

Exploración y visualización de datos en Educación Matemática con Jupyter y Python

José Miguel Contreras García
Universidad de Granada, jmcontreras@ugr.es

Cristina Hidalgo Castillo
Universidad de Granada, chidcas@correo.ugr.es

Resumen: Este artículo explora el uso de Python a través de Jupyter Notebook como herramienta educativa para la visualización y análisis de datos en el ámbito de la enseñanza de matemáticas. A través de ejemplos prácticos, se muestra cómo Jupyter permite crear gráficos variados (barras, líneas, sectores, histogramas, entre otros) usando bibliotecas de Python como Matplotlib y Seaborn, facilitando el análisis exploratorio de datos. Dirigido a profesores, el artículo destaca la interactividad de Jupyter para que los estudiantes puedan experimentar y desarrollar habilidades en programación y visualización. Así, Jupyter se presenta como una plataforma accesible y eficaz para promover la alfabetización de datos en educación obligatoria.

Palabras clave: Visualización de datos, Jupyter Notebook, Python, educación matemática, alfabetización en datos, aprendizaje interactivo.

Data Exploration and Visualization in Mathematics Education with Jupyter and Python

Abstract: This study examines the use of Python through Jupyter Notebook as an educational tool for data visualization and analysis in mathematics education. Through practical examples, it illustrates how Jupyter enables the creation of diverse graphical representations (bar, line, pie charts, histograms, among others) using Python's libraries such as Matplotlib and Seaborn, thereby facilitating exploratory data analysis. Aimed at educators, the article focuses on Jupyter's interactive features, which allow students to experiment and develop skills in programming and visualization. Consequently, Jupyter is positioned as an accessible and effective platform to foster data literacy in compulsory education.

Key words: Data visualization, Jupyter Notebook, Python, mathematics education, data literacy, interactive learning.

1. INTRODUCCIÓN

La visualización de datos es una habilidad fundamental en la alfabetización de datos. Para comprender un conjunto de datos, es crucial explorar y conocer en profundidad las variables que lo componen. Aunque no se espera que un estudiante sea un experto en los datos que maneja, siempre será necesario investigarlos a fondo antes de realizar cualquier otro análisis. La visualización de datos es la herramienta más importante para lograr esto, por lo que es una habilidad esencial que debe dominarse en la educación obligatoria.

Jupyter con Python es una herramienta poderosa para la visualización y análisis de datos que combina un entorno interactivo de desarrollo con el lenguaje de programación Python.

Jupyter Notebook es un software de código abierto que permite a los usuarios crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Esta herramienta es ampliamente utilizada en educación, ciencia de datos e investigación, y es gratuita para todos los usuarios. Desarrollada como parte del Proyecto Jupyter, esta plataforma es compatible con numerosos lenguajes de programación (aunque Python es el más común) y proporciona un entorno intuitivo para la exploración de datos, el desarrollo de modelos y la documentación de procesos de análisis.

Jupyter Notebook es una herramienta poderosa para el análisis de datos interactivo, que permite a los usuarios crear gráficos y visualizaciones dinámicas que facilitan la comprensión de patrones y tendencias en los datos. Su interfaz, accesible desde un entorno web, es fácil de usar y altamente compatible con bibliotecas populares de Python como Matplotlib, Pandas, y Seaborn, lo que la hace ideal tanto para principiantes como para profesionales en ciencia de datos. Además, Jupyter se integra fácilmente con otras herramientas y plataformas de desarrollo, ofreciendo un ecosistema robusto para el análisis de datos y la creación de contenido interactivo.

En el ámbito educativo, Jupyter resulta particularmente valioso para los profesores (Kluyver et al., 2016), ya que permite la preparación de materiales didácticos interactivos y la inclusión de ejemplos prácticos que los estudiantes pueden explorar y modificar de forma autónoma, todo dentro del mismo entorno. Su diseño, basado en celdas de código y texto, permite a los educadores combinar explicaciones teóricas con ejercicios prácticos, promoviendo un enfoque de aprendizaje activo. Esto no solo enriquece el proceso de enseñanza, sino que también brinda a los estudiantes una experiencia práctica en programación, análisis de datos y visualización, fomentando su capacidad para experimentar y aprender de manera independiente y personalizada.

En el contexto de la visualización de datos, Jupyter se integra de manera eficiente con bibliotecas como Matplotlib, Plotly y Seaborn, permitiendo a los usuarios generar gráficos y visualizaciones directamente en la aplicación. Estos gráficos pueden variar desde representaciones simples, como histogramas o diagramas de dispersión, hasta visualizaciones más complejas e interactivas en 3D, facilitando una exploración visual más profunda. Estas capacidades de visualización son esenciales para el análisis exploratorio de datos (EDA, por sus siglas en inglés), ya que ayudan a identificar patrones, relaciones y anomalías dentro de los conjuntos de datos.

Siguiendo el enfoque del documento de referencia de la primera sección, titulado “Explorando la visualización de datos” (Molina-Portillo et al., 2024), este trabajo tiene como objetivo ofrecer a los profesores de matemáticas de las etapas de Secundaria y Bachillerato una guía práctica sobre el uso de Python mediante Jupyter Notebook para explorar datos mediante su visualización. Se presenta una selección de ejemplos que ilustran cómo adaptar la complejidad de las visualizaciones y el análisis a diferentes niveles de demanda cognitiva y habilidades digitales, destacando las posibilidades de esta herramienta para enriquecer la enseñanza de matemáticas y el análisis visual de datos.

2. ANTECEDENTES

El uso de Jupyter Notebook en educación ha sido ampliamente documentado, destacando su capacidad para facilitar un aprendizaje interactivo, a través de sus notebooks, y reproducible en diversas disciplinas, especialmente en matemáticas y ciencias de la computación.

Por ejemplo, Barba et al. (2019) describen el uso de Jupyter Notebook como una herramienta educativa que puede mejorar el aprendizaje en diversas disciplinas, incluyendo matemáticas, debido a su naturaleza interactiva. Los autores destacan cómo los notebooks permiten a los estudiantes y profesores combinar explicaciones teóricas con código ejecutable y visualizaciones, creando un entorno de aprendizaje dinámico y exploratorio, particularmente útil en la educación secundaria y superior.

Granger y Pérez (2021) también exploran el papel de Jupyter en la educación, subrayando cómo su capacidad de integrar código y texto ha sido beneficiosa para la enseñanza de materias como Matemáticas y Ciencia de Datos. Argumentan que el enfoque interactivo de Jupyter permite a los estudiantes entender mejor los conceptos teóricos a través de la experimentación práctica, facilitando un entorno de aprendizaje activo y personalizado.

En un estudio más específico, Gajdoš et al. (2022) analizan el uso de Jupyter Notebook en combinación con SageMath para la enseñanza de contenidos matemáticos. Este trabajo demuestra la efectividad de los notebooks como herramienta educativa al permitir a los estudiantes explorar y visualizar conceptos matemáticos complejos de manera interactiva, promoviendo así una comprensión más profunda.

Desde una perspectiva de ciencia de datos, Perkel (2018) en *Nature* resalta que Jupyter Notebook ha emergido como la herramienta preferida por los científicos de datos debido a su capacidad para soportar la computación interactiva. Este artículo destaca que los notebooks no solo facilitan la exploración de datos, sino que también permiten la comunicación clara de resultados a través de una combinación de texto explicativo y código ejecutable. Esto también se aplica al ámbito educativo, en el que los notebooks se utilizan para enseñar conceptos matemáticos y de programación de manera visual e interactiva.

La arquitectura de Jupyter, que se centra en la interactividad y la facilidad de uso, también ha sido objeto de atención en la literatura. Ragan-Kelley et al. (2014) discuten cómo el diseño de Jupyter permite un enfoque unificado en la investigación computacional, desde la exploración interactiva hasta la comunicación y publicación de resultados. Esta capacidad ha sido fundamental para su adopción en entornos educativos, en los que se valora la capacidad de los estudiantes para explorar y manipular conceptos matemáticos y científicos de manera práctica.

Rule et al. (2018) exploran cómo los notebooks computacionales, incluidos los Jupyter Notebook, facilitan tanto la exploración como la explicación en el aprendizaje. Destacan su utilidad en la enseñanza de programación y matemáticas al integrar código, visualizaciones y narrativas que permiten una comprensión más profunda y dinámica de conceptos complejos.

Finalmente, en Thompson y Mills (2020) se utiliza Python a través de Jupyter Notebooks en un aula de secundaria para la exploración de datos. En el estudio se muestra que el uso de Python puede ser útil a la hora de visualizar y analizar datos reales. Sin embargo, concluye que en este nivel educativo se debe utilizar de manera guiada ya que los estudiantes pueden llegar a sentirse confusos dado que se trata de un lenguaje de programación al que no suelen estar acostumbrados.

Estos estudios proporcionan una muestra sobre los beneficios de usar Jupyter Notebook en la educación, especialmente en matemáticas y programación. Su capacidad para integrar contenido diverso y su enfoque en la interactividad y reproducibilidad hacen de Jupyter una herramienta poderosa para educadores y estudiantes, promoviendo un aprendizaje más profundo y significativo.

Para este estudio tomamos en consideración la fundamentación teórica descrita en Molina-Portillo et al. (2024)

3. METODOLOGÍA

Para ilustrar el uso del lenguaje de programación Python en la visualización y análisis de datos, este estudio adopta un enfoque exploratorio-descriptivo centrado en el análisis del uso de Tecnologías de la Información y la Comunicación (TIC) entre menores de 10 a 15 años (INE, 2024). Basándonos en los datos presentados en el primer estudio del monográfico, que ofrece una descripción detallada de las fuentes, selección de casos, variables y resultados documentados, este apartado se enfoca en explicar los métodos de análisis y visualización empleados.

El trabajo se lleva a cabo en Jupyter Notebook, una plataforma que facilita la combinación de código, gráficos y documentación en un único entorno. En el notebook, se incluyen inicialmente instrucciones sobre cómo cargar el conjunto de datos, junto con una breve descripción de cada variable relevante. Los datos analizados proceden de la Encuesta sobre Equipamiento y Uso de TIC (2023), y contienen información sobre el acceso y uso de TIC en el hogar por menores, según el número de niños en la familia y su distribución geográfica en España, desglosada por comunidades y ciudades autónomas.

Para representar visualmente la información, se han seleccionado diversos tipos de gráficos que permiten explorar y describir los datos desde diferentes ángulos y niveles de complejidad.

Se muestran ejemplos de los siguientes tipos de gráficos:

- Gráfico de barras: ideal para mostrar comparaciones entre comunidades autónomas y el acceso a TIC según la cantidad de menores en el hogar.
- Gráfico de líneas: útil para observar tendencias en el tiempo o variaciones por regiones en el uso de diferentes tipos de TIC.
- Gráfico de sectores (o de pastel): se emplea para representar la proporción de hogares con acceso a diferentes tecnologías, facilitando una vista de la distribución general.
- Histograma: ayuda a comprender la frecuencia de uso de TIC en diferentes franjas de edad.
- Gráfico de dispersión: utilizado para analizar relaciones entre variables, como el uso de TIC frente al número de menores por hogar o por comunidad.
- Diagrama de cajas y bigotes: proporciona una visión detallada de la variabilidad y distribución de los datos de acceso a TIC por región.

Cada visualización se acompaña de una explicación detallada sobre el uso de bibliotecas de Python específicas, como Matplotlib, Pandas y Seaborn, adaptadas al tipo de gráfico en cuestión. Estas explicaciones guían al lector en el proceso de creación de cada gráfico, ajustando los parámetros y configuraciones para obtener representaciones visuales claras y precisas.

Para fomentar el aprendizaje activo, cada sección incluye una tarea práctica que invita al lector a aplicar los conceptos aprendidos y realizar ajustes adicionales a las visualizaciones. A continuación de cada tarea, se presenta una solución detallada que facilita la comprensión del proceso y refuerza las habilidades prácticas en el uso de Jupyter Notebook y sus bibliotecas para la visualización de datos.

Dada la limitación del texto, toda la descripción de los ejemplos se realiza en el ejemplo virtual, describiendo solo una parte de uno de ellos en este trabajo.

