

Appendix A: 15 Most Common Issues

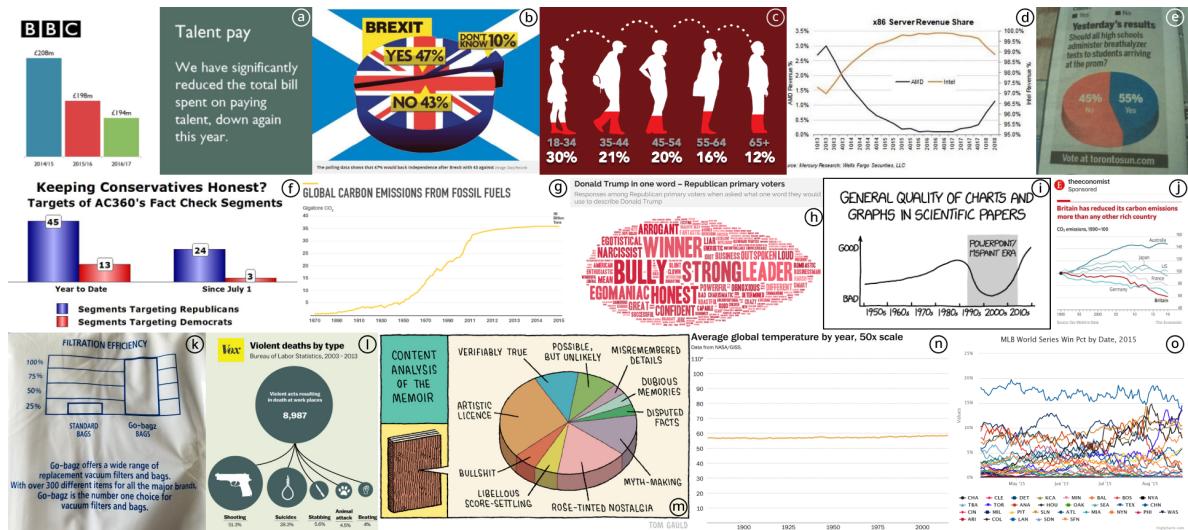


Figure 1: Top 15 issues. These issues are (a) Truncated Axis, (b) 3D, (c) Missing Title, (d) Dual Axis, (e) Misrepresentation, (f) Missing Axis, (g) Inconsistent Tick Intervals, (h) Missing Legend, (i) Not Data, (j) Selective Data, (k) Dubious Data, (l) Area Encoding, (m) Missing Value Labels, (n) Inappropriate Axis Range, and (o) Overusing Colors.

Appendix B: Description of Example Visualizations of Each Tag

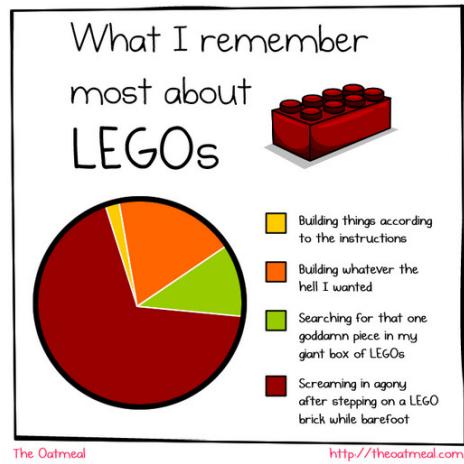


Figure 2a: Stage: Input Category: Garbage-in Tag: Not Data

Tag Description: The chart is not intended to be drawn according to the real data values.

Visualization Description: This humorous visualization reflects the author's most vivid memories of LEGOs. It gives extra weight to the memory of stepping on them as this is a painful experience likely shared by the audience. It is not intended to be based on real data.

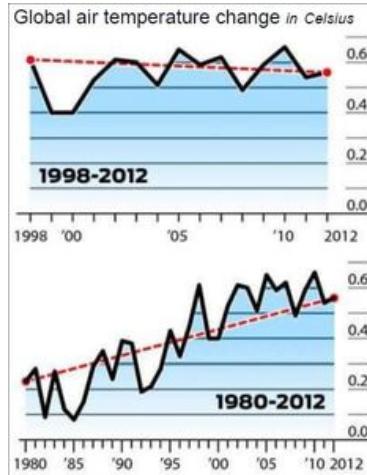


Figure 2b: Stage: Input Category: Garbage-in Tag: Selective Data

Tag Description: Selective data is data that has been cherry-picked to tell a particular story or support a specific claim.

Visualization Description: By selecting only the period from 1998 to 2012, this visualization obscures the overall trend of increasing temperature, thus misleading the audience.

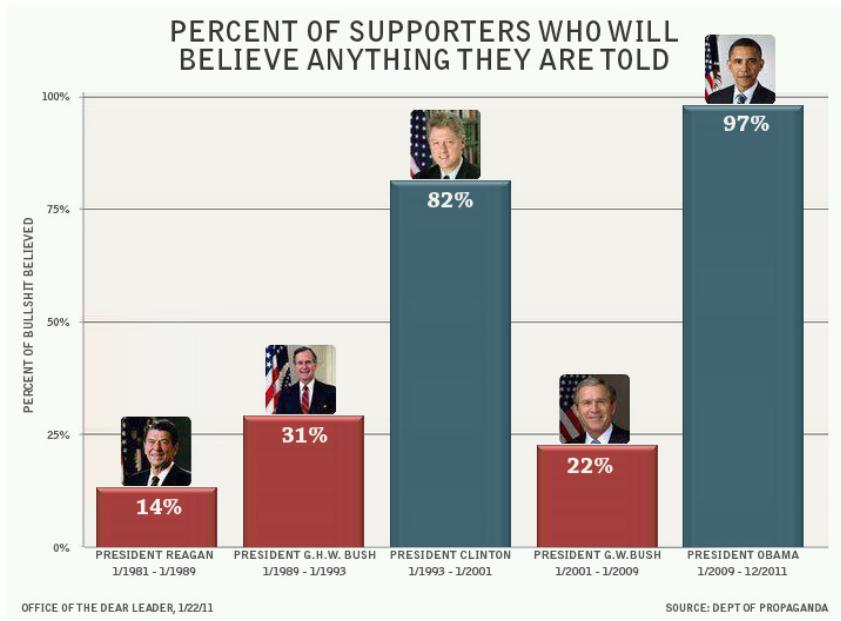


Figure 2c: Stage: Input Category: Garbage-in Tag: Dubious Data

Tag Description: In some instances, a dubious source is cited for the underlying data associated with the visualization, calling the credibility of the data into question.

Visualization Description: The data source cited for the political visualization is “DEPT OF PROPAGANDA” and it claims to be from the “OFFICE OF THE DEAR LEADER”. No attempt has been made to hide the lack of credibility of these claims.



Figure 2d: Stage: Input Category: Garbage-in Tag: Non Sequitur

Tag Description: The visualization does not attempt conveying usable information. It has been created purely for decorative purposes.

Visualization Description: The visualization, which serves only as a decorative backdrop for the presenter, looks like data visualization, but upon closer inspection, it is not.



Figure 2e: Stage: Input Category: Garbage-in Tag: Too Few Data Points

Tag Description: It is clear that the underlying dataset for the visualization has too few data points to be meaningful.

Visualization Description: In this COVID-19 visualization for India, the line chart on the television screen shows a steep line, suggesting a significant rise in cases. However, according to the data presented, there was only one new confirmed COVID-19 case between the 7th of March and the 16th of March.

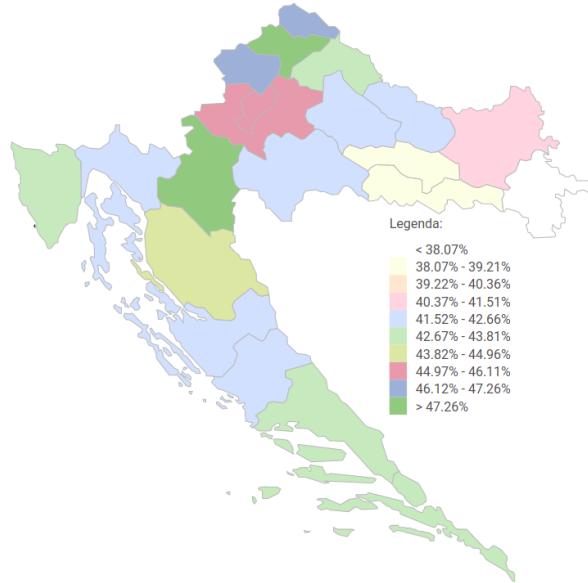


Figure 2f: Stage: Input Category: Lie with Statistics Tag: Discretized Continuous Variable

Tag Description: A discretized continuous variable is when a continuous variable is transformed into a categorical variable by cutting into groups, thus exaggerating the difference between boundary cases.

Visualization Description: This visualization has split the continuous variable into multiple categories. In an extremely misleading case, the difference between a red region and a blue region can be as little as 0.01%.

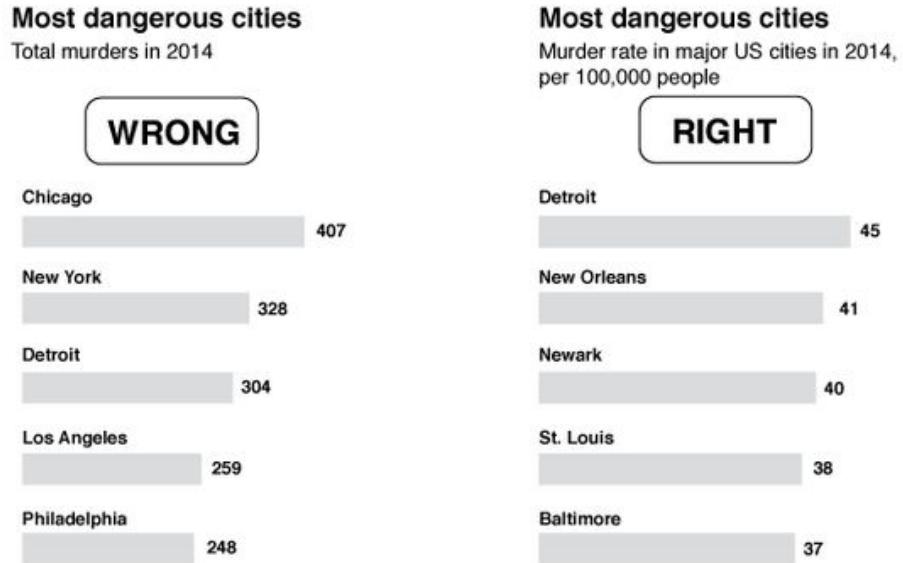


Figure 2g: **Stage: Input Category: Lie with Statistics Tag: Missing Normalization**

Tag Description: The visualization compares numbers in absolute values rather than normalized values, thus creating an incorrect narrative and misleading the audience.

Visualization Description: By ranking the cities by the absolute number of murders for a given year, the author of this visualization completely disregards the most important factor: the population of those cities.

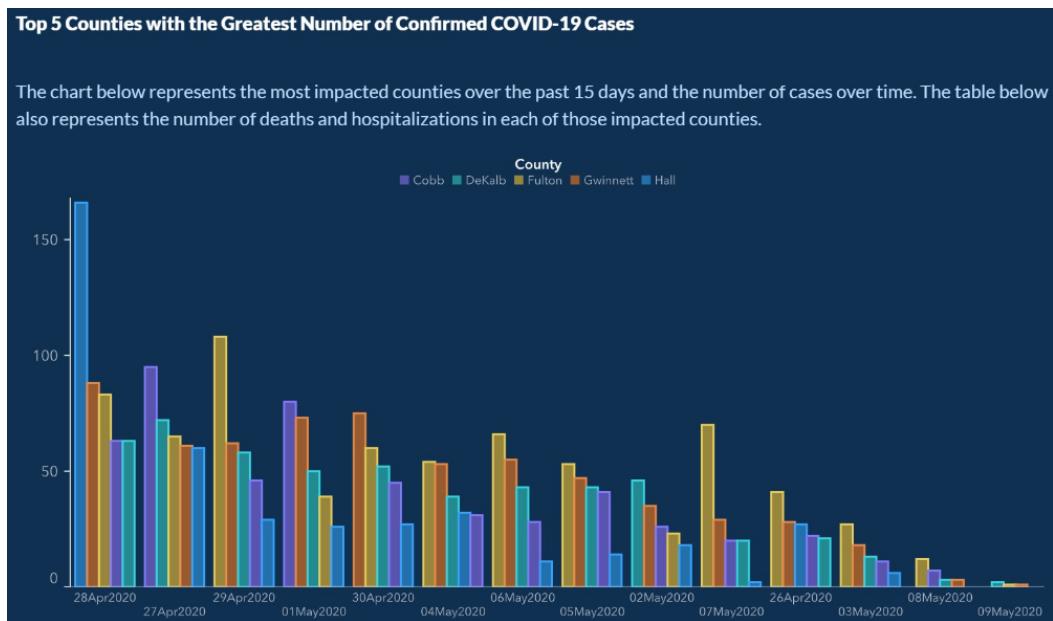
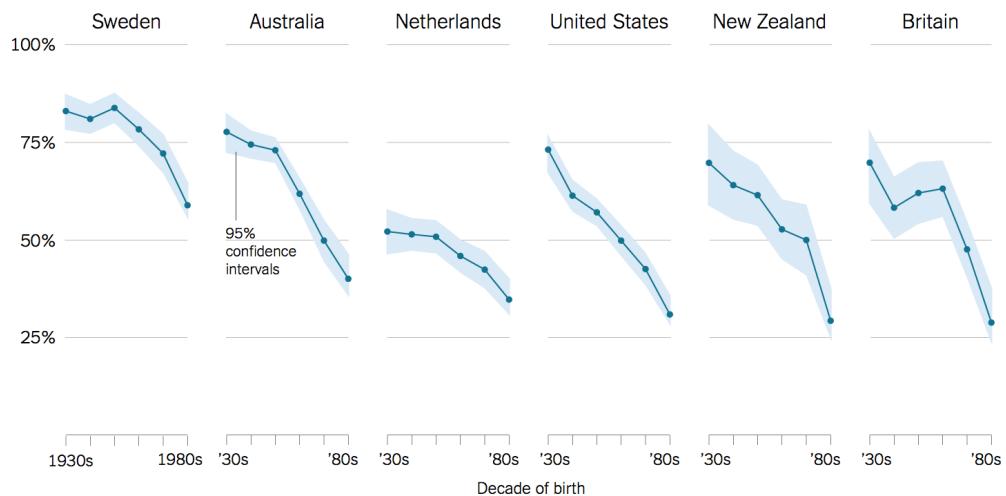


Figure 2h: **Stage: Input Category: Lie with Statistics Tag: Inappropriate Item Order**

Tag Description: In some cases, the items are arranged in an unconventional order, misleading the audience or creating confusion.

Visualization Description: In this COVID-19 cases visualization, the dates on the x-axis are ordered arbitrarily, switching between various dates in April and May, to create a decreasing trend that does not exist.

Percentage of people who say it is “essential” to live in a democracy



Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” Journal of Democracy | By The New York Times

Figure 2i: Stage: Input Category: Lie with Statistics Tag: Inappropriate Metric

Tag Description: In an inappropriate metric, the author carefully selects a metric, threshold, or definition that will help support an unconvincing claim.

Visualization Description: For this visualization on the topic of democracy, only questionnaire responses rated 10 are counted as meaning it is “essential” to live in a democracy, while ratings of 9 or less are counted as meaning it is “non-essential” to live in a democracy.

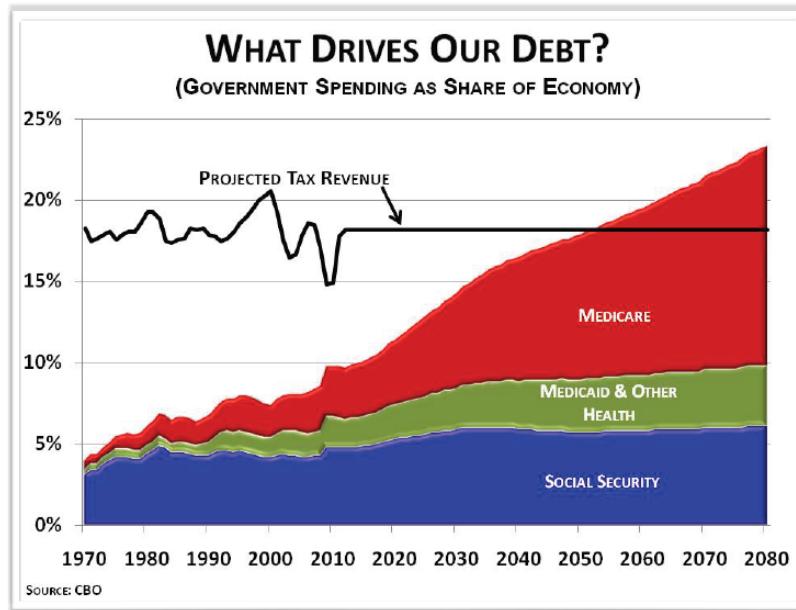


Figure 2j: Stage: Input Category: Lie with Statistics Tag: Questionable Prediction

Tag Description: Questionable prediction refers to instances in which the data series extends beyond the present time without stating how the data is projected.

