Homework 2

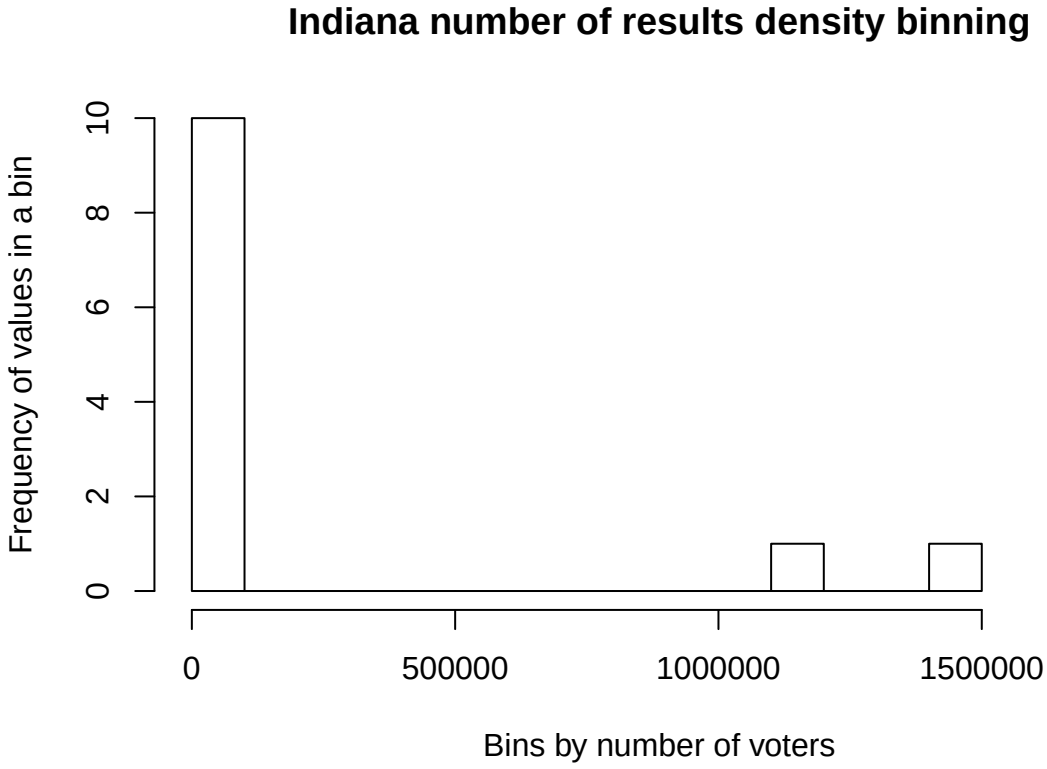
Dima Galat

23 October 2016

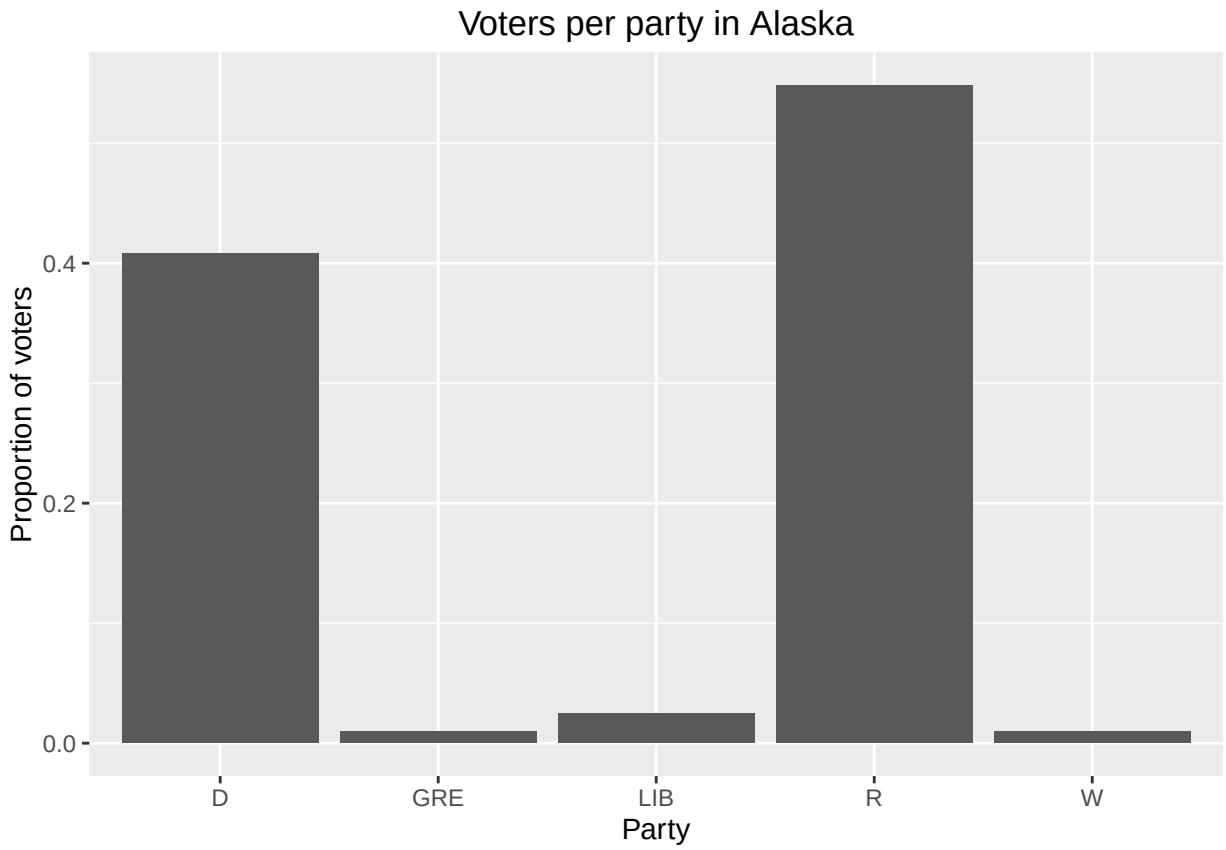
Question 1

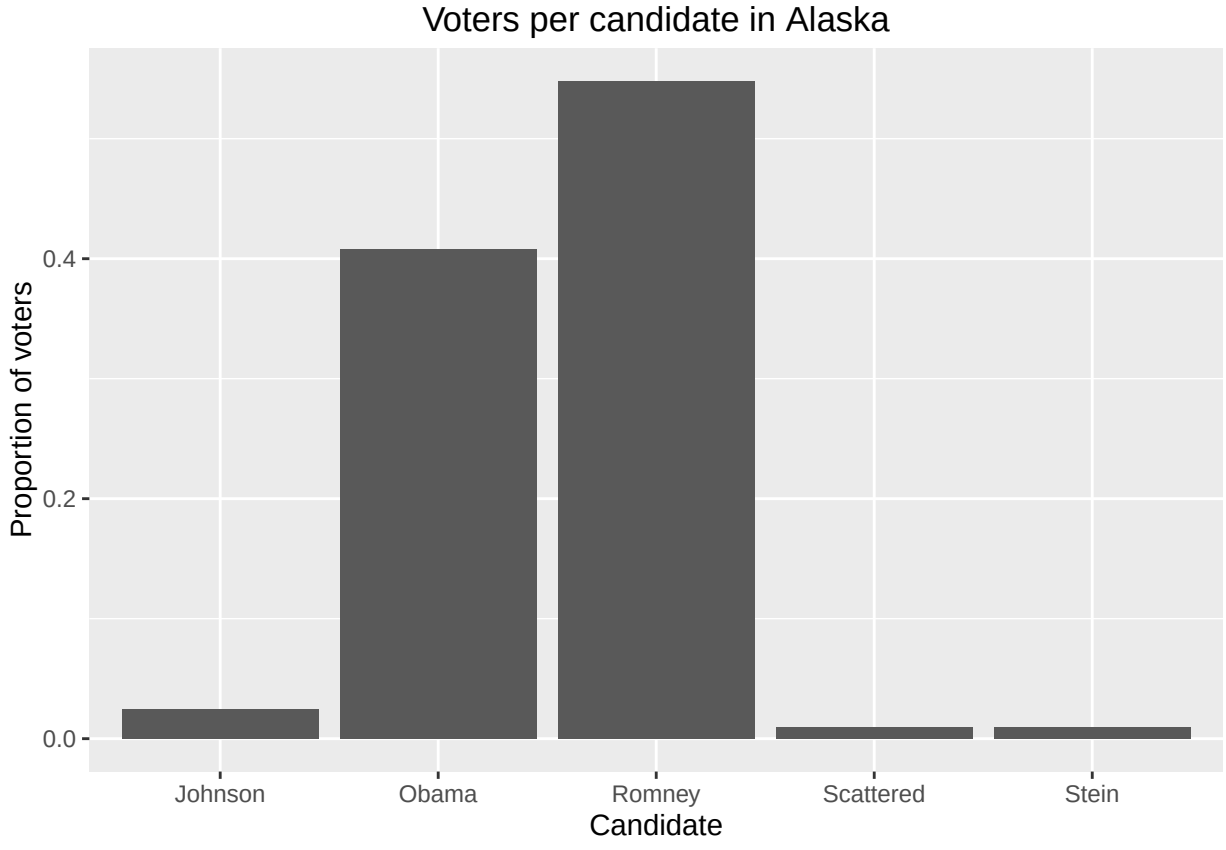
Creating histograms for quantitative variables

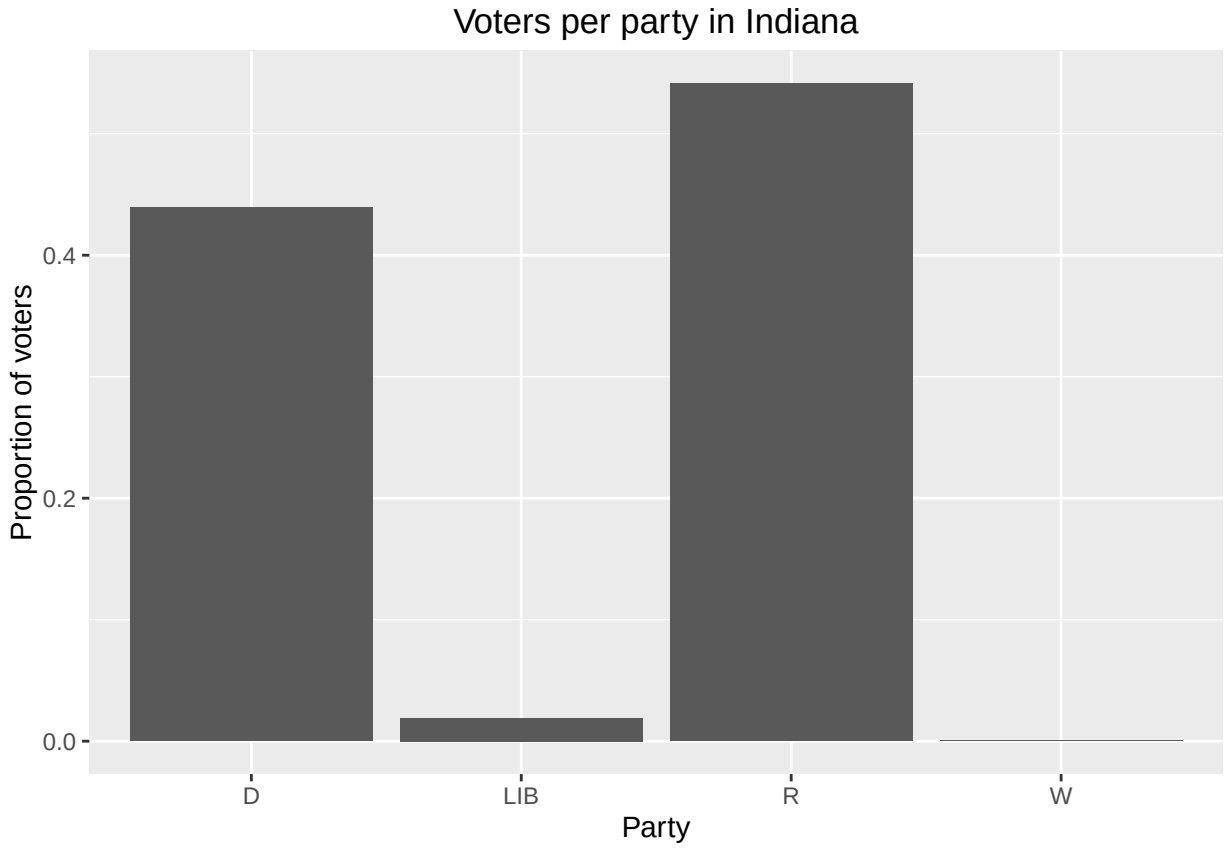
- This is the histogram that you requested:

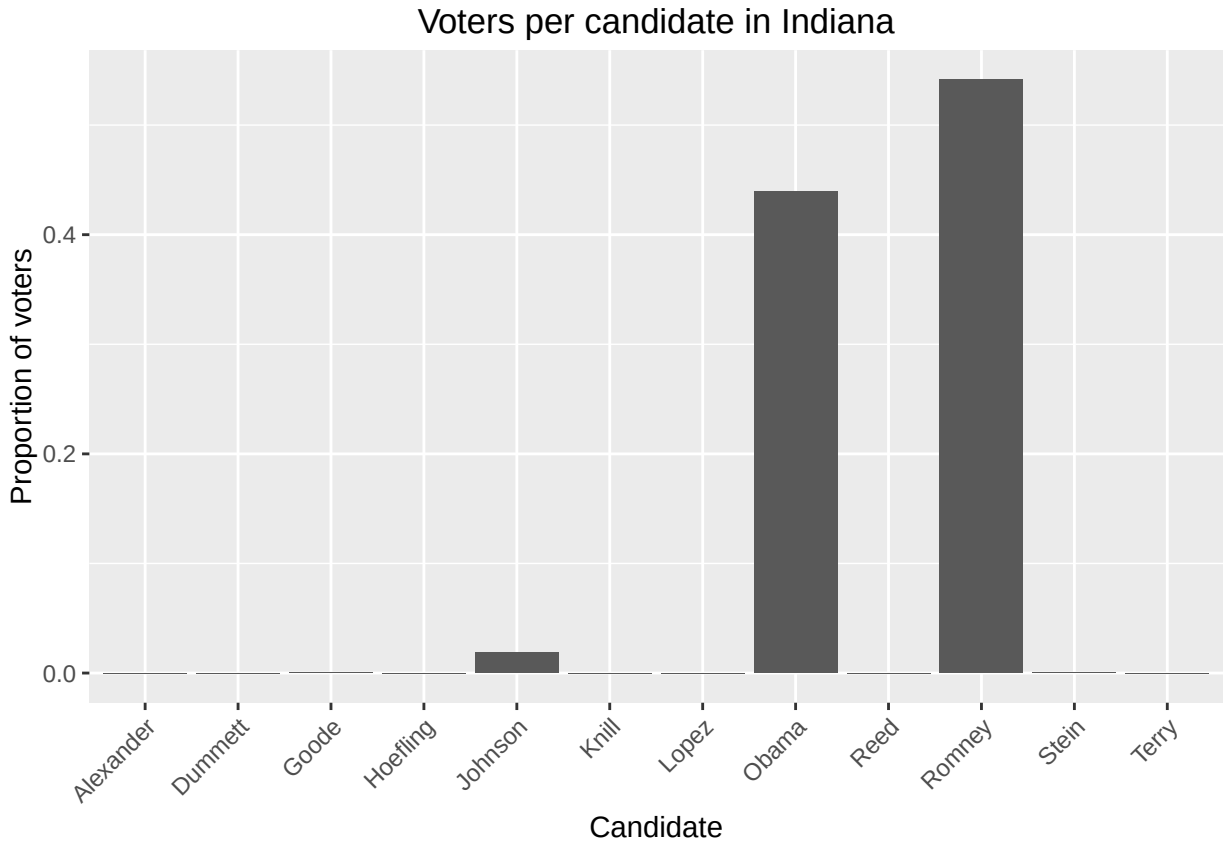


- These are boxplots for categorical variables that we have (party and candidate):


Voters per party in Alaska
Proportion of voters
0.4
0.2
0.0
D
GRE
LIB
R
W
Party

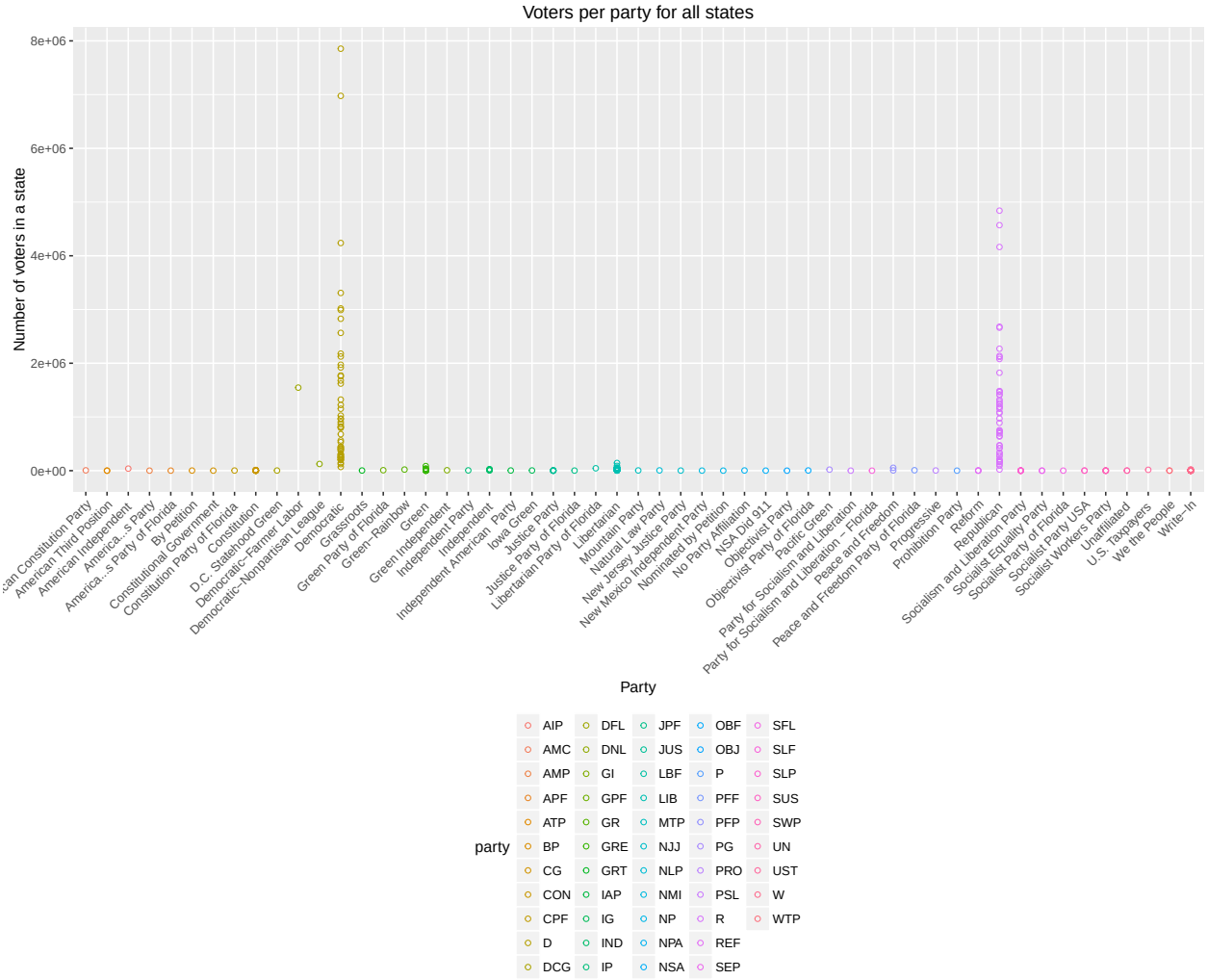

Voters per candidate in Alaska
Proportion of voters
0.4
0.2
0.0
Johnson
Obama
Romney
Scattered
Stein
Candidate

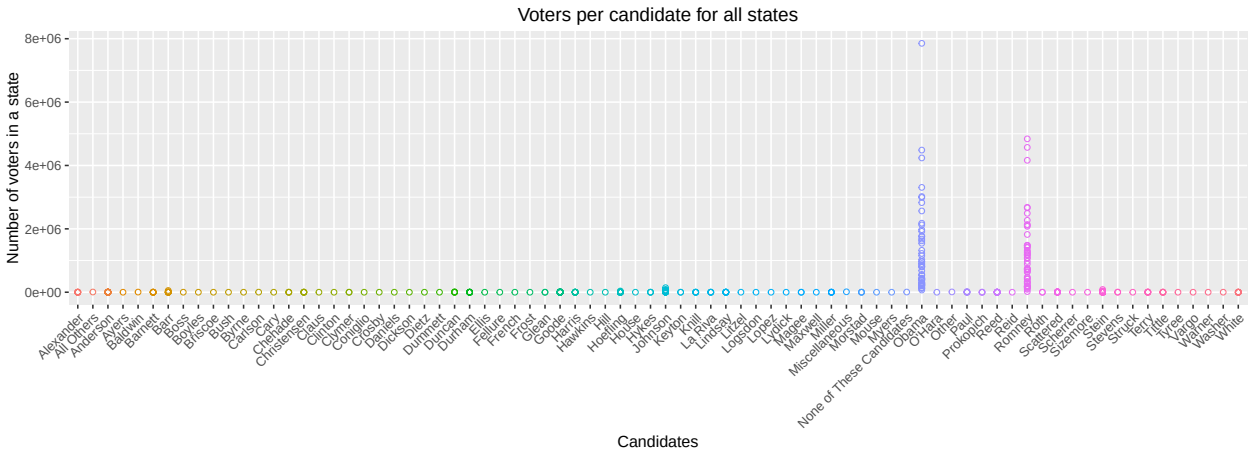

Voters per party in Indiana
Proportion of voters
0.4
0.2
0.0
D
LIB
R
W
Party


Voters per candidate in Indiana
Proportion of voters
0.4
0.2
0.0
Alexander
Dunmett
Goode
Hoefling
Johnson
Knill
Lopez
Obama
Reed
Romney
Stein
Terry
Candidate

Question 2

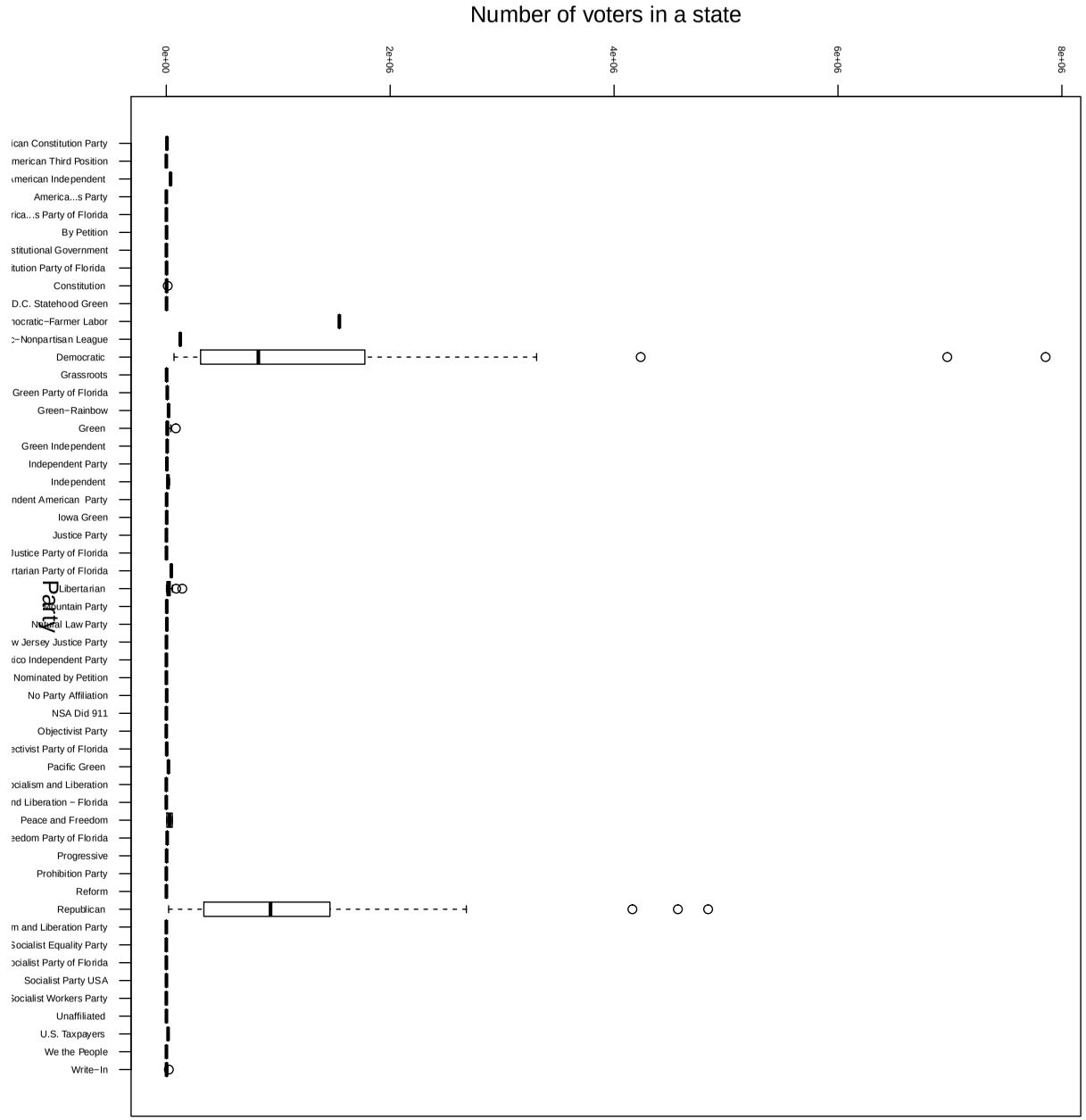
Creating box and scatter plots for quantitative variables

