HDL Case-Study (Business Analytics)

Group : Sharang Biradar, N11407631

Parth Parmar, N19007204

Malav Gandhi, N10033700

Ganesh Parayangattil, N18808269

Question 1)

How do these two chains make their decisions about where to have store locations? What are the major criteria that drive this decision, and can you provide a very brief rationalization for each? Essentially, this question gets at how the two chains are similar in their decision making.

Answer:

Lowe's Decision Making Model:

```
> ld<-lm(Lcount~r1+r2+pop_2000+pop_2010+income_2000+income_2010+
pct_U18_2000+pct_U18_2010+pctcollege_2000+pctcollege_2010+ownhome_2000+ ownhome_2010+density_2000+density_2010+pctwhite
_2000+pctwhite_2010+    pctblack_2000+pctblack_2010)
> summary(ld)

Call:
lm(formula = Lcount ~ r1 + r2 + pop_2000 + pop_2010 + income_2000 +
    income_2010 + pct_U18_2000 + pct_U18_2010 + pctcollege_2000 +
    pctcollege_2010 + ownhome_2000 + ownhome_2010 + density_2000 +
    density_2010 + pctwhite_2000 + pctwhite_2010 + pctblack_2000 +
    pctblack_2010)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4099 -0.2842 -0.1138  0.1548  5.2716

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.010e-01  2.297e-01   3.486 0.000496 ***
r1              -4.156e-02  1.882e-02  -2.208 0.027299 *
r2               1.785e-02  4.517e-02   0.395 0.692678
pop_2000        -1.226e-05  4.798e-07 -25.549  < 2e-16 ***
pop_2010         1.457e-05  4.488e-07  32.471  < 2e-16 ***
income_2000      1.663e-05  3.705e-06   4.489 7.42e-06 ***
income_2010     -1.322e-05  2.902e-06  -4.555 5.44e-06 ***
pct_U18_2000    -3.258e-02  8.639e-03  -3.771 0.000165 ***
pct_U18_2010     3.039e-02  8.186e-03   3.712 0.000209 ***
pctcollege_2000  5.811e-04  6.004e-03   0.097 0.922906
pctcollege_2010  1.456e-02  5.621e-03   2.589 0.009658 **
ownhome_2000     3.800e-03  6.442e-03   0.590 0.555349
ownhome_2010    -1.772e-02  6.583e-03  -2.692 0.007138 **
density_2000     4.808e-05  2.773e-04   0.173 0.862364
density_2010    -1.544e-04  2.676e-04  -0.577 0.564021
pctwhite_2000    1.940e-02  4.237e-03   4.578 4.89e-06 ***
pctwhite_2010   -1.533e-02  4.307e-03  -3.560 0.000376 ***
pctblack_2000    1.702e-02  5.940e-03   2.865 0.004198 **
pctblack_2010   -8.838e-03  6.026e-03  -1.467 0.142525
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6479 on 3118 degrees of freedom
  (9 observations deleted due to missingness)
Multiple R-squared:  0.7521,    Adjusted R-squared:  0.7507
F-statistic: 525.5 on 18 and 3118 DF,  p-value: < 2.2e-16
```

Above snapshot has a linear regression model to determine Lowe's Decision Making to open new stores across the United States.

We have recognized Population, Income, under 18 residents, College going residents and classification of demographics based on race. We have considered the same because each category has its own requirements which stores like Lowe's and Home Depot can fulfill.

The model has a R-Squared Value of 0.7521 which shows that the predictability is good.

The model has 8 3 starred parameters and 3 2 starred parameters. which shows how complex the decision making process might be.

According to this model, Population, Income and Percentage of Under 18 residents are major factors that Lowe's considers before opening a new store. Surprisingly, Percentage of White residents is important factor too as per this model.

But, ethically and morally speaking no company or a firm would consider racial bias when opening a new store. So we would like to remove percentage White and Black from consideration for future calculations.

This model has almost all the factors which the dataset represents, so it represents a holistic approach to calculate the decision making process.

