

Herramientas Computacionales para la investigación

Tarea N° 4 - Data Visualization

Maria Emilia Bullano y Gastón García Zavaleta

14 de julio 2021

Introducción

En este trabajo se corrigen una serie de gráficos realizados en clase y en prácticos anteriores a partir de las sugerencias encontradas en Schwabish (2014), que se pueden resumir en los siguientes tres principios básicos:

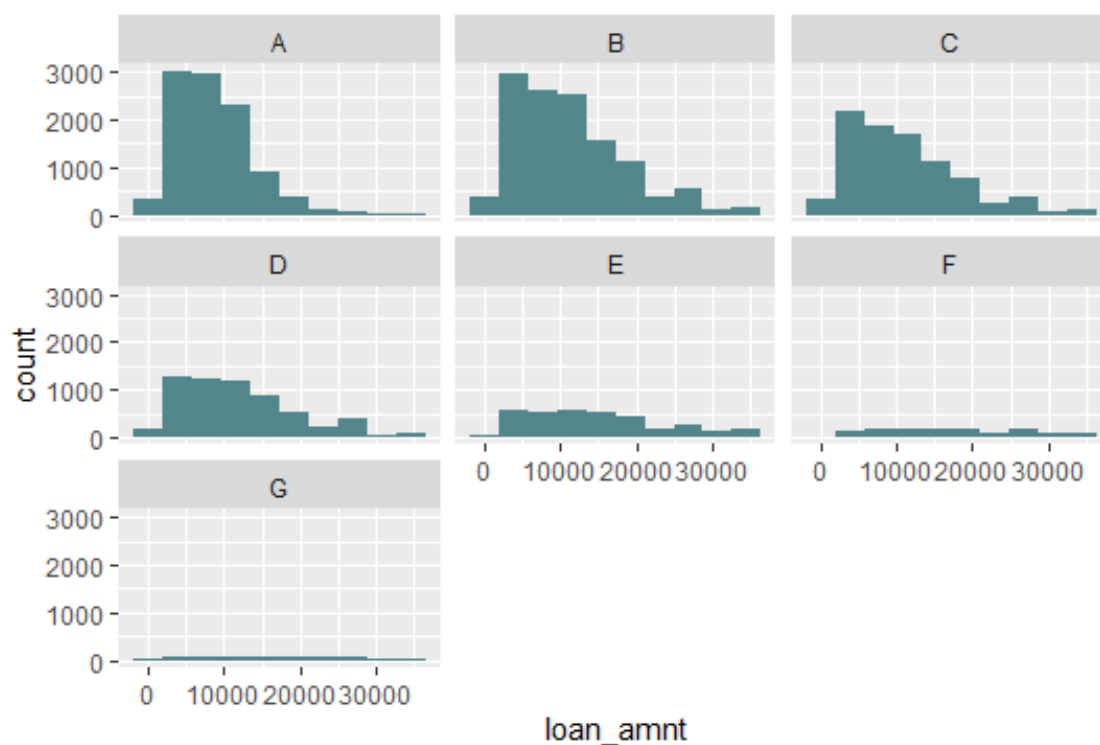
1. Mostrar los datos. Son la parte más importante de un gráfico y deben exponerse de manera clara.
2. Minimizar la confusión. Se debe evitar el uso de elementos visuales innecarios, que puedan desviar la atención de lo verdaderamente importante.
3. Integrar el gráfico y el texto. Los gráficos deben incluir información suficiente para ser autocontenidos y funcionar como un complemento del texto.

Corrección de gráficos

1. Cantidad de clientes según monto de préstamo por categoría.

En esta primera sección se corrige la Figura 1 - realizada en clase - que busca mostrar la cantidad de clientes de cada categoría que reciben préstamos de un determinado monto. Esta figura consiste de 7 histogramas, uno para cada categoría.

