

As you work on your projects, many of you are thinking about using data from the Boston Open Data Portal.

Remember that you've developed skills earlier in this class that can apply here.

To work with city files in geo-mapping or analysis software, you may need to clean up the data a little bit. The city tends to distribute geodata as locations like "(42.35158,-71.06024)" as a single column, but you want latitude and longitude to be separate columns.

You can do this in Excel or another program.

An R snippet to clean up your CSV in place.

But here's some code that will automatically do this for you, *if* the csv that you read in has a file called `Location` formatted in parentheses.

(If it's called something else, you'll have to modify the third line of the big block).

Note that the filename here is just the one that I used on *my* computer; you'll have to change it.

```
library(tidyverse)
filename = "~/Downloads/Issued_Moving_Truck_Permits.csv";

filename %>%
  read_csv(file) %>%
  mutate(clean_loc = Location %>%
    gsub("[\\(\\)]", "", .) %>% strsplit(",") %>%
  mutate(
    lat = map_chr(clean_loc, ~ .[1]) %>% as.numeric,
    long = map_chr(clean_loc, ~ .[2]) %>% as.numeric) %>%
  select(-clean_loc) %>%
  write_csv(paste0("cleaned_", file))
```

Point in polygon analysis.

If you have a lot of points (moving permits, say) and want to count how many appear in some polygons, you can do this in QGIS. Some online instructions are [here](#)