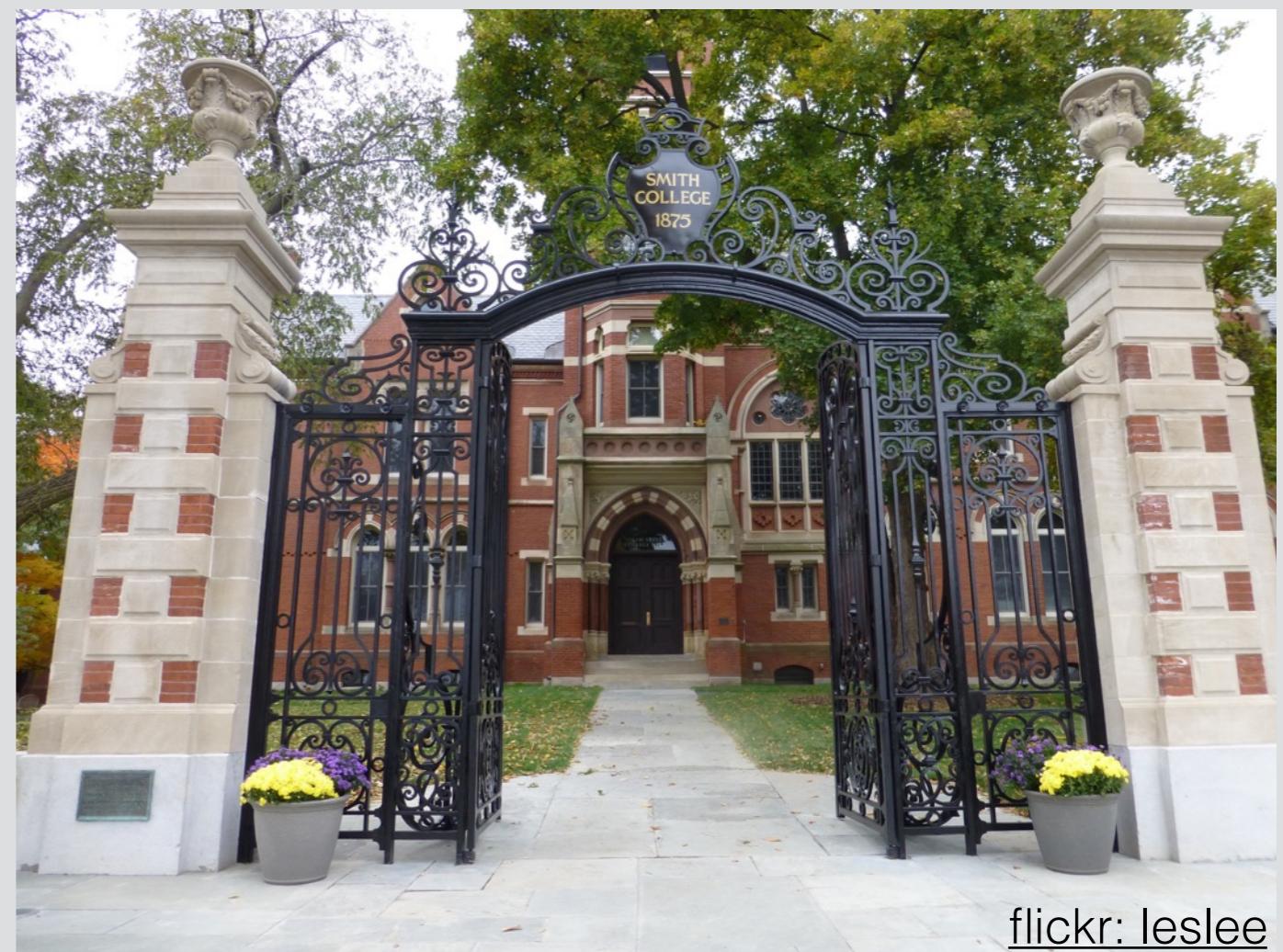


Using Data Visualization to Find Patterns in Multidimensional Data

Amelia McNamara [@AmeliaMN](#)

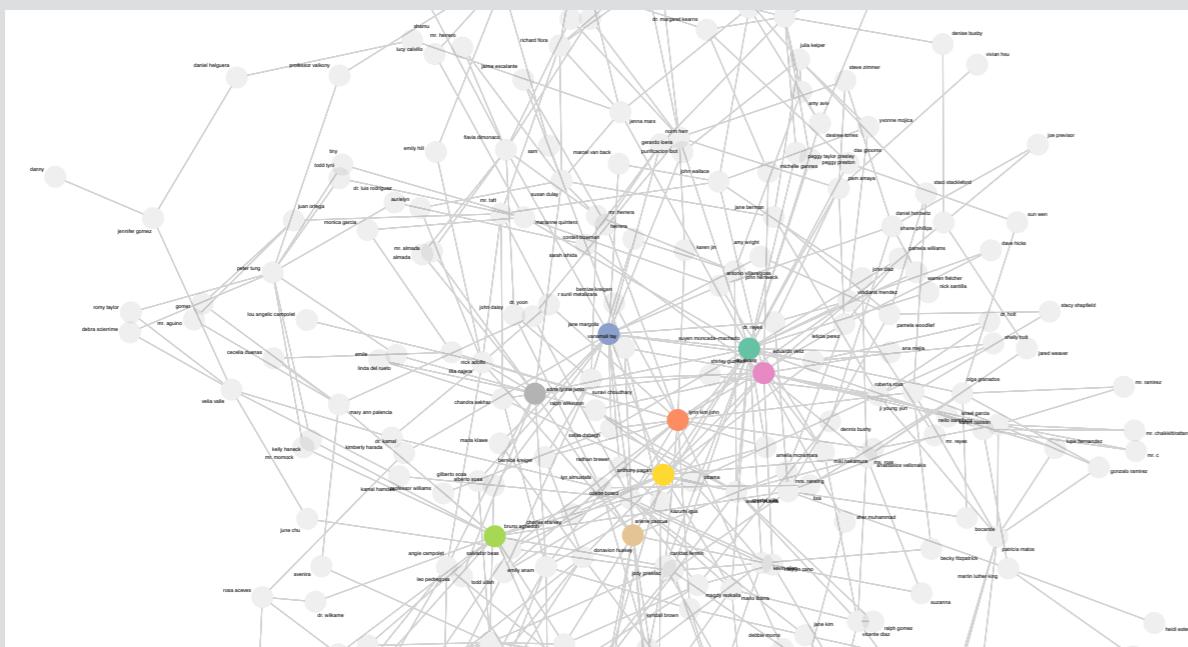
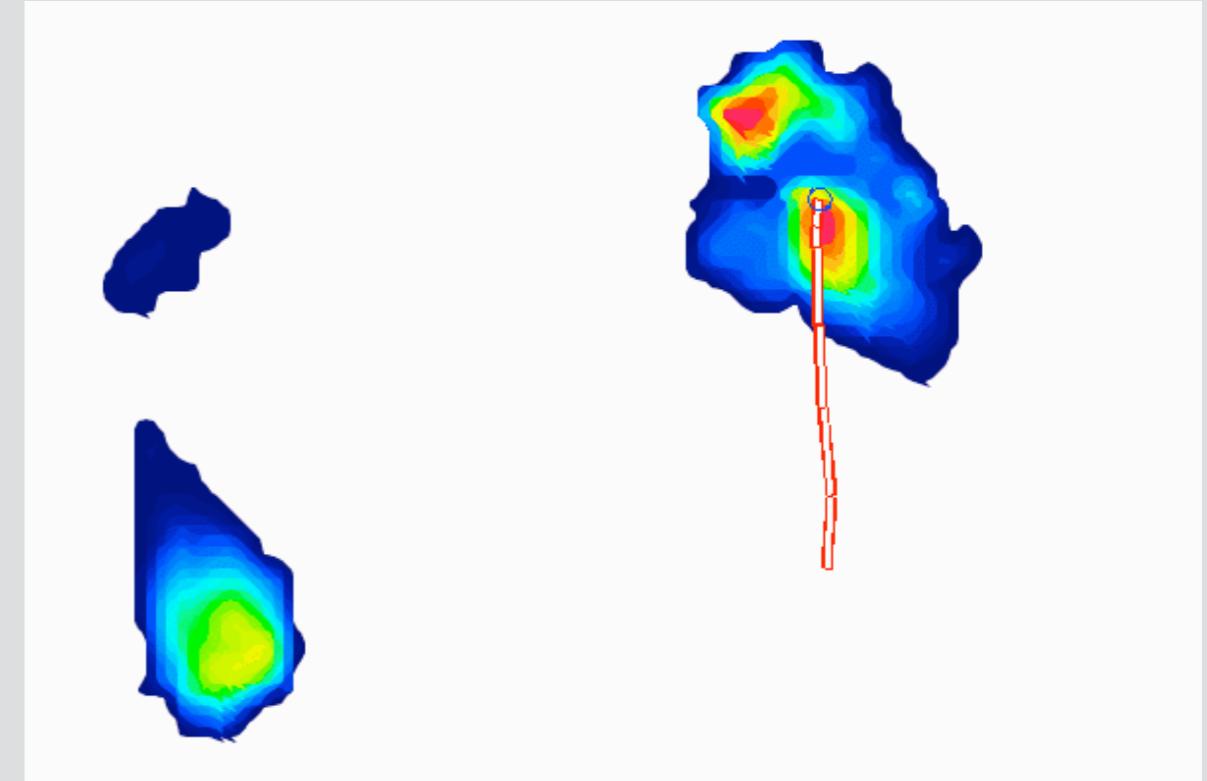
About me

- BA in Mathematics and English from Macalester College
- PhD in Statistics from UCLA
- Visiting Assistant Professor of Statistical and Data Sciences at Smith College
- Research at the intersection of statistical computing, statistics education, data visualization



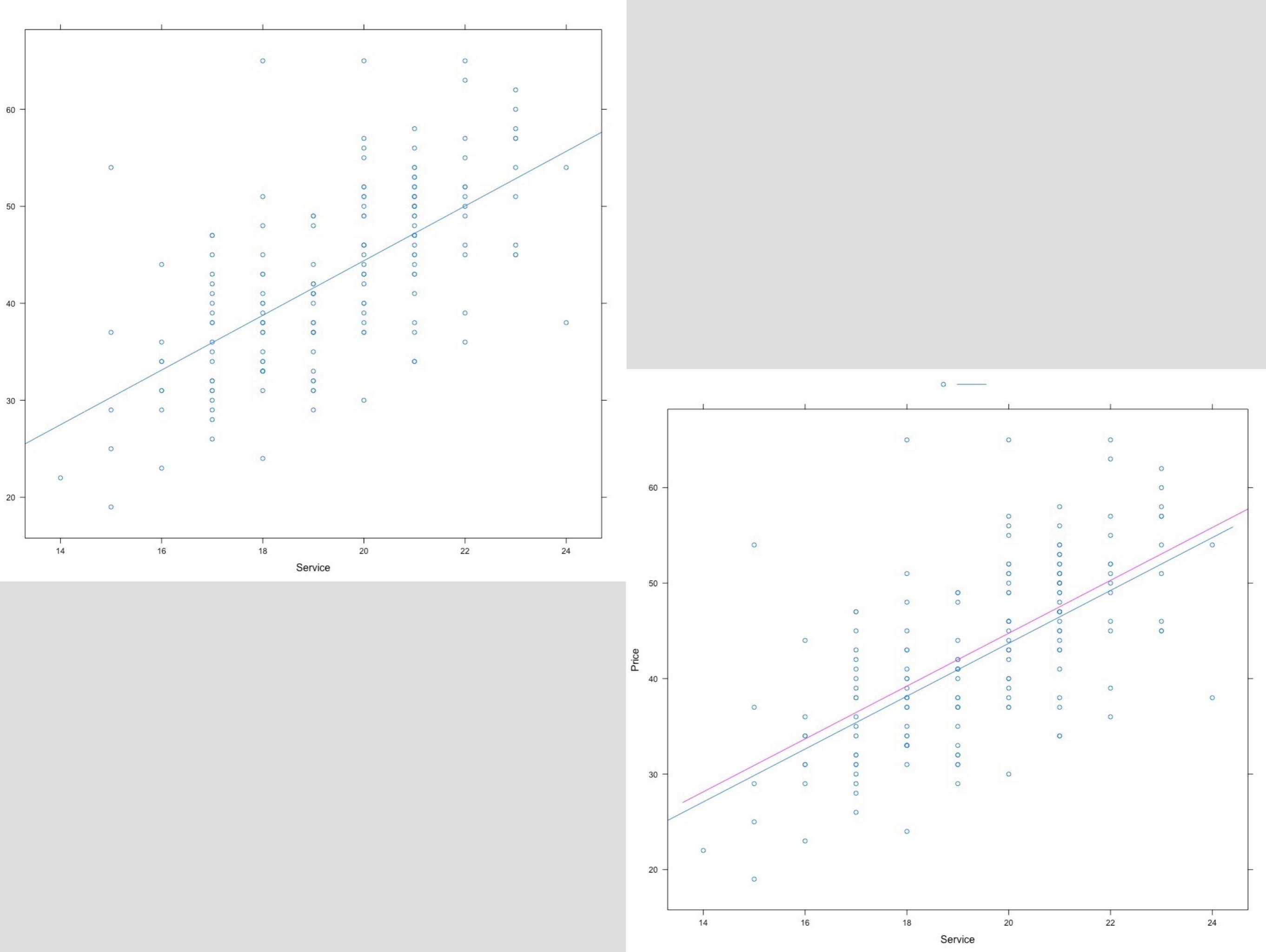
[flickr: leslee](#)

