

Example of correlation

Alexander Valencia

2023-03-10

```
#source: https://blog.revolutionanalytics.com/2017/05/the-datasaurus-dozen.html  
#In case you need to install the package  
#install.packages("datasauRus")  
library("tidyverse")  
library("datasauRus")
```

The datasaurus dataset contains more than 12 different datasets with very similar characteristics. Based on the summary below, do you have any expectations on how similar the datasets should look like?

```
datasaurus_dozen|>  
  group_by(dataset)|>  
  summarise(mean_x = round(mean(x),2),  
            mean_y = round(mean(y),2),  
            sd_x = round(sd(x),2),  
            sd_y = round(sd(y),2),  
            cor = round(cor(x,y),2))
```



```
## # A tibble: 13 x 6  
##   dataset    mean_x mean_y sd_x  sd_y cor  
##   <chr>      <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 away      54.3  47.8  16.8  26.9 -0.06  
## 2 bullseye  54.3  47.8  16.8  26.9 -0.07  
## 3 circle    54.3  47.8  16.8  26.9 -0.07  
## 4 dino      54.3  47.8  16.8  26.9 -0.06  
## 5 dots      54.3  47.8  16.8  26.9 -0.06  
## 6 h_lines   54.3  47.8  16.8  26.9 -0.06  
## 7 high_lines 54.3  47.8  16.8  26.9 -0.07  
## 8 slant_down 54.3  47.8  16.8  26.9 -0.07  
## 9 slant_up   54.3  47.8  16.8  26.9 -0.07  
## 10 star     54.3  47.8  16.8  26.9 -0.06  
## 11 v_lines  54.3  47.8  16.8  26.9 -0.07  
## 12 wide_lines 54.3  47.8  16.8  26.9 -0.07  
## 13 x_shape  54.3  47.8  16.8  26.9 -0.07
```

However this is how data looks like.

```
ggplot(datasaurus_dozen, aes(x = x, y = y, colour = dataset)) +  
  geom_point() +  
  theme_void() +  
  theme(legend.position = "none") +  
  facet_wrap(~dataset, ncol = 3)
```

