CAS RPM 2014 Washington DC | PM-14

Introduction to Nonparametric Regression

UBI Analytics for Mileage & Daytime Discounts™

Ryan N. Morrison
Founder & CEO | True Mileage, Inc.

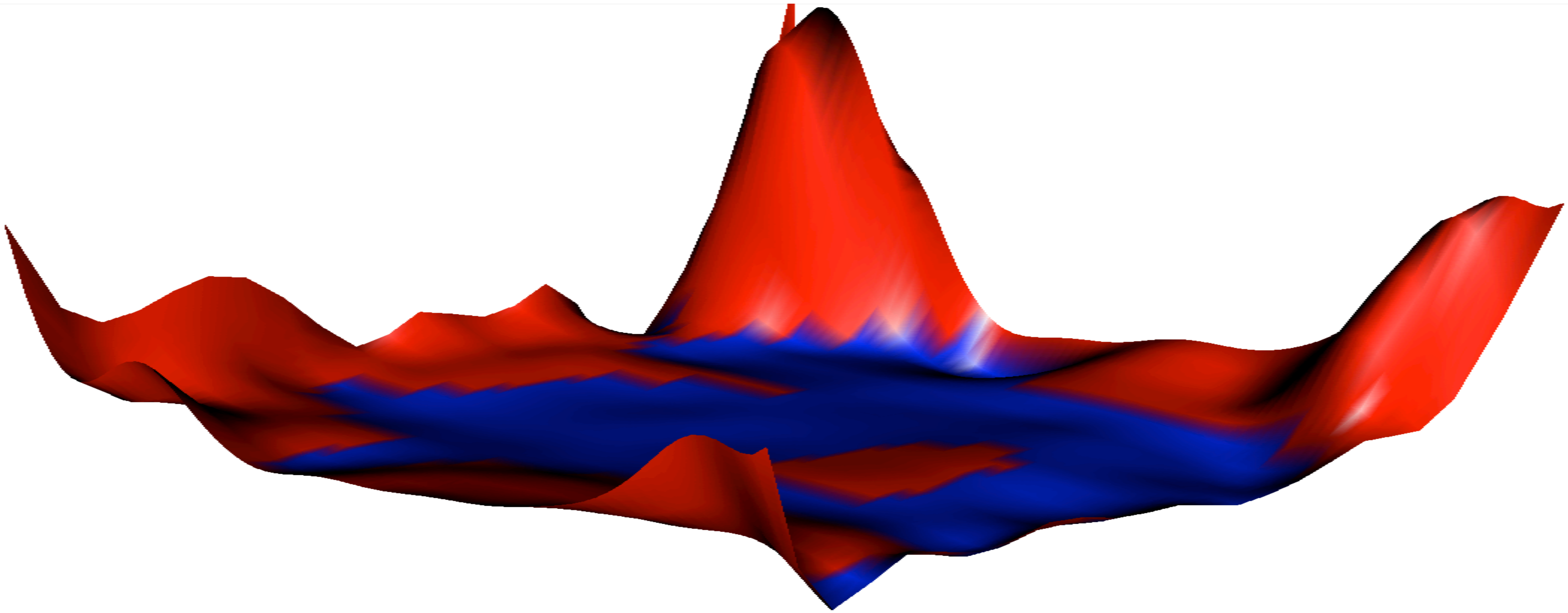
Daniel Hernandez-Stumpfhauser PhD
Lead Statistician | True Mileage, Inc.

Agenda

1) About us
2) Intro to Nonparametric Regression
3) Mileage Discount Analytics™
4) Daytime Discount Analytics™

TRUE MILEAGE

Nonparametric Regression



Are male or female drivers safer?

This nonparametric surface will answer that question today!

TRUE MILEAGE

UBI Issues

Technology

Devices and data transfer.

-Telecomm Fees

-Privacy Issues

-6 Months, 30%

Analytics

How big should discounts be?

-Historic Data

-Rating Plan Associations

-Time to Implement

TRUE MILEAGE

Solution

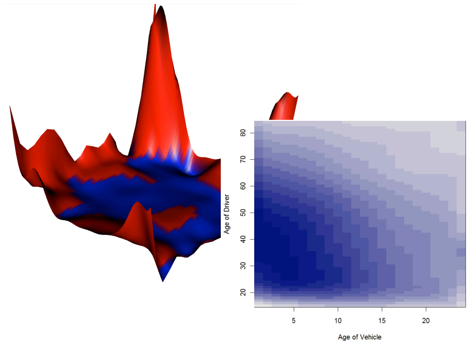
Technology

Devices and data transfer.

- -Devices 25% less
- -Transfer 100% less
- -Privacy Sensitive
- -Discounts 50-60%

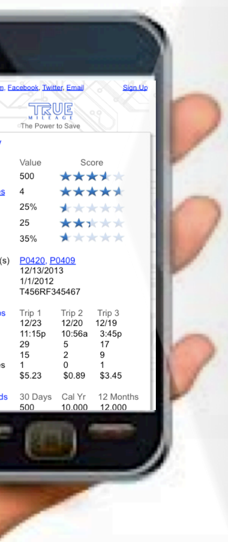
Analytics

How big should discounts be?



- -National Database
- -Accounts for Rating Plan
- -Ready Immediately

TRUE MILEAGE



Summary

30 Days

	Value	Score
Mileage	500	
Hard Brakes	4	
Late Night	25%	
MPG	25	
Rush Hour	35%	

Error Codes	P0420, P0409
Transfer	12/13/2013
Install	1/1/2012
VIN	T456RF345467

Recent Trips	Trip 1	Trip 2	Trip 3
Date	12/20	12/20	12/19
Start	11:15p	10:56a	3:45p
Minutes	29	5	17
Miles	15	2	9
Hard Brakes	1	0	1
Cost	$5.23	$0.89	$3.45