Statistics and data are everywhere

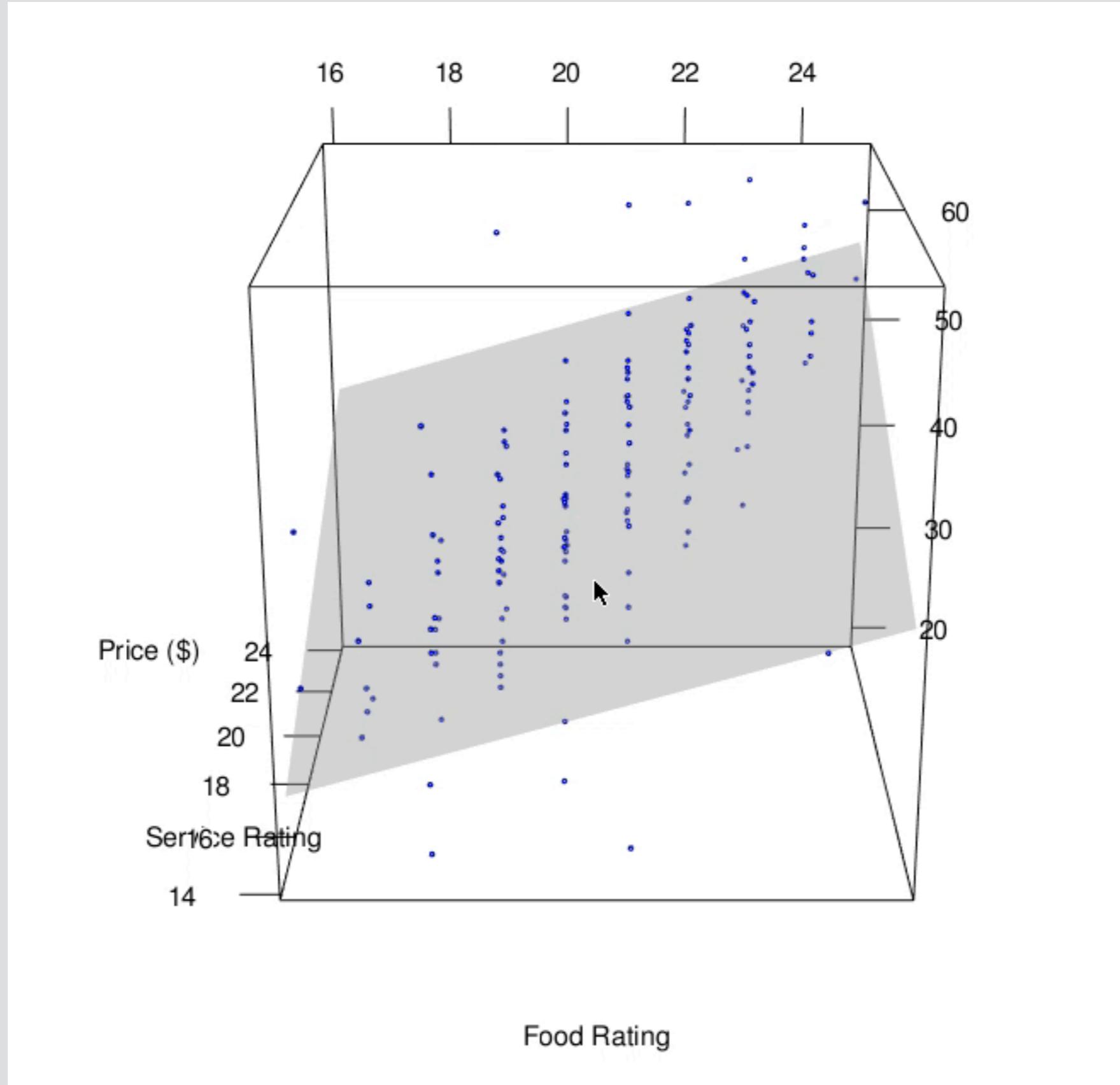


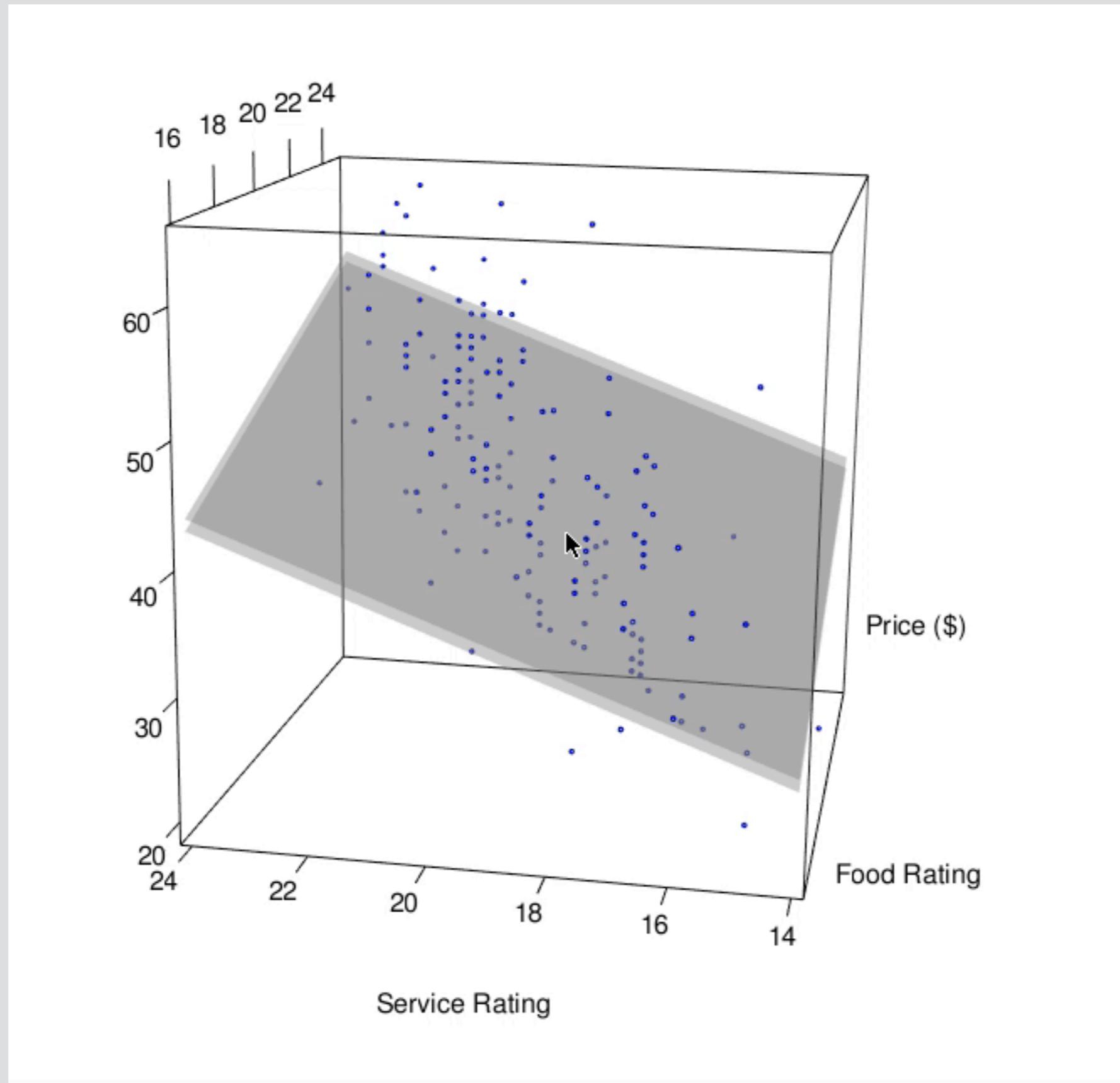
Big data is tall,
multidimensional
data is wide

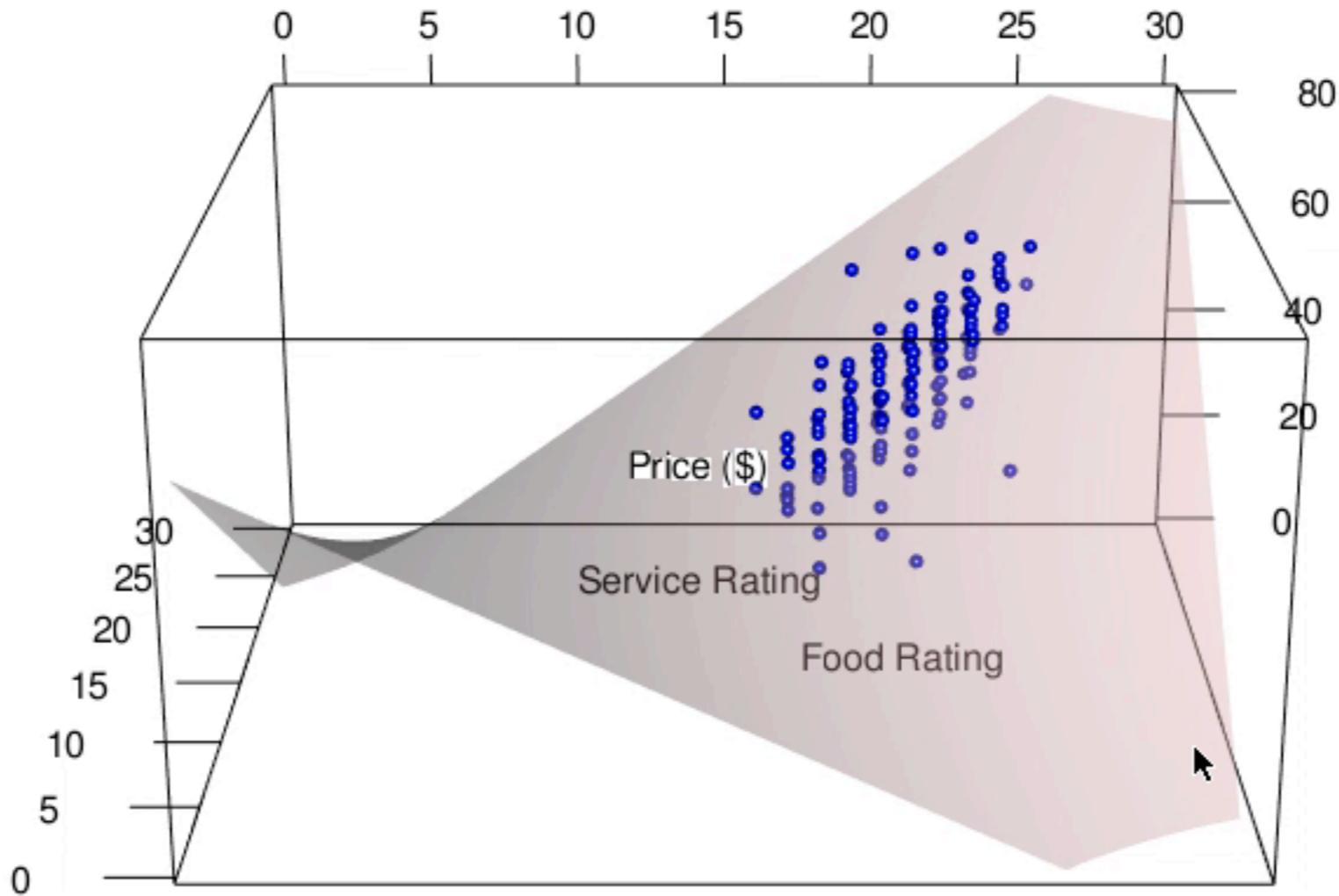
Data can be both!



```
library(rgl)
```







