

diamonds

Chris Parrish

January 19, 2016

diamonds

references:

- Cannon, et al., Stat2, chapter 05, example 5.2

Import the data.

```
data <- read.csv("Diamonds2.csv", header=TRUE)
head(data, 4)
```

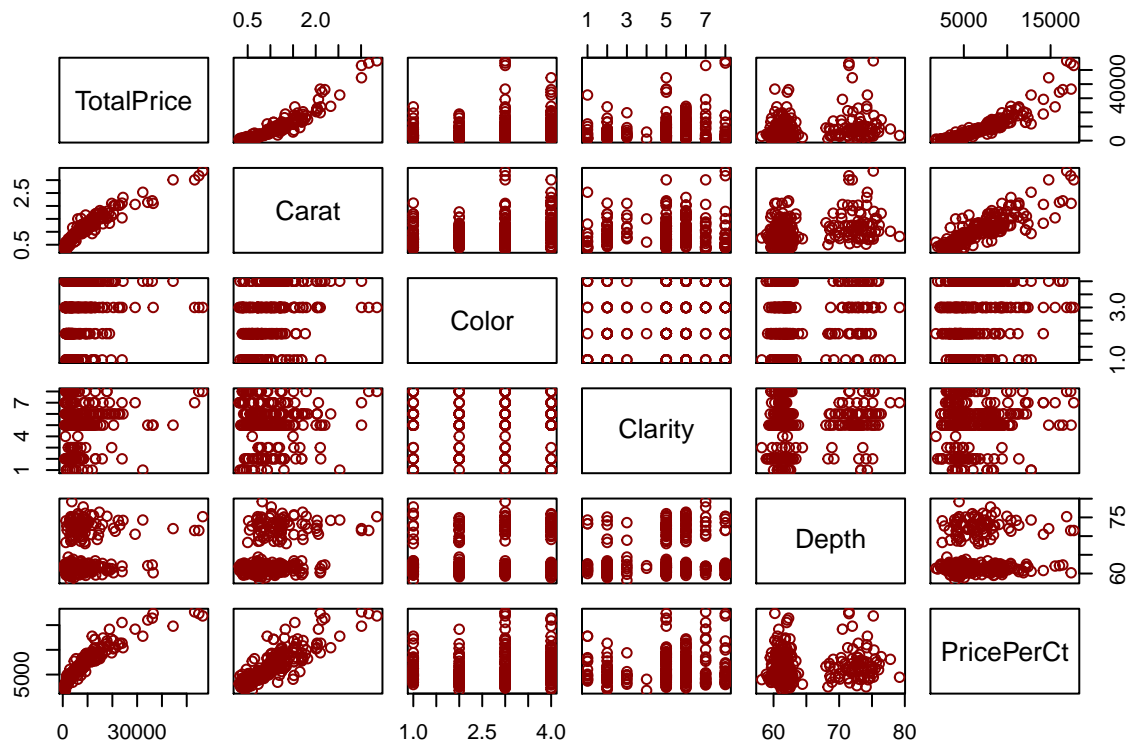
```
##   Carat Color Clarity Depth PricePerCt TotalPrice
## 1  1.08   E    VS1   68.6   6693.3     7228.8
## 2  0.31   F    VVS1  61.9   3159.0     979.3
## 3  0.32   F    VVS1  60.8   3159.0    1010.9
## 4  0.33   D     IF   60.8   4758.8    1570.4
```

```
dim(data)
```

```
## [1] 307  6
```

Scatterplot matrix.

```
pairs(~ TotalPrice + Carat + Color + Clarity + Depth + PricePerCt, data=data,
      col="darkred")
```



