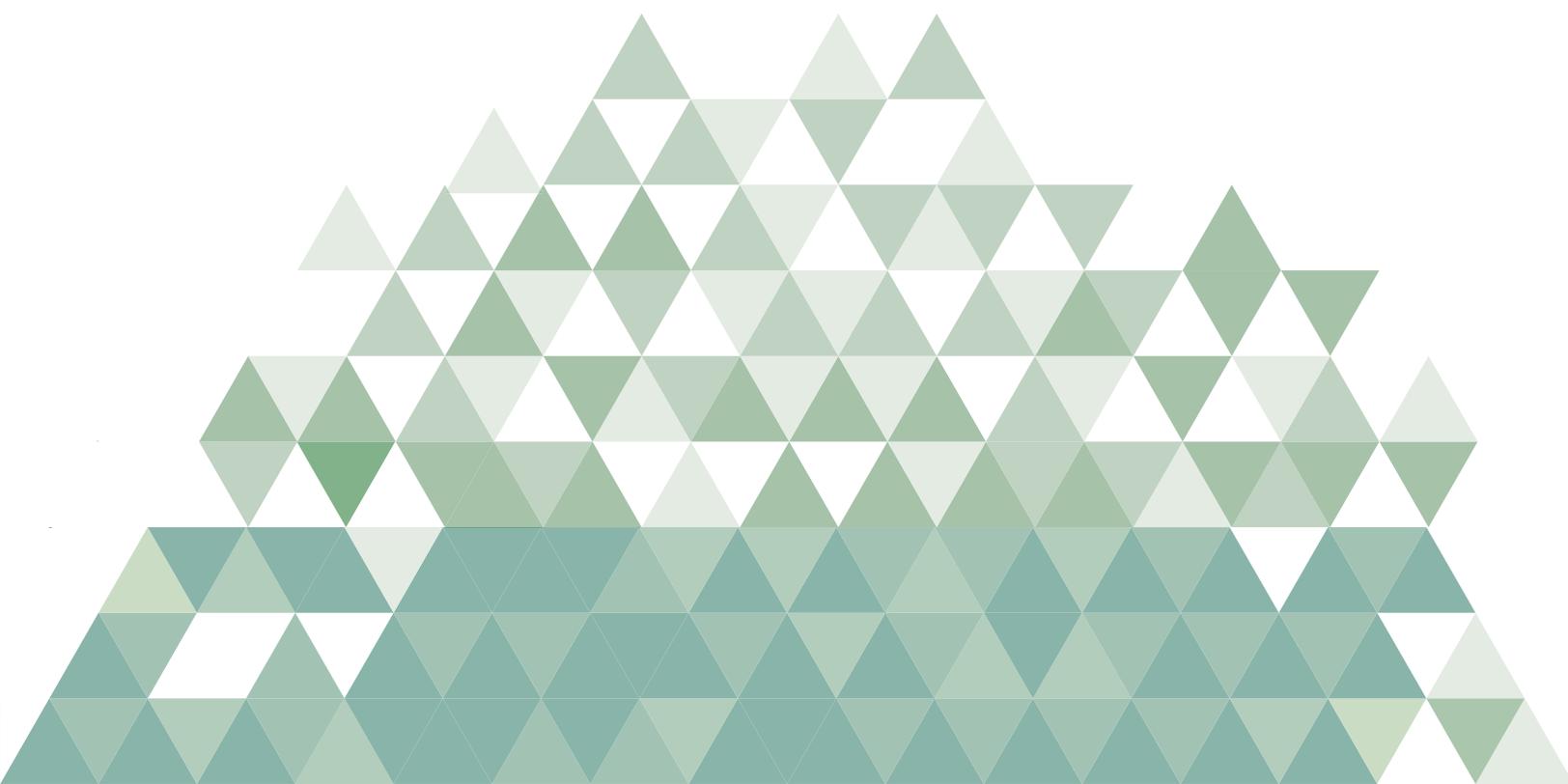


SUMA Y SIGUE MATEMÁTICA EN LÍNEA

MATERIAL PEDAGÓGICO COMPLEMENTARIO

MATERIAL PEDAGÓGICO COMPLEMENTARIO

FICHAS TALLER 3:
VISUALIZACIÓN Y ANÁLISIS DE DATOS.



