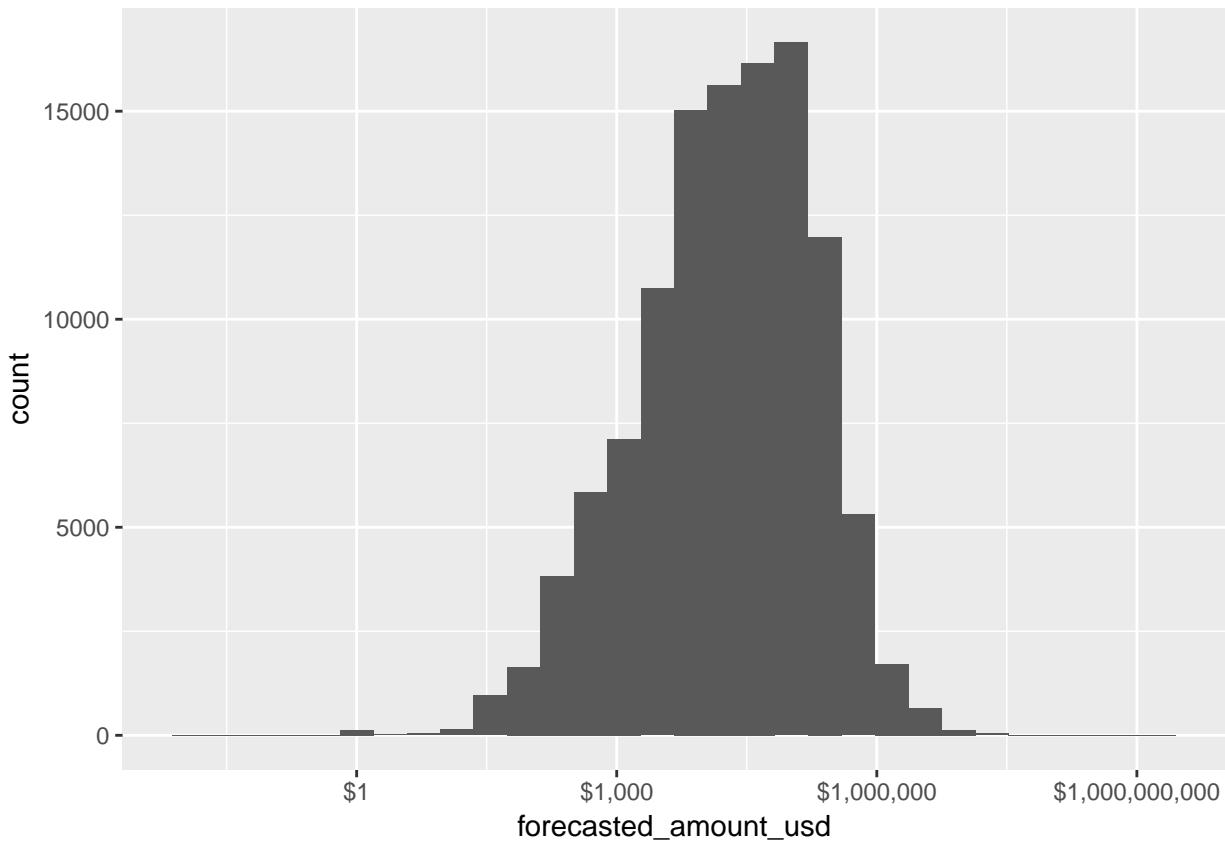


Gasbarro Capstone Project: Opportunities Analysis - Statistics - Section 5

November 14, 2018

STAGE: Count of Forecasted Dollar Amount (USD)

This first plot shows the count (number of deals) and where they fall in a range of US dollars (\$) for the projected Forecasted Dollar Amount (what the deal was proposed to the customer for). As you can see, the vast majority of them fall in a range between USD 500 and USD 1 million, with the concentration being between USD 1,000 and 900,000.

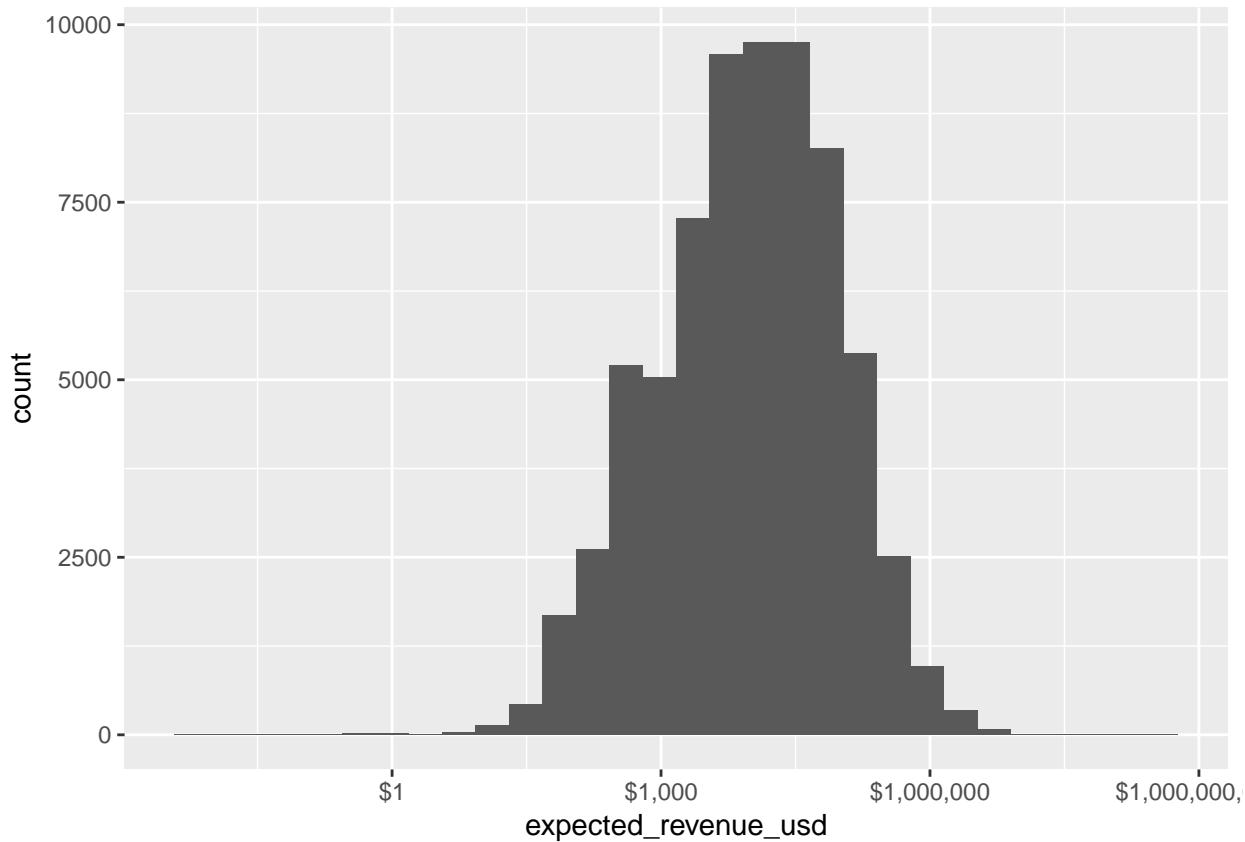


STAGE: Count of Expected Revenue Amount (USD)

The Expected Revenue is the amount of money we expect to make from the deal when it closes and commits. A deal might be forecasted for USD 500,000, and the expected revenue might be USD 250,000 (meaning there are USD 250,000 of costs).

The bar plot below for Expected should align with the plot above for Forecasted. Are there noticeable differences in the shape or outline of these plots? If they are closely match, that indicates that expected revenue is aligned with the forecasted amount (meaning, costs are controlled proportionate to the deal). If there is significant difference, this would suggest that there is a misalignment between what was Forecasted and what we Expect to make. For example: If Expected Revenue looks significantly lower, that might indicate

that the costs to staff the project increased, meaning our internal costs consumed more money, and the expected revenue decreased proportionately.

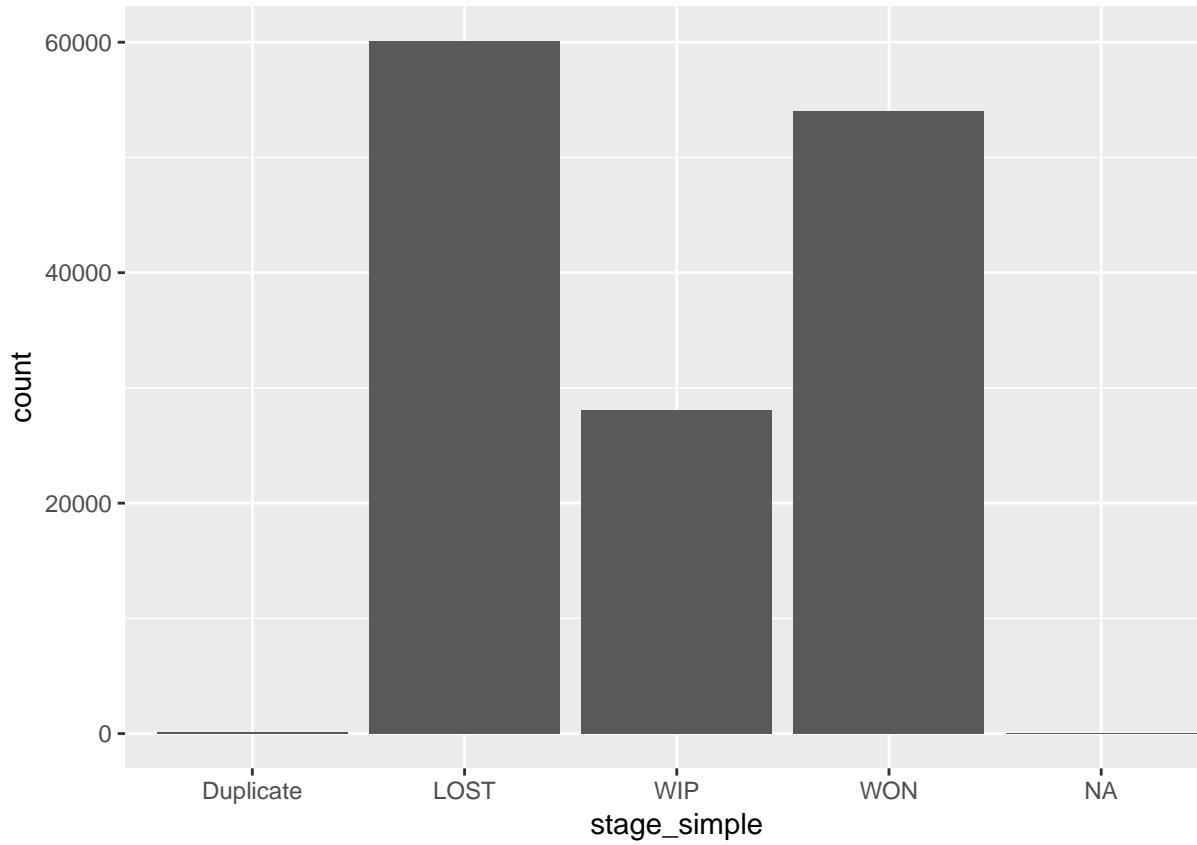


Is there a way to mount one graph on the other?

STAGE: Count by Year

One of the original questions our group asked is, “What is overall number of deals we WON AND CLOSED vs. DROPPED and LOST?”

To answer this question, plot a bar graph that shows the 5 categories in Stage (simple_stage) and the number of deals in each category. As you can see in the chart below, only 3 of the categories (WON, LOST, WIP) have the

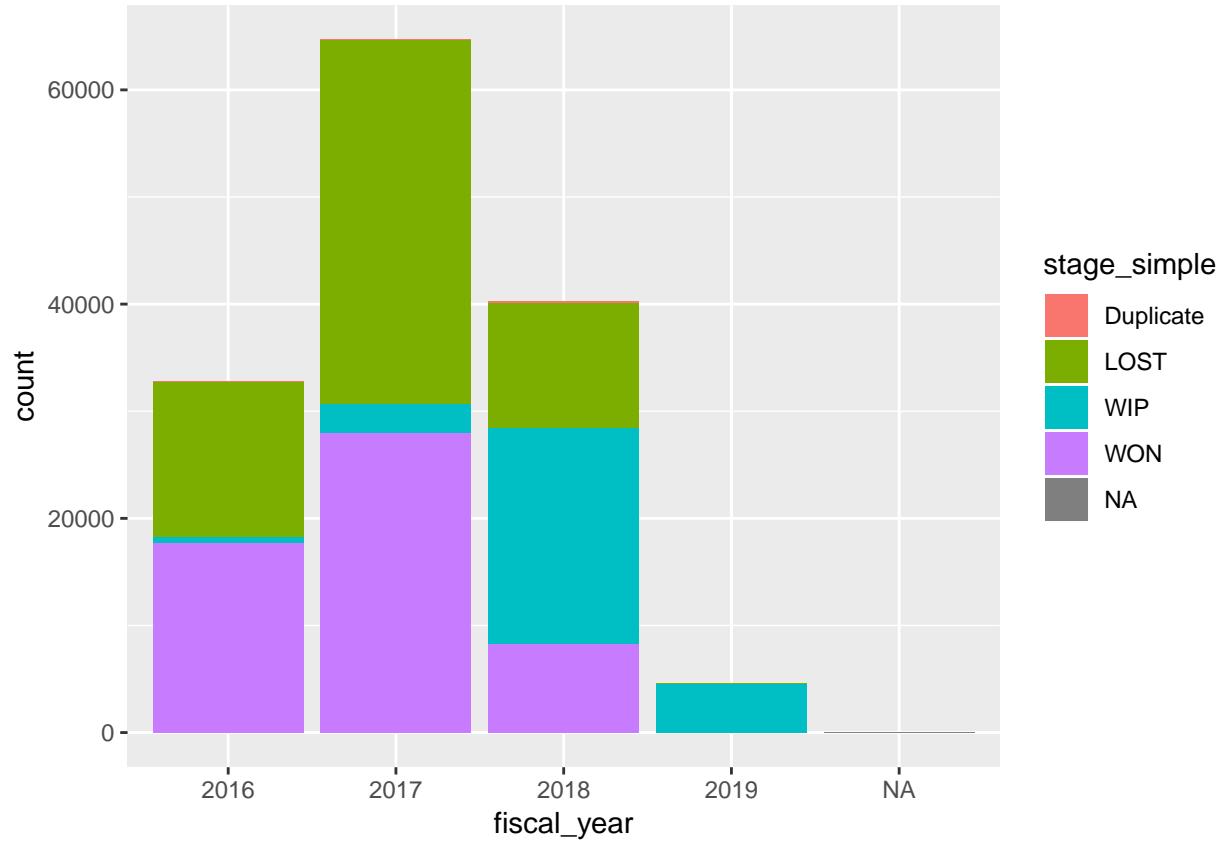
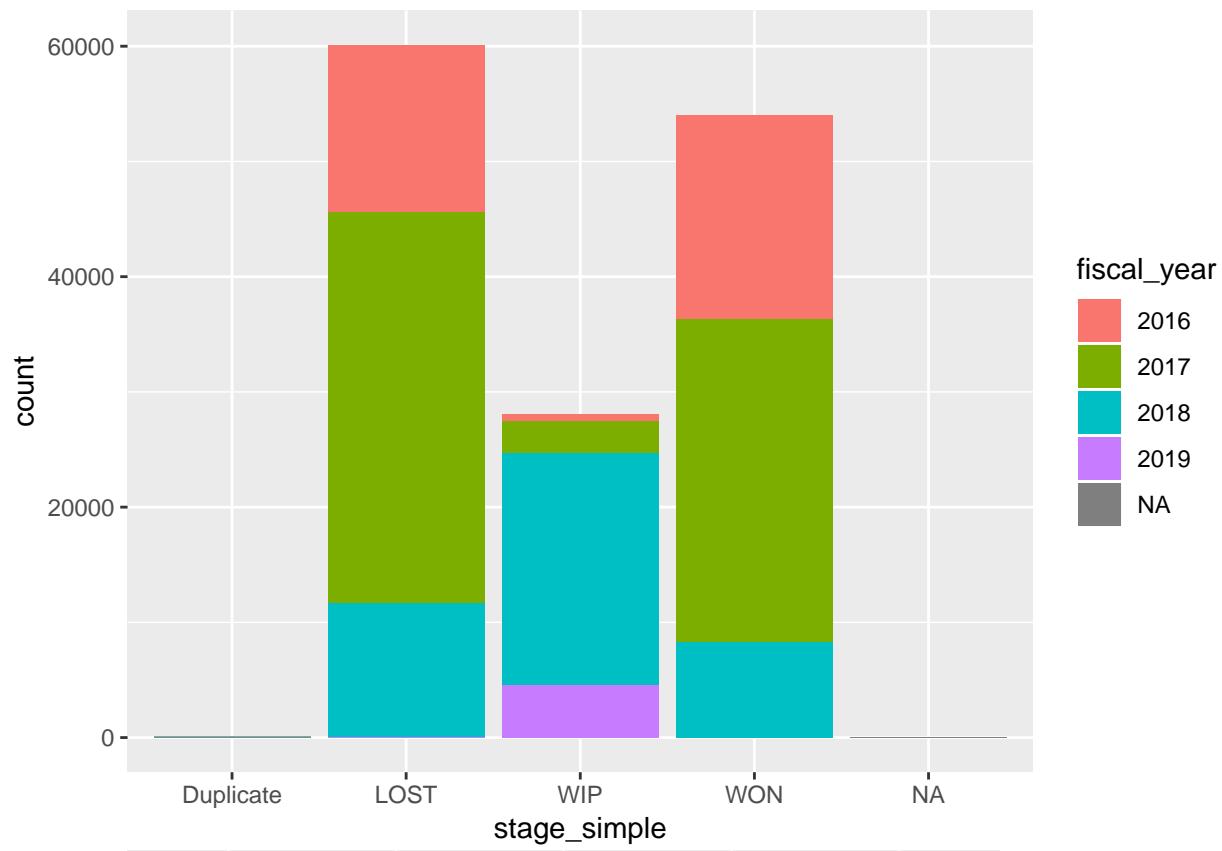


majority of the deals.

Then we wanted to know, “What is overall number of deals WON AND CLOSED, vs. DROPPED and LOST, grouped by YEAR?”

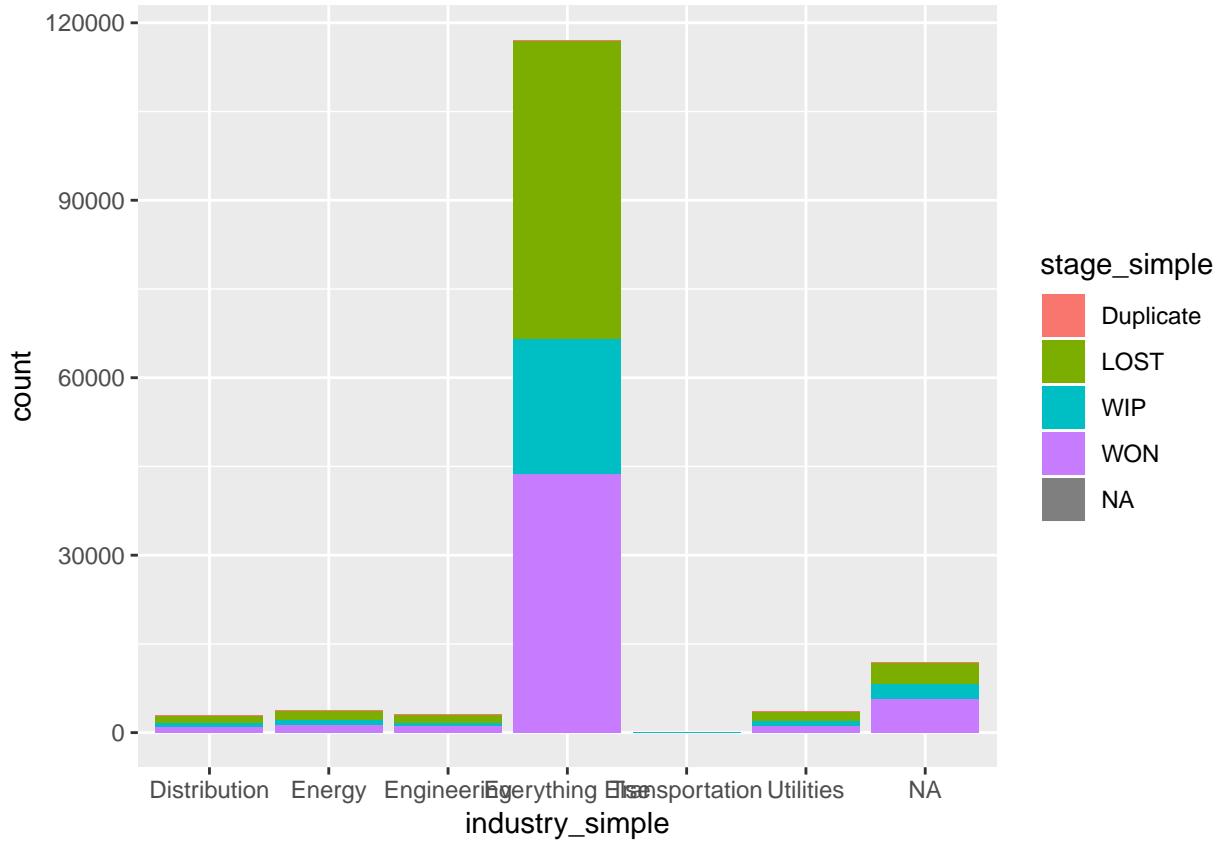
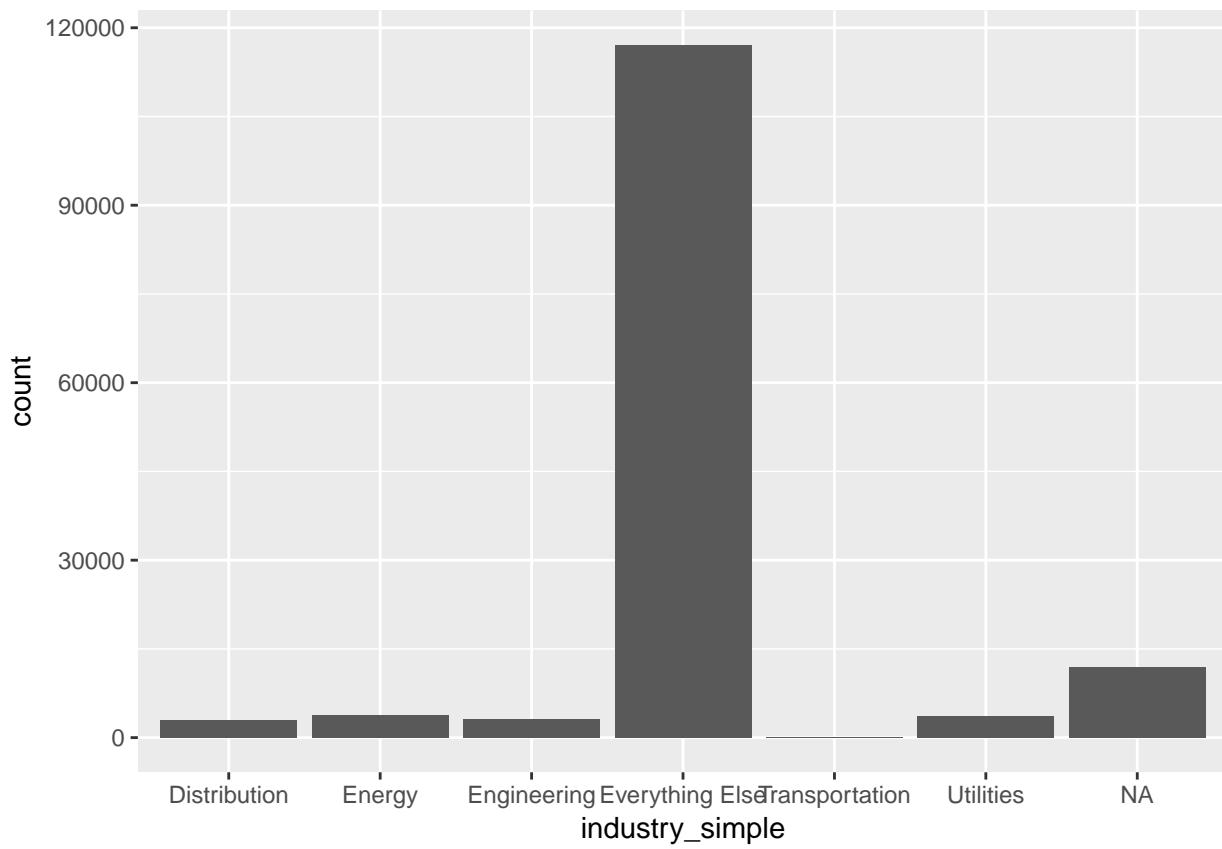
Below are two bar graphs showing the same information.

1. The first chart simply adds a fill to the previous chart (in which the Stage is our baseline X axis). This clearly shows 2017 had the highest counts of WON and LOST deals, more than 2016 or 2018.
2. The first chart changes the X axis to the year, and instead uses the fill to identify the stage. This makes it less easier to quickly identify the year which had the highest counts, but does read nicely, and might be used for other charts (e.g. Industry).



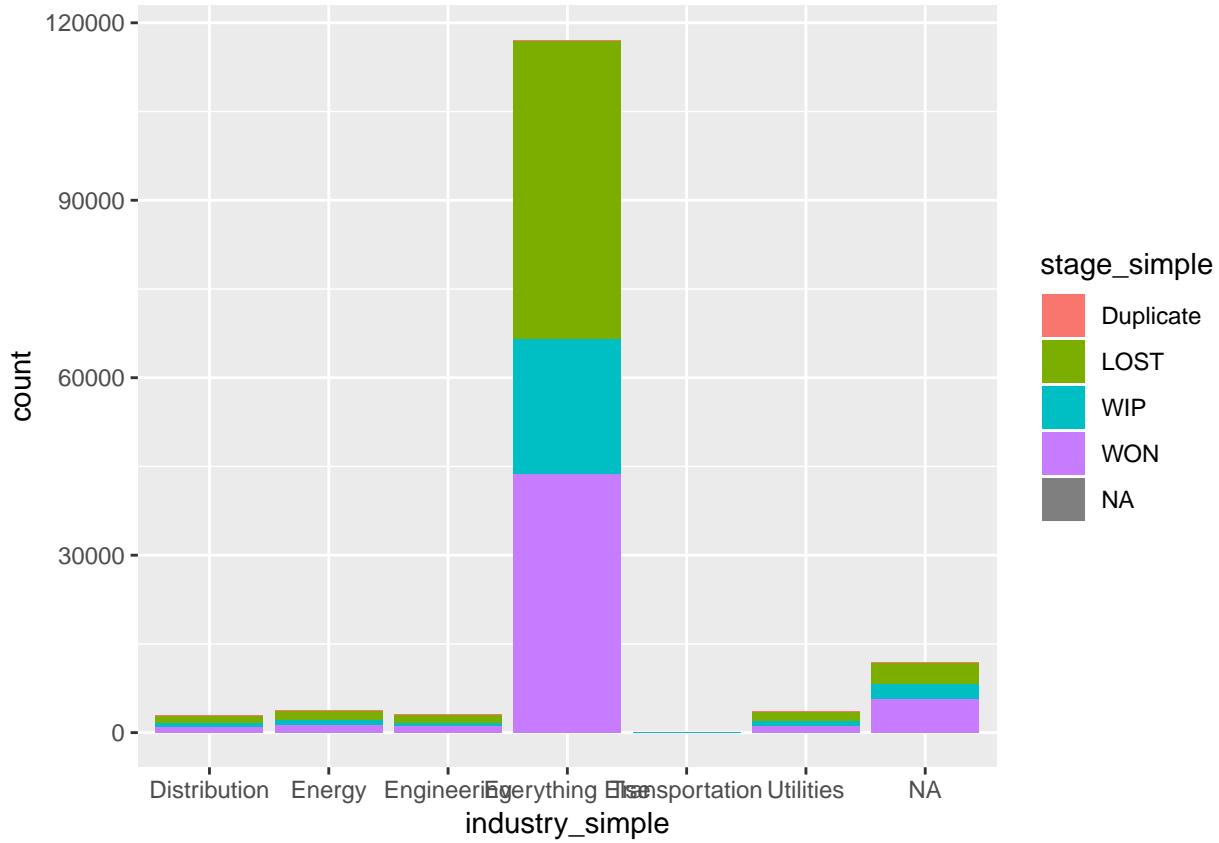
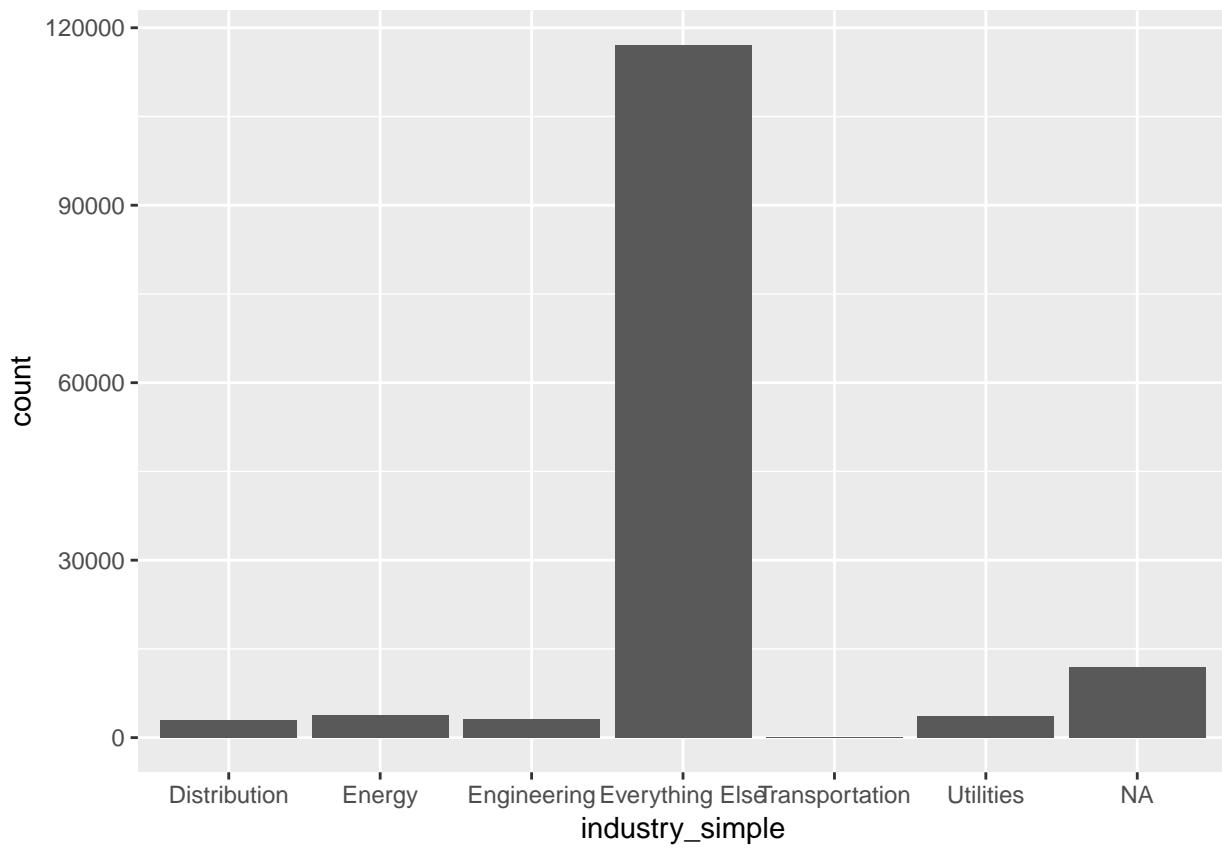
STAGE: Count by Industry

We use the last bar graph as the basis for looking at the count by Stage across the industries. However, this chart is nearly useless: “Everything Else” dominates, while our areas of focus (Distribution, Engineering, Utilities) reduced to very small amounts.



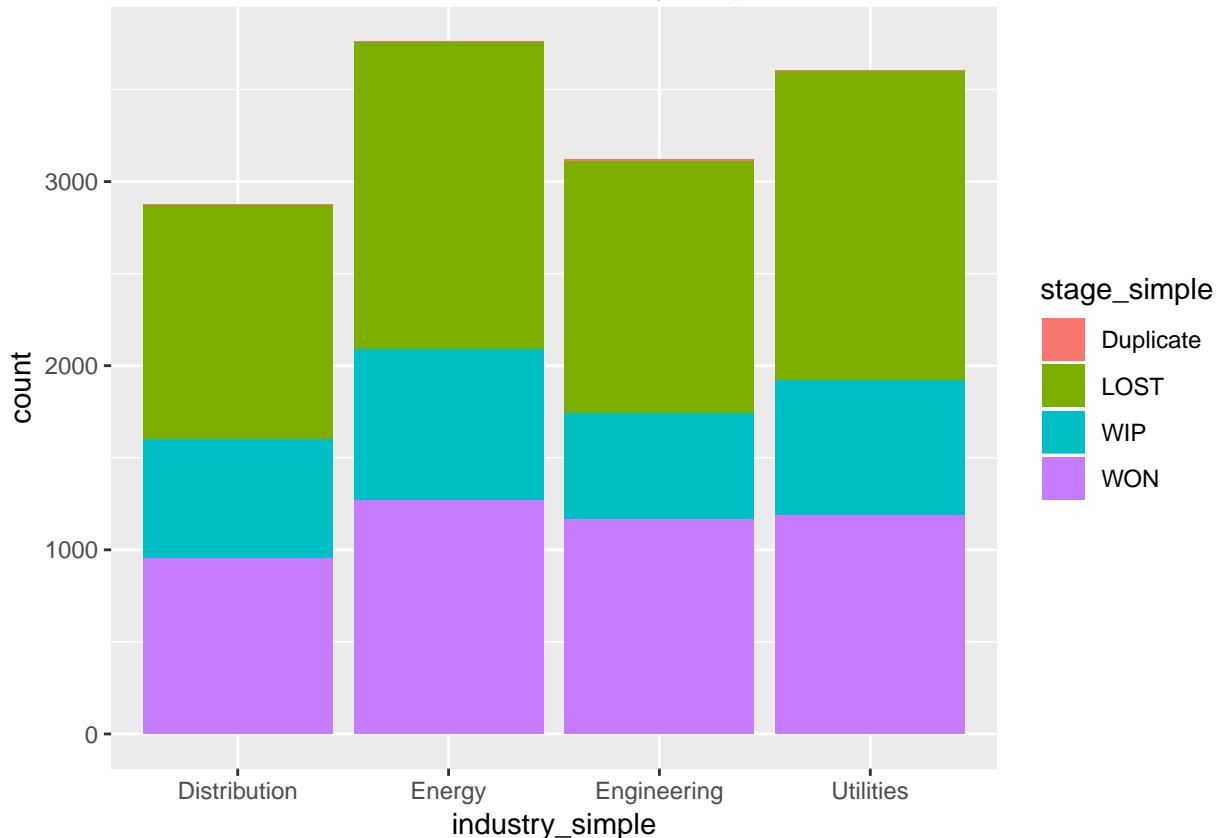
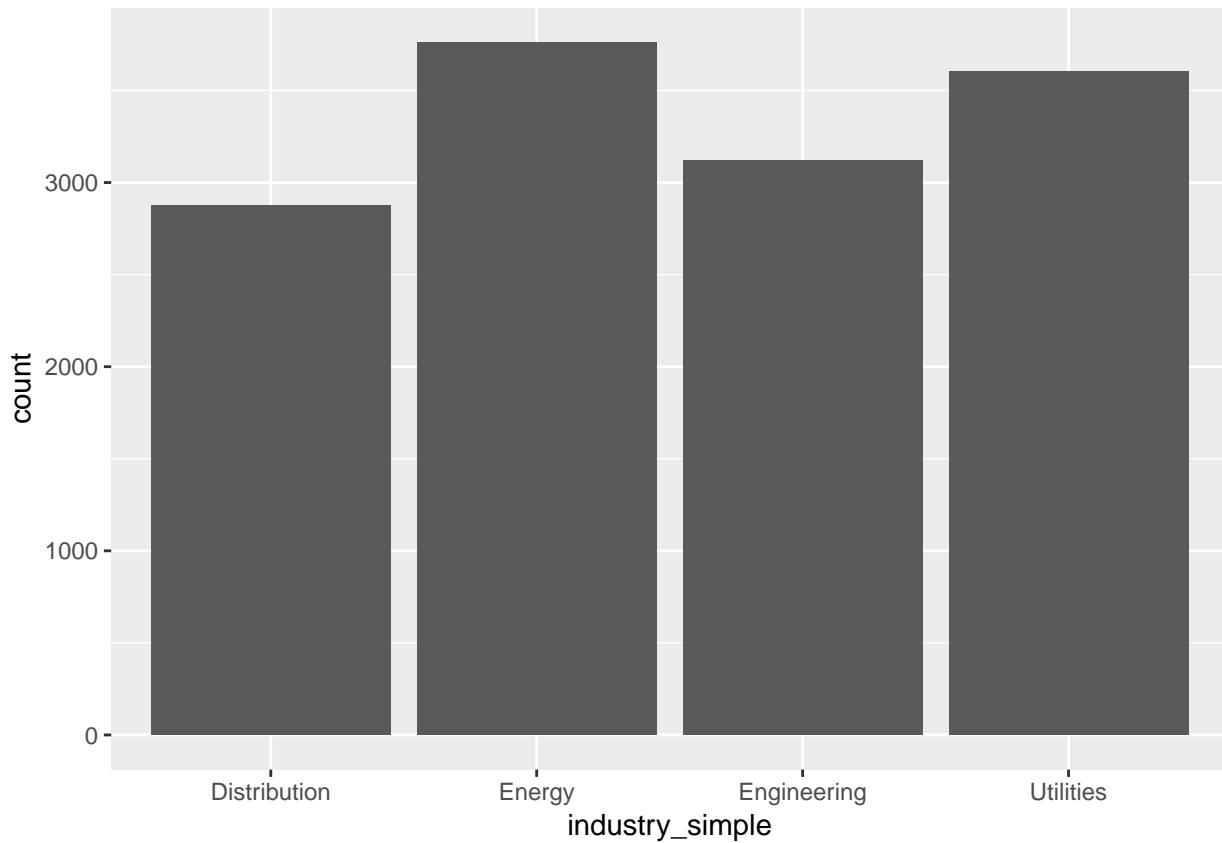
Update Industry Simple with Energy

