A Chicken in Every Pot, and a Social Scientist in Every Classroom: Political Tools for Creation and Collaboration

Ruben R. Puentedura, Ph.D.

Please rank the following ice-cream flavors, from most to least preferred:

- Chocolate
- Vanilla
- Strawberry
- Lemon

Condorcet Voting

Building Fair Recommendation Systems

- Suppose that seven people are asked to vote on three textbooks, so as to recommend the “best” one. Here’s how they voted:

Voter 1	A
Voter 2	A
Voter 3	B
Voter 4	B
Voter 5	C
Voter 6	C
Voter 7	C

- So, textbook C should be the chosen one, right? Not so fast…

Building Fair Recommendation Systems (cont.)

- Had you asked voters about their full preference ranking, this would have been the result:

Voter 1	A	B	C
Voter 2	A	B	C
Voter 3	B	A	C
Voter 4	B	A	C
Voter 5	C	A	B
Voter 6	C	A	B
Voter 7	C	A	B

- In other words, B acted as a “spoiler” for A, allowing the (overall inferior) C to be ranked first!

What Can We Do About This?

- This scenario (and others like it) can never be **completely** avoided (*Arrow's Impossibility Theorem*);

- However, simple plurality systems are **worse** than many other voting schemes in terms of the problems they cause.

- A scheme that is significantly **better** than plurality systems: *Condorcet Voting*.

Condorcet Voting

- Voters rank the options in order of preference;
- For each pair of options, the number of times one is preferred to the other is counted;
- If an option is preferred over all others, it is the winner.
- Ties and ambiguities can be handled separately by a range of methods, should they arise.

Condorcet Voting - An Example

- Using our results from before:

Voter 1	A	B	C
Voter 2	A	B	C
Voter 3	B	A	C
Voter 4	B	A	C
Voter 5	C	A	B
Voter 6	C	A	B
Voter 7	C	A	B

- In pairwise contests:
  A is preferred to B 5 times; B is preferred to A 2 times
  A is preferred to C 4 times; C is preferred to A 3 times
  B is preferred to C 4 times; C is preferred to B 3 times

- So A is the winner.

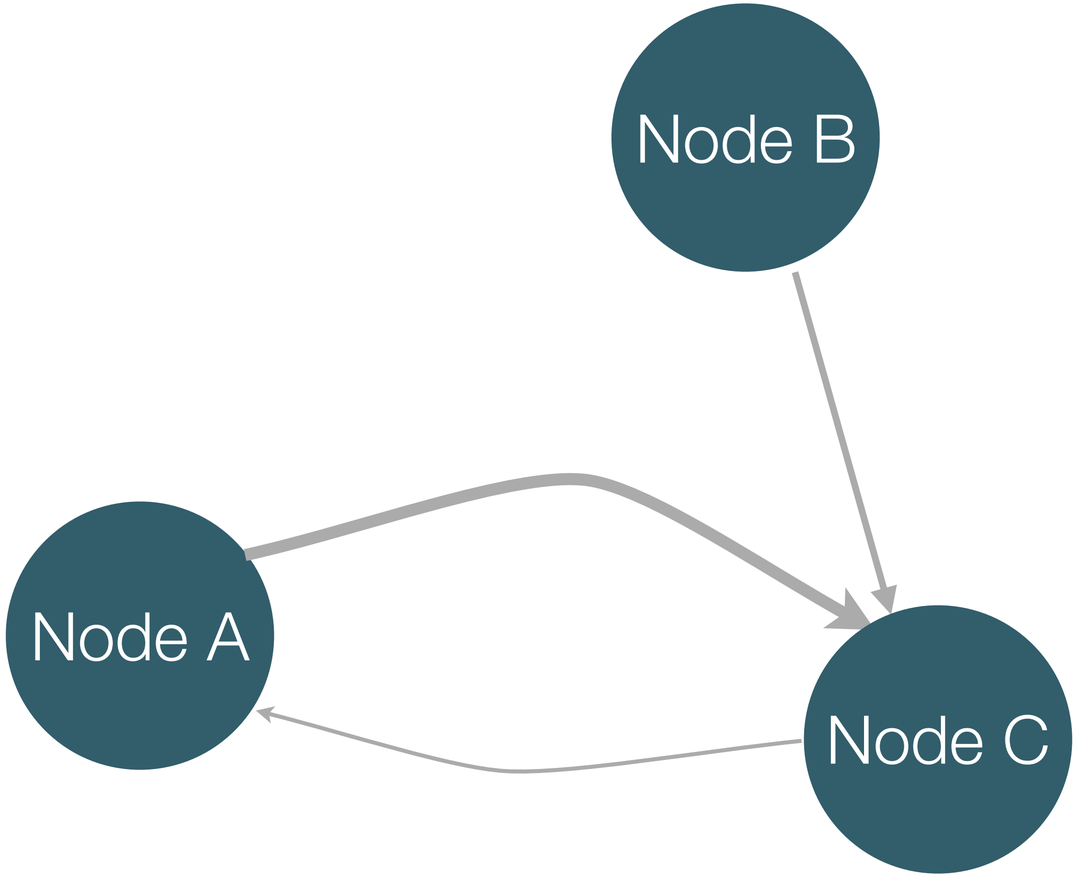
Arrow's Five Fairness Criteria

- Always-A-Winner
  - Every sequence of preference lists produces at least one winner
- Condorcet Winner Criterion
  - If there is a Condorcet winner, then it is the social choice
- Pareto Condition
  - If everyone prefers x to y, then y is not a social choice
- Monotonicity
  - If someone changes their vote in such a way that it favors the social choice, the social choice should not change
- Independence of Irrelevant Alternatives
  - If someone changes their vote in such a way that it changes the ranking of an alternative not in the social choice set, but does not place it over an option in the social choice set, the new alternative should not now be in the social choice set

Social Network Analysis

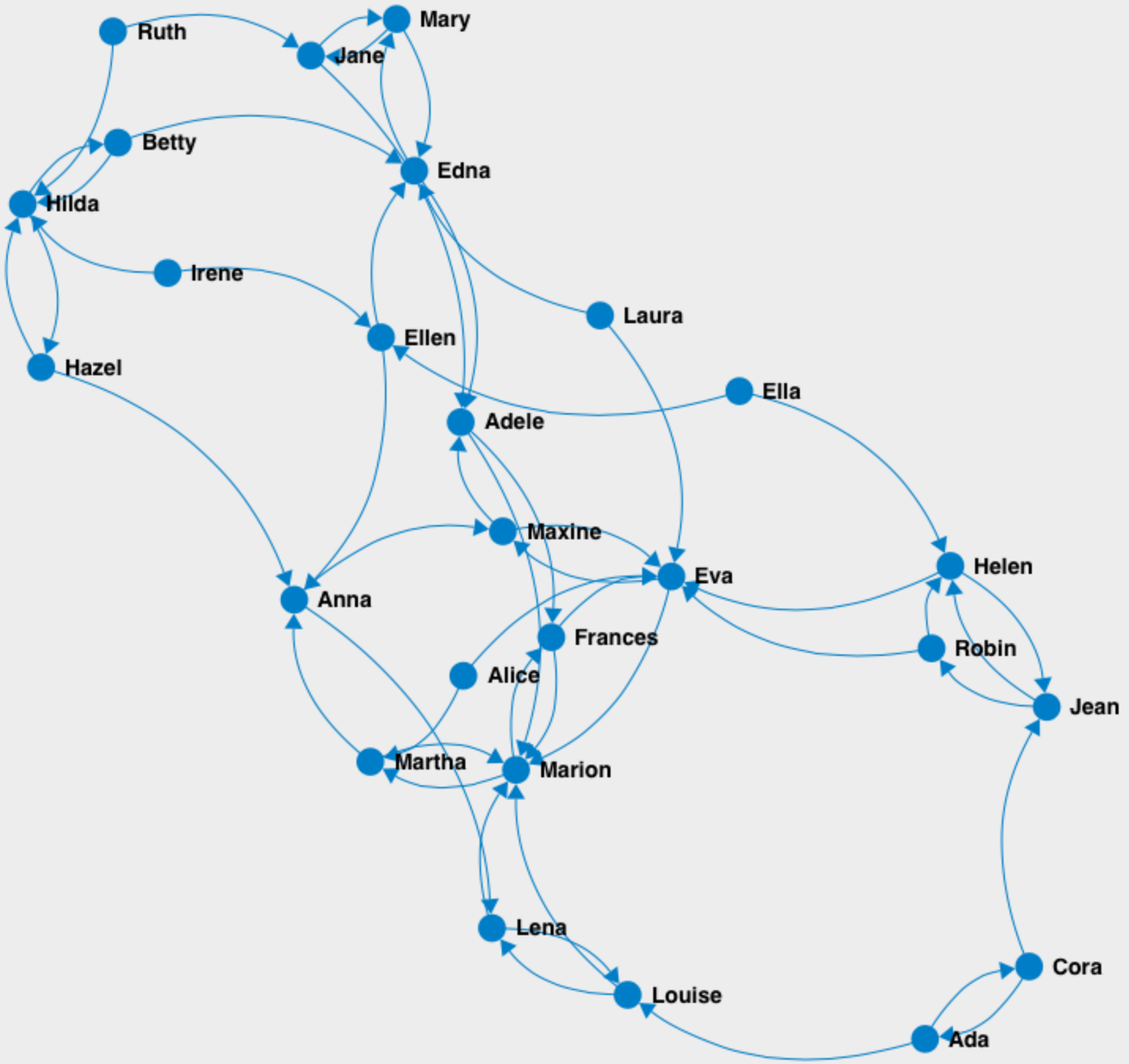
What Is a Network?

