Gerrymandering: Exploring the Data

An, Anderson, and Deck

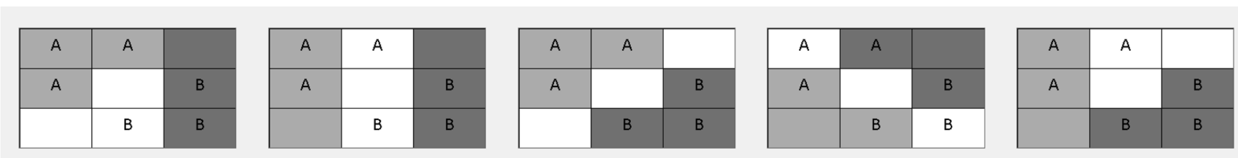
4/4/2021

We are interested in a more reasonable naming device to provide a better way to think about each map.

For the gerrymandered maps we refer to them as $Gerry_i$ for $i \in \{A, B\}$ where i identifies the player for whom the map is gerrymandered (Player A is advantaged in $Gerry_A$). That is, Map 1 will be $Gerry_B$ and Map 5 will be $Gerry_A$.

As the remaining maps are symmetric at the player level we reference $Sym_{d,z}$ for $d \in \{1, 3\}$ and $z \in \{1, 3\}$ where d denotes the number of competitive districts and z denotes the number of zones within each competitive district. That is, Map 2 will be $Sym_{1,1}$, Map 3 will be $Sym_{1,3}$, and Map 4 will be $Sym_{3,1}$.

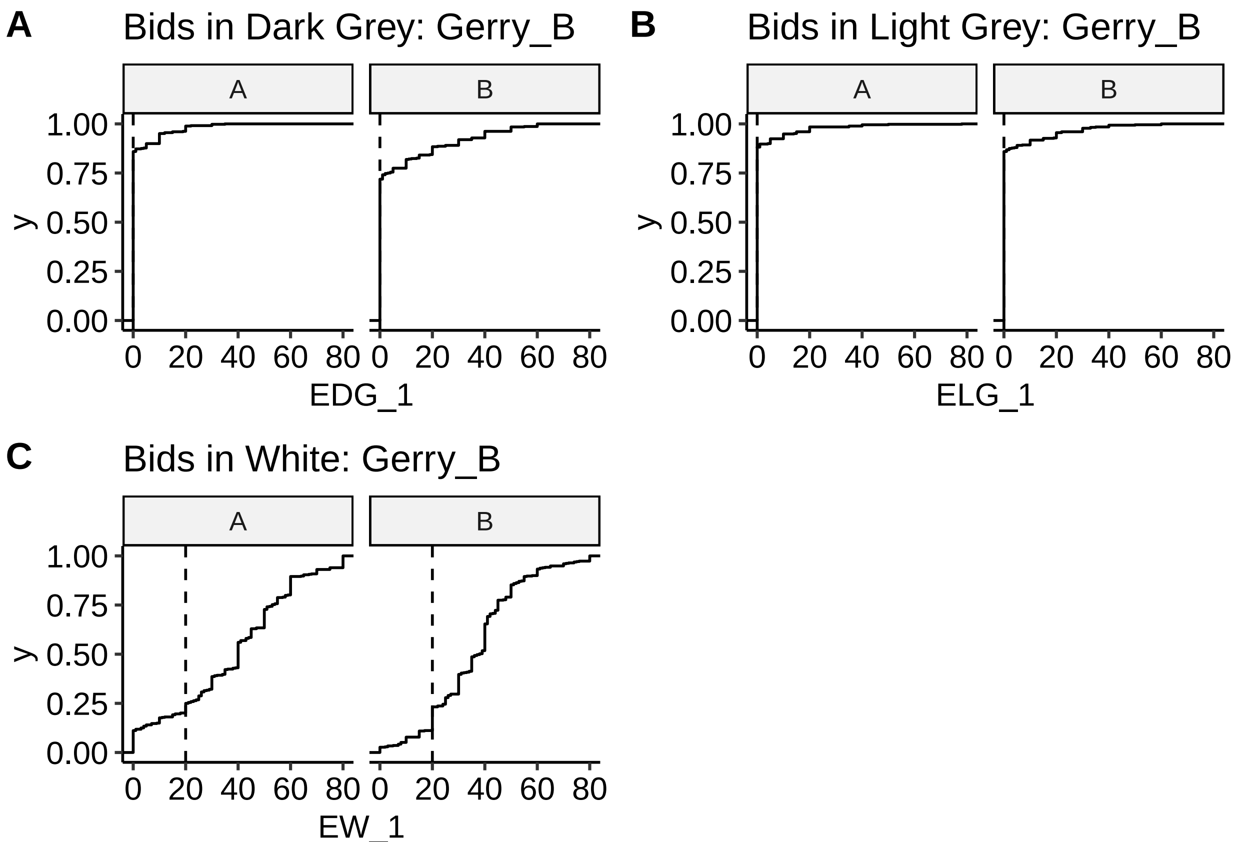
For reference:



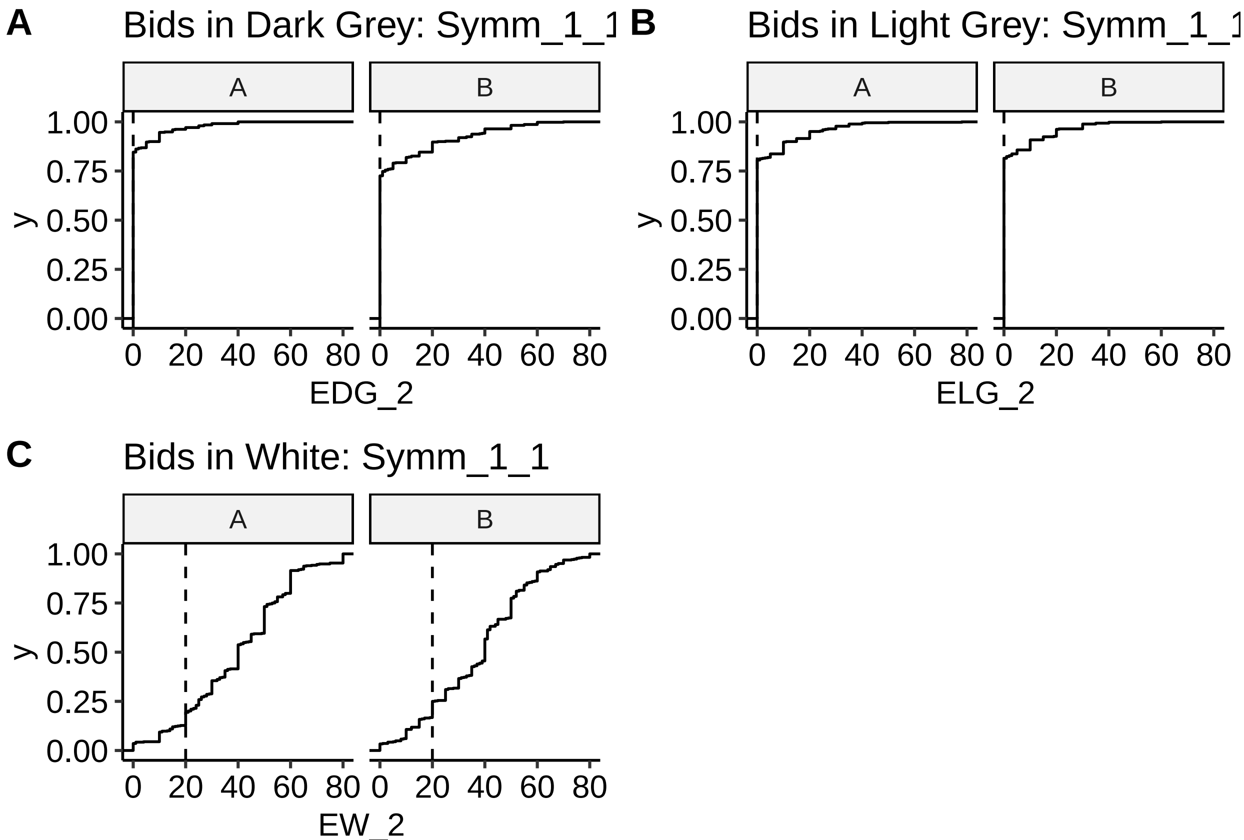
Reading from left to right we have Gerry_B, Symm_1_1, Symm_1_3, Symm_3_1, and Gerry_A.

Note that in Gerry_A, Gerry_B, Symm_1_1, and Symm_1_3 the white district is the only competitive district in the sense that only the competition within the white district determines whether a subject wins that Map. The exception is Symm_3_1 in which it is logical to bid in any district as no district is guaranteed a victor.

