PS1

Jo Denby

4/16/2018

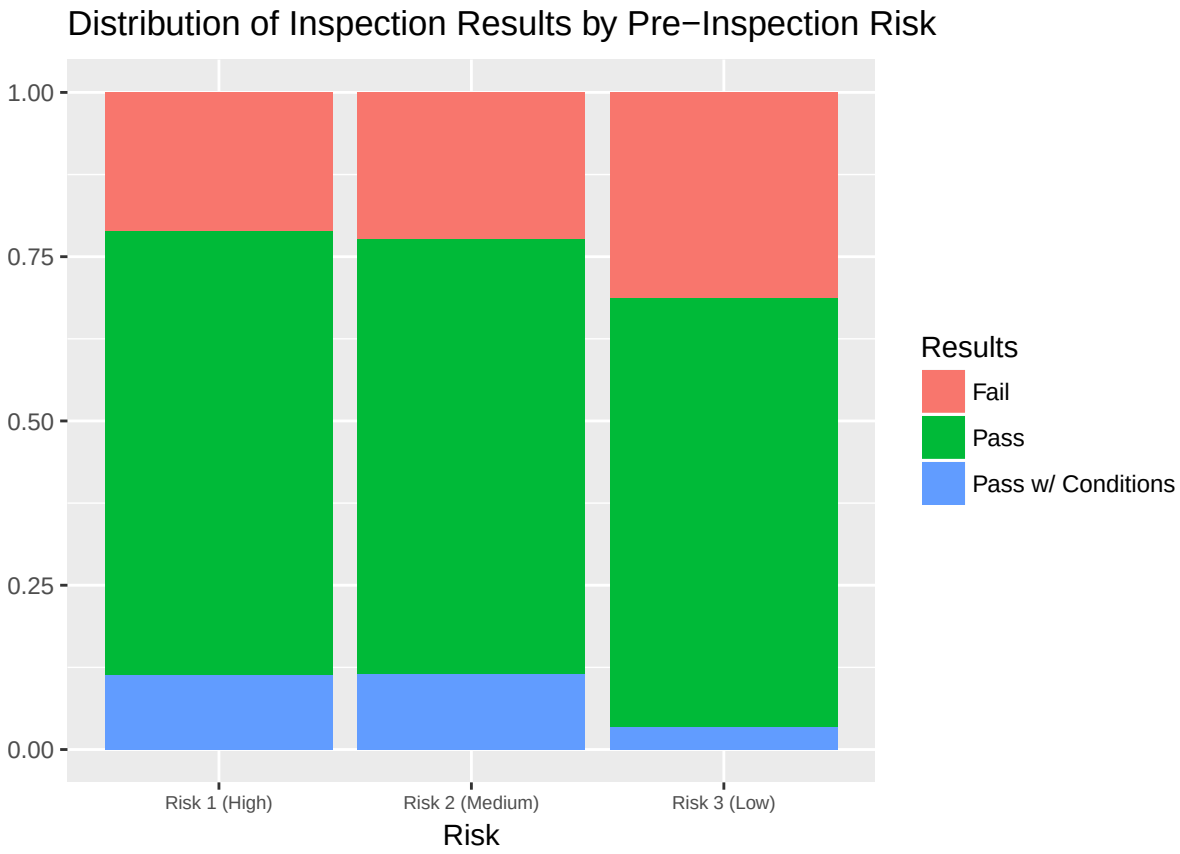
Part 1.

1. The data source (Food Inspection records for establishments within the City of Chicago since January 1, 2010) is hosted by the City of Chicago as part of their Data Portal. It's easily accessible by browsing or searching the portal through its category structure or search bar. While it's not immediately obvious who specifically maintains the dataset, the site provides a link to the administrator's profile (cocadmin) as well as a means to contact the administrator. The dataset itself comes from the City of Chicago's Department of Public Health.
2. There don't seem to be many key papers or publications that cite this dataset. Harris et al., (2014) make some use of the dataset in their work, but it is far from the main focus of the research endeavor. Instead, the City of Chicago itself (along with a team of volunteer researchers) is spearheading a research effort to forecast inspection outcomes by training machine learning models using components of this dataset. Their work can be viewed at https://chicago.github.io/food-inspections-evaluation/.
3. The dataset includes a description section outlining how the data are collected. Once inspections are conducted by officials employed by the Chicago Department of Public Health's Food Protection Program, the results are added to the database. After this initial submission, all inspection outcomes are reviewed and approved by an Environmental Health Practitioner licensed by the State of Illinois.
4.

Summary Table	
Number of States	1
Number of Cities/Neighborhoods	45
Est. Number of Establishments	34145
Number of Inspections	167109
Failure Rate	19%
Unconditional Pass Rate	59%
Percent High Risk Establishments	70%
Percent Routine Inspection (Canvass)	53%

5/6.

```
food %>%
  filter(!is.na(Risk)) %>%
  filter(Risk != 'All') %>%
  filter(Results %in% c("Pass","Fail","Pass w/ Conditions")) %>%
  count(Results, Risk) %>%
  ggplot(data = ., aes(Risk, n, fill=Results)) +
  geom_col(position = 'fill') +
  labs(title = 'Distribution of Inspection Results by Pre-Inspection Risk',
       y = "") +
  theme(axis.text.x = element_text(size = 7 ))
```




```
food %>%
  filter(`Inspection Type` %in% c('Complaint', "Canvass", "License")) %>%
  filter(Results %in% c("Pass", "Fail","Pass w/ Conditions")) %>%
  count(Results, `Inspection Type`) %>%
  ggplot(aes(`Inspection Type`, n, fill = Results)) +
  geom_col(position = 'fill') +
  labs(title = 'Distribution of Inspection Results by Inspection Type',
       y= "")
```

Based on these plots, two interesting insights come to light. From the first, it's notable that establishments deemed 'Low Risk' seem to have the greatest chance of failing their inspection! In the next, among the most common inspection types, establishments are most likely to fail their inspection as part of their initial license application.