Visualization Description: In this visualization of the national debt in America, beyond 2010, the projected tax revenue rate becomes a constant line, which is a highly unlikely occurrence. Furthermore, no projection method is stated for the spending projection.

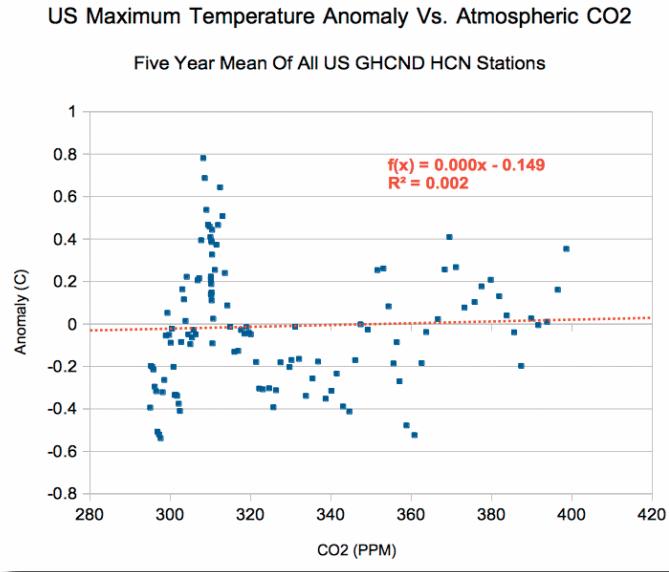
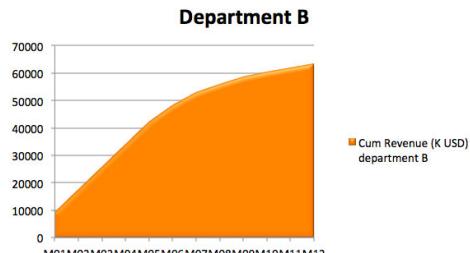
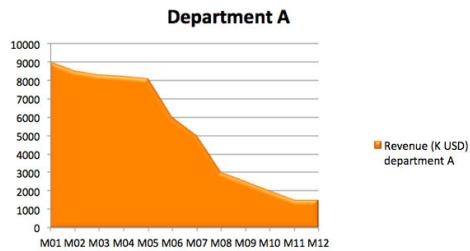


Figure 2k: Stage: Input Category: Lie with Statistics Tag: Trend Line on Random Data

Tag Description: Trend line on random data occurs means the regression line is drawn on weakly correlated data, causing the audience to be misled into thinking there is a trend.

Visualization Description: In this visualization of maximum temperature anomalies in the US, drawing a regression line by connecting random data points through a scattering of data points is not a credible approach. The line drawn essentially means nothing.

Where would you put your money?



Answer: Both are doing bad. The numbers are exactly the same.

Figure 2l: Stage: Input Category: Lie with Statistics Tag: Inappropriate Use of Accumulation

Tag Description: A cumulative measure is used on data that can only accumulate to hide the declining trend.

Visualization Description: The graph for Department A depicts revenue while the graph for Department B depicts cumulative revenue. At a glance, the audience is likely to miss this distinction. In reality, the revenue of Department B is actually declining.



Figure 2m: Stage: Input Category: Lie with Statistics Tag: Inappropriate Aggregation Granularity
Tag Description: Inappropriate aggregation granularity is related to cases where the level of granularity results in an inaccurate interpretation of the data. A higher or lower level of granularity may be needed to accurately reflect the data.

Visualization Description: This choropleth map concerning the percentage of the population using the language “Irish Gaelic” misleads the audience by displaying the data on a country level. Therefore, the information for the Northern Ireland region on this map is inaccurate.

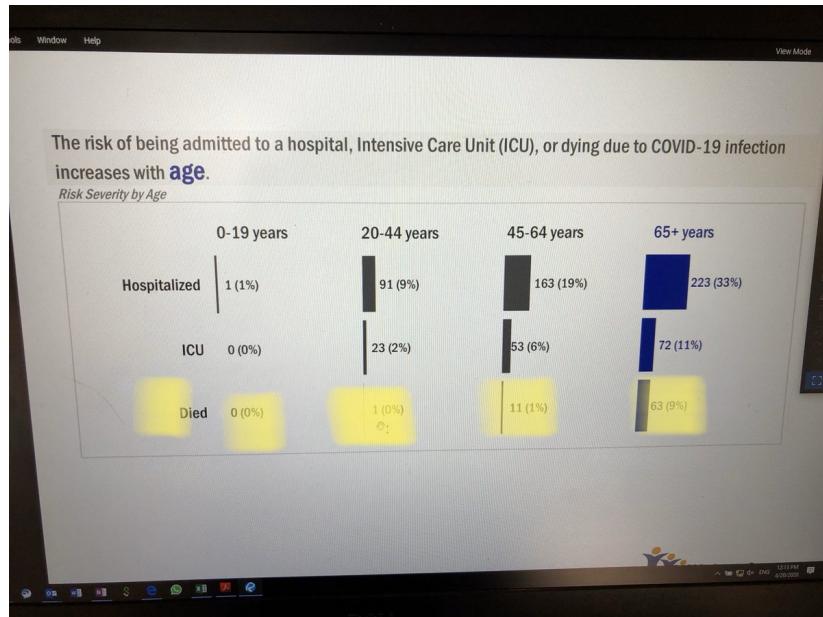


Figure 2n: Stage: Input Category: Lie with Statistics Tag: Two-way Normalization
Tag Description: Normalization can be arbitrarily done, whether vertically or horizontally on a two-way table.

Visualization Description: In this visualization of COVID-19 hospitalizations and deaths, the percentage values in the chart can be interpreted arbitrarily by the reader.

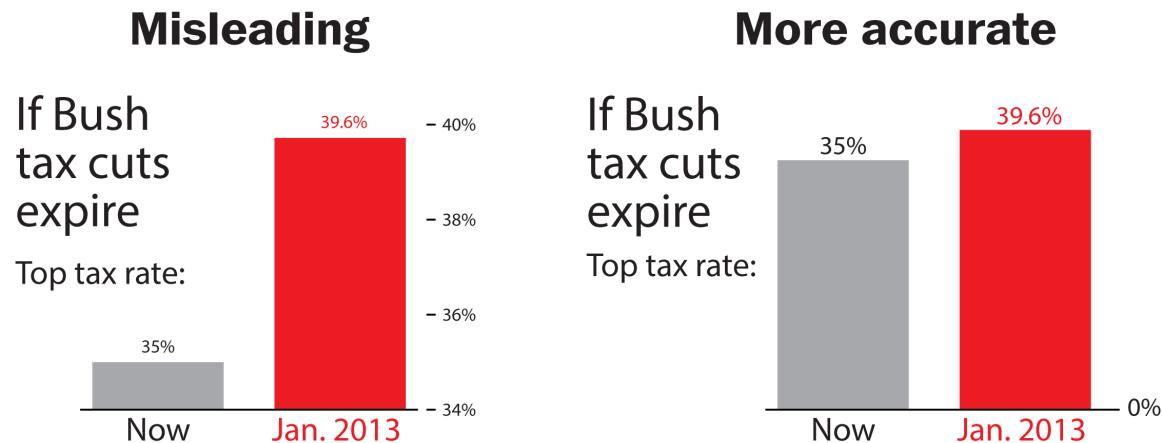


Figure 3a: Stage: Visualization Design Category: Choice of Axis Tag: Truncated Axis

Tag Description: In a truncated axis, the axis does not start from zero or is truncated in the middle, resulting in an exaggerated difference between the two bars.

Visualization Description: By truncating the axis in this visualization of expiring tax cuts in the USA, the author exaggerates the difference between the two bars, thus misleading the audience by creating the idea of a significant rise in the top tax rate.

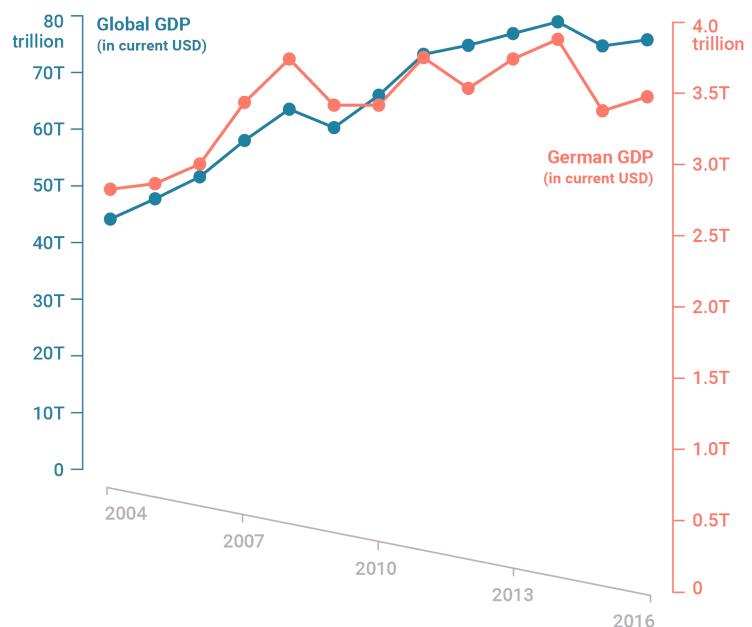


Figure 3b: Stage: Visualization Design Category: Choice of Axis Tag: Dual Axis

Tag Description: Dual axis is when two independent axes are layered on top of each other with inappropriate scaling. This results in a misleading narrative about the relationship between the two.

Visualization Description: In this visualization comparing global GDP to the German GDP, separate scales are used for each of the two axes, thus misleading the audience.

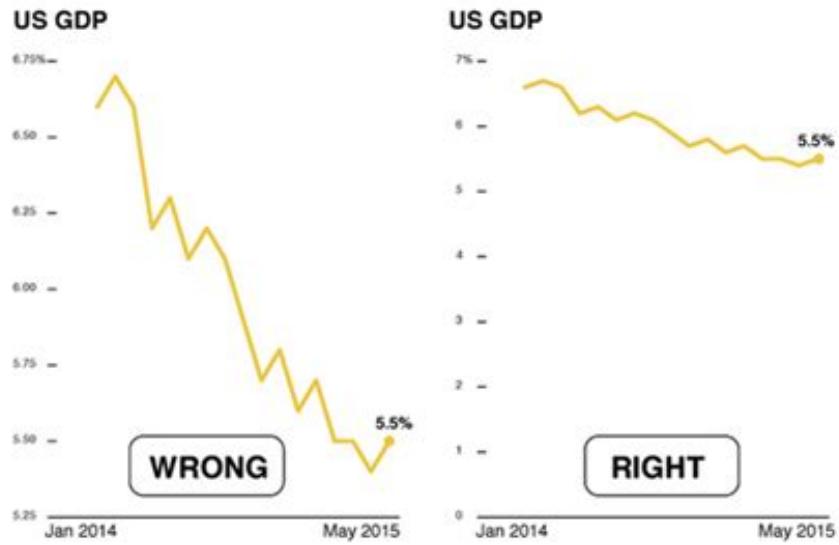


Figure 3c: Stage: Visualization Design **Category:** Choice of Axis **Tag:** Inappropriate Axis Range **Tag Description:** In the case of an inappropriate axis range, the axis range is either too broad or too narrow to accurately visualize the data, allowing changes to be minimized or maximized depending on the author's intention.

Visualization Description: In this visualization, using a different axis range creates the illusion of a much more severe decline in US GDP.

states compare in terms of their reliance on federal funding. The states deemed "most dependent" by the analysis are bright red on the map, those "least dependent" are bright green. You can move your cursor around on the map to see how each state ranks. (There were some ties.)

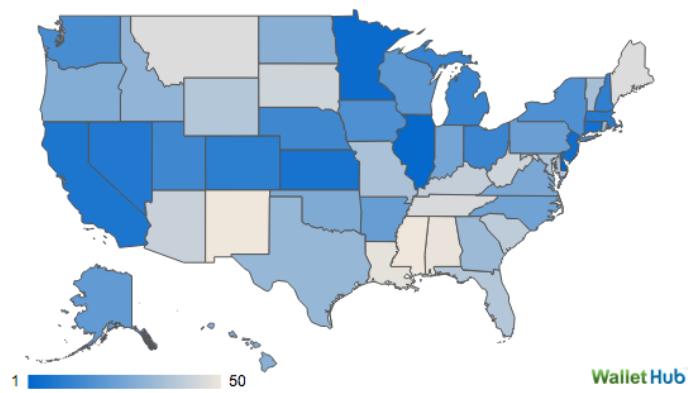


Figure 3d: Stage: Visualization Design **Category:** Choice of Axis **Tag:** Inverted Axis **Tag Description:** An inverted axis is when an axis is oriented in an unconventional direction and the perception of the data is reversed, thus misleading or confusing the audience.

Visualization Description: The solid blue color represents the least dependent states, while the most dependent states are in a light-grey color. This flips the convention of saturated colors representing a stronger intensity and less saturated colors representing the weaker intensity of the measured metric.

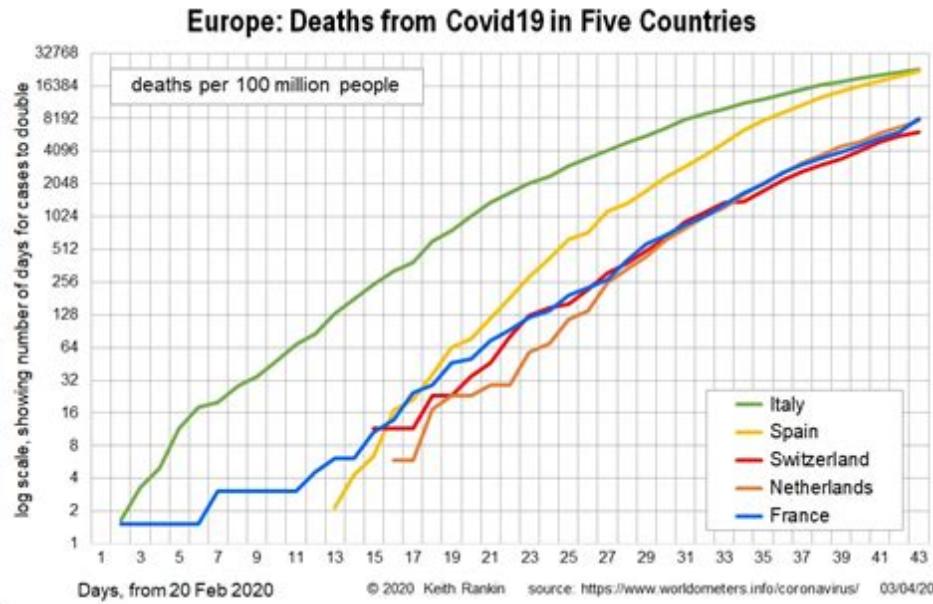


Figure 3e: Stage: Visualization Design **Category:** Choice of Axis

Tag: Log Scale **Tag Description:** Log scale is applied to non-exponential data.

Visualization Description: People who are not aware of the log scale used in this visualization of deaths from COVID-19 will incorrectly interpret the curves as flattening.

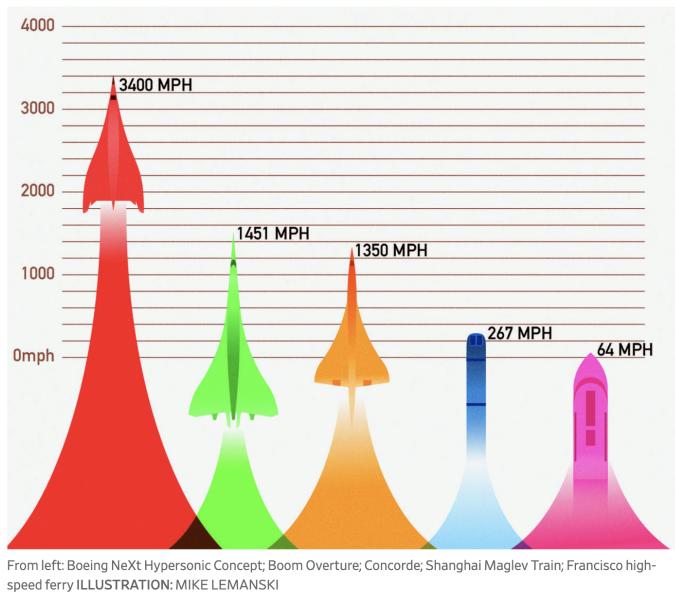


Figure 3f: Stage: Visualization Design **Category:** Choice of Axis

Tag: Extended Axis **Tag Description:** An extended axis is one that starts below zero, resulting in the difference between the measured items being understated.