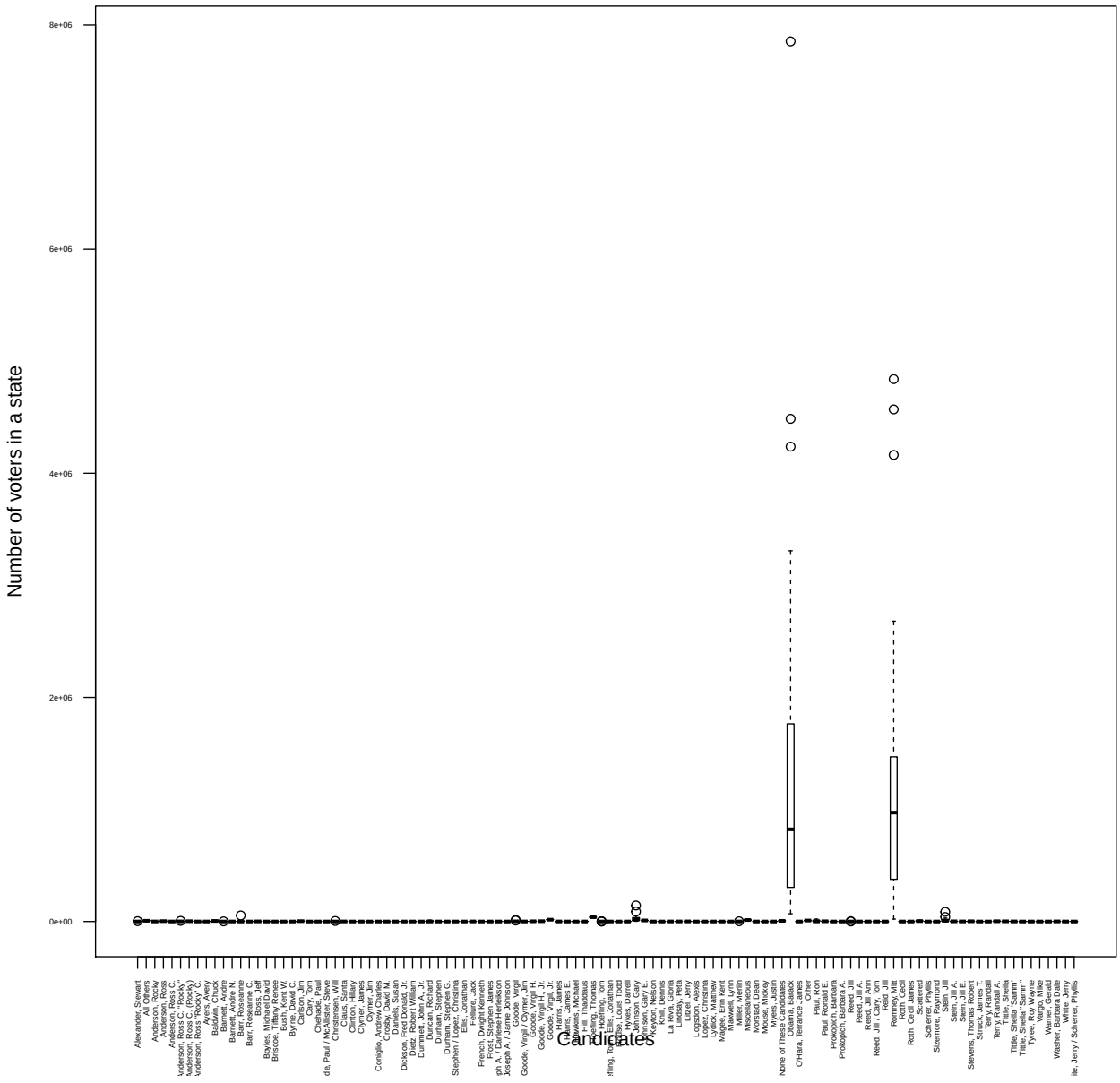




q2$'LAST NAME'