Home Depot's Decision Making Model

```
> hd<-lm(HDcount~r1+r2+pop_2000+pop_2010+income_2000+income_2010+
pct_U18_2000+pct_U18_2010+pctcollege_2000+pctcollege_2010+ownhome_2000+ ownhome_2010+density_2000+density_2010+pctwhite
_2000+pctwhite_2010+    pctblack_2000+pctblack_2010)
> summary(hd)

Call:
lm(formula = HDcount ~ r1 + r2 + pop_2000 + pop_2010 + income_2000 +
    income_2010 + pct_U18_2000 + pct_U18_2010 + pctcollege_2000 +
    pctcollege_2010 + ownhome_2000 + ownhome_2010 + density_2000 +
    density_2010 + pctwhite_2000 + pctwhite_2010 + pctblack_2000 +
    pctblack_2010)

Residuals:
     Min      1Q   Median      3Q      Max
-10.9312  -0.2027  -0.0334  0.1043   6.8931

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -7.196e-01  2.124e-01  -3.387 0.000714 ***
r1              -9.478e-04  1.740e-02  -0.054 0.956566
r2              -2.033e-03  4.176e-02  -0.049 0.961180
pop_2000        -6.719e-06  4.436e-07 -15.146  < 2e-16 ***
pop_2010         1.255e-05  4.149e-07  30.256  < 2e-16 ***
income_2000      1.054e-05  3.426e-06   3.077 0.002109 **
income_2010     -1.169e-06  2.683e-06  -0.436 0.663020
pct_U18_2000     1.024e-02  7.988e-03   1.282 0.199842
pct_U18_2010    -1.075e-02  7.569e-03  -1.420 0.155763
pctcollege_2000  1.297e-02  5.551e-03   2.336 0.019532 *
pctcollege_2010  8.235e-05  5.198e-03   0.016 0.987360
ownhome_2000     7.052e-03  5.957e-03   1.184 0.236564
ownhome_2010    -4.877e-03  6.087e-03  -0.801 0.423057
density_2000    -6.725e-04  2.564e-04  -2.623 0.008764 **
density_2010     5.056e-04  2.474e-04   2.044 0.041086 *
pctwhite_2000    1.816e-02  3.918e-03   4.634 3.73e-06 ***
pctwhite_2010   -1.884e-02  3.982e-03  -4.730 2.34e-06 ***
pctblack_2000   -1.762e-02  5.492e-03  -3.208 0.001352 **
pctblack_2010    1.900e-02  5.571e-03   3.411 0.000657 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5991 on 3118 degrees of freedom
  (9 observations deleted due to missingness)
Multiple R-squared:  0.9173,    Adjusted R-squared:  0.9169
F-statistic:  1923 on 18 and 3118 DF,  p-value: < 2.2e-16
```

:

As described in the snapshot, this is a linear regression model that represents Home Depot's decision making process.

We have recognized Population, Income, under 18 residents, College going residents and classification of demographics based on race. We have considered the same because each category has its own requirements which stores like Lowe's and Home Depot can fulfill.

This model has a R-Squared value: 0.9173, hence predictability of the model is pretty good. The model has five 3 starred parameters, and three 2 starred, which shows how complex the decision making process might be.

The factors which are the most important for Home Depot are: population, percentage of white residents and percentage of Black residents.

