

# VISUALIZACIÓN DE DATOS EN RSTUDIO

MAURICIO HUERTA AGUAR

[MAURICIO.HUERTA@PUCV.CL](mailto:MAURICIO.HUERTA@PUCV.CL)

ESCUELA DE INGENIERÍA INDUSTRIAL  
PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

CICLO DE MINICURSOS

SOCHE, 28 Y 30 DE OCTUBRE DE 2020

# VISUALIZACIÓN DE DATOS EN RSTUDIO

MAURICIO HUERTA AGUAR

[MAURICIO.HUERTA@PUCV.CL](mailto:MAURICIO.HUERTA@PUCV.CL)

ESCUELA DE INGENIERÍA INDUSTRIAL

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

CICLO DE MINICURSOS

SOCHE, 28 Y 30 DE OCTUBRE DE 2020

# AGENDA

INTRODUCCIÓN A LA VISUALIZACIÓN DE DATOS

BUENAS VS MALAS REPRESENTACIONES

DIMENSIONES Y ESTÉTICA

TIPOS DE GRÁFICOS

PROPÓSITO DE LA VISUALIZACIÓN

INTRODUCCIÓN A GGPLOT2

VISUALIZACIÓN INTERACTIVA

# INTRODUCCIÓN A LA VISUALIZACIÓN DE DATOS

# IMAGÍNENSE POR UN MOMENTO. . .

- ❖ UN LIBRO ACADÉMICO SIN ILUSTRACIONES
- ❖ UN ARTÍCULO CIENTÍFICO SIN FIGURAS





**EL APRENDIZAJE SE  
VOLVERÍA MÁS DIFÍCIL**

**INCLUSO PODRÍAMOS  
CONCLUIR  
EQUIVOCADAMENTE**



**ERROR**

**LOS COEFICIENTES SON FÁCILES DE  
INTERPRETAR**

**LOS COEFICIENTES SON FÁCILES DE  
INTERPRETAR**

**¿SEGUROS/AS?**

# ¿SEGUROS/AS?

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

# ¿SEGUROS/AS?

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

N = 11

mean of X's = 9.0

mean of Y's = 7.5

equation of regression line:  $Y = 3 + 0.5X$

standard error of estimate of slope = 0.118

t = 4.24

sum of squares  $\sum (X - \bar{X})^2 = 110.0$

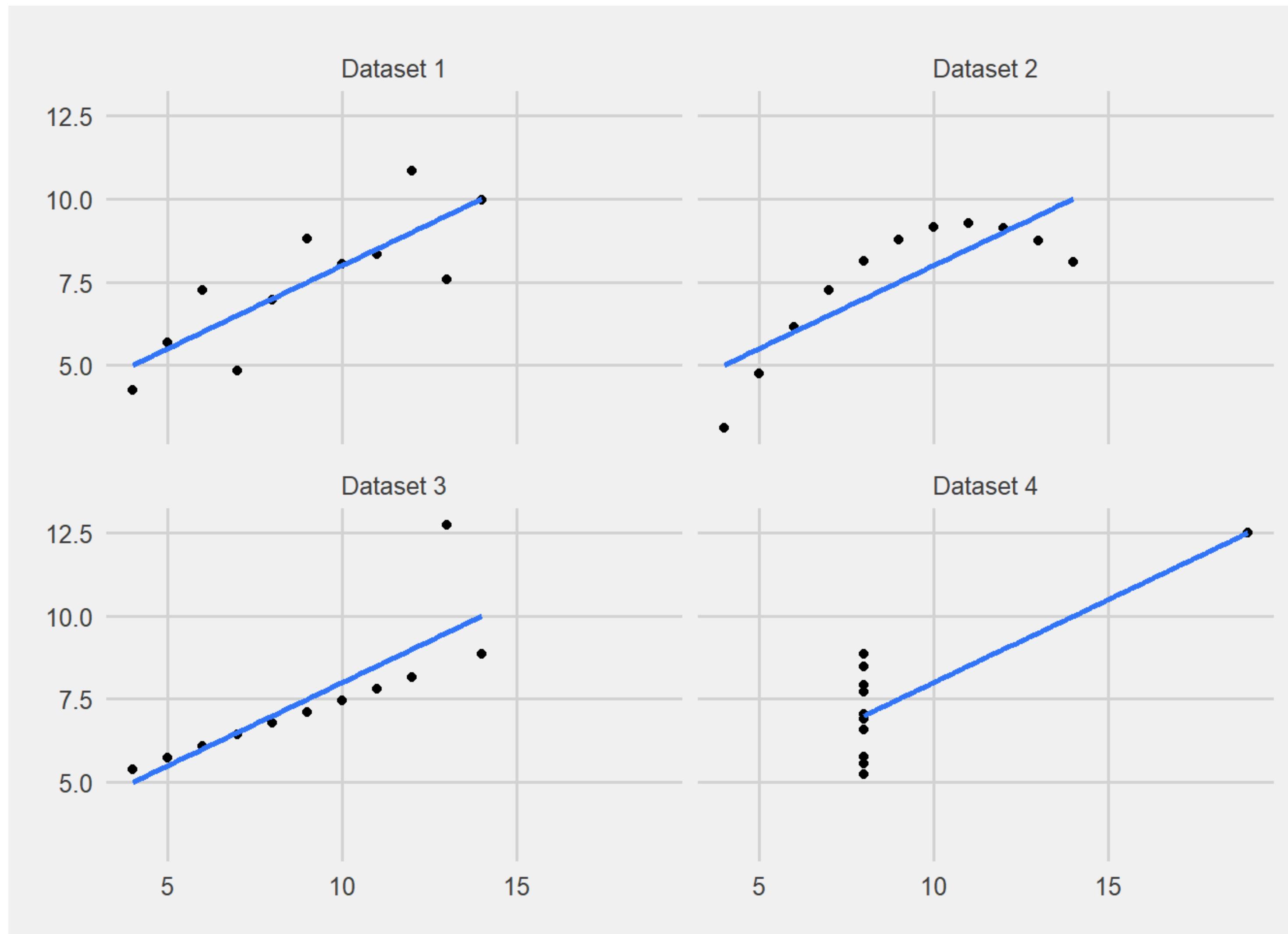
regression sum of squares = 27.50

residual sum of squares of Y = 13.75

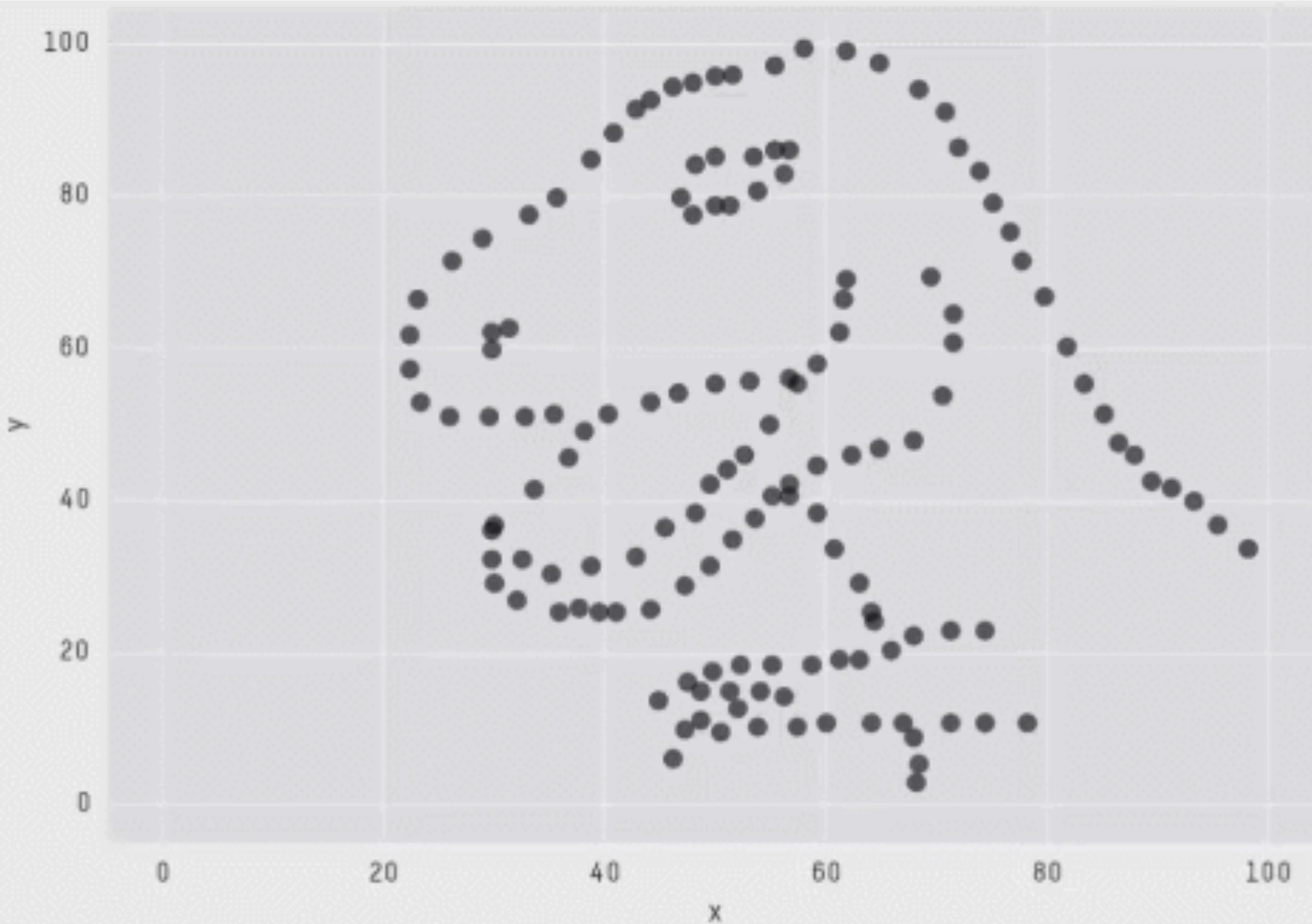
correlation coefficient = .82

$r^2 = .67$

# ¿SEGUROS/AS?



# ¿SEGUROS/AS?



X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD : 16.7649829  
Y SD : 26.9342120  
Corr. : -0.0642526



**LA VISUALIZACIÓN DE DATOS  
ES UNA PARTE FUNDAMENTAL  
PARA COMUNICAR  
CONOCIMIENTO**

# UNA IMAGEN VALE MÁS QUE MIL PALABRAS

Todos lo sabemos. La educación es la madre de todas las batallas. Es la cuna de la igualdad de oportunidades. La principal vía de movilidad social. Es en la educación donde debemos ganar la batalla del futuro. Por eso, la hemos puesto en el corazón de nuestro gobierno.

Pero esta lucha por la calidad y equidad de la educación no se gana con discursos o promesas. Tampoco en la calle. Se gana en la sala de clases y en los hogares.

El éxito está en invertir más, pero también en exigir más. Por eso, estamos construyendo una alianza sólida que requiere el compromiso de todos. De los alumnos y profesores en la sala de clases, de los directores en escuelas y liceos, de los padres en sus hogares, de los alcaldes en sus comunas, de los parlamentarios en el Congreso, y ciertamente del Ministerio de Educación y el Presidente en La Moneda.

Queremos familias informadas y comprometidas. Nos alegra que más de un millón de familias hayan firmado el Contrato de Honor, que refleja su testimonio y compromiso. Se la van a jugar por la educación de sus hijos.

Este fue un año de grandes avances en educación. Parto por agradecer a este Congreso por haber aprobado, con amplias mayoría, la reforma educacional y la nueva institucionalidad, que crea la Agencia y Superintendencia de Educación.

Reitero mi voluntad de que la educación siempre sea tratada como una política de Estado. Podremos tener diferencias, pero el futuro de nuestros niños requiere unidad.

De lo mucho que se ha hecho, quiero destacar la Beca Vocación de Profesor, que premia a los buenos alumnos que quieren estudiar Pedagogía. Escúchenme bien: todo joven con más de 600 puntos en la PSU, que quiera dedicar su vida a ser profesor y a educar a nuestros niños más vulnerables, podrá estudiar Pedagogía ¡completamente gratis! ¡Eso es tomarse en serio la educación!

Destaco también el proyecto de ley que aumentará en un 20 por ciento la subvención preferencial, los 30 liceos de excelencia ya funcionando. El aumento en las horas de Lenguaje y Matemática, las nuevas pruebas Simce de Inglés y Educación Física. También el aumento en la asignación de excelencia pedagógica, que premia a los buenos profesores, los incrementos de sueldos y nuevas atribuciones para los directores de colegios, el bono de hasta 20 millones de pesos para mejorar las condiciones de retiro de los profesores en edad de jubilar y el bono de reconocimiento de hasta dos millones de pesos para los profesores ya jubilados con pensiones bajas, entre otros.

El esfuerzo de años ya está rindiendo frutos. La última prueba Simce nos dio dos grandes alegrías. Tanto en calidad como en equidad. Mención especial merece el aumento en el Simce de Lectura. Y también la reducción significativa en la brecha de resultados entre los estudiantes de colegios particulares y municipales. Sin duda, un gran logro de nuestros niños y profesores.

Las tareas para este año están claras. Mejoramos en Lectura, pero tenemos que reforzar Matemática. Y para ser en el futuro un país bilingüe, necesitamos un Plan Nacional de Inglés.

La educación pública de excelencia, al nivel del Instituto Nacional o del Carmela Carvajal, tiene que seguir llegando a todas las regiones. Por eso, en marzo próximo inauguraremos otros 30 liceos Bicentenario de Excelencia, multiplicando las oportunidades para los niños meritorios.

Pero no queremos que nadie se quede atrás. Por eso estamos apoyando con un plan especial a los mil colegios donde hay más dificultades de aprendizaje.

Seguiremos aumentando la subvención escolar preferencial para llegar a más sectores de clase media, que hasta ahora siempre han estado postergados. Tendremos un nuevo Simce de Tecnología de la Información. Mejoraremos la alimentación que nuestros niños reciben en las escuelas. Independiente en el

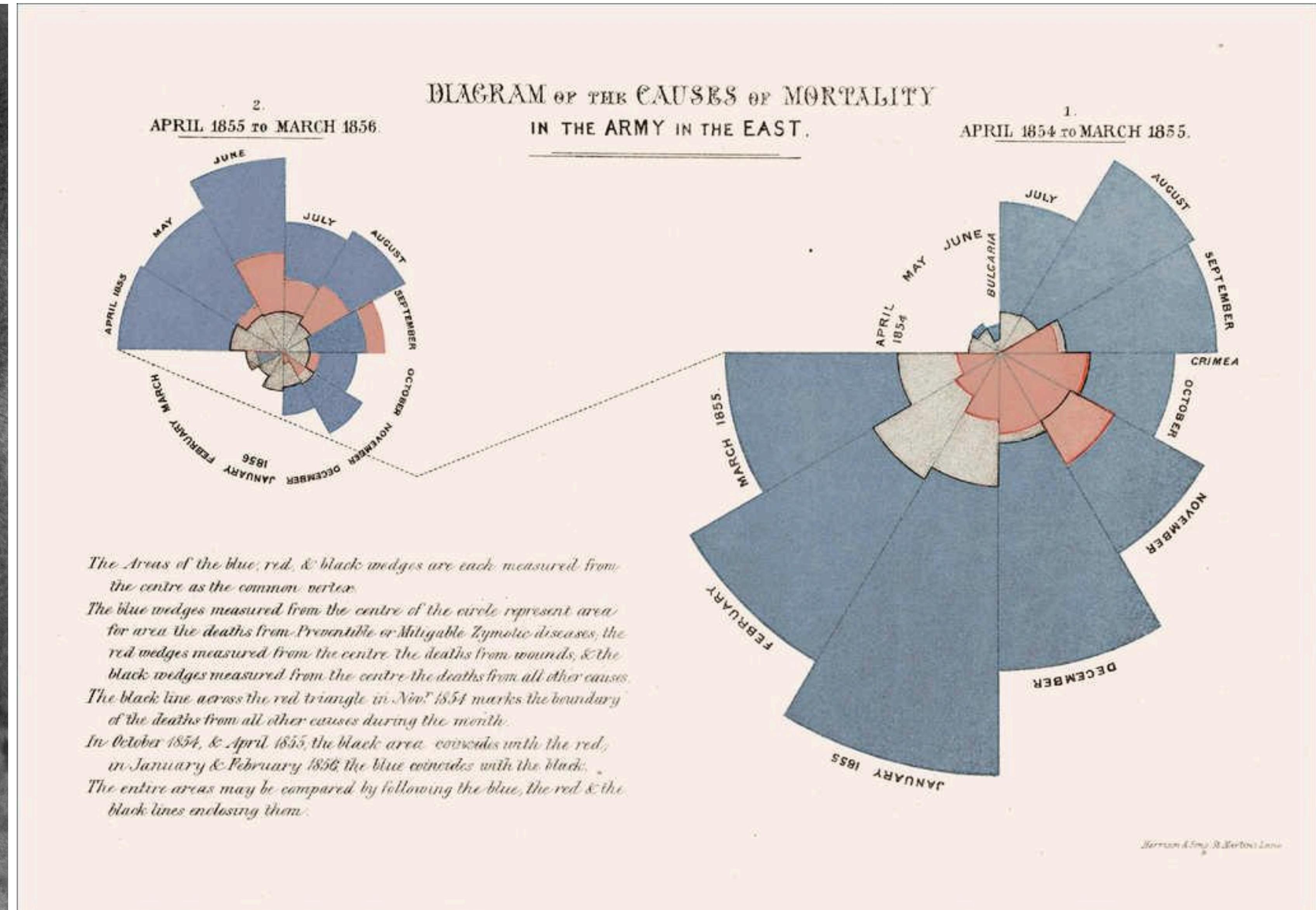


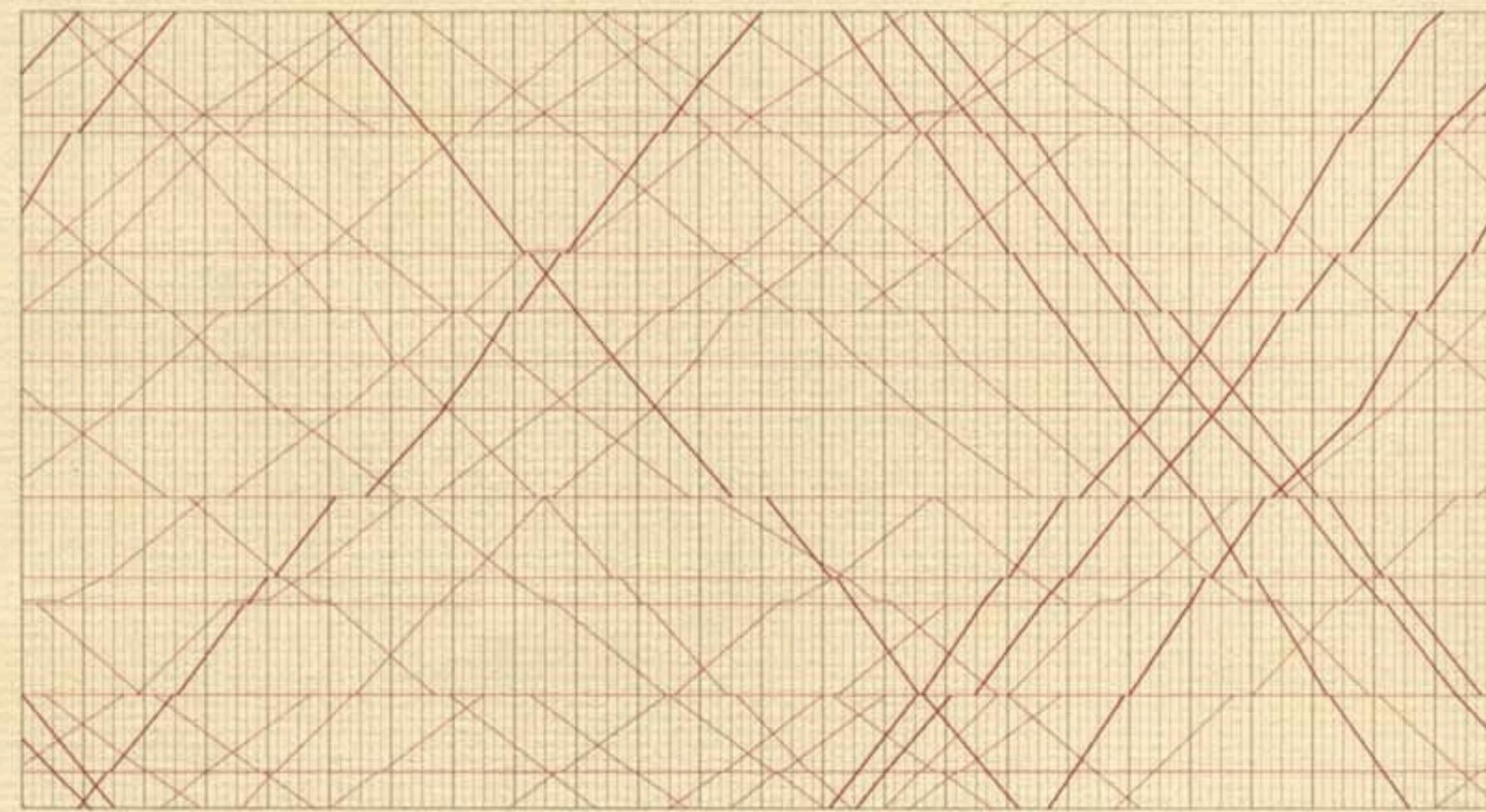
EN LA ACTUALIDAD...

**ESTAMOS RODEADOS DE  
UNA ENORME CANTIDAD  
DE DATOS**

**¿CÓMO PODEMOS  
PRESENTARLOS DE  
MANERA EFICIENTE?**

# INTRODUCCIÓN A LA VISUALIZACIÓN DE DATOS





# The Visual Display of Quantitative Information

EDWARD R. TUFTE

**“LA EXCELENCIA EN LOS  
GRÁFICOS CONSISTE EN IDEAS  
COMPLEJAS COMUNICADAS  
CON CLARIDAD, PRECISIÓN Y  
EFICIENCIA”.**

**EDWARD TUFTE (1983)**

# GRÁFICO

- ▶ **[ETIMOLOGÍA]:** DEL LATÍN “**GRAPHICUS**” (DIBUJADO MAGISTRALMENTE)
- ▶ **[DEFINICIÓN]:** REPRESENTACIÓN DE **DATOS** NUMÉRICOS O DE CANTIDADES QUE SE HACE POR MEDIO DE **DIBUJOS**, **COORDENADAS**, **ESQUEMAS** O **LÍNEAS** QUE REFLEJAN LA **RELACIÓN** QUE EXISTE ENTRE DICHOS DATOS

**¿QUÉ ROL CUMPLEN  
ENTONCES LOS  
GRÁFICOS?**

- ▶ LOS GRÁFICOS PERMITEN EXPLORAR Y APRENDER SOBRE LA ESTRUCTURA DE LOS DATOS RECOPILADOS
- ▶ LAS BUENAS VISUALIZACIONES DE DATOS FACILITAN LA COMUNICACIÓN DE LAS IDEAS Y HALLAZGOS A OTRAS PERSONAS

**¿CÓMO DEBERÍA SER  
UN GRÁFICO FORMAL?**

# UN GRÁFICO DEBERÍA

- ▶ MOSTRAR LOS DATOS.
- ▶ INDUCIR AL VISUALIZADOR A PENSAR EN LO SUSTANCIAL MAS QUE EN LA METODOLOGÍA, DISEÑO GRÁFICO, TECNOLOGÍA DE PRODUCCIÓN U OTRAS COSAS.

# UN GRÁFICO DEBERÍA

- ▶ NO DISTORSIONAR LO QUE “DICEN” DATOS.
- ▶ PRESENTAR MUCHOS NÚMEROS EN UN **ESPACIO PEQUEÑO**.
- ▶ ANIMAR A LA VISTA A COMPARAR DATOS.
- ▶ TENER UN **PROPÓSITO CLARO**

# **PRINCIPIOS DE INTEGRIDAD GRÁFICA**

# PRINCIPIOS DE INTEGRIDAD GRÁFICA

- LA REPRESENTACIÓN DE FRECUENCIAS, PLASMADA FÍSICAMENTE EN LA SUPERFICIE DEL GRÁFICO EN SÍ, DEBE SER DIRECTAMENTE PROPORCIONAL A LAS CANTIDADES NUMÉRICAS EXPRESADAS.

# PRINCIPIOS DE INTEGRIDAD GRÁFICA

- ▶ SE DEBE UTILIZAR UN **ETIQUETADO CLARO, DETALLADO Y COMPLETO** PARA EVITAR LAS **DISTORSIONES GRÁFICAS** Y LA **AMBIGÜEDAD**. ESCRIBA EXPLICACIONES DE LOS DATOS EN EL GRÁFICO MISMO. ETIQUETE EVENTOS IMPORTANTES EN LOS DATOS.

# PRINCIPIOS DE INTEGRIDAD GRÁFICA

- ▶ MOSTRAR **VARIACIÓN DE DATOS**, NO VARIACIÓN DE DISEÑO.
- ▶ EN LAS SERIES TEMPORALES, LAS UNIDADES DE MEDIDA MONETARIA **DESINFLADAS Y ESTANDARIZADAS** SON CASI SIEMPRE MEJORES QUE LAS UNIDADES NOMINALES.

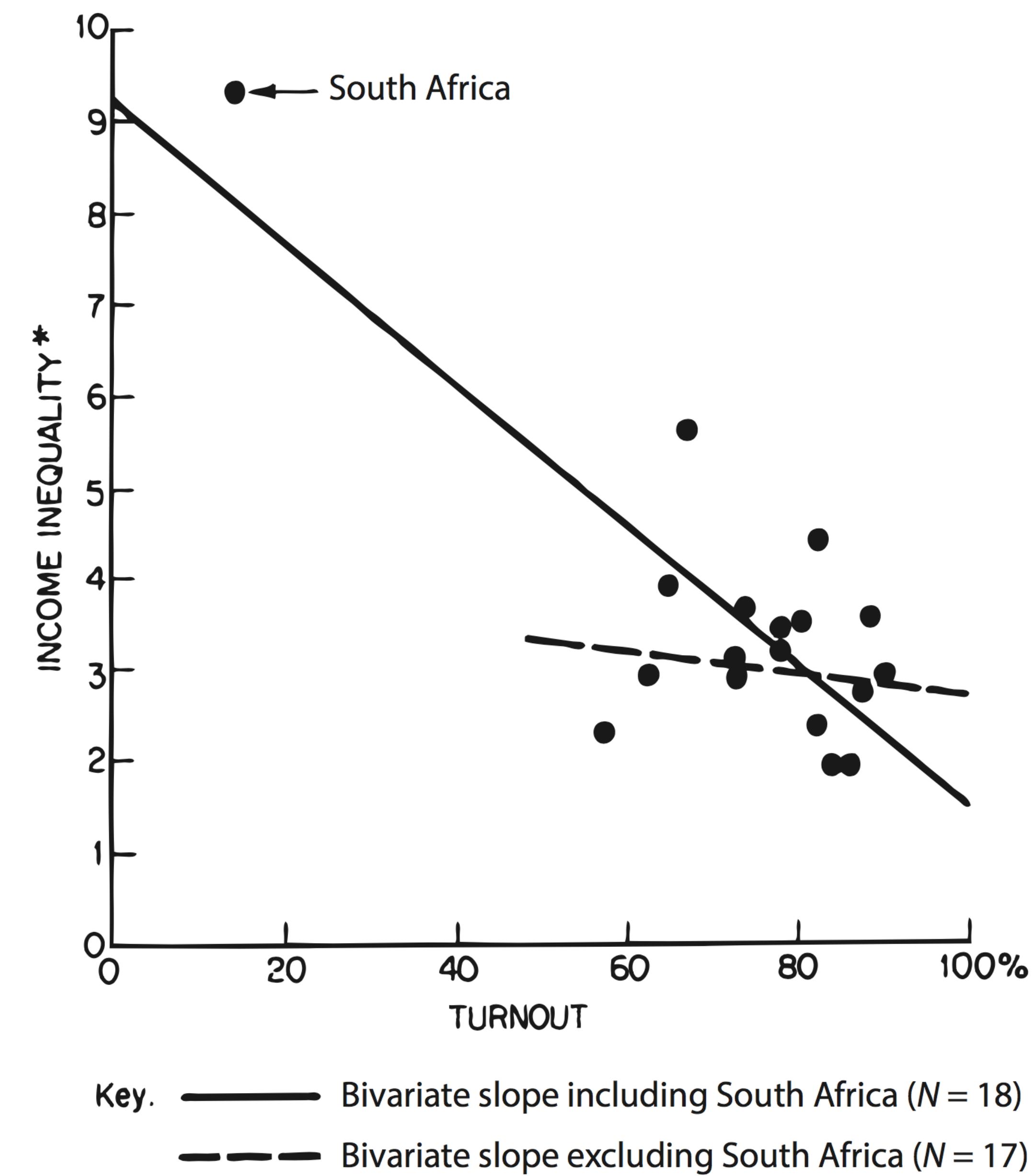
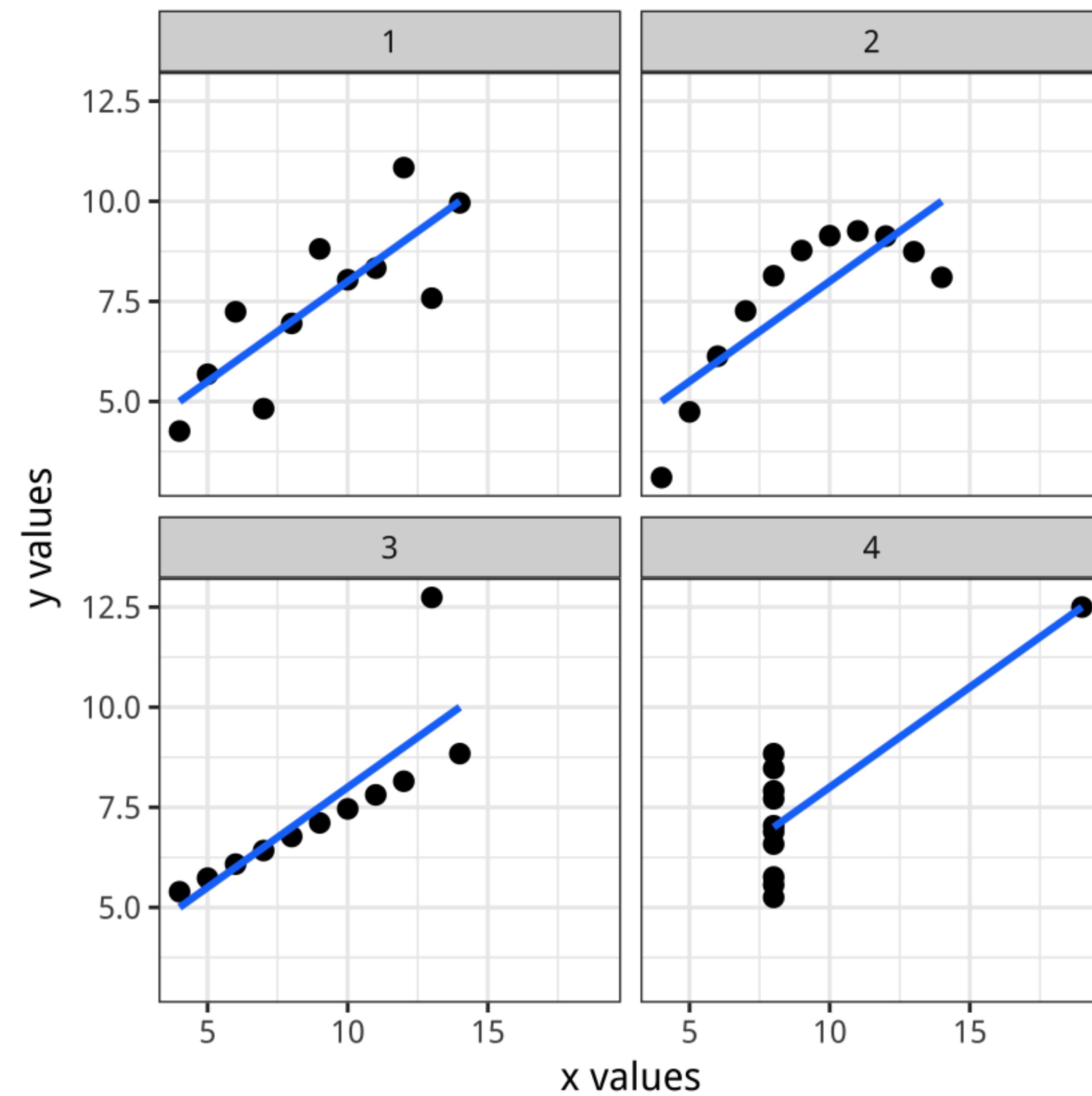
# PRINCIPIOS DE INTEGRIDAD GRÁFICA

- ▶ EL NÚMERO DE DIMENSIONES PORTADORAS DE INFORMACIÓN (VARIABLES) REPRESENTADAS **NO DEBE EXCEDER** EL NÚMERO DE DIMENSIONES EN LOS DATOS.
- ▶ LOS GRÁFICOS NO DEBEN CITAR DATOS **FUERA DE CONTEXTO**.

