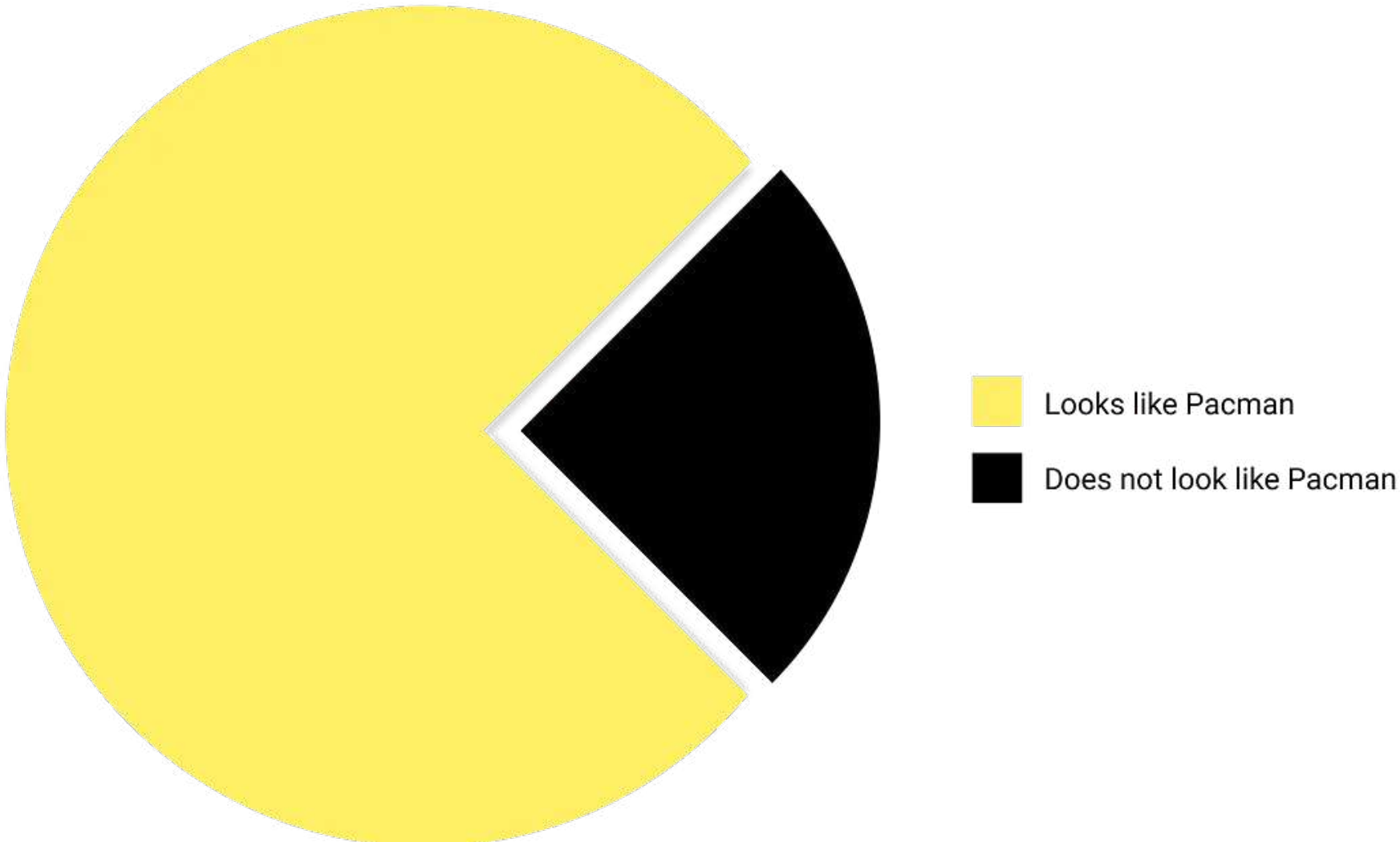
One of the few good reasons to use a Pie Chart



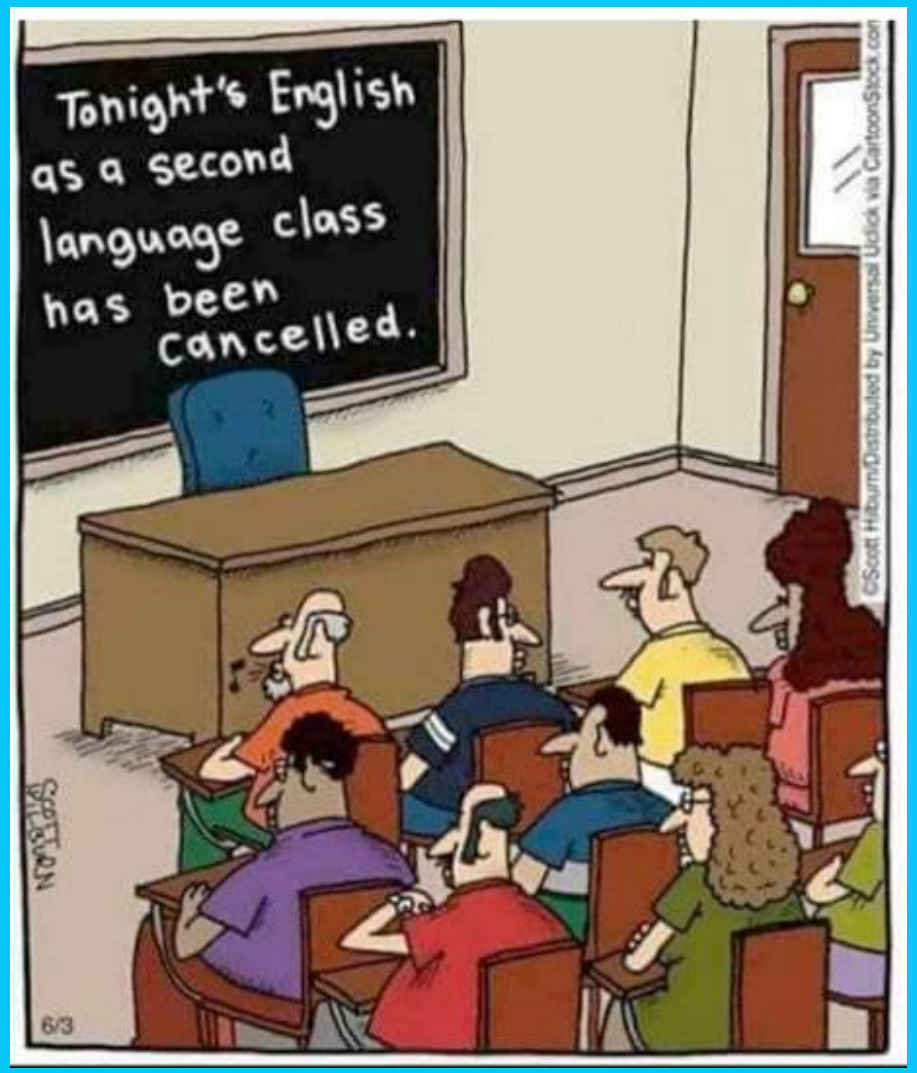
Un gráfico (bien hecho) dice más que mil palabras

Hernán E. Grecco
hgrecco@df.uba.ar

Laboratorio 4 | Abril de 2023

Departamento de Física
.UBAexactas

#1 Conocer a tu audiencia



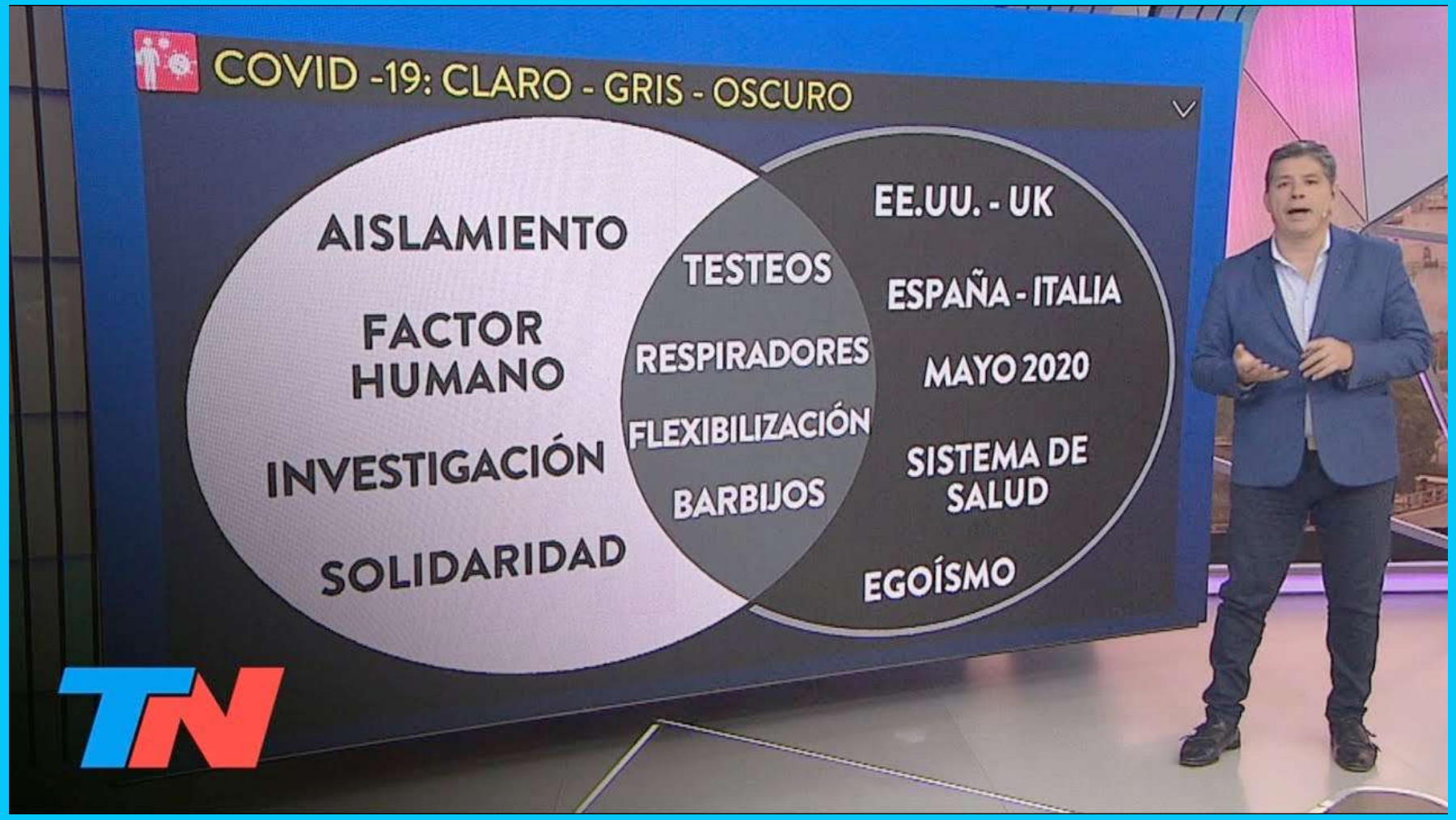
¿A quién le hablo?

Colegas de otras disciplinas
Colegas de mi misma disciplinas
Periodistas científicos
Público en general
Estudiantes
…

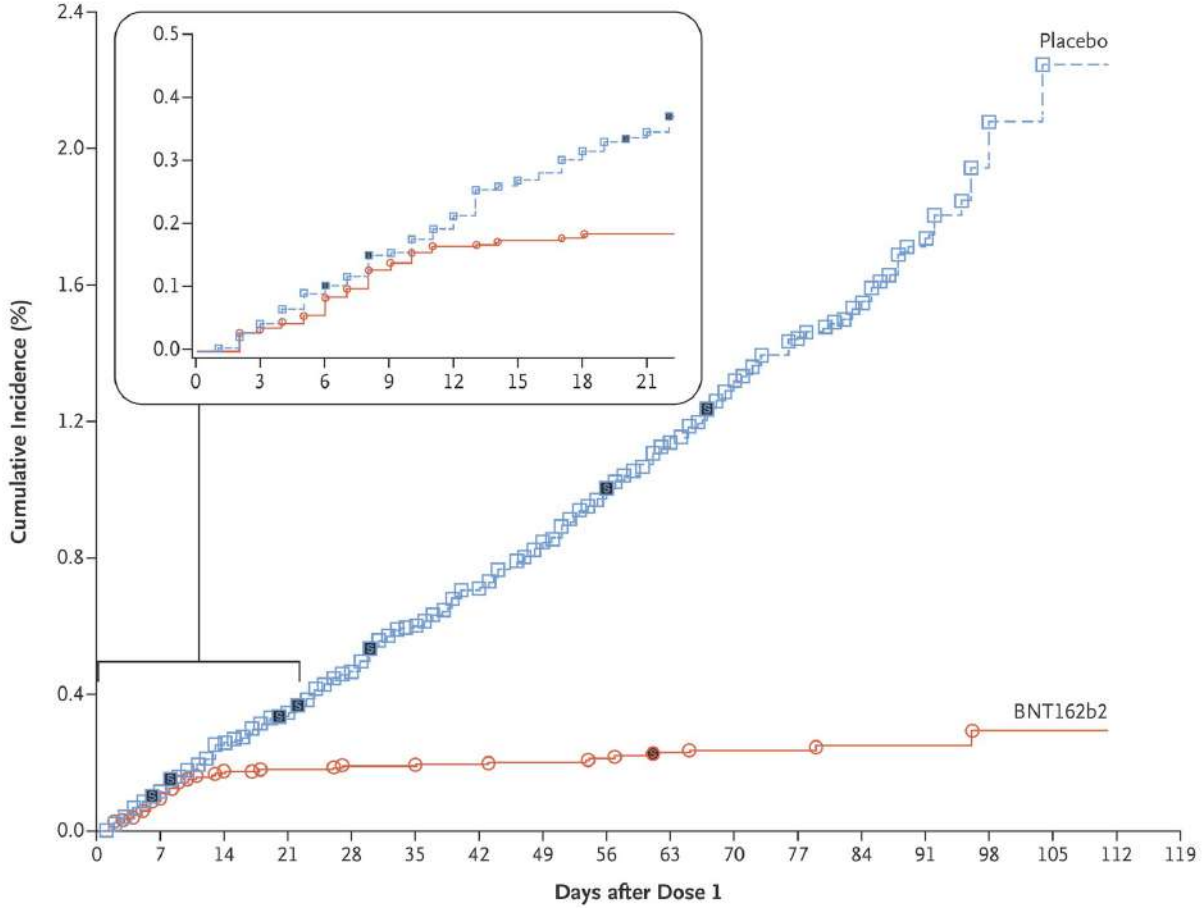
¿Por qué medio?

Charlas en congresos científicos
Posters en congresos científicos
Publicaciones científicas
Clases
Publicaciones en medios
…

#2 Definir un mensaje claro

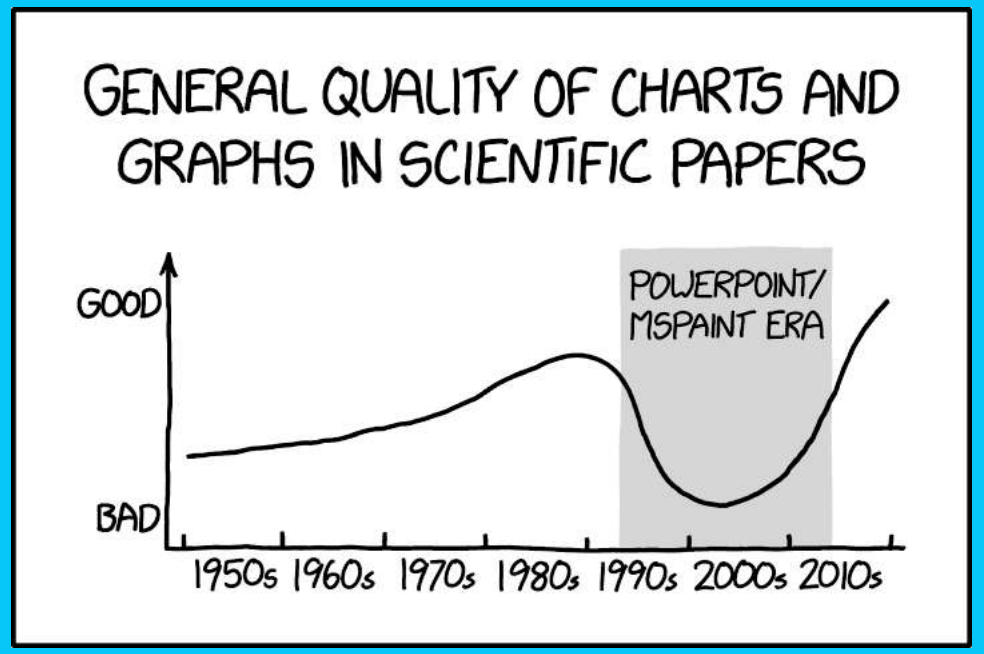


Escribir en una **oración** (sujeto y predicado)
enunciando la **conclusión** (conectada en los datos)

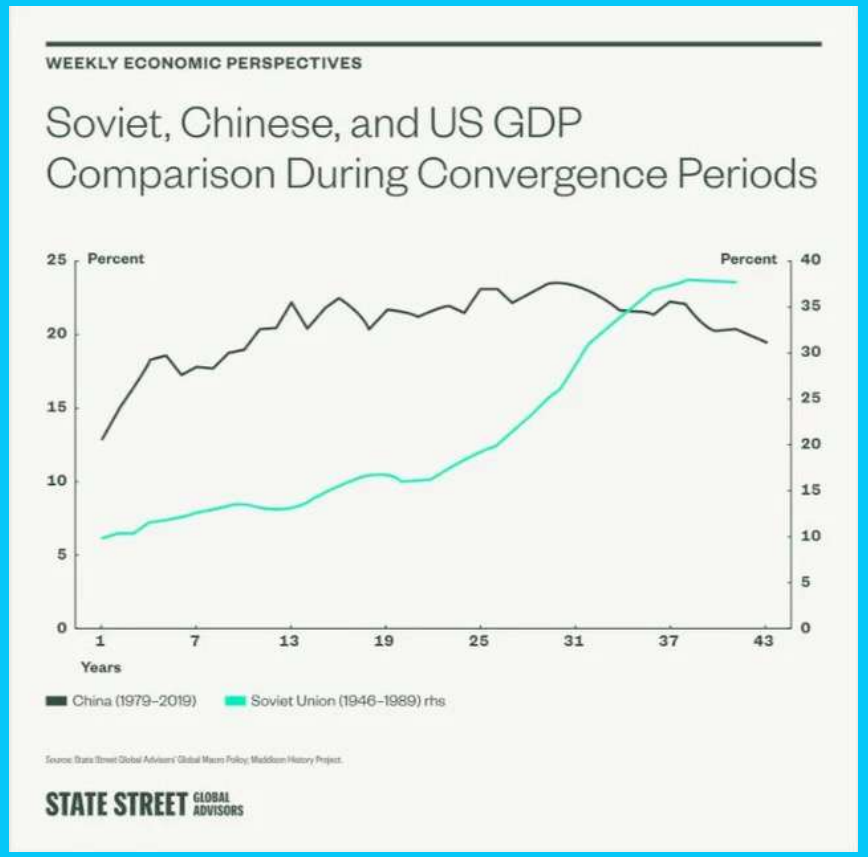


#3 Usar herramientas adecuadas

(y no abusar de ellas)



https://xkcd.com/1945/




MY FAVORITE BARS
(In Percentage of Awesomeness)
MacLaren's
12%
27%
3%
King Coin Bar
P & G

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith, Chris Campbell, Ian Bott, Liz Faunce, Graham Parrish, Billy Ehrenberg-Shannon, Paul McCallan, Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar
A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar
Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine
Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line
The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot
The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline
A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot
Usually used to show how the relationship between 2 variables has changed over time.

Bubble
Like a scatterplot but adds additional detail by sizing the circles according to a third variable.

XY heatmap
A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar
Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column
See above.

Ordered proportional symbol
Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot
Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope
Perfect for showing how ranks have changed over time or vary between categories.

Lollipop
Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Bump
Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram
The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot
A simple way of showing the change or range (min/max) of data across multiple categories.

Dot strip plot
Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Barcode plot
Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Boxplot
Summarise multiple distributions by showing the median (centre) and range of the data.

Violin plot
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple averages).

Population pyramid
A standard way for showing the age and sex breakdown of a population/distribution; effectively back to back histograms.

Cumulative curve
A good way of showing how unequal a distribution is; y axis is always cumulative frequency, x axis is always a measure.

Frequency polygons
For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm
Use to emphasise individual points in a distribution. Points can be sized to an additional variable. Best with medium-sized datasets.

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line
The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column
Columns work well for showing change over time – but usually best with only one series of data at a time.

Column + line timeline
A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope
Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart
Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Candlestick
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projections)
Use to show the uncertainty in future projections – usually this grows the further forward to projection.

Connected scatterplot
A good way of showing changing data for two variables – where there is a relatively clear pattern of progression.

Calendar heatmap
A great way of showing temporal patterns (daily, weekly, monthly) – at the expense of showing precision in quantity.

Priorities timeline
Great when date and duration are key elements of the story in the data.

Circle timeline
Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Vertical timeline
Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart, use when seeing changes in proportions over time is more important than individual values.

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation, volumes in general

Column
The standard way to compare the size of things. Must always start at 0 on the axis.

Bar
See above. Good when the data are not time series and labels have long category names.

Paired column
As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar
See above.

Marimekko
A good way of showing the size and proportion of data all in the same time – as long as the data are not too complicated.

Proportional symbol
Use when there are big variations between values and/or seeing fine differences between data is not so important.

Isotype (pictogram)
Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Lollipop
Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Radar
A space-efficient way of showing value of multiple variables – but make sure they are organised in a way that makes sense to reader.

Parallel coordinates
An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.

Bullet
Good for showing a measurement against the context of a target or performance range.

Grouped symbol
An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column/bar
A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko
A good way of showing the size and proportion of data all in the same time – as long as the data are not too complicated.

Pie
A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut
Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Treemap
Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Voronoi
A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Arc
A hemicycle, often used for visualising parliamentary composition by number of seats.

Gridplot
Good for showing % information, they work best when used on whole numbers and work well in small multiple layout form.

Venn
Generally only used for schematic representation.

Waterfall
Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)
The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional symbol (count/magnitude)
Use for totals rather than rates – be wary that small differences in data will be hard to see.

Flow map
For showing unambiguous movement across a map.

Contour map
For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values.

Equalised cartogram
Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Scaled cartogram (value)
Stretching and shrinking a map so that each area is sized according to a particular value.

Dot density
Used to show the location of individual events/actions – make sure to annotate any patterns the reader should see.

Heat map
Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information relationship graphs.

