

120 : gui.plot.box

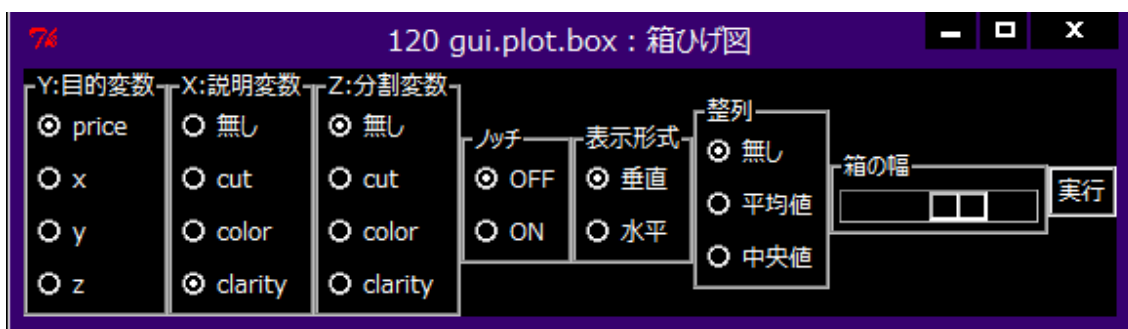
機能 : Tukey の箱ひげ図を描きます。

目的変数 : price ダイヤモンドの価格
説明変数 : cut 研磨
 color 色調
 clarity 透明度
分割変数 : 説明変数と同じ

gui.plot.box の起動

```
outcome <- data.frame(price)
predictor <- data.frame(cut, color, clarity)
f.predictor <- data.frame(cut, color, clarity)
```

120 : gui.plot.box : 箱ひげ図



目的変数 : Y 軸に割り当てる連続量
説明変数 : X 軸に割り当てるファクタ
分割変数 : グループ分けファクタ
ノッチ : ノッチの有無
表示形式 : 垂直 (標準)、水平 (Y と X の入れ替え)
整列 : 無し : 水準の順位、平均値、中央値
箱の幅 : 箱の横幅を調整
実行 : 実行ボタン

目的変数： price (ダイヤモンドの価格)
説明変数： 無し
分割変数： 無し

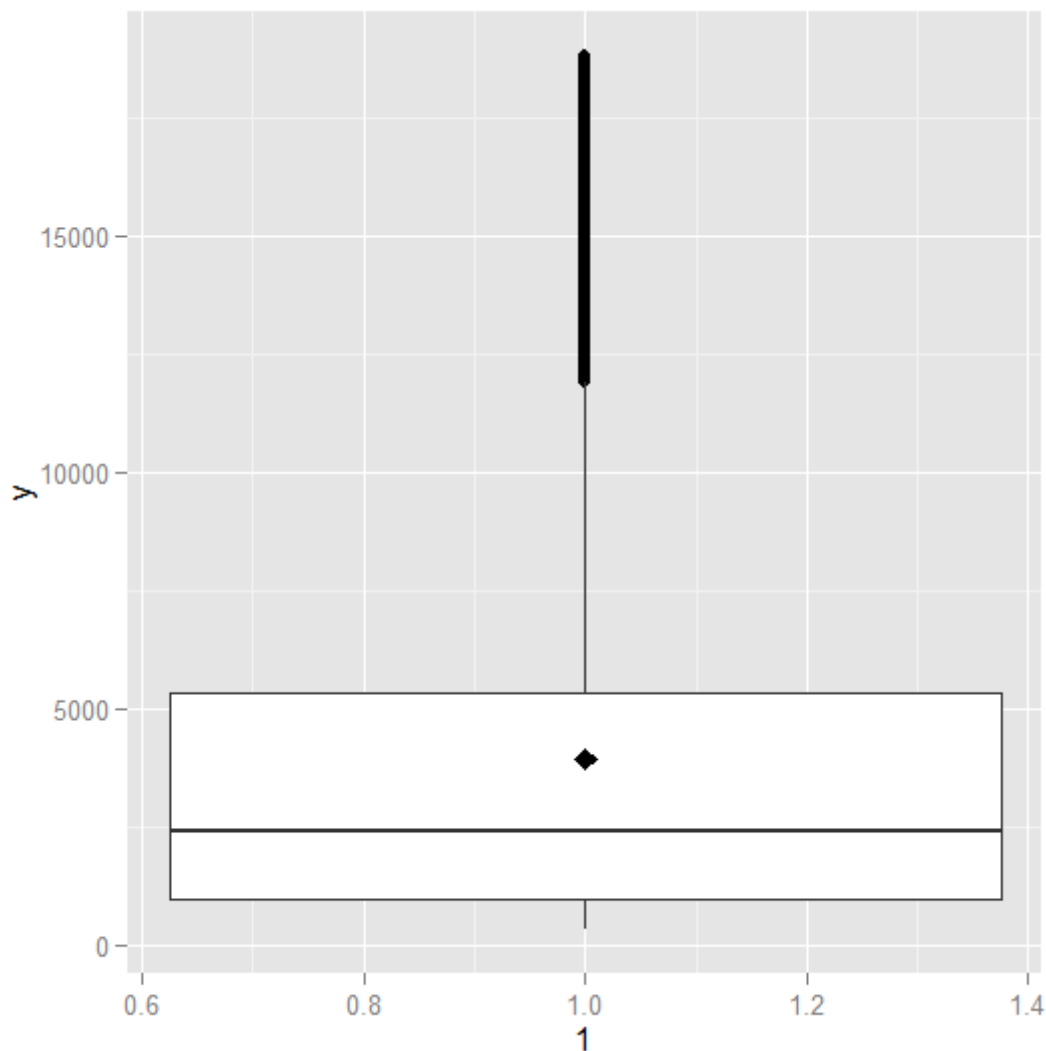


図 1 . Tukey の箱ひげ図の説明 : price (ダイヤモンド価格) の分布をみます

箱中央の線： 中央値 (ダイヤモンド価格の中央値)
箱の中の黒点： 平均値 (ダイヤモンド価格の平均値)
箱の上部の線： 25%分位値 (価格の高い方から 25%点の価格)
箱の下部の線： 75%分位値 (価格の高い方から 75%点の価格)
箱外部の黒丸点： 外れ値