**UN GRÁFICO DEBE SER  
AUTOIINFORMATIVO**

DEBES MIRAR

TUS DATOS



# **BUENAS VS MALAS REPRESENTACIONES**

**¿QUÉ HACE QUE UNA  
VISUALIZACIÓN SEA  
MALA?**

CONTINENT

Africa   Americas   Asia   Europe   Oceania

O

22.5

45

67.5

90

Years

Africa

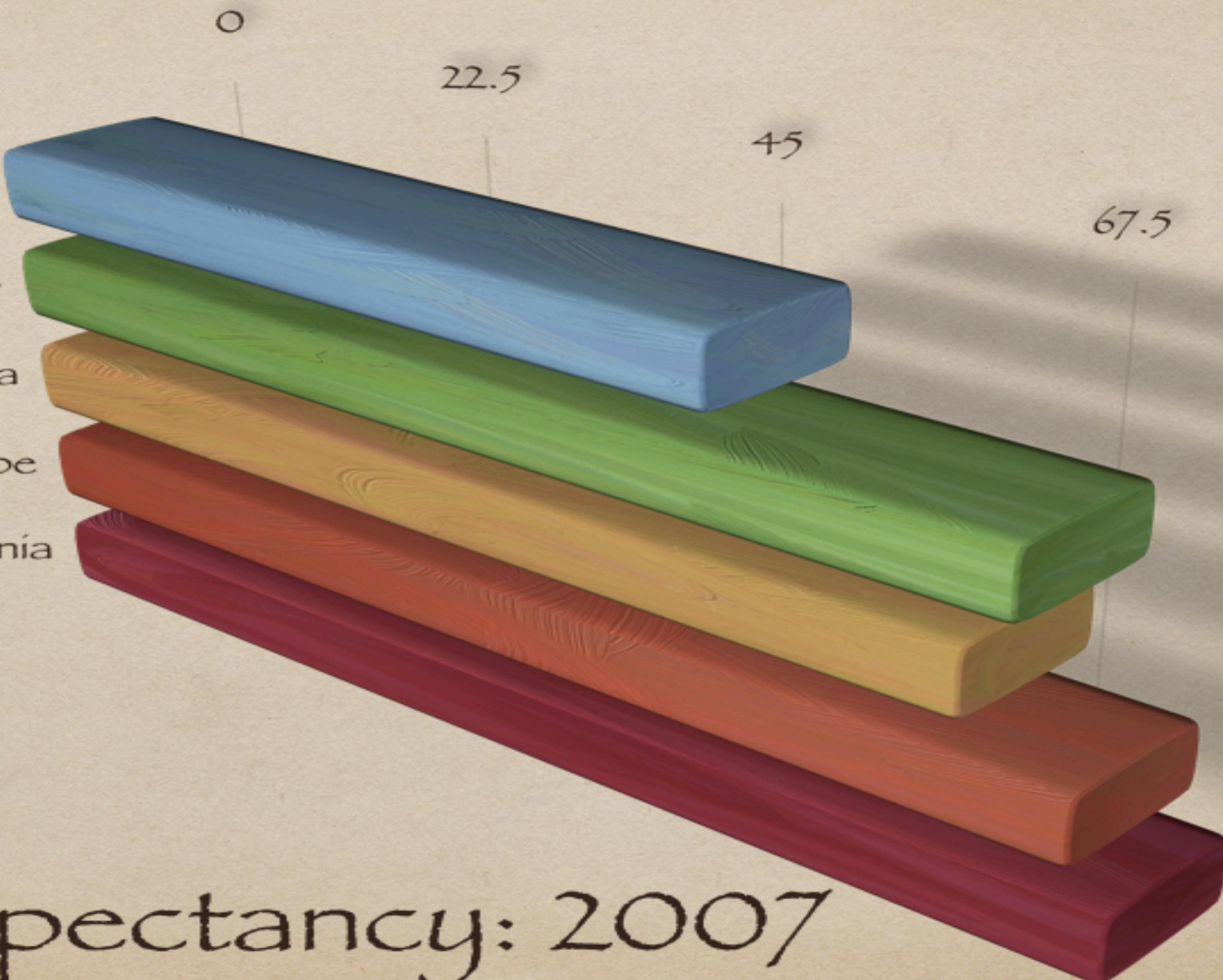
Americas

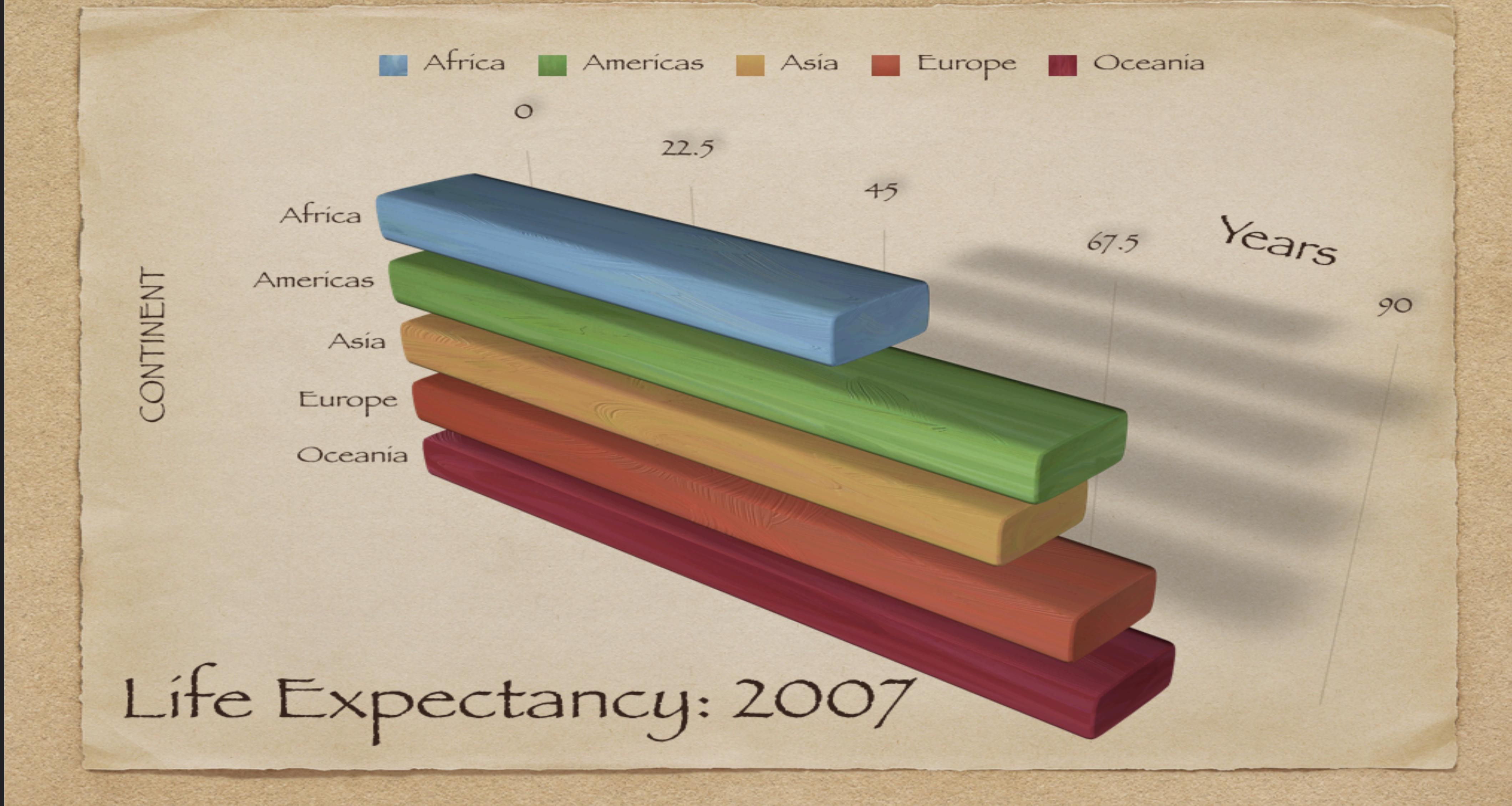
Asia

Europe

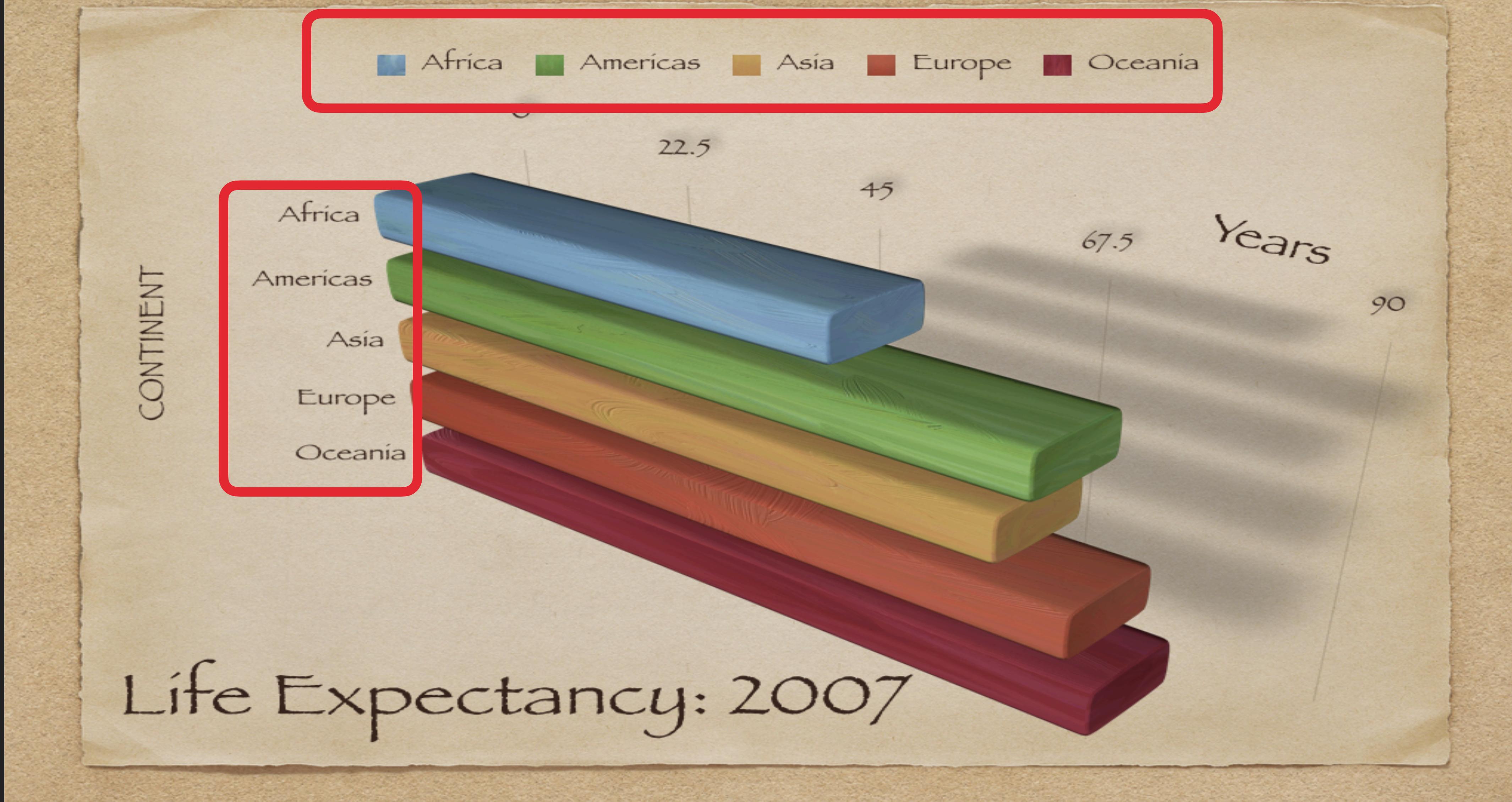
Oceania

Life Expectancy: 2007

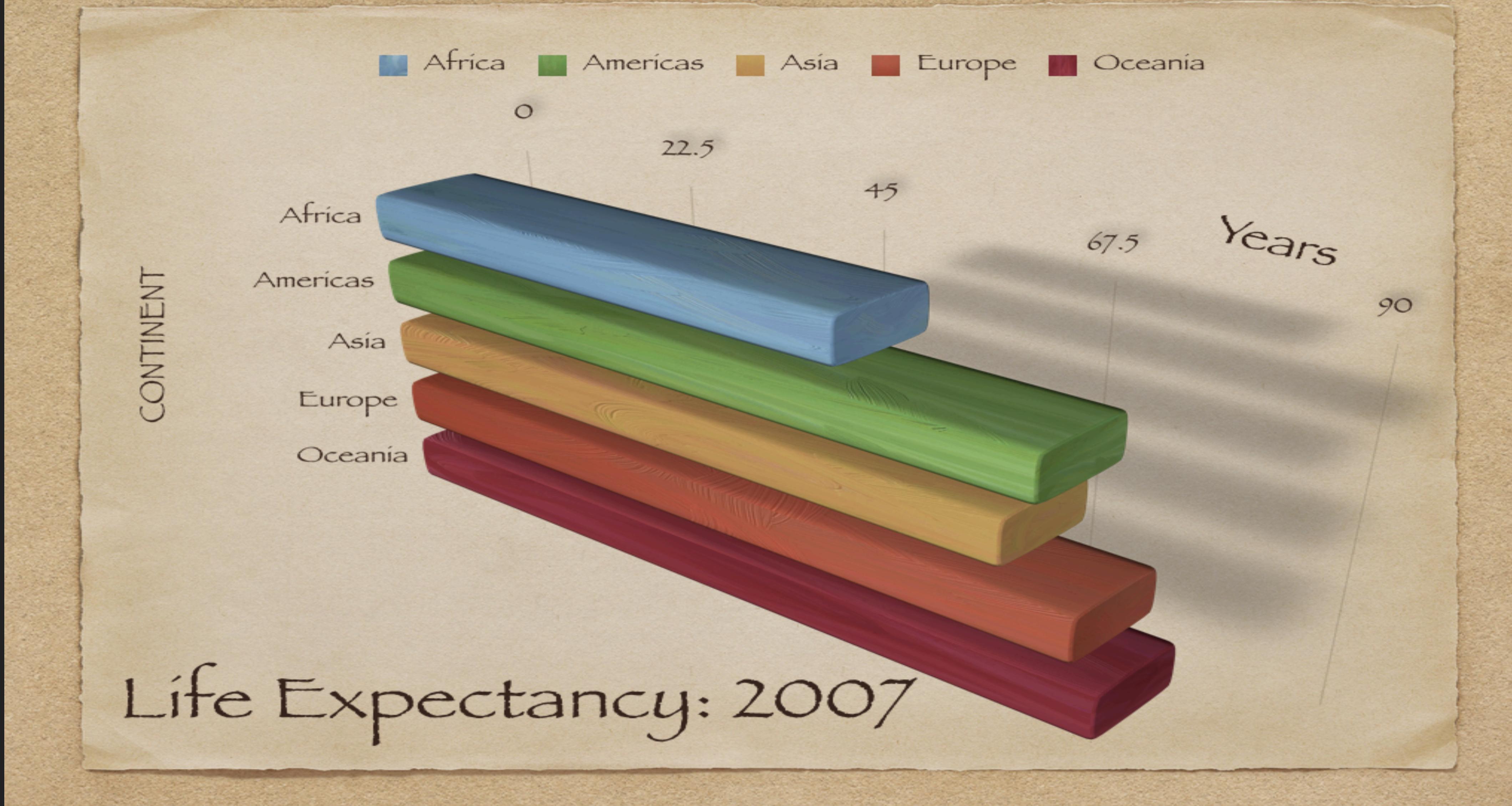




EXCESO DE TINTA GASTADA EN BASURA NO INFORMATIVA

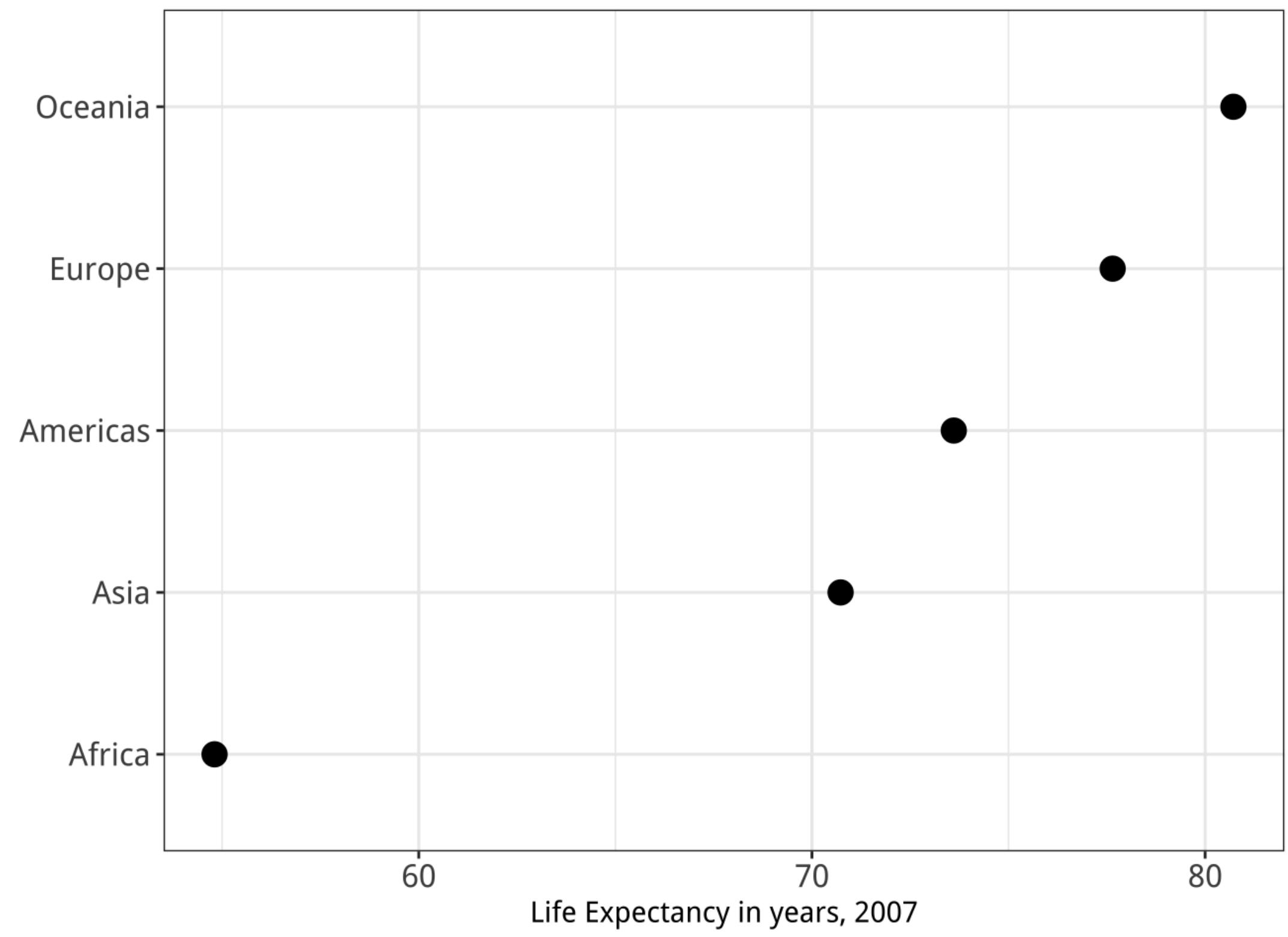
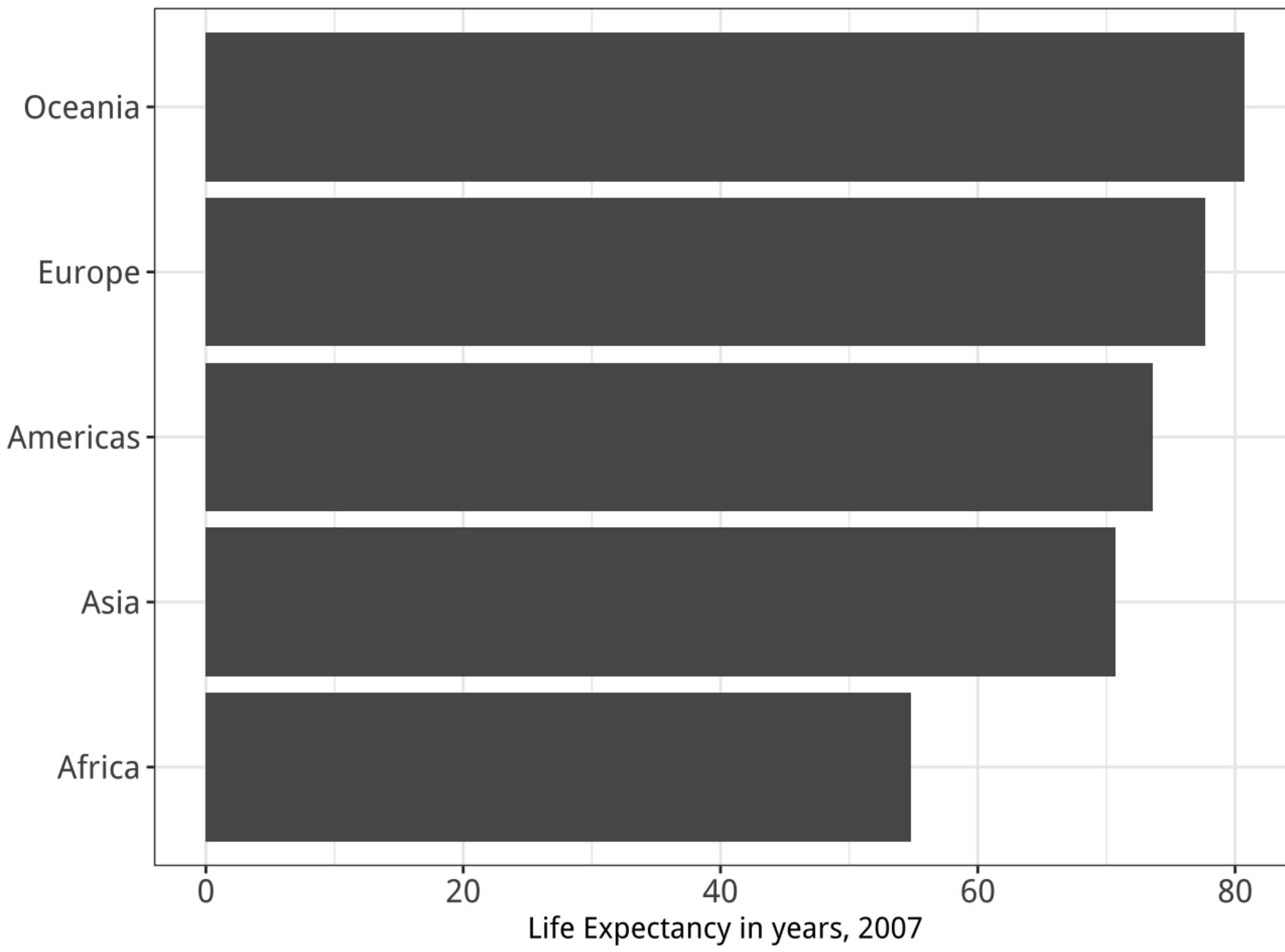


ETIQUETAS REPETIDAS



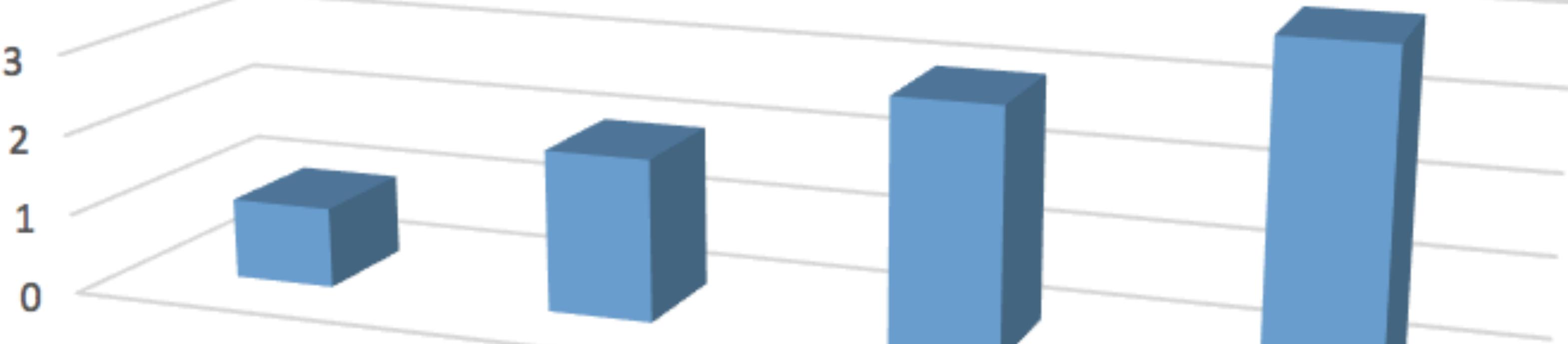
PROBLEMA DE PERCEPCIÓN DEBIDO A BARRAS 3D INNECESARIAS

