

Intro Stats in the 21st Century

Dick De Veaux
Williams College
USCOTS 15
May 28, 2015

WHAT KEEPS ME UP AT NIGHT



- Data Scientists teaching our course
- Students think that Statistics is irrelevant for their lives and work
- Students think that Statistics is essentially univariate
- We continue to change the course around the edges
 - With Resampling the exception

5/28/15

USCOTS 15

THE PROBLEM

- We teach the wrong stuff
- We teach it the wrong way
- We teach it in the wrong order
- I don't have any answers – but I'll keep asking the questions

5/28/15

WHAT DO I WANT THEM TO TAKE AWAY?

- The Idea that Statistics is
 - Relevant
 - Intuitive
 - Cool
 - A Powerful Method for Solving Problems in the World
- Armed with a Healthy (and Informed) Skepticism for
 - Data Quality
 - Models
 - Inference and Conclusions

5/28/15

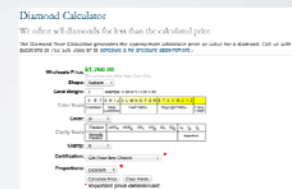
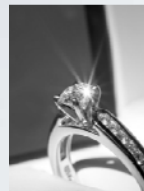
HOW TO GET TO THE COOL STUFF?

- Start with it
- Introduce models early
 - Motivate univariate questions from more complex models
 - Socratic method
- What to leave out
 - Mathematics of Sampling distributions
 - We just need the concept
 - Getting too under the hood with methods
 - They don't need to "produce" everything – especially in the first course

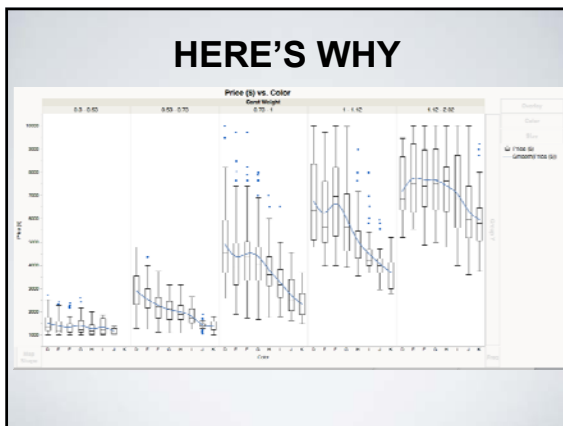
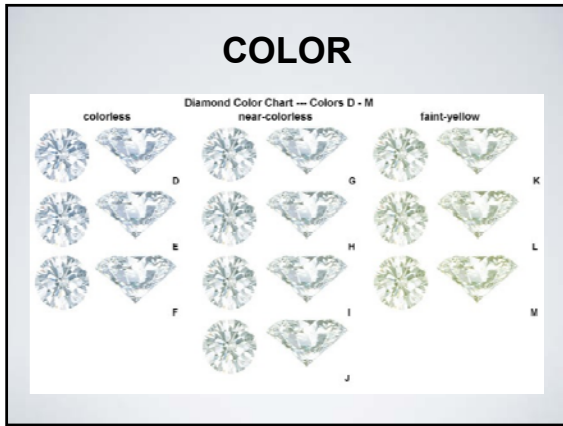
5/28/15