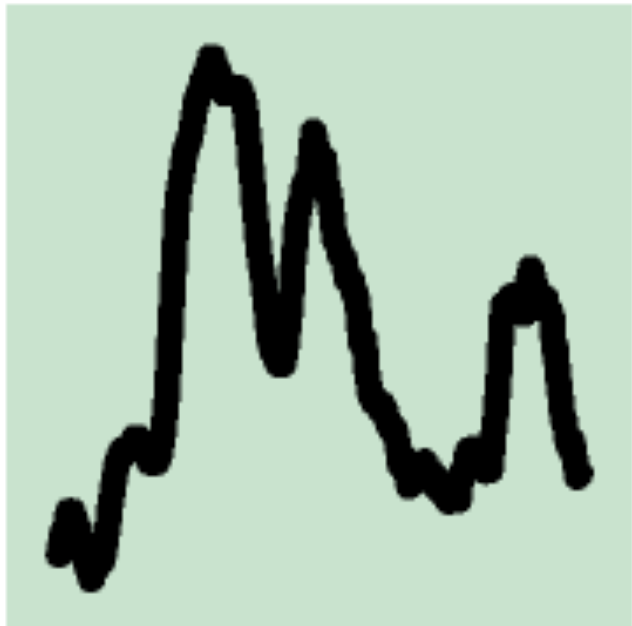
Sankey
Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall
Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord
A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

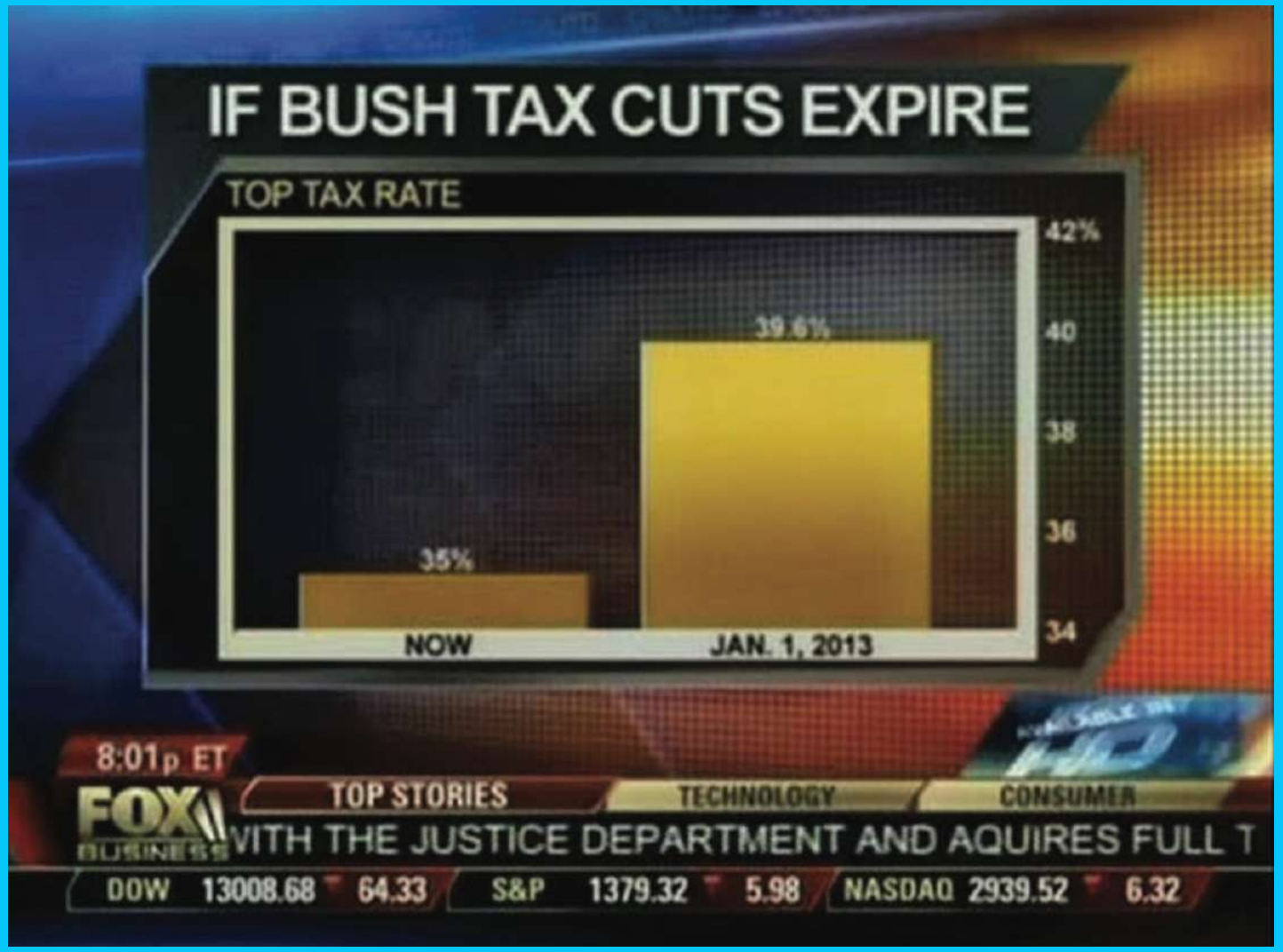
Network
Used for showing the strength and inter-connectedness of relationships of varying types.

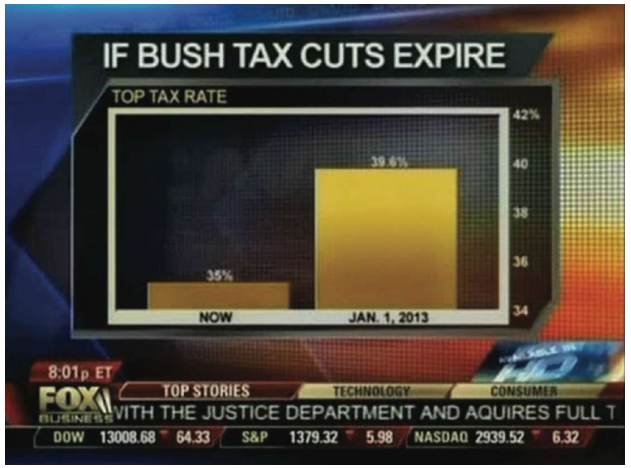
Line

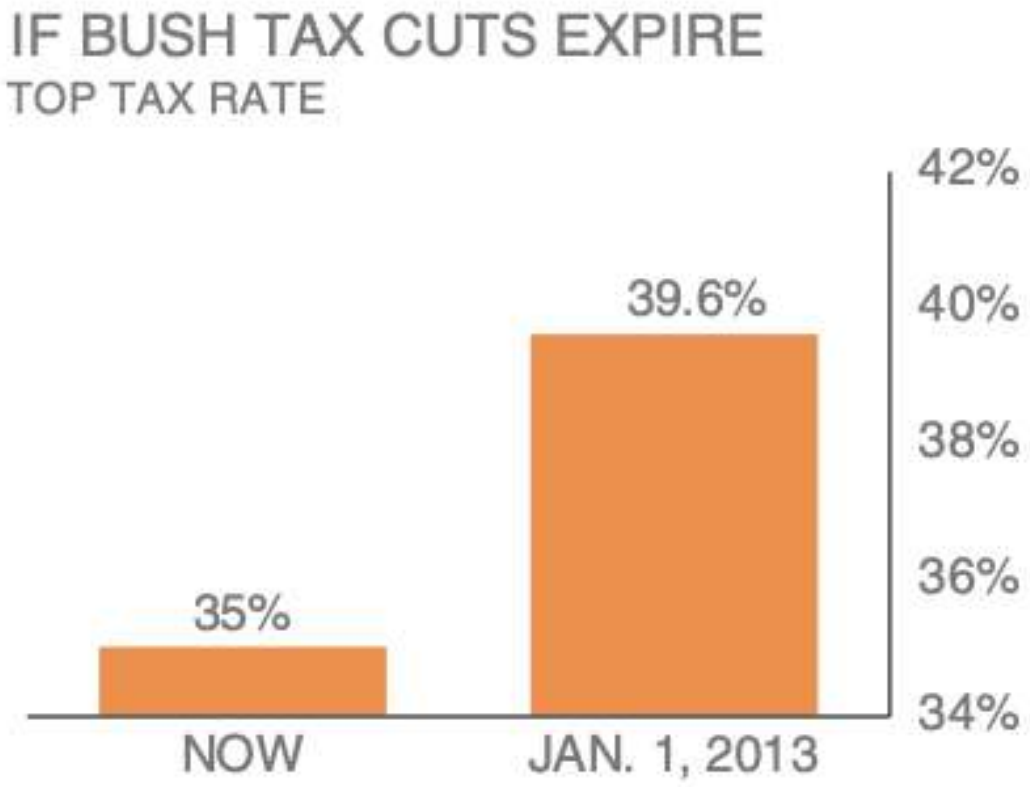


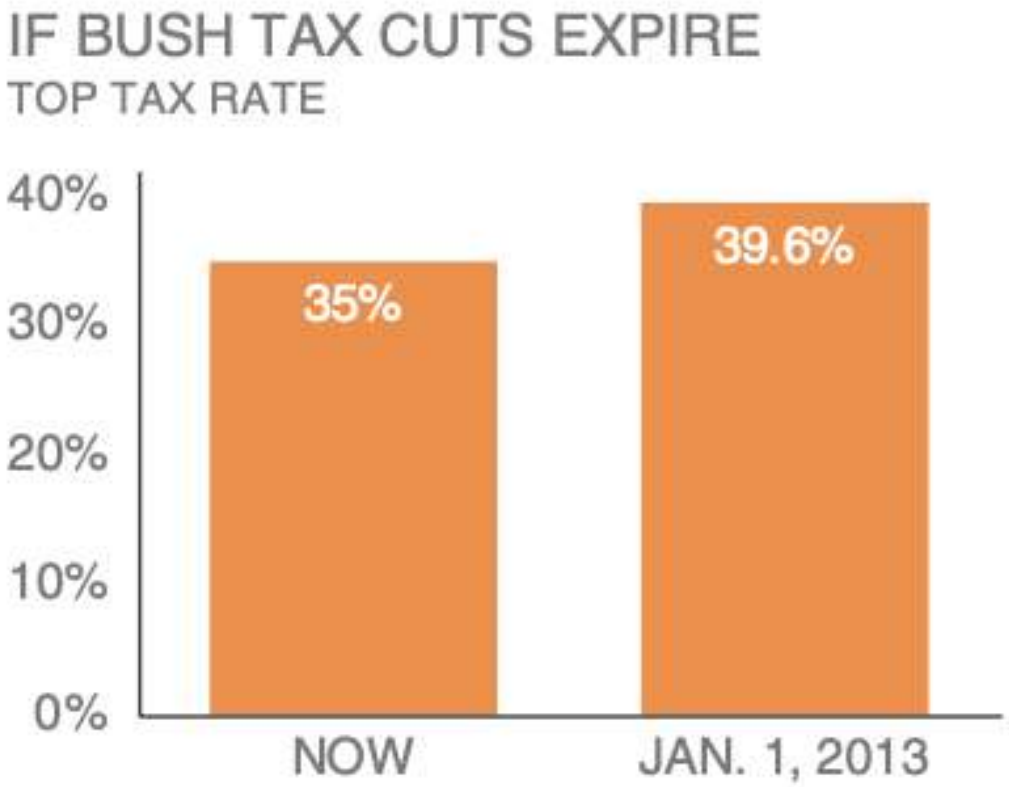
The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

#4 Graficar los datos fielmente



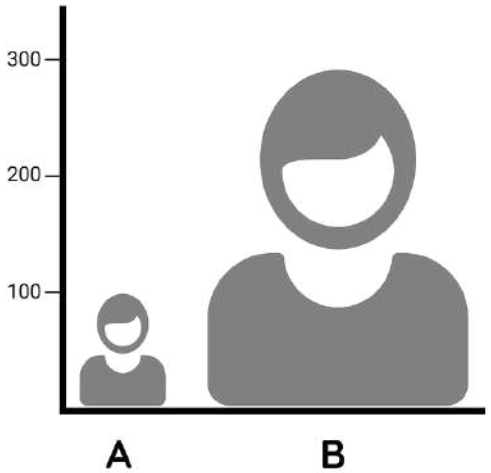

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
42%
40
38
36
34
39.6%
35%
NOW
JAN. 1, 2013
8:01p ET
TOP STORIES
TECHNOLOGY
CONSUMER
WITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T
DOW 13008.68 64.33
S&P 1379.32 5.98
NASDAQ 2939.52 6.32


IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
42%
40%
38%
36%
34%
39.6%
35%
NOW
JAN. 1, 2013

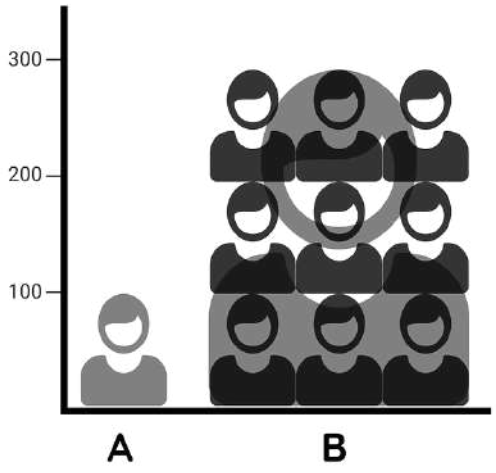

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
40%
30%
20%
10%
0%
39.6%
35%
NOW
JAN. 1, 2013


WELFARE VS. FULL TIME JOBS
108.6M
101.7M
PEOPLE ON WELFARE
PEOPLE WITH A FULL TIME JOB
SOURCE: CENSUS BUREAU, 2011
FOX NEWS channel
RED SOX BEAT ST. LOUIS CARDINALS 4-2 TO EVEN WORLD SERIES AT T
mediamatters.org

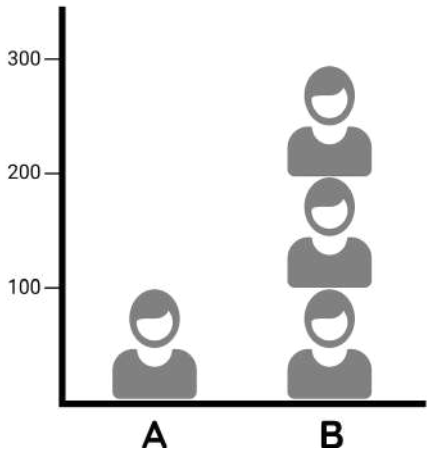
Misleading Pictograms

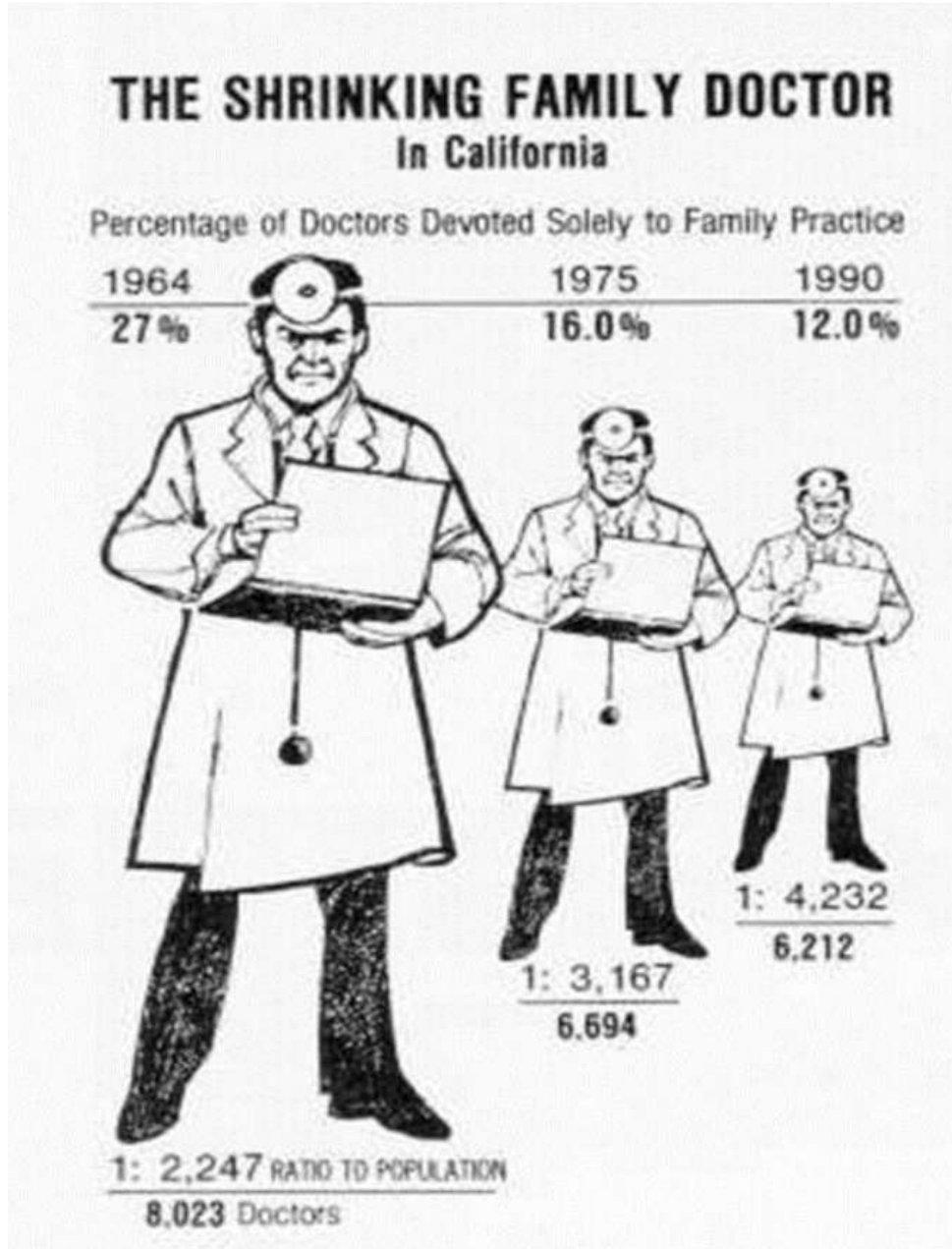


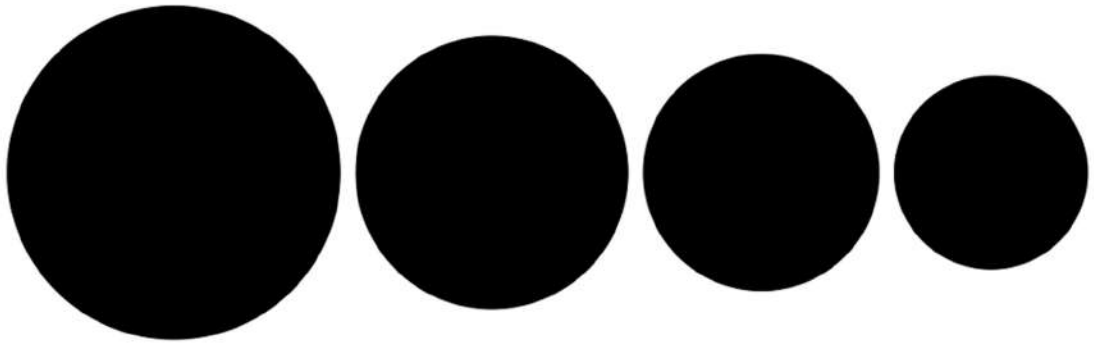
Comparison



Fixed Pictograms

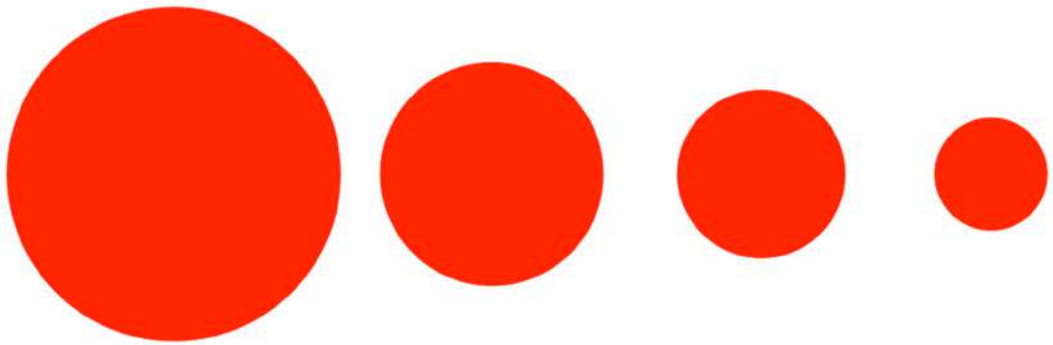






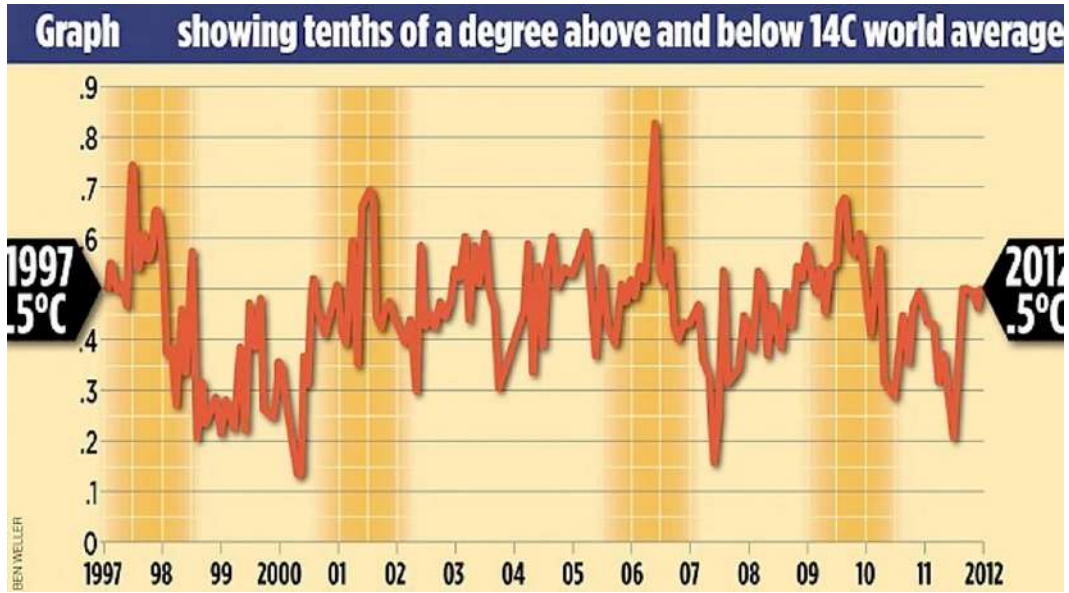
Relative size using disc area

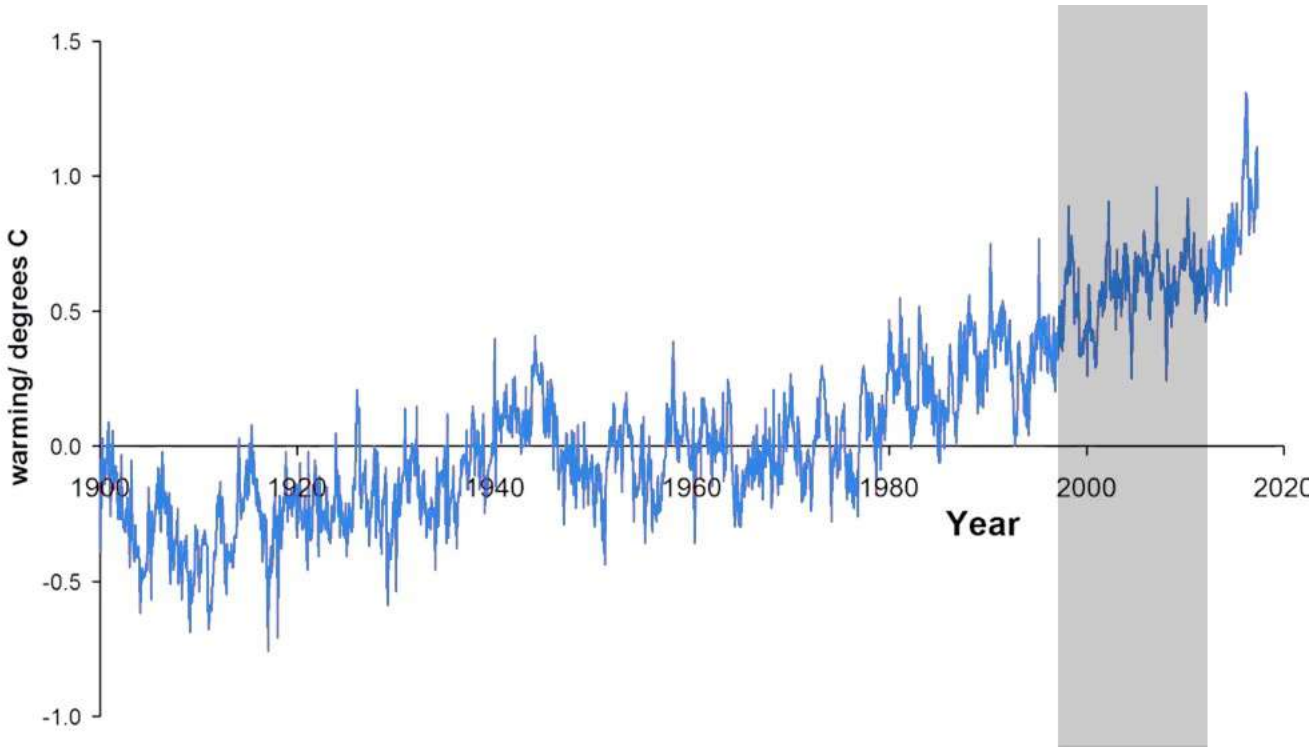
Relative size using disc radius



Relative size using full range

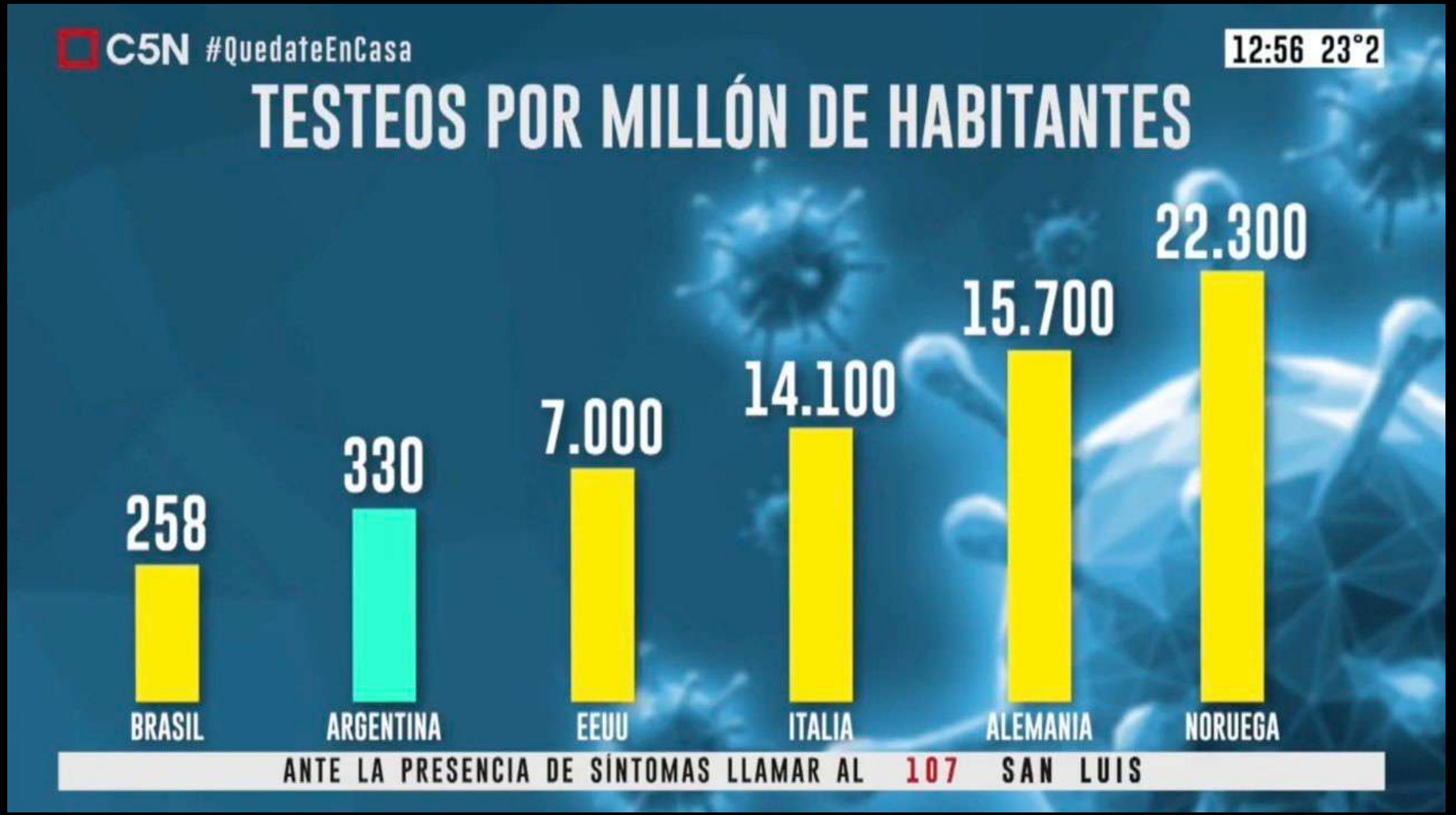
Relative size using partial range


Graph showing tenths of a degree above and below 14C world average
1997
.5°C
.9
.8
.7
.6
.4
.3
.2
.1
0
1997 98 99 2000 01 02 03 04 05 06 07 08 09 10 11 2012
BEN WELLER


warming/ degrees C
1.5
1.0
0.5
0.0
-0.5
-1.0
1900
1920
1940
1960
1980
2000
2020
Year