# UNA MEJOR REPRESENTACIÓN



# **EL PROBLEMA DE LA PERCEPCIÓN**

3D Column Charts in Excel are Awesome



A  
B  
C  
D

# DISTORSIÓN DE LA INFORMACIÓN



# DATOS MAL REPRESENTADOS



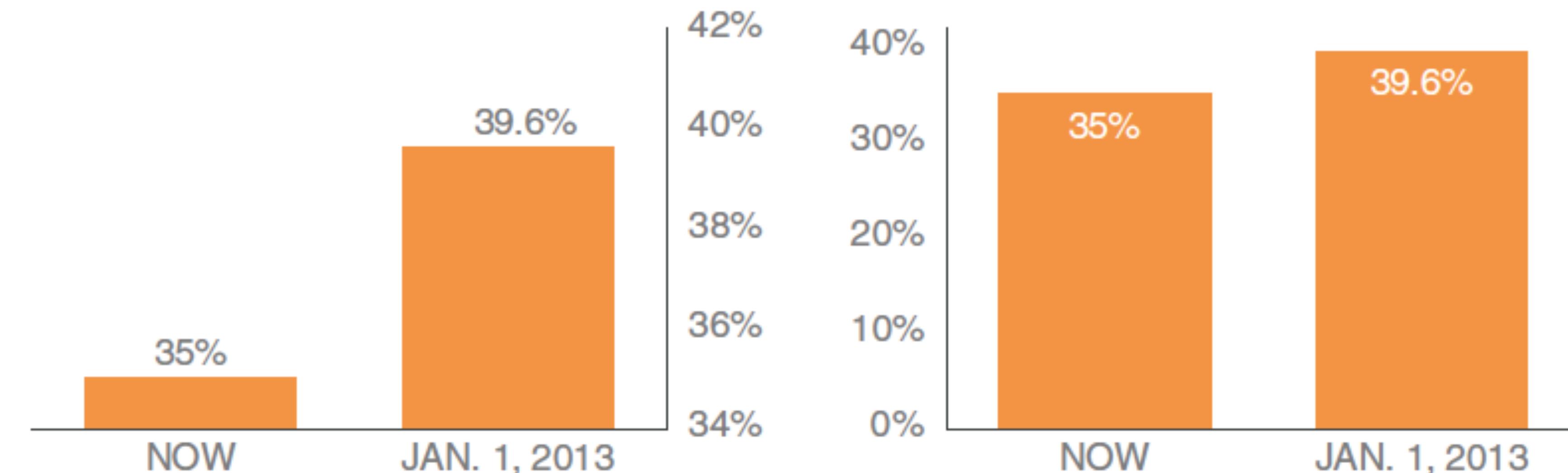
## IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



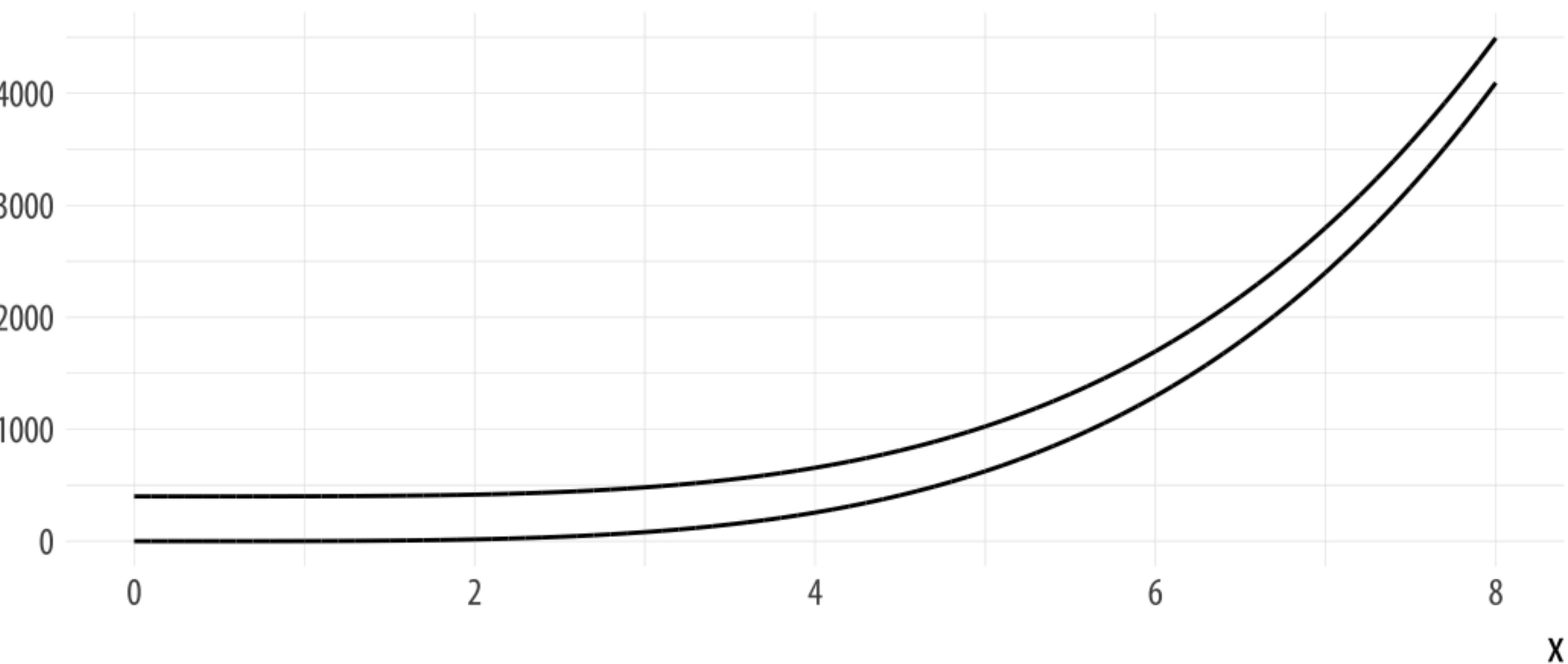
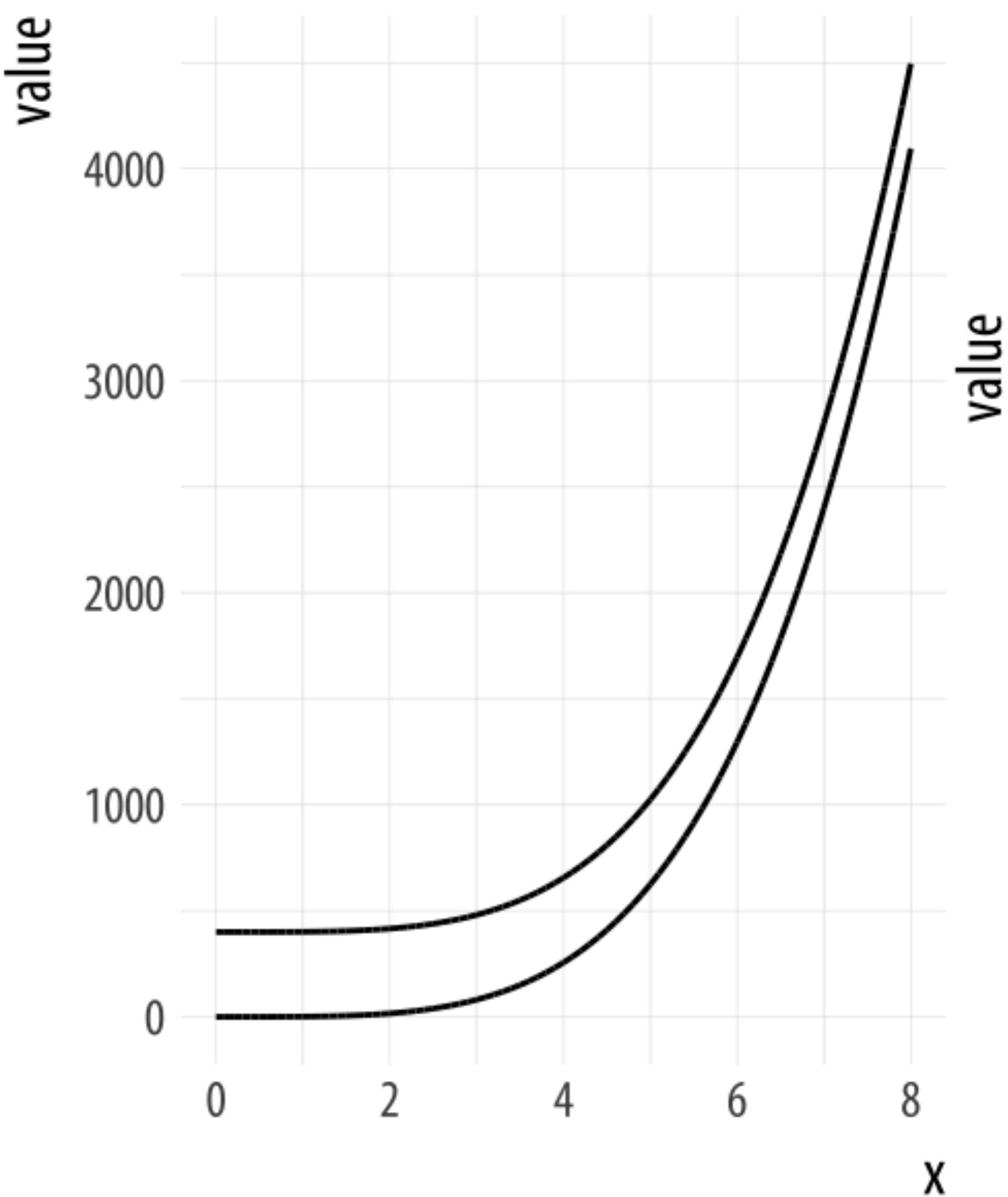
**Non-zero baseline:** as originally graphed

IF BUSH TAX CUTS EXPIRE  
TOP TAX RATE



**Zero baseline:** as it should be graphed

IF BUSH TAX CUTS EXPIRE  
TOP TAX RATE



# NO SÓLO EN EL ANÁLISIS DE DATOS

<https://cutt.ly/mapa-percepcion>

<https://thetruesize.com/>

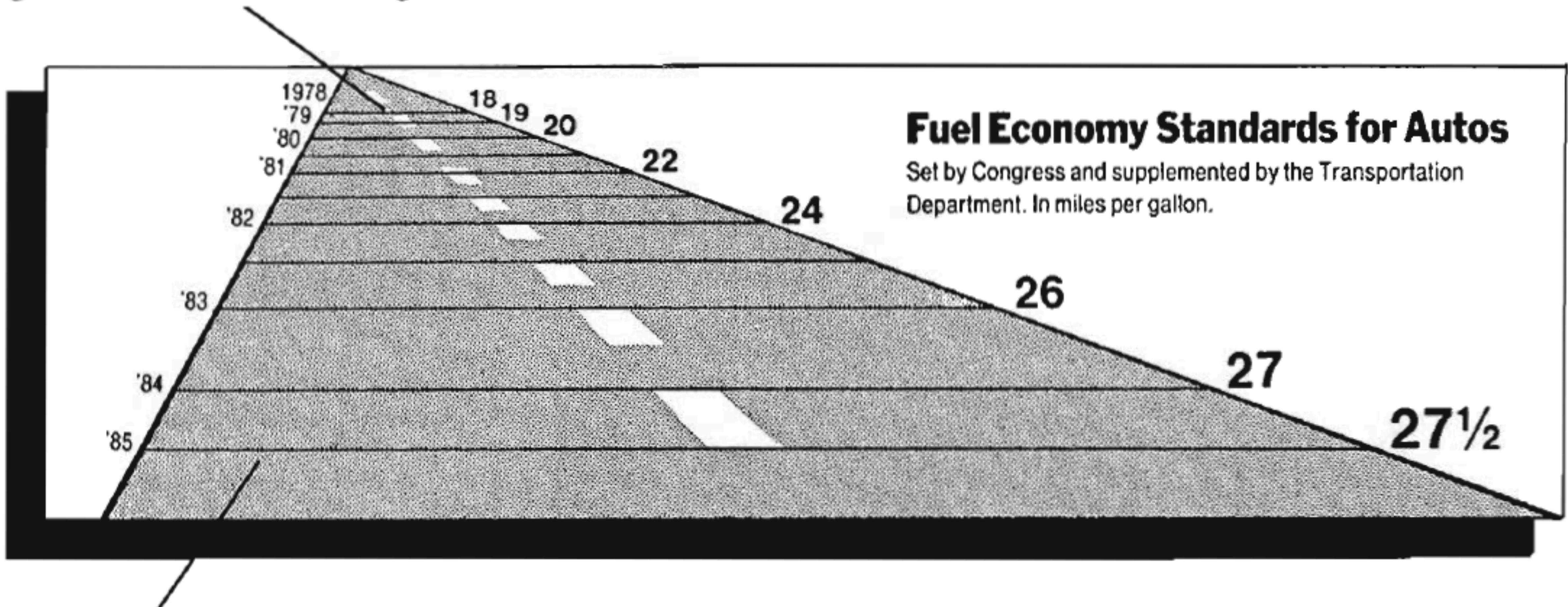
# **EL EXCESO DE BASURA INFORMATIVA**

# MONSTROUS COSTS

Total House and Senate campaign expenditures,  
in millions



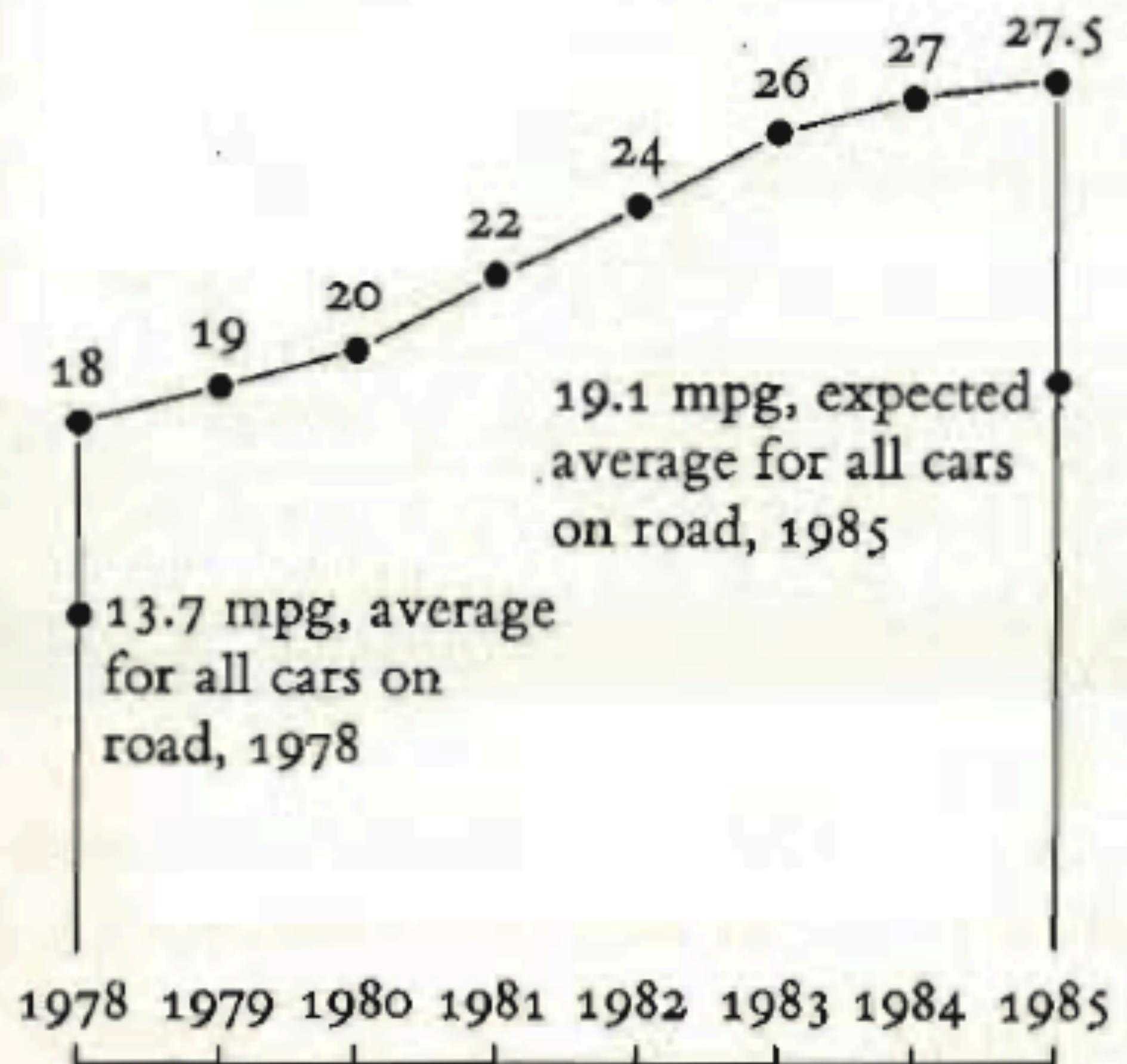
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



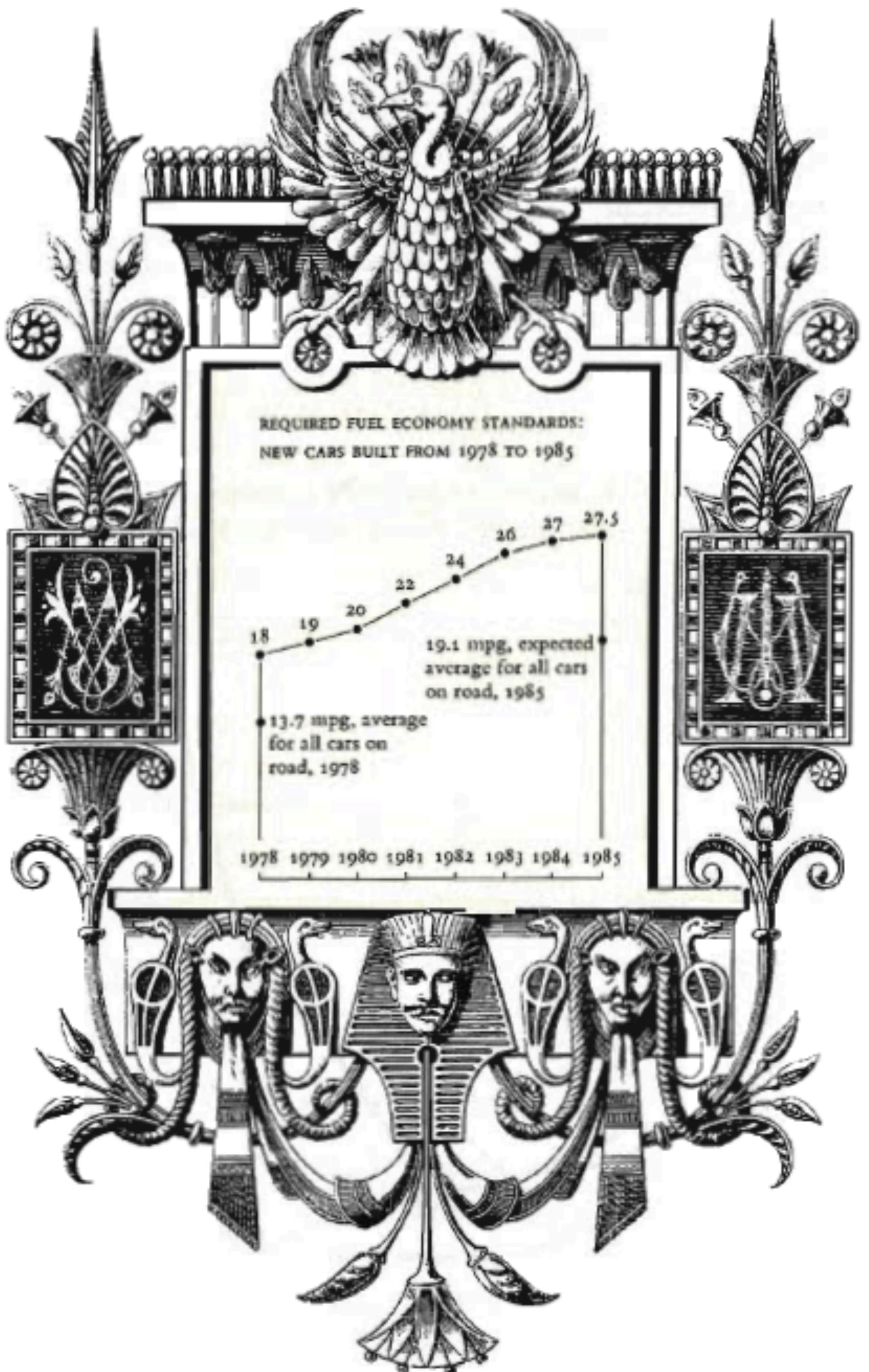
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

