

Visualización de Datos

A continuación, se presenta tres diferentes técnicas de visualización de datos mediante la utilización de flourish y jupyter notebooks.

Primera Visualización – Nube de Palabras o Word Cloud

1. [Puntos 1 y 2] Presentaros siguiendo el esquema de la PEC 1.

Mi nombre es Erika Martinez Soria y estoy presentando la primera técnica de visualización Word Cloud o Nube de Palabras de la PEC2 de la asignatura de Visualización de Datos del Máster Universitario de Ciencia de Datos de la UOC.

2. [Puntos 1 y 2: 5%] Indicad donde está colgada la visualización y cómo acceder.

La visualización está creada en la página web flourish y se accede mediante este link:

<https://public.flourish.studio/visualisation/12000792/>

3. [Punto 3: 30%] Definid cada técnica de visualización de forma general: nombre, origen, descripción/funcionamiento, ejemplos de aplicación, etc.

Un Word o Tag Cloud es una representación visual de datos de texto, que se utiliza para representar metadatos de palabras clave. Por ejemplo, un Word Cloud de los países del mundo, donde el tamaño del territorio determinaría el tamaño o color del nombre de cada país. También se usa para analizar discursos, por ejemplo, de campañas electorales. En este caso, la importancia de cada palabra se muestra con el tamaño o el color de la fuente que son proporcionales a su frecuencia. Por lo tanto, es útil para ver rápidamente los términos más destacados.

4. [Punto 4: 10%]

- Describid el tipo de datos que se pueden representar con cada técnica ¿datos cuantitativos, cualitativos?
- ¿Qué estructura tienen que tener para cada técnica?).
- Explicar las limitaciones en cuanto a datos (¿hay medida mínima y máxima del juego de datos para cada técnica?).

Para esta técnica de visualización los datos cualitativos y cuantitativos. La estructura de estos contiene las palabras claves de un texto junto a su frecuencia de aparición. O como mencionamos, los metadatos cuantitativos de palabras claves, como, por ejemplo, la población de países.

Algunas de las limitaciones de esta técnica son que las palabras largas se enfatizan sobre las palabras cortas. También, que no es adecuado para un entendimiento analítico, sino más estético. Y solo es útil si el tamaño de la muestra es muy grande.

5. [Punto 5: 20%] Hacer una representación con cada una de las técnicas usando un conjunto de datos abiertos (veáis más abajo algunas fuentes de datos abiertos). En

total hay que hacer 3 representaciones simples con 3 conjuntos de datos escogidos por el estudiante.

6. [Punto 6: 30%] Comentar brevemente las tres representaciones indicando qué se representa y qué o qué aspectos muestra o demuestra cada representación.

Los datos contienen los 150 países con mayor territorio en kilómetros cuadrados. Como podemos ver, mientras más territorio más grande es el nombre del país.

7. [Punto 7: 5%] 250 – 300 palabras. Cada una de las tres presentaciones deben tener una duración de unos 2-3 minutos. En total, el vídeo (único) con las tres técnicas no puede exceder de 7 minutos. Se penalizará salir de ese rango, porque se valora la capacidad de síntesis y de comunicación

Segunda Visualización: Gráfico de Íconos o Icon Chart

1. [Puntos 1 y 2] Presentaros siguiendo el esquema de la PEC 1.

Mi nombre es Erika Martinez Soria y estoy presentando la segunda técnica de visualización Icon Chart o Gráfico de Íconos de la PEC2 de la asignatura de Visualización de Datos del Máster Universitario de Ciencia de Datos de la UOC.

2. [Puntos 1 y 2: 5%] Indicad donde está colgada la visualización y cómo acceder.

La visualización está creada en la página web flourish y se accede mediante este link:

<https://public.flourish.studio/visualisation/12011822/>

3. [Punto 3: 30%] Definid cada técnica de visualización de forma general: nombre, origen, descripción/funcionamiento, ejemplos de aplicación, etc.

Un Icon Chart o Gráfico de íconos, o también conocido como pictograma, usa íconos para brindar una vista general atractiva de pequeños conjuntos de datos discretos. Los íconos representan una categoría de los datos; por ejemplo, los datos sobre la población de países usarían íconos de personas, donde el nombre de cada país sería la categoría y cada icono dentro de la categoría podría representar 1 millón de personas. Los conjuntos de datos se comparan uno al lado del otro en columnas o filas de iconos, y así poder comparar cada categoría entre sí. El uso de iconos tiene la ventaja de ayudar a superar diferencias de idioma, cultura y educación.