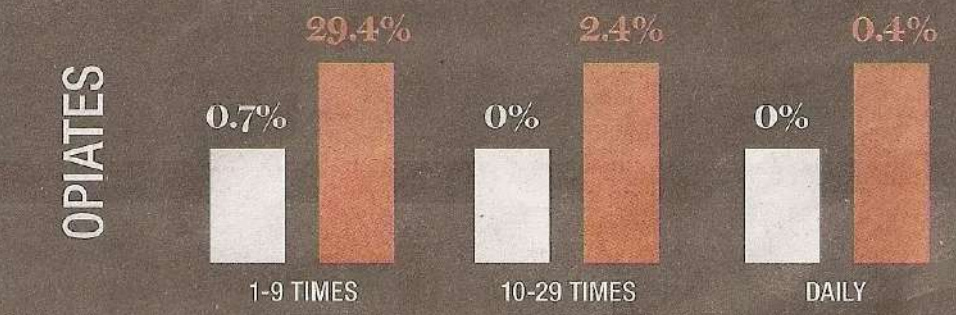
#5 Tener consistencia interna

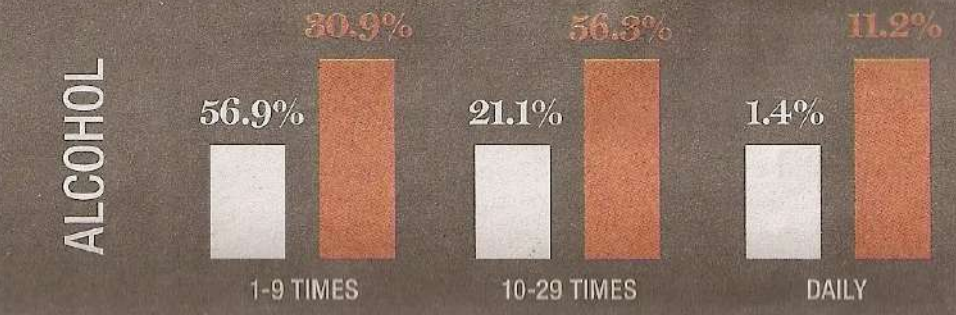

COVID -19: CLARO - GRIS - OSCURO
AISLAMIENTO
FACTOR HUMANO
INVESTIGACIÓN
SOLIDARIDAD
TESTEOS
RESPIRADORES
FLEXIBILIZACIÓN
BARBIJOS
EE.UU. - UK
ESPAÑA - ITALIA
MAYO 2020
SISTEMA DE SALUD
EGOÍSMO
TN

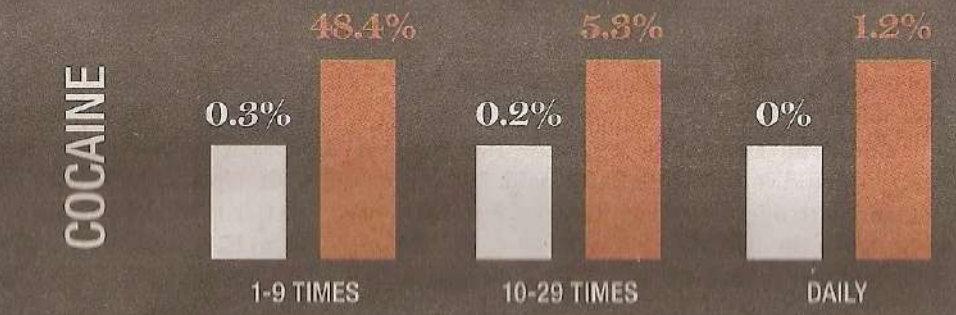

C5N #QuedateEnCasa
12:56 23°2
TESTEOS POR MILLÓN DE HABITANTES
258
330
7.000
14.100
15.700
22.300
BRASIL
ARGENTINA
EEUU
ITALIA
ALEMANIA
NORUEGA
ANTE LA PRESENCIA DE SÍNTOMAS LLAMAR AL 107 SAN LUIS

BY THE NUMBERS

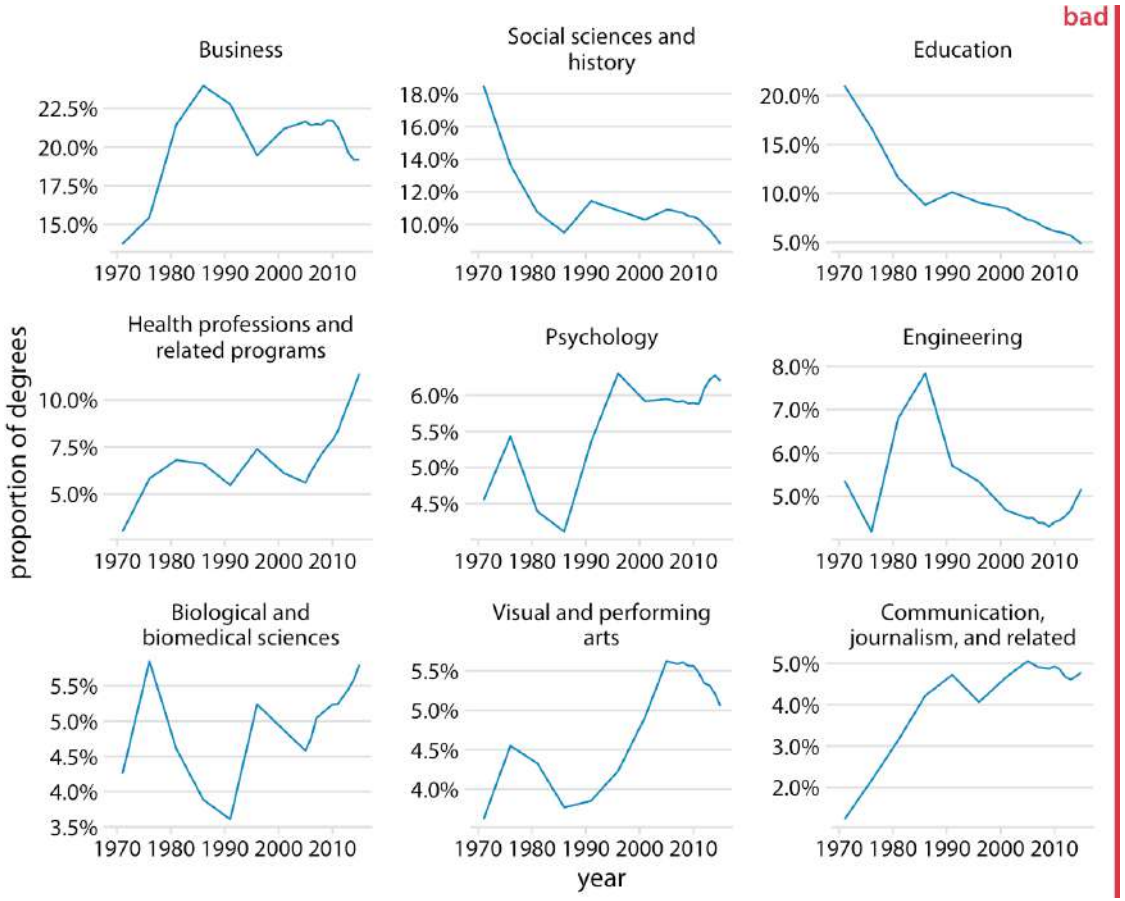
The National Collegiate Health Assessment was taken by 1,000 UCSB students in Spring 2009. Participants were asked how frequently they used substances over the past 30 days. Numbers in white reflect actual student use, while red numbers indicate perceived substance use. The average age of participants was 20 years and approximately 99 percent were full-time students.

