

SF Trees

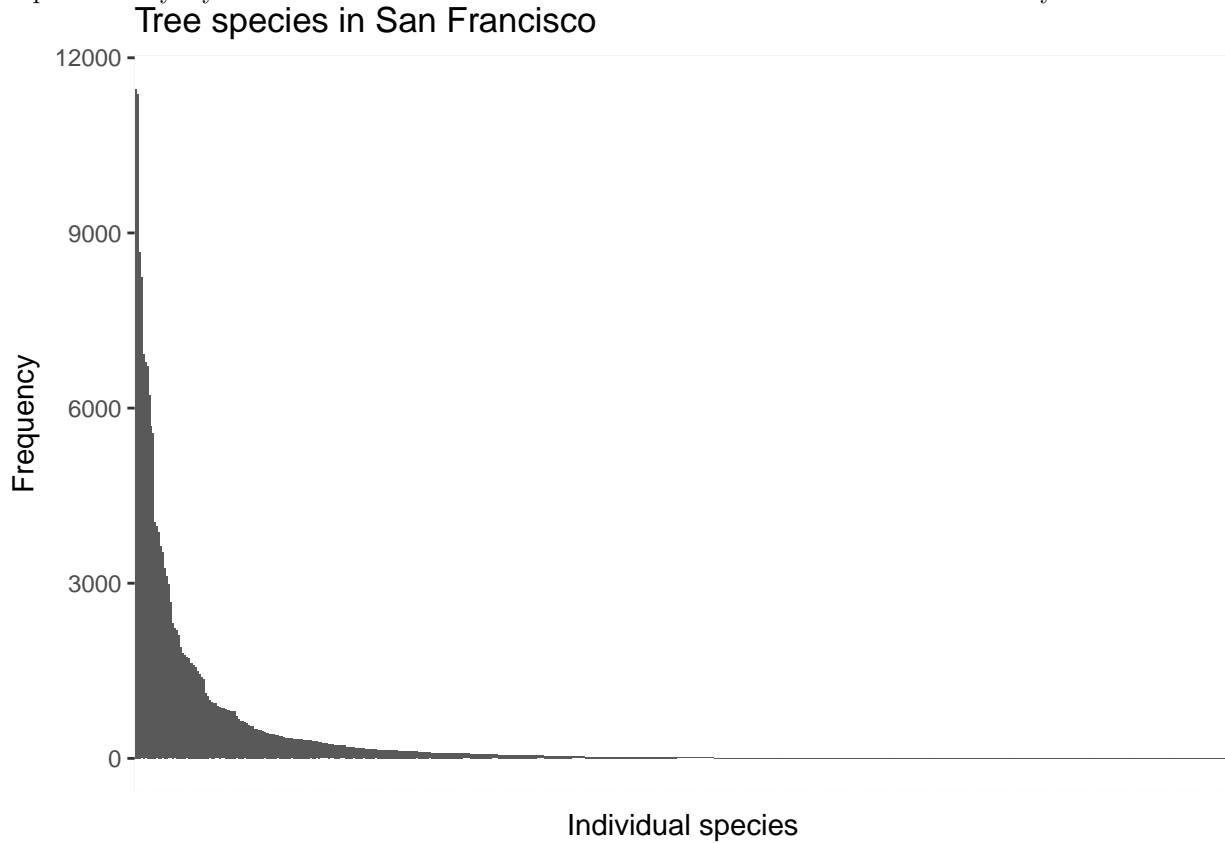
Load data

Missing values

```
## [1] 0.01511976
```

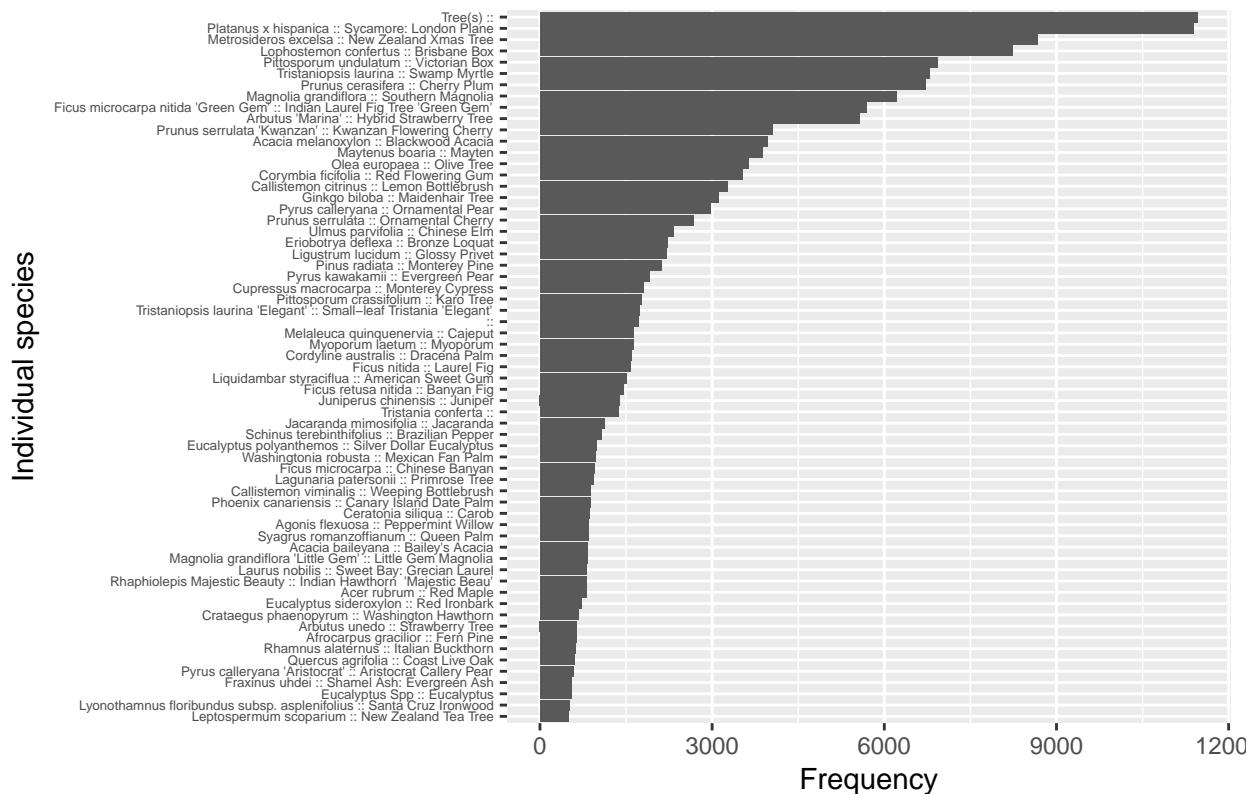
There is 1.5% missing values in this data set; this isn't a lot and I will ignore them.

Next, build a plot of a number of trees by tree species, leave species name blank; we just want to understand where to focus and what to ignore. In this particular case, looks like if we can focus attention at species that are represented by say more than ~500 trees. We will use this as a cut-off for a more detailed analysis and take a look.



Now let's take a look at the species where tree count is over the desired threshold:

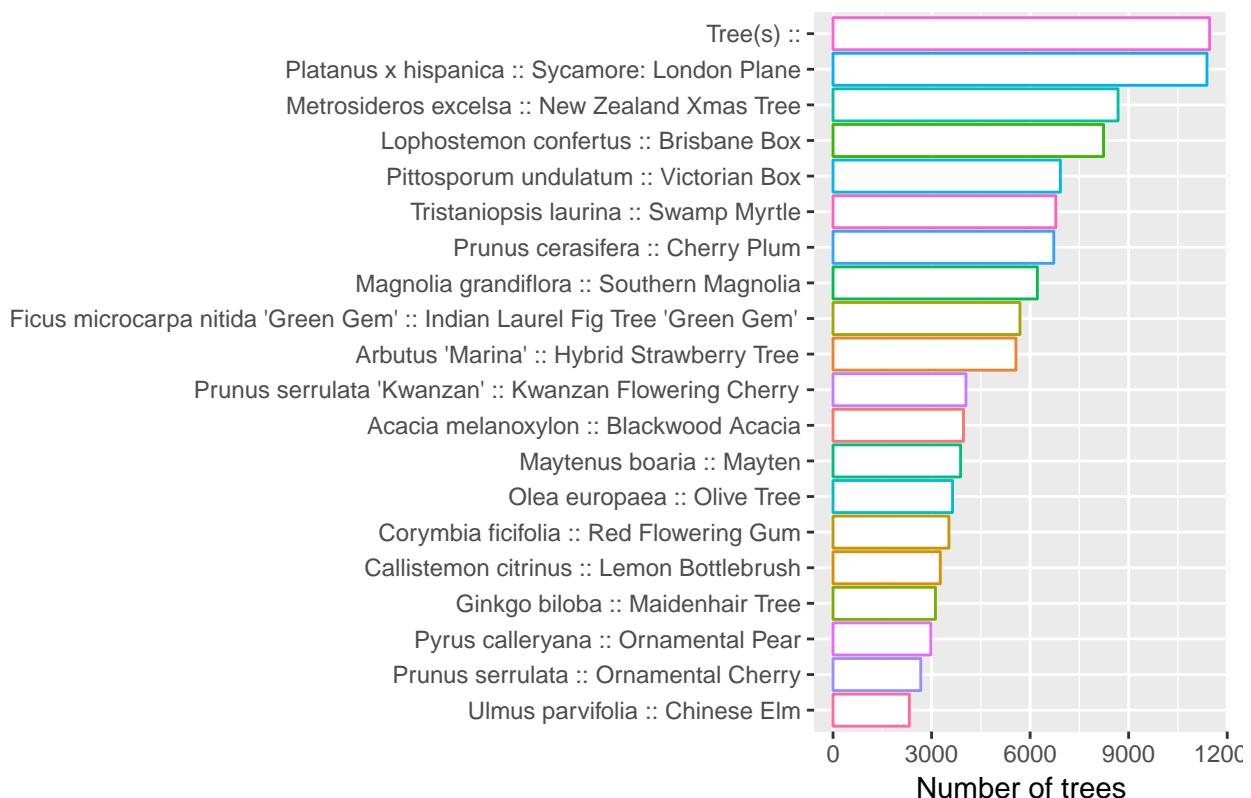
Tree species in San Francisco > threshold



Two things catch my eye: the first row says ‘Tree(s)’ and close to the middle there is a ‘species’ with no name counting about 1,700 trees. These are trees where no species name has been assigned.

Still the purpose here is to play with visualization tools available, so let’s focus on the top 20 trees and build some colored pretty plots:

Tree species in San Francisco

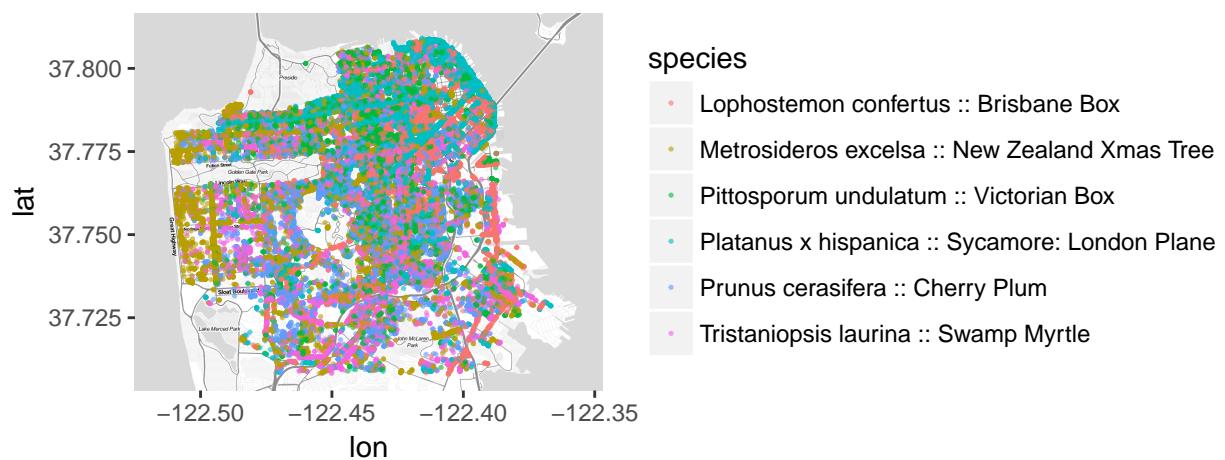


Say we have examined the top species and decided now to take a closer look at the top 6 species, excluding non-named species.

Let's build graphs with specific tree species, each tree species marked by color

In addition, build density trees for:

- all 6 species
- separate graph for each species



Top 6 tree species density in San Francisco



Top 6 tree species density in Brisbane

