The following text is taken from the capstone project report chapter on Exploratory Data Analysis:

3. EXPLORATORY DATA ANALYSIS

The three final dataframes produced during the data wrangling process were used to conduct visual and exploratory analysis. The main purpose of this phase was to: 1) quality check the data, and identify outliers and issues requiring further data cleaning and wrangling; 2) gather general statistical information, identify correlations and trends, and develop insights about the wells and their production characteristics; and 3) begin to build a basis for predicting future well behavior.

All plotting was done within Jupyter notebooks using either matplotlib or the plotnine ggplot package for python.

Rollup DataFrame EDA

In the analysis of the surface location shapefile, we found that there were 71116 surface locations within the study area. Analysis of the rollup dataframe for all the wells listed in the production spreadsheets indicates that there are a total of 37116 unique wells listed with entries in those spreadsheets over the 21 years analyzed. This suggests that more than half of the wells existing in the study area have been in production over the 1999 - 2019 time period, although as will be shown in the following bar charts, some of these wells are of a status that suggests they have not yet produced.
The bar chart below provides a breakdown of the number of producing wells by county within each of the 16 counties included within the study area. Clearly the bulk of the producing wells are in Weld County, where the giant Wattenberg Field exists.

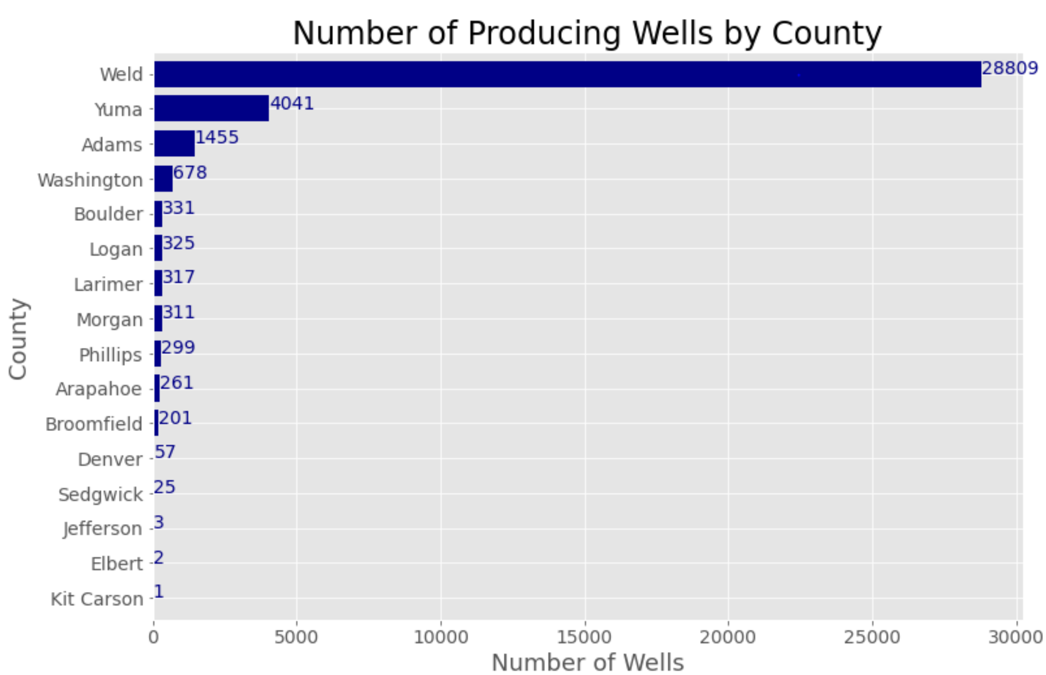

Fig 5: Horizontal bar chart showing the count of producing wells by county from 1999 through 2019 within the study area.

The following barchart shows a breakdown of the type of wells producing within the study area. The bulk of wells producing over the 21 year time period are non-horizontal wells.

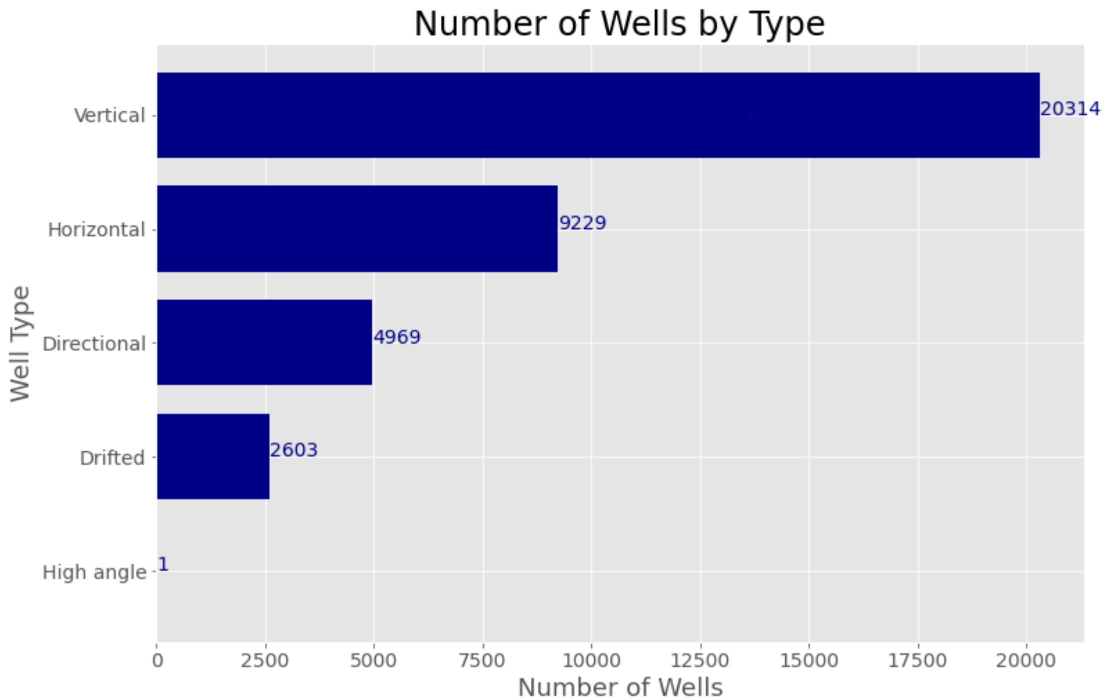

Fig 6: Horizontal bar chart showing the count of producing wells by well type (horizontal vs vertical and near vertical variants).

The following barchart shows a breakdown of the type of each well's final status. The bulk of wells are still actively producing.

