

Name: Niko Kortström
Student number: 014154573

Exercise 5

First I mapped each needed value into it's own RDD. Then I calculated the Pearson correlations of required values with the function of previous exercise:

Correlation between energy rate and cpu load: 0.00897893228638918

Correlation between energy rate and screen brightness: -2.550540607267431E-4

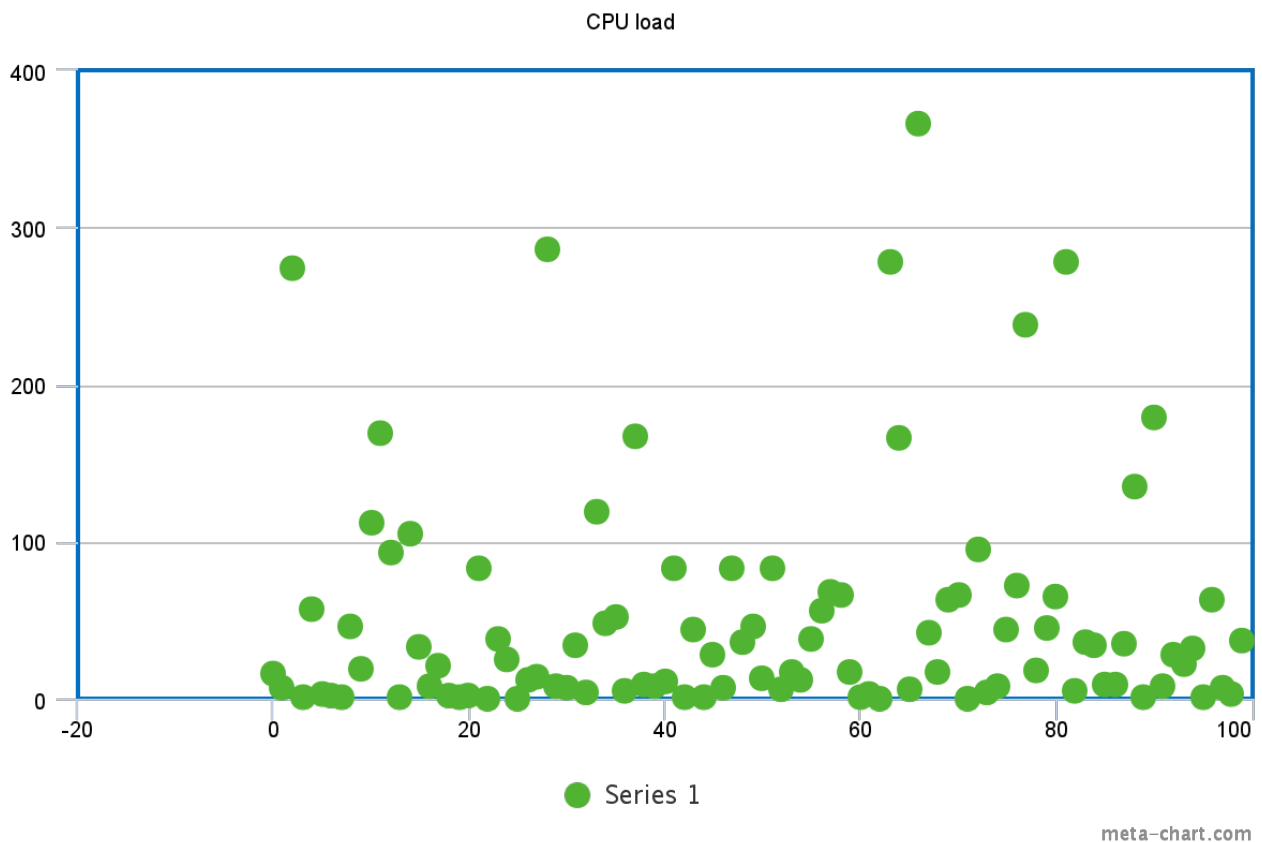
Correlation between energy rate and wifi link speed: 0.002433415320945404

