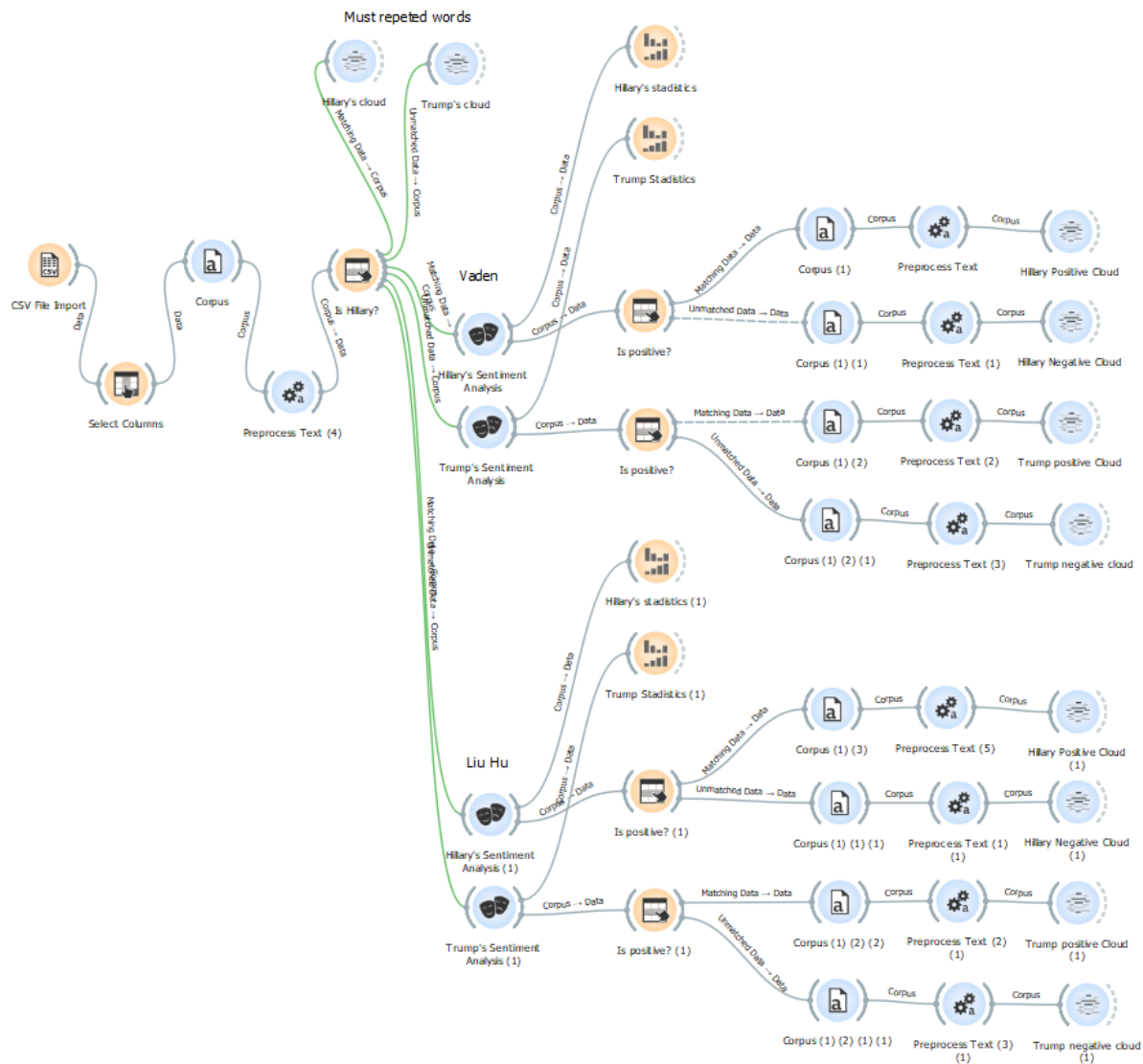


Análisis de Sentimiento en Redes Sociales

Elecciones USA 2016: Clinton vs Trump

Para este proyecto realizará el análisis de sentimiento de los tweets publicados por Hillary Clinton y Donald Trump que se encuentran en el dataset de Orange.



Conteste las siguientes preguntas:

1. ¿Cuáles son las palabras que más se repiten en los tweets de **Clinton**?

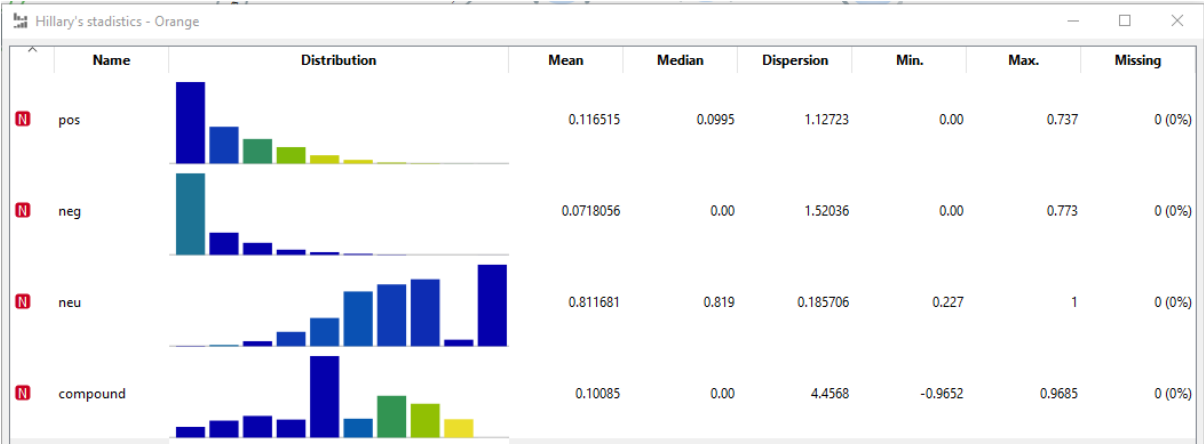
Realizamos un proceso de limpieza de las palabras a analizar de forma que no tuviéramos números, conexiones lingüísticas como “and, of, the”, restos de hipervínculos y caracteres especiales como comas y puntos.

Con el conjunto de palabras limpio era de esperarse que sus propios nombres estuvieran al tope de la lista.

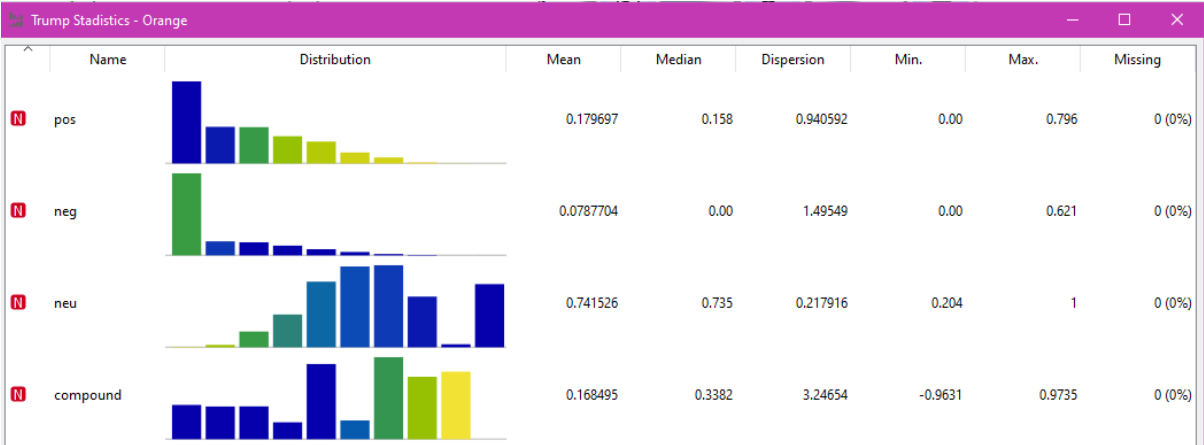
Algo a destacar de la nube de palabras de Hilary es su uso de palabras como “women”, “People”, “america”, “country” y “families”

3. ¿Cuál es el sentimiento promedio en los tweets de Clinton vs Trump?
Nos basamos en el análisis con el método de Vaden. En este podemos ver como Clinton tiende a no tuitear negativamente y mantenerse neutra. Trump tiende a tuitear con mayor frecuencia de forma negativa.

a. Clinton



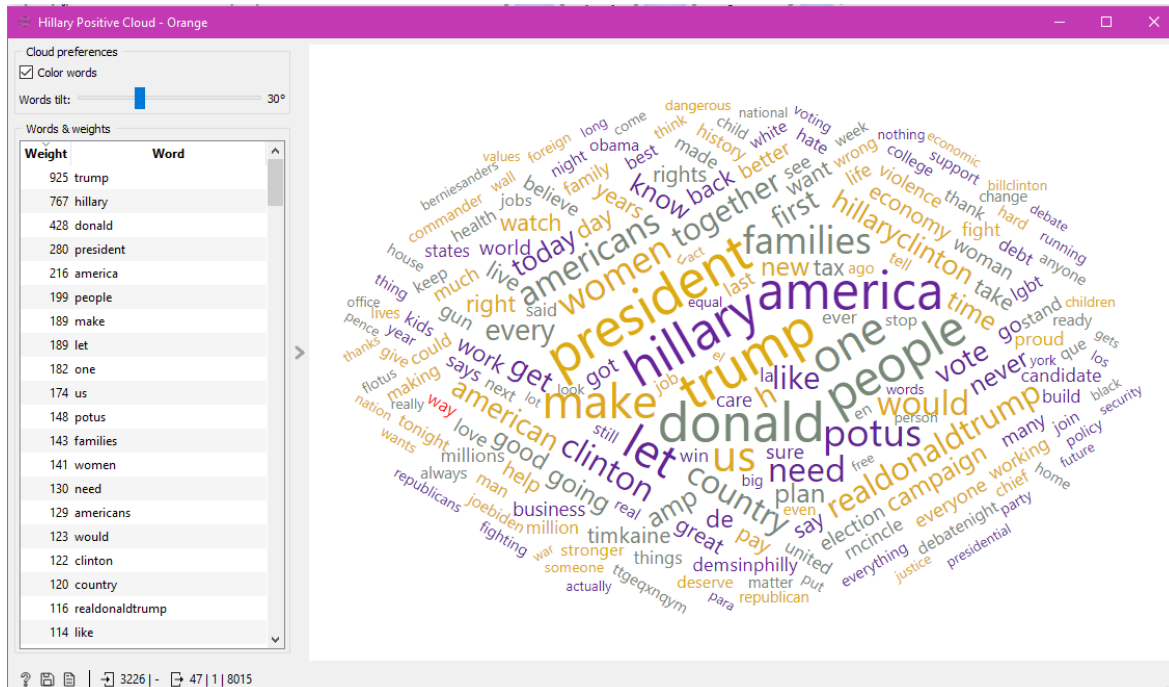
b. Trump



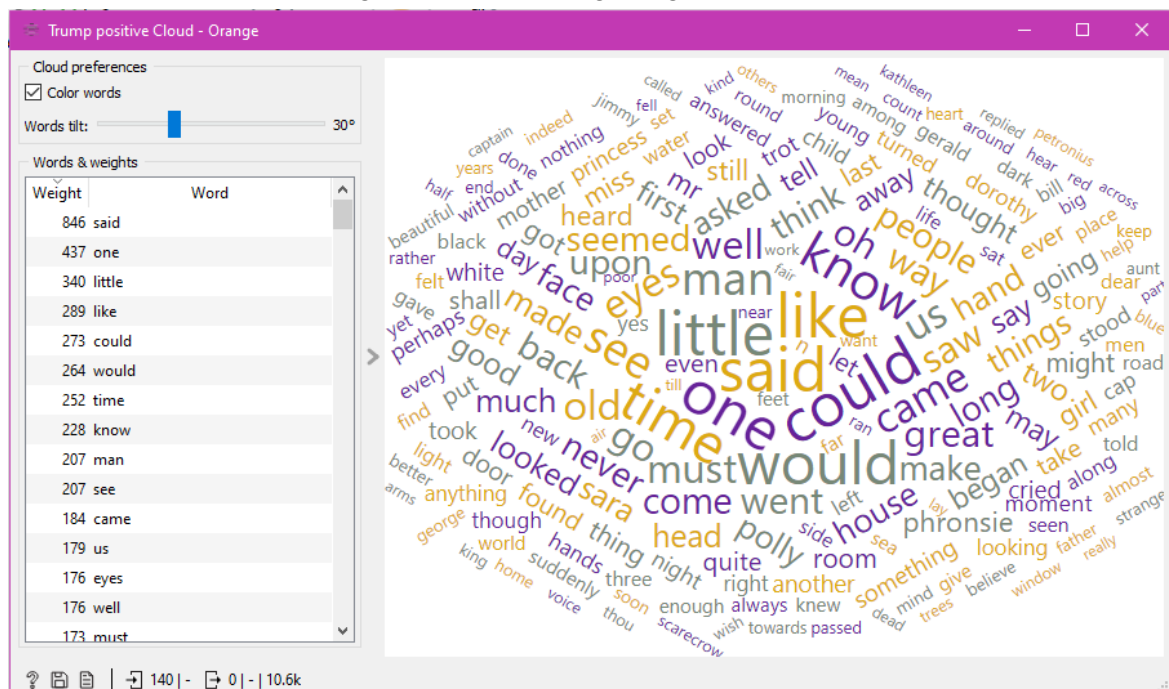
4. ¿Qué palabras son las que predominan en los tweets **positivos** de Clinton y Trump?