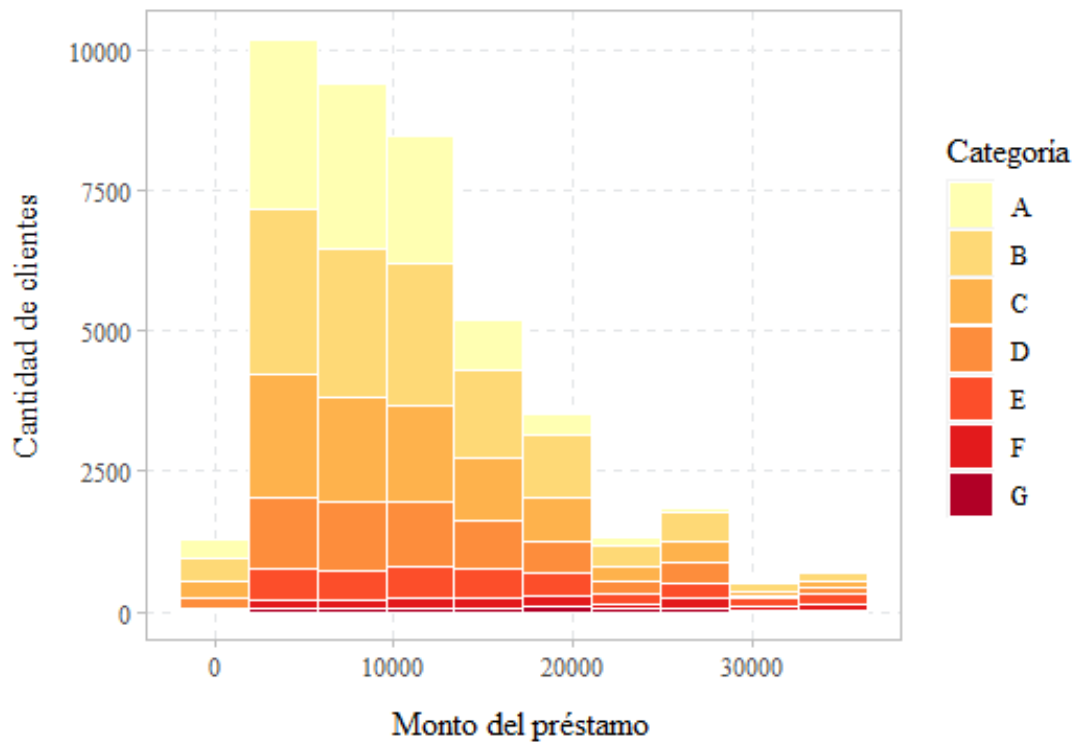
Figura 1: Cantidad de clientes según monto de préstamo por categoría.



En términos generales, consideramos que este gráfico cumple con el principio 2 de Schwabish (2014) porque es un gráfico simple, sin elementos visuales innecesarios. Sin embargo, no cumple los principios 1 y 3, debido a que no muestra adecuadamente los datos - por ejemplo, para los grupos E, F y G resulta difícil visualizar la distribución por una cuestión de escala - y contiene leyendas confusas que obligan al lector a recurrir al texto para entender la información presentada.

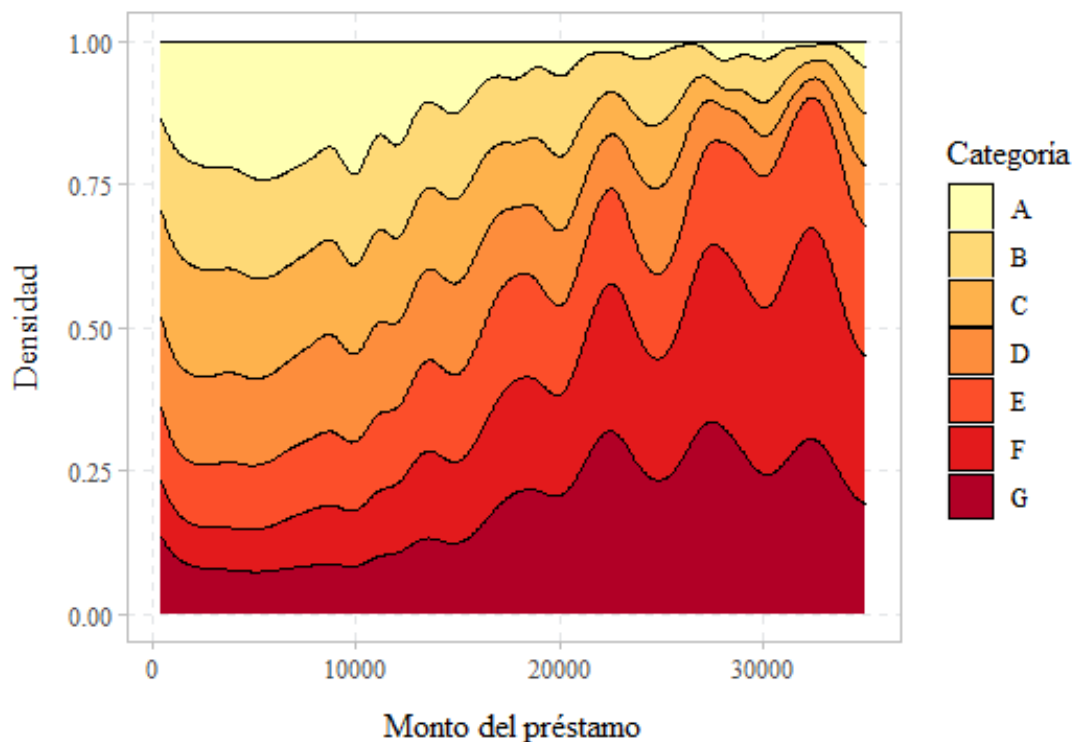
Con el objetivo de solucionar estos problemas se proponen dos gráficos alternativos. El primero (Figura 2) es un gráfico de barras que muestra, para distintos niveles de monto del préstamo, la cantidad de clientes existentes de cada categoría.