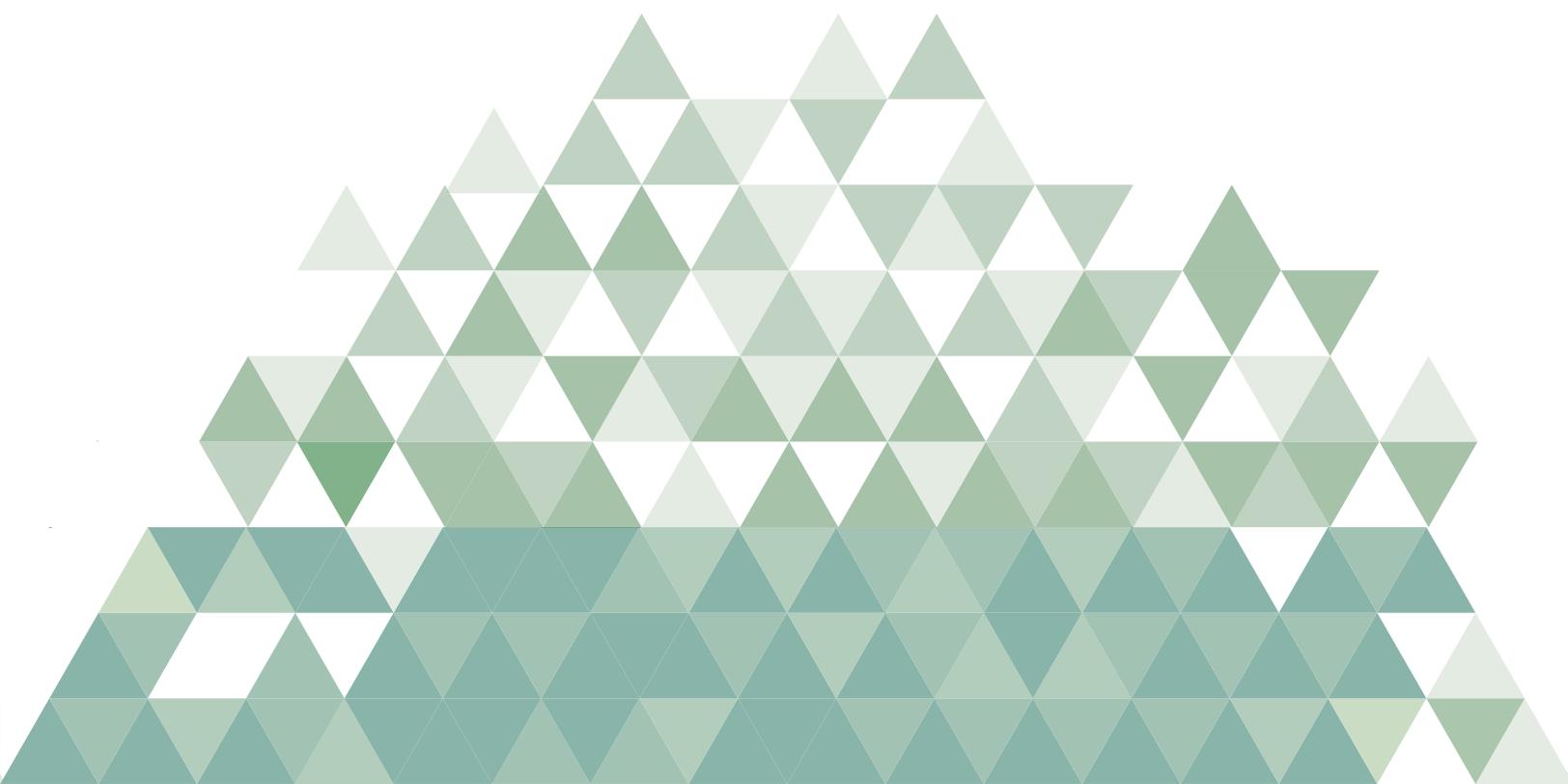
INTRODUCCIÓN

En este taller hemos reflexionado acerca de la relevancia de la visualización en el análisis de los datos, abordando las características de las principales representaciones gráficas, las que constituyen, junto con las tablas, herramientas fundamentales para iniciar el trabajo estadístico.

Las actividades de este taller nos han permitido reconocer las ventajas y limitaciones de cada una de estas representaciones, así como las dificultades y errores que se pueden presentar en su construcción, lectura e interpretación. Además, hemos distinguido los diferentes niveles requeridos para desarrollar la competencia de lectura e interpretación de gráficos.

Los contenidos abordados en las fichas son los siguientes:

- Gráficos.
 - Pictograma.
 - Gráfico de barras.
 - Gráfico de barras agrupadas y apiladas.
 - Gráfico circular.
 - Gráfico de puntos.
 - Diagrama de tallo y hojas.
 - Gráfico de líneas.
 - Histograma.
- Lectura de gráficos.





1- Gráficos

Un **gráfico** es una representación visual de la distribución de un conjunto de valores numéricos, en la que se utilizan distintos recursos para representar las frecuencias asociadas a las categorías de una variable.

Uno de los objetivos de los gráficos es comunicar información de una manera sintética, explícita y fácil de comprender.

Existen distintos tipos de gráficos, tales como los pictogramas, los gráficos de barras y los gráficos circulares. Todos ellos pueden utilizarse para representar grandes cantidades de datos.

Otros tipos de representaciones gráficas, tales como los gráficos de puntos y los diagramas de tallo y hojas, solo son útiles cuando las cantidades de datos por representar son relativamente pequeñas.



Comentarios

- La elección y construcción de un gráfico requiere la toma de ciertas decisiones que finalmente determinarán el grado de efectividad y fidelidad con que la información será visualmente comunicada. Estas decisiones deben considerar, al menos:
 - la naturaleza de los datos disponibles (tipo, tamaño).
 - la información que se espera sea visualmente comunicada.
- Es importante que los docentes tengan en cuenta que la lectura, la interpretación y la construcción de los distintos tipos de gráficos que pondrán a disposición de sus estudiantes, corresponden a actividades de muy diversos niveles de complejidad.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.

Actividad: Todas las actividades.

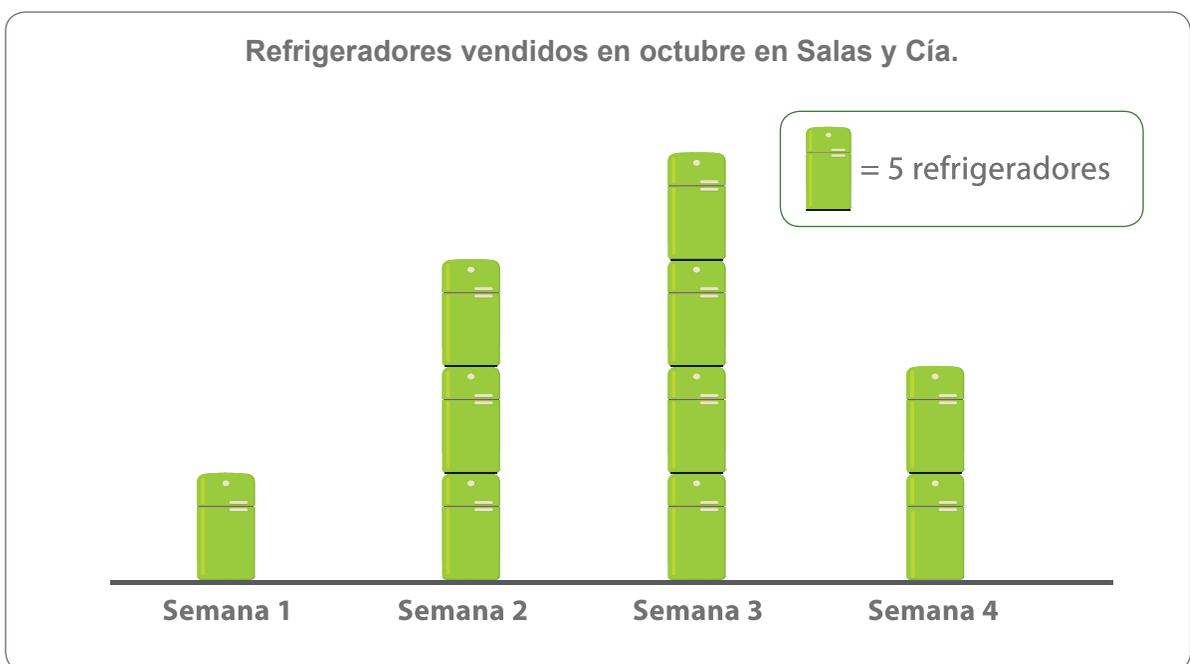


2- Pictograma

Un **pictograma** es una forma visual de mostrar la distribución de un conjunto de datos, en la que se utilizan figuras (dibujos o símbolos) para representar las frecuencias asociadas a las categorías de una variable.

