Predicting Investment in Renewable Energy for 2020

Table of Contents

Objective

Renewable Energy Investments of each US State for 2020

Chevron is looking for renewable energy businesses, initiative, and start-ups to potentially invest in. The greater the investment in the state, the more likely Chevron is to find collaborators and companies of interest. Finding collaborators becomes trickier every year as the national investment for clean energy has greatly increased in the United States.

01.

Data Cleaning and Processing

Deleting NA Values

2797	**2795**	BDFDB	US	2016	202797	United States		
2798	**2796**	BDPRP	US	2016	37327	United States		
2799	**2797**	BFFDB	US	2016	2295111	United States		
2800	**2798**	BFPRP	US	2016	404308	United States		
2801	**2799**	CLPRB	US	2016	14538027	United States		
2802	**2800**	CLPRK	US	2016	19.96	United States		
2803	**2801**	CLPRP	US	2016	728364	United States		
2804	**2802**	COPRK	US	2016	5.722	United States		
2805	**2803**	EMFDB	US					
2806	**2804**	ENPRP	US					
4550	**4548**	NGMPK	X3	2017	1.15	Federal Offshore - Gulf of Mexico		
4551	**4549**	NGMPP	X3	2017	1060453	Federal Offshore - Gulf of Mexico		
4552	**4550**	PAPRB	X3	2017	3511644	Federal Offshore - Gulf of Mexico		
4553	**4551**	PAPRP	X3	2017	613602	Federal Offshore - Gulf of Mexico		
4554	**4552**	TEPRB	X3	2017	4731270	Federal Offshore - Gulf of Mexico		
4555	**4553**	COPRK	X5	2017	5.723	Federal Offshore - Pacific		
4556	**4554**	PAPRB	X5	2017	32701	Federal Offshore - Pacific		

Adjusting Layout

	A	B	C	D	E	F	G	H	I	J	K
1		State	Year	BDFDB	BDPRP	BFFDB	BFPRP	CLPRB	CLPRK	CLPRP	COPRK
2	1	Alabama	2015	1933	356	1933	356	331420	25.122	13193	5.717
3	2	Alabama	2016	1906	351	1906	351	247632	25.68	9643	5.722
4	3	Alabama	2017	1585	292	1585	292	326748	25.407	12861	5.723
5	4	Alabama	2018	1652	304	1652	304	370533	25.065	14783	5.706
6	5	Alabama	2019	1494	275	1494	275	350506	24.816	14124	5.698
7	6	Alaska	2015	21	4	21	4	17747	15.073	1177	5.717
8	7	Alaska	2016	27	5	27	5	13942	14.957	932	5.722
9	8	Alaska	2017	29	5	29	5	14365	14.978	959	5.723
10	9	Alaska	2018	15	3	15	3	13752	15.253	902	5.706
11	10	Alaska	2019	0	0	0	0	14867	15.252	975	5.698
12	11	Arizona	2015	12	2	6602	1157	146450	21.522	6805	5.717
13	12	Arizona	2016	57	10	6204	1089	116678	21.516	5423	5.722
14	13	Arizona	2017	0	0	6584	1155	134024	21.543	6221	5.723
15	14	Arizona	2018	0	0	6758	1184	140759	21.489	6550	5.706
16	15	Arizona	2019	0	0	3026	531	82222	21.398	3843	5.698

Repeating Columns

MSN	Description	Unit
BDFDB	Biomass inputs (feedstock) to the production of biodiesel	Billion Btu
BDPRP	Biodiesel production	Thousand barrels
BFFDB	Biomass inputs (feedstock) to the production of biofuels	Billion Btu
BFPRP	Biofuels production	Thousand barrels
CLPRB	Coal production	Billion Btu
CLPRK	Factor for converting coal production from physical units to Btu	Million Btu per short ton
CLPRP	Coal production	Thousand short tons
COPRK	Factor for converting crude oil production from physical units to Btu	Million Btu per barrel
EMFDB	Biomass inputs (feedstock) to the production of fuel ethanol	Billion Btu