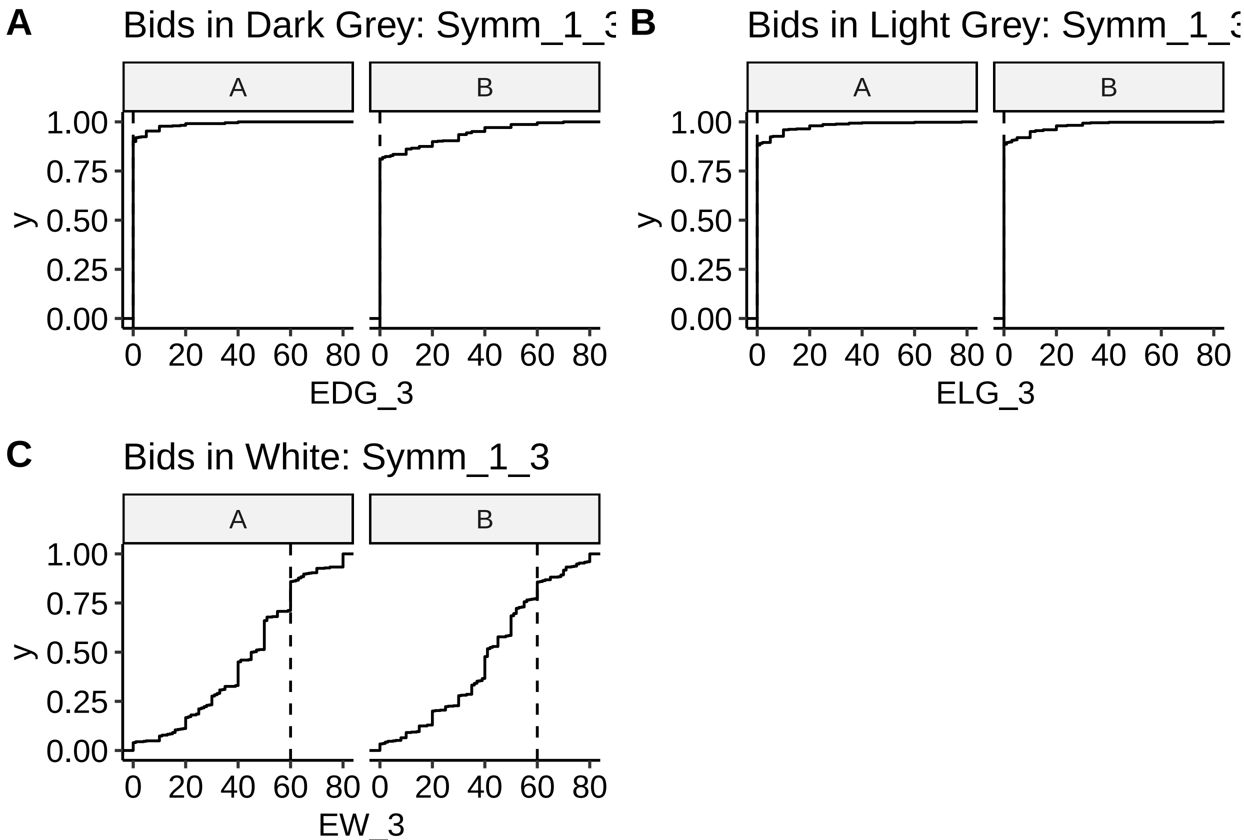
Gerry_B



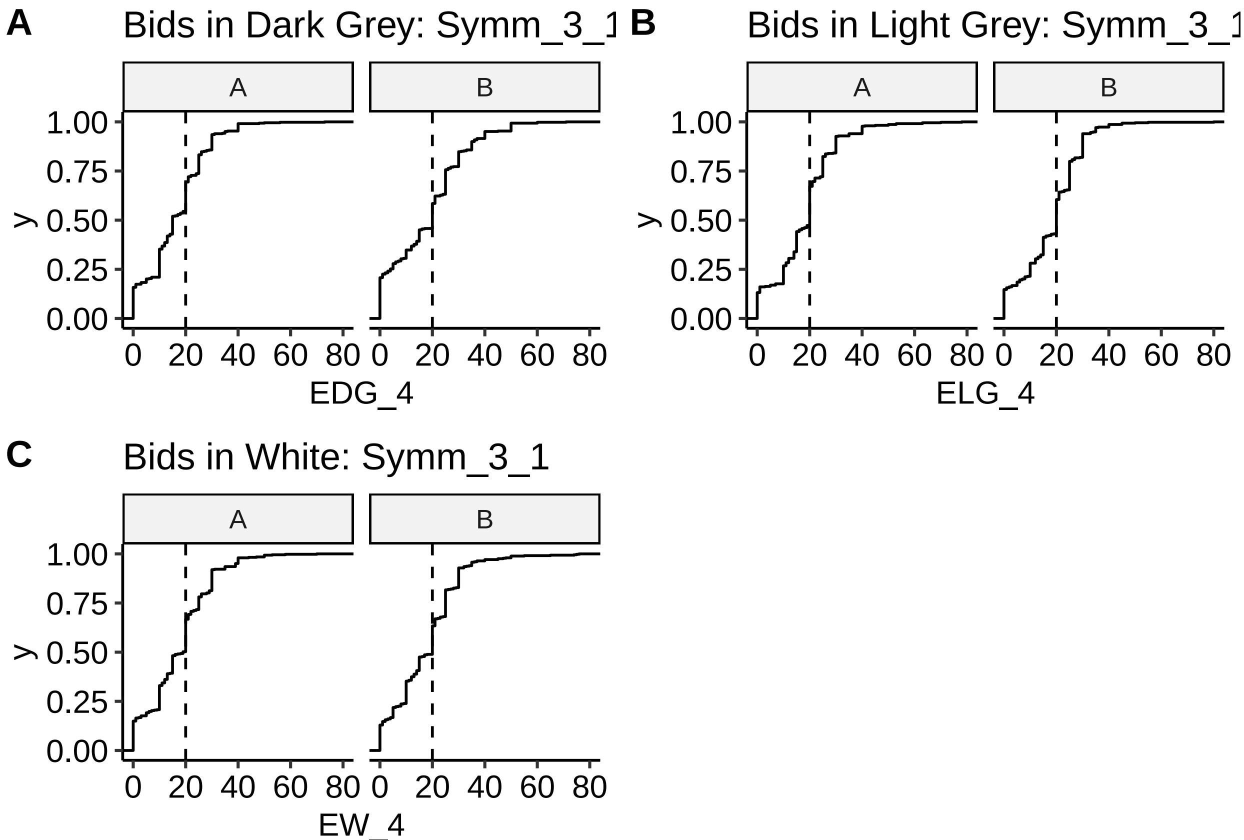
$Symmetric_{1,1}$



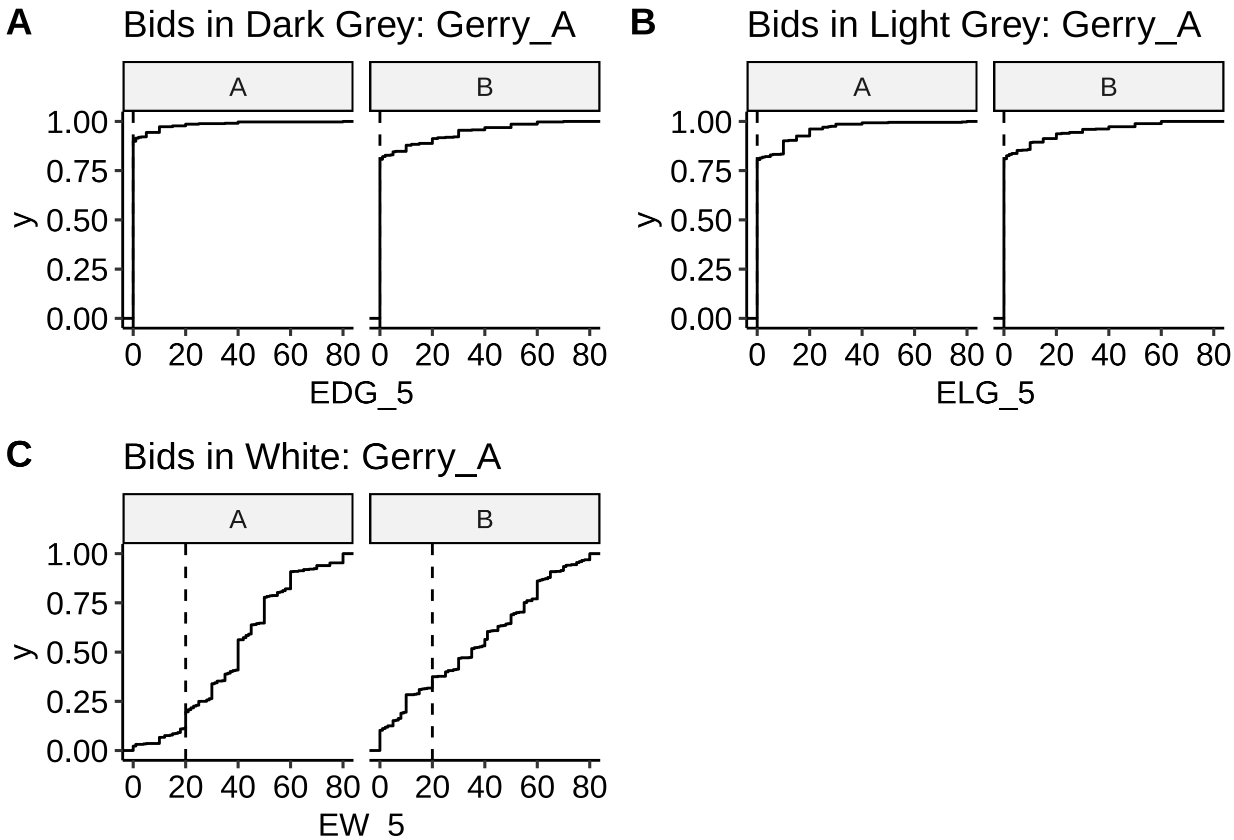
$Symmetric_{1,3}$



$Symmetric_{3,1}$

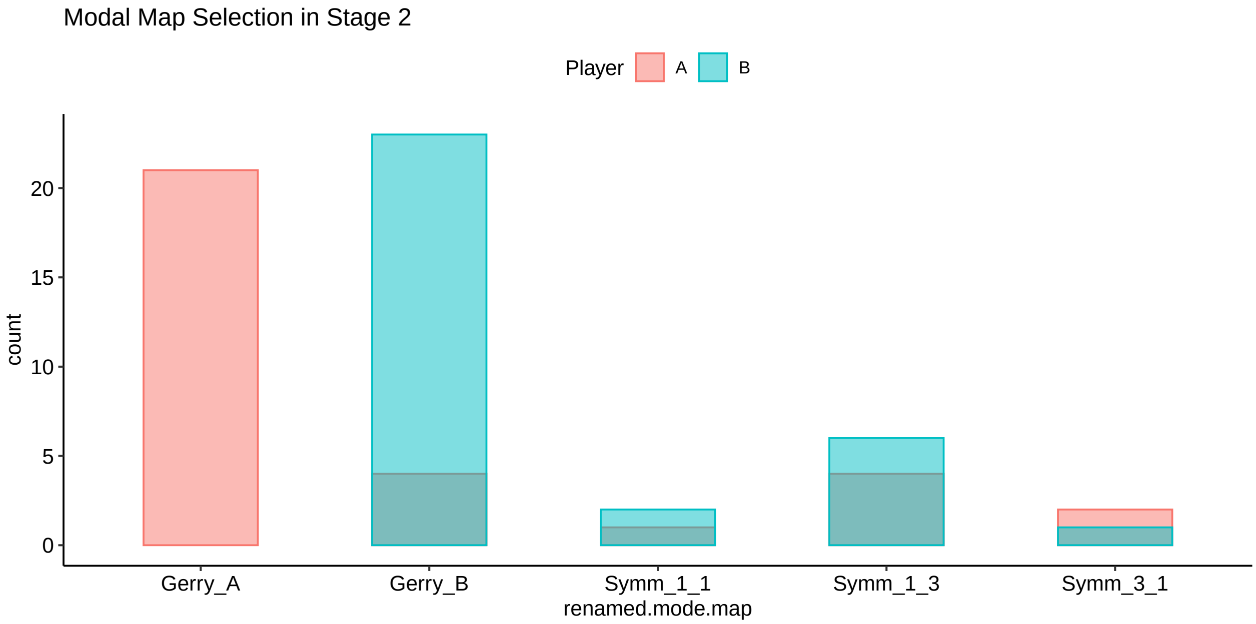


Gerry_A

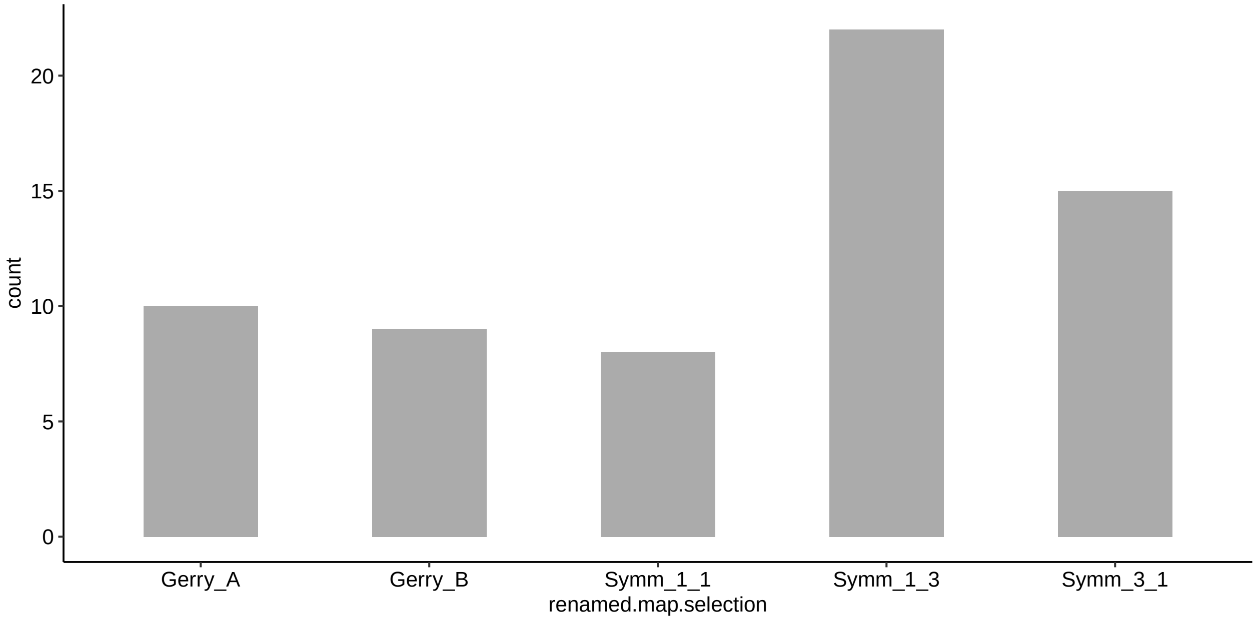


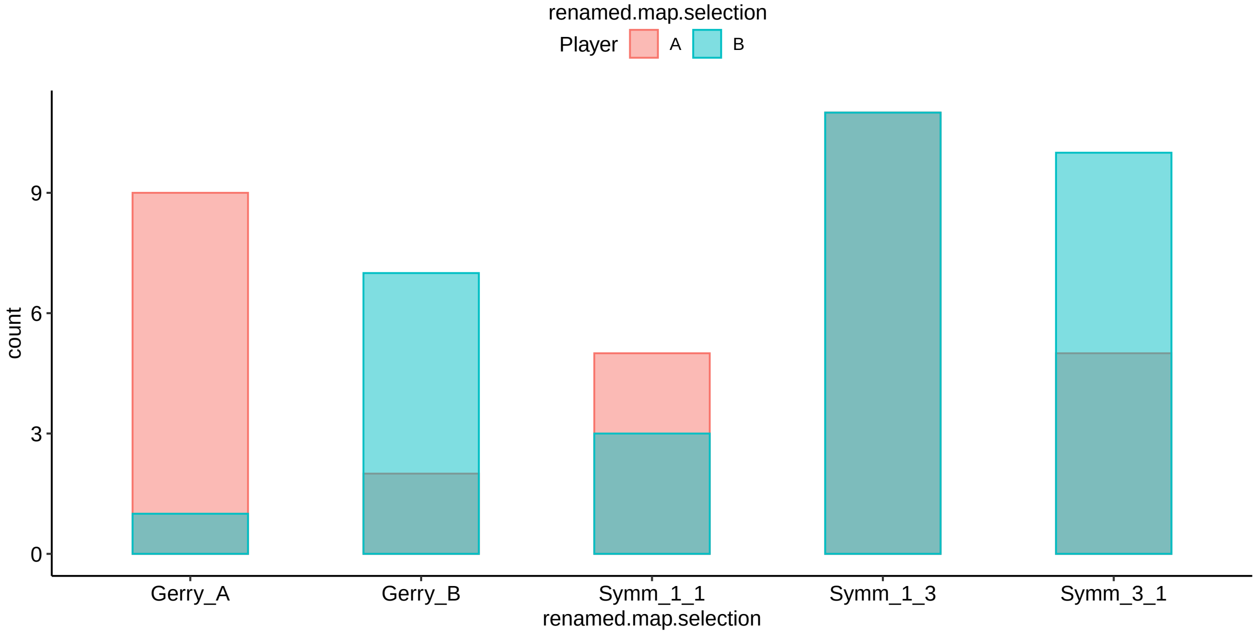
A
Bids in Dark Grey: Gerry_A
A
B
1.00
0.75
0.50
0.25
0.00
y
0
20
40
60
80
EDG_5
B
Bids in Light Grey: Gerry_A
A
B
ELG_5
C
Bids in White: Gerry_A
A
B
EW_5

