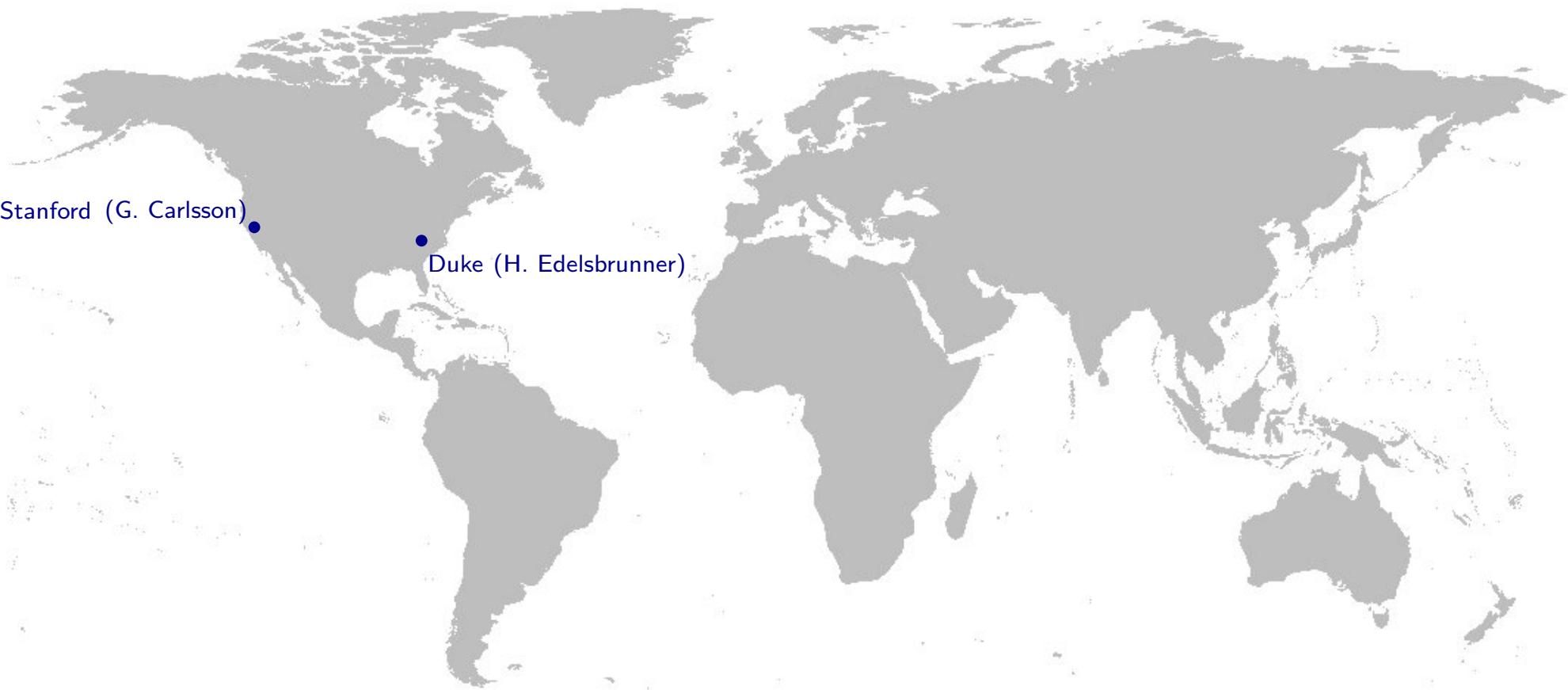
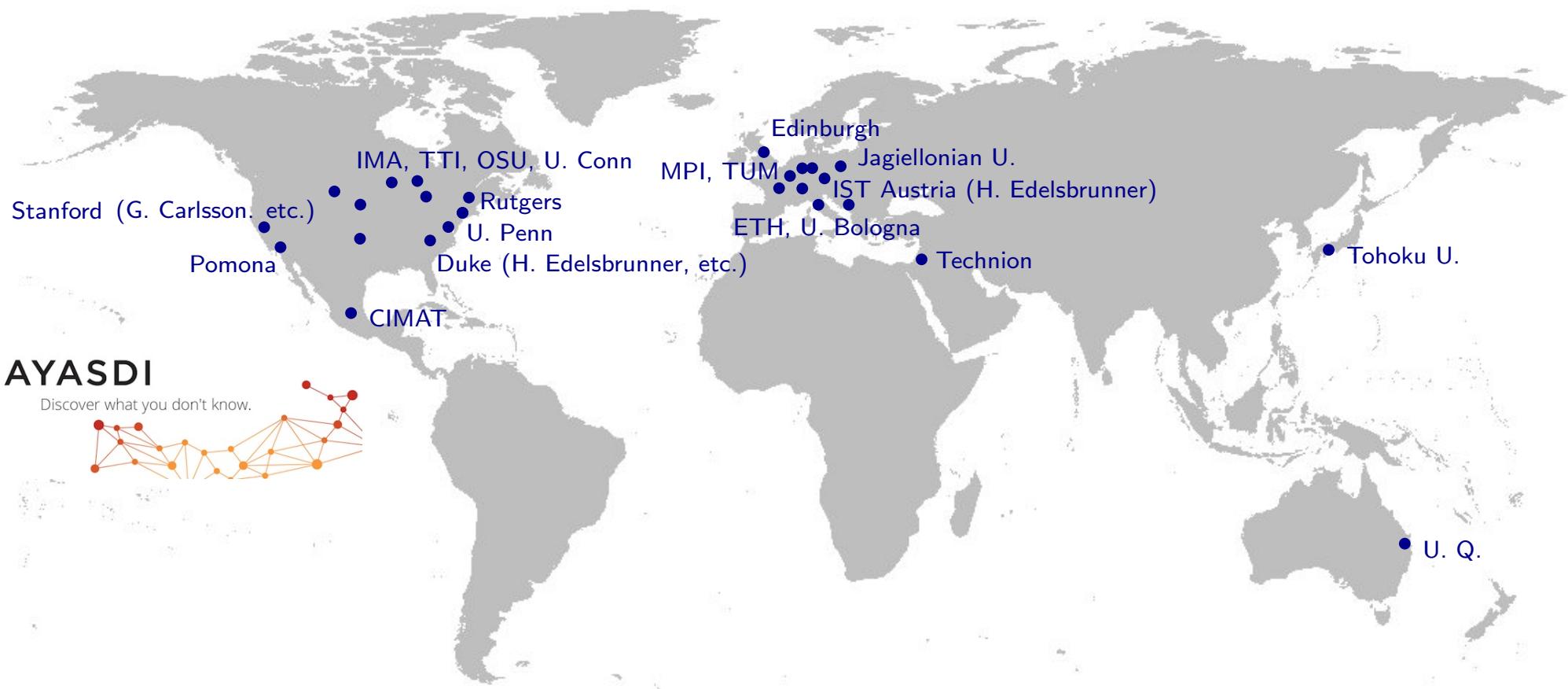


The TDA community (as of 2002)



- 2 research groups (5-10 researchers)

The TDA community (as of 2016)



- 50-100 researchers working on theoretical foundations
- 200-300 researchers at the interface with applications
- very successful applications and company (Ayasdi)

Mapper

Eurographics Symposium on Point-Based Graphics (2007)
M. Botsch, R. Pajarola (Editors)

Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition

A Original Point Cloud

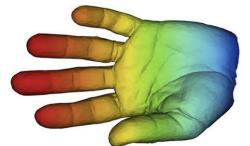


Gurjeet Singh¹, Facundo Mémoli² and Gunnar Carlsson^{†2}

¹ Institute for Computational and Mathematical Engineering, Stanford University, California, USA.

² Department of Mathematics, Stanford University, California, USA.

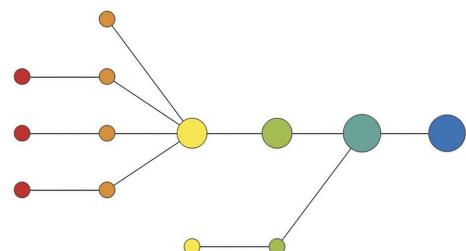
B Coloring by filter value



C Binning by filter value



D Clustering and network construction



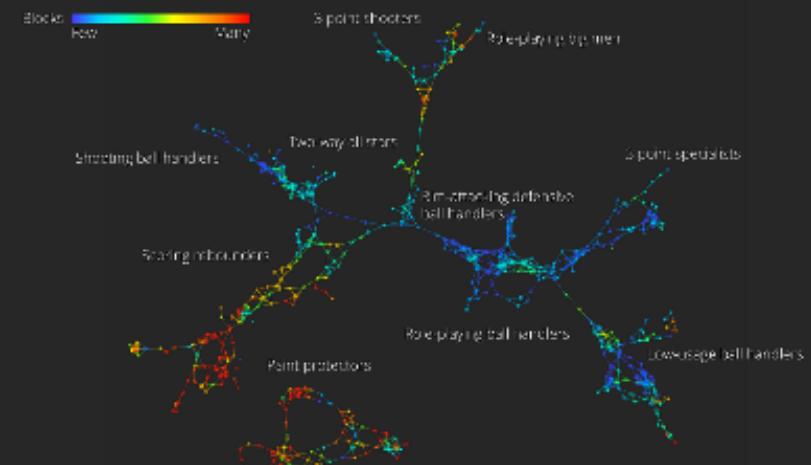
Abstract

We present a computational method for extracting simple descriptions of high dimensional data sets in the form of simplicial complexes. Our method, called Mapper, is based on the idea of partial clustering of the data guided by a set of functions defined on the data. The proposed method is not dependent on any particular clustering algorithm, i.e. any clustering algorithm may be used with Mapper. We implement this method and present a few sample applications in which simple descriptions of the data present important information about its structure.

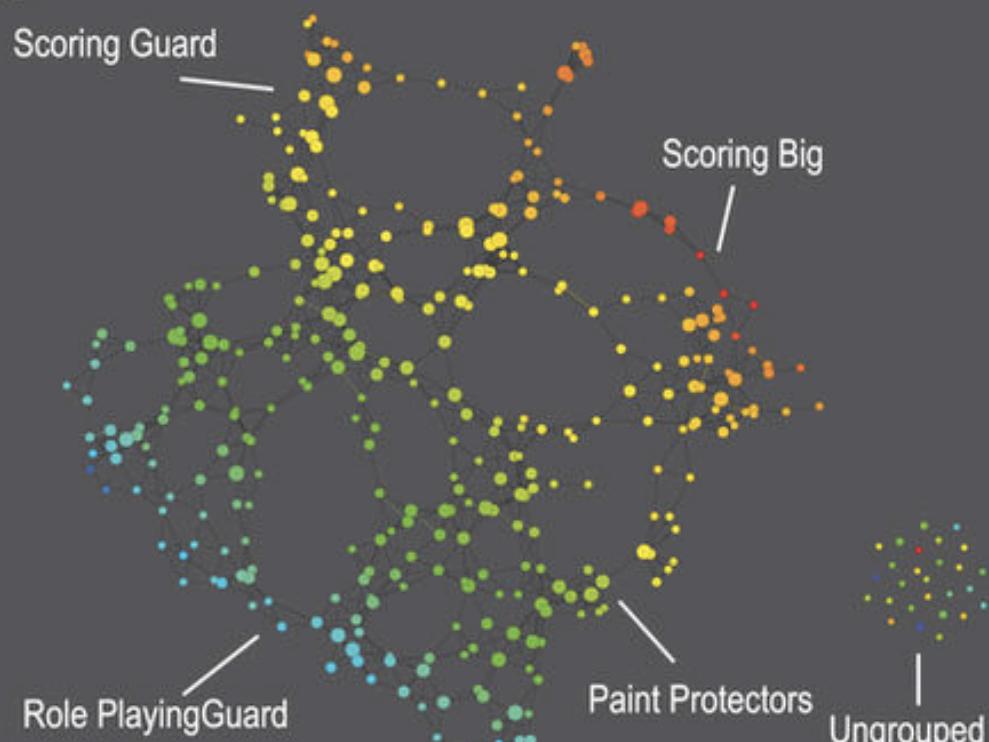
Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modelling.

