

74.24 / 95.07 Teoría de la Programación



Grupo: 8

Lenguaje: R

|  |  |
| --- | --- |
| *Alumno* | *Padrón* |
| Juan Ignacio Colombo | 103471 |
| Ignacio Janeiro | 103550 |
| Federico Andres Ramos | 101640 |
| Martin Andres Suarez | 101540 |

1. Origen

R fue creado por Ross Ihaka y Robert Gentleman en la universidad de Auckland.R fue creado en el 1991 y fue anunciado en 1993. En 1995 usaron una licencia GNU convirtiendo R en software gratuito.Finalmente en el 2000 se lanzó la versión beta estable.El lenguaje se basó en una implementación de dos lenguajes existentes, S y Scheme.Debido a su similitud con S y el hecho de que sea compatible con cualquier sistema operativo facilitó su migración.El lenguaje surge debido a que sus creadores buscaban una forma de enseñar estadística a sus alumnos en la facultad.

2. Para qué sirve

R es un lenguaje de programación cuyo uso está mayormente orientado a las estadísticas, debido a la gran cantidad de herramientas que posee.R es un lenguaje que cuenta con más de 10.000 paquetes en su repositorio CRAN.Algunos paquetes como por ejemplo readr facilitan trabajar con grandes cantidades de información desordenada y así poder ordenar la misma y estructurarla para así poder entender y trabajar mejor con la misma.R también tiene una gran capacidad gráfica, el cual permite hacer gráficos de manera sencilla y entendibles a través de librerías como ggplot o plotly.En fin R es un lenguaje con gran capacidad para el análisis de datos y con gran utilidad para el análisis estadístico, no solo por su gran cantidad de paquetes que facilitan el manejo de datos, sino que también al ser un lenguaje de programación, permite crear sus propias funciones y tratar la información al gusto de uno.

En R los objetos son almacenados en la memoria física.Para manejar la información, el lenguaje requiere que se almacene toda la información en la memoria, es por esto que cuando se quiere trabajar con grandes cantidades, como por ejemplo Big Data, R deja de ser una buena opción, como lo es Python.Otro aspecto a tener en cuenta es que si se requiere trabajar con mucha información y se quiere poder procesarla rápidamente, R y sus paquetes no sería la mejor opción ya está, MATLAB o Python que son lenguajes muchos más rápidos a la hora de procesar información.

3. Características Básicas

3.1 Características principales

R es un lenguaje con licencia GNU, es decir que es libre, gratuito y abierto.En el caso de R en se trabaja con paquetes gratuitos, en vez de librerías como por ejemplo, pygame una librería de Python para crear juegos.R tiene la capacidad de crear gráficos, basado en Latex.También cuenta con una gran cantidad de herramientas estadísticas como por ejemplo modelos lineales y no lineales, tests estadísticos y algoritmos de clasificación y agrupamiento.R tiene la capacidad de ser integrado con distintas bases de datos.

3.2 Paradigmas

R es un lenguaje que al igual que Oz es multiparadigma, ya que soporta varios paradigmas, y el usuario puede utilizar el que le parezca más conveniente para resolver el problema.

R soporta los paradigmas:

1-Funcional

2-Vectorial

3-Imperativo

4-Procedural

5-POO

3.3 Tipado

R es de tipado dinámico, con la particularidad de que al declarar una variable numérica automáticamente le asigna un double.Ser un lenguaje de tipado dinámico implica que una variable puede tomar cualquier valor, ya sea un entero, un flotante, un carácter en distintos momentos.A diferencia de un tipado estático donde hay que declarar de qué tipo va a ser la variable y este tipo es inalterable.Algunos lenguajes de tipado dinámico como R son, Python, Javascript o Oz, mientras que algunos de tipado estático pueden ser C o C++.

3.4 Lenguaje interpretado

R es un lenguaje de programación del tipo interpretado.Un lenguaje interpretado es aquel que se ejecuta directamente, instrucción a instrucción, es decir que no se requiere que pase por un proceso de compilación, sino que tenemos un intérprete, que lee la instrucción en tiempo real y la ejecuta.Algunos beneficios de ser un lenguaje interpretado es que el código es multiplataforma y también tiene un aumento en rendimiento.Algunos lenguajes interpretados son Java, Python, Ruby.En el caso de lenguajes compiladores se encuentran OZ, C, C++.

3.5 Estructuras de control de flujo

Los bucles son un método de control de flujo para repetir una tarea o conjunto de tareas en un dominio.

Un ejemplo para obtener la suma acumulativa de un vector de valores en R utilizando un bucle for es:

x <- 1:4

acumulative\_sum <- 0

for (i in x){

cumulative\_sum <- cumulative\_sum + x[i]

}

acumulative\_sum

Otro bucle utilizado es el while, en el siguiente ejemplo se muestra como imprimir un vector de 5 elementos.

> x <- 1

> while(x < 5) {

x <- x+1;

print(x);

}

Finalmente existe el bucle repeat, el cual itera sobre una porcion de codigo hasta encontrarse con un break.En el siguiente ejemplo se muestra como imprime nuevamente un vector de 5 elementos.

x <- 1

repeat {

print(x)

x = x+1

if (x == 6){

break

}

}

4. Características Avanzadas:

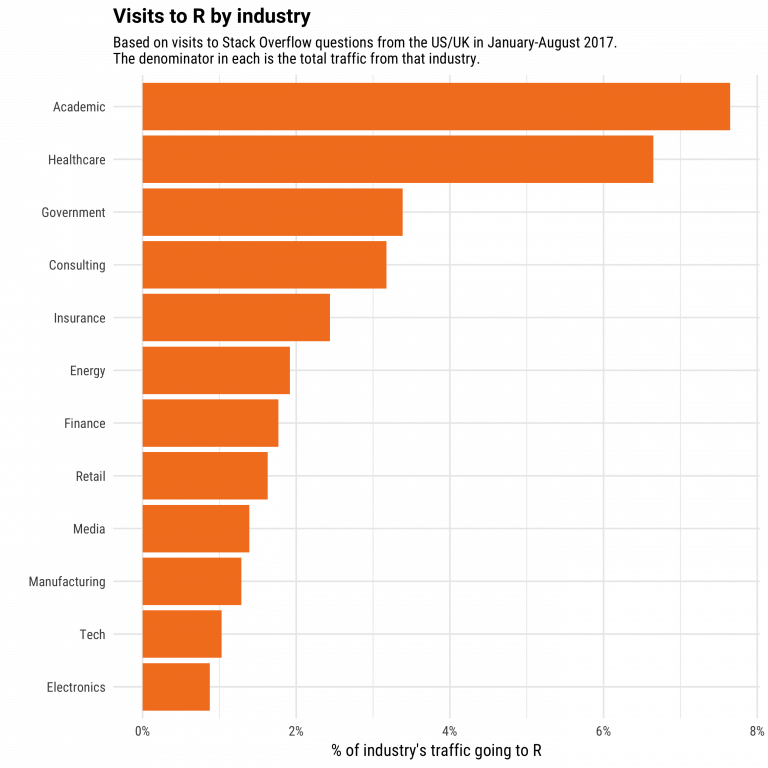
**Manejo de memoria: (desventaja)**

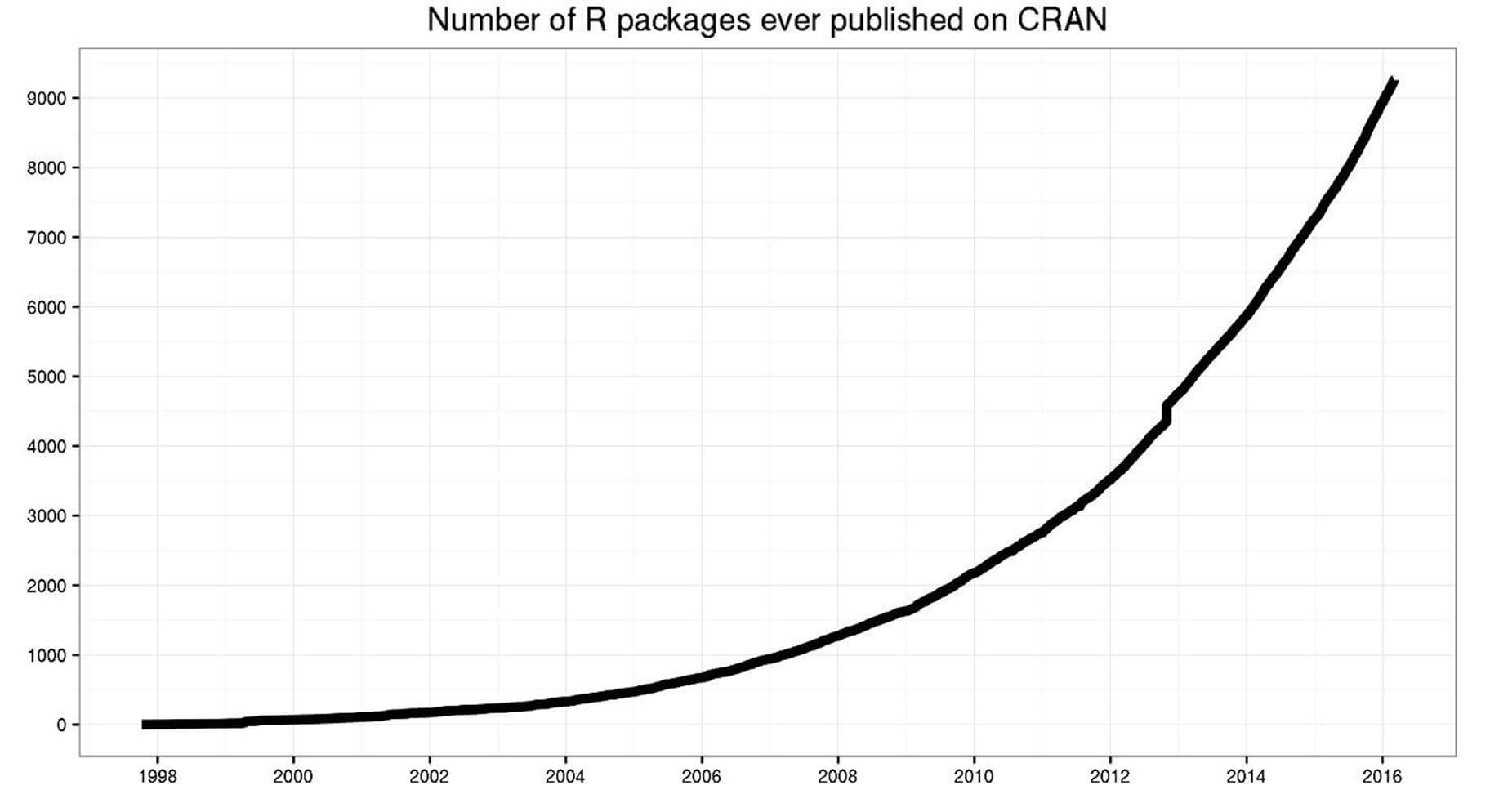
<http://learnr.usu.edu/base_r/environment/2_7_memory_management.php>

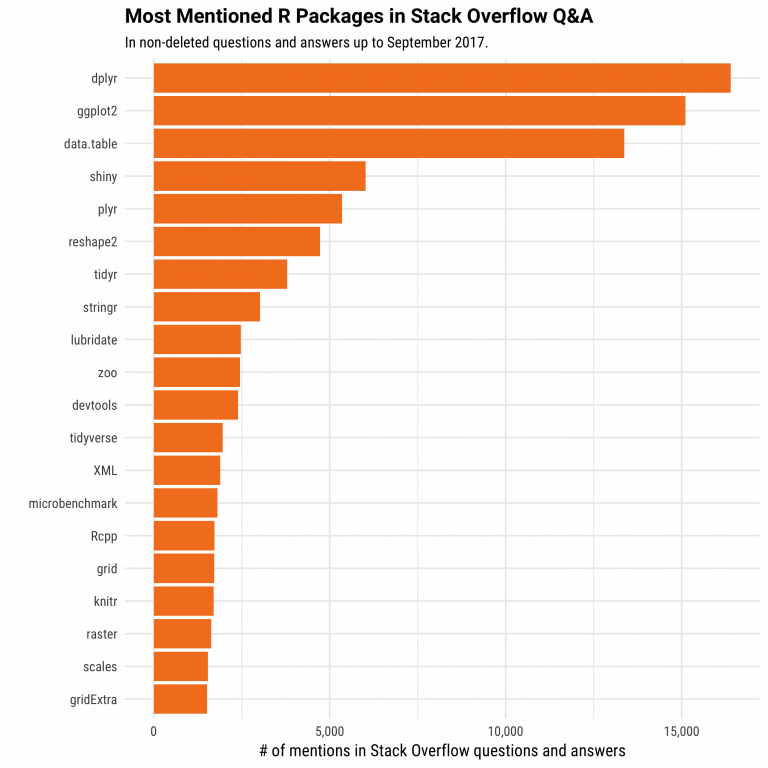
**Concurrencia: Buscar**

**Errores:**

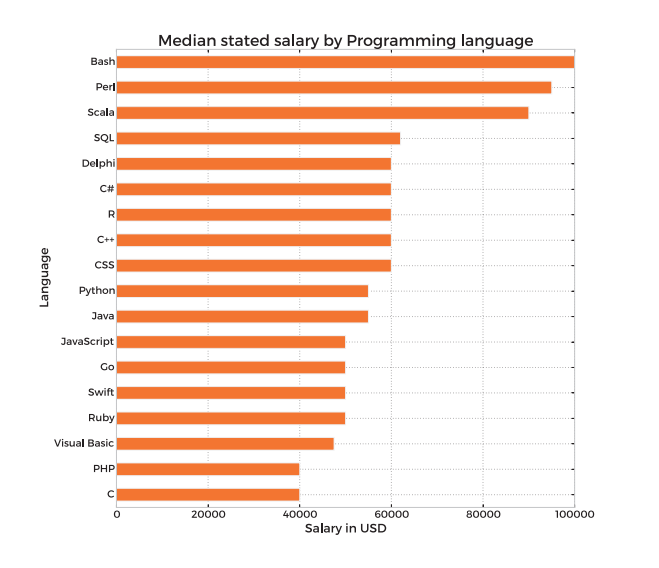
5. Estadísticas del lenguaje

5.1 Uso de R según industria:

5.2 Cantidad de Paquetes:

5.3 Paquetes mas usados:

5.4 R como uno de los lenguajes más usados

5.5 R como uno de los lenguajes mejores pagos

6. Comparación con otros lenguajes similares

6.1 Comparación con S

Contrario a S, R adoptó un modelo de evaluación, en el cual las definiciones en funciones anidadas son de scope estático.Esto quiere decir que las entidades, solo pueden ser usadas desde el bloque al que pertenecen o desde bloques internos.En S el valor de las variables libres está determinado por un conjunto de variables globales, en cambio en R está determinado por el ambiente en el que se encuentra.Al usar un scope estático, R al iniciar el programa, agarra una porción de memoria y almacena los objetos en la misma.R también maneja su propia memoria, agrandando o achicando según necesite. Al tener todo en memoria, esto genera que R funcione más rápido que S ya que S almacena los objetos en diferentes partes de memoria, normalmente en el .data. El aspecto negativo de R en este caso es que si el programa se rompe, se pierde el trabajo actual, es por esto que si se está trabajando en algo importante, se recomienda ir guardando y restaurando la memoria, aunque pueda ralentizar el proceso.

6.2 Comparación con Python

Python y R son dos lenguajes de programación open-source.R es un lenguaje mayormente utilizado por académicos y estadísticos.R también tiene uno de los ecosistemas más ricos al momento de analizar información, esto se debe a la cantidad de paquetes disponibles para cualquier tipo de análisis que se requiera.Una de las características principales que diferencian a R es la capacidad de mostrar el resultado de manera clara y sencilla, utilizando herramientas como Rstudio.En cambio python si bien posee las mismas funcionalidades, es más utilizado cuando se tiene grandes cantidades de datos, debido a la que el código en Python es más fácil de mantener y más robustos que R.

6.2.1 Usabilidad

R es una mejor opción para aquellos que no tiene experiencia en la programación debido a la facilidad del lenguaje.Esto también resulta en una mayor facilidad al momento de usar funciones complejas en R que Python.Por último los modelos estadísticos en R se pueden implementar con muy pocas líneas de código.En cambio Python por su parte es un lenguaje más conveniente para gente con experiencia previa en la programación.Debido a la sintaxis simple que posee, hacer programas y debuggear se vuelve más fácil y entendible.R además de contar con una gran cantidad de paquetes, los mismos pueden ser instalados con una única línea de código y se puede entenderlos fácilmente debido a la documentación de R.En cambio en Python se requiere instalar las librerías y sus dependencias para poder usarlas correctamente.

6.2.2 Ventajas y Desventajas de cada uno

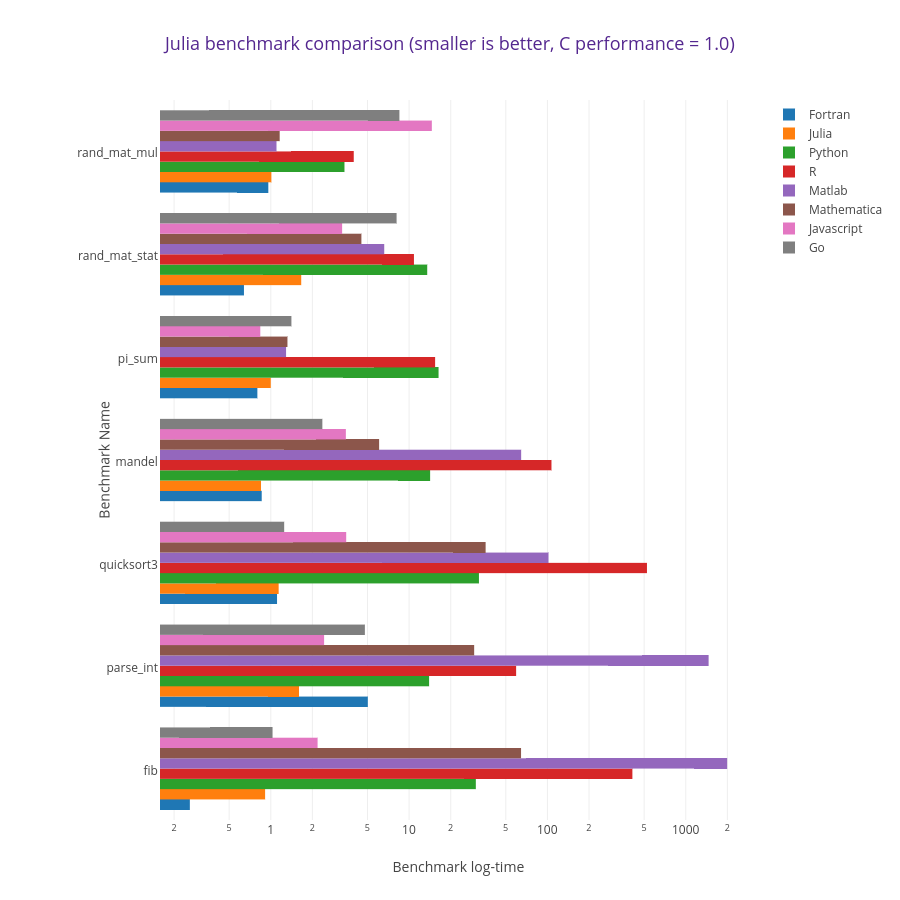
Una de las ventajas de R es que es considerada la mejor herramienta para generar gráficos y visualizaciones.También posee varias funcionalidades para el análisis de datos y es muy útil para el análisis estadístico.R posee un ambiente llamado Rstudio en donde se puede editar la información, debuggear y visualizar los gráficos.Python por otro lado es un lenguaje con mucha mas capacidad que simplemente el analisis de informacion.Es un lenguaje muy poderoso para la matemática computacional y también para entender y aplicar algoritmos.

