



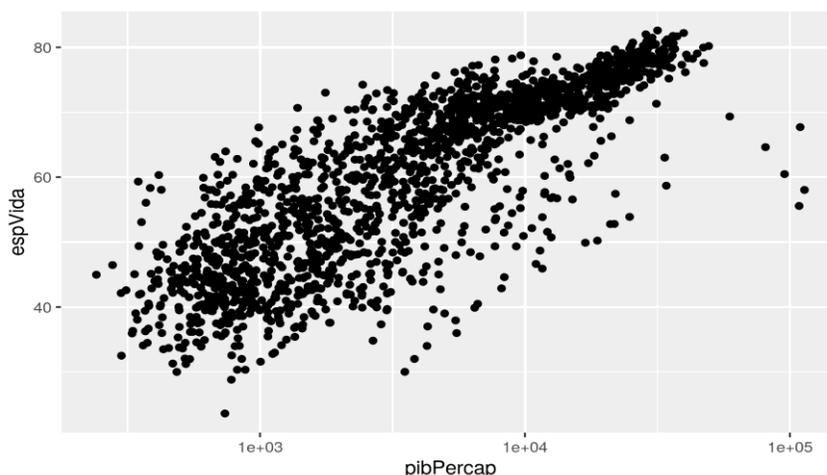
Diseñar visualizaciones de datos para informes estadísticos

Una de las habilidades de comunicación que se espera que llegues a dominar como estudiante de Estadística es la elaboración de **gráficos** para **informes que reportan resultados de análisis estadísticos**. A diferencia de las visualizaciones que creamos para explorar nuestros datos, aquellas que son parte de un informe cumplen un rol clave, ya que nos permiten respaldar y validar tanto el análisis como sus conclusiones. Para cumplir con ese objetivo, una visualización de datos debe tener una serie de características que le permiten cumplir su propósito de manera efectiva y eficiente.

En este recurso, conocerás **estrategias para producir visualizaciones de datos para informes estadísticos**. Para ello, analizaremos una serie de ejemplos y evaluaremos cómo mejorarlos. Al final de este documento, encontrarás un enlace con el código para reproducir los gráficos que hemos incluido.

I. ¿Qué representan los datos que estamos visualizando?

Imagina que para un informe estadístico que estás elaborando necesitas presentar un gráfico que muestre la relación entre el PIB per cápita (en dólares) y la esperanza de vida para datos a nivel mundial que van de 1952 a 2007. Para ello, decides hacer un diagrama de dispersión (o *scatterplot*), que es un tipo de visualización que permite mostrar los valores de dos variables. El gráfico que te entrega por defecto el programa estadístico con el que estás trabajando es el siguiente:



El gráfico cumple el propósito de mostrar la relación que te interesa y es posible que te parezca que está listo para ser incluido en tu informe. Pero ¿qué tan claro será para otras personas?

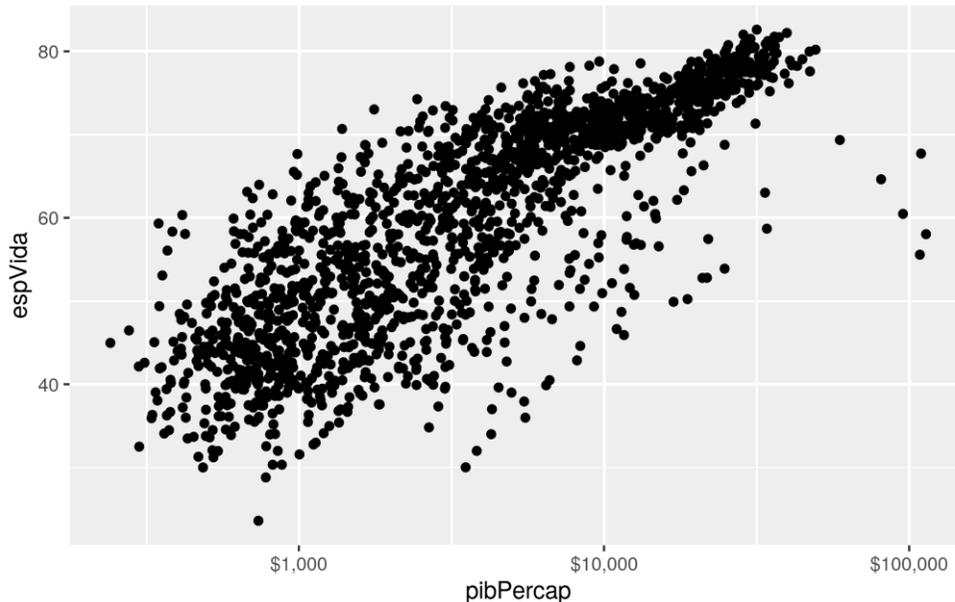
Los *scatterplots* son gráficos muy comunes para alguien que estudia o trabaja en Estadística, pero ¿qué pasa si quienes leerán tu informe provienen de otras áreas?

