Predictor selection and variation information calculation

2019-1-24

There are 45 parameters in the merged data, that is continuous or categorical with at lease 5 degrees.

The I selected one of them into the k-means algorithm and calculated the VI.

One predictor selected

There are 45 parameters in the merged data, that is continuous or categorical with at lease 5 degrees.

Then I selected one of them into the k-means algorithm and calculated the VI.

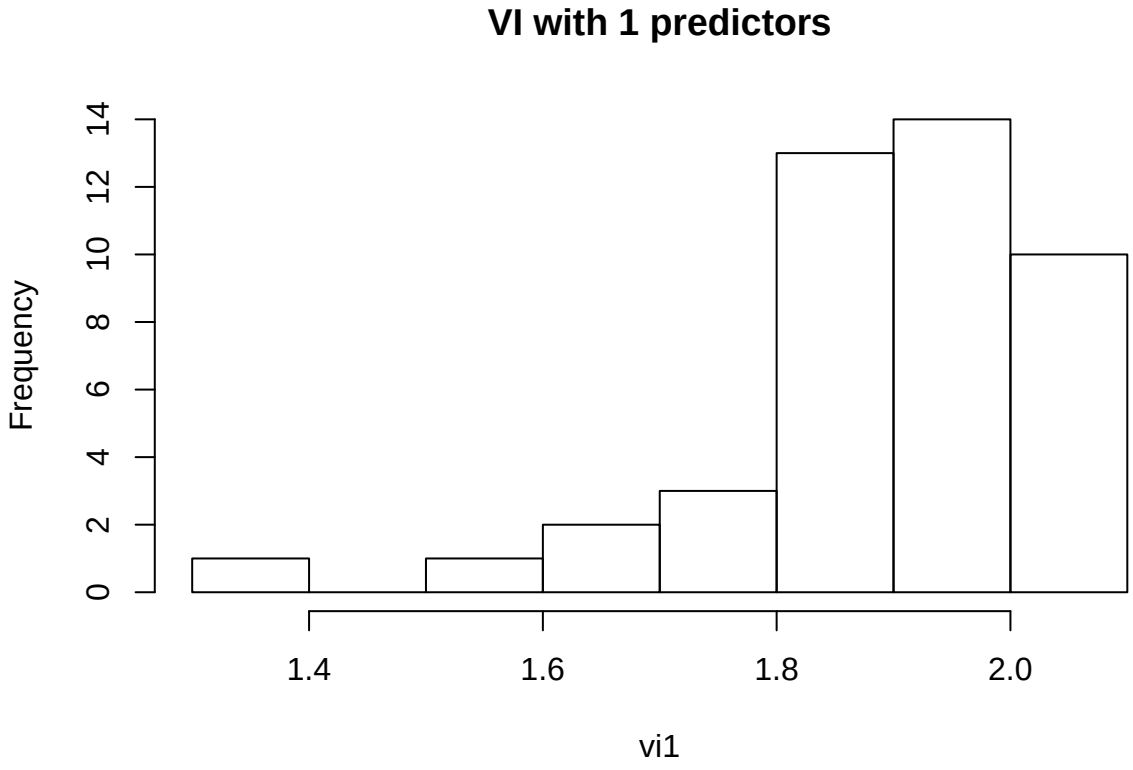
The minimum VI value is:

```
## [1] 1.331118
```

The predictors that got the minimum value is:

```
##                   names       vi
## 37 demo_maritial_status 1.331118
```

The histgram is:



Two predictor selected

There are 45 parameters in the merged data, that is continuous or categorical with at lease 5 degrees.

Then I selected two of them into the k-means algorithm and calculated the VI.

