Northwestern University
Master of Science in Data Science

By: Ali Gowani

Table of Contents

Section 1:

Given the variables in this dataset, which variables can be considered explanatory (X) and which considered response (Y)? Can any variables take on both roles? What is the population of interest for this problem (yes – this is a trick question!)?

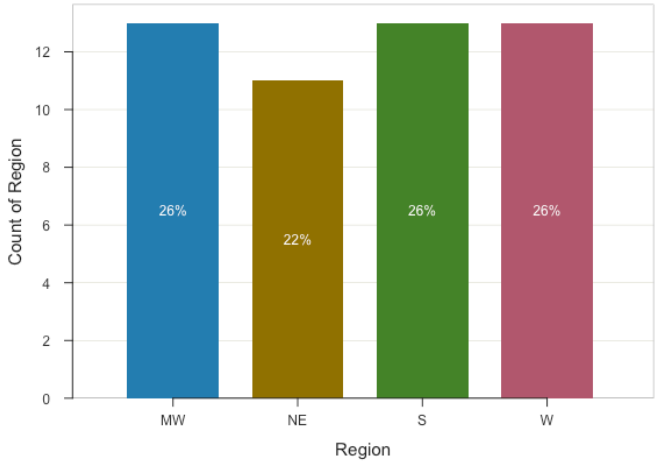
In this dataset, we are looking at 50 observations, each representing a State in the United States. There are 13 variables for each observation. The variables in this dataset can represent both explanatory variable (X) and response variable (Y). However, there are few variables, such as State, Region and Population that are considered demographic variables.

Explanatory and Response Variables:

1. HouseholdIncome
2. HighSchool
3. College
4. Smokers
5. PhysicalActivity
6. Obese
7. NonWhite
8. HeavyDrinkers
9. TwoParents
10. Insured

Demographic Variables:

1. State
    a. Indicates the each, unique observation.
2. Region



a.

b. As it can be seen in the chart above that the Region variable does not really provide any meaningful information that would be beneficial to our response variable as the data is a view at the State and Region level.

3. Population
    a. This is the number of people in each state and the corresponding variables are an average of this population.

Out dataset is quite small as it is census data at a State and Region level so it may get difficult to predict the value of a response variable in a meaningful manner. The population of interest does not exist in this dataset as the data provided is a summation of the population for the respective State and Region. See Appendix A.

Section 2:

For the duration of this assignment, let's have HOUSEHOLDINCOME be the response variable (Y). Also, please consider the STATE, REGION and POPULATION variables to be demographic variables. Obtain basic summary statistics (i.e. n, mean, std dev.) for each variable. Report these in a table. Then, obtain all possible scatterplots relating the non-demographic explanatory variables to the response variable (Y).

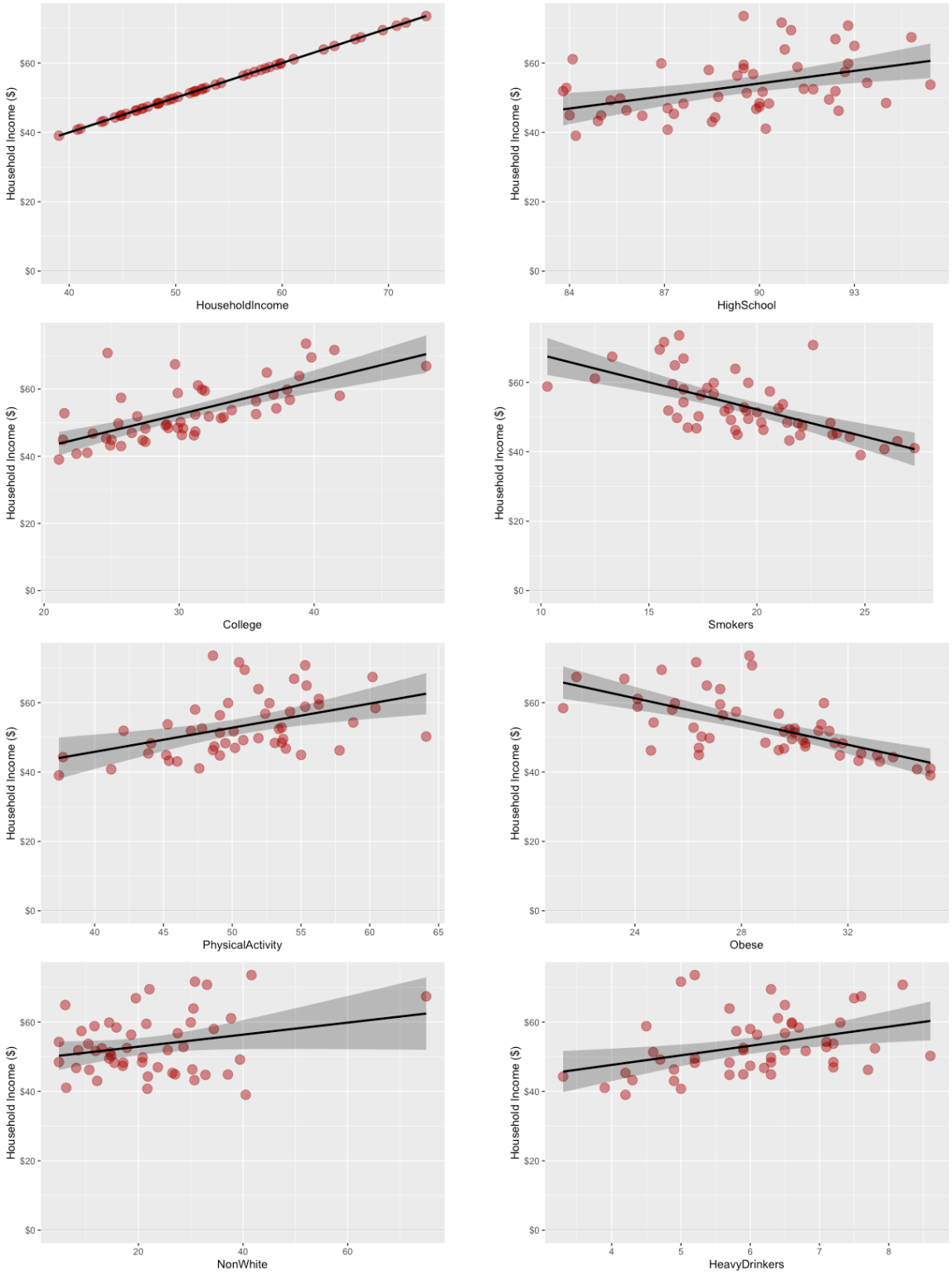
Below is the table of descriptive statistics for the 10 variables. We will not consider demographic variables in our model.