Periods	Cal. Yr.	12 Mo.	All

Agenda

1) About us

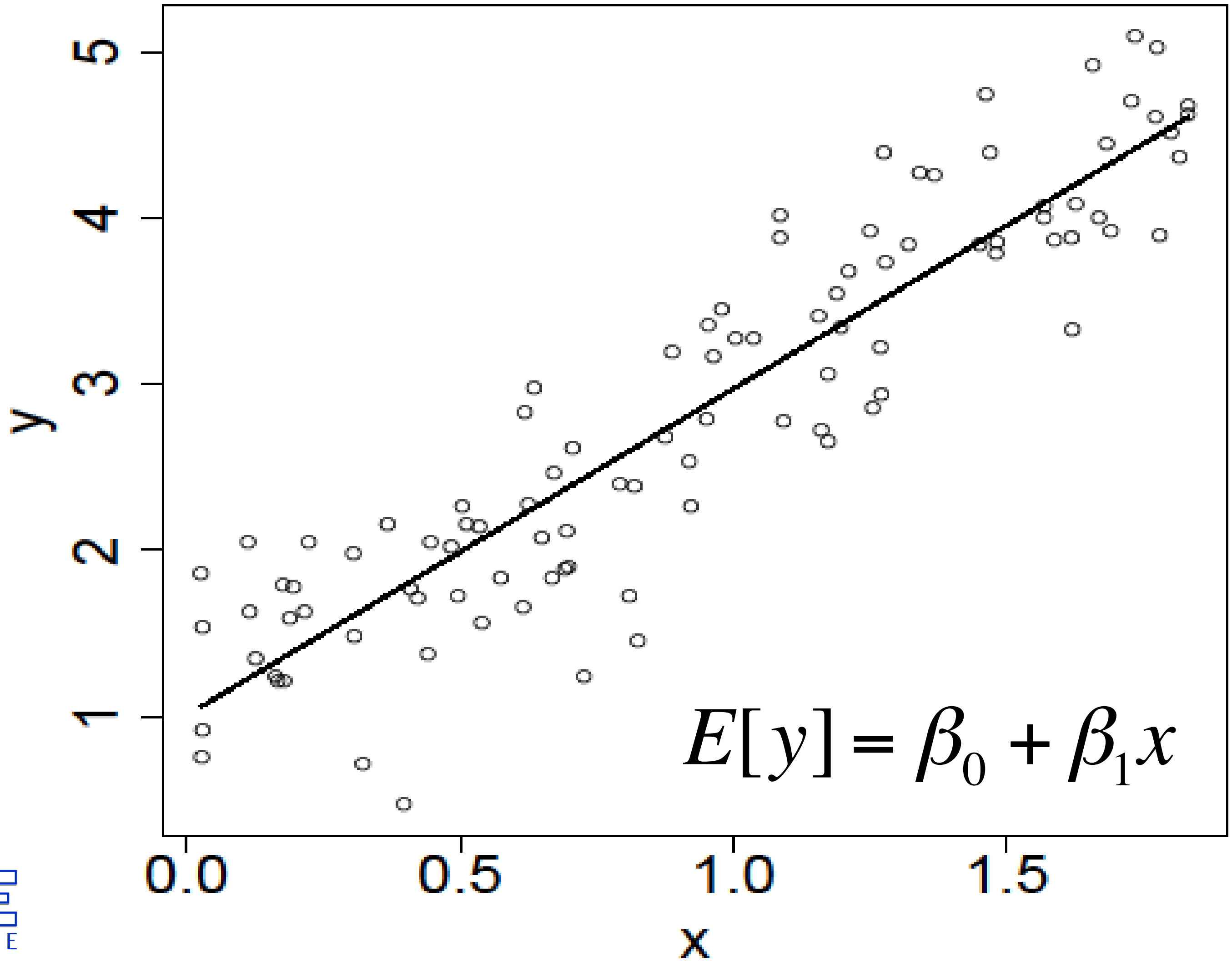
2) Intro to Nonparametric Regression

3) Mileage Discount Analytics™

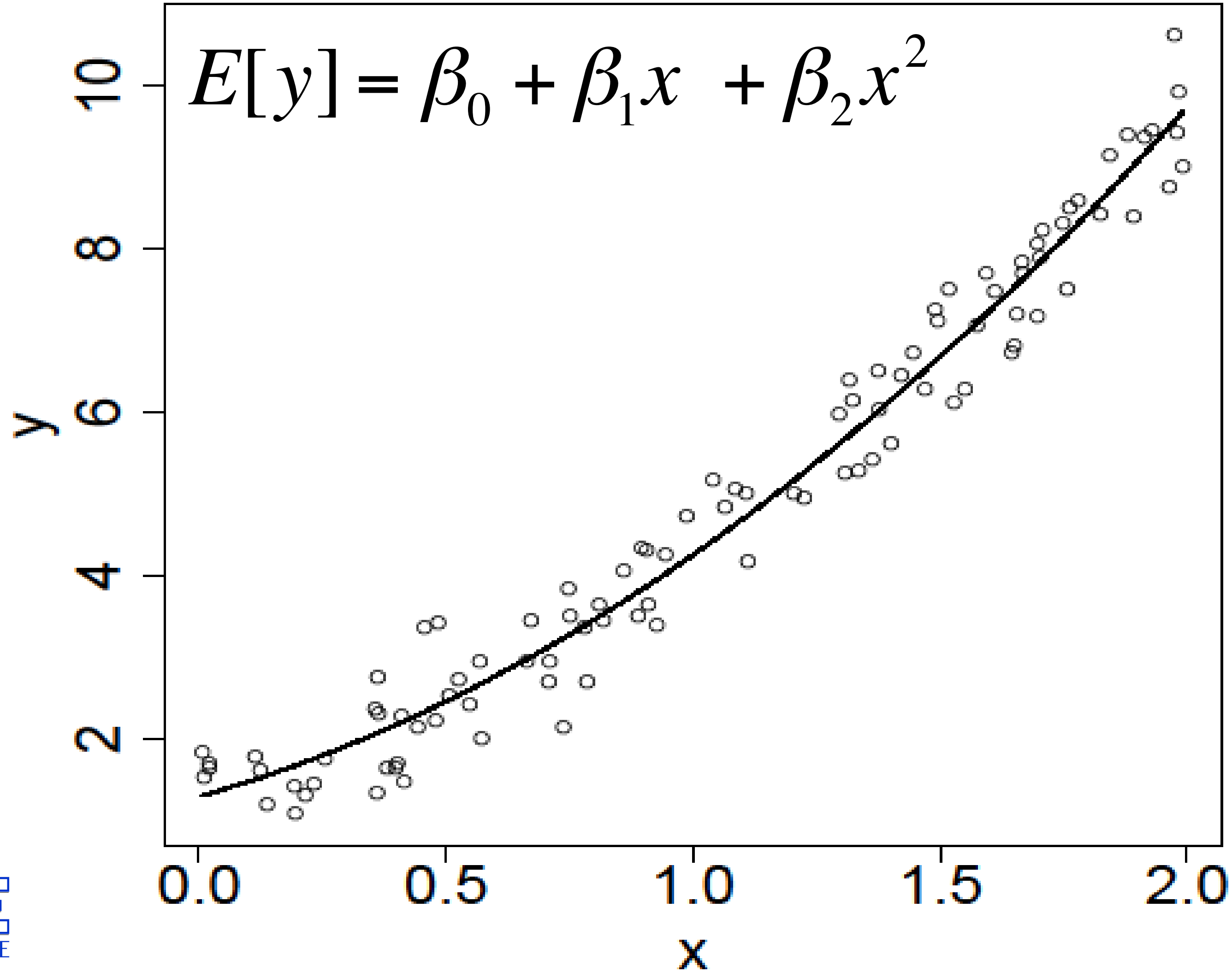
4) Daytime Discount Analytics™

TRUE MILEAGE

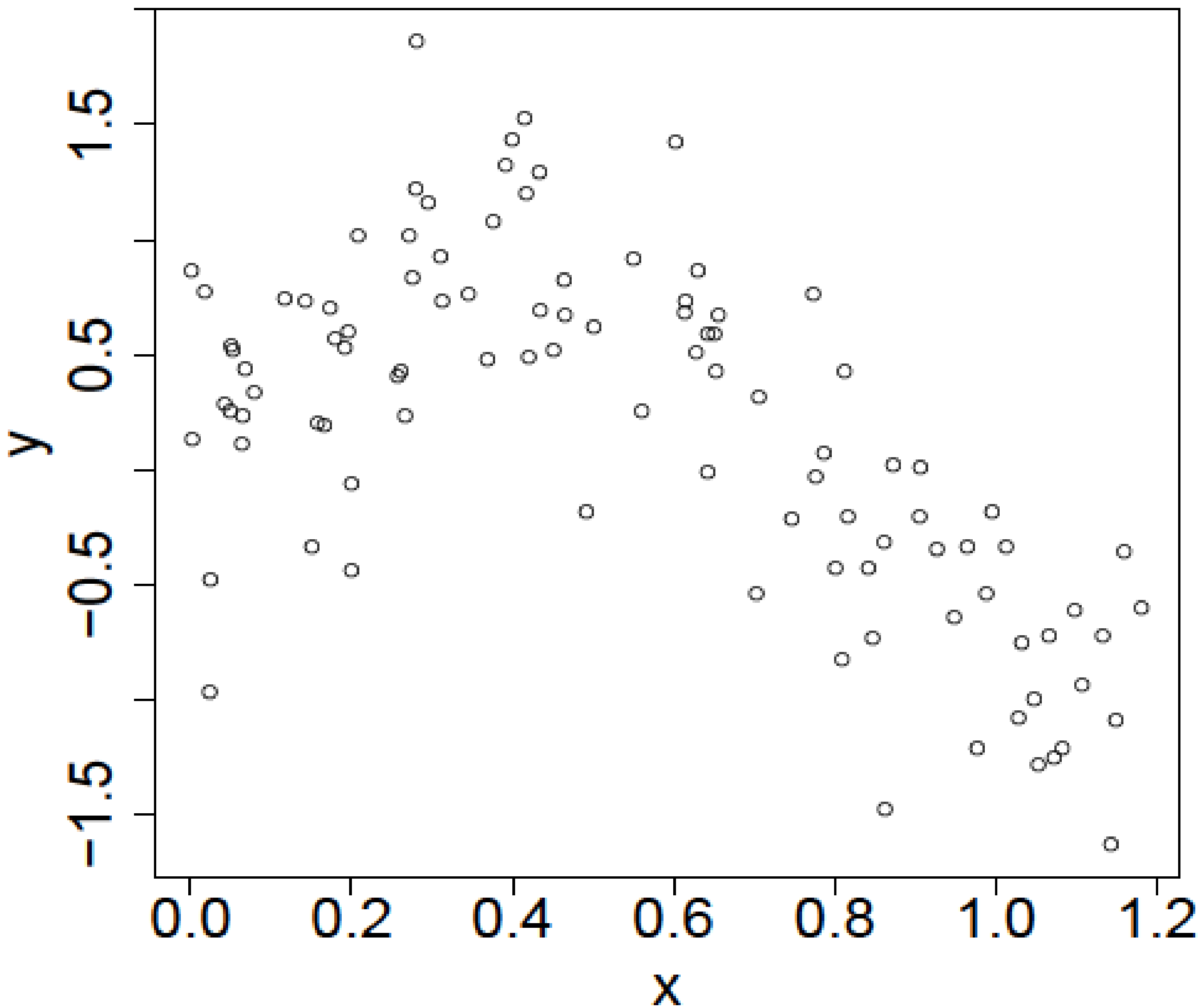
Linear Regression



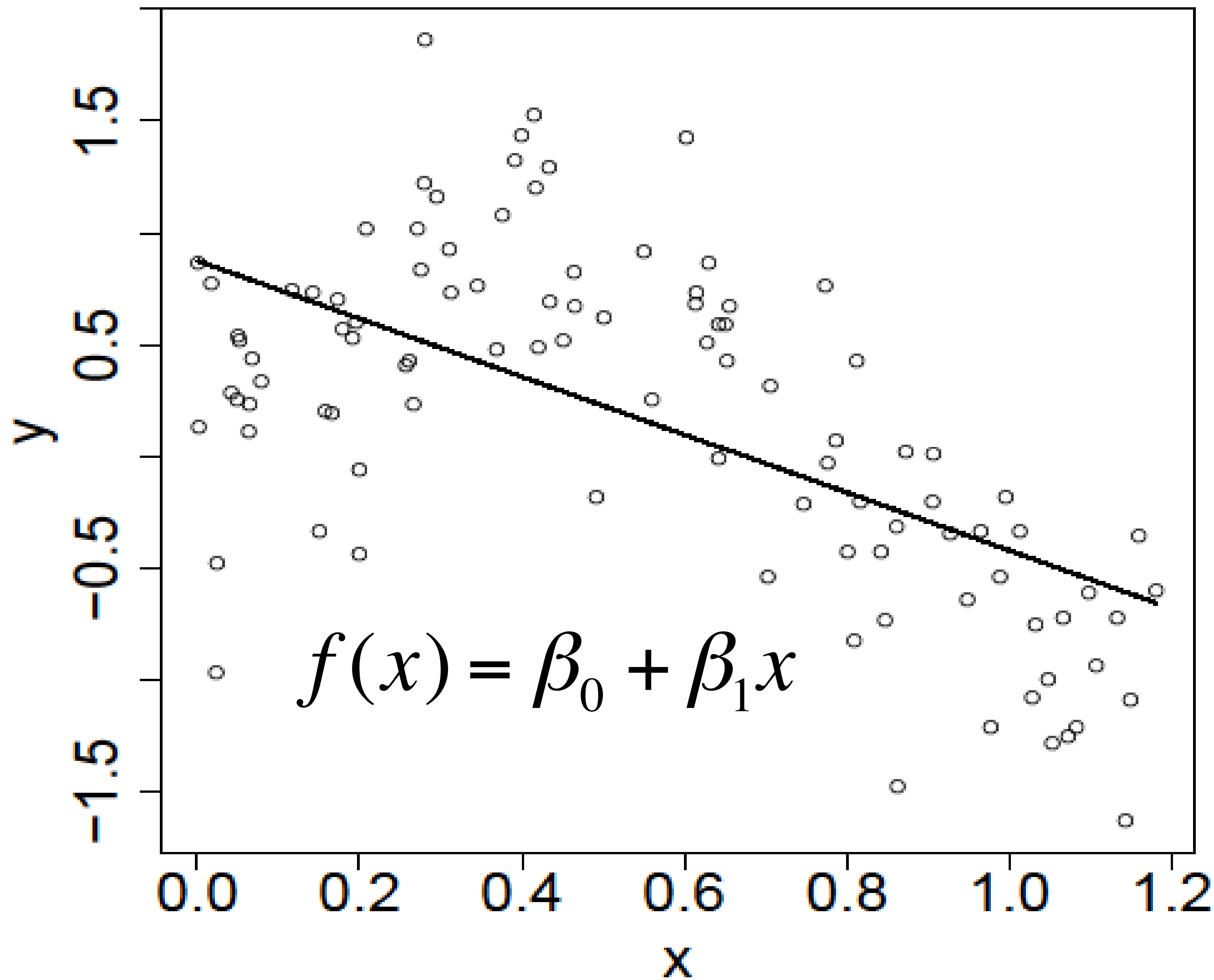
Polynomial Regression



$$E[y] = \sin(4x)$$

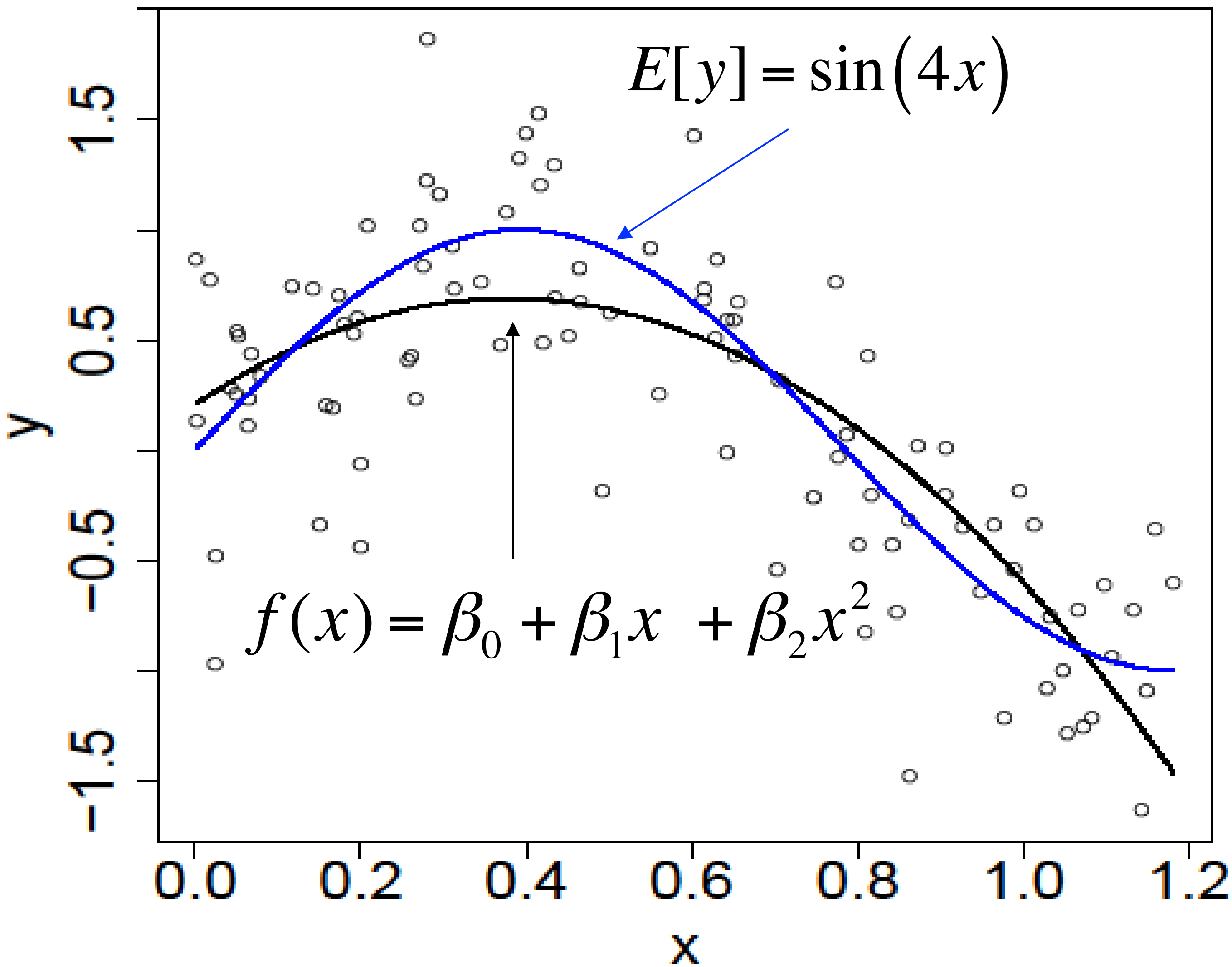


Linear Regression



$$f(x) = \beta_0 + \beta_1 x$$

Polynomial Regression



Nonparametric Regression

Goals

- Given a scatterplot;
- We want to find a function $f(\mathbf{x})$ that best predicts the dependent variable y

