Antitrust Notice

- 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.

CLUSTERING ENDGAME

Dave Clark
Diana Rangelova
Munich Reinsurance America, Inc.

CLRS – September 2019

Munich RE

Agenda

- Introduction
- Visualization of Multidimensional Data
- Clustering Methods Applied to Overlapping Groups
- Outliers and Noise
- Practical Considerations

Introduction
What is Clustering?

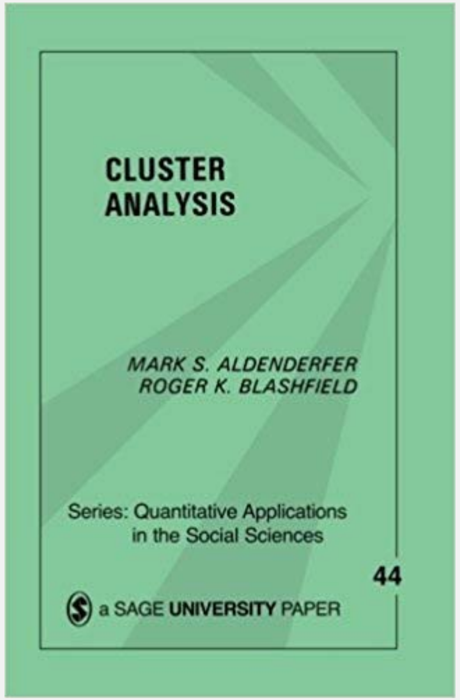
- A cluster is a group of similar objects
- Clustering is an unsupervised learning technique: No need to define the groups in advance
- It is essential to assess the usefulness and meaning of the identified groups

Source: https://www.nasa.gov Image Credit: ESA/Hubble & NASA

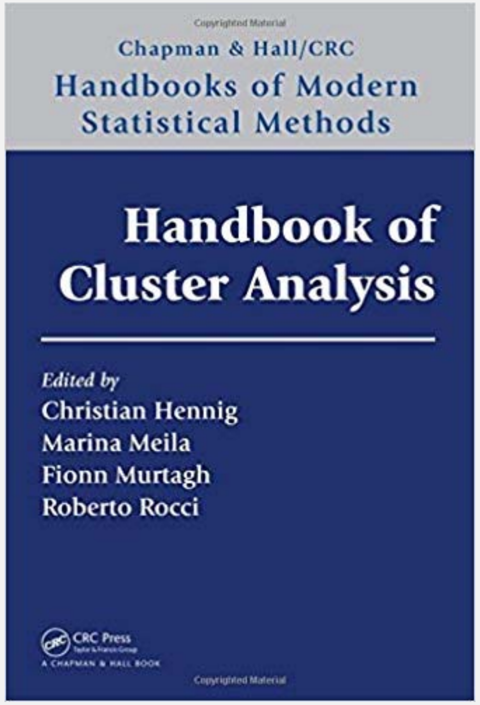
Introduction
Publications on Clustering

Munich RE

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



1984 - **88 pages**



2015 - **773 pages**

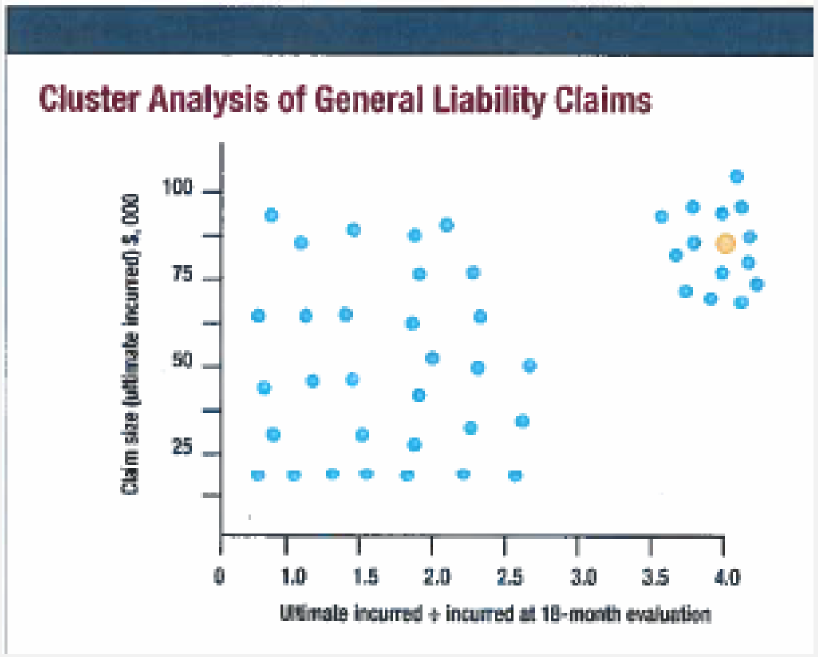
Why Clustering?

- What questions could be answered with cluster analysis?
  - Exploratory analysis
  - Test the data homogeneity
  - Find a benchmark
- What kind of data can be clustered?
  - Segments, contracts, claims…
  - Counties, regions…
  - Loss development patterns, loss ratios, severity, frequency, etc.

Introduction
What Does Reserving Data Look Like?

Munich RE

- Text book example



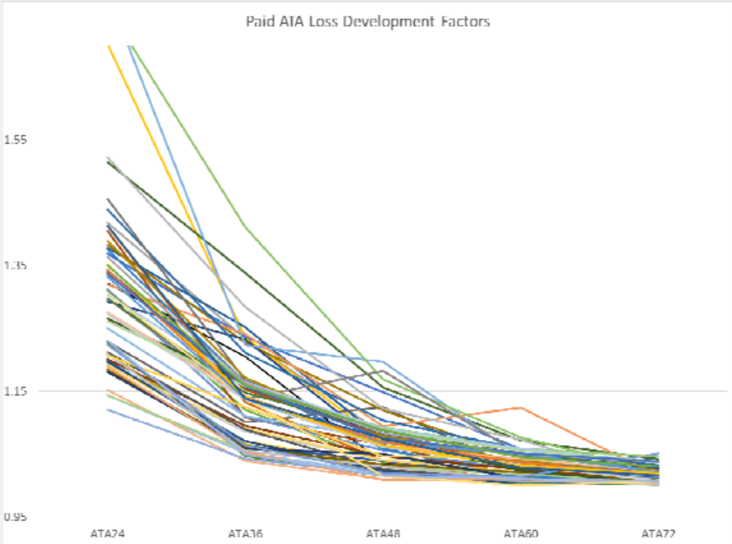
Source: Associates in Data Analytics (AIDA) 181 textbook

- One or two dimensions
- No outliers
- Distinct clusters

Introduction

What Does Reserving Data Look Like?

- Real data example
  - Multidimensional observations
  - Overlapping clusters
  - Outliers and noise are present



- Schedule P Example:
  - CAS Schedule P data for Loss Reserving [10]
  - 3 lines: CAL, PPAL,WC
  - 20 observations per line
  - Each observation represents a company

Paid ATA Loss Development Factors

Line	12	24	36	48	60	72
CAL	1.87	1.32	1.20	1.04	1.04	1.01
CAL	1.99	1.42	1.23	1.08	1.03	1.02
			...			
PPAL	2.26	1.21	1.07	1.02	1.01	1.00
PPAL	1.78	1.20	1.06	1.04	1.02	1.01
			...			
WC	2.22	1.34	1.16	1.09	1.06	1.05
WC	2.47	1.44	1.21	1.10	1.06	1.03
			...			

Agenda

Munich RE

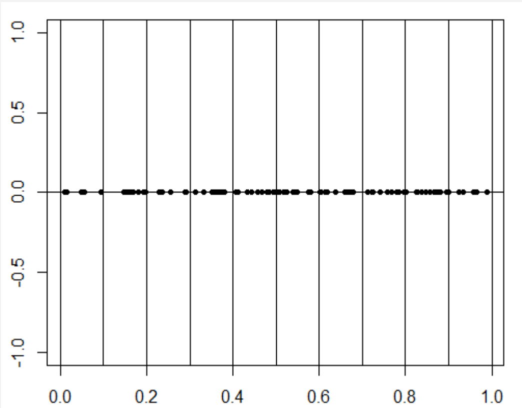
- Introduction
- Visualization of Multidimensional Data
  - Statistical Challenge
  - Why it is Important to Visualize Data
  - Dimension Reduction Techniques
- Clustering Methods Applied to Overlapping Groups
- Dealing with Outliers and Noise
- Practical Considerations

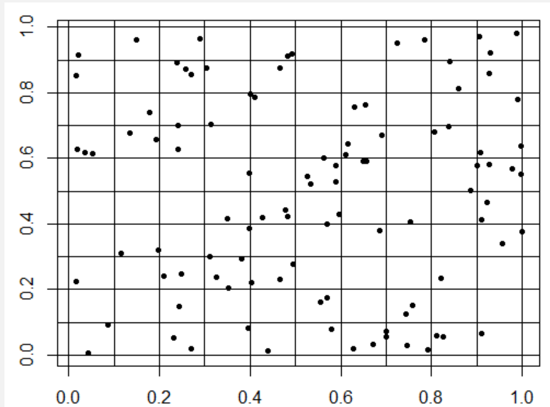
Visualization
Statistical Challenge

- *"A high dimensional space is a lonely place"*

Bernhard Scholkopf

Randomly generated 100 points in 1D and 2D





- *"…Thus, it is often said, "**in high dimensional spaces, distances between points become relatively uniform**." In such cases, the notion of the nearest neighbor of a point is meaningless…." [8]*

Visualization

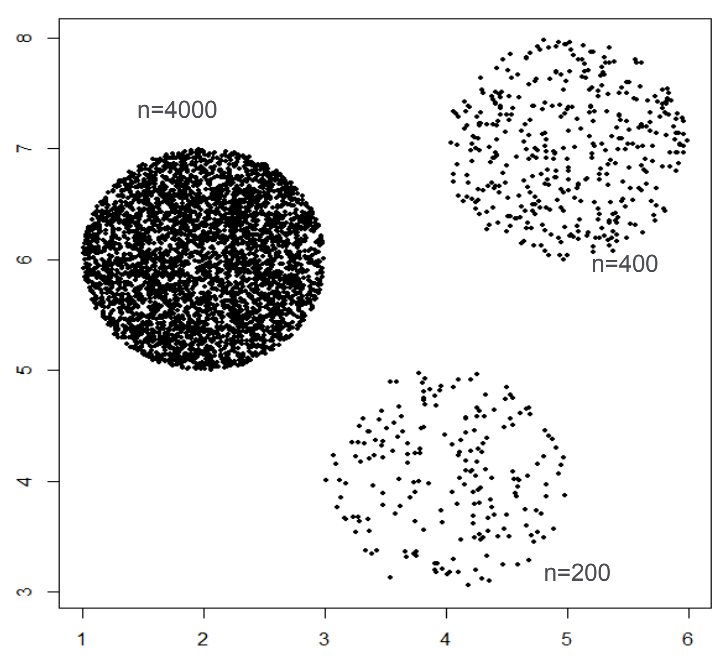
Why Is It Important to Visualize Data?

- Choose the most appropriate clustering model for your data
  - Are the clusters spherical? (K-means)
  - Are the clusters overlapping? (Fuzzy clustering, Gaussian Mixture Models)
  - Noise points (Density-based clustering)
  - Select the number of clusters
- Explain clusters and communicate results

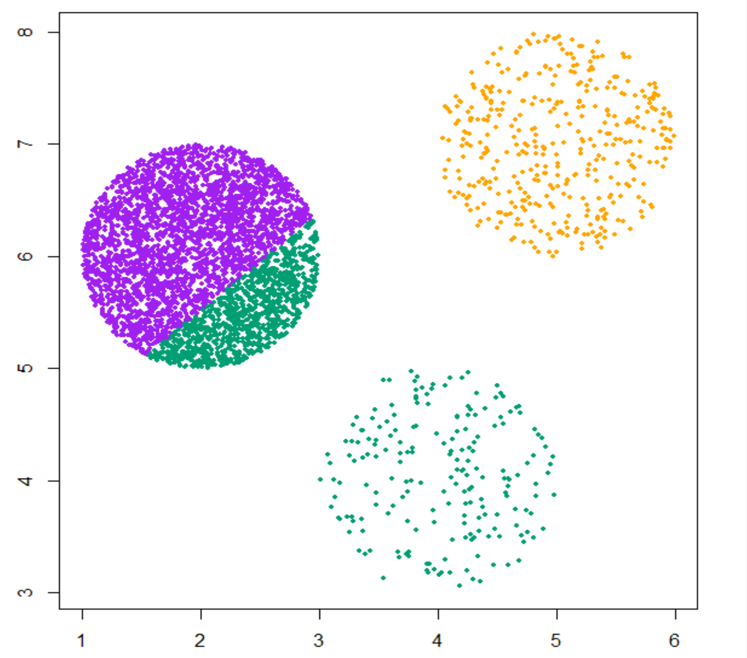
Visualization
Why Is It Important to Visualize Data?

- Validate results

Original data



K-means result

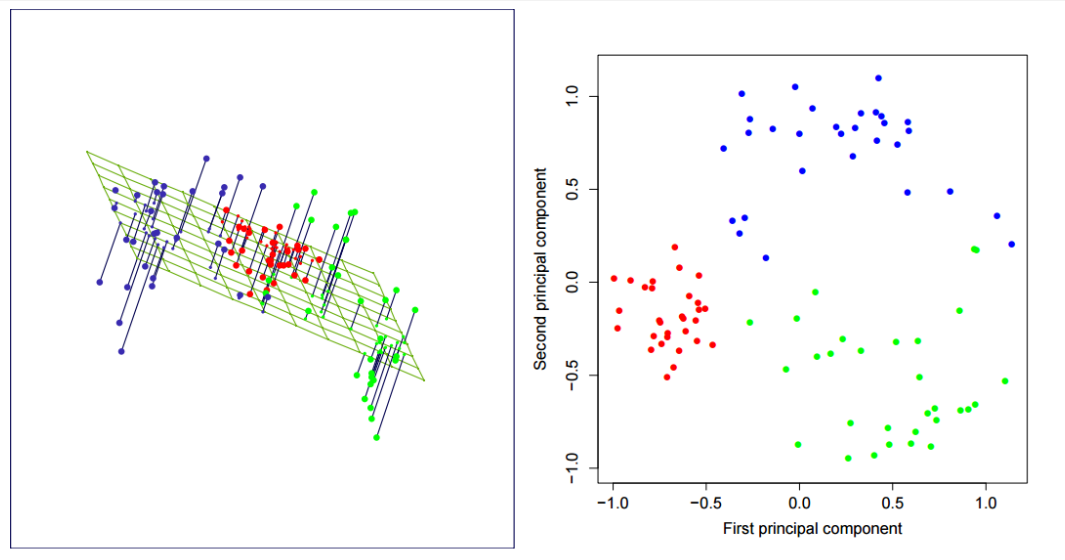


Visualization
Dimension Reduction Techniques

- 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.