グラフの解釈：

ダイヤモンド価格は、高い方に強く歪んだ分布をしています。その為、平均値は、中央値より高い方へとずれています。外れ値が非常に多いのが特徴として挙げられます。

目的変数： price (ダイヤモンドの価格)
説明変数： cut (研磨：水準、Fair, Good, Verygood, Premium, Ideal)
分割変数： 無し

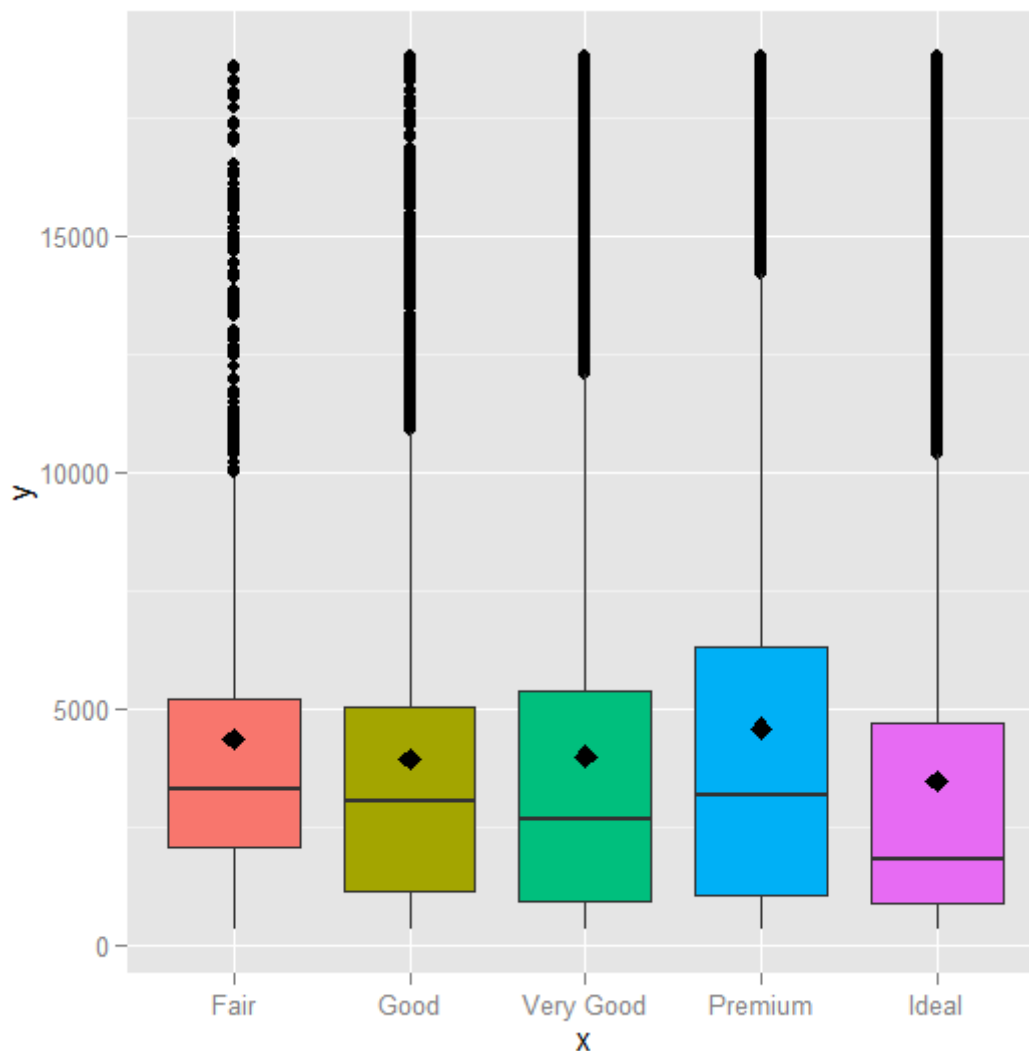


図2. Cut (研磨の水準ごとの箱ひげ図)

いずれも、ダイヤモンドの価格は、高い方に大きく歪んだ分布をしています。平均値、中央値とも大きな差異はありません。研磨が、Fairの群に較べて、Premiumの群では、IQRが大きく、ダイヤモンド価格の分布のばらつきが大きい事を示しています。

