

Replication Code for Massive Election Fraud?: A Compendium of Statistically Fallacies in Claims about the 2020 Presidential Election

Jonathan Cervas

Bernard Grofman

2023-11-23

Accepted, Statistics and Public Policy

Setup

Set directories where data will be read from or written to

```
dir.download <- "/Users/cervas/Downloads"
dir.git <- "/Users/cervas/My Drive/GitHub/Data Files"
dir.online.git <- "https://raw.githubusercontent.com/jcervas/Data"
dir.paper <- "/Users/cervas/My Drive/GitHub/jcervas.github.io/2023/SPP"
dir.data <- paste0(dir.paper, "/data")
dir.figures <- paste0(dir.paper, "/figures")
dir.gis <- paste0(dir.paper, "/GIS")
```

Read in Functions used in other projects

```
source("https://raw.githubusercontent.com/jcervas/R-Functions/main/seatsvotes.R")
```

```
## Seats-Votes Function - v1.0
```

```
source("https://raw.githubusercontent.com/jcervas/R-Functions/main/sv-hyp.R")
source("https://raw.githubusercontent.com/jcervas/R-Functions/main/GERRYfunctions.R")
```

```
##
##
## .....
##   LOADING FUNCTIONS. . . . .
##   .....
```

Set years examined

```
## [1] 1868 1872 1876 1880 1884 1888 1892 1896 1900 1904 1908 1912 1916 1920 1924
## [16] 1928 1932 1936 1940 1944 1948 1952 1956 1960 1964 1968 1972 1976 1980 1984
## [31] 1988 1992 1996 2000 2004 2008 2012 2016 2020
```

Load Data

Read 2020 Presidential election data by county, via: <https://observablehq.com/@charliesmart/dorling-cartogram>

```
## ST GEOID NAME STATEFP state_name county_name votes_gop votes_dem
## 1 IA 19107 Keokuk 19 Iowa Keokuk County 3797 1414
## 2 IA 19189 Winnebago 19 Iowa Winnebago County 3707 2135
## 3 KS 20093 Kearny 20 Kansas Kearny County 1134 255
## 4 KS 20123 Mitchell 20 Kansas Mitchell County 2454 547
## 5 KS 20187 Stanton 20 Kansas Stanton County 607 147
## 6 KY 21005 Anderson 21 Kentucky Anderson County 9661 3348
## total_votes diff per_gop per_dem per_point_diff
## 1 5303 2383 0.7160098 0.2666415 0.4493683
## 2 5970 1572 0.6209380 0.3576214 0.2633166
## 3 1413 879 0.8025478 0.1804671 0.6220807
## 4 3039 1907 0.8075025 0.1799934 0.6275090
## 5 767 460 0.7913950 0.1916558 0.5997392
## 6 13254 6313 0.7289120 0.2526030 0.4763090
```

Read Shapefiles

US Census Bureau's County Shapefile

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/cervas/My Drive/GitHub/Data Files/GIS/Tigerline/TIGER2020PL/counties/tl_2020pl_counties.shp", layer: "COUNTY"
## with 3142 features
## It has 17 fields
## Integer64 fields read as strings: ALAND20 AWATER20
```

```
## OGR data source with driver: GeoJSON
## Source: "/Users/cervas/My Drive/GitHub/Data Files/GIS/Tigerline/TIGER2020PL/counties-cartographic/counties-cartographic.geojson", layer: "COUNTY"
## with 3234 features
## It has 12 fields
## Integer64 fields read as strings: ALAND AWATER
```

NYTs County Shapefile

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/cervas/My Drive/GitHub/Data Files/GIS/NYT/2020/counties-albers-med/counties.shp", layer: "COUNTY"
## with 3153 features
## It has 7 fields
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/cervas/My Drive/GitHub/Data Files/GIS/NYT/2020/counties-albers-med/state_labels.shp", layer: "STATE"
## with 51 features
## It has 13 fields
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/cervas/My Drive/GitHub/Data Files/GIS/NYT/2020/counties-albers-med/states.shp", layer: "STATE"
## with 51 features
## It has 7 fields
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/cervas/My Drive/GitHub/Data Files/GIS/NYT/2020/counties-albers-med/statelines.shp",
## with 107 features
## It has 2 fields
```

Coattails

Plot Coattails over time (not used)

```
## Half the Population in X Counties
pres.top.cnty <- pres.cnty.2020.decrease[cumsum(pres.cnty.2020.decrease$total)<sum(pres.cnty.2020.decrease$total)/2,]
dim(pres.top.cnty)[1] # 150 counties have half the votes
```

```
## [1] 151
```

```
## Reverse
pres.top.cnty.rev <- pres.cnty.2020.increase[cumsum(pres.cnty.2020.increase$total)<sum(pres.cnty.2020.increase$total)/2,]
dim(pres.top.cnty.rev)[1] # 3001 have the other half
```

```
## [1] 3000
```

```
sum(pres.cnty.2020.decrease$total[1:150]) ## Population of largest 150 counties
```

```
## [1] 78974022
```

```
sum(pres.cnty.2020.increase$total[1:3001]) # Population of smallest 3001 counties
```

```
## [1] 79227659
```

