

Homework 4: Rodent Reports

Introduction

Now that you have harnessed the power of repetition, we can work with real-sized data sets! This is a very exciting step in our collective programming journey as blossoming data scientists.

Content Expectations

Note that though you are welcome to use content "beyond" what we have covered through week 4, you are by no means required to or expected to use this content.

For this homework, we have covered through:

- reading files with loops
- for loops
- lists

You are not required to define any functions other than `main()`.

Style Expectations

You are expected to follow the style guide: https://course.ccs.neu.edu/ds2000/ds2000_fall2022_styleguide.pdf

Files

For you to complete your homework, we'll be providing you with these files:

- `rodents_311_2021_small.csv`
- `rodents_311_2021.csv`

For your homework submission, submit these files:

- `rodent_reports.py` + `boston_rodents.pdf` + `neighborhoods.pdf`
 - `boston_rodents.pdf` is the map generated from the full dataset
 - `neighborhoods.pdf` is the bar chart generated from the full dataset
 - **(update 10/5/2022)** you may turn in `.png` or `.jpg` files instead, but make sure that they are called `boston_rodents.png` or `boston_rodents.jpg` / `neighborhoods.png` or `neighborhoods.jpg`

Make sure to note:

- Each `.py` file should have one (and only one!) `main()`
- We'll continue to use an autograder for this homework that will automatically run your code for you with different combinations of inputs!
 - Watch the demonstration of what you should see when you submit to Gradescope on Piazza.

You are welcome to upload a file multiple times, but only your last submission will be graded. Verify that all files have been submitted at the same time.

See the grading rubric for more information:

https://course.ccs.neu.edu/ds2000/ds2000_fall2022_gradingpolicy.pdf

Question 1 (The only question on this HW) (rodent_reports.py)

For this homework, we're working with the real live [311 data that the city of Boston provides](#) for rodent reports.¹ You'll analyze a number of different aspects of this dataset. This is very much a real-live, large, dataset with local connections so get ready to dig in!

Given the provided `rodents_311_2021_small.csv` and `rodents_311_2021.csv` files, you'll generate a map of rodent reports in Boston and use bar charts to compare the relative number of rodent reports in different neighborhoods in Boston.

In summary, your program will:

- Define a constant at the top of your file that stores the filename of the data file you are reading.
 - To switch between the small and full datasets, you should *only* need to edit this value.
- Read in the provided data.
 - Ignore rows that have no neighborhood.
 - Read the data into three lists, one per column of data (one list holding all the latitude values, one list holding all the longitude values, one holding all the neighborhood names)
 - All three lists should have the same number of values.
- Generate a well-labeled scatter plot of the latitude and longitude data:
 - Add a point to the map representing Northeastern. Place a marker that is easy to identify at the appropriate latitude/longitude.
 - Use the [color, alpha, and marker optional parameters](#) to experiment with different ways to display the data such that it is aesthetically pleasing to you.
 - Make sure that your graph has a legend by using `plt.legend()`.
 - Make sure that when you plot your data you provide the `label` optional parameter to include it in the legend.
- Print the total number of rodent reports in the data.
- Find the different unique neighborhoods represented in this dataset. Print these in sorted order.
- Count the number of rodent reports in each neighborhood and do the following:
 - Generate a well-labeled bar chart comparing all neighborhoods in Boston.
 - Print the average number of rodent reports for Boston neighborhoods.

Restrictions:

- you may not use the pandas library on this homework
- you may not use a set data structure for this homework
- you may not use a dict data structure for this homework
- **(update 10/5/2022)** you may not use the numpy library on this homework

Recommendations:

- The `rodents_311_2021_small.csv` is for development purposes. It has 99 lines of data instead of 3000+. We recommend using this file to develop your program, but your final file must run on the full data set!
- You'll want to store a list of the unique neighborhoods in a variable. This will make your life significantly easier when generating your bar chart.
- Use `plt.savefig(filename, bbox_inches="tight")` to save your plots. You'll end up with much higher resolution and higher-quality plots than if you take a screenshot of spyder!

¹ We want to know what neighborhoods in Boston might be the best ones to have a feline guardian in.