**PODRÍA  
HABERSE  
PRESENTADO DE  
ESTA MANERA**

REQUIRED FUEL ECONOMY STANDARDS:  
NEW CARS BUILT FROM 1978 TO 1985

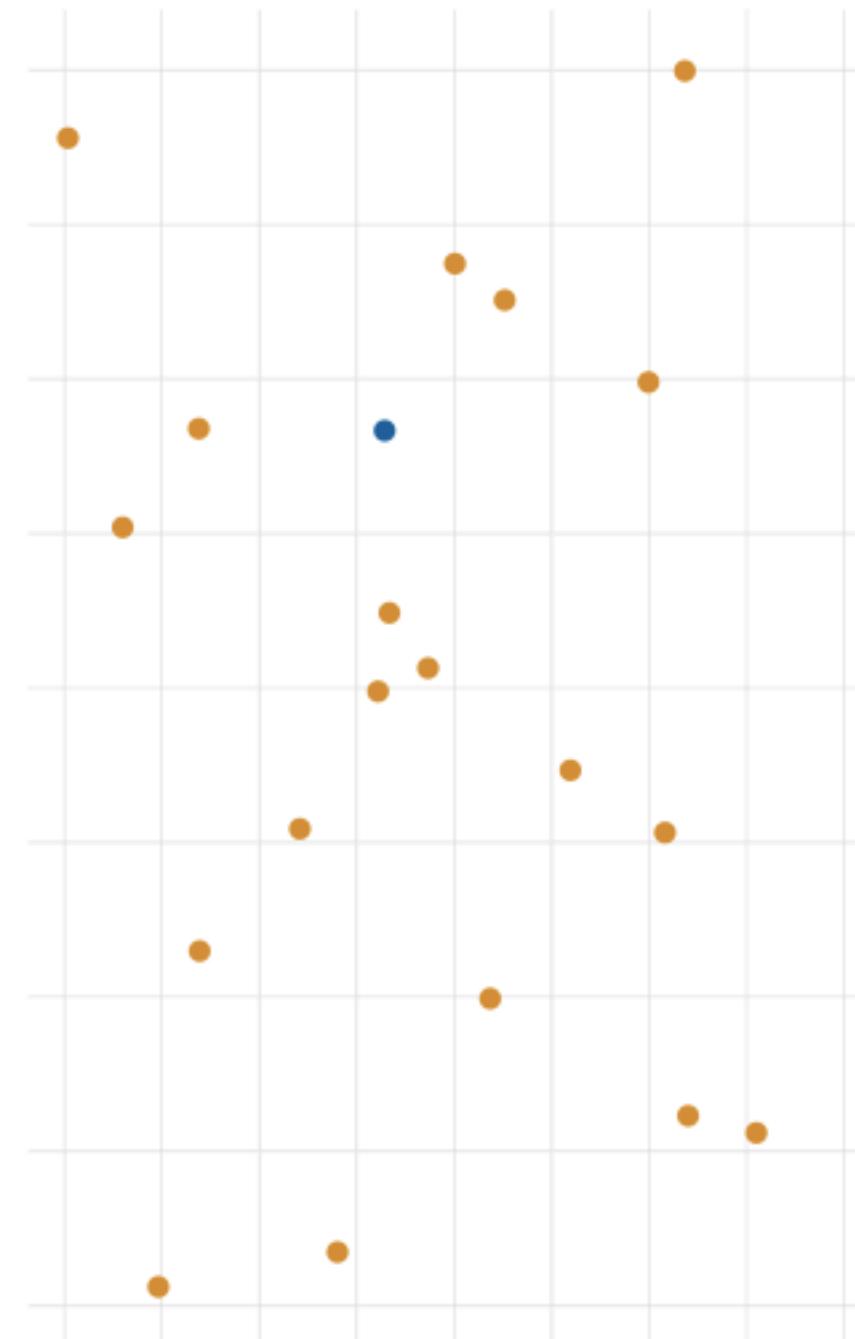


A VECES ES  
FACTIBLE INCLUIR  
DECORACIONES,  
PERO NO EN EXCESO

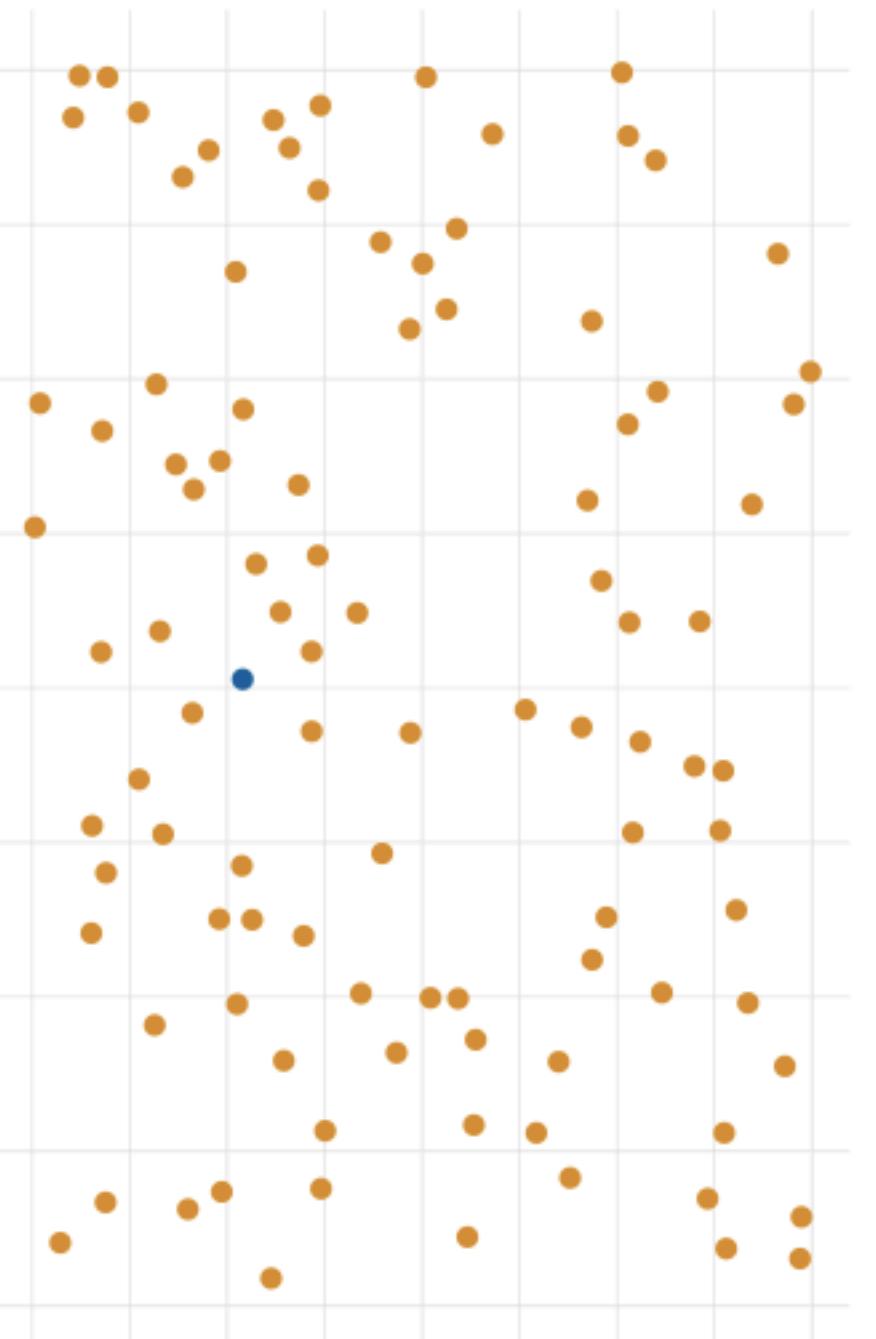


**UN USO INADECUADO DE  
NUEVAS VARIABLES Y  
FALTA DE LEYENDAS**

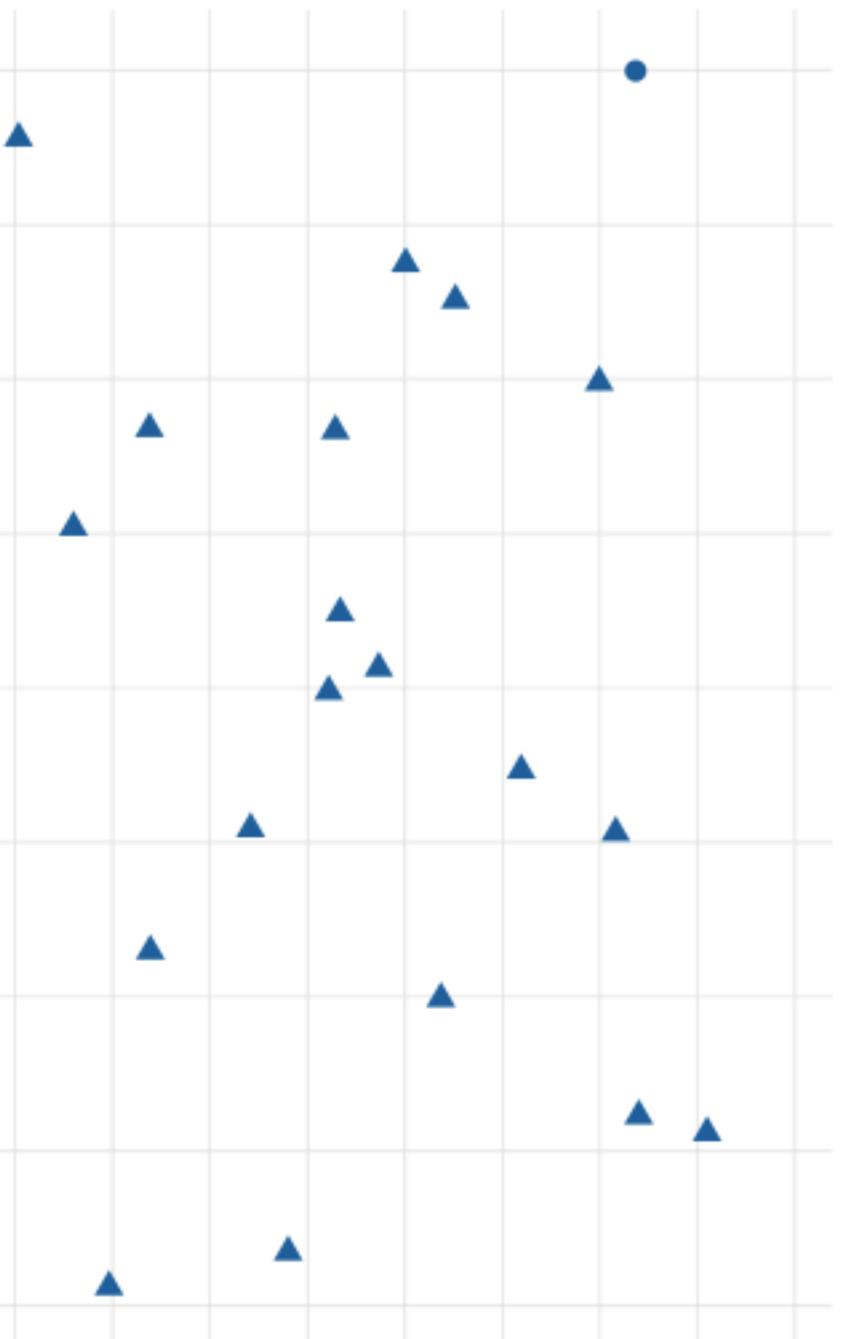
Color Only, N=20



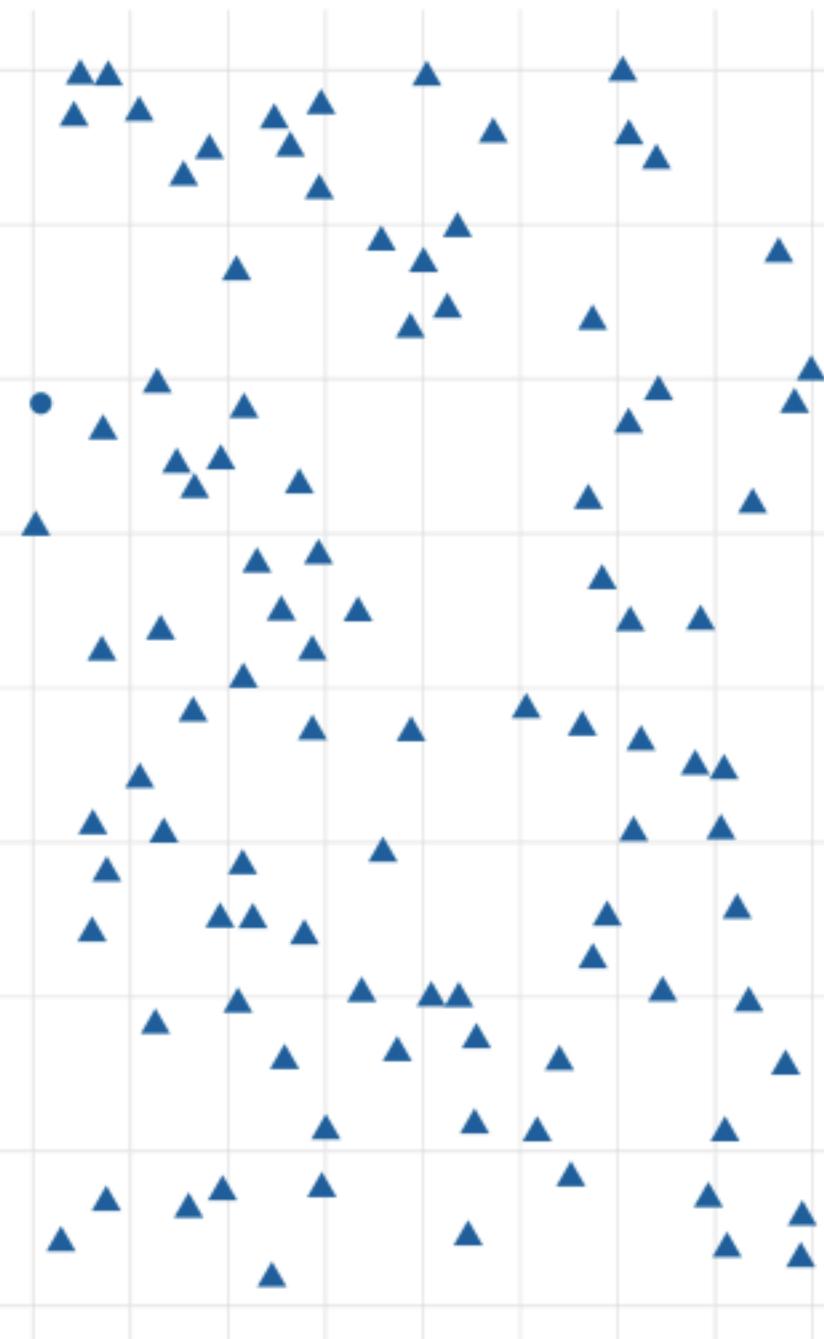
Color Only, N=100



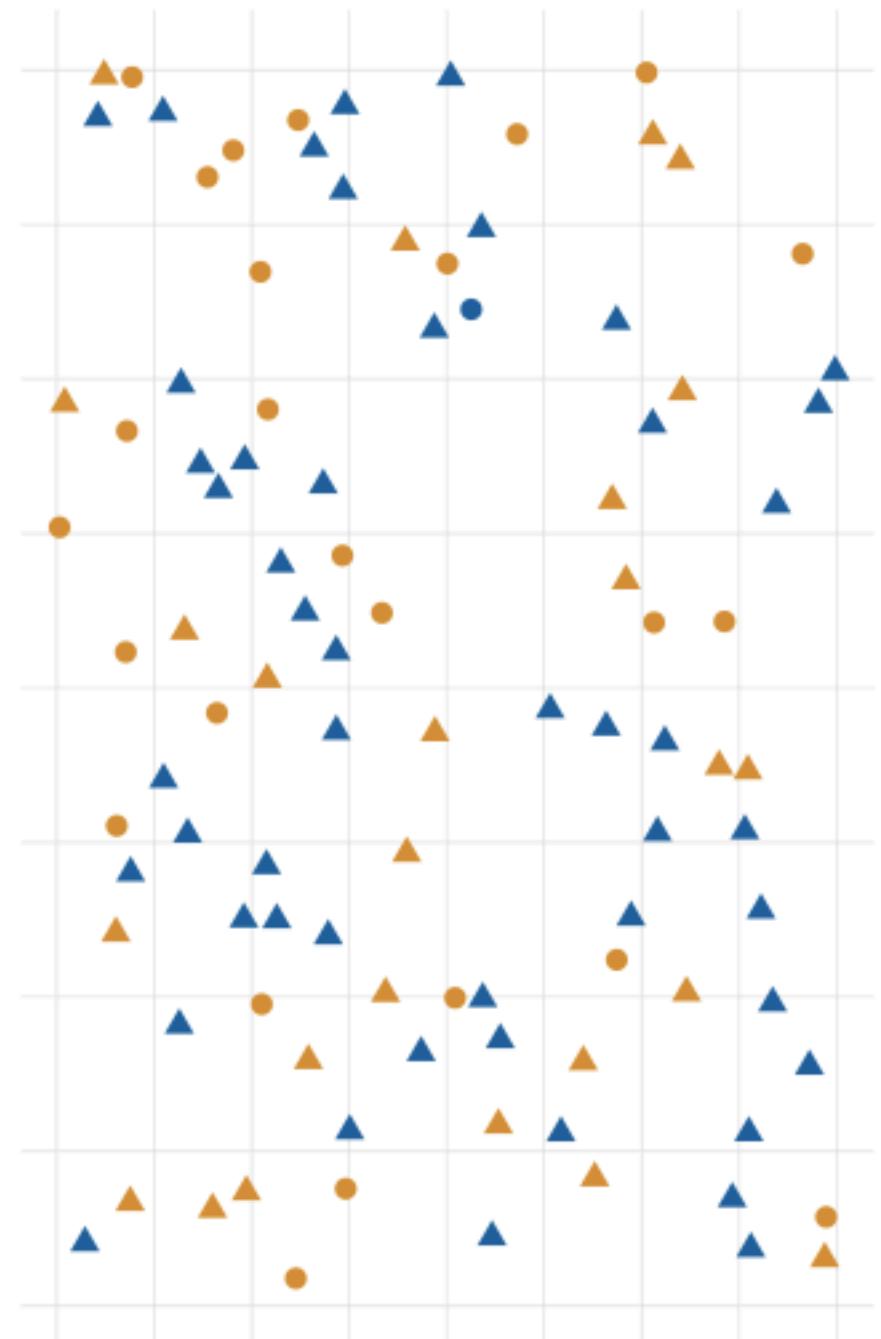
Shape Only, N=20

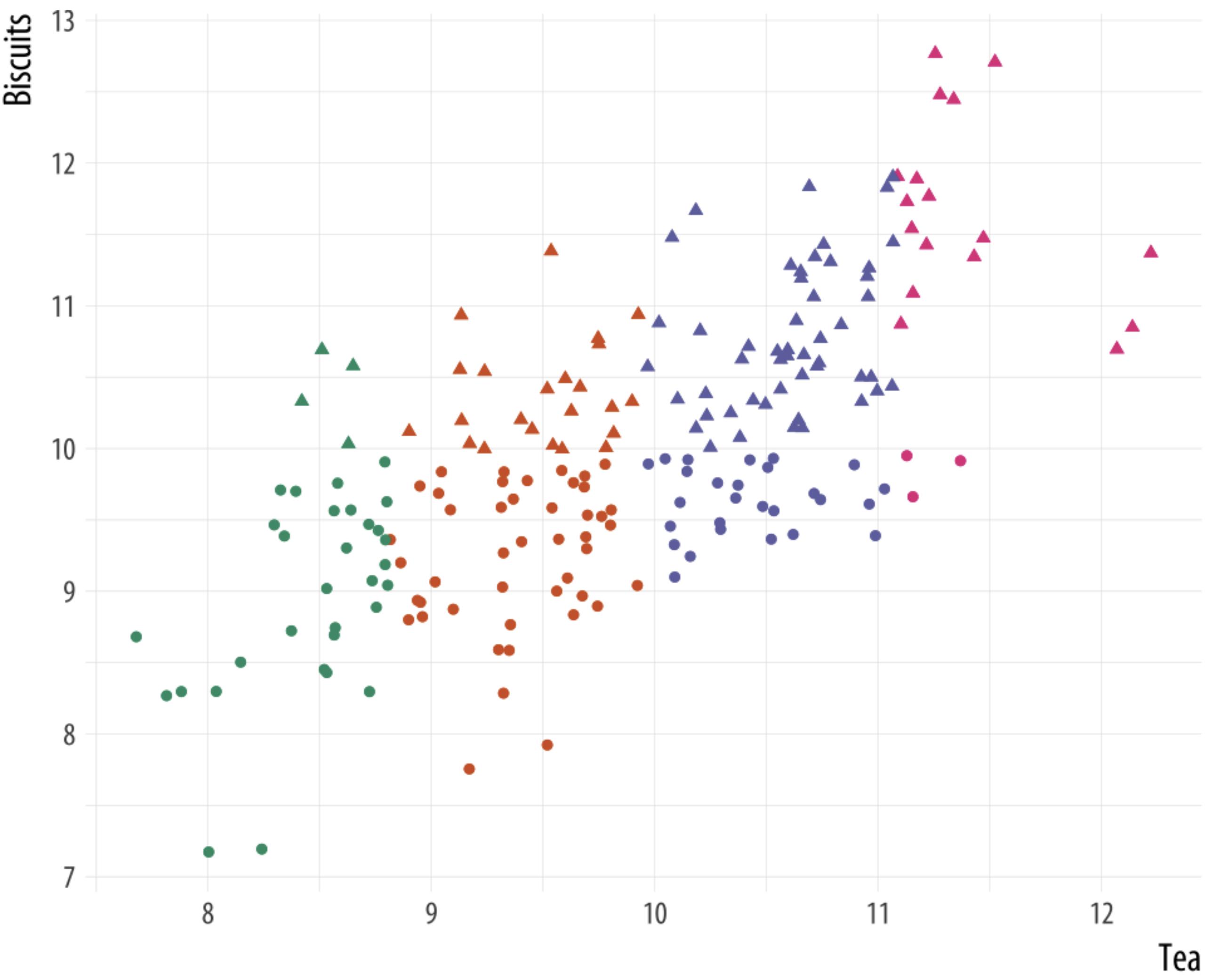
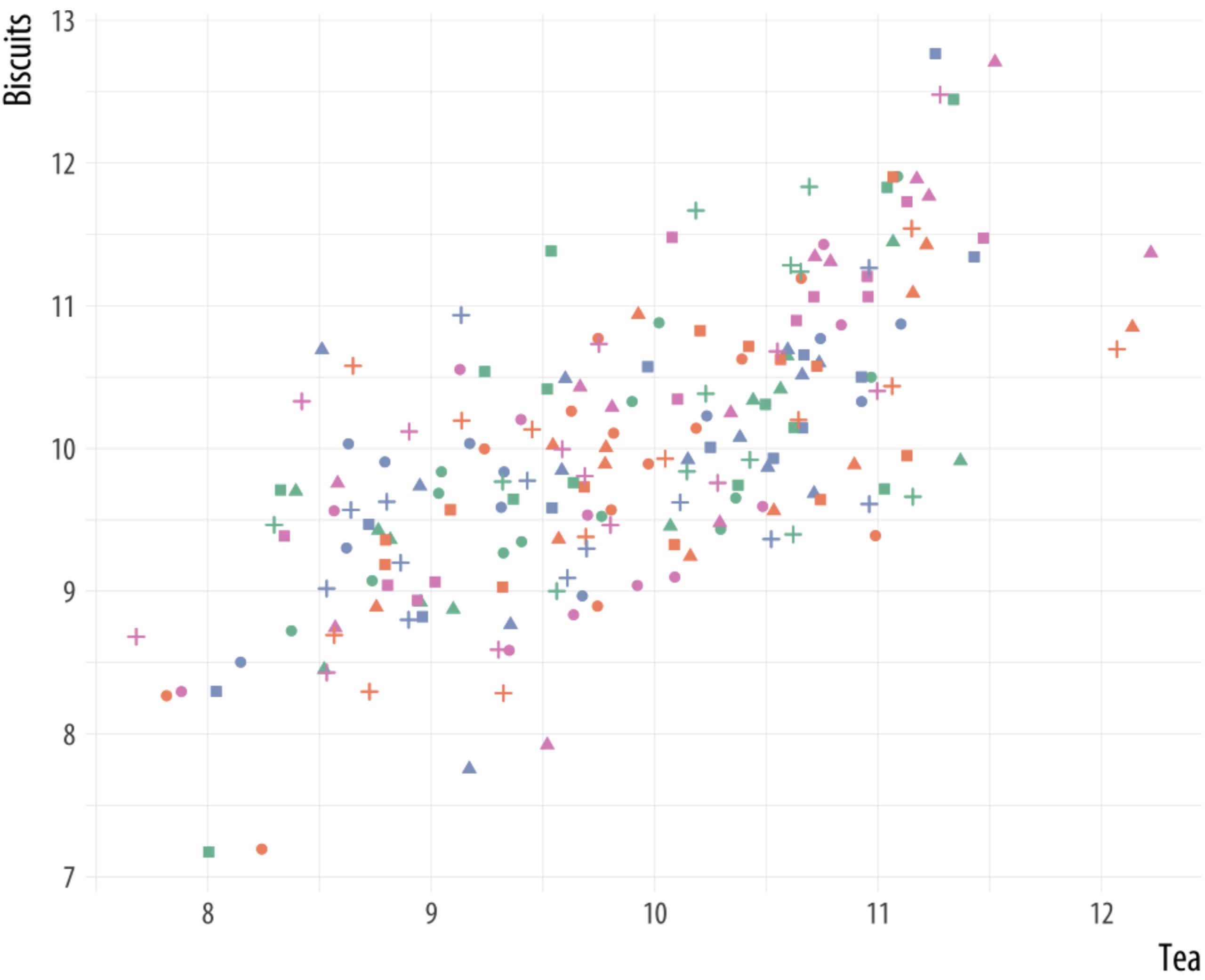


Shape Only, N=100



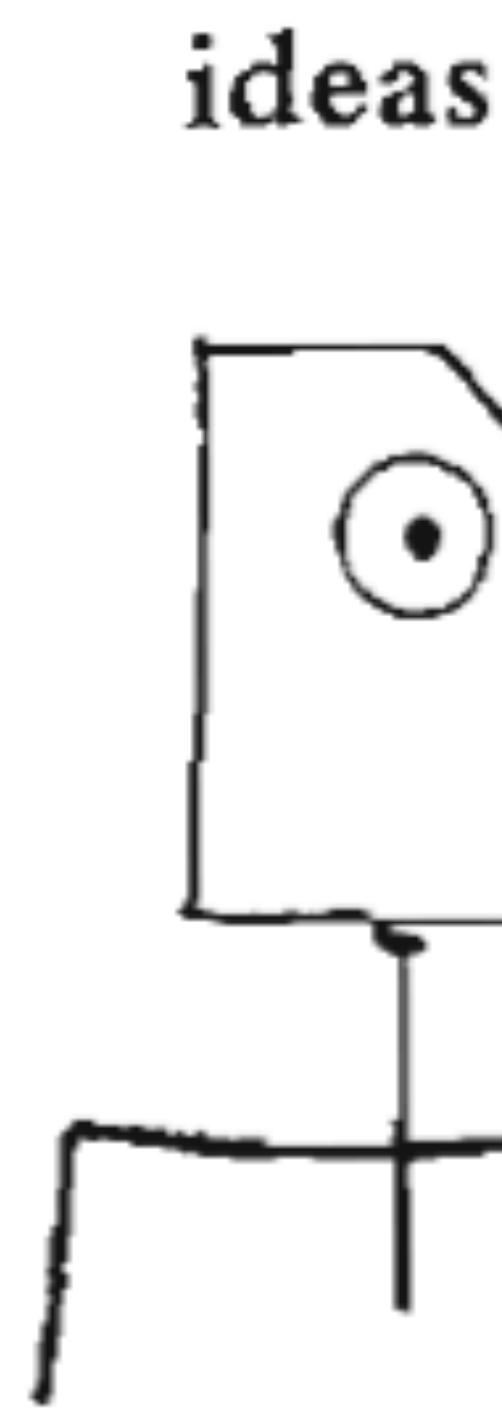
Color & Shape, N=100





# UN BUEN GRÁFICO

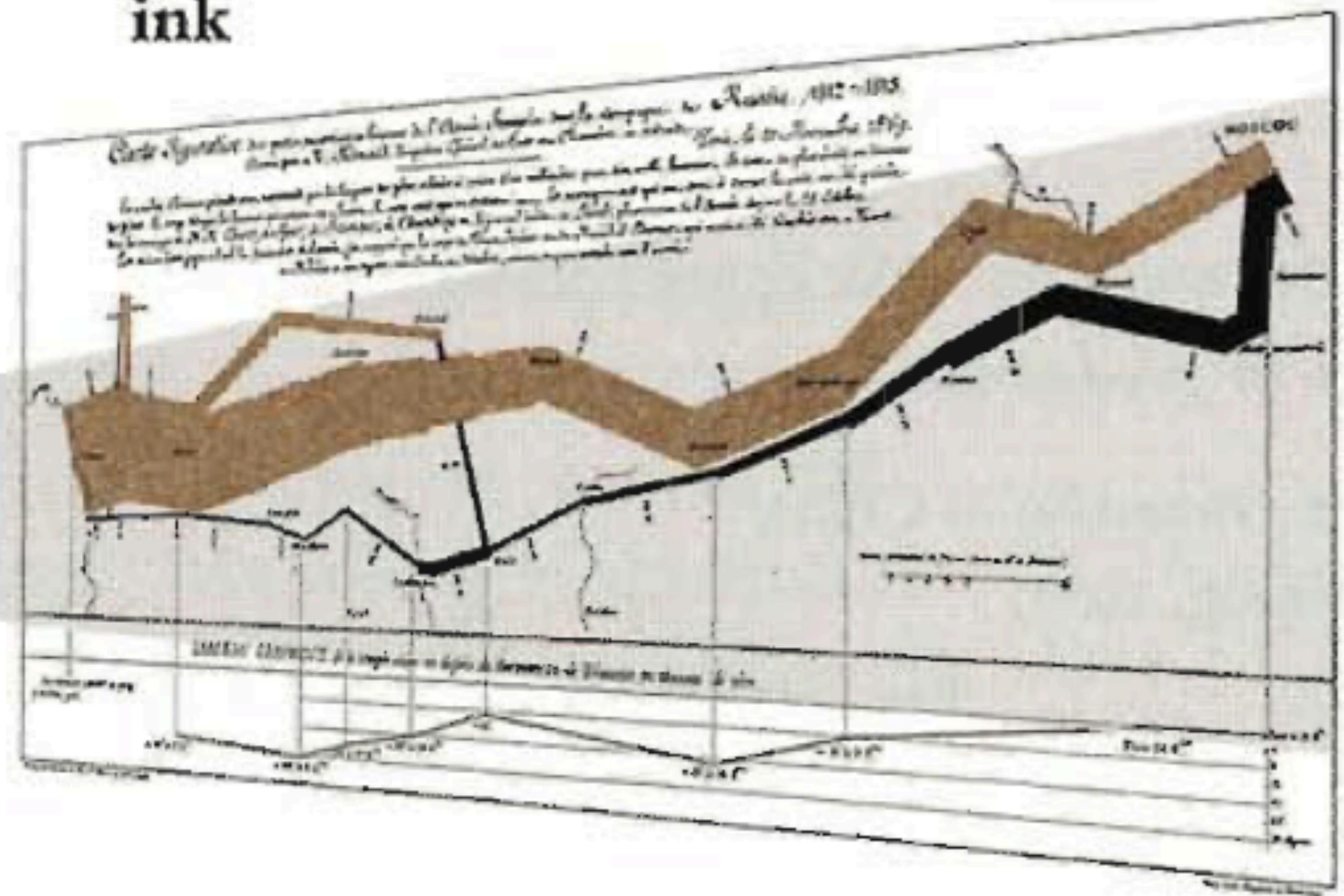
- ▶ MAXIMIZA LA **INFORMACIÓN**
- ▶ MINIMIZA **TINTA Y ESPACIO**
- ▶ COMUNICA EN EL MENOR **TIEMPO POSIBLE**



time

ink

space



# Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Mohilow et qui rejoignirent Orscha et Witebsk, avaient toujours marché avec l'armée.

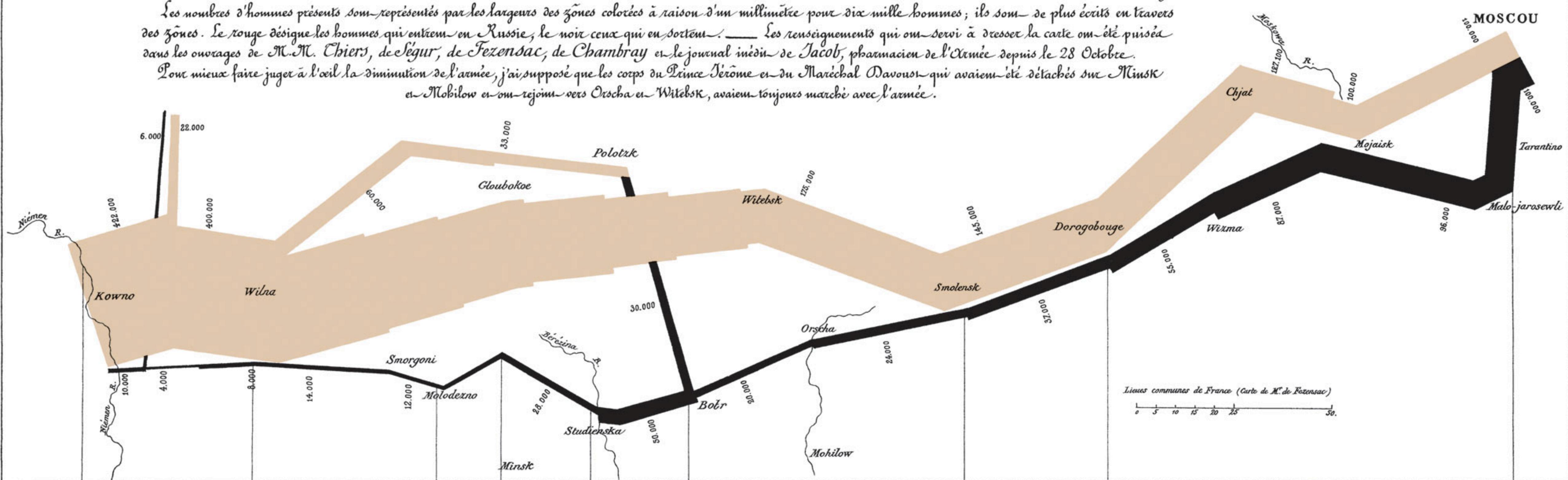
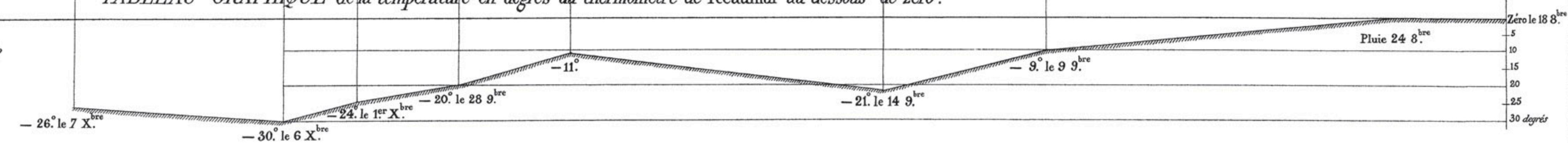


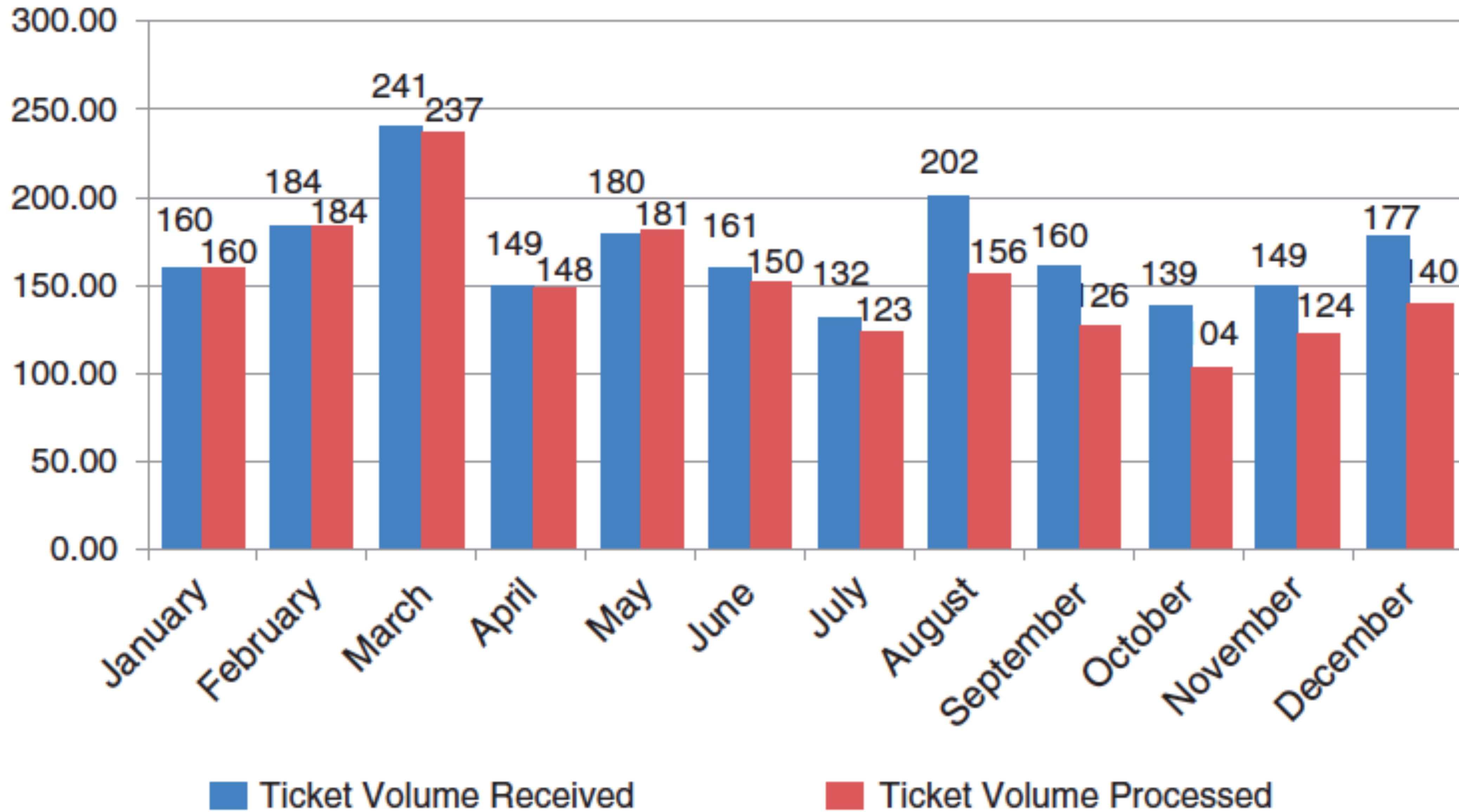
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop  
le Niemen gelé.



