**Predicting Reasons for Crosswalk in American’s Health Insurance Plans**

1. Background:

I will be conducting a data science project on the “Health Insurance Marketplace”, a dataset published by the Centers for Medicare & Medicaid Services (CMS) which is also part of [Kaggle’s public datasets](https://www.kaggle.com/hhsgov/health-insurance-marketplace).

Prior to beginning into the analysis, we will begin with a very brief explanation of the Health Insurance Marketplace in the United States.[[1]](#footnote-1)

1. The Health Insurance Marketplace is for who people who don’t have health coverage.
   1. If you don’t have health insurance through a job, Medicare, Medicaid, the Children’s Health Insurance Program (CHIP), or another source that provides [qualifying coverage](https://www.healthcare.gov/glossary/minimum-essential-coverage), the Marketplace can help you get covered.
   2. If you have job-based insurance: You can buy a plan through the Marketplace, but you’ll pay full price unless [your employer’s insurance doesn’t meet certain standards](https://www.healthcare.gov/have-job-based-coverage/change-to-marketplace-plan/#affordable). Most job-based plans do meet the standards.
   3. If you have Medicare: You can’t switch to Marketplace insurance, supplement your coverage with a Marketplace plan, or buy a Marketplace dental plan.
2. What you pay for insurance depends on your income – and you’ll probably save:
   1. Your savings depend on your estimate of your expected income for 2016, not your income for 2015. Based on the income estimate you put on your application, we’ll tell you if you qualify for one of these:
      1. A health insurance plan with savings

Most people who apply qualify for a [premium tax credit](https://www.healthcare.gov/glossary/premium-tax-credit) that lowers their monthly insurance bill. Some also save on out-of-pocket costs like [deductibles](https://www.healthcare.gov/glossary/deductible) and [co- payments](https://www.healthcare.gov/glossary/co-payment). The plans are offered by private insurance companies with a range of prices and features. All plans cover [Essential health benefits](https://www.healthcare.gov/coverage/what-marketplace-plans-cover/), [Pre-existing conditions](https://www.healthcare.gov/coverage/pre-existing-conditions/), including pregnancy & [Preventive care](https://www.healthcare.gov/coverage/preventive-care-benefits/). You can [add dental to a health plan](https://www.healthcare.gov/coverage/dental-coverage/), but you don’t have to. You can’t buy a dental plan unless you enroll in a health plan.

* + 1. Medicaid and the Children’s Health Insurance Program (CHIP)

[Medicaid and CHIP](https://www.healthcare.gov/medicaid-chip/) provide free or low-cost coverage to millions of people and families with limited income, disabilities, and some other situations.

3) If you don’t have health insurance, you may have to pay a fee:

a. Most people must have [qualifying health coverage](https://www.healthcare.gov/glossary/minimum-essential-coverage) or pay a fee with their 2016 federal

taxes.

* 1. If you don’t have coverage in 2016, you’ll pay a penalty of either 2.5% of your income, or $695 per adult ($347.50 per child) — whichever is higher.

1. Problem:

There has been uncertainty on whether to crosswalk from a previous year plan into an upcoming year plan. But over the years, industry experts have often supported crosswalk as they are the quick-and-dirty fix that everyone is looking for as updating the individual’s systems and processes to accept and process 1CD-10 codes is better for the future.

As a result, crosswalks have happened more frequently but individuals can still specify reasons for cross walking if they do choose to crosswalk into the upcoming year insurance plan from an existing current year insurance plan. We will perform an exploration of the reasons for cross walking from individual’s application form and make a predictive model based on them. The findings from the analysis will be used to propose new amendments to insurance providers in terms of coverage.

1. Health Insurance Marketplace Dataset

There are seven files that make upp the Marketplace PUF . We will be using the dataset “The Plan ID Crosswalk PUF (CW-PUF)”. The purpose of the CW-PUF is to map QHPs and SADPs offered through the Marketplaces in 2014 to plans that will be offered through the Marketplaces in 2015.

These data either originate from the Plan Crosswalk template (i.e., template field), an Excel-based form used by issuers to describe their plans in the QHP application process, or were generated by CCIIO for use in data processing (i.e., system-generated). Each record relates to a mapping between a plan offered in 2014 and a plan offered in 2015 at the county or county-zip code level.

