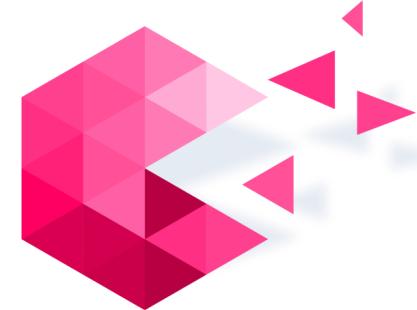


le
campus
numérique
in the ALPS

DATA VIZ

1. Graphs classiques
2. Pièges à éviter
3. Process de création d'un graph
4. Usage des couleurs
5. Cas d'études



le
campus
numérique
in the ALPS

Graphs classiques

Quand les utiliser

Graph classiques



df.plot



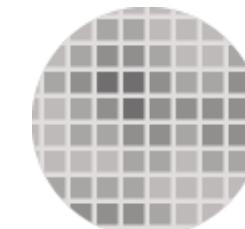
df.plot.area



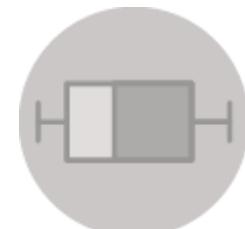
df.plot.area



df.plot.scatter



heatmap



df.boxplot



df.plot.bar



df.plot.barch



df.plot.bar
(stacked=True)



df.plot.hist



df.plot.pie

De quelles données je dispose ?

- Variable quantitative
- Variable qualitative

Qu'est ce que je souhaite montrer ?

- Distribution
- Relation
- Classement
- Proportion
- Evolution

Graph classiques: évolution



Line plot

- Montre l'évolution d'une ou plusieurs variables quantitatives
- Implique un lien entre deux points => ne fonctionne pas pour les variables qualitatives
 - Les points sont ordonnés
 - Très utilisé pour les séries temporelles



Area plot

Accentue la tendance ..



Stacked area plot

Montre à la fois l'évolution du total des variables quantitatives et l'importance de chaque variable

Graph classiques: classement



Bar plot

Montre un lien entre des variables qualitatives et quantitatives

- Une catégorie = une barre
- Taille de la barre = valeur numérique

Très connu => facile à lire pour tout le monde
L'oeil compare facilement la taille des barres
L'ordre est important



Horizontal bar plot

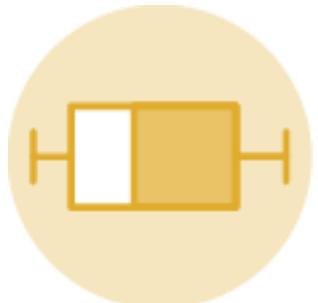
Souvent plus simple à lire à l'horizontal,
Pratique si les noms de categories sont longs

Graph classiques: distributions



Hist

- Représente la distribution d'une variables quantitative
- La variable est découpée en plusieurs classes (bins)
 - Nombre d'observation par classe = hauteur de chaque barre
 - Penser à faire varier la taille des classes



Boxplot

- Résume simplement la distribution d'une ou plusieurs variables numérique

Graph classiques: relations



Scatter plot

Montre la relation entre deux variables quantitatives



Heatmap

Vue générale d'un tableau contenant des valeurs quantitatives

- Réduit le temps de traitement par rapport à un tableau
- Attention aux couleurs
- Penser à permuter les colonnes / lignes pour mieux visualiser des tendances

Graph classiques: proportions



Pie plot

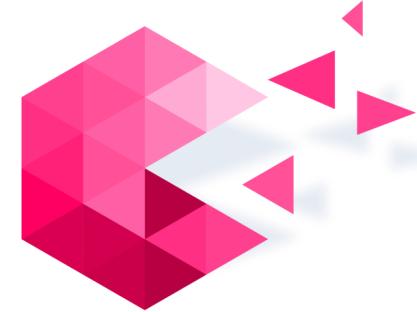
Illustre des proportions pour des variables qualitatives

- La somme doit être égale à 100% pour montrer une proportion
- Mal aimé ...



Stacked bar plot

Illustre une répartition et un total
Attention à l'ordre



le
campus
numérique
in the ALPS

Graphs classiques

Pièges à éviter

Une mine d'informations:
<https://www.data-to-viz.com/caveats.html>

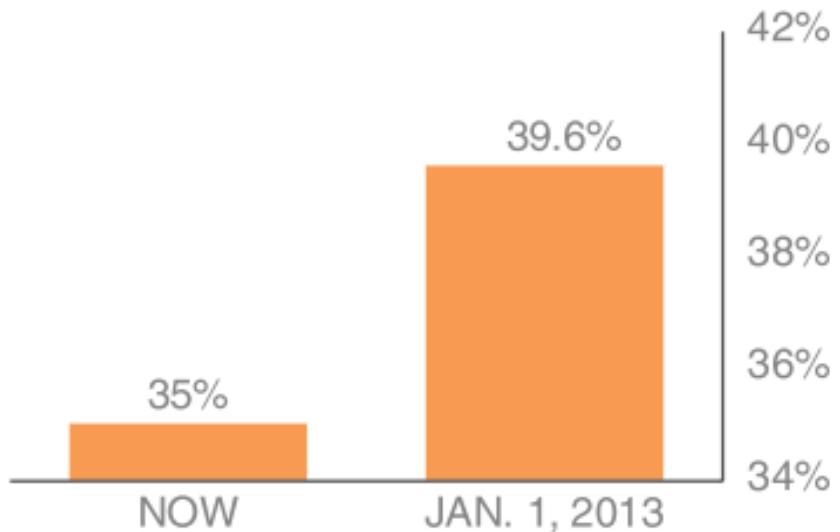
Bar graph



Bar graph: baseline

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

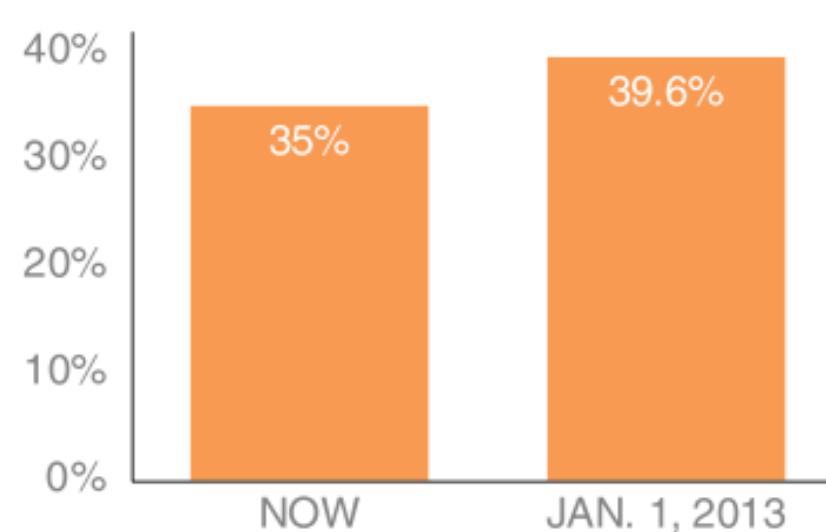
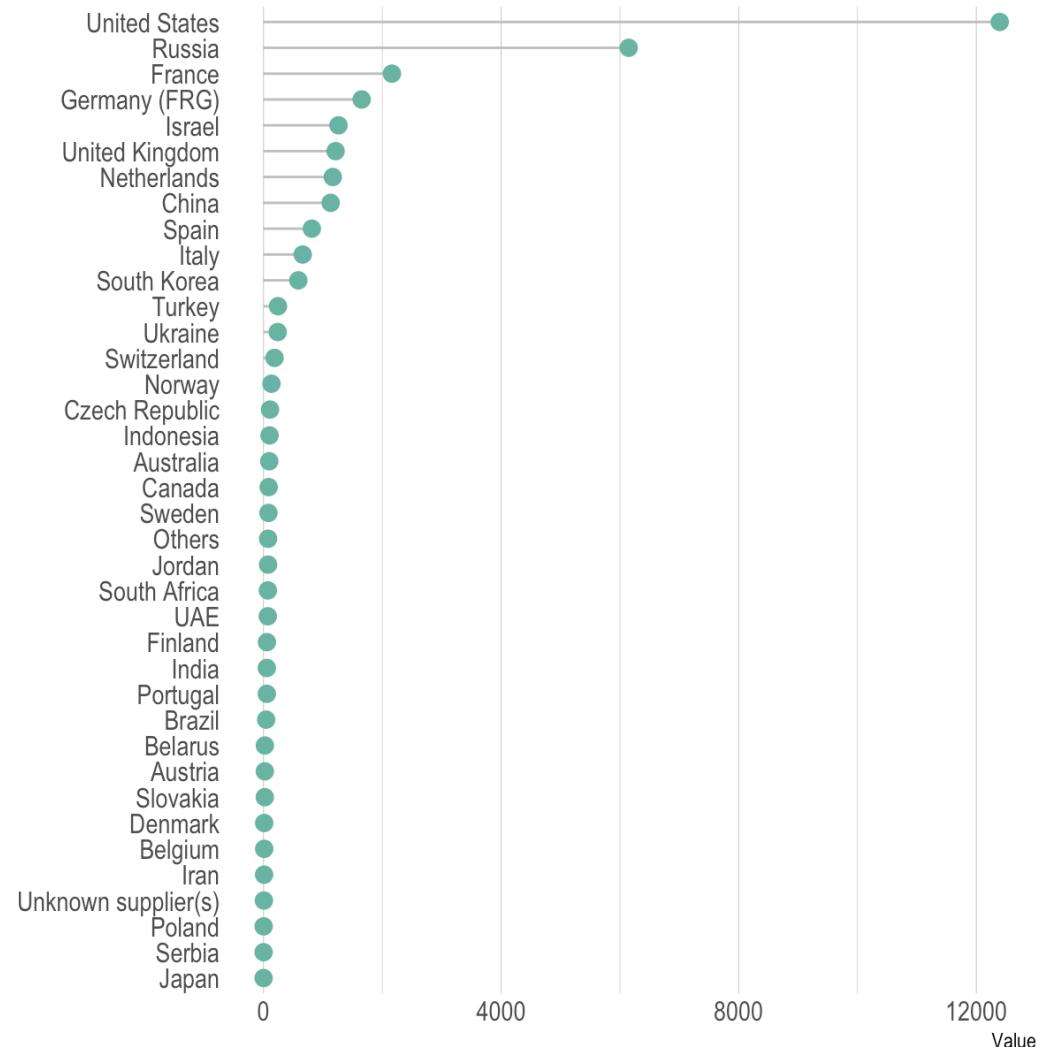
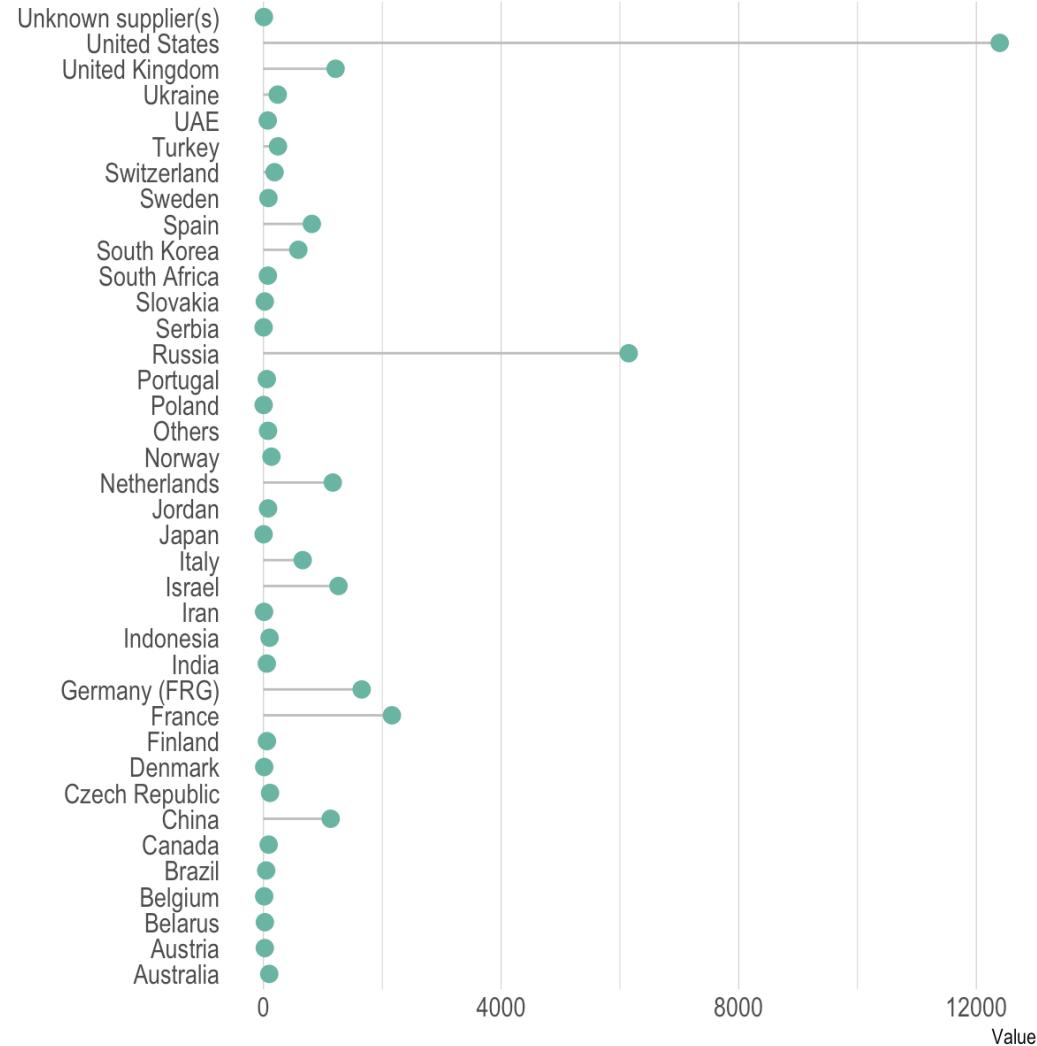


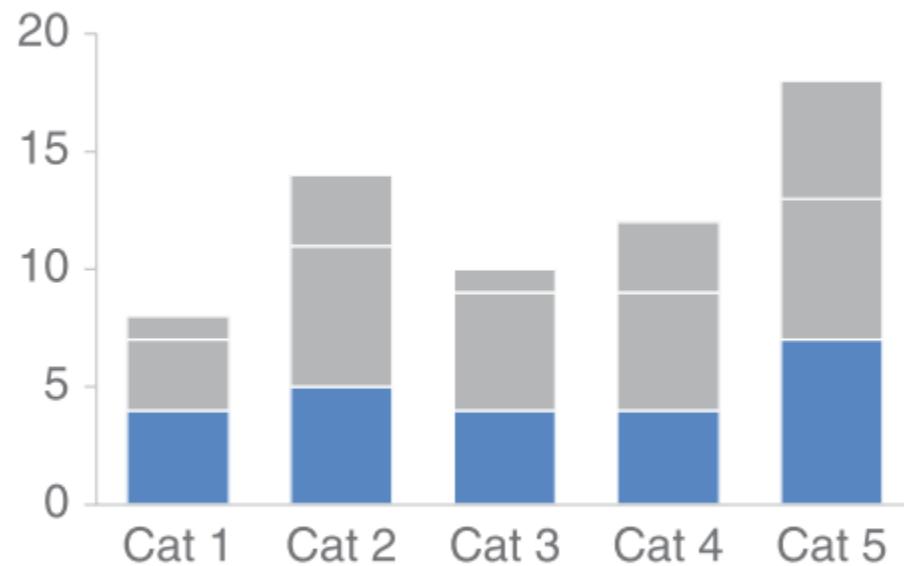
FIGURE 2.13 Bar charts must have a zero baseline

Bar graph: order your data



Bar graph: order your data

Comparing **these** is easy



Comparing **these** is hard

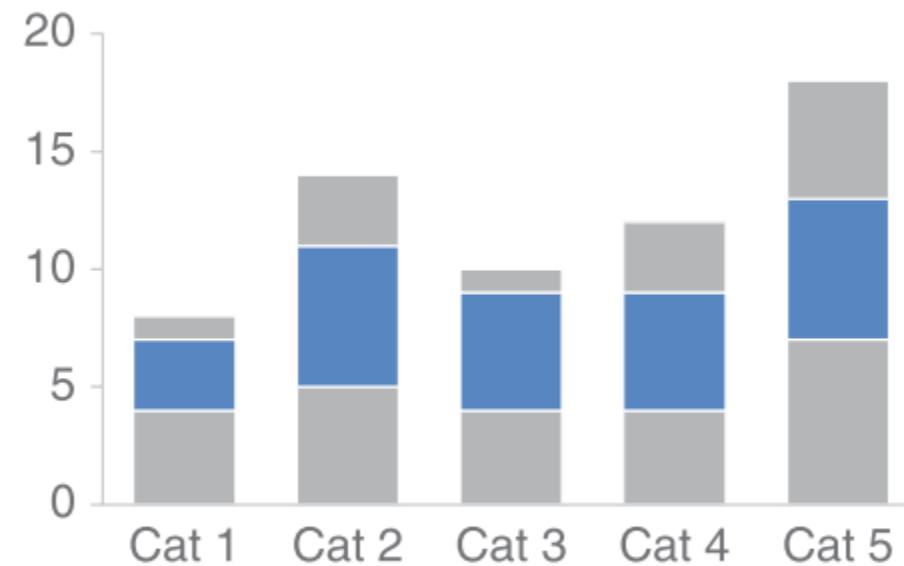


FIGURE 2.16 Comparing series with stacked bar charts

Bar graph

NOT IDEAL

TOP RANKED UNIS

MIT	
HARVARD	
STANFORD	
CALTECH	

NOT IDEAL

CANDIDATE	USD ▾
DANIEL	9.53M
EGON	2.01M
CORY	1.24M
AMY	1.12M
BILL	0.34M

BETTER

TOP RANKED UNIS

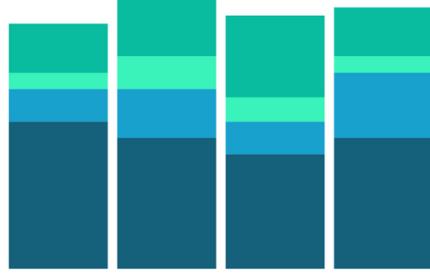
1	MIT
2	HARVARD
3	STANFORD
4	CALTECH

BETTER

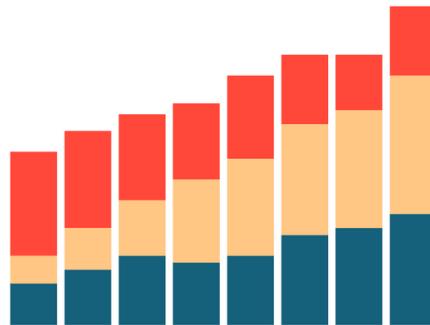
CANDIDATE	USD ▾
DANIEL	9.53M 
EGON	2.01M 
CORY	1.24M 
AMY	1.12M 
BILL	0.34M 

Stacked bar

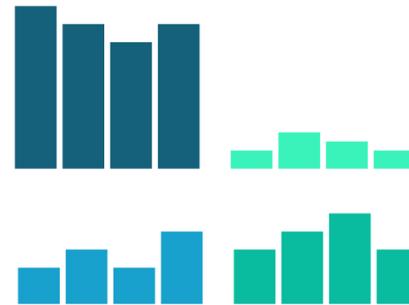
NOT IDEAL



NOT IDEAL



BETTER



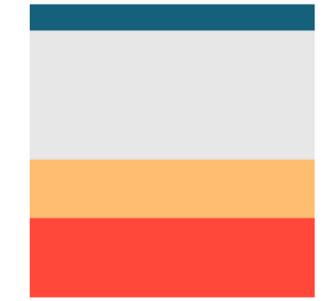
BETTER



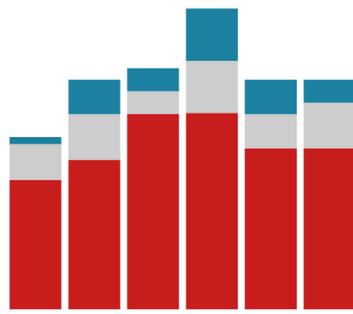
! Time data !