- A network is a collection of items (*nodes*) where connections among the items are indicated by links (*edges*):

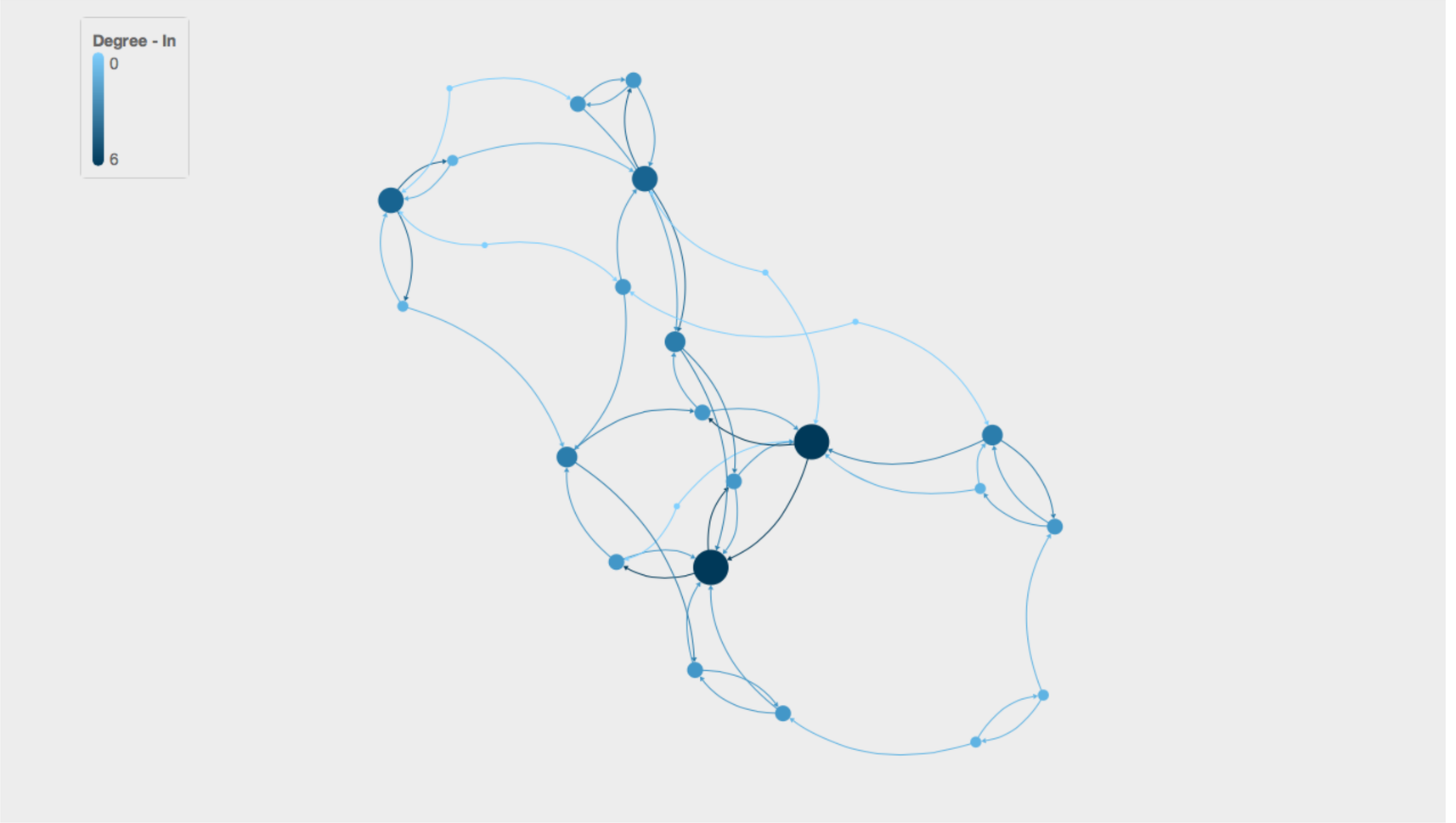


Social Network Analysis – Collecting Data

- **The key components:**
  - The **actors**
  - Their **connections**
  - The **directionality** (if any) of the connections
  - The **strength** (if any) of the connections
- **Looking at the actors:**
  - **Whole Network:** Do we choose a full set for which we then determine connections?
  - **Egocentric Network:** Do we choose a subset (egos) for which we then determine connections to other actors (alters)?
- **Choosing the actors:**
  - **Positional Approach:** based upon set membership or shared attributes
  - **Event-Based Approach:** based upon participation in a set of activities
  - **Relational Approach:** based upon social connectedness
- **Collecting the data:**
  - All the data sources discussed before, plus:
  - Electronic network data from e.g. email, discussion forums

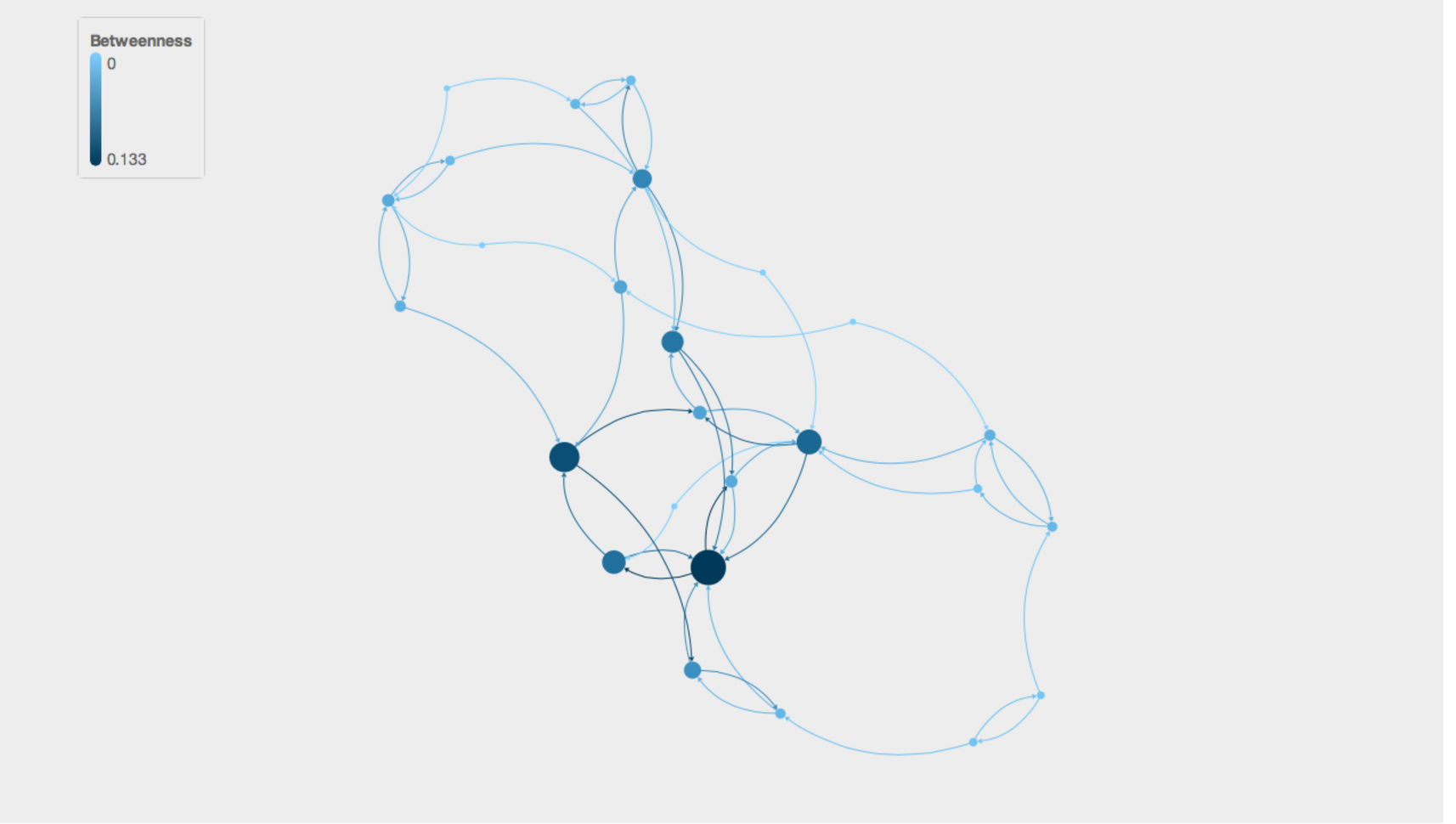

Mary
Ruth
Jane
Betty
Edna
Hilda
Irene
Laura
Ellen
Hazel
Ella
Adele
Maxine
Eva
Helen
Anna
Frances
Robin
Alice
Jean
Martha
Marion
Lena
Cora
Louise
Ada

Degree

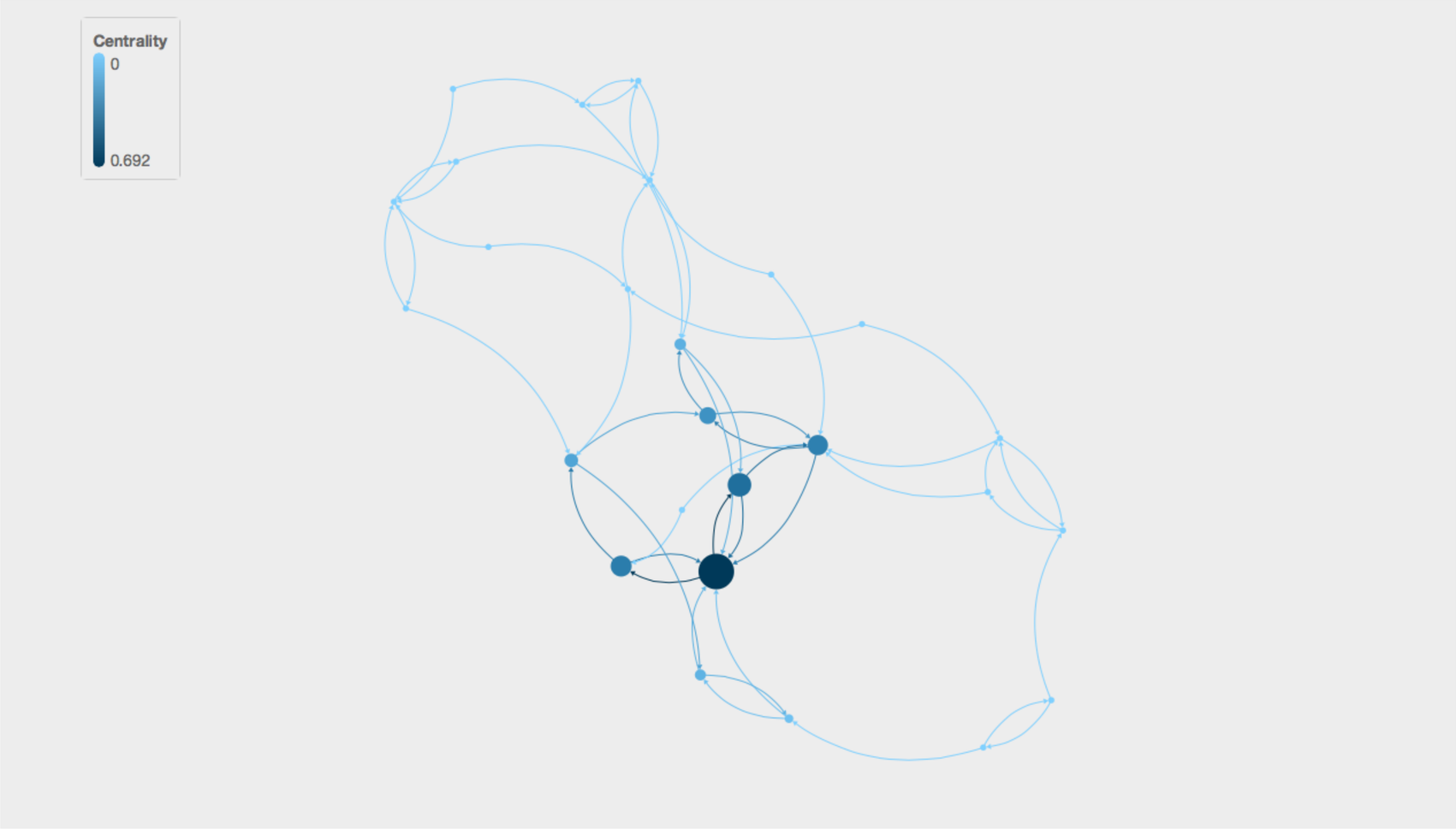


Data Source: Dining-table partners in a dormitory at a New York State Training School – Coded as a Pajek network data file by W. de Nooy (2001)
Originally published in J.L. Moreno, *The Sociometry Reader*. The Free Press (1960) – Research by H.H. Jennings and J.L. Moreno