Visualization Description: In this visualization of transport speed, the hypersonic plane's speed is more than 50 times that of the train, but the height in the visualization is only doubled, thus misleading the audience.

Tracking Coronavirus in New Mexico

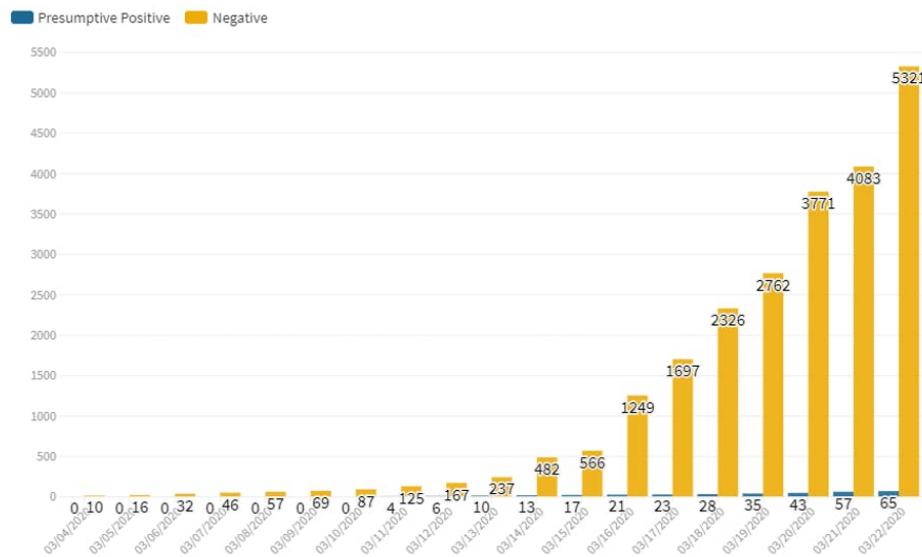


Figure 3g: Stage: Visualization Design **Category:** Choice of Axis **Tag:** Data of Different Magnitudes **Tag Description:** Indistinguishable differences in a series result from two series of data with different magnitudes plotted on the same chart.

Visualization Description: In this visualization of COVID-19 in New Mexico, the blue bar is completely dominated by the yellow bar. As a result, the increases in the blue bar over time are almost imperceptible.

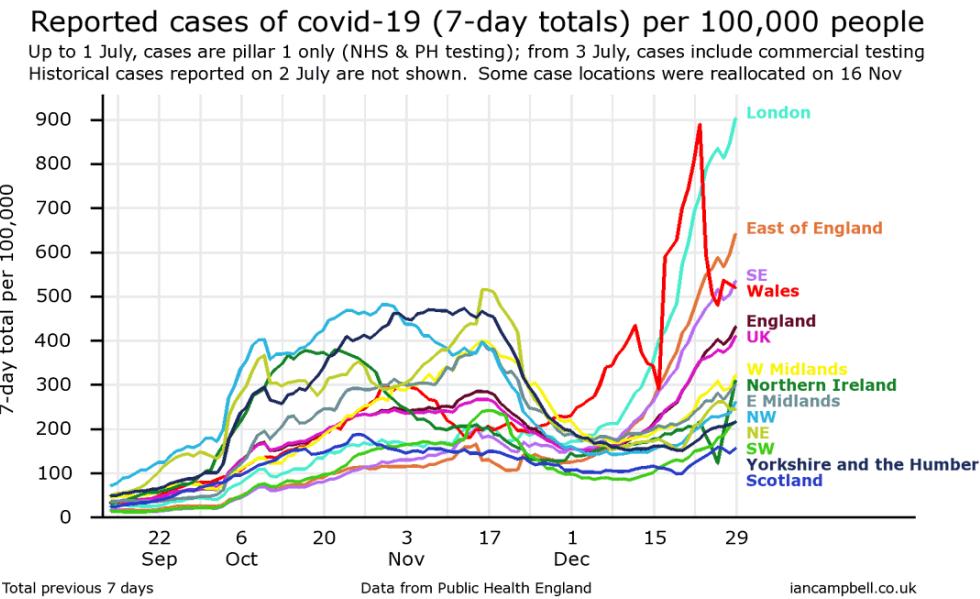


Figure 3h: Stage: Visualization Design **Category:** Choice of Axis **Tag:** Linear Scale on Exponential Data

Tag Description: The chart keeps its default linear scale and does not apply log scale on exponential data.

Visualization Description: The spread of the virus is an exponential process, but this visualization of COVID-19 cases uses a linear scale. A linear scale cannot capture whether or not the spread is slowing down or speeding up.

Points and lines for categorical variables

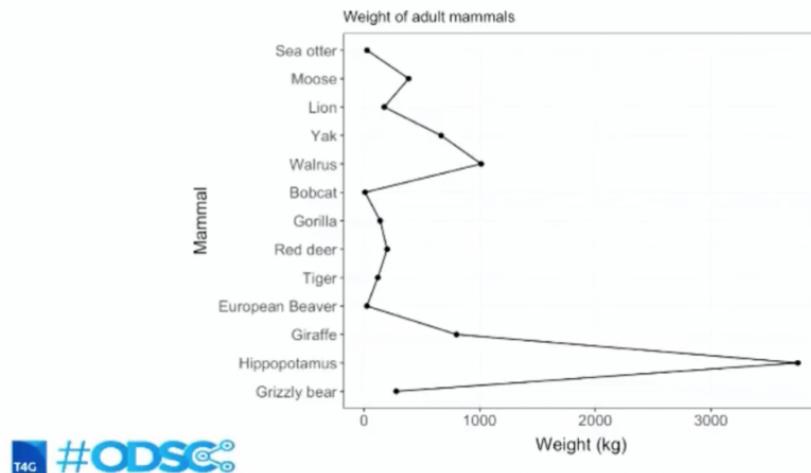


Figure 3i: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Inappropriate Use of Line Chart

Tag Description: A line chart is deemed inappropriate when used in an unconventional way or in a way that results in incorrect interpretation of the data or intentionally misleading the audience. Examples are encoding a categorical variable on one of the axes or encoding the time dimension on the y-axis.

Visualization Description: In this visualization of the weight of adult mammals, although the slope of the line in the line chart is visually prominent, in terms of the data, it holds no meaning.

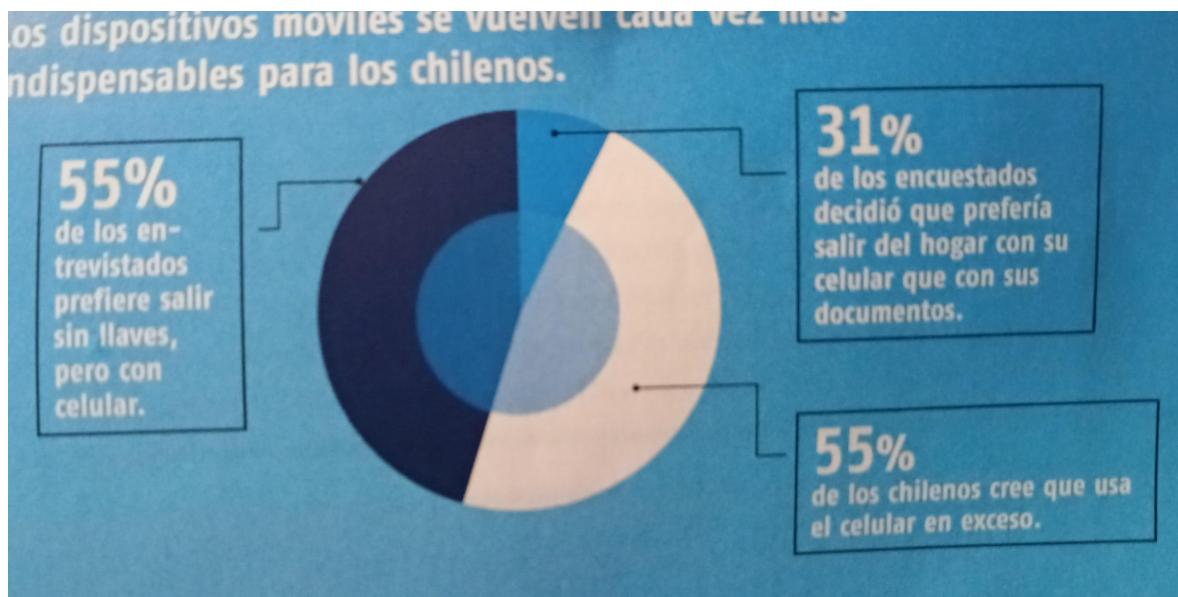


Figure 3j: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Inappropriate Use of Pie Chart

Tag Description: When a pie chart is used for non-part-to-whole data, it creates confusion for the audience, who may misinterpret the significance of a given section.

Visualization Description: The data has no part-to-whole relationship in this pie chart. The three sections of the pie chart add up to 141% rather than 100%, creating confusion and potentially misleading the audience.

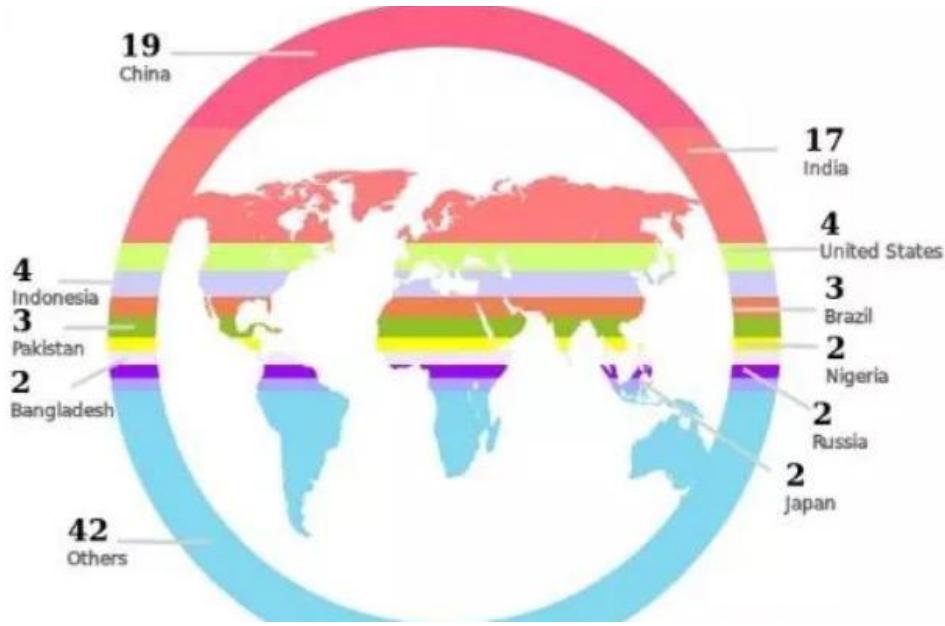


Figure 3k: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Confusing Chart Type
Tag Description: Confusing charts are often those that partly belong to a common chart type but also deviate from them.

Visualization Description: The above chart combines a map with a donut chart, but it also encodes the values using the height of horizontal stripes. This deviates significantly from what the reader of this visualization would expect and may result in confusion and misinterpretation.

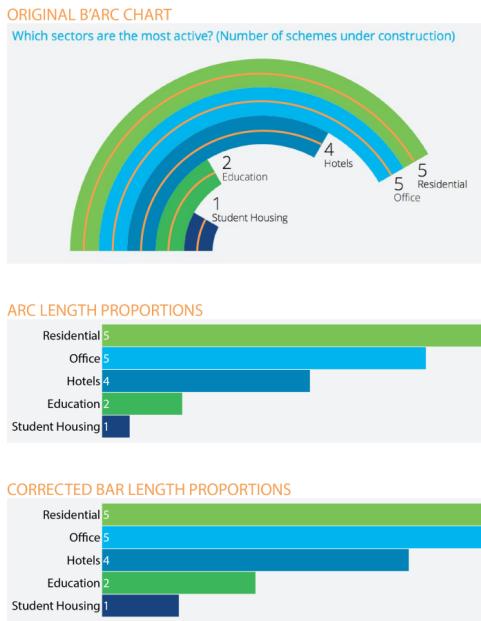


Figure 3l: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Misusing Circular Layout
Tag Description: A circular layout is deemed misused when it attempts to use curved bars in a comparison or renders length-encoded lines in curves, resulting in an inaccurate comparison for the reader.

Visualization Description: In this visualization of the activity levels of the specified industries, curved bar lengths should not be used in the context of a comparison as this is an inaccurate and unclear visualization method.

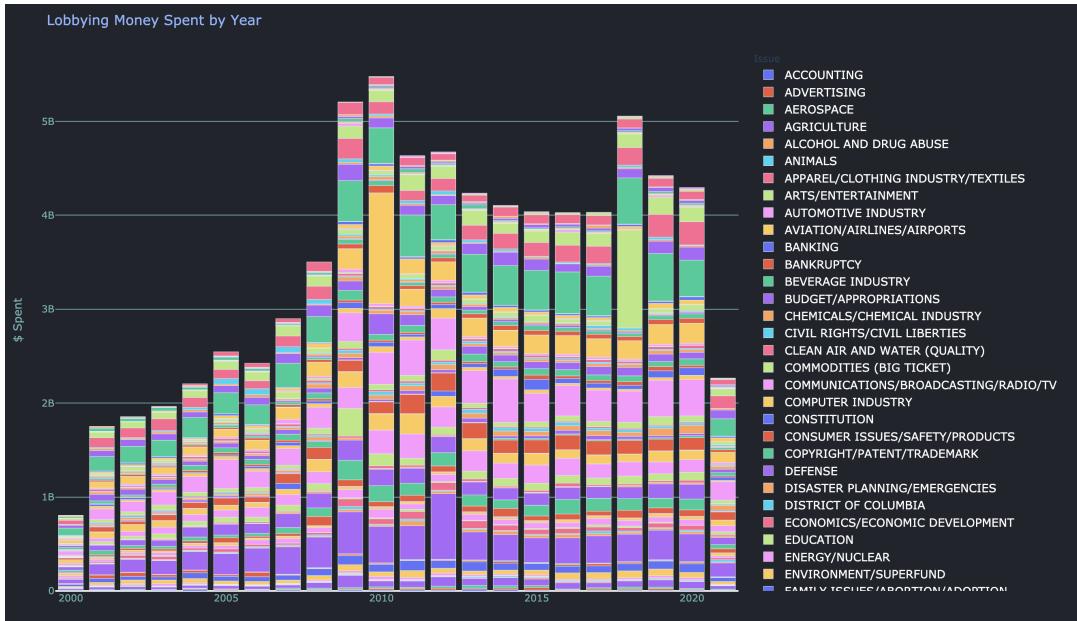


Figure 3m: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Inappropriate Use of Stacked

Tag Description: Inappropriate use of stacked simply means too many layers have been stacked upon each other, making the entire visualization incomprehensible for the reader.

Visualization Description: In this visualization of lobbying money spent by year, it is clear that far too many layers have been stacked upon each other. As a result, the colors are indistinguishable, and the reader cannot draw any meaningful insights from the data presented.

