Antitrust Notice

Munich RE

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Cluster Analysis in Loss Reserving

Dave Clark
Diana Rangelova
Christina Oda

CLRS - September 2017

Munich RE

Image: used under license from shutterstock.co

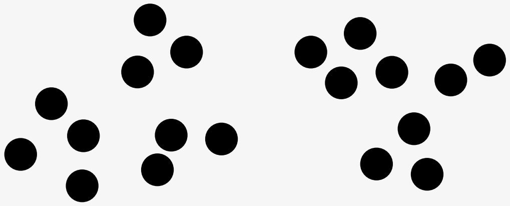
Agenda

Munich RE

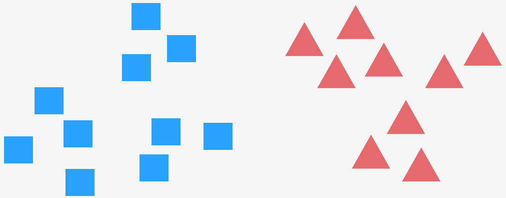
1. Introduction: Purpose of clustering in reserving
2. How to find clusters
    a) Cluster Analysis
    b) Principal Component Analysis (PCA)
    c) Data transformation (curve fitting)
3. Practical considerations and observations
4. Next steps

Clustering

- Clustering is about finding groups in a set of objects
  - The objects in a group should be similar and groups should be different from each other
  - No need to define the groups in advance (i.e. unsupervised learning)
  - Essential to assess the usefulness and meaning of the identified groups



Original data



Two clusters

Introduction
Why Clustering?

- What reserving questions could be answered with cluster analysis?
  - Test the data homogeneity
  - Find a benchmark
  - Identify drivers of development

- What kind of data can be clustered?
  - Segments, contracts or claims
  - Loss development patterns, loss ratios, severity, frequency…

How to Find Clusters?

- Exploratory Data Analysis
  - Cluster analysis
  - Principal Component Analysis (PCA)
  - Data transformation (curve fitting)

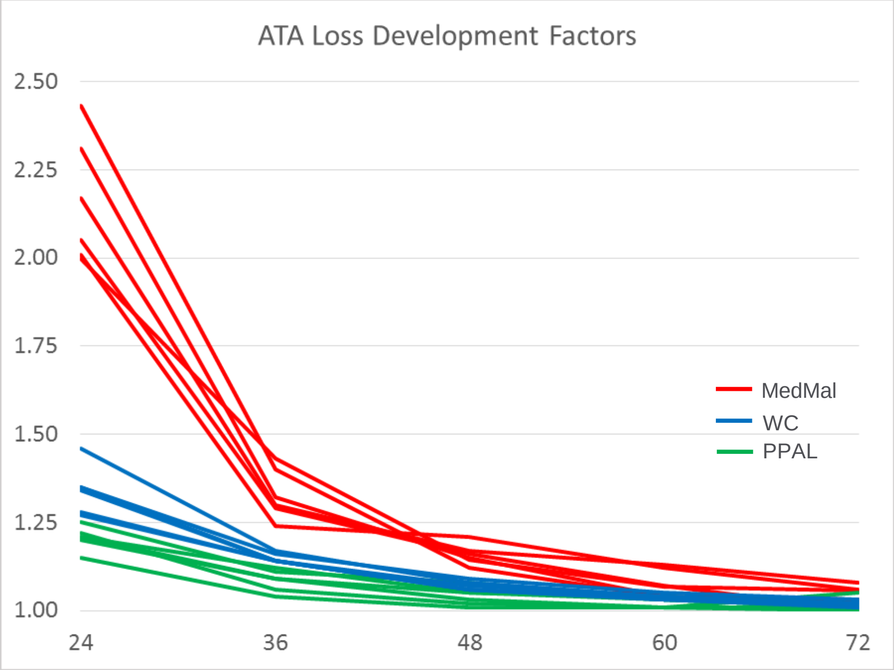
Introduction
Schedule P Example

Co.	Line	Ownership	Geographic	Distribution
1	MedMal	Mutual	Regional	Direct, Ind Agency
2	MedMal	Stock	National	Direct, Ind Agency
3	PPAL	Stock	National	MGA, Ind Agency
4	PPAL	Stock	Regional	Ind Agency
5	WC	Stock	National	MGA
6	WC	Mutual	Regional	Ind Agency

…

Co.	24	36	48	60	72
1	2.01	1.24	1.21	1.12	1.06
2	2.05	1.29	1.16	1.07	1.00
3	1.20	1.09	1.05	1.03	1.01
4	1.15	1.04	1.01	1.01	1.00
5	1.34	1.14	1.07	1.04	1.02
6	1.28	1.14	1.06	1.04	1.02

…



Introduction
Where to Start?

Explanatory Variables

Variables used for clustering, PCA, …

Co.	Line	Ownership	Geographic	Distribution	24	36	48	60	72
1	MedMal	Mutual	Regional	Direct, Ind Agency	2.01	1.24	1.21	1.12	1.06
2	MedMal	Stock	National	Direct, Ind Agency	2.05	1.29	1.16	1.07	1.00
3	PPAL	Stock	National	MGA, Ind Agency	1.20	1.09	1.05	1.03	1.01
4	PPAL	Stock	Regional	Ind Agency	1.15	1.04	1.01	1.01	1.00
5	WC	Stock	National	MGA	1.34	1.14	1.07	1.04	1.02
6	WC	Mutual	Regional	Ind Agency	1.28	1.14	1.06	1.04	1.02
			…				…		

Cluster Analysis
How to Find Clusters?

- Exploratory Data Analysis
  - Cluster Analysis
  - Principal Component Analysis (PCA)
  - Data transformation (curve fitting)

Types of Clustering

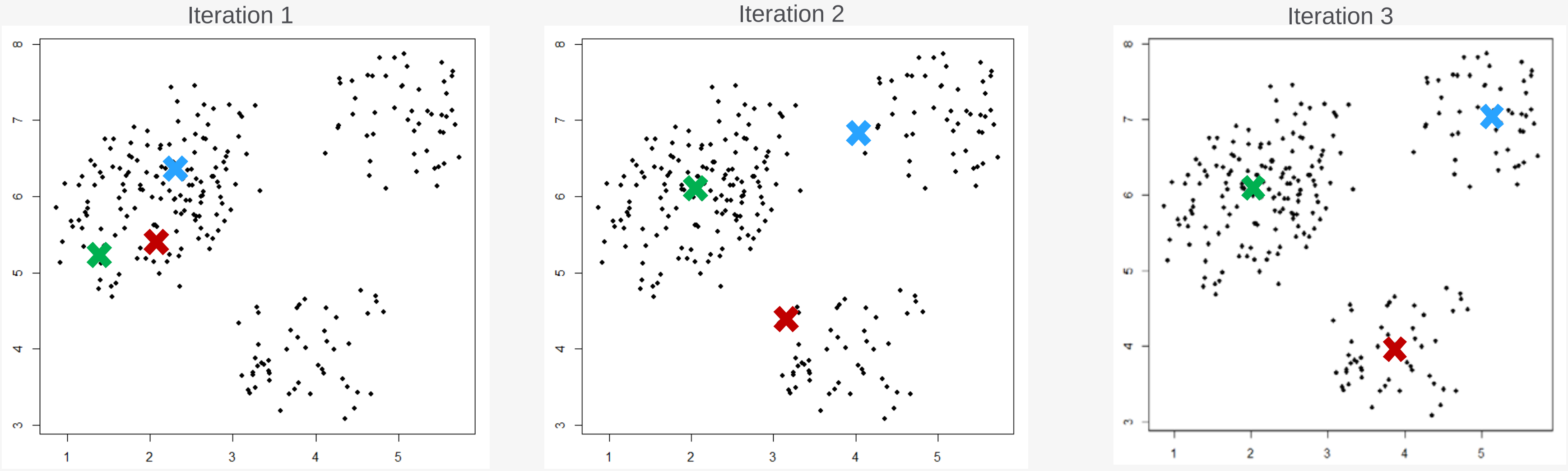
Munich RE

- Types of clustering algorithms
  - Hierarchical vs. Partitioned
  - Hard vs. Soft (ex: K-means vs. Fuzzy C-means)
  - Complete vs. Partial
  - Density Based Clusters (ex: DBSCAN)

- **K-means** partitions the data in a user-specified number of clusters (K), in which each observation belongs to the cluster with the nearest mean.

Cluster Analysis
How does K-means work?

- Initiate the centroids
- Assign points to the closest centroid
- Recalculate new centroid
- Iterate until no point has to be reassigned



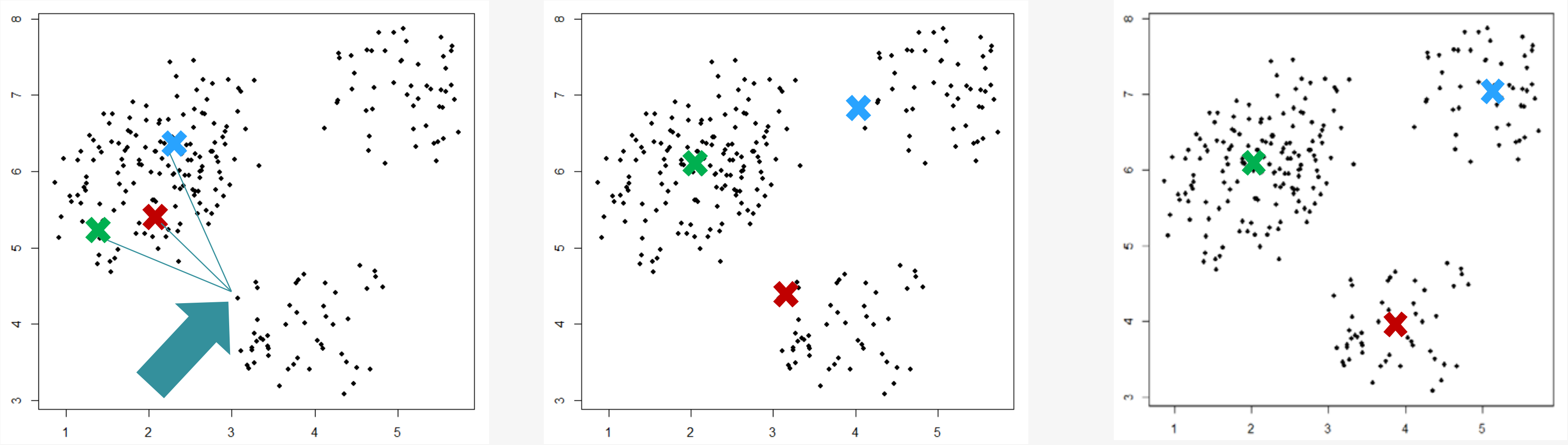
Cluster Analysis
How does K-means work?

