

VISUALIZATION AS AUDIT EVIDENCE

Qi Liu, Heejae Lee, Zamil S. Alzamil
RADAR In-Person Meeting
October 2nd, 2017

OVERVIEW

Visualizations are created based on Government Financial Statement to perform the following audit procedures:

Risk Assessment Procedures

- Compare different state
 - Cluster analysis to select states with similar features to compare
 - Dynamic visualization for comparison among states over years
- Single state analysis
 - Fluctuation analysis
 - Regression Analysis
 - Correlation analysis to select appropriate variables to perform regression analysis
 - 2D and 3D visualization for regression analysis

Substantive Testing Procedures

- Income Tax over population – analysis by year to test completeness and valuation assertions

CLUSTERING METHODOLOGY

All of the clustering methodologies in this analysis have been conducted using *K*-means & *Hierarchical* clustering.

The variables used the analysis are as follow:

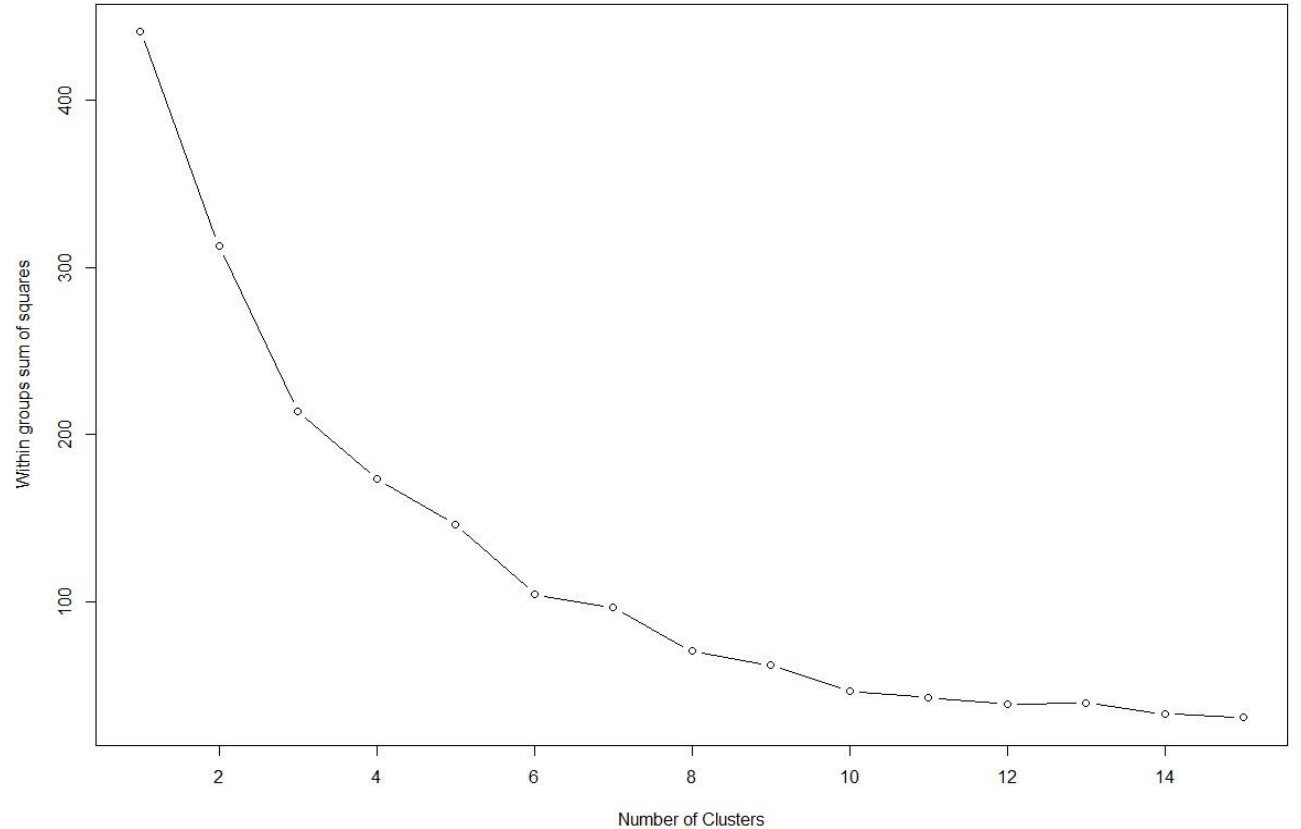
1. Total General Fund Revenues.
2. Excess (Deficiency) of Revenues over Expenditures.
3. Total Operating Expenses.
4. Education Expenses.
5. Net Change in Fund Balance.
6. General Fund Total Other Financing Sources.
7. General Fund Transfers to Other Funds.
8. General Fund Transfers from Other Funds.
9. Pension Expense.