Compare 2016 and 2020 by county (not used)

```
a <- dplyr::full_join(pres.cnty.2016, pres.cnty.2020, by="fips")
tail(a)
counties.16.20 <- a[complete.cases(a),] # Problems with Alaska

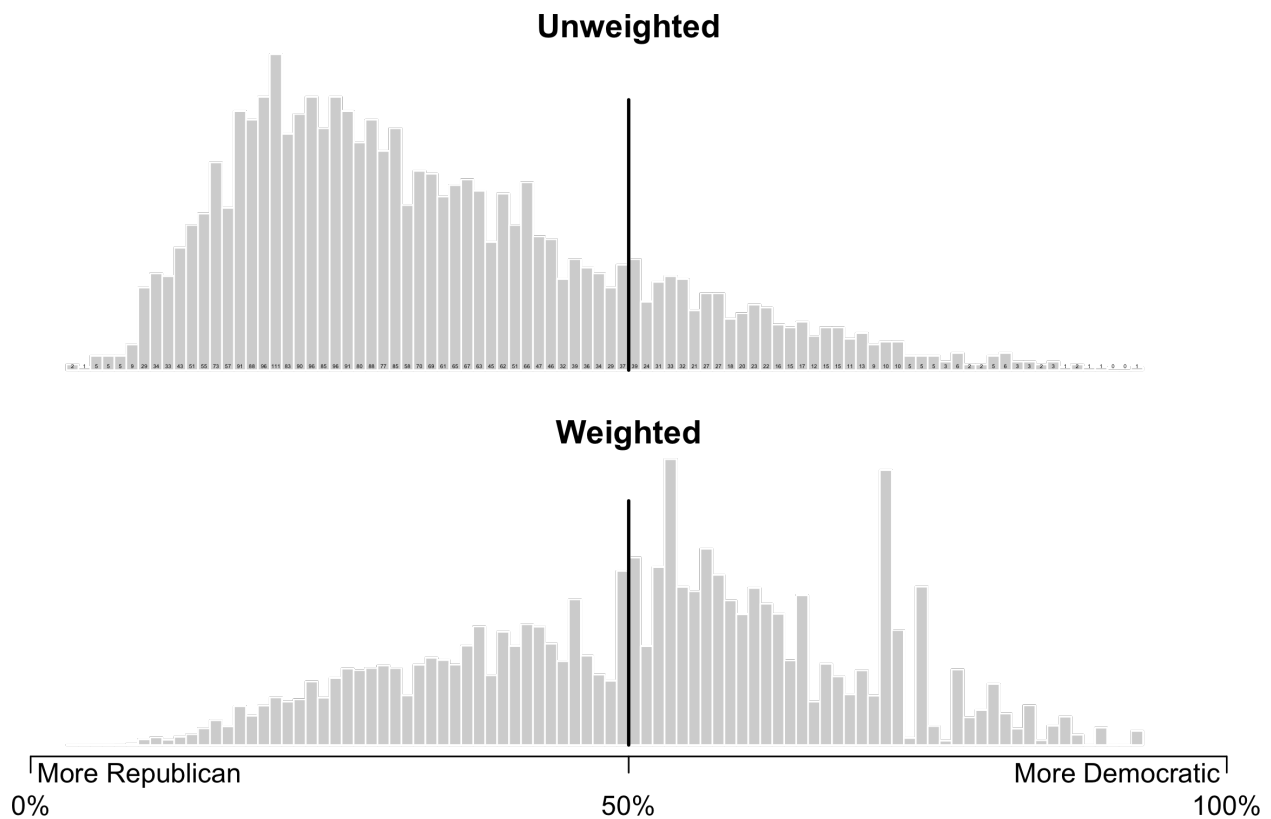
plot(
  two_party(counties.16.20$dem2016,counties.16.20$gop2016),
  two_party(counties.16.20$dem2020,counties.16.20$gop2020),
  xlab="Clinton 2016 County Vote Share",
  ylab="Biden 2020 County Vote Share",
  col="#33333333")
abline(0,1)

summary(lm(two_party(counties.16.20$dem2020,counties.16.20$gop2020) ~ two_party(counties.16.20$dem2016,counties.16.20$gop2016)))
```

This time with raw votes (not used)

```
plot(
  counties.16.20$dem2016-counties.16.20$gop2016,
  counties.16.20$dem2020-counties.16.20$gop2020,
  xlab="Clinton Advantage 2016 County Vote",
  ylab="Biden Advantage 2020 County Vote",
  col="#33333333")
abline(0,1)
```