External Sources

Reasons for Adding dataset

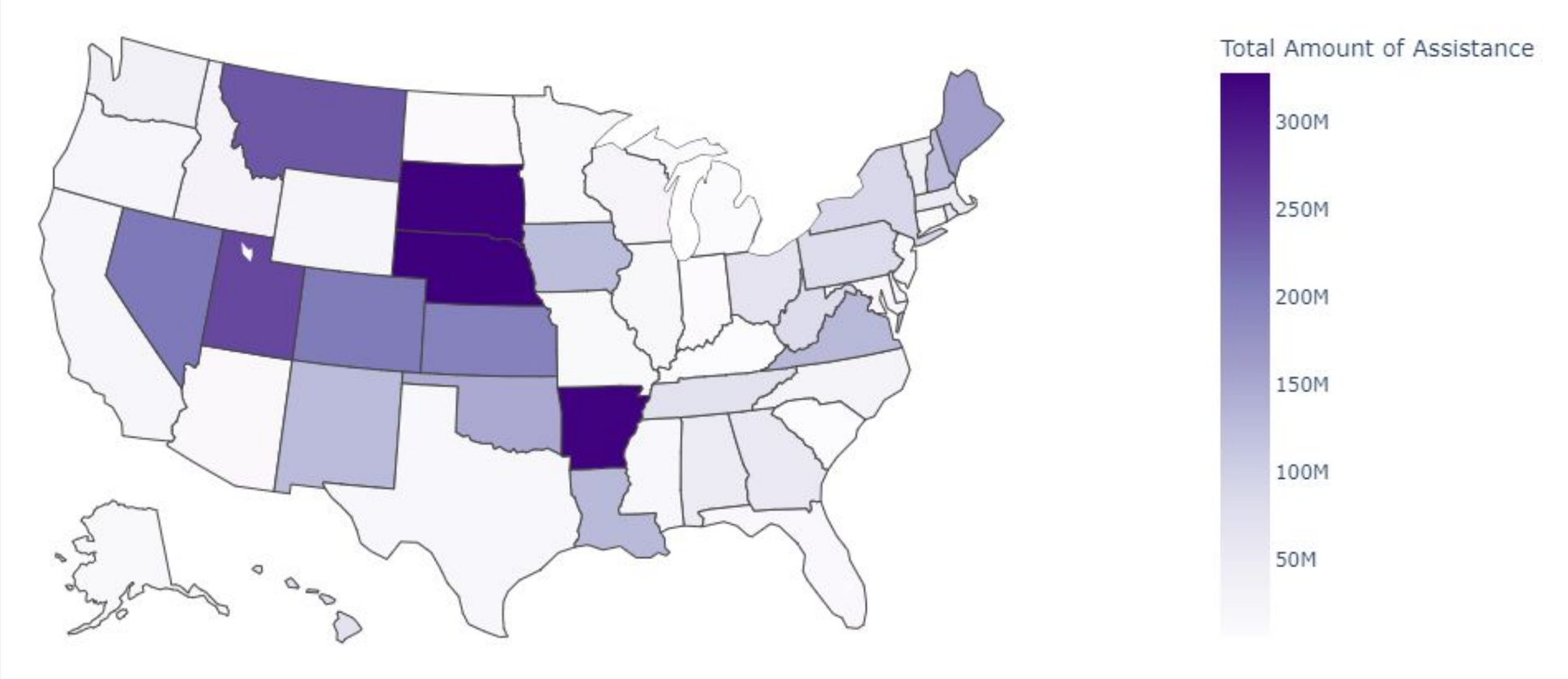
1. Limited dataset in original data, poor prediction power
2. Other factors can influence federal investment on renewable energy

   (energy disbursement, energy price, population, GDP)

