



Graficas que grafiquen... ¡¡¡por favor!!!

Las representaciones graficas permiten representar, como su nombre lo indica, gráficamente la información del comportamiento de una variable. La ventaja que tienen por sobre las medidas de resumen (media, varianza, y percentiles entre otras) es su impacto visual. Un gráfico construido de manera correcta puede ser un gran aporte a cualquier nota periodística que la amerite ya que la información contenida en él puede ser (debe poder ser) analizada de manera intuitiva por el lector.

Existen distintos tipos de gráficos y la selección de uno el otro dependerá, en principio, del tipo de variable que se está representando

Tipos de variables

Las variables se dividen en dos grupos:

- Cualitativas o categóricas
- Cuantitativas

Las variables cualitativas están relacionadas a características de los individuos bajo estudio y los valores que estas toman corresponden a categorías por esta razón es que también reciben el nombre de categóricas.

A modo de ejemplo consideremos la base de datos de la Encuesta Nacional del Empleo (ENE) (Fuente:

<http://www.ine.cl/docs/default-source/laborales/ene/antecedentes-metodologicos/enfoque-estad%C3%ADstico-lanzamiento-de-la-nueva-encuesta-nacional-de-empleo.pdf?sfvrsn=4>)

Dentro de la información que se releva en este estudio esta región y comuna donde habita el encuestado, sexo, relación con el jefe de hogar, nivel educacional más alto aprobado, nacionalidad entre otras. Todos estas corresponden a variables de tipo cualitativas.

Por ejemplo, para la variable relación con el jefe de hogar las posibles respuestas o categorías de respuesta (Nombre del código) son:

Variable	Descripción de la variable	Cód.	Nombre del código
parentesco	Identifica el tipo de parentesco de cada miembro del hogar respecto al jefe(a) de hogar	1	Jefe(a) de Hogar
		2	Cónyuge
		3	Conviviente
		4	Hijo(a) o Hijastro(a)
		5	Yerno o Nuera
		6	Nieto(a)
		7	Hermano(a) o Cuñado(a)
		8	Padres o Suegros
		9	Otro Pariente
		10	No Pariente
		11	Servicio Doméstico
		12	Fuera del Hogar
		13	Fallecido

Para la variable nivel educacional más alto aprobado, las posibles respuestas o categorías de respuestas son:

Variable	Descripción de la variable	Cód.	Nombre del código
nivel	Identifica el nivel educacional más alto aprobado, agrupando a las personas que terminaron el nivel, así como las que no lo terminaron	0	Nunca estudió
		1	Sala cuna
		2	Kínder
		3	Básica o primaria
		4	Media común
		5	Media técnico profesional
		6	Humanidades
		7	Centro formación técnica
		8	Instituto profesional
		9	Universitario
		10	Post títulos
		11	Magíster
		12	Doctorado
		14	Normalista
999	Nivel Ignorado		

Observe que en este caso las posibles respuestas se pueden ordenar ya que sabemos, por ejemplo, que una persona que completó la educación media esta en un nivel de instrucción más alto que aquella que solo completó la enseñanza básica.

Si las posibles respuestas de una variable cualitativa se pueden ordenar diremos que esta es una variable cualitativa ordinal. De lo contrario diremos que es nominal.

Las variables cuantitativas son variables numéricas. Ejemplo de variables cuantitativas que se relevan en la Encuesta Nacional de Empleo son: edad y horas habituales de trabajo.

Las variables cuantitativas a su vez también se dividen en dos grupos:

- Variables cuantitativas discretas. Una variable cuantitativa discreta es aquella cuyo recorrido se puede contar así estemos infinitamente contando. En palabras más simples podemos decir que cuando una variable es cuantitativa discreta si su conjunto de valores posibles o recorrido queda definido, por ejemplo, por {1, 2, 3...}.

Algunos ejemplos de variables cuantitativas discretas son el número de hijos que tiene una persona, el número de semestres en que un alumno tarda en terminar su carrera, etc.

- Variables cuantitativas continuas. Una variable cuantitativa continua es aquella cuyos valores posibles o recorrido consisten en un intervalo de la recta numérica. Por ejemplo, la estatura de una persona, en metros, es una variable continua y podemos decir que esta va de 1,40 mts. a 2,00 mts.

