

# Predicting Food Inspection Outcomes in Chicago

Luke Farewell, Jake Gober, Sam Green, Jeremy Welborn

Computer Science 109a / Statistics 121a, Harvard University

## Objectives

Chicago's Department of Public Health is responsible for inspecting 15,000 food establishments across the city. Our goal was to reduce the amount of time required to discover critical violations:

- Aggregate and clean useful data sources.
- Build models for probabilities of failure.
- Optimize cost of inspection by recommending inspections in a probability-ranked order.

## Data: Sources and Cleaning

Food Inspections	Business Activities
Weather	Crime Reports
Sanitation Complaints	Business Location

Sources: City of Chicago Open Data Portal, NOAA.

- Collate inspections with businesses for history.
- Process free text columns.
- Bucket spatial data with a grid (see visualizations).

## Exploration (I)

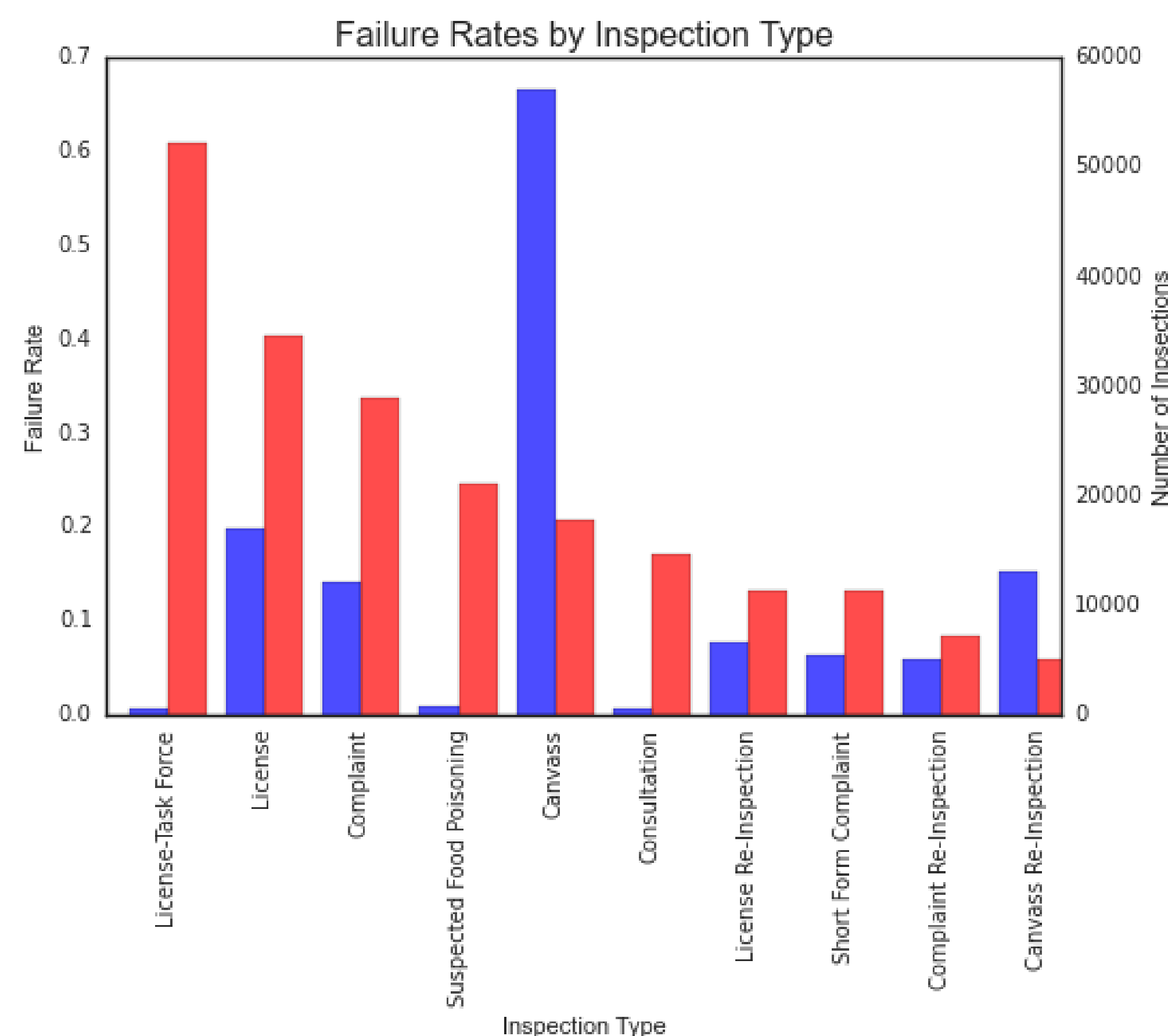


Figure 1: Scoring Function

## Exploration (II)

Model success curve 1.

