

# Data Visualization Model Visualization

Bob Stine

Department of Statistics, Wharton

# Question

What concepts do you find hard to teach?

What concepts do they find hard to learn?

# Question

What concepts do you find hard to teach?

Normal quantile plots

Anything to do with hypothesis testing

Categorical variables in regression

Interactions in regression

What concepts do they find hard to learn?

Normal quantile plots

Anything to do with hypothesis testing

Categorical variables in regression

Interactions in regression

# Question

What concepts do you find hard to teach?

- Normal quantile plots
- Anything to do with hypothesis testing
- Categorical variables in regression
- Interactions in regression

What concepts do they find hard to learn?

- Normal quantile plots
- Anything to do with hypothesis testing
- Categorical variables in regression
- Interactions in regression

# Curve Filling

# Normal Quantile Plots

Why bother with these?

Just look at histograms, boxplots

Need to go further

Asking a lot to make ‘judgement’.

Statistics has enough heuristics already.

QQ plots

introduce diagnostic plots

preview hypothesis testing

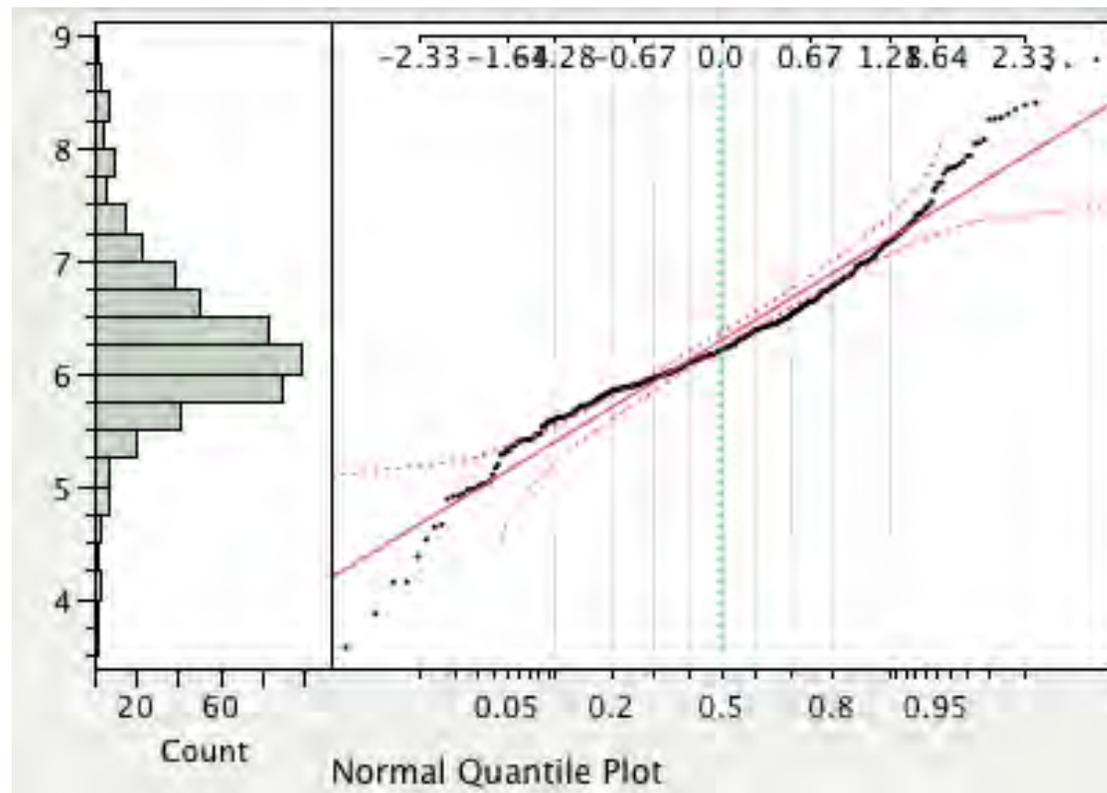
# Hard to Teach

## Conceptual black box

Early in the course, around histograms

Normal score is complex formula

Computer does the work by magic

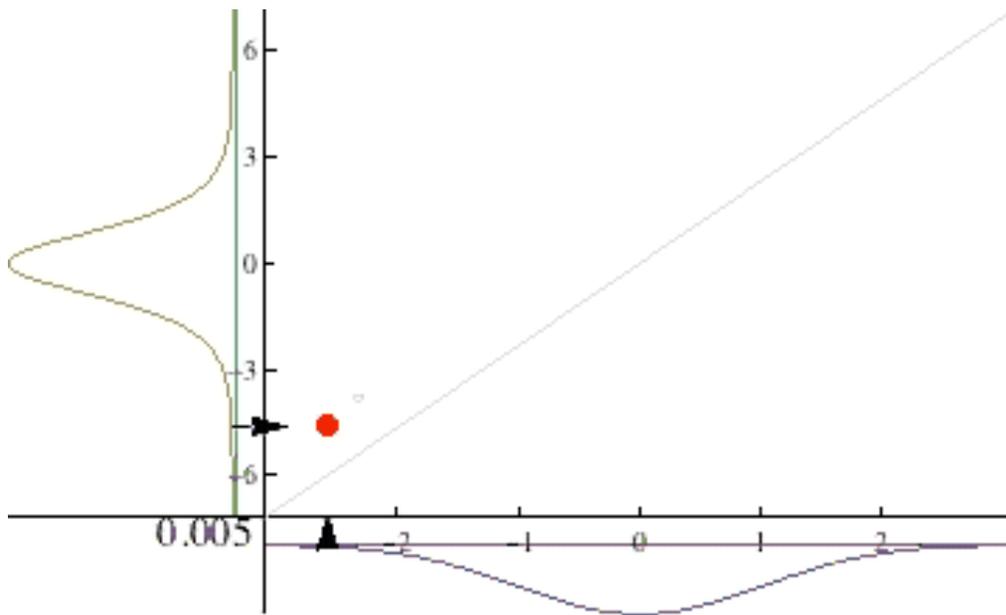




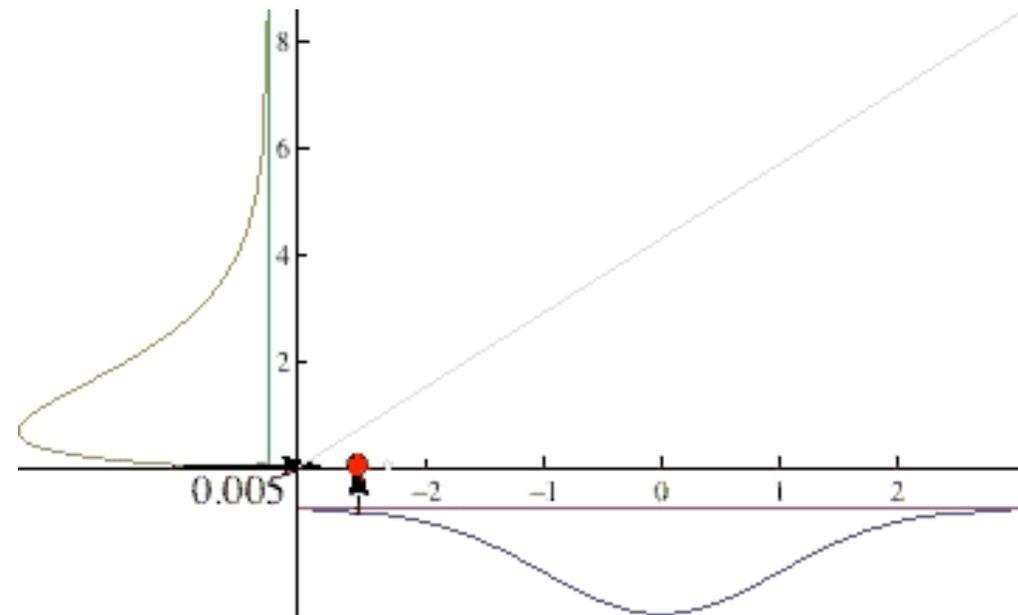
# Non-normal?

