Business Analytics

Case Study

Pushpak Jain (N16771623)

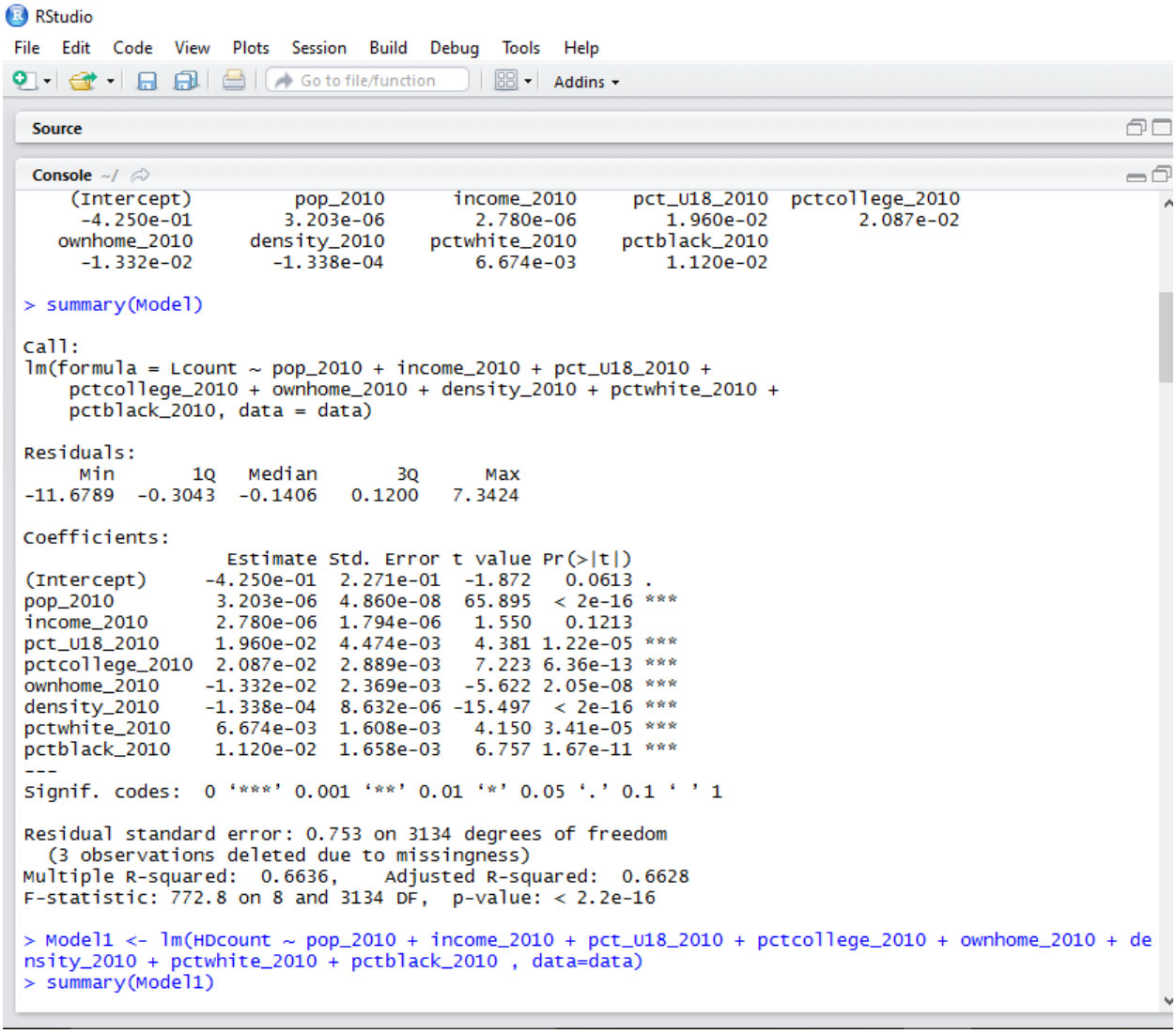
Ushnik Dasgupta (N17591222)

Harshitha Maithry (N12854198)

Ashwin Rajendran (N16477173)

Q1.

The linear model for the no. of Lowe's stores (dependant variable):




```
    (Intercept)         pop_2010      income_2010     pct_U18_2010  pctcollege_2010
     -4.250e-01        3.203e-06        2.780e-06        1.960e-02        2.087e-02
   ownhome_2010     density_2010    pctwhite_2010    pctblack_2010
     -1.332e-02       -1.338e-04        6.674e-03        1.120e-02

> summary(Model)

Call:
lm(formula = Lcount ~ pop_2010 + income_2010 + pct_U18_2010 +
    pctcollege_2010 + ownhome_2010 + density_2010 + pctwhite_2010 +
    pctblack_2010, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-11.6789  -0.3043  -0.1406   0.1200   7.3424

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -4.250e-01  2.271e-01  -1.872   0.0613 .
pop_2010        3.203e-06  4.860e-08  65.895  < 2e-16 ***
income_2010     2.780e-06  1.794e-06   1.550   0.1213
pct_U18_2010    1.960e-02  4.474e-03   4.381 1.22e-05 ***
pctcollege_2010 2.087e-02  2.889e-03   7.223 6.36e-13 ***
ownhome_2010   -1.332e-02  2.369e-03  -5.622 2.05e-08 ***
density_2010   -1.338e-04  8.632e-06 -15.497  < 2e-16 ***
pctwhite_2010   6.674e-03  1.608e-03   4.150 3.41e-05 ***
pctblack_2010   1.120e-02  1.658e-03   6.757 1.67e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.753 on 3134 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.6636,    Adjusted R-squared:  0.6628
F-statistic: 772.8 on 8 and 3134 DF,  p-value: < 2.2e-16

> Model1 <- lm(HDcount ~ pop_2010 + income_2010 + pct_U18_2010 + pctcollege_2010 + ownhome_2010 + de
nsity_2010 + pctwhite_2010 + pctblack_2010 , data=data)
> summary(Model1)
```

The linear model for the number of Lowe's outlets depend significantly on all the independent variables (for year 2010) exept for income levels in 2010 and gives us a model strength of 66.28%.