Figure 3 - Histogram of the 2020 Presidential Election Results, by county



Biden Counties vs. Trump Counties

```
## Biden Counties
county.2020.biden <- county.2020[county.2020$votes_dem > county.2020$votes_gop,]
sum(county.2020.biden$votes_dem) # Biden Votes
```

```
## [1] 59019426
```

```
sum(county.2020.biden$votes_gop) # Trump Votes
```

```
## [1] 33564182
```

```
sum(county.2020.biden$diff) # Difference
```

```
## [1] -25455244
```

```
## Trump Counties
county.2020.trump <- county.2020[county.2020$votes_dem < county.2020$votes_gop,]
sum(county.2020.trump$votes_dem) # Biden Votes
```

```
## [1] 22245568
```

```
sum(county.2020.trump$votes_gop) # Trump Votes
```

```
## [1] 40644014
```

```
sum(county.2020.trump$diff) # Difference
```

```
## [1] 18398446
```

```
## Trump most votes, county
```

```
county.2020[order(county.2020$votes_gop, decreasing=T),][1:10,]
```

##	ST	GEOID	NAME	STATEFP	state_name	county_name	votes_gop
## 49	CA	06037	Los Angeles	6	California	Los Angeles County	1145530
## 31	AZ	04013	Maricopa	4	Arizona	Maricopa County	995665
## 1003	TX	48201	Harris	48	Texas	Harris County	700630
## 878	CA	06059	Orange	6	California	Orange County	676498
## 2993	CA	06073	San Diego	6	California	San Diego County	600094
## 364	IL	17031	Cook	17	Illinois	Cook County	558269
## 1681	FL	12086	Miami-Dade	12	Florida	Miami-Dade County	532833
## 1046	CA	06065	Riverside	6	California	Riverside County	448702
## 1207	NV	32003	Clark	32	Nevada	Clark County	430930
## 1908	TX	48439	Tarrant	48	Texas	Tarrant County	409741
##		votes_dem	total_votes	diff	per_gop	per_dem	per_point_diff
## 49		3028885	4263443	-1883355	0.2686866	0.7104317	-0.44174509
## 31		1040774	2069475	-45109	0.4811196	0.5029169	-0.02179732
## 1003		918193	1640818	-217563	0.4270004	0.5595947	-0.13259423
## 878		814009	1521725	-137511	0.4445600	0.5349252	-0.09036521
## 2993		964650	1601722	-364556	0.3746555	0.6022581	-0.22760254
## 364		1725973	2321399	-1167704	0.2404882	0.7435055	-0.50301736
## 1681		617864	1156816	-85031	0.4606031	0.5341074	-0.07350434
## 1046		527945	996156	-79243	0.4504335	0.5299823	-0.07954879
## 1207		521852	972510	-90922	0.4431111	0.5366032	-0.09349210
## 1908		411567	834697	-1826	0.4908859	0.4930735	-0.00218762

Statewide Vote

Combine 2020 data with Shapefiles

```
## Warning in rgdal::writeOGR(counties.shp, dir.gis, "us2020", driver = "ESRI
## Shapefile", : Field names abbreviated for ESRI Shapefile driver
```

```
## Registered S3 method overwritten by 'geojsonlint':
```

```
## method from
## print.location dplyr
```

Exit Polls

Table 3

exit.2016

```
##           White Black Hispanic Asian Other
## proportion_vote 0.70 0.12    0.11 0.04 0.03
## Democratic      0.37 0.89    0.66 0.65 0.56
## Republican      0.57 0.08    0.28 0.27 0.36
```

exit.2020

```
##           White Black Hispanic Asian Other
## proportion_vote 0.67 0.13    0.13 0.04 0.04
## Democratic      0.41 0.87    0.65 0.61 0.55
## Republican      0.58 0.12    0.32 0.34 0.41
```

Demographic and Election Results

```
##      White      Black  Hispanic      Asian      Other
## 5740108.0 492009.3 902017.0 437341.6 328006.2
```

```
##      White      Black  Hispanic      Asian      Other
## -105086374 -20389894 -19977976 -6020326 -6083698
```

```
##      White      Black  Hispanic      Asian      Other
## 10479386.9 4195543.6 5562236.0 870415.8 2237108.1
```

```
##      White      Black  Hispanic      Asian      Other
## 7034729.0 1159477.6 2381260.1 678615.2 1122218.2
```

```
##      White      Black  Hispanic      Asian      Other
## 8123287.3 3322116.8 3465117.2 312282.8 1189408.7
```

Table 4 - Change in Non-Hispanic White Votes between 2016 and 2020

tab4

```
##           2016      2020 Difference
## Trump           54531026 61565755    7034729
## Clinton/Biden     35397332 43520620    8123287
## Other              5740108  1061479   -4678629
## Non-Hispanic White Votes 95668466 106147853  10479387
## Minority Votes      41000771 52281778   11281007
## All Votes          136669237 158429631   21760394
```