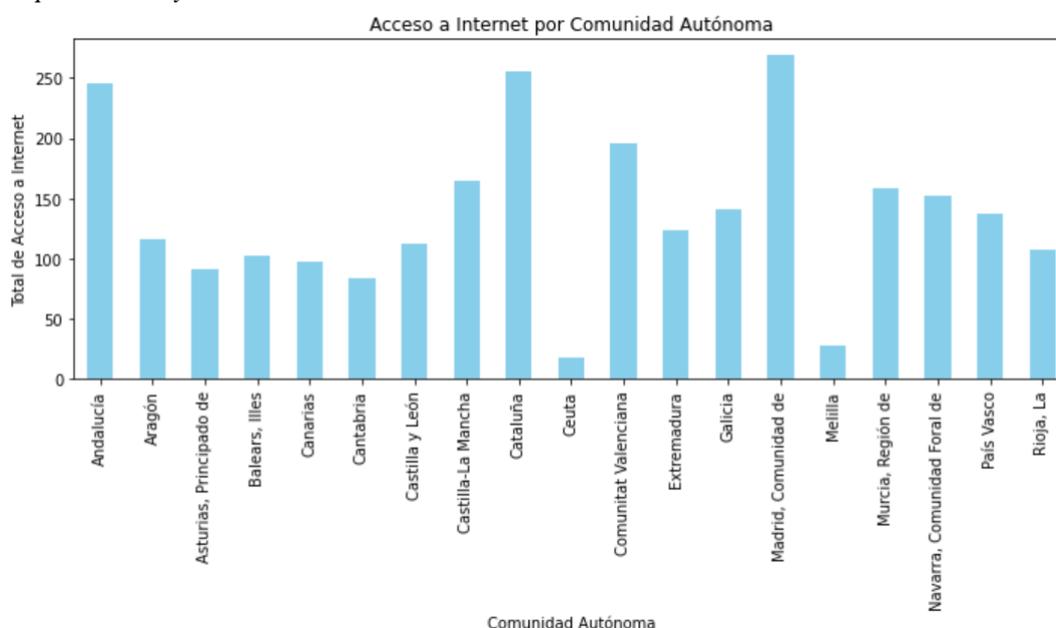
4. RESULTADOS

En los resultados se utiliza Jupyter Notebook para la visualización de los distintos ejemplos diseñados para mostrar la potencialidad de éste, que se pueden encontrar en el enlace <https://github.com/jmcontrerasgarcia/EDUCACION-EN-VISUALIZACION-DE-DATOS>. Al utilizar un lenguaje de programación algo complejo, como es Python, este notebook se enfoca a partir de la etapa de Secundaria. Los gráficos que se utilizan son el gráfico de barras, gráfico de líneas, gráfico de sectores, histograma, gráfico de dispersión y diagrama de cajas y bigotes. Estos recogen los datos de la Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares para el año 2023. A continuación, mostramos algunos ejemplos que se pueden realizar usando este software. Dada la limitación de espacio, mostramos un solo ejemplo de gráfico (los demás ejemplos los puede encontrar en el anterior enlace).

En la Figura 1, se muestra un gráfico de barras creado con la biblioteca Matplotlib en Python, el cual representa el acceso a Internet por comunidad autónoma.

Figura 1

Gráfico de barras que representa el acceso a Internet por comunidad autónoma en la librería Matplotlib de Python

