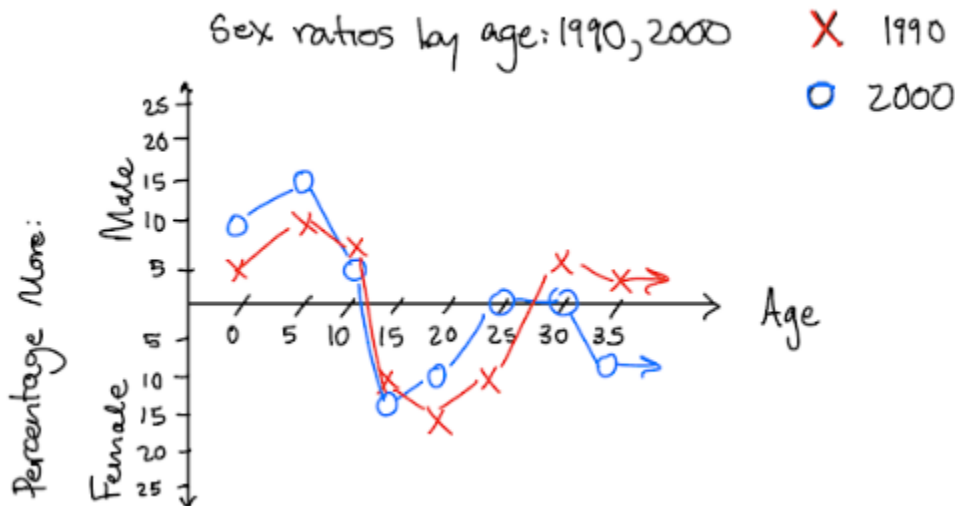
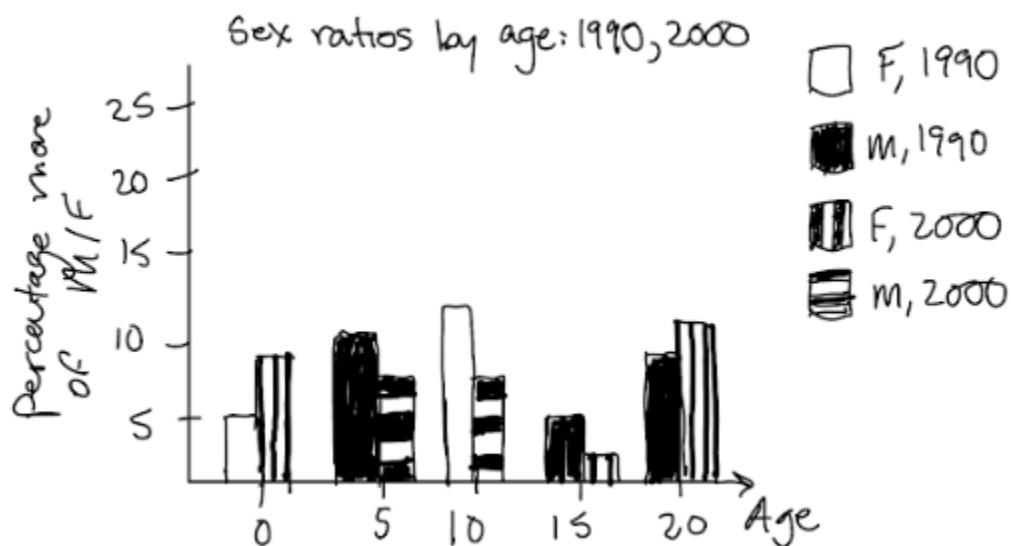


Sketching Visualizations

Guiding Question: What does the ratio of males to females look like with age, and has it changed over the past century?

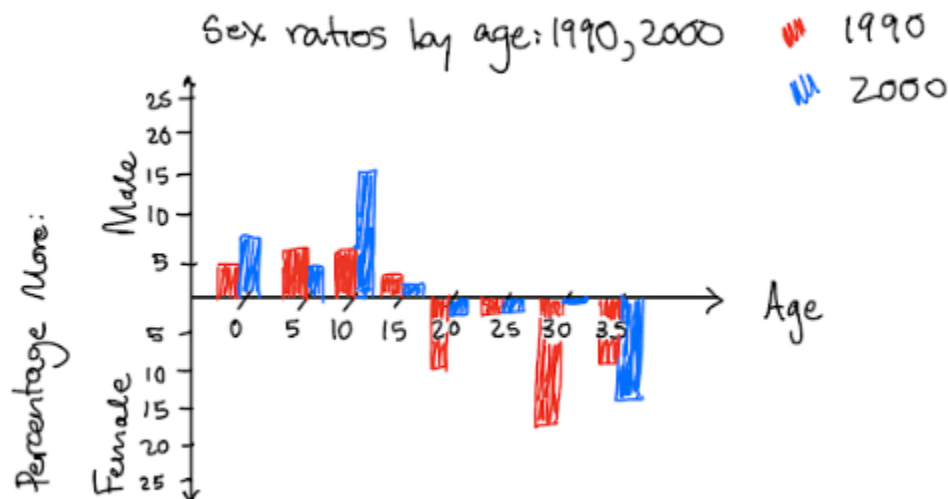


- Ability to track each 1900 and 2000 census data separately
- Able to view each age bucket independently
- See easily if a group is more male or female dominated
- Ability to compare years against each other