Recall, Player A should pick Gerry_A and Player B should pick Gerry_B if they are choosing the map that gives them the best chance of winning.



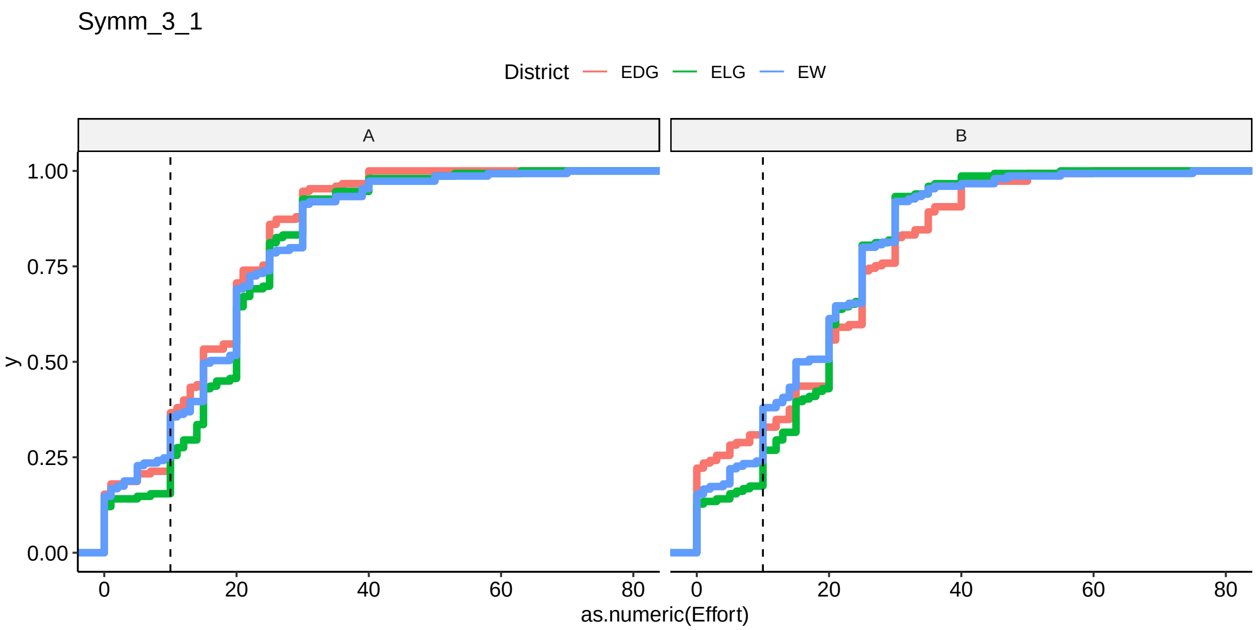
The first figure depicts the map choices during the final stage for all participants. The second figure is of interest because we might have spillover from the previous stage whereby participants choose the map they have been choosing without really paying attention to the implications. . . or they could just be flipping the coin that they are the "incumbent" after randomization occurs.