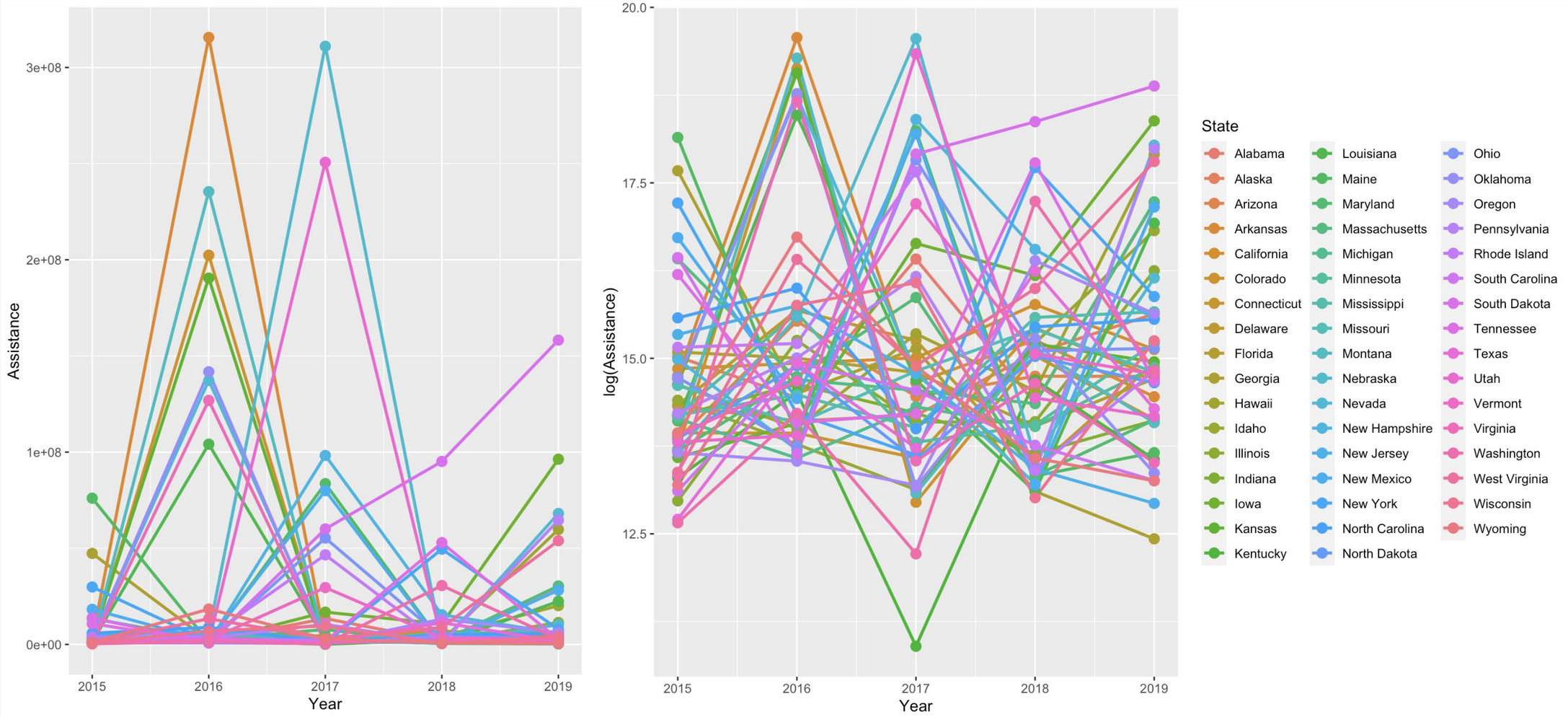
02.

