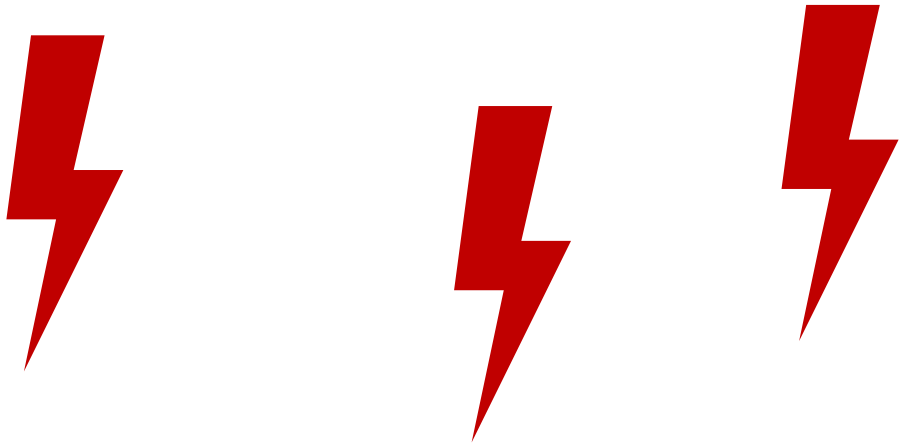
Quantile Regression for Data Science

Gib Bassett

Lightning Talk



UIC

Quantile Regression

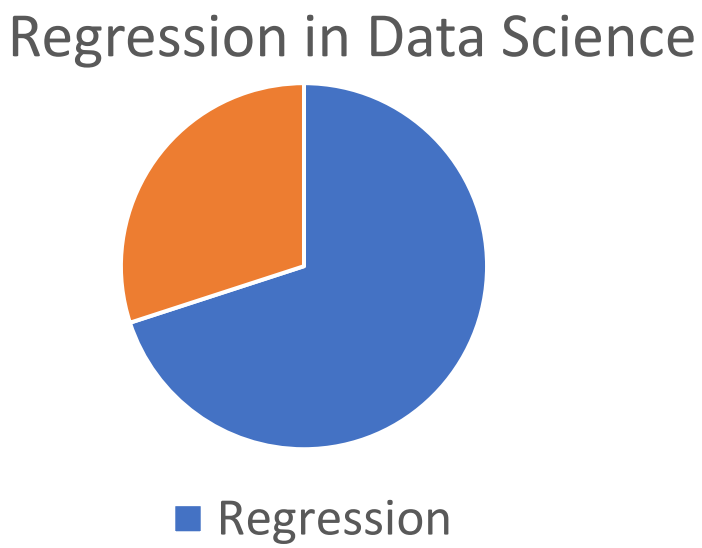
Detects data features otherwise missed

Powerful tool for analyzing data

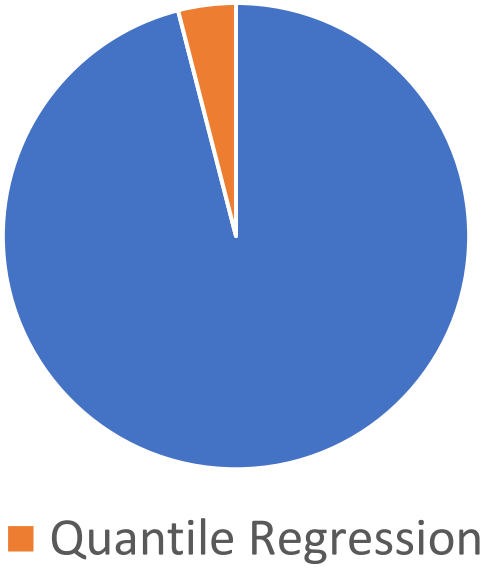
A Better Mousetrap

The Current Situation in Data Science

Standard Regression Analysis in Data Science: COMMON



Quantile Regression: RARE



Reason

Quantile Models:

- *not Understood*
- *not Applicable*
- *not Estimable*

Standard Regression Analysis:
Conditional Expected Values

Standard Regression Analysis:

Conditional Expected Values

Quantile Regression Analysis:

Conditional Quantiles

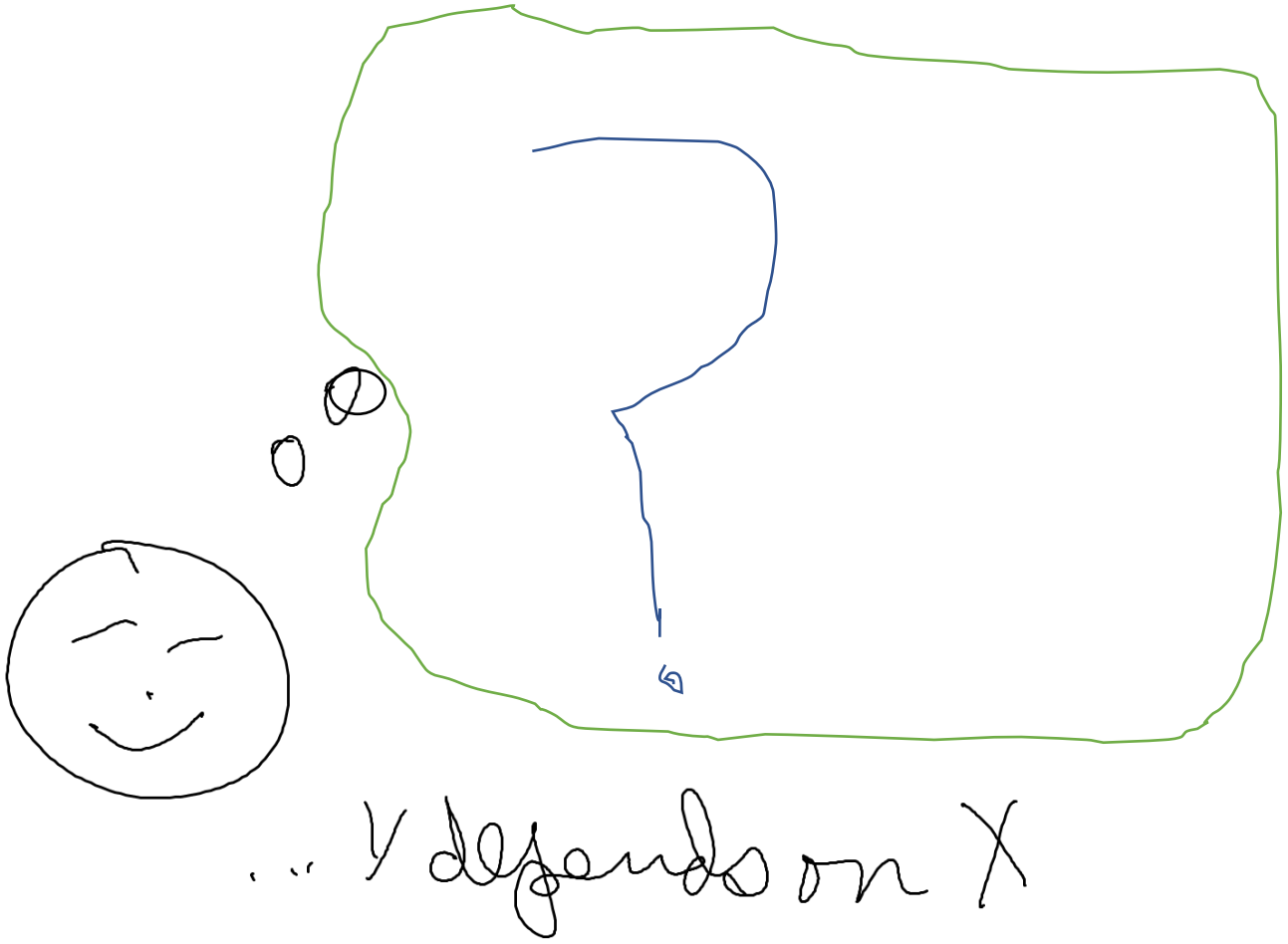
Expected and Quantile Effects can be very different:

It matters!

Regression
“History”

Wondering about Data?

How does y depend on x?