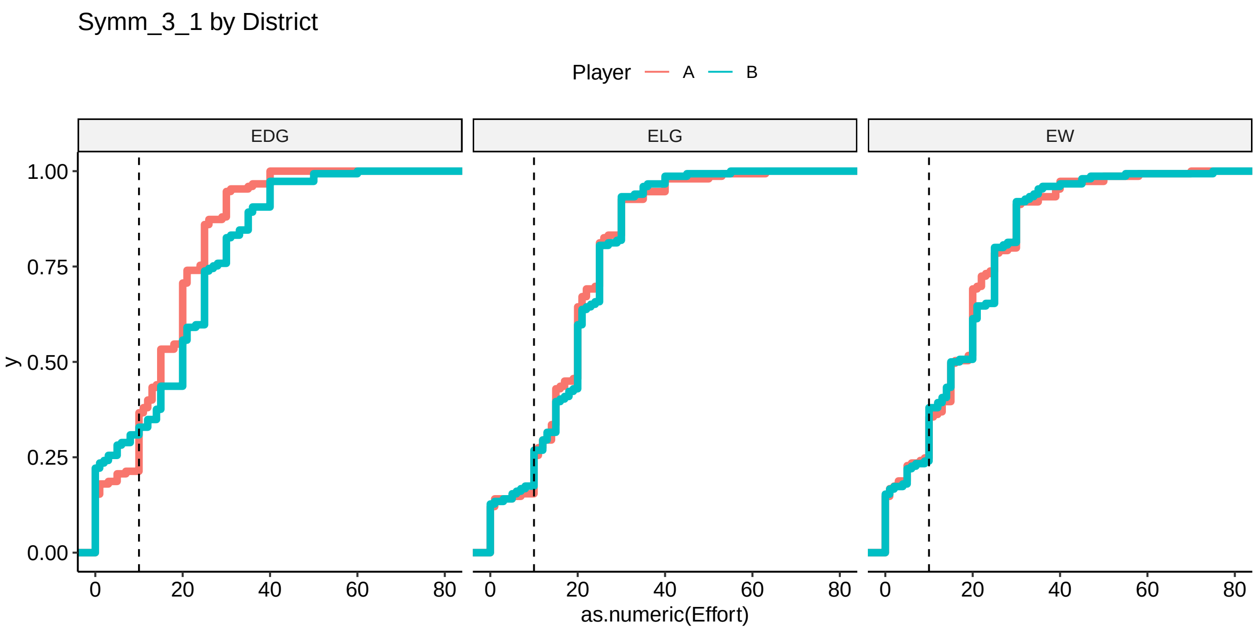


Now we are addressing:

1) For the map where they should be bidding on every region, I would like to see player A's 3 CDFs overlaid on top of each other because there's no reason for them to differ but it's hard to tell in the version you sent.
2) I would also like to see player B's CDFs overlaid

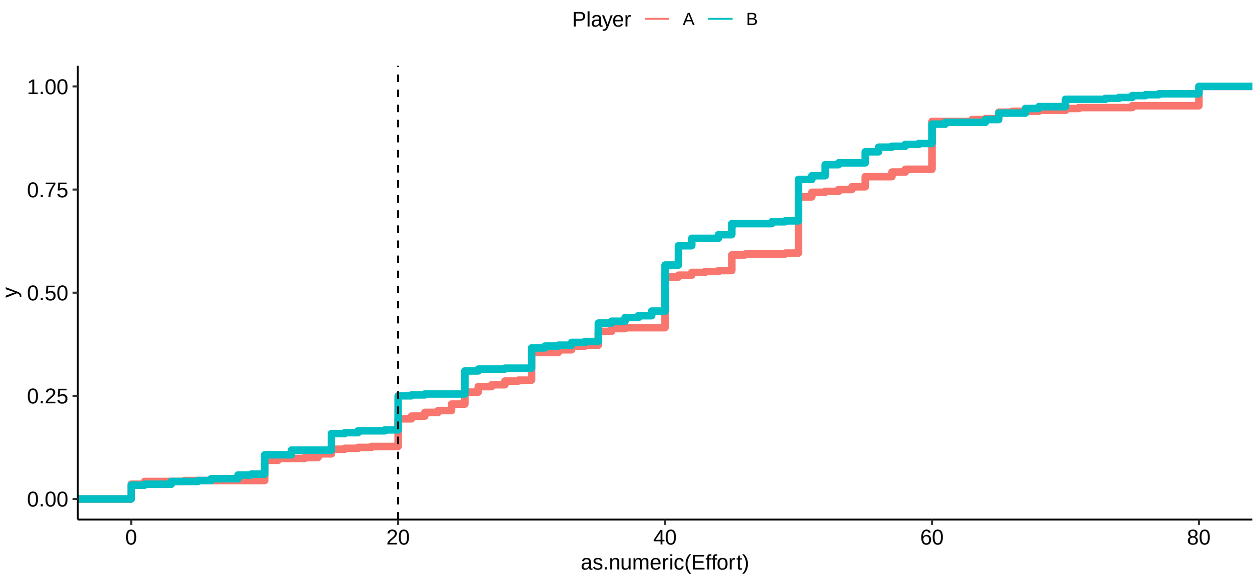


3), 4), & 5) Separately I would like to overlay player A and B's CDFs for districts 1-3 in the map where they bid on all districts since there's no reason for these to differ.

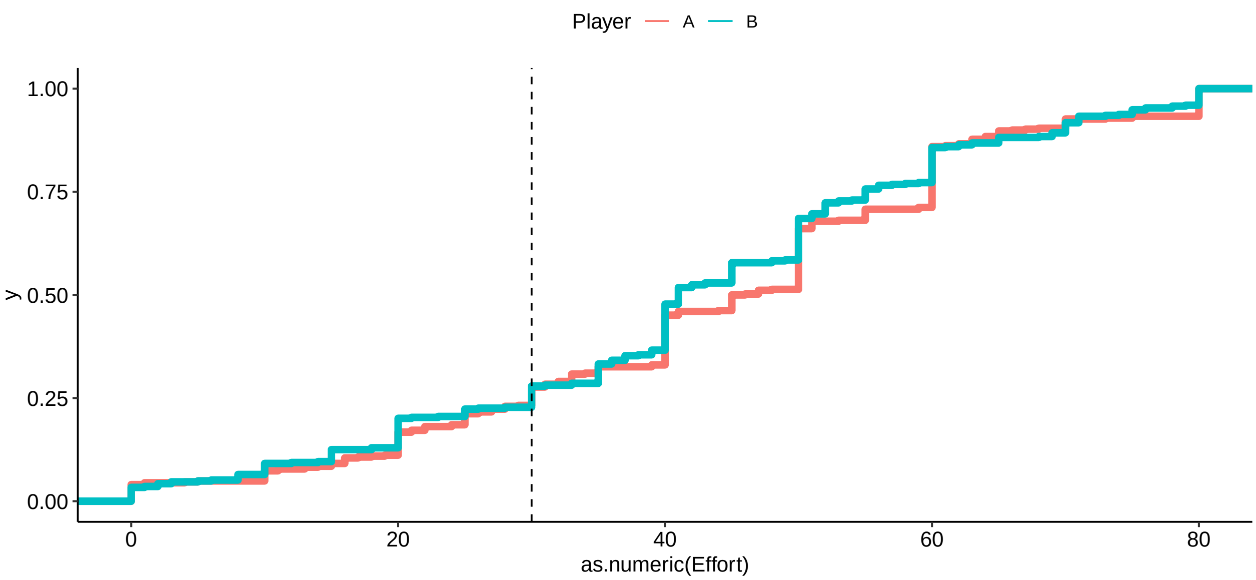


6) & 7) On each of the two maps where the players are symmetric and only bidding on one district I would like to see their CDF‘s overlaid.