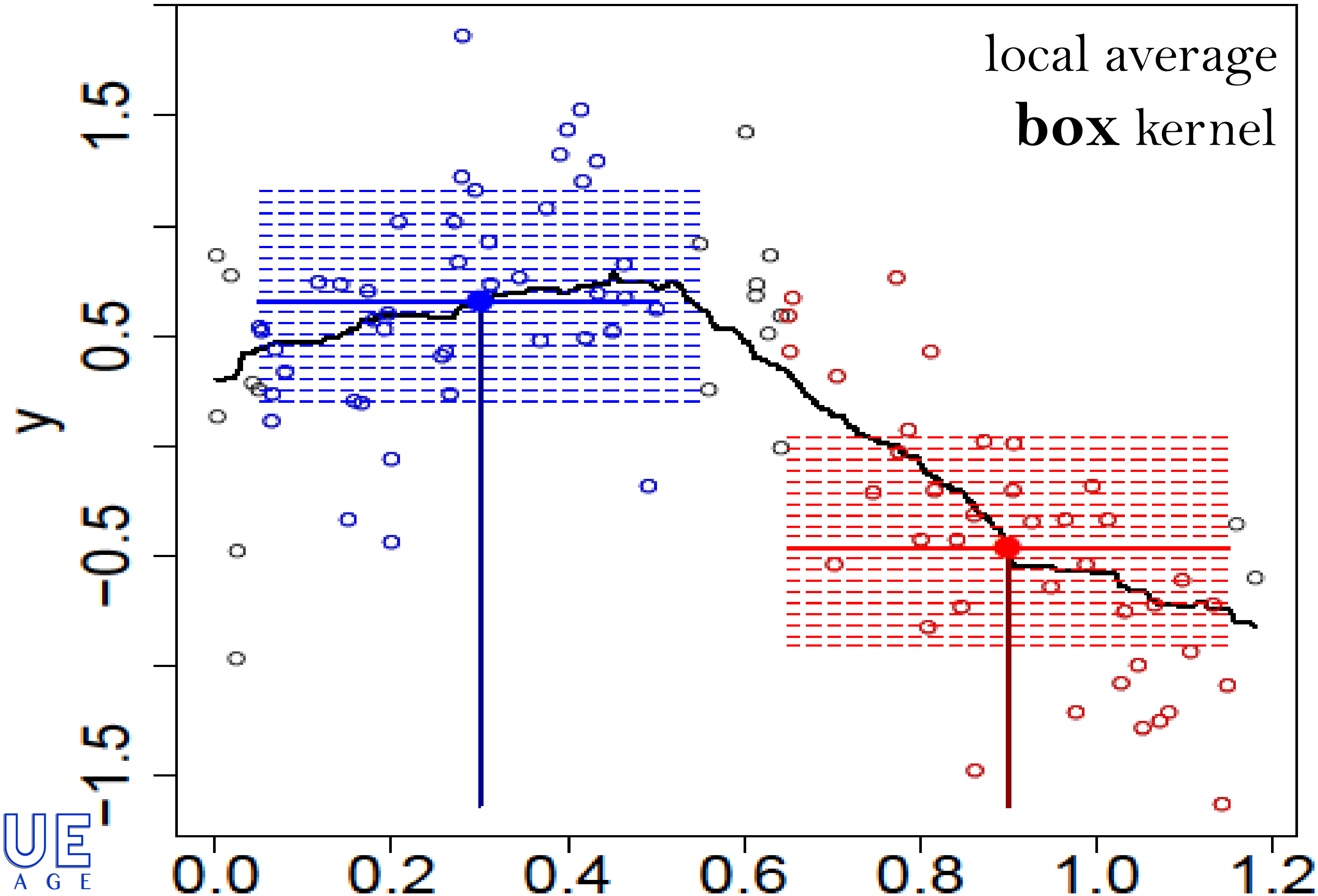
TRUE MILEAGE

Nonparametric Regression

- Estimate smooth regression function $f(X)$ at each target point X_0
- Use only those observations close to the target point X_0
- Smooth localization is achieved using a *kernel* $K_\lambda(X_0, X_i)$
- The *width* of the neighborhood λ controls the smoothness, bias and variance.

TRUE MILEAGE

Nonparametric Regression



Nonparametric Regression

- Nadaraya-Watson kernel-weighted average

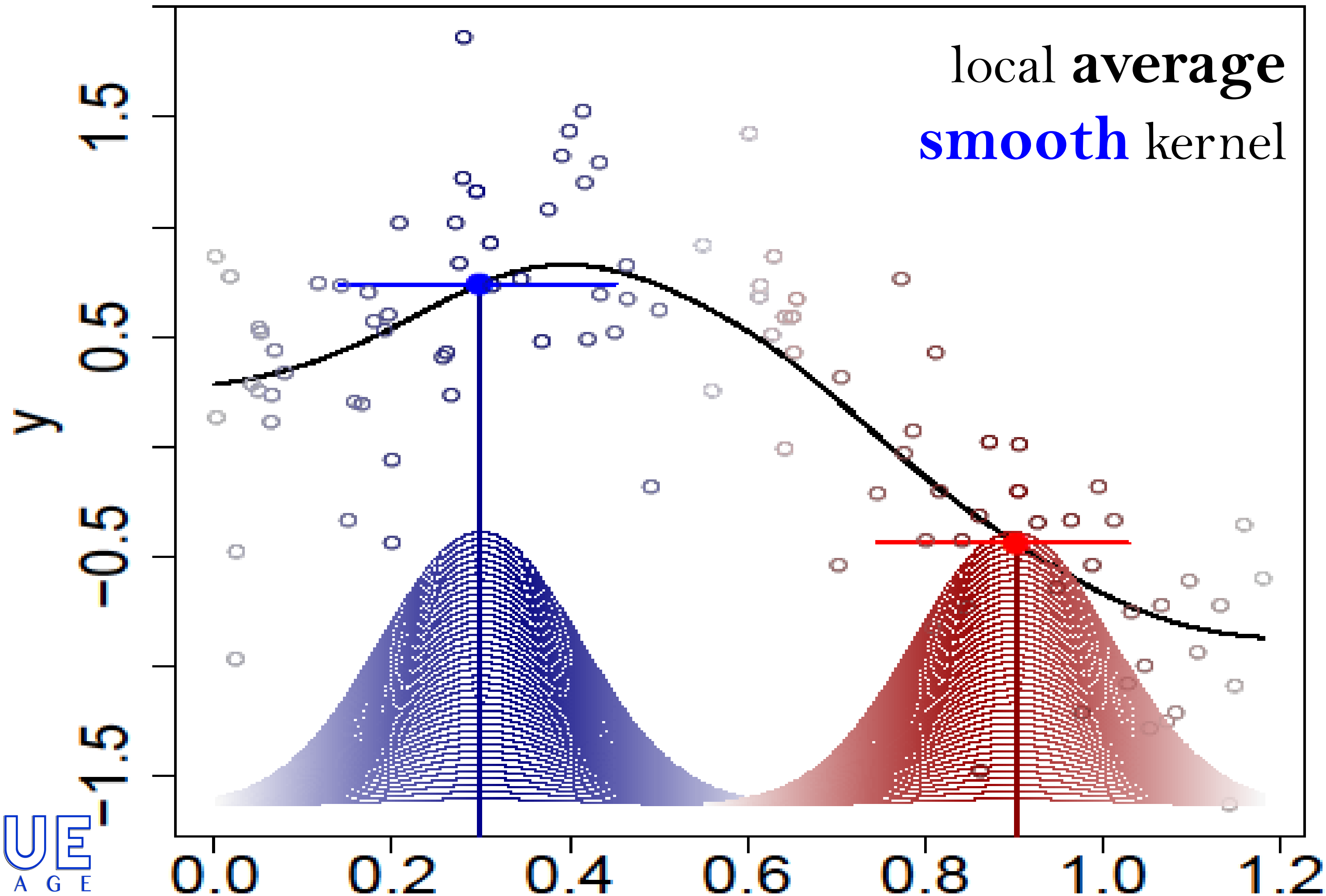
$$\hat{f}(x_0) = \sum_{i=1}^{n} \left(\frac{K_\lambda(x_0, x_i)}{\sum_{i=1}^{n} K_\lambda(x_0, x_i)} \right) y_i$$