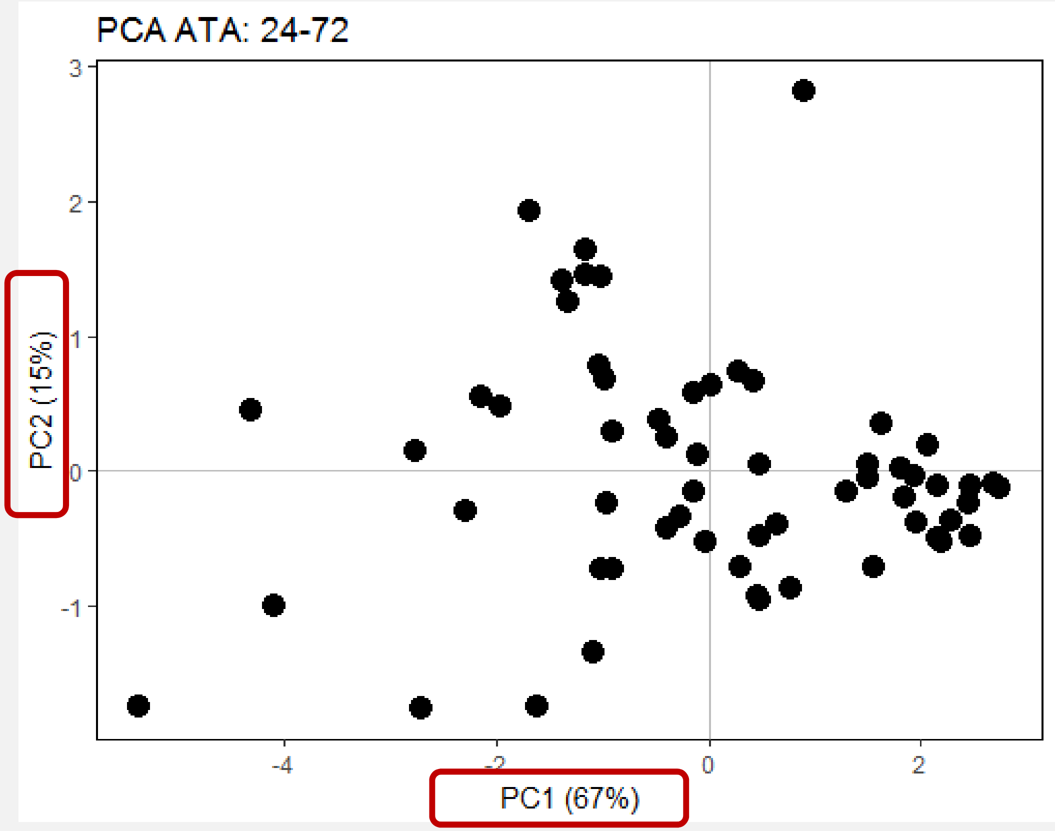
Source: The Elements of Statistical Learning

PCA
Principal Component Analysis

- Think about viewing a galaxy from "above" rather than the side: what angle do we want in order to get the most understanding of the "shape" of the galaxy?

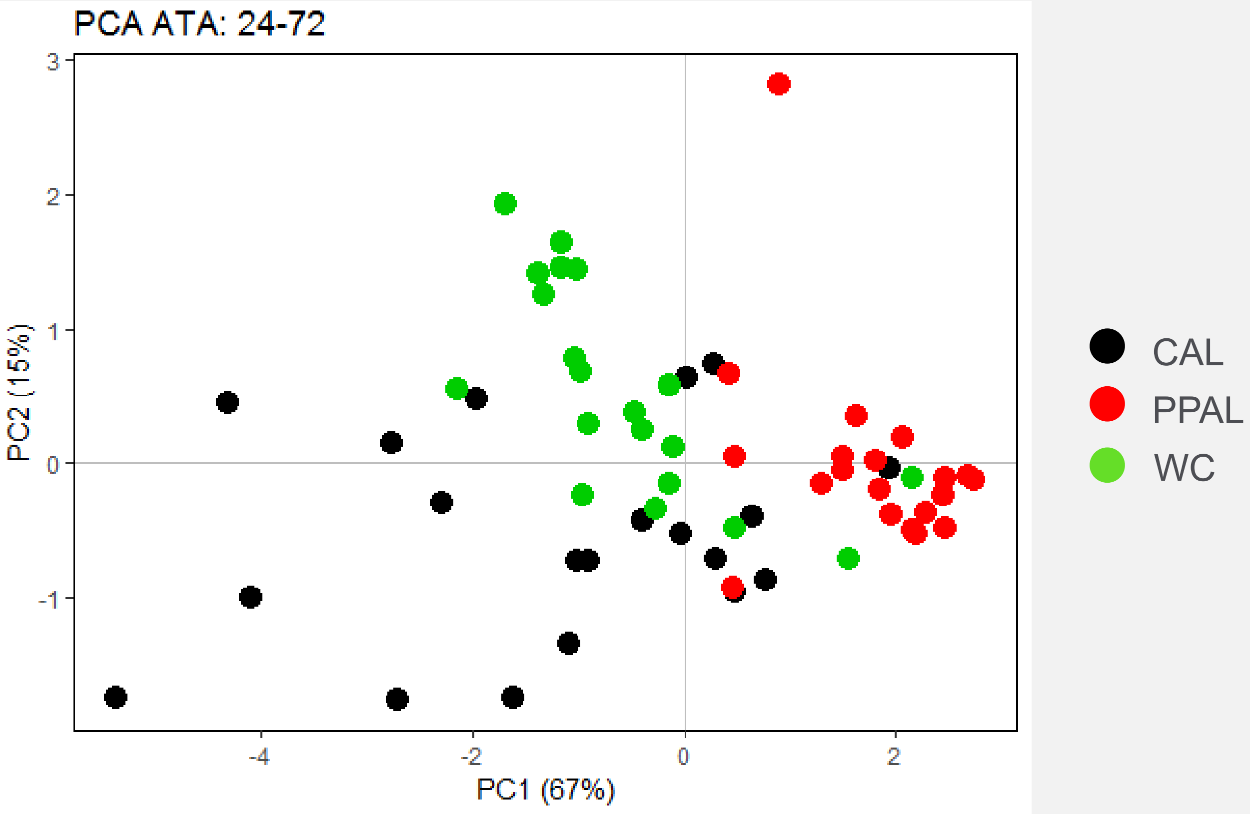
Source: https://www.nasa.gov/feature/goddard/2017/a-new-angle-on-two-spiral-galaxies-for-hubbles-27th-birthday;
Credits: NASA, ESA, and M. Mutchler (STScI)

PCA
Schedule P example: Visualization



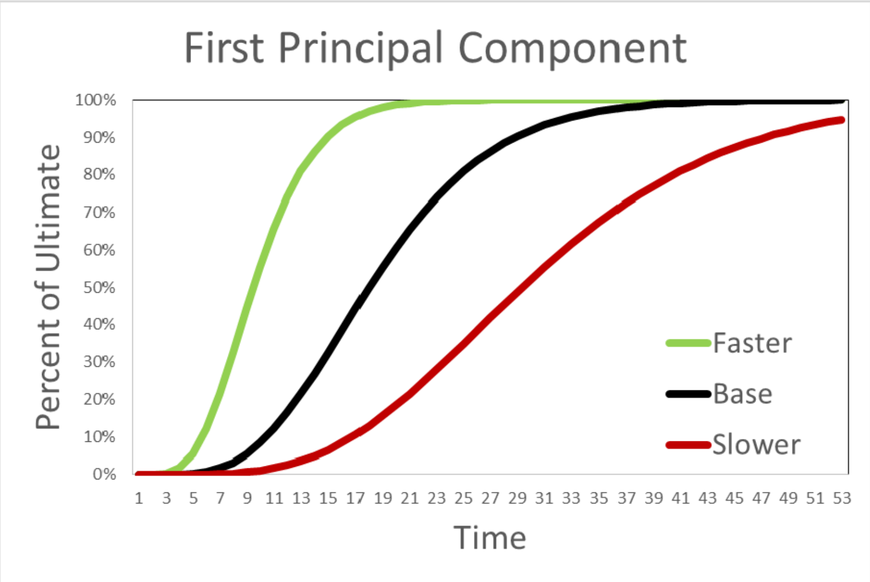
PCA

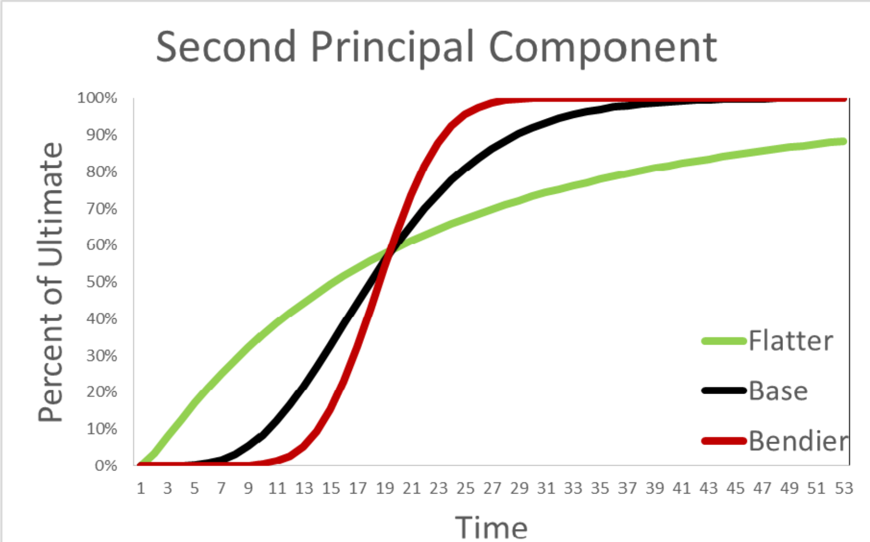
Schedule P example: Visualization - LOB



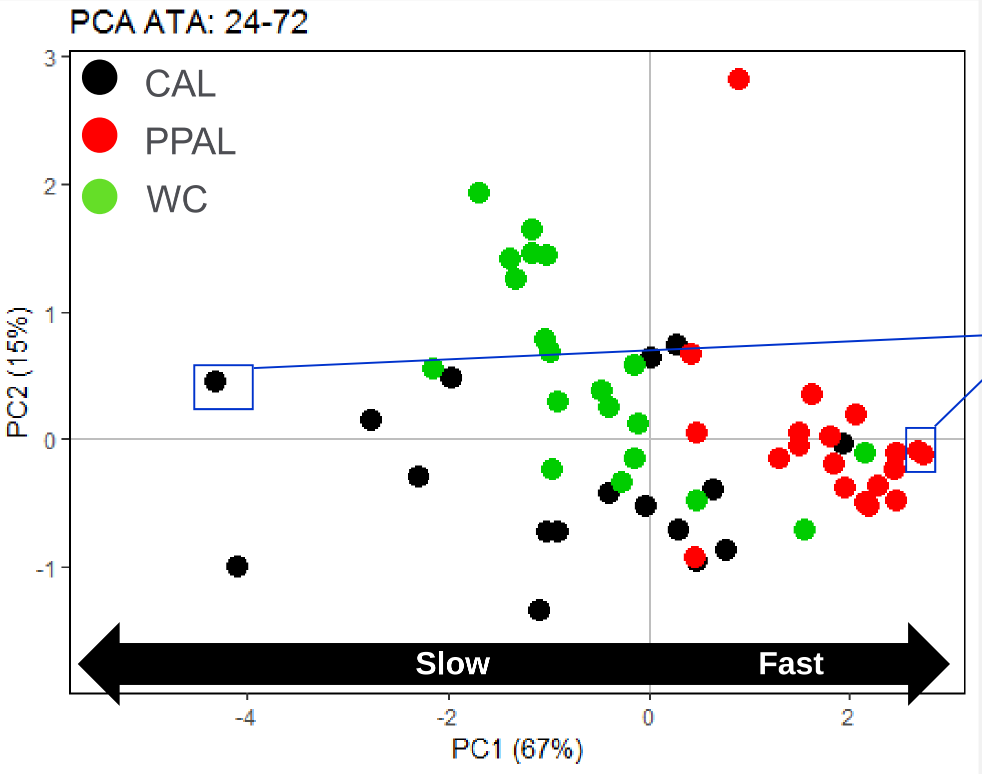
PCA
Interpretation

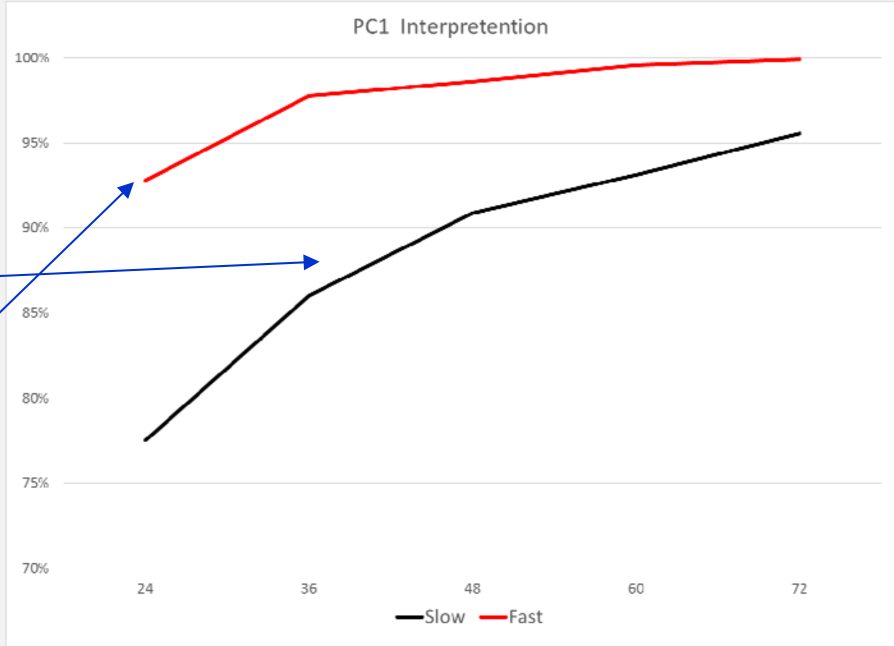
- PCA provides an opportunity for interpretation
  - PC1 captures the mean loss development
  - PC2 indicates a change in the loss curve shape



