# Simulation with algorithm 2

Sayani Gupta

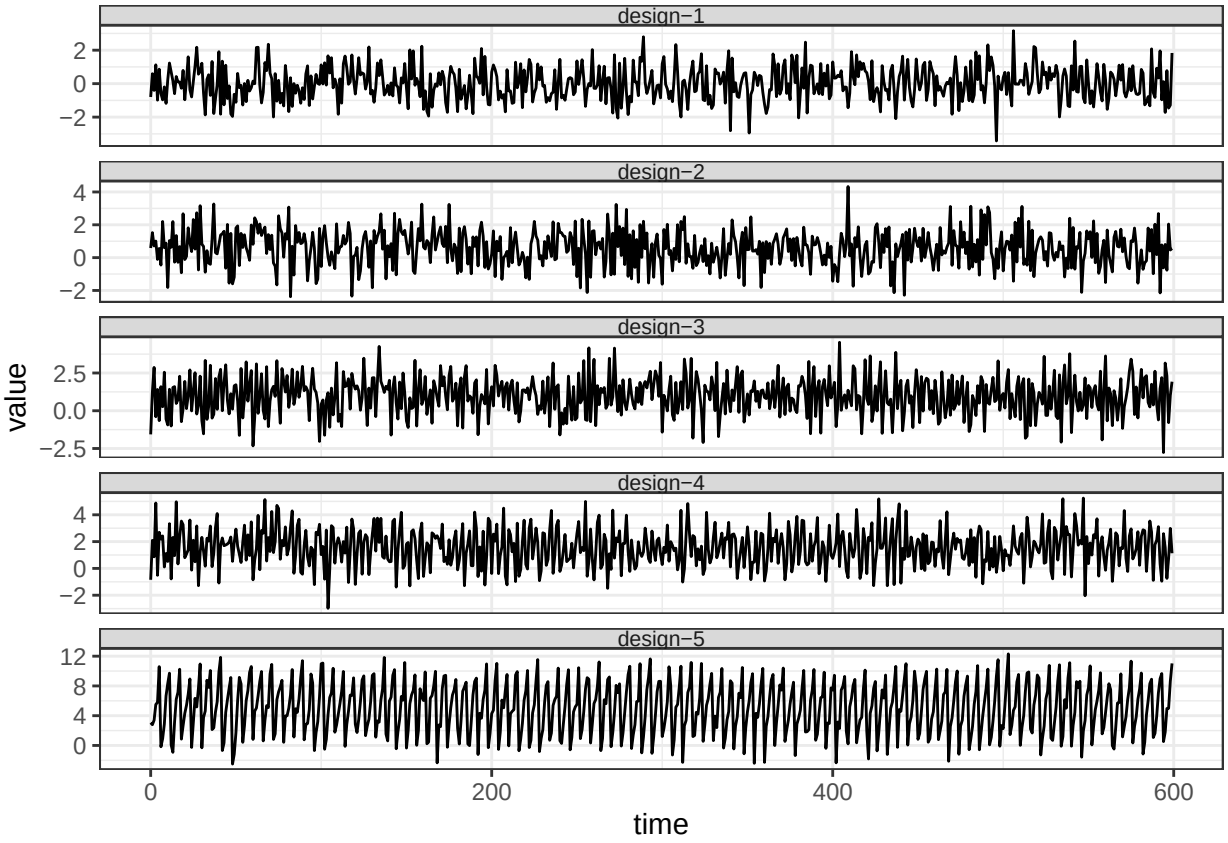
## Simulation setup

Three circular granularities g1, g2 and g3 are considered with levels 2, 3 and 4 respectively. Many time series are created using the four designs below, each of which is iterated five times. We anticipate to have four clusters, each with five time series conforming to the same design, once we execute the clustering.