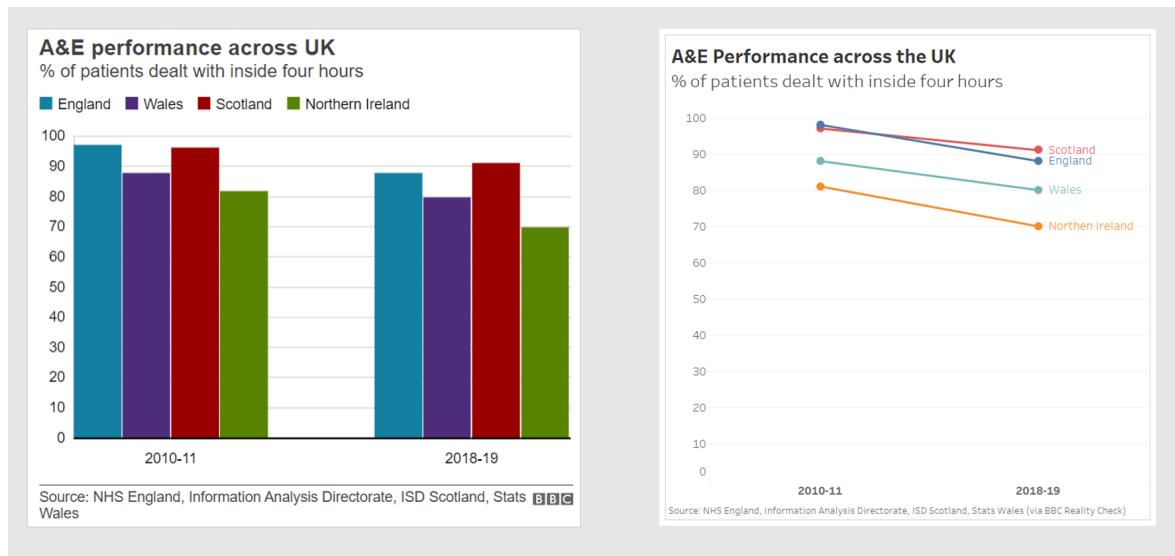


Figure 3n: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Inappropriate Use of Bar Chart

Tag Description: When bar charts are used unconventionally, the trends, patterns, and insights can be lost in the design, resulting in the reader being unable to understand the information or completely misinterpreting the data presented.

Visualization Description: In the case of this visualization of Accident and Emergency (A&E) performance across the UK, it's clear that a line chart better represents the data. The pattern, a downward trend, is not evident from the bar chart, but when the data is presented differently, this pattern is revealed.

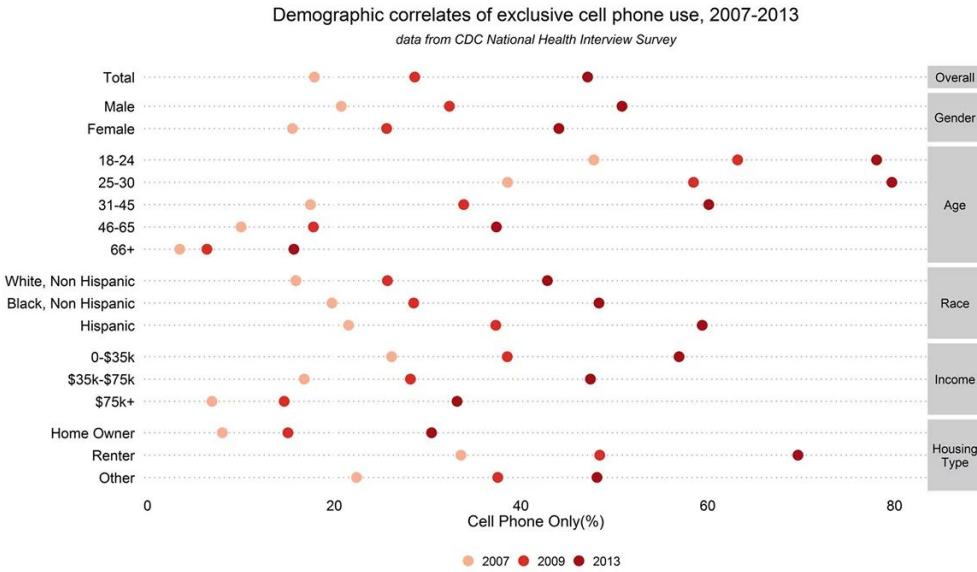


Figure 3o: **Stage:** Visualization Design **Category:** Choice of Chart **Tag:** Inappropriate Use of Scatterplot

Tag Description: When a scatterplot is used unconventionally, it can be difficult for the reader to spot trends and patterns or correctly interpret the visualized data. Encoding a categorical variable on one of the axes is an example.

Visualization Description: While this chart on cellphone use may look similar to a scatterplot at a glance, it is not a scatterplot at all. As a result, the data is difficult and time-consuming to interpret correctly.

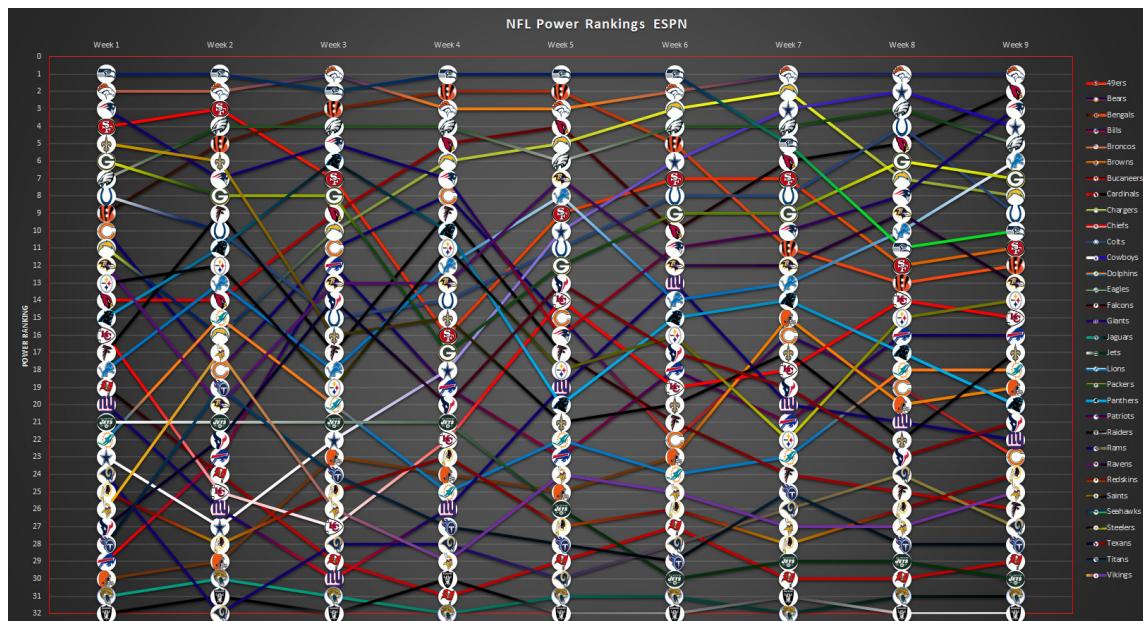


Figure 3p: **Stage:** Visualization Design **Category:** Color Mess **Tag:** Overusing Colors
Tag Description: Overusing colors refers to using too many colors in one chart.

Visualization Description: In this chart on NFL power rankings, it is clear that too many colors are being used. Furthermore, some of the colors gradually change along the x-axis, shifting from white to purple or red to orange, for example. This makes it almost impossible for the reader to draw accurate conclusions from the presented data.



Figure 3q: Stage: Visualization Design Category: Color Mess Tag: Indistinguishable Colors
Tag Description: The same color or very similar colors are used for the different categories. Readers inevitably mistake one category for another.

Visualization Description: In the case of this visualization, an incompatible visualization type was selected for the presentation medium. Printed in black and white for a newspaper, the colors for each state on this map of America are completely indistinguishable.

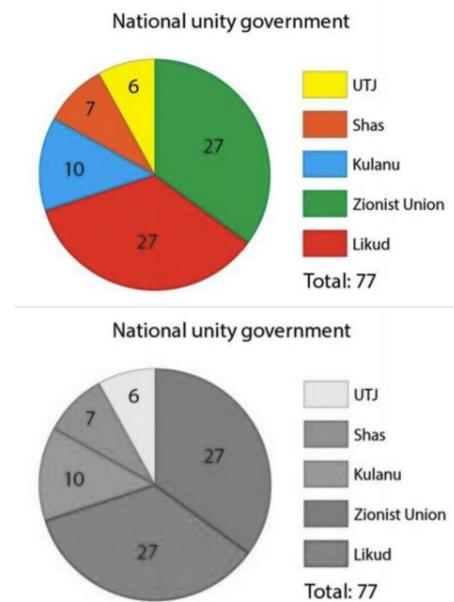


Figure 3r: Stage: Visualization Design Category: Color Mess Tag: Color Blind Unfriendly
Tag Description: For color-blind people, a Color Blind Unfriendly visualization is similar to a visualization with indistinguishable colors. It cannot be interpreted correctly by color-blind people and is open to being misinterpreted.

Visualization Description: A colorful visualization, like the pie charts presented above, can be indistinguishable for color-blind people. To combat this, select colors with different saturation levels for categorical data, not just different color hues. This will at least create some distinction between the categories.

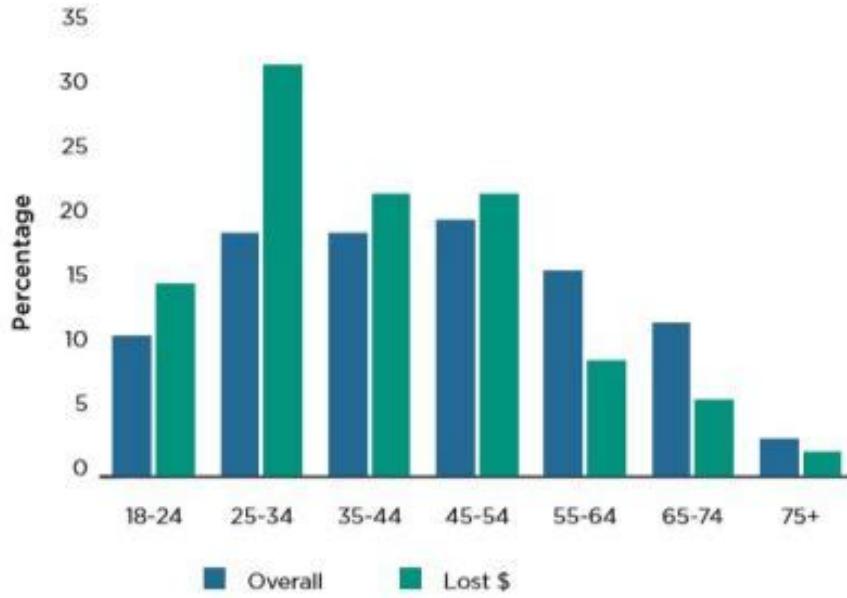


Figure 4a: Stage: Plotting Category: Incomplete Chart Tag: Missing Title

Tag Description: In some instances, no title is provided for the visualization. As a result, the reader misses out on the entire context of the visualization and cannot access the information presented in it.

Visualization Description: With the title of the above visualization omitted, what exactly this bar chart is referring to remains a complete mystery.

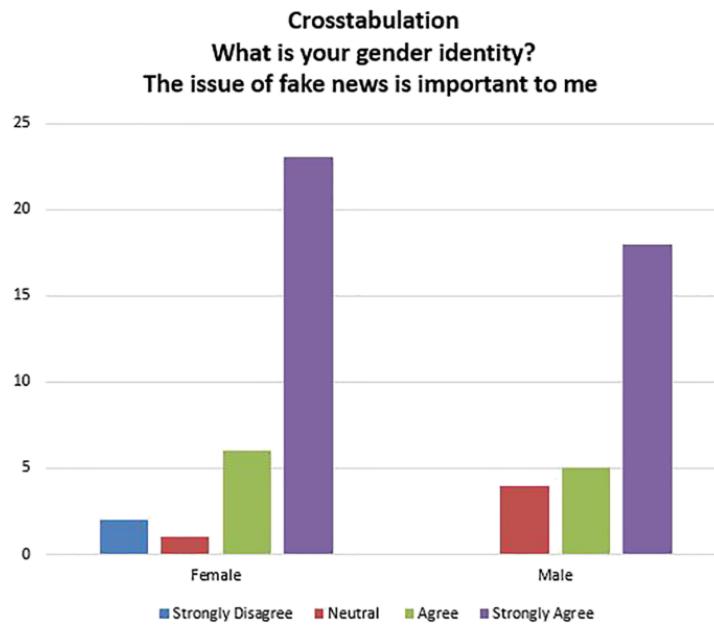


Figure 4b: Stage: Plotting Category: Incomplete Chart Tag: Missing Axis Title

Tag Description: When no axis title is provided, the meaning of all the data related to that axis is lost.

Visualization Description: In this visualization of gender identity, the figures on the y-axis have no title. They could represent anything from the number of people questioned to the age of those people to a percentage or something completely different. Thus, no valid insights can be drawn from this bar chart.

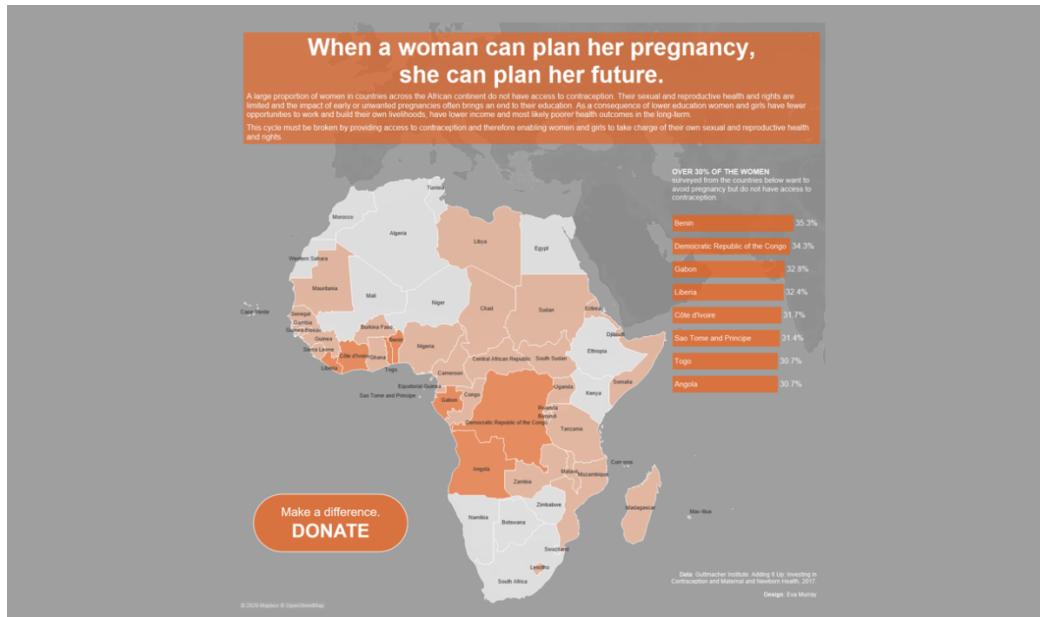


Figure 4c: Stage: Plotting Category: Incomplete Chart Tag: Missing Legend

Tag Description: Omitting a legend for a visualization makes it impossible for the reader to understand the visual encoding.

Visualization Description: In this choropleth map, color encoding is used but no legend has been provided. Thus, the reader has no way to discern the information related to the colors used for each section.

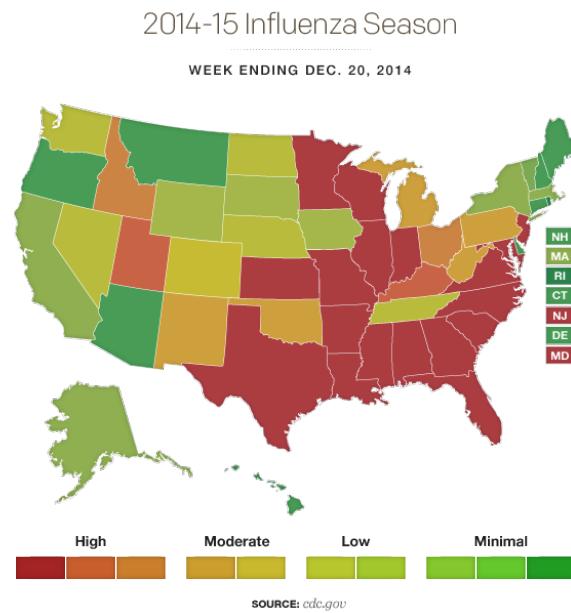


Figure 4d: Stage: Plotting Category: Incomplete Chart Tag: Missing Value Labels

Tag Description: In instances where no value labels have been included on charts that have no coordinated axis, the audience is left with no way extract information from the visualization.

Visualization Description: In this chart on the 2014–15 influenza season, the terms “high,” “moderate,” “low” and “minimal” have not been defined with a value label or an annotation on the map, making it impossible for the reader to interpret or verify the data accurately.

