

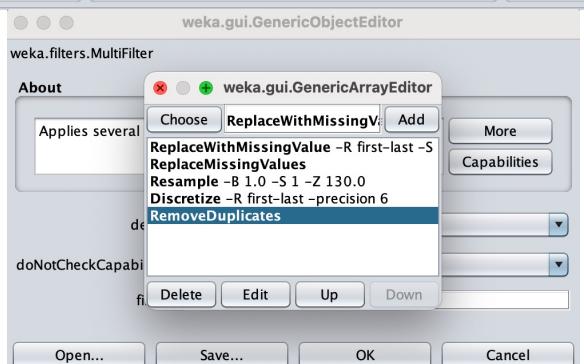
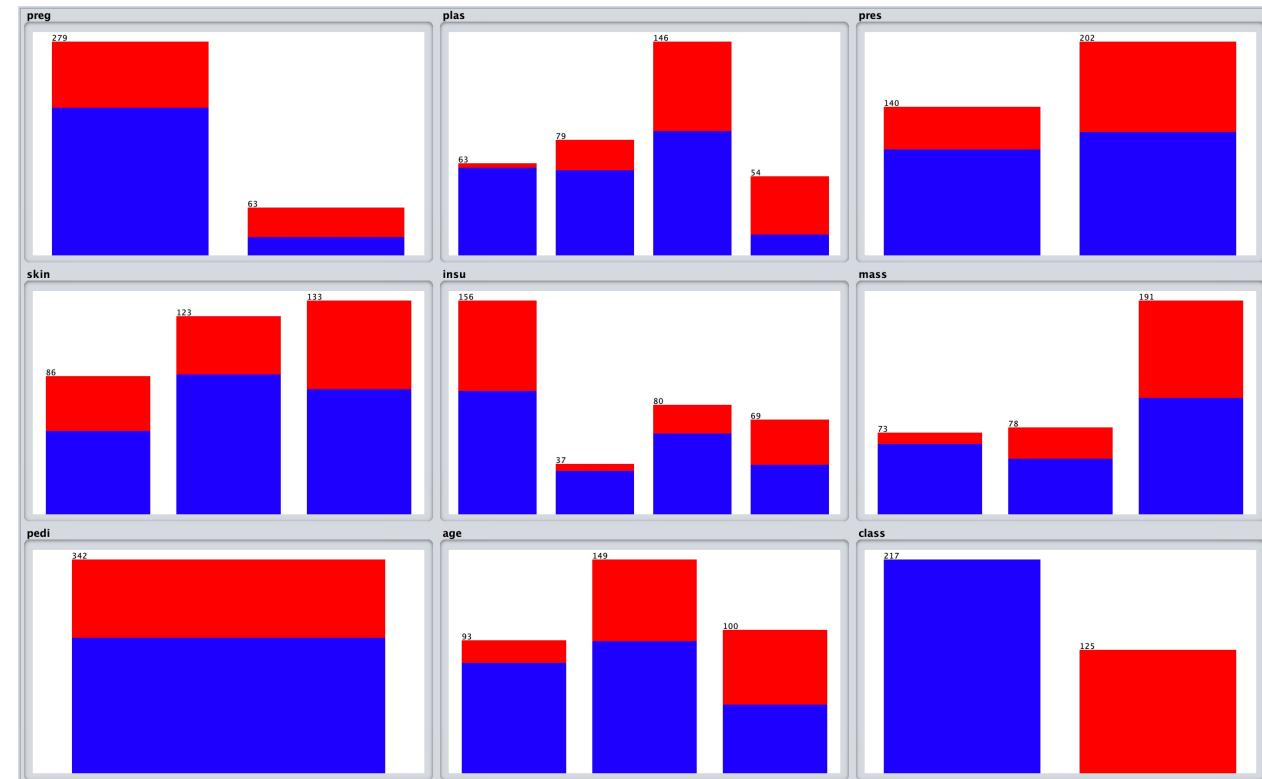
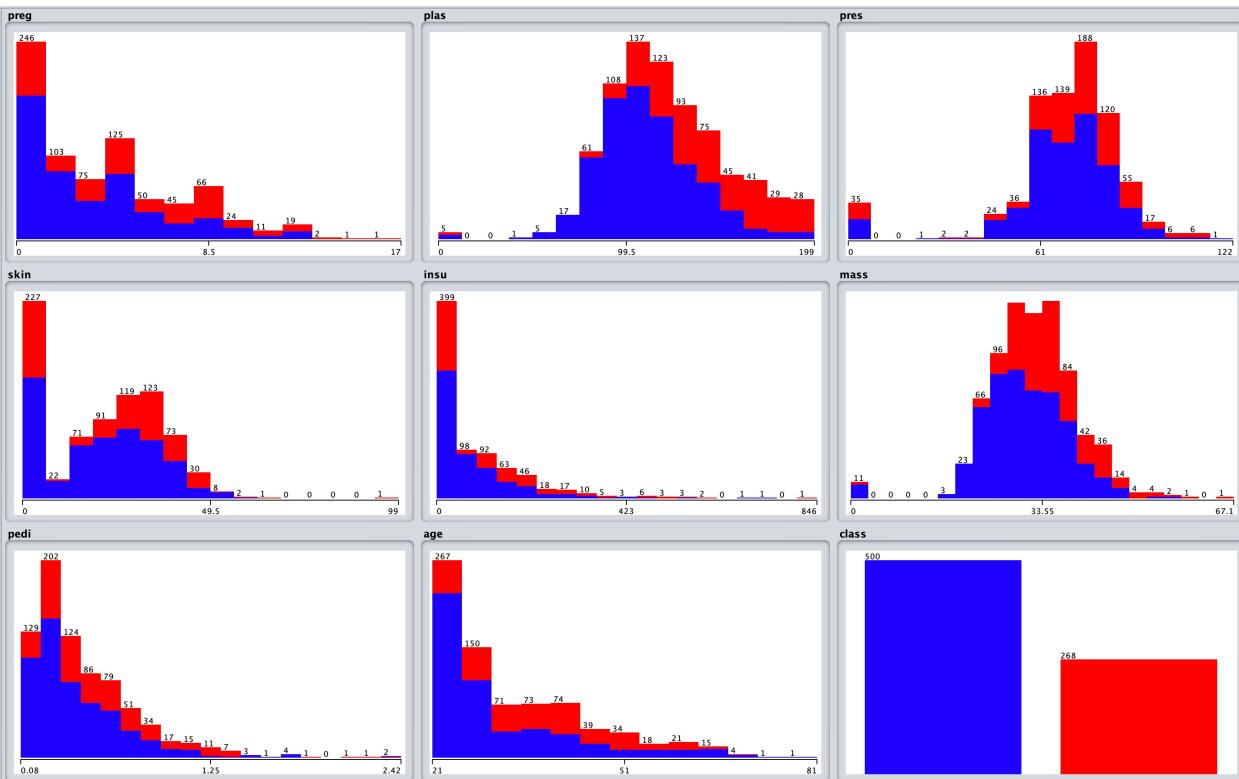
# Exercise 3

Repeat both exercises 1 and 2, considering both SMOTE and simple oversampling as resampling methods and compare the achieved results (the minority class should contain the same number than instances of the initial majority class).

To compare the achieved results, ***remove the duplicated instances*** and return, for each resampling method, the final number of instances and check the class distribution.

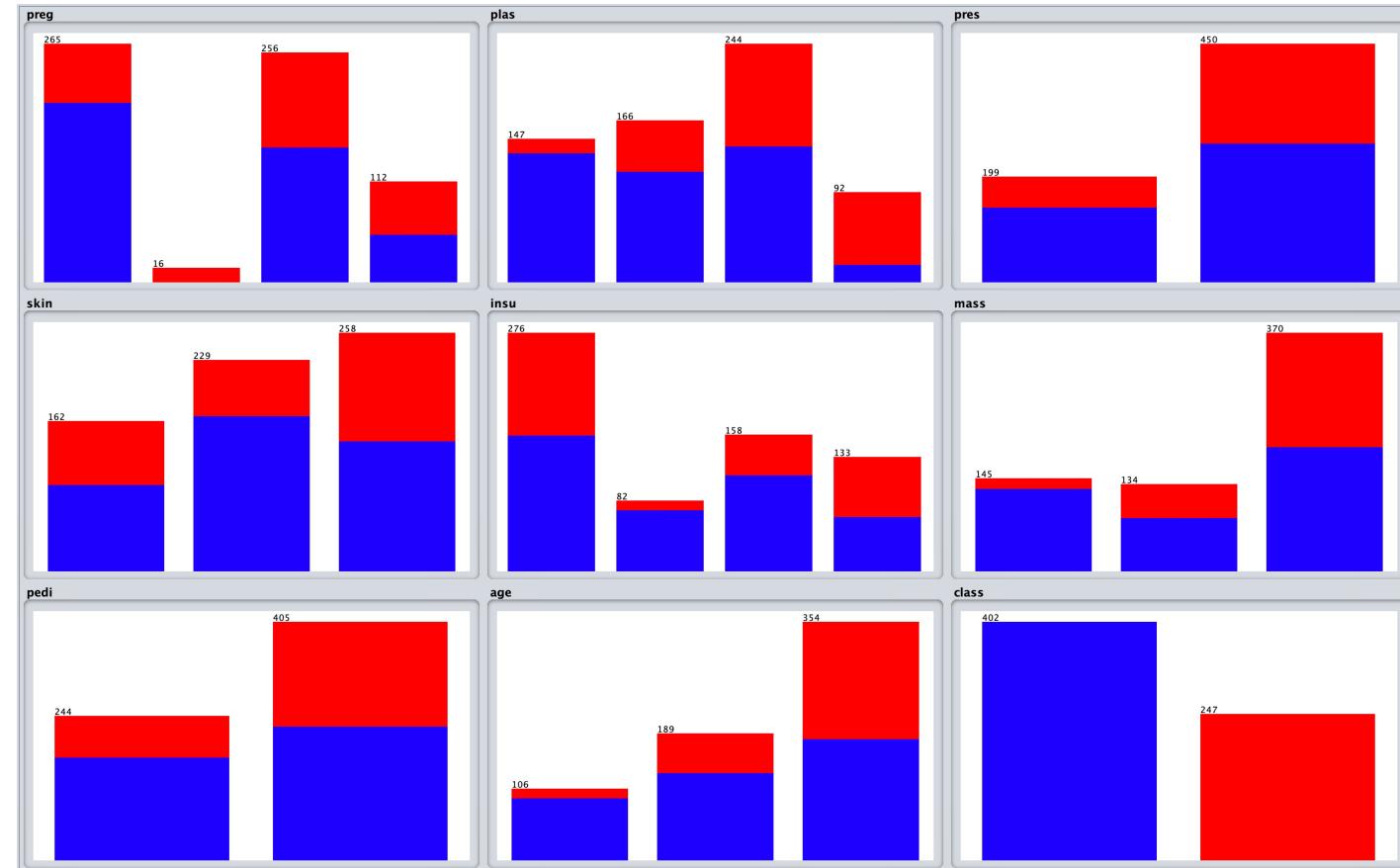
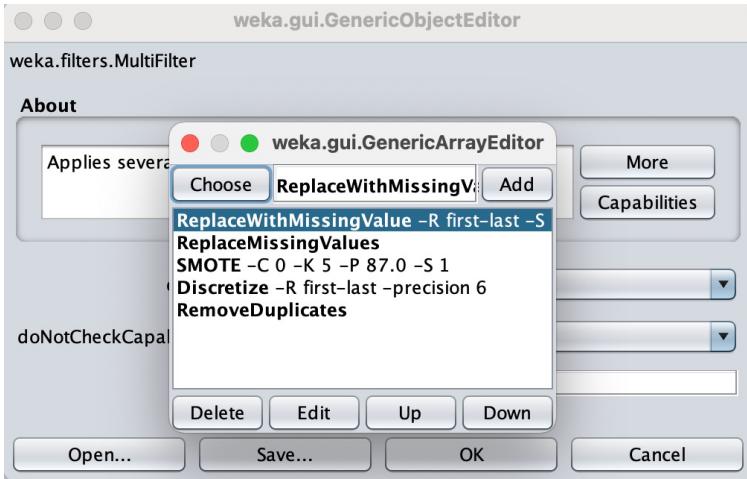
Adopt the multi-filter for solving the exercises.

# Diabetes.arff with multifilter (Resample)



We can see the starting data-set on the left  
 The results, on the right, are obtained by applying the multifilter  
 The total amount of instances change from 768 to 342

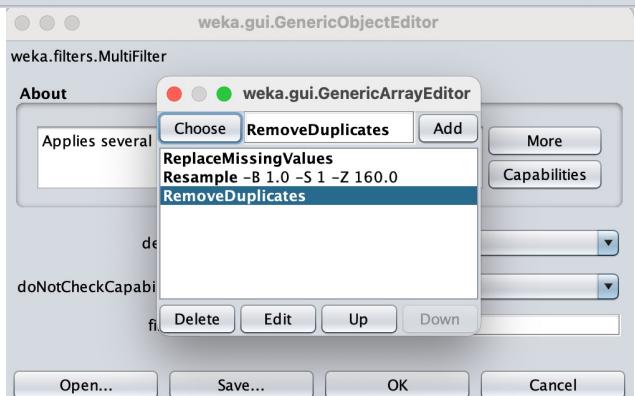
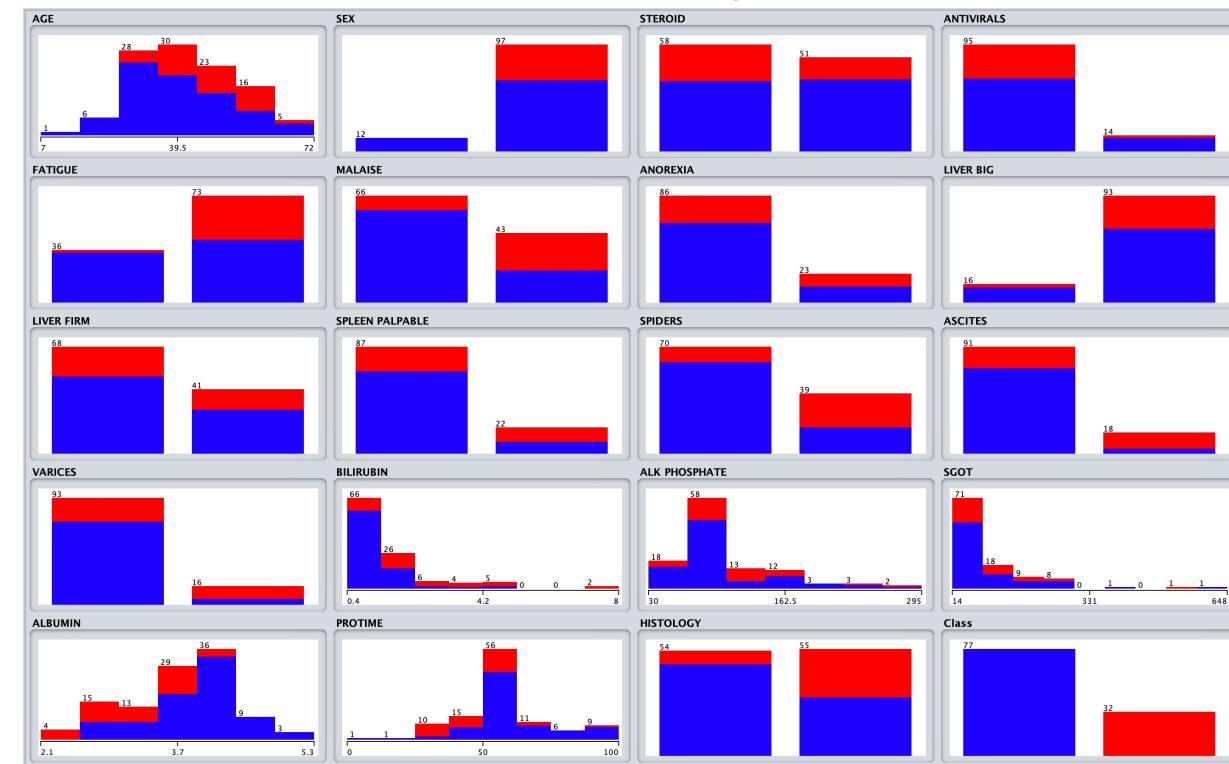