Figura 2: Cantidad de clientes según monto de préstamo.



Este gráfico resulta ser visualmente más atractivo y resume la información de los 7 histogramas previos en 1 solo. Se solucionan los problemas asociados al principio 1, pero no es tan claro que se solucionen los asociados al principio 3. Nuevamente, resulta difícil observar la distribución de los clientes pertenecientes a las categorías E, F y G. Por este motivo, se construye otro gráfico (Figura 3), que soluciona este problema.

Figura 3: Proporción de clientes según monto de préstamo.

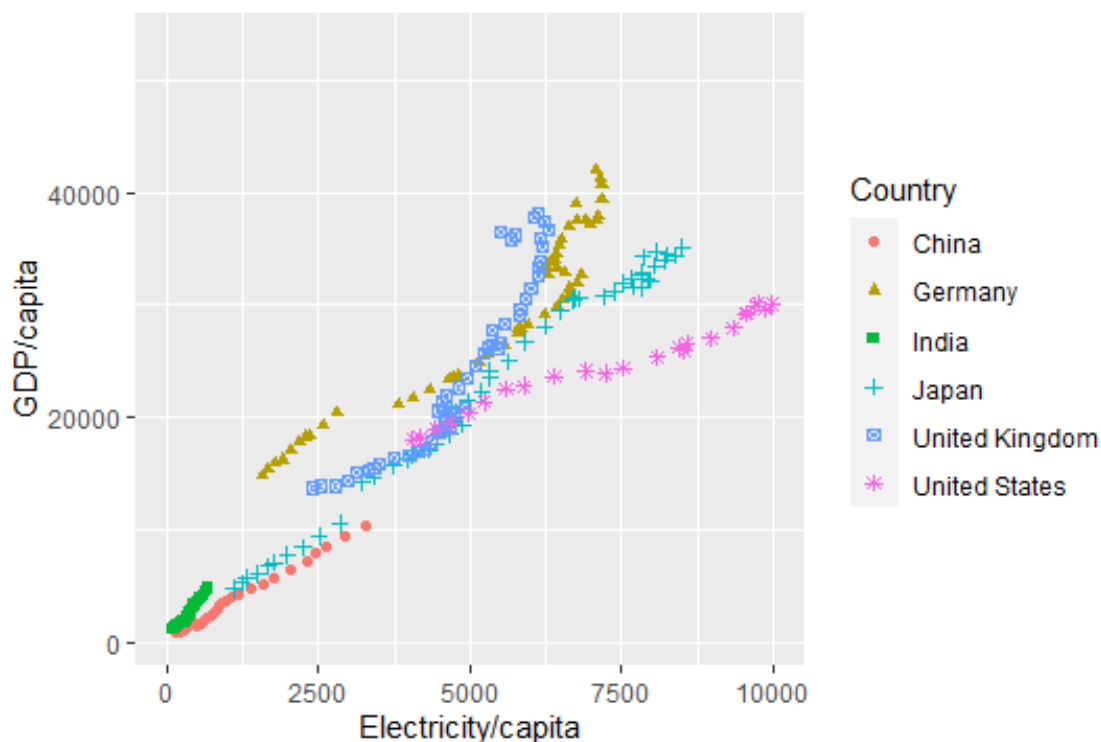


Resulta ahora sencillo observar la distribución de los clientes pertenecientes a las mencionadas categorías. El costo de esta mejora está en que se pierde información respecto a la proporción de clientes existentes de cada categoría. Por lo tanto, la conveniencia de utilizar la Figura 2 o la Figura 3 dependerá de cual sea la información que se prioriza presentar.