K-means clustering: The aim of the *K*-means clustering algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized. Meaning that each observation (record in a table) belongs to the cluster with the nearest mean. ^{1, 2}

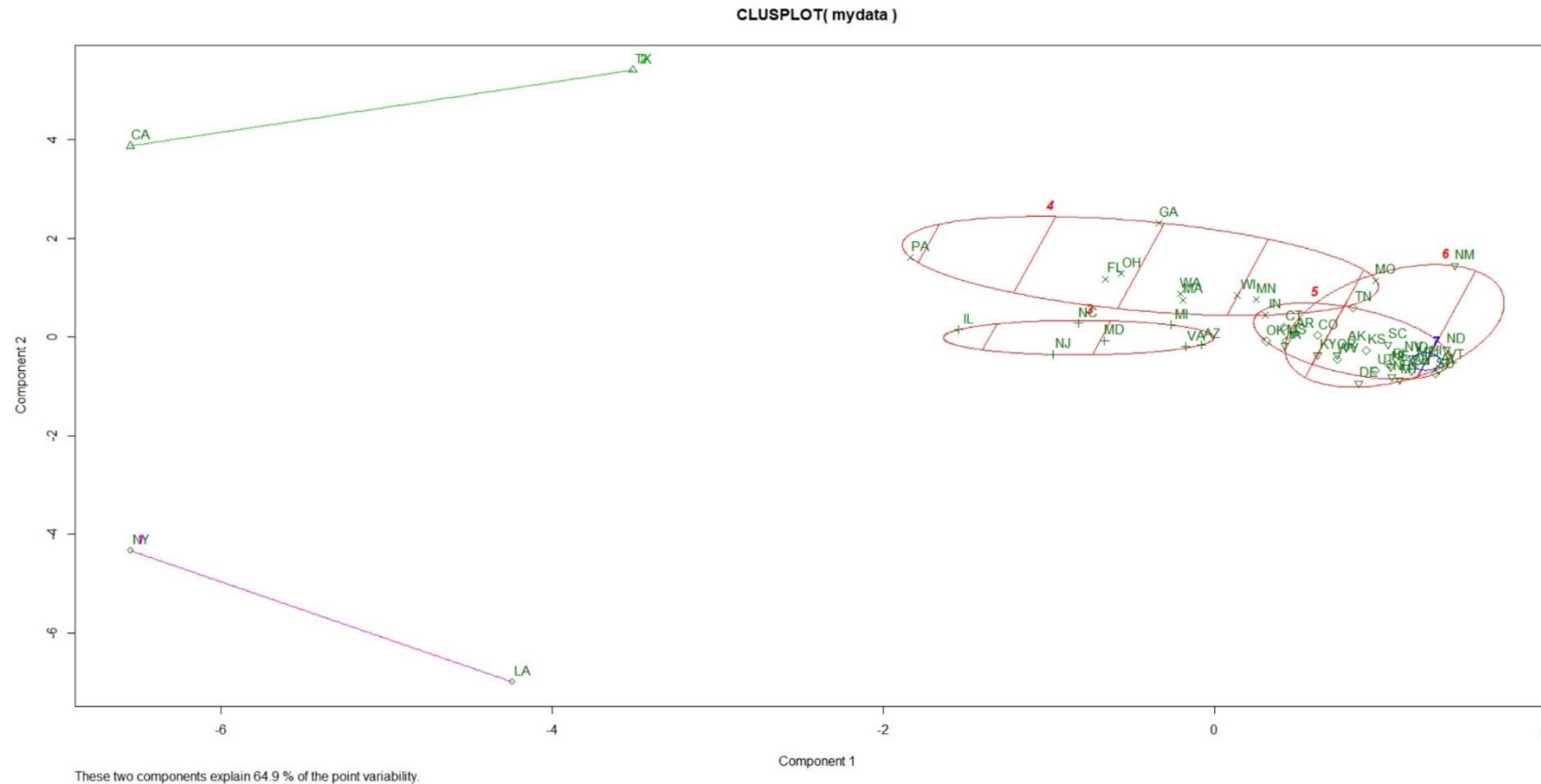
1) K-MEANS CLUSTERING - AVERAGE

The figure on the right side shows that **6 clusters** would be a good fit.

This method is called “the within clusters sum of squares” or the Elbow method which is a method of interpretation and validation of consistency of points within each cluster. It is performed by computing the within clusters sum of squares designed to help determine the optimal number of clusters.



2) CLUSTER PLOT: 2D REPRESENTATION OF CLUSTERS SOLUTION (K-MEANS “6 CLUSTERS”)



CONT'D

As shown from the previous figure, the states are clustered as follow:

Cluster	Members
#1	AR, AZ, DE, IA, ID, KY, MT, ND, NE, NH, NM, OR, SC, VA, VT, WY
#2	AK, AL, CO, CT, IN, KS, ME, MN, MO, MS, NV, OK, RI, SD, TN, UT, WV
#3	LA, NY
#4	CA, TX
#5	HI
#6	FL, GA, IL, MA, MD, MI, NC, NJ, OH, PA, WA, WI

3) HIERARCHICAL CLUSTERING

In data mining and statistics, **hierarchical clustering** (also called **hierarchical cluster analysis** or HCA) is a method of **cluster analysis** which seeks to build a **hierarchy of clusters**.

