

Can we talk about the gender pay gap?

A Chart Analysis and Critique

Jess Reingold

Bubble Chart: Pay Gap For Every Occupation

There are several charts in *The Washington Post* article, [Can we talk about the gender pay gap?](#) However, I chose to analyze and critique the first bubble chart under the section, “Claim: Women choose lower-paying jobs.” The chart shows the difference in pay between jobs that women are more likely to hold, and jobs that men are more likely to hold. In general, a bubble chart shows “two dimensions of data: one, a numerical value visualized in the scale of its circular bubbles, and the second in what each bubble represents. Simply put, larger bubbles equal larger values” (Starr, 2015). The bubble chart depicting the gender pay gap in various occupations does just that. It shows what the bubbles represent, men and women, in form of two different colored bubbles, and a numerical value, the amount of women or men, in the bubble size. Therefore, it is definitely the right kind of chart for the visualizing the data set.

Originally, I found the bubble chart enticing because of its use of color and dynamic nature. The charts in the article are dynamic and interactive, meaning more data explanations appear when users hover over the bubbles, and when they start to scroll, because there is a parallax design in place. However, upon trying to read the chart, I ran into some flaws. While second bubble chart, the one that appears as users scroll past the first bubble chart, has remedied some of these flaws, I believe it is important to talk about the problematic aspects of the first bubble chart, because that is the one that users initially see.

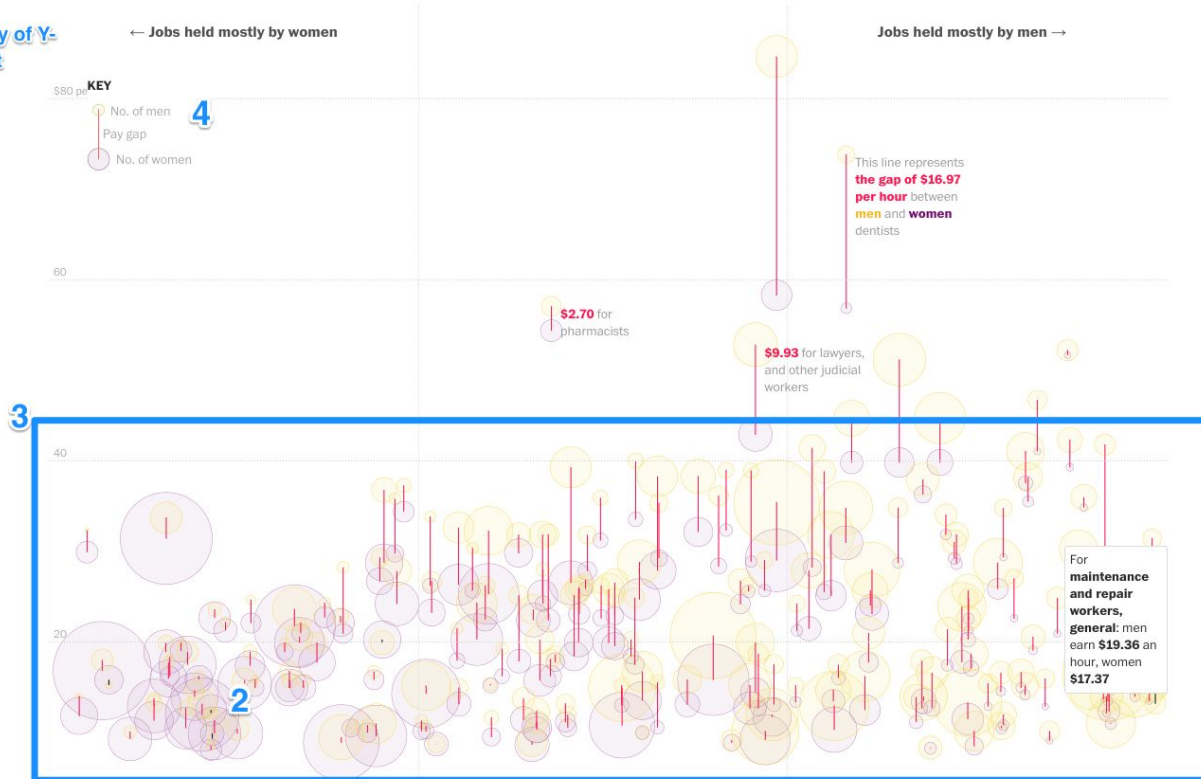
1) Quadrants are not distinct.

2) Data is not distinguishable, legible, or clean

3) Lack of preattentive visual attributes, lack of the use of color intensity and saturation

4) Legend is in the way of Y-axis; blends into chart

Here's the **pay gap for every occupation**, arranged by how many women do each job, with highest share of **women** at left and the highest share of **men** on the right.¹



Flaw 1: Lack of Quadrants

The chart is “arranged by how many women do each job, with highest share of **women** at left and the highest share of **men** on the right.” Left and right are always areas on a chart, and once you have looked at it for a bit, you can see that the majority of the purple circles are on the left and the majority of the yellow circles are on the left. However, having clearer “majority women” and “majority men” areas, or at least a more distinct line going down the middle, or other breaks in the chart, could help the user better visualize where each gender falls in the opposing areas. Having these greater distinctions between the areas on the chart would better show “where items fall within qualitative ranges based on two quantitative values” (Few, 2013, pg. 137). The qualitative ranges are “Jobs held mostly by women” and “Jobs held mostly by men”, and the two quantitative values are the number of women or men in each job, and the amount of money they make per hour. Having more discernable areas would show, for example, how large the gap is on the right side of the chart, in the jobs that men mostly hold.