Further examples look at distributions with common deviations from normality

## Fat Tails



## Skewness



# Model Profiling

# Regression Models

Simple regression is fun

Look at the picture

Lots of intuitive examples

Interpretation

Multiple regression

Which picture

Collinearity is not so intuitive

Worse when add categorical variables

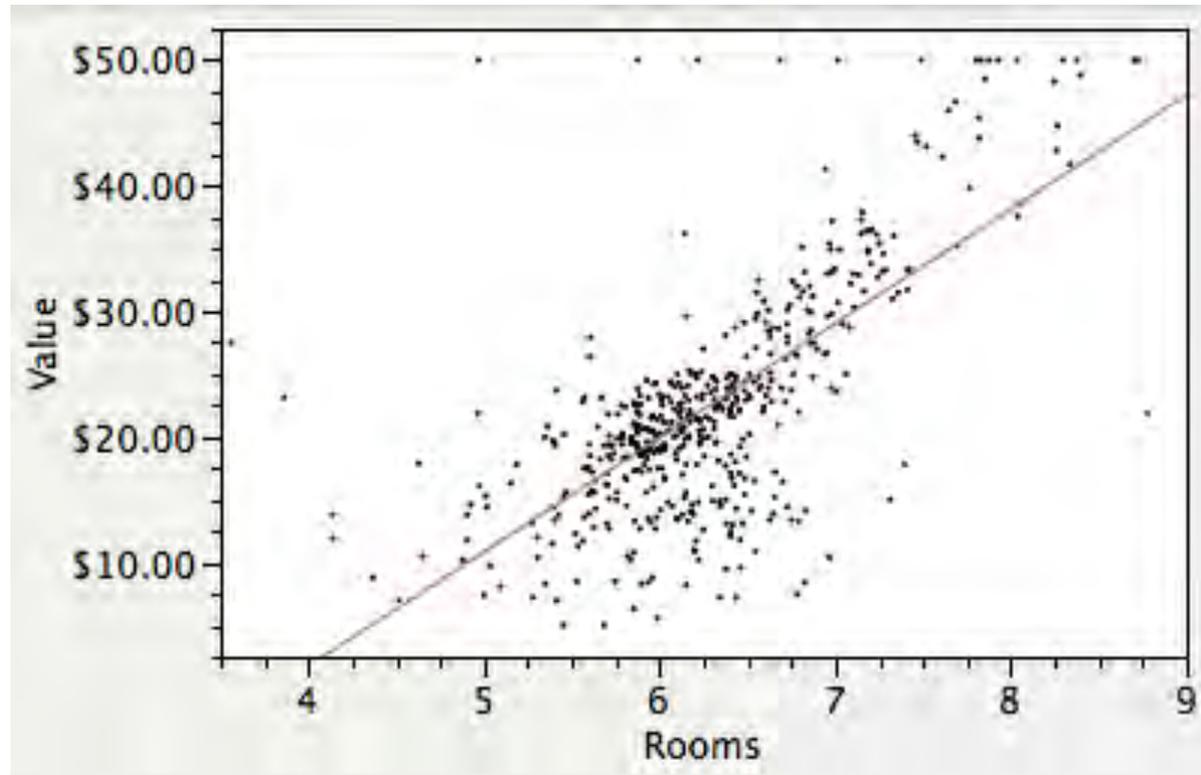
Too many nuanced, subtle tasks in complex situation



# Example

What's this model tell us?

Avg Price  
in \$1000s  
in 1970s



$$\text{Est Value} = -35 + 9 \text{ Avg Num Rooms}$$

# Example

What's this model tell us?



Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	52.439	13.090	4.01	<.0001*
Rooms	5.644	0.358	15.78	<.0001*
(Rooms-6.28463)*(Rooms-6.28463)	2.526	0.237	10.65	<.0001*
Charles River[Away]	-5.697	0.984	-5.79	<.0001*
NOx	-70.826	10.682	-6.63	<.0001*
Tax Rate	0.064	0.011	5.76	<.0001*
Pupil/Teacher	-2.600	0.6		
(NOx-0.5547)*Charles River[Away]	55.986	10.9		
Charles River[Away]	-0.073	0.0		
Charles River[Away]	1.751	0.6		