Important to notice, Home Depot considers density as an important factor while Lowe's disregards it. On the contrary, it ignores Under 18 and College going residents whereas Lowe's takes it into consideration.
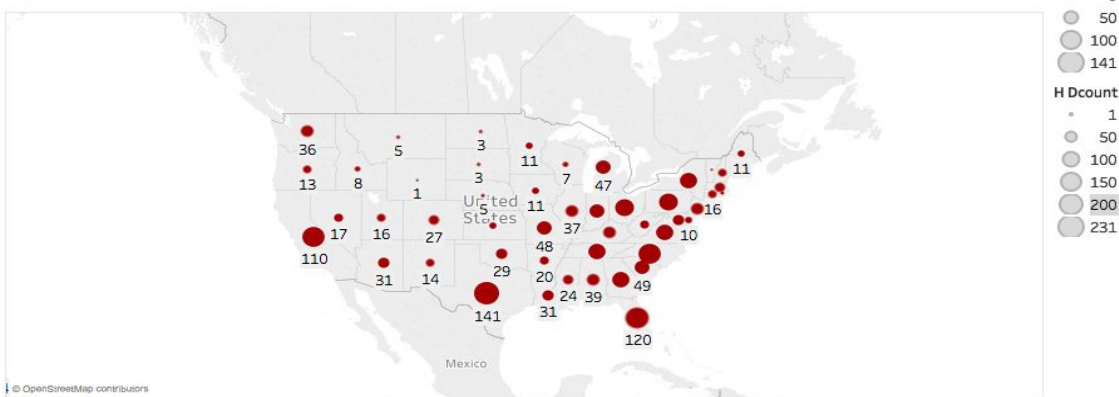
Question 2)

Are there ways in which the two chains are different in the types of locations they target? What are those differences, and why do you think that they may be apparent in the data? Characterize the targeting strategies for each of the two chains.
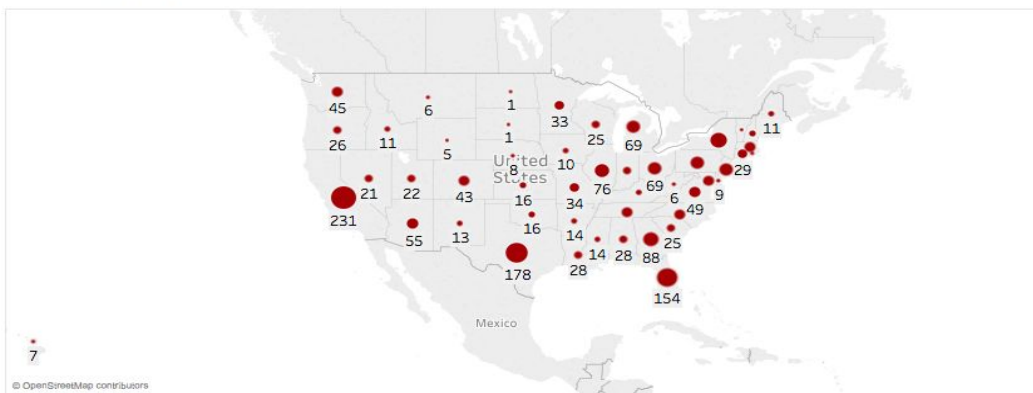
Answer:

Following is the distribution of stores across the United States. Larger the dots greater the presence in those areas.
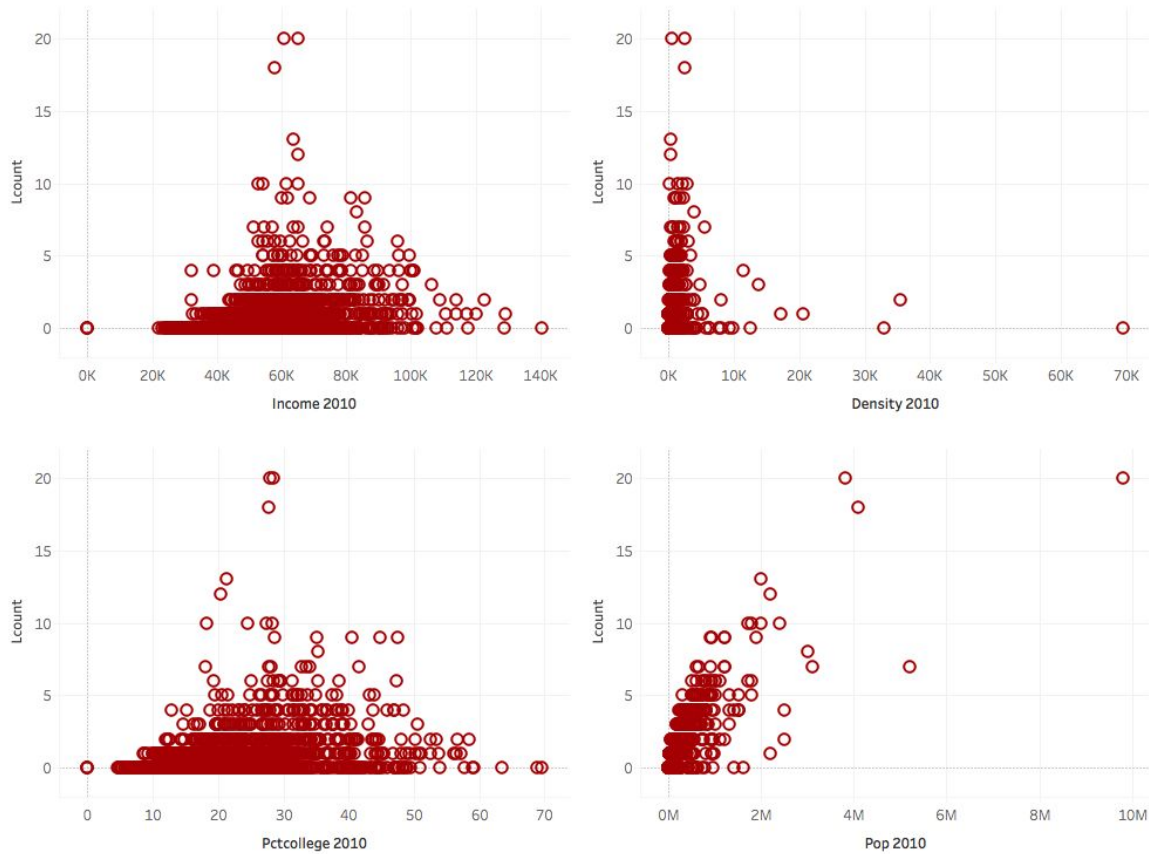

Easy to Notice that Lowe's focus is more on the East Coast, so is Home Depot's, although deriving from the size of the dots, Lowes cover those markets better than Home Depot because they have more number of stores.

Home Depot concentrates more on the obvious metropolitan areas of the United States (signified by large dots).

Lowe Stores

Above are some of the correlations between the number of Lowe's Stores and the factors it is based on. The factors are selected as per the regression model we have used to answer the first question. Also, 2010 numbers are chosen because they are having more impact as compared to the 2000 numbers.

