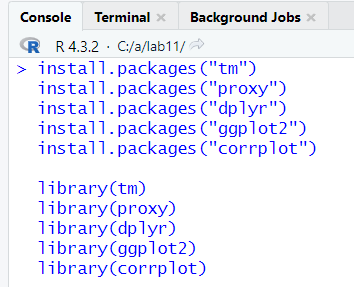
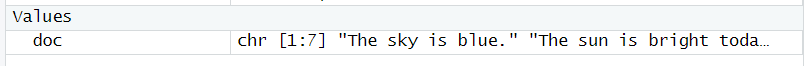
LAB 11:

Install và import thư viện  


Tạo một vector có tên là **doc** để chứa danh sách các câu văn bản

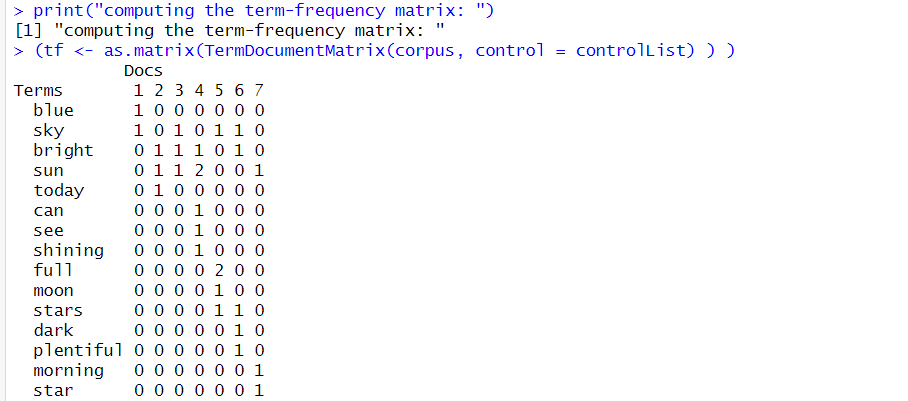




Tạo đối tượng corpus để thực hiện các phương pháp tiền xử lí văn bản



Tạo ma trận tần số từ đối tượng corpus

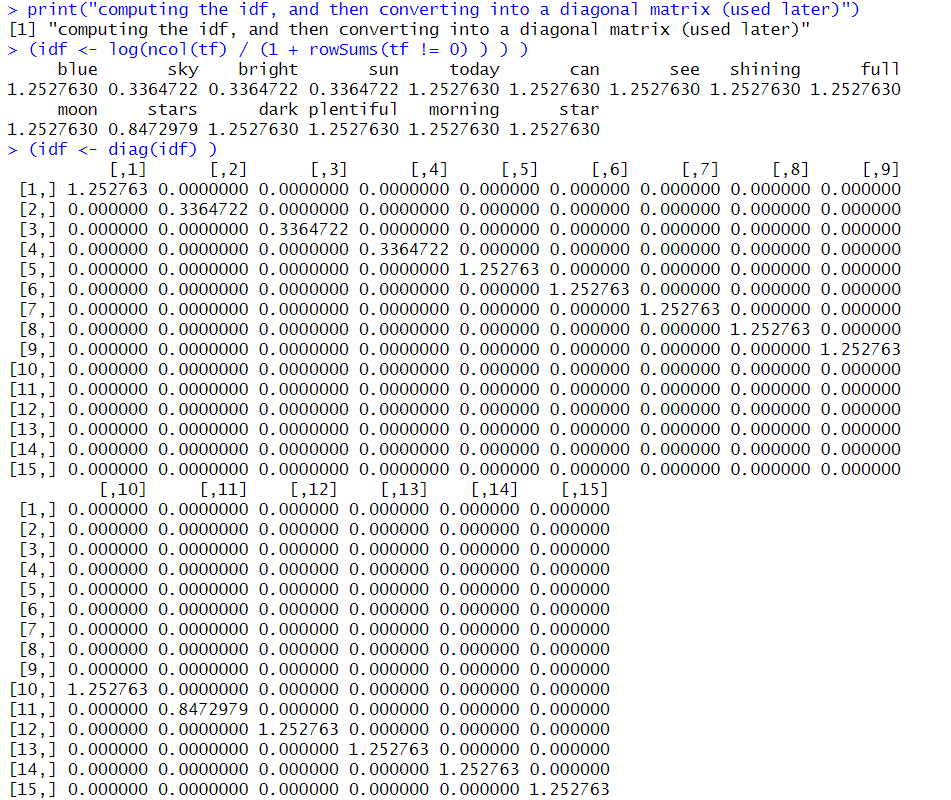


Vẽ biểu đồ hiển thị tương quan giữa các từ trong ma trận tần số **tf** đã tính toán từ văn bản

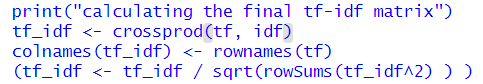




Tính toán idf và chuyển nó thành ma trận đường chéo



Tính toán tf\_idf và chuẩn hóa các vector dòng của ma trận TF-IDF



vẽ biểu đồ hiển thị mức độ tương quan giữa các từ trong ma trận TF-IDF