Maps

Create Maps

Figure 4 - Choropleth Plot, 2020 Presidential Election by county

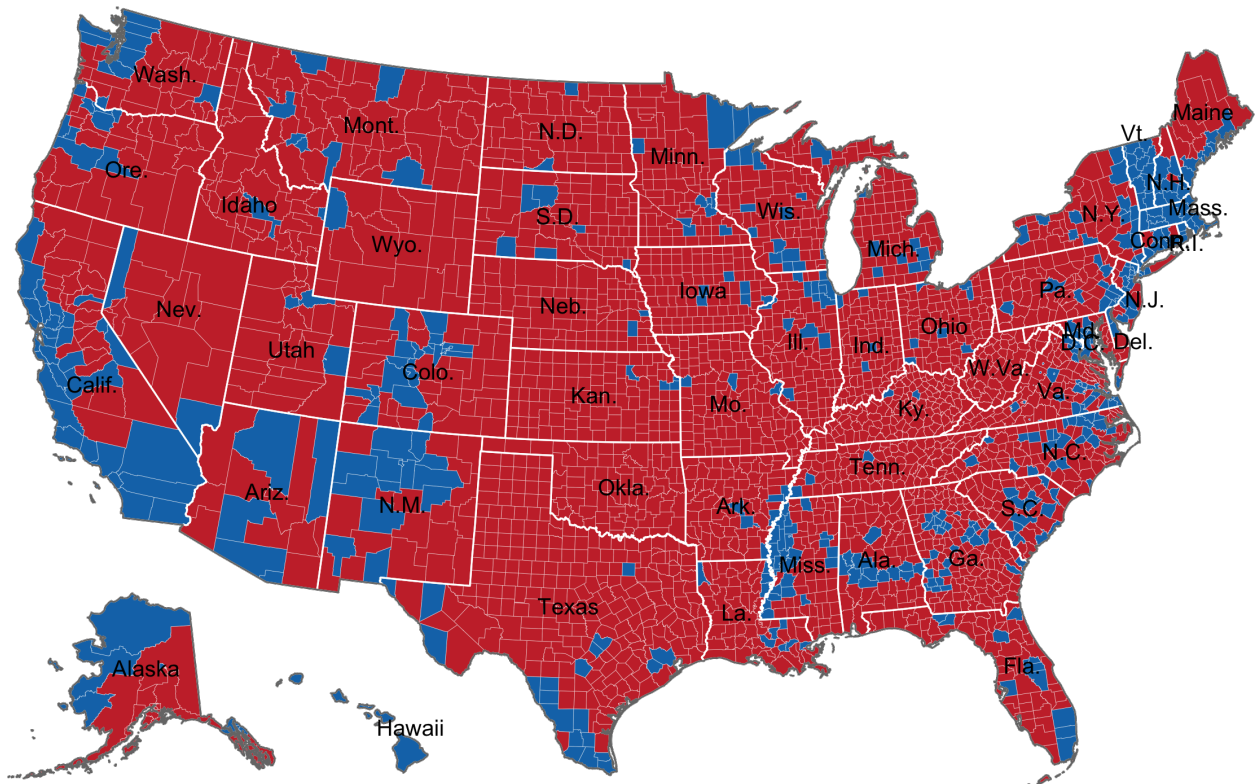
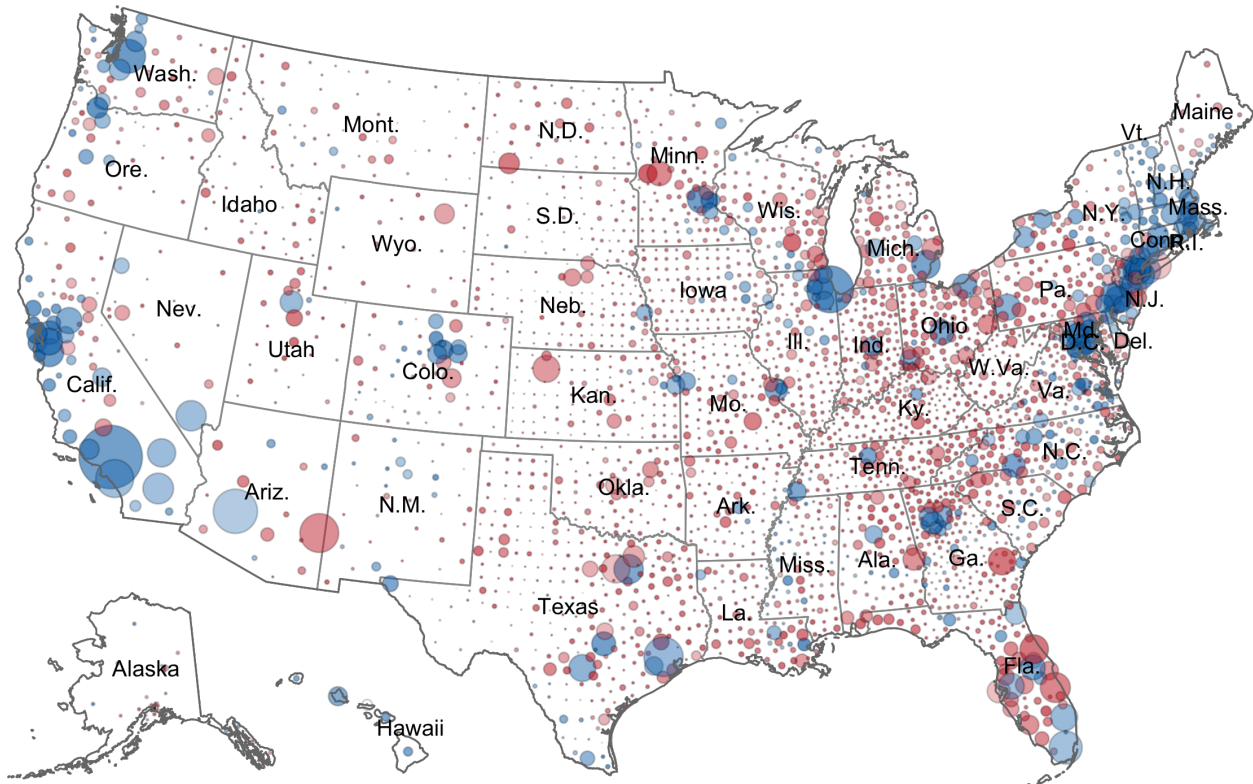


Figure 5 – Bubble Plot, 2020 Presidential Election by county

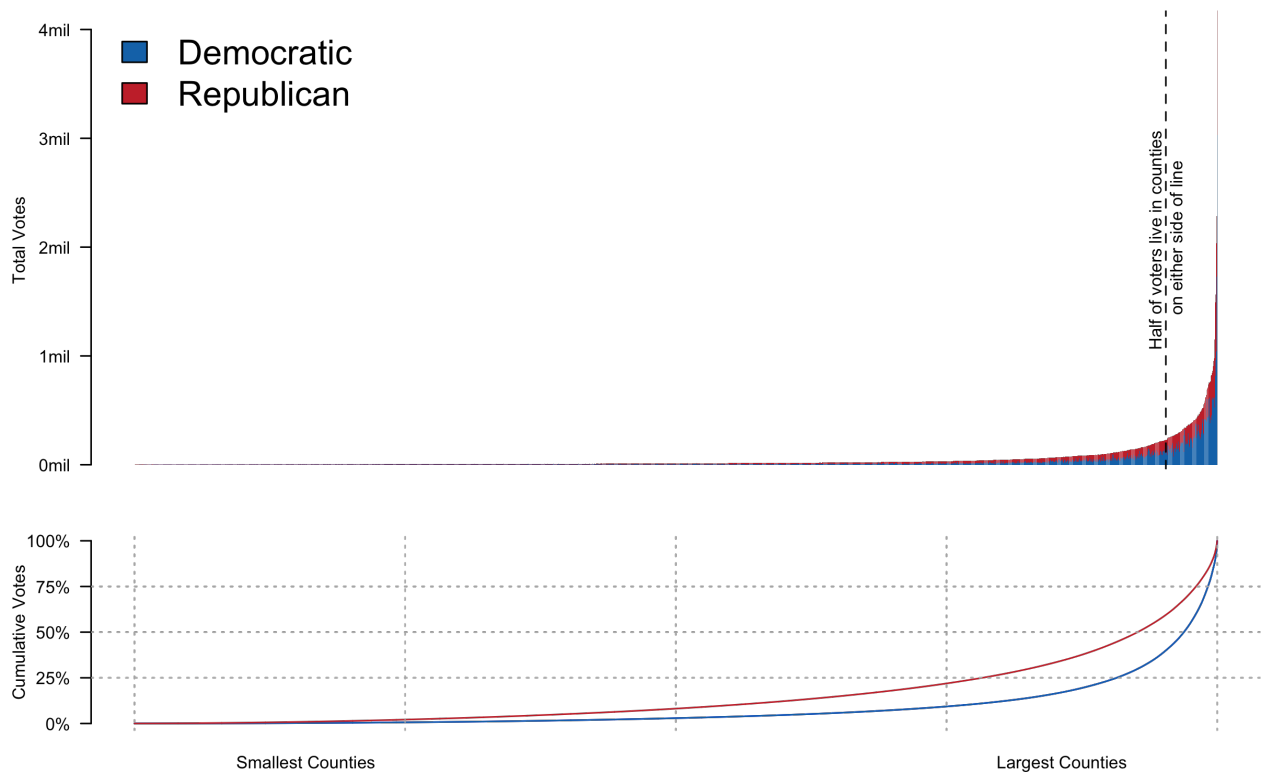


#Make Choropleth Plot in mapshaper.org

FIGURE 2A and 2B - Choropleth Plot, 2020 Presidential Election by county; Bubble Plot, 2020 Presidential Election by county

```
# mapshaper -i "/Users/cervas/My Drive/GitHub/Data Files/GIS/ NYT/counties-albers-med.json"
# -i "/Users/cervas/Downloads/county_2020.csv" string-fields=GEOID name=data
# -join target=counties data keys=GEOID,GEOID
# -each target=counties 'marginper = per_dem-0.5'
# -each target=counties 'absmargin = Math.abs(per_point_diff)'
# -each 'absmargin = Math.abs(per_point_diff)'
# -style target=counties r='Math.sqrt(total_votes) * 0.008'
# -sort absmargin descending
# -style target=counties opacity=1 fill='per_point_diff > 0 ? "#cc0000" : "#0061aa"'
# -innerlines name=counties_style
# -style target=counties_style stroke="#ddd" stroke-width=0.15
# -style target=states stroke="#000" fill=none
# -o "/Users/cervas/Downloads/us_chor.svg" target=counties,states,state_labels
# -points target=counties inner name=points
# -style opacity=0.5 fill='per_point_diff > 0 ? "#cc0000" : "#0061aa"'
# -o "/Users/cervas/Downloads/us_bubble.svg" target=points,states,state_labels
```


Figure 2 – Votes in each County



Kent County, Michigan 2020 election data plotted as Ayyadurai shows it.

```
##
## Call:
## lm(formula = I(kent$GOP_Split - kent$GOP_Straight) ~ kent$GOP_Straight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.149350 -0.039117 -0.002273  0.039378  0.128631
##
## Coefficients:
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept)    0.09535    0.01019   9.355 <0.0000000000000002 ***