Vuelve a mirar el gráfico y reflexiona sobre las siguientes preguntas:

- ¿Qué elementos del gráfico podrían resultar poco claros para alguien de otra área o para una persona no familiarizada con un diagrama de dispersión?
- ¿Qué aspectos podrían ser difíciles de interpretar para alguien que está hojeando tu informe y se encuentra con esta visualización? ¿Sabemos inmediatamente qué es lo que se está representando en ella?

¿Qué podríamos editar?

Uno de los elementos que probablemente te haya parecido que podía ser poco claro para alguien de otra área es la notación utilizada en la variable “pibPercap”. Puede que para ti sea evidente qué significa $1e+03$, pero para alguien que viene de otra disciplina eso podría ser un enigma. Uno de los primeros cambios que podríamos hacer, entonces, es editar la forma en que se representan los valores de esa variable en nuestro gráfico:

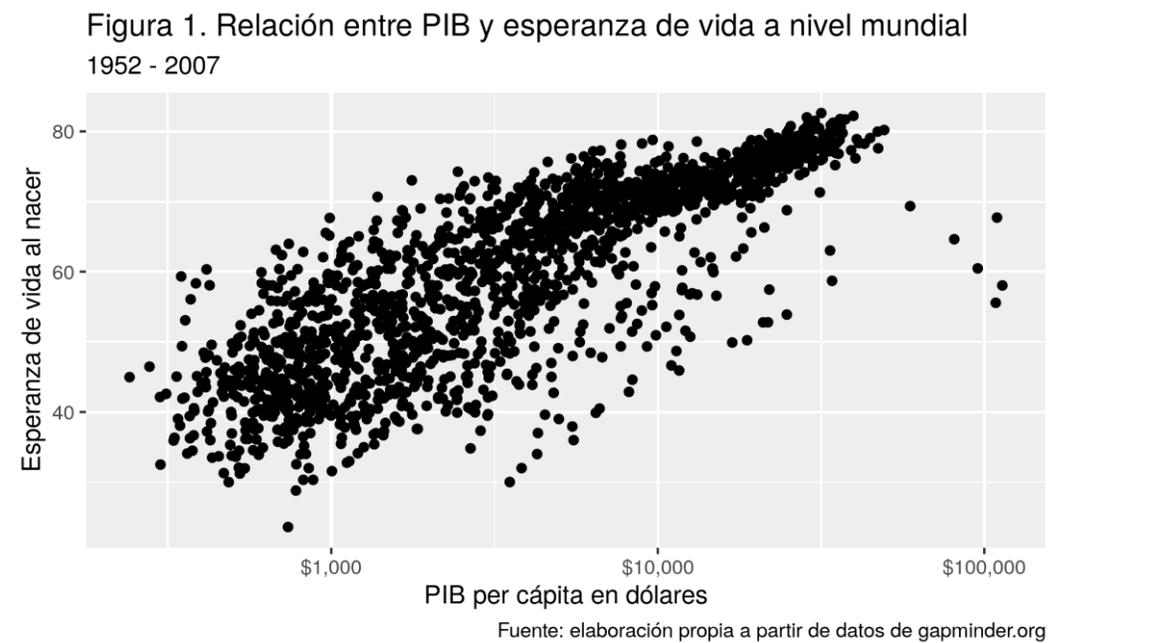


Cuando estamos explorando nuestros datos solemos generar muchos gráficos que nos sirven para entender el comportamiento de las variables con las que trabajamos. En esos casos, lo que nos ofrece por defecto el programa estadístico suele ser suficiente. Sin embargo, incluso si tu informe será leído por personas de tu área, es preferible editar los valores. Muchas veces las figuras son lo primero que miramos, ya que nos permiten

hacernos una idea acerca de qué trata el informe. Es por eso que es importante tratar de que toda la información contenida en él pueda ser rápidamente comprensible.

