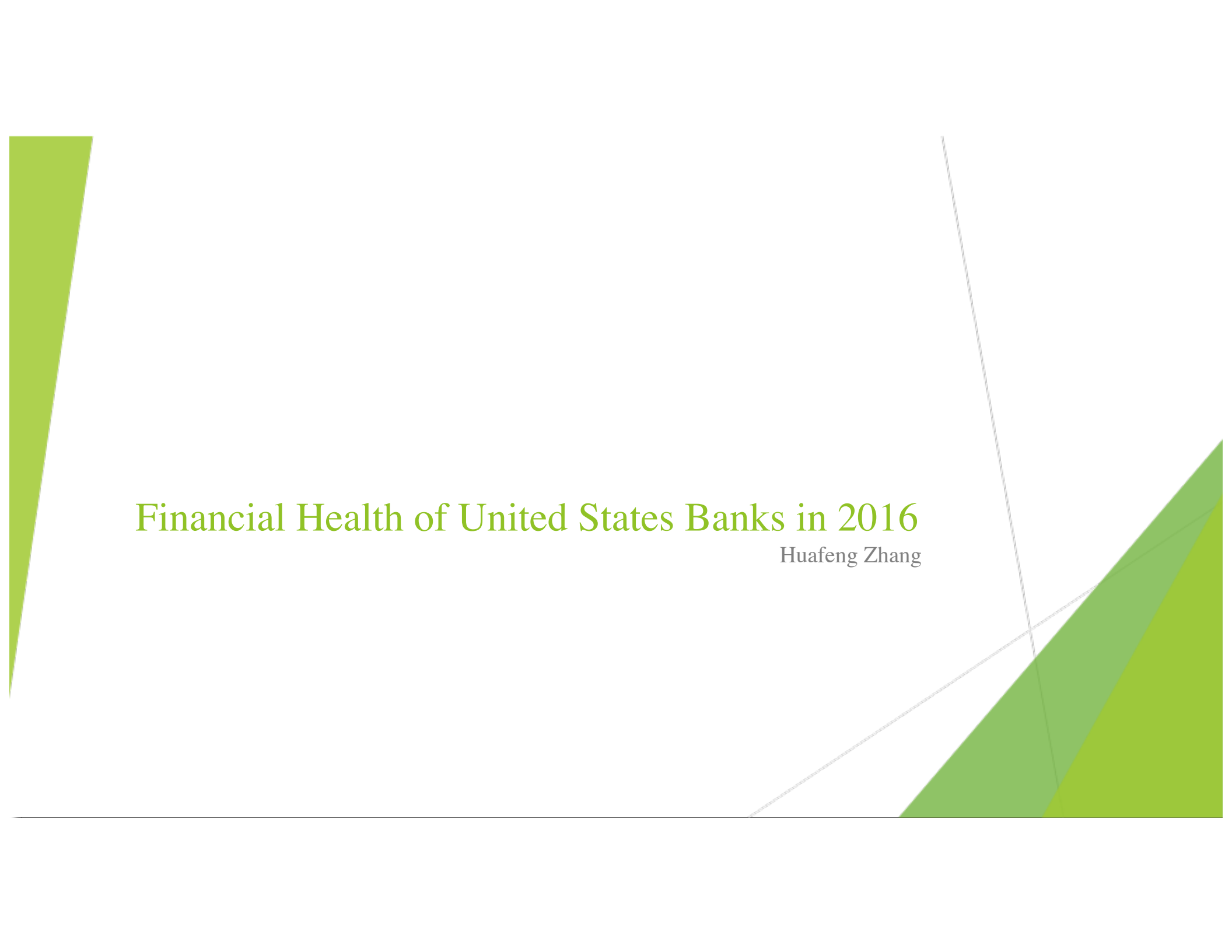




Financial Health of United States Banks in 2016

Huafeng Zhang

Contents

- Background
- Data
- Methods
- Results
- Conclusions



Background

- Banks play a vital role in keeping the economy going by attracting savings and granting loans. It is essential for economists as well as politicians to understand what banks do, and how healthy their financial systems are.
- A bank's financial health can be measured by its investment effectiveness, profitability, loan and debt health, capital stability, solvency.[1]



This study aims to understand financial health by performing a clustering analysis of all the banks in the United States based on their financial health in 2016.