2. Relación entre producto y consumo de electricidad per cápita.

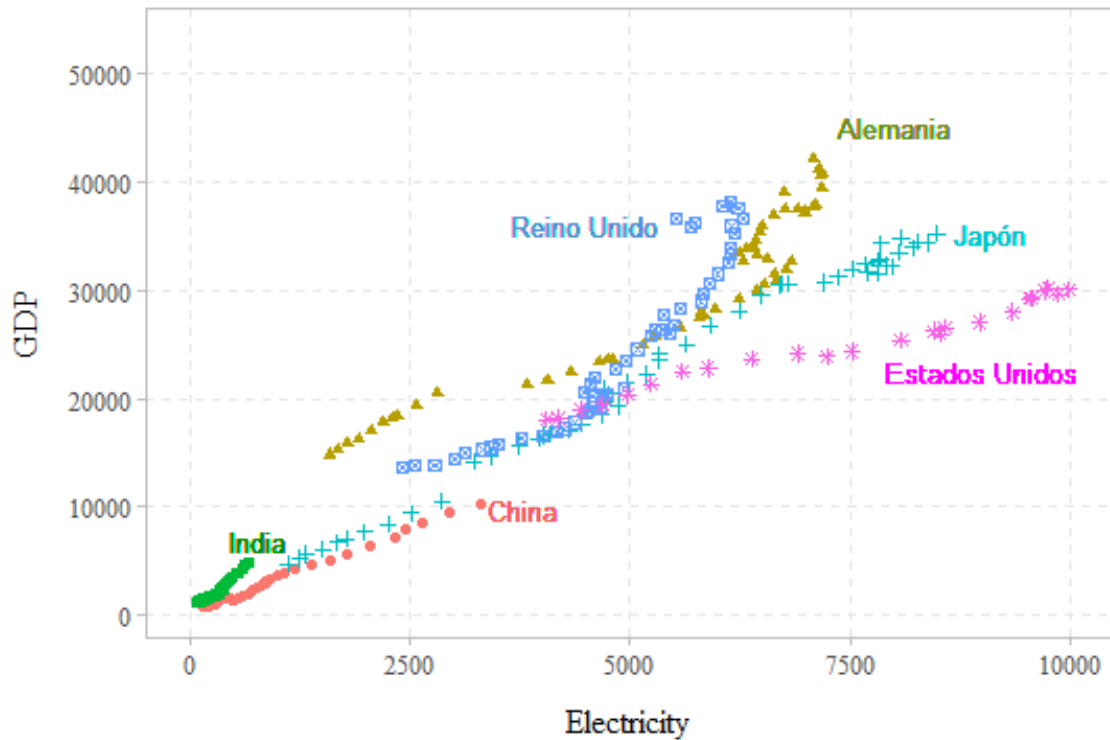
En esta sección se corrige la Figura 4 - realizada en clase - que busca mostrar la relación entre el consumo de electricidad per cápita y el producto per cápita para distintos países.

Figura 4: Relación entre producto y consumo de electricidad per cápita por país.



Consideramos que este gráfico cumple en general con los principios mencionados. Sin embargo, resulta incómodo que las etiquetas se ubiquen por fuera del mismo. Como consecuencia, decidimos realizar la Figura 5, en la cual no solo buscamos mejorar la estética sino que además incorporamos las etiquetas al lado de sus correspondientes series, de modo de facilitarle la tarea al lector. Además, eliminamos información de las leyendas de los ejes que resulta redundante dado el título (/cápita).