目的変数： price (ダイヤモンドの価格)
説明変数： color (色調：水準、D, E, F, G, H, I, J)
分割変数： 無し

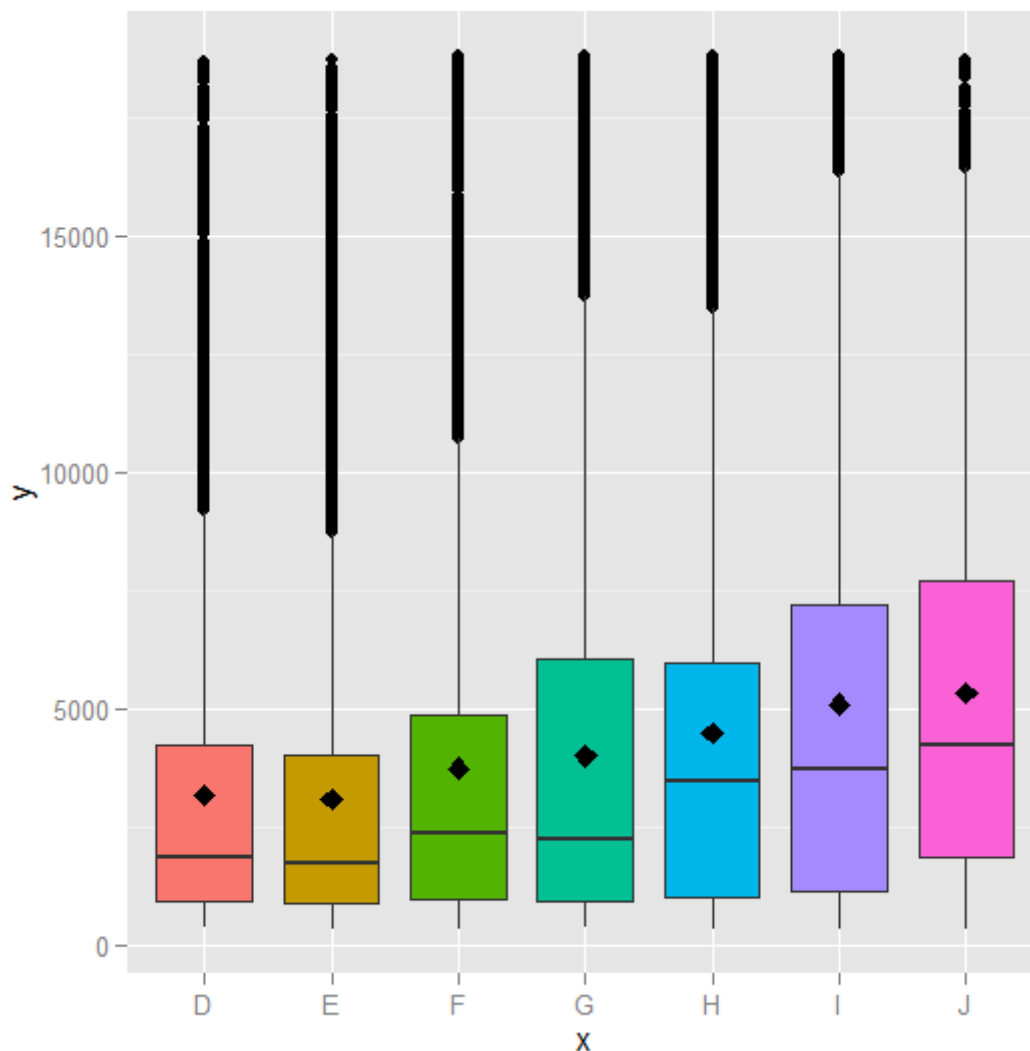


図3. Color (色調の水準ごとの箱ひげ図)

ダイヤモンドの価格が高い方に大きく歪んでいるのは、同じです。平均値は、D から J に水準が変化するとともに、増加（高く）しています。また、中央値も同じ傾向を見せています。

目的変数： price (ダイヤモンドの価格)
説明変数： clarity (透明度)
分割変数： 無し

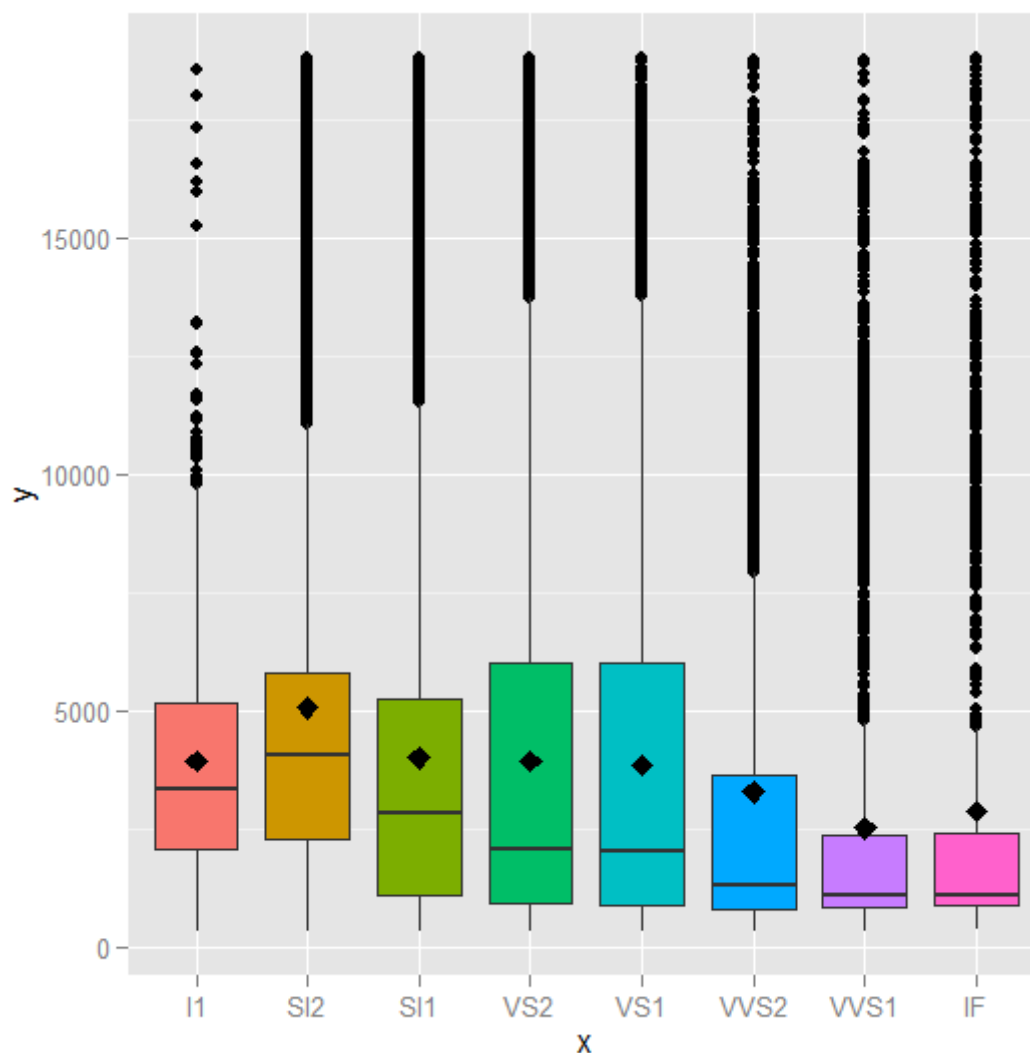


図4. Clarity (透明度水準ごとの箱ひげ図)

I1~VS1 までは、比較的ダイヤモンド価格が高いですが、VVS2~IF では、低くなっています。また、VVS1、IF では、平均値が 25%分位点より大きくなっており、分布の歪みが顕著である事が判ります。