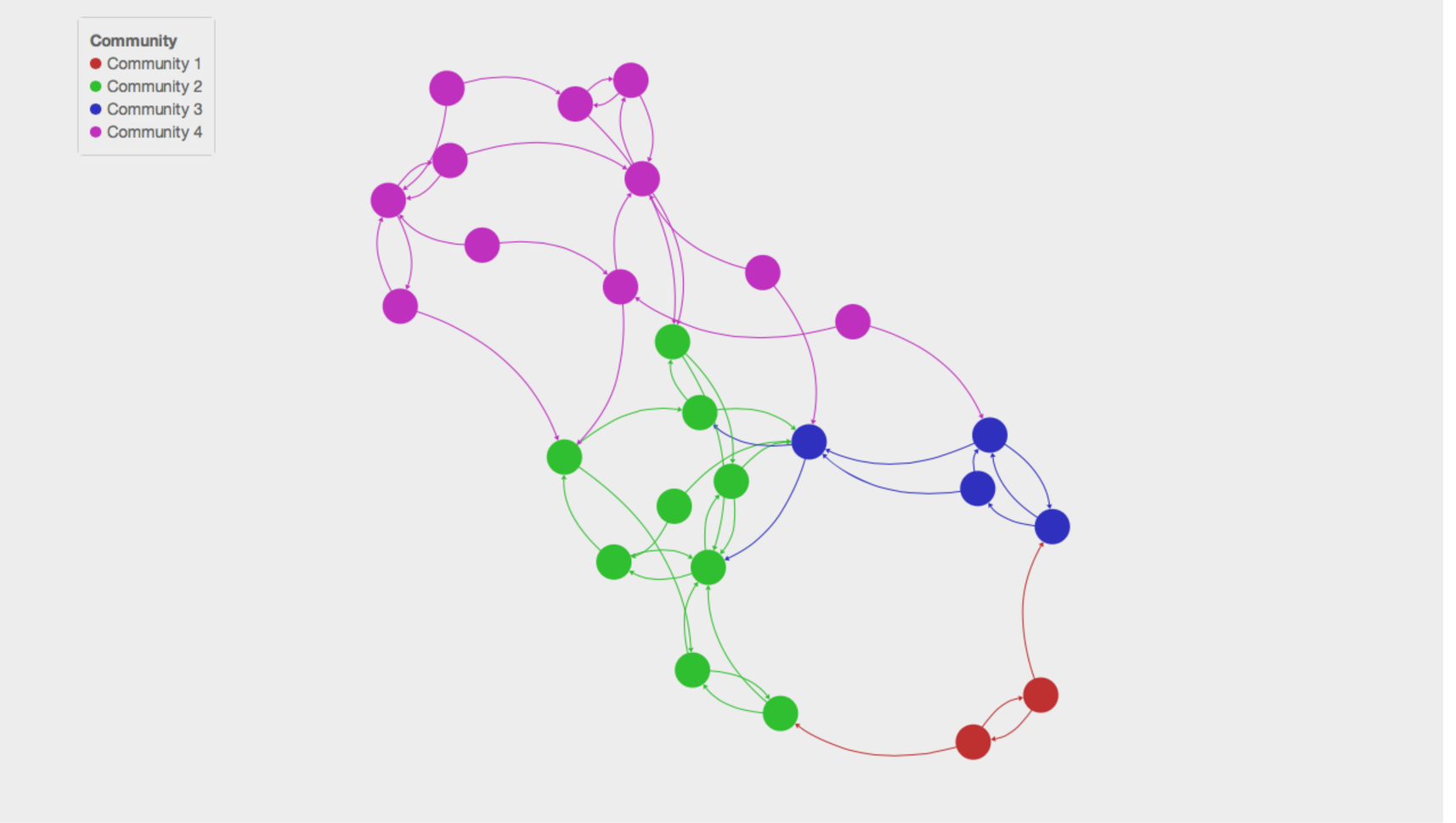
Betweenness



Eigenvector Centrality



Community

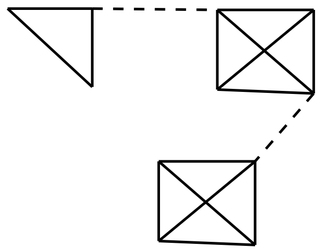


3 Networks for Today

- **Dining-table partners in a dormitory at a New York State Training School**
  - Research by H.H. Jennings and J.L. Moreno, published in J.L. Moreno, *The Sociometry Reader*. Glencoe (Ill.), The Free Press, 1960, p. 35. Network file coded by W. de Nooy.
- **Student government discussion network**
  - V. Hlebec, “Recall versus recognition: comparison of two alternative procedures for collecting social network data.” (in A. Ferligoj & A. Kramberger (Eds.), *Developments in Statistics and Methodology*. Ljubljana: FDV, 1993). Network file coded by V. Batagelj.
- **Modern math method diffusion**
  - R.O. Carlson, *Adoption of Educational Innovations* (Eugene: University of Oregon, Center for the Advanced Study of Educational Administration, 1965, p. 19). Network file coded by W. de Nooy.

Granovetter and “The Strength of Weak Ties”

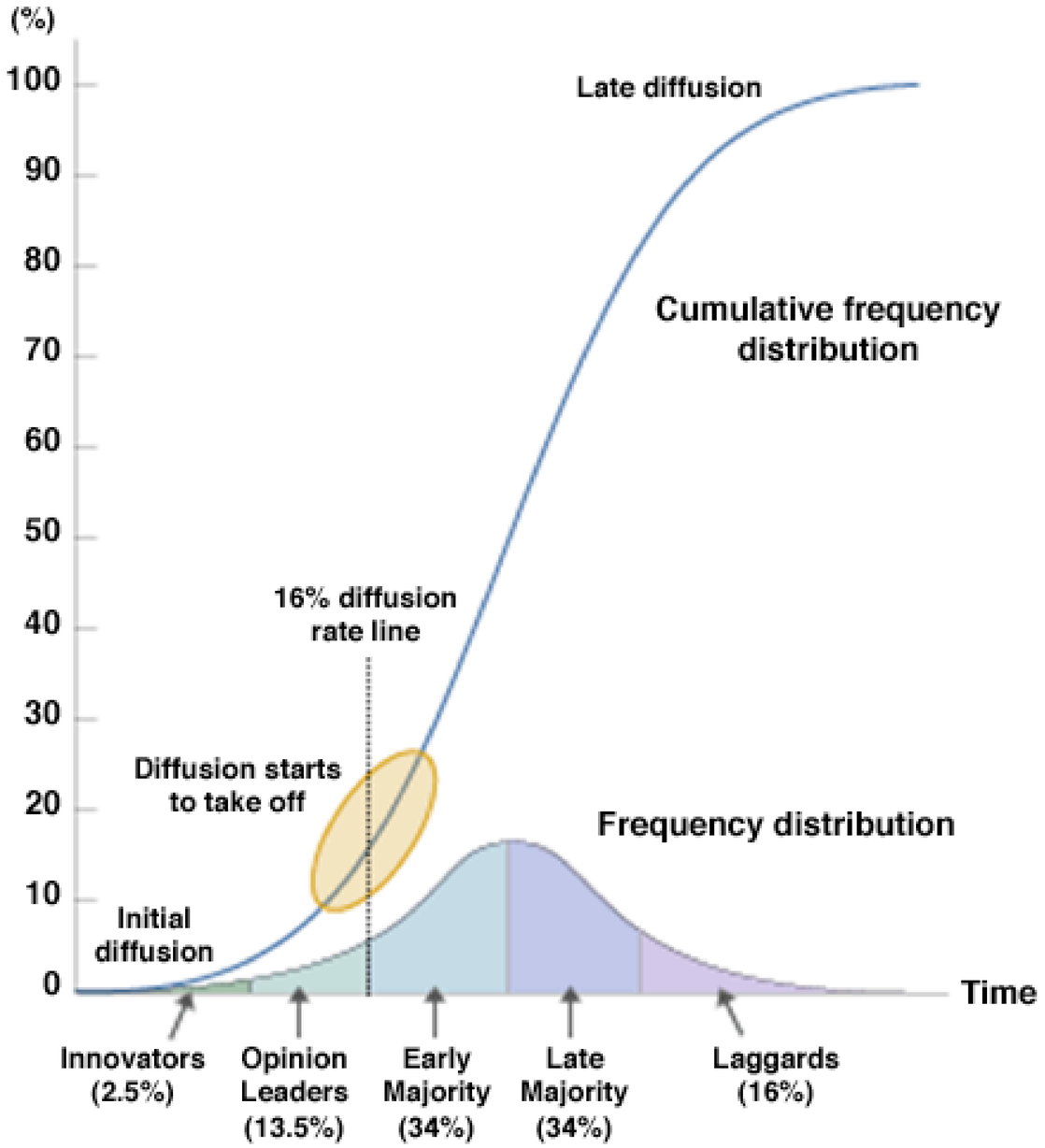
- In a 1973 paper (inspired by his 1970 thesis research on job mobility), Granovetter described the phenomenon of the “weak tie”:
  - Weak ties act as bridges between network clusters that are otherwise unconnected - the picture looks like this:



- Granovetter found that, while more blue-collar workers find out about new jobs through personal contacts than by any other method, most of these contacts were weak ties.

Granovetter and Resilience

- Weak ties are *essential* to the creation of resilient communities.
- Example: two communities faced with disintegration via an urban renewal process:
  - West End (Gans 1962): all ties in the community strong (primarily family, close friends-based)
    - The West End was unable to even form an organization to fight the process of urban renewal.
  - Charlestown (Keyes 1969): a mix of strong and weak ties in the community (both family and organizations)
    - Charlestown successfully organized against the same urban renewal plan.
- Weak ties provide *flexibility* and *resources*.



Everett M. Rogers. *Diffusion of Innovations, 5th Edition*. New York:Free Press, 2003.