DIAMONDS

- Four C's:
 - Carat, Color, Cut, Clarity



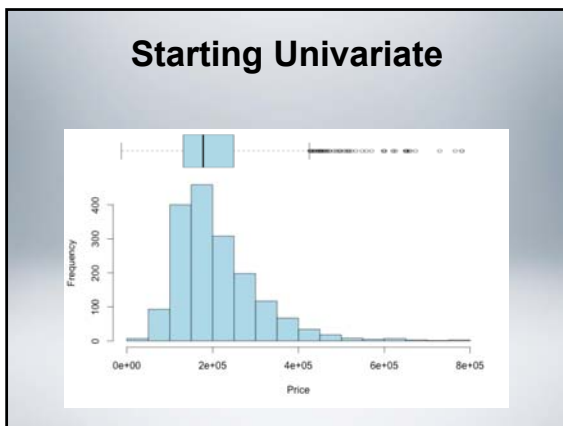
5/28/15

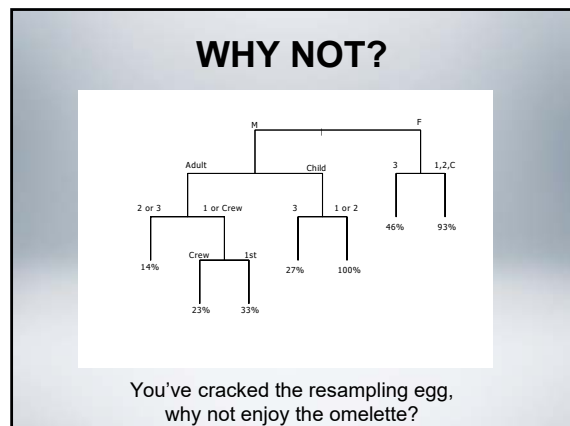
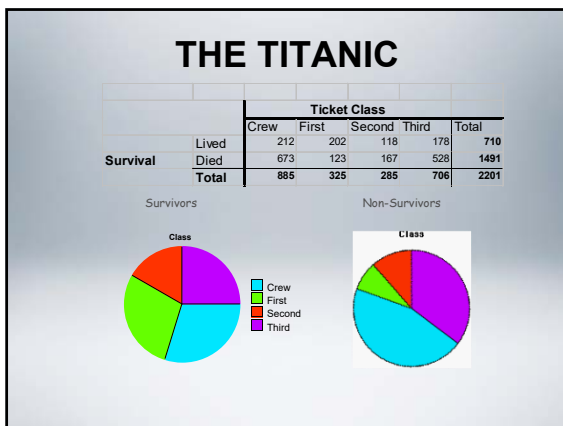
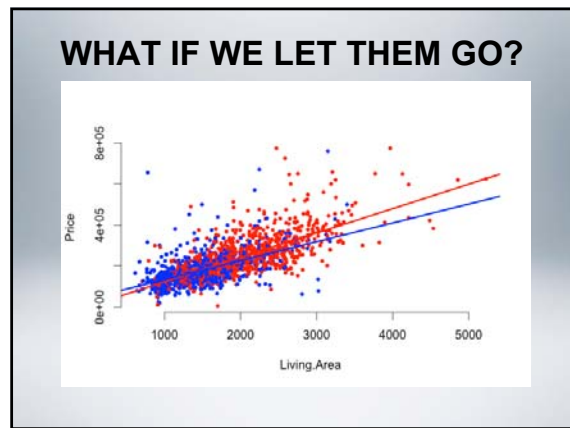
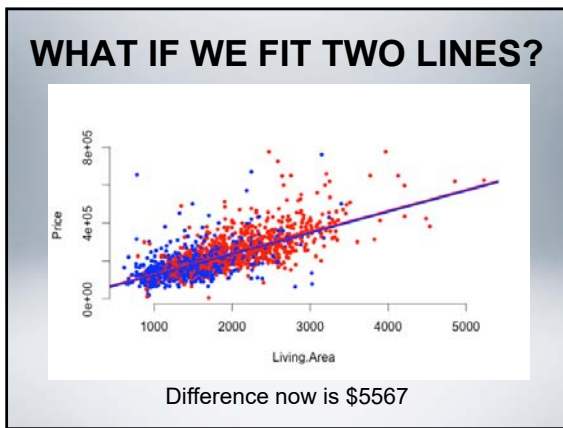
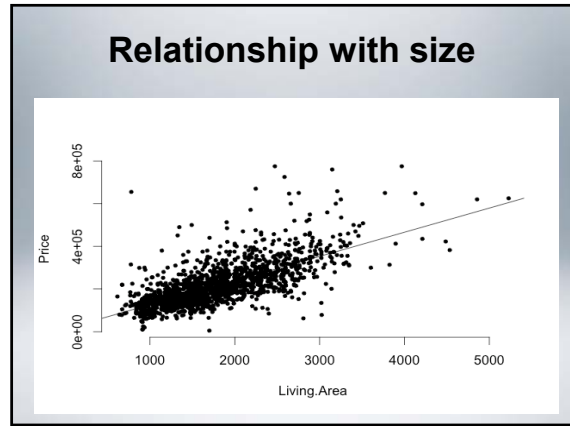
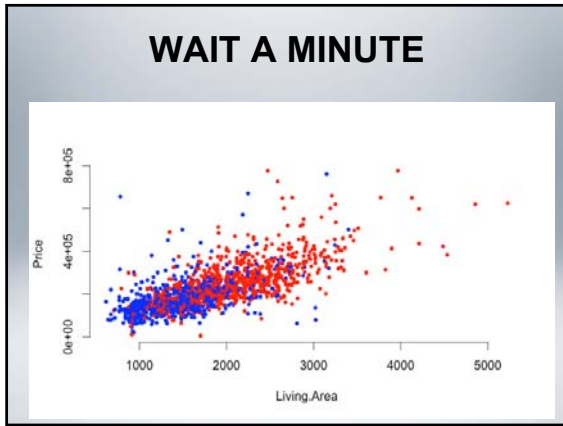


HOW MUCH IS A FIREPLACE WORTH?

The Data

- 1729 randomly selected homes from the county records of Saratoga New York about 2005.
- Conclusions are only about the Northeast...





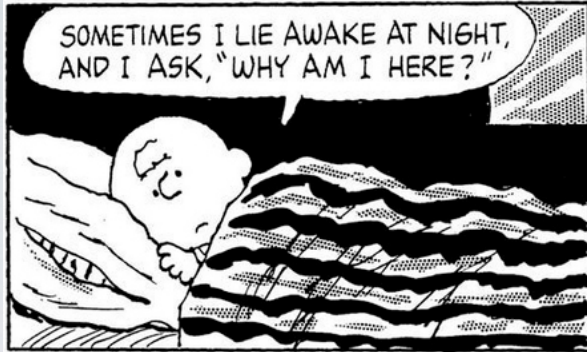
WHERE ARE WE?

- Introduce models early
 - Motivate univariate questions from more complex models
 - Not the other way around!
- Use complex questions and models to drive topics
- Statistics is more than a collection of tools
 - Which ones are really important?
- How much should we ask them to “produce”?

Intro Stats in the 21st Century

Dick De Veaux
Williams College
USCOTS 15
May 28, 2015

WHAT KEEPS ME UP AT NIGHT



- Data Scientists teaching our course
- Students think that Statistics is irrelevant for their lives and work
- Students think that Statistics is essentially univariate
- We continue to change the course around the edges
 - With Resampling the exception

THE PROBLEM

- We teach the wrong stuff
- We teach it the wrong way
- We teach it in the wrong order
- I don't have any answers – but I'll keep asking the questions

WHAT DO I WANT THEM TO TAKE AWAY?

- The Idea that Statistics is
 - Relevant
 - Intuitive
 - Cool
 - A Powerful Method for Solving Problems in the World
- Armed with a Healthy (and Informed) Skepticism for
 - Data Quality
 - Models
 - Inference and Conclusions

HOW TO GET TO THE COOL STUFF?

- Start with it
- Introduce models early
 - Motivate univariate questions from more complex models
 - Socratic method
- What to leave out
 - Mathematics of Sampling distributions
 - We just need the concept
 - Getting too under the hood with methods
 - They don't need to “produce” everything – especially in the first course

DIAMONDS

- Four C's:
 - Carat, Color, Cut, Clarity



Diamond Calculator

We often sell diamonds for less than the calculated price.