- The Gaussian density function is a popular choice of kernel

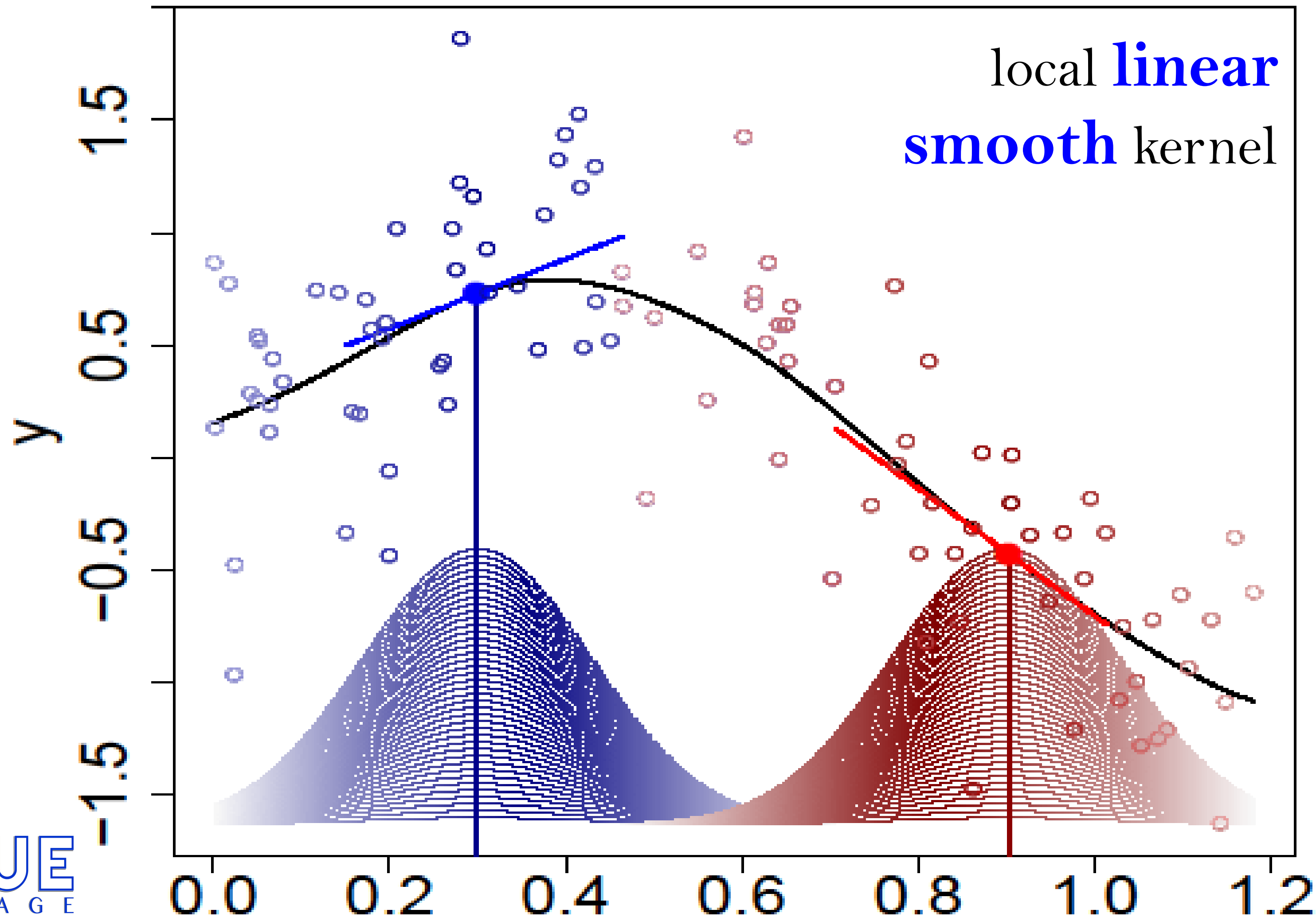
$$K_\lambda(x_0, x) = \phi\left(\frac{x - x_0}{\lambda} \right)$$

TRUE MILEAGE

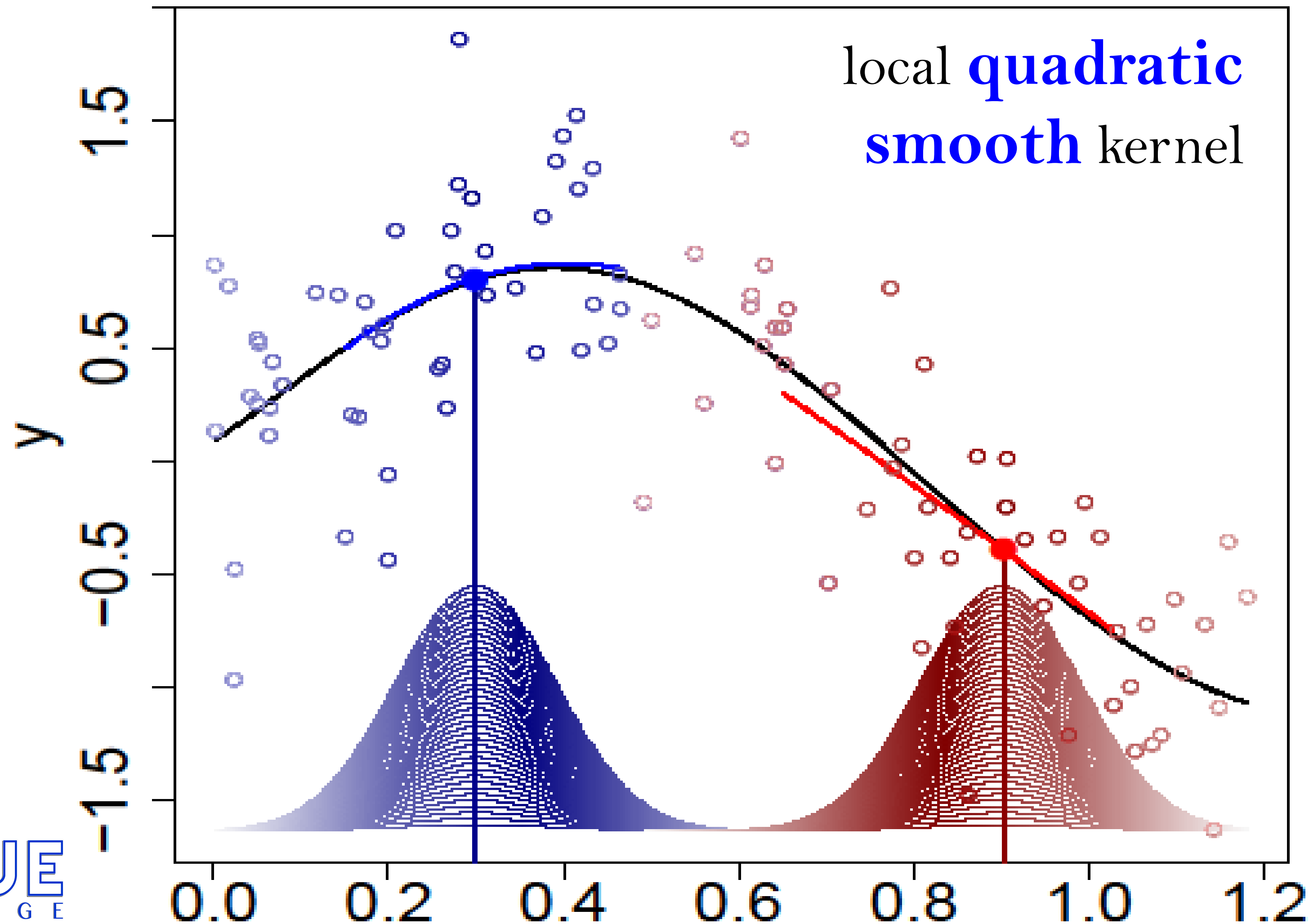
Nonparametric Regression



Nonparametric Regression



Nonparametric Regression



Nonparametric Regression

- Kernel local regression generalizes naturally to higher dimensions.
- Let $b(X)$ be a vector of polynomial terms in X
  e.g.
  $$b(X) = \left(1, X_1, X_2, X_1^2, X_2^2, X_1 X_2\right)$$
- Let $K_\lambda(x_0, x_i)$ be a 2-dimensional kernel.
  At each $x_0 \in \Re^2$ minimize,
  $$\sum_{i=1}^{n} K_\lambda(x_0, x_i)\left[y_i - b(x_i)^T \beta(x_0)\right]^2$$

TRUE MILEAGE

Agenda

1) About us

2) Intro to Nonparametric Regression

3) Mileage Discount Analytics™

4) Daytime Discount Analytics™

TRUE MILEAGE

Mileage Discount Analytics™

Example 1:

Mileage	Discount
500	45%
3,500	26%
6,000	19%
8,500	16%
11,000	11%
13,000	7%
16,000+	1%

TRUE MILEAGE

Mileage Discount Analytics TM

Example 2:

Mileage Up To	Discount
2,500	54%
5,000	39%
7,500	34%
10,000	26%
12,500	18%
15,000	13%
15,000 +	7%

TRUE MILEAGE

Mileage Discount Analytics™

Rating variable with the strongest mileage relationship?

- Driver Age
- Urban vs. Rural
- Vehicle Type
- Driver Gender
- Drivers/Vehicles
- Vehicle Age

TRUE MILEAGE

Mileage Discount Analytics™

Rating variable with the strongest mileage relationship?

- Driver Age
- Urban vs. Rural
- Vehicle Type
- Driver Gender
- Drivers/Vehicles
- Vehicle Age

TRUE MILEAGE

Mileage Discount Analytics TM

Driver Age

- 18 yr ~ 11,000
- 48 yr ~ 13,000
- 70 yr ~ 9,000

Vehicle Age

- New ~ 14,000
- Old ~ 8,000

TRUE MILEAGE

Mileage Discount Analytics™

Driver Age

- 18 yr ~ 11,000
- 48 yr ~ 13,000
- 70 yr ~ 9,000

Vehicle Age

- New ~ 14,000
- Old ~ 8,000

Should a 10,000 mile vehicle get a discount?

TRUE MILEAGE

Mileage Discount Analytics™

Should a 10,000 mile vehicle get a discount?

Driver Age

- 18 yr ~ 11,000
- 48 yr ~ 13,000
- 70 yr ~ 9,000

Vehicle Age

- New ~ 14,000
- Old ~ 8,000

Not always! It would be a double discount for older drivers and vehicles.

TRUE MILEAGE

Mileage Discount Analytics™

How do we resolve the double discounting issue?

Rating Mileage: The mileage a vehicle is effectively being charged for in an existing rating plan.

- Discount vehicles only if below their rating mileage

Rating Mileage = function(Vehicle Age, Driver Age)

Rating Mileage Model

1) Data: Unbiased national data set with hundreds of thousands of mileage observations.

2) Variables: The most predictive rating variables are driver age and vehicle age.

3) Goal: Estimate rating mileage, the mileage a vehicle is effectively charged for through a typical rating plan.

TRUE MILEAGE

Rating Mileage Model

- We model annual mileage averages y_i as

$$E[y_i] = f\left(x_{1i}, x_{2i}\right)$$

where $\left(x_{1i}, x_{2i}\right)$ is driver age and vehicle age, respectively.