Stacked bar

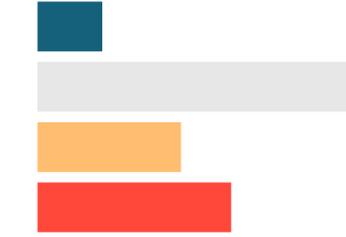
NOT IDEAL



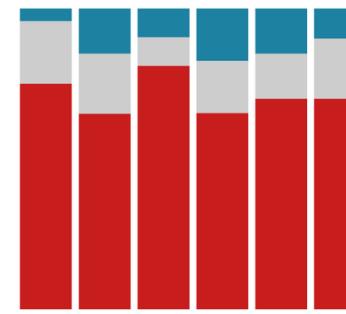
NOT IDEAL



BETTER



BETTER



!\\ Consider stacking
percentage

Bar graph

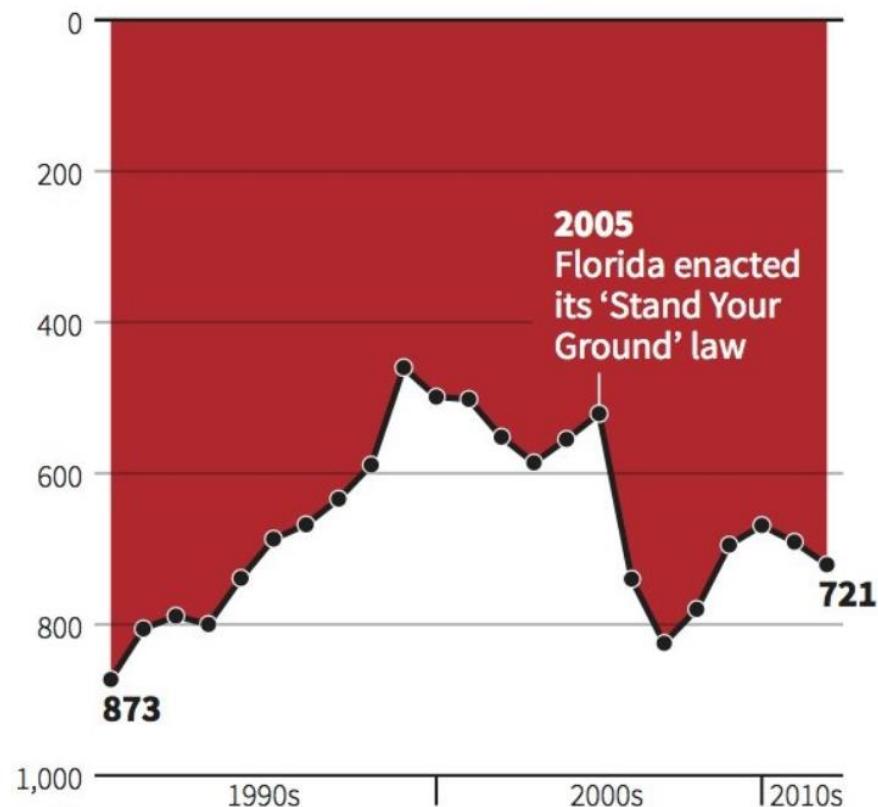
Référence:

- [Storytelling with data](#), a data vizualization guide for professionnals, Cole Nussbaumer Knafic
- https://www.data-to-viz.com/caveat/order_data.html
- <https://blog.datawrapper.de/guide-what-to-consider-when-creating-tables/>
- <https://blog.datawrapper.de/stacked-column-charts/>

Line graph: axis

Gun deaths in Florida

Number of murders committed using firearms



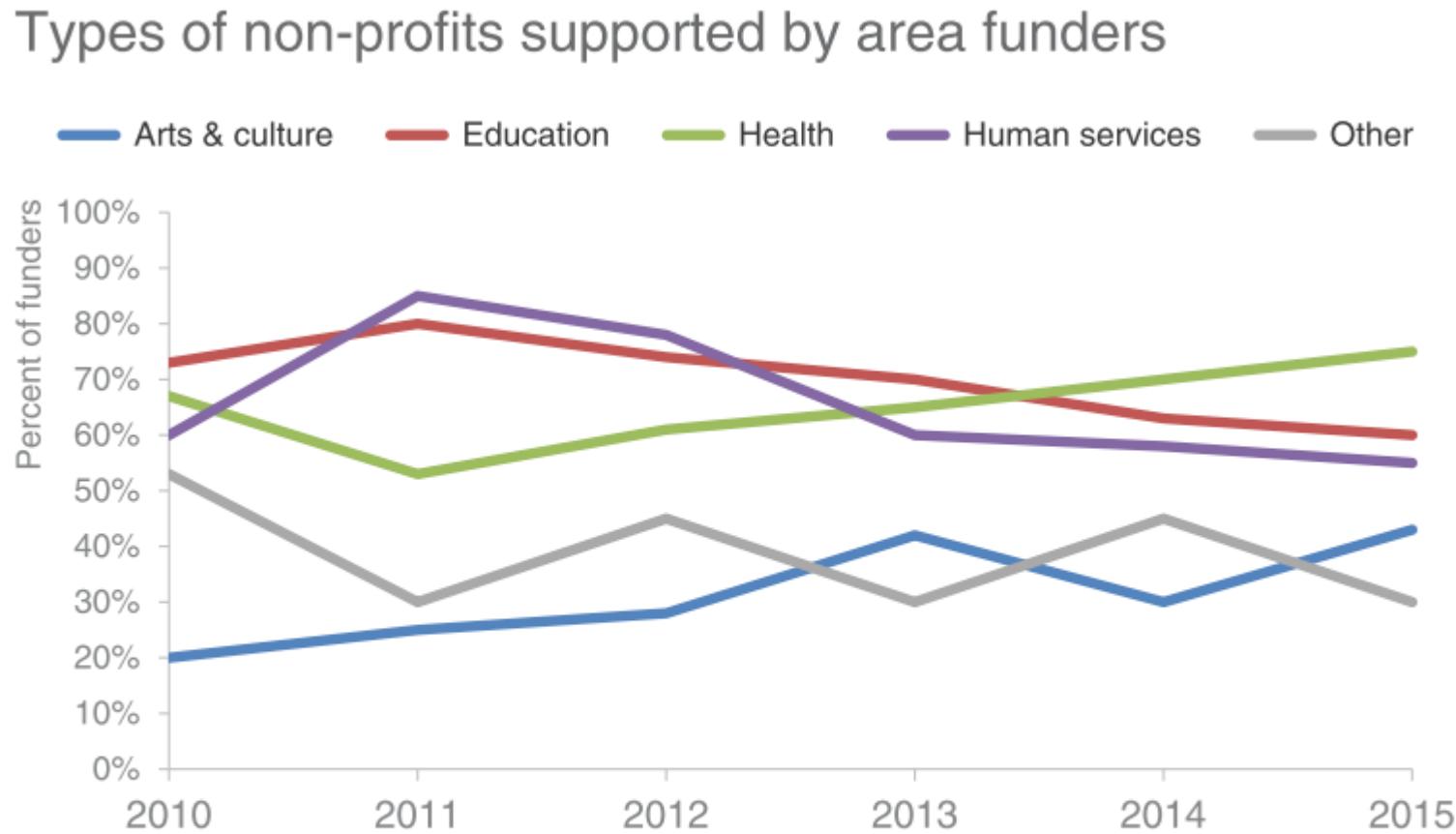
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

https://www.data-to-viz.com/caveat/counter_intuitive.html

Line graph: spaghetti



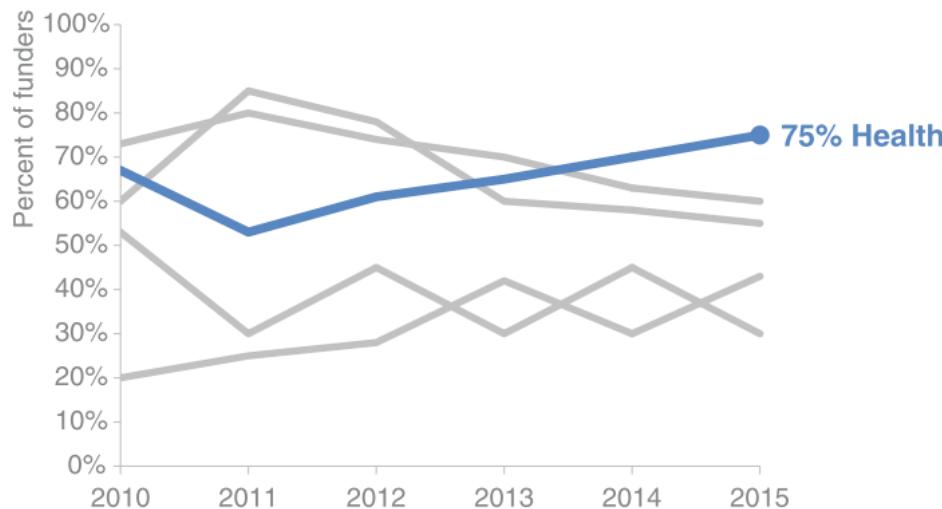
Data is self-reported by funders; percents sum to greater than 100 because respondents can make multiple selections.

FIGURE 9.21 The spaghetti graph

Line graph: spaghetti

Emphasize one line at a time

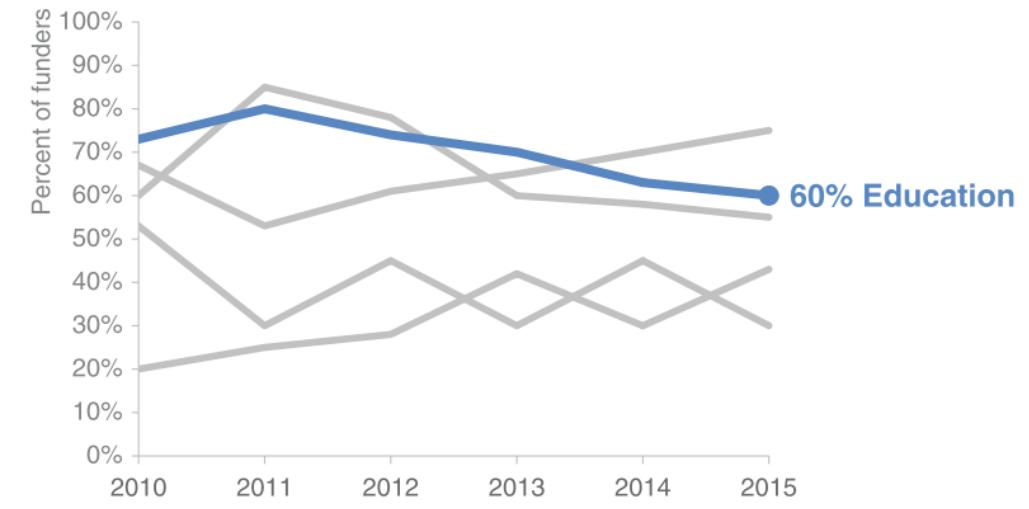
Types of non-profits supported by area funders



Data is self-reported by funders; percents sum to greater than 100 because respondents can make multiple selections.

FIGURE 9.22 Emphasize a single line

Types of non-profits supported by area funders



Data is self-reported by funders; percents sum to greater than 100 because respondents can make multiple selections.

FIGURE 9.23 Emphasize another single line

Line graph: spaghetti

Separate spatially

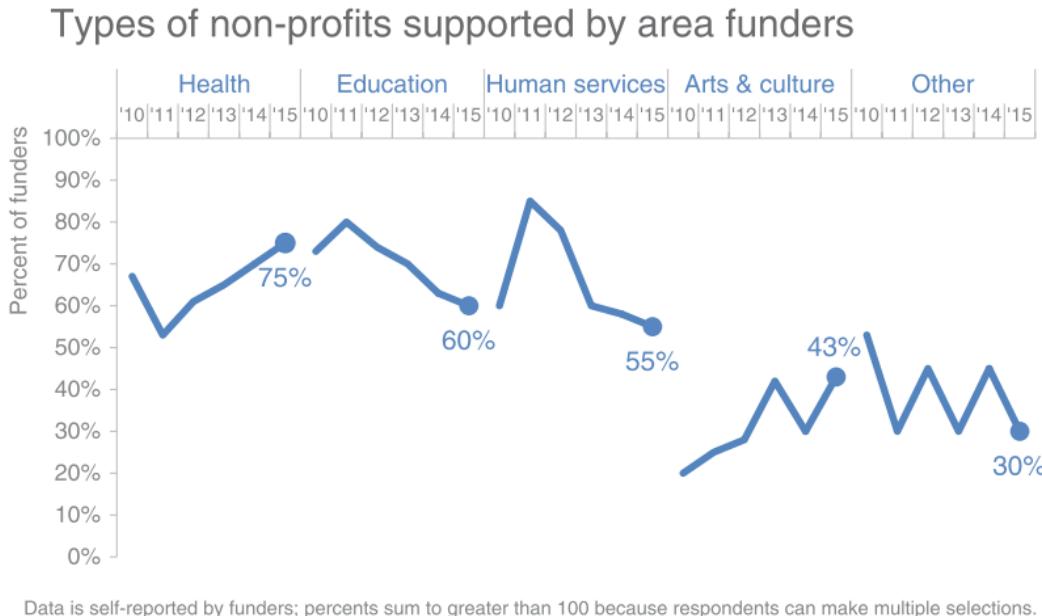


FIGURE 9.25 Pull the lines apart horizontally

Types of non-profits supported by area funders

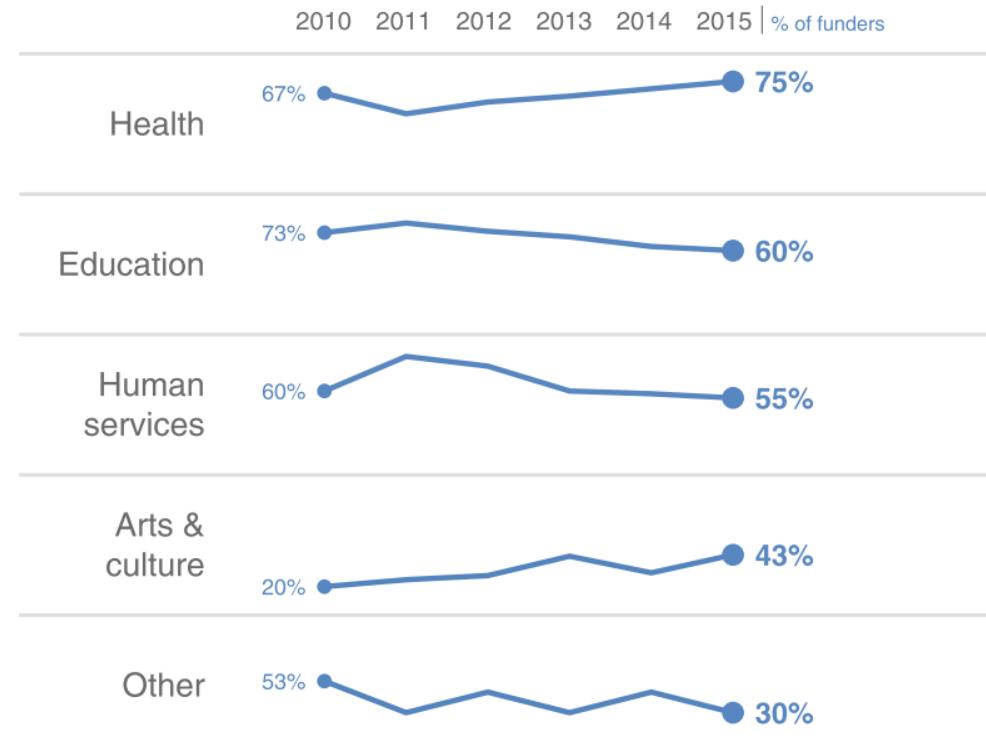


FIGURE 9.24 Pull the lines apart vertically

Line graph: spaghetti

Combine both approaches

Types of non-profits supported by area funders

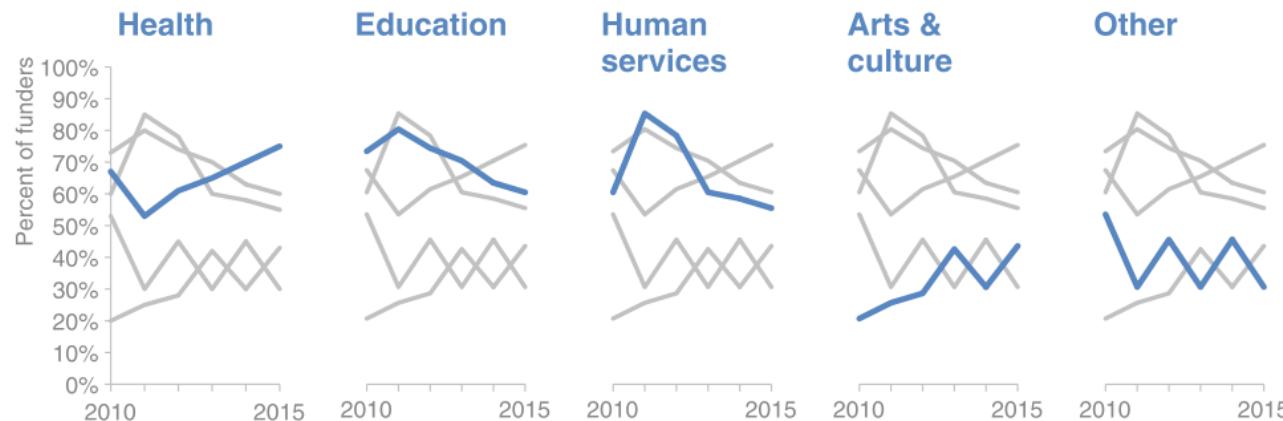


FIGURE 9.27 Combined approach, with horizontal separation

Types of non-profits supported by area funders

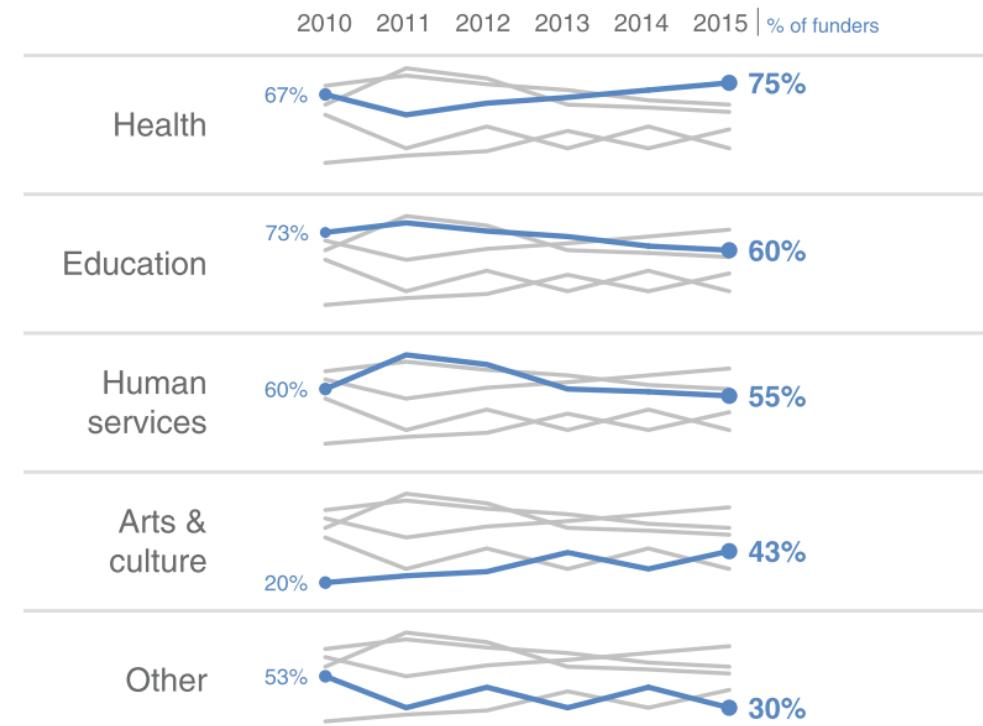


FIGURE 9.26 Combined approach, with vertical separation

Line graph: spaghetti

Référence:

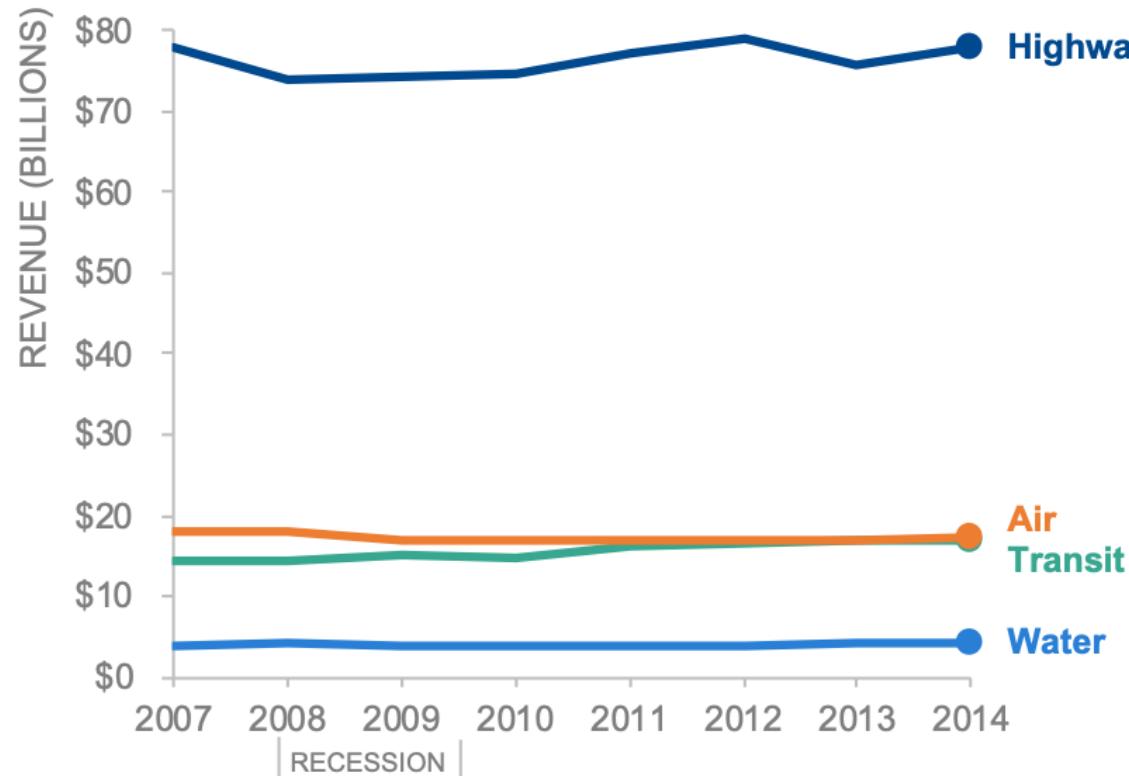
Storytelling with data, a data vizualization guide for professionnals,
Cole Nussbaumer Knafic

A lire également:

<https://www.data-to-viz.com/caveat/spaghetti.html>

Line graph: range

State and local revenue by transportation mode

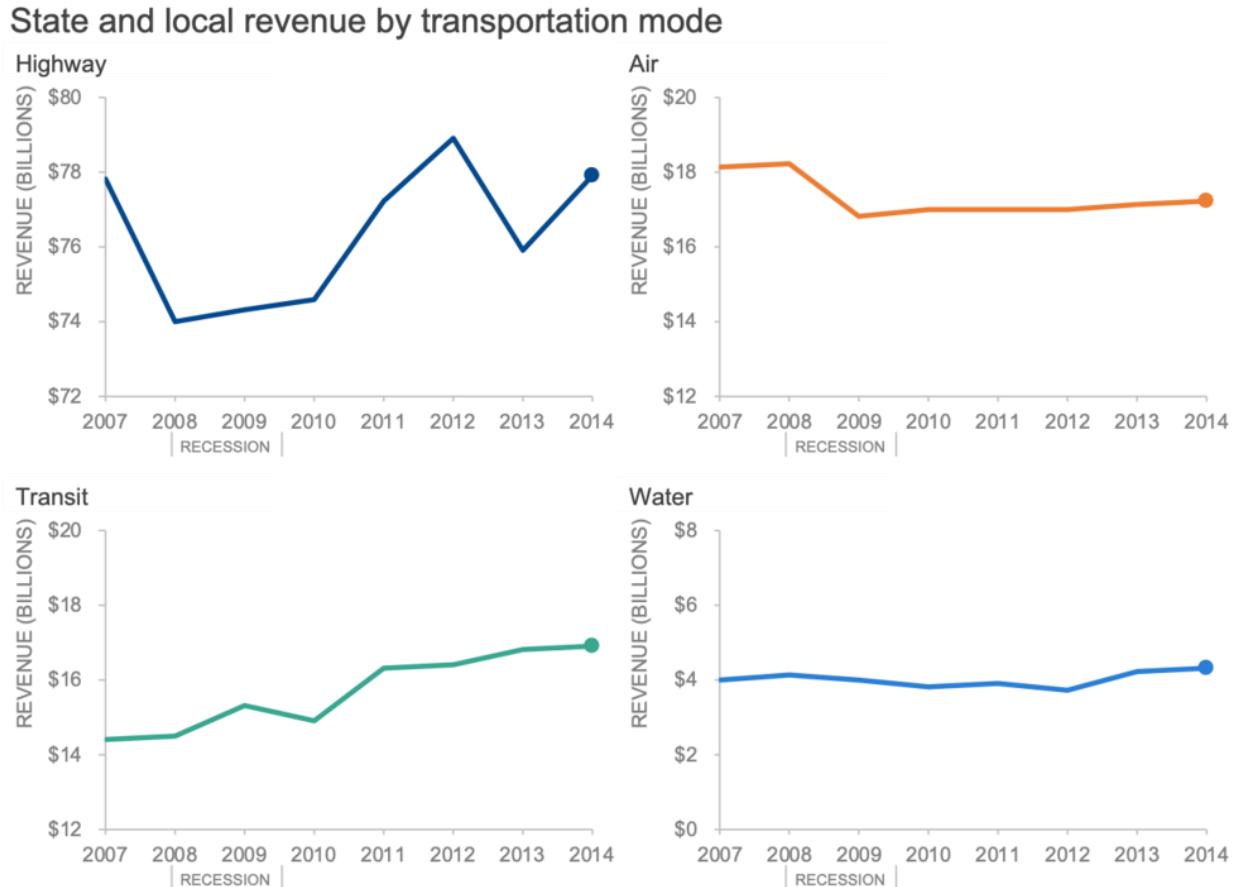


Common challenge:

When one data series is so large relatively to the others that a single scale makes it nearly impossible to see any details.

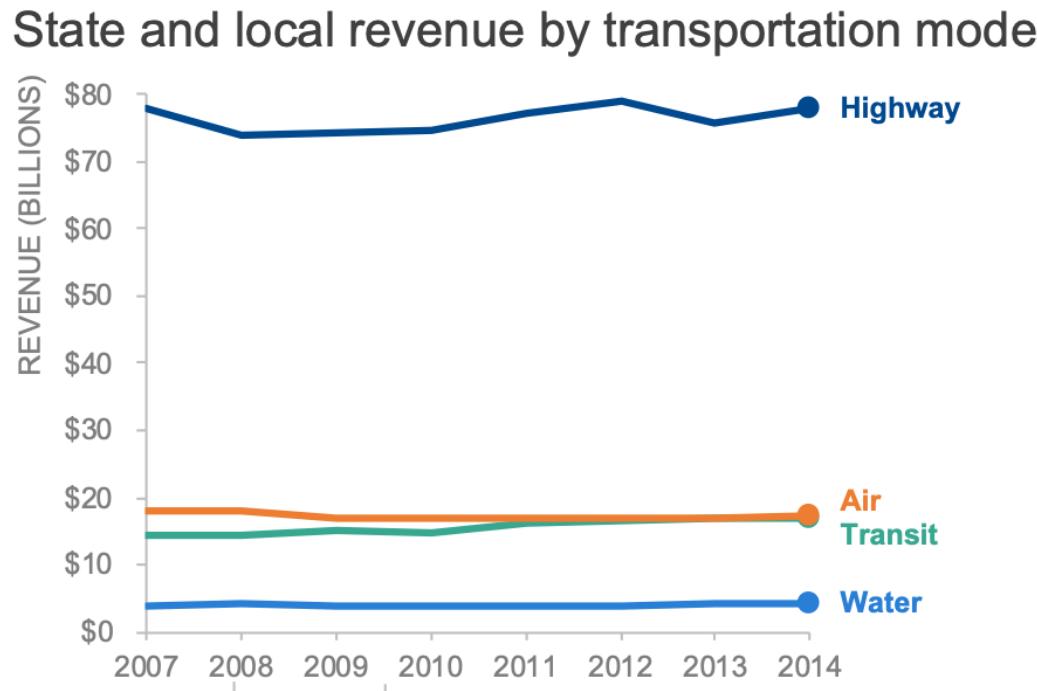
SOURCE: Bureau of Transportation Statistics 2018 Report, Trends in State and Local Own Source Revenue by Mode, 2007 to 2014 (Billions of Chained 2009 Dollars)

Line graph: range



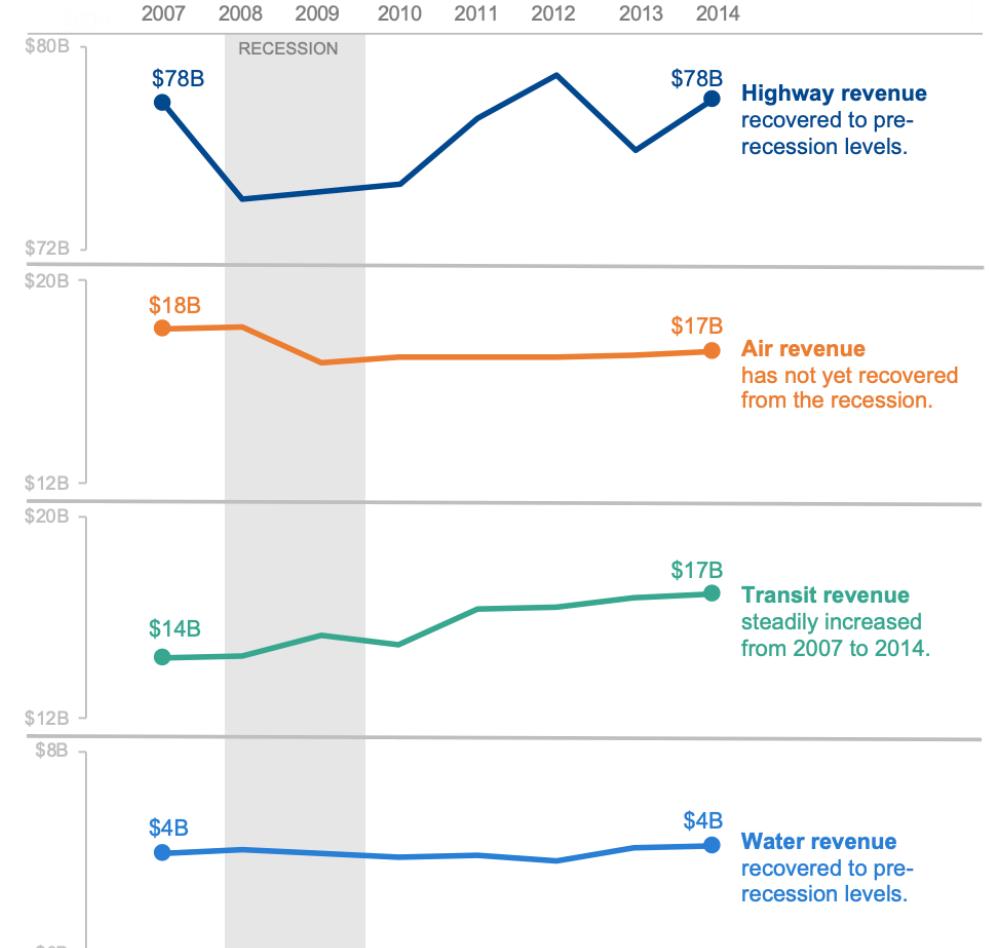
SOURCE: Bureau of Transportation Statistics 2018 Report, Trends in State and Local Own Source Revenue by Mode, 2007 to 2014 (Billions of Chained 2009 Dollars)

Line graph: range



SOURCE: Bureau of Transportation Statistics 2018 Report, Trends in State and Local Own Source Revenue by Mode, 2007 to 2014 (Billions of Chained 2009 Dollars)

State and local revenue by transportation mode



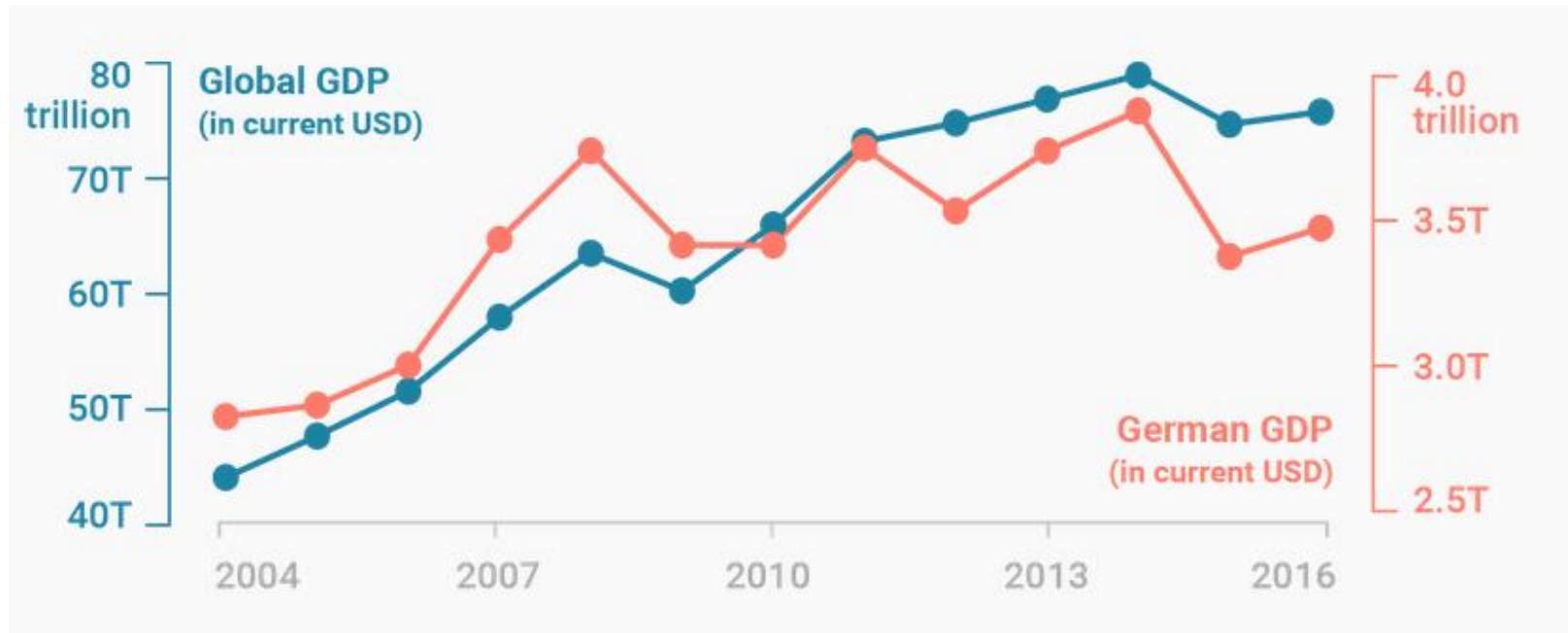
SOURCE: Bureau of Transportation Statistics 2018 Report, Trends in State and Local Own Source Revenue by Mode, 2007 to 2014 (Billions of Chained 2009 Dollars)

Line graph: range

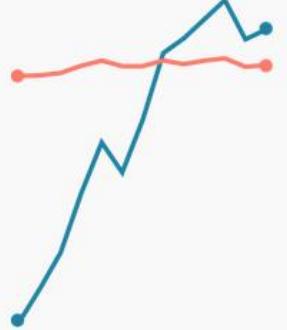
Référence:

<http://www.storytellingwithdata.com/blog/2020/1/21/a-small-multiple-makeover-for-a-big-range-problem>

Line graph: dual axis



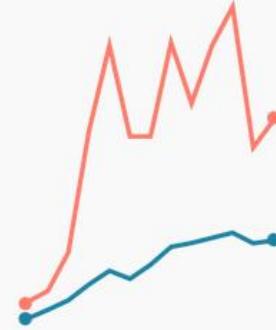
Line graph: dual axis



Orange steady,
Blue massively increasing.



Blue steady,
Orange increasing.



Both started at the same
level, but Orange increased
far more than Blue.



Both started at the same
level, but Blue increased far
more than Orange.



Both started with the
same increase, then Blue
raced to the top.



Both steady.

A lire : <https://blog.datawrapper.de/dualaxis/>

Area graph

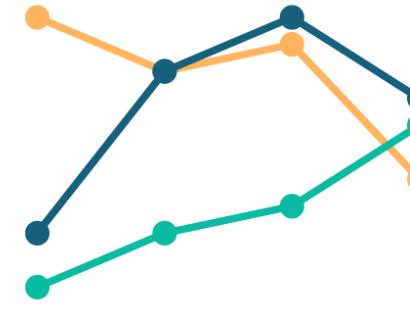
NOT IDEAL



NOT IDEAL



BETTER

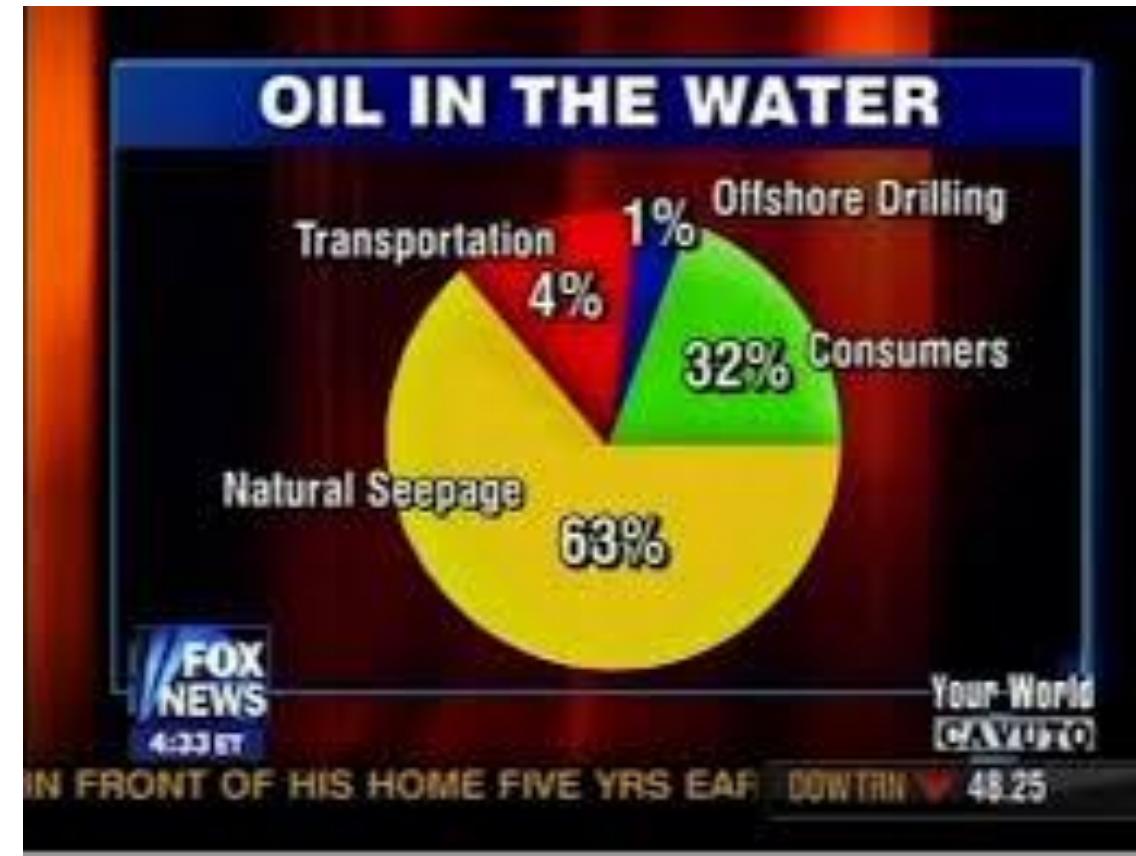
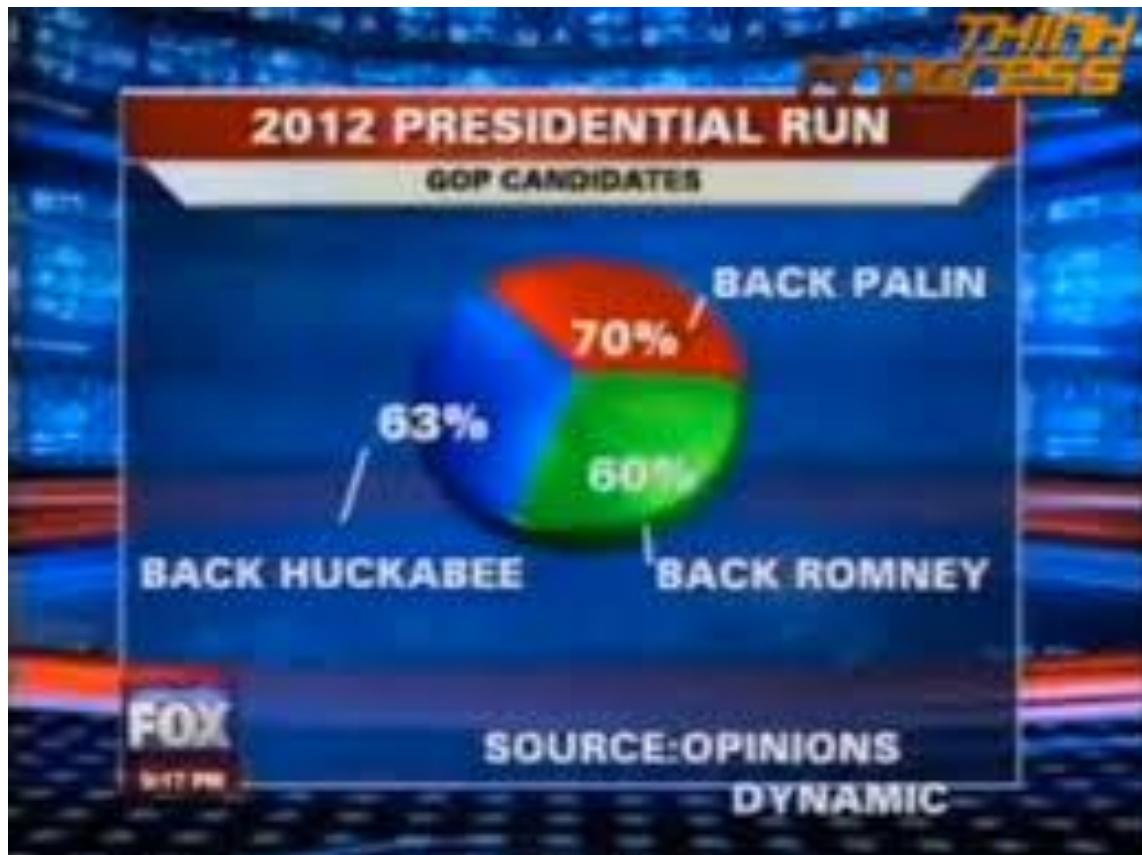


BETTER



<https://blog.datawrapper.de/area-charts/>

Pie graph



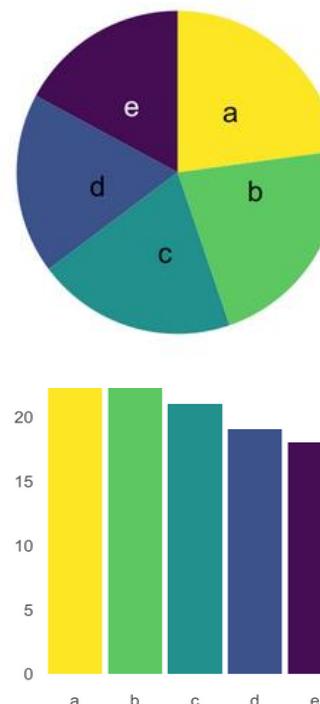
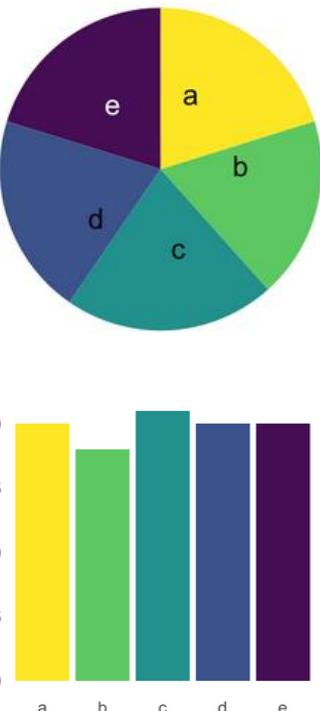
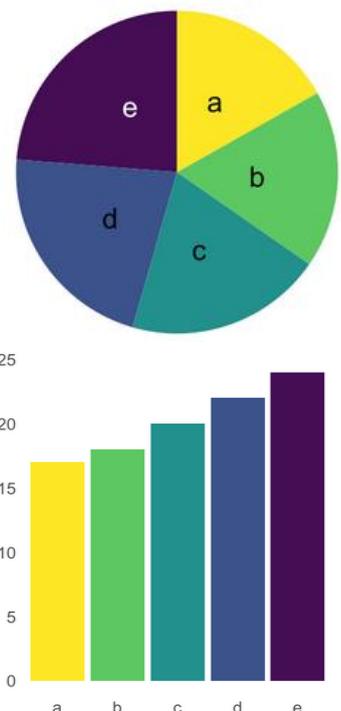
Pie graph

Efficace pour comparer rapidement des proportions
autours de 25%, 50%, ou 75%



Pie graph

C'est une autre histoire lorsqu'on souhaite comparer des proportions similaires ...