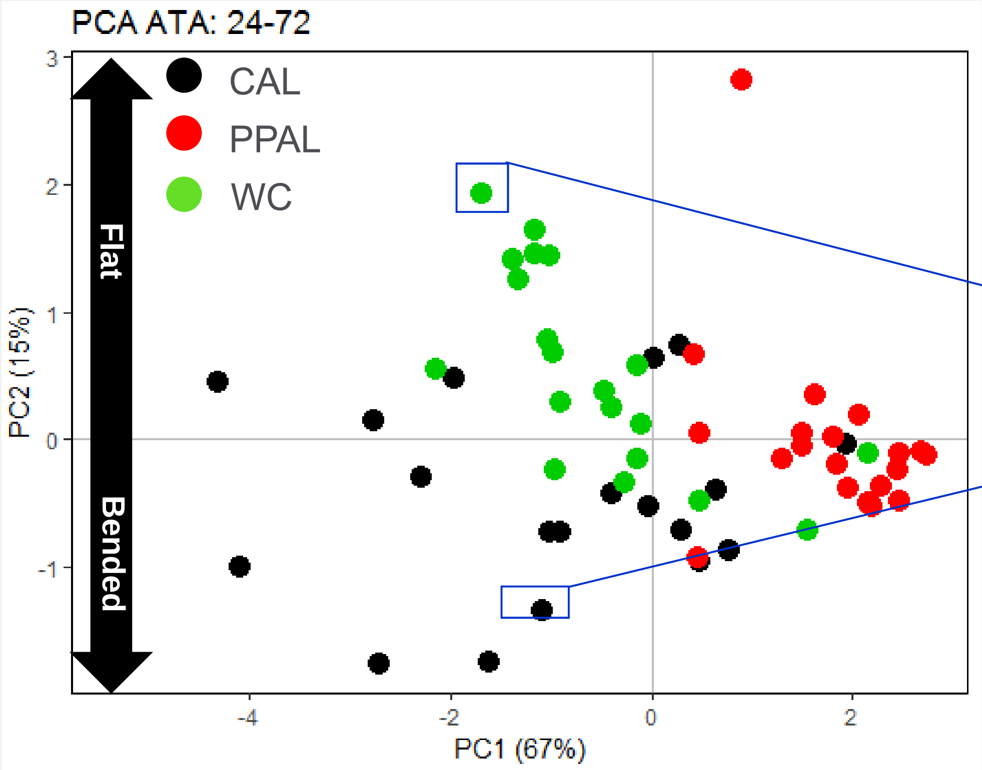


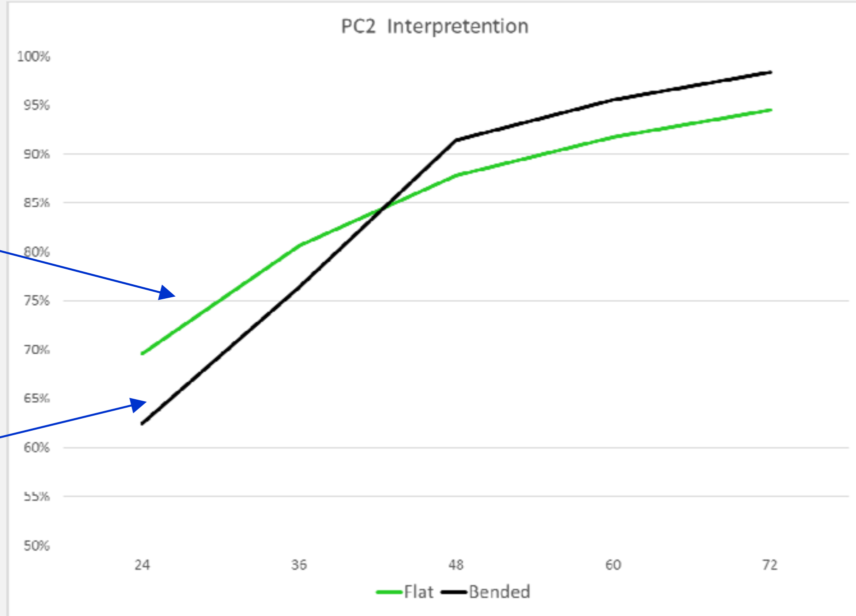
PCA
PC1 Interpretation





PCA
PC2 Interpretation





Visualization
Dimension Reduction Techniques

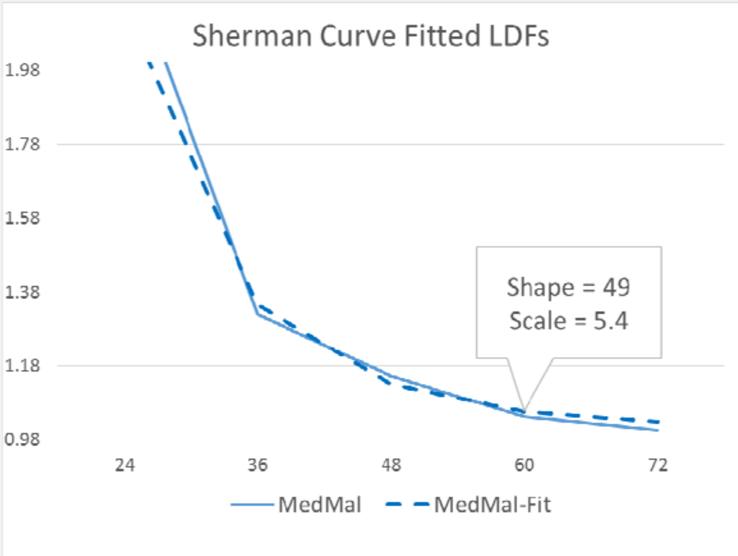
- Principal Component Analysis (PCA)
- Data Transformation (Curve Fitting)

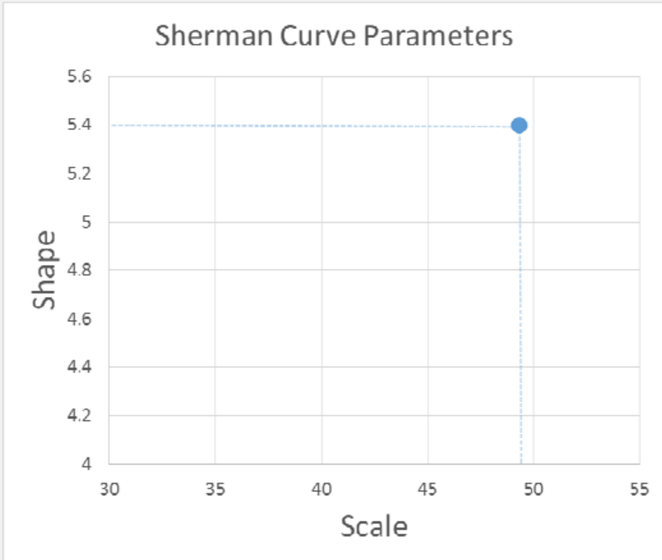
Sherman Curve

Munich RE

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

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





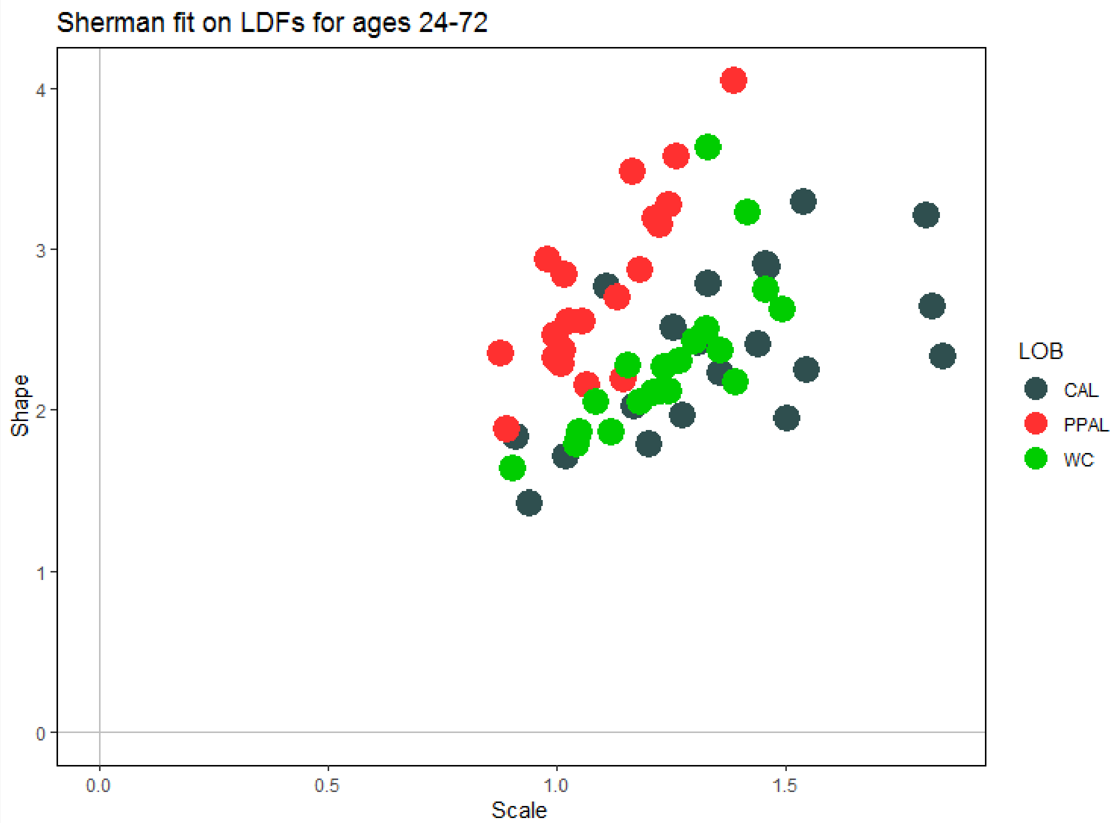
Data Transformation

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



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, not on maintaining the distances between points

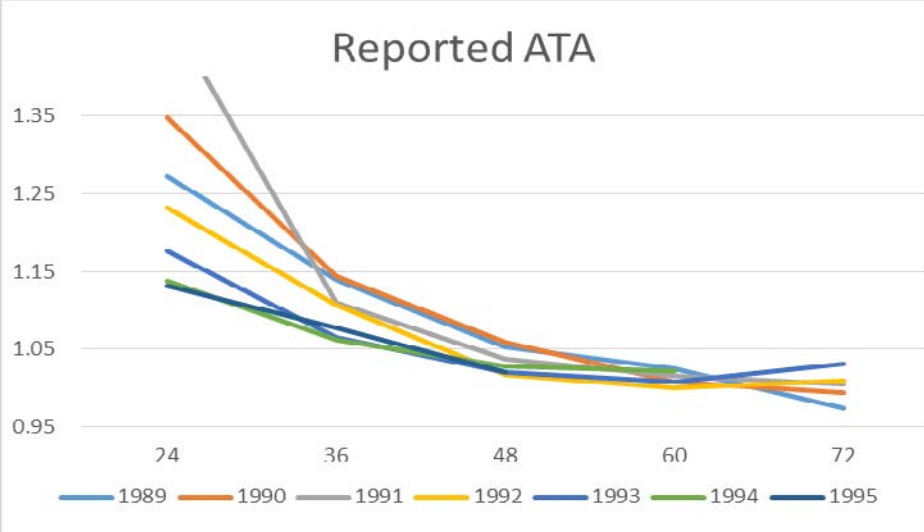
Data Transformation

Another Schedule P Example

Munich RE

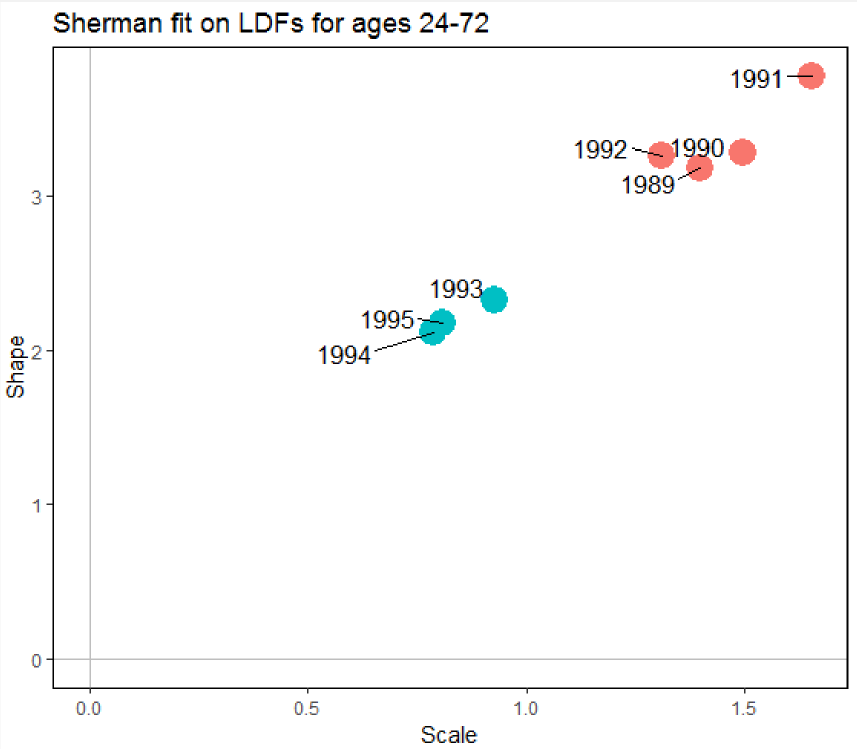
Reported ATA					
	24	**36**	**48**	**60**	**72**
1989	1.27	1.14	1.05	1.03	0.97
1990	1.35	1.14	1.06	1.01	0.99
1991	1.48	1.11	1.04	1.02	1.01
1992	1.23	1.11	1.02	1.00	1.01
1993	1.18	1.06	1.02	1.01	1.03
1994	1.14	1.06	1.03	1.02	
1995	1.13	1.08	1.02		