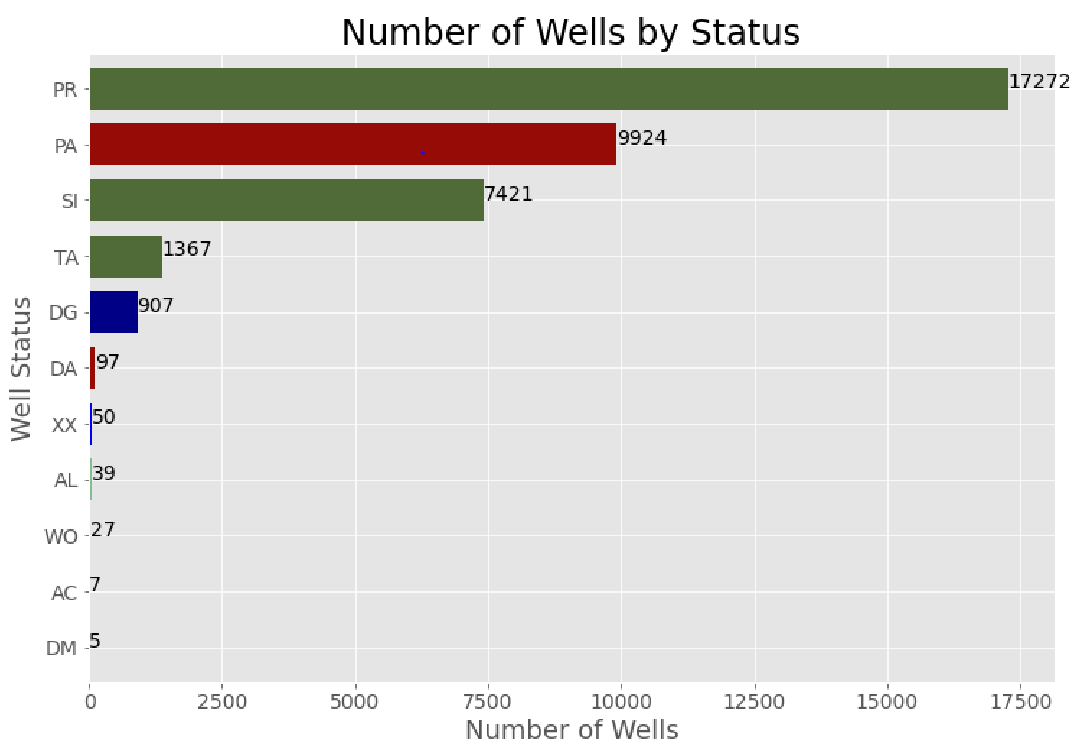

Fig 7: Horizontal bar chart showing the count of producing wells by their last status. Those shaded in green are considered active. Those shaded in red are considered abandoned and those shaded in blue are in progress (permitted, drilling, active location, etc.)

Each well was also tagged with an 'Active' or 'Abandoned' flag, given it's status. The breakdown of active versus abandoned wells is shown in Figure 8 below.

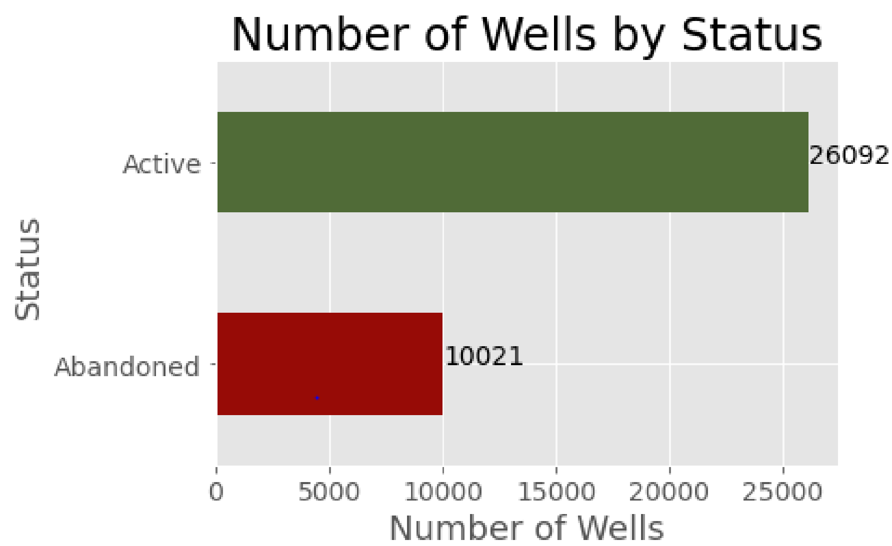

Fig 8: Horizontal bar chart showing the count of producing wells by active or abandoned category. Wells with final status of 'DA' or 'PA' have gotten an abandoned flag (1). Wells with final status of 'PR', 'SI', 'TA', 'WO', or 'DM' have gotten an active flag (0). Wells with final status of 'DG', 'XX', 'AL', or 'AC' were not included in the count.

The following stacked horizontal bar chart shows a breakdown of producing wells by well type (horizontal or vertical) and active / abandoned category. There is a much larger percentage of abandoned vertical wells than there are horizontal wells. In fact, only 120 horizontal wells have been abandoned in the study area.

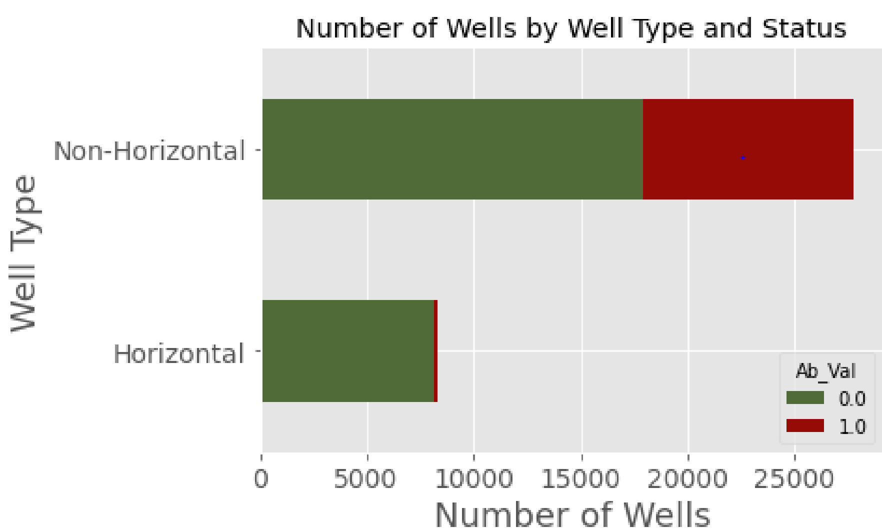

Fig 9: Stacked horizontal bar chart showing the count of producing wells by well type and active (0) or abandoned (1) category (color-coded).

The following horizontal bar chart shows a breakdown of operators by producing well count. The top 20 producers are shown.

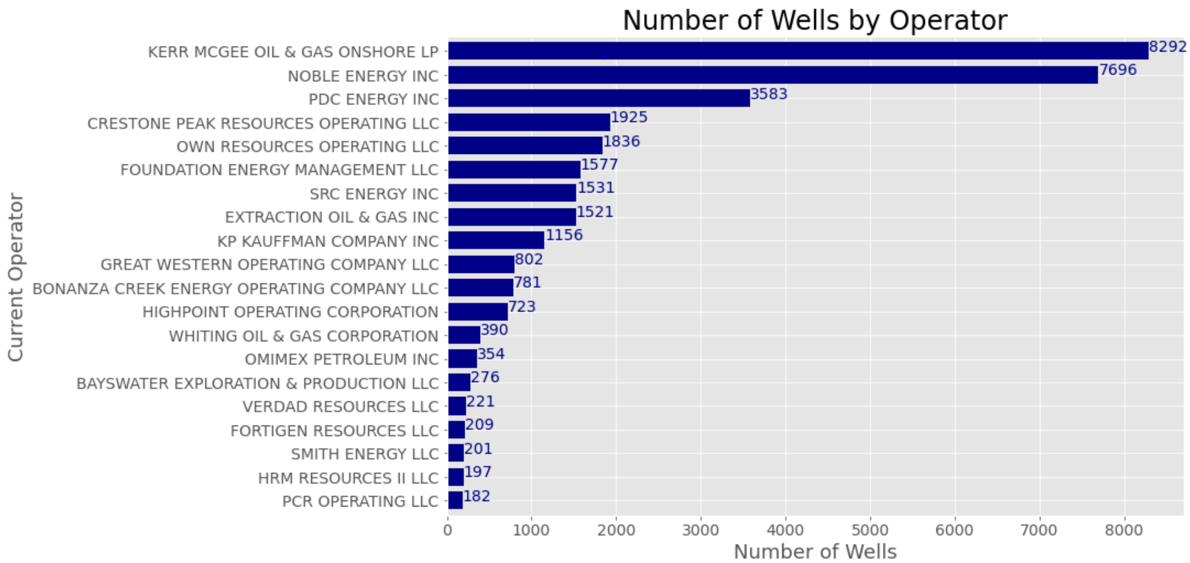

Fig 9: Horizontal bar chart showing the count of producing wells by current operator, with the top 20 most common operators shown.

Need to add in additional stacked bar charts showing production by unit and well type and scatter plots of production by unit.

QGIS Map-Based EDA

Need to work on this section.

Time Series DataFrame EDA

The time series provides an excellent data source for reviewing the production information over time. Figure 5 shows the total production of gas, oil, and water over the time period provided in the production spreadsheets. Production of all entities has increased over time. A noticeable steep increase in oil and gas production can be seen in 2014. The oil and water rates then decline over the 2015 to 2017 time period with gas production flattening. Since 2017, production of all fluids have risen sharply.

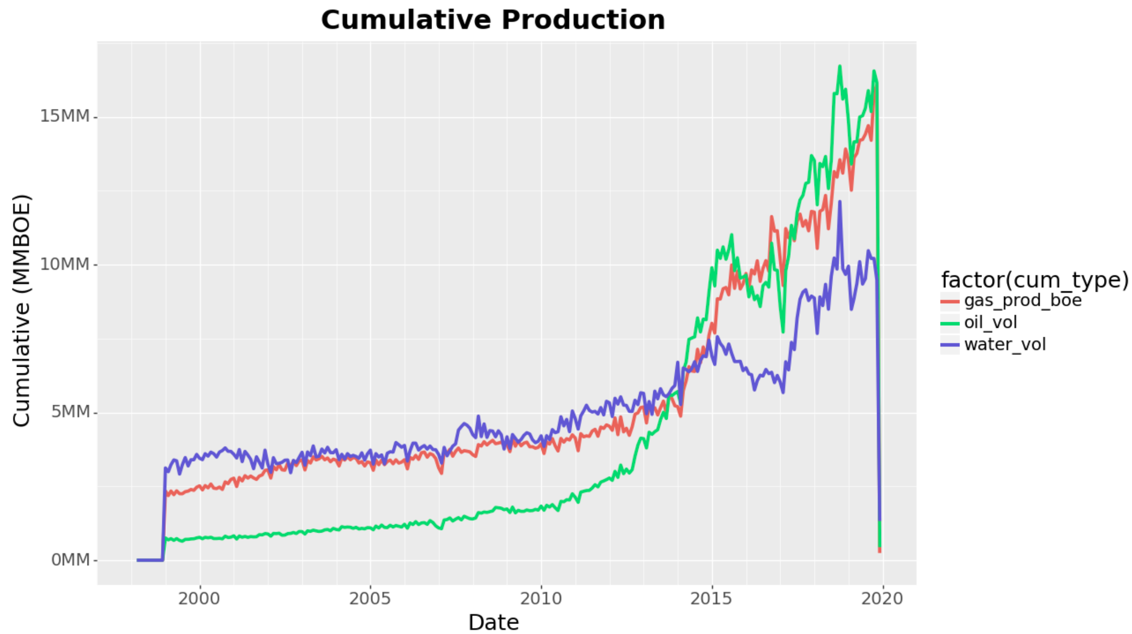

Fig 10: Cumulative production of gas, oil and water (MMBOE) over the production file time period (1999-2019).

Figure 11 shows the same information, but with the production entities stacked.

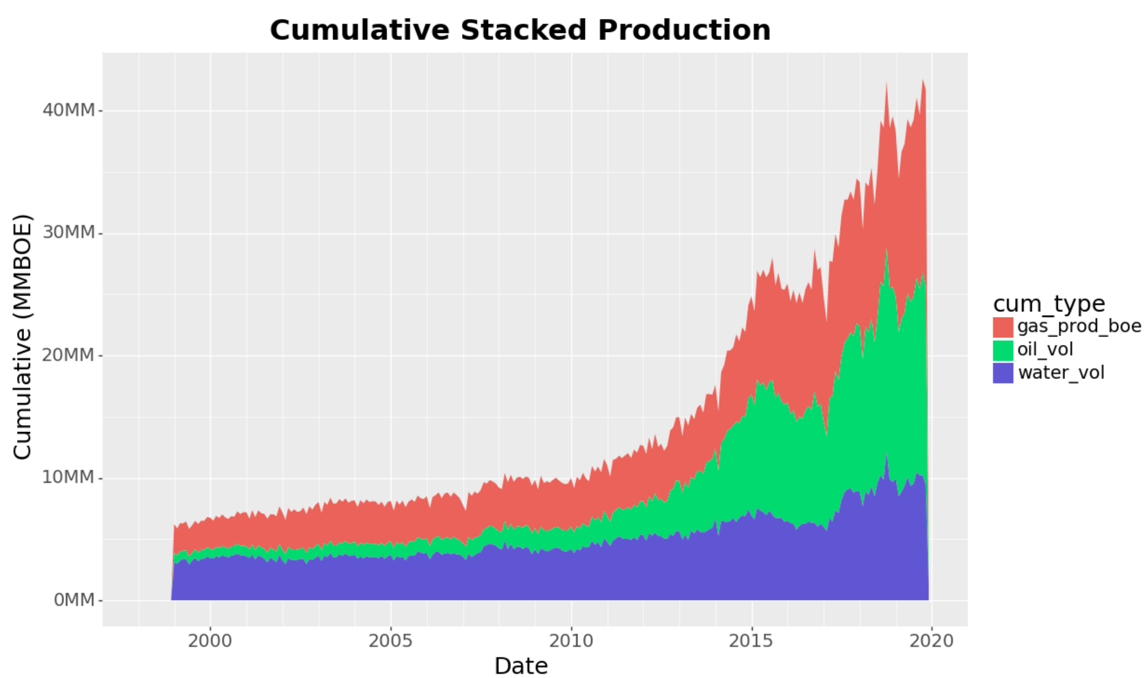

Fig 11: Cumulative Stacked Production of gas, oil and water (MMBOE) over the production file time period (1999-2019).

The following two figures show the total production over time for all non-Horizontal and Horizontal wells broken out separately. The same y-axis scale has been used as in Figure 6 above to understand the relative production from these 2 well type categories as a part of the total production.

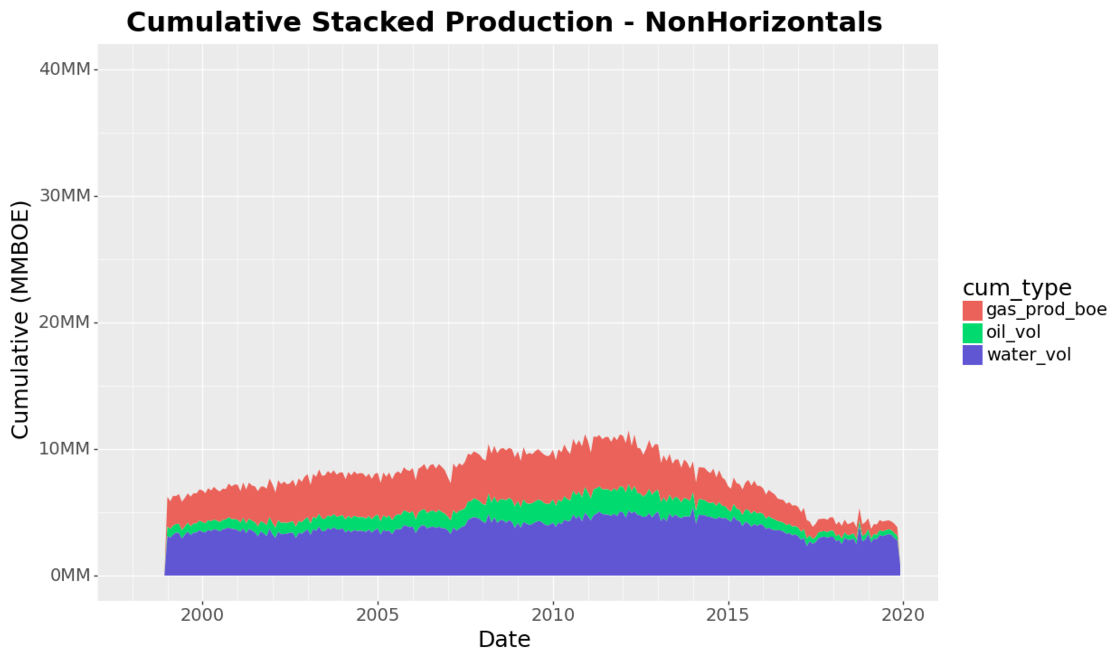

Fig 12: Cumulative Stacked Production of gas, oil and water (MMBOE) for non-Horizontal wells over the production file time period (1999-2019).

