Data Visualization HWK 6

Isaiah Thompson Ocansey

2023-04-12

```
df<-read.csv(file = "serialdat.csv",sep=",", header = TRUE)
head(df);dim(df)
```

	SUMOvar <chr>	X10.x.copies <int>	Replicate.1 <dbl>	Replicate.2 <dbl>	Replicate.3 <dbl>	Average.Cq <dbl>
1	S1V1	6	16.27132	16.19231	16.36603	16.27655
2	S1V1	5	20.14263	20.12184	20.05466	20.10638
3	S1V1	4	23.07819	23.10269	22.86079	23.01389
4	S1V1	3	25.53921	25.51511	25.41548	25.48993
5	S1V1	2	26.05758	25.99988	26.04024	26.03257
6	S1V1	1	26.23620	26.03428	26.19077	26.15375

6 rows

```
## [1] 42  6
```

The data contains information about gene variant transcriptions.There were three replicates of the variant transcriptions and a final column where the three replicates were averaged. The categorical variable included is SUMOver-this has seven classes of genes labelled in the format S1V1, S1V2, S1V3, S2V1, S2V2, S3V1,S3V2.In all, the data has 42 observations and 6 variables.For the purposes of visualizing associations, we take two continuous variables which are Replicate 1 and Replicate 2 and observe their associations within the different classes of genes.

```
dat<-df[,c(1,3,4)]
head(dat)
```

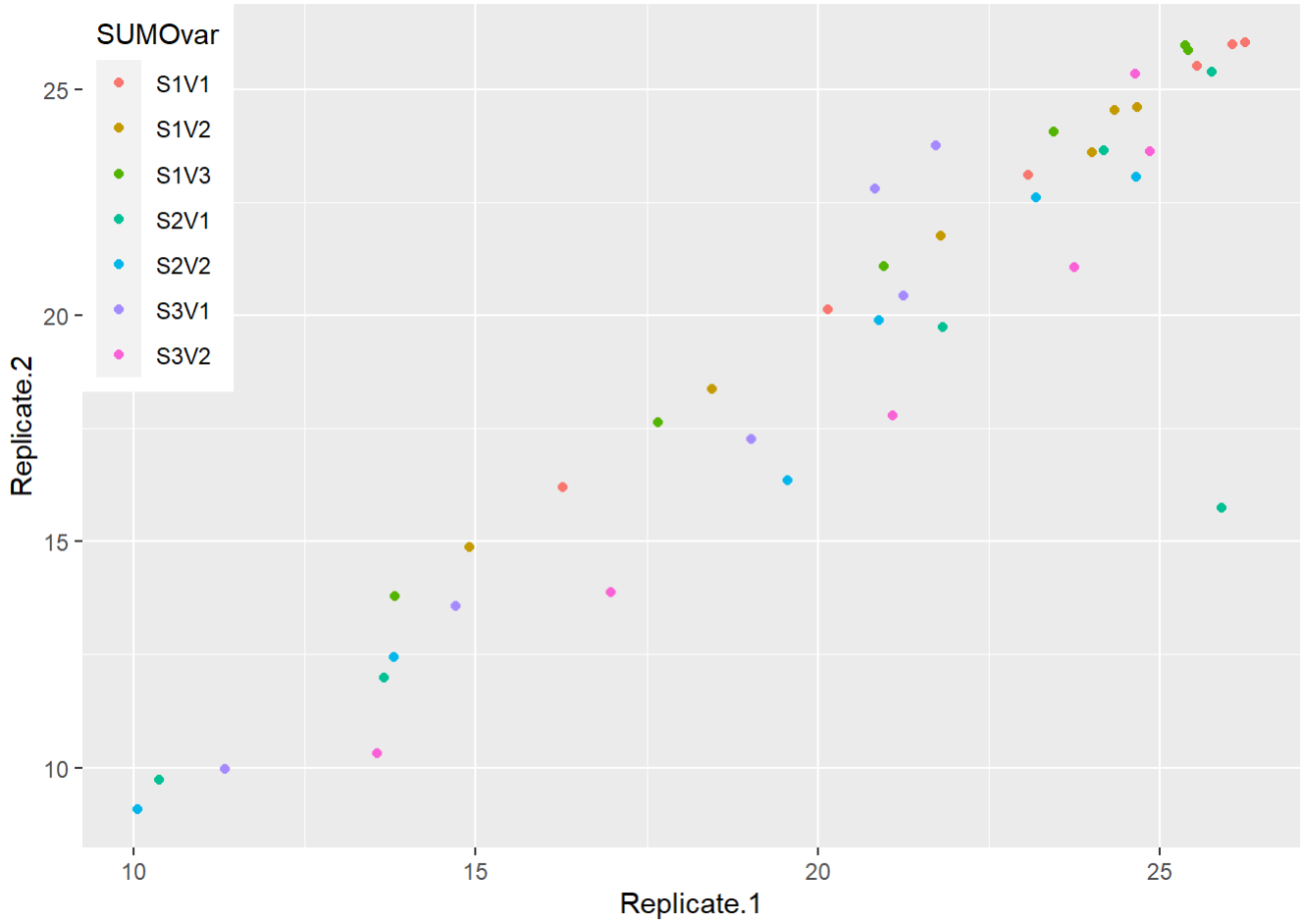
	SUMOvar <chr>	Replicate.1 <dbl>	Replicate.2 <dbl>
1	S1V1	16.27132	16.19231
2	S1V1	20.14263	20.12184
3	S1V1	23.07819	23.10269
4	S1V1	25.53921	25.51511
5	S1V1	26.05758	25.99988
6	S1V1	26.23620	26.03428

