

# Politics Are Afoot!

Da Qi Ren

## The Setup

There is *a lot* of money that is spent in politics in Presidential election years. So far, estimates have the number at about \$11,000,000,000 (11 billion USD). For context, in 2019 Twitter's annual revenue was about \$3,500,000,000 (3.5 billion USD).

## The work

Install the package, `fec16`.

```
## install.packages('fec16')
```

This package is a compendium of spending and results from the 2016 election cycle. In this dataset are 9 different datasets that cover:

- **candidates:** candidate attributes, like their name, a unique id of the candidate, the election year under consideration, the office they're running for, etc.
- **results\_house:** race attributes, like the name of the candidates running in the election, a unique id of the candidate, the number of **general\_votes** garnered by each candidate, and other information.
- **campaigns:** financial information for each house & senate campaign. This includes a unique candidate id, the total receipts (how much came in the doors), and total disbursements (the total spent by the campaign), the total contributed by party central committees, and other information.

## Your task

Describe the relationship between spending on a candidate's behalf and the votes they receive.

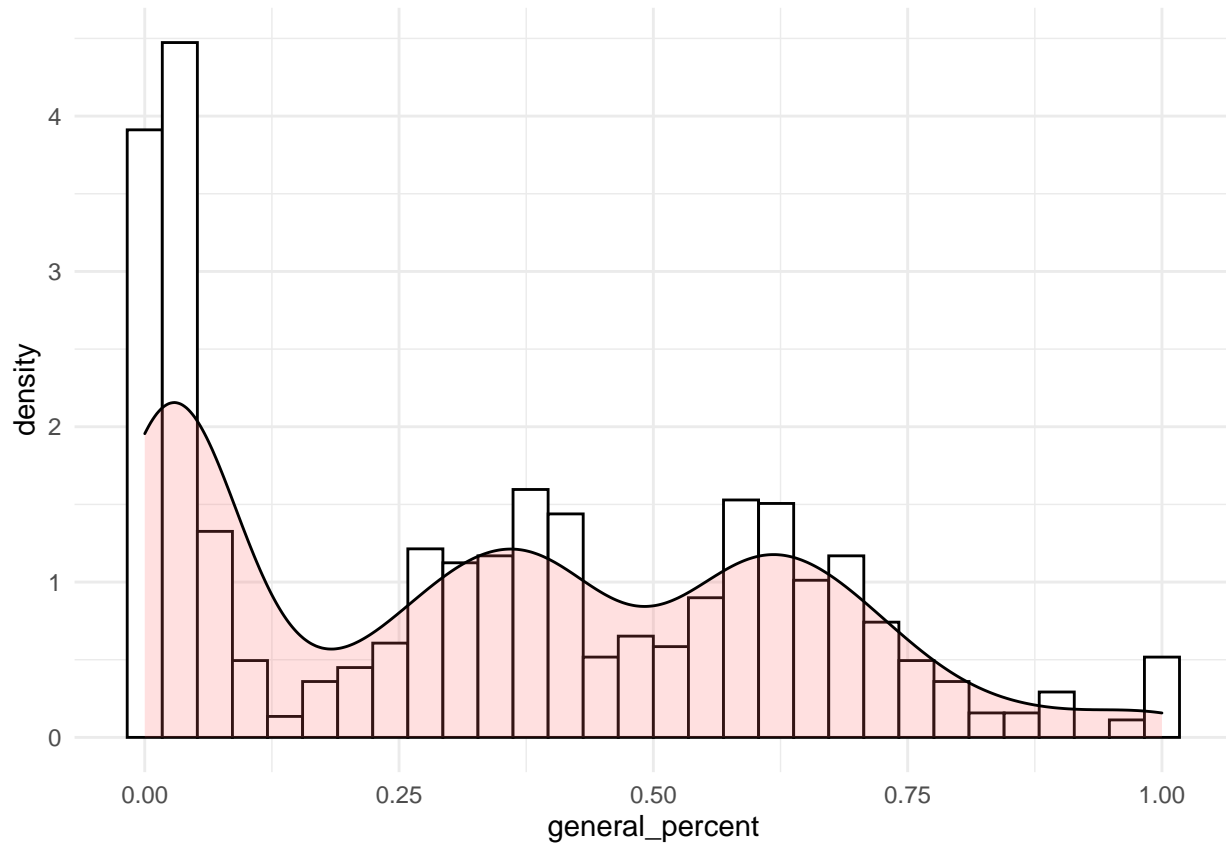
## Your work

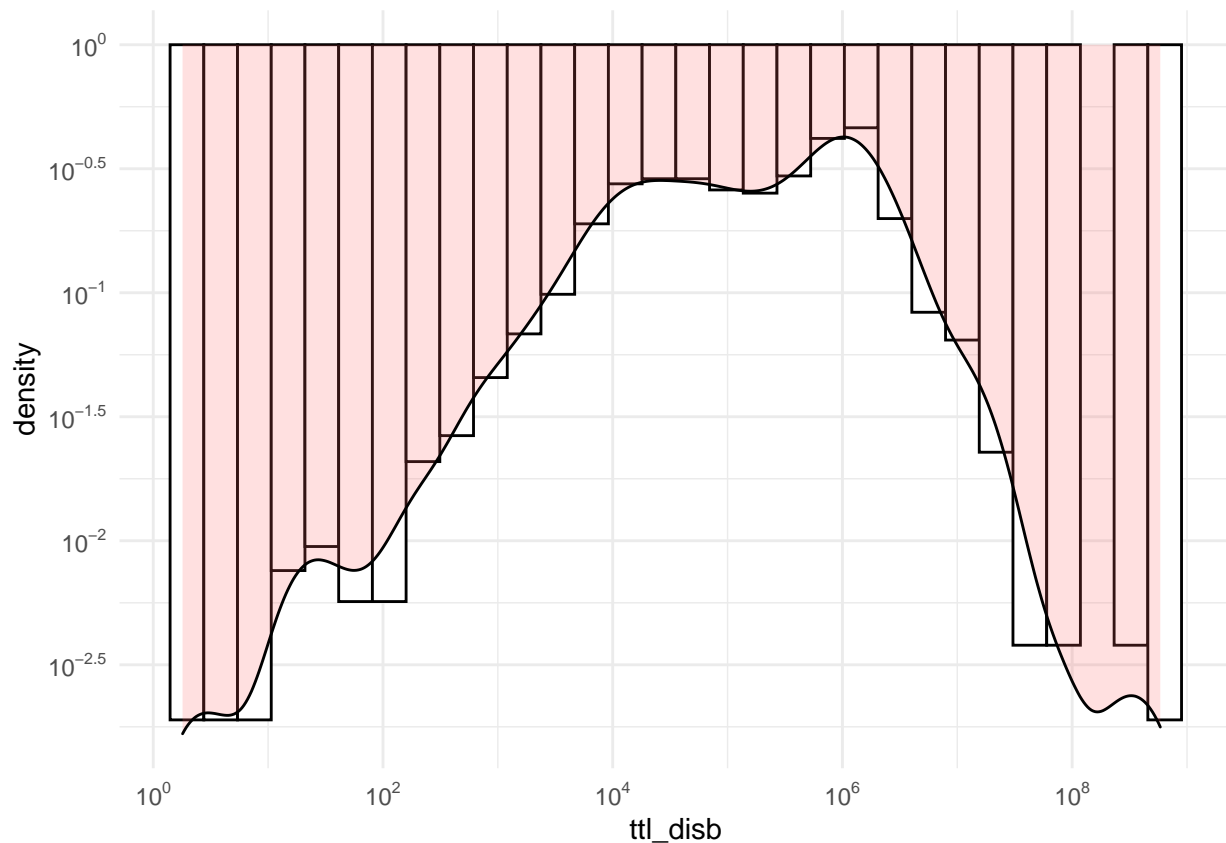
- We want to keep this work *relatively* constrained, which is why we're providing you with data through the `fec16` package. It is possible to gather all the information from current FEC reports, but it would require you to make a series of API calls that would pull us away from the core modeling tasks that we want you to focus on instead.
- Throughout this assignment, limit yourself to functions that are within the **tidyverse** family of packages: `dplyr`, `ggplot`, `patchwork`, and `magrittr` for wrangling and exploration and `base`, `stats`, `sandwich` and `lmtest` for modeling and testing. You do not *have* to use these packages; but try to limit yourself to using only these.

```
candidates <- fec16::candidates
results_house <- fec16::results_house
campaigns <- fec16::campaigns
```

## 1. What does the distribution of votes and of spending look like?

- (3 points) In separate histograms, show both the distribution of votes (measured in `results_house$general_percent` for now) and spending (measured in `t11_disb`). Use a log transform if appropriate for each visualization. How would you describe what you see in these two plots?





## 2. Exploring the relationship between spending and votes.

2. (3 points) Create a new dataframe by joining `results_house` and `campaigns` using the `inner_join` function from `dplyr`. (We use the format `package::function` – so `dplyr::inner_join`.)

```
nrow(results_house)
```

```
## [1] 2110
```

```
nrow(campaigns)
```

```
## [1] 1898
```

```
d1 <- inner_join(results_house, campaigns, by = NULL)
```

```
## Joining, by = "cand_id"
```

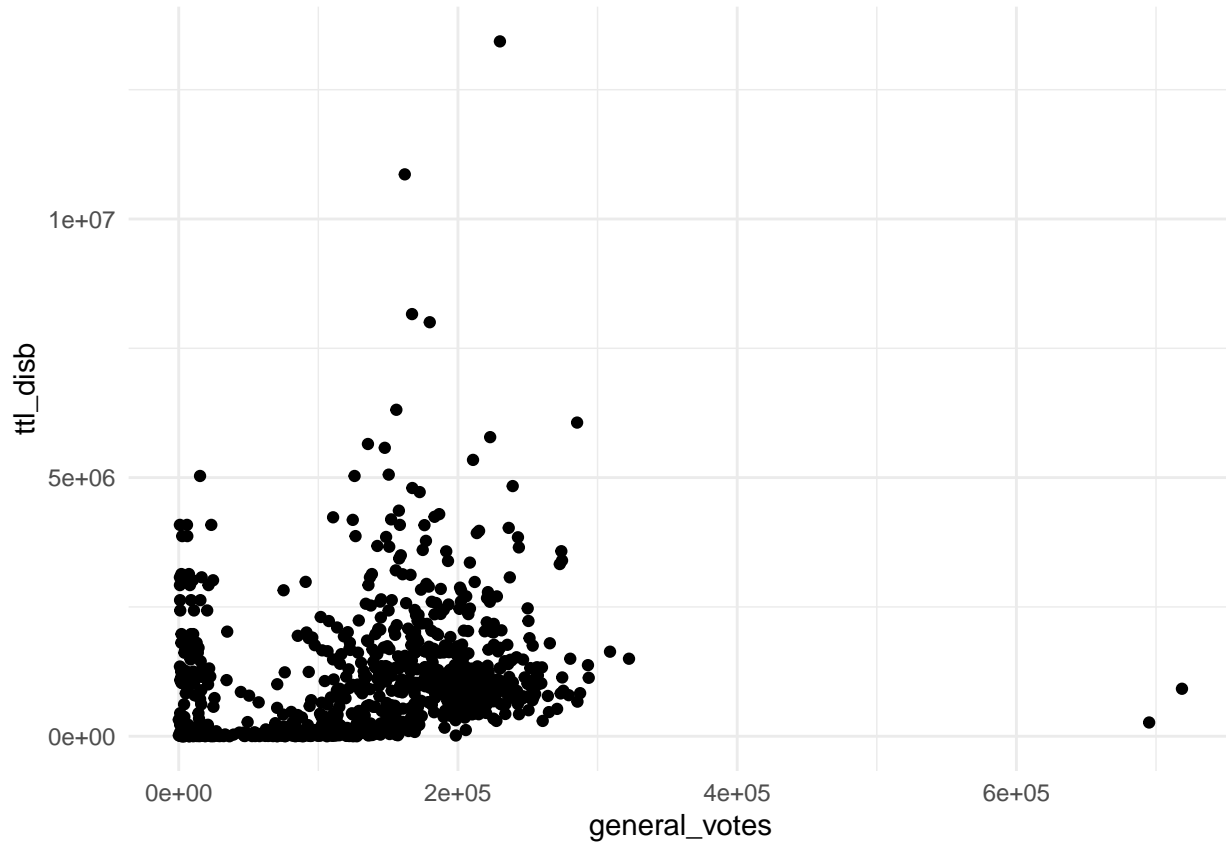
```
nrow(d1)
```

```
## [1] 1342
```

3. (3 points) Produce a scatter plot of `general_votes` on the y-axis and `ttl_disb` on the x-axis. What do you observe about the shape of the joint distribution?

```
ggplot(d1, aes(x=general_votes, y=ttl_disb)) + geom_point()
```

```
## Warning: Removed 462 rows containing missing values (geom_point).
```

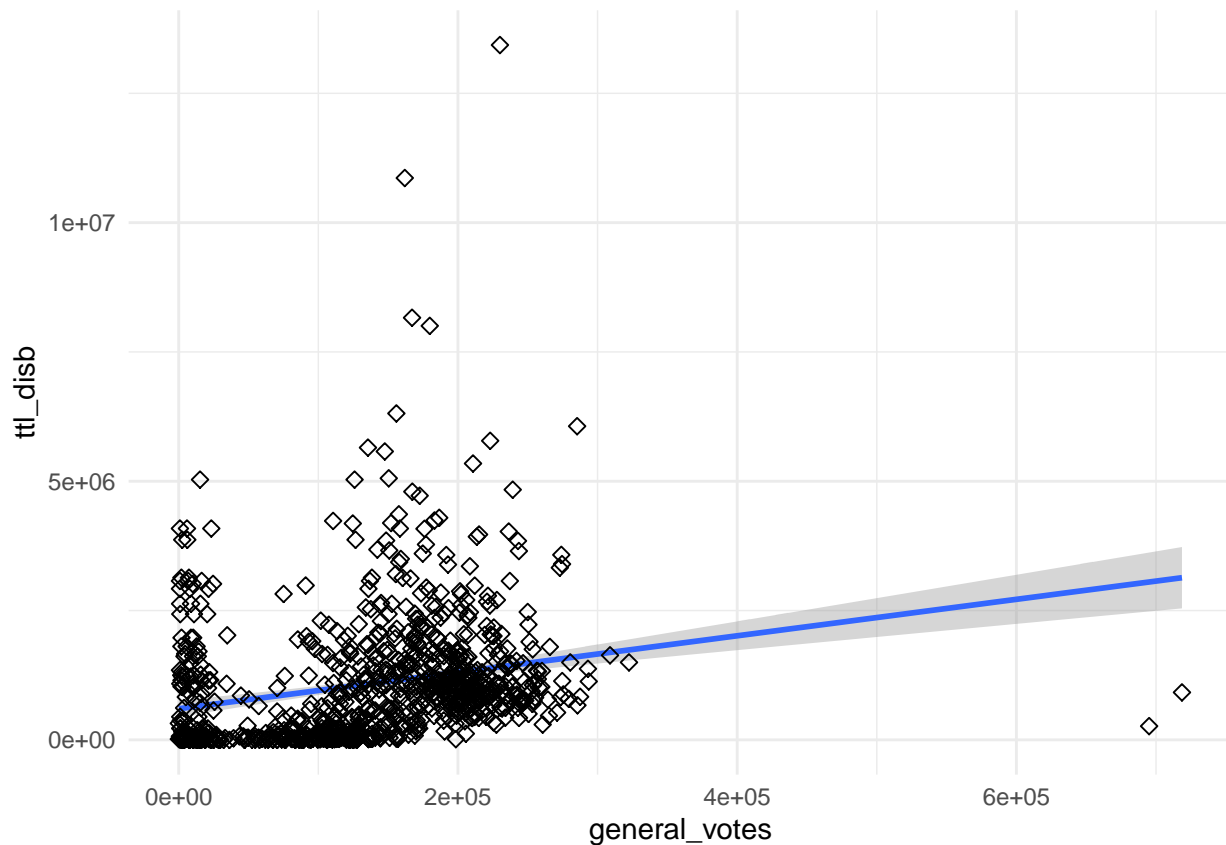


```
sp <- ggplot(d1, aes(x=general_votes, y=ttl_disb )) +  
  geom_smooth(method=lm)+  
  geom_point(size=2, shape=23)  
  
sp
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

```
## Warning: Removed 462 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 462 rows containing missing values (geom_point).
```



4. (3 points) Create a new variable to indicate whether each individual is a “Democrat”, “Republican” or “Other Party”.

- Here’s an example of how you might use `mutate` and `case_when` together to create a variable.

```
starwars %>%
  select(name:mass, gender, species) %>%
  mutate(
    type = case_when(
      height > 200 | mass > 200 ~ "large",
      species == "Droid" ~ "robot",
      TRUE ~ "other"
    )
  )
```

Once you’ve produced the new variable, plot your scatter plot again, but this time adding an argument into the `aes()` function that colors the points by party membership. What do you observe about the distribution of all three variables?