- Compare census data for each year side by side

- Percentage is only on the positive side of the axis
- Visually see if a year is more male or female dominated
- Fit all age groups onto same axis
- Same scale between years



- See different census years side by side
- If male dominated, bar and magnitude on top side. If female dominated, bar and magnitude on bottom side
- Same x and y axis scales
- Easy to differentiate years and ratio

Reflection on sketches:

Write one paragraph that reflects on all 3 of your sketches overall. Compare your designs with each other: what are their strengths and weaknesses? What new directions might you explore in the next phase (e.g., synthesizing elements from every sketch, or describing why one of them is significantly better than the others?).

My initial desire in designing these visualizations was to be able to readily distinguish if an age group is dominated by one or the other sex. In charts 1 and 3, I did this by placing the bars on either side of the X-axis depending on the ratio of males to females for a given age group. Chart 2 does this with different patterns in each bar depending on both the year displayed and which sex is the majority. I found this aspect of the data to be easier to distinguish in charts 1 and 3.

Another thing that I wanted to do was to be able to compare the census data from the 1900s and 2000s without too much effort. In charts 2 and 3, I did this by placing the bar data for each year side by side, while in chart 1 I plotted lines instead, and so I was able to stack the data within the same vertical plane. I don't necessarily think the data should be plotted as a line as in chart 1, because that implies some linearity, while the data is bucketed into discrete age groups. I think the four different bar encodings on chart 2 might be too much — at least it is in black and white. Some color to distinguish the categories might help, but the color choices would have to be deliberate to separate Male from Female and 1900 from 2000 while maintaining some continuity. I think chart three is probably the strongest, but I don't really like it plotted as percentages, and I don't appreciate the colors that I chose to separate 1900 and 2000. In the next phase, I'll explore how to communicate the total population of each sex per age group, how to better communicate that for the 1900s and 2000s, and how I might change my Y-axis.

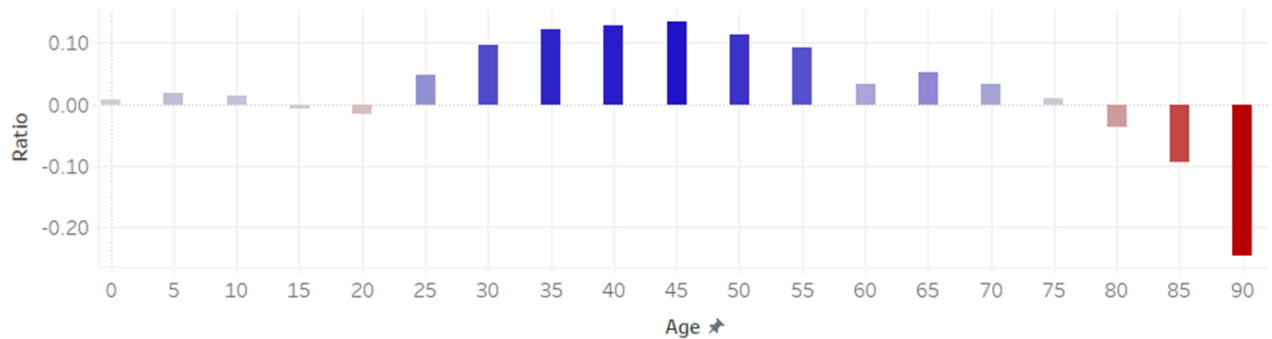
Final Design

Guiding Question: I'd like to continue analyzing my question in part one, and see if I can come up with a better visualization.