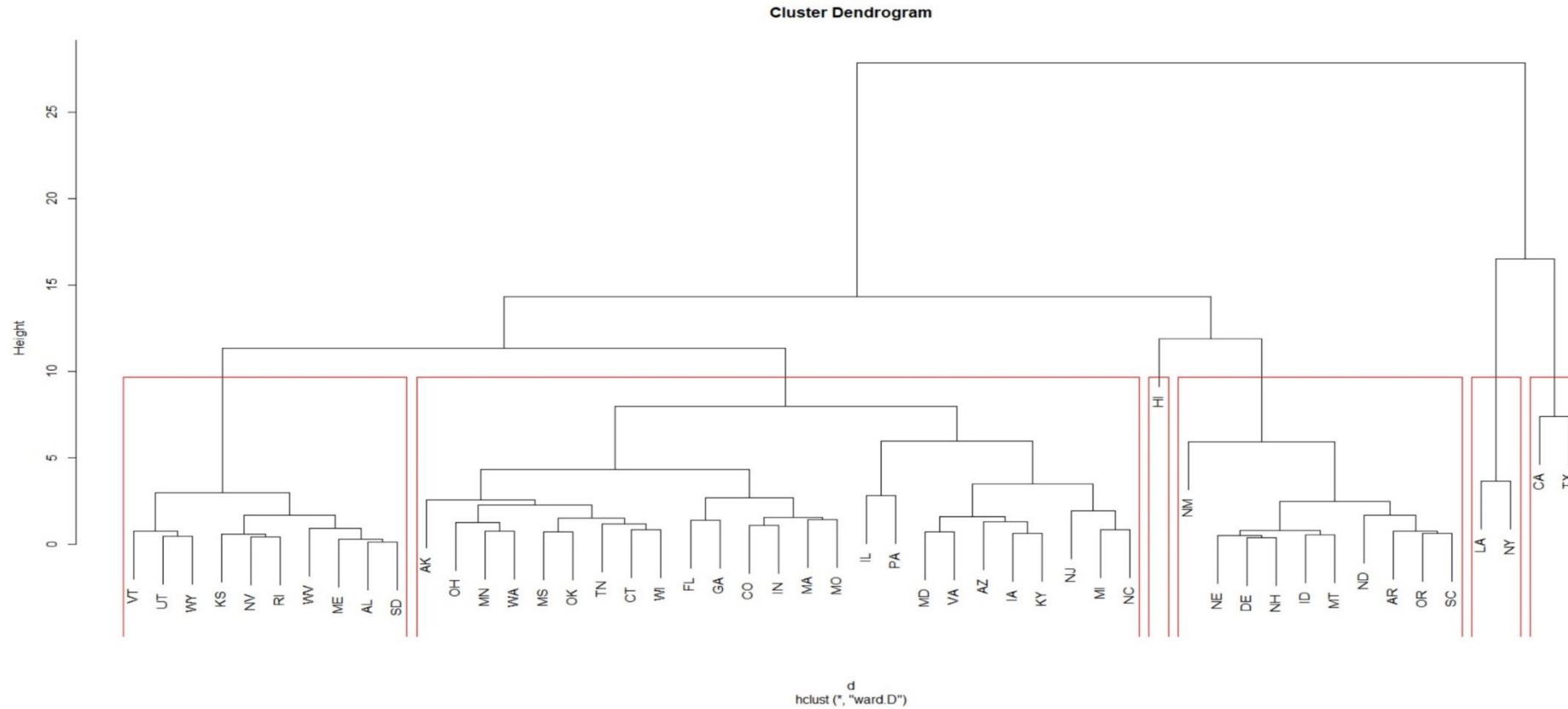
We used ward.D method with Euclidean as a measure of distance.

Ward.D is an **Agglomerative** which means a "bottom up" approach:

- Each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Ward.D is a method that finds minimum variance by minimizing the total within-cluster variance.

5) CLUSTER PLOT: A DENDROGRAM REPRESENTATION OF CLUSTERS SOLUTION (CONT'D)