- Initiate the centroids
- Assign points to the closest centroid
- Recalculate new centroid
- Iterate until no point has to be reassigned

Iteration 1

Iteration 2

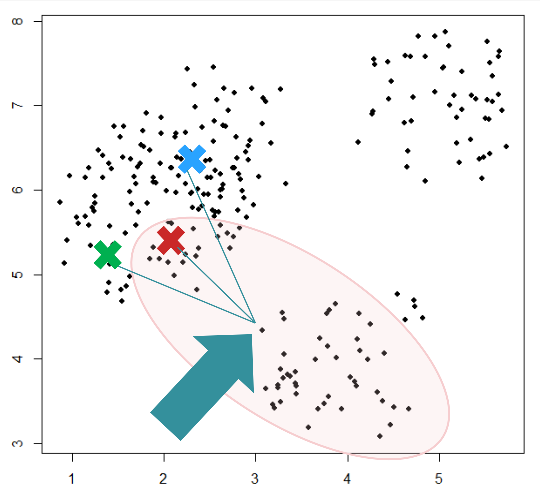
Iteration 3



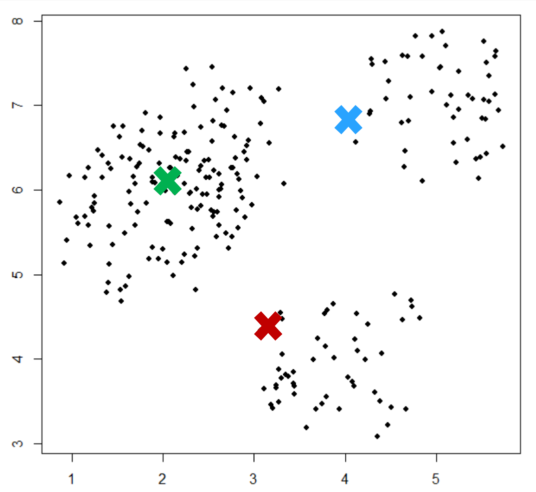
Cluster Analysis
How does K-means work?

- ➢ Initiate the centroids
- ➢ Assign points to the closest centroid
- ➢ Recalculate new centroid
- ➢ Iterate until no point has to be reassigned

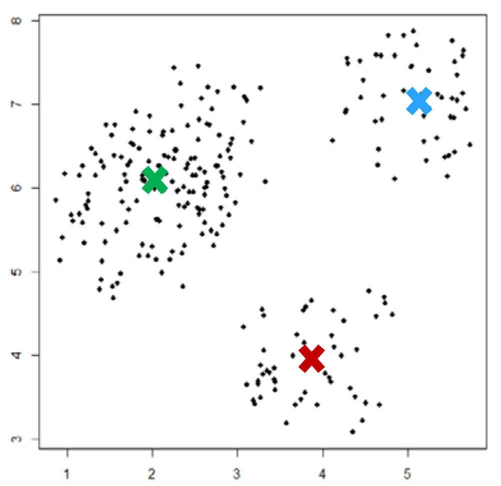
Iteration 1



Iteration 2



Iteration 3

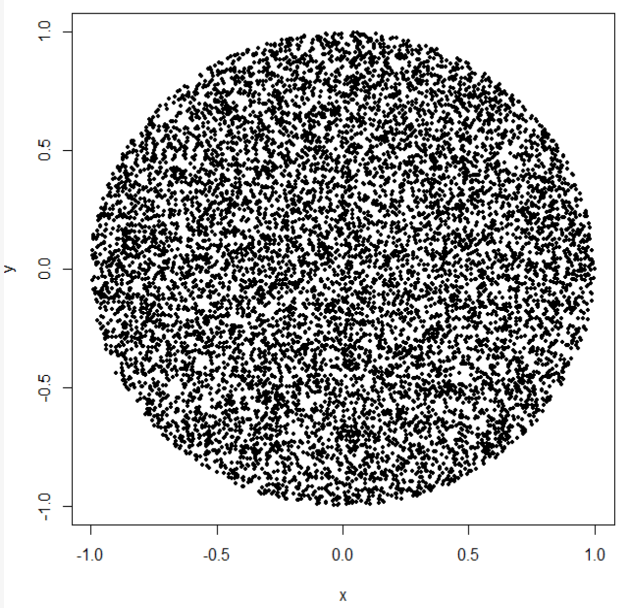


Cluster Analysis
Pros & Cons

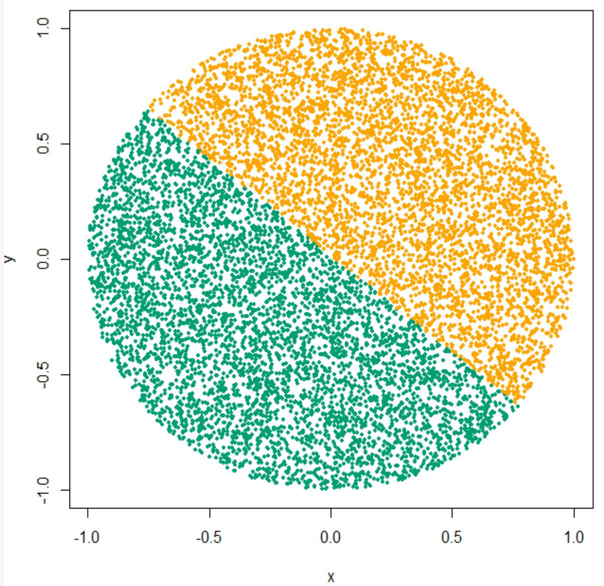
- K-means is simple, fast and efficient
- K-means does not perform well when:
  - There are no natural clusters
  - Clusters are of different size
  - Clusters are not spherical
  - Outliers exist