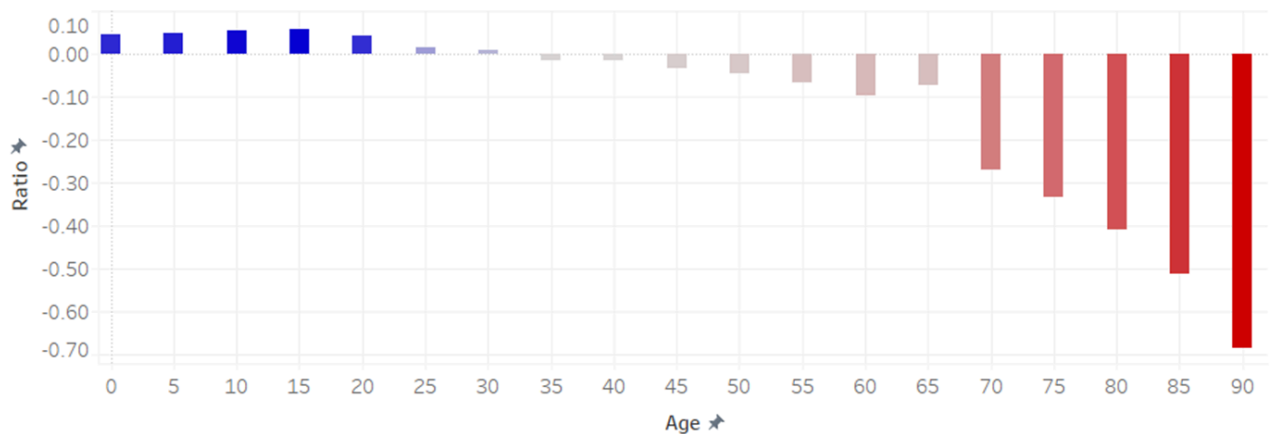
Static Visualization:

Male to Female Sex Ratio by age group; 1900s and 2000s census data

1900s census data



2000s census data



Writeup:

This visualization represents the ratio of males to females in a given age group in the 1900 and 2000 census year categories. I decided to use a bar chart, as in sketches 2 and 3, because I felt it was easier to visually analyze than the line chart in sketch 1. The bar chart also accounts for the fact that the age groups are discrete and that there should be no continuity between them (the line chart failed in this). In sketch 2 I used markings to highlight if a group was dominated by more males or more females. Instead of doing that here, I decided to use blues and reds on a gradient. More blue means more male; more red means more female. The

gradient is nice because it tends toward a whitish-purple in the middle, which I feel is a valuable addition that helps convey that for a given age group the ratio between sexes is not too extreme. As in each of the sketches, I decided to continue plotting my age groups on the X-axis. However, in my static visualization I changed my Y-axis. In the sketches, I was plotting the percentage of males for a group — $\# \text{males} / \text{total population}$. In my final visualization, I plotted the ratio of the genders — $\# \text{males} / \# \text{females}$, adjusted to fall around 0. I felt that this would give better insight into the trends I wanted to analyze. As in sketches 1 and 3, the values are plotted on either side of 0: positive values indicate a group that has more males while negative values indicate a group with more females.

My first goal in making this visualization was to easily communicate whether an age group had more males or females, and my second goal was to make it easy to compare the data over the two census years. I think that having the data span both positive and negative values, combined with the very clear color gradient, makes figuring out if an age group had more males or females very easy to do. I had some struggles in figuring out how I should go about comparing years. In my opinion, sketch 2 was too cluttered, and I wasn't confident that I'd be able to come up with a color set that makes my messaging clear and easy to distinguish. In sketch 3, I used two different colors and placed the date by year side-by-side within each age group. This was nice because it made comparing the values very easy. However, when I went to implement this in Tableau, I realized that I'd have to pick two distinct color gradients. That would have been ok, but not optimal. I feel that the blue/red gradient is a very standard choice of colors to differentiate sexes, and I figured that my plot could get confusing if one gradient was blue/red (very clear), while the other was purple/orange or green/yellow, which wouldn't be immediately recognizable. In addition to thinking about the colors on my plots, I also was limited by my Tableau ability. I was able to get my charts aligned onto a dual, synchronized axis, but I couldn't figure out how to offset the bars from each other. After a while, I moved on, and decided that having two charts would be better anyway. Mostly, this is because having one plot per chart is a lot easier to read, and having only the blue/red gradient rather than the blue/red next to a green/orange or some other color combination was a lot more easy to digest and made accomplishing my first goal easier. To achieve my second goal, I had to make sure to align my tick marks and axis scaling, and to ensure that my plots were vertically aligned. Initially I was worried that it would take too much effort to compare the two plots, but I no longer feel that way and am confident that it was the right choice. I also elected to keep the negative values on the Y-axis. I am not sure that this was the right decision, but again I was limited by my Tableau ability and couldn't figure out how to put the absolute value on both sides of my X-axis while retaining my positive and negative data points. The final thing that I think may need improvement is the "key" that I'm using to explain colors. The default legend that Tableau provided had the blue/red gradient, but it assigned each end the highest and lowest data values, respectively. I couldn't get tableau to remove those, and I felt that they confused the key, so I decided to just include the Male (blue) and Female (red) encoding into my title.

Overall, I'm fairly happy with how this turned out. I think that it represents a significant improvement in my Tableau skills from Assignment 2 (which was very rough for me), while still showing that there are a lot of things to improve. As with Assignment 2, I felt that most of my

graph-making was limited by my Tableau skills... There are certainly many things on this plot that I'd change if I knew how to, but overall I find my progress encouraging. I feel that I accomplished the goals that I set out to in designing my visualization, and that I was able to effectively answer my guiding question.