error_analysis

W. Evan Durno

February 24, 2015

load data

```
ecoli = read.table("ecoliErrs1.tsv",row.names=NULL)[,-1]
nitma = read.table("nitmaErrs1.tsv",row.names=NULL)[,-1]
pelag = read.table("pelagErrs1.tsv",row.names=NULL)[,-1]
```

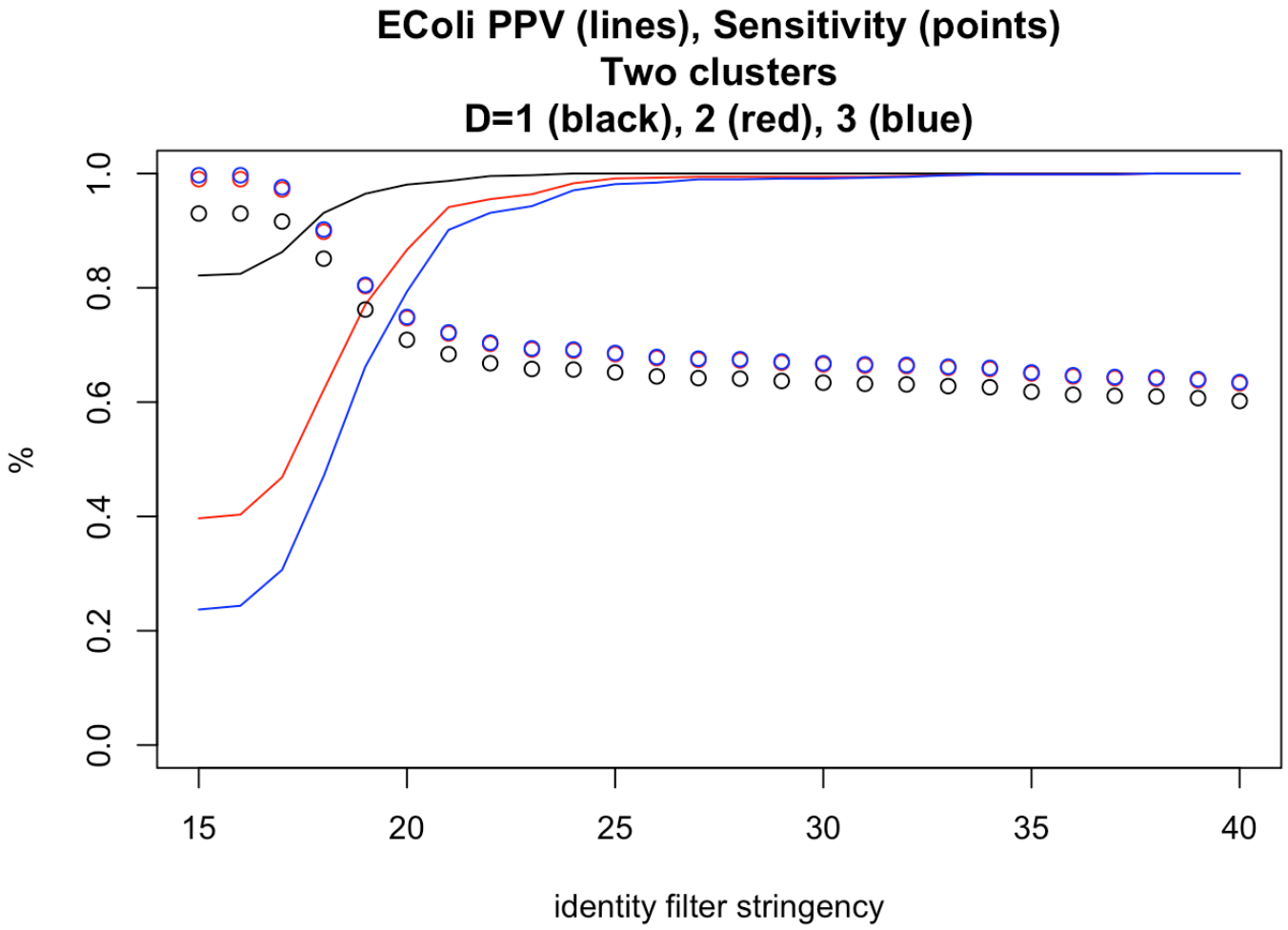
Plotting functions

```
ppv = function(errs,orgStr)
{
idx = 2*(1:156)-1
PPV = errs[idx,]
