Gerrymandering: Exploring the Data

An, Anderson, and Deck

4/4/2021

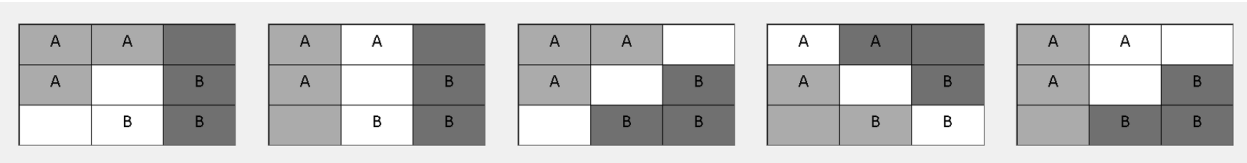
Definitely not the prettiest of graphics, but we can revisit the design.

We first look at CDFs of bids for each district in each map for Players A and B.

Next we depict the modal map selection by each subject during stage 2 in a simple bar graph.

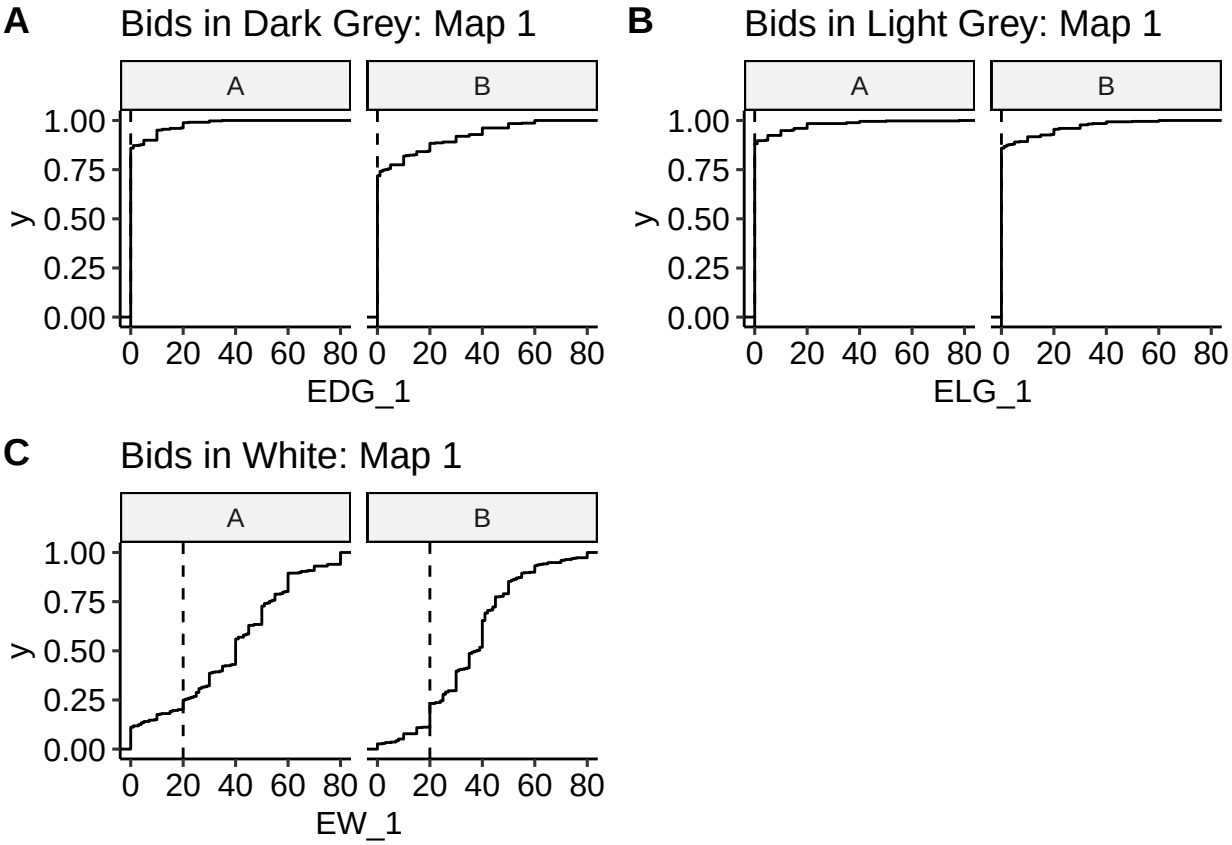
Lastly we depict the final stage distribution of map selection for the entire sample and by Player (A or B).

For reference:

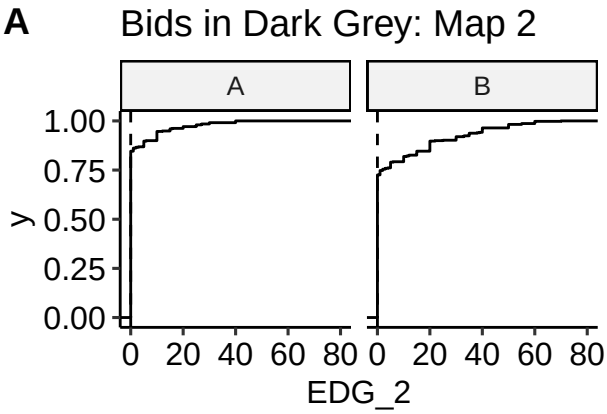


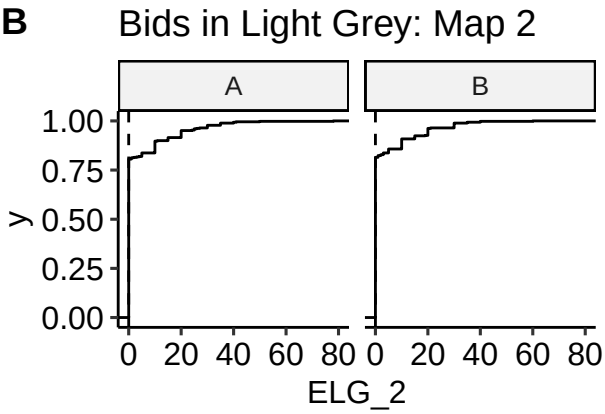
Note that in Maps 1, 2, 3, and 5 the white district is the only competitive district in the sense that only the competition within the white district determines whether a subject wins that Map. The exception is Map 4 in which it is logical to bid in any district as no district is guaranteed a victor.