Group statistics: Carat ~ Color.

```
n <- with(data, tapply(Carat, Color, length))
mean <- with(data, round(tapply(Carat, Color, mean), 3))
sd <- with(data, round(tapply(Carat, Color, sd), 3))
diamond.statistics <- cbind(n, mean, sd)
diamond.statistics
```

```
##      n mean  sd
## D 52 0.822 0.392
## E 82 0.775 0.287
## F 87 1.057 0.594
## G 86 1.168 0.503
```

```
grand.mean <- cbind(n = length(data$Carat),
                    mean = mean(data$Carat),
                    sd = sd(data$Carat))
rownames(grand.mean) <- c("Total")
grand.mean <- round(grand.mean, 3)
grand.mean
```

```
##           n mean  sd
## Total 307 0.973 0.494
```

First model: ANOVA with aov

```
diamond.aov1 <- aov(Carat ~ Color, data=data)
diamond.aov1
```

```
## Call:
## aov(formula = Carat ~ Color, data = data)
##
## Terms:
##              Color Residuals
## Sum of Squares  8.29937 66.35939
## Deg. of Freedom      3      303
##
## Residual standard error: 0.4679828
## Estimated effects may be unbalanced
```

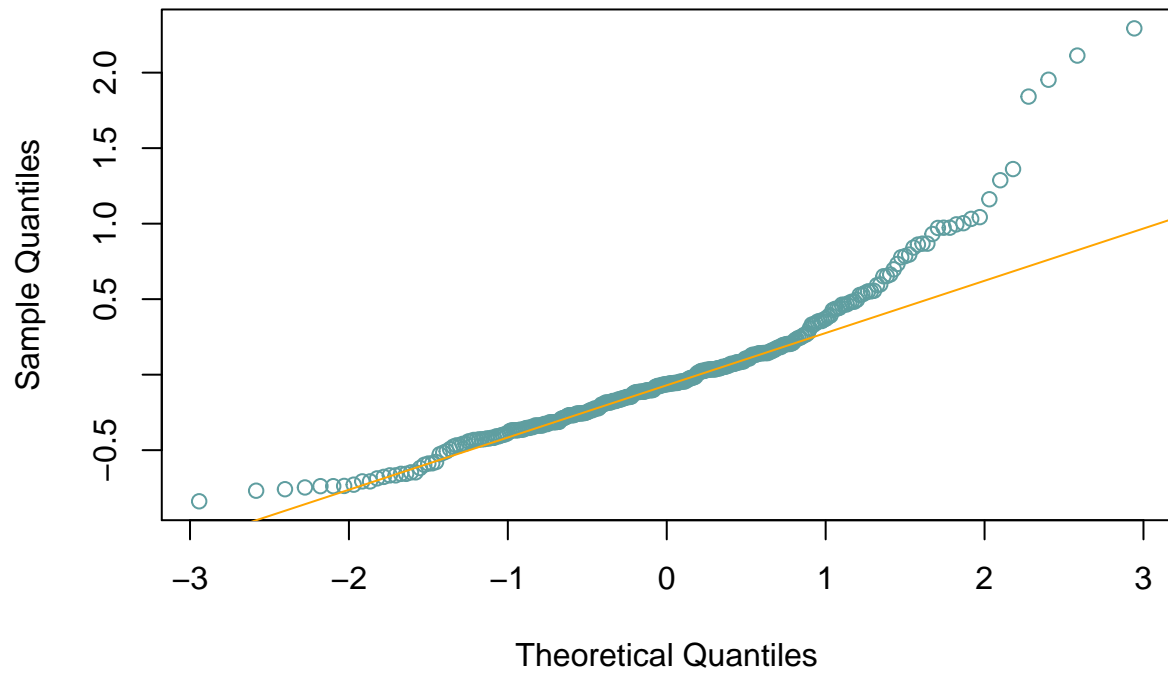
```
options(show.signif.stars=FALSE)
summary(diamond.aov1)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Color          3   8.30   2.767  12.63 8.4e-08
## Residuals    303  66.36   0.219
```

