

lecture 11: big data and multidimensional data

October 30, 2017

Announcements

- Mini grad school fair Wednesday
- Algorithmic Accountability workshop on Otelia Cromwell Day

Pizza will be served!
November 1st, 4:30pm, Seelye 109

Statistical and Data Science Program's
Mini-Graduate School Fair

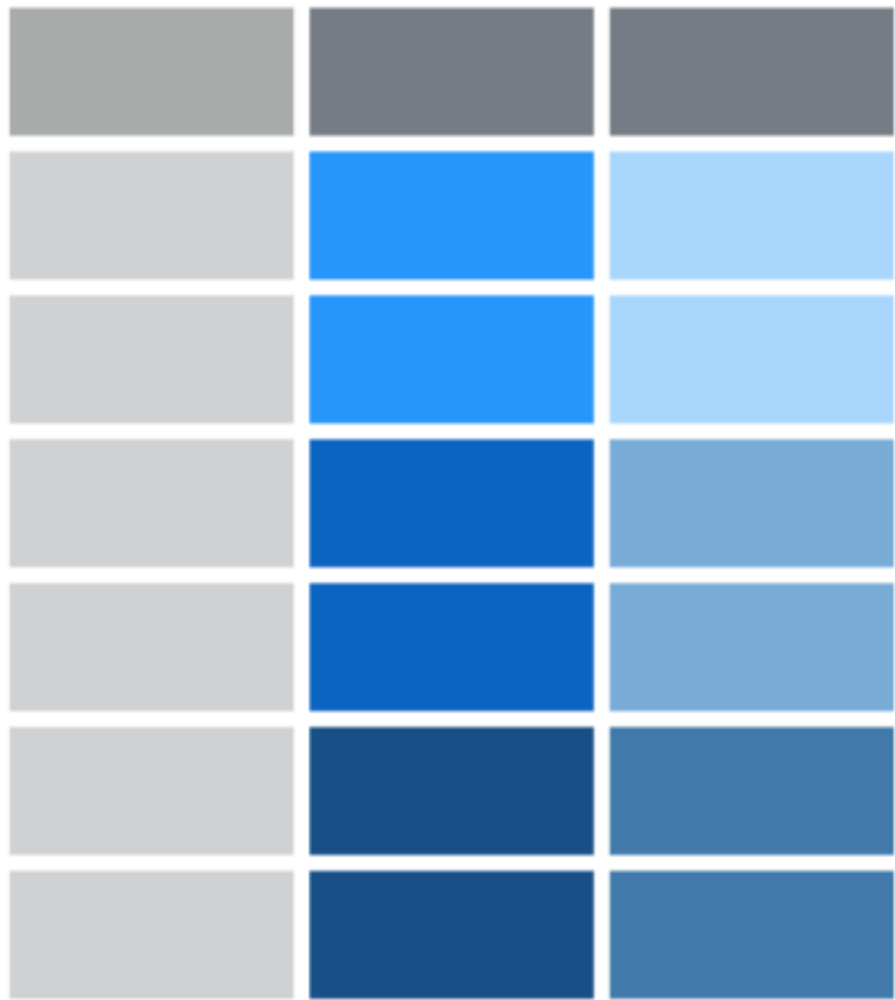
D. Betsy McCoach, PhD University of Connecticut, Measurement, Evaluation, and Assessment	Stephanie Eckman, PhD, Smith '94 University of Maryland, Joint program in Survey Methodology
Jean Wu, PhD Brown University, Biostatistics	Leontine Alkema, PhD UMass Amherst, Biostatistics

Discussion: museum and special collections visits

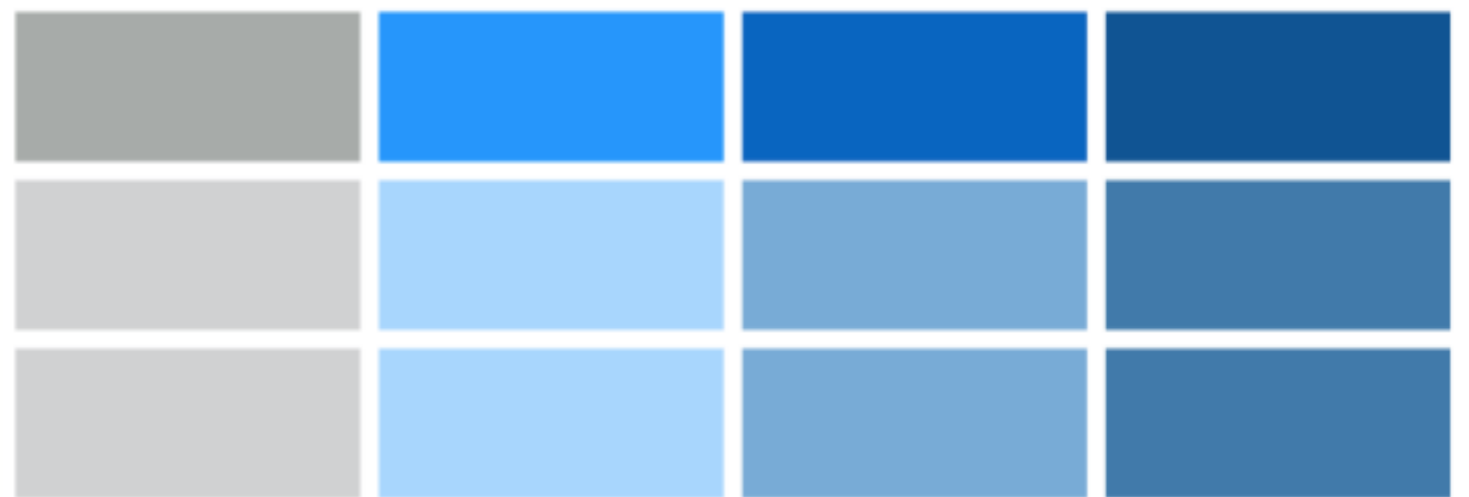
What surprised you?

What was your favorite part?