Granovetter and Innovation

- Rogers (1962) indicates that innovators are “marginal”, while early adopters are more integrated within the social system.
- Granovetter suggests that innovators are connected to early adopters via weak ties:
  - Kerckhoff and Back (1968): innovators are almost never cited as “friends”.
  - Korte and Milgram (1970): weak interracial ties are more effective in bridging social distance than strong ties.
- Weak ties provide a *medium* for innovation and change to propagate.

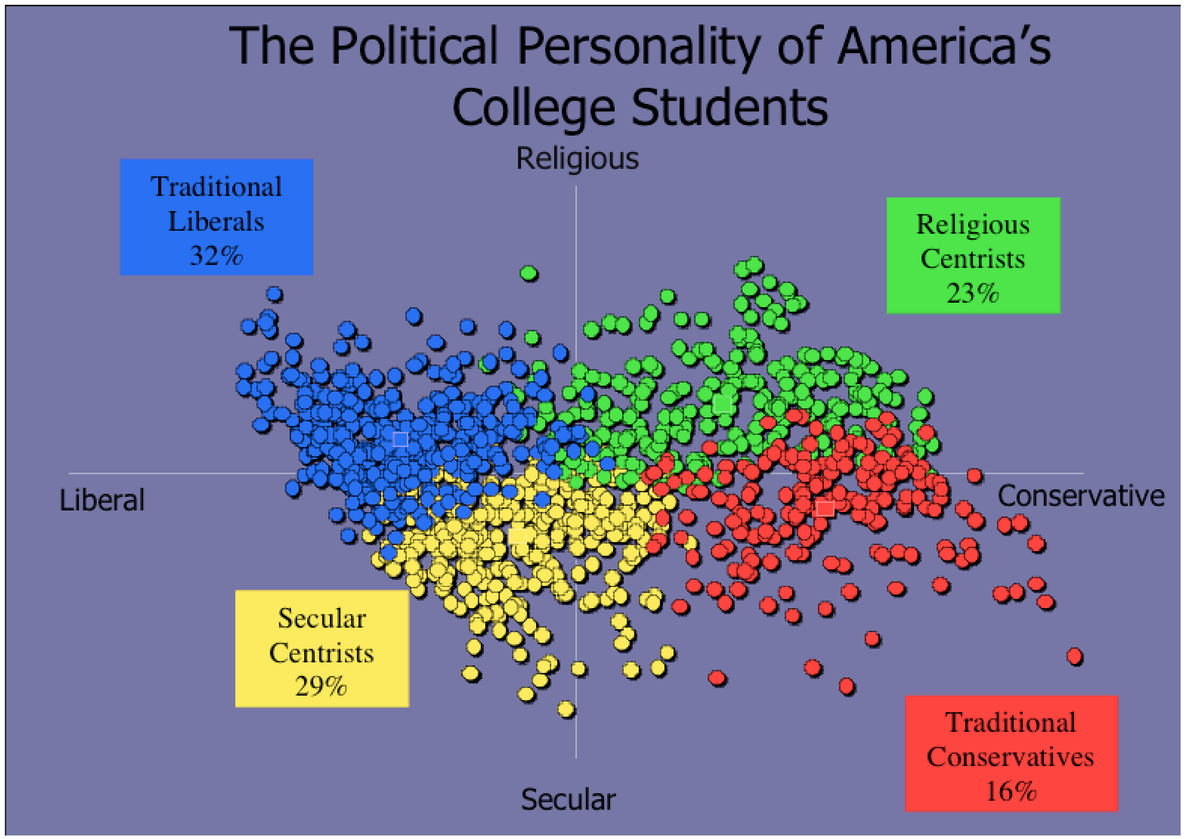
Principal Components Analysis

Factor Analysis and the Network Space

- Factor analysis reorganizes and simplifies data by expressing it in terms of a few underlying axes:
  - For instance, a table containing the flight times between all 50 USA state capitals would contain 49 separate pieces of data for each capital - a rather unwieldy data set;
  - After a suitable factor analysis, the entire data set could be expressed in terms of two axes, which would correspond to latitude and longitude.

Example: the April 2004 Harvard Institute of Politics/SDS Prime Survey of College Undergraduates

- Eleven questions on political topics can be summarized by two axes:

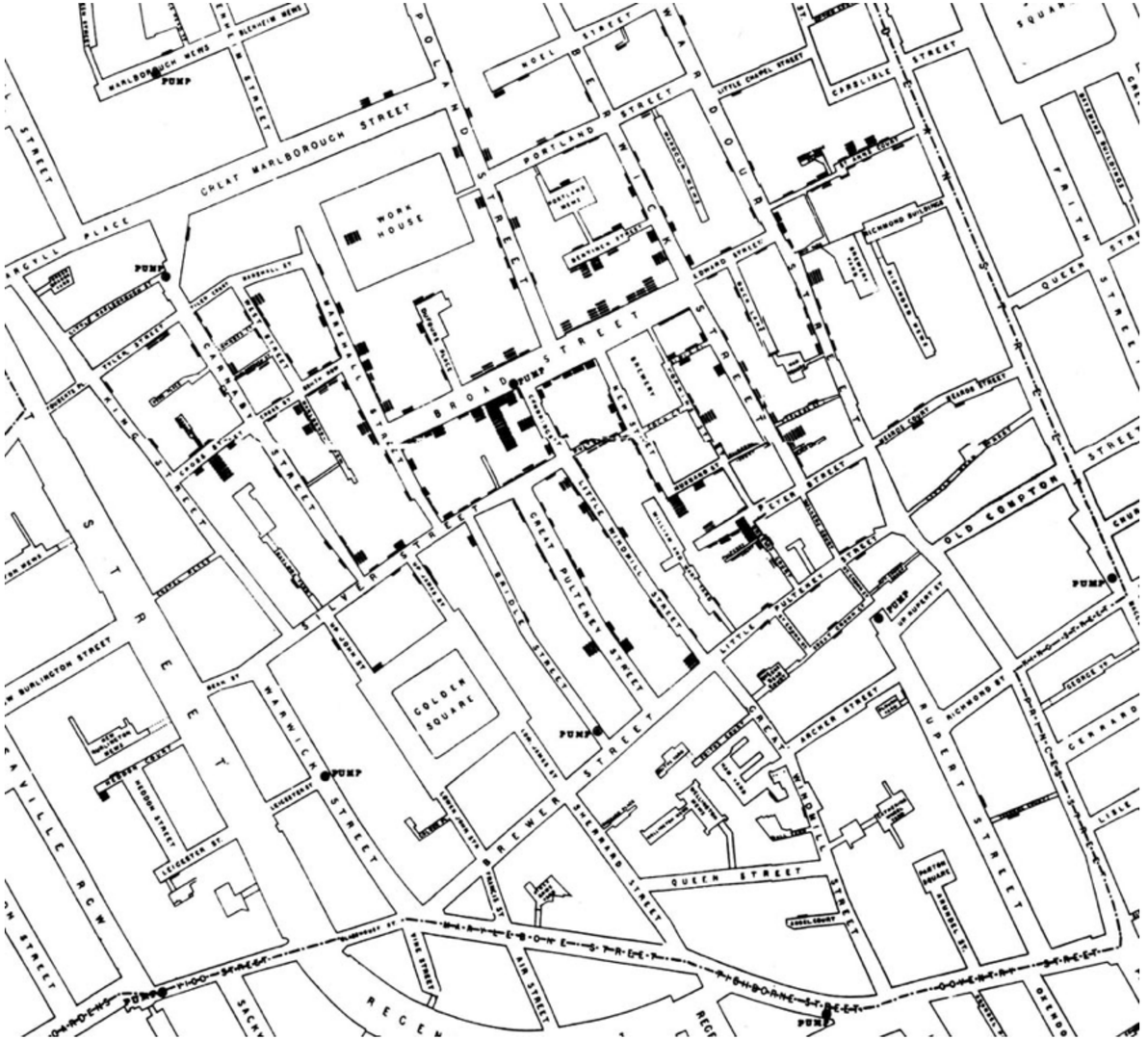


Available online at: http://www.iop.harvard.edu/programs/other/natlsurvey/april_2004.pdf

```
> library(stats)
> MyData <- read.table("mydata.txt")
> MyData.pca <- prcomp(MyData,center = TRUE,scale. = TRUE)
> print(MyData.pca)
> summary(MyData.pca)
> plot(MyData.pca,type="l")
```

