



Tablero de control (Dashboard) de correlación consultas en Google y VIH

UNIVERSIDAD EL BOSQUE
MAESTRÍA DE ESTADÍSTICA APLICADA Y CIENCIA DE
DATOS
ESTADÍSTICA Y PROBABILIDAD – PROGRAMACIÓN EN R

Presentado por:

Mario Benedetti
Nelson López
Oscar Vanegas
Luis Acuña
Diana Parra Galvis

Presentación de la temática

Virus de la Inmunodeficiencia Humana
(VIH) y comorbilidades

**Virus que ataca el
sistema
Inmunológico del
cuerpo**

**Contagio por
intercambio de
fluidos
corporales
infectados
(sangre, leche
materna, semen,
etc)**

**Personas con
TAR y carga
vímica suprimida,
no transmiten el
virus a sus
parejas sexuales.**

**Fase más
avanzada es el
Sida**



38.4 millones de personas en el mundo
tienen VIH

2.1 millones en LAC

180.000 casos de personas con VIH en
Colombia



Colombia

Epidemia en aumento desde 1984.

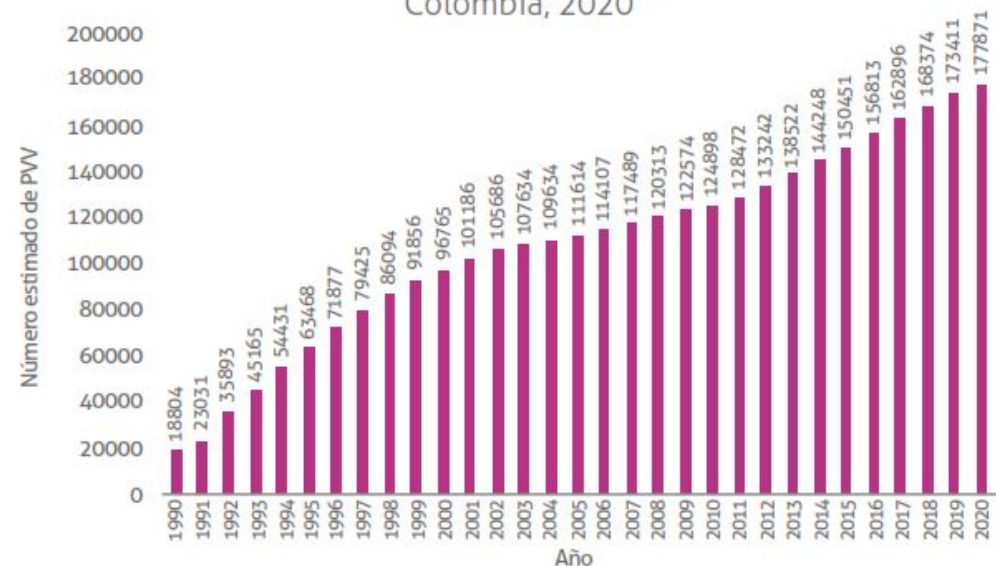
2019: Epidemia concentrada según grupos poblacionales de mayor riesgo : HSH, Mujeres TRANS, PID, TS

2020: 78,5% hombres del total de casos notificados al SIVIGILA.

2020: Prevalencia estimada de 0.44% en personas de 15 a 49 años.

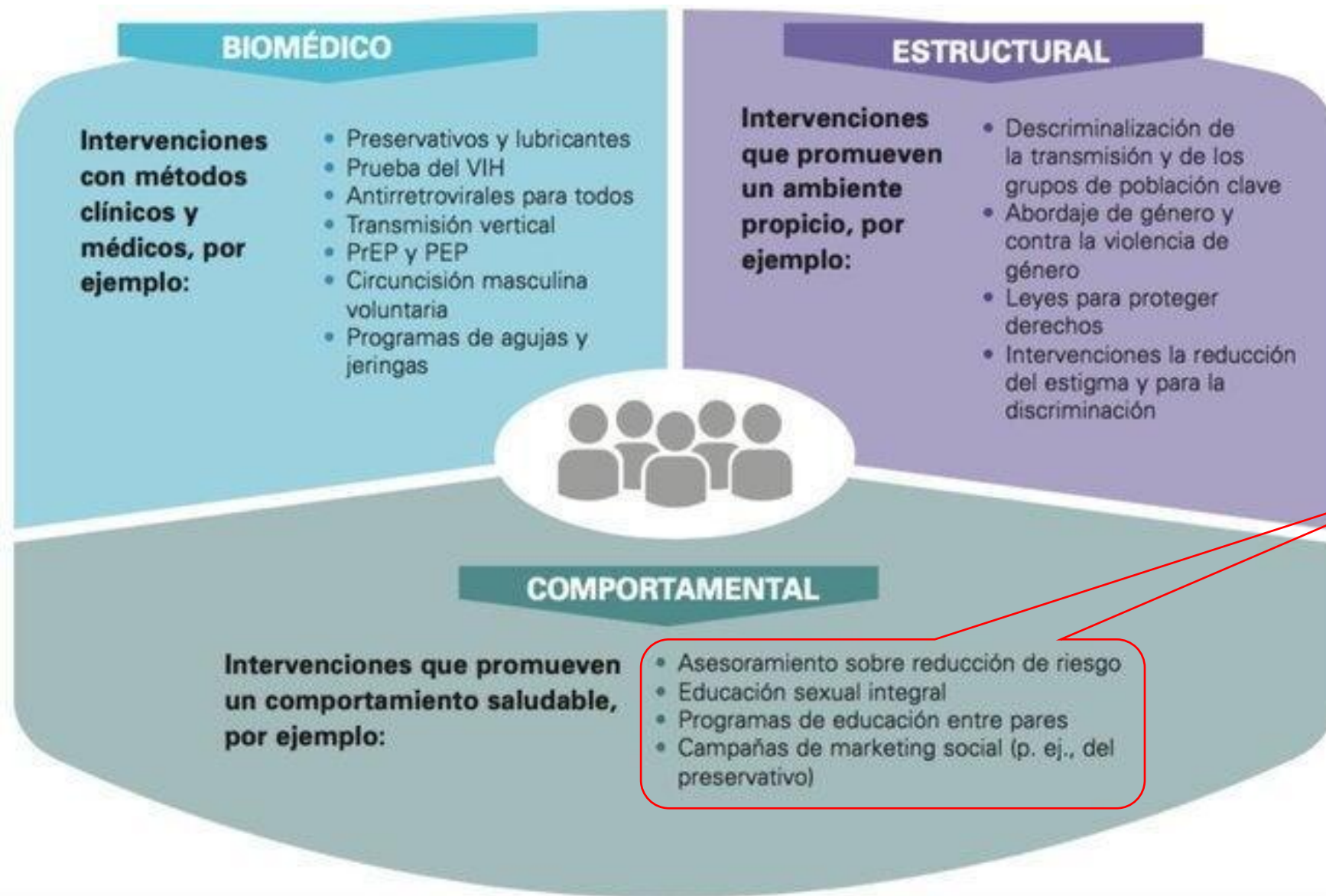
El 77% de los casos de VIH se observan en las personas entre 15 a 49 años.

Gráfico 1. Estimación de población que vive con VIH con base en actualización de Spectrum. Colombia, 2020



Fuente: Ministerio de Salud y Protección Social y ENTerritorio. Boletín de la situación del VIH de acuerdo con estimaciones realizadas en Spectrum. 2021 (Pendiente publicación).

Prevención combinada de la infección por el VIH



Las cifras de infecciones siguen en aumento.

Servicios de salud estigmatizantes y discriminatorios.

Falta de mecanismos efectivos para fortalecer la demanda inducida a los servicios de salud.