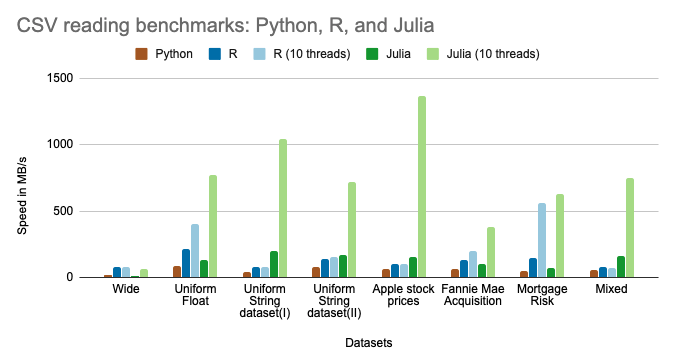
Debido a la gran cantidad de paquetes disponibles en R , esto puede resultar en que encontrar el paquete correcto demore demasiado tiempo.También si no se programa correctamente en R, esto puede generar que el programa no sea eficiente y tarde mucho en realizar tareas.Por último no es tan popular como lo es Python.Por otro lado Python no cuenta con tantos paquetes como R y la visualización de resultados se dificulta más.Debido a que en Python los errores aparecen en tiempo de ejecución, se requiere de test más rigurosos.

6.3 Comparación con Julia

Julia es otro lenguaje open-source al igual que R.Este lenguaje más reciente es mayormente usado para el análisis de datos, al igual que R.Las mayores diferencias es que al ser un lenguaje moderno Julia no cuenta con la cantidad de paquetes, no posee las herramientas para debuggear suficientes comparadas a R o Python. La gran ventaja que presenta Julia es su gran velocidad de procesamiento de la información, mucho mayor a R o Python.Entonces si la velocidad es algo importante para nosotros y necesitamos analizar grandes cantidades de datos, es probable que nos convenga utilizar Julia.Por otro lado si no nos importa tanto la velocidad, sino que buscamos resultados mejores y mejores visualizaciones de los mismos, R al ser un lenguaje más antiguo y contar con muchos mas paquetes para trabajar con la información, probablemente sea más conveniente usarlo, antes que Julia.

7. Benchmarks





8. Casos de Estudio

8.1 E-Commerce

Las empresas del comercio online es una de las que más utilizan data science.Debido a que las empresas tienen que trabajar con varios tipos de información, la cual puede o no estar estructurada , y en diversos formatos.R es una herramienta muy útil al momento de recolectar esta información y poder analizarla. También es muy utilizada para analizar los productos de venta cruzada, es decir cuanto un cliente compra un producto en particular, se utiliza R para recomendar productos similares que le podrían interesar.R también es utilizado por esta industria para los procedimientos estadísticos como el modelado lineal.Estos son necesarios para analizar las compras realizadas por los clientes y predecir las ventas de productos.

8.2 Redes Sociales

R es muy utilizado para el análisis de sentimientos de los usuarios y otras formas de minería de datos.R es muy utilizado para analizar posibles consumidores de productos y marcarlos como clientes para ciertos productos.Una herramienta muy popular es SocialMediaMineR mediante la cual se puede tomar una variedad de URLs y determinar la popularidad y el alcance de las mismas en las redes sociales.

8.3 Salud

Compañías farmacéuticas, genéticas y epidemiológicas utilizan mucho R.R es muy usado para realizar ensayos preclínicos y analizar los datos de seguridad de los medicamentos.R poseen un paquete muy popular llamado Bioconductor el cual posee varias herramientas para analizar los datos genómicos.También se utiliza para el modelado estadístico en el campo de la epidemiología donde los científicos de datos analizan y predicen la propagación de enfermedades.

8.4 Bancos

Los bancos utilizan R para modelar el riesgo de crédito y otras formas de analis de riesgos.Mortgage Haircut Model es un modelo utilizado para hacerse cargo de una propiedad en caso de un incumplimiento de préstamo. Este paquete implica la distribución del precio de venta, la volatilidad del precio de venta y el cálculo de déficit esperado.R también es utilizado junto con Hadoop para facilitar el análisis de calidad del cliente, la segmentación del cliente y la retención.El Bank of America, utiliza R para la información financiera.Con la ayuda del lenguaje el banco puede analizar las pérdidas financieras y utilizar las herramientas de visualización que posee R. J.P Morgan es otro que utiliza R, en su caso es utilizado para la detección de fraudes, gracias a los algoritmos de machine learning, les es más fácil predecir los fraudes y las irregularidades en los patrones de transacción.Por último R tambien es usado para calcular el valor de por vida de los clientes, para la segmentación de clientes y el análisis preventivo en tiempo real.