Gerber and Green Experiment

Gerber and Green (2003)

In this appendix, I first describe the experimental data from the paper. The original sample size for the experiment consists of 18933 individuals from 6 different cities in the United States.

Data preparation

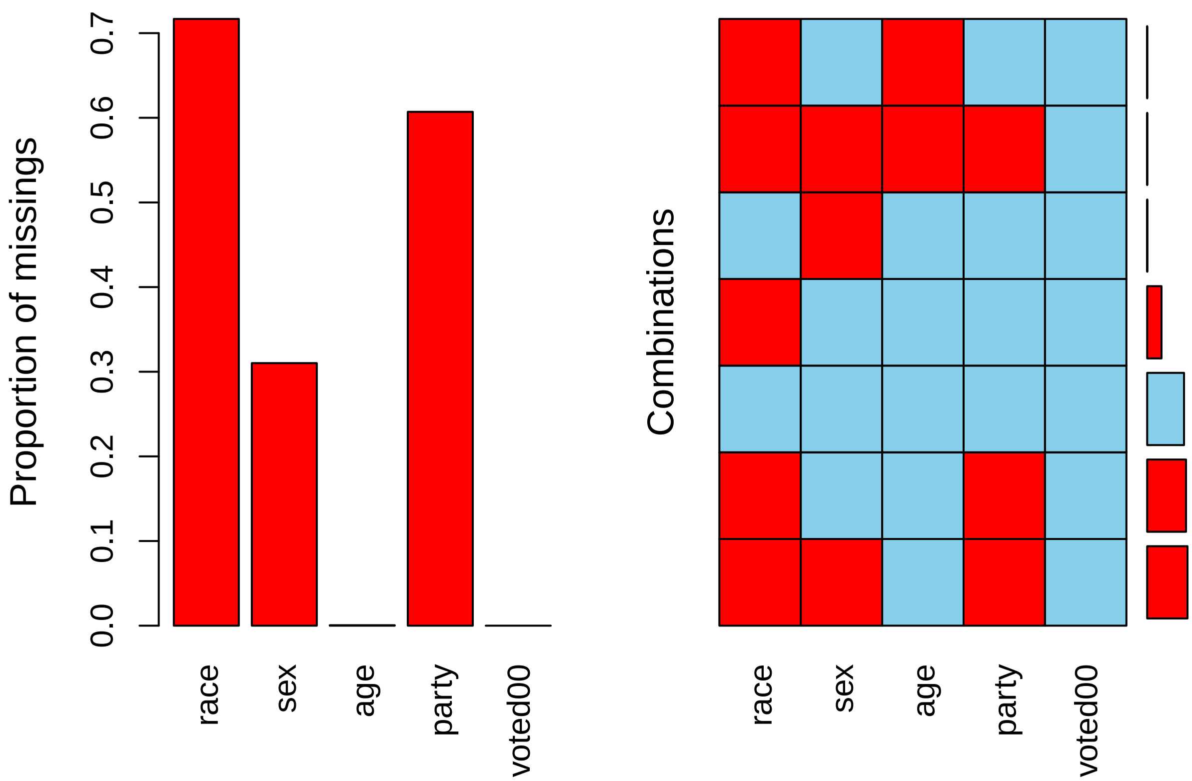
Variables

The dependent variable `voted00` tells us whether the person voted in the 6 November election in 2001. The treatment indicator `treatmen` tells us whether the individual was encouraged to be visited by the canvassers. Additionally, the experiment records six background variables about every individual; race, sex, age, party affiliation, turnout in the 2000 election and turnout in the 1999 election.

In our prediction problem using the experimental data, we do not use the 1999 election variable. According to our knowledge, no elections ocurred in all of the concerned cities in 1999.

Missing data

There is missing data in the experimental data. In R, the missing data was first recoded to the R's standard NA format and the patterns are visualised in the Figure ().



Rather than looking at aggregate patterns of missingness, what is more of an interest for our specific case are the patterns of missing values in each of the locations. We proceed with complete case analysis for cases where the missingness is ~1% on a covariate for a given location. The table below shows the proportion of missing values for a given covariate for every location.

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
```

```
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

	race	sex	age	party	voted00
Raleigh	0	0.30	0.00	0	0
Bridgeport	100	0.00	0.11	0	0
Detroit	100	1.11	0.00	100	0
Minneapolis	100	100.00	0.00	100	0
St Paul	100	100.00	0.32	100	0

In the complete case analysis we dicarded 78 rows from the dataset.

Other data transformations

As the next step in the data transformation process, we delete one invalid data observation with the value of 2001 for age.

Additionally, we reclassify subjects who answered with Unknown to their sex to 'NA' and excluded. This simplifies the matching procedure since with only two sexes, the variable can be operationalised as a dummy variable.

Data exploration

The varying individual characteristics between locations

This

Age and Voting in the 2000 election

	Raleigh	Bridgeport	Detroit	Minneapolis	St Paul
age	45.61	44.78	49.35	39.66	41.72
voted00	0.72	0.44	0.61	0.54	0.82

Sex

	F	M	U
Raleigh	0.51	0.49	NA
Bridgeport	0.61	0.38	0.01
Detroit	0.61	0.39	NA

Party affiliation

	D	I	R
Raleigh	0