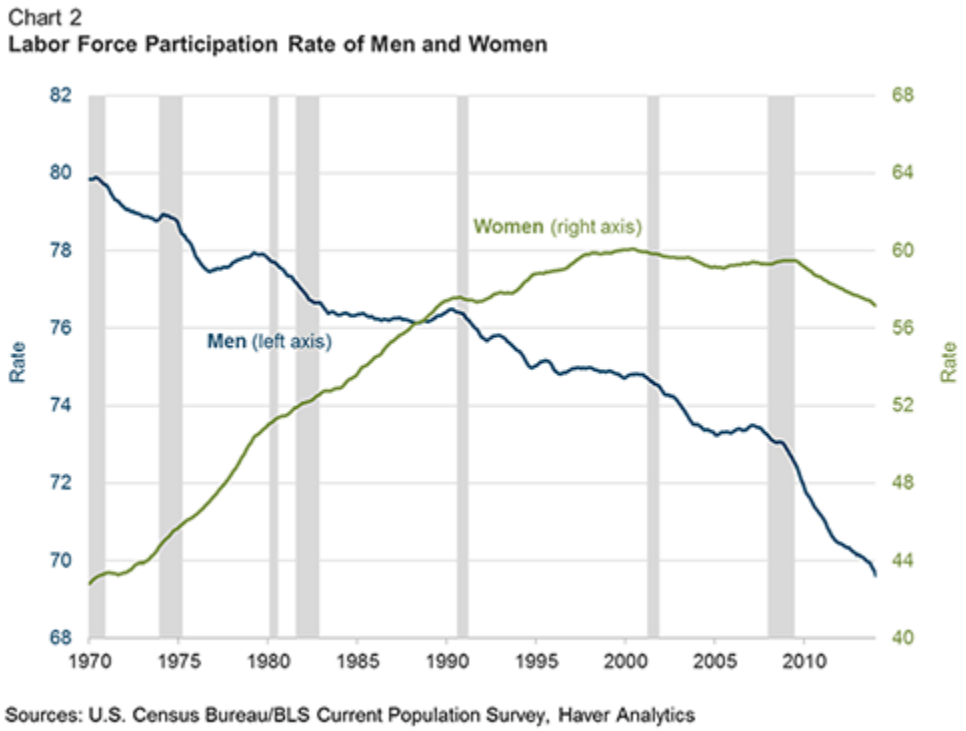
Graphic Redesign Project Report

Introduction

For my bad graph redesign project, I decided to redesign and improve a graph which I have had saved on my phone for years now as a useful example of how not to create a data visualization. An example of what at first looks like a good graph until you take some time to inspect it carefully. The bad graph is the chart below:



This chart is originally from a blog post titled "A Closer Look at Post-2007 Labor Force Participation Trends" published on April 8th, 2014 on the Federal Reserve Bank of Atlanta's *macroblog* [1]. This particular chart I am redesigning is only the 2nd of 11 charts in that blog article. The blog post was authored by Melinda Pitts, John Robertson, and Ellyn Terry with explicit acknowledgements to Tomaz Cajner and Bruce Fallick as well.

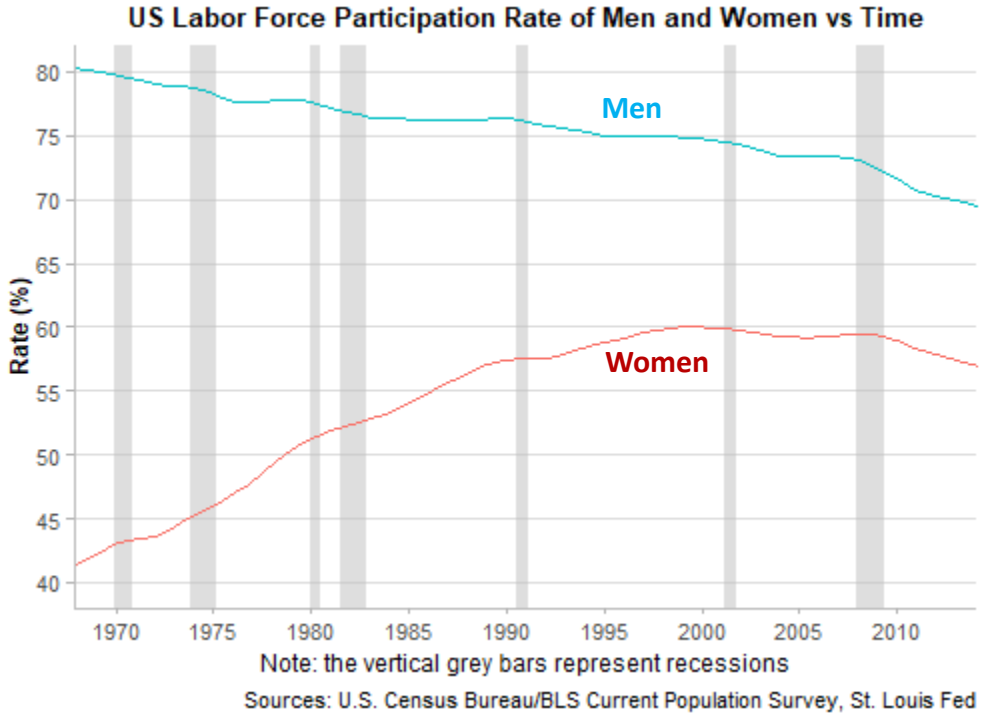
I originally found the bad graph I used around 5 or 6 years ago, maybe from a Google search of labor force participation or maybe in a Facebook post, I cannot remember exactly. The reason I saved it on my phone is because when I first looked at it, it deceived me. On first inspection, I thought the graph was saying that the labor force participation rate of women in the United States had overtaken that of men around 1987, 1988 or 1989 when the green line in the graph intersects the blue line.