PPVD1K2 = PPV[ (PPV[,2] == 2) * (PPV[,3] == 1) > 0 , 6 ]
PPVD2K2 = PPV[ (PPV[,2] == 2) * (PPV[,3] == 2) > 0 , 6 ]
PPVD3K2 = PPV[ (PPV[,2] == 2) * (PPV[,3] == 3) > 0 , 6 ]
Sens = errs[idx, 4 ] / rowSums(errs[ idx , 4:5 ])
SensD1K2 = Sens[ (errs[idx,2] == 2) * (errs[idx,3] == 1) > 0 ]
SensD2K2 = Sens[ (errs[idx,2] == 2) * (errs[idx,3] == 2) > 0 ]
SensD3K2 = Sens[ (errs[idx,2] == 2) * (errs[idx,3] == 3) > 0 ]
plot( 15:40 , PPVD1K2 , ylim=c(0,1) , type='l' , ylab="%" , xlab="identity filter str
ingency" , main=paste(orgStr,"PPV (lines), Sensitivity (points)\nTwo clusters\nD=1 (b
lack), 2 (red), 3 (blue)") )
lines( 15:40 , PPVD2K2 , col="red" )
lines( 15:40 , PPVD3K2 , col="blue" )
points( 15:40 , SensD1K2 )
points( 15:40 , SensD2K2 , col="red" )
points( 15:40 , SensD3K2 , col="blue" )
}

npv = function(errs,orgStr)
