



Máster Ciencia de Datos - *Data Science*

Visualización de Datos / Profesor colaborador: Jaume Nualart Vilaplana

Fernando Antonio Barbeiro Campos - fbarbeiro@uoc.edu

A5: Análisis de una visualización de datos (PEC2)

Entrega prevista: 21 de Noviembre de 2018.

Índice General

A5: Análisis de una visualización de datos (PEC2)	0
Índice General	1
Introducción	2
Enunciado	2
Respuestas	4
Buena Práctica	4
Justificativa por se tratar de una buena práctica	5
Mala Práctica	7
Justificativa por se tratar de una mala práctica	9
Propuesta del rediseño	10
Referencias bibliográficas	13

Introducción

Enunciado

En esta actividad cada estudiante deberá buscar y seleccionar dos visualizaciones, una como ejemplo de buena práctica de visualización de datos y otra como ejemplo de mala práctica. Cada crítica deberá constar de:

- Datos sobre la visualización: dónde y cuándo ha sido usada (en qué medio, web, etc.) y un enlace a la visualización, así como una o más capturas de pantalla que la muestren.
- Una descripción de la pregunta que "contesta" la visualización. Por ejemplo: "¿Cómo evoluciona la tasa de paro juvenil en España?", es la pregunta que contesta el siguiente gráfico:

TASA DE PARO JUVENIL

En % de la población activa



- Una descripción de no más de una página (para cada visualización) justificando por qué se considera que se trata de una buena o mala visualización de datos. Entre otros, se debe incluir si la visualización consigue responder o no a la pregunta anterior y porqué.

Además, para la visualización considerada como mala práctica, se deberá hacer una propuesta de rediseño. Es decir, se deberán proponer mejoras que ayuden a responder mejor la pregunta que debe responder la visualización. Podréis describir las mejoras que incorporaríais (¿utilizaríais otra visualización totalmente distinta? ¿cambiaríais los colores? ¿quitaríais esa molesta leyenda que oculta parcialmente el contenido?, etc.). Se valorarán muy positivamente las entregas que incorporen el propio rediseño, aunque sea hecho a mano con lápiz y papel y lo incorporen al documento como una imagen fotografiada o escaneada.

En el [blog de Stephen Few](#) se pueden encontrar ejemplos de críticas de visualizaciones y de rediseños para que os inspiréis. No os limitéis a reproducir, intentad obviar los ejemplos "típicos".

Respuestas

1. Buena Práctica

Mi ejemplo de una increíble visualización de datos fue diseñado por [1] Bostock, Carter y Tse (2014) y fue publicada también en 2014 en “**The Upshot**” que es una empresa basada en datos, centrada en la política, análisis política y económica (empresa del grupo “**The New York Times**”).

Por aquí, el enlace de la visualización:

https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0

Y una captura de pantalla de la misma:

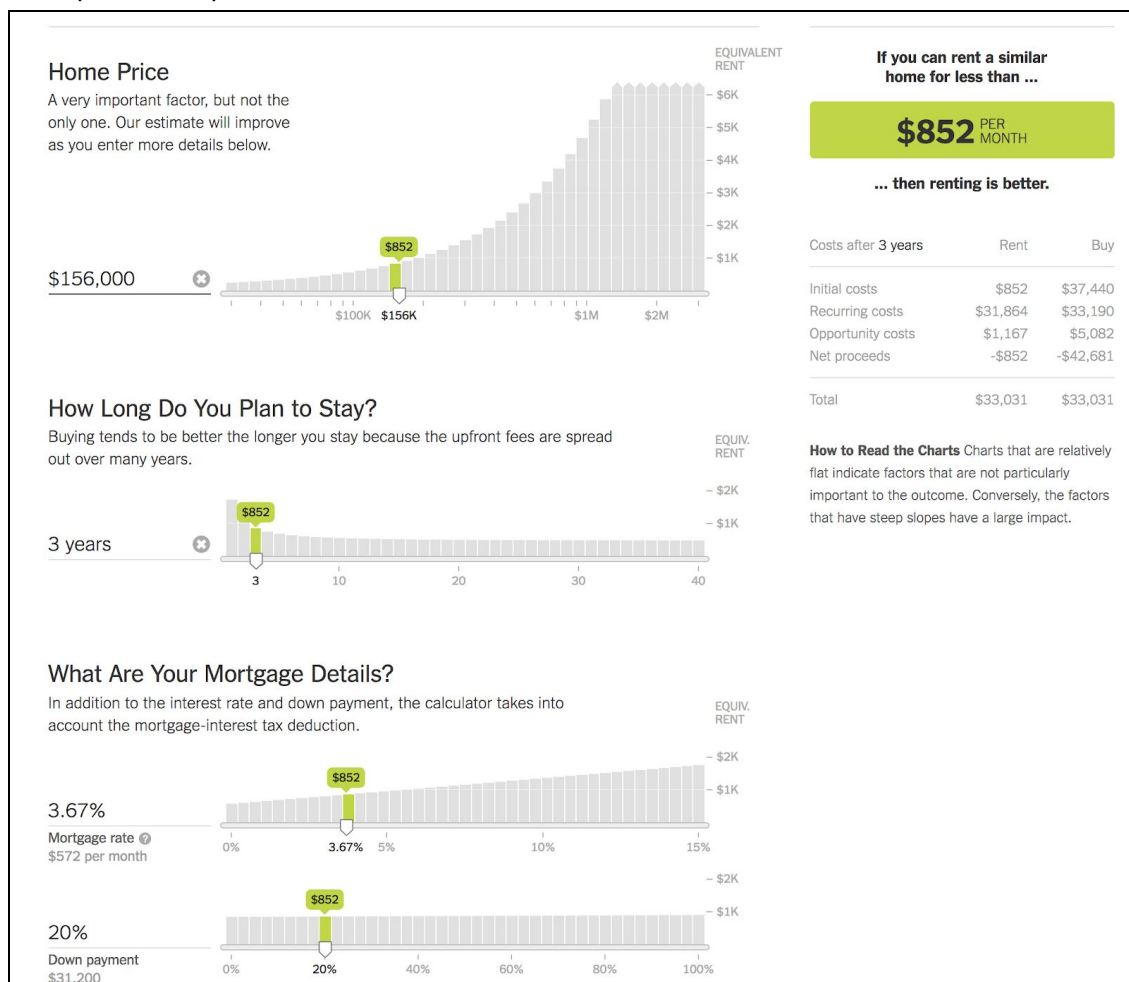


Figura 1: Buena práctica de visualización de datos. [1] Bostock, Carter y Tse (2014)

Un detalle interesante que el título de la web-page donde está publicada es precisamente la pregunta que la visualización busca contestar: “*Is It Better to Rent or Buy?*” o “**¿Es mejor alquilar o comprar?**”.

La idea es una calculadora de presupuesto para ayudar a los posibles propietarios de viviendas a saber si será mejor para ellos alquilar un piso o poseerlo de manera absoluta.

La calculadora incluye varios gráficos inclinados. Cada cuadro incluye un factor que afectará la cantidad que tendrá que pagar, como el coste individual de su vivienda y las tasas de su hipoteca.

Una escala móvil a lo largo de la parte inferior de cada tabla le permite ingresar datos diferentes, cambiando el "coste de alquiler por mes" en la lateral. Si puede encontrar una casa similar para alquilar por esa cantidad por mes o menos, es más rentable simplemente alquilar la casa.

a. Justificativa por se tratar de una buena práctica

Basado en la experiencia que tuve con la dataviz en foco y también considerando los puntos de una buena práctica de dataviz de [2] *Korada (2018)*, los puntos que aclaran que se trata de una buena práctica son:

- Los autores conocen perfectamente la audiencia y los graficos estan diseñados a medida para tal.
- El dashboard es clasificado (desde mi punto de vista, **analítico**) ofreciendo una variedad distinta de enfoques de investigación para un tema central específico con vistas contextuales corolarias limitadas. Aquí, basta mirar la capacidad de customizar cada variable del gráfico.
- El diseño trae soluciones interactivas

- Esta visualización es una herramienta increíblemente completa y útil

Dados los elementos comentados, claramente la visualización de datos contesta perfectamente la pregunta al que se propone a atacar. Dicho, haciendo uso de otra captura de pantalla, abajo, podemos ver que precisamente hace recomendaciones precisas en términos de la pregunta.