目的変数： price (ダイヤモンドの価格)
説明変数： cut
分割変数： color

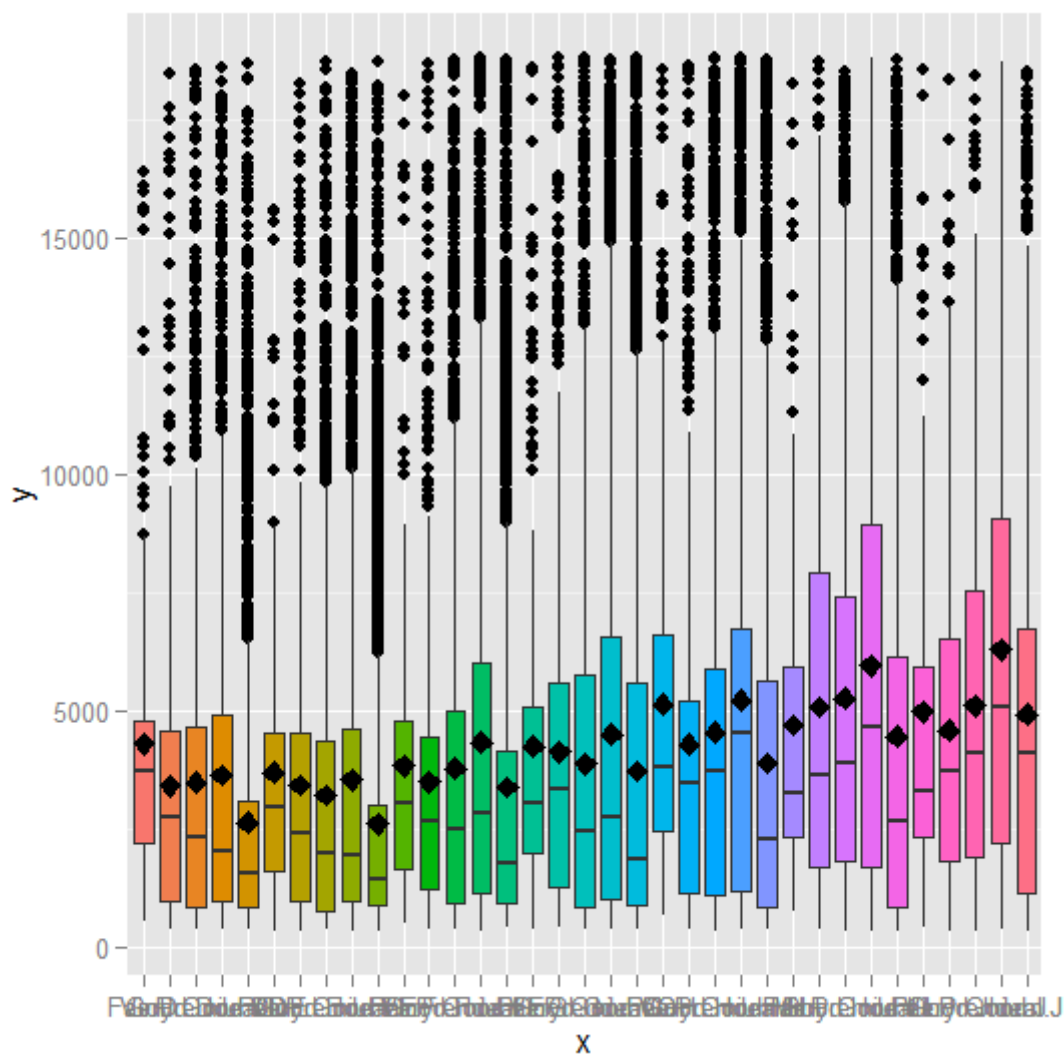


図 5. Cut の水準を color でグループ化した箱ひげ図

X 軸のラベルが重複して判読できませんので、X 軸と Y 軸を入れ替えます。

目的変数： price (ダイヤモンドの価格)