Cluster Analysis
No natural clusters

Original data

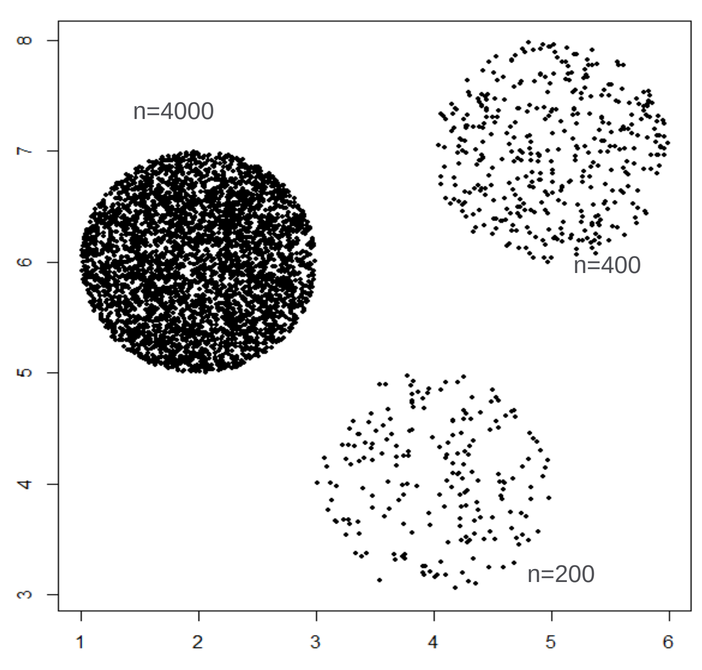


K-means result

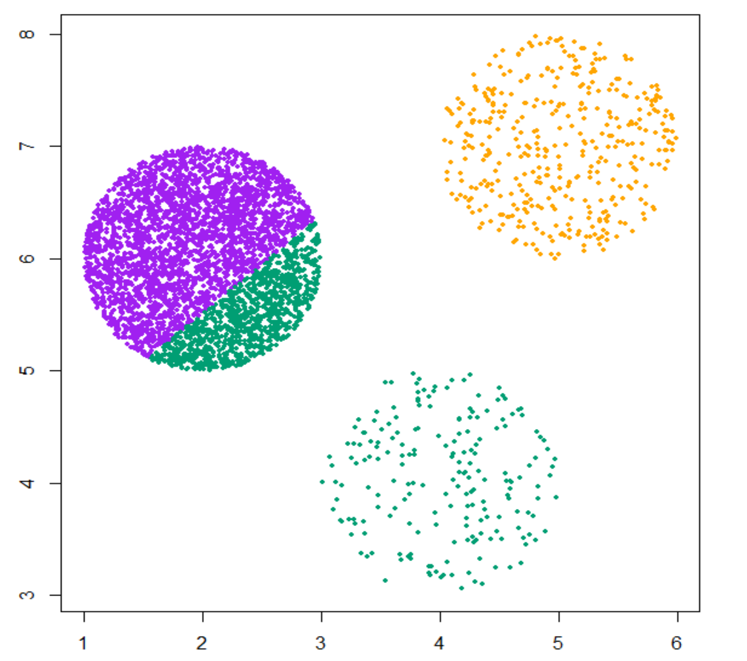


Cluster Analysis
Clusters of different size

Original data

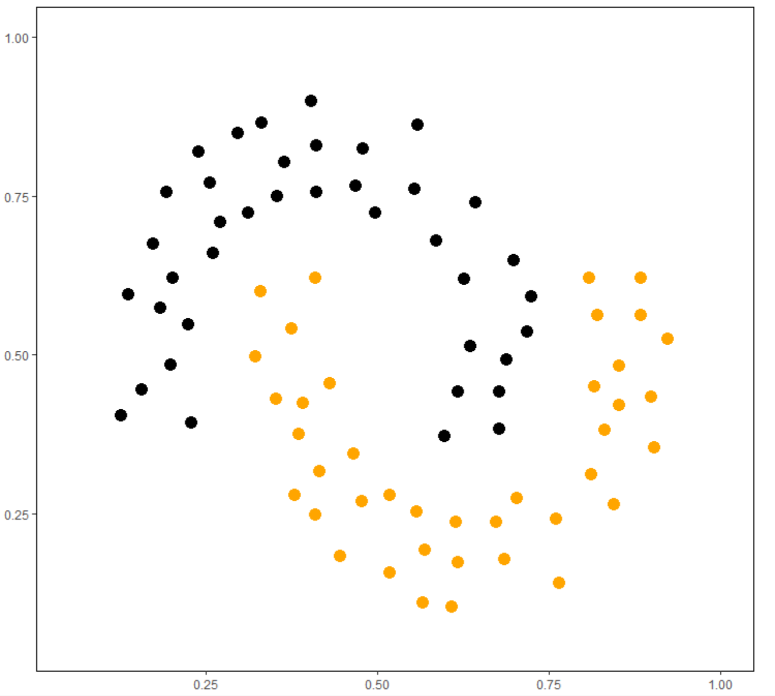


K-means result

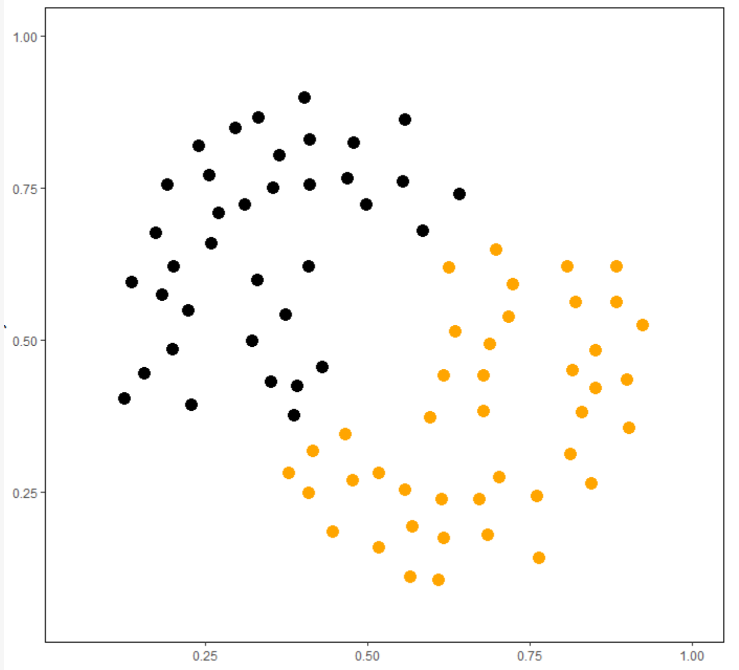


Cluster Analysis
Non spherical clusters

Original data



K-means result



Source: Introduction to Data Mining [5]

Cluster Analysis
Outliers

- Outliers make the centroid less representative
- Eliminate outliers prior to clustering
- K-medoids: variation of K-means where the centroids are actual data points

Cluster Analysis
How to perform the Clustering?

- **Use scaled and centered data for clustering**

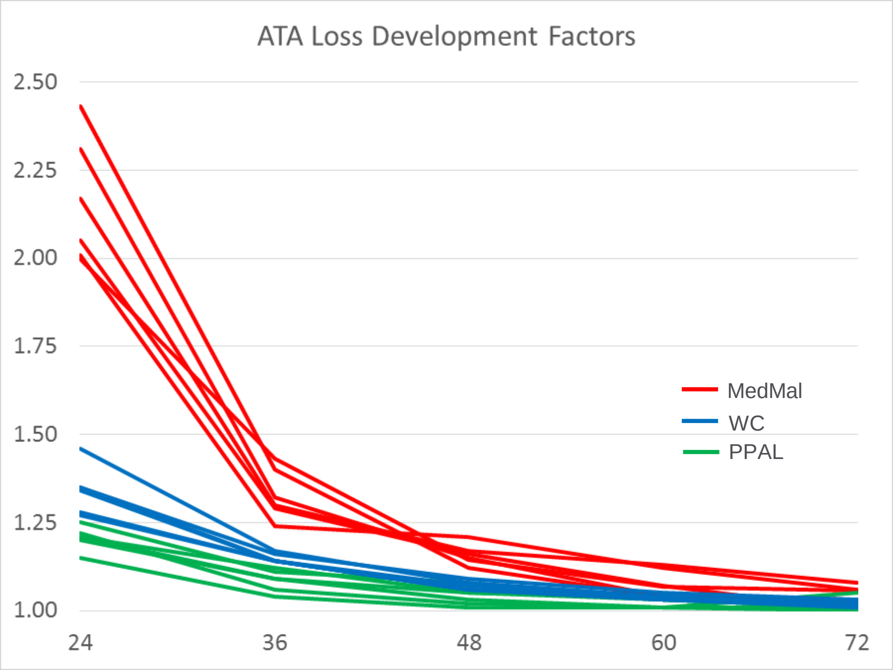
- **R package ‘*cluster*’**
  - K-means: *kmeans*(data, k=2, …)
  - K-medoids: *pam*(data, k=2,…)
  - Fuzzy clustering: *fanny*(data, k=2,…)

- **SAS**
  - Proc FASTCLUS

    https://support.sas.com/documentation/cdl/en/statugclustering/61759/PDF/default/statugclustering.pdf

Cluster Analysis

Schedule P example: Cluster Analysis



LOB	K-means 2 clusters	K-means 3 clusters	K-medoids 3 clusters
MedMal	1	1	1
MedMal	1	1	1
MedMal	1	2	1
MedMal	1	1	1
MedMal	1	2	1
MedMal	1	2	1
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
WC	2	3	3
WC	2	3	3
WC	2	3	3
WC	2	3	3
WC	2	3	3
WC	2	3	3

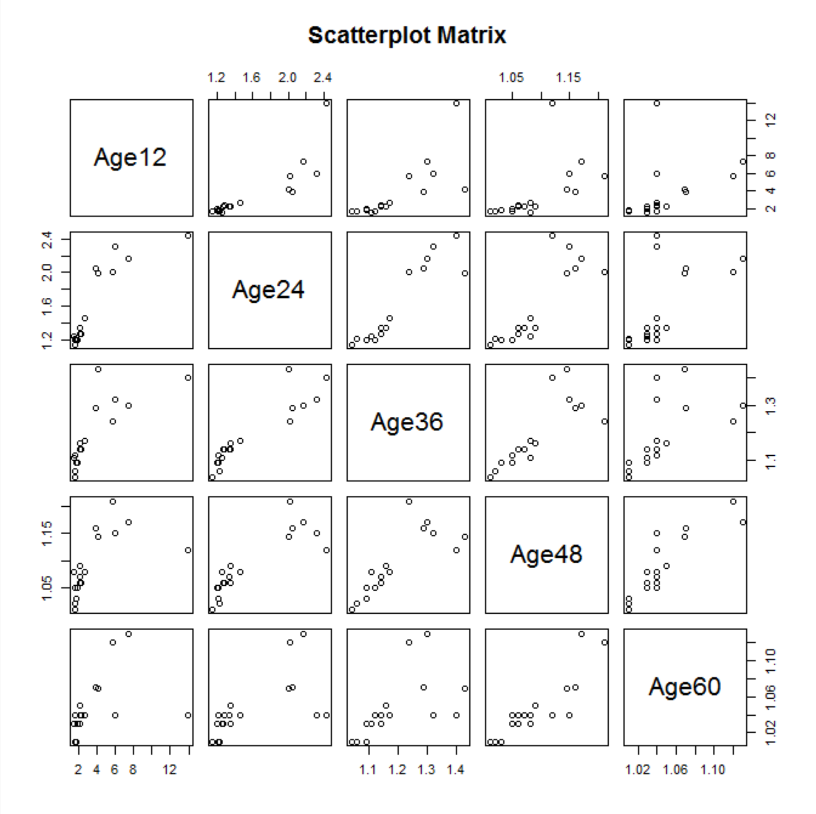
Cluster Analysis

Too Many Dimensions

Data

12	24	36	48	60	72
5.70	2.01	1.24	1.21	1.12	1.06
3.86	2.05	1.29	1.16	1.07	1.00
1.92	1.20	1.09	1.05	1.03	1.01
1.64	1.15	1.04	1.01	1.01	1.00
2.19	1.34	1.14	1.07	1.04	1.02
2.33	1.28	1.14	1.06	1.04	1.02

…



PCA
How to Find Clusters?

- Exploratory Data Analysis
  - Cluster Analysis
  - Principal Component Analysis (PCA)
  - Data transformation (curve fitting)

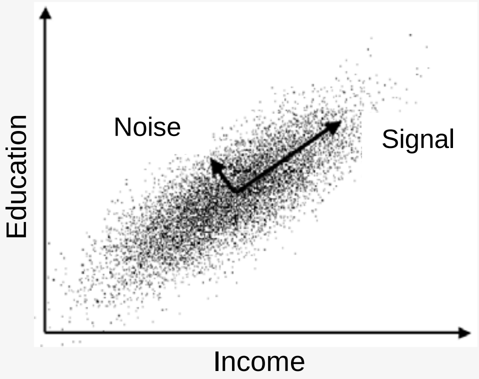
- PCA reduces the dimensions of the data and keeps the signal

Example: Socioeconomic status

Redundant Information

Correlated variables

Noise & Signal

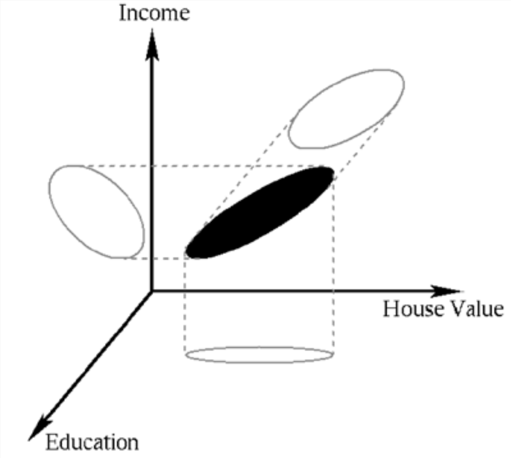


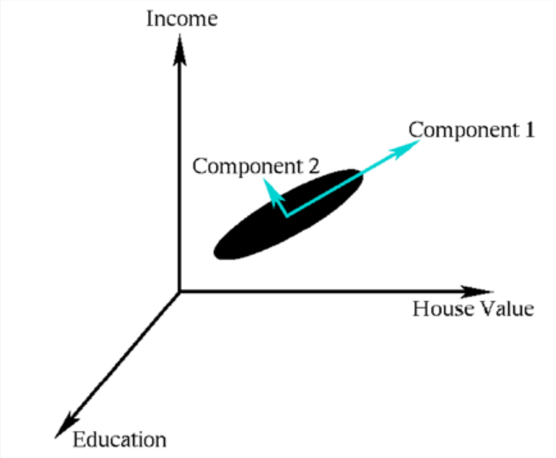
*Education = c*Income*

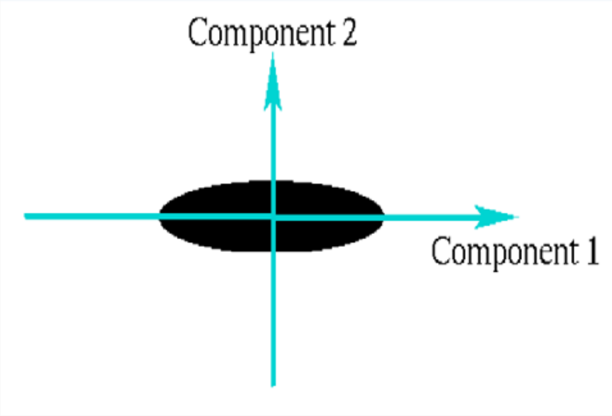
PCA
How does PCA work?