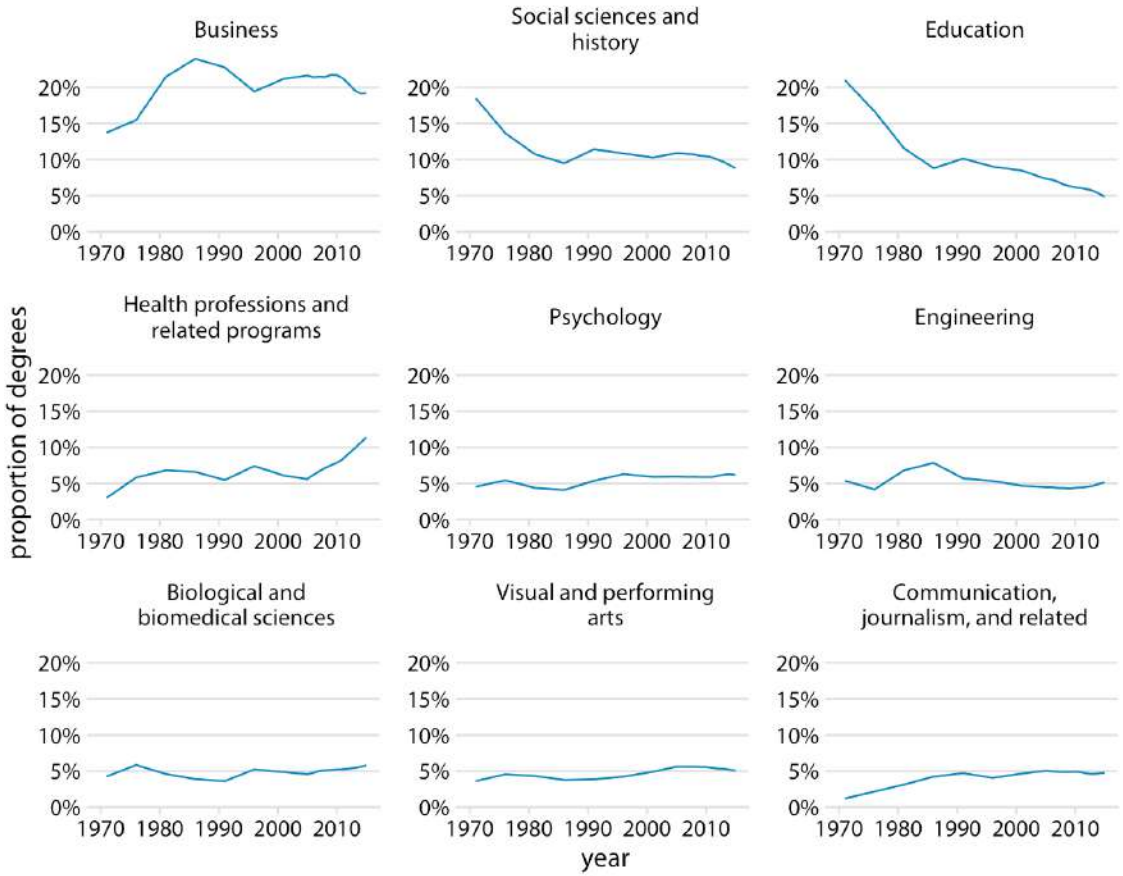




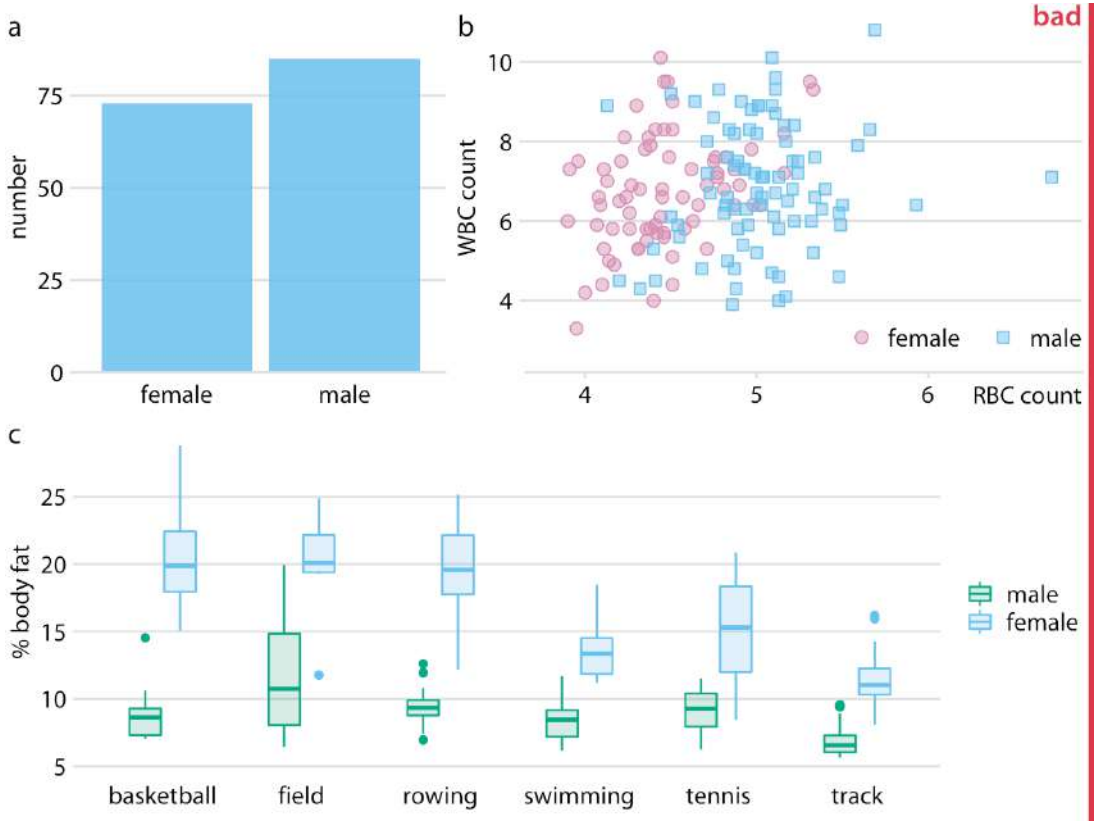


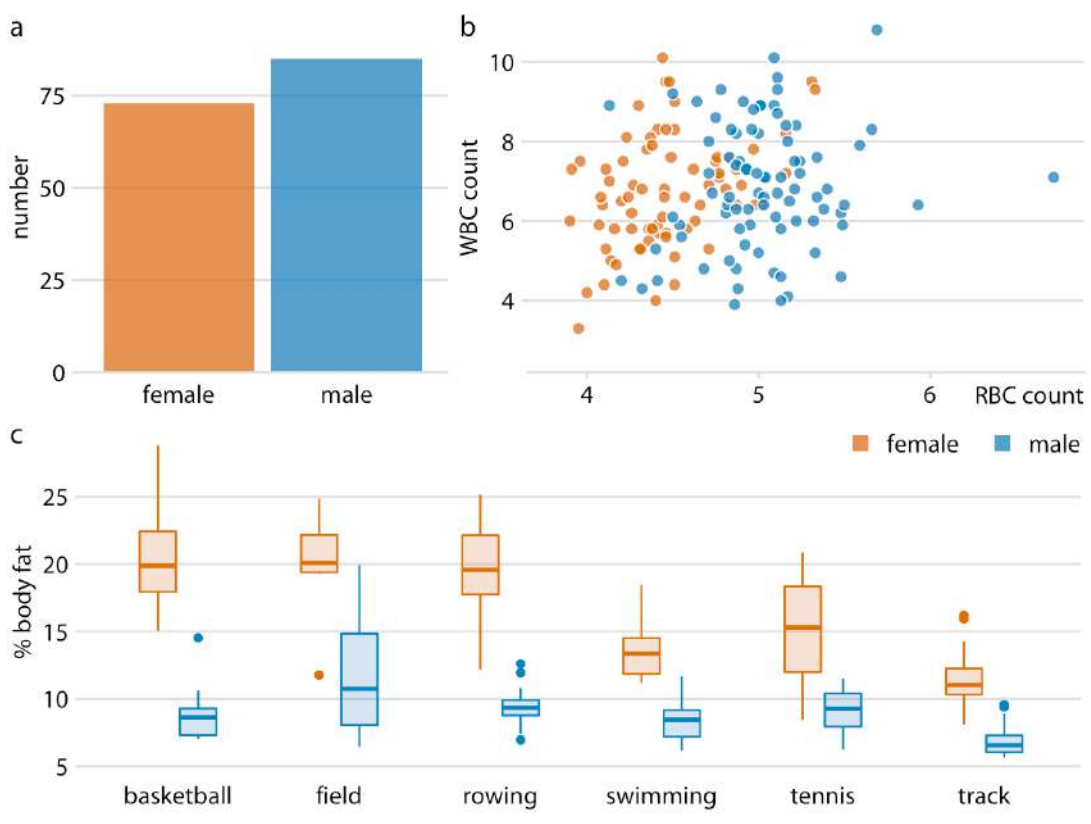
Antes y después



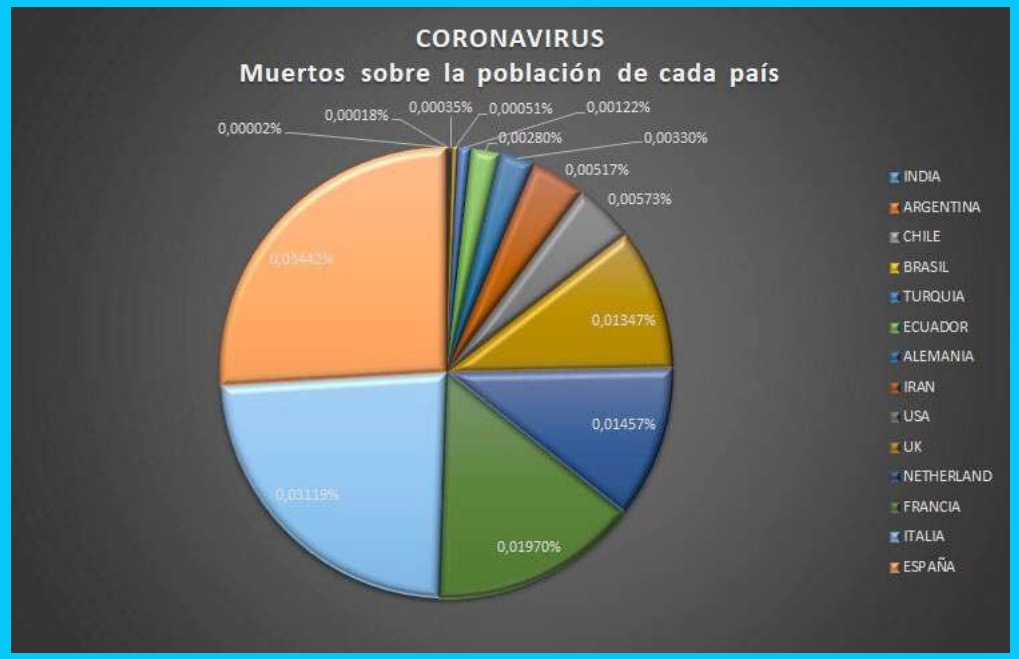


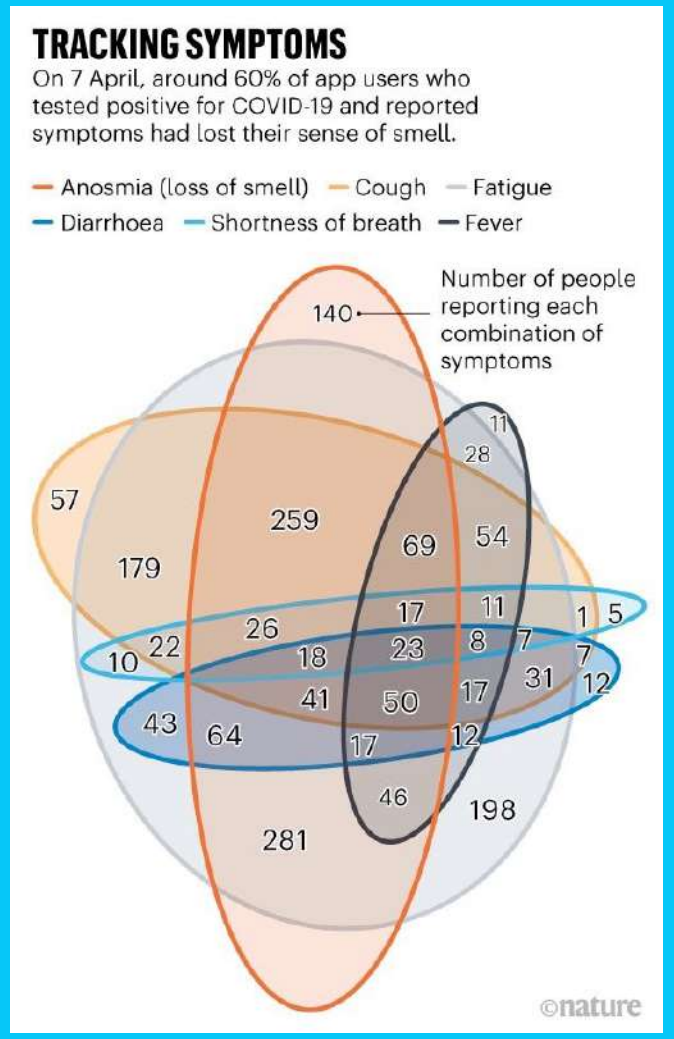
Antes y después



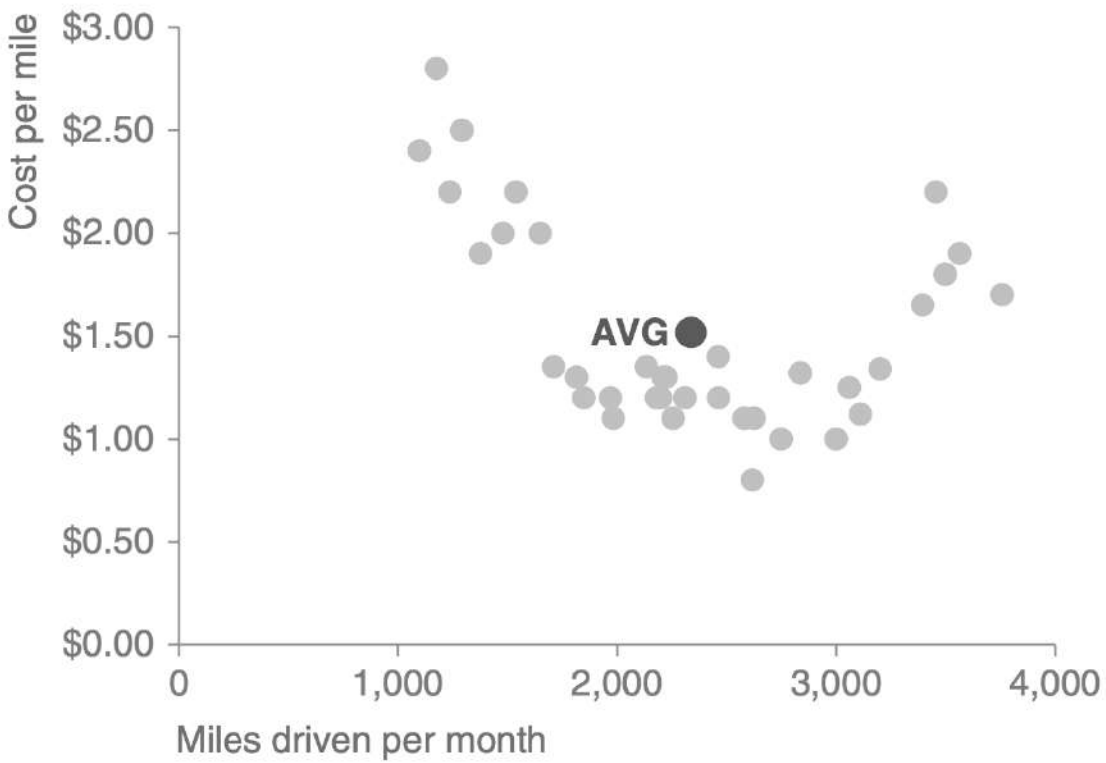


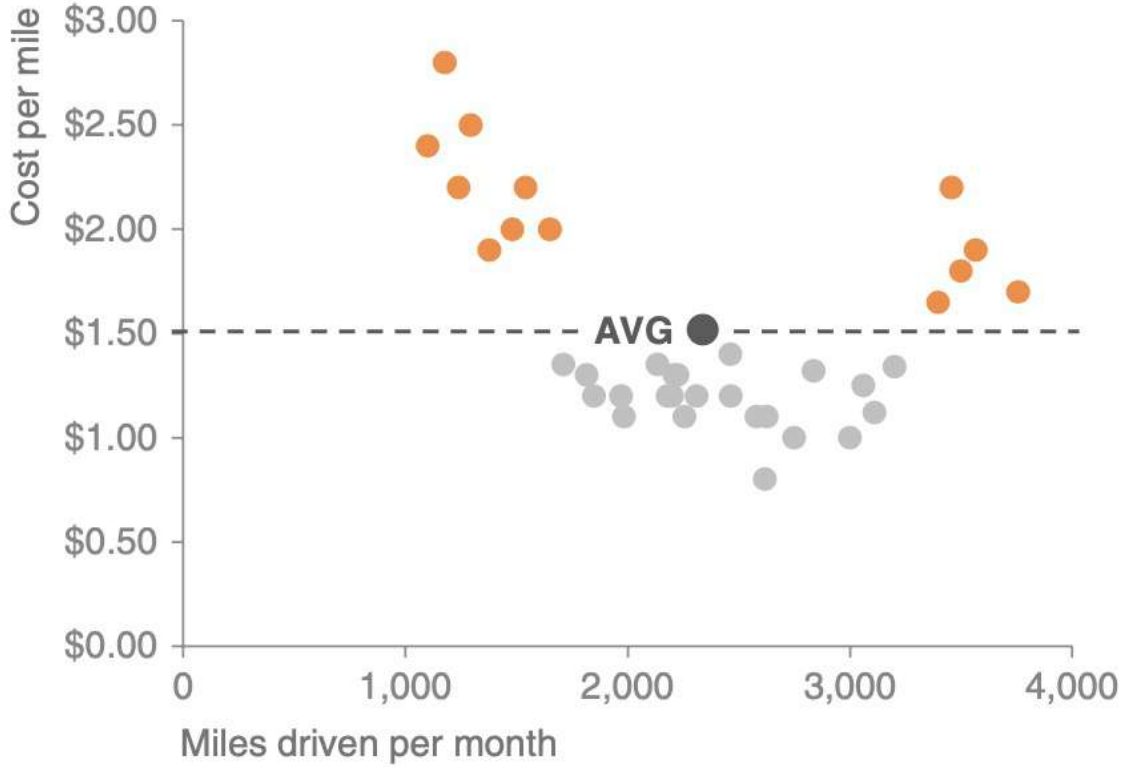
#6 Simplificar y jerarquizar



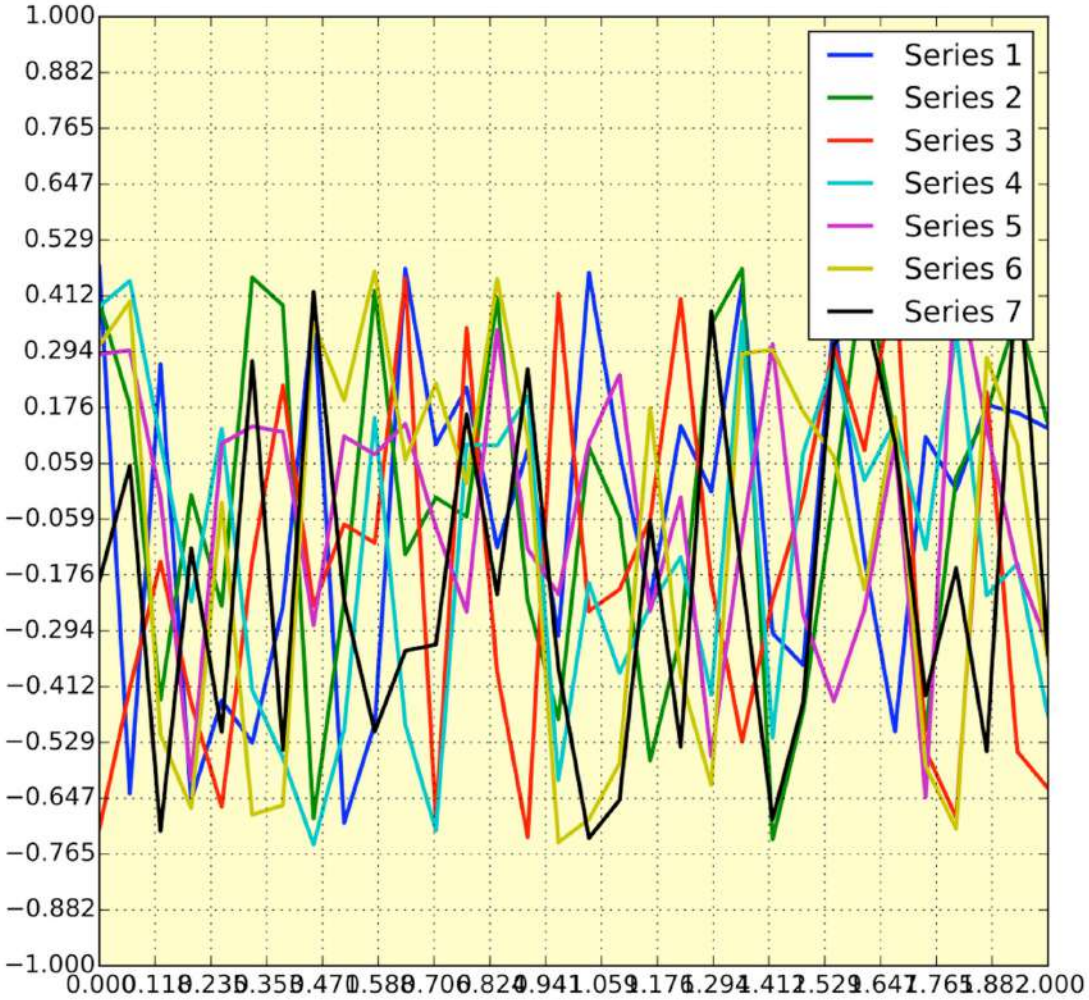


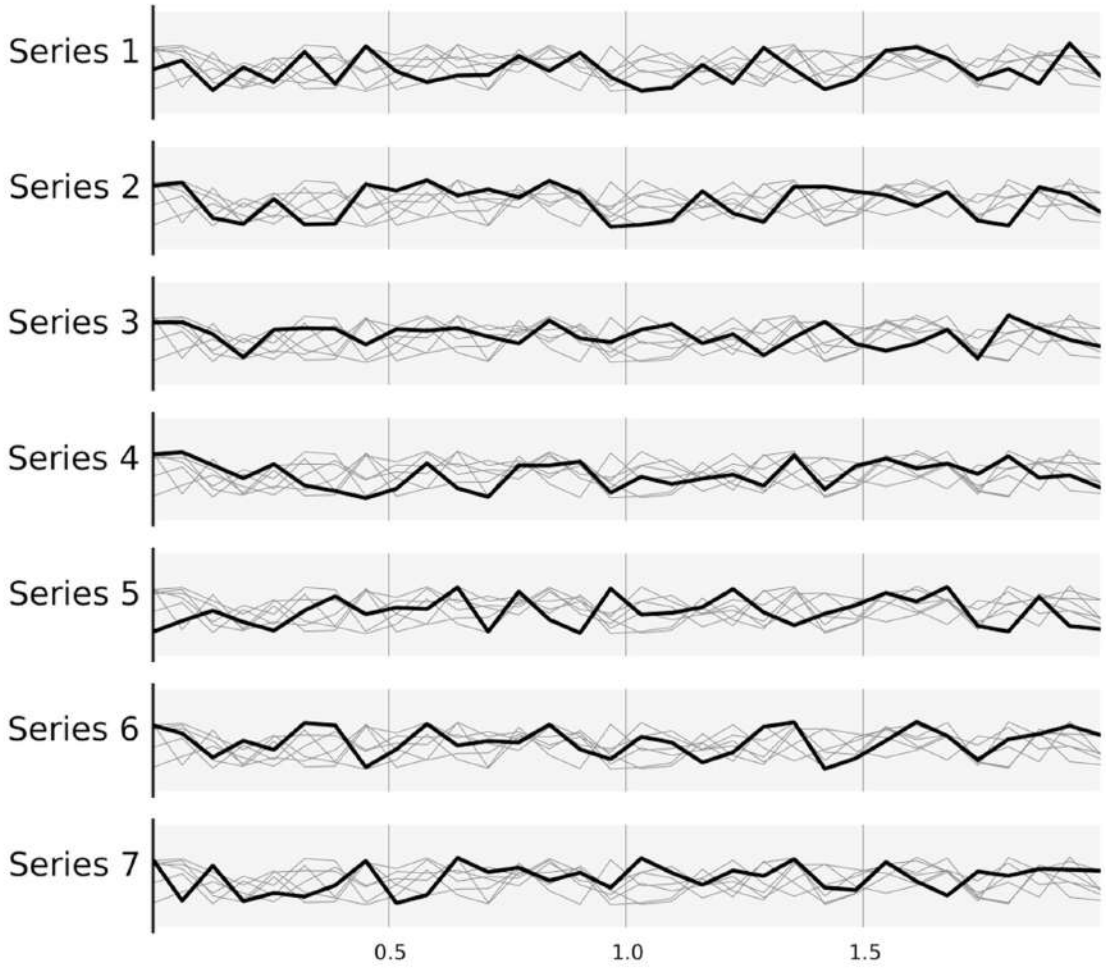
Antes y después



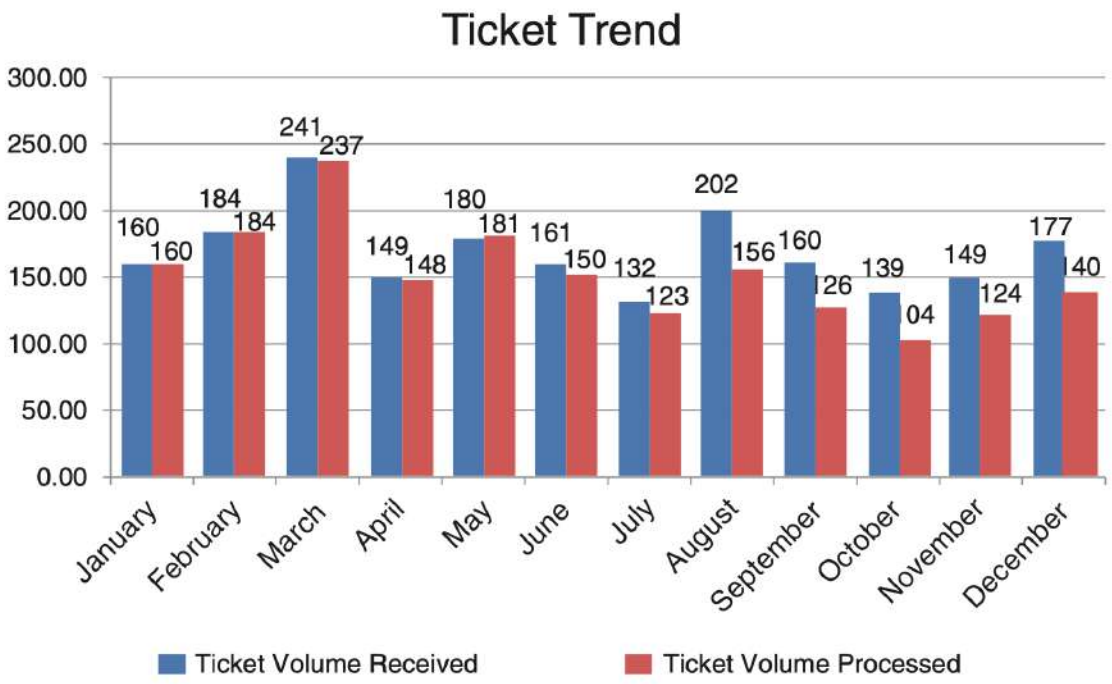


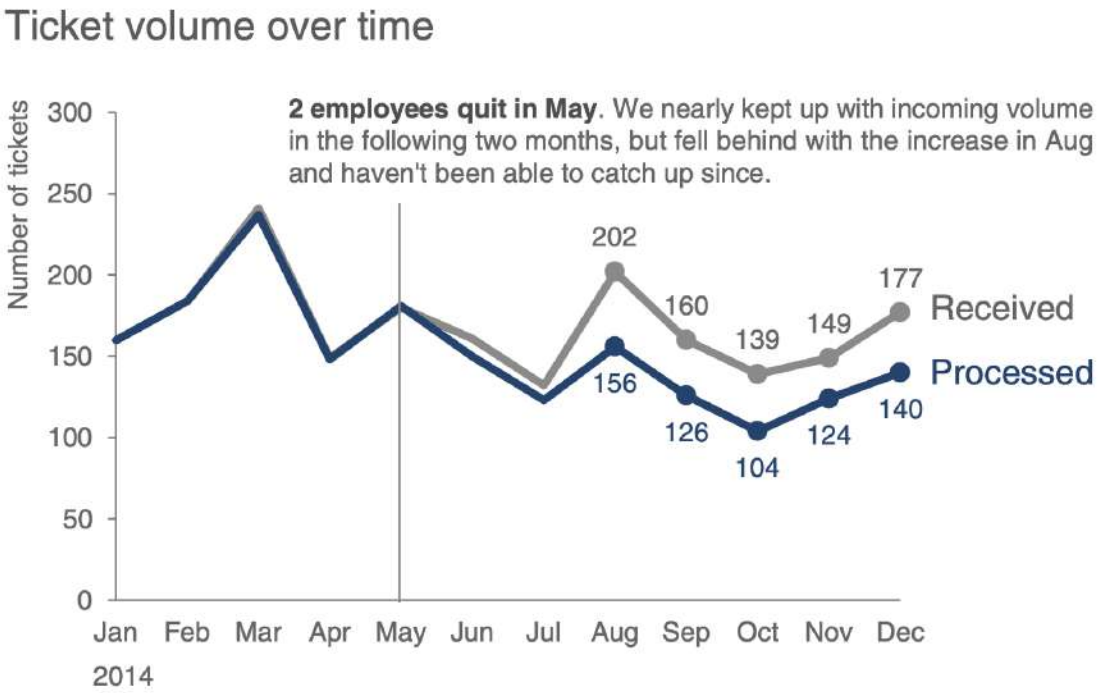
Antes y después



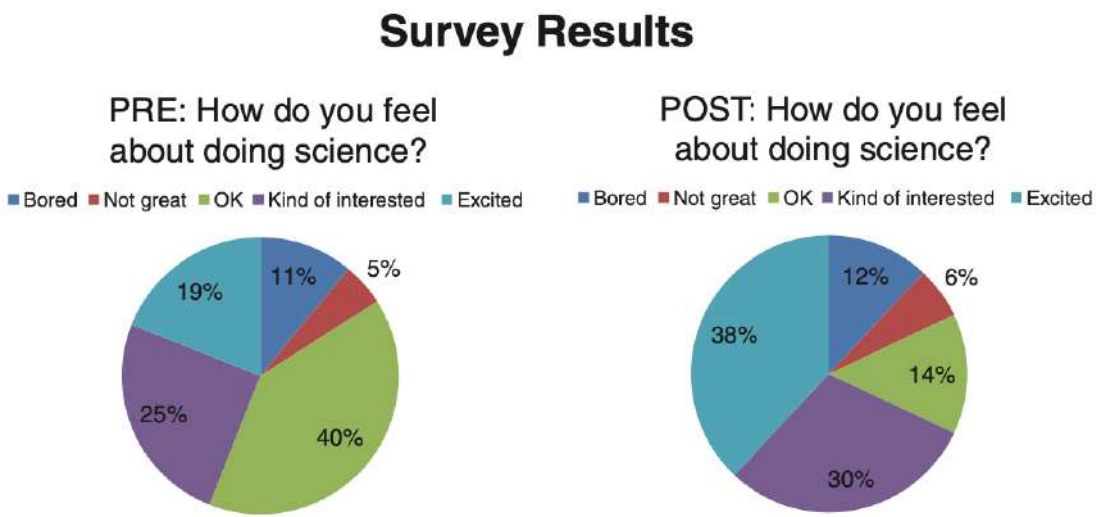


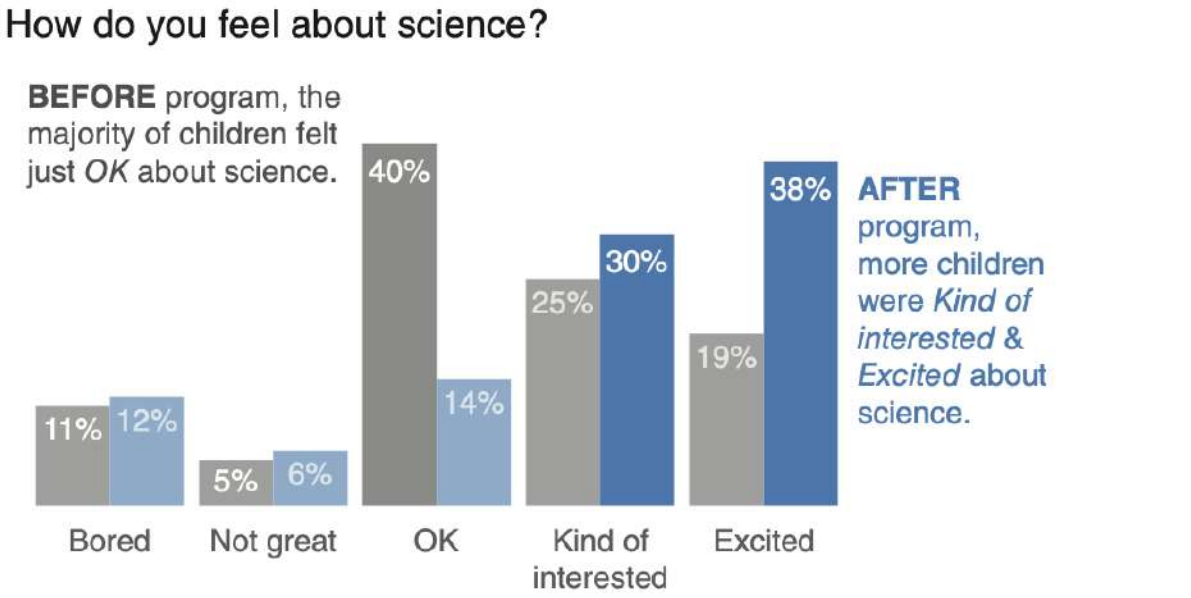
Antes y después





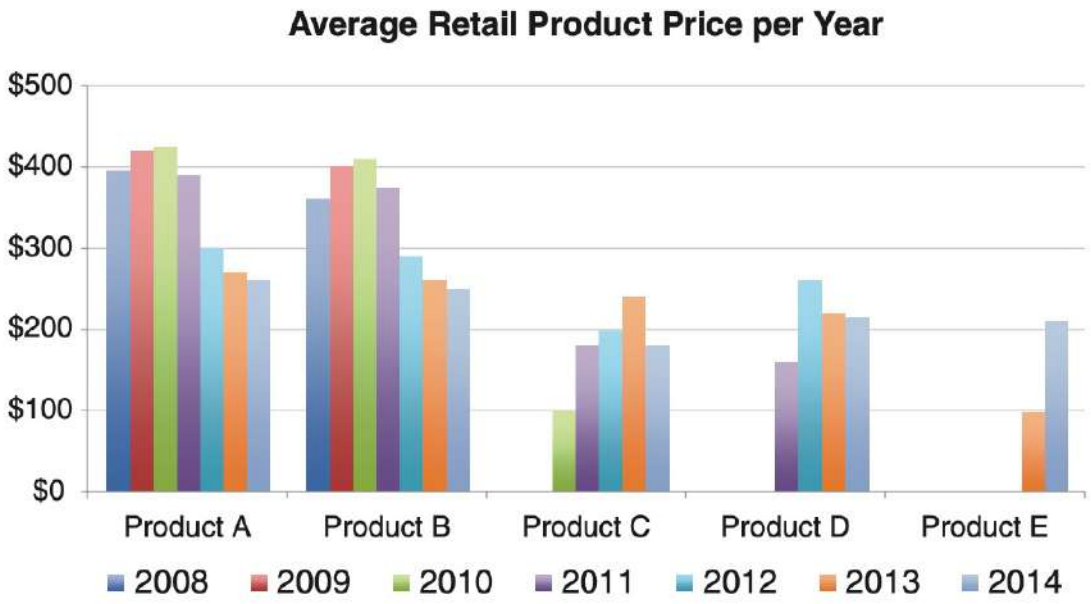
Antes y después

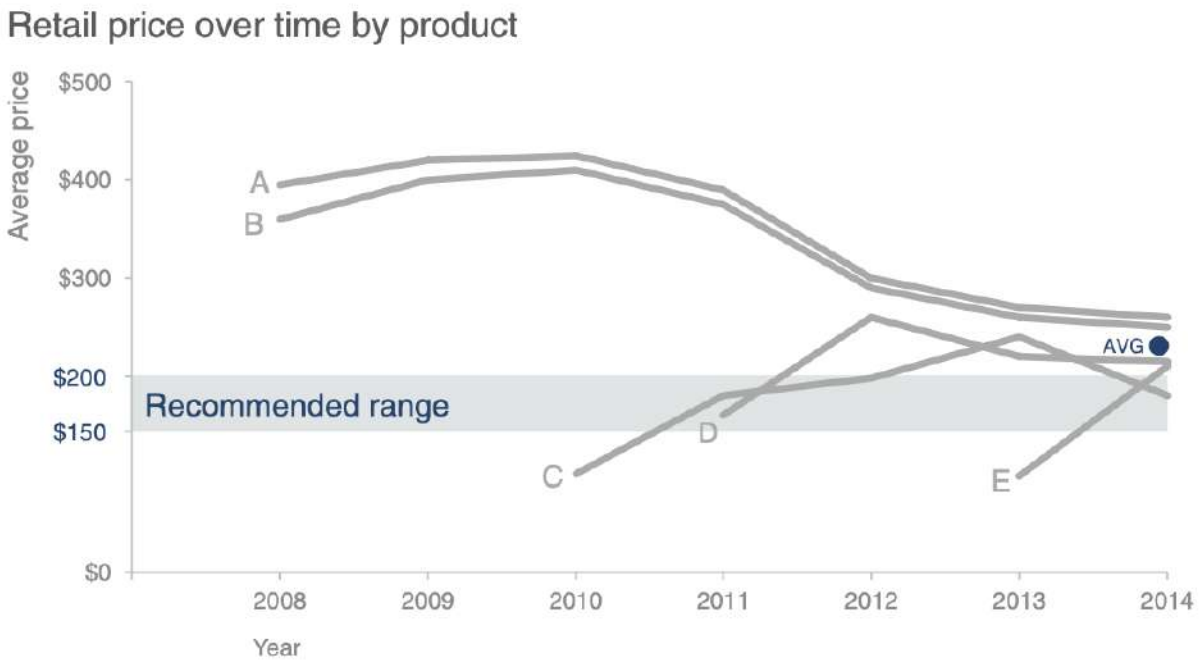




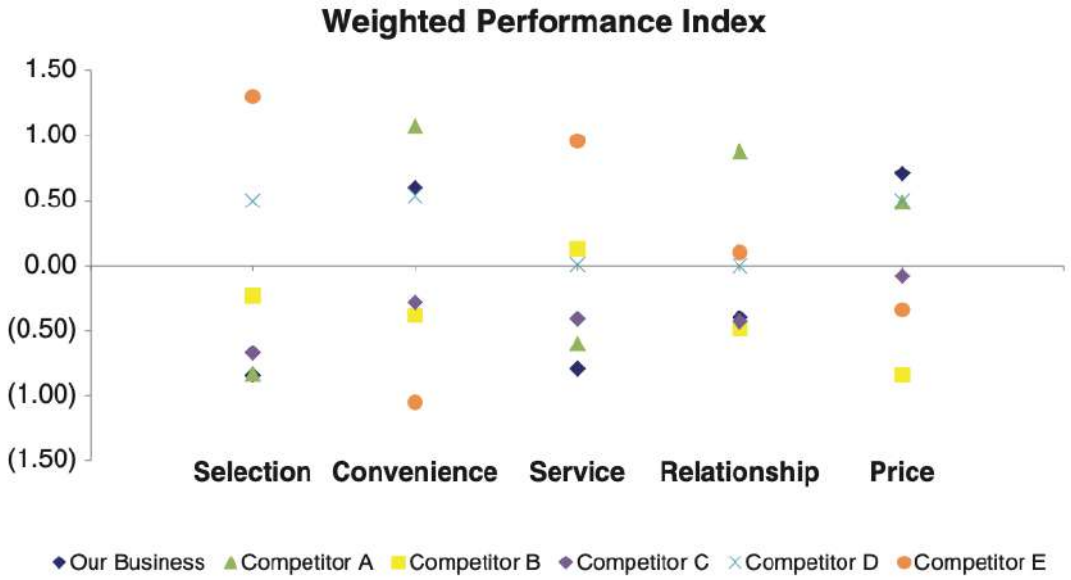
Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Antes y después

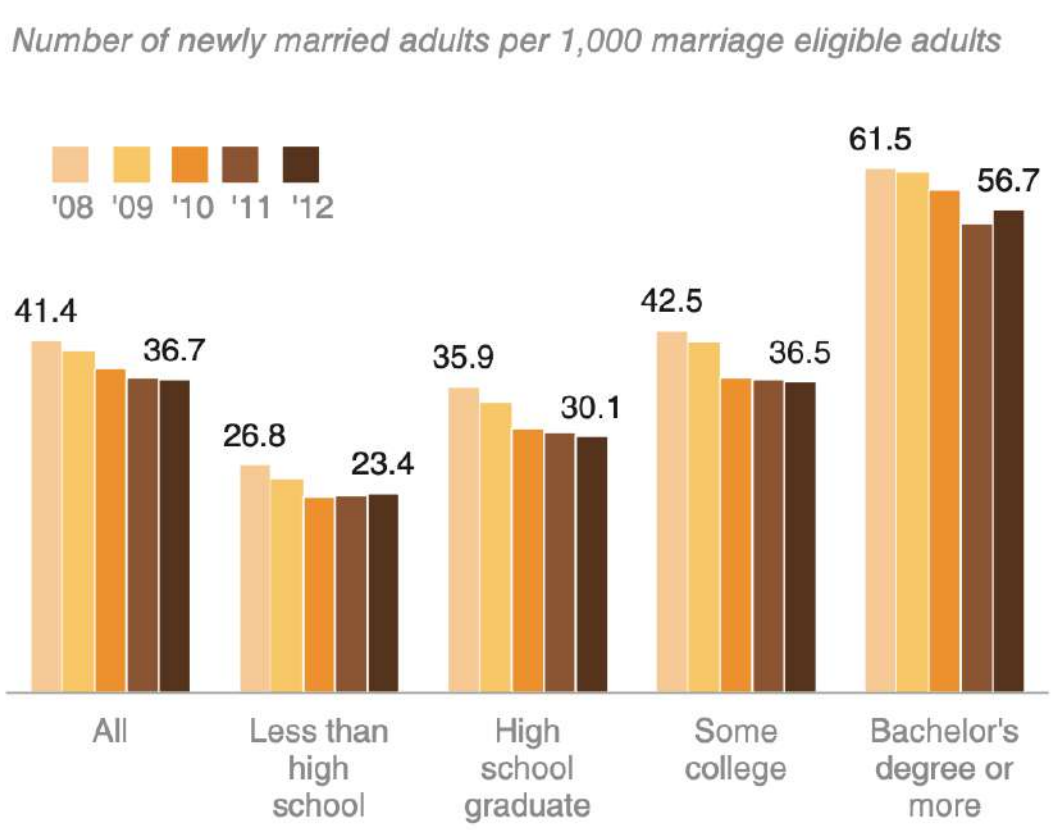


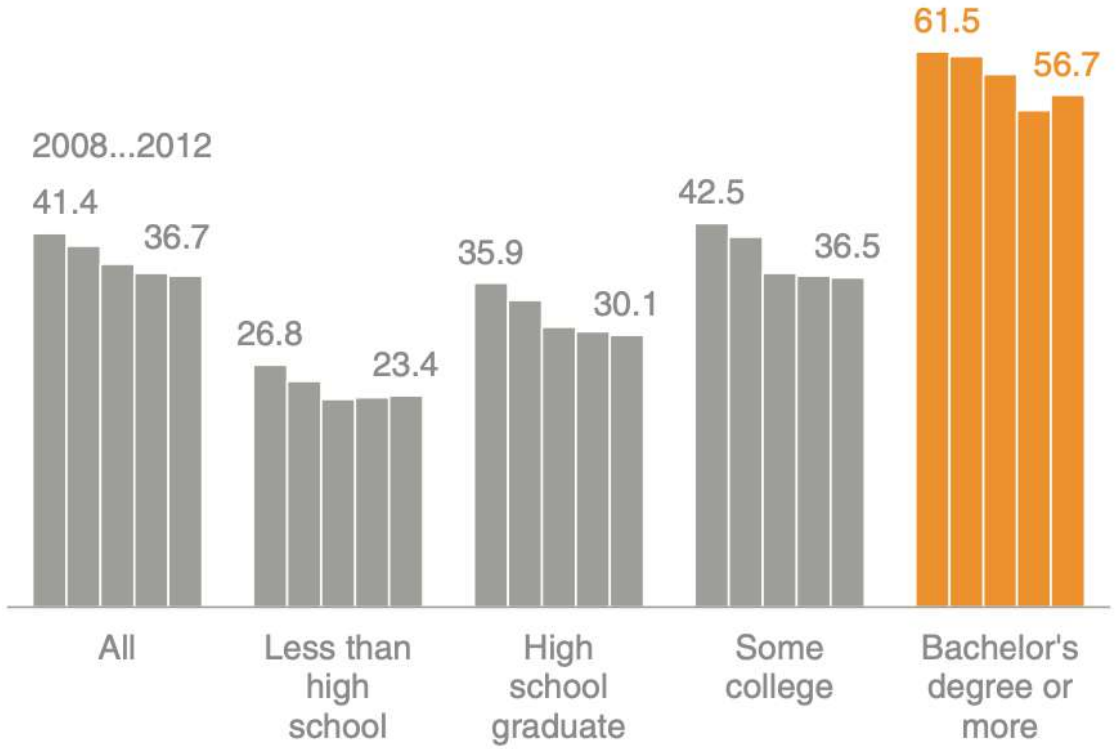


Antes y después

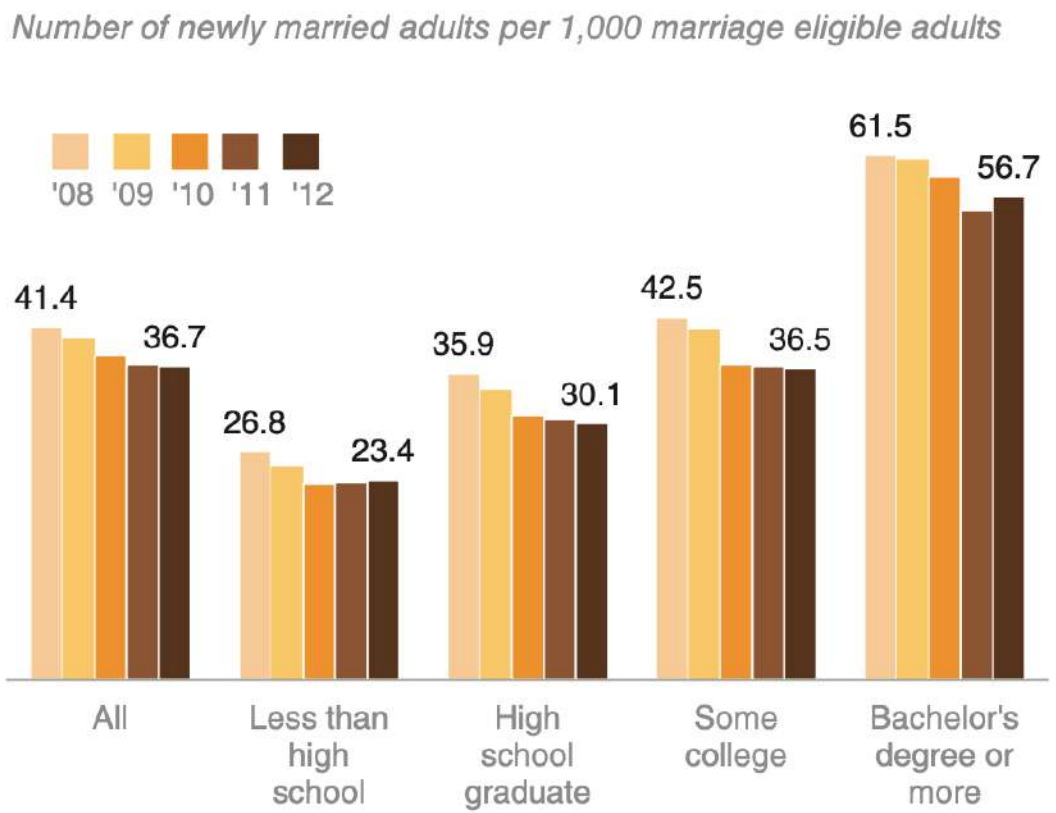


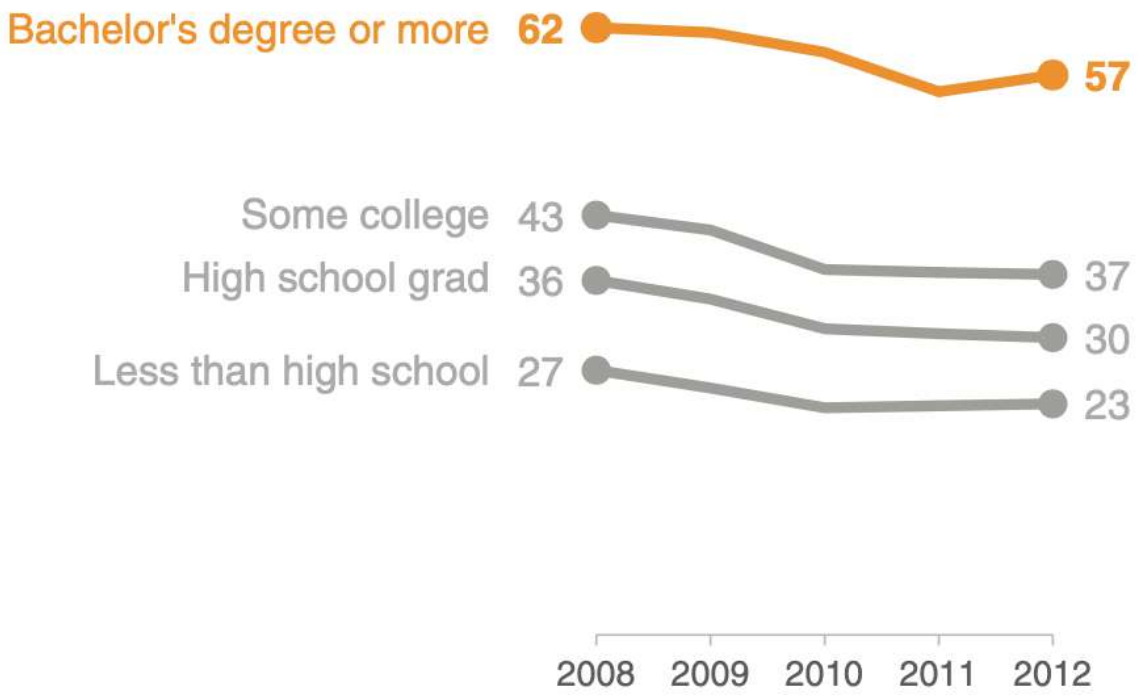
Antes y después





Antes y después, después





#7 Escribir un pie de figura informativo

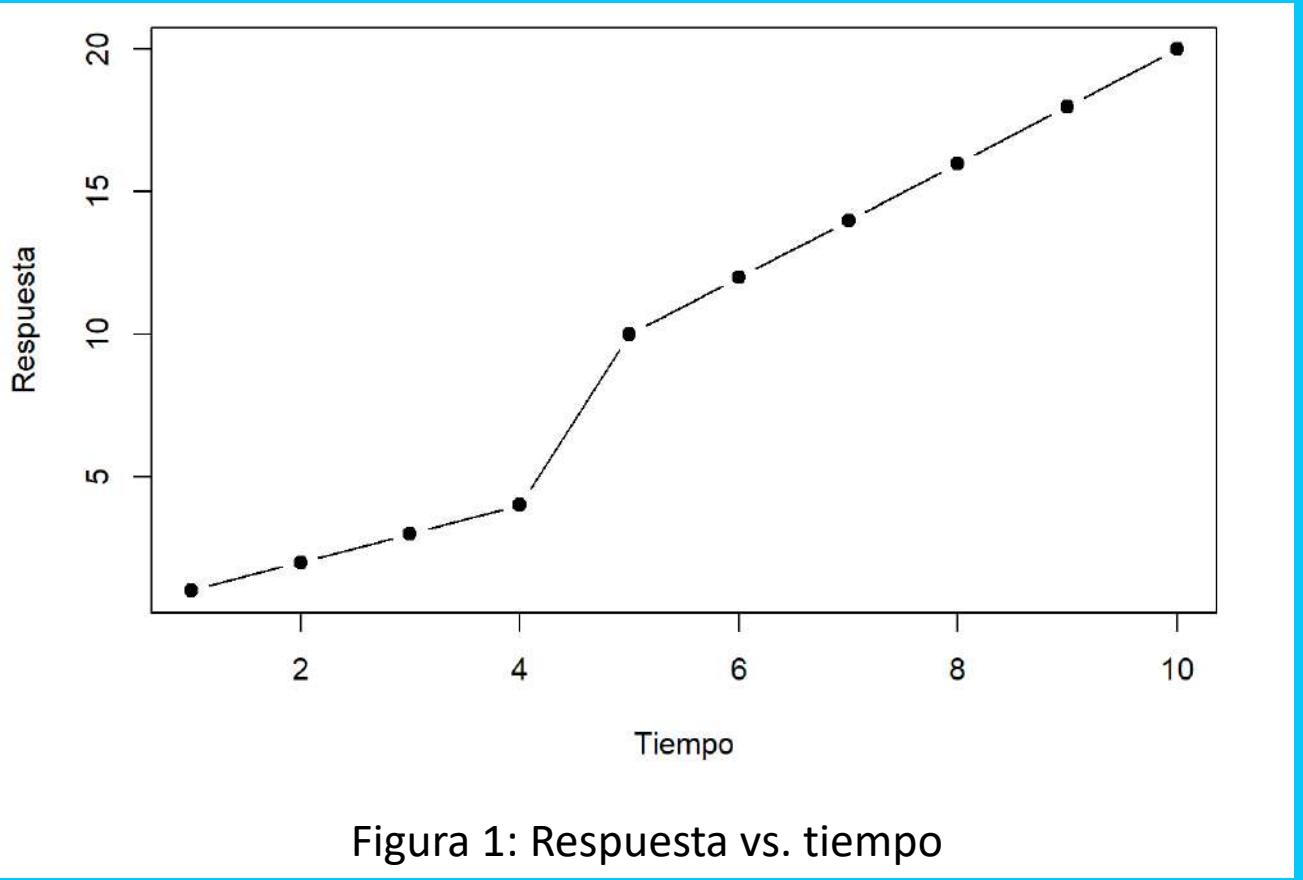


Figura 1: Respuesta vs. tiempo

Traditional

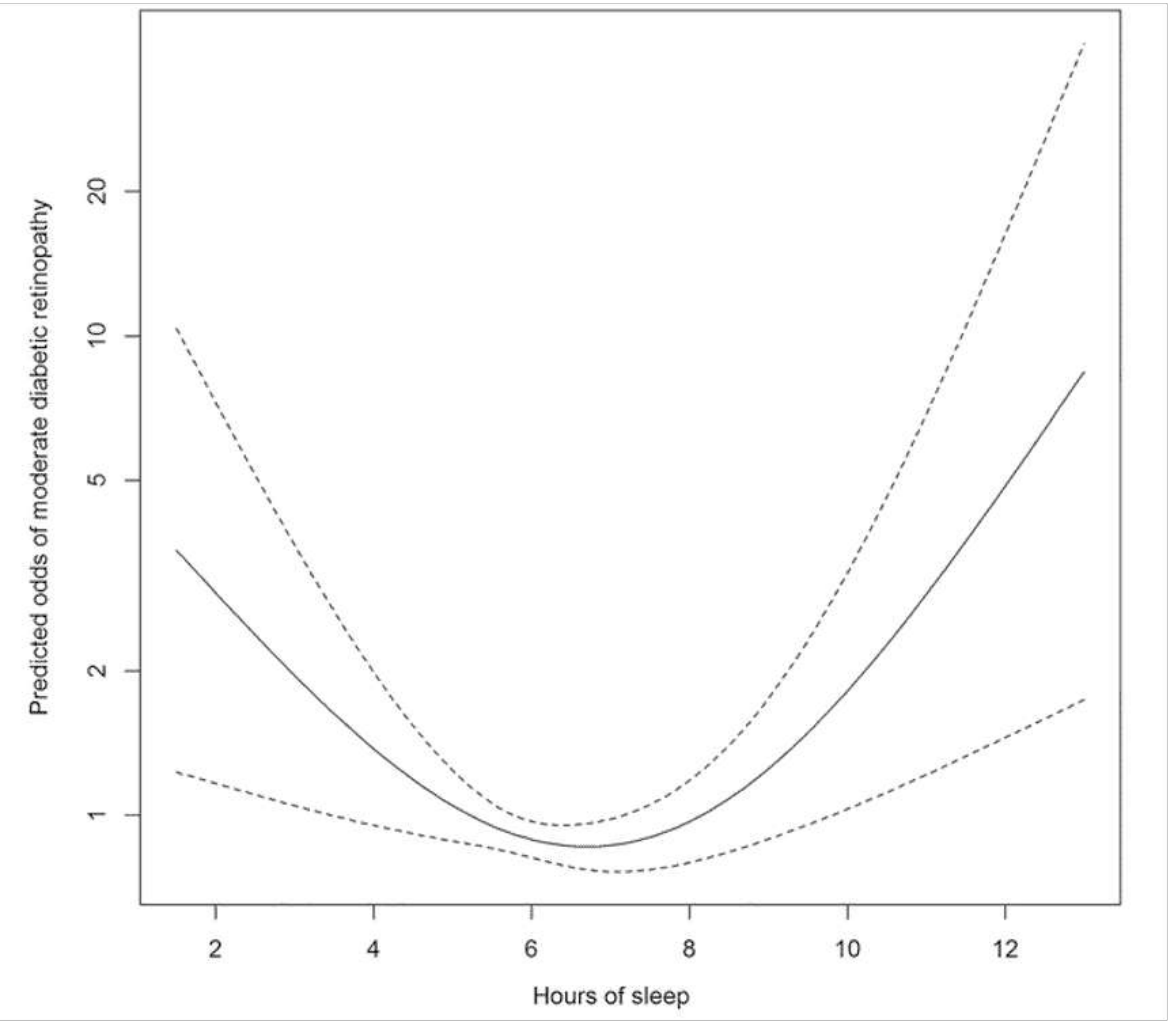


Fig. 1. Multivariable-adjusted odds of moderate diabetic retinopathy according to sleep duration.

VS.

New style

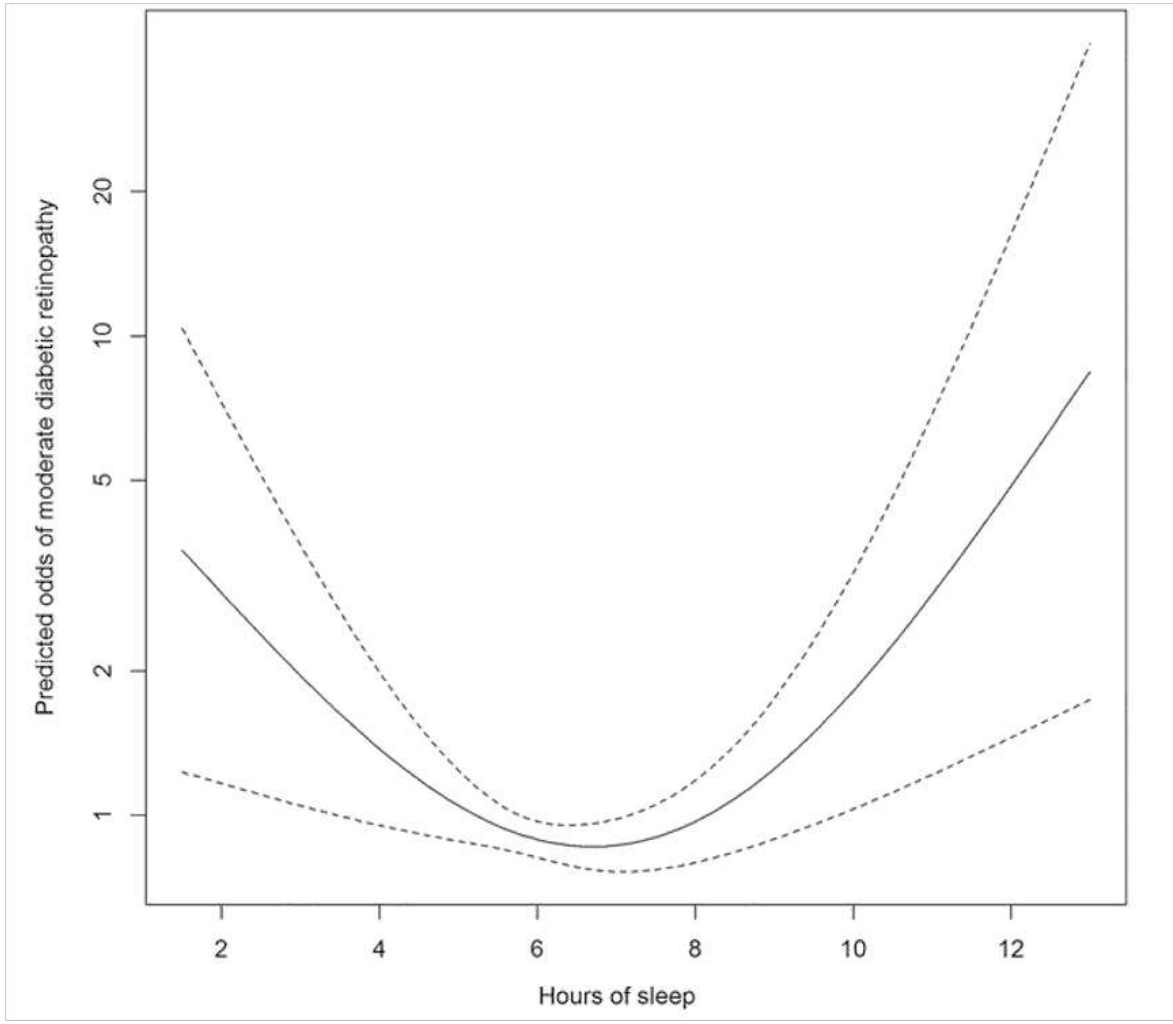
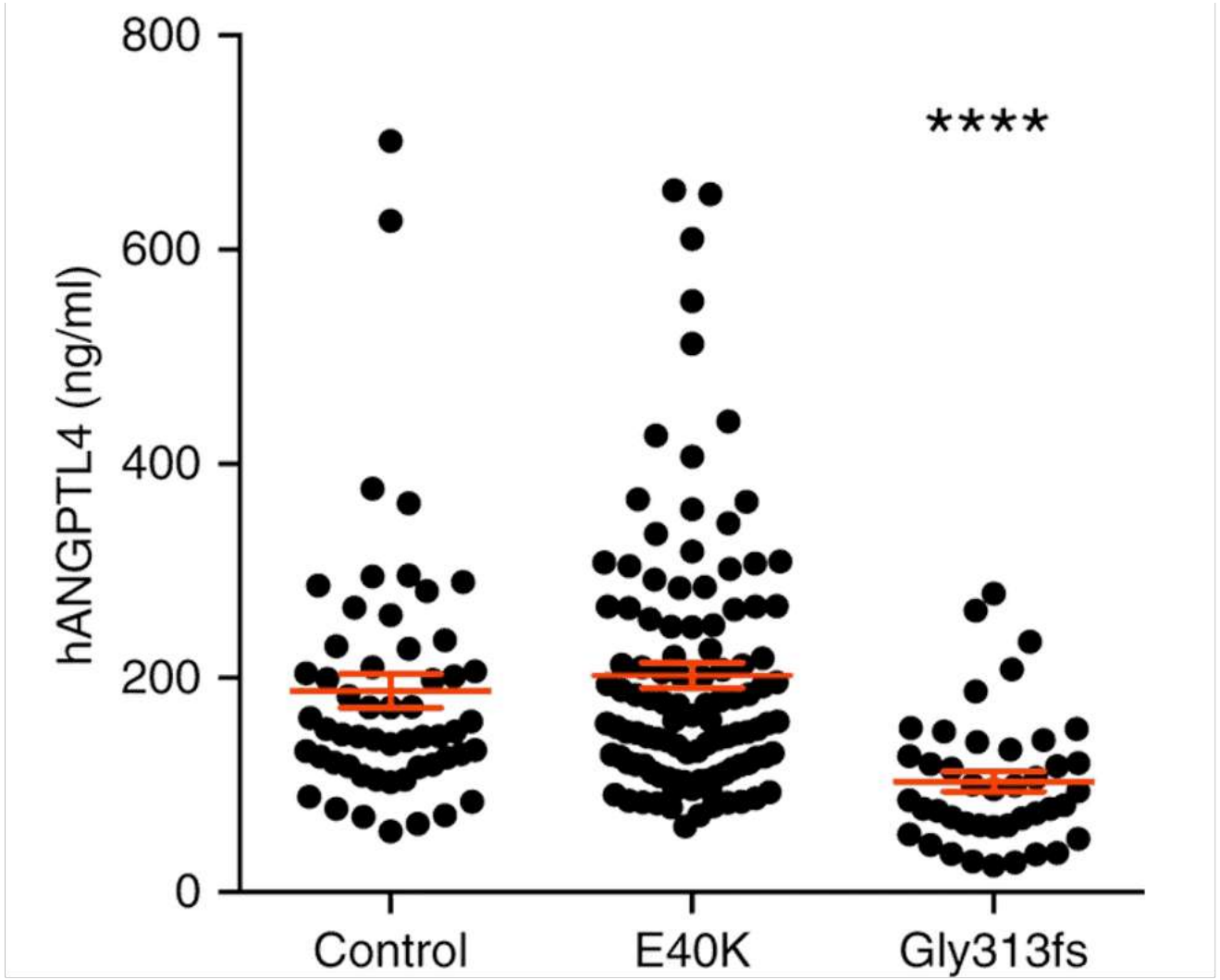


Fig. 1. A total sleep duration of 6–8 h per day was associated with the lowest risk of moderate diabetic retinopathy.