Five indicators in measuring banks' financial health

(based on suggestions from a bank examiner)

- **Net interest margin (NIM)** measures a bank's investment effectiveness. A negative NIM suggests that the bank has not invested its funds efficiently.
- **Return on assets (ROA)** measures a bank's annual profitability. The higher ROA is, the more profitable the bank is.
- **Net charge-offs to loans (NCOL)** is high, the bank has "bad debt" or "poor quality loans" that are unlikely to be recovered by the bank.
- **Tier 1 capital ratio (CCLR)** measures a bank's capital stability. The higher the CCLR is, the higher is the likelihood of the bank withstanding negative shocks to its balance sheet.
- **Total risk-based capital ratio (TRBCR)** measures how well a bank is prepared to protect its investors and depositors from riskiness

Data

- Data are all collected from the FDIC website.[2]
- These data sets consist of 5,922 banks in the United States and sixteen variables.
- There were 85 missing values in one of data set, all from the columns of the five variables of interest. The missing values are removed, which led to 48 banks being removed from the data set.
- Three banks with extreme high TRBCR and one bank with extremely big ROA are removed from the data set in an effort to avoid the clustering of 5,870 banks distorted by the four small banks.
- All values of the five variables of interest are scaled to avoid bias caused by the difference in magnitudes of the variables.

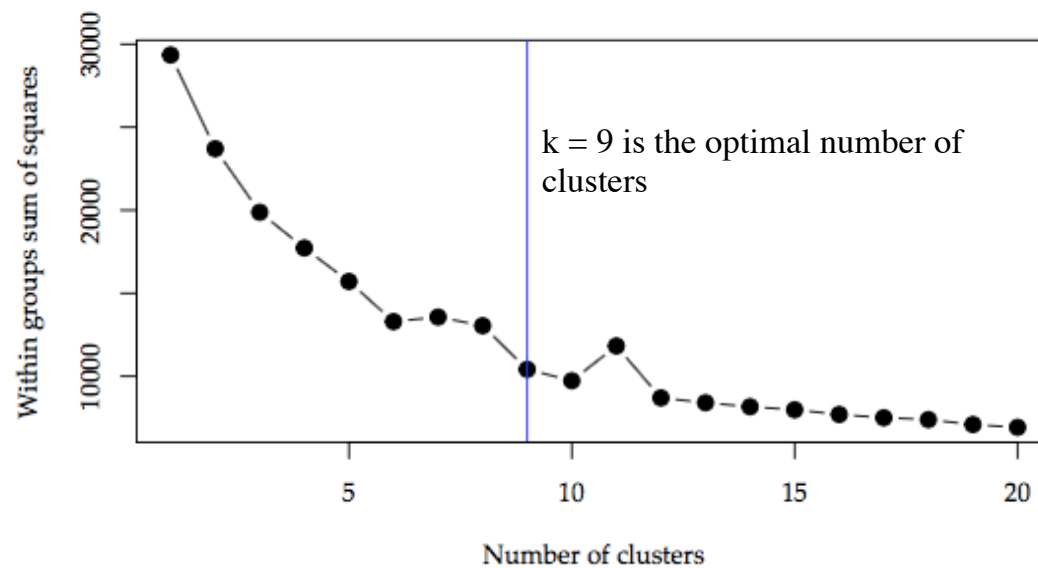
Methods: k-means clustering

Reasons for using k-means clustering

- **Familiarity:** k-means clustering method is widely used in the financial field so that it will be easier for most of the audiences to understand this study.
- **Simplicity:** Given 5,870 banks in the data set, a k-means clustering method is a good starting point for exploratory analysis.
- **Flexibility:** it is easier and faster to change to use other clustering techniques depending on the results of the k-means clustering compared to using other clustering methods, such as hierarchical clustering methods.

