Design Studio 2

Part 1 - Analysis

What trends do you see in the data?

There are more of the sources with estimates towards the end of the time scale. Also the differences between the estimates become closer over time. This seems perfectly logical as in relatively recent(historically anyway) years we have become more accurate in estimate population, as well as more interested in doing so. For much of human history we do not have much information or data to base the population estimates on, so it is understandable that there is so much variance between these estimates for the earlier years.

The populations start climbing steadily around the 1700, and where as previously they had been gradually increasing, in this later part of the graph, from the 1700's onward, the population climbs rapidly. This gives us, at the end of the graph, a rather abrupt curve and a steep slope to the line.

Analyze how big the differences between various estimates are. Do you see a trend, i.e., do the differences become smaller or larger over time?

Differences become smaller over time. As explained above.

Think about these differences relative to the estimates at the respective time points and in absolute terms. When are the uncertainties the largest in absolute, when in relative terms?

The absolute differences between the estimates is greater towards the end of the graph, while the relative differences are less towards the end. This makes sense because the estimates at the beginning of the chart are not as accurate, and organizations probably have varying ideas on how to achieve these estimates. This makes the relative differences more at the start. Towards the end of our time scale, the population estimates become more accurate, and probably also more standardized in how they are achieved. This makes the relative differences much lower. However, the population itself is drastically higher towards the end of our time scale, thus making the absolute difference much higher at the end of the chart, even while the relative difference remains low.

Do you think you can faithfully represent the uncertainty and the data in the same plot? Why, or why not?

I think this would be a challenge because of how steeply the graph climbs towards the end of the scale. It is hard to show were are the estimates and thus uncertainty lies since in terms of absolute differences, they are very great at the end of the scale but the relative differences are so close, it can be hard to see. However, if you adjust the scale to accommodate the end of the graph, then the start of the graph becomes oddly represented. Or, if you change the scale completely to show the uncertainty and not the absolute differences, then you loose the ability to see the overview of the data which has an interest of its own. Better, I think to leave the overview plot of an average of the estimates, and supplement this with additional charts that can better show the variations between these estimates.

What effect do you think will the linear interpolation have on the uncertainty?

The uncertainty for the interpolation will depend upon the source, and will change relative to

that. It will depend on the uncertainty of the data source being considered, since the interpolation is being calculated based on the

reported numbers before and after that date. The sources that do have reporting

data, however, will probably have more accurate estimates, based on how they calculate the data and what they all take into consideration for it.

If a population estimate has a large uncertainty for a date they did report, and the next date is an interpolation, then if the data following the interpolation has a high uncertainty, the interpolation will as well. Even if the date following has a lower uncertainty than the proceeding, since the estimate will fall in-between the two, it will still have a high uncertainty. However, if both the data points before and after the interpolation have low uncertainties, then that interpolation will as well.

Is linear interpolation a suitable method for this data?

Yes, I think so. since the data we are looking to preform the interpolation on is using 2 data points, this is the simplest way to obtain this value. I suppose another interpolation scale would be more accurate, but also more complex.

Part 2-Sketching

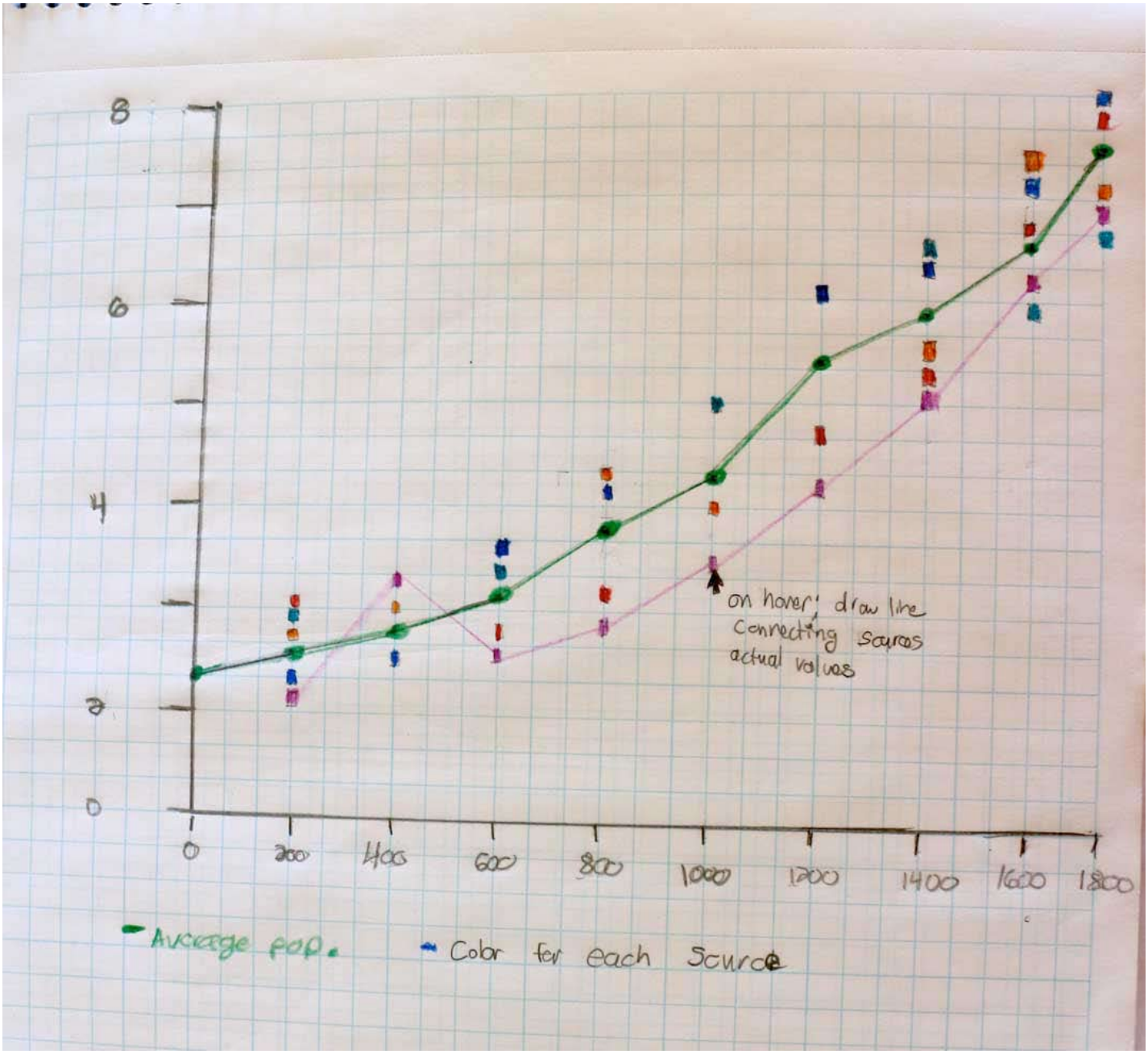
Part 3 - Group Reflection

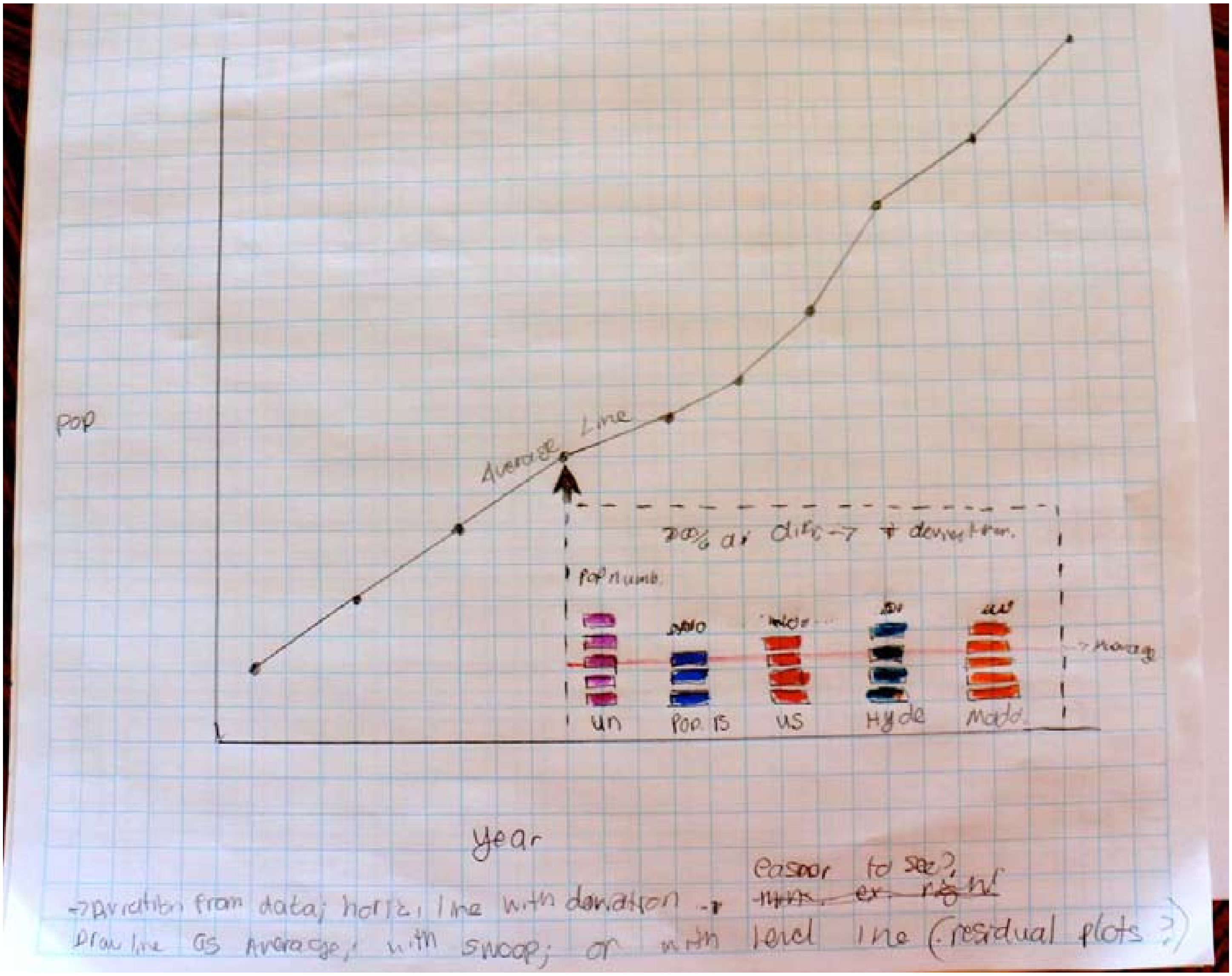
A lot of the Design Studio discussion dealt with how to deal with the data sets for such a visualization, as well as which scales to use. There were varying techniques for working with the data, with some people have 10 different data sets, one for each source with regular values, and one for each source with the interpolated values. It seems one of challenges is how to show the general overview of the data for all of the years, and still show the divergences from the average for each population estimate. Many ideas were submitted about creating visualizations that allowed for panning and brushing thus being able to effectively show both and overview and the details. Concerns about this were raised due to the complexity and the time required to finish the other parts of the homework as well. Other ideas included having tooltips on the overview, to allow the viewer to see the details this way. This seemed more workable with the time we have and also should solve some of the visual problems. Discussion also happened about weither the interpolated data should be included in this, and if it should be represented. It seems the consensus at the end was yes to both. Some suggested the idea of displaying only the values on the later section of the chart, since there is such a difference in population between the start and the finish of our time scale and it gets very challenging to effectively show such a spread. People seemed very uncertain of this, and I personally felt completely leaving off the start of the chart changes what we are looking at, and you loose the perspective of how steeply the population climbed in the later part of the chart compared to the slower and steadier increases earlier on. From a source perspective, you also loose a comparison of how the differences (absolute and relative),change and vary across time. Ultimately, it comes down to what questions your graph is asking and what story you are trying to tell. Data isn't always clean and pretty and you shouldn't simply change what data you are using to try and make it so. Differing opinions were also raised as to which scales to use to display the data, should it be keep in a scale of population/time, or should it represent time and a percent? I'm not sure we reached a consensus on which visualization was ideal, it was challenging for everyone to share/see the visualizations