It took me a while and multiple times looking at the graph to realize that the line for the labor force participation rate of men has a different vertical axis, the vertical axis on the left side of the graph, than the line for the labor force participation rate of women does which is the vertical axis on the right side of the graph. In fact, there is exactly zero overlap between the two vertical axes because the

vertical axis for men starts at 68 percent while the vertical axis for women ends at 68 percent. I have showed this graph to multiple people over the years and very few of them have ever noticed this important detail immediately. Many of them did not notice the different vertical axes until I pointed it out to them. This might not be a problem if this feature of the chart was pointed out in the blog post which the chart is from, but it was not.

Redesigned and improved version of the graph

The final result of my redesign of the bad graph of male and female labor force participation rates in the United States from 1970 until 2013:



In order to complete this redesign, I used the tidyverse, ggplot2, and lubridate packages in RStudio. It is important to point out that the men and women labels for the labor force participation rate lines were added using text boxes here in Microsoft Word, not in RStudio, for the sake of simplicity and efficiency. Whether or not this Rplot looks simple, that is an eye of the beholder sort of thing, it took me somewhere from 25 to 35 hours of coding to complete.

As you can see, I have fixed the main issue, namely, the issue of using a different vertical axis for the line representing the labor force participation of men than the line representing that of women in my redesign of the original graph. I have done this by just expanding the vertical axis so that both lines fit on the same graph with the same position along a scale measurement. I made sure to include guidelines for the sake of improved clarity as well, but not too many, by default, ggplot2 was putting twice as many gridlines as there are now which was just too many. I also adjusted the horizontal axis so that every 5 years were included as opposed to only every 10 years as it came out by default.

For the majority of the time I spent working on the graph, I had trouble deciding whether I wanted to go with a version of it which used a 12 month moving average of the monthly labor force participation rate data or a smoothed version of the raw monthly LFPR figures using ggplot's geom_smooth function for each LFPR line. The 12-month MA version looked jagged, but the bad graph in the original blog post mentioned at the end of it that all of their charts used 12 month moving averages of the labor force participation rates which I thought was a good idea to replicate in my improved version of their second chart. In the end, I decided to do both, so the graph you see if a smoothed 12 month moving average of the raw LFPR figures.

Special Efforts and Acknowledgements

The first special effort I made was that after completing what I thought was my final version of the redesigned graph using only data from 1970 to 2014, I was not satisfied with the fact that the LFPR lines did not go to the ends of the graph on either end because ggplot2 includes a little extra space to the left and right of where you instruct it to end by default, so the only way to remedy this issue was to go back to the website I got my dataset from originally and redownload the labor force participation rate data csv files from men and women from 1965 to 2020 instead so that the lines on my redesigned graph would not end before the graph ends. I did this on Tuesday after I thought I was finished.