Some motivation

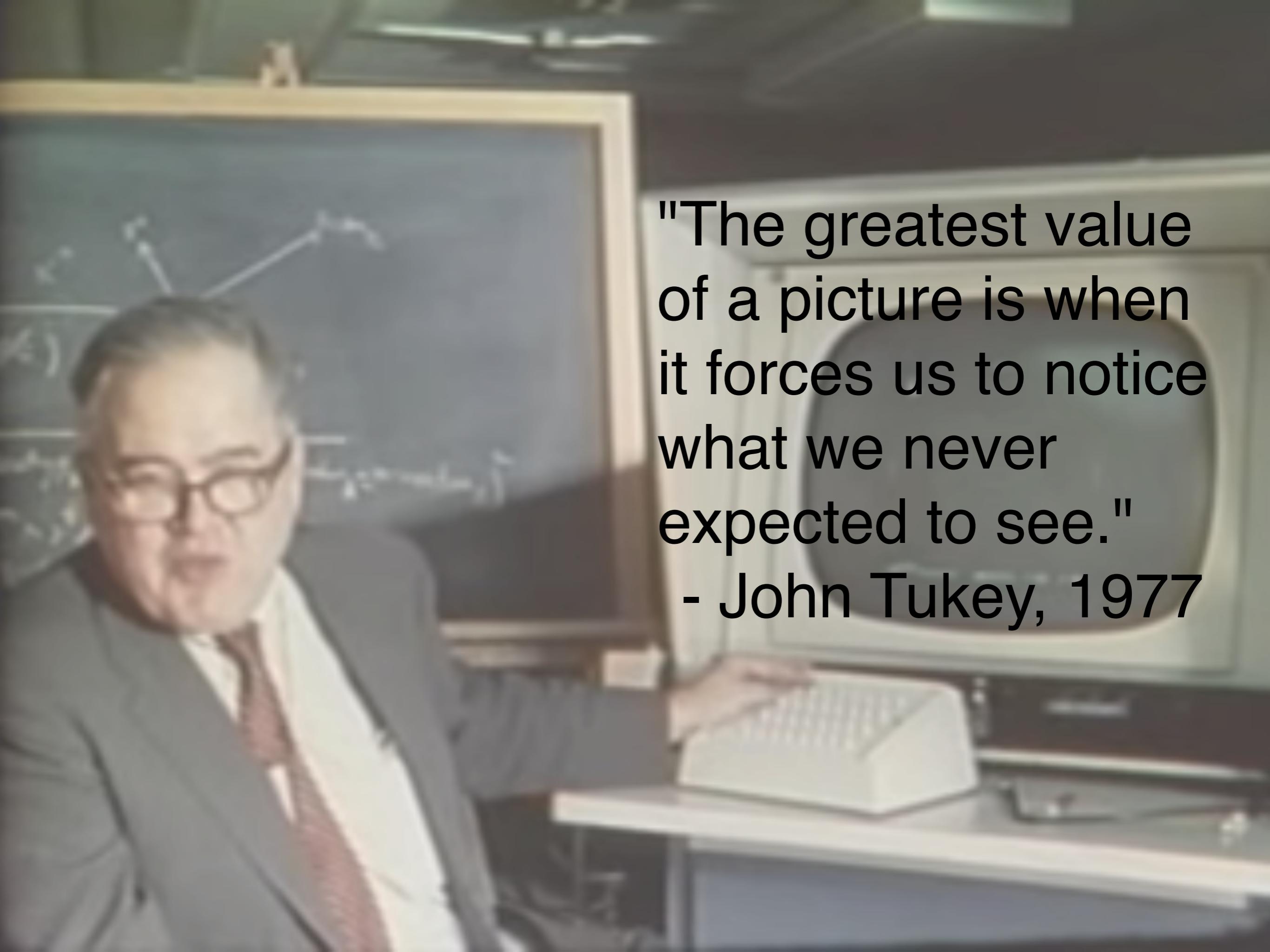
“Everybody says that statistics should be taught.
But how?

Statistics are not simply figures. It is said that nothing lies like figures except facts. You want statistics to tell you the truth. You can find truth there if you know how to get at it, and romance, human interest, humor and fascinating revelations as well.”

- Joseph Pulitzer,
The Power of Public Opinion, 1904

“We will be remiss in our duty to our students if we do not see that they learn to use the computer more easily, flexibly, and thoroughly than we ever have; we will be remiss in our duties to ourselves if we do not try to improve and broaden our own uses.”

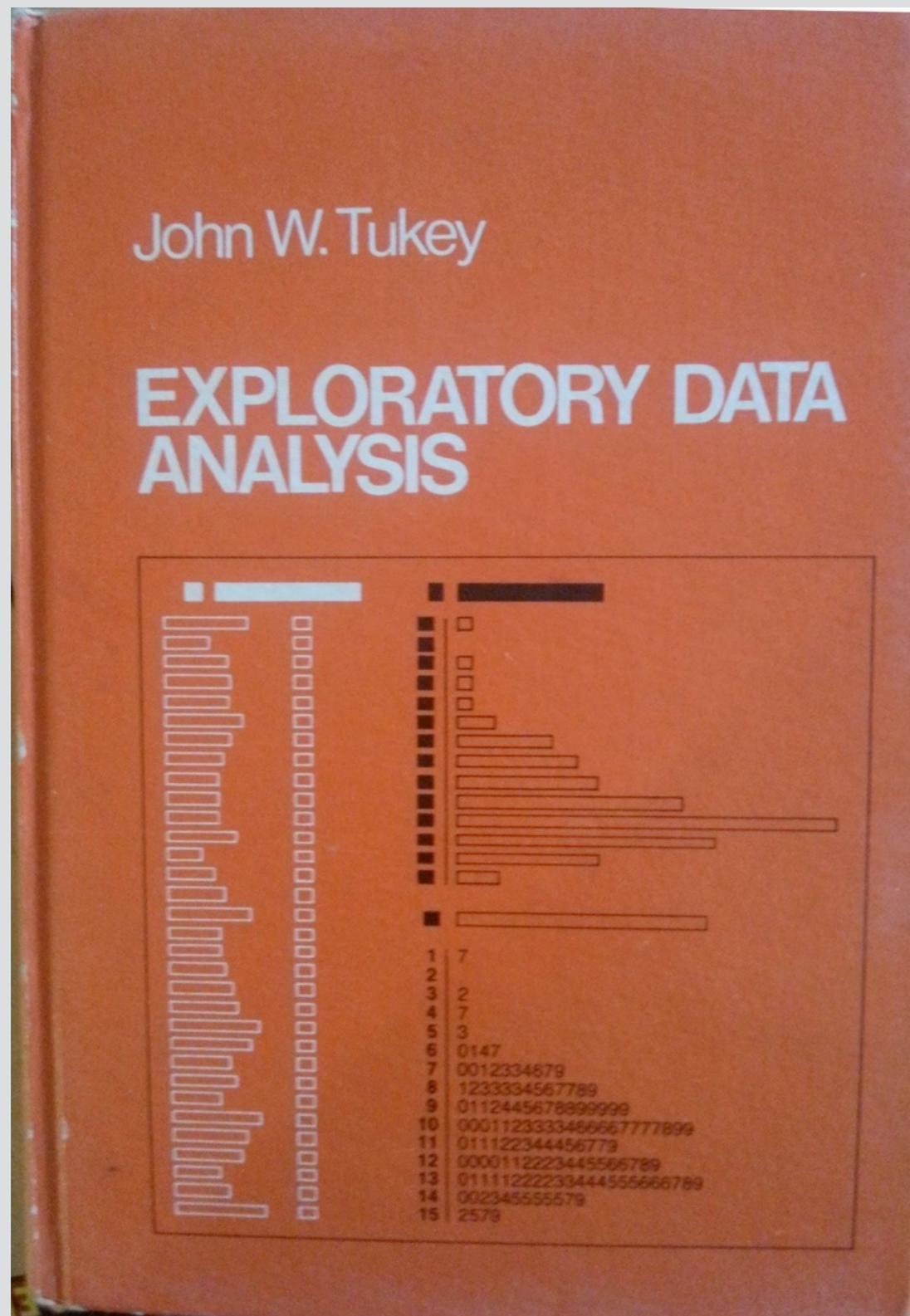
- John Tukey,
The Technical Tools of Statistics
talking about the class of 1970

A black and white photograph of a man with glasses, wearing a suit and tie, looking towards a chalkboard. The chalkboard has some faint, illegible markings. In the foreground, there's a stack of papers or books.

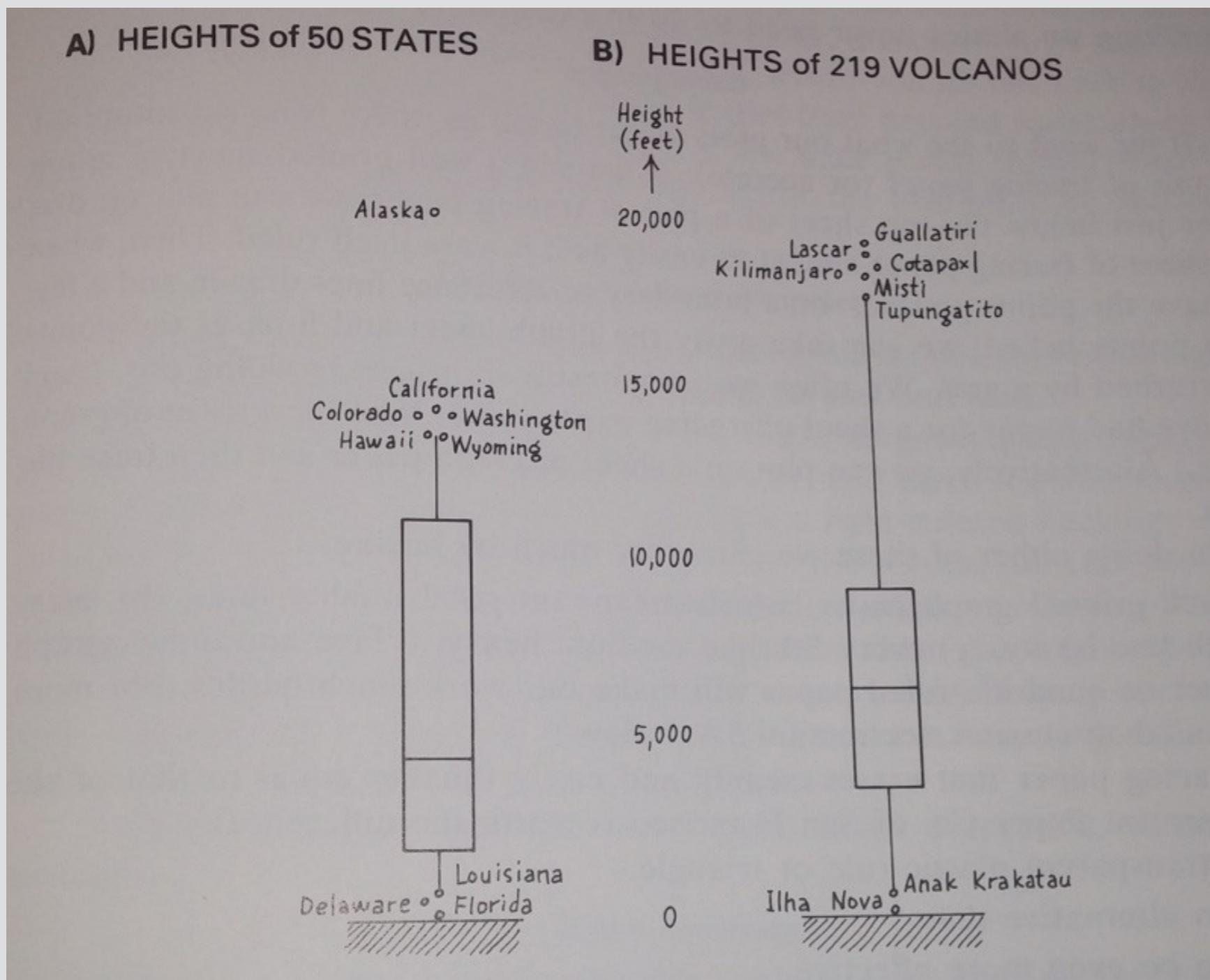
"The greatest value
of a picture is when
it forces us to notice
what we never
expected to see."

