



Máster Ciencia de Datos – *Data Science*

Visualización de Datos / Profesor colaborador: Jaume Nualart Vilaplana

Fernando Antonio Barbeiro Campos – [fbarbeiro@uoc.edu](mailto:fbarbeiro@uoc.edu)

## **A7: Creación de una visualización de datos con Tableau (PEC3)**

Entrega prevista: 19 de Diciembre de 2018.

# Índice General

<b>Índice General</b>	<b>1</b>
<b>Introducción</b>	<b>2</b>
Enunciado	2
<b>Respuestas</b>	<b>3</b>
URL (de Tableau Public)	3
Explicación sobre la pregunta que contesta el dashboard y las elecciones hechas	3
Referencias bibliográficas	4

# Introducción

## Enunciado

En esta actividad cada estudiante deberá crear una visualización de datos siguiendo la metodología usada en el ejemplo propuesto en la actividad A6.

Concretamente, se propone mejorar la visualización mostrada aquí:

<http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>

Los datos, adjuntados en esta actividad en formato Excel / LibreOffice, muestran robos de datos (*data breaches*) que han sufrido distintas empresas. El objetivo es crear varias visualizaciones, coordinadas en un dashboard, que permitan entender el método usado en la violación de datos, las empresas, y la cantidad de datos robados.

En caso de querer crear una visualización con otro conjunto de datos distinto de complejidad similar, es necesario obtener la validación previa del consultor. Podéis usar como ejemplos diferentes conjuntos de datos disponibles en el recurso Statista, accesible a través de la Biblioteca de la UOC: Biblioteca -> Recursos -> Bases de datos -> letras -> Statista. <https://0-www-statista-com.catalog.uoc.edu/>

Cada alumno deberá entregar un documento con la URL (de Tableau Public) y el dashboard final explicando, con una extensión de no más de media página, la pregunta que contesta el dashboard propuesto y las elecciones a nivel de visualización que se han tomado. Debéis adjuntar también los ficheros de libros de trabajo twb, los libros de trabajo empaquetados twbx, y los archivos de datos y sus fuentes (opcional) extra, si habéis agregado algún dato más para mejorar el contexto de la visualización.

# Respuestas

## 1. URL (de Tableau Public)

Por aquí, el enlace de la visualización en Tableau Public:

<https://public.tableau.com/views/FBarbeiroDataVizPec3/DataBreachesDashboard>

## 2. Explicación sobre la pregunta que contesta el dashboard y las elecciones hechas

Es crucial entender todas las perspectivas de *data breaches* si queremos evitar los mismos problemas en el futuro, por lo tanto, la sabiduría de cómo evitar cada método, bien como la evolución de cada ataque a través del tiempo es un elemento importante para tratar el tema. No obstante, conocer los errores de empresas y aprender de ellos también es primordial, por ello el tema de las historias y cantidad de datos.

Sin más preámbulos, la pregunta que nos proponemos a responder es: **¿Cuáles son el método usado en la violación de datos, las empresas, y la cantidad de datos robados?**

---

Personalmente, creo que la visualización de punto de partida estaba interesante, sin embargo, no era debidamente informativa en el ámbito de diferenciar los métodos de vulnerabilidades y ni tampoco era efectiva en el aspecto de diferenciar cantidades (pues habían inúmeras bolas que dificulta entender donde están las principales diferencias).

Mi decisión por usar **barras apiladas** ataca la dificultad de definición de los métodos de vulnerabilidad dejando evidentes por colores cada método y empresa. Además, hay la posibilidad de filtro por método, que aporta bastante para aislar en un escenario de estudios más profundizados sobre el problema (dicho, una vez cambiado, el filtro se va a aplicar para todos las *sheets* del *dashboard*).

Mantuve la idea del gráfico de burbujas pero con enfoque de aclarar los sectores afectados por ataques (como no son muchos sectores distintos, es sencillo visualizar la amplitud de cada sector).

Por último, también he añadido un último gráfico de líneas para pérdidas de información por año y método de ataque.

Concluyendo: con el dashboard creado, además de visualizar las cantidades en todos los gráficos (atención para el efecto de *mouseover* del primer) he logrado obtener una visualización que añade información de sectores y habilita el usuario a entender y sobretodo tener una visión también historial de los métodos y empresas que sufrieron ataques como proponía la pregunta principal.

### 3. Referencias bibliográficas

[1] **Information is Beautiful. (2018).** “*Balloon Race: Data Breaches - LATEST*”. *Information is Beautiful*. [artículo en línea]. [Fecha de consulta: 10 de Diciembre del 2018].

<<https://docs.google.com/spreadsheets/d/1iOoIJJMRG-7t1GT-mr4smaTTU7988yXVz8nPlwaJ8Xk/edit#gid=2>>