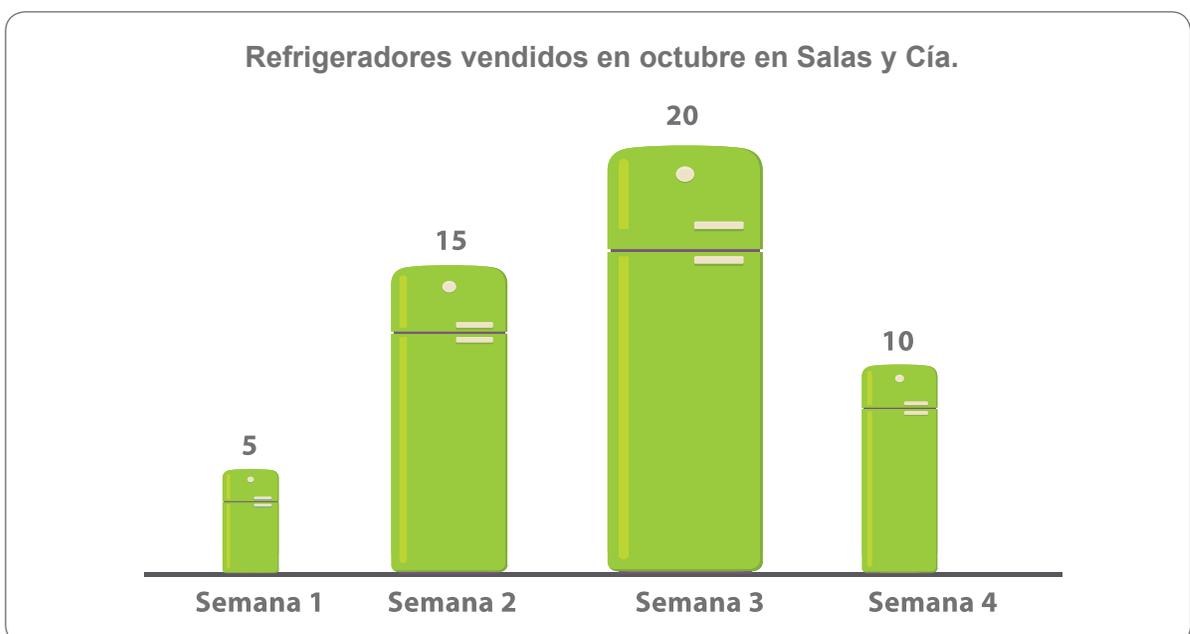
Se distinguen dos tipos de pictogramas:

- **Con escala:** La frecuencia asociada a cada categoría se representa mediante una fila o columna de figuras, en la que cada una de ellas representa una misma cantidad. La escala es la relación entre la figura y la cantidad que representa. Por ejemplo, en el siguiente pictograma cada figura representa a 5 refrigeradores:



Todas las figuras deben ser del mismo tamaño y equidistantes. Este tipo de pictogramas es el más usual.

- **Sin escala:** La frecuencia asociada a cada categoría se representa mediante una sola figura, cuyo tamaño, generalmente medido por la altura, está en correspondencia con dicha frecuencia. Debido a que las figuras representan distintos valores, es necesario indicarlos de forma explícita sobre cada figura o en un eje, como se muestra en el siguiente ejemplo:





Comentarios

- Es conveniente iniciar el aprendizaje usando pictogramas con escala en los que cada figura represente a un solo objeto.
- Muchas veces sucede que las frecuencias que se desea representar a través de un pictograma no son múltiplos de la escala, por lo que su representación resulta ser aproximada. Para mejorar la aproximación se suele usar el dibujo de media figura, pero no se recomienda introducir otras fracciones, para no complejizar la lectura del gráfico.
- Corresponde a los docentes procurar que los estudiantes comprendan la importancia de que en un pictograma con escala todas las figuras tengan el mismo tamaño y estén a igual distancia entre ellas, de modo que las alturas de las columnas guarden relación con las frecuencias respectivas. El uso de hojas de papel cuadriculado puede ayudar a lograr este objetivo.
- En los pictogramas sin escala las figuras más altas también son más anchas, a fin de conservar sus proporciones. Esto puede conducir a interpretaciones erróneas sobre la relación entre las cantidades representadas, la que depende solo de la altura de las figuras, no de su área.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.
Actividad: Estudiando la inmigración en Chile.



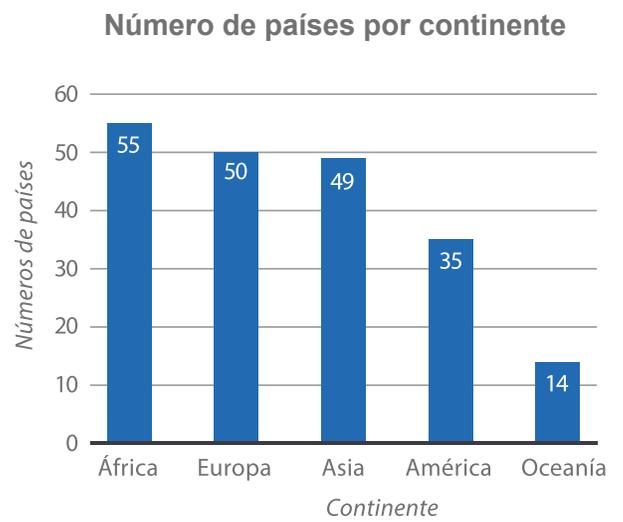
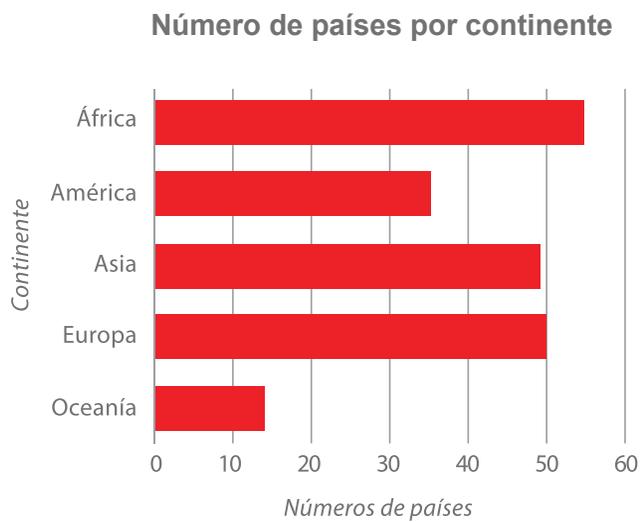
3- Gráfico de barras