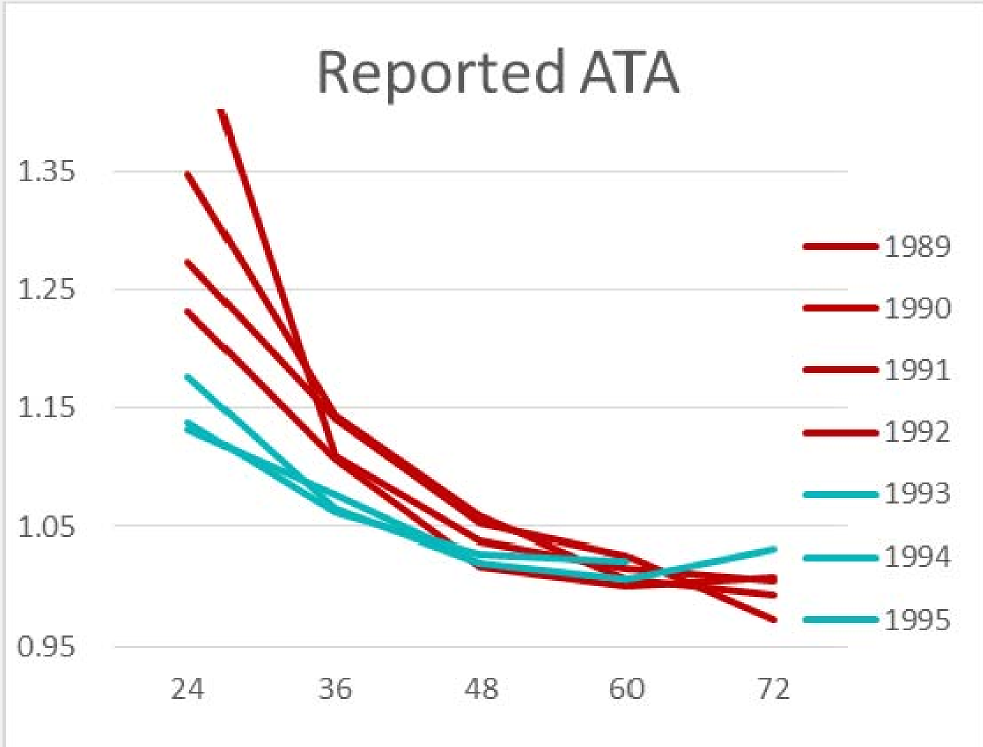
Source: CAS Schedule P Reported LDF – CAL



Data Transformation

Another Schedule P Example





Agenda

Munich RE

- Introduction
- Visualization
- Clustering Methods Applied to Overlapping Groups
  - K-means
  - Fuzzy Clustering
  - Gaussian Models
- Dealing with Outliers and Noise
- Practical considerations

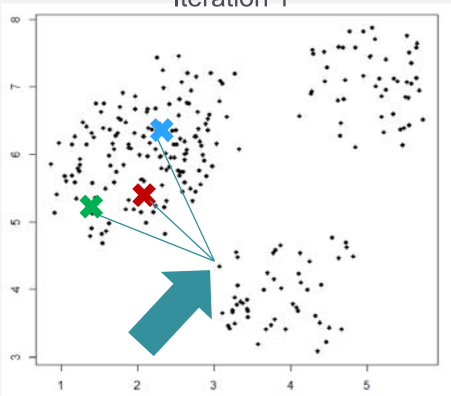
Clustering Methods Applied to Overlapping Groups
K-means

- K-means is simple, fast and efficient

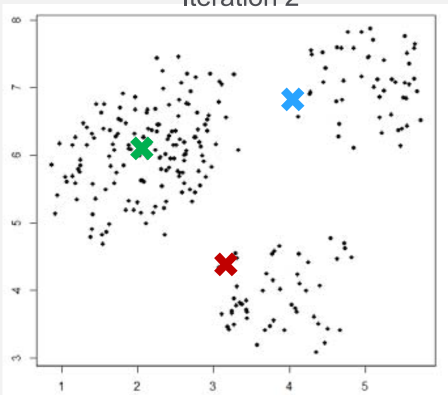
- How does K-means work?
  - Initiate the centroids
  - Assign points to the closest centroid
  - Recalculate new centroid
  - Iterate until no point left to be reassigned

- In R, use kmeans() from package "stats"

Iteration 1



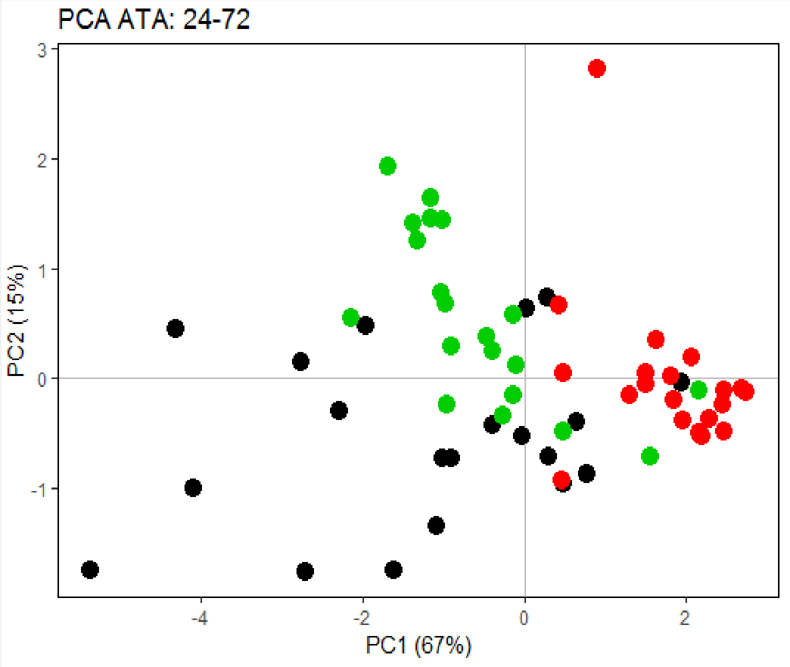
Iteration 2



Clustering Methods Applied to Overlapping Groups

K-means

- K-means <u>does not perform</u> well when:
  - **There are no natural distinct clusters** ➡
  - Clusters are of different size
  - Clusters are not roughly spherical
  - Outliers exist



Clustering Methods Applied to Overlapping Groups

Fuzzy Clustering: Schedule P Example

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
- Results are obtained using R function *fanny()* from the package "cluster"

LOB	Fuzzy 1	Fuzzy 2	Fuzzy 3
CAL	68%	18%	14%
CAL	67%	27%	6%
CAL	49%	37%	14%
CAL	31%	64%	6%
		…	
PPAL	2%	1%	97%
PPAL	9%	3%	88%
PPAL	4%	2%	94%
PPAL	2%	1%	96%
		…	
WC	16%	80%	4%
WC	17%	81%	3%
WC	75%	21%	4%
WC	65%	31%	4%
		…	

Clustering Methods Applied to Overlapping Groups

Gaussian Mixture Models: Motivation

Munich RE

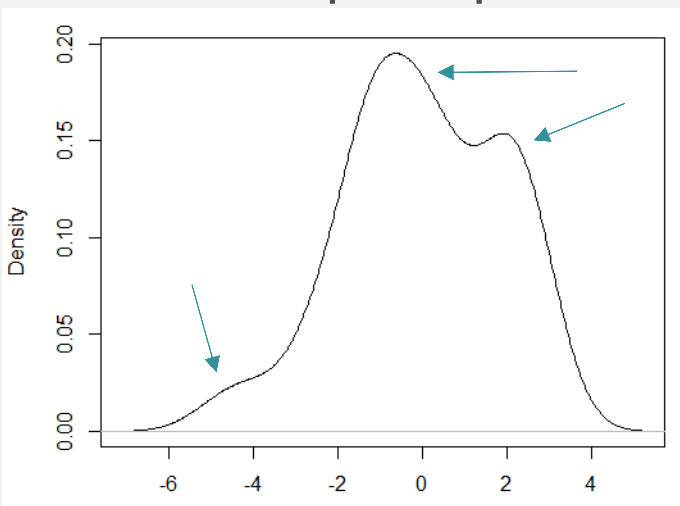
- **Probabilistic clustering:**
  - Each cluster is represented by a distribution
  - All observations are described by a mixture of these distributions
  - Well defined mathematical structure allows for:
    - Probabilistic assignments to clusters (soft clustering)
    - Generation of new points from a given cluster
    - Hypothesis testing
  - Allows for overlapping, non-spherical clusters, and clusters with varying size
  - Danger of overfitting and inappropriate distribution selection

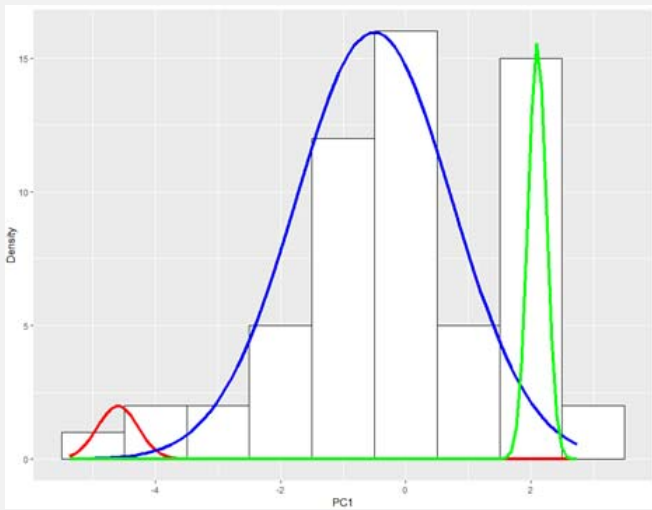
Gaussian Mixture Models: PC1 Density

- One dimensional example: using PC1 of our Schedule P example

- Fit a Gaussian distribution for each cluster

First Principal Component





Clustering Methods Applied to Overlapping Groups

Gaussian Mixture Models: Schedule P Example

Munich RE

- GMM work well for overlapping, non-spherical clusters, and clusters with varying size
- Results were obtained using R package "*Mclust*". Multiple other options are possible *(ex: mixtools, Rmixmod…)*
- Bayesian Information Criterion is used to determine the number of clusters.

LOB	1	2	3
CAL	93%	7%	0%
CAL	98%	2%	0%
CAL	0%	100%	0%
CAL	17%	83%	0%
		...	
PPAL	0%	0%	100%
PPAL	0%	0%	100%
PPAL	0%	0%	100%
PPAL	0%	0%	100%
		...	
WC	98%	2%	0%
WC	98%	2%	0%
WC	100%	0%	0%
WC	98%	2%	0%
		...	

Agenda

- Introduction
- Multidimensional Data
- Clustering Methods Applied to Overlapping Groups
- Outliers and Noise
  - Recognizing outliers and noise points
  - Dealing with outliers and noise points
- Practical considerations

Outliers and Noise Points

Recognizing Outliers and Noise points

- Types of outliers / noise
  - Points that are very different from the rest
  - Points that are too small
  - Erroneous points

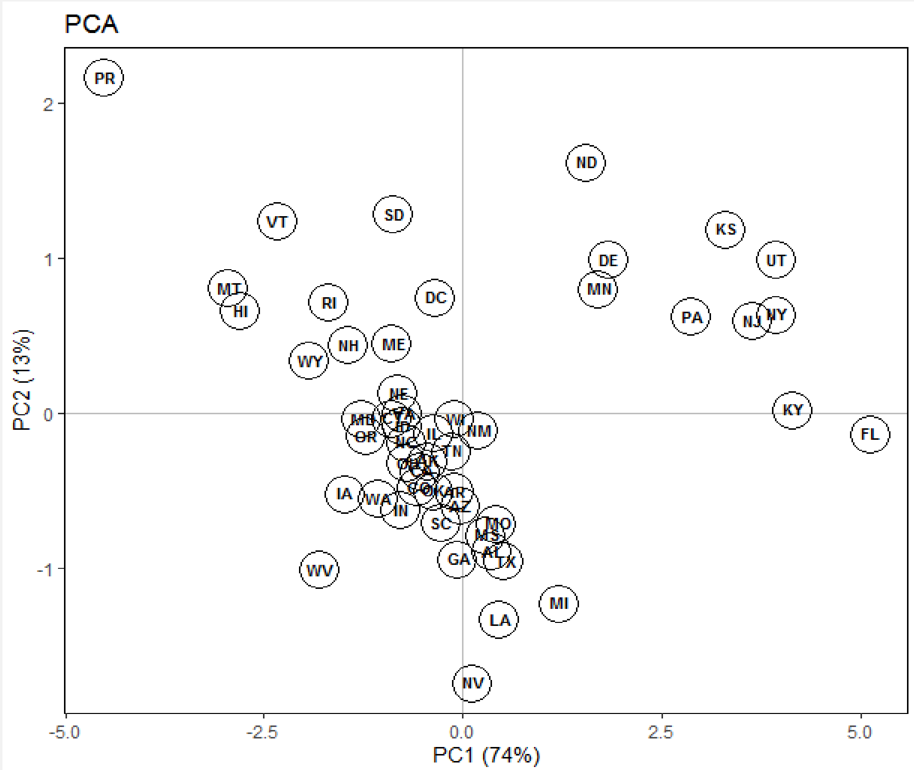
- Recognising Outliers and Noise points:
  - Visualization of the data
  - Increase the number of groups to detect and isolate small clusters
  - Fuzzy clustering: outliers are “equally remote” to all clusters. They will have similar membership to all clusters

