

Midterm Project Presentation

Tianchi(Ava) Zhang

October 26, 2016

Part 1. Dataset Description

County-level Oil and Gas Production in the U.S. dataset is a national county-level database by Economic Research Service researchers. It includes disaggregated data from State agencies EIA (Energy Information Administration), county-level production statistics by States, and well-level databases. This dataset provides gas and oil production for counties in the lower 48 States, from 2000 to 2011. It also includes variables related to rural-urban, metro-nonmetro factors. More information is available on website <http://www.ers.usda.gov/data-products/county-level-oil-and-gas-production-in-the-us/documentation-and-maps/>.

Part 2. Tidy Data

Step I. Revise Column Headers

After reading the raw dataset, the first thing we did is to change a couple of column headers to simpler and more readable ones. For example, we revised the column header “Rural_Urban_Continuum_Code_2013” to “Population Level”.

Step II. Separate Multiple Types in One Table

Since there are two types of data in the raw table, we decided to split the table into two. One contains information about County, including FIPS, State, Population Level and other metro/non-metro variables. Since the fourth column already showed “County” in its header, we eliminated the word “County” from the column content. The other table includes information about county-level oil and gas production each year from 2000 to 2011.

Step III. Separate Multiple Variables in One Column

The second table has multiple variables stored in one column, like “oil2000” for example, so we melted the column headers into rows: “production” and “year”. So now the “production” column includes production types which are either oil or gas, and the “year” column includes data year ranging from 2000 to 2011.

Part 3. Data Summary

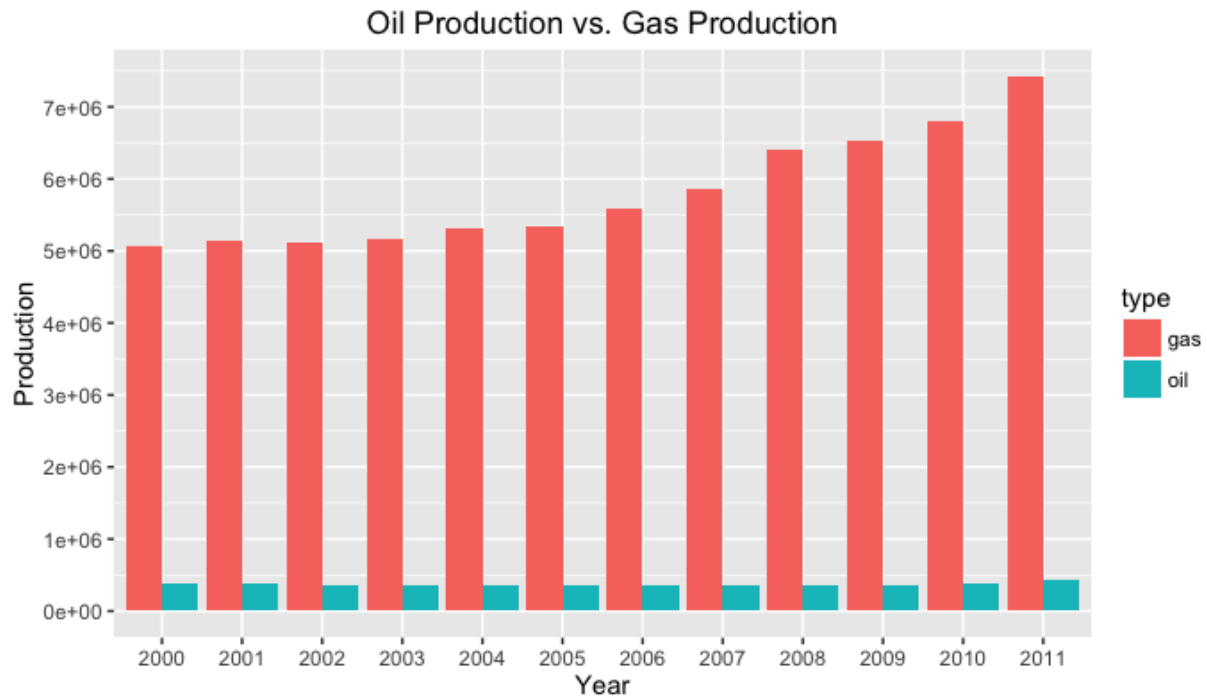
For the preparation of future data visualization, we grouped Counties by its population level, introduced a new variable called “County Level” (ranges from level 1 to 9), and created a couple of new tables. The first three tables are average annual oil and gas production for each county, nationwide mean production in each year, and average oil and gas production for counties grouped by each county level (population level). Then we tidied the third table, since it has multiple variables in one column. After that, we created three more tables, which are mean oil and gas production for each county in terms of population level, mean oil production for each county in each year, and mean gas production for each county in each year.