{
idx = 2*(1:156)
PPV = errs[idx,]
PPVD1K2 = PPV[ (PPV[,2] == 2) * (PPV[,3] == 1) > 0 , 6 ]
PPVD2K2 = PPV[ (PPV[,2] == 2) * (PPV[,3] == 2) > 0 , 6 ]
PPVD3K2 = PPV[ (PPV[,2] == 2) * (PPV[,3] == 3) > 0 , 6 ]
Sens = errs[idx, 5 ] / rowSums(errs[ idx , 4:5 ])
SensD1K2 = Sens[ (errs[idx,2] == 2) * (errs[idx,3] == 1) > 0 ]
SensD2K2 = Sens[ (errs[idx,2] == 2) * (errs[idx,3] == 2) > 0 ]
SensD3K2 = Sens[ (errs[idx,2] == 2) * (errs[idx,3] == 3) > 0 ]
plot( 15:40 , PPVD1K2 , ylim=c(0.975,1) , type='l' , ylab="%" , xlab="identity filter
stringency" , main=paste(orgStr,"NPV (lines), Specificity (points)\nTwo clusters\nD=1
(black), 2 (red), 3 (blue)") )
lines( 15:40 , PPVD2K2 , col="red" )
lines( 15:40 , PPVD3K2 , col="blue" )
points( 15:40 , SensD1K2 )
points( 15:40 , SensD2K2 , col="red" )