Un **gráfico de barras** es una forma visual de mostrar la distribución de un conjunto de datos, en la que se usan barras para representar las frecuencias de cada categoría.

Las barras son de un mismo ancho, elegido arbitrariamente. Las frecuencias pueden estar registradas en un eje paralelo a las barras, o bien sobre ellas. Se distinguen dos tipos de gráficos de barras, de acuerdo a la orientación de estas: **vertical** y **horizontal**.

Ejemplo:

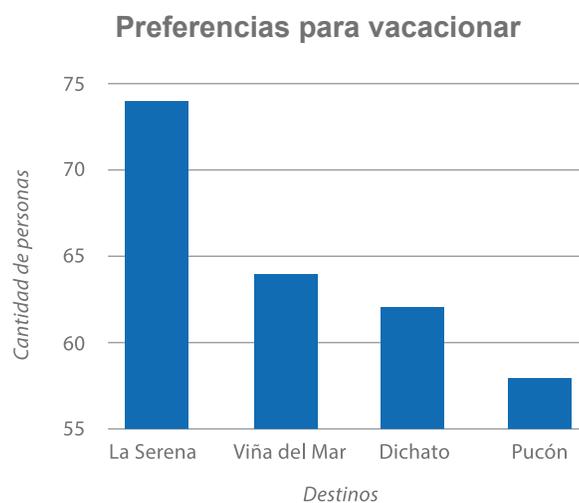
Los siguientes gráficos de barras con orientaciones horizontal y vertical, respectivamente, muestran el número de países que hay en los distintos continentes:





Comentarios

- Los gráficos de barras tienen la ventaja de permitir comparar de forma rápida las frecuencias de las distintas categorías y de ayudar a visualizar el comportamiento de los datos a través de la forma de la distribución.
- Una limitación de los gráficos de barras reside en que, en general, no es fácil distinguir qué parte del total representan las frecuencias de las distintas categorías.
- Con el fin de facilitar la comunicación de la información representada, en la construcción de un gráfico de barras se debe procurar:
 - indicar de forma precisa el título del gráfico, los nombres de las categorías, las frecuencias marcadas en el eje a una escala dada y la fuente de los datos.
 - seleccionar el rango de frecuencias cubierto por el eje, de modo que no sea mucho mayor que la frecuencia más alta (la barra más larga).
- Si en un gráfico de barras se quiere acentuar la visibilidad de las diferencias entre las cantidades que representan las barras, se puede cubrir un intervalo más restringido en el eje. Por ejemplo, el gráfico siguiente comienza en 55 personas.



- Por el contrario, si lo que se quiere es preservar las proporciones relativas entre los largos de las barras, el intervalo cubierto por el eje debe partir desde cero. En este ejemplo, la cantidad de personas que prefieren vacacionar en La Serena no corresponde al doble de las que prefieren vacacionar en Viña del Mar, lo que se apreciaría claramente si el eje vertical comenzara en cero.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.
Actividad: Estudiando la inmigración en Chile.

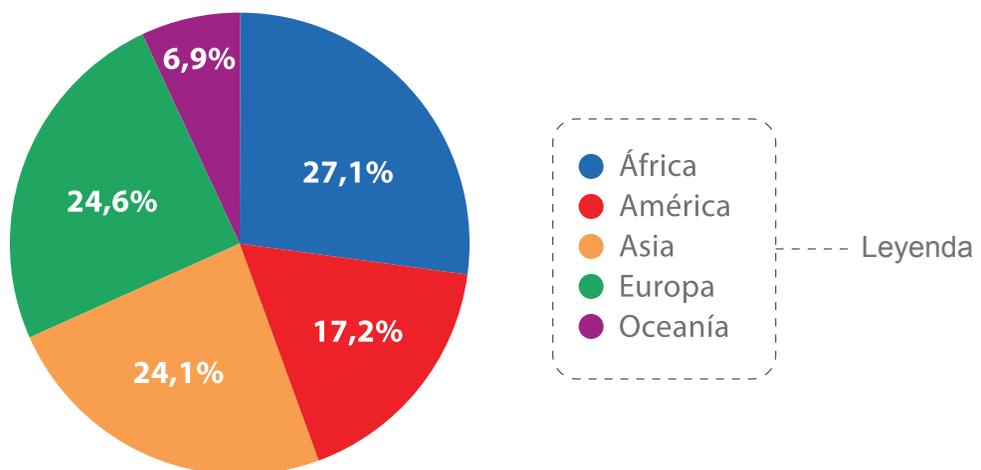


4- Gráfico circular

Un **gráfico circular**, o también llamado de **torta**, es una forma visual de mostrar la distribución de un conjunto de datos. Se basa en un círculo dividido en sectores circulares, cuyas áreas son proporcionales a las frecuencias asociadas a las categorías correspondientes.

En estos gráficos debe indicarse explícitamente la categoría que cada sector representa, lo que generalmente se resume en una lista conocida como **leyenda**. La proporción asociada a cada sector suele indicarse por medio de porcentajes.