```
d2<-d1 %>%
  dplyr::select(cand_pty_affiliation, general_votes, ttl_disb, state) %>%
  na.omit() %>%
  mutate(
    can_party = case_when(
      cand_pty_affiliation=="REP" ~ "REP",
```

```

    cand_pty_affiliation=="DEM" ~ "DEM",
    TRUE ~ "Other"
  )
)

d2<-d2 %>% dplyr::select(can_party, general_votes, ttl_disb, state)

```

```

#Y = d2$general_votes
library(MASS)

```

```

##
## Attaching package: 'MASS'

## The following object is masked from 'package:patchwork':
##
##   area

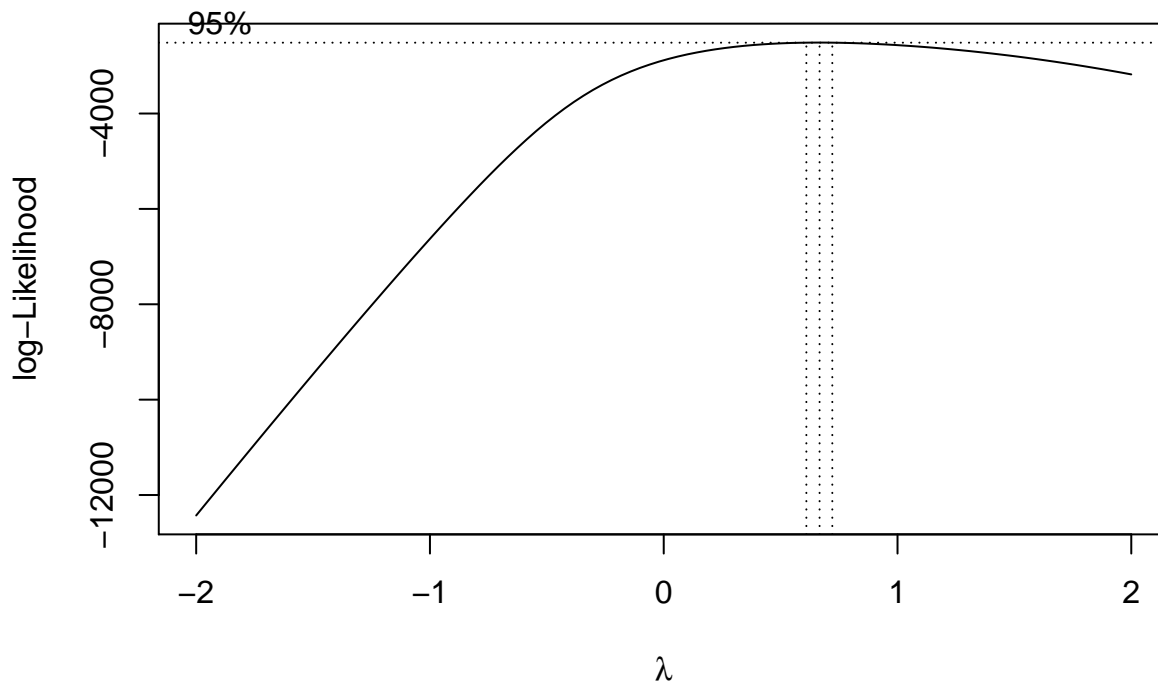
## The following object is masked from 'package:dplyr':
##
##   select

```

```

b <- boxcox(general_votes ~ ttl_disb + state + can_party, data = d2)

```



```

b

```

```

## $x
## [1] -2.00000000 -1.95959596 -1.91919192 -1.87878788 -1.83838384 -1.79797980
## [7] -1.75757576 -1.71717172 -1.67676768 -1.63636364 -1.59595960 -1.55555556
## [13] -1.51515152 -1.47474747 -1.43434343 -1.39393939 -1.35353535 -1.31313131

```

```
## [19] -1.27272727 -1.23232323 -1.19191919 -1.15151515 -1.11111111 -1.07070707
## [25] -1.03030303 -0.98989899 -0.94949495 -0.90909091 -0.86868687 -0.82828283
## [31] -0.78787879 -0.74747475 -0.70707071 -0.66666667 -0.62626263 -0.58585859
## [37] -0.54545455 -0.50505051 -0.46464646 -0.42424242 -0.38383838 -0.34343434
## [43] -0.30303030 -0.26262626 -0.22222222 -0.18181818 -0.14141414 -0.10101010
## [49] -0.06060606 -0.02020202 0.02020202 0.06060606 0.10101010 0.14141414
## [55] 0.18181818 0.22222222 0.26262626 0.30303030 0.34343434 0.38383838
## [61] 0.42424242 0.46464646 0.50505051 0.54545455 0.58585859 0.62626263
## [67] 0.66666667 0.70707071 0.74747475 0.78787879 0.82828283 0.86868687
## [73] 0.90909091 0.94949495 0.98989899 1.03030303 1.07070707 1.11111111
## [79] 1.15151515 1.19191919 1.23232323 1.27272727 1.31313131 1.35353535
## [85] 1.39393939 1.43434343 1.47474747 1.51515152 1.55555556 1.59595960
## [91] 1.63636364 1.67676768 1.71717172 1.75757576 1.79797980 1.83838384
## [97] 1.87878788 1.91919192 1.95959596 2.00000000
##
## $y
## [1] -12428.469 -12186.090 -11944.150 -11702.670 -11461.669 -11221.167
## [7] -10981.185 -10741.746 -10502.875 -10264.601 -10026.952 -9789.962
## [13] -9553.666 -9318.104 -9083.319 -8849.362 -8616.285 -8384.153
## [19] -8153.033 -7923.007 -7694.166 -7466.614 -7240.478 -7015.893
## [25] -6793.031 -6572.083 -6353.272 -6136.876 -5923.190 -5712.589
## [31] -5505.491 -5302.367 -5103.782 -4910.323 -4722.659 -4541.495
## [37] -4367.527 -4201.440 -4043.886 -3895.341 -3756.210 -3626.712
## [43] -3506.825 -3396.466 -3295.307 -3202.940 -3118.898 -3042.639
## [49] -2973.625 -2911.333 -2855.251 -2804.907 -2759.869 -2719.724
## [55] -2684.115 -2652.701 -2625.173 -2601.255 -2580.683 -2563.225
## [61] -2548.663 -2536.796 -2527.446 -2520.441 -2515.634 -2512.880
## [67] -2512.054 -2513.041 -2515.733 -2520.037 -2525.865 -2533.138
## [73] -2541.789 -2551.751 -2562.971 -2575.398 -2588.988 -2603.703
## [79] -2619.508 -2636.377 -2654.282 -2673.204 -2693.127 -2714.035
## [85] -2735.920 -2758.771 -2782.584 -2807.356 -2833.085 -2859.770
## [91] -2887.413 -2916.015 -2945.580 -2976.110 -3007.608 -3040.078
## [97] -3073.521 -3107.938 -3143.332 -3179.703
```

```
lambda <- b$x
lik <- b$y
bc <- cbind(lambda, lik)
bc[order(~lik),]
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type 'language'
```

```
##      lambda      lik
## [1,] -2.000000 -12428.47
## [2,] -1.959596 -12186.09
```

```
lambda <- 0.67
d2$lamvotes <- (d2$general_votes^lambda-1)/lambda

m1 <- lm(lamvotes ~ ttl_disb + state + can_party, data = d2)
summary(m1)
```

```
##
```

```

## Call:
## lm(formula = lamvotes ~ ttl_disb + state + can_party, data = d2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6487.0  -871.4   -33.6    798.5   4647.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.648e+03  8.976e+02   4.065 5.27e-05 ***
## ttl_disb      3.178e-04  3.618e-05   8.785 < 2e-16 ***
## stateAL       7.792e+02  9.739e+02   0.800  0.42388
## stateAR       9.543e+02  1.060e+03   0.900  0.36838
## stateAS      -3.322e+03  1.157e+03  -2.871  0.00419 **
## stateAZ       1.806e+02  9.535e+02   0.189  0.84985
## stateCA      -7.052e+01  9.046e+02  -0.078  0.93788
## stateCO       9.369e+02  9.575e+02   0.978  0.32814
## stateCT      -9.164e+02  9.536e+02  -0.961  0.33686
## stateDC       2.624e+03  1.552e+03   1.691  0.09126 .
## stateDE       1.440e+03  1.267e+03   1.137  0.25598
## stateFL       4.873e+02  9.114e+02   0.535  0.59303
## stateGA       1.046e+03  9.444e+02   1.108  0.26829
## stateGU      -2.682e+03  1.267e+03  -2.116  0.03462 *
## stateHI       4.567e+02  1.157e+03   0.395  0.69322
## stateIA       9.324e+02  1.001e+03   0.931  0.35210
## stateID       9.668e+02  1.098e+03   0.880  0.37893
## stateIL       7.593e+02  9.245e+02   0.821  0.41171
## stateIN       4.468e+02  9.475e+02   0.472  0.63735
## stateKS      -3.782e+02  9.905e+02  -0.382  0.70266
## stateKY       9.407e+02  9.739e+02   0.966  0.33438
## stateLA      -1.339e+03  9.420e+02  -1.421  0.15567
## stateMA       1.355e+03  9.509e+02   1.425  0.15459
## stateMD       8.252e+02  9.576e+02   0.862  0.38910
## stateME       8.248e+02  1.097e+03   0.752  0.45241
## stateMI       5.628e+02  9.266e+02   0.607  0.54377
## stateMN       1.244e+03  9.555e+02   1.302  0.19343
## stateMO       1.273e+03  9.626e+02   1.323  0.18628
## stateMP      -1.863e+02  1.564e+03  -0.119  0.90521
## stateMS       6.250e+02  1.035e+03   0.604  0.54595
## stateMT       1.041e+03  1.272e+03   0.819  0.41330
## stateNC       9.988e+02  9.310e+02   1.073  0.28363
## stateND       4.843e+02  1.158e+03   0.418  0.67597
## stateNE       7.490e+02  1.097e+03   0.683  0.49500
## stateNH      -4.540e+02  1.061e+03  -0.428  0.66874
## stateNJ       3.003e+02  9.357e+02   0.321  0.74832
## stateNM       9.043e+01  1.060e+03   0.085  0.93202
## stateNV      -1.676e+02  9.904e+02  -0.169  0.86568
## stateNY      -2.202e+03  9.032e+02  -2.438  0.01499 *
## stateOH       7.680e+02  9.253e+02   0.830  0.40680
## stateOK       8.868e+02  1.060e+03   0.837  0.40310
## stateOR       1.549e+03  1.016e+03   1.525  0.12767
## statePA       1.034e+03  9.260e+02   1.116  0.26469
## statePR       6.674e+03  1.164e+03   5.735 1.37e-08 ***
## stateRI      -4.742e+02  1.097e+03  -0.432  0.66576

```



```
## stateSC      -6.127e+02  9.445e+02  -0.649  0.51670
## stateSD      7.991e+02  1.267e+03   0.631  0.52834
## stateTN      4.779e+02  9.579e+02   0.499  0.61801
## stateTX      1.757e+02  9.121e+02   0.193  0.84733
## stateUT     -1.187e+02  1.001e+03  -0.119  0.90568
## stateVA      8.729e+02  9.374e+02   0.931  0.35199
## stateVI     -2.840e+03  1.552e+03  -1.830  0.06768 .
## stateVT      2.513e+03  1.552e+03   1.619  0.10585
## stateWA      5.526e+02  9.501e+02   0.582  0.56097
## stateWI      7.334e+02  9.476e+02   0.774  0.43922
## stateWV     -2.422e+02  1.016e+03  -0.238  0.81167
## stateWY     -3.161e+02  1.158e+03  -0.273  0.78496
## can_partyOther -2.738e+03  2.023e+02 -13.536 < 2e-16 ***
## can_partyREP  -1.582e+00  9.055e+01  -0.017  0.98607
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1267 on 821 degrees of freedom
## Multiple R-squared:  0.5346, Adjusted R-squared:  0.5017
## F-statistic: 16.26 on 58 and 821 DF,  p-value: < 2.2e-16
```

```
#d2$state <- as.numeric(d2$state)
#d2$can_party <- as.numeric(d2$can_party)

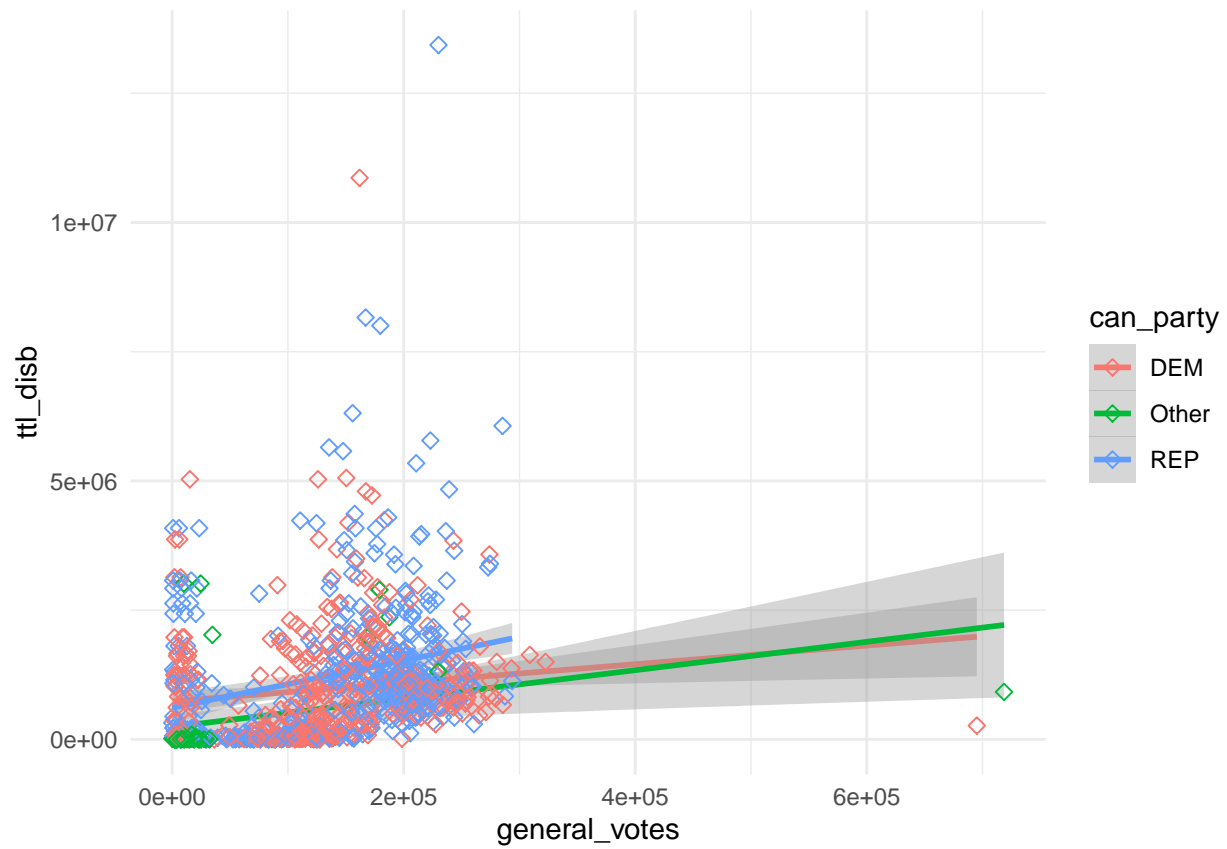
#d2

#write.csv(d2, "d2.csv")

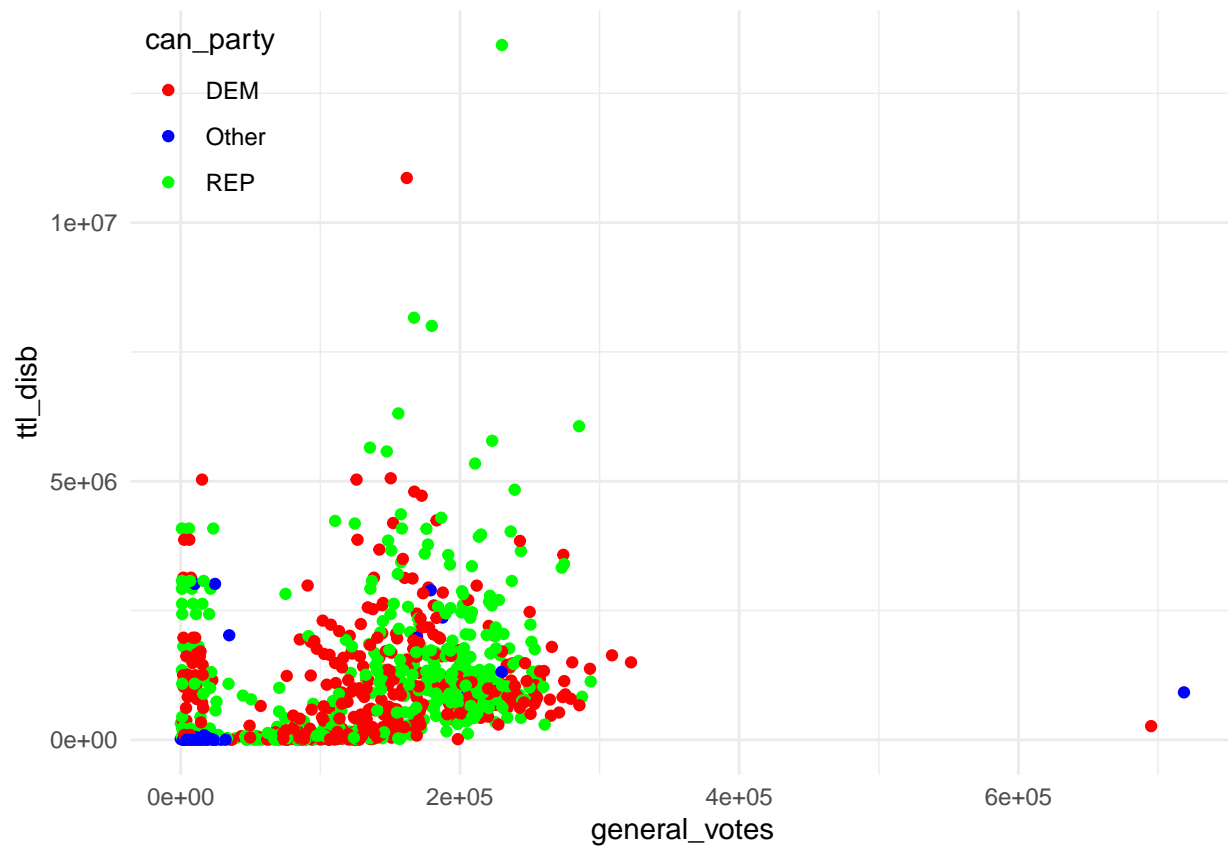
#head(d2)

