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 Development

Dave Clark
Diana Rangelova

Ratemaking, Product and Modeling (RPM) Seminar - March 2018

Munich RE

Agenda

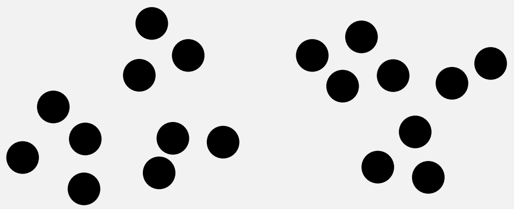
Munich RE

1. Introduction
2. How to find clusters:
   a) Cluster analysis
   b) Principal Component Analysis (PCA)
   c) Data transformation (curve fitting)
3. Practical considerations and observations
4. Soft clustering and credibility

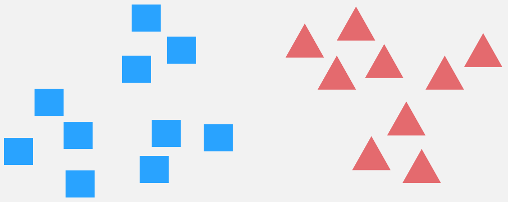
Introduction
Clustering

Munich RE

- 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



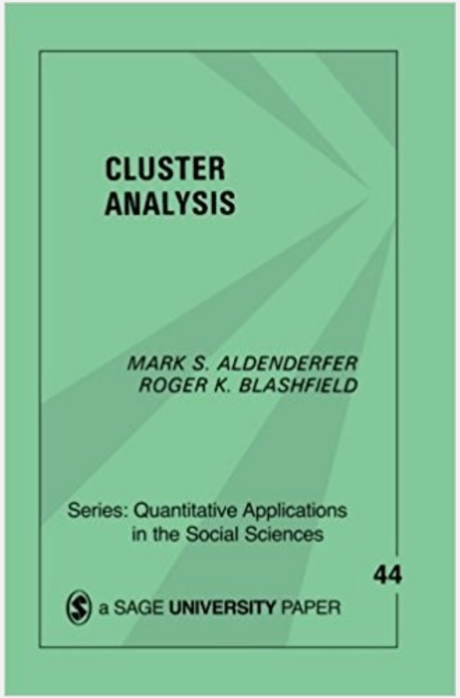
Original data



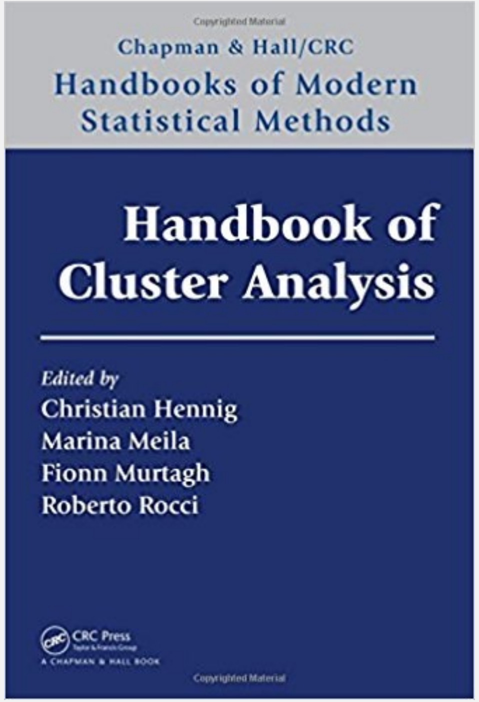
Two clusters

Introduction
Why Clustering?

Cluster Analysis has grown rapidly, especially as computer software has become more readily available.



1984 - 88 pages



2015 - 773 pages

Introduction
Why Clustering?

- What 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
  - State (for classification ratemaking)
  - Loss development patterns, loss ratios, severity, frequency…

Introduction

How to Find Clusters?

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

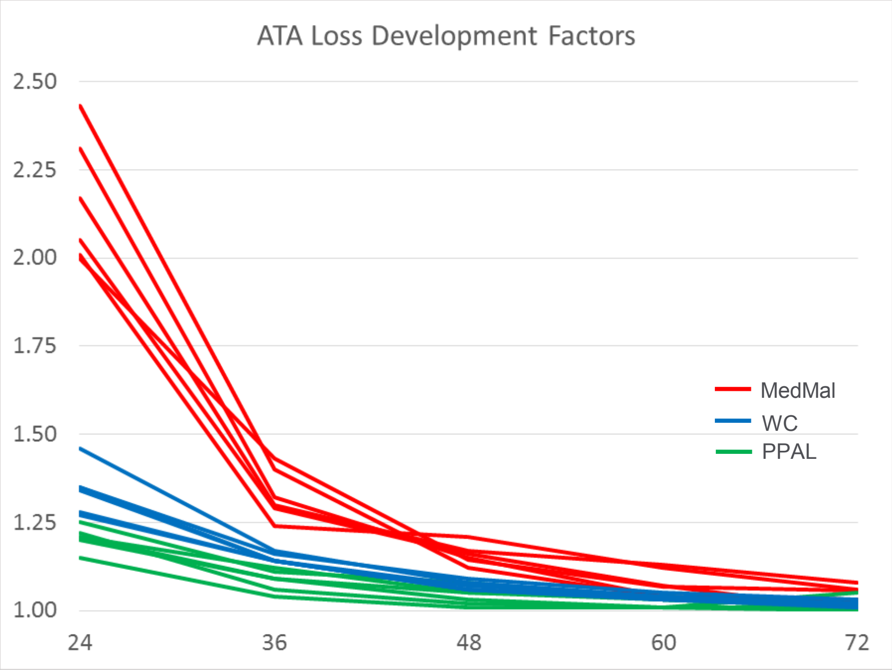
Introduction
Schedule P Example

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)

Cluster Analysis
Types of Clustering

- 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.

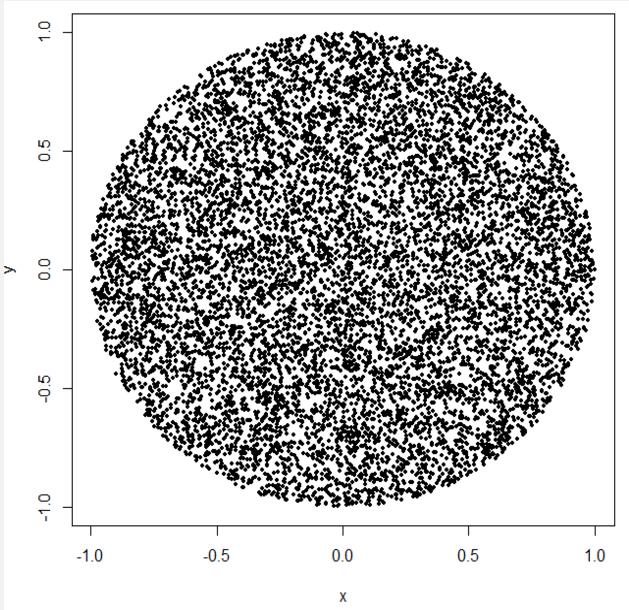
K-means: Pros & Cons

Munich RE

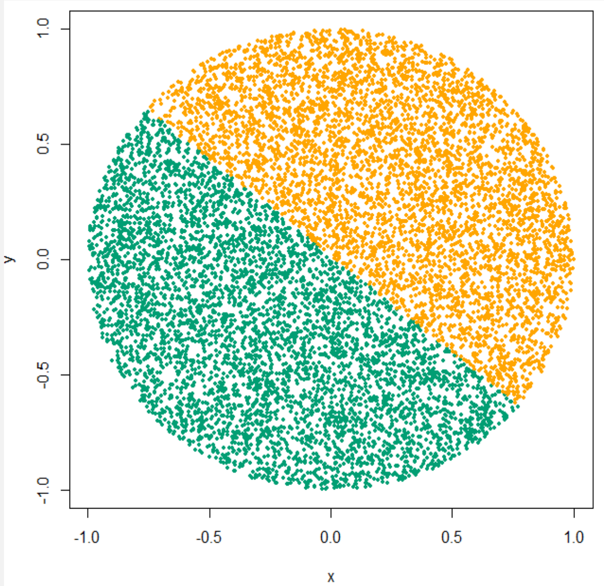
- ➢ K-means is simple, fast and efficient
- ➢ K-means does not perform well when:
  - There are no natural clusters
  - Clusters are of different size
  - Clusters are not spherical
  - Outliers exist

Cluster Analysis
No natural clusters

Original data

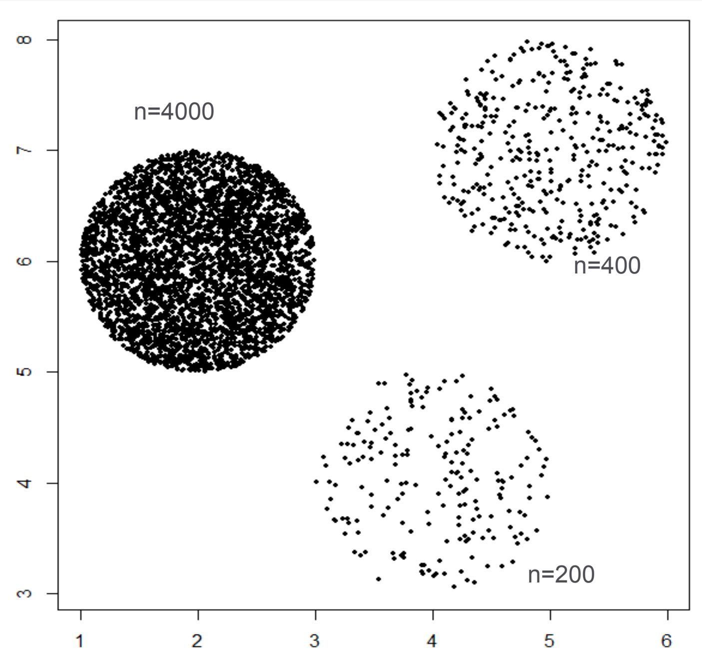


K-means result

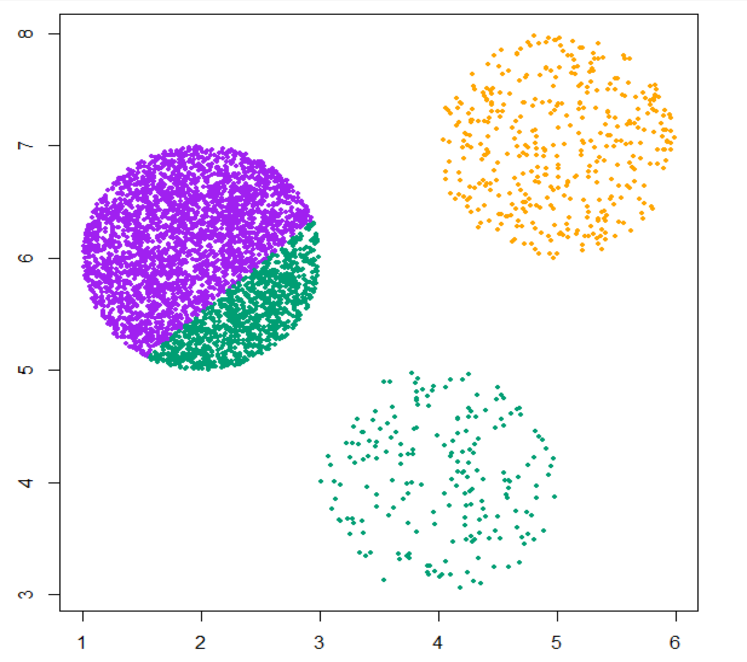


Cluster Analysis
Clusters of different size

Original data

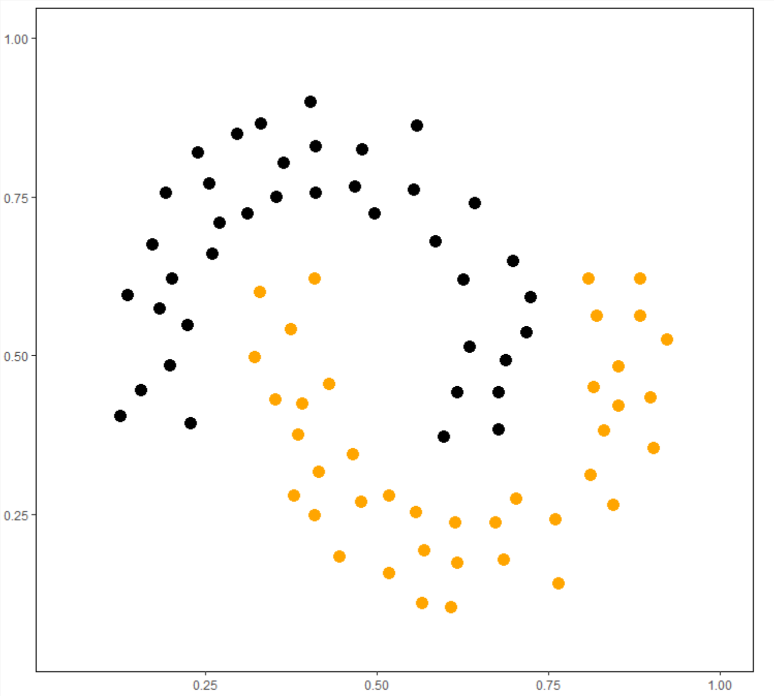


K-means result

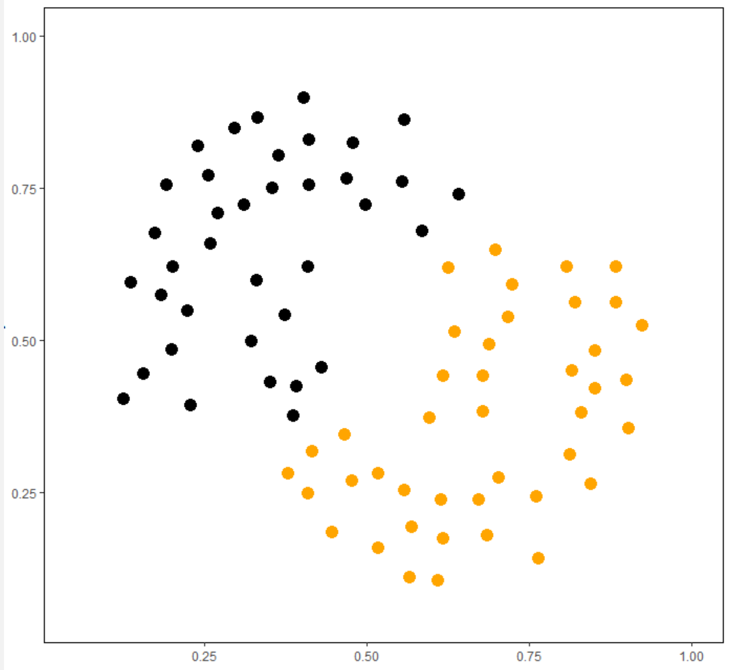


Cluster Analysis
Non spherical clusters

Original data



K-means result



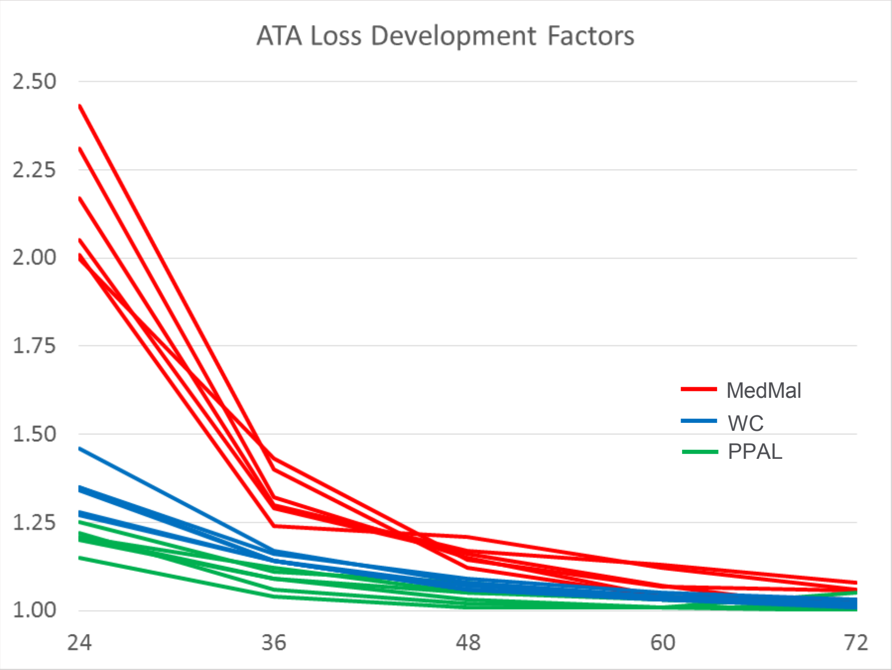
Source: Introduction to Data Mining [5]

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

Schedule P example: Cluster Analysis

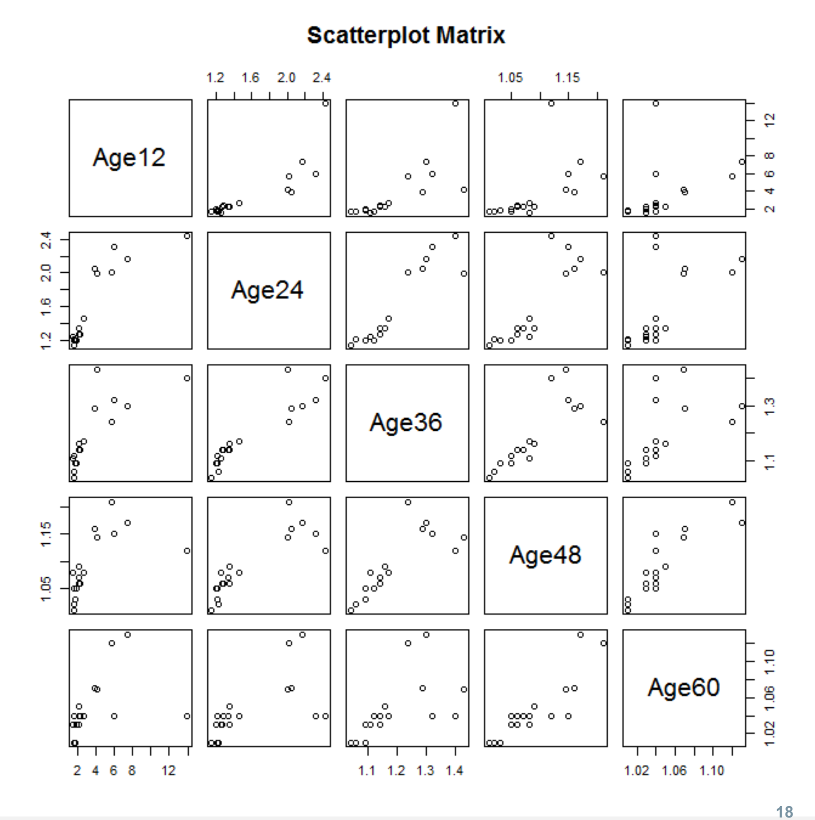


LOB	K-means 2 clusters	K-means 3 clusters	K-medoids 3 clusters
MedMal	1	1	1
MedMal	1	1	1
MedMal	1	2	1
MedMal	1	1	1
MedMal	1	2	1
MedMal	1	2	1
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
PPAL	2	3	2
WC	2	3	3
WC	2	3	3
WC	2	3	3
WC	2	3	3
WC	2	3	3
WC	2	3	3

Cluster Analysis
Too Many Dimensions

- Difficulty visualizing more than two dimensions for validation purposes

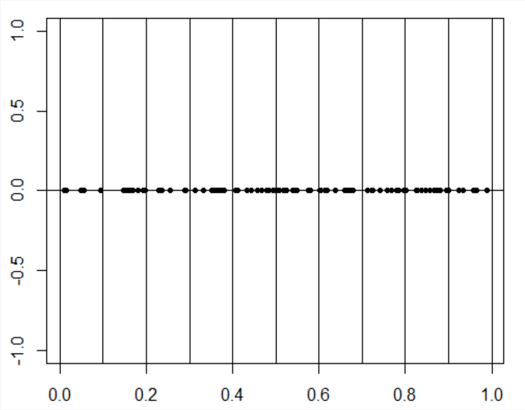
12	24	36	48	60	72
5.70	2.01	1.24	1.21	1.12	1.06
3.86	2.05	1.29	1.16	1.07	1.00
1.92	1.20	1.09	1.05	1.03	1.01
1.64	1.15	1.04	1.01	1.01	1.00
2.19	1.34	1.14	1.07	1.04	1.02
2.33	1.28	1.14	1.06	1.04	1.02
		…			

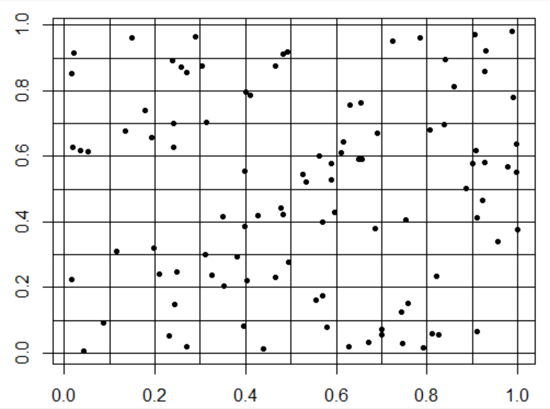


Cluster Analysis
Too Many Dimensions

- Data gets “lost in space”

Randomly generated 100 points in 1D and 2D





- *“In high dimension spaces, distances between points become relatively uniform.”* The performance of clustering algorithms relying on L_1 (sum of absolute values) or L_2 (Euclidian) metrics in high dimensional data may be compromised.

Source: M. Steinbach, L. Ertoz, V. Kumar, “The Challenges of Clustering High Dimensional Data” [7]

PCA
How to Find Clusters?

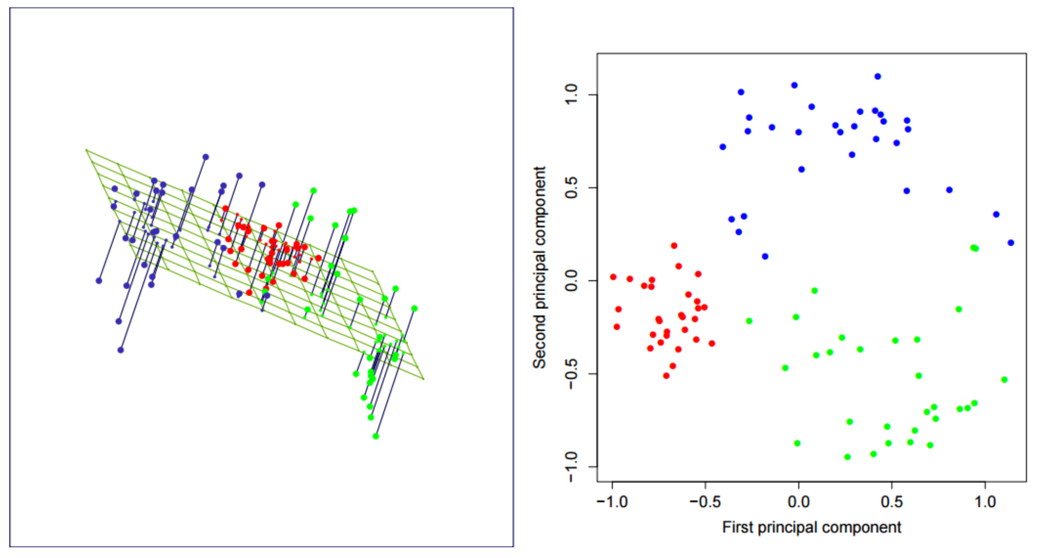
Munich RE

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

PCA

Principal Component Analysis

- **PCA stretches and rotates data** with the goal to derive the best possible k-dimensional representation of the Euclidean distance among objects.



Source: The Elements of Statistical Learning

PCA
How to perform a PCA?

Data (scaled and centered) → Correlation matrix → Eigenvalues & Eigenvectors

Data (scaled and centered) x Eigenvectors = Principal Components

PCA

Interpretation

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

Data (scaled and centered) x Eigenvectors = Principal Components

Co	24	36	48	60	72
1	0.99	0.46	2.07	2.18	1.43
2	1.08	0.89	1.17	0.66	-1.04
3	-0.83	-0.82	-0.75	-0.60	-0.78
4	-0.94	-1.30	-1.39	-1.14	-0.96
			…		

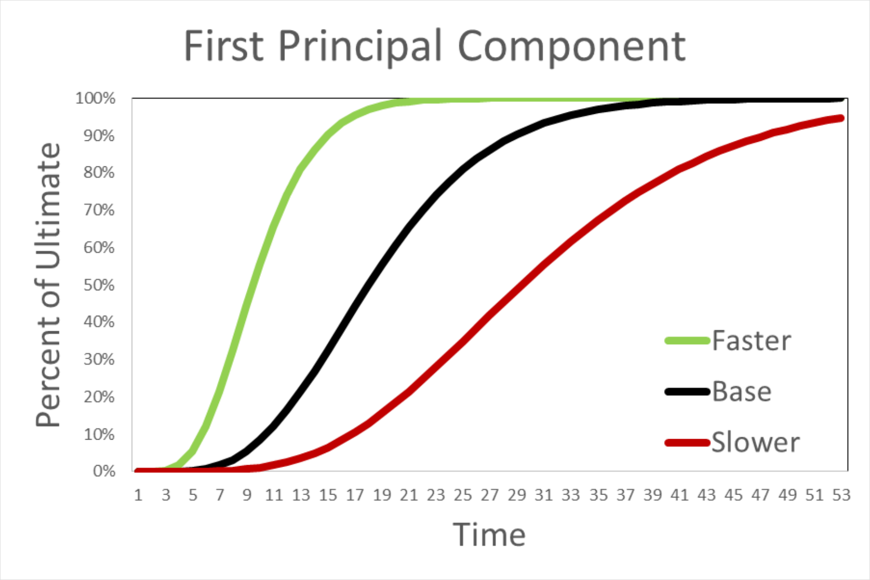
x

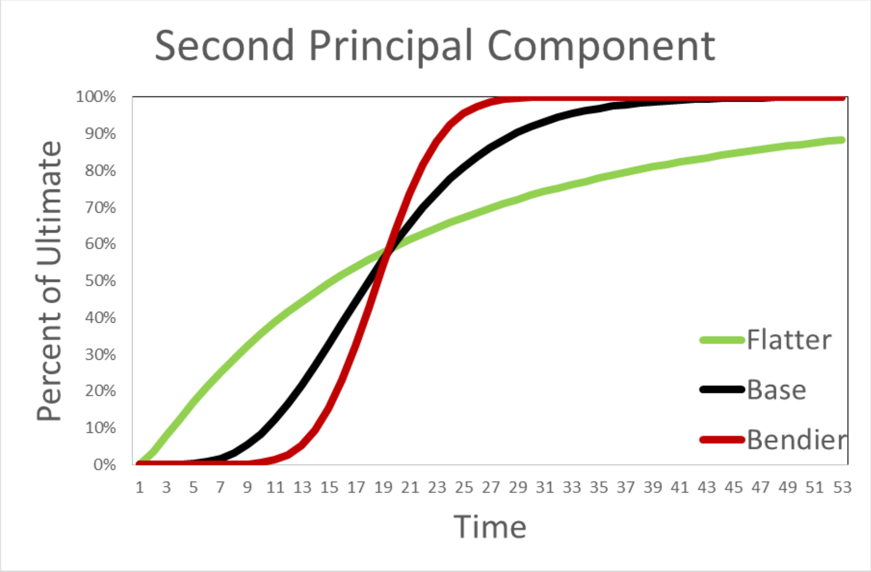
Dim	1	2
24	0.47	-0.39
36	0.46	-0.38
48	0.49	-0.12
60	0.46	0.36
72	0.34	0.75

=

Co	PC1	PC2
1	3.18	1.04
2	1.45	-1.44
3	-1.67	-0.07
4	-2.57	-0.10
	…	

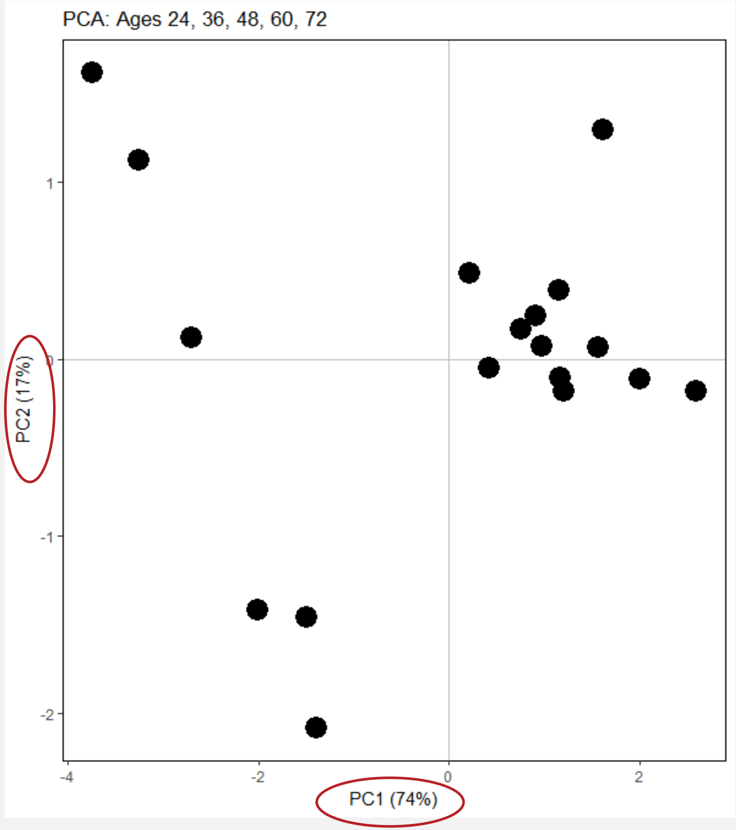
PCA
Interpretation





PCA

Schedule P example: Visualization



PCA

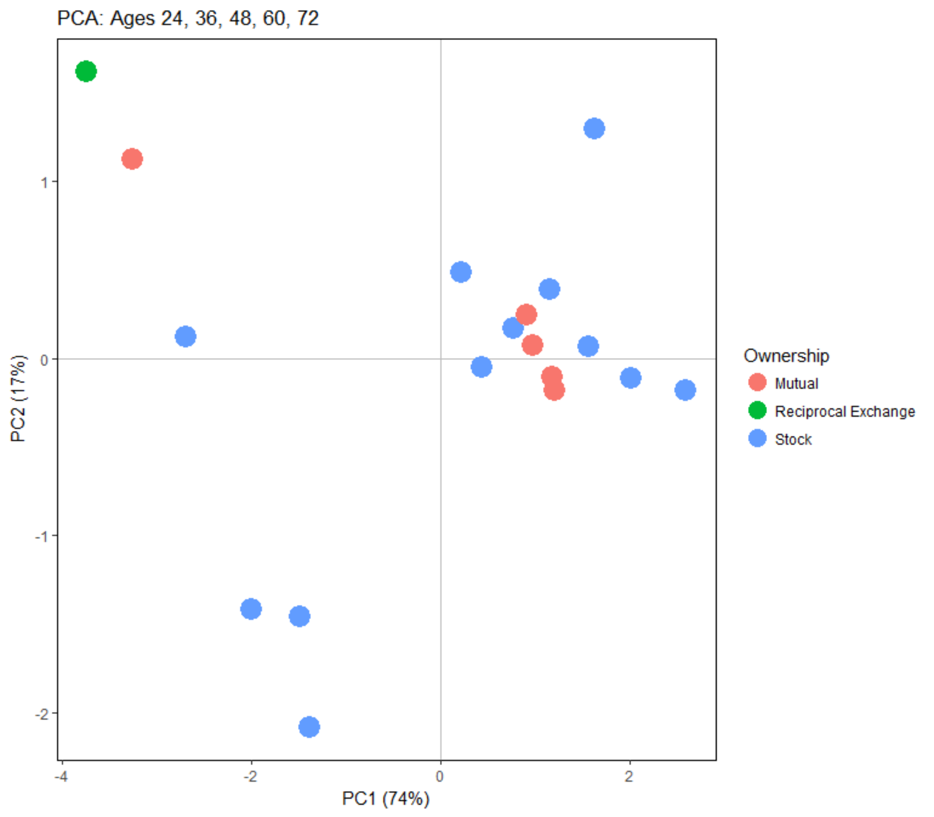
Explanatory Variables

Explanatory Variables

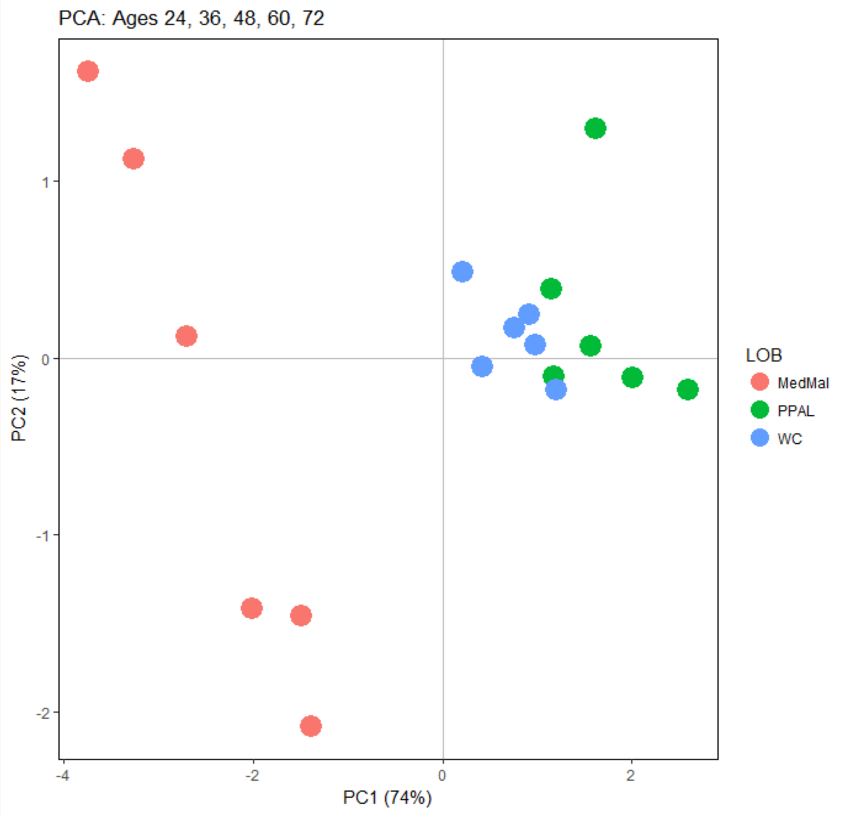
Co.	Line	Ownership	Geographic	Distribution
1	MedMal	Mutual	Regional	Direct, Ind Agency
2	MedMal	Stock	National	Direct, Ind Agency
3	PPAL	Stock	National	MGA, Ind Agency
4	PPAL	Stock	Regional	Ind Agency
5	WC	Stock	National	MGA
6	WC	Mutual	Regional	Ind Agency
			…	

PCA

Schedule P example: Visualization - Ownership



PCA
Schedule P example: Visualization - LOB



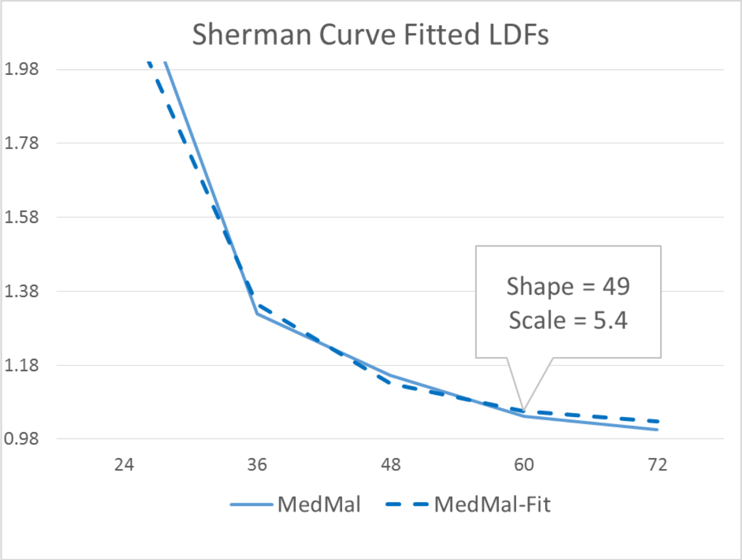
Data Transformation
How to Find Clusters?

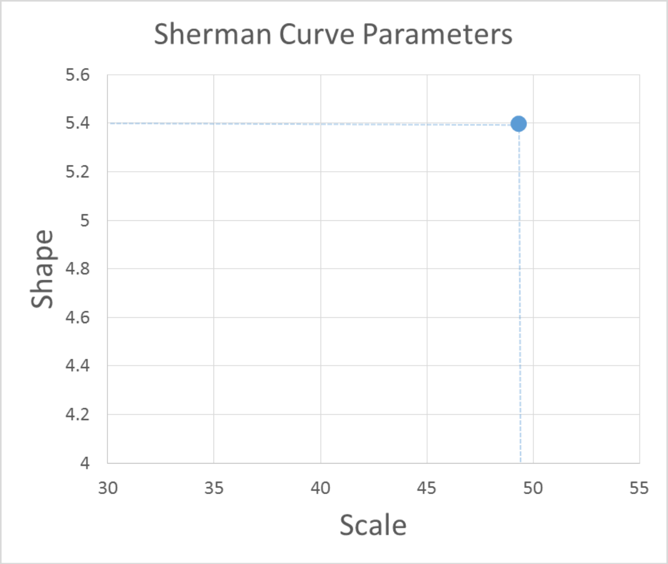
- Exploratory Data Analysis
  - Cluster analysis
  - Principal Component Analysis (PCA)
  - Data transformation (curve fitting)

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?

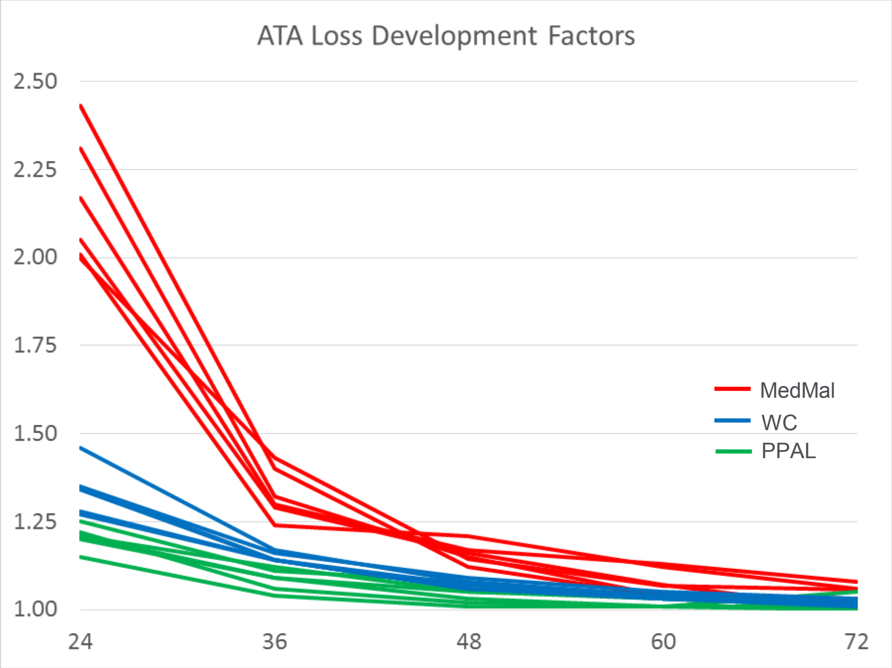
- Sherman recommends estimating the parameters by using log-linear regression
  - All actual age-to-age factors must be strictly greater than 1
  - Fitting a logged value rather than actual amounts
- GLM to the rescue!
  - Apply GLM with log-link on actual data

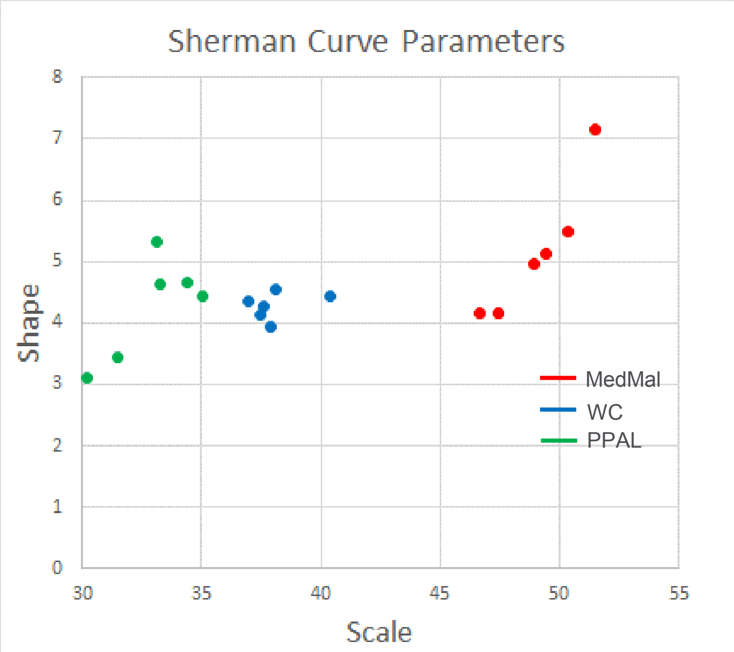
Data Transformation
Pros & Cons

Munich RE

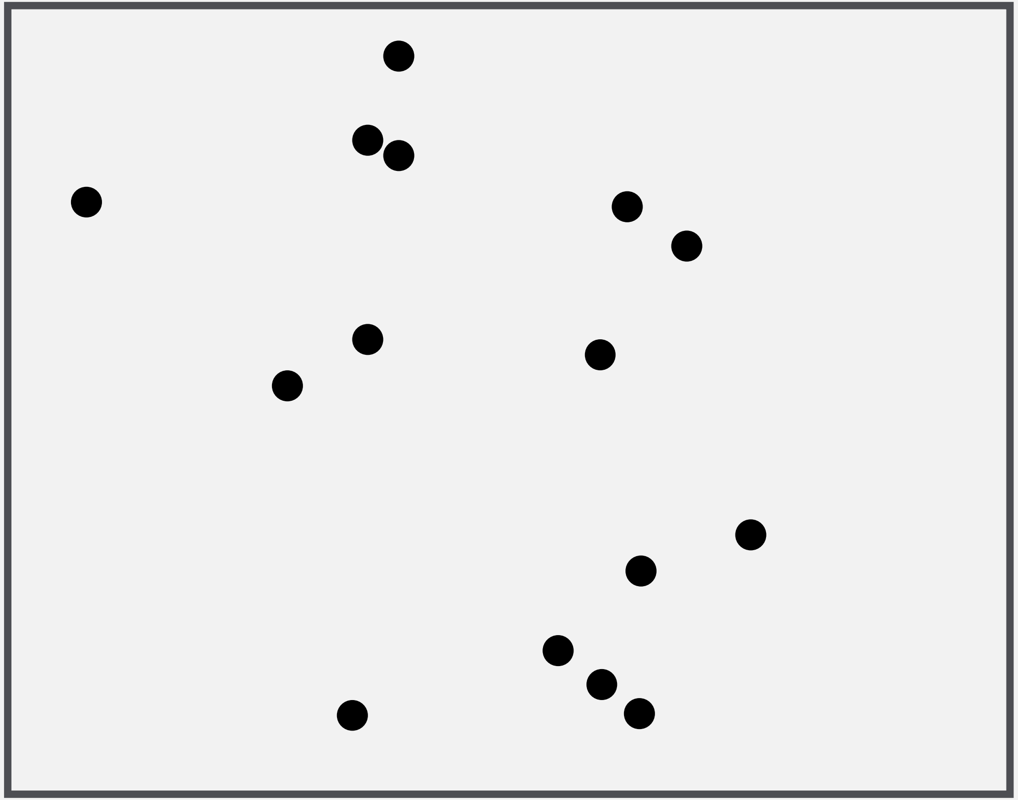
- Allows comparison of loss development patterns of different sizes
- Does not work well for flat curves
- The focus is on the fit and not on maintaining the distances between points

Data Transformation
Schedule P example: Sherman curve





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
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

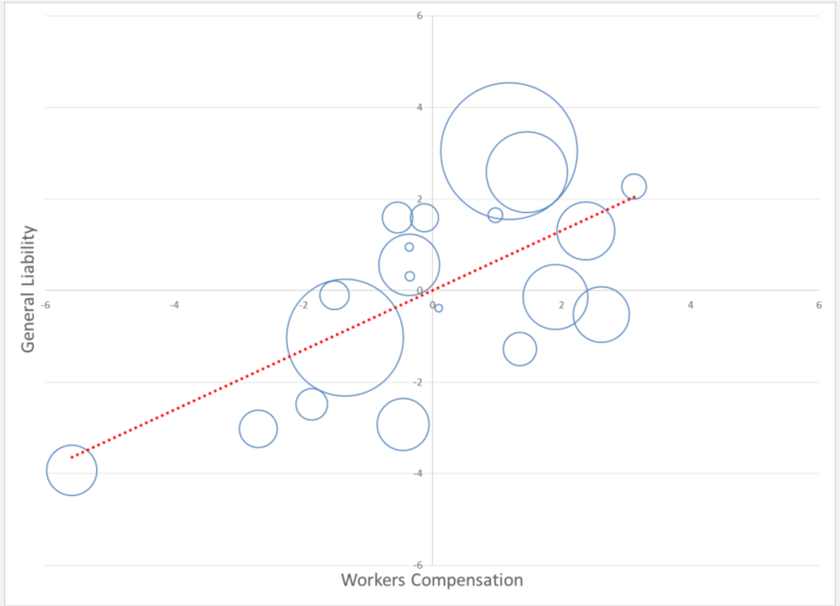
First principal component for WC/GL

Munich RE

- PCA on Reported loss

1988 - 1997

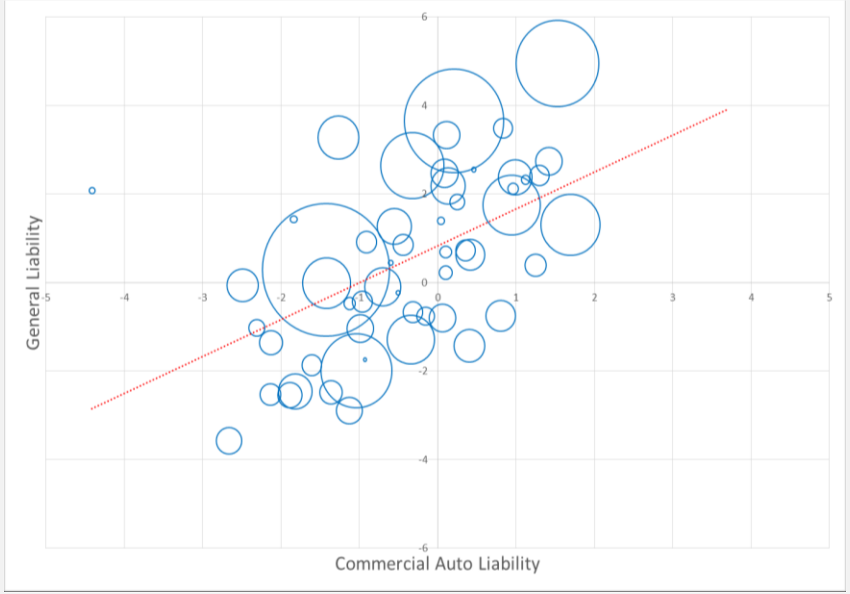
1998 - 2007



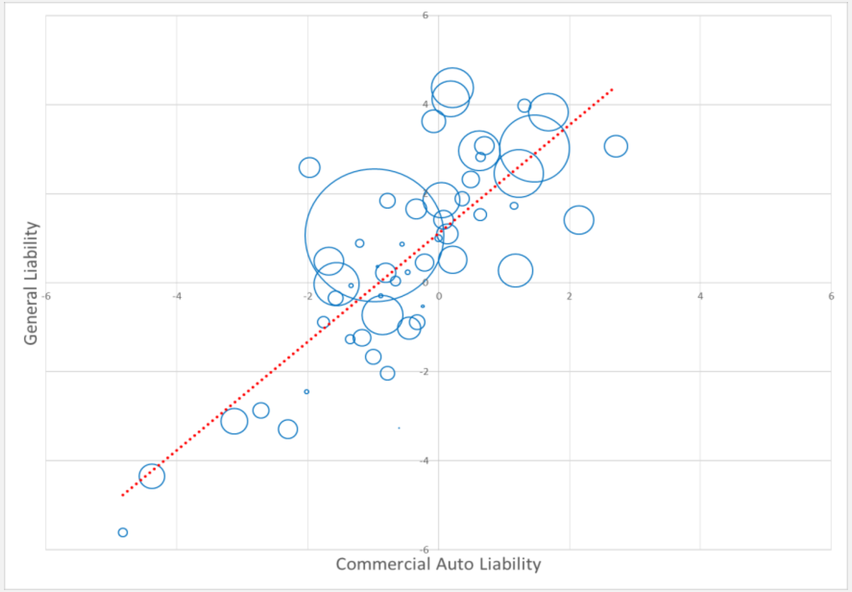
Practical Considerations
First principal component for CAL/GL