It also appears we missed something in our additional cleaning operation: Where is the “Energy” sector? Turns out, we lumped it with “Everything Else” incorrectly. So we’ll perform an updated operation and refresh our graphs.



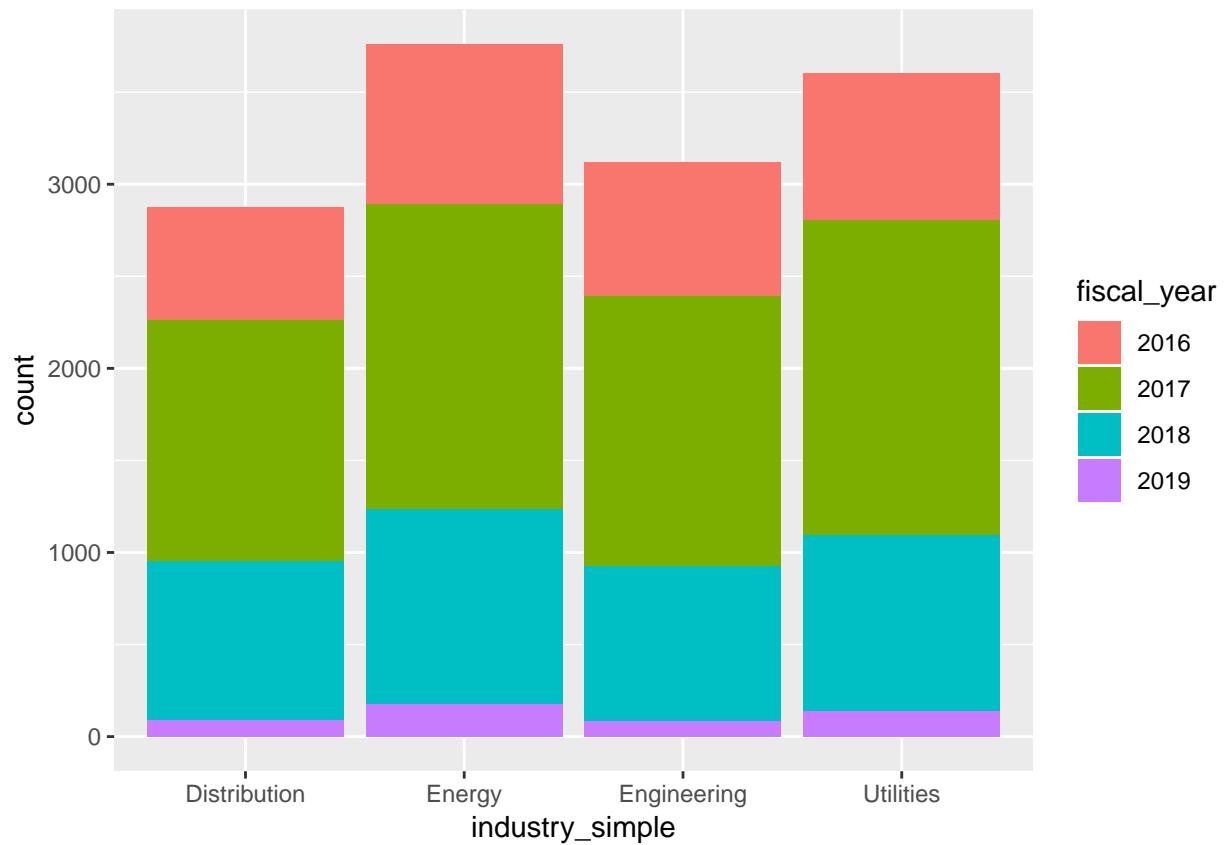
Now let's filter out "Everything Else" and concentrate on the Industries we selected: "Energy", "Utilities",

“Engineering”, “Distribution”

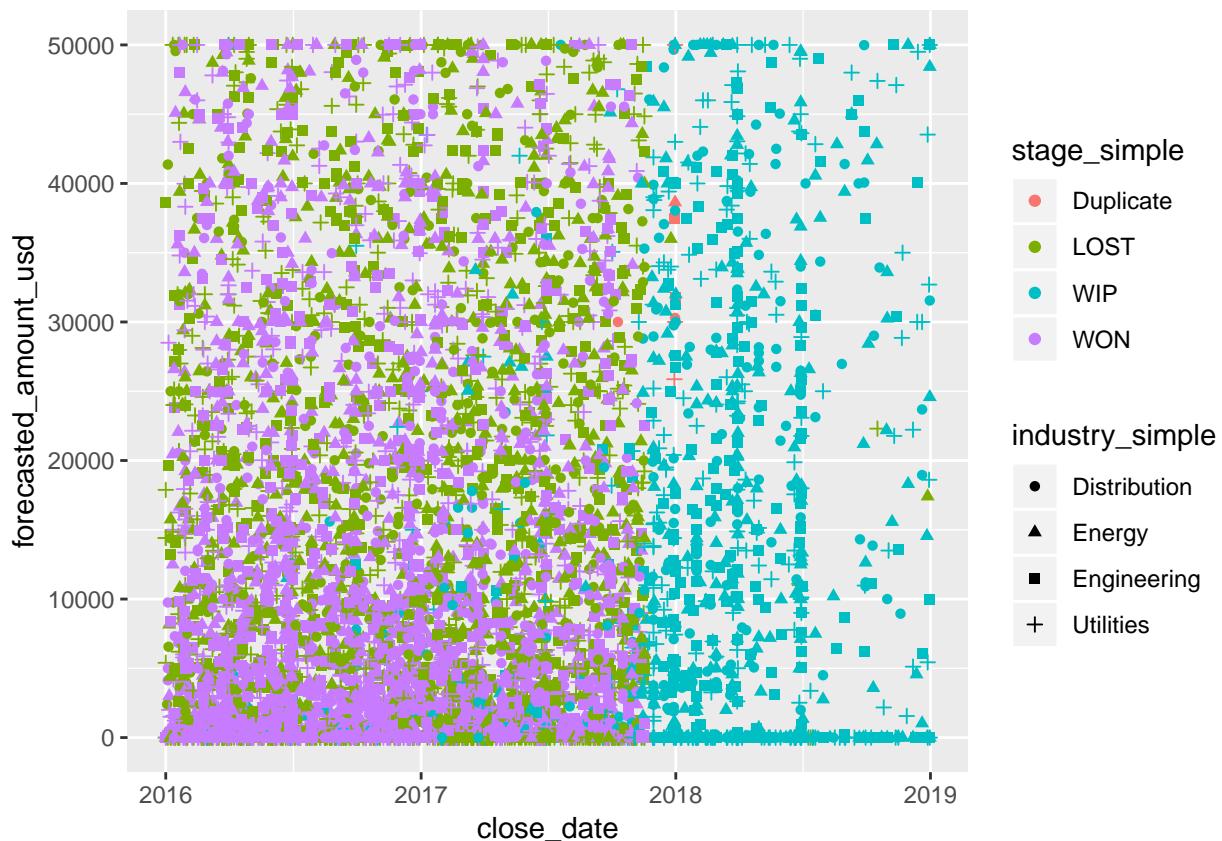


Energy (which comprises Oil, Gas, Fossil and Mineral industries) and Utilities are the sectors with the most deals.

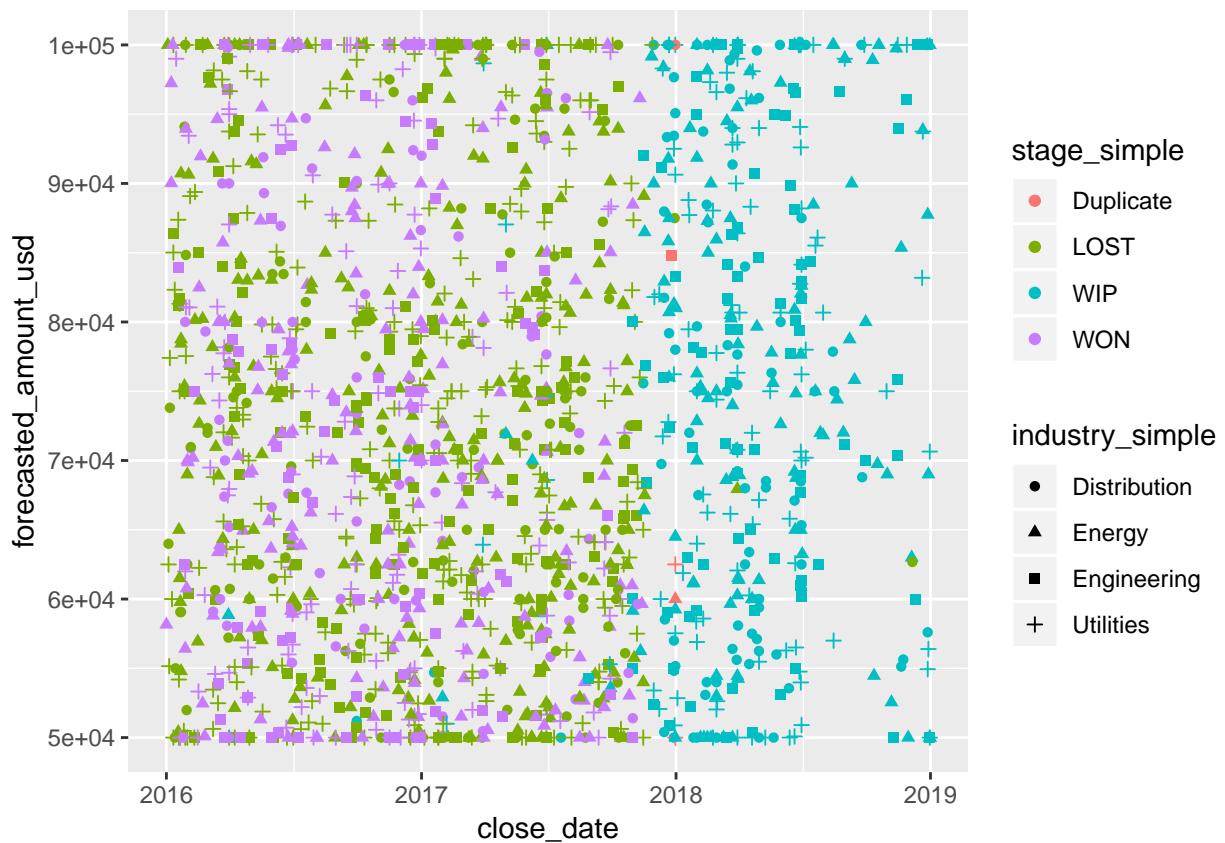
What if we break down by year?



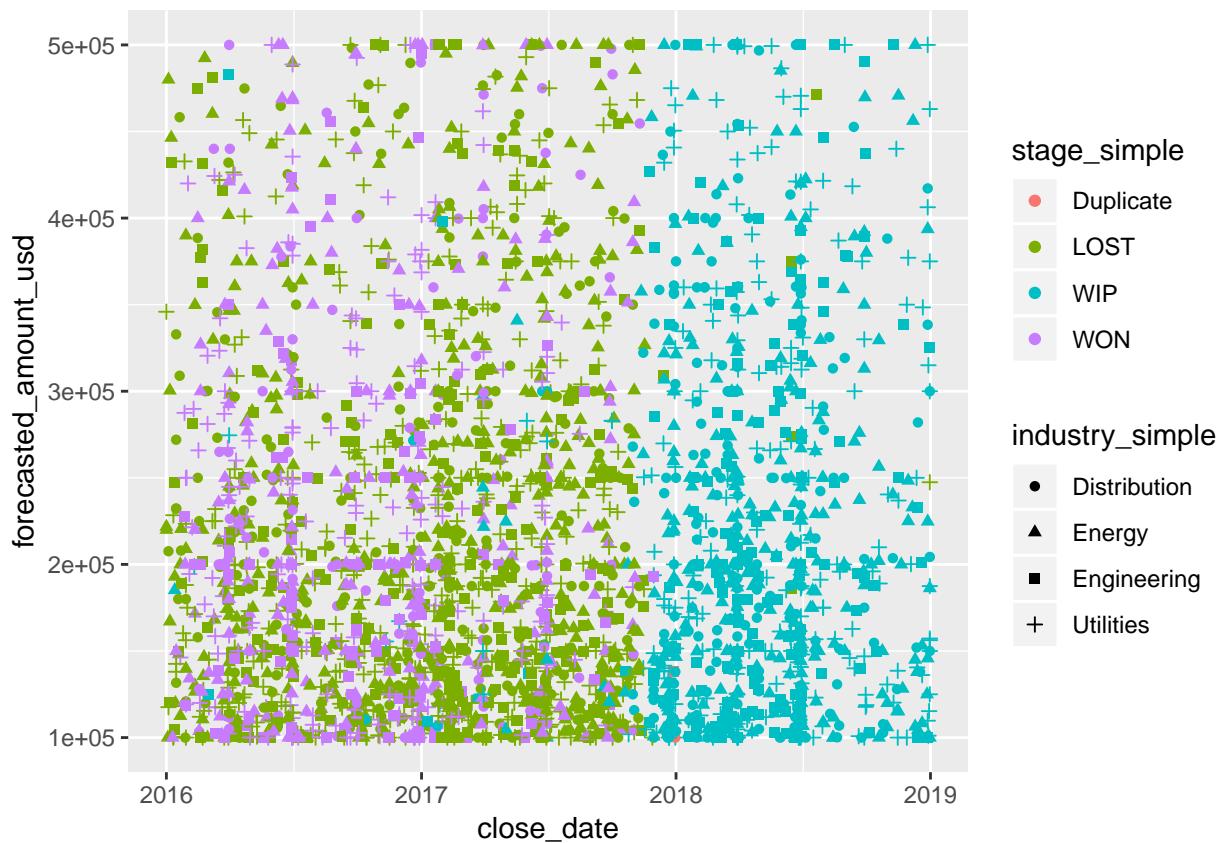
Size of Opportunities: Up to \$50K



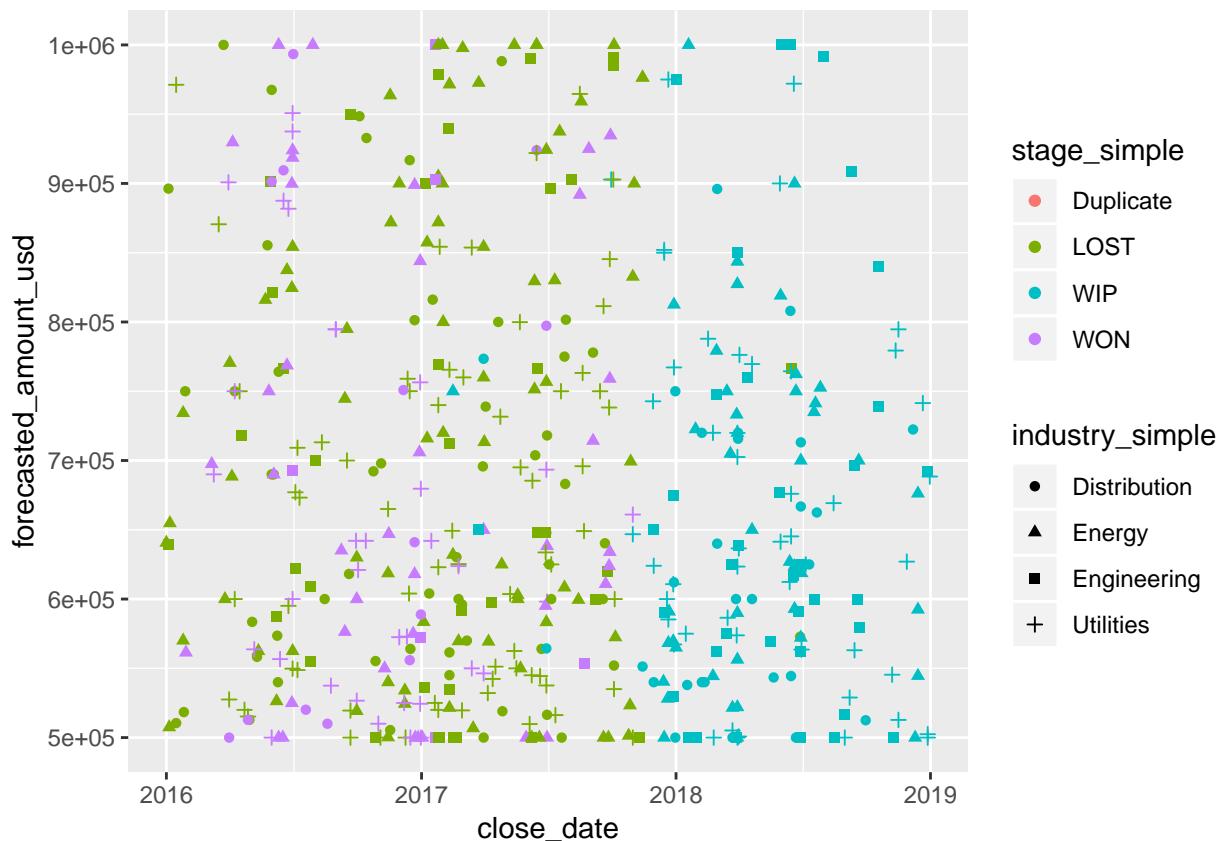
Size of Opportunities: \$50K to \$100K

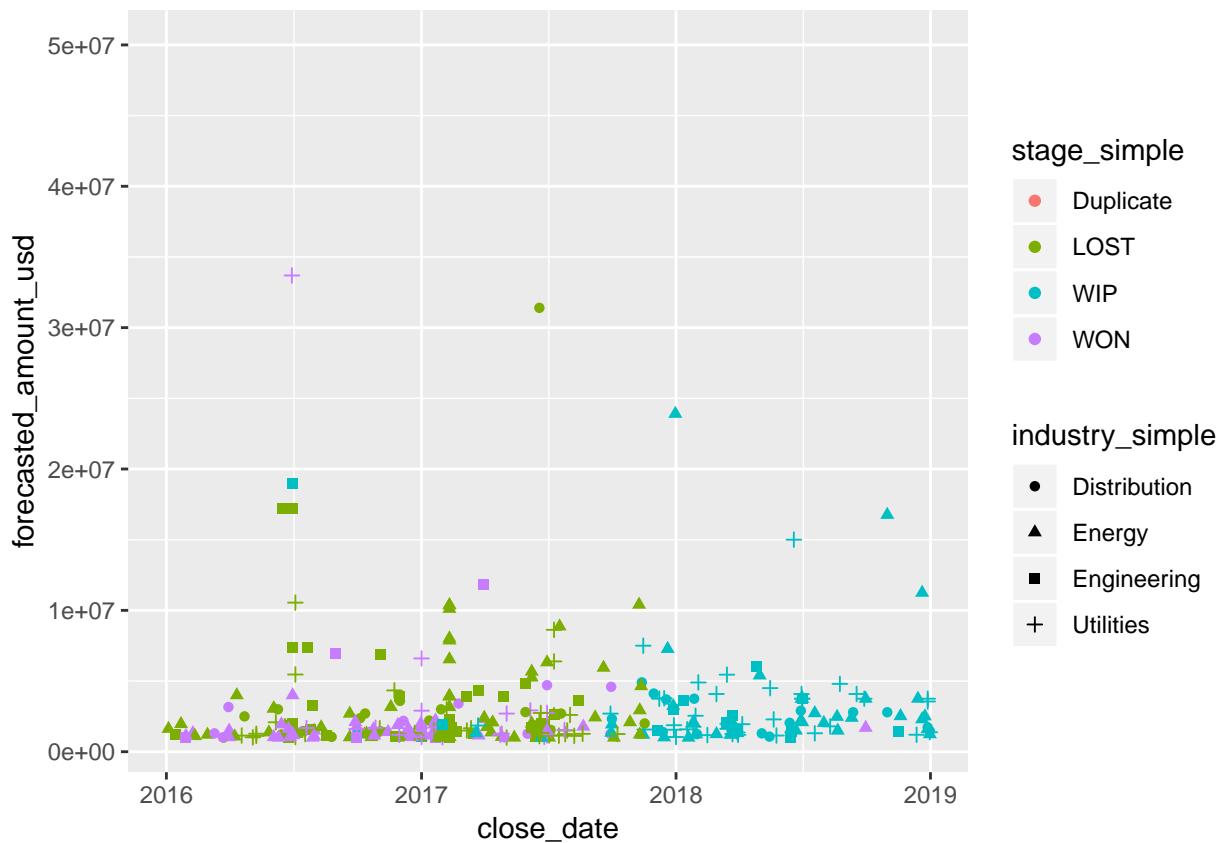


Size of Opportunities: \$100K to \$500K

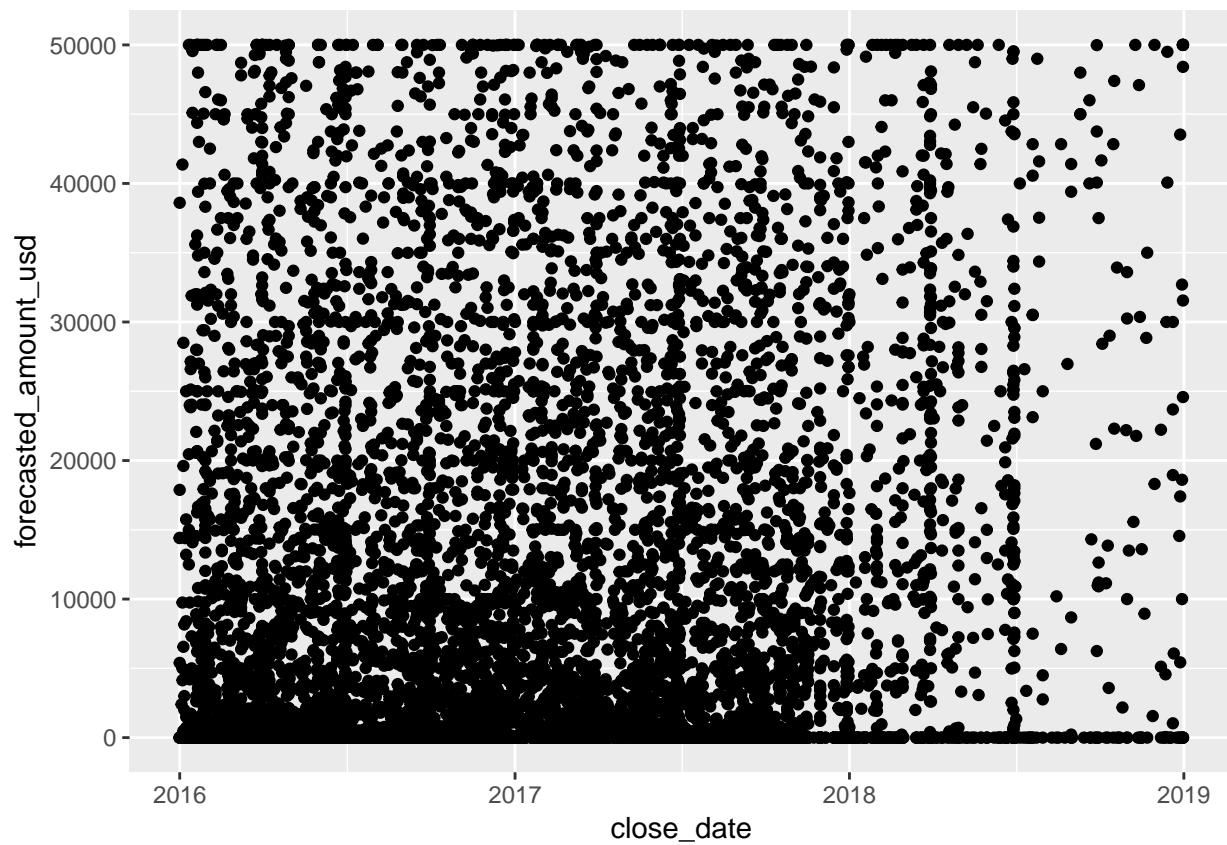


Size of Opportunities: \$500K to \$1M

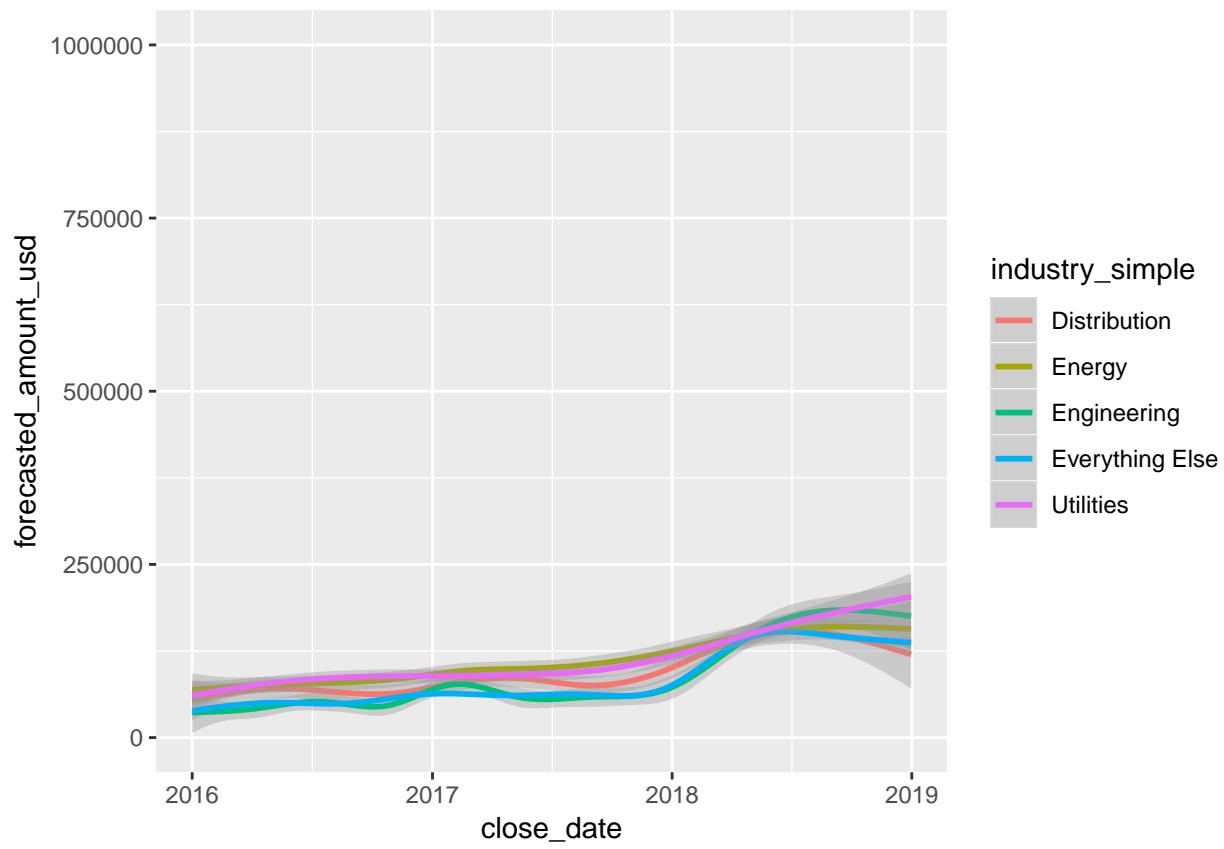




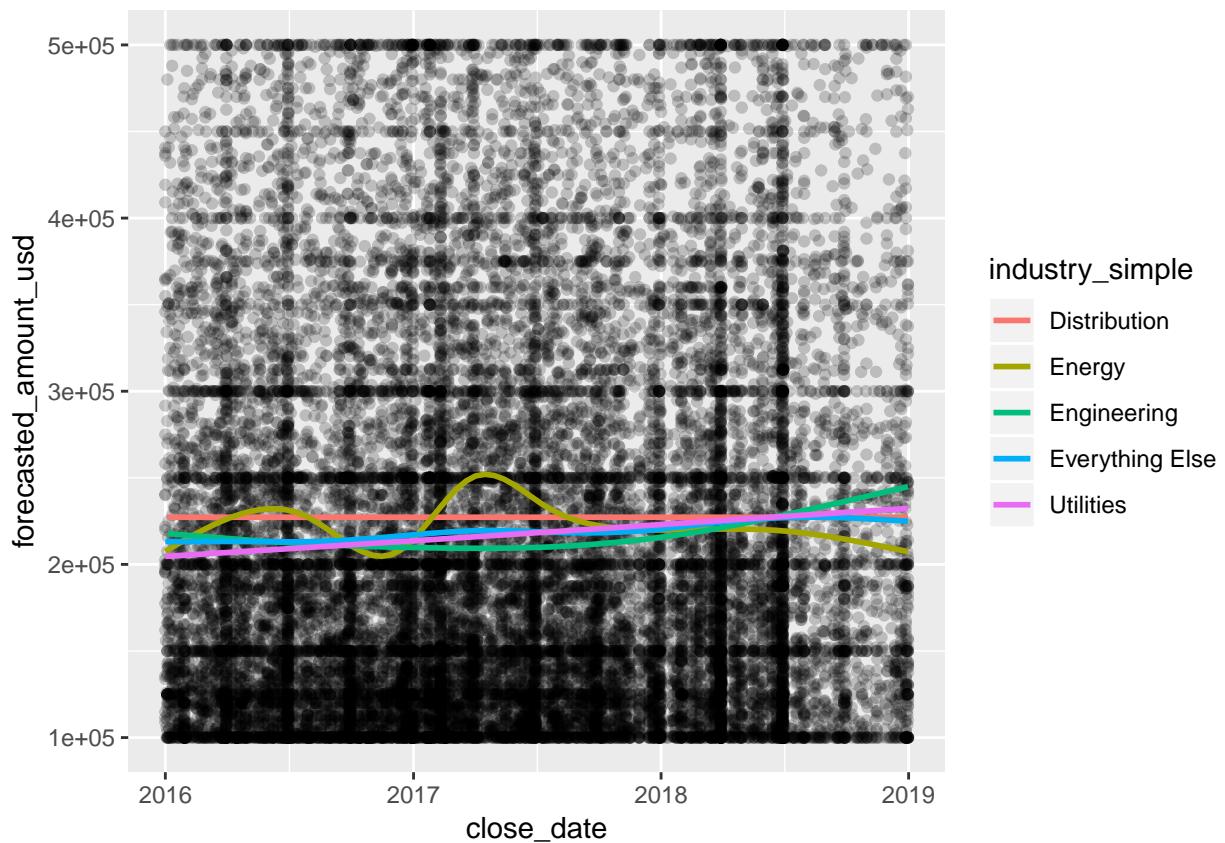
WON Opportunities: Up to \$50K Try to smooth and see if pattern emerges Take out the colors and shaping, and try to get into a smooth line



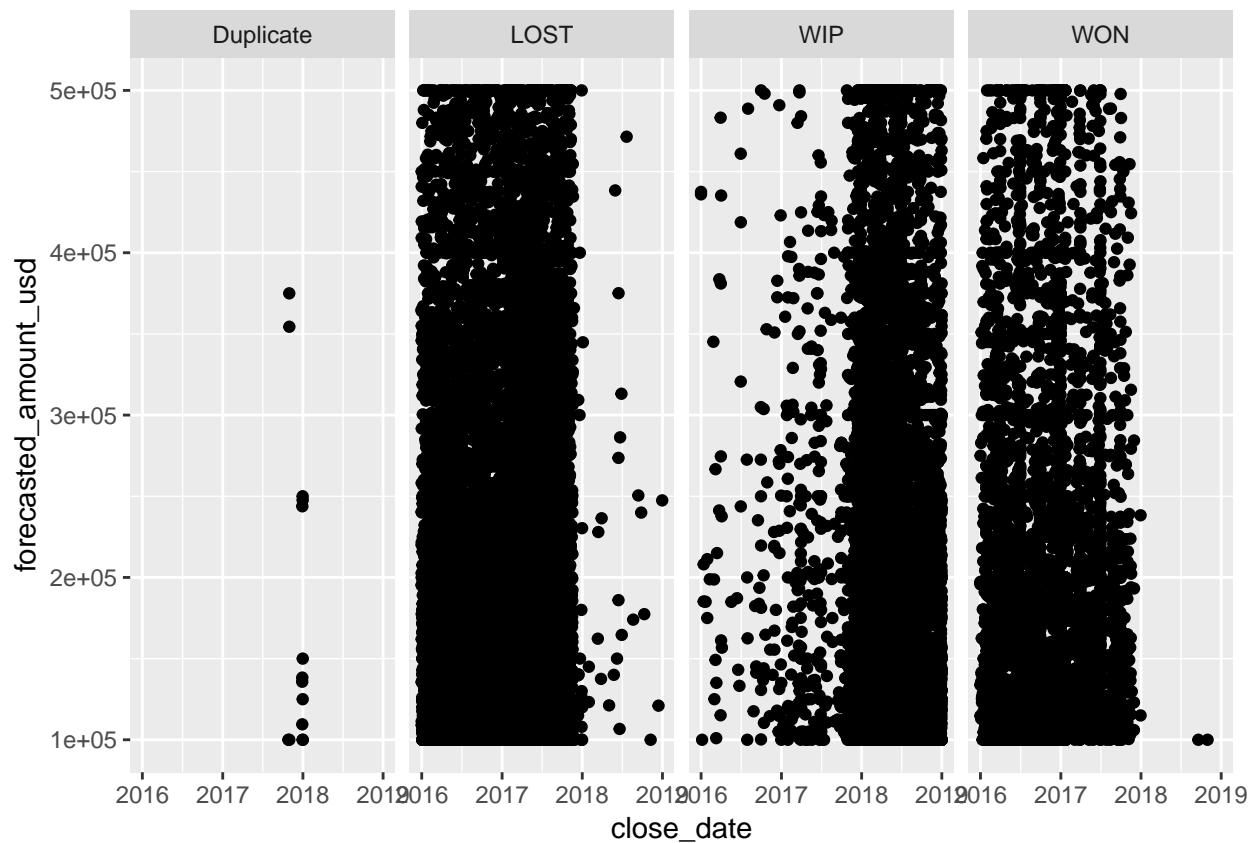
WON Opportunities: Up to \$1M Smoothing work - now remove the points and just see the line, colored by industry An interesting pattern emerges: 1. Utilities dollar amounts continuously increase over time 2. Distribution and Engineering dollar amounts have gone down