10 Positions of Basketball

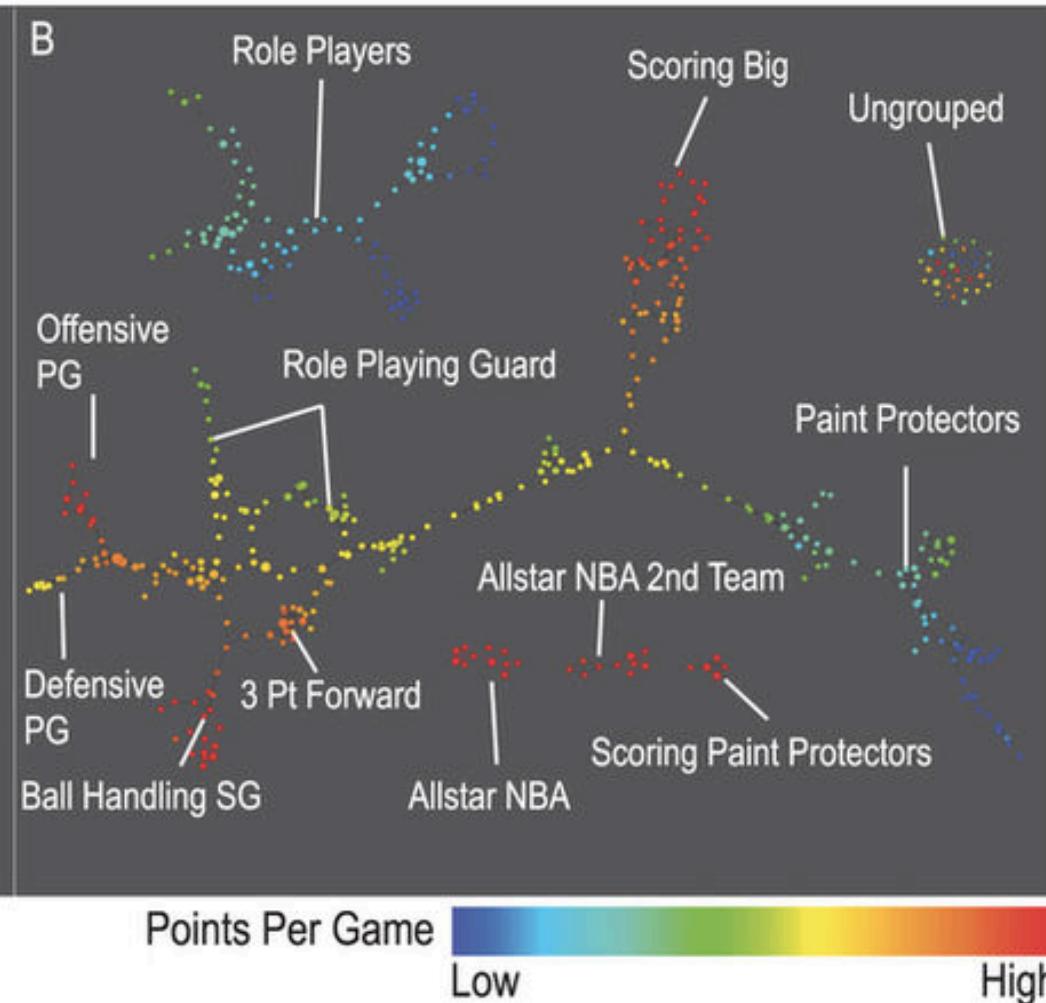
AYASDI



A

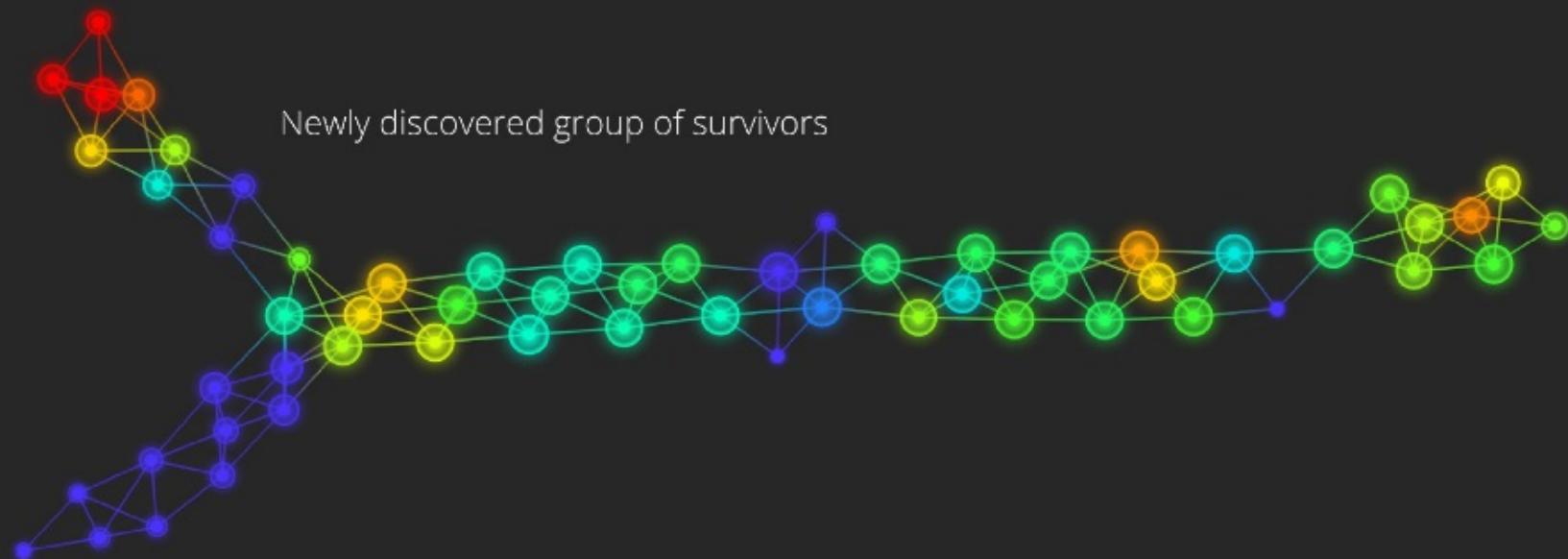


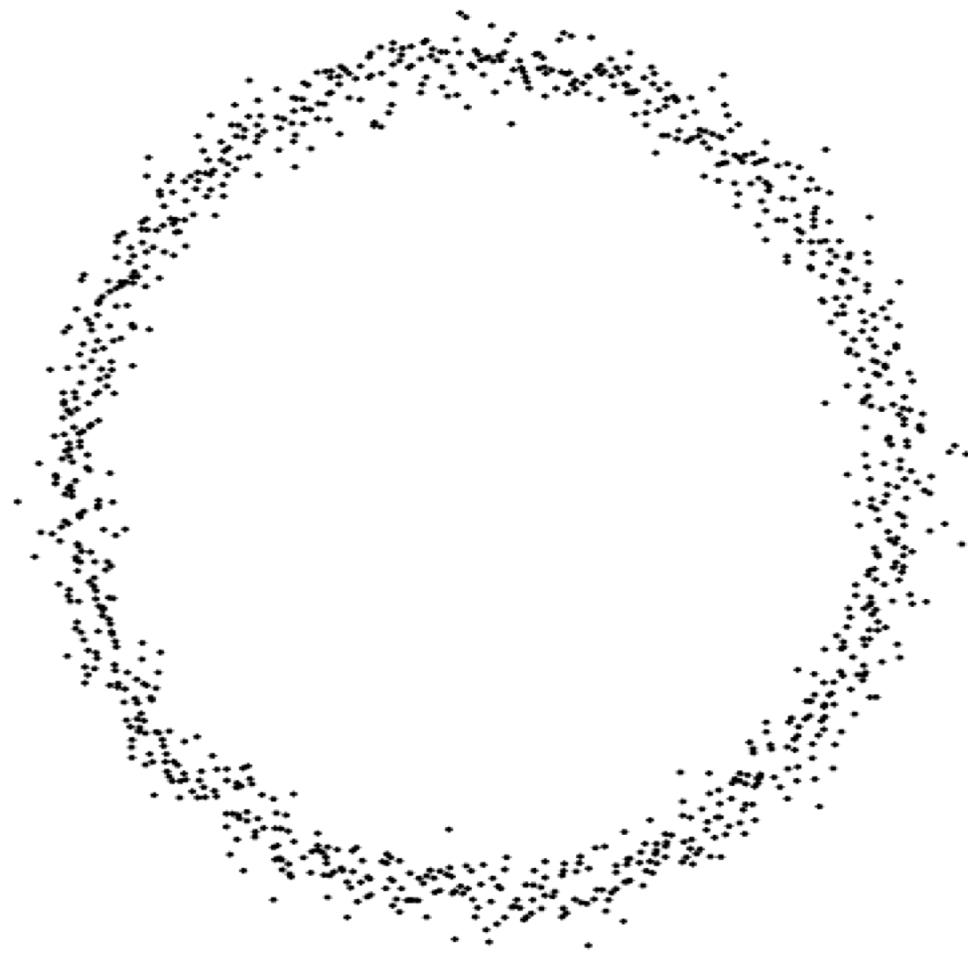
B

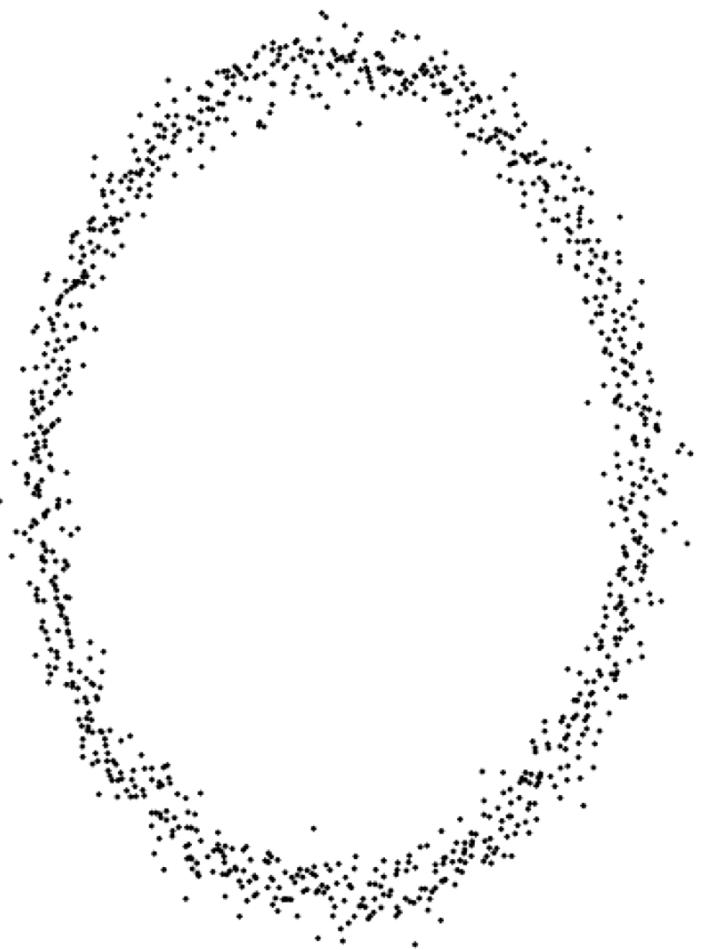


Mortality

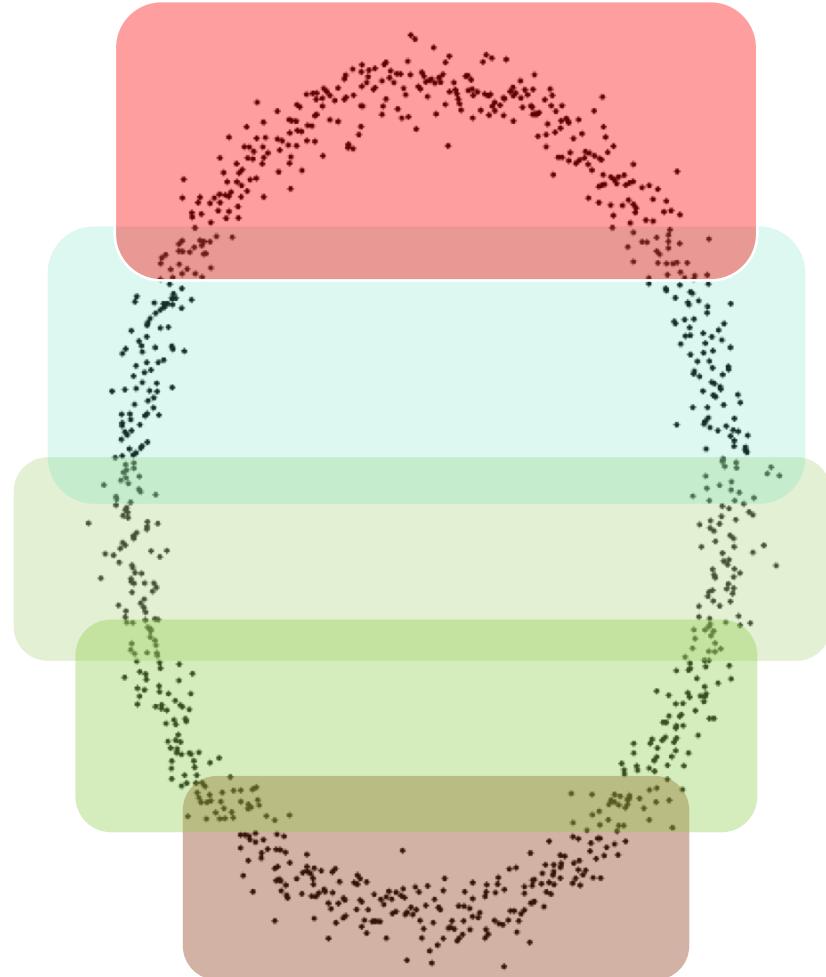
Low High

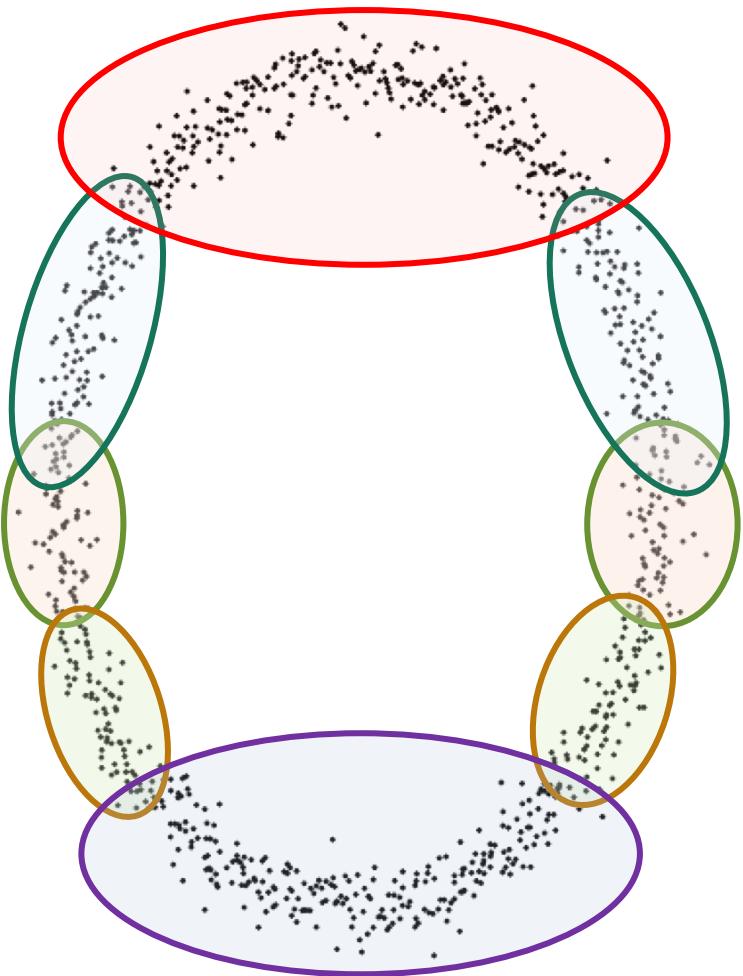
A horizontal color bar representing mortality risk. It transitions from blue on the left (labeled 'Low') to red on the right (labeled 'High').