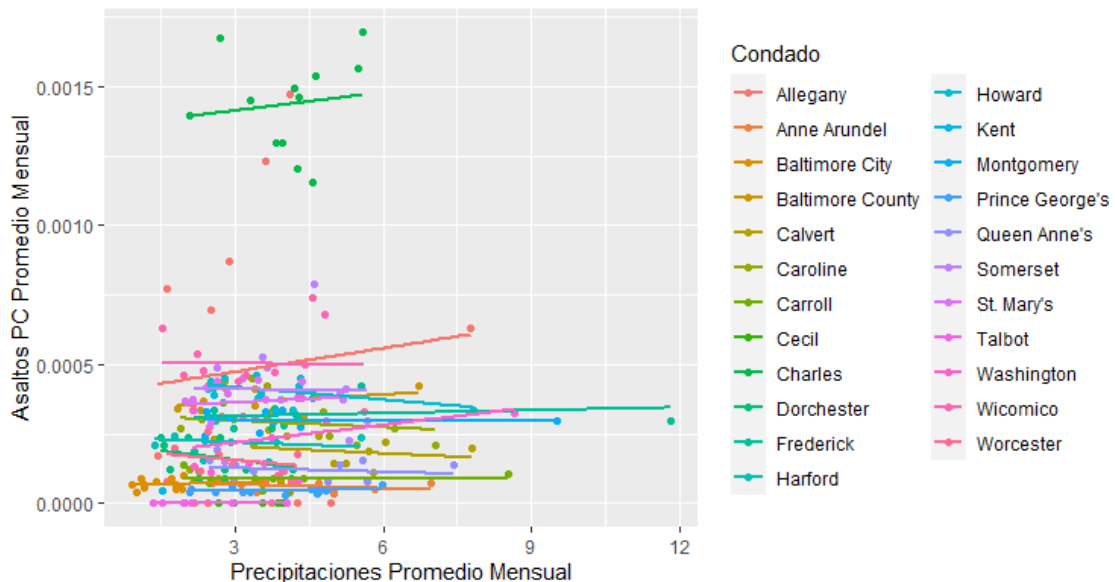
Figura 5: Relación entre producto y consumo de electricidad per cápita por país.



3. Correlación entre cantidad de asaltos y precipitaciones.

En esta sección se corrige la Figura 6 - perteneciente a la entrega de la Tarea 3 de esta materia - que busca mostrar la relación entre la cantidad de asaltos per cápita y las precipitaciones para cada condado del estado de Maryland, Estados Unidos.

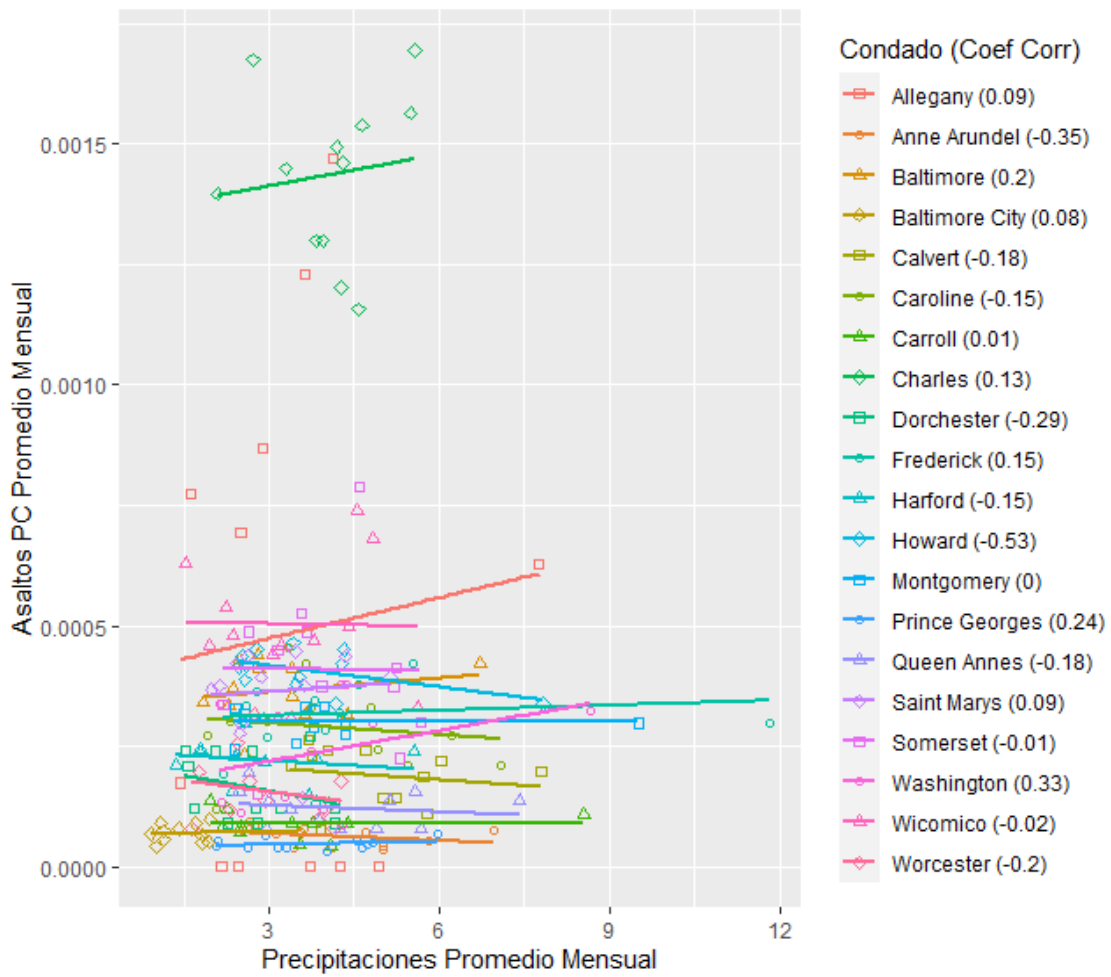
Figura 6: Correlación entre asaltos per cápita y precipitaciones promedio por condado.