a. Vaden

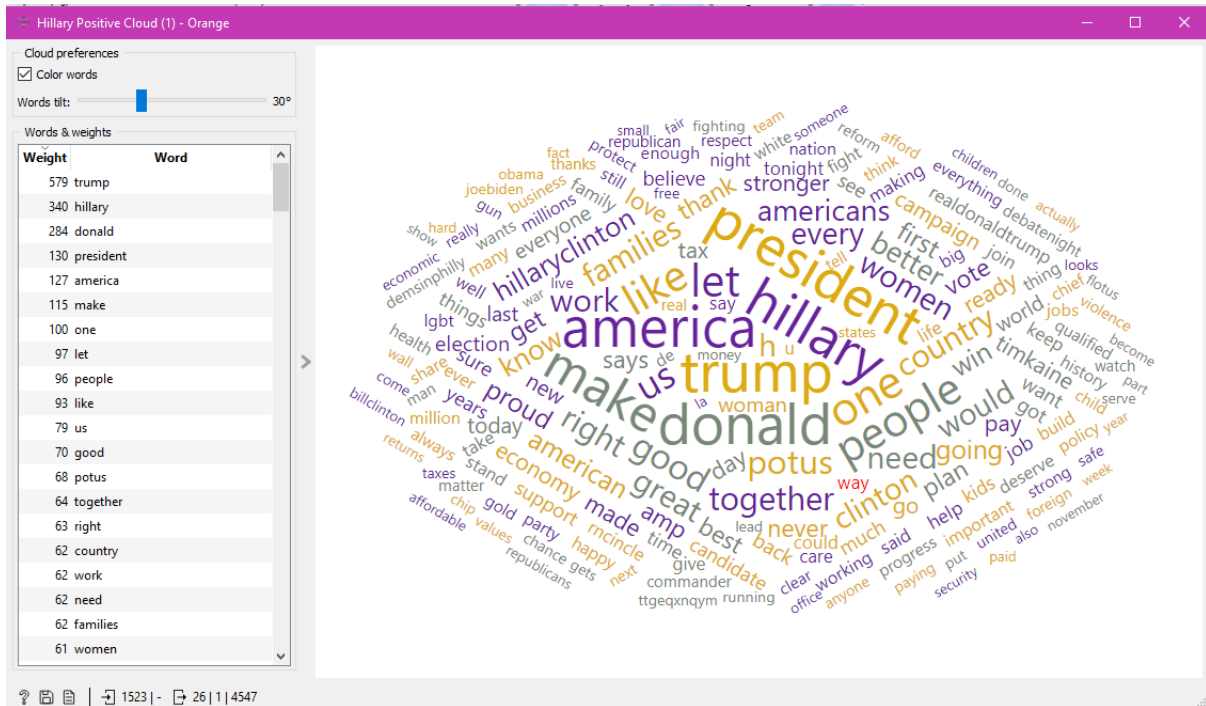
- i. **Clinton** tiende a hablar con sus posibles votantes usando palabras como “people”, “families”, “american”. Su discurso está enfocado en acciones, por eso tiene las palabras “make”, “need”, “would” entre las más usadas.