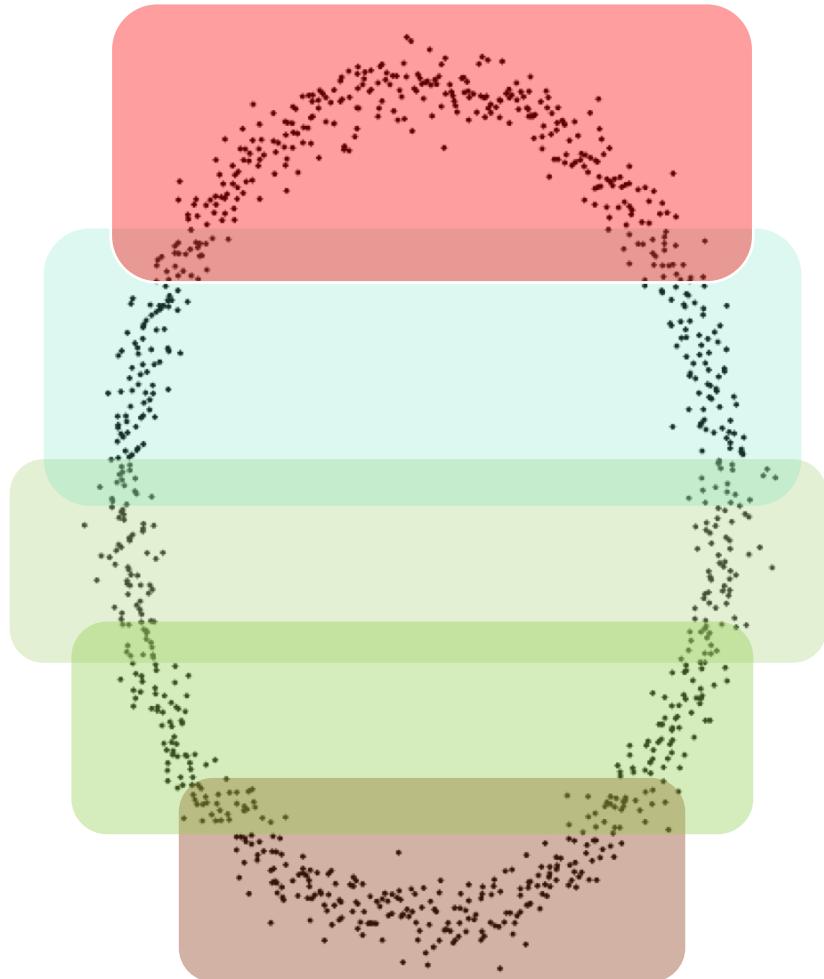


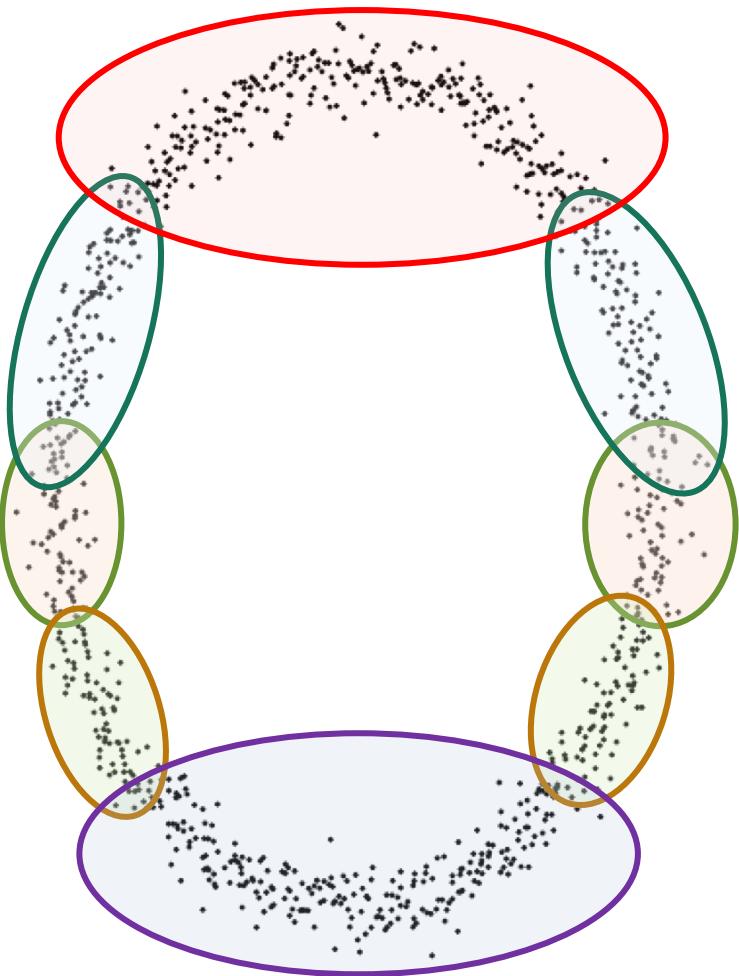
Divide into
bins



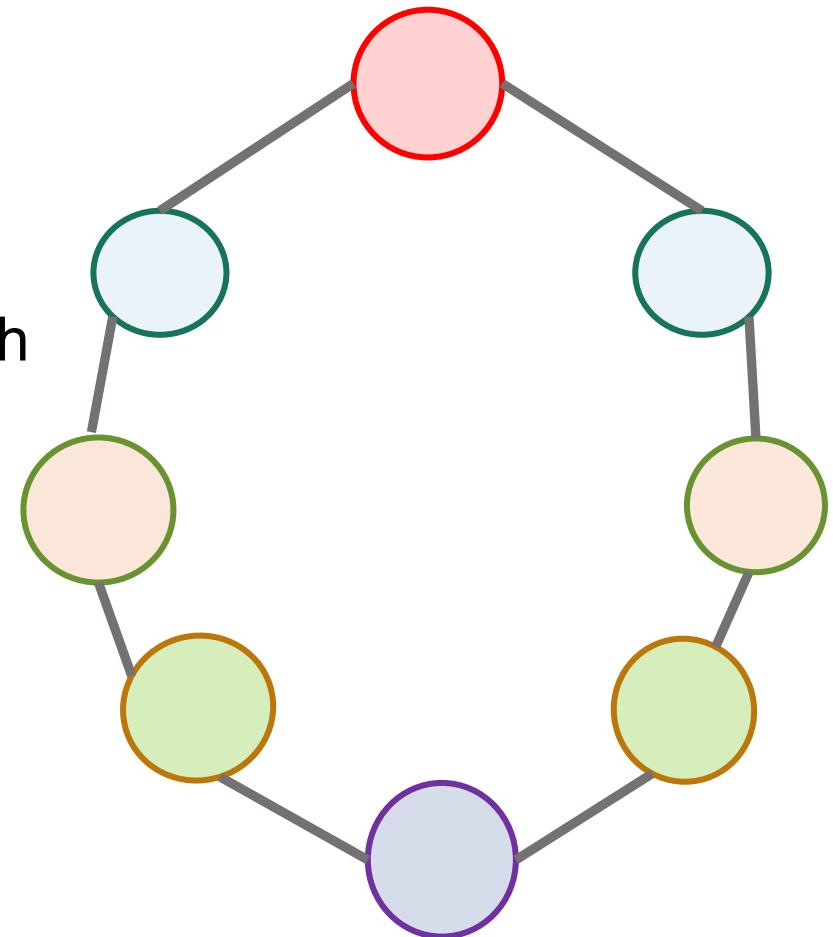


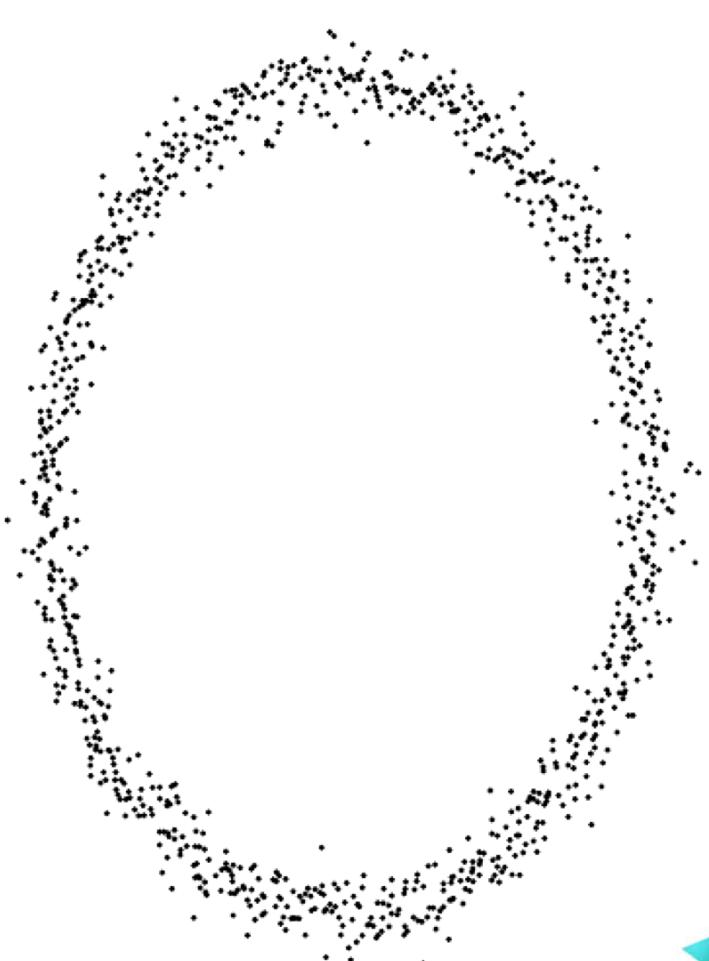
Cluster
each bin



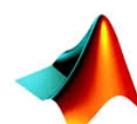


Obtain a graph
from clusters

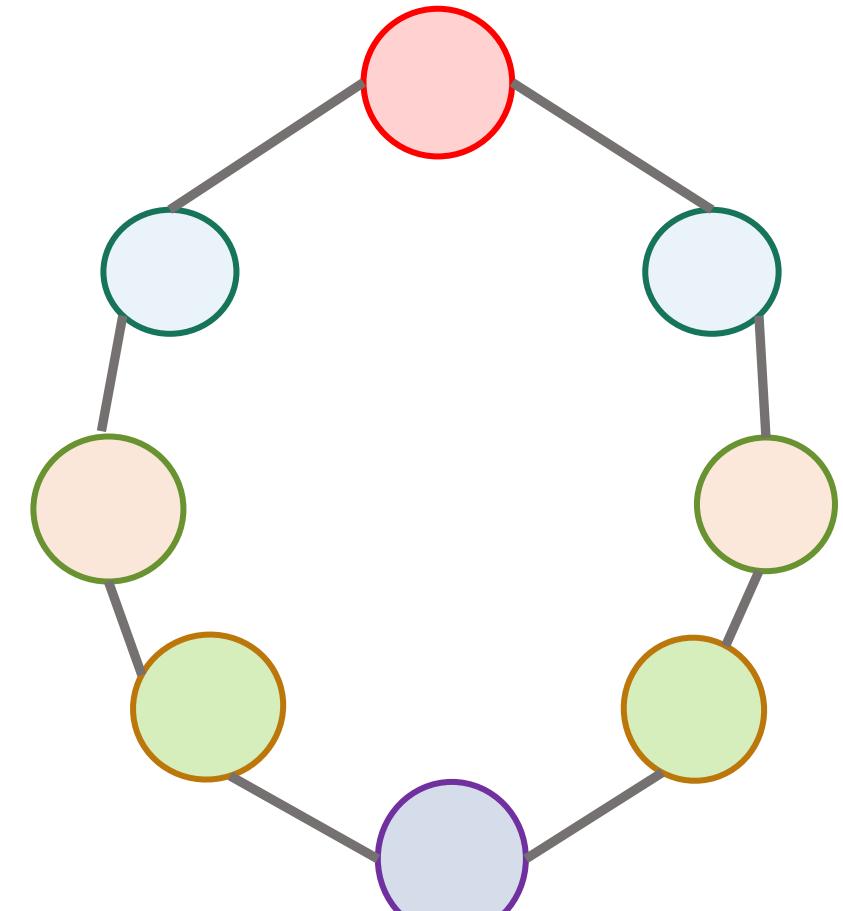




Analyze

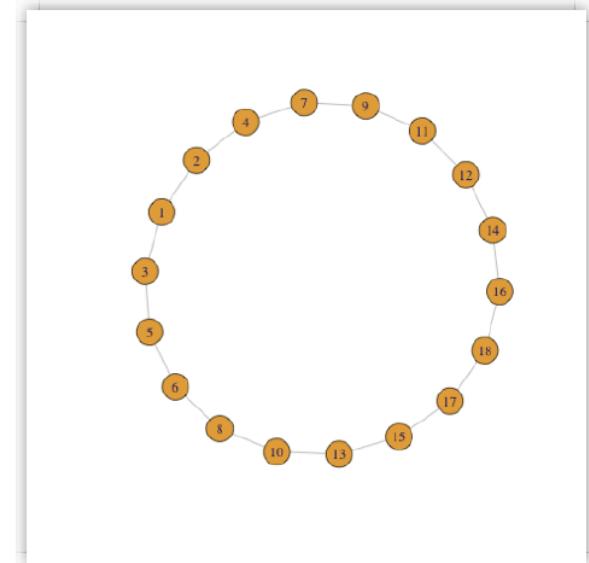
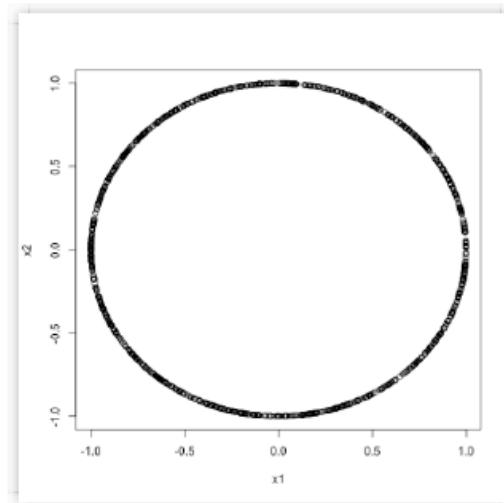