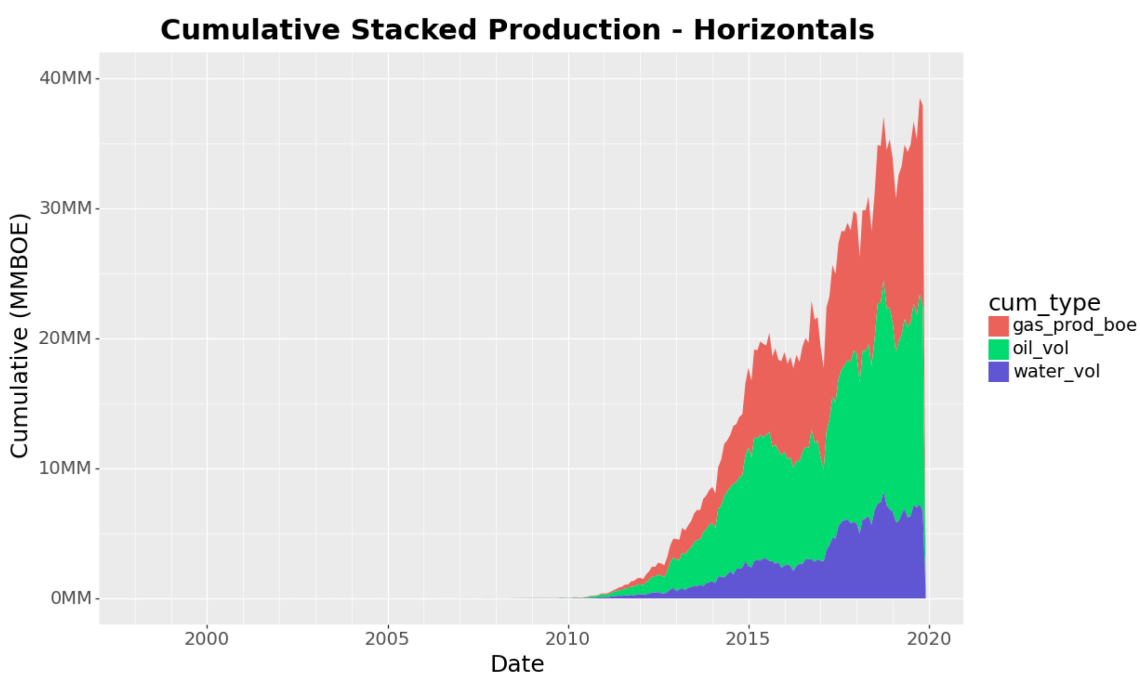

Fig 13: Cumulative Stacked Production of gas, oil and water (MMBOE) for Horizontal wells over the production file time period (1999-2019).

These plots indicate that non-Horizontal production peaked toward the end of 2011 and has continued to decline since that time, while horizontal production, since its inception in 2010, has quickly grown to account for the majority of the fluid produced over the last five years.

The same cumulative stacked production plots were also created for the non-Horizontal and Horizontal well population by producing formation. Figure 14 below indicates that the JSND and

NB-CD units have accounted for the majority of the production from non-horizontal wells in the study area. CODL production peaked in 2007 and has been declining since that time, but this has been compensated by the increasingly common practice of commingling Niobrara and Codell production together (NB-CD). The latter peaked near the end of 2012 and has continued to decline. Note also the substantial amount of water production from the JSND and DSND, and the very low oil production (mostly gas) in the NBRR non-horizontals.

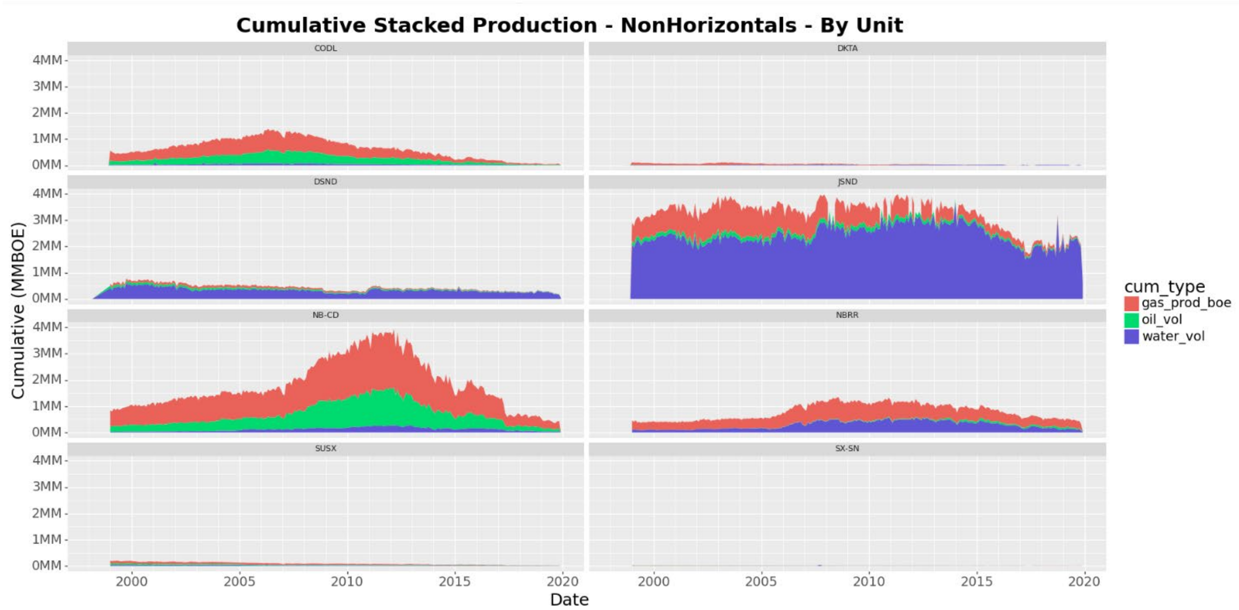

Fig 14: Cumulative Stacked Production of gas, oil and water (MMBOE) for non-Horizontal wells over the production file time period (1999-2019) grouped by producing formation.

Figure 15 below indicates that the NBRR unit is by far the largest producing unit from the horizontal well population. The CODL production is quite small in contrast.

