Statistics for the 21st Century
Are We Teaching the Right Course?

Dick De Veaux
Williams College

Danny Kaplan
Macalester College

Our Theme

- We teach the wrong stuff
- We teach it the wrong way
- We teach it in the wrong order
- We don’t have the answers — but we’re good at asking the questions :)

What to Teach

- We should
  - Teach people what you want them to know
  - Have them do what you want them to be able to do
  - Get them to ask the questions they will need to ask in their other courses and eventually in their work

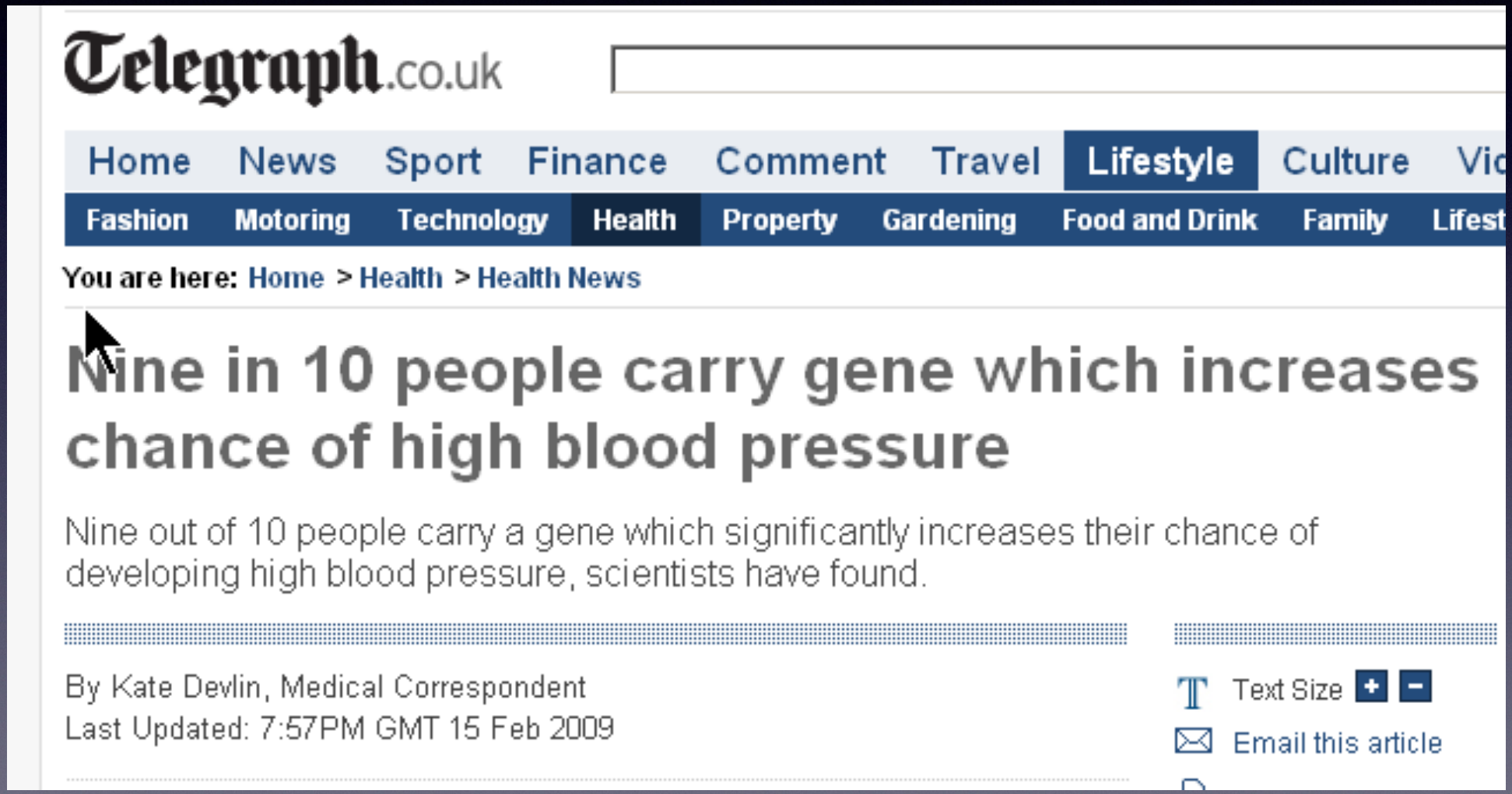
Topics

- Communicating Statistical Ideas
- Decision Making under Uncertainty
- Modeling (simple and complex)
- Data (and their issues)
- Inference and Causation

Communication

10% with genetic variant that reduces risk of high blood pressure

Huh?



Telegraph.co.uk

Home News Sport Finance Comment Travel Lifestyle Culture Vic

Fashion Motoring Technology Health Property Gardening Food and Drink Family Lifest

You are here: Home > Health > Health News

Nine in 10 people carry gene which increases chance of high blood pressure

Nine out of 10 people carry a gene which significantly increases their chance of developing high blood pressure, scientists have found.