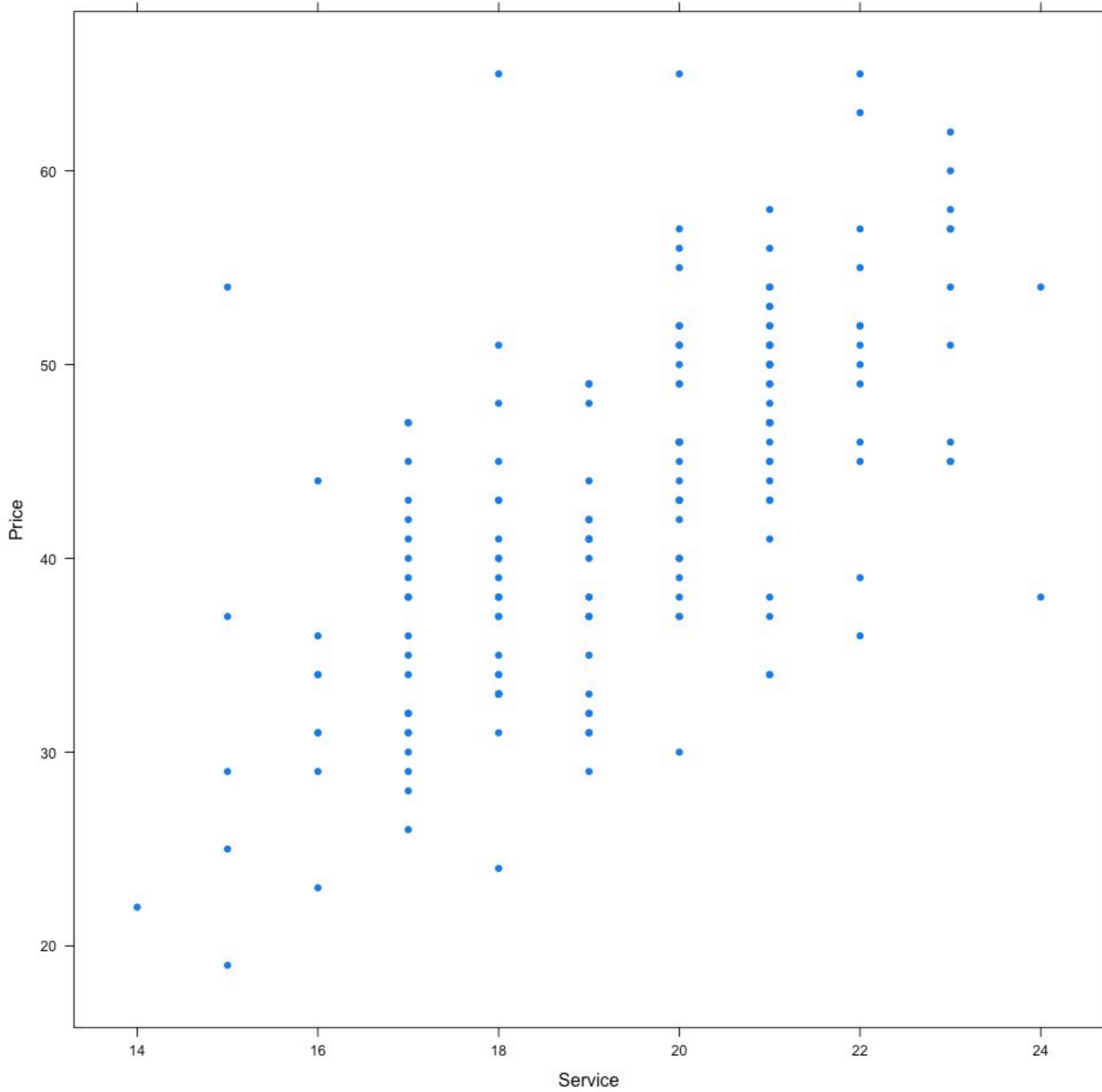
Do you have questions?



Big data is tall,
multidimensional
data is wide

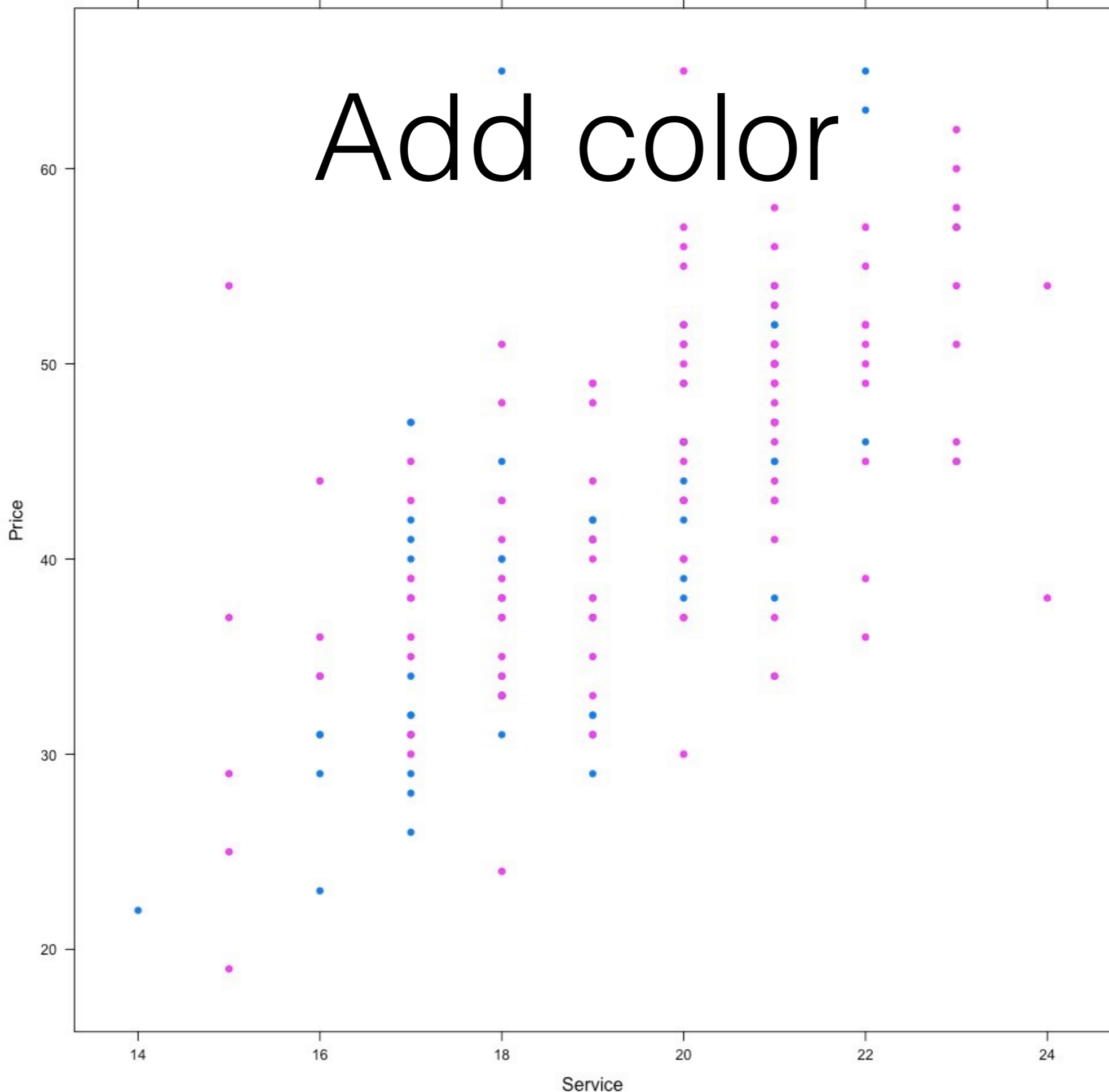


Data can be both!

