Record your design decisions

Adding a provinces table with the full name of province as well as their corresponding code:

- This schema was added into this table so that when data is inserted into other tables, the rows that require provinces related information can simply use the codes, reducing the amount of writing (good for efficiency).

- The provinces table also allows us to avoid using nulls in the other tables. We list all the provinces that we are observing in this table; so even if a certain province does not have an associated value in another table; we simply don't include it in that table; the user will however still know that we are observing all the provinces.

- For that reason, we also added the following foreign key constraint: FOREIGN KEY (location) REFERENCES provinces(code). This would ensure that only data for province codes was being added into other tables; rather than nonsensical codes.

Adding the check constraint: CHECK (ratioMale + ratioFemale >=98 and ratioMale + ratioFemale <= 102)

- For tables that had ratioMale and ratioFemale attributes, we realized that the sum of these 2 attributes is not equivalent to 100%. They were slightly above or slightly below that point, which can be caused by rounding values up or down. So to ensure our data could go into the table without allowing nonsensical values to be inserted, we ensured that the addition of the 2 ratio would add to about a 100 (>= 98 and <= 100).

Decided not to follow the comment for only including 1 ratio:

- We were advised to only include ratioMale or ratioFemale within our tables. We chose not to do that as our data, in some rows, for many of the tables did not add up to a 100. Also, we wanted the ratios to be explicitly present for purposes of comparison between male and female occupations.

Removing ratioMarried/ratioCommonlaw to ratioCouple ratioNotInCouple:

- We were interested in seeing if partnership has an association with occupation. So we removed the differentiation between ratioCoupleCommonLaw vs ratioCoupleMarried to be represented simply by ratioCouple.

GDP >= 0 constraint:

- This constraint allows us to ensure that we do not get negative values for GDP; however we did not put in an upper limit as that is hard to predict and may cause problem for future insertions.

Cleaning Up Data:

Removing all unnecessary columns:

- We ended up removing unnecessary columns that did not have to do with our schema in all 4 of the tables that we used. For example, relationship status had ratios for couple married, and those that were in common law relationships, but we only used data from the total number of couples columns, as that was the only piece of data we needed.

Ratios:

- We wanted to work in ratios, as seeing percentages is much easier then to have to compare large numbers of some whole value, Therefore, we ended up acquiring the ratios by performing calculations; for example, for the relationshipStatus table, we ended up dividing the total number of people in a relationship in different provinces respectively, by the number of people eligible to be in a relationship to acquire the ratio.

Null Value

- For the seniors table, there were values that were not present, so we removed those null values and did not include them in the table.

Territories

- The territories had very inconsistent data. For some data sets, we had data the territories being represented as 1 row, and for others we had a row for each territory, and for other data sets we had no territories at all. So due to this inconsistency, we made sure to only focus on the provinces; for which data was present in each table.

Unreasonable Ratios

- We also had some data that had unreasonable ratios that added up to 107 for Saskatchewan in the seniors table. We did not include that province in the table.