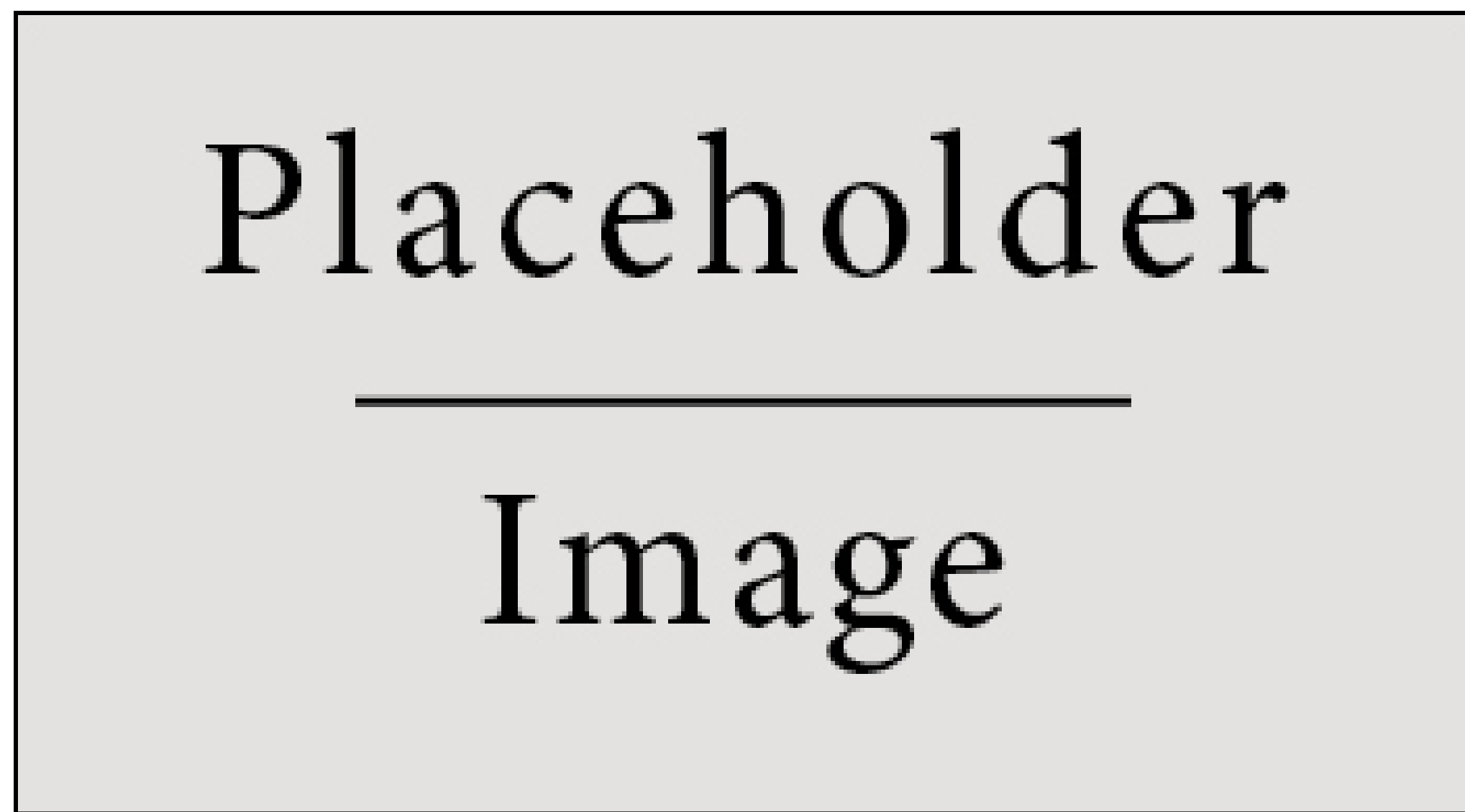


Figure 2: Scoring Function

## Results (I)

Model success curve 2.