Part 4. Data Visualization

Based on these tidy tables we created in Part 3, we generated six different graphs in order to prepare for the future analysis.

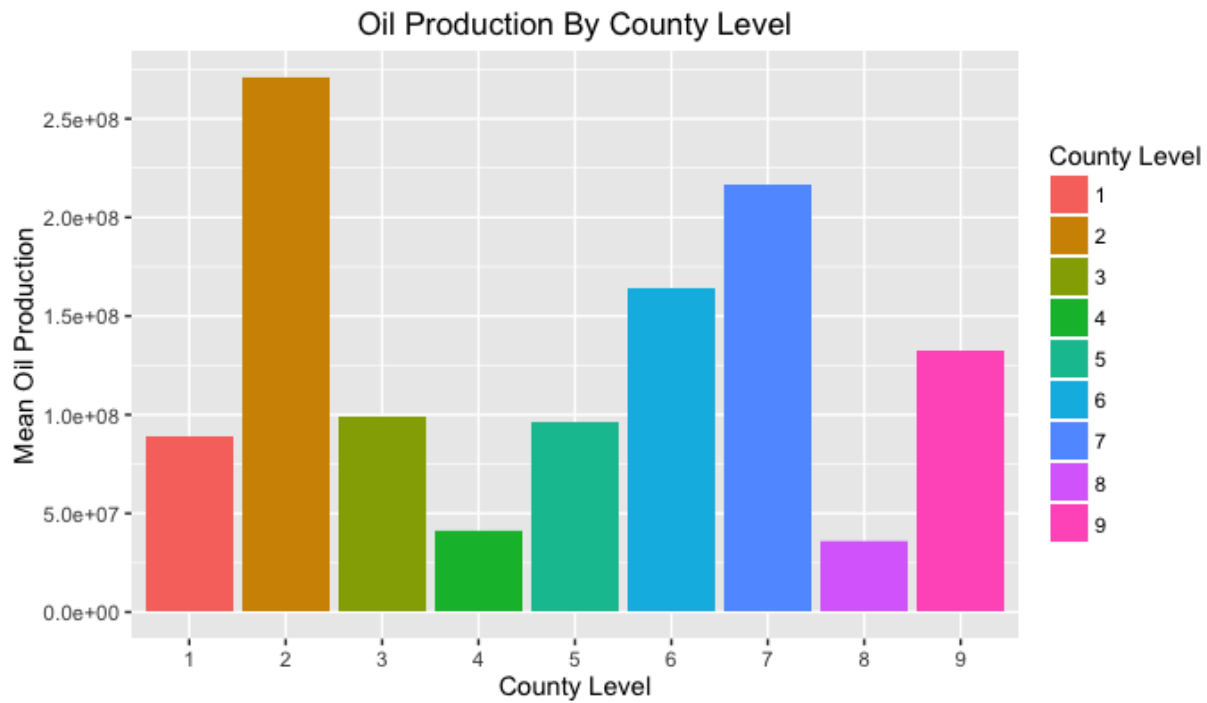
Graph 1. National Mean Oil & Gas Production In Each Year

The following is a double bar chart aiming to compare the national mean oil and gas production in each year.