Consideramos que esta figura no respeta el segundo principio porque resulta visualmente poco atractiva y confusa. Debido a que el objetivo de este gráfico era mostrar los coeficientes de correlación entre las variables a nivel condado, decidimos construir la Figura 7, en

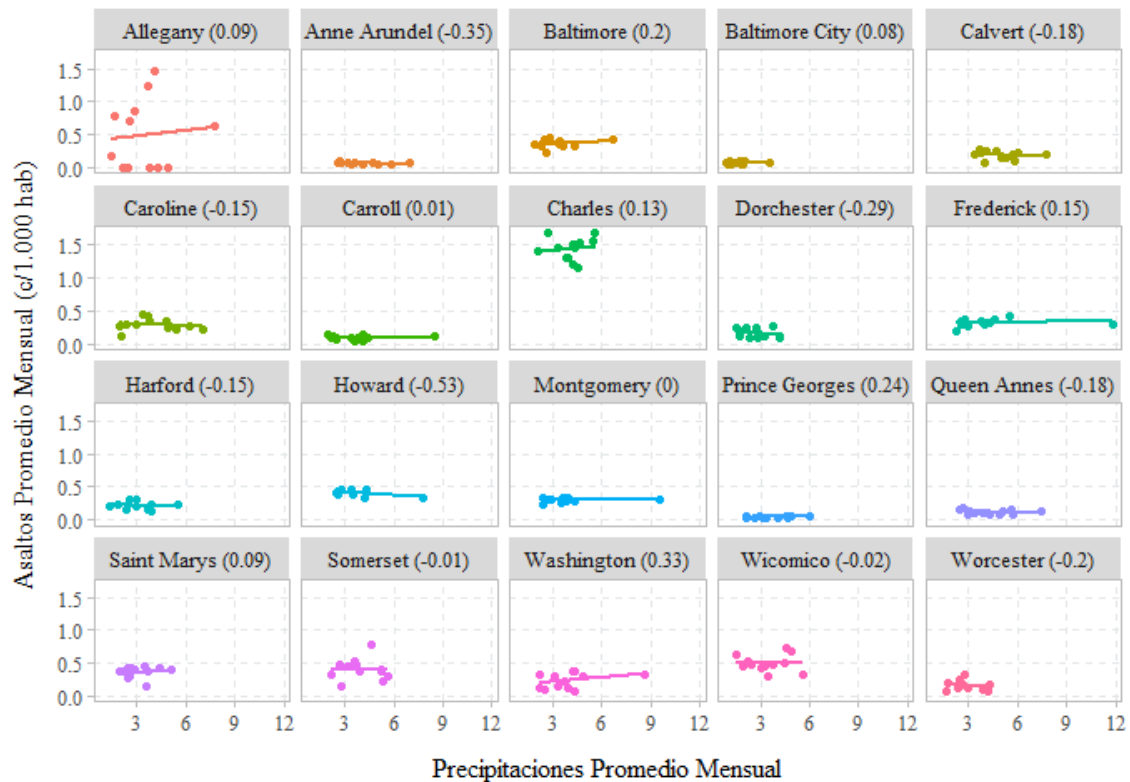
la cual incorporamos distintas formas a las observaciones de cada condado para hacerlas más distinguibles y mostramos el coeficiente de correlación de cada condado al lado de su etiqueta.

