

Introducción a la visualización de datos

Análisis y Exploración de datos | ITFS24
Prof. Martín Pasztetnik

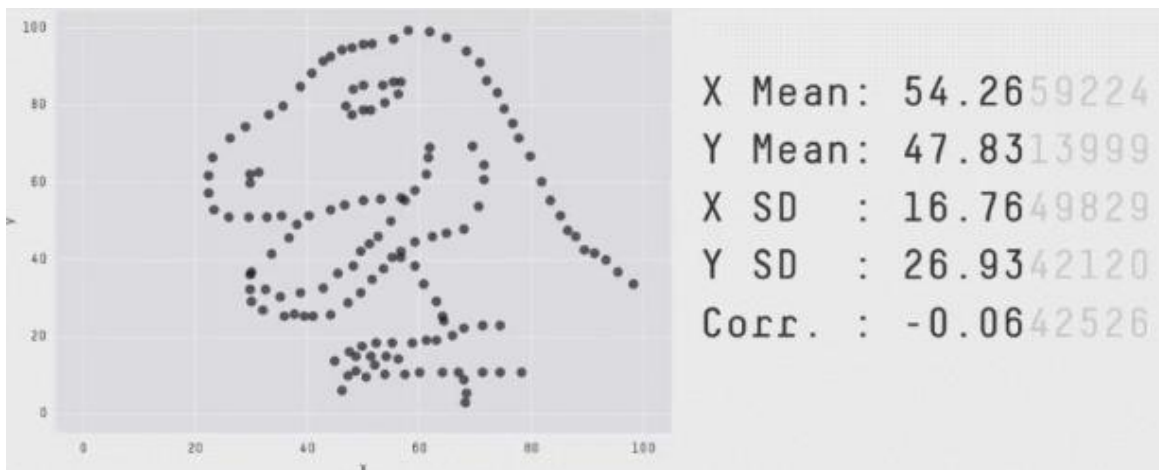




¿Por qué visualizar datos?

1) Para explorar (Jet Thorp)

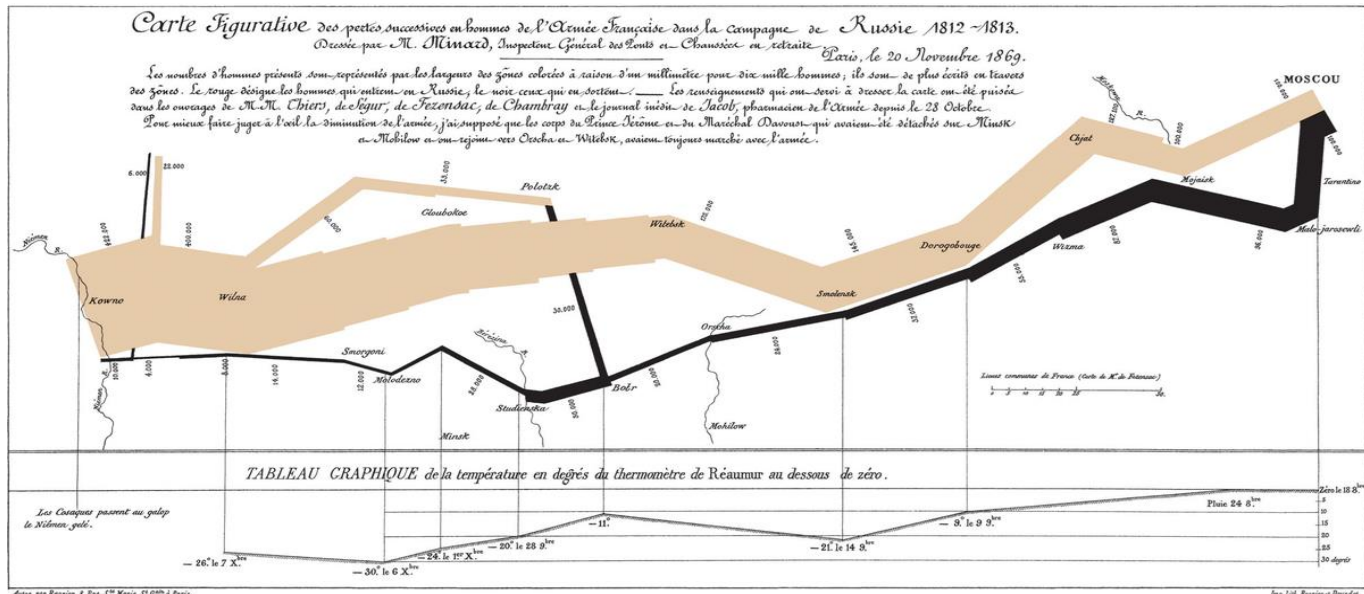
A veces los estadísticos resumen no son suficientes para entender la estructura de los datos.



¿Por qué visualizar datos?

2) Para comunicar (Edward Tufte)

Patrones y fenómenos complejos pueden comprenderse mucho mejor a partir de visualizaciones bien diseñadas.





Temas

1. Entender el contexto
2. Elegir una visual adecuada
3. Eliminar el desorden
4. Percepción y diseño



1. Contexto

Antes de comenzar a crear visualizaciones o comunicaciones de datos, es crucial entender el **contexto en el que se necesita comunicar**. Se discuten las diferencias entre análisis **exploratorio y explicativo**, enfatizando la importancia de pasar de explorar datos a explicar conclusiones específicas. Es común el error de mostrar análisis exploratorio cuando se debería mostrar análisis explicativo. Se destaca la importancia de concentrarse en comunicar la **información relevante para la audiencia**, en lugar de presentar todos los datos sin discernimiento.



¿Quién, qué, cómo?

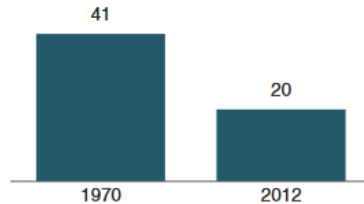
- **¿Quién es tu audiencia?** Comprender quién es tu audiencia y cómo te perciben es fundamental. Identificar puntos en común para asegurarte de que reciban tu mensaje.
- **¿Qué quieres que tu audiencia sepa o haga?** Debes tener claridad sobre lo que deseas que tu audiencia comprenda o realice. Considera cómo comunicarás tu mensaje y el tono general de la comunicación.
- **¿Cómo puedes utilizar los datos para respaldar tu punto?** Una vez que hayas respondido de manera concisa las dos primeras preguntas, puedes avanzar a esta.

2. Elegir una visual adecuada

a. Texto simple

Children with a "Traditional" Stay-at-Home Mother

*% of children with a married
stay-at-home mother with a
working husband*



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013



20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970



El tener números no siempre justifica un gráfico. Se emplea mucho espacio para solo dos números. El gráfico no contribuye mucho a la interpretación de los datos y puede distorsionar la percepción visual de las diferencias. En este caso, una simple oración sería suficiente

2. Elegir una visual adecuada

b. Tablas

Cuando se tiene uno o dos números que se quieren comunicar, es mejor utilizar los números directamente. Pero cuando hay más datos para mostrar, por lo general conviene usar una tabla o un gráfico.



Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ



Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ



Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

2. Elegir una visual adecuada

c. Gráficos

Mientras que las tablas interactúan con nuestro sistema verbal, los gráficos interactúan con nuestro sistema visual, que es más rápido en procesar información. Esto significa que un gráfico bien diseñado generalmente transmitirá la información más rápidamente que una tabla bien diseñada.