CONT'D

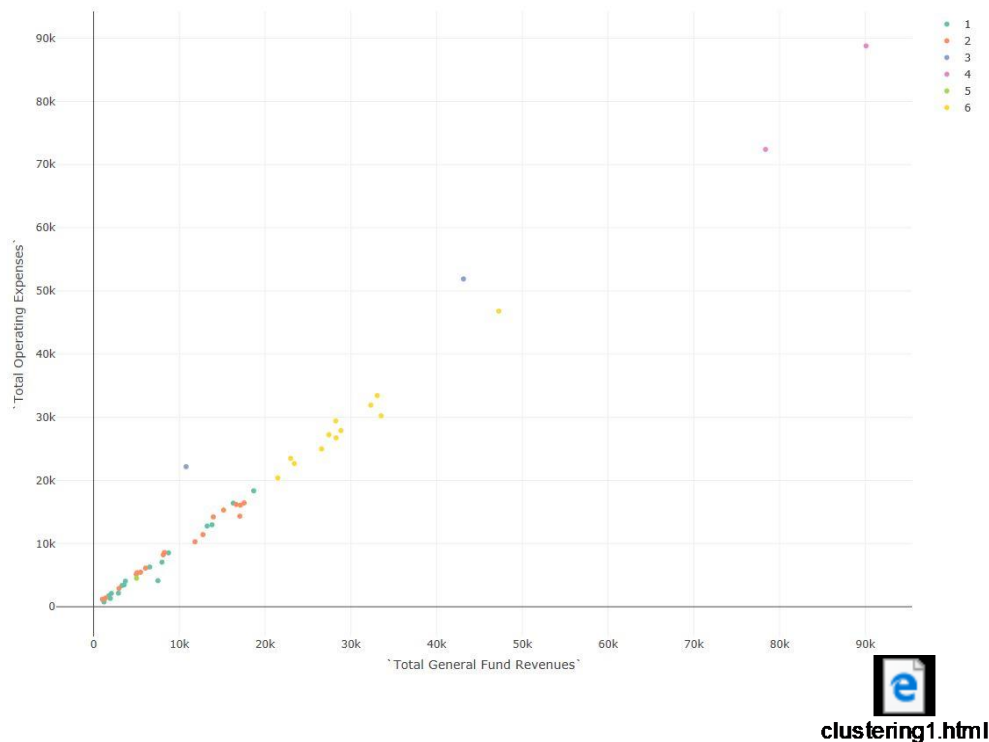
As shown from the previous figure, the states are clustered (grouped) as follow:

Cluster	Members
#1	VT, UT, WY, KS, NV, RI, WV, ME, AL, SD
#2	AK, OH, MN, WA, MS, OK, TN, CT, WI, FL, GA, CO, IN, MA, MO, IL, PA, MD, VA, AZ, IA, KY, NJ, MI, NC
#3	HI
#4	NM, NE, DE, NH, ID, MT, ND, AR, OR, SC
#5	LA, NY
#6	CA, TX

COMPARING CLUSTERING RESULTS

Cluster	K-mean	Hierarchical
#1	AR, DE, ID, MT, ND, NE, NH, NM, OR, SC, AZ, IA, KY, VA, VT, WY	AR, DE, ID, MT, ND, NE, NH, NM, OR, SC
#2	AK, KS, ME, NV, RI, SD, UT, WV, AL, CO, CT, IN, MN, MO, MS, OK, SD, TN,	AL, KS, ME, NV, RI, SD, UT, VT, WY, WV,
#3	LA, NY	LA, NY
#4	CA, TX	CA, TX
#5	HI	HI
#6	FL, GA, IL, MA, MD, MI, NC, NJ, OH, PA, WA, WI	FL, GA, IL, MA, MD, MI, NC, NJ, OH, PA, WA, WI, AK, MN, MS, OK, TN, CT, CO, IN, MO, VA, AZ, IA, KY,