By Kate Devlin, Medical Correspondent
Last Updated: 7:57PM GMT 15 Feb 2009

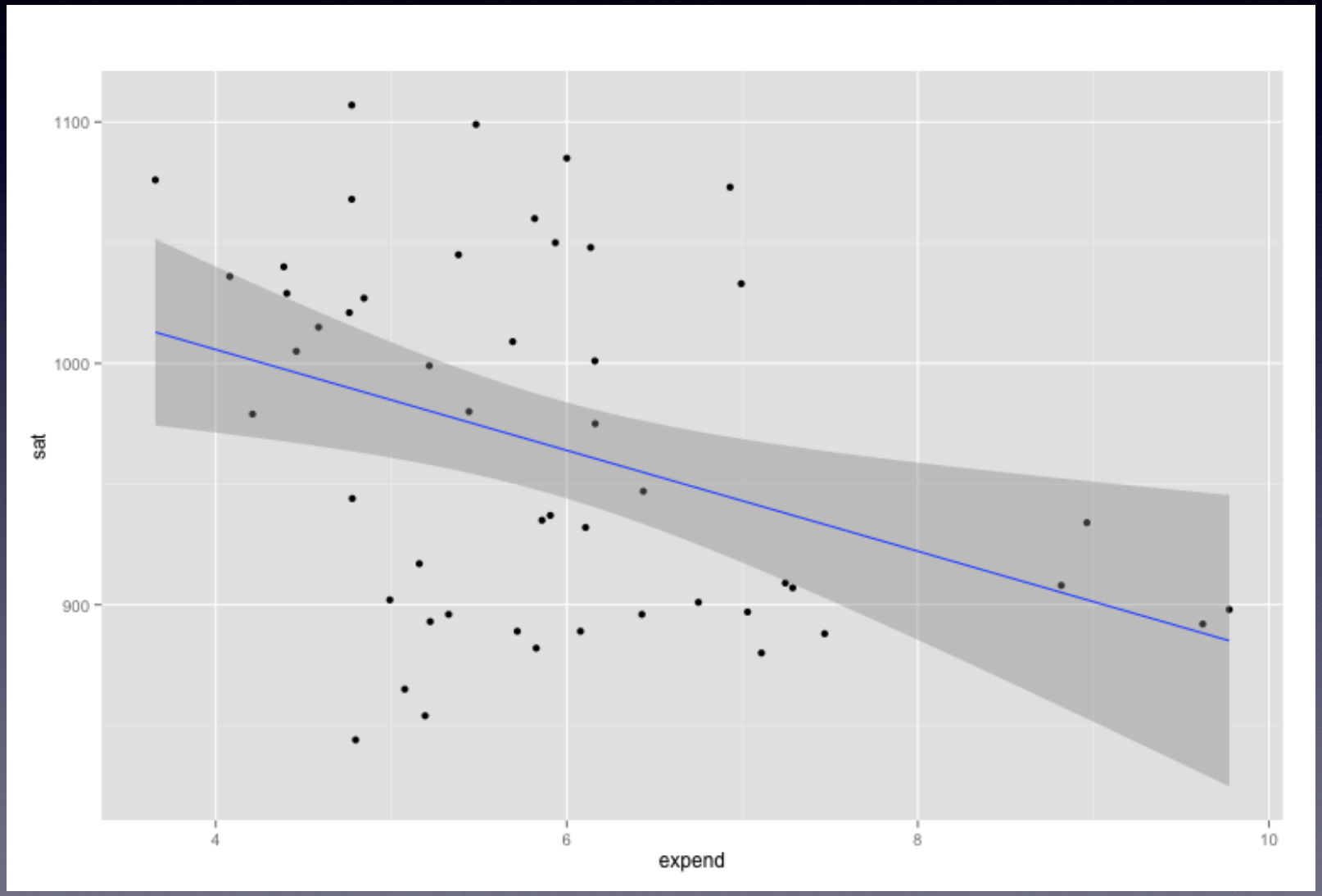
Text Size + -

Email this article

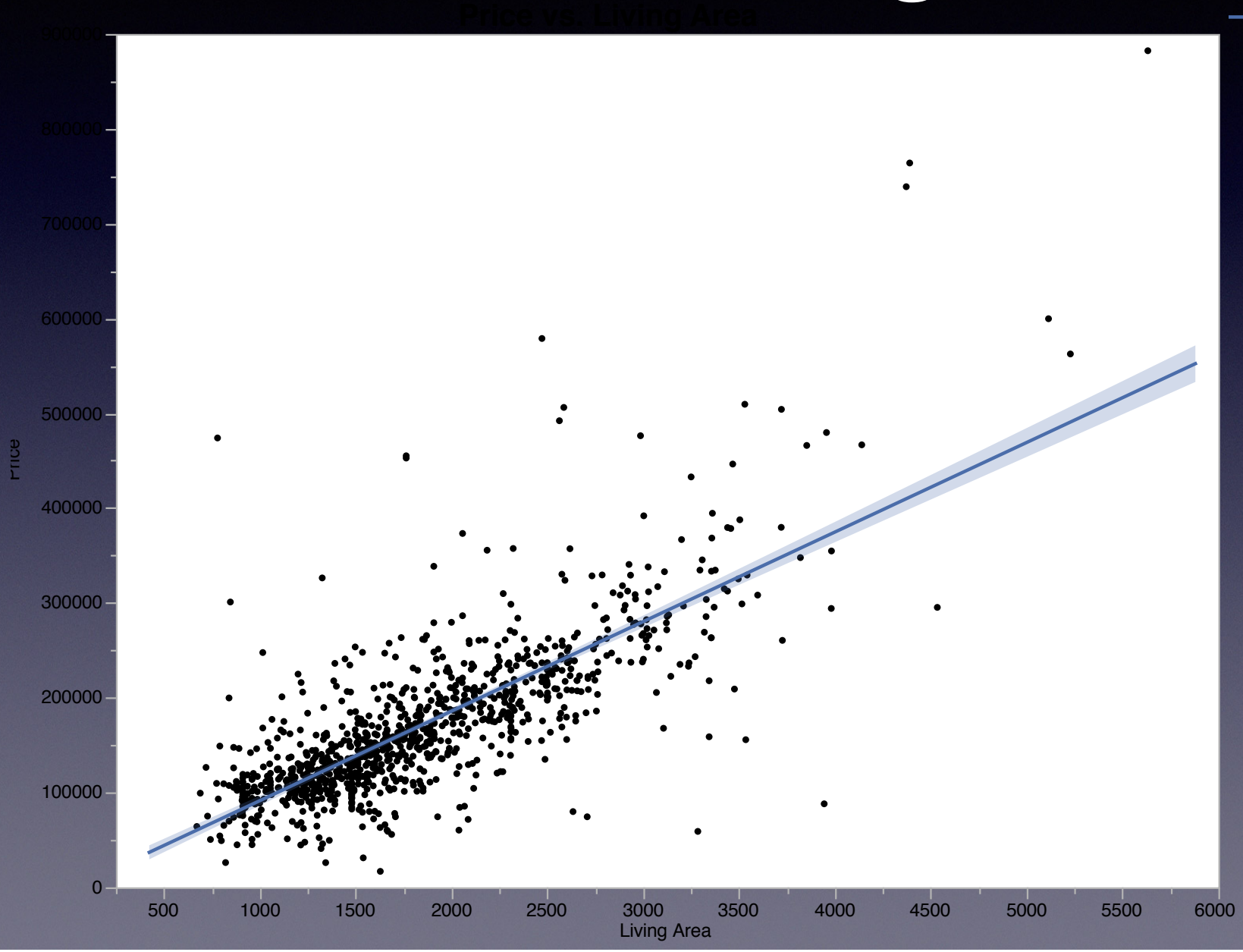
Communication

- Think about the audience
  - Spectators, referees or players? Roxy Peck
- We should teach the concepts in a way that illustrates how they will communicate what they've learned