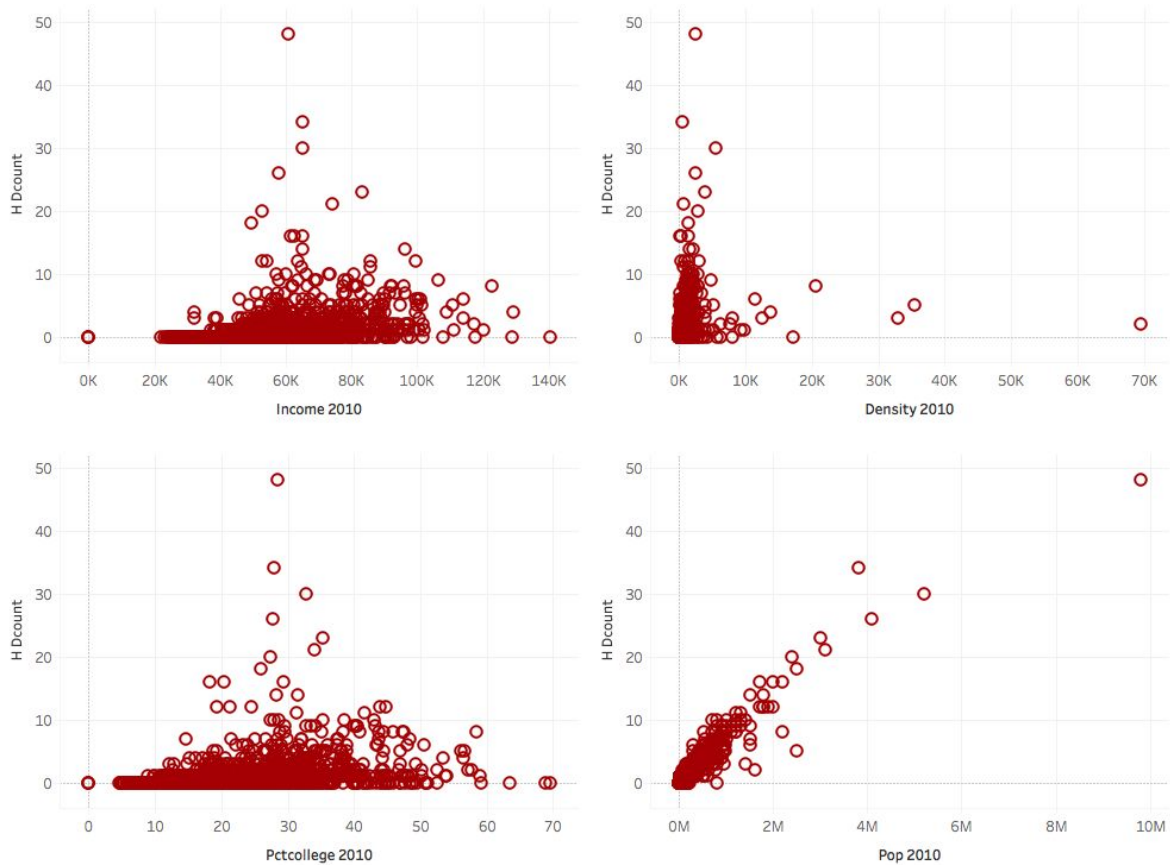An ideal situation for a Lowe's location would be :

A location with:

Household Income - $65K per annum

Percentage of college going residents: 25-30%

Population: 0.6-1 Million

Density: 1500 residents/ sq. mil

Home Depot Stores

Above are some of the correlations between the number of Home Depot Stores and the factors it is based on. The factors are selected as per the regression model we have used to answer the first question. Also, 2010 numbers are chosen because they are having more impact as compared to the 2000 numbers.

An ideal situation for a Home Depot location would be :

A location with:

Household Income - $65K per annum
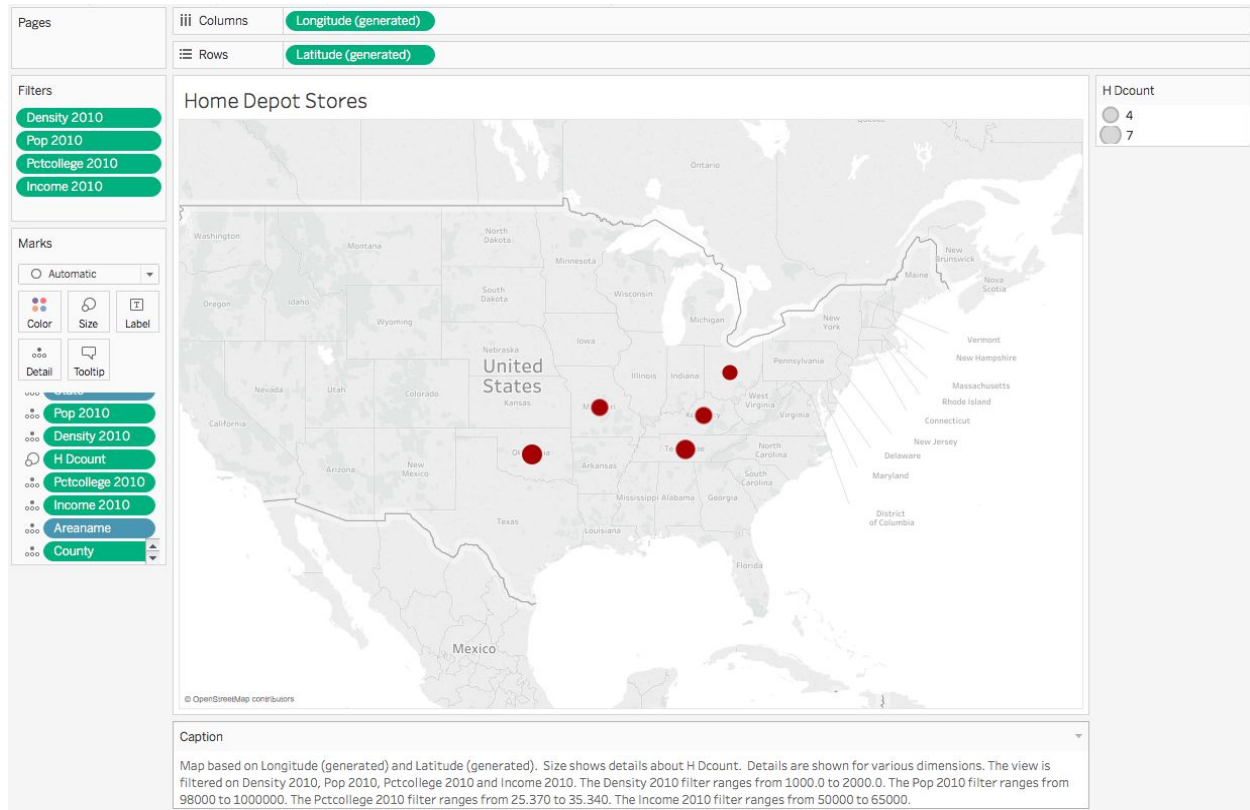
Percentage of college going residents: 30-35%