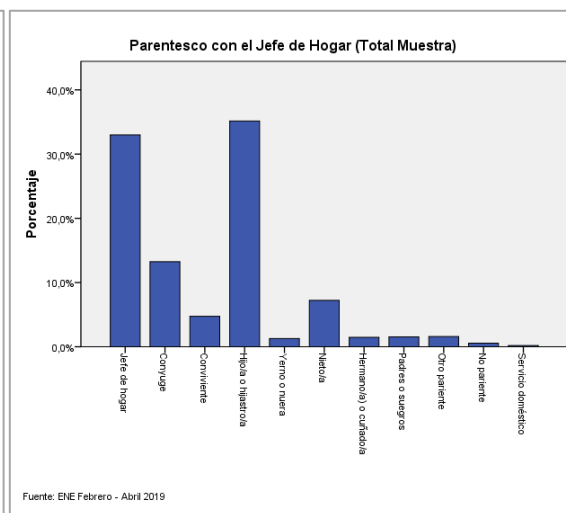
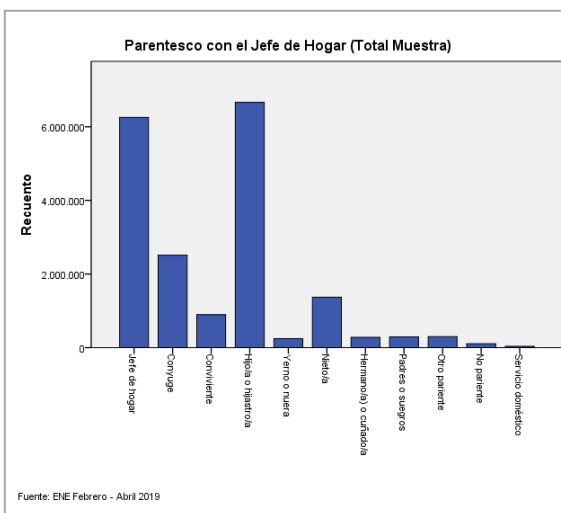
En general las variables continuas son aquellas que surgen de “medir”; el peso en kilogramos de una persona, toneladas exportadas de un determinado producto, la cantidad de lluvia caída en milímetros, etc.

De acuerdo con el tipo de variable las representaciones graficas que se recomiendan son:

Variable cualitativa nominal

- Diagrama de barras
- Diagrama de sectores circulares también conocido como el grafico de torta. Este solo se recomienda si la variable que estamos analizando tiene pocas categorías de respuesta ya que de lo contrario puede volverse ilegible.

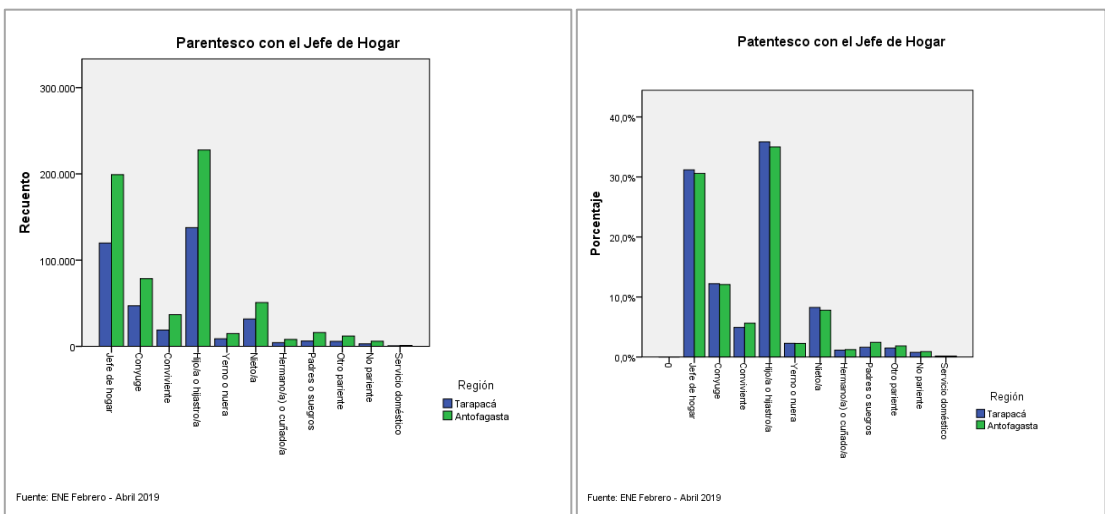
Veamos un ejemplo tomando la variable Relación con el Jefe de Hogar de la ENE.



