Teen pregnancy and childbearing bring substantial social and economic costs through immediate and long-term impacts on teen parents and their children. According to the [CDC’s website](http://www.cdc.gov/teenpregnancy/about/):

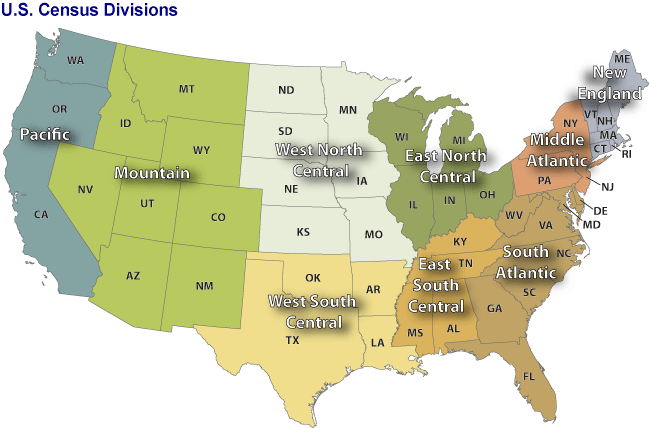
* In 2010, teen pregnancy and childbirth accounted for at least $9.4 billion in costs to U.S. taxpayers for increased health care and foster care, increased incarceration rates among children of teen parents, and lost tax revenue because of lower educational attainment and income among teen mothers.
* Pregnancy and birth are significant contributors to high school dropout rates among girls. Only about 50% of teen mothers receive a high school diploma by 22 years of age, whereas approximately 90% of women who do not give birth during adolescence graduate from high school.
* The children of teenage mothers are more likely to have lower school achievement and to drop out of high school, have more health problems, be incarcerated at some time during adolescence, give birth as a teenager, and face unemployment as a young adult.

These effects continue for the teen mother and her child even after adjusting for those factors that increased the teenager’s risk for pregnancy, such as growing up in poverty, having parents with low levels of education, growing up in a single-parent family, and having poor performance in school.

Understandably then, lowering teenage fertility rates (TFRs) is of great public health importance. It might first be important to understand where TFRs are the highest, and among which groups of people.

**Data**

The data for this problem were obtained from the CDC’s data base query system “[WONDER](http://wonder.cdc.gov/natality.html).” There are three total data files for this problem. The first two contain information on teen fertility rates; one on the state level (CDC-Fertility-Data-State.csv) and one on the Census Division level (CDC-Fertility-Data-Division.csv; for Census Divisions, see figure below). Both of these file contains columns for the region (State or Division, depending on the file;), year (2003-2015), race, number of births to female teens and number of female teens in each region. **Problem:** Some counts in the State data file are “suppressed” because they were too small, in order to protect patient confidentiality. For these cells, the number of teen pregnancies must be *estimated* somehow. One possibility is to use the information in the Census Division to which the state belongs. This requires a data merge, which can be done by way of the **third** data file: Division-State-link.csv.



Your tasks are as follows:

* Find a way to estimate the number of births to female teens for the year/race/state combinations that are “Suppressed.” This process of filling in missing values in an intelligent way is called *imputing.*
* Create a visualization that somehow assesses the quality of your imputation.
* Given the state-level data with the imputed values, create a single dashboard that answers the following:

1. How have TFRs changed over time, both on the state-by-state level and overall?
2. Are there racial gaps in the TFRs? How have any gaps changed over time?
3. Which states have the best TFRs, and which states have the worst?
4. Which states have improved the most from 2003 to 2015?
5. Which states have the most racial disparity in TFRs?

Include at least one filter and one highlight action in your final dashboard.