6 rows

The above data set is a subset of the original serial data with only two Replicates;1 and 2 and the categorical variable SUMOvar.

```
library(ggplot2)

ggplot(df, aes(x = Replicate.1, y = Replicate.2, color=SUMOvar)) +
  geom_point() +
  #scale_color_manual(values = df$SUMOvar[1:6]) +
  theme(legend.position=c(0,1), legend.justification=c(0,1))
```



It can be observed from the above scatter plot that overall, all the different gene classes has linear associations with replicate 1 and replicate 2 except for some potential outlier of the gene class S2V1.Particularly, We can observe that the different gene classes tend to cluster together more as both replicates increase.

```
rules <- apriori(dat, parameter = list(support = 0.01, confidence = 0.5))
```

```
## Warning: Column(s) 1, 2, 3 not logical or factor. Applying default
## discretization (see '? discretizeDF').
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5    0.01      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[13 item(s), 42 transaction(s)] done [0.00s].
## sorting and recoding items ... [13 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [71 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

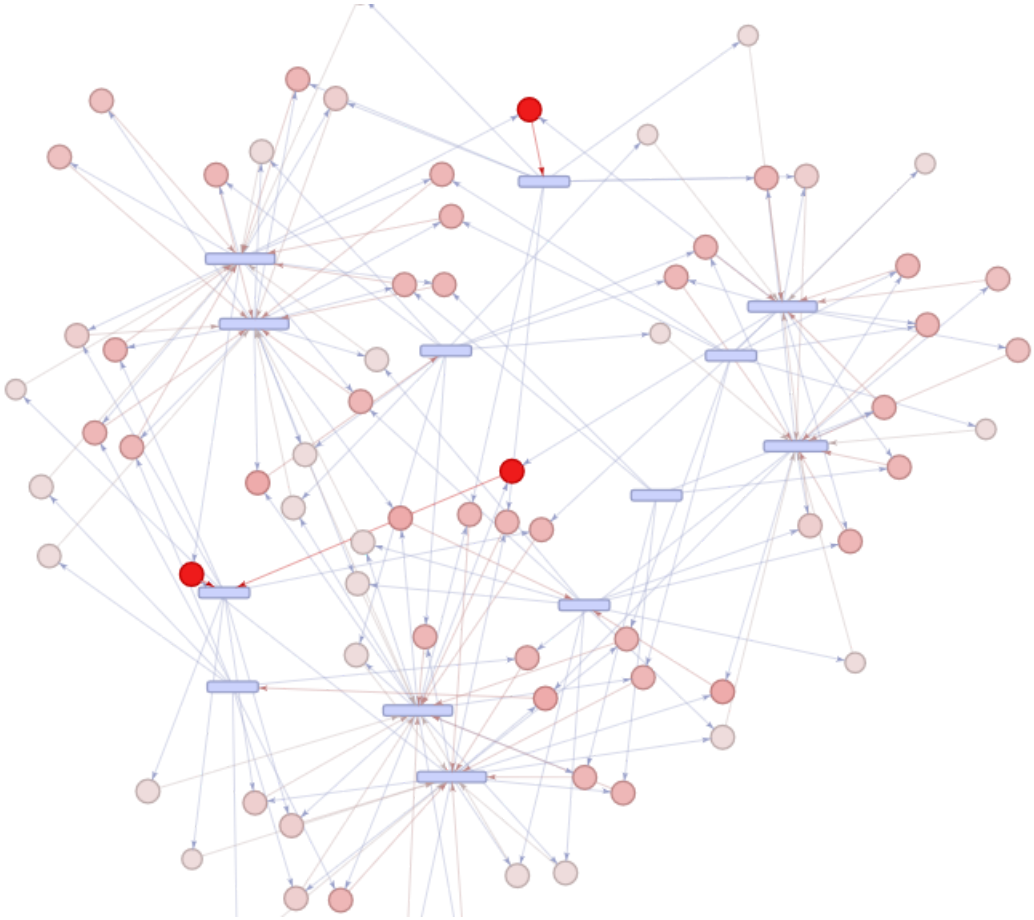
```
rules
```