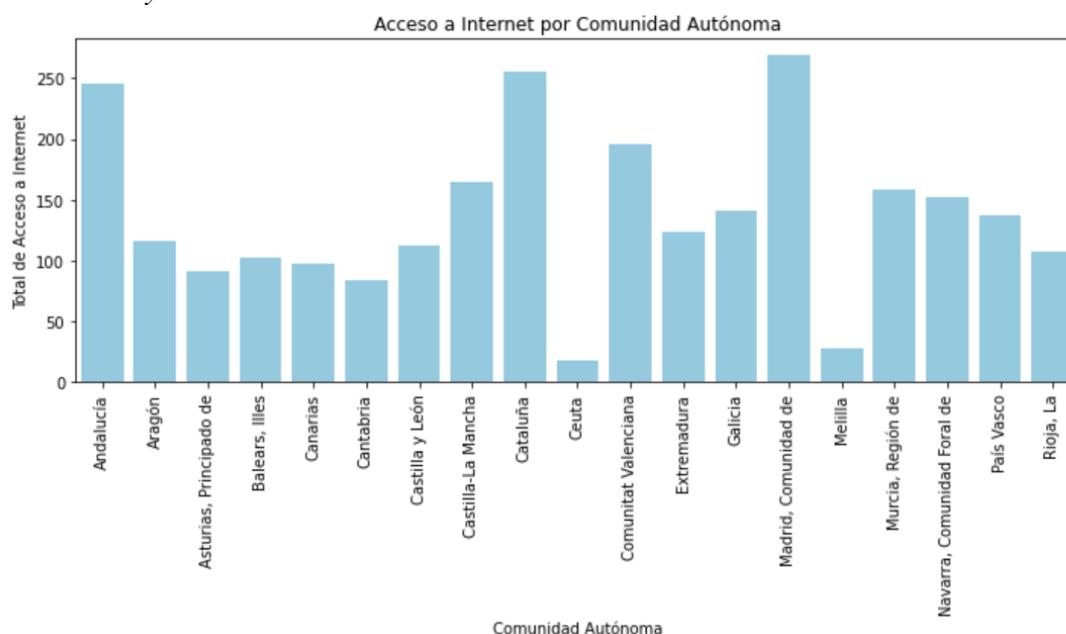


Fuente: elaboración propia

La Figura 2 presenta el mismo gráfico, pero utilizando la biblioteca Seaborn. Aunque ambas herramientas generan gráficos similares, Seaborn ofrece ciertas ventajas, como estilos predeterminados y una mayor simplicidad al gestionar datos. Mientras que Matplotlib requiere configuraciones manuales para ciertos efectos visuales, Seaborn facilita la creación de gráficos estéticamente agradables con menos líneas de código, gracias a sus paletas de colores predefinidas y su mejor integración con estructuras de datos de Pandas. Ambas bibliotecas son útiles, pero Seaborn puede resultar más conveniente para quienes buscan una presentación visual más atractiva con un código más conciso.

Figura 2

Gráfico de barras que representa el acceso a Internet por comunidad autónoma en la librería Seaborn de Python



Fuente: elaboración propia

En la Figura 3 se muestra un gráfico de barras creado con Seaborn, el cual destaca por la mejora en la estética visual y la legibilidad de los datos. A diferencia de las figuras anteriores, este gráfico utiliza una paleta de colores gradientes que facilita la distinción entre las diferentes comunidades autónomas. Además, se han añadido etiquetas numéricas en cada barra, lo que permite una lectura rápida de los valores sin necesidad de referirse al eje horizontal. Estas mejoras hacen que la información sea más accesible y fácil de interpretar, contribuyendo a una experiencia visual más agradable y clara.

En la Figura 4 se presentan diversas visualizaciones de datos que reflejan diferentes aspectos del uso de Internet y tecnologías de la información en España en menores de 10 a 15 años. El gráfico de sectores (arriba a la izquierda) muestra la distribución del acceso a Internet por niveles educativos, proporcionando una visión clara de cómo varía el acceso según la formación. A su lado, un diagrama de cajas y bigotes resalta las variaciones en el acceso a Internet por diferentes comunidades autónomas, mostrando la mediana y los rangos intercuartílicos de los datos. El gráfico de puntos (arriba a la derecha) permite visualizar el uso de Internet por provincias, ofreciendo una perspectiva clara de la relación entre las provincias y el acceso a la red. En la parte inferior, un histograma compara la distribución de edades en relación con el acceso a Internet, y un gráfico de barras agrupadas (abajo a la derecha) compara el uso de Internet por provincias, facilitando una rápida comparación entre ellas. Estos gráficos, demuestran la versatilidad de las herramientas de visualización de datos en Jupyter Notebook y su capacidad para presentar diferentes dimensiones de un conjunto de datos de manera clara y comprensible.

5. CONCLUSIONES

Jupyter Notebook se ha consolidado como una herramienta poderosa y versátil para la visualización y análisis de datos, particularmente relevante en los ámbitos de ciencia de datos y educación. Su entorno de desarrollo interactivo facilita la integración de código, texto explicativo, gráficos y resultados en un solo documento, lo cual mejora significativamente la experiencia de aprendizaje y la comunicación de hallazgos. Este trabajo muestra cómo Jupyter puede utilizarse para desarrollar visualizaciones que permiten una comprensión profunda de los datos, proporcionando al profesorado de matemáticas un recurso de gran valor para abordar la alfabetización de datos en distintos niveles educativos.

La compatibilidad de Jupyter con bibliotecas de visualización populares de Python, como Matplotlib y Seaborn, permite a los usuarios crear gráficos de alta calidad y personalizarlos para satisfacer distintas necesidades analíticas y pedagógicas. Matplotlib ofrece una flexibilidad completa en la configuración de gráficos, mientras que Seaborn facilita el diseño de visualizaciones más estéticas con menor cantidad de código, integrándose naturalmente con estructuras de datos de Pandas. Esta dualidad de opciones brinda a los educadores la libertad de elegir la biblioteca que mejor se adapte a su nivel de experiencia y a los objetivos específicos de su enseñanza.