Symm_1_1: White District

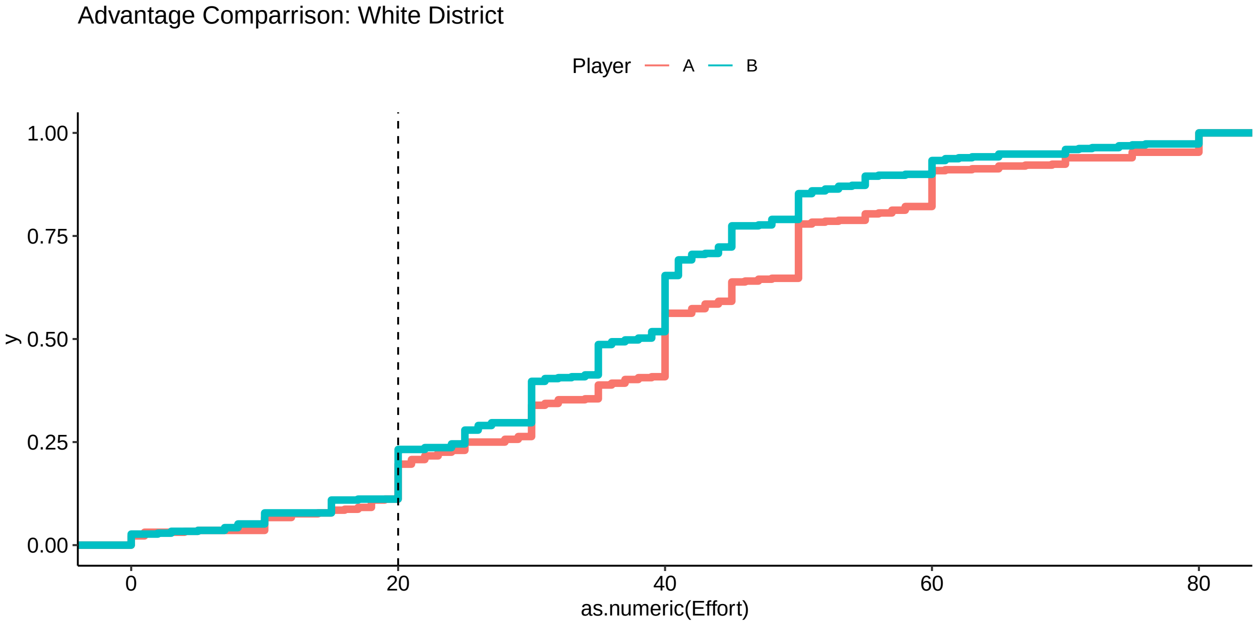


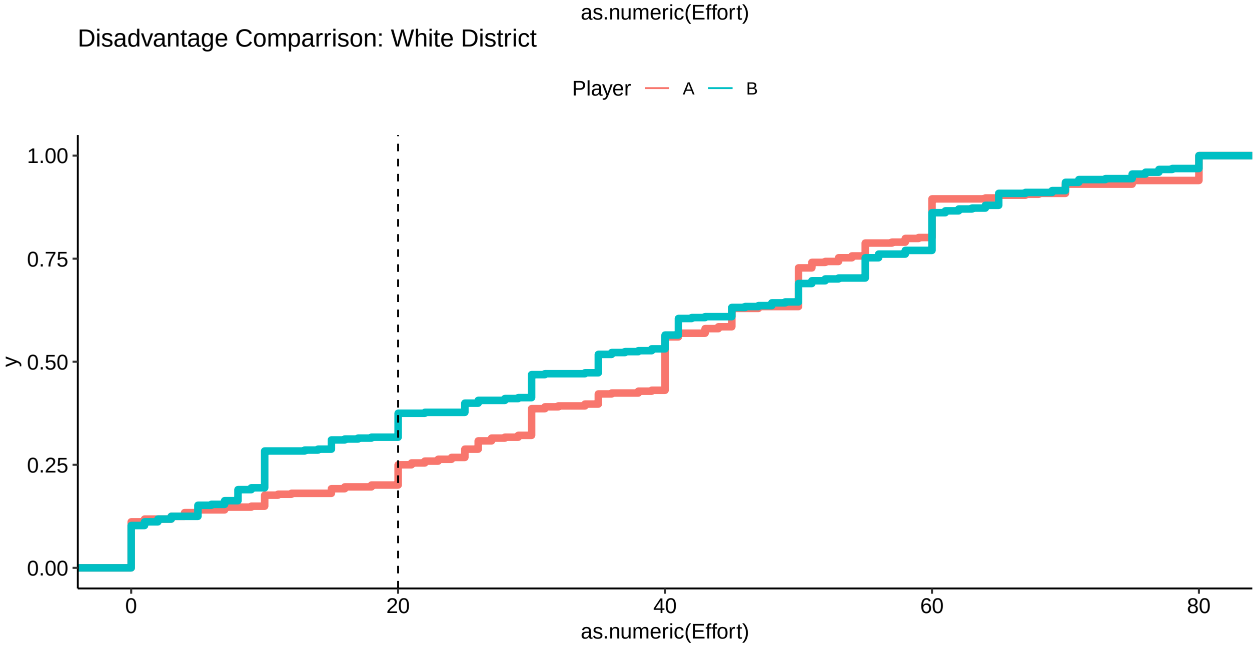
Symm_1_3: White District



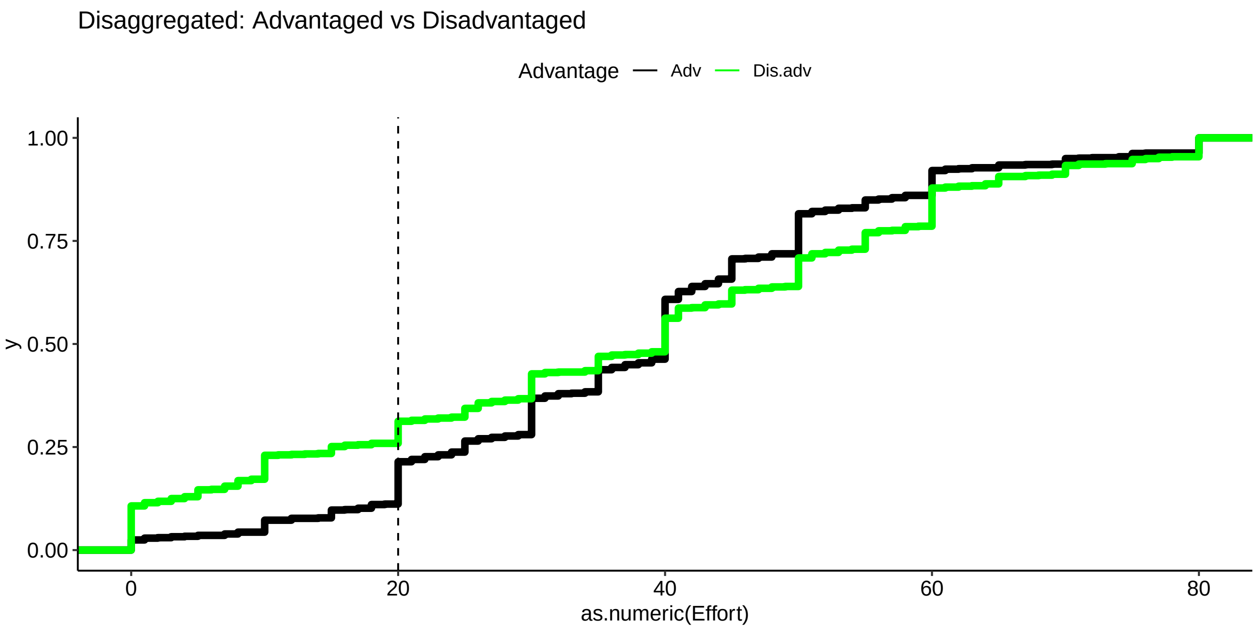
8) Overlay the CDFs of the advantage player in Gerry_B and the advantage player in Gerry_A.

9) Overlay the CDFs of the disadvantaged player in Gerry_B and the disadvantaged player in Gerry_A.





10) Assuming the two CDFs in 8) look the same and the two CDFs in 9) look the same, then make a combined advantaged CDF and a combined disadvantaged CDF and overlay those so we me can easily see how being advantaged matters.



****

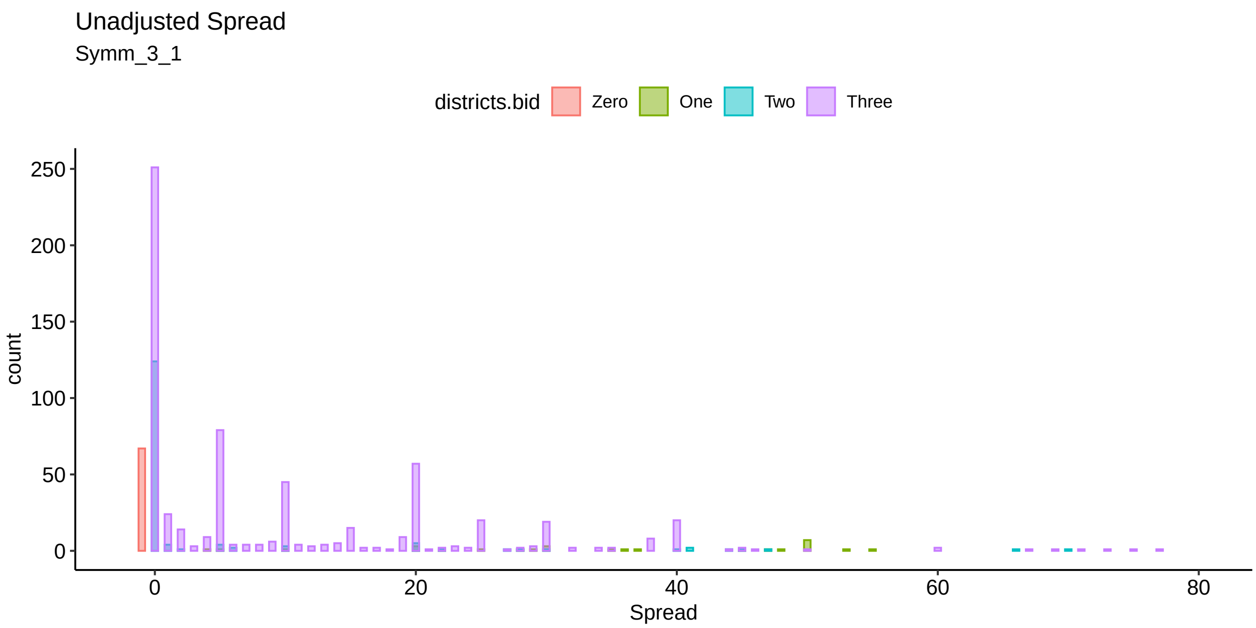
To be added as of 2021-04-07

[DONE]- One other small improvement to all the CDF figures would be to add a vertical line at the theoretical prediction for that map.