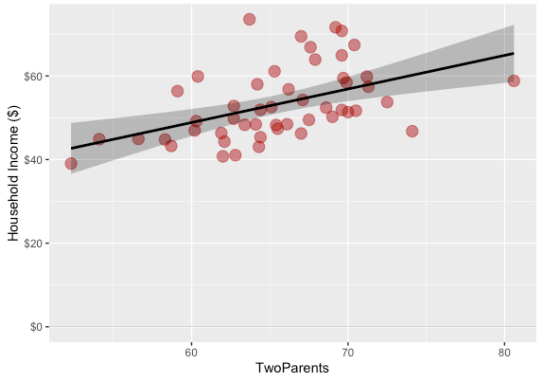
Descriptive Statistics:

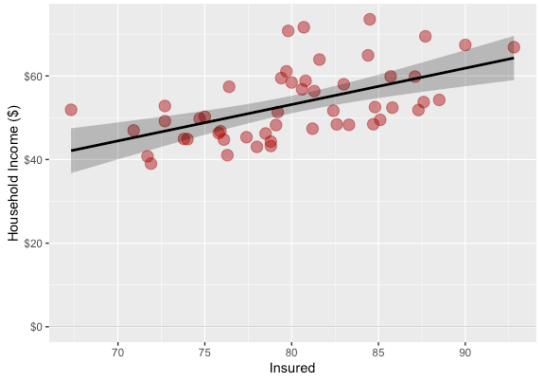
Variables	N	Std.Dev	Mean	Median	Min	Max
College	50	6.08	30.83	30.15	21.10	48.30
HeavyDrinkers	50	1.18	6.05	6.15	3.30	8.60
HighSchool	50	3.11	89.32	89.70	83.80	95.40
HouseholdIncome	50	8.69	53.28	51.76	39.03	73.54
Insured	50	5.49	80.15	79.90	67.30	92.80
NonWhite	50	12.69	22.16	20.75	4.80	75.00
Obese	50	3.37	28.77	29.40	21.30	35.10
PhysicalActivity	50	5.51	50.73	50.65	37.40	64.10
Smokers	50	3.52	19.32	19.05	10.30	27.30
TwoParents	50	5.17	65.52	65.45	52.30	80.60

Below are 9 scatter plots for the explanatory variables against the response variable HouseholdIncome. This sort of a correlation matrix allows us to determine whether a relationship, both positive or negative, exists between the specific explanatory variables and the response variable. It appears that there is a relationship with most variables. College, HeavyDrinkers, HighSchool, Insured, NonWhite, PhysicalActivity and TwoParents variables seems to have a positive linear relationship with the response variable HouseholdIncome, while Obese and Smokers variables seem to have a negative linear relationship.

Scatter Plots of Variables Compared to HouseholdIncome:





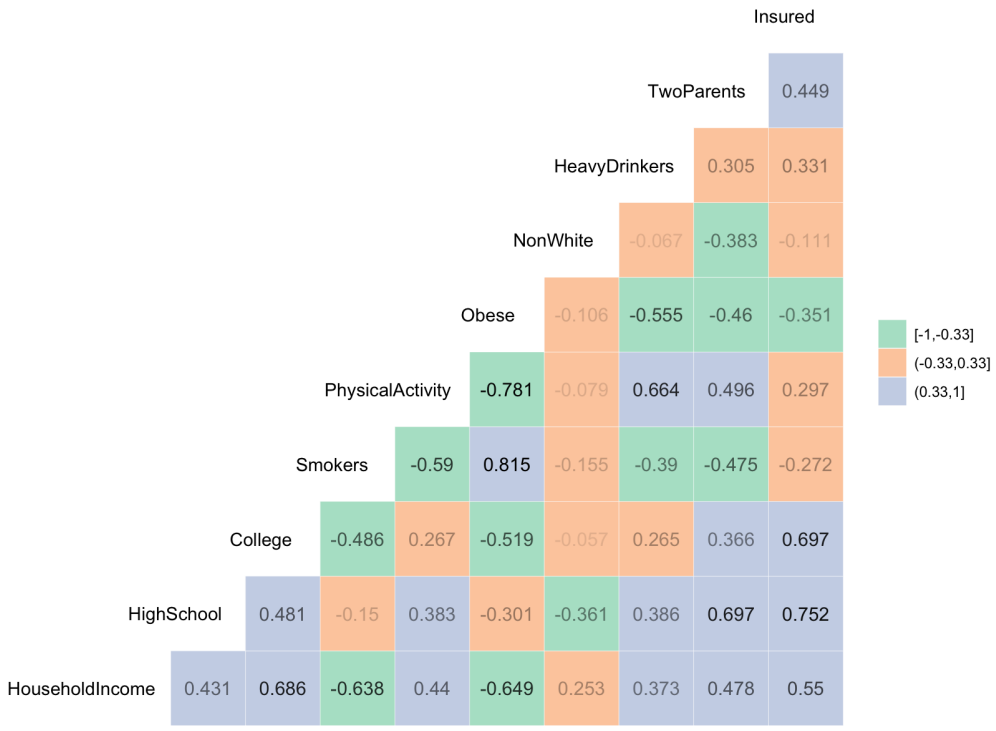


Section 3:

Obtain all possible pairwise Pearson Product Moment correlations of the non-demographic variables with Y and report the correlations in a table. Given the scatterplots from step 2) and the correlation coefficients, is simple linear regression an appropriate analytical method for this data? Why or why not?

Based on the correlation plot below, it seems that few variables (e.g.: Smoker, Obese and College) have a strong relationship to response variable HouseholdIncome. College has the strongest positive correlation of .686 and Obese has the strongest negative correlation of -.649 to HouseholdIncome.