Necesidad de estrategias de identificación de las poblaciones en los territorios.

Aumento del uso de mecanismos electrónicos para conocer acerca de las enfermedades (automedicación).

Para 2020 los grupos de edad de 12 a 24 años (84.1%) y de 24 a 54 años (76.3%), aumentaron el uso de internet.

¿Existe correlación entre el comportamiento de las consultas semanales de los términos relacionados con la enfermedad del VIH/Sida en el buscador Google y la notificación de los casos del evento en el año 2020?

Objetivos del proyecto

OBJETIVO GENERAL

Establecer la correlación entre el comportamiento de las consultas semanales de términos relacionados con la enfermedad del VIH/Sida en el buscador Google y la notificación de los casos del evento en el año 2020.

OBJETIVOS ESPECÍFICOS

- Observar el comportamiento de las consultas de términos relacionados al VIH/Sida realizadas en el buscador (Google) de manera semanal.
- Describir el comportamiento semanal de la notificación de los casos de VIH/Sida en el sistema de información nacional.
- Contrastar a través de un Dashboard el comportamiento de las consultas en Google respecto a la notificación del evento VIH/Sida semanal en el sistema de información nacional.

Fuentes de información

1. SIVIGILA
2. Google Trends

Fuentes de información

1. Sistema nacional de vigilancia en salud pública –SIVIGILA-



Estructura de datos

Base de datos de escritorio: Los datos migran semanalmente (o de manera inmediata según corresponda) por correo electrónico en la cascada definida por el INS.



INSTITUTO
NACIONAL DE
SALUD

Modularidad:

- Notificación individual de casos: registro de variables de identificación del paciente y sociodemográficas, registro de la notificación (evento a notificar), datos complementarios (específicos según evento a reportar).

Salida de información en archivo plano según evento a visualizar.

Variables tipo categórico o numérico según corresponda.

Fuentes de información

1. Sistema nacional de vigilancia en salud pública –SIVIGILA-



Fuente de información administrada por el INS

Recopila diariamente información acerca de los EISP

Alimentada en los servicios de salud de todo el país



INSTITUTO
NACIONAL DE
SALUD

Cuenta con información sociodemográfica y propia de cada evento

Observación y análisis objetivo, sistemático y constante de los eventos en salud

Consolida información de más de 85 EISP

Fuentes de información

2. Tendencias de búsqueda en web por términos (Google)



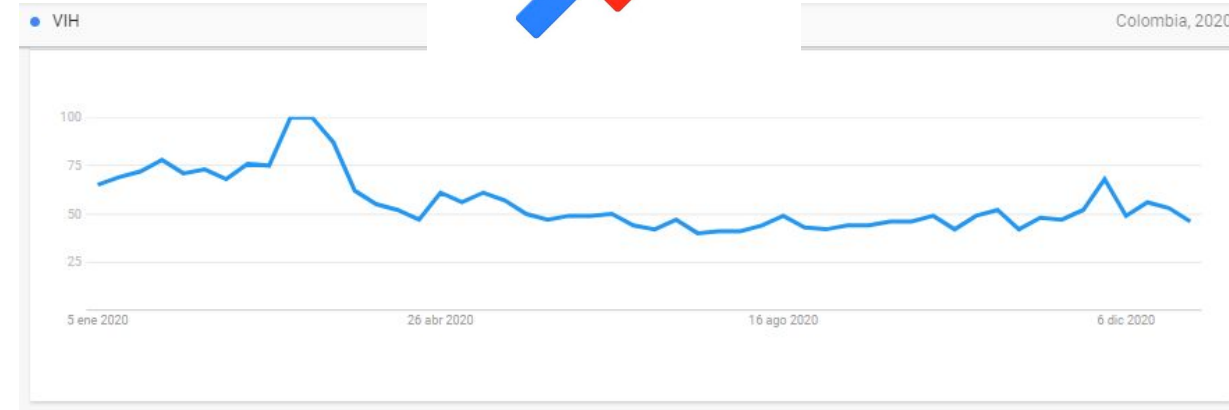
Variable:

Se calcula de manera automática un índice de interés del término de búsqueda, en un intervalo entre 0 y 100, siendo 100 el máximo durante un período y 0 el mínimo en el mismo intervalo de tiempo.

