Details

1. Reading Data (15 pts)

We will be using the data from the Saturday, November 25, 2017 report. Download that data and load it into pandas. A description of the fields can be found here. Load the data and make sure it has a field in timestamp (datetime64) format that combines the information in the `DATE` and `TIME` columns. Fix the column names so that extra trailing spaces (e.g. on the `EXITS` column) are not included. Finally, set the index to reflect the hierarchy defined above (UNIT, C/A, SCP) plus the `DATE_TIME` column.

Hints:

- There are multiple ways to create the `DATE_TIME` column. Check the `read_csv` documentation or consider the `to_datetime` method.
- You should use a `MultiIndex` to index the data, but `set_index` handles this easily.

2. Differences (15 pts)

The `ENTRIES` and `EXITS` fields hold raw counts that **do not** reset to zero each week. We would like to know how many entries and exits there are in the 4-hour periods. To calculate this, we need to calculate the difference between neighboring rows that have the same (UNIT, C/A, SCP) key. Create `NUM_ENTRIES` and `NUM_EXITS` columns that store these numbers.

Hints:

- The `shift` method will be useful.
- It will be easier to use groupby when doing the shift as it will respect boundaries between subunits. The `level` argument will help define the subunits.
- Most of the counters count up, but there are some that count down. How should you handle those cases? Fix this for extra credit.

3. Resampling (15 pts)

One problem with the numbers from the previous question is that they are sampled at different times. Resample the `ENTRIES` and `EXITS` columns to an hourly rate and interpolate it to fill in the missing values. Use the "pchip" interpolation method as it will preserve monotonicity. Again, this should be done in groups using groupby, but the `apply` function will allow the use of arbitrary interpolate methods. Now, recompute the `NUM_ENTRIES` and `NUM_EXITS` columns from Part 2.

Hints:

- Use `reset_index` to clear the UNIT, C/A, and SCP levels of the index as this makes the resample and interpolate methods used in the `apply` function more straightforward. Add the index back after performing the interpolation via `set_index`.

4. Visualization (15 pts)

Now, create a visualization that plots the resampled entry and exit counts for a particular station. Specifically, let's examine the unit corresponding to Chambers St. (unit `R029`). Sum the `NUM_ENTRIES` and `NUM_EXITS` columns for **all** of the control areas (and all their subunits).

Finally, plot the number of entries and exits over the weeklong period covered by the data by drawing two line plots on the same axes. Your final image should look something like the following: