Nonparametric Statistics

Noé Debrois, Person Code 10949145, ID 242751

2024-01-12

Exercise 1

Question 1.1

Synthetic description of assumptions, methods, and algorithms
The median is the deepest point in the sense of the Tukey depth. ROM (Romania) has the highest tukey depth. Then PAN. Then MEX. **Results and brief discussion**

```
## GDP_per_capita  CO2_emissions            Gini
##    8297.4836207      3.9184102      -0.9365389

## GDP_per_capita  CO2_emissions            Gini
##    7.958584e+03   2.449052e+00   7.643719e-02

## GDP_per_capita  CO2_emissions            Gini
##    6752.5521777      5.9668835      -0.5905878
```

Question 1.2

Synthetic description of assumptions, methods, and algorithms
We can build automatically the bagplot matrix using the aplpack library. We need a bagplot MATRIX since we have three dimensions (GDP, CO2 and Gini) and since a bagplot only compares two dimensions at a time. Then we plot only the bagplot based on the two dimensions of interest and extract outliers using the $pxy.outlier argument. We can retrieve the name of the using by looking in the df_red dataframe.

Results and brief discussion

Bagplot
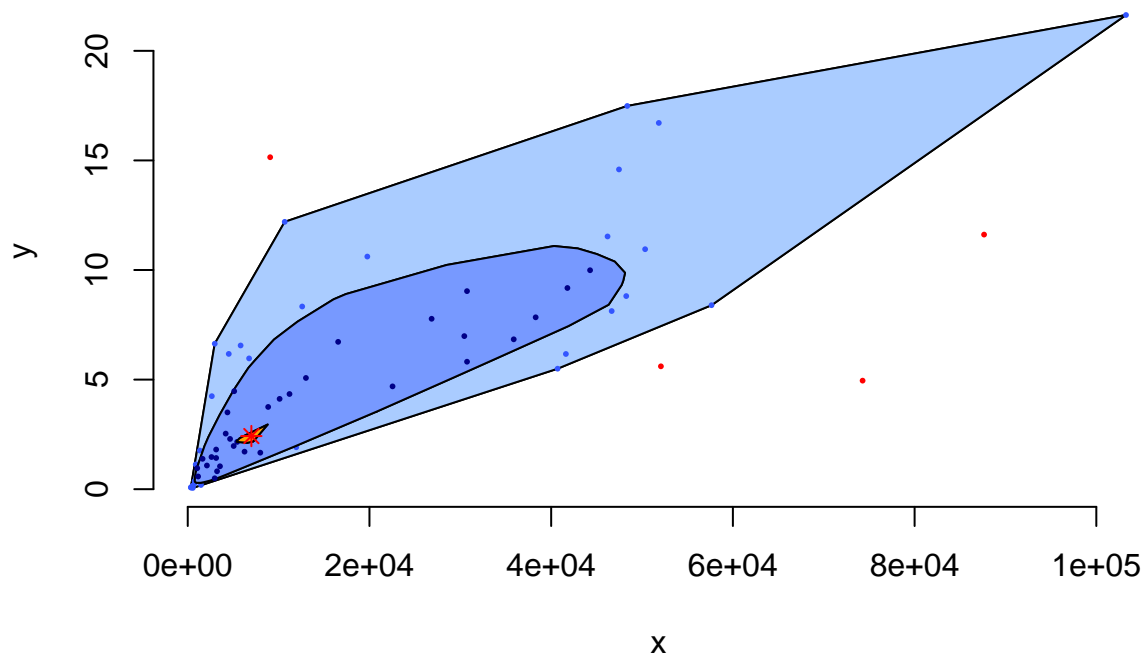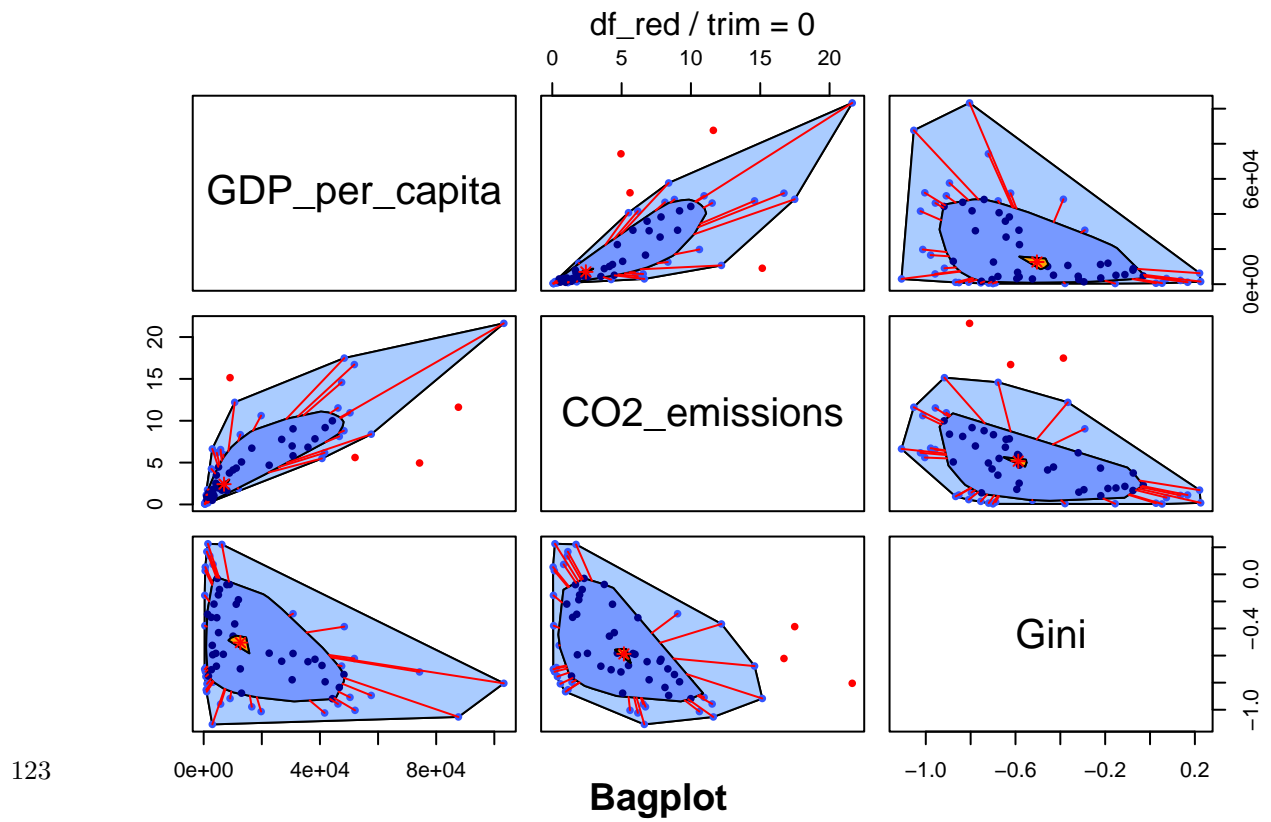

```
## 35 46 59 60
## 35 46 57 58

##    CountryCode GDP_per_capita CO2_emissions        Gini
## 35         KAZ        9070.65      15.14648 -0.9168508
```

```
##    CountryCode GDP_per_capita CO2_emissions       Gini
## 46         NOR       87646.27       11.61619 -1.053258

##    CountryCode GDP_per_capita CO2_emissions       Gini
## 61         THA       5111.909       4.470426 -0.4305291

##    CountryCode GDP_per_capita CO2_emissions       Gini
## 62         TUN       4176.611       2.535619 -0.5836203
```

Question 1.3

Synthetic description of assumptions, methods, and algorithms
Test statistics will be : Wilks Lambda. Note that wilk's lambda is significant for small values. It is sufficient
to change its sign to use it in a permutation test. We are in a multivariate framework,so wee permute the
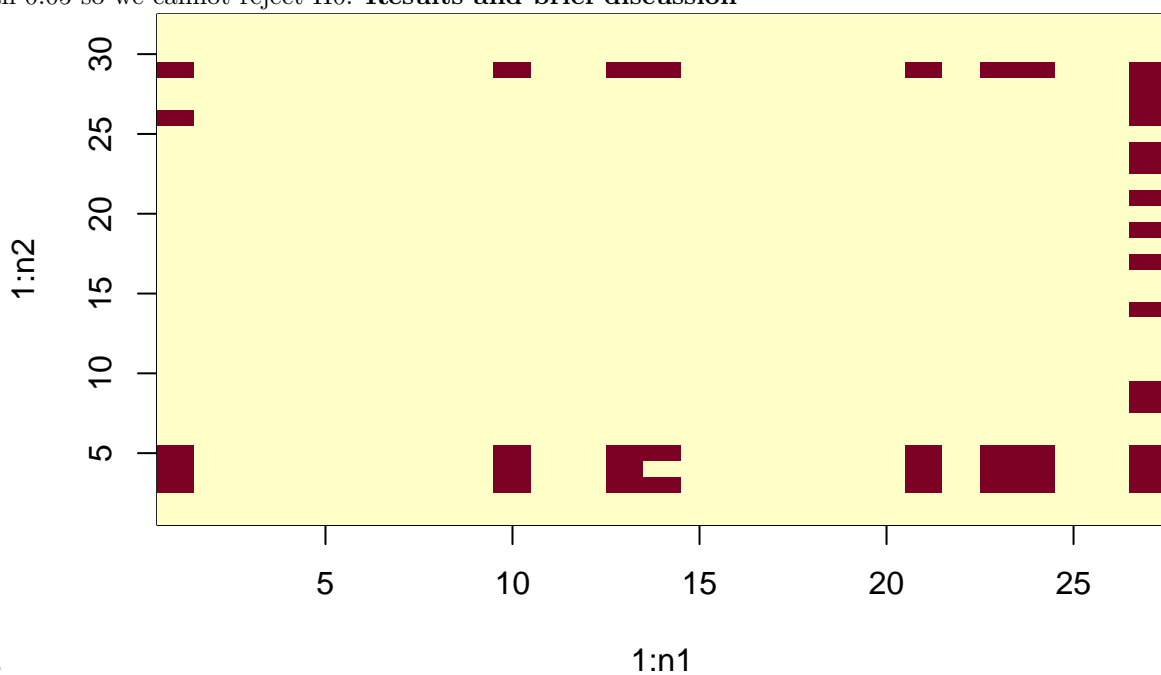labels associated to each unit.

Results and brief discussion
123

Question 1.4

Synthetic description of assumptions, methods, and algorithms
We perform here a TWO-SIDED MANN-WHITNEY U-TEST. H0: P(CO2_high_income >
CO2_middle_income) = 0.5 H1: P(CO2_high_income > CO2_middle_income) != 0.5 The pvalue is bigger
than 0.05 so we cannot reject H0. **Results and brief discussion**

123