```
## set of 71 rules
```

we proceed to observe the association rules in both replicates characterized by the gene classes. From the above, we have 71 different association rules

```
plot(rules, method = "graph", engine = "htmlwidget")
```

Select by id

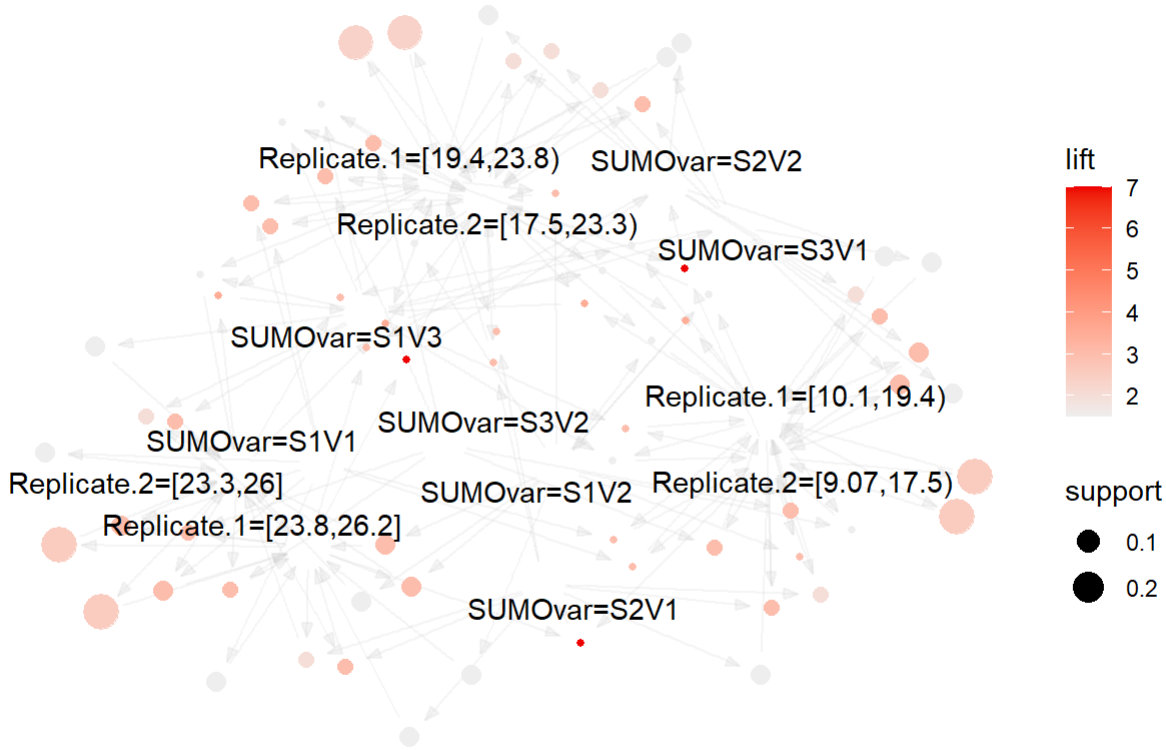


From the above plot, we could view the association rules using either the replicates or the classes of genes.

```
plot(rules, method="graph", control=list(type="itemsets"))
```

```
## Warning: Unknown control parameters: type
```

```
## Available control parameters (with default values):
## layout    =  stress
## circular  =  FALSE
## ggraphdots    =  NULL
## edges     =  <environment>
## nodes     =  <environment>
## nodetext  =  <environment>
## colors    =  c("#EE0000FF", "#EEEEEEFF")
## engine    =  ggplot2
## max   =  100
## verbose   =  FALSE
```


Replicate.1=[19.4,23.8)
SUMOvar=S2V2
Replicate.2=[17.5,23.3)
SUMOvar=S3V1
SUMOvar=S1V3
Replicate.1=[10.1,19.4)
SUMOvar=S3V2
SUMOvar=S1V1
Replicate.2=[23.3,26]
SUMOvar=S1V2
Replicate.2=[9.07,17.5)
Replicate.1=[23.8,26.2]
SUMOvar=S2V1
lift
7
6
5
4
3
2
support
0.1
0.2