Interpretabilidad:

Se realizarán transformaciones para calcular el aumento o disminución entre el interés de las búsquedas, con una variación entre una semana epidemiológica y la siguiente.

Google Trends



Estructura de datos

Los datos se pueden extraer de la plataforma después de introducir términos de búsqueda. Se sugieren de manera automática búsquedas relacionadas con un término. Se presentan datos con agregación nacional y desagregación temporal a nivel de semanas. Exportables a CSV.

Fuentes de información



Términos de Búsqueda

Se emplearán operadores lógicos “AND” “OR” combinados entre los términos de búsqueda definidos de acuerdo al objetivo del estudio

Términos de Búsqueda

- VIH
- SIDA
- Enfermedad
- Transmisión
- Sexual
- Sífilis
- Gonorrea
- VPH/Papiloma
- Protección
- Prevención
- Cura
- Casos
- Salud
- Retroviral
- Medicamentos
- Tratamientos
- Inmunodeficiencia

Metodología



Metodología

Validación con portadores de la enfermedad

1. Selección de los posibles términos utilizados por personas infectadas o casos sospechosos



Metodología

2. Descarga de datos relacionados a la frecuencia (“Interés”) con la que son buscados los términos de interés.

Fuente de datos: Google Trends

determinado tema. Para cualquier **intervalo** de tiempo seleccionado siempre encontraremos un valor 100. Esto indica que durante ese período el término ha experimentado el máximo de relevancia posible dentro del **intervalo** seleccionado. De esta forma, el resto de valores relativos solo indican la relevancia en comparación a ese período máximo. EJ:





Metodología

Fuente de datos: Google Trends y
SIVIGILA

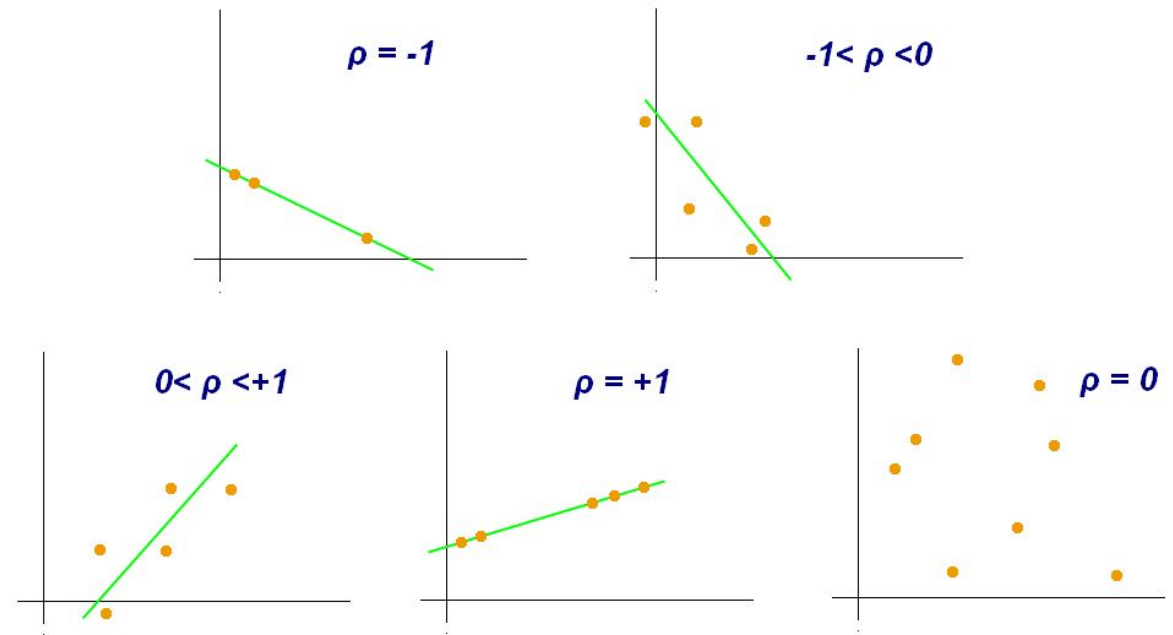
3. Limpieza de datos y preparación de la base
de datos.