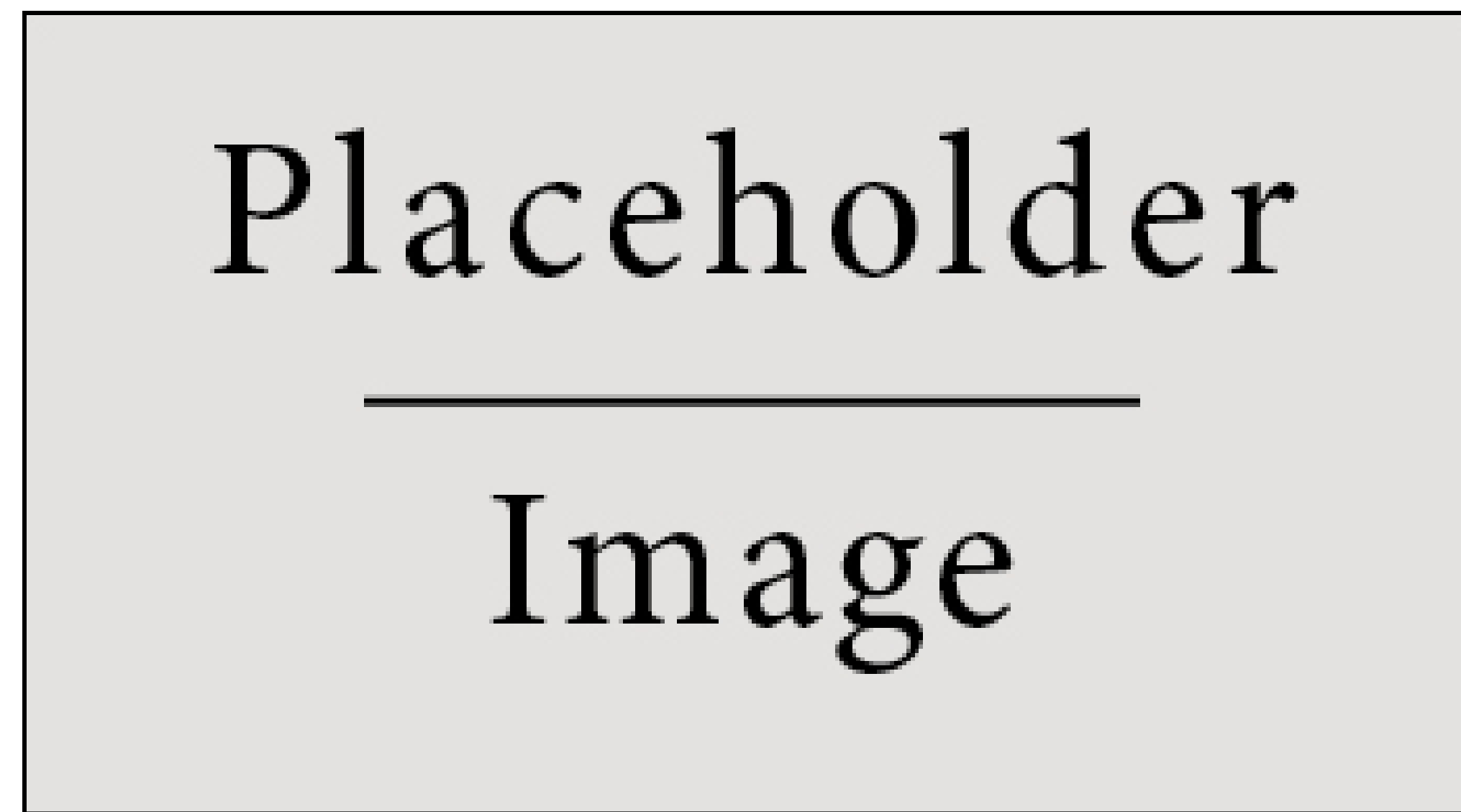


Figure 3: Scoring Function

## Results (II) & Discussion

Nunc tempus venenatis facilis. **Curabitur suscipit** consequat eros non porttitor. Sed a massa dolor, id ornare enim. Fusce quis massa dictum tortor **tincidunt mattis**. Donec quam est, lobortis quis pretium at, laoreet scelerisque lacus. Nam quis odio enim, in molestie libero. Vivamus cursus mi at *nulla elementum sollicitudin*. Maccenas ultricies feugiat velit non mattis. Fusce tempus arcu id ligula varius dictum.

- Curabitur pellentesque dignissim
- Eu facilisis est tempus quis
- Duis porta consequat lorem



Figure 5: This is a result curve.

