**Extracción de datos de Google**

Google Trends es una herramienta gratuita y de acceso público que permite analizar las tendencias de búsqueda en Google de diferentes temas o palabras clave. La herramienta busca proporcionar información sobre los cambios de popularidad de dichos términos en el tiempo o entre regiones. Los datos están disponibles a partir del año 2004.

Google Trends se basa en una muestra aleatoria de búsquedas realizadas por los usuarios y presenta los resultados en una escala de 0 a 100, que muestra la popularidad relativa para un término de búsqueda en relación con otros términos, en un período, o entre regiones especificadas, representando 100 el pico máximo de popularidad.

**Construcción de la base y método de extracción**

Para la extracción de datos, se construyó una base de datos de candidatos presidenciales latinoamericanos entre 2004 y 2022 en la que se incluyeron las siguientes variables:

* País
* Año de la elección
* Nombre Completo del Candidato
* Apellido del Candidato
* Nombre del partido
* Fecha de la elección

En este caso, la extracción se realiza con el paquete *gtrendsR* versión 1.5.1.90000 (Massicotte y Eddelbuettel, 2023) para el lenguaje de programación R.

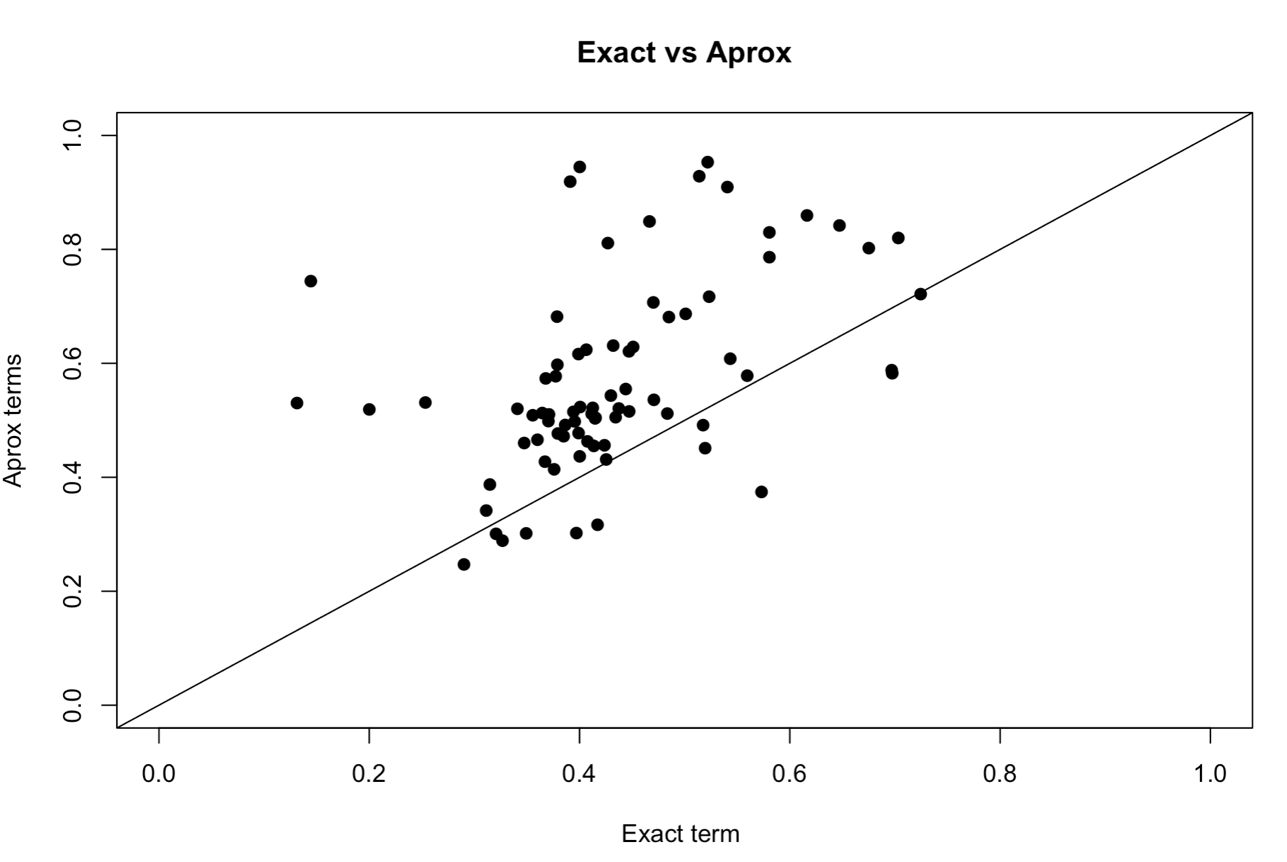
Se busca comprar la relevancia relativa entre las búsquedas al candidato y al partido del candidato para cada elección realizada en Latinoamérica entre 2004 y 2022. Los parámetros a configurar fueron los siguientes:

**1. Keyword**

Se realiza la extracción para cada pareja candidato-partido, con el fin de obtener la relevancia relativa entre estos dos términos.

Dos opciones: términos exactos o búsquedas relacionadas. Google proporciona la posibilidad de extraer las búsquedas realizadas al término

**Gráfico 1.** Comparación del indicador de personalismo sistémico con extracciones con términos exactos o aproximados.



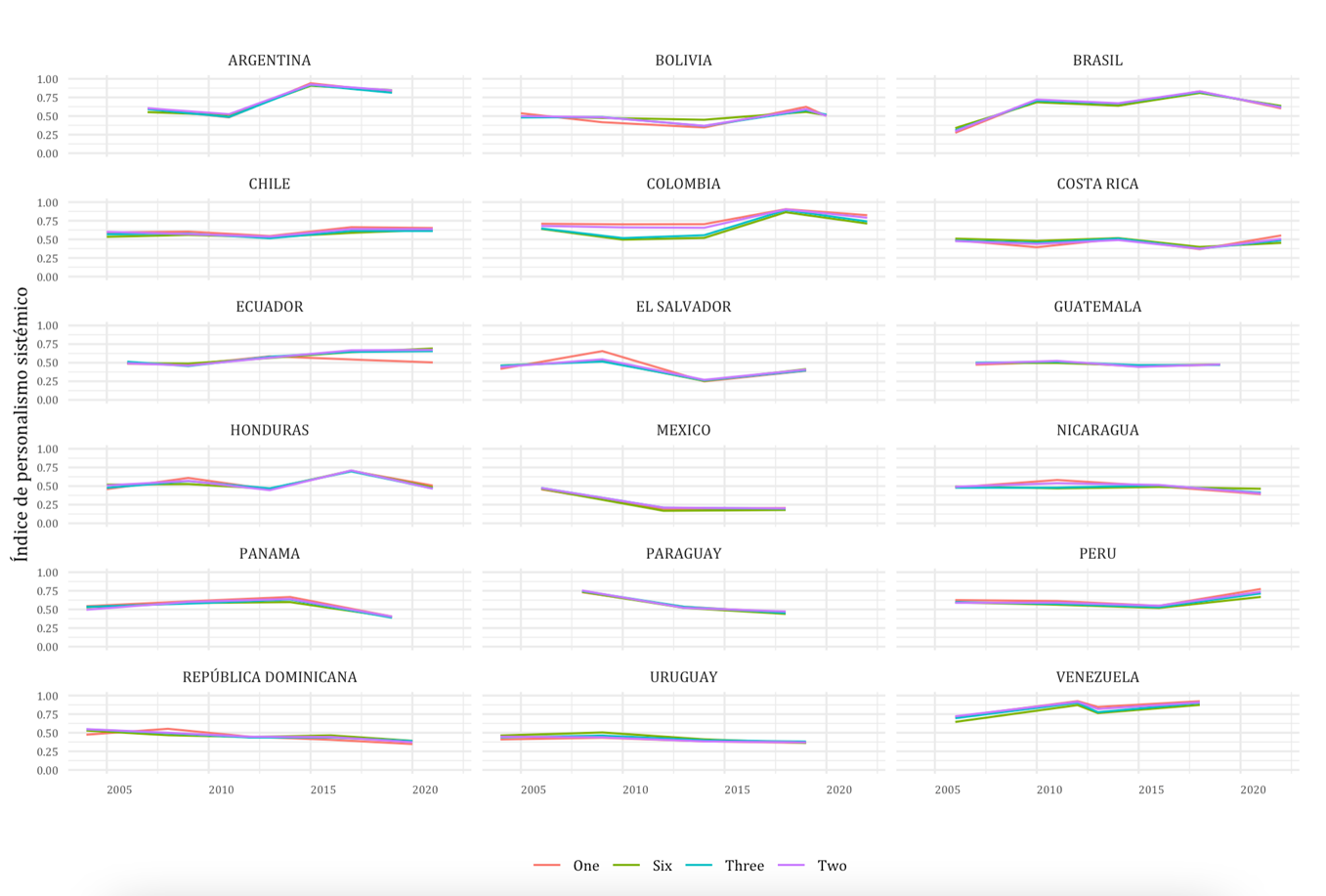
Como se puede ver, hay diferencias importantes entre realizar la extracción con términos exactos o permitiendo a Google Trends incluir las aproximaciones. En la mayoría de los casos, la extracción con los términos aproximados arroja valores mayores del indicador en comparación con solo incluir el término exacto.