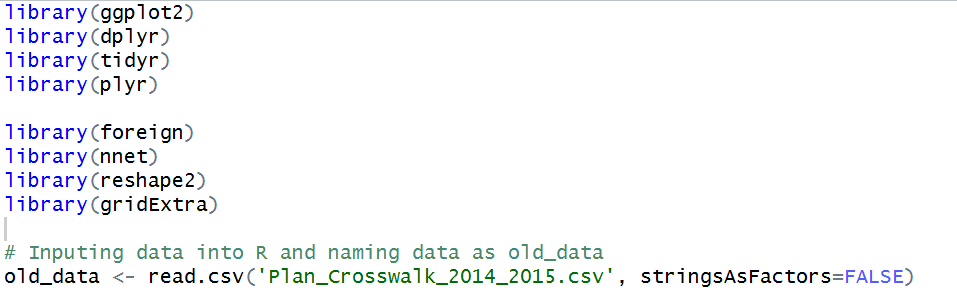
3.1 Data Wrangling

The complete R code can be found on [Vivan Raaj’s GitHub repository](https://github.com/titanum456/springboardfoundationofdatascience/blob/master/Capstone%20Project) .

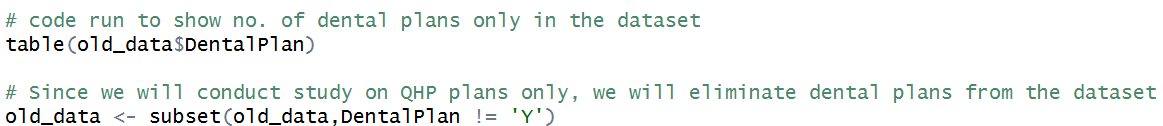
The Plan ID Crosswalk PUF (CW-PUF) consists of many different variables but not all of them are relevant for our analysis. The list of the relevant variable for our analysis are shown below:

|  |  |
| --- | --- |
| State | All 50 state abbreviations + 9 territory abbreviations |
| Dental Plan | Indicator that plan is a stand-alone dental plan (SADP) |
| PlanID\_2014 | Fourteen-character alpha-numeric code that identifies the 2014 insurance plan in the Health Insurance Oversight System (HIOS) |
| IssuerID\_2014 | Five-digit numeric code that identifies the issuer organization in HIOS for the 2014 plan |
| MultistatePlan\_2014 | Categorical indicator of whether the 2014 plan is a multi-state plan |
| MetalLevel\_2014 | Metal level, or coverage category, of the 2014 plan based on its actuarial value.  (Platinum , Gold, Silver, Bronze, Catastrophic, High, Low ).  Values of High and Low are only applicable for dental plans; values other than High and Low are only applicable to medical plans |
| FIPSCode | Federal Information Processing Standards (FIPS) code assigned to the county within a state that is included in the plan’s service area |
| CrosswalkLevel | Categorical indicator of the crosswalk level.  (0,1, 2, 3, 4, 5)  A value of 0 equals Crosswalking to the same Plan ID; a value of 1 equals Crosswalking at the Plan ID level; a value of 2 equals Crosswalking at the Plan ID and county coverage level; a value of 3 equals Crosswalking at the zip-code level for one or more counties; a value of 4 equals Discontinue 2014 plan with no cross walk (no re-enrollment option in 2015); a value of 5 equals 2014 Plan withdrawn prior to certification |
| Reason For Crosswalk | Categorical indicator of the reason for the crosswalk  (0,1 ,2 ,3, 4, 5, 6)  A value of 0 equals Renewing the same product/plan combination using the same 2014 Plan ID; a value of 1 equals Renewing the same product/plan combination using a different 2015 Plan ID; a value of 2 equals Renewing product or renewal in a different plan within the product; a value of 3 equals Continuing product with no plan available in the particular service area under that product and enrollment in a different product; a value of 4 equals Continuing product with no plan available in the particular service area under that product and no enrollment option; a value of 5 equals Discontinuing product and no enrollment option; a value of 6 equals Discontinuing product and enrollment into a different product |
| PlanID\_2015 | Fourteen-character alpha-numeric code that identifies the 2015 insurance plan in HIOS .  A value of “00000XX0000000” indicates the 2014 plan is not mapped to a 2015 plan for the given county or county-zip code because the product is discontinued and the issuer is not offering auto-enrollment |
| IssuerID\_2015 | Five-digit numeric code that identifies the issuer organization in HIOS for the 2015 plan. A value of “00000” indicates the 2014 plan is not mapped to a 2015 plan for the given county or county-zip code because the product is discontinued and the issuer is not offering auto-enrollment |
| MultistatePlan\_2015 | Categorical indicator of whether the 2015 plan is a multi-state plan  (Y,N, X)  A value of X indicates the issuer did not provide a crosswalk plan ID |
| MetalLevel\_2015 | Metal level, or coverage category, of the 2015 plan based on its actuarial value  ( Platinum ,Gold, Silver ,Bronze, Catastrophic, High, Low, X )  A value of X indicates the issuer did not provide a crosswalk plan ID; values of High and Low are only applicable for dental plans; values other than High and Low are only applicable to medical plans |
| ChildAdultOnly\_2015 | The types of child enrollment options for the 2015 plan  (0,1, 2, X)  A value of 0 means the 2015 plan allows adult-only and child-only enrollment; a value of 1 means the 2015 plan allows only child-only enrollment; a value of 2 means the 2015 plan allows only adult-only enrollment; a value of X indicates the issuer did not provide a crosswalk plan ID |