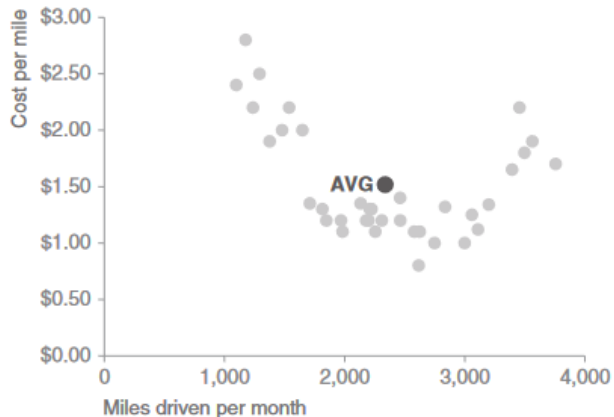


2. Elegir una visual adecuada

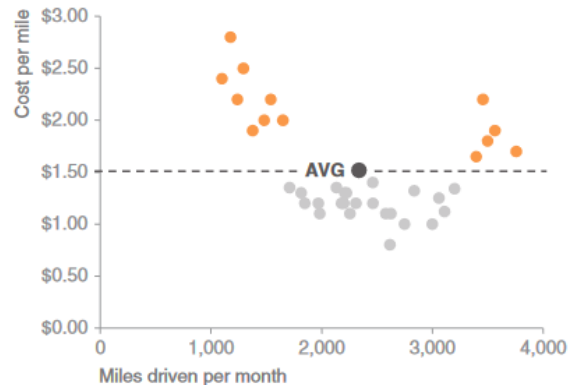
c. Gráficos: Puntos / Scatterplot

Los gráficos de dispersión pueden ser útiles para mostrar la relación entre dos variables, ya que te permiten codificar datos simultáneamente en un eje horizontal (x) y vertical (y) para ver si existe alguna relación y de qué tipo.

Cost per mile by miles driven



Cost per mile by miles driven



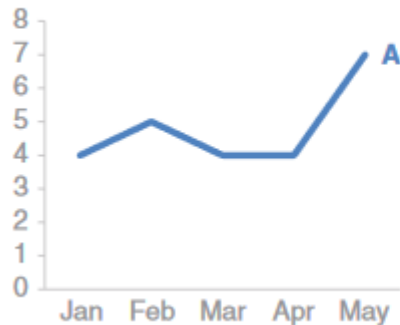


2. Elegir una visual adecuada

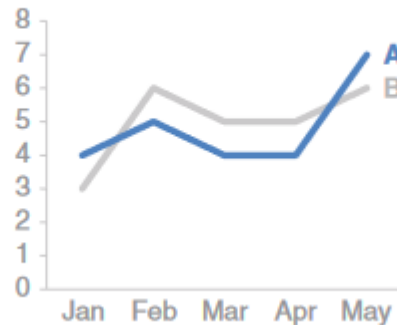
c. Gráficos: Líneas

Los gráficos de líneas se utilizan principalmente para representar **datos continuos**. Al estar los puntos físicamente conectados mediante la línea, implica una relación entre los puntos que puede no tener sentido para datos categóricos (un conjunto de datos clasificados o divididos en diferentes categorías). A menudo, nuestros datos continuos están en alguna unidad de tiempo: días, meses, trimestres o años.

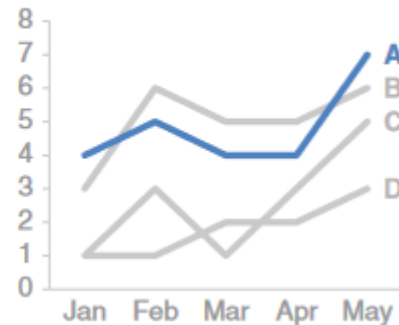
Single series



Two series



Multiple series



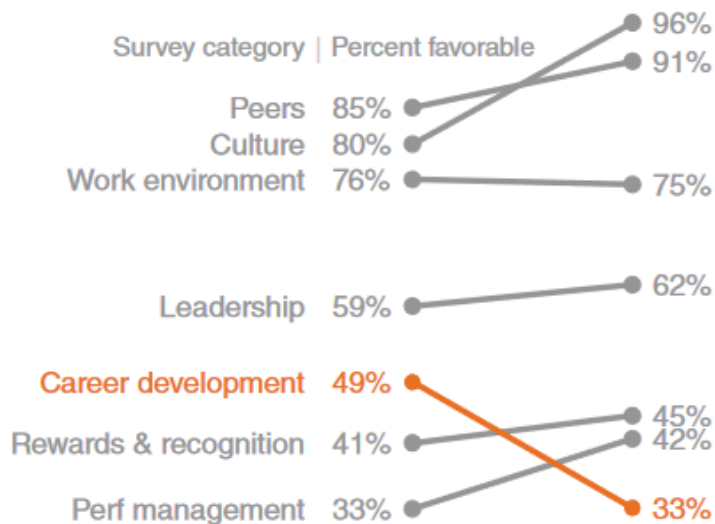


2. Elegir una visual adecuada

c. Gráficos: Pendiente / Slopegraph

Los slopegraphs son útiles cuando tienes dos períodos de tiempo o puntos de comparación y deseas mostrar rápidamente aumentos y disminuciones relativas o diferencias entre varias categorías entre los dos puntos de datos.

Employee feedback over time



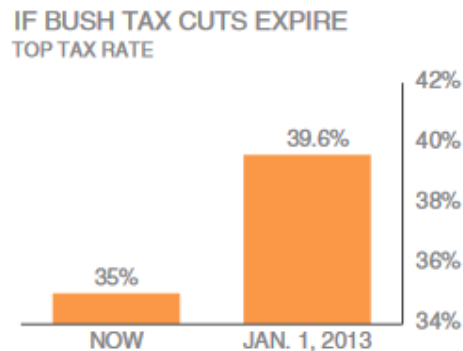


2. Elegir una visual adecuada

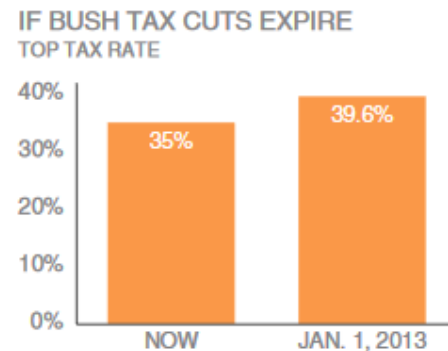
c. Gráficos: Barras

Los gráficos de barras son fáciles de leer para nuestros ojos. Nuestros ojos comparan los extremos de las barras, por lo que es fácil ver rápidamente cuál es la categoría más grande, cuál es la más pequeña y también la diferencia incremental entre las categorías. Es importante destacar que, debido a cómo nuestros ojos comparan los extremos relativos de las barras, es importante que los gráficos de barras siempre tengan una línea base de cero (donde el eje x cruza el eje y en cero), de lo contrario, se obtiene una comparación visual falsa.

Non-zero baseline: as originally graphed



Zero baseline: as it should be graphed



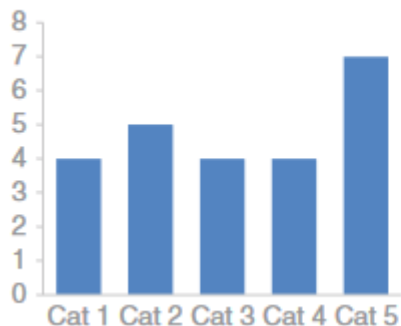


2. Elegir una visual adecuada

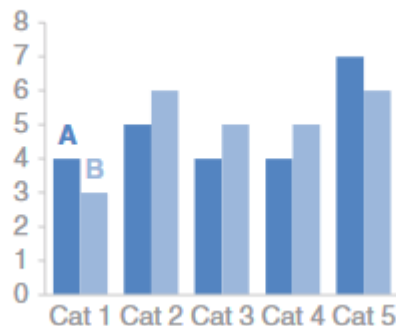
c. Gráficos: Barras verticales

El gráfico de barras básico es el de barras verticales, o sea, el de columnas. Al igual que con los gráficos de líneas, podés tener barras verticales de una sola serie, dos series o varias. Pero hay que tener cuidado,, ya que a medida que se agregan más series de datos, se complica enfocarse en una sola y sacar conclusiones.