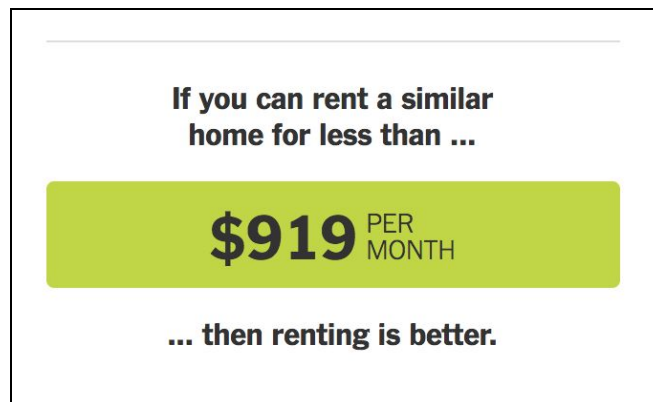


Figura 2: Buena práctica contestando a la pregunta principal. [1] *Bostock, Carter y Tse (2014)*

Un último punto interesante, después que había encontrado la dataviz aquí defendida, por casualidad he encontrado también en los comentarios de [3] *Darling (2016)* que clasificaba la dataviz como una de las más guays de 2016 (aunque el proyecto es originalmente de 2014).

2. Mala Práctica

La visualización que he elegido ha sido publicada en Octubre de 2018 vía web (puede que también haya salido en periódicos, teniendo en cuenta que el grupo que publicó actúa en los 2 sectores). Es del grupo “**Folha de São Paulo**”, o simplemente “**Folha**”, que es un periódico brasileño, el segundo diario de mayor circulación en Brasil. El enlace para la visualización está abajo:

<http://temas.folha.uol.com.br/eleicoes-2018-em-graficos/resultado/onda-de-direita-toma-o-pais.shtml>

P.D.: La parte de texto está en Portugués, sin embargo, el gráfico que explicaré contiene pocas palabras, de las cuales haré una traducción al explicar. De manera complementaria, aunque creo que el gráfico que enseñaré trae malas prácticas, en el propio enlace podemos ver ejemplos de dataviz que están sí bastante adecuados.

La principal pregunta que se busca contestar con la visualización es: **¿Cómo se distribuye los votantes de cada candidato en Brasil?**

Sin más preámbulos, vamos a una captura de pantalla que muestra la visualización:

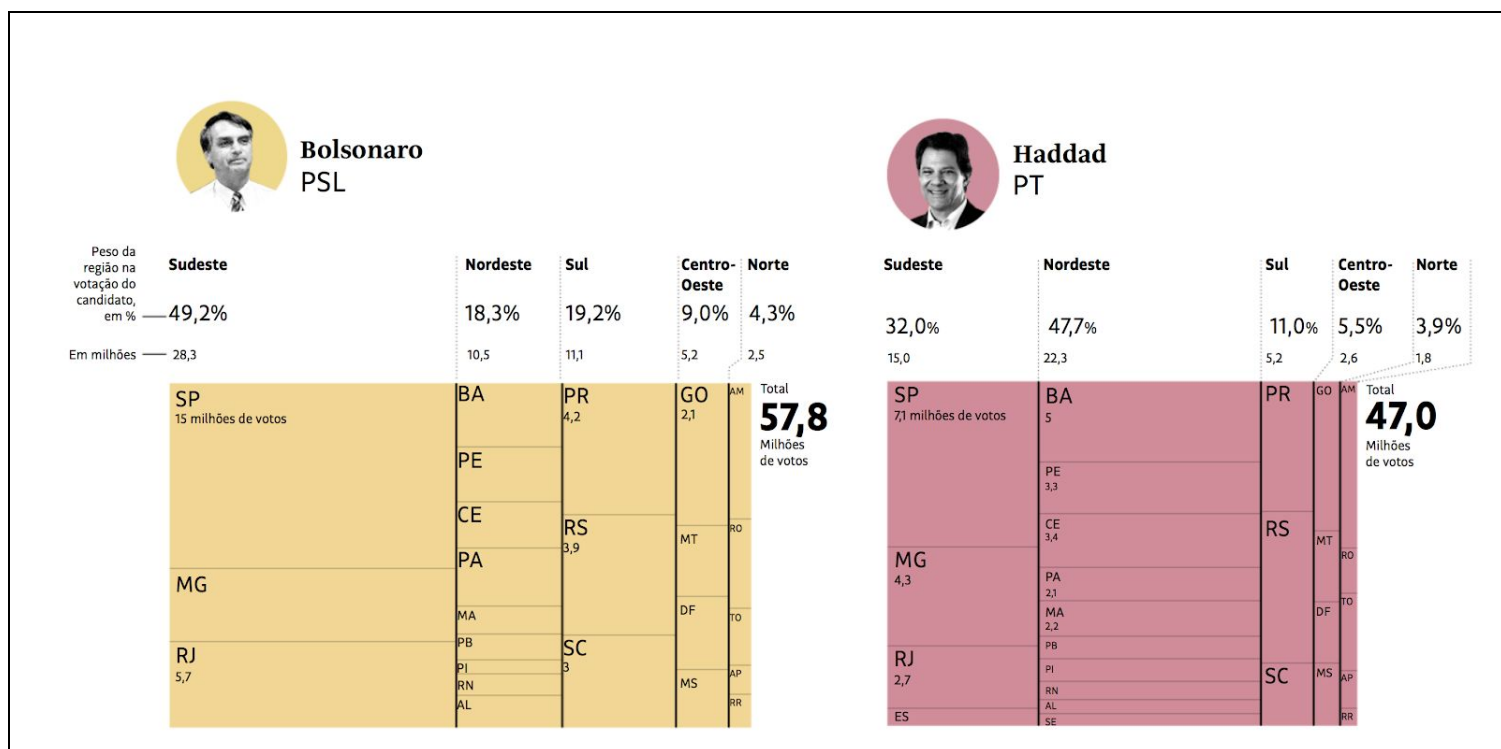


Figura 3: Visualización con mala práctica. [4] *Folha de São Paulo* (2018)

Antes de seguir con la justificativa, algunas explicaciones para no perjudicar el entendimiento:

- Bolsonaro y Haddad eran, obviamente, los candidatos.
- Hay 2 divisiones principales en el gráfico:
 - Regiones - Brasil está formado por 5 regiones (Sudeste, Nordeste, Sul, Centro-Oeste y Norte)
 - Los acrónimos, como SP, MG, etc representan cada estado dentro de una región.

Haciendo una analogía, sería como las comunidades autónomas y las provincias que tenemos en España.

- Los números presentados están en millones.

a. Justificativa por se tratar de una mala práctica

Entre los elementos que me hacen creer que no se trata de una buena práctica están:

- Hay incongruencia en la cantidad de estados representados, o sea, mientras para el candidato Bolsonaro (amarillo) notamos 24 estados, para el candidato Haddad (rojo) tenemos 26.
- La cantidad de votos por muchas veces tiene que ser inferida (o calculada) del total. Es decir, si necesito comparar para el estado de MG, tendría que calcular la resta del candidato en amarillo para poder comparar con el candidato en rojo – el mismo problema se repite por todo el gráfico.
- La proporción de los elementos presentados no es correcta. Por ejemplo, en la región “*Centro Oeste*” donde Haddad tiene precisamente la mitad de los votos de Bolsonaro, respectivamente 2.7 y 5.2 millones, el gráfico tiene una variación en pixels que no representa la mitad – en pixels, el adecuado sería 94 px del amarillo para 47px del rojo (y tenemos 94 amarillo para 42 rojo). Lo mismo ocurre para otras regiones como sur y también para estados individualmente.