説明変数： cut

分割変数： color

表現形式： 水平

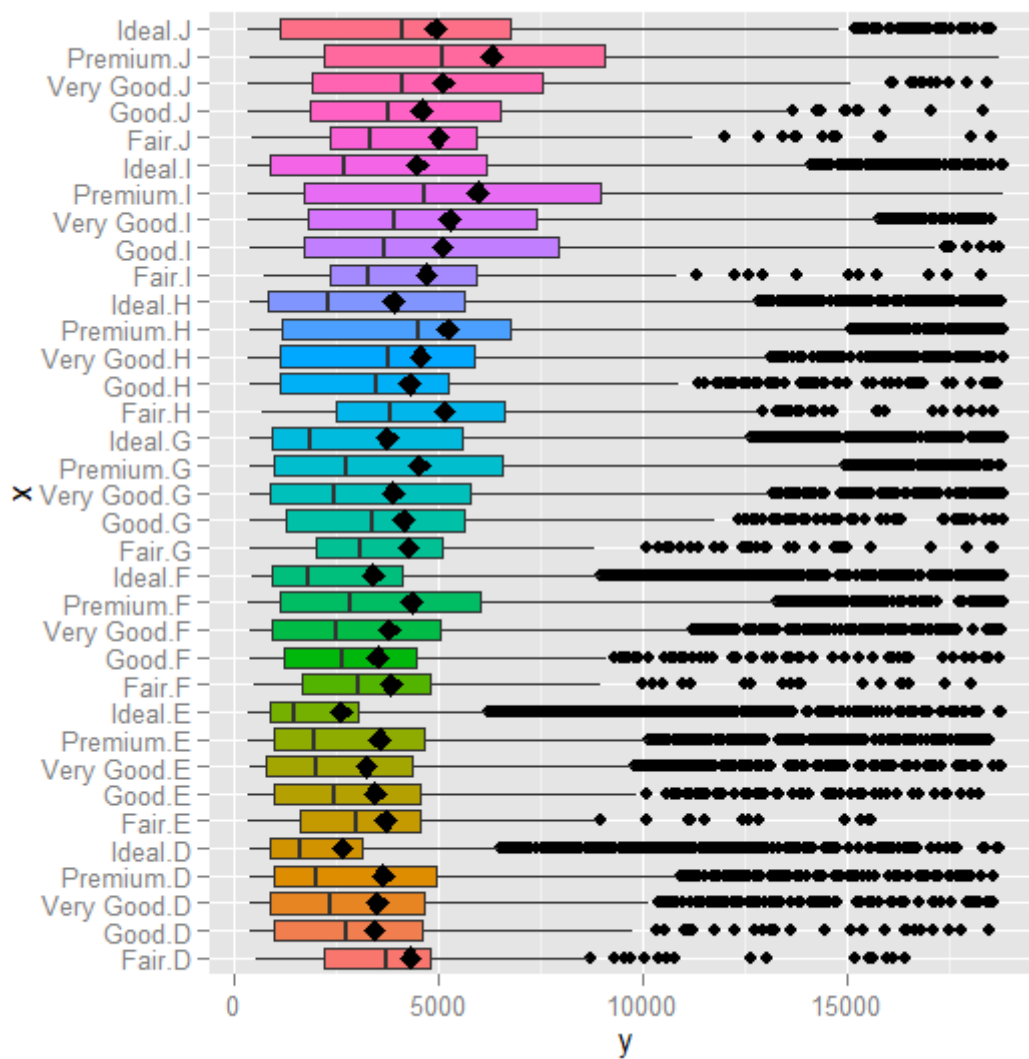


図6. Color (色調) でグループ化された cut 水準ごとの箱ひげ図

Color が、D から J になるほど、ダイヤモンドの価格が大きくなる様に見えますが、はっきりした傾向は窺えません。

目的変数： price (ダイヤモンドの価格)
 説明変数： color
 分割変数： cut
 表現形式： 水平

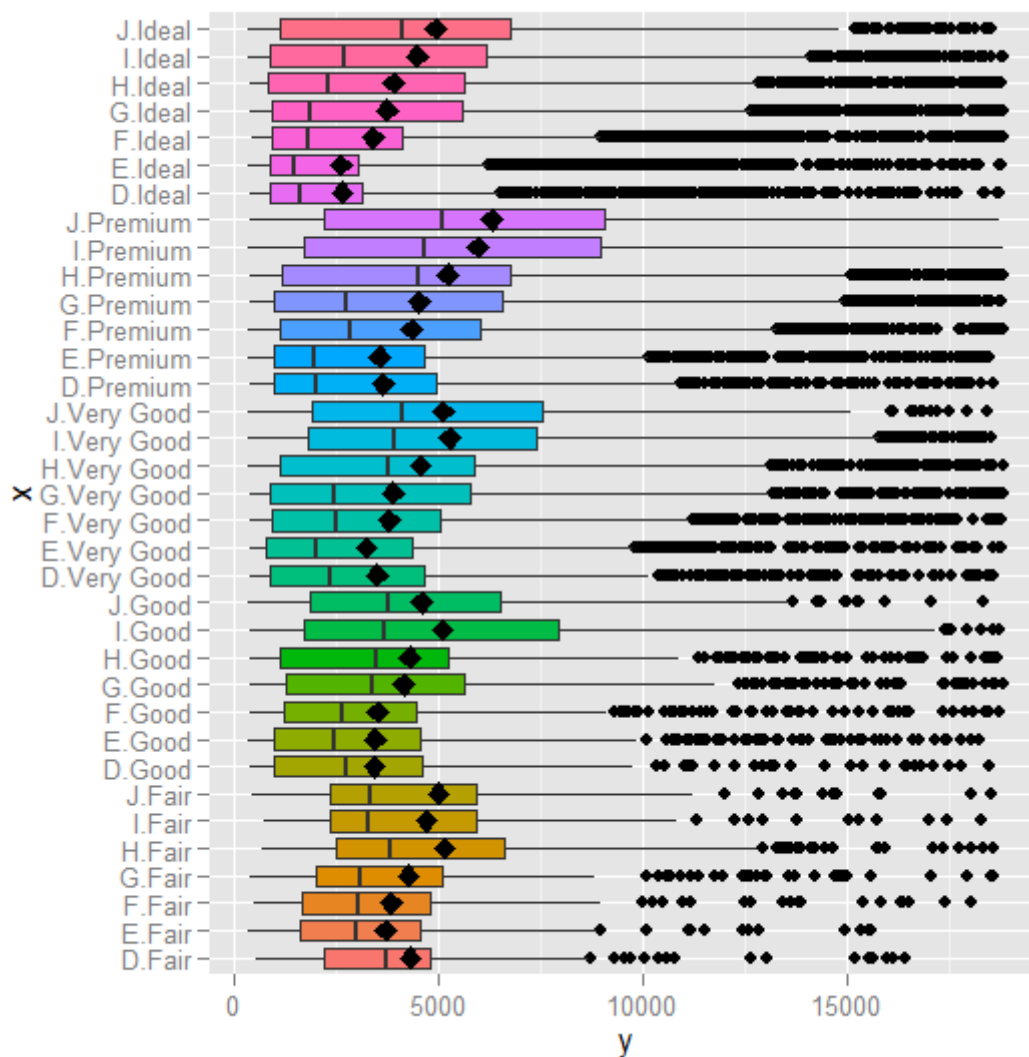


図7. Cut でグループ化した color 水準とダイヤモンド価格の箱ひげ図

前図の cut と color を入れ替えた箱ひげ図です。Cut が同じであれば、Color が、D から J に変化するに従って、ダイヤモンド価格が大きくなる傾向が判ります。特に、cut が、Premium あるいは、Ideal である場合、この傾向は顕著になっています。