Observe que hay algunos elementos importantes que deben ir en el gráfico de manera que para el lector sea simple entender la información que presenta. En el eje Y se especifica el dato que se esta mostrando, ya sea el recuento (o frecuencia absoluta) o importancia de cada categoría dentro del total (porcentaje).

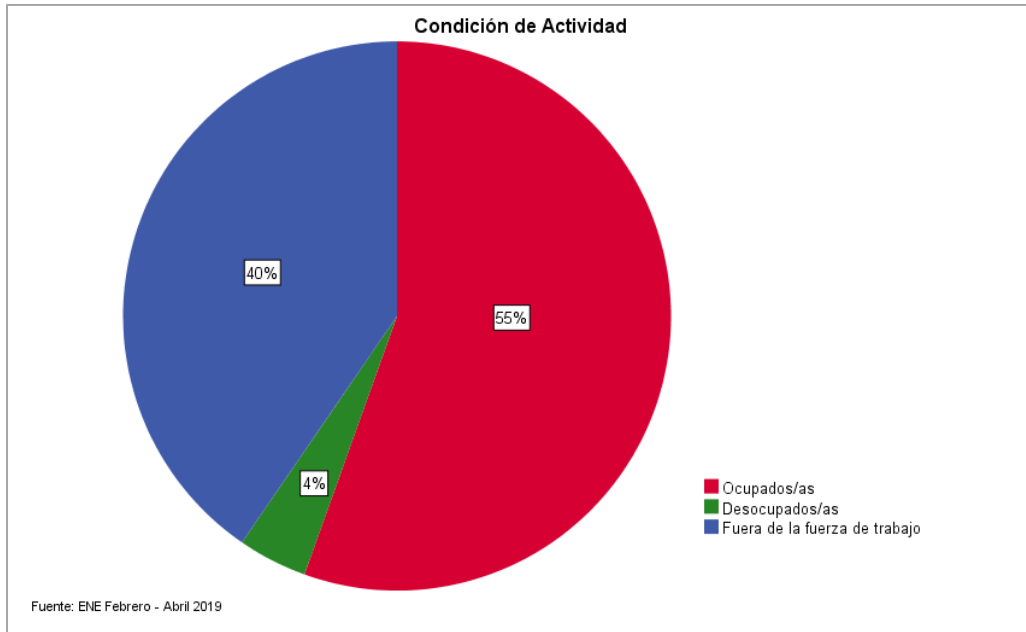
En este caso el título refleja exactamente la información que esta en el eje X por lo que el nombre de este se puede omitir. La fuente de información es otro elemento que no puede faltar ya sea dentro del grafico como en este caso, bajo él o dentro del cuerpo del artículo.

Como ya se mencionó, la única diferencia entre los gráficos que presenta es que el primero muestra el recuento o frecuencia, es decir, la cantidad de veces que ha sido mencionada una categoría, en tanto que el segundo muestra el porcentaje o la importancia de cada categoría dentro del total. Observe que la forma es exactamente la misma, por lo que en este caso podría dar lo mismo si se muestra un gráfico o el otro. Los porcentajes cobran importancia cuando estoy comparando el comportamiento de una variable dentro de dos grupos. Por ejemplo, podemos comparar el comportamiento de la variable Parentesco con Jefe de Hogar en la primera y segunda región:



Note como cambia la forma de los gráficos cuando pasamos de un representar las frecuencias a representar los porcentajes y en consecuencia lo que podemos interpretar y/o concluir de ellos. En el primero, podríamos decir que en la región de Antofagasta (columna verde) hay más hogares donde el jefe de hogar está casado (parentesco con el jefe de hogar cónyuge) que en la región de Tarapacá. Hay que tener en cuenta que el número de casos en la región Antofagasta es 650.573 en tanto que en la región de Tarapacá solo hay 383.870 casos, por lo que se podría esperar un comportamiento como el que muestra la figura. Ahora, si miramos el grafico donde se encuentran representados los porcentajes, la situación cambia absolutamente y no se observa ninguna diferencia muy relevante en cuanto al comportamiento de esta variable en ambas comunas.

Aquí aparece un nuevo elemento muy importante en las representaciones gráficas, **la leyenda**, esta es fundamental para entender que información representa cada barra o cada sector circular en el caso de que se construya un diagrama de sectores circulares como el que se muestra a continuación.



Los números en los cuadros blancos corresponden a la importancia de cada categoría dentro del total. No es requisito que estén en un gráfico de sectores circulares, pero sin duda aportan a tener una mejor comprensión de este.

Variables cualitativas ordinales.

Suelen usarse las mismas que en el caso de las variables cualitativas nominales. En el caso del diagrama de barras se debe respetar el orden intrínseco de la variable cualitativa ordinal. En lo personal, no recomiendo el uso de diagrama de sectores circulares pues en él no es fácil de observar la jerarquía de las categorías de este tipo de variables.

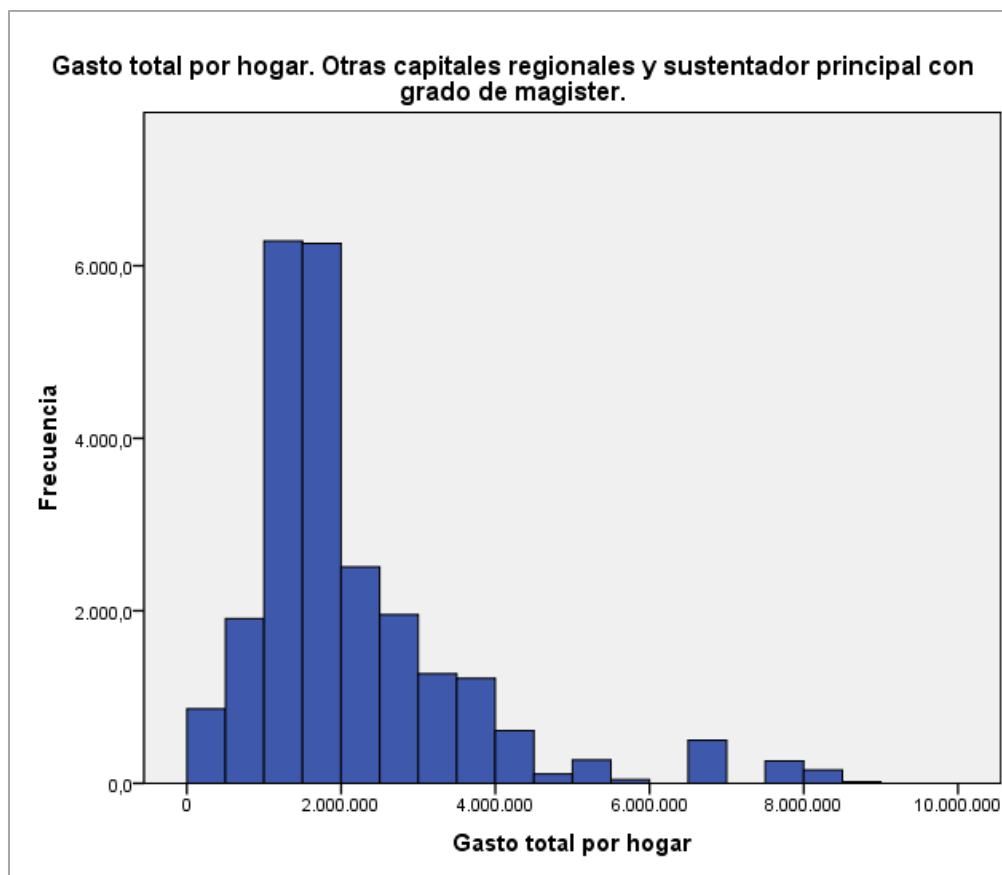
Variables cuantitativas discretas.

Si el rango de estas es pequeño, suelen usarse también los diagramas de barras. En el caso de que el rango sea grande las variables cuantitativas discretas suelen tratarse como si fueran variables cuantitativas continuas.