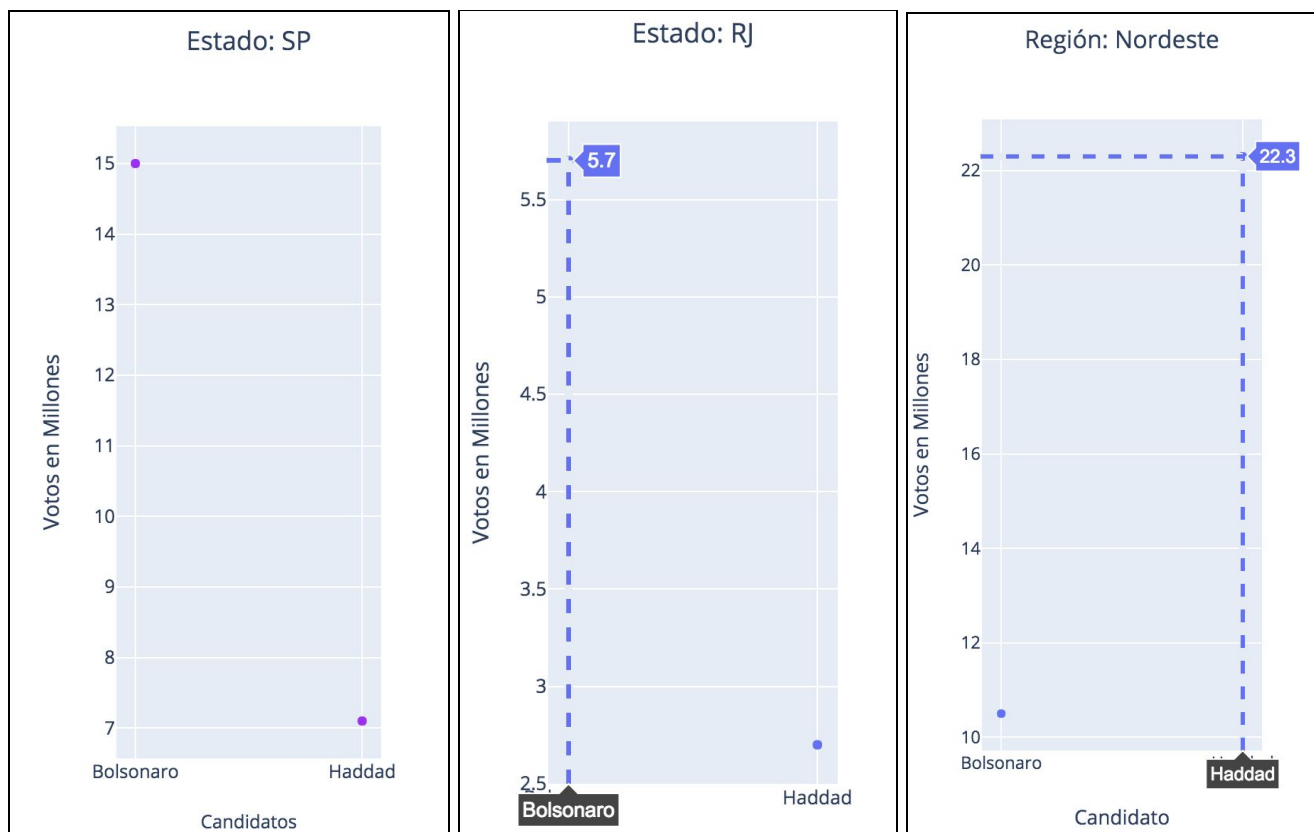
Por los motivos mencionados arriba, concluimos que la visualización solo consigue responder parcialmente a la pregunta: **¿Cómo se distribuye los votantes de cada candidato en Brasil?**

Especialmente porque las imprecisiones de proporción pasan una información incorrecta y porque quedan sitios no representados (o con información que se omitió) no dejando visualmente claro para el lector que se ha querido transmitir.