- Finds the most meaningful basis to re-express complex data
  - Minimizes redundancy by using orthogonal components
  - Maximizes signal by taking a linear combination of the dimensions

New Coordinate Basis







What are PC1 & PC2?

Munich RE

- Principal Component are linear combinations of the original data dimensions
- How to find them?

Eigenvalues & Eigenvectors

Eigenvector

$$A \; x = \lambda \; x$$

Square matrix

Eigenvalue

- Most square matrixes with n dimensions have n eigenvectors
- Each eigenvector has an eigenvalue
- The magnitude of the eigenvalues is an indicator of how much variance is captured by each eigenvector

PCA

How to perform a PCA?

Data

Correlation matrix

Eigenvalues & Eigenvectors

With Excel					
Co.	**24**	**36**	**48**	**60**	**72**
1	2.01	1.24	1.21	1.12	1.06
2	2.05	1.29	1.16	1.07	1.00
3	1.20	1.09	1.05	1.03	1.01
4	1.15	1.04	1.01	1.01	1.00
…			…		

With R					
Co.	**24**	**36**	**48**	**60**	**72**
1	2.01	1.24	1.21	1.12	1.06
2	2.05	1.29	1.16	1.07	1.00
3	1.20	1.09	1.05	1.03	1.01
4	1.15	1.04	1.01	1.01	1.00
…			…		

AGE	**24**	**36**	**48**	**60**	**72**
24					
36	=*CORREL*(Var1,Var2,…)				
48					
60					
72					

prcomp(data, scale=TRUE,…)

VBA code for Eigenvalue/vectors:
http://www.freevbcode.com/ShowCode.asp?ID=9209

PCA
How to perform a PCA?

Data x Eigenvalues & Eigenvectors = Principal Components

Co.	24	36	48	60	72
1	2.01	1.24	1.21	1.12	1.06
2	2.05	1.29	1.16	1.07	1.00
3	1.20	1.09	1.05	1.03	1.01
4	1.15	1.04	1.01	1.01	1.00
…			…		

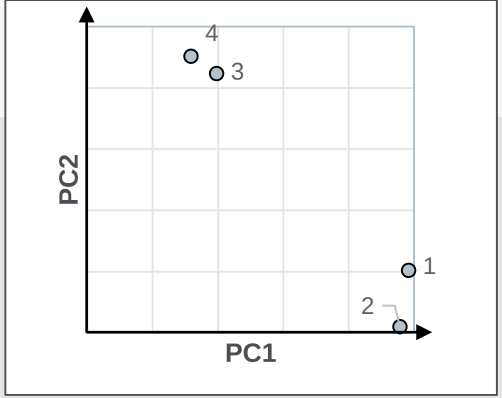
x

Dim	1	2
24	0.47	(0.39)
36	0.46	(0.38)
48	0.50	(0.11)
60	0.46	0.35
72	0.33	0.75

=

Co.	PC1	PC2
1	2.98	(0.20)
2	2.96	(0.29)
3	2.40	0.12
4	2.32	0.15
…	…	

Scatterplot



PCA
Interpretation

Munich RE

- PCA provides an opportunity for interpretation
  - PC1 captures the mean development
  - PC2 indicates a change in the curve shape

Co.	24	36	48	60	72
1	2.01	1.24	1.21	1.12	1.06
2	2.05	1.29	1.16	1.07	1.00
3	1.20	1.09	1.05	1.03	1.01
4	1.15	1.04	1.01	1.01	1.00
…	…				

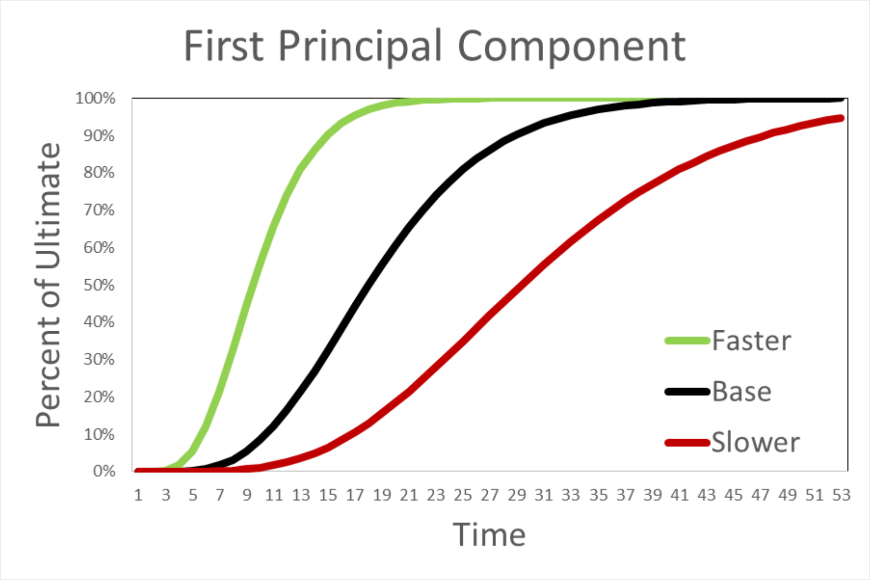
x

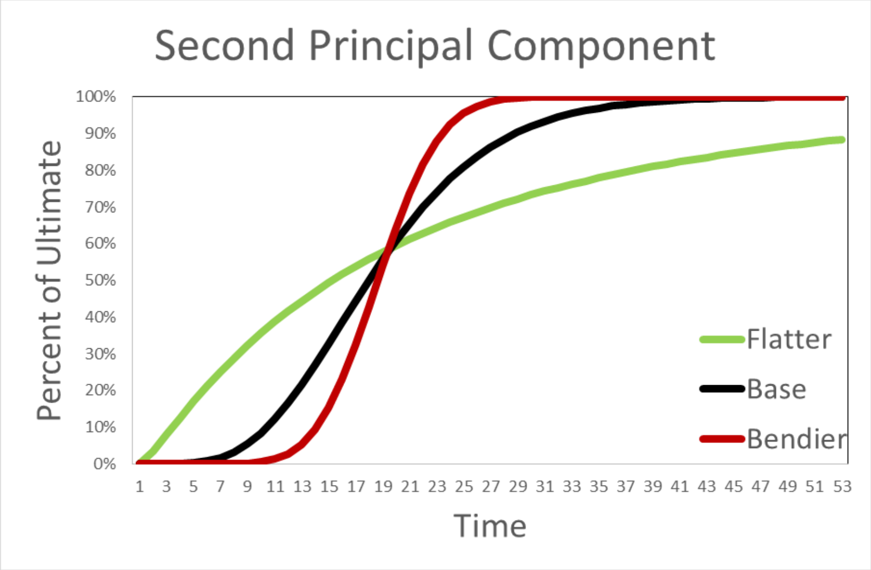
Dim	1	2
24	0.47	(0.39)
36	0.46	(0.38)
48	0.50	(0.11)
60	0.46	0.35
72	0.33	0.75

=

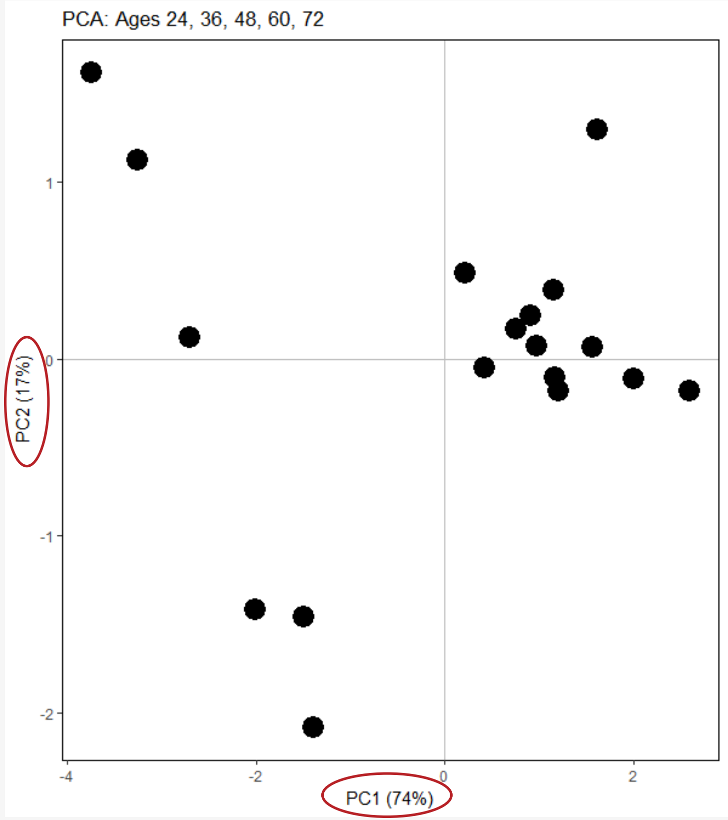
Co.	PC1	PC2
1	2.98	(0.20)
2	2.96	(0.29)
3	2.40	0.12
4	2.32	0.15
…	…	