Declarative title → Plasma ANGPTL4 levels were reduced in p.G313fs carriers.

Methods → ANGPTL4 plasma levels were measured in fasted serum from 86 heterozygous p.E40K, 42 heterozygous p.G313fs variant carriers, and 55 controls matched for age, sex, and body mass index.

Statistical information → Statistics performed by unpaired *t*-test with Welch's correction, comparing each variant carriers group to controls, ****$p < 0.0001$

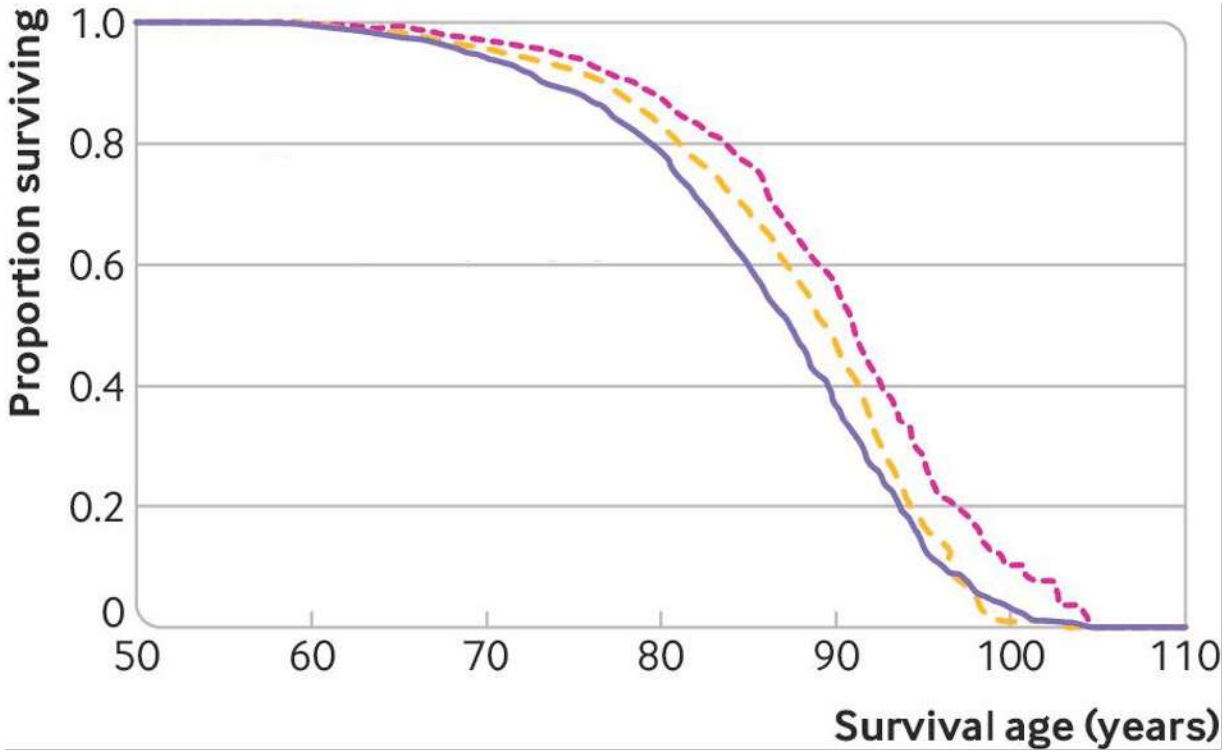


Fig. 1. Arts engagement had a protective association with longevity in older adults. Adjusted for demographic, socioeconomic, health related, behavioural, and social confounding factors. Solid blue line represents adults who never engaged with arts activities; yellow dashed line represents those who infrequently engaged with arts activities; pink dashed line represents those who frequently engaged with arts activities.

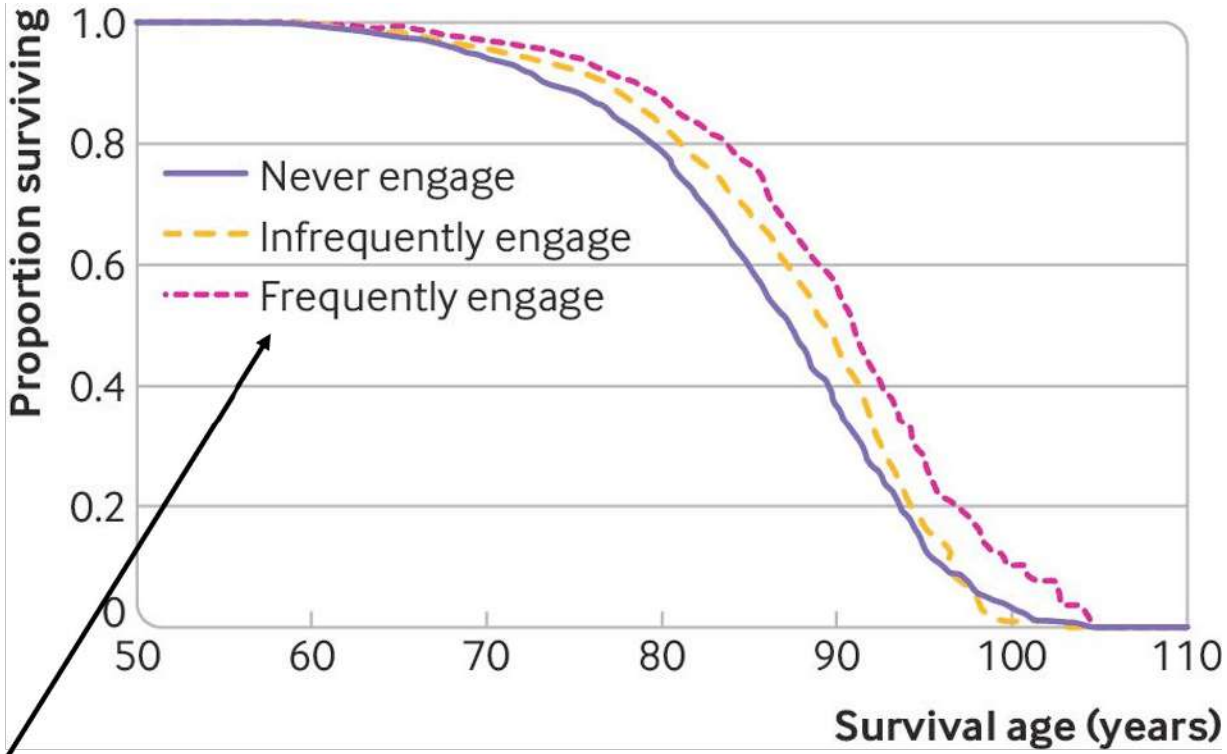
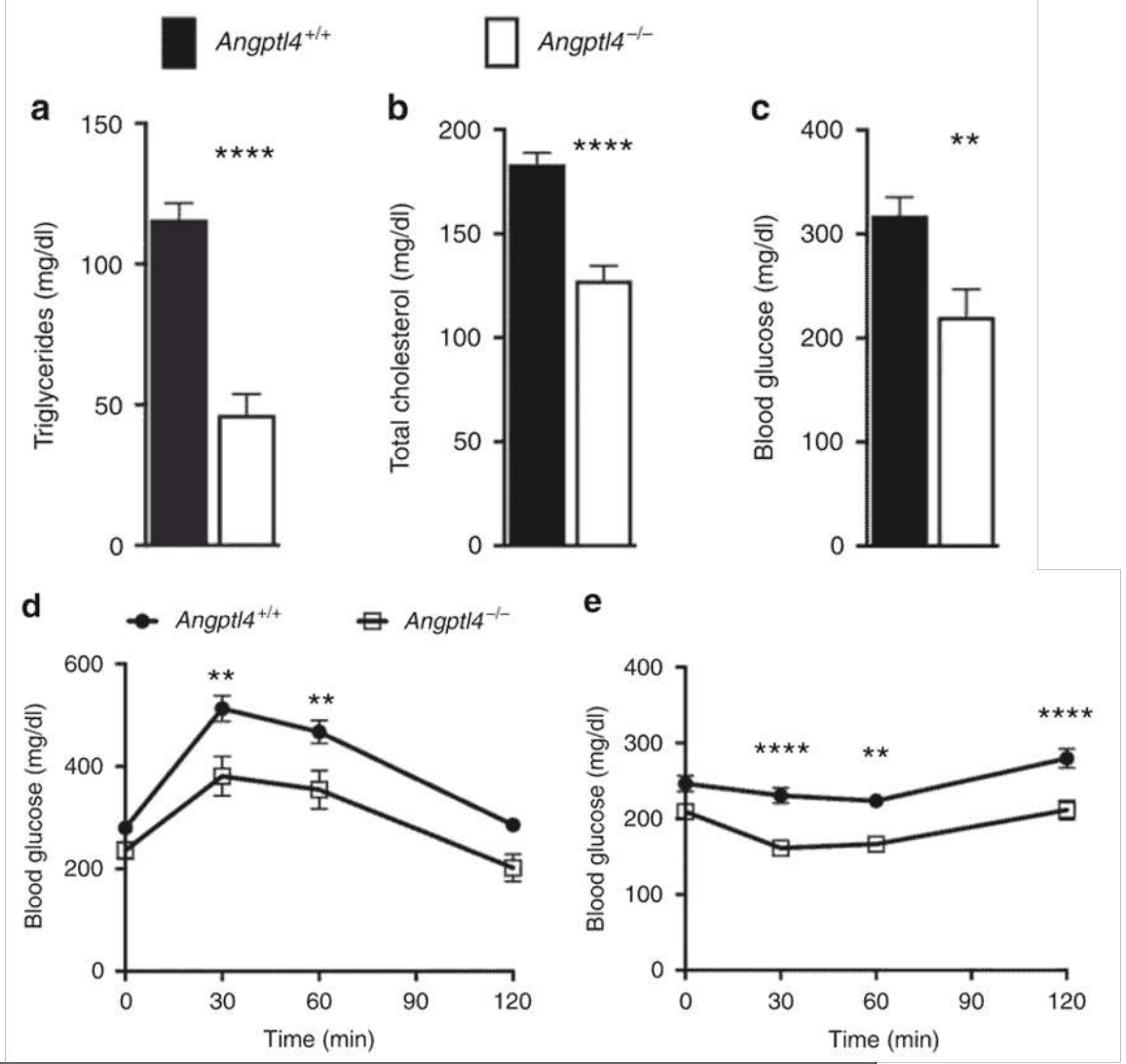


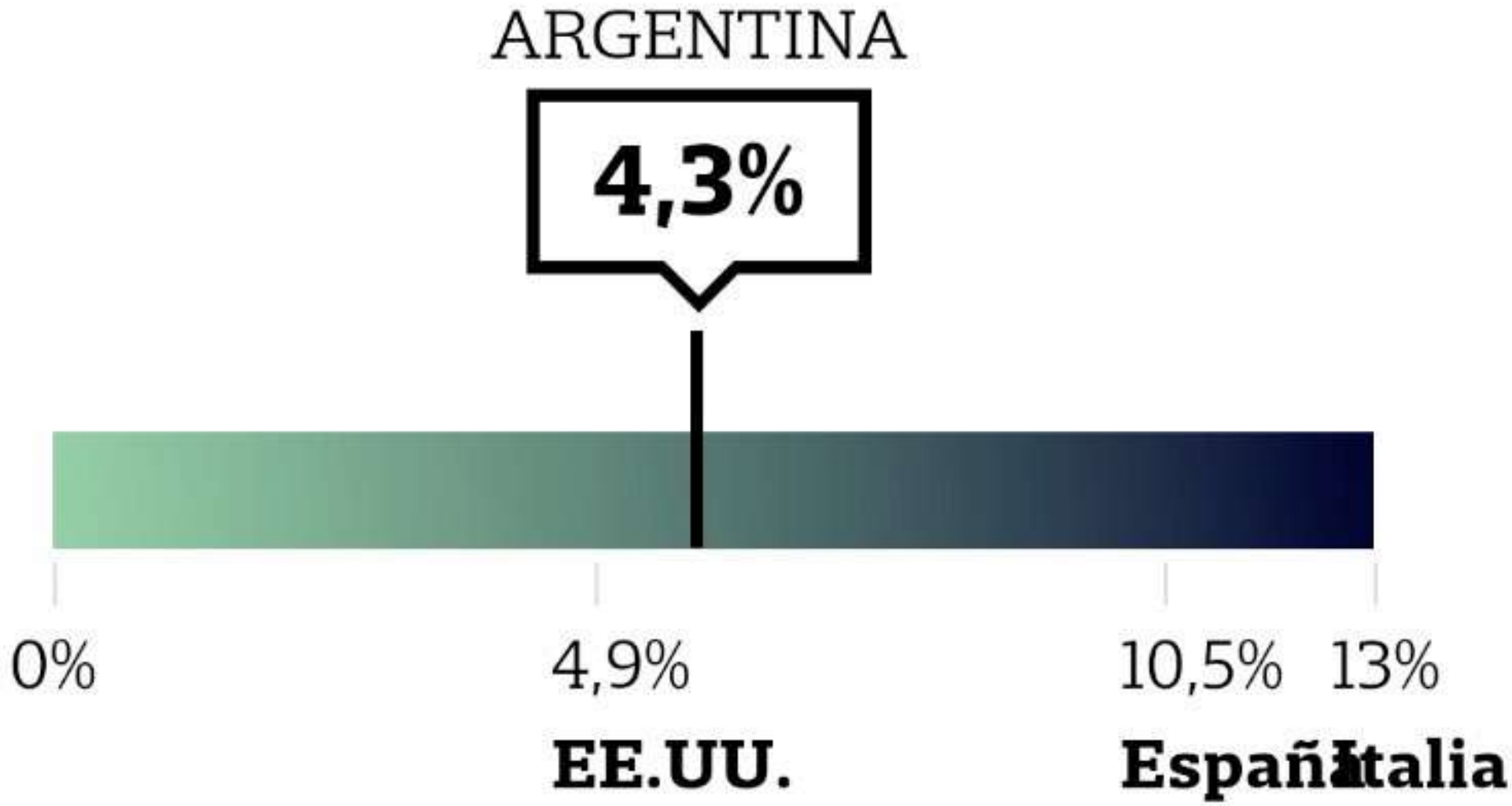
Fig. 1. Arts engagement had a protective association with longevity in older adults. Adjusted for demographic, socioeconomic, health related, behavioural, and social confounding factors.



Overarching declarative title →

Figure 2. *Angptl4*$^{-/-}$ mice had improved glucose homeostasis. (**a**) Serum triglycerides, (**b**) total cholesterol, and (**c**) blood glucose levels in *Angptl4*$^{-/-}$ and littermate control mice on a high-fat diet for 9 weeks. (**d**) Oral glucose tolerance test and (**e**) insulin tolerance test in the animals described in (**a**–**c**). All groups had 9–11 animals. Values are mean ± SEM. Statistical analysis by Welch's t-test (**a**) and 2-way ANOVA with Sidak's post-test (**d**, **e**), $^{**}p < 0.001$, $^{****}p < 0.0001$. The study was conducted in three different cohorts of mice, with qualitatively similar results in each replicate.

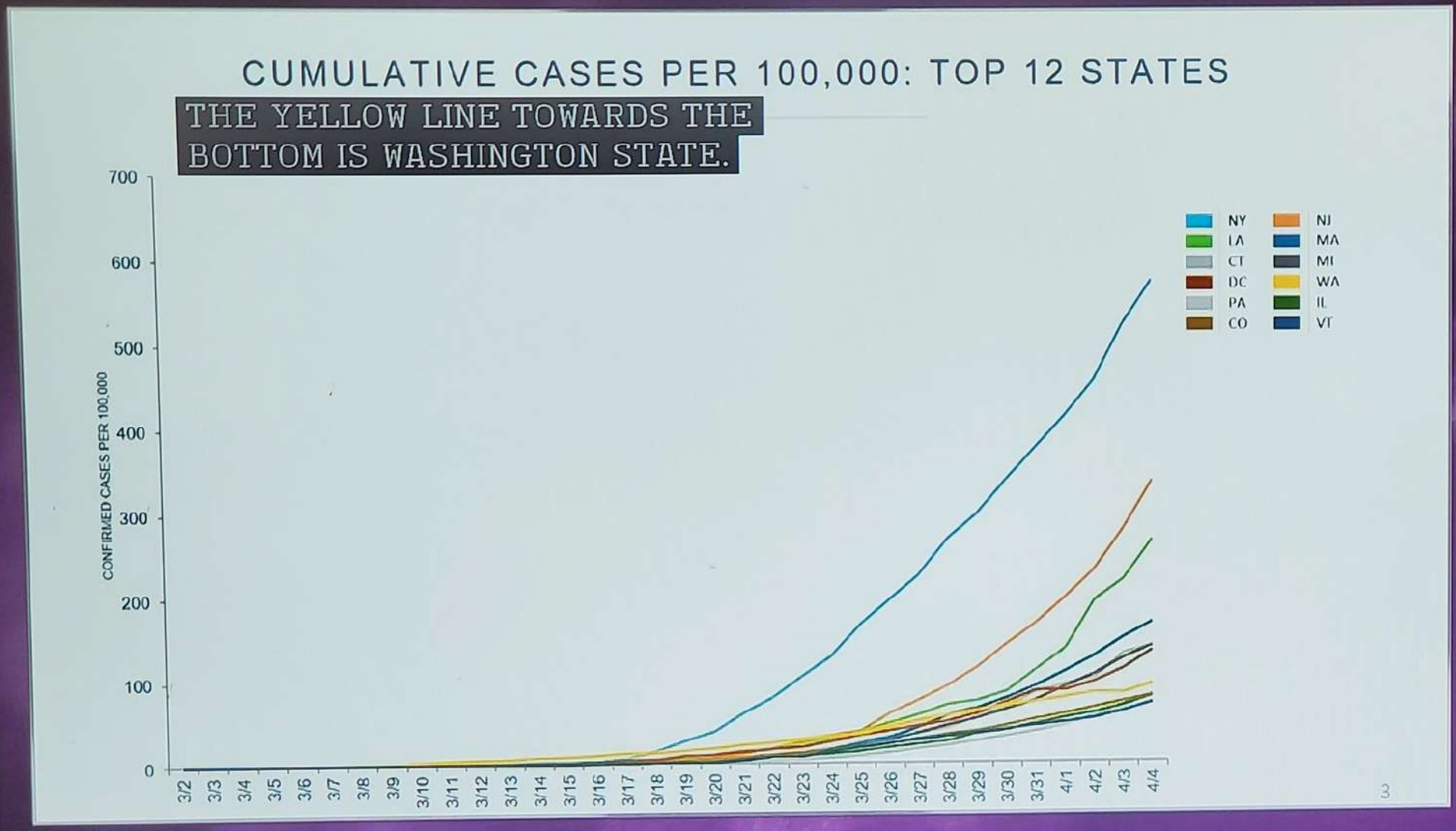
Más ejemplos


ARGENTINA
4,3%
0%
4,9%
10,5%
13%
EE.UU.
España
Italia

Crecimiento anual del PBI peruano
2001 - 2021

Digan lo que digan,
las cifras no mienten.


