Additional Encodings

As seen earlier in the lesson, we typically try to use position on the x- and y-axes to encode, or depict the value of variables. If we have more than two variables, however, we have to start considering other visual encodings for the additional variables.

In general, **color and shape** are best for **categorical** variables, while the **size of marker** can assist in adding additional **quantitative data**, as we demonstrated here.

Only use these additional encodings when absolutely necessary. Often, overuse of these additional encodings suggest you are providing too much information in a single plot. **Instead, it might be better to break the information into multiple individual messages**, so the audience can understand every aspect of your message. You can also build in each aspect one at a time, which you saw in the previous lesson with Hans Rosling. This feels less overwhelming than if you just saw this plot all at once.

Extra: Code

Some of the plots in this presentation were created using the programming language R, and a very popular library known as **ggplot2**. Though this is beyond the scope of this course, the code used to create these visualizations is provided below:

```
install.packages('ggplot2')
library(ggplot2)

df = read.csv(file.choose()) #select your dataset
df2 = head(df, 30)



qplot(df2$Math_SAT, df2$Verbal_SAT, xlab = 'Math SAT Score'
```