Probablemente también notaste que si nos encontramos con este gráfico mientras ojeamos un documento sería difícil saber qué es lo que está representando: no tiene título y los ejes tienen los nombres de las variables tal como aparecen en el set de datos. Quizás `pibPerCap` podría llegar a ser entendible para alguien que mira el informe por primera vez. Pero ¿qué pasa con `espVida`? Además, ¿es claro a primera vista cuál es el período que abarca la muestra?

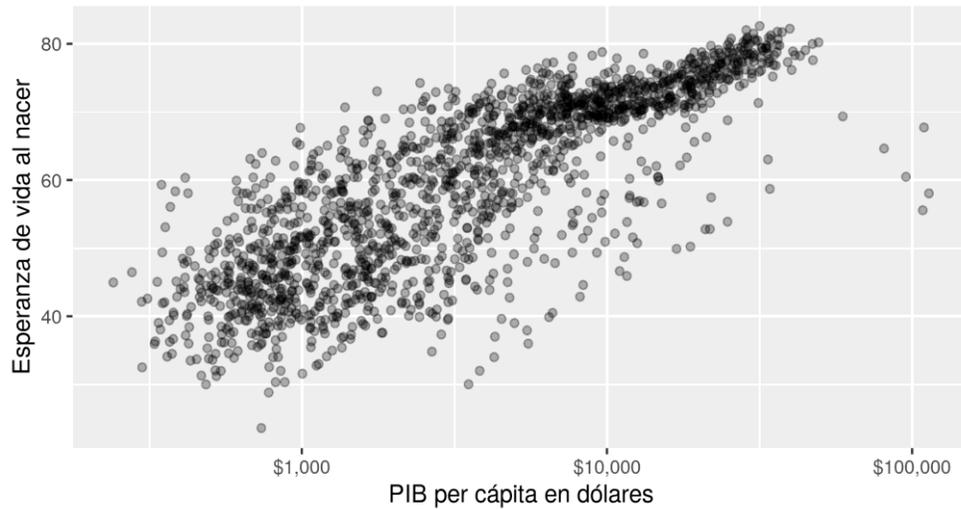
En un informe los gráficos deben llevar título y etiquetas en los ejes que orienten a quienes leen sobre qué es lo que se está visualizando, tal como en la Figura 1, que corresponde a una nueva versión de nuestro *scatterplot* inicial. En ella, además, agregamos información relativa a la fuente de los datos.



Nuestro gráfico pareciera estar listo para poder ser incluido en el informe: tiene un título en el que queda claro qué es lo que se está visualizando, etiquetas en los ejes que permiten interpretar con facilidad los datos y valores en un formato legible para cualquier persona. Pero ¿qué pasa con los puntos? ¿Hay algún elemento que podría dificultar la interpretación de cómo se distribuyen?

Como nuestro gráfico contiene muchas observaciones, hay zonas en la que se podría estar produciendo *overplotting*, es decir, en las que tenemos muchas observaciones superpuestas. Esto hace que a simple vista no sea fácil distinguir las zonas en las que se concentra mayor cantidad de puntos. Una manera de resolver esto es regulando el nivel de transparencia de los puntos, tal como se muestra en esta nueva versión de la Figura 1:

Figura 1. Relación entre PIB y esperanza de vida a nivel mundial
1952 - 2007

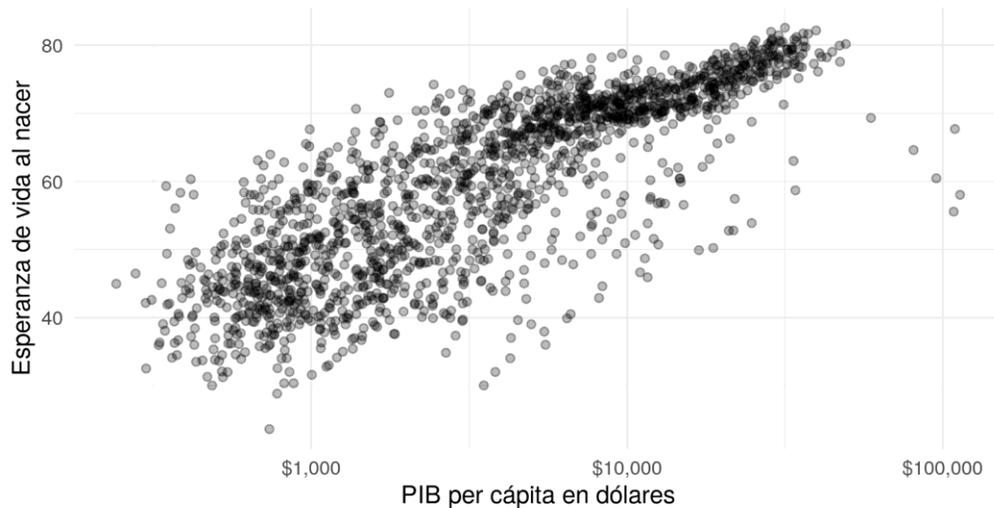


Fuente: elaboración propia a partir de datos de gapminder.org

Gracias a la modificación que hicimos en el nivel de transparencia de los puntos, es posible advertir con mayor facilidad las zonas del gráfico en las que hay mayor densidad de datos. Si comparas este gráfico con la primera versión que hicimos, notarás que ahora es posible entender con mayor claridad cuáles son las variables que se están visualizando.

Una última modificación que podríamos hacer es editar el color del fondo. ¿Por qué crees que se hace esto? Es posible que quienes lean tu informe lo hagan en formato impreso. Al dejar el fondo blanco hacemos que sea más fácil distinguir los puntos que conforman el diagrama de dispersión.

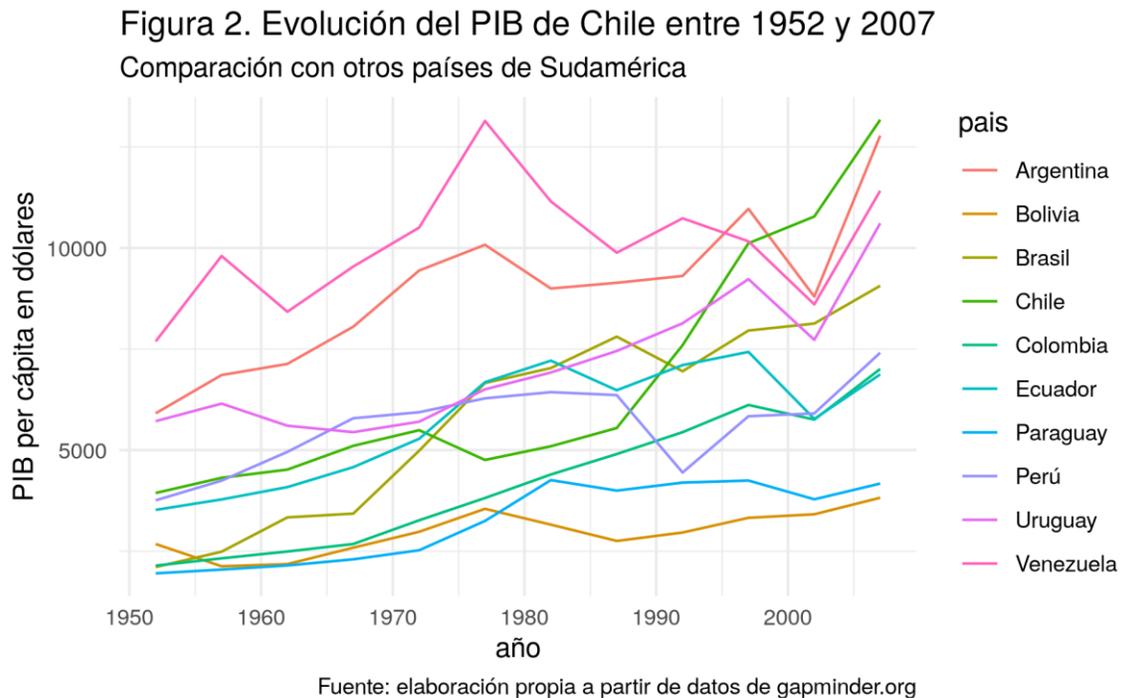
Figura 1. Relación entre PIB y esperanza de vida a nivel mundial
1952 - 2007