sp <- ggplot(d2, aes(x=general_votes, y=ttl_disb, color=can_party)) +
  geom_smooth(method=lm)+
  geom_point(size=2, shape=23)
sp
```

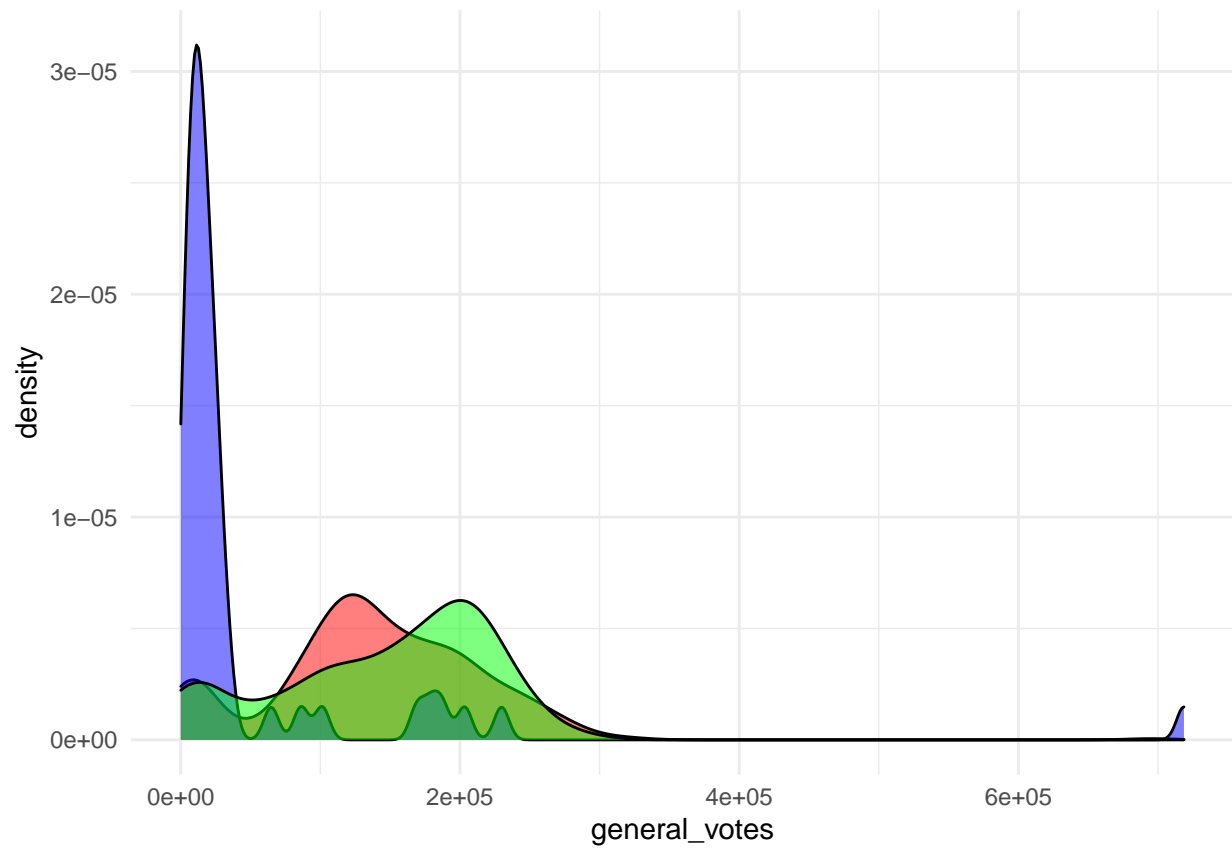
```
## 'geom_smooth()' using formula 'y ~ x'
```



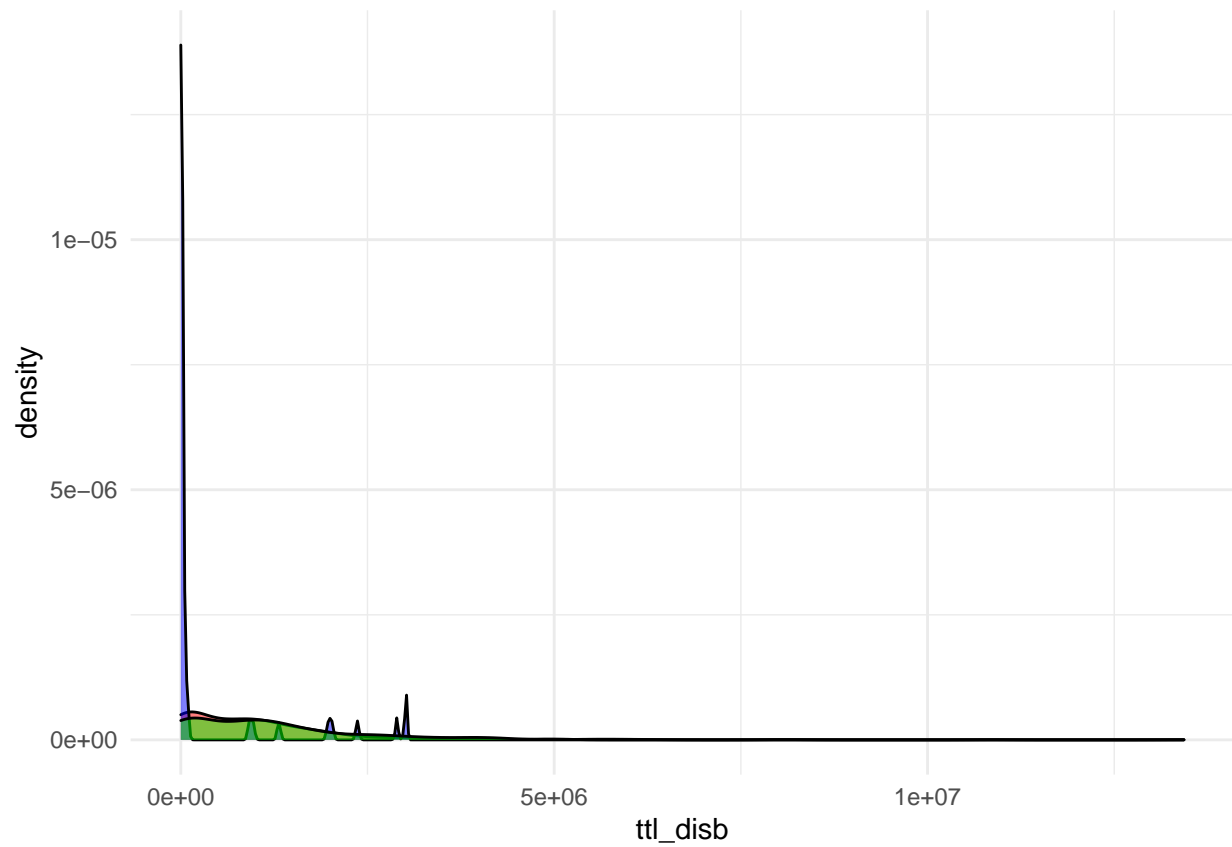
```
p1<-ggplot(d2, aes(x=general_votes, y=ttl_disb, color=can_party)) +
  geom_point() +
  scale_color_manual(values = c("red", "blue", "green")) +
  theme(legend.position=c(0,1), legend.justification=c(0,1))
p1
```



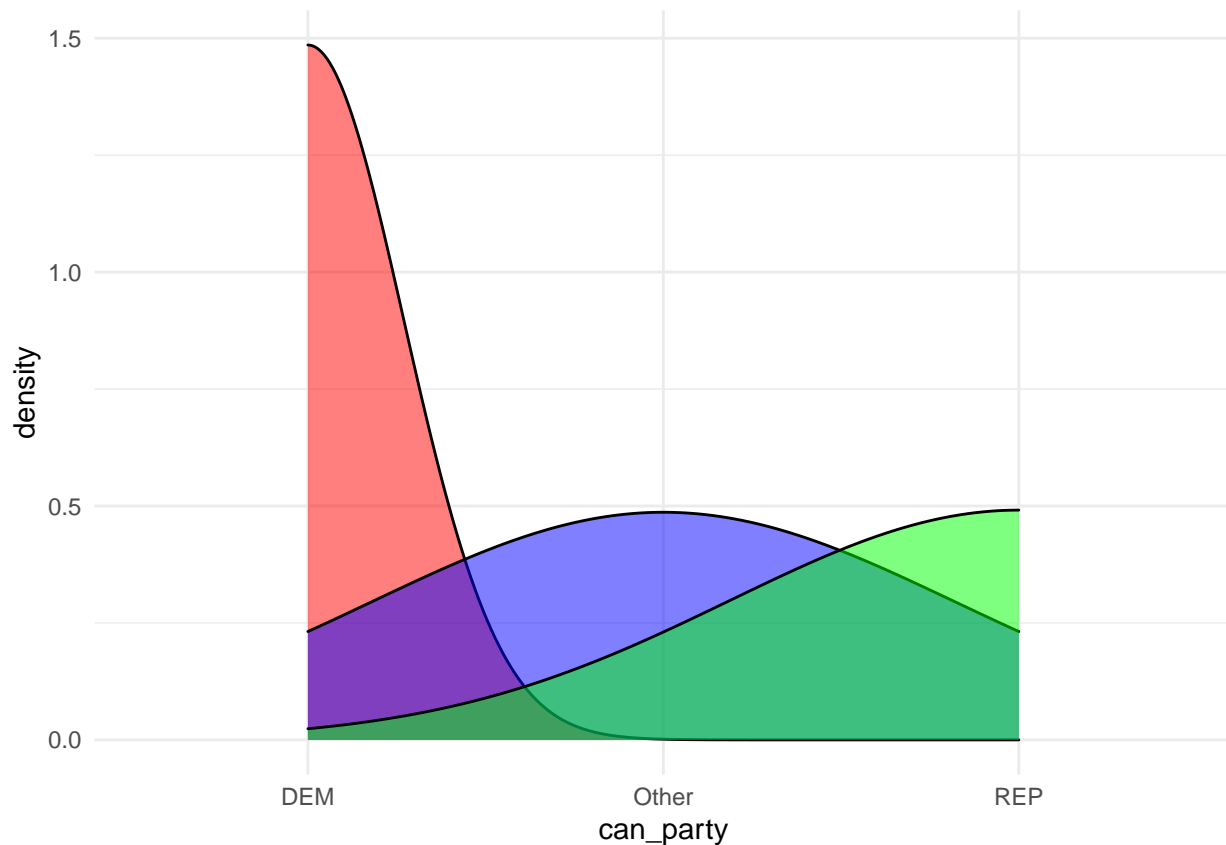
```
p2<-ggplot(d2, aes(x=general_votes, fill=can_party)) +  
  geom_density(alpha=.5) +  
  scale_fill_manual(values = c("red", "blue", "green")) +  
  theme(legend.position = "none")  
p2
```



```
# Marginal density plot of y (right panel)
p3<-ggplot(d2, aes(x=ttl_disb, fill=can_party)) +
  geom_density(alpha=.5) +
  scale_fill_manual(values = c("red", "blue", "green")) +
  theme(legend.position = "none")
p3
```



```
p3<-ggplot(d2, aes(x=can_party, fill=can_party)) +
  geom_density(alpha=.5) +
  scale_fill_manual(values = c("red", "blue", "green")) +
  theme(legend.position = "none")
p3
```



```
#sp + geom_density_2d()
```

```
#summary(d1)
```

## Produce a Descriptive Model

5. (5 Points) Given your observations, produce a linear model that you think does a good job at describing the relationship between candidate spending and votes they receive. You should decide what transformation to apply to spending (if any), what transformation to apply to votes (if any) and also how to include the party affiliation.

```
d2[d2 == -Inf] <- 0
```

```
sdat <- d2[, c("general_votes", "ttl_disb")]
```

```
imp <- preProcess(sdat, method = c("knnImpute"), k = 5)
```

```
sdat <- predict(imp, sdat)
```

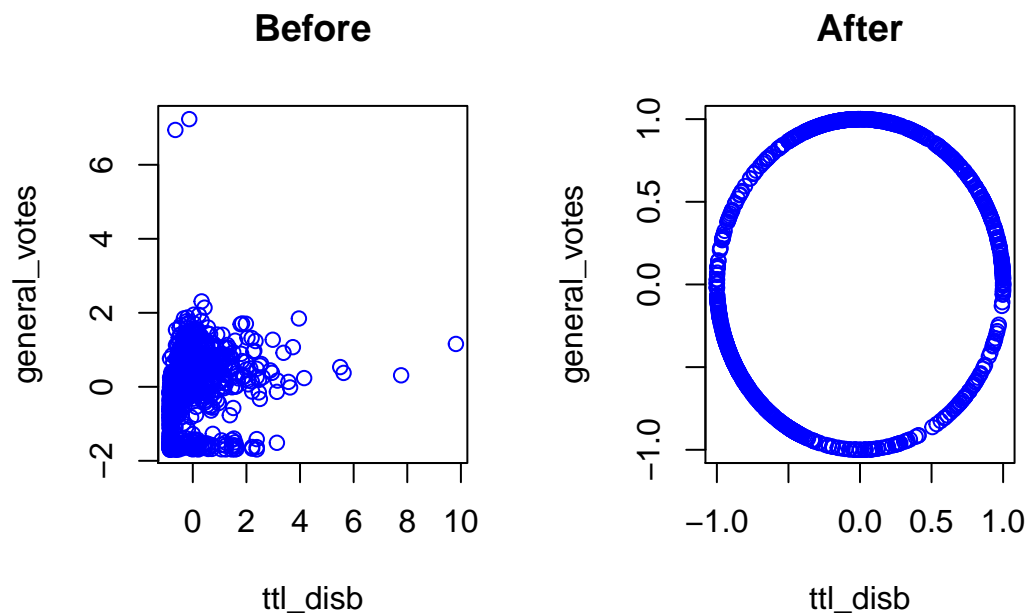
```
transformed <- spatialSign(sdat)
```

```
transformed <- as.data.frame(transformed)
```

```
par(mfrow = c(1, 2), oma = c(2, 2, 2, 2))
```

```
plot(general_votes ~ ttl_disb, data = sdat, col = "blue", main = "Before")
```

```
plot(general_votes ~ ttl_disb, data = transformed, col = "blue", main = "After")
```



```
d2$tvotes<-transformed$"general_votes"
d2$tdisb<-transformed$"ttl_disb"

summary(d2)
```

##	can_party	general_votes	ttl_disb	state
##	Length:880	Min. : 55	Min. : 0	Length:880
##	Class :character	1st Qu.: 88229	1st Qu.: 102276	Class :character
##	Mode :character	Median :142597	Median : 830659	Mode :character
##		Mean :136932	Mean : 1084565	
##		3rd Qu.:198290	3rd Qu.: 1527533	
##		Max. :718591	Max. :13433669	
##	lamvotes	tvotes	tdisb	
##	Min. : 20.38	Min. :-1.00000	Min. :-1.0000	
##	1st Qu.: 3070.96	1st Qu.: -0.65905	1st Qu.: -0.7263	
##	Median : 4236.68	Median : 0.07400	Median : -0.2163	
##	Mean : 3903.53	Mean : 0.07698	Mean : -0.1272	
##	3rd Qu.: 5284.38	3rd Qu.: 0.90077	3rd Qu.: 0.4287	
##	Max. :12523.23	Max. : 1.00000	Max. : 1.0000	

```
#d2<-transformed
```

```
write.csv(d2, "d2.csv")
summary(d2)
```

##	can_party	general_votes	ttl_disb	state
##	Length:880	Min. : 55	Min. : 0	Length:880
##	Class :character	1st Qu.: 88229	1st Qu.: 102276	Class :character
##	Mode :character	Median :142597	Median : 830659	Mode :character
##		Mean :136932	Mean : 1084565	
##		3rd Qu.:198290	3rd Qu.: 1527533	
##		Max. :718591	Max. :13433669	

```
##      lamvotes      tvotes      tdisb
## Min.   : 20.38   Min.   :-1.00000   Min.   :-1.0000
## 1st Qu.: 3070.96 1st Qu.: -0.65905   1st Qu.: -0.7263
## Median : 4236.68 Median : 0.07400   Median : -0.2163
## Mean   : 3903.53 Mean    : 0.07698   Mean    : -0.1272
## 3rd Qu.: 5284.38 3rd Qu.: 0.90077   3rd Qu.: 0.4287
## Max.   :12523.23 Max.    : 1.00000   Max.    : 1.0000
```

```
# set the 'method' option
trans <- preProcess(d2, method = c("center", "scale"))
# use predict() function to get the final result
d3 <- predict(trans, d2)

summary(d3)
```

```
##      can_party      general_votes      ttl_disb      state
## Length:880      Min.   :-1.70236   Min.   :-0.8619   Length:880
## Class :character 1st Qu.: -0.60573   1st Qu.: -0.7806   Class :character
## Mode  :character Median : 0.07045   Median : -0.2018   Mode  :character
##                  Mean    : 0.00000   Mean    : 0.0000
##                  3rd Qu.: 0.76311   3rd Qu.: 0.3520
##                  Max.    : 7.23415   Max.    : 9.8139
##      lamvotes      tvotes      tdisb
## Min.   :-2.1641   Min.   :-1.455716   Min.   :-1.3285
## 1st Qu.: -0.4640   1st Qu.: -0.994862   1st Qu.: -0.9119
## Median : 0.1857   Median : -0.004027   Median : -0.1356
## Mean    : 0.0000   Mean    : 0.000000   Mean    : 0.0000
## 3rd Qu.: 0.7696   3rd Qu.: 1.113486   3rd Qu.: 0.8460
## Max.    : 4.8038   Max.    : 1.247607   Max.    : 1.7156
```

```
write.csv(d3, "d3.csv")
```

```
summary(d3)
```

```
##      can_party      general_votes      ttl_disb      state
## Length:880      Min.   :-1.70236   Min.   :-0.8619   Length:880
## Class :character 1st Qu.: -0.60573   1st Qu.: -0.7806   Class :character
## Mode  :character Median : 0.07045   Median : -0.2018   Mode  :character
##                  Mean    : 0.00000   Mean    : 0.0000
##                  3rd Qu.: 0.76311   3rd Qu.: 0.3520
##                  Max.    : 7.23415   Max.    : 9.8139
##      lamvotes      tvotes      tdisb
## Min.   :-2.1641   Min.   :-1.455716   Min.   :-1.3285
## 1st Qu.: -0.4640   1st Qu.: -0.994862   1st Qu.: -0.9119
## Median : 0.1857   Median : -0.004027   Median : -0.1356
## Mean    : 0.0000   Mean    : 0.000000   Mean    : 0.0000
## 3rd Qu.: 0.7696   3rd Qu.: 1.113486   3rd Qu.: 0.8460
## Max.    : 4.8038   Max.    : 1.247607   Max.    : 1.7156
```

```
write.csv(d3, "d3.csv")
#d2$disb <- log(d2$tdisb)
#d2$votes <- log(d2$tvotes)
```



```
d3$logdisb <- log(d3$tdisb)
```

```
## Warning in log(d3$tdisb): NaNs produced
```

```
d3$logvotes <- log(d3$tvotes)
```

```
## Warning in log(d3$tvotes): NaNs produced
```

```
#d3 <- na.omit(d3)
```

```
#d2[which(!is.finite(d2))] <- 0  
#d2 <- d2[is.finite(rowSums(d2)),]  
#d2[d2 == -Inf] <- 0  
#d3[d3 == -Inf] <- 0  
#data_new <- d2 # Duplicate data  
#d2[is.na(d2$disb) | d2$disb == "Inf"] <- NA # Replace NaN & Inf with NA  
#d3 <- data_new  
#head(d2)  
#head(d2$disb)  
#d3<-d3%>%na.omit()
```

```
#only center and scale R2 = 0.5116
```

```
fit0 <- lm(d3$general_votes ~ d3$ttl_disb + d3$state + d3$can_party)  
summary(fit0)
```

