

**UNIVERSIDAD CATOLICA BOLIVIANA “SAN PABLO”**  
**SEDE SANTA CRUZ**  
**INGENIERIA DE SOFTWARE**



**INFORME FINAL**

**MATERIA:** Probabilidad y Estadística I  
**DOCENTE:** Ph.D. Sergio Daniel Moreira Ascarrunz  
**ESTUDIANTES:** Josué Balbontín  
Alejandro Ramírez Vallejos  
Fernando Terrazas  
**FECHA:** 28/11/2024

**Santa Cruz-Bolivia**

# INFORME FINAL

## Miembros

Nombre	Correo Electrónico	Número de Teléfono
Josue Balbontín	<a href="mailto:josue.balbontin@ucb.edu.bo">josue.balbontin@ucb.edu.bo</a>	75041970
Alejandro Ramírez	<a href="mailto:alejandro.ramirez.v@ucb.edu.bo">alejandro.ramirez.v@ucb.edu.bo</a>	76380296
Fernando Terrazas	<a href="mailto:fernando.terrazas@ucb.edu.bo">fernando.terrazas@ucb.edu.bo</a>	79943071

## Tema

Comparación y análisis de la longitud de las noticias periodísticas digitales en Bolivia, con análisis específico de las noticias publicadas los días domingo entre 2010 y 2023.

## Preguntas de interés

- ¿Se han adaptado los periódicos digitales a los malos hábitos de lectura de la generación actual en Bolivia?
- ¿Ha disminuido la longitud de una publicación periodística digital entre el 3 de enero de 2010 y el 31 de diciembre de 2023 en Bolivia?

## Investigaciones secundarias

Se tomará como referencia investigaciones que analicen los niveles de lectura en Bolivia y América Latina, para respaldar la pregunta de interés relacionada con los hábitos de lectura de las generaciones actuales. Entre las referencias se incluyen:

- Franco, M. P. (2020). *Desarrollo educativo y ciudadanía: universitarios y lectura en Santa Cruz de la Sierra, Bolivia*. Encrucijada Americana, 12(1), 39-53.
- Monak, Lenin (2013). *Comportamiento lector y hábitos de lectura: Una comparación de resultados en algunos países de América Latina*. CERLALC.

- Beatriz, Paula (2023). *Más de la mitad de los niños de la región no alcanza la suficiencia en capacidad lectora*. UNIFRANZ.

## Unidad de observación

Noticias periodísticas digitales de Bolivia publicadas en domingo entre los años 2010 y 2023.

## Variables

- **Variables cuantitativas:** cantidad de letras, cantidad de palabras, cantidad de oraciones, cantidad de párrafos.
- **Variables categóricas:** nombre del periódico emisor de la noticia, año de emisión del periódico digital.
- **Datos adicionales:** URL de cada periódico digital analizado, para registro de las fuentes investigadas.

## Rango de respuestas

Los rangos de respuesta para las variables cuantitativas y el nombre del emisor de la noticia variarán entre cada artículo. Se tiene certeza de que los valores de la variable “año” fluctuarán entre 2010 y 2023.

## Población

Noticias digitales publicadas los días domingo entre los años 2010 y 2023.

## Muestra

**Muestreo aleatorio simple:** en esta investigación, se seleccionarán de forma aleatoria domingos entre los años 2010 y 2023. Para cada año, de los 52 domingos disponibles, se elegirá aleatoriamente un 60% del total anual. Por cada domingo seleccionado, se analizará una noticia elegida aleatoriamente.

## Marco muestral

El marco muestral incluirá todas las noticias bolivianas publicadas en línea entre 2010 y 2023, accesibles a través de motores de búsqueda comunes, como Google.

## Tecnología

Se utilizará el lenguaje C++ para la extracción de variables de cada periódico digital y para organizar los datos en formato .csv, después de la corrección de errores. Posteriormente, se empleará el lenguaje R para la creación de gráficos y el análisis estadístico de los datos recolectados.