The Diamond Price Calculator generates the approximate wholesale price or value for a diamond. Call us with questions at 703-536-3600 or to [schedule a no-pressure appointment](#).

Wholesale Price: **\$1,260.00**

Our prices are often less than this.

Shape:

Carat Weight: example: 0.98 or 1.73 or 3.95

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Colorless	Near Colorless	Faint Yellow	Very Light Yellow		Fancy Yellow																	

Color:

Flawless	VS ₁	VS ₂	VS ₁	VS ₂	S ₁	S ₂	I ₁	I ₂	I ₃	
Internally Flawless								Imperfect		

Clarity:

Certification: *

Proportions: *

* Important price determinant

COLOR

Diamond Color Chart --- Colors D - M

colorless

near-colorless

faint-yellow



D



G



K



E



H



L



F



I

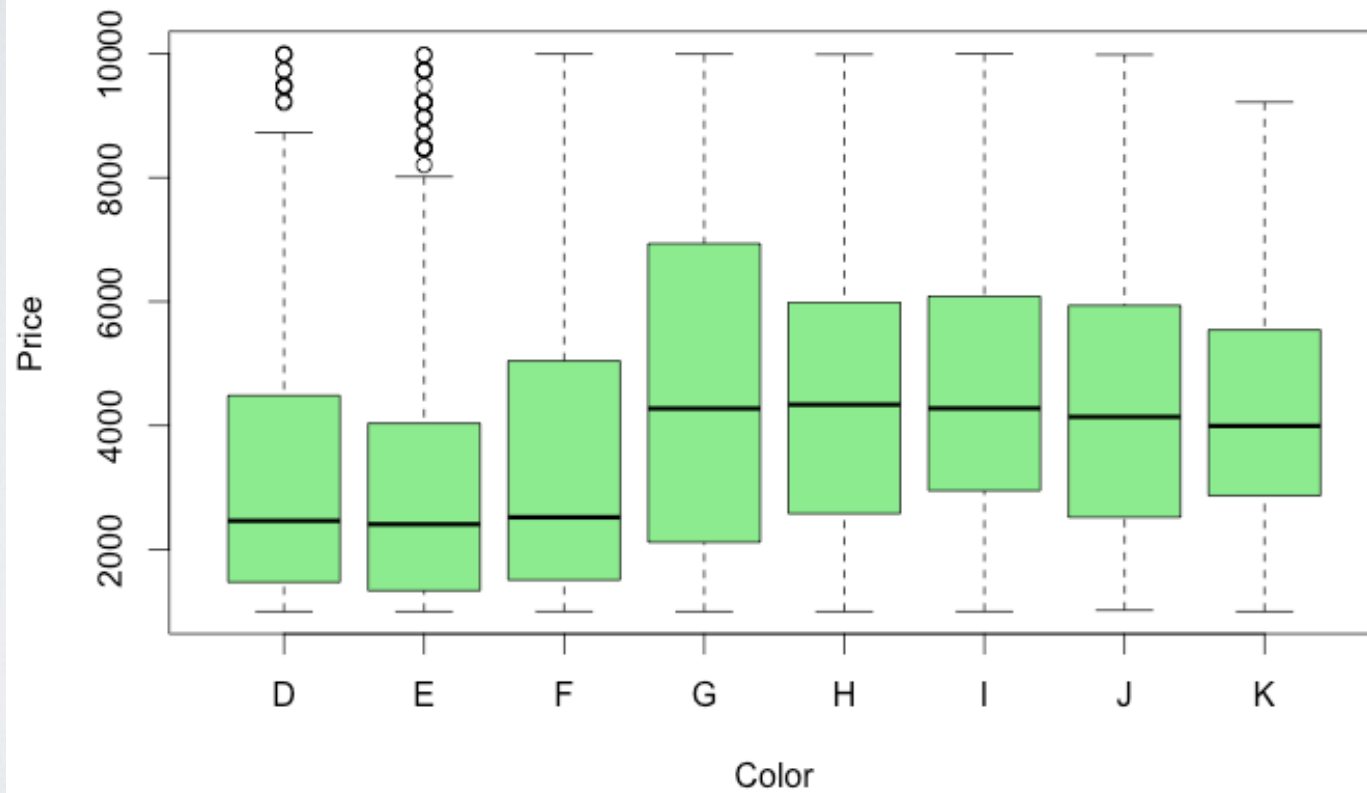


M

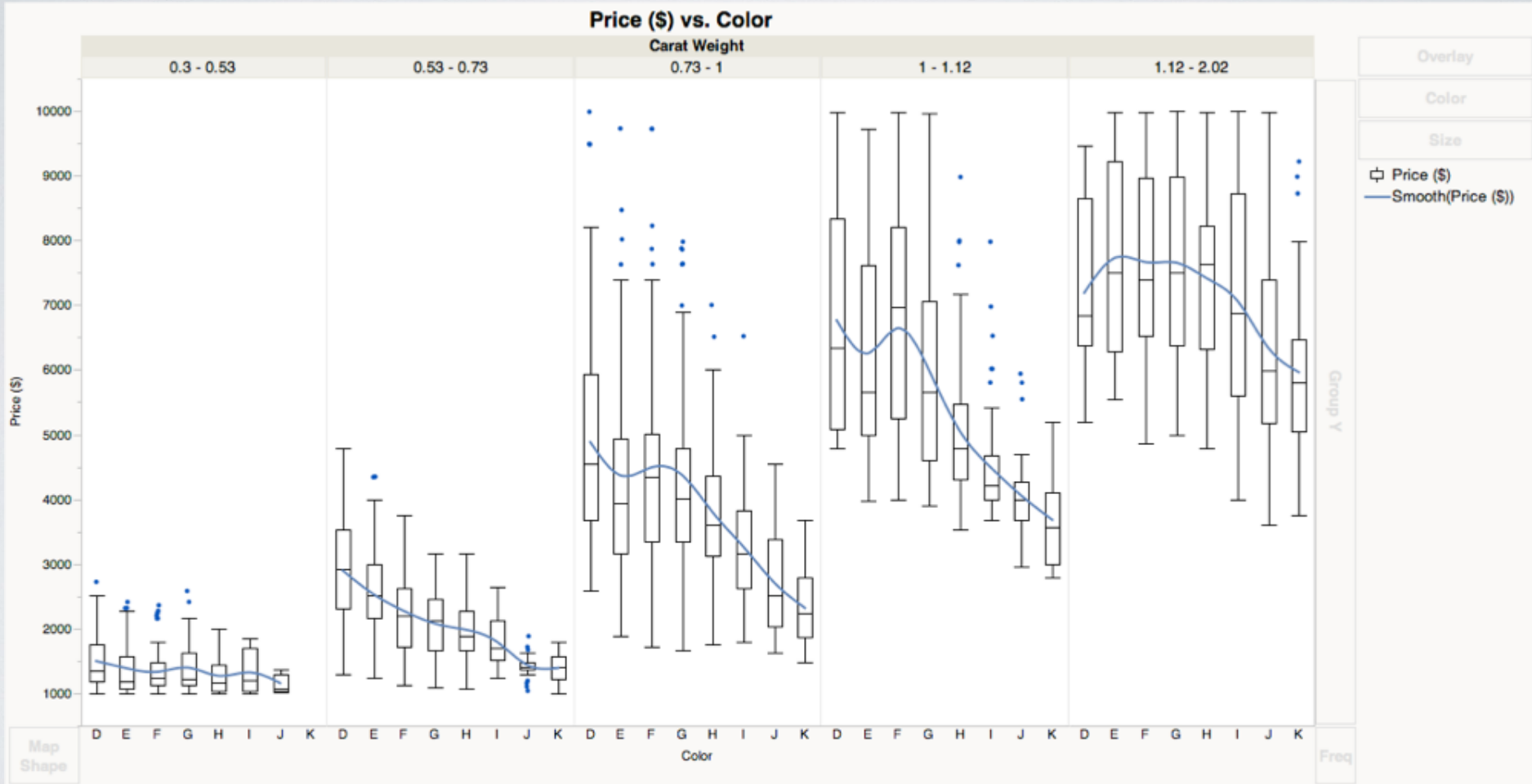


J

PRICE BY COLOR



HERE'S WHY

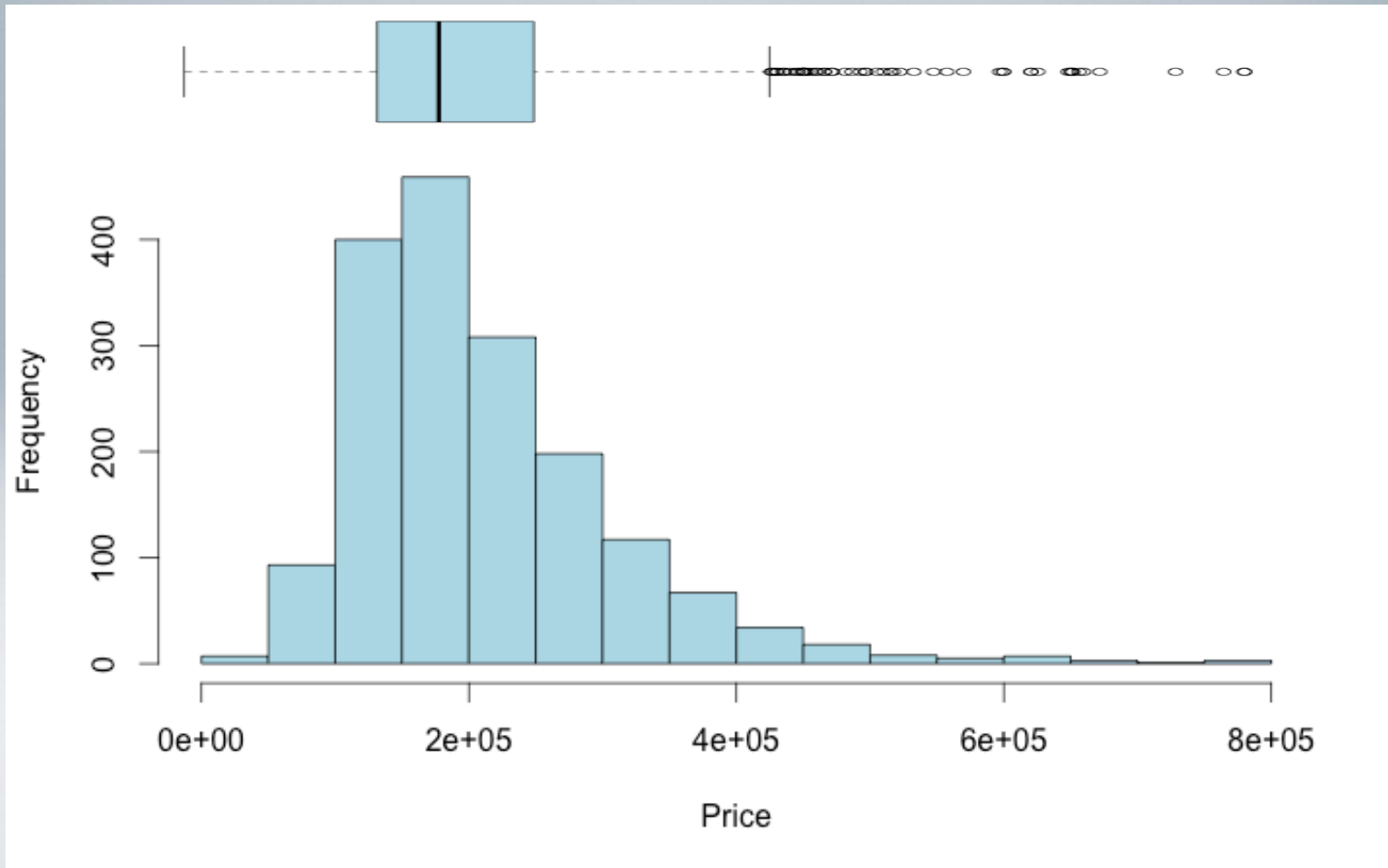


HOW MUCH IS A FIREPLACE WORTH?