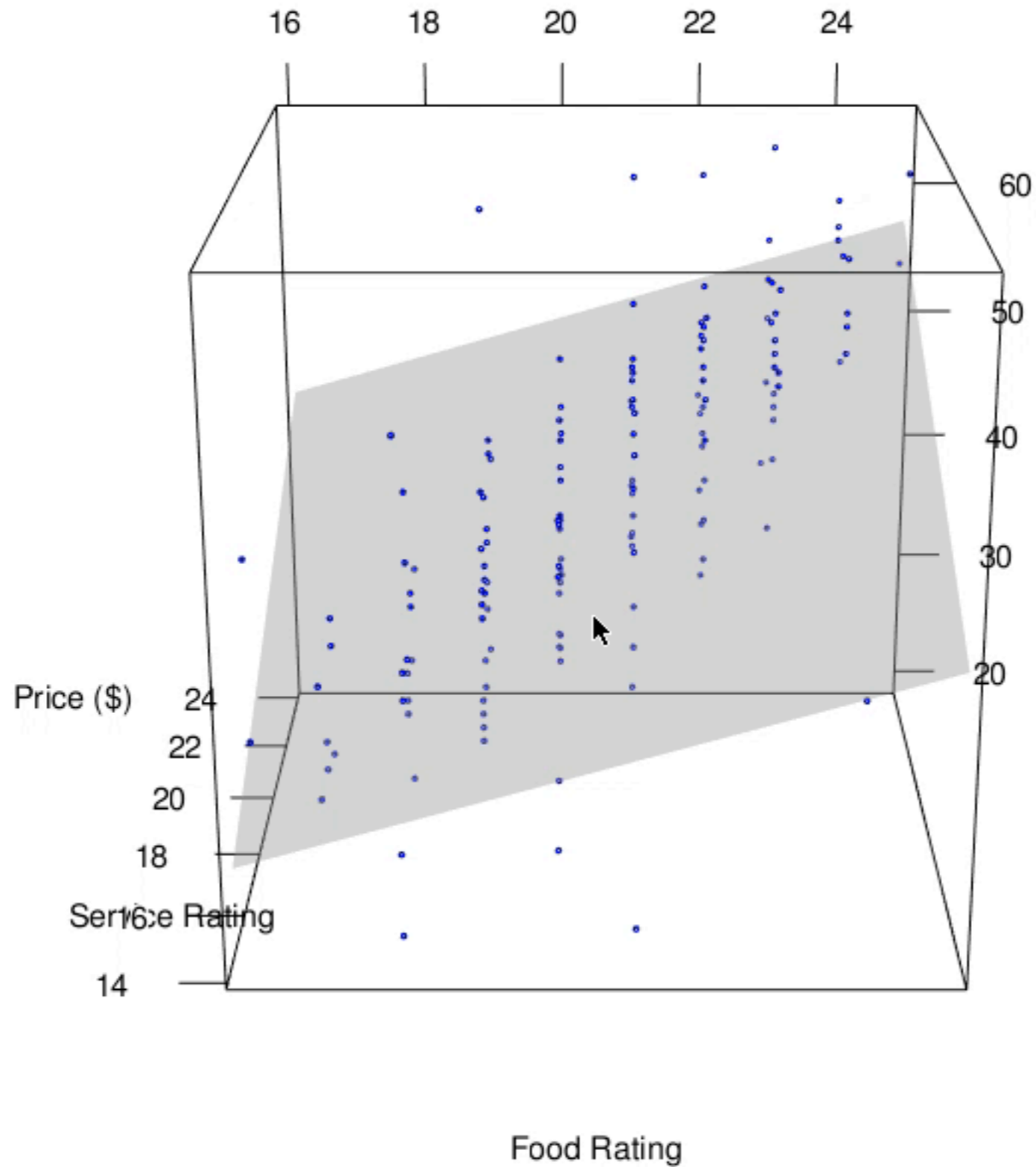


What if you want to
visualize more variables?