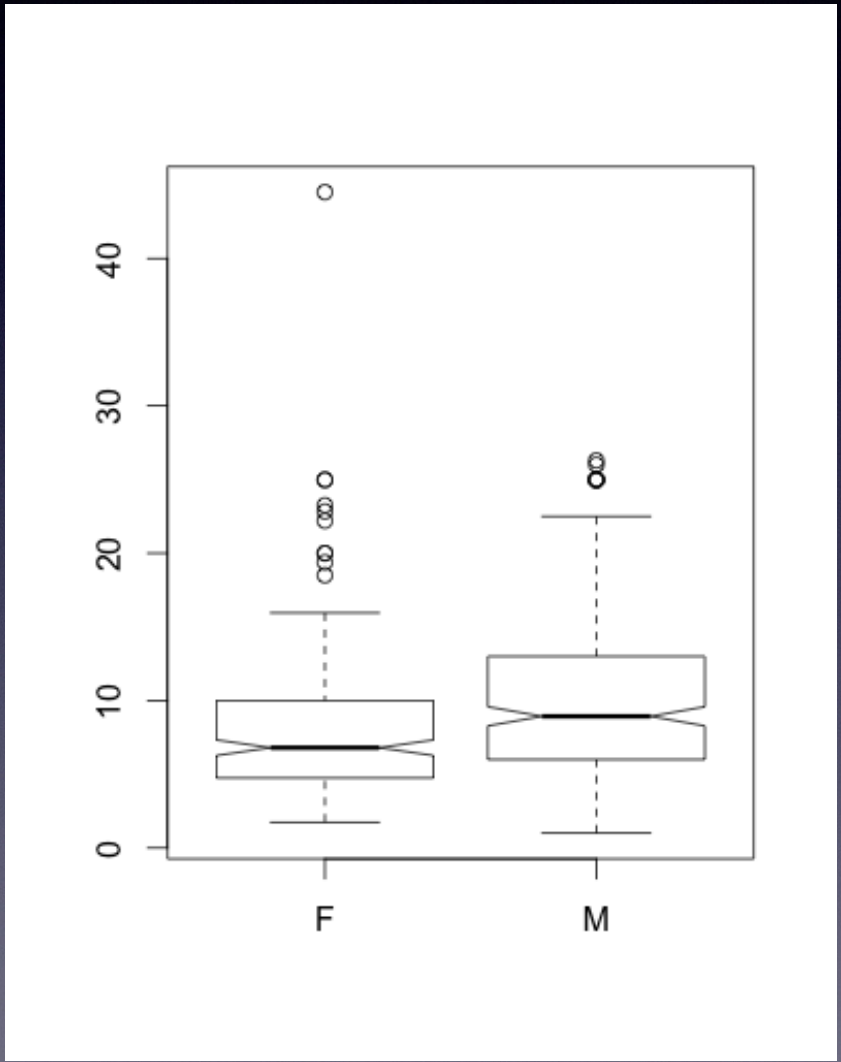
SAT vs. Sch Expenditure



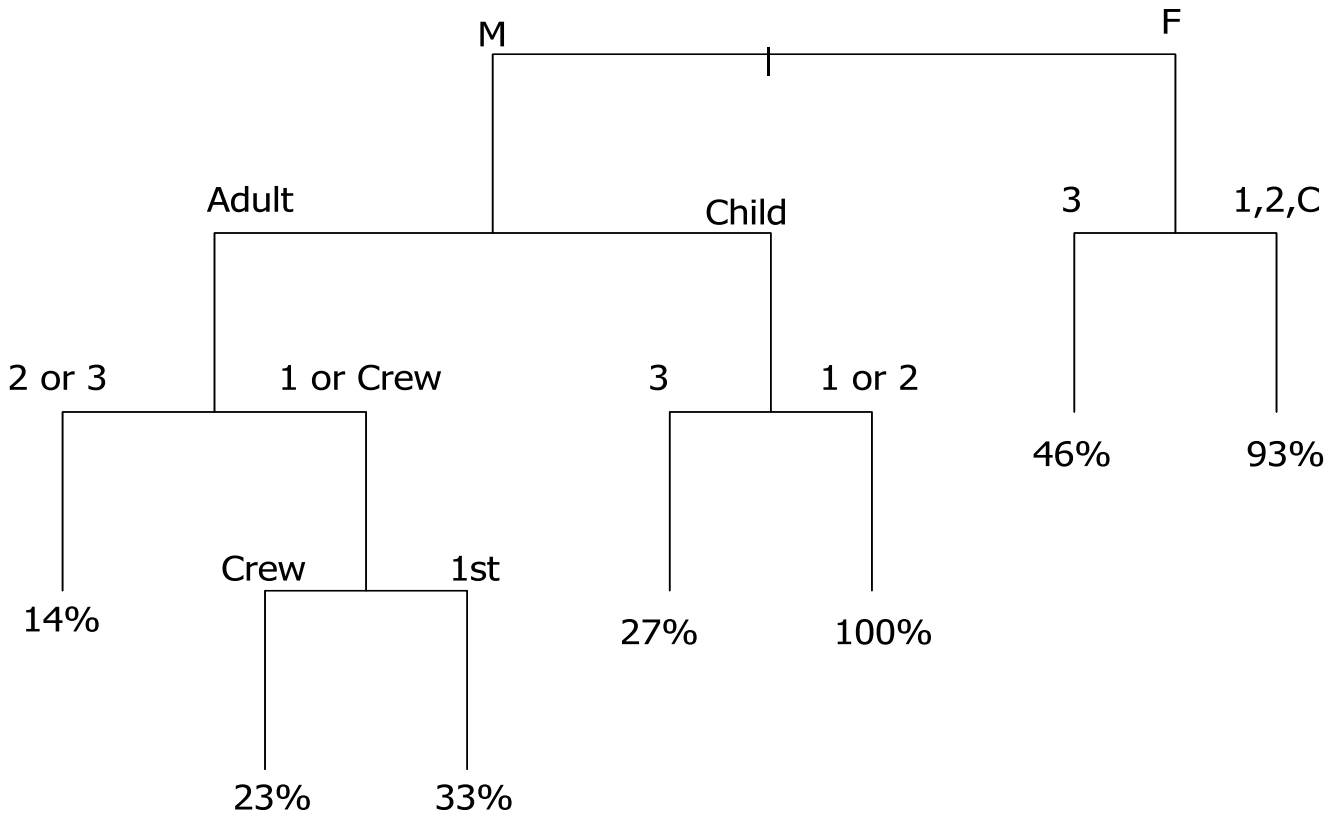
Price vs. Living Area



Wage vs. Sex



Shouldn’t we Include this?