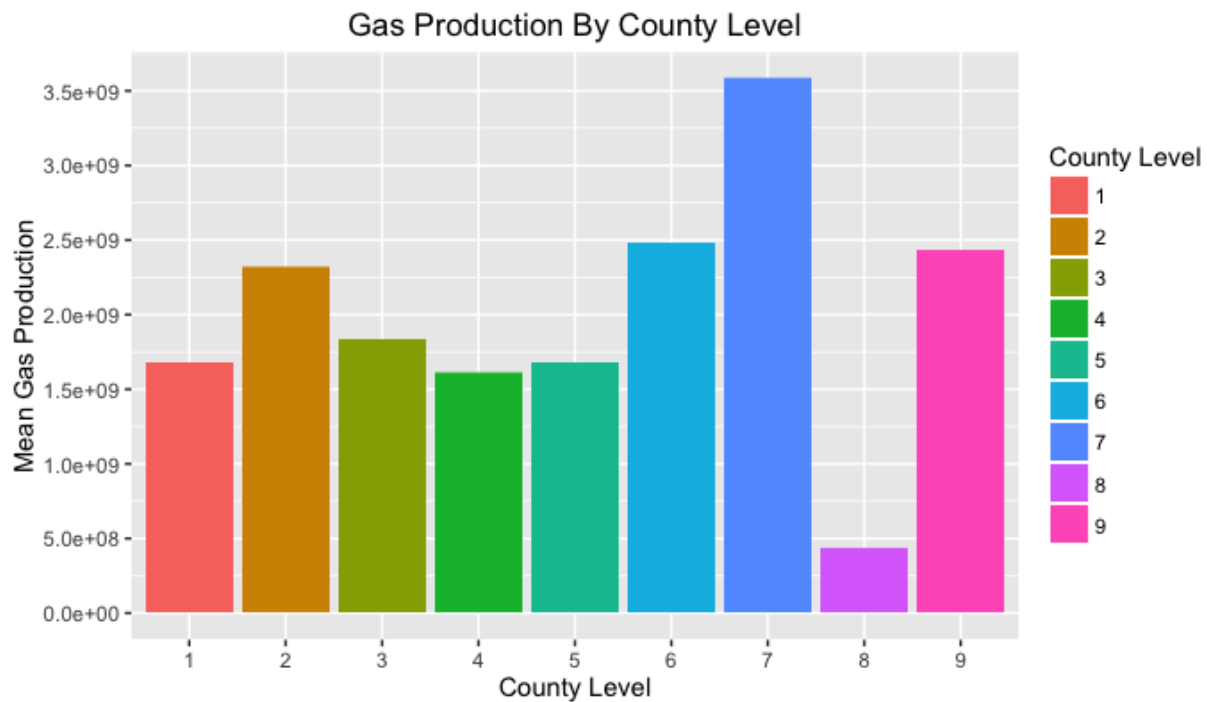


Graph 2. Oil(Gas) Production Based On Each County Level

The following is a bar graph that designed to show the mean oil production based on each county level.