Methods: k-means clustering

Find the optimal number of clusters by the elbow method



Methods: k-means clustering

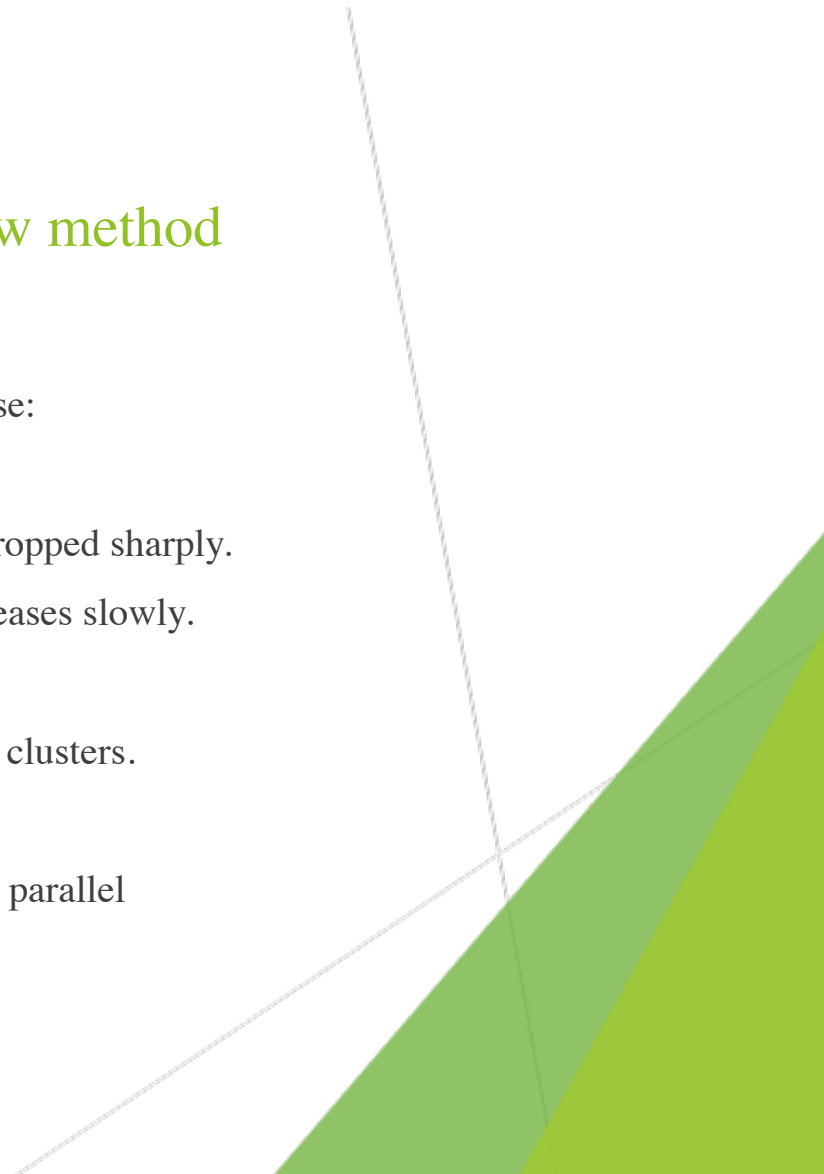
Find the optimal number of clusters by the elbow method

The figure suggests that $k = 9$ is an optimal choice for clustering, because:

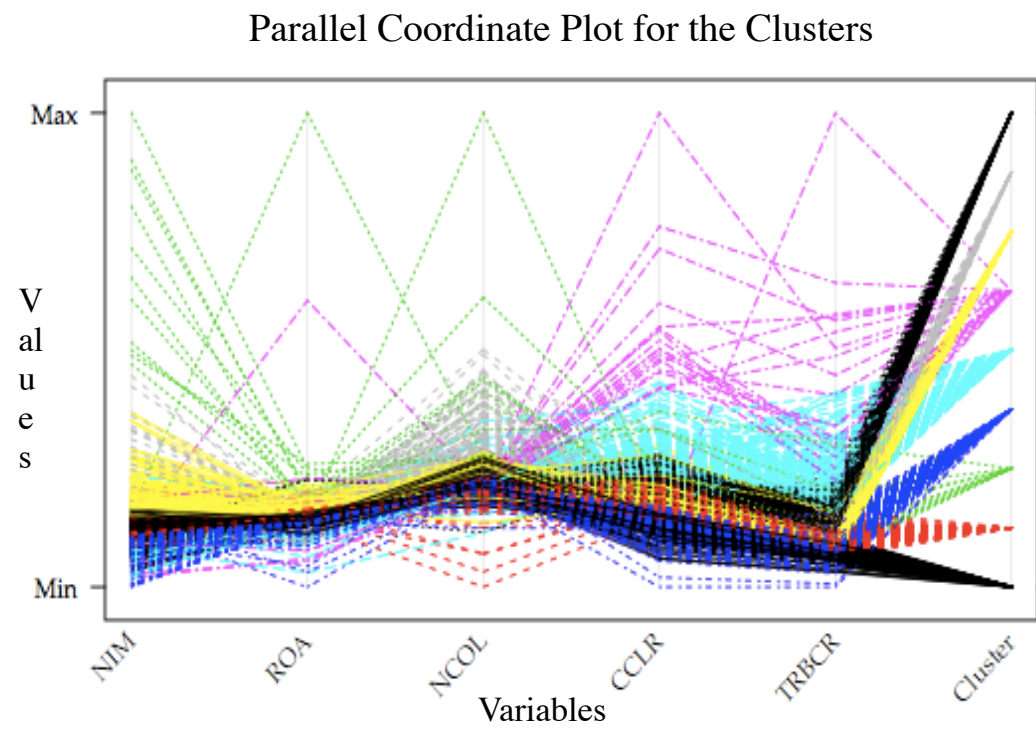
- When there are less than 9, the variance explained by the clusters dropped sharply.
- When there are more than 9, the sum of the variance explained decreases slowly.

This indicates that $k = 9$ should be the optimal choice for the number of clusters.

- Although $k = 6$ also appears to be a reasonable choice, the resulting parallel coordinates plot showed no clear pattern in the clustering.