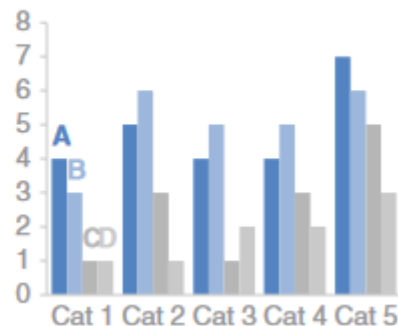
Single series



Two series



Multiple series



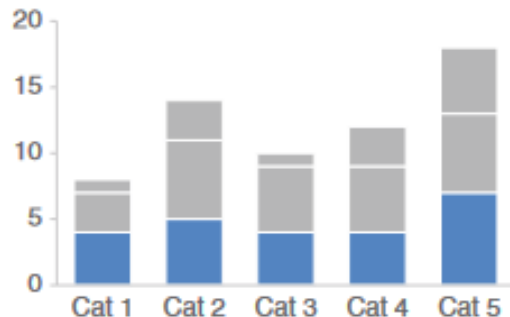


2. Elegir una visual adecuada

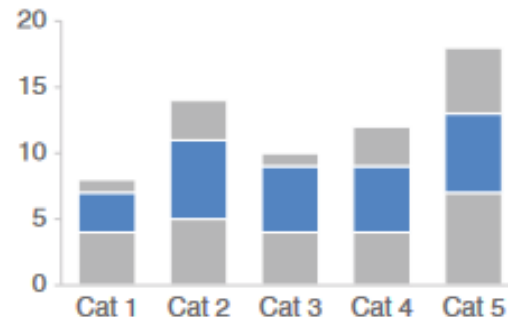
c. Gráficos: Barras apiladas / Stacked bars

Los casos de uso para los gráficos de barras verticales apiladas son más limitados. Están diseñados para permitirte comparar totales entre categorías y también ver las partes subcomponentes dentro de una categoría dada.

Comparing **these** is easy



Comparing **these** is hard

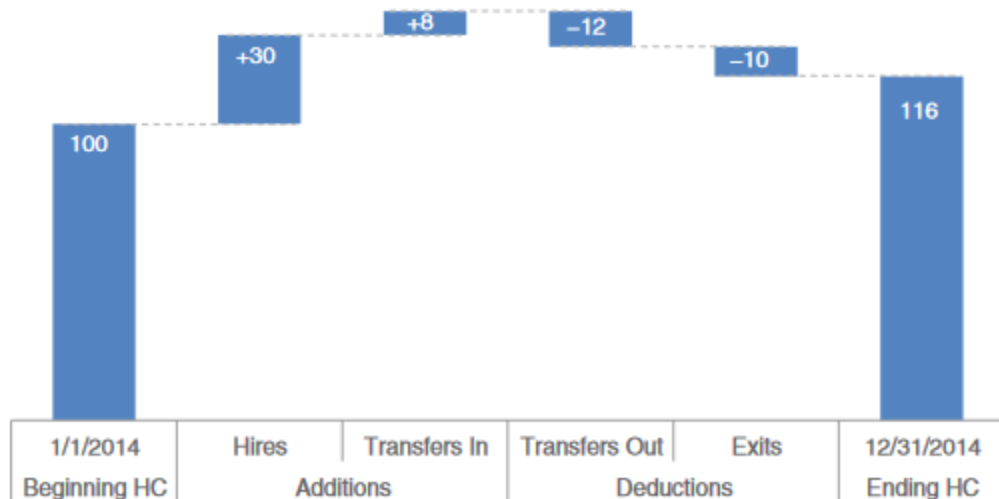




2. Elegir una visual adecuada

c. Gráficos: Cascada / Waterfall

El gráfico de cascada puede utilizarse para separar las partes de un gráfico de barras apiladas y enfocarse en una a la vez, o para mostrar un punto de partida, aumentos y disminuciones, y el punto final resultante.



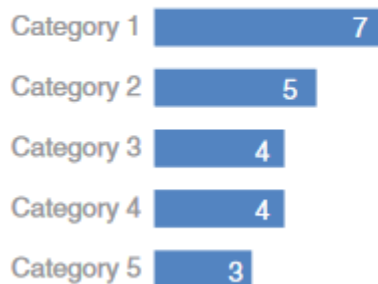


2. Elegir una visual adecuada

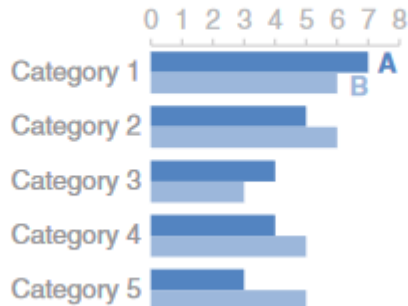
c. Gráficos: Barra horizontal

El gráfico de barras horizontales es especialmente útil si los nombres de tus categorías son largos, ya que el texto se escribe de izquierda a derecha, como la mayoría de las audiencias leen, lo que hace que tu gráfico sea legible para tu audiencia. Además, debido a la forma en que normalmente procesamos la información, comenzando en la esquina superior izquierda y haciendo zigzags con nuestros ojos a través de la pantalla o la página, la estructura del gráfico de barras horizontales hace que nuestros ojos encuentren primero los nombres de las categorías antes que los datos reales.

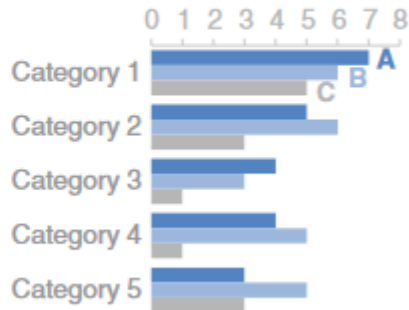
Single series



Two series



Multiple series

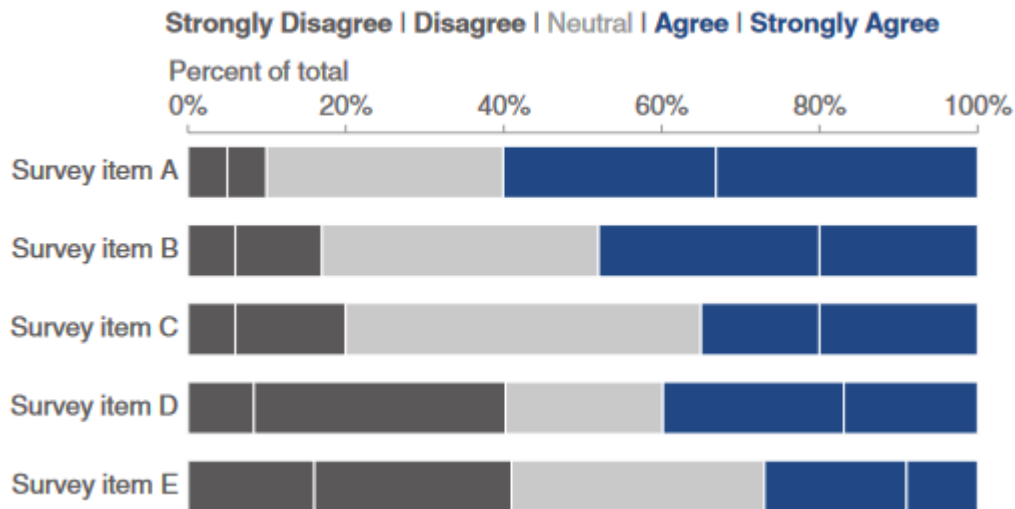




2. Elegir una visual adecuada

c. Gráficos: Barra horizontal apilada / Stacked

Al igual que el gráfico de barras verticales apiladas, los gráficos de barras horizontales apiladas se pueden utilizar para mostrar los totales en diferentes categorías y también dar una idea de las partes subcomponentes. Se pueden estructurar para mostrar tanto valores absolutos como que sumen el 100%.