Population: 0.6-1 Million

Density: 2000 residents/ sq. mil

Question 3)

What counties appear underserved in the data, by one or both store chains? Where would you expect Home Depot to open its next 2-3 locations? Lowe's?

Answer:

Map based on Longitude (generated) and Latitude (generated). Size shows details about H Dcount. Details are shown for various dimensions. The view is filtered on Density 2010, Pop 2010, Pctcollege 2010 and Income 2010. The Density 2010 filter ranges from 1000.0 to 2000.0. The Pop 2010 filter ranges from 98000 to 1000000. The Pctcollege 2010 filter ranges from 25.370 to 35.340. The Income 2010 filter ranges from 50000 to 65000.

View Data: Sheet 12

☑ Show aliases   Copy   Export All

State	Areaname	County	Density 2010	H Dcount	Income 2010	Pctcollege 2010	Pop 2010
TN	Shelby	47157	1215.500	6	55513	27.50000	927644.00
TN	Davidson	47037	1243.300	6	57601	33.60000	626681.00
OK	Oklahoma	40109	1013.800	7	54056	28.50000	718633.00
OK	Tulsa	40143	1058.100	5	57865	29.10000	603403.00
OH	Summit	39153	1312.600	4	61579	28.70000	541781.00
MO	Jackson	29095	1115.300	5	58778	26.90000	674158.00
KY	Jefferson	21111	1948.100	5	58869	28.10000	741096.00