QuedateConC5N
CASOS ANUNCIADOS POR DIA
INFECTADOS
80
70
60
50
40
30
20
10
0
1
0
1
6
1
3
5
2
2
10
11
11
8
13
19
31
30
67



CORONAVIRUS PANDEMIC

GLOBALLY

TOTAL CASES
1,270,069
DEATHS
69,309

IN THE UNITED STATES

TOTAL CASES
335,524
DEATHS
9,562

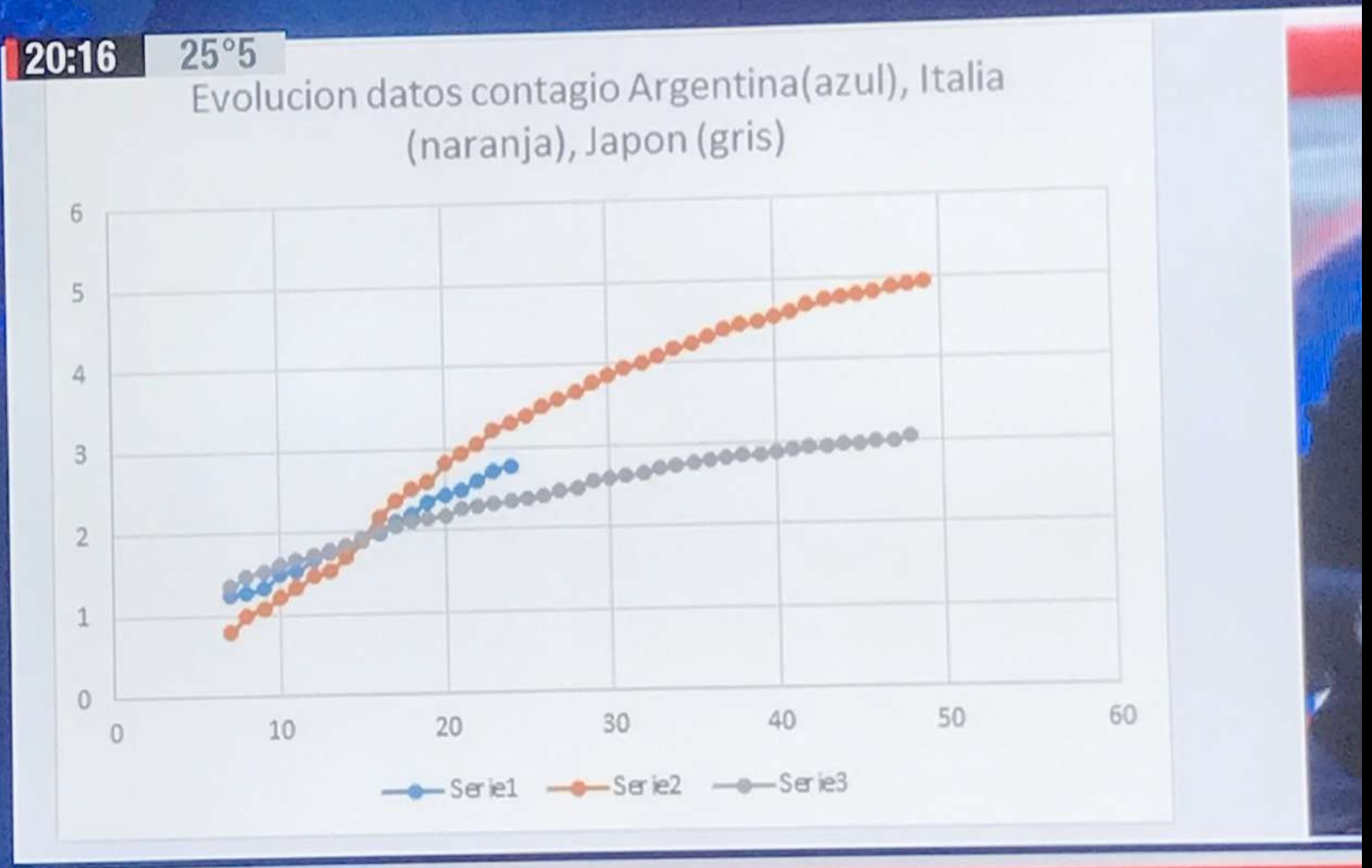
SOURCE: JOHNS HOPKINS UNIVERSITY

LIVE
CNN
7:20 PM ET

)N ADMITTED TO HOSPITAL, DOWNING STREET SAYS ▶ JOHNSON TESTED

SITUATION ROOM

20:16 25°5

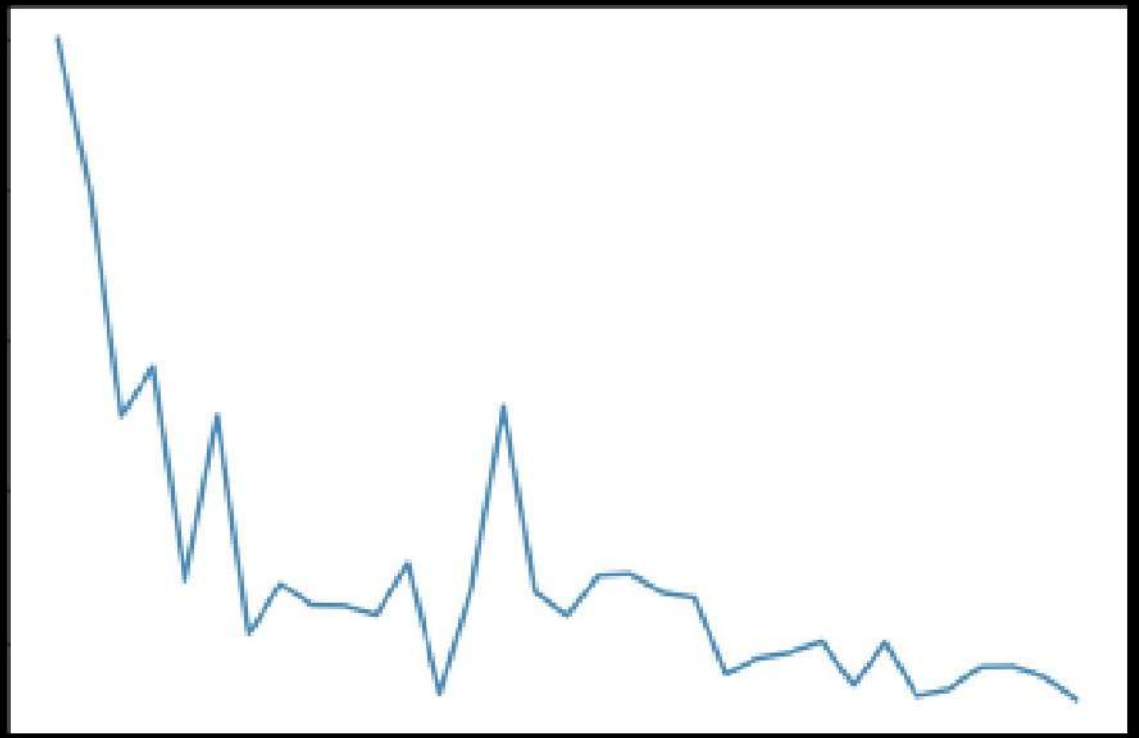


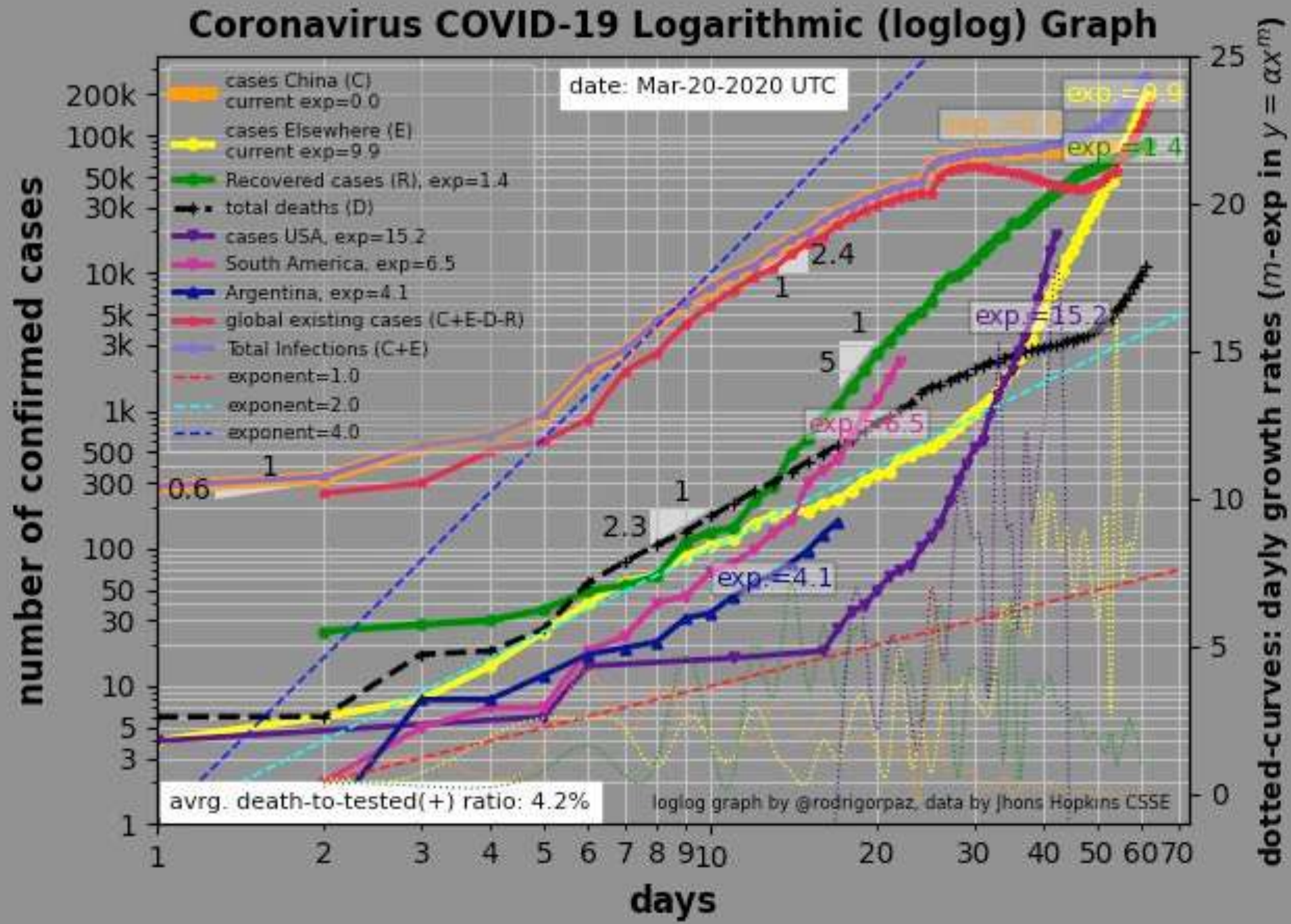
CIFRAS GLOBALES: 640.589 INFECTADOS Y

TODAS LAS DUDAS ACERCA DEL CORO

CORONAVIRUS | ALEMANIA | INFECTADOS: 53340 | FALLECIDOS: 399 | EE.UU | INFEC

Este gráfico indica el porcentaje de infectados NUEVOS de #COVID19 cada día sobre el TOTAL de infectados en España
#sePuede




Coronavirus COVID-19 Logarithmic (loglog) Graph
date: Mar-20-2020 UTC
cases China (C)
current exp=0.0
cases Elsewhere (E)
current exp=9.9
Recovered cases (R), exp=1.4
total deaths (D)
cases USA, exp=15.2
South America, exp=6.5
Argentina, exp=4.1
global existing cases (C+E-D-R)
Total Infections (C+E)
exponent=1.0
exponent=2.0
exponent=4.0
exp.=0.9
exp.=1.4
exp.=15.2
exp.=6.5
exp.=4.1
0.6
1
2.4
1
1
5
2.3
1
number of confirmed cases
dotted-curves: dayly growth rates (m-exp in y = αx^m)
days
avrg. death-to-tested(+) ratio: 4.2%
loglog graph by @rodrigorpaz, data by Jhons Hopkins CSSE

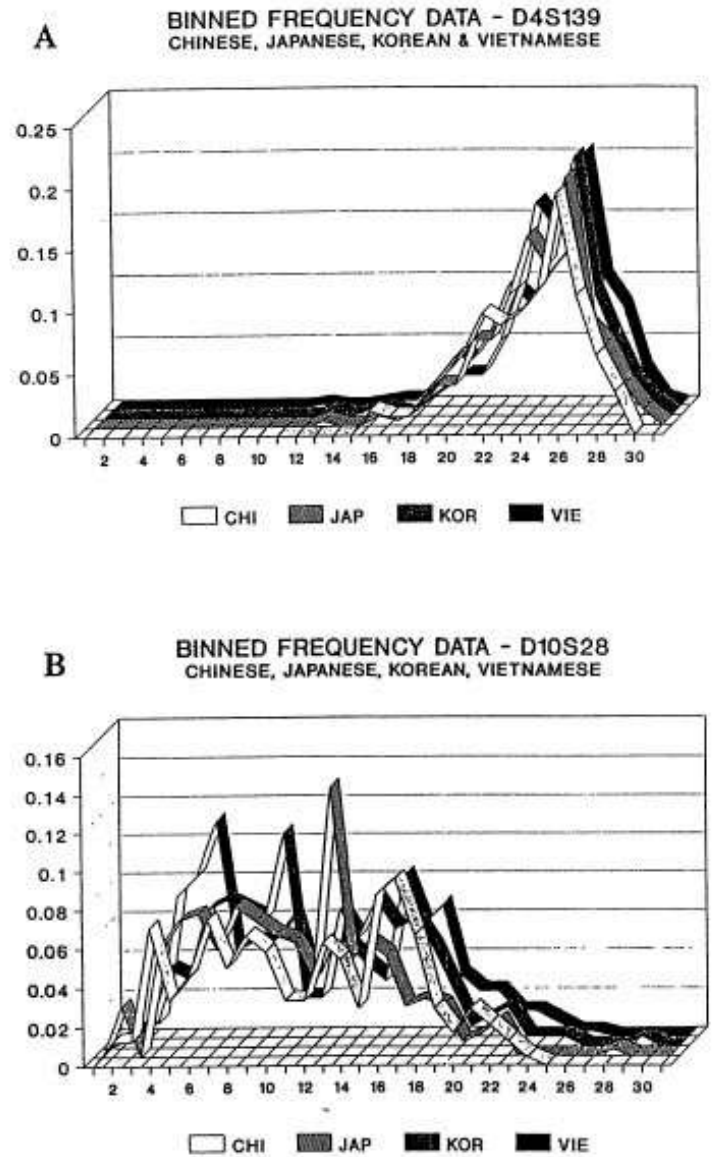


FIG. 4. *Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are 103, 125, 93 and 215 for D4S139 and 120, 137, 100 and 193 for D10S28. The horizontal axis is the bin number; bins are not of equal length.*

Roeder K (1994)
DNA fingerprinting: A review of the controversy (with discussion).
Statistical Science 9:222-278, Figure 4

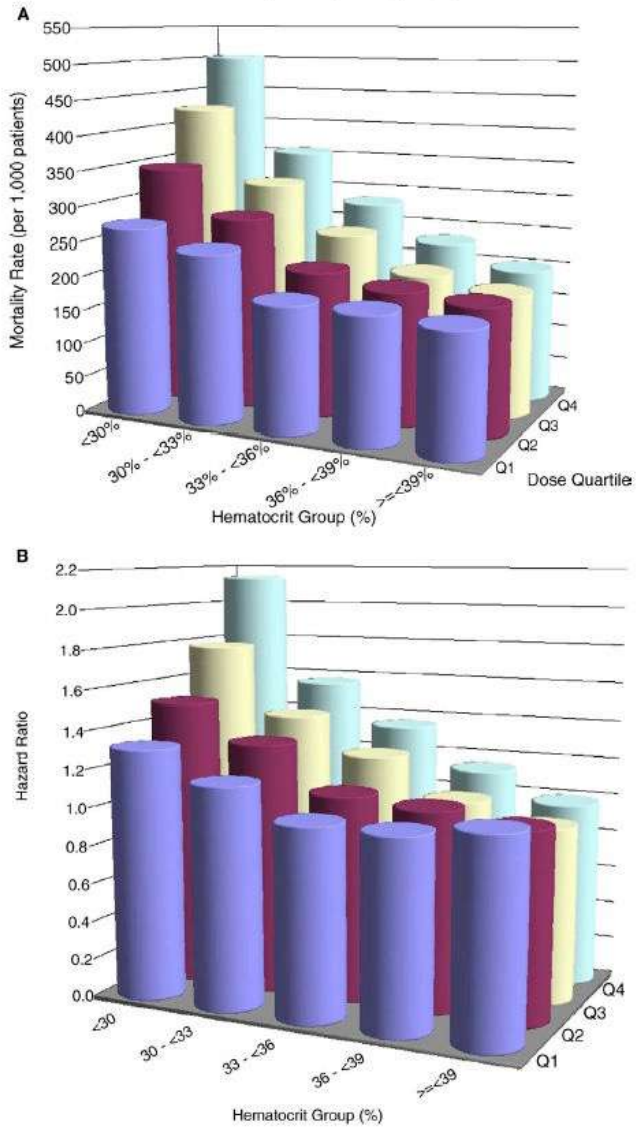


Fig. 2. (A) Unadjusted 1-year mortality rates by hematocrit group disaggregated by epoetin dose quartile. Within each epoetin dose quartile, there is a trend toward increasing mortality as the observed study hematocrit decreases, most notably in the fourth quartile (>21,692 units/wk). Similarly, there is a trend toward increasing mortality as the epoetin dose increases within each observed study hematocrit range, most notably in the lowest (<30%) hematocrit range. (B) Relative risk of death by hematocrit group disaggregated by epoetin dose quartile. For the three lowest observed study hematocrit ranges, compared with the reference group, there is a trend toward higher relative risk of mortality within each hematocrit range as the epoetin dose increases and within each dose quartile as the hematocrit range decreases. For the two highest hematocrit ranges, compared with the reference group, the relative risk of mortality varies, depending on the specific hematocrit range and dose quartile.

Cotter DJ, et al. (2004)
Hematocrit was not validated as a surrogate endpoint for survival amoung epoetin-treated hemodialysis patients.
Journal of Clinical Epidemiology 57:1086-1095, Figure 2

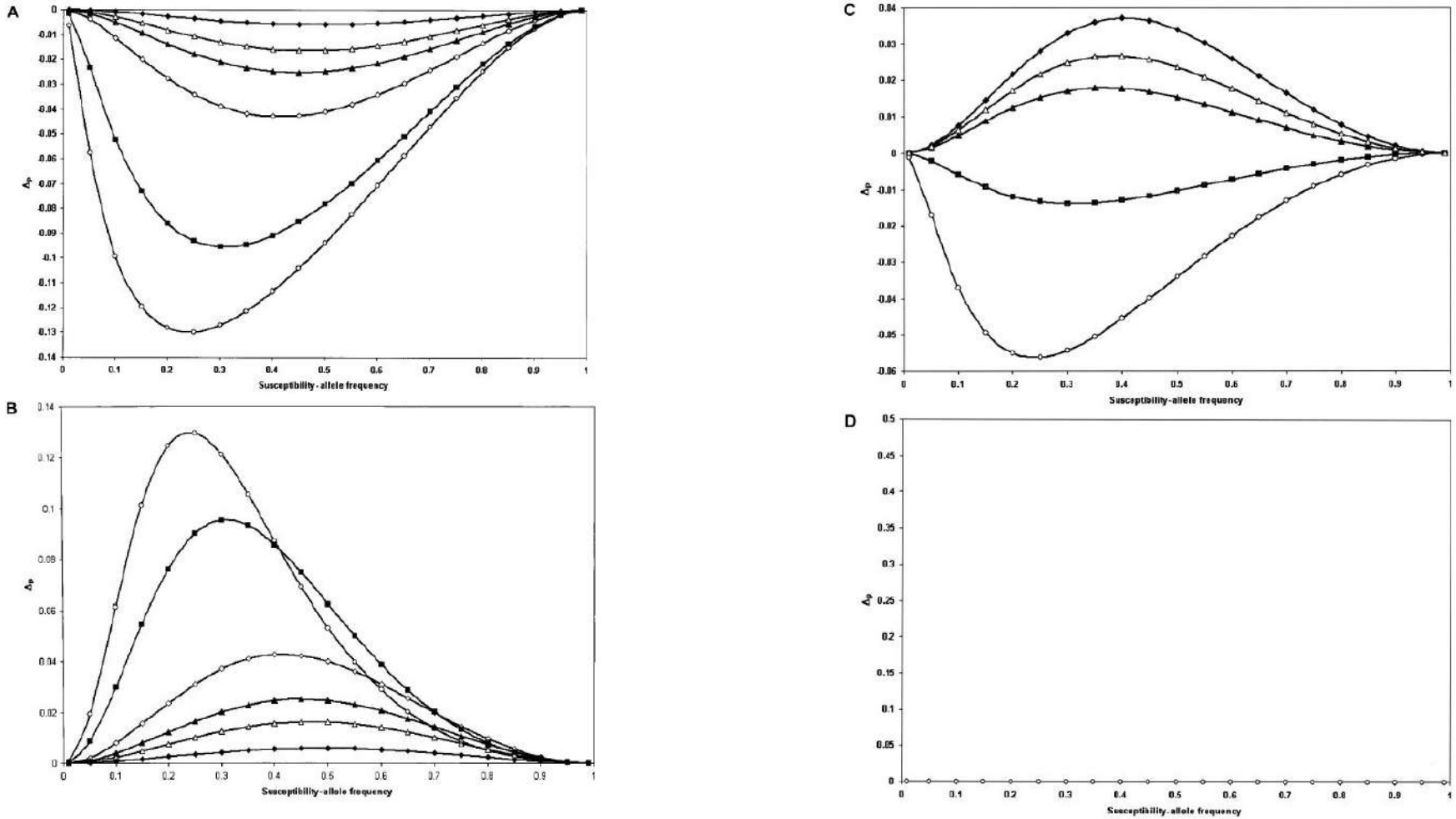


Figure 1 Δ_p plotted versus the susceptibility-allele frequency for patients. *A*, *B*, and *D*, Data points are as follows: $\gamma = 1.1$ (*blackened diamonds*), $\gamma = 1.3$ (*unblackened triangles*), $\gamma = 1.5$ (*blackened triangles*), $\gamma = 2$ (*unblackened diamonds*), $\gamma = 5$ (*blackened squares*), and $\gamma = 10$ (*unblackened circles*). *A*, Dominant model. *B*, Recessive model. *C*, Additive model. Since $\gamma < 2$ would not satisfy our definition of an additive model as $\gamma = 2\beta$ and $\beta > 1$, the data points in *C* are as follows: $\gamma = 2.2$ ($\beta = 1.1$) (*blackened diamonds*), $\gamma = 2.6$ ($\beta = 1.3$) (*unblackened triangles*), $\gamma = 3$ ($\beta = 1.5$) (*blackened triangles*), $\gamma = 5$ (*blackened squares*), $\gamma = 2$ (*unblackened diamonds*). *D*, Multiplicative model.

Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005)
Rational inferences about departures from Hardy-Weinberg equilibrium.
American Journal of Human Genetics 76:967-986, Figure 1

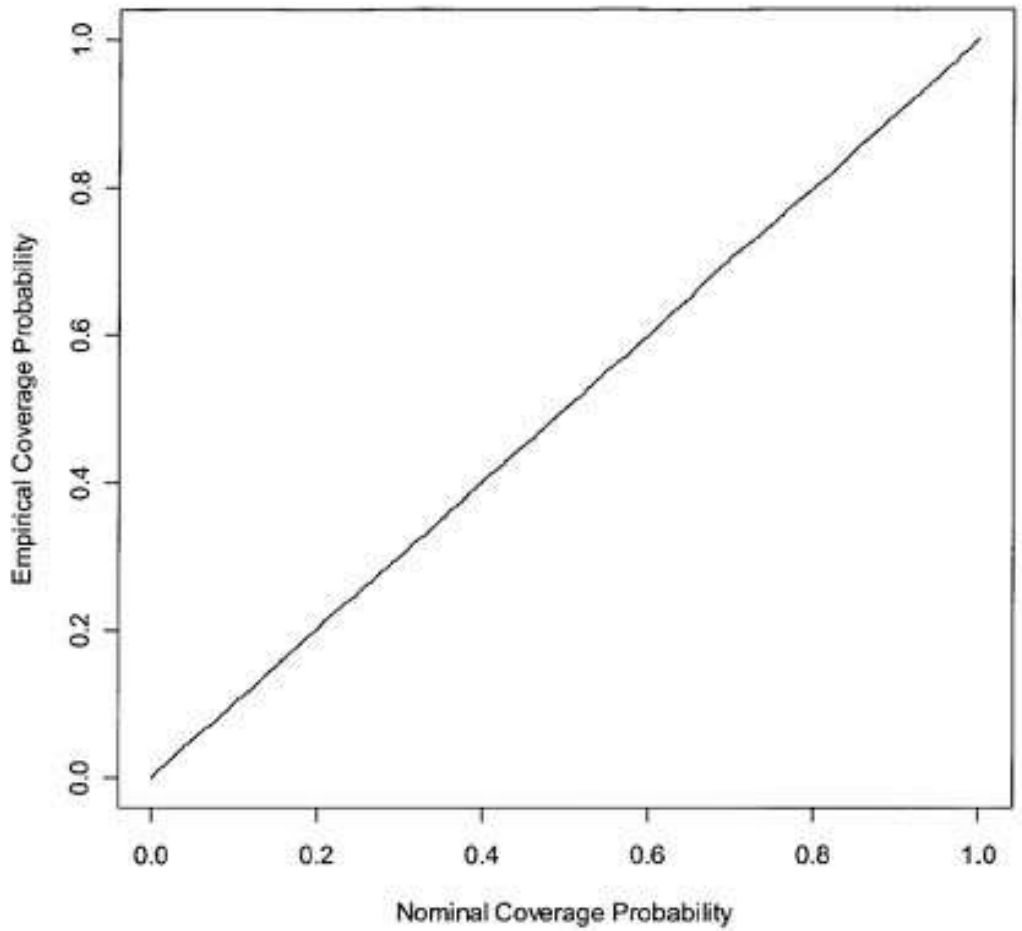
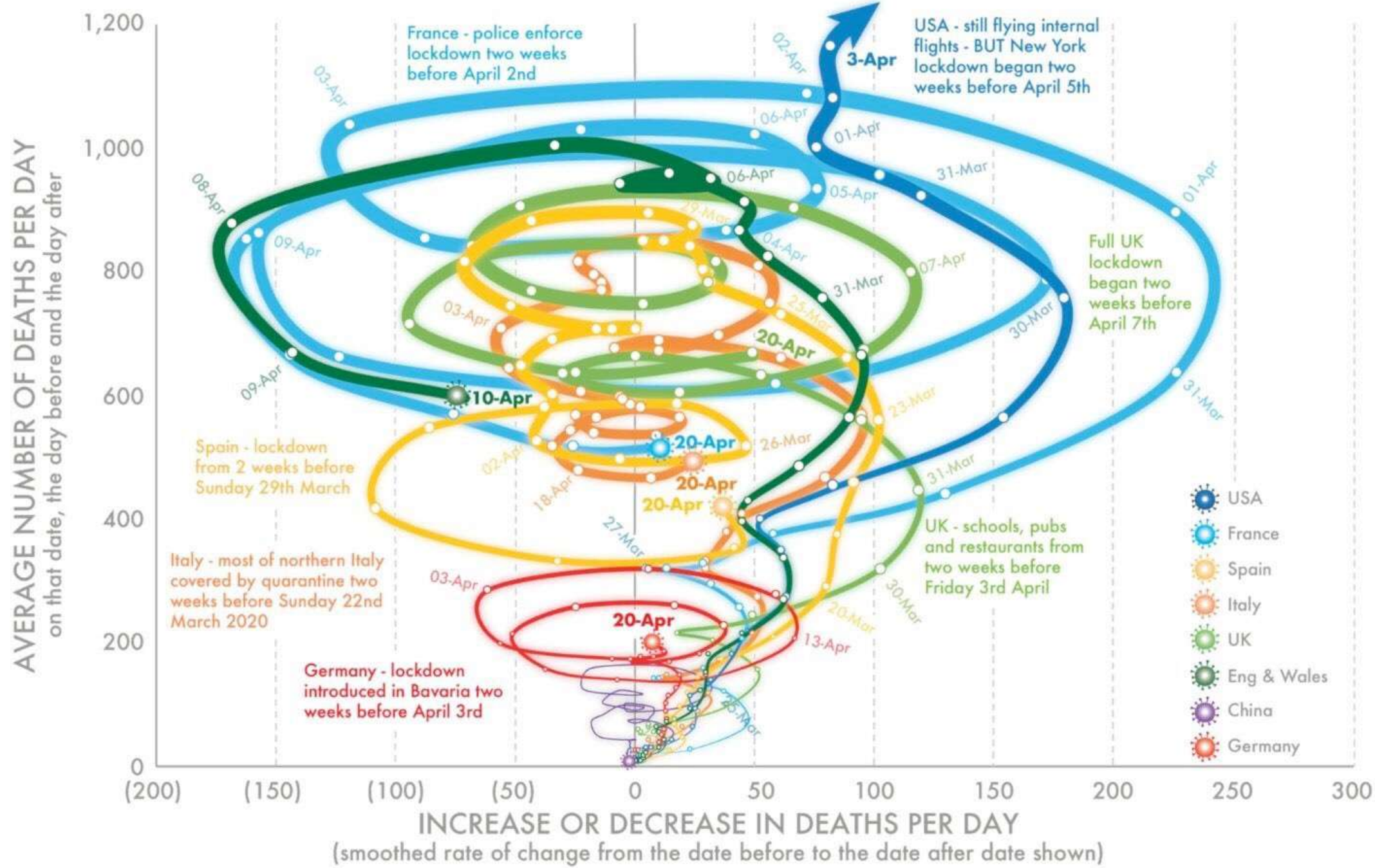


Figure 1 Empirical coverage of CIs for the relative-risk parameter β of haplotype 01100. Results are based on 10,000 simulated data sets with the same haplotype frequencies as the FUSION data. Haplotype 01100 has a multiplicative effect on disease risk, with $\beta = 0.35$.