Fuente: elaboración propia a partir de datos de gapminder.org

2. Enfocar la atención

En la Figura 2 hemos aplicado algunos de los principios que revisamos anteriormente, como la inclusión de un título explicativo y de etiquetas y formatos de valores comprensibles para un público amplio.

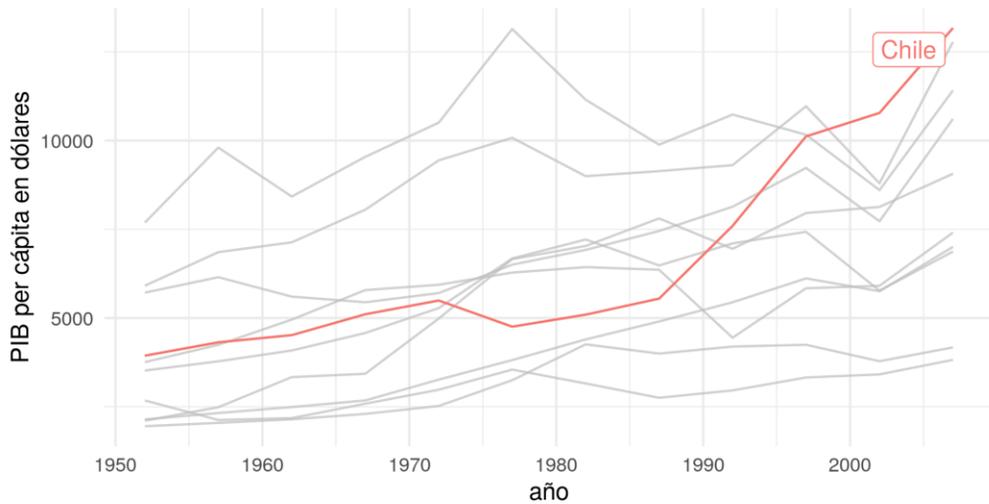


Como se indica en el título, el objetivo de este gráfico es comparar la evolución del PIB de Chile con el de otros países de Sudamérica. ¿Hay algún aspecto de la forma en que se visualizan los datos que podría hacer difícil establecer esta comparación?

Probablemente hayas notado que no es sencillo distinguir cuál es la línea que representa a Chile. El gráfico que entrega por defecto el programa utilizado le asigna la misma importancia a todos los datos; sin embargo, el propósito con el que lo hemos construido es establecer una comparación. Es por ello que deberíamos diseñar nuestro gráfico para que quienes lean puedan enfocar su atención en ello.

A continuación, encontrarás una nueva versión de la Figura 2. En ella hemos modificado los colores y agregamos una etiqueta para indicar con mayor claridad cuál es la línea que corresponde a Chile. Estos dos elementos permiten orientar la atención precisamente en aquello indicado en el título del gráfico: la evolución del PIB de nuestro país.