4. [Punto 4: 10%]
 - Describid el tipo de datos que se pueden representar con cada técnica ¿datos cuantitativos, cualitativos?
 - ¿Qué estructura tienen que tener para cada técnica?).
 - Explicar las limitaciones en cuanto a datos (¿hay medida mínima y máxima del juego de datos para cada técnica?).

Para esta técnica de visualización los datos deben ser cuantitativos. La estructura de estos contiene los nombres de las categorías en una columna y los datos numéricos en otras.

Algunas limitaciones del pictograma son que, al usarlo con grandes conjuntos de datos, los valores en el gráfico son difíciles de contar. También está el hecho de que si se muestran íconos parciales puede llegar a confundir lo que representan. Y finalmente, aunque puede ser una alternativa divertida a un gráfico de barras tradicional, esta no es tan precisa.

5. [Punto 5: 20%] Hacer una representación con cada una de las técnicas usando un conjunto de datos abiertos (veáis más abajo algunas fuentes de datos abiertos). En total hay que hacer 3 representaciones simples con 3 conjuntos de datos escogidos por el estudiante.
6. [Punto 6: 30%] Comentar brevemente las tres representaciones indicando qué se representa y qué o qué aspectos muestra o demuestra cada representación.

El conjunto de datos es pequeño, contiene 5 países europeos y su población representada con los íconos.

7. [Punto 7: 5%] 250 – 300 palabras. Cada una de las tres presentaciones deben tener una duración de unos 2-3 minutos. En total, el vídeo (único) con las tres técnicas no puede exceder de 7 minutos. Se penalizará salir de ese rango, porque se valora la capacidad de síntesis y de comunicación

Tercera Visualización: Gráfico de Enjambre o Beeswarm Plot.

1. [Puntos 1 y 2] Presentaros siguiendo el esquema de la PEC 1.

Mi nombre es Erika Martinez Soria y estoy presentando la tercera técnica de visualización Beeswarm plot o gráfico de enjambre de la PEC2 de la asignatura de Visualización de Datos del Máster Universitario de Ciencia de Datos de la UOC.

2. [Puntos 1 y 2: 5%] Indicad donde está colgada la visualización y cómo acceder.

La visualización se encuentra en github y se accede mediante este link:

https://github.com/emartinezsor/M2.859/blob/main/PEC2_Visualizacion3.ipynb

3. [Punto 3: 30%] Definid cada técnica de visualización de forma general: nombre, origen, descripción/funcionamiento, ejemplos de aplicación, etc.

Un gráfico de enjambre muestra el tamaño de un grupo de elementos u observaciones al agrupar visualmente cada punto de datos individual. Este está diseñado para mostrar la distribución subyacente de los datos evitando que se organicen en la visualización de una forma que se superpongan, a diferencia de otros gráficos de dispersión como las franjas o strip charts.

4. [Punto 4: 10%]

- Describid el tipo de datos que se pueden representar con cada técnica ¿datos cuantitativos, cualitativos?
- ¿Qué estructura tienen que tener para cada técnica?
- Explicar las limitaciones en cuanto a datos (¿hay medida mínima y máxima del juego de datos para cada técnica?).

Para esta técnica de visualización los datos son cuantitativos y cualitativos. La estructura de se centra en la variable categórica situada en el eje x combinada con una variable numérica en el eje y, y opcionalmente con otras variables numéricas o categóricas representadas con los colores del gráfico.

Una de las limitaciones del este gráfico es que no visualiza bien conjuntos grandes de datos ya que los puntos que representan cada observación congestionarían la visualización.

5. [Punto 5: 20%] Hacer una representación con cada una de las técnicas usando un conjunto de datos abiertos (veáis más abajo algunas fuentes de datos abiertos). En total hay que hacer 3 representaciones simples con 3 conjuntos de datos escogidos por el estudiante.
6. [Punto 6: 30%] Comentar brevemente las tres representaciones indicando qué se representa y qué o qué aspectos muestra o demuestra cada representación.

El conjunto de datos es del Titanic del repositorio de Seaborn. En el eje x se encuentra la variable categórica con las mujeres y hombres, el eje y con el precio de pasaje y los colores indican si sobrevivieron o no. La gráfica muestra como a pesar del costo de su pasaje, la mayoría de hombres no sobrevivió, al contrario de las mujeres.

7. [Punto 7: 5%] 250 – 300 palabras. Cada una de las tres presentaciones deben tener una duración de unos 2-3 minutos. En total, el vídeo (único) con las tres técnicas no puede exceder de 7 minutos. Se penalizará salir de ese rango, porque se valora la capacidad de síntesis y de comunicación