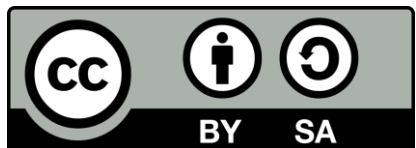




# Web scraping con R: Análisis bibliométrico de la tesis de pregrado en psicología en Perú

---

Jorge Huanca-Martinez  
*18 de noviembre de 2023*



# Análisis bibliométrico

---

La investigación psicooncológica en el Perú: un estudio bibliométrico de artículos publicados en revistas científicas peruanas (2006-2016)

Revistas	Artículos totales revisados	Artículos con contenidos sobre psicooncología: frecuencias y porcentajes, respectivamente.
Revista de Psicología (RP-PUCP)	148	0 (0%)
Revista Libe-rabit	208	0 (0%)
Revista de Investigación en Psicología (RIP)	290	2 (0.7%)
Revista de Psicología (RP-UCV)	112	0 (0%)
Revista de Psicología (RP-UCSP)	149	4 (2.68%)
Revista de Psicología (RP-UCSP)	45	0 (0%)
Revista de Psicología de Arequipa (RP-AQP)	65	2 (3.08%)
	<b>TOTAL 1017</b>	<b>8 (0.79%)</b>

Caycho-Rodríguez et al. (2017)

The scientific production of medical students in Lima, Peru

Table 3. Characteristics of papers published by students (N = 174).

Characteristics	N (%)
Language	
Spanish	128 (73.6)
English	45 (25.9)
Spanish and English	1 (0.6)
Has as first author a medical student	
No	84 (48.3)
Yes	90 (51.7)
Type of article	
Original	78 (44.8)
Letter to the editor	58 (33.3)
Case report	26 (14.9)
Other	12 (6.9)
Article design*	
Descriptive	22 (28.2)
Analytic	41 (52.6)
Experimental	10 (12.8)
Quasi-experimental	5 (6.4)

Urrunaga et al. (2020)

# Oportunidades del análisis bibliométrico

Líneas de investigación poco exploradas

Menos recursos (económicos y humanos)

Mayor control de la veracidad de los datos

Insumo para trabajos futuros

¿POR QUÉ ES  
UNA BUENA  
ALTERNATIVA?