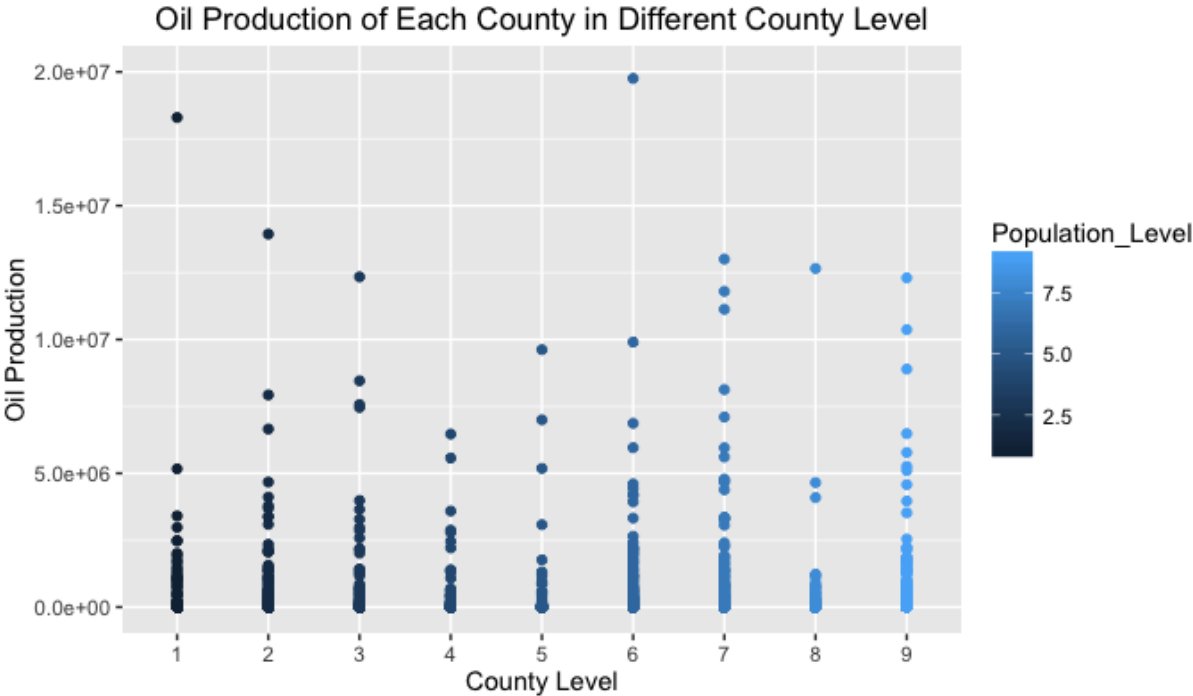


The following is a bar graph that designed to show the mean gas production based on each county level.



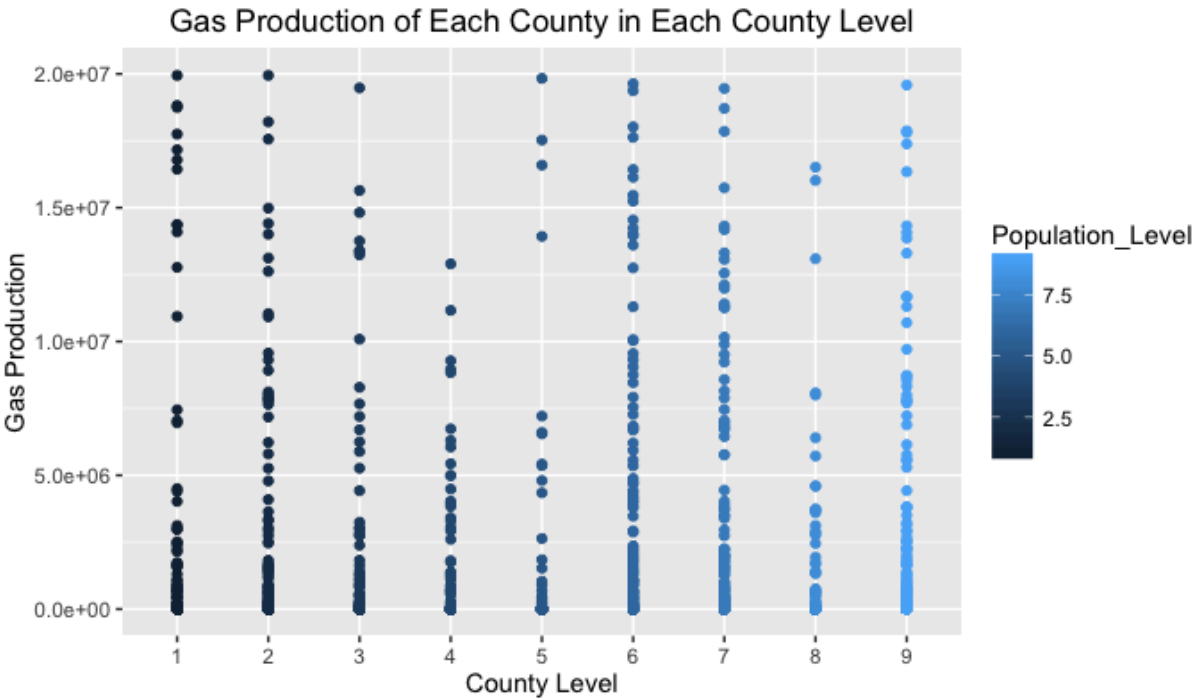
Graph 3. Oil(Gas) Production Of Each County In Each County Level

The following is a scatter plot that presents the distribution of the oil production of each County in each County



level.

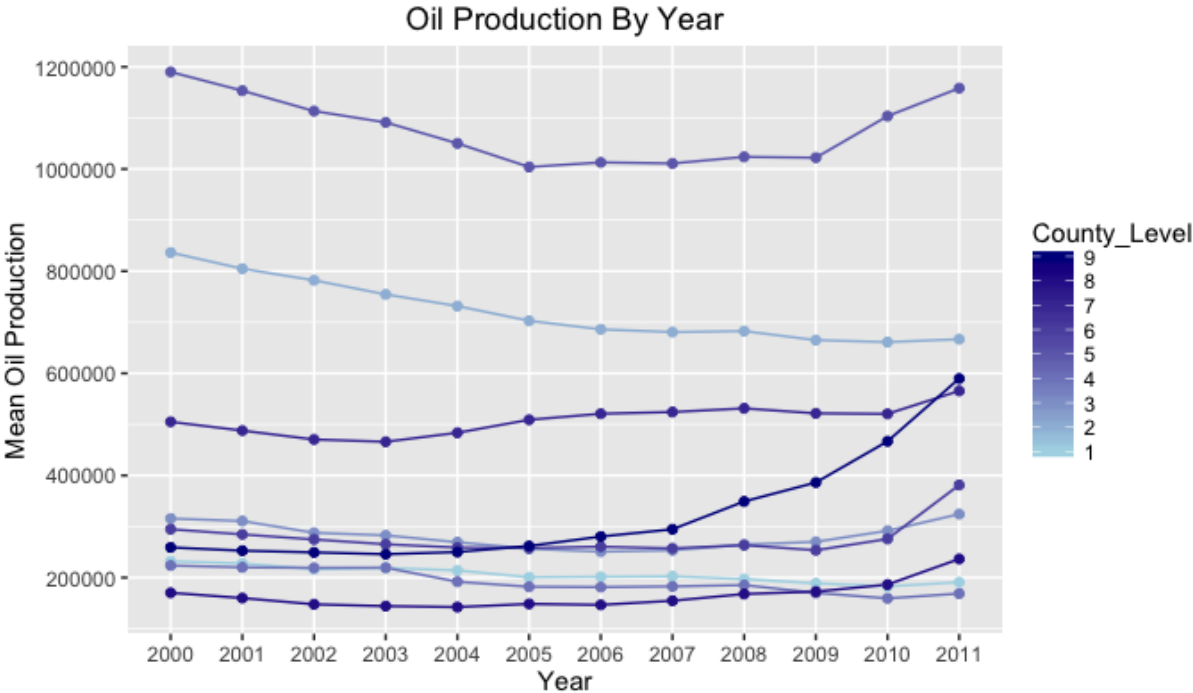
The following is a scatter plot that presents the distribution of the gas production of each County in each County



level.

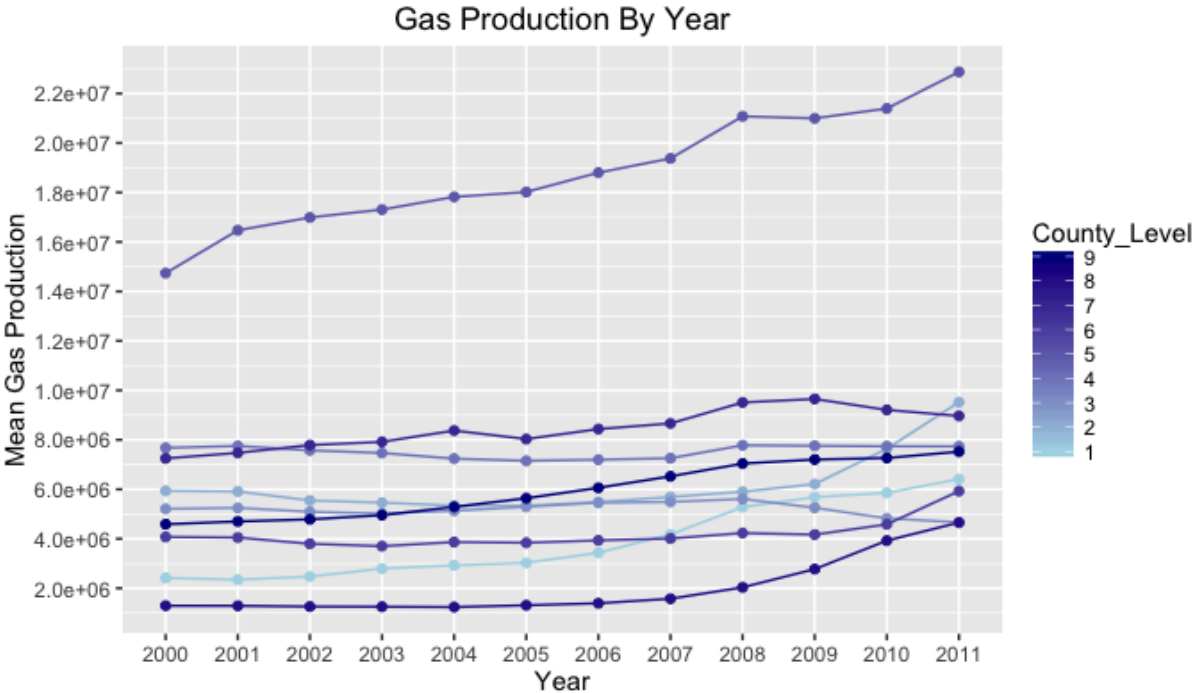
Graph 4. Annual Oil(Gas) Production For Each County Level