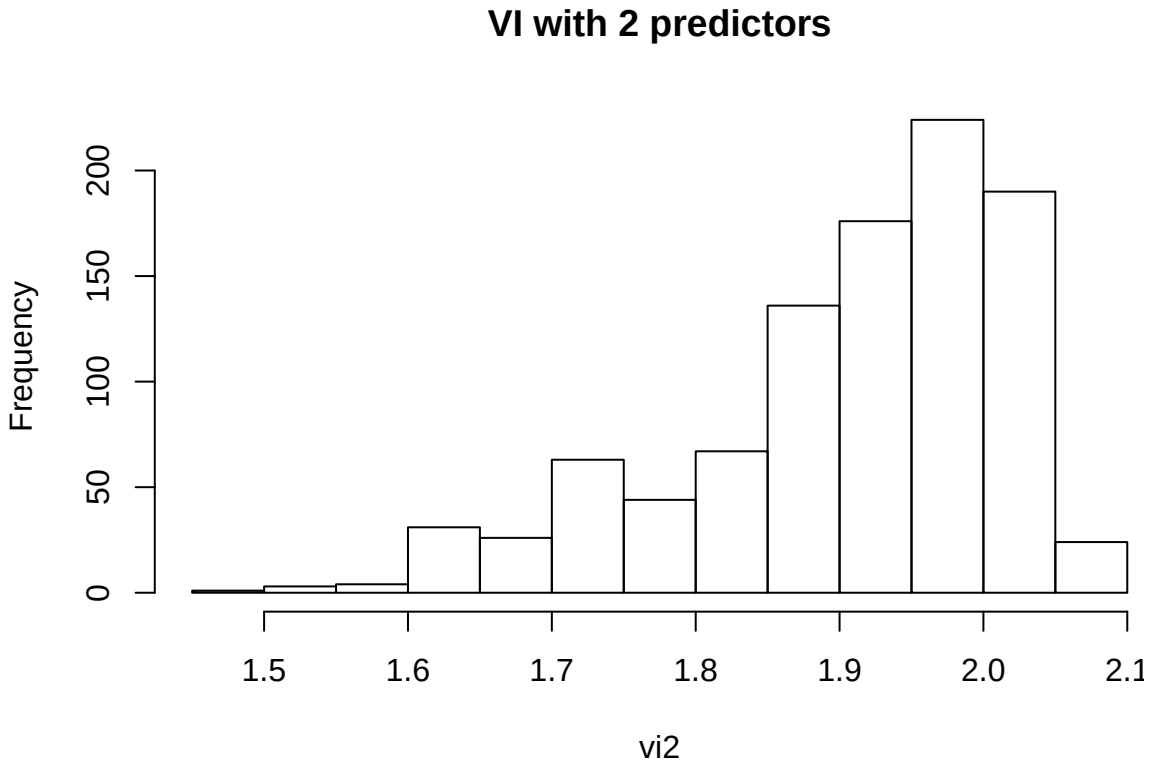
The minimum VI value is:

```
## [1] 1.47107
```

The predictors that got the minimum value is:

```
##          names1            names2      vi
## 734 eeg.w0_2062 demo_educa_status 1.47107
```

The histgram is:



Three predictor selected

There are 45 parameters in the merged data, that is continuous or categorical with at lease 5 degrees.

Then I selected three of them into the k-means algorithm and calculated the VI.