RISK ASSESSMENT PROCEDURE — COMPARING DIFFERENT STATES

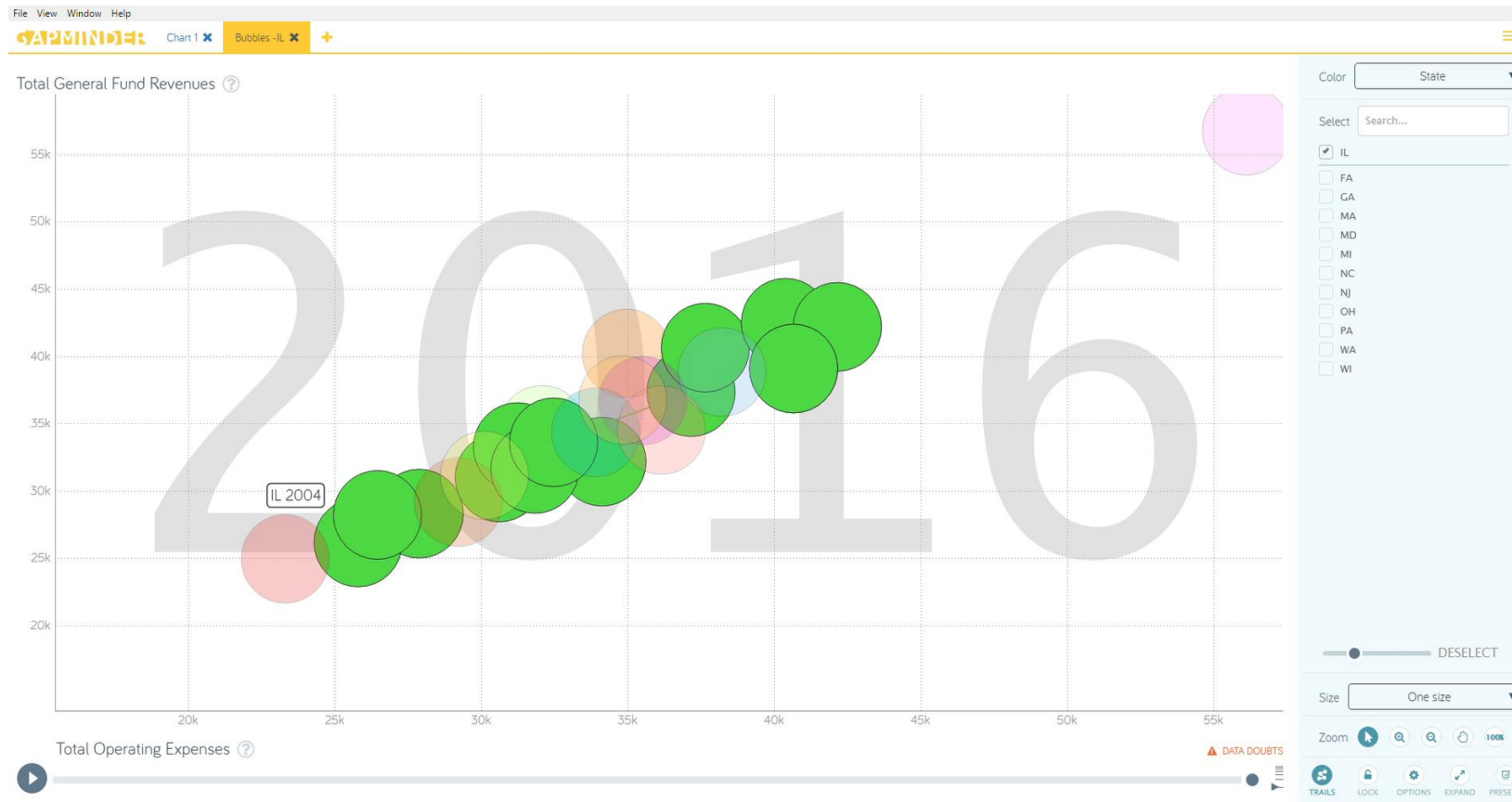


Comparing the relationship between Total General Funds Revenue and Total Operating Expenses in all clusters.

Different clusters have different numbers in Total General Funds Revenue and Total Operating Expenses.

RISK ASSESSMENT PROCEDURE – COMPARING DIFFERENT STATES

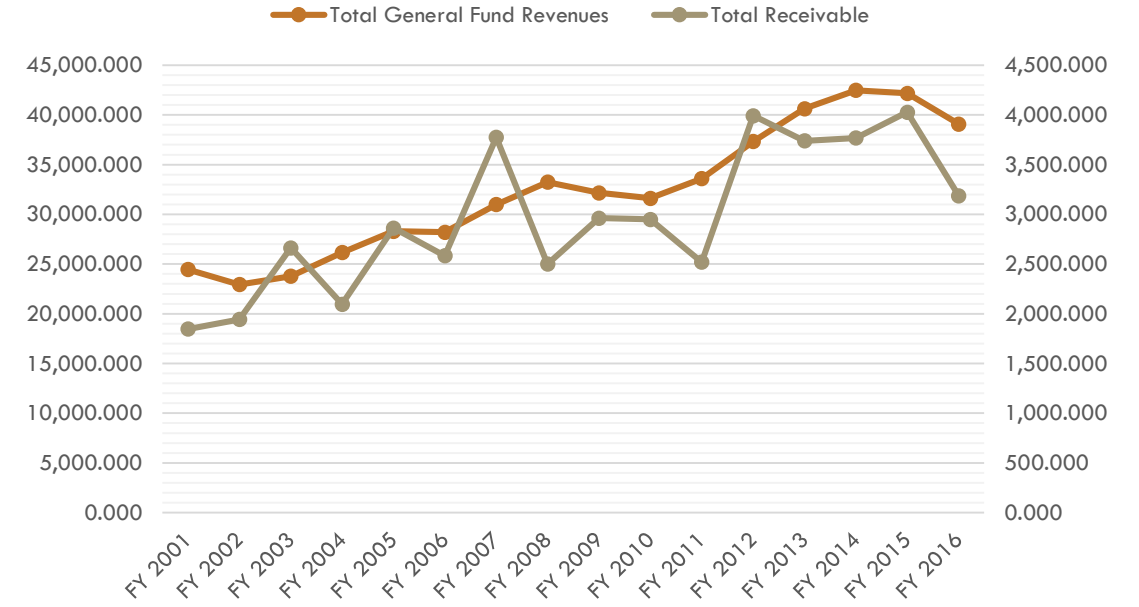
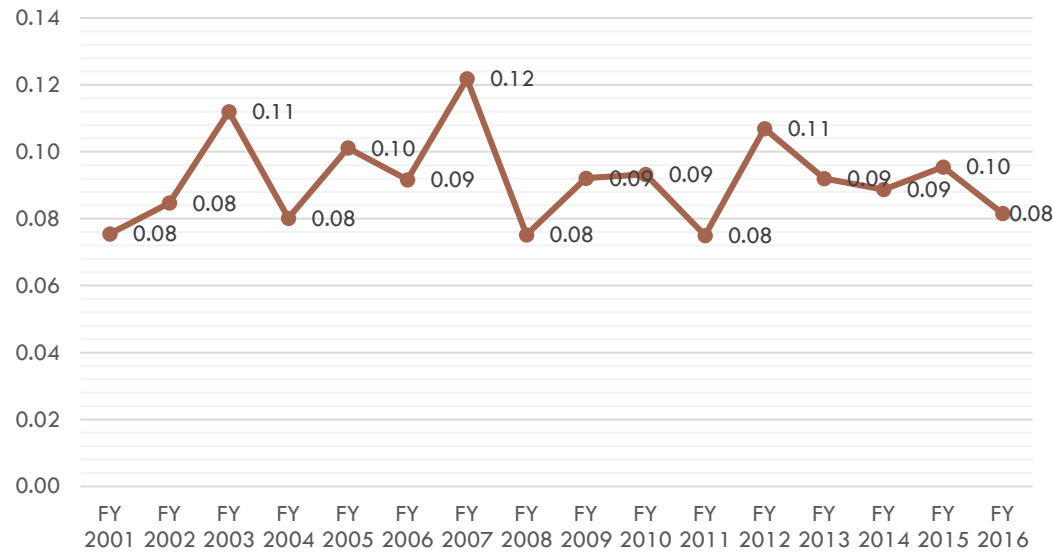
Comparing the change of Total General Funds Revenue and Total Operating Expenses over years in cluster 6 (Illinois is used as target state).