- John Tukey, 1977

Exploratory Data Analysis (EDA)



Exploratory Data Analysis (EDA)

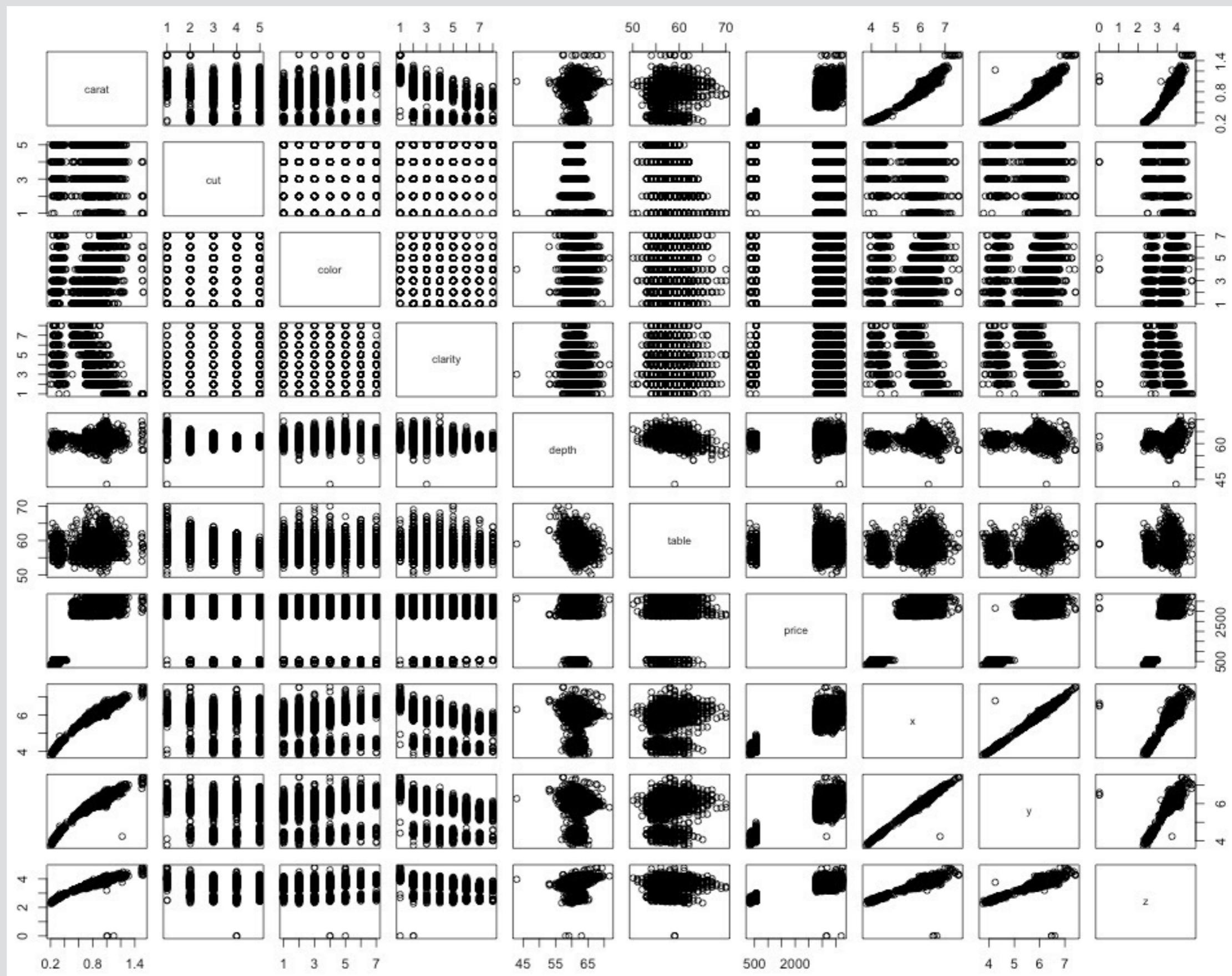


Goal 1:

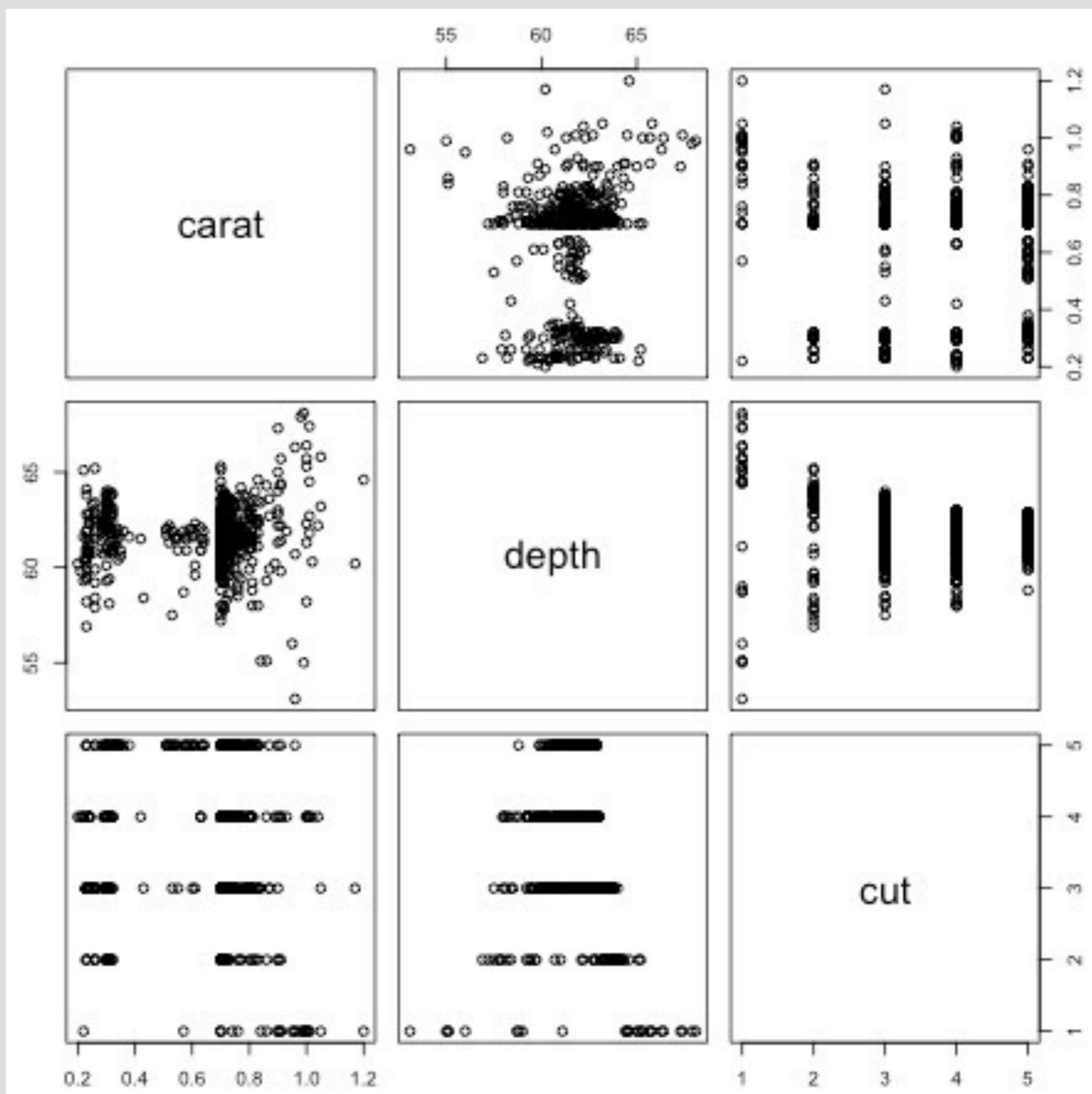
Allow a person with a computer to be better than John Tukey with a pencil

```
plot(diamonds)
```

Pairs plot

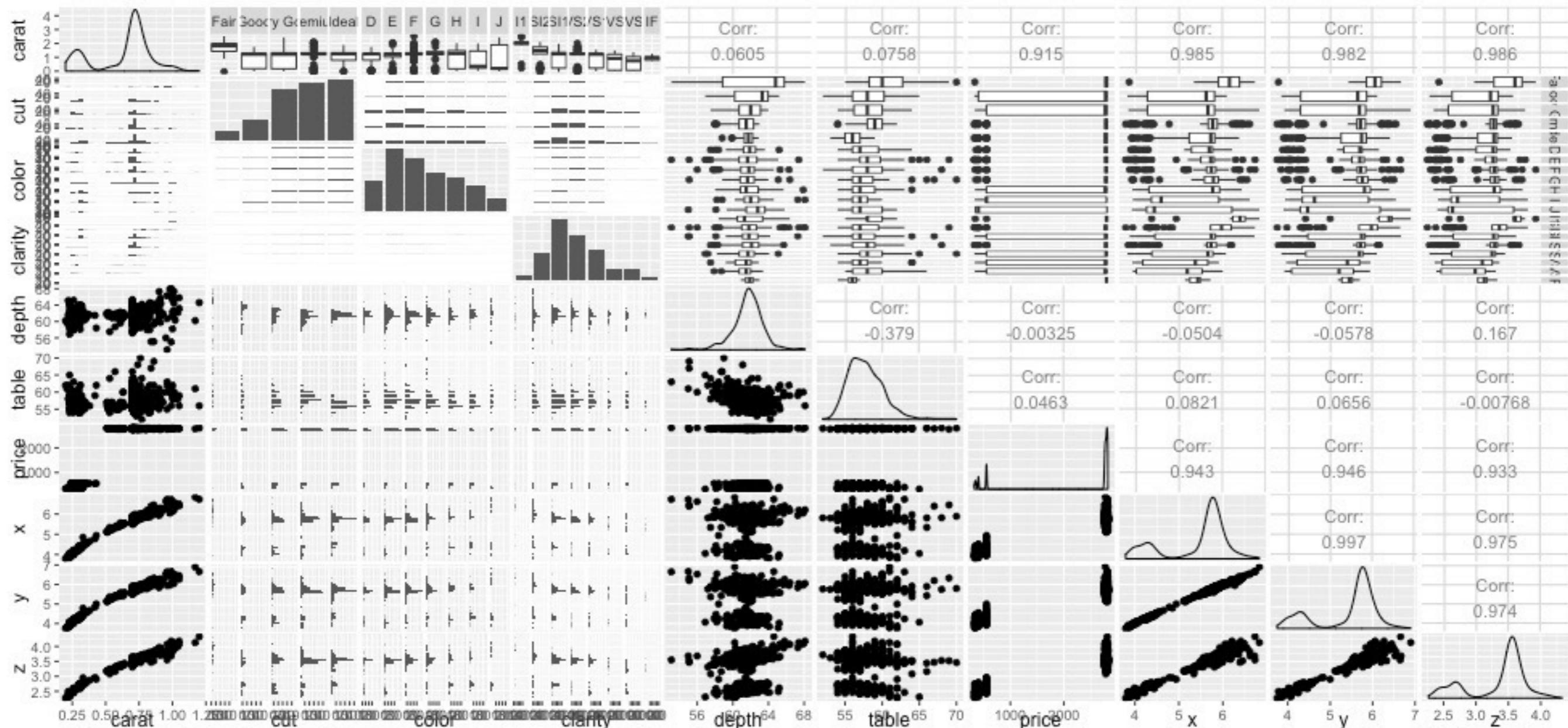


```
select(diamonds, carat, depth, cut) %>%  
plot()
```