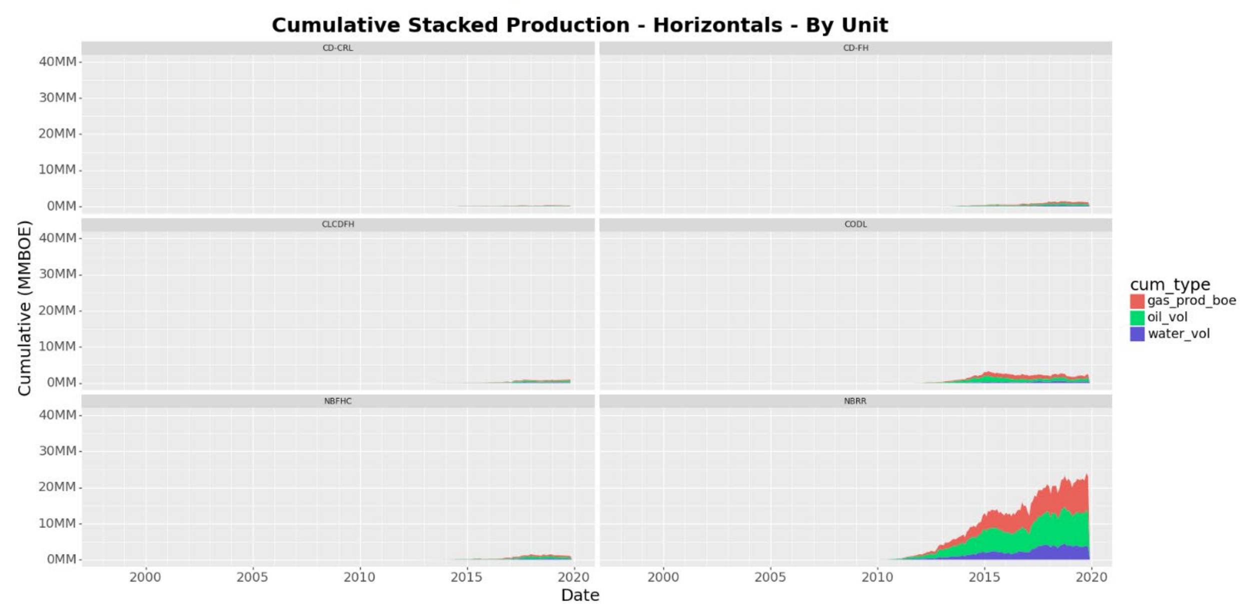

Fig 15: Cumulative Stacked Production of gas, oil and water (MMBOE) for Horizontal wells over the production file time period (1999-2019) grouped by producing formation.

The stacked plots below contrast further the NBRR, CODL and NB-CD production broken out by well type.

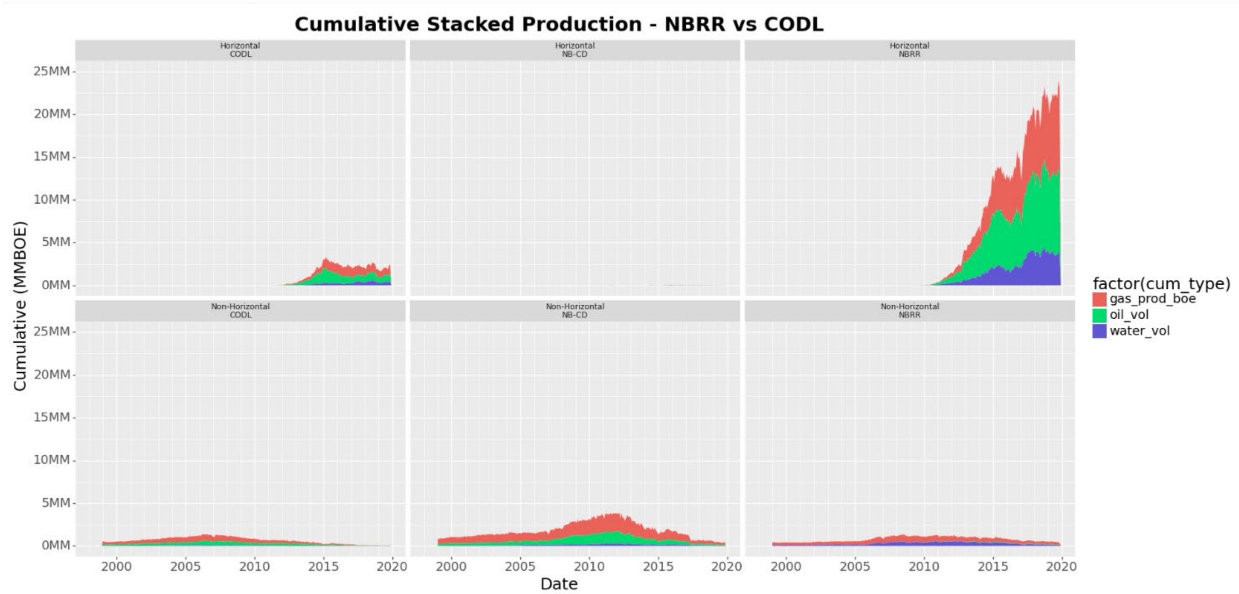

Fig 16: Cumulative Stacked Production of gas, oil and water (MMBOE) for all NBRR and CODL producing wells over the production file time period (1999-2019) grouped by producing formation and well type (Horizontal and Non-Horizontal).

To understand the production decline trends in both the NBRR and CODL by individual horizontal wells, the production was broken out by quintiles, first using the first 6 months initial production (with first 3 months production removed). For each quintile, the production was then plotted for each well over time starting with the first month of production. The values were also averaged for each quintile. (Need the count of wells in each quintile)

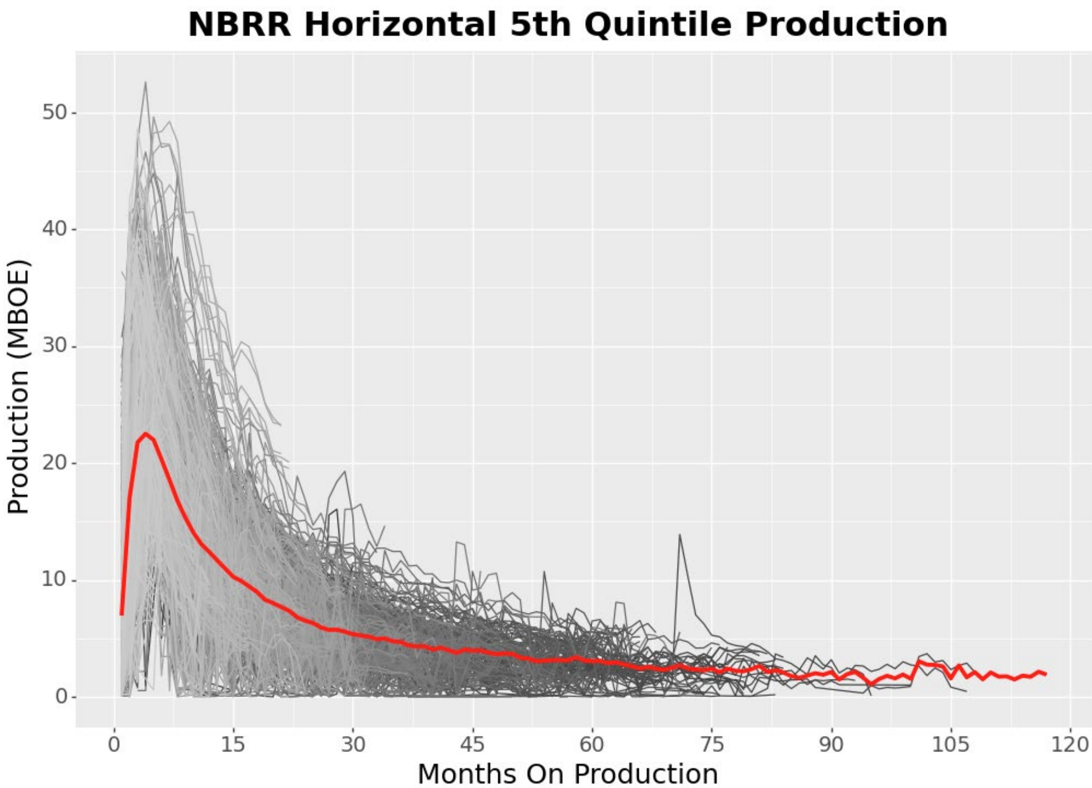

Fig 17: NBRR 5th Quintile Production by Well Since First Month. Red line is average of all wells by month.

