

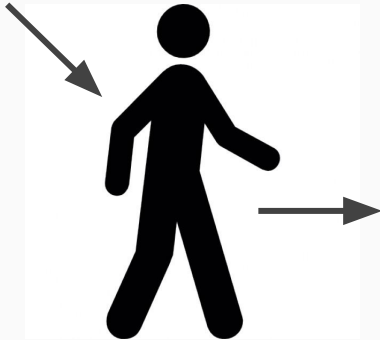
# WalkSafr

## Using data to improve your personal safety on the streets

**By Francois Charest, Insight Fellow,  
October 2015**

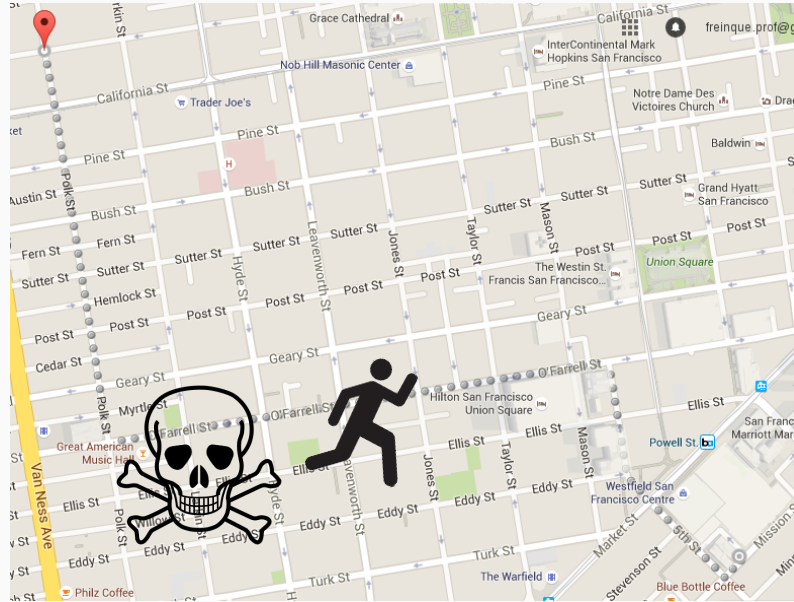
# Problem

I was visiting San Francisco a few months ago...