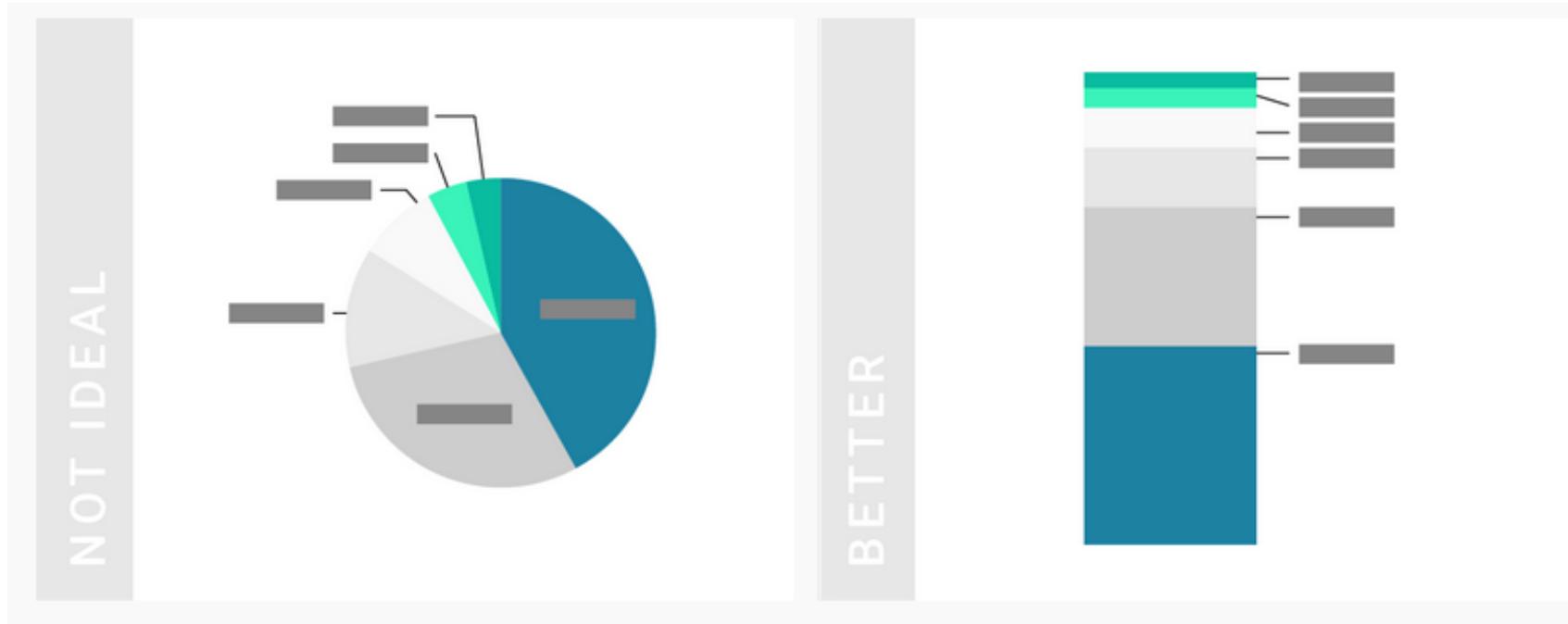


L'oeil humain n'est pas très efficace pour comparer des angles

Il l'est beaucoup plus pour comparer des hauteurs de barres

Pie graph

Il devient illisible si trop de catégories ...



Pie graph

Attention à la 3D ... toujours trompeuse !

Supplier Market Share



FIGURE 2.21 Pie chart

Supplier Market Share

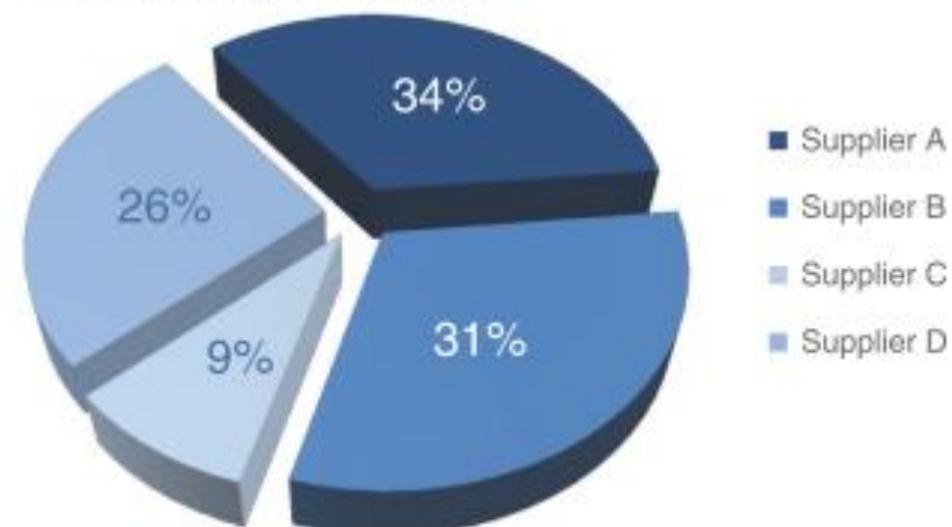


FIGURE 2.22 Pie chart with labeled segments

Pie graph

Références:

- <https://blog.datawrapper.de/pie-charts/>
- <https://www.data-to-viz.com/caveat/pie.html>
- <https://www.businessinsider.com.au/pie-charts-are-the-worst-2013-6?r=US&IR=T>

Boxplot

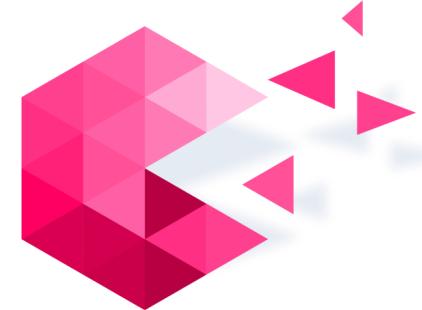
A lire :

<https://www.data-to-viz.com/caveat/boxplot.html>

Graph classiques

Une mine d'informations: <https://www.data-to-viz.com/>

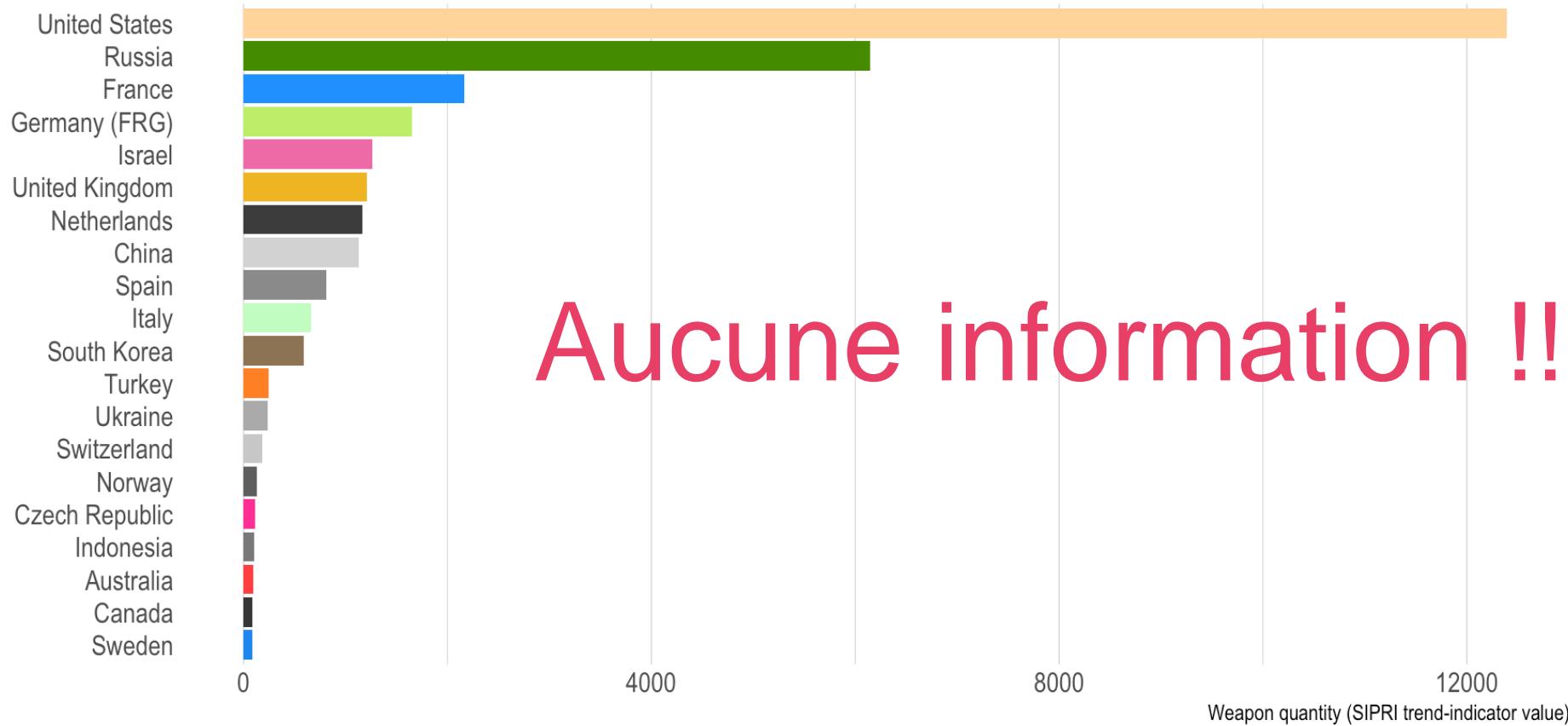




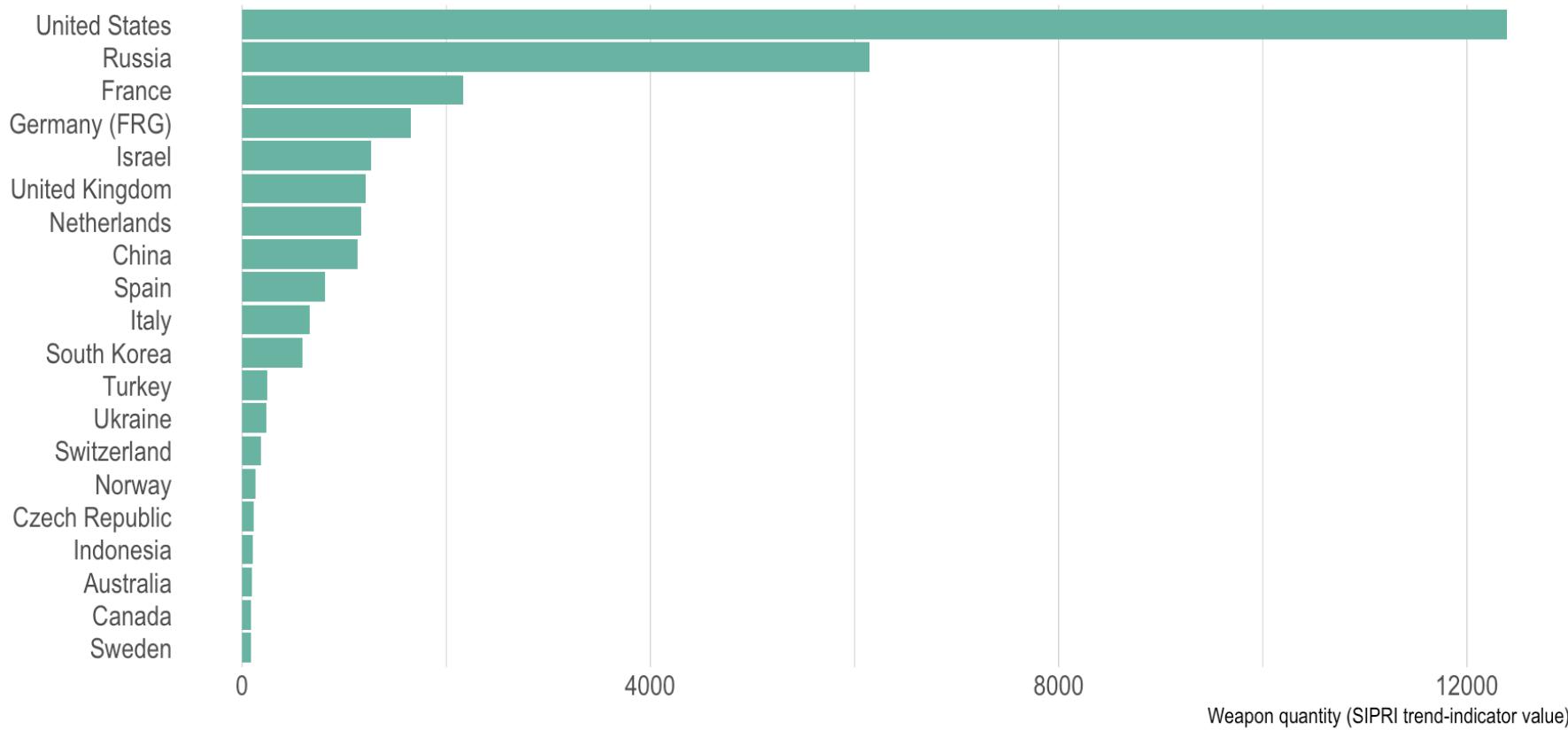
le
campus
numérique
in the ALPS

Usage des couleurs

Usage des couleurs



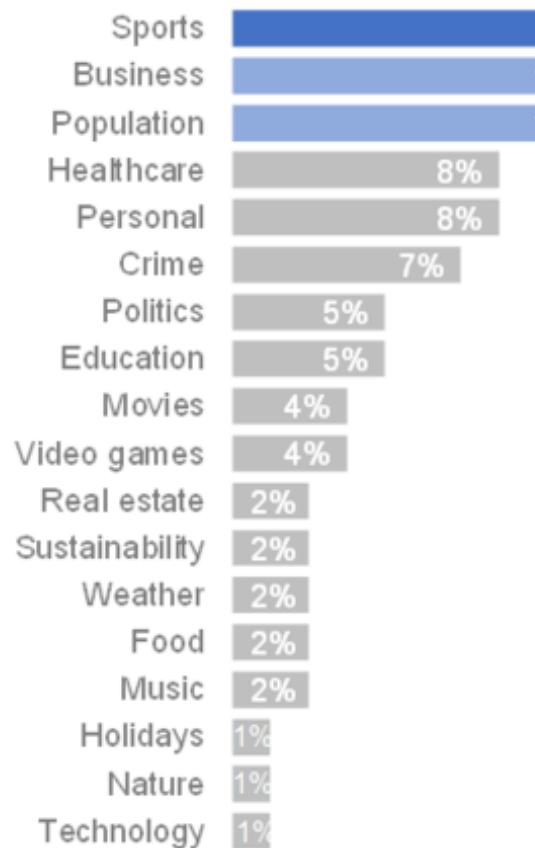
Usage des couleurs



Usage des couleurs

March #SWDchallenge readers are **most interested in sports**

% OF TOTAL

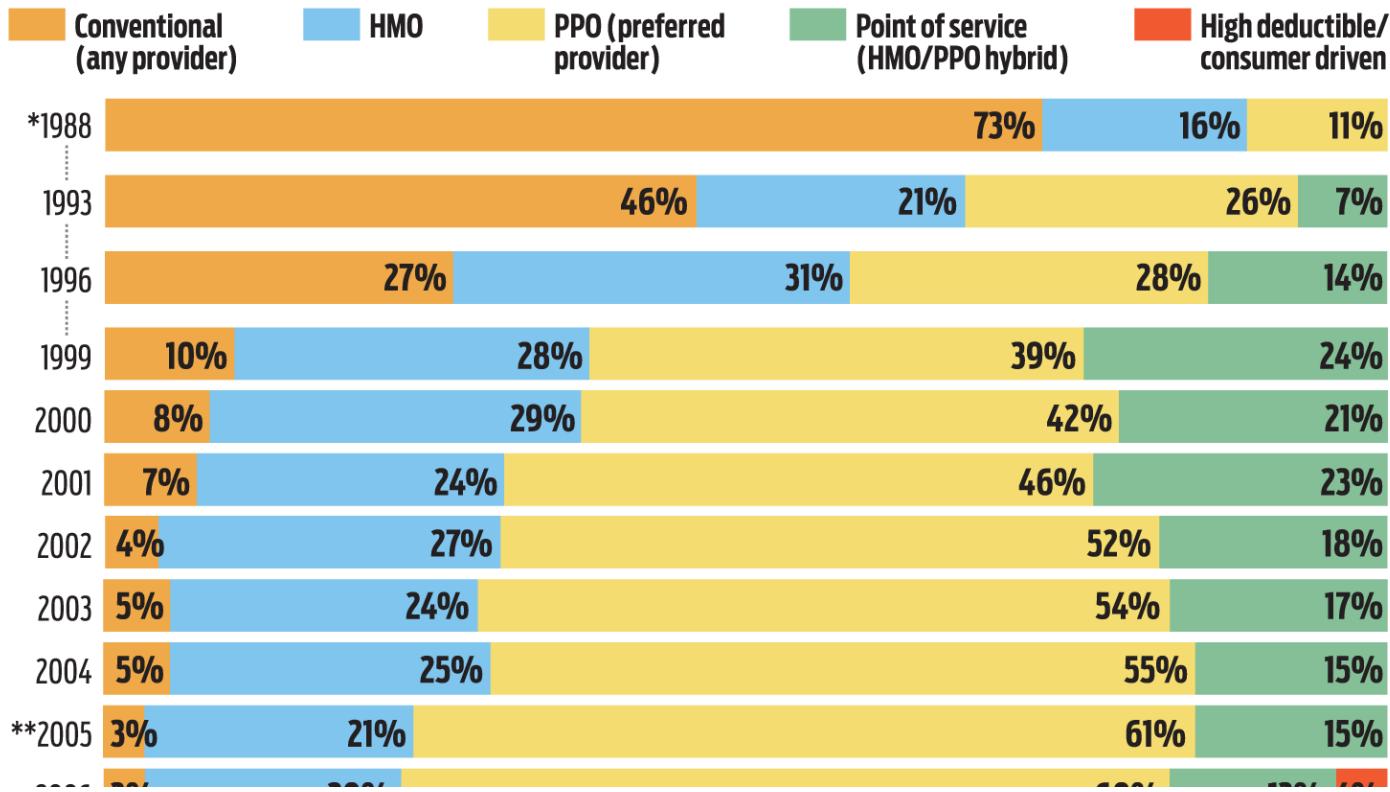


Nearly 50% of visualizations were about **sports**, **business** or **population trends**.

