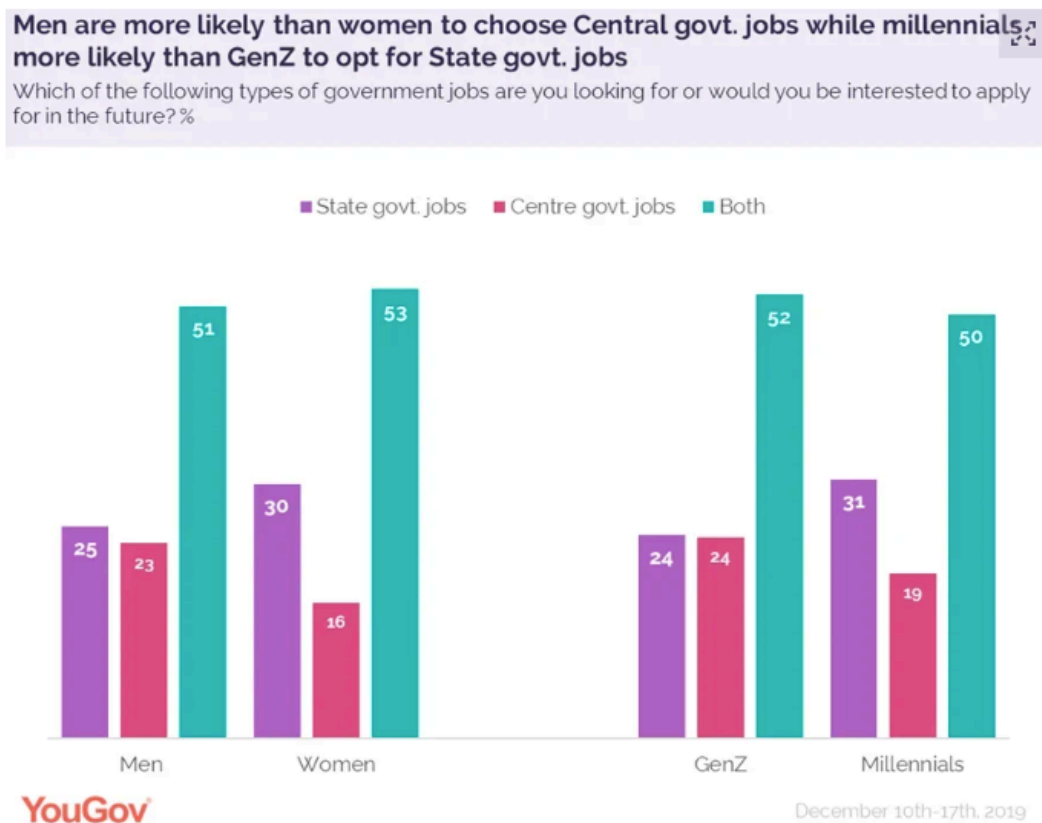


Data Visualization Critique Paper

Amrapali Samanta

Original Data Visualization:



Story from original visualization:

a. What you think the story the visualization is trying to tell is

The visualization report is from an analytics company 'YouGov' on preferred Govt jobs in India by small town residents. I think the story reveals that more than half of the residents prefer both State as well as Central govt. jobs. Within that a higher percentage of men than women (23% vs 16%) are likely to choose Central over State government jobs. Among the generations, a higher number of millennials (31%) than GenZ (24%) showed interest in the State government jobs.

Storytelling:

b. How effectively you think it conveys that story

I think this graph does somewhat convey the story from the title. But there are some thoughts that come up as there is less clarity on what 'both' mean as it is not total of Men having both state and govt job. Also, normally we would represent such kind of data in total % but here as it says more than half prefer any kind of govt jobs, it is not clear that the total value for example '51' in Men group doesn't clarify if there is any other category. As per the title and the story is concerned, I do understand that it shows the total for each preference of job, along with separate values for Men, Women, GenZ and Millennials. The graph shows the values of the type of Govt jobs opted by those categories in percentage (as said before in article).

c. Whether it's being misleading in some way (intentionally or not!)

I do not think the graph is misleading intentionally but in terms of the unit of calculation. Also, the 'Both' values makes no sense and gives no story from the graph as per the title. In the passage it does say that more than half of the residents prefer both the jobs but it is not clear if the author wants to convey anything more from the value of 'Both'. The unit here is percentage which is not mentioned explicitly anywhere inside the graph. This could be misleading for a generic audience with less statistical modelling knowledge. The spelling of legend for Central govt. jobs is wrong. Also I feel the graph has more clutter which takes more time to get the story.

d. How well it holds up the visualization principles we've learned about

The graph does have a context. It does a good job spelling all variables correctly. The values mentioned are also clear to understand. The bars separately represent correct number of people in all the categories. It also shows the legends of the graph with color coding correctly

and the different colors also help us to understand the different categories though I feel there is too many colors.

New improved visualization

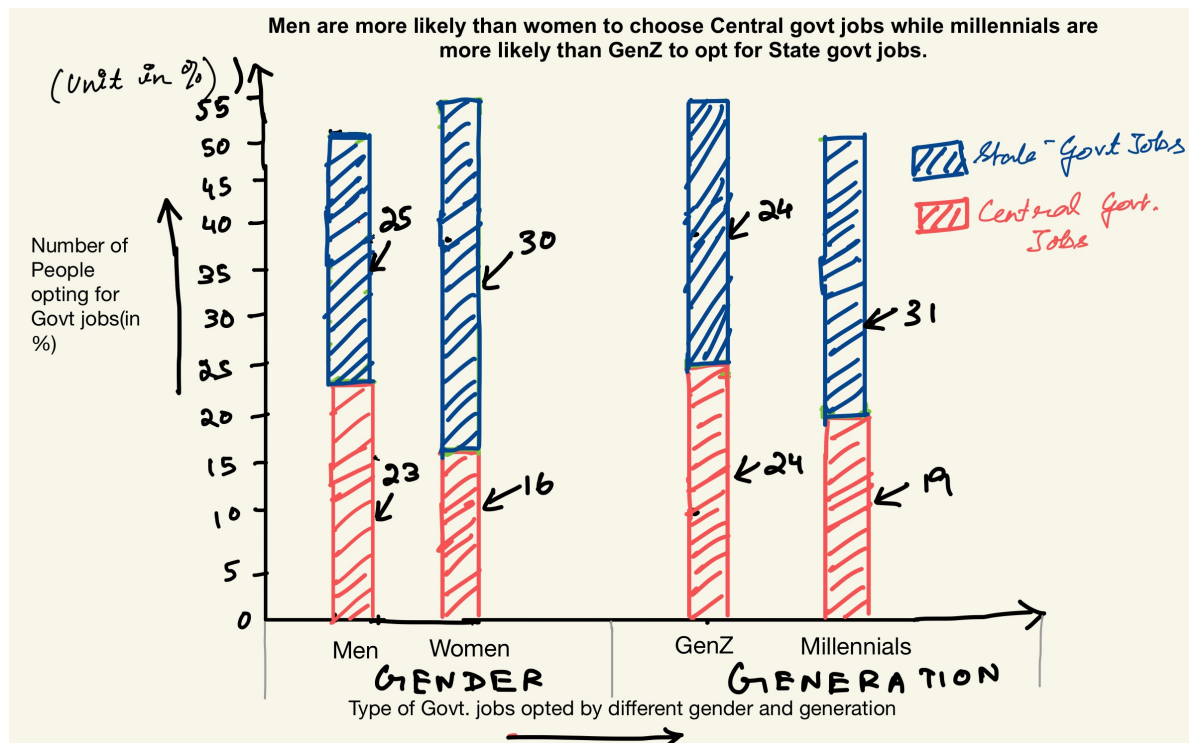
e. How it could be improved

I think this graph can be improved in a few ways as below:

I would add axes titles for easy understanding of what does the graph signify. We can also add the % sign after the values or we can explicitly mention in y-axis that the calculation is in percentage(%). Also here the graph shows too many information as the graph is shown as a single graph with 4 groups 'men, 'women', 'GenZ', 'millennials' but it also shows the totals of each category . This makes a total of 12 bars. I think the number of bars could be made less keeping the story of the visualization intact. When the number of bars are reduced, the color palette could also be chosen accordingly to have less number of colors in the graph. And may be we can take off the 'Both' category as it does not add value as per title.

New Graph:

The new graph is below:



f. Why your improved version (from step 4) improves it

I try to convey in one single graph my approach as it signifies the comparison being told in story.

My graph does obey the principles of visualization like correct axes, axes titles, legends, title of the graph, correct labels, showing the totals of each category, color coding the categories and mentioning in legends, proximity while grouping, correct contrasting color options. I have added the axes titles with proper markings starting from 0 and added the unit of measurement for y-axis. The legends are given as per the type of Govt jobs preferred. In the x-axis I have put a name of main category as 'Gender' and 'Generation' under which the subgroups are placed. There is a light colored demarcation to divide the main categories and keep similar sub-groups closer i.e. men and women belong to Gender.

I have created grouped stacked bar chart. The groups are based on Gender(men, women) and Generation(GenZ, Millennials). This helps to bring into notice the men vs women bars and GenZ vs Millennials bars. Also, the graph looks much simpler and does convey the same moral of the story. The story is mostly focussing on that men prefer Central govt jobs more than women and Millennials prefer State govt. jobs more than GenZ. I have also omitted the 'Both' bars which was showing the totals in original graph. Instead I have stacked all similar categories and marked with different color the two different jobs within each bar. The bars are marked with a preferred contrast color of red/blue. Appropriate sectional data labels based on the number of people taking those jobs are marked as data callout like 23% men under Central govt category.

This new graph conveys same information with more clarity and because of all the above improvements I would say this is a better graph.