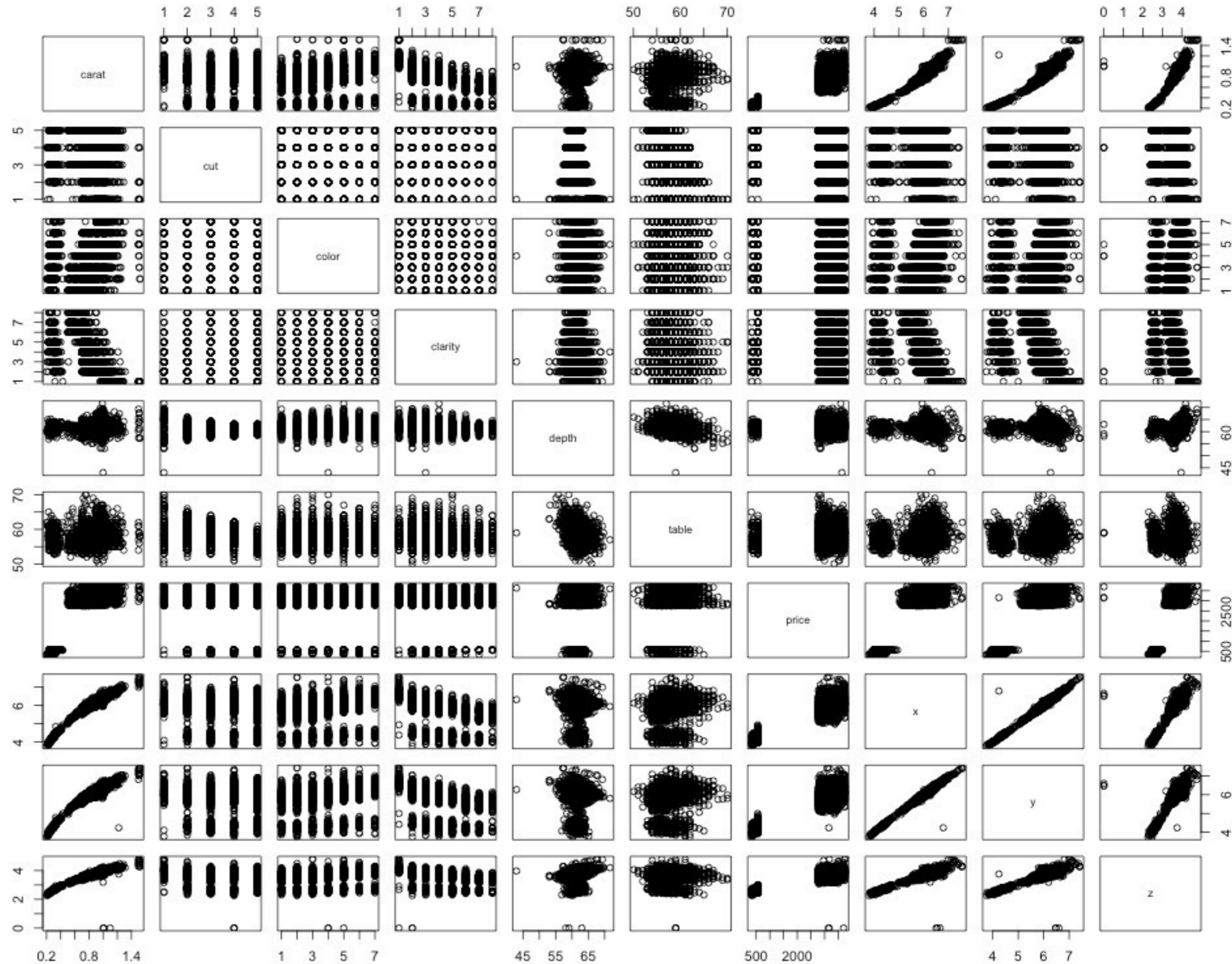
Add color



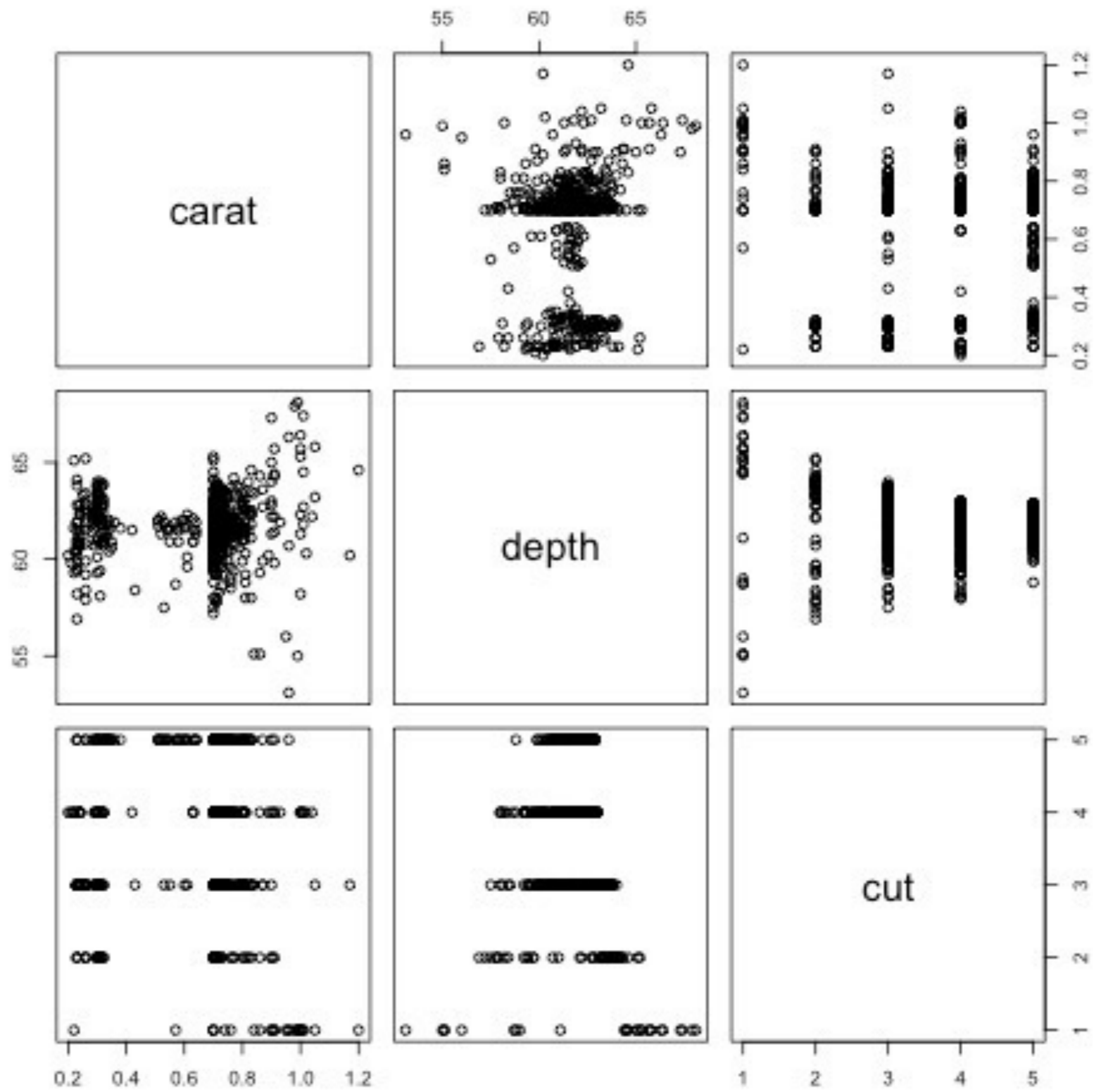
Add a 3rd dimension



Make a scatterplot matrix

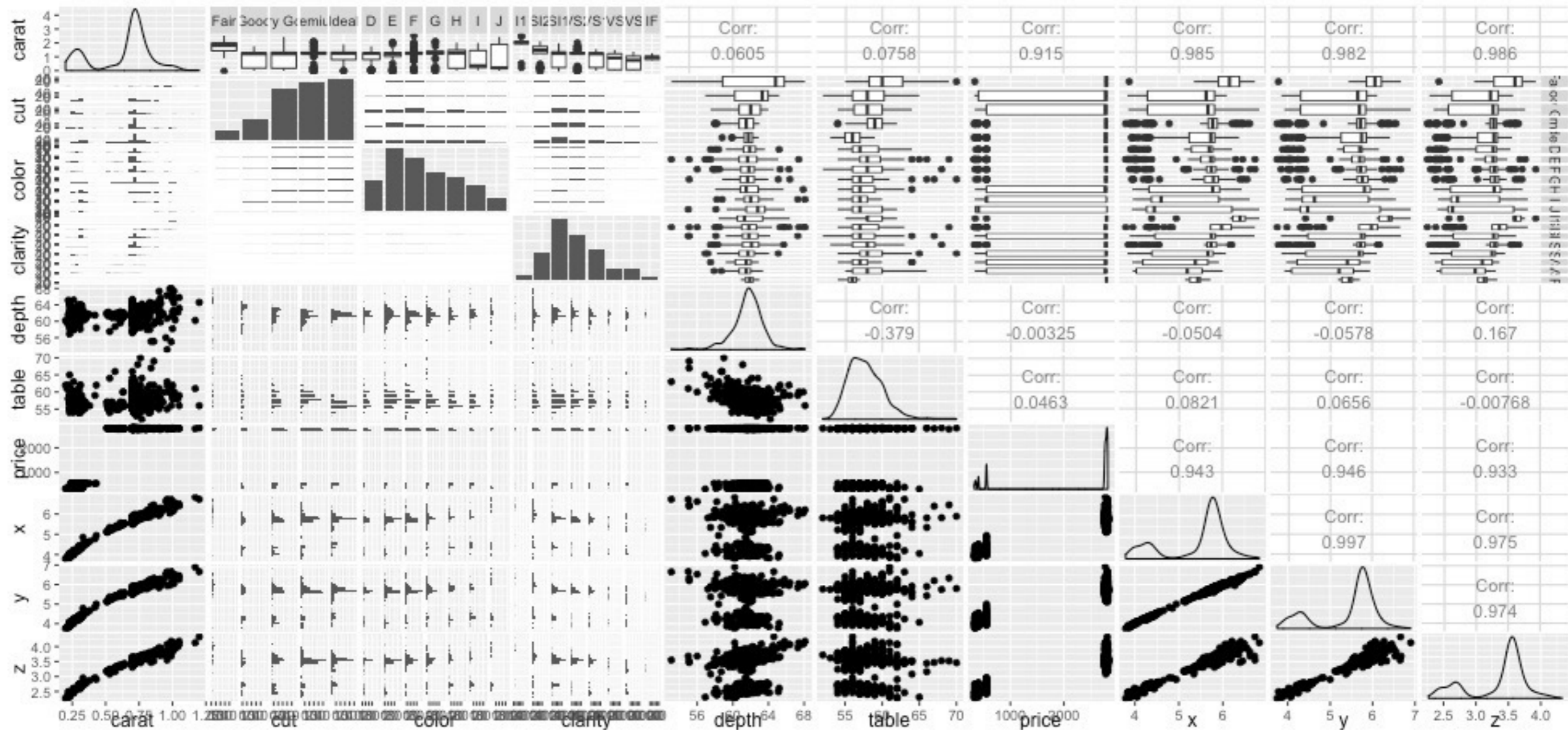