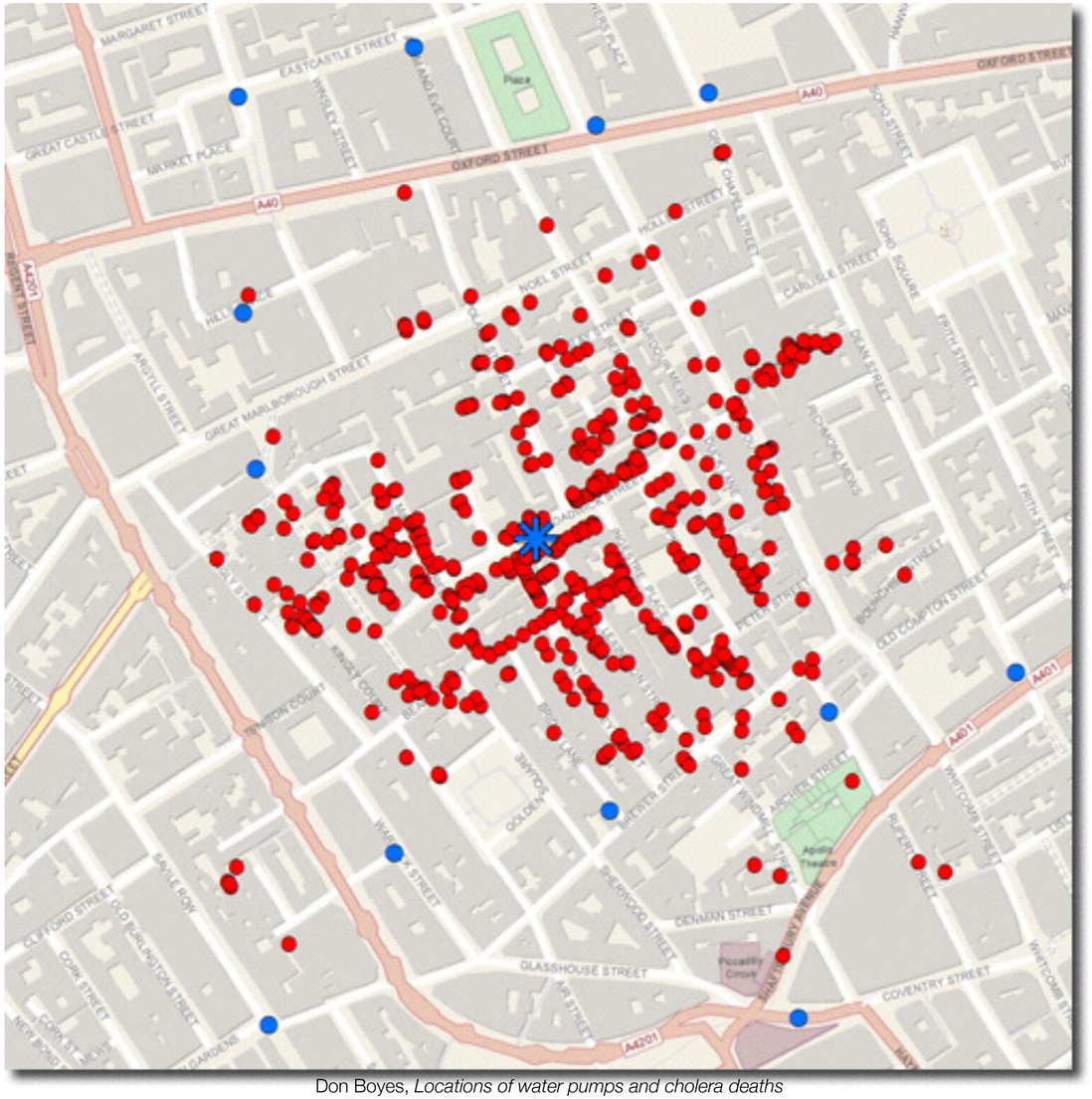
Cluster Analysis

Why Clusters Matter: John Snow and Cholera



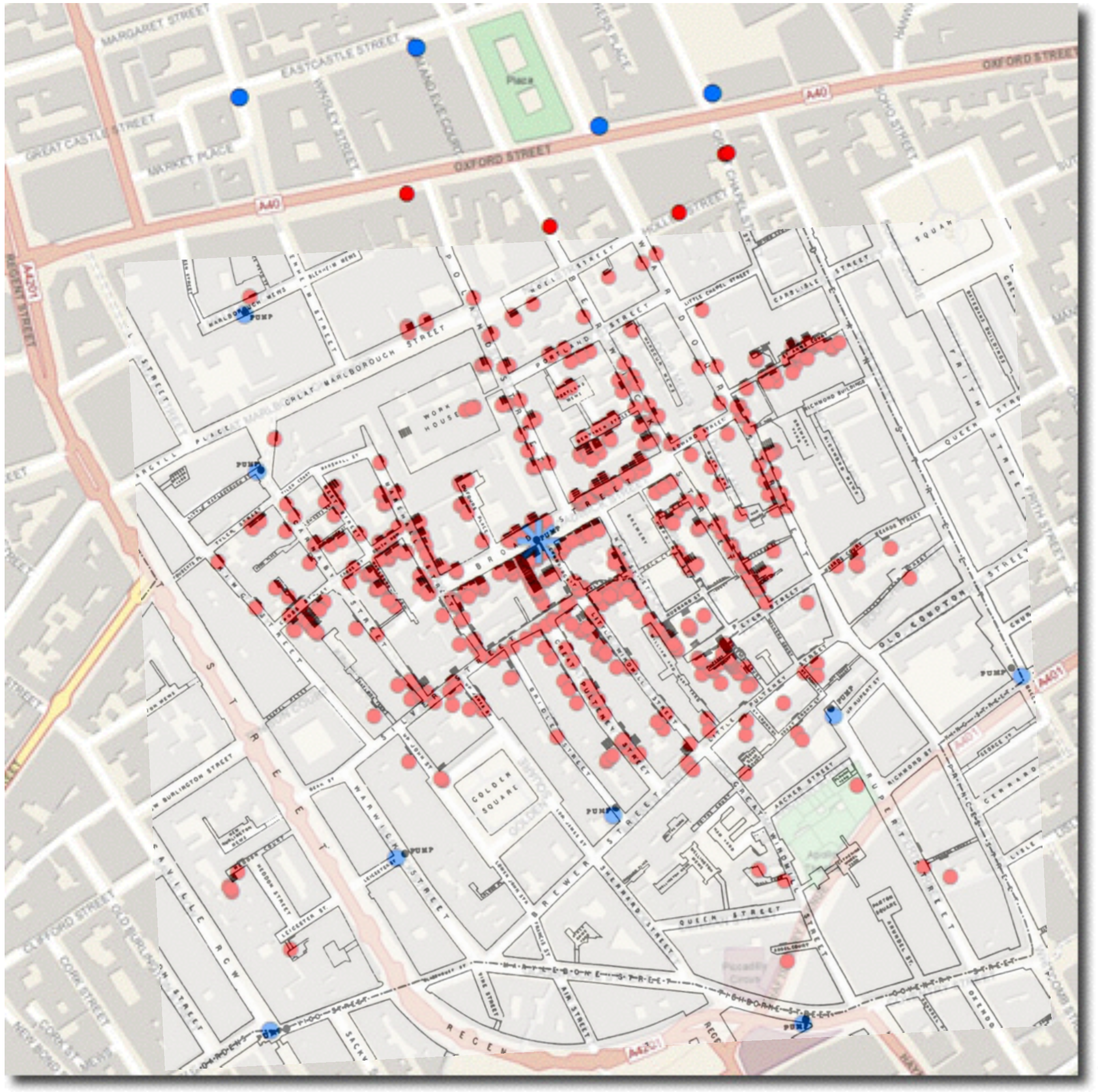
John Snow, *On the Mode of Communication of Cholera (1855)*

Why Clusters Matter: John Snow and Cholera



Don Boyes, *Locations of water pumps and cholera deaths*
http://donboyes.com/2011/10/14/john-snow-and-serendipity/pumps-and-deaths-drop/

Why Clusters Matter: John Snow and Cholera

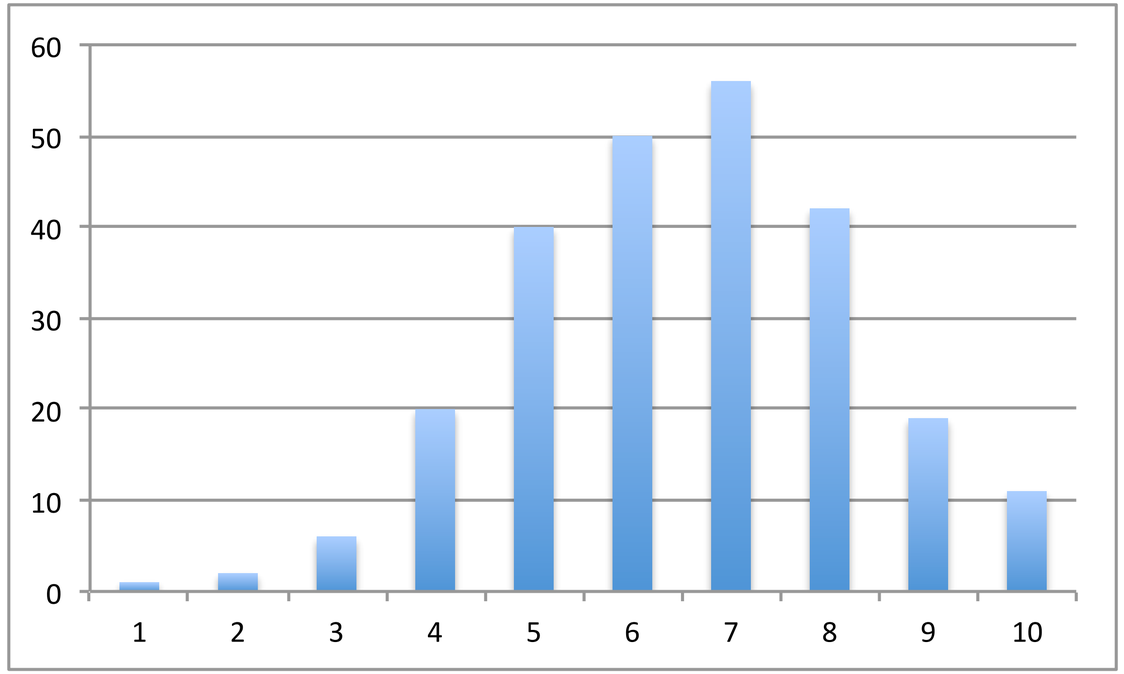


DEATH'S DISPENSARY.

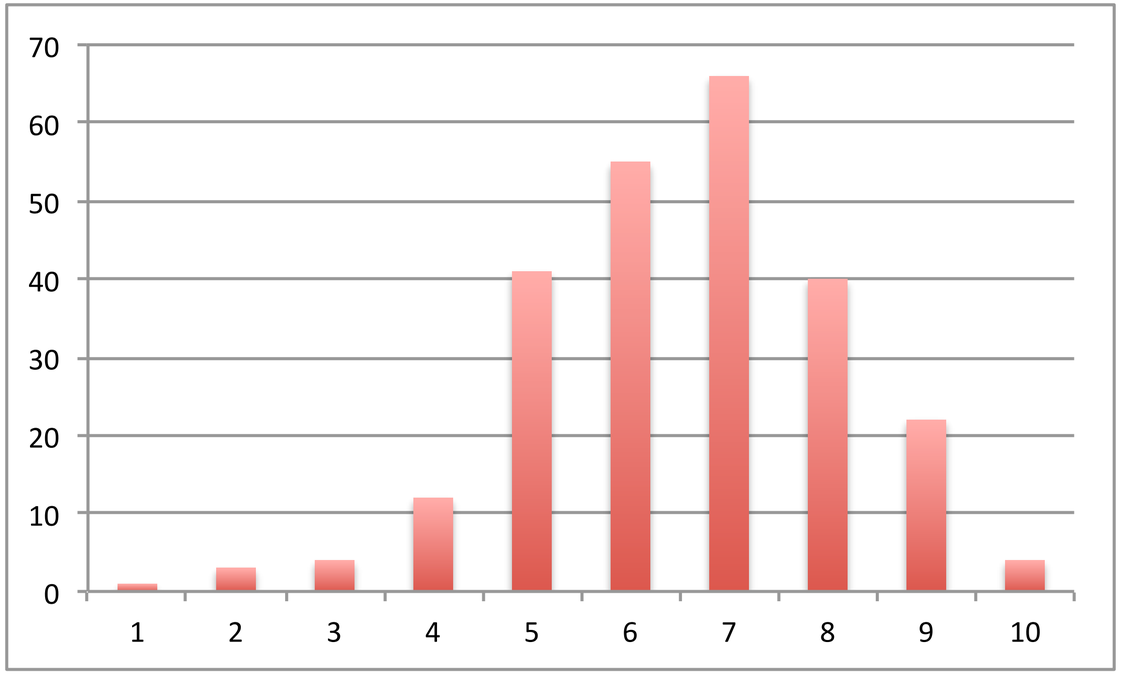
OPEN TO THE POOR, GRATIS, BY PERMISSION OF THE PARISH.

