MSDS 6372 : Principal Component Analysis

***Glass Identification : Refractive Index (RI) Analysis***

**1. Introduction**

“*Now you see it, now you don't. Glass is a bit of a riddle. It's hard enough to protect us, but it shatters with incredible ease. It's made from opaque sand, yet it's completely transparent. And, perhaps most surprisingly of all, it behaves like a solid material... but it's also a sort of weird liquid in disguise! You can find glass wherever you look: most rooms in your home will have a glass window and, if not that, perhaps a glass mirror... or a glass lightbulb. Glass is one of the world's oldest and most versatile human-created materials.”* Chris Woodford September 17, 2016, <http://www.explainthatstuff.com/glass.html>

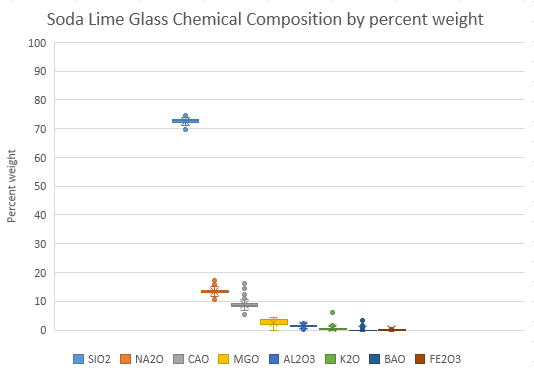
Soda Lime glass is the most common type of glass used for windows and glass containers. Its primary components are Silica (Quartz sand), Sodium Oxide (Soda Ash), and Calcium Oxide (Lime). (Soda-Lime Glass n.d.) Soda Lime glass dates as far back as 2000 B.C. It is made in batch and continuous casting processes with refining temperatures as high as . (IMI-NFG course on Processing Glass 2015)When the mixture cools, it completely transforms into a different structure that is not quite solid. It becomes a material that scientists refer to as an **amorphous solid**, a cross between a solid and a liquid. Usually other minerals are added into the liquid glass to change its working temperature, hardness, chemical properties, or color based on the purpose of the glass.

Refractive index (RI) determines how much light is bent or refracted when entering or exiting a transparent material or in our case glass. The purpose of this observational study is to understand the different chemical compositions of glass and determine if refractive index can be predicted based on composition. This study looks at seven chemical compositions of Soda Lime Silica Glass using a dataset from The University of California, Irvine which contains 214 observations of 7 different glass types composed of 8 minerals by % weight. (UCI Machine Learning Repository n.d.)

**2. Exploratory Analysis**

The dataset includes compositions for Building and Vehicle windows (float and non-float processed) containers, tableware, and headlamps. (Float Glass process n.d.) SAS proc Means was used to summarize the data. Note that Silica, Soda, and Lime make up 95% of the composition by weight



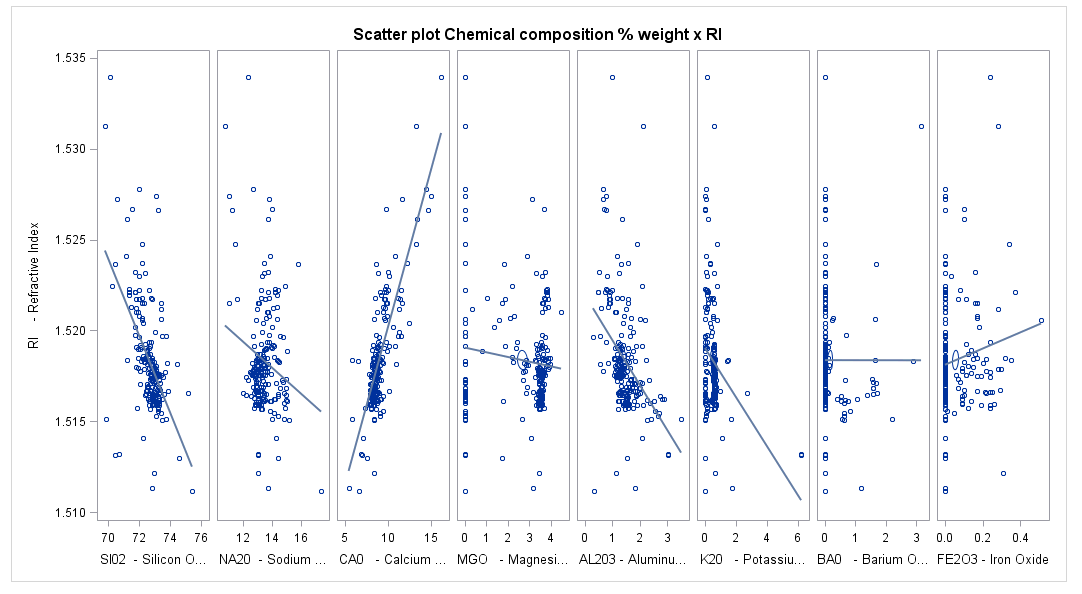
The box and whisker plot shows the variability of each mineral in the UCI glass dataset as a percent of weight.