Outliers and Noise Points
ISO Example

Munich RE

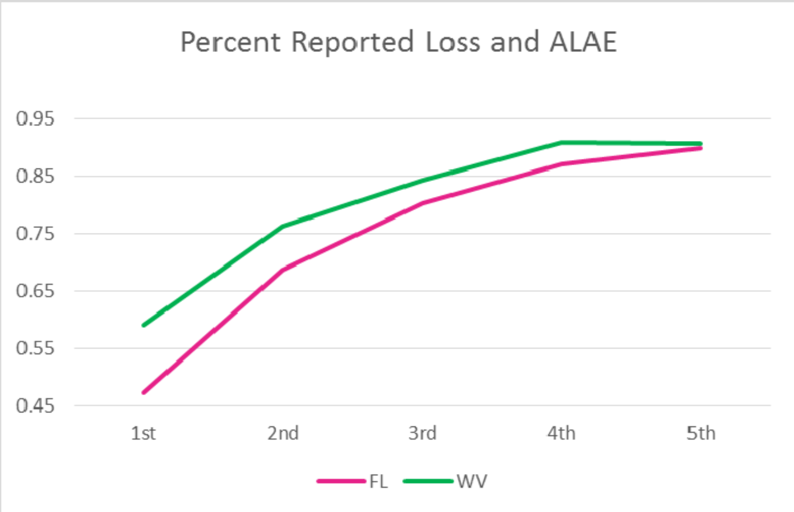
- ISO Commercial Auto patterns by State (51 obs. incl. Puerto Rico)
- Reported Loss and ALAE for AYs 2013-2017
- Percentage of Ultimate Loss – centered and standardized

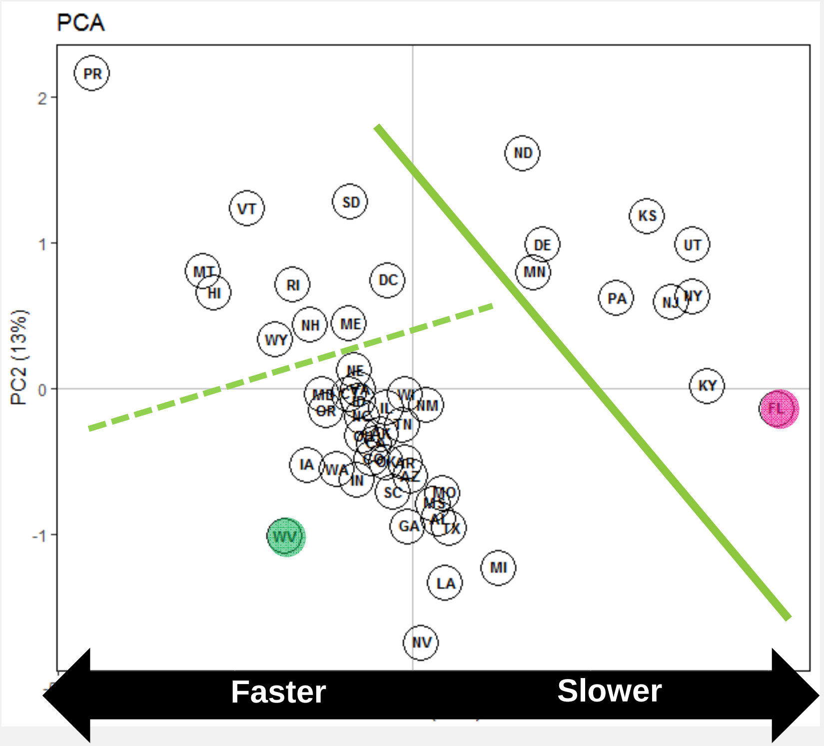
How many clusters are there and what is the Explanatory Variable?



Outliers and Noise Points
ISO Example

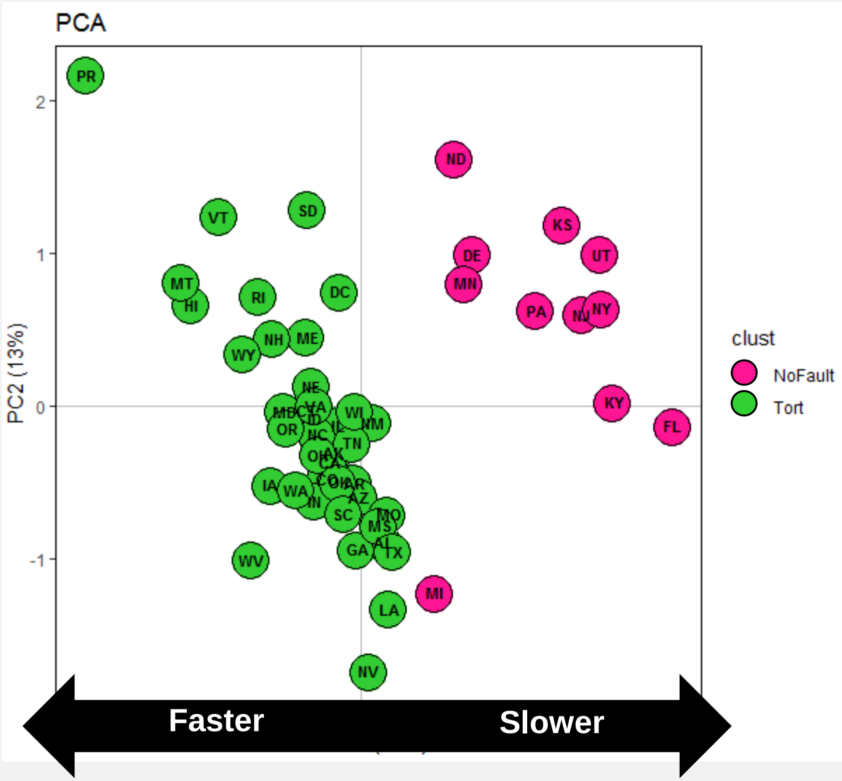
- 2-3 clusters
- One with slower pattern than the other
- One more spread than the other





Outliers and Noise Points
ISO Example

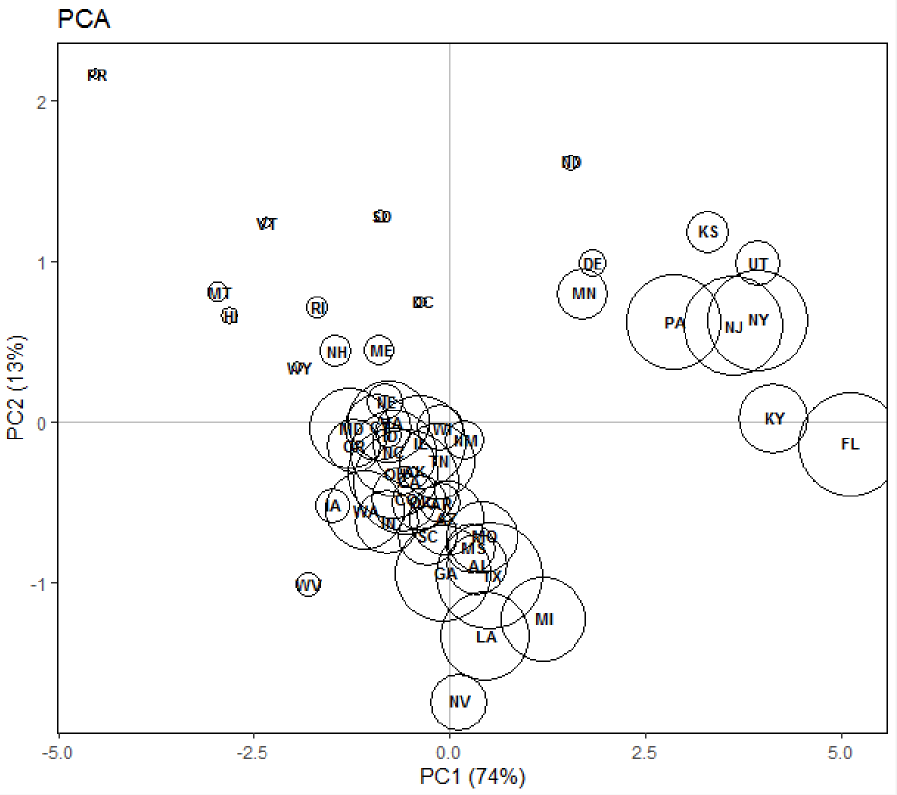
- Noticeable differences in the patterns of the states that have adopted No-Fault auto insurance laws
- Most “No-Fault” states have slower patterns



Outliers and Noise Points

ISO Example –Weights

- The weight for each state is based on the rank of the average ultimate loss for AYs 2013-2017
- Natural clusters become even more clear



Recognizing Outliers and Noise points

- Recognising Outliers and Noise points:
  - Visualization of the data
  - Increase the number of groups to detect and isolate small clusters
  - Fuzzy clustering: outliers are "equally remote" to all clusters. They will have similar membership to all clusters

Outliers and Noise Points
How to deal with them?

- Remove outliers before clustering
- Partial clustering algorithms that leave noise/outlier points outside the clusters (DBSCAN)
- Some methods are more robust than others when outliers are present (ex: K-medoids)
- Clustering with weights

Outlier and Noise Points
K-Medoids

- Similar to K-means but uses real data points as centroids for the clusters
- K-medoids is minimizing the distance to the “median” of the cluster and this makes it more robust.
- Its robustness is unlikely to work for:
  - Multi-dimensional space
  - Many outliers points

Outliers and Noise Points

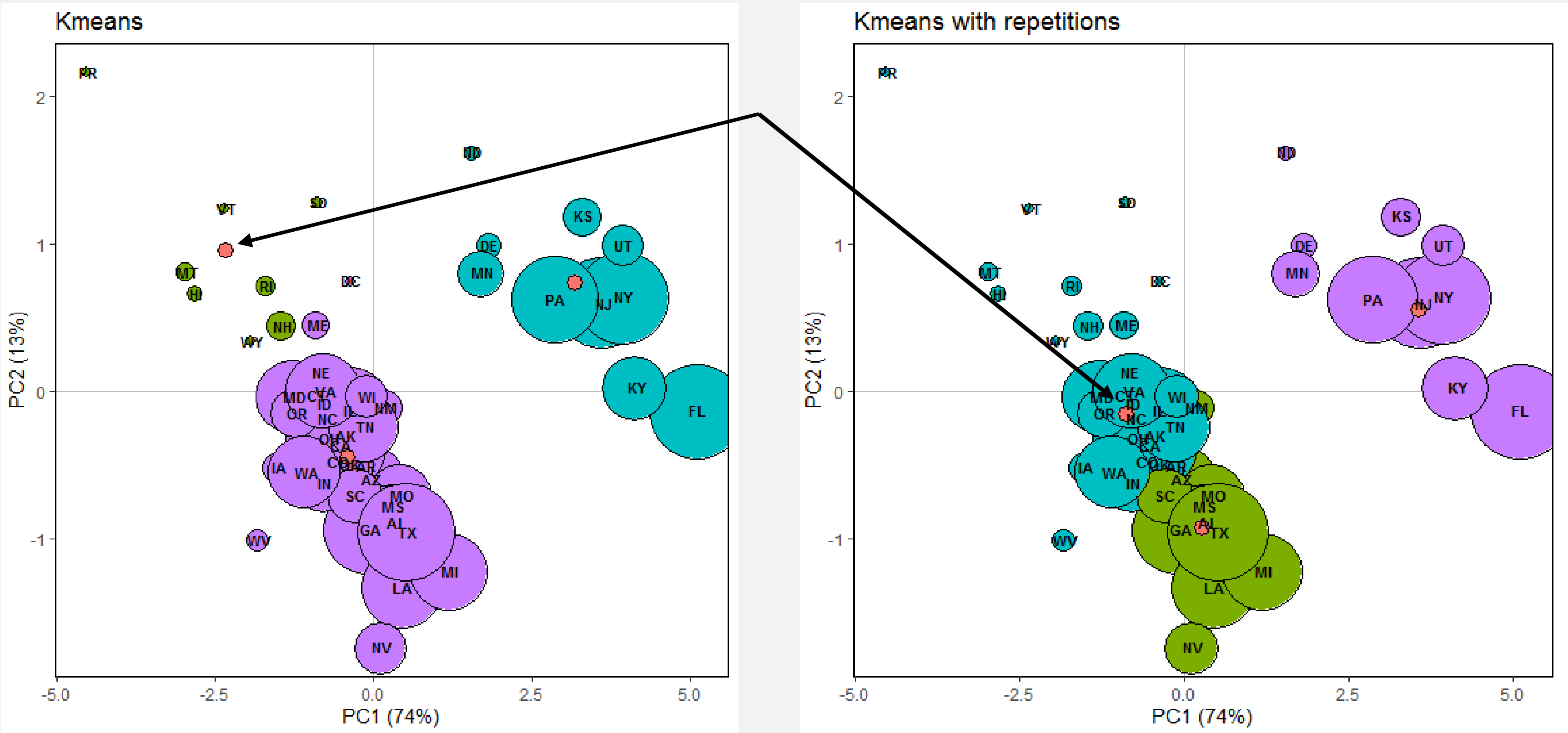
Clustering with Weights

- Easy way to introduce weights in the clustering model is to repeat several times the more important points
- ISO example: Repeat the observation based on the rank of their premium or ultimate values
  - TX is the largest of 51 observations