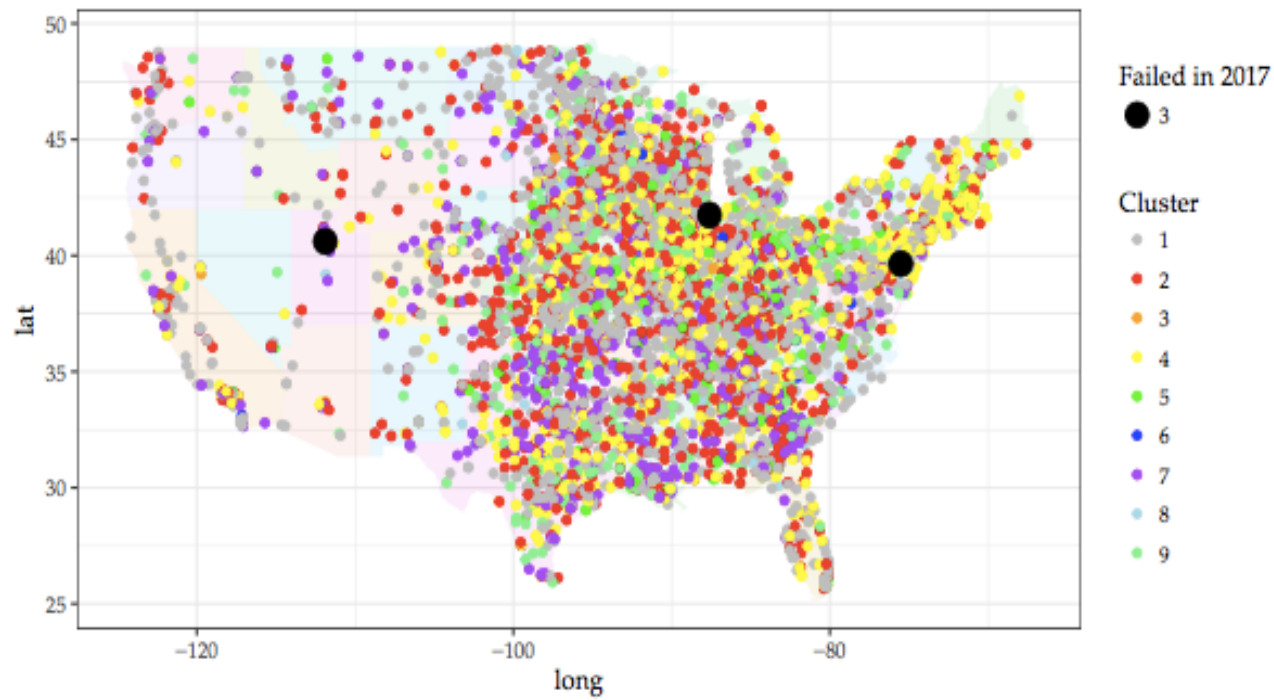
Characteristics of the clusters



Interpretation of the clusters

- **Cluster 1** (black solid line): “normal in making investments and profits, but not reliable or stable.”
- **Cluster 2** (red line): similar to Cluster 1, but with lower mean net charge-offs to loans.
- **Cluster 3** (green line): “most profitable but also most risky.”
- **Cluster 4** (blue line): “most unhealthy financially.”
- **Cluster 5** (light blue line): “less effective, profitable, but stable, reliable.”
- **Cluster 6** (pink line): “most healthy financially.”
- **Cluster 7** (yellow line): “on track, but potentially growing risky.”
- **Cluster 8** (grey line): “least profitable.”
- **Cluster 9** (black dash line): “on track.”