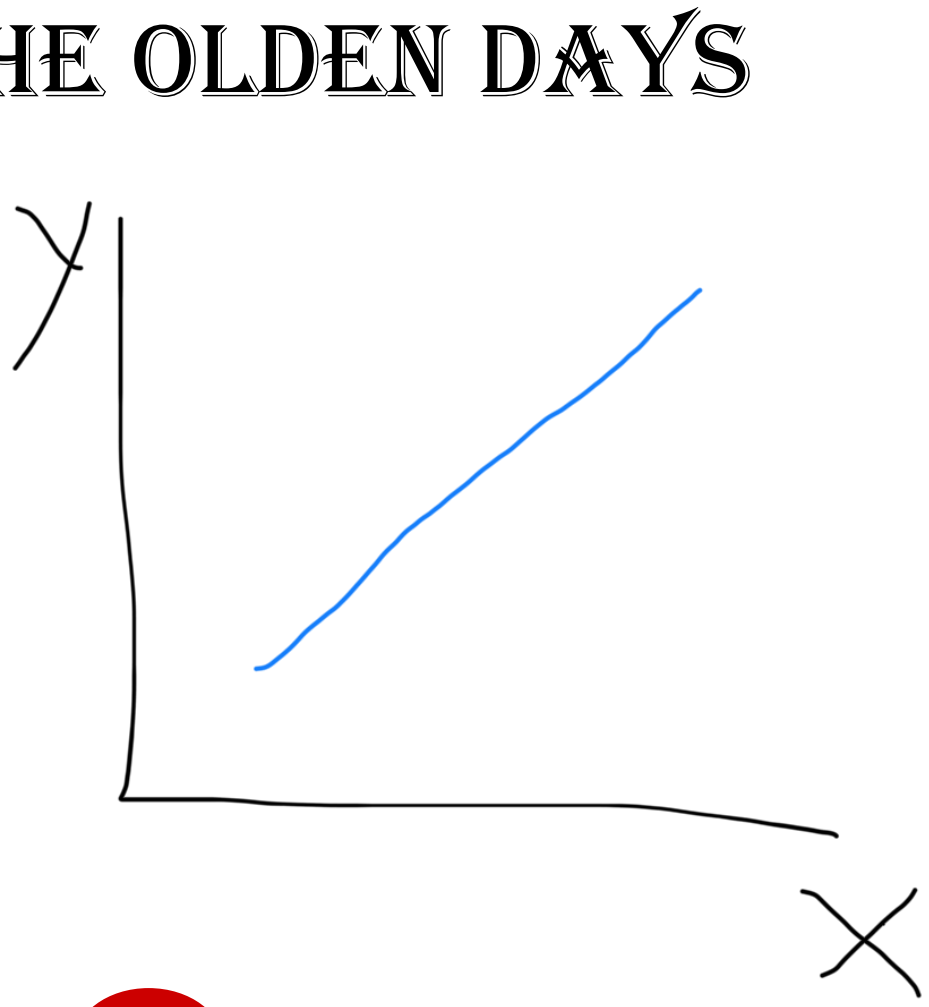
Wondering about Data?

How does y depend on x?

- **How does y=GPA of UIC freshman depends on their x=ACT, c.p.**
- How does y=his wage depend on x=education, c.p.
- How does y=her wage depend x=education, c.p.
- How does y=Google's stock price depend on the x= Market, c.p.?
- How does y=income distribution depend on x=tax rates, c.p.?

