Our goal is to predict which restaurant inspections in the city of Chicago will result in critical violations. To address this problem, we are examining data from three sources.

First, we are looking at the records of the food inspection outcomes since 2010. This database contains all the food inspections done within Chicago over that period, which amounts to over 130,000 inspections. For each inspection, we have the result: a pass, a pass with conditions, or a critical violation. Furthermore, we have access to the establishment name, location, type (restaurant vs café, etc), as well as details on the inspection failure, if any.

We are combining the inspection data with weather records from the same period, obtained from National Oceanic and Atmospheric Administration (NOAA). These data were recorded at Chicago O’Hare international airport; we are using the daily high, low, and average temperature, the amount of precipitation, the amount of snowfall, and the average wind speed. From a previous analysis done by the City of Chicago, we believe the temperature data may predict inspection outcomes to some degree.

Finally, we are also examining 311 sanitation complaint records from the City of Chicago. We downloaded all the 311 sanitation complaints in the city, of which there were 112,000, or about 45/day. These complaints may reflect areas of the city where sanitation standards are reduced and therefore may predict where restaurants will fail their inspections.

We first examined whether the inspection outcomes depended on nearby sanitation complaints. To do this, we took each inspection record and found all the sanitation complaints that occurred within the previous week and within a small local distance (about a 3km radius). We then plotted a histogram of the number of complaints among those inspections with critical violations and compared it to a histogram for the inspections that passed or passed with conditions (Fig 1). We found that those inspections that passed had fewer recent local 311 complaints (histogram has more mass to the left; green line). Conversely, those inspections that failed were likely to have more sanitation complaints within the past week (red line). However, this effect is quite subtle; a classifier based on this information alone would be quite poor.

Next, we examined whether the inspection outcome depended on the weather. We plotted a histogram of the day’s high temperature for each inspection that critically failed (Fig 2A, panel A, red line). We did the same for those inspections that passed with conditions (blue line) and those that passed outright (green line). We found that the distribution of temperatures was skewed toward hot days for inspections that failed. That is, inspections were more likely to fail on hot days than on moderate ones. Interestingly, both very hot days and very cold days did not seem to affect inspection outcome; only above-freezing to 70° days were more likely to pass while 80° to 90° days were more likely to fail. However, as with the number of nearby sanitation complaints, this effect is very subtle.

The daily low was even less predictive than the daily high (Fig 2, panel B). Similarly, the day’s wind speed and precipitation had not predictive power at all.