Se opta por proseguir con los términos exactos, para esto se utiliza la pareja apellido del candidato y nombre del partido para cada extracción.

**2. Time**

Esta variable permite definir la ventana temporal para la extracción.

Las variaciones entre períodos más largos de extracción (seis meses) y más cortos (un mes) son poco significantes en la mayoría de los casos. La excepción es Colombia, debido a la desincronización de las elecciones parlamentarias respecto a las presidenciales. Para evitar este inconveniente y para agilizar la extracción de datos se optó por extraer solamente el último mes previo a la elección presidencial.

**Gráfico 2.** Comparación con diferentes períodos definidos para la extracción.

**3. Geo**

Corresponde a un código de dos letras que identifica al país en cuestión. La extracción se restringe a las búsquedas realizadas solamente en el país en el que se lleva a cabo la elección.

**5. Category**

Se define *category = 396*, que restringe la extracción a los resultados que Google clasificó en la categoría “política”.

**4. Gprop**

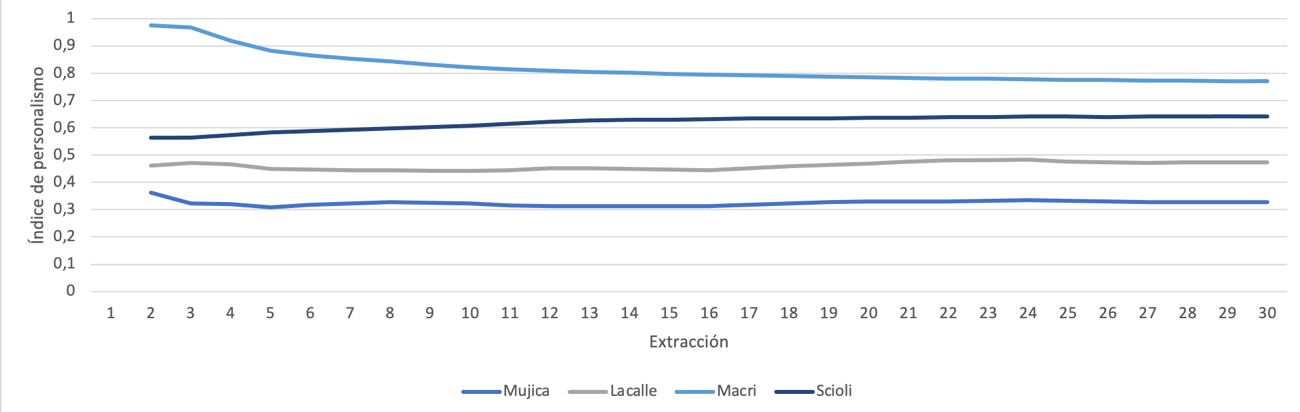
Se define *gprop = “web”*. Solamente se consideran las búsquedas en el motor principal de búsqueda de Google, no aquellas que corresponden a imágenes o realizadas en YouTube.

**Número de extracciones**

Google Trends se basa en una muestra aleatorias de búsquedas para computar el indicador que permite extraer. Para disminuir la probabilidad de que la extracción realizada sea producto de una muestra aleatoria que se ubique en un extremo de la distribución de probabilidad, se consideró conveniente realizar varias extracciones y luego promediarlas.

El gráfico siguiente muestra como varía el indicador según la cantidad de extracciones realizadas. Se estimó conveniente realizar al menos 15 extracciones para asegurar la estabilidad del mismo.

**Gráfico 3** . Variación del indicador por número de extracciones promediadas



**Cálculo del indicador**

La extracción arroja como resultados dos vectores de interés relativo *– hits –* uno que corresponde al nombre del partido y otro al del candidato. Para cada extracción país-elección, este indicador varía entre 0 y 100, siendo 100 el punto de interés máximo.

El indicador de personalismo – *Ipers* – a nivel de candidato es la proporción de los hits del candidato sobre la suma total de hits (candidato y su partido) entre 30 días anteriores a la elección y el día de la misma *– d –.*

Los candidatos independientes adquieren el valor 1 en el índice de a nivel de candidato.

El indicador de personalismo a nivel sistémico se calcula agregando el indicador anterior ponderado por la proporción de votos – *V* – obtenida por cada partido – *i –* .

El resultado es un valor único por elección/país para cada elección llevada a cabo en Latinoamérica entre 2004 y 2022.

Definidos así, ambos indicadores varían entre 0 y 1, indicando 0 un nulo personalismo (solamente búsquedas al partido) y correspondiendo la unidad a una personalización absoluta (solamente búsquedas al candidato).