on the "Filter" button.
- On the menu that pops up, click on "Add a New Filter Condition".
- Choose "Date of Census" but change the "is" to be "is after".
- Click in the box below and a calendar will pop up. Highlight January 1, 2015.
- Click the check box to the left of the data.
- It will take a few seconds (it's a large file) but the rows on the left will be filtered to be all counts after January 1, 2015.

To download the file,

- Click on the "Export" button.
- Under "Download", choose "CSV".
- The download will begin automatically (files are usually stored in "Downloads" folder).

Move your CSV file to the directory you save your programs. Open with Excel (or your favorite spreadsheet program) to make sure it downloaded correctly. Look at the names of the columns since those will correspond to series we can plot.

### Explore the Data

Using what you learned in Lab 1, plot how the total number of individuals in shelter has changed over time. To do this, write code that:

- opens `DHS_Daily_report.csv` and store a link to it in a variable called `homeless` (a *file handle*)
- create a plot where the x values are the "Date of Census" column and the y values are the "Total Individuals in Shelter" column
- add a title to your plot

*What's happening in your plot? Is the number of people in shelter increasing or decreasing overall?*

- *What happens if you leave off the  $y = \text{"Total Individuals in Shelter"}$ ?*

Next make a plot using the Families with Children column for the y value. You do not have to reopen the data file.

*What's happening in your plot? Is the number of families in shelter increasing or decreasing overall?*

While numbers of both individuals and families in shelter are increasing overall, this is not the whole story. We can do column-wise operations. These columns are called *series*.

For example, in the Shell window, try printing:

```
print(homeless['Total Individuals in Shelter'])
```

What happens when you print:

```
print(homeless['Total Individuals in Shelter']*2)
```

You can also use multiple columns in a calculation:

```
print(homeless['Total Adults in Shelter']/homeless['Total  
Individuals in Shelter'])
```

prints out the fraction of the total number of people in shelter who are adults.

We can save that series by creating a new column for it:

```
homeless['FractionAdults'] = homeless['Total Adults in  
Shelter']/homeless['Total Individuals in Shelter']
```

and then can use it to create a new graph:

```
homeless.plot(x = 'Date of Census', y = 'FractionAdults')
```

This still shows an increasing trend.

For some good news:

- *Display a plot that shows the fraction of the total population that is families, over time. Is it increasing or decreasing?*