Número de países por continente



Comentarios

- En un gráfico circular generalmente se usan porcentajes, pero también pueden usarse frecuencias o fracciones. Es importante tener en cuenta que
 - si se usan porcentajes, la suma total de estos debe ser 100.
 - si se usan frecuencias, la suma de los valores de los sectores debe ser igual al total de datos.
 - si se usan fracciones, la suma de todas ellas debe ser 1.
- Para los estudiantes de Educación Básica que aún no conocen porcentajes, el trabajo con estos gráficos ofrece una interesante oportunidad para ejercitar fracciones.
- Los gráficos circulares tienen la ventaja, respecto a los de barras, de facilitar la comparación de las frecuencias de las categorías con respecto al total de los datos. No obstante, cuando se usan porcentajes, los cuales suelen ser aproximados, no siempre es posible recuperar los valores precisos de las frecuencias, lo que puede dificultar la comparación entre algunas categorías.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.
Actividad: Estudiando la inmigración en Chile.



5- Gráfico de puntos

Un **gráfico de puntos** es un recurso visual en el que se utilizan puntos para representar las frecuencias asociadas a un conjunto de categorías. Las categorías se marcan en un eje vertical (u horizontal), y a cada categoría se le asocia una fila (o columna) con una cantidad de puntos igual a la frecuencia correspondiente a esa categoría. Estas categorías pueden ser cualitativas o cuantitativas.

Por ejemplo, si en un puesto de venta de jugos se registra, durante una hora, cada vaso vendido con un punto, se obtiene un gráfico como el siguiente.



En síntesis, un gráfico de puntos es una tabla de conteo en la que las marcas están igualmente espaciadas. Esto, además de facilitar el conteo de los puntos para conocer la frecuencia asociada a cada categoría, permite también comparar las frecuencias de acuerdo a la longitud de su representación.



Comentarios

- La limitación principal del gráfico de puntos reside en que, como cada punto corresponde a un dato, solo es posible utilizarlo cuando la cantidad de datos es relativamente pequeña.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.

Actividad: Visualizando la inmigración en las escuelas de una comuna.

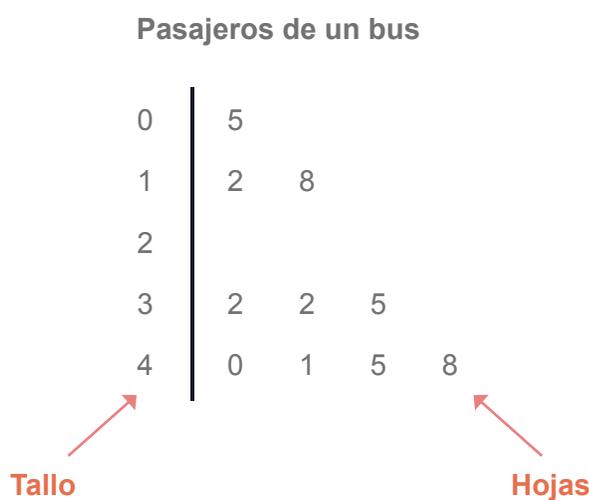


6- Diagrama de tallo y hojas

Un **diagrama de tallo y hojas** es una forma compacta de representar un conjunto de datos numéricos. El valor de cada dato es separado en una parte que se registra en una **hoja** (los últimos dígitos) y otra parte que se registra en el **tallo** (conformado por el bloque de dígitos restantes).

Los valores del tallo se escriben hacia abajo, y los valores de las hojas en filas hacia la derecha. El tallo es usado para agrupar los datos, y cada hoja permite leer los datos individuales en cada grupo. Por ejemplo, si en un bus viajan 10 personas cuyas edades son las siguientes:

5 - 12 - 18 - 32 - 32 - 35 - 40 - 41 - 45 - 48, su representación en un diagrama de tallo y hojas será:



En el tallo colocamos el número de decenas de cada edad y en las hojas el número de unidades. Los números se ordenan en cada hoja de menor a mayor. Si dos o más datos coinciden, se repiten en la misma hoja la cantidad de veces que corresponda. Por otra parte, los números que van en los tallos se escriben en orden creciente y correlativo de arriba hacia abajo, aún si no existen datos para algunos de ellos.



Comentarios

- Cabe destacar que en el diagrama de tallo y hojas los números registrados en el tallo representan agrupaciones a las que pertenecen los datos, cuyos valores se especifican en las hojas. Cualquier otro tipo de información sobre los datos no aparece explícitamente.
- Al igual que en el gráfico de puntos, en este diagrama cada dato está individualmente representado, lo que limita su uso a una cantidad relativamente pequeña de datos. Este diagrama, además, requiere datos numéricos que se puedan descomponer de acuerdo a la estructura del sistema de numeración decimal.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.

Actividad: Visualizando la inmigración en las escuelas de una comuna.



7- Datos agrupados

Al estudiar la distribución de una variable cuantitativa, diremos que los **datos** están **agrupados** cuando los valores de la variable se organizan en intervalos. A cada intervalo o **clase** se le asocia el número de datos que se encuentran en él, es decir, su frecuencia. El largo de cada intervalo se conoce como **amplitud**.

Ejemplo: Los siguientes datos corresponden al número de consultas en la biblioteca de un curso de 20 niños durante un mes.

Datos no agrupados: 2 - 4 - 0 - 1 - 3 - 3 - 4 - 6 - 2 - 8 - 9 - 12 - 5 - 6 - 3 - 8 - 5 - 9 - 10 - 3

Datos agrupados en 4 clases:

Amplitud	Cantidad de consultas (clases)	Cantidad de niños (frecuencia)
2 ←	0 - 2	4
2 ←	3 - 5	8
2 ←	6 - 8	4
3 ←	9 - 12	4
	Total	20

Trabajar con datos agrupados facilita el análisis de la distribución cuando las observaciones toman una gran cantidad de valores distintos.



Comentarios

- Las clases pueden tener o no la misma amplitud, de acuerdo a la distribución de los datos.
- Al agrupar los datos en clases, estas deben tener al menos las siguientes dos características:
 - ser exhaustivas: todos los datos deben pertenecer a alguna clase.
 - ser mutuamente excluyentes: cada dato debe pertenecer a una sola clase.
- Para evitar ambigüedades al agrupar datos de variables continuas, es necesario precisar si la clase incluye o no a sus extremos, lo que depende de la convención que se esté usando. Una convención posible consiste en considerar que un intervalo incluye el extremo inferior, pero no al superior, salvo en el caso del último intervalo, que también incluye al extremo superior.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.