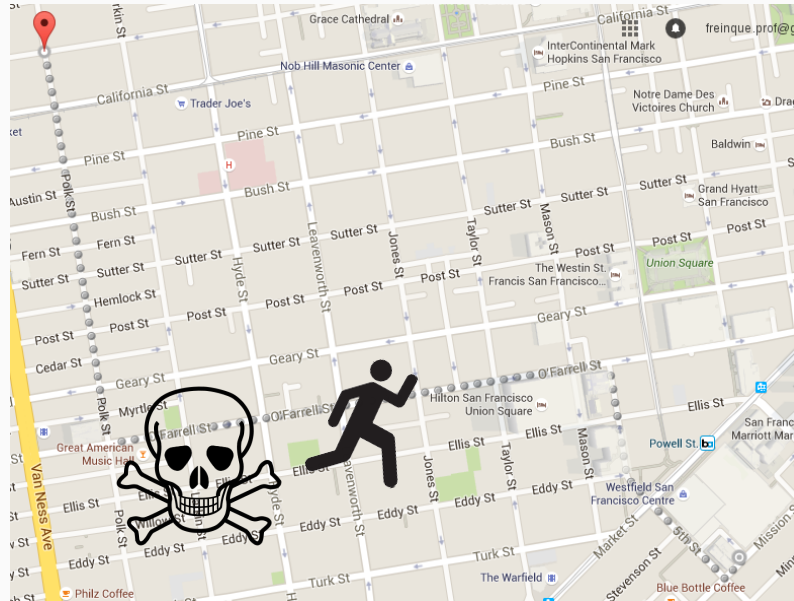
# Problem

**SF person:** “Maybe you should avoid those blocks on your way back.”



# Problem

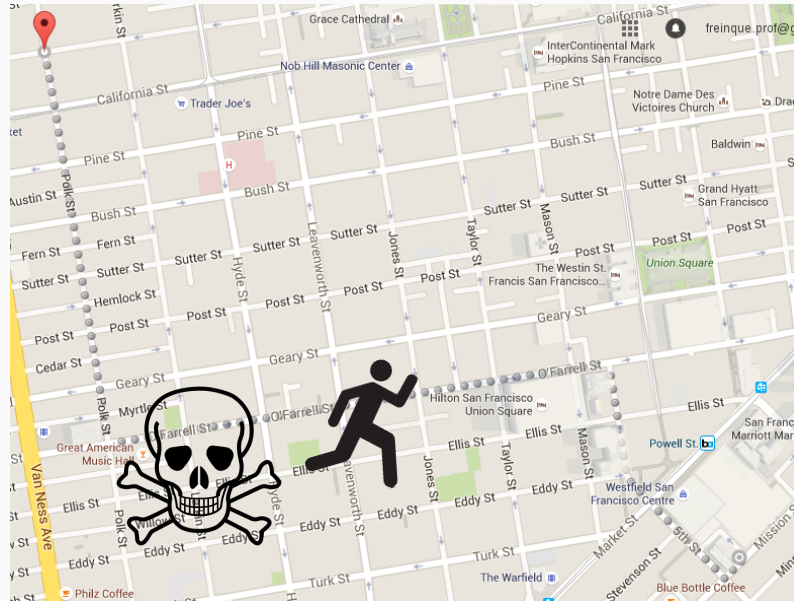
**SF person:** “Maybe you should avoid those blocks on your way back.”



**Was that right?**

# Problem

**SF person:** “Maybe you should avoid those blocks on your way back.”



**Was that right?**

**If yes, could I use data to provide such valuable advice?**

# Demo

Let's see....

