Wei Tang
Dec 26 · 2 min read

# An exploration of Diamond dataset by using R

**What is the Diamond dataset?**

Prices of 50,000 round cut diamonds

**Description**

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

**Usage**

diamonds

**Format**

A data frame with 53940 rows and 10 variables:

price

price in US dollars (\$326–\$18,823)

carat

weight of the diamond (0.2–5.01)

cut

quality of the cut (Fair, Good, Very Good, Premium, Ideal)

color

diamond colour, from J (worst) to D (best)

clarity

a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))

x

length in mm (0–10.74)

y

width in mm (0–58.9)

z

depth in mm (0–31.8)

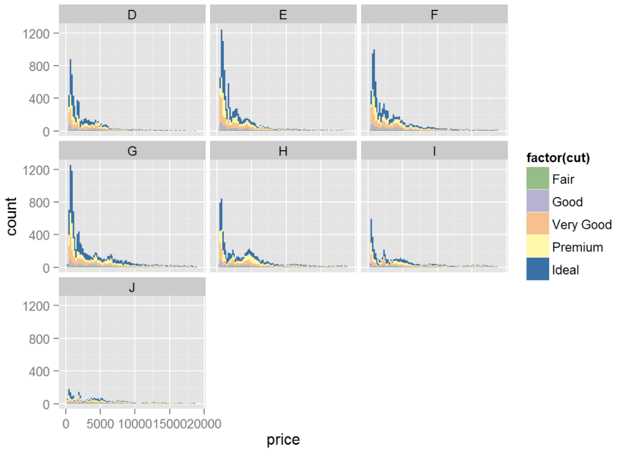
depth

total depth percentage = z / mean(x, y) = 2 * z / (x + y) (43–79)

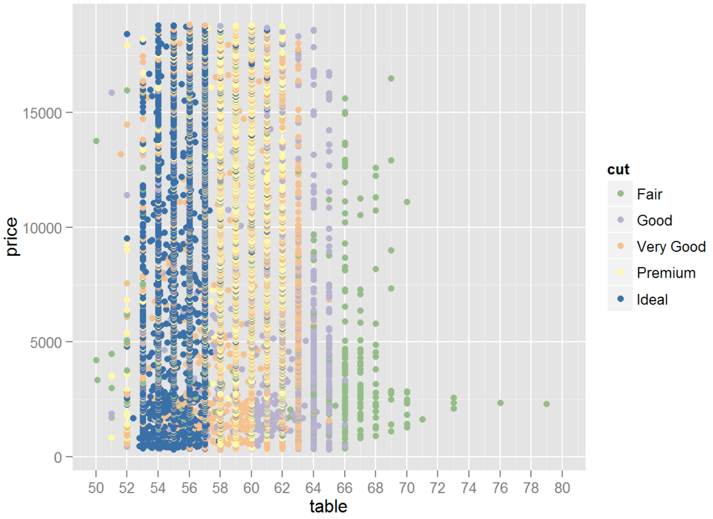
table

width of top of diamond relative to widest point (43–95)

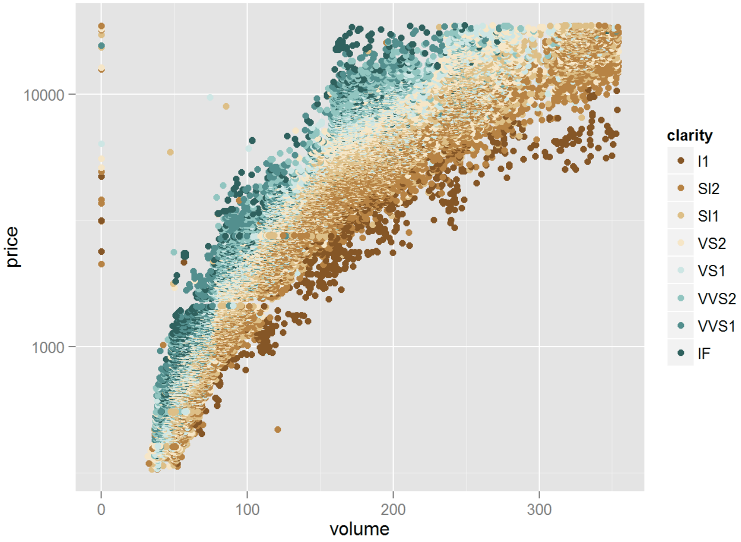
## Price Histograms with Facet and Color