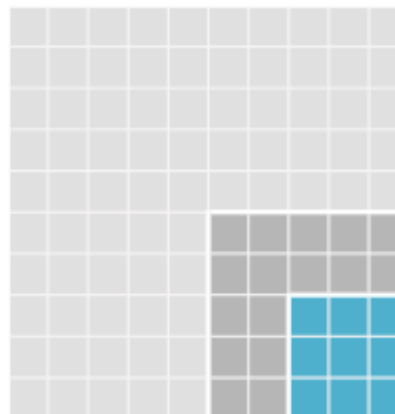


2. Elegir una visual adecuada

c. Gráficos: Area

Los gráficos de área se usan para representar datos numéricos en un plano bidimensional, donde el área bajo la curva o línea indica la magnitud de los valores. Aunque menos efectivos que otros tipos de visualizaciones debido a la dificultad para evaluar valores cuantitativos en dos dimensiones, se utilizan en situaciones donde se necesite visualizar números de magnitudes muy diferentes para aprovechar la dimensión adicional que ofrecen.

Interview breakdown



Out of every 100
phone screens...

we bring 25
candidates onsite
for interviews...

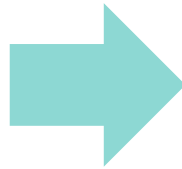
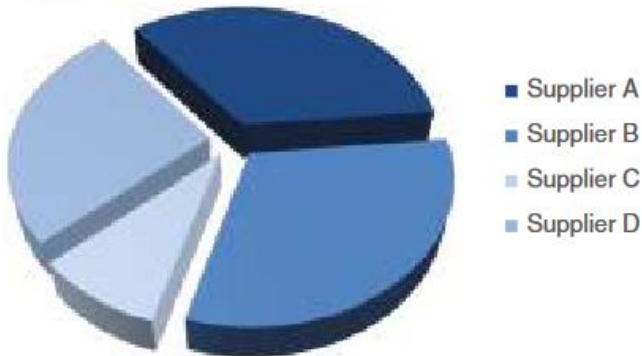
and
extend 9 offers.

2. Elegir una visual adecuada

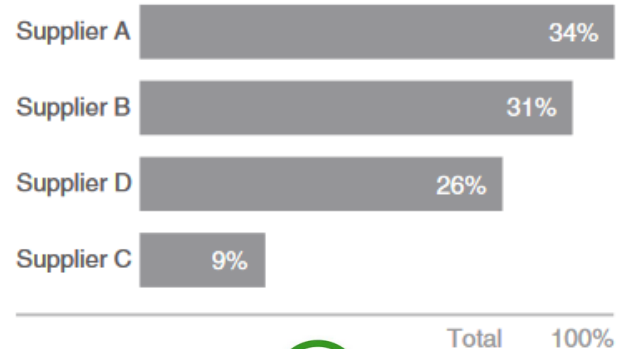
d. Lo que hay que evitar #1

También hay algunos tipos específicos de gráficos y elementos que se deben evitar: gráficos circulares, gráficos en 3D y ejes y secundarios.

Supplier Market Share

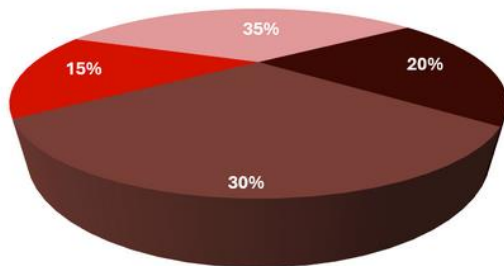
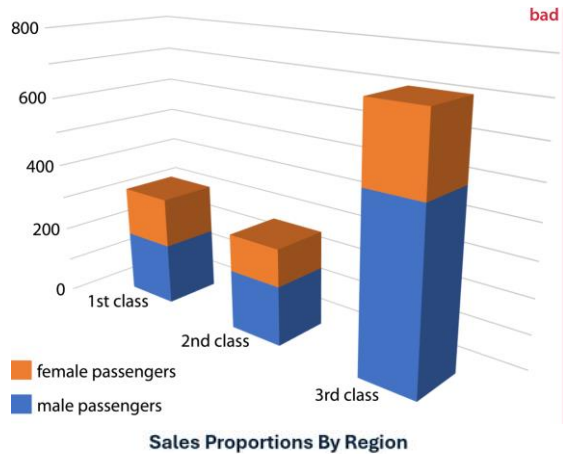


Supplier Market Share



2. Elegir una visual adecuada

d. Lo que hay que evitar #2 | 3D = FALOPA



2. Elegir una visual adecuada

d. Lo que hay que evitar #3

Segundo eje y-secundario

El uso de un eje y secundario generalmente no es una buena idea. A veces puede ser útil representar datos que están en unidades completamente diferentes contra el mismo eje x. Esto a menudo da lugar al uso de un eje y secundario: otro eje vertical en el lado derecho del gráfico.

Secondary y-axis

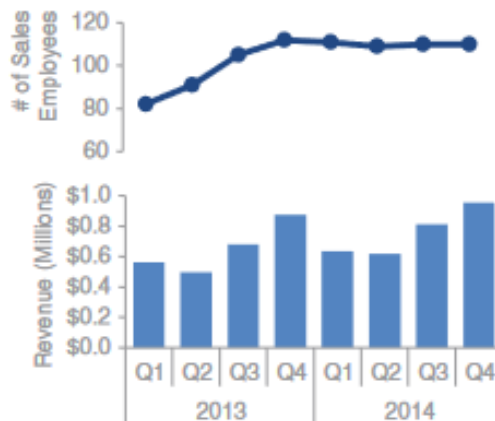


1. No mostrar el segundo eje y. En su lugar, etiqueta directamente los puntos de datos que pertenecen a este eje.
2. Separar los gráficos verticalmente y tener un eje y separado para cada uno (ambos a lo largo de la izquierda), pero aprovechar el mismo eje x en ambos.

Alternative 1: label directly



Alternative 2: pull apart vertically





2. Elegir una visual adecuada

Conclusiones

- En muchos casos, no hay una única representación visual correcta; a menudo hay diferentes tipos de visuales que podrían satisfacer una necesidad específica.
- Es fundamental tener claramente articulada esa necesidad, basándose en la pregunta rectora **¿Qué necesitas que tu audiencia sepa?**
- La clave es elegir una representación visual que permita transmitir claramente el mensaje.
- Si te preguntas "¿Cuál es el gráfico adecuado para mi situación?", la respuesta siempre es la misma: aquel que sea más fácil de leer para tu audiencia.
- Una forma fácil de probar esto es mostrar tu visualización a un amigo o colega y que describan dónde se enfocan, qué ven, qué observaciones hacen y qué preguntas tienen. Esto te ayudará a evaluar si tu visualización cumple su objetivo o, en caso contrario, te indicará en qué áreas debes hacer cambios.



3. Eliminar el desorden

Sobrecarga cognitiva

- La carga cognitiva es el **esfuerzo mental** para procesar información.
- Experiencias comunes de carga cognitiva incluyen diapositivas abrumadoras y gráficos confusos.
- Queremos evitar sobrecargar a nuestra audiencia con información innecesaria.
- La carga cognitiva afecta la capacidad de comprensión de la audiencia.
- Como **diseñadores de información**, debemos ser conscientes de cómo utilizamos el poder de **procesamiento mental de nuestra audiencia**.
- Es crucial evitar la carga cognitiva excesiva que no contribuye a la comprensión de la información.