PCA
Interpretation



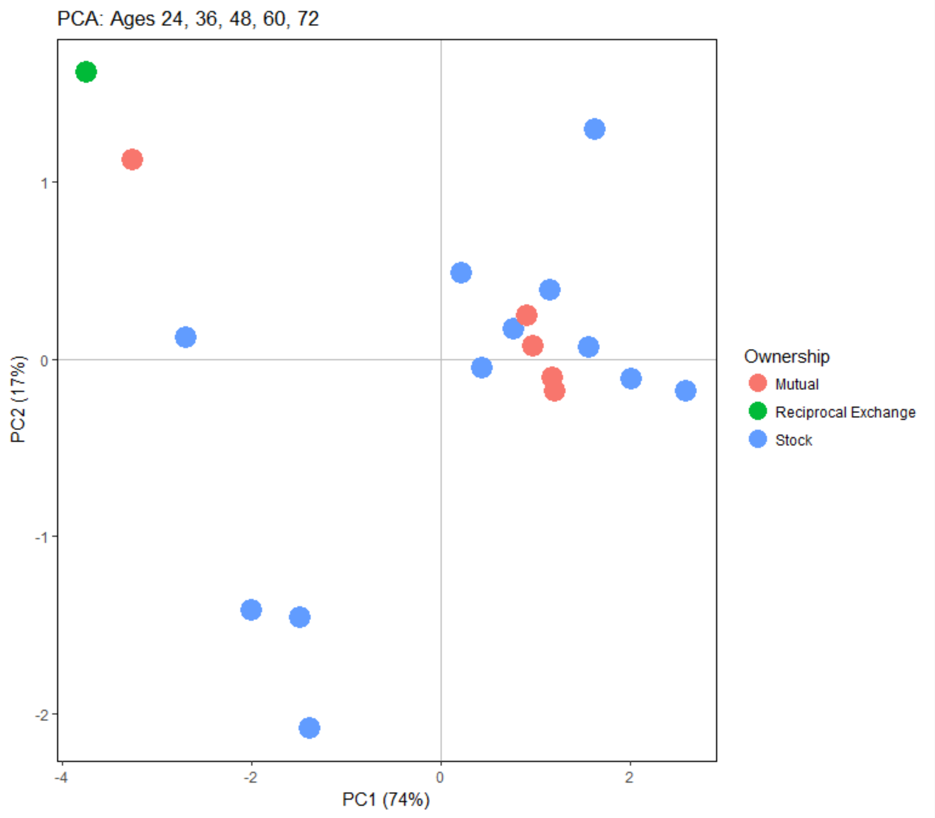


PCA
Visualization

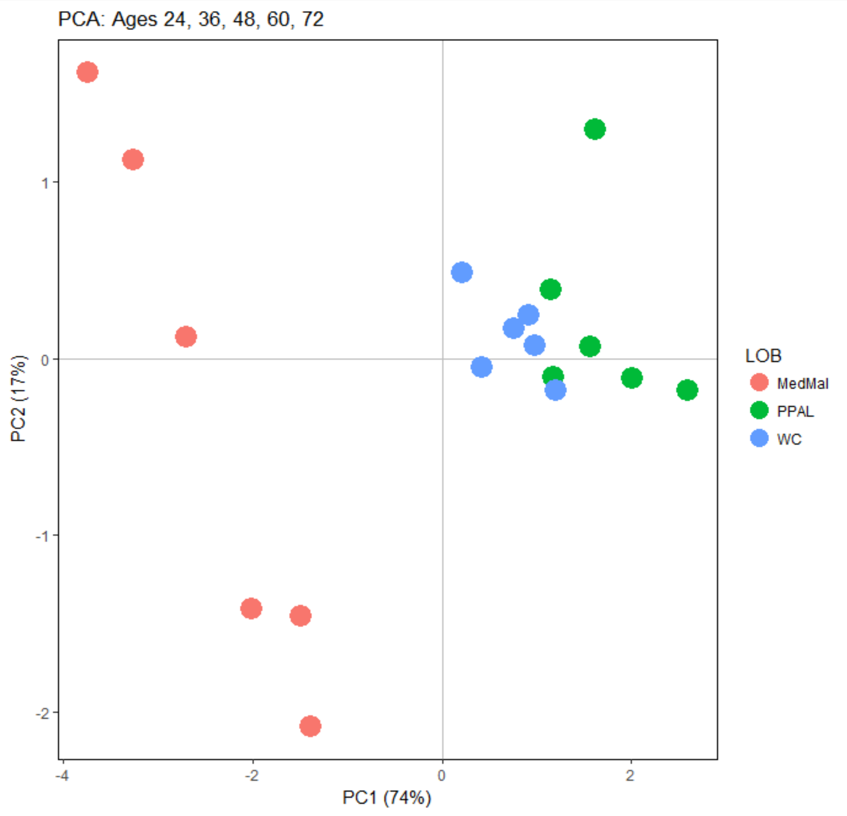


PCA

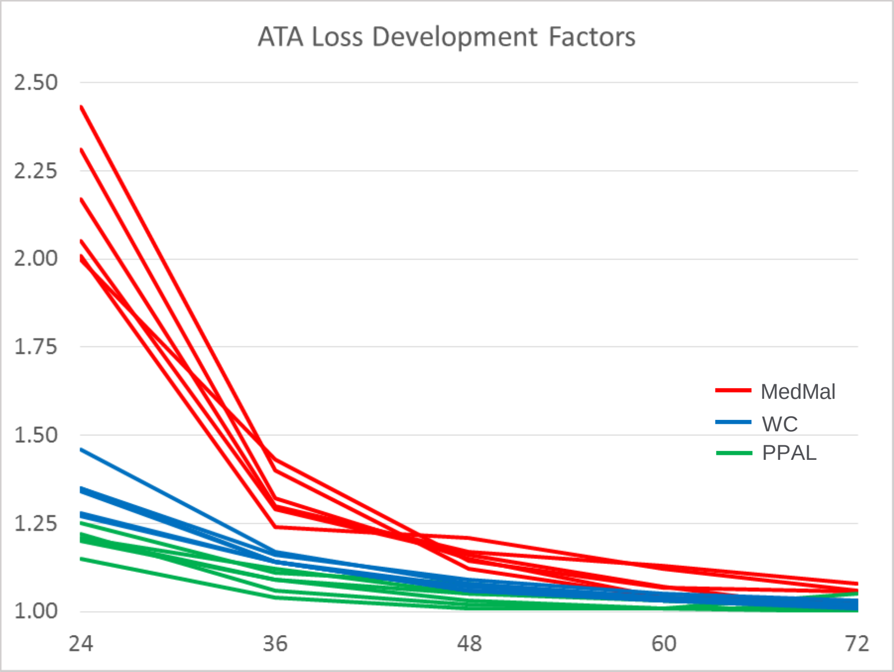
Visualization - Ownership

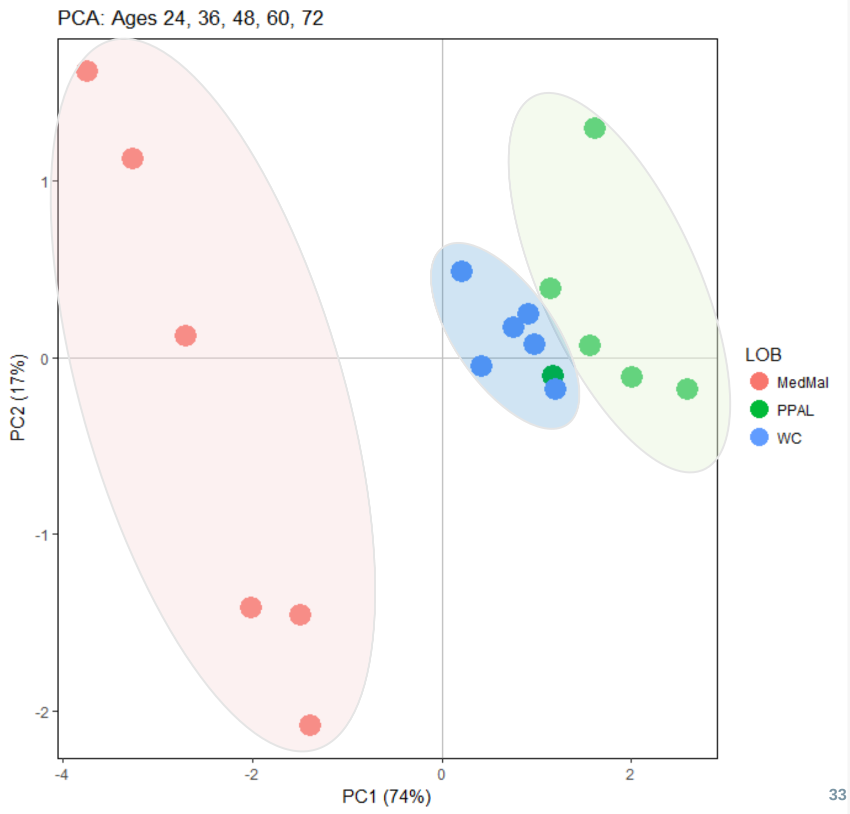


PCA
Visualization - LOB



PCA
Schedule P example: PCA





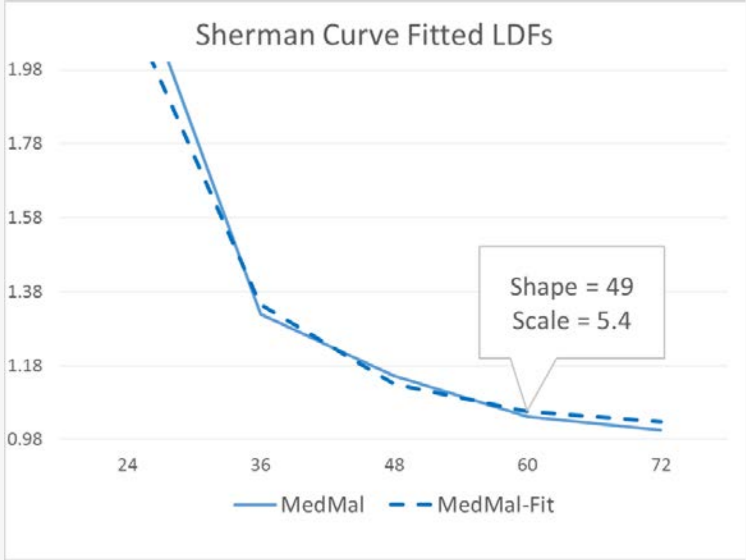
Data Transformation
How to Find Clusters?

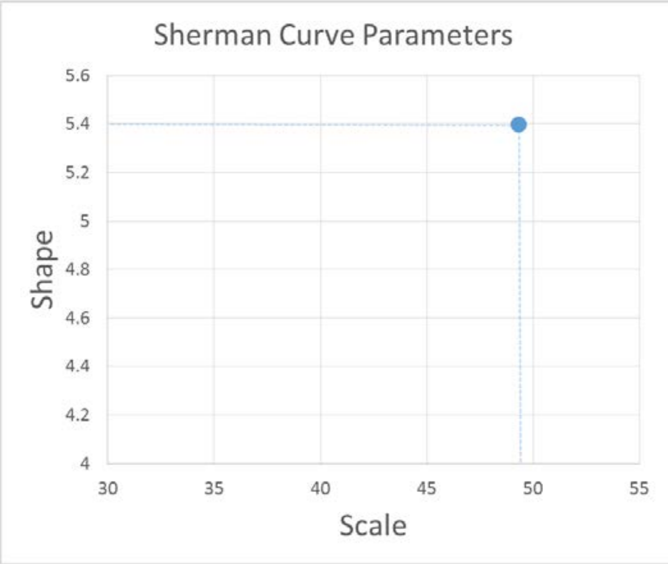
- Exploratory Data Analysis
  - Cluster Analysis
  - Principal Component Analysis (PCA)
  - Data transformation (curve fitting)