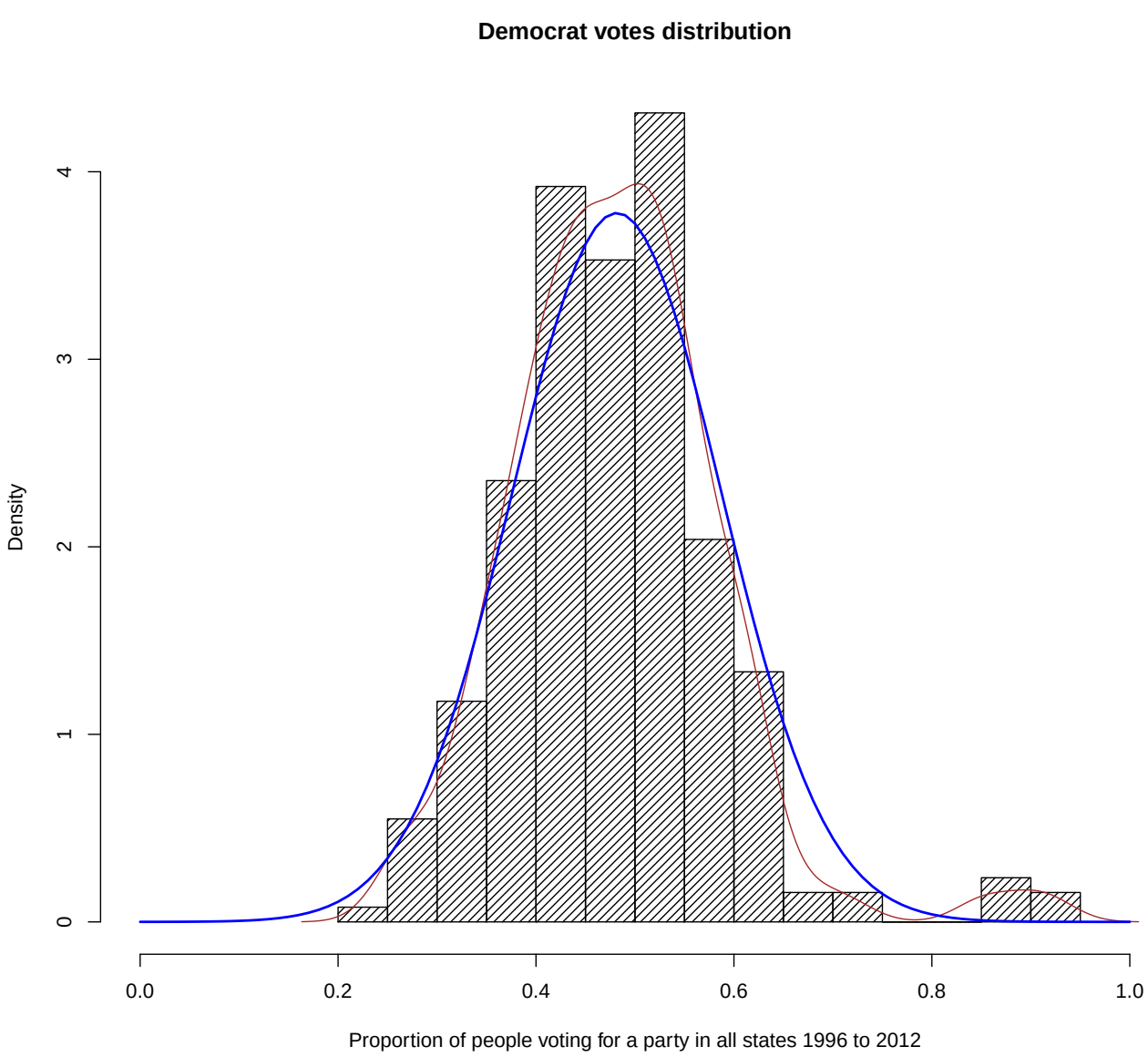
- Alexander, Stewart
- All Others
- Anderson, Rocky
- Anderson, Ross
- Anderson, Ross C.
- Anderson, Ross C. "Rocky"
- Anderson, Ross C. (Rocky)
- Anderson, Ross "Rocky" C.
- Ayers, Avery
- Baldwin, Chuck
- Barnett, Andre
- Barnett, Andre N.
- Barr, Roseanne
- Barr, Roseanne C.
- Boss, Jeff
- Boyles, Michael David
- Briscoe, Tiffany Renee
- Bush, Kent W.
- Byrne, David C.
- Carlson, Jim
- Cary, Tom
- Chehade, Paul
- Chehade, Paul / McAllister, Steve
- Christensen, Will
- Claus, Santa
- Clinton, Hillary
- Clymer, James
- Clymer, Jim
- Coniglio, Andrew Charles
- Crosby, David M.
- Daniels, Susan
- Dickson, Fred Donald, Jr.
- Dietz, Robert William
- Dummett, John A., Jr.
- Duncan, Richard
- Durham, Stephen
- Durham, Stephen G.
- Durham, Stephen / Lopez, Christina
- Ellis, Jonathan
- Fellure, Jack
- French, Dwight Kenneth
- Frost, Stephen James
- Glean, Joseph A. / Darlene Herleikson
- Glean, Joseph A. / Jamie Johnson
- Goode, Virgil
- Goode, Virgil / Clymer, Jim
- Goode, Virgil H.
- Goode, Virgil H., Jr.
- Goode, Virgil, Jr.
- Harris, James
- Harris, James E.
- Hawkins, Michael
- Hill, Thaddaus
- Hoefling, Thomas
- Hoefling, Tom
- Hoefling, Tom / Ellis, Jonathan
- House, Louis Todd
- Hykes, Darrell
- Johnson, Gary
- Johnson, Gary E.
- Keyton, Nelson
- Knill, Dennis
- La Riva, Gloria
- Lindsay, Peta
- Litzel, Jerry
- Logsdon, Alexis
- Lopez, Christina
- Lydick, Matthew
- Magee, Erin Kent
- Maxwell, Lynn
- Miller, Merlin
- Miscellaneous
- Morstad, Dean
- Mouse, Mickey
- Myers, Justin
- None of These Candidates
- Obama, Barack
- O'Hara, Terrance James
- Other
- Paul, Ron
- Paul, Ronald E.
- Prokopich, Barbara
- Prokopich, Barbara A.
- Reed, Jill
- Reed, Jill A.
- Reed, Jill Ann
- Reed, Jill / Cary, Tom
- Reid, Jill
- Romney, Mitt
- Roth, Cecil
- Roth, Cecil James
- Scattered
- Scherrer, Phyllis
- Sizemore, Raymond
- Stein, Jill
- Stein, Jill A.
- Stein, Jill E.
- Stevens, Thomas Robert
- Struck, James T.
- Terry, Randall
- Terry, Randall A.
- Tittle, Sheila
- Tittle, Sheila 'Samm'
- Tittle, Sheila "Samm"
- Tyree, Roy Wayne
- Vargo, Mike
- Warner, Gerald
- Washer, Barbara Dale
- White, Jerry
- White, Jerry / Scherrer, Phyllis