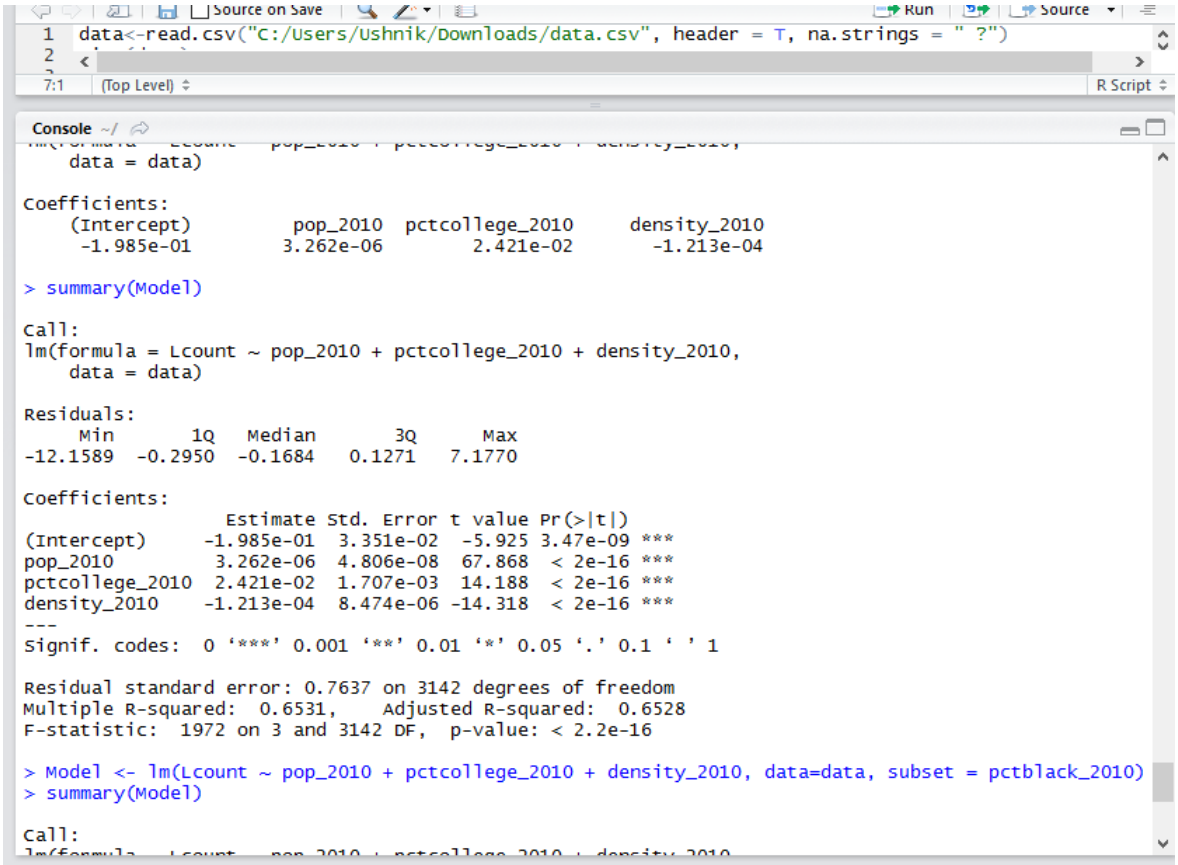
Logically, we exclude:

- percentages of white and black people (irrelevant business approach),
- income levels (not significant from the model above),
- percentage under 18 (would not provide a great deal of leverage especially because it is a home improvement store that retails home appliances),
- owning a home (negative relation from the model above makes no logical sense. Increasing number of houses should increase sales and should translate into greater number of stores)

We include:

- population (higher the footfall higher the sales),
- percentage college (direct relationship of income levels with increasing education that enhances the purchsing power),
- densities (residents per square mile – New York is highly dense (one store would be accessible by many), whereas cities in California are much less dense (accessibility to a particular store reduces; and therefore calls for more no. of stores)

The model returns a strength of 65.28% with these independent variables as shown in the screenshot below:



However, when we include the population (screenshot below) as the only independent variable, we get a model strength of 61.37%. This communicates that Lowe's takes population as the most significant factor in making decisions regarding their stores.

```
> data<-read.csv("C:/Users/Ushnik/Downloads/data.csv", header = T, na.strings = " ?")
> Model <- lm(Lcount ~ pop_2010, data = data)
> summary(Model)

Call:
lm(formula = Lcount ~ pop_2010, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-12.0925  -0.3003  -0.2552   0.2765   7.4192

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.234e-01  1.505e-02   14.84   <2e-16 ***
pop_2010    3.252e-06  4.600e-08   70.69   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8055 on 3144 degrees of freedom
Multiple R-squared:  0.6138,    Adjusted R-squared:  0.6137
F-statistic:  4998 on 1 and 3144 DF,  p-value: < 2.2e-16

> |
```

