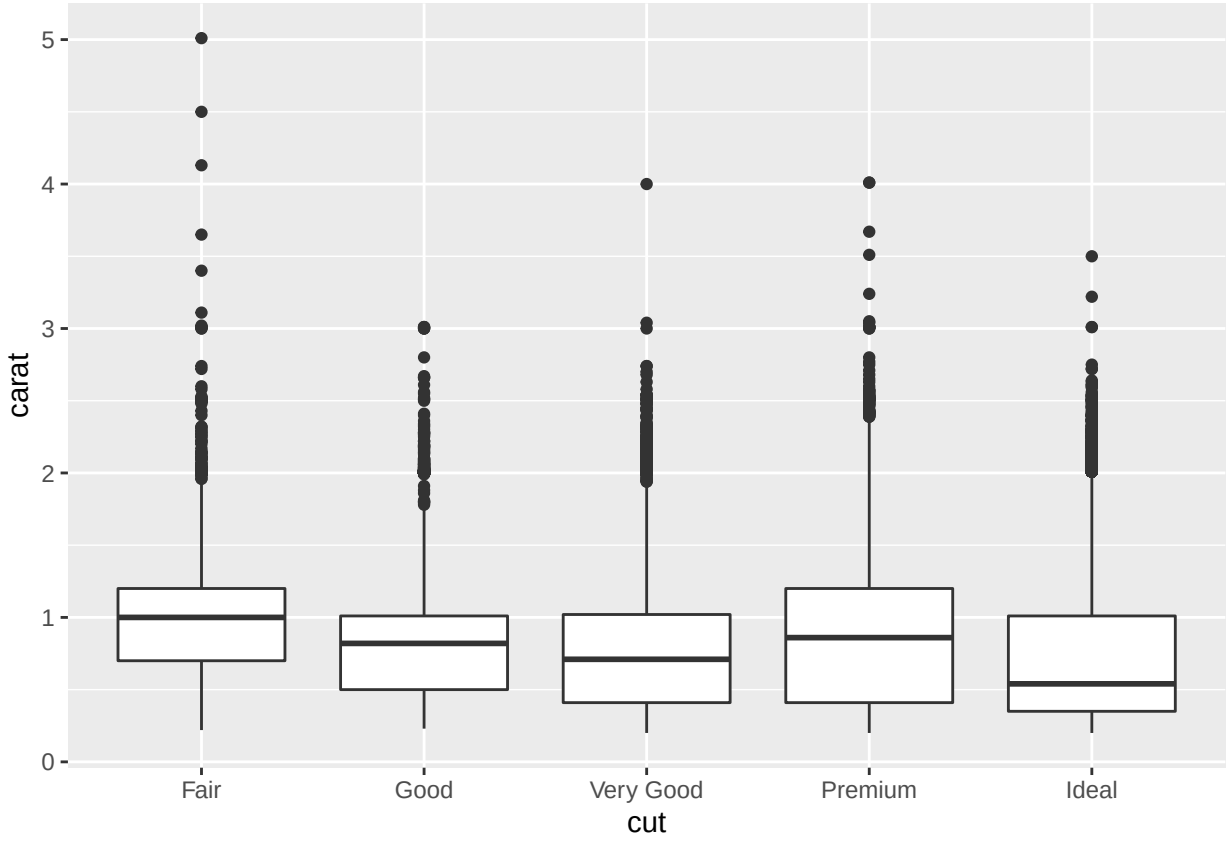
Diamond sizes

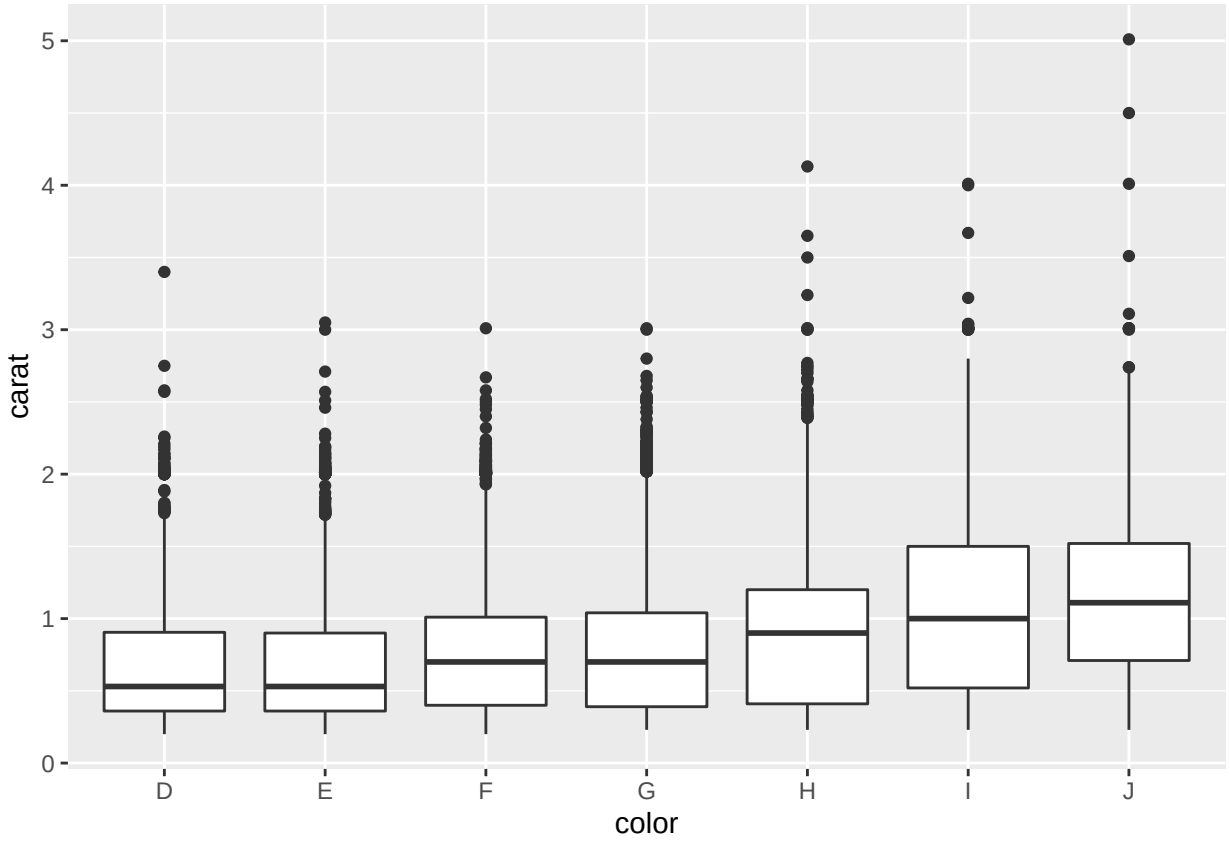
2016-08-25

Size and Cut, Color and Clarity

