Gerber and Green Experiment

## Gerber and Green 2003

Exploratory data analysis

In this experiment there is total number of 18 933 individuals from 6 different cities in the United States.

The dependent and independent variables

The dependent variable is whether the person voted in the 6 November election in 2001. There are six independent variables; race, sex, age, party affiliation, turnout in the 2000 election and turnout in the 1999 election. The treatment indicator variable

Weird values

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   31.00   42.00   45.25   56.00 2001.00    2487
```
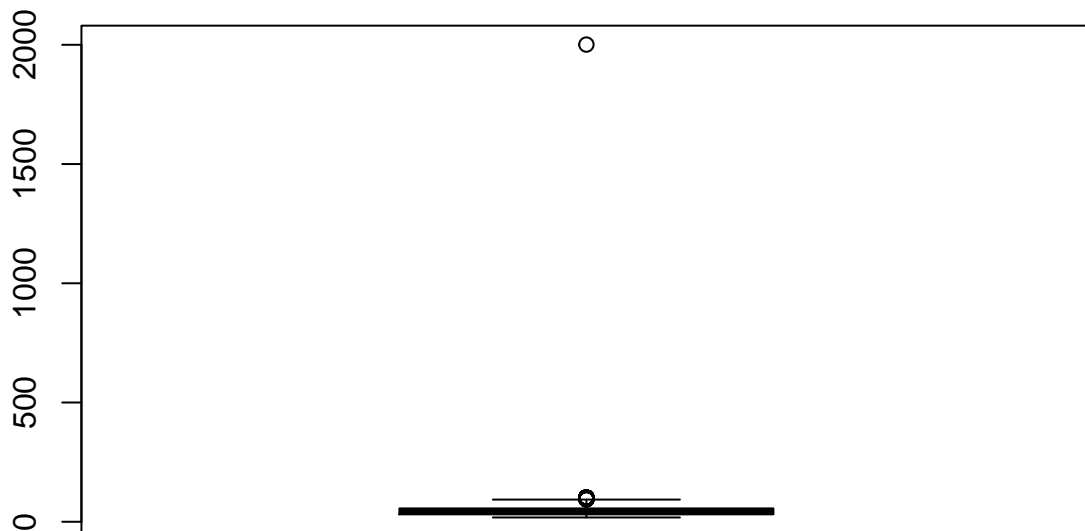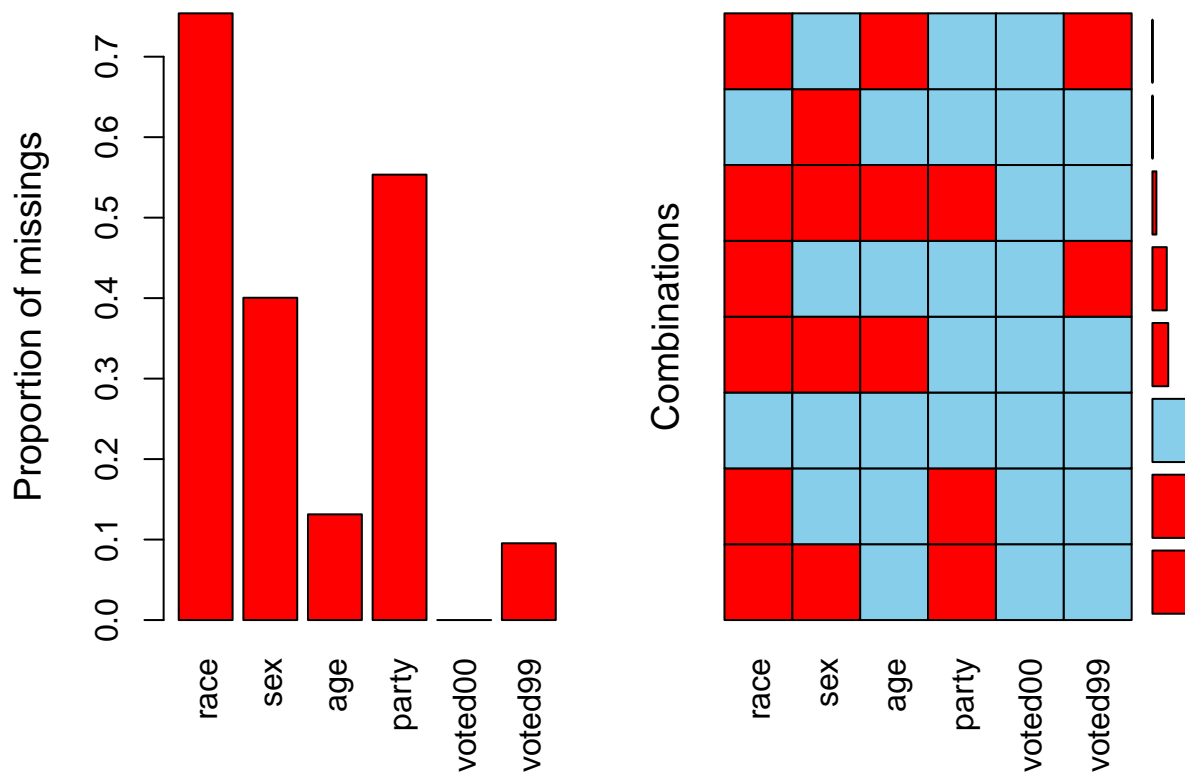