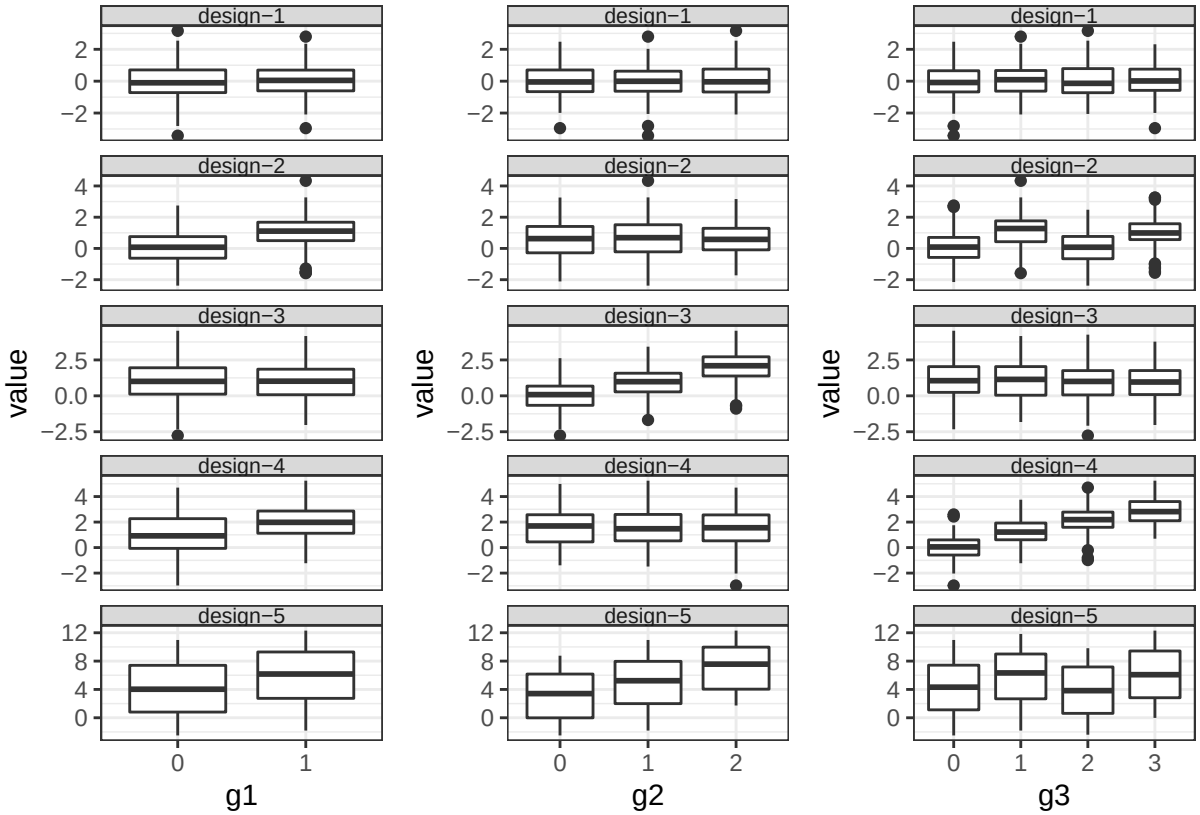
*Algorithm* Compute wpd for each granularity.The distance matrix is then computed with 20 time series and wpd for three granularities as the variable. Euclidean distances are computed and then hierarchical clustering is applied on them. (It is expected that wpd will be close to zero if there is no change in distribution across categories)

| design | g1 | g2 | g3 |
|---|---|---|---|
| design-1 | no | no | no |
| design-2 | yes | no | yes |
| design-3 | no | yes | no |
| design-4 | yes | no | yes |
| design-5 | yes | yes | yes |