The following is a line graph that shows the fluctuation of annual oil production for each county level in each



year.

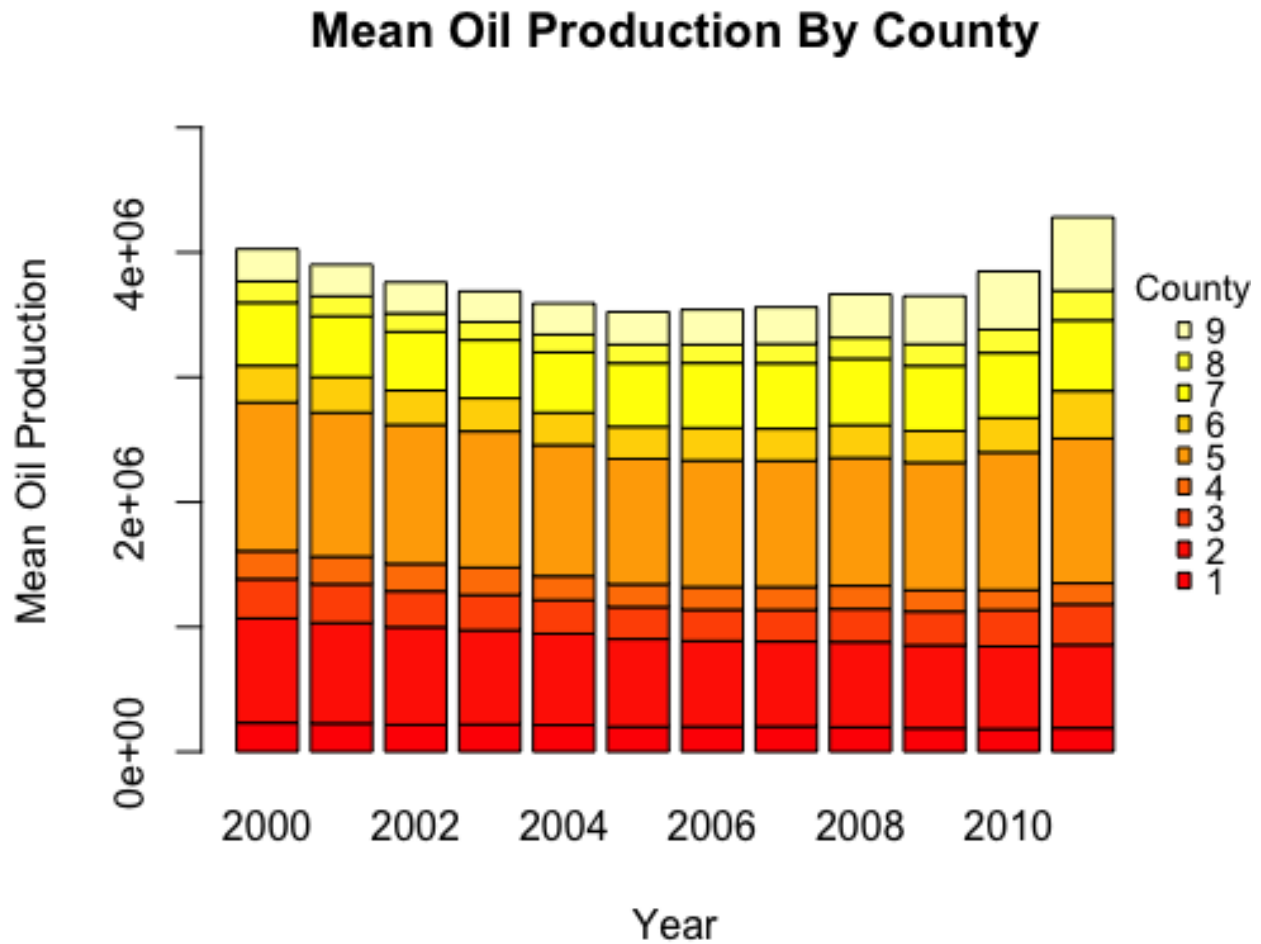
The following is a line graph that shows the fluctuation of annual gas production for each county level in each



year.