```
##             city           precinct    zip race party sex  age turf
## 13846 MINNEAPOLIS MINNEAPOLIS W- 6 P- 8 55407                2001  103
##       voted01 voted00 voted99 family famsize represen reached other goaway
## 13846       1       0       0   4506       1        1       0     0      0
```

```
##       nothome bad cant nothing contact treatmen primary
## 13846       0   0    0       0       0        0       0
```
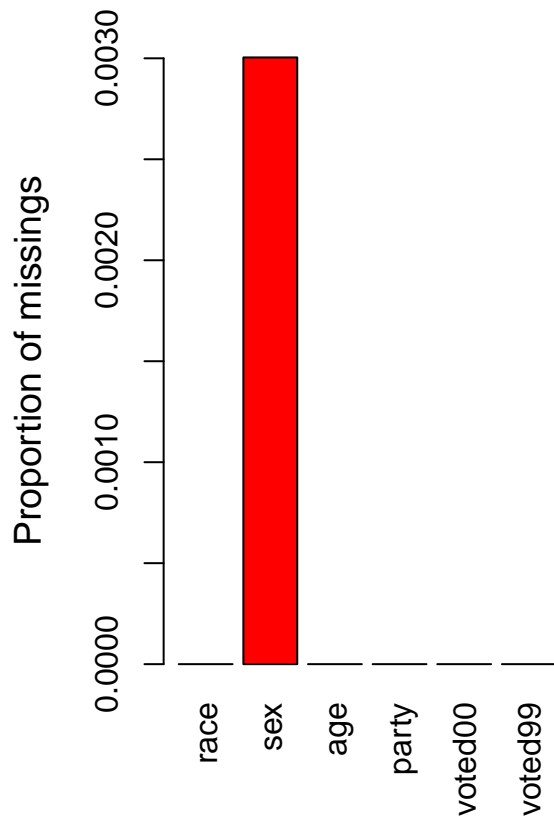
Missing data

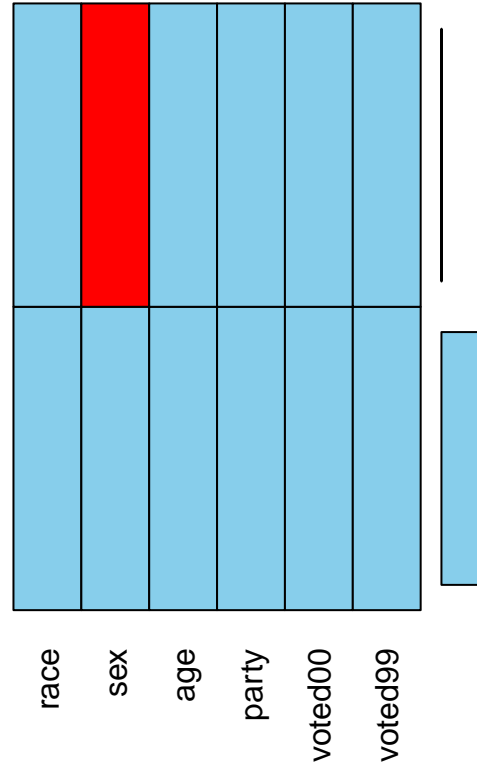There is missing data in the experimental files - we need to recode them to the NA format to further inspect them.
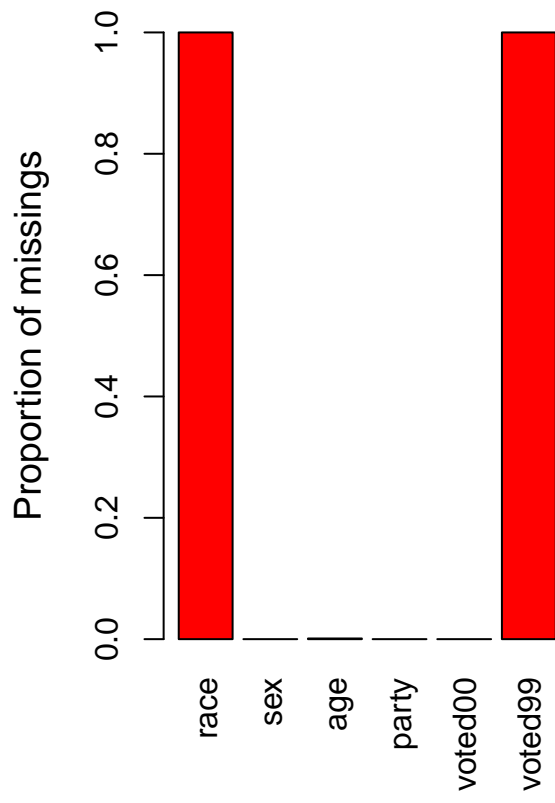