`plot(diamonds)`



```
select(diamonds, carat, depth, cut) %>%  
plot()
```

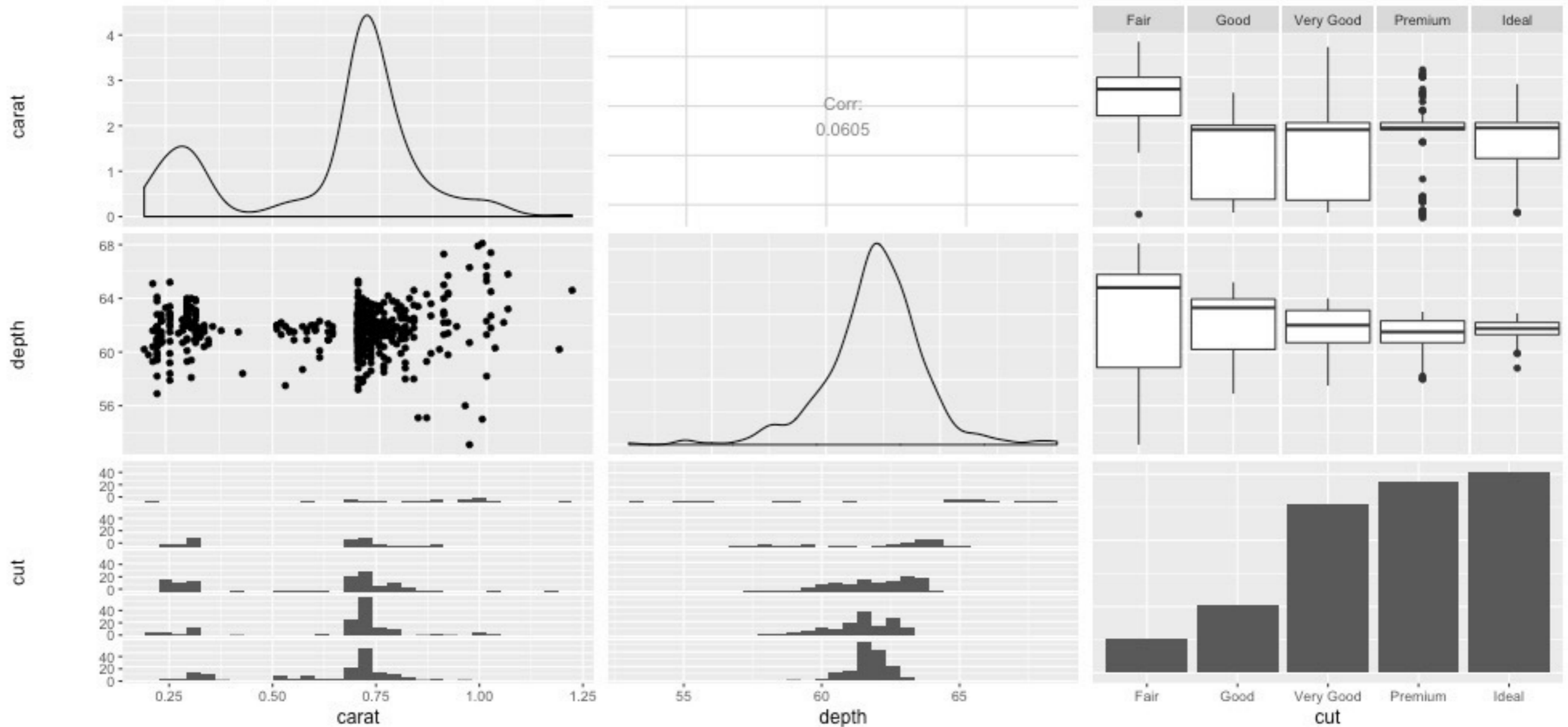
Generalized pairs plot



```
library(GGally)
ggpairs(diamonds)
```