Data Exploration

US Choropleth Map for Total Assistance



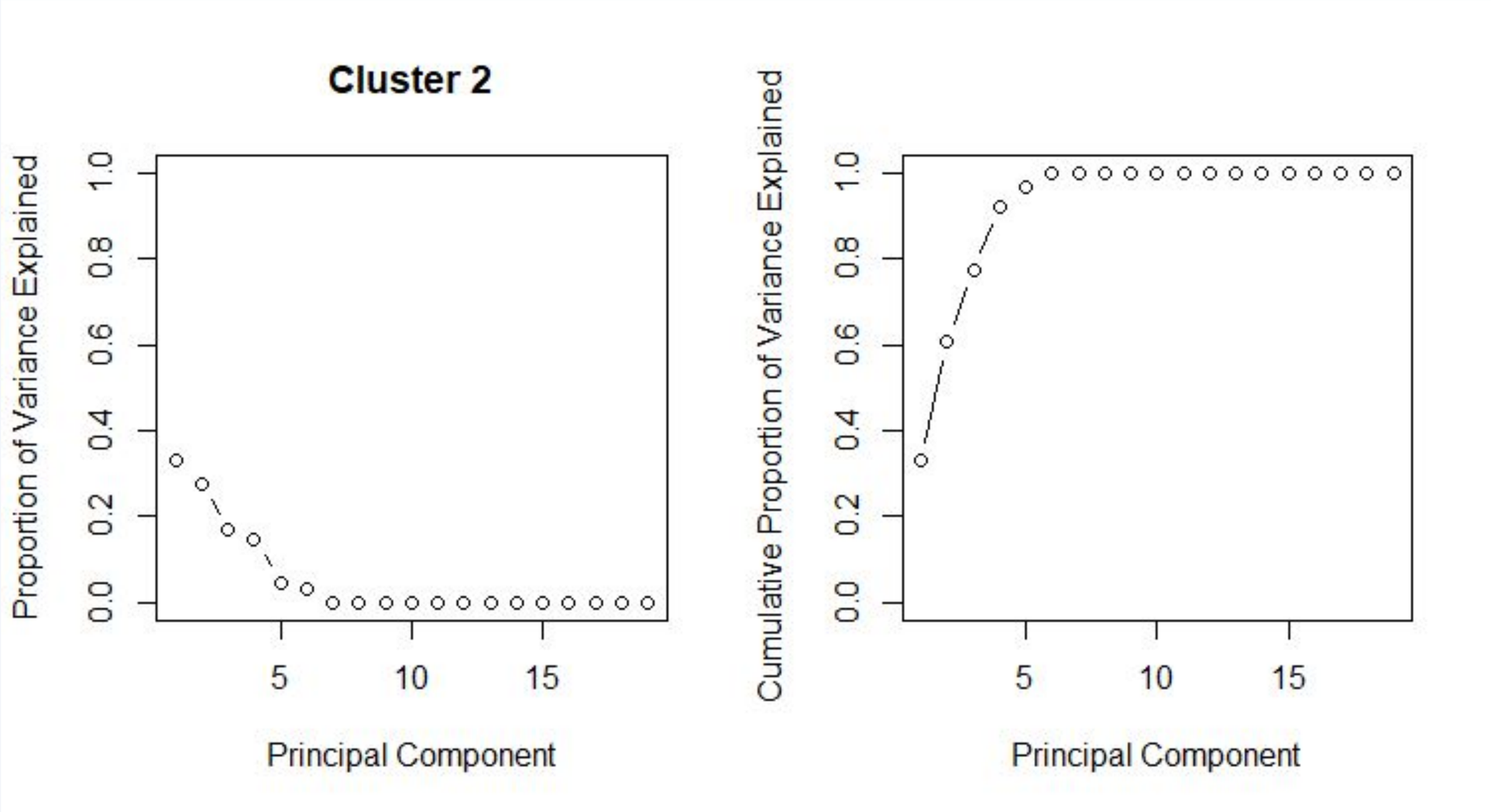
Assistance Per State For Each Year



03.