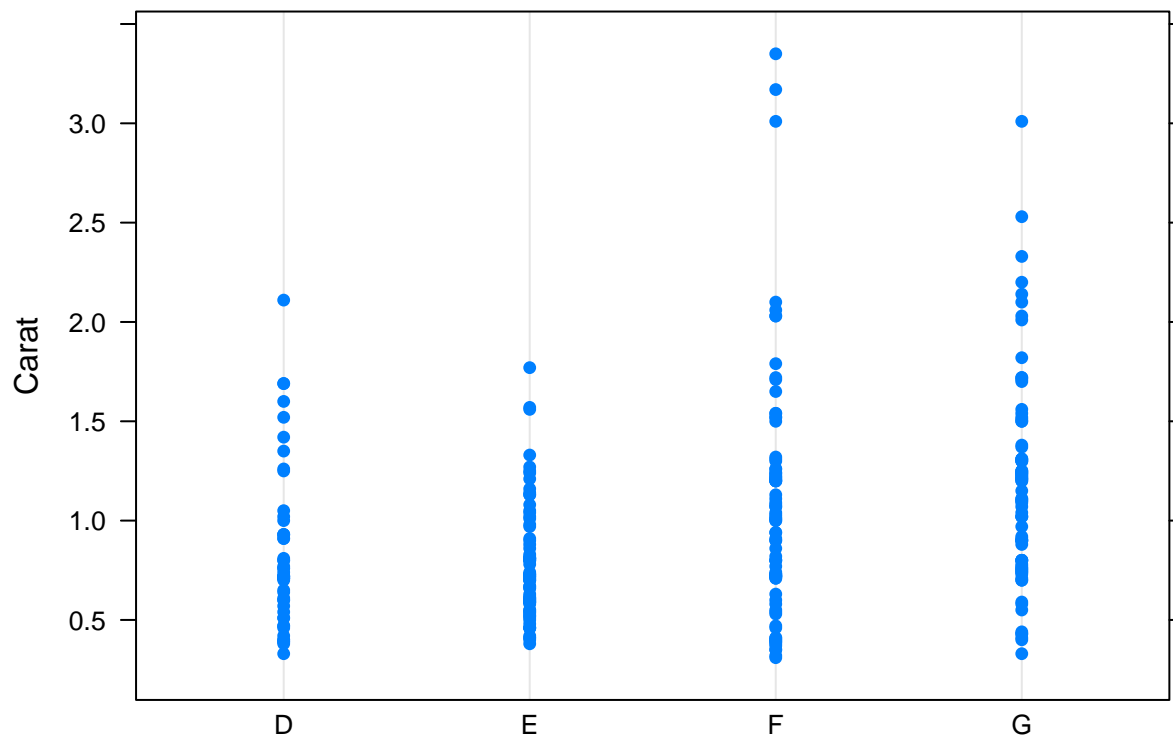
Residuals.

```
qqnorm(resid(diamond.aov1), col="cadetblue")
qqline(resid(diamond.aov1), col="orange")
library(lattice)
```