**GRÁFICOS INEFICIENTES**  
**HAY EN TODAS PARTES**

# Ticket Trend

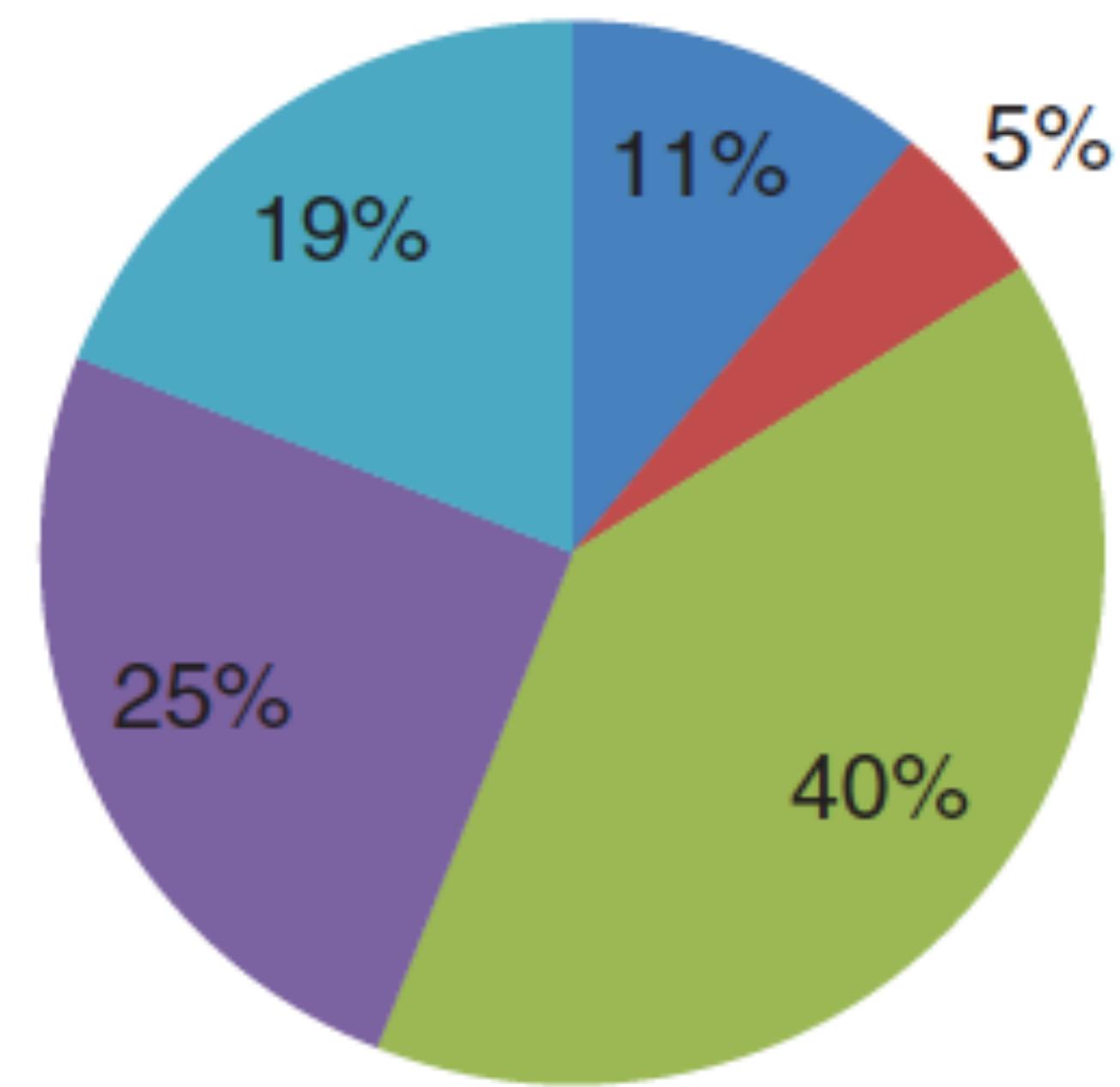


# Ticket volume over time



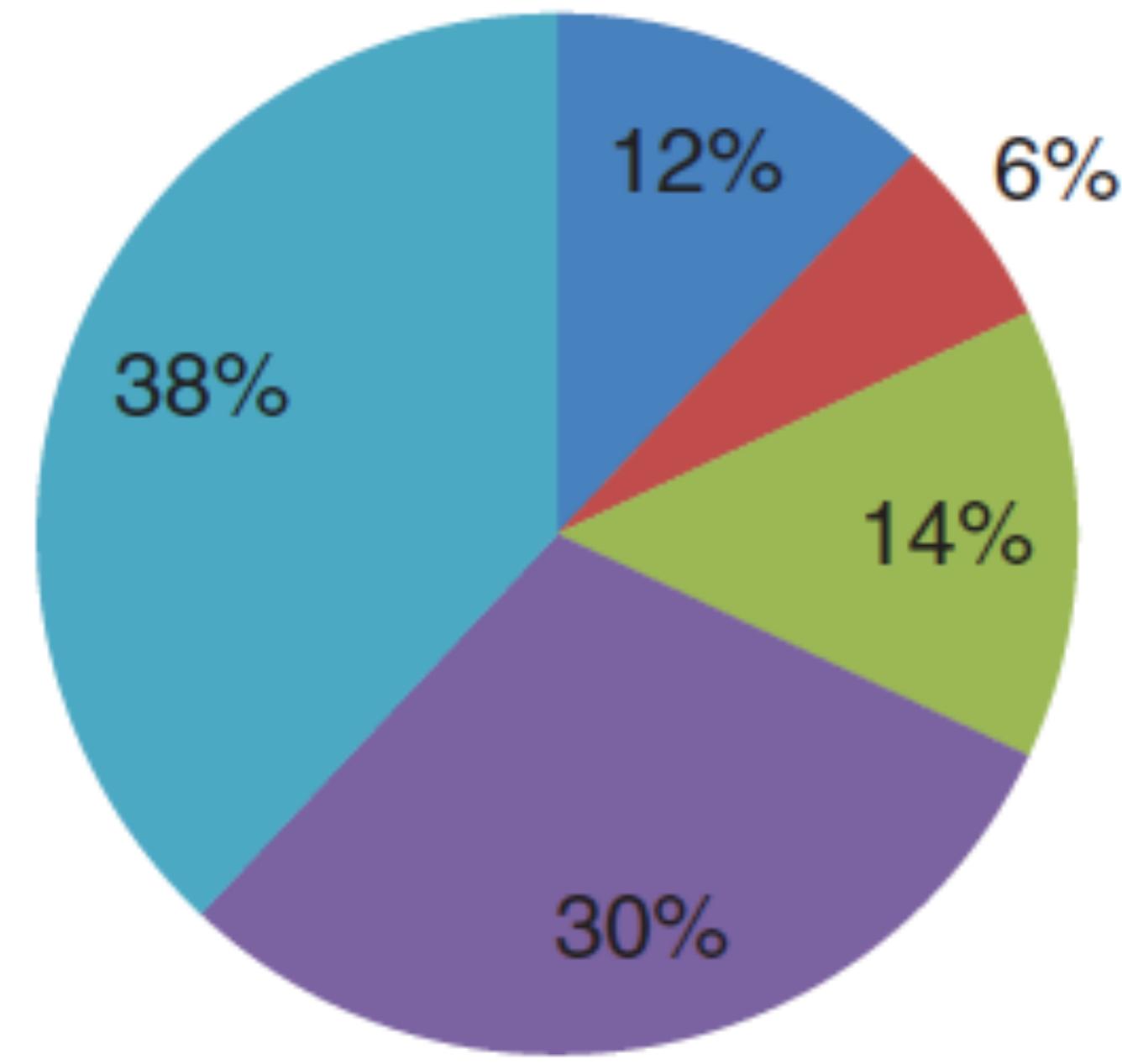
## PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



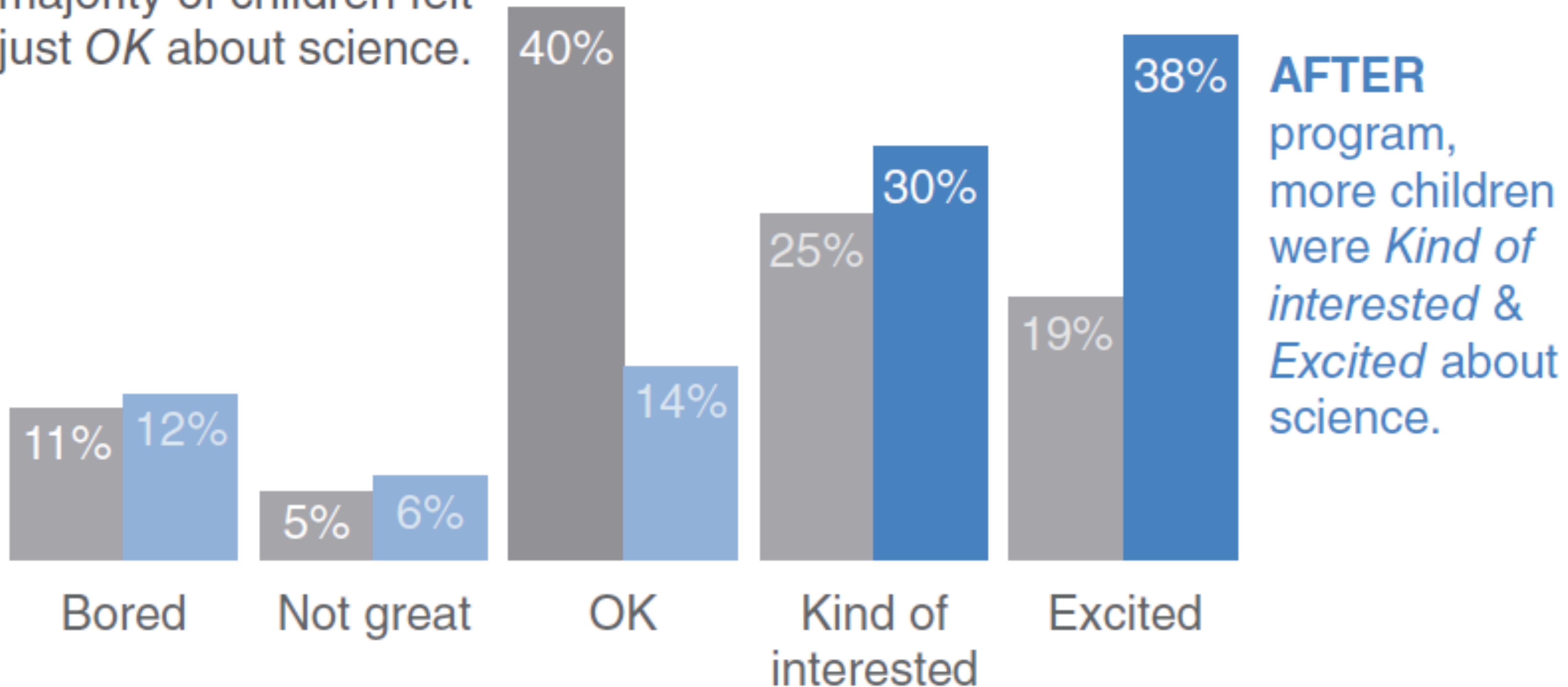
## POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



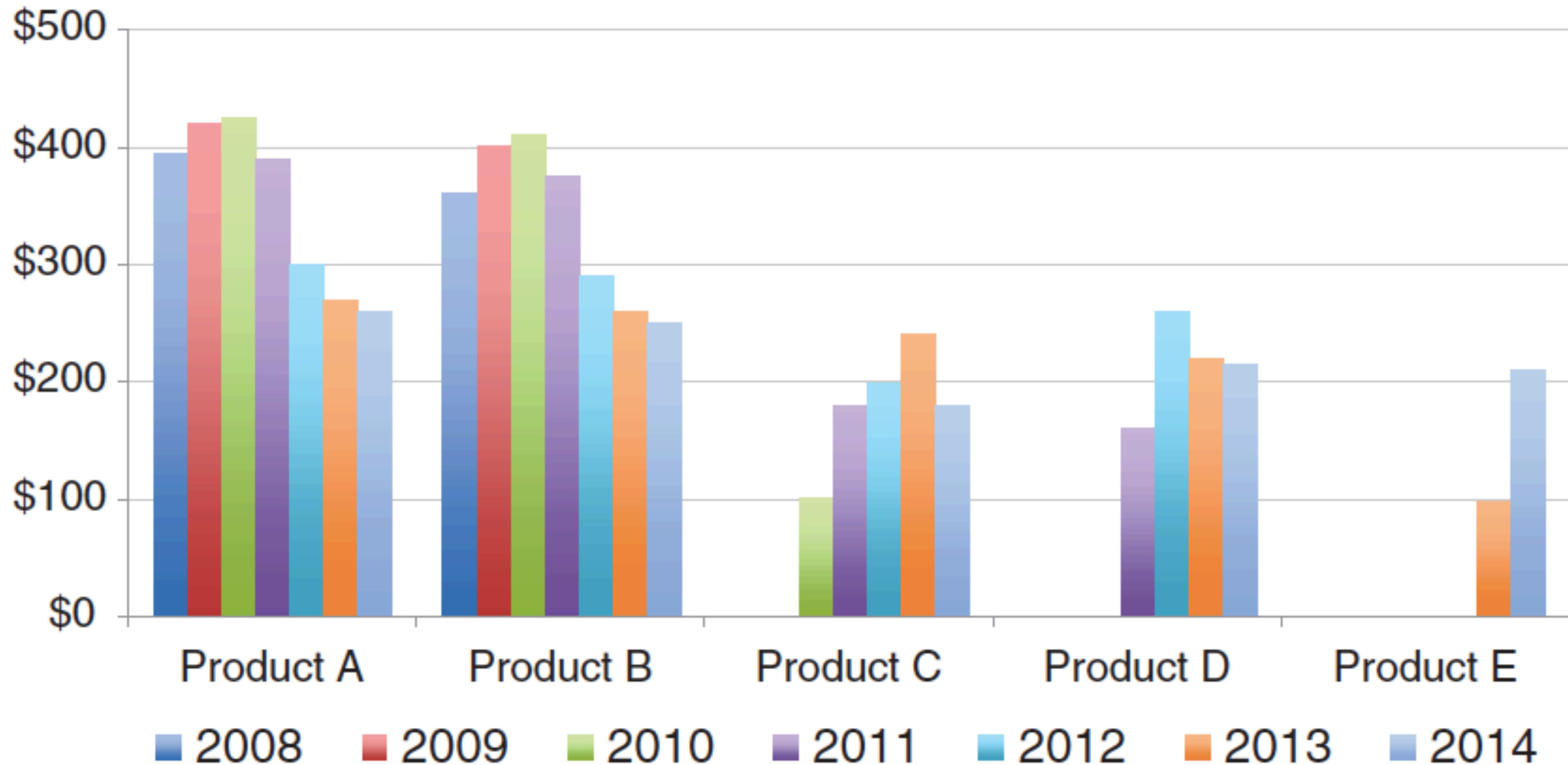
# How do you feel about science?

**BEFORE** program, the majority of children felt just *OK* about science.



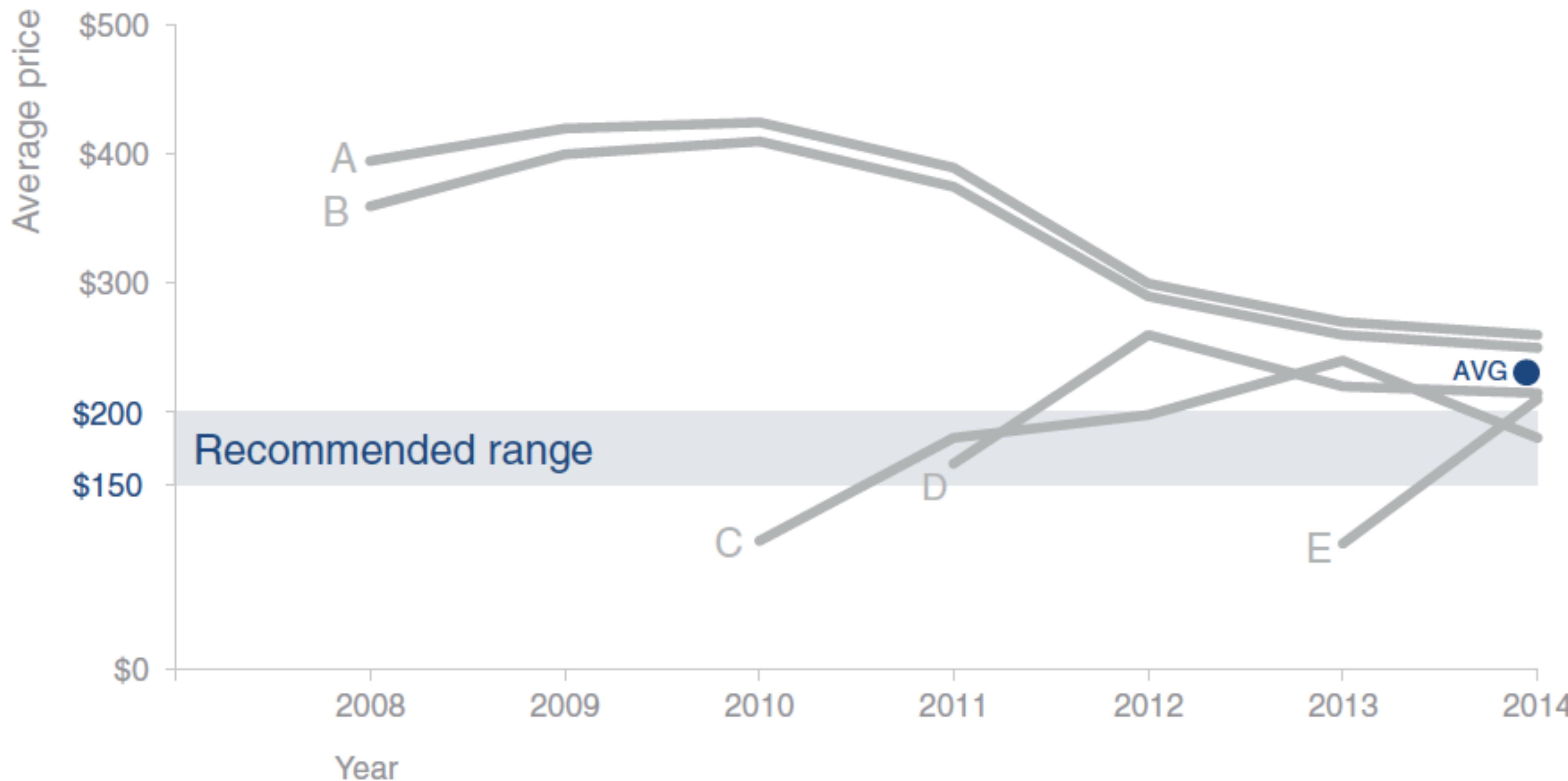
**AFTER** program,  
more children  
were *Kind of  
interested &  
Excited* about  
science.

# Average Retail Product Price per Year



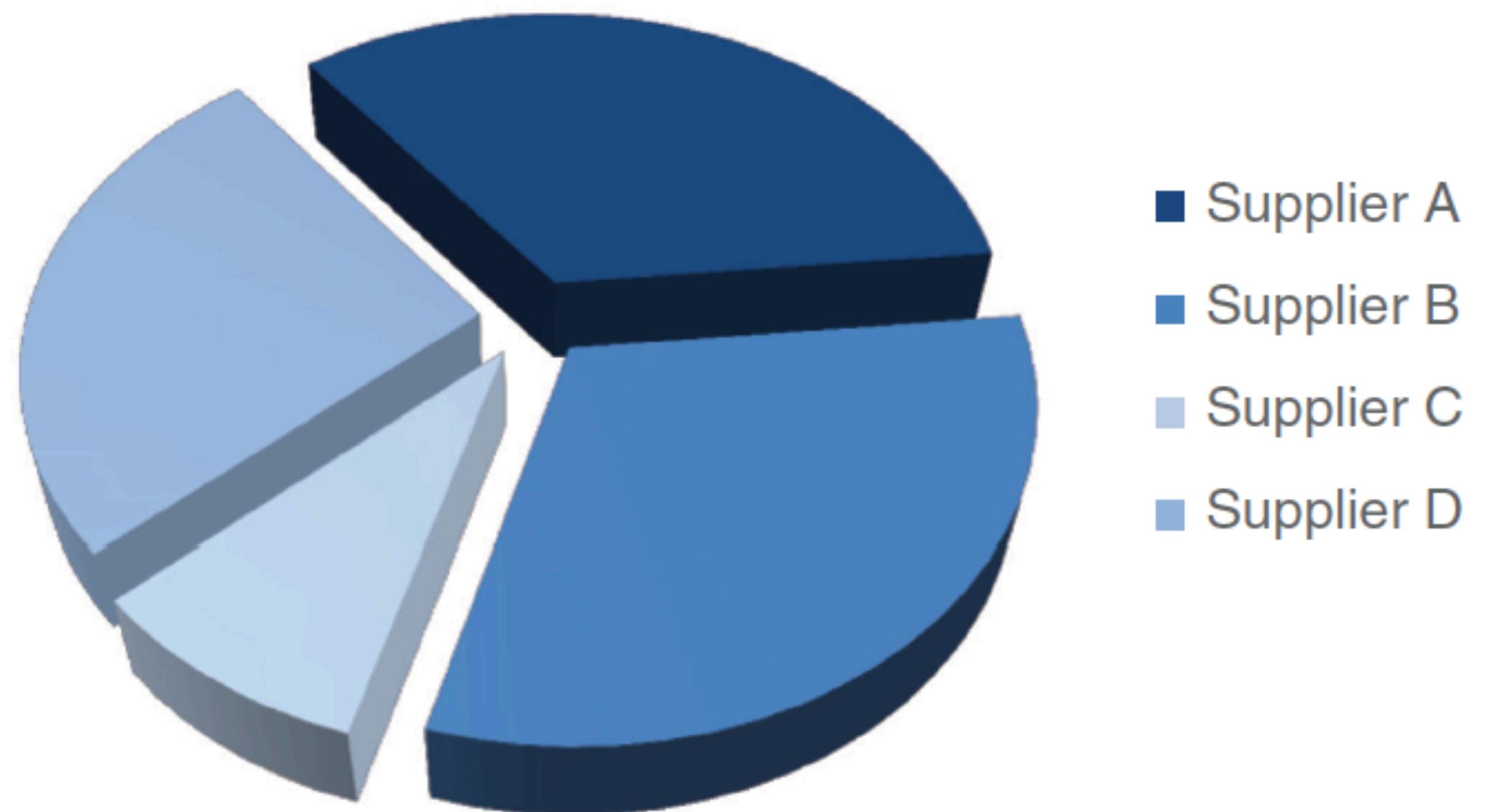
To be competitive, we recommend introducing our product *below* the \$223 average price point in the **\$150–\$200 range**

### Retail price over time by product



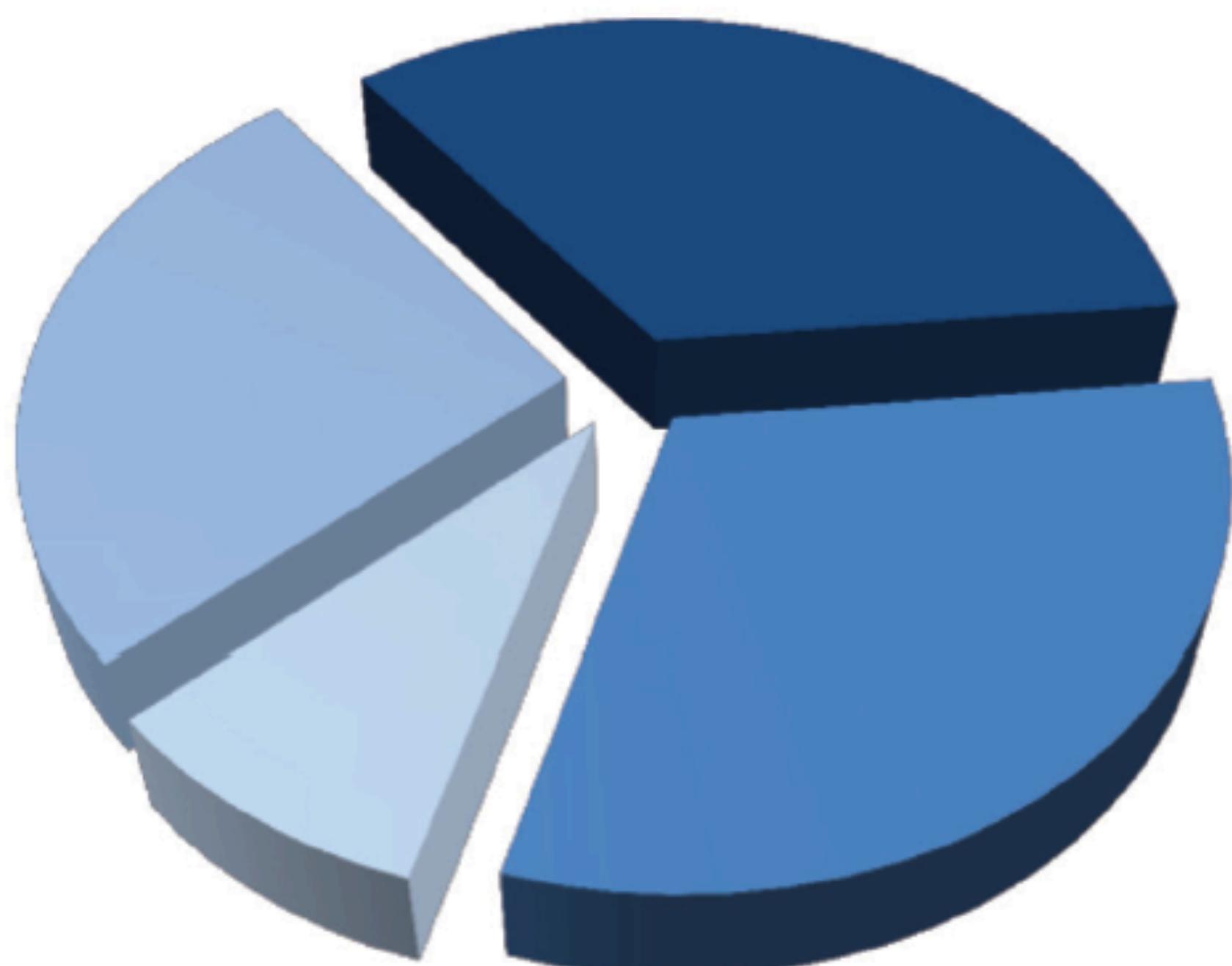
**EN GENERAL, LOS  
GRÁFICOS CIRCULARES  
SON MALVADOS**