目的変数： price (ダイヤモンドの価格)
 説明変数： color
 分割変数： cut
 表現形式： 水平
 整列： 平均値

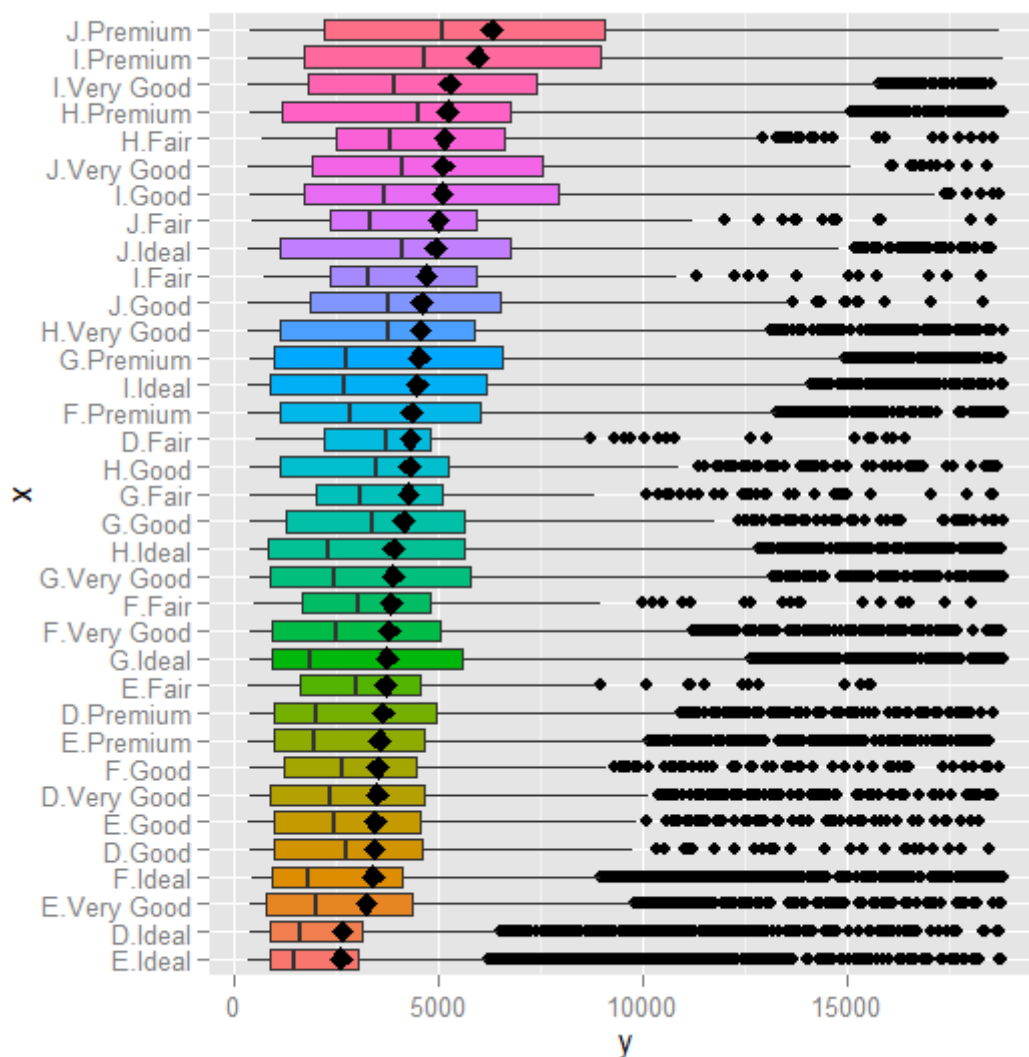


図8. Cut でグループ化した color 水準とダイヤモンド価格の平均値で整列した箱ひげ図

ダイヤモンド価格の平均値によって整列させた箱ひげ図です。Cut が Premium で、color が、H, I, J のダイヤモンドの平均価格が高くなっています。逆に、cut が Ideal で、color が D, E のダイヤモンドは平均価格が小さくなっています。