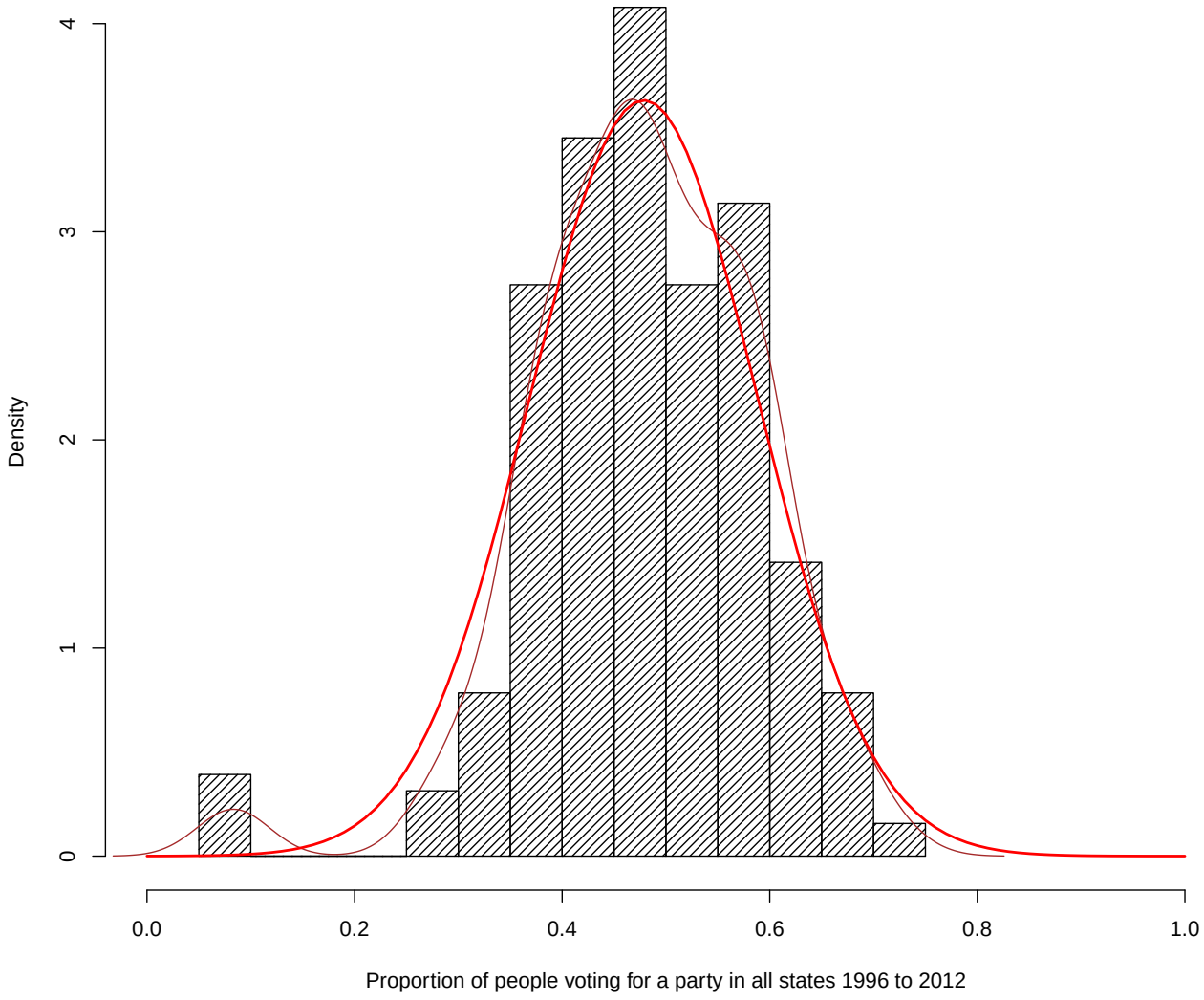

Number of voters in a state
0e+00
2e+06
4e+06
6e+06
8e+06
ican Constitution Party
merican Third Position
\merican Independent
America...s Party
rica...s Party of Florida
By Petition
stitutional Government
itution Party of Florida
Constitution
D.C. Statehood Green
nocratic−Farmer Labor
c−Nonpartisan League
Democratic
Grassroots
Green Party of Florida
Green−Rainbow
Green
Green Independent
Independent Party
Independent
ndent American Party
Iowa Green
Justice Party
Justice Party of Florida
rtarian Party of Florida
Libertarian
ountain Party
Natural Law Party
w Jersey Justice Party
xico Independent Party
Nominated by Petition
No Party Affiliation
NSA Did 911
Objectivist Party
ectivist Party of Florida
Pacific Green
ocialism and Liberation
nd Liberation − Florida
Peace and Freedom
eedom Party of Florida
Progressive
Prohibition Party
Reform
Republican
m and Liberation Party
Socialist Equality Party
ocialist Party of Florida
Socialist Party USA
Socialist Workers Party
Unaffiliated
U.S. Taxpayers
We the People
Write−In
Party