➤ PCA on Reported loss

1988 - 1997



1998 - 2007



Soft Clusters and Credibility

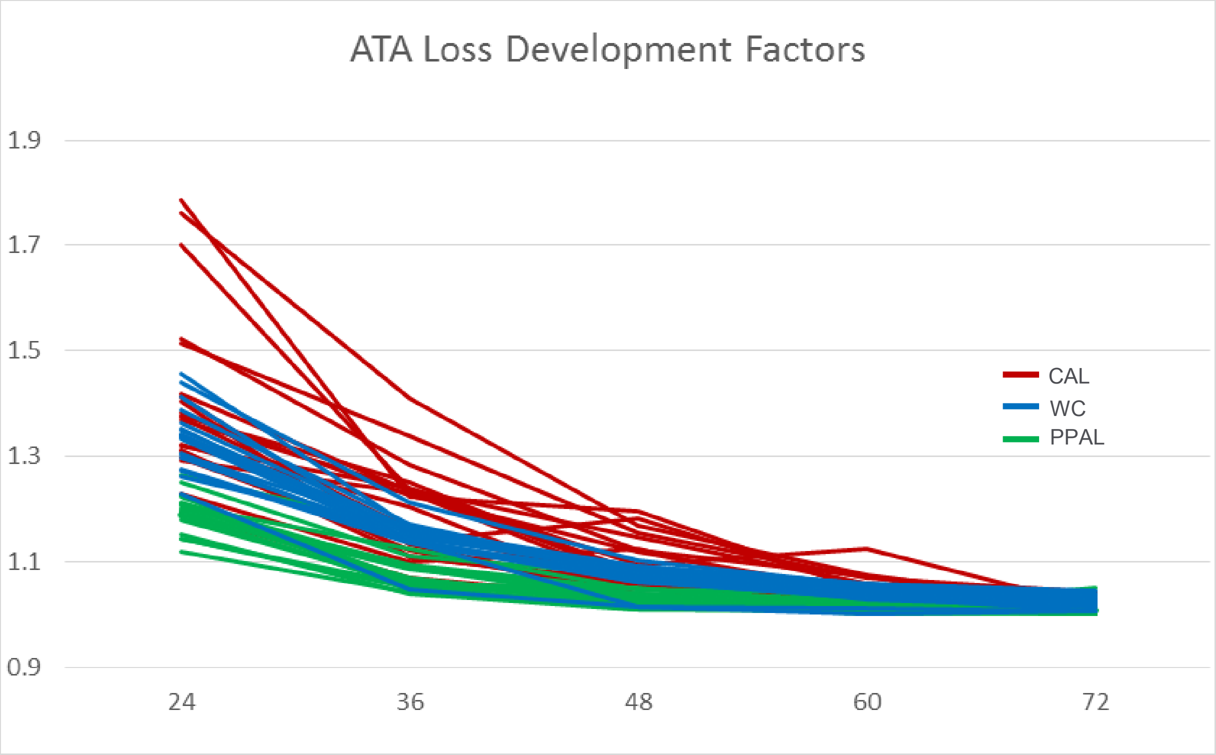
Munich RE

- 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
- In R, use *fanny(data, k=2,…)* from package "cluster" for fuzzy clustering
- Gaussian Mixed Models can also produce soft clusters

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%

Soft Clusters and Credibility

Schedule P: New Example

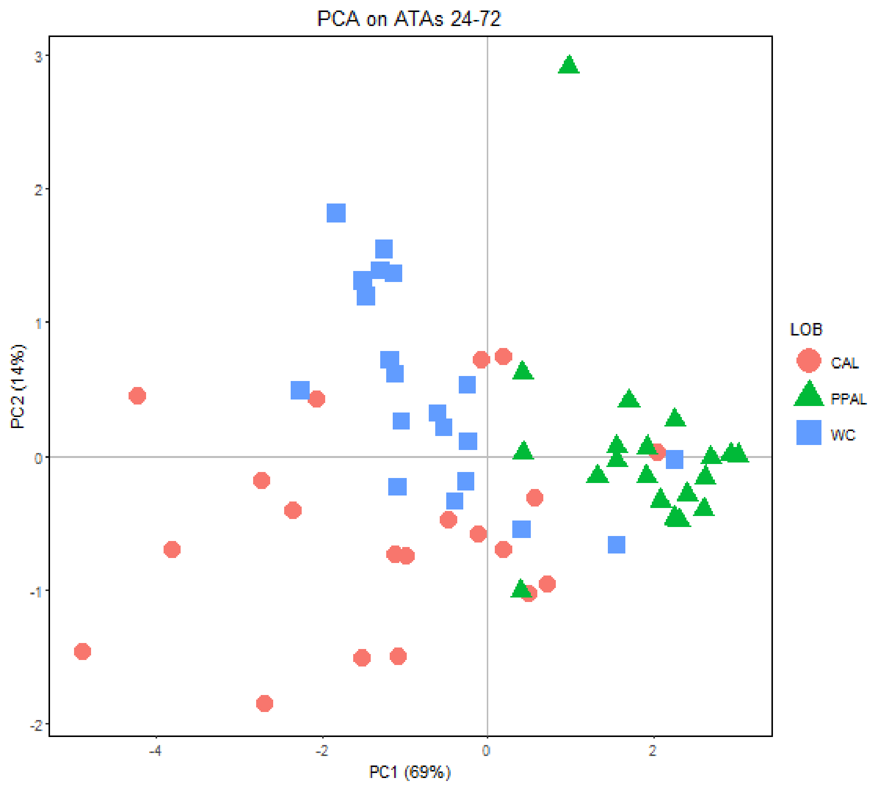


Soft Clusters and Credibility
K-means and PCA

- No real distance between clusters
- K-means is challenged with overlapping clusters

K-means

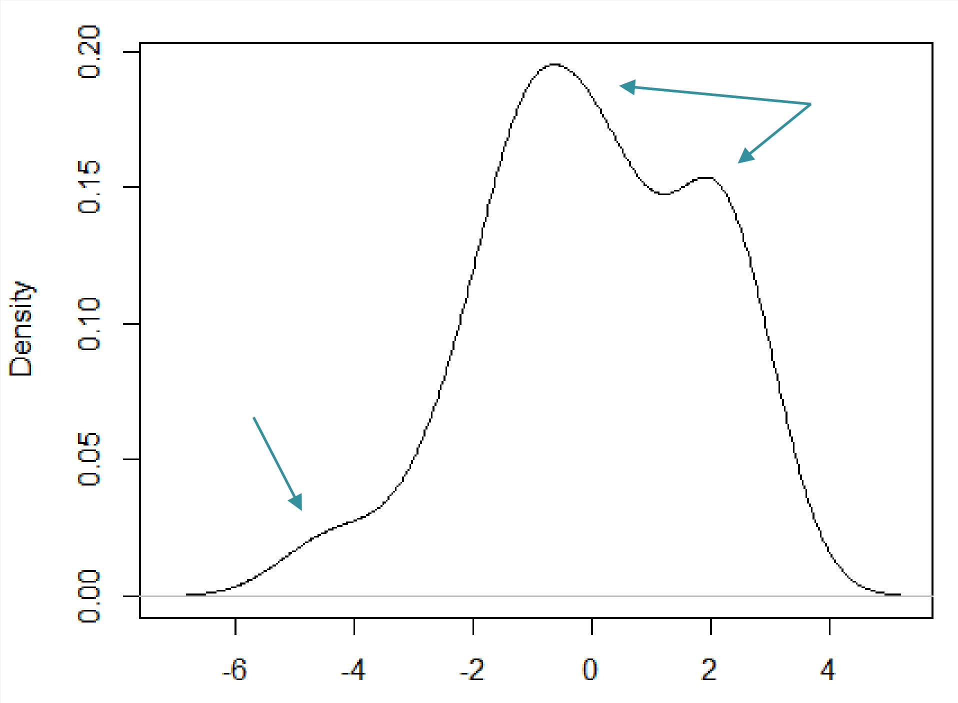
	1	2	3
CAL	**8**	3	**9**
PPAL	0	**16**	4
WC	1	2	**17**



