

Universidad de Costa Rica

Facultad de Ciencias

Escuela de Matemática

Herramientas De Ciencia De Datos I CA-0204

Reporte del Proyecto Individual

Estudiante: Daniela Prado Vargas

Carné: C26070

Profesor: Joshua Cervantes Artavia

Grupo: 2

II Semestre, Año 2025

1. Introducción

El siguiente proyecto de investigación se basa en el análisis de datos provenientes del Wellcome Global Monitor 2020, un estudio internacional que contiene información sobre la percepción y confianza que tienen las personas en la ciencia, el gobierno y otras instituciones relacionadas con la salud. El conjunto de datos original incluye información de casi todos los países alrededor del mundo, sin embargo, en este proyecto se trabajó únicamente con los países de América, abarcando tanto América Latina como Norteamérica. El objetivo principal fue explorar las tendencias de confianza, afectación del covid en la economía de los hogares, y cuánto apoyan las personas que se ayuden a otros países en la prevención y cura de enfermedades.

2. Explicación del código

2.1. Capítulo 0: Preparación de datos

El primer paso del proyecto consistió en la limpieza de los datos para su respectivo análisis utilizando el paquete de tidyverse para la transformación de las tablas que se iban a manipular. Además, se usaron los paquetes scales y ggrepel para la personalización de los gráficos, ya que los mismos funcionan para una mejor visualización. Los datos elegidos del Wellcome Global Monitor 2020, se cargaron desde un archivo CSV, y se declaró un vector con una paleta de colores para mantener la estética visual durante todo el proyecto.

Para facilitar el manejo de los datos, se utilizó la función rename() del paquete tidyverse, la cual se permitió reemplazar los nombres originales de las variables por etiquetas descriptivas. Esto hizo que las columnas fueran más fáciles de comprender, por ejemplo, transformando códigos como W1 en conocimiento_ciencia o WPID_RANDOM en id_unico, basándose en el diccionario de variables del Wellcome Global Monitor (Wellcome, 2020).

A continuación, se filtraron los datos para mantener únicamente los países de interés, como se mencionó antes: los de América Latina y Norteamérica, utilizando la función filter(). Además, se seleccionaron solo las variables relevantes para el análisis, ya de esta forma el DataFrame iba a ser más compacto y fácil de trabajar.

El siguiente paso fue recodificar las de variables. Para ello se utilizaron varias funciones de R, especialmente mutate() y case_when(). La función case_when() transforma los valores de las variables según con condiciones específicas, logrando así convertir los códigos numéricos en categorías menos complicadas de leer. Por ejemplo, los niveles de confianza (antes codificados como 1, 2, 3, 4) se recodificaron a “Mucho”, “Algo”, “Poco” y “Nada”. De manera similar, las variables de tipo sí/no se recodificaron incluyendo la opción “NS/NR” para los casos donde el encuestado no sabía o no quería responder. También, se aplicaron recodificaciones para grupos de edad, nivel educativo, ingresos y estado laboral, para facilitar las comparaciones durante el desarrollo del trabajo.

Para variables con nombres que empezaran igual, como todas las relacionadas con confianza (confianza_*) o con la confianza en fuentes de información sobre COVID (covid_confia_*), se utilizó la función across() junto con un mutate(). Esto permitió aplicar la misma

recodificación a varias variables de manera eficiente, sin tener que escribir una recodificación para cada una por separado.

En el tratamiento de valores faltantes, se reemplazaron los datos ausentes con “NS/NR” según correspondiera, usando `ifelse()` dentro de `mutate()`, a excepción de la variable “ingreso_hogar_quintil”, ya que en esta se aplicó un valor medio a todos los valores faltantes. Esto, para asegurar que los cálculos posteriores y las visualizaciones no se vean afectados por los valores nulos y que todas las categorías sean consistentes.

2.2. Capítulo I: El impacto desigual del Covid-19 en América

En el primer capítulo, el análisis comienza con el gráfico 1.1, el cual muestra el porcentaje de trabajadores que perdieron su empleo o negocio debido al COVID-19, agrupado por el nivel de ingreso del país, según el Banco Mundial. Para esto, se utilizó la función `filter()` para asegurarse de que solo se incluyeran los casos en los que los trabajadores respondieron afirmativamente o negativamente a la pregunta sobre si perdieron su trabajo. Luego, se aplicó un `group_by()` para agrupar los datos de acuerdo con el nivel de ingreso del país, categorizado en “bajo”, “medio” o “alto”. Posteriormente, se empleó la función `summarize()` para calcular el total de respuestas por grupo y el número de trabajadores que indicaron haber perdido su empleo. El porcentaje de trabajadores que perdieron su trabajo se calculó dividiendo el número de trabajadores que respondieron afirmativamente entre el total de trabajadores del grupo, multiplicado por 100. Finalmente, el gráfico se generó con `ggplot2` utilizando la función `geom_col()`, que crea un gráfico de barras, donde cada barra representa el porcentaje de trabajadores que perdieron su empleo dentro de cada nivel de ingreso.

Para la tabla 1.1, se creó con el objetivo de mostrar cómo distintas consecuencias económicas afectaron a los trabajadores. Se utilizó la función `group_by()` para agrupar los datos por país y la función `summarize()` para calcular el porcentaje de trabajadores que experimentaron las consecuencias del COVID, tales como la pérdida del empleo, un paro temporal, la reducción de las horas de trabajo o la disminución de ingresos. Cada una de estas consecuencias se calculó como el porcentaje de respuestas afirmativas, es decir, las personas encuestadas que respondieron “sí”, utilizando la función `mean()` para obtener la media de las respuestas por cada país. Luego, los países se ordenaron según el porcentaje de trabajadores que perdieron su empleo o negocio, utilizando la función `arrange()`, lo que permitió tener una mejor visualización de los datos.

Y por último, en el gráfico 1.2, nos enfocamos en analizar la pérdida de empleo según el quintil de ingreso dentro de cada país. Primero, se utilizó la función `mutate()` para ordenar la variable `ingreso_hogar_quintil`, que indica el quintil de ingreso al que pertenece cada hogar. Luego, se aplicó un `group_by()` y un `summarize()` para calcular el porcentaje de trabajadores que perdieron su empleo dentro de cada quintil de ingreso. Al igual que en los pasos anteriores, se usó la función `mean()` para obtener el porcentaje de respuestas afirmativas dentro de cada quintil. Finalmente, se generó el gráfico con `geom_bar()`, creando un gráfico de barras horizontales, donde cada barra representa el porcentaje de trabajadores que perdieron su empleo, segmentado por quintil de ingreso.

2.3. Capítulo II: Apoyo público al gasto global en la prevención y cura de enfermedades

En el capítulo 2, se comenzó con el gráfico 2.1, cuyo objetivo era mostrar la distribución de las opiniones de las personas sobre el gasto gubernamental en salud. En particular, se quería saber si las personas creen que el dinero debería gastarse para ayudar a otros países a prevenir y curar enfermedades, o si solo debe invertirse si las enfermedades representan un riesgo para el propio país. Para ello, se transformaron los datos a formato largo seleccionando las columnas relevantes, que eran "Gobierno ayuda a todo el mundo" y "Gobierno ayuda solo al país", que contenían las respuestas a las preguntas sobre el gasto en salud. Luego, utilizando la función `pivot_longer()`, se transformaron estas columnas a un formato largo para facilitar el cálculo de las respuestas. Después, con un `group_by()` y `summarize()`, se agruparon los datos por pregunta y respuesta, y se contó cuántas personas respondieron cada opción. Usando `mutate()`, se calculó el porcentaje de respuestas para cada combinación de pregunta y respuesta, de manera similar a los gráficos anteriores. Finalmente, se utilizó `factor()` para ordenar los niveles de las respuestas de acuerdo con su duración. Para generar el gráfico, se empleó `ggplot2` con la función `geom_bar()`, creando un gráfico de barras donde cada barra representaba el porcentaje de respuestas para cada tipo de gasto gubernamental.

A continuación, se construyó el gráfico 2.2, que compara el acuerdo con dos enfoques de gasto gubernamental en salud por país. En este gráfico, se comparó el porcentaje de personas que estaban de acuerdo con ayudar a otros países o con mantener los gastos solo para su propio país. Primero, se calculó el porcentaje de acuerdo por país y tipo de ayuda, seleccionando las columnas relacionadas con las respuestas de los países para ambos enfoques de gasto. Luego, se transformaron estas columnas a formato largo utilizando `pivot_longer()`, al igual que en el gráfico anterior. Se filtraron solo las respuestas de "mucho" y "algo", que indicaban el acuerdo, y se calcularon los porcentajes para cada país y tipo de ayuda con `group_by()` y `summarize()`, seguido de `mutate()` para calcular el porcentaje. Después, se separaron los datos en dos grupos: uno para los países que respondieron "mucho" o "algo" sobre el gasto en salud global, y otro para los países que respondieron sobre el gasto solo dentro de su propio país. Para unir estos datos, se utilizó `inner_join()` y luego `pivot_longer()` para fusionar los dataframes en uno solo, de manera que cada fila representara el porcentaje para cada país y tipo de gasto. Como había muchos países, se creó un índice para dividirlos en dos grupos para mejorar la visualización. Finalmente, se generó un gráfico de barras agrupado por país y tipo de ayuda.