Dimension Reduction

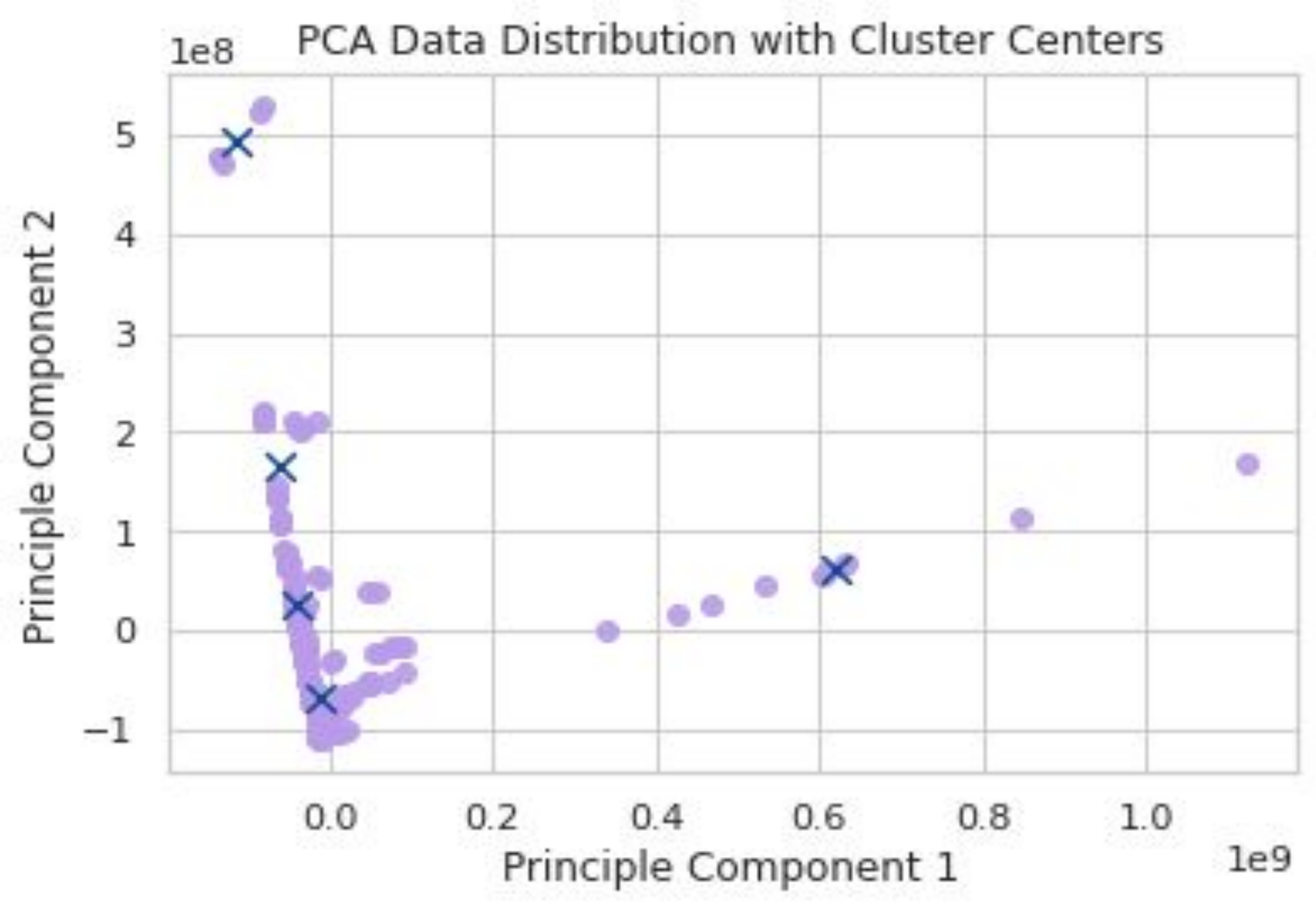
Principal Component Analysis



04.

