

Asignatura	Datos del alumno	Fecha
Herramienta de Visualización	Apellidos: Aranguren Ubierna	11/05/2025
	Nombre: Andoni	

Actividad Grupal: Análisis tendencias en D3.js

Exploración Interactiva del Catálogo de Steam con D3.js

Autores

Andoni Aranguren Ubierna - Jesús Rugarcía Sandía - Alejandro Adrián López González

Resumen

En este trabajo se muestra la **capacidad de D3.js para convertir un conjunto de datos recopilados de la plataforma Steam en gráficos que nos permitan dar explicación a varias preguntas**. Partiendo del *Steam Games Dataset* (Kaggle, ~73 000 títulos publicados entre 1997 y 2025), construimos seis gráficos avanzados cuya nomenclatura en D3 es: *Chord Diagram*, *Pie Chart dinámico*, *Zoomable Sunburst*, *Treemap jerárquico*, *Word Cloud* y *Race Bar Chart*. Estos gráficos han sido elegidos por cumplir distintos objetivos de análisis visual. Algunos son casi exclusivos de D3.js mientras que otros son más comunes en paquetes como matplotlib o matlab.

La interacción, el *drill-down* y la fusión de vistas permiten responder preguntas clásicas del mercado del videojuego (peso de los géneros, concentración editorial,

diversidad lingüística o sinergias entre géneros populares) **sin abandonar el navegador**. En conjunto, las visualizaciones confirman que:

- El ecosistema de videojuegos está dominado por un grupo de *publishers* (Treemap + Sunburst) aunque los géneros *indie-friendly* irrumpen en casi todas las ramas de la jerarquía.
- Los géneros Action, Adventure y Casual forman un núcleo fuerte del catálogo, dada su elevada interconexión. A través del Chord Diagram se observa que estos géneros no sólo coexisten frecuentemente entre sí, sino que además tienen vínculos significativos con estilos como Indie o Singleplayer. Este hecho sugiere que los títulos más representativos combinan acción o exploración con accesibilidad (Casual) y profundidad narrativa (Singleplayer), una fórmula recurrente en las listas de tendencias.
- El soporte de lenguajes es todavía un cuello de botella: el análisis (Bar chart race) muestra un fuerte monopolio del inglés desde que hay datos de publicación estando siempre a la cabeza de lenguajes preferidos.
- Se encuentra una diversidad de palabras bastante uniforme entre las palabras más comunes de las descripciones de los juegos con “game”, “new”, “world” y “experience” como palabras más frecuentes entre los juegos de las 20 grandes publishers (Wordcloud).

Todas las transformaciones necesarias para tener un dataset limpio, se han realizado en *Python + Pandas* y se proyectan en el navegador gracias a D3. El resultado demuestra, la capacidad de **D3 como herramienta efectiva** para trabajar de manera visual con la riqueza analítica del dominio de los videojuegos cuya facturación global supera ya al cine y la música juntos.

Además, el potencial de D3 se dispara si incluimos un planteamiento de desarrollo que incluya la aplicación de filtros de datos ya que permiten personalizar la experiencia de usuario a la hora de responder a preguntas de mercado concreta con, por ejemplo: Un Product Owner de una desarrolladora que quiere elegir los idiomas a los que le conviene traducir su juego y elegir una descripción del mismo puede

filtrar los datos para que sólo tenga en cuenta juegos exitosos (con muchos jugadores) con la categoría “aventura”.

Introducción

La industria moderna del videojuego es una “cancha de juego” en la que tanto las grandes desarrolladoras como equipos de una sola persona compiten en la creación de juegos exitosos, creando anualmente miles de títulos nuevos. Para analizar esta industria es por lo tanto necesario moverse a la misma velocidad que al ritmo de los **datos** (Battiato, S., 2021) y (Goh, E., Al-Tabbaa, O., 2023): conteos de usuarios concurrentes, *reviews* minuto a minuto, descargas de *DLC*, *micro-transactions*, “wishlist velocity”... Cada título llega acompañado de decenas de variables de mercado definidas por datos, técnicos y sociales que, puestos en común, **perfilan tendencias en el ocio interactivo contemporáneo**.