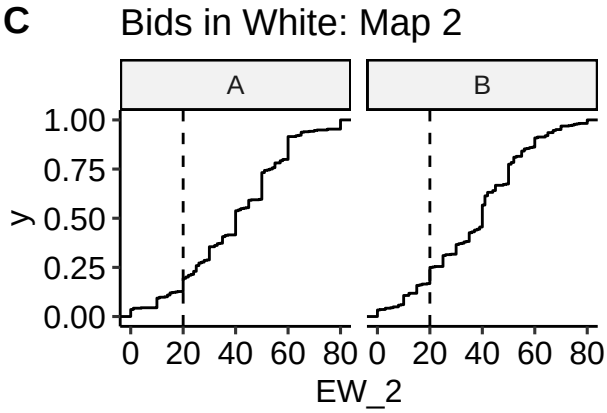
Map 1



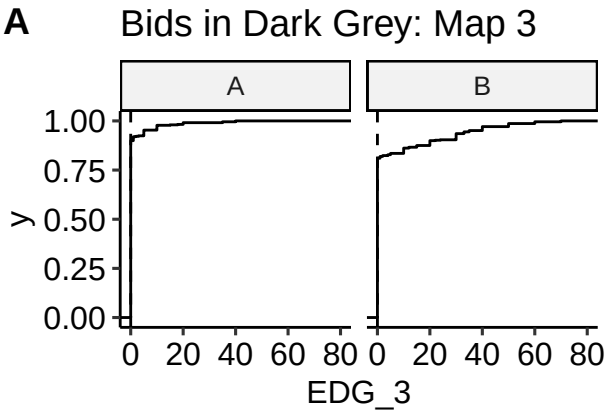
Map 2

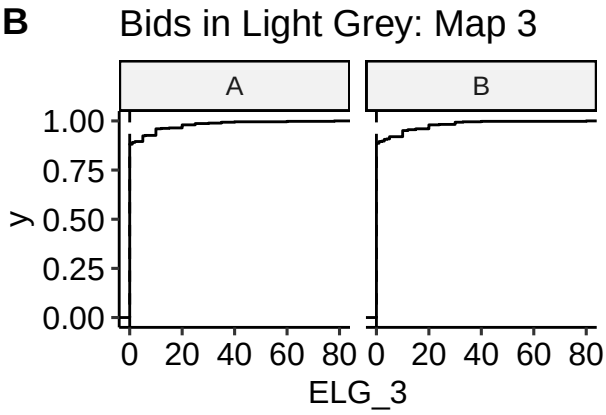

A
Bids in Dark Grey: Map 2
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EDG_2

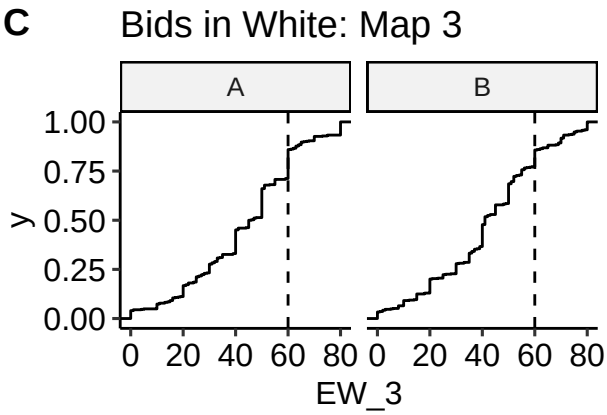

B
Bids in Light Grey: Map 2
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
ELG_2


C
Bids in White: Map 2
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EW_2

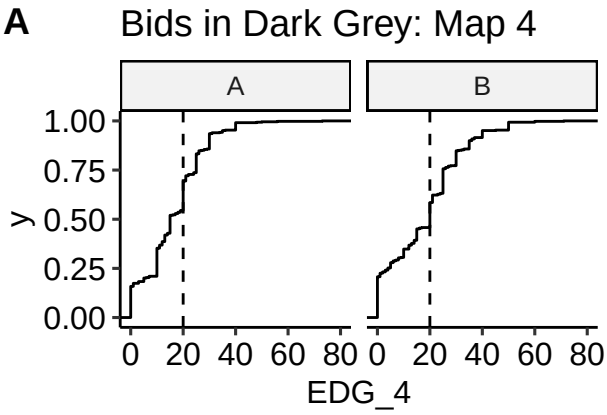
Map 3

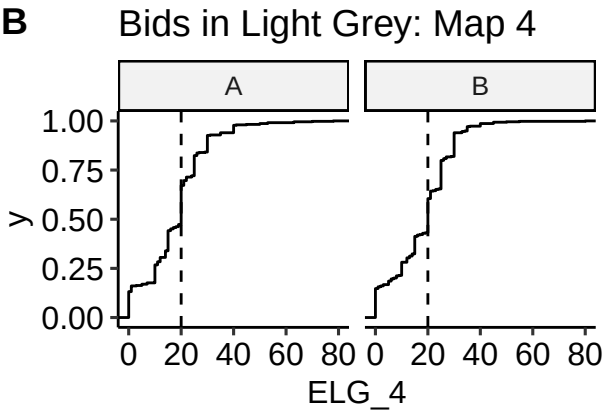

A
Bids in Dark Grey: Map 3
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EDG_3


B
Bids in Light Grey: Map 3
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
ELG_3


C
Bids in White: Map 3
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EW_3

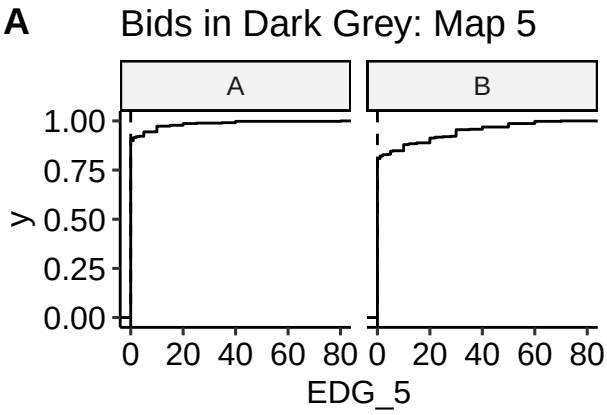
Map 4

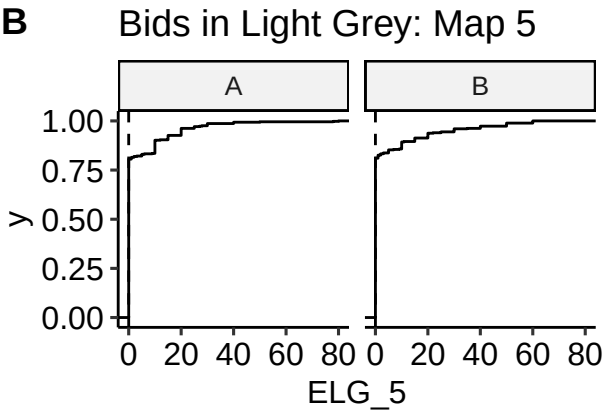

A
Bids in Dark Grey: Map 4
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EDG_4

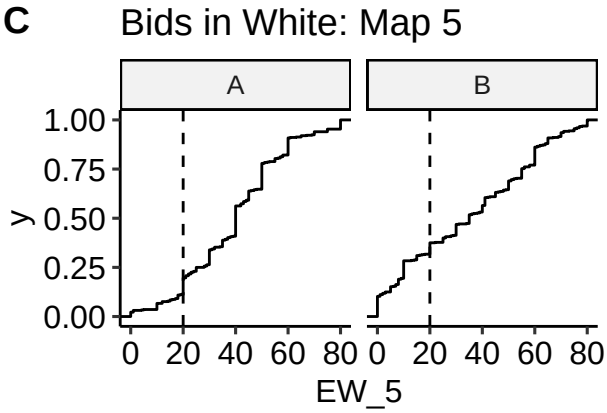

B
Bids in Light Grey: Map 4
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
ELG_4


C
Bids in White: Map 4
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EW_4

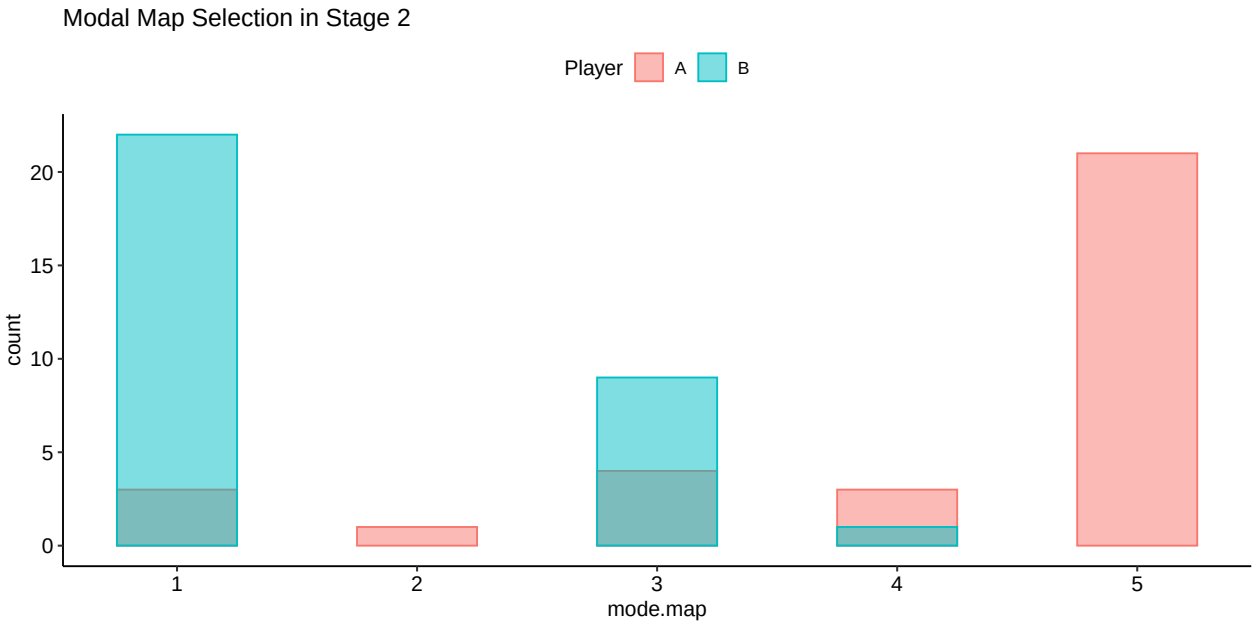
Map 5


A
Bids in Dark Grey: Map 5
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EDG_5

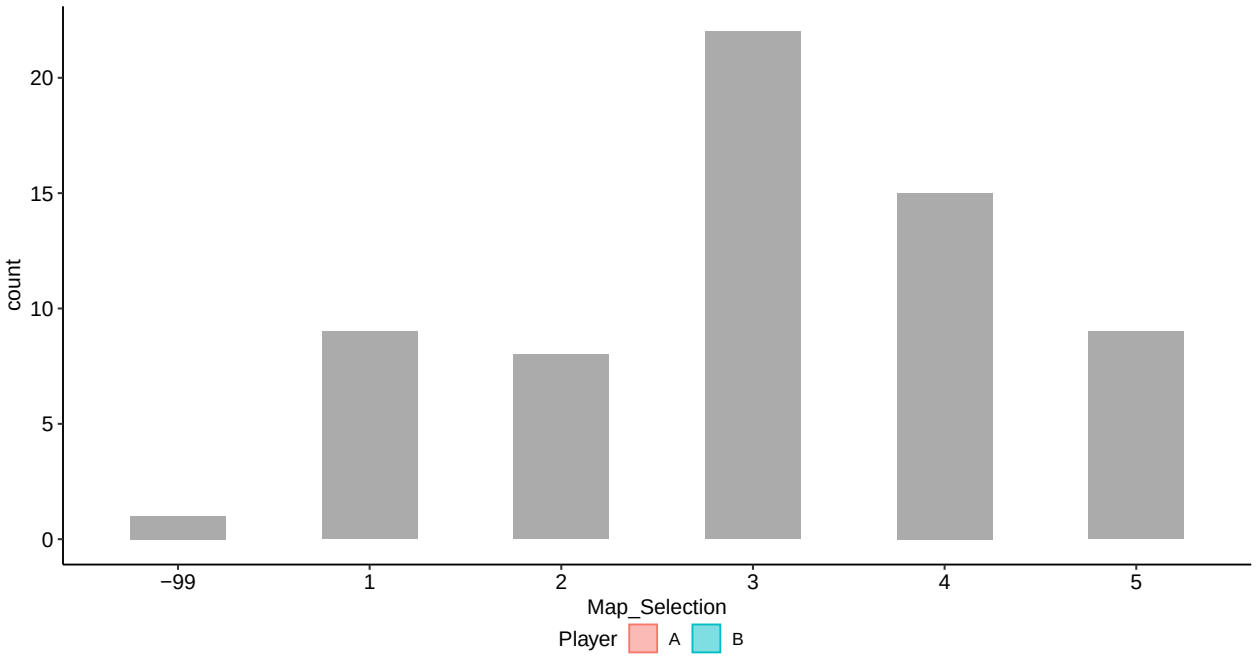

B
Bids in Light Grey: Map 5
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
ELG_5