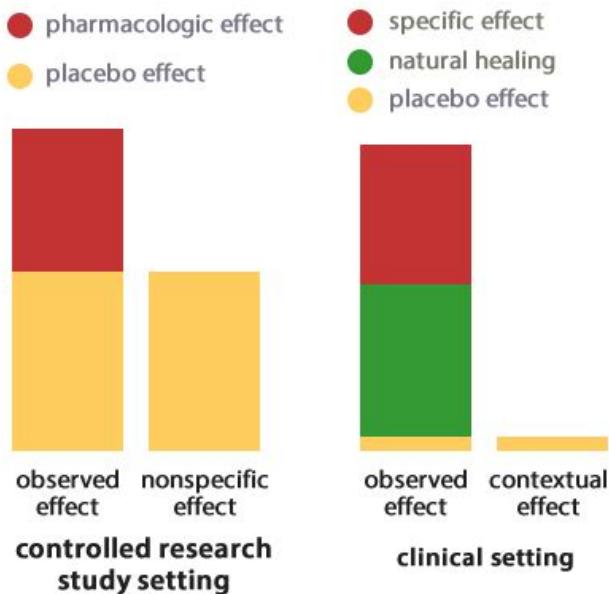


Figure 4e: Stage: Plotting Category: Incomplete Chart Tag: Missing Axis

Tag Description: Missing axis means that one or more axes are missing, and no value is retrievable.
Visualization Description: The y-axis has not been included in the above bar chart, making the visualization entirely incomprehensible. No meaningful information can be extracted from the data presented.

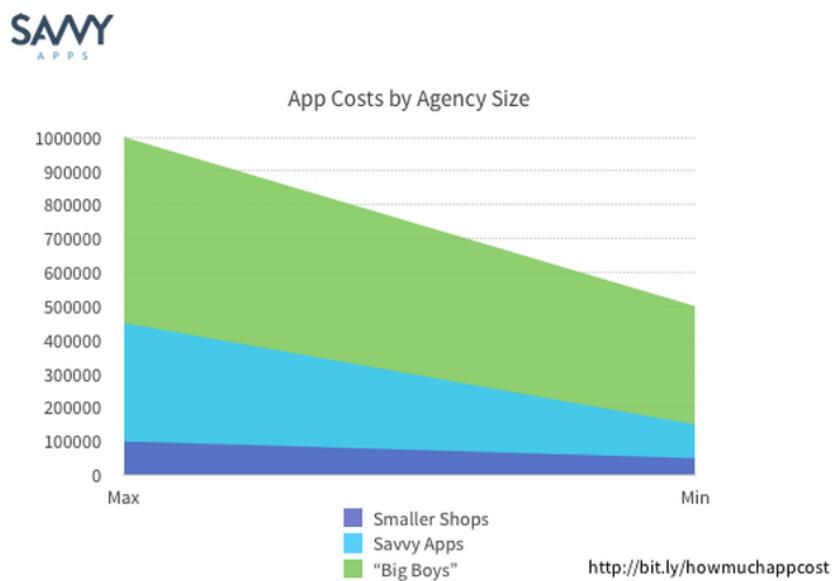


Figure 4f: Stage: Plotting Category: Incomplete Chart Tag: Missing Axis Ticks

Tag Description: When the axis does not include the appropriate tick marks, it is impossible for the reader to retrieve the exact values accurately.

Visualization Description: In this visualization of app costs, tick marks have been omitted along the x-axis. Thus, it is impossible to know where the divisions lie, rendering the axis meaningless.

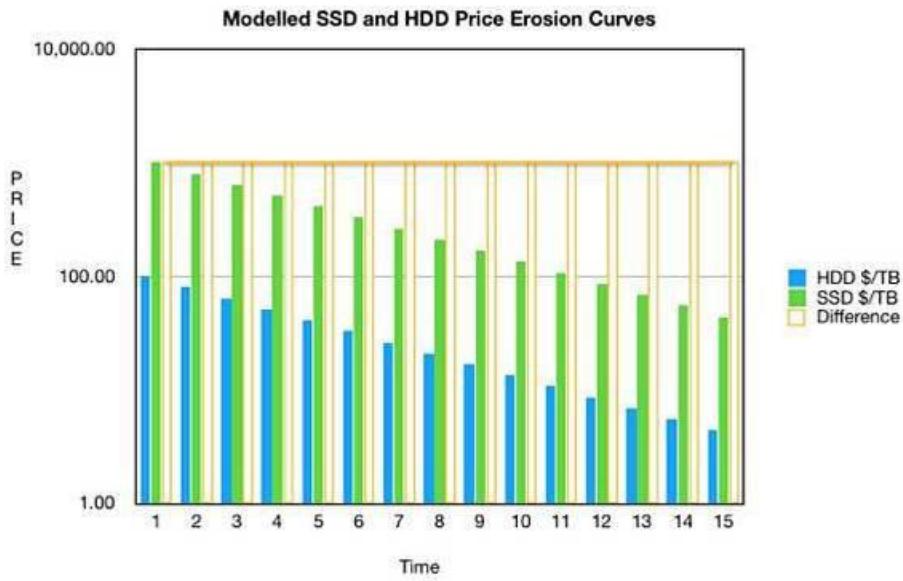


Figure 4g: **Stage:** Plotting **Category:** Incomplete Chart **Tag:** Missing Units
Tag Description: When units are omitted, the meaning of the numbers presented on the axis is completely ambiguous.
Visualization Description: In this visualization of price erosion curves, no details have been provided on the units associated with the figures for both axes, rendering them meaningless for the reader.

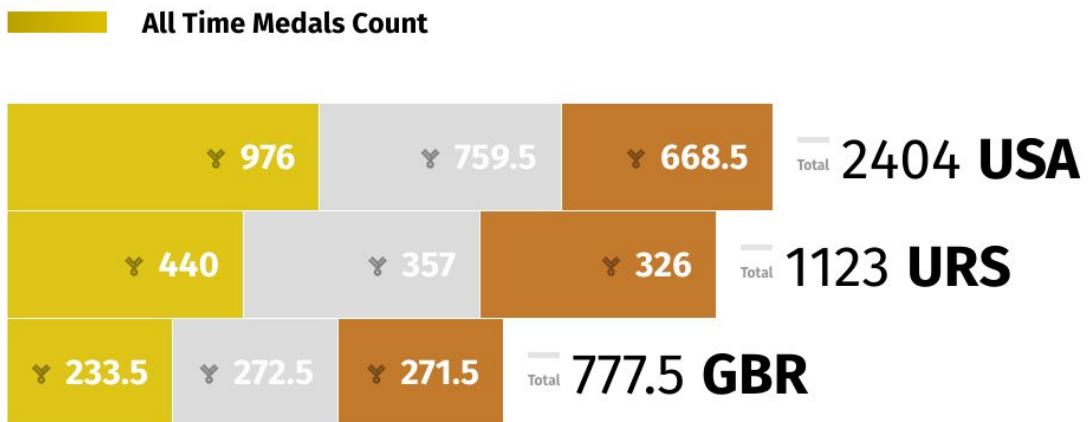
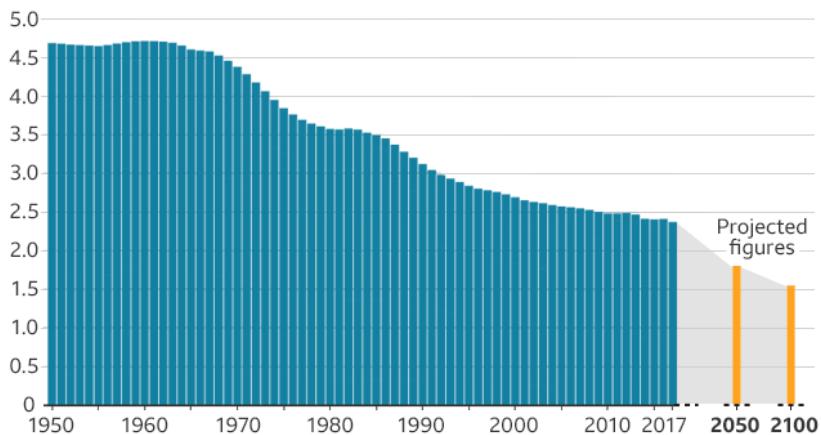


Figure 4h: **Stage:** Plotting **Category:** Inconsistency **Tag:** Misrepresentation
Tag Description: Misrepresentation occurs when the value labels provided do not match the visual encoding. For example, the data values may be drawn disproportionately or not to scale, thus intentionally or accidentally to cause the data to be misrepresented.
Visualization Description: The data is misrepresented in this medal count visualization because the bars and sub-bars have not been drawn to scale.

Women are having fewer children

Global fertility rate (livebirths per woman)



Source: Institute for Health Metrics and Evaluation at the University of Washington

BBC

Figure 4i: **Stage:** Plotting **Category:** Inconsistency **Tag:** Inconsistent Tick Intervals

Tag Description: Inconsistent axis ticks refer to cases with varying intervals between the ticks.

Visualization Description: In this visualization of the global fertility rate, the years at the right end of the graph are 2050 and 2100, a 50-year interval, but this interval is similar in length on the axis to the five-year intervals. Although there are visual cues that prompt the reader to pay particular attention to these two projected data points, it is still misleading.

The Middle Class Tax Target

The amount of total taxable income (left scale) for all filers by adjusted gross income level for 2008

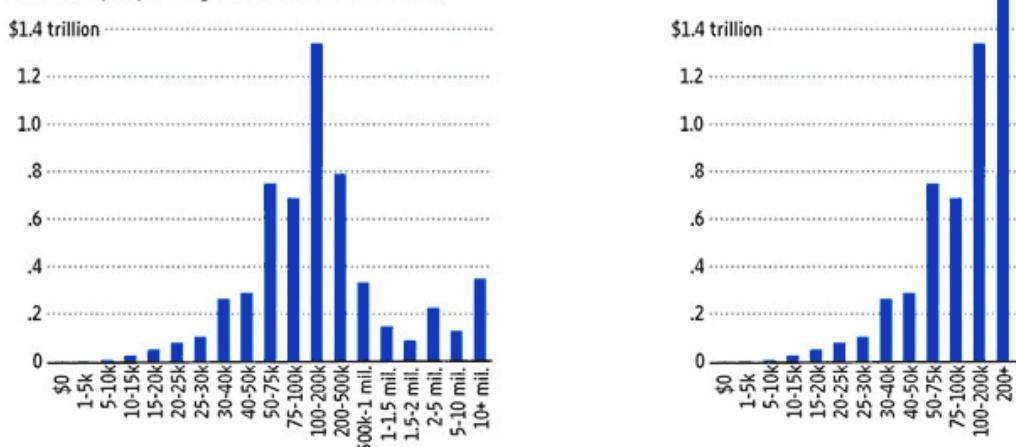
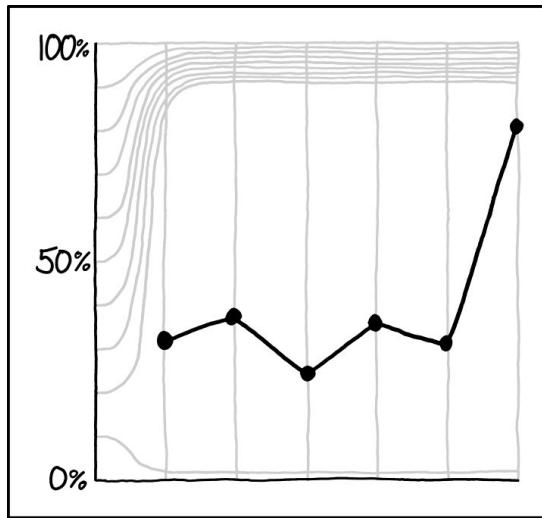


Figure 4j: **Stage:** Plotting **Category:** Inconsistency **Tag:** Inconsistent Binning Size

Tag Description: Inconsistent binning size occurs when there are variations in the boundaries of the binning groups.

Visualization Description: In the case of this tax target visualization, the income groups are binned unevenly, which may cause confusion and errors in interpretation.



PEOPLE HAVE WISED UP TO THE "CAREFULLY CHOSEN Y-AXIS RANGE" TRICK, SO WE MISLEADING GRAPH MAKERS HAVE HAD TO GET CREATIVE.

Figure 4k: Stage: Plotting Category: Inconsistency Tag: Changing Scale

Tag Description: Changing scale refers to instances where the scale being used changes midway through the visualization.

Visualization Description: In the above visualization, which it must be noted is clearly not based on real data, the scale changes from the first data point. This illustrates how this distortion trick can be used in real visualizations.

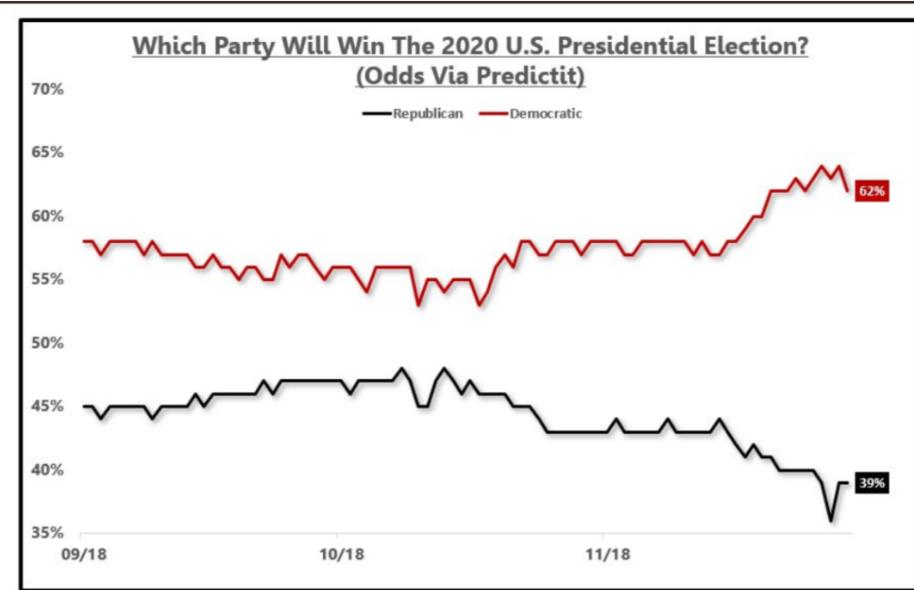


Figure 4l: Stage: Plotting Category: Inconsistency Tag: Violating Color Convention

Tag Description: When the colors selected do not match our collective expectation for the category they represent, it is deemed that they violate color convention, which can impact comprehension.

Visualization Description: Conventionally, the Republican party is represented by the color red, and the Democratic party is represented by the color blue. In this visualization of the 2020 American presidential election, the two are switched, breaking the color convention and potentially misleading the reader.

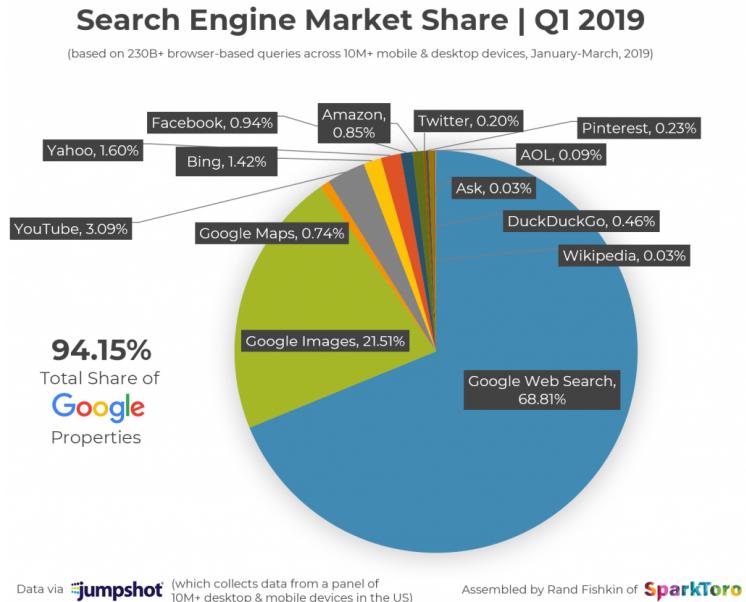


Figure 4m: **Stage: Plotting Category: Inconsistency Tag: Inconsistent Grouping**

Tag Description: Inconsistent grouping is when some of the entities being visualized are grouped while others are not.

Visualization Description: In the case of this visualization, Google is split into multiple groups—Google Web Search, Google Images, Google Maps, and YouTube, while Bing, which provides similar services, and the other entities all remain grouped. This could mislead the reader into thinking that Google has a smaller percentage of the pie than it really does.