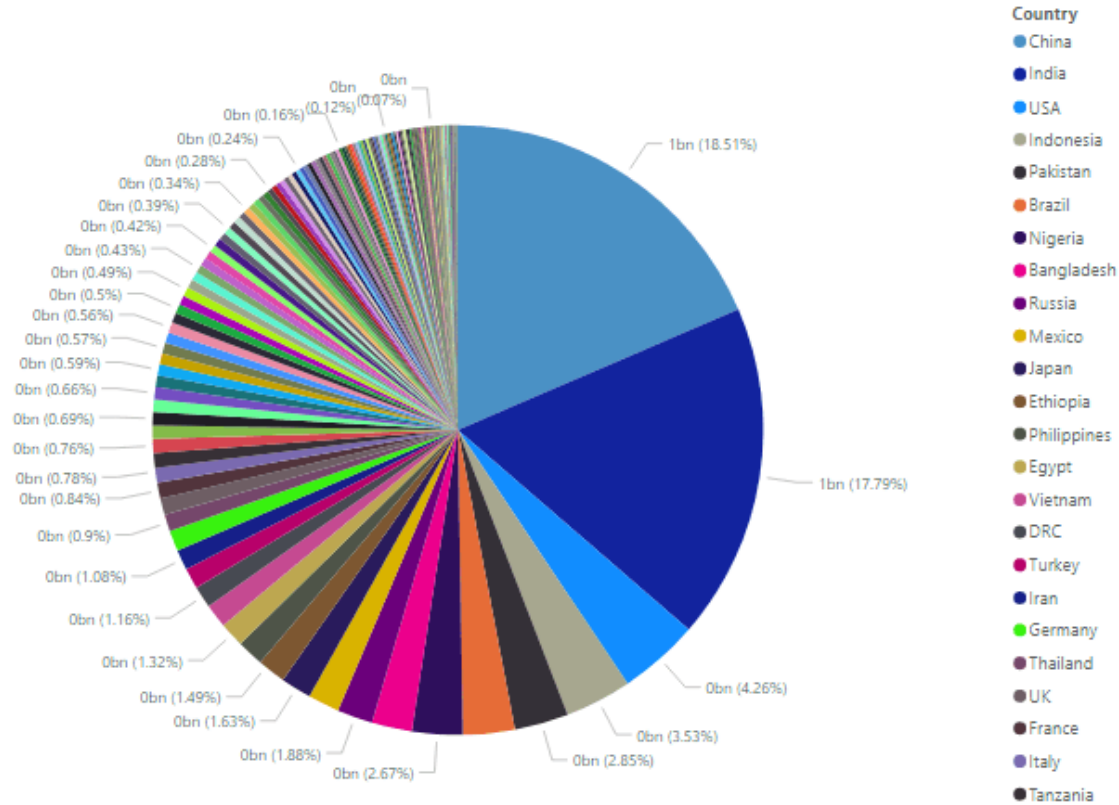


3. Eliminar el desorden

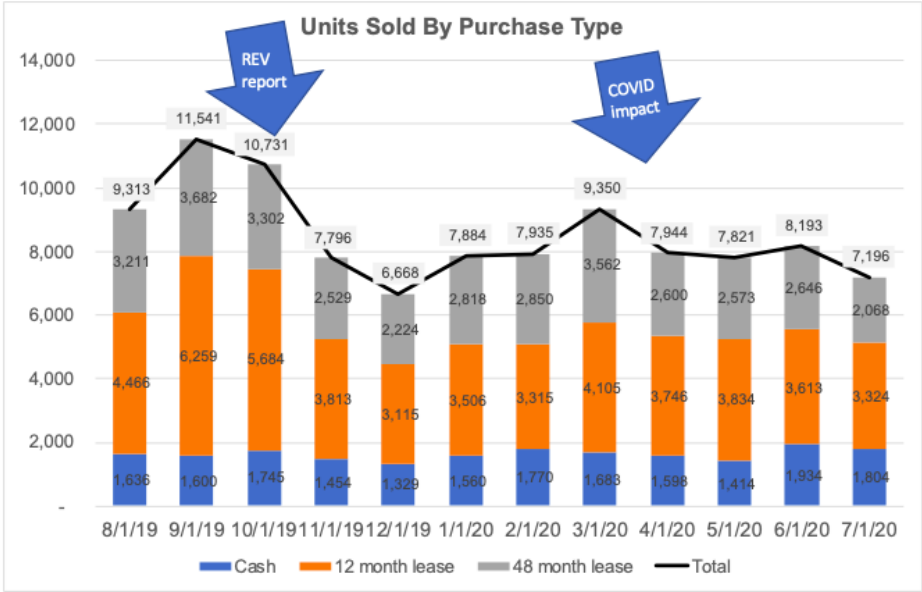
Siempre es complejo visualizar muchas variables o dimensiones. Tanto sea para entender o comunicar, hay algunos principios básicos para ayudarnos a lograr visualizaciones bien diseñadas, claras e intelectualmente honestas sin empantanarse en el intento. Para ello debemos:

- Evitar la ambigüedad
- Evitar la distorsión (es decir, el engaño)
- Evitar la distracción

Population by Country and Country

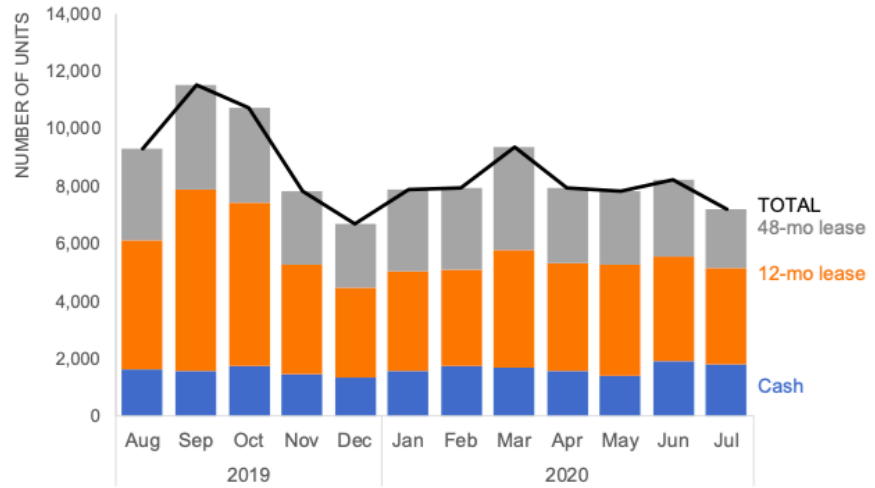


BEFORE



AFTER

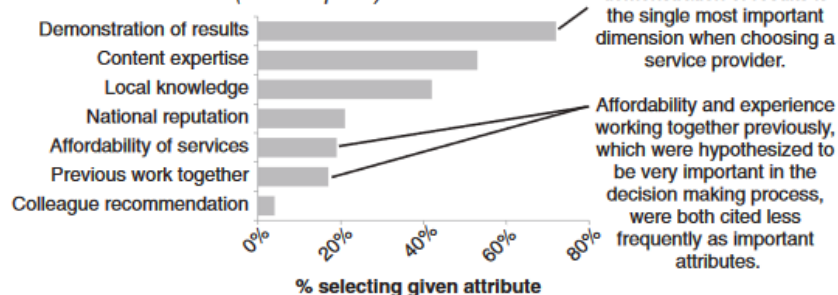
Sales over time by purchase type





Demonstrating effectiveness is most important consideration when selecting a provider

In general, what attributes are the most important to you in selecting a service provider?
(Choose up to 3)

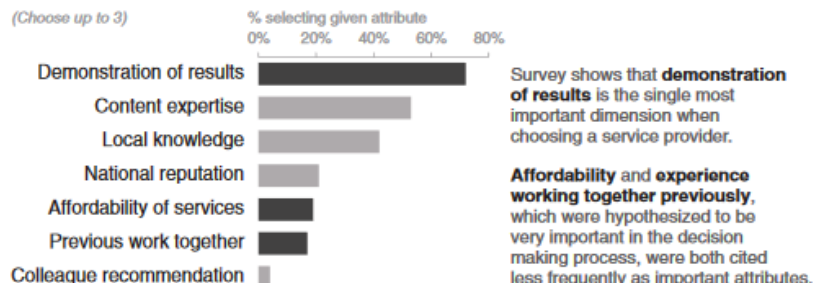


Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes** are the most important to you in selecting a service provider?