Decision Making

- Tukey
  - Decisions vs. Conclusions
    - Business vs. Science
- Conditional Probability

Bayes

$$P(\mathbf{A}_i \mid \mathbf{B}) = \frac{P(\mathbf{B} \mid \mathbf{A}_i)P(\mathbf{A}_i)}{\sum_j P(\mathbf{B} \mid \mathbf{A}_j)P(\mathbf{A}_j)}$$

Bayes

$$P(\mathbf{A}_i \mid \mathbf{B}) = \frac{P(\mathbf{B} \mid \mathbf{A}_i)P(\mathbf{A}_i)}{\sum_j P(\mathbf{B} \mid \mathbf{A}_j)P(\mathbf{A}_j)}$$

Screening

100000 people
1000 have disease

5% false pos
0.1% false neg

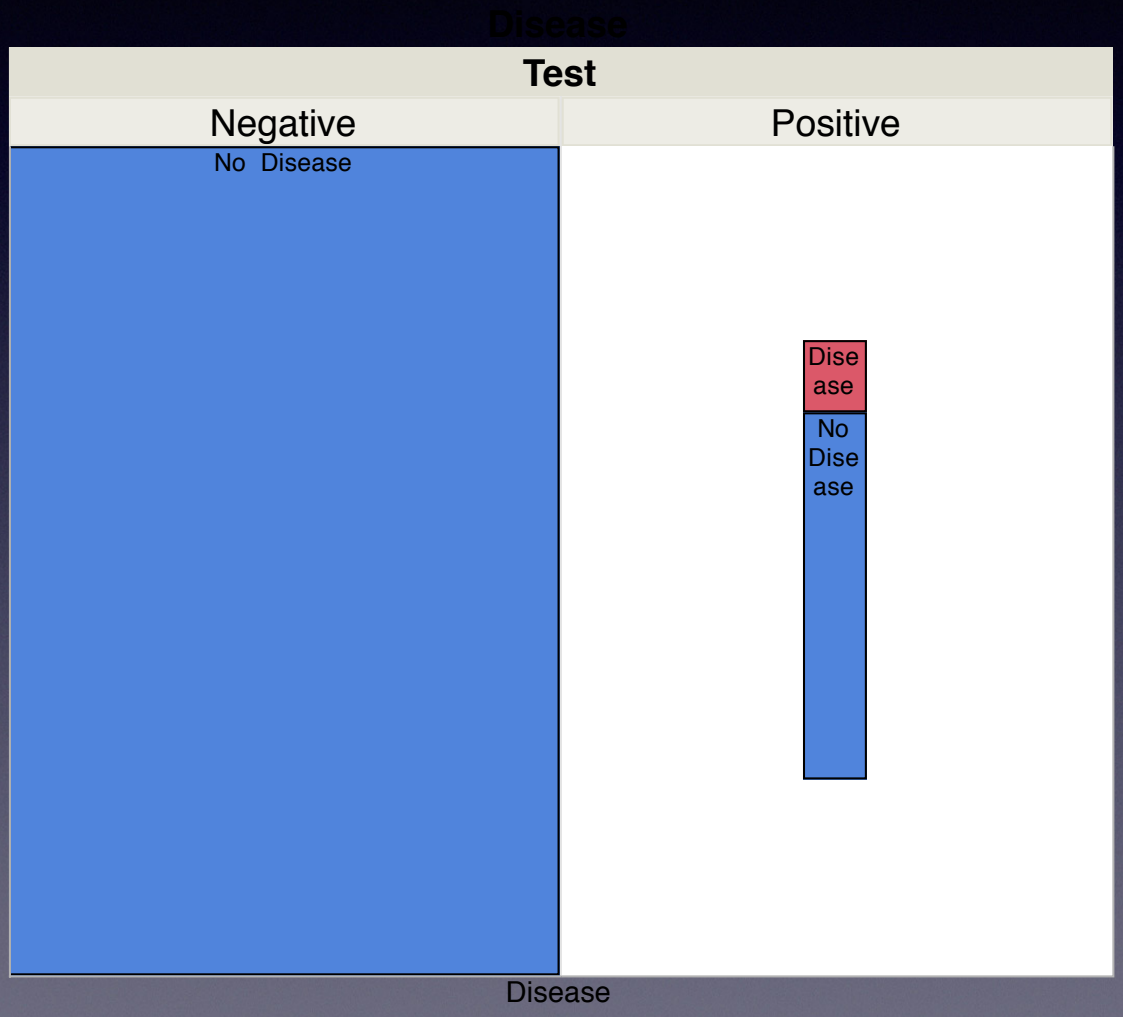
4950 false pos
990 true positives

Screening

100000 people
1000 have disease

5% false pos
0.1% false neg

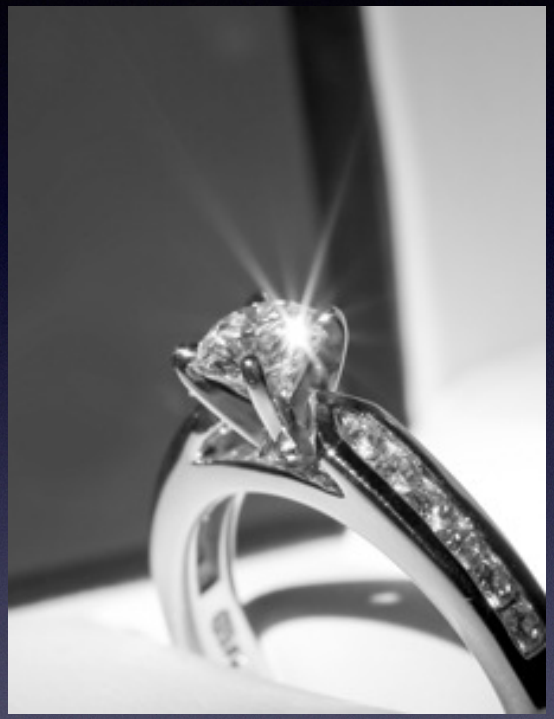
4950 false pos
999 true positives



Complex Models

- Diamonds
  - The 4 C's
- Graduation Rates
  - Simpson's Paradox?