Variables cuantitativas continuas.

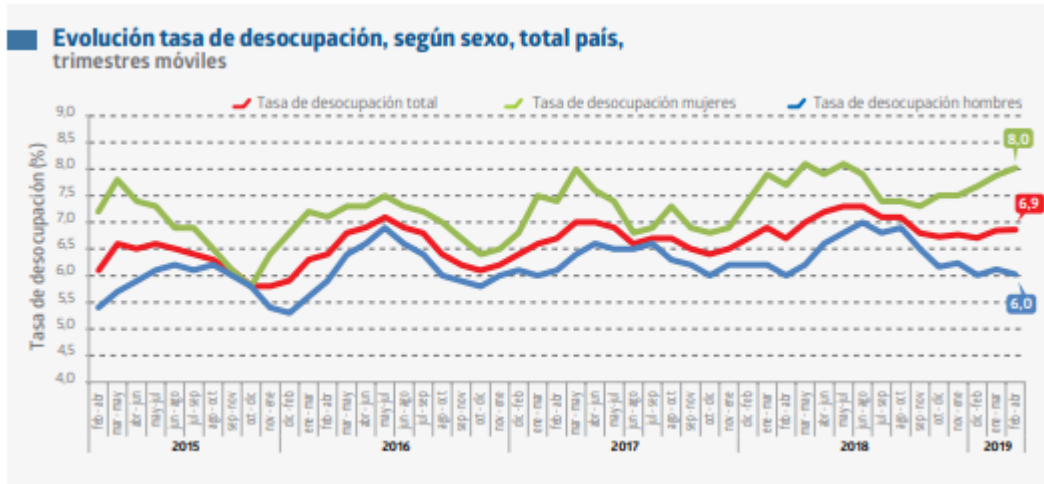
Las variables cuantitativas continuas suelen representarse con histogramas. El histograma es un gráfico de barras juntas. La base de estas barras corresponden a subdivisiones del rango de la variable.

Consideremos la base de datos de la Encuesta de Presupuestos Familiares (EPF) que se encuentra en <https://www.ine.cl/estadisticas/ingresos-y-gastos/epf>. Consideremos la variable gasto total del hogar, en otras capitales regionales y donde el sustentador principal tiene grado de magister. El histograma sería el siguiente:



En este caso el gasto máximo es aproximadamente \$9.000.000. El rango de la variable que va de 0 a 9.000.000 se ha subdividido en 18 subintervalos de ancho igual a 500.000. Del gráfico, se desprende por ejemplo que más de 6.000 hogares de las otras capitales regionales cuyo sustentador principal tiene grado de magister tienen un gasto total entre \$1.500.000 y \$2.000.000.

Muchas veces se quiere representar la evolución de una variable en el tiempo. Si son muchos periodos, el grafico que se recomienda en estos casos es un gráfico de líneas. Veamos el siguiente ejemplo donde se muestra la evolución de la tasa de desocupación por trimestres móviles entre febrero del 2015 y abril del 2019 (Fuente: https://www.ine.cl/docs/default-source/boletines/empleo/2019/espanol/bolet%3%ADn-empleo-nacional-trimestre-m%3%B3vil-febrero-abril-2019.pdf?sfvrsn=85e35dd2_6):