RISK ASSESSMENT PROCEDURE – SINGLE STATE

FLUCTUATION ANALYSIS

Total Receivable/Total Revenue Ratio



Ratio	Total Receivable/Total Revenues Ratio	(Million Dollars)	Total Revenues	Total Receivables
average	0.091685	average	32320.54	2963.314
Standard Deviation	0.013581	Standard Deviation	6525.375399	723.139986

RISK ASSESSMENT PROCEDURE — SINGLE STATE CORRELATION ANALYSIS

Show the relationships among different variables

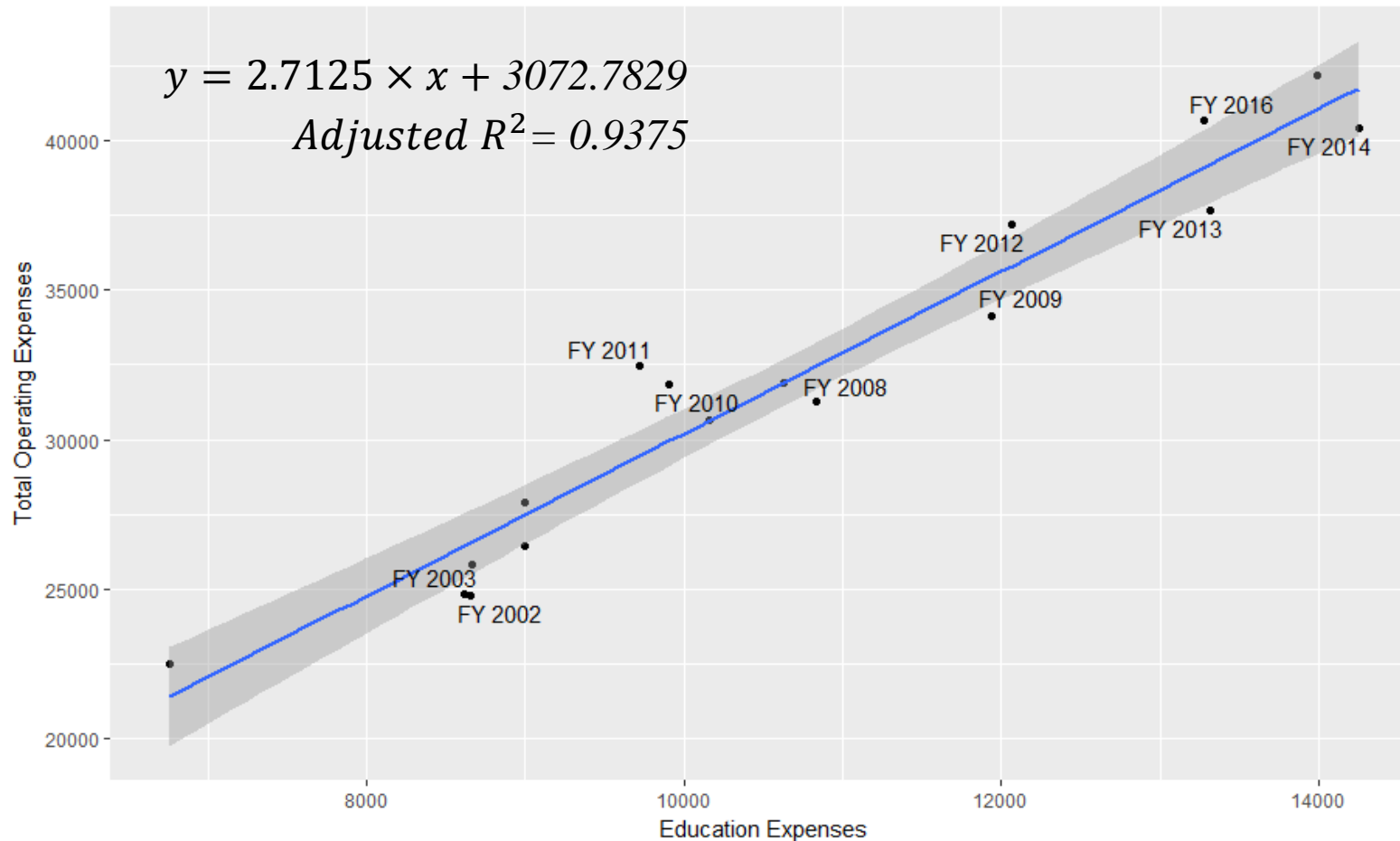
Provide information in selecting variables for regression models