```
##
```

```
## Call:
```

```
## lm(formula = d3$general_votes ~ d3$ttl_disb + d3$state + d3$can_party)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -5.1708 -0.4949 -0.0652  0.4587  3.3568
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   -0.082259   0.511990  -0.161   0.8724  
## d3$ttl_disb    0.237502   0.025991   9.138 < 2e-16 ***  
## d3$stateAL     0.524188   0.556001   0.943   0.3461  
## d3$stateAR     0.603292   0.605358   0.997   0.3193  
## d3$stateAS    -1.383035   0.660605  -2.094   0.0366 *  
## d3$stateAZ     0.141268   0.544374   0.260   0.7953  
## d3$stateCA    -0.001843   0.516456  -0.004   0.9972  
## d3$stateCO     0.623378   0.546654   1.140   0.2545  
## d3$stateCT    -0.368689   0.544442  -0.677   0.4985  
## d3$stateDC     1.793898   0.886133   2.024   0.0433 *  
## d3$stateDE     0.939545   0.723195   1.299   0.1943  
## d3$stateFL     0.366283   0.520334   0.704   0.4817  
## d3$stateGA     0.702187   0.539163   1.302   0.1932  
## d3$stateGU    -1.245282   0.723480  -1.721   0.0856 .
```

```

## d3$stateHI      0.287047  0.660672  0.434  0.6641
## d3$stateIA      0.607270  0.571730  1.062  0.2885
## d3$stateID      0.637111  0.627011  1.016  0.3099
## d3$stateIL      0.511702  0.527825  0.969  0.3326
## d3$stateIN      0.319652  0.540964  0.591  0.5548
## d3$stateKS     -0.122698  0.565483 -0.217  0.8283
## d3$stateKY      0.632212  0.556006  1.137  0.2558
## d3$stateLA     -0.609228  0.537817 -1.133  0.2576
## d3$stateMA      0.943688  0.542895  1.738  0.0825 .
## d3$stateMD      0.569534  0.546716  1.042  0.2978
## d3$stateME      0.543333  0.626397  0.867  0.3860
## d3$stateMI      0.406391  0.529043  0.768  0.4426
## d3$stateMN      0.736585  0.545504  1.350  0.1773
## d3$stateMO      0.847183  0.549584  1.541  0.1236
## d3$stateMP      0.078546  0.892744  0.088  0.9299
## d3$stateMS      0.421057  0.590658  0.713  0.4761
## d3$stateMT      0.801066  0.726121  1.103  0.2703
## d3$stateNC      0.650975  0.531516  1.225  0.2210
## d3$stateND      0.356967  0.661283  0.540  0.5895
## d3$stateNE      0.481916  0.626401  0.769  0.4419
## d3$stateNH     -0.122774  0.605531 -0.203  0.8394
## d3$stateNJ      0.232794  0.534187  0.436  0.6631
## d3$stateNM      0.071177  0.605059  0.118  0.9064
## d3$stateNV     -0.072257  0.565452 -0.128  0.8983
## d3$stateNY     -0.981809  0.515662 -1.904  0.0573 .
## d3$stateOH      0.526224  0.528275  0.996  0.3195
## d3$stateOK      0.574023  0.605256  0.948  0.3432
## d3$stateOR      1.039290  0.579825  1.792  0.0734 .
## d3$statePA      0.697339  0.528689  1.319  0.1875
## d3$statePR      5.364448  0.664425  8.074  2.42e-15 ***
## d3$stateRI     -0.261912  0.626441 -0.418  0.6760
## d3$stateSC     -0.187505  0.539246 -0.348  0.7281
## d3$stateSD      0.546415  0.723236  0.756  0.4502
## d3$stateTN      0.344451  0.546902  0.630  0.5290
## d3$stateTX      0.132274  0.520761  0.254  0.7996
## d3$stateUT     -0.041692  0.571725 -0.073  0.9419
## d3$stateVA      0.578824  0.535162  1.082  0.2798
## d3$stateVI     -1.299244  0.886173 -1.466  0.1430
## d3$stateVT      1.725428  0.886068  1.947  0.0518 .
## d3$stateWA      0.362064  0.542431  0.667  0.5047
## d3$stateWI      0.524065  0.541033  0.969  0.3330
## d3$stateWV     -0.105322  0.580157 -0.182  0.8560
## d3$stateWY     -0.108791  0.661207 -0.165  0.8694
## d3$can_partyOther -1.373386  0.115500 -11.891 < 2e-16 ***
## d3$can_partyREP  0.020657  0.051698  0.400  0.6896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7232 on 821 degrees of freedom
## Multiple R-squared:  0.5116, Adjusted R-squared:  0.477
## F-statistic: 14.82 on 58 and 821 DF, p-value: < 2.2e-16

```

```
#only original data R2 = 0.5116
```

```
fit1 <- lm(d3$tvotes ~ d3$tdisb + d3$state + d3$can_party)
```

```
summary(fit1)
```

```
##
## Call:
## lm(formula = d3$tvotes ~ d3$tdisb + d3$state + d3$can_party)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9250 -0.5917 -0.1259  0.6115  2.2425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.53412    0.56508  -0.945   0.3448
## d3$tdisb        0.33184    0.02853  11.632 <2e-16 ***
## d3$stateAL      0.90811    0.61387   1.479   0.1394
## d3$stateAR      1.28399    0.66844   1.921   0.0551 .
## d3$stateAS     -0.64196    0.72909  -0.880   0.3788
## d3$stateAZ      0.52615    0.60110   0.875   0.3817
## d3$stateCA      0.32277    0.57006   0.566   0.5714
## d3$stateCO      0.93381    0.60335   1.548   0.1221
## d3$stateCT      0.20533    0.60098   0.342   0.7327
## d3$stateDC      1.80627    0.97780   1.847   0.0651 .
## d3$stateDE      1.50202    0.79829   1.882   0.0603 .
## d3$stateFL      0.78630    0.57452   1.369   0.1715
## d3$stateGA      1.26777    0.59533   2.130   0.0335 *
## d3$stateGU     -0.66194    0.79852  -0.829   0.4074
## d3$stateHI      1.05850    0.72962   1.451   0.1472
## d3$stateIA      1.00622    0.63078   1.595   0.1111
## d3$stateID      1.05131    0.69235   1.518   0.1293
## d3$stateIL      0.78899    0.58249   1.355   0.1759
## d3$stateIN      0.76522    0.59704   1.282   0.2003
## d3$stateKS      0.31897    0.62400   0.511   0.6094
## d3$stateKY      1.03751    0.61395   1.690   0.0914 .
## d3$stateLA     -0.30775    0.59374  -0.518   0.6044
## d3$stateMA      1.13443    0.59947   1.892   0.0588 .
## d3$stateMD      0.81136    0.60339   1.345   0.1791
## d3$stateME      0.86074    0.69098   1.246   0.2132
## d3$stateMI      0.85638    0.58389   1.467   0.1428
## d3$stateMN      1.12232    0.60152   1.866   0.0624 .
## d3$stateMO      1.25931    0.60681   2.075   0.0383 *
## d3$stateMP      0.50667    0.98516   0.514   0.6072
## d3$stateMS      1.07169    0.65226   1.643   0.1008
## d3$stateMT      0.54871    0.79814   0.687   0.4920
## d3$stateNC      1.12190    0.58724   1.910   0.0564 .
## d3$stateND      0.66496    0.73004   0.911   0.3626
## d3$stateNE      0.79872    0.69108   1.156   0.2481
## d3$stateNH      0.25598    0.66803   0.383   0.7017
## d3$stateNJ      0.73254    0.58962   1.242   0.2144
## d3$stateNM      0.55368    0.66779   0.829   0.4073
## d3$stateNV      0.10998    0.62388   0.176   0.8601
## d3$stateNY     -0.43376    0.56888  -0.762   0.4460
## d3$stateOH      0.96918    0.58350   1.661   0.0971 .
## d3$stateOK      1.20276    0.66805   1.800   0.0722 .
```

```
## d3$stateOR      1.41977      0.63999      2.218      0.0268 *
## d3$statePA      1.00513      0.58356      1.722      0.0854 .
## d3$statePR      1.71829      0.73315      2.344      0.0193 *
## d3$stateRI     -0.10839      0.69262     -0.156      0.8757
## d3$stateSC      0.40481      0.59552      0.680      0.4969
## d3$stateSD      0.73557      0.79804      0.922      0.3569
## d3$stateTN      0.93943      0.60406      1.555      0.1203
## d3$stateTX      0.60212      0.57490      1.047      0.2952
## d3$stateUT      0.42026      0.63094      0.666      0.5055
## d3$stateVA      1.04359      0.59094      1.766      0.0778 .
## d3$stateVI     -0.71038      0.97788     -0.726      0.4678
## d3$stateVT      1.77837      0.97767      1.819      0.0693 .
## d3$stateWA      0.82053      0.59877      1.370      0.1710
## d3$stateWI      0.87144      0.59718      1.459      0.1449
## d3$stateWV      0.35967      0.64058      0.561      0.5746
## d3$stateWY      0.14112      0.72959      0.193      0.8467
## d3$can_partyOther -1.10702      0.12663     -8.742     <2e-16 ***
## d3$can_partyREP   0.07709      0.05712      1.350      0.1775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7978 on 821 degrees of freedom
## Multiple R-squared:  0.4055, Adjusted R-squared:  0.3635
## F-statistic: 9.655 on 58 and 821 DF,  p-value: < 2.2e-16
```

```
#only original, log(spending) data R2 = 0.6041
fit2 <- lm(d3$logvotes ~ d3$logdisb + d3$state + d3$can_party)
summary(fit2)
```

```
##
## Call:
## lm(formula = d3$logvotes ~ d3$logdisb + d3$state + d3$can_party)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5784 -0.1343  0.1078  0.3683  1.2149
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.09120    0.77418   0.118  0.90632
## d3$logdisb     -0.33277    0.04225  -7.877 1.34e-13 ***
## d3$stateAL     -0.23316    0.85792  -0.272  0.78604
## d3$stateAR     -0.02316    1.08210  -0.021  0.98295
## d3$stateAZ     -2.15197    0.88740  -2.425  0.01608 *
## d3$stateCA     -0.92438    0.78493  -1.178  0.24015
## d3$stateCO     -0.62587    0.82227  -0.761  0.44735
## d3$stateCT     -0.19042    0.84650  -0.225  0.82222
## d3$stateDE     -0.09853    1.08852  -0.091  0.92796
## d3$stateFL     -0.54024    0.78608  -0.687  0.49261
## d3$stateGA     -0.38604    0.81496  -0.474  0.63617
## d3$stateHI     -0.04283    1.08844  -0.039  0.96865
## d3$stateIA     -0.63451    0.83956  -0.756  0.45057
## d3$stateID     -1.61350    1.10854  -1.456  0.14689
## d3$stateIL     -0.80246    0.80126  -1.001  0.31765
```

```

## d3$stateIN      -0.62371      0.81419   -0.766   0.44444
## d3$stateKS      -0.40702      0.85548   -0.476   0.63468
## d3$stateKY      -0.21287      0.83948   -0.254   0.80005
## d3$stateLA      -0.15868      0.88449   -0.179   0.85778
## d3$stateMA      -0.29874      0.84730   -0.353   0.72472
## d3$stateMD      -0.28610      0.83479   -0.343   0.73212
## d3$stateME      -1.57267      0.93926   -1.674   0.09542 .
## d3$stateMI      -0.44925      0.79757   -0.563   0.57380
## d3$stateMN      -0.90759      0.81168   -1.118   0.26467
## d3$stateMO      -0.31278      0.83142   -0.376   0.70712
## d3$stateMS      -0.73468      0.89237   -0.823   0.41120
## d3$stateMT      -0.49793      0.93917   -0.530   0.59650
## d3$stateNC      -0.17823      0.83831   -0.213   0.83182
## d3$stateND      -0.41024      1.08575   -0.378   0.70590
## d3$stateNE      -2.44656      1.08222   -2.261   0.02472 *
## d3$stateNH      -0.36919      0.88723   -0.416   0.67772
## d3$stateNJ      -0.73246      0.80465   -0.910   0.36363
## d3$stateNM      -0.37483      0.94482   -0.397   0.69194
## d3$stateNV      -2.49600      0.93928   -2.657   0.00843 **
## d3$stateNY      -0.72344      0.79572   -0.909   0.36422
## d3$stateOH      -0.03717      0.81841   -0.045   0.96381
## d3$stateOK      -0.05756      0.88413   -0.065   0.94815
## d3$stateOR      -0.36298      0.86101   -0.422   0.67373
## d3$statePA      -0.32428      0.79132   -0.410   0.68233
## d3$statePR      -0.73667      0.96309   -0.765   0.44512
## d3$stateSC      -0.12043      0.88440   -0.136   0.89180
## d3$stateSD      -0.18375      1.08213   -0.170   0.86531
## d3$stateTN      -0.10755      0.85576   -0.126   0.90009
## d3$stateTX      -0.30567      0.78517   -0.389   0.69742
## d3$stateUT      -0.02155      1.08274   -0.020   0.98414
## d3$stateVA      -0.41240      0.81388   -0.507   0.61284
## d3$stateWA      -0.38631      0.81849   -0.472   0.63739
## d3$stateWI      -0.77909      0.82122   -0.949   0.34377
## d3$stateWV      -0.87210      0.93791   -0.930   0.35344
## d3$stateWY      -1.14576      1.08220   -1.059   0.29084
## d3$can_partyOther 0.14347      0.37385    0.384   0.70152
## d3$can_partyREP   -0.10835      0.11863   -0.913   0.36202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7652 on 229 degrees of freedom
## (599 observations deleted due to missingness)
## Multiple R-squared:  0.4161, Adjusted R-squared:  0.2861
## F-statistic: 3.2 on 51 and 229 DF, p-value: 1.33e-09

#only original, log(spending) data R2 = 0.6173
fit3 <- lm(d3$tvotes ~ d3$logdisb + d3$state + d3$can_party)
summary(fit3)

##
## Call:
## lm(formula = d3$tvotes ~ d3$logdisb + d3$state + d3$can_party)
##
## Residuals:

```

```

##      Min      1Q   Median      3Q      Max
## -2.11222 -0.33336  0.03292  0.34268  2.05060
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.42987    0.49818   -0.863   0.3888
## d3$logdisb     -0.15998    0.03303  -4.843 1.93e-06 ***
## d3$stateAL      1.12846    0.58738   1.921   0.0555 .
## d3$stateAR      1.22425    0.86012   1.423   0.1555
## d3$stateAZ      0.27724    0.60832   0.456   0.6489
## d3$stateCA      0.63061    0.50769   1.242   0.2150
## d3$stateCO      1.12533    0.56290   1.999   0.0464 *
## d3$stateCT      0.29122    0.54971   0.530   0.5966
## d3$stateDE      1.54443    0.85985   1.796   0.0733 .
## d3$stateFL      0.95164    0.51815   1.837   0.0671 .
## d3$stateGA      1.25050    0.55577   2.250   0.0251 *
## d3$stateHI      1.53706    0.85998   1.787   0.0748 .
## d3$stateIA      0.98529    0.58689   1.679   0.0941 .
## d3$stateID      0.64330    0.87592   0.734   0.4632
## d3$stateIL      0.99870    0.53350   1.872   0.0620 .
## d3$stateIN      0.95555    0.54908   1.740   0.0827 .
## d3$stateKS      0.59664    0.58834   1.014   0.3112
## d3$stateKY      1.26124    0.58832   2.144   0.0327 *
## d3$stateLA      0.03711    0.57333   0.065   0.9484
## d3$stateMA      1.45431    0.58831   2.472   0.0139 *
## d3$stateMD      1.09044    0.56315   1.936   0.0536 .
## d3$stateME      0.67406    0.70226   0.960   0.3378
## d3$stateMI      0.98147    0.53020   1.851   0.0650 .
## d3$stateMN      1.04827    0.54899   1.909   0.0570 .
## d3$stateMO      1.32506    0.57616   2.300   0.0220 *
## d3$stateMS      1.13229    0.64451   1.757   0.0798 .
## d3$stateMT      0.98405    0.70211   1.402   0.1619
## d3$stateNC      1.17561    0.58821   1.999   0.0464 *
## d3$stateND      1.22774    0.86119   1.426   0.1549
## d3$stateNE      0.43542    0.70232   0.620   0.5357
## d3$stateNH      0.65143    0.61168   1.065   0.2876
## d3$stateNJ      1.10465    0.54350   2.032   0.0429 *
## d3$stateNM      1.15542    0.70345   1.643   0.1014
## d3$stateNV      0.30672    0.60849   0.504   0.6145
## d3$stateNY     -0.50606    0.50198  -1.008   0.3141
## d3$stateOH      1.35313    0.56303   2.403   0.0168 *
## d3$stateOK      1.33483    0.64147   2.081   0.0382 *
## d3$stateOR      1.31537    0.60784   2.164   0.0311 *
## d3$statePA      1.12256    0.52617   2.133   0.0336 *
## d3$statePR      1.37527    0.71619   1.920   0.0556 .
## d3$stateSC      1.28319    0.64059   2.003   0.0459 *
## d3$stateSD      1.07829    0.86025   1.253   0.2109
## d3$stateTN      1.27389    0.60868   2.093   0.0371 *
## d3$stateTX      0.77441    0.51592   1.501   0.1342
## d3$stateUT      0.55615    0.64070   0.868   0.3860
## d3$stateVA      1.20026    0.55442   2.165   0.0311 *
## d3$stateWA      1.12739    0.54867   2.055   0.0406 *
## d3$stateWI      1.02751    0.55457   1.853   0.0648 .
## d3$stateWV      0.90715    0.70250   1.291   0.1974

```

```
## d3$stateWY      0.63747    0.86045    0.741    0.4593
## d3$can_partyOther -0.16367    0.26750   -0.612    0.5410
## d3$can_partyREP   0.13746    0.08183    1.680    0.0939 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7013 on 350 degrees of freedom
## (478 observations deleted due to missingness)
## Multiple R-squared:  0.526, Adjusted R-squared:  0.4569
## F-statistic: 7.616 on 51 and 350 DF, p-value: < 2.2e-16
```

```
summary(d2$state)
```

```
##      Length      Class      Mode
##      880 character character
```

```
d2$disb <- log(d2$t1l_disb)
d2$votes <- log(d2$general_votes)

write.csv(d2, "d2.csv")

#d2[which(!is.finite(d2))] <- 0
#d2 <- d2[is.finite(rowSums(d2)),]
d2[d2 == -Inf] <- 0

#data_new <- d2                                # Duplicate data

#d2[is.na(d2$disb) | d2$disb == "Inf"] <- NA # Replace NaN & Inf with NA

#d3 <- data_new

head(d2)
```

```
## # A tibble: 6 x 9
##   can_party general_votes t1l_disb state lamvotes  tvotes  tdisb  disb votes
##   <chr>          <dbl>    <dbl> <chr>    <dbl>    <dbl>    <dbl> <dbl> <dbl>
## 1 REP            208083 1172750. AL      5458.    0.997    0.0789 14.0   12.2
## 2 REP            134886 1850536. AL      4082.   -0.0418    0.999   14.4   11.8
## 3 DEM            112089   36844 AL      3605.   -0.348   -0.938   10.5   11.6
## 4 REP            192164 1071289. AL      5174.    1.00   -0.0154 13.9   12.2
## 5 DEM             94549    7348 AL      3217.  -0.524   -0.852    8.90   11.5
## 6 REP            235925 1394461. AL      5937.    0.981    0.196   14.1   12.4
```