```
## [1] "Raleigh"
```

Proportion of missings
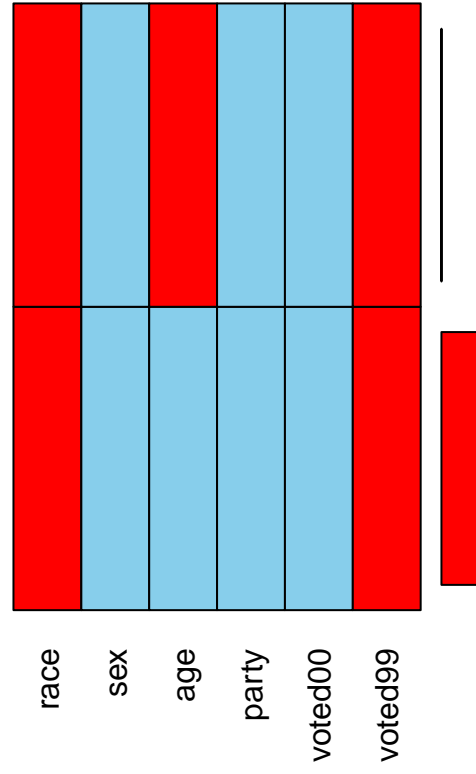
Combinations

race sex age party voted00 voted99

[1] "Bridgeport"

```
## [1] "DETROIT"
```

Proportion of missings — barplot showing race, sex, age, party, voted00, voted99

Combinations — showing race, sex, age, party, voted00, voted99

```
## [1] "COLUMBUS"
```
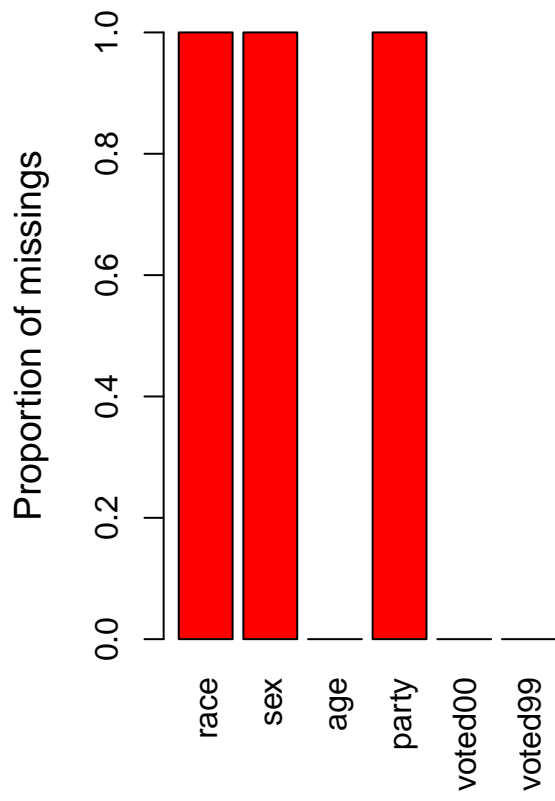
```
## [1] "MINNEAPOLIS"
```