Correlation between energy rate and wifi signal strength: -0.017981212297007488

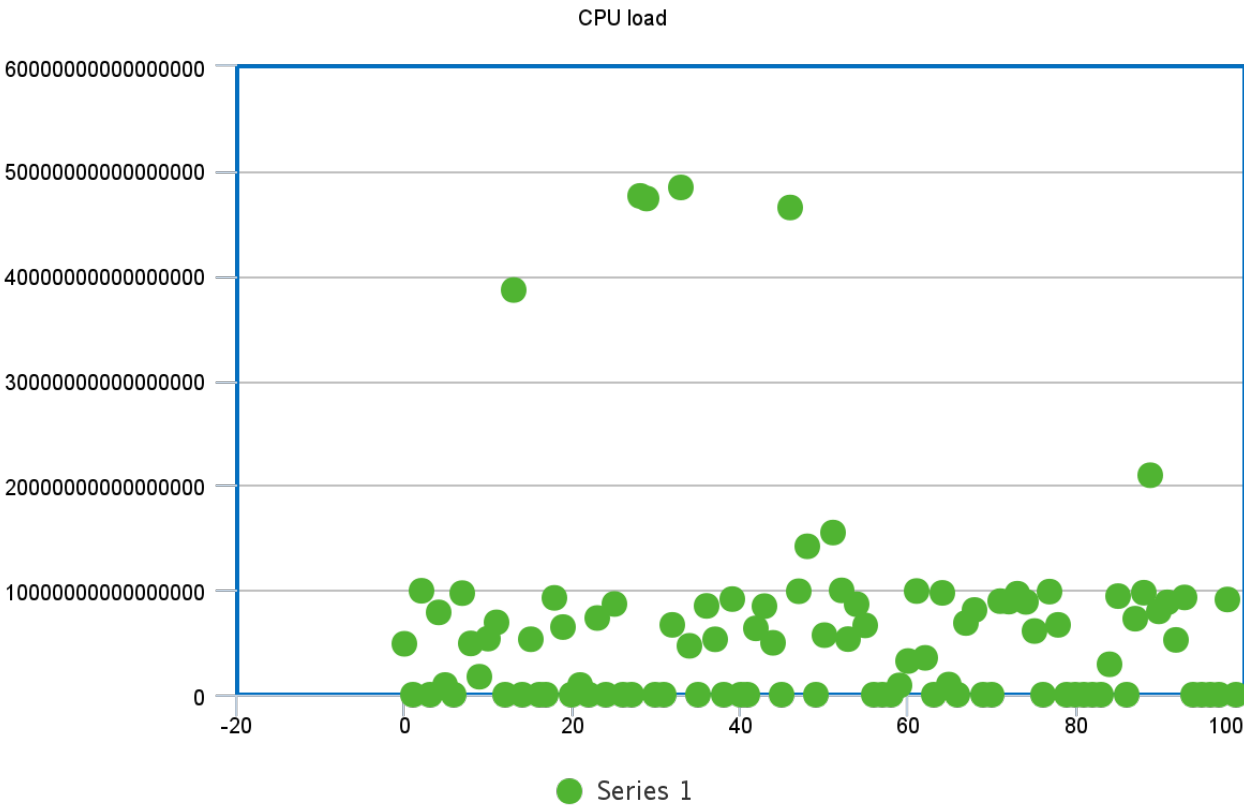
This would indicate that the correlations between these values were really small. The most significant ones are the negative correlation of energy rate and wifi signal strength and the correlation between energy rate and cpu load.

I then used this tool <https://www.meta-chart.com/scatter#> to create the charts for 100 first values of each of these values. I think the charts are hard to look into since all different values are pretty widely scattered. Still the two previously mentioned relationships do pop out as energy rate chart differs clearly the most from wifi signal strength and some similarities can be seen between it and cpu load.

Energy rate:

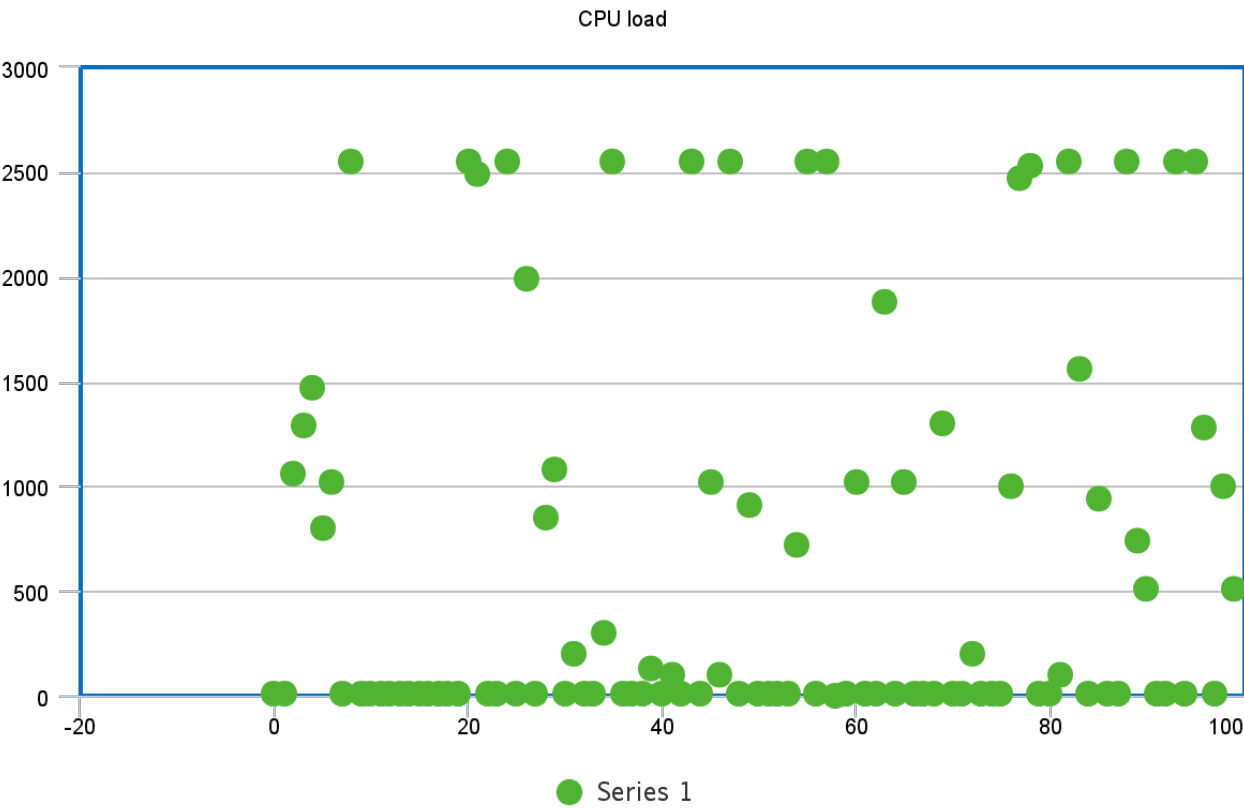


CPU load:



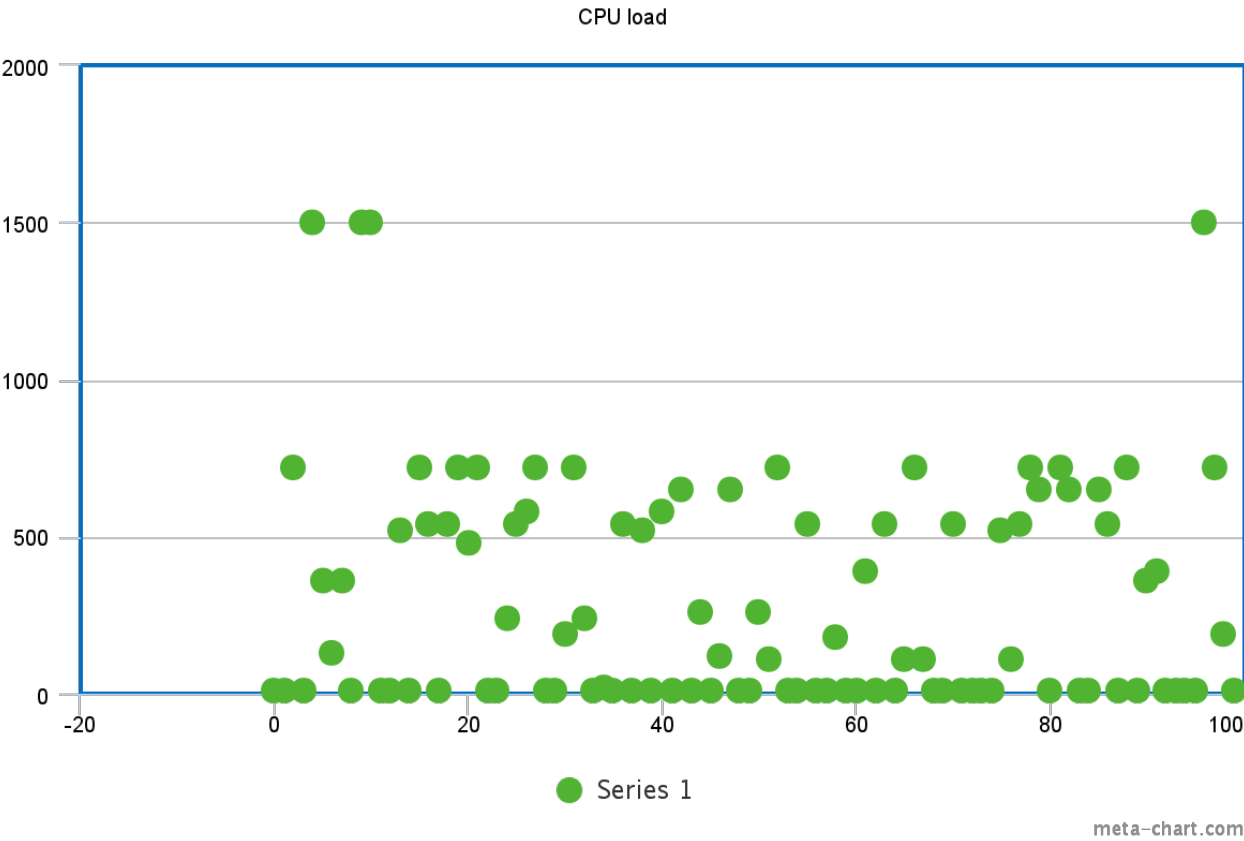
meta-chart.com

Screen brightness:

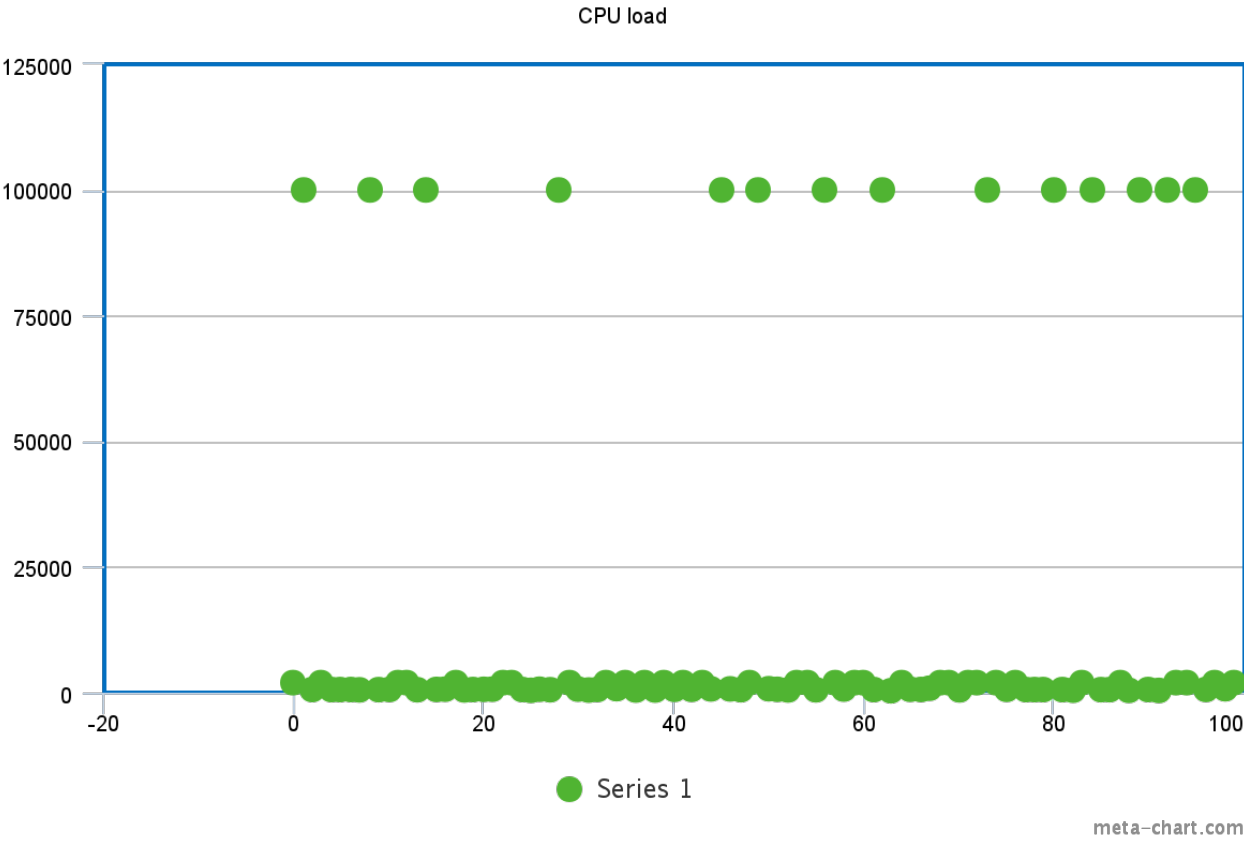


meta-chart.com

Wifi link speed:



Wifi signal strength:



Exercise 6

I again started by parsing the populations needed into their own RDDs. I then created functions for calculating needed new correlations that applied function of exercise 4 to get the correlation coefficient of pairs needed to calculate the new ones. Results were then again written to a text file. They were:

Multiple correlation between energy rate, wifi link speed and wifi signal strength:
0.018195961609565138

Partial correlation between energy rate, wifi link speed and wifi signal strength:
0.002422706199917082

Semi partial correlation between energy rate, wifi link speed and wifi signal strength:
0.002422699026866817