```
head(d2$disb)
```

```
## [1] 13.974862 14.430986 10.514448 13.884374  8.902183 14.148019
```

```
#d3<-d3%>%na.omit()
```

```
fit0 <- lm(d2$general_votes ~ d2$t1l_disb + d2$state + d2$can_party)
summary(fit0)
```

```
##
## Call:
## lm(formula = d2$general_votes ~ d2$ttl_disb + d2$state + d2$can_party)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -415756  -39794   -5242   36879  269903
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.139e+05  4.120e+04   2.763  0.00585 **
## d2$ttl_disb     1.518e-02  1.661e-03   9.138 < 2e-16 ***
## d2$stateAL      4.215e+04  4.471e+04   0.943  0.34607
## d2$stateAR      4.851e+04  4.867e+04   0.997  0.31926
## d2$stateAS     -1.112e+05  5.312e+04  -2.094  0.03660 *
## d2$stateAZ      1.136e+04  4.377e+04   0.260  0.79531
## d2$stateCA     -1.482e+02  4.153e+04  -0.004  0.99715
## d2$stateCO      5.012e+04  4.395e+04   1.140  0.25447
## d2$stateCT     -2.964e+04  4.378e+04  -0.677  0.49848
## d2$stateDC      1.442e+05  7.125e+04   2.024  0.04325 *
## d2$stateDE      7.554e+04  5.815e+04   1.299  0.19425
## d2$stateFL      2.945e+04  4.184e+04   0.704  0.48167
## d2$stateGA      5.646e+04  4.335e+04   1.302  0.19316
## d2$stateGU     -1.001e+05  5.817e+04  -1.721  0.08558 .
## d2$stateHI      2.308e+04  5.312e+04   0.434  0.66406
## d2$stateIA      4.883e+04  4.597e+04   1.062  0.28847
## d2$stateID      5.123e+04  5.041e+04   1.016  0.30988
## d2$stateIL      4.114e+04  4.244e+04   0.969  0.33260
## d2$stateIN      2.570e+04  4.350e+04   0.591  0.55475
## d2$stateKS     -9.865e+03  4.547e+04  -0.217  0.82828
## d2$stateKY      5.083e+04  4.471e+04   1.137  0.25585
## d2$stateLA     -4.898e+04  4.324e+04  -1.133  0.25764
## d2$stateMA      7.588e+04  4.365e+04   1.738  0.08254 .
## d2$stateMD      4.579e+04  4.396e+04   1.042  0.29784
## d2$stateME      4.369e+04  5.037e+04   0.867  0.38598
## d2$stateMI      3.268e+04  4.254e+04   0.768  0.44261
## d2$stateMN      5.922e+04  4.386e+04   1.350  0.17730
## d2$stateMO      6.812e+04  4.419e+04   1.541  0.12358
## d2$stateMP      6.315e+03  7.178e+04   0.088  0.92991
## d2$stateMS      3.385e+04  4.749e+04   0.713  0.47613
## d2$stateMT      6.441e+04  5.838e+04   1.103  0.27026
## d2$stateNC      5.234e+04  4.274e+04   1.225  0.22102
## d2$stateND      2.870e+04  5.317e+04   0.540  0.58947
## d2$stateNE      3.875e+04  5.037e+04   0.769  0.44191
## d2$stateNH     -9.872e+03  4.869e+04  -0.203  0.83938
## d2$stateNJ      1.872e+04  4.295e+04   0.436  0.66310
## d2$stateNM      5.723e+03  4.865e+04   0.118  0.90638
## d2$stateNV     -5.810e+03  4.546e+04  -0.128  0.89835
## d2$stateNY     -7.894e+04  4.146e+04  -1.904  0.05726 .
## d2$stateOH      4.231e+04  4.248e+04   0.996  0.31949
## d2$stateOK      4.615e+04  4.867e+04   0.948  0.34321
## d2$stateOR      8.356e+04  4.662e+04   1.792  0.07343 .
## d2$statePA      5.607e+04  4.251e+04   1.319  0.18754
## d2$statePR      4.313e+05  5.342e+04   8.074  2.42e-15 ***
```



```
## d2$stateRI      -2.106e+04  5.037e+04  -0.418  0.67599
## d2$stateSC      -1.508e+04  4.336e+04  -0.348  0.72814
## d2$stateSD       4.393e+04  5.815e+04   0.756  0.45016
## d2$stateTN       2.770e+04  4.397e+04   0.630  0.52899
## d2$stateTX       1.064e+04  4.187e+04   0.254  0.79956
## d2$stateUT      -3.352e+03  4.597e+04  -0.073  0.94188
## d2$stateVA       4.654e+04  4.303e+04   1.082  0.27975
## d2$stateVI      -1.045e+05  7.125e+04  -1.466  0.14300
## d2$stateVT       1.387e+05  7.124e+04   1.947  0.05184 .
## d2$stateWA       2.911e+04  4.361e+04   0.667  0.50465
## d2$stateWI       4.214e+04  4.350e+04   0.969  0.33301
## d2$stateWV      -8.468e+03  4.665e+04  -0.182  0.85599
## d2$stateWY      -8.747e+03  5.316e+04  -0.165  0.86935
## d2$can_partyOther -1.104e+05  9.287e+03 -11.891 < 2e-16 ***
## d2$can_partyREP   1.661e+03  4.157e+03   0.400  0.68957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58140 on 821 degrees of freedom
## Multiple R-squared:  0.5116, Adjusted R-squared:  0.477
## F-statistic: 14.82 on 58 and 821 DF,  p-value: < 2.2e-16
```

```
fit1 <- lm(d2$general_votes ~ d2$disb + d2$state + d2$can_party)
summary(fit1)
```

```
##
## Call:
## lm(formula = d2$general_votes ~ d2$disb + d2$state + d2$can_party)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -378949  -35379   -1422    30616   228002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -78916.4    39004.1  -2.023   0.0434 *
## d2$disb         15114.2     880.1   17.174 < 2e-16 ***
## d2$stateAL      56260.9    40260.6   1.397   0.1627
## d2$stateAR      57802.7    43824.5   1.319   0.1876
## d2$stateAS     -72654.9    47899.7  -1.517   0.1297
## d2$stateAZ      29836.2    39421.4   0.757   0.4494
## d2$stateCA      17128.9    37399.7   0.458   0.6471
## d2$stateCO      64120.1    39580.5   1.620   0.1056
## d2$stateCT     -19862.7    39417.3  -0.504   0.6145
## d2$stateDC     146804.2    64142.8   2.289   0.0223 *
## d2$stateDE      81261.9    52351.8   1.552   0.1210
## d2$stateFL      47406.9    37680.6   1.258   0.2087
## d2$stateGA      76113.3    39053.8   1.949   0.0516 .
## d2$stateGU     -85174.9    52378.5  -1.626   0.1043
## d2$stateHI      24440.2    47823.8   0.511   0.6095
## d2$stateIA      56284.4    41384.9   1.360   0.1742
## d2$stateID      64117.5    45397.5   1.412   0.1582
## d2$stateIL      53099.1    38209.2   1.390   0.1650
## d2$stateIN      39081.9    39168.9   0.998   0.3187
```

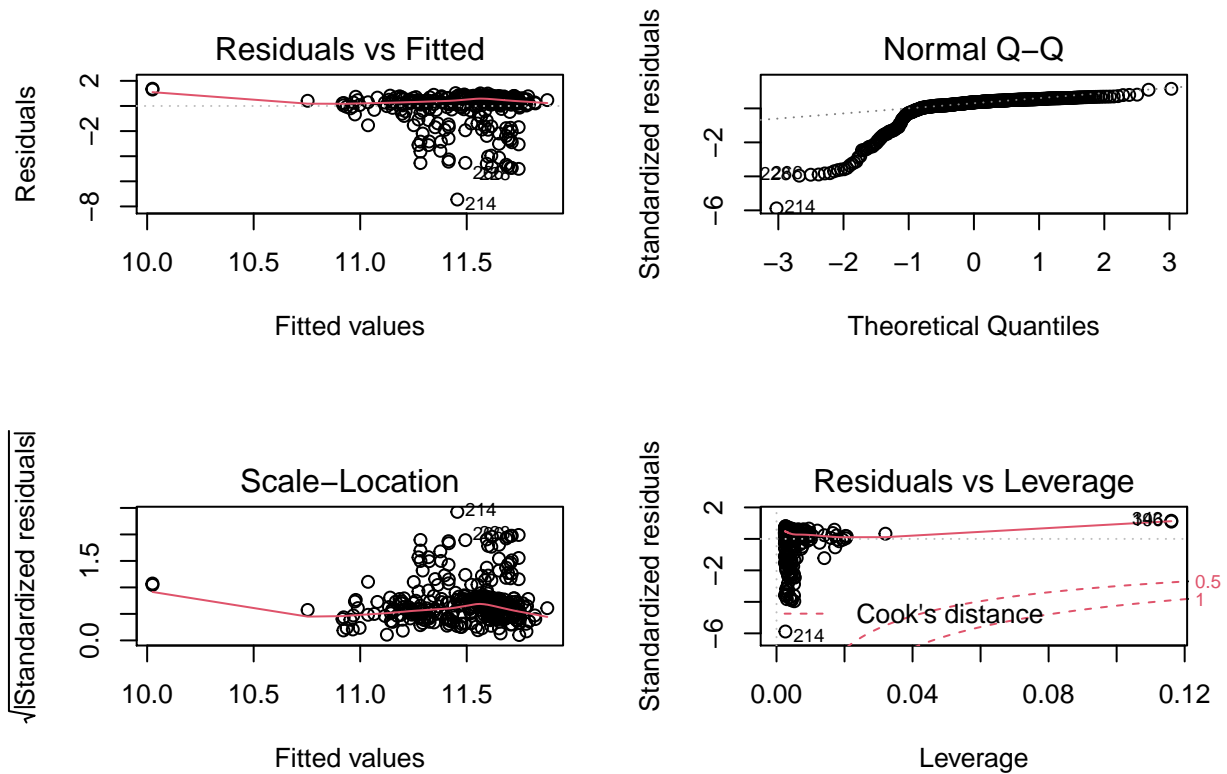
```
## d2$stateKS      -5399.2    40934.3  -0.132    0.8951
## d2$stateKY      66803.2    40264.3   1.659    0.0975 .
## d2$stateLA     -36807.4    38941.3  -0.945    0.3448
## d2$stateMA      84732.1    39304.1   2.156    0.0314 *
## d2$stateMD      53793.9    39578.7   1.359    0.1745
## d2$stateME      59654.7    45336.0   1.316    0.1886
## d2$stateMI      50454.6    38311.6   1.317    0.1882
## d2$stateMN      67285.5    39467.7   1.705    0.0886 .
## d2$stateMO      74373.2    39785.7   1.869    0.0619 .
## d2$stateMP     -3586.2    64623.1  -0.055    0.9558
## d2$stateMS      70879.1    42832.0   1.655    0.0983 .
## d2$stateMT      94261.3    52358.5   1.800    0.0722 .
## d2$stateNC      72652.2    38502.5   1.887    0.0595 .
## d2$stateND      44297.4    47882.1   0.925    0.3552
## d2$stateNE      42828.1    45344.4   0.945    0.3452
## d2$stateNH     -13691.4    43832.0  -0.312    0.7548
## d2$stateNJ      38246.2    38684.9   0.989    0.3231
## d2$stateNM      12176.5    43801.2   0.278    0.7811
## d2$stateNV      1680.0     40934.3   0.041    0.9673
## d2$stateNY     -71680.4    37323.9  -1.920    0.0551 .
## d2$stateOH      59912.9    38261.6   1.566    0.1178
## d2$stateOK      52178.6    43815.3   1.191    0.2340
## d2$stateOR      91212.5    41974.8   2.173    0.0301 *
## d2$statePA      64973.1    38275.0   1.698    0.0900 .
## d2$statePR     434255.9    48096.7   9.029 < 2e-16 ***
## d2$stateRI       2223.8    45377.7   0.049    0.9609
## d2$stateSC       3311.5    39057.5   0.085    0.9325
## d2$stateSD      53147.5    52347.7   1.015    0.3103
## d2$stateTN      56307.3    39638.8   1.421    0.1558
## d2$stateTX      27554.8    37713.9   0.731    0.4652
## d2$stateUT       8442.1    41387.1   0.204    0.8384
## d2$stateVA      57819.4    38747.1   1.492    0.1360
## d2$stateVI     -98991.8    64146.4  -1.543    0.1232
## d2$stateVT     138293.3    64139.7   2.156    0.0314 *
## d2$stateWA      43843.3    39278.6   1.116    0.2647
## d2$stateWI      65675.6    39171.6   1.677    0.0940 .
## d2$stateWV       76.6     41999.6   0.002    0.9985
## d2$stateWY     16878.2    47890.3   0.352    0.7246
## d2$can_partyOther -72265.7    8767.5  -8.242 6.65e-16 ***
## d2$can_partyREP   683.6     3727.5   0.183    0.8545
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52350 on 821 degrees of freedom
## Multiple R-squared:  0.6041, Adjusted R-squared:  0.5761
## F-statistic: 21.6 on 58 and 821 DF, p-value: < 2.2e-16
```

```
## boxcox test
#library(MASS)
#boxcox(general_votes~poly(disb,2),          data = d2)

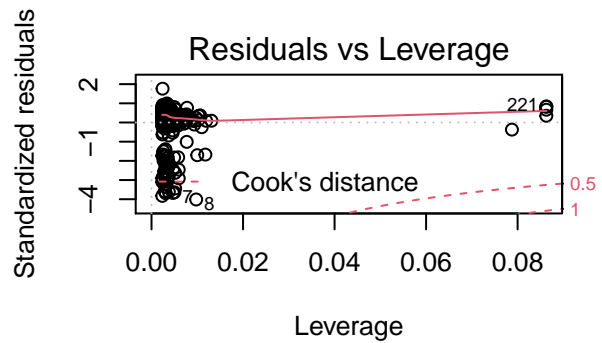
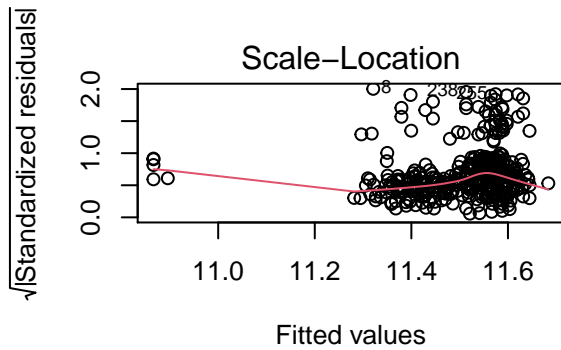
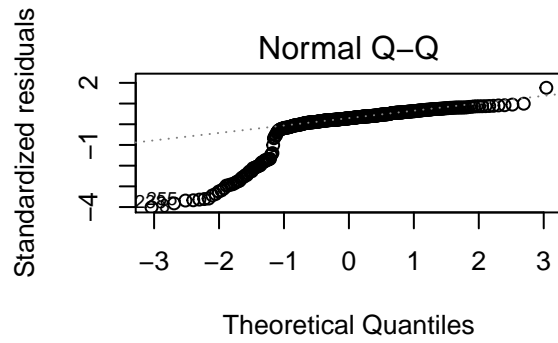
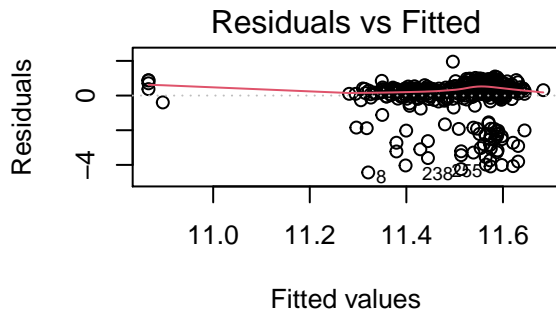
g1 <- filter(d2, can_party == "REP")
```

```
g2 <- filter(d2, can_party == "DEM")
g3 <- filter(d2, can_party == "Other")
```

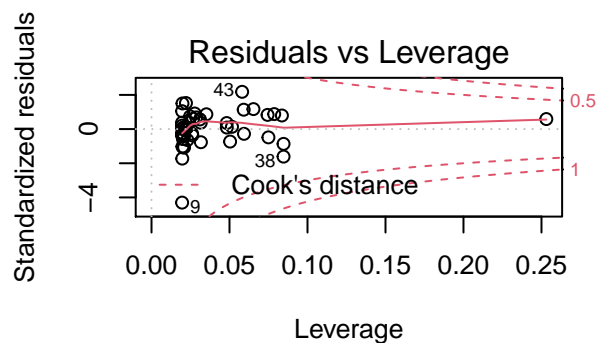
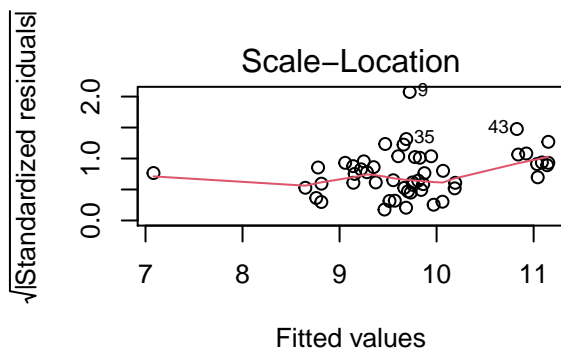
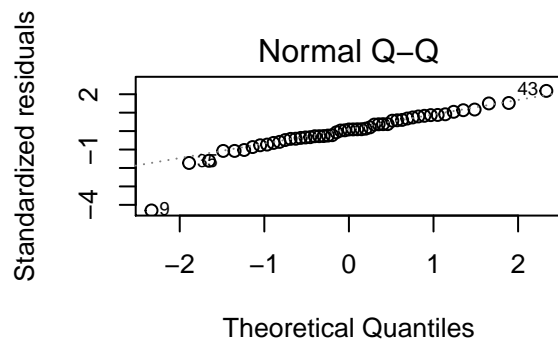
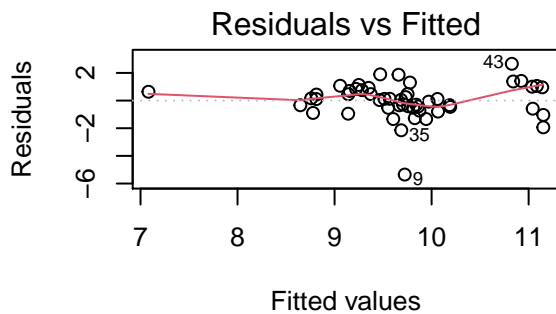
```
fit <- lm(g1$votes ~ g1$disb)
par(mfrow=c(2,2))
plot (fit)
```