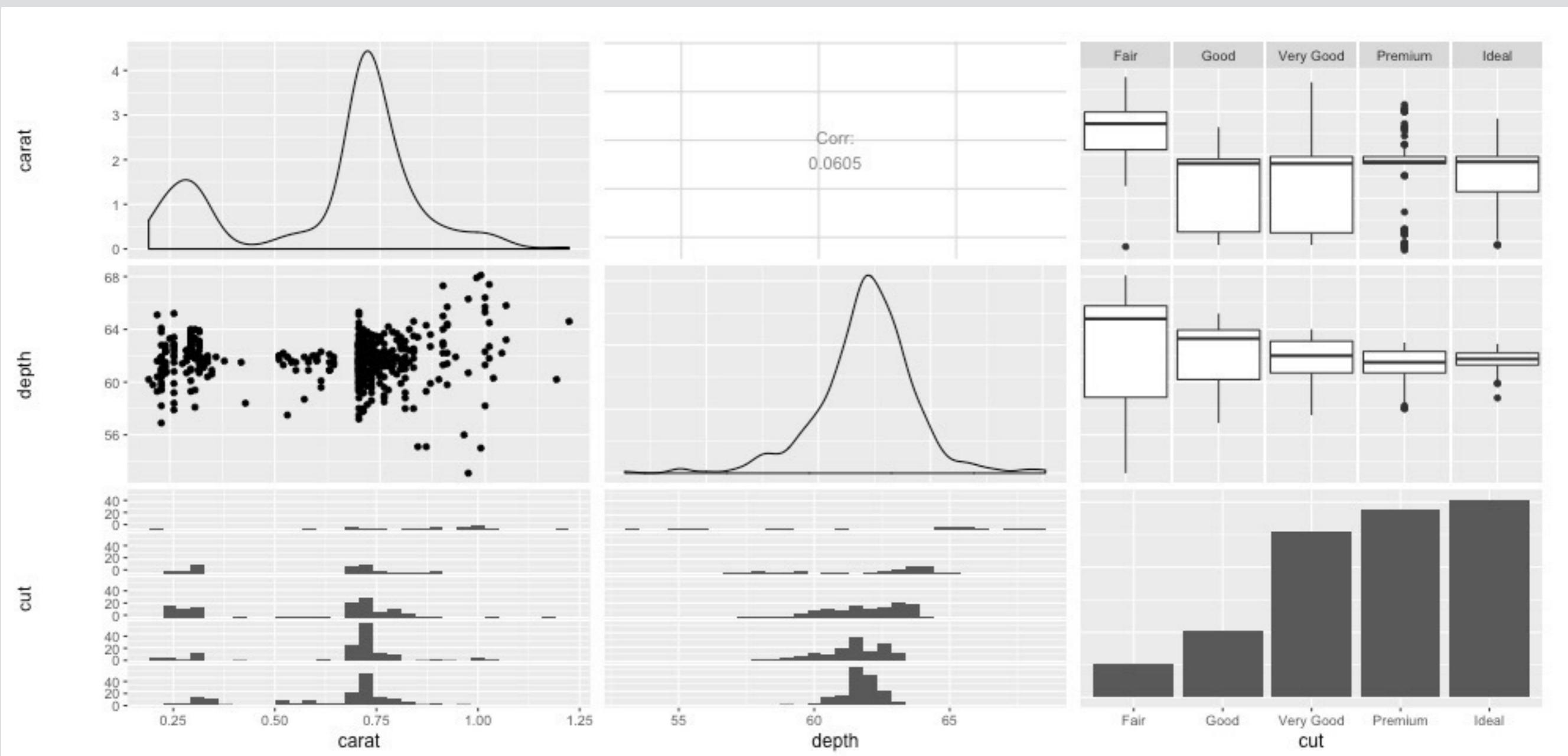
Emerson, J. W., Green, W. A., Schloerke, B., Crowley, J., Cook, D., Hofmann, H., and Wickham, H. (2013). The generalized pairs plot. *Journal of Computational and Graphical Statistics*, 22(1):79–91. <http://bit.ly/gpairs>

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



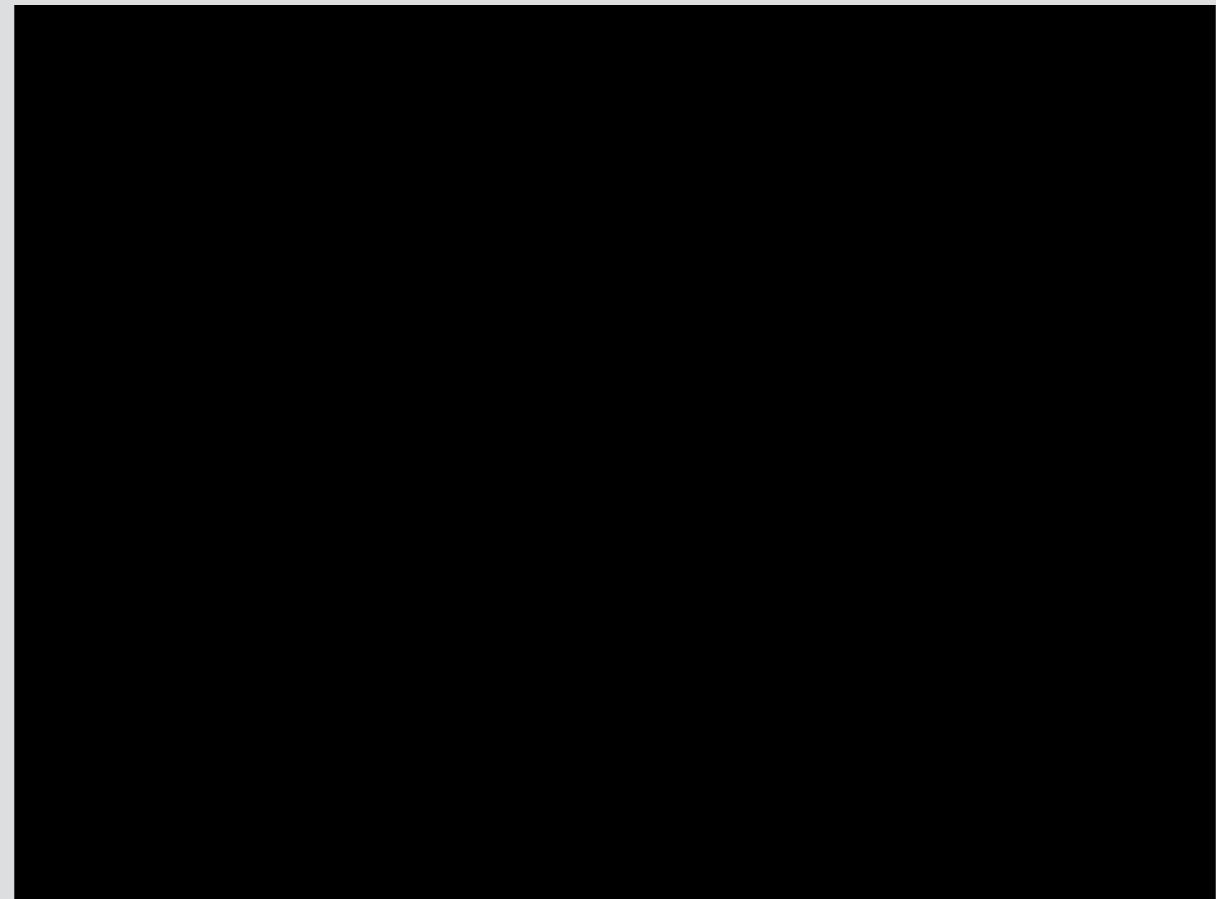
Generalized
pairs plot

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



Generalized
pairs plot

prim9



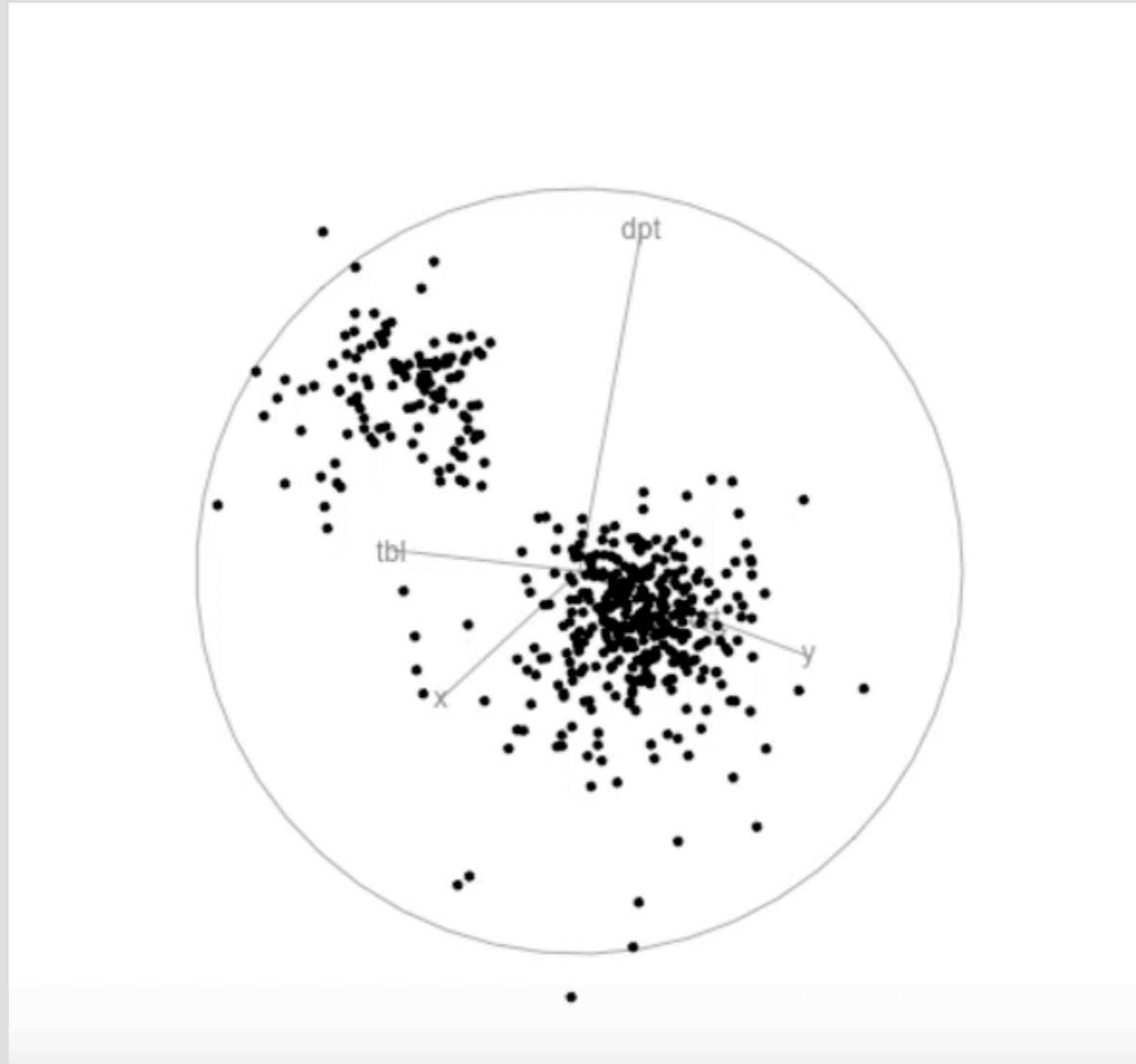
Goal 2:

Allow a person with a computer to be better than John Tukey with a computer

The grand tour

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).