# How?

GoogleMapAPIs



polygonal chains

$$\{p_i\} = \{p_1, p_2, \dots, p_l\}$$

# How?

GoogleMapAPIs



polygonal chains

$$\{p_i\} = \{p_1, p_2, \dots, p_l\}$$

SF crime data

SF OpenData

SF population data



KDE

cross-validation

local crime density

local population density

$$f(\mathbf{x}) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$



# How?

GoogleMapAPIs



polygonal chains

$$\{p_i\} = \{p_1, p_2, \dots, p_l\}$$

integration

relative crime/danger levels

$$f(\{p_i\}) = \sum_{i=1}^l f(p_i)$$

$$\text{relative\_level}(\{p_i\}, \{q_i\}) = \frac{f(\{p_i\})}{f(\{q_i\})}$$

SF crime data

SF OpenData

SF population data



KDE

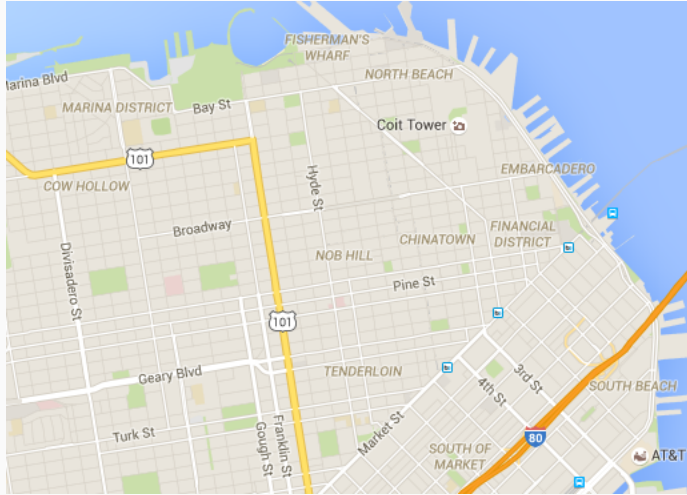
cross-validation

local crime density

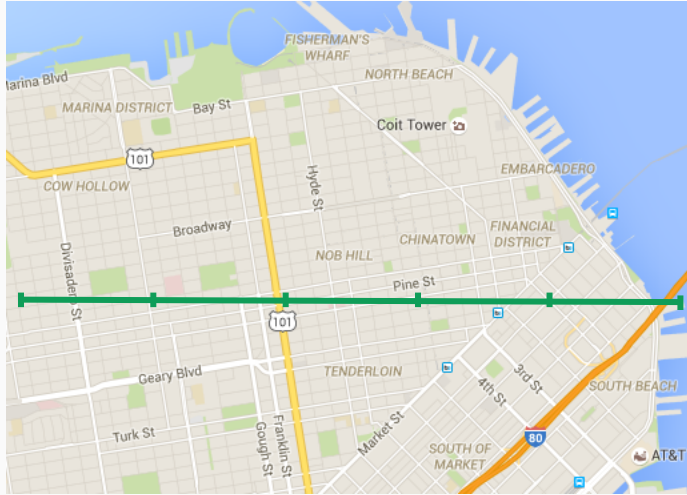
local population density

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

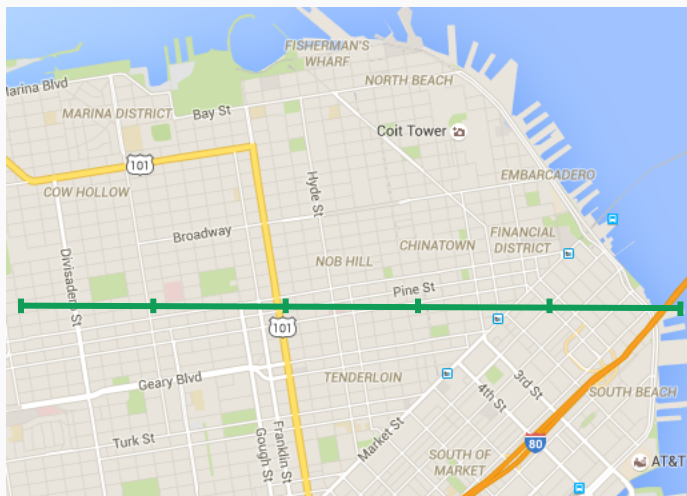
# How well?



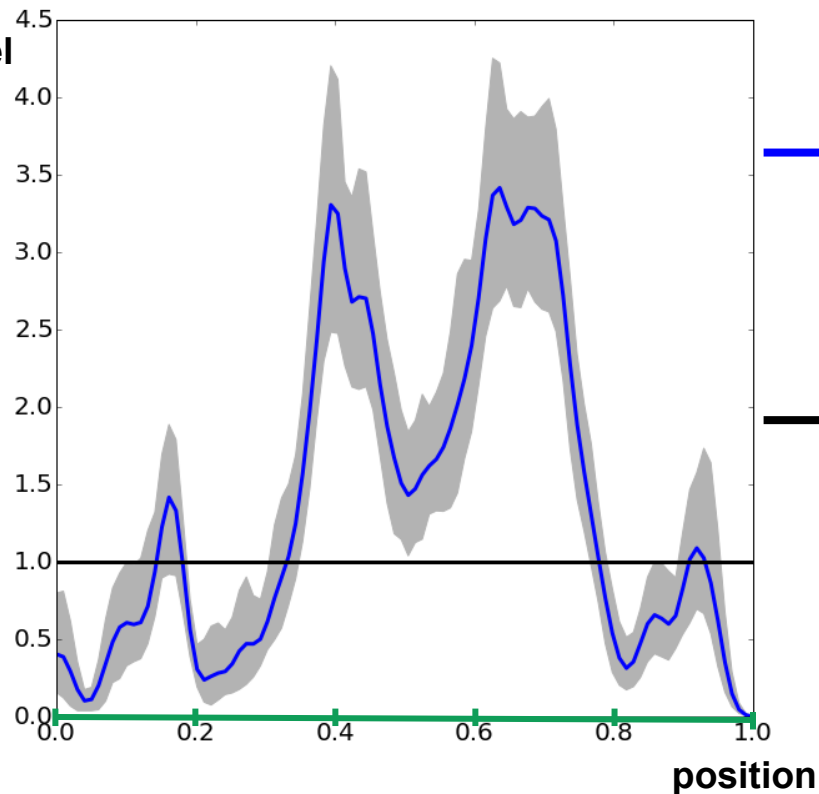
# How well?