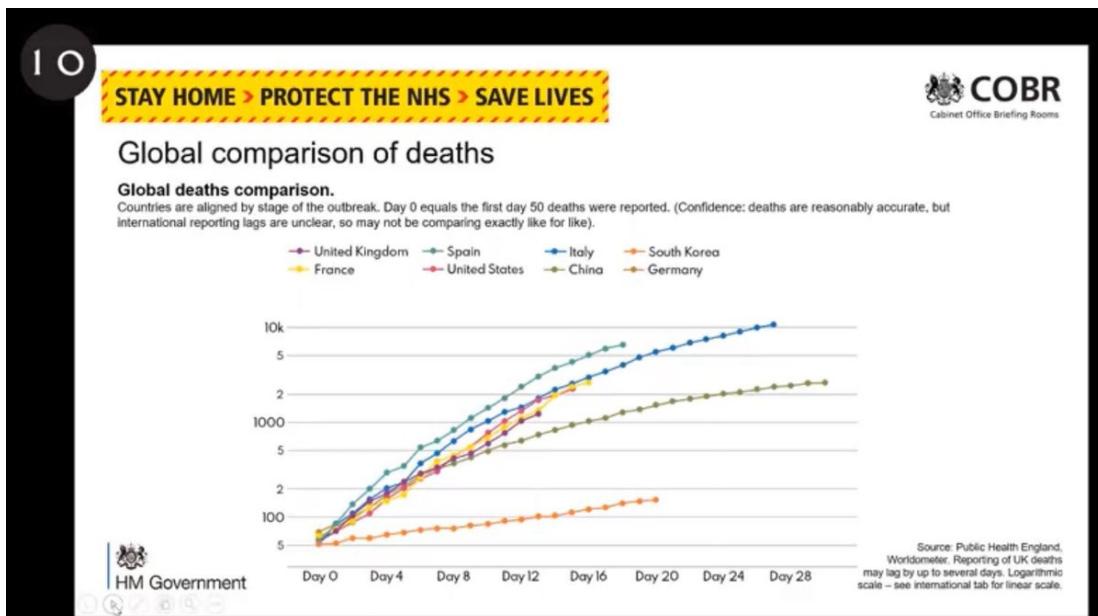


Figure 4n: **Stage: Plotting Category: Inconsistency Tag: Inconsistent Tick Labels**

Tag Description: Inconsistent tick labels mean that the tick labels along the axis are not all presented in the same format.

Visualization Description: In order to be clear and consistent, in this global comparison of deaths, the tick labels should be 50, 100, 200, 500, 1000, 2000, 5000, and 10000. An inconsistent approach to these labels will inevitably cause confusion.

The White House projects 100,000 to 200,000 Covid-19 deaths

At least. And outside experts say that's a reasonable estimate.

By Brian Resnick | @b_resnick | brian@vox.com | Mar 31, 2020, 7:40pm EDT

f t SHARE

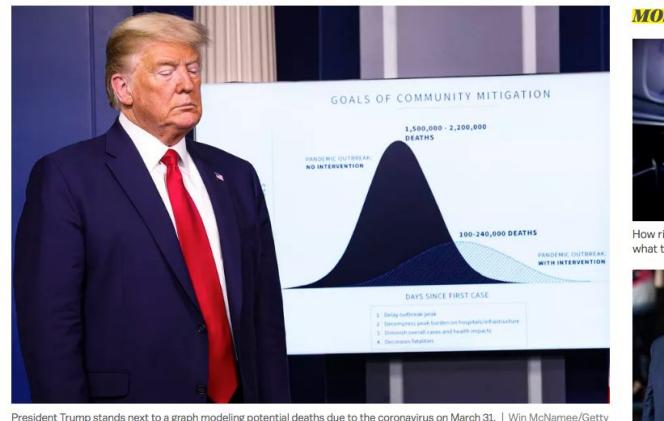


Figure 4o: Stage: Plotting Category: Inconsistency Tag: Inconsistent Value Labels

Tag Description: Inconsistent value labels refers to instances where a visualization's value labels are inconsistently annotated.

Visualization Description: In the visualization above, we can see that the label “100-240,000 deaths” is inconsistent with the label at the peak of the curve, which uses the full number each time: “1,500,000-2,200,000 deaths”. This allows the label to be misread as one hundred deaths instead of one hundred thousand deaths, which is a significant difference.

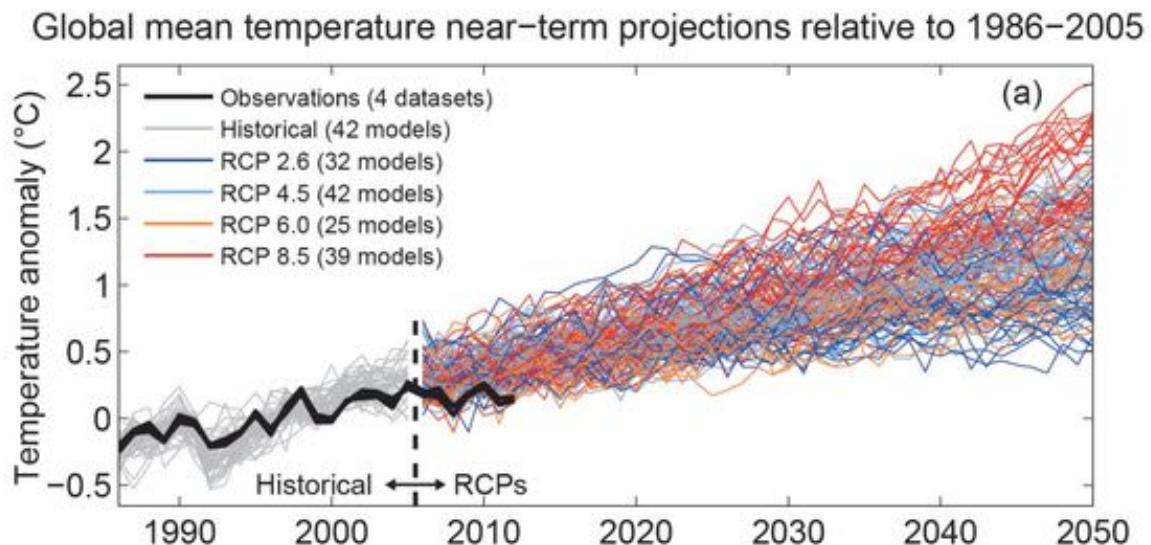


Figure 4p: Stage: Plotting Category: Chaotic Canvas Tag: Cluttering

Tag Description: Cluttering means too many data points or series presented together in one place.

Visualization Description: In this visualization of temperature projections, the lines are so muddled together that drawing any meaningful insights from the data would be impossible.

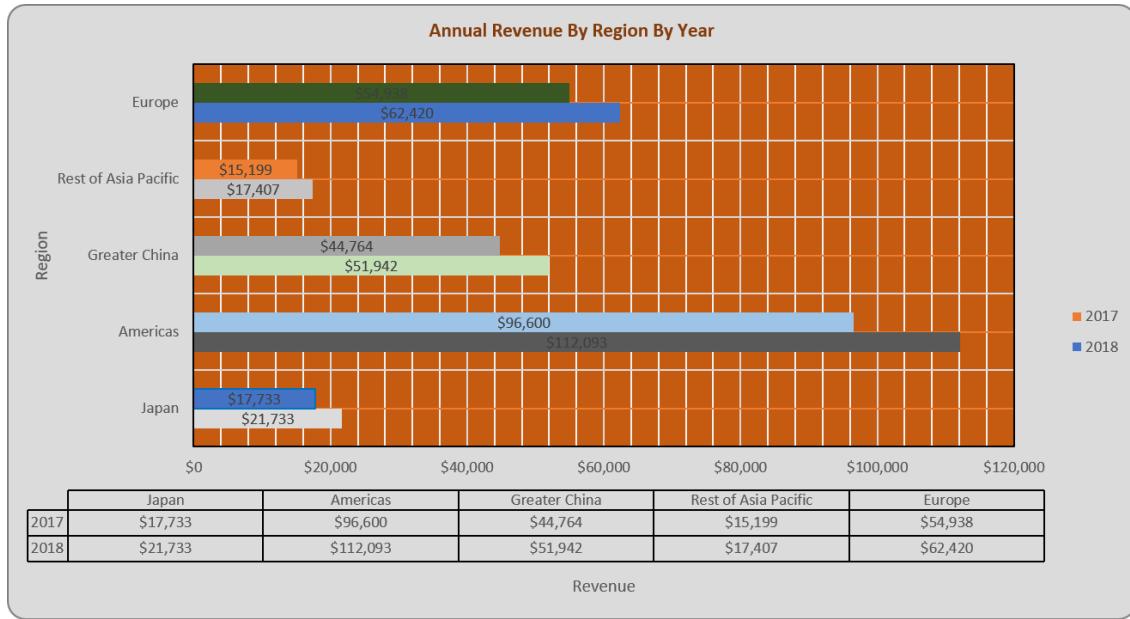


Figure 4q: Stage: Plotting Category: Chaotic Canvas Tag: Confusing Legend

Tag Description: The legend associated with a chart is not understandable due to the author's ordering, description, or a mismatch of the visual marks.

Visualization Description: In the above financial visualization, the legend indicates the relevant years, but years are not used anywhere in the chart. There is a clear mismatch, which confuses the reader and renders the legend meaningless.

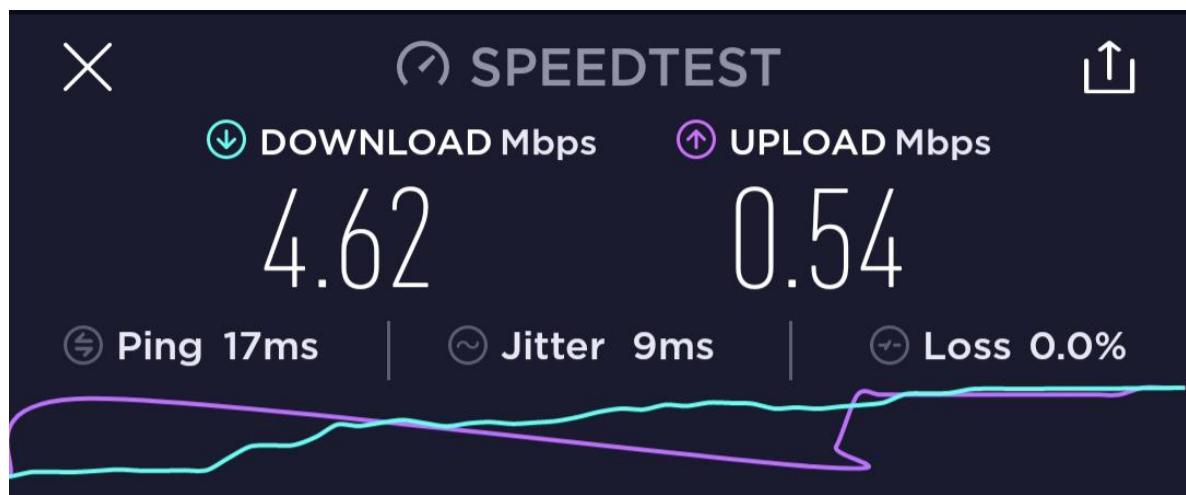


Figure 4r: Stage: Plotting Category: Chaotic Canvas Tag: Plotting Error

Tag Description: A plotting error is a glitch in a chart that occurs as a result of software bugs or other reasons.

Visualization Description: In this line chart, we can see that the line has curved backward in time, and there are no other reasons than a bug in the program that can explain it.

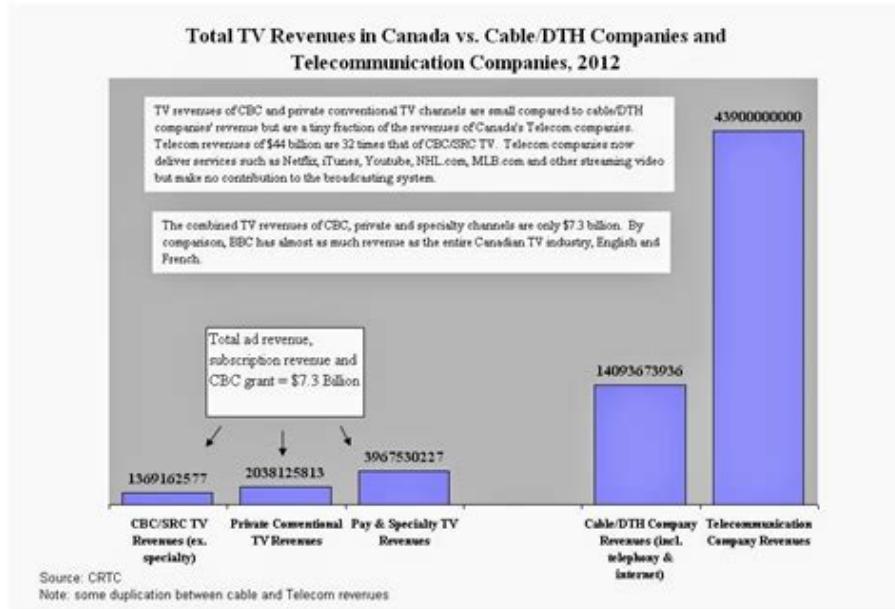


Figure 4s: Stage: Plotting Category: Chaotic Canvas Tag: Missing Abbreviation

Tag Description: An abbreviation is deemed missing when the text is unnecessarily long, and an abbreviation could have been applied.

Visualization Description: The values are presented in their full form in the above visualization. Although this is very precise, it does make it very difficult for the reader to comprehend the information and make comparisons.



Figure 4t: Stage: Plotting Category: Chaotic Canvas Tag: Misalignment

Tag Description: In the instance of misalignment, items or labels can intentionally or accidentally be misaligned. This results in poor readability, making performing an accurate comparison difficult.

Visualization Description: In this visualization of political support growth, the starting point of the bars is not aligned. They do not have the same baseline.

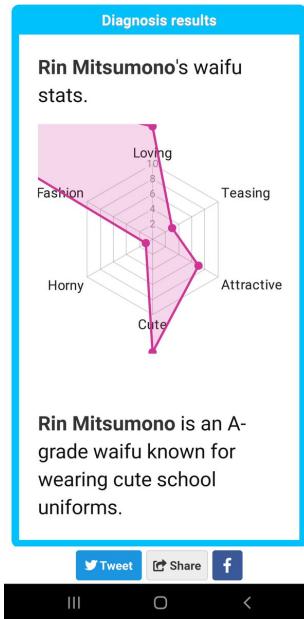


Figure 4u: **Stage:** Plotting **Category:** Chaotic Canvas **Tag:** Plotting Out of Chart

Tag Description: Plotting out of the chart occurs when the data value presented in the chart extends beyond the axis range.

Visualization Description: In the above radar chart, the values associated with “cute,” “fashion,” and “loving” extend beyond the axis range, rendering the data point meaningless.

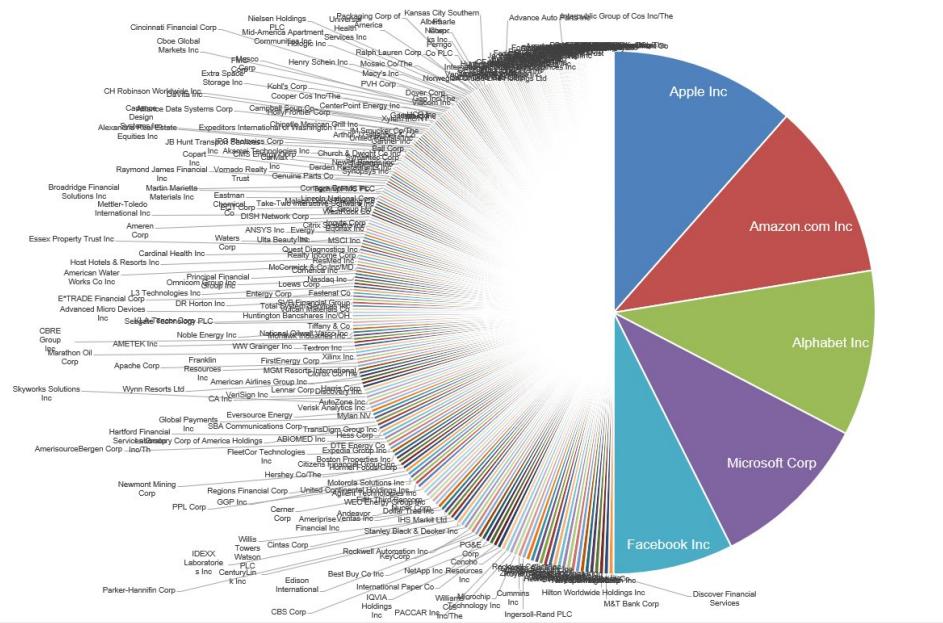


Figure 4v: **Stage:** Plotting **Category:** Chaotic Canvas **Tag:** Illegible Text

Tag Description: Illegible text refers to text that overlaps itself or other text and thus becomes unreadable.