MATLAB®



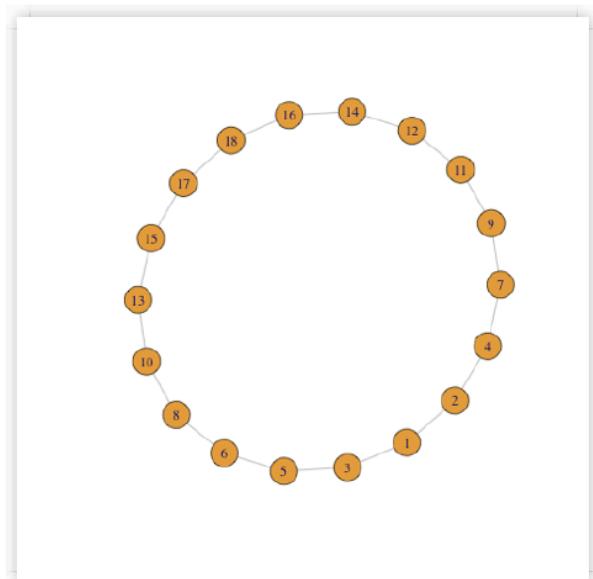
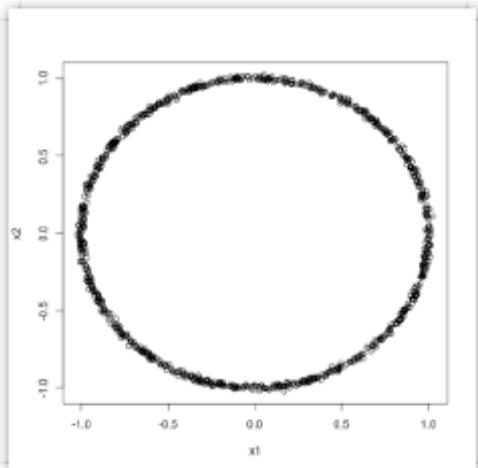
mapper1D Circulo

```
plot(datos_circulo_uniforme)
```



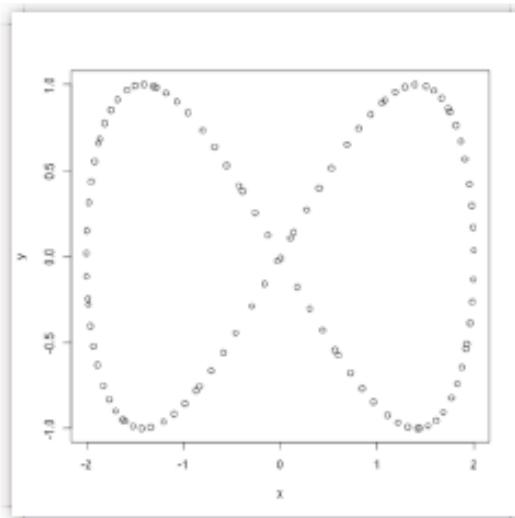
mapper1D Circulo con ruido

```
plot(datos_circulo_con_ruido)
```



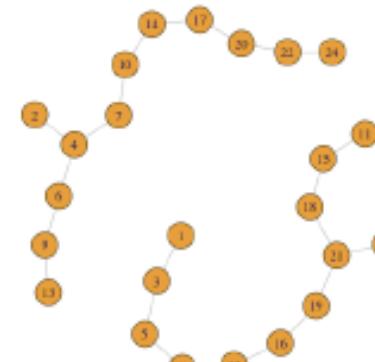
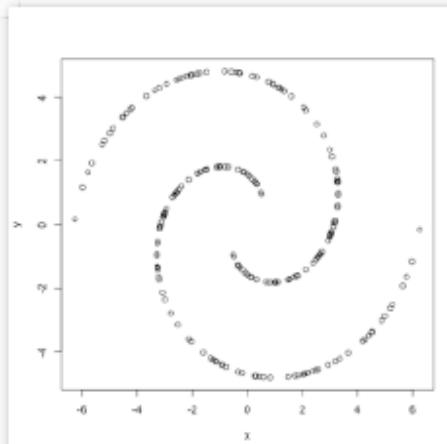
mapper1D Ocho

```
plot(datos_ocho)
```



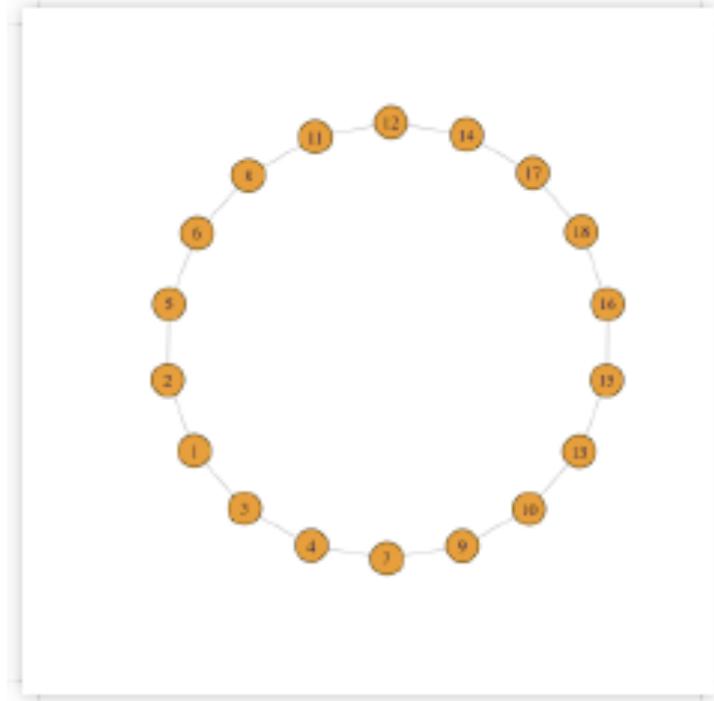
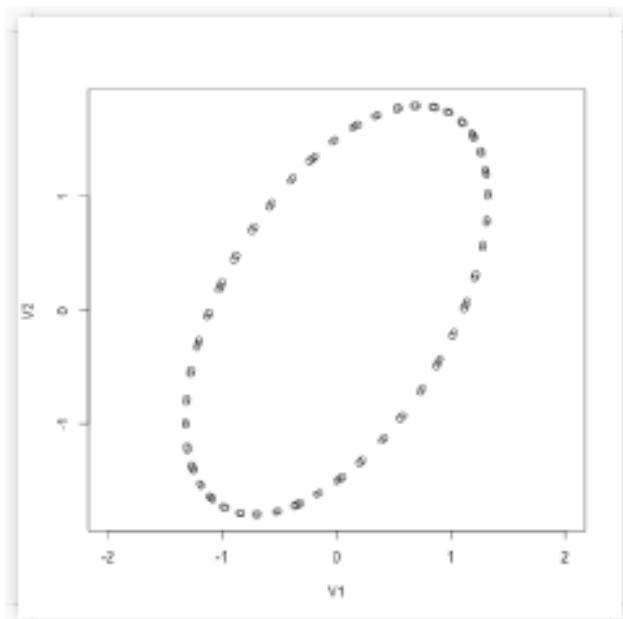
mapper1D Dos Espirales

```
plot(datos_espirales)
```



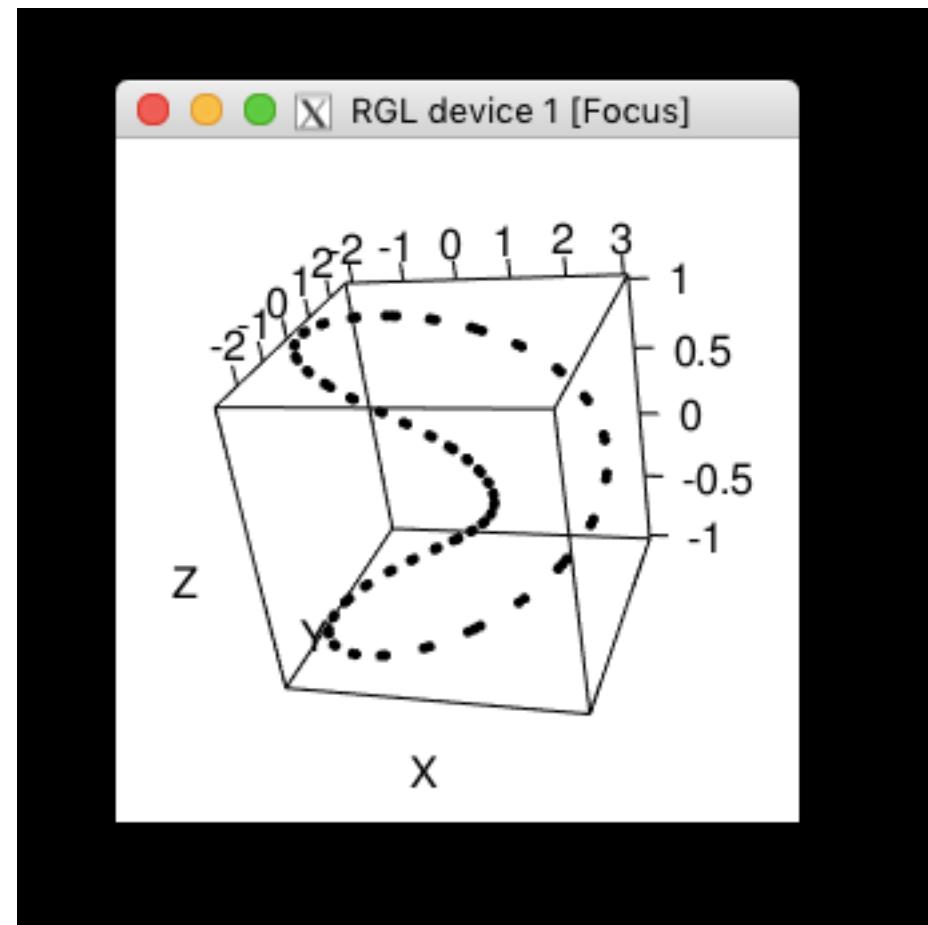
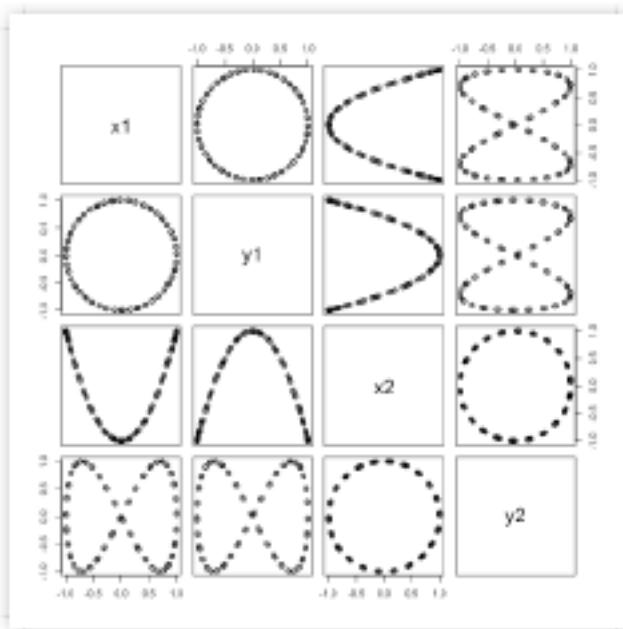
mapper1D Transformaciones lineales

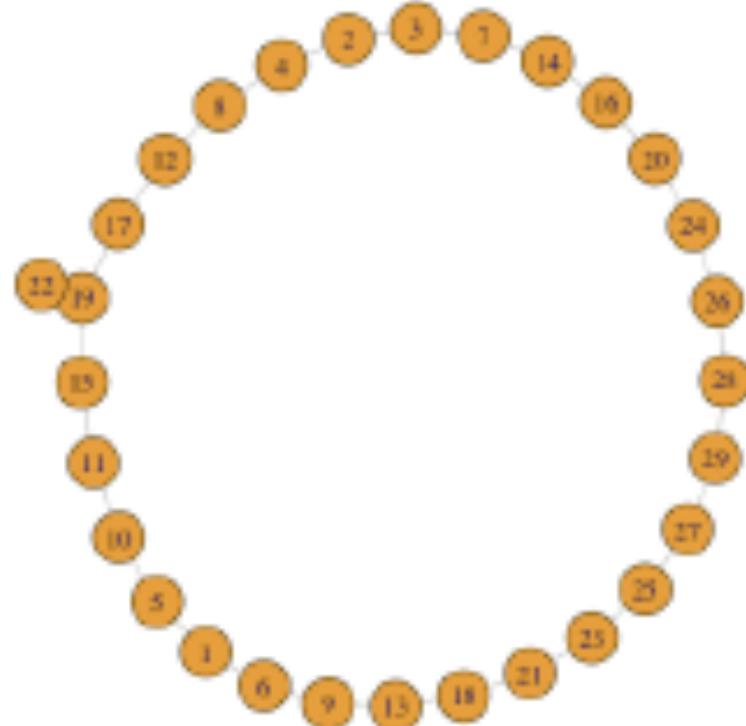
```
plot(datos_transformados_elipse,  
asp=1) # note M is the image of  
circle C under map A
```

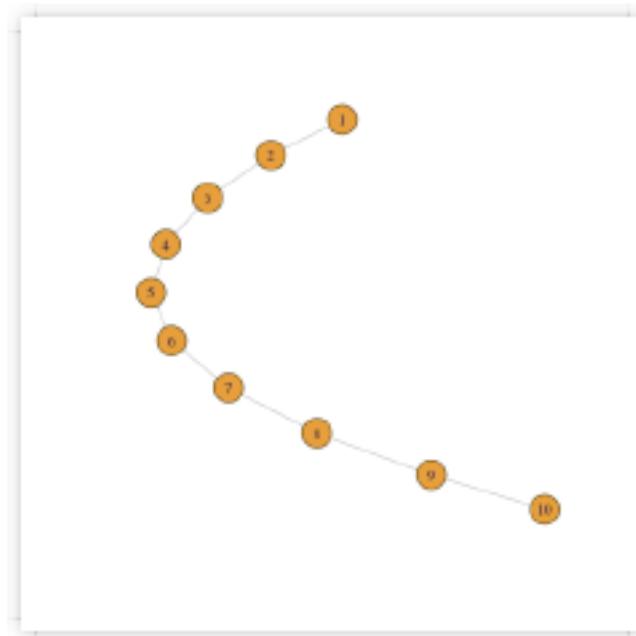
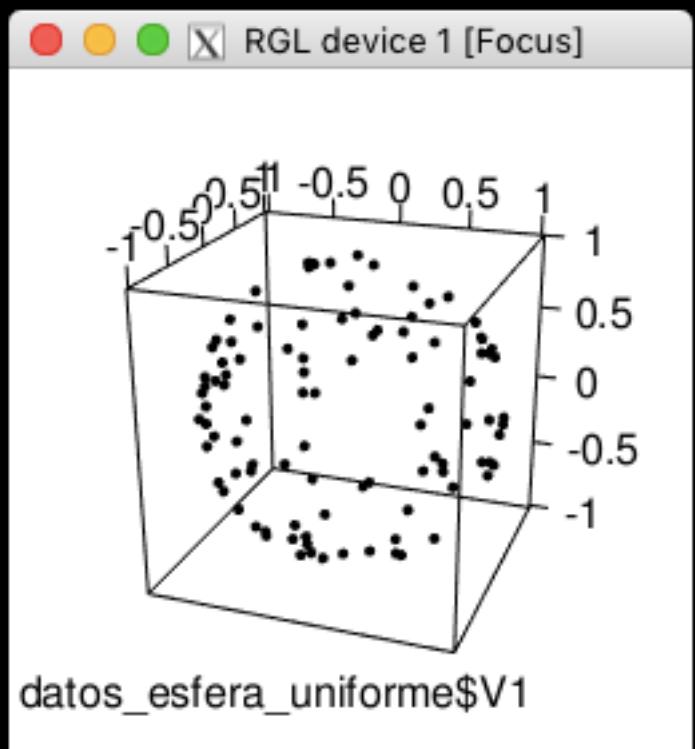