Image Credit: Justin Cormack

Some Simple Data…

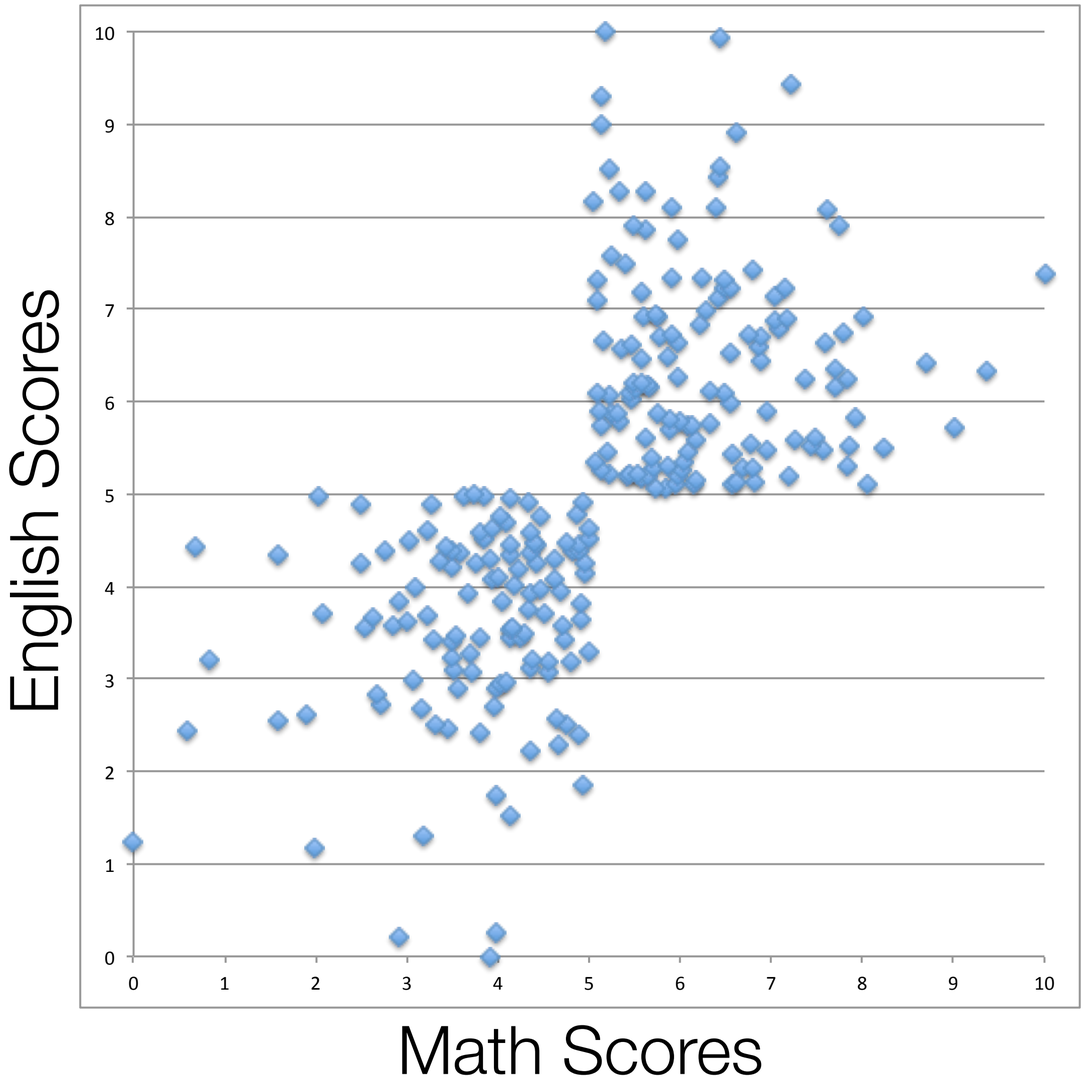


English Scores



Math Scores

…That Isn't That Simple



The OECD Pisa 2012 Data - Reading (OECD Average: 496)

Country	Score
Shanghai-China	570
Hong Kong-China	545
Singapore	542
Japan	538
Korea, Republic of	536
Finland	524
Ireland	523
Canada	523
Chinese Taipei	523
Poland	518
Estonia	516
Liechtenstein	516
New Zealand	512
Australia	512
Netherlands	511
Switzerland	509
Macao-China	509
Belgium	509
Vietnam	508
Germany	508

Country	Score
France	505
Norway	504
United Kingdom	499
United States	498
Denmark	496
Czech Republic	493
Italy	490
Austria	490
Latvia	489
Hungary	488
Spain	488
Luxembourg	488
Portugal	488
Israel	486
Croatia	485
Sweden	483
Iceland	483
Slovenia	481
Lithuania	477
Greece	477

Country	Score
Turkey	475
Russian Federation	475
Slovak Republic	463
Cyprus	449
Serbia, Republic of	446
United Arab Emirates	442
Chile	441
Thailand	441
Costa Rica	441
Romania	438
Bulgaria	436
Mexico	424
Montenegro, Republic of	422
Uruguay	411
Brazil	410
Tunisia	404
Colombia	403
Jordan	399
Malaysia	398
Indonesia	396

Country	Score
Argentina	396
Albania	394
Kazakhstan	393
Qatar	388
Peru	384

U.S. States	Score
Massachusetts	527
Connecticut	521
Florida	492

The OECD Pisa 2012 Data - Mathematics (OECD Average: 494)

Country	Score
Shanghai-China	613
Singapore	573
Hong Kong-China	561
Chinese Taipei	560
Korea, Republic of	554
Macao-China	538
Japan	536
Liechtenstein	535
Switzerland	531
Netherlands	523
Estonia	521
Finland	519
Canada	518
Poland	518
Belgium	515
Germany	514
Vietnam	511
Austria	506
Australia	504
Ireland	501

Country	Score
Slovenia	501
Denmark	500
New Zealand	500
Czech Republic	499
France	495
United Kingdom	494
Iceland	493
Latvia	491
Luxembourg	490
Norway	489
Portugal	487
Italy	485
Spain	484
Russian Federation	482
Slovak Republic	482
United States	481
Lithuania	479
Sweden	478
Hungary	477
Croatia	471

Country	Score
Israel	466
Greece	453
Serbia, Republic of	449
Turkey	448
Romania	445
Cyprus	440
Bulgaria	439
United Arab Emirates	434
Kazakhstan	432
Thailand	427
Chile	423
Malaysia	421
Mexico	413
Montenegro, Republic of	410
Uruguay	409
Costa Rica	407
Albania	394
Brazil	391
Argentina	388
Tunisia	388

Country	Score
Jordan	386
Colombia	376
Qatar	376
Indonesia	375
Peru	368

U.S. States	Score
Massachusetts	514
Connecticut	506
Florida	467

The OECD Pisa 2012 Data - Science (OECD Average: 501)

Country	Score
Shanghai-China	580
Hong Kong-China	555
Singapore	551
Japan	547
Finland	545
Estonia	541
Korea, Republic of	538
Vietnam	528
Poland	526
Canada	525
Liechtenstein	525
Germany	524
Chinese Taipei	523
Netherlands	522
Ireland	522
Australia	521
Macao-China	521
New Zealand	516
Switzerland	515
Slovenia	514

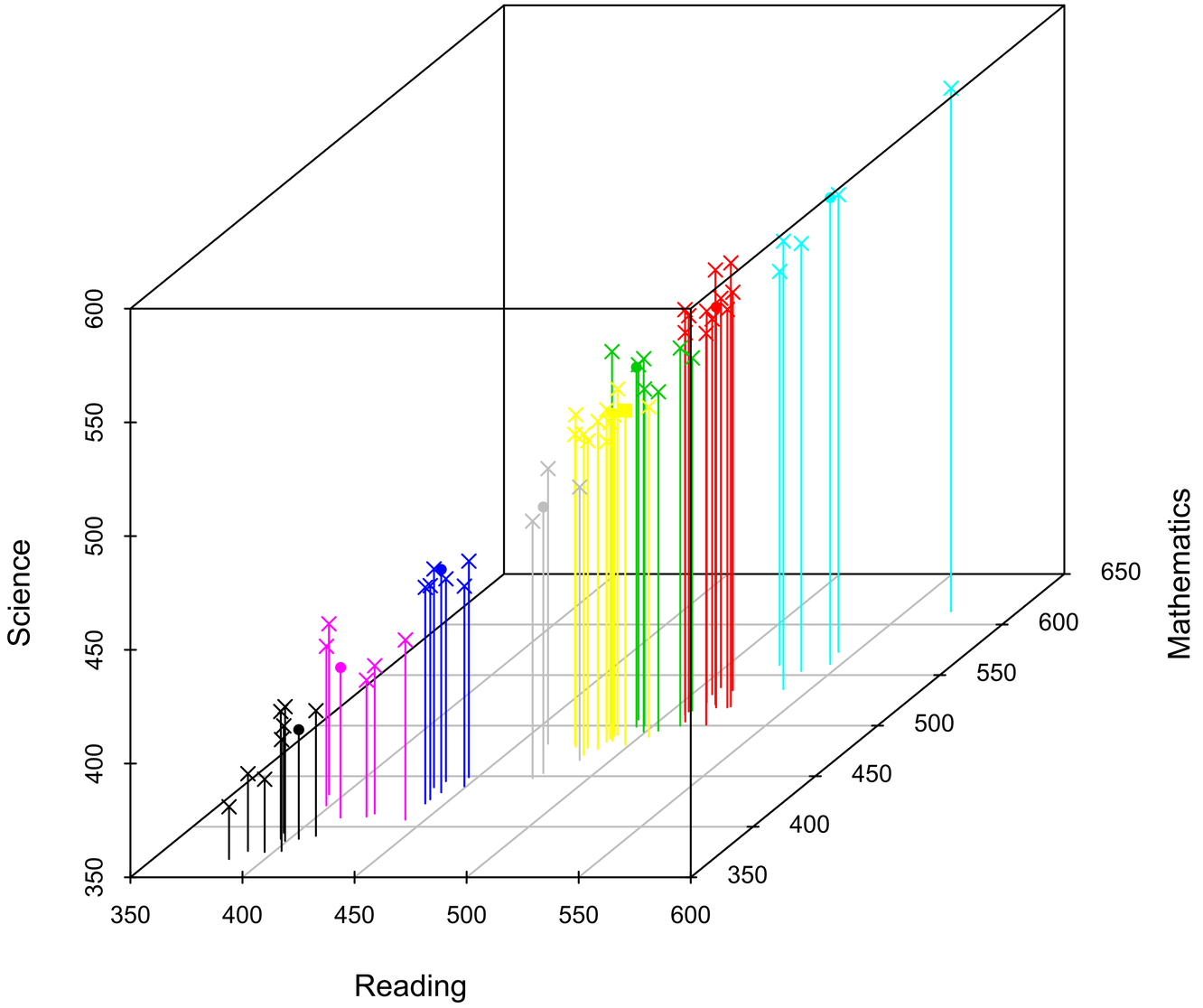
Country	Score
United Kingdom	514
Czech Republic	508
Austria	506
Belgium	505
Latvia	502
France	499
Denmark	498
United States	497
Spain	496
Lithuania	496
Norway	495
Hungary	494
Italy	494
Croatia	491
Luxembourg	491
Portugal	489
Russian Federation	486
Sweden	485
Iceland	478
Slovak Republic	471