## kent$GOP_Straight -0.40097    0.01961 -20.449 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05898 on 250 degrees of freedom
## Multiple R-squared:  0.6258, Adjusted R-squared:  0.6243
## F-statistic: 418.2 on 1 and 250 DF, p-value: < 0.00000000000000022
##
## Call:
## lm(formula = I(kent$DEM_Split - kent$DEM_Straight) ~ kent$DEM_Straight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -0.139628 -0.037758 0.000147 0.035940 0.149718
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)    0.270533   0.009905   27.31 <0.0000000000000002 ***
## kent$DEM_Straight -0.362538   0.018551  -19.54 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05474 on 250 degrees of freedom
## Multiple R-squared:  0.6044, Adjusted R-squared:  0.6028
## F-statistic: 381.9 on 1 and 250 DF, p-value: < 0.00000000000000022

##
## Call:
## lm(formula = kent$GOP_Split ~ kent$GOP_Straight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.149350 -0.039117 -0.002273  0.039378  0.128631
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)    0.09535   0.01019   9.355 <0.0000000000000002 ***
## kent$GOP_Straight 0.59903   0.01961  30.549 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05898 on 250 degrees of freedom
## Multiple R-squared:  0.7887, Adjusted R-squared:  0.7879
## F-statistic: 933.2 on 1 and 250 DF, p-value: < 0.00000000000000022

##
## Call:
## lm(formula = kent$DEM_Split ~ kent$DEM_Straight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.139628 -0.037758  0.000147  0.035940  0.149718
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)    0.270533   0.009905   27.31 <0.0000000000000002 ***
## kent$DEM_Straight 0.637462   0.018551   34.36 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05474 on 250 degrees of freedom
## Multiple R-squared:  0.8253, Adjusted R-squared:  0.8246
## F-statistic: 1181 on 1 and 250 DF, p-value: < 0.00000000000000022

```

Figure 6 – Kent County, Michigan 2020 election data plotted as Ayyadurai shows it.