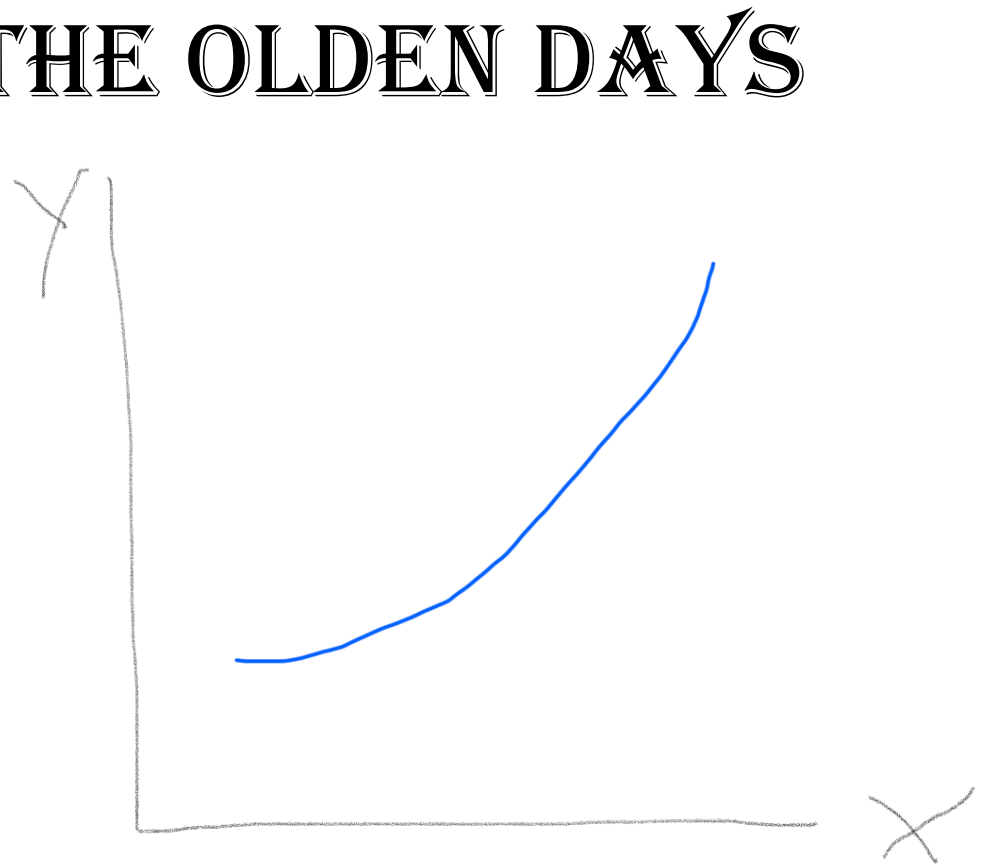
Wondering about Data?

How does y depend on x?



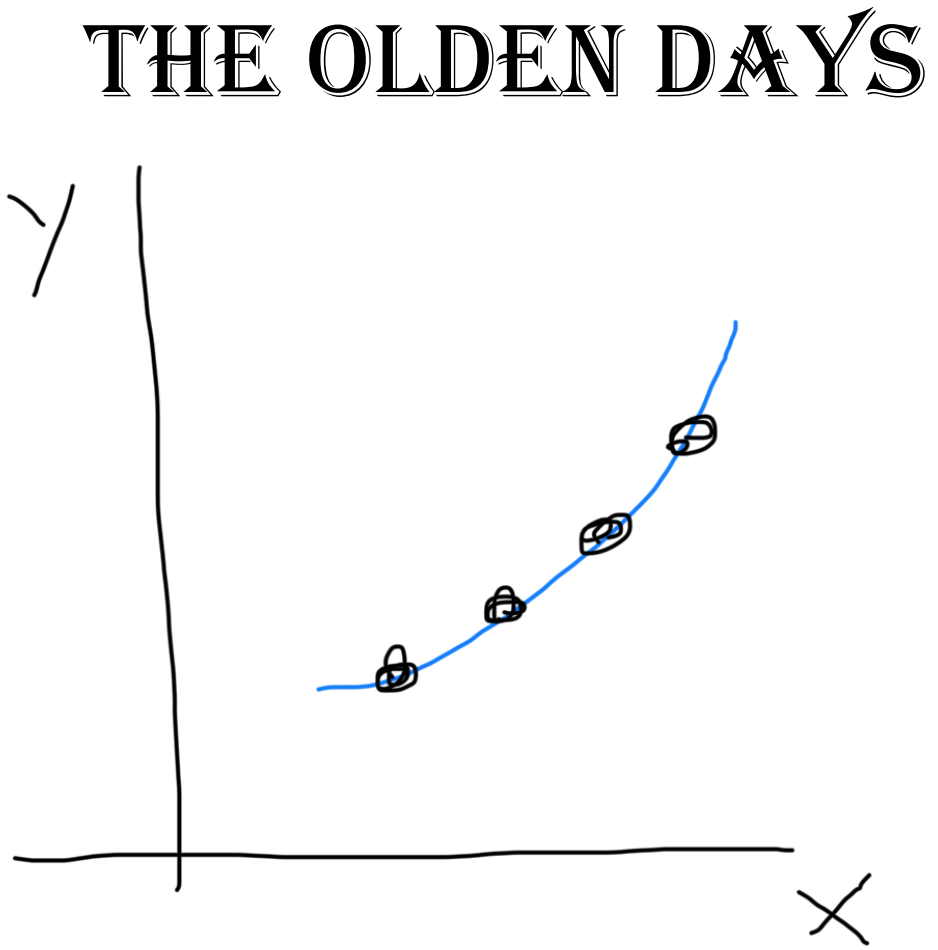
Wondering about Data?

