Datasaurus Exercise

The challenge in to gain some insight into the data in the 'Datasaurus Dozen' file. This dataset has 1846 rows of data (and a first row of column headers). It has three columns: dataset, x and y. There dataset column is a text datatype with thirteen distinct values e.g. away, bullseye, circle. The x and y columns are decimal numbers A few sample rows are shown below.

dataset	x	y
dino	55.3846	97.1795
dino	51.5385	96.0256
dino	46.1538	94.4872
dino	42.8205	91.4103
dino	40.7692	88.3333
dino	38.7179	84.8718
dino	35.641	79.8718
dino	33.0769	77.5641

Each of the 'datasets' has the same values of summary statistics, the same mean and standard deviation for both x and y, and the same correlation between x and y. Each also has the same number of rows.

dataset	Average of x	Average of y	Standard deviation of x	Standard deviation of y	Corr_x_y	Number of datasaurus dozen rows
v_lines	54.27	47.84	16.71	26.84	-0.069	142
high_lines	54.27	47.84	16.71	26.84	-0.069	142
bullseye	54.27	47.83	16.71	26.84	-0.069	142
slant_down	54.27	47.84	16.71	26.84	-0.069	142
star	54.27	47.84	16.71	26.84	-0.063	142
circle	54.27	47.84	16.70	26.84	-0.068	142
wide_lines	54.27	47.83	16.71	26.84	-0.067	142
away	54.27	47.83	16.71	26.84	-0.064	142
slant_up	54.27	47.83	16.71	26.84	-0.069	142
dino	54.26	47.83	16.71	26.84	-0.064	142
h_lines	54.26	47.83	16.71	26.84	-0.062	142
dots	54.26	47.84	16.71	26.84	-0.060	142
x_shape	54.26	47.84	16.71	26.84	-0.066	142
Total	**54.27**	**47.84**	**16.71**	**26.84**	**-0.066**	**1846**

Is there any substantial difference between the data in these 13 datasets?

Once you have completed your analysis, you may want to take a look at these two links

https://itsalocke.com/datasaurus/articles/datasaurus

http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html