**Código C++ (GitHub):** <https://lc.cx/cCujs6>

**Base de Datos Depurada:** <https://n9.cl/pov0c>

**Base de Datos Sin Depurar:** <https://n9.cl/pxutm>

## Justificación del tipo de estudio

Nuestra pregunta de interés es: “¿Se han adaptado los periódicos digitales a los hábitos de lectura de la generación actual?”. Para llevar a cabo este análisis y obtener resultados representativos, es fundamental considerar la longitud de las noticias a lo largo de un período significativo.

Además, optamos por el muestreo aleatorio simple como técnica de muestreo. Ya que procederemos con la elección aleatoria de domingos identificando cada uno con el número correspondiente en el período anual. Luego obtendremos 1 noticia de cada domingo aleatoriamente. Esto nos garantizará una muestra sin sesgo y tendrá representatividad de la población.

## Precauciones para tomar en cuenta

- Si en un domingo seleccionado al azar no contiene 1 noticia disponibles para el análisis, se elegirá otro domingo aleatoriamente.
- Para el análisis estadístico solo consideraremos la sección del conjunto de información sobre la noticia, omitiendo así partes de las noticias como los

derechos de autor, publicidad, y todo lo que los usuarios no leerían con relación a la noticia.

### **Estadística descriptiva preliminar**

- Análisis de relación entre cantidad de letras, palabras, oraciones y párrafos, con el año individualmente (4 histogramas).
- Analizar la diferencia de cantidad de letras entre los periódicos más conocidos en la actualidad con el año (Datos agrupado-Cajas y bigotes).
- Verificar correlación entre años y el total de las letras de cada noticia (Grafico de puntos y grafico con series de tiempo).
- Analizar relación entre longitud de letras, palabras, oraciones, y párrafos (grafico de barras y grafico de puntos-contingencia).
- Mostrar aumento medio (mediana) de longitud de cada periodo de tiempo (Histograma).

### **Finalización del trabajo pendiente**

Para completar la recolección de datos, los 14 años de estudio se dividirán entre los tres integrantes del grupo: Josué Balbontín, Alejandro Ramírez y Fernando Terrazas. Cada uno será responsable de la depuración de datos de sus años asignados, con la recolección y depuración de datos programada entre el 1 y el 7 de noviembre, y la finalización del informe para el 30 de noviembre.

## Conclusiones Investigativas

En un mundo cada vez más digitalizado, donde la información se consume de manera instantánea y fragmentada, los medios de comunicación se enfrentan al desafío de adaptarse a los nuevos hábitos de lectura de la audiencia. Este estudio se propuso analizar cómo ha evolucionado la longitud de las noticias digitales en Bolivia entre los años 2010 y 2023, un período marcado por la consolidación de internet y la proliferación de dispositivos móviles.

Nuestro objetivo principal fue determinar si los medios de comunicación bolivianos han ajustado la extensión de sus contenidos a las preferencias de un público cada vez más exigente y menos capacitado, que busca información concisa y de fácil acceso. Para ello, se llevó a cabo un análisis cuantitativo de una muestra aleatoria de noticias publicadas los domingos en diversos periódicos digitales bolivianos.



A través del análisis de variables como la cantidad de letras, palabras, oraciones y párrafos, se buscó identificar tendencias en la evolución de la longitud de las noticias a lo largo del tiempo. Asimismo, se exploraron las posibles relaciones entre la longitud de los textos y factores como el tipo de noticia, el medio de comunicación y el año de publicación.

Los resultados obtenidos en esta investigación aportan valiosa información sobre la adaptación de los medios de comunicación bolivianos al nuevo entorno digital y tienen implicaciones importantes para comprender cómo se consume la información en el país.

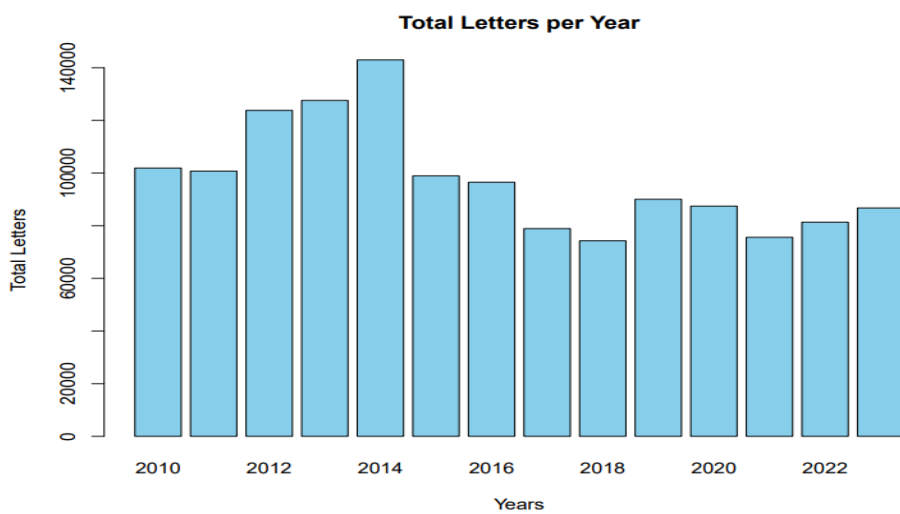
A continuación, presentaremos los principales hallazgos de nuestro estudio investigativo.

Al finalizar esta presentación, se espera que los asistentes tengan una visión más clara sobre la evolución de los formatos periodísticos digitales en Bolivia y las implicaciones de estos cambios para los periodistas, los medios de comunicación y el público en general.

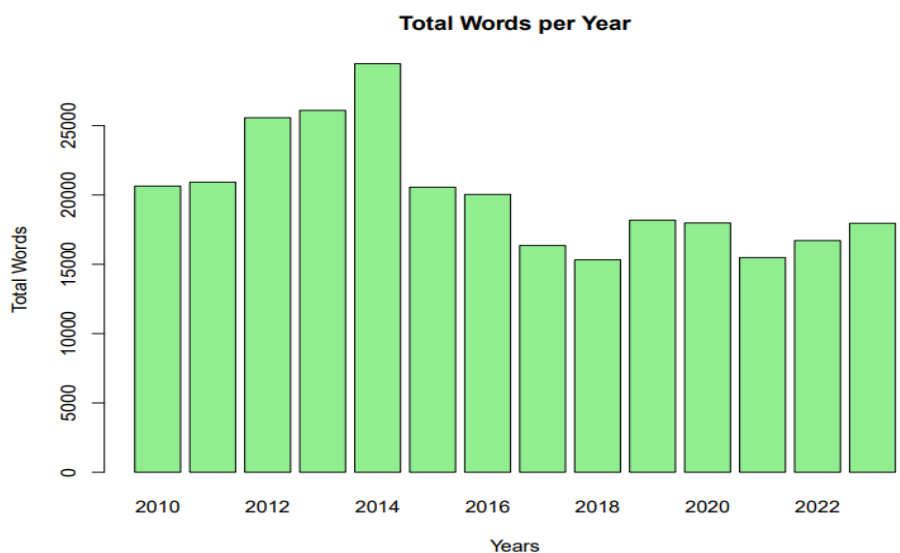
## **Presentación de datos Estadísticos e interpretaciones**

A continuación, ilustraremos gráficamente las conclusiones de nuestra investigación basada en la longitud de las noticias periodísticas bolivianas. Se informará y responderá a los objetivos de la investigación basada en datos recopilados entre el 60% de los domingos de cada periodo anual.

A continuación, se observa el histograma de letras por año. Este no presenta una curva normal, sino que presenta una distribución sesgada a la derecha. El año con mayor número de letras en las noticias digitales fue el 2014 con aproximadamente 140000. Vemos un incremento de letras en los años 2010 hasta 2014, posteriormente una caída abrupta en 2015 y ligeramente hasta 2018. Finalmente se estabilizo hasta 2023.

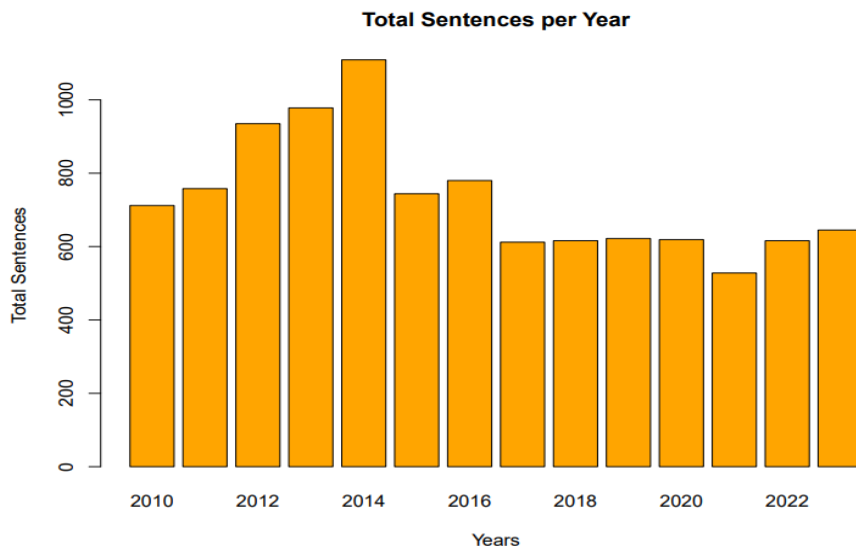


El histograma de palabras por año carece de curva normal, por lo que presenta una distribución sesgada a la derecha y muy semejante al histograma de letras por año. El año con mayor número de palabras en las noticias digitales fue el 2014 con aproximadamente 25000. Vemos un incremento de palabras en los años 2010 hasta 2014, posteriormente una caída en 2015 hasta las 20000 palabras y siguió ligeramente hasta 2018. Finalmente se estabilizo hasta 2023.

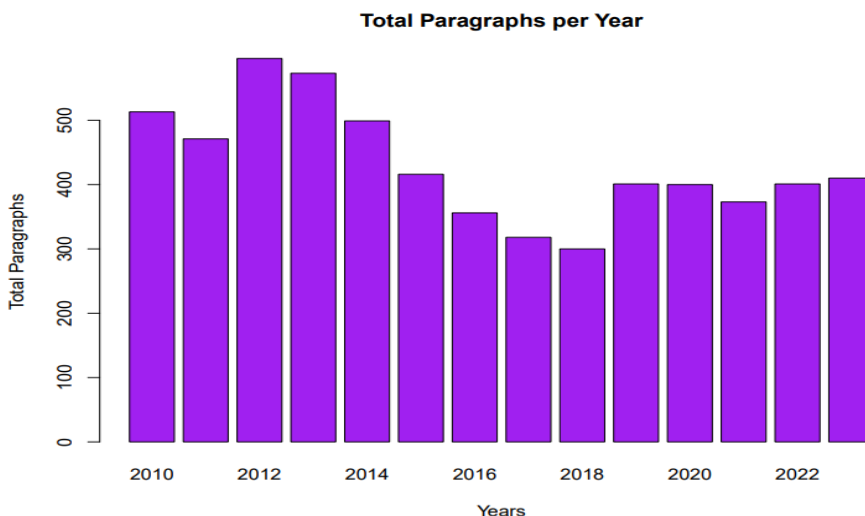




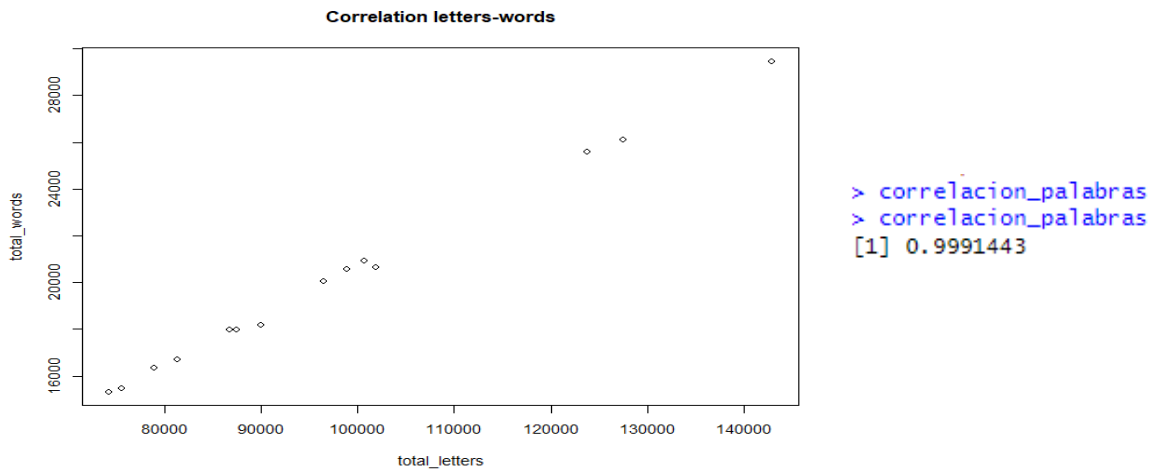
El histograma de oraciones por año no presenta una curva normal, sino que presenta una distribución sesgada a la derecha y semejante al histograma de letras y oraciones por año. El año con mayor número de oraciones en las noticias digitales fue el 2014 con aproximadamente 1100. Vemos un incremento de oraciones en los años 2010 hasta 2014, posteriormente una fuerte caída en 2015 y se continuo ligeramente hasta 2017. Posteriormente se estabilizo hasta 2023.



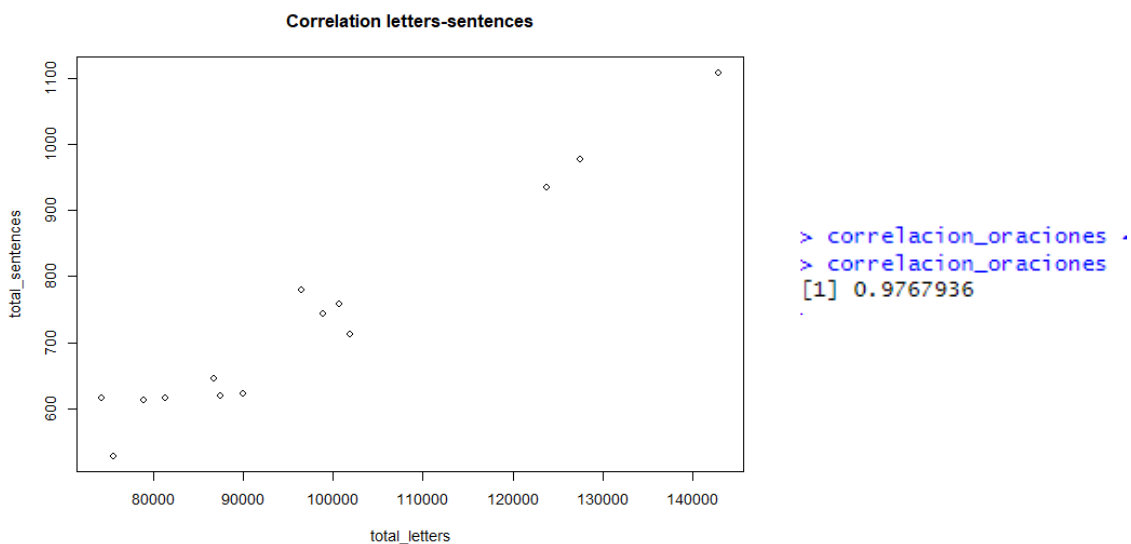
El histograma de párrafos por año no presenta una curva normal, sino que presenta una distribución sesgada a la derecha. El año con mayor número de párrafos en las noticias digitales fue el 2012 con aproximadamente 600. Vemos un incremento de párrafos en los años 2010 hasta 2012, posteriormente una caída progresiva desde 2013 a 2018. En 2019 aumento ligeramente el número de párrafos y posteriormente se estabilizo hasta 2023.



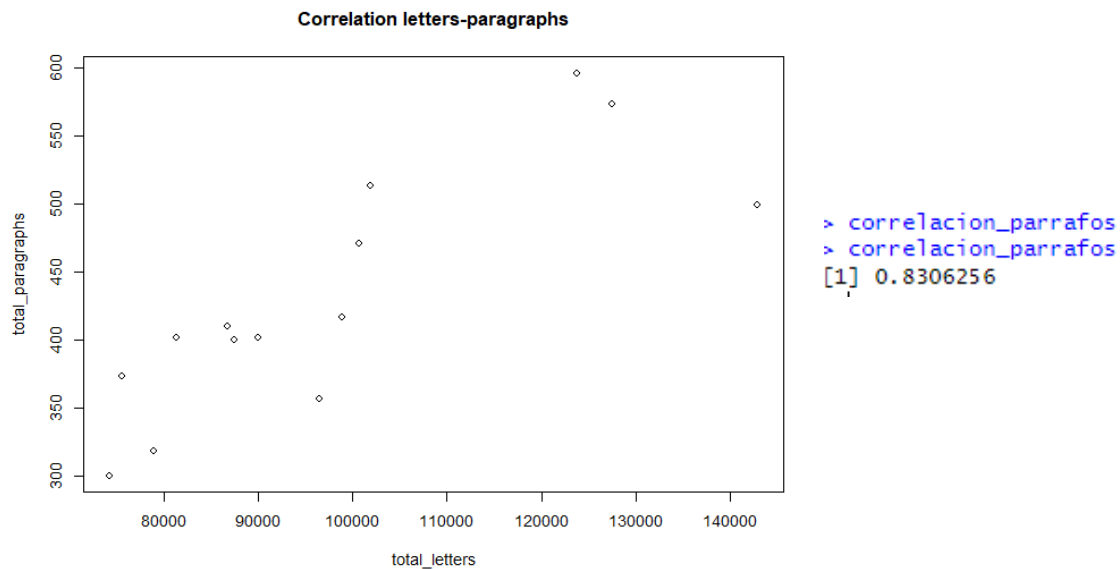
El presente diagrama de desviación nos muestra la relación entre la cantidad de letras y palabras. Podemos ver una correlación muy fuerte de 0.999 con dirección positiva, expresando que cada vez que el numero de palabras aumenta, entonces muy probablemente el número de letras también aumentara.



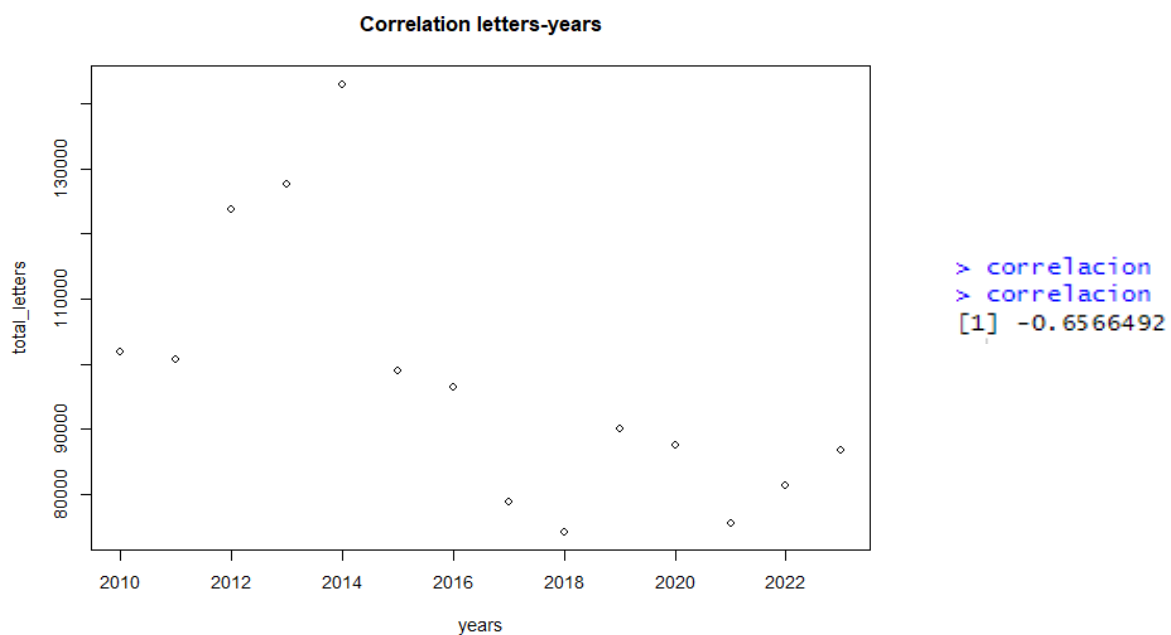
El grafico de desviación muestra una relación muy fuerte entre la cantidad de letras y oraciones de las noticias digitales, presentando un índice de correlación de 0.977 con dirección positiva. Lo que quiere decir que, si la cantidad de letras aumenta, entonces es muy probable que la cantidad de oraciones igual aumente.