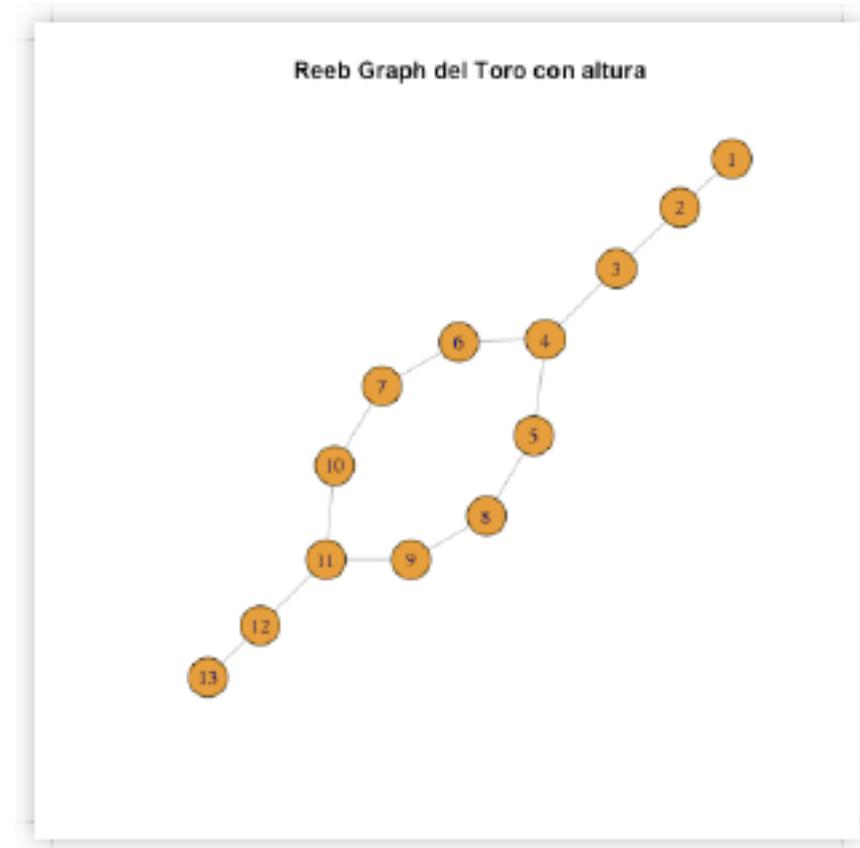
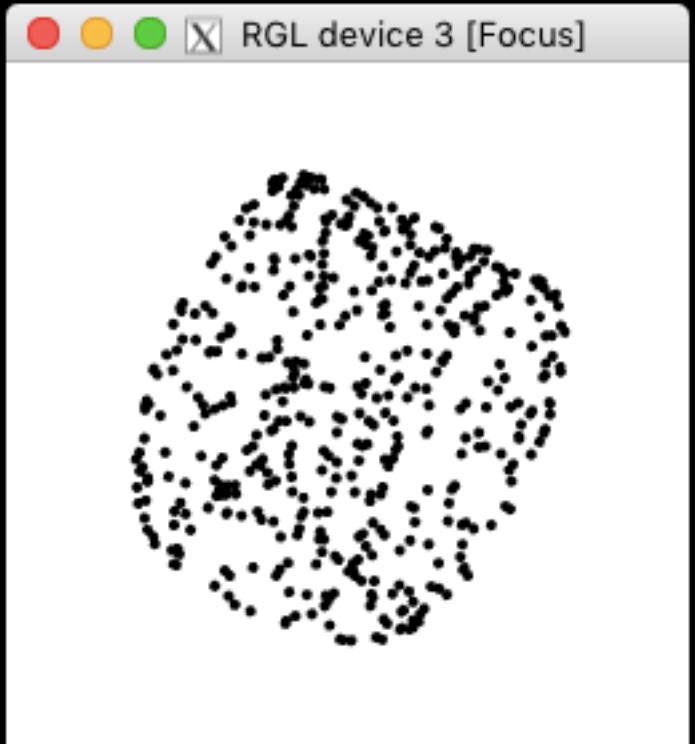
mapper1D Curva momento (trig)

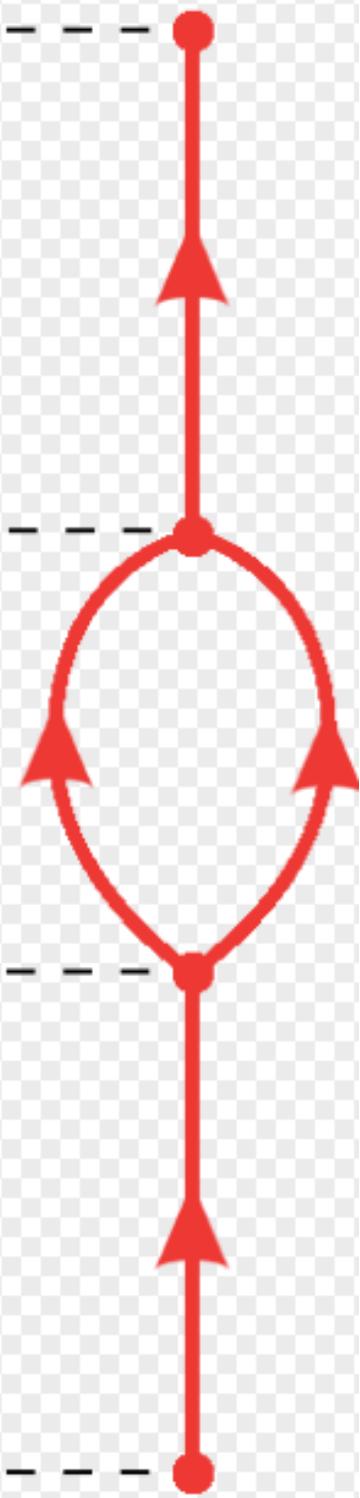
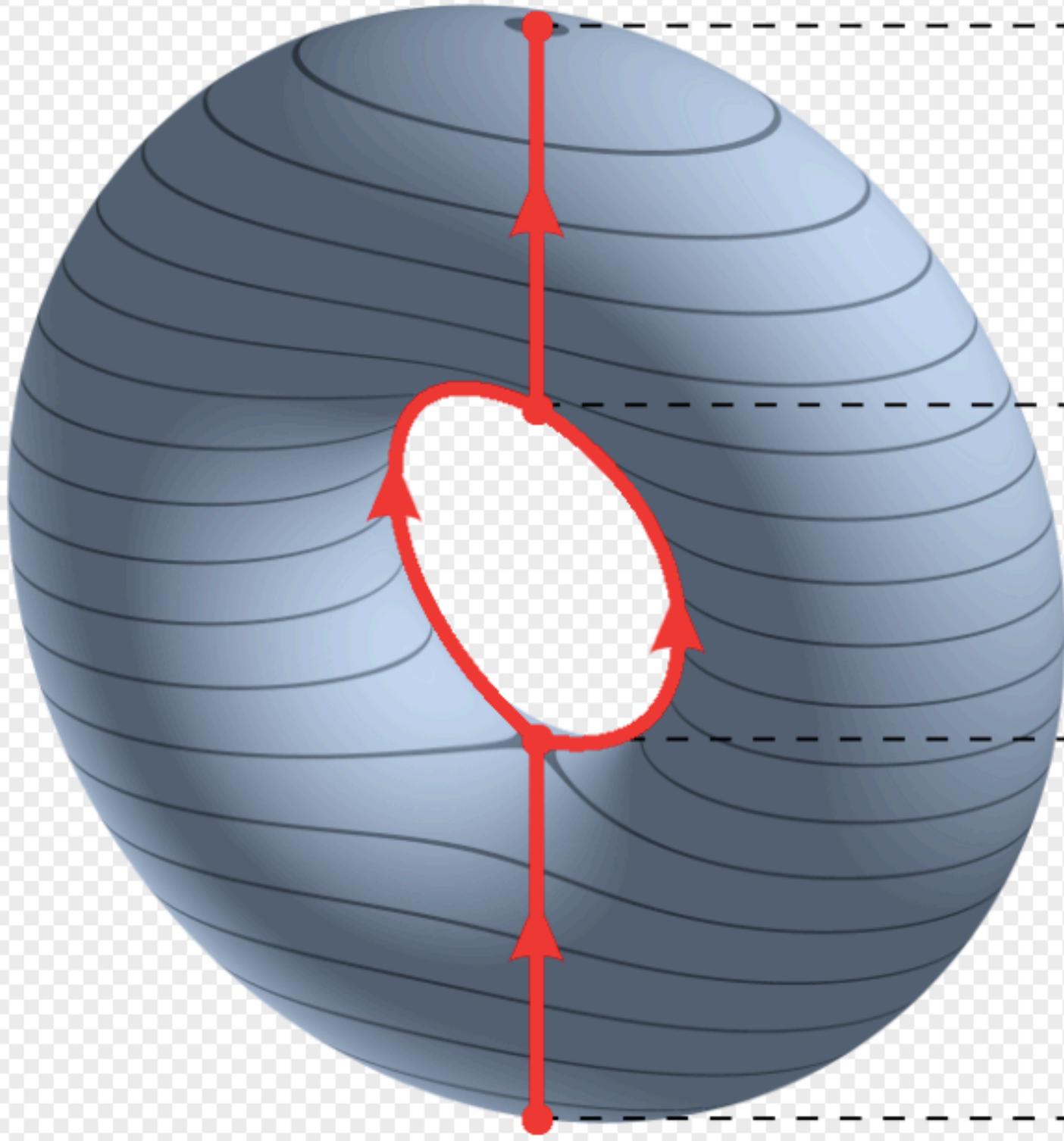
```
#generar datos curva de momento
trigonometrica en R4
datos_curva_momento<-
data.frame(x1=cos((1:numpoints)),y1=sin((1:numpoints)),x2=cos(2*
(1:numpoints)),y2=sin(2*
(1:numpoints)))
```