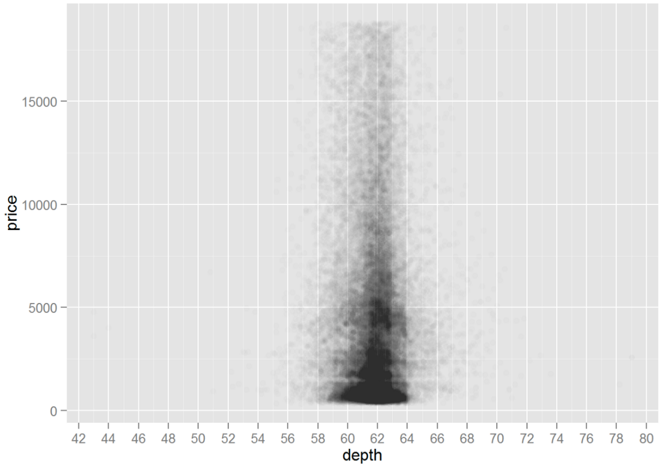
## Price vs. Table Colored by Cut



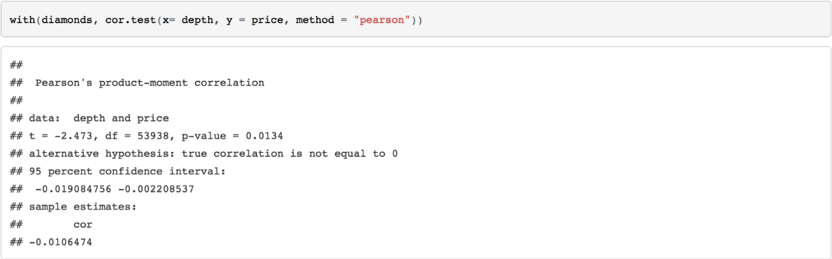
## Price vs. Volume and Diamond Clarity



## price vs. depth



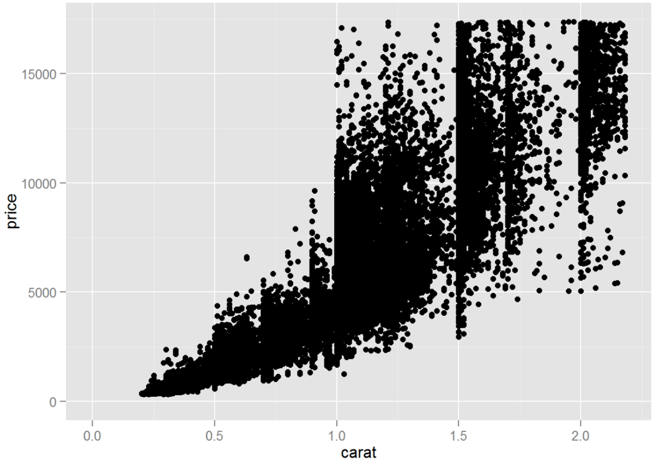
## Correlation — price and depth



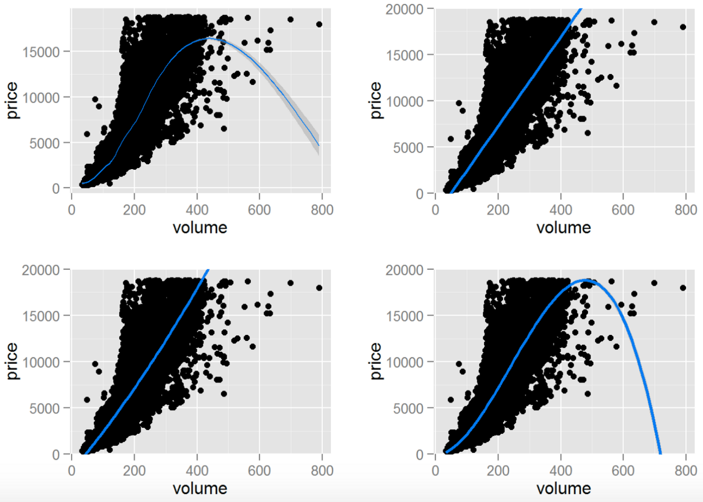
```
with(diamonds, cor.test(x= depth, y = price, method = "pearson"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  depth and price
## t = -2.473, df = 53938, p-value = 0.0134
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.019084756 -0.002208537
## sample estimates:
##        cor
## -0.0106474
```

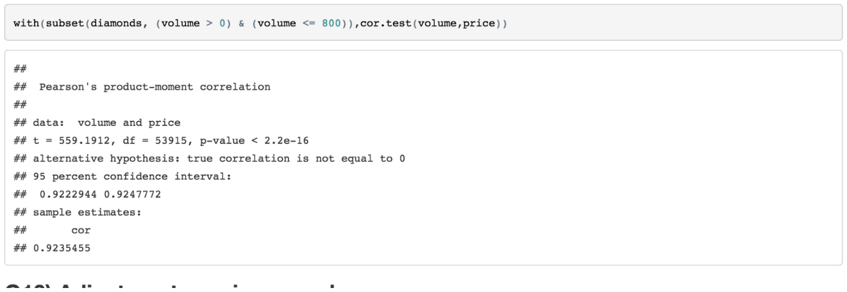
## price vs. carat



# price vs. volume



# Correlation of price and volume



```
with(subset(diamonds, (volume > 0) & (volume <= 800)),cor.test(volume,price))
```

```
##
##  Pearson's product-moment correlation
##
## data:  volume and price
## t = 559.1912, df = 53915, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9222944 0.9247772
## sample estimates:
##       cor
## 0.9235455
```