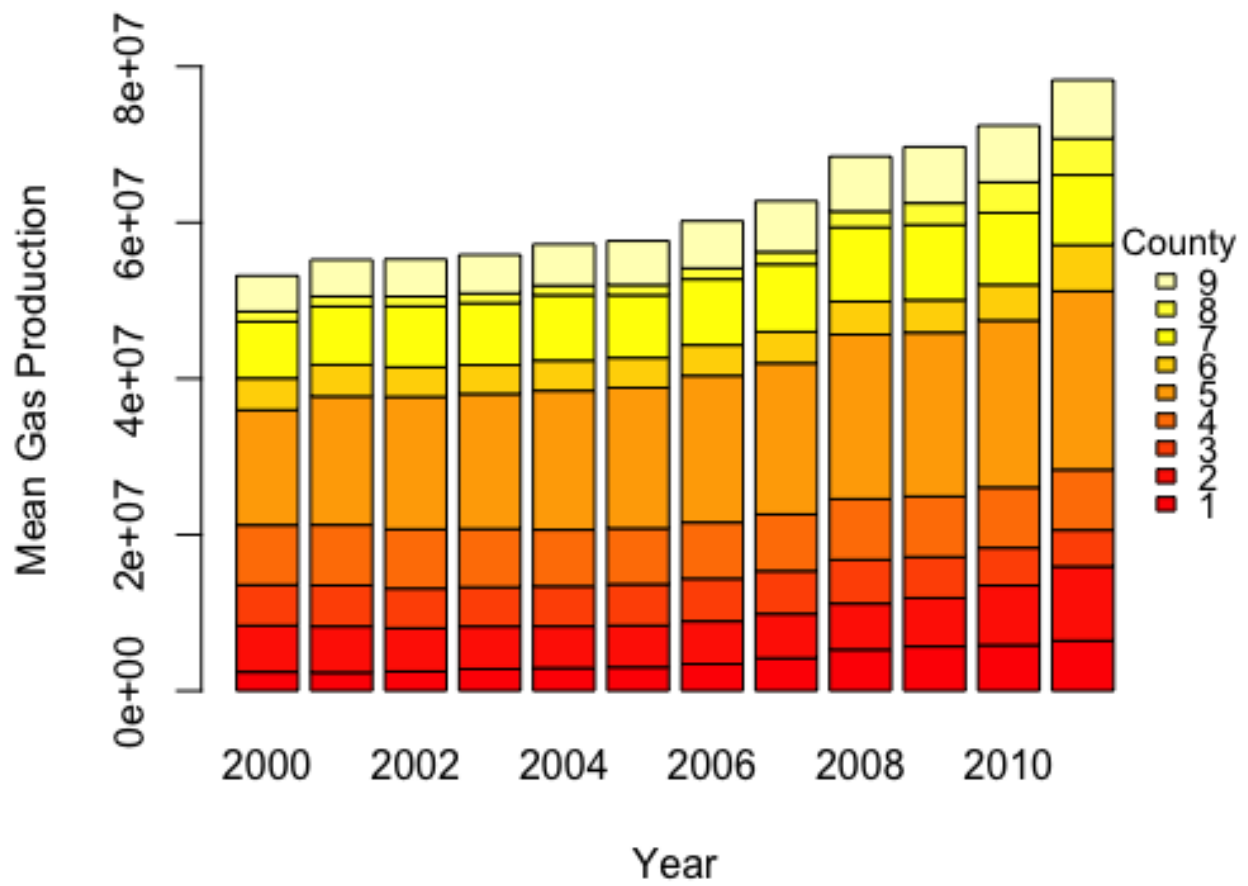
Graph 5. Mean Oil(Gas) Production By County

The following is a stacked bar chart that conveys the mean oil production for each county level in each year.