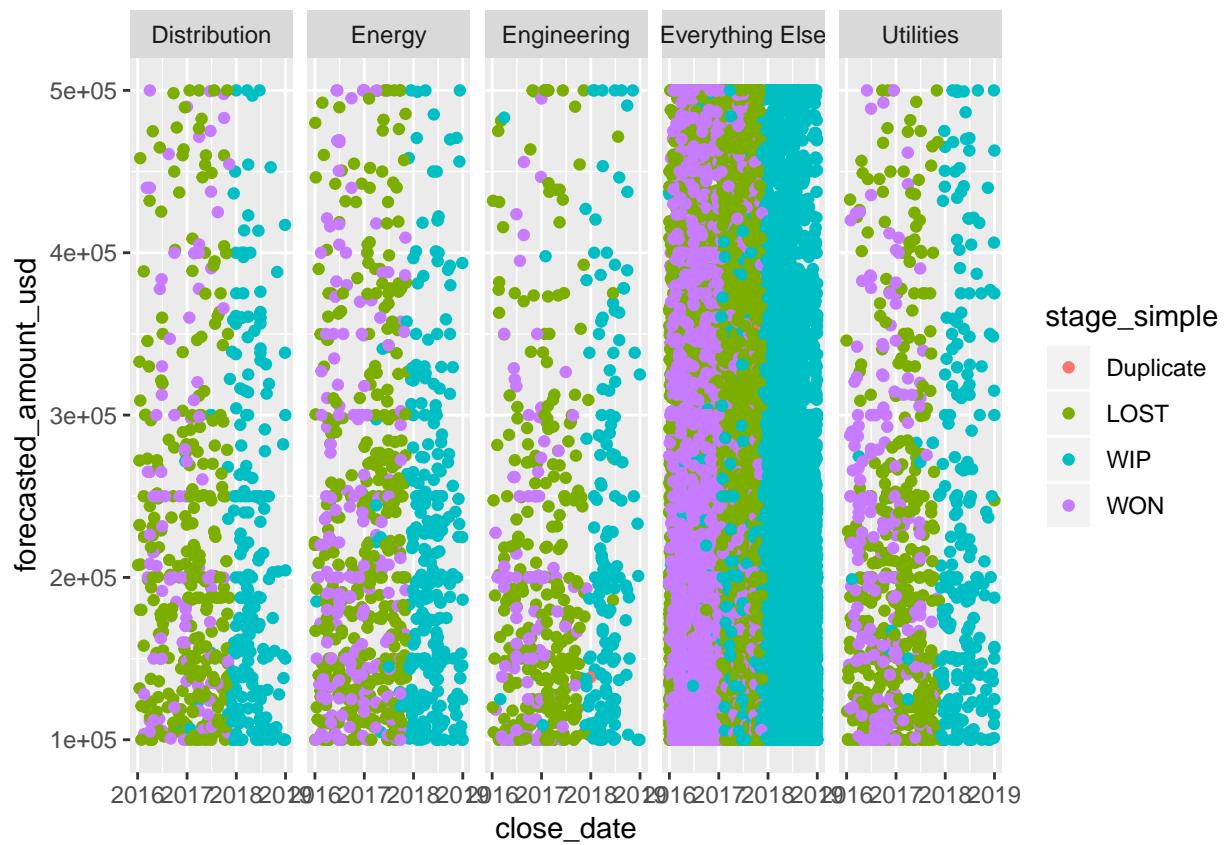
HELP!!! How can get the lines to display WON, LOSS, WIP, as opposed to industry? When we change the following to stage_simple, no lines appear



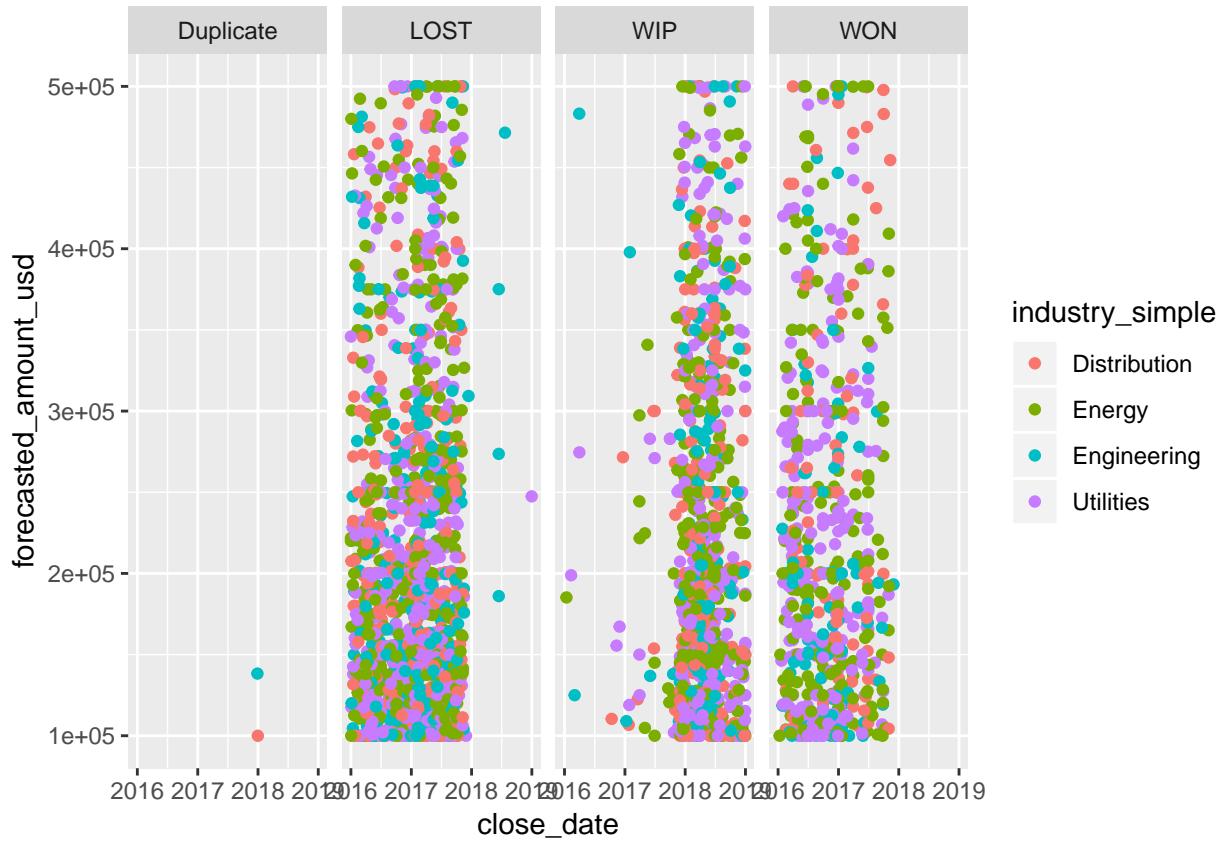
Try another track: Use facet_grid to compare across Stages



BETTER! Now use facet_grid to compare across Industry



BETTER! Move Industry to color, and Stage to the facet



Grouping and Summarizing Part 1

In order to group and summarize the dataset more effectively with `tidyverse`, we need to convert some of the data (such as Close Date) into Discrete data variables (such as Month and Year).

Grouping and Summarizing Part 2

Now that we have date variables that can be grouped, we'll group some sets of data, and plot them.

```
## # A tibble: 27 x 6
##   industry                      sumWon meanWon medianWon maxWon minWon
##   <chr>                         <dbl>   <dbl>     <dbl>   <dbl>   <dbl>
## 1 Aerospace & Defense          3.18e7  58423.    3532.  3.40e6 -2.53e4
## 2 Automotive                     4.90e7  66110.    10000  2.88e6 -4.99e3
## 3 Computer Hardware              2.96e7  30935.    6221.  2.23e6 -1.78e4
## 4 Computer Software              6.81e8  134966.   1728.  5.22e8 -3.21e6
## 5 Consumer Packaged Goods       4.51e7  22560.    1725.  4.87e6 -3.50e4
## 6 Distribution & Transport~     7.52e7  78695.    6864.  4.71e6 -1.71e4
## 7 Education                      3.24e7  32046.    3826.  2.77e6 -1.98e5
## 8 Energy                          1.41e8  111241.   18688. 4.00e6 -3.67e5
## 9 Engineering & Constructi~    5.92e7  50702.    3886.  1.18e7 -1.90e3
## 10 Federal Public Sector         2.45e7  46614.    3798.  2.22e6  0.
## # ... with 17 more rows

## # A tibble: 6 x 6
##   industry_simple      sumWon meanWon medianWon maxWon minWon
##   <chr>                  <dbl>   <dbl>     <dbl>   <dbl>   <dbl>
## 1 Distribution        75232661.  78695.    6864.  4706310 -17126.
## 2 Energy                141387504. 111241.   18688. 4000188 -367000
## 3 Engineering         59169236.  50702.    3886. 11800000 -1895
```

```

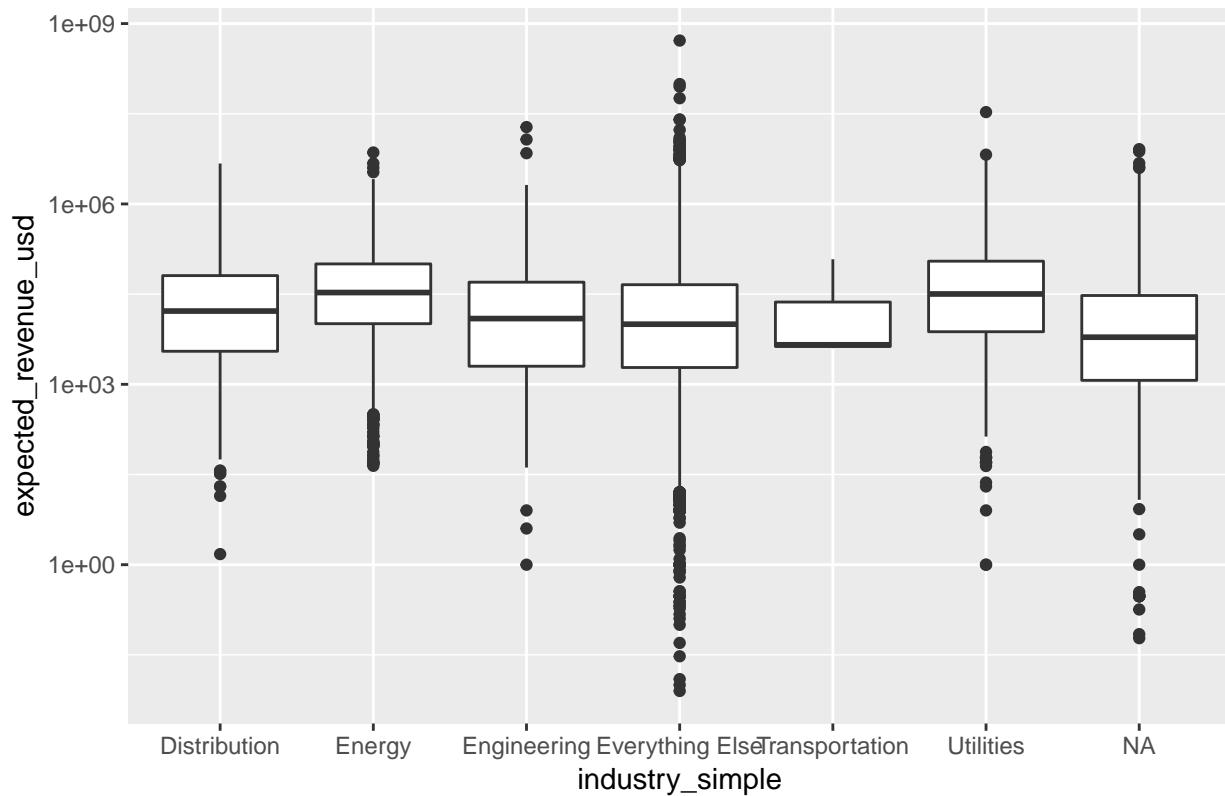
## 4 Everything Else 2802057077. 64063.      3832. 521876775 -3214097
## 5 Utilities       159260898. 133833.     14845. 33687033. -420000
## 6 <NA>            223138202. 39285.      2492    4085822. -1141460.

