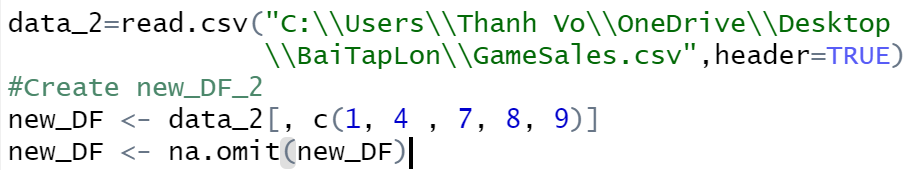
# II.Phần riêng:

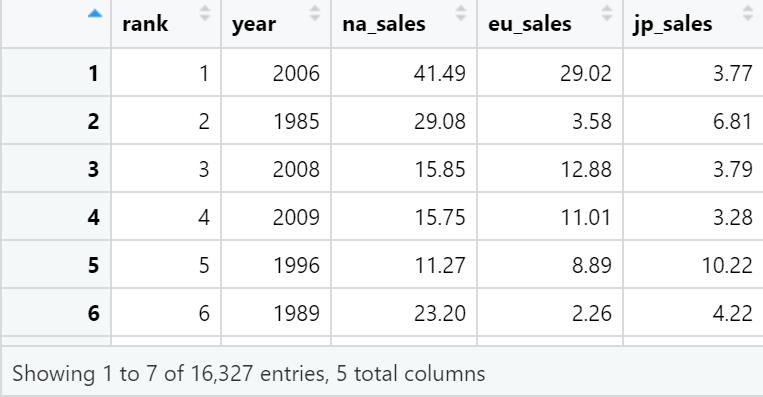
Tập tin GameSales.csv là tập số liệu về doanh thu bán game toàn cầu trong giai đoạn (1983-2017)

**Mục tiêu:** xây dựng mô hình hồi quy bội (multiple regression) để phân tích tác động của yếu tố do ta lựa chọn tác động lên na\_sales và diễn giải kết quả.

**Data cleaning:**

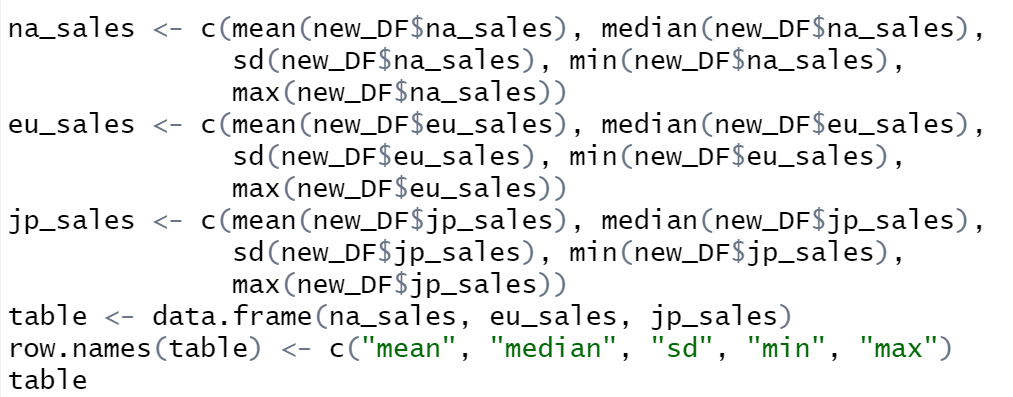


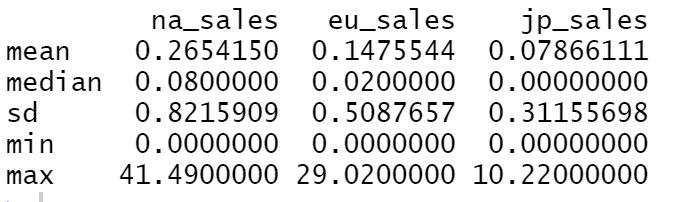
Đọc file dữ liệu và chọn các biến chính mà ta quan tâm. Sau đó loại bỏ hàng có giá trị NA.



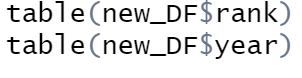
**Data visualization:**

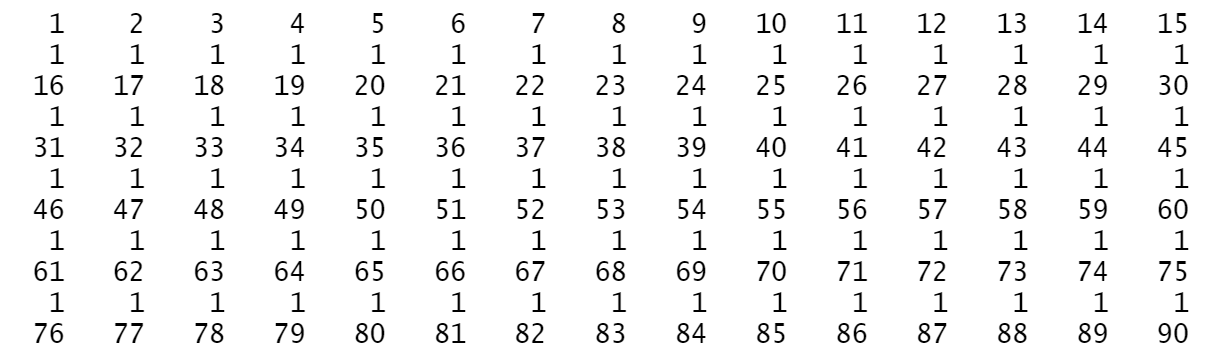
Đối với các biến liên tục, ta tính các giá trị thống kê mô tả bao gồm: trung bình, trung vị, độ lệch chuẩn, giá trị lớn nhất và giá trị nhỏ nhất. Xuất kết quả dưới dạng bảng.

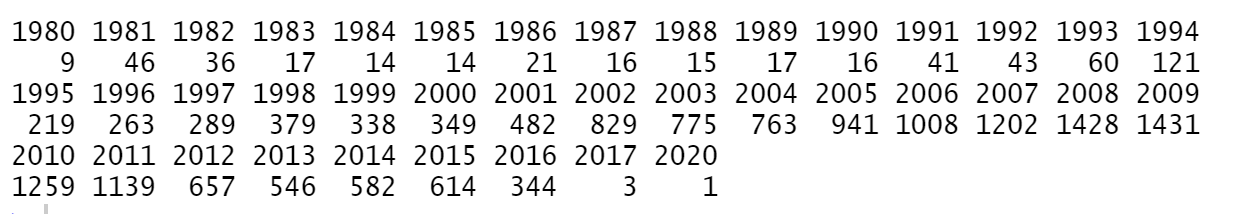
****

****

Đối với các biến phân loại, lập một bảng thống kê số lượng cho từng chủng loại:







Dùng hàm hist() để vẽ đồ thị phân phối của biến na\_sales dưới dạng log.