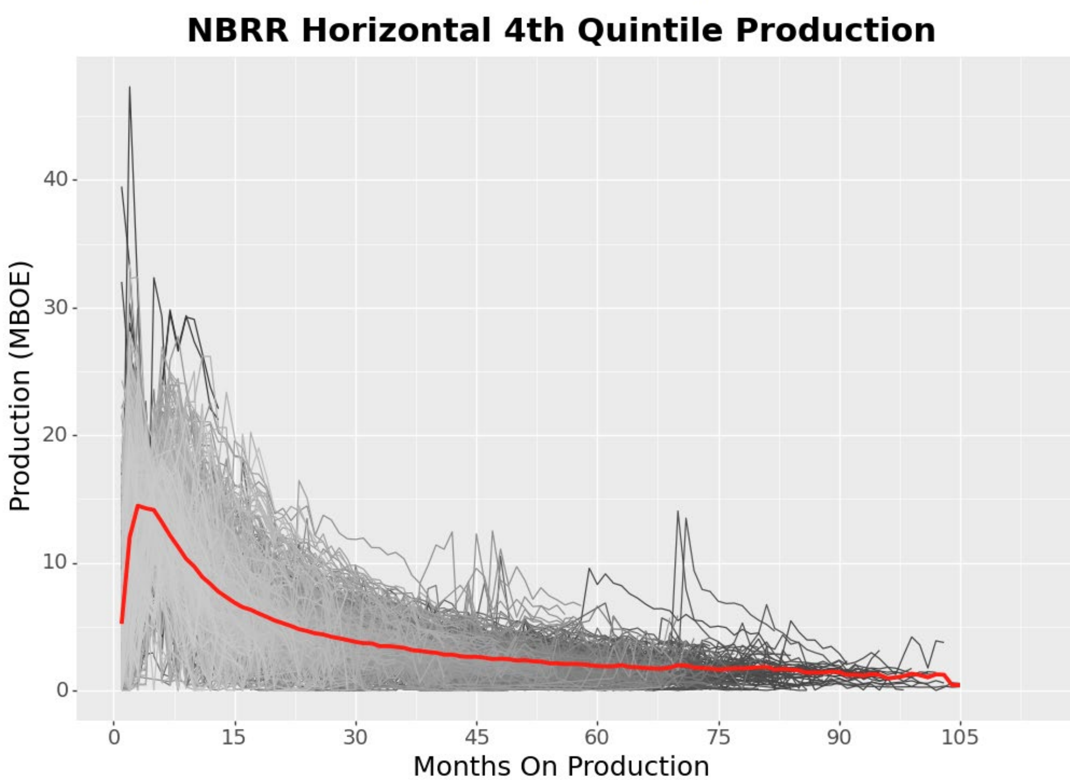

Fig 18: NBRR 4th Quintile Production by Well Since First Month. Red line is average of all wells by month.

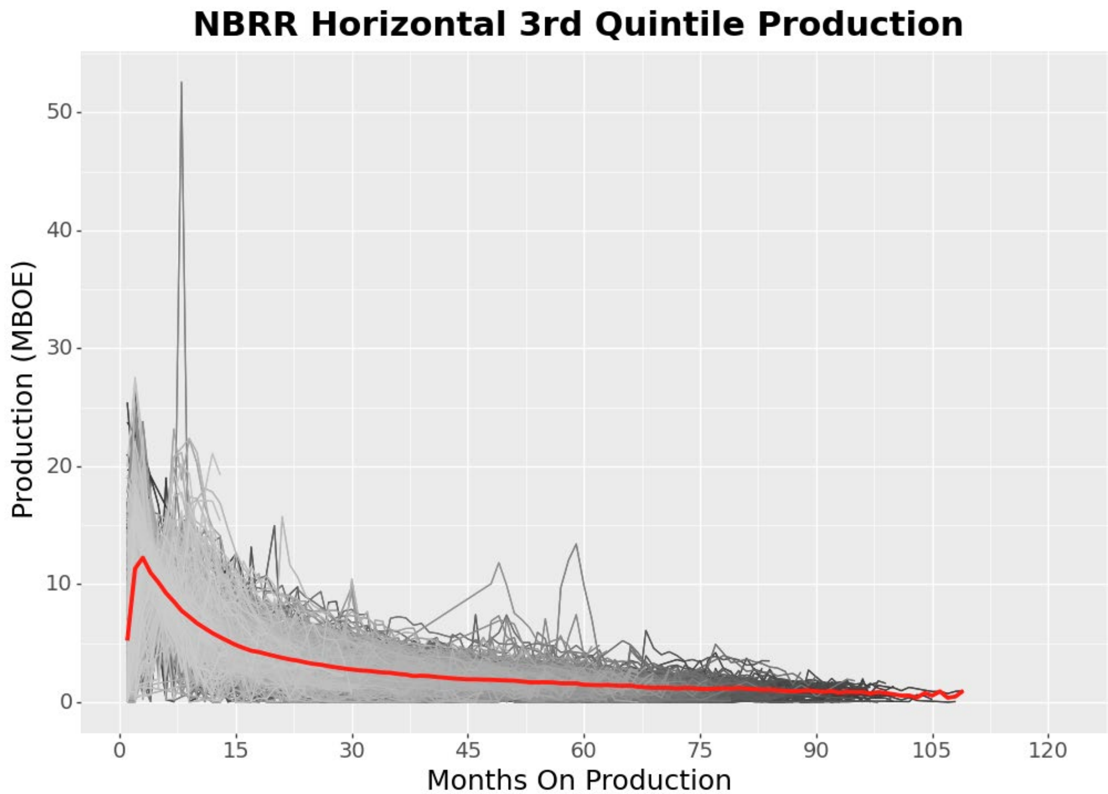

Fig 19: NBRR 3rd Quintile Production by Well Since First Month.Red line is average of all wells by month.

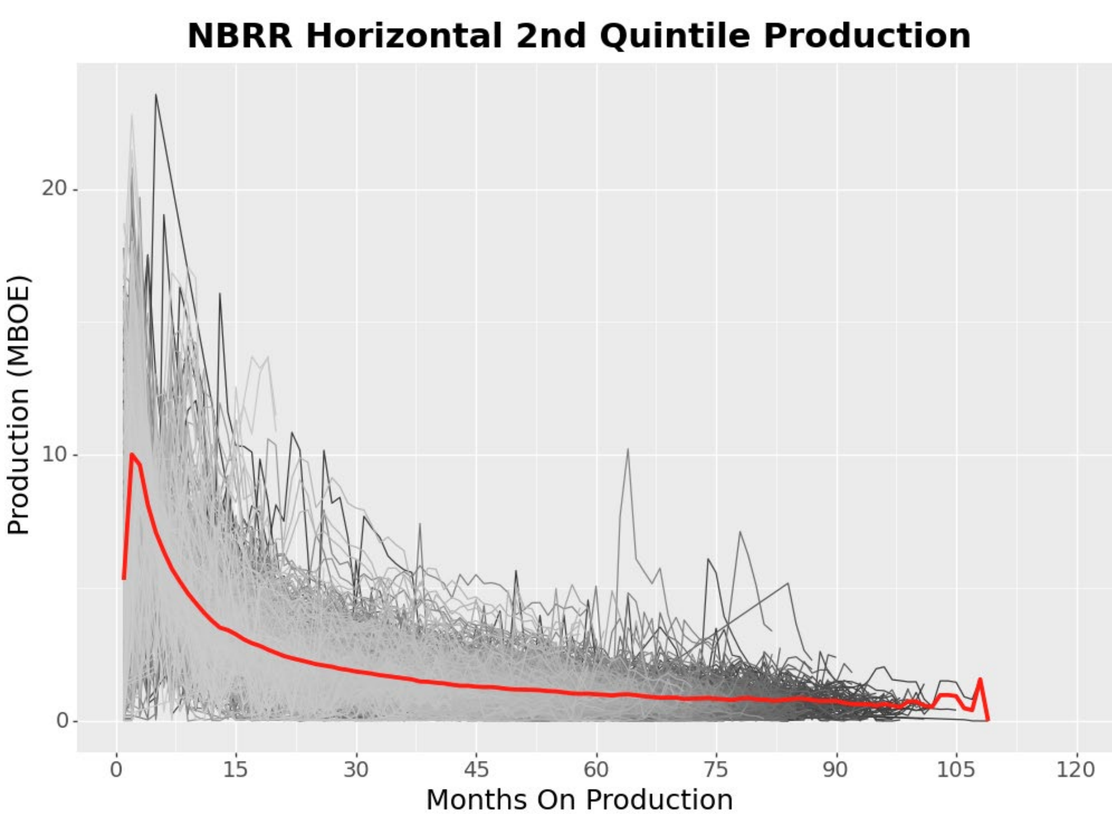

Fig 20: NBRR 2nd Quintile Production by Well Since First Month. Red line is average of all wells by month.