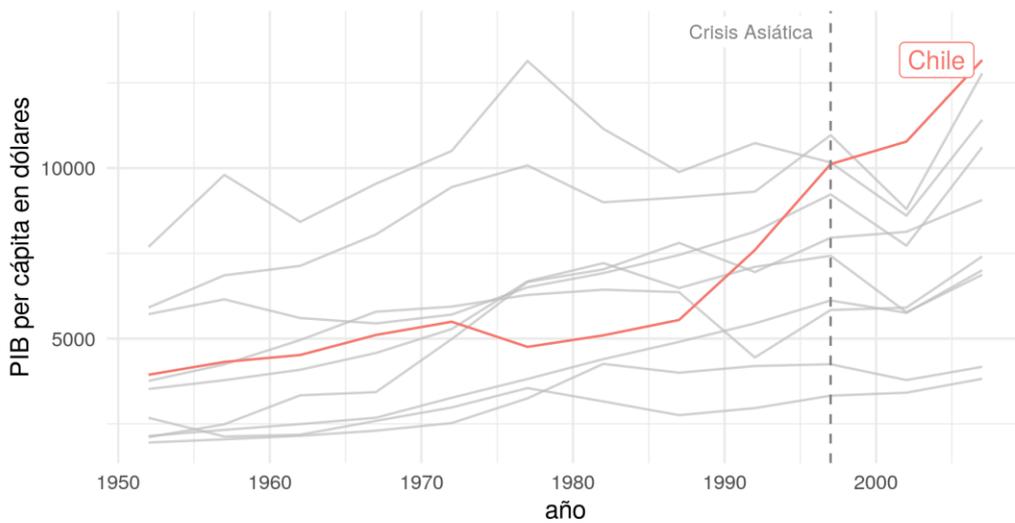
Figura 2. Evolución del PIB de Chile entre 1952 y 2007
Comparación con otros países de Sudamérica



Fuente: elaboración propia a partir de datos de gapminder.org

Las anotaciones como las que acabamos de hacer con el nombre del país son muy útiles cuando queremos guiar la atención hacia un elemento de nuestro gráfico. No solo podemos utilizar este recurso para hacer que quienes leen identifiquen rápidamente una variable o una observación, sino también como ayuda para interpretar los resultados. Por ejemplo, si en tu informe hicieras alusión al efecto que tuvo la Crisis Asiática en la región a partir de 1997, sería necesario estimar en el eje dónde está ese valor para poder llegar a esa misma conclusión. ¿Por qué mejor no anotar nuestro gráfico con esa información como en esta nueva versión del gráfico?

Figura 2. Evolución del PIB de Chile entre 1952 y 2007
Comparación con otros países de Sudamérica



Fuente: elaboración propia a partir de datos de gapminder.org

Como ves, las anotaciones no solo sirven para ayudar a quienes leen a enfocarse en los aspectos clave de los datos que estamos visualizando al mirar el gráfico por primera vez, sino que también son un apoyo para las explicaciones e interpretaciones que elaboramos en el texto de nuestro informe.

3. Explicar lo que estamos viendo

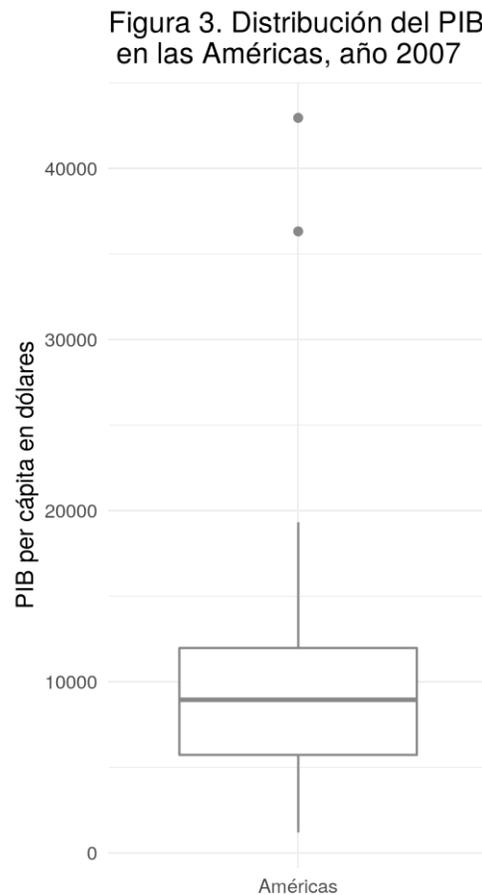
Tal como lo indica su título, la Figura 3 muestra un diagrama de caja o *boxplot* de la variable PIB per cápita para un subconjunto de nuestro datos: las Américas en el año 2007. Este gráfico sería muy útil para mostrar que los datos tienen una distribución simétrica y que hay dos *outliers* (Estados Unidos y Canadá).