Average of the years were used from

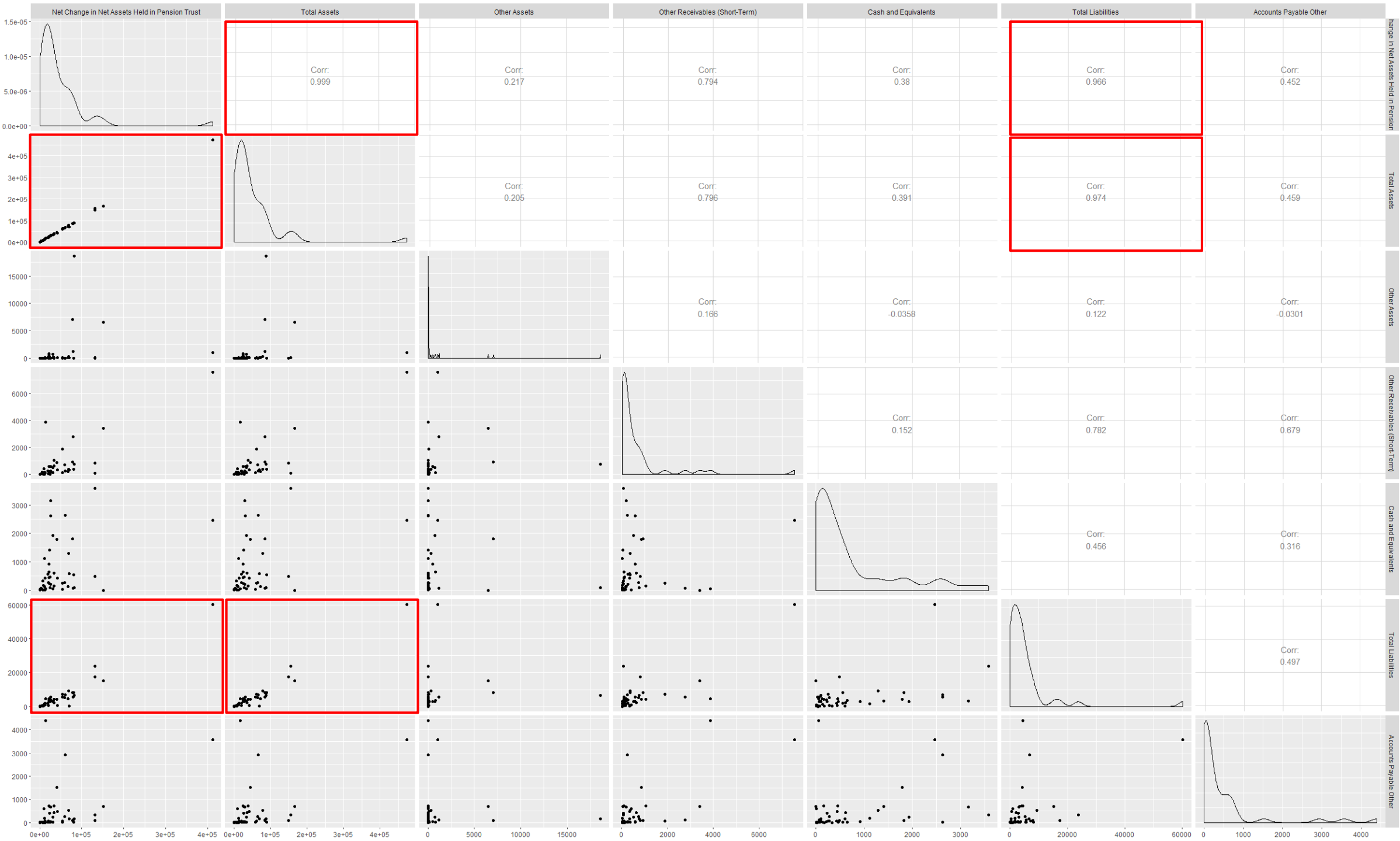
- General Fund Income Statement (FY 2000-FY 2016)
- Pension Fund Income Statement (FY 2004-FY 2016)
- Pension Fund Balance Sheet (FY 2004-FY 2016)