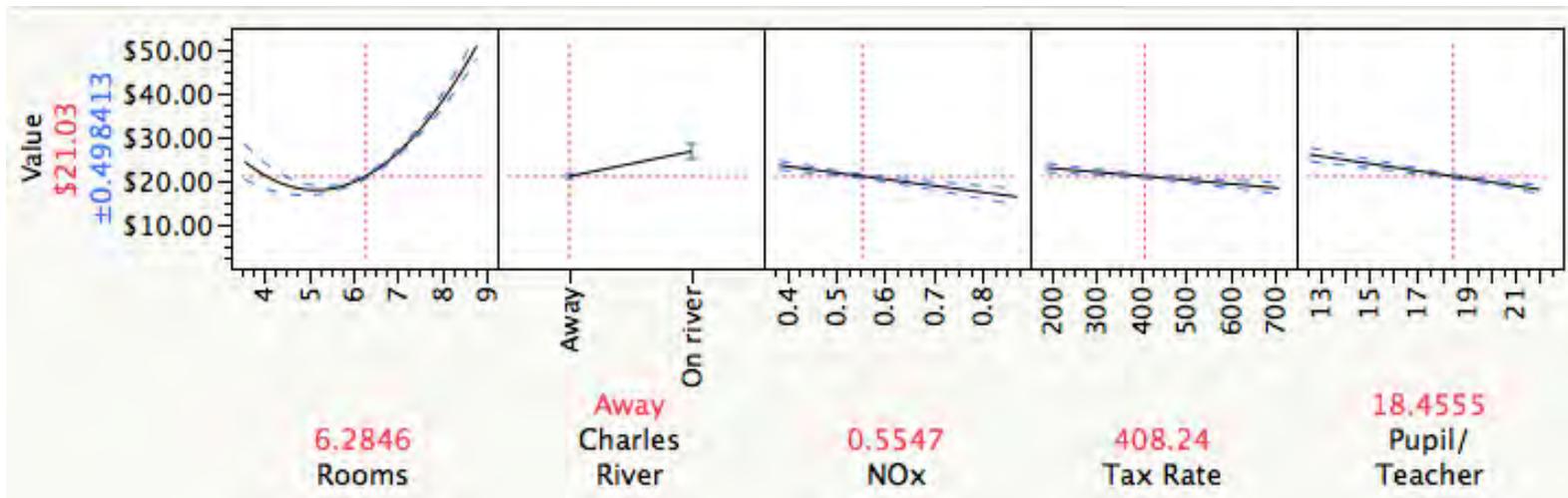
Need a better presentation...

# Profile of Model

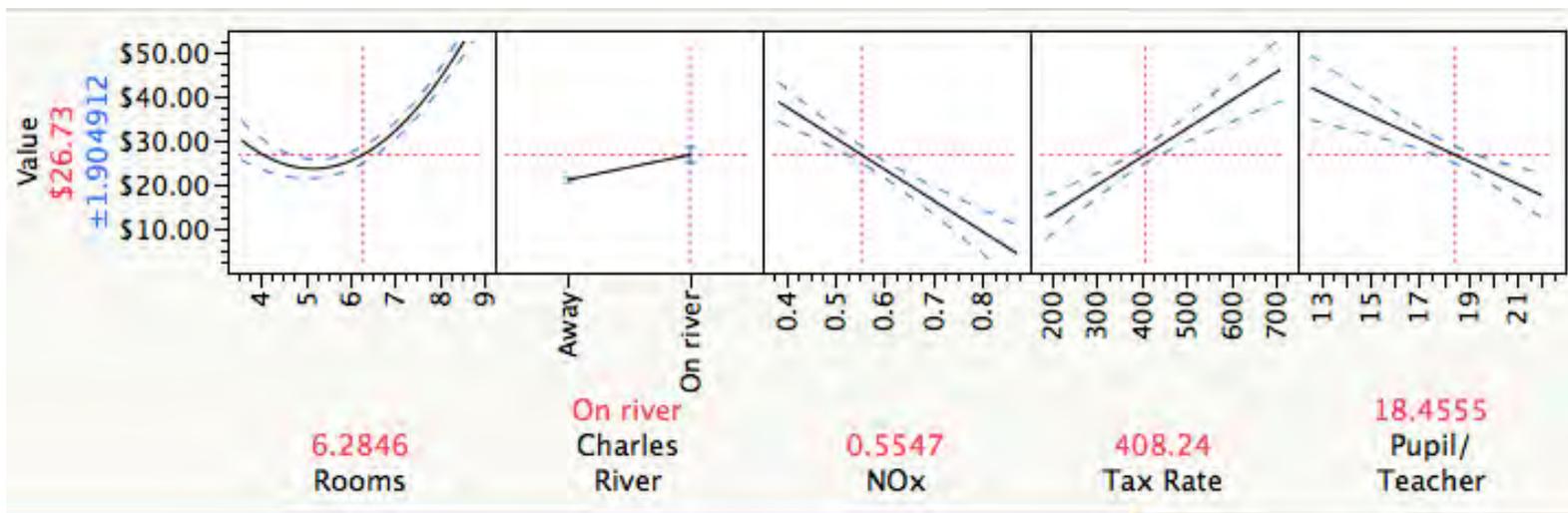
Partial effect conditional on others...

Interactive tool

Away from



On river



# Closing Remarks

Static presentation is less compelling

At what point does the course become too oriented toward using software?

Thanks!