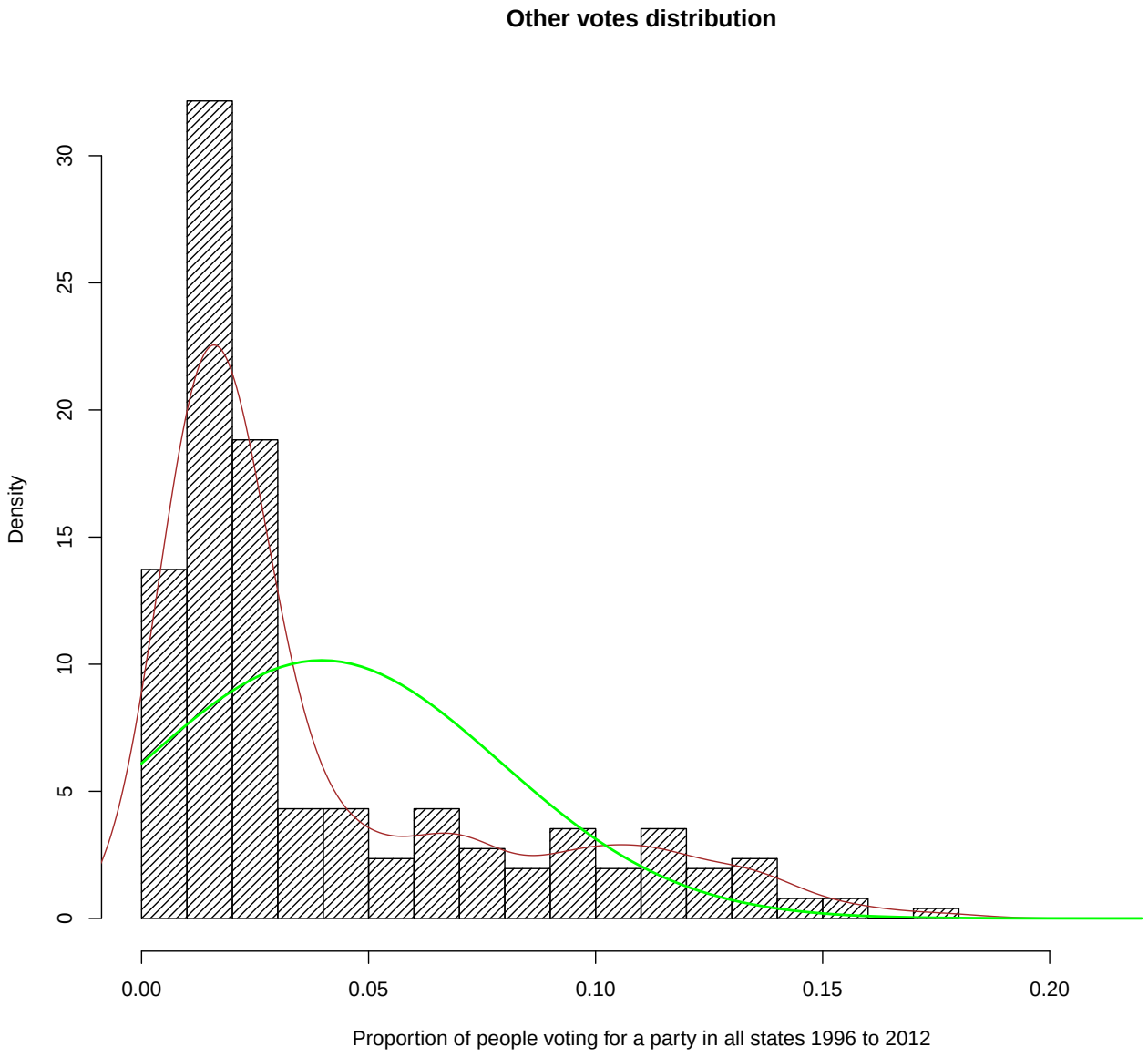


Question 3

Create an overlay based on party red and blue lines as shown in Lynda.com exercise


Democrat votes distribution
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of people voting for a party in all states 1996 to 2012


Republican votes distribution
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of people voting for a party in all states 1996 to 2012



Question 4

Calculate the frequencies of wins for each party win with all states.

```
## Number of republican wins: 24
```

```
## Number of democrat wins: 27
```

Question 5

Create a summary using descriptive statistics off voter turnout in all states. Return the values of percentage for each party.

```
## Democratic turnout summary:
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.3934  0.4789  0.4675  0.5698  0.8976
```

```
## Republican turnout summary:
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.3791  0.4678  0.4519  0.5693  0.7106
```

```
## Other turnout summary:
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.005455 0.018430 0.028750 0.080640 0.057580 1.000000
```

Question 6

Using a single proportion: Hypothesis test and confidence interval

Lynda shows an example for comparing independent samples, whereas we only have one sample, since democrats and republican votes are not independent. We could've used the function below, but it would've been incorrect.

```
prop.test(x, n, p = 0.5, conf.level = 0.95, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x out of n, null probability 0.5
## X-squared = 0.078431, df = 1, p-value = 0.7794
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3315787 0.6140026
## sample estimates:
##         p
## 0.4705882
```

Hence, I wrote some code to do a Binomal Proportion test that makes sense here.

Since, Ho: p0=0.05 vs Ha: p0 is not 0.05, use 2 sided test; Z~Norm(0,1)

```
## Accept Ho, Not enough evidence to reject Ho at 5%
```

- Errors for extra credit

Type 1 error, that we reject H0, given that H0 is true- will be used, if we reject NULL

Type 2 error, rejecting Ha, given that Ha is true; same as the probability that observed value is not the rejection region when Ha is true

```
## The value of the alpha is 5%
```

```
## The value of the beta is 29.9%
```

Question 7

Using a single mean: Hypothesis test and confidence interval

Claim: Democrates will lose the election with mu < 0.5

Null Hypothesis is that Republicans will score 50% of voters, therefore win

Alternative hypothesis: Republicans will get less than 50 % of voters and lose

Test to perform at alpha 5% significance:

```
Ho: mu = 0.5001
```

```
Ha: mu < 0.5001
```

```
## [1] -1.76131
```

```
## Statistical evidence suggests that Ho cannot be rejected at 5%, because the sample does not signific
```

Question 8

Using a single categorical variable: One sample chi-square test

This is a goodness of fit test, we are testing that votes are equally distributed between parties, this is an equivalent of saying mu = 0.5 Ho: votes are equally distributed between both parties Ha: votes are not equally distributed between both parties

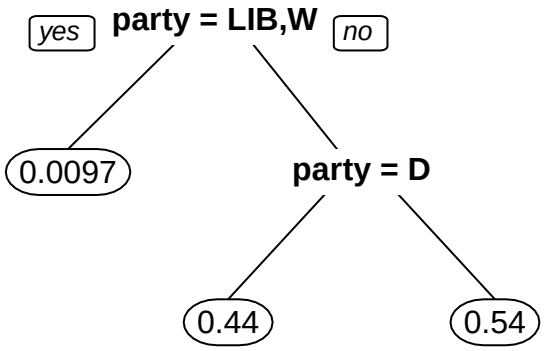
```
##
##  Chi-squared test for given probabilities
##
## data:  voters
## X-squared = 1.5474, df = 1, p-value = 0.2135
```

```
## Statistical evidence suggests that Ho cannot be rejected at 5%, because the sample does not signific
```

Question 9

Create a decision tree in your state based on voter turnout.

Regression Tree based on Party voting turnout in Indiana



Question 10

Create a scatter plot with liner regression line to show each counties voting pattern for each party.