Clustering

K-means Clustering

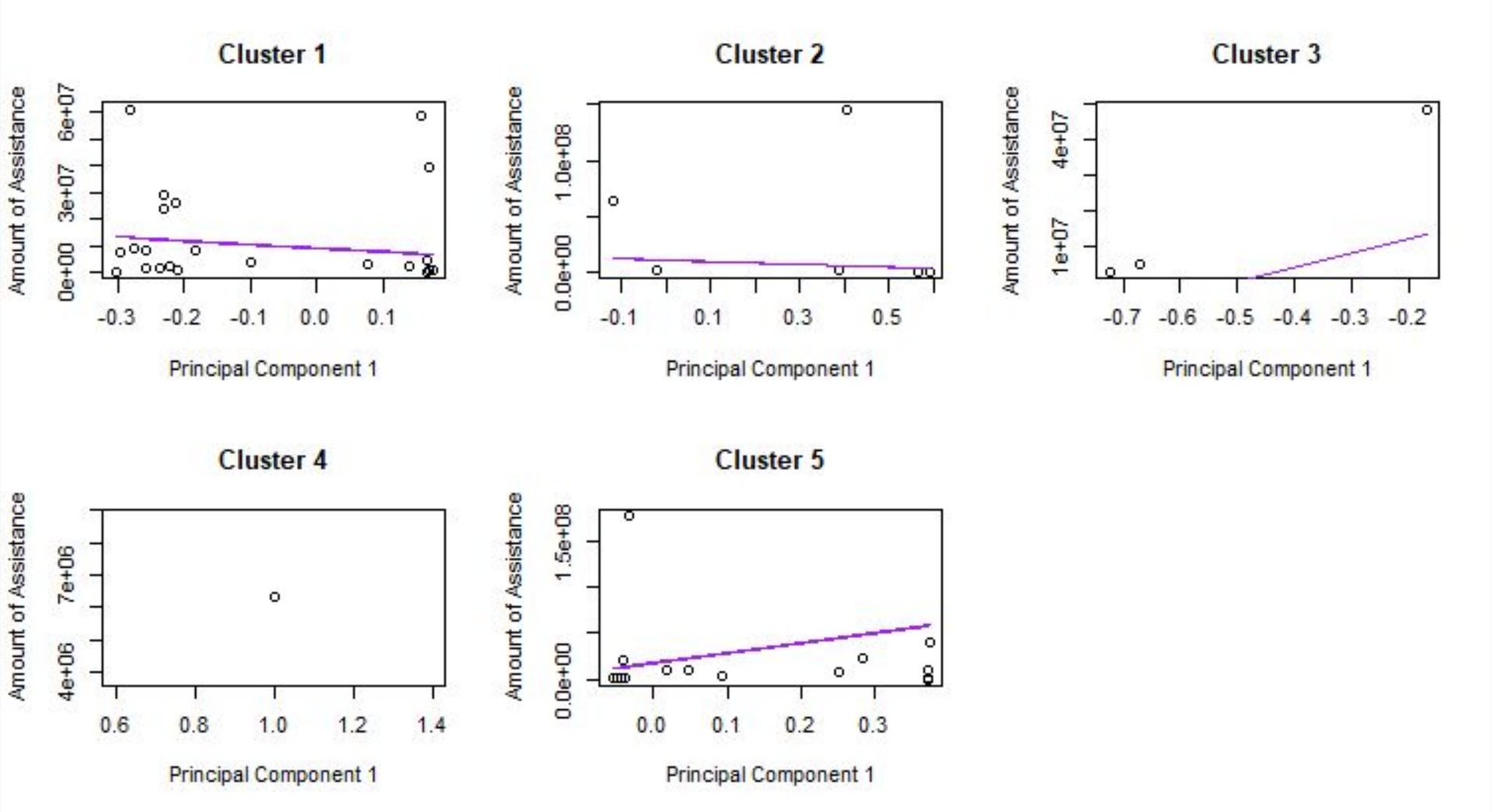


Silhouette Score

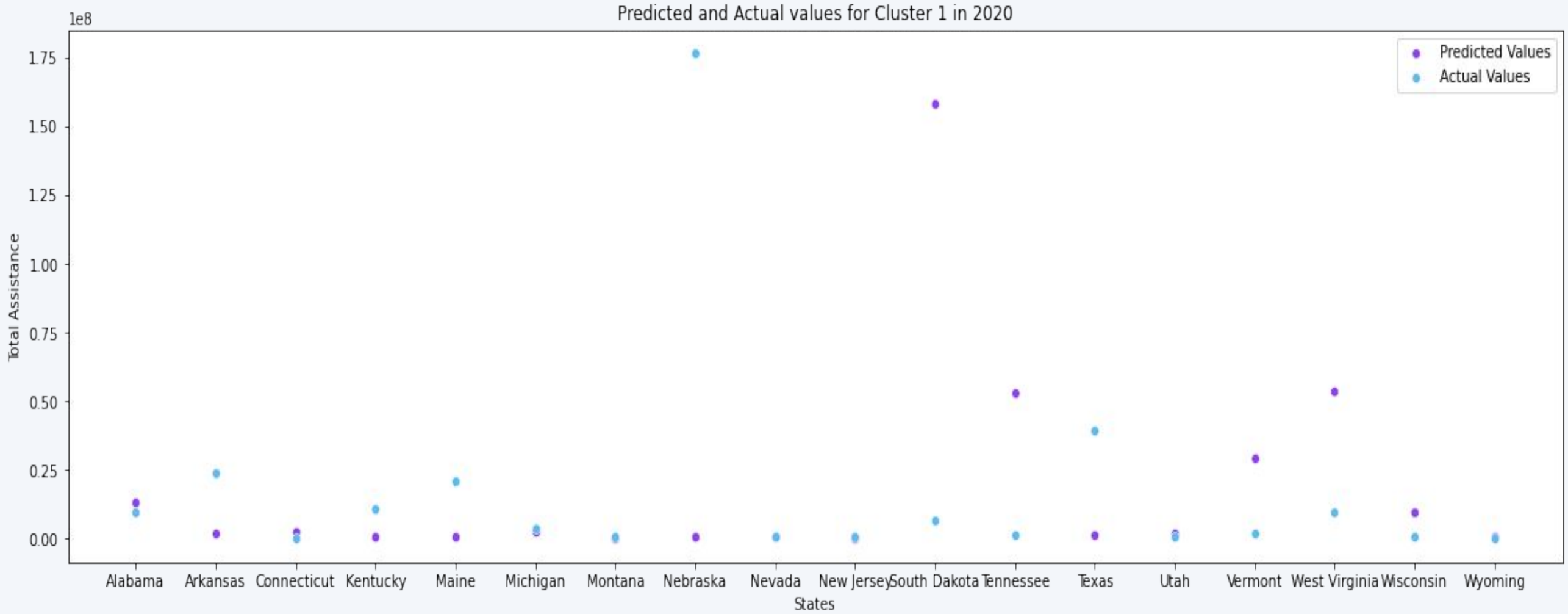
05.