## Choosing a Scoring Function

To measure *rank accuracy* and match the recommendation setting, we selected **log loss** as our scoring function.

$$-\frac{1}{n} \sum_{i=1}^n [y_i \log(p_i) - (1 - y_i) \log(1 - p_i)] \quad (1)$$

for  $n$  observations, where the  $i$ th observation is of correct class  $y_i \in \{0, 1\}$  which our model assigns probability  $p_i$ .

## Exploration (III): Spatial Predictors & Neighborhood Dynamics

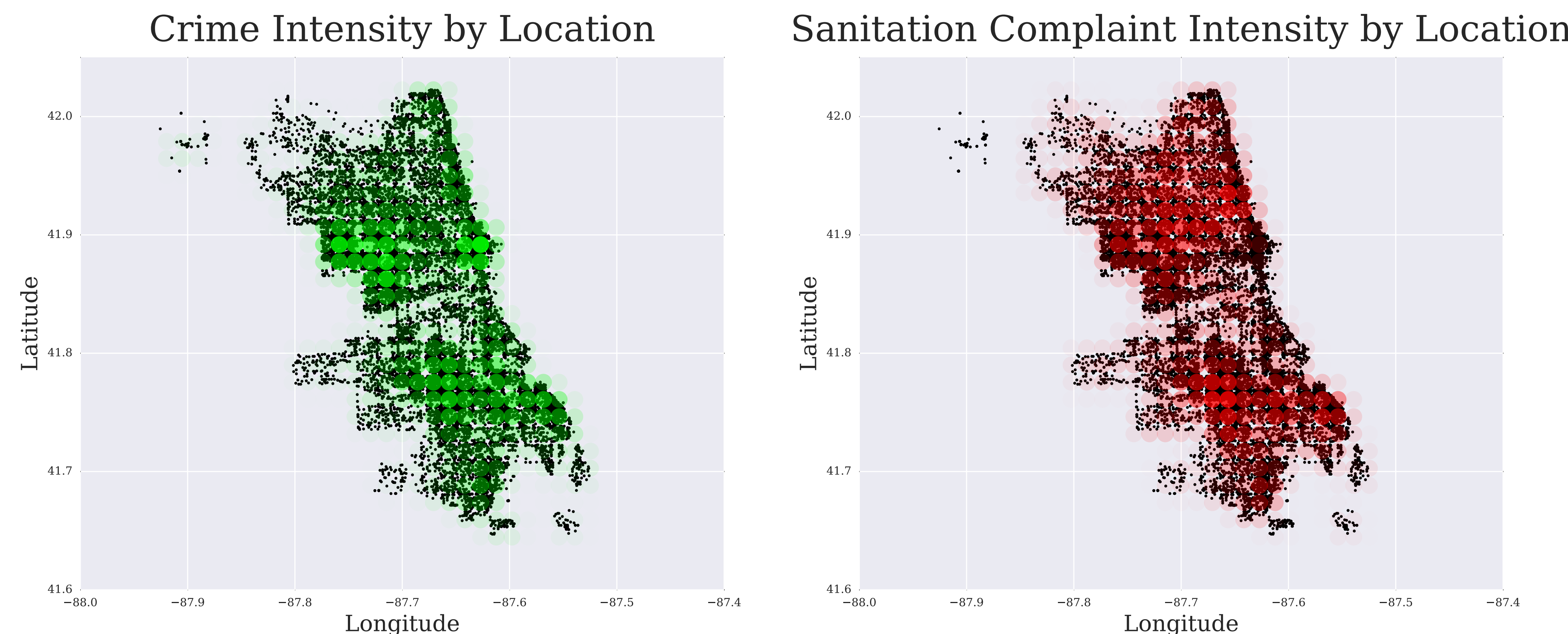


Figure 4: Crime (left) and Sanitation Complaint (Right) by Location. Darker shades represent a larger fraction of observations.

## Next Steps

LATIN LATIN LATIN LATIN LATIN LATIN  
LATIN LATIN LATIN LATIN LATIN LATIN  
LATIN LATIN LATIN LATIN

## Contact Information

- Web: [fggw.github.io/foodinspections/](https://fggw.github.io/foodinspections/)
- Code: [github.com/fggw/foodinspections/](https://github.com/fggw/foodinspections/)
- Contact: {lfarewell, jgober, samuelgreen, jeremywelborn}@college.harvard.edu

## References

- <https://chicago.github.io/food-inspections-evaluation/>