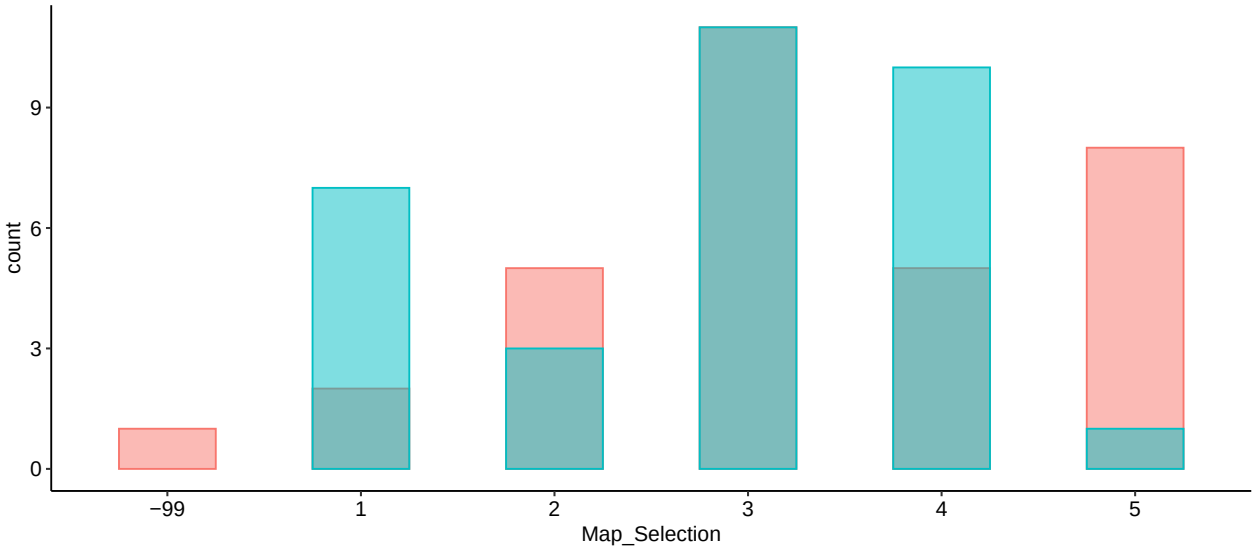

C
Bids in White: Map 5
A
B
1.00
0.75
0.50
0.25
0.00
y
0 20 40 60 80
0 20 40 60 80
EW_5

Recall, Player A should pick Map 5 and Player B should pick Map 1 if they are choosing the map that gives them the best chance of winning.



Notice, in the below figures the singular choice of -99 represents the individual (subject 4 from session 8) that excused themselves from the experiment during this part of their session. 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.



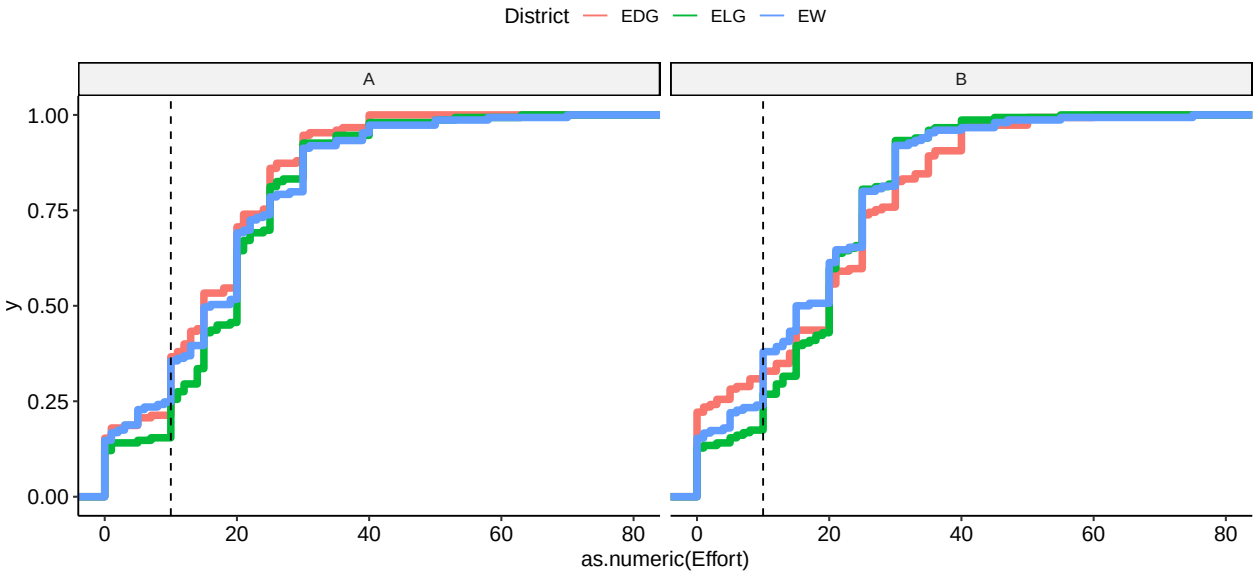


Now we are addressing:

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.

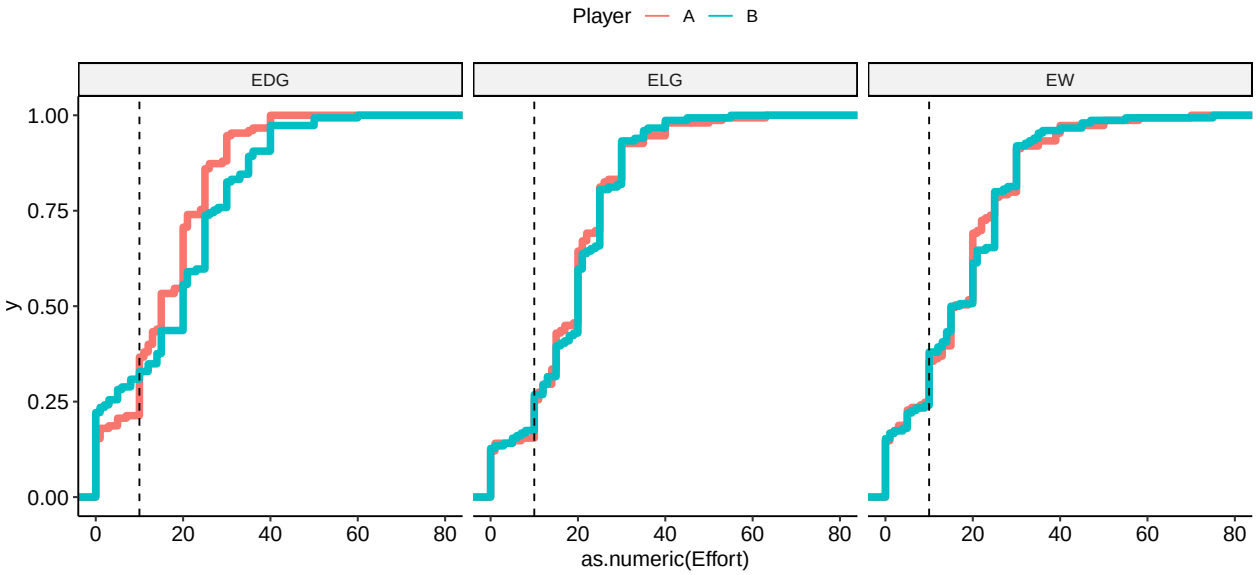
2) I would also like to see player B's CDFs overlaid

Map 4



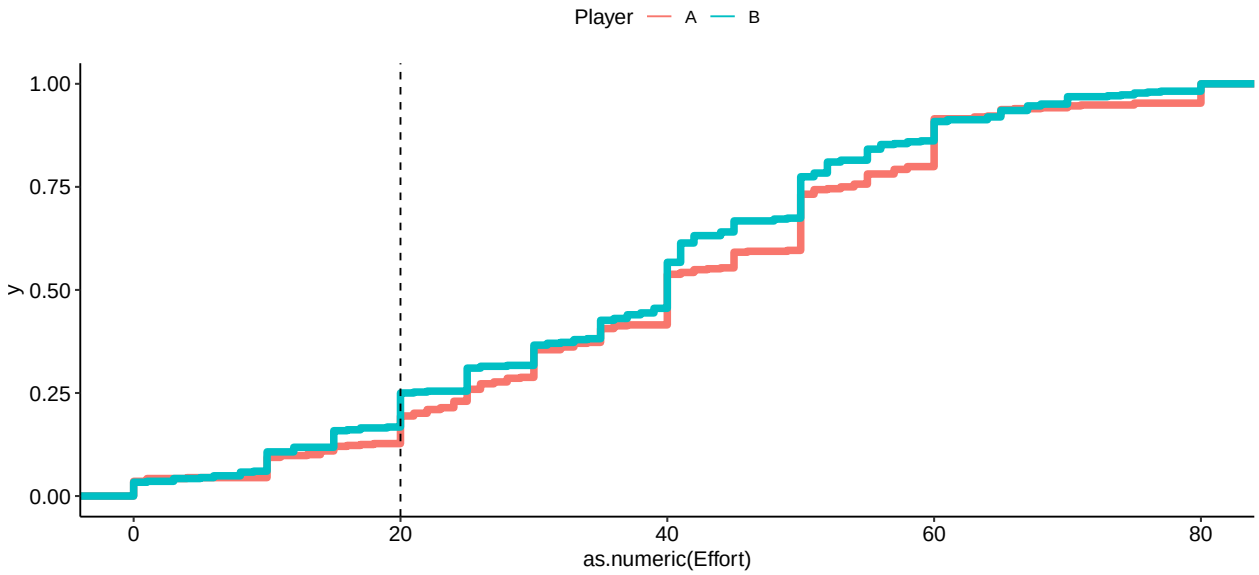
3), 4), & 5) Separately I would like to overlay player A and B's CDFs for districts 1-3 in the map where they bid on all districts since there's no reason for these to differ.

Map 4 by District

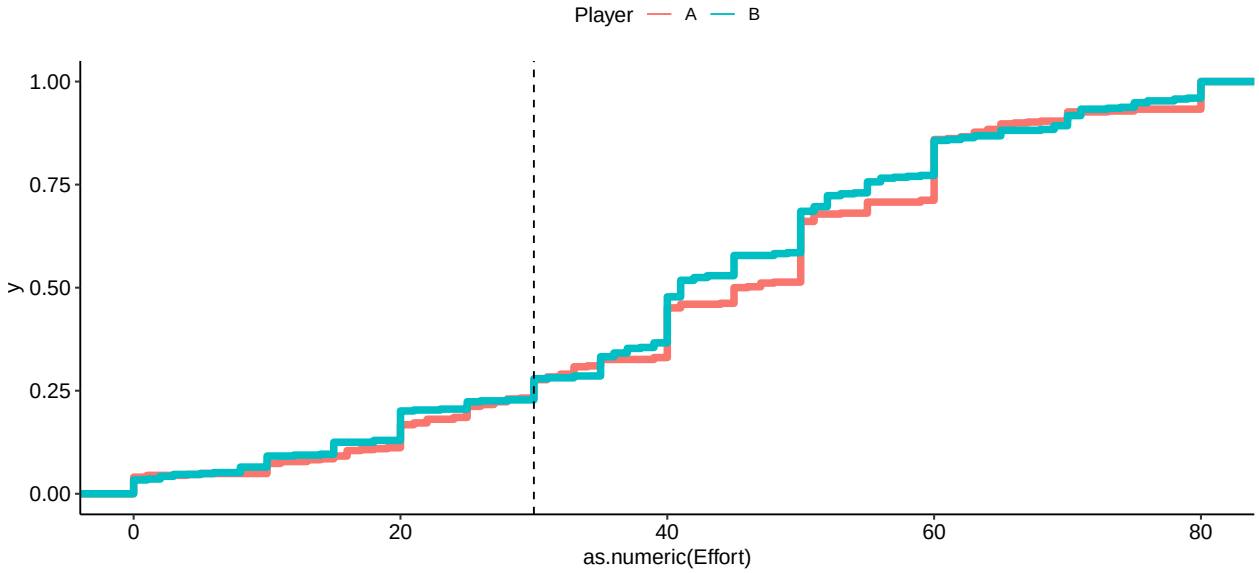


6) & 7) On each of the two maps where the players are symmetric and only bidding on one district I would like to see their CDF's overlaid.