- We model y_i as having variance equal to σ^2 / n_i

TRUE MILEAGE

Rating Mileage Data

Vehicle \ Driver	**15**	**16**	**…**	**84**
1	10,598	12,771	…	8,786
2	14,335	12,385	…	8,633
…	…	…	…	…
24	8,513	8,882	…	6,703

TRUE MILEAGE

Rating Mileage Model

- At each point $\mathbf{x}_0 = (x_{1,0}, x_{2,0})$ we estimate $f(x_{1,0}, x_{2,0})$ via kernel methods

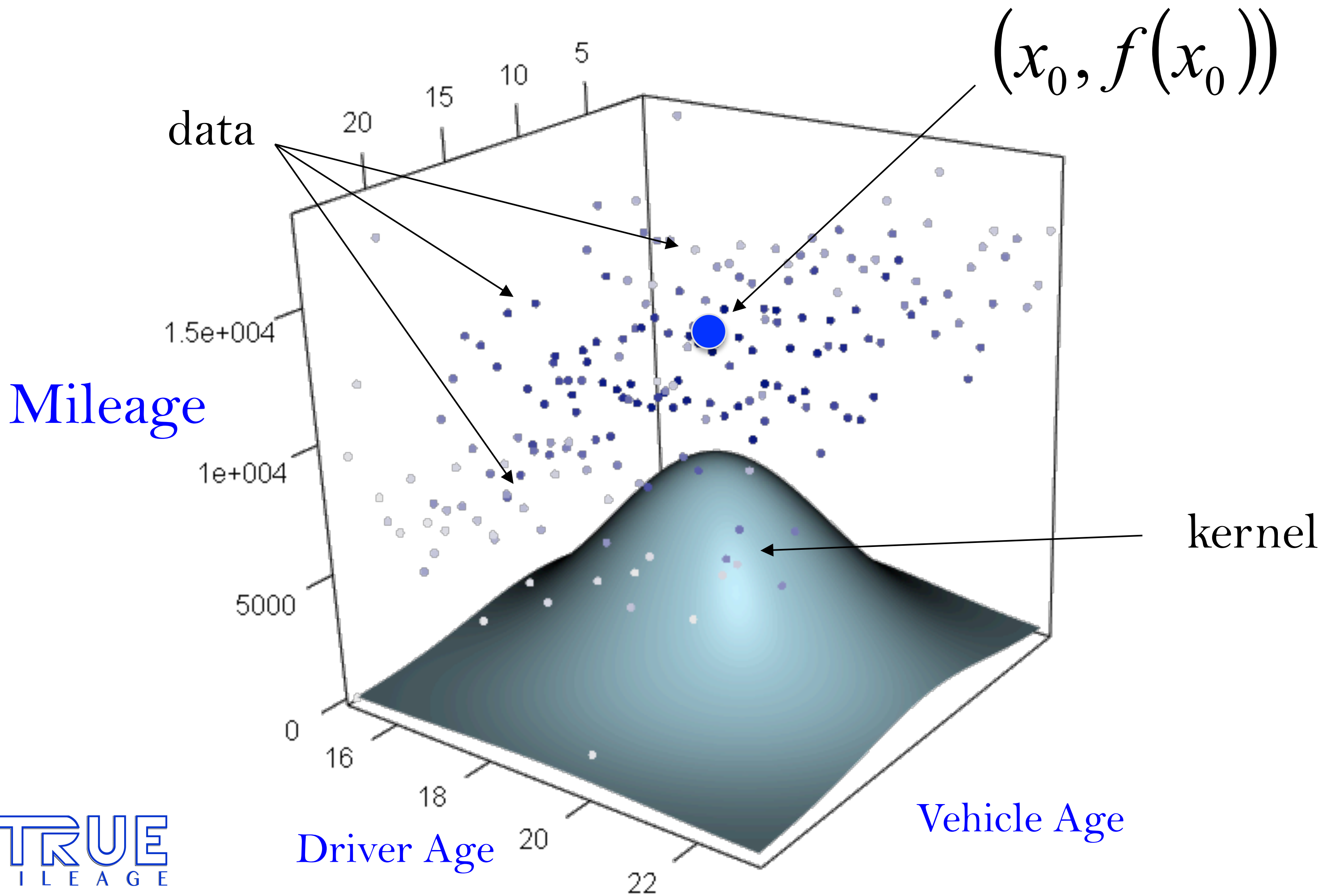
$$\hat{f}(x_{1,0}, x_{2,0}) = b(x_{1,0}, x_{2,0})^T \hat{\beta}_0$$

where $b(x_{1,0}, x_{2,0})^T = (1, x_{1,0}, x_{2,0}, x_{1,0}^2, x_{2,0}^2, x_{1,0}x_{2,0})$
and $\hat{\beta}_0$ minimizes

$$\sum_{i=1}^{n} K_\lambda(\mathbf{x}_0, \mathbf{x}_i) n_i \left[y_i - b(x_{1i}, x_{2i})^T \beta_0 \right]^2$$