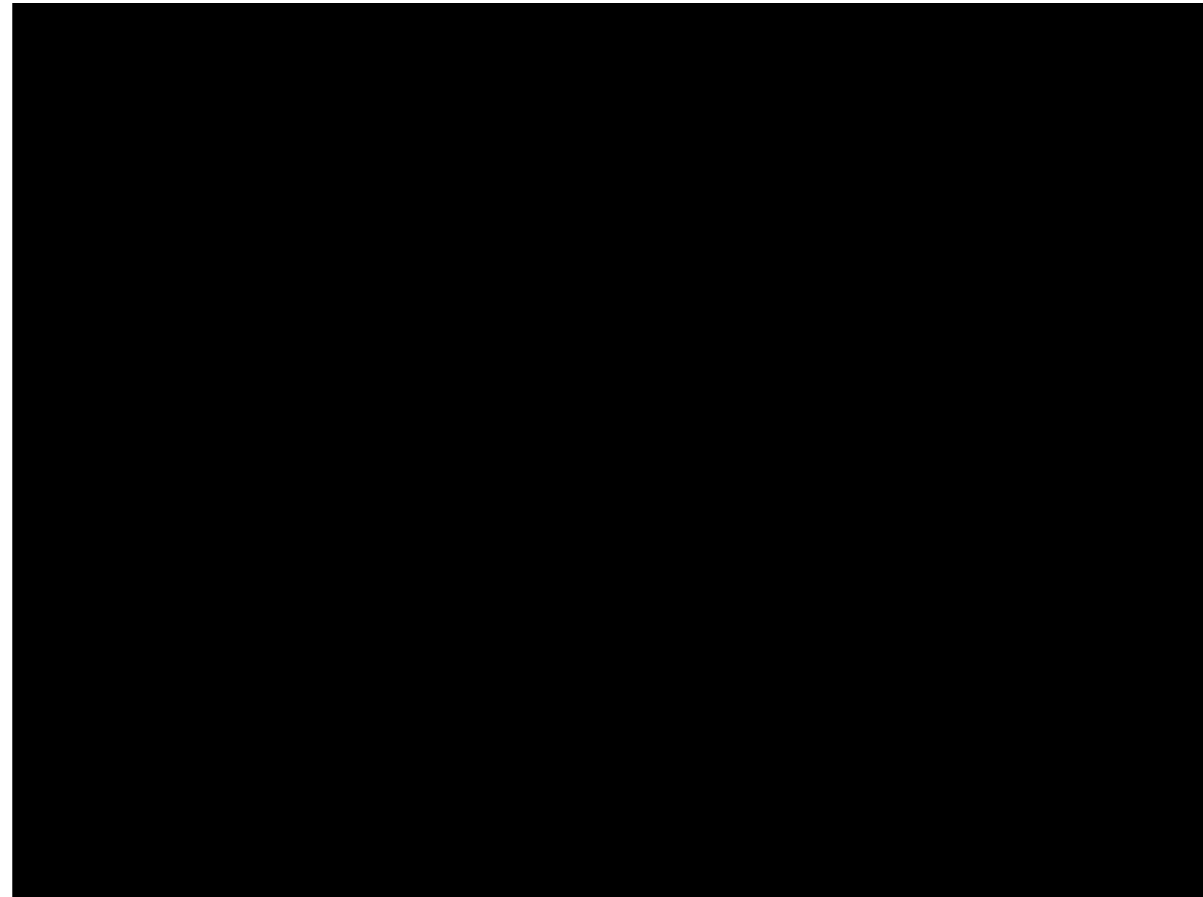
Emerson, J. W., Green, W. A., Schloerke, B., Crowley, J., Cook, D., Hofmann, H., and Wickham, H. (2013). The generalized pairs plot. *Journal of Computational and Graphical Statistics*, 22(1):79–91. <http://bit.ly/gpairs>

Generalized pairs plot

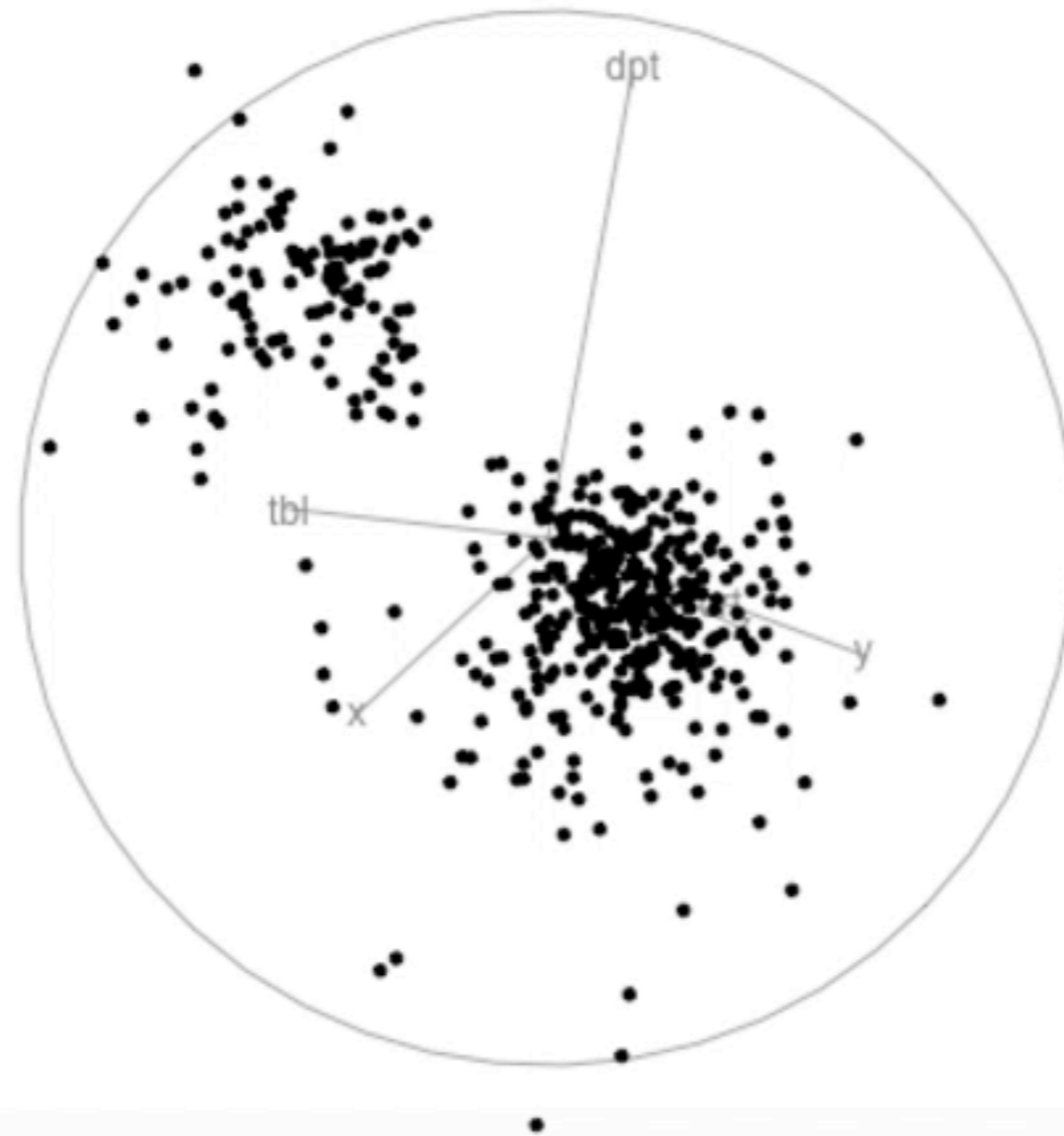


```
select(diamonds, carat, depth, cut) %>%  
ggpairs()
```