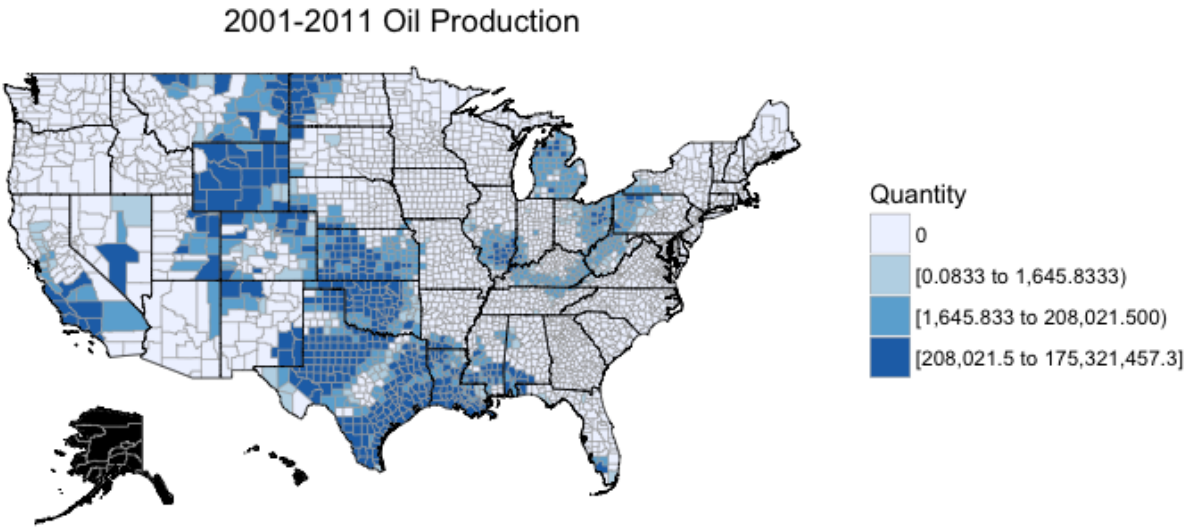
The following is a stacked bar chart that conveys the mean gas production for each county level in each year.

Mean Gas Production By County



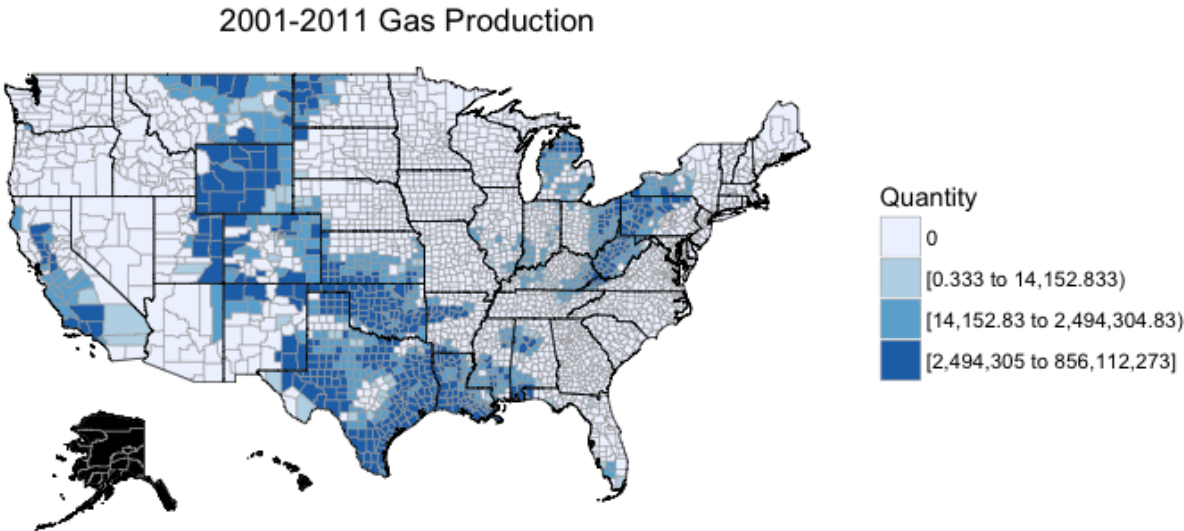
Graph 6. Total Oil(Gas) Production Map For Each County

Next is a U.S national map that displays the distribution of different quantity levels in total oil production of



each County.

Next is a U.S national map that displays the distribution of different quantity levels in total gas production of



each County.