Diamond Prices



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: Radiant

Carat Weight: .3 example: 0.98 or 1.73 or 3.95

Color Scale

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: D

Clarity Scale

Flawless	VVS_1 VVS_2	VS_1 VS_2	SI_1 SI_2	I_1 I_2 I_3
Internally Flawless				Imperfect

Clarity: IF

Certification: GIA (Your Best Choice) *

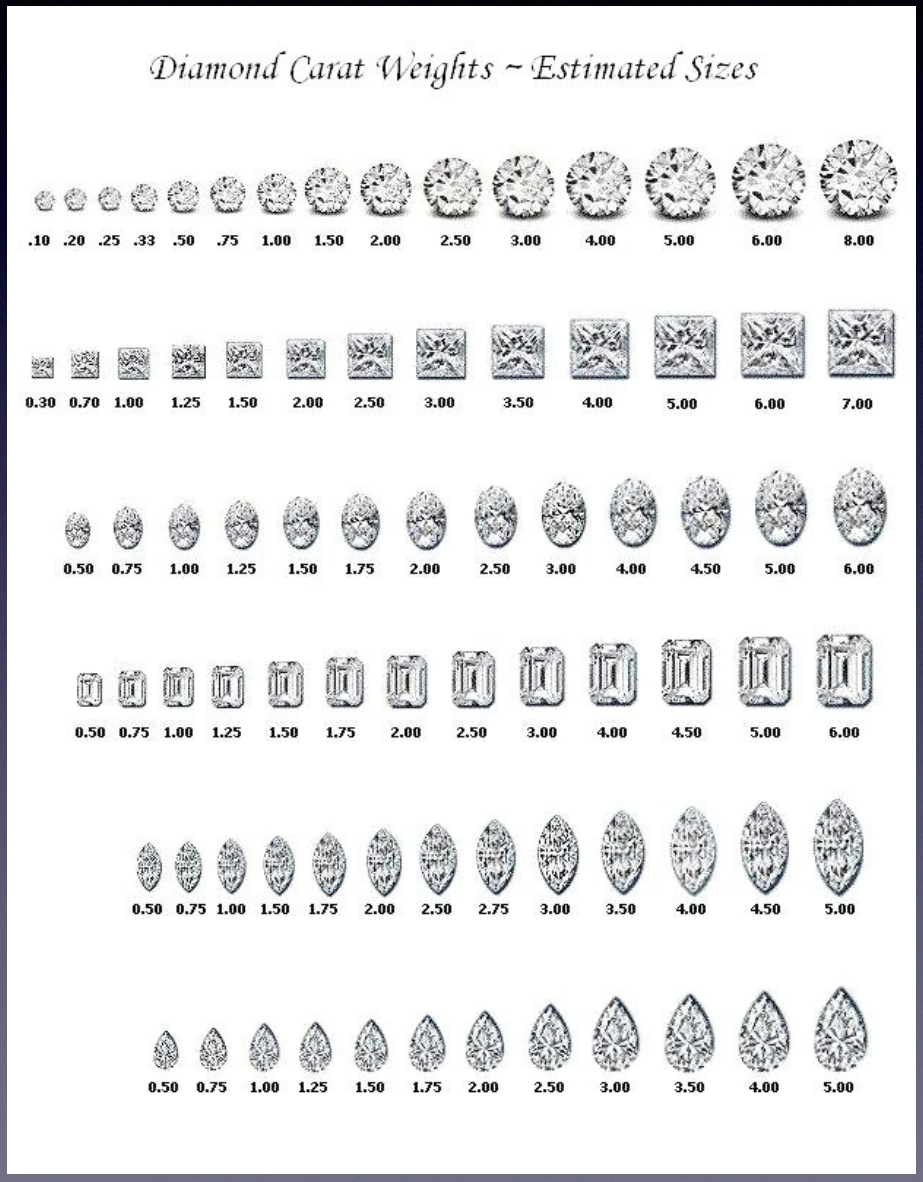
Proportions: Excellent *

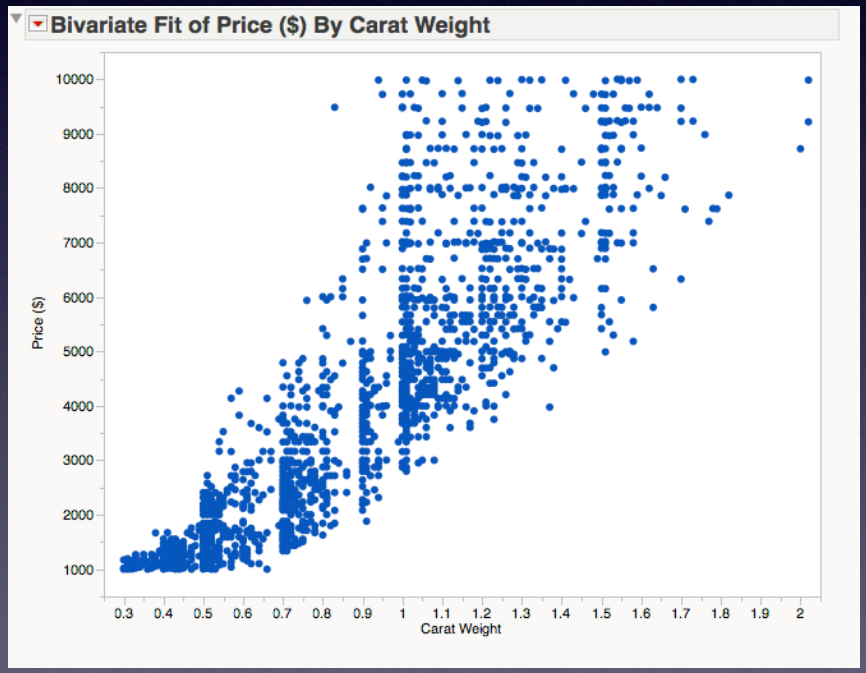
Calculate Price Clear Fields

* Important price determinant

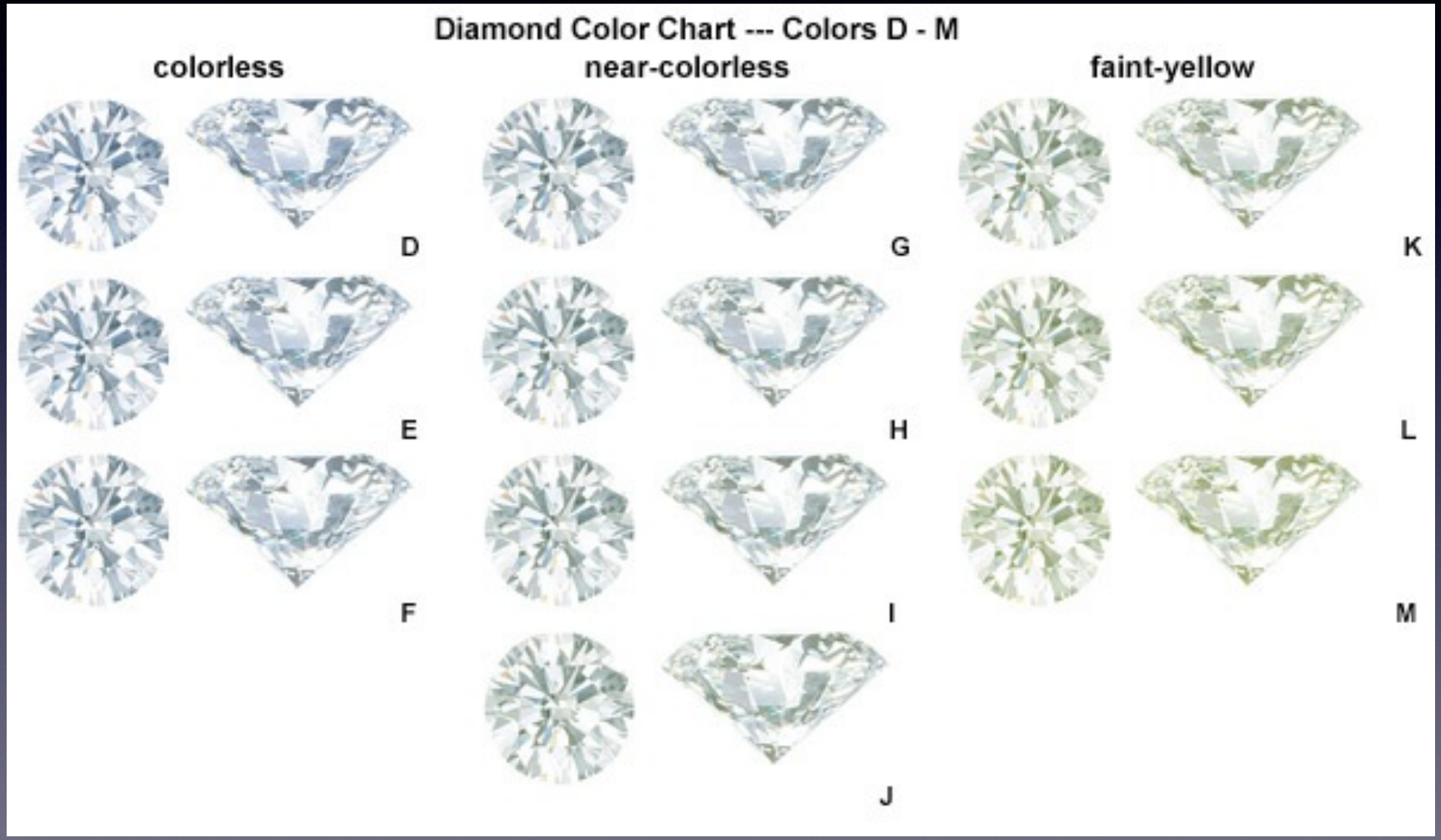