# How well?



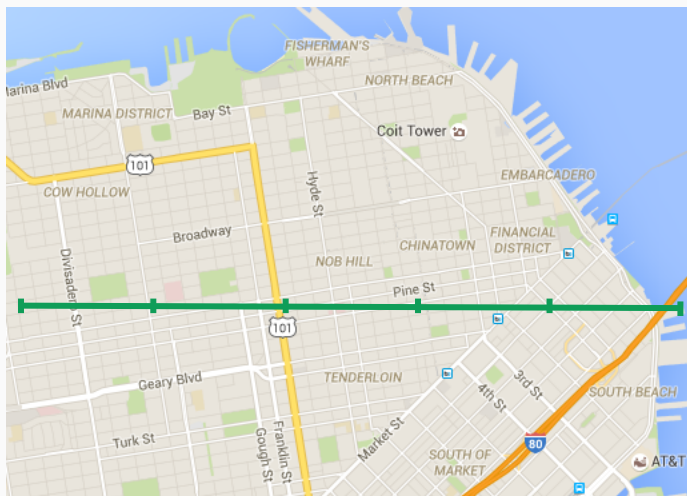
Predicted crime rate along the line at 4pm



KDE last 20 days

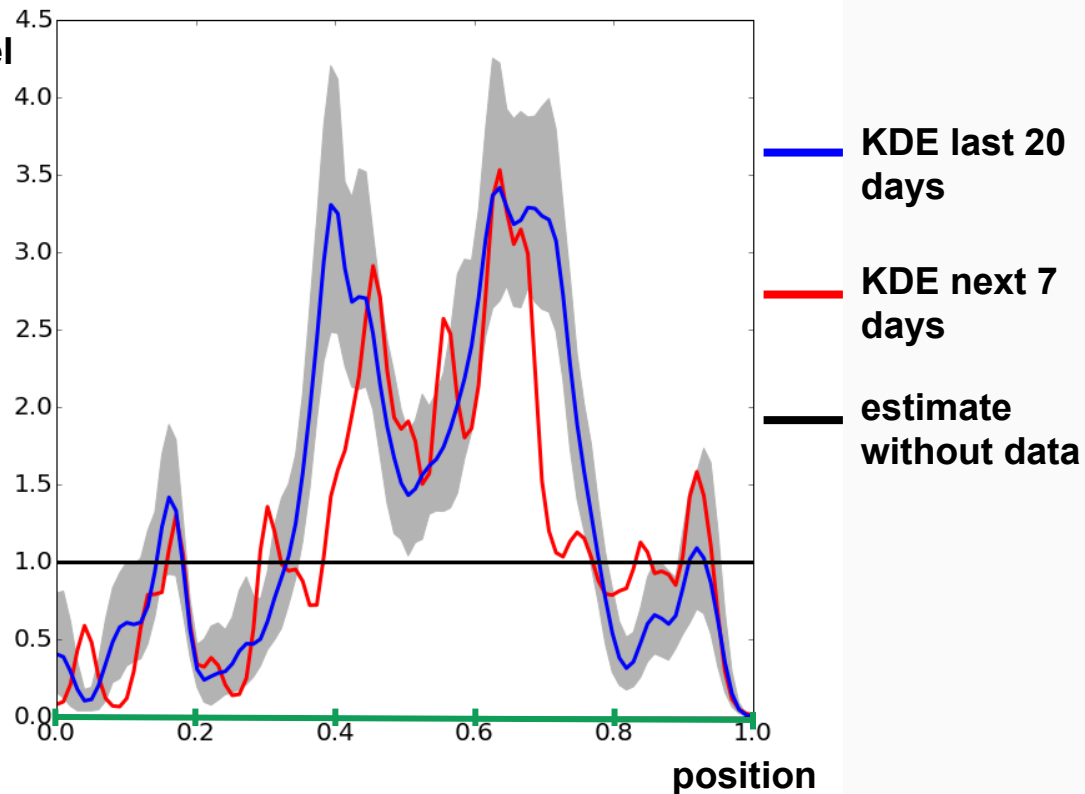
estimate without data