## Supplier Market Share

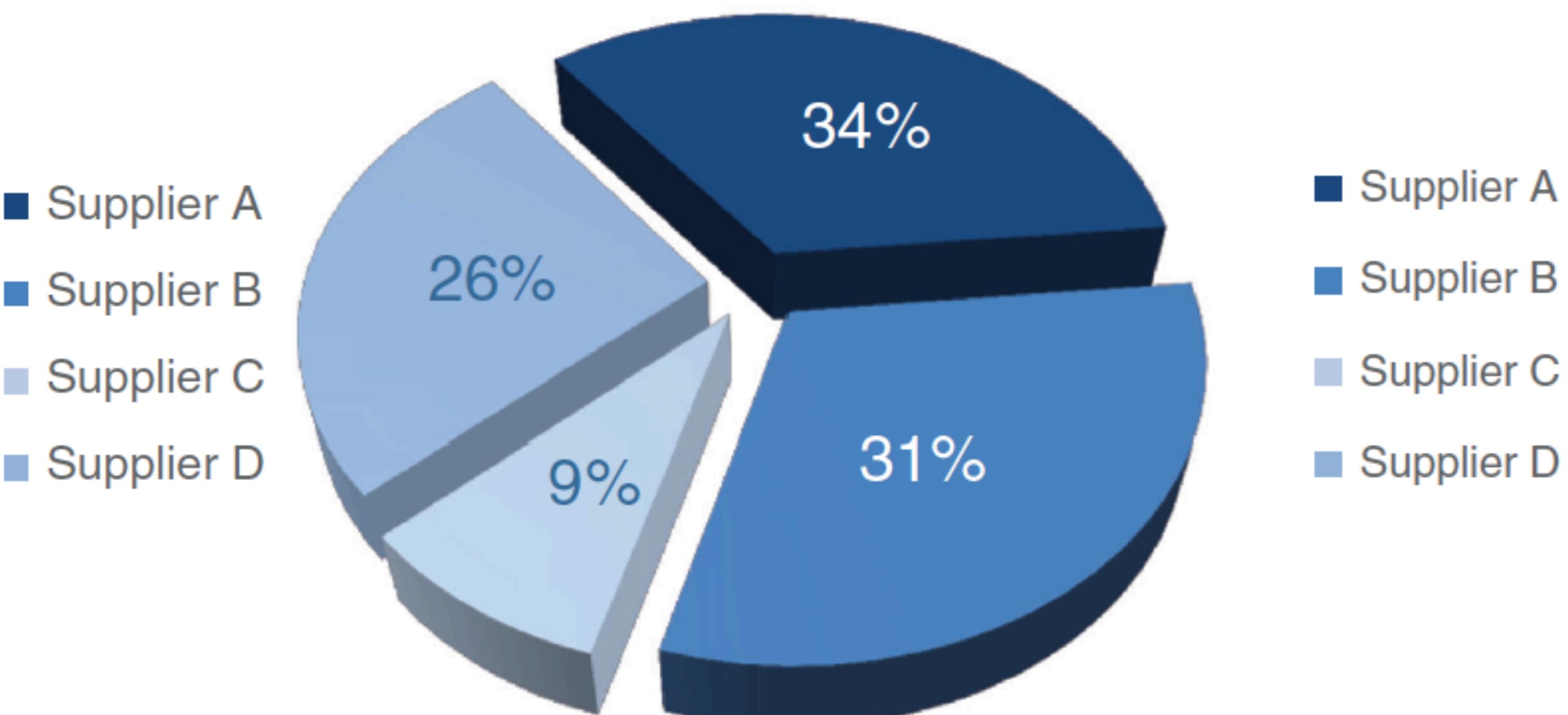


# INCLUSO AÑADIENDO ETIQUETAS

Supplier Market Share

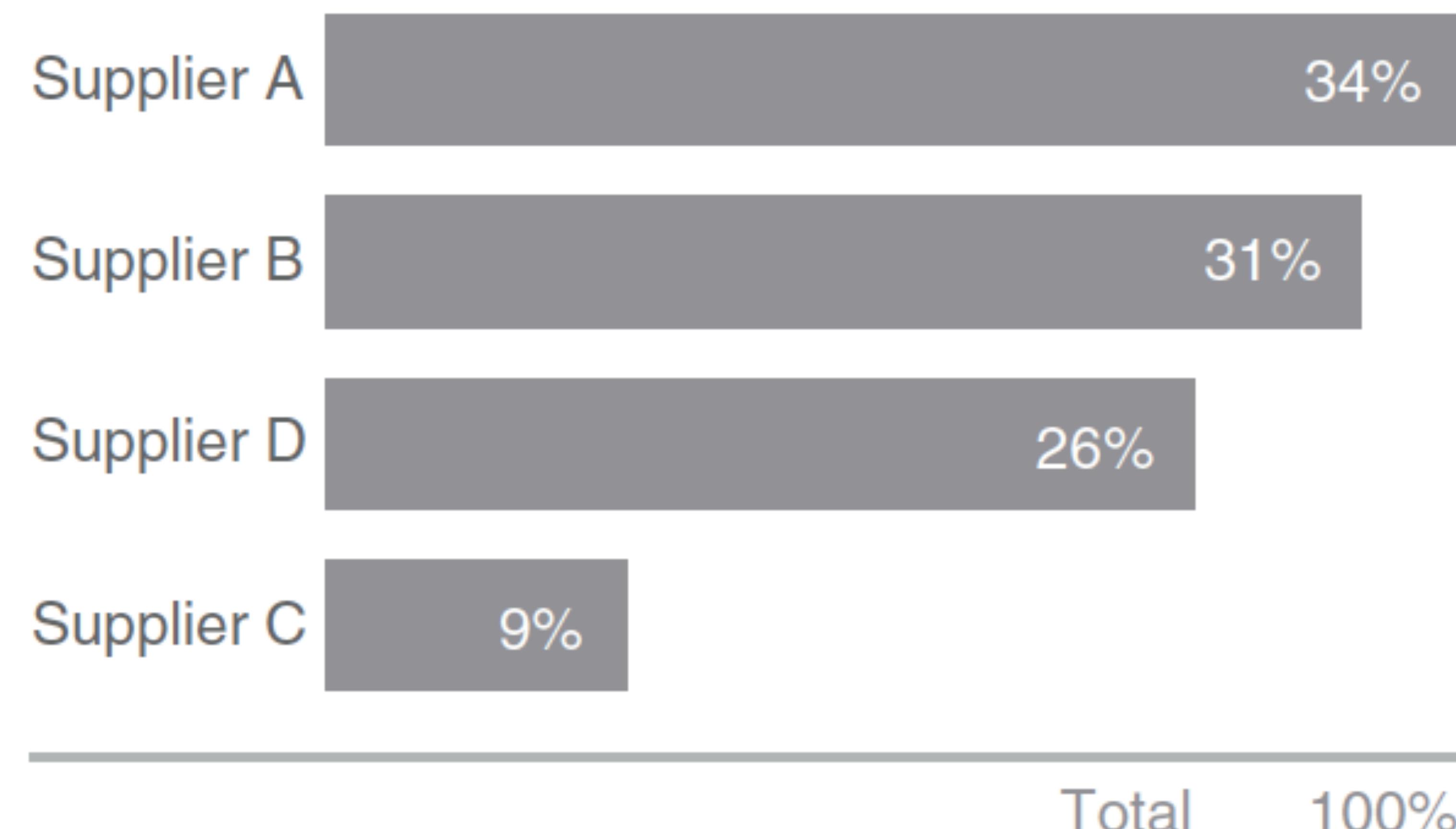


Supplier Market Share

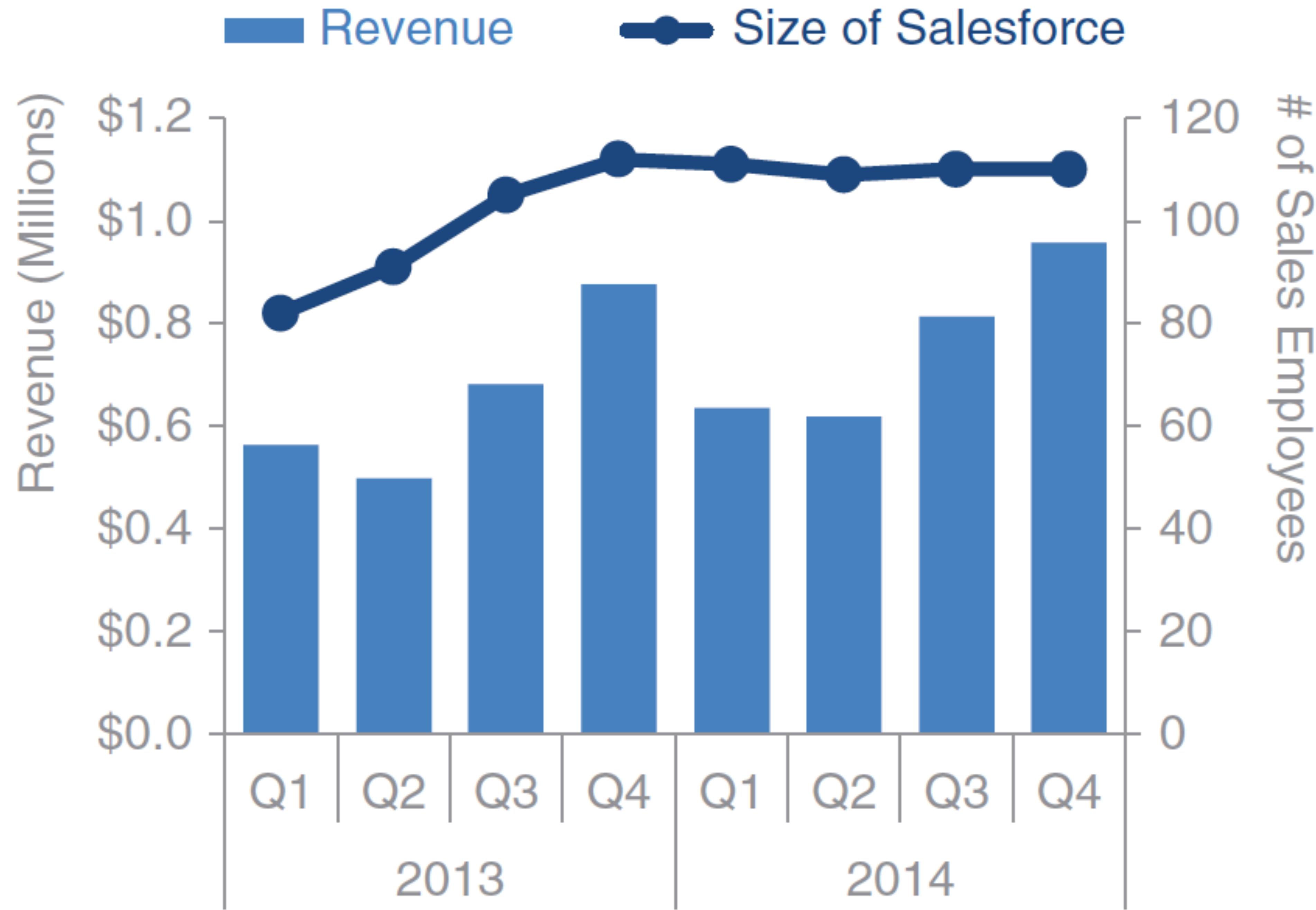


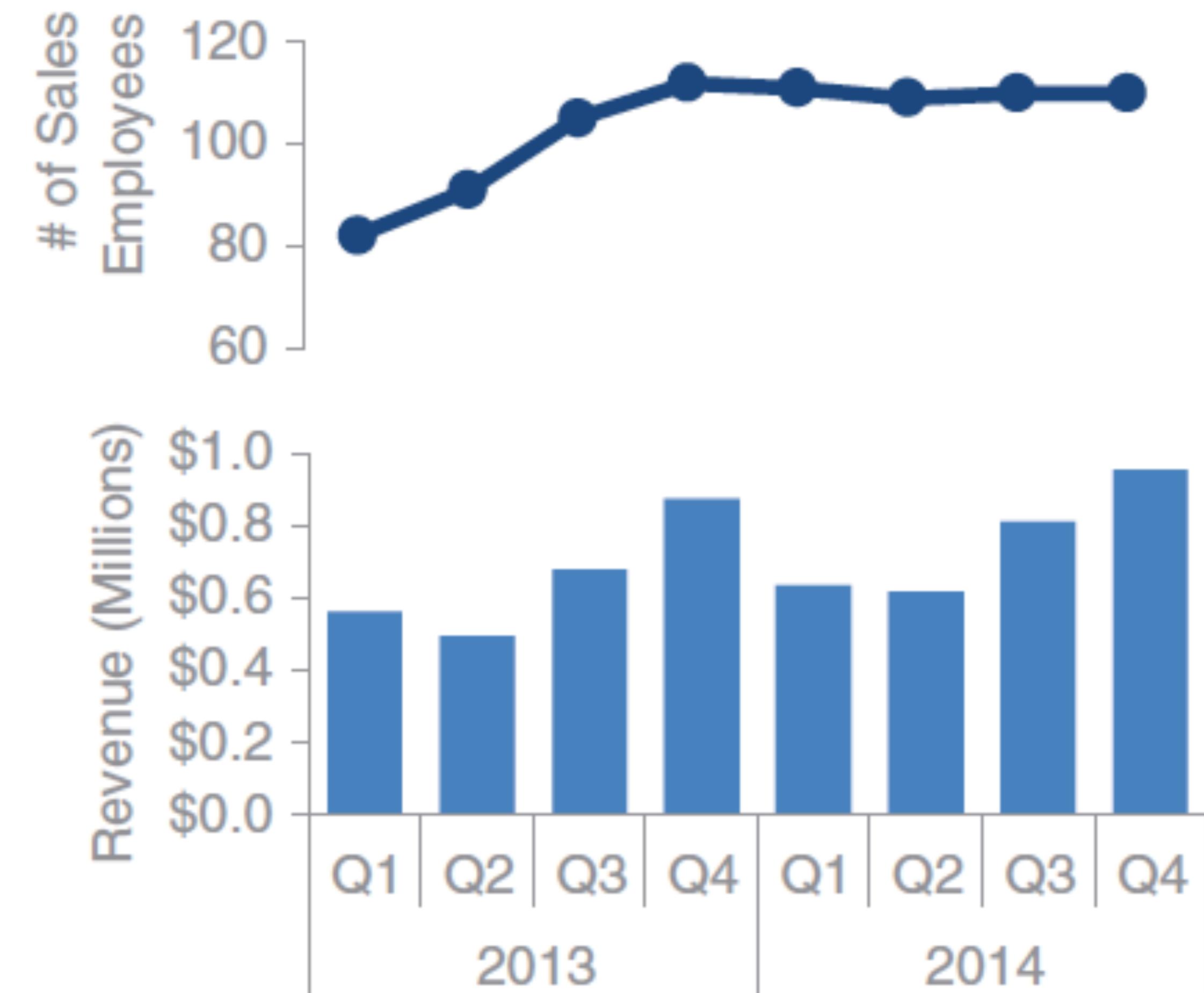
# UN GRÁFICO DE BARRAS ES UNA MEJOR OPCIÓN

Supplier Market Share



**UN DOBLE USO DEL EJE  
Y GENERALMENTE NO ES  
BUENA IDEA**





**ESTÉTICA**

**UTILICEMOS CORRECTAMENTE  
OTROS RECURSOS**

**DEBEMOS PONER ÉNFASIS EN  
LO QUE DESEAMOS COMUNICAR**

# CONTAR EL NÚMERO DE 3'S

756395068473

658663037576

860372658602

846589107830

# CONTAR EL NÚMERO DE 3'S

756395068473

658663037576

860372658602

846589107830

756**3**9506847**3**

6586630**3**7576

860**3**72658602

8465891078**3**0

# ¿CÓMO LLAMAR LA ATENCIÓN?



Orientation



Shape



Line length



Line width



Size



Curvature



Added marks



Enclosure



Hue



Intensity



Spatial position



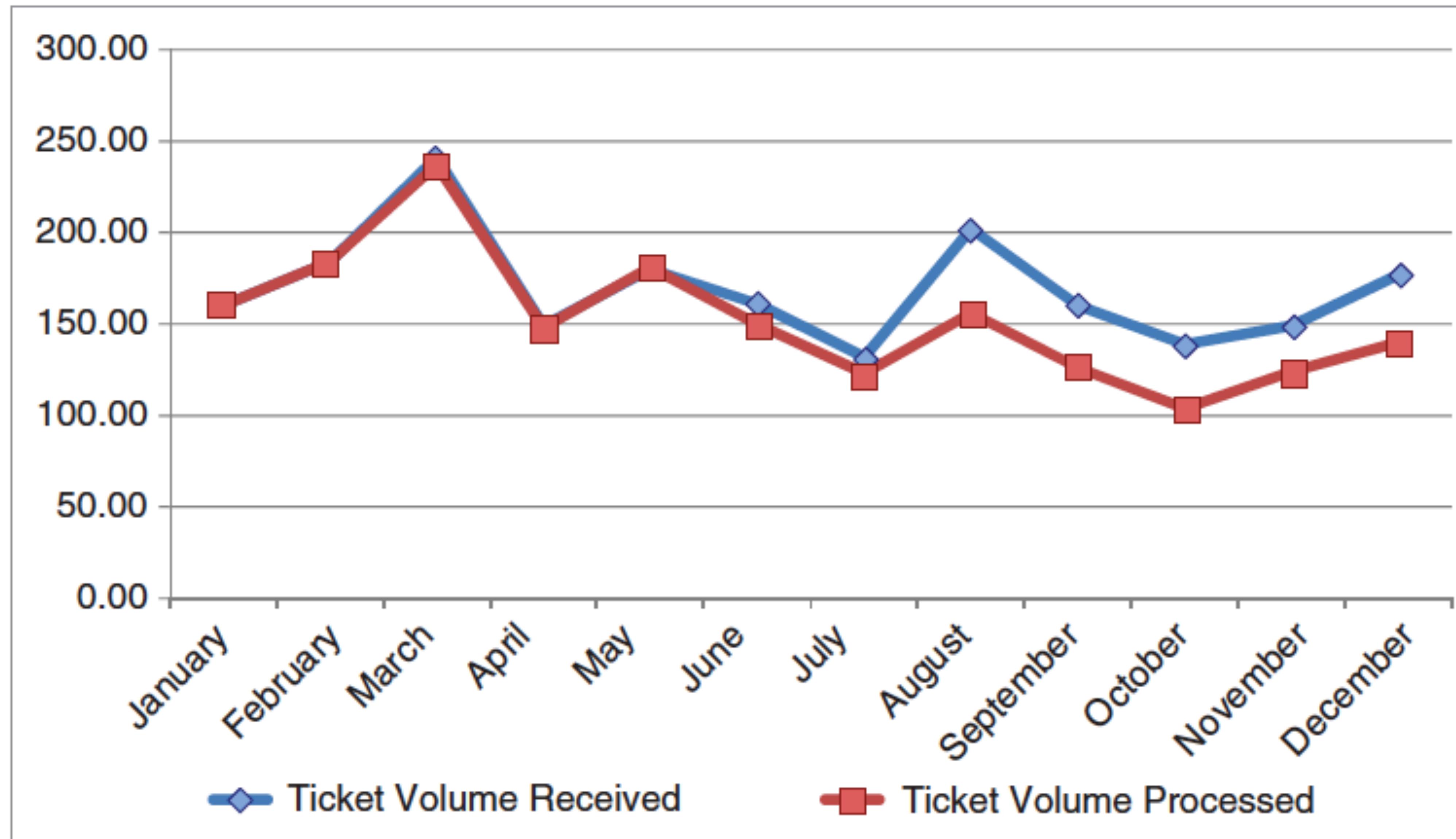
Motion

# **¿CÓMO LLAMAR LA ATENCIÓN?**

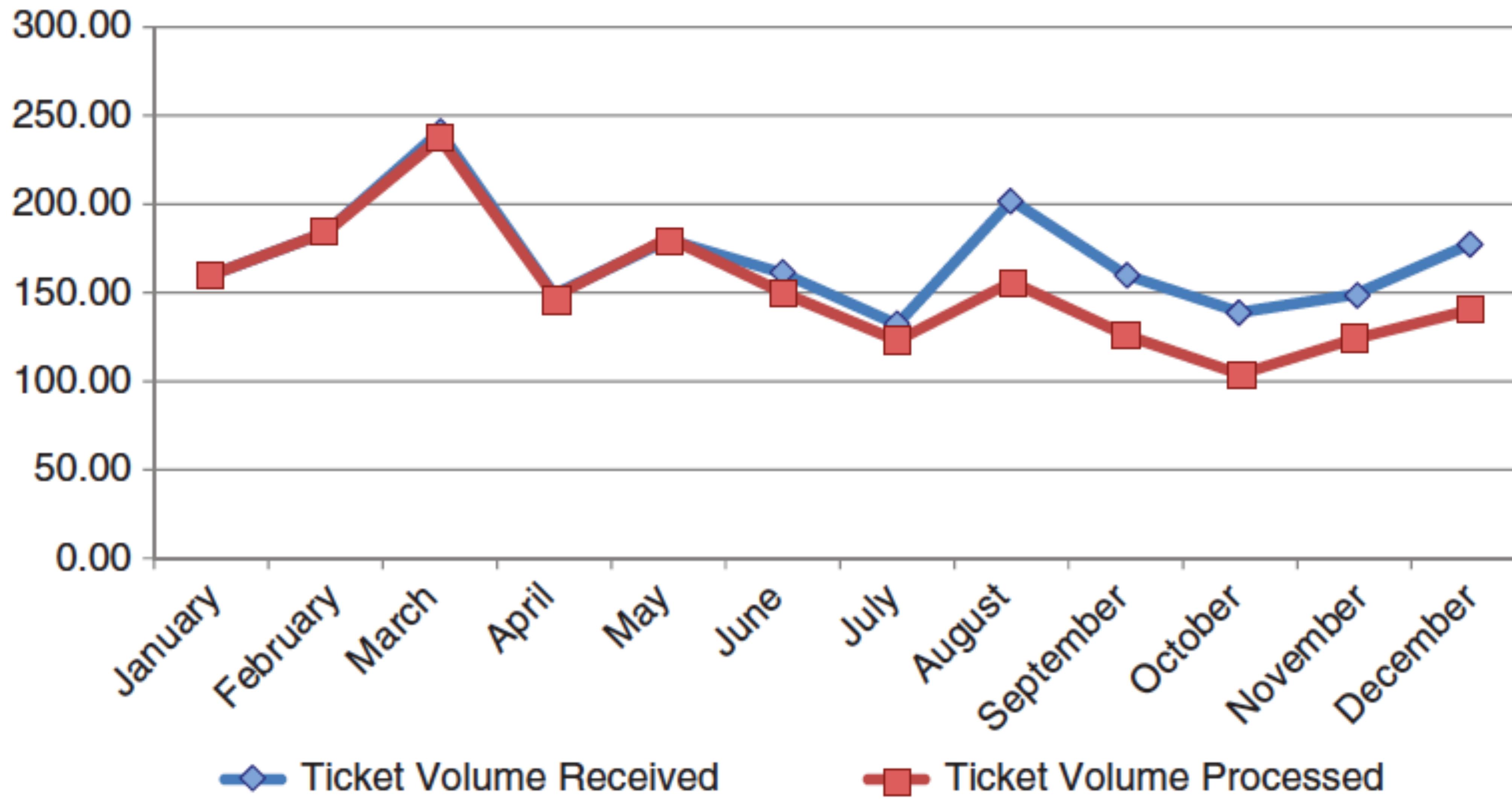
**RECUERDEN:**

**ELIMINAR LA BASURA PARA NO  
GENERAR DISTRACCIONES**

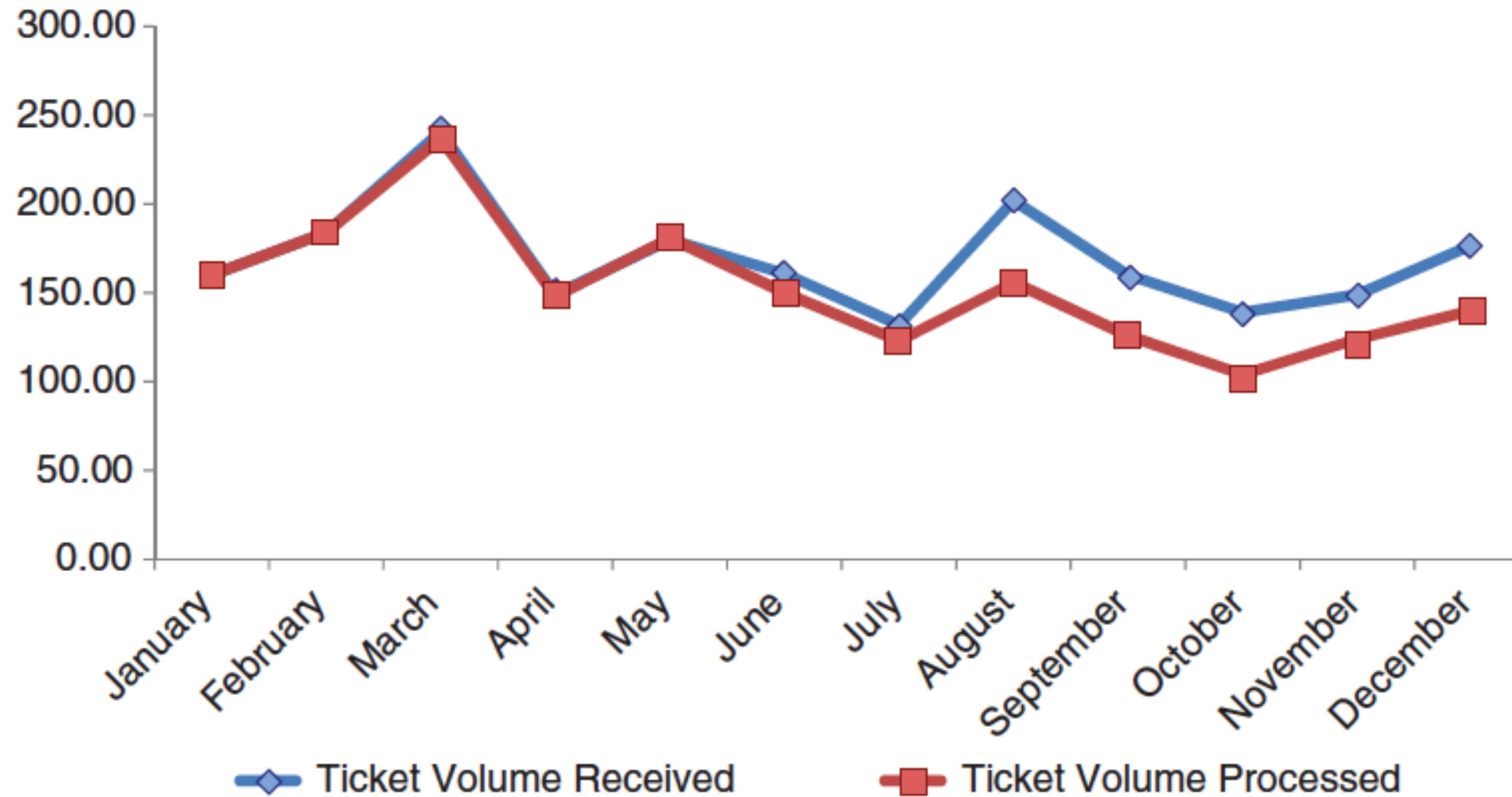
# ¿CÓMO LLAMAR LA ATENCIÓN?



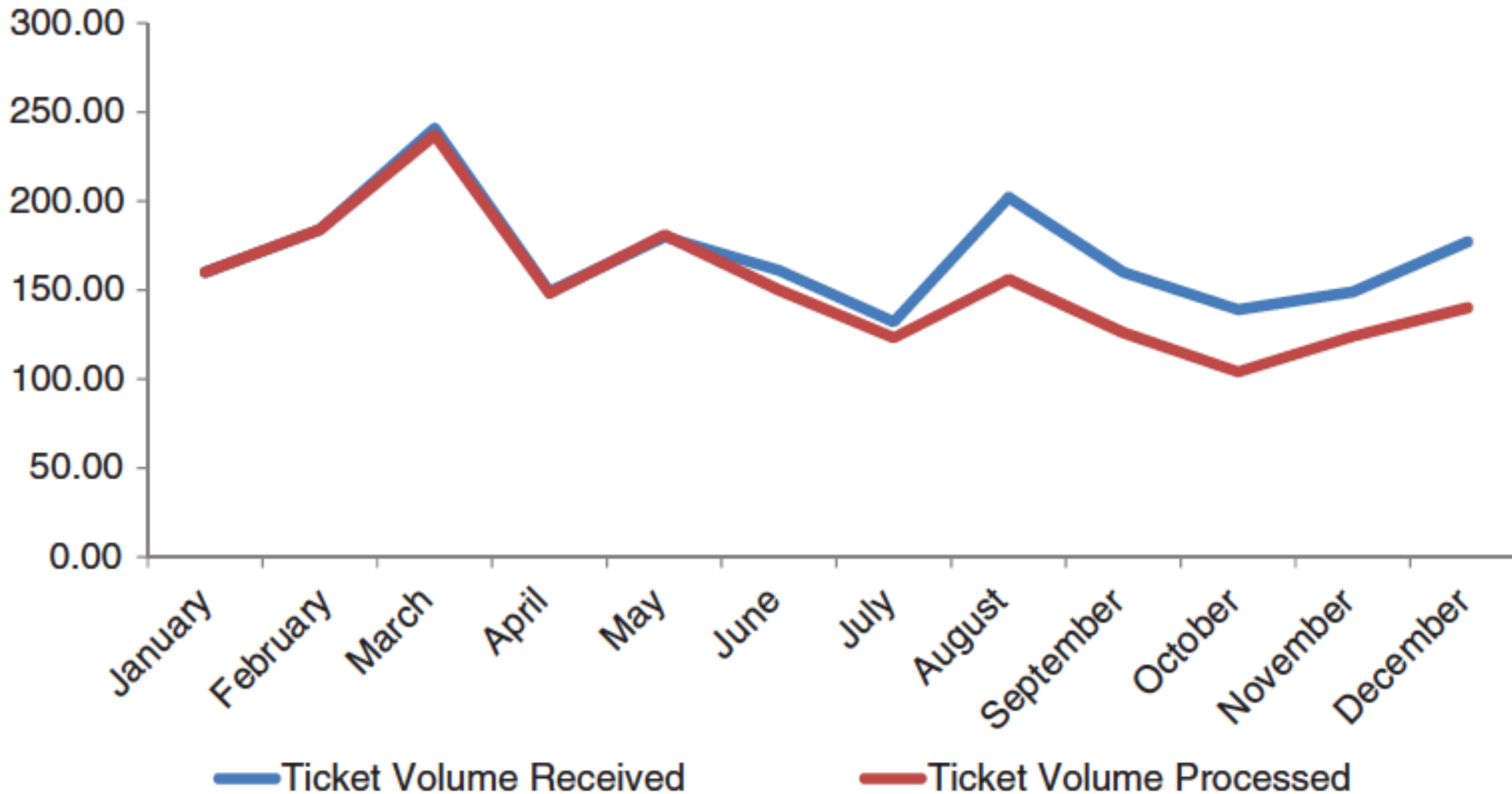
# ¿CÓMO LLAMAR LA ATENCIÓN?