Carat Color Cut Clarity

Carat

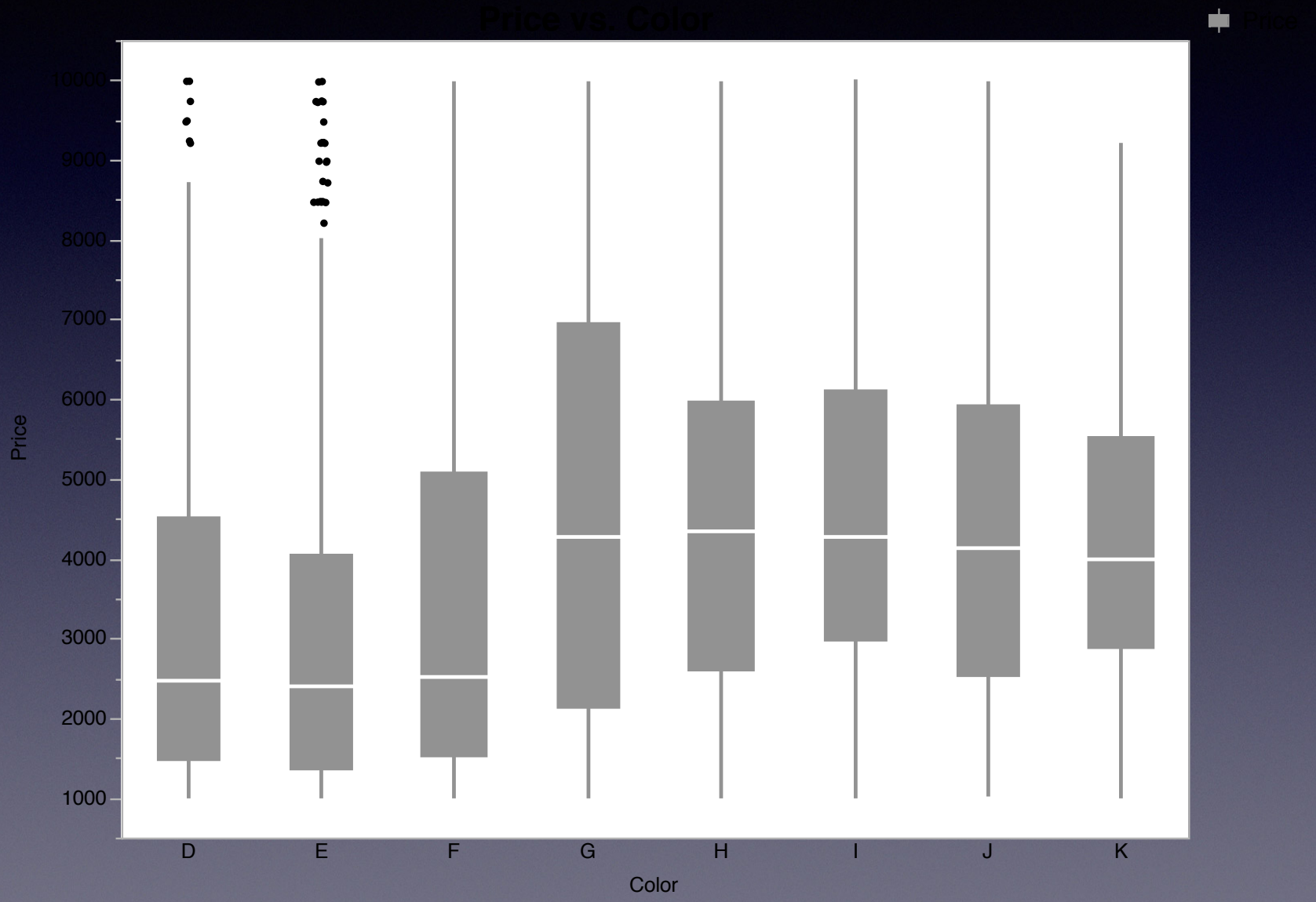

Diamond Carat Weights ~ Estimated Sizes
.10 .20 .25 .33 .50 .75 1.00 1.50 2.00 2.50 3.00 4.00 5.00 6.00 8.00
0.30 0.70 1.00 1.25 1.50 2.00 2.50 3.00 3.50 4.00 5.00 6.00 7.00
0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.50 3.00 4.00 4.50 5.00 6.00
0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.50 3.00 4.00 4.50 5.00 6.00
0.50 0.75 1.00 1.50 1.75 2.00 2.50 2.75 3.00 3.50 4.00 4.50 5.00
0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.50 3.00 3.50 4.00 5.00