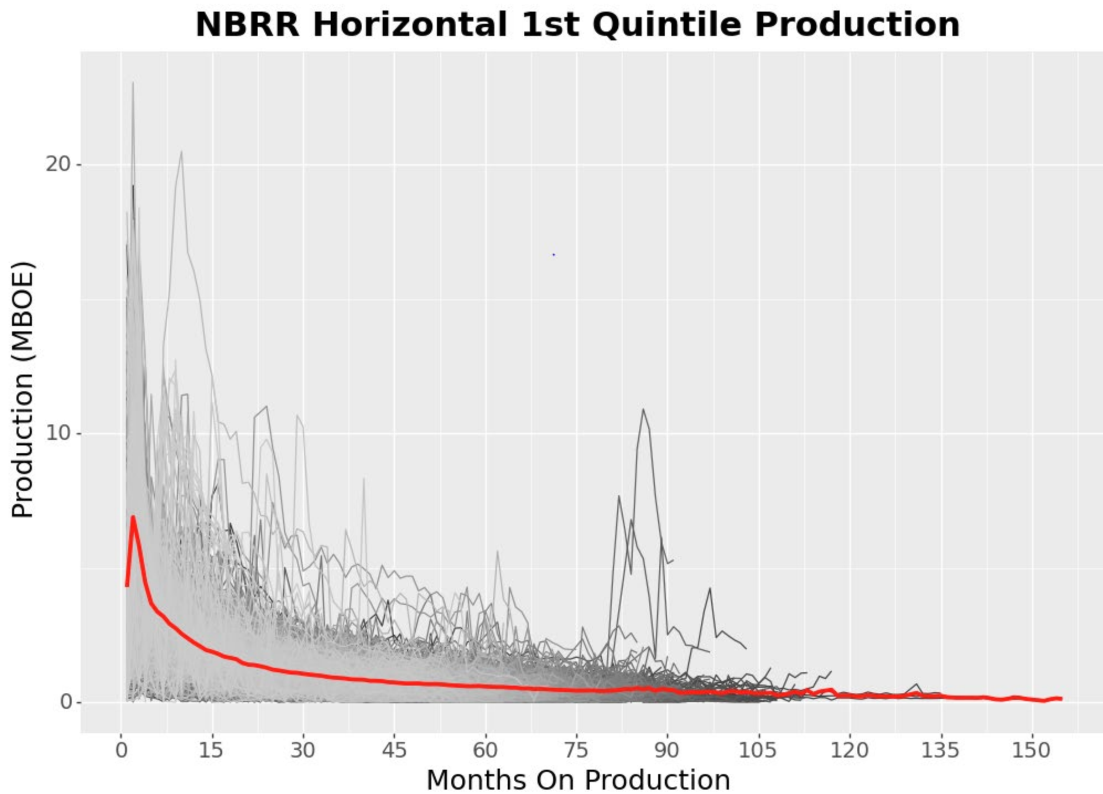

Fig 21: NBRR 1st Quintile Production by Well Since First Month.Red line is average of all wells by month.

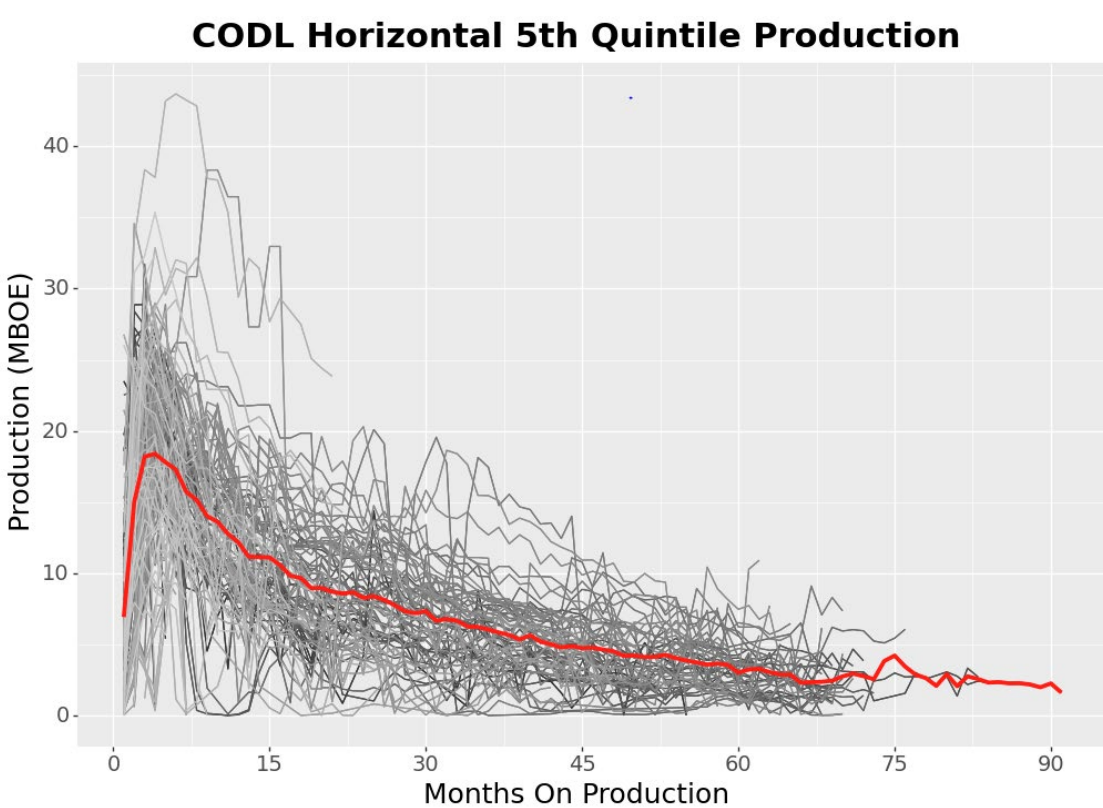

Fig 22: CODL 5th Quintile Production by Well Since First Month. Red line is average of all wells by month.