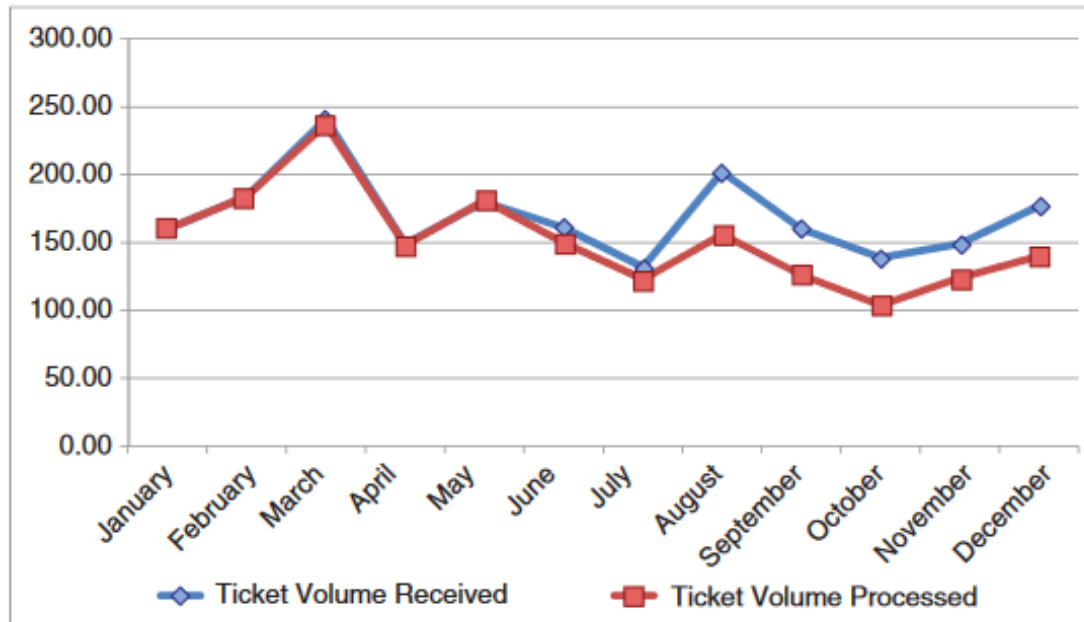
(Choose up to 3)



Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

3. Eliminar el desorden

Paso a paso #1

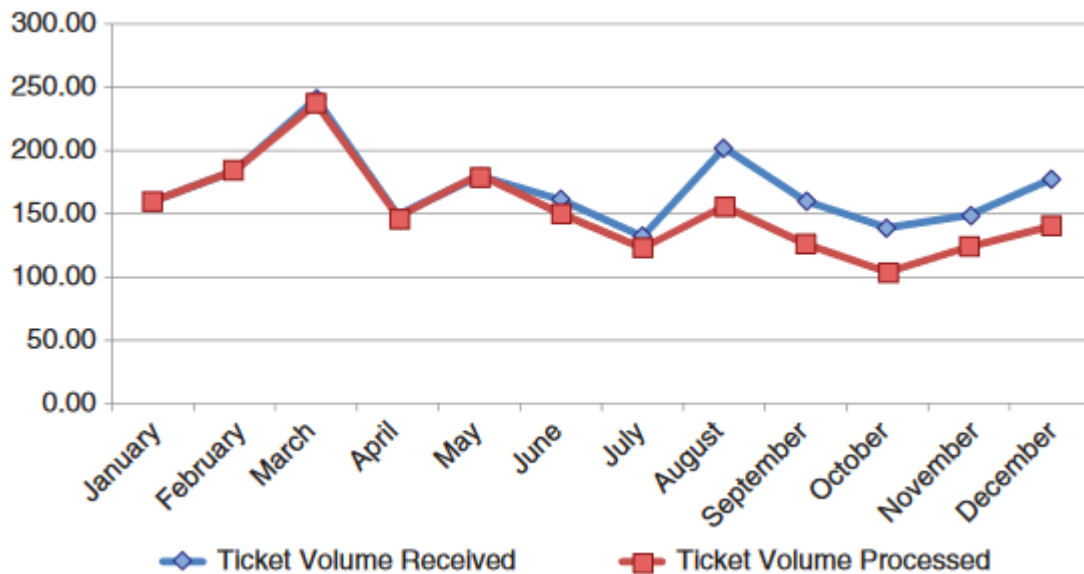




3. Eliminar el desorden

Paso a paso #2

1. Remover el borde

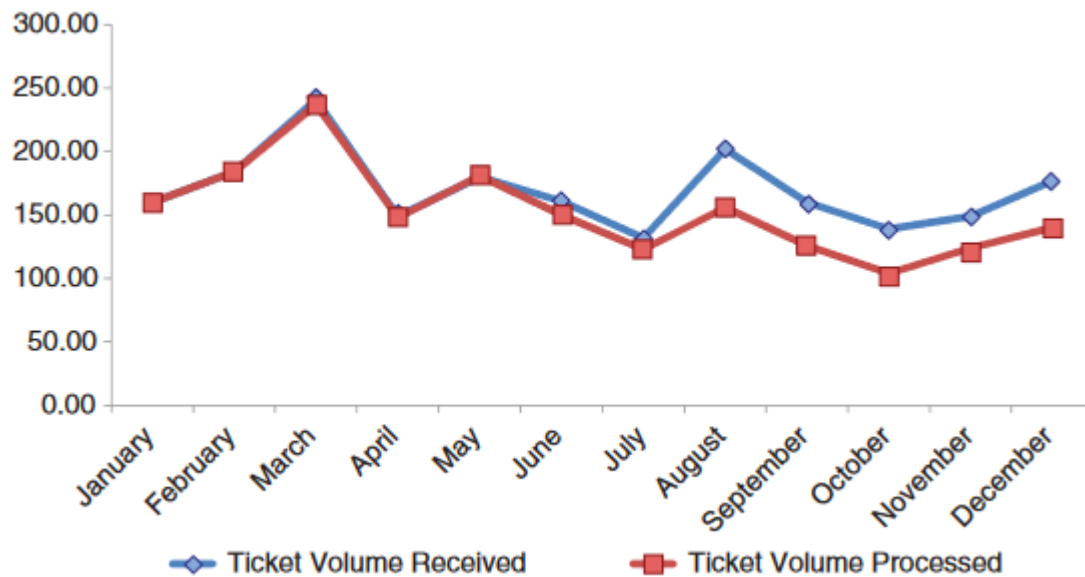




3. Eliminar el desorden

Paso a paso #3

2. Remover la grilla

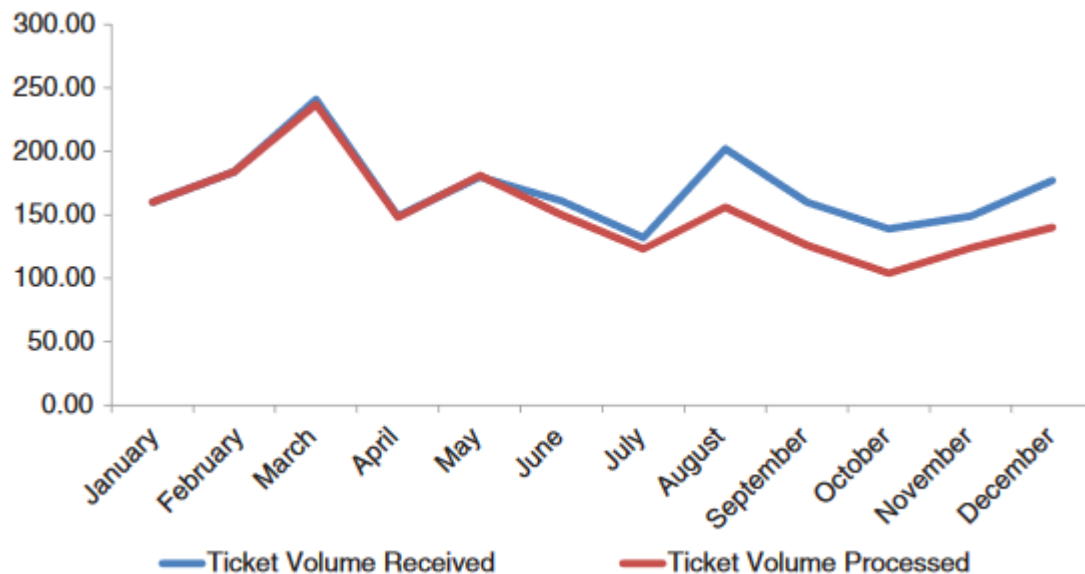




3. Eliminar el desorden

Paso a paso #4

3. Remover los puntos

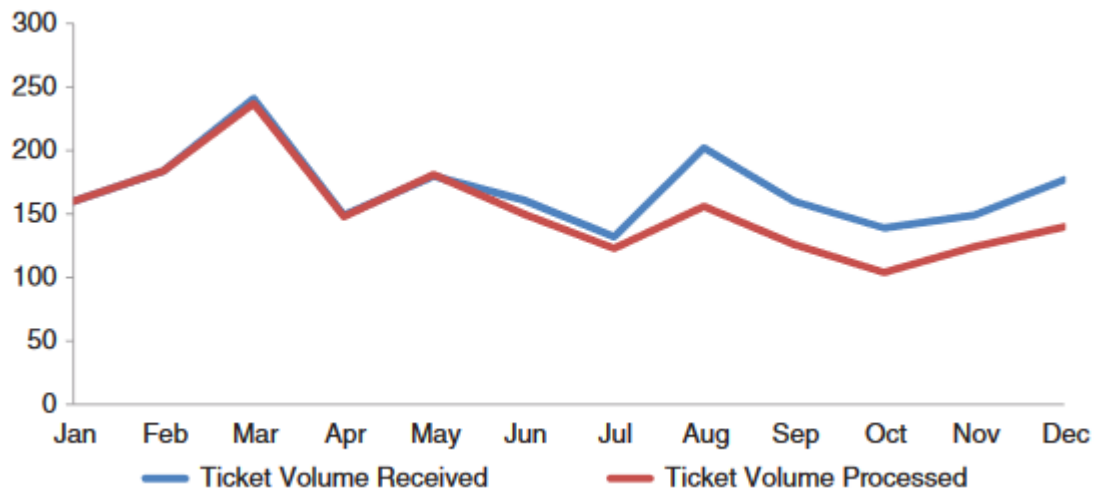




3. Eliminar el desorden

Paso a paso #5

3. Simplificar los ejes X-Y

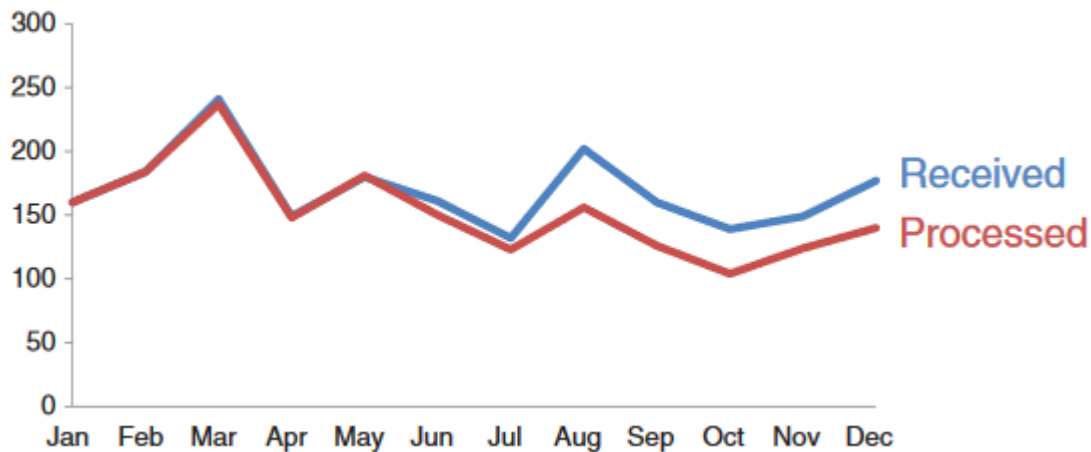




3. Eliminar el desorden

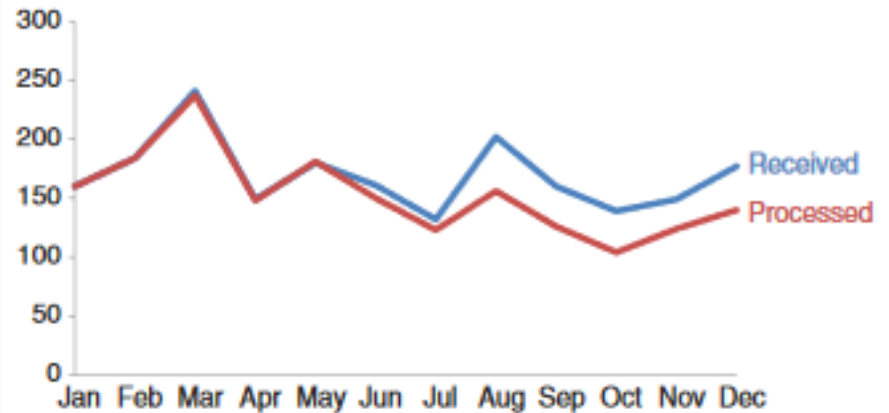
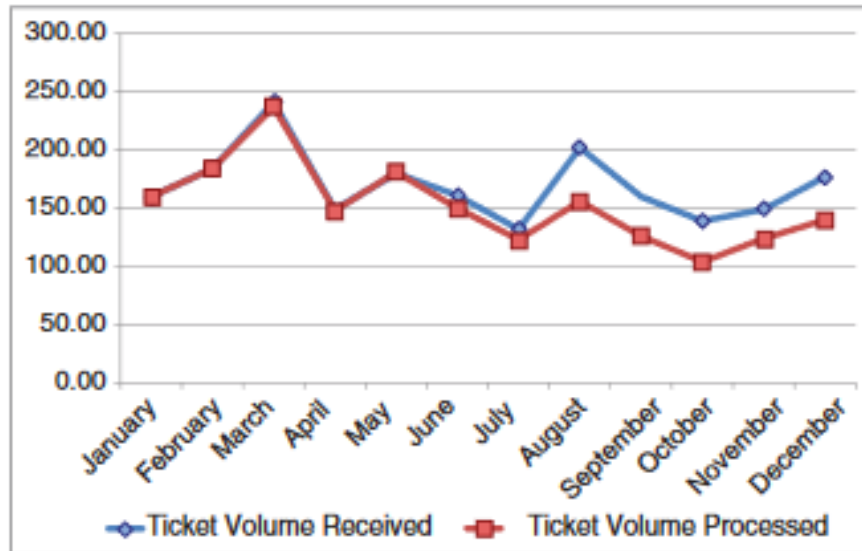
Paso a paso #6