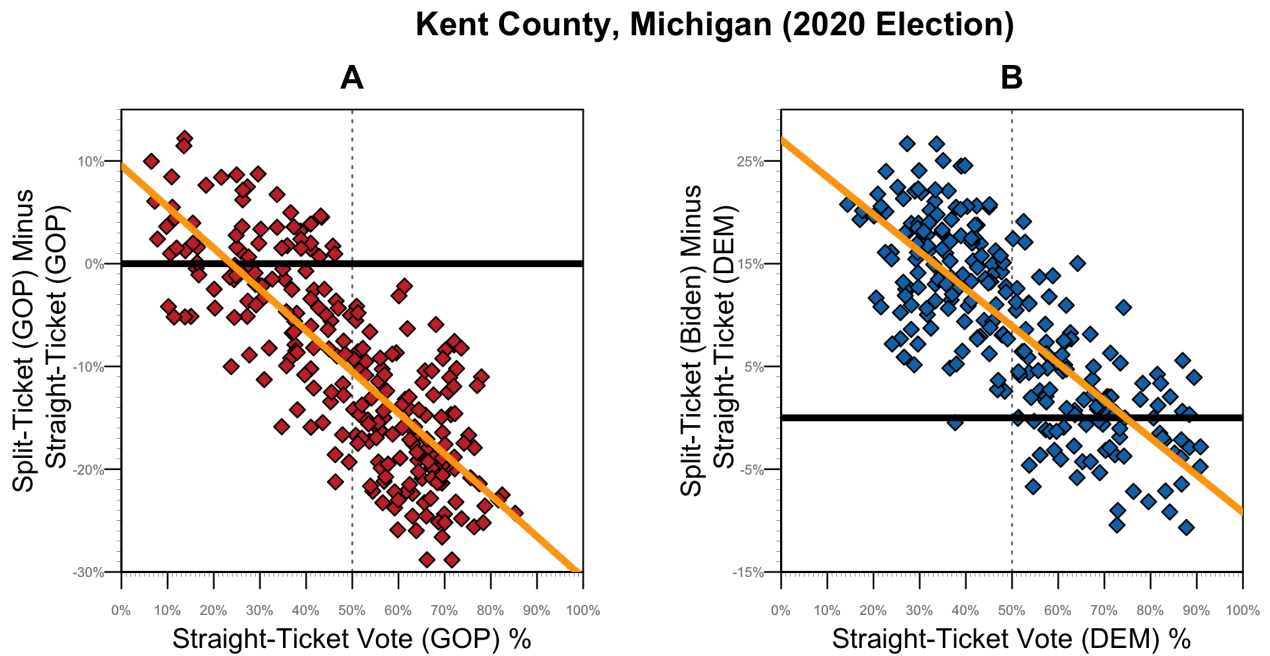
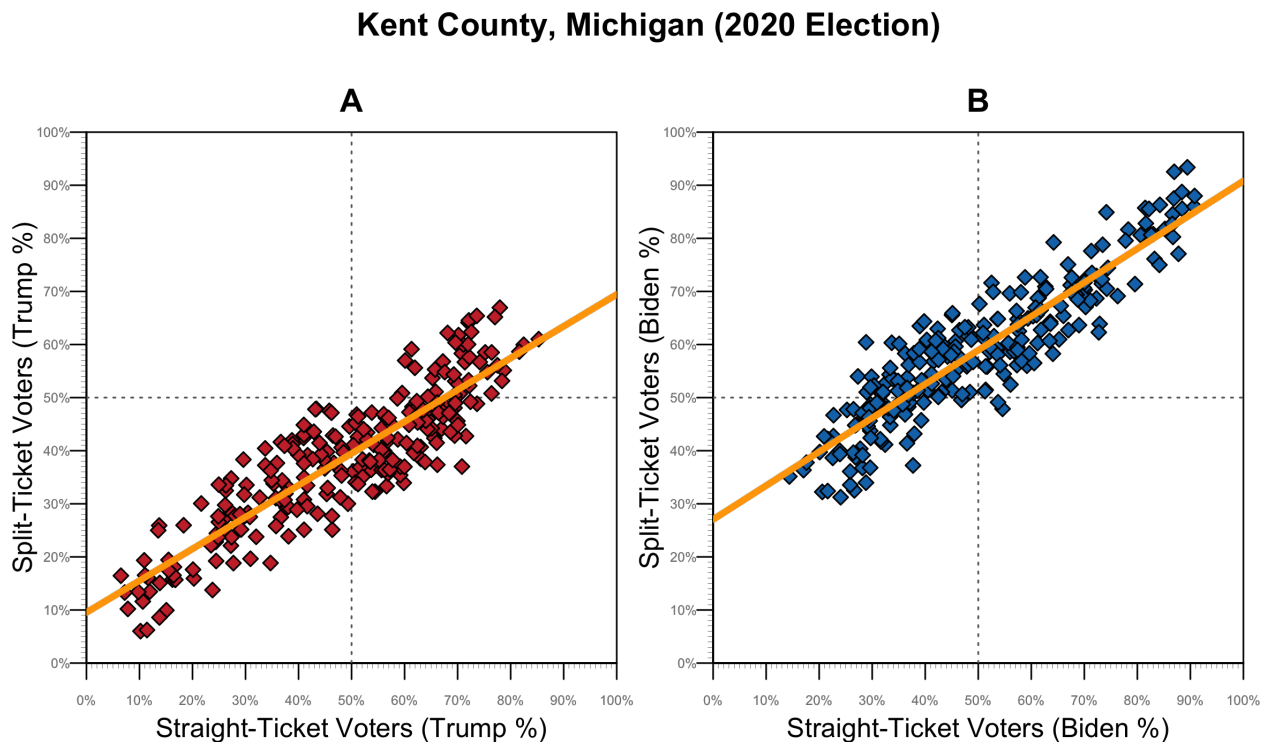


Figure 7 – Kent County, Michigan Precinct comparison between Trump Straight-ticket and Trump Split-Ticket Support



Birthday Paradox

- Original Problem: How many people do you need in order for the probability that at least two people have the same birthday to exceed 0.5?