Actividad: Visualizando la inmigración en las escuelas de una comuna.



8- Histograma

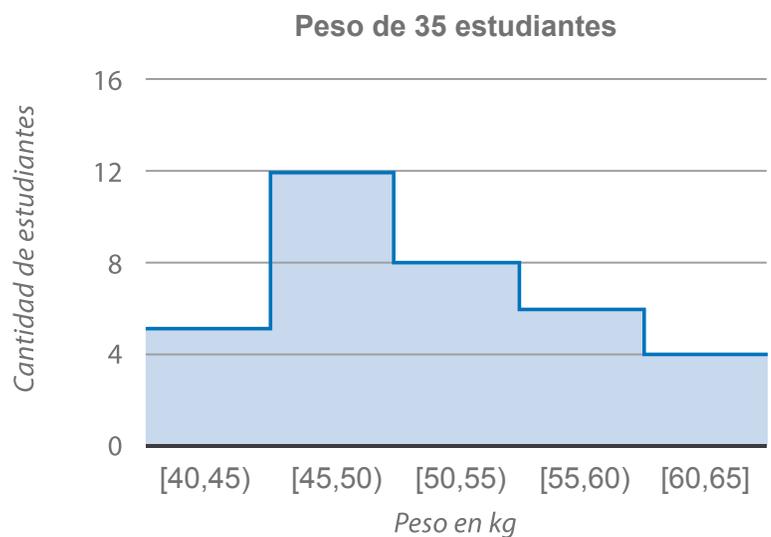
Un **histograma** es un gráfico que utiliza barras para representar la distribución de frecuencias de una variable cuantitativa, cuyas categorías son intervalos consecutivos.

Estas clases se representan en el eje horizontal del gráfico y cada una constituye la base de la barra, cuya altura representa la frecuencia correspondiente.

Si bien es similar al gráfico de barras, en el histograma las barras se disponen juntas, lo que facilita la visualización global de la distribución.

Ejemplo:

Peso (kg)	Cantidad de estudiantes
[40,45)	5
[45,50)	12
[50,55)	8
[55,60)	6
[60,65]	4

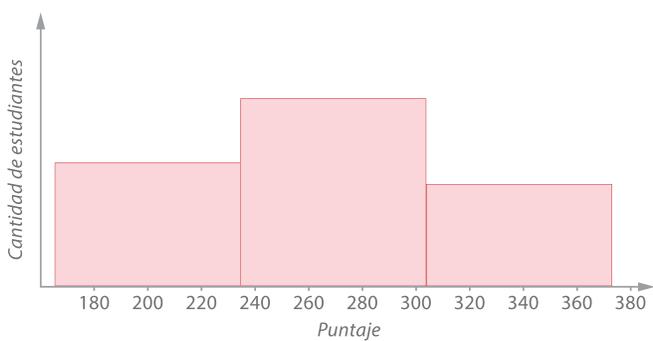


El histograma es especialmente útil para variables continuas, o variables cuantitativas discretas con gran número de valores, agrupadas en intervalos o clases.

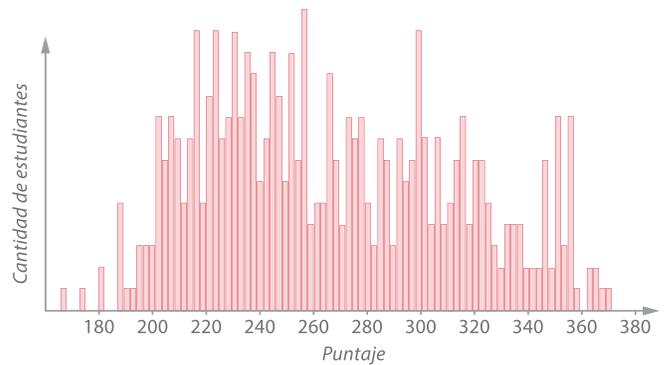
Comentarios

- Por su naturaleza, en el histograma las barras son siempre contiguas y su ancho se corresponde con la amplitud de los intervalos. En el gráfico de barras, en cambio, las barras se presentan separadas, y su ancho se escoge arbitrariamente.
- Al agrupar datos en intervalos de igual tamaño, a mayor número de intervalos, menor será la amplitud, y viceversa. Al construir un histograma, la elección del número de intervalos incide, por tanto, en qué tan detallada se puede observar la distribución de los datos.
- Así, en el gráfico de la izquierda, con solo tres clases, es difícil apreciar la variabilidad de los datos. En el de la derecha, en cambio, el gran número de clases impide observar regularidades en la variabilidad y, por lo tanto, interpretar de manera adecuada los datos.

Distribución de puntajes en una prueba estandarizada



Distribución de puntajes en una prueba estandarizada



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.
 Actividad: Visualizando la inmigración en las escuelas de una comuna.



