

## EL PAPEL DE LOS RECURSOS INFORMÁTICOS EN EL PROCESAMIENTO Y ANÁLISIS DE DATOS EN LA INVESTIGACIÓN CIENTÍFICA

### *The role of IT resources in data processing and data processing and analysis in scientific research*

GALI MONPUE<sup>a</sup>

Recibido: 12/04/24 · Aceptado: 12/05/24

**Cómo citar:** Monpue, G. (2024). El papel de los recursos informáticos en el procesamiento y análisis de datos en la investigación científica. *Ciencia, Ingenierías y Aplicaciones*, 7(1), 1–6. <https://doi.org/10.22206/cyap.2024.v7i1.3116>

#### Resumen

*Esta nota técnica explora tres recursos informáticos fundamentales para el procesamiento y análisis de datos en la investigación científica: R, Python y Tableau. Se discute cómo estas herramientas se utilizan en diferentes disciplinas científicas y en qué etapas de la investigación son más útiles. R es un lenguaje estadístico que ofrece una amplia gama de funciones para análisis estadístico y visualización de datos. Python, por otro lado, es un lenguaje de propósito general con bibliotecas especializadas en análisis de datos y aprendizaje automático. Finalmente, Tableau es una plataforma de visualización de datos que permite crear gráficos interactivos y paneles de control para comunicar los resultados de la investigación de manera efectiva.*

**Palabras clave:** Análisis de datos; investigación; herramientas computacionales; visualización.

#### Abstract

*This technical note explores three fundamental computational resources for data processing and analysis in scientific research: R, Python, and Tableau. It discusses how these tools are used across different scientific disciplines and at which stages of research they are most useful. R is a statistical language that provides a wide range of functions for statistical*

---

<sup>a</sup> METXI. Universidad Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo, República Dominicana.  
ORCID: 0000-0002-3337-8179, Correo-E: [gmonpue@intec.edu.do](mailto:gmonpue@intec.edu.do), [galim1680@gmail.com](mailto:galim1680@gmail.com)



*analysis and data visualization. Python, on the other hand, is a general-purpose language with libraries specialized in data analysis and machine learning. Finally, Tableau is a data visualization platform that allows creating interactive charts and dashboards to effectively communicate research findings.*

**Keywords:** Data analysis; research; computational tools; visualization.

## 1. Introducción

Como perfectamente explican Kelleher y sus colegas, en la era digital, el procesamiento y análisis de datos se han convertido en componentes esenciales de la investigación científica en prácticamente todos los campos (Kelleher, et al. 2020). La cantidad cada vez mayor de datos disponibles presenta desafíos significativos, pero también oportunidades para obtener información valiosa. En este ensayo, exploraremos tres recursos informáticos fundamentales para el procesamiento y análisis de datos en la investigación científica: R, Python y Tableau. Discutiremos cómo estas herramientas se utilizan en la práctica y su utilidad en diferentes etapas de la investigación (Kelleher, et al. 2020). Es importante subrayar que esta nota técnica se sustenta en la literatura ya existente o estudios previos realizados por investigadores expertos en el tema.

## 2. R: Un Lenguaje Estadístico para Análisis de Datos

Según Chambers y Hohman, R es un lenguaje de programación y un entorno de software específicamente diseñado para el análisis estadístico y la representación gráfica de datos (Chambers, 2008; Hohman, et al. 2018). Es una herramienta poderosa para la investigación científica debido a su amplia gama de paquetes y funciones estadísticas disponibles. R permite a los investigadores realizar análisis descriptivos, inferenciales y predictivos, así como generar visualizaciones avanzadas para explorar y comunicar resultados.

De igual modo comprenden Krzywinski y Peng que la utilidad de R se extiende a diversas disciplinas, desde las ciencias sociales hasta la biología y la economía (Krzywinski, et al. 2014; Peng, 2011). Por ejemplo, la utilización de R en biología ha avanzado significativamente en la

comprensión de los datos genéticos y los estudios de expresión génica. Se han desarrollado varios paquetes R para mejorar diferentes aspectos de la investigación biológica. Por ejemplo, el paquete MultiWgCNA permite analizar las redes de coexpresión génica asociadas a enfermedades en diferentes condiciones espacio-temporales, (Tommasini & Fogel, 2023). Además, el paquete MRPC ayuda a aprender los gráficos causales con mayor precisión, particularmente en los datos genómicos, al incorporar el principio de la aleatorización mendeliana, (Badsha et al., 2021). Además, la herramienta DNA Architect permite la visualización de los datos de interacción de la cromatina y los componentes de la red, lo que ofrece información valiosa sobre la expresión génica (Righelli et al., 2022). Además, el paquete FACtORs proporciona una metodología novedosa para los estudios de asociación funcional en la investigación de la transcriptómica, lo que ayuda a validar los perfiles de expresión génica en todos los estudios y reduce el error de tipo I en los análisis posteriores, (Ramirez et al., 2020).

Definitivamente, la flexibilidad y la capacidad de personalización de R lo convierten en una herramienta indispensable para la investigación científica.