```
## [1] "ST PAUL"
```
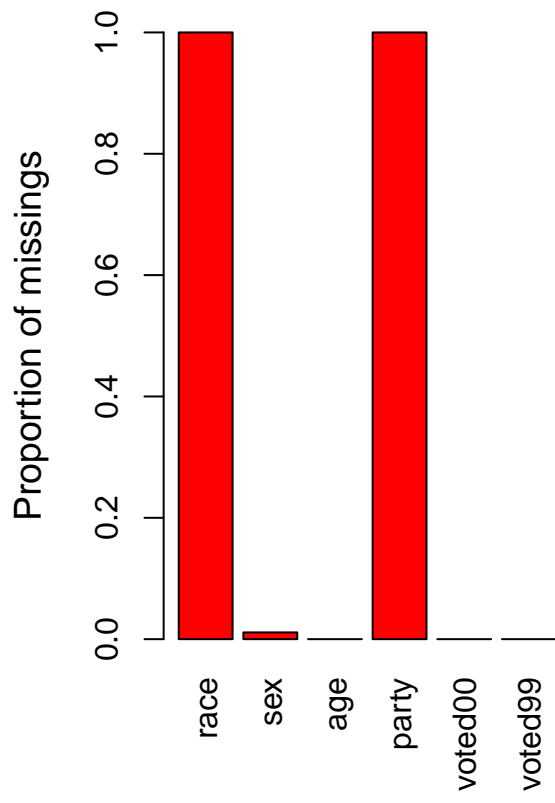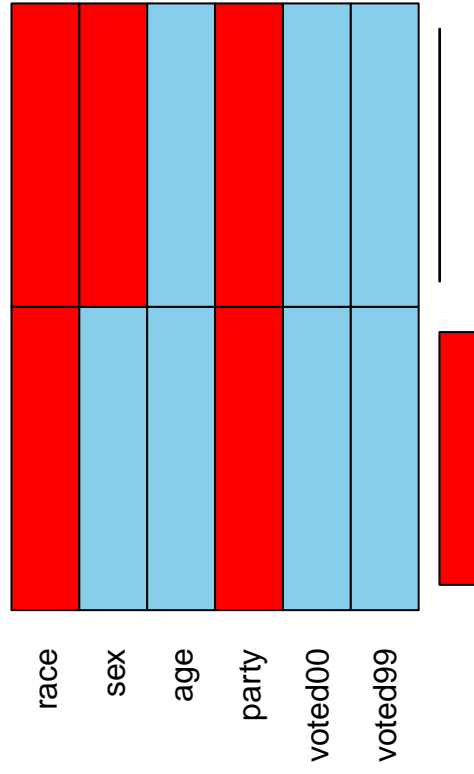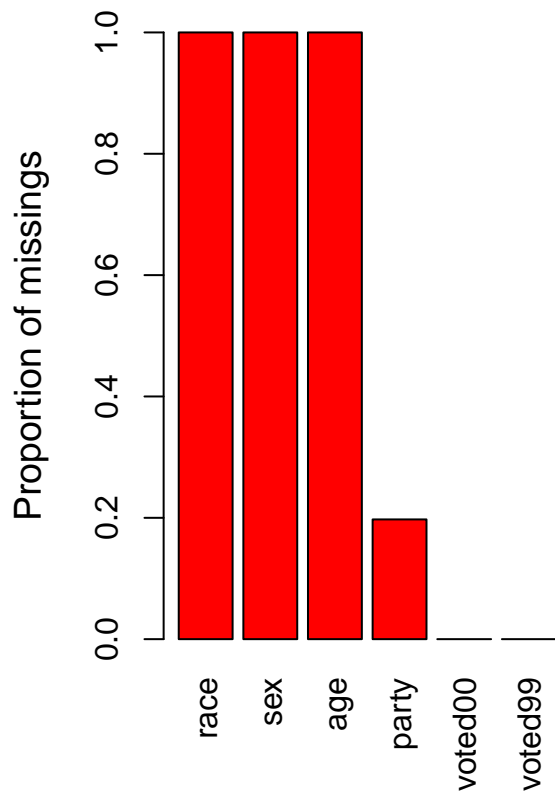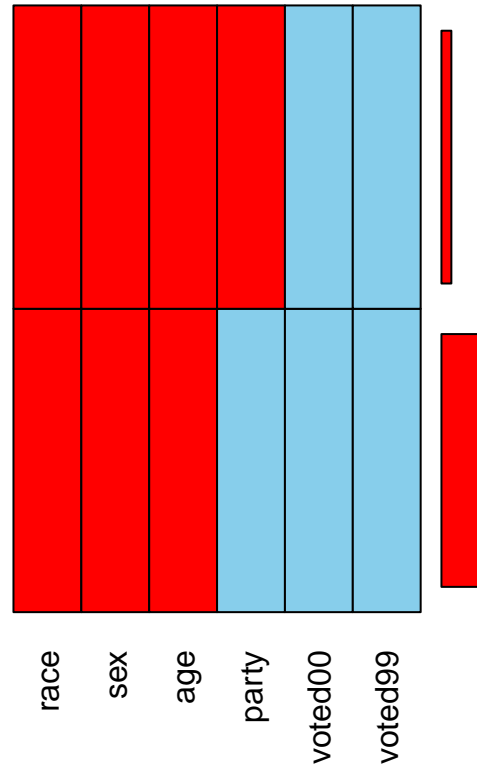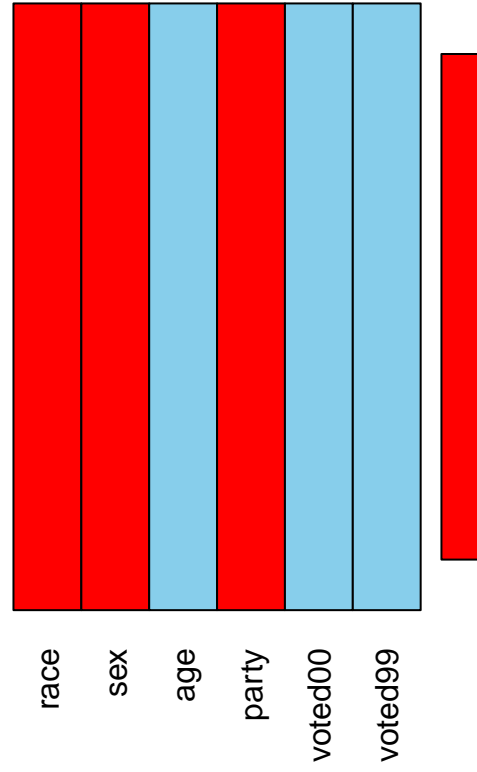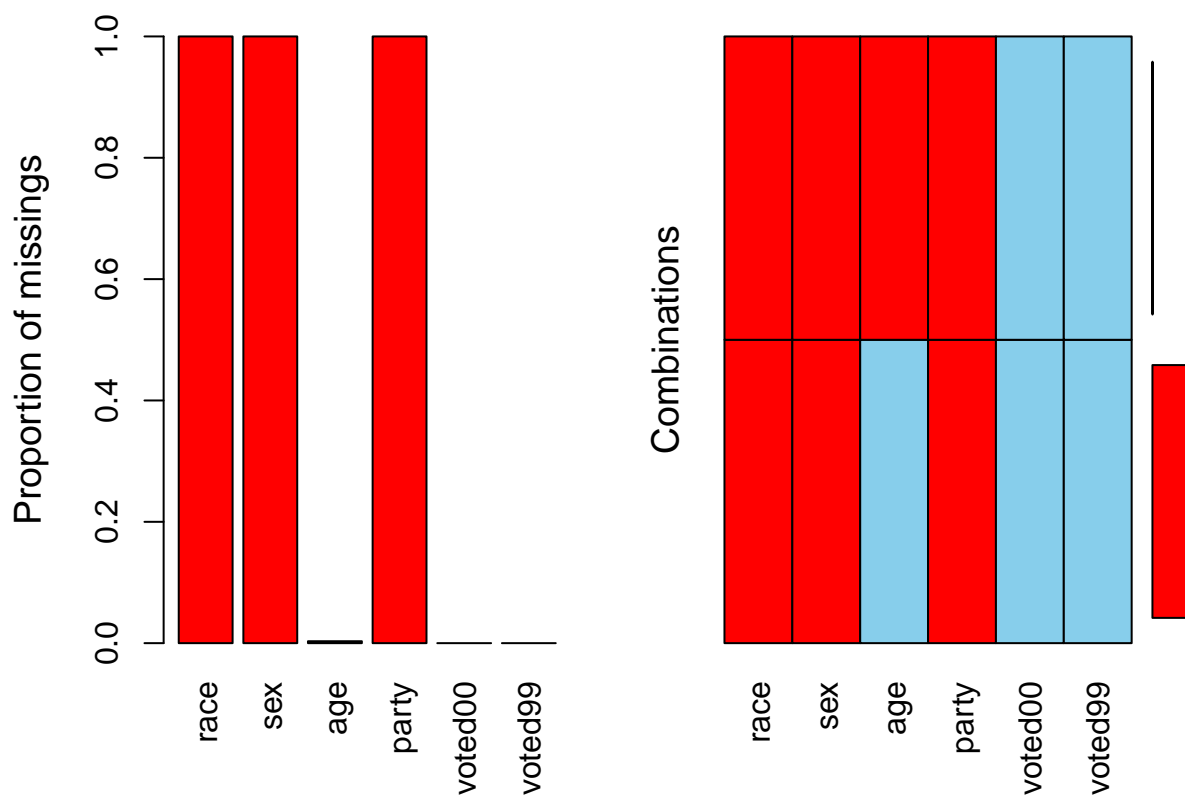
From these plots we conclude that Columbus is not useful at all.

St Paul and Minneapolis only useful together. I will try this.

Detroit, Raleigh, Bridgeport.

The rationale for putting Minneapolis and St Paul together

- Same state
- Same sampling composition
- x Variables to be used: age, voted00, voted99

Dealing with the one weird individual who is 2001 years old

Detroit, Raleigh, Bridgeport

- Bridgeport seems a bit different; very low turnout. Different composition and so on.
- Detroit, Raleigh is a stretch but let's try.

Understanding the differences in background covariates in Minnesota
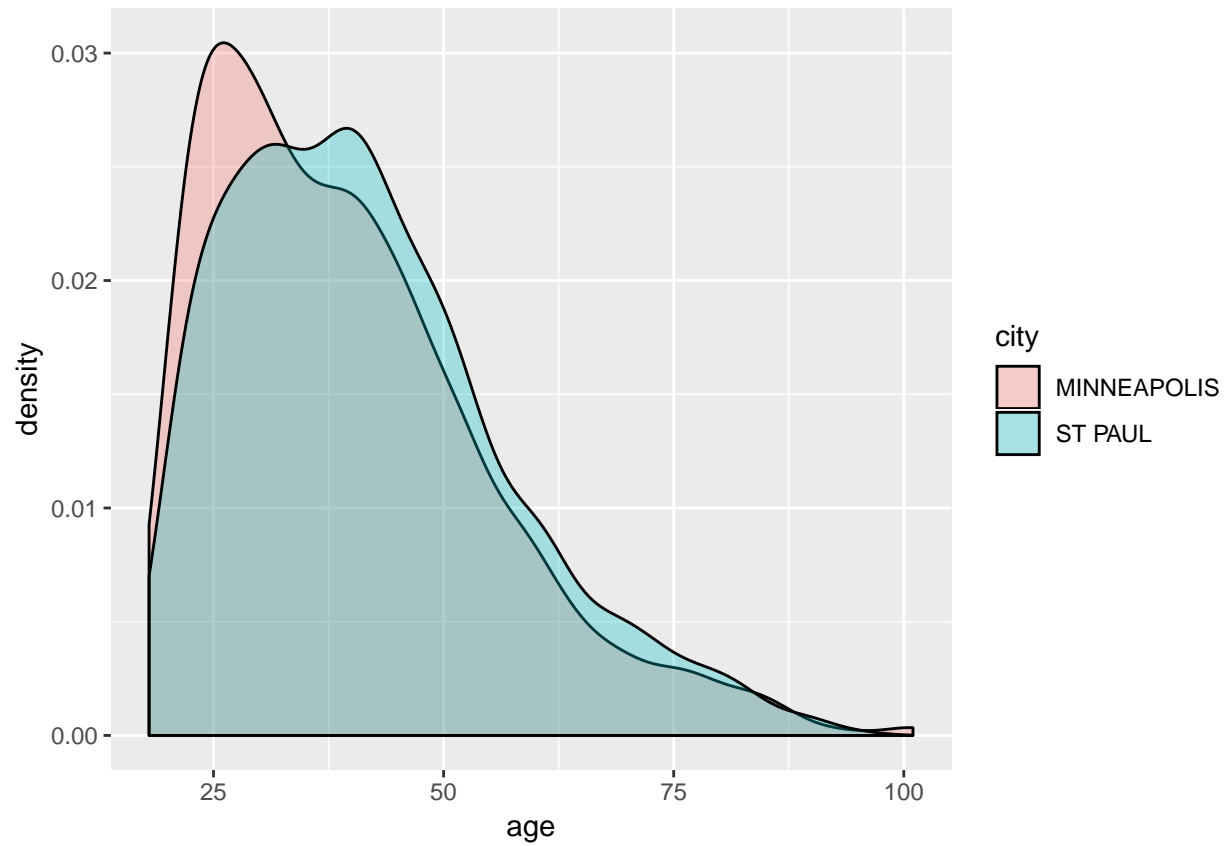

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  1.0000  1.0000  0.8174  1.0000  1.0000

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.0000  1.0000  0.5389  1.0000  1.0000
```
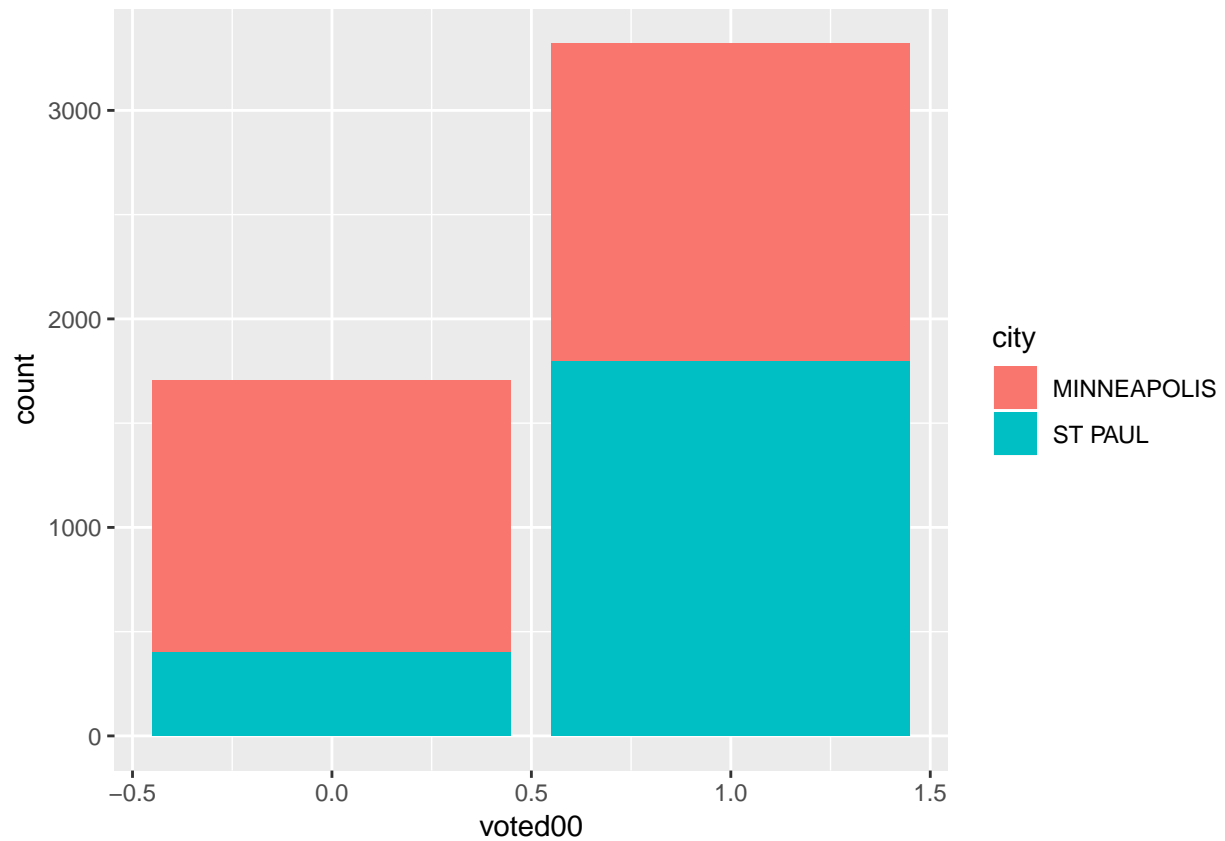
Rescale the variables

Making prediction in Minnesota

First, we prepare the data. Let $D = 0$ be Minneapolis and $D = 1$ St Paul.

Now we run causal Match just once to see

Histogram of d0$age
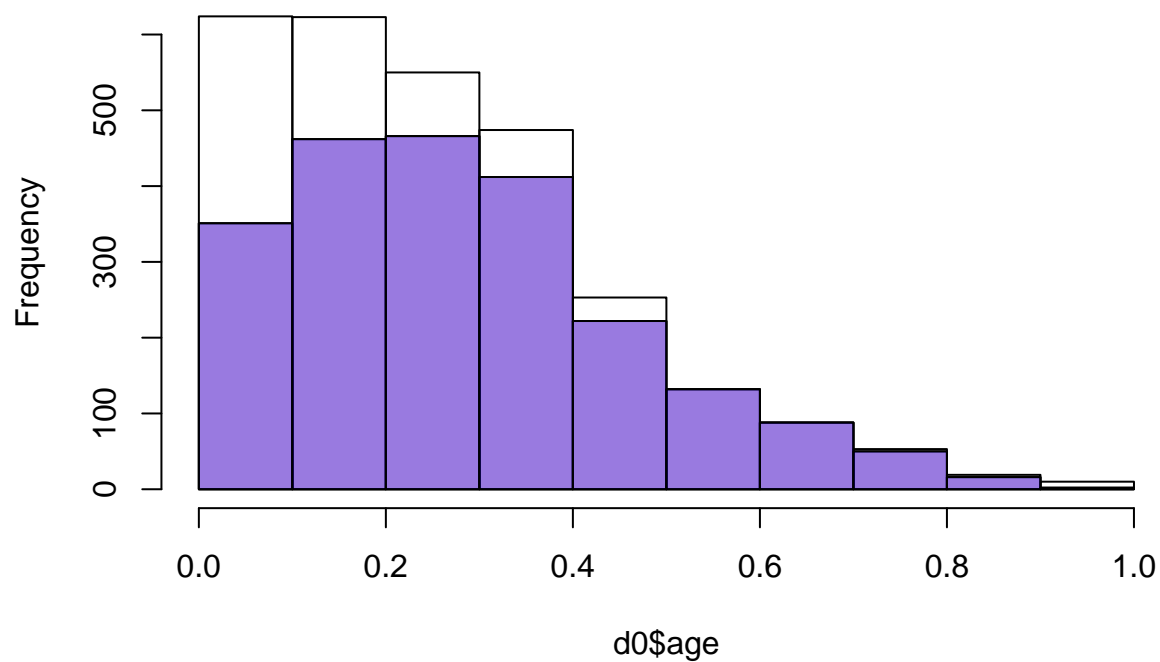
Now we obtain the final predictions:

```
## [1] "Building trees ..."
## [1] "Tree 1"
## [1] 2
## [1] "CT"
## [1] "Tree 2"
## [1] 2
## [1] "CT"
## [1] "Tree 3"
## [1] 2
## [1] "CT"
## [1] "Tree 4"
## [1] 2
## [1] "CT"
## [1] "Tree 5"
## [1] 2
## [1] "CT"
## [1] "Tree 6"
## [1] 2
## [1] "CT"
## [1] "Tree 7"
## [1] 2
## [1] "CT"
## [1] "Tree 8"
## [1] 2
## [1] "CT"
```

```
## [1] "Tree 9"
## [1] 2
## [1] "CT"
## [1] "Tree 10"
## [1] 2
## [1] "CT"
## [1] "Tree 11"
## [1] 2
## [1] "CT"
## [1] "Tree 12"
## [1] 2
## [1] "CT"
## [1] "Tree 13"
## [1] 2
## [1] "CT"
## [1] "Tree 14"
## [1] 2
## [1] "CT"
## [1] "Tree 15"
## [1] 2
## [1] "CT"
## [1] "Tree 16"
## [1] 2
## [1] "CT"
## [1] "Tree 17"
## [1] 2
## [1] "CT"
## [1] "Tree 18"
## [1] 2
## [1] "CT"
## [1] "Tree 19"
## [1] 2
## [1] "CT"
## [1] "Tree 20"
## [1] 2
## [1] "CT"
## [1] "Tree 21"
## [1] 2
## [1] "CT"
## [1] "Tree 22"
## [1] 2
## [1] "CT"
## [1] "Tree 23"
## [1] 2
## [1] "CT"
## [1] "Tree 24"
## [1] 2
## [1] "CT"
## [1] "Tree 25"
## [1] 2
## [1] "CT"
## [1] "Tree 26"
## [1] 2
## [1] "CT"
```

```
## [1] "Tree 27"
## [1] 2
## [1] "CT"
## [1] "Tree 28"
## [1] 2
## [1] "CT"
## [1] "Tree 29"
## [1] 2
## [1] "CT"
## [1] "Tree 30"
## [1] 2
## [1] "CT"
## [1] "Tree 31"
## [1] 2
## [1] "CT"
## [1] "Tree 32"
## [1] 2
## [1] "CT"
## [1] "Tree 33"
## [1] 2
## [1] "CT"
## [1] "Tree 34"
## [1] 2
## [1] "CT"
## [1] "Tree 35"
## [1] 2
## [1] "CT"
## [1] "Tree 36"
## [1] 2
## [1] "CT"
## [1] "Tree 37"
## [1] 2
## [1] "CT"
## [1] "Tree 38"
## [1] 2
## [1] "CT"
## [1] "Tree 39"
## [1] 2
## [1] "CT"
## [1] "Tree 40"
## [1] 2
## [1] "CT"
## [1] "Tree 41"
## [1] 2
## [1] "CT"
## [1] "Tree 42"
## [1] 2
## [1] "CT"
## [1] "Tree 43"
## [1] 2
## [1] "CT"
## [1] "Tree 44"
## [1] 2
## [1] "CT"
```

```