# How well?

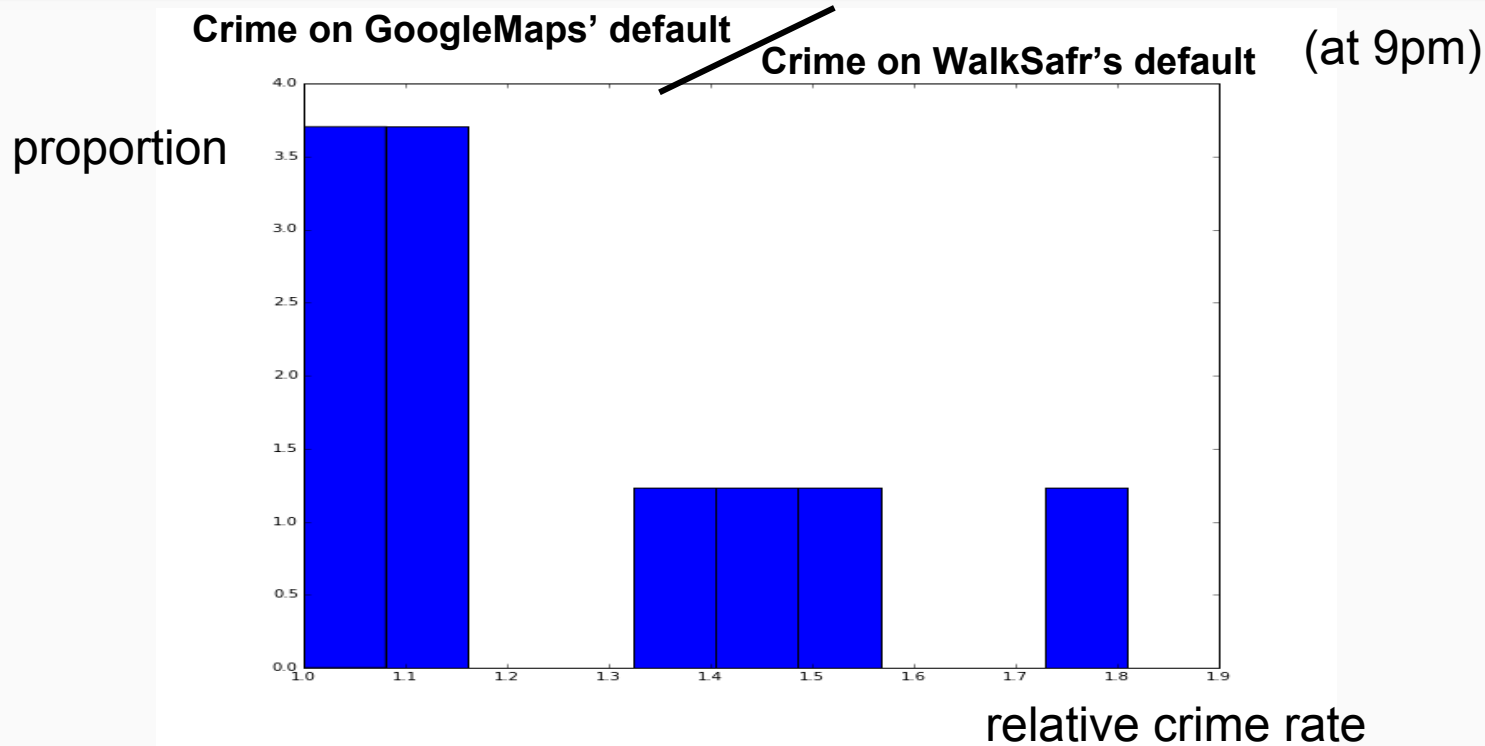


crime level

Predicted crime rate along the line at 4pm



# Routes 20% safer than GoogleMaps

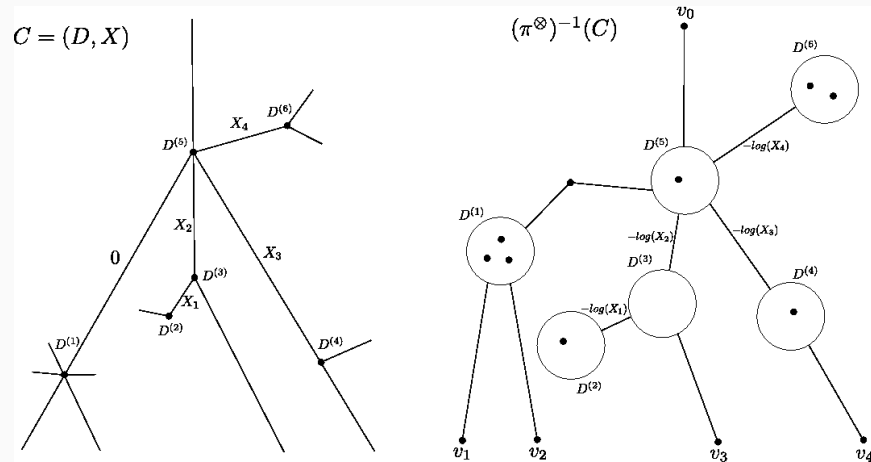
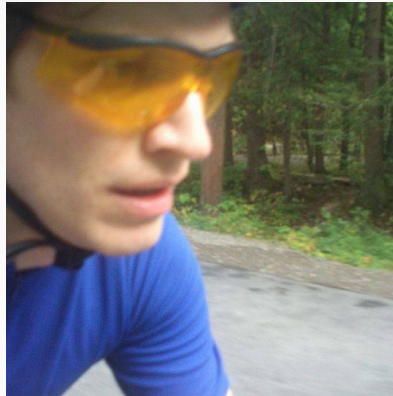