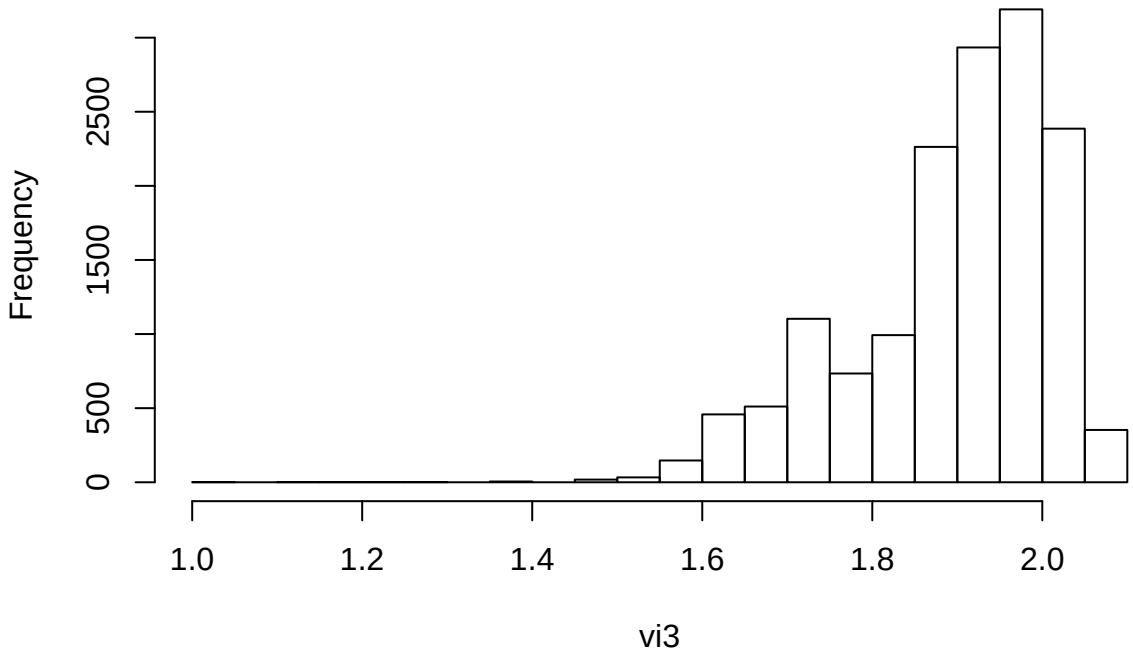
The minimum VI value is:

```
## [1] 1.128081
```

The predictors that got the minimum value is:

```
##            names1       names2 names3       vi
## 11689 eeg.w0_2059 fMRI.w0_1593   Race 1.128081
```

The histgram is:



Four predictor selected

There are 45 parameters in the merged data, that is continuous or categorical with at lease 5 degrees.

Then I selected four of them into the k-means algorithm and calculated the VI.

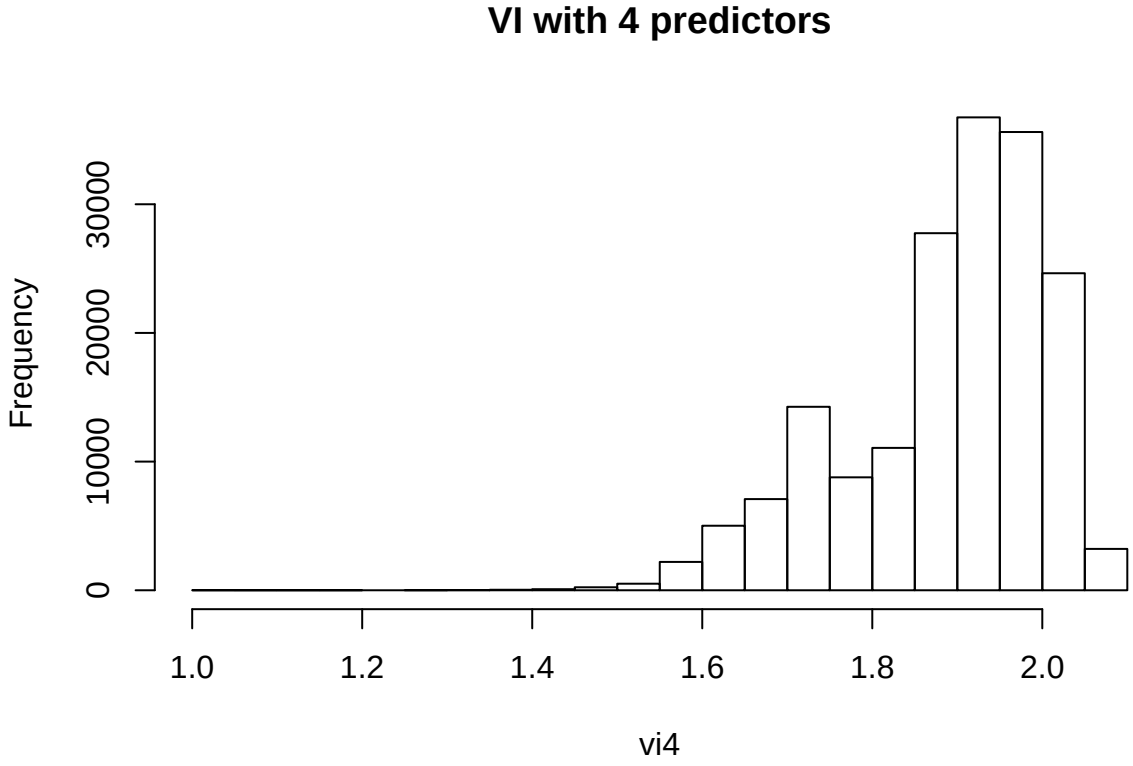
The minimum VI value is:

```
## [1] 1.013434
```

The predictors that got the minimum value is:

```
##              names1       names2              names3          names4
## 171964 fMRI.w0_1671 fMRI.w0_1729 demo_monthly_income qids_eval_total
##              vi
## 171964 1.013434
```

The histgram is:



The combination of two parameters:

I tried to add two predictors together: a new predictor is $a_1x_1 + a_2x_2$

Here, I tried $3.5 * x_1 + x_2$ and got a low VI value:

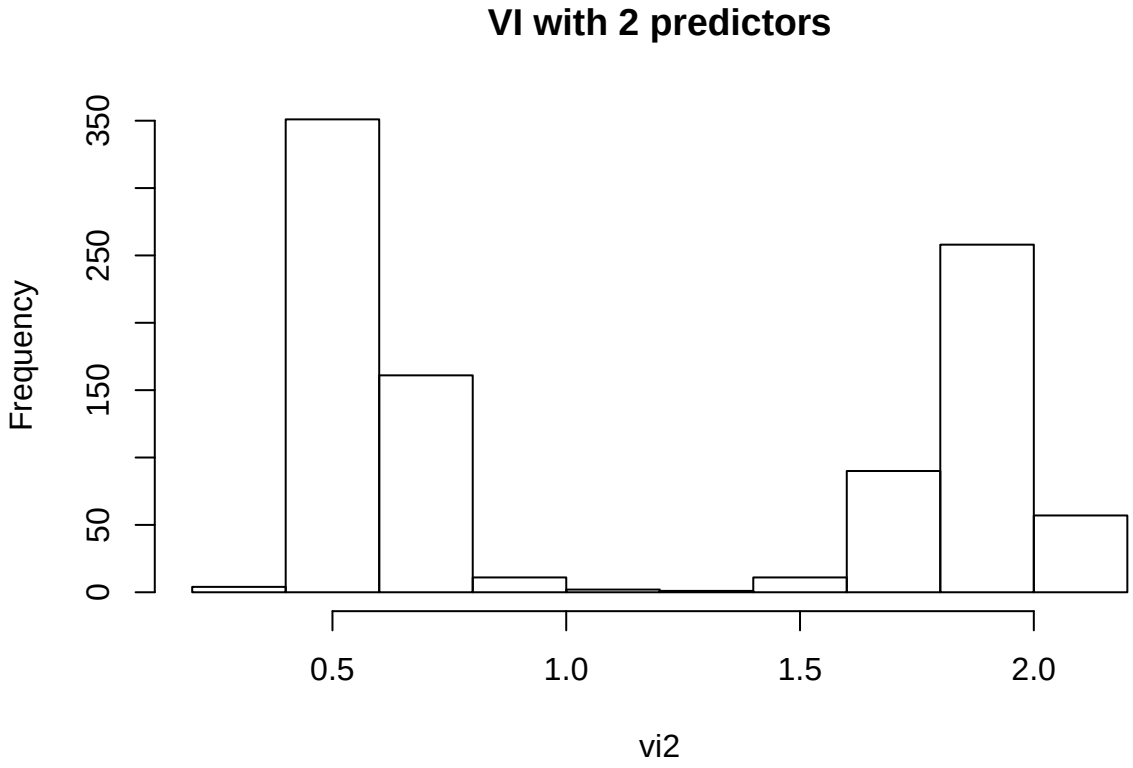
The minimum VI value is:

```
## [1] 0.2491987
```

The predictors that got the minimum value is:

```
##          names1          names2        vi
## 664 eeg.w0_2042 demo_hous_numbr 0.2491987
```

The histgram is:


VI with 2 predictors
Frequency
0
50
150
250
350
0.5
1.0
1.5
2.0
vi2