```

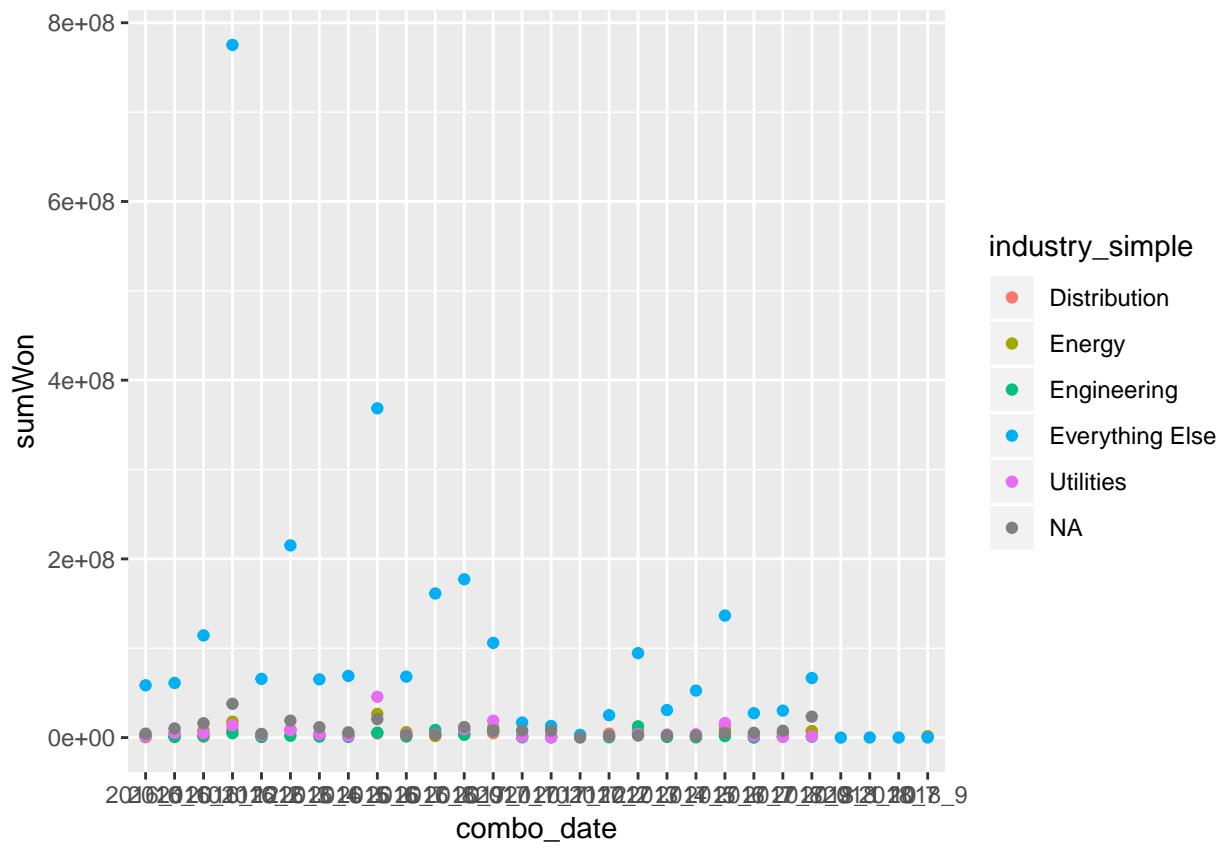
The following chart appears to show the following for each Industry: - black line is the MEDIAN - top is 75th percentile - bottom is 25th percentile - the “whiskers” represent the additional range of values - the dots represent the outliers

The graph seems to show that the Energy industry ranges are in alignment with the median and distributions for other industries.

Expected Revenue By Industry

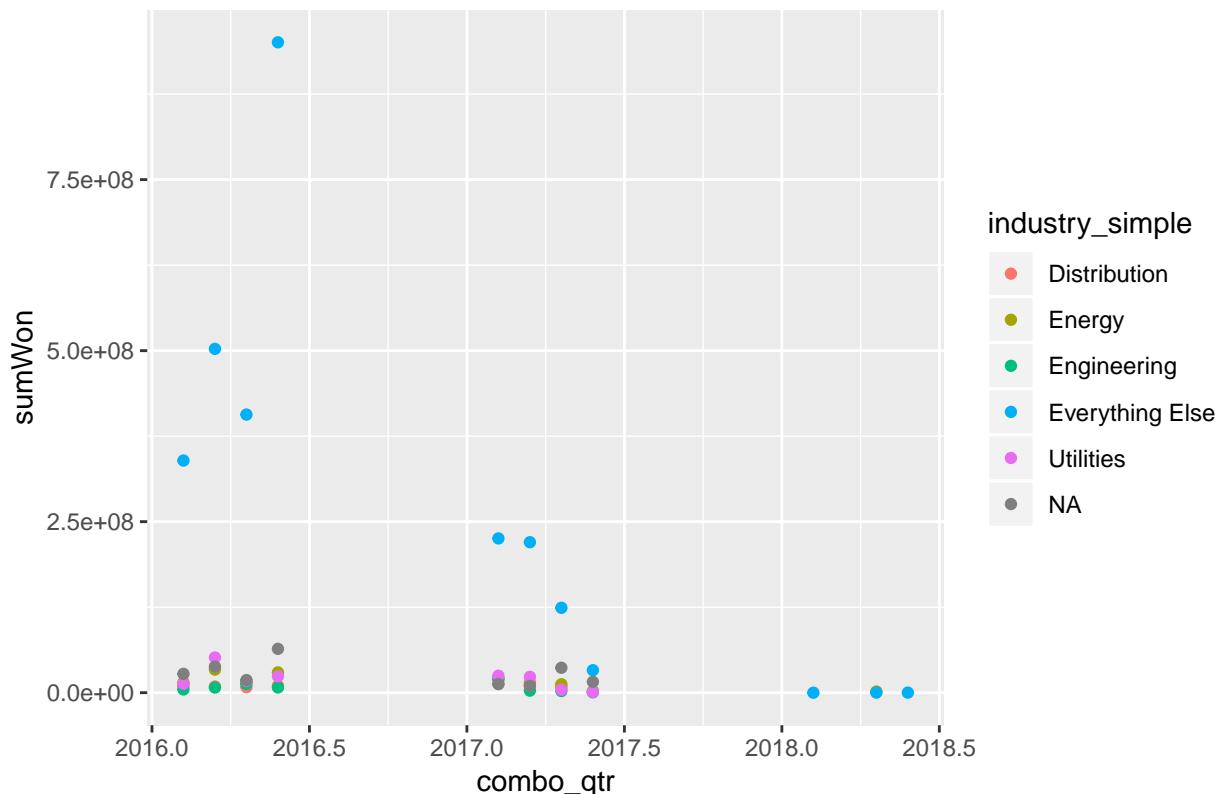


Now let's try grouping by Years and Months

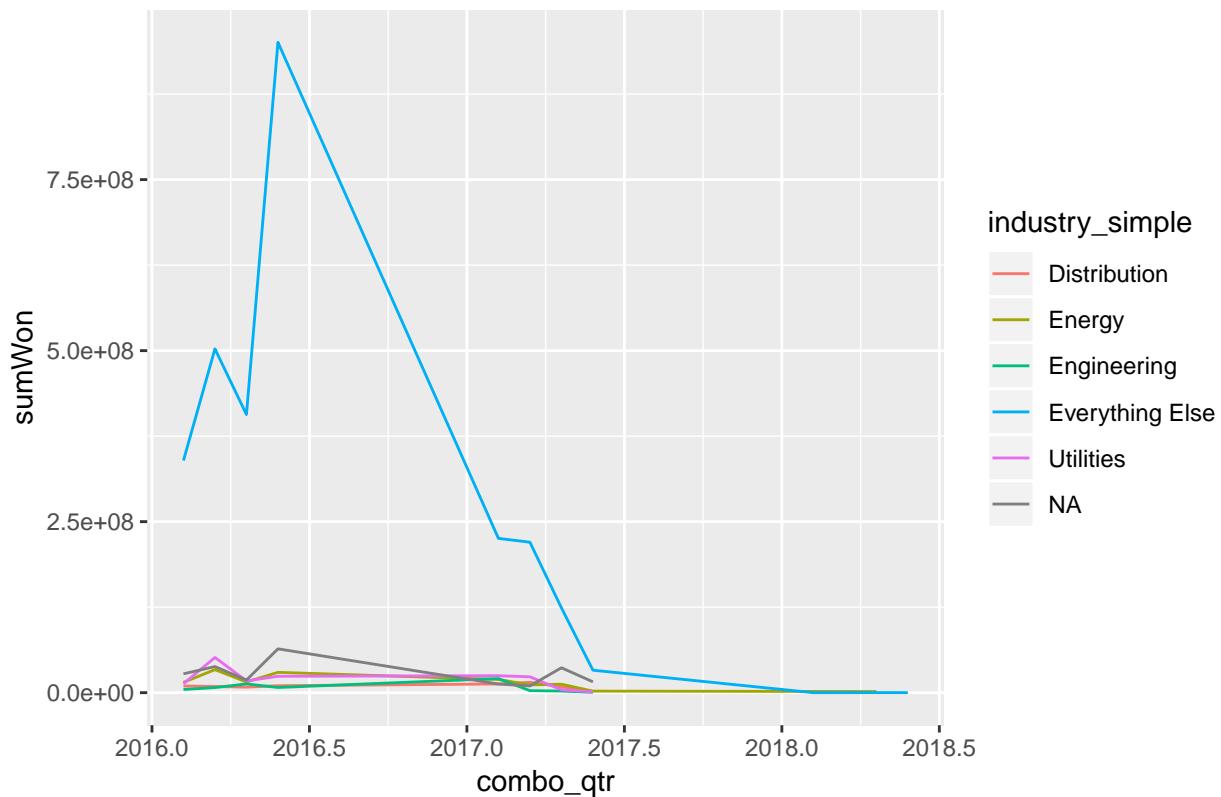


Still too crowded, unless we get a larger screen or layout. Let's try adding a new column for Year + Fiscal Quarter.

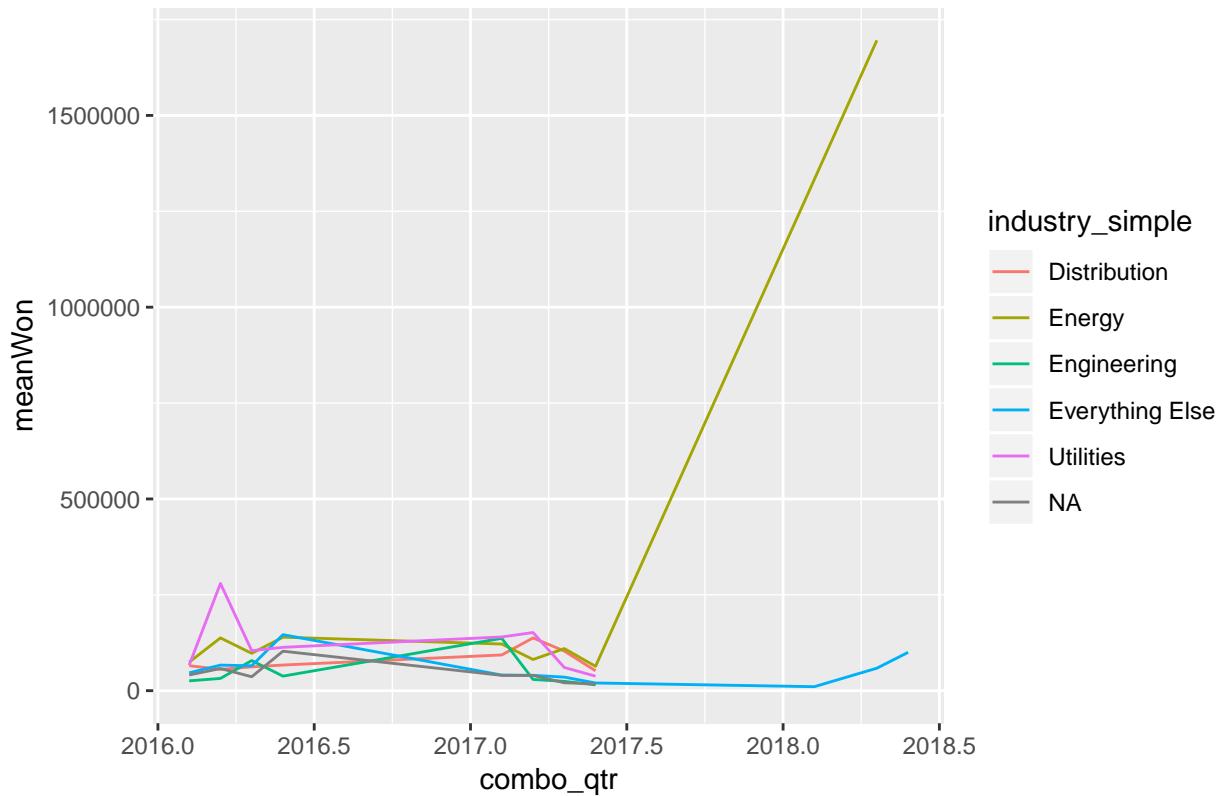
Sum Amounts of WON by Fiscal Quarter 2016–2018



Sum Amounts of WON by Fiscal Quarter 2016–2018



Mean Amounts of WON by Fiscal Quarter 2016–2018



Median Amounts of WON by Fiscal Quarter 2016–2018