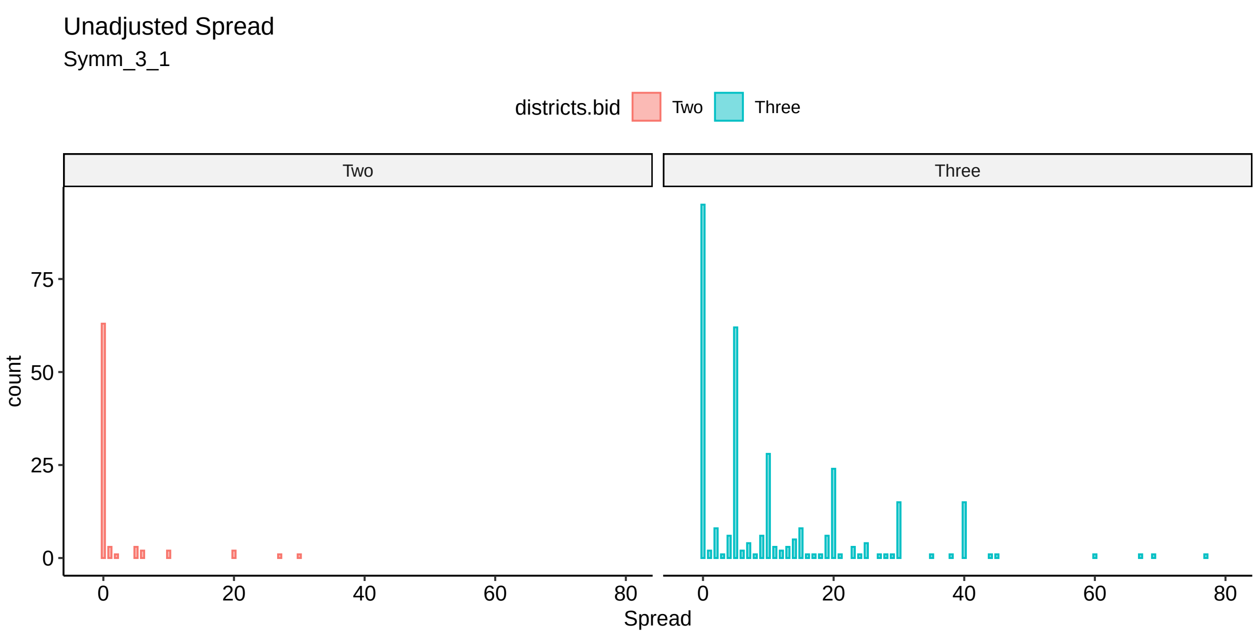
[DONE]- It looks like on Symm_3_1 there is a fair amount of zero bids placed on each map. My guess is that we have lots of instances where people bid on ONLY TWO MAPS. Could you find the proportions of cases (bid tripled by a person in a period) in Symm_3_1 where the person bid 0 on all three districts (that is in a period bid 0,0,0), bid 0 on one district (so 0,x,y or x,0,y, or x,y,0 for x,y>0), bid 0 on two districts, and bid 0 on none of the districts? My guess is that there are lots of cases where they bid 0 on one map.

```
## # A tibble: 1 x 9
##   n.records n.all.zeros n.one.district n.two.districts n.three.distric~
##       <int>       <dbl>          <dbl>           <dbl>            <dbl>
## 1       896          67             29             155              645
## # ... with 4 more variables: pct.zeros <dbl>, pct.bid.one <dbl>,
## #   pct.bid.two <dbl>, pct.bid.three <dbl>
```

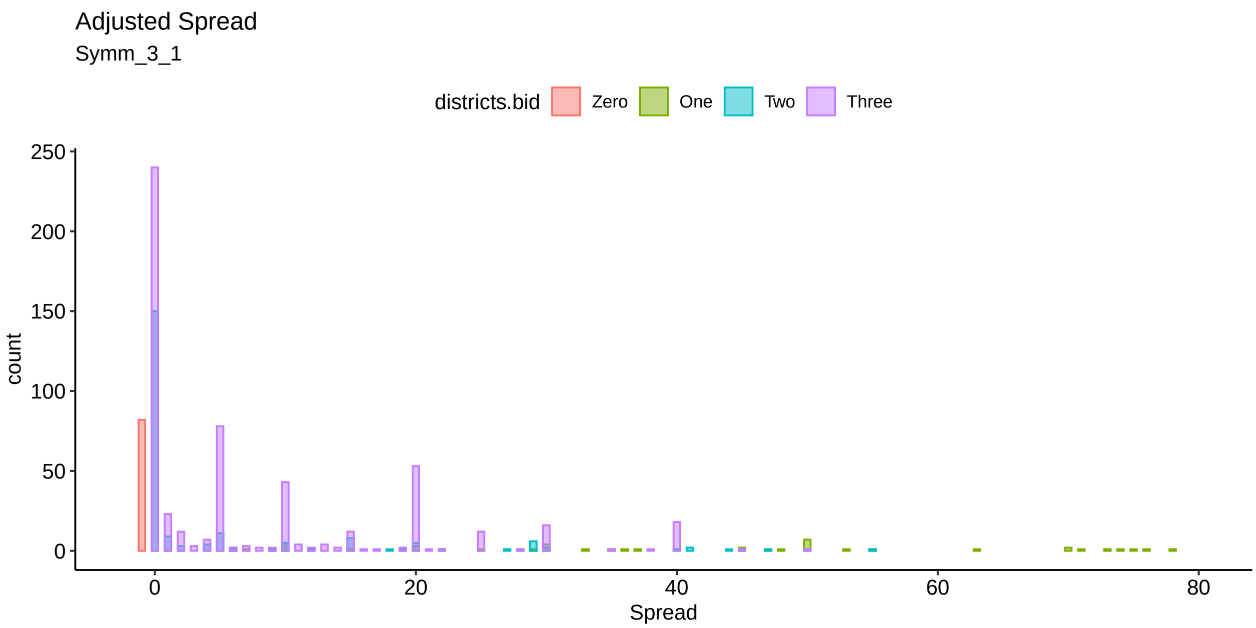
[DONE]- Look at "spread" of own bids across Symm_3_1 (max bid in any district of Symm_3_1 - min bid in any district in Symm_3_1); we'd like to see this overall (graph?) and just in the cases they bid a positive amount on everything then, for the case they only bid on 2, look at the max minus the median

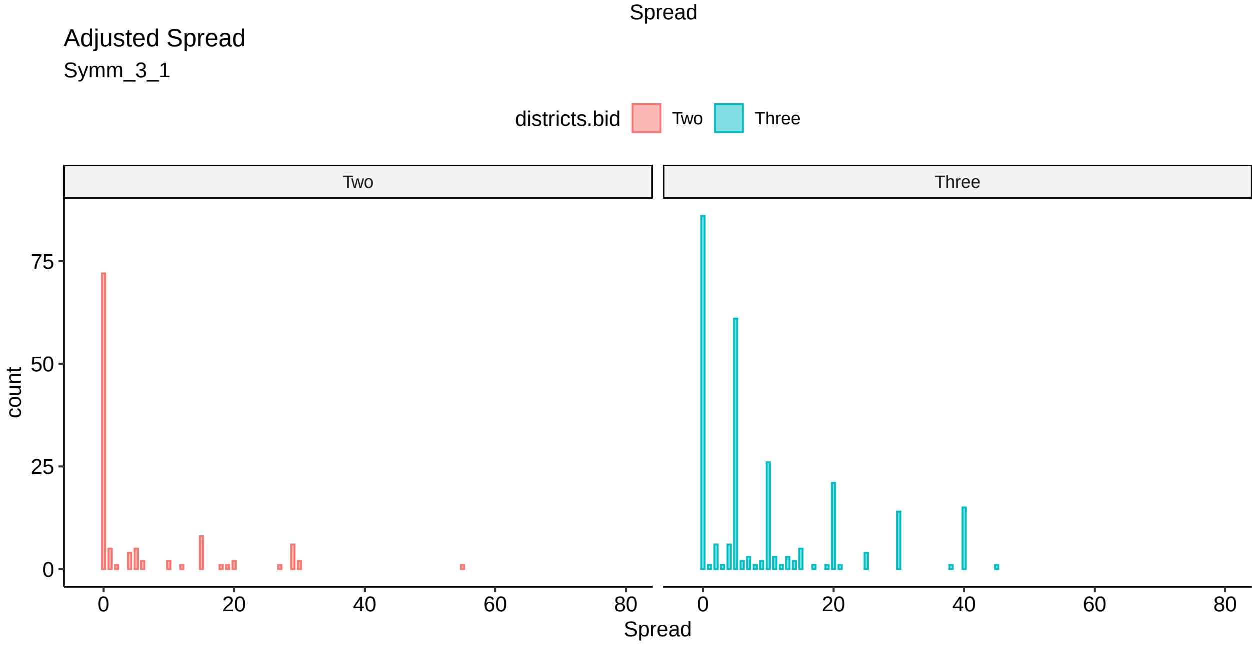


Separate graphs for bidding in two and separate for bidding in three (maybe under table with pct of Zero, One, Two, and Three bids in Symmetric_Map_3,1)




```
## # A tibble: 1 x 9
##   n.records n.all.zeros n.one.district n.two.districts n.three.distric~
##       <int>       <dbl>          <dbl>           <dbl>            <dbl>
## 1       896          82             47             218              549
## # ... with 4 more variables: pct.zeros <dbl>, pct.bid.one <dbl>,
## #   pct.bid.two <dbl>, pct.bid.three <dbl>
```





[DONE]- As a first pass, we should run a K-S tests to see if the various pairs of distributions you overlaid are the same.