    =>repeat TX values 51 times
  - PR is the smallest

    =>PR will be in the data only once

Outlier and Noise Points
Clustering with Weights



This slide includes copyrighted material of Insurance Services Office, Inc. with its permissions

Agenda

Munich RE

- Introduction
- Visualization of Multidimensional Data
- Cluster Analysis
- Dealing with Outliers and Noise
- Practical Considerations
  - Correlations between LOB
  - Identifying drivers of loss development

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

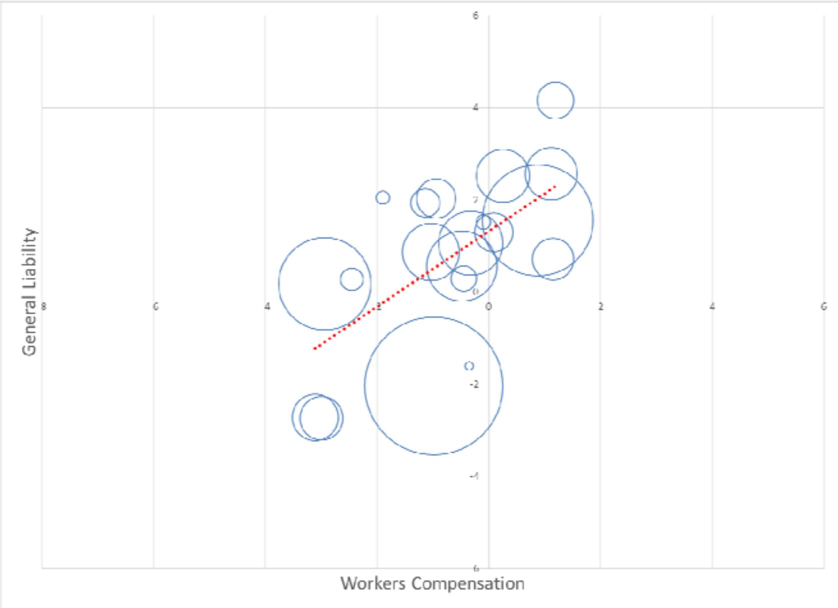
- Check if historical dependency is preserved in more recent years

First Principal Component for WC/GL

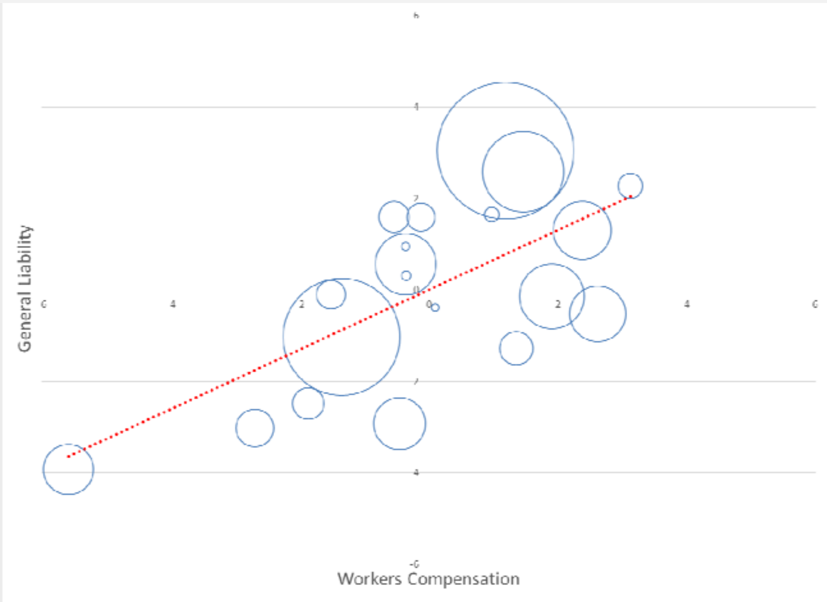
Munich RE

- PCA on Reported loss

1988 - 1997



1998 - 2007

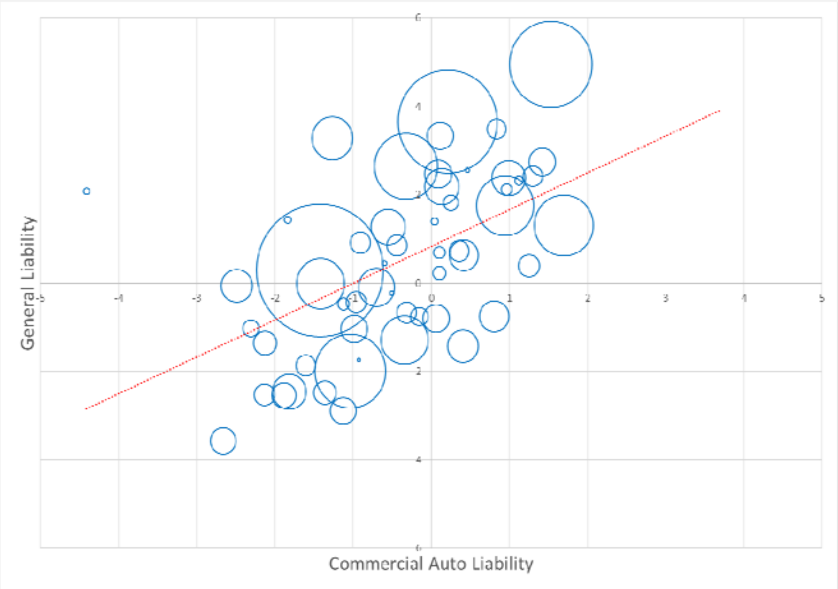


Note: bubble size corresponds to a company's average yearly premium volume

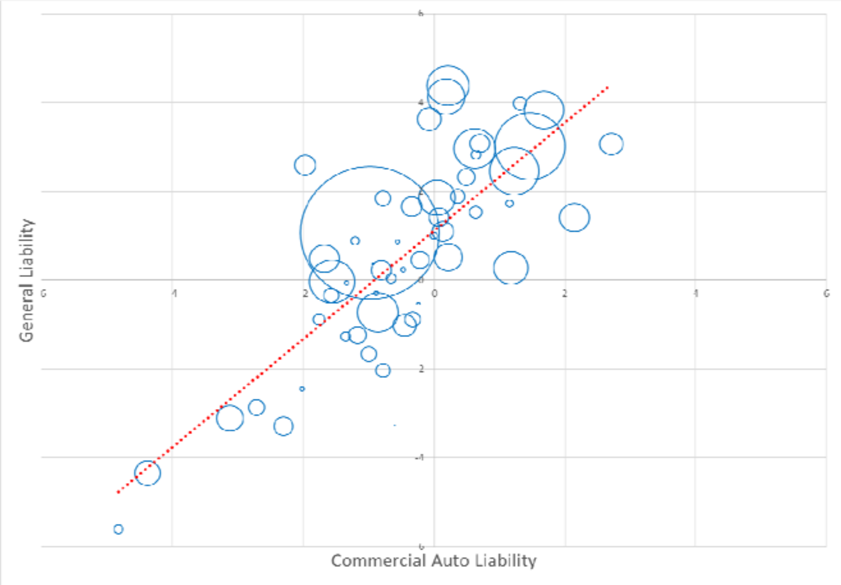
First Principal Component for CAL/GL

➤ PCA on Reported loss

1988 - 1997



1998 - 2007



Note: bubble size corresponds to a company's average yearly premium volume

Agenda

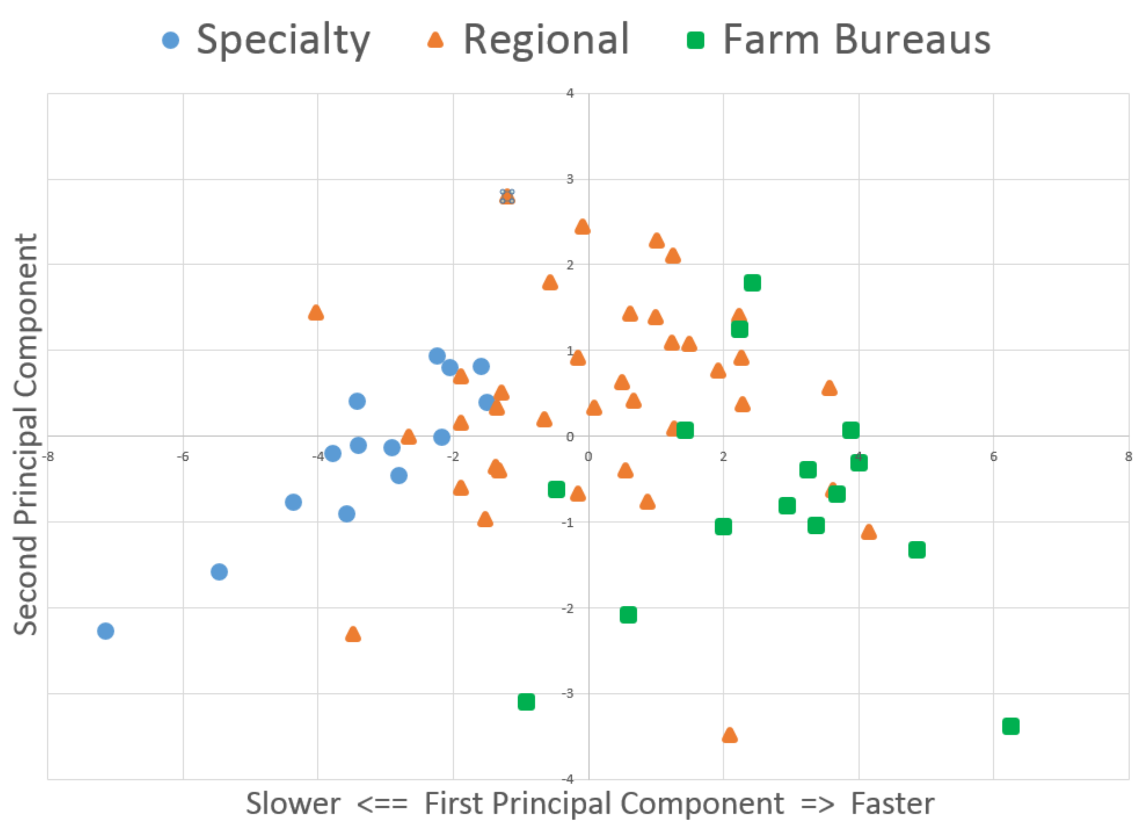
- Introduction
- Visualization of Multidimensional Data
- Cluster Analysis
- Dealing with Outliers and Noise
- Practical Considerations
  - Correlations between LOB
  - Identifying drivers of loss development

Practical Considerations

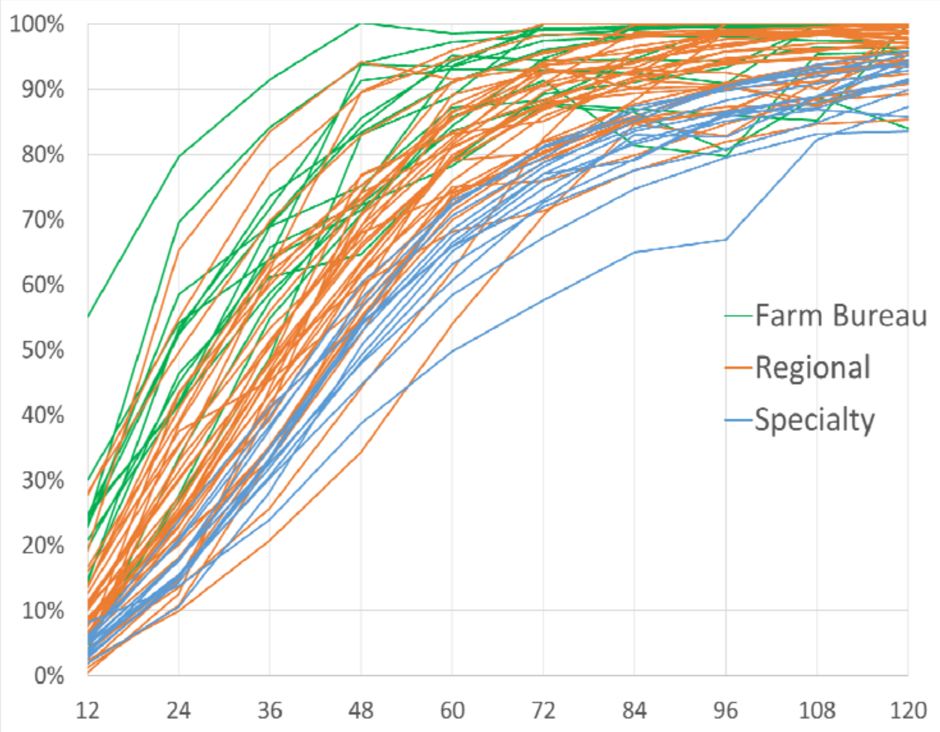
Visualization: Finding the Right Variables

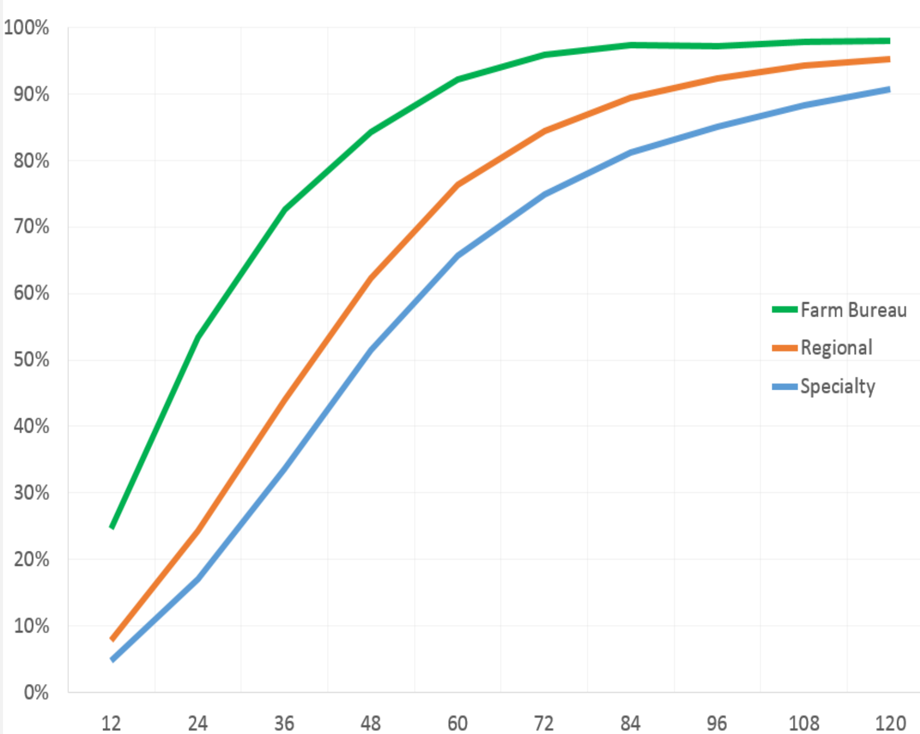
Munich RE

- Schedule P & SNL company profile
- GL paid development
  - 15 Farm bureaus
  - 14 Specialty
  - 37 Regional
- Loss data is from 2009 to 2019