prim9



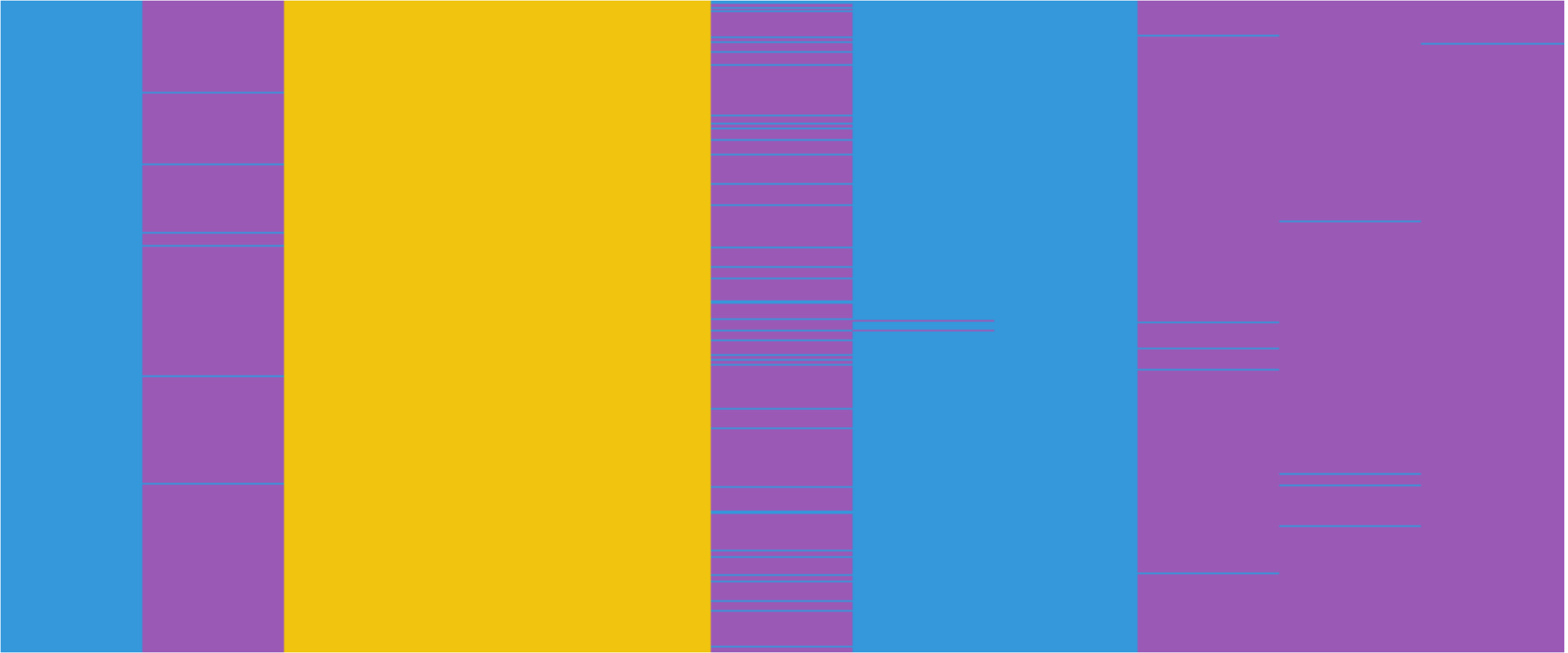
The Grand Tour



```
library(tourr)
select(diamonds, carat, depth, table, x, y, z) %>%
  animate(grand_tour(), display = display_xy(), fps=15)
```