# Tesis de pregrado

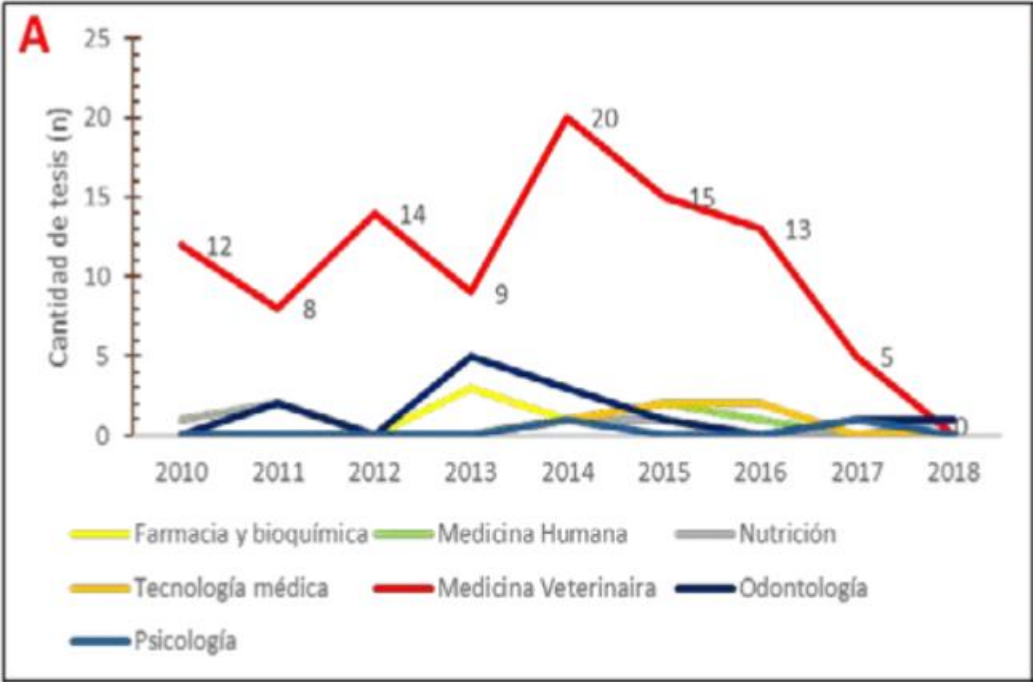
---

# Publicación científica y características de las tesis de salud de universidades del departamento de Puno, Perú

Variable	Categoría	Frecuencia	Porcentaje (%)
Carrera profesional	Enfermería	224	46,50
	Medicina	121	25,10
	Nutrición	89	18,50
	Psicología	48	10,00
Año de aprobación	2016	81	16,80
	2017	205	42,50
	2018	196	40,70
Universidad de procedencia	UNA	311	64,50
	UANCV	93	19,30
	UPeU	73	15,10
	UPSC	5	1,00
Tipo de población	Escolar	115	23,90
	Universitario	41	8,50
	Centro de salud	70	14,50
	Comunidad	26	5,40
	Hospitalario	174	36,10
	Penitenciario	6	1,20
	Otro	50	10,40

Mamani-Benito et al. (2020)

# Tesis sustentadas y publicadas por estudiantes de las ciencias de la salud en Perú



Castro-Rodriguez et al. (2020)

# ¿De dónde extraer la información?



## Universidad Peruana Cayetano Heredia


[Repositorio Institucional](#) → [Documentos para optar grados y títulos](#) → [Título Profesional](#)

Psicología

Listar por


- [Por fecha de publicación](#)
- [Autores](#)
- [Títulos](#)
- [Materias](#)

Búsqueda en esta colección:  



## renati

registro nacional de trabajos de investigación

 [Nave](#)

Buscar trabajos conducentes a grados y títulos

### Ranking de instituciones

Universidad César Vallejo	121273
Universidad Católica los Ángeles de Chimbote	25670
Pontificia Universidad Católica del Perú	21853

### Ranking de autores

Chenet Zuta, Manuel Enrique	7
Pesantes Arriola, Genaro Christian	7
Guerra García, Pedro Robinson	6
Subelete Auccacusi, Edith	6

### Ranking de asesores

La Rosa Feijoo, Oscar Calixto	713
Murriel Santolalla, Luis Alberto	626
Calvanapón Alva, Flor Alicia	578
Manrique Placido, Juana	558

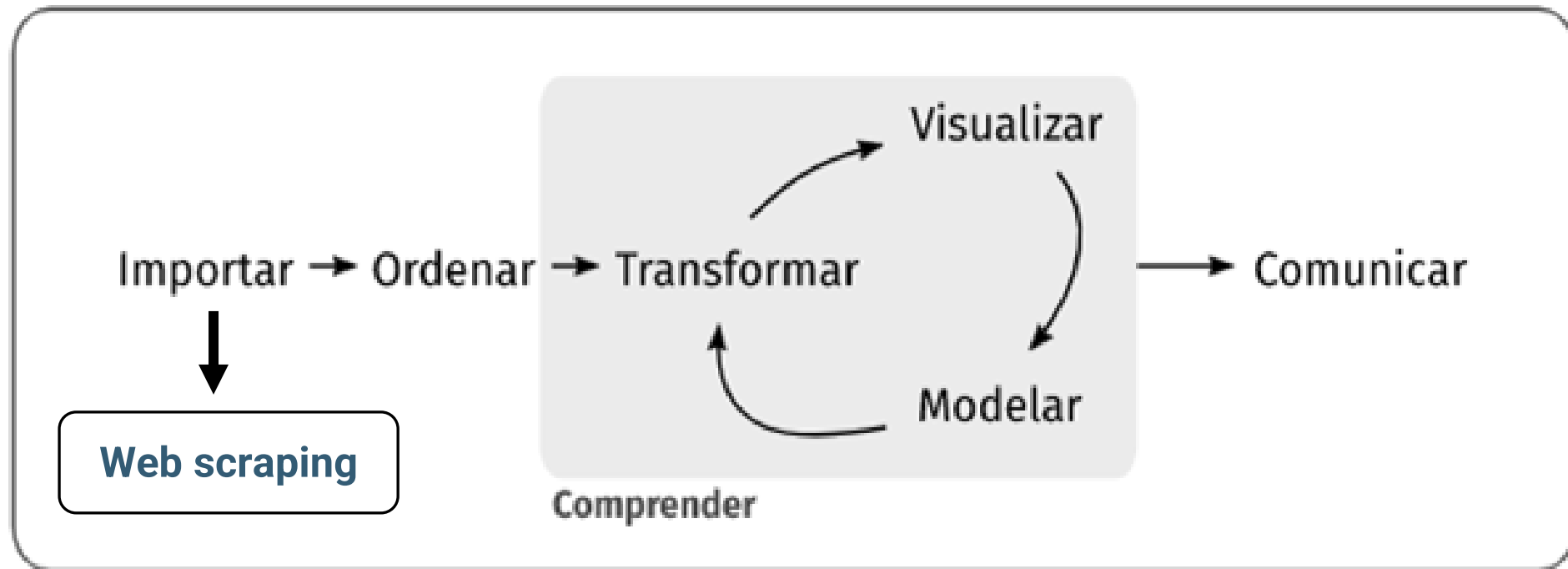
# Web Scraping

---



# ¿Cuál es el aporte del webscraping?

Flujo de análisis de datos:



Wickham et al. (2023)

# Webscraping como nuevo enfoque

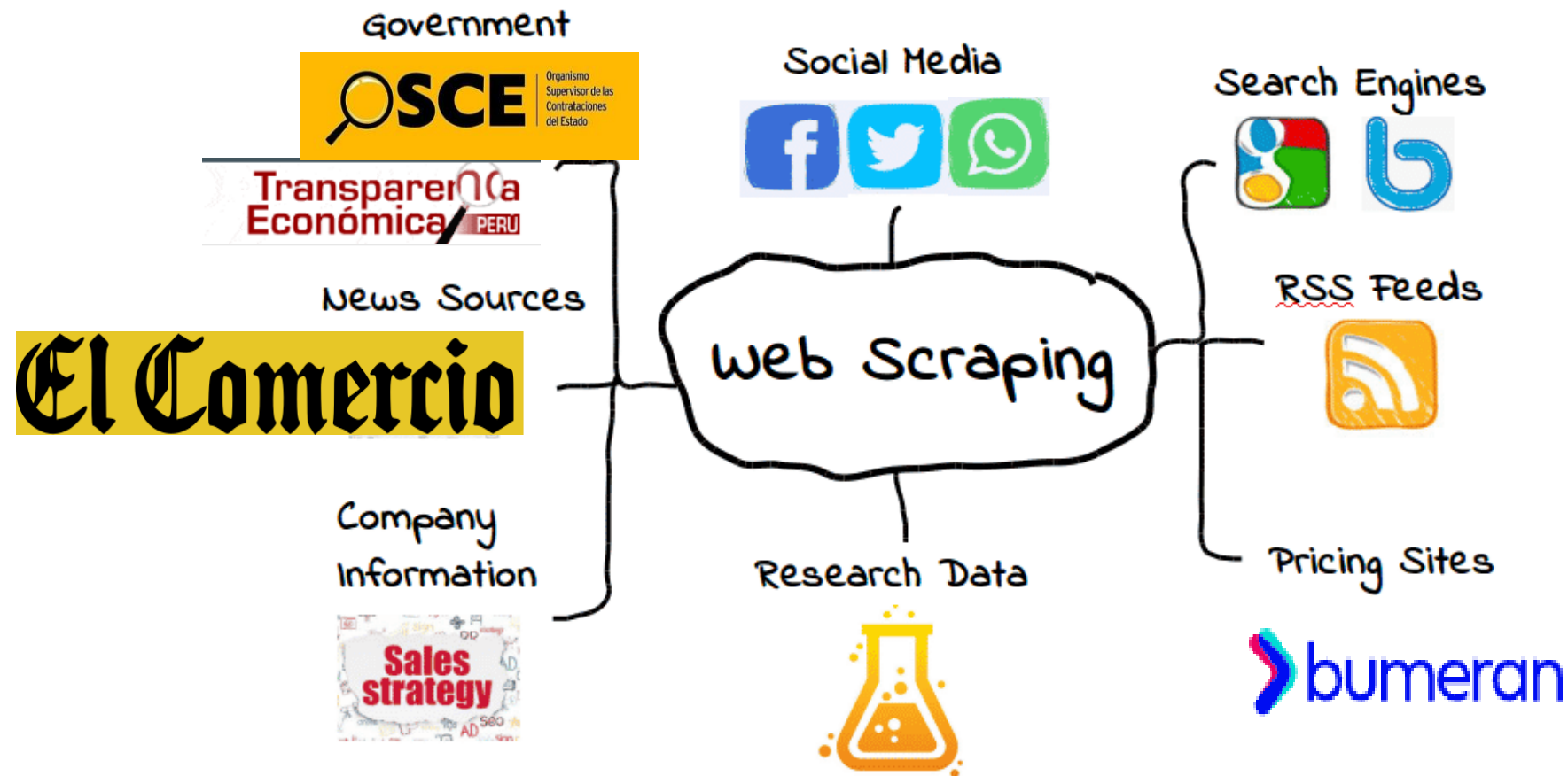
- Data estructurada
- Extracción manual
- Mayor tiempo de extracción
- Más propenso a errores
- Menos riesgos ante cambios de la estructura de la web

## Enfoque tradicional

- Data semi o no estructurada
- Uso de scrapers o API's
- Menor tiempo de extracción
- Menos propenso a errores
- Más riesgos ante cambios en la estructura web

## Enfoque datos web

# ¿Qué se puede scrapear?



# Limitaciones: Consideraciones legales y éticas

Consultar  
condiciones legales  
de la web

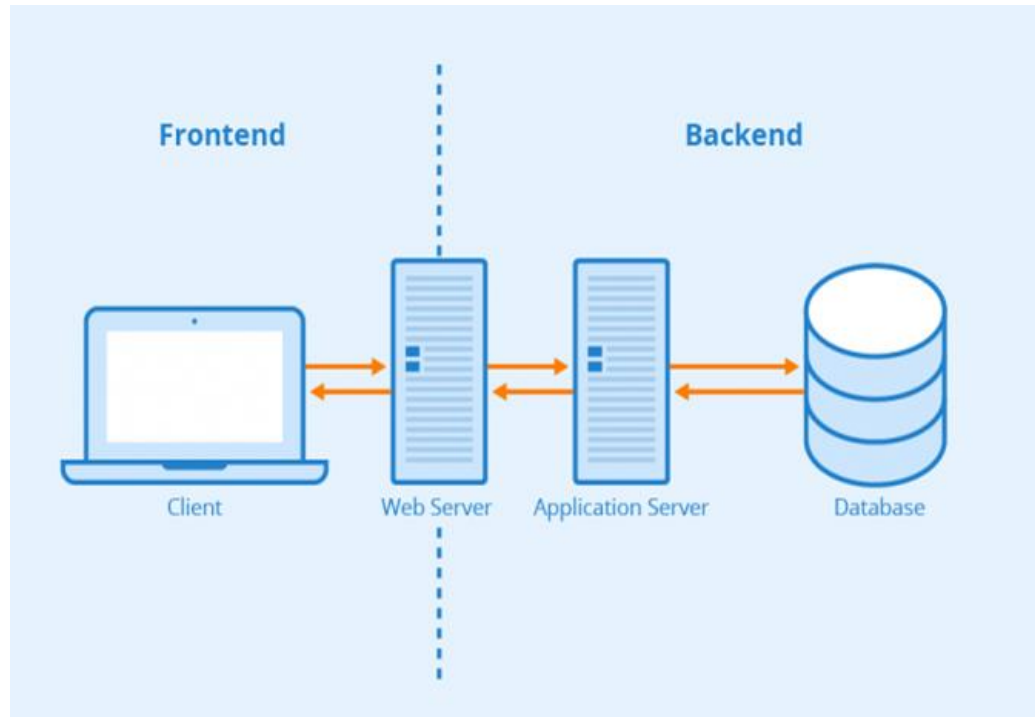
Sistemas de  
protección web

Dificultad para  
acceder a datos  
demográficos  
específicos

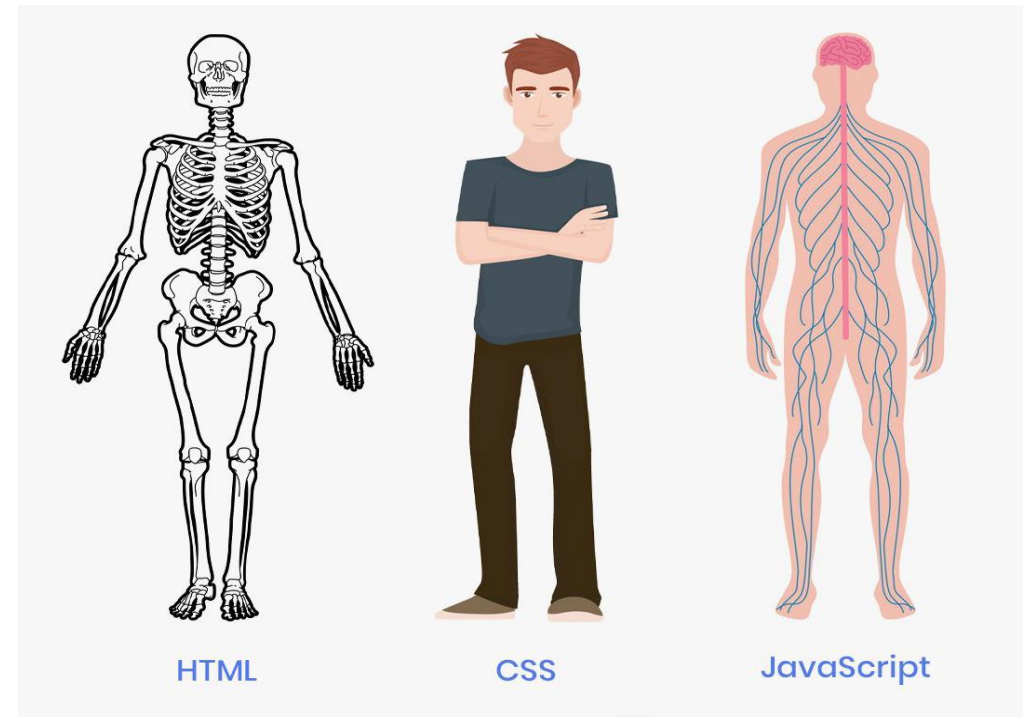


# Página web

## Componentes:



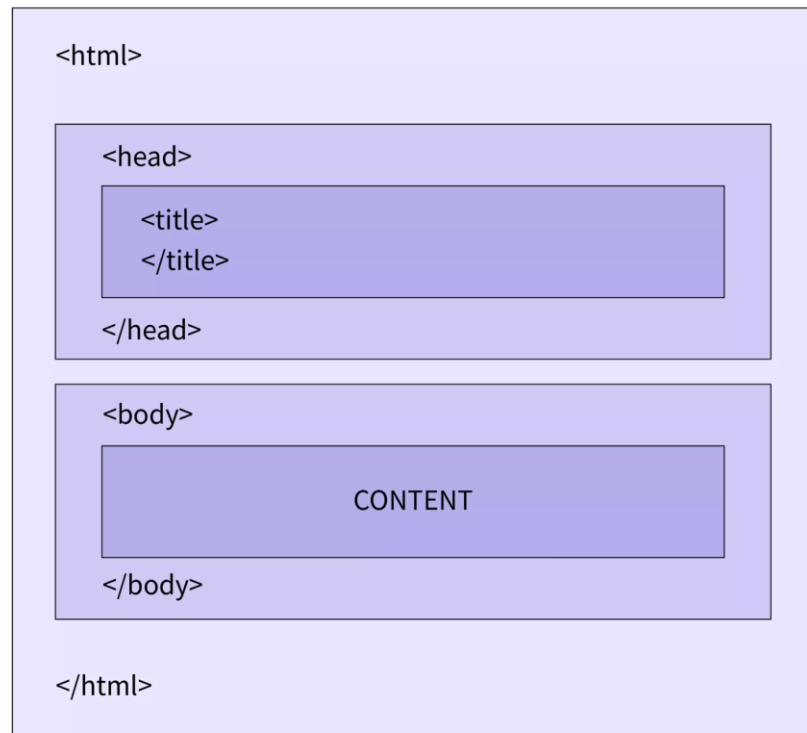
## Elementos frontend:



# HTML

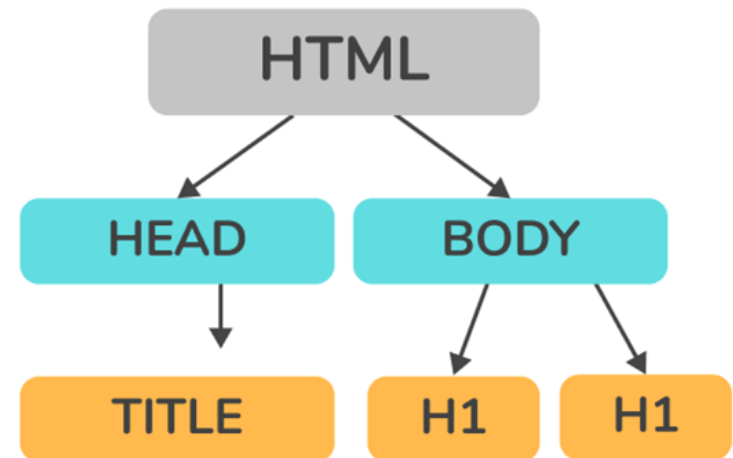
Describe los elementos de una página mediante etiquetas caracterizadas por paréntesis angulares (<>).

1. El documento siempre comienza y termina utilizando las etiquetas `<html>` y `</html>`.
2. El body constituye la parte visible del documento HTML.



## Tree Structure

```
<!DOCTYPE html>  
<html>  
  <head>  
    <title></title>  
  <head>  
  <body>  
    <h1></h1>  
    <h1></h1>  
  <body>  
</html>
```



# HTML Etiquetas

Tag	Description
<code>&lt;h1&gt;</code> to <code>&lt;h6&gt;</code>	Defines HTML headings
<code>&lt;ul&gt;</code>	Defines an unordered list
<code>&lt;ol&gt;</code>	Defines an ordered list
<code>&lt;p&gt;</code>	Defines a paragraph
<code>&lt;a&gt;</code>	It is termed as anchor tag and it creates a hyperlink or link.
<code>&lt;div&gt;</code>	It defines a division or section within HTML document.
<code>&lt;strong&gt;</code>	It is used to define important text.
<code>&lt;table&gt;</code>	It is used to present data in tabular form or to create a table within HTML document.
<code>&lt;td&gt;</code>	It is used to define cells of an HTML table which contains table data
<code>&lt;iframe&gt;</code>	Defines an inline frame

# Referencias

- Castro-Rodríguez, Y., Hinojosa-Añorga, M., Torres-Robles, G., Roca-Sacramento, C., & Rojas-Ortega, R. (2020). Tesis sustentadas y publicadas por estudiantes de las ciencias de la salud en Perú. *Edumecentro*, 12(1), 15-29.
- Caycho-Rodríguez, T., Ventura-León, J., Barboza-Palomino, M., Jara-Ávalos, S., & Gallegos, W (2017). La investigación psicooncológica en el Perú: un estudio bibliométrico de artículos publicados en revistas científicas peruanas (2006-2016).
- Mamani-Benito, O., Ramos Vilca, G., Rodrigo Ccaso, D., Quispe Mayta, F., & Chipana Paye, N. (2020). Publicación científica y características de las tesis de salud de universidades del departamento de Puno, Perú. *Horizonte Médico (Lima)*, 20(3).
- Urrunaga-Pastor, D., Alarcon-Ruiz, C. A., Heredia, P., Huapaya-Huertas, O., Toro-Huamanchumo, C., Acevedo-Villar, T., Arestegui-Sánchez, L., Taype-Rondán, Á., & Mayta-Tristán, P. (2020). The scientific production of medical students in Lima, Peru. *Heliyon*, 6(3), e03542.
- Wickham, H., Cetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc.



# ¡MUCHAS GRACIAS!



[jorgeahm98@gmail.com](mailto:jorgeahm98@gmail.com)



[www.linkedin.com/in/jorgehuanca](https://www.linkedin.com/in/jorgehuanca)



<https://github.com/jorgehmartinez>

---

Jorge Huanca-Martinez  
*18 de noviembre de 2023*