Bivariate Fit of Price ($) By Carat Weight
Price ($)
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2
Carat Weight

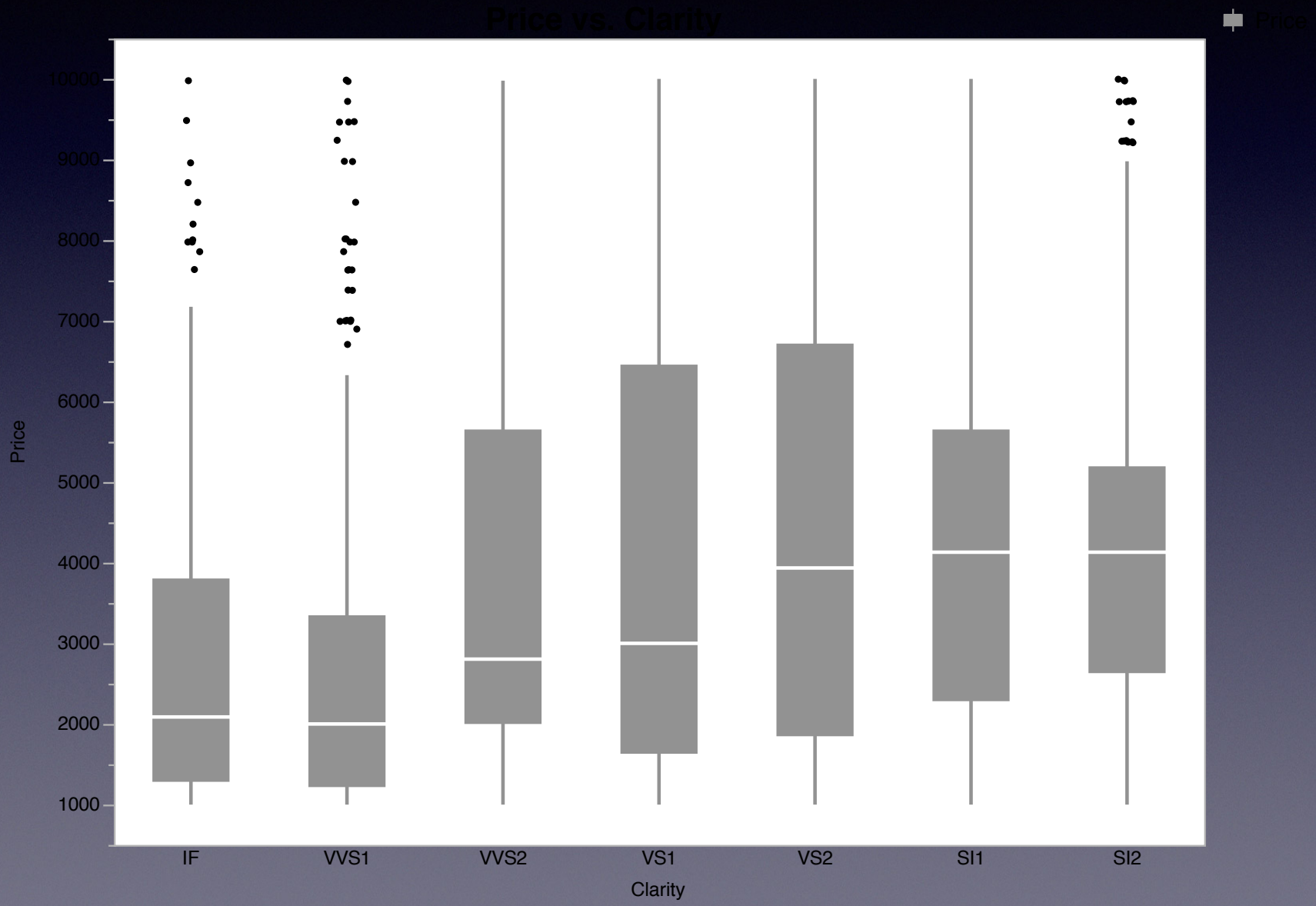
Color


Diamond Color Chart --- Colors D - M
colorless
near-colorless
faint-yellow
D
E
F
G
H
I
J
K
L
M

