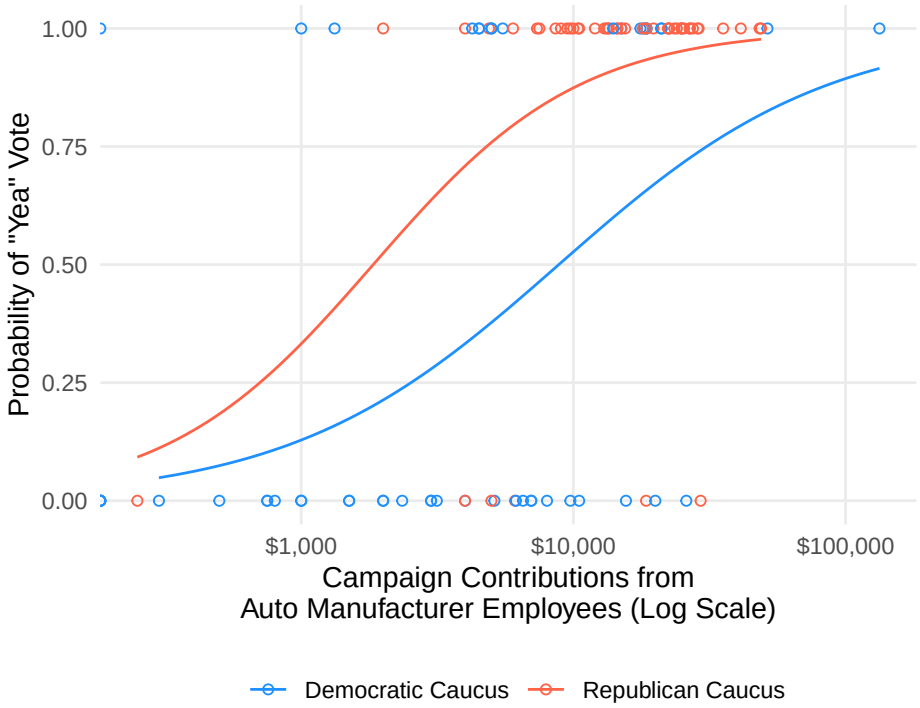
Skill Task 3
Graphics

PS 811: Statistical Computing

Due February 28, 2020

The `CAFE` dataset describes how Senators voted for a fuel efficiency standards bill (The "Corporate Average Fuel Economy" Bill) and their campaign contributions from individuals who work for auto manufacturers (to date). Use this dataset to visualize the relationship between contributions (x) and the Senators' votes (y). You can emulate the graphic in this document (I discuss helpful tips below), or you can put your own spin on how you want to visualize the information. Ensure that your graphic looks professional. For example, use informative labels instead of the *blah* variable names from the dataset.



Tips

Recoding variables: As we saw from Skills Task 2, this dataset contains Republican, Democratic, and Independent senators. In practice, Independents tend to "caucus" with one of the two major parties, meaning they coordinate activities and vote together. In this dataset, the only Independent senator (Jim Jeffords) caucused with the Democrats. Instead of using the `Party_Code` variable in the data, create a `Caucus` variable that codes Independents as Democrats.

Create a binary (0/1) `Yea_Vote` variable as you did for Skills Task 2. This will serve as the Y variable.

Geoms: In addition to points, I plotted a line using `geom_smooth()`. When the *y* variable is binary, linear models tend to behave weirdly, so I used a nonlinear *logistic* regression line. You can create this yourself by copying the code below and adding it to your plot.

```
geom_smooth(
  method = "glm",
  method.args = list(family = binomial(link = "logit"))
)
# glm means "generalized linear model"
# `family` and `link` set the probability model/link function
```

Aesthetics: You will notice that I mapped `color` according to the `Caucus` variable, and then I manually set the colors. Note also that I transformed the *x* axis to the log scale (`scale_x_log10()`). This isn't necessary, but it can sometimes be helpful for displaying dollar values, which often have large outliers. I modified the tick labels using `labels = scales::dollar_format()`, which formats tick labels with a `$` and commas.

Other possibilities: Consider the following.

I played with plotting the two party caucuses in different facet panels, but eventually didn't. I also experimented with different point shapes, "jittering" the point positions, keeping the confidence intervals from `geom_smooth()` (I eventually didn't keep them), a title and subtitle for the plot, and so on.

Experiment with the things we learned in class/in the notes to make your plot look nice!