Disaggregated Map 2: White District



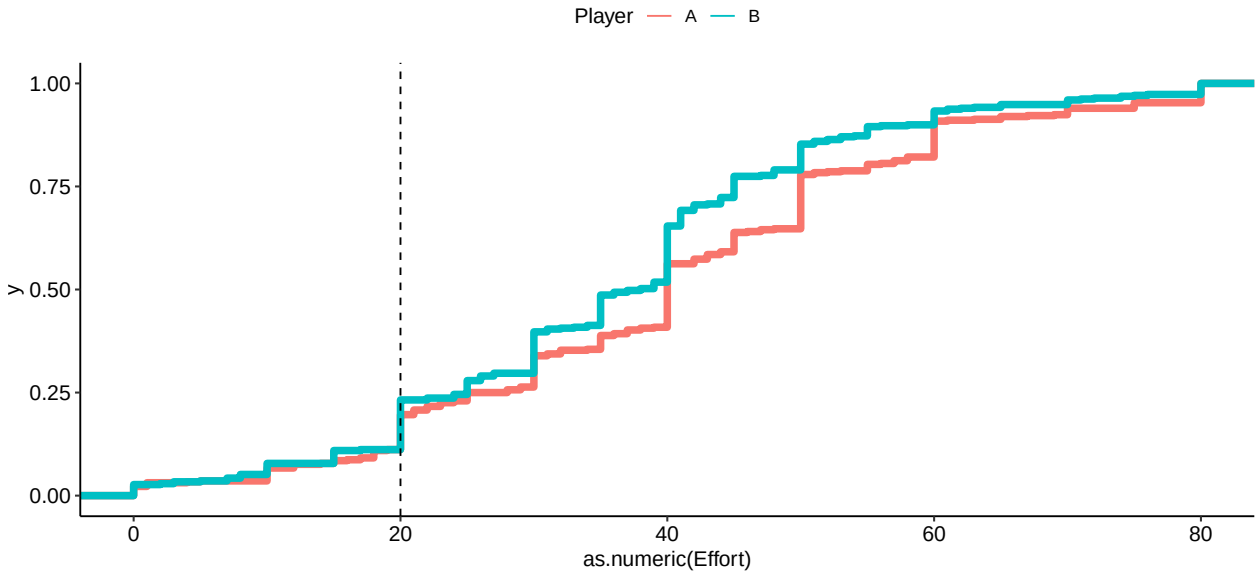
Disaggregated Map 3: White District



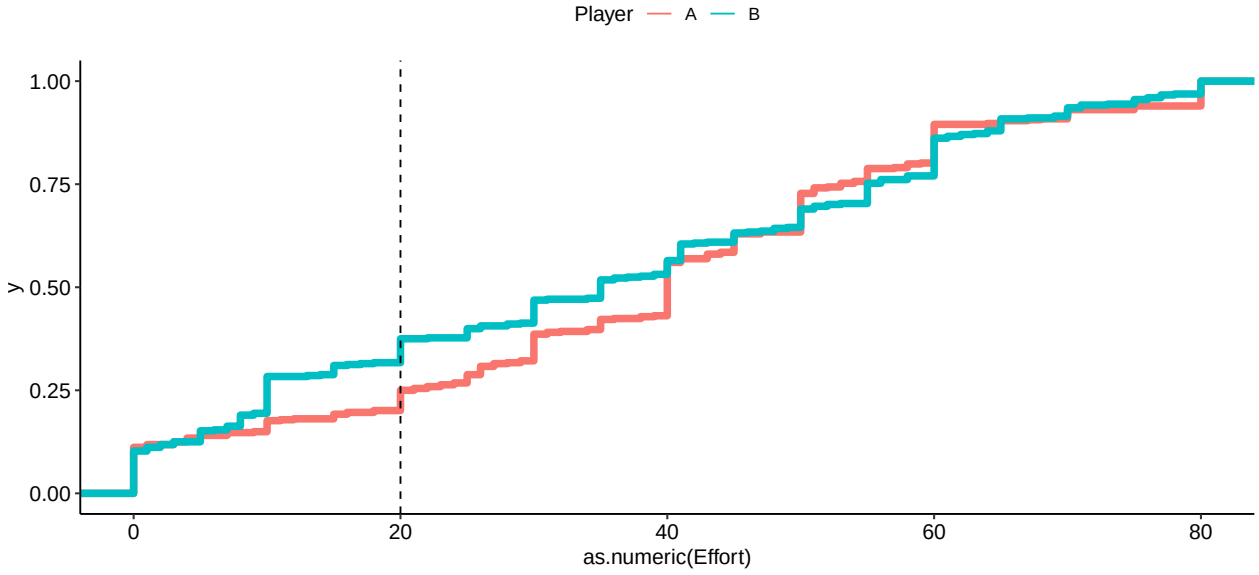
8) Overlay the CDFs of the advantage player in map 1 and the advantage player in map 5.

9) Overlay the CDFs of the disadvantaged player in map 1 and the disadvantaged player in map 5.