## [1] "Tree 45"
## [1] 2
## [1] "CT"
## [1] "Tree 46"
## [1] 2
## [1] "CT"
## [1] "Tree 47"
## [1] 2
## [1] "CT"
## [1] "Tree 48"
## [1] 2
## [1] "CT"
## [1] "Tree 49"
## [1] 2
## [1] "CT"
## [1] "Tree 50"
## [1] 2
## [1] "CT"
## [1] "Tree 51"
## [1] 2
## [1] "CT"
## [1] "Tree 52"
## [1] 2
## [1] "CT"
## [1] "Tree 53"
## [1] 2
## [1] "CT"
## [1] "Tree 54"
## [1] 2
## [1] "CT"
## [1] "Tree 55"
## [1] 2
## [1] "CT"
## [1] "Tree 56"
## [1] 2
## [1] "CT"
## [1] "Tree 57"
## [1] 2
## [1] "CT"
## [1] "Tree 58"
## [1] 2
## [1] "CT"
## [1] "Tree 59"
## [1] 2
## [1] "CT"
## [1] "Tree 60"
## [1] 2
## [1] "CT"
## [1] "Tree 61"
## [1] 2
## [1] "CT"
## [1] "Tree 62"
## [1] 2
## [1] "CT"
```

```
## [1] "Tree 63"
## [1] 2
## [1] "CT"
## [1] "Tree 64"
## [1] 2
## [1] "CT"
## [1] "Tree 65"
## [1] 2
## [1] "CT"
## [1] "Tree 66"
## [1] 2
## [1] "CT"
## [1] "Tree 67"
## [1] 2
## [1] "CT"
## [1] "Tree 68"
## [1] 2
## [1] "CT"
## [1] "Tree 69"
## [1] 2
## [1] "CT"
## [1] "Tree 70"
## [1] 2
## [1] "CT"
## [1] "Tree 71"
## [1] 2
## [1] "CT"
## [1] "Tree 72"
## [1] 2
## [1] "CT"
## [1] "Tree 73"
## [1] 2
## [1] "CT"
## [1] "Tree 74"
## [1] 2
## [1] "CT"
## [1] "Tree 75"
## [1] 2
## [1] "CT"
## [1] "Tree 76"
## [1] 2
## [1] "CT"
## [1] "Tree 77"
## [1] 2
## [1] "CT"
## [1] "Tree 78"
## [1] 2
## [1] "CT"
## [1] "Tree 79"
## [1] 2
## [1] "CT"
## [1] "Tree 80"
## [1] 2
## [1] "CT"
```

```
## [1] "Tree 81"
## [1] 2
## [1] "CT"
## [1] "Tree 82"
## [1] 2
## [1] "CT"
## [1] "Tree 83"
## [1] 2
## [1] "CT"
## [1] "Tree 84"
## [1] 2
## [1] "CT"
## [1] "Tree 85"
## [1] 2
## [1] "CT"
## [1] "Tree 86"
## [1] 2
## [1] "CT"
## [1] "Tree 87"
## [1] 2
## [1] "CT"
## [1] "Tree 88"
## [1] 2
## [1] "CT"
## [1] "Tree 89"
## [1] 2
## [1] "CT"
## [1] "Tree 90"
## [1] 2
## [1] "CT"
## [1] "Tree 91"
## [1] 2
## [1] "CT"
## [1] "Tree 92"
## [1] 2
## [1] "CT"
## [1] "Tree 93"
## [1] 2
## [1] "CT"
## [1] "Tree 94"
## [1] 2
## [1] "CT"
## [1] "Tree 95"
## [1] 2
## [1] "CT"
## [1] "Tree 96"
## [1] 2
## [1] "CT"
## [1] "Tree 97"
## [1] 2
## [1] "CT"
## [1] "Tree 98"
## [1] 2
## [1] "CT"
```

```
## [1] "Tree 99"
## [1] 2
## [1] "CT"
## [1] "Tree 100"
## [1] 2
## [1] "CT"

## [1] 2201   100
```

The tau_1^{PRED} from causal match is 0.0305319 and from causal forest 0.0263373. Their respective errors are 2.008498 and 3.37339. The NPE is 6.1686766.