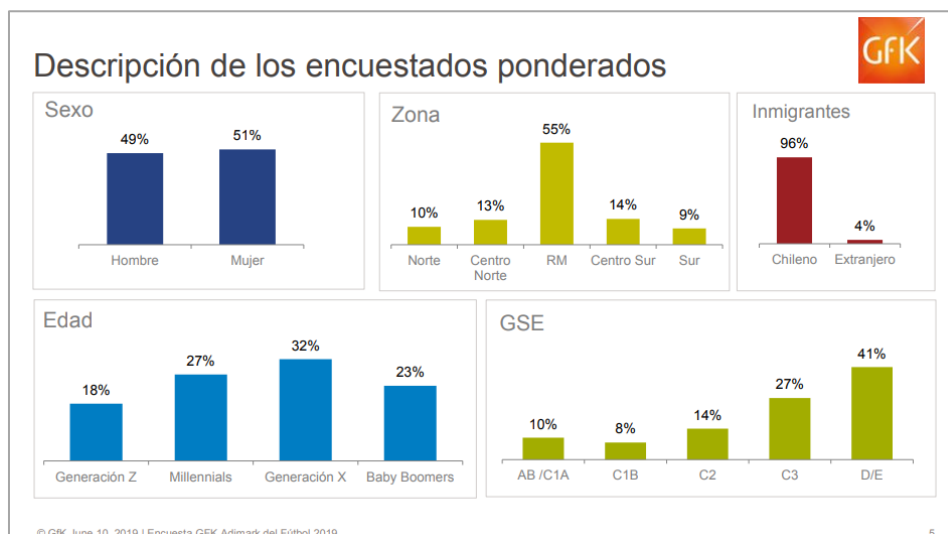


Hoy en día existen muchos paquetes estadísticos o aplicaciones con las que se pueden construir representaciones graficas. Estas aplicaciones nos salvan de no cometer errores que solían cometerse cuando los gráficos se hacían “a mano”. Algunos de los errores más frecuentes son:

- Valores errados o ausentes en las escalas
- Altura de las barras o sectores circulares no proporcionales a las frecuencias
- Barras de anchos diferentes ya sea en el diagrama de barras o en el histograma.

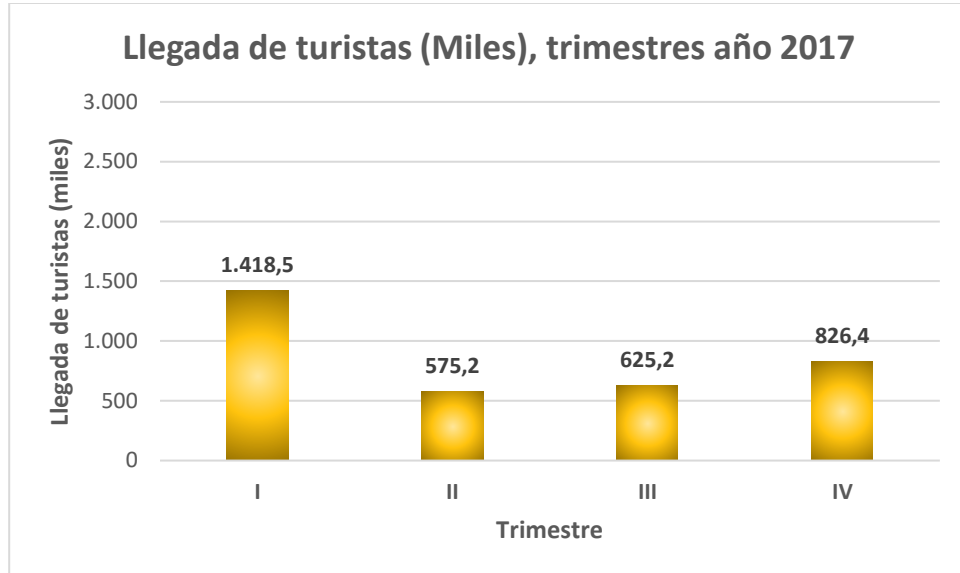
A pesar de lo anterior, dada la flexibilidad que pueden tener algunas aplicaciones, los errores más comunes que suelen cometerse hoy en día guardan relación con la escala. Aquí les dejo un par de recomendaciones al respecto:

- Deben ir identificados y con las escalas correspondiente ambos ejes, X e Y, del grafico. A veces, por tema de diseño, en los gráficos de barra se elimina el eje Y. En esos casos se debe poner el valor; frecuencia o porcentaje que representa cada barra. Veamos este ejemplo de la 7ª ENCUESTA GfK DEL FÚTBOL CHILENO 2019 (Fuente: https://www.gfk.com/fileadmin/user_upload/country_one_pager/CL/20190611_Encuesta_GfK_del_Futbol_2019_VF.pdf)



- Selecciona una escala apropiada para los datos tomando en cuenta el valor máximo que toma la variable.