2.4. Capítulo III: Confianza y valoración percibida de la ciencia en el contexto del Covid-19

En el capítulo 3, se comenzó con el gráfico 3.1, el cual analiza cómo las personas perciben que diferentes fuentes, como líderes religiosos, familia, gobierno, OMS y médicos, toman decisiones sobre el coronavirus basándose en asesoramiento científico. La visualización muestra el porcentaje de respuestas para cada fuente y cómo se distribuyen entre las diferentes categorías. Primero, se transformaron los datos a formato largo utilizando `pivot_longer()`, lo que permitió convertir las columnas relacionadas con la confianza en un

formato más manejable, donde cada fila representaba una combinación de fuente, quién tomó la decisión y la respuesta, mostrando cómo percibieron las personas la base científica de esas decisiones. Luego, se recodificaron las respuestas con la función `factor()` para asegurarse de que las categorías se ordenaran correctamente. A continuación, se agruparon los datos por fuente y respuesta, y se contó cuántas respuestas hubo en cada categoría utilizando `summarize()`. Posteriormente, con `mutate()`, se calculó el porcentaje de respuestas para cada fuente, dividiendo la cantidad de respuestas en cada categoría entre el total de respuestas para esa fuente. Finalmente, se generó un gráfico de barras con `ggplot2` y `geom_bar()` para mostrar la relación entre las respuestas.

Para la tabla 3.1, que muestra el porcentaje de personas que creen que las decisiones de los diferentes actores están basadas en asesoramiento científico, centrado en las respuestas "mucho", primero se transformaron los datos a formato largo utilizando `pivot_longer()`, con cada fila representando una combinación de país y fuente. Luego, se calculó el porcentaje de respuestas "mucho" agrupando los datos por país y fuente. Usando `summarize()`, se calculó el porcentaje de personas que respondieron "mucho" en relación con cada actor, y se utilizó `mean()` para obtener este valor. Posteriormente, se reorganizaron los datos con `pivot_wider()` para presentar cada actor como una columna separada en la tabla.

Por último, para la tabla 3.2, que compara la creencia de que las decisiones del gobierno y la OMS sobre el coronavirus están basadas en asesoramiento científico, se seleccionaron las columnas relacionadas con la confianza en la OMS y el gobierno y se utilizaron `pivot_longer()` para convertirlas a formato largo. Luego, se calculó el porcentaje de respuestas con `summarize()` y `mutate()`, siguiendo el mismo proceso de los gráficos anteriores. Después, se calculó la diferencia entre la confianza en la OMS y el gobierno para facilitar el análisis y se ordenaron los resultados en función de esta diferencia.

Para el gráfico 3.2, se analizó la relación entre la confianza en los científicos y la percepción de que la ciencia beneficia al país. Primero, se transformaron los datos a formato largo usando `pivot_longer()`, para combinar las respuestas sobre la confianza en los científicos y los beneficios de la ciencia. Luego, se calculó el porcentaje de respuestas positivas por país utilizando `summarize()` y `mutate()`. Con los datos procesados, se creó un gráfico de dispersión, donde el eje X representa el porcentaje de confianza en los científicos y el eje Y la percepción de que la ciencia beneficiará al país. Se añadió una línea de regresión utilizando `geom_smooth()` para identificar las tendencias generales, y finalmente, se etiquetaron los nombres de los países con `geom_text_repel()` para mejorar la visualización.

En el gráfico 3.3, el análisis siguió un enfoque similar, pero en este caso se centró en la relación entre la confianza en el gobierno y la percepción de que el gobierno toma decisiones basadas en la ciencia. Los datos se agruparon por país y se calcularon los porcentajes de respuestas afirmativas para ambas variables. Luego, se visualizó la relación con un gráfico de dispersión, similar al del gráfico anterior, destacando la correlación entre ambas variables e incluyendo una línea de regresión.

Finalmente, en la tabla 3.3, se calcularon los porcentajes para mostrar los países donde las personas son más propensas a decir que los líderes del gobierno no valoran la opinión de los

científicos. Para ello, primero se calcularon los porcentajes de personas que respondieron "mucho" o "algo" en comparación con los que respondieron "poco" o "nada". Los resultados se organizaron por país y se calculó la diferencia entre estos porcentajes. La tabla se formateó para mostrar los porcentajes de "mucho" o "algo" y "poco" o "nada", junto con la diferencia entre ambos, y los países se ordenaron según esa diferencia.