## Raw plots


design−1
design−2
design−3
design−4
design−5
value
time
0
200
400
600

## Designs (Distribution of simulated data across different granularities)




## Five iterations of each design (changing seeds)

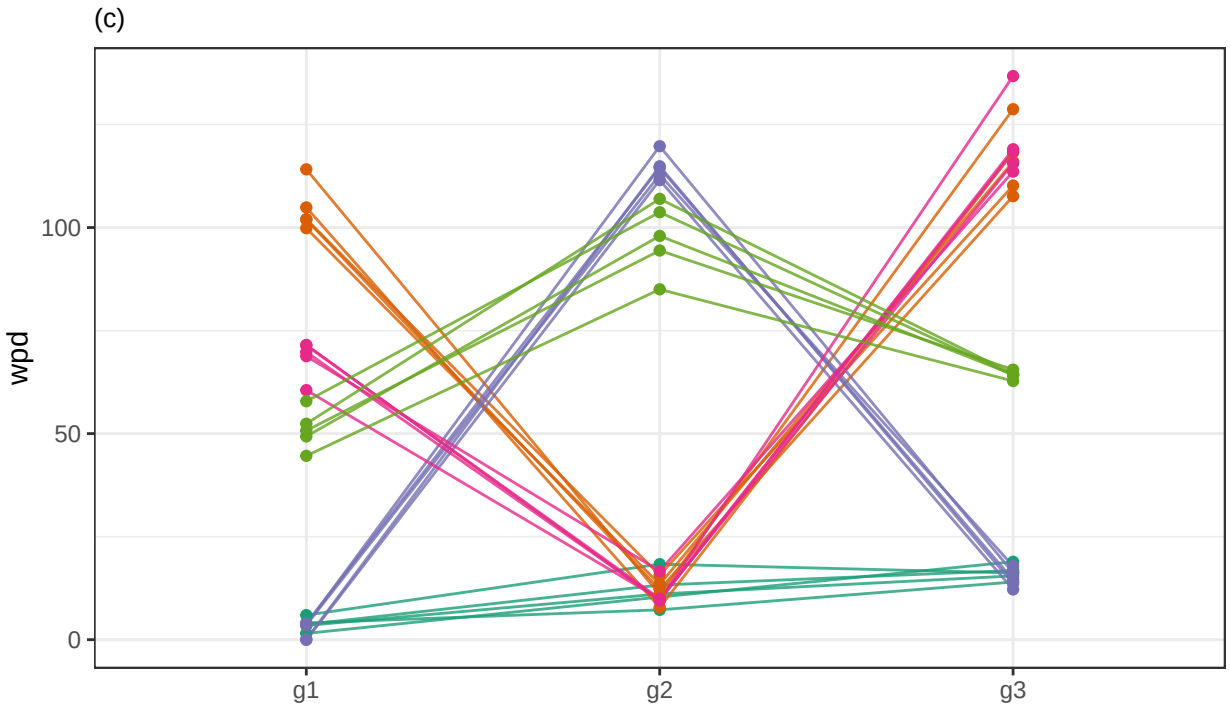
## Table of wpd across designs and wpd

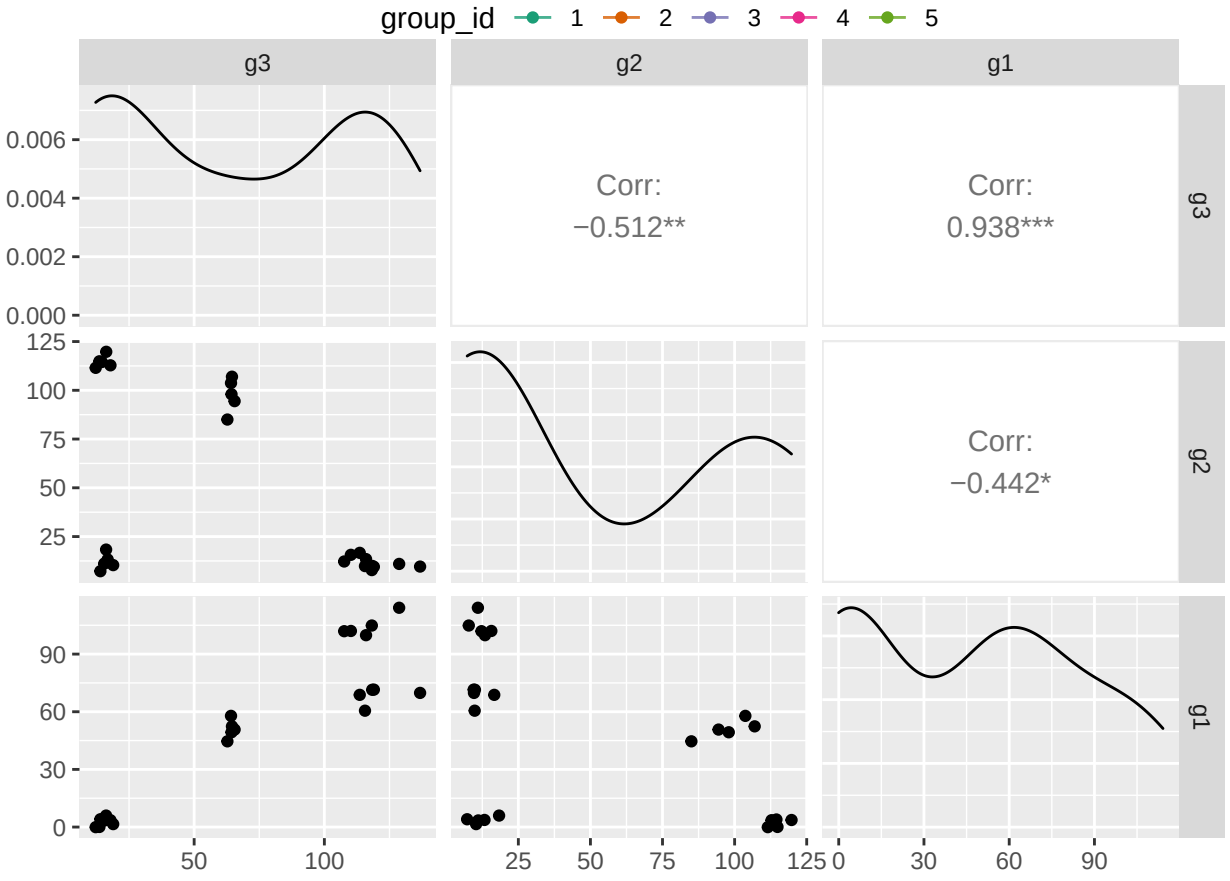
| data_id | g3 | g2 | g1 |
|---|---|---|---|
| design-1-1 | 15.471 | 11.095 | 3.450 |
| design-1-2 | 13.994 | 7.241 | 4.058 |
| design-1-3 | 16.196 | 18.318 | 5.966 |
| design-1-4 | 16.795 | 13.256 | 3.692 |
| design-1-5 | 18.907 | 10.351 | 1.525 |
| design-2-1 | 107.613 | 12.214 | 101.916 |
| design-2-2 | 110.182 | 15.639 | 102.042 |
| design-2-3 | 118.215 | 7.807 | 104.875 |
| design-2-4 | 115.995 | 13.435 | 99.841 |
| design-2-5 | 128.711 | 10.970 | 114.123 |
| design-3-1 | 14.546 | 114.497 | 3.924 |
| design-3-2 | 13.623 | 114.872 | 0.062 |
| design-3-3 | 17.866 | 112.808 | 3.554 |
| design-3-4 | 16.243 | 119.741 | 3.665 |
| design-3-5 | 12.196 | 111.521 | -0.051 |
| design-4-1 | 118.368 | 9.938 | 71.425 |
| design-4-2 | 115.595 | 9.859 | 60.551 |

| data_id | g3 | g2 | g1 |
|---|---|---|---|