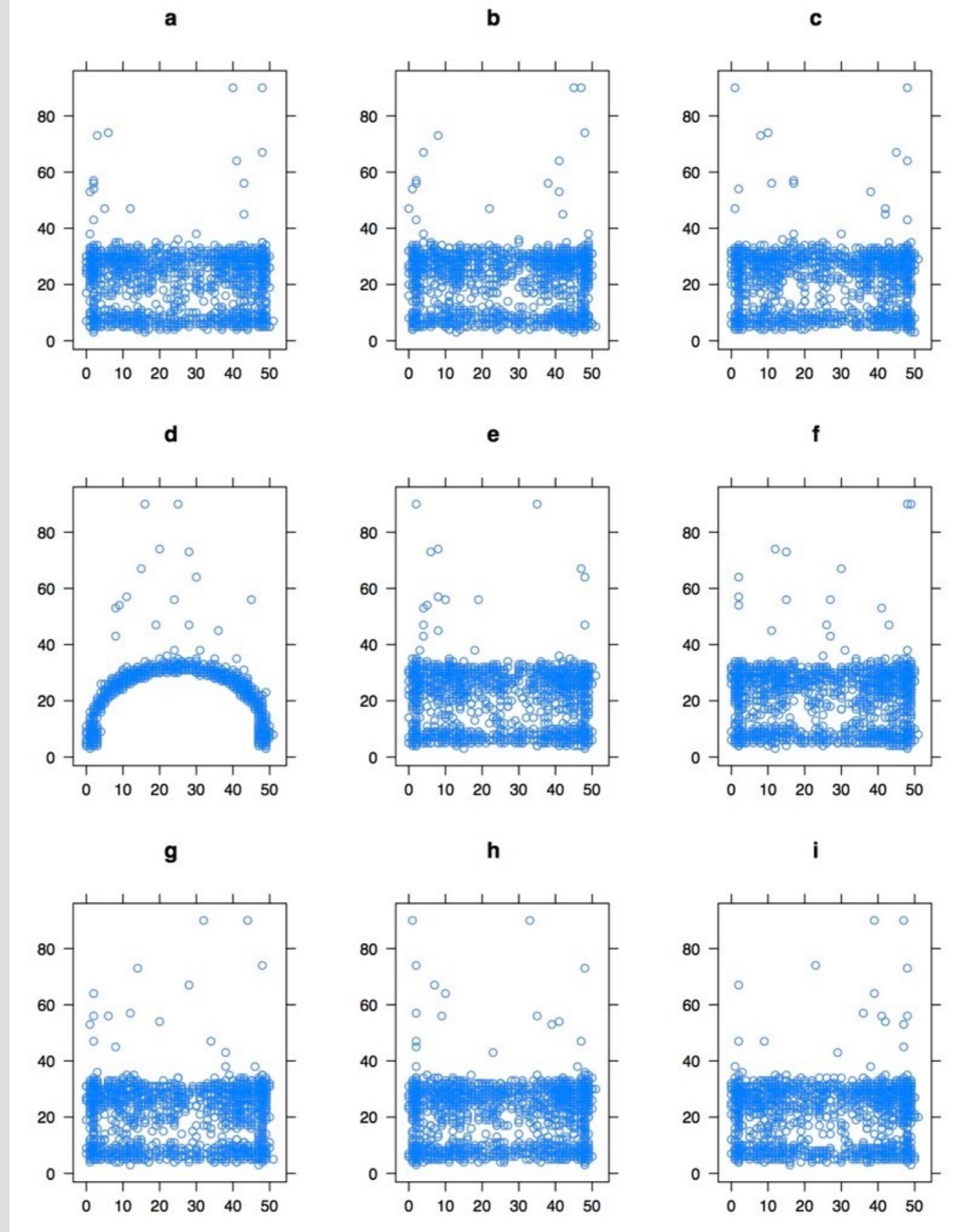
```
library(tourrr)
select(diamonds, carat, depth, table, x, y, z) %>%
  animate(grand_tour(), display = display_xy(),fps=15)
```



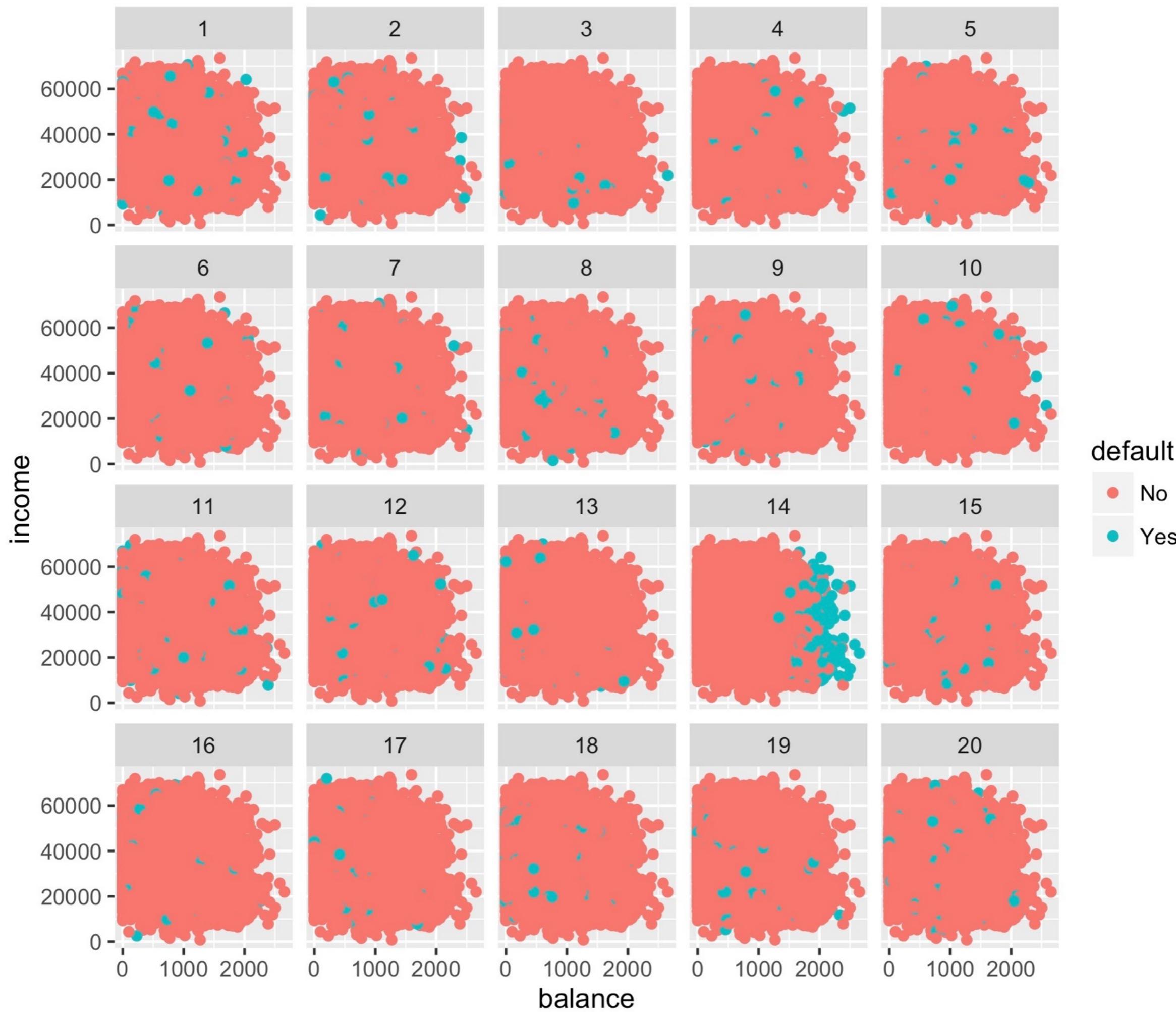
Graphical inference

Hadley Wickham, Dianne Cook, Heike Hofmann, and Andreas Buja. (2010). Graphical Inference for Infovis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6).
http://bit.ly/graphical_inference

