Applying Statistical Analysis

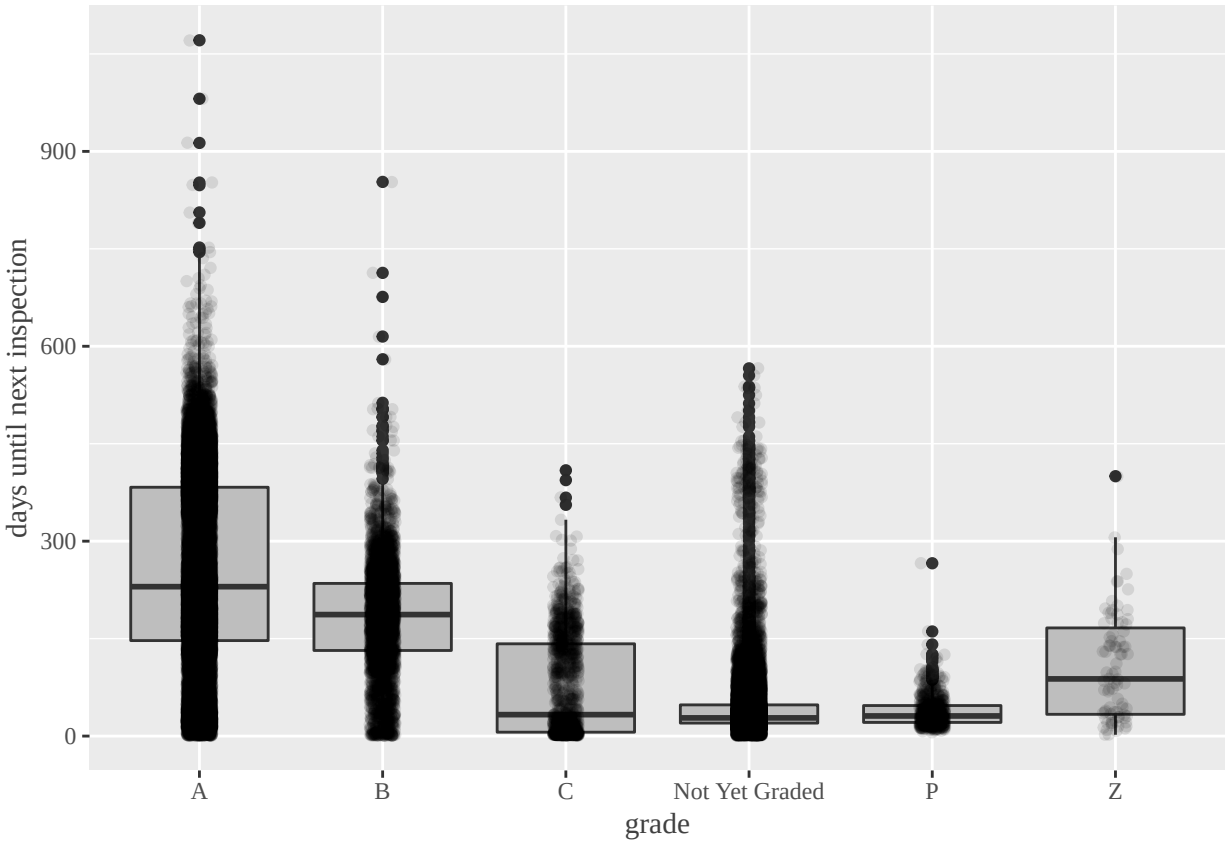
Prediction of restaurant inspections in NYC

```
## -- Attaching packages ------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0

## -- Conflicts ---------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

1 Introduction

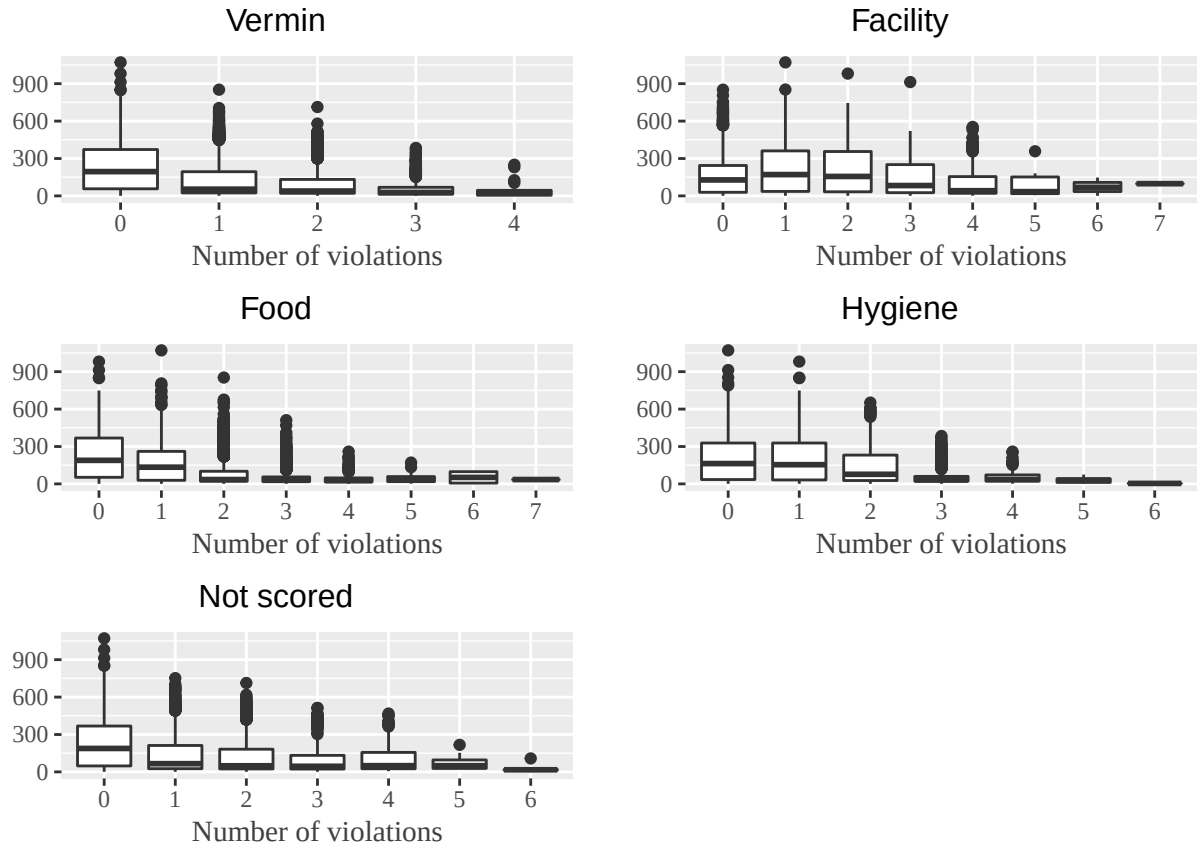
In the following, I will describe two instances where I used statistical analysis in my project. First, there is an exploration between one of the predictors, namely *grade* and the target, *days_until_next_inspection*. Next, I will carry out a similar descriptive analysis for *violation_group* and the target.

2 Days vs. grade



Here, we see boxplots of the target feature for each level of *grade*. I added some random 'horizontal' noise to the observations as a remedy against overplotting.The median number of days decreases as we move from A to C.

3 Days vs. violation groups



The median number of days until the next inspection decreases as the number of violations per inspection increases.