Should we include this?

How to interpret it?

Scatter plots show a strong neg ative slope for Silica and a strong positive relationship for Calcium Oxide on Refractive index.



**3. Validation of assumptions**

For purposes of this study, it is assumed that the refractive index of soda lime glass can be predicted using statistical tools such as principal components and regression. Although there are only 8 minerals in the dataset, the objective of this project is to use principal components to reduce the number of independent variables.

**4. Principal Components Analysis (PCA)**

A Principal components analysis was performed using the SAS Princomp procedure (SAS princomp n.d.)

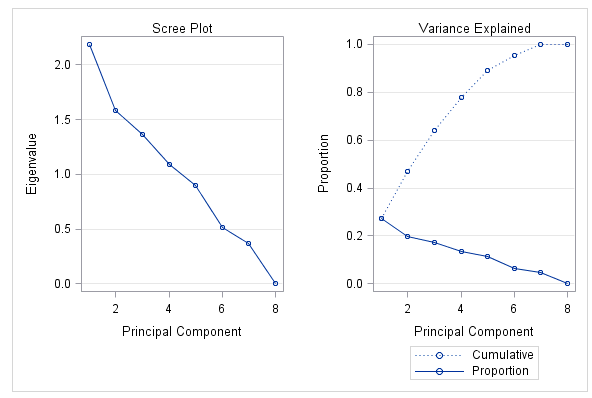
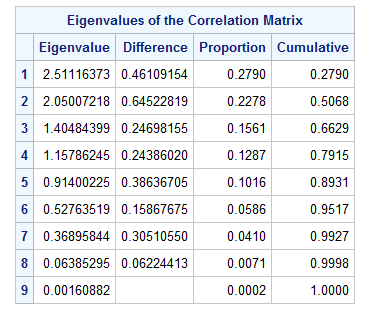
Principal components were calculated using SAS PROC Princomp with all independent variables (Mineral composition) and the dependent variable RI (Refractive Index)

title1 "Full PCA with dependent RI included";

**proc** **princomp** data = glass out=glassPC\_all plots=all;

var RI SIO2 NA2O CAO MGO AL2O3 K2O BAO FE2O3;

**run**;



Choosing the factors with Eigenvalues >= 1 gives 4 principal components. The scree plot and variance explained plots show the first two factors explaining 50% of the total variance with four factors explaining 79% of the variance.