average ~ 1.25

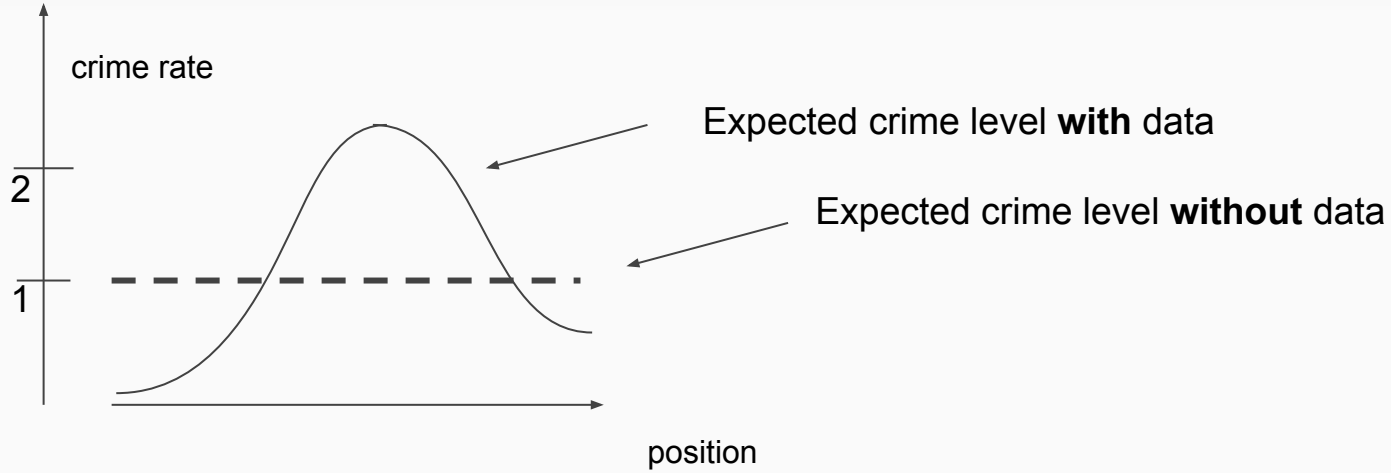
# François Charest

Ph.D. Math (UofMontreal)

PostDoc (ColumbiaU)

