Diamonds Data Set:

This dataset has 53,940 data points with 10 variables where price is the response varibale

| Variable | Variable Description | Variable Type | Range of values |
| --- | --- | --- | --- |
| price | In US dollars | Numeric | 326-18,823 |
| carat | Weight of diamond | Numeric | 0.2-5.01 |
| cut | Quality of cut | Ordered Factor | Fair,Good,Very Good,Premium,Idea |
| color | Color of Diamond | Ordered Factor | J,I,H,G,F,E,D |
| clarity | How Clear the diamond is | Ordered Factor | I1, SI2, SI1, VS2, VS1, VVS2, VVS1, IF |
| x | Length in mm | Numeric | 0-10.74 |
| y | Width in mm | Numeric | 0-58.9 |
| Z | Depth in mm | Numeric | 0–31.8 |
| depth | Total Depth percentage = z / mean(x, y)  = 2 \* z / (x + y) | Numeric | 43-79 |
| table | width of top of diamond relative to widest point (43–95) | Numeric | 43–95 |

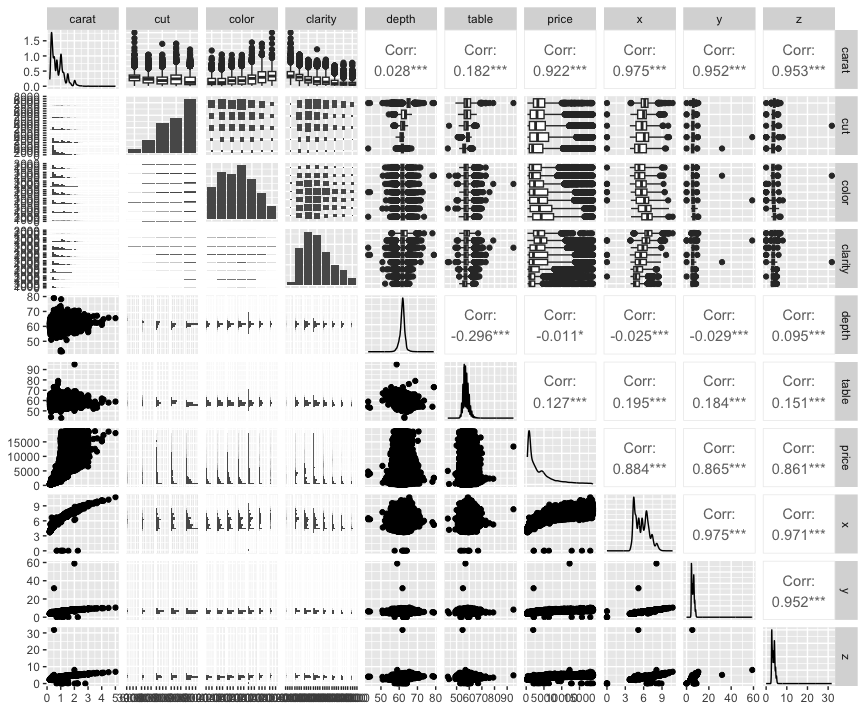
Pair Wise Association between all the variables.

Correlation Matrix between all the numeric variables

**carat depth table price x y z**

**carat 1.00000000 0.02822431 0.1816175 0.9215913 0.97509423 0.95172220 0.95338738**

**depth 0.02822431 1.00000000 -0.2957785 -0.0106474 -0.02528925 -0.02934067 0.09492388**

**table 0.181****61755 -0.29577852 1.0000000 0.1271339 0.19534428 0.18376015 0.15092869**

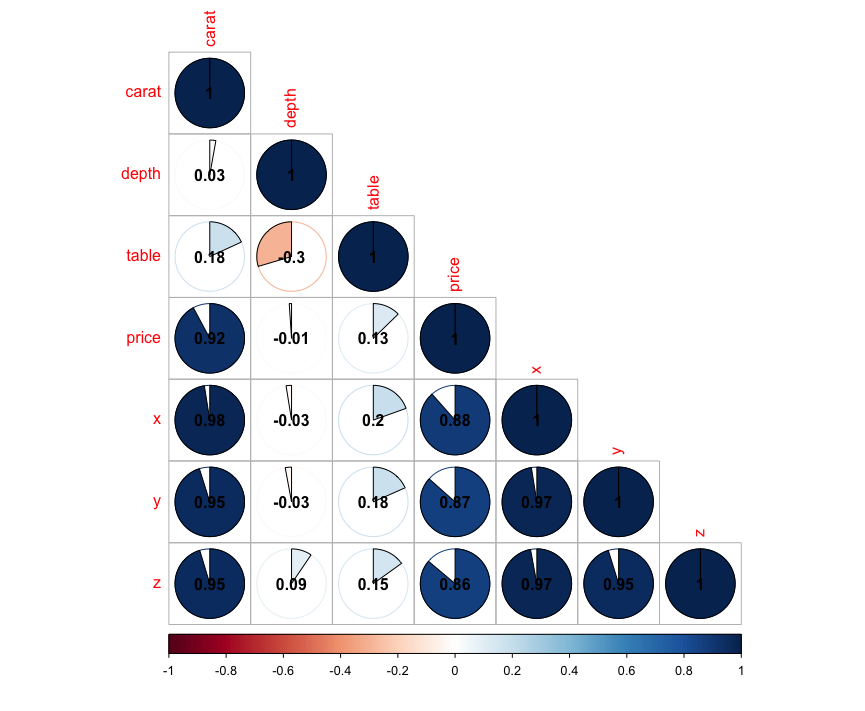
**price 0.92159130 -0.01064740 0.1271339 1.0000000 0.88443516 0.86542090 0.86124944**

**x 0.97509423 -0.02528925 0.1953443 0.8844352 1.00000000 0.97470148 0.97077180**

**y 0.95172220 -0.02934067 0.1837601 0.8654209 0.97470148 1.00000000 0.95200572**

**z 0.95338738 0.09492388 0.1509287 0.8612494 0.97077180 0.95200572 1.00000000**

**Pie Chart describing the correlation between variables**

****

X,Y,Z are highly correlated to carat. So we are using Carat as a predictor for price

where as depth and table are not are good options as a predictor

Predictors are  
carat, price, clarity, cut, color