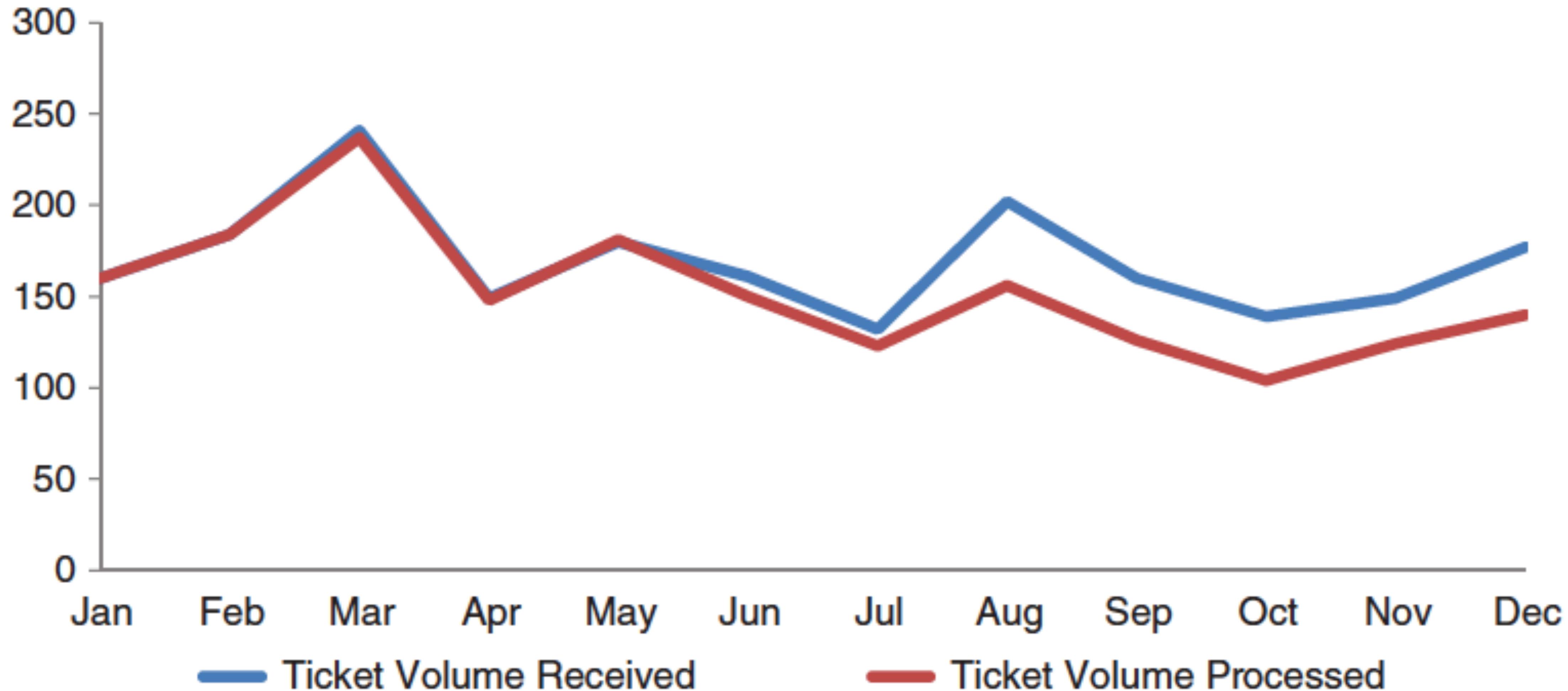
# ¿CÓMO LLAMAR LA ATENCIÓN?



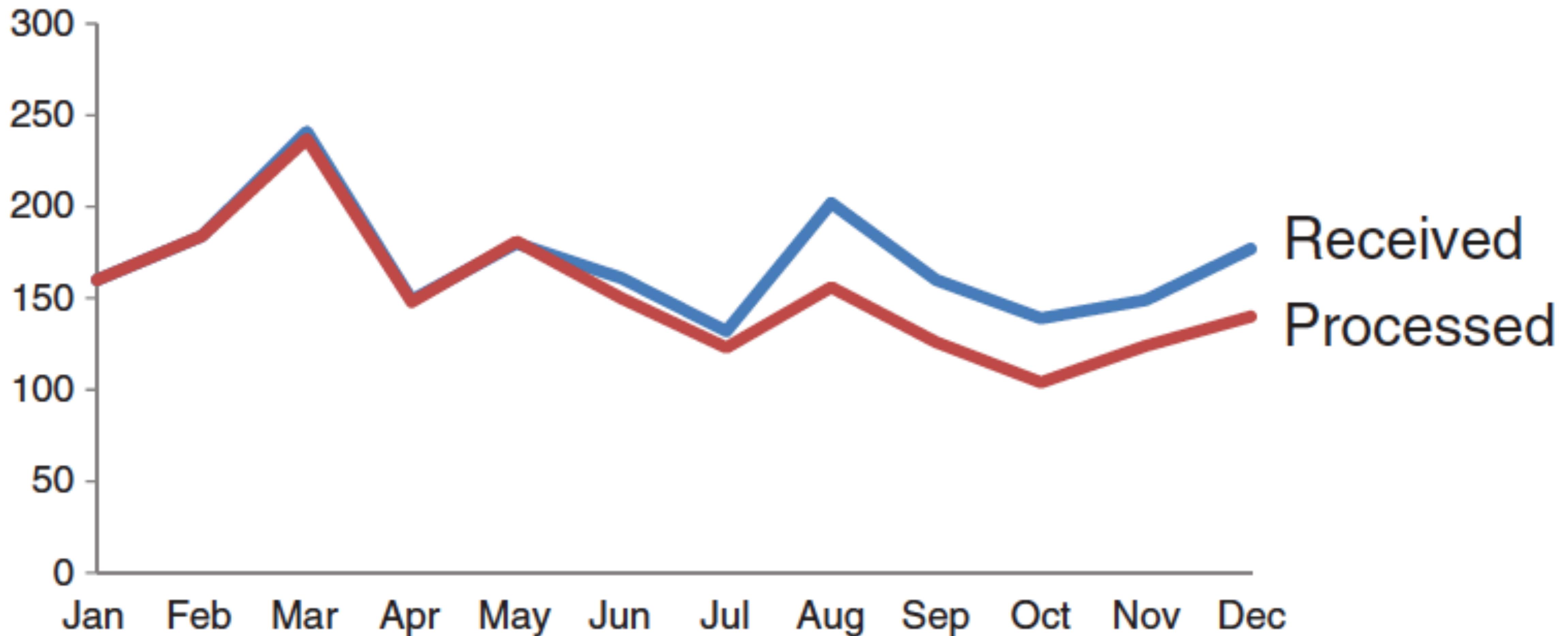
# ¿CÓMO LLAMAR LA ATENCIÓN?



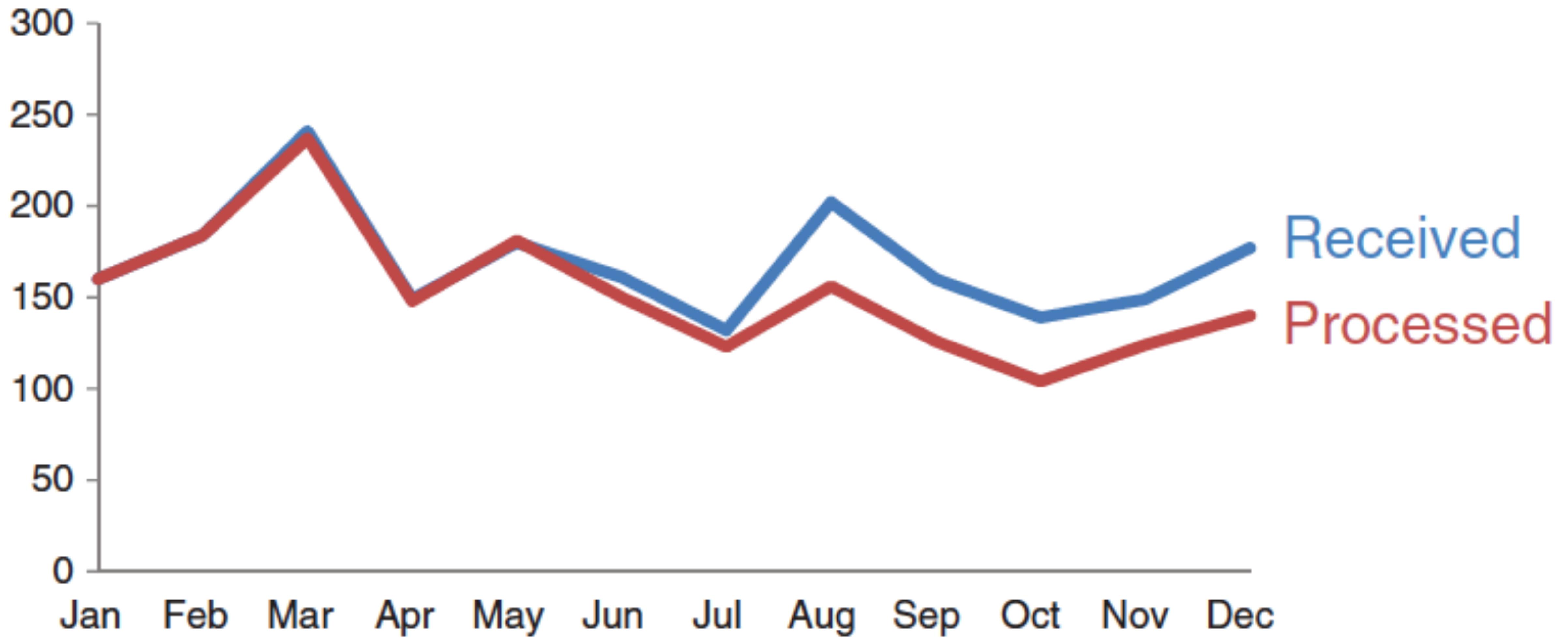
# ¿CÓMO LLAMAR LA ATENCIÓN?



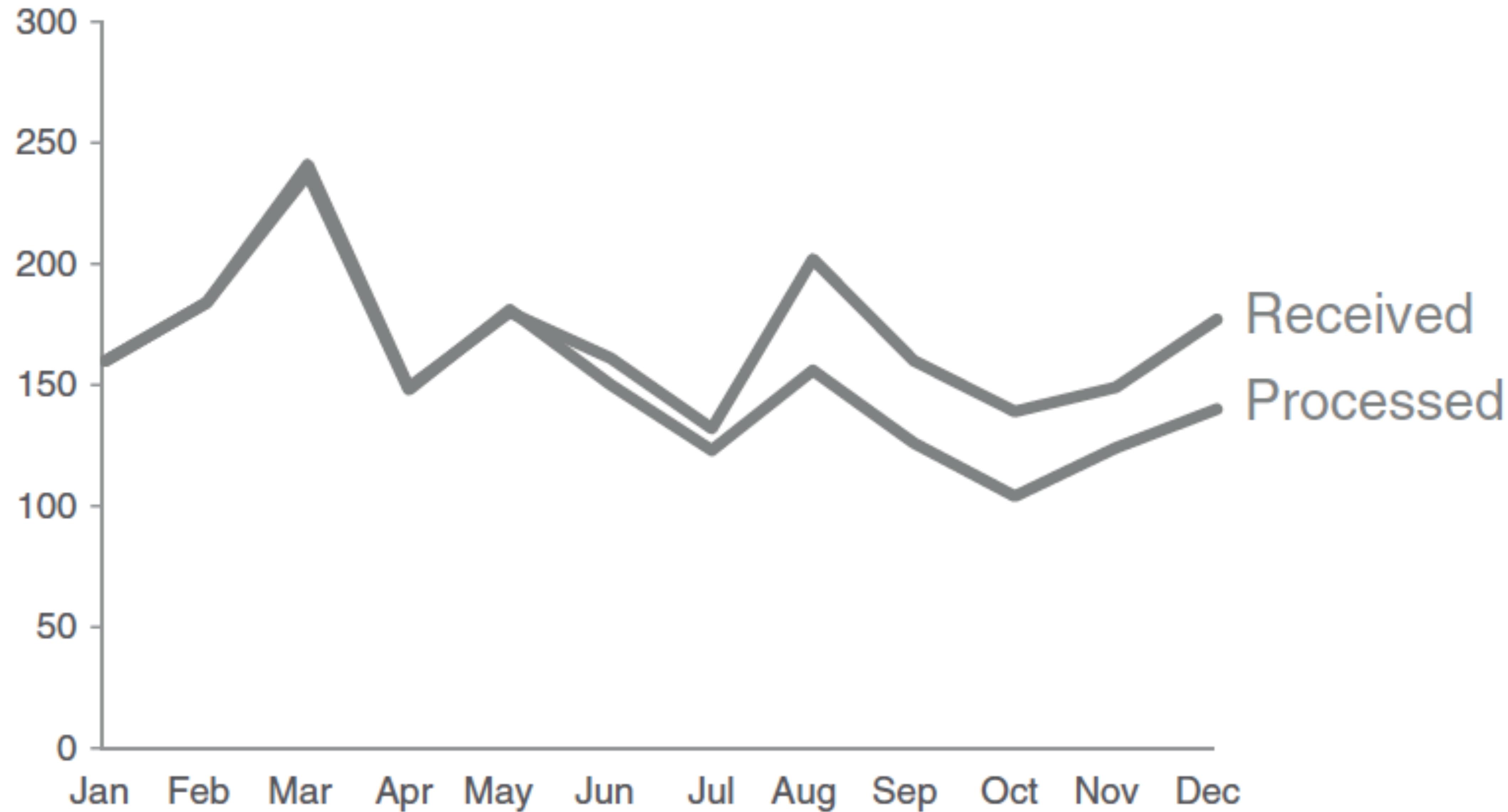
# ¿CÓMO LLAMAR LA ATENCIÓN?



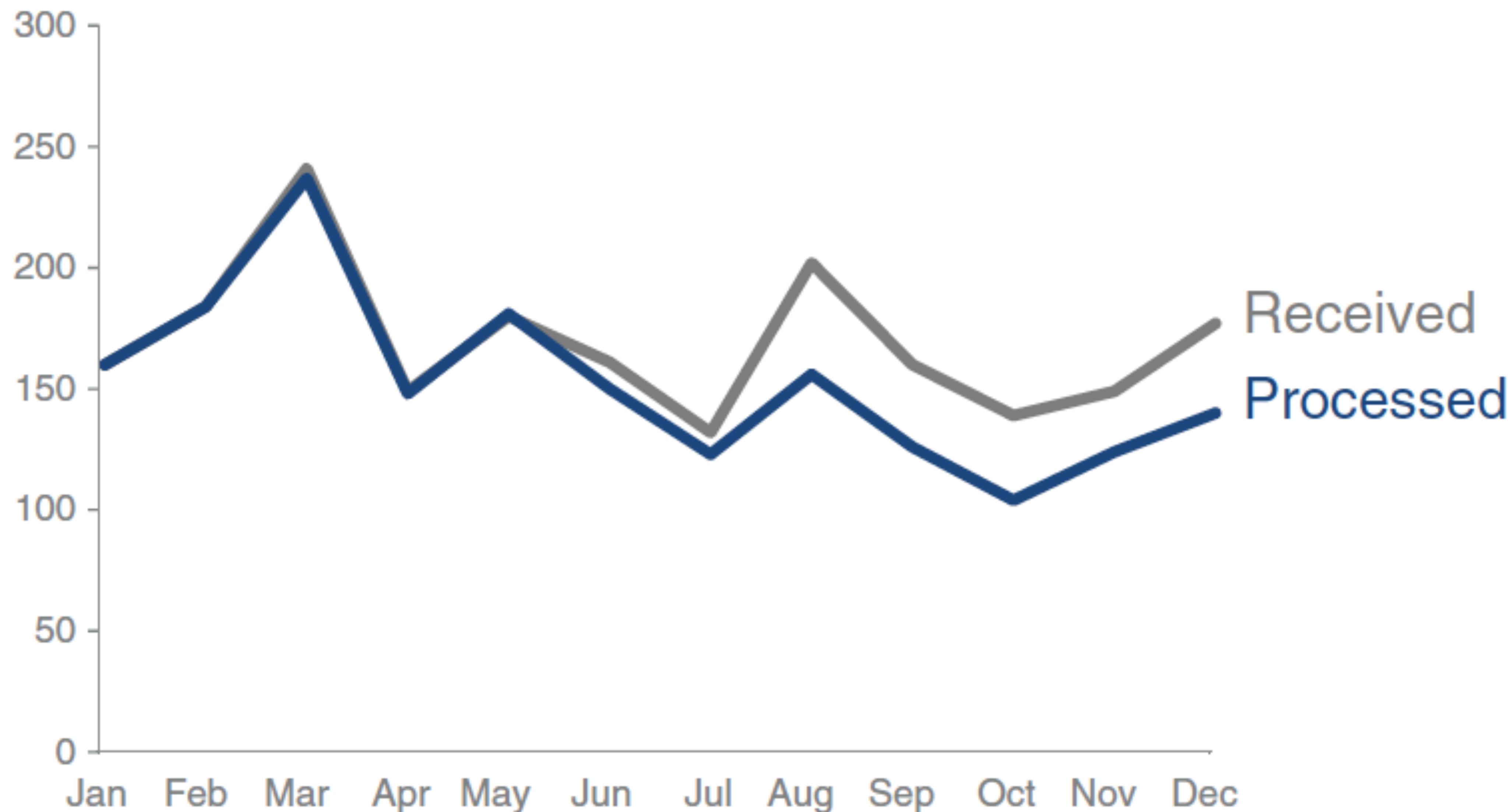
# ¿CÓMO LLAMAR LA ATENCIÓN?



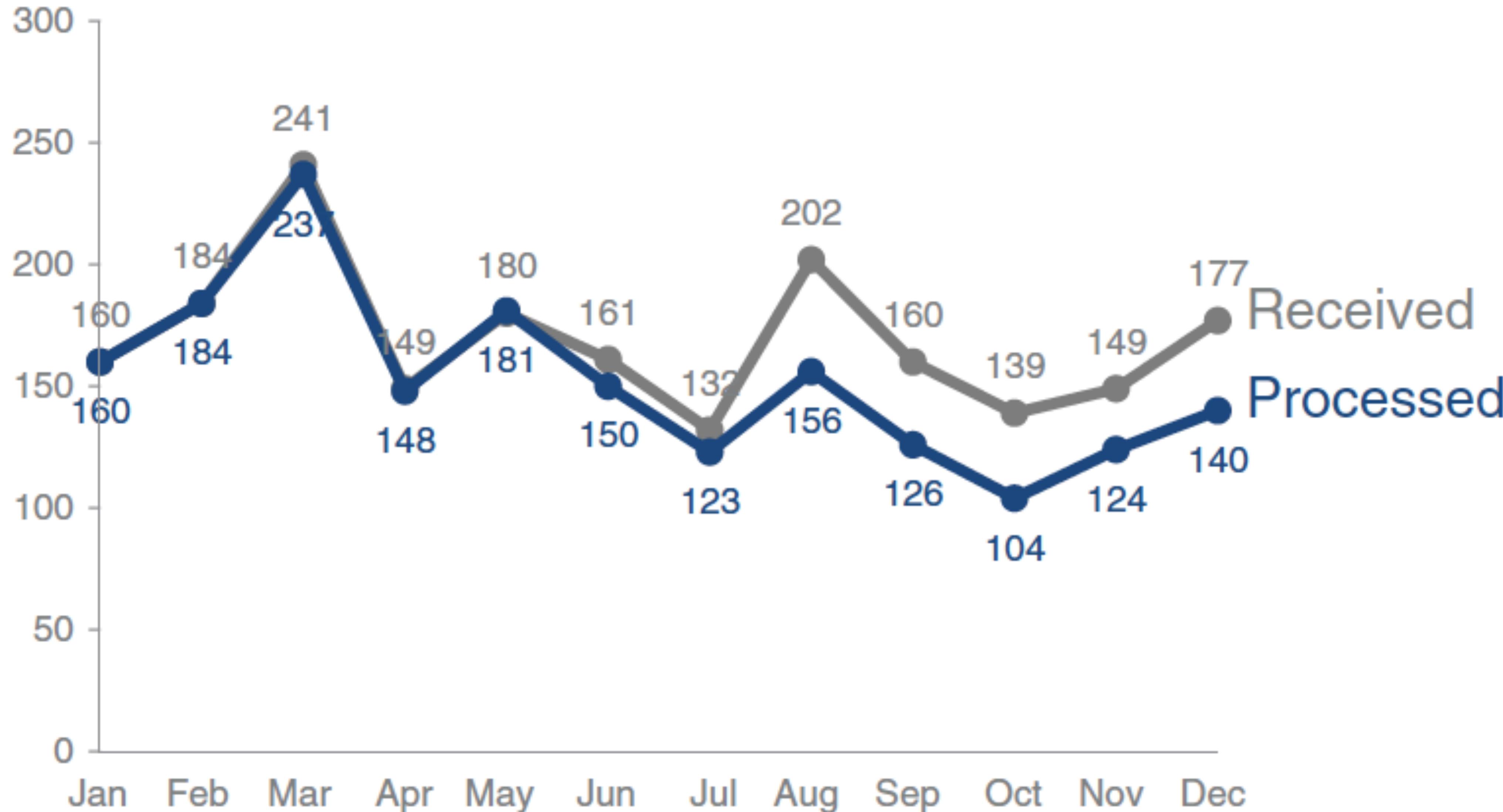
# ¿CÓMO LLAMAR LA ATENCIÓN?



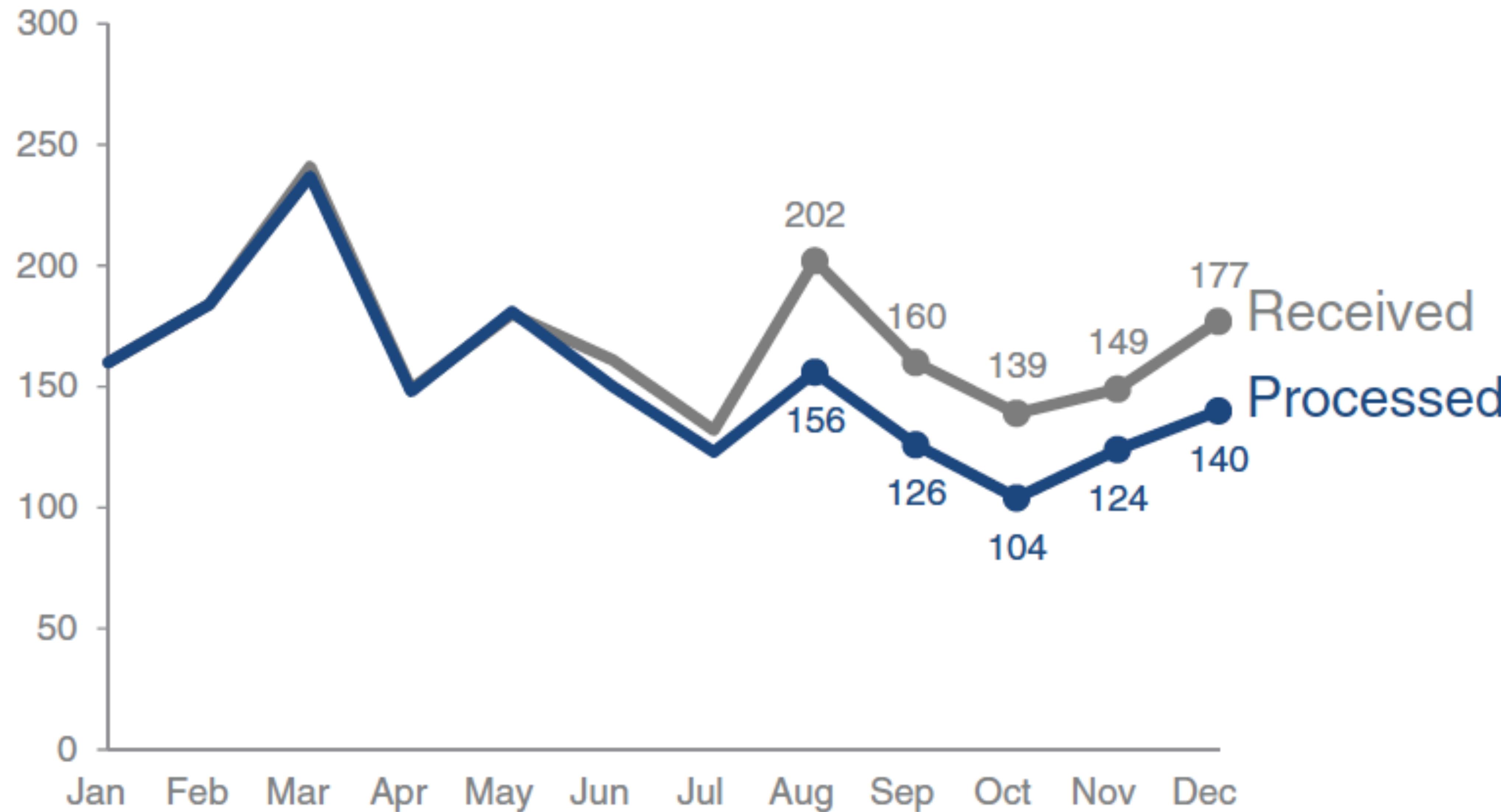
# ¿CÓMO LLAMAR LA ATENCIÓN?