Sin embargo, una tabla de 70 000 filas y 90 columnas genera la vez un gran valor y un obstáculo a la hora de trabajar con esos datos de manera visual. Aquí entra **D3.js**:

- **D3 data-driven**: D3 sigue un enfoque centrado en los datos, lo que lo hace ideal para trabajar con conjuntos de datos variados. Su gran ventaja es que permite vincular directamente los datos del CSV con los elementos visuales de la página, y aplicar transformaciones de forma sencilla y declarativa.
- **Colección de layouts**: Treemap, Partition, Chord, Pie, Pack, Force... Que son a la vez gráficos de gran atractivo visual y algoritmos de disposición espacial que, cuando se aplican a los datos de los videojuegos, desvelan estructura de mercado, jerarquía editorial o vecindad temática.
- **Animación e interacción sin plugins**: filtros, *hover*, *zoom*... refuerzan la exploración. El observador puede interactuar con el gráfico para verificar hipótesis.

Motivación y justificación de la fuente de datos

El objetivo de esta actividad es demostrar el potencial de D3.js, consideraciones del desarrollo de la plataforma escalable y una estructura modular. Elegir el *Steam Games Dataset público* (Martin Bustos Roman, 2022) cumple tres objetivos:

1. **Volumen y variedad:** datos numéricos densos (picos de jugadores), cadenas ruidosas (idiomas), listas múltiples (géneros, etiquetas) y fechas. Ideal para exprimir las funciones de *nesting*, *roll-up* o *hierarchy*.
2. **Relevancia socioeconómica:** el mercado global del videojuego roza los 190kM\$ (Newzoo, 2024) y se halla en plena ebullición *indie* vs. *AAA*. Las conclusiones que se pueden sacar son un buen reto para *business-intelligence* real.
3. **Facilidad Explicativa:** al mostrar información visual de títulos como *Fortnite* (battle royale) frente a *League of Legends* (MOBA), o la importancia de un gigante como *Valve* frente a un estudio indie como *Innersloth*, el lector reconoce de inmediato los referentes y puede centrarse en lo que revelan las visualizaciones.

Objetivos del estudio

1. **Ilustrar patrones de concentración editorial** mediante visualizaciones jerárquicas (Sunburst, Treemap) que permita ver las apuestas y diversificación de los gigantes de la industria.
2. **Detectar sinergias de género** con un Chord Diagram que cuantifique la co-ocurrencia en fichas de Steam que puedan guiar en la creación de títulos con temáticas consolidadas o novedosas.
3. **Comparar la popularidad de los idiomas y el esfuerzo de traducción** de los publishers para tomar mejores decisiones a la hora de elegir partidas de localización de los títulos.

4. **Visualizar la diversidad lingüística y frecuencia de palabras en descripciones**
con una nube de palabras que aúne frecuencia y accesibilidad global para entender el estilo de escritura de los publishers más exitosos.

En los siguientes apartados se presenta el detalle de cada gráfico, la lógica de su selección dentro del catálogo de D3 y las conclusiones finales.

Metodología

Los datos brutos son los del *Steam Games Dataset publicado en Kaggle* (Martin Bustos Roman, 2022). Con Python y pandas se realiza una limpieza mínima (fechas coherentes, separación de listas, depuración de nulos) y, a continuación, se extrae para cada gráfico el subconjunto exacto de columnas que necesita: géneros y su co-ocurrencia para el Chord Diagram, jerarquía de publishers para el Sunburst y el Treemap, conteos por categoría para el Pie Chart y frecuencias de palabras de descripciones para la Word Cloud o conteo de idiomas disponibles por año para el Bar chart race. Estos datasets adaptados se guardan como pequeños ficheros JSON o TXT, uno por visualización, para posteriormente usarlos en cada visualización.

A la hora de presentar se usa una página **HTML principal desde la que se puede acceder a cada una de las gráficas**. En éstas cada bloque de script importa su JSON o TXT y, mediante **D3.js v7**, invoca el layout apropiado (treemap, chord, pie, word-cloud...).