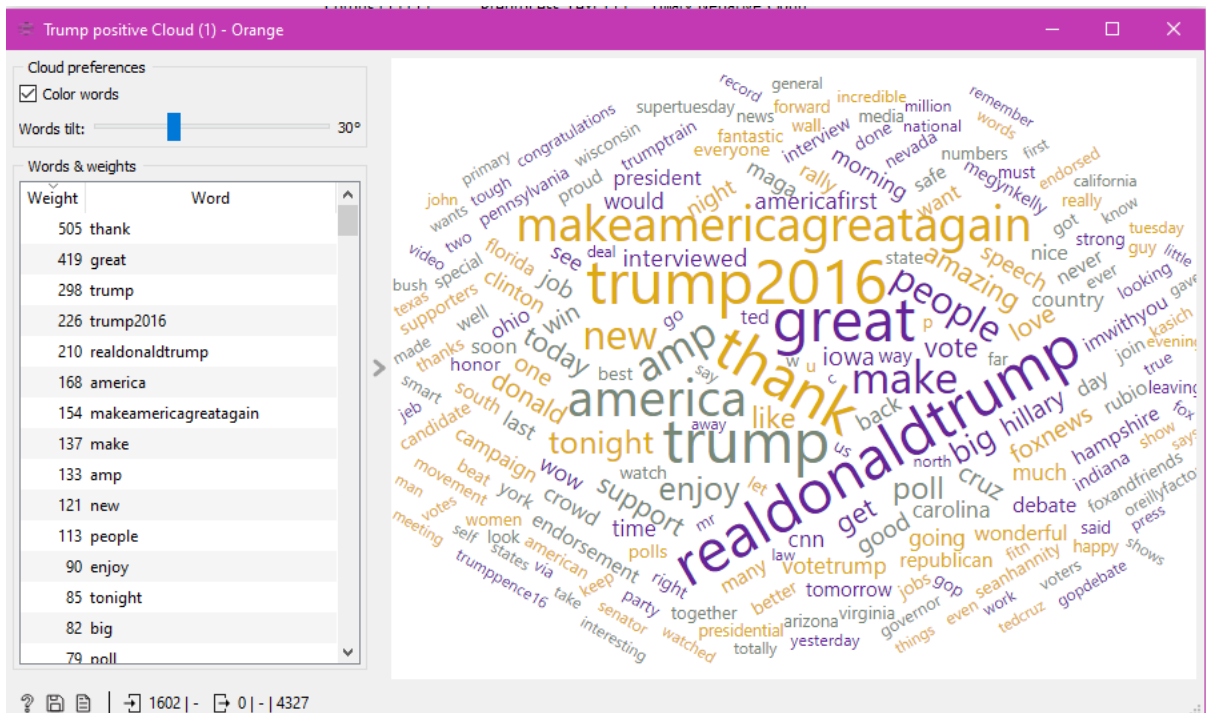
- ii. **Trump** tiene un patrón bastante curioso, a pesar de tener su nombre como una de las palabras más repetida, al momento de separar y analizar las palabras con connotación positiva, ya no vemos su nombre. Parece ser que Trump tuiteando sobre sí mismo es catalogado como un rasgo negativo.



i. **Clinton** sigue teniendo palabras enfocadas a sus propuestas como las más usadas y sigue teniendo un lenguaje con sus votantes para hacerlos sentir incluidos (“together”, “work”, “families”, etc) la diferencia es que seguimos viendo sus nombres dentro de las palabras más usadas.



ii. **Trump** ahora sí contiene dentro de las palabras más usadas su nombre en las múltiples combinaciones.