Metodología

Software: R Studio

4. Calcular la correlación entre las variables:

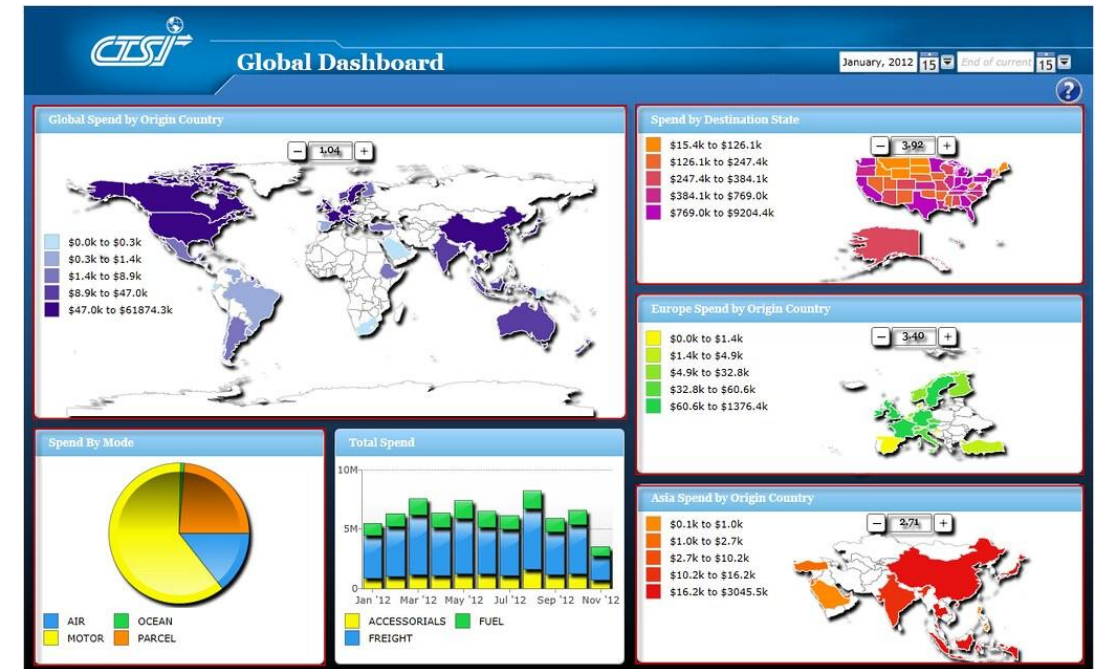
Casos notificados de VIH vs Términos buscados



5. Elaboración de Dashboard dinámicos

Metodología

Software: R Studio - Shiny





Metodología

6. Análisis de resultados y Conclusiones



~~PLAN A:~~ ¿Existe correlación entre el comportamiento de las consultas en buscador (Google) semanales con términos relacionados con la enfermedad del VIH/Sida y el comportamiento de la notificación de los casos en el año 2020?

PLAN B: ¿Existen diferencias entre el número de casos notificados al sistema de vigilancia en salud pública del país del evento VIH/Sida con respecto a las variables **género, estrato socioeconómico y nacionalidad** por departamento en el año 2020?