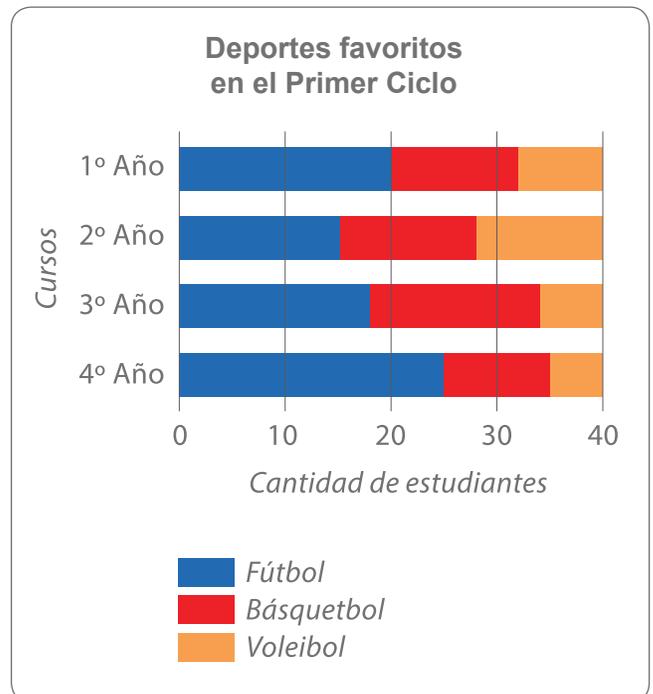
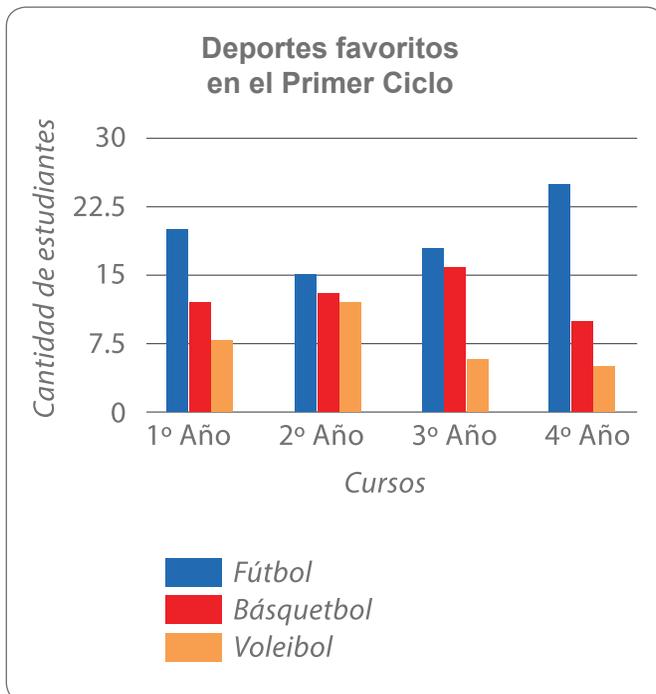
9- Gráfico de barras agrupadas y apiladas

Los gráficos de barras agrupadas y apiladas resultan útiles para comparar varias distribuciones.

El **gráfico de barras agrupadas** se usa para visualizar las frecuencias registradas en una tabla de doble entrada. En él, las barras correspondientes al mismo valor de la variable se ubican una al lado de la otra. Según lo que se busca visualizar, la agrupación de las barras se puede hacer tomando los valores de cualquiera de las variables de la tabla.

El **gráfico de barras apiladas** es similar al de barras agrupadas, salvo que las barras en lugar de agruparse se apilan una sobre la otra. En general, este gráfico es usado cuando se desea comparar las frecuencias de las categorías respecto al total de datos.

Por ejemplo, los resultados de una encuesta aplicada a los estudiantes del primer ciclo de enseñanza básica de una escuela, en la que se pedía que cada uno eligiera su deporte preferido entre fútbol, básquetbol y voleibol, han sido representados en los siguientes gráficos.



En estos gráficos la comparación directa es posible, ya que los cuatro cursos tenían la misma cantidad de estudiantes. Si la cantidad de estudiantes de los cursos hubiese variado, habría sido necesario representar las frecuencias relativas, para poder compararlas.

En ambos gráficos la longitud de las barras indica la frecuencia o cantidad de estudiantes que han preferido un deporte determinado, y el color de las barras muestra cuál es el deporte preferido, de acuerdo a lo señalado en la leyenda.



Comentarios

- Es importante notar que en estos gráficos cada grupo de barras, o cada pila, permite visualizar la distribución de las preferencias de los estudiantes de cada uno de los cursos.
- Pero ambos gráficos permiten también comparar las preferencias de los distintos cursos por un determinado deporte. Por ejemplo, las preferencias por el fútbol, al considerar las cuatro barras azules de cada gráfico.
- Los gráficos son herramientas estadísticas que buscan transmitir información de manera visual, por lo que es imprescindible que no presenten ambigüedades o particularidades que dificulten la lectura o la comparación de la información.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.

Actividad: ¿Cómo ha variado la inmigración en las escuelas de la comuna?

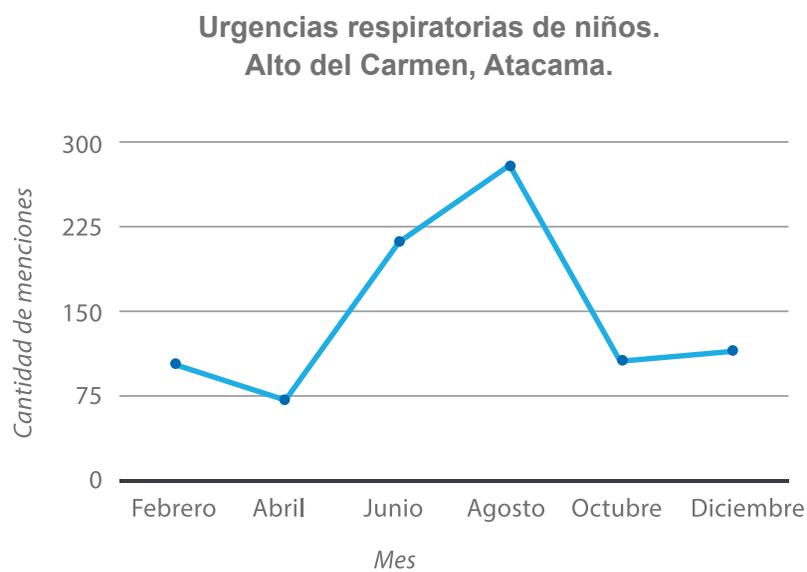


10- Gráfico de líneas

Un **gráfico de líneas** es una representación visual de un conjunto de valores asociados a distintas **categorías cuantitativas**. Este tipo de gráficos se utiliza para estudiar la variación de los valores, usualmente a través del tiempo.