Data Transformation
Sherman Curve

- Sherman proposed a curve that fits to the typical LDF pattern

$$ATA_t = 1 + \left(\frac{Scale}{t + c}\right)^{Shape}$$



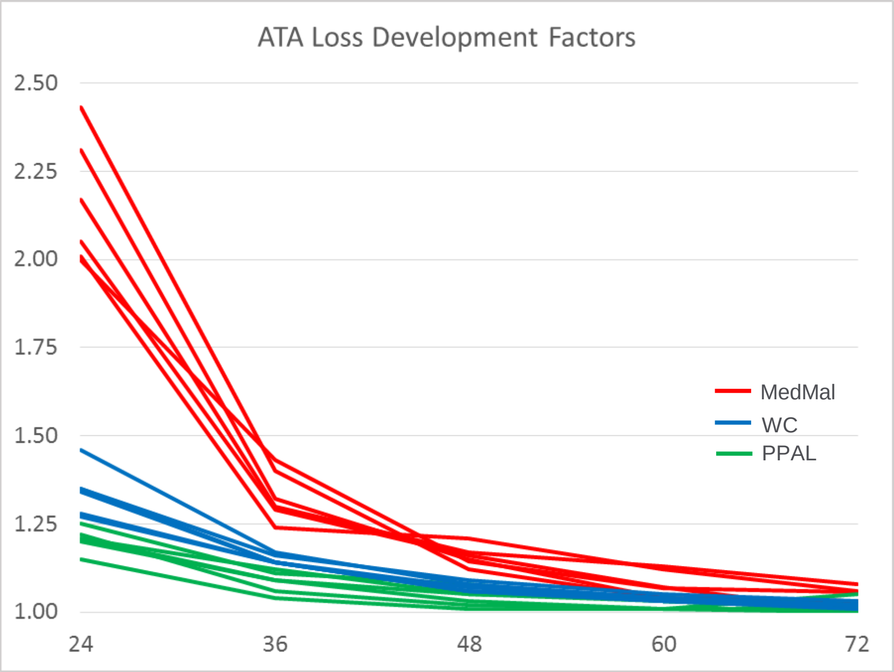


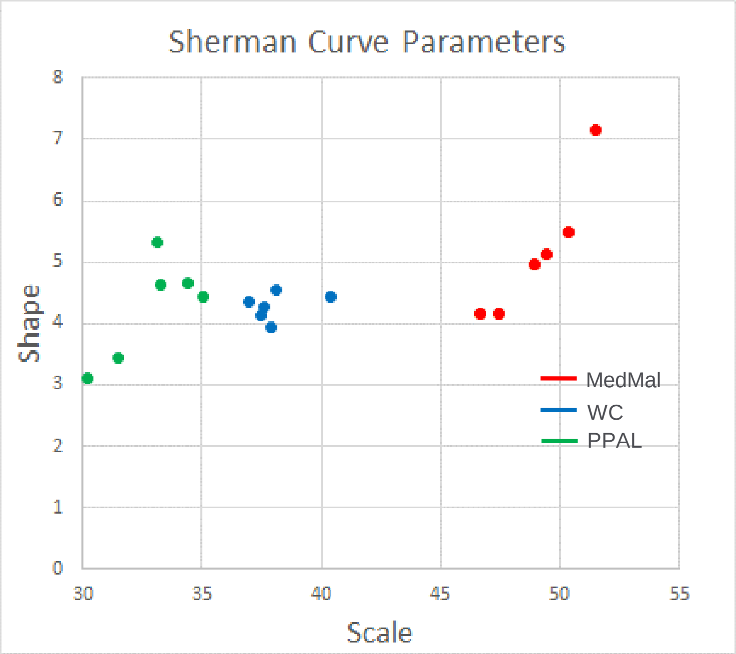
Data Transformation

How to estimate the parameters?

- Sherman recommends estimating the parameters by using log-linear regression
  - All actual age-to-age factors must be strictly greater than 1
  - Fitting a logged value rather than actual amounts
- GLM to the rescue!
  - Apply GLM with log-link on actual data

Data Transformation
Schedule P example: Sherman curve





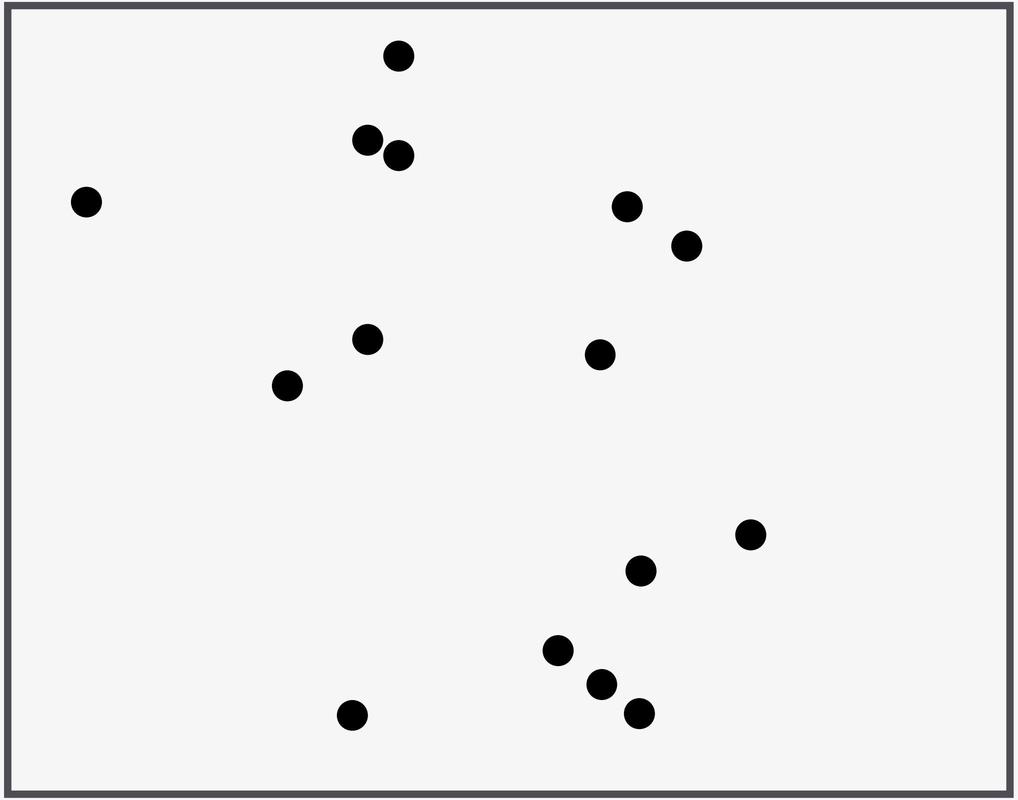
Data Transformation

Pros & Cons

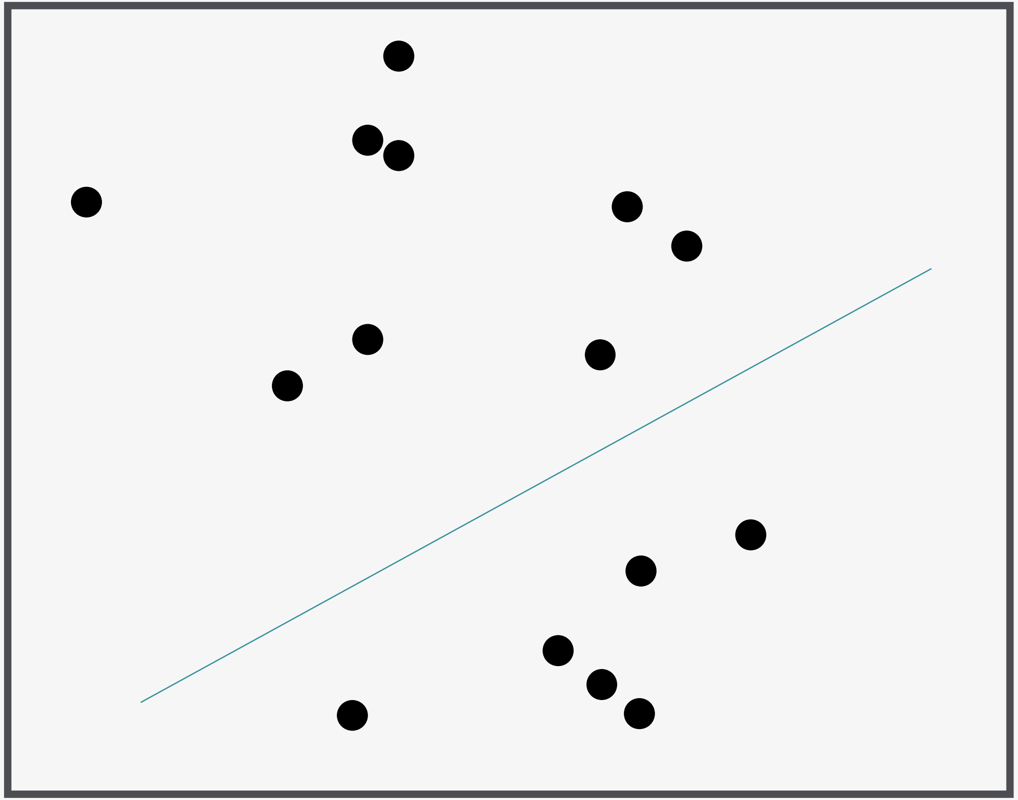
- Allows comparison of loss development patterns of different sizes
- Does not work well for flat curves
- The focus is on the fit and not on maintaining the distances between points

Practical Considerations

How Many Clusters Do You See?



Practical Considerations
How Many Clusters Do You See?



Practical Considerations
The Coins Experiment

Practical Considerations

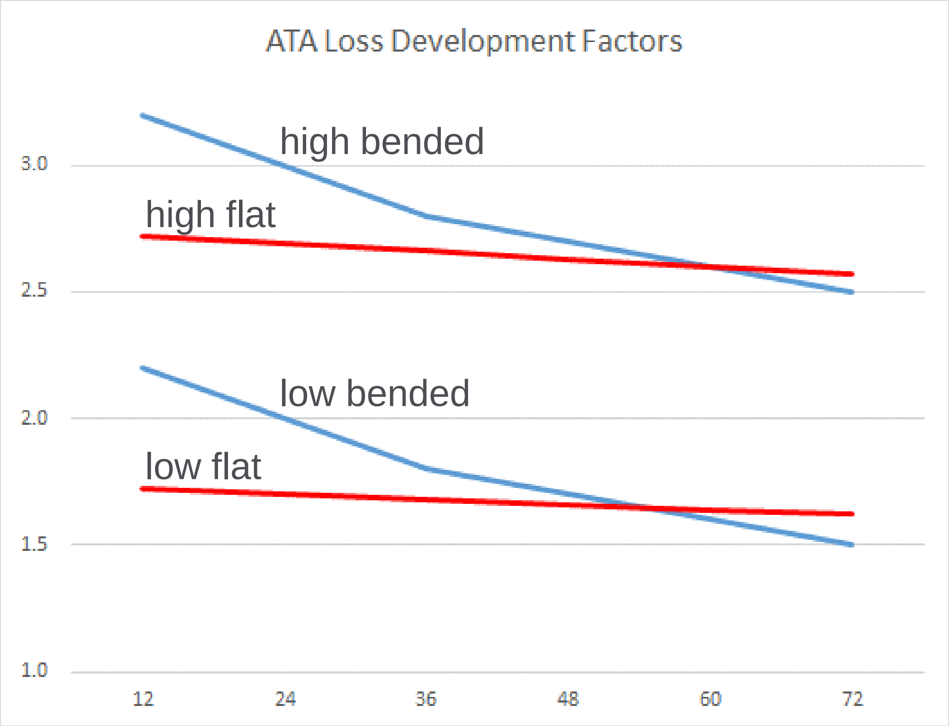
Clustering Illusion

Munich RE

"The predisposition to detect patterns and make connections is what leads to discovery and advance. The problem, however, is that this tendency is so strong and so automatic that we sometimes detect patterns when they do not exist."