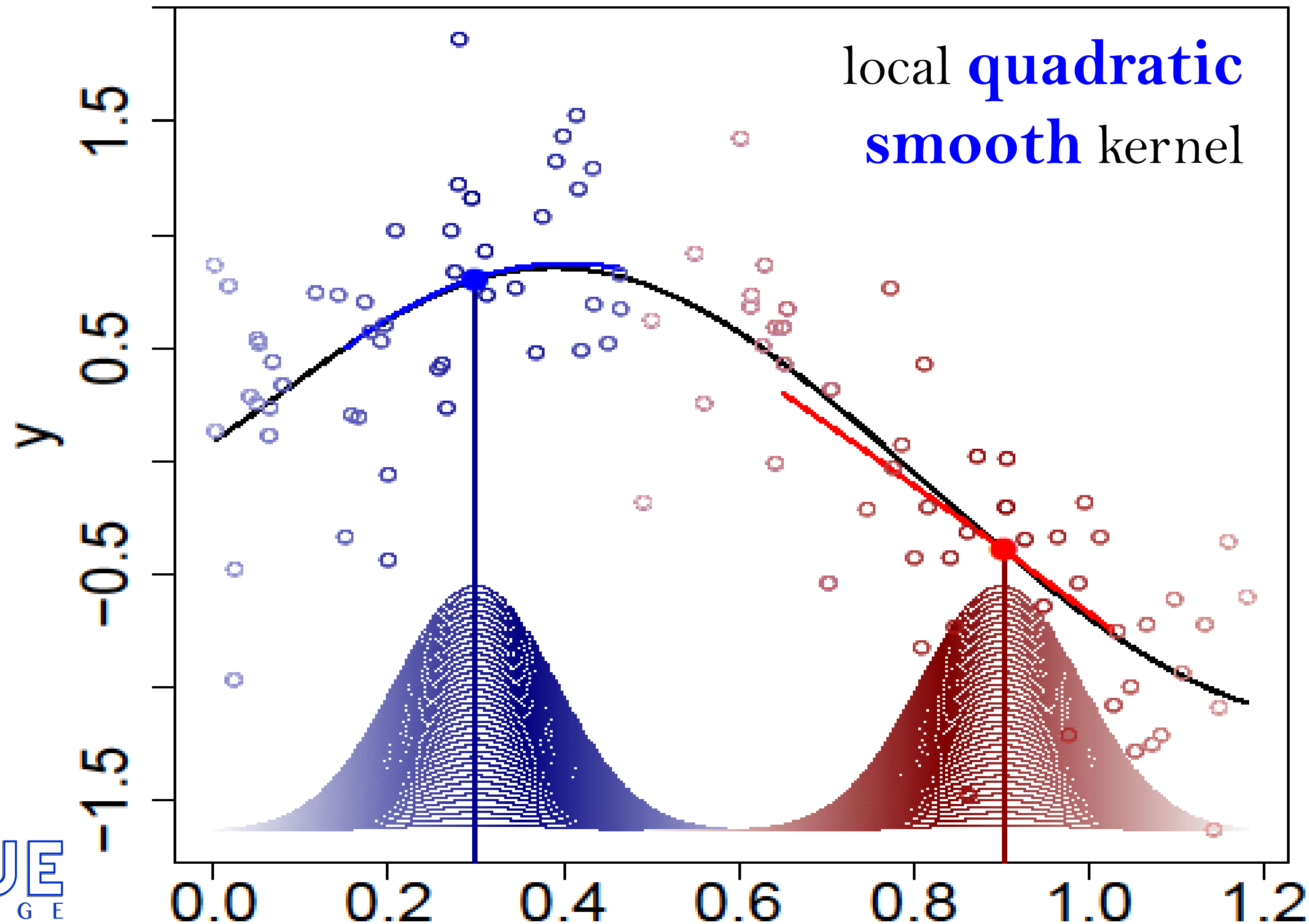
TRUE MILEAGE

Rating Mileage Model

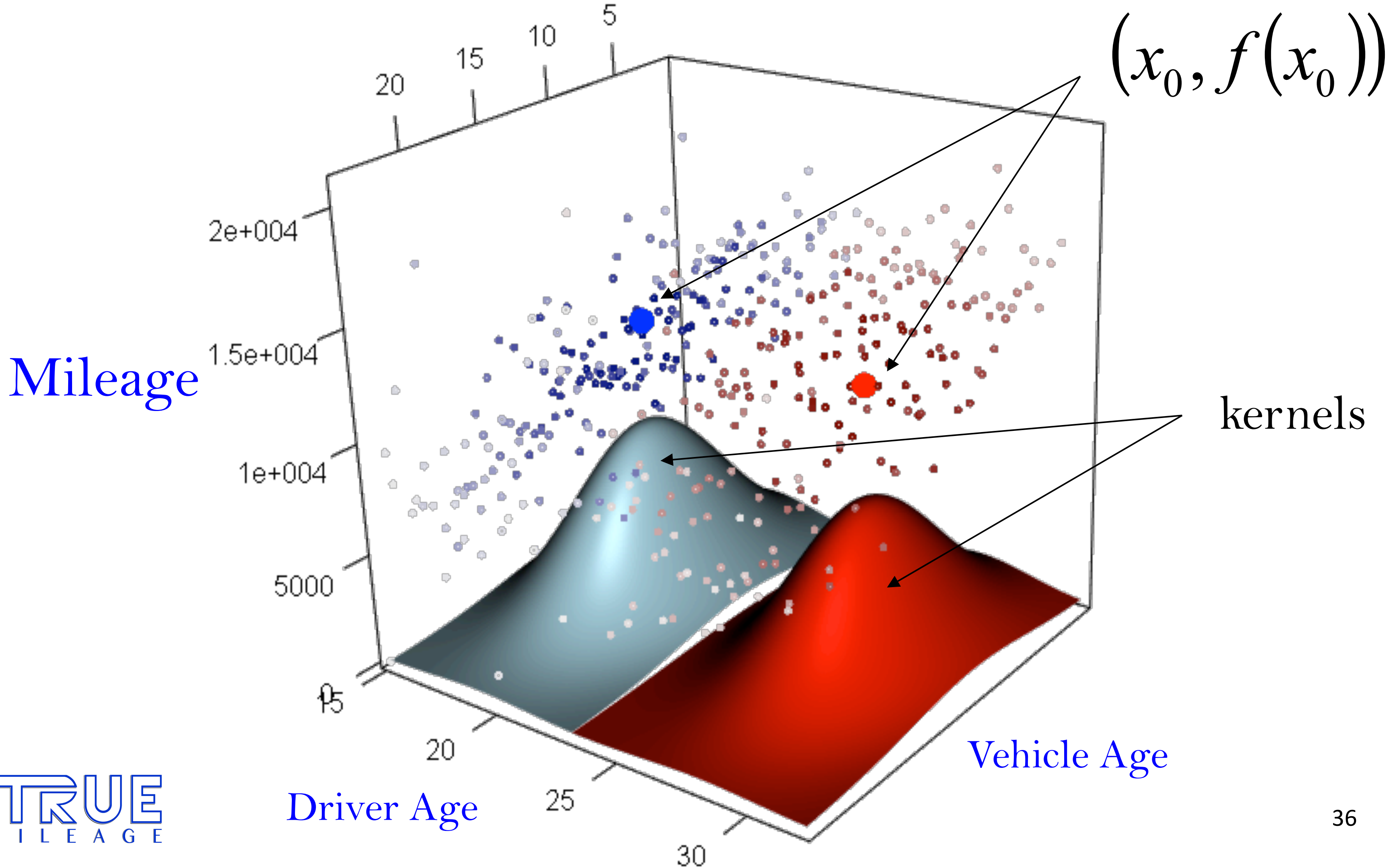


TRUE MILEAGE

Nonparametric Regression

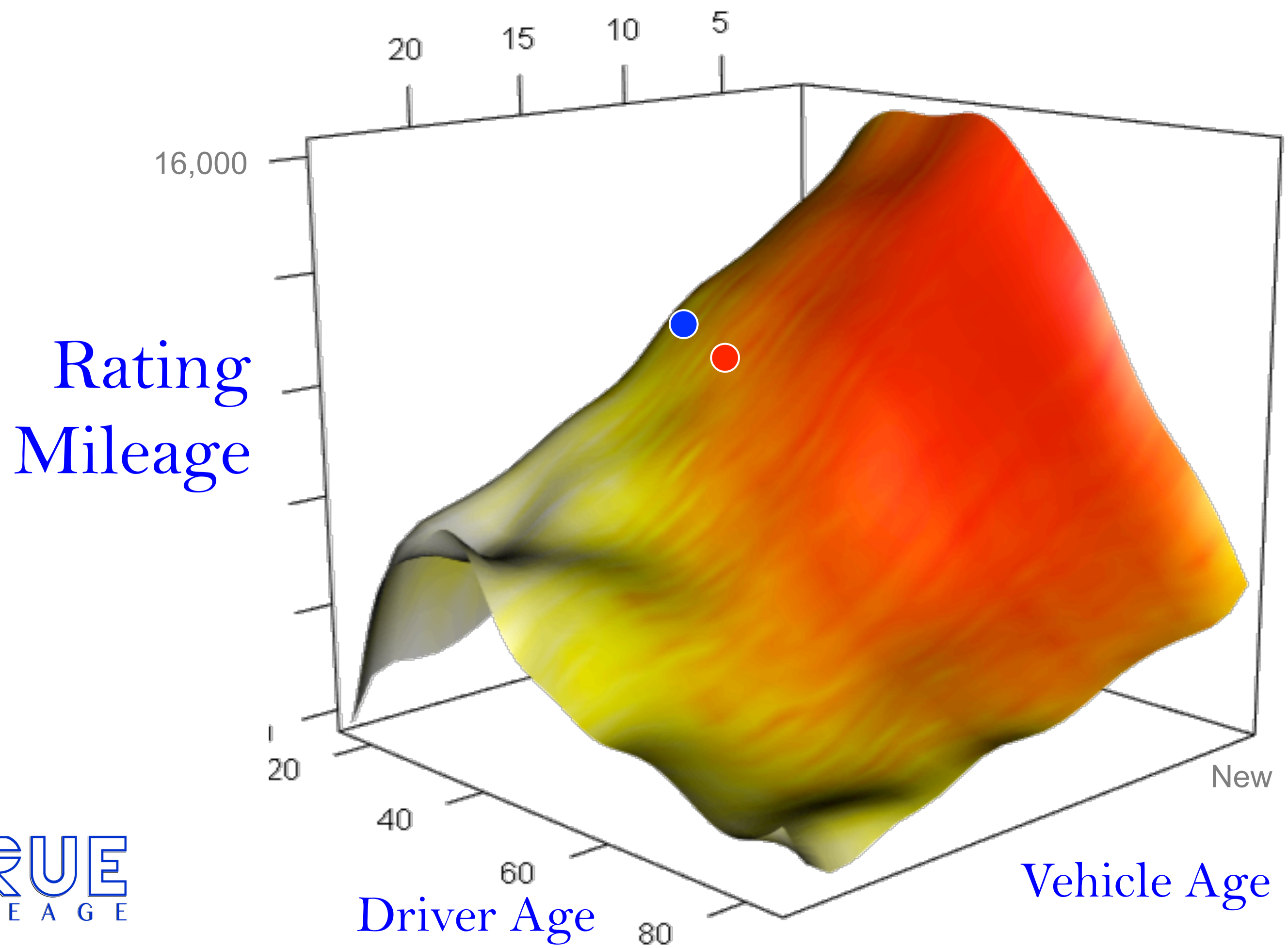


Rating Mileage Model



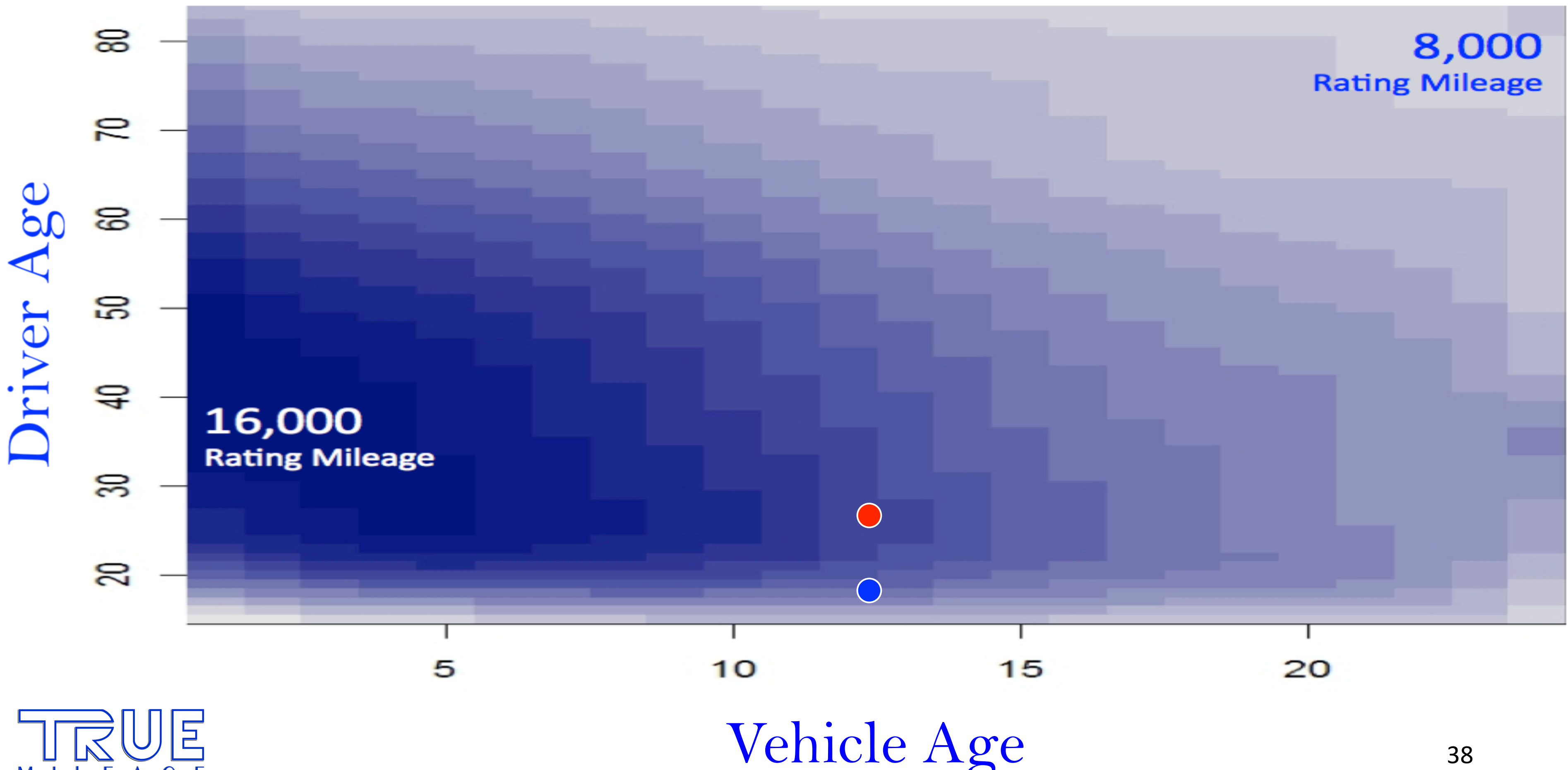
Rating Mileage Model

Results: 3D



Rating Mileage Model

Results: 2D



TRUE MILEAGE

Mileage Discount Analytics™

To eliminate double discounts use:

$$Max\ Discount \cdot \left(1 - \frac{Mileage}{\mathbf{Rating\ Mileage}}\right)$$

Example 1:
(new car and mid-age driver)

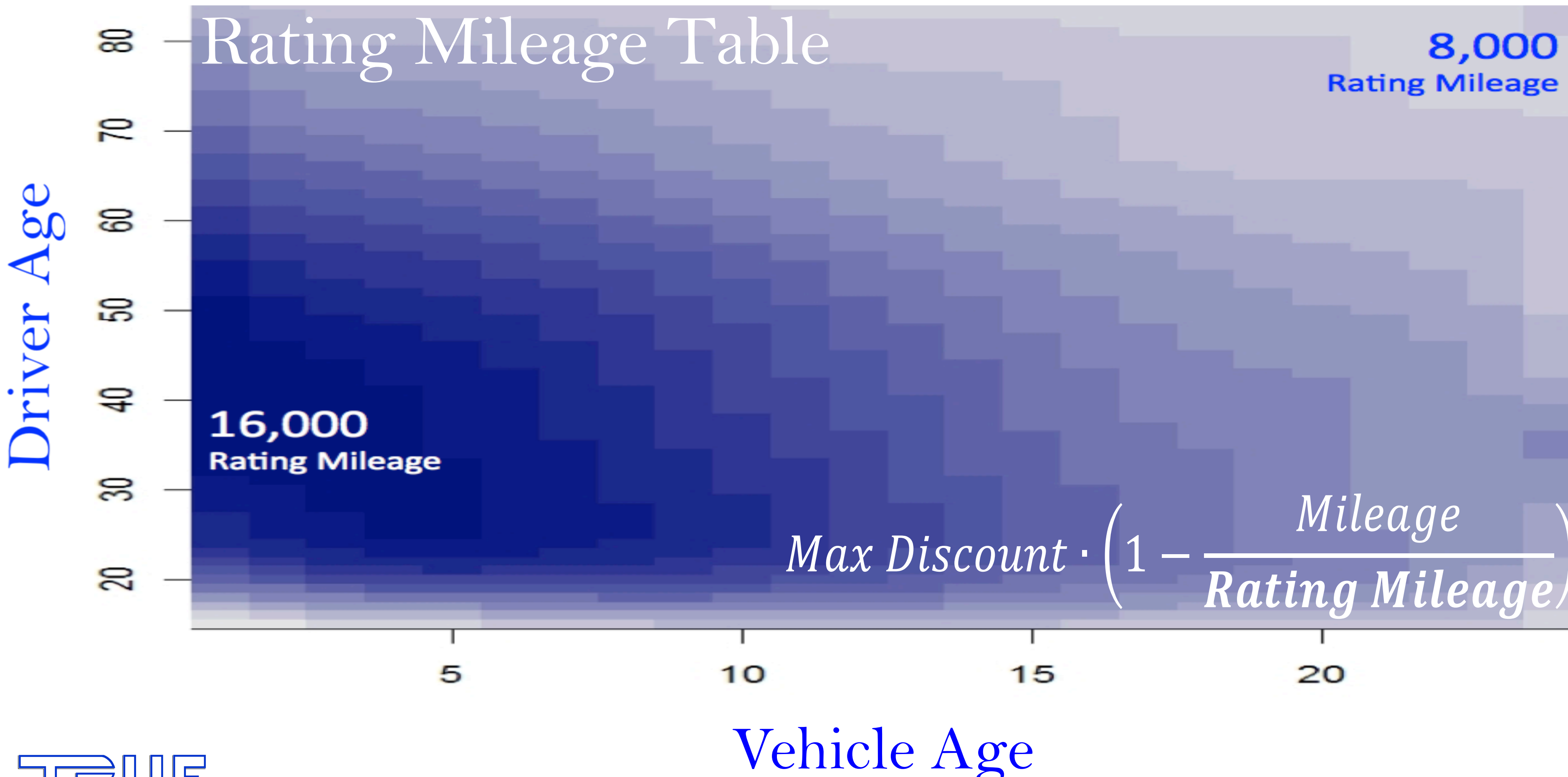
$$50\% \cdot \left(1 - \frac{10{,}000}{\mathbf{16{,}000}}\right) = \mathbf{19\%}$$