Este planteamiento modularizado de los recursos de D3.js junto con los archivos de estilo y los scripts de carga permiten expandir el visor de una forma escalable y controlada. Cada carpeta de gráfico de D3.js puede ser incorporada en el proyecto simplemente haciendo un fork al repositorio donde está disponible, sustituir el archivo html por el genérico que se ha preparado que incluye los estilos y la barra de navegación y por último añadir los archivos que actualmente se encuentran en la raíz del sistema pero podrían ser sustituidos por: enlaces a los [archivos](#) colgados en [github](#), enlaces a un repositorio de artifactory si el archivo es más grande o una petición dinámica a una API de consulta que permita añadir algún filtro.

Resultados

Pie Chart - Distribución global de etiquetas de géneros

Este gráfico permite las categorías más comunes de los juegos y la cantidad de jugadores que tienen. En este caso se han elegido las 15 categorías más comunes.

Conclusión

El diagrama circular (Figura 1) confirma una **gran concentración**: formando alrededor del 12% de los datos. Además, si tomamos las 5 categorías más comunes ya vemos que forman más del 50% de las 15 categorías más comunes de Steam.

Este desequilibrio sugiere que, aunque publicar sea barato, los creadores siguen tendiendo hacia temas probados, reforzando la competición directa y el “ruido” de descubrimiento.

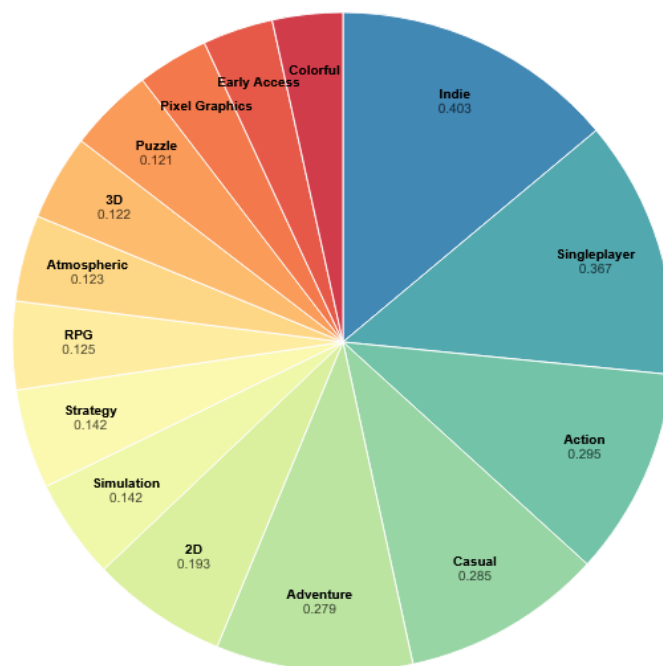


Figura 1 – Pie Chart. Géneros más populares

Chord Diagram - Co-ocurrencia de géneros

El **chord diagram** (Figura 2) visualiza 5 géneros más comunes y traza relaciones entre ellos para entender la coincidencia de estos géneros en los distintos juegos. Dicho de otra manera describe las veces en las que los géneros aparecen a la vez.. El ancho de cada cuerda es proporcional al número de juegos que comparten ambos géneros.

Conclusión

Los géneros están altamente interconectados. Las probabilidades de que los géneros coincidan son casi equiprobables rondando entre el 0.6% y el 0.8% de las veces. Esto destaca que los géneros más comunes suelen siempre ir de la mano. Esto podría llevar a pensar que es así para todos los casos. Sin embargo, existen más de 200 categorías de juegos y muchas no tienen esta cantidad de interconexión. Este gráfico, idealmente, podría ser filtrado de forma dinámica para elegir distintas categorías y poder comparar las interconexiones entre géneros de distintas categorías. Esto podría llevar a la conclusión de que los géneros más comunes son los que más se interconectan entre sí y que los menos comunes no tienen esta interconexión.

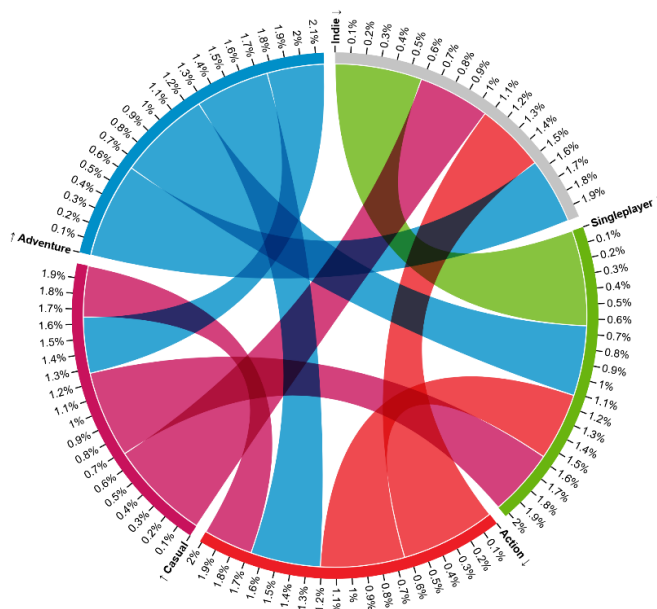


Figura 2 – Chord Diagram. Relación entre los generos más populares

Sunburst - Jerarquía editorial

Este gráfico permite ver la proporción de jugadores por desarrolladora y juego, pero no permite ver la proporción de jugadores por desarrolladora y el número de juegos que tiene cada desarrolladora.

La **sunburst** (Figura 3) muestra tres niveles: publisher → sub-estudio → juego. El zoom revela:

1. **Valve** posee pocos segmentos exteriores pero todos desmesuradamente gruesos (*CS:GO*, *Dota 2*, *Team Fortress 2*).
2. **Ubisoft** aparece atomizada (Montreal, Massive, Reflections, etc.), estrategia de producción distribuida que da como resultado un catálogo extenso pero con liderazgo difuso.
3. **Embracer Group** –a través de THQ Nordic, Deep Silver, Gearbox. Destaca por cantidad de lanzamientos, no por tráfico; es un ejemplo de conglomerado “paraguas”.

Conclusión

Cuatro desarrolladoras hacen el 50% de las 20 desarrolladoras con más jugadores en sus juegos. Valve es la principal, con más de un 25% de los jugadores. Es muy probable que esto se deba a que es la propietaria de la plataforma STEAM y pueda publicitar aquí sus juegos.



Figura 3. Sunburst. Mayores publishers

Treemap - Dependencia de títulos estrella

Este gráfico permite visualizar mejor la distribución de jugadores por desarrolladora y juego, ya que el gráfico de sunburst no permite ver la proporción de jugadores por desarrolladora y juego.

Conclusión

CSGO contiene un porcentaje 12% de los jugadores de la plataforma. Otras desarrolladoras que contienen una gran porción de los jugadores están divididas en distintos juegos excepto PUBG, Terraria, Rust, Among Us y algunos otros juegos cuyas desarrolladoras no tienen más que un juego con grandes números de jugadores como es el caso de Re-Logic mantiene relevancia con **un único juego (Terraria)**, recordatorio de que un estudio minúsculo puede desafiar a los gigantes si genera suficiente boca-oreja.