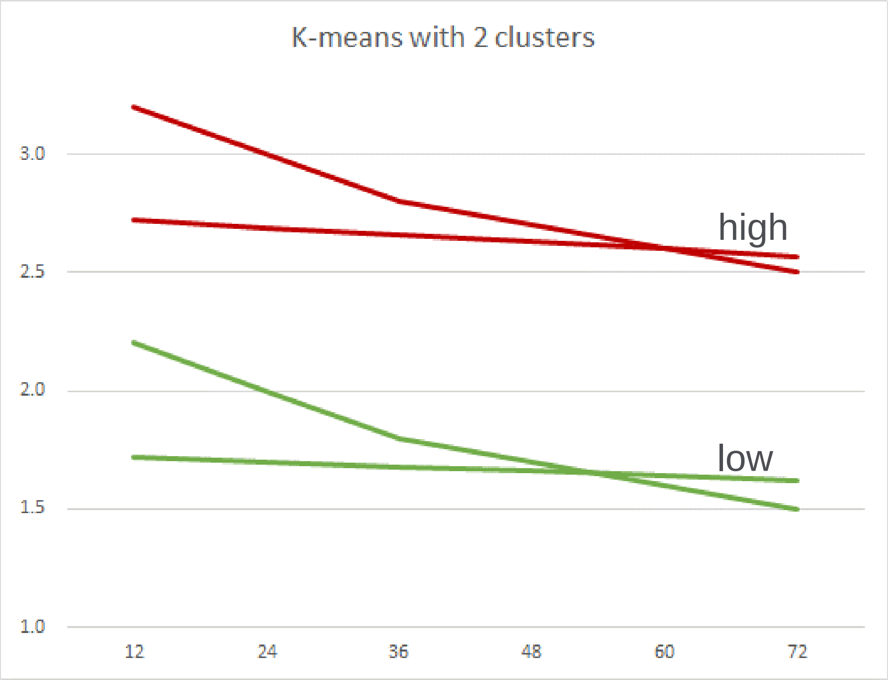
T. Gilovich, "How We Know What Isn't So - The Fallibility of Human Reason in Everyday Life"

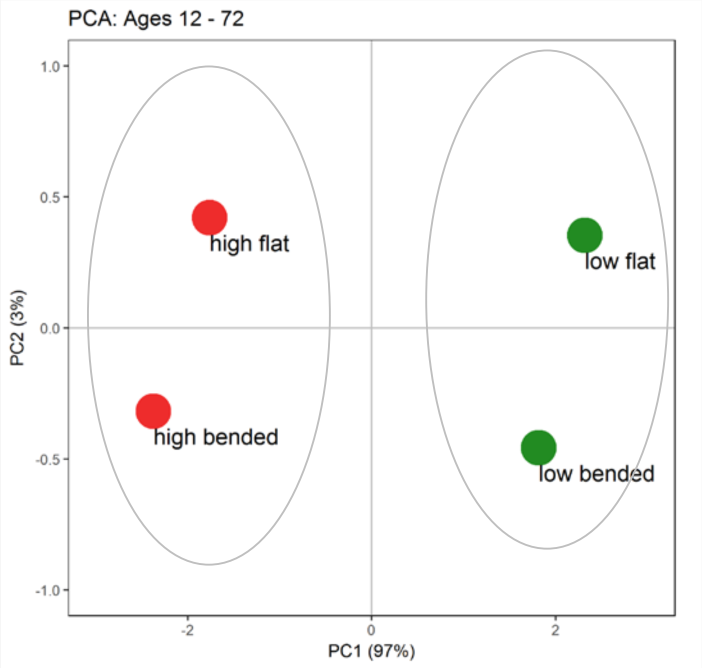
Practical Considerations
Finding the Right Question



Practical Considerations
Magnitude Clusters

- K-means and PCA on original data identify “magnitude clusters”





Practical Considerations
Variable Transformation

Munich RE

- Data enrichment
  - Cluster on parameters from PCA and Sherman fit
  - Include new variables
- Emphasizing similarities of interest

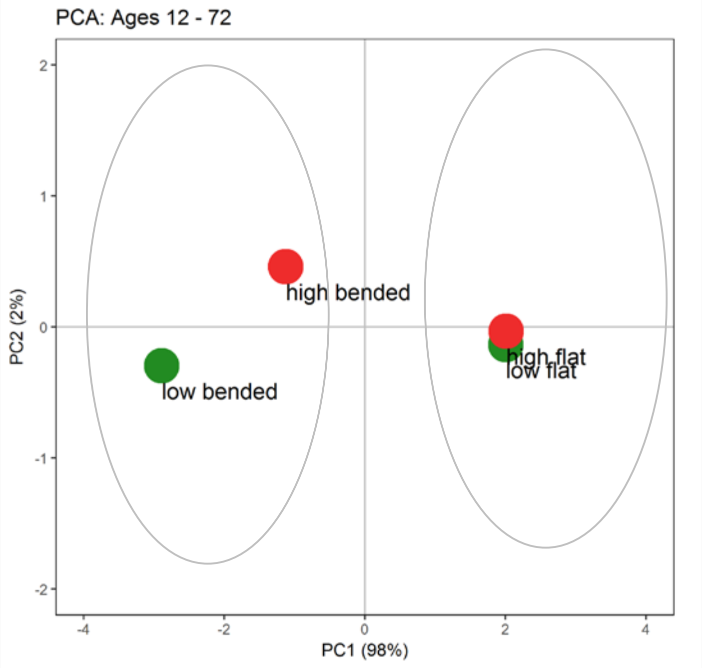
Level	Shape	12	24	36	48	60	72	Mean
low	bended	2.2	2.0	1.8	1.7	1.6	1.5	1.8
high	bended	3.2	3.0	2.8	2.7	2.6	2.5	2.8
low	flat	1.7	1.7	1.7	1.7	1.6	1.6	1.7
high	flat	2.7	2.7	2.7	2.7	2.6	2.6	2.6

New Variables = ATA / Mean

Level	Shape	12	24	36	48	60	72
low	bended	1.2	1.1	1.0	0.9	0.9	0.8
high	bended	1.1	1.1	1.0	1.0	0.9	0.9
low	flat	1.0	1.0	1.0	1.0	1.0	1.0
high	flat	1.0	1.0	1.0	1.0	1.0	1.0

Practical Considerations
Shape Clusters

- K-means and PCA on transformed variables identify “shape clusters”



Practical Considerations

Correlations between lines of business

- Compare the first principal component for two different lines, written by the same company

- Schedule P data for loss reserving posted on the CAS website
  - 54 companies with CAL and GL lines
  - 20 companies with WC and GL lines
  - Data is from 1988 to 1997

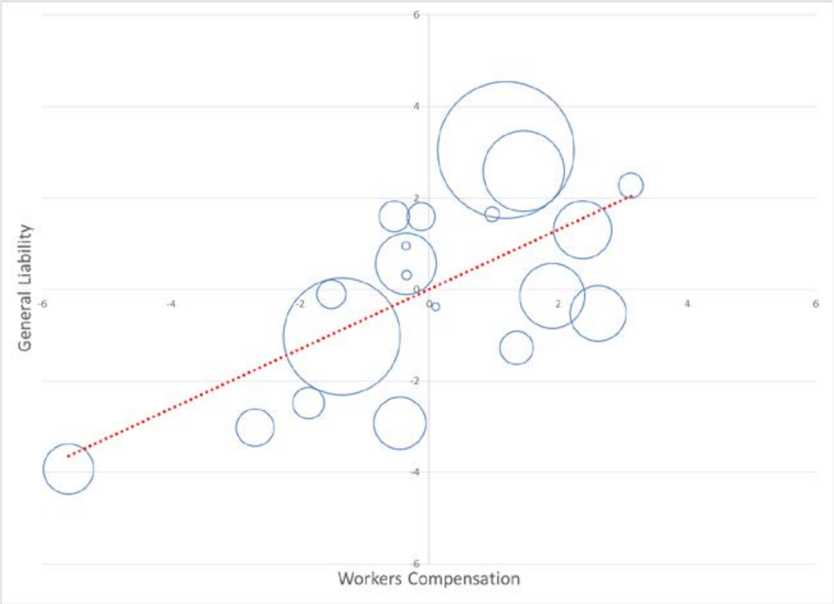
- Check if historical dependency is preserved in more recent years