Usage des couleurs

Growth in high-deductible health plans

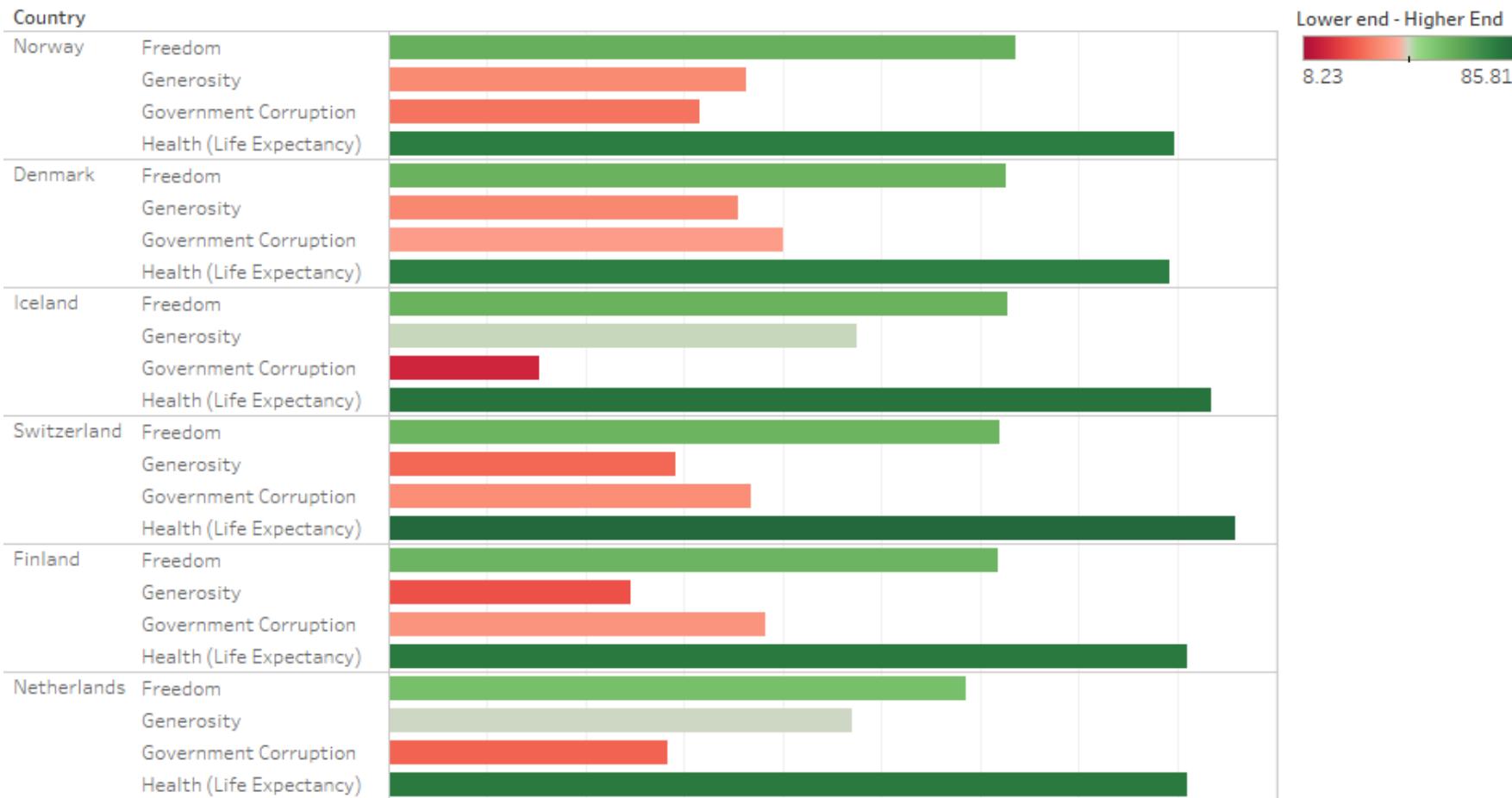
A 2015 Kaiser Family Foundation report found that the number of employers who offer only a consumer-driven health plan (often referred to as a high-deductible plan) has grown from 4 percent in 2006 to 24 percent in 2015, making it the second-most popular option. Health-care analysts project that percentage will increase in 2016. The graphic shows by year the percentage of each plan type offered by companies that provide only one plan to employees.



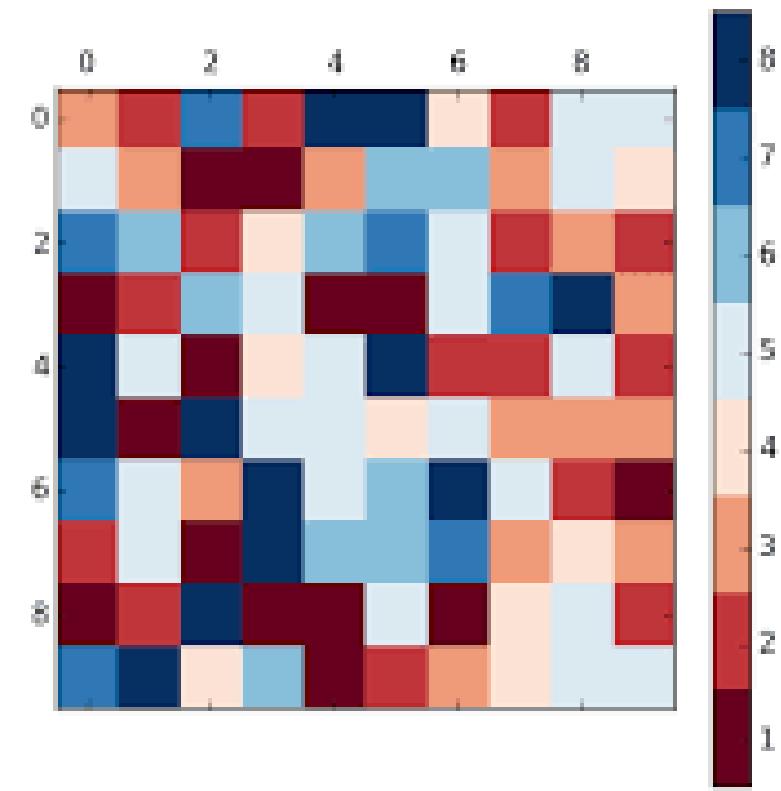
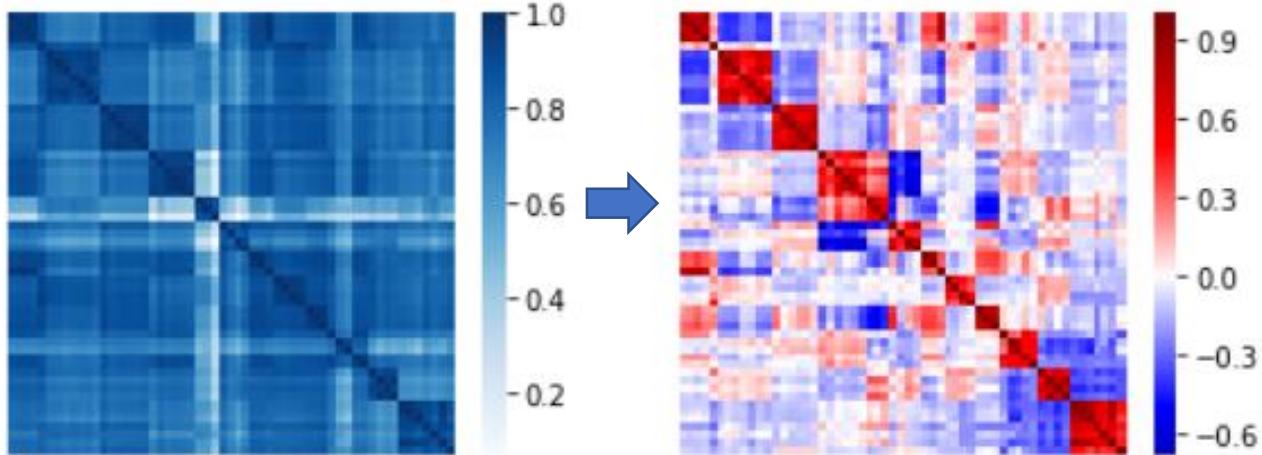
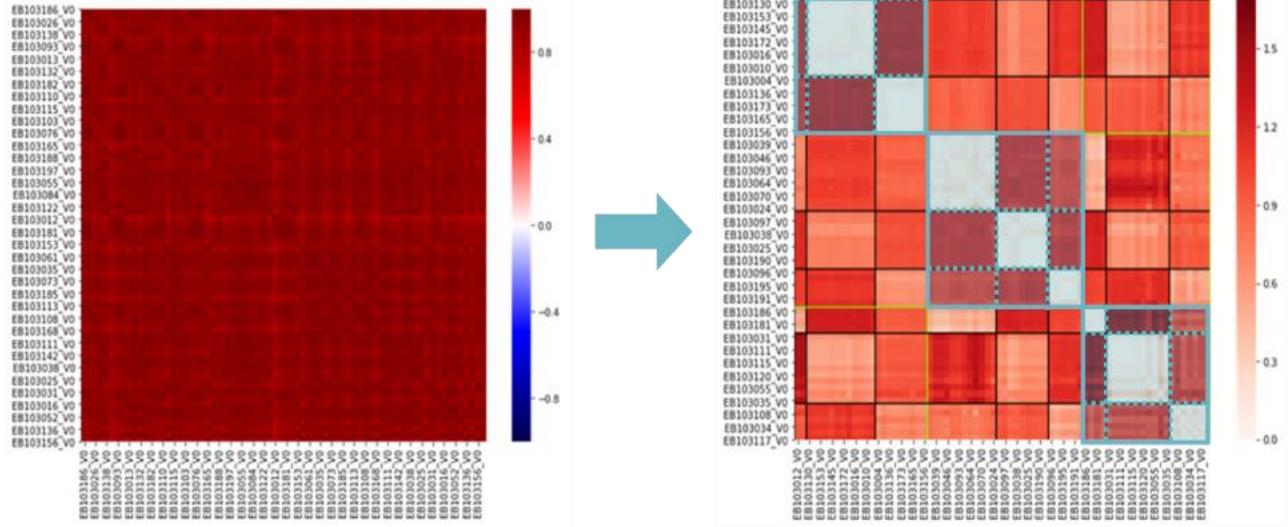
Usage des couleurs

What makes a Country **happy**?

Below are the top 20 happy countries. There are different parameters that contribute to the happiness of a country. *Red* indicates lower and *Green* indicates higher for different categories that weigh a country's happiness.



Usage des couleurs



Usage des couleurs

Lorsqu'on ajoute des couleurs à un graph, toujours réfléchir à ce qu'elles apportent !!!

Les couleurs sont utiles pour:

- Montrer des catégories
- Mettre en avant un élément
- Montrer un gradient



Usage des couleurs

Références:

https://www.data-to-viz.com/caveat/color_com_nothing.html

<http://www.storytellingwithdata.com/blog/2018/3/9/bring-on-the-bar-charts>

A lire:

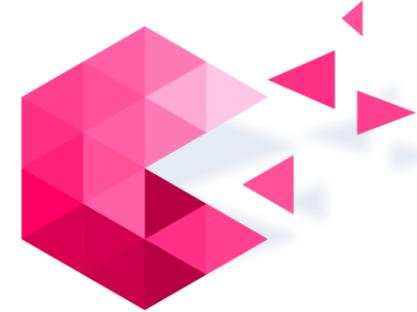
<https://blog.datawrapper.de/colors/>

Pour aller plus loin:

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

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

<https://www.storytellingwithdata.com/blog/2020/5/6/picking-the-right-colors>



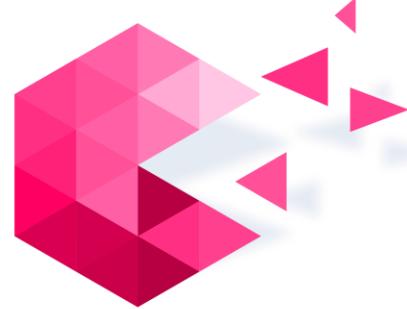
le
campus
numérique
in the ALPS

Création d'un graph

Un exemple

A lire:

<https://blog.datawrapper.de/better-charts/>



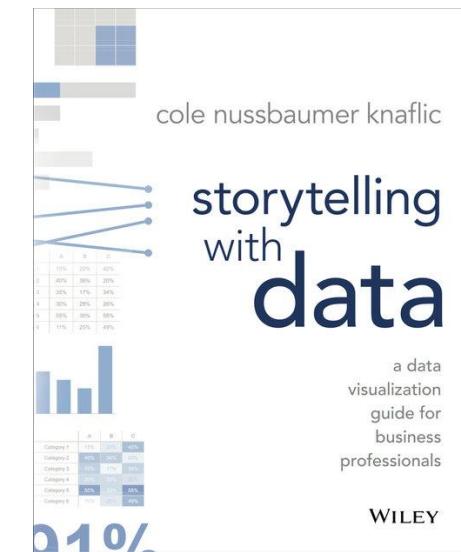
le
campus
numérique
in the ALPS

Transformation de graphs

Cas d'études « corrigés »

Référence:

[Storytelling with data: a data visualization guide for professionals](#), Cole Nussbaumer Knafic



Mettre en avant:

La part des segments 3, 4 et 5 dans la population et parmi nos clients.

Distribution by customer segment

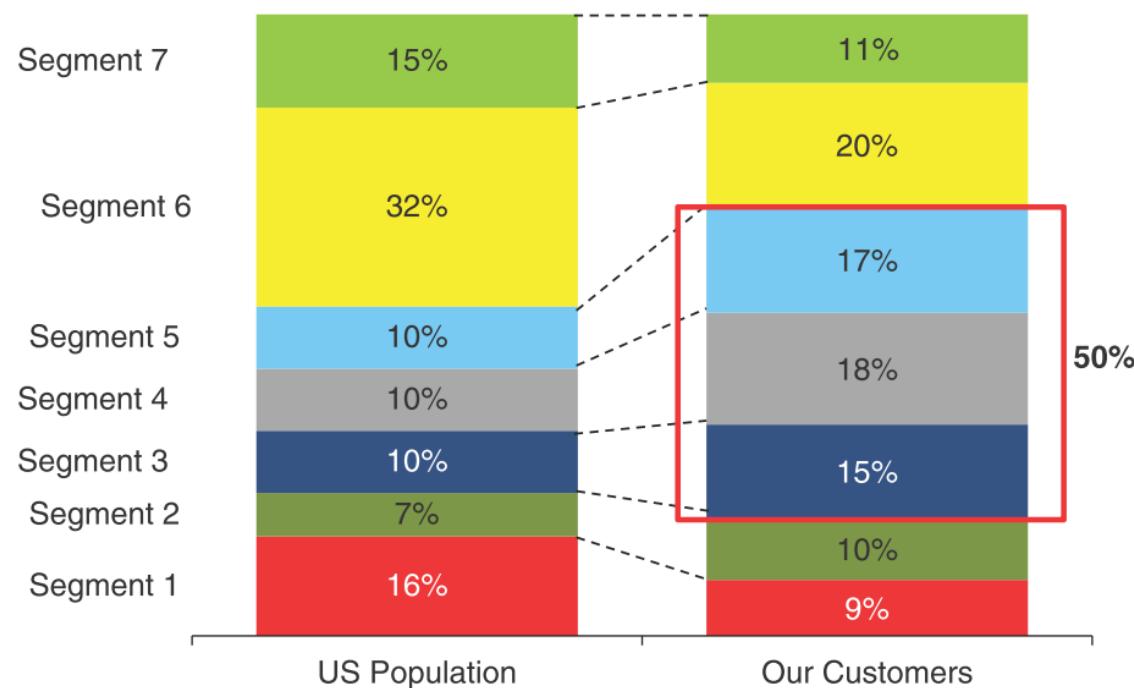


FIGURE 5.12 Unaesthetic design

Distribution by customer segment

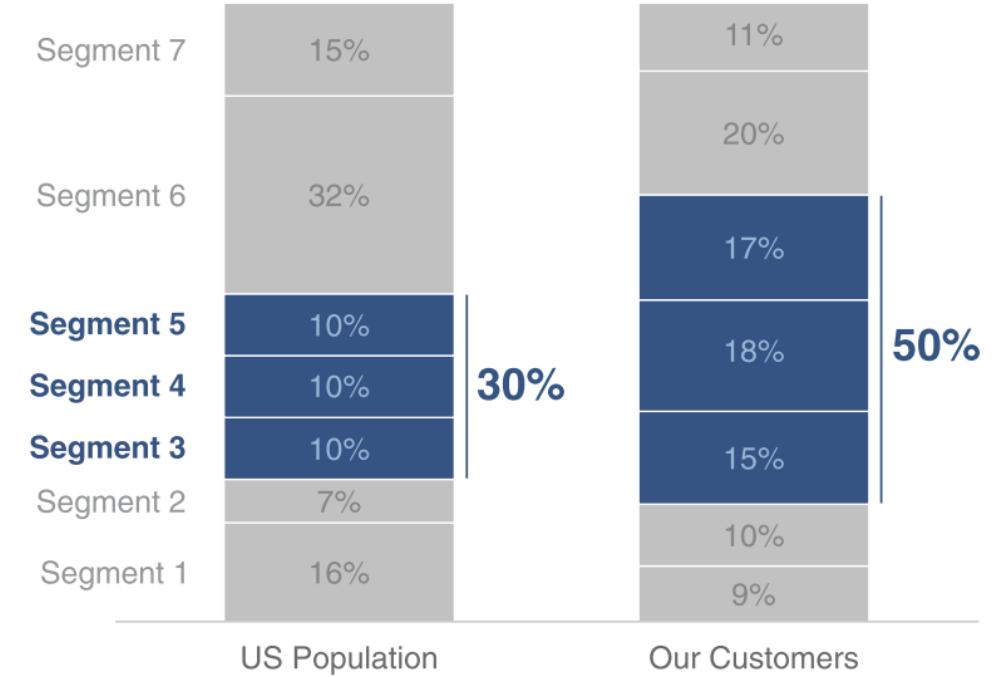


FIGURE 5.13 Aesthetic design

Mettre en avant:

- Les préoccupations de conception supérieures à 10 pour 1000
- Particulièrement celles qui concernent le bruit

Top 10 design concerns

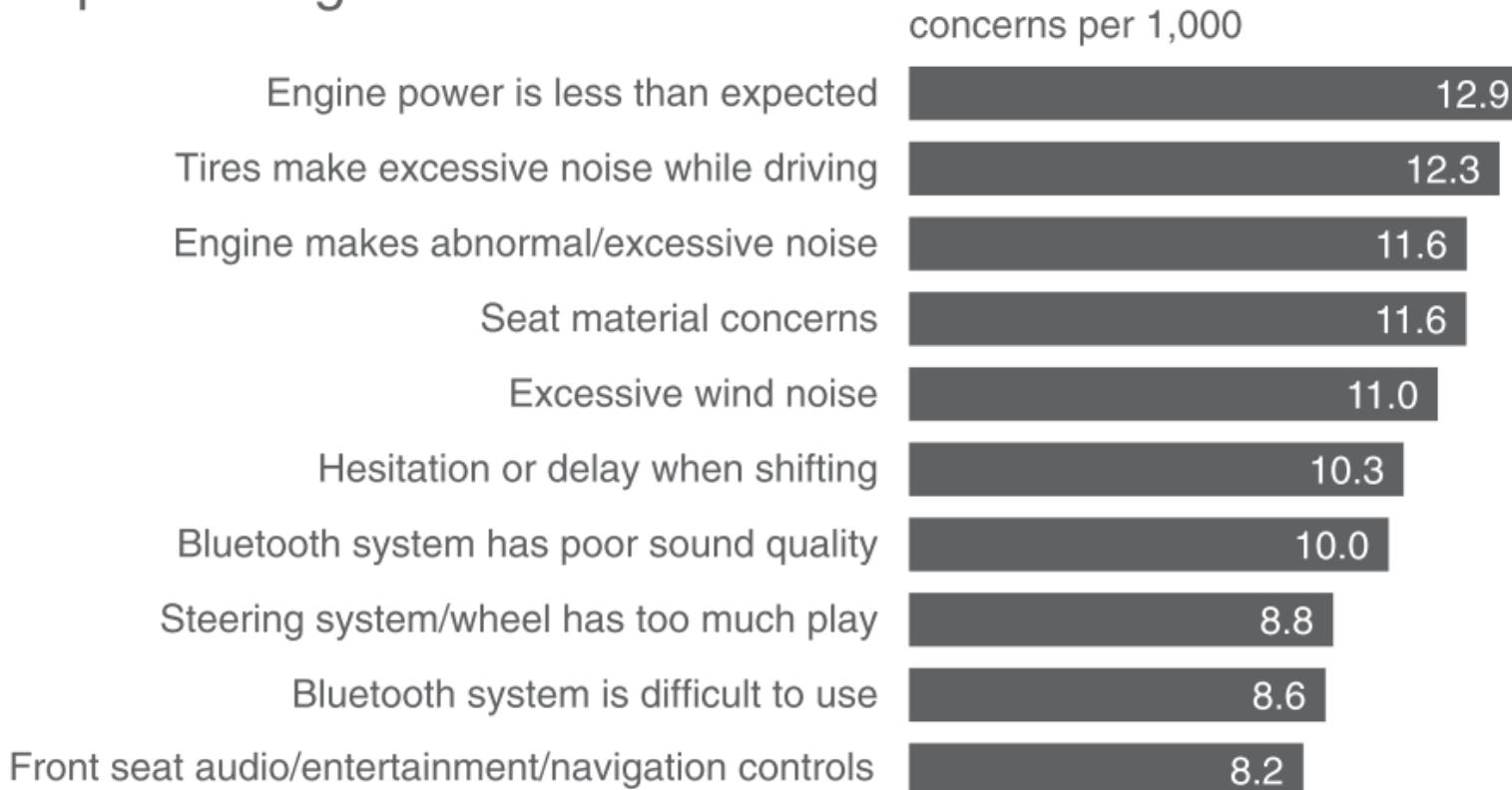


FIGURE 4.7 Original graph, no preattentive attributes

Top 10 design concerns

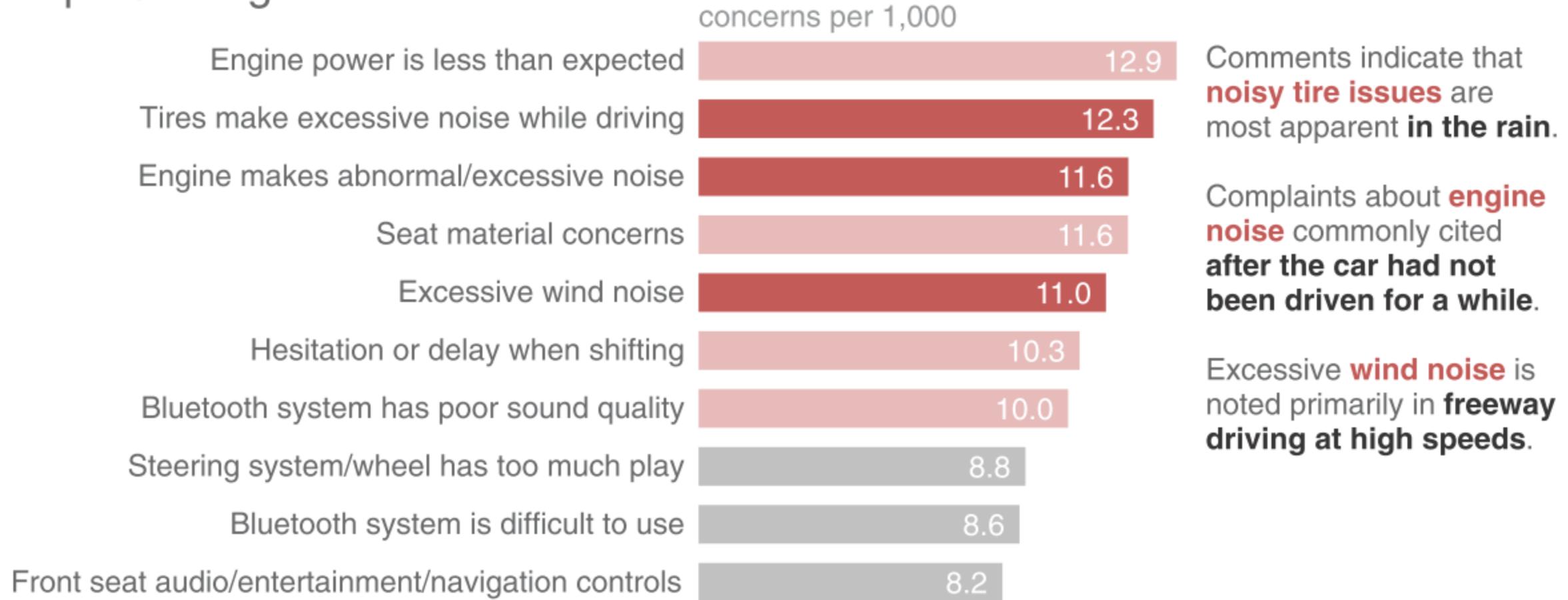
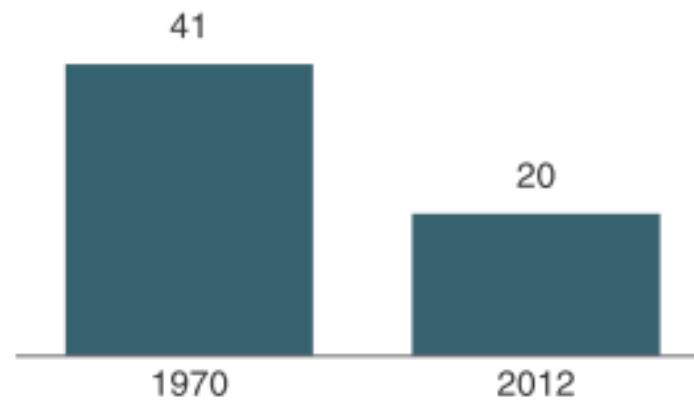


