1.We need variables to represent the volume of opioid prescribed and drug overdose deaths.

To show the volume of opioid prescribed, we need DOSAGE\_UNIT\*QUANTITY to calculate the volume of opioid

To show the drug overdose deaths, we need Drug\_Induced\_Death per county per year.

2.For samples, we should have date for each county per year(at least from 2009-2012 because). For Florida part, the year data should at least include 2009-2012. For Texas, the year data should at least include 2005-2008. For Washington, should at least include 2010-2013. Since some counties may have the same name, it is better to obtain FIPS Code for each county from other data sources.

3.A single row should be like

County | Year |Effective\_Volume(Opioid) | Overdoes deaths

4.We need to clean and process the original data:

a. make sure each column data have the same unit

b. Calculate the variable we want (like how to obtain effective\_opioid\_volume prescribed in each county per year?)

c. separate drug overdose induced deaths from other deaths

d. Obtain year/month from the transaction\_date

5.Where can we find FIPS code?