points( 15:40 , SensD3K2 , col="blue" )
}
```

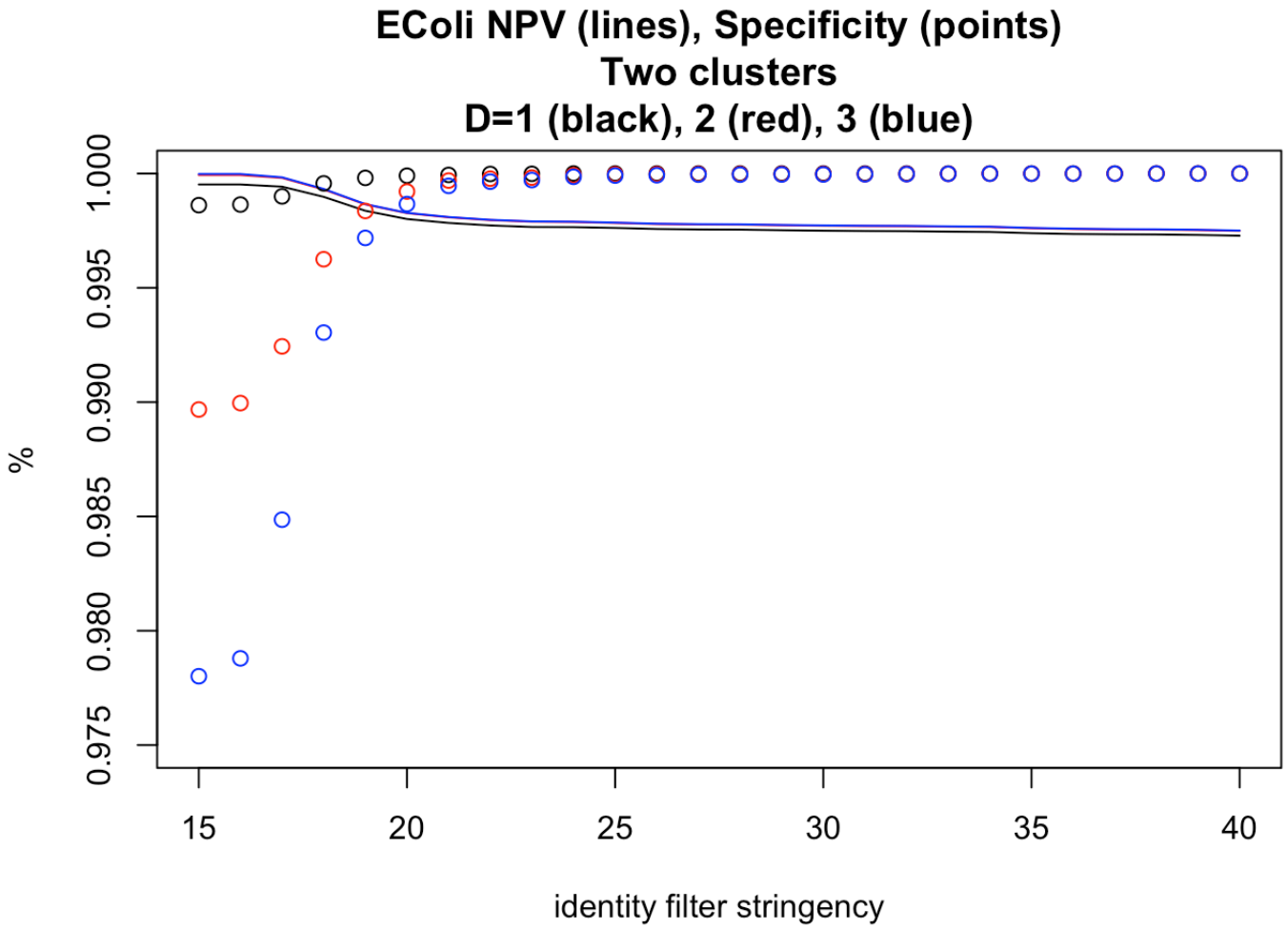
Ecoli PPV and sensitivity

```
ppv(ecoli, "EColi")
```



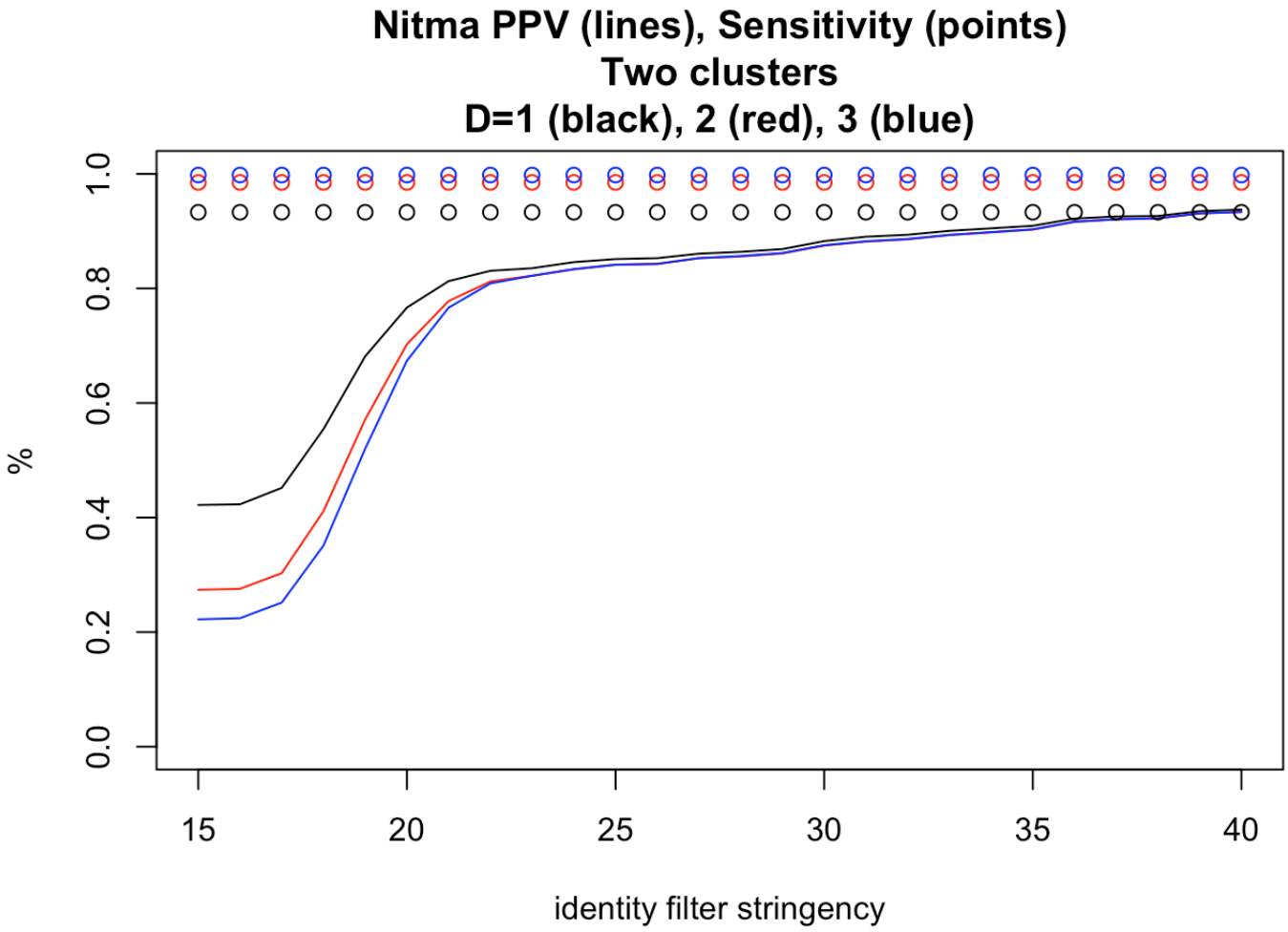
Ecoli NPV and specificity

```
npv(ecoli, "EColi")
```



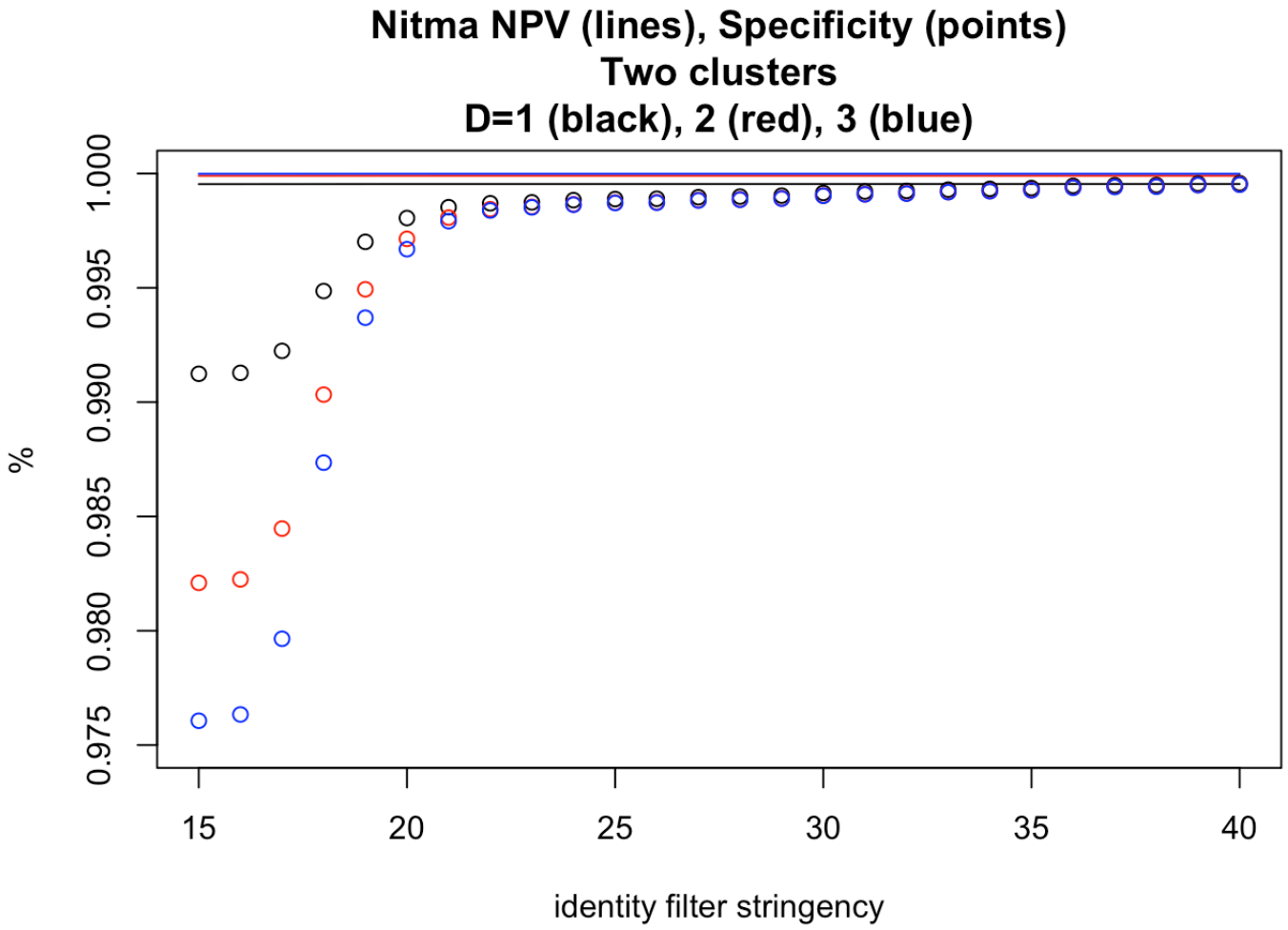
Nitma PPV and sensitivity

```
ppv(nitma,"Nitma")
```



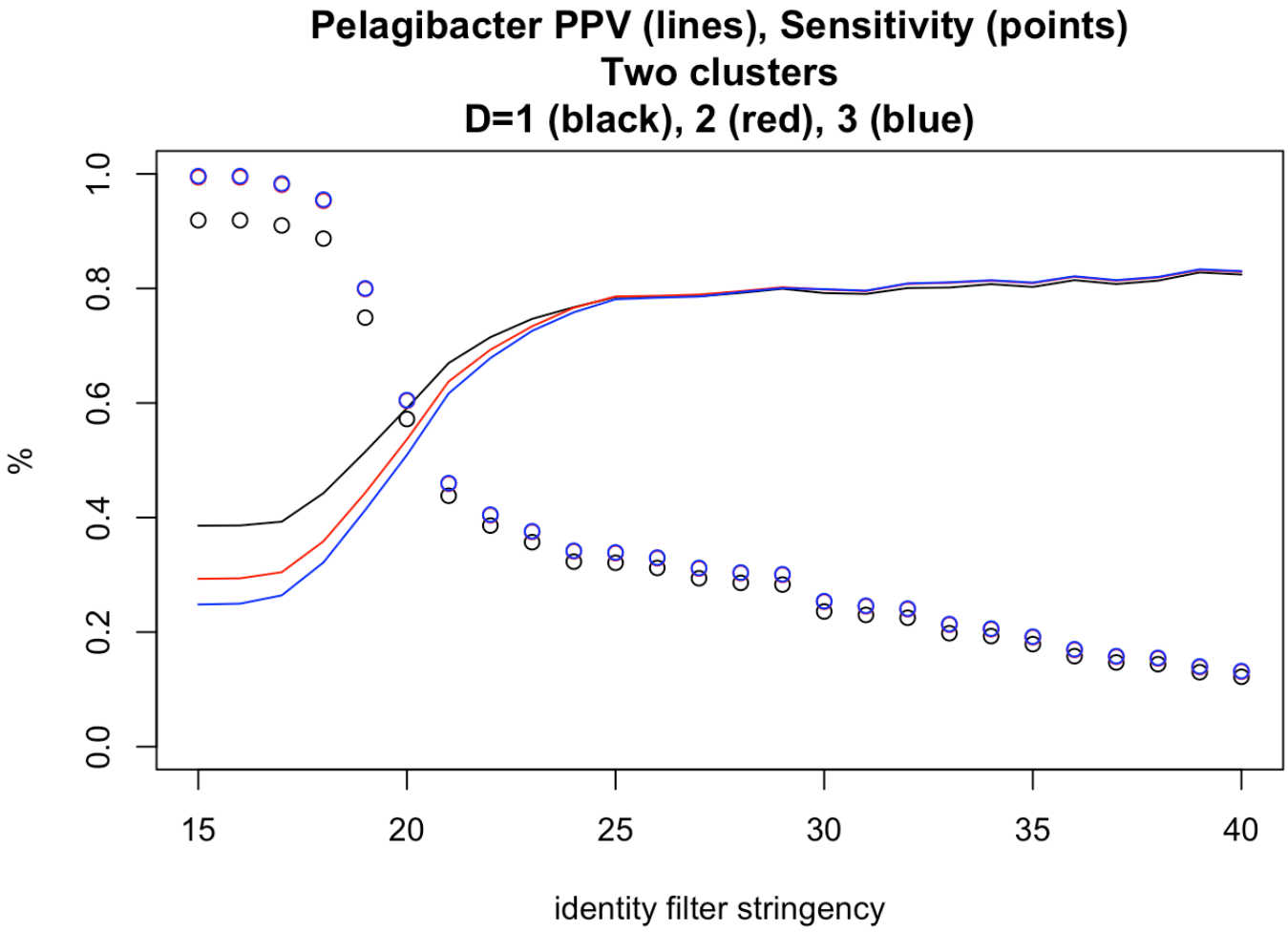