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(EDG4A)) and as.numeric(unlist(EDG4B))
## D = 0.10938, p-value = 0.009408
## alternative hypothesis: two-sided
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(ELG4A)) and as.numeric(unlist(ELG4B))
## D = 0.069196, p-value = 0.2337
## alternative hypothesis: two-sided
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(EW4A)) and as.numeric(unlist(EW4B))
## D = 0.035714, p-value = 0.9375
## alternative hypothesis: two-sided
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(EW2A)) and as.numeric(unlist(EW2B))
## D = 0.087054, p-value = 0.06707
## alternative hypothesis: two-sided
```

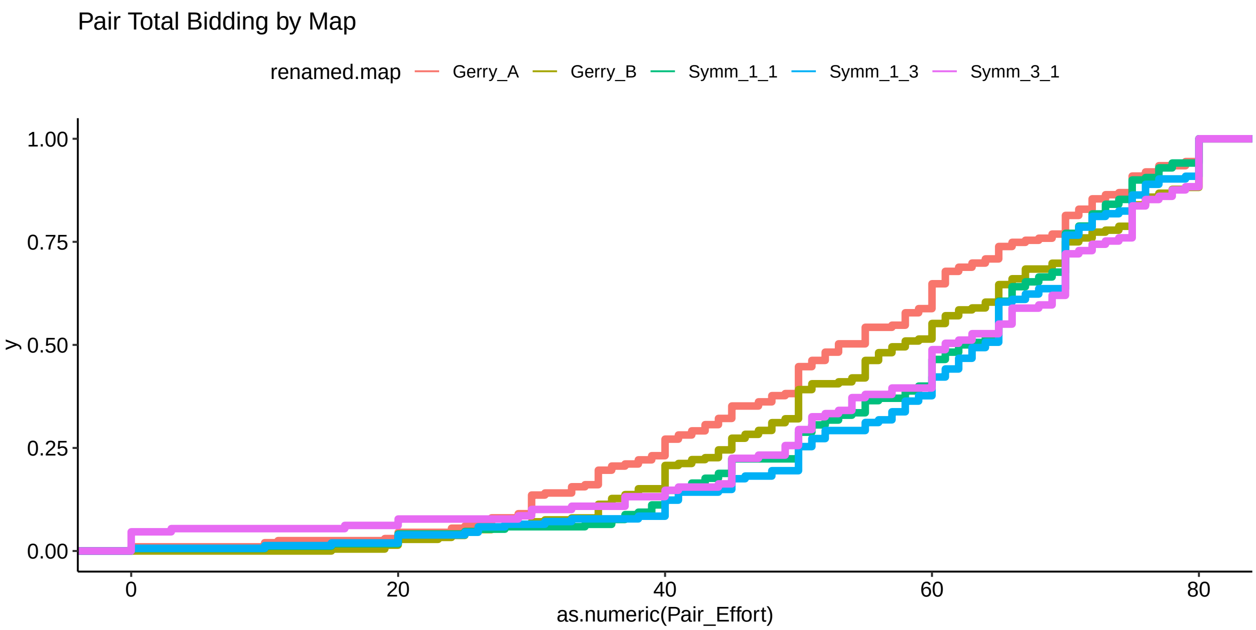
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(EW3A)) and as.numeric(unlist(EW3B))
## D = 0.078125, p-value = 0.1298
## alternative hypothesis: two-sided
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(ADV.A)) and as.numeric(unlist(ADV.B))
## D = 0.14286, p-value = 0.000214
## alternative hypothesis: two-sided
```

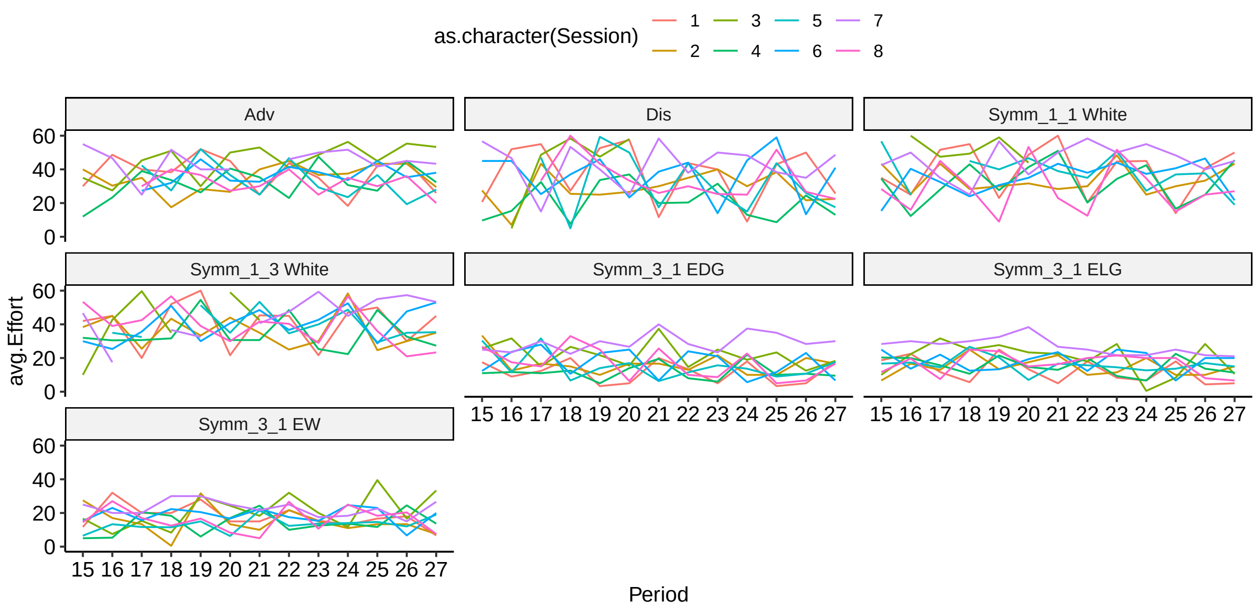
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(Dis.ADV.A)) and as.numeric(unlist(Dis.ADV.B))
## D = 0.125, p-value = 0.001824
## alternative hypothesis: two-sided
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(ADV.All)) and as.numeric(unlist(Dis.ADV.All))
## D = 0.15848, p-value = 3.369e-10
## alternative hypothesis: two-sided
```

[DONE]- Also, since it seems that things are symmetric, it would be good to make a single graph that has the cdfs of total pair level investment by map (here a pair in a period is an observation). That way we can see if more is spent on some maps than others.



[DONE]- One thing that would be good to do is for each kind of choice (advantaged in map 1 or 5, disadvantaged in map 1 or 5, white in map 2, white in map 3, all regions in map 4) take the average across all subjects in a period. Then plot a time series of those averages. This should include phase 1 and 2 so we can see if map selection impacted bidding on maps.



[DONE]- A small cosmetic point is to make sure you keep the x-axis fixed to make comparisons between graphs easier. It is not a big deal for this, just something to do in general. In the first part of the document you have some that include 80 and some that don't.

[DONE]- Average bid on each district on each map by role

```
## # A tibble: 6 x 4
## # Groups:   Player, Map [1]
##   Player Map   District avg.Effort
##   <chr>  <chr> <chr>         <dbl>
## 1 A      1     EDG            1.68
```

```
## 2 A       1      ELG             1.76
## 3 A       1      EW             37.8
## 4 A       1      pEDG            6.50
## 5 A       1      pELG            2.66
## 6 A       1      pEW            36.2
```

[DONE]- Percent gerrymandering in stage 2

```
## [1] 0.6875
```

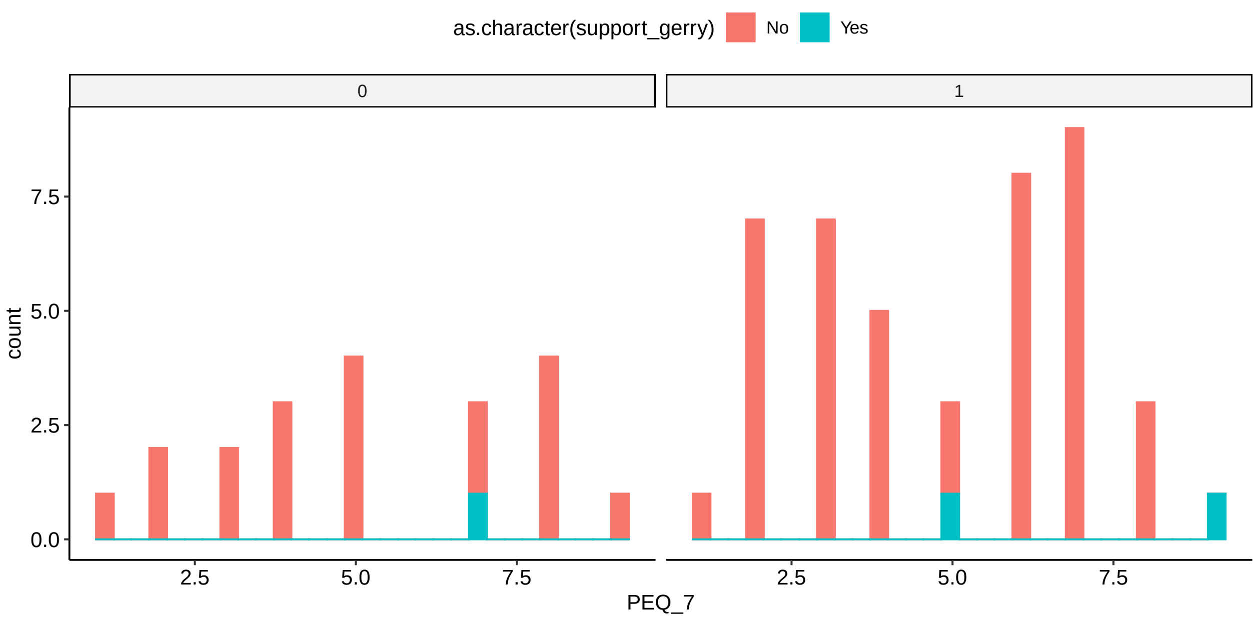
[DONE]- Percentage picking each map in stage 3

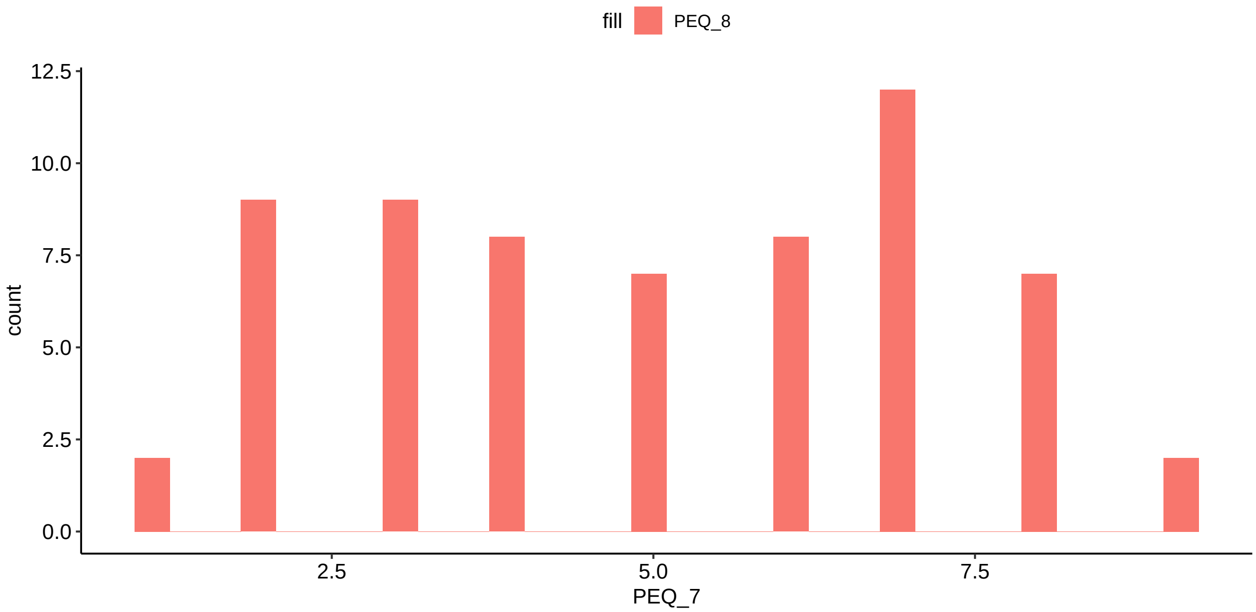
```
## # A tibble: 6 x 3
##   Map_Selection     n pct.of.pop
##           <dbl> <int>      <dbl>
## 1           -99     1          2
## 2             1     9         14
## 3             2     8         12
## 4             3    22         34
## 5             4    15         23
## 6             5     9         14
```

[DONE]- Rank sum test looking at whether or not their political views influence whether they gerrymander or not. . . ?

Before the rank sum test let's recall the PEQ relevant for the test.

PEQ_7: "On a scale of 1 to 9, how would you describe your political views with 1 being extremely liberal (i.e. to the left of the Democratic Party), 5 being centrist (i.e. falling between the Democratic Party and the Republican Party), and 9 being extremely conservative (i.e. to the right of the Republican party)." (multiple choice; 1 - 9)