```
fit1 <- lm(g2$votes ~ g2$disb)
par(mfrow=c(2,2))
plot (fit1)
```



```
fit2 <- lm(g3$votes ~ g3$disb)
par(mfrow=c(2,2))
plot(fit2)
```



```
summary(fit)
```

```
##
## Call:
## lm(formula = g1$votes ~ g1$disb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4485  0.1288  0.4484  0.6419  1.3730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.02502    0.43189  23.212 < 2e-16 ***
## g1$disb      0.11290    0.03245   3.479 0.000557 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.268 on 404 degrees of freedom
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.02669
## F-statistic: 12.11 on 1 and 404 DF,  p-value: 0.0005571
```

```
summary(fit1)
```

```
##
## Call:
## lm(formula = g2$votes ~ g2$disb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4355  0.0615  0.3208  0.5981  1.9554
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.86598    0.32627  33.304 <2e-16 ***
## g2$disb      0.05047    0.02509   2.011  0.0449 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.111 on 421 degrees of freedom
## Multiple R-squared:  0.009519, Adjusted R-squared:  0.007166
## F-statistic: 4.046 on 1 and 421 DF,  p-value: 0.04491
```

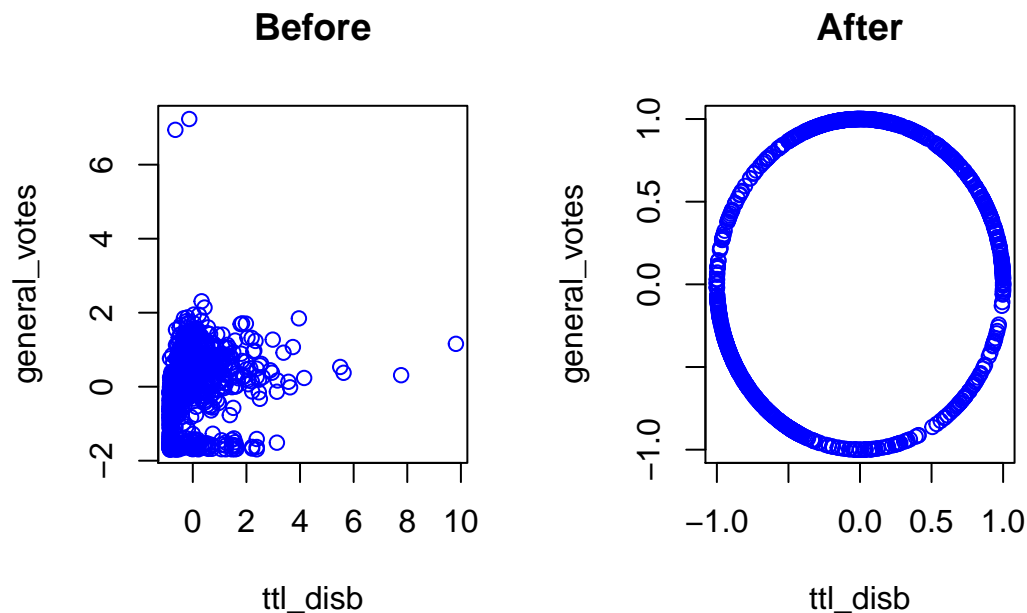
```
summary(fit2)
```

```
##
## Call:
## lm(formula = g3$votes ~ g3$disb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.3535 -0.5197  0.1090  0.7988  2.6582
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.08213    0.63215  11.203 4.06e-15 ***
## g3$disb      0.27274    0.06218   4.387 6.10e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.257 on 49 degrees of freedom
## Multiple R-squared:  0.282, Adjusted R-squared:  0.2673
## F-statistic: 19.24 on 1 and 49 DF, p-value: 6.103e-05
```

```
d2[d2 == -Inf] <- 0
```

```
sdat <- d2[, c("general_votes", "ttl_disb")]
imp <- preProcess(sdat, method = c("knnImpute"), k = 5)
sdat <- predict(imp, sdat)
transformed <- spatialSign(sdat)
transformed <- as.data.frame(transformed)
par(mfrow = c(1, 2), oma = c(2, 2, 2, 2))
plot(general_votes ~ ttl_disb, data = sdat, col = "blue", main = "Before")
plot(general_votes ~ ttl_disb, data = transformed, col = "blue", main = "After")
```



```
#d2<-transformed
```

```
d2[d2 == -Inf] <- 0
```

```
head(d2)
```

```
## # A tibble: 6 x 9
##   can_party general_votes ttl_disb state lamvotes tvotes tdisb disb votes
##   <chr>         <dbl>     <dbl> <chr>   <dbl>   <dbl>   <dbl> <dbl> <dbl>
## 1 REP           208083 1172750. AL      5458.   0.997   0.0789 14.0   12.2
```

```
## 2 REP          134886 1850536. AL          4082. -0.0418  0.999  14.4  11.8
## 3 DEM          112089   36844 AL          3605. -0.348 -0.938  10.5  11.6
## 4 REP          192164 1071289. AL          5174.  1.00 -0.0154 13.9  12.2
## 5 DEM           94549   7348 AL          3217. -0.524 -0.852   8.90  11.5
## 6 REP          235925 1394461. AL          5937.  0.981  0.196  14.1  12.4
```

```
head(d2$disb)
```

```
## [1] 13.974862 14.430986 10.514448 13.884374  8.902183 14.148019
```

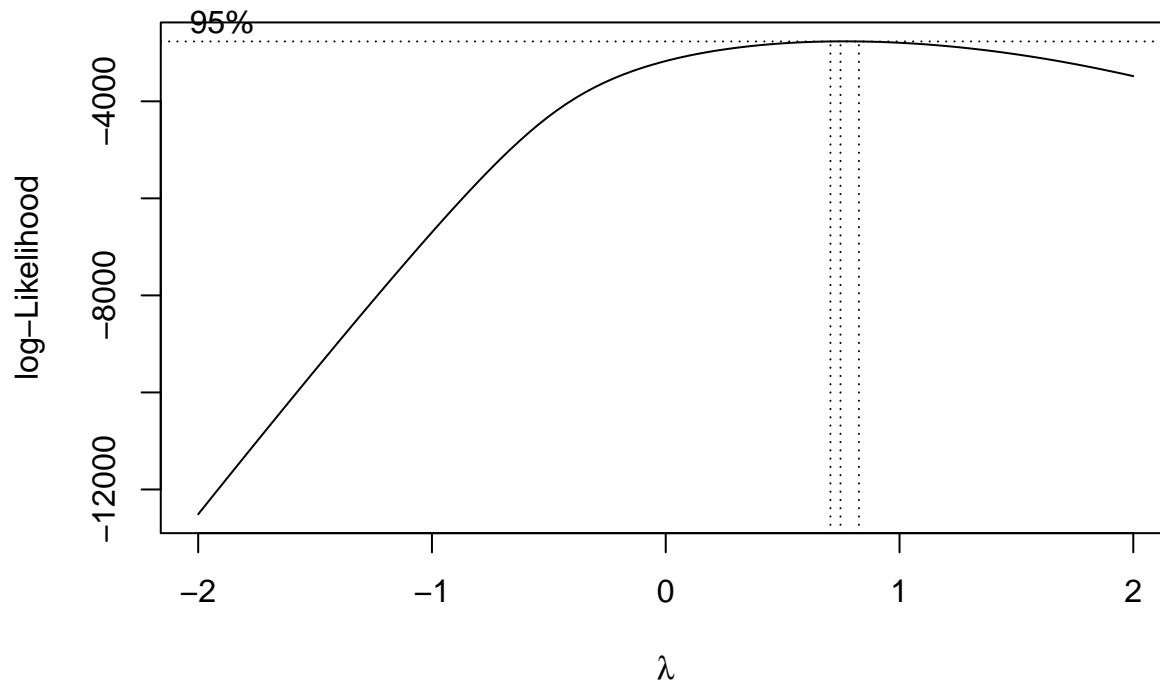
```
#d3<-d3%>%na.omit()
```

```
fit <- lm(d2$general_votes ~ d2$disb)
summary(fit)
```

```
##
## Call:
## lm(formula = d2$general_votes ~ d2$disb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -170750  -34066    7653   45029  568412
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -46697     14420  -3.238  0.00125 **
## d2$disb         14339       1109  12.928 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 73740 on 878 degrees of freedom
## Multiple R-squared:  0.1599, Adjusted R-squared:  0.159
## F-statistic: 167.1 on 1 and 878 DF, p-value: < 2.2e-16
```

```
## boxcox test
```

```
library(MASS)
boxcox(general_votes~poly(disb,2),
       data = d2)
```



```
# g0 <- d2
# g0$votes <- log10(g0$general_votes)
# g0$disb <- log10(g0$t1l_disb)
# g0[g0 == -Inf] <- 0

g0 <- d2
g0$votes <- g0$general_votes
g0$disb <- g0$t1l_disb
g0[g0 == -Inf] <- 0

g1 <- filter(d2, can_party == "REP")
g1$votes <- g1$general_votes*g1$general_votes
g1$disb <- log(g1$t1l_disb)
g1[g1 == -Inf] <- 0

g2 <- filter(d2, can_party == "DEM")
g2$votes <- g2$general_votes*g2$general_votes
g2$disb <- log(g2$t1l_disb)
g2[g2 == -Inf] <- 0

g3 <- filter(d2, can_party == "Other")
g3$votes <- g3$general_votes
g3$disb <- log(g3$t1l_disb)
g3[g3 == -Inf] <- 0

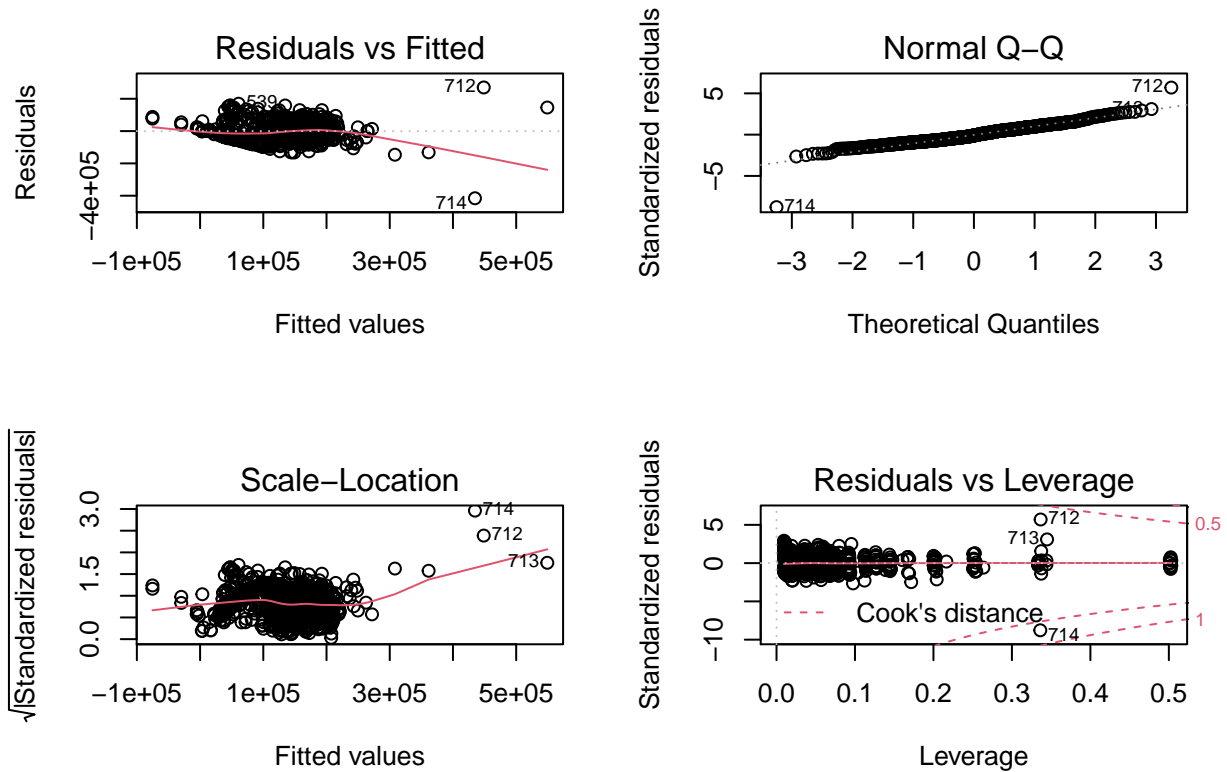
write.csv(g1, "g1.csv")
write.csv(g2, "g2.csv")
write.csv(g3, "g3.csv")
```



```
fit0 <- lm(g0$votes ~ g0$disb + g0$state + g0$can_party )
par(mfrow=c(2,2))
plot (fit0)
```

```
## Warning: not plotting observations with leverage one:
```

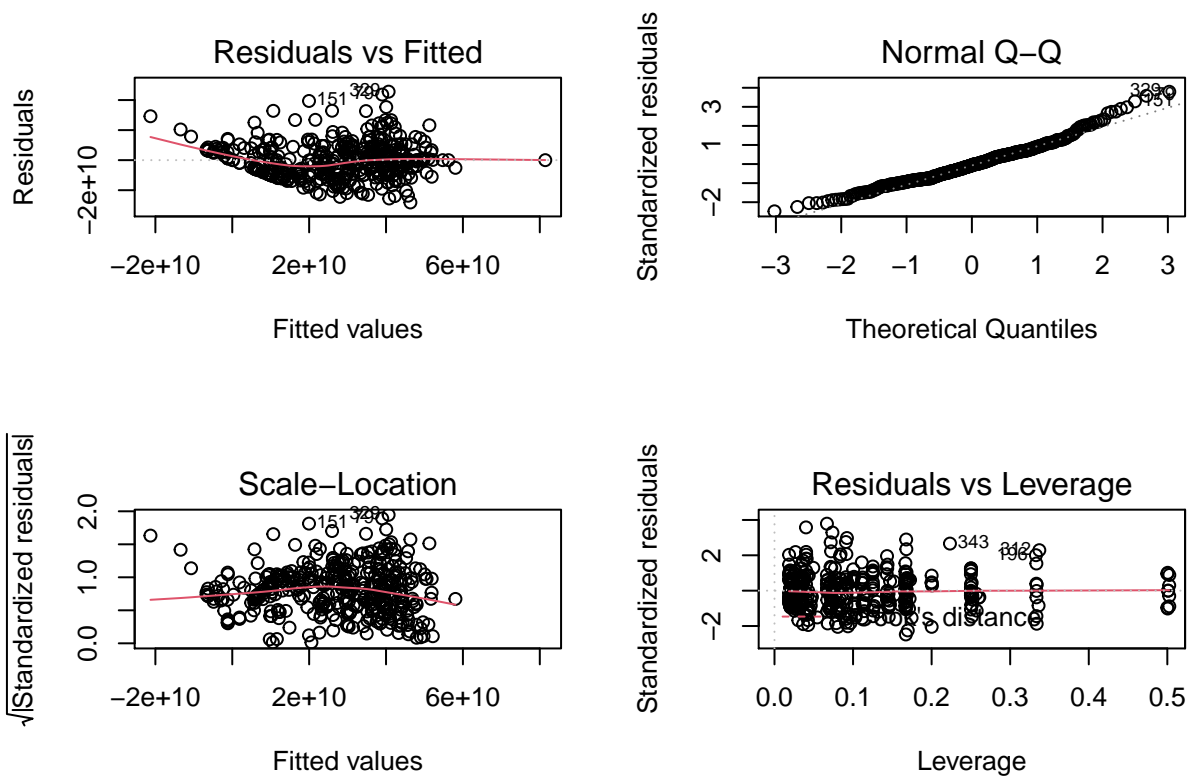
```
## 168, 640, 815, 837
```



```
fit1 <- lm(g1$votes ~ g1$disb + g1$state )
par(mfrow=c(2,2))
plot (fit1)
```

```
## Warning: not plotting observations with leverage one:
```

```
## 7, 8, 75, 113, 205, 293, 338, 406
```

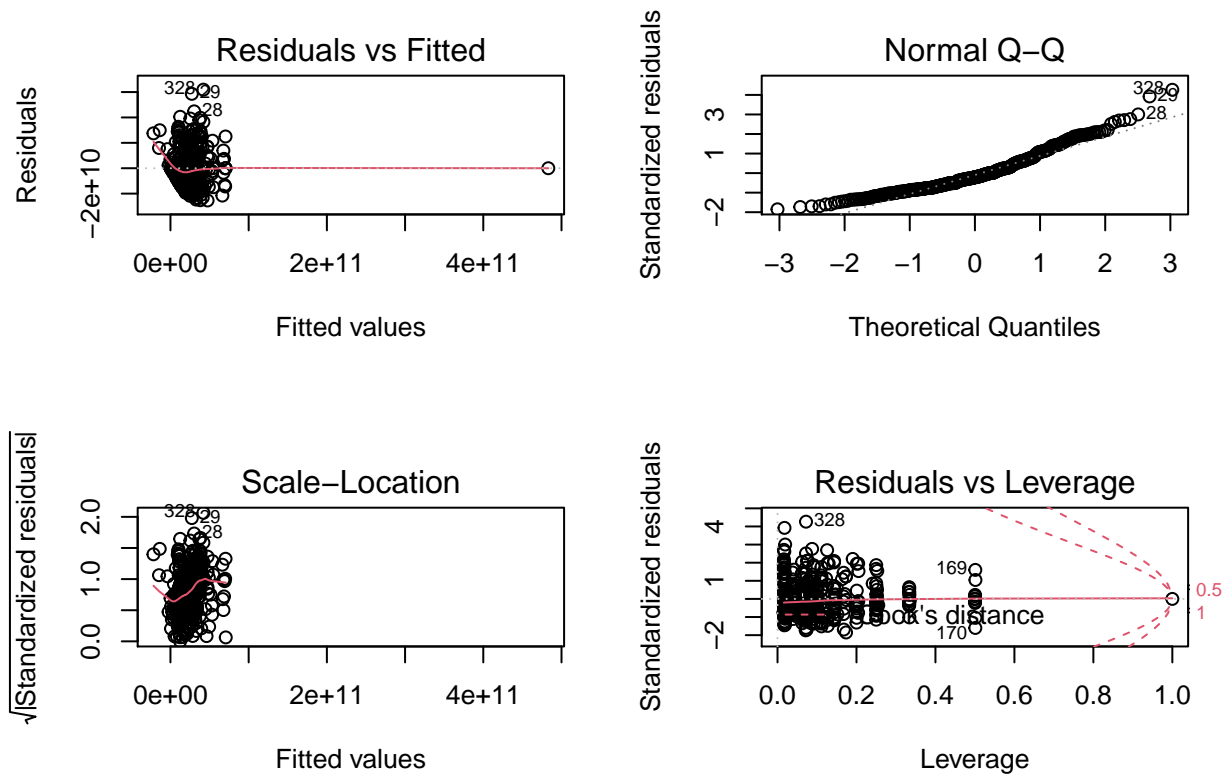