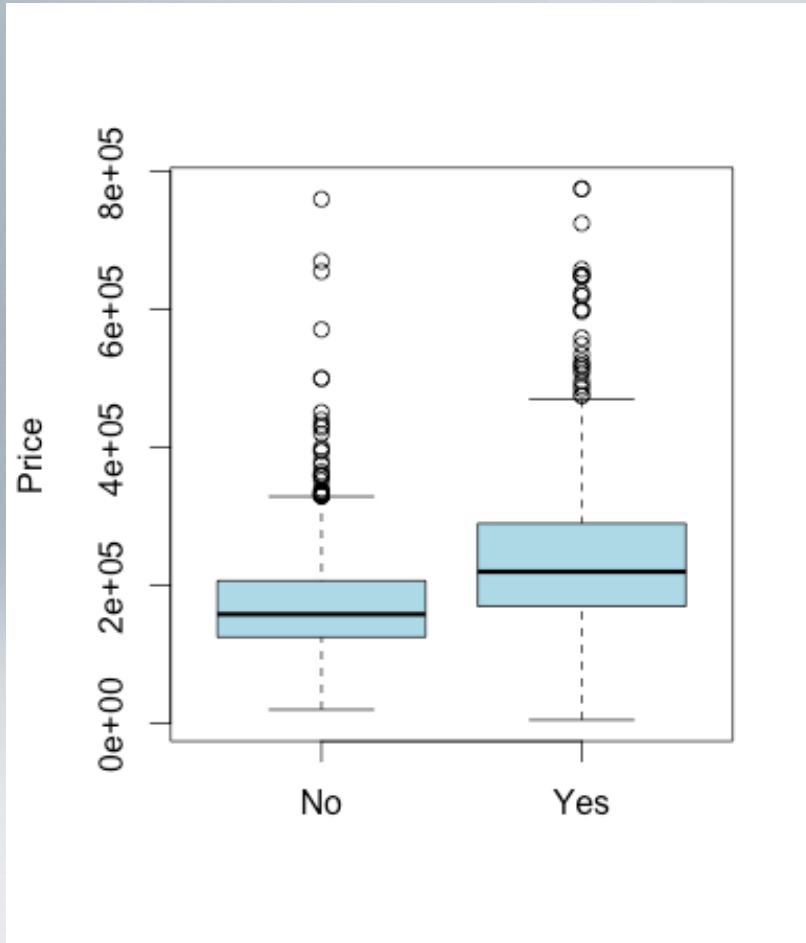
The Data

- 1729 randomly selected homes from the county records of Saratoga New York about 2005.
- Conclusions are only about the Northeast...

Starting Univariate



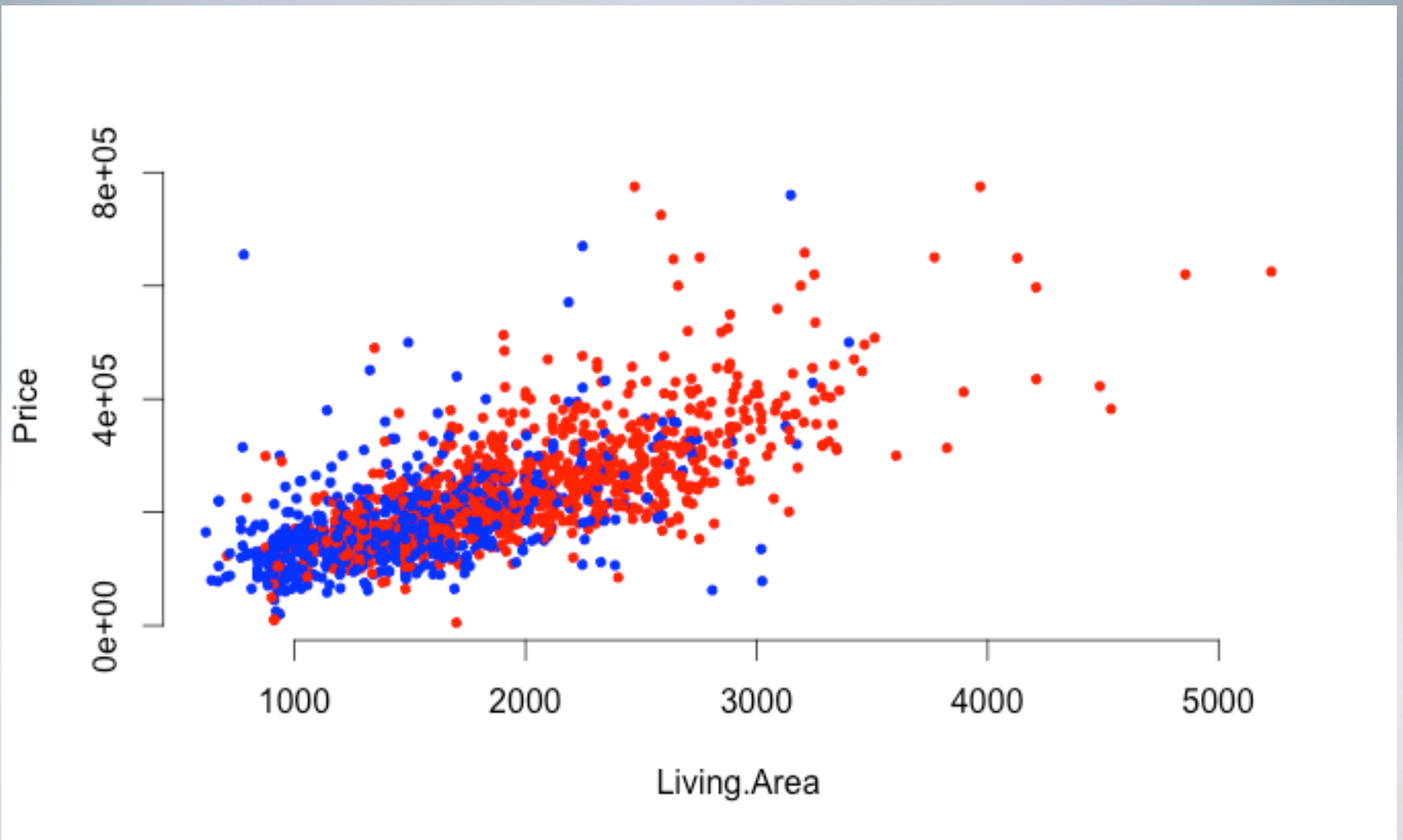
PRICE BY FIREPLACE?