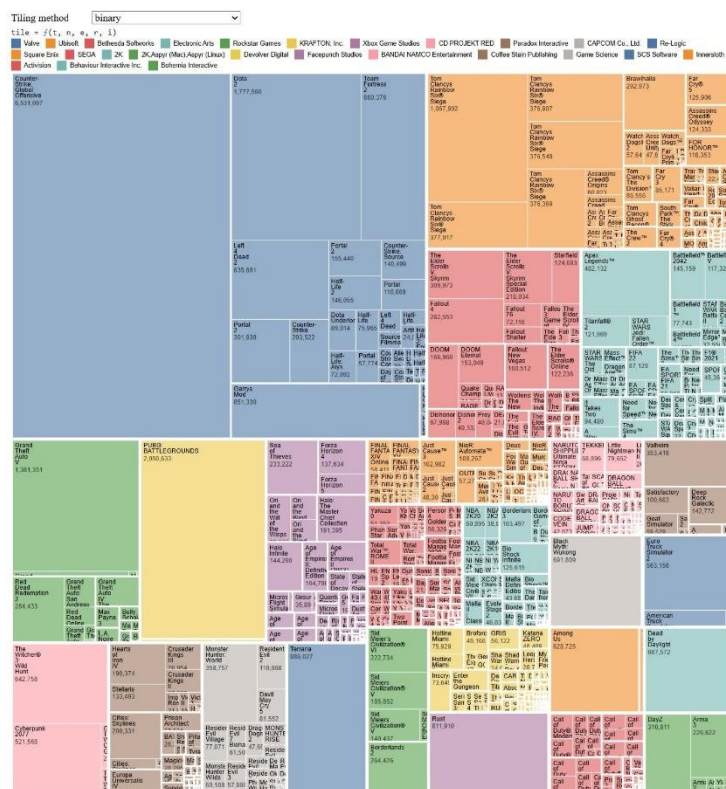


Figura 4. Treemap. Mayores publishers

Wordcloud – Descripciones más comunes

Este gráfico visualiza las palabras más comunes de las descripciones de un set de juegos seleccionado. Estos juegos son aquellos de las 10 publicadoras con más jugadores que se han analizado en el zoomable sunburst.

Conclusión

Se encuentra una diversidad de palabras bastante uniforme entre las palabras más comunes de las descripciones de los juegos con “game”, “new”, “world” y “experience” como palabras más frecuentes entre los juegos de las 20 grandes publishers (Wordcloud).



Figura 5. Wordcloud. Palabras más communes en descripciones

Race Chart Bar – Evolución de idiomas disponibles por año

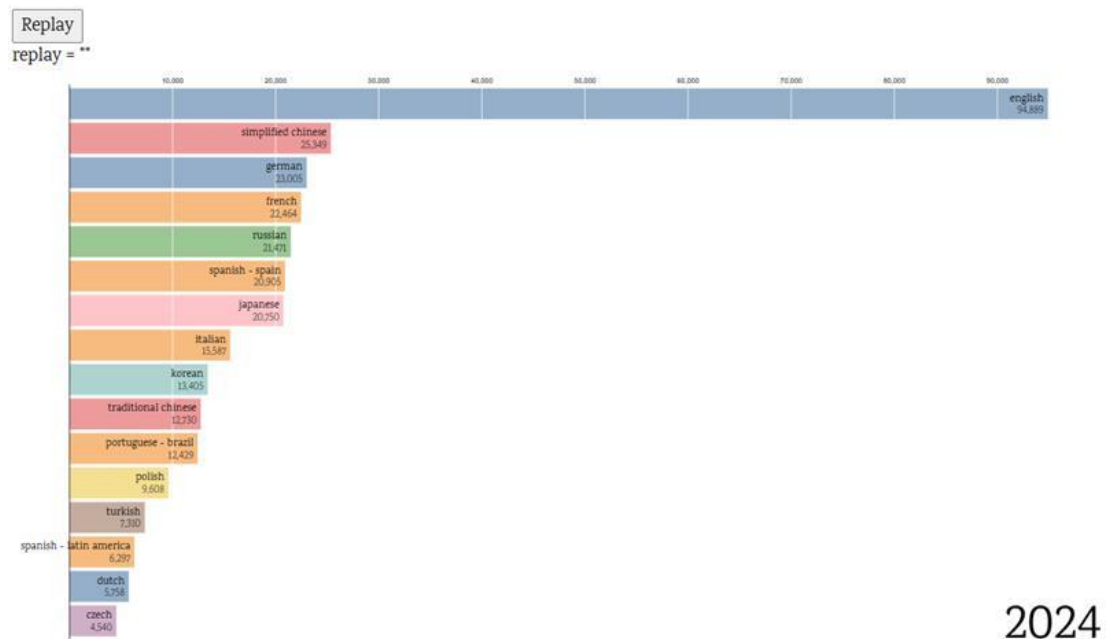


Figura 5. Bar Chart Race. Animación de lenguajes más comunes a través de los años 1997-2025.