```
fit2 <- lm(g2$votes ~ g2$disb + g2$state )
par(mfrow=c(2,2))
plot (fit2)
```

```
## Warning: not plotting observations with leverage one:
## 6, 91, 92, 126, 130, 211, 212, 307, 322, 341, 356, 389, 401, 423
```

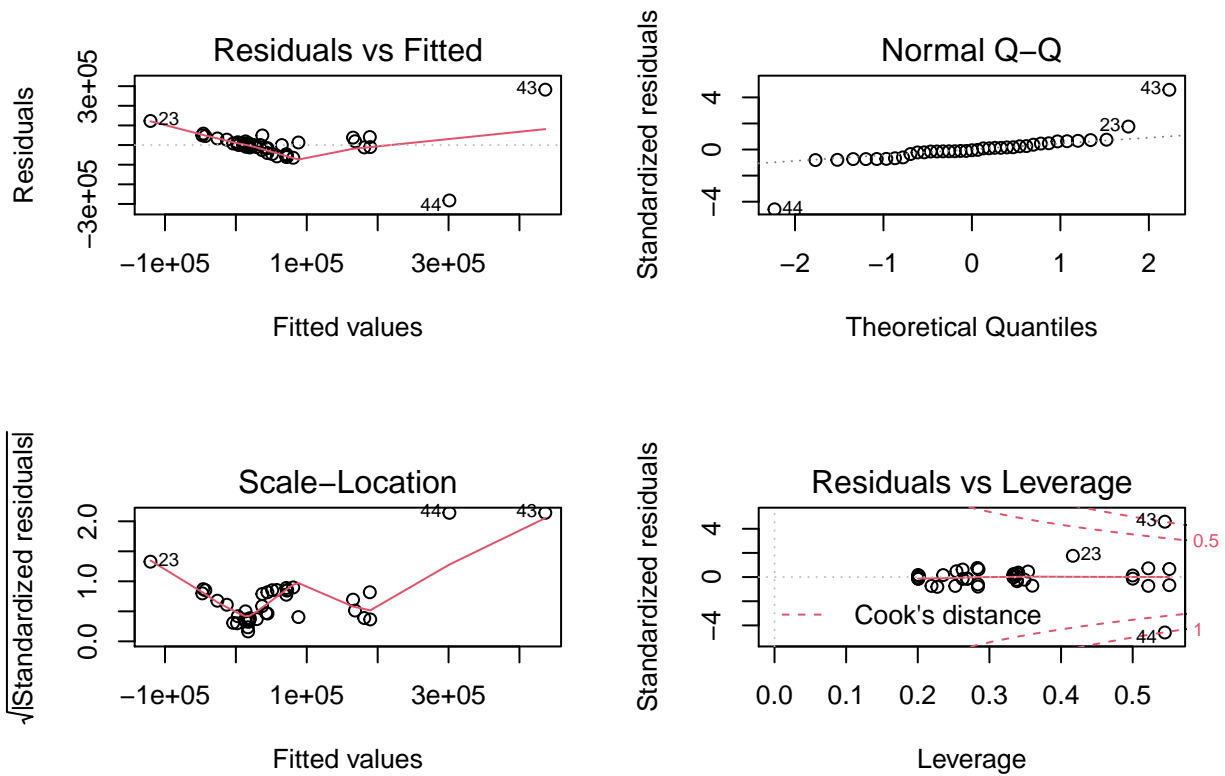
```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



```
fit3 <- lm(g3$votes ~ g3$disb + g3$state )
par(mfrow=c(2,2))
plot (fit3)
```

```
## Warning: not plotting observations with leverage one:
## 1, 7, 15, 16, 31, 32, 39, 40, 45, 46, 47, 51
```



```
summary(fit0)
```

```
##
## Call:
## lm(formula = g0$votes ~ g0$disb + g0$state + g0$can_party)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -415756  -39794   -5242    36879   269903
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.139e+05  4.120e+04   2.763  0.00585 **
## g0$disb        1.518e-02  1.661e-03   9.138 < 2e-16 ***
## g0$stateAL     4.215e+04  4.471e+04   0.943  0.34607
## g0$stateAR     4.851e+04  4.867e+04   0.997  0.31926
## g0$stateAS    -1.112e+05  5.312e+04  -2.094  0.03660 *
## g0$stateAZ     1.136e+04  4.377e+04   0.260  0.79531
## g0$stateCA    -1.482e+02  4.153e+04  -0.004  0.99715
## g0$stateCO     5.012e+04  4.395e+04   1.140  0.25447
## g0$stateCT    -2.964e+04  4.378e+04  -0.677  0.49848
## g0$stateDC     1.442e+05  7.125e+04   2.024  0.04325 *
## g0$stateDE     7.554e+04  5.815e+04   1.299  0.19425
## g0$stateFL     2.945e+04  4.184e+04   0.704  0.48167
## g0$stateGA     5.646e+04  4.335e+04   1.302  0.19316
## g0$stateGU    -1.001e+05  5.817e+04  -1.721  0.08558 .
## g0$stateHI     2.308e+04  5.312e+04   0.434  0.66406
## g0$stateIA     4.883e+04  4.597e+04   1.062  0.28847
```

```

## g0$stateID      5.123e+04  5.041e+04   1.016  0.30988
## g0$stateIL      4.114e+04  4.244e+04   0.969  0.33260
## g0$stateIN      2.570e+04  4.350e+04   0.591  0.55475
## g0$stateKS     -9.865e+03  4.547e+04  -0.217  0.82828
## g0$stateKY      5.083e+04  4.471e+04   1.137  0.25585
## g0$stateLA     -4.898e+04  4.324e+04  -1.133  0.25764
## g0$stateMA      7.588e+04  4.365e+04   1.738  0.08254 .
## g0$stateMD      4.579e+04  4.396e+04   1.042  0.29784
## g0$stateME      4.369e+04  5.037e+04   0.867  0.38598
## g0$stateMI      3.268e+04  4.254e+04   0.768  0.44261
## g0$stateMN      5.922e+04  4.386e+04   1.350  0.17730
## g0$stateMO      6.812e+04  4.419e+04   1.541  0.12358
## g0$stateMP      6.315e+03  7.178e+04   0.088  0.92991
## g0$stateMS      3.385e+04  4.749e+04   0.713  0.47613
## g0$stateMT      6.441e+04  5.838e+04   1.103  0.27026
## g0$stateNC      5.234e+04  4.274e+04   1.225  0.22102
## g0$stateND      2.870e+04  5.317e+04   0.540  0.58947
## g0$stateNE      3.875e+04  5.037e+04   0.769  0.44191
## g0$stateNH     -9.872e+03  4.869e+04  -0.203  0.83938
## g0$stateNJ      1.872e+04  4.295e+04   0.436  0.66310
## g0$stateNM      5.723e+03  4.865e+04   0.118  0.90638
## g0$stateNV     -5.810e+03  4.546e+04  -0.128  0.89835
## g0$stateNY     -7.894e+04  4.146e+04  -1.904  0.05726 .
## g0$stateOH      4.231e+04  4.248e+04   0.996  0.31949
## g0$stateOK      4.615e+04  4.867e+04   0.948  0.34321
## g0$stateOR      8.356e+04  4.662e+04   1.792  0.07343 .
## g0$statePA      5.607e+04  4.251e+04   1.319  0.18754
## g0$statePR      4.313e+05  5.342e+04   8.074  2.42e-15 ***
## g0$stateRI     -2.106e+04  5.037e+04  -0.418  0.67599
## g0$stateSC     -1.508e+04  4.336e+04  -0.348  0.72814
## g0$stateSD      4.393e+04  5.815e+04   0.756  0.45016
## g0$stateTN      2.770e+04  4.397e+04   0.630  0.52899
## g0$stateTX      1.064e+04  4.187e+04   0.254  0.79956
## g0$stateUT     -3.352e+03  4.597e+04  -0.073  0.94188
## g0$stateVA      4.654e+04  4.303e+04   1.082  0.27975
## g0$stateVI     -1.045e+05  7.125e+04  -1.466  0.14300
## g0$stateVT      1.387e+05  7.124e+04   1.947  0.05184 .
## g0$stateWA      2.911e+04  4.361e+04   0.667  0.50465
## g0$stateWI      4.214e+04  4.350e+04   0.969  0.33301
## g0$stateWV     -8.468e+03  4.665e+04  -0.182  0.85599
## g0$stateWY     -8.747e+03  5.316e+04  -0.165  0.86935
## g0$can_partyOther -1.104e+05  9.287e+03 -11.891 < 2e-16 ***
## g0$can_partyREP   1.661e+03  4.157e+03   0.400  0.68957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58140 on 821 degrees of freedom
## Multiple R-squared:  0.5116, Adjusted R-squared:  0.477
## F-statistic: 14.82 on 58 and 821 DF,  p-value: < 2.2e-16

```

```
summary(fit1)
```

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

```

## lm(formula = g1$votes ~ g1$disb + g1$state)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -2.815e+10 -8.813e+09 -3.549e+08  6.976e+09  4.558e+10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.780e+10  1.339e+10  -2.824  0.00502 **
## g1$disb      4.389e+09  3.491e+08  12.570 < 2e-16 ***
## g1$stateAL   2.081e+10  1.345e+10   1.548  0.12261
## g1$stateAR   1.472e+10  1.392e+10   1.057  0.29121
## g1$stateAS  -1.141e+10  1.764e+10  -0.647  0.51798
## g1$stateAZ   7.702e+09  1.321e+10   0.583  0.56021
## g1$stateCA  -1.213e+09  1.262e+10  -0.096  0.92352
## g1$stateCO   1.745e+10  1.332e+10   1.310  0.19091
## g1$stateCT  -2.819e+09  1.348e+10  -0.209  0.83451
## g1$stateDE   1.347e+10  1.762e+10   0.765  0.44500
## g1$stateFL   1.699e+10  1.270e+10   1.337  0.18193
## g1$stateGA   1.724e+10  1.296e+10   1.330  0.18436
## g1$stateGU  -1.353e+10  1.763e+10  -0.767  0.44343
## g1$stateIA   1.872e+10  1.392e+10   1.345  0.17955
## g1$stateID   2.803e+10  1.525e+10   1.838  0.06692 .
## g1$stateIL   1.186e+10  1.301e+10   0.912  0.36251
## g1$stateIN   1.220e+10  1.321e+10   0.924  0.35610
## g1$stateKS  -1.339e+09  1.364e+10  -0.098  0.92184
## g1$stateKY   2.486e+10  1.331e+10   1.867  0.06273 .
## g1$stateLA  -8.544e+09  1.289e+10  -0.663  0.50788
## g1$stateMA   2.607e+09  1.396e+10   0.187  0.85204
## g1$stateMD   4.015e+09  1.346e+10   0.298  0.76563
## g1$stateME   1.164e+10  1.525e+10   0.763  0.44567
## g1$stateMI   1.207e+10  1.301e+10   0.928  0.35396
## g1$stateMN   1.309e+10  1.331e+10   0.983  0.32633
## g1$stateMO   2.717e+10  1.345e+10   2.020  0.04412 *
## g1$stateMS   2.436e+10  1.398e+10   1.742  0.08234 .
## g1$stateMT   5.069e+10  1.762e+10   2.878  0.00425 **
## g1$stateNC   1.711e+10  1.292e+10   1.323  0.18653
## g1$stateND   3.177e+10  1.761e+10   1.804  0.07201 .
## g1$stateNE   1.409e+10  1.438e+10   0.980  0.32792
## g1$stateNH  -8.923e+09  1.525e+10  -0.585  0.55887
## g1$stateNJ   5.136e+09  1.313e+10   0.391  0.69583
## g1$stateNM  -2.122e+09  1.526e+10  -0.139  0.88946
## g1$stateNV  -2.764e+09  1.392e+10  -0.199  0.84277
## g1$stateNY  -1.740e+10  1.257e+10  -1.384  0.16710
## g1$stateOH   2.203e+10  1.289e+10   1.709  0.08835 .
## g1$stateOK   1.505e+10  1.392e+10   1.081  0.28045
## g1$stateOR   2.319e+10  1.438e+10   1.612  0.10787
## g1$statePA   1.731e+10  1.286e+10   1.346  0.17919
## g1$stateRI   7.234e+07  1.533e+10   0.005  0.99624
## g1$stateSC   1.359e+10  1.345e+10   1.010  0.31296
## g1$stateSD   2.849e+10  1.761e+10   1.618  0.10655
## g1$stateTN   1.653e+10  1.315e+10   1.257  0.20954
## g1$stateTX   6.193e+09  1.266e+10   0.489  0.62515
## g1$stateUT   7.225e+09  1.392e+10   0.519  0.60410

```

```
## g1$stateVA 1.569e+10 1.306e+10 1.201 0.23042
## g1$stateWA 5.199e+09 1.348e+10 0.386 0.69987
## g1$stateWI 2.383e+10 1.345e+10 1.772 0.07729 .
## g1$stateWV -3.932e+08 1.438e+10 -0.027 0.97820
## g1$stateWY -1.792e+09 1.761e+10 -0.102 0.91901
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.245e+10 on 355 degrees of freedom
## Multiple R-squared: 0.6461, Adjusted R-squared: 0.5962
## F-statistic: 12.96 on 50 and 355 DF, p-value: < 2.2e-16
```

```
summary(fit2)
```

```
##
## Call:
## lm(formula = g2$votes ~ g2$disb + g2$state)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.538e+10 -1.073e+10 -2.444e+09  7.456e+09  6.188e+10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.076e+10  1.600e+10  -1.922 0.055324 .
## g2$disb      3.097e+09   3.868e+08   8.008 1.56e-14 ***
## g2$stateAL   1.405e+10  1.654e+10   0.850 0.396151
## g2$stateAR   1.135e+10  2.136e+10   0.531 0.595464
## g2$stateAS   4.557e+08  1.852e+10   0.025 0.980386
## g2$stateAZ   1.033e+10  1.612e+10   0.641 0.522153
## g2$stateCA   1.432e+10  1.520e+10   0.942 0.346879
## g2$stateCO   2.555e+10  1.611e+10  1.586 0.113565
## g2$stateCT   6.029e+09  1.588e+10   0.380 0.704452
## g2$stateDC   6.065e+10  2.131e+10  2.846 0.004681 **
## g2$stateDE   4.136e+10  2.131e+10  1.941 0.053040 .
## g2$stateFL   1.665e+10  1.536e+10   1.084 0.278957
## g2$stateGA   3.523e+10  1.632e+10  2.158 0.031575 *
## g2$stateGU  -6.583e+09  2.132e+10  -0.309 0.757684
## g2$stateHI   1.150e+10  1.740e+10   0.661 0.508951
## g2$stateIA   1.720e+10  1.685e+10   1.021 0.307958
## g2$stateID   4.271e+09  2.131e+10   0.200 0.841305
## g2$stateIL   2.302e+10  1.553e+10  1.482 0.139199
## g2$stateIN   1.132e+10  1.628e+10   0.695 0.487225
## g2$stateKS   2.107e+09  1.741e+10   0.121 0.903706
## g2$stateKY   1.378e+10  1.687e+10   0.817 0.414583
## g2$stateLA   6.148e+09  1.654e+10   0.372 0.710317
## g2$stateMA   5.687e+10  1.588e+10  3.580 0.000389 ***
## g2$stateMD   3.498e+10  1.611e+10  2.172 0.030530 *
## g2$stateME   2.645e+10  1.845e+10  1.433 0.152633
## g2$stateMI   1.801e+10  1.564e+10  1.151 0.250297
## g2$stateMN   2.327e+10  1.685e+10  1.381 0.168017
## g2$stateMO   1.891e+10  1.652e+10  1.145 0.253021
## g2$stateMS   1.375e+10  1.846e+10   0.745 0.456887
## g2$stateMT   2.729e+10  2.131e+10  1.280 0.201256
```

```
## g2$stateNC 2.698e+10 1.572e+10 1.716 0.087011 .
## g2$stateND -7.935e+08 2.132e+10 -0.037 0.970329
## g2$stateNE 4.028e+09 2.131e+10 0.189 0.850209
## g2$stateNH 1.443e+10 1.846e+10 0.782 0.434738
## g2$stateNJ 1.708e+10 1.575e+10 1.085 0.278757
## g2$stateNM 1.056e+10 1.740e+10 0.607 0.544415
## g2$stateNV 4.560e+09 1.685e+10 0.271 0.786798
## g2$stateNY -1.247e+09 1.519e+10 -0.082 0.934636
## g2$stateOH 1.800e+10 1.563e+10 1.151 0.250283
## g2$stateOK 6.966e+09 2.134e+10 0.326 0.744237
## g2$stateOR 4.134e+10 1.685e+10 2.454 0.014594 *
## g2$statePA 2.885e+10 1.560e+10 1.849 0.065201 .
## g2$statePR 4.752e+11 2.132e+10 22.293 < 2e-16 ***
## g2$stateRI 5.501e+09 1.845e+10 0.298 0.765818
## g2$stateSC 3.419e+09 1.571e+10 0.218 0.827864
## g2$stateSD 8.423e+09 2.131e+10 0.395 0.692931
## g2$stateTN 1.212e+10 1.687e+10 0.718 0.473015
## g2$stateTX 9.048e+09 1.541e+10 0.587 0.557360
## g2$stateUT -2.441e+08 1.685e+10 -0.014 0.988453
## g2$stateVA 2.226e+10 1.574e+10 1.414 0.158272
## g2$stateVI -8.469e+09 2.131e+10 -0.397 0.691344
## g2$stateVT 5.865e+10 2.131e+10 2.752 0.006209 **
## g2$stateWA 2.067e+10 1.580e+10 1.308 0.191848
## g2$stateWI 2.545e+10 1.599e+10 1.592 0.112293
## g2$stateWV 1.758e+09 1.743e+10 0.101 0.919686
## g2$stateWY -1.132e+09 2.132e+10 -0.053 0.957693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.507e+10 on 367 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.7438
## F-statistic: 23.27 on 55 and 367 DF, p-value: < 2.2e-16
```

```
summary(fit3)
```

```
##
## Call:
## lm(formula = g3$votes ~ g3$disb + g3$state)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -282119  -11977         0   13659  282119
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -153401    110453  -1.389  0.17667
## g3$disb         22085       6306   3.502  0.00169 **
## g3$stateFL    -54719       99904  -0.548  0.58856
## g3$stateID     18427      130043   0.142  0.88841
## g3$stateIL    -2650      102013  -0.026  0.97947
## g3$stateIN   -26225      105545  -0.248  0.80572
## g3$stateKS   -78968      129323  -0.611  0.54675
## g3$stateMA   -45073      105309  -0.428  0.67217
## g3$stateMD   -44706      129001  -0.347  0.73171
```



```
## g3$stateMI      32926      101862    0.323  0.74909
## g3$stateMN      14323      103691    0.138  0.89120
## g3$stateMO      31203      111852    0.279  0.78248
## g3$stateMP     -77640      129149   -0.601  0.55294
## g3$stateND       3729      129615    0.029  0.97727
## g3$stateNH    -132536      132253   -1.002  0.32551
## g3$stateNJ     -27077      112109   -0.242  0.81104
## g3$stateNV     -56156      128984   -0.435  0.66689
## g3$stateNY    -104945      103055   -1.018  0.31790
## g3$stateOH     -48102      111696   -0.431  0.67027
## g3$statePR     286648      111806    2.564  0.01648 *
## g3$stateTN      30708      131593    0.233  0.81731
## g3$stateTX     -1954      129482   -0.015  0.98807
## g3$stateWI     -8531      105822   -0.081  0.93636
## g3$stateWV    -81526      129332   -0.630  0.53396
## g3$stateWY      23580      130886    0.180  0.85843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 91200 on 26 degrees of freedom
## Multiple R-squared:  0.6474, Adjusted R-squared:  0.322
## F-statistic: 1.989 on 24 and 26 DF, p-value: 0.04476
```

```
#d2$disb <- log(d2$tll_disb)
#d2$votes <- log(d2$general_votes)

write.csv(d2, "d2.csv")

#d2[which(!is.finite(d2))] <- 0
#d2 <- d2[is.finite(rowSums(d2)),]
d2[d2 == -Inf] <- 0

#data_new <- d2                                # Duplicate data

#d2[is.na(d2$disb) | d2$disb == "Inf"] <- NA # Replace NaN & Inf with NA

#d3 <- data_new

head(d2)
```

```
## # A tibble: 6 x 9
##   can_party general_votes tll_disb state lamvotes  tvotes  tdisb  disb votes
##   <chr>          <dbl>    <dbl> <chr>    <dbl>    <dbl>    <dbl> <dbl> <dbl>
## 1 REP            208083 1172750. AL      5458.    0.997    0.0789 14.0   12.2
## 2 REP            134886 1850536. AL      4082.   -0.0418    0.999   14.4   11.8
## 3 DEM            112089  36844 AL      3605.   -0.348   -0.938   10.5   11.6
## 4 REP            192164 1071289. AL      5174.    1.00   -0.0154 13.9   12.2
## 5 DEM             94549   7348 AL      3217.  -0.524   -0.852    8.90  11.5
## 6 REP            235925 1394461. AL      5937.    0.981    0.196   14.1   12.4
```

```
head(d2$disb)
```

```
## [1] 13.974862 14.430986 10.514448 13.884374  8.902183 14.148019
```

```
#d3<-d3%>%na.omit()
```

```
fit <- lm(d2$general_votes ~ d2$disb)
```

```
summary(fit)
```

```
##
```

```
## Call:
```

```
## lm(formula = d2$general_votes ~ d2$disb)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -170750  -34066    7653   45029  568412
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   -46697     14420   -3.238  0.00125 **  
## d2$disb         14339       1109  12.928 < 2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 73740 on 878 degrees of freedom
```