How does y depend on x?



Wondering about Data?

How does y depend on x?



Not *deterministic*

Instead:
EXPECTED VALUE
depends on x.

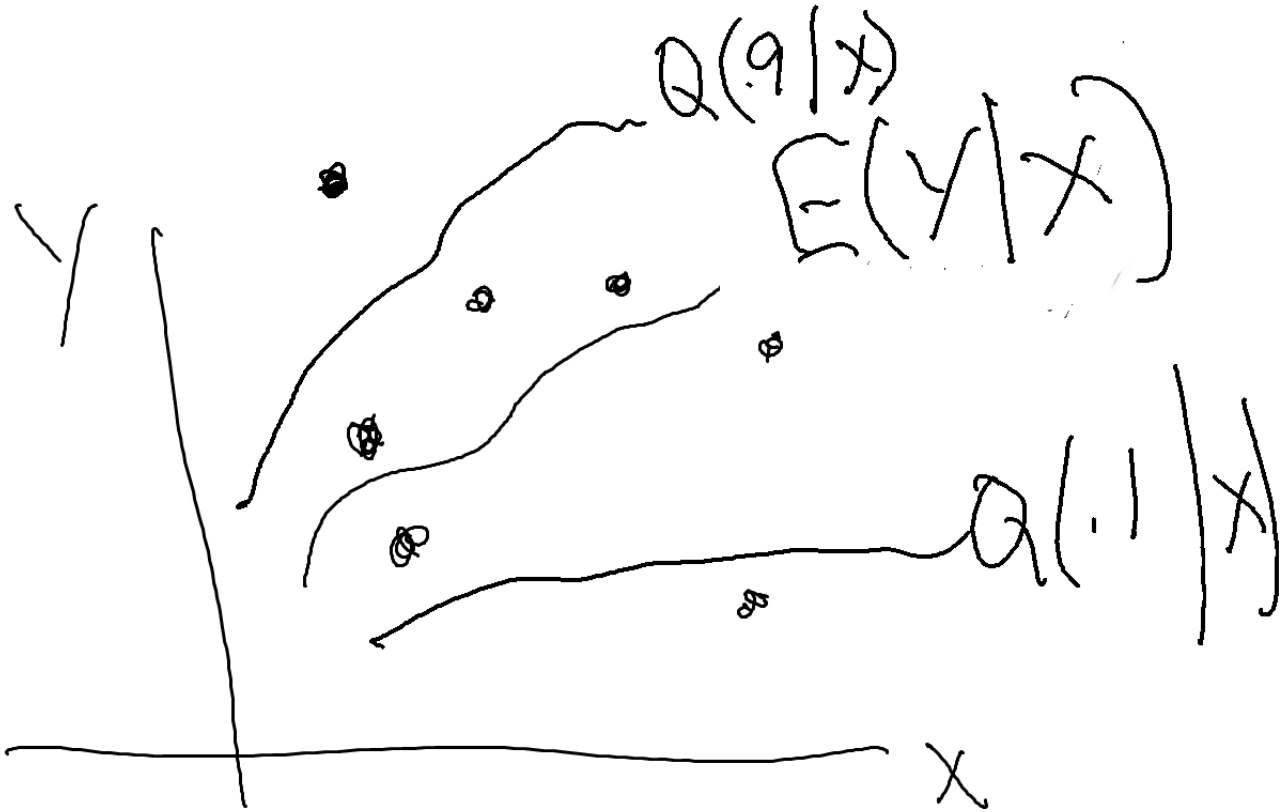
MODERN

DISTRIBUTION depends on x.