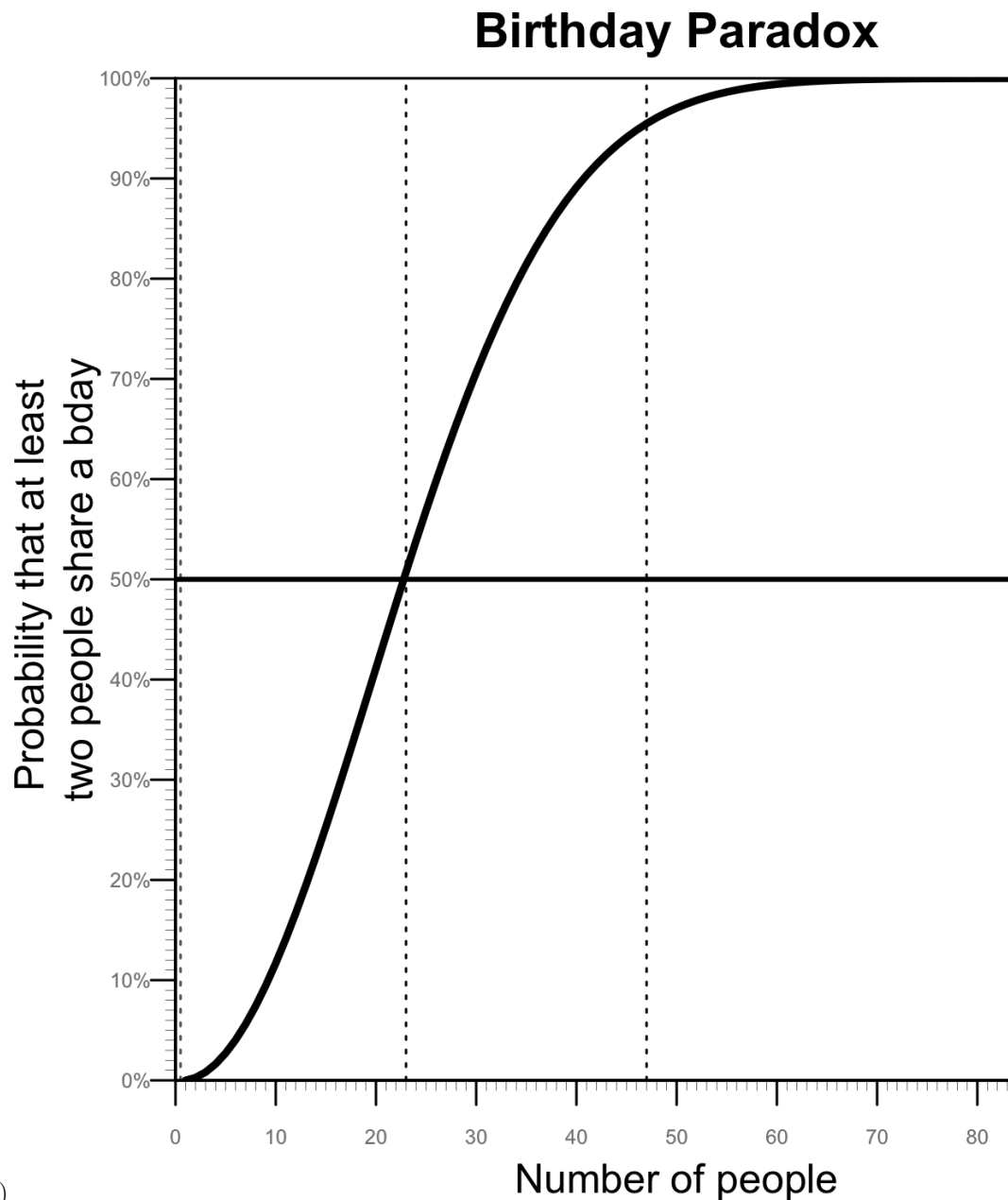
– Derivation for the original question:

$$1 - P(\text{everyone has different birthday})$$

$$= 1 - \frac{365 P_k}{365^k} = 1 - \frac{365!}{365^k (365 - k)!}$$

```
c(bday[10], bday[23], bday[68])
```

```
##          10          23          68
## 0.1169482 0.5072972 0.9987264
```



Birthday Paradox Plot (not used)