Regression

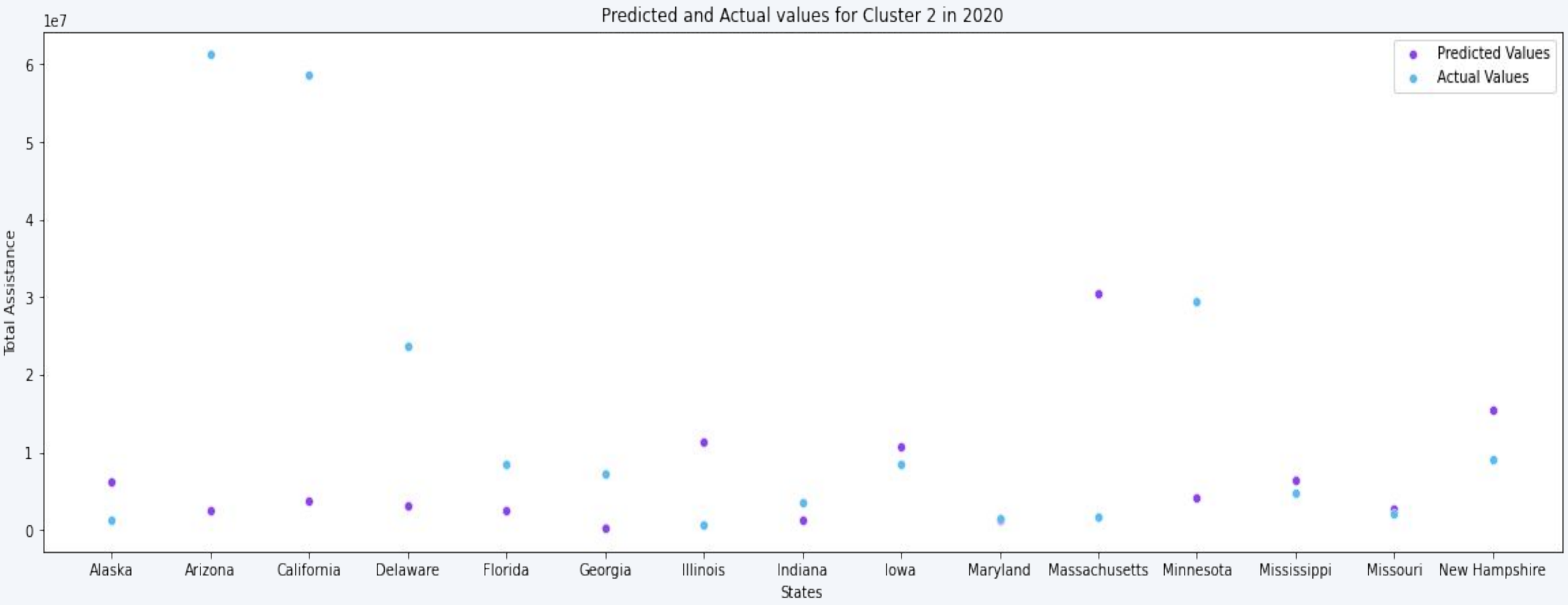
Simple Linear Regression



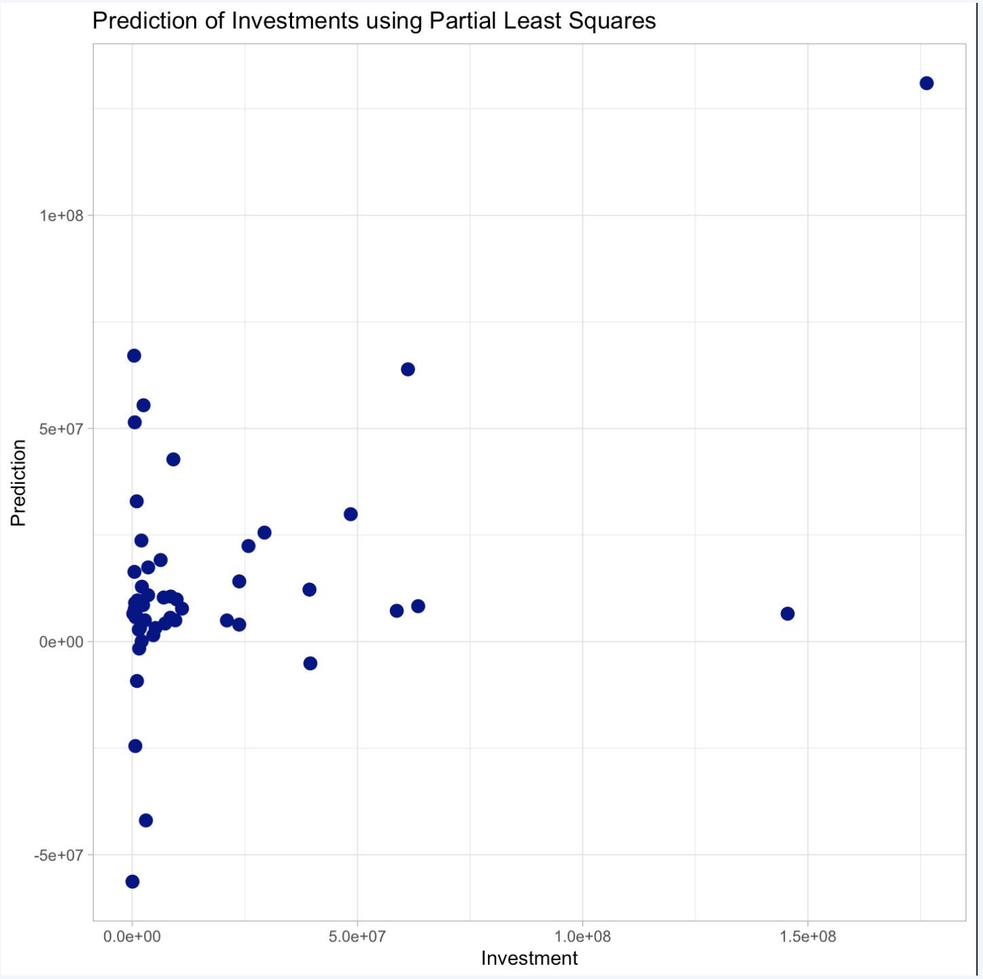
Random Forest



Random Forest



Partial Least Squares



- PLS - regression technique used for few observations but many variables
- Only used data provided

Limitations

- Assumes independence within each cluster
- Cannot be used for single-state clusters (two)
- Prediction included negative values
- Large RMSE

06.

Results and Accuracy

Root Mean Square Error

Cluster	Model	RMSE
Cluster 2	SLR; full model	3.297159
Cluster 2	SLR; no Net Summer Capacity and Generation, GDP, Pop	10.68734
Cluster 2	SLR; no Total Retail Sales, Net Summer Capacity and Generation, GDP, Pop	0.4187101
Cluster 2	SLR; no Net Summer Capacity and Generation, Disbursement, GDP, Pop	379.8185
Cluster 2	SLR; no Net Summer Capacity and Generation, Total Retail Sales	0.4464448
Cluster 2	Random Forest Classifier	14986823

Results

2020 Predictions

- Regression Models
  - More accurate with states that had lower historical investment

Significant Variables

- Disbursement
  - During cross validation, showed up consistently in SLR models with smallest RMSE

Where to Invest?

- South Dakota
- Texas
- Colorado
- West Virginia
- Tennessee
- New Hampshire