3. Resultados

En el primer capítulo se observa que Perú fue el país más afectado por la pandemia, con el 49% de los trabajadores perdiendo su empleo y más de 60% experimentando reducciones en horas o ingresos. Otros países de América Latina, como Ecuador, Colombia y Bolivia, también sufrieron un impacto significativo, con más del 30% de los trabajadores perdiendo sus empleos. En contraste, los efectos en países como Canadá y Estados Unidos fueron mucho menores, con menos del 10% de los trabajadores perdiendo su empleo y alrededor del 25-28% experimentando reducciones en sus ingresos. Además, la relación entre el nivel de ingresos y la vulnerabilidad económica es clara: los países con ingresos más bajos y los hogares con menos recursos fueron los más afectados, lo que resalta la mayor vulnerabilidad de los sectores más pobres durante la crisis.

En el segundo capítulo, el análisis del gasto en salud muestra que la mayoría de la población prefiere que el gasto público en salud se destine a proteger sus propios intereses nacionales. Sin embargo, existe un apoyo significativo hacia la cooperación internacional, aunque con una preferencia por la protección interna. En América Latina, países como Perú, Argentina y Bolivia priorizan el gasto para proteger a sus propios ciudadanos, mientras que países de Norteamérica, como Estados Unidos y Canadá, muestran una mayor inclinación a apoyar la cooperación internacional en salud, especialmente en situaciones de crisis global. Esto refleja una división entre las prioridades nacionales de América Latina y la disposición de los países norteamericanos para colaborar en iniciativas de salud globales.

Y por último, en el tercer capítulo, el análisis de la confianza en la ciencia y en las decisiones gubernamentales relacionadas con el Covid-19 revela que, en general, los médicos y la OMS son las fuentes más confiables, con países como Uruguay, Canadá y Costa Rica mostrando niveles de confianza superiores al 70%. En cambio, los líderes religiosos y la familia son considerados fuentes menos científicas, con menor confianza en la mayoría de los países. Además, la relación entre la confianza en los científicos y la percepción de que la ciencia beneficia al país es positiva, indicando que en los países con mayor confianza en los científicos, la percepción de los beneficios de la ciencia es también más alta. Por otro lado, la confianza en el gobierno y la percepción de que sus decisiones están basadas en ciencia varía significativamente entre países. Mientras que en Uruguay y Canadá la confianza en el gobierno se asocia estrechamente con la percepción de decisiones basadas en ciencia, en países como Estados Unidos y Brasil se observa una desconexión, lo que sugiere una menor relación entre la confianza en el gobierno y la percepción de decisiones informadas científicamente.

4. Recomendaciones

Se recomienda fortalecer las políticas laborales en países con mayor impacto económico, como Perú, Bolivia y Ecuador, promoviendo medidas de protección laboral ante emergencias, como empleo temporal, redes de apoyo económico para trabajadores informales y programas de capacitación. Además, es crucial reducir la vulnerabilidad de los hogares, especialmente los de menores ingresos.

Es necesario incrementar la cooperación internacional en salud. A pesar de la tendencia en América Latina de priorizar los intereses nacionales, se debe equilibrar con la importancia de la solidaridad internacional, especialmente en situaciones de crisis global como la pandemia. Aumentar la participación en iniciativas globales de salud y fomentar alianzas con organizaciones internacionales ayudará a mejorar el acceso a recursos y tecnologías en la región.

Fomentar la confianza en la ciencia y las autoridades sanitarias es esencial, especialmente en países como Venezuela, República Dominicana y Brasil, donde esta confianza es baja. Para ello, se debe implementar una estrategia de comunicación transparente y promover la educación científica y campañas informativas en salud pública. Además, involucrar a los líderes comunitarios en la difusión de información, especialmente en áreas rurales, será clave para combatir la desinformación.

En cuanto a la transparencia gubernamental, es fundamental reforzar la rendición de cuentas en países con baja confianza en el gobierno, como Estados Unidos y Brasil. Implementar auditorías independientes, generar informes accesibles y promover la participación ciudadana en la toma de decisiones ayudará a mejorar la percepción pública y aumentar la credibilidad del gobierno. Además, fortalecer las relaciones con organizaciones científicas garantizará que las decisiones sanitarias estén basadas en evidencia sólida.

Finalmente, es vital priorizar el apoyo a los sectores más vulnerables. Se deben implementar políticas económicas inclusivas que brinden apoyo directo a los hogares de bajos ingresos, como subsidios, y acceso a servicios básicos. Crear programas de empleo social y fortalecer las redes de seguridad social ayudará a mitigar el impacto económico de futuras crisis y a reducir la desigualdad.

Referencias:

Wellcome. (2020). Wellcome global monitor: Covid-19, 2020. Wellcome Trust.
<https://wellcome.org/insights/reports/wellcome-global-monitor-covid-19/2020>