Example 2:
(older car or older driver)

$$50\% \cdot \left(1 - \frac{10{,}000}{\mathbf{10{,}000}}\right) = \mathbf{0\%}$$

TRUE MILEAGE

Mileage Discount Analytics™



TRUE MILEAGE

State adjustments to the national rating mileage table also recommended and available.

Agenda

1) About us

2) Intro to Nonparametric Regression

3) Mileage Discount Analytics™

4) Daytime Discount Analytics™

Daytime Discount Analytics™

1) Data: Unbiased national data set with hundreds of thousands of mileage observations and accidents.

2) Variables: Predictive rating variables used are driver age, driver gender, and hour.

3) Goal: Estimate the typical and actual risk for every combination of driver age, gender, and hour.

TRUE MILEAGE

Daytime Discount Analytics™

- The average loss for a general cell is y_i

$$E[y_i] = \exp\{m(x_{1i}, x_{2i})\}$$

- $m(x_{1i}, x_{2i})$ is an unknown function of interest of the predictor variables; driver's age and hour.

- We transform time of day to $\cos(x_2)$ and $\sin(x_2)$.

- Models run separately for males and females.

TRUE MILEAGE

Daytime Discount Analytics™

- We include first order and second order terms,

$$m(x_{10}, x_{20}) = b(x_{10}, \cos(x_{20}), \sin(x_{20}))^T \beta(x_{10}, x_{20})$$

- Local likelihood approach,

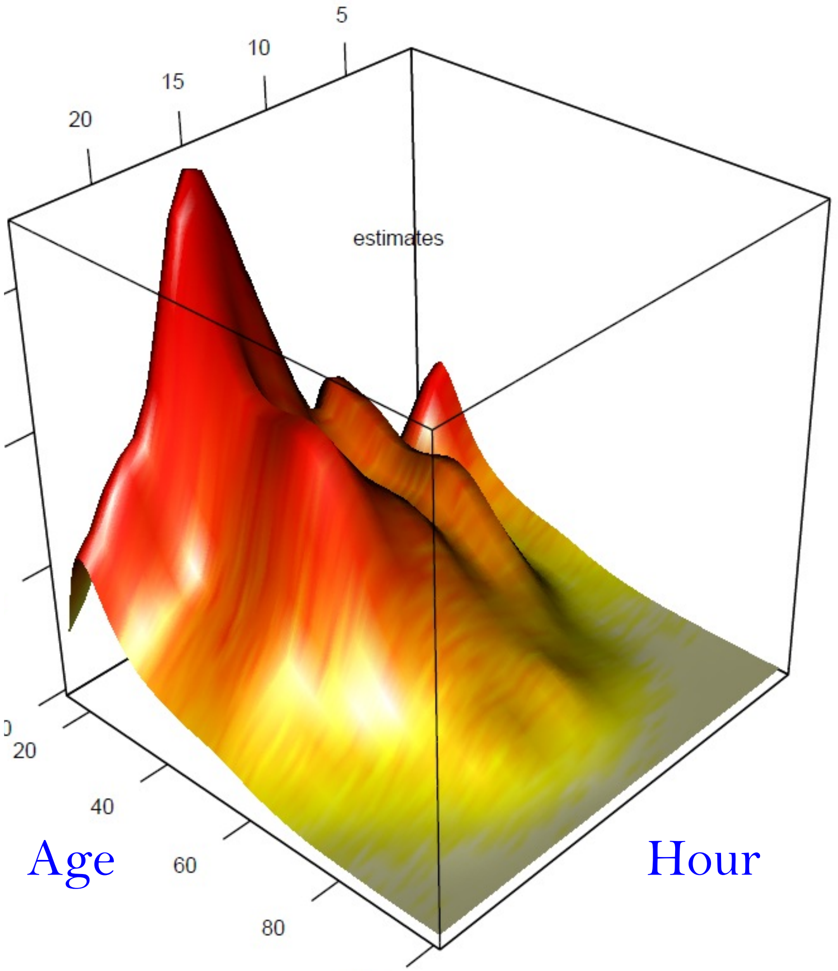
- We multiply kernel by sample size weights n_i to account for different sample sizes