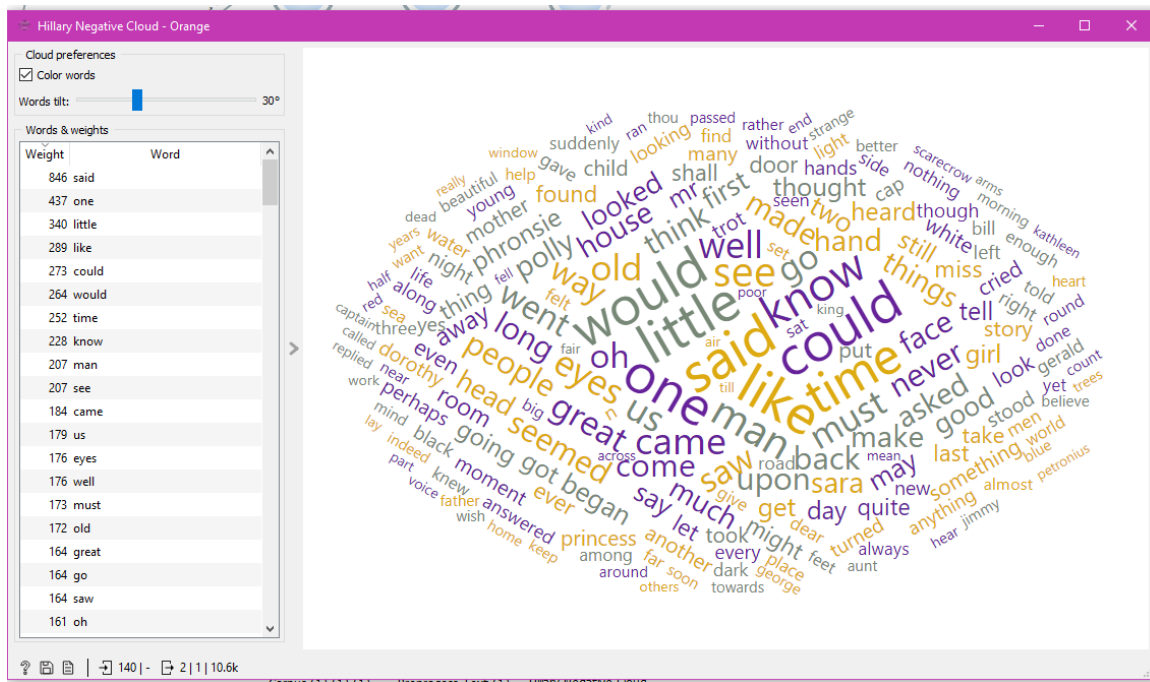


5. ¿Qué palabras son las que predominan en los tweets **negativos** de Clinton y Trump?

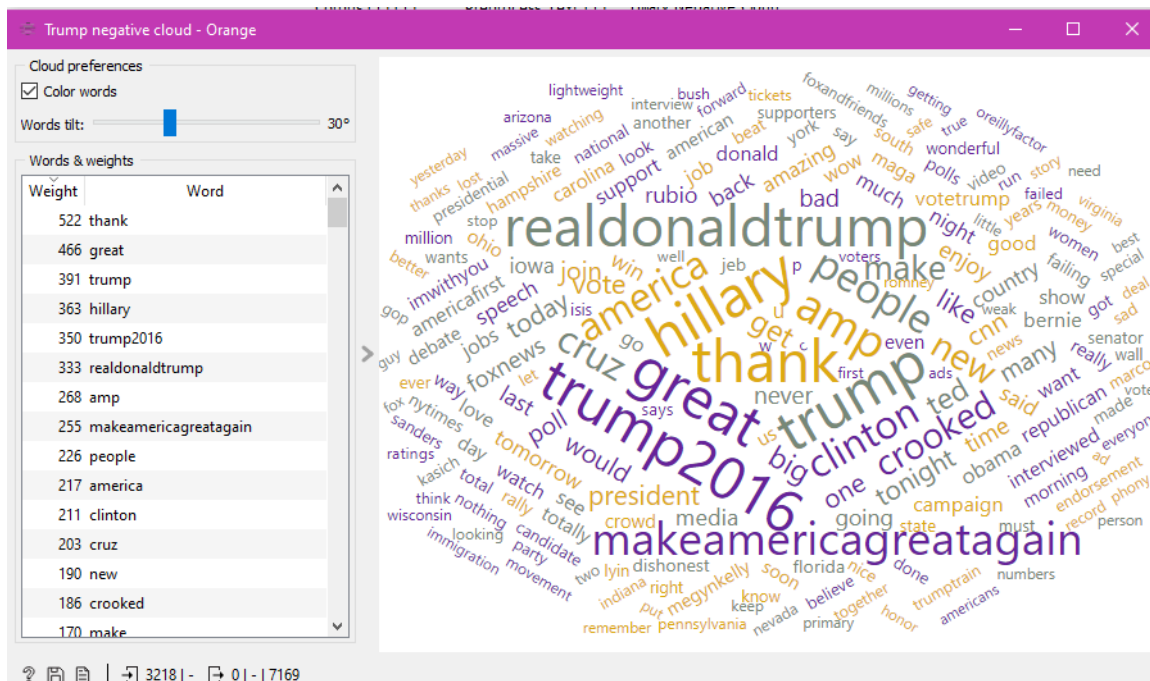
a. **Vaden**

- i. **Clinton** tiene catalogado como negativo la palabra “man”, no estamos seguros de porque ocurre esto pero su campaña usaba un discurso feminista y de empoderamiento femenino para atraer votantes, puede que por eso ocurra esto.