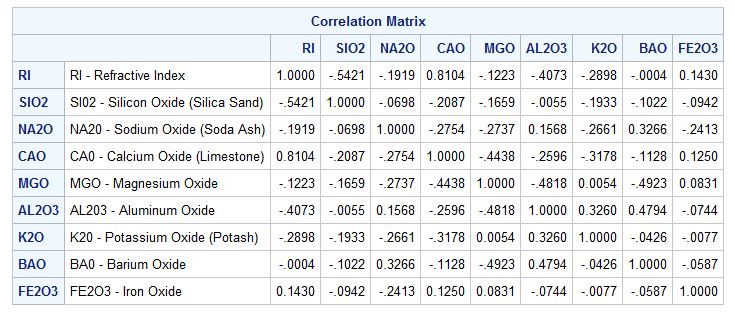
A second analysis was done with just the four main principal components:

title1 "Principal Components N=4 ";

**proc** **princomp** data = glass out=glassPC plots = all N=**4 ;**

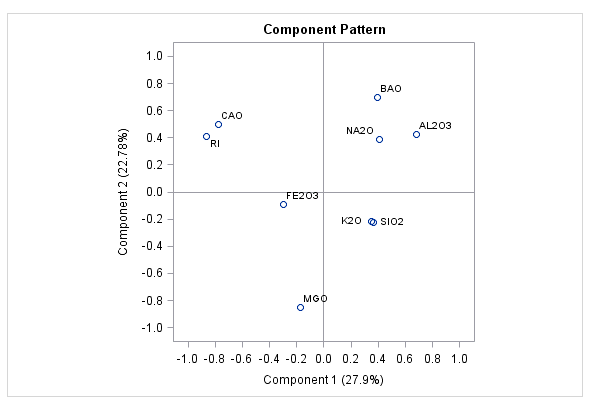
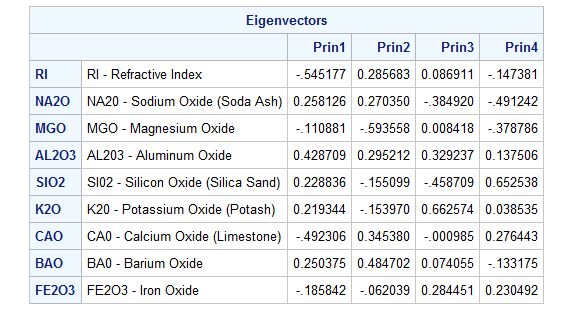
var RI NA2O MGO AL2O3 SIO2 K2O CAO BAO FE2O3; **run**;

The correlation matrix shows several high correlations. Calcium Oxide has a high positive correlation and silicon oxide has a high negative correlation to RI.



The Eigenvector table shows the variables contributing to each principle component.

I thought the plot would match the values in the eigenvector table, they have the right direction but not quite the same value, not sure how to interpret this. And thought the grouping meant something but I can’t find where I saw that.



**Regression Analysis of the Principal Components**

Regression analysis of the principal components….

**6. Conclusion**

**Statistical Conclusion** :

**APPENDIX**

**SAS CODE**

|  |
| --- |
| **data** glass;  infile 'glass.csv' dlm=',' firstobs=**2** dsd;  input ID RI NA20 MG0 AL203 SI02 K20 CA0 BA0 FE203 Type ;  **run**;  **proc** **print** data=glass ; **run**; |

**References I think these should be put in as citations and show up in the bibliography.**

<http://www.explainthatstuff.com/glass.html>

<http://hypertextbook.com/facts/2002/SaiLee.shtml>

Use References Insert citation to add new citations then right click the bibliography and select update

\*\*\*\*\*\*\*\*\*\*\*\*

# Bibliography

n.d. *Float Glass process.* http://www.pilkington.com/pilkington-information/about+pilkington/education/float+process/step+by+step.htm.

n.d. *History of Glass.* https://en.wikipedia.org/wiki/History\_of\_glass.

2015. *IMI-NFG course on Processing Glass.* http://www.lehigh.edu/imi/teched/GlassProcess/Lectures/Lecture02\_Hubert\_Melting%20and%20fining%20processes.pdf.

n.d. "SAS princomp." *OKState.edu.* https://www.okstate.edu/sas/v8/saspdf/stat/chap52.pdf.

n.d. *Soda-Lime Glass.* https://en.wikipedia.org/wiki/Soda-lime\_glass.

n.d. *UCI Machine Learning Repository.* http://archive.ics.uci.edu/ml/datasets/Glass+Identification.