

UNIVERSIDAD DEL VALLE DE GUATEMALA

CC3084 - Data Science

Sección 10

MsC. Lynette García



Laboratorio 8: Visualizaciones estáticas

SEBASTIAN ARISTONDO PEREZ 20880

JOSE DANIEL GONZALEZ CARRILLO 20293

GUATEMALA, 9 de octubre de 2018

Decisiones para construir la visualización

Explicación de conjunto de datos

El conjunto de datos que se utilizó fue una colección de tweets relacionados a los candidatos por la presidencia de Guatemala. Además de incluir múltiples tweets como se indicó, también contiene el nombre de usuario de la persona que realizó el tweet, la cantidad de likes y retweets, los hashtags del tweet, los usuarios mencionados e información extra como cierta metadata. En realidad, son dos conjuntos de datos, uno para Sandra Torres y otro para Bernardo Arévalo, pero ambos tienen los mismos campos, por lo que se unieron.

Decisiones del diseño basadas en la teoría de color

Al momento de realizar la visualización nos topamos con 3 puntos claves los cuales nos permitieron elegir los colores a utilizar. El más importante era que estábamos presentando resultados sobre partidos políticos de Guatemala por lo tanto debíamos evitar cualquier color de los partidos involucrados, de esta manera la visualización permanecería imparcial. El segundo punto trata sobre la confianza de transmitir información, al ser material delicado debíamos elegir colores que nos ayudarán a llegar al público, por lo tanto el azul que simboliza la confianza debía formar parte de nuestra paleta. El tercer punto se trata del color que el mundo asoció hacia las redes sociales, al investigar vimos que las personas asocian el color azul con la redes.

Dado que teníamos un color que sí o sí debía de estar, utilizamos la teoría de la tríada para elegir el resto de colores. Si vemos el círculo cromático podemos ver que utilizando una tríada cerrada el azul combina con el naranja y el amarillo. Para el amarillo se utilizó un tinte bastante cercano al blanco y el naranja fue más cercano al hue.

Decisiones del uso de los gráficos

Los gráficos que se usaron fueron de barras y de pie. Por un lado, se usaron gráficos de barra para representar la popularidad de los usuarios en la red social. Con esto, se buscó evidenciar la cantidad de tweets realizados por usuarios en los datos relacionados con Sandra Torres y con Bernardo Arévalo. De esta manera, se tuvo una forma clara de comunicar los usuarios más activos en la red social y evidenciar tendencias en ambos conjuntos de datos. Por otra parte, se usaron gráficos de pie para mostrar el sentimiento de los tweets. Se tomó la decisión de usar este gráfico debido a que se tenían pocas clases por representar y se buscó enfatizar el sentimiento más frecuente, apelando a la variable preatencional del tamaño. Esto se realizó con ambos candidatos.

¿Por qué se presentan los resultados en ese orden?

Es necesario transmitir los resultados poco a poco, por esta razón se optó por primero mostrar el contexto de las redes sociales de ambos candidatos. Explicar la presencia que tuvo cada uno y cómo es que estos interactúan entre sí. Una vez se contextualizó sobre la cantidad de retweets y su magnitud se procedió a explicar cuál fue la distribución de los sentimientos encontrados en todos los tweets. De esta forma es posible contar una historia con los datos y volver la información mucho más digerible.