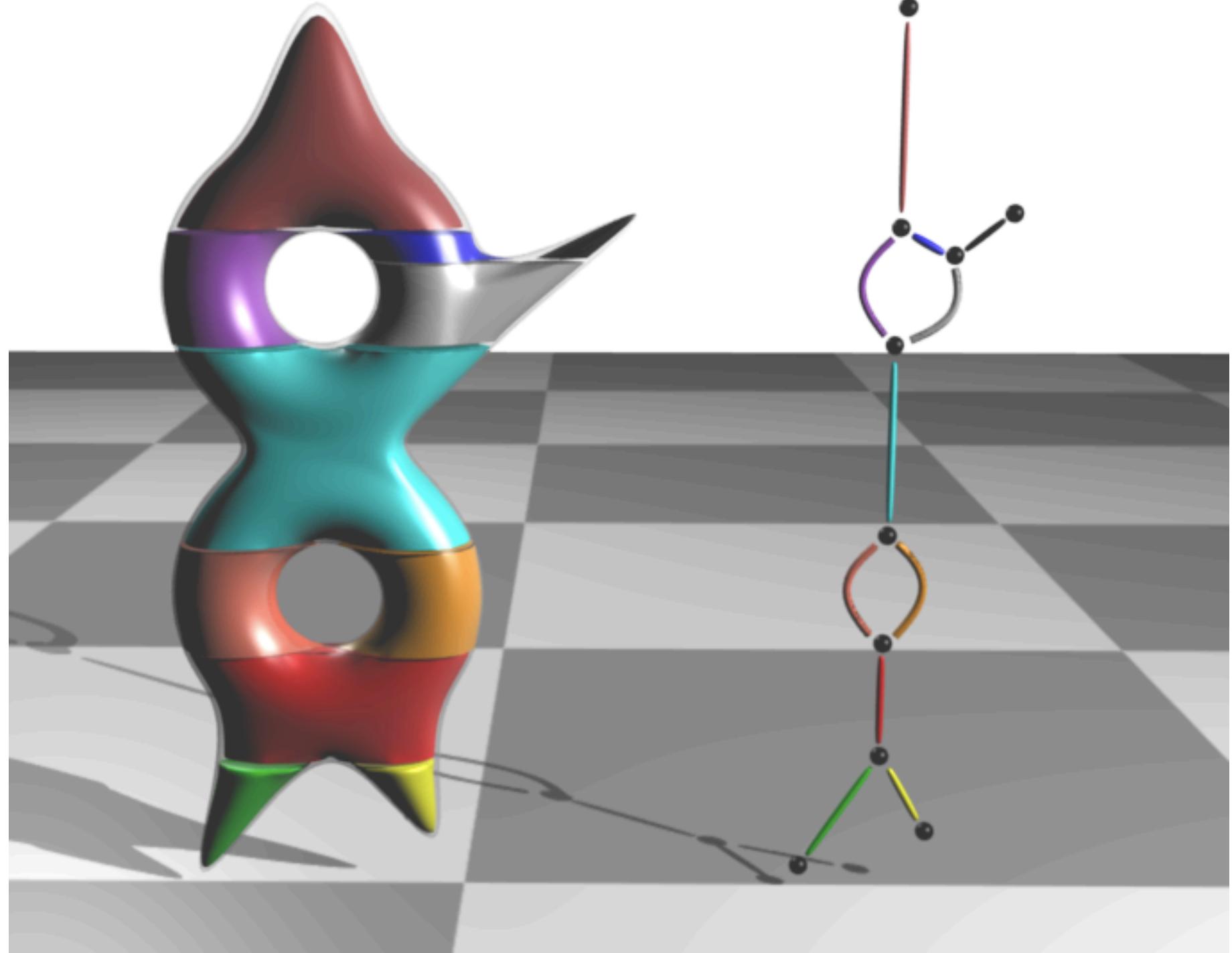


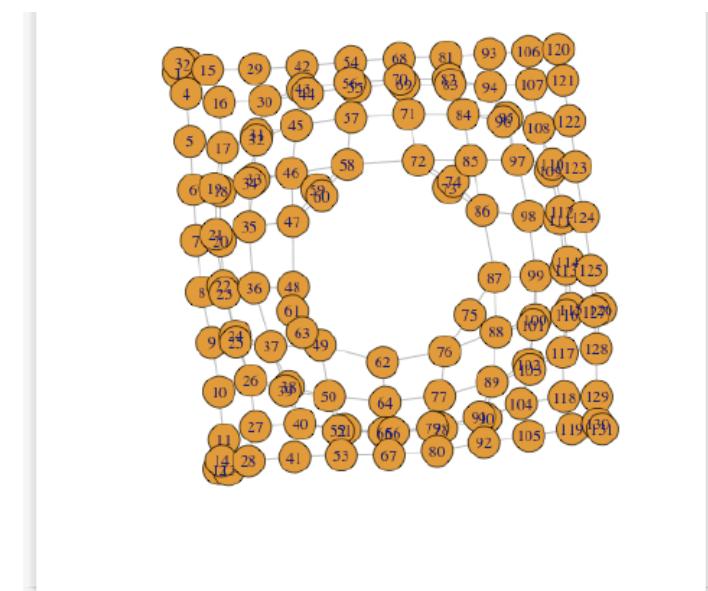
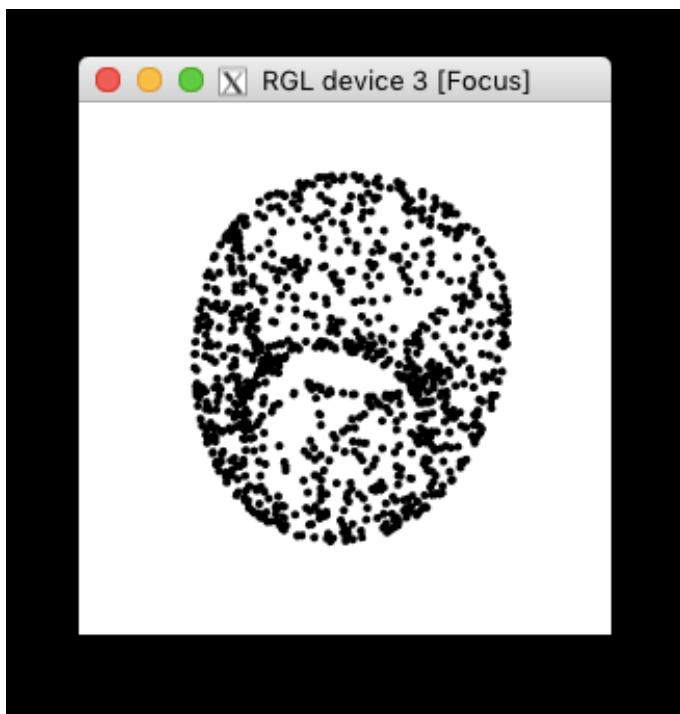
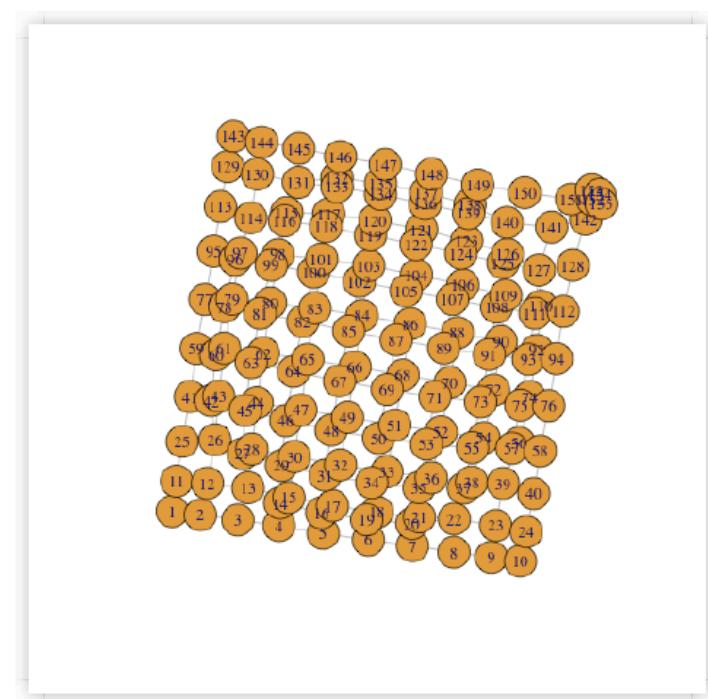
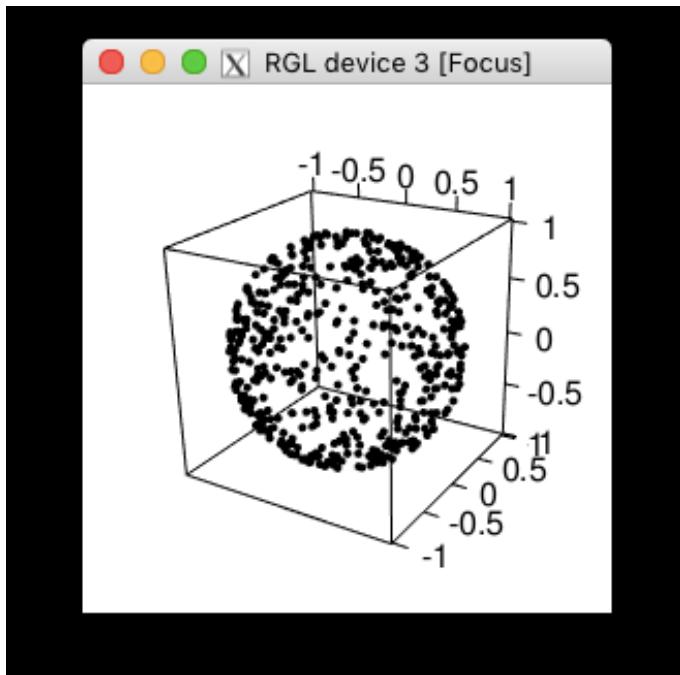












Barclays Premier League 2011-2012 Club Line Up



Arsenal



Aston Villa



Blackburn



Bolton



Chelsea



Everton



Fulham



Liverpool



Man City



Man Utd



Newcastle



Norwich



QPR



Stoke City



Sunderland



Swansea City



Tottenham



West Brom

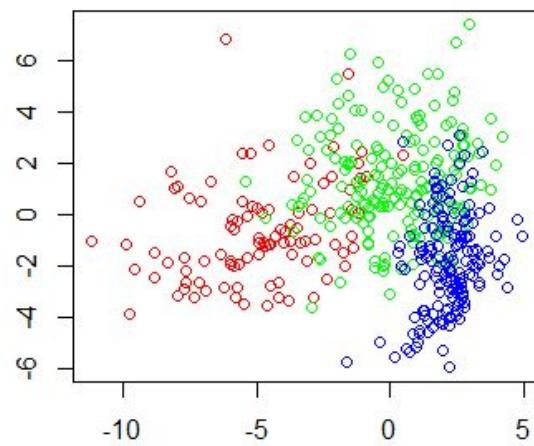


Wigan Athlethic



Wolves

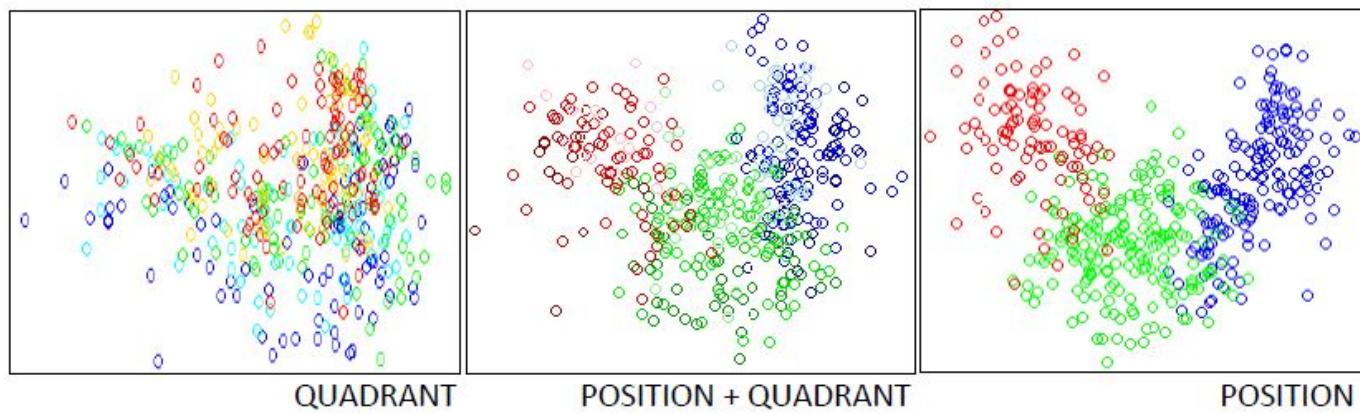
PCA datos jugadores Liga Premiere 2011-2012
(DATOS OPTA 210 características por jugador por partido)



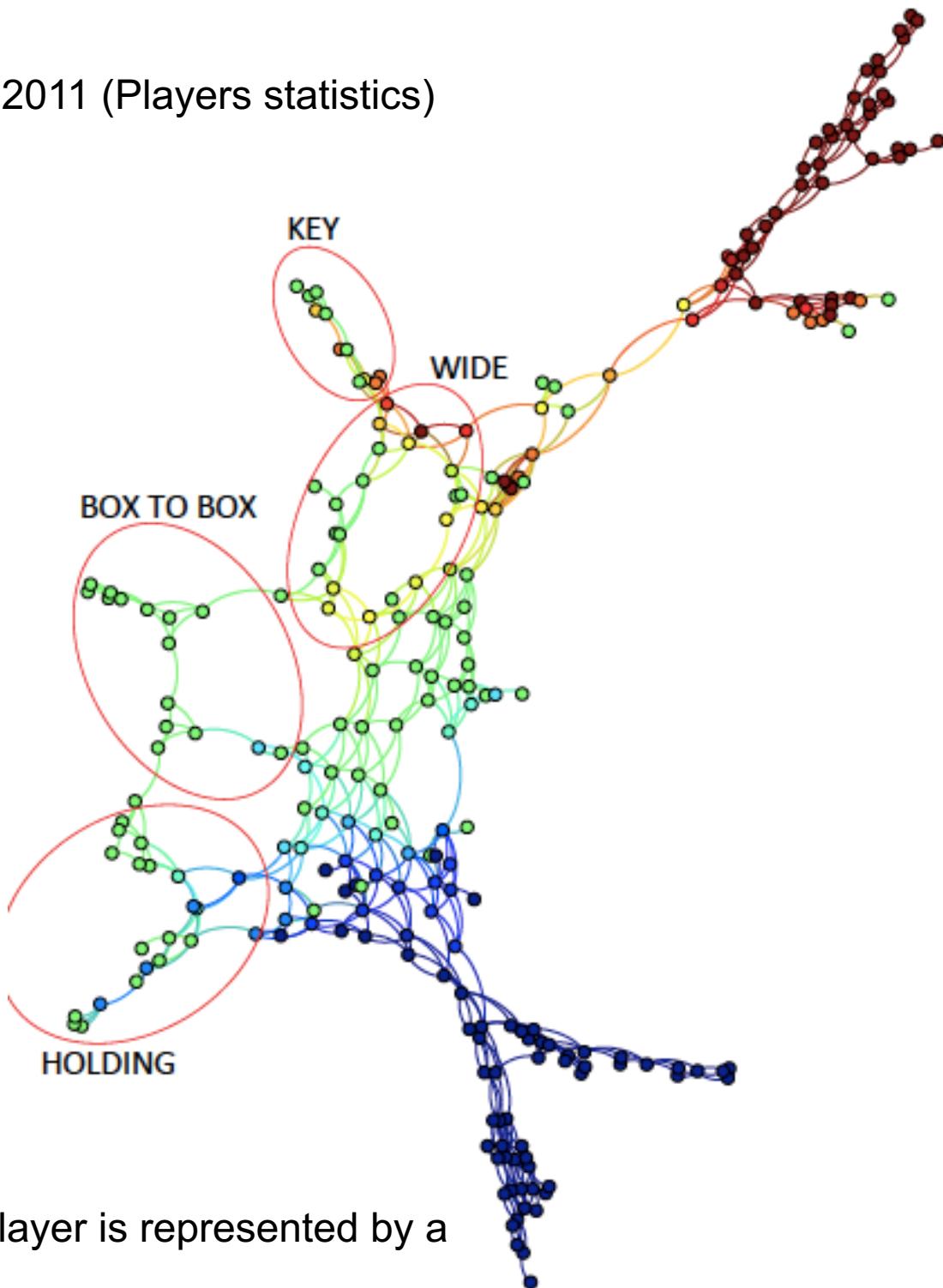
Trabajo en conjunto con David Perdomo

Rojo defensores
Verde Mediocampistas
Morado Atacantes

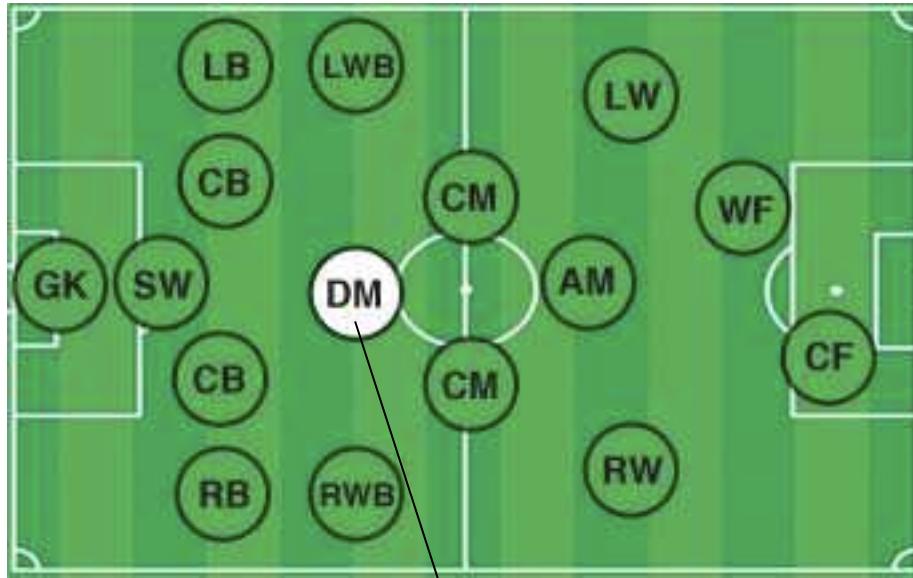
Métricas aprendidas por optimización semidefinida