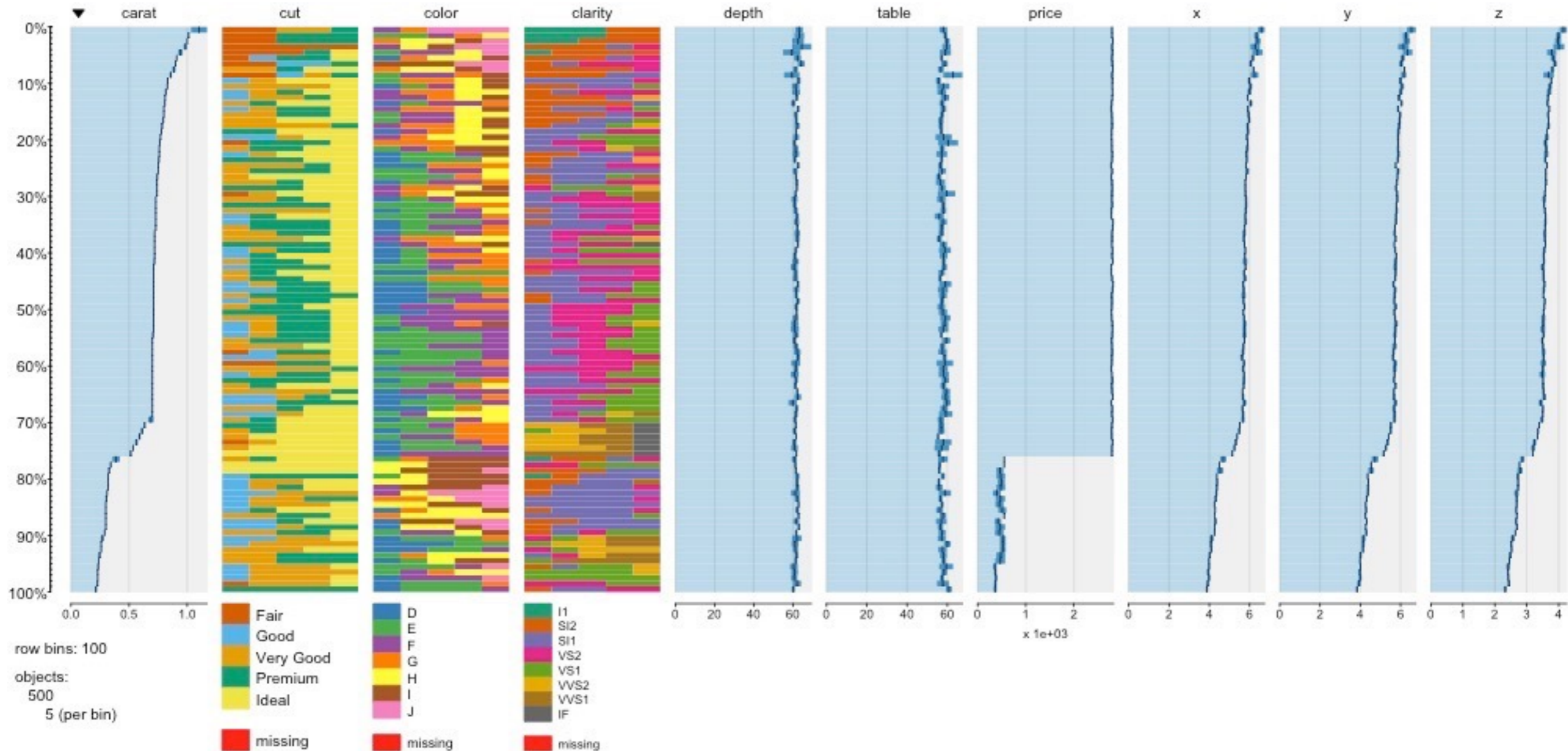
Wickham, H., Cook, D., Hofmann, H., and Buja, A. (2011). tourr: An R package for exploring multivariate data with projections. *Journal of Statistical Software*, 40(2).

CSV fingerprint



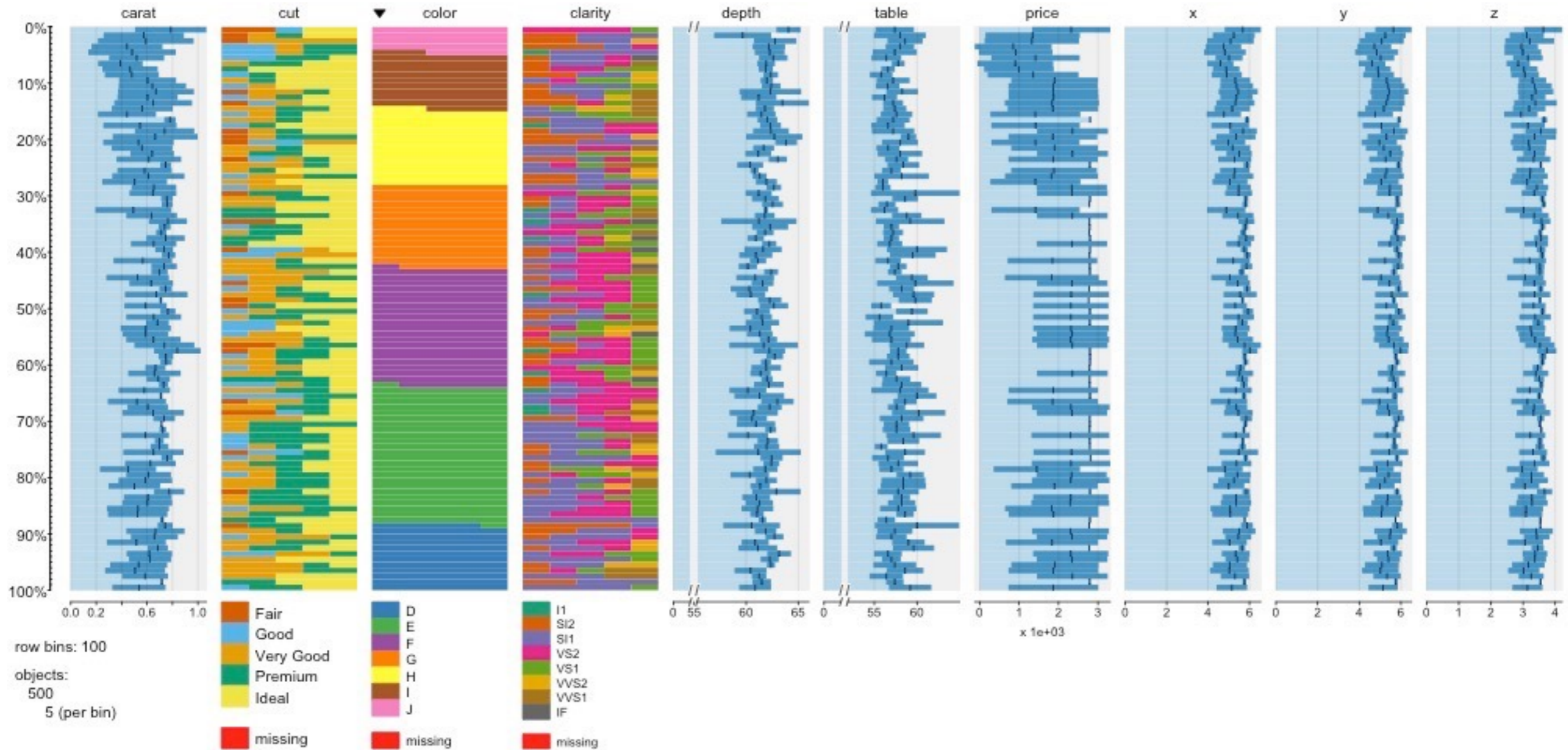
String Empty Decimal Integer

Tableplots



```
library(tabplot)  
tableplot(diamonds)
```


Tableplots



```
tableplot(diamonds, sortCol="color")
```

Lab: visualizing multidimensional data

Lab: visualizing multidimensional data

Use the college data posted on the course website and #lab5

- Create a higher-dimensional data visualization using either:
 - 3D (<https://help.plot.ly/tutorials/#graph3d>)
 - Sub-plots (<https://help.plot.ly/subplot-layouts/>)
- Post in #lab5, along with some comments about how effective you believe your visualization is