Soft Clusters and Credibility

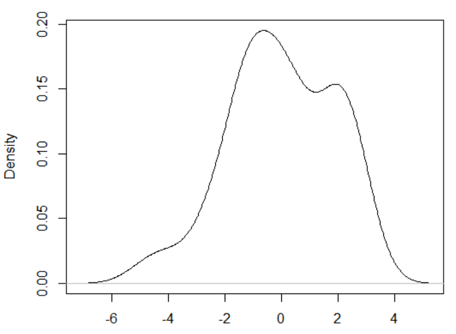
PC1 Density

First Principal Component



Soft Clusters and Credibility
The Theory

- Assume that the density of the data *(y)* is described by a mixture of number *(g)* of component densities *f(y) in* some unknown proportions (π).



$$pdf = \sum_{i=1}^{g} \pi_i \, f_i(y)$$

- For clustering, ***g*** will be the number of clusters
- Calculate the posterior probability (Bayes Theorem) that an observation y_j belongs to the *i-th* component of the mixture:

$$\tau_i(y_j) = \frac{\pi_i f_i(y_j)}{f(y_j)}$$

- If we assume that the data in the clusters is independent and normally distributed, we can use a Gaussian Mixture Model (GMM).

Soft Clusters and Credibility
Results

- GMM outperforms K-means when clusters are overlapping.
- These results were obtained using R package "*Rmixmod*". Multiple other options are possible *(ex: mclust, mixtools…)*
- Bayesian Information Criterion is used to determined the number of clusters.

Probabilities

	1	2	3
CAL	0%	0%	100%
CAL	0%	0%	100%
		…	
PPAL	0%	100%	0%
PPAL	0%	100%	0%
		…	
WC	93%	0%	7%
WC	94%	0%	6%
		…	

Gaussian Mixed Model

	1	2	3
CAL	1	1	**18**
PPAL	0	**17**	3
WC	**17**	1	2

Pros and Cons of GMM

Munich RE

Pros:

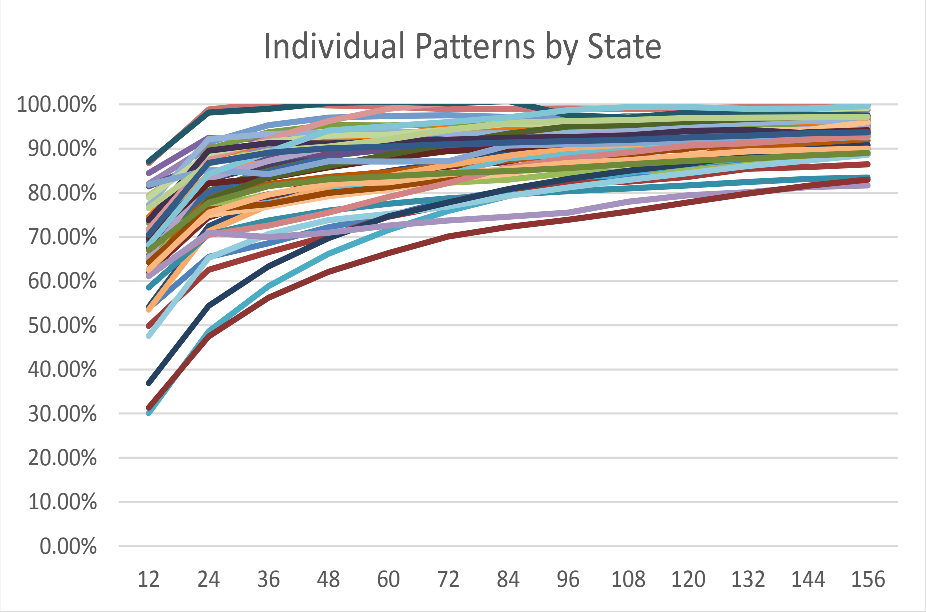
- It is a well defined statistical model allowing for hypothesis testing and validation
- Compared to K-means, it considers the cluster covariance
- It produces soft clusters

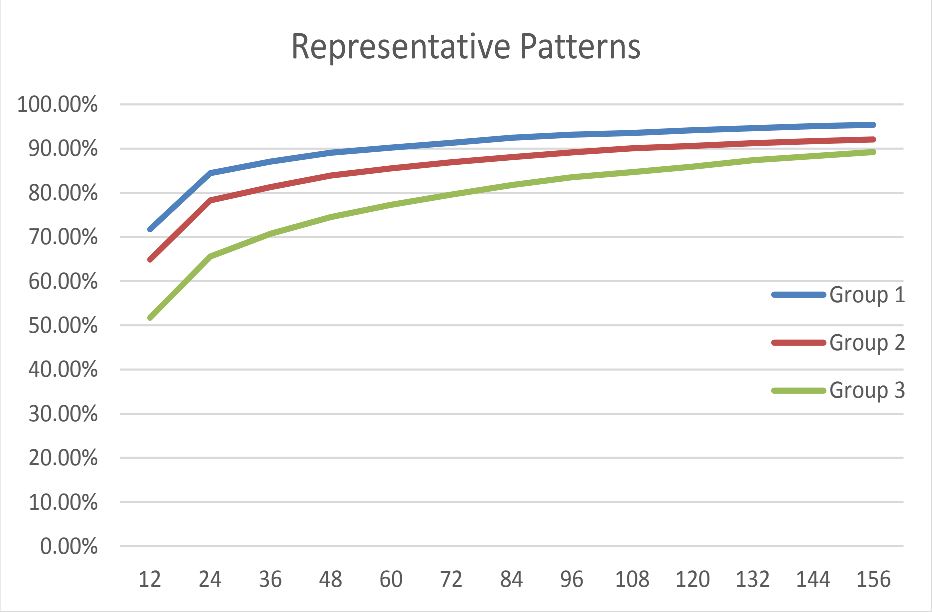
Cons:

- GMMs are not robust against outliers. It is recommended to use t-distribution if outliers are present
- Over-parameterization: Everything can be fitted with enough number of mixture components

Soft Clusters and Credibility
Practical Applications

We use “soft clustering” to collapse a large number of patterns into a smaller set of representative groups.





Soft Clusters and Credibility
Practical Applications

State	Pr (G_1|X_i)	Pr (G_2|X_i)	Pr (G_3|X_i)
Alabama			
Alaska			
Arizona			
Arkansas			
California			
⋮	⋮	⋮	⋮
Wyoming			
===========	======	======	======
Grand Total	Pr (G_1)	Pr (G_2)	Pr (G_3)

Soft clustering can be estimated using Mixed Models or Fuzzy clustering.

Membership in any benchmark Group is estimated using Bayes' Theorem.

$$\Pr(G_j|X) = \frac{\Pr(G_j) \cdot f(X|G_j)}{\sum_{j=1}^{3} \Pr(G_j) \cdot f(X|G_j)}$$

Example: Loss development for Alaska=

$\Pr(G_1|X_{Alaska})$ * Benchmark Group 1

+ $\Pr(G_2|X_{Alaska})$ * Benchmark Group 2

+ $\Pr(G_3|X_{Alaska})$ * Benchmark Group 3

Conclusion
Key Takeaways

- Clustering techniques help us obtain a better understanding of the loss development:
  - 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

- Investigate connection between Soft Clustering and Credibility Theory

Selected References

Munich RE

1. D. Clark (2017) **Estimation of Inverse Power Parameters via GLM**, *Actuarial Review, May-June 2017* , https://ar.casact.org/estimation-of-inverse-power-parameters-via-glm/

2. T. Hastie, R. Tibshirani, J. Friedman (2009) **The Elements of Statistical Learning - Data Mining, Inference, and Prediction**, Springer https://web.stanford.edu/~hastie/Papers/ESLII.pdf

3. 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/

4. C. Hennig, M.Meila, F. Murtagh, R.Rocci (2017) **Handbook of Cluster Analysis,** CRC Press

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. M. Steinbach, L. Ertoz, V. Kumar, **“The Challenges of Clustering High Dimensional Data”**, https://www-users.cs.umn.edu/~kumar001/papers/high_dim_clustering_19.pdf

8. J. VanderPlas, “**Python Data Science Handbook”,** O'Reilly Media, http://shop.oreilly.com/product/0636920034919.do

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

Thank you!