目的変数： price (ダイヤモンドの価格)
 説明変数： color
 分割変数： cut
 表現形式： 水平
 整列： 中央値

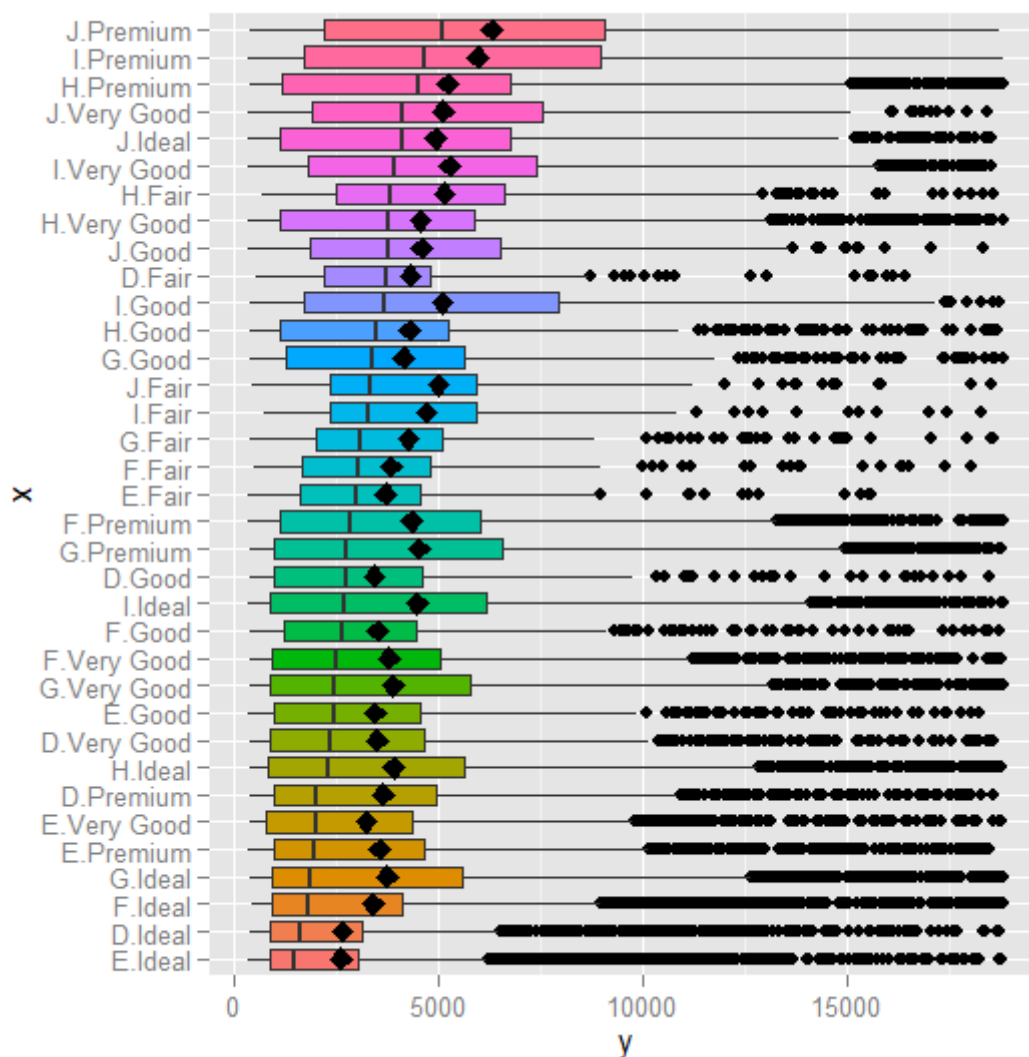


図9. Cut でグループ化した color 水準とダイヤモンド価格の中央値で整列した箱ひげ図

ダイヤモンド価格の中央値で整列した箱ひげ図です。前図と大きな変化はありませんが、かなりの順位の入替えがあります。

目的変数： price (ダイヤモンドの価格)
 説明変数： color
 分割変数： cut
 表現形式： 水平
 整列： 中央値
 ノッチ： ON

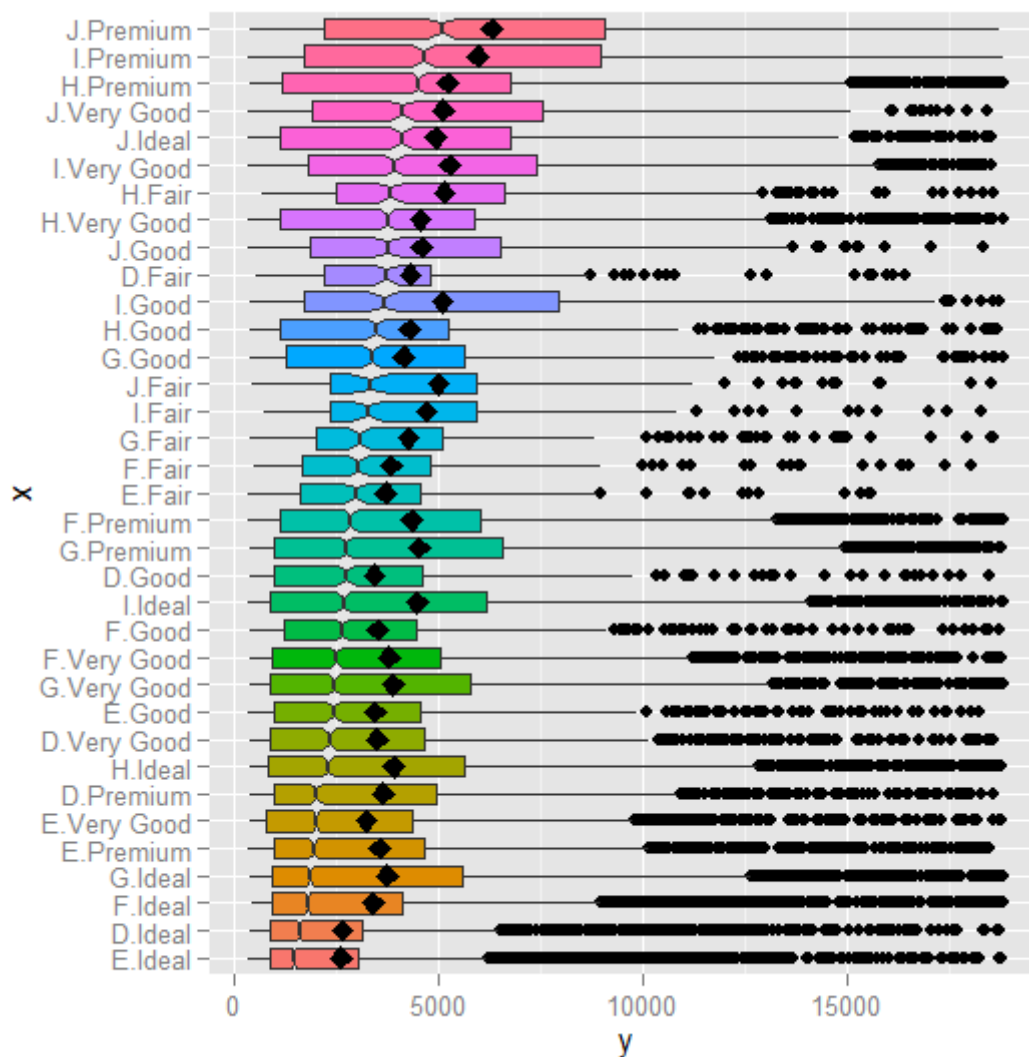