Este gráfico muestra los idiomas más comunes de los videojuegos presentes en Steam desde el año 1997 al 2025 y que al ver en la página HTML muestra la evolución en forma de animación.

Conclusión

El inglés domina con diferencia como idioma principal en Steam, superando ampliamente al resto de lenguas presentes. Aun así, destacan idiomas como chino simplificado, alemán, francés o español, que muestran una fuerte presencia en los últimos años. Llama la atención la fragmentación del español entre su variante de España y la de Latinoamérica, o la del chino simplificado y tradicional que nos permite tener más detalle de la localización de los usuarios. La diversidad de idiomas disponibles sobre todo a partir de inicios de los 2000 refleja una tendencia a un catálogo globalizado y una industria que prioriza cada vez más la adaptación cultural del contenido.

Conclusiones

El estudio ha demostrado que, con un diseño adecuado de visualizaciones y un tratamiento del dato, es posible transformar un dataset complejo como el de Steam en una narrativa visual coherente e interactiva. El uso de D3.js ha sido clave: sus layouts jerárquicos, matriciales y semánticos han permitido extraer de los datos crudos diferentes niveles de información, desde la estructura editorial hasta la recurrencia léxica, y presentarlos de forma intuitiva para el usuario final.

Entre los principales hallazgos destacan:

- **Concentración editorial:** un número reducido de desarrolladoras de videojuegos domina el sector en cuanto a cantidad de jugadores. Valve sobresale por volumen y estabilidad de usuarios, en parte debido a su control de la propia plataforma.
- **Sinergias entre temáticas:** La fuerte presencia conjunta de Acción, Adventura, Casual y Indie en los datos indica que muchos títulos combinan estos géneros como una forma de diseño común y una estrategia para atraer a más jugadores sin perder su estilo propio.
- **Dependencia de Superventas en la Plataforma:** El Treemap muestra que unos pocos juegos muy populares, como CS:GO, PUBG o Terraria, concentran la mayoría de los jugadores. Esto sugiere que el éxito está muy concentrado en unos cuantos títulos, lo que genera dudas sobre cómo se mantiene el interés en el resto del catálogo.
- **Descripciones más de juego exitosos:** La nube de palabras generada a partir de las descripciones de los juegos más populares muestra una repetición constante de términos como “game”, “world”, “experience” o “multiplayer”. Esto indica que muchos títulos usan un lenguaje simple, centrado en la inmersión, la amplitud del entorno y el juego en línea.
- **El inglés reina en la plataforma:** El idioma inglés sigue siendo el idioma universal. Prácticamente todos los juegos de Steam están disponibles en inglés y se mantiene año tras año en el idioma principal de los juegos de la plataforma. Aunque el chino ha adelantado al resto de los primeros 6 idiomas

en los últimos 5 años lo cual parecer indicar el status quo de la próxima década.

Finalmente, se ha evidenciado que el uso combinado de Python para la transformación de datos y D3.js para su visualización permite construir herramientas analíticas altamente expresivas sin necesidad de frameworks más complejos. La arquitectura propuesta usando D3.js usando D3.js resulta replicable en otros contextos donde se requiera análisis exploratorio con orientación a negocio o UX visual.

Referencias

1. Goh, E., Al-Tabbaa, O., & Khan, Z. (2023). *Unravelling the complexity of the Video Game Industry*. Telematics & Informatics Reports, 12, 100100.
2. Newzoo (2024). *Global Games Market Report*.
3. Martin Bustos Roman. (2022). Steam Games Dataset [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DS/2109585>
4. Battiato, S. & Weber, J. (2021). *Visual Mapping of Game Genres using t-SNE*. IEEE VIS Workshop.