Summary   Full Data   7 rows

Suggestion: Home Depot Stores should consider opening its stores in the above states and the respective areas mentioned, considering that they are underserved.
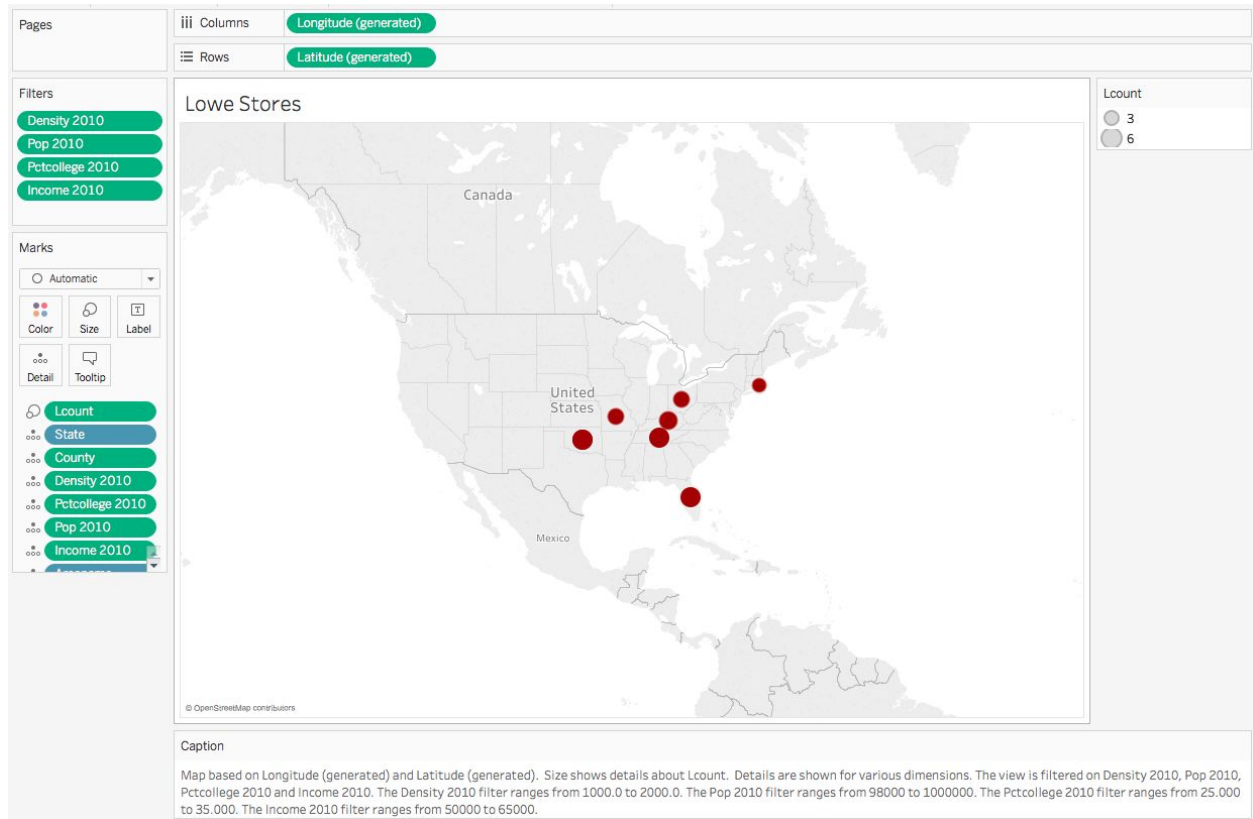
Caption: Map based on Longitude (generated) and Latitude (generated). Size shows details about Lcount. Details are shown for various dimensions.

The view is filtered on Density 2010, Pop 2010, Pctcollege 2010 and Income 2010. The Density 2010 filter ranges from 1000.0 to 2000.0. The Pop 2010 filter ranges from 98000 to 1000000. The Pctcollege 2010 filter ranges from 25.000 to 35.000. The Income 2010 filter ranges from 50000 to 65000.

		View Data: Sheet 12 (2)					
☑ Show aliases						Copy	Export All
Sta ▼	Areaname	County	Density 2010	Income 2010	Lcount	Pctcollege 2010	Pop 2010
TN	Davidson	47037	1243.300	57601	5	33.60000	626681.00
TN	Shelby	47157	1215.500	55513	6	27.50000	927644.00
RI	Providence	44007	1530.300	62108	3	25.10000	626667.00
OK	Oklahoma	40109	1013.800	54056	6	28.50000	718633.00
OK	Tulsa	40143	1058.100	57865	6	29.10000	603403.00
OH	Summit	39153	1312.600	61579	4	28.70000	541781.00
MO	Jackson	29095	1115.300	58778	4	26.90000	674158.00
KY	Jefferson	21111	1948.100	58869	5	28.10000	741096.00
FL	Duval	12031	1133.900	59520	6	25.10000	864263.00

Summary	Full Data						9 rows

Suggestion: Lowe's should consider opening their stores in the states and areas mentioned above, considering that they are underserved.