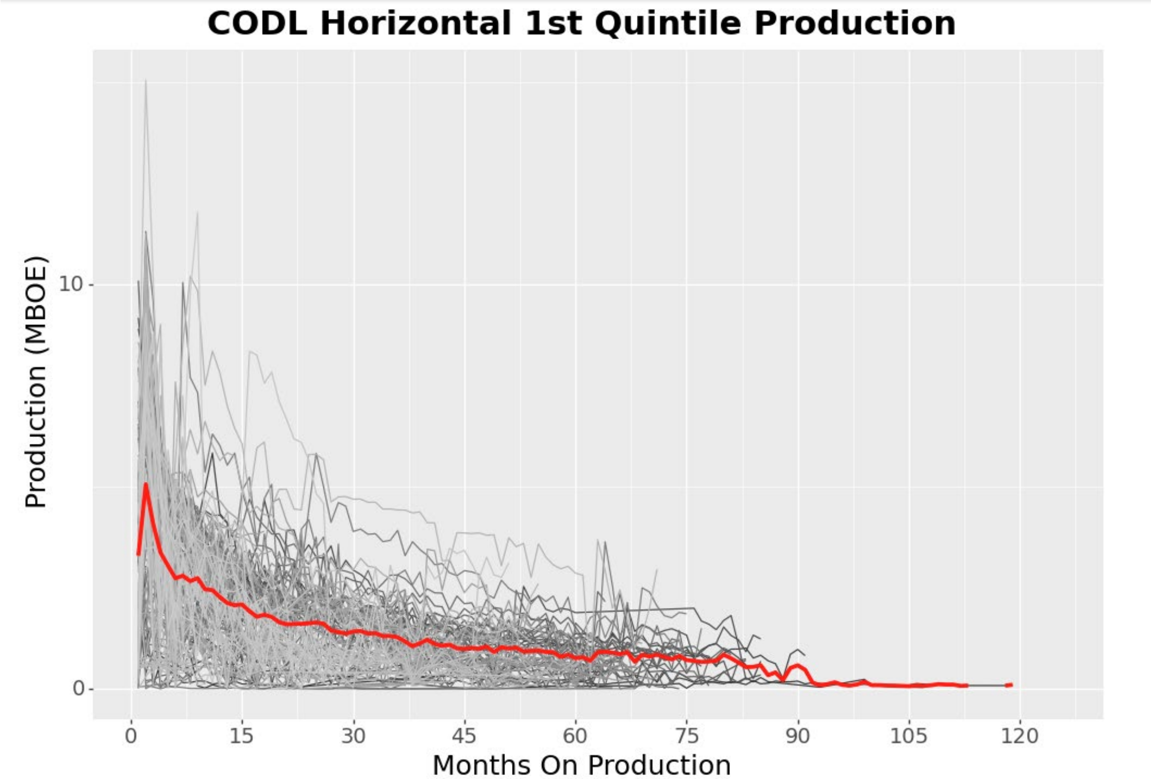

Fig 23: CODL 1st Quintile Production by Well Since First Month. Red line is average of all wells by month.

Even though it is clear that each NBRR quintile shows sequentially less initial production, the longevity of the wells across quintiles is similar. Comparison between the NBRR and CODL indicates that even though there are fewer CODL wells and less production, the best 5th quintile CODL wells are almost on par with the best 5th quintile NBRR wells in terms of initial production rates. The 5th quintile CODL wells, however, have not produced as long.

The following figures show the production for only the few horizontal wells that have been abandoned, for both the NBRR and CODL. Only 3 CODL wells have been abandoned as of the beginning of 2020.

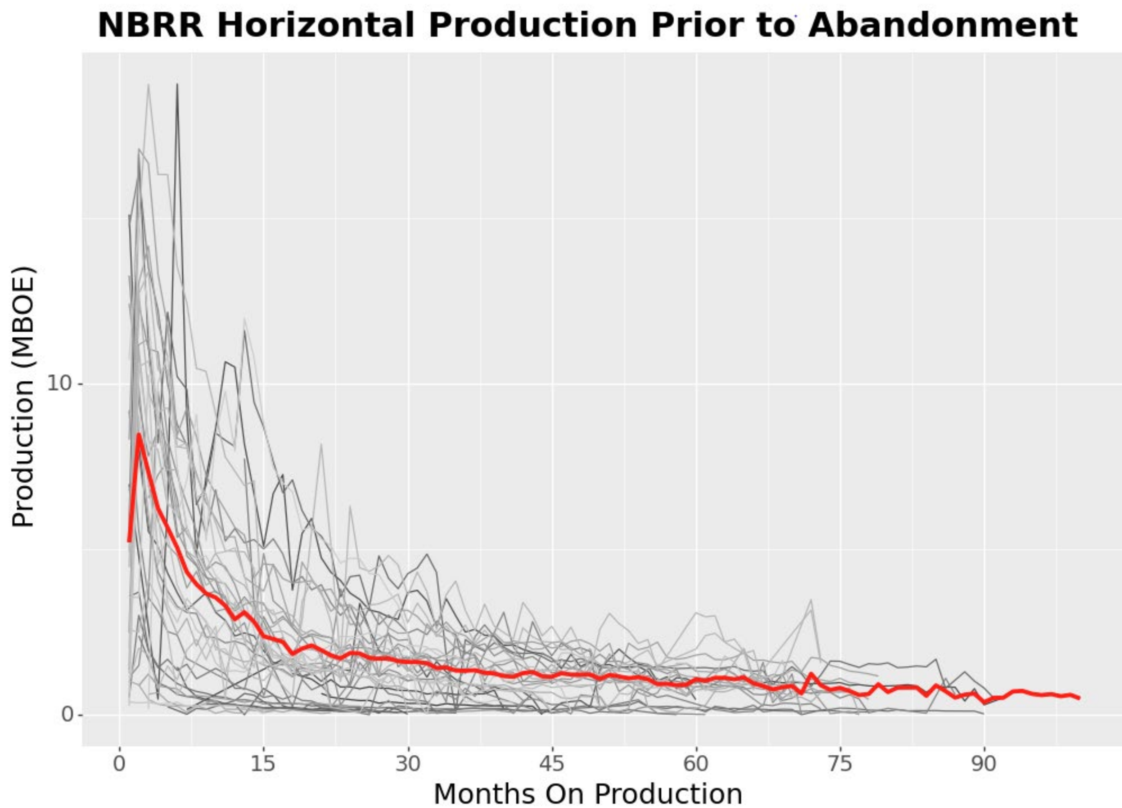

Fig 24: NBRR Production by Well Since First Month only for wells that have since been abandoned. Red line is average of all wells by month.

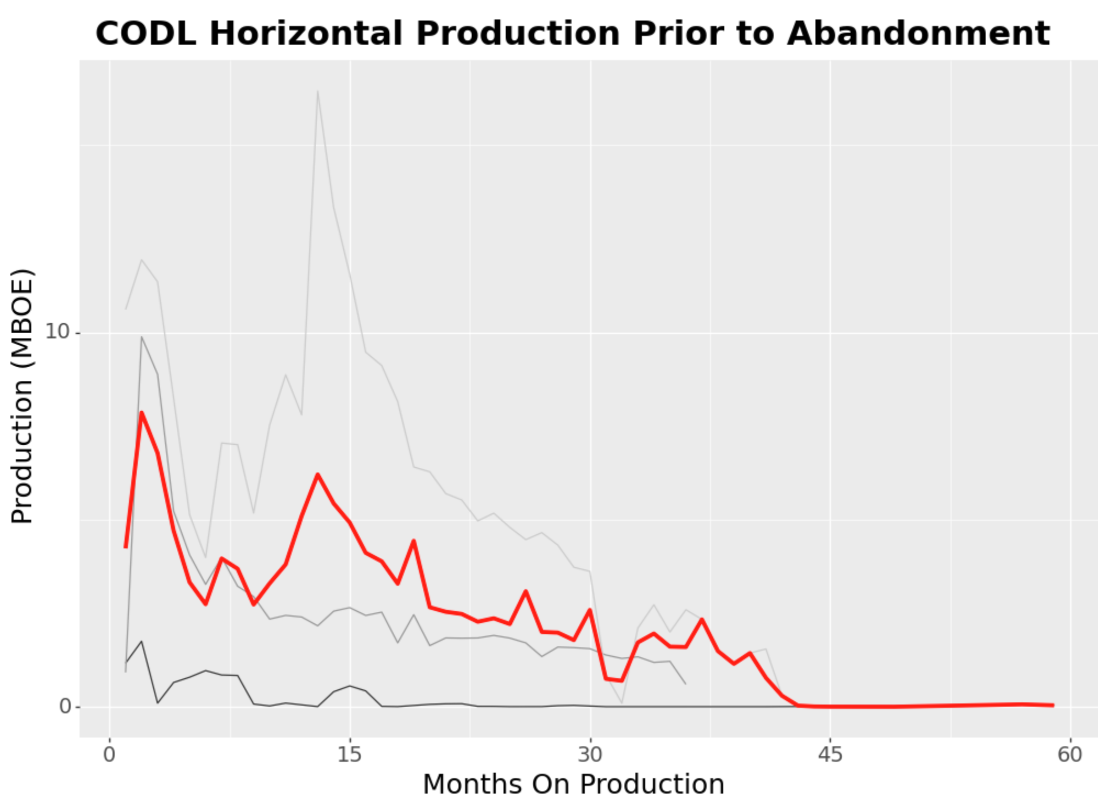

Fig 25: CODL Production by Well Since First Month only for wells that have since been abandoned. Red line is average of all wells by month.

Further Work - Questions Remaining to Be Answered

Remove outliers?; compare and contrast other quintilized population - which is better