Country	Score
Israel	470
Greece	467
Turkey	463
United Arab Emirates	448
Bulgaria	446
Chile	445
Serbia, Republic of	445
Thailand	444
Romania	439
Cyprus	438
Costa Rica	429
Kazakhstan	425
Malaysia	420
Uruguay	416
Mexico	415
Montenegro, Republic of	410
Jordan	409
Argentina	406
Brazil	405
Colombia	399

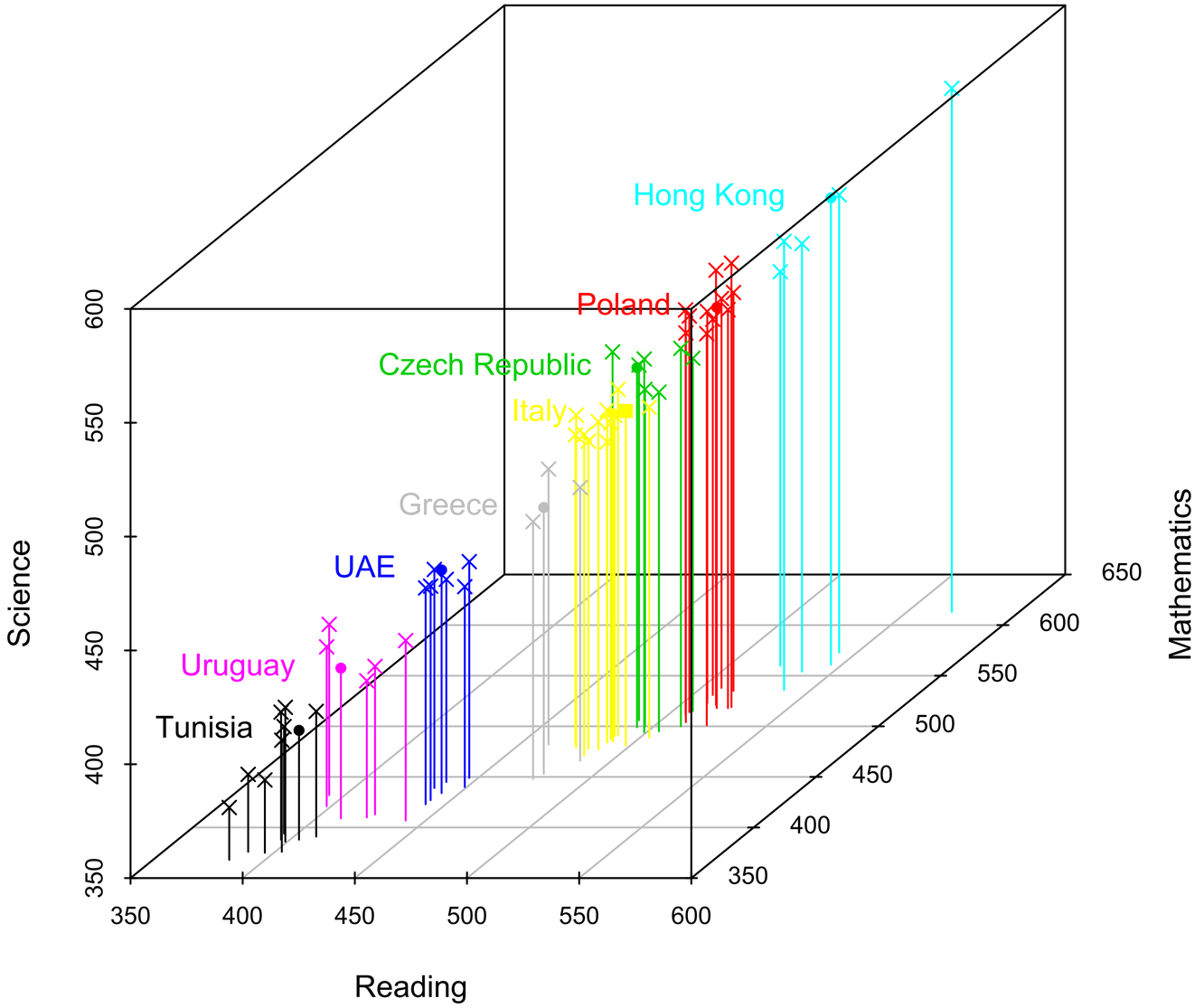
Country	Score
Tunisia	398
Albania	397
Qatar	384
Indonesia	382
Peru	373

