

SAS CODES

Objective:

To enhance and format the scatter plot graph by using PROC SGPLOT.

1. Removed the nulls values from scatter plot
2. Numbers formatted in X-axis and Y-axis.
3. Axis labeled in English.
4. Labeling the clusters.

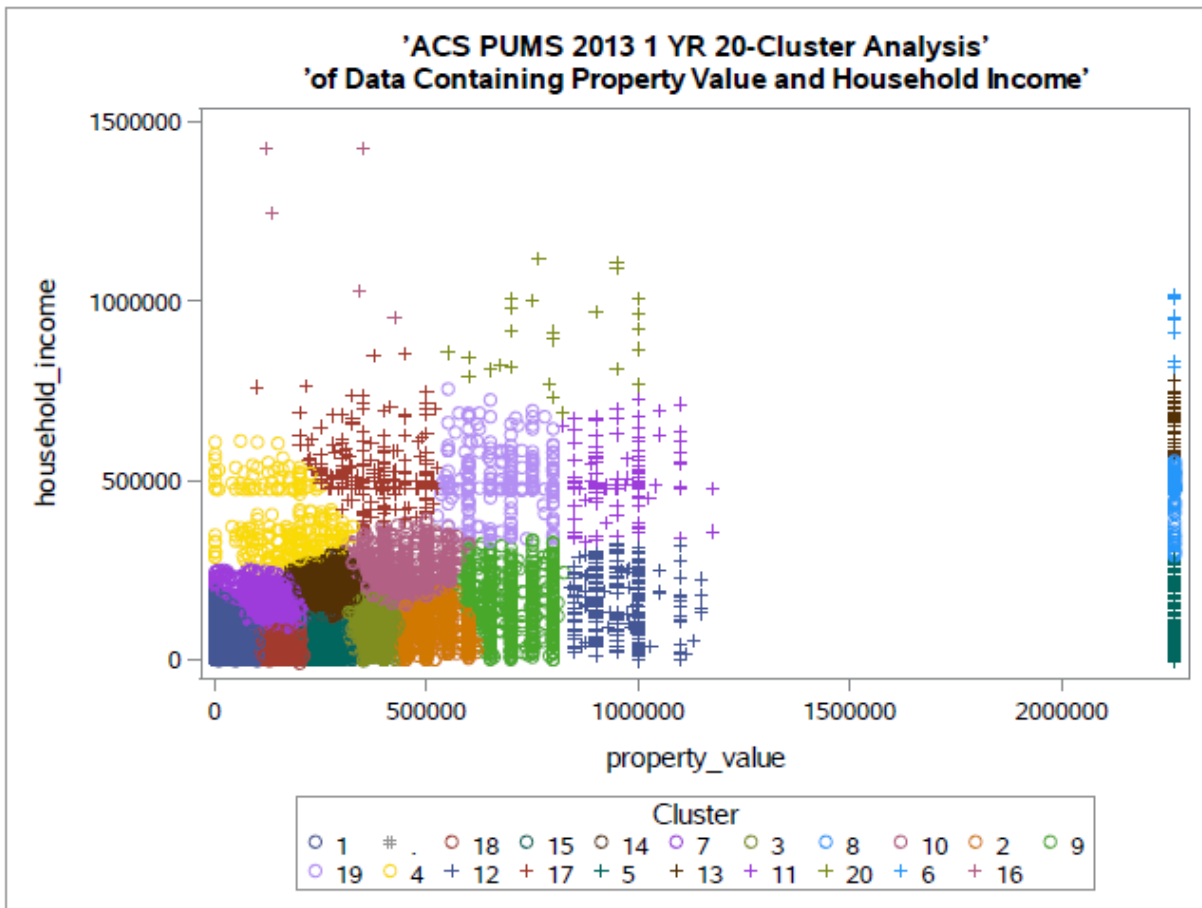
Before formatting:

Code:

```
proc fastclus data=census.psam_h17_subset1
radius=0 replace=full
converge=0 maxiter=200
maxclusters=20
OUTSTAT=census.psam_h17_subset1_20clusters_stat
OUT=census.psam_h17_subset1_20clusters
distance;
id SERIALNO;
var VALP HINCP;
run;

proc sgplot;
scatter y=HINCP x=VALP / group=cluster;
title 'ACS PUMS 2013 1 YR 20-Cluster Analysis';
title2 'of Data Containing Property Value and Household Income';
run;
```

Output:



After Formatting:

Code:

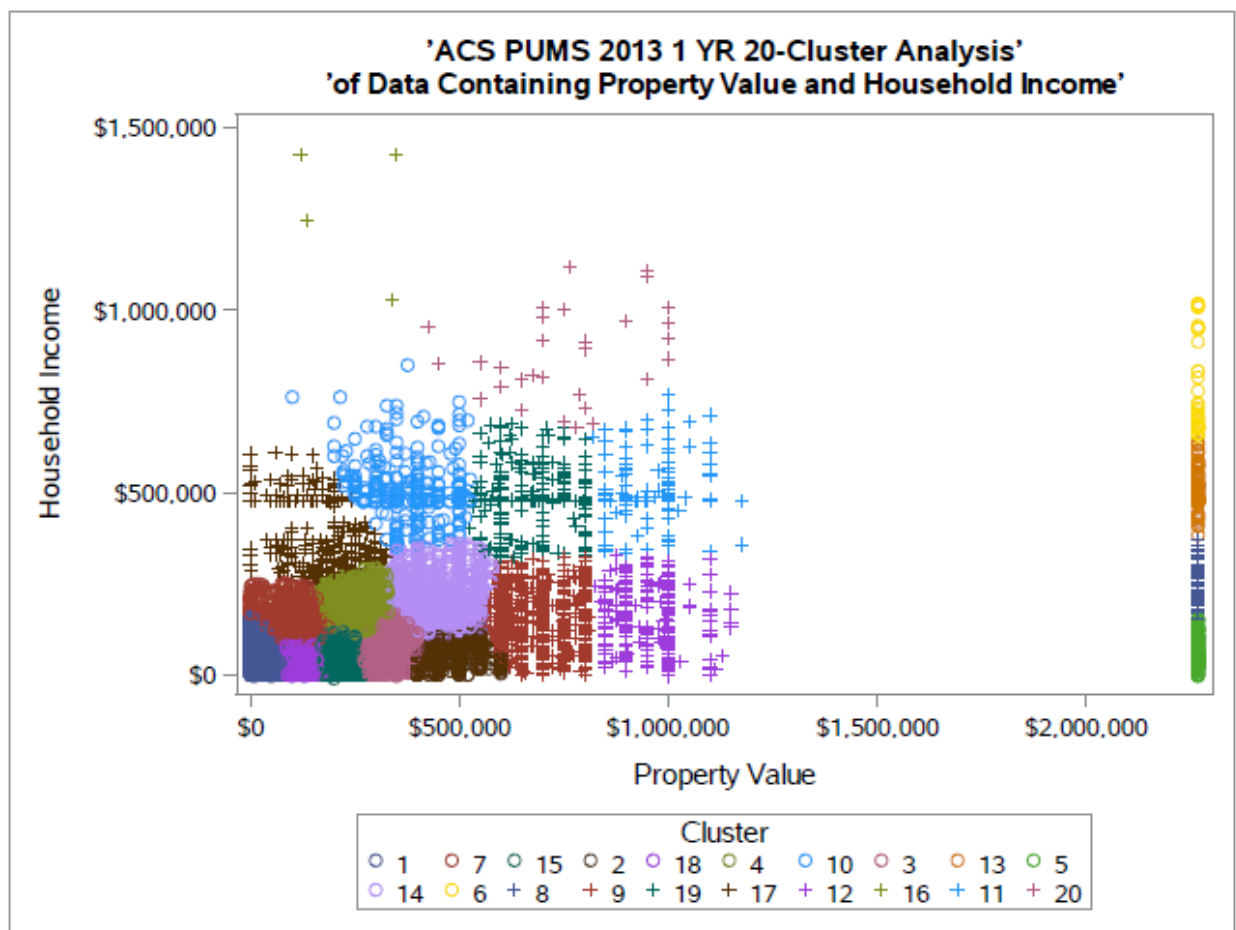
```
proc fastclus data=census.psam_h17_subset1
radius=0 replace=full
converge=0 maxiter=200
maxclusters=20
OUTSTAT=census.psam_h17_subset1_20clusters_stat
OUT=census.psam_h17_subset1_20clusters
distance;
id SERIALNO;
var VALP HINCP;
run;
```

```

proc sgplot;
scatter y=HINCP x=VALP / group=cluster nomissinggroup;
axis label= 'Property Value' tickvalueformat=dollar12.;
yaxis label= 'Household Income' tickvalueformat=dollar12.;
title 'ACS PUMS 2013 1 YR 20-Cluster Analysis';
title2 'of Data Containing Property Value and Household Income';
run;

```

Output:



Data Labeling:

Code:

```

proc fastclus data=census.psam_h17_subset1

```

```

radius=0 replace=full
converge=0 maxiter=200
maxclusters=20
OUTSTAT=census.psam_h17_subset1_20clusters_stat
OUTSEED=census.psam_h17_subset1_20clusters
distance;
id SERIALNO;
var VALP HINCP;
run;

proc sgplot;
scatter y=HINCP x=VALP / group=cluster datalabel=cluster nomissinggroup;
axis label= 'Property Value' tickvalueformat=dollar12.;
axis label= 'Household Income' tickvalueformat=dollar12.;
title 'ACS PUMS 2013 1 YR 20-Cluster Analysis';
title2 'of Data Containing Property Value and Household Income';
run;

```

Output:

