Homework 1

Imputing missing data and detecting outliers

The Russet data set. In 1964 Russet tried to find the relation between political instability of countries and the economical and agricultural inequality.

The collected data refer to 47 countries and 9 variables on the period after the Second World War (~1945-1962).

List of manifest variables (indicators)
- GINI: concentration index (Lorenz curve)
- FARM: % of small farmers with 50% of land
- RENT: % of farmers not cultivating their own land
- GNPR: GNP per capita ($ 1955).
- LABO: % active population in Agriculture
- INST: total number of prime ministers during 1945-1961
- ECKS: index of violent conflicts during 1946-1961
- DEAT: deaths during demonstrations 1950-1962
- DEMO: 1=stable, 2=instable, 3=dictatorship

Russet B.M. (1964) Inequality and instability. World Politics, 21, pp 442-454.

Using the Russet data set, do the following exercises:

1. Impute missing values and save the result obtained.
2. Write the R script to find multivariate outliers and detect them.