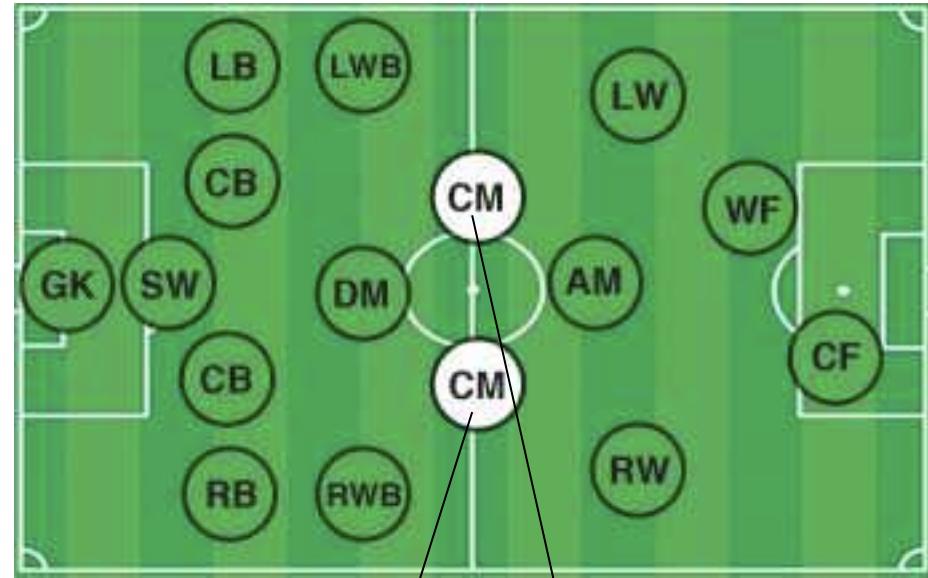
Premier League 2011 (Players statistics)



450 players, each player is represented by a vector in \mathbb{R}^{210}

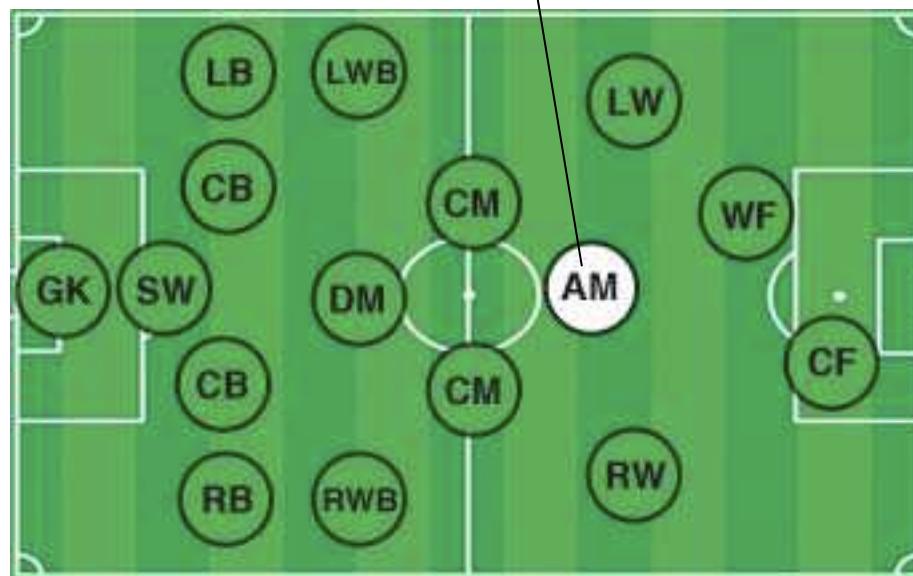


Holding Midfielder



Box to Box Midfielder

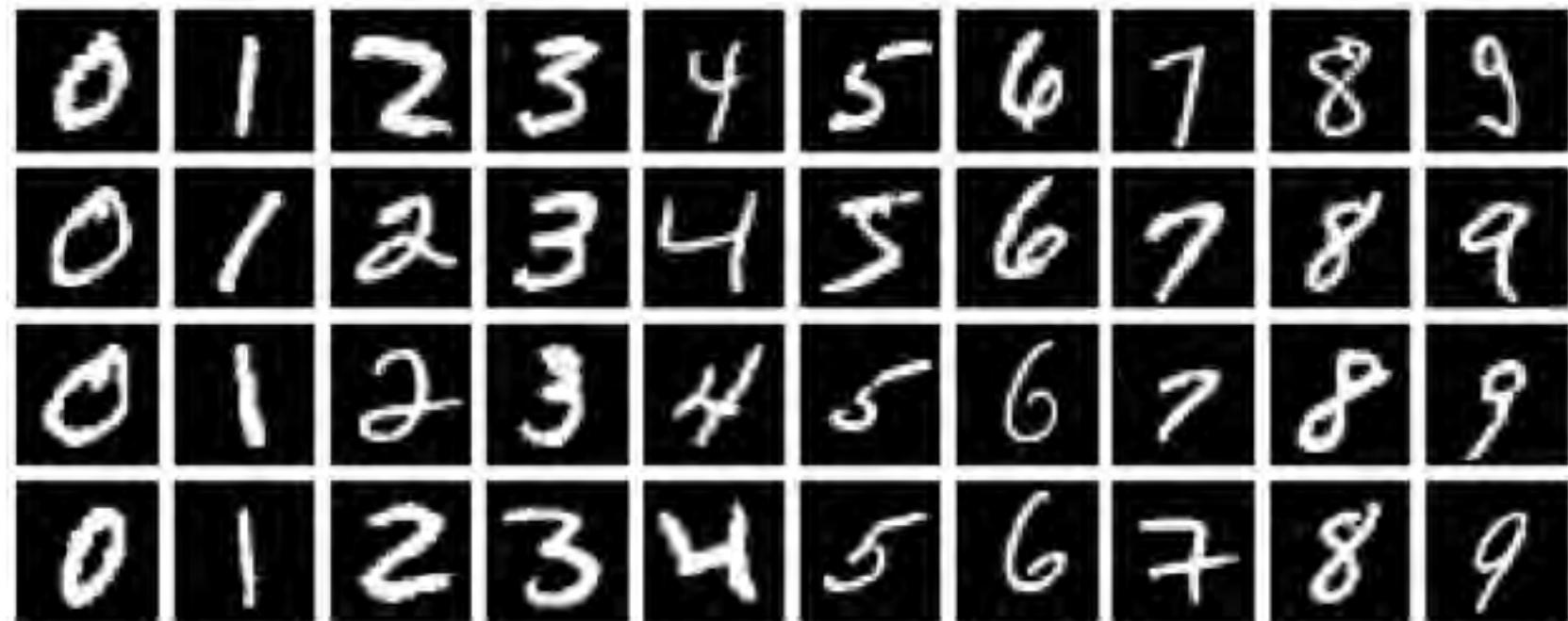
Key Midfielder



Wide Midfielder



MNIST (Hand written digits)

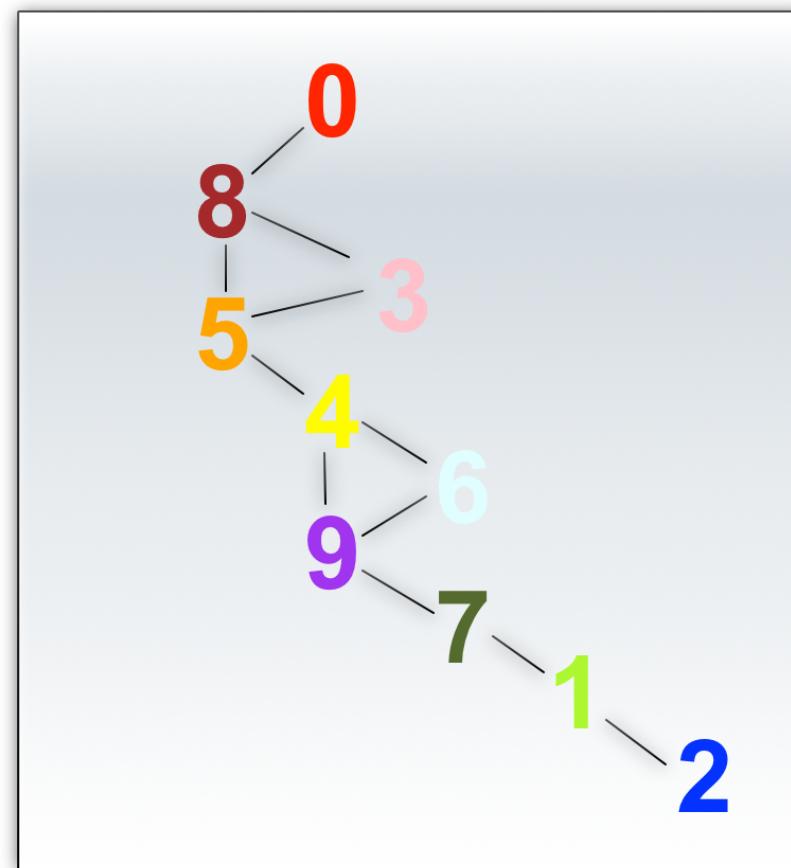
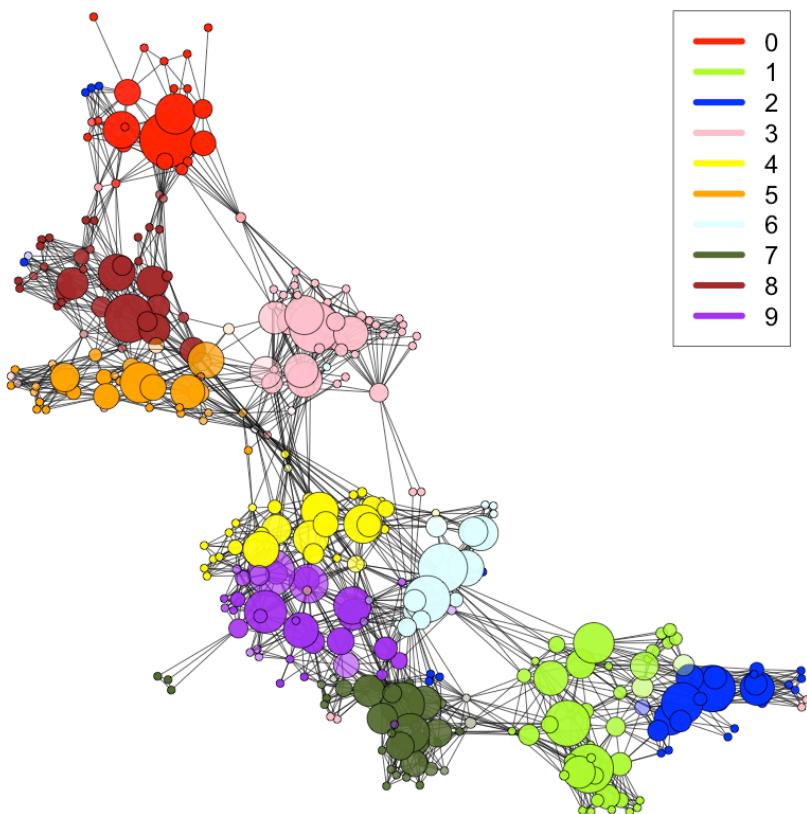


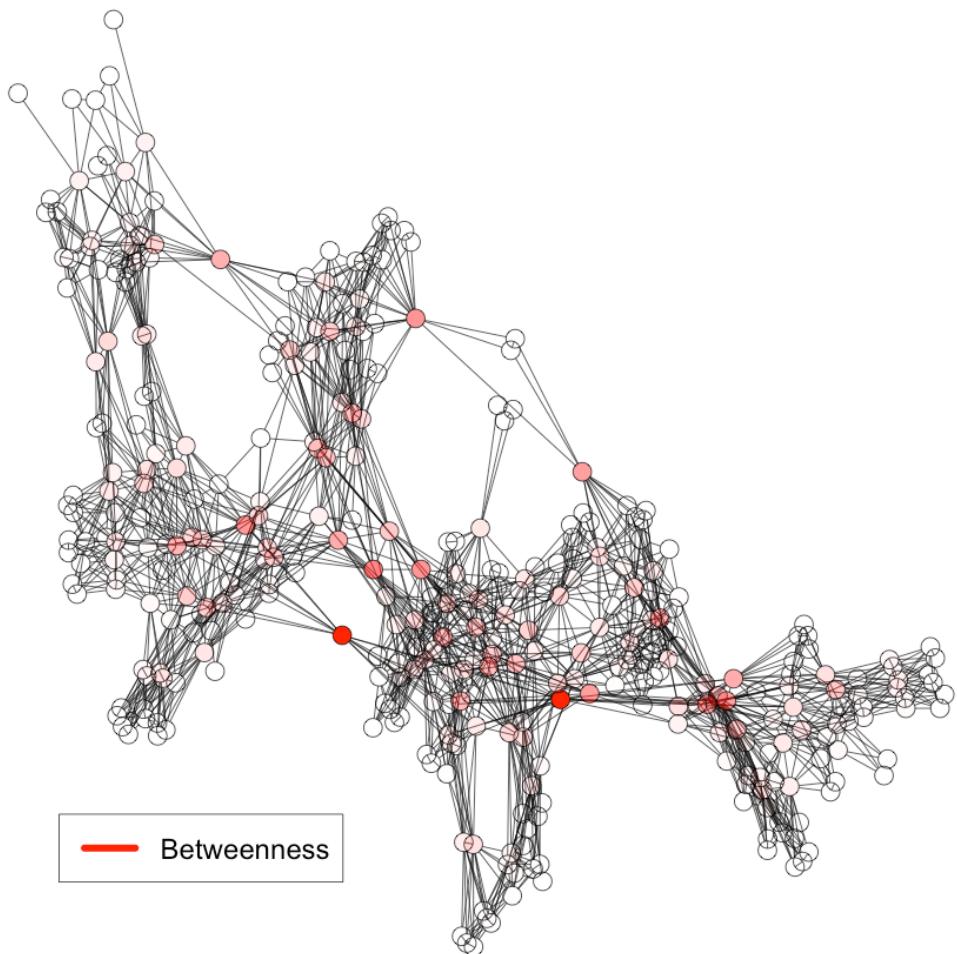
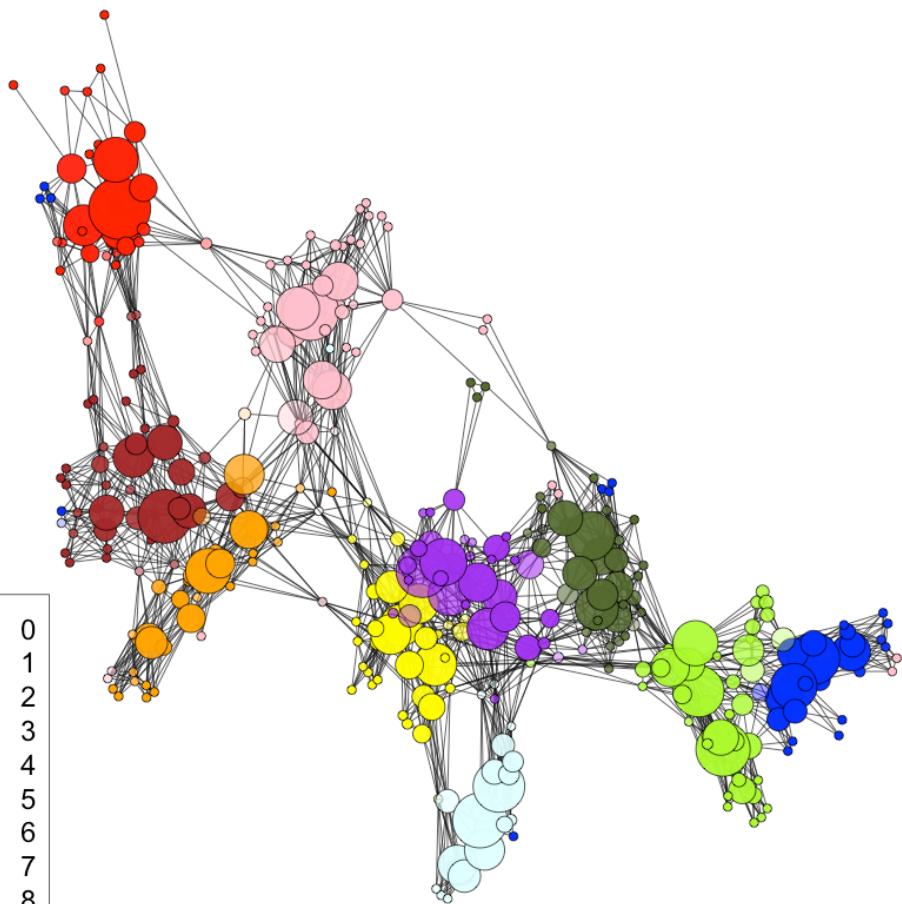
Yann LeCunn, Corinna Cortes and Christopher J.C Burges