El diagrama presenta una relación fuerte entre la cantidad de letras y párrafos en una noticia digital. Expresa una dirección positiva con un índice de correlación de 0.831. Por lo que es bastante probable que, si la cantidad de letras aumenta, entonces también lo haga los párrafos en las noticias.

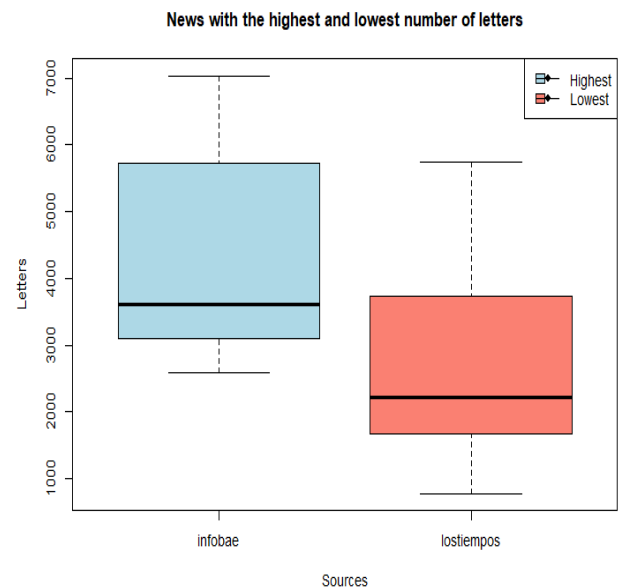
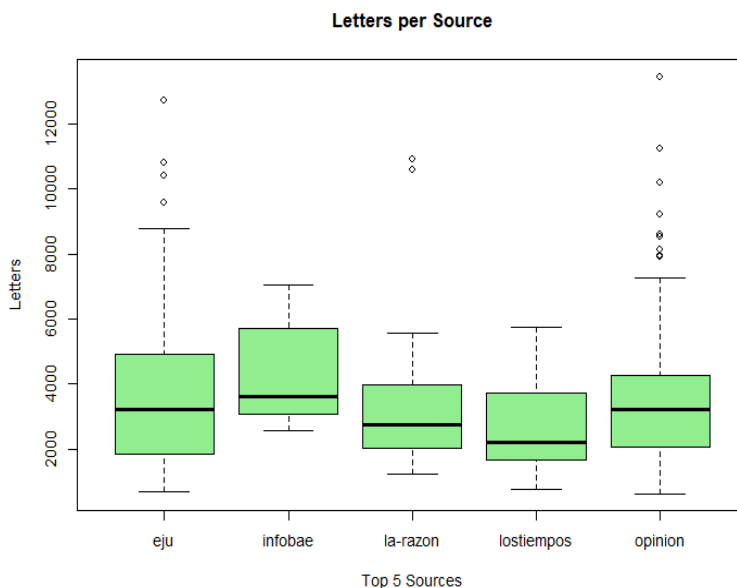


El diagrama entre letras y años nos muestra una relación moderada entre la cantidad de letras y los años de las noticias. Podemos observar una correlación de -0.657 de dirección negativa. Por lo que podríamos decir que la relación presentada es: mientras mayor sea el año de la noticia, menos letras tendrá la noticia.