- You may find the following functions helpful (this list is not exhaustive, nor are you required to use these functions):
 - `ls.sort()` - puts a list into sorted order. Does not have a return value.
 - `ls.count(value)` - given a value, returns the number of times that it occurs in this list
 - `ls.index(value)` - given a value, returns the first index that it occurs at in this list
 - `sum(ls)` - given a list, returns the sum
 - `len(ls)` - given a list, returns the number of values in it
 - the `in` keyword - tests for membership in a list (e.g. `3 in [1, 2, 3, 5]` evaluates to `True`)
 - `plt.xticks(rotation = integer from 0 to 180)` - rotates the directionality of x labels in a graph

Example output using the **SMALL DATASET**:

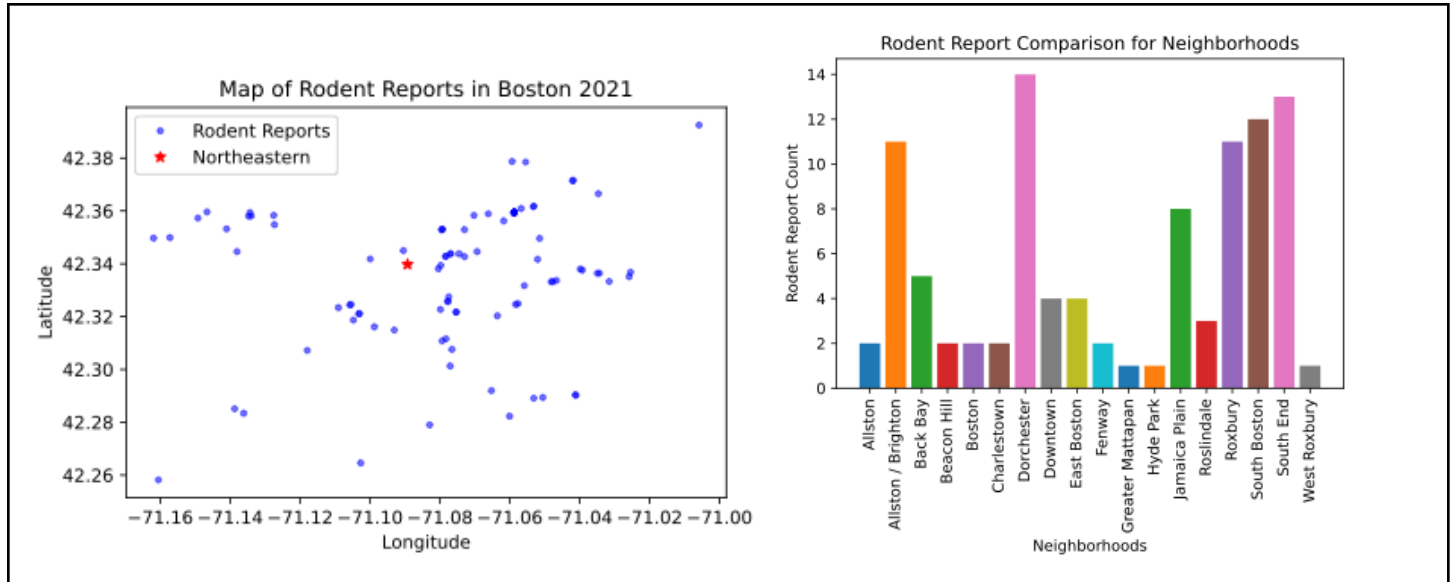
Total rodent reports assigned to a valid neighborhood: 98

Neighborhoods:

- Allston
- Allston / Brighton
- Back Bay
- Beacon Hill
- Boston
- Charlestown
- Dorchester
- Downtown
- East Boston
- Fenway
- Greater Mattapan
- Hyde Park
- Jamaica Plain
- Roslindale
- Roxbury
- South Boston
- South End
- West Roxbury

Average number of rodent reports 5.444444444444445

Example generated plots for the **SMALL DATASET**



Example output using the **FULL DATA SET** (text only, images *not* included in this write-up):

```
Total rodent reports assigned to a valid neighborhood: 3569
```

```
Neighborhoods:
```

```
Allston
```

```
Allston / Brighton
```

```
Back Bay
```

```
Beacon Hill
```

```
Boston
```

```
Brighton
```

```
Charlestown
```

```
Dorchester
```

```
Downtown
```

```
East Boston
```

```
Fenway
```

```
Greater Mattapan
```

```
Hyde Park
```

```
Jamaica Plain
```

```
Mattapan
```

```
Mission Hill
```

```
Roslindale
```

```
Roxbury
```

```
South Boston
```

```
South End
```

```
West Roxbury
```

```
Average number of rodent reports 169.95238095238096
```