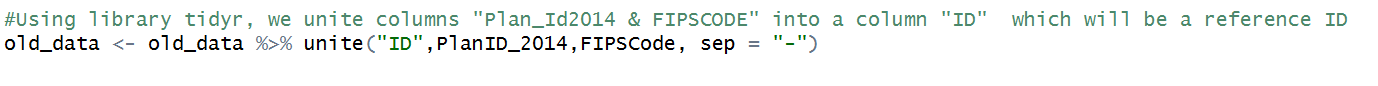
The first step after downloading the dataset as to load into R. Prior to that, we would load the R library packages into the R environment.



Based on our understanding of the healthcare plan that you can’t buy a dental plan without a health plan. On this basis, we will remove all records that contains health plans with dental plans as we consider it as an interference to the data.

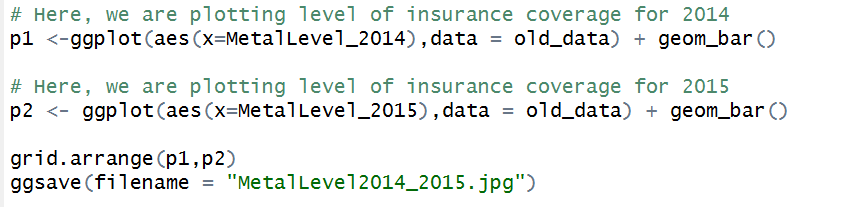


Next step is to identify each row as an individual id. As the plan id is not unique for each individual, for the purpose of this analysis, we will create a reference ID by uniting the columns “Plan\_ID2014 & FIPSCODE” into “ID”