U.S. States	Score
Massachusetts	527
Connecticut	521
Florida	485

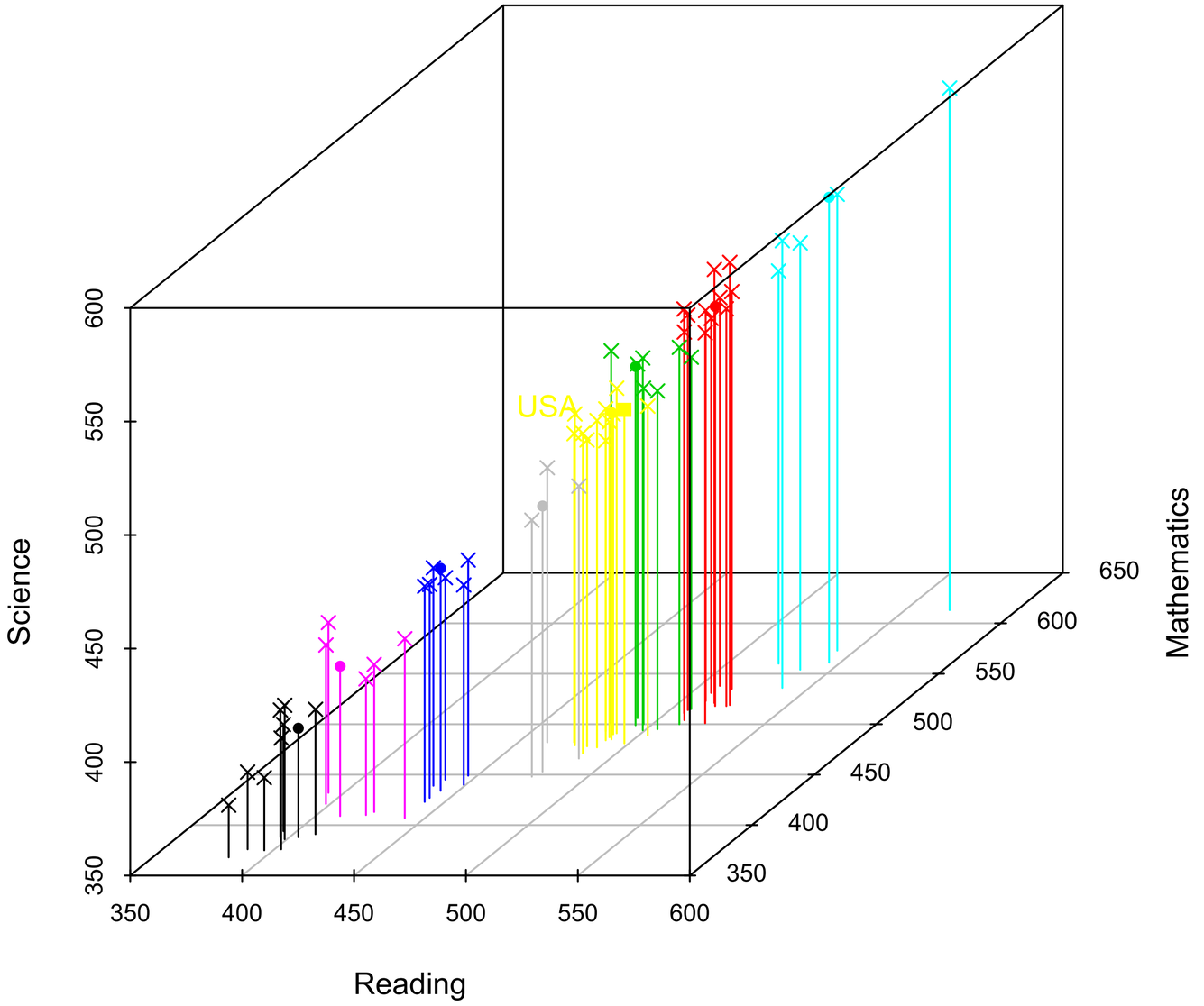
OECD PISA 2012



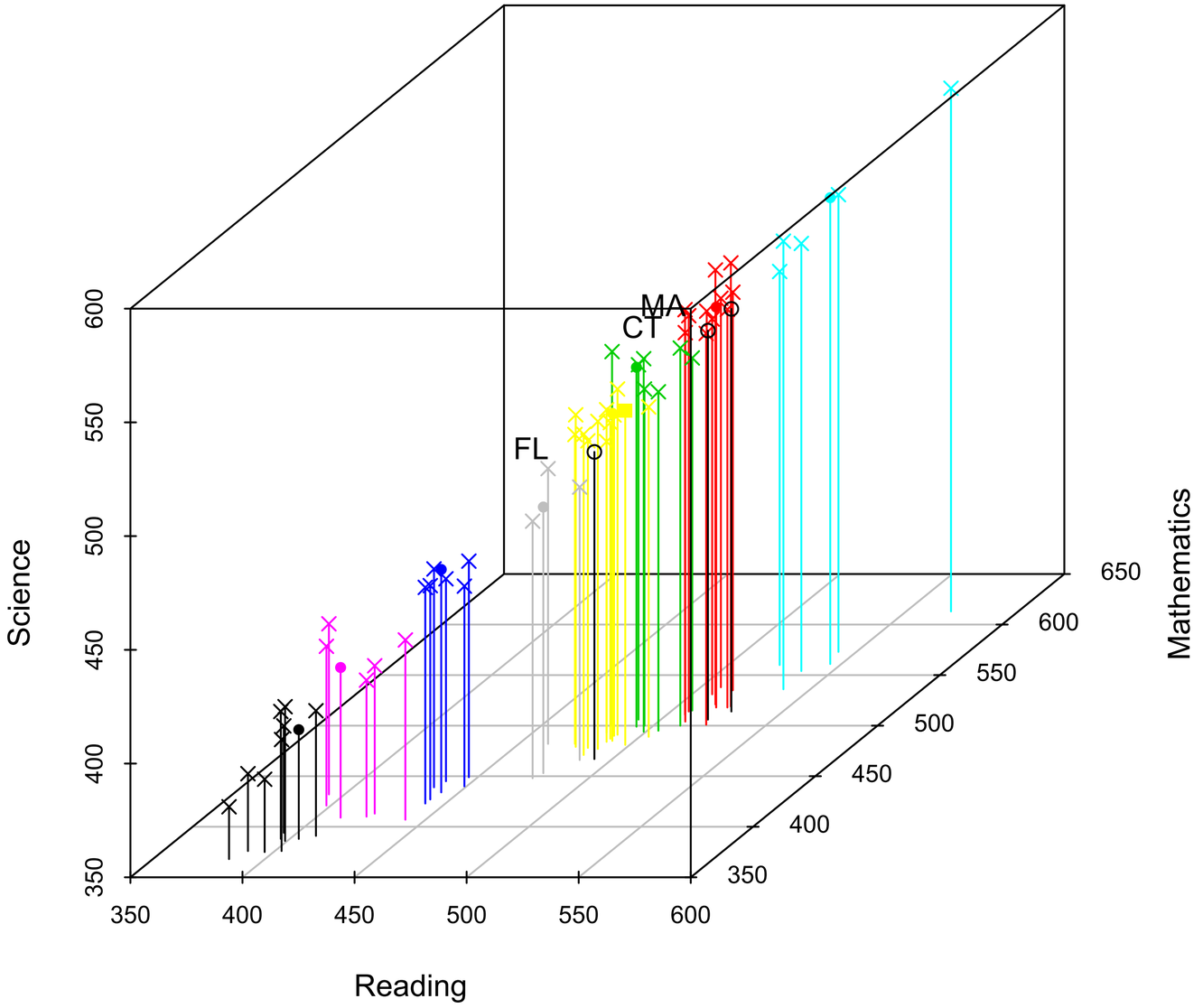
OECD PISA 2012



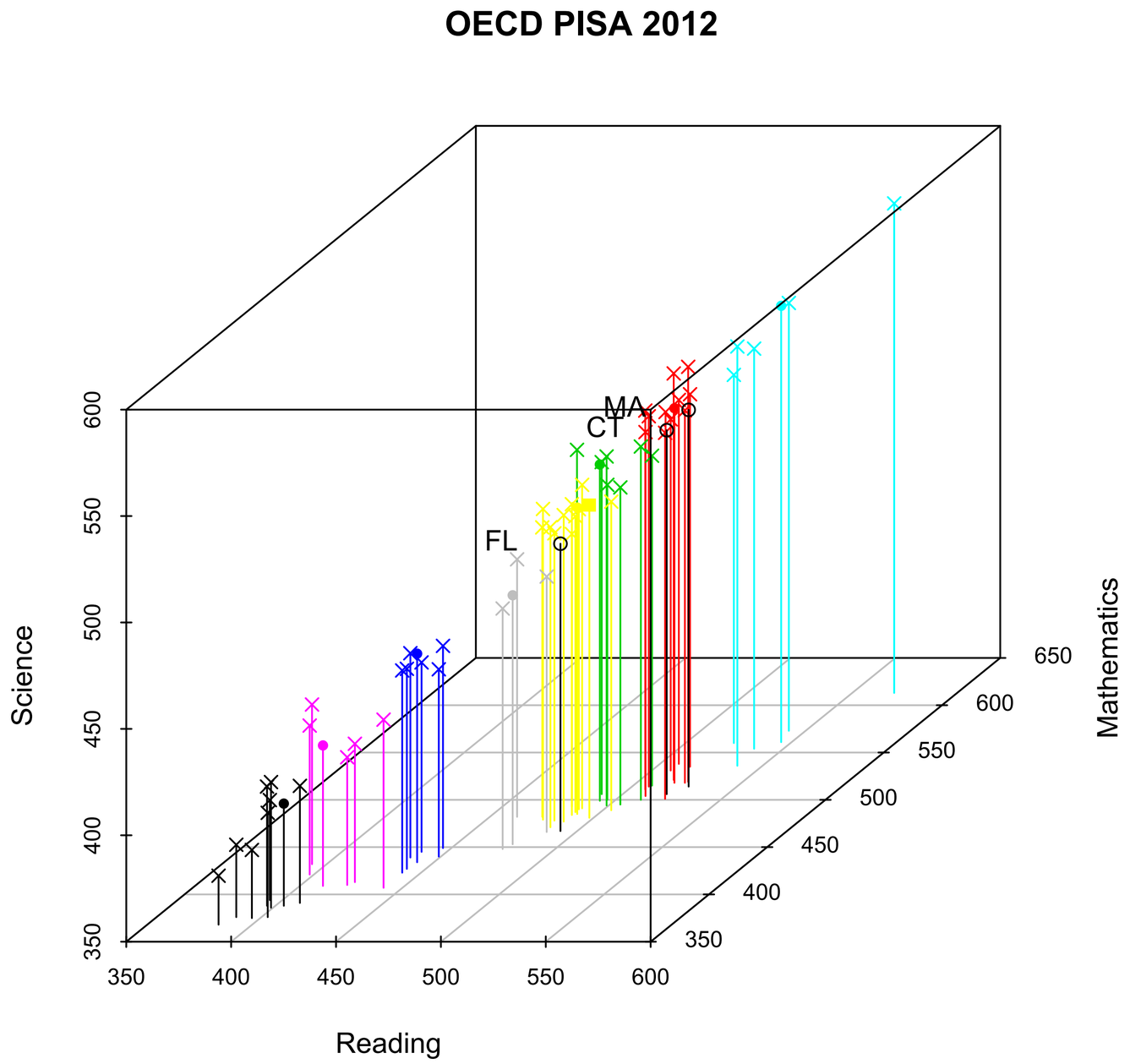
OECD PISA 2012



OECD PISA 2012

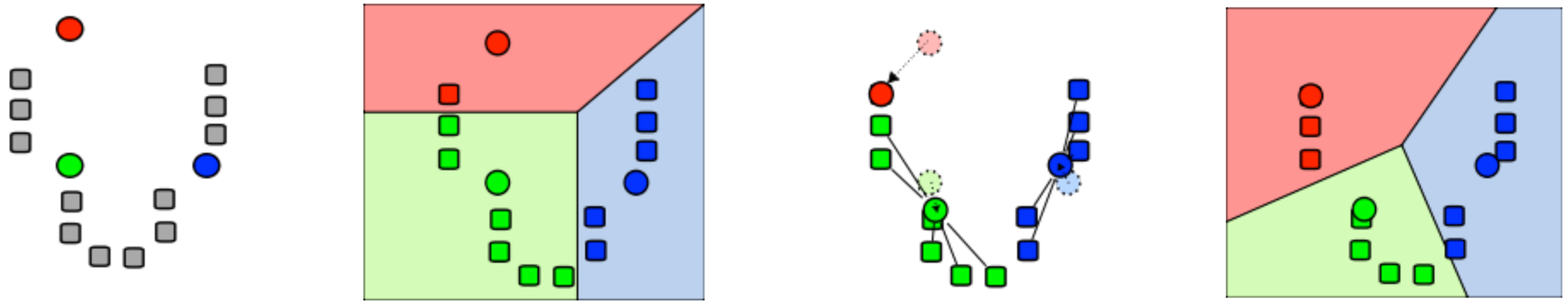


Croatia
Hungary
Iceland
Italy
Latvia
Lithuania
Luxembourg
Norway
Portugal
Russian Federation
Spain
Sweden
USA



Australia
Canada
Estonia
Finland
Germany
Ireland
Liechtenstein
Macao-China
Netherlands
Poland
Switzerland
Vietnam

Finding Groups: k-Means Clustering

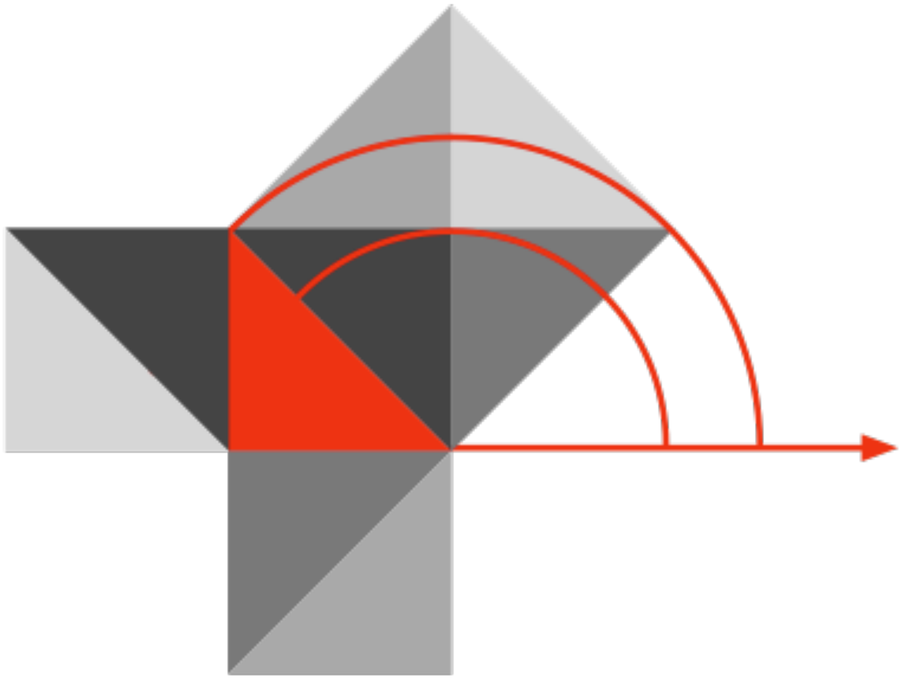


```
> library(cluster)
> MyData <- read.table("mydata.txt")
> mydatapamfour <- pam(MyData,4)
> summary(mydatapamfour)
> plot(mydatapamfour)
```

Additional Bibliography

- **General:**
  - Hanneman, R.N. and Riddle M. (2005). *Introduction to Social Network Methods*. Available online at: http://faculty.ucr.edu/~hanneman/
  - Kline, P. (1994). *An Easy Guide to Factor Analysis*. New York, New York: Routledge.
  - Taylor, A.D. and Pacelli, A.M. (2008). *Mathematics and Politics: Strategy, Voting, Power and Proof - Second Edition*. New York, New York: Springer.
- **Weak Ties:**
  - Granovetter, M.S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78 (May), 1360-1380.
  - Granovetter, M.S. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, Volume 1, 201-233.
  - Rogers, E.M. (2003). *Diffusion of Innovations, Fifth Edition*. New York, New York: The Free Press.
- **Social Media and Teams:**
  - Muller, M., Ehrlich, K., Matthews, T., Perer, A. A., Ronen, I., and Guy, I. Diversity among enterprise online communities: Collaborating, teaming, and innovating through social media. 2012 ACM SIGCHI Conf. on Human Factors in Computing Systems (2012). Available online at: http://perer.org/papers/adamPerer-CHI2012.pdf

Hippasus



Blog: http://hippasus.com/blog/
Email: rubenrp@hippasus.com
Twitter: @rubenrp