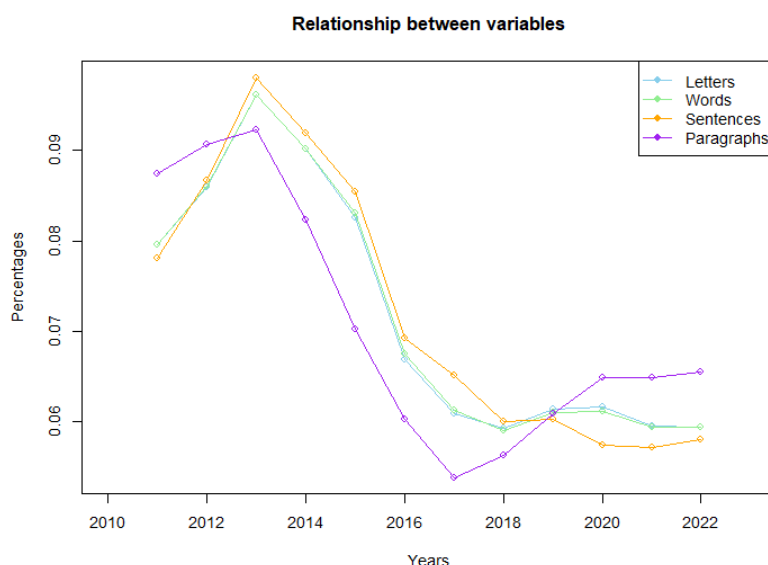


El gráfico de caja y bigotes nos muestra una comparación de la cantidad de letras generadas por cinco fuentes distintas (eju, infobae, la-razon, lostiempos y opinion). El grafico muestra que el 50% de las letras de cada noticia rondan los 2000 y 5000. El número de valores extremos se extiende más para las noticias deaju y opinión, además que sus amplios bigotes muestran alta volatilidad en sus valores sobre las demás noticias. Por otro lado, la mediana aproximada para la cantidad de letras de las 5 noticias rondaría las 3000.

Por otro lado, podemos ver que la noticia con mayor numero de letras es infobae, con aproximadamente 3800 letras por noticia en mediana y que el 50% de sus noticias rondan las 3200 y 5800 letras por noticia. Por otro lado, la noticia con menor numero de letras es lostiempo, con una mediana de 220 letras aproximadamente y que el 50% de sus noticias tienen entre 1700 a 3600 letras aproximadamente.



Finalmente, el presente gráfico muestra una comparación de cuatro variables relacionadas con la longitud de las noticias (letras, palabras, oraciones y párrafos) a lo largo de un rango de años (de 2010 a 2023), con el objetivo de analizar su comportamiento respecto a cada variable en el tiempo. Podemos observar una fuerte correlación entre las 4 variables cuantitativas. Observamos un aumento constante en el año 2013 y posteriormente una caída considerable hasta el 2017 donde se estabilizo hasta 2023.



En síntesis tras una recopilación de datos en base a la longitud de las noticias digitales entre los años 2010 a 2023, podemos finalmente resolver nuestras preguntas de interés de la investigación.

¿Ha disminuido la longitud de una publicación periodística digital entre el 3 de enero de 2010 y el 31 de diciembre de 2023 en Bolivia?

Efectivamente, encontramos que en el año 2010 el promedio de las noticias rondaba los 1959.788 letras. Así mismo, en 2014 se tenía un promedio de 2749.173 y finalmente en 2023 se obtiene un promedio de 1668.192 letras por noticia periodística digital.

Por lo tanto, se encuentra una caída en la longitud de las noticias digitales de Bolivia del 14.88% del 2010 al 2023 y una caída del 39.32% del 2014 al 2023, lo que concluye con una afirmación a una de nuestras preguntas de interés.

¿Se han adaptado los periódicos digitales a los malos hábitos de lectura de la generación actual en Bolivia?

Para responder esta pregunta, hay que tomar en cuenta los malos hábitos de lectura de los bolivianos. Según encuestas de la universidad Unifranz Santa Cruz, el 48% de las personas en Cochabamba, El Alto, La Paz y Santa Cruz admite que no ha leído un libro hace más de un año. Estos datos recopilados entre los años 2015, 2016, 2017 y 2018.

Tomando en cuenta la preocupante cifra de los lectores en territorio boliviano, ahora analicemos los datos los cuales obtuvimos de nuestra investigación.

Podemos observar en los datos presentados que hay una correlación muy fuerte entre las longitudes de las noticias (letras, palabras, oraciones y párrafos). Además de que existe un índice de correlación moderado de dirección negativo entre los años y las letras de cada noticia. Por esto podemos decir que existe mientras mayor sean los años, menor serán la cantidad de letras que tendrá una noticia.

Tomando en cuenta que los hábitos lectores de Bolivia son solo del 52% de la población. Podríamos determinar que efectivamente existe una relación entre los malos hábitos lectores de los bolivianos y la disminución de la longitud de las noticias periodísticas digitales.