Practical Considerations
Visualization: Finding the Right Variables





Conclusion
Key Takeaways

Munich RE

- 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

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. P. Coretto, C. Hennig (2017) **Robust improper maximum likelihood: tuning, computation, and a comparison with other methods for robust Gaussian clustering,** Journal of the American Statistical Association 111(516), pp. 1648--1659
5. C. Hennig, M.Meila, F. Murtagh, R.Rocci (2017) **Handbook of Cluster Analysis,** CRC Press
6. 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
7. 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
8. 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
9. J. VanderPlas, "**Python Data Science Handbook", **O'Reilly Media, http://shop.oreilly.com/product/0636920034919.do
10. CAS Schedule P data for Loss Reserving: http://www.casact.org/research/index.cfm?fa=loss_reserves_data

Thank you!

daveclark@munichreamerica.com
drangelova@munichreamerica.com

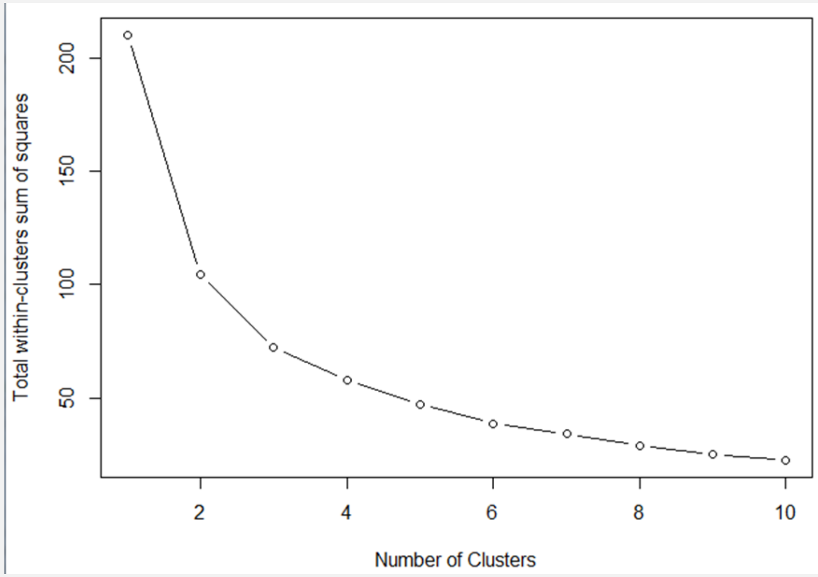
Munich RE

Cluster Analysis
K-means Algorithm

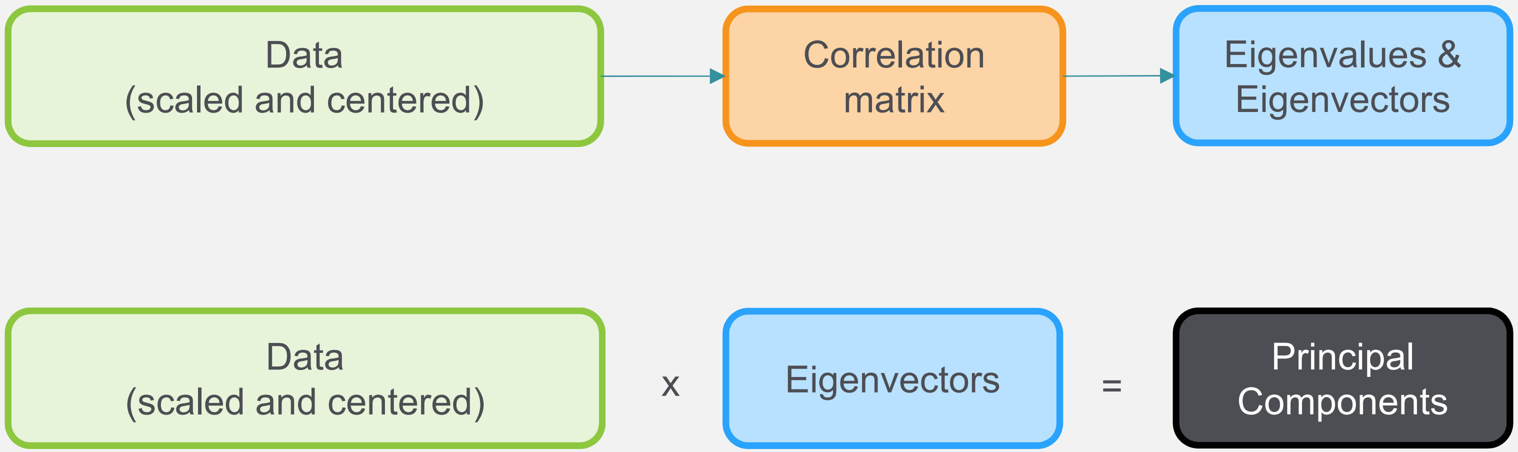
Munich RE

- 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

- No definitive answer for selecting K
  - Scree plot: locate the sharpest drop in within-cluster sum of squares



PCA
How to perform a PCA?



Clustering Methods Applied to Overlapping Groups

Munich RE

Fuzzy Clustering

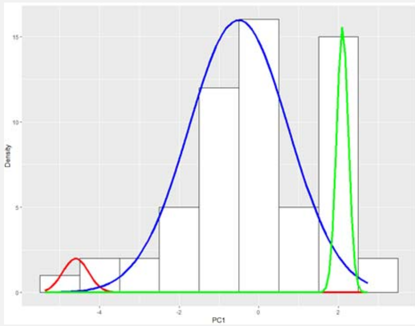
- Fuzzy clustering is an iterative process that optimizes a cost function (similar to K-means) and at each iteration recalculates a membership function.
- Fuzzy: $min: \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^{m} d_{ik}^{2}$ $where$ $u_{ij} = \frac{d_j^{-\frac{2}{m-1}}}{\sum_{k=1}^{c} d_k^{-\frac{2}{m-1}}}$
- K-means: $min: \sum_{i=1}^{n} \sum_{k=1}^{c} d_{ik}^{2}$
- d_{ik}^{2}: squared Euclidean distance
- m: controls the fuzziness (m>1, m→1 increases the crispiness of the cluster)
- u_{ik}: membership degree of the i-th object to the k-th cluster

Clustering Methods Applied to Overlapping Groups

Gaussian Mixture Models: 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:

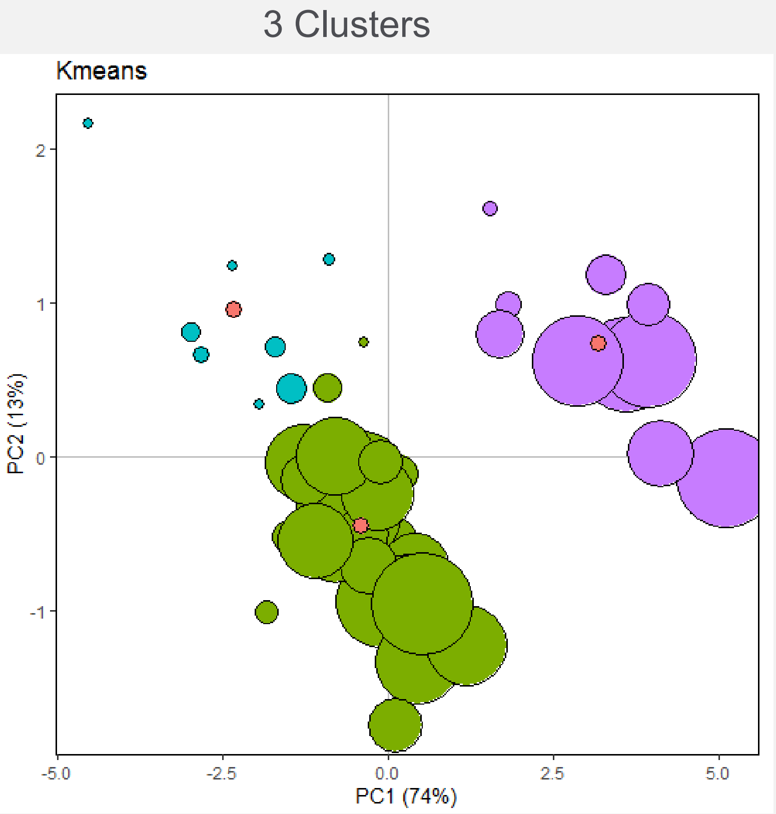
$$\tau_i(y_j) = \frac{\pi_i f_i(y_j)}{f(y_j)}$$

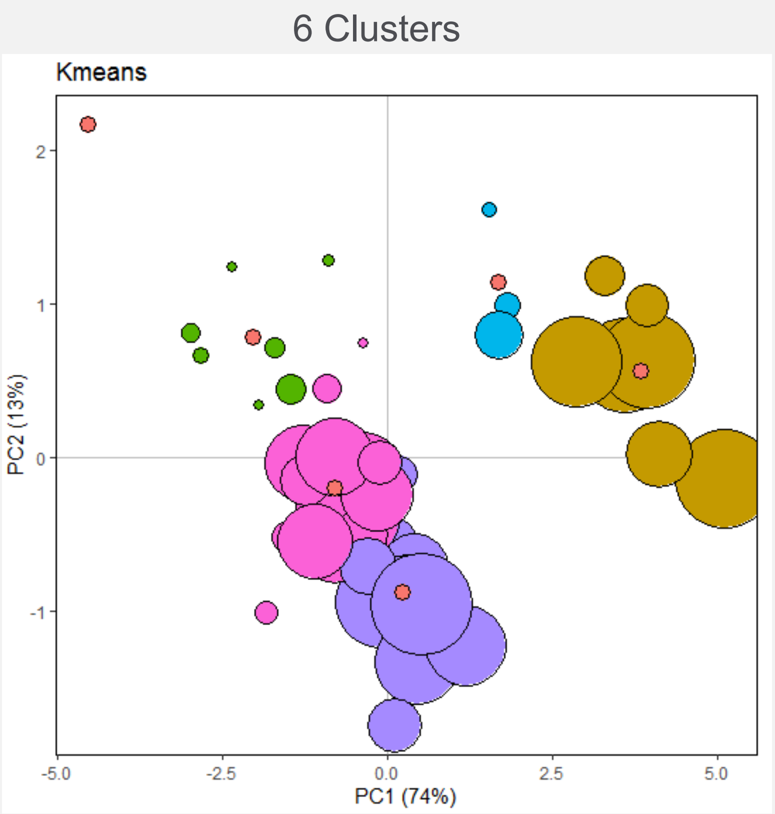
- If we assume that the data in the clusters is independent and normally distributed, we can use a Gaussian Mixture Model (GMM).

Outliers and Noise Points

Recognising Outliers: Large Number of Clusters

Munich RE



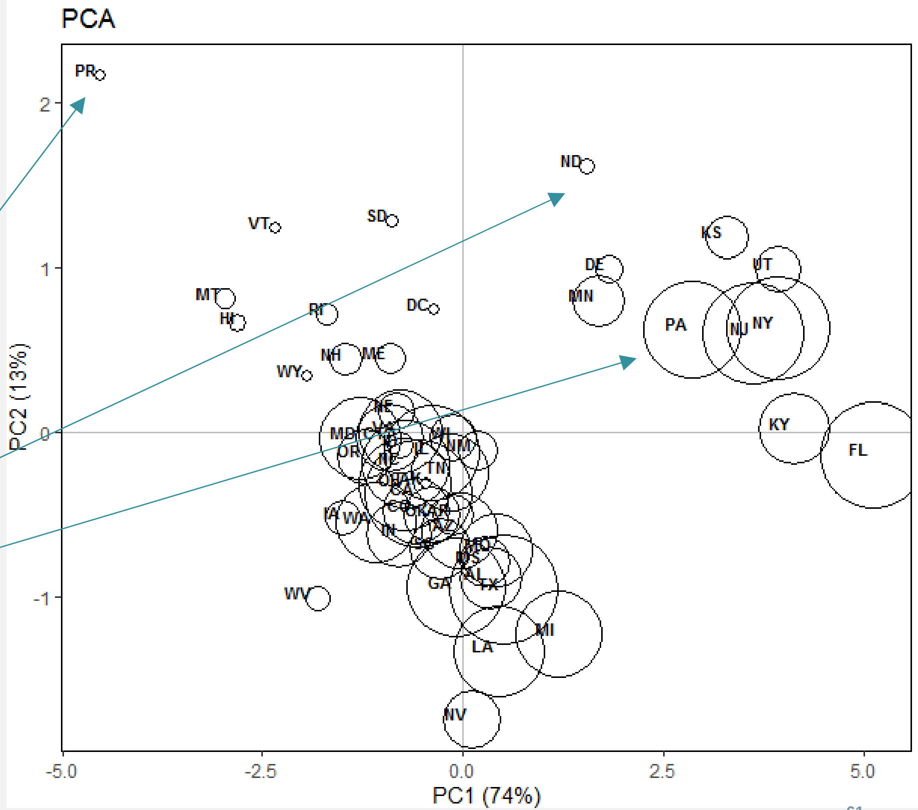


Outlier and Noise Points
Recognising Outliers: Fuzzy Clustering

Munich RE

- Outliers are "equally remote" to all clusters.

m=2	Fuzzy1	Fuzzy2	Fuzzy3
PR	34%	**45%**	21%
ND	29%	27%	**44%**
PA	12%	10%	**77%**



Practical Considerations
What Are the Drivers of Loss Development?