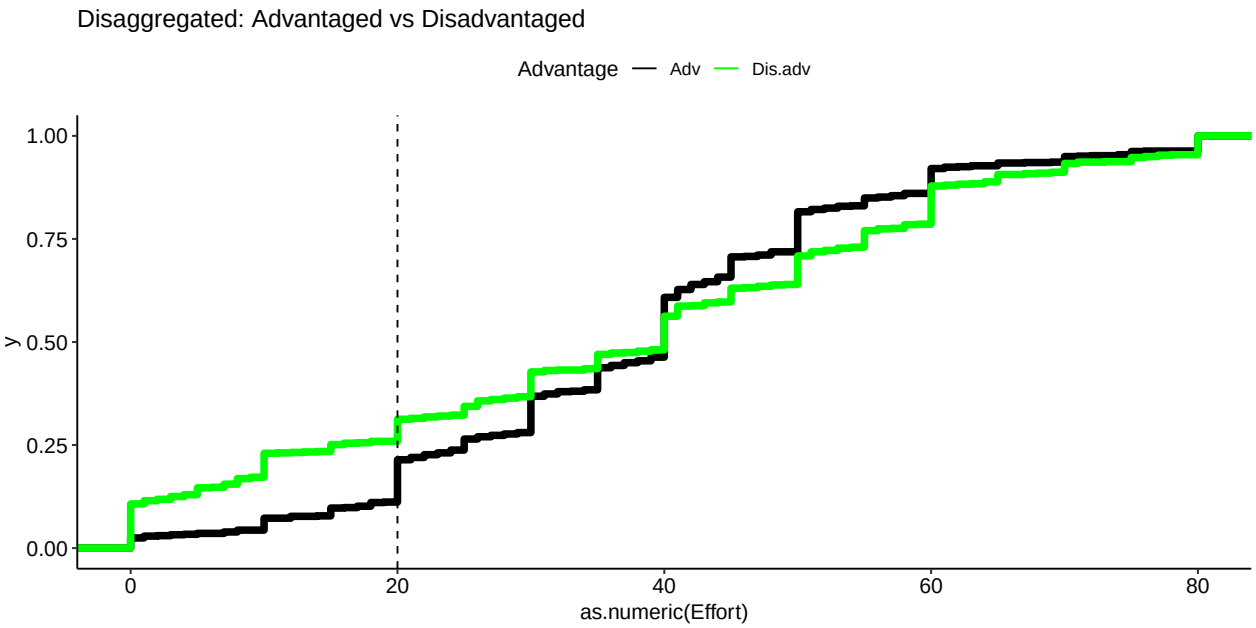
Advantage Comparrison: White District



Disadvantage Comparrison: White District



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.



****

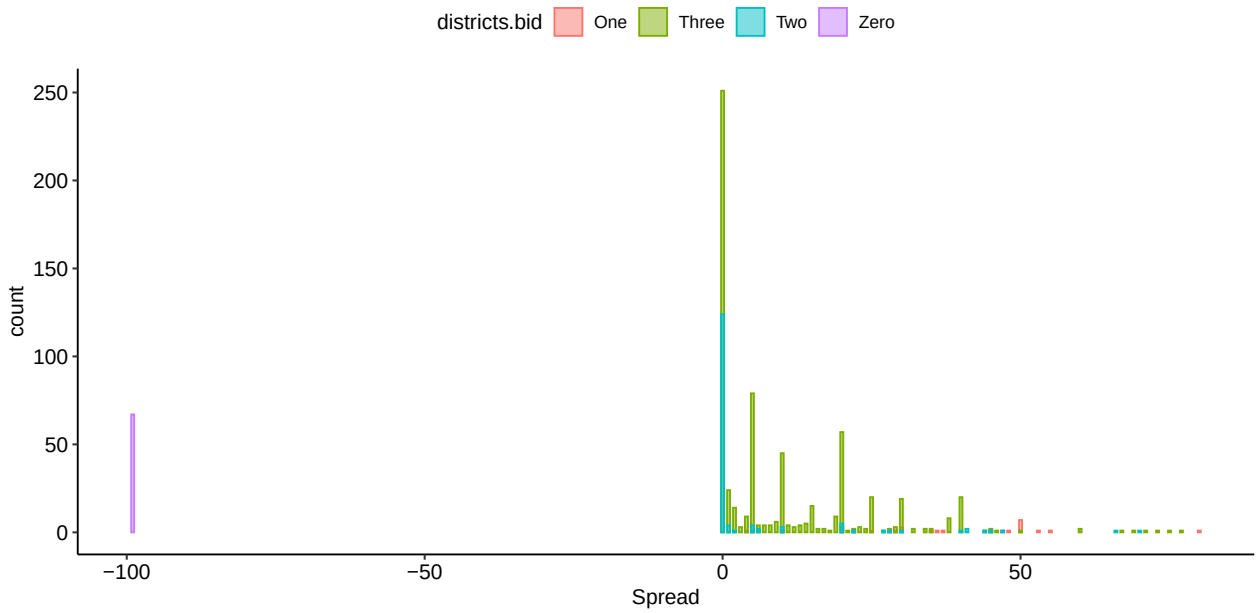
To be added as of 2021-04-07

[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 map 4 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 map 4 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 map 4 (max bid in any district of Map 4 - min bid in any district in Map 4); 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



[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
```

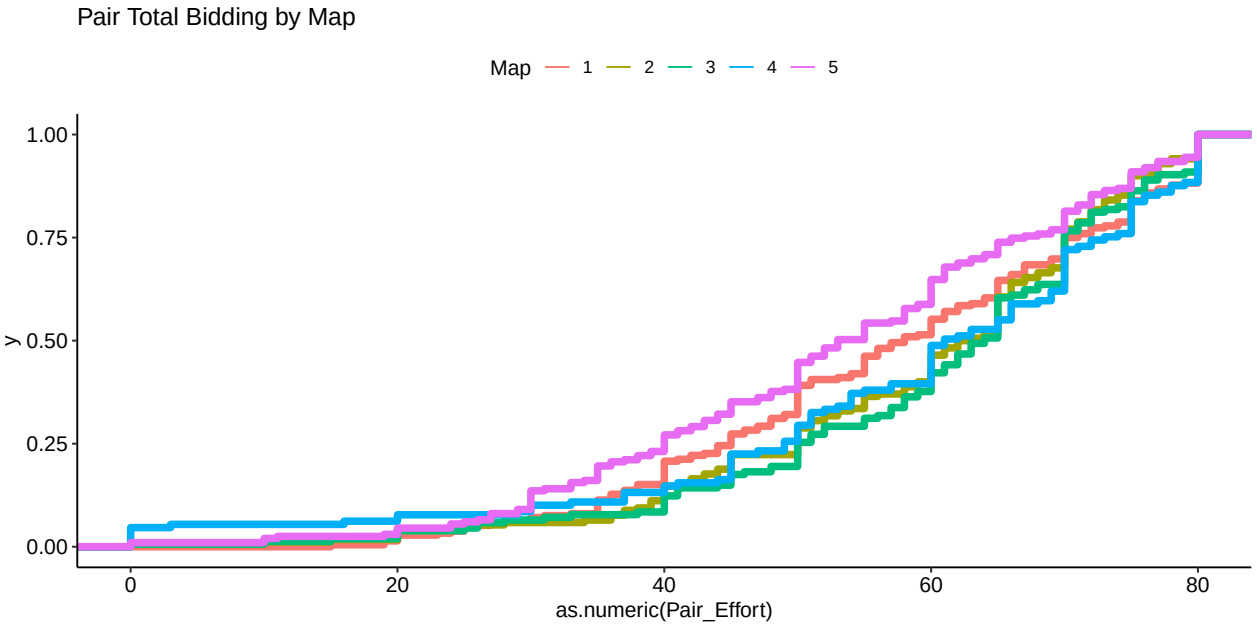
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(EW3A)) and as.numeric(unlist(EW3B))
## D = 0.078125, p-value = 0.1298
## alternative hypothesis: two-sided
```

```
##
##  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
```