Vale resaltar que el problema está basado también en la definición de [5] *Cleveland y McGill (1984)* que clasifican la percepción visual de área (dataviz en cuestión) como una de las menos precisas en un ranking de las mejores interpretadas por nuestro cerebro. Veamos la propuesta de rediseño.

b. Propuesta del rediseño

Aun teniendo en cuenta el ranking de [5] *Cleveland y McGill (1984)*, una visualización donde la posición es presente suele ser el mecanismo más sencillo de percepción visual, por lo tanto, propondría una aproximación donde las diferencias de votos en los datos quedarían completamente clara por estado. Un ejemplo de que haría yo:



Obviamente, generaría un poco más de trabajo para crear los gráficos, sin embargo, creo que lograríamos contestar efectivamente a la pregunta sobre la diferencia por estado y, además, por región - facilitando en entendimiento y el objetivo de transmitir la información.

Otro punto que he tenido en cuenta ha sido en poner adecuadamente los *labels* en cada eje - especialmente porque si clasificamos simplemente por colores, no diseñamos para la deficiencia de la visión, es decir, cuando elegimos los colores

para una visualización, debemos acordarnos que una buena proporción de nuestros lectores puede, como un ejemplo, ser daltónicos.

Consideramos esto como la primera versión de mejora para atacar los problemas que inicialmente había listado.

Pero, honestamente, considerando que son 27 estados, más 5 regiones, la idea de crear 32 gráficos distintos parece demasiado fatigosa y lo es.

Entonces, me ha ocurrido de mejorar aún más la exhibición que teníamos hasta aquí (que resultó en la necesidad de buscar las informaciones que faltaban de la visualización original – comento simplemente para dejar claro que la dataviz inicial no estaba completa).

Afortunadamente, al buscar información de los votantes por estados, he encontrado una segunda dataviz que considero bastante más interesante (y tiene completa similitud con lo que buscaba yo para el diseño final). Vamos a echar un vistazo a la solución final:

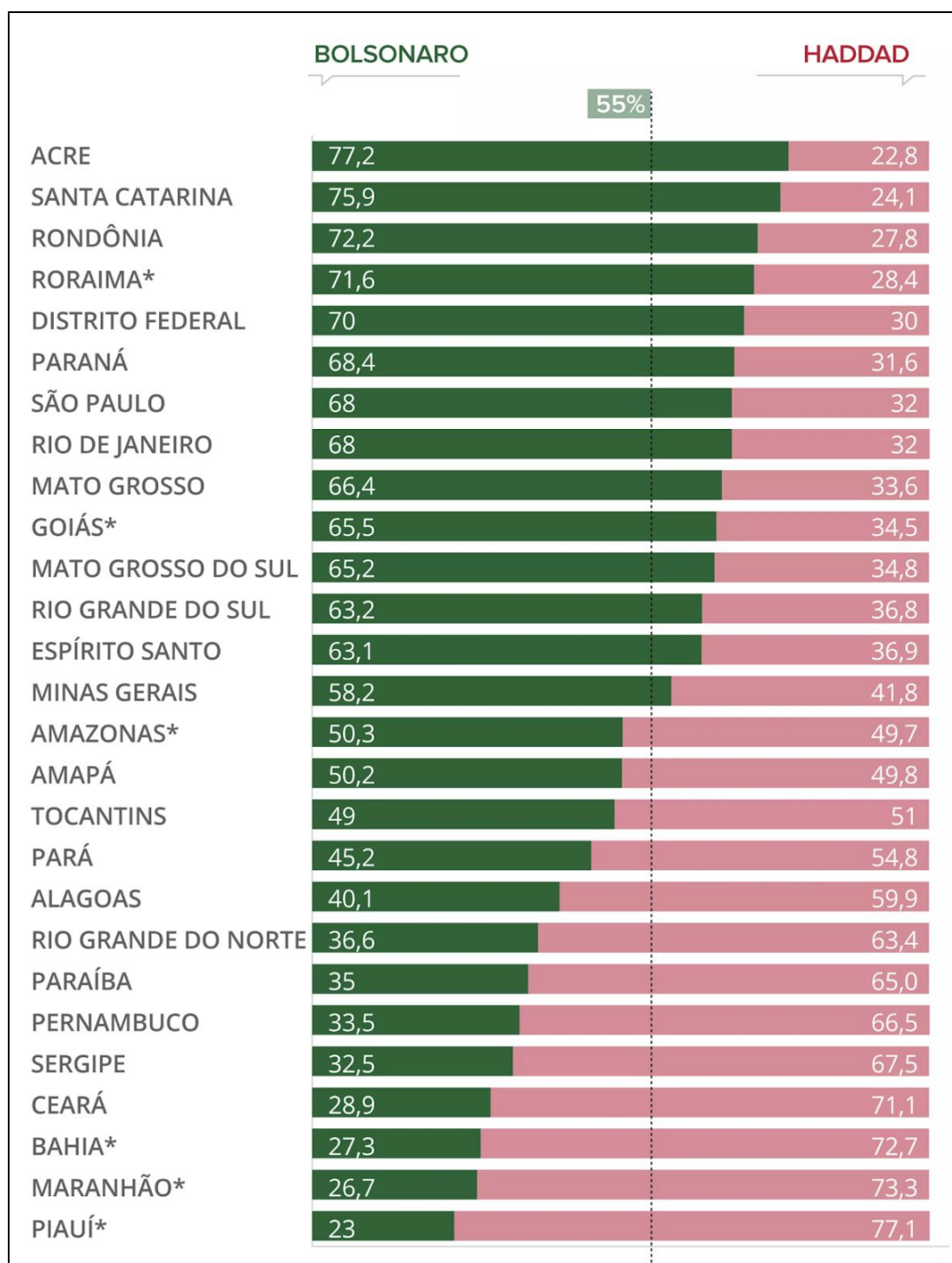


Figura 4: Ejemplo completo de dataviz después de aprimorarla. [6] Sorano (2018)

3. Referencias bibliográficas

- [1] **Bostock, M. Carter, S. y Tse, A. (2014).** *"Is It Better to Rent or Buy?"*. The Upshot - New York Times. [artículo en línea]. [Fecha de consulta: 10 de Noviembre del 2018].
<https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0>
- [2] **Korada, P. (2018).** *"5 Best Practices for Data Visualization"*. Oracle + DataScience.com. [artículo en línea]. [Fecha de consulta: 14 de Noviembre del 2018].
<<https://www.datascience.com/blog/data-visualization-best-practices>>
- [3] **Darling, K. (2016).** *"15 Cool Information Graphics and Data Viz from 2016"*. [artículo en línea]. [Fecha de consulta: 14 de Noviembre del 2018].
<<https://blog.visme.co/best-information-graphics-2016/>>
- [4] **Folha de São Paulo (2018).** *"Eleições 2018 em gráficos - Onda de direita toma o país"* - [artículo en línea]. [Fecha de consulta: 15 de Noviembre del 2018].
<<http://temas.folha.uol.com.br/eleicoes-2018-em-graficos/resultado/onda-de-direita-toma-o-pais.shtml>>
- [5] **Cleveland, W. y McGill, R. (1984).** *"Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods"*. Journal of the American Statistical Association. Vol. 79, No. 387 (Sep. 1984), pp. 531-554. [artículo en línea]. [Fecha de consulta: 16 de Noviembre del 2018].
<<https://www.jstor.org/stable/2288400>>

[6] **Sorano, V. (2018).** “*No 2º turno, Bolsonaro tem melhor desempenho no Acre; Haddad, no Piauí*”. G1 - Globo.com. [artículo en línea]. [Fecha de consulta: 17 de Noviembre del 2018].

<<https://g1.globo.com/politica/eleicoes/2018/eleicao-em-numeros/noticia/2018/10/28/no-2o-turno-bolsonaro-tem-melhor-desempenho-n-acre-haddad-no-piaui.html>>