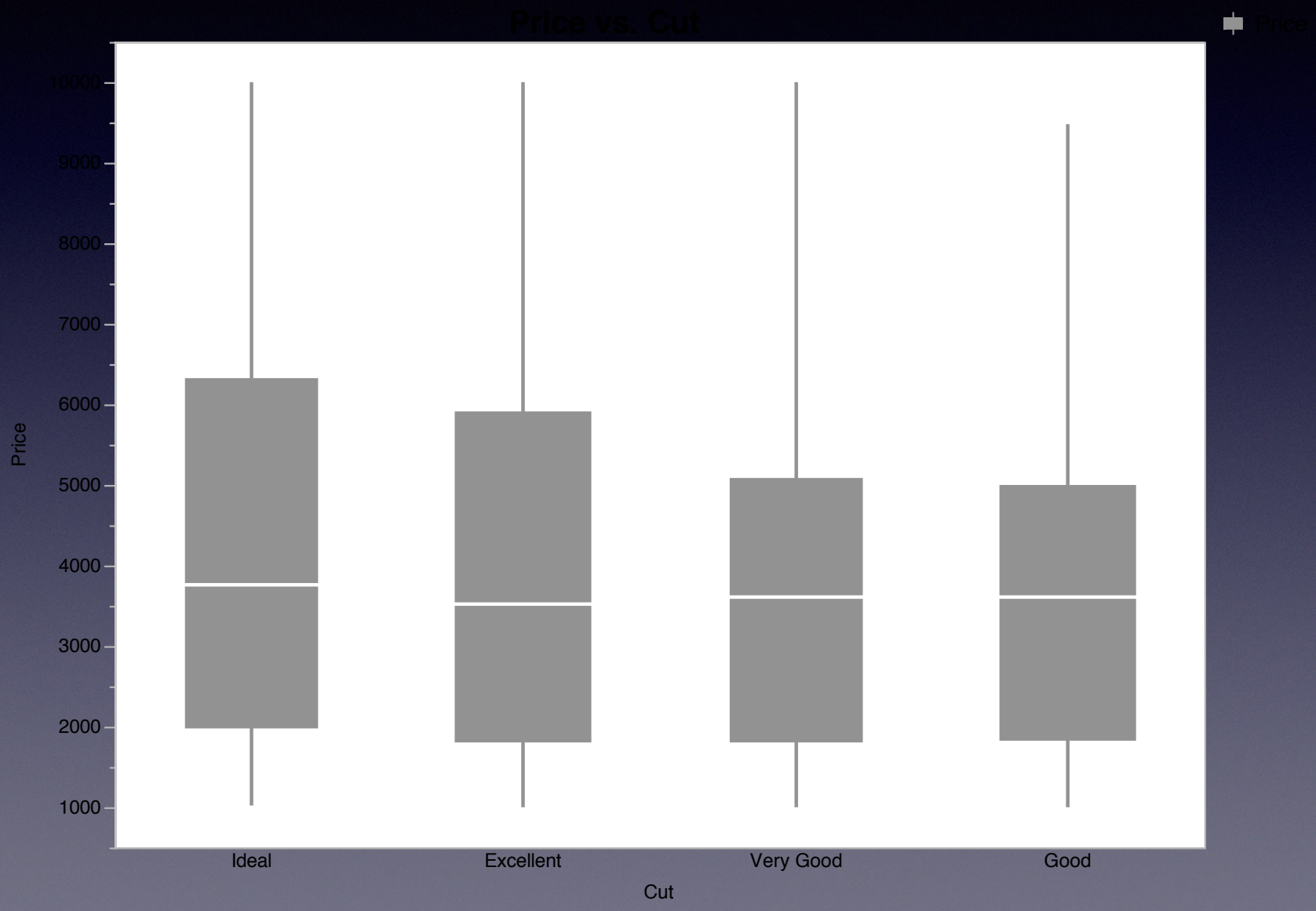
Price vs. Color



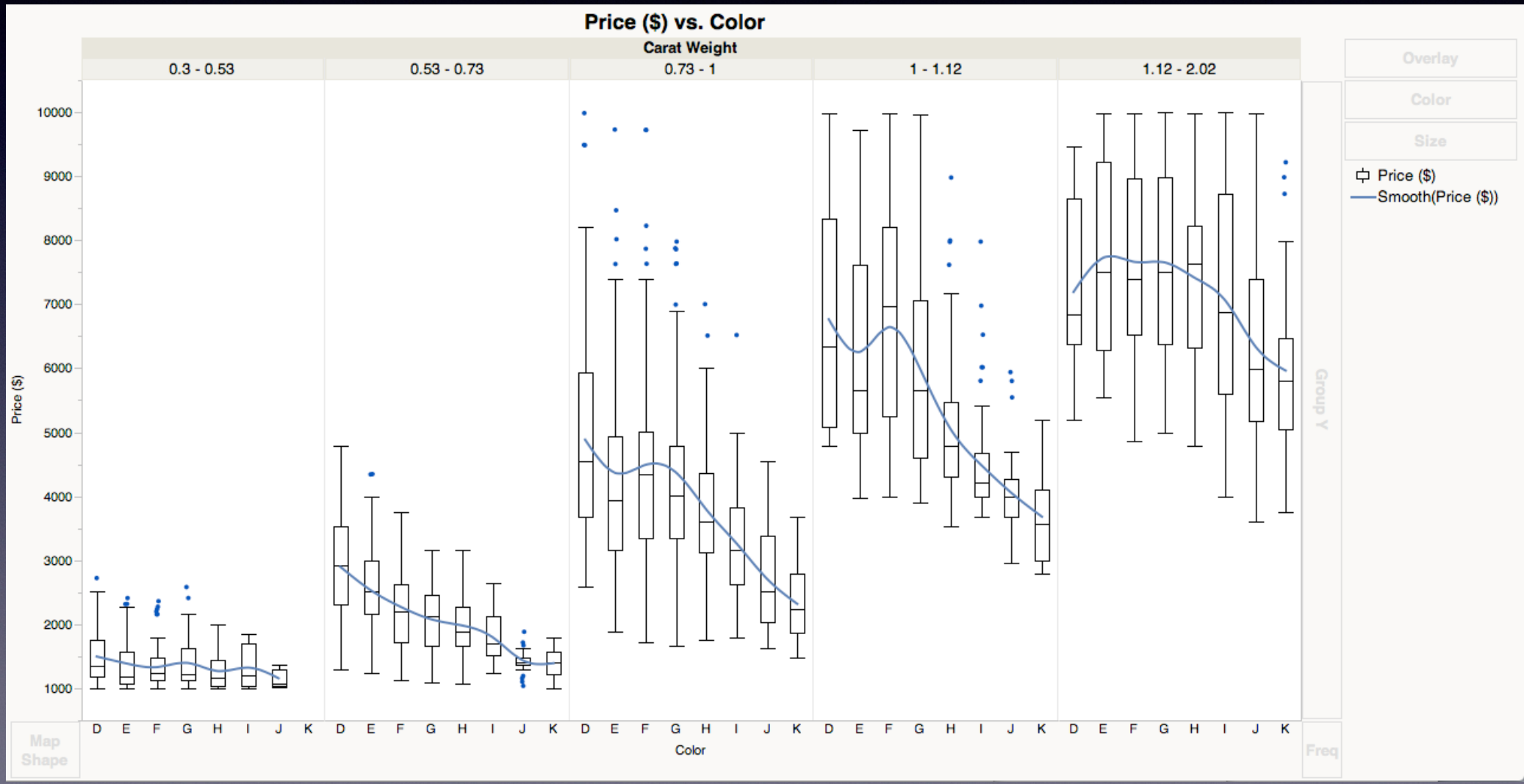
Price vs. Clarity



Price vs. Cut


Price vs. Cut
Price
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
Price
Ideal
Excellent
Very Good
Good
Cut

This is why


Price ($) vs. Color
Carat Weight
0.3 - 0.53
0.53 - 0.73
0.73 - 1
1 - 1.12
1.12 - 2.02
Price ($)
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
D E F G H I J K
Color
Overlay
Color
Size
Price ($)
Smooth(Price ($))
Group Y
Freq
Map Shape

Data

- Data are
  - Messy
  - Wrong
  - Missing
  - From different Places
- What will their data look like?
- Basic Operations on Data
  - Select, group, merge, transform, create

Inference and Causation

- What are Parameters?
- What are we Doing?
- How to make plausible conclusions about causality?
  - Adjustment
  - Mechanism
    - Inform their next steps
    - Statistical Thinking

How

- What we shouldn't be teaching
  - What can go
- Formulas vs. Notation

$$SE(\hat{y}_v) = \sqrt{SE^2(b_1) \times (x_v - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}.$$

$$b_1 = r\frac{s_y}{s_x}.$$

How II

- Teaching the Course Backward
- Computation and Notation
- Simulation and Resampling

Why Don’t we Do This?

- Environment — inertia
- Textbooks
- Support
- Background
- Proofiness

Where do we go from Here?

- Communication
- Data
- Inference
- Models