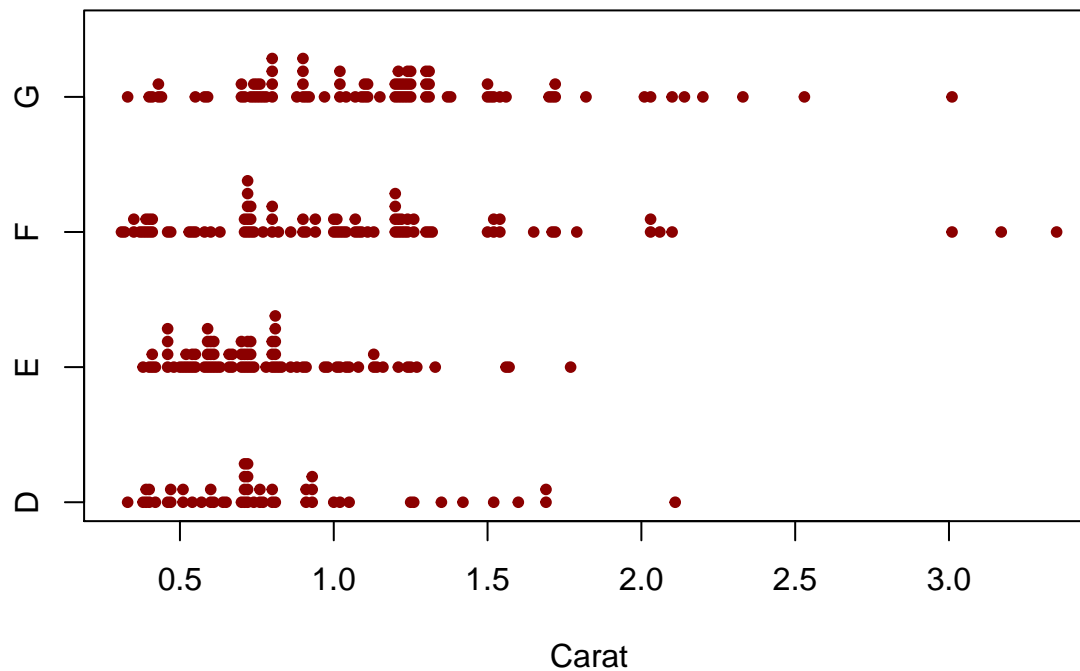
Normal Q-Q Plot



```
dotplot(Carat ~ Color, data=data, jitter=TRUE)
```



```
stripchart(Carat ~ Color, data=data,
           pch=20, col="darkred", method="stack")
```



Second model: ANOVA with aov and log transformation

```
diamond.aov2 <- aov(log(Carat) ~ Color, data=data)
diamond.aov2
```

```
## Call:
##   aov(formula = log(Carat) ~ Color, data = data)
##
## Terms:
##              Color Residuals
## Sum of Squares  7.61754  60.38248
## Deg. of Freedom    3      303
##
## Residual standard error: 0.4464102
## Estimated effects may be unbalanced
```

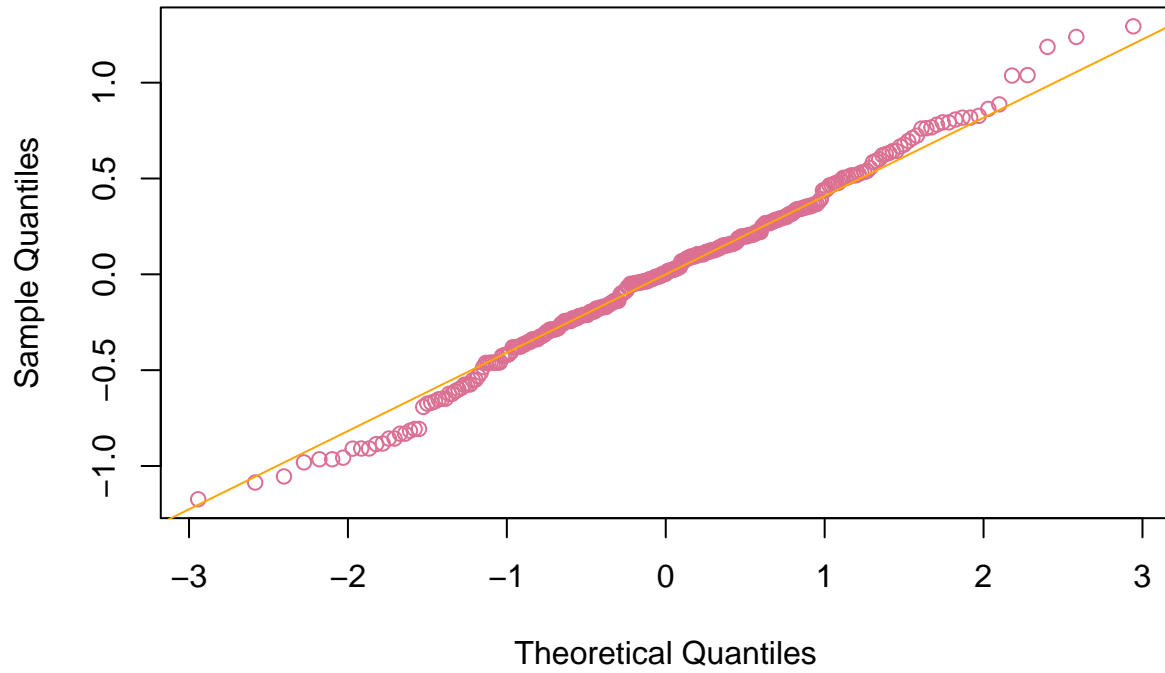
```
summary(diamond.aov2)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Color          3   7.62  2.5392  12.74 7.28e-08
## Residuals    303  60.38  0.1993
```

