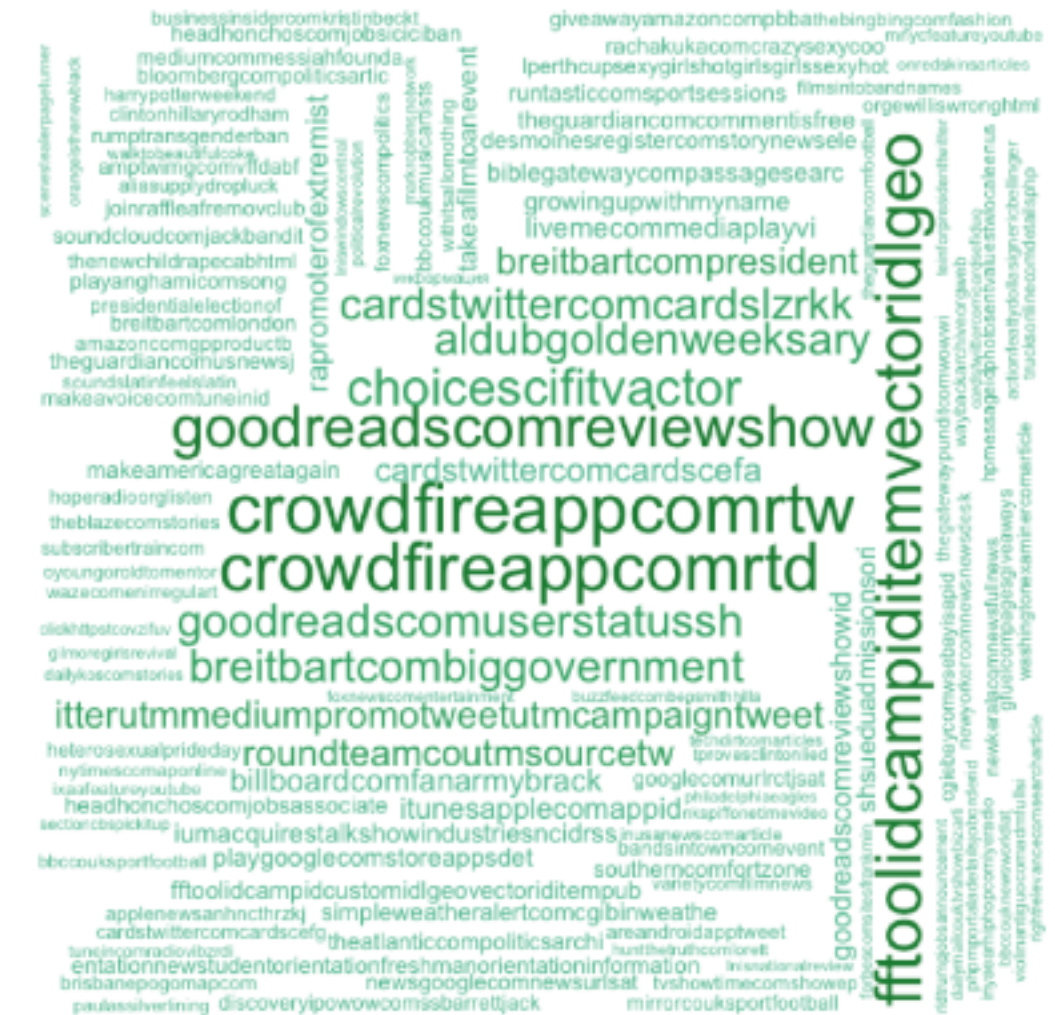
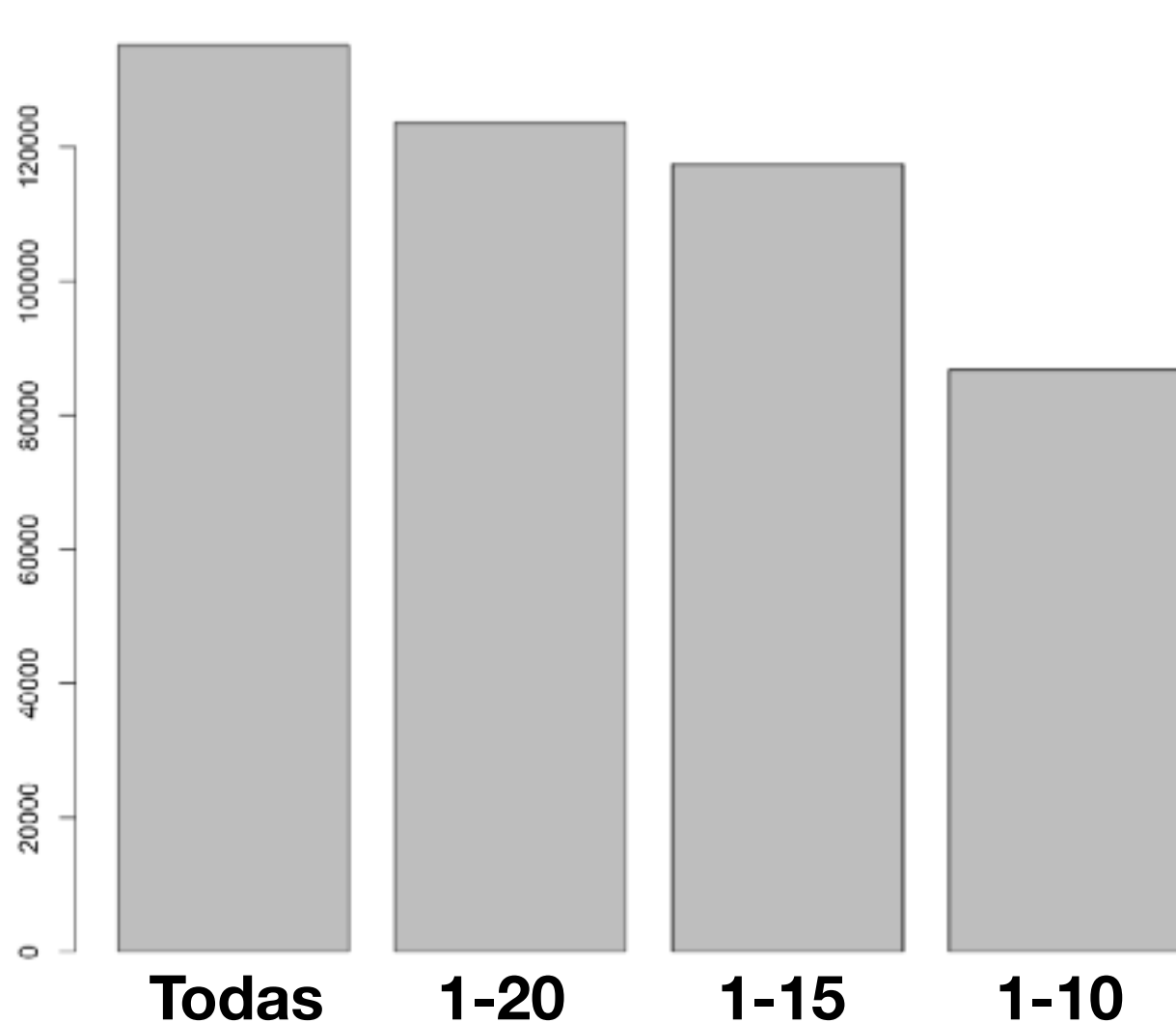


Preprocesado: Selección de datos

Es necesario obtener la **matriz de términos**.

Problema: Matriz muy grande imposible de mantener en memoria.

Solución: Eliminar elementos poco comunes (Shalton).



Preprocesado: Bigramas