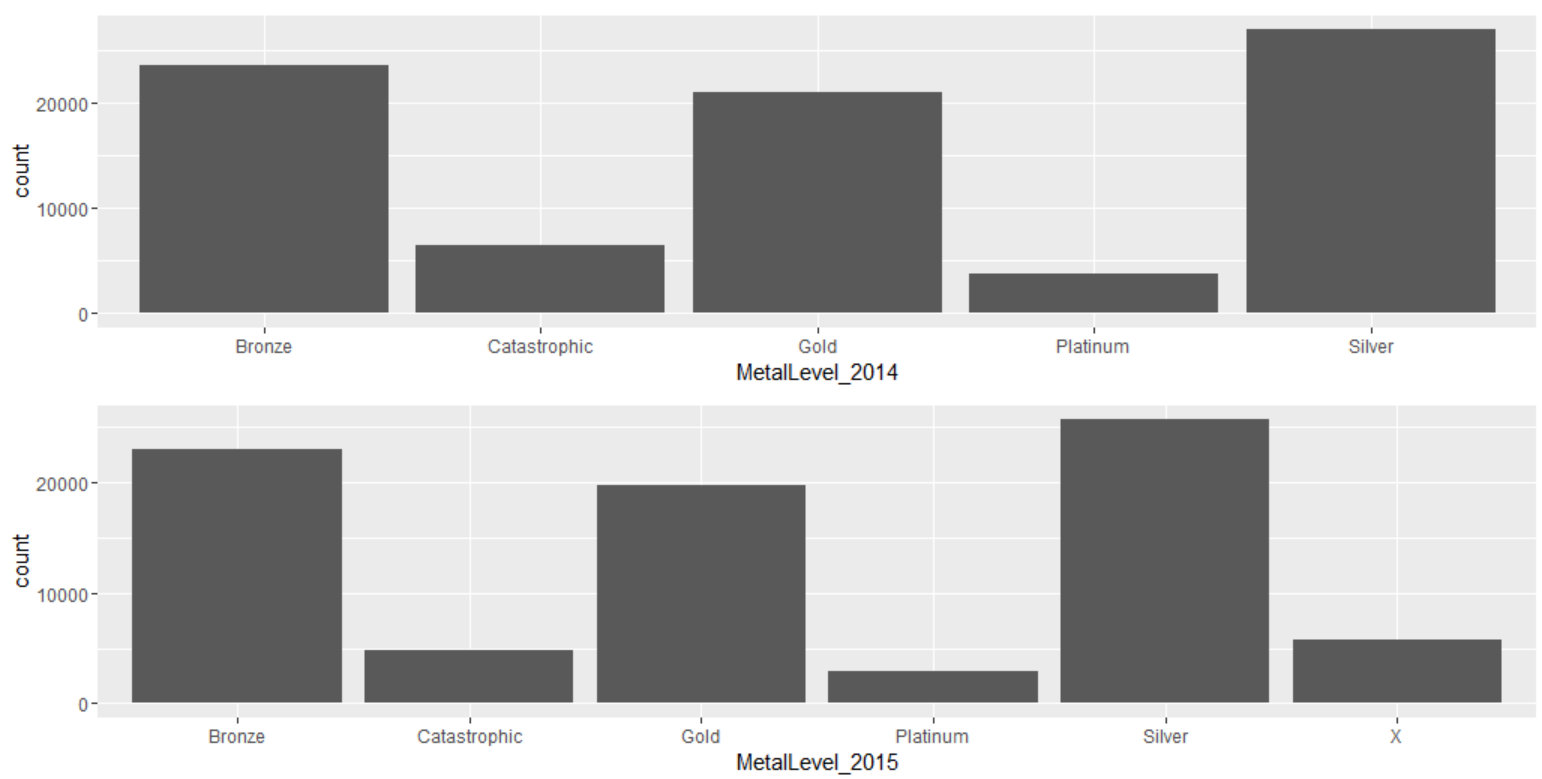


3.2 Data Exploration

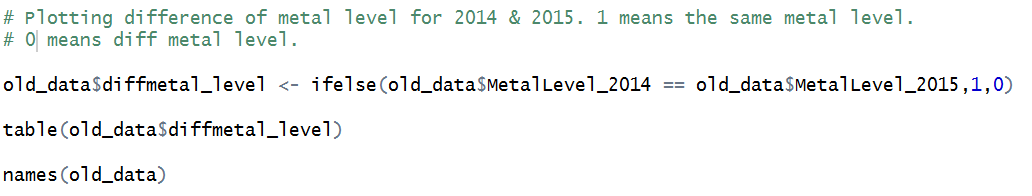
The cleaning of the data is complete at this point and we will move into data exploration. We first plot a bar plot of the “MetalLevel\_2014” & “MetalLevel\_2015” to analyze the distribution of the coverage category based on the insurance actuarial value.



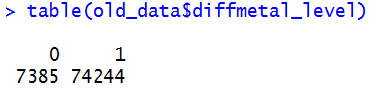
Output:



The analysis of the bar plot above shows that the highest premium insurance, Silver has more been included in more insurance plans as compared to cheapest premium insurance, Bronze. However, we would like to get the difference in metal level for 2014 & 2015 as individuals may change the metal level in their application form for crosswalk to 2015 plan.

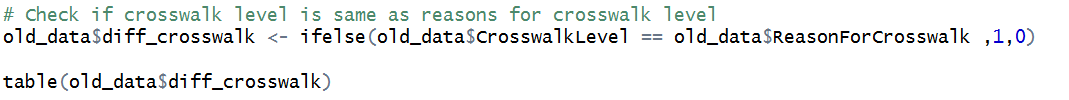


Output:

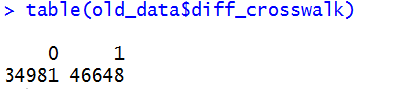


The output above shows that there are more individuals maintaining their metal levels than switching them. Thus, we can consider this variable as a possible independent variable that determines the reason for crosswalk.

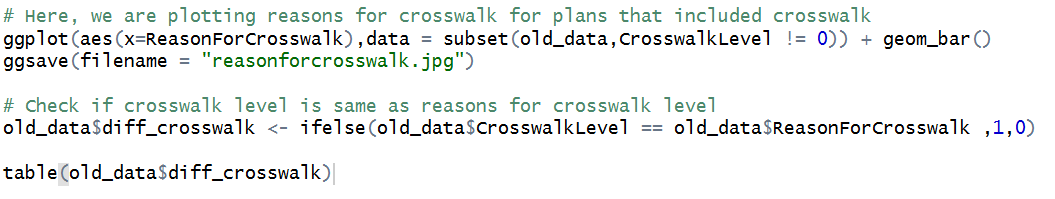
Next, we consider another variable which could have an effect on the reason for crosswalk which is the variable “CrosswalkLevel”. My initial assumption was that the crosswalk level equals the reason for crosswalk but however after checking for similarities between the crosswalk level and reasons for crosswalk, the observation was that there were almost as many dissimilarities as the similarities. Thus, we will also consider the variable “ CrosswalkLevel” as a possible independent variable that determines the reason for crosswalk.



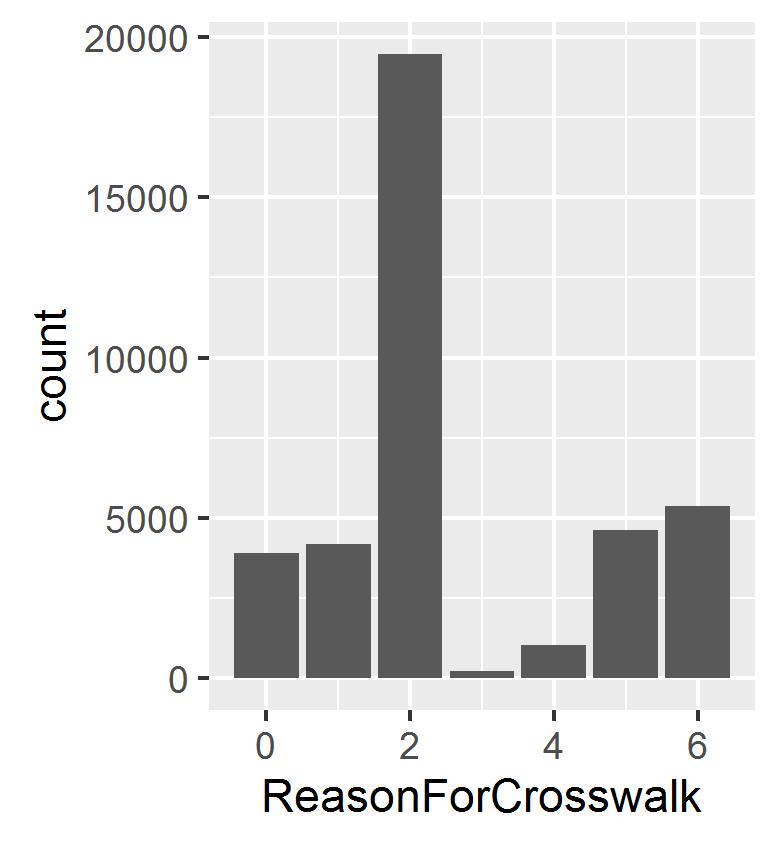
Output:



Next, we omit out observations for individuals whom opted out from crosswalk. We will subset the data to include “CrosswalkLevel != 0”.



Output:



As observed above, majority of the individuals chose 2 which references to “Renewing product or renewal in a different plan within the product “. Generally, this translates to the individuals being not too comfortable with the previous year plan and would like further addition/omission of benefits.

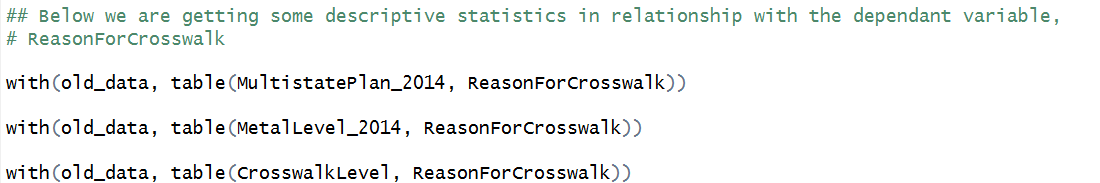
The observed output should be heavily weighted by all parties involved in healthcare as it justifies a general consensus within the public that they are not too comfortable with their existing plans.

4. Prediction Algorithms

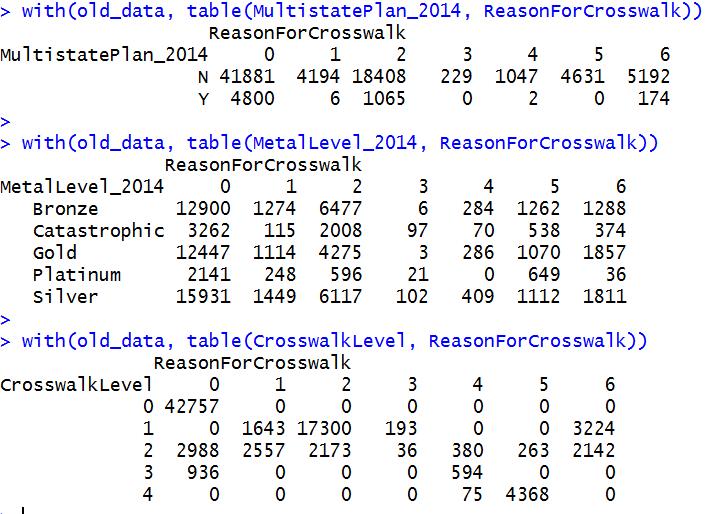
We noticed a trend in which most individuals opted for option “2” when filling up their application form for the 2016 insurance plan. However, we are not able to determine their reasons as whether they require addition/omission of clauses for the upcoming year insurance plan. To facilitate the prediction of “Reasons for Crosswalk” in future Plan ID Crosswalk, we use “Multinomial Logistic Regression” to model nominal outcome of between 0 to 6 for the “Reasons for Crosswalk”.

4.1 Prediction Model 1:

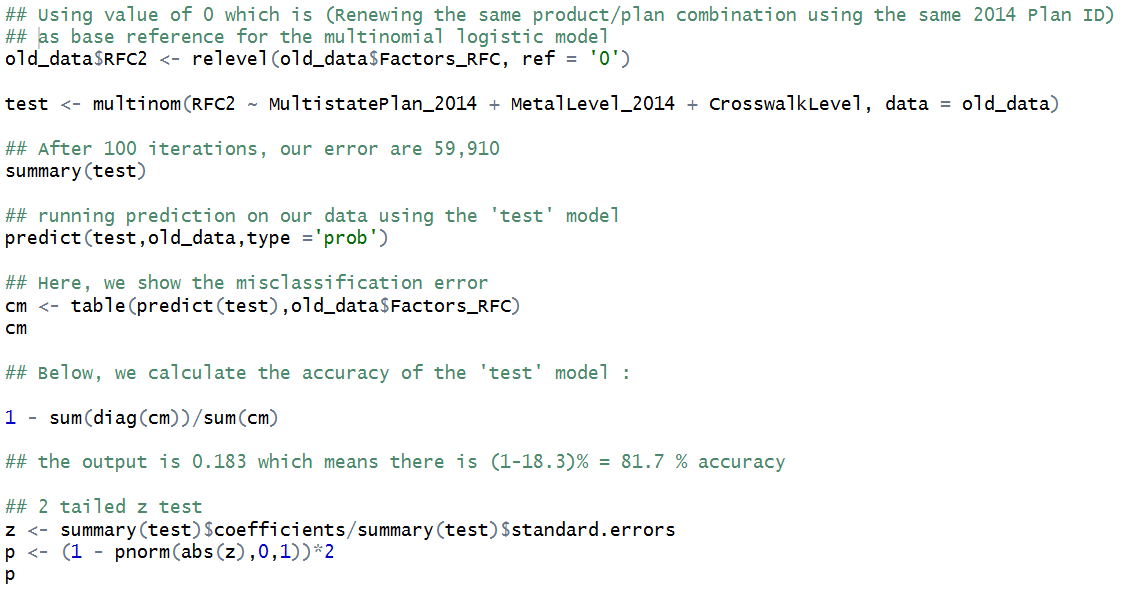
For the 1st model, we will use the independent variables MultistatePlan\_2014, MetalLevel\_2014 & CrosswalkLevel and the dependent variable ReasonforCrosswalk. We added the 3rd independent variable: MultistatePlan\_2014 as it is an general factor that influences the individual’s decision when filling up the Plan ID Crosswalk.



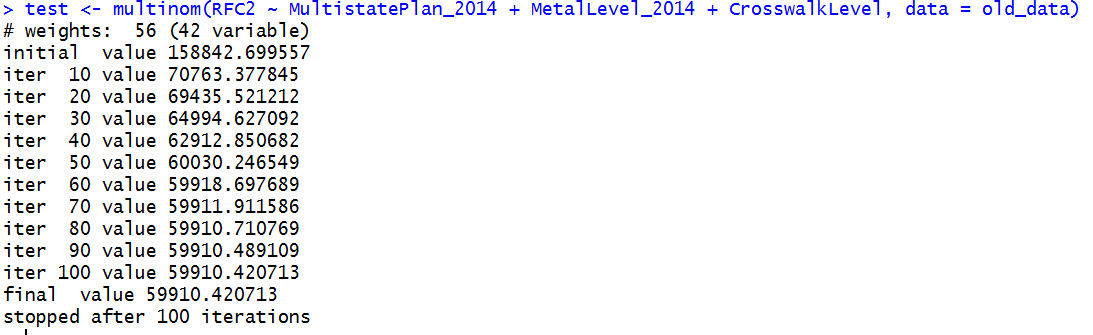
Output:

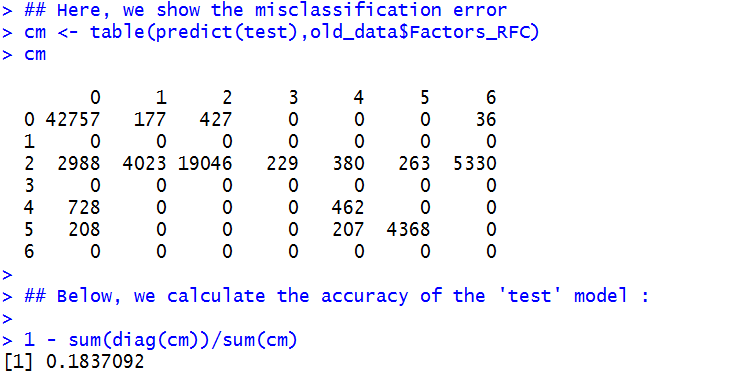


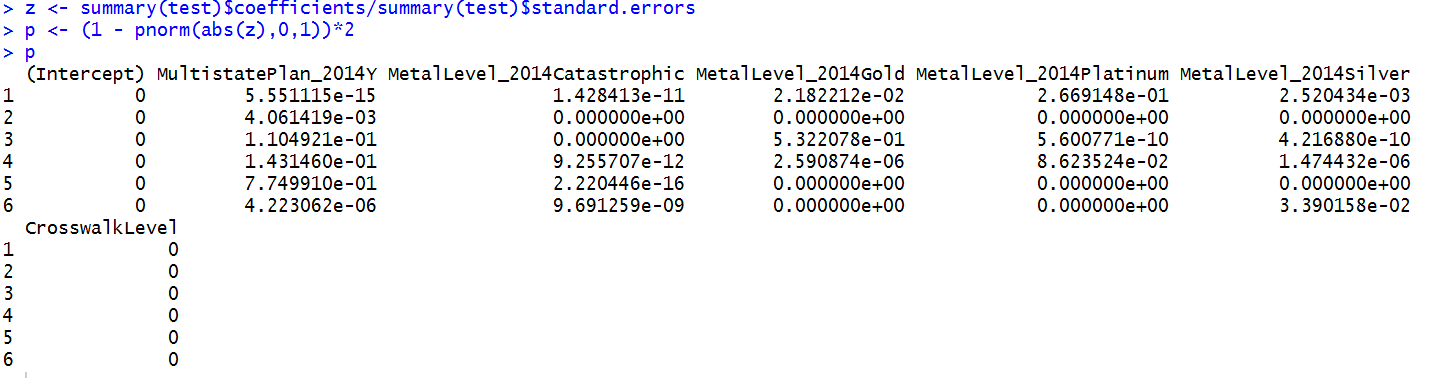
The output shows a scattered distribution among all independent variable values and the reason for crosswalk values. The below lines of code shows the steps to obtain the error value of these independent variable combined together and the accuracy level of the models.



Output:



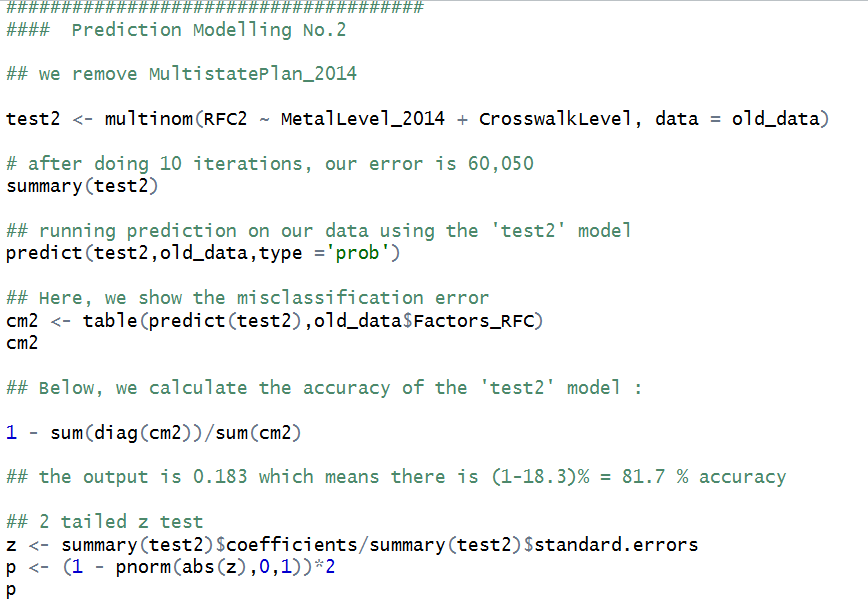




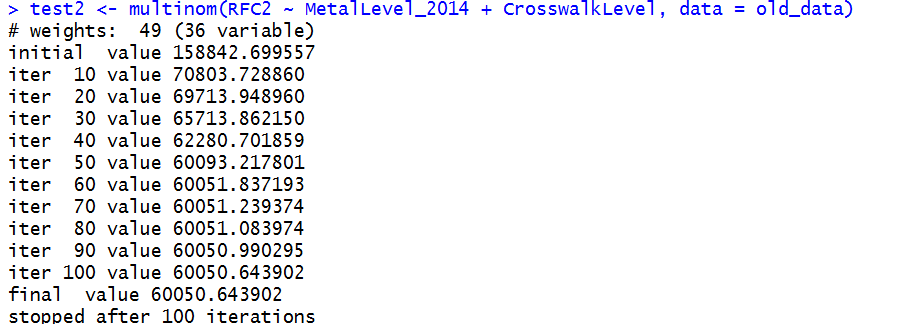
Our observation from the above lines of code is that prediction model 1 has 81.7% accuracy.