Flaw 2: Illegible Data Representation

Although the bubble chart makes use of three visual layers including data in the form of bubbles and numbers when you hover, non-data items in the form of explanations of the chart, and the background (Few, 2012, pg. 142), there are areas in the chart that are hard to navigate to. For example, in the lower left corner of the chart, there are a bunch of purple circles overlapping each other. In the context of the chart it makes sense, but it still makes it difficult for the user to find out more information about the bubbles, because they would have to hover over the bubble in just the right place. The chart is cluttered, and could benefit from being more legible, clean, by favoring simplicity over complexity (Nussbaumer, 2016, 21:28).

Showing the data for every occupation from the data set is important, but not to the expense of cluttered visualizations. It may be boring, but a list of the pay gaps in each job could have gone at the end, and the bubble chart could have focused on a select few occupations so readers get an understanding of the data that supports the analysis of the gender pay gap. When the user scrolls past this initial bubble chart, the second one appears, and does in fact pick out certain data points to focus on, but I believe this is the chart the designers should have started with.

Flaw 3: Lack of Preattentive Visual Attributes

Before the second chart highlights data points, the bubbles show no differences in intensity of color. Users looking at bubble charts are used to seeing “varying intensities of the same color to represent values” (Few, 2012, pg. 96). At first glance, it is confusing as to why nothing on the bubble chart really stands out. There is nothing drawing users into the chart. The intensity, or saturation of color “helps our eyes and brains more quickly target the potential points of interest” (Knafllic, 2015, pg. 42). Having a kind of “heat map” helps reduce the mental processing it takes to find interesting trends or areas or differences in the data on the chart (Knafllic, 2015, pg. 42). Heat maps of color intensity or saturation are examples of preattentive visual attributes. Preattentive visual attributes “grab and direct our readers’ attention” (Rinker, 2013) and show readers, or users, what is important, or what stand out in the data. Without preattentive attributes, users are “left to process all of the data” (Knafllic, 2015, pg. 110). Unfortunately, this is what is happening on this bubble chart. There are no preattentive visual attributes at first, and users are left to look at and comb through the entire set of data.

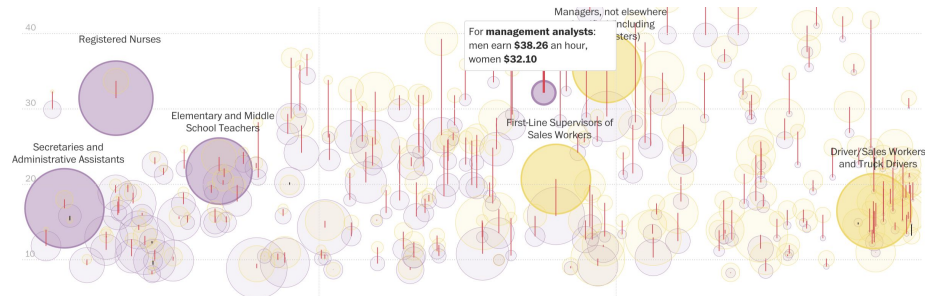
Flaw 4: Legend Placement

Legends are very important when it comes to reading things like charts and maps. The information they hold helps users decipher what they are seeing. Typically, legends are found outside of the chart, but this actually increases the mental processing it takes to read and interpret the chart, because the user's eyes have to travel farther when going back and forth between the legend and the data. Therefore, "the closer the legend is to the data values, the easier it is to read the chart" (Few, 2012, pg. 245). The legend on the bubble chart is not outside the chart boundaries, and is close to the data values, which is good in terms of mental processing, but as the user scrolls, it covers an important piece of information on the Y-axis. As seen in the screenshot, covering up the what the numbers on the left mean (\$ per hour for pay), makes it difficult to understand the quantitative data on the Y-axis, and it modifies the perception of the chart's shape (Few, 2012, pg. 245). Keeping the legend stagnant as the chart transforms, or as the user scrolls even just the tiniest bit would help it stand out amongst the data points so it does not get hidden or looped in with them.

Conclusion

— — —

The story of the gender pay gap is important, and many graphs and charts on the subject have been published. Despite the chart's flaws, a bubble chart is the right kind of chart for the data set containing the gendered pay gap across different occupations. By narrowing the scope of the initial chart to keep it clean, moving the legend, adding visual cues to grab the user's attention, and adding more distinguishable quadrants, the chart could be more user-friendly by making it less overwhelming and easier for users to quickly parse the information displayed. If I were able to give the designers advice, I would suggest they remove this chart, the first bubble chart, entirely, and start the dynamic charts with the second bubble graph that has various data points drawing in users' attentions with its several more saturated bubbles. The bubble chart would still have all of the same information as it does now, but it would be self-remedying some of its flaws. All of the information on the chart is important, but it is not going to create a lasting impact on users if they scroll past it and end up briefly stopping on the second chart.



Sources

Few, S. (2012). *Show Me the Numbers*. 2nd ed. Burlingame, CA: Analytics Press

Few, S. (2013). *Information Dashboard Design*. 2nd ed. Burlingame, CA: Analytics Press

Knaflic, C.N. (2015). *Storytelling with Data*. Hoboken, NJ: John Wiley & Sons Inc.

Nussbaumer, C. (2016). Declutter Your Data Visualizations. YouTube. Retrieved from <https://www.youtube.com/watch?v=X79o46W5pII&feature=youtu.be>

Rinker, T. (2013). Data Visualization: Seeing the Story in the Data and Learning to Effectively Communicate. YouTube. Retrieved from <https://www.youtube.com/watch?v=IHHPO4VhucY>

Starr, B. (2015). How to Design Bubble Charts. Visage. Retrieved from <https://visage.co/data-visualization-101-bubble-charts/>

Xaquín G.V. (2017). Can we talk about the gender pay gap?. The Washington Post. Retrieved from https://www.washingtonpost.com/graphics/2017/business/women-pay-gap/?utm_term=.8360593eb785