FIGURE 4.9 Create a visual hierarchy of information

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

Adapted from PEW RESEARCH CENTER

FIGURE 2.2 Stay-at-home moms original graph

Simplifier

20%

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

FIGURE 2.3 Stay-at-home moms simple text makeover

Objectif du graph:

Comparer la performance de « notre business » par rapport à d'autres concurrents dans plusieurs catégories (selection, convenience, service, relationship, price)



FIGURE 3.15 Original graph

Performance overview

■ Our business

- Competitor A
- Competitor B
- Competitor C
- Competitor D
- Competitor E

Weighted performance index | relative rank

Price



Convenience



Relationship



Service



Selection



FIGURE 3.16 Revamped graph, using contrast strategically

Objectif du graph:

Montrer l'impact sur la qualité du travail (ticket reçus – tickets traités) de la perte de 2 employés (en mai)

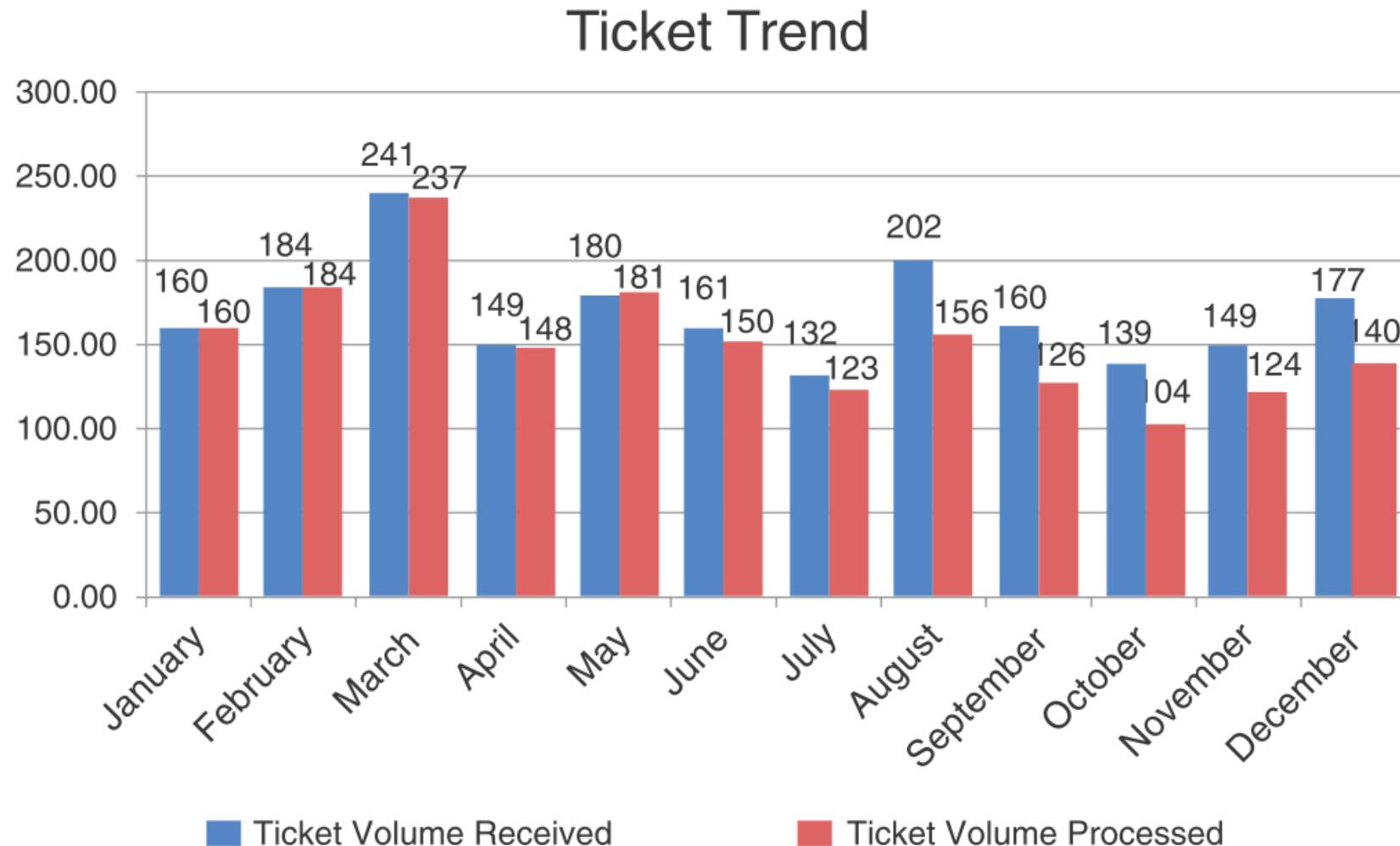


FIGURE 0.2 Example 1 (before): showing data

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time

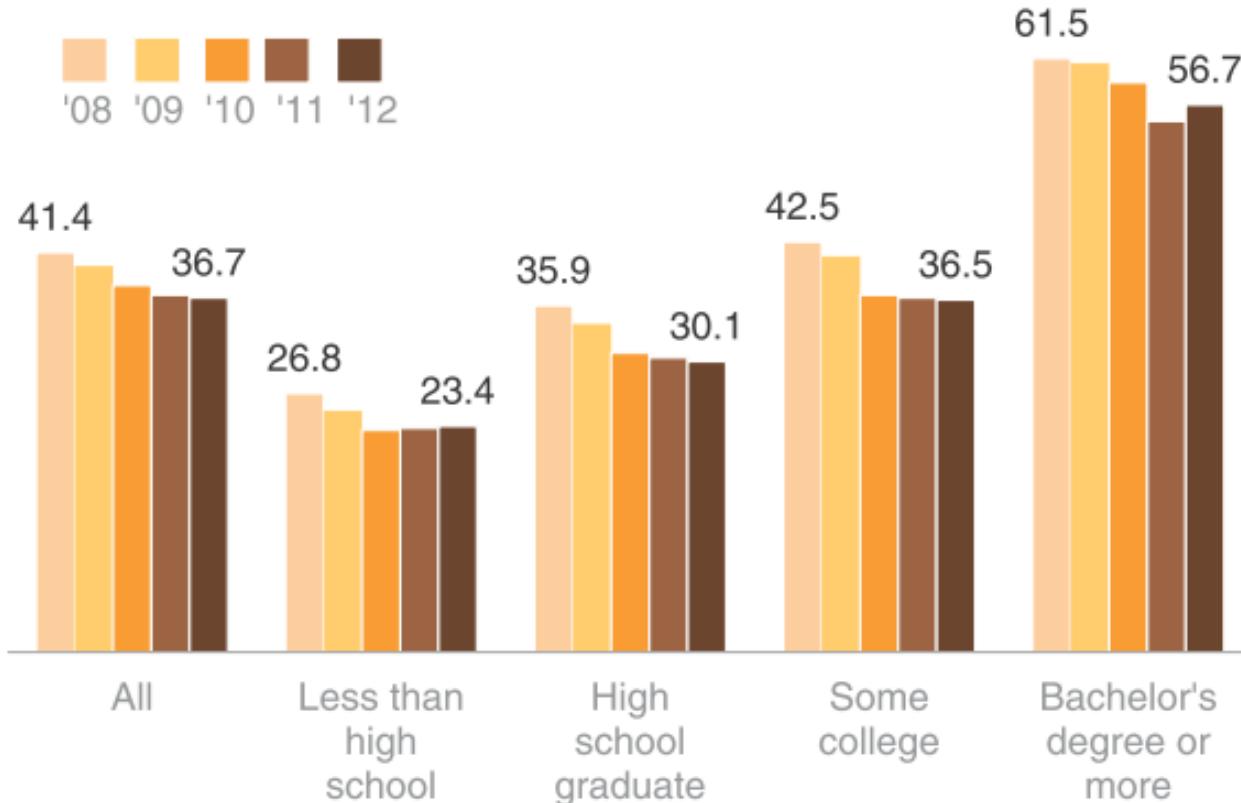


Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

FIGURE 5.10 Add action title and annotation

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

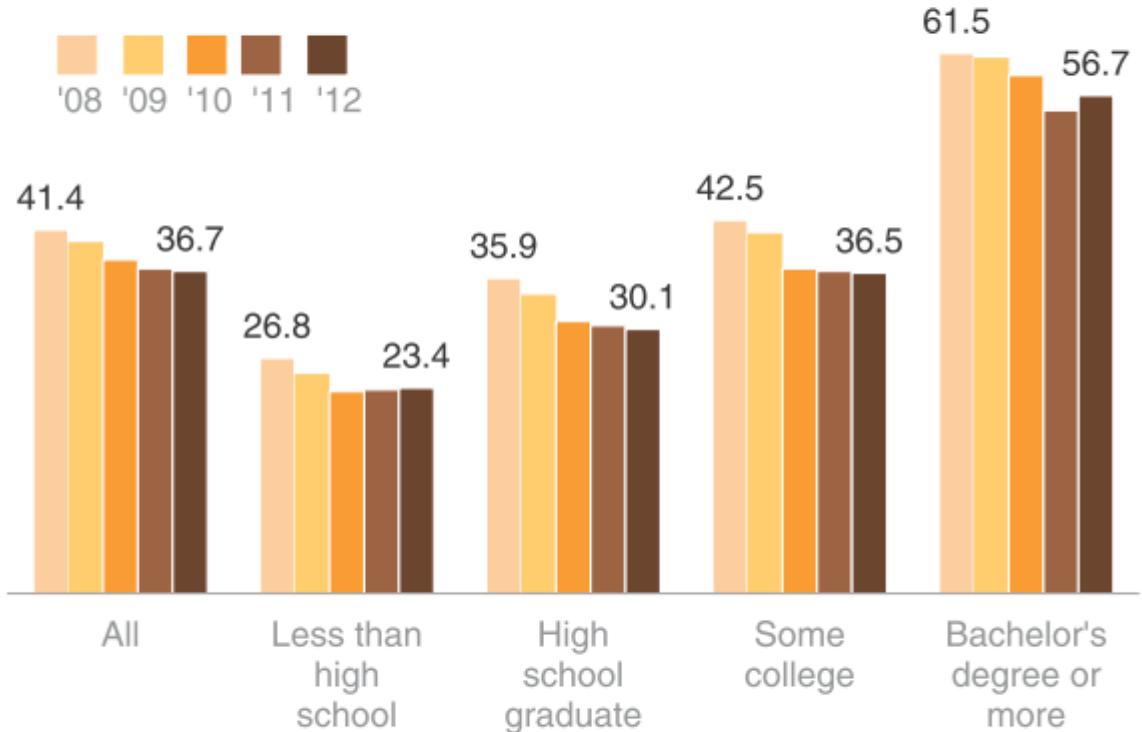
Adapted from PEW RESEARCH CENTER

Objectif du graph:

- Visualiser l'évolution des mariages selon le niveau d'éducation
- Mettre en avant l'évolution du groupe ayant au moins un Bachelor
- Tenter d'éliminer au maximum les distractions pour attirer l'attention au bon endroit

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



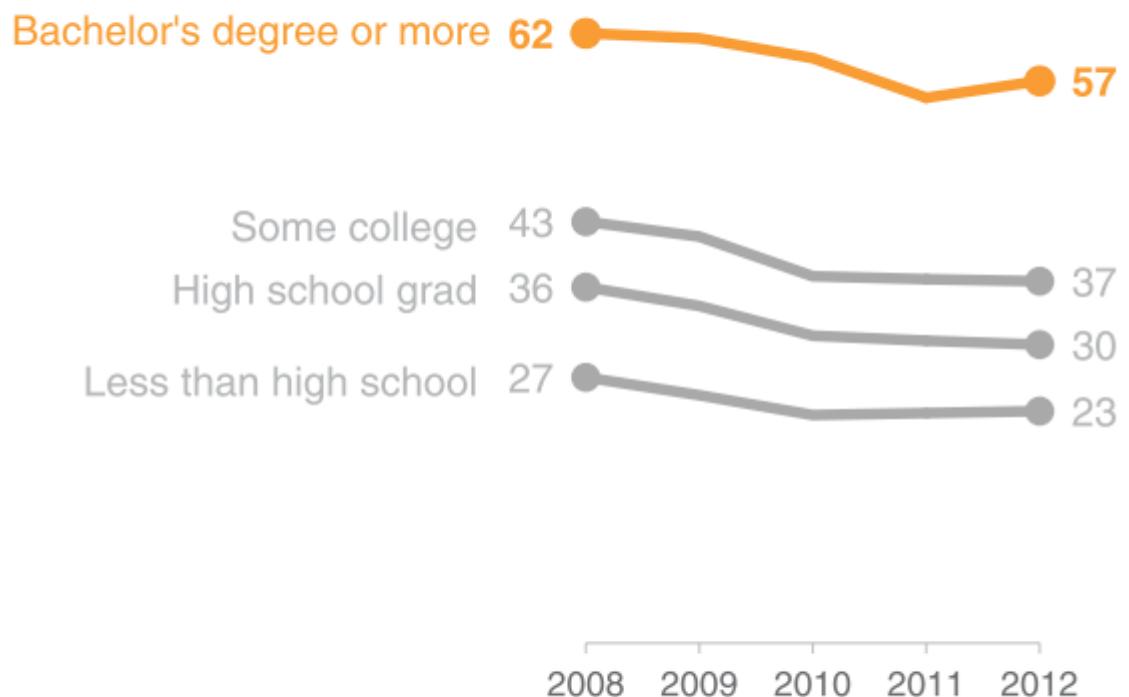
Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Adapted from PEW RESEARCH CENTER

New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Adapted from PEW RESEARCH CENTER

Price has declined for all products on the market since the launch of Product C in 2010

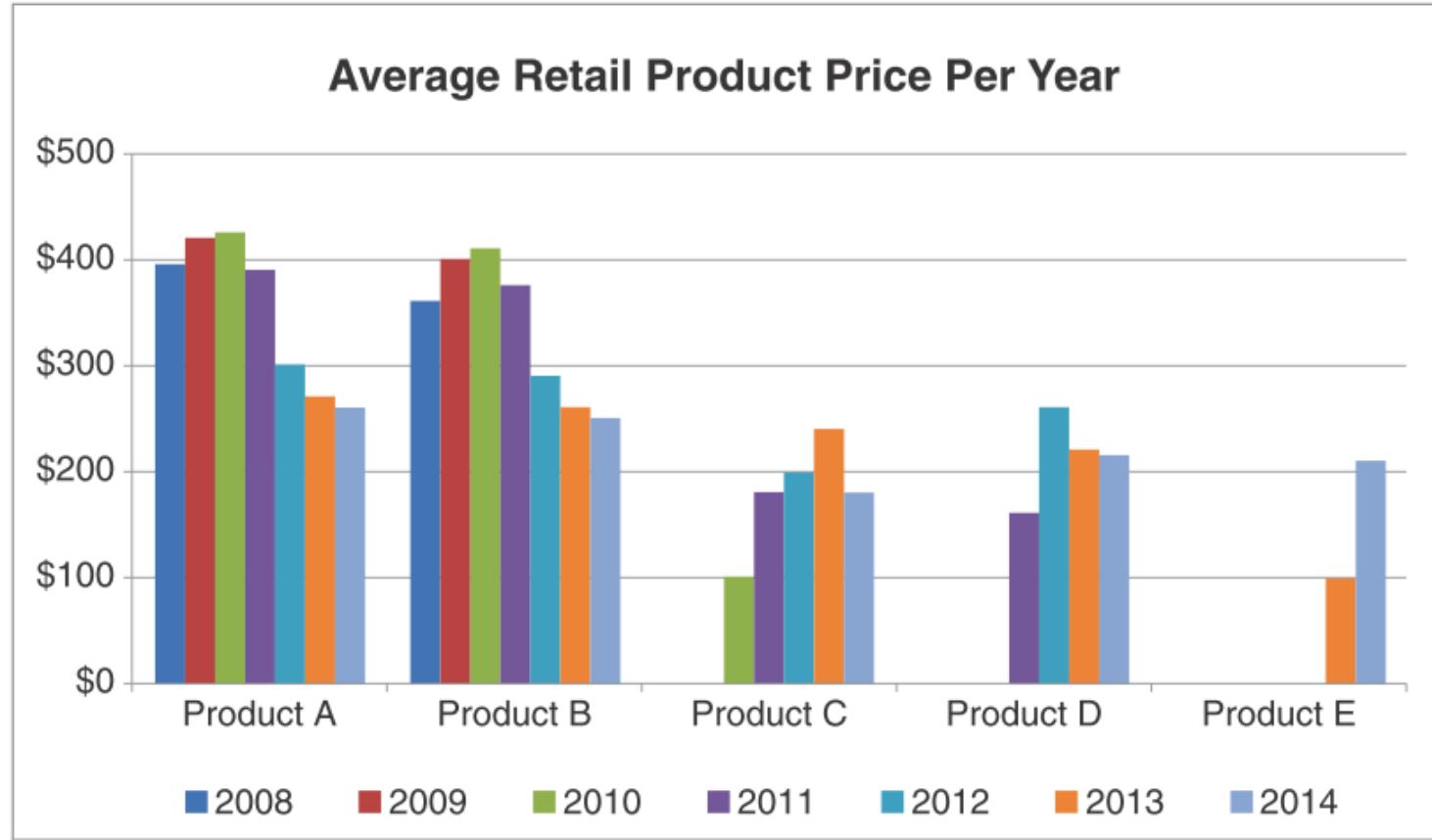


FIGURE 8.1 Original visual

Créer une série de graphs qui permettent d'aboutir à un positionnement de prix:

- Price of existing products declined after the launch of product C
- Structure of price evolution for all products
- As of 2014, retail prices have converged across products
- Price recommandation for a new product

After the launch of Product C in 2010, **the average retail price of existing products declined.**