Permuted graphics



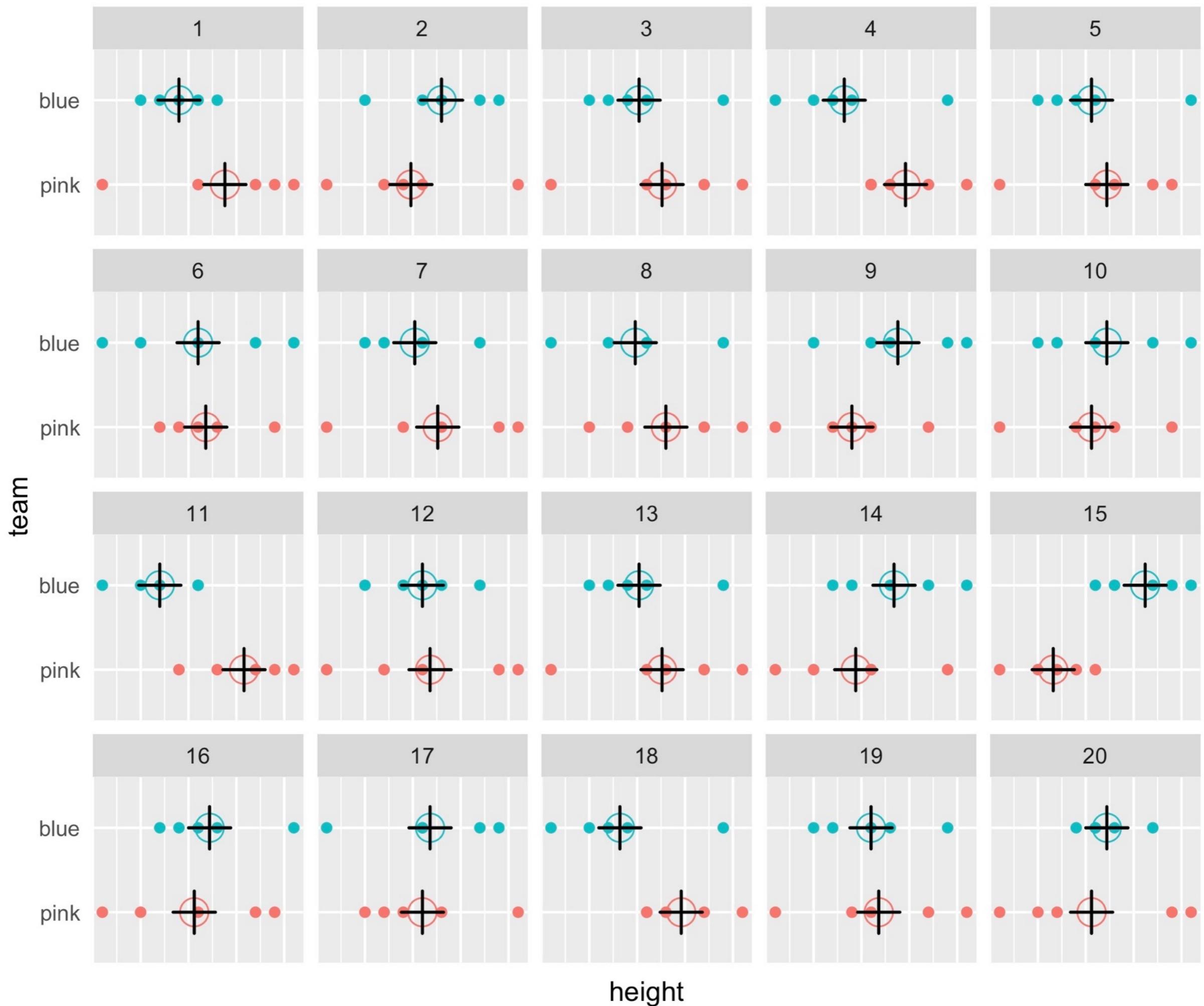




```
library(ggplot2)
ggplot(Default) +
  geom_point(aes(x=balance, y=income, col=default))

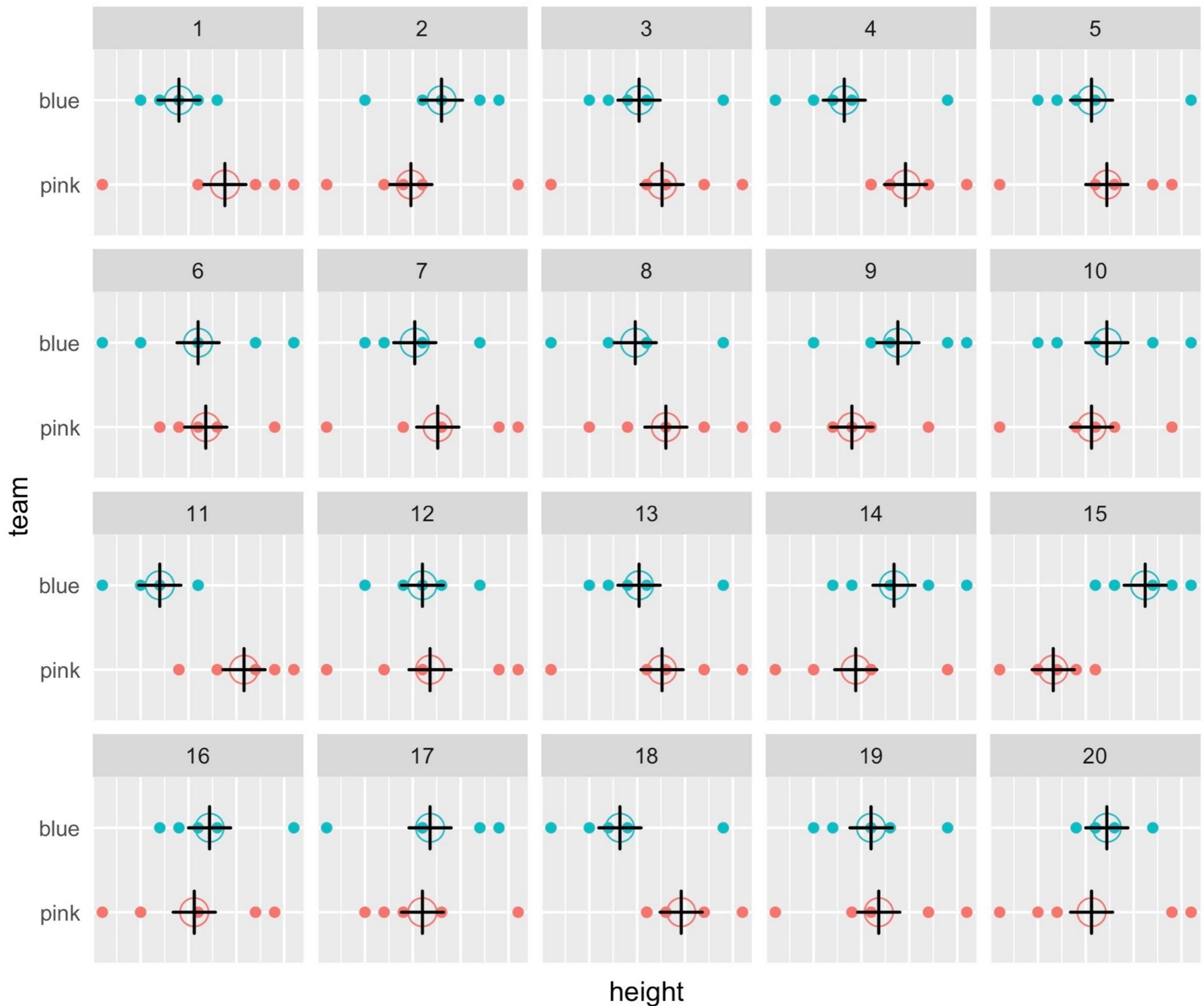
library(nullabor)
ggplot(lineup(null_permute('default'), Default)) +
  geom_point(aes(x=balance, y=income, col=default)) +
  facet_wrap(~.sample)

decrypt("01CE bQTQ Aw GWPATAWw d")
[1] "True data in position 14"
```



```
decrypt("01CE bQTQ Aw GWPATAWw J")
```

```
[1] "True data in position 5"
```

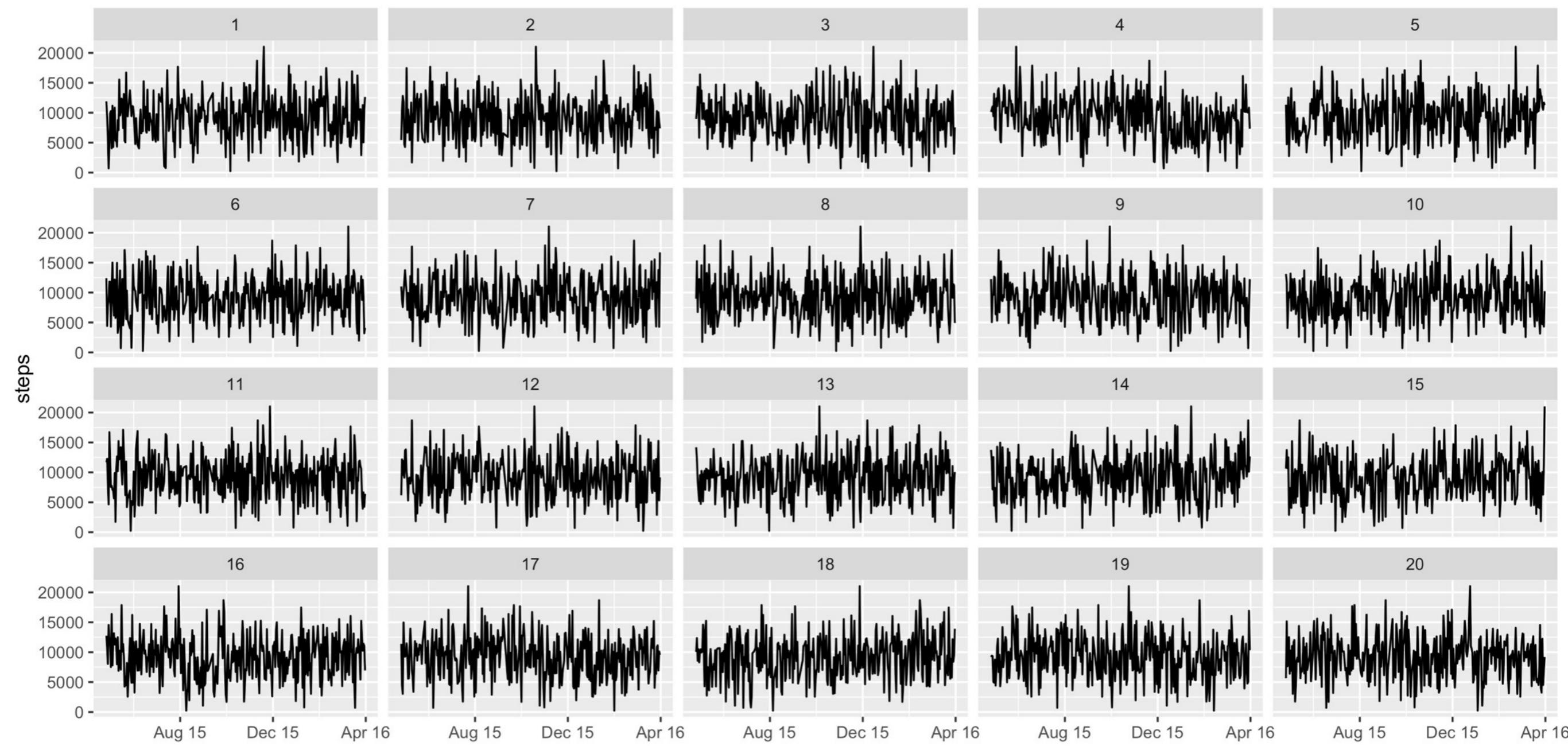






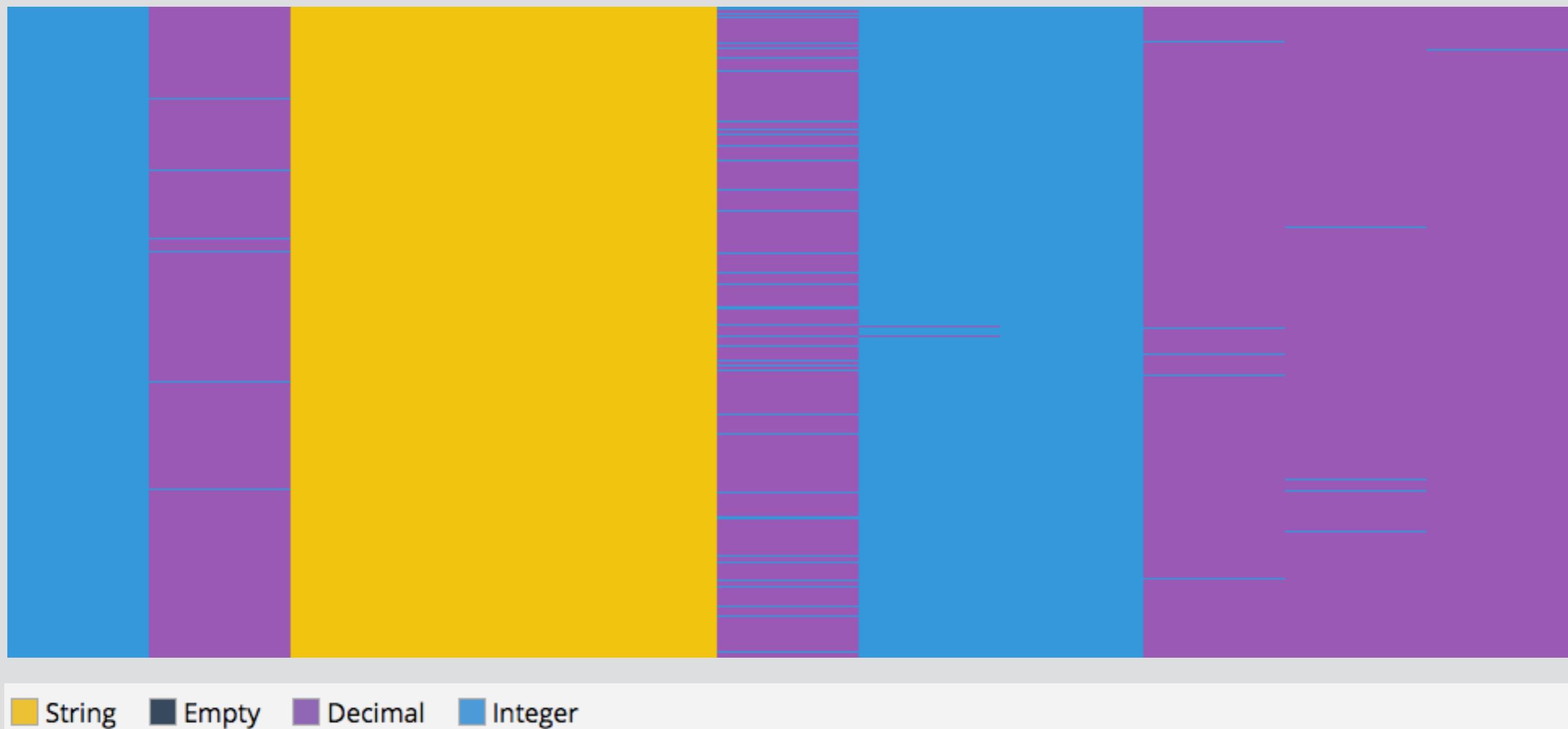
```
decrypt("OlCE bQTQ Aw GWPATAWw y")
```

```
[1] "True data in position 4"
```