Visualization Description: In the above visualization, there are so many text labels on the left side that many overlap and become completely unreadable. The reader would struggle to extract meaningful information as a result.

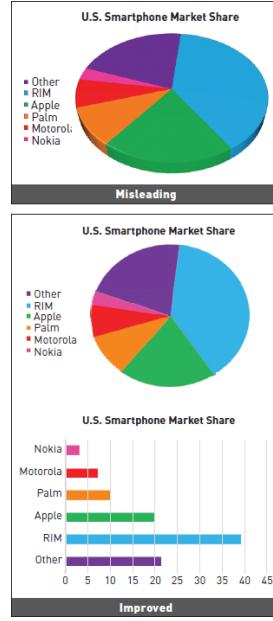


Figure 5a: Stage: Perception Category: Visual Illusion Tag: 3D

Tag Description: For 3D, the closer something is, the larger it appears, despite being the same size in 3D perspective.

Visualization Description: In the pie charts of smartphone market share above, the green section of the pie chart is larger in the 3D pie than its true size in the 2D pie.

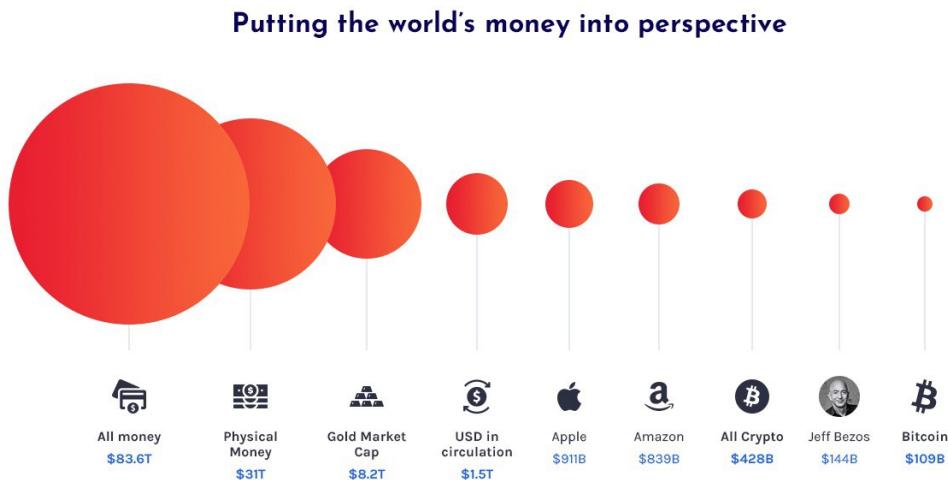
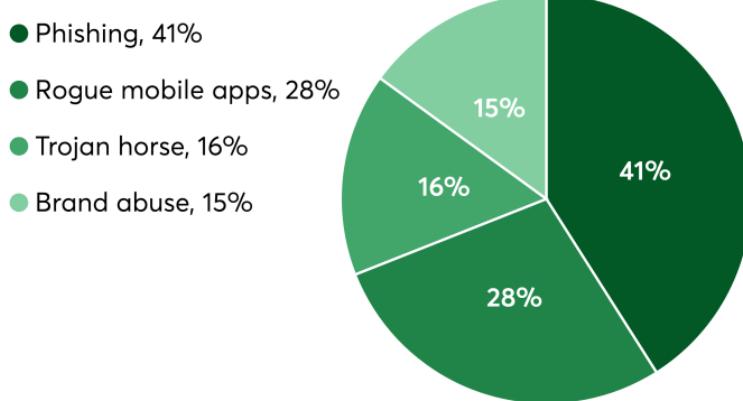


Figure 5b: Stage: Perception Category: Visual Illusion Tag: Area Encoding

Tag Description: According to Stevens' power law, the exponents are ~ 0.7 for area, which means linearly encoding the values as areas leads the readers to consistently underestimate the values. We should avoid using area encoding when there are better alternatives.

Visualization Description: In the visualization above, the amounts of money are encoded by area. Comparison on the size of the circles is inaccurate, and there are other better alternatives, for example, a bar chart.

Top global fraud types in Q2 2018



Source: RSA

Figure 5c: Stage: Perception Category: Visual Illusion Tag: Ineffective Color Scheme

Tag Description: In some cases, the color scheme selected is not effective for the encoded data. Examples of this can include rainbow colors, categorical colors on sequential data, and sequential colors on categorical data.

Visualization Description: In the above visualization, we can see that “Phishing,” “Rogue mobile apps,” “Trojan horse,” and “Brand abuse” are distinct categorical values. However, the author has chosen to use a sequential color scheme, which is not the most effective way to present this data.



Figure 5d: Stage: Perception Category: Visual Illusion Tag: Pictorial Area Encoding

Tag Description: Pictorial area encoding refers to the use of pictorial images as a method of encoding data—for example, using logos to represent companies.

Visualization Description: In this visualization, it is clear that Starbucks' sales is one-tenth of McDonald's sales, but the difference in area (in terms of how the logos are depicted on the graph) is far greater than 10 times.



Figure 5e: Stage: Perception Category: Visual Illusion Tag: Inappropriate Use of Smoothing
Tag Description: Line smoothing techniques cause interpolation between data points and can even create non-existent data points.

Visualization Description: The seven-day trend chart above indicates that the highest temperature is reached at midnight between Monday and Tuesday, which is completely incorrect.



Figure 5f: Stage: Perception Category: Visual Illusion Tag: Distractive Value Labels
Tag Description: The size of the value label can provide visual cues that can dominate the data-encoded geometric objects and affect the perceived data size.

Visualization Description: In the above visualization, the larger font size used for the value label 43% is compelling, and we perceive the bar that it annotates as larger. Additionally, the axis is truncated.

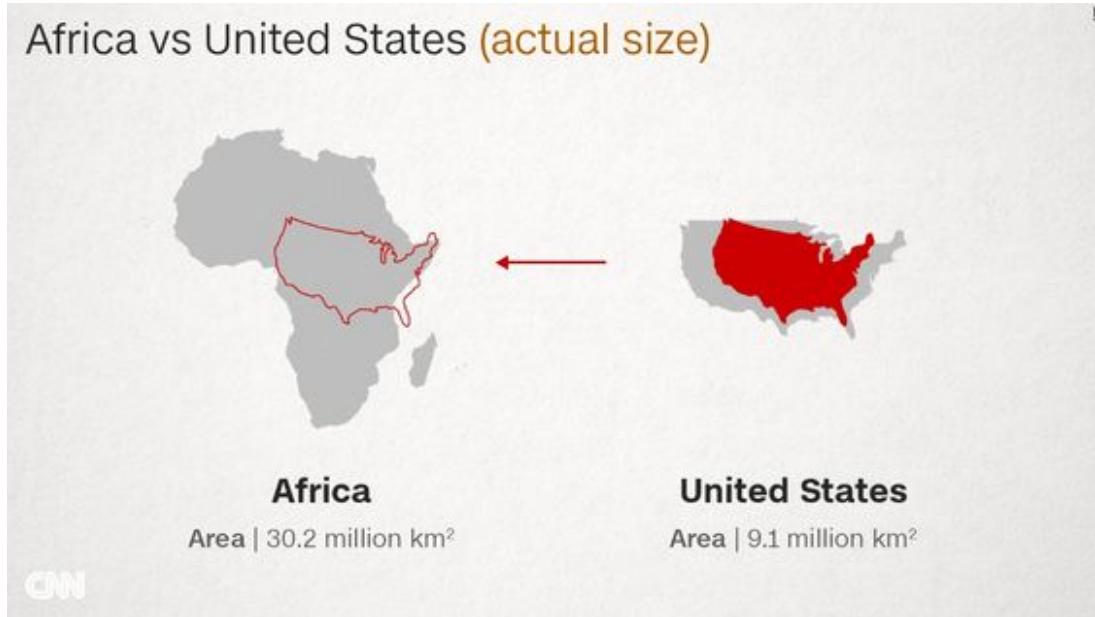


Figure 5g: Stage: Perception **Category:** Visual Illusion **Tag:** Map Projection Distortion

Tag Description: When we project a 3D object, such as the spherical Earth, onto a 2D plane, distortion of the area is unavoidable, making it impossible for the areas to be directly compared.

Visualization Description: In the African map visualization above, we can see a comparison of the areas of Africa and America. The commonly used Mercator projection distorts the size of geographic areas, causing the area of Africa to be understated frequently.

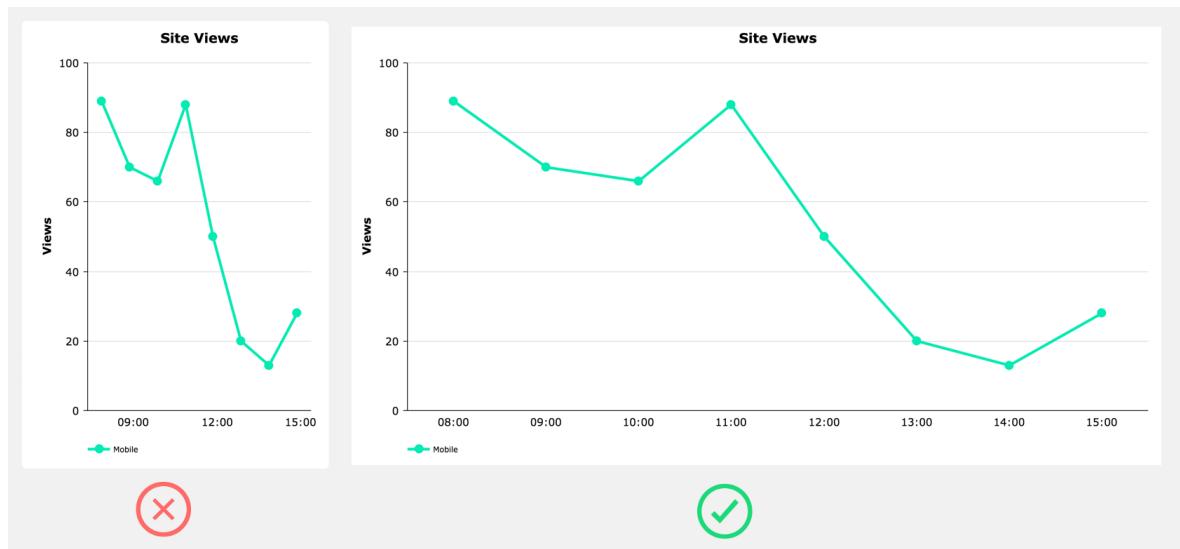


Figure 5h: Stage: Perception **Category:** Visual Illusion **Tag:** Inappropriate Aspect Ratio

Tag Description: In some instances, the choice of aspect ratio can affect the reader's perception of the chart.

Visualization Description: In this site views visualization, the aspect ratio drastically changes the perceived severity of the decline.

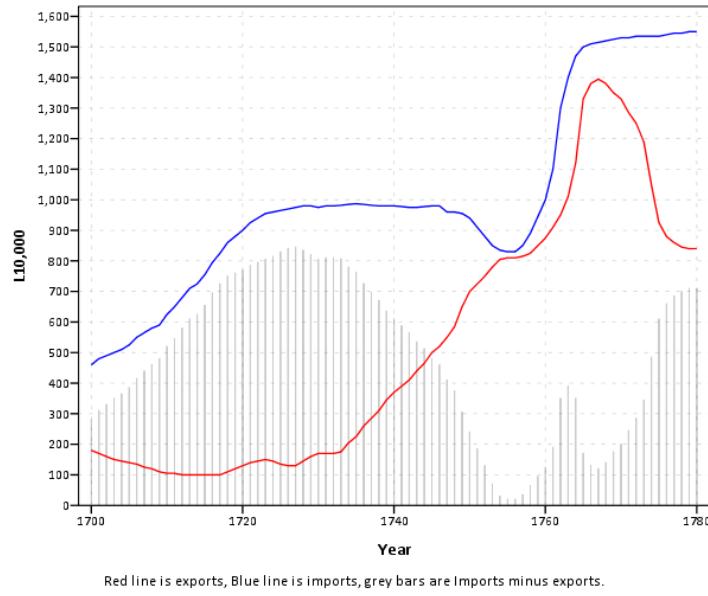


Figure 5i: Stage: Perception Category: Visual Illusion Tag: Sine Illusion

Tag Description: Instead of comparing only the two curves in terms of vertical deviations from one another, we tend to compare the curves in terms of the nearest location.

Visualization Description: The light-grey bars at the bottom indicate the differences between the two lines. For the period from 1762 to 1765, we perceive the differences between the two lines to be more or less constant, but it is actually enlarging.

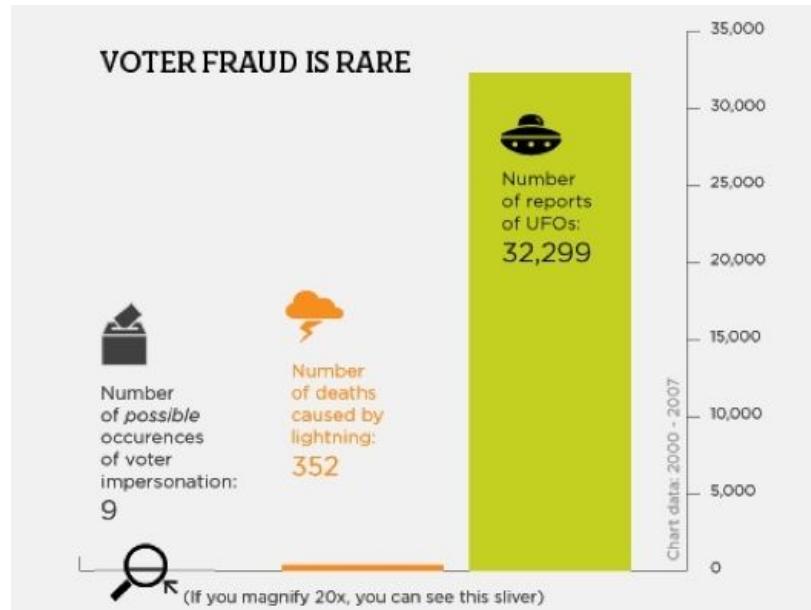


Figure 5j: Stage: Interpretation Category: False Linkage Tag: Invalid Comparison

Tag Description: An invalid comparison occurs when we link unrelated items by comparison.

Visualization Description: This visualization confidently declares that "Voter Fraud is Rare" by comparing it with deaths by lightning and the number of UFO reports, which are entirely illogical comparisons.

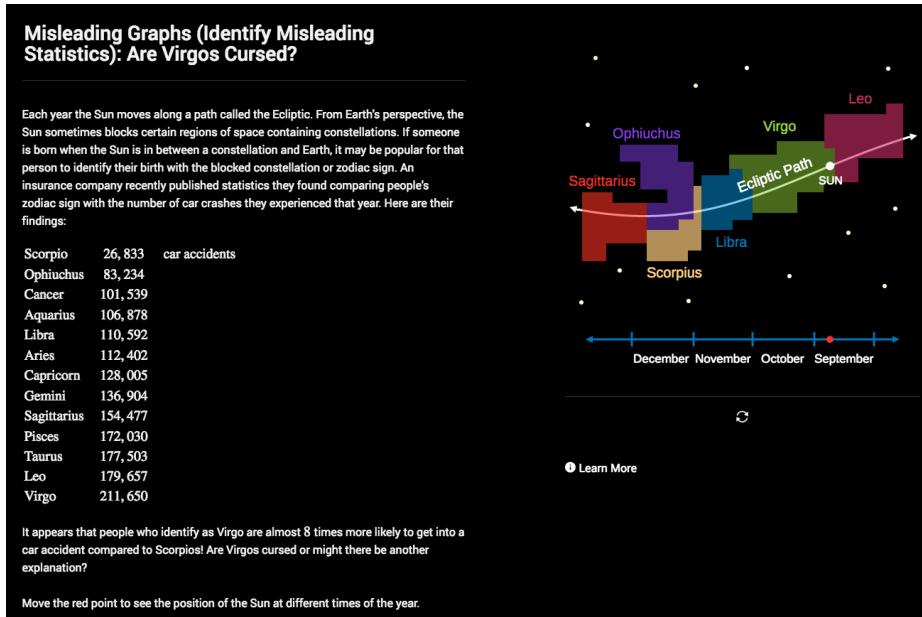


Figure 5k: Stage: Interpretation Category: False Linkage Tag: Correlation Not Causation
Tag Description: Correlation does not mean causation. Events may be falsely linked to explain the data or give a cause to the observed data.
Visualization Description: In this visualization of zodiac signs, an attempt is made to link people's zodiac sign with the number of car crashes they have experienced. The two do not have a causation relationship and suggesting causation is misleading.