| design-4-3 | 113.580 | 16.643 | 68.790 |
| design-4-4 | 136.751 | 9.612 | 69.808 |
| design-4-5 | 118.999 | 9.516 | 71.511 |
| design-5-1 | 64.519 | 106.959 | 52.402 |
| design-5-2 | 64.154 | 103.724 | 57.853 |
| design-5-3 | 64.341 | 97.969 | 49.319 |
| design-5-4 | 62.766 | 85.010 | 44.611 |
| design-5-5 | 65.525 | 94.433 | 50.724 |

## Clustering of designs

| group | data_id |
|---|---|
| 1 | design-1-1 |
| 1 | design-1-2 |
| 1 | design-1-3 |
| 1 | design-1-4 |
| 1 | design-1-5 |
| 2 | design-2-1 |
| 2 | design-2-2 |
| 2 | design-2-3 |
| 2 | design-2-4 |
| 2 | design-2-5 |
| 3 | design-3-1 |
| 3 | design-3-2 |
| 3 | design-3-3 |
| 3 | design-3-4 |
| 3 | design-3-5 |
| 4 | design-4-1 |
| 4 | design-4-2 |
| 4 | design-4-3 |
| 4 | design-4-4 |
| 4 | design-4-5 |
| 5 | design-5-1 |
| 5 | design-5-2 |
| 5 | design-5-3 |
| 5 | design-5-4 |
| 5 | design-5-5 |


(c)
wpd
100
50
0
g1
g2
g3


group_id
1
2
3
4
5
g3
g2
g1
0.006
0.004
0.002
0.000
Corr:
−0.512**
Corr:
0.938***
125
100
75
50
25
Corr:
−0.442*
90
60
30
0
50
100
25
50
75
100
125
0
30
60
90

**First try Euclidean distance**

**Second try Mahanttan and see if cluster separation is better**

**Try PCA before clustering**