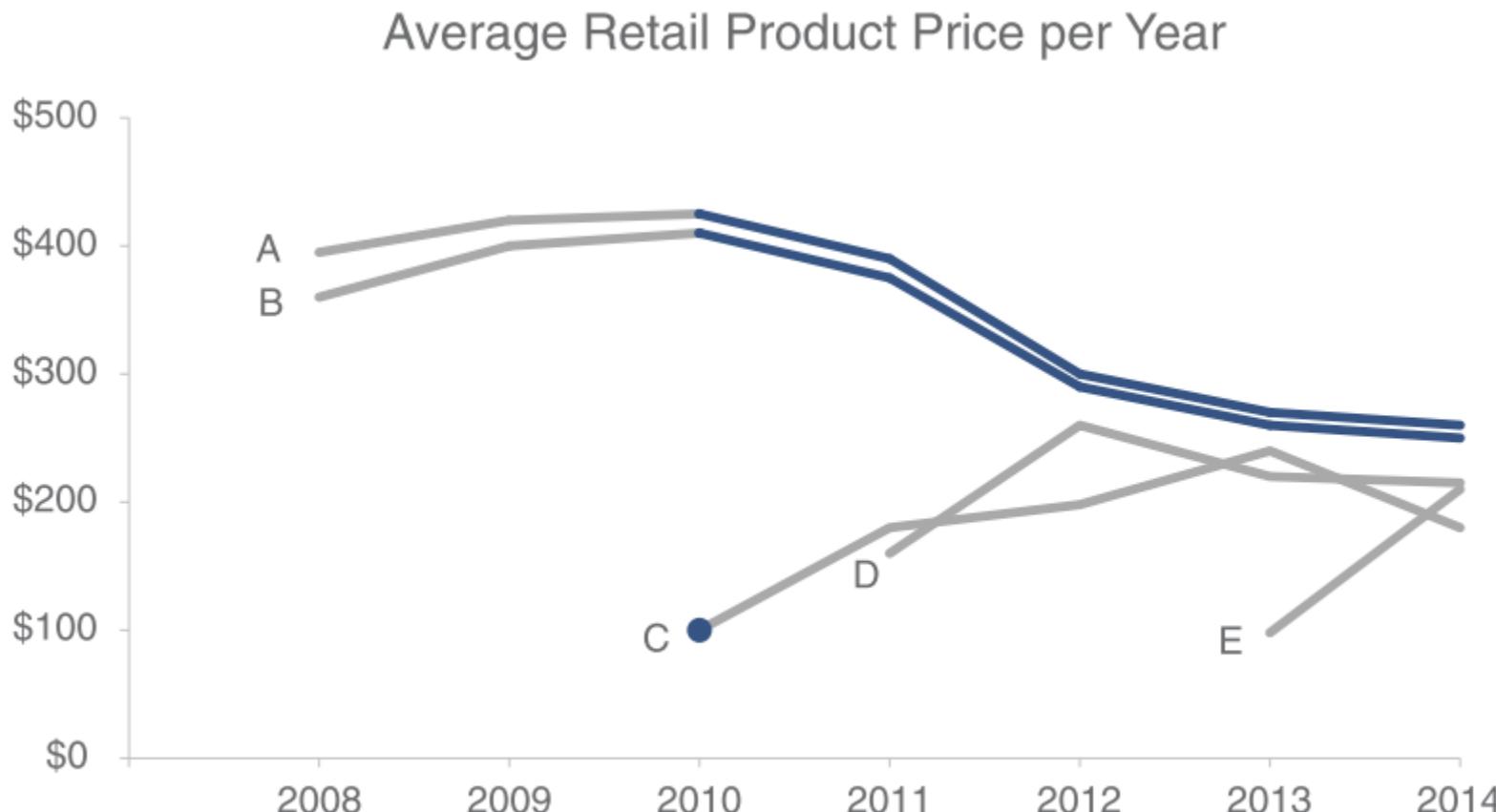


FIGURE 8.7 Focus the audience's attention

...but all have **increased in price** since their respective launches

Retail price over time

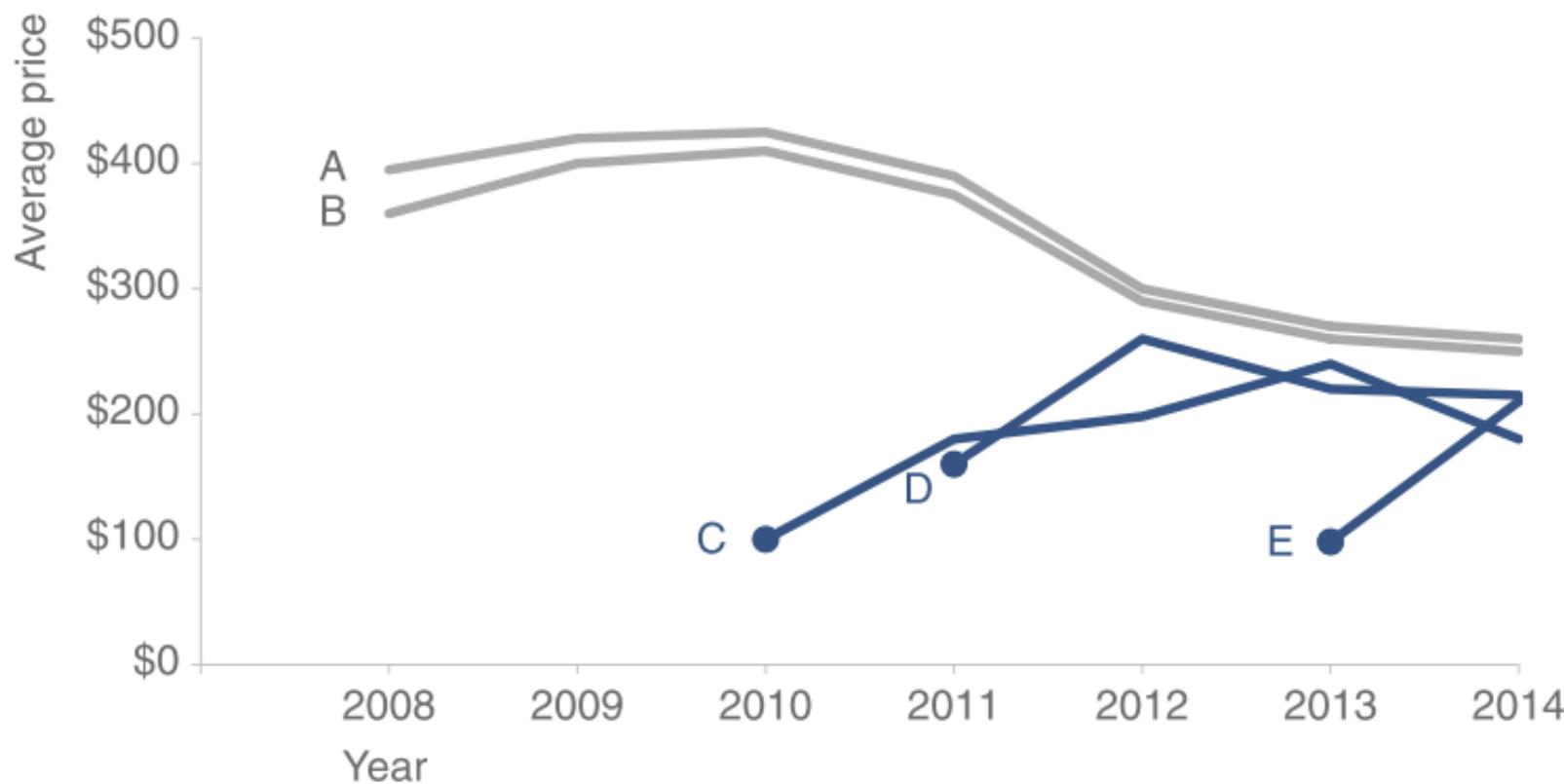


FIGURE 8.16

In fact, with the launch of a new product in this space, we tend to see an **initial price increase**, followed by a **decrease** over time

Retail price over time

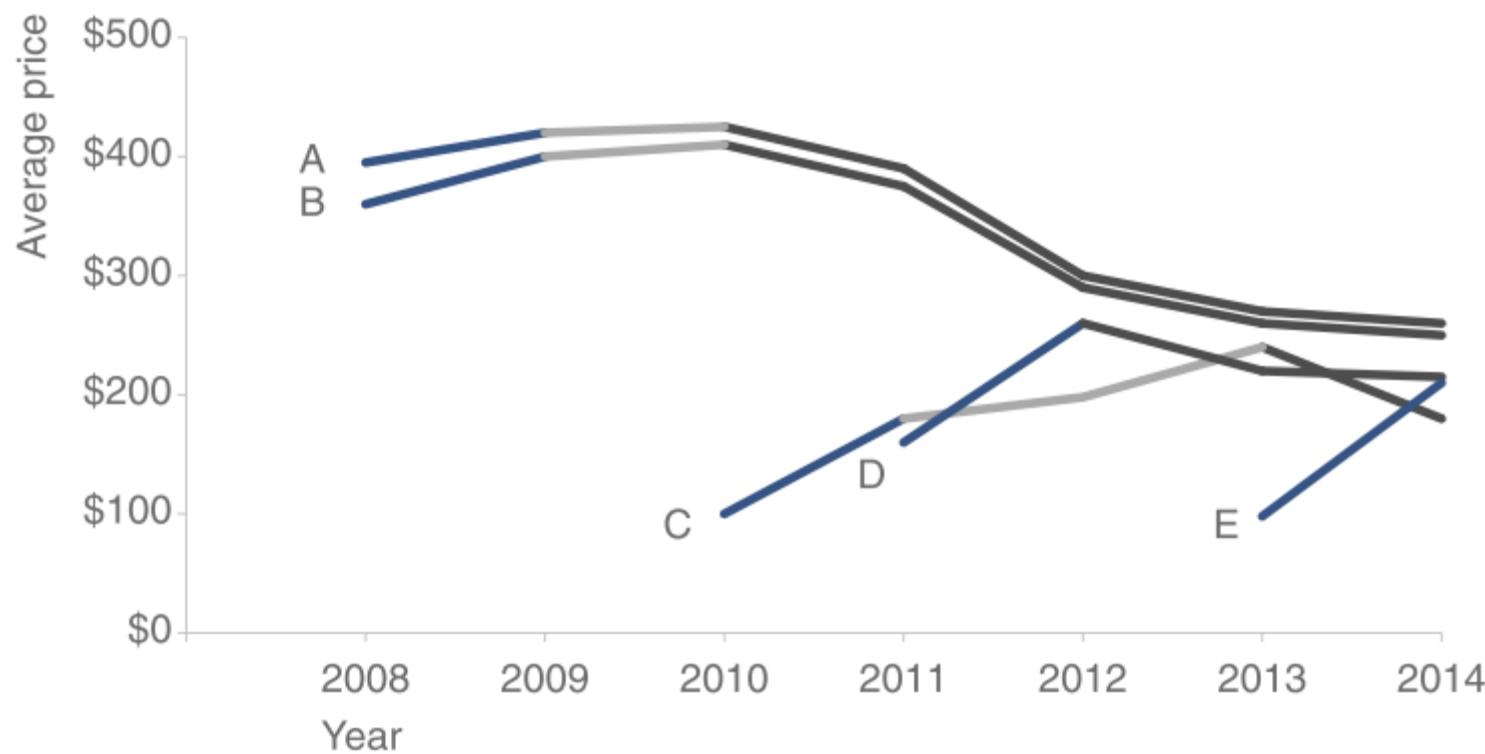


FIGURE 8.17

As of 2014, retail prices have converged, with an **average retail price of \$223**, ranging from a low of \$180 (C) to a high of \$260 (A)

Retail price over time

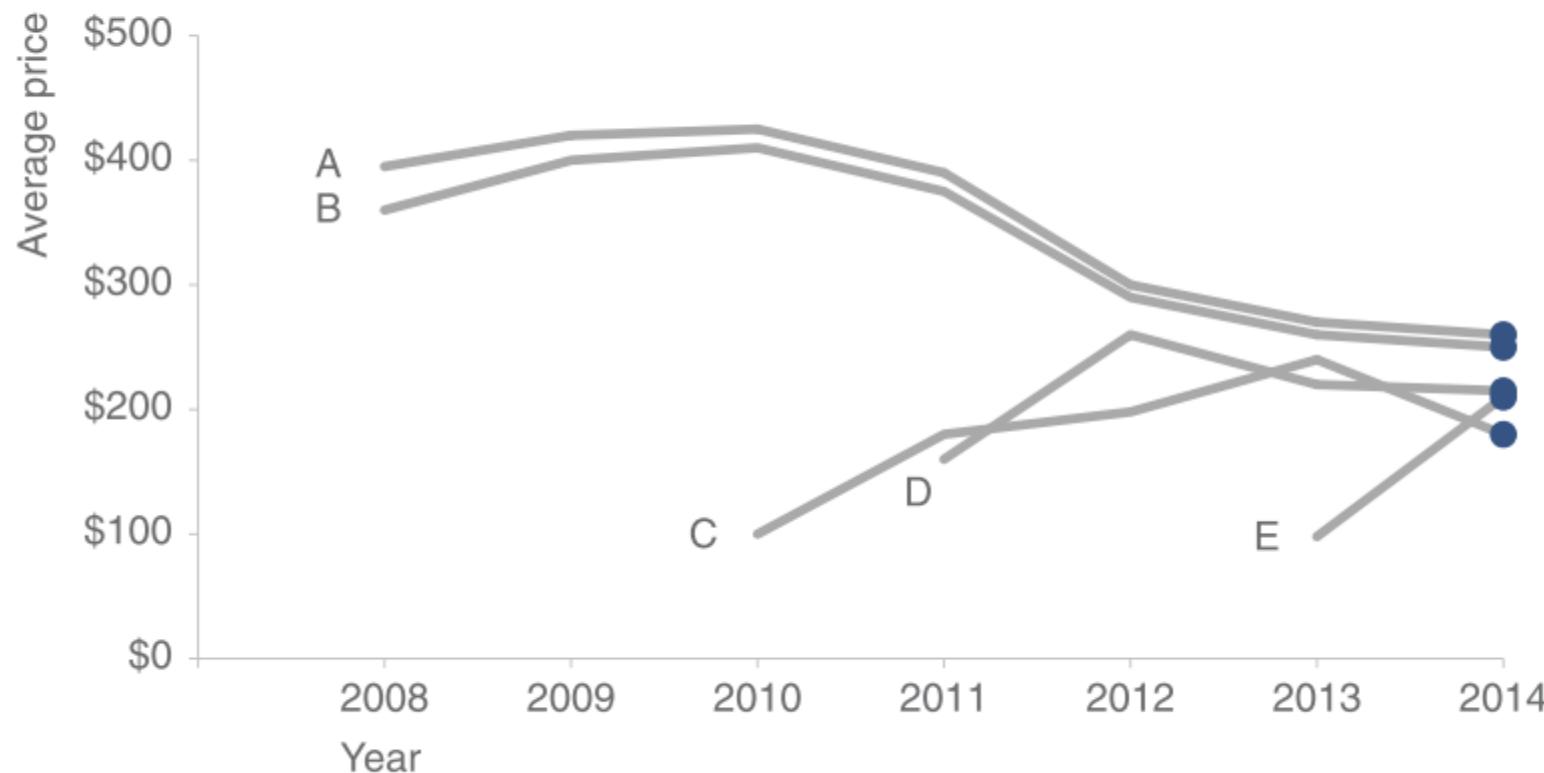


FIGURE 8.18

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

Retail price over time

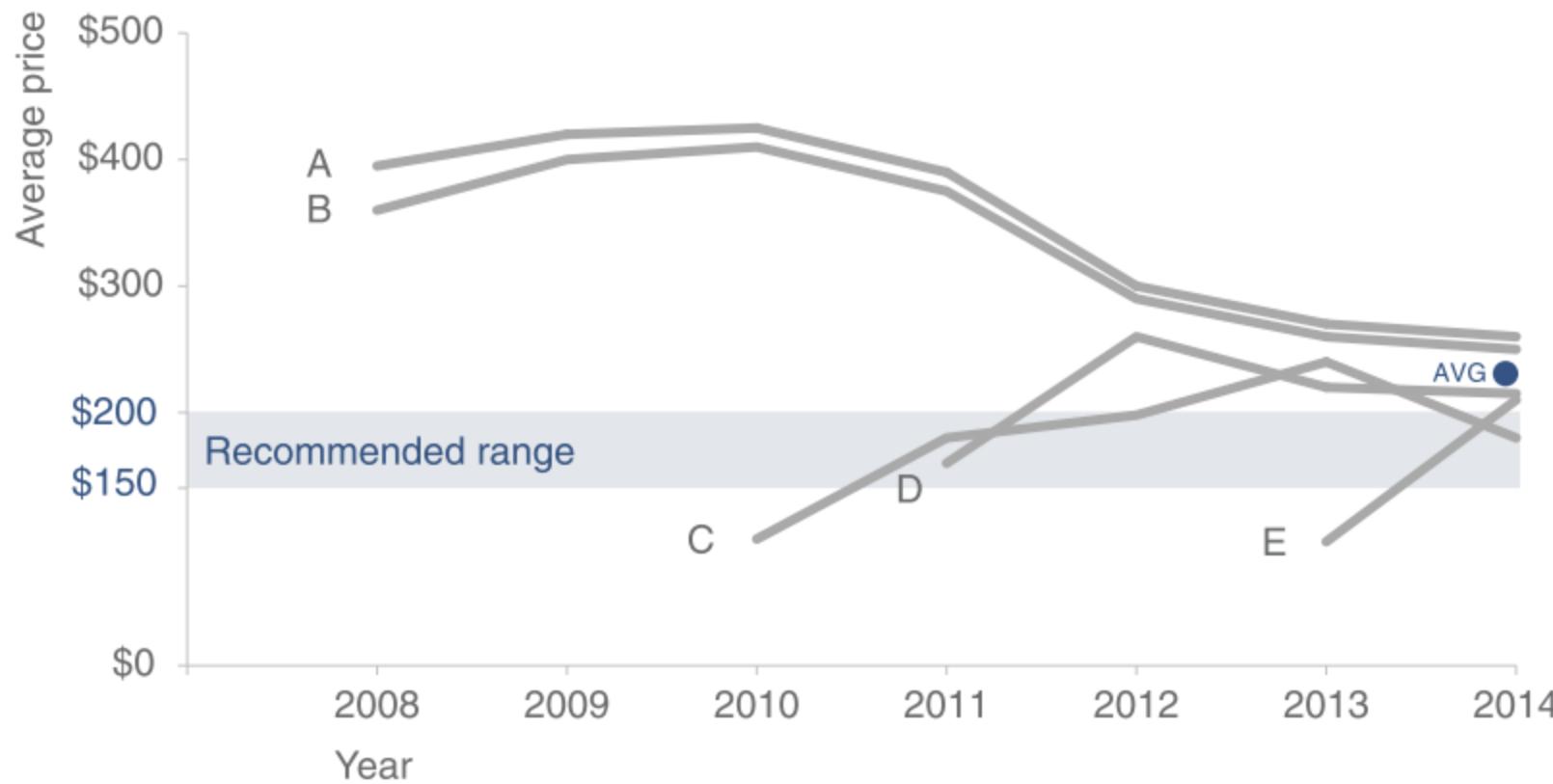


FIGURE 8.19

How satisfied have you been with each of these features?

■ Have not used ■ Not satisfied at all ■ Not very satisfied ■ Somewhat satisfied ■ Very satisfied ■ Completely satisfied

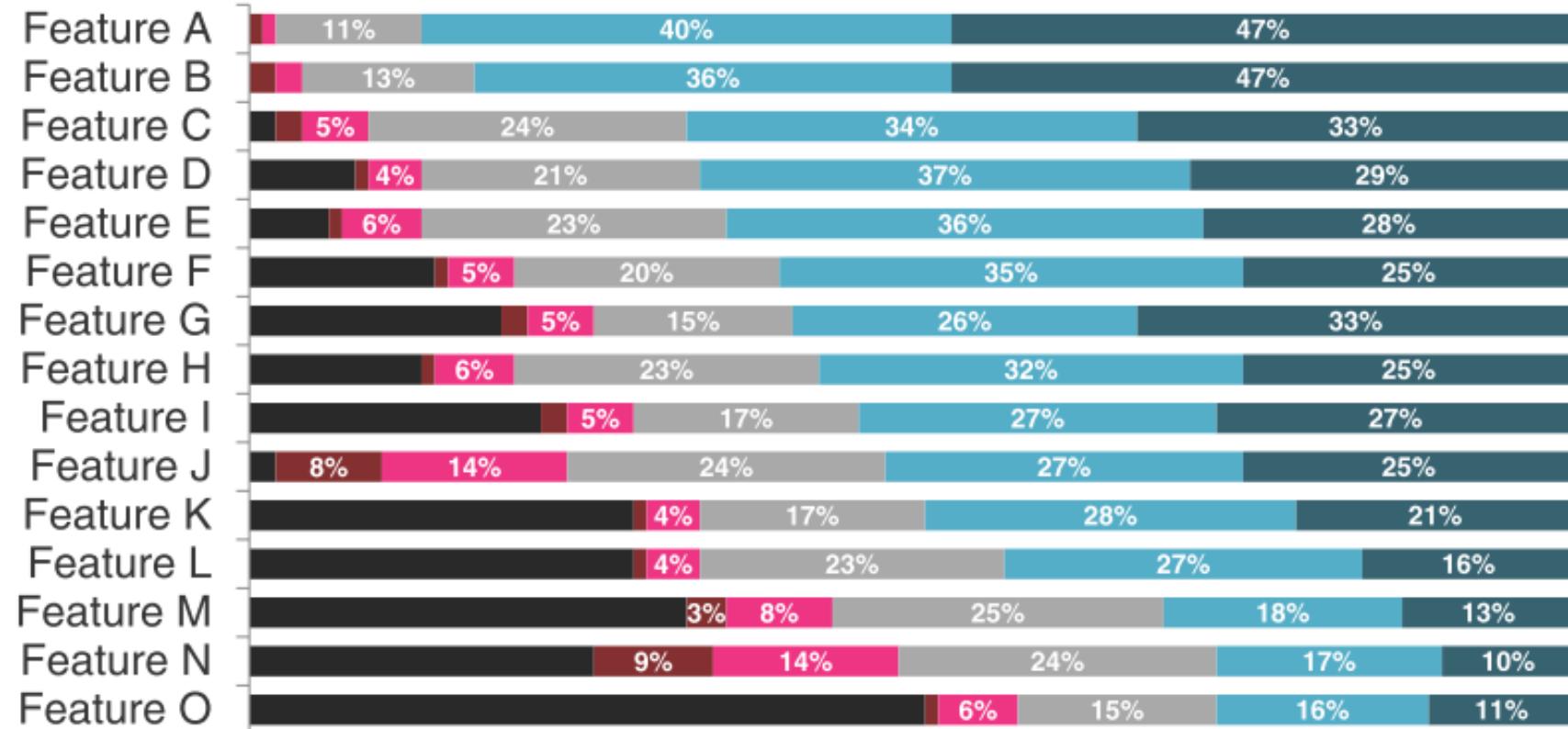


FIGURE 9.12 User satisfaction, original graph

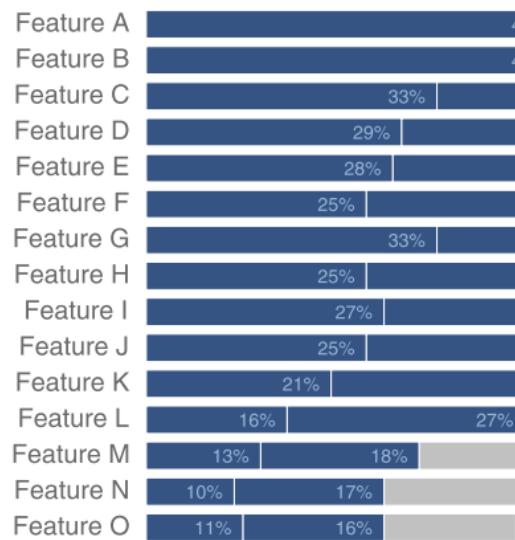
Créer des graphs qui illustrent au mieux:

- Quelles features sont appréciées des utilisateurs ?
- Quelles features ne sont pas appréciées des utilisateurs ?
- Quelles features sont utilisées par les utilisateurs ?
- Les 3 à la fois ?