60,000 training digits and 10,000 test digits.

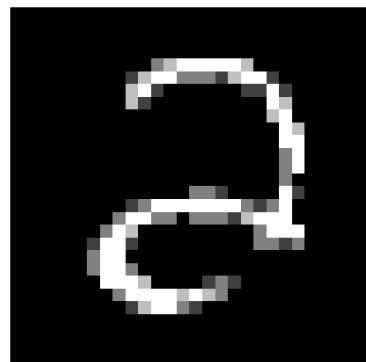
Each is a 28x28 image with gray intensity between 0 and 255 and can be represented as a vector in \mathbb{R}^{784} .

MNIST (Hand written digits)

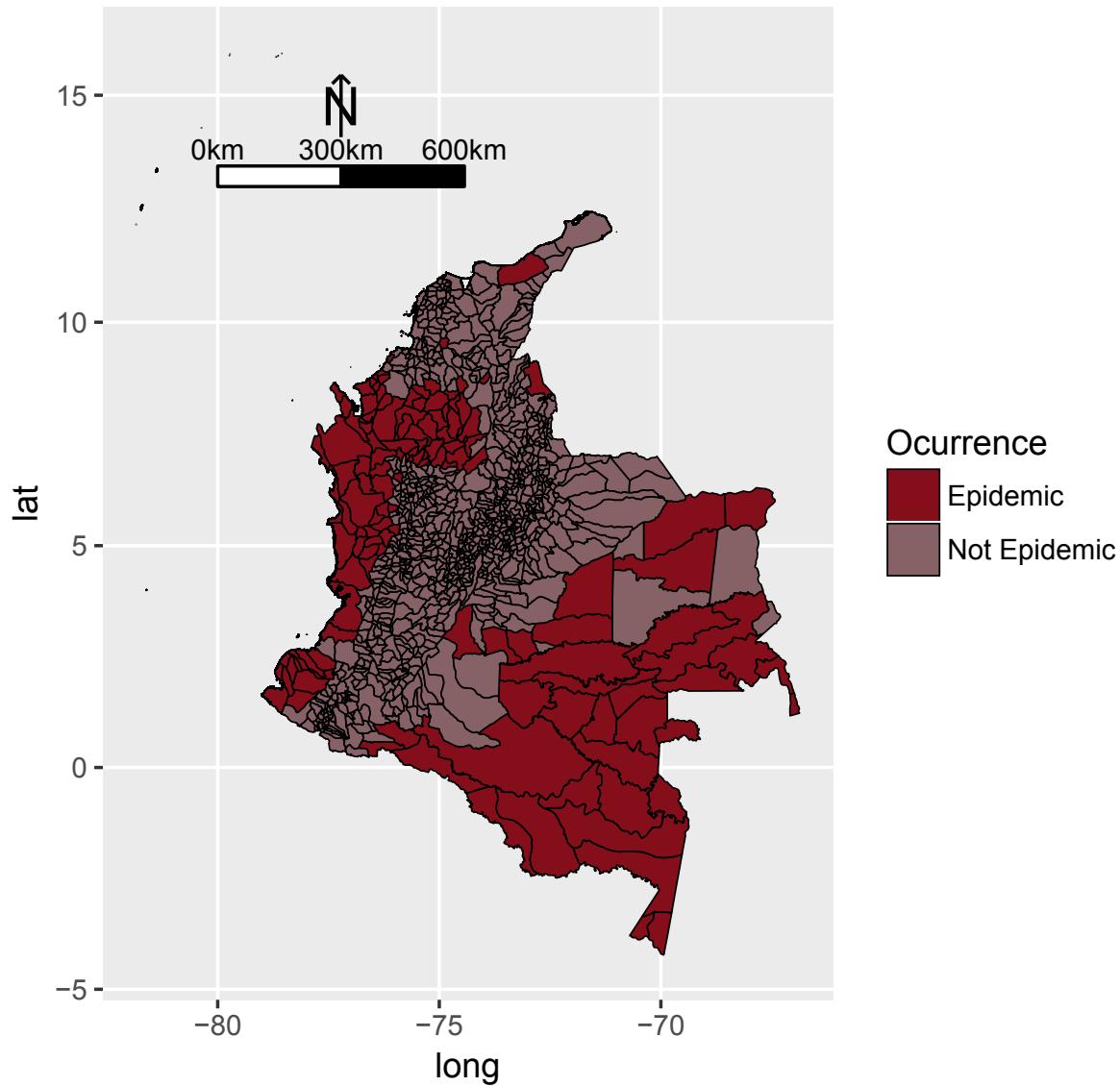




— Betweenness



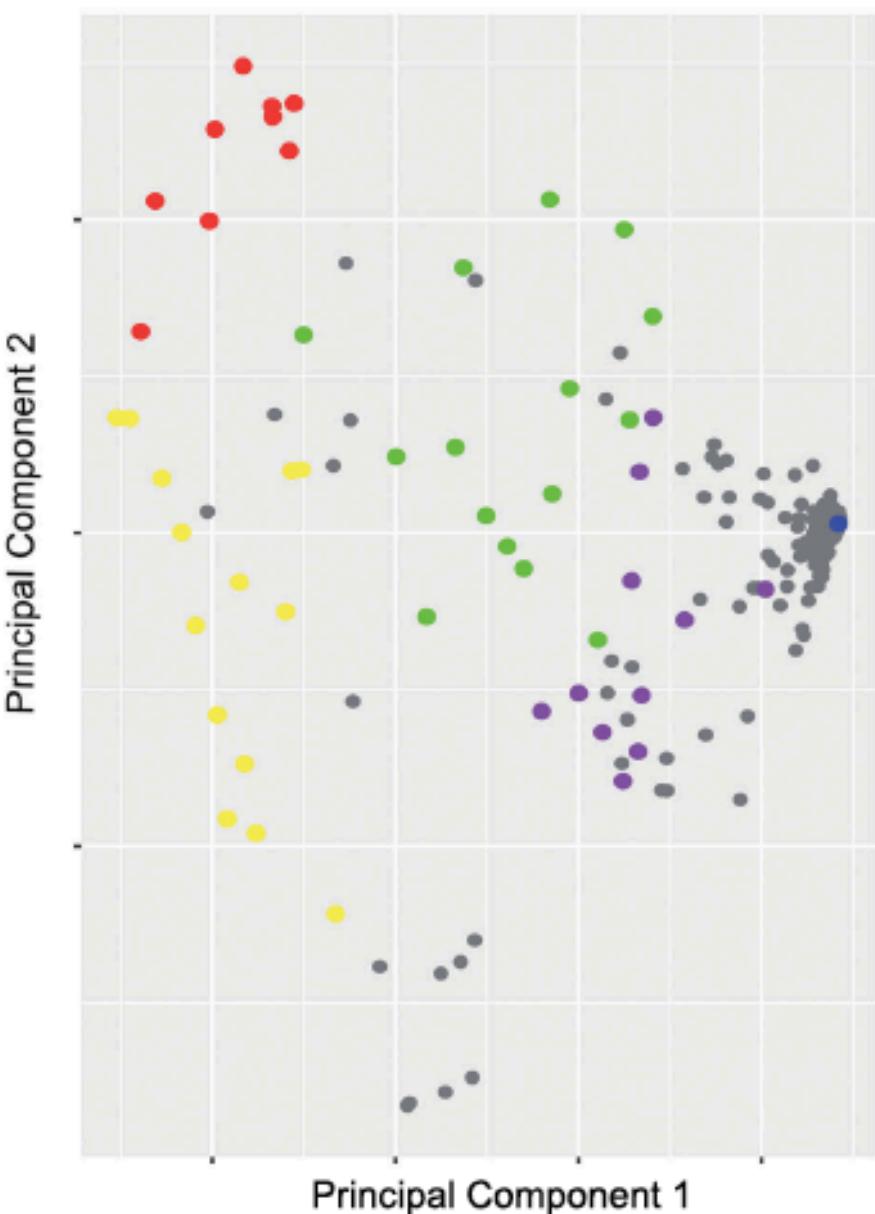
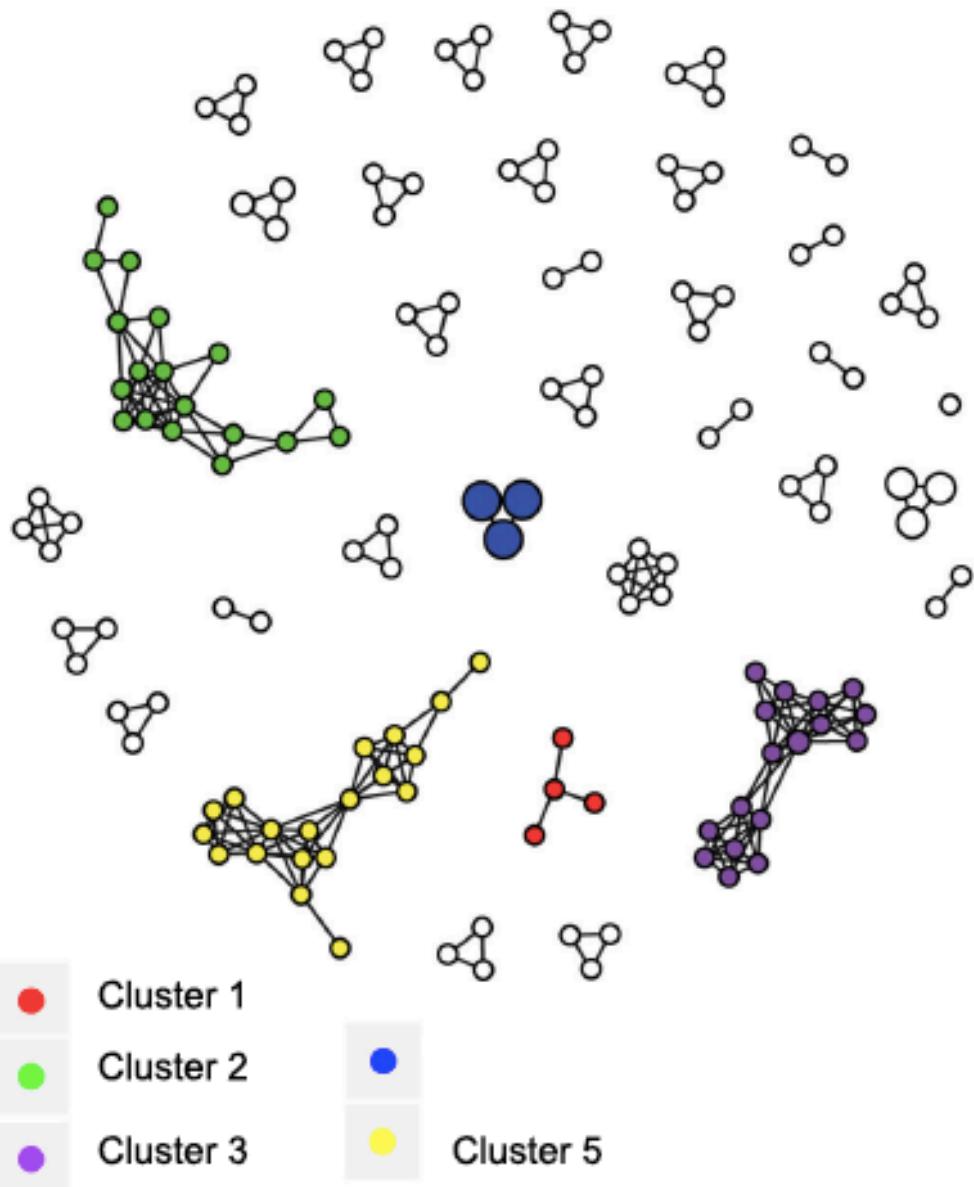
Epidemic Clusters for Malaria (All Parasites)



1156 municipalities, 450 weeks 2007-2015.
Each municipality can be represented as a vector in \mathbb{R}^{450}



Malaria epidemics 2007-2015



Selected Clusters