Epstein MP, Satten GA (2003)
Inference on haplotype effects in case-control studies using unphased genotype data.
American Journal of Human Genetics 73:1316-1329, Figure 1



DannyDorling.org. Illustration by Kirsten McClure @orpheuscat

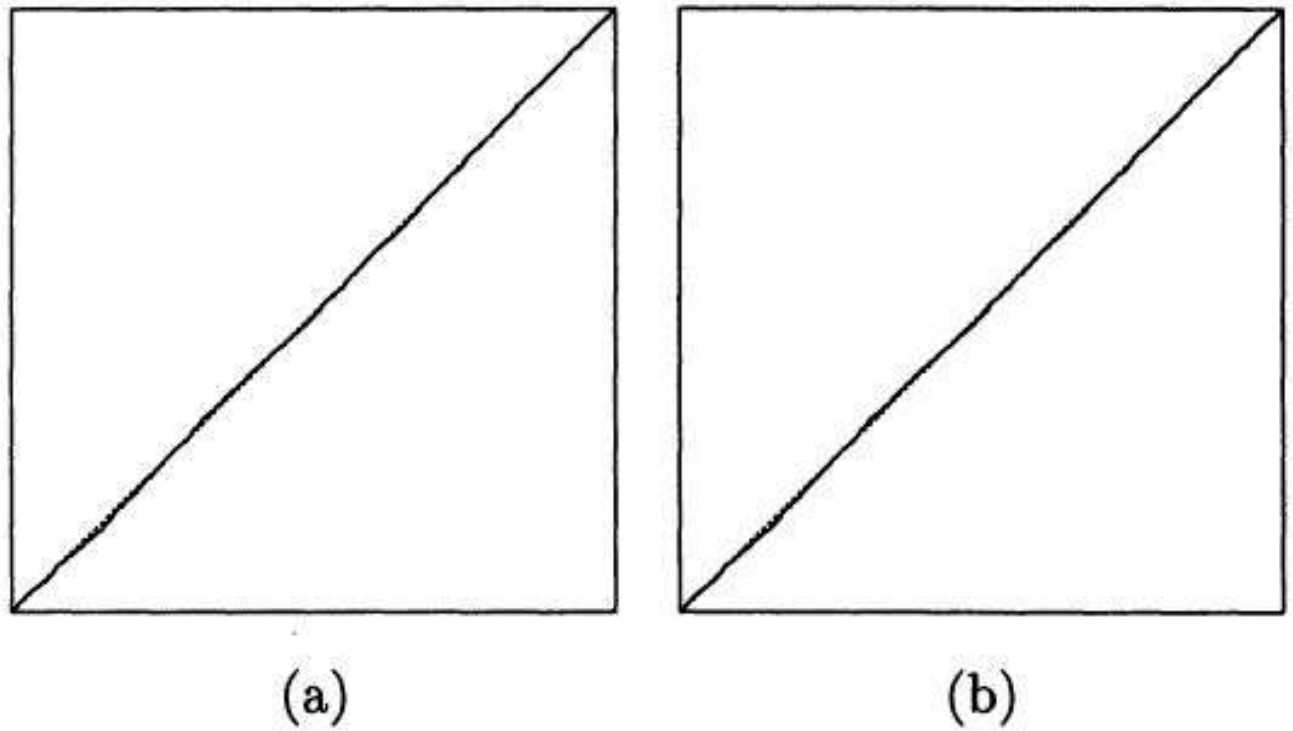


Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

Mykland P, Tierney L, Yu B (1995)
Regeneration in Markov chain samplers.
Journal of the American Statistical Association 90:233-241, Figure 1

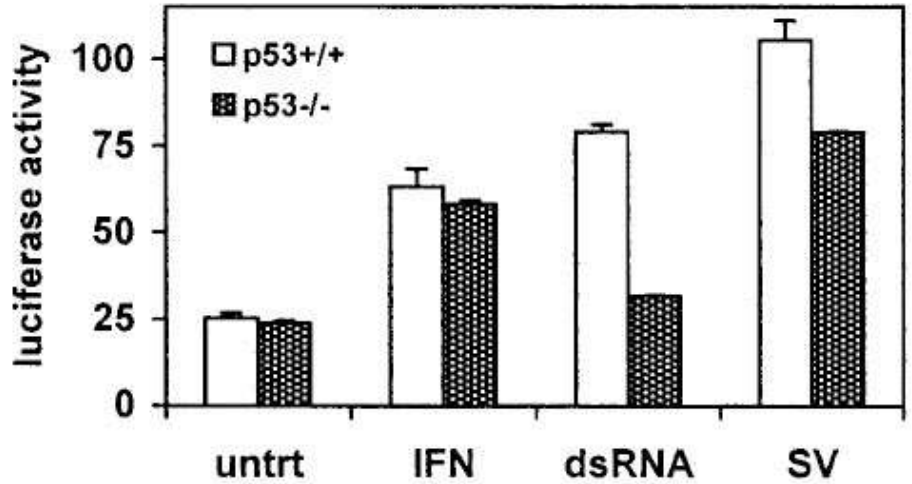


FIG. 4. ISG15 promoter activity mimics endogenous ISG15 mRNA regulation by p53, dsRNA, and virus. Cells (6×10^5 HCT 116) were seeded in 32-mm plates and allowed to attach overnight. Cells were transfected with 500 ng of pGL3/ISG15-Luc, 50 ng of pRL null (Promega), and 450 ng of pcDNA3 for carrier DNA by using Lipofectamine Plus (Life Technologies) following the manufacturer's instructions. Twenty-four hours posttransfection, the medium was aspirated and replaced with medium containing either 1,000 U of IFN-α/ml, 50 μg of dsRNA/ml, or Sendai virus (multiplicity of infection, 10). Cells were incubated for 12 h and then lysed, and luciferase assays were performed. Luciferase activity was assessed on 20 μl of each lysate as directed by the supplier (Dual Luciferase Kit, Promega) using a TD 20/20 luminometer (Turner Designs). Luciferase activity is presented as the ratio of firefly activity to renilla activity to control for differences in transfection efficiency. Each data point is the mean of triplicate samples ± the standard error; the data presented are representative of four independent experiments.

Hummer BT, Li XL, Hassel BA (2001)
Role for p53 in gene induction by double-stranded RNA.
J Virol 75:7774-7777, Figure 4

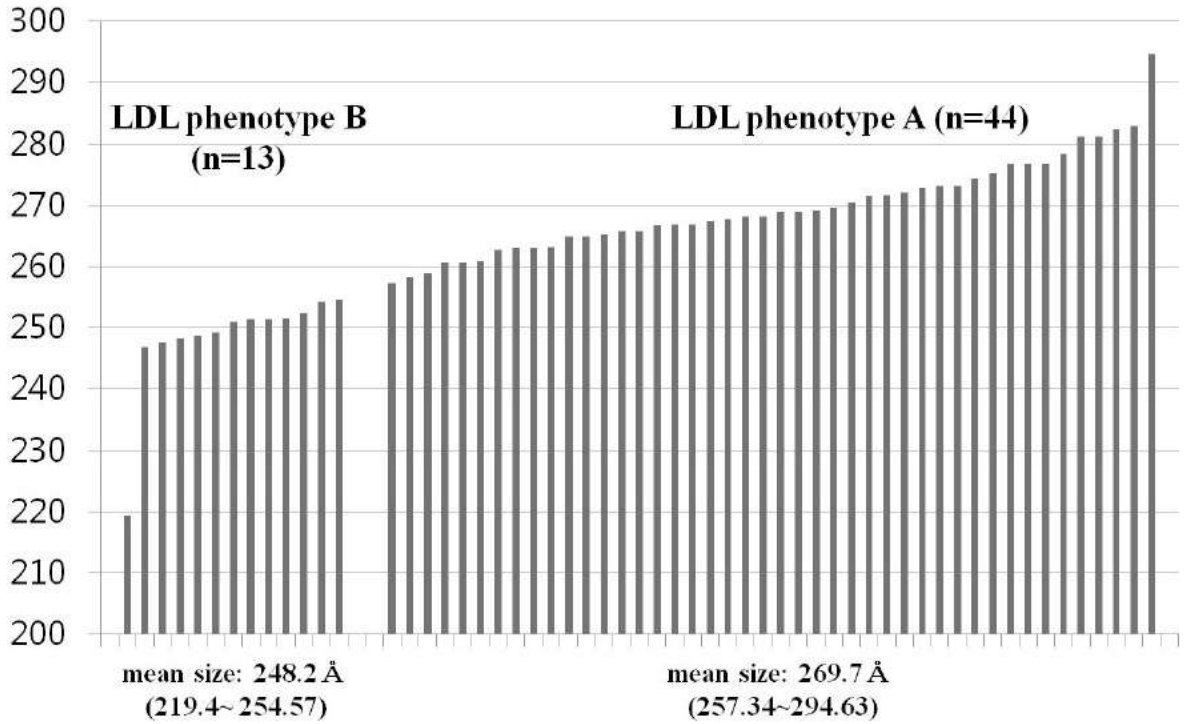


Fig. 1. Distribution of low-density lipoprotein (LDL) particle size in all study subjects (LDL phenotypes A and B). *LDL phenotype A group* (mean size: 269.7 Å, n = 44), subjects with buoyant-mode profiles [peak LDL particle diameter $\geq$ 264 Å] including intermediate LDL subclass pattern [256 Å $\leq$ peak LDL particle diameter $\leq$ 263 Å]; *LDL phenotype B group* (mean size: 248.2 Å, n = 13), subjects with dense-mode profiles [peak LDL particle diameter $\leq$ 255 Å]

Kim OY, et al. (2012)
Higher levels of serum triglyceride and dietary carbohydrate intake are associated with smaller LDL particle size in healthy Korean women.
Nutrition Research and Practice 6:120-125, Figure 1

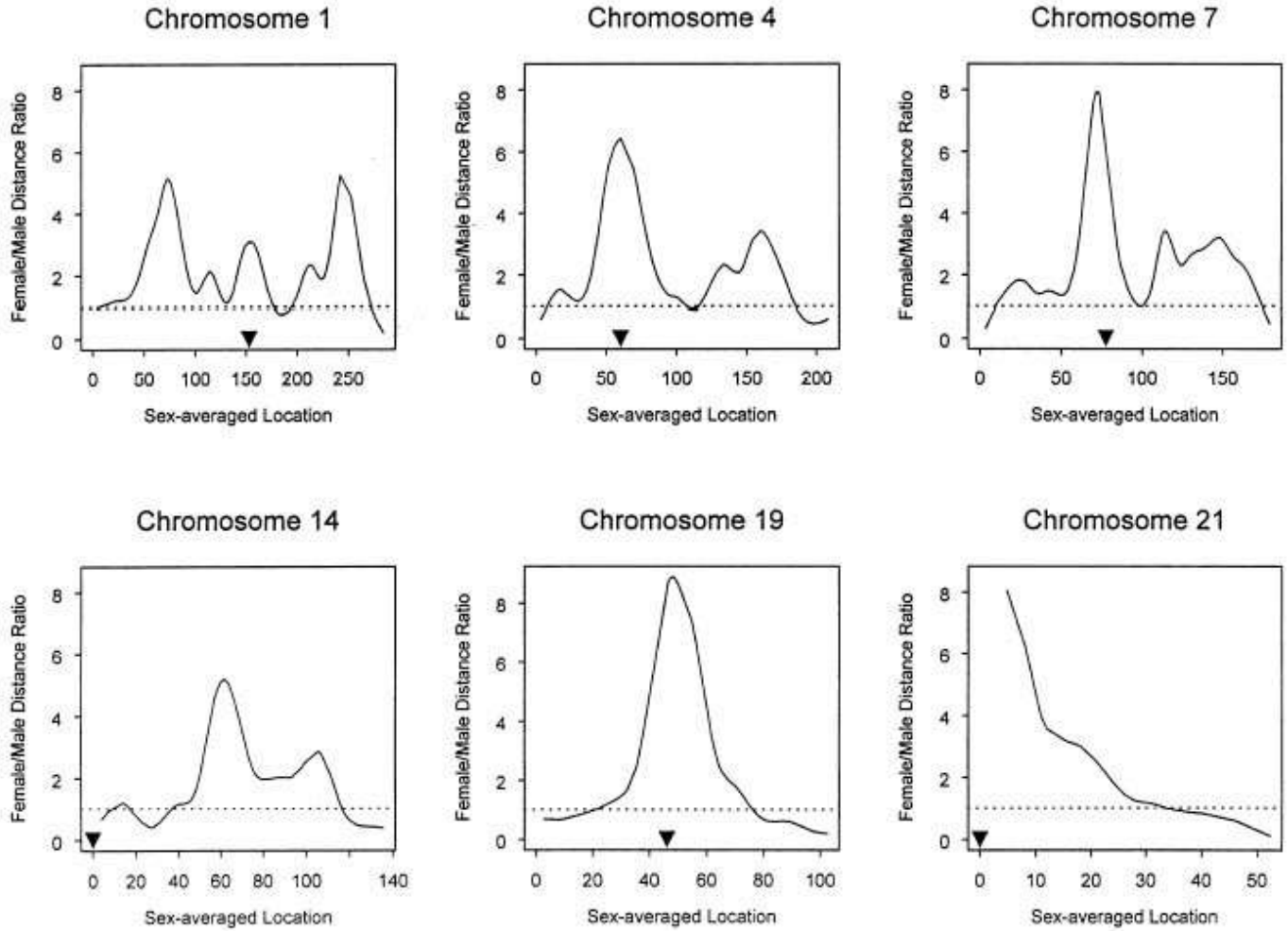
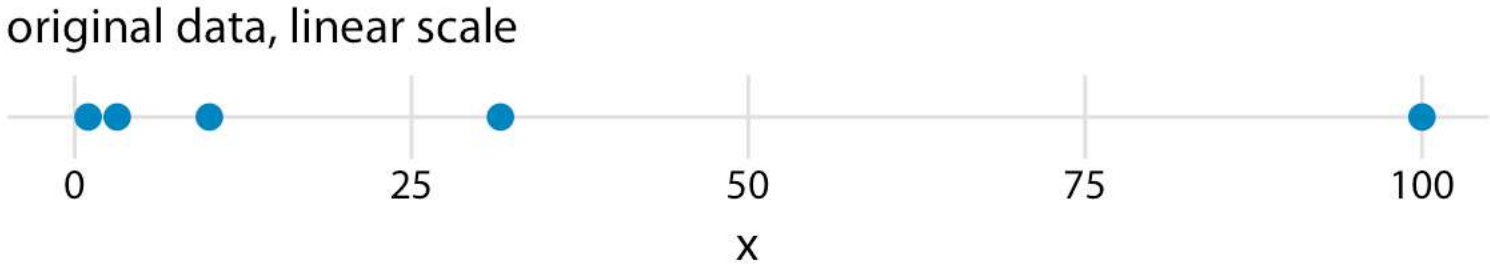
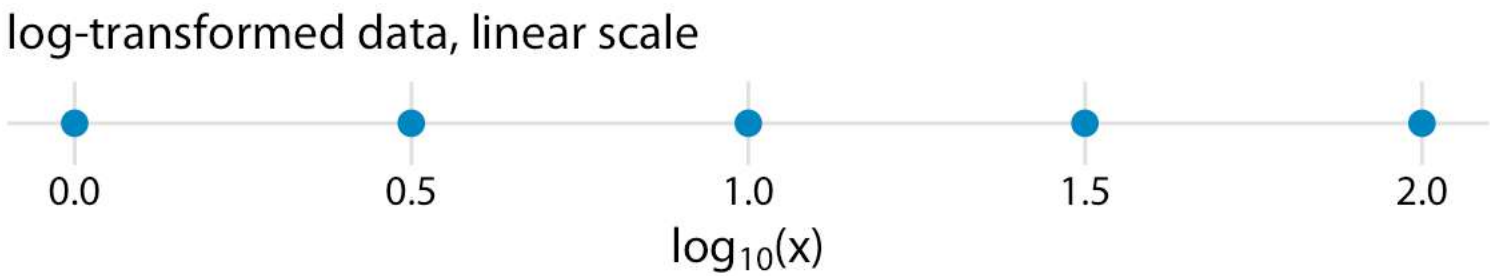
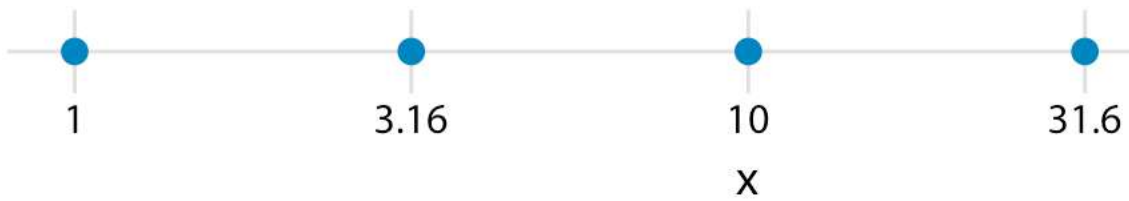


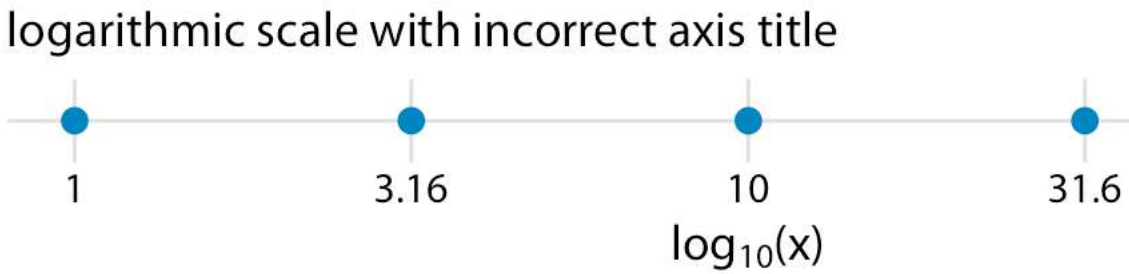
Figure 1 Plots of the female:male genetic-distance ratio against sex-averaged genetic location (in cM) along six selected chromosomes. Approximate locations of the centromeres are indicated by the triangles. The dashed lines correspond to equal female and male distances.

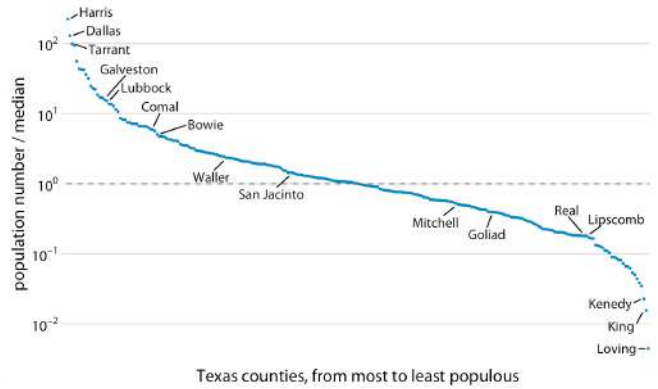
Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998)
Comprehensive human genetic maps: Individual and sex-specific variation in recombination.
American Journal of Human Genetics 63:861-869, Figure 1

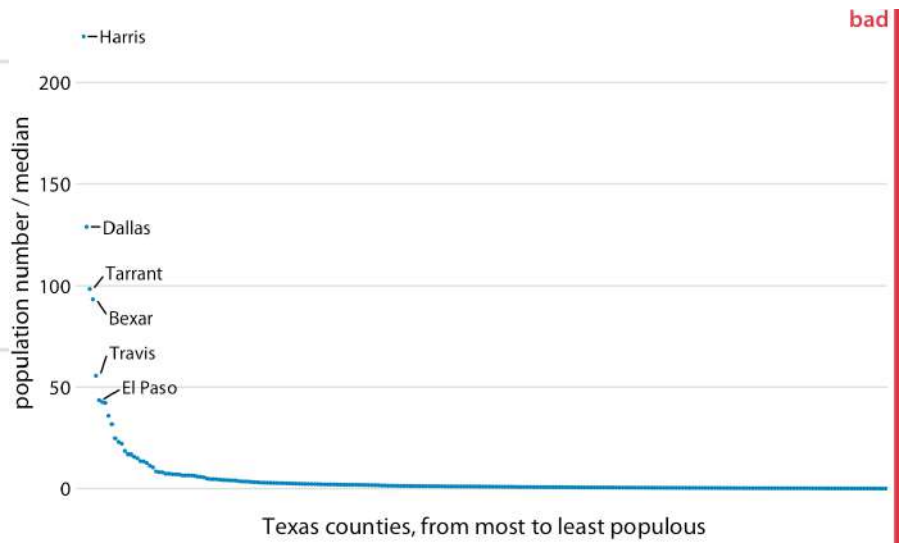

original data, linear scale
0
25
50
75
100
x


log-transformed data, linear scale
0.0
0.5
1.0
1.5
2.0
log10(x)


original data, logarithmic scale
1
3.16
10
31.6
x


logarithmic scale with incorrect axis title
1
3.16
10
31.6
log10(x)


Harris
Dallas
Tarrant
Galveston
Lubbock
Comal
Bowie
Waller
San Jacinto
Mitchell
Goliad
Real
Lipscomb
Kenedy
King
Loving
population number / median
Texas counties, from most to least populous


bad
Harris
Dallas
Tarrant
Bexar
Travis
El Paso
200
150
100
50
0
population number / median
Texas counties, from most to least populous

#1 Conocer a tu audiencia
#2 Definir un mensaje claro
#3 Usar herramientas adecuadas
(y no abusar de ellas)
#4 Graficar los datos fielmente
#5 Tener consistencia interna
#6 Simplificar y jerarquizar
#7 Escribir un pie de figura informativo

¿Cuál es el mensaje
que querés transmitir?

¿Cómo es la mejor manera
de hacerlo fiel a los datos?

Referencias

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4161295/

https://clauswilke.com/dataviz/

https://ft.com/vocabulary

https://www.internationalscienceediting.com/how-to-write-a-figure-caption/

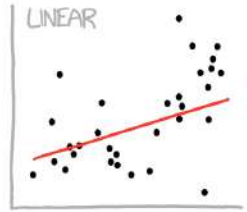
https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/

How to Lie with Statistics (Darrell Huff)

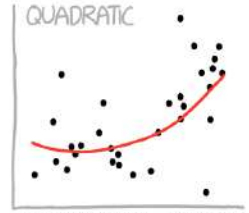
Story telling with data (Cole Nussbaumer Knaflic)

…

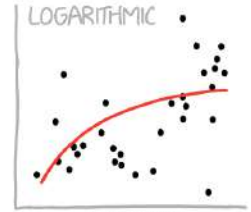
CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND



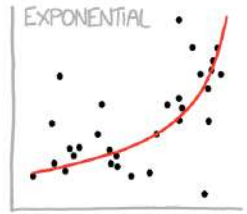
"HEY, I DID A REGRESSION."



"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."



"LOOK, IT'S TAPERING OFF!"



"LOOK, IT'S GROWING UNCONTROLLABLY!"

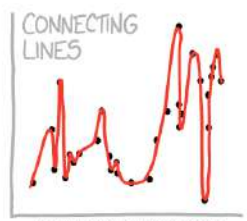
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."

"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."

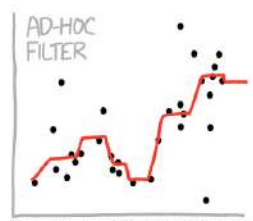
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."

"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."

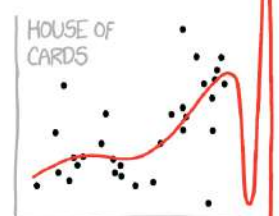
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."



"I CLICKED 'SMOOTH LINES' IN EXCEL."



"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"



"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE- *WAIT NO NO DON'T EXTEND IT AAAAAA!!*"