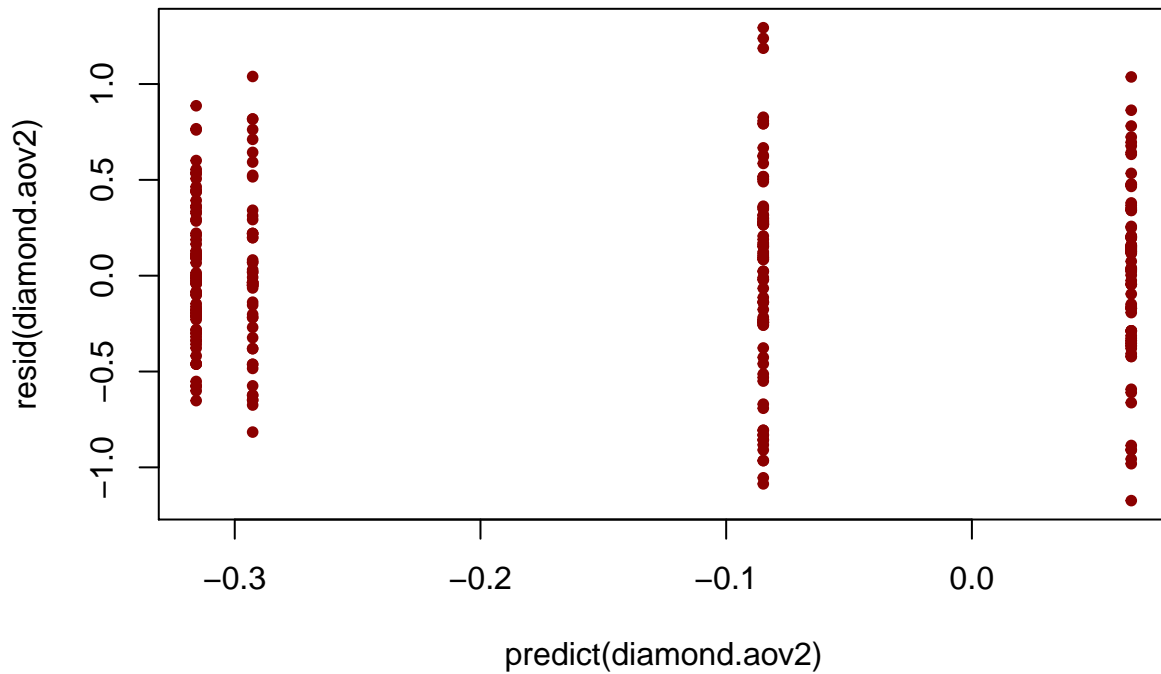
Residuals.

```
qqnorm(resid(diamond.aov2), col="palevioletred")
qqline(resid(diamond.aov2), col="orange")
```

Normal Q-Q Plot



```
plot(predict(diamond.aov2), resid(diamond.aov2),  
     pch=20, col="darkred")
```



```
std.dev <- diamond.statistics[, 3]  
std.dev
```

```
##      D      E      F      G
```

```
## 0.392 0.287 0.594 0.503
```

```
ratio <- max(std.dev) / min(std.dev)  
ratio
```

```
## [1] 2.069686
```