```
## [1] 29
```

```
## [1] 28
```

```
## [1] 0.6875
```

Now, onto the rank sum test.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  PEQ_7 by as.character(gerry)
## W = 487, p-value = 0.4966
## alternative hypothesis: true location shift is not equal to 0
```

So we fail to reject the null that the political preference is the same regardless of whether they actually gerrymandered.

What about based on whether they *support* gerrymandering? (a.k.a PEQ_8)

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  PEQ_7 by support_gerry
## W = 44, p-value = 0.1319
## alternative hypothesis: true location shift is not equal to 0
```

Also fail to reject the null that political preference is the same regardless of whether they support gerrymandering.

[DONE]- political beliefs and saying gerrymandering (**done above**; no diff. between gerrymandering and politics)

[DONE]- how either of those answers depend on whether they actually gerrymander (**above** = no diff. b/w support gerry and politics; **below** = no diff in support of gerrymandering based on whether actually gerry)

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  PEQ_8 by as.character(gerry)
```

```
## W = 438, p-value = 0.9527
## alternative hypothesis: true location shift is not equal to 0
```

[DONE]- Z test of whether observations are same for number of people selecting whether they support gerrymandering or not (same # of people in both camps; probably going to be diff given the distribution between y and n)

(^In Sig.)

[DONE]- Of the people who say they don't support it, what % actually did it

```
nrow(subset(gerry_and_politics, PEQ_8 == 2 & gerry == 1))/nrow(subset(gerry_and_politics, PEQ_8 == 2))
```

```
## [1] 0.6885246
```

[DONE]- for the same split, did they say they like gerrymandering or not proportionately (are the proportions the same) ????????? Only have 3 that say support gerrymandering… is this enough to make any determination?

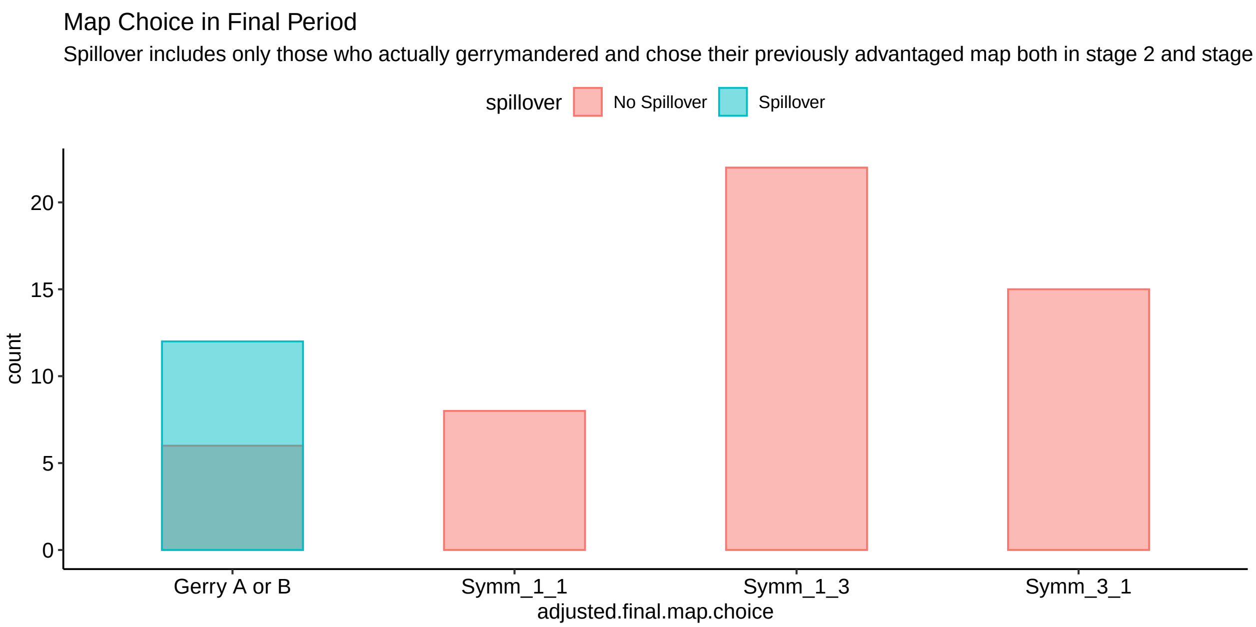
(Do you like it as a function of whether you actually did it)

```
## [1] 0.04545455
```

```
## [1] 0.05
```

This is for the bar graph

[]- when they don't know who they are which maps are they choosing - distinguish b/w people choosing gerrymandered map based on if they are choosing it after having chosen it in previous periods - 4 bars; gerrymander A and B on one column (two colored bars; one color is "gerrymandered for self" other color "gerrymandered for other") - Some people like gerrymandered maps even not knowing who they are - Some pick gerrymander for self (have been picking the map for themselves in previous round)



[DONE]- Regression from Deck's notes

$$Effort = \alpha + \beta_1 Player_B + \beta_2 Map_2 + \beta_3 Map_2 Player_B + \beta_4 Map_3 + \beta_5 Map_3 Player_B + \beta_6 Map_4 + \beta_7 Map_4 Player_B + \beta_8 Map_5 + \beta_9$$

```
##
## Call:
```

```
## lm(formula = Effort ~ Player_B + Map_1 + Map_1 * Player_B + Map_3 +
##     Map_3 * Player_B + Map_4 + Map_4 * Player_B + Map_5 + Map_5 *
##     Player_B, data = regress_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -53.35 -14.55   1.25  15.45  38.78
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     45.3058     1.0328  43.865  < 2e-16 ***
## Player_B         1.2879     1.4607   0.882  0.37795
## Map_1           -4.0893     1.4607  -2.800  0.00514 **
## Map_3            0.9732     1.4607   0.666  0.50526
## Map_4            6.2656     1.4607   4.290 1.83e-05 ***
## Map_5           -0.7589     1.4607  -0.520  0.60338
## Player_B:Map_1   2.8571     2.0657   1.383  0.16669
## Player_B:Map_3   1.1830     2.0657   0.573  0.56687
## Player_B:Map_4   0.4866     2.0657   0.236  0.81378
## Player_B:Map_5  -2.6674     2.0657  -1.291  0.19667
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.86 on 4470 degrees of freedom
## Multiple R-squared:  0.02517,    Adjusted R-squared:  0.02321
## F-statistic: 12.83 on 9 and 4470 DF,  p-value: < 2.2e-16
```

[Done?]- Regression of average bid as function of period with dummy variable for Map selection phase (periods 25,26,27)

(so we just want the impact on the map selection phase on the average map level bids)

```
##
## Call:
## lm(formula = avg.effort ~ Selection_Stage, data = period.averages)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7209 -2.9553 -0.6116  2.7478 10.3572
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       48.096      0.574  83.796  < 2e-16 ***
## Selection_Stage   -5.234      1.195  -4.381 4.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.059 on 63 degrees of freedom
## Multiple R-squared:  0.2335, Adjusted R-squared:  0.2214
## F-statistic: 19.19 on 1 and 63 DF,  p-value: 4.555e-05
```