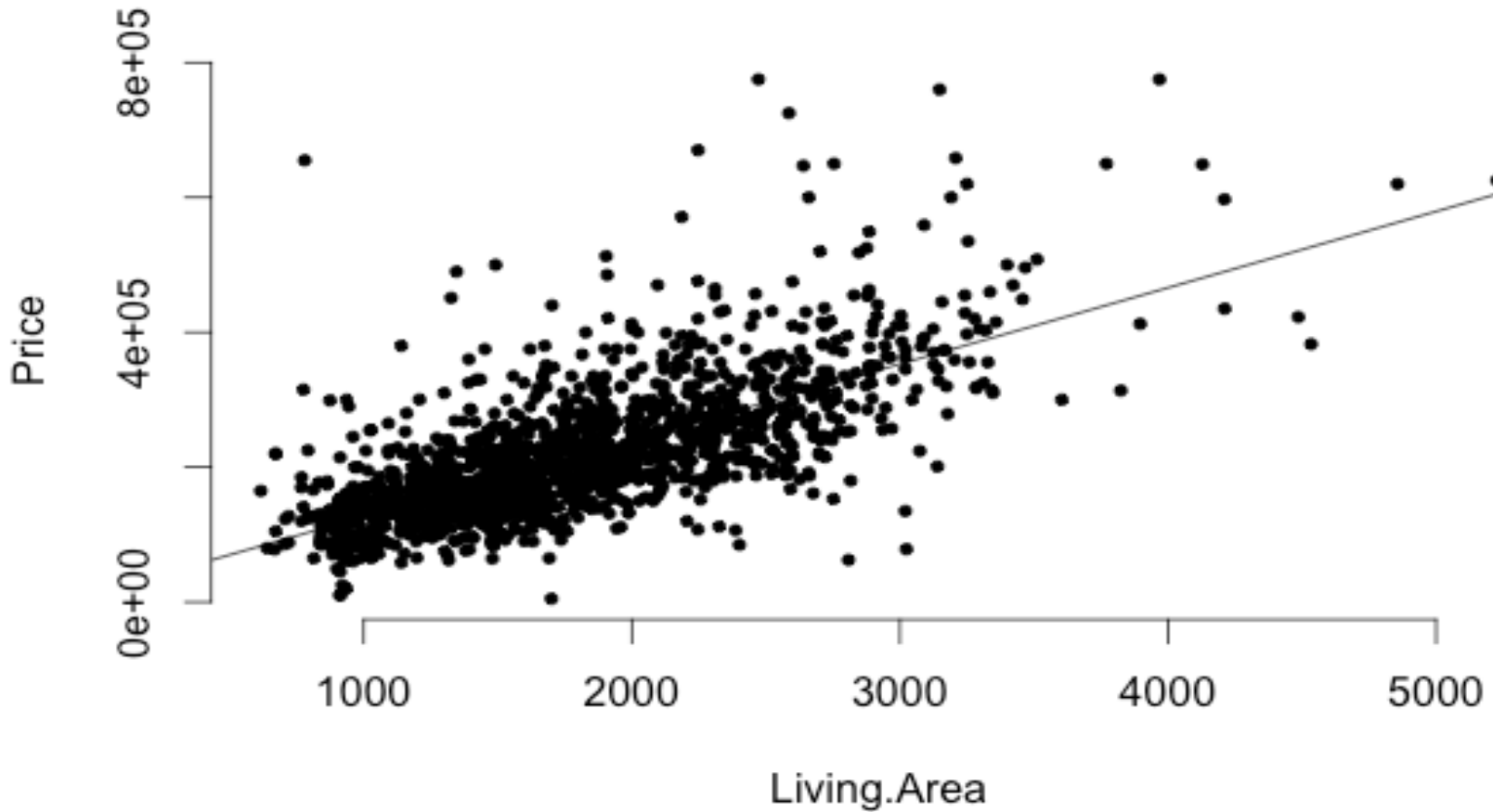
No 740 \$174,653
Yes 988 \$239,914

Difference \$65,620

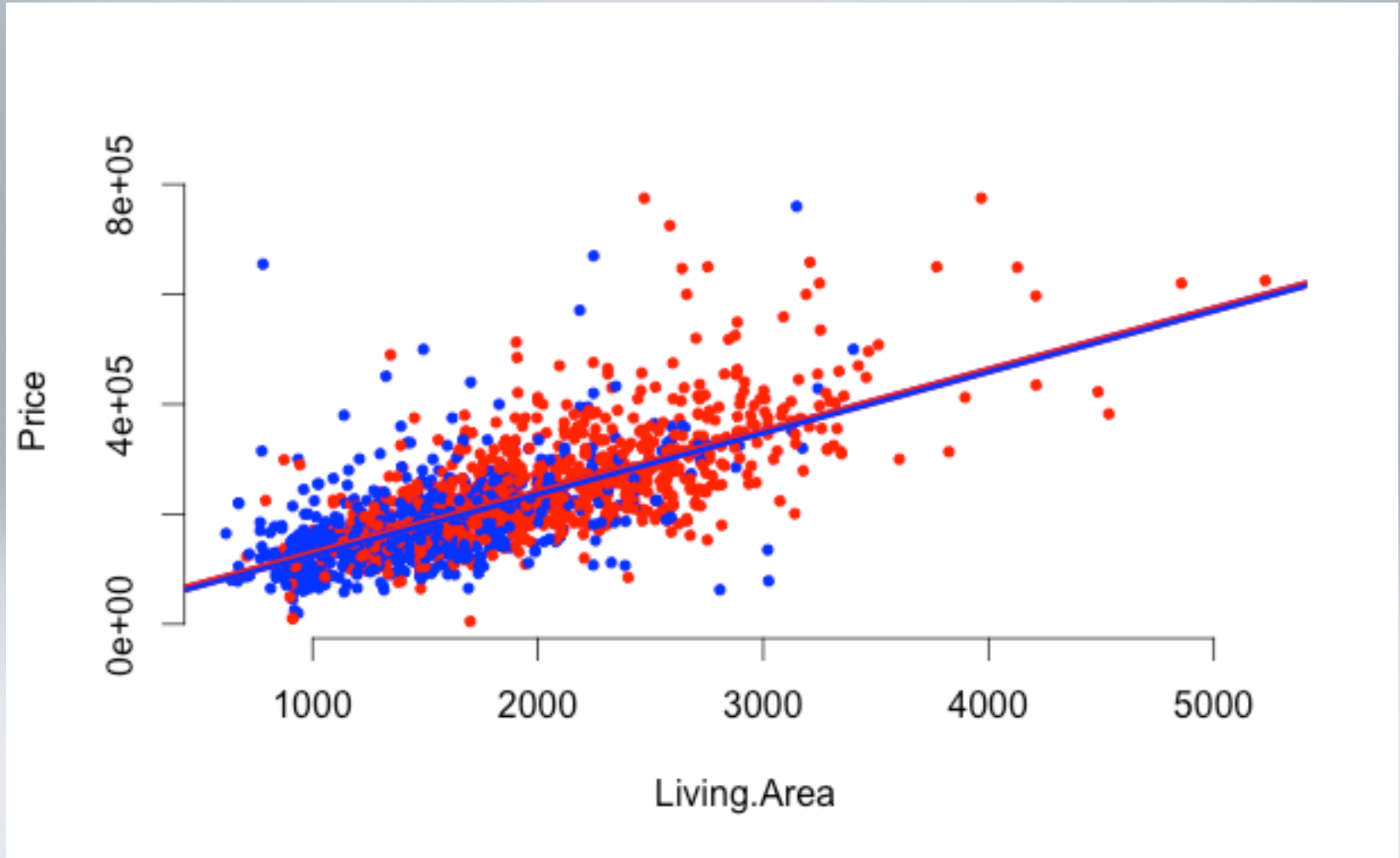
WAIT A MINUTE



Relationship with size

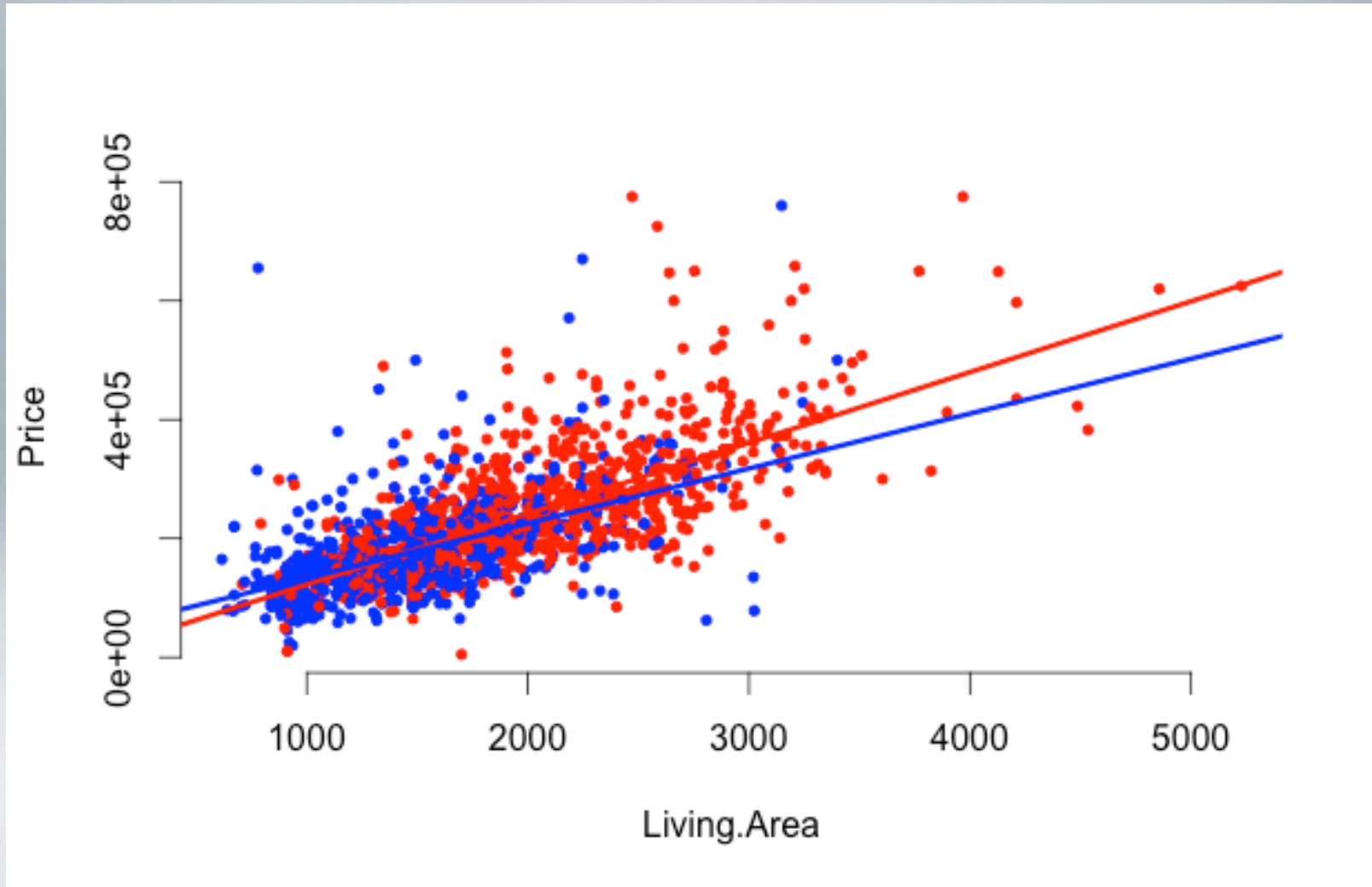


WHAT IF WE FIT TWO LINES?



Difference now is \$5567

WHAT IF WE LET THEM GO?