CSV fingerprint

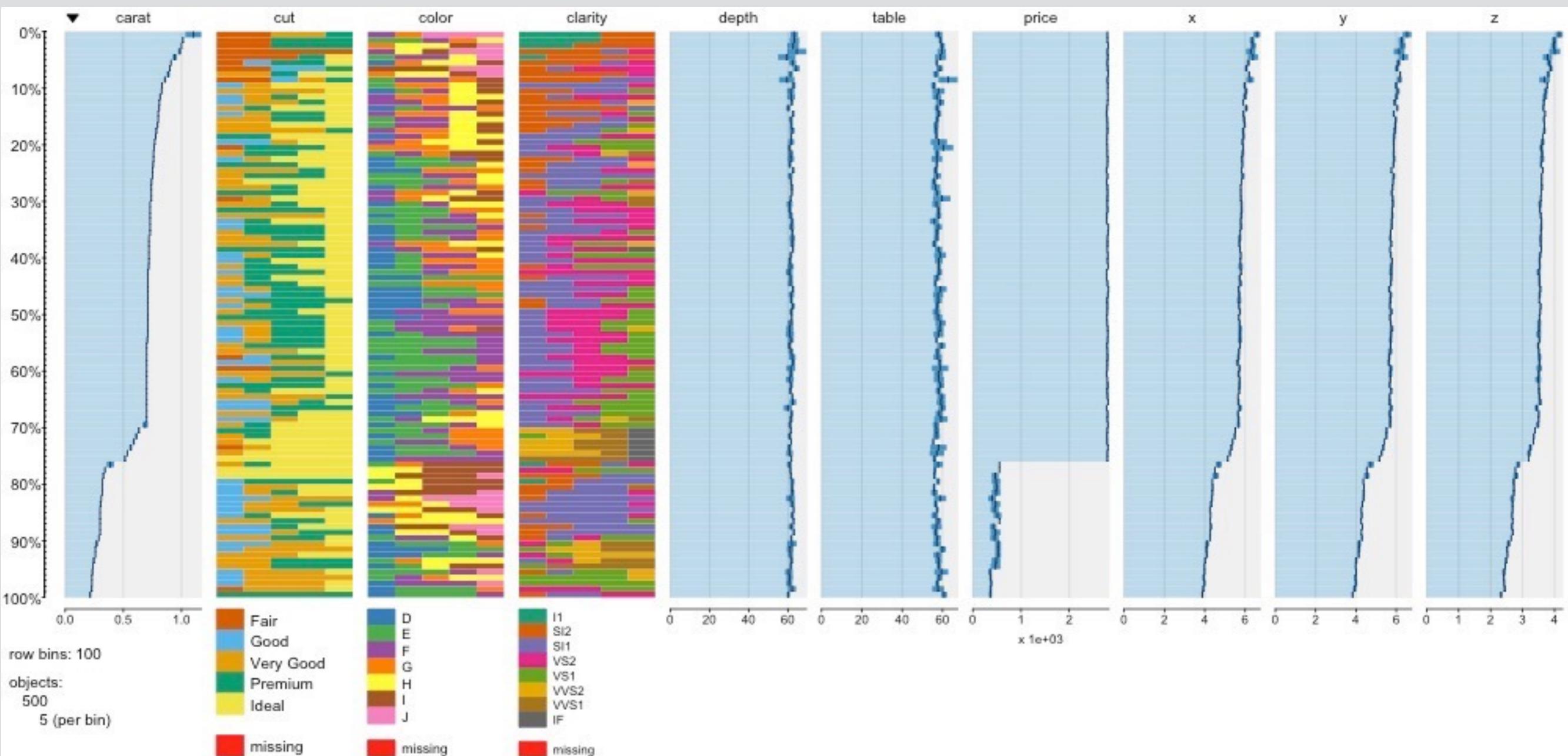
Powell, V. CSV Fingerpint. www.bit.ly/csvfingerprint



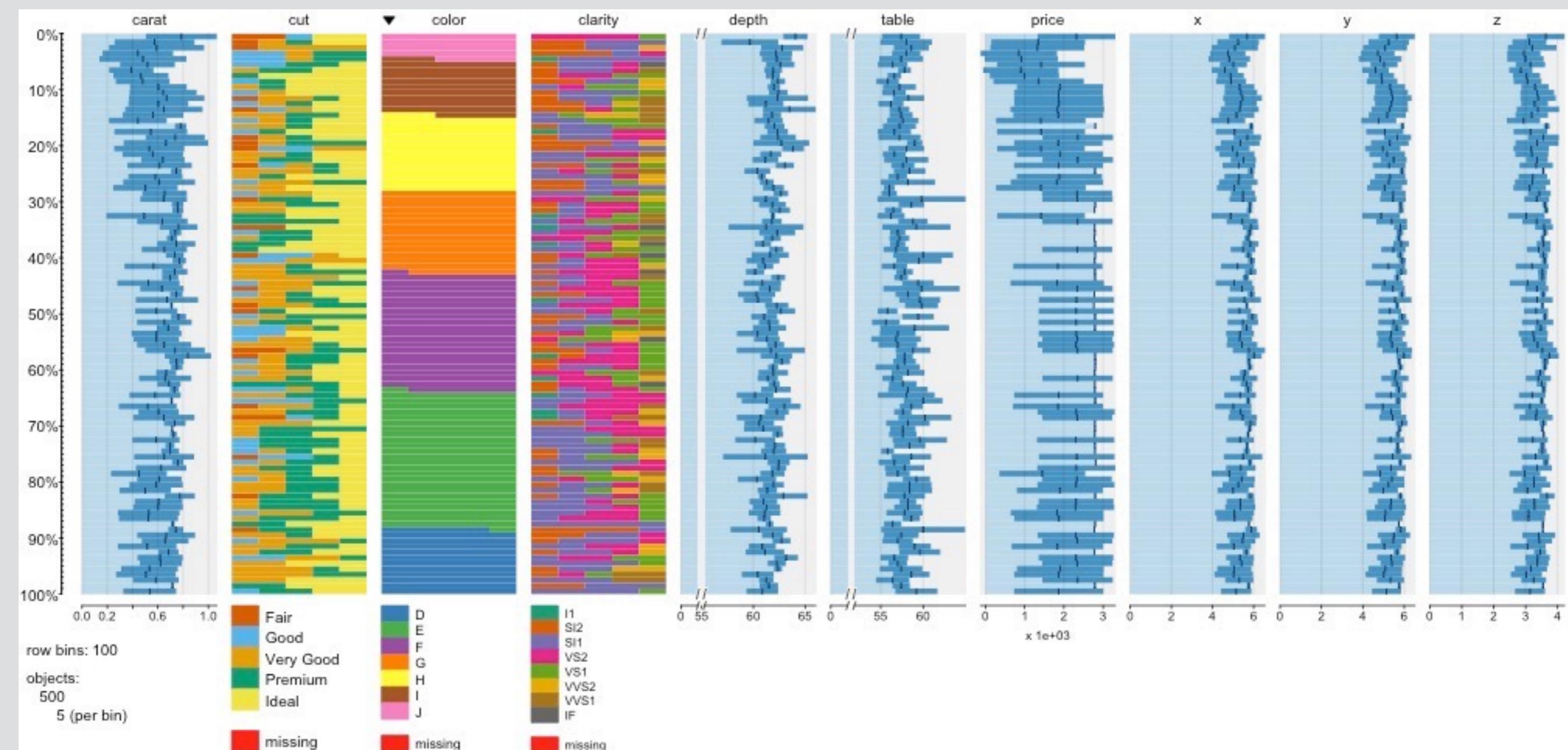
Tableplots

Tennekes, M., de Jonge, E., and Daas, P. J., H. (2013). Visualizing and inspecting large datasets with tableplots. Journal of Data Science, 11(2013):43-58. <http://bit.ly/tabplot>

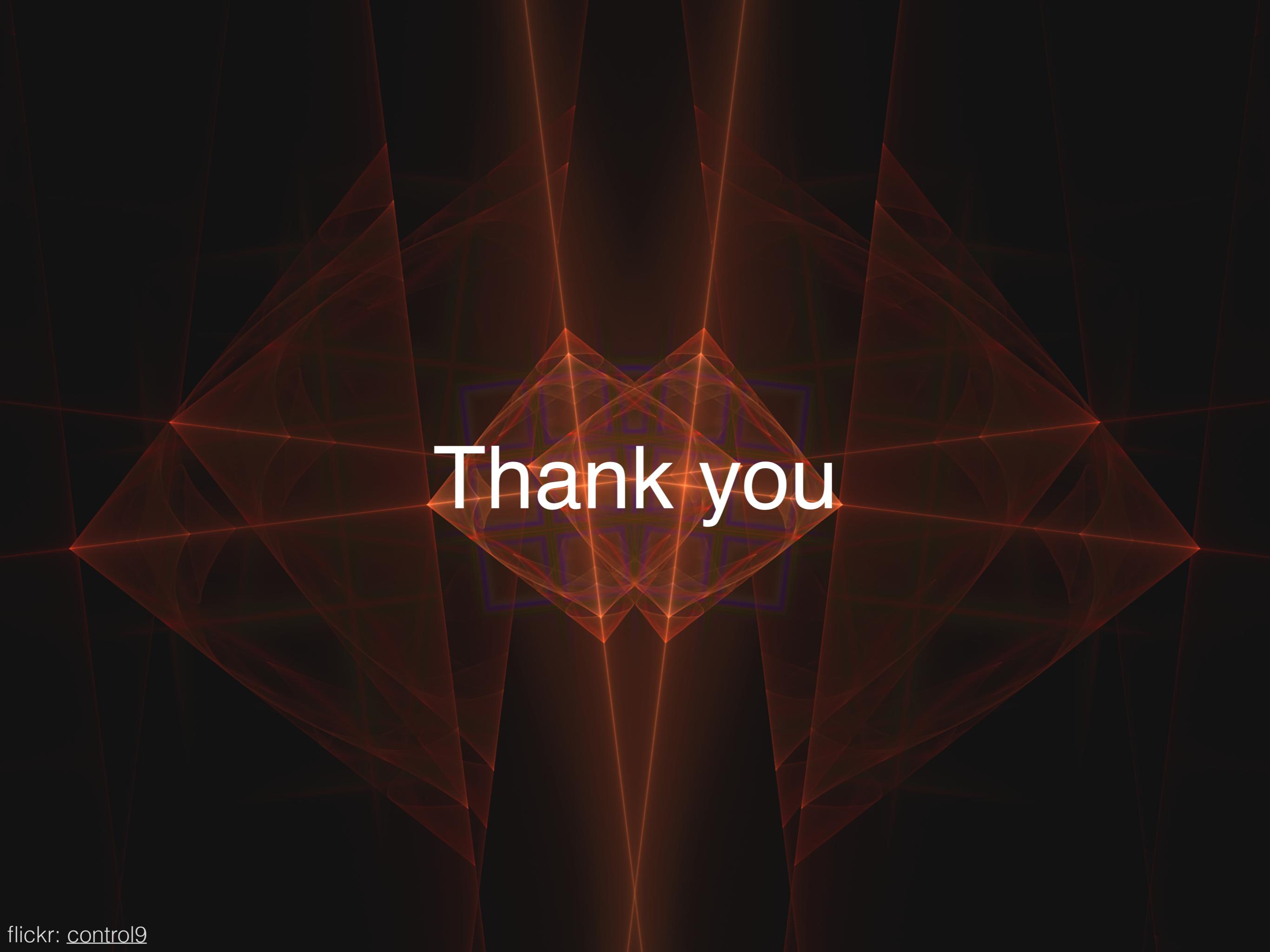
```
library(tabplot)
tableplot(diamonds)
```



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



Clearly, there is room
for computers to
give us even more
super powers!



A dark background featuring a complex, glowing red fractal pattern. The pattern consists of numerous thin, translucent red lines forming a three-dimensional, branching structure that resembles a network or a crystal lattice. The intensity of the red light varies, creating a sense of depth and motion. In the center of this pattern, the words "Thank you" are written in a large, white, sans-serif font.

Thank you