Features A & B top user satisfaction

Product X User Satisfaction: Features

■ Completely satisfied ■ Very satisfied ■ Somewhat satisfied ■ Not very satisfied ■ Not satisfied at all ■ Have not used



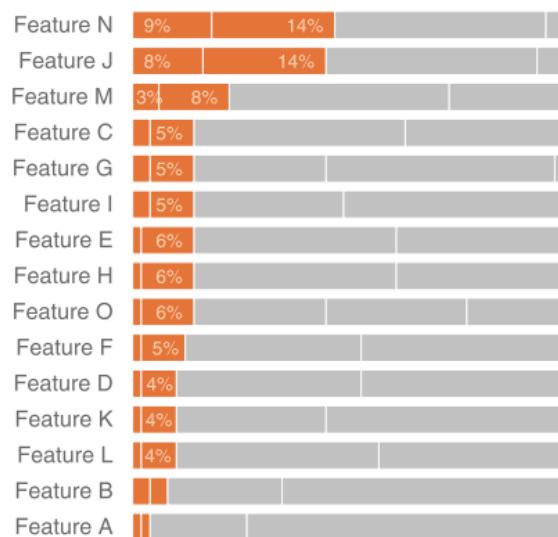
Responses based on survey question "How satisfied have you been with each of these features?".
Need more details here to help put this data into context: How many people completed survey? What proportion of users does this represent? Do those who completed survey look like the overall population, demographic-wise?

FIGURE 9.13 Highlight the positive side of the data

Users least satisfied with Features N & J

Product X User Satisfaction: Features

■ Not satisfied at all ■ Not very satisfied ■ Somewhat satisfied ■ Very satisfied ■ Completely satisfied



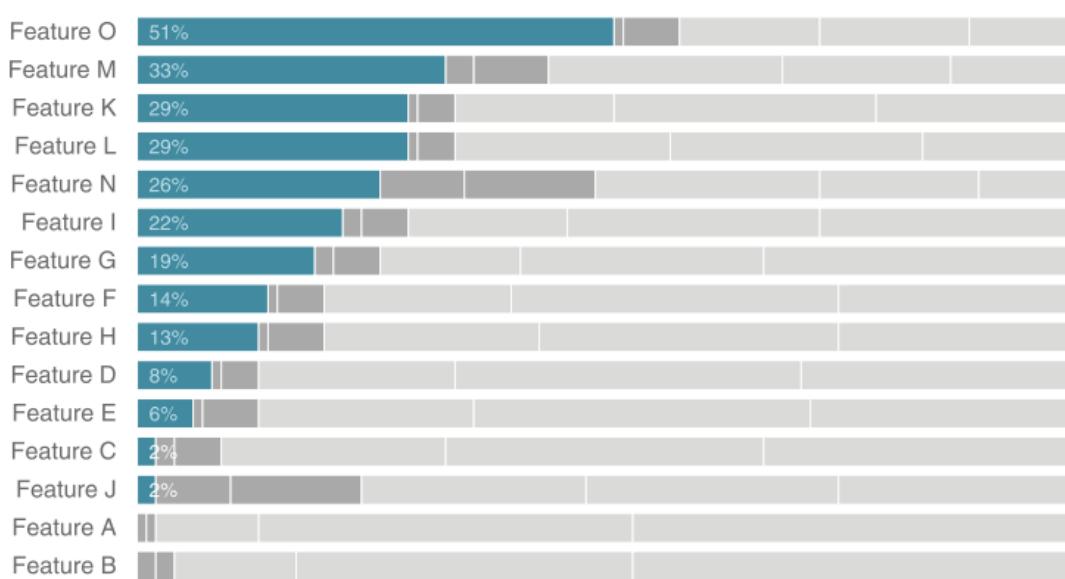
Responses based on survey question "How satisfied have you been with each of these features?".
Need more details here to help put this data into context: How many people completed survey? What proportion of users does this represent? Do those who completed survey look like the overall population, demographic-wise?

FIGURE 9.14 Highlight dissatisfaction

Feature O is least used

Product X User Satisfaction: Features

■ Have not used ■ Not satisfied at all ■ Not very satisfied ■ Somewhat satisfied ■ Very satisfied ■ Completely satisfied



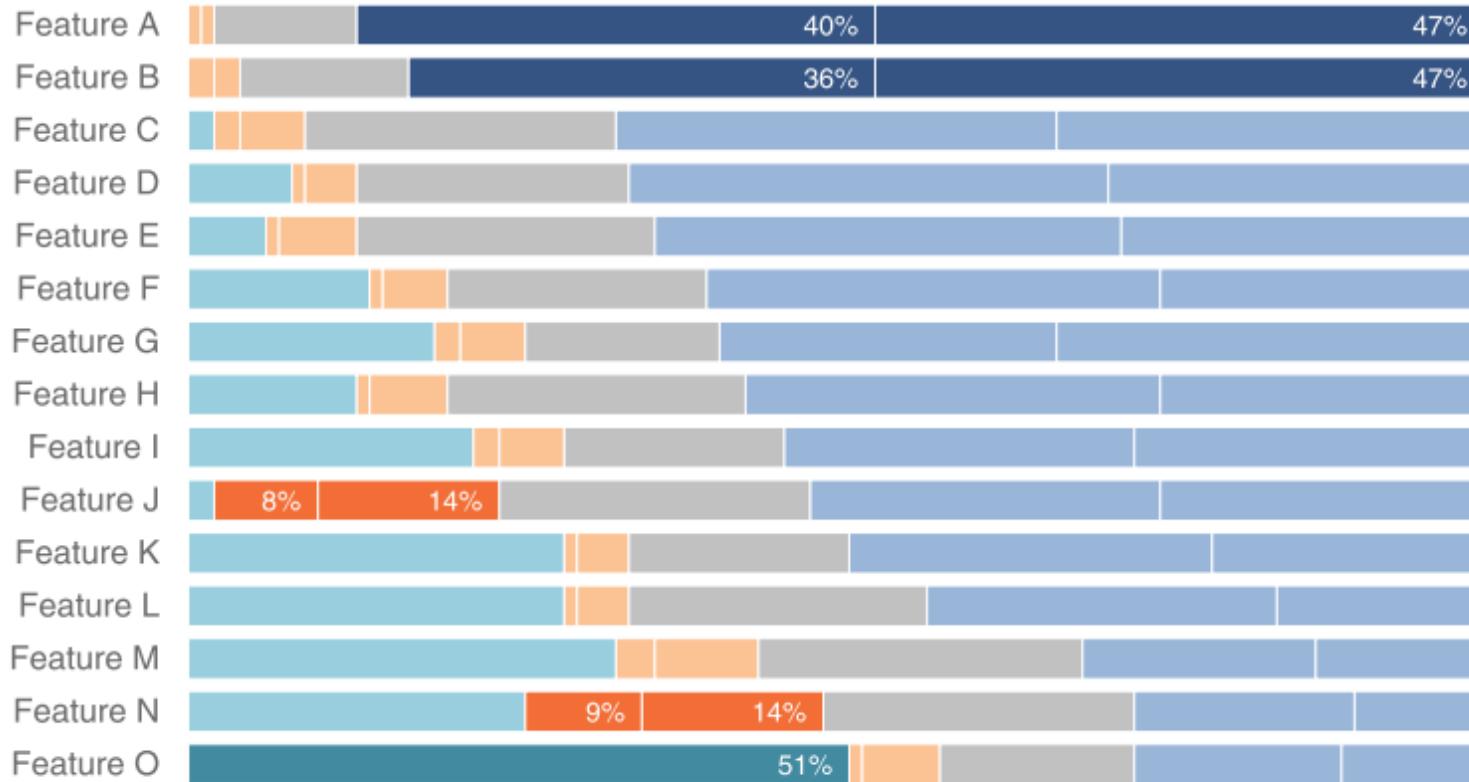
Responses based on survey question "How satisfied have you been with each of these features?".
Need more details here to help put this data into context: How many people completed survey? What proportion of users does this represent? Do those who completed survey look like the overall population, demographic-wise? When was the survey conducted?

FIGURE 9.15 Focus on unused features

User satisfaction varies greatly by feature

Product X User Satisfaction: Features

■ Have not used ■ Not satisfied at all ■ Not very satisfied ■ Somewhat satisfied ■ Very satisfied ■ Completely satisfied



Features A and B continue to top user satisfaction

Users are least satisfied with Features J and N; what improvements can we make here for a better user experience?

Feature O is least used. What steps can we proactively take with existing users to increase utilization?

Responses based on survey question "How satisfied have you been with each of these features?"

Need more details here to help put this data into context: How many people completed survey? What proportion of users does this represent?

Do those who completed survey look like the overall population, demographic-wise? When was the survey conducted?

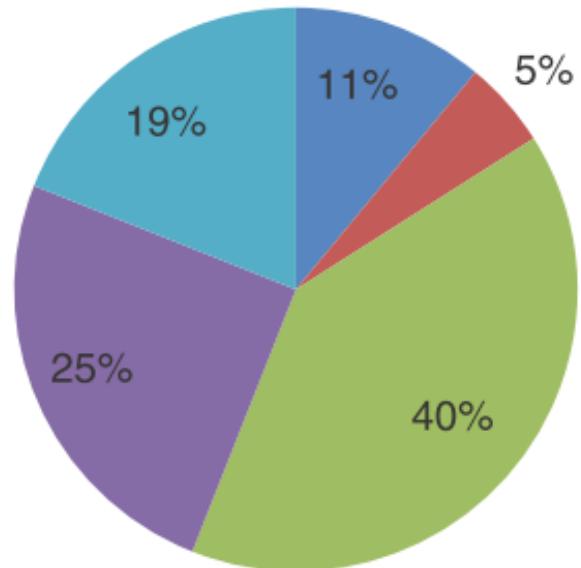
FIGURE 9.20 Comprehensive visual

Illustrer au mieux le succès du programme

Survey results: summer learning program on science

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

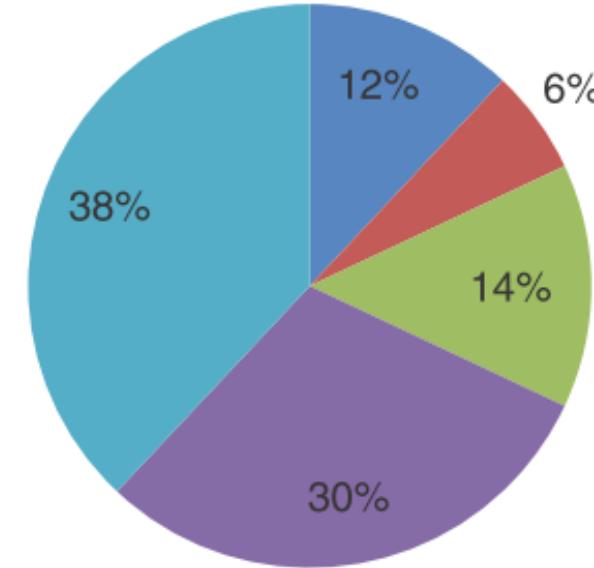


FIGURE 9.28 Original visual

Pilot program was a success

After the pilot program,

68%

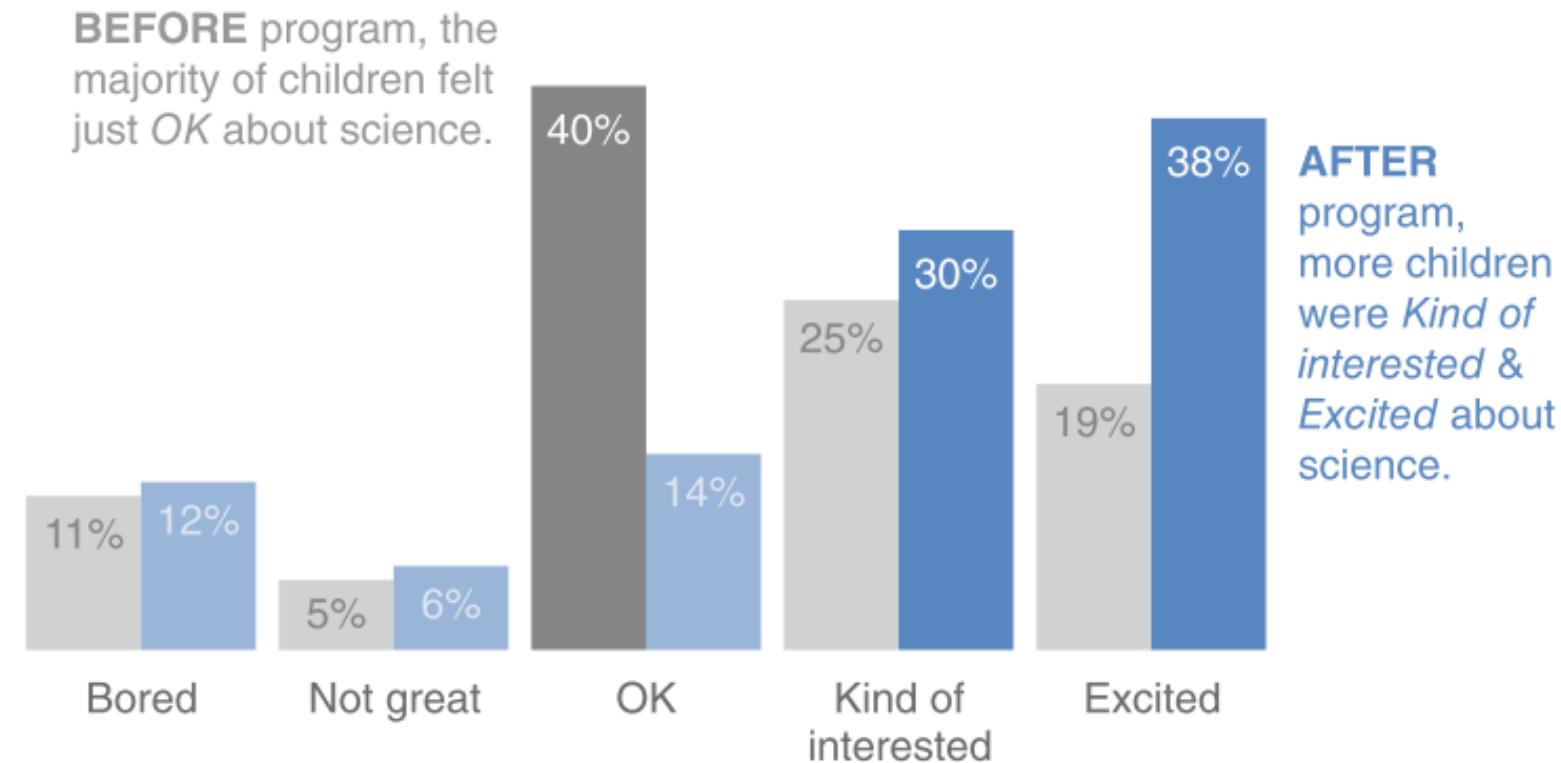
of kids expressed interest towards science,
compared to 44% going into the program.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.29 Show the numbers directly

Pilot program was a success

How do you feel about science?

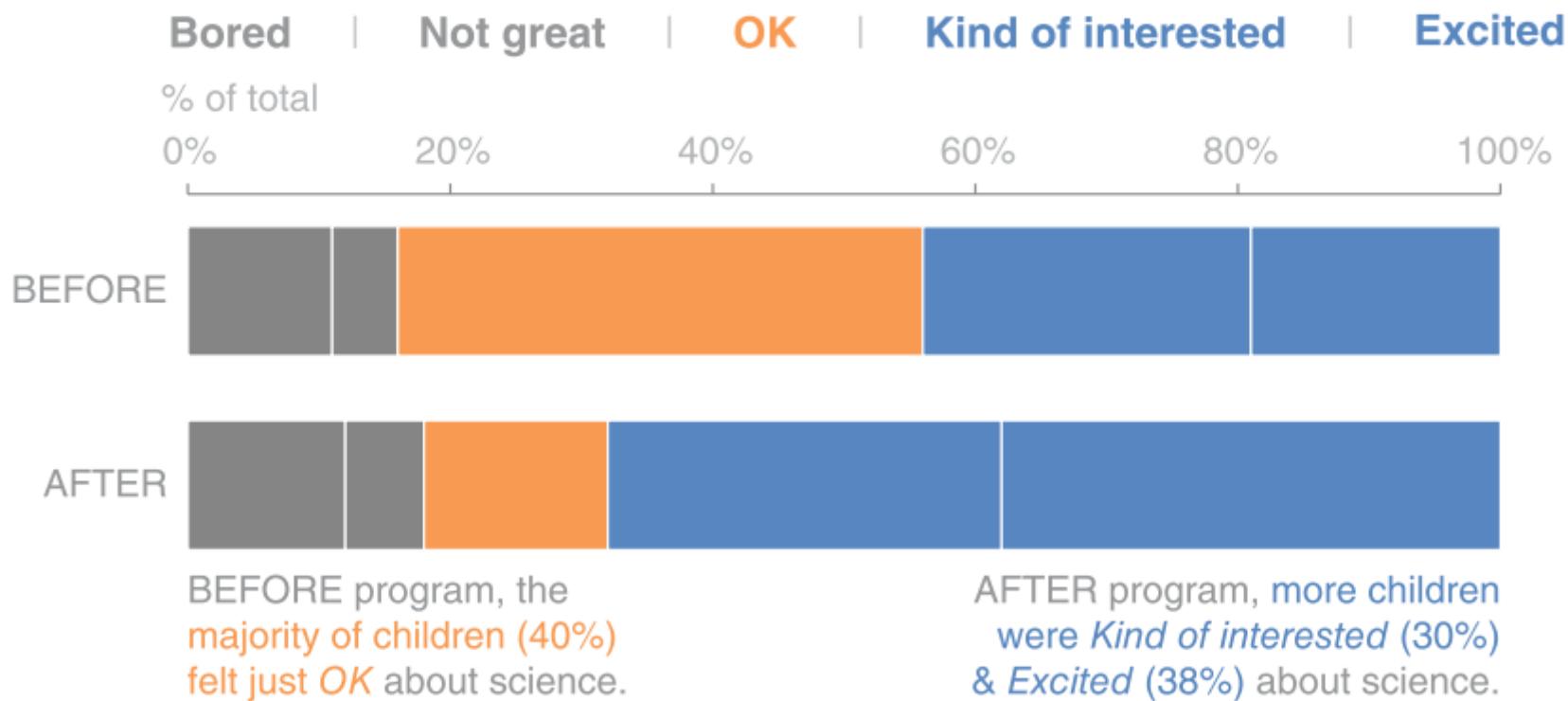


Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.30 Simple bar graph

Pilot program was a success

How do you feel about science?

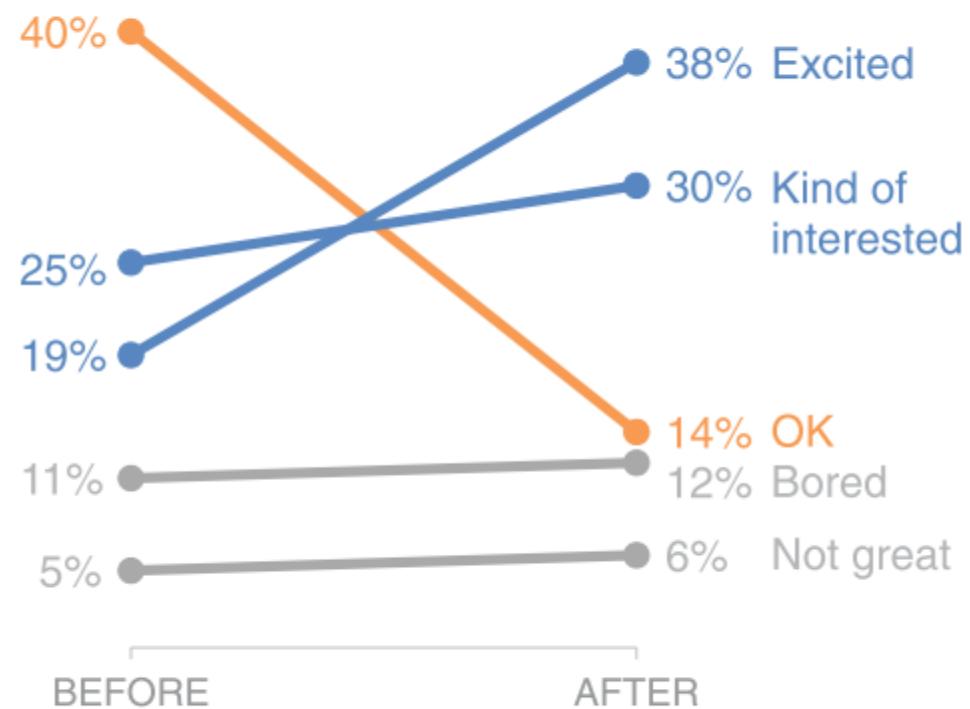


Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.31 100% stacked horizontal bar graph

Pilot program was a success

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.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.32 Slopegraph

D'autres « makeover » ...

<https://www.makeovermonday.co.uk/makeovers/makeovers-2016/>