Distribution of the banks by cluster



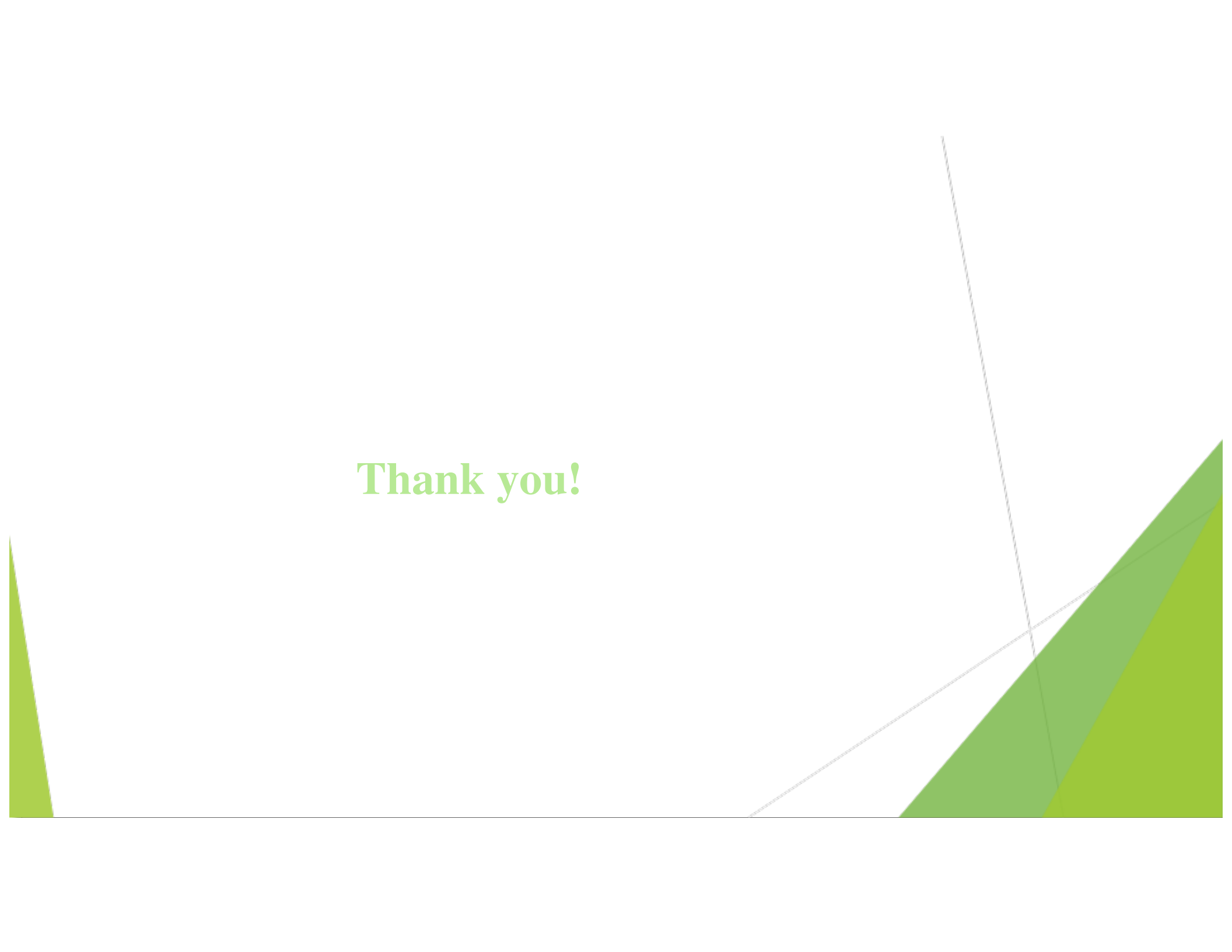
Summary

- There are nine groups of banks in terms of their financial health.
- 1,224 banks are found be in the cluster that represented the most unhealthy institutions.
- The banks that failed in the first quarter of 2017 are all found in the most unhealthy cluster.
- Only 19 banks are in the cluster described as most healthy.

References

- [1] Necas, S. (2016). Financial health of a commercial insurance company and its coherences. Trendy Ekonomiky a Managementu, 10(25), 22-32. doi:<http://dx.doi.org.proxybz.lib.montana.edu/10.13164/trends.2016.25>
- [2] Each depositor insured to at least \$250,000 per insured bank. (2017). Retrieved April 7, 2017, from <https://www5.fdic.gov/sdi/main.asp?formname=customddownload>
- [3] Ashcraft, A. (2005). Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks. The American Economic Review, 95(5), 1712-1730. Retrieved from <http://www.jstor.org.proxybz.lib>
- [4] An Act (2010). Retrieved April 11, 2017, from https://www.gpo.gov/fdsys/pkg/PLAW-111publ203/html/PLA_111publ203.htm
- [5] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [6] Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5
- [7] Taiyun Wei and Viliam Simko (2016). corrplot: Visualization of a Correlation Matrix. R package version 0.77. <https://CRAN.R-project.org/package=corrplot>

Thank you!

An abstract geometric design featuring several overlapping triangles in shades of green and yellow. A thin, light gray line runs diagonally across the composition, intersecting the triangles. The background is a solid light gray.