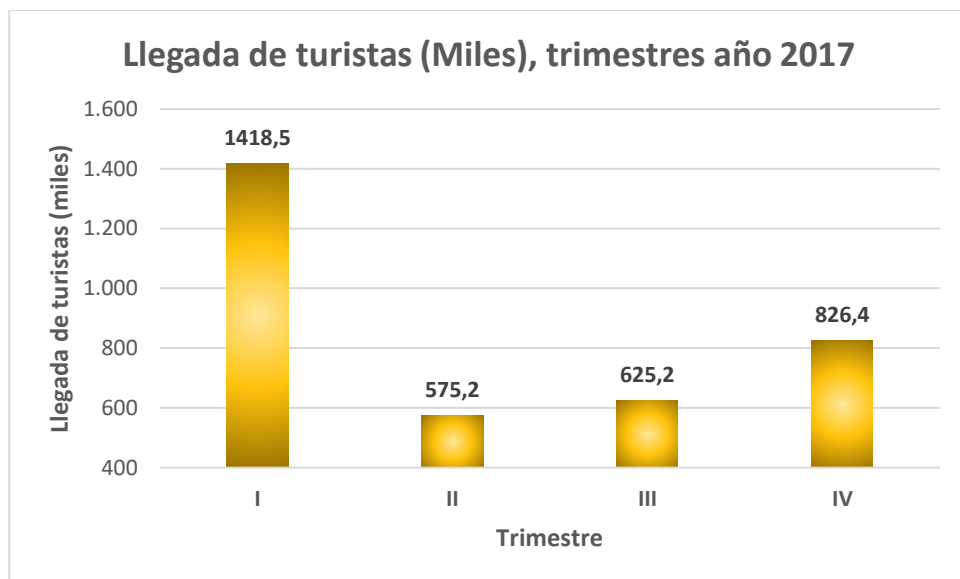
✗ **Incorrecto**



Note que el máximo valor observado es 1.418, 5 por lo que, si el rango del eje llega hasta 3.000, no se aprecian en su totalidad las variaciones de trimestre a trimestre que presenta la variable Llegada de turistas.

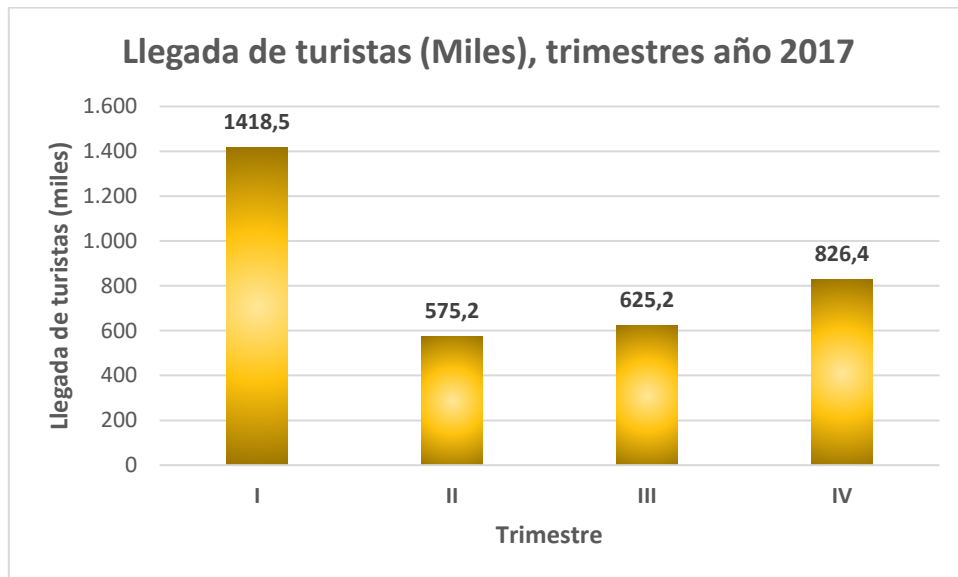
- No cortes el eje, en especial cuando se trata de diagramas de barra. En otras palabras, el eje Y siempre debe partir de 0.

✗ **Incorrecto**



Observe que visualmente da la impresión de que caída en la llegada de turistas a Chile que se da del primer al segundo semestre del 2107 fuera mucho más de lo que realmente fue que el incremento del tercer al cuarto semestre fue mucho más de lo que efectivamente fue.

✓ Correcto



Espero que estos consejos les sirvan y construyan gráficos que grafiquen... ¡¡por favor!!