Nitma NPV and specificity

```
npv(nitma,"Nitma")
```



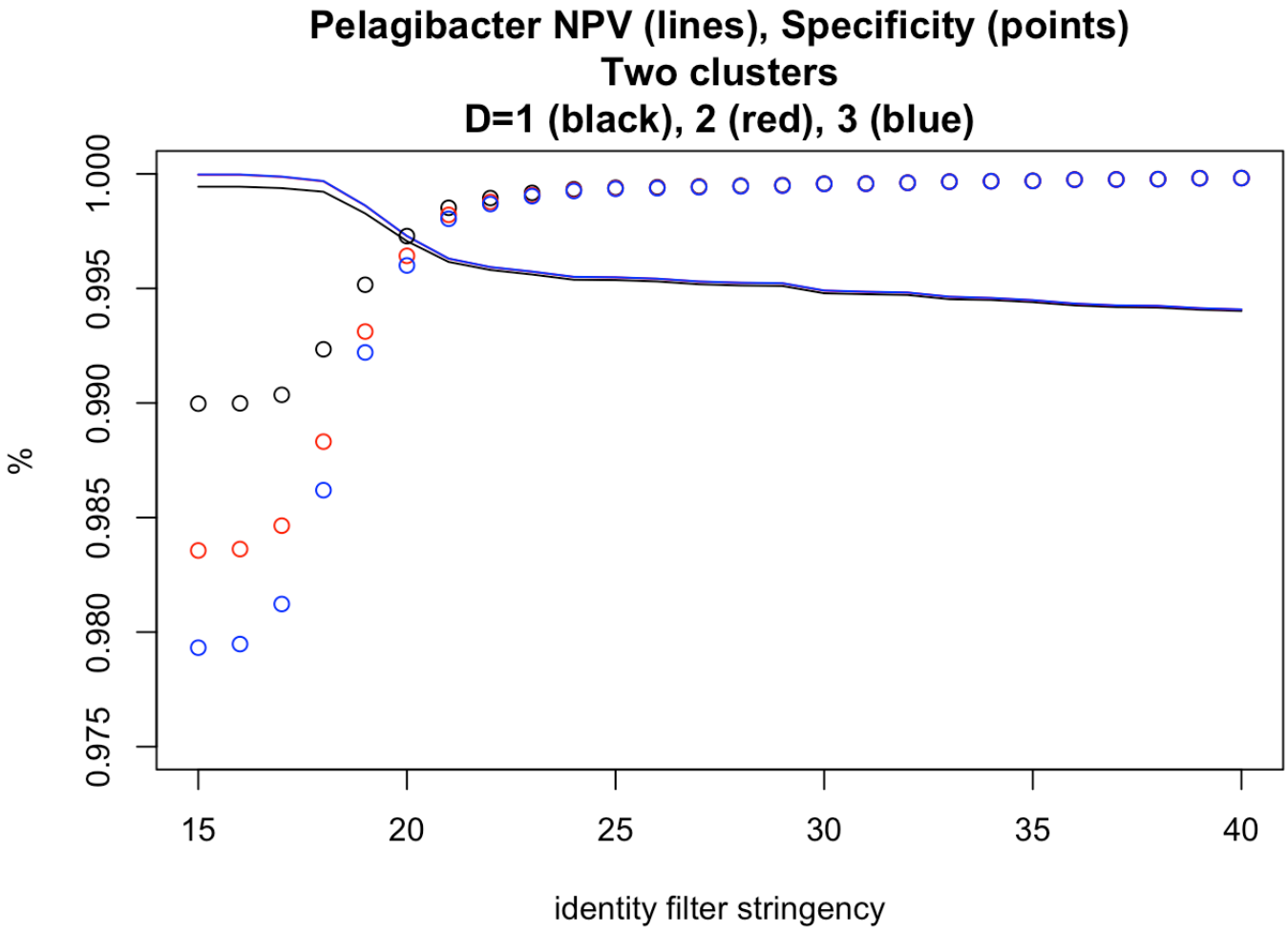
Pelag PPV and sensitivity

```
ppv(pelag,"Pelagibacter")
```



Pelag NPV and specificity

```
npv(pelag,"Pelagibacter")
```


Pelagibacter NPV (lines), Specificity (points)
Two clusters
D=1 (black), 2 (red), 3 (blue)
%
0.975
0.980
0.985
0.990
0.995
1.000
15
20
25
30
35
40
identity filter stringency