El enfoque visual que ofrece Jupyter no solo facilita el análisis exploratorio de datos, sino que también permite presentar la información de manera clara y accesible. A lo largo del trabajo (especialmente en el enlace anexo), hemos mostrado cómo diferentes tipos de gráficos, como barras, líneas, sectores, histogramas, gráficos de dispersión y diagramas de cajas y bigotes, se pueden emplear para visualizar y comprender distintos aspectos de los datos sobre el uso de Tecnologías de la Información y la Comunicación (TIC) en menores. Estas visualizaciones, además de ser intuitivas, permiten al usuario identificar patrones, tendencias y posibles anomalías en los datos de forma rápida y efectiva.

Para los profesores de matemáticas, Jupyter Notebook representa una herramienta didáctica valiosa, ya que ofrece un entorno en el que los estudiantes pueden interactuar con datos reales, crear sus propias visualizaciones y explorar diferentes perspectivas de análisis de forma autónoma. Al poder manipular celdas de código en tiempo real, los estudiantes pueden experimentar y ver los efectos de cada cambio de manera inmediata, lo cual fomenta el pensamiento crítico y el aprendizaje activo. Asimismo, el diseño en celdas de Jupyter facilita la organización del contenido en secciones claras, combinando explicaciones teóricas con ejercicios prácticos y fomentando así una comprensión integral del análisis de datos.

Este trabajo también pone de manifiesto la importancia de la alfabetización de datos en la educación obligatoria. En un mundo cada vez más impulsado por los datos, la capacidad de explorar y representar visualmente la información es una competencia esencial que debe desarrollarse desde edades tempranas. Herramientas como Jupyter no solo acercan a los estudiantes al lenguaje de programación Python, sino que también les proporcionan un entorno amigable para la práctica de habilidades fundamentales en el análisis de datos, alineado con los diferentes niveles de demanda cognitiva y habilidades digitales. Como limitación, Python puede ser confuso si se utiliza sin tener conocimientos previos o si no se utiliza de manera guiada.

En conclusión, Jupyter Notebook es más que una simple herramienta de programación; es una plataforma integral para la educación en ciencia de datos, que permite a los docentes integrar visualización y análisis de datos en sus lecciones de manera accesible y efectiva. Esta capacidad de combinar teoría y práctica en un solo entorno contribuye a una enseñanza más

dinámica y adaptada a las necesidades de los estudiantes actuales, preparándolos para un futuro en el que la capacidad de analizar y comunicar datos será cada vez más relevante.

REFERENCIAS BIBLIOGRÁFICAS

- Barba, L. A., Barker, L. J., Blank, D. S., Brown, J., Downey, A. B., George, T., Heagy, L. J., Mandli, K. T., Moore, J. K., Lippert, D., Niemeyer, K. E., Watkins, R. R., West, R. H., Wickes, E., Willing, C., y Zingale, M. (2019). *Teaching and learning with Jupyter*. Recuperado: <https://jupyter4edu.github.io/jupyter-edu-book>, 1-77.
- Gajdoš, A., Hanč, J., y Hančová, M. (2022). Interactive Jupyter Notebooks with SageMath in Number Theory, Algebra, Calculus, and Numerical Methods. En *2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA)* (pp. 178-183). IEEE.
- Granger, B., y Pérez, F. (2021). Jupyter: Thinking and Storytelling with Code and Data. *Computing in Science & Engineering*, 23(2), 7-14.
- Instituto Nacional de Estadística [INE] (2024). Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares, año 2023. <https://tinyurl.com/2bq2py8r>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., y Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. En *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87-90). IOS press.
- Molina-Portillo, E., Contreras, J.M., Molina-Muñoz, D. y Cabrera, R. (2024). Explorando la visualización de datos. *Epsilon*, 118, 9-20.
- Perkel, J. M. (2018). Why Jupyter is data scientists' computational notebook of choice. *Nature*, 563(7732), 145-146.
- Ragan-Kelley, M., Perez, F., Granger, B., Kluyver, T., Ivanov, P., Frederic, J., y Bussonnier, M. (2014). The Jupyter/IPython architecture: a unified view of computational research, from interactive exploration to communication and publication. En *Proceedings of the 13th Python in Science Conference (SciPy)*, 51-56.
- Rule, A., Tabard, A., y Hollan, J. D. (2018). Exploration and explanation in computational notebooks. En *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Thompson, J., Wu, S. P., y Mills, J. (2020). The use of computer programming in a secondary mathematics class. En *2020 ASEE Virtual Annual Conference Content Access*.