Positive mean effect

No quantile(.1) effect

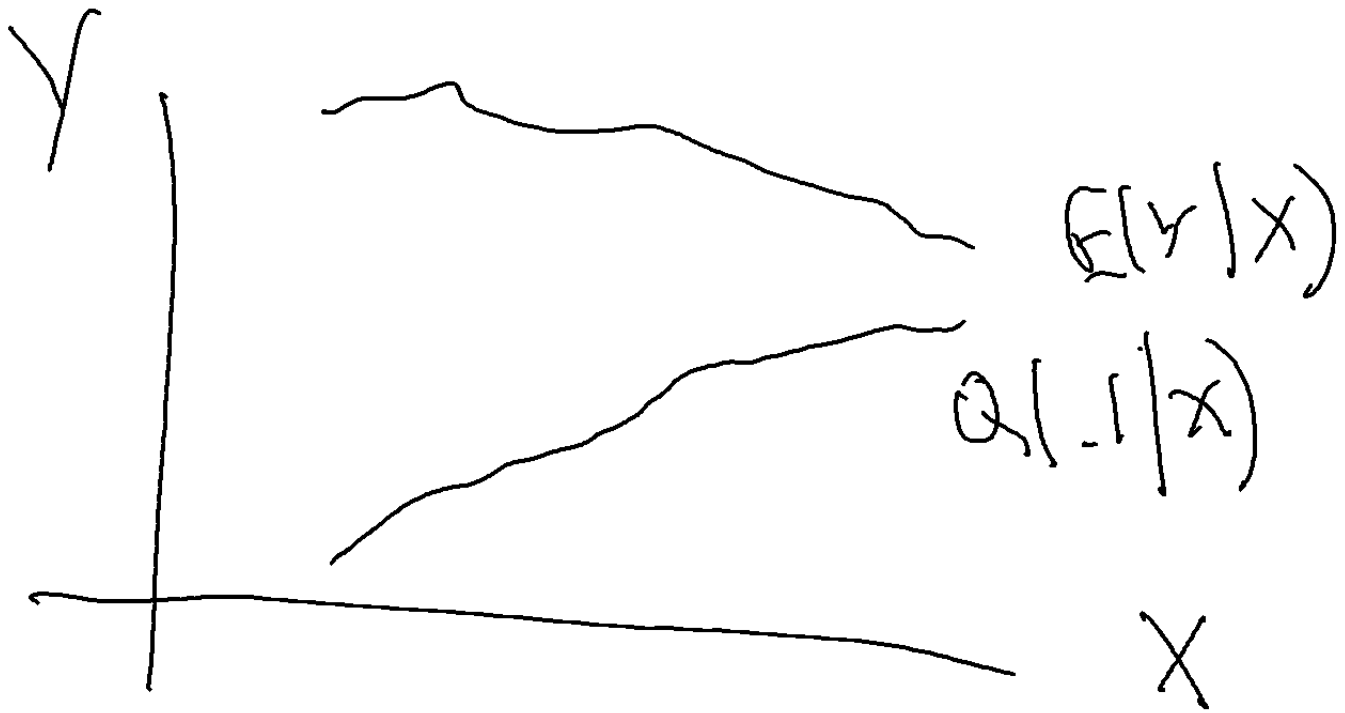
Does ACT matter?



Zero mean effect

Positive quantile(.1) effect

Does ACT matter?



So,
Does y depend on X?

Does y depend on X?

Yes or no:
depending on where in the distribution

What to do?

Quantile Regression Estimation

Reveals how the conditional distribution depends on x

How to Estimate?

---in 30 seconds……

Think about the mean this way:

$$min_b \sum (y_i - b)^2$$

(Not that way: $b = n^{-1} \sum y_i$)

Think about the θ^{th} quantile this way:

$$min_b \sum\{\theta|y_i - b|^+ + (1-\theta)|y_i - b|^-\}$$

(Not that way:

$b = the\ \ \theta^{th}\, smallest\ y_i, i = 1, \ldots, n.$

Software

R

Stata

SAS

Etc.

Quantile Regression for Data Science

Use for any “y depends on x” Model

Detects Data Features otherwise missed

Quantile Regression for Data Science

Thanks!