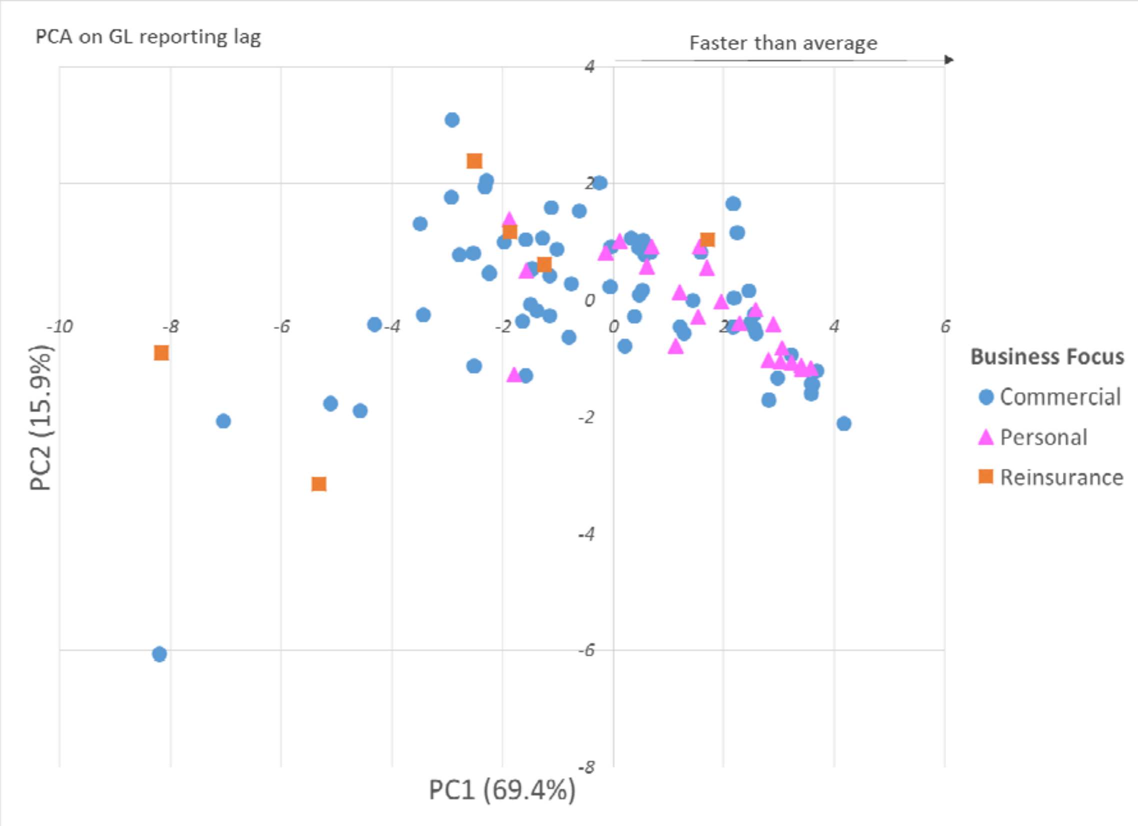
Munich RE

- Identify potential predictors
  - Business focus (Commercial, Personal, Reinsurance)
  - Ownership (Stock, Mutual, Others)
  - Distribution channel (Broker vs Non-Broker)
  - Geography (Regional vs National)

- Schedule P GL data & SNL company profile
  - Top 100 insurers by market share
  - Loss data is from 2008 to 2017

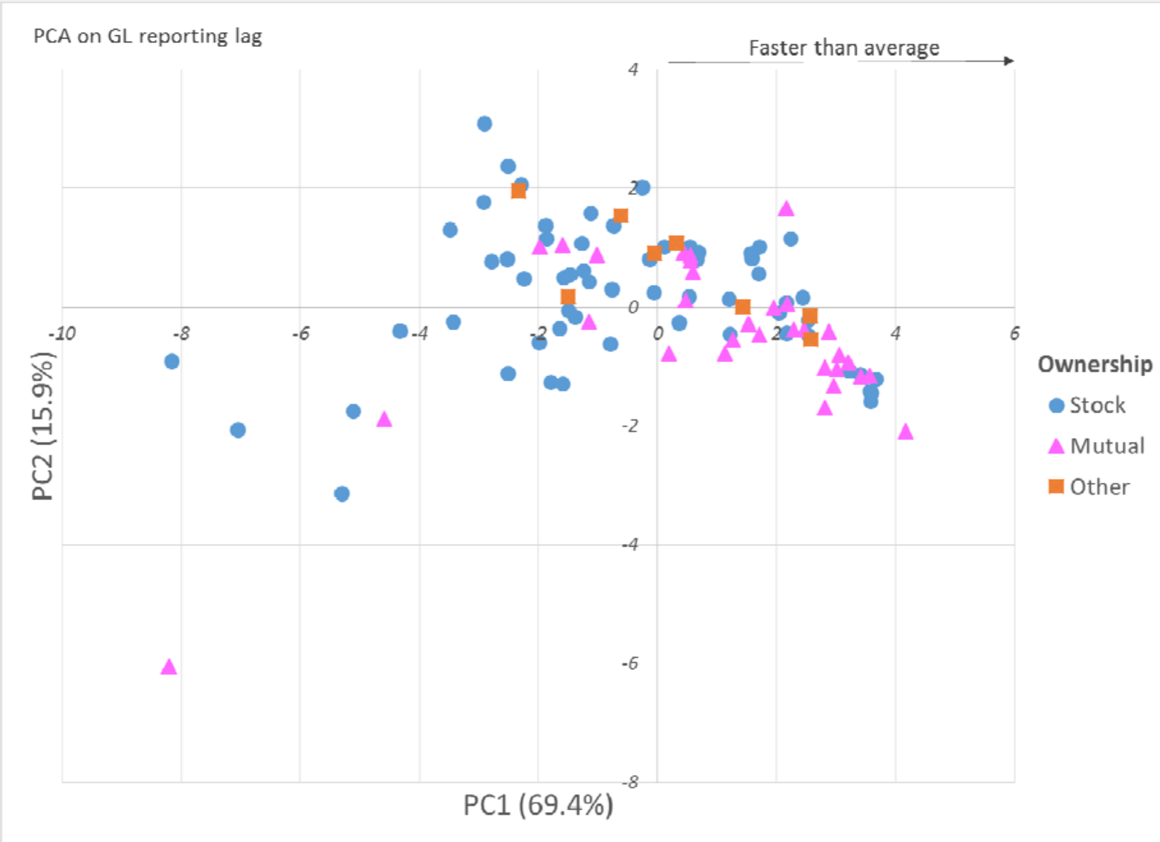
Practical Considerations
Visualization: Are the Explanatory Variables Logical?

Munich RE



Practical Considerations
Visualization: Are the Explanatory Variables Logical?

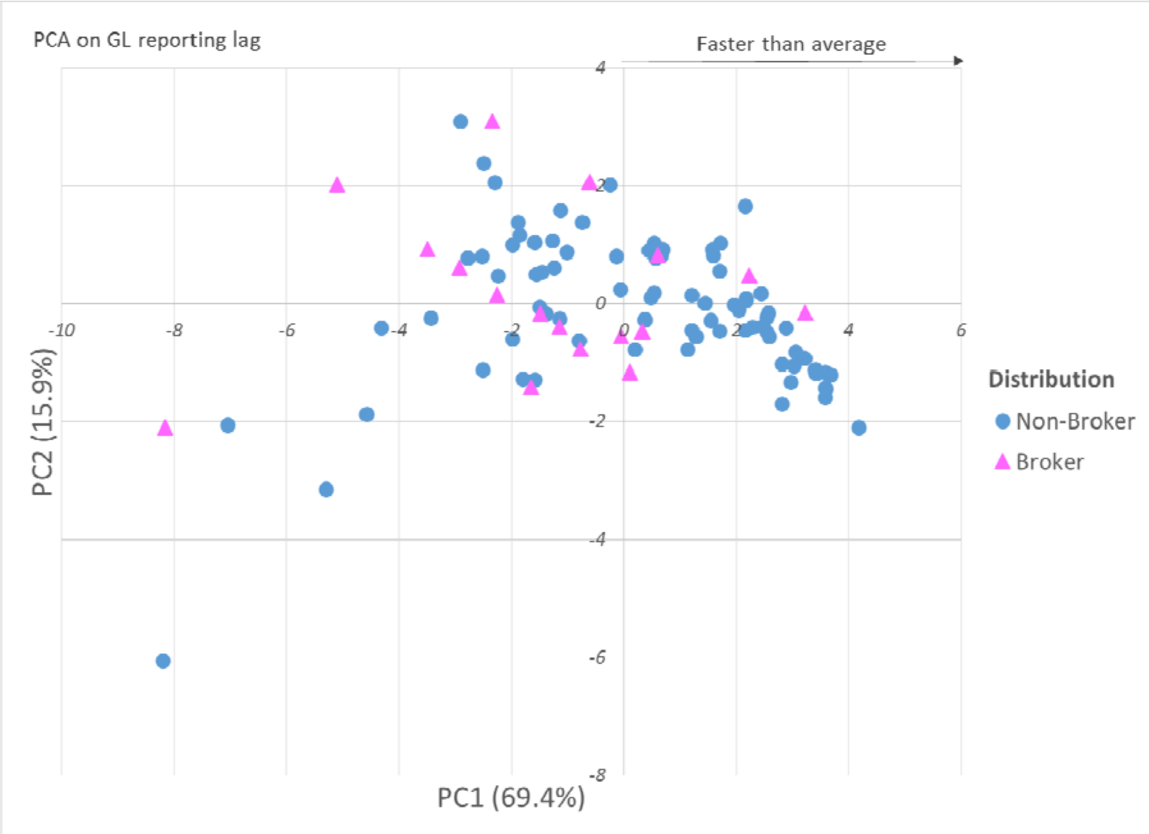
Munich RE



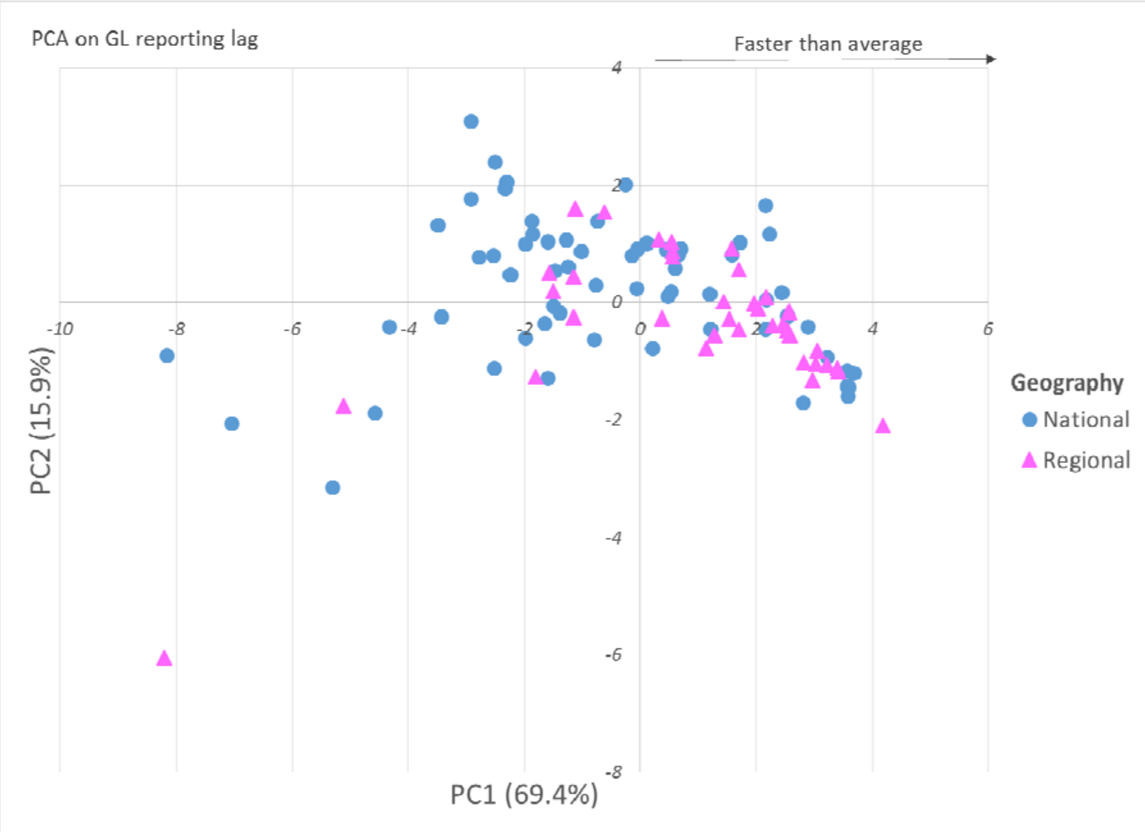
Practical Considerations

Visualization: Are the Explanatory Variables Logical?

Munich RE



Practical Considerations
Visualization: Are the Explanatory Variables Logical?



R Packages

Important R packages:

- Package “stats” *(kmeans, prcomp,…)* - https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
- Package “cluster” *(pam, fanny,…)* - https://cran.r-project.org/web/packages/cluster/cluster.pdf
- Package “factoextra” (get_eigenvalue, fviz_cluster,…) - https://cran.r-project.org/web/packages/factoextra/factoextra.pdf
- Package “ggplot2” - https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf
- Package “mclust”(*mclust*) - https://cran.r-project.org/web/packages/mclust/mclust.pdf