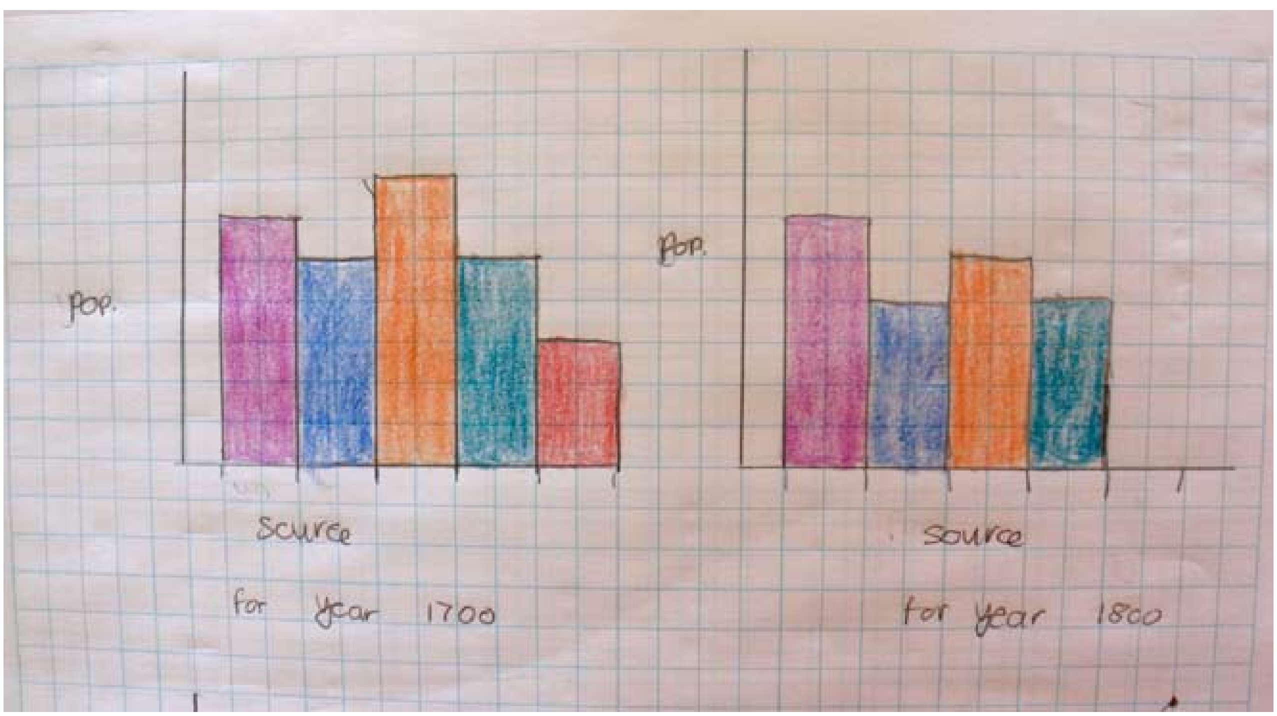
in the online format, so a lot of the discussion was more about what a good visualization of this should include. We agreed on important point to consider.

1. How to effectively display the overview and the detail/difference comparisons.
It seemed most people were in favor of using tooltips or drawing additional graphs to show this.
2.Interpolated data should be used, and marked as such.
3. Which scales to use should be considered.

Sketches:




pop
Average Line
un
Pop. IS
US
Hyde
Madd.
year
easier to see?
→ Deviation from data; horiz. line with deviation →
Draw line as Average; with swoop; or with level line (residual plots?)


pop.
source
for year 1700
pop.
source
for year 1800