TRUE MILEAGE

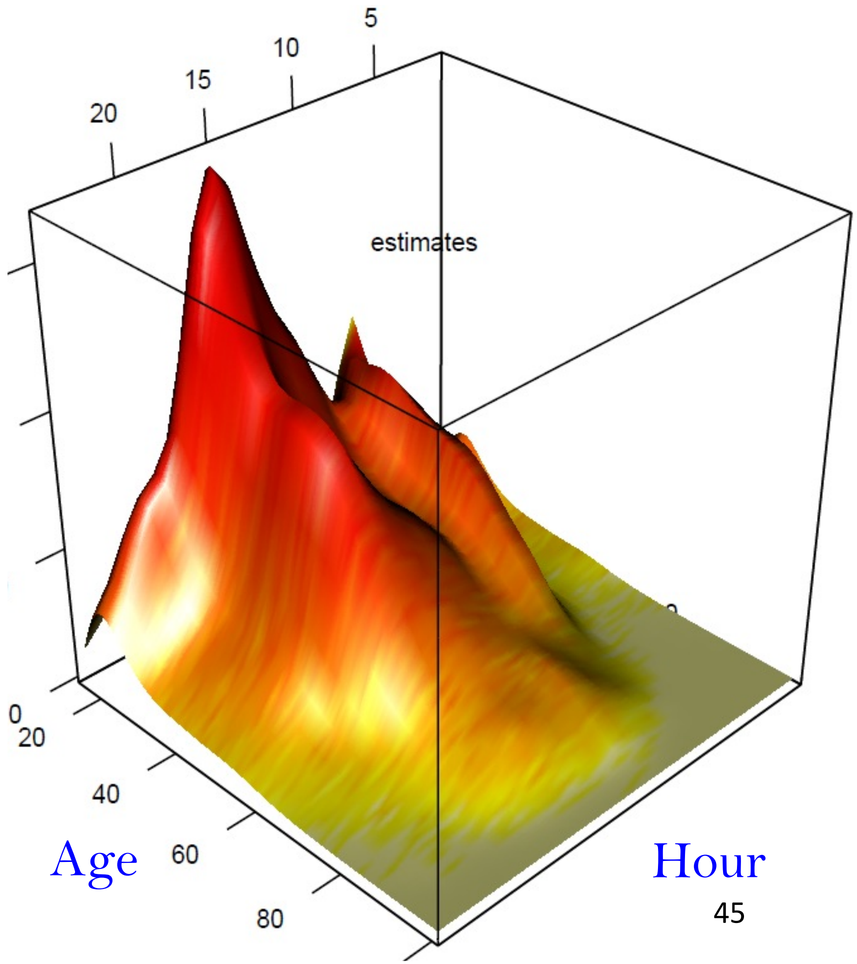
Daytime Discount Analytics ™

Loss Models

Males



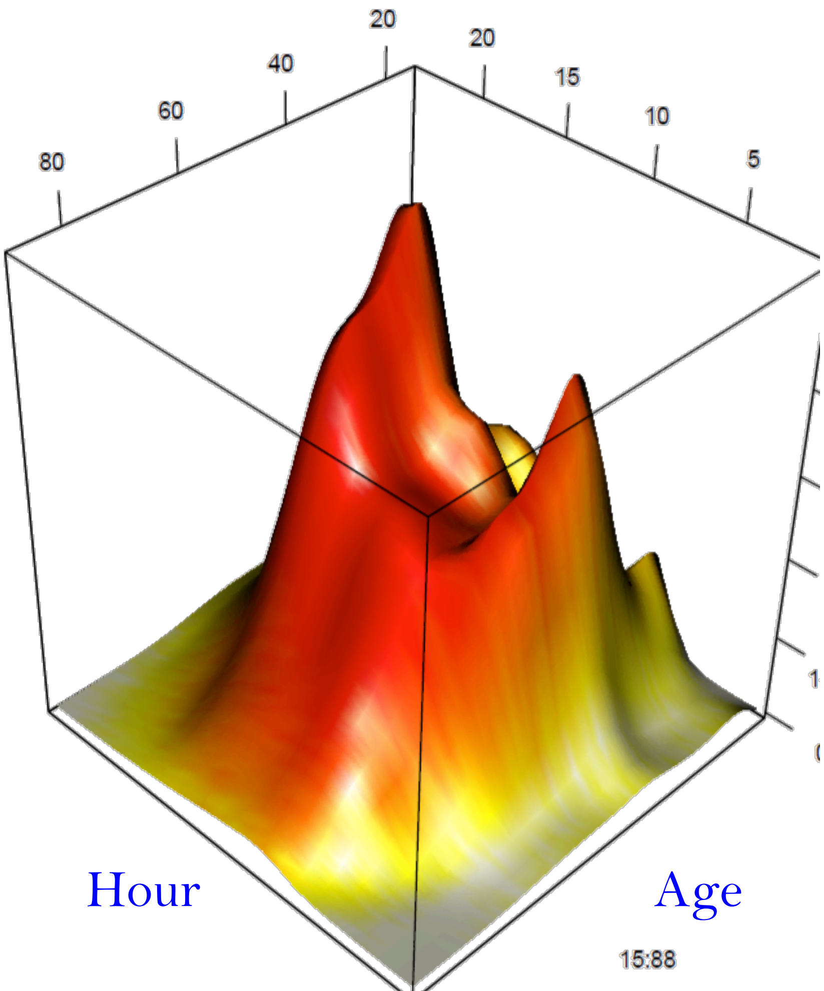
Females



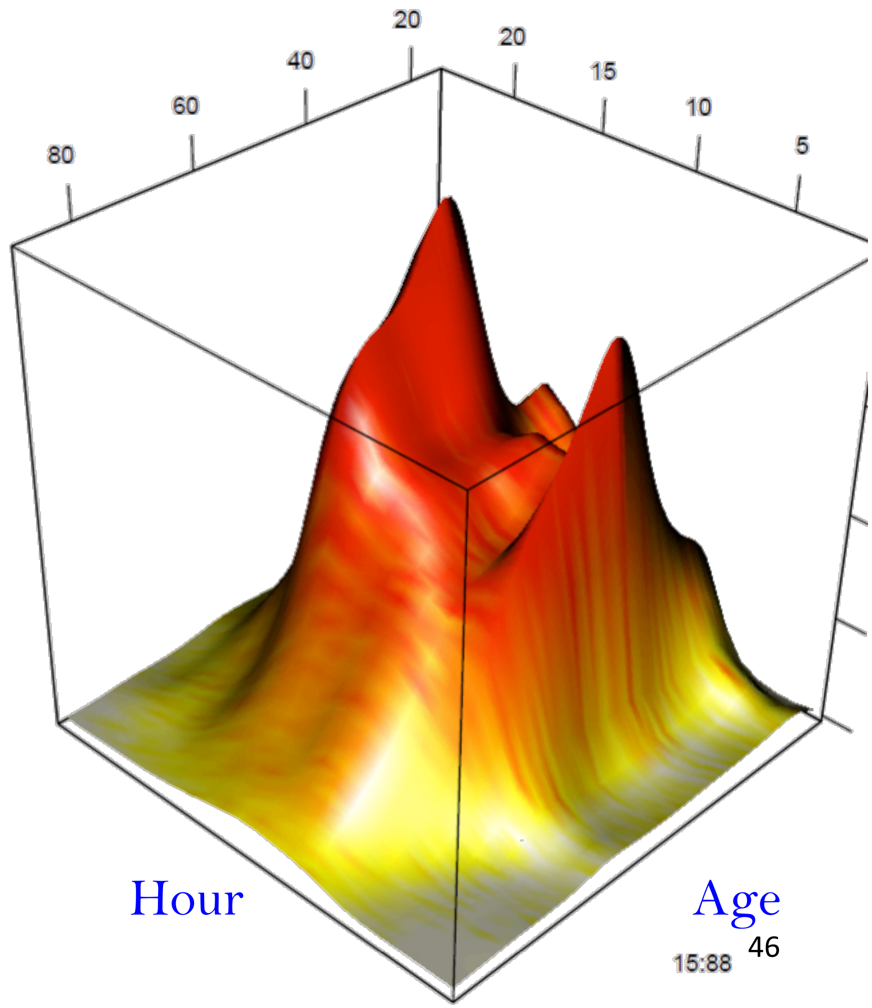
Daytime Discount Analytics™

Distribution Models

Males



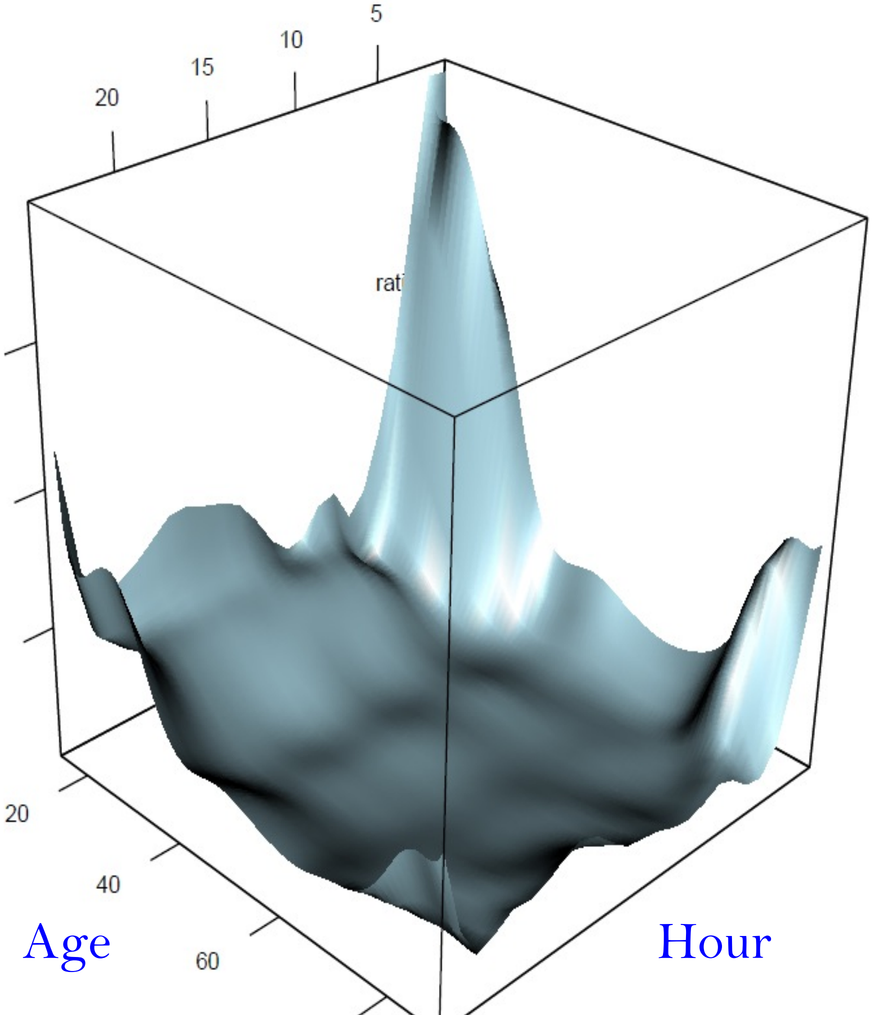
Females



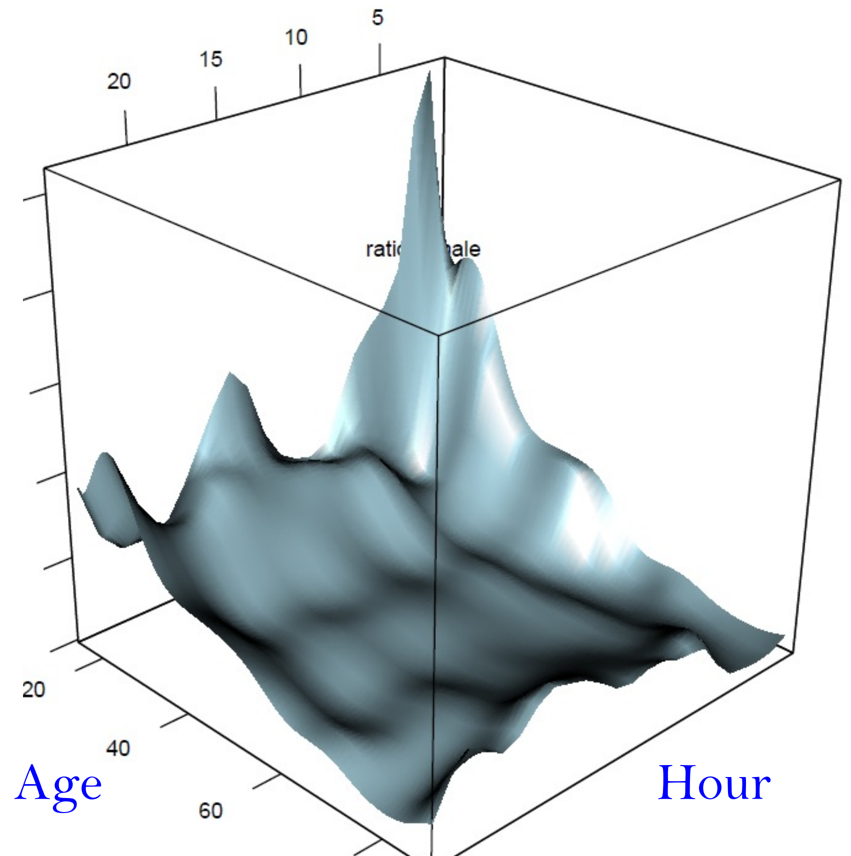
Daytime Discount Analytics™

Risk Models

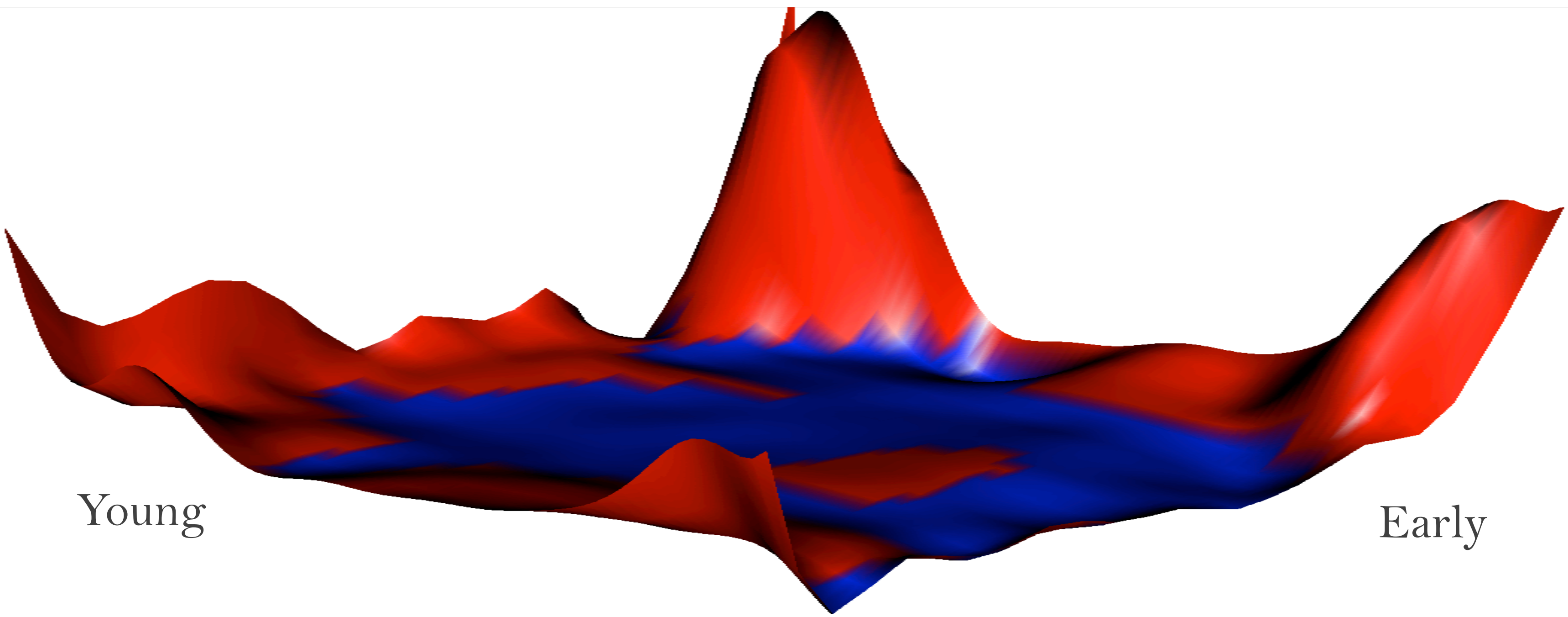
Males



Females



Nonparametric Regression



Are male or female drivers safer?

Red = Females Safer | Blue = Males Safer

TRUE MILEAGE

Thank you!

Ryan N. Morrison
Founder & CEO | True Mileage, Inc.

Daniel Hernandez-Stumpfhauser PhD
Lead Statistician | True Mileage, Inc.

Visit True Mileage at Booth #5