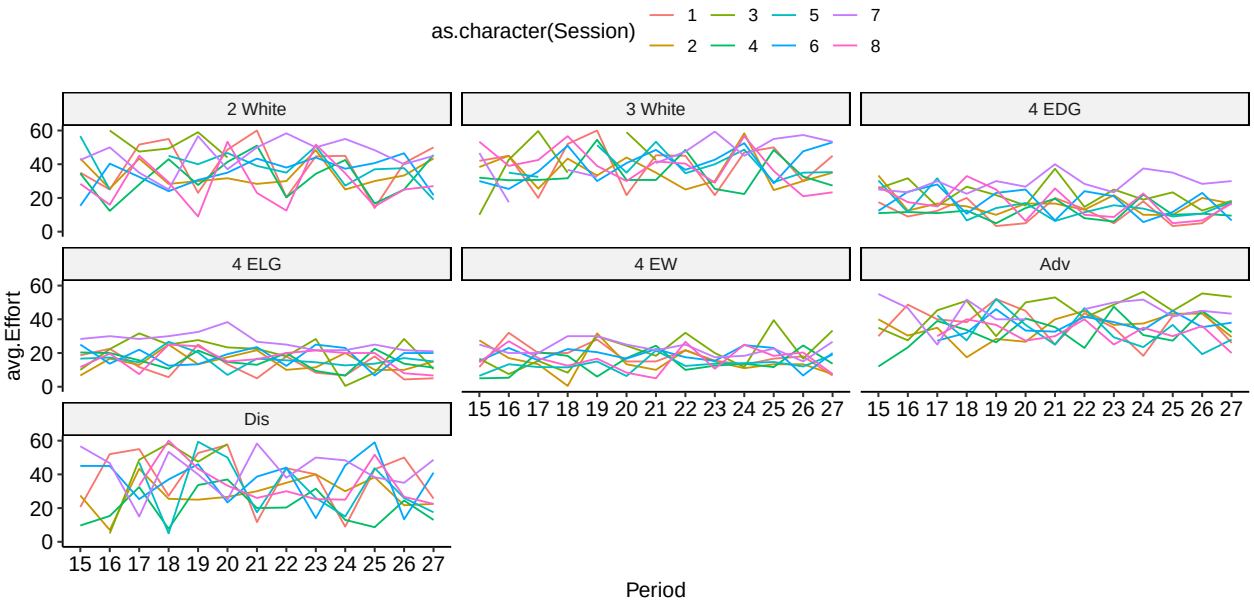
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  as.numeric(unlist(Dis.ADV.A)) and as.numeric(unlist(Dis.ADV.B))
## D = 0.125, p-value = 0.001824
## alternative hypothesis: two-sided
```

```
##
##  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] 0.671875
```

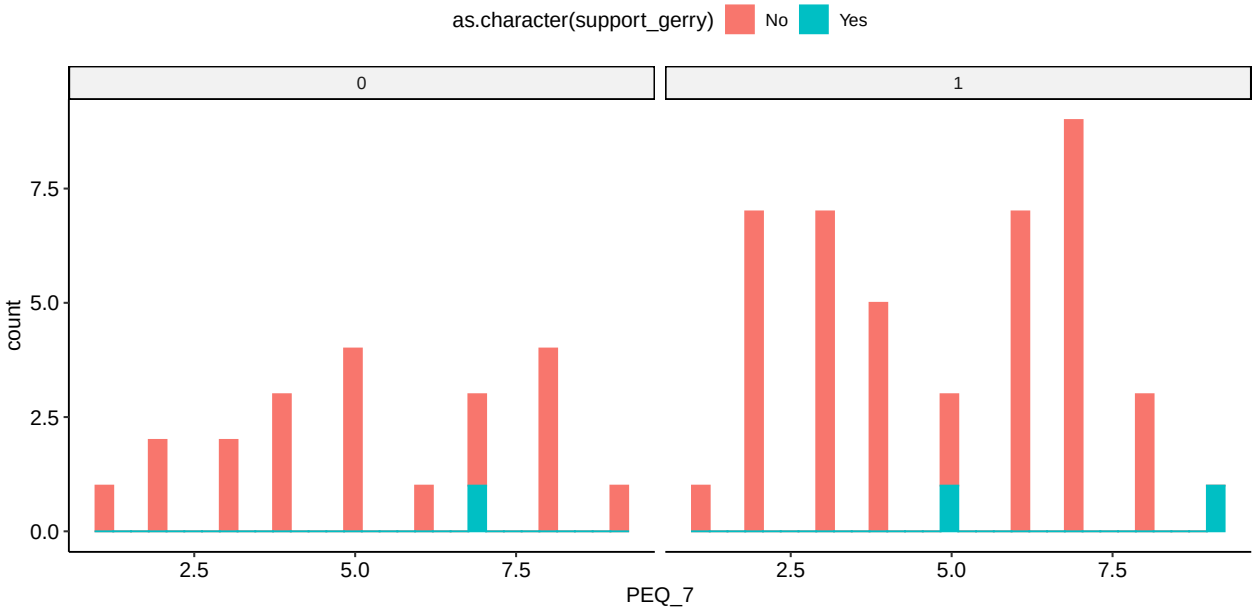
[DONE]- Percentage picking each map in stage 3

```
## # 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)



Now, onto the rank sum test.

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  PEQ_7 by as.character(gerry)
## W = 505.5, p-value = 0.44
## 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 = 451, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

[]- T 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)

[]- Look at for gerrymanders, is their political response different than the political response of the non-gerrymanders (can gather two vectors; yes support group and no support then are their politics the same)

[]- for the same split, did they say they like gerrymandering or not (are the proportions the same)

[]- comparing the two survey responses; the relationship between gerrymandering and political views

[]- when they don't know who they are which maps are they choosing

[]- Regression from Deck's notes