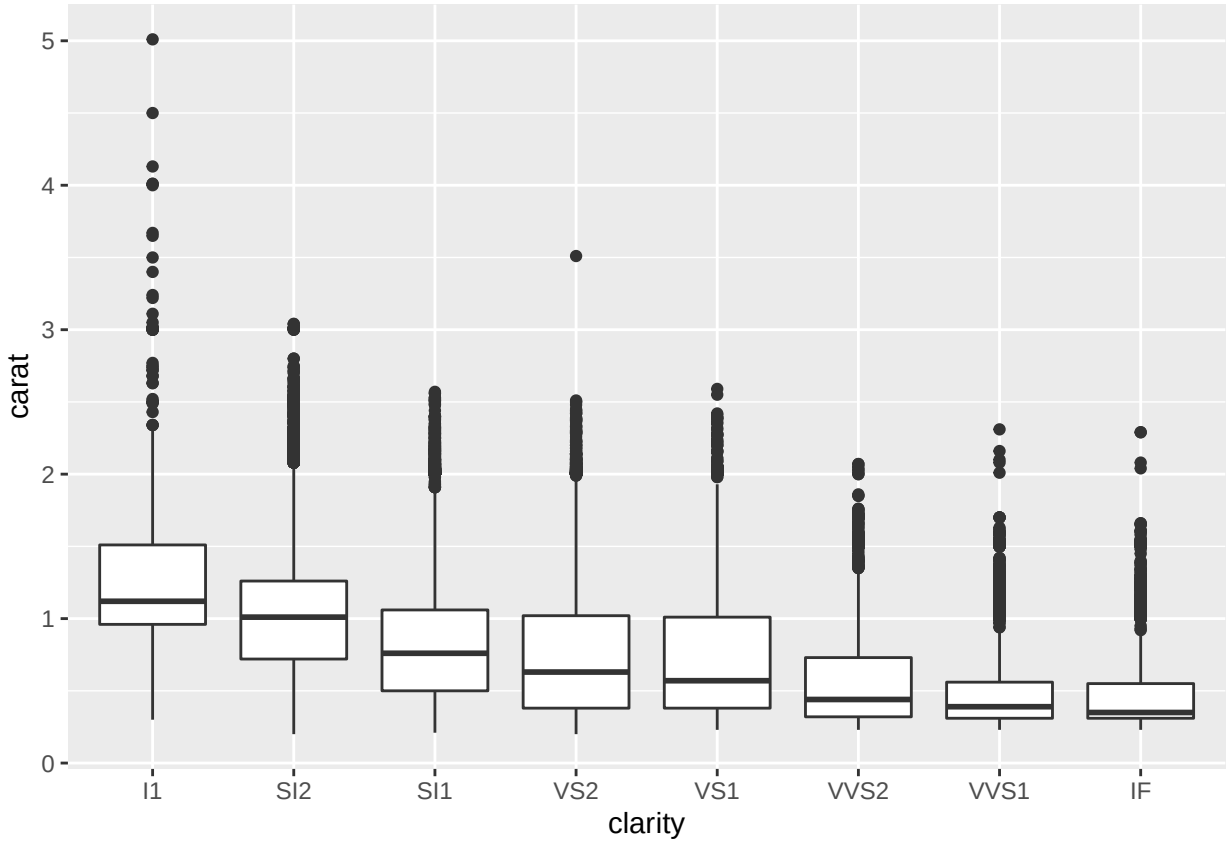
Diamonds with lower quality cuts (cuts are ranked from "Ideal" to "Fair") tend to be larger.



Likewise, diamonds with worse color (diamond colors are ranked from J (worst) to D (best)) tend to be larger:



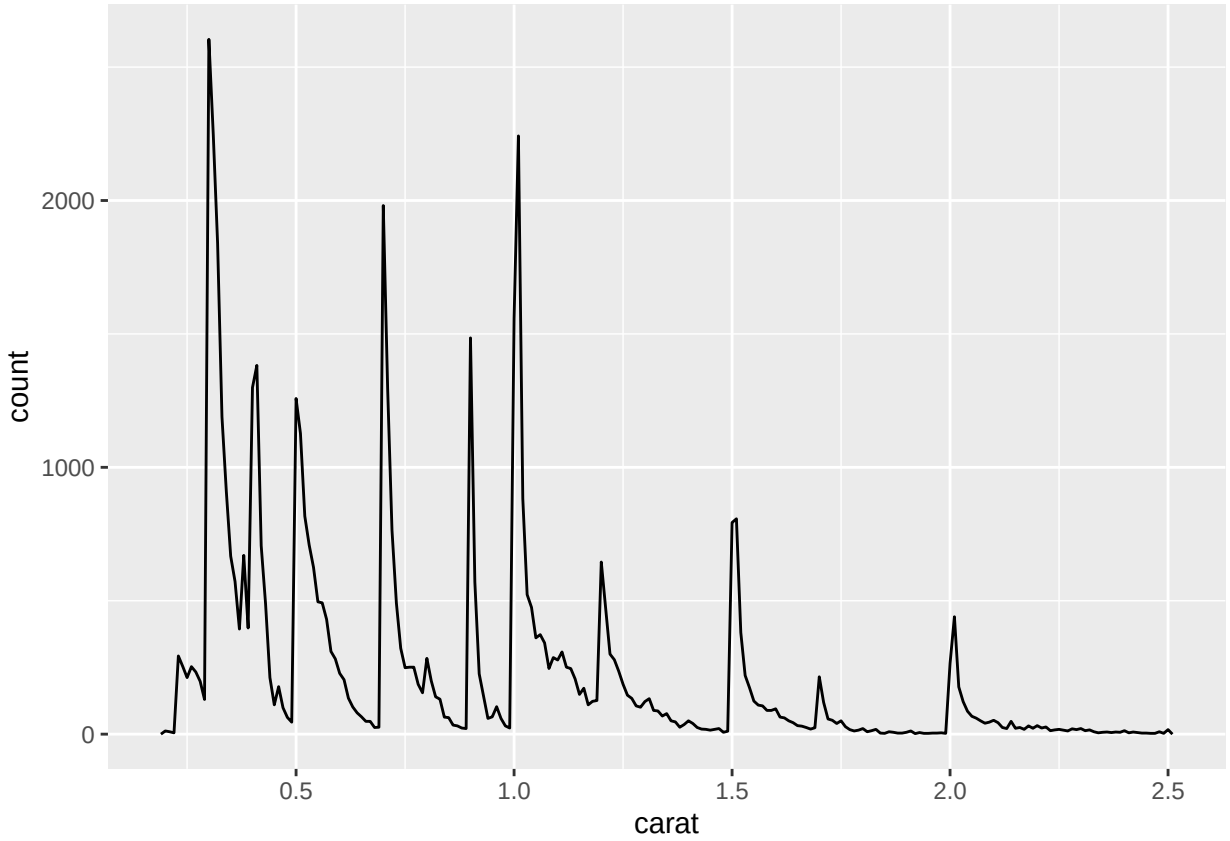
The pattern present in cut and color os also present in clarity. Diamonds with worse clarity (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best)) tend to be larger:



This patterns are consistent with there being a profitability threshold for retail diamonds that is a function of carat, clarity, color, cut and other characteristics. A diamond may be profitable to sell if a poor value of one feature, for example, poor clarity, color or cut, is be offset by a good value of another feature, such as large size. This can be considered an example of Berkson's paradox.

Largest diamonds

We have data about 53,940 diamonds. Only 126 (0.2%) are larger than 2.5 carats. The distribution of the remainder is shown below:



The frequency distribution of diamond sizes is marked by spikes at whole-number and half-carat values, as well as several other carat values corresponding to fractions.

The largest twenty diamonds (by carat) in the dataset are,

Table 1: The largest 20 diamonds in the `diamonds` dataset.

carat	cut	color	clarity
5.01	Fair	J	I1
4.50	Fair	J	I1
4.13	Fair	H	I1
4.01	Premium	I	I1
4.01	Premium	J	I1
4.00	Very Good	I	I1
3.67	Premium	I	I1
3.65	Fair	H	I1
3.51	Premium	J	VS2
3.50	Ideal	H	I1
3.40	Fair	D	I1
3.24	Premium	H	I1
3.22	Ideal	I	I1
3.11	Fair	J	I1
3.05	Premium	E	I1
3.04	Very Good	I	SI2
3.04	Premium	I	SI2
3.02	Fair	I	I1

carat	cut	color	clarity
3.01	Premium	I	I1
3.01	Premium	F	I1

Most of the twenty largest datasets are in the lowest clarity category (“I1”), with one being in the second best category (“VVS2”) The top twenty diamonds have colors ranging from the worst, “J”, to best, “D”,categories, though most are in the lower categories “J” and “I”. The top twenty diamonds are more evenly distributed among the cut categories, from “Fair” to “Ideal”, although the worst category (Fair) is the most common.