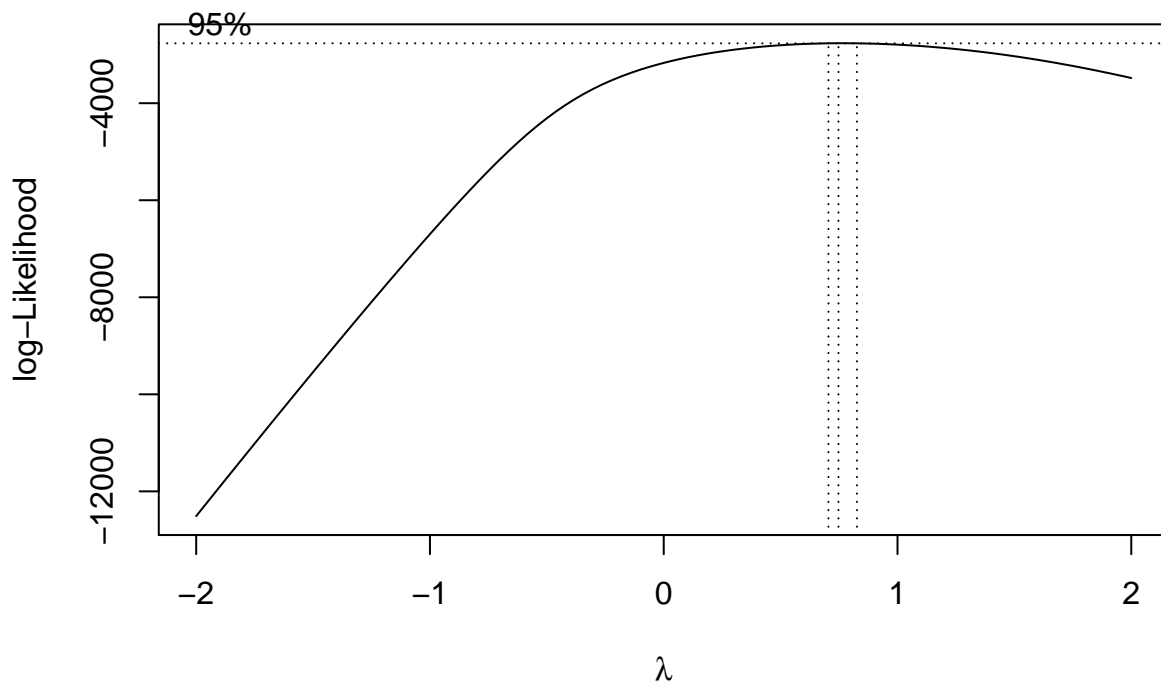
```
## Multiple R-squared:  0.1599, Adjusted R-squared:  0.159
```

```
## F-statistic: 167.1 on 1 and 878 DF,  p-value: < 2.2e-16
```

```
## boxcox test
```

```
library(MASS)
```

```
boxcox(general_votes~poly(disb,2),  
      data = d2)
```



```

g0 <- d2
g0$votes <- log10(g0$general_votes)
g0$disb <- log10(g0$t1_disb)
g0[g0 == -Inf] <- 0

g1 <- filter(d2, can_party == "REP")
g1$votes <- g1$general_votes*g1$general_votes
g1$disb <- log(g1$t1_disb)
g1[g1 == -Inf] <- 0

g2 <- filter(d2, can_party == "DEM")
g2$votes <- g2$general_votes*g2$general_votes
g2$disb <- log(g2$t1_disb)
g2[g2 == -Inf] <- 0

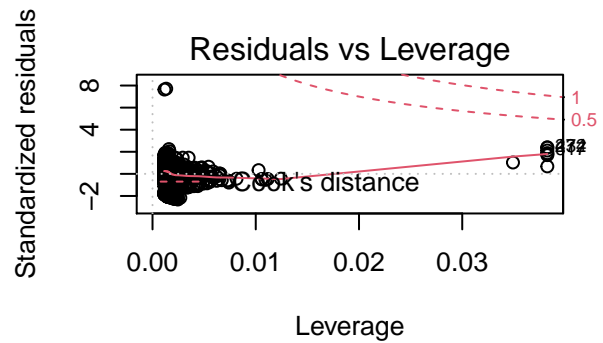
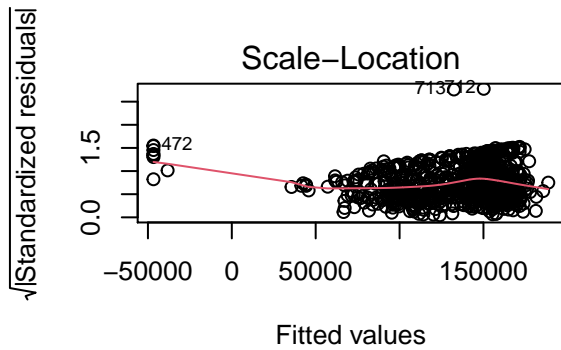
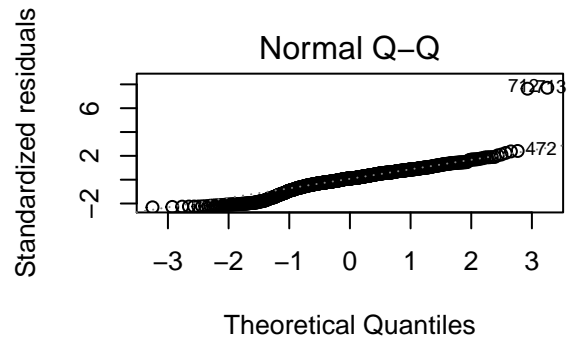
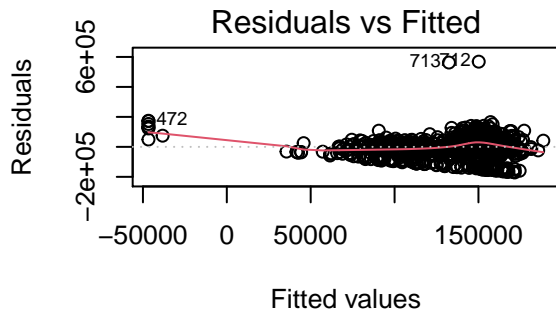
g3 <- filter(d2, can_party == "Other")
g3$votes <- g3$general_votes
g3$disb <- log(g3$t1_disb)
g3[g3 == -Inf] <- 0

write.csv(g1, "g1.csv")
write.csv(g2, "g2.csv")
write.csv(g3, "g3.csv")

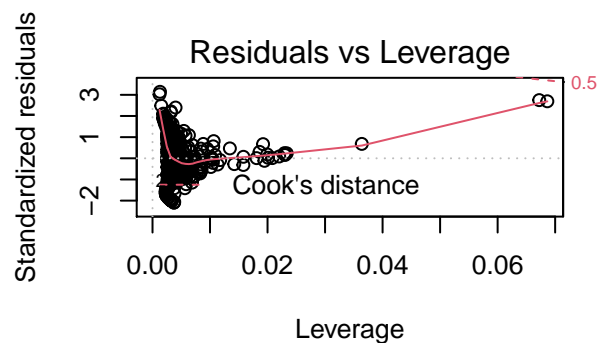
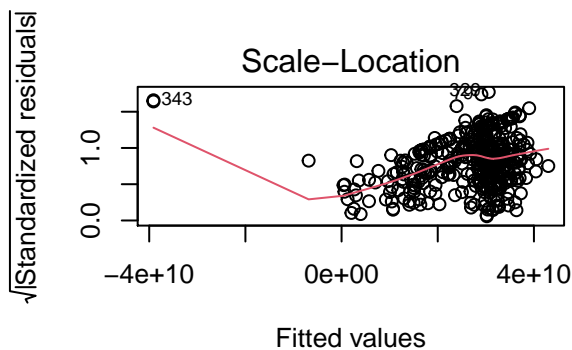
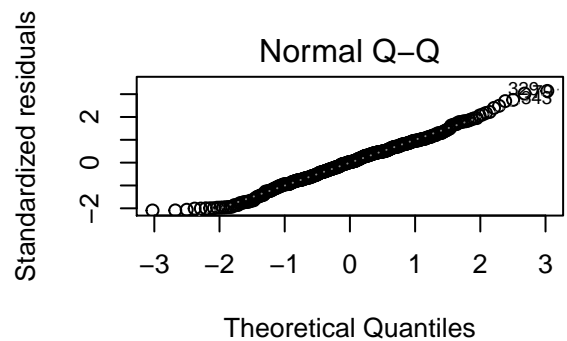
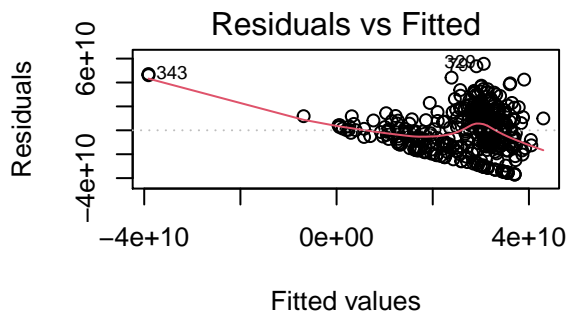
fit0 <- rlm(g0$votes ~ g0$disb)
par(mfrow=c(2,2))
plot (fit)

fit1 <- rlm(g1$votes ~ g1$disb)
par(mfrow=c(2,2))
plot (fit)

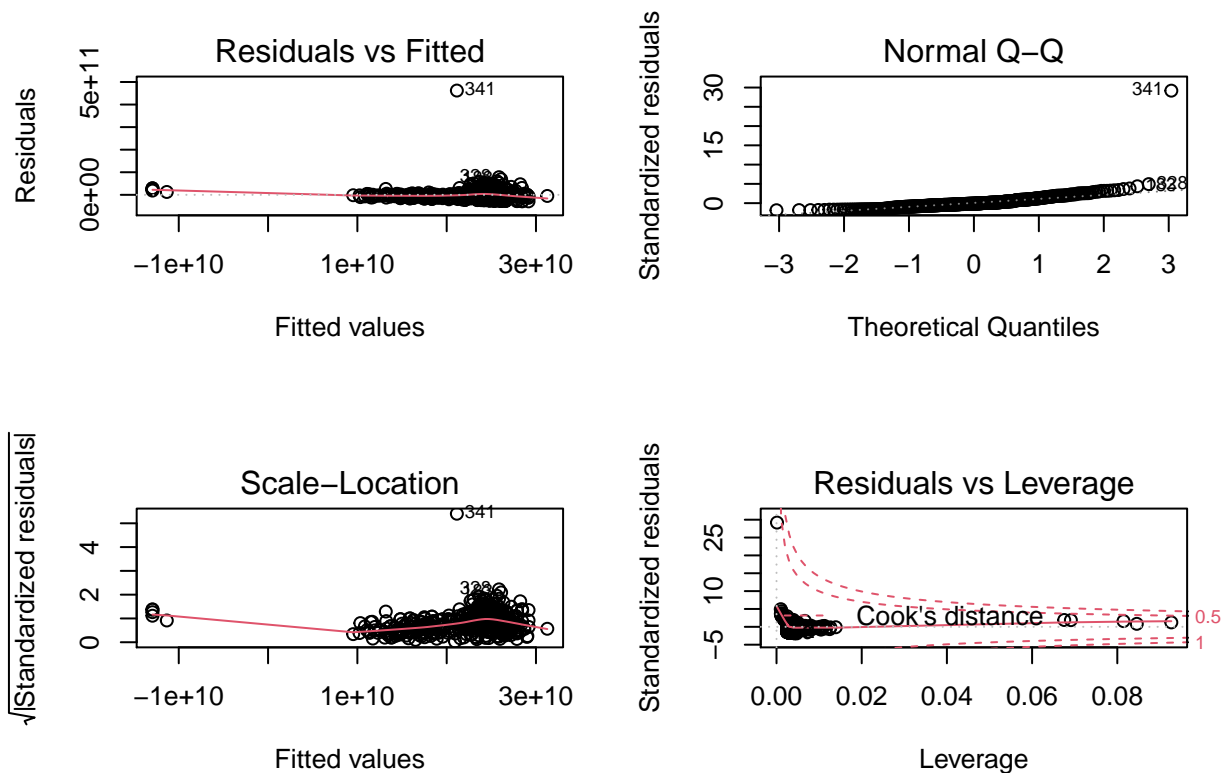
```



```
fit2 <- rlm(g2$votes ~ g2$disb)
par(mfrow=c(2,2))
plot(fit1)
```



```
fit3 <- rlm(g3$votes ~ g3$disb)
par(mfrow=c(2,2))
plot(fit2)
```



```
summary(fit0)
```

```
##
## Call: rlm(formula = g0$votes ~ g0$disb)
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.36130 -0.12353  0.03335  0.13016  0.88942
##
## Coefficients:
##              Value      Std. Error t value
## (Intercept)  4.2162     0.0414    101.8323
## g0$disb      0.1609     0.0073     21.9398
##
## Residual standard error: 0.1917 on 878 degrees of freedom
```

```
summary(fit1)
```

```
##
## Call: rlm(formula = g1$votes ~ g1$disb)
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.704e+10 -1.201e+10  4.285e+08  1.181e+10  5.564e+10
##
```

```
## Coefficients:
##           Value      Std. Error  t value
## (Intercept) -3.911640e+10  6.067250e+09 -6.447100e+00
## g1$disb      5.002444e+09  4.558440e+08  1.097400e+01
##
## Residual standard error: 1.772e+10 on 404 degrees of freedom
```

```
summary(fit2)
```

```
##
## Call: rlm(formula = g2$votes ~ g2$disb)
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.893e+10 -9.863e+09 -2.154e+09  1.217e+10  4.620e+11
##
## Coefficients:
##           Value      Std. Error  t value
## (Intercept) -1.294200e+10  5.318627e+09 -2.433300e+00
## g2$disb      2.728824e+09  4.089823e+08  6.672200e+00
##
## Residual standard error: 1.583e+10 on 421 degrees of freedom
```

```
summary(fit3)
```

```
##
## Call: rlm(formula = g3$votes ~ g3$disb)
## Residuals:
##      Min       1Q   Median       3Q      Max
## -30841 -10191  -2653   11235  682215
##
## Coefficients:
##           Value      Std. Error  t value
## (Intercept) -15663.5931    8732.4093   -1.7937
## g3$disb      3790.2361     858.8929    4.4129
##
## Residual standard error: 15330 on 49 degrees of freedom
```

6. (3 points) Interpret the model coefficients you estimate.

- Tasks to keep in mind as you're writing about your model:
  - At the time that you're writing and interpreting your regression coefficients you'll be *deep* in the analysis. Nobody will know more about the data than you do, at that point. *So, although it will feel tedious, be descriptive and thorough in describing your observations.*
  - It can be hard to strike the balance between: on the one hand, writing enough of the technical underpinnings to know that your model meets the assumptions that it must; and, on the other hand, writing little enough about the model assumptions that the implications of the model can still be clear. We're starting this practice now, so that by the end of Lab 2 you will have had several chances to strike this balance.

```
#lm(d2$general_votes ~ b1*d2$ttl_disb + b2)
```