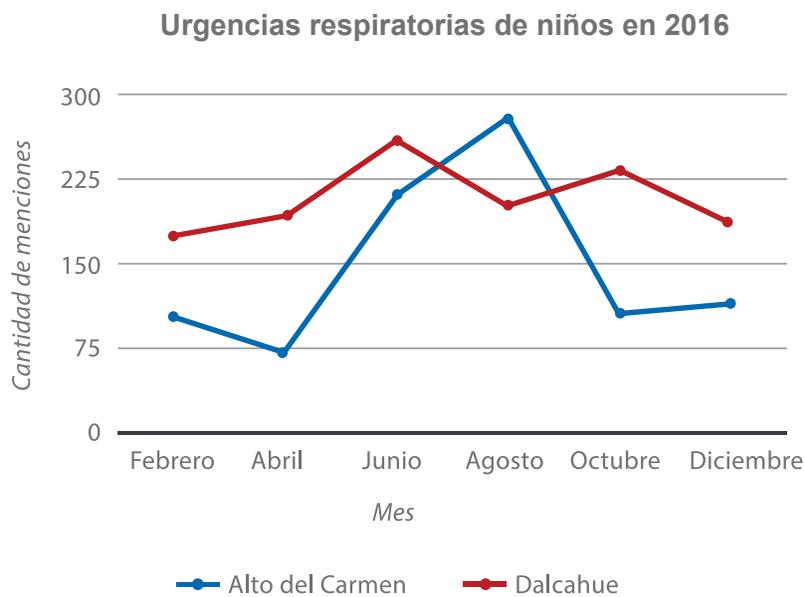
En un gráfico de líneas, cada valor se representa con un punto ubicado sobre la categoría correspondiente en el plano, a una altura que corresponde a ese valor. Además, para facilitar la visualización de la variación de los valores, se traza una línea recta entre los puntos asociados a categorías consecutivas.

El siguiente gráfico representa la evolución de la cantidad de atenciones de urgencia por afecciones respiratorias de niños de 5 a 14 años durante los seis bimestres de 2016, en el consultorio Alto del Carmen, en Atacama.



Comentarios

- Los segmentos entre dos puntos sucesivos se interpretan como conjuntos de valores que hipotéticamente se asocian a categorías intermedias. Por ejemplo, se puede pensar que en el mes de mayo el consultorio atendió aproximadamente a 130 niños con urgencias respiratorias. Esto corresponde a una **interpolación** de un valor.
- Del mismo modo, el gráfico de líneas permite asignar, también hipotéticamente, valores asociados a categorías previas o posteriores al período en que se efectuó el registro. Por ejemplo, se puede suponer que en el mes de enero de 2017 la cantidad de este tipo de urgencias atendidas pudo ser, aproximadamente, de 150. Esto corresponde a una **extrapolación** de un valor.
- Tanto el gráfico de barras agrupadas como el de líneas permiten comparar datos correspondientes a dos o más variables, puesto que estos pueden ser incluidos en una misma representación y utilizando una misma escala. Así, es posible comparar la cantidad de atenciones a urgencias respiratorias de niños de 5 a 14 años en el consultorio de Alto del Carmen, en Atacama, con la del consultorio de Dalcahue, en Chiloé.



- Sin embargo, mientras que el gráfico de barras agrupadas representa la distribución de los datos recopilados, el gráfico de líneas muestra la evolución de estos datos, lo que permite identificar más claramente tendencias o patrones.

Ubicación: Módulo 1

Taller: Visualización y análisis de datos.
 Actividad: ¿Cómo ha variado la inmigración en las escuelas de la comuna?



11- Lectura de gráficos

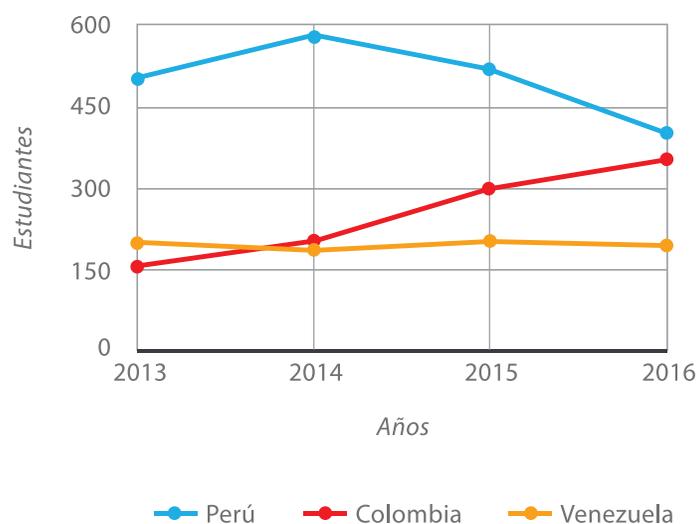
Un objetivo muy importante de las representaciones gráficas es comunicar información, por lo que es necesario aprender a leer e interpretar los datos contenidos en un gráfico.

Típicamente se distinguen tres niveles:

- **Nivel elemental:** lectura directa e inmediata del gráfico, en la que no se interpreta la información contenida en él. Implica una comprensión básica de la forma en que los datos están representados en el gráfico.
- **Nivel intermedio:** lectura que involucra la interpretación e integración de los datos que muestra el gráfico. Implica observar el conjunto de los datos representados y detectar relaciones entre ellos mediante comparaciones directas u operaciones simples.
- **Nivel avanzado:** lectura que permite realizar predicciones e inferencias que van más allá de los datos que aparecen en el gráfico. Implica hacer estimaciones relativas al comportamiento de las variables estudiadas a partir de los datos presentes en el registro gráfico.

A continuación se ejemplifican preguntas que, para ser respondidas, requieren distintos niveles de lectura:

Cantidad de estudiantes extranjeros por país y por año



Nivel elemental: ¿En qué año hubo aproximadamente 300 estudiantes colombianos en las escuelas consideradas en el gráfico?

Nivel intermedio: ¿Aumentó cada año la cantidad de estudiantes extranjeros para alguna de estas nacionalidades?

Nivel avanzado: ¿Es esperable que en el año 2017 haya más estudiantes colombianos que peruanos y venezolanos en este grupo de escuelas?



Comentarios

- La distinción entre estos niveles de lectura constituye una eficaz herramienta pedagógica para graduar la dificultad de las actividades propuestas a los estudiantes.
- Por ejemplo, se espera que en los primeros cursos los niños sean capaces de responder y formular preguntas de nivel elemental, para luego, en la medida que avanzan a los siguientes cursos, ir abordando los otros dos niveles de lectura.
- Por otro lado, cada vez que estudian un nuevo gráfico, es importante organizar las actividades siguiendo esta secuencia de niveles de lectura.



Ubicación: Módulo 1

Taller: Visualización y análisis de datos.

Actividad: ¿Cómo ha variado la inmigración en las escuelas de la comuna?