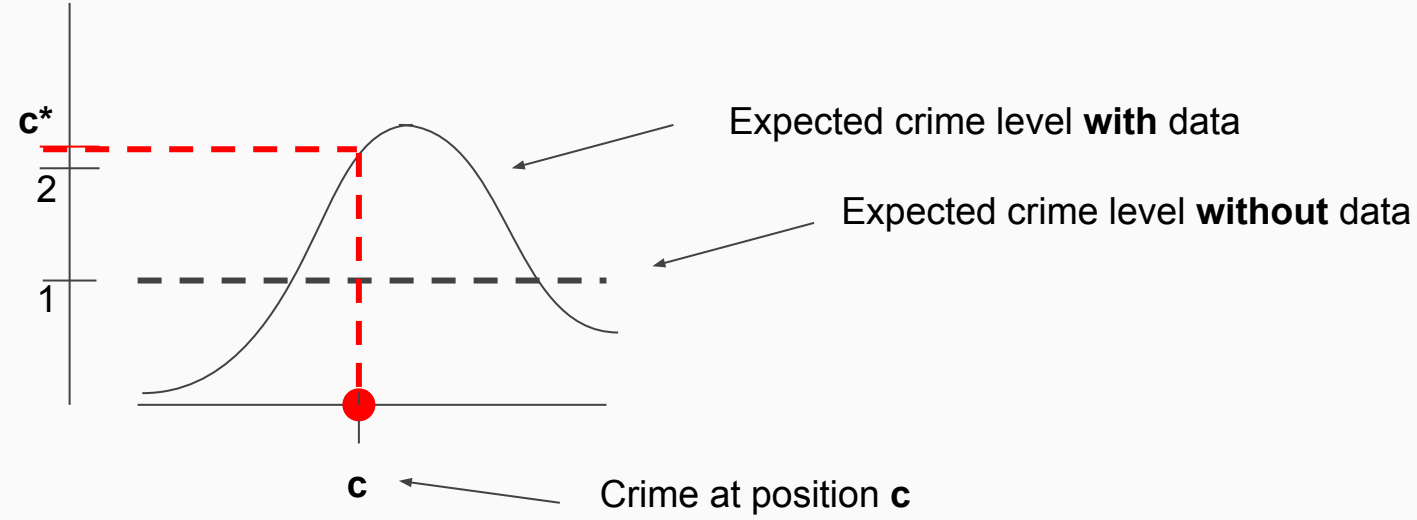


# How well?

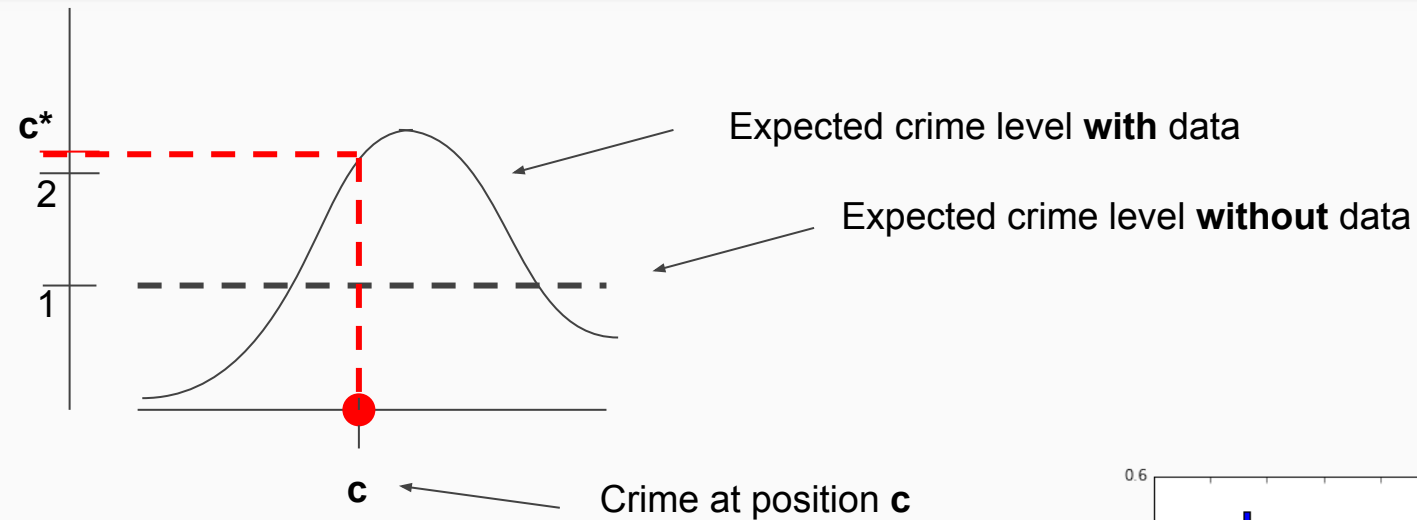




# How well?

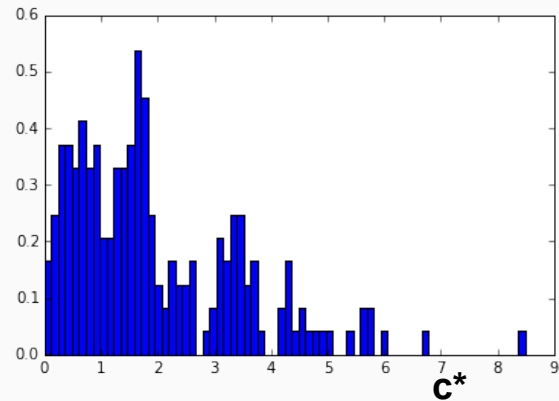


# How well?



performance :=  $E(c^*) \sim 2$

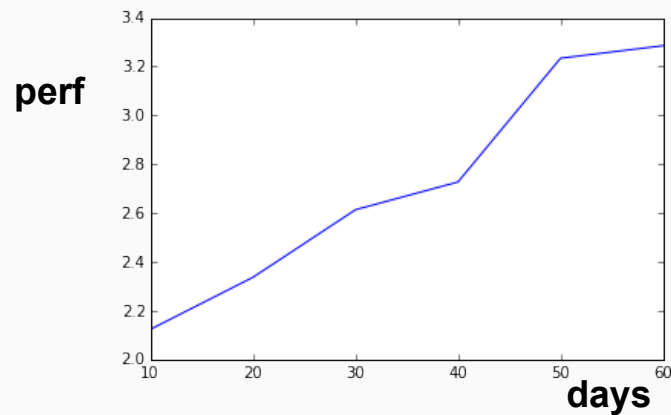
prop



# Observations

**Q) How far in the past should we consider data to predict future crimes?**

Given the sparsity of the considered data, as long as computationally possible:



**Q) How helpful are the last hours?**

They seem to be.