The linear model for the no. of Home Depot stores (dependant variable):

```
Console ~/
F-statistic: 772.8 on 8 and 3134 DF,  p-value: < 2.2e-16

> Model1 <- lm(HDcount ~ pop_2010 + income_2010 + pct_U18_2010 + pctcollege_2010 + ownhome_2010 + de
nsity_2010 + pctwhite_2010 + pctblack_2010 , data=data)
> summary(Model1)

Call:
lm(formula = HDcount ~ pop_2010 + income_2010 + pct_U18_2010 +
    pctcollege_2010 + ownhome_2010 + density_2010 + pctwhite_2010 +
    pctblack_2010, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.7191  -0.2138  -0.0456   0.0987   9.7852

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -9.044e-01  1.923e-01  -4.703 2.67e-06 ***
pop_2010         6.312e-06  4.116e-08 153.365  < 2e-16 ***
income_2010      7.123e-06  1.519e-06   4.689 2.87e-06 ***
pct_U18_2010     8.222e-03  3.789e-03   2.170   0.0301 *
pctcollege_2010  1.639e-02  2.447e-03   6.699 2.48e-11 ***
ownhome_2010     2.850e-03  2.006e-03   1.421   0.1555
density_2010    -1.638e-04  7.310e-06 -22.411  < 2e-16 ***
pctwhite_2010   -1.808e-03  1.362e-03  -1.327   0.1845
pctblack_2010    1.540e-03  1.404e-03   1.097   0.2727
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6376 on 3134 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.9059,    Adjusted R-squared:  0.9057
F-statistic:  3772 on 8 and 3134 DF,  p-value: < 2.2e-16
```

This model for the number of Home Depot stores depend significantly on population, income levels, percentage college, and density of the area in 2010.

Excluding the insignificant variables, we get a model strength of 90.5%:

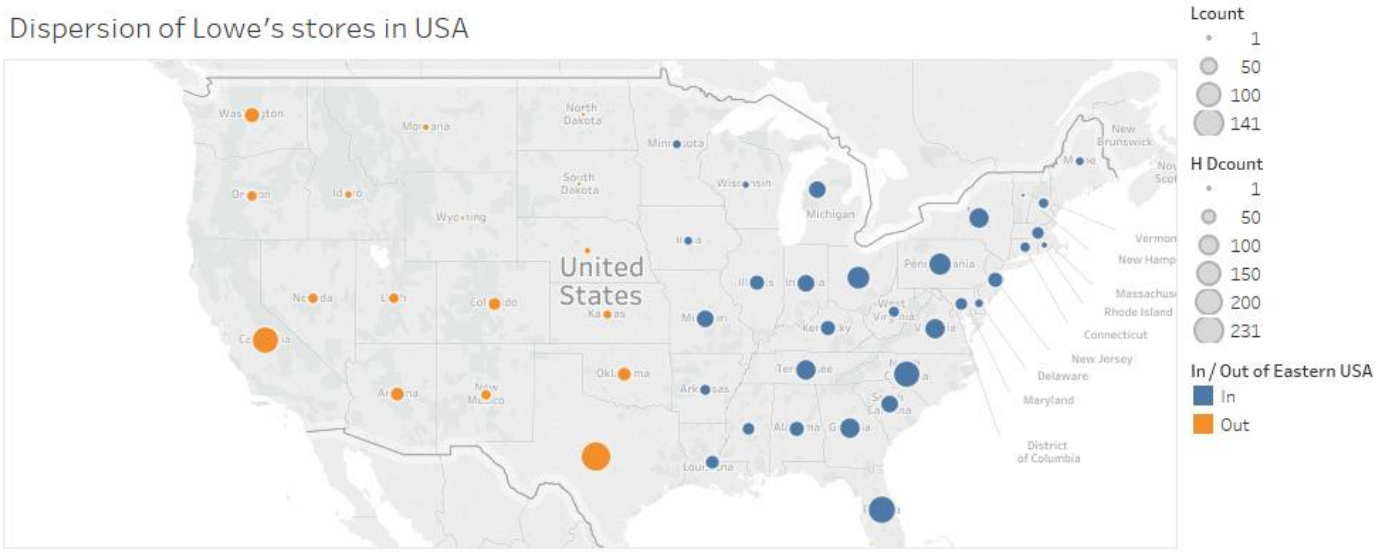
```
> Model1 <- lm(HDcount ~ pop_2010 + pctcollege_2010 + density_2010 + income_2010, data=data)
> summary(Model1)

Call:
lm(formula = HDcount ~ pop_2010 + pctcollege_2010 + density_2010 +
    income_2010, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.9599  -0.2184  -0.0423   0.0963   9.7188

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -6.034e-01  5.085e-02 -11.866  < 2e-16 ***
pop_2010         6.339e-06  4.030e-08 157.296  < 2e-16 ***
pctcollege_2010  1.471e-02  2.140e-03   6.877 7.35e-12 ***
density_2010    -1.644e-04  7.110e-06 -23.126  < 2e-16 ***
income_2010      6.951e-06  1.357e-06   5.122 3.21e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6396 on 3141 degrees of freedom
Multiple R-squared:  0.9051,    Adjusted R-squared:  0.905
F-statistic:  7491 on 4 and 3141 DF,  p-value: < 2.2e-16

>
```