Lo que notamos es que las palabras que con Trump se califican como positivas, con Hilary se califican como negativas.



- ii. **Trump** tiene su propio nombre calificado como una palabra negativa junto con su famoso slogan “makeAmericaGreatAgain”.

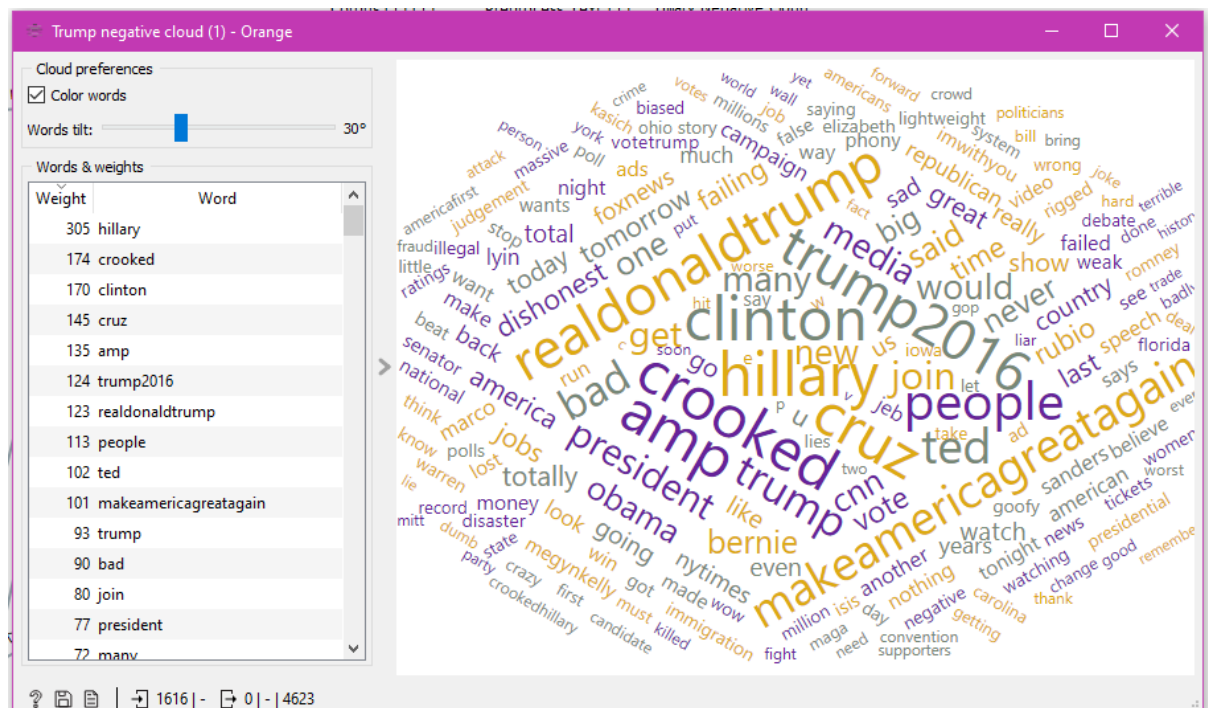


b. Liu-Hu

- i. **Clinton** tiene categorizadas como negativas cuando tuitea usando el nombre de Trump o cuando usa su propio nombre.



- ii. **Trump** ahora contiene el nombre de Ted Cruz categorizada como negativa junto con su nombre y demás personas que atacaba en su campaña como son Obama y Bernie Sanders.



6. ¿Qué diferencias se observan al analizar los tweets con el método de Liu Hu vs Vaden?

El método Vaden tiende a parecer que estuviera más inclinado a Hilary ya que categoriza su nombre como algo positivo mientras que pone a Trump, por el simple hecho de ser nombrado, como algo negativo.

Por otro lado, el método de Liu Hu no pone ninguno de los nombres como algo positivo o negativo sino que los mantiene en ambos espectros pero categoriza las palabras con las que estos venían acompañados.

7. Conclusiones

En lo personal, me costó trabajo llegar a conclusiones viendo solamente las palabras, aun estando categorizadas una palabra depende bastante de su contexto. Otro problema que tuve que solucionar fue que en twitter las imágenes se ponen como links y cualquier documento adjunto se hace de esa forma por lo que tuve que limpiar la información.