Si bien esa conclusión te puede parecer muy obvia porque a lo largo de la carrera has tenido que interpretar y elaborar muchos diagramas de caja, ¿qué crees que pasaría con una persona que no está familiarizada con este tipo de gráfico? ¿Podría llegar fácilmente a esas mismas conclusiones por sí sola? Si tus datos tuviesen una distribución asimétrica, ¿qué crees que se imaginaría alguien que nunca ha visto un *boxplot* si mencionas que los datos están sesgados hacia la derecha?

El uso de gráficos de barra o de línea está muy extendido, por lo que es poco frecuente que en un informe tengamos que explicar cómo interpretarlos, incluso si es para un público amplio. Sin embargo, hay gráficos que no están tan extendidos, por lo que si crees que quienes leerán tu informe podrían no conocer alguna de las formas de visualización que has elegido, lo mejor sería entregar algunas indicaciones sobre cómo interpretarlos.

Por ejemplo, si el informe que incluye este *boxplot* va a leer un público no tan especializado, podríamos explicar la figura así:

Tal como se observa en la Figura 3, el PIB per cápita de los países de América del Norte, Centro y Sur presenta dos casos atípicos, representados por los puntos en el gráfico: Estados Unidos, con un PIB de 43000 dólares, y Canadá, con uno de 36000. La mediana, que corresponde a la línea horizontal que atraviesa la caja, la divide en dos mitades, lo que indica que la distribución de los datos es simétrica.



Fuente: elaboración propia a partir de datos de gapminder.org

Además de los diagramas de caja, ¿qué otros tipos de visualizaciones crees que no serían tan fáciles de interpretar para alguien que no es de Estadística?

Accesibilidad

Al diseñar una visualización de datos es importante considerar también que nuestros datos tienen que ser accesibles. Describir lo que se observa en un gráfico, como hicimos con la Figura 3, no solo ayuda a entender mejor el *boxplot* para quienes no están familiarizados con ese tipo de visualización, sino que es un apoyo para personas con algún tipo de discapacidad visual.

Existen herramientas que permiten simular cómo una persona daltónica vería nuestros documentos e imágenes. A continuación, te recomendamos dos:

- [Let's get color blind](#) es una extensión de Firefox que te permite simular cómo personas con distintos tipos de daltonismo ven una determinada imagen.
- [Color Oracle](#) es una aplicación disponible para Windows, Mac y Linux que cumple el mismo propósito.

Prueba alguna de estas aplicaciones y fíjate qué pasa con la primera versión de la Figura 2. ¿Es posible distinguir todas las líneas de colores? ¿Para qué tipo de daltonismo podría ser difícil distinguir la línea roja de la segunda y tercera versión de esa figura? Al final de este documento encontrarás el código para reproducir los gráficos. Explora qué combinaciones de colores podrían resolver estos problemas.

¡Ahora estás listo/lista para incluir visualizaciones en tus informes estadístico!

Recuerda

- ✓ Las visualizaciones de datos son un elemento clave de cualquier informe estadístico.
- ✓ Ayuda a tus lectores a enfocarse en los elementos más importantes a través del uso del color o de anotaciones.
- ✓ Dependiendo del nivel de conocimiento de tu audiencia, puede que sea necesario explicar cómo interpretar los datos.
- ✓ Considera que entre quienes leerán tu informe puede haber personas con daltonismo o con algún tipo de discapacidad visual. Chequear los colores y describir los datos son dos estrategias para hacer sus visualizaciones accesibles.

Sobre este material

Autora: Riva Quiroga

Edición: Patricia Lobato Concha

Revisión de recurso desde teoría basada en MEDA*: Verónica Sánchez Gibbons

Revisión y edición final: Claudia Avendaño y Karen Urrejola

*PRAC se basa en la metodología de Materiales Educativos de Aprendizajes (MEDA) para la sistematización y presentación del contenido de sus recursos.

Puedes descargar el código para reproducir los gráficos de este recurso en [este enlace](#).

Puedes encontrar más recursos como este en el sitio web del PRAC



PROGRAMA DE APOYO A LA
COMUNICACIÓN ACADÉMICA