4. Simplificar la leyenda



3. Eliminar el desorden

Paso a paso #7



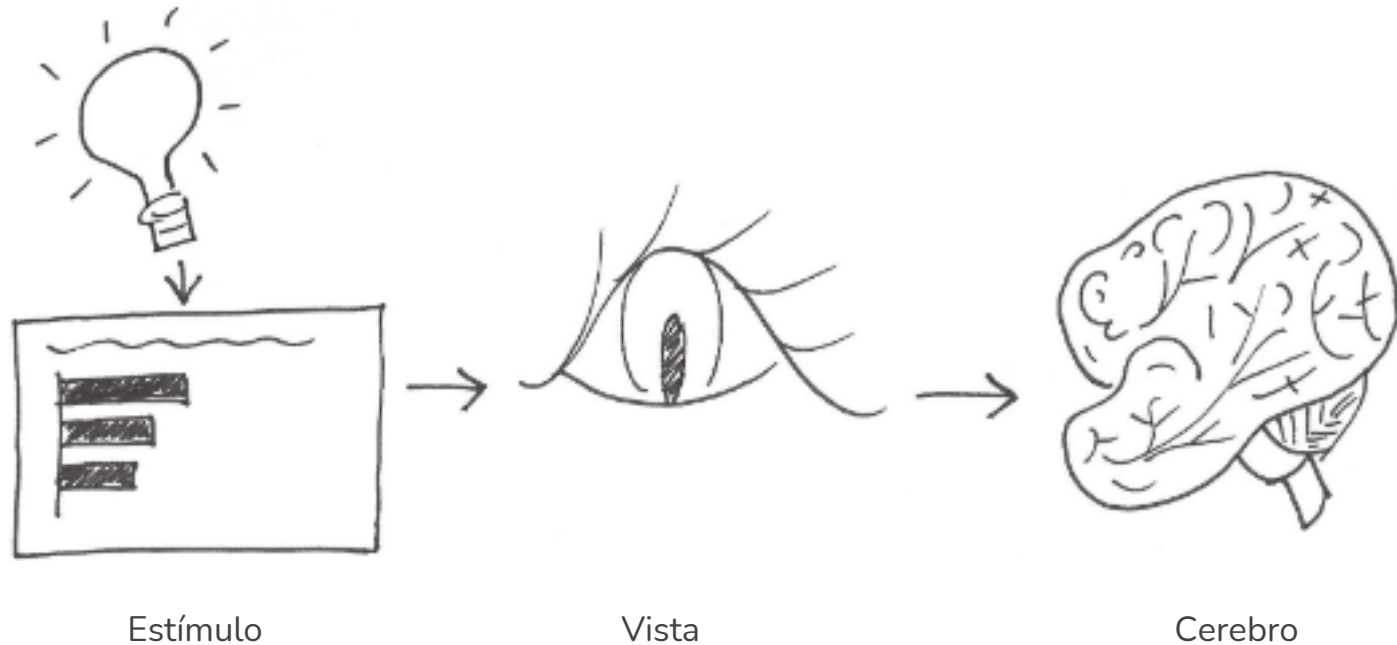


3. Eliminar el desorden

Conclusiones

- **Presentar información genera carga cognitiva en la audiencia.**
- El desorden visual aumenta la carga cognitiva y dificulta la transmisión del mensaje.
- Utiliza la alineación y el espacio en blanco para facilitar la interpretación de tus visuales.
- Utiliza el contraste estratégicamente.

4. Percepción y Diseño





4. Percepción y Diseño

Atributos de la atención

756395068473

658663037576

860372658602

846589107830

¿Cuántos número 3 hay
en la figura?



4. Percepción y Diseño

Atributos de la atención

756**3**9506847**3**
65866**3**0**3**7576
860**3**72658602
8465891078**3**0

6!

El atributo de la intensidad del color, en este caso, hace que los "3" sean lo único que destaca como distintivo del resto. Nuestro cerebro es rápido para percibir esto y es notable. Significa que, si utilizamos los atributos estratégicamente, pueden ayudarnos a permitir que nuestra audiencia vea lo que queremos que vean antes de que siquiera sepan que lo están viendo.



4. Percepción y Diseño

Atributos de la atención



Orientation



Shape



Line length



Line width



Size



Curvature



Added marks



Enclosure



Hue



Intensity



Spatial position



Motion



4. Percepción y Diseño

Atributos de la atención en texto

No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Color

What are we doing well? Great Products. **These products are clearly the best in their class.** Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Bold

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Italics

What are we doing well? Great Products. These products are clearly the best in their class. *Replacement parts are shipped when needed.* You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!



4. Percepción y Diseño

Atributos de la atención en texto

Size

What are we doing well? Great Products. These products are the best in their class. Replacement parts are shipped when needed. You sent gaskets

without me having to ask.

Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Outline (enclosure)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

Separate spatially

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask.

Problems are resolved promptly.

Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Underline (added marks)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

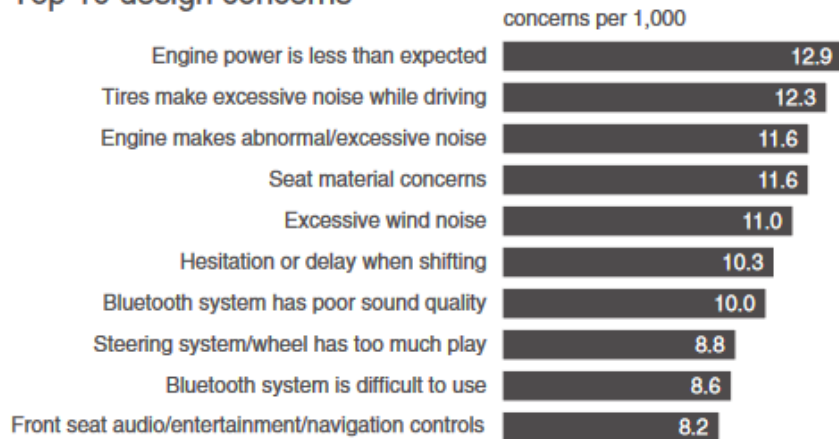
You have a great company – keep up the good work!



4. Percepción y Diseño

Atributos de la atención en gráficos

Top 10 design concerns



Top 10 design concerns

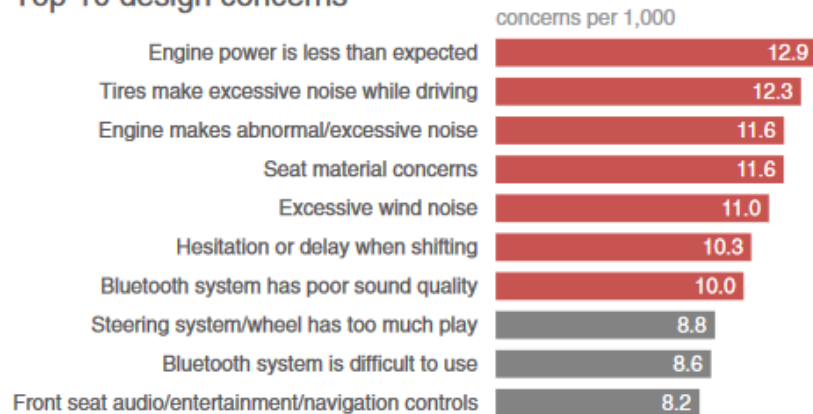


FIGURE 4.8 Leverage color to draw attention

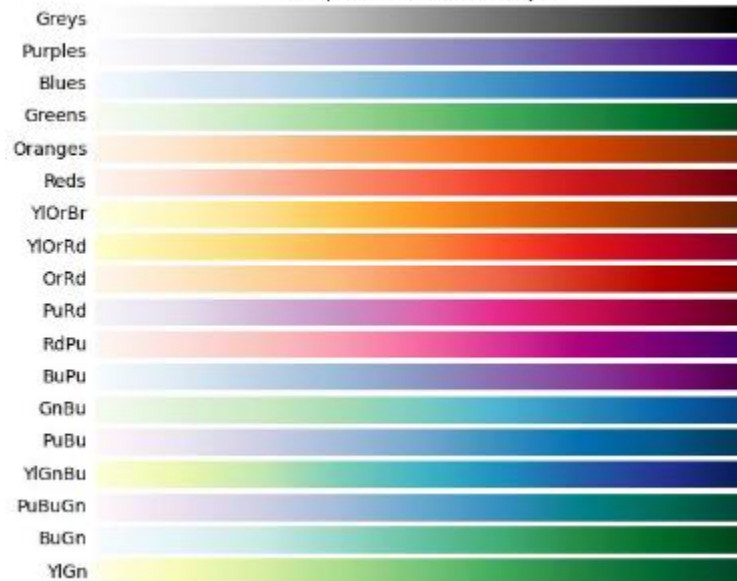


Color

Qualitative colormaps



Sequential colormaps



Las paletas también comunican

Las paletas no son solo estéticas. También comunican información. Puede ser positivo como el caso del rojo y azul en los partidos de EEUU. Puede ser de modo negativo estigmatizando (utilizar rojo con sus connotaciones en mapas de pobreza, distribución del ingreso, etc) o el caso de usar rosa y celeste para visualizar por sexo y como puede reproducir estereotipos nocivos.



<https://blog.datawrapper.de/gendercolor/>

El daltonismo es más común de lo que pensamos

El 5% de la población tiene daltonismo. En una audiencia de 20 personas, es probable que al menos 1 se encuentre con dificultades para distinguir colores en la visualización si no se usan paletas acordes.

Normal vision



Protanopia
Red-Blind



Deuteranopia
Green-Blind



Trianopia
Blue-Blind