RISK ASSESSMENT PROCEDURE — SINGLE STATE

REGRESSION ANALYSIS



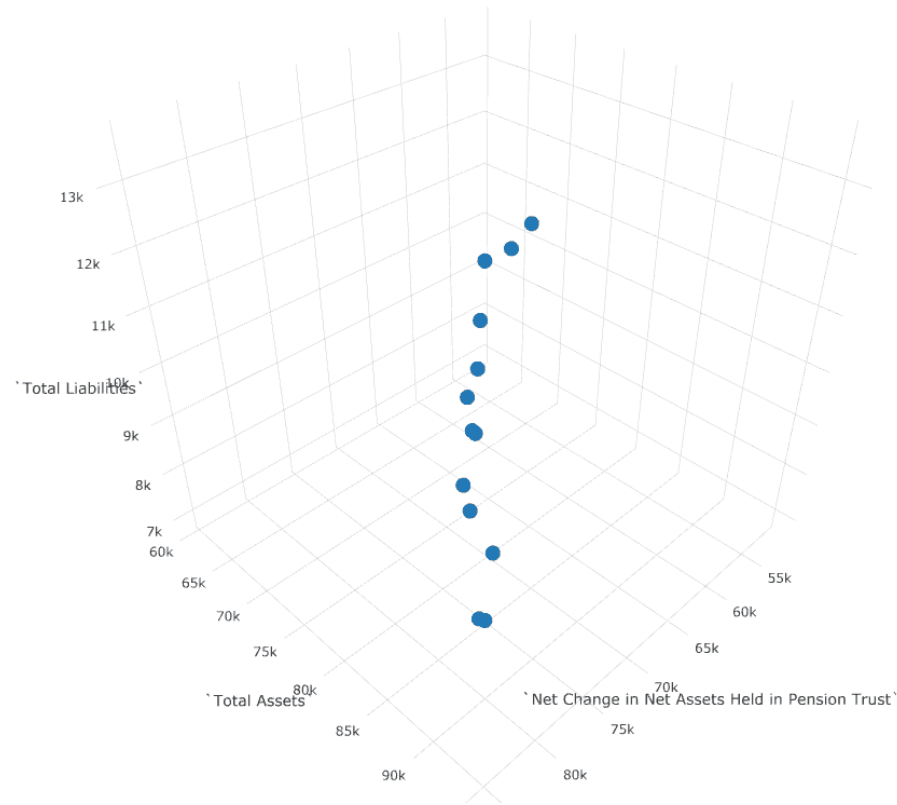
Pension Fund Balance Sheet



Risk Assessment Procedure – Single State

3D Visualization

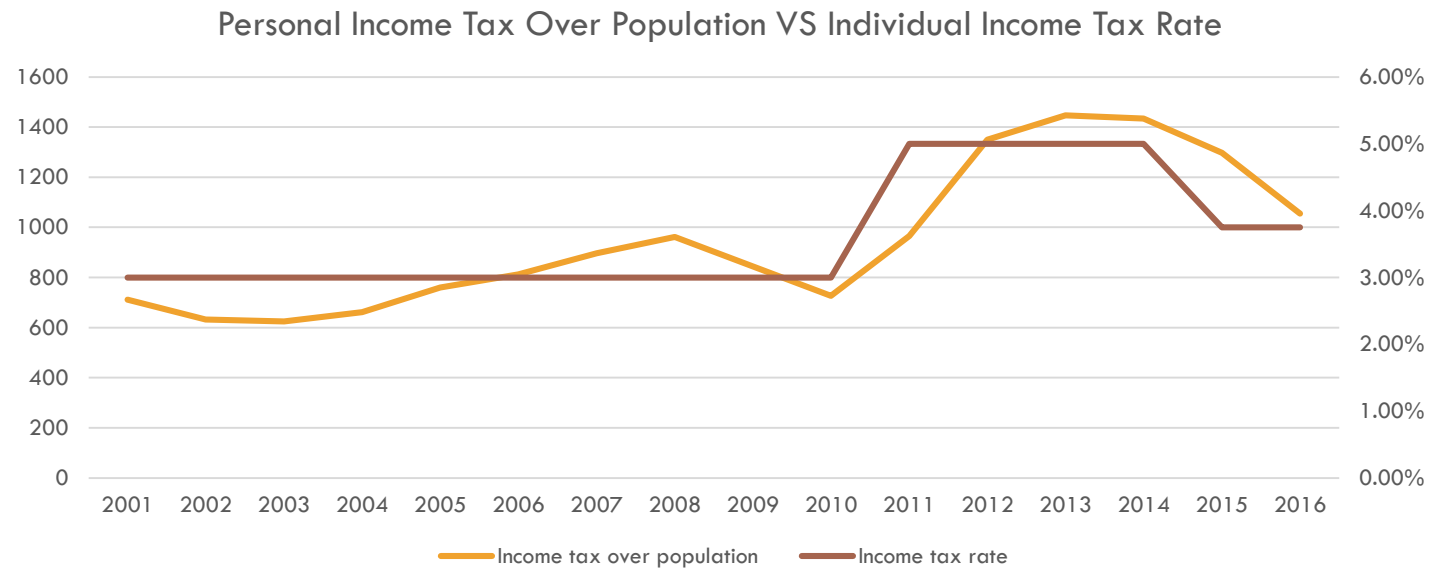
00000000



SUBSTANTIVE PROCEDURES

Receivables/revenue:

- Income Tax over population – analysis by year to test completeness and valuation assertions



FUTURE WORK

- Collecting local government data from Illinois Comptroller website to perform more substantive testing.
- Continue to work with the two governmental audit experts to prepare more visualizations for the substantive procedures with the detailed municipal data we just collected. Interview with the two experts will be scheduled after the in-person meeting.
- Work with another board member to develop a visualized audit program.