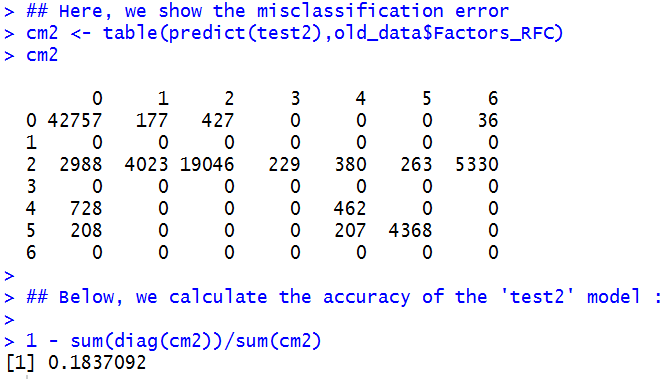
4.1 Prediction Model 2:

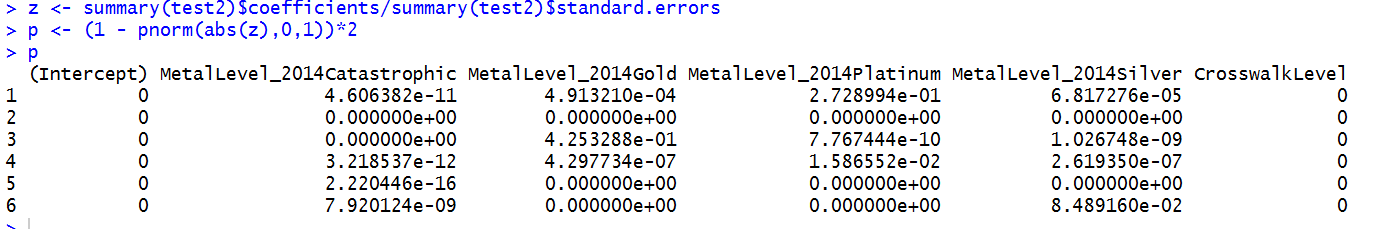
The accuracy value of 81.7% of prediction model 1 is quite good however we would try to re-create the model with one less independent variable in which we will omit out, MultiStatePlan\_2014. Below, we will run the code:



Output :







We also get the same accuracy value 81.7% for the 2nd model.

4. Results & Discussions

Our results may be summarized as follows:

* From the data exploration, we identified that three independent variable; MetalLevel\_2014, CrosswalkLevel & MultistatePlan\_2014 correlates with the ReasonForCrosswalk variable.
* There is common pattern whereby most individuals opted for option “2” when choosing the ReasonForCrosswalk in the Plan ID Crosswork Template. (“2” = Renewing product or renewal in a different plan within the product”)
* Generally, this translates to the individuals being not too comfortable with the previous year plan and would like further addition/omission of benefits.
* Based on both prediction models, as both outputs the same level of accuracy, we will opt for prediction model 2 as “the general rule within data science is to choose the models with less variables if different models exhibit same statistical values”.

4. Recommendations

The observed output from the data exploration should be considered in more serious weight by all parties involved in healthcare as it justifies a general consensus within the public that they are not too comfortable with their existing plans. A couple of reasons for this general consensus maybe due to their previous year’s insurance plan in terms of whether it caters for multi states plans, cost of the insurance plan which is referenced to the metal level and finally their selection of crosswalk.

My recommendations:

1. To Include multistate plan as a standard clause in every insurance plan as more Americans will be more inclined to purchase them. The only stumbling block will be the different state policies in regards to insurance plans.
2. Individuals may be too confused/overwhelmed when choosing coverage for insurance plans hence the unequal pattern seen in the metal level bar plot. My recommendations to solve this problem is to reduce no. of options for coverage plan which may entail more simplified package of insurance plan. Thus those that pay for low insurance premium may not necessarily have to pay for higher health cost but instead slightly lower health cost.

1. 4 Health Insurance Marketplace Tips from Healthcare.gov. (n.d.). Retrieved April 03, 2016, from

   https://www.healthcare.gov/quick-guide/one-page-guide-to-the-marketplace/ [↑](#footnote-ref-1)