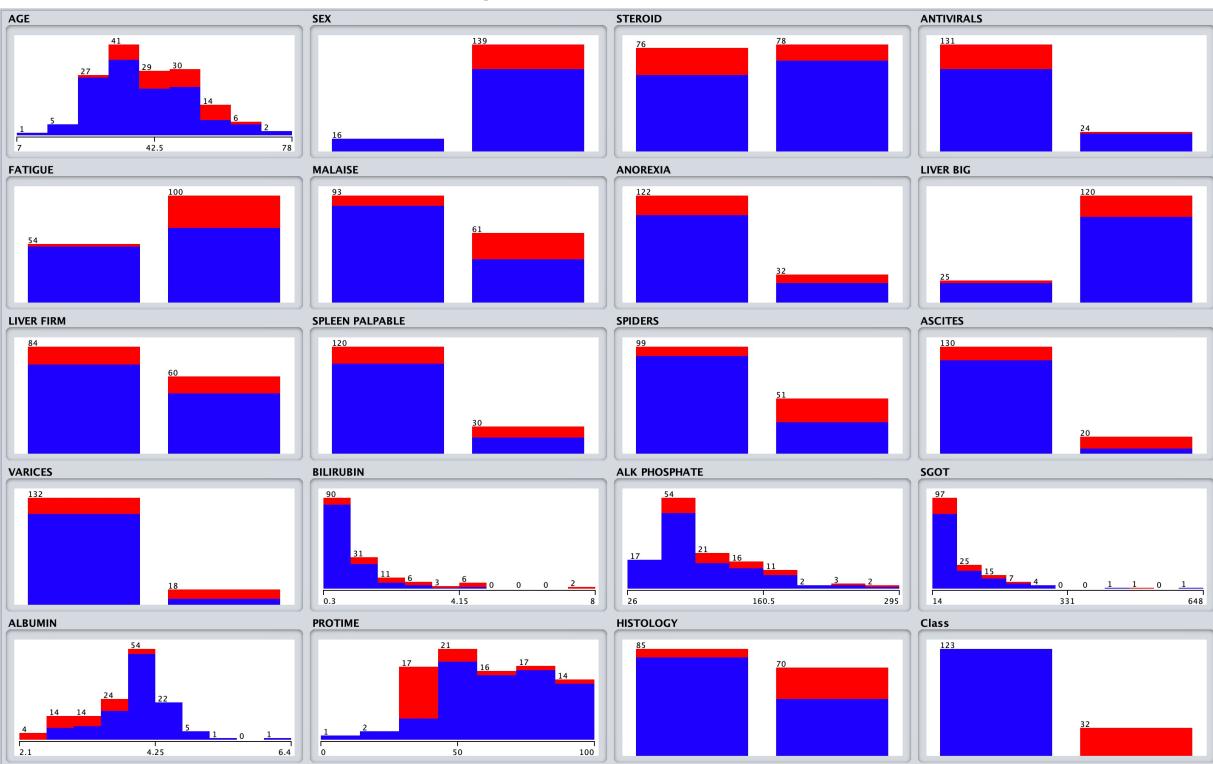
# Diabetes.arff with multifilter (Smote)



In this case we have applied the the Smote filter instead of using the Resample one

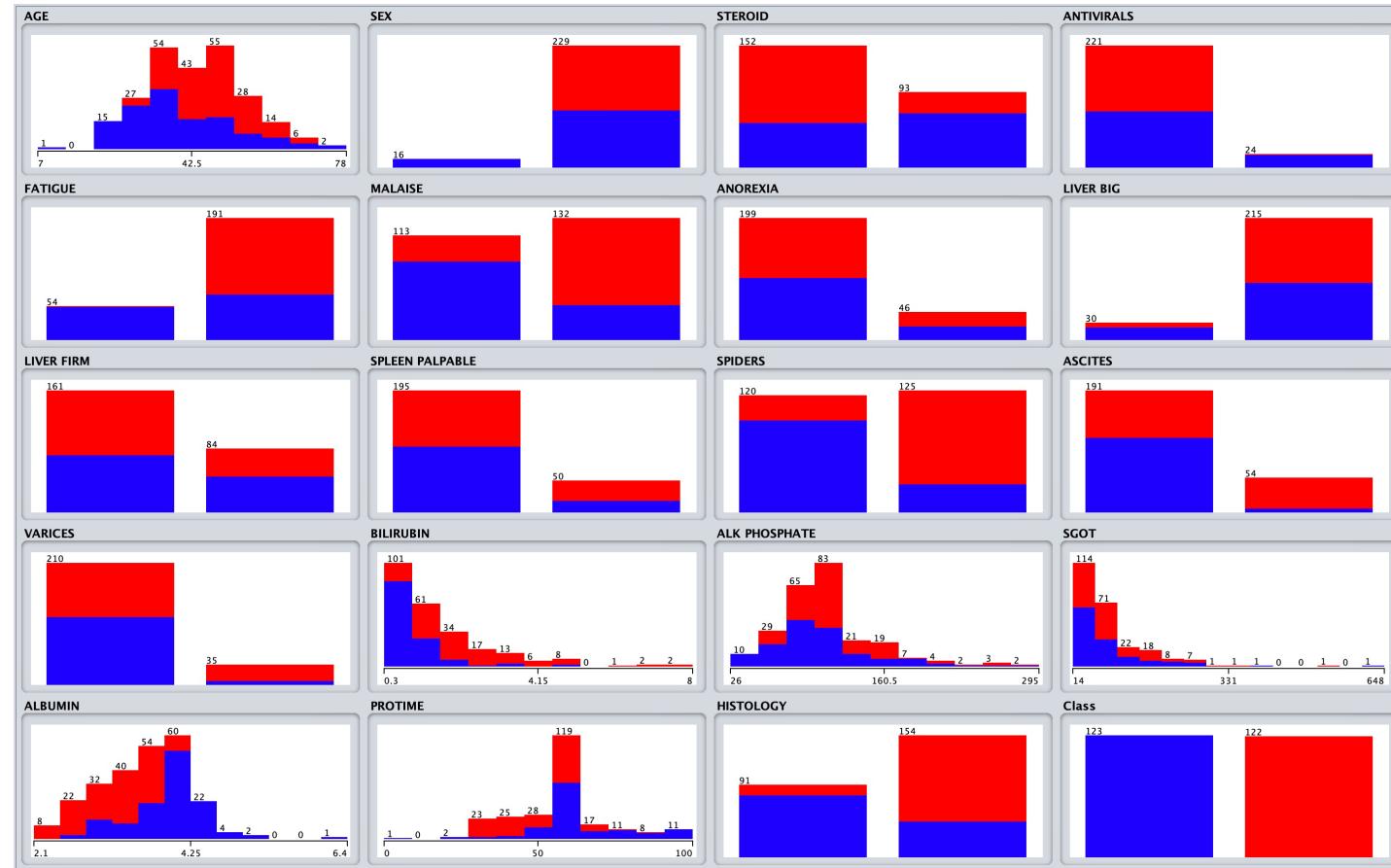
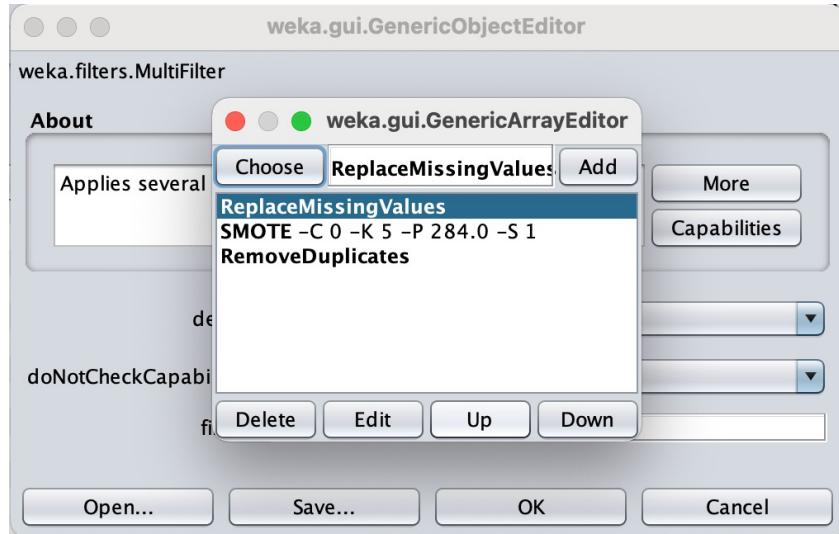
We leave to 0 the classValue parameter so the filter will auto-detect the minority class, we set the percentage to 87 in order to obtain and equal distribution of the two classes. After the removing of duplicates the total instances are 649

# Hepatitis.csv with multifilter (Resample)



As in the Diabetes case, on the left, the starting data-set and on the right the result obtained using the multifilter  
The total instances varieties from 155 to 109

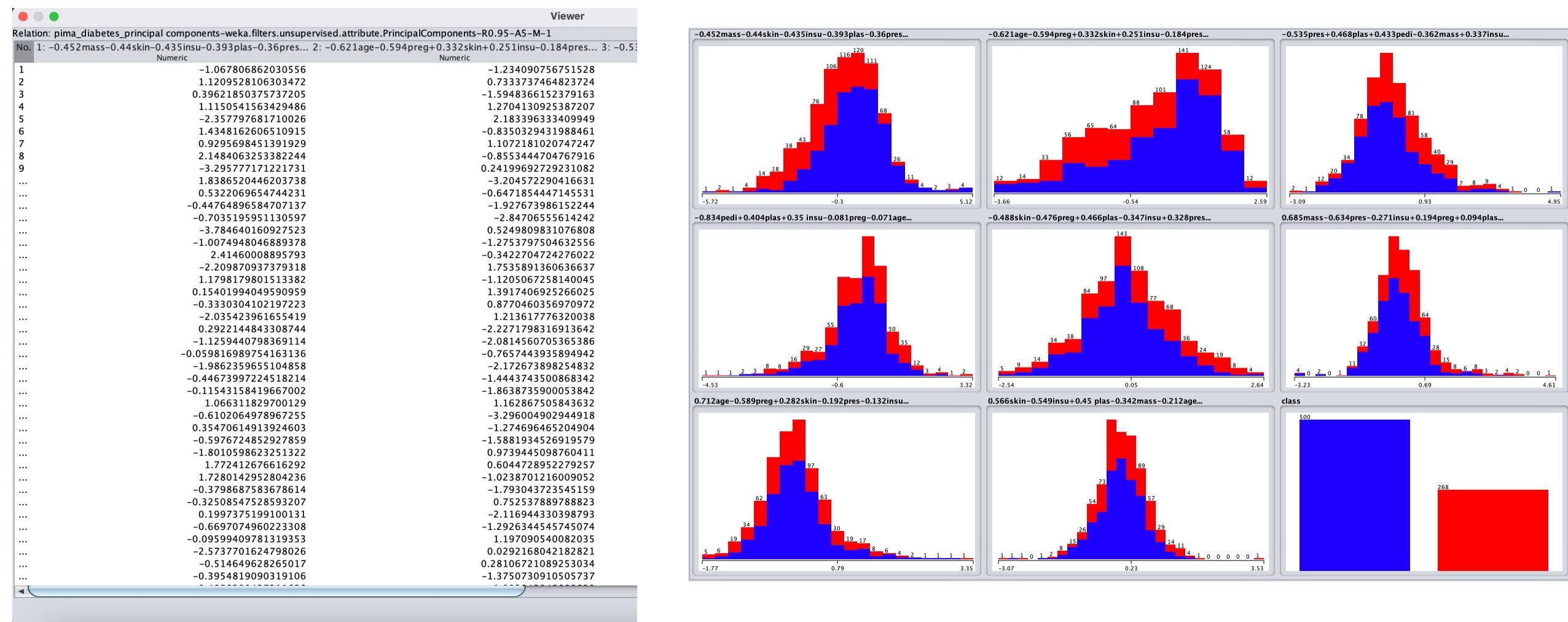
# Hepatitis.csv with multifilter (Smote)



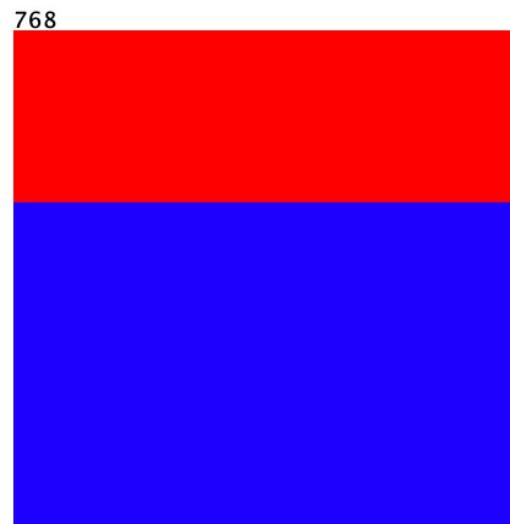
This results are obtained using Smote filter instead of Resample filter.

The total of unique instances are : 245

# Diabetes.arff PCA

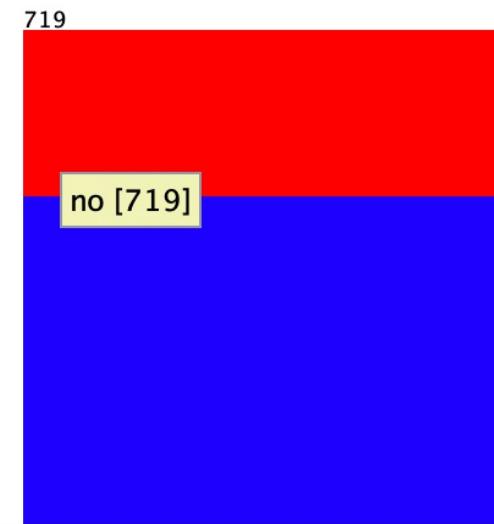


# Diabetes.arff InterquartileRange



There aren't extreme values

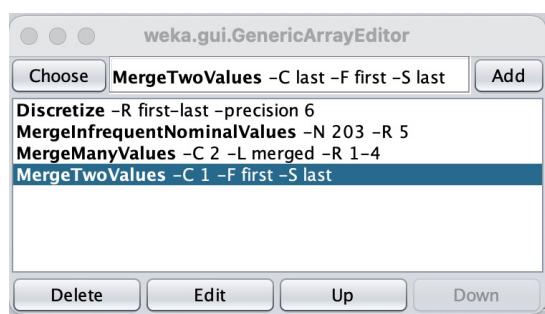
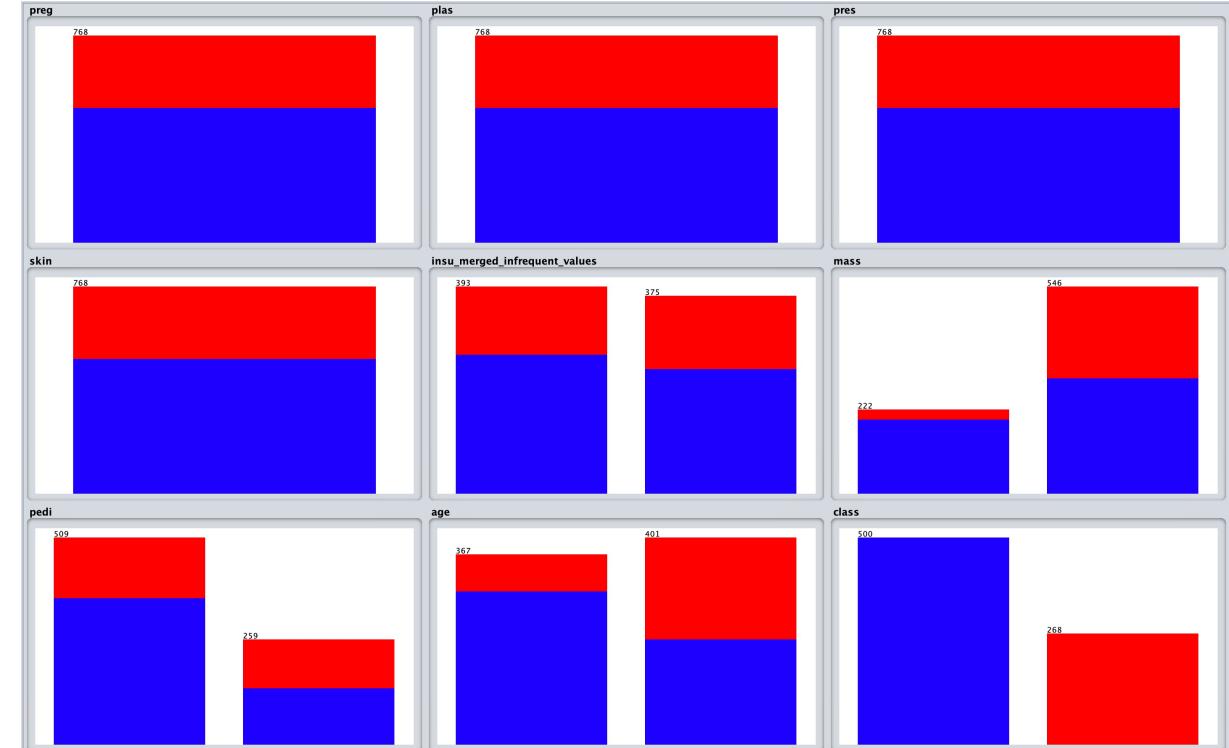
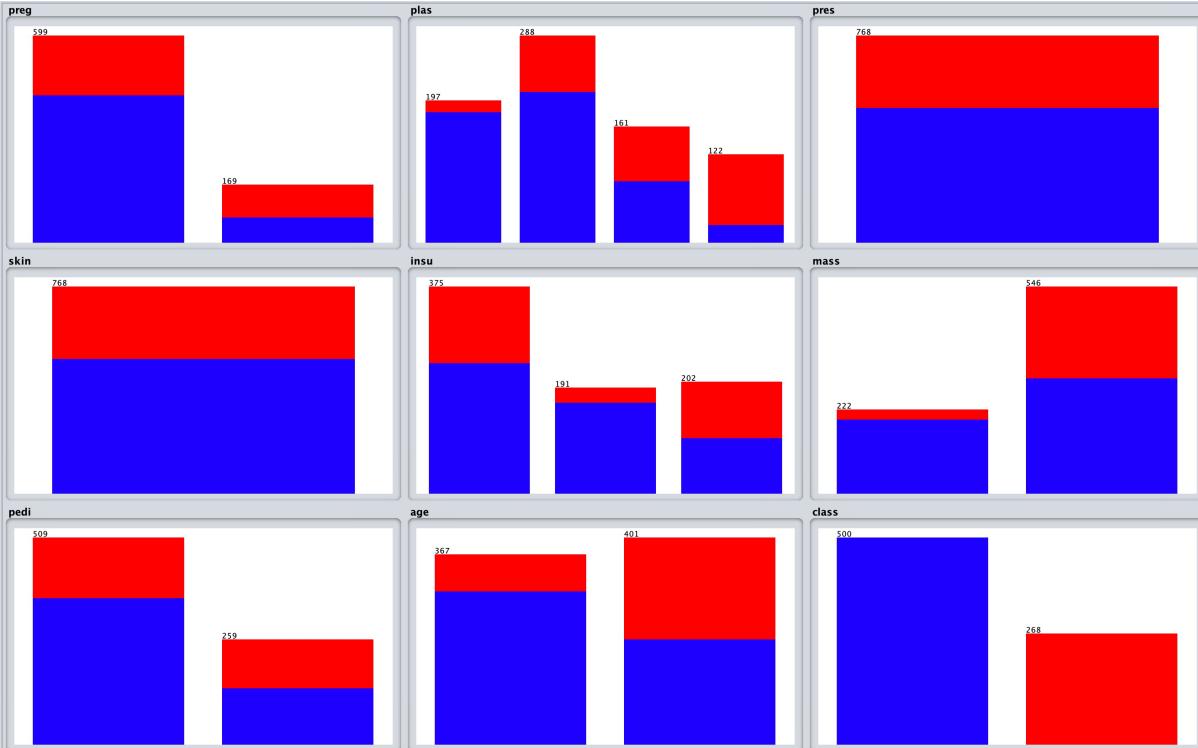
0



There are 49 outliers

49

# Diabetes.arff MergeFilter



**MergeInfrequentNominalValues** : Merge all bins that don't contain at least 203 instances (in this case), the filter is applied to the 5<sup>th</sup> attribute

**MergeManyValues** : Merge the bins (1 to 4) of the attribute 2

**MergeTwoValues** : Merge two bins of the attribute 1