Practical Considerations
First principal component for WC/GL

➢ PCA on Reported loss

1988 - 1997

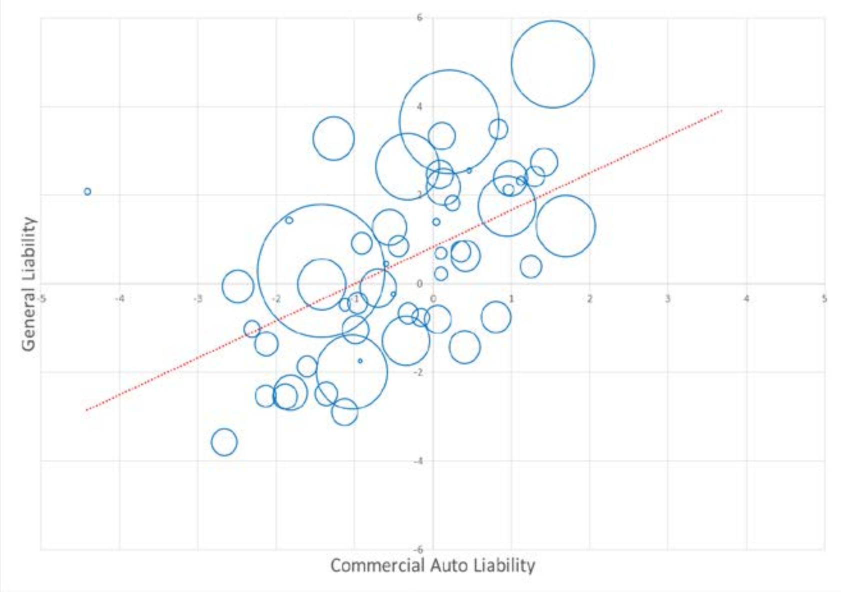
1998 - 2007



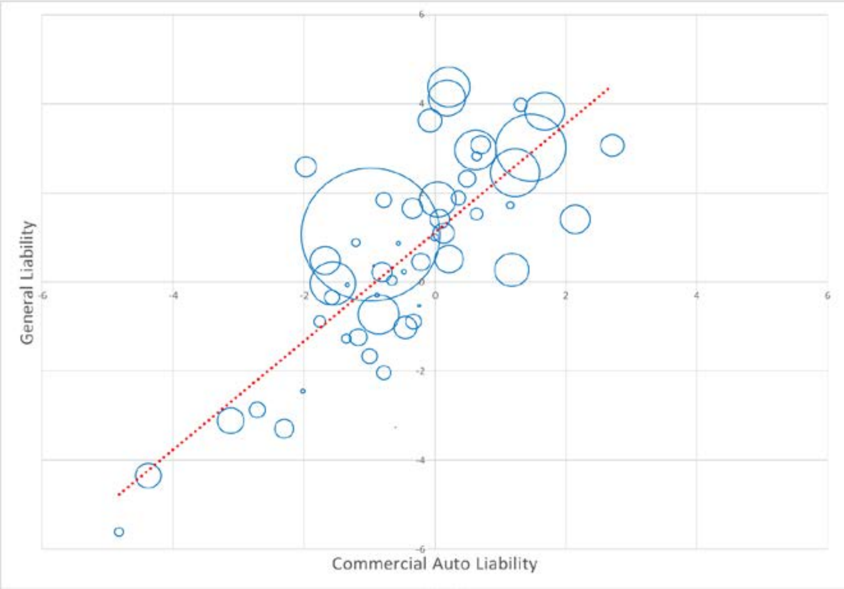
Practical Considerations
First principal component for CAL/GL

- PCA on Reported loss

1988 - 1997



1998 - 2007



Conclusion
Next Steps

- Investigate factors causing correlation between classes of business by company
  - Stock vs. Mutual
  - Regional vs. National
- Investigate connection between Fuzzy Clustering and “Mixed Models” (what actuaries know as credibility theory)

Conclusion
Soft Clustering

Munich RE

- Soft (a.k.a. fuzzy) clustering allows each data point to belong to more than one cluster
- Membership grades are assigned to each data point

LOB	Fuzzy 1 (MedMal)	Fuzzy 2 (PPAL)	Fuzzy 3 (WC)
MedMal	45%	27%	28%
MedMal	54%	22%	24%
MedMal	66%	17%	18%
MedMal	46%	26%	28%
MedMal	65%	17%	18%
MedMal	66%	17%	18%
PPAL	6%	57%	38%
PPAL	12%	51%	37%
PPAL	16%	44%	40%
PPAL	8%	55%	37%
PPAL	5%	45%	49%
PPAL	6%	49%	44%
WC	5%	51%	44%
WC	5%	41%	54%
WC	9%	36%	56%
WC	5%	34%	61%
WC	5%	37%	58%
WC	13%	36%	51%

Conclusion
Key Takeaways

Munich RE

- Clustering techniques help us obtain a better estimate of reserves:
  - Explore the structure of data
  - Go beyond “just” practical grouping of data
  - Identify variables impacting the development

- Each method has strengths and weaknesses
  - Look for robustness between methods

Selected References

1. K. Boudt, J. Cornelissen, C. Croux (2012) **The Gaussian rank correlation estimator: robustness properties**, *Statistics and Computing, March 2012, Volume 22, Issue 2, pp 471–483,* http://feb.kuleuven.be/public/ndbae06/pdf-files/grcort.pdf

2. D. Clark (2017) **Estimation of Inverse Power Parameters via GLM**, *Actuarial Review, May-June 2017 ,* http://ar.casact.org/actuarialreview/may_june_2017/MobilePagedArticle.action?articleId=1112695#articleId1112695

3. T. Hastie, R. Tibshirani, J. Friedman (2009) **The Elements of Statistical Learning - Data Mining, Inference, and Prediction**, Springer https://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf

4. C. Hennig (2015) **Clustering strategy and method selection,** In C. Hennig, M. Meila, F. Murtagh, and R. Rocci (Eds.). *Handbook of Cluster Analysis*. Chapman and Hall/CRC, http://www.homepages.ucl.ac.uk/~ucakche/

5. P. Tan, M. Steinbach, V. Kumar (2005) **Cluster Analysis: Basic Concepts and Algorithms,** In P. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining,* Pearson Addison Wesley, http://www-users.cs.umn.edu/~kumar/dmbook/index.php

6. J. Shlens (2003) **A Tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition,** arXiv preprint arXiv:1404.1100, 2014, https://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf

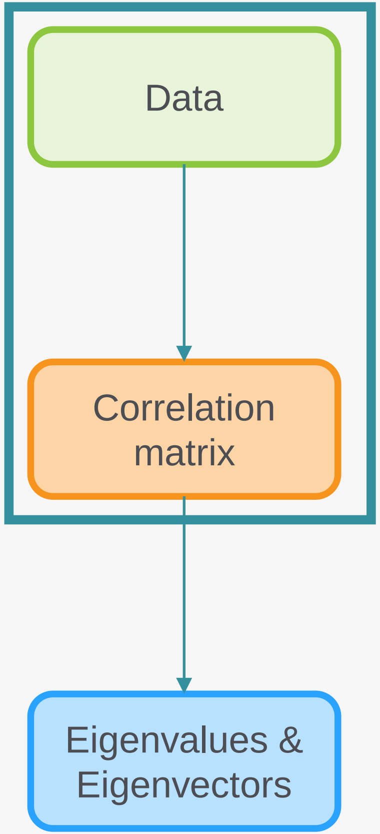
7. CAS Schedule P data for Loss Reserving: http://www.casact.org/research/index.cfm?fa=loss_reserves_data

Thank you!

Appendix
PCA-Gaussian Rank Correlation (GRC)

- PCA uses a covariance-variance matrix of the data
- GRC uses an alternative matrix based on ranked scores of the data
- GRC is more robust in the presence of outliers

Appendix
How GRC works?



1 Data – age to age factors

ATA 1	ATA 2	ATA 3
5.99	2.31	1.32
5.13	2.24	1.68
1.92	1.2	1.09
1.64	1.15	1.04
2.19	1.34	1.14
2.33	1.28	1.14
2.25	1.35	1.16

2 Rank each observation

ATA 1	ATA 2	ATA 3
1	1	2
2	2	1
6	6	6
7	7	7
5	4	4
3	5	4
4	3	3

3 Rescale / compute percentiles

ATA 1	ATA 2	ATA 3
93%	93%	79%
79%	79%	93%
21%	21%	21%
7%	7%	7%
36%	50%	50%
64%	36%	50%
50%	64%	64%

4 Calculate Guassian rank scores

ATA 1	ATA 2	ATA 3
1.47	1.47	0.79
0.79	0.79	1.47
-0.79	-0.79	-0.79
-1.47	-1.47	-1.47
-0.37	0.00	0.00
0.37	-0.37	0.00
0.00	0.37	0.37

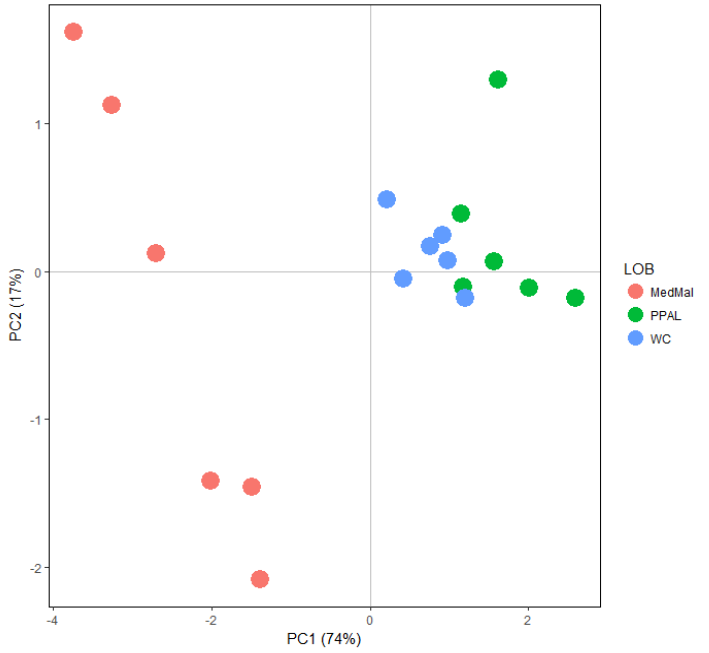
Correlation matrix based on Gaussian rank

	ATA 1	ATA 2	ATA 3
ATA 1			
ATA 2		CORREL(Var1, Var2, …)	
ATA 3			

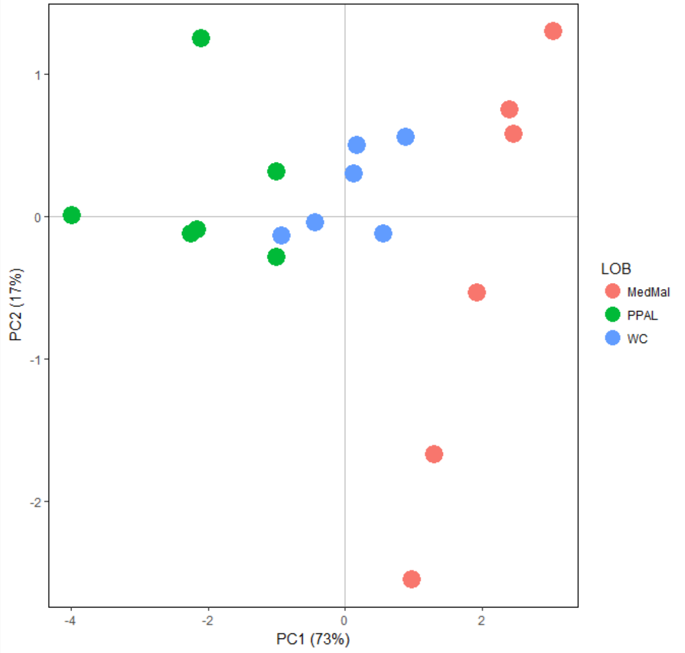
Appendix
PCA-Gaussian Rank Correlation Example

- Schedule P example

Traditional PCA



PCA – Gaussian Rank Correlation



Appendix
Determining the Number of Cluster