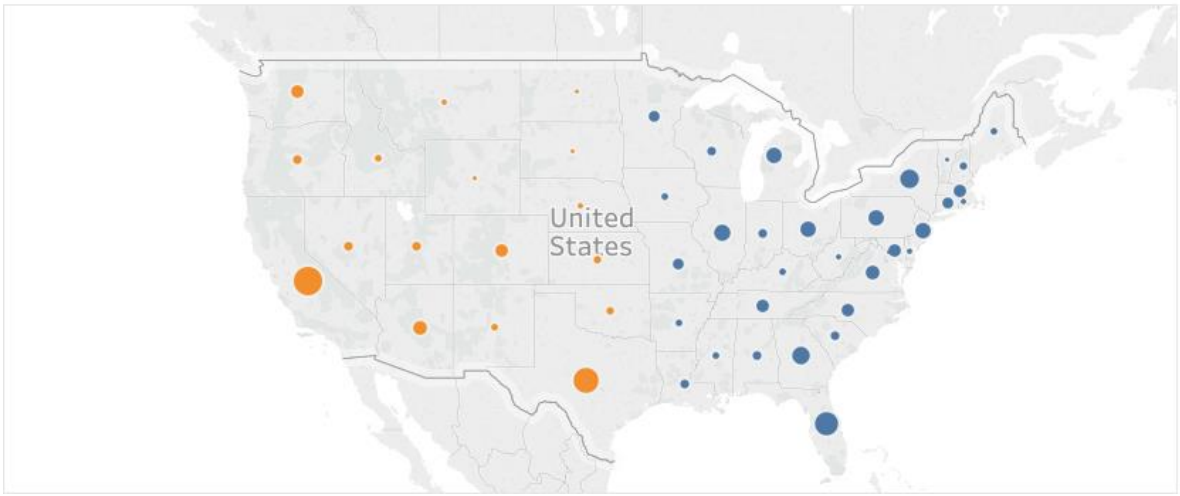
Furthermore, when we include population (screenshot below) as the only independent variable, we get a model strength of 88.27%. This communicates that HD also takes population as the most significant factor in making decisions regarding their stores.

```
Console ~/
F-statistic:  4998 on 1 and 3144 DF,  p-value: < 2.2e-16

> Model <- lm(HDcount ~ pop_2010, data = data)
> summary(Model)

Call:
lm(formula = HDcount ~ pop_2010, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.1981  -0.1910  -0.0903  -0.0237  10.2651

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.247e-03  1.328e-02   0.621    0.535
pop_2010    6.244e-06  4.059e-08 153.846   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7107 on 3144 degrees of freedom
Multiple R-squared:  0.8827,    Adjusted R-squared:  0.8827
F-statistic: 2.367e+04 on 1 and 3144 DF,  p-value: < 2.2e-16

>
```

Both the chains therefore target population (most important factor for both), percentge college, densities and use statistics of these variables in their decisions of opening new stores. However, unlike Lowe's, Home Depot also takes income levels into consideration in addition to the other three factors.

Q2.





There is a concentration of both Lowe's as well as Home Depot stores on the Eastern coast of USA.

Summary of Lowe's stores in Eastern USA (31 states with a total of 1227 stores; total population of 201,008,271):

View Data: Sheet 44

Show aliases

In / Out of Eastern USA	State	Latitude (generated)	Longitude (generated)	Lcount	Pop 2010
In	WV	38.6300	-80.7300	18	1,852,994.00
In	WI	44.5000	-90.0000	7	5,686,986.00
In	VT	44.0000	-72.7400	2	625,741.00
In	VA	37.7700	-78.2400	66	8,019,298.00
In	TN	35.7500	-86.2500	60	6,346,105.00
In	SC	34.0000	-81.0000	49	4,625,364.00
In	RI	41.7500	-71.4900	5	1,052,567.00
In	PA	40.8600	-77.9000	79	12,653,025.00
In	OH	40.2500	-83.0000	83	11,592,968.00
In	NY	43.0000	-75.4900	67	19,338,817.00
In	NJ	40.1600	-74.4900	39	8,791,894.00
In	NH	43.6600	-71.4900	13	1,316,470.00
In	NC	35.5000	-80.0000	111	9,535,483.00
In	MS	32.7500	-89.7500	24	2,967,297.00
In	MO	38.2500	-92.5000	48	5,988,927.00
In	MN	46.2500	-94.2500	11	5,351,500.00
In	MI	44.2500	-85.5000	47	9,860,694.00
In	ME	45.5000	-69.2400	11	1,328,361.00
In	MD	39.0200	-76.6300	28	5,773,552.00
In	MA	42.3300	-71.8200	27	6,544,544.00
In	LA	31.0000	-92.0000	31	4,533,372.00
In	KY	37.7900	-84.8700	41	4,339,367.00
In	IN	40.0000	-86.2500	44	6,483,802.00
In	IL	40.0000	-89.2500	37	12,835,957.00
In	IA	42.0000	-93.5000	11	3,046,355.00
In	GA	32.7500	-83.5000	63	9,687,653.00
In	FL	28.4200	-81.5500	120	18,661,493.00
In	DE	39.0000	-75.4900	10	897,934.00
In	CT	41.6600	-72.6600	16	3,574,097.00
In	AR	34.7500	-92.5000	20	2,915,918.00
In	AL	32.7500	-86.7500	39	4,779,736.00

Summary of Lowe's stores in Western USA (18 states with a total of 338 stores; total population of 81,725,117):

View Data: Sheet 44

Show aliases

In / Out of Eastern USA	State	Latitude (generated)	Longitude (generated)	Lcount	Pop 2010
Out	WY	43.0000	-107.5000	1	563,626.00
Out	WA	47.5000	-120.5000	36	6,693,291.00
Out	UT	39.2500	-111.7500	16	2,734,230.00
Out	SD	44.5000	-100.2500	3	814,180.00
Out	OR	44.0000	-120.5000	13	3,831,074.00
Out	OK	35.5000	-97.5000	29	3,751,351.00
Out	NV	39.2500	-116.7500	17	2,749,282.00
Out	NM	34.5000	-106.0000	14	2,059,179.00
Out	NE	41.5000	-99.7500	5	1,826,341.00
Out	ND	47.5000	-100.0000	3	672,591.00
Out	MT	47.0000	-109.7500	5	989,415.00
Out	KS	38.5000	-98.5000	11	2,853,118.00
Out	ID	44.0739	-114.5420	8	1,567,582.00
Out	HI	20.7900	-156.3400	4	1,360,301.00
Out	CO	39.0000	-105.5000	27	5,029,196.00
Out	CA	37.2500	-119.7500	110	37,145,229.00
Out	AZ	34.5000	-111.5000	31	6,374,900.00
Out	AK	64.0000	-150.0000	5	710,231.00

Summary of Home Depot stores in Eastern USA (32 states with a total of 1240 stores; total population of 201,609,994):

View Data: Sheet 43

Show aliases

In / Out of Eastern USA	State	Latitude (generated)	Longitude (generated)	H Dcount	Pop 2010
In	WV	38.6300	-80.7300	6	1,852,994.00
In	WI	44.5000	-90.0000	25	5,686,986.00
In	VT	44.0000	-72.7400	3	625,741.00
In	VA	37.7700	-78.2400	49	8,019,298.00
In	TN	35.7500	-86.2500	39	6,346,105.00
In	SC	34.0000	-81.0000	25	4,625,364.00
In	RI	41.7500	-71.4900	8	1,052,567.00
In	PA	40.8600	-77.9000	69	12,653,025.00
In	OH	40.2500	-83.0000	69	11,592,968.00
In	NY	43.0000	-75.4900	100	19,338,817.00
In	NJ	40.1600	-74.4900	67	8,791,894.00
In	NH	43.6600	-71.4900	20	1,316,470.00
In	NC	35.5000	-80.0000	39	9,535,483.00
In	MS	32.7500	-89.7500	14	2,967,297.00
In	MO	38.2500	-92.5000	34	5,988,927.00
In	MN	46.2500	-94.2500	33	5,351,500.00
In	MI	44.2500	-85.5000	69	9,860,694.00
In	ME	45.5000	-69.2400	11	1,328,361.00
In	MD	39.0200	-76.6300	41	5,773,552.00
In	MA	42.3300	-71.8200	44	6,544,544.00
In	LA	31.0000	-92.0000	28	4,533,372.00
In	KY	37.7900	-84.8700	14	4,339,367.00
In	IN	40.0000	-86.2500	24	6,483,802.00
In	IL	40.0000	-89.2500	76	12,835,957.00
In	IA	42.0000	-93.5000	10	3,046,355.00
In	GA	32.7500	-83.5000	88	9,687,653.00
In	FL	28.4200	-81.5500	154	18,661,493.00
In	DE	39.0000	-75.4900	9	897,934.00
In	DC	38.9100	-77.0000	1	601,723.00
In	CT	41.6600	-72.6600	29	3,574,097.00
In	AR	34.7500	-92.5000	14	2,915,918.00
In	AL	32.7500	-86.7500	28	4,779,736.00

Summary of Home Depot stores in Western USA (19 states with a total of 712 stores; total population of 106,862,007):

View Data: Sheet 43

Show aliases

In / Out of Eastern USA	State	Latitude (generated)	Longitude (generated)	H Dcount	Pop 2010
Out	WY	43.0000	-107.5000	5	563,626.00
Out	WA	47.5000	-120.5000	45	6,693,291.00
Out	UT	39.2500	-111.7500	22	2,734,230.00
Out	TX	31.2500	-99.2500	178	25,136,890.00
Out	SD	44.5000	-100.2500	1	814,180.00
Out	OR	44.0000	-120.5000	26	3,831,074.00
Out	OK	35.5000	-97.5000	16	3,751,351.00
Out	NV	39.2500	-116.7500	21	2,749,282.00
Out	NM	34.5000	-106.0000	13	2,059,179.00
Out	NE	41.5000	-99.7500	8	1,826,341.00
Out	ND	47.5000	-100.0000	1	672,591.00
Out	MT	47.0000	-109.7500	6	989,415.00
Out	KS	38.5000	-98.5000	16	2,853,118.00
Out	ID	44.0739	-114.5420	11	1,567,582.00
Out	HI	20.7900	-156.3400	7	1,360,301.00
Out	CO	39.0000	-105.5000	43	5,029,196.00
Out	CA	37.2500	-119.7500	231	37,145,229.00
Out	AZ	34.5000	-111.5000	55	6,374,900.00
Out	AK	64.0000	-150.0000	7	710,231.00

Summary of location and population for both stores:

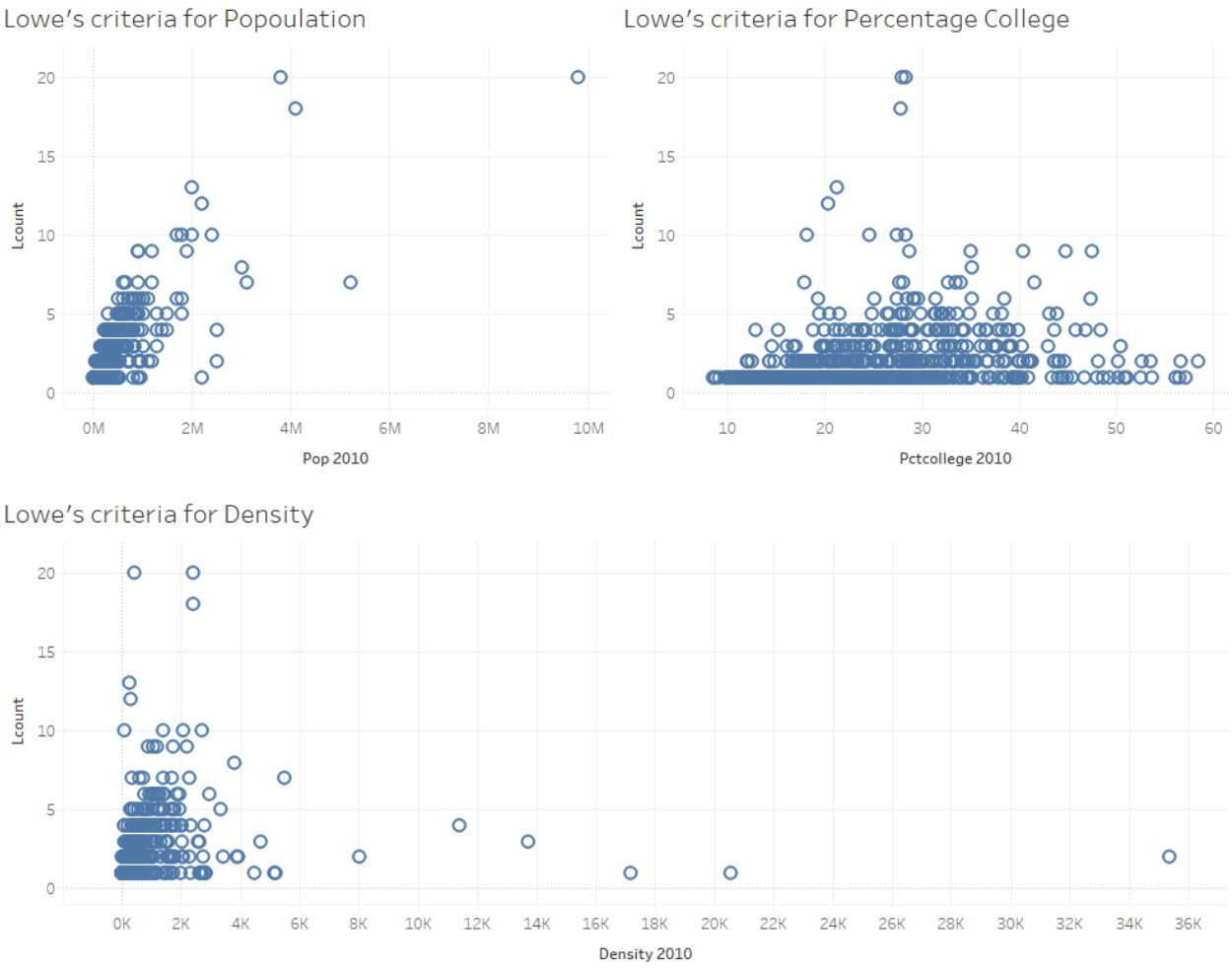
Store Name (Location)	No. of States	No. of Stores	Total population (2010)
Lowe's (Eastern USA)	31	1227	201,008,271
Home Depot (Eastern USA)	32	1240	201,609,994
Lowe's (Western USA)	18	338	81,725,117
Home Depot (Western USA)	19	712	106,862,007

Location strategies and number of stores are therefore quite alike for both Lowe's and Home Depot in East USA. However, the number of Lowe's stores on the Western part of the country is about 50% lesser than that of Home Depot stores.

The difference in the number of stores in the Eastern and Western parts of the country for both Lowe's and Home Depot is made apparent from the population differences (population is the most significant factor as explained in Q1) as identified above.

Furthermore, Home Depot serves a population which is 25,136,890 greater than that of Lowe's in the western part of the country. This in turn explains the greater number of Home Depot stores in the west.

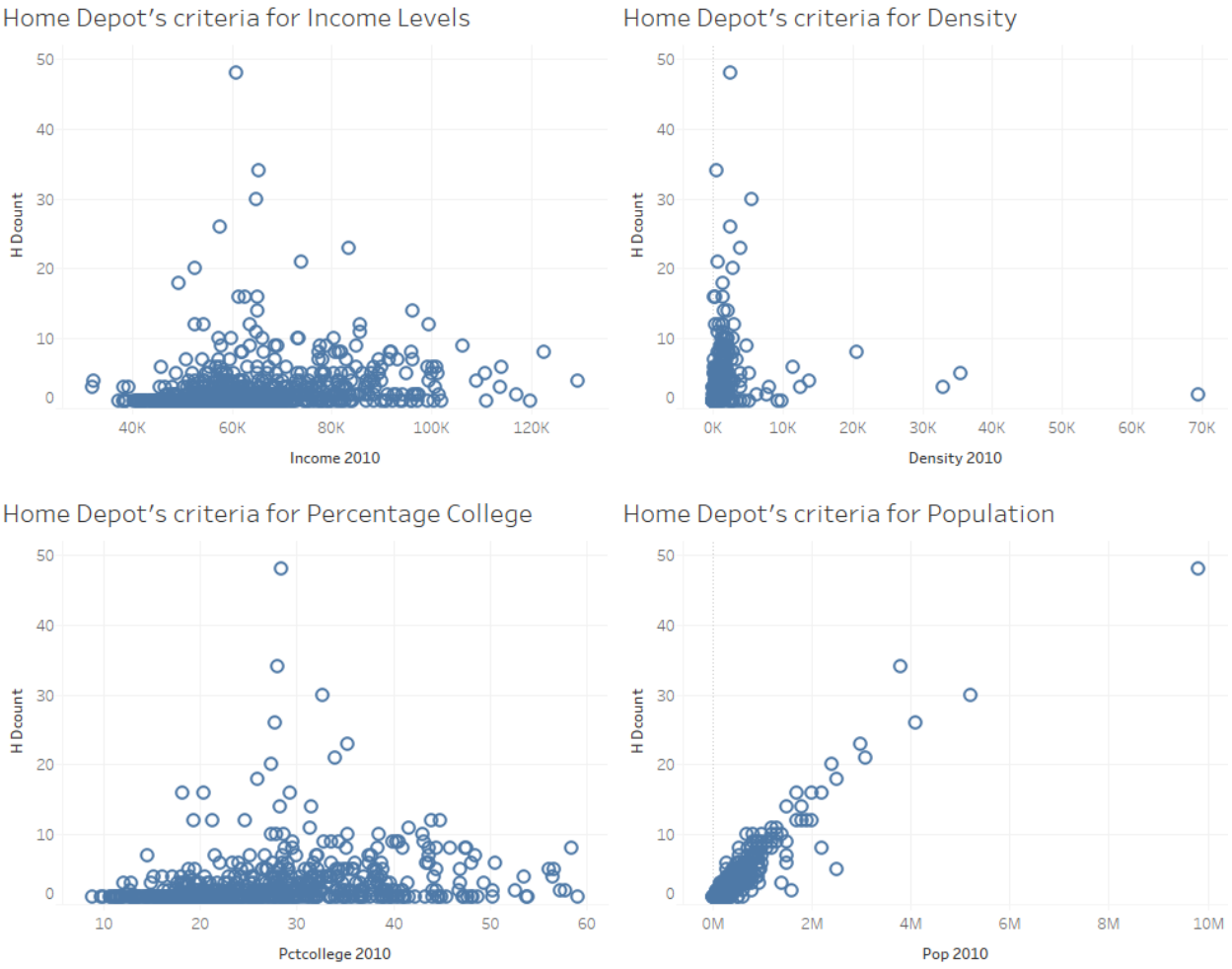
Statistics of the number Lowe's stores based on the following variables (2010):



A majority of the stores lie within the following ranges of values for each variable:

- Population : 8500 to 1 million
- Percentage college : 10% to 40%
- Density : 0 to 2000 residents/sq. mile

Statistics of the number Home Depot stores based on the following variables (2010):

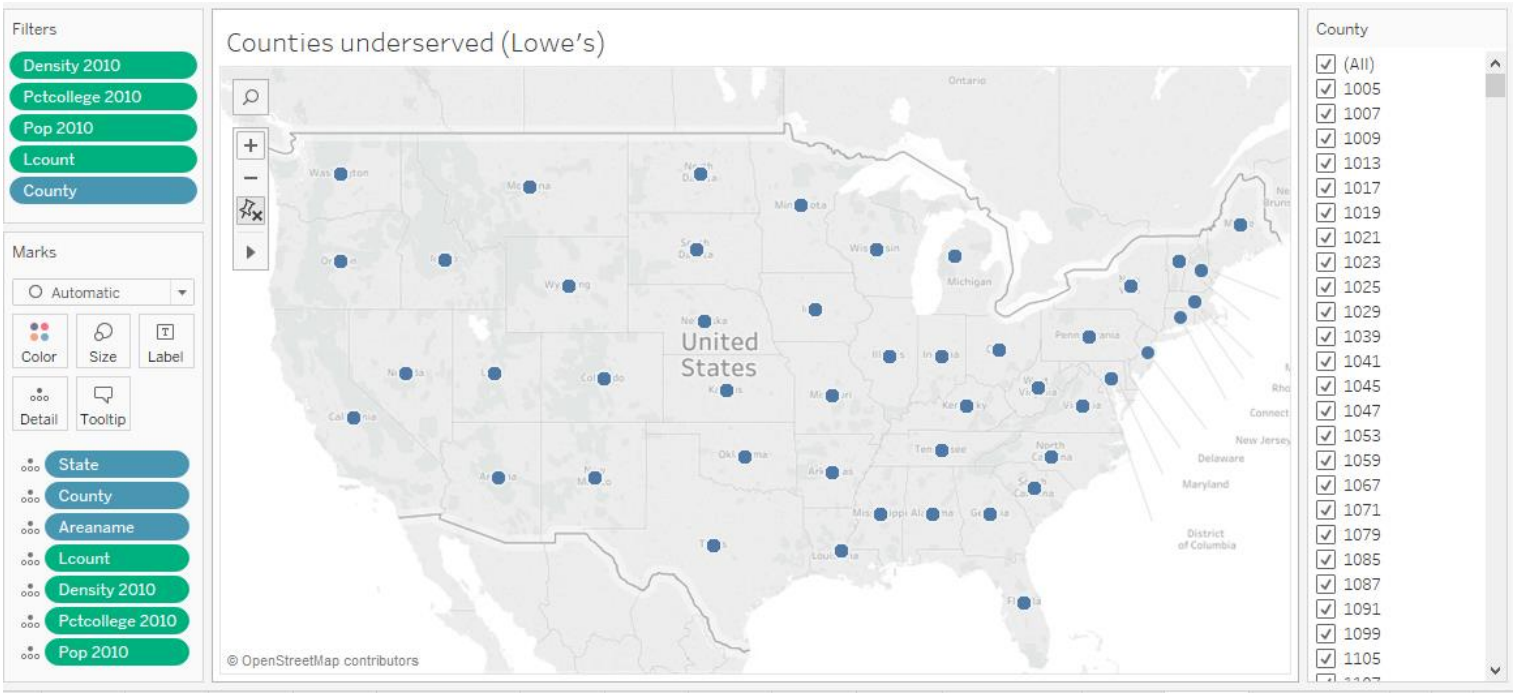


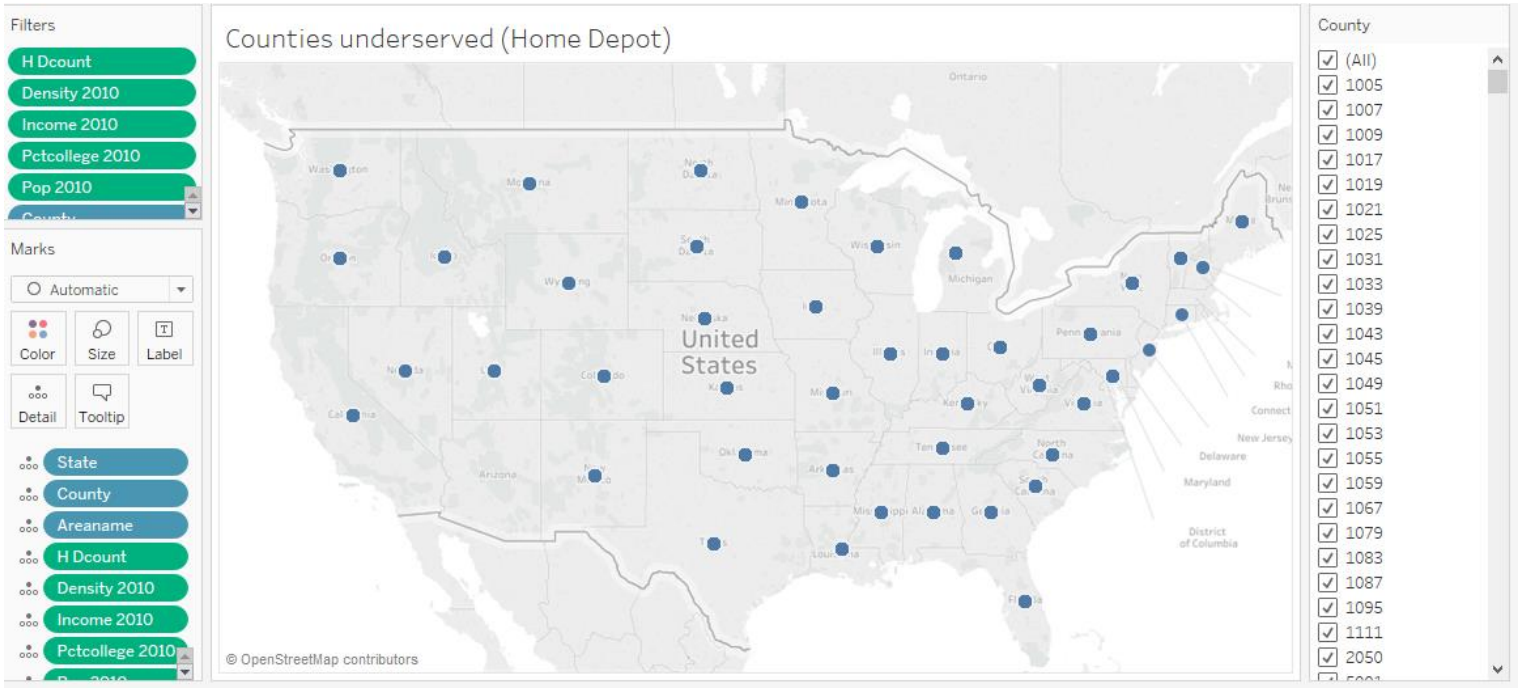
A majority of the stores lie within the following ranges of values for each variable:

- Income Levels : $40,000 to $100,000
- Density : 0 to 5000 residents/sq. mile
- Percentage college : 10% to 40%
- Population : 16,000 to 1.3 million

Q3.

The counties that are underserved assuming the range of values in Q2:





Filtering the range of values of the variables as obtained in Q2, and further manipulating the population factor (population is the most significant factor as shown in Q1) to achieve accurate results, we feel that:

- Lowe's should look to open its stores in Santa Barbara (County: 6083; population: 423,895) and Waukesha (County: 55133; population: 389,891)

Location for Lowe's to open its next stores

State	County	Areaname	
CA	6083	Santa Barbara	0 423895
	6087	Santa Cruz	0 262382
MA	25001	Barnstable	0 215888
WI	55133	Waukesha	0 389891

Lcount and Pop 2010 broken down by State, County and Areaname. Details are shown for Density 2010 and Pctcollege 2010. The view is filtered on Density 2010, Pctcollege 2010, Pop 2010 and Lcount. The Density 2010 filter ranges from 0.0 to 1000.0. The Pctcollege 2010 filter ranges from 30.000 to 40.000. The Pop 2010 filter ranges from 200000 to 1000000. The Lcount filter ranges from 0 to 0.

- Home Depot should look to open its stores in Richmond (County: 51760; population: 204,214) and Union (County: 37179; population: 201,292)

Location for Home Depot to open its next stores

State	County	Areaname	
GA	13245	Richmond	0 200549
NC	37179	Union	0 201292
VA	51760	Richmond	0 204214

H Dcount and Pop 2010 broken down by State, County and Areaname. Details are shown for Density 2010, Pctcollege 2010 and Income 2010. The view is filtered on H Dcount, Income 2010, Density 2010, Pctcollege 2010 and Pop 2010. The H Dcount filter ranges from 0 to 0. The Income 2010 filter ranges from 40000 to 100000. The Density 2010 filter ranges from 0 to 5000. The Pctcollege 2010 filter ranges from 10.00 to 40.00. The Pop 2010 filter ranges from 200000 to 1300000.