図10. Cut でグループ化した color 水準とダイヤモンド価格の中央値で整列した箱ひげ図
ノッチ ON

前図にノッチを付けたものです。

目的変数： y (詳細不明)
説明変数： color
ノッチ： ON

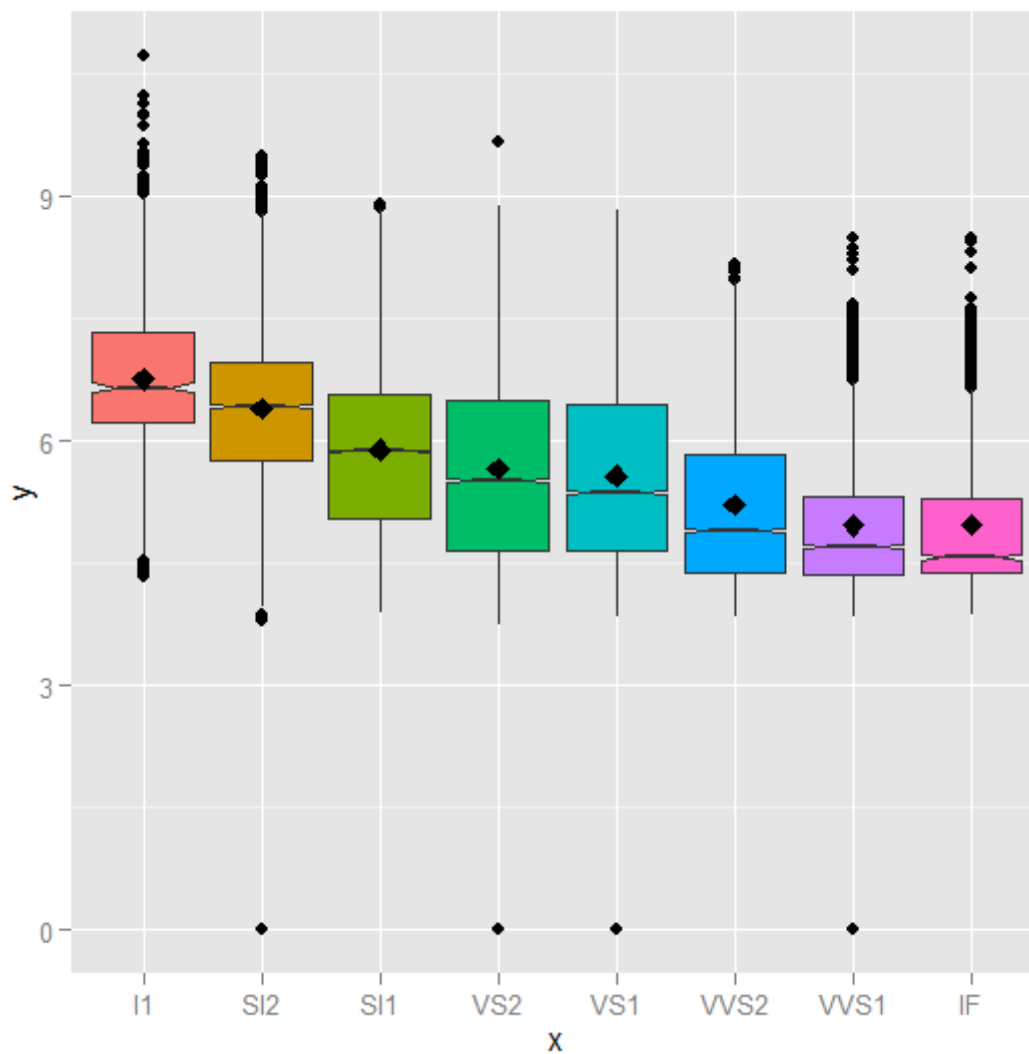


図 1 1. clarity 水準ごとの y (詳細不明) の箱ひげ図

目的変数： y (詳細不明)
説明変数： color
ノッチ： ON

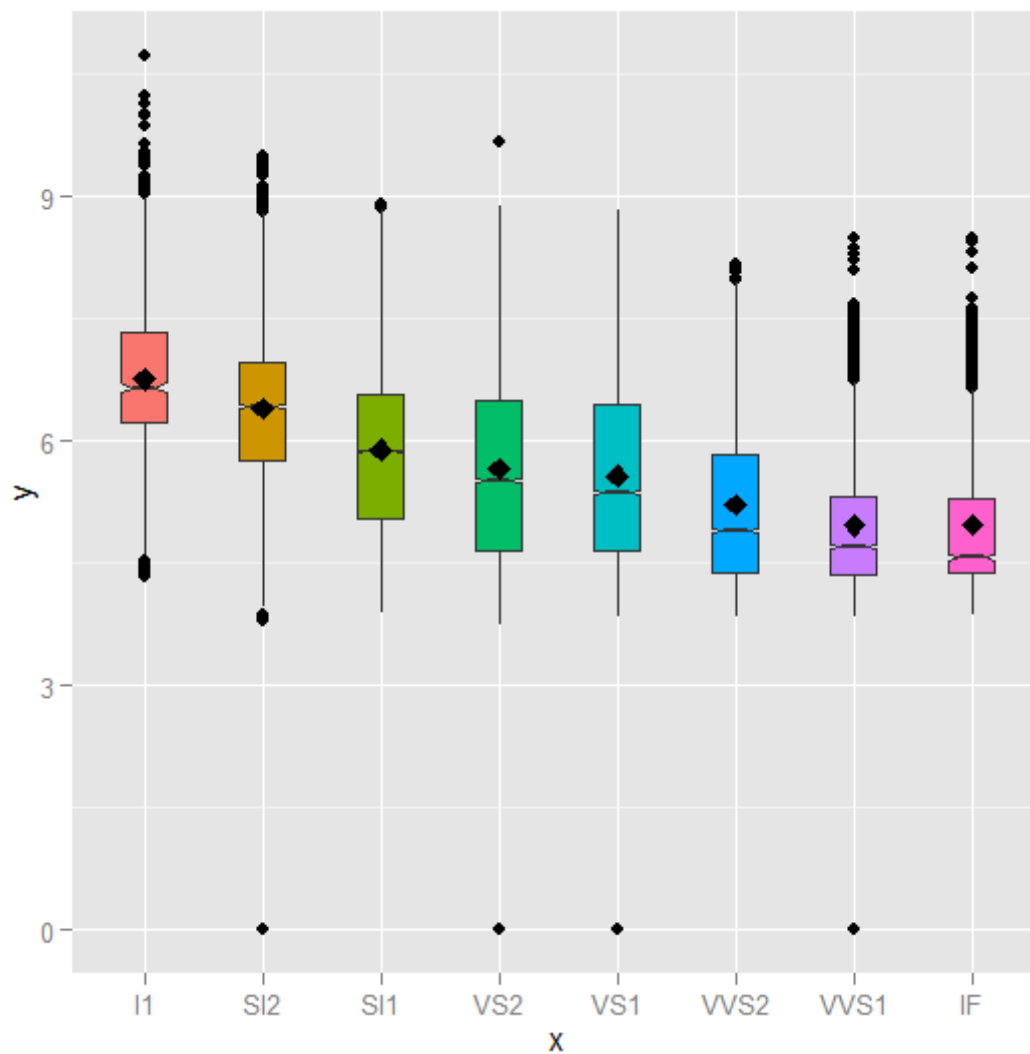


図 1 2. clarity 水準ごとの y (詳細不明) の箱ひげ図

前図と同じものです。スライダーによって、箱ひげ図の横幅を細くしました。