Figura 7: Correlación entre asaltos per cápita y precipitaciones promedio por condado.



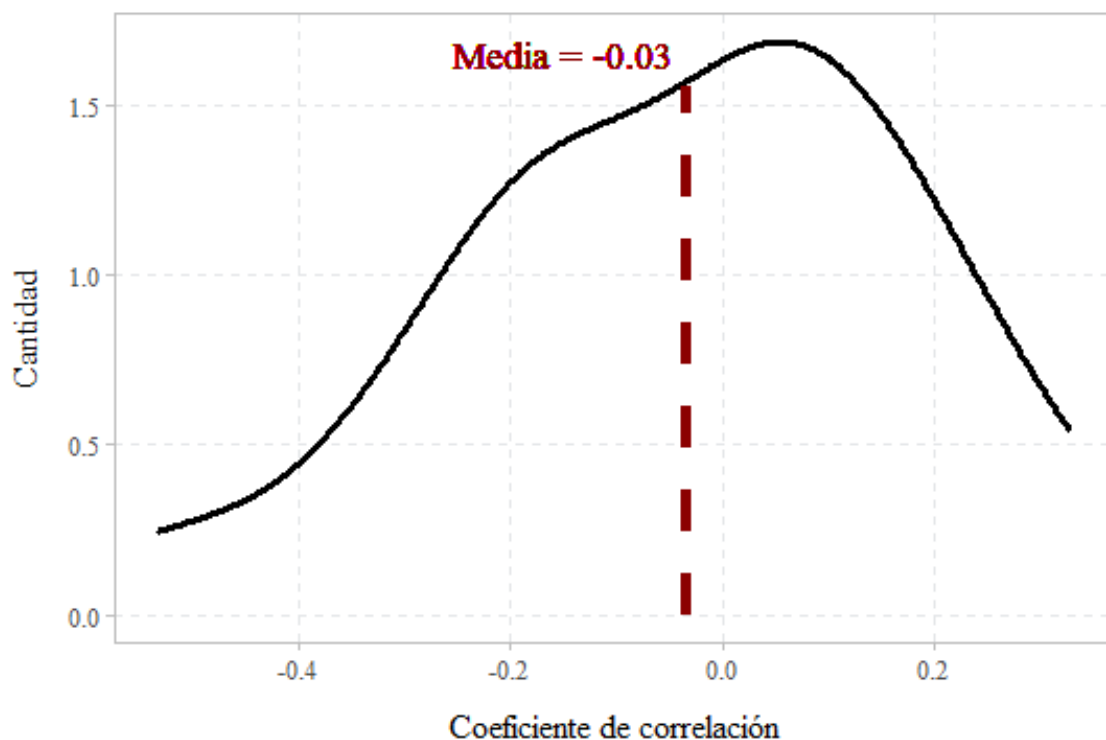
Consideramos que este gráfico sigue siendo confuso, principalmente porque la concentración de observaciones en el extremo inferior izquierdo dificulta la identificación de las distintas rectas. Por este motivo proponemos un gráfico distinto (Figura 8), en el cual se genera un diagrama de dispersión para cada uno de los condados y se muestra la cantidad de asaltos cada 1000 habitantes como variable, para facilitar la interpretación de los resultados.

Figura 8: Correlación entre asaltos per cápita y precipitaciones promedio por condado.



Este gráfico resuelve los problemas de la Figura 7. Sin embargo, dado que el principal objetivo era mostrar que el coeficiente de correlación entre las variables es cercano a cero para todos los condados, decidimos crear la Figura 9, que resume la distribución de los coeficientes de correlación para los distintos condados.

Figura 9: Correlación entre asaltos per cápita y precipitaciones promedio por condado.



Es posible observar que los coeficientes de correlación se ubican en valores cercanos al cero para todos los condados, con una media de -0.03 y una baja dispersión. Este gráfico presenta la información relevante de una forma sencilla, pero no permite mostrar información acerca de como es este coeficiente en cada condado.