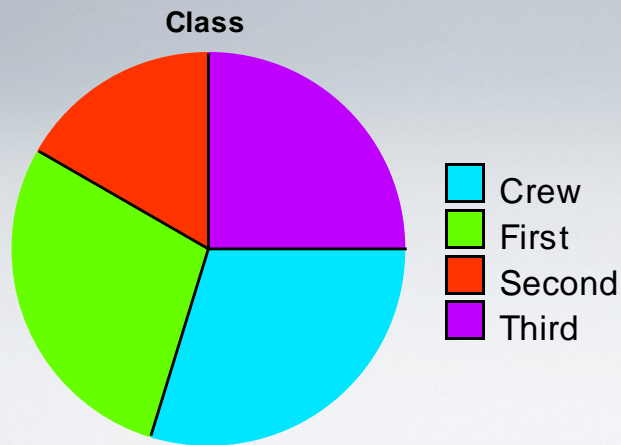


THE TITANIC

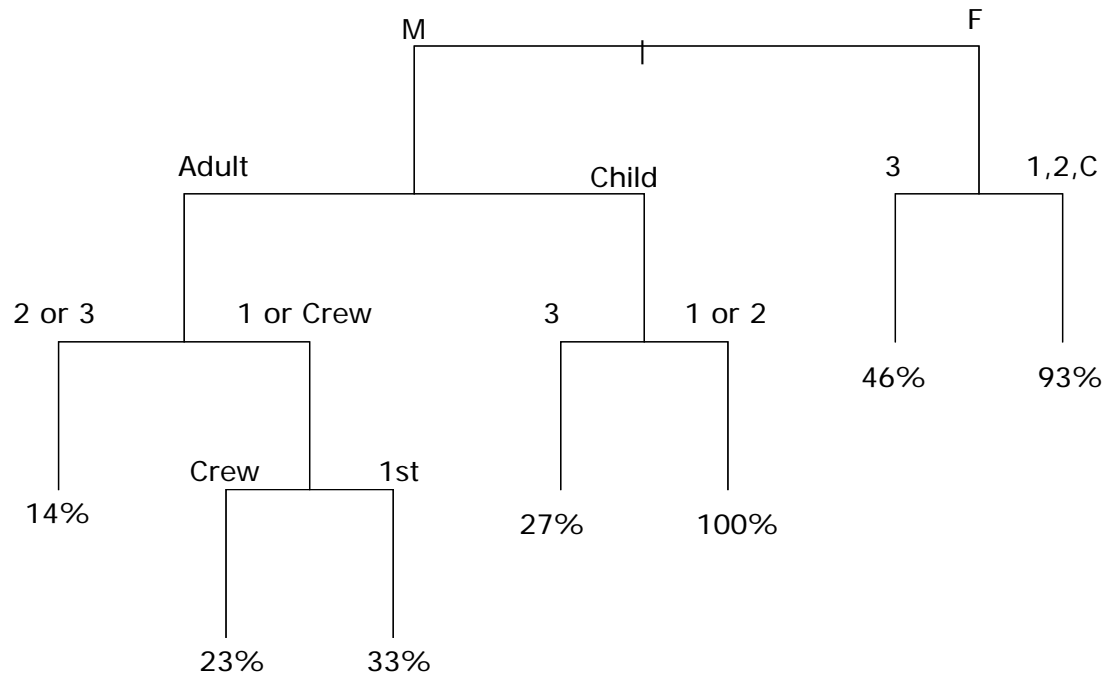
		Ticket Class				Total
		Crew	First	Second	Third	
Survival	Lived	212	202	118	178	710
	Died	673	123	167	528	1491
	Total	885	325	285	706	2201

Survivors

Non-Survivors



WHY NOT?



You've cracked the resampling egg,
why not enjoy the omelette?

WHERE ARE WE?

- Introduce models early
 - Motivate univariate questions from more complex models
 - Not the other way around!
- Use complex questions and models to drive topics
- Statistics is more than a collection of tools
 - Which ones are really important?
- How much should we ask them to “produce”?