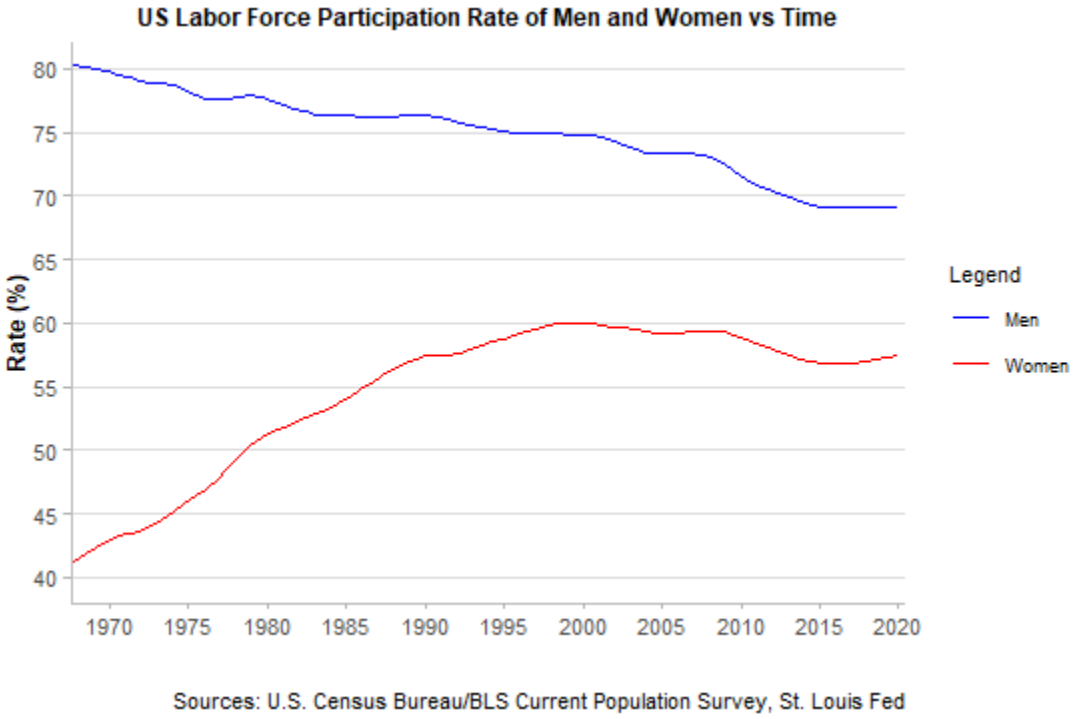
To my great surprise, it was difficult for me to be able to track down a downloadable dataset of the labor force participation rate for the United States that comes separated by men and women or even datasets that have a column for sex so that I could separate them by sex myself. You can see for yourself that this is the case with the BLS Data Finder on the Bureau of Labor Statistic's website.

Furthermore, many easy-to-use labor force participation data tools online only have data going back to the 1990s. For instance, the highly intuitive and interactive chart building tool with the option to download the data on the Labor Force Participation Dynamics webpage created by the authors of the article I got the bad graph from only has data going back to 1998. Another Current Population Survey data tool I found online only has data going back to 1976.

Eventually, last week, I reached out to my longtime friend, the economist and econometrician, Dr. Antony Davies of Duquesne University in Pennsylvania who knew where to find the datasets I needed and sent me the links. I could not have completed this project without his help. The place to find labor force participation rate data that is broken down by sex going back all the way to the 1940s is on The Federal Reserve Bank of St. Louis's FRED economic data website[2][3].

Alternative version of the redesign

I also included one alternative version of my redesign of the bad graph I chose for this project below. This version has a legend and uses more of the available time series by extending the year out to 2020 instead of 2012 because the blog post the original version came from is trying to explore the longer term effects The Great Recession had on labor force participation rates, so including more recent data than was available to the authors back in 2014 should help to explore the longer term effects better. In this version, I also did not include the grey recession bars because they are a little distracting and confusing if you are not used to seeing them.



References

[1] Pitts, Melinda, Robertson, John and Ellyn Terry. "A Closer Look at Post-2007 Labor Force Participation Trends." *Macroblog*, April 8, 2014. https://www.frbatlanta.org/blogs/macroblog/2014/04/08/a-closer-look-at-post-2007-labor-force-participation-trends.aspx

[2] https://fred.stlouisfed.org/series/LNS11300001?fbclid=IwAR20n9K-sZaDYUYz37XPWY76zQu8TC0lK32OcRE6K8xnzfCbTo82zwCOsQ4

[3] https://fred.stlouisfed.org/series/LNS11300002?fbclid=IwAR2V0FKuXGZTFvmmd-o1BwrZCNLHuQp6o_XWAMvhDp0OxC5dPvN3lh_7Kl0

[4] https://ggplot2.tidyverse.org/

[5] https://cran.r-project.org/web/packages/lubridate/index.html