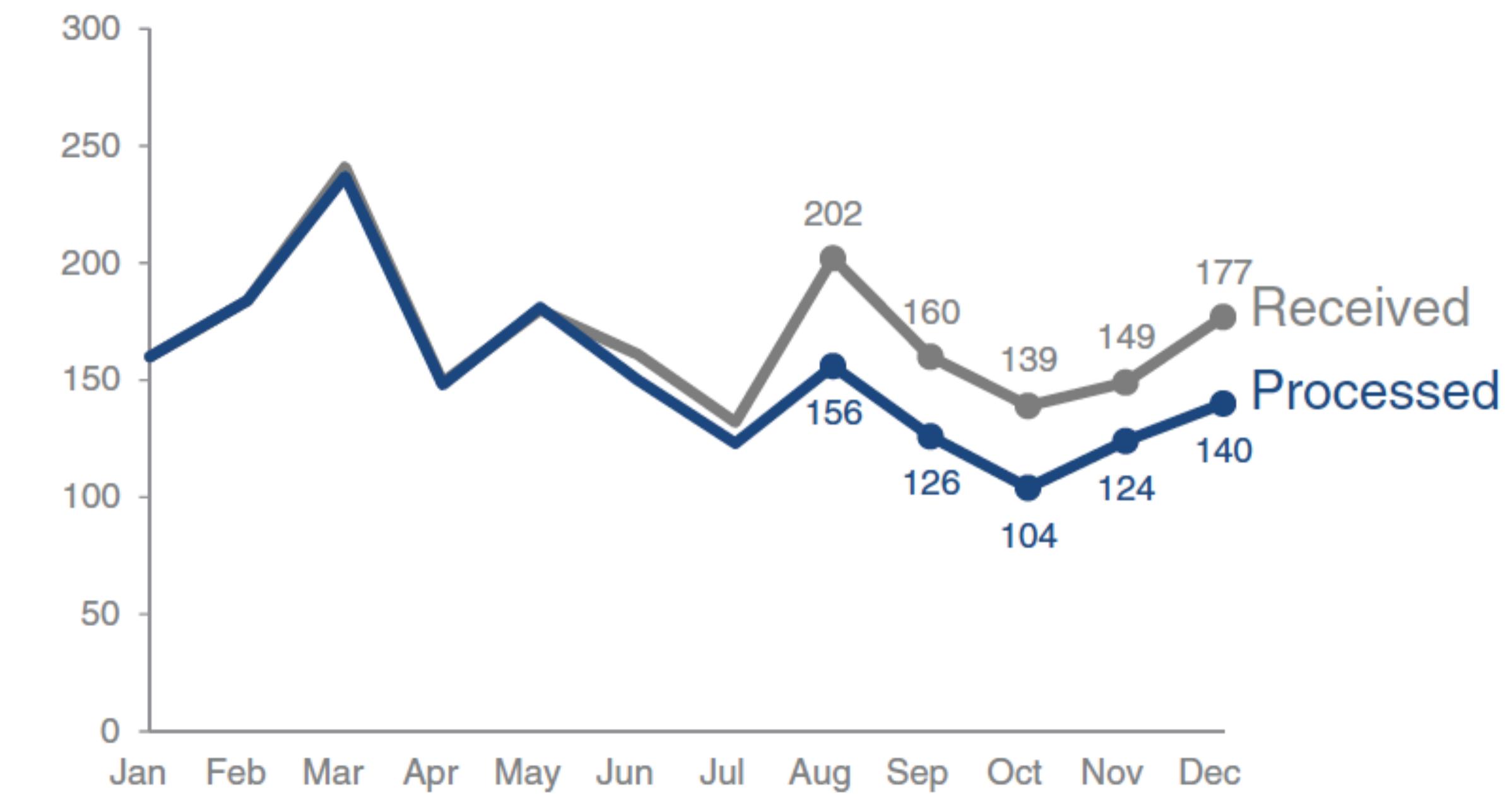
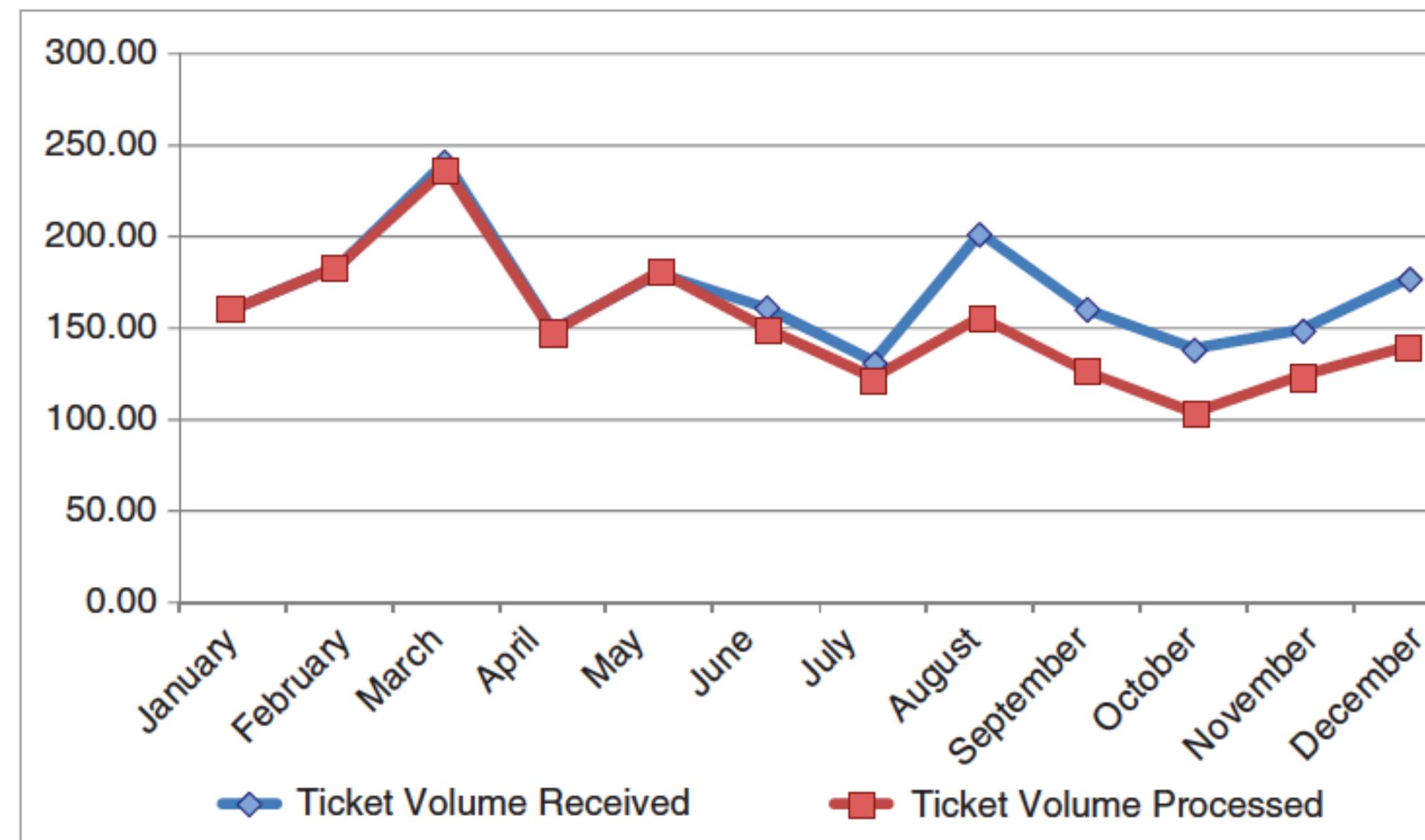
# ¿CÓMO LLAMAR LA ATENCIÓN?



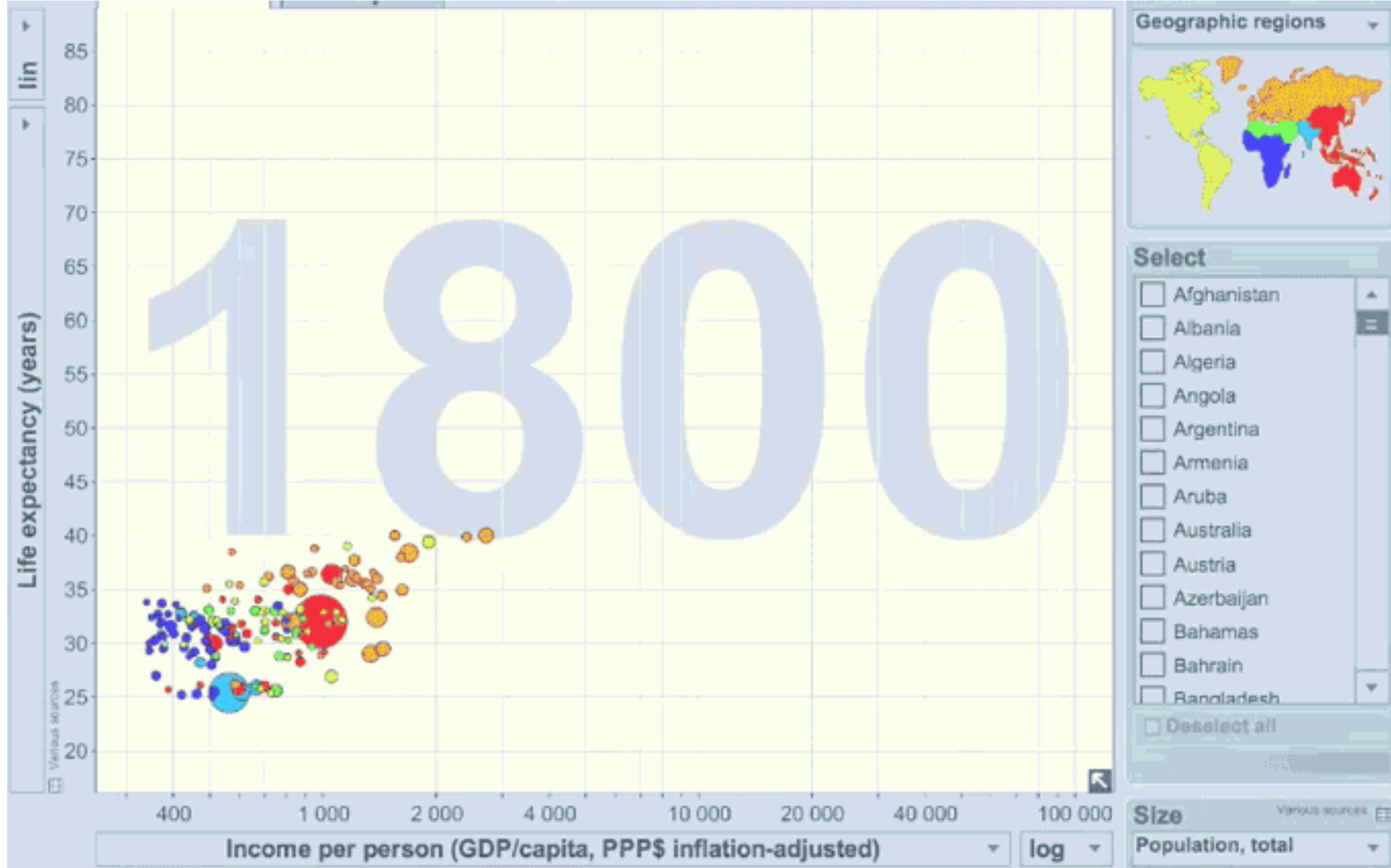
# ¿CÓMO LLAMAR LA ATENCIÓN?

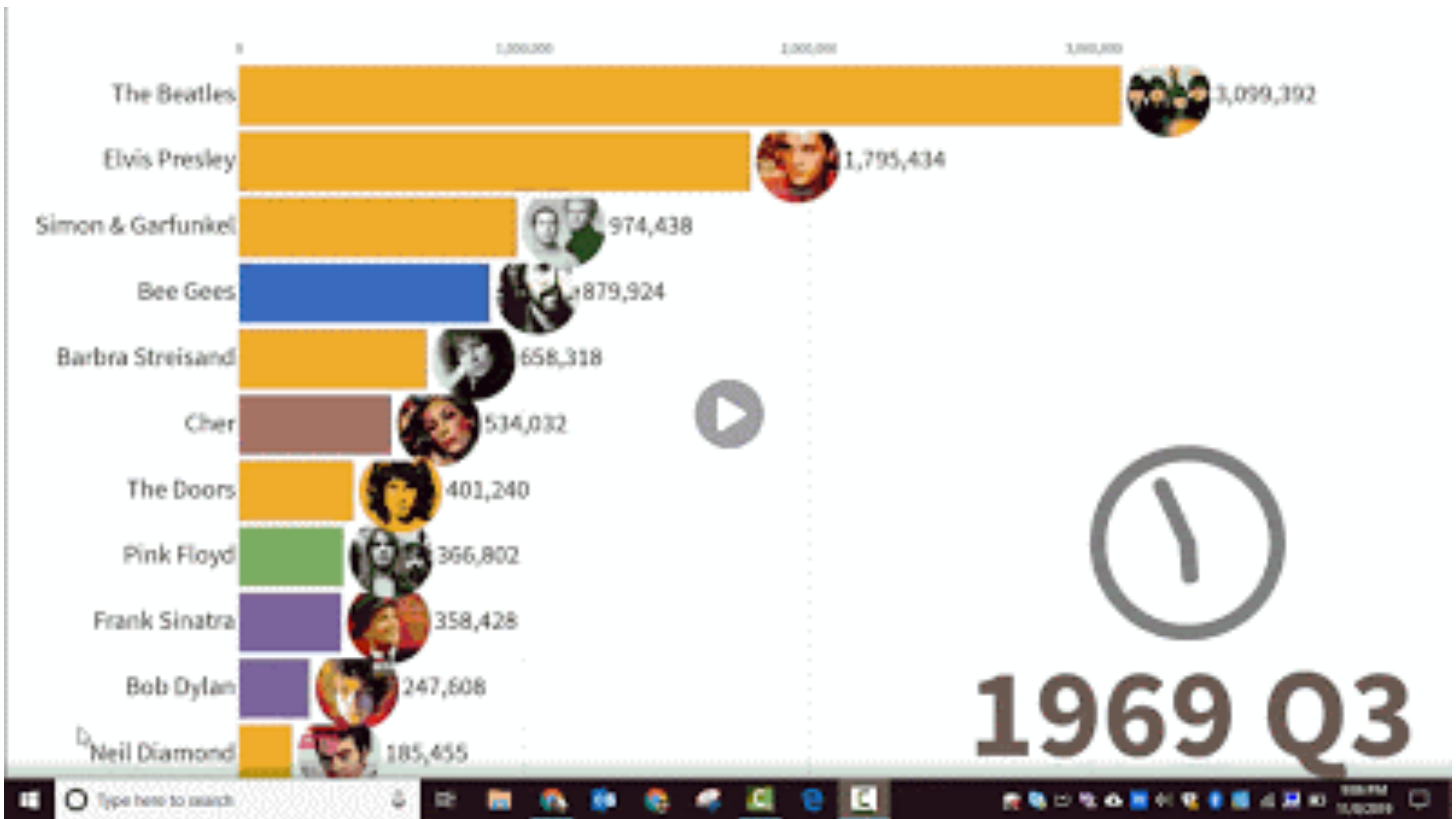


# ¿CÓMO LLAMAR LA ATENCIÓN?



**MOSTRAR LA EVOLUCIÓN A  
TRAVÉS DE GRÁFICOS  
ANIMADOS ES OTRA  
BUENA OPCIÓN**





# ALGUNOS TIPOS DE GRÁFICOS



# from Data to Viz

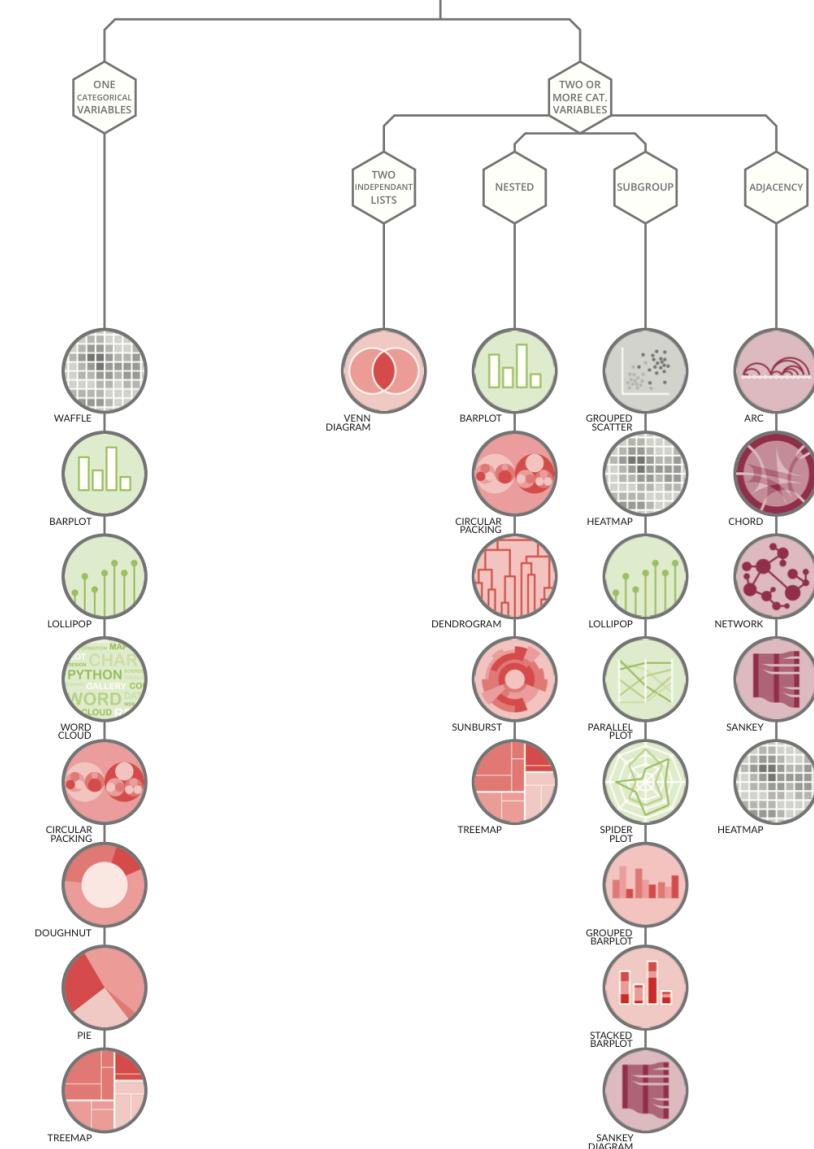
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- 1 Identify what type of data you have.
- 2 Go to the corresponding decision tree and follow it down to a set of possible charts.
- 3 Choose the chart from the set that will suit your data and your needs best.

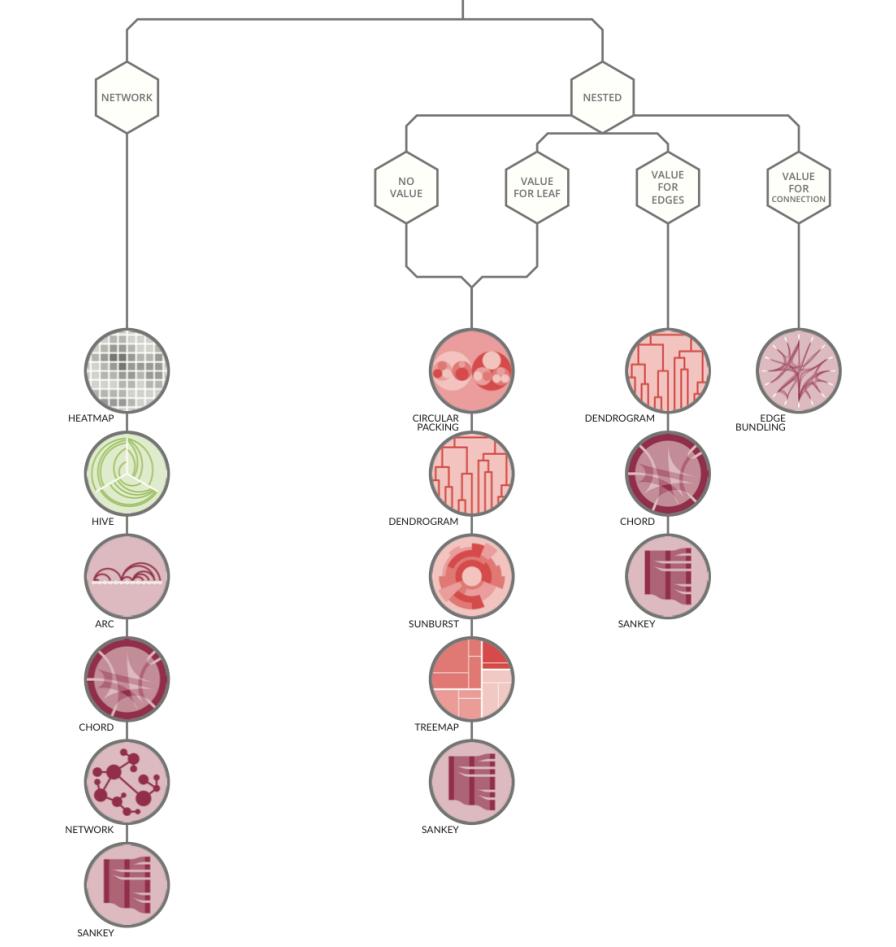
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

[data-to-viz.com](http://data-to-viz.com)

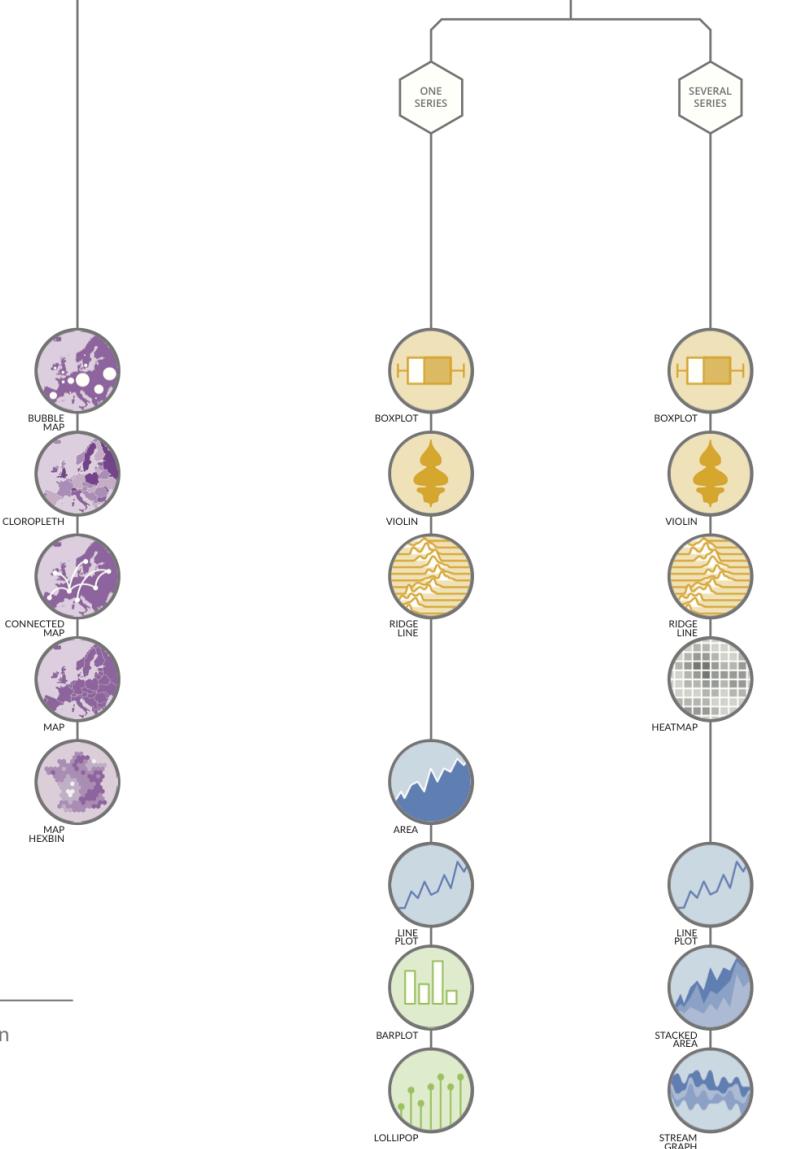
## CATEGORIC



## RELATIONAL



## MAP



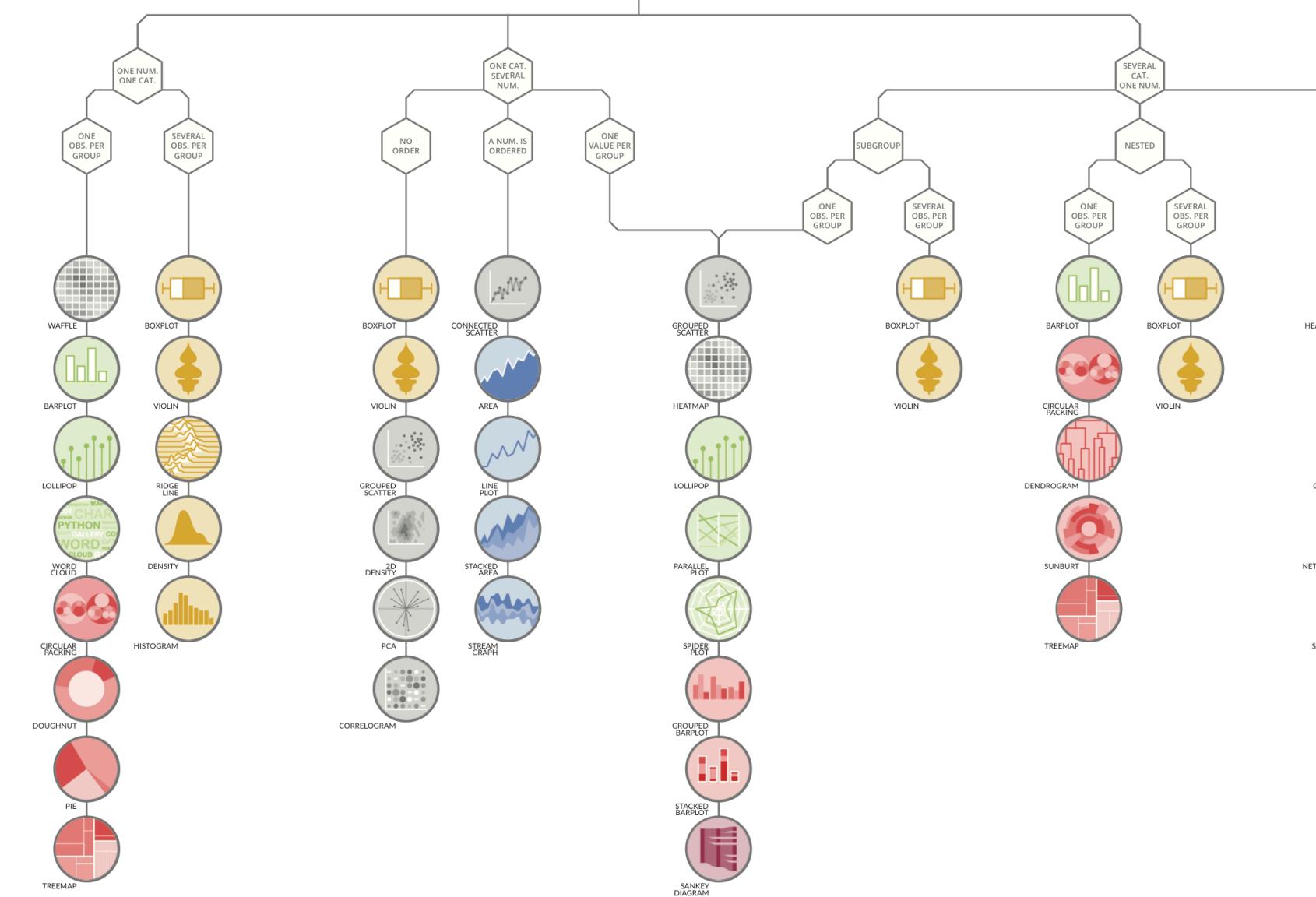
## TIME SERIES



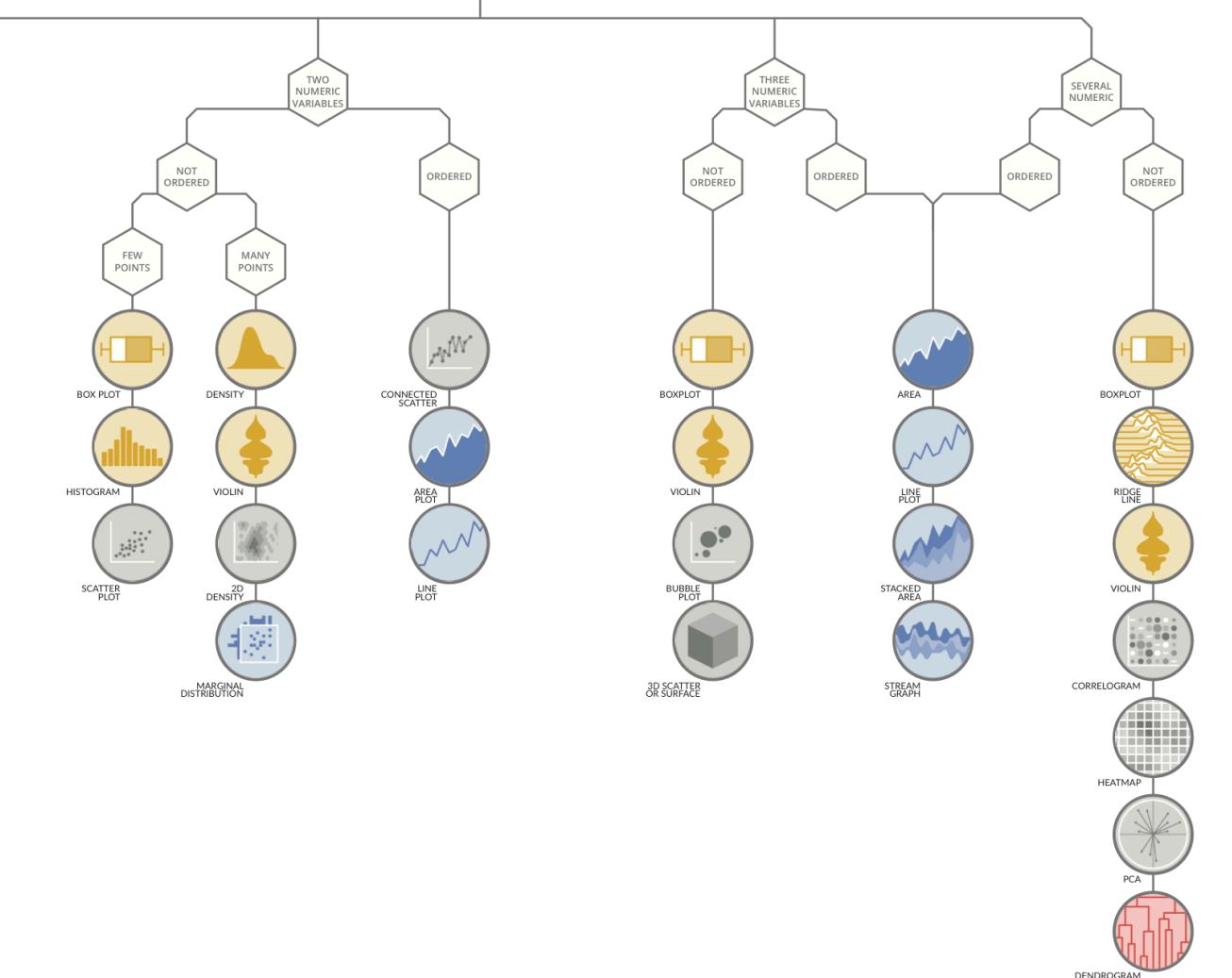
### WHAT DO YOU WANT TO SHOW ?

- Distribution
- Evolution
- Maps
- Ranking
- Part of a whole

## CATEGORIC AND NUMERIC



## NUMERIC





from Data to Viz

[HTTPS://WWW.DATA-TO-VIZ.COM/](https://www.data-to-viz.com/)

# VISUALIZACIÓN DE DATOS EN RSTUDIO

MAURICIO HUERTA AGUAR

[MAURICIO.HUERTA@PUCV.CL](mailto:MAURICIO.HUERTA@PUCV.CL)

ESCUELA DE INGENIERÍA INDUSTRIAL  
PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

CICLO DE MINICURSOS

SOCHE, 28 Y 30 DE OCTUBRE DE 2020