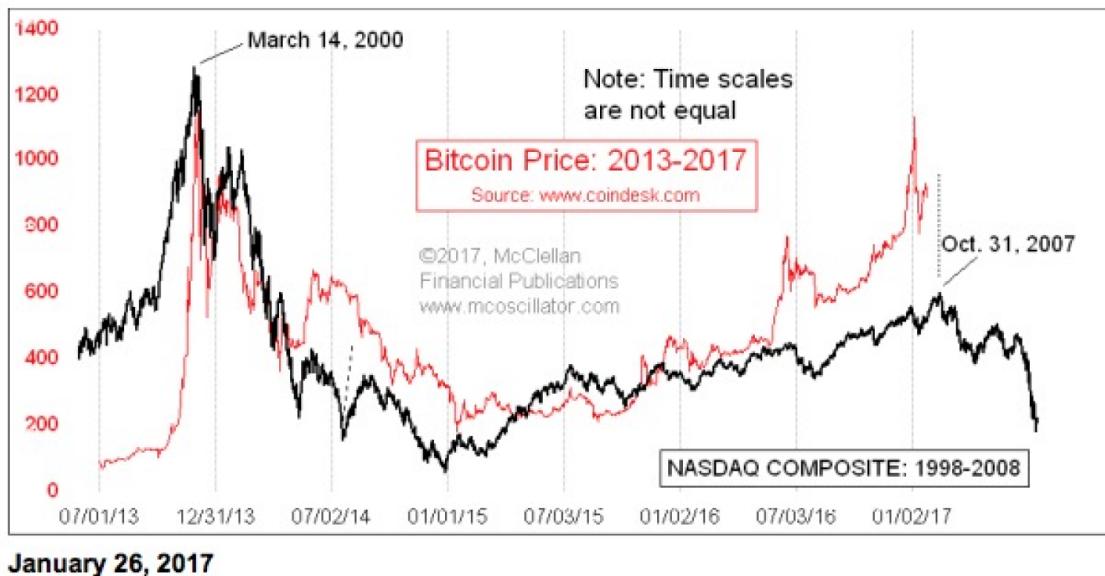


Figure 5l: Stage: Interpretation Category: False Linkage Tag: Pattern Seeking
Tag Description: Pattern seeking refers to seeking a correlation from historical data to predict whether a past event will happen again in the future, just because the pattern matches.
Visualization Description: In this financial visualization, the similar pattern between the two lines holds no indication of how the other line will develop in the future.

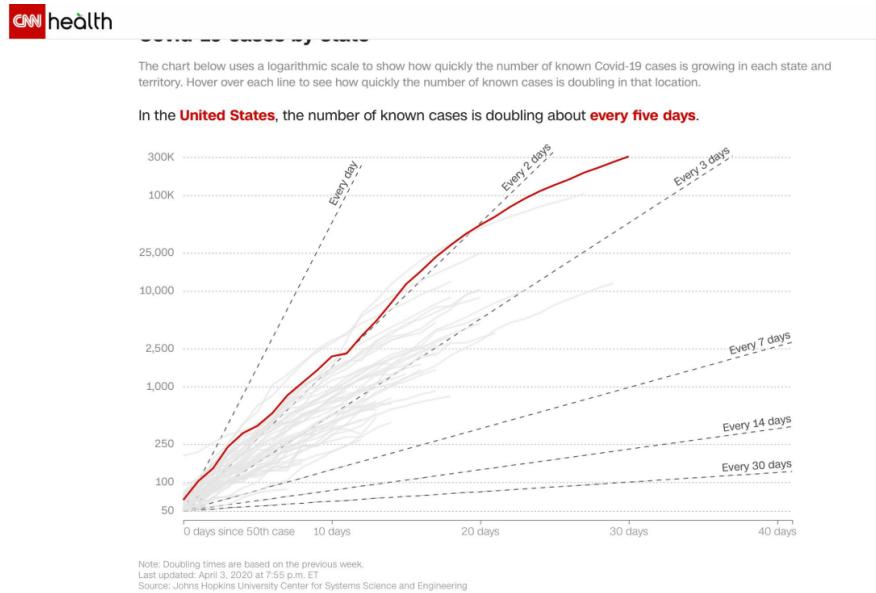


Figure 5m: Stage: Interpretation Category: Power of Words Tag: Misleading Claim

Tag Description: Authors may use visualizations to make a misleading claim. This claim is often delivered via their interpretation in the associated text description.

Visualization Description: The chart above shows that the number of COVID-19 cases in the USA doubled every two to three days. However, the author's text states that this occurs every five days, thus intentionally or accidentally misleading the reader with an incorrect claim.

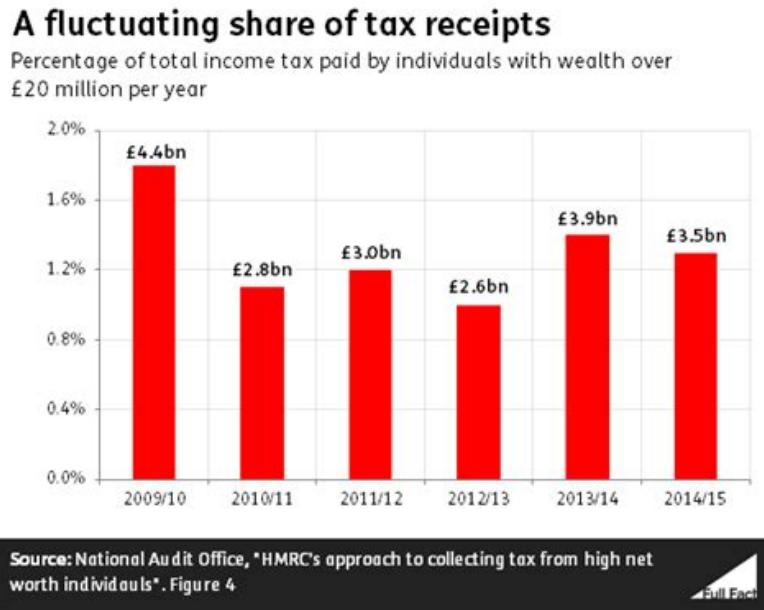


Figure 5n: Stage: Interpretation Category: Power of Words Tag: Misleading Annotation

Tag Description: The annotation in a chart provides misleading information instead of aids to the reader to comprehend the chart.

Visualization Description: In the case of this visualization, the annotation on the bars are the absolute amounts of tax paid, but the reader should instead be guided to read the percentages that the bars represent.

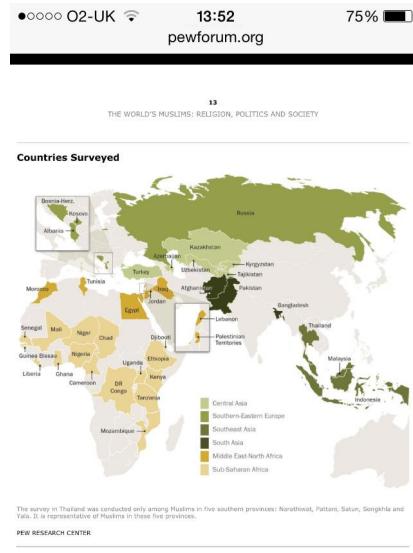


Figure 5o: Stage: Interpretation Category: Power of Words Tag: Misleading Title

Tag Description: A misleading title is one that does not match the message presented in the chart.

Visualization Description: The map includes only Africa, Asia, and Europe, but the title states “world” while missing the Americas and Oceania.

Healthcare Victim Organization Relationship to Insiders

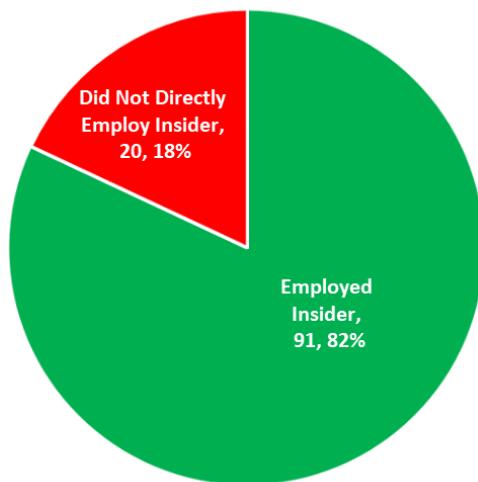


Figure 5p: Stage: Interpretation Category: Power of Words Tag: Misleading Value Labels

Tag Description: Value labels can be misleading when they confuse the reader or fail to communicate clearly.

Visualization Description: In this pie chart, the value labels 20 and 91 are the absolute numbers, while 18% and 82% are the encoded value of the pie sections. This unclear communication may mislead the reader and cause errors in interpretation.

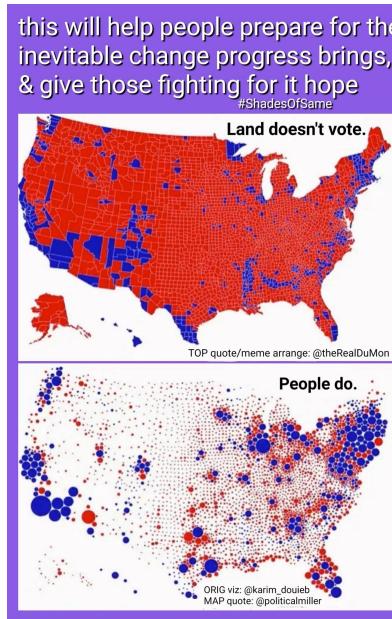


Figure 5q: **Stage:** Interpretation **Category:** Hidden Information **Tag:** Hidden Distribution
Tag Description: There is an underlying distribution not shown in the visualization, which inevitably results in a misperception of the true picture.

Visualization Description: In this political map visualization, the use of dichotomous colors obscures the underlying voting distribution. The visual marks of blue/red color on the map conceal all of the votes for the minority candidate in that region.

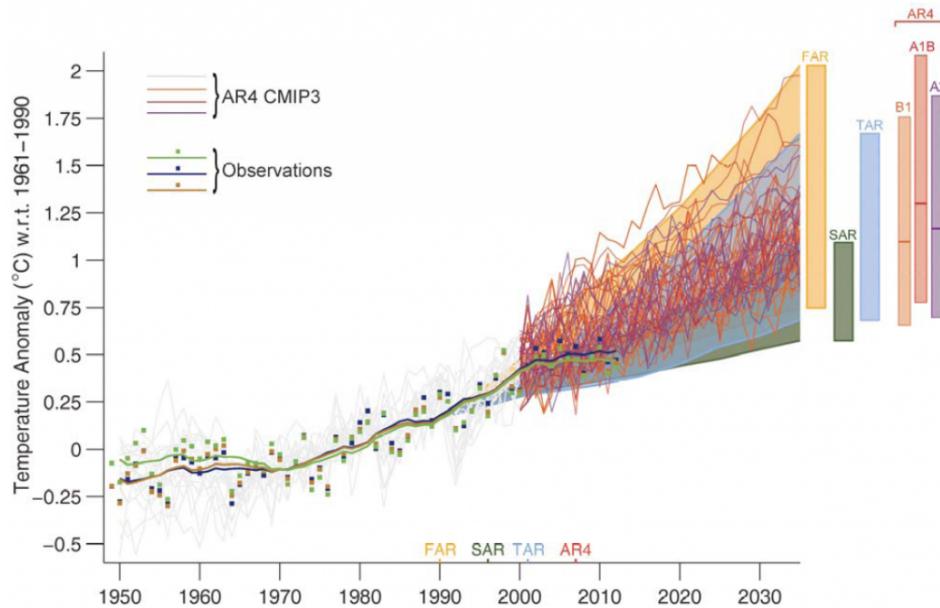


Figure 5r: **Stage:** Interpretation **Category:** Hidden Information **Tag:** Overplotting
Tag Description: Overplotting occurs when the visual marks that are being overlapped are hidden from the reader, resulting in understated areas or obscured data points.

Visualization Description: In this visualization, the areas are completely overlapping each other and the true range is obscured.

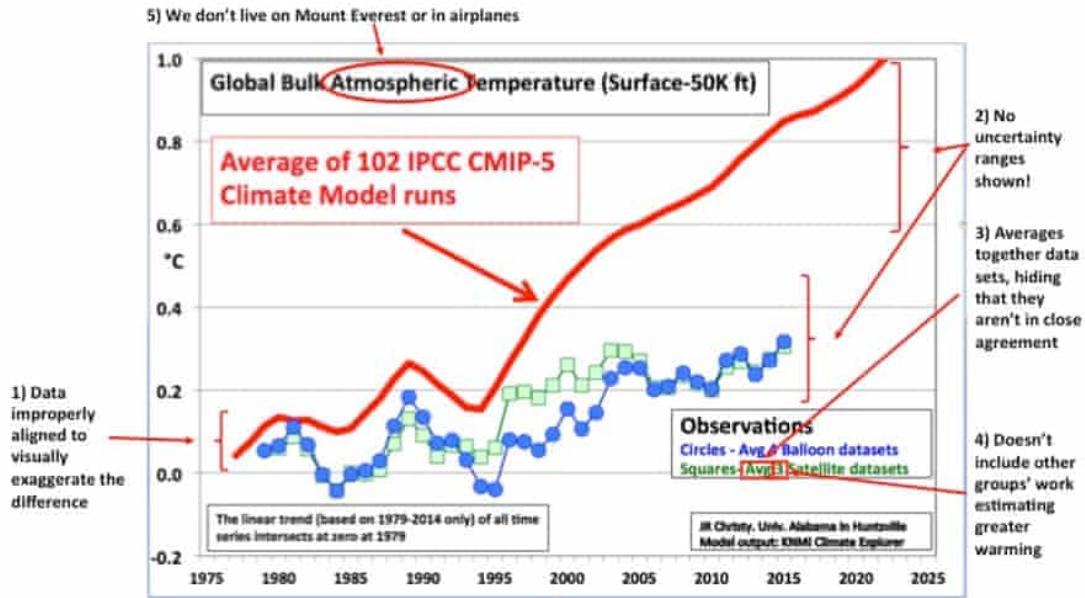


Figure 5s: Stage: Interpretation Category: Hidden Information Tag: Hidden Uncertainty

Tag Description: Hidden uncertainty is the case that uncertainty is not being visually represented when it matters.

Visualization Description: In this atmospheric temperature-related visualization, the uncertainty range of the prediction is not depicted in the chart. This could mislead the reader into thinking that no uncertainty exists.

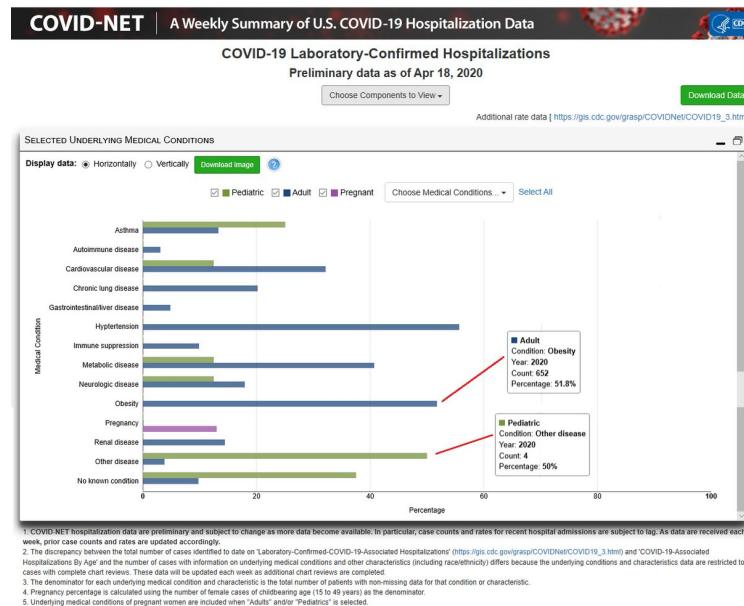


Figure 5t: Stage: Interpretation Category: Hidden Information Tag: Hidden Population Size

Tag Description: In some cases, the population size is highly relevant to the statistical summary, but not shown in the visualization.

Visualization Description: The two highlighted bars in this visualization have similar mean percentages but the underlying population size is actually very different, which could incorrectly alter the reader's interpretation of the data.