### **3. Python: Un Lenguaje de Propósito General para Análisis de Datos**

Python es un lenguaje de programación de propósito general que se ha vuelto cada vez más popular en el ámbito científico debido a su versatilidad y facilidad de uso. A través de bibliotecas como NumPy, Pandas y Matplotlib, Python proporciona un conjunto completo de herramientas para el procesamiento, análisis y visualización de datos. Además, el crecimiento de bibliotecas de aprendizaje automático como TensorFlow y scikit-learn ha ampliado aún más las capacidades de Python en la investigación científica (VanderPlas, 2016).

En este sentido, para VanderPlas y Wilke, la utilidad de Python en la investigación científica es evidente en una amplia variedad de aplicaciones (VanderPlas, 2016; Wilke, 2019). Desde la bioinformática hasta la física, los investigadores utilizan Python para manipular datos experimentales, realizar análisis estadísticos y desarrollar modelos predictivos. Desde la bioinformática hasta la física, los investigadores utilizan Python

para manipular datos experimentales, realizar análisis estadísticos y desarrollar modelos predictivos (McKinney, 2010). Su sintaxis clara y legible, junto con una amplia comunidad de desarrolladores, hacen de Python una opción atractiva para aquellos que desean aprovechar el poder de la informática en su investigación.

#### **4. Tableau: Visualización Interactiva para la Comunicación de Resultados**

Según Wickham, Tableau es una plataforma de visualización de datos que permite a los usuarios crear gráficos interactivos, paneles de control y tableros dinámicos a partir de conjuntos de datos complejos (Wickham & Wickham, 2016a; Wickham & Wickham, 2016b). Aunque no es un lenguaje de programación en sí mismo, Tableau es una herramienta invaluable para la investigación científica al facilitar la comunicación efectiva de resultados,.

Los investigadores pueden utilizar Tableau para crear visualizaciones interactivas que resalten patrones y tendencias en los datos. Estas visualizaciones pueden ser compartidas fácilmente con colegas, tomadores de decisiones y el público en general, lo que facilita la comprensión y la difusión de los hallazgos de investigación. Además, Tableau ofrece capacidades de análisis avanzadas, como el filtrado dinámico y la segmentación de datos, que permiten a los investigadores profundizar en sus datos y descubrir ideas nuevas (Wickham & Grolemund, 2017).

#### **5. Conclusión**

En conclusión, los recursos informáticos desempeñan un papel fundamental en el procesamiento y análisis de datos en la investigación científica. R y Python proporcionan herramientas poderosas para realizar análisis estadísticos y modelado de datos, mientras que Tableau facilita la comunicación efectiva de resultados a través de visualizaciones interactivas. Al aprovechar estas herramientas, los investigadores pueden explorar datos, descubrir patrones y comunicar hallazgos de manera más efectiva, lo que contribuye al avance del conocimiento en una amplia variedad de disciplinas científicas.

## Referencias

- Badsha, M. B., Martin, E. A., & Fu, A. Q. (2021). MRPC: An R Package for Inference of Causal Graphs. *Frontiers in Genetics, 12*, 651812. <https://doi.org/10.3389/fgene.2021.651812>
- Chambers, J. M. (2008). *Software for data analysis: programming with R*. New York: Springer.
- Hohman, F., Kahng, M., Pienta, R., & Chau, D. H. (2018). Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE transactions on visualization and computer graphics, 25*(8), 2674-2693.
- Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press.
- Krzywinski, M., & Altman, N. (2014). Visualizing samples with box plots. *Nature methods, 11*(2), 119-120.
- McKinney, W. (2010, June). Data structures for statistical computing in Python. *SciPy, 445*(1), 51-56.
- Peng, R. D. (2011). Reproducible research in computational science. *Science, 334*(6060), 1226-1227.
- Ramirez, R. N., Bedirian, K., Gray, S. M., & Diallo, A. (2020). DNA Rchitect: An R based visualizer for network analysis of chromatin interaction data. *Bioinformatics, 36*(2), 644–646. <https://doi.org/10.1093/bioinformatics/btz608>.
- Righelli, D., Weber, L. M., Crowell, H. L., Pardo, B., Collado-Torres, L., Ghazanfar, S., Lun, A. T. L., Hicks, S. C., & Risso, D. (2022). SpatialExperiment: Infrastructure for spatially-resolved transcriptomics data in R using Bioconductor. *Bioinformatics, 38*(11), 3128–3131. <https://doi.org/10.1093/bioinformatics/btac299>.
- Tommasini, D., & Fogel, B. L. (2023). multiWGCNA: An R package for deep mining gene co-expression networks in multi-trait expression data. *BMC Bioinformatics, 24*(1), 115. <https://doi.org/10.1186/s12859-023-05233-z>
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media, Inc.

- Wickham, H., & Wickham, H. (2016a). *Data analysis*. Springer International Publishing.
- Wickham, H., & Wickham, H. (2016b). *Getting Started with ggplot2. ggplot2: Elegant graphics for data analysis*. Springer International Publishing.
- Wickham, H., & Grolemund, G. (2017). *R for data science: Import. Tidy, transform, visualize, and model data*. O'Reilly Media.
- Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.