Pearson Product Moment Correlation Plot:



Pearson Product Moment Correlation Table for HouseholdIncome:

HouseholdIncome	Vairable	corr	p_value
HouseholdIncome	College	0.6855909	0.0000026
HouseholdIncome	Obese	-0.6491116	0.0000086
HouseholdIncome	Smokers	-0.6375225	0.0000124
HouseholdIncome	Insured	0.5496786	0.0001643
HouseholdIncome	TwoParents	0.4776443	0.0010582
HouseholdIncome	PhysicalActivity	0.4404166	0.0025332
HouseholdIncome	HighSchool	0.4308448	0.0031397
HouseholdIncome	HeavyDrinkers	0.3730143	0.0105501
HouseholdIncome	NonWhite	0.2529418	0.0829036

Given the scatter plots and the correlations coefficients, we may need to think about a linear regression model (with more than 1 explanatory variable in our model) to see whether it can provide a more meaningful outcome. At this stage, I am particular interested to see how a mix of the variables positive and negative relationship (e.g.: College, Obese and Smokers) interact with our response variable.

It is interesting to note that for the most part, all explanatory variables have a linear relationship to HouseholdIncome and there does not seem to be any non-linear relationship, where the observations to those variables are all over the plot.

Section 4:

Fit a simple linear regression model to predict Y using the COLLEGE explanatory variable. Use the base STAT lm(Y~X) function. Why would you want to start with this explanatory variable? Call this Model 1. Report the results of Model 1 in equation form and interpret each coefficient of the model in the context of this problem. Report the ANOVA table and model fit statistic, R-squared. Use the summary statistics from steps 2) and 3) to verify, by hand computation, the estimates for the slope and intercept.

College is a good explanatory variable to start our model with as it has the highest correlation to our target response variable. After running our model, we can clearly see that the R^2 value of .47 is a decent place to start but we should think about whether this model can be improved by incorporating other features.

Below are the summary statistics, ANOVA table, R Squared, along with, the manual calculations of the slope and intercept for Model 1.

Model 1 Fit: Statistics

```
Call:
lm(formula = HouseholdIncome ~ College, data = mydata_sub)

Residuals:
   Min     1Q Median     3Q    Max
-7.319 -4.245 -2.203  2.652 23.484

Coefficients:
            Estimate Std. Error t value     Pr(>|t|)
(Intercept)  23.0664     4.7187   4.888 0.0000117800
College       0.9801     0.1502   6.525 0.0000000394

Residual standard error: 6.392 on 48 degrees of freedom
Multiple R-squared:   0.47,     Adjusted R-squared:  0.459
F-statistic: 42.57 on 1 and 48 DF,  p-value: 0.00000003941
```

Model 1 Fit: ANOVA Table

```
Analysis of Variance Table

Response: HouseholdIncome
           Df Sum Sq Mean Sq F value        Pr(>F)
College     1 1739.4 1739.36  42.572 0.00000003941
Residuals 48 1961.1   40.86
```

Model 1: R Squared

```
Model 1 R Squared:  0.4700349
```

Model 1: Slope

```
> cat("Model 1 Slope: ", cor(mydata_sub$College, mydata_sub$HouseholdIncome) * (sd
(mydata_sub$HouseholdIncome) / sd(mydata_sub$College)))
Model 1 Slope:  0.9801441
```

Model 1: Intercept

```
> mean(mydata_sub$HouseholdIncome) - (model1_slope * mean(mydata_sub$Colleg
e))
[1] 23.06644
```

Section 5:

Write R-code to calculate and create a variable of predicted values based on Model 1. Use the predicted values and the original response variable Y to calculate and create a variable of residuals (i.e. residual = Y

– Y_hat = observed minus predicted) for Model 1. Using the original Y variable, the predicted, and/or residual variables, write R-code to:

- *Square each of the residuals and then add them up. This is called sum of squared residuals, or sums of squared errors.*
- *Deviate the mean of the Y's from the value of Y for each record (i.e. Y – Y_bar). Square each of the deviations and then add them up. This is called sum of squares total.*
- *Deviate the mean of the Y's from the value of predicted (Y_hat) for each record (i.e. Y_hat – Y_bar). Square each of these deviations and then add them up. This is called the sum of squares due to regression.*
- *Calculate a statistic that is: (Sum of Squares due to Regression) / (Sum of squares Total)*

Verify and note the accuracy of the ANOVA table and R-squared values from the regression printout from part 4), relative to your computations here.

SSE (Sums of Squared Errors) or Sum of Squared Residuals:

```
> mydata_sub_sec5$model_1_pred <- predict.lm(model1, mydata_sub)
> mydata_sub_sec5$model_1_residual <- mydata_sub_sec5$HouseholdIncome - mydata_sub
_sec5$model_1_pred
> mydata_sub_sec5$model_1_residual_sqrd <- mydata_sub_sec5$model_1_residual ^ 2
> sum(mydata_sub_sec5$model_1_residual_sqrd)
[1] 1961.13
```

SST (Sum of Squares Totals):

```
> mydata_sub_sec5$model_1_mean_dev <-mydata_sub_sec5$HouseholdIncome - mean(mydata
_sub_sec5$HouseholdIncome)
> mydata_sub_sec5$model_1_mean_dev_sq <- mydata_sub_sec5$model_1_mean_dev ^ 2
> sum(mydata_sub_sec5$model_1_mean_dev_sq)
[1] 3700.488
```

SSR (Sum of Squares Residuals):

```
> mydata_sub_sec5$model_1_yhat_devbar <- mydata_sub_sec5$model_1_pred - mean(mydat
a_sub_sec5$HouseholdIncome)
> mydata_sub_sec5$model_1_yhat_dev_sq <- mydata_sub_sec5$model_1_yhat_devbar ^ 2
> sum(mydata_sub_sec5$model_1_yhat_dev_sq)
[1] 1739.359
```

R^2 (R Squared):

```
> sum(mydata_sub_sec5$model_1_yhat_dev_sq) / sum(mydata_sub_sec5$model_1_mean_dev_
sq)
[1] 0.4700349
```

We can see that R Squared value of 0.47 and the ANOVA table (above) matches the regression output from Section 4. In general, the results show that the R Squared value explains 47% of the variability in our model. In addition, with a low p-value, we can reject the null hypothesis and there is a stronger evidence to favor the alternative hypothesis. The high t-value give us enough evidence to also reject the null hypothesis.

Lastly, the coefficient for College variable is .98 and this means that for every 1 unit increase in College, we can expect to see a .98 increase in HouseholdIncome. In general, this makes sense as College educated population would mean higher HouseholdIncome. However, we should be careful as it does not mean that if people continue to get educated that their HouseholdIncome will rise indefinitely. Conversely, it does not mean that a HouseholdIncome of greater than 0 cannot exist without a College degree. We should also, do a "sense check" when evaluating these statistics and understand them with the proper context.

Section 6:

Fit a multiple linear regression model to predict Y using COLLEGE and INSURED as the explanatory variables. Use the base lm(Y~X) function. Call this Model 2. Report the results of Model 2 in equation form, interpret each coefficient of the model in the context of this problem, and report the model fit statistic, R-squared. How have the coefficients and their interpretations changed? Calculate the change in R-squared from Model 1 to Model 2 and interpret this value. For this specific problem, is it OK to use the hypothesis testing results to determine if the additional explanatory variable should be retained or not? Think statistically using first principals. Discuss. NOTE: The topic of hypothesis testing in regression is the focus of Module 2 – you should NOT need to read anything about hypothesis testing to answer this.

We have now incorporated an additional explanatory variable of Insured into our Model 1, which included College. This model is interesting for two reasons: the t-value of .651 in Model 2 compared to 4.88 in Model 1 and the high p-value of .518 in Model 2 compared to .00001 in Model 1. Unlike Model 1, there is not enough evidence for us to reject the null hypothesis.

In this model, we added a new variable thinking it would provide additional information to make a greater impact to our model but when we compared the R Squared values, we do not see this to hold true. When we only had 1 variable of College, the R Squared value was .47 and then we added Insured to our model. By doing so, we would think that adding a new variable should improve our R Squared value, but it only made an impact of .01, to a total of .48. Therefore, we can conclude that Insured is probably not a worthwhile variable to consider in future models as we evaluate explanatory variables.

Summary Model 2: coefficients, fit statistic, r-squared, etc.:

```
> summary(model2)

Call:
lm(formula = HouseholdIncome ~ College + Insured, data = mydata_sub)

Residuals:
   Min     1Q Median     3Q    Max
-6.918 -4.545 -2.125  4.357 22.709

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.6728    14.8628   0.651 0.518339
College       0.8411     0.2098   4.010 0.000216
Insured       0.2206     0.2321   0.950 0.346759

Residual standard error: 6.398 on 47 degrees of freedom
Multiple R-squared:   0.48,     Adjusted R-squared:  0.4579
F-statistic: 21.69 on 2 and 47 DF,  p-value: 0.0000002116
```

R-Squared: Model 1 vs Model 2

```
> summary(model2)$r.squared - summary(model1)$r.squared
[1] 0.009993449
```

Section 7:

In a sequential fashion, continue to add in the non-demographic variables into the prediction model, one variable at a time. Make a table summarizing the change in R-squared that is associated with each variable added. Based on this information, what variables should be retained for a "best" predictive model? What criteria seems appropriate to you?

During this problem, practice interpreting coefficients for each model. Do any of the interpretations become counter intuitive as you fit more and more complex models? What does, or would, this mean for the model being developed? You do not need to report all of the coefficient interpretations, but this is a general question to contemplate and skill to use in model determination. Please write a short summary of your conclusions here.

It is interesting to note that as we add more features into our model, that our R Squared value increases. This makes sense as we are adding more information to our model and expecting it to improve its predication capabilities for the target response variable HouseholdIncome. R Squared assumes that every explanatory variable explains the variation in the target response variable. However, Adjusted R

Squared gives a percentage for the variation explained by only those variables that actually impact the target response variable.

As the table below shows, the more explanatory variables we add, the higher our R Squared value increases. However, our Adjusted R Squared value increases initially as we add more explanatory variables to explain the variation in our model, but it starts to decreases after the third variable. This indicates that adding more features (in this order) to our model does not help explain the variation to our target response variable and can actually, decrease it. We should evaluate different combinations of explanatory variables to see how the Adjusted R Squared value gets impacted.

Table: R^2 and Adjusted R^2 for each additional variable

Features	Features_R2	Features_R2_Adjusted
College	0.4700	0.4590
College + Insured	0.4800	0.4579
College + Insured + Smokers	0.6104	0.5850
College + Insured + Smokers + PhysicalActivity	0.6136	0.5793
College + Insured + Smokers + PhysicalActivity + TwoParents	0.6184	0.5751
College + Insured + Smokers + PhysicalActivity + TwoParents + HeavyDrinkers	0.6204	0.5674
College + Insured + Smokers + PhysicalActivity + TwoParents + HeavyDrinkers + Obese	0.6309	0.5694
College + Insured + Smokers + PhysicalActivity + TwoParents + HeavyDrinkers + Obese + Highschool	0.6311	0.5591

Section 8:

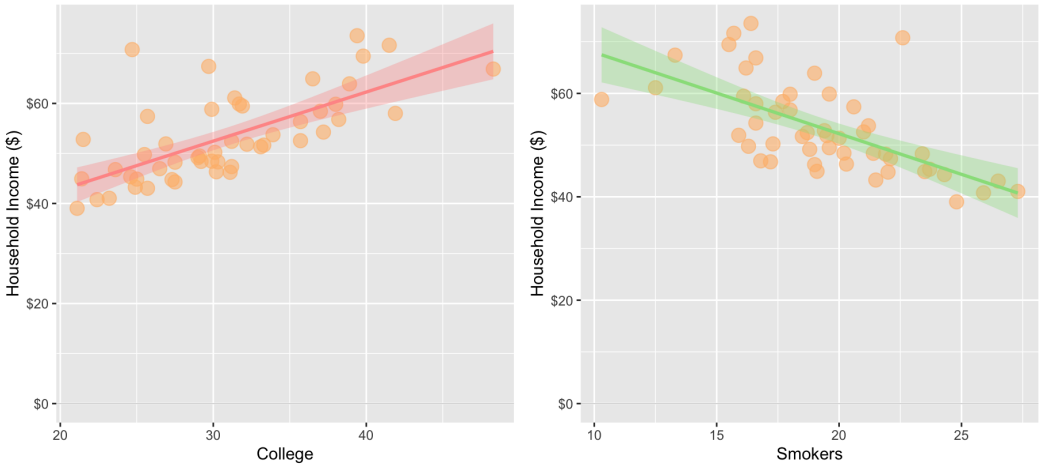
Now that you have a sense of which explanatory variables contribute to explaining HOUSEHOLDINCOME, refit a model using only the set of variables you consider to be appropriate to model Y. Report this model, interpret the coefficients, and interpret R-squared in the context of this problem. Discuss why is it necessary to refit this model.

As I was looking through the correlation plot and the scatter plots, I thought that we should look at three variables in particular: College, Smokers and Obese. I initially ran the model with two explanatory variables: College and Obese. I thought that with the higher variability of Obese values (-.649) to the HouseholdIncome that it would be better for our model. I was going to stop there but then thought, let me try to run a quick model with College and Smokers as explanatory variables. To my surprise, it produced a better model with a higher Adjusted R Squared value.

I also ran some random models, but they were producing results that did not make sense, as some coefficients became negative or not significant.

I was finished with this assignment but then a thought came to mind: since we only have a small number of observations, can't we run through all the combinations of 2 or 3 explanatory variables to see which yielded a positive result to a more robust linear regression model? See Section 9! ☺

Plot: Refit Model with College and Smokers variables



Summary Statistics: Refit Model with College and Smokers variables

```
Call:
lm(formula = HouseholdIncome ~ College + Smokers, data = mydata_sub)

Residuals:
    Min      1Q  Median      3Q     Max
-7.5549 -3.2223 -1.7403  0.7376 25.0169

Coefficients:
            Estimate Std. Error t value    Pr(>|t|)
(Intercept)  50.5892     8.4703   5.973 0.000000296 ***
College       0.7035     0.1525   4.614 0.000030619 ***
Smokers      -0.9832     0.2631  -3.738    0.000503 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.671 on 47 degrees of freedom
Multiple R-squared:  0.5915,    Adjusted R-squared:  0.5741
F-statistic: 34.02 on 2 and 47 DF,  p-value: 0.0000000007305
```

Section 9:

There are 9 variables not including our target response variable. Since we had only used 2 explanatory variables in our previous two models that we should try a model with three such variables. In addition, when we were adding variables, our table showed the Adjusted R Square value increasing until the third variable. If we tried three explanatory variables in our model, then there would be 84 different combinations (as order does not matter). I did not think I would have enough time to try all of them or randomly guess which of three variables would yield the best result.

After trying many packages and hitting errors, I was able to parse some code to go through the combinations of different variables to yield the highest Adjusted R Squared value. See Appendix B. To my surprise, the variables that yield the best result were not the ones I was thinking about: NonWhite and TwoParents. The explanatory variable of College seemed obvious and our initial exploratory data analysis showed that it would be meaningful to our model. However, NonWhite and TwoParents completely threw me off and I don't think I would have even guessed including them in my model.

By using these three variables to predict our target response variable of HouseholdIncome, our model was able to generate an Adjusted R Squared value of .688. It also generated a low p-value of .0089 so we were able to reject our null hypothesis as this indicates that there is a significant difference between our factors.

Plot: Refit Model with College, NonWhite and TwoParents variables

Summary Statistics: Refit Model with College, NoneWhite and TwoParents variables

```
Call:
lm(formula = HouseholdIncome ~ College + TwoParents + NonWhite,
    data = mydata_sub)

Residuals:
   Min     1Q Median     3Q    Max
-6.753 -3.452 -1.170  2.152 15.723

Coefficients:
             Estimate Std. Error t value     Pr(>|t|)
(Intercept) -27.64061   10.12129  -2.731      0.00892 **
College       0.78053    0.12326   6.332 0.0000000917 ***
TwoParents    0.76175    0.15661   4.864 0.0000138476 ***
NonWhite      0.31360    0.05951   5.270 0.0000035318 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.858 on 46 degrees of freedom
Multiple R-squared:  0.7066,    Adjusted R-squared:  0.6875
F-statistic: 36.93 on 3 and 46 DF,  p-value: 0.000000000002633
```

Section 10:

Given what you've learned from this modeling endeavor, what overall conclusions do you draw? What is the "Story" contained in this data? What have you learned? What are your Prescriptive Recommendations for action based on this evidence? Finally, feel free to reflect on what you've learned from a modeling perspective.

Conclusion:

In this assignment, our focus was to learn linear regression and to how to interpret the summary statistics of different models. We learned about truly understanding R Squared, Adjusted R Squared, T value, P value, coefficients, slopes, intercepts, etc. and how they interact with each other. We executed a correlation plot and created a table against our target variable HouseholdIncome to get a sense of which variables might explain the model the best. Our initial model only included one variable but then we added more explanatory variables to see how this would impact our model and whether it would add meaningful information to generate a more robust model. We evaluated whether it helped explain the variation in our model by looking at the R Squared and Adjusted R Squared values. In regards to the intercept, we learned that a negative value may not make much sense with our dataset. We needed to keep in mind the range of our dataset to make sure we understood the results of our model and can apply them with context. While we evaluated various variables to include in our model, we truly need to

understand how each variable and set of variables interact with the target response variable to build the best model.

Reflection:

Wow! I am spending way too much time on these assignments but am really enjoying them and learning a great deal.

My perception initially was that we can look at the correlation plot to select the variables with the highest correlation, whether negatively or positively correlated compared to our target variable of HouseholdIncome. As we got deeper into the assignment, it made me realize that perhaps, what I perceived to be good predictors may not be ideal as I too can be biased in my understanding. In addition, variables with the highest correlation may not yield the best model as the interactions between the explanatory variables might be more meaningful than just 1-on-1 comparison of variables to the response variable.

I really enjoyed learning how to code through the iteration of all the possible combinations of the variables against the target response variable. I realize that this may not always be possible when we may have more than 20 or 30 variables and hundreds of thousands or millions of observations. Perhaps instead of randomly picking explanatory variables, we break the dataset into a smaller set and run the combinations of all the features on it to see what the results show and see if it questions our biases and assumptions.

In the end, I learned to apply the various statistics with some context and learned that biases in our understanding can limit the robustness of our model.

Appendix

A: Data Quality Check

Dimensions: 50 x 13

Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Missing
1	State [factor]	1. Alabama 2. Alaska 3. Arizona 4. Arkansas 5. California 6. Colorado 7. Connecticut 8. Delaware 9. Florida 10. Georgia [40 others]	1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 1 (2.0%) 40 (80.0%)	0 (0%)
2	Region [factor]	1. MW 2. NE 3. S 4. W	13 (26.0%) 11 (22.0%) 13 (26.0%) 13 (26.0%)	0 (0%)
3	Population [numeric]	Mean (sd) : 6.4 (7.2) min < med < max: 0.6 < 4.5 < 38.8 IQR (CV) : 5.1 (1.1)	50 distinct values	0 (0%)
4	HouseholdIncome [numeric]	Mean (sd) : 53.3 (8.7) min < med < max: 39 < 51.8 < 73.5 IQR (CV) : 11.9 (0.2)	50 distinct values	0 (0%)
5	HighSchool [numeric]	Mean (sd) : 89.3 (3.1) min < med < max: 83.8 < 89.7 < 95.4 IQR (CV) : 4.5 (0)	44 distinct values	0 (0%)
6	College [numeric]	Mean (sd) : 30.8 (6.1) min < med < max: 21.1 < 30.1 < 48.3 IQR (CV) : 9.4 (0.2)	45 distinct values	0 (0%)
7	Smokers [numeric]	Mean (sd) : 19.3 (3.5) min < med < max: 10.3 < 19.1 < 27.3 IQR (CV) : 4.8 (0.2)	45 distinct values	0 (0%)
8	PhysicalActivity [numeric]	Mean (sd) : 50.7 (5.5) min < med < max: 37.4 < 50.6 < 64.1 IQR (CV) : 6.5 (0.1)	43 distinct values	0 (0%)
9	Obese [numeric]	Mean (sd) : 28.8 (3.4) min < med < max: 21.3 < 29.4 < 35.1 IQR (CV) : 4.7 (0.1)	42 distinct values	0 (0%)
10	NonWhite [numeric]	Mean (sd) : 22.2 (12.7) min < med < max: 4.8 < 20.8 < 75 IQR (CV) : 16.9 (0.6)	48 distinct values	0 (0%)
11	HeavyDrinkers [numeric]	Mean (sd) : 6 (1.2) min < med < max: 3.3 < 6.2 < 8.6 IQR (CV) : 1.6 (0.2)	31 distinct values	0 (0%)
12	TwoParents [numeric]	Mean (sd) : 65.5 (5.2) min < med < max: 52.3 < 65.5 < 80.6 IQR (CV) : 6.8 (0.1)	45 distinct values	0 (0%)
13	Insured [numeric]	Mean (sd) : 80.1 (5.5) min < med < max: 67.3 < 79.9 < 92.8 IQR (CV) : 8.3 (0.1)	48 distinct values	0 (0%)

B: Combination of 3 Variables (Predictors)

Table with variations of 3 different predictors compared to our target variable (HouseholdIncome) and sorted by Adjusted R^2.

(Intercept)	College	HeavyDrinkers	HighSchool	Insured	NonWhite	Obese	PhysicalActivity	Smokers	TwoParents	R^2	Adjusted R^2
-27.6406	0.7805	NA	NA	NA	0.3136	NA	NA	NA	0.7618	0.7066	0.6875
53.2857	0.7361	NA	NA	NA	0.1676	-0.9180	NA	NA	NA	0.6459	0.6228
42.7149	0.7605	NA	NA	NA	0.1576	NA	NA	-0.8474	NA	0.6419	0.6186
-2.8228	0.8916	NA	NA	NA	0.2139	NA	0.4707	NA	NA	0.6380	0.6144
-47.3595	NA	NA	1.3408	NA	0.2376	NA	NA	-1.2623	NA	0.6205	0.5958
-53.2325	0.8041	NA	0.8479	NA	0.2704	NA	NA	NA	NA	0.6161	0.5910
11.4582	0.5685	NA	0.4950	NA	NA	NA	NA	-1.0308	NA	0.6153	0.5902
50.5892	0.7035	NA	NA	NA	NA	NA	NA	-0.9832	NA	0.5915	0.5741
33.2208	0.4997	NA	NA	0.3050	NA	NA	NA	-1.0243	NA	0.6104	0.5850
16.5171	NA	NA	NA	0.7018	0.1558	NA	NA	-1.1873	NA	0.6083	0.5827
62.0310	0.6823	NA	NA	NA	NA	-1.0353	NA	NA	NA	0.5877	0.5702
33.1641	0.6671	NA	NA	NA	NA	NA	NA	-0.8447	0.2422	0.6071	0.5814
60.6642	0.6531	NA	NA	NA	NA	-0.5622	NA	-0.5873	NA	0.6065	0.5808
41.9620	0.6439	NA	NA	NA	NA	-0.8880	NA	NA	0.2597	0.6060	0.5804
10.3200	0.9207	1.6465	NA	NA	0.2087	NA	NA	NA	NA	0.6017	0.5757
44.7182	0.6889	0.7659	NA	NA	NA	NA	NA	-0.8957	NA	0.6005	0.5744
38.1780	0.7075	NA	NA	NA	NA	NA	0.1782	-0.8153	NA	0.5998	0.5737

48.0362	0.5332	NA	NA	0.2340	NA	-1.0409	NA	NA	NA	0.5990	0.5728
33.2738	NA	NA	NA	0.6466	0.1694	-1.2363	NA	NA	NA	0.5984	0.5722
37.2613	0.6156	NA	0.2942	NA	NA	-1.0161	NA	NA	NA	0.5962	0.5699
58.9863	0.6838	0.2544	NA	NA	NA	-0.9846	NA	NA	NA	0.5886	0.5617
62.2616	0.6819	NA	NA	NA	NA	-1.0387	-0.0024	NA	NA	0.5877	0.5609
-18.1654	NA	NA	1.1464	NA	0.2392	-1.2600	NA	NA	NA	0.5820	0.5548
26.9023	NA	NA	NA	0.6424	NA	NA	NA	-1.2996	NA	0.5591	0.5404
40.2192	NA	NA	NA	0.5888	NA	-0.6262	NA	-0.8344	NA	0.5779	0.5503
17.8781	1.0041	NA	NA	NA	0.2008	NA	NA	NA	NA	0.5557	0.5368
0.2038	0.8241	NA	NA	0.2873	0.2097	NA	NA	NA	NA	0.5725	0.5446
30.4007	NA	NA	NA	NA	0.2569	-1.0583	NA	NA	0.7270	0.5678	0.5396
-56.6096	NA	NA	NA	0.6052	0.3311	NA	NA	NA	0.8249	0.5670	0.5388
10.8082	NA	NA	0.3008	0.5117	NA	NA	NA	-1.3153	NA	0.5641	0.5357
4.1708	0.8743	NA	NA	NA	NA	NA	0.4368	NA	NA	0.5412	0.5217
21.2149	NA	NA	NA	0.6056	NA	NA	NA	-1.2363	0.1131	0.5621	0.5336
45.3566	NA	NA	NA	0.5805	NA	-1.3419	NA	NA	NA	0.5394	0.5198
-6.6844	0.8141	NA	NA	NA	NA	NA	0.3234	NA	0.2818	0.5608	0.5321
25.6833	NA	0.3066	NA	0.6265	NA	NA	NA	-1.2664	NA	0.5605	0.5318
25.6835	NA	NA	NA	0.6391	NA	NA	0.0220	-1.2807	NA	0.5592	0.5305
71.0392	NA	NA	NA	0.5885	NA	-1.7234	-0.3024	NA	NA	0.5538	0.5247

-1.5490	0.8432	NA	NA	NA	NA	NA	NA	NA	0.4401	0.5294	0.5094
-3.1820	0.8037	1.1511	NA	NA	NA	NA	NA	NA	0.3774	0.5508	0.5215
35.4046	NA	NA	NA	0.5256	NA	-1.2478	NA	NA	0.1778	0.5472	0.5177
-3.0305	0.7964	NA	NA	0.1296	NA	NA	0.4214	NA	NA	0.5446	0.5149
-4.4461	NA	NA	0.9589	NA	NA	NA	NA	-1.4456	NA	0.5213	0.5010
-5.1183	0.8485	NA	0.1237	NA	NA	NA	0.4177	NA	NA	0.5426	0.5128
4.6911	0.8677	0.3641	NA	NA	NA	NA	0.3871	NA	NA	0.5426	0.5127
48.8076	NA	-0.4105	NA	0.5942	NA	-1.4135	NA	NA	NA	0.5415	0.5116
46.0455	NA	NA	-0.0121	0.5855	NA	-1.3423	NA	NA	NA	0.5394	0.5094
16.9850	NA	NA	NA	NA	0.2493	NA	NA	-0.9169	0.7400	0.5393	0.5093
14.1057	NA	NA	0.8437	NA	NA	-0.5805	NA	-1.0085	NA	0.5367	0.5065
16.2683	0.9022	1.5219	NA	NA	NA	NA	NA	NA	NA	0.5094	0.4886
11.7472	0.8720	NA	-0.2161	NA	NA	NA	NA	NA	0.5183	0.5322	0.5016
-4.8612	0.8043	NA	NA	0.0699	NA	NA	NA	NA	0.4235	0.5304	0.4997
-7.8384	NA	NA	1.1505	NA	NA	NA	NA	-1.5457	-0.1798	0.5256	0.4947
-2.4369	NA	NA	1.0137	NA	NA	NA	-0.1027	-1.5331	NA	0.5237	0.4927
-4.0833	NA	0.1172	0.9441	NA	NA	NA	NA	-1.4323	NA	0.5215	0.4903
23.0664	0.9801	NA	NA	NA	NA	NA	NA	NA	NA	0.4700	0.4590
8.5781	0.8229	1.4235	NA	0.1339	NA	NA	NA	NA	NA	0.5129	0.4812

1.1652	0.8622	1.3801	0.1925	NA	NA	NA	NA	NA	NA	0.5127	0.4809
57.7565	NA	NA	0.8527	NA	NA	-2.0109	-0.4493	NA	NA	0.5120	0.4801
-7.0506	0.8896	NA	0.3684	NA	NA	NA	NA	NA	NA	0.4834	0.4614
31.0230	NA	NA	0.7237	NA	NA	-1.4731	NA	NA	NA	0.4822	0.4602
9.6728	0.8411	NA	NA	0.2206	NA	NA	NA	NA	NA	0.4800	0.4579
-30.3565	NA	1.6512	NA	NA	0.3410	NA	NA	NA	1.0088	0.4959	0.4630
-39.7125	NA	NA	NA	0.7770	0.2279	NA	0.5061	NA	NA	0.4955	0.4626
-80.6881	NA	NA	0.8118	NA	0.3720	NA	NA	NA	0.8123	0.4932	0.4602
68.6833	NA	NA	NA	NA	NA	-1.4048	NA	NA	0.3817	0.4620	0.4391
98.6637	NA	NA	NA	NA	NA	-1.8847	-0.4187	NA	0.4591	0.4879	0.4545
32.4184	NA	-0.5091	0.7701	NA	NA	-1.5588	NA	NA	NA	0.4852	0.4517
32.8013	NA	NA	0.5903	NA	NA	-1.4181	NA	NA	0.1305	0.4849	0.4513
70.1671	NA	NA	NA	NA	NA	-0.8873	NA	-0.6544	0.3248	0.4848	0.4512
-5.4286	0.8533	NA	0.2802	0.0921	NA	NA	NA	NA	NA	0.4844	0.4507
93.9102	NA	NA	NA	NA	0.1138	-1.0208	NA	-0.7135	NA	0.4833	0.4496
97.3040	NA	NA	NA	NA	NA	-0.9950	NA	-0.7971	NA	0.4564	0.4333
97.1538	NA	NA	NA	NA	0.1275	-1.6233	NA	NA	NA	0.4556	0.4324
-32.1894	NA	NA	NA	NA	0.3323	NA	0.3135	NA	0.9494	0.4800	0.4461

-28.6159	NA	NA	NA	NA	0.3499	NA	NA	NA	1.1316	0.4508	0.4274
101.4449	NA	NA	NA	NA	NA	-1.6742	NA	NA	NA	0.4213	0.4093
53.7155	NA	NA	NA	NA	NA	NA	NA	-1.3081	0.3790	0.4458	0.4222
116.0258	NA	NA	NA	NA	NA	-1.3033	-0.2114	-0.7520	NA	0.4633	0.4283
68.5133	NA	0.0164	NA	NA	NA	-1.4018	NA	NA	0.3815	0.4620	0.4269
83.6593	NA	NA	NA	NA	NA	NA	NA	-1.5725	NA	0.4064	0.3941
124.8759	NA	NA	NA	NA	NA	-2.0162	-0.2679	NA	NA	0.4326	0.4085
111.2367	NA	NA	NA	NA	0.1163	-1.8278	-0.1567	NA	NA	0.4592	0.4239
80.0969	NA	NA	NA	NA	0.1081	NA	NA	-1.5121	NA	0.4308	0.4065
93.0623	NA	0.3554	NA	NA	NA	-0.9076	NA	-0.8190	NA	0.4580	0.4227
92.4136	NA	0.3946	NA	NA	0.1321	-1.5450	NA	NA	NA	0.4575	0.4221
48.6839	NA	0.8673	NA	NA	NA	NA	NA	-1.2162	0.3487	0.4572	0.4218
68.6031	NA	1.2829	NA	NA	0.1236	NA	NA	-1.3364	NA	0.4558	0.4203
74.3896	NA	1.0827	NA	NA	NA	NA	NA	-1.4315	NA	0.4246	0.4001
-97.3102	NA	NA	1.3482	NA	0.3085	NA	0.4601	NA	NA	0.4500	0.4142
99.8665	NA	0.1359	NA	NA	NA	-1.6479	NA	NA	NA	0.4216	0.3970
-26.5007	NA	1.6673	NA	0.8082	0.2223	NA	NA	NA	NA	0.4471	0.4110
51.8389	NA	NA	NA	NA	NA	NA	0.0416	-1.2783	0.3667	0.4462	0.4101
63.6763	NA	NA	NA	NA	0.1257	NA	0.2311	-1.2890	NA	0.4441	0.4078
73.0171	NA	NA	NA	NA	NA	NA	0.1552	-1.4293	NA	0.4127	0.3878

124.4647	NA	0.6907	NA	NA	NA	-1.9936	-0.3549	NA	NA	0.4375	0.4008
-74.0688	NA	NA	0.8316	0.5861	0.2750	NA	NA	NA	NA	0.4333	0.3964
-25.6827	NA	NA	NA	0.9251	0.2176	NA	NA	NA	NA	0.4018	0.3764
74.2771	NA	1.0771	NA	NA	NA	NA	0.0023	-1.4301	NA	0.4246	0.3871
-29.2748	NA	NA	NA	0.7266	NA	NA	0.4794	NA	NA	0.3864	0.3603
-91.5608	NA	1.5072	1.4427	NA	0.3103	NA	NA	NA	NA	0.4128	0.3745
-35.9668	NA	NA	NA	0.6301	NA	NA	0.3639	NA	0.3096	0.4086	0.3700
-103.9360	NA	NA	1.6803	NA	0.3221	NA	NA	NA	NA	0.3778	0.3513
-31.7614	NA	NA	NA	0.6640	NA	NA	NA	NA	0.4857	0.3688	0.3419
9.4756	NA	NA	-0.8833	0.9386	NA	NA	NA	NA	0.7245	0.3960	0.3566
-30.4643	NA	1.2555	NA	0.5987	NA	NA	NA	NA	0.4300	0.3936	0.3540
-18.1301	NA	NA	-0.2119	0.8105	NA	NA	0.5003	NA	NA	0.3887	0.3488
-28.4399	NA	0.2623	NA	0.7184	NA	NA	0.4447	NA	NA	0.3871	0.3471
-16.9889	NA	1.5866	NA	0.7571	NA	NA	NA	NA	NA	0.3431	0.3152
-16.4004	NA	NA	NA	0.8695	NA	NA	NA	NA	NA	0.3021	0.2876
-12.5280	NA	1.6197	-0.0810	0.7892	NA	NA	NA	NA	NA	0.3435	0.3007
-22.6402	NA	NA	0.1130	0.8214	NA	NA	NA	NA	NA	0.3029	0.2732
-2.1145	NA	1.8544	NA	NA	NA	NA	NA	NA	0.6744	0.2852	0.2548

-6.1780	NA	NA	NA	NA	NA	NA	0.4258	NA	0.5778	0.2831	0.2526
11.8094	NA	NA	NA	NA	0.1984	NA	0.7308	NA	NA	0.2773	0.2466
-49.3282	NA	NA	0.8596	NA	NA	NA	0.5092	NA	NA	0.2746	0.2437
0.6845	NA	NA	NA	NA	NA	NA	NA	NA	0.8028	0.2281	0.2121
-35.6670	NA	NA	0.4800	NA	NA	NA	0.4125	NA	0.3837	0.2982	0.2524
-5.2002	NA	1.1792	NA	NA	NA	NA	0.2547	NA	0.5866	0.2973	0.2515
-21.3986	NA	1.6942	0.3156	NA	NA	NA	NA	NA	0.5533	0.2913	0.2451
13.0402	NA	1.1196	NA	NA	0.1999	NA	0.5725	NA	NA	0.2901	0.2439
-32.2724	NA	NA	0.5326	NA	NA	NA	NA	NA	0.5797	0.2468	0.2147
-46.1866	NA	0.5473	0.8275	NA	NA	NA	0.4386	NA	NA	0.2776	0.2304
-41.7623	NA	1.7950	0.9426	NA	NA	NA	NA	NA	NA	0.2358	0.2032
18.0414	NA	NA	NA	NA	NA	NA	0.6947	NA	NA	0.1940	0.1772
-54.3476	NA	NA	1.2050	NA	NA	NA	NA	NA	NA	0.1856	0.1687
31.5432	NA	2.8955	NA	NA	0.1911	NA	NA	NA	NA	0.2166	0.1833
19.2565	NA	1.0656	NA	NA	NA	NA	0.5437	NA	NA	0.2056	0.1718
36.6088	NA	2.7581	NA	NA	NA	NA	NA	NA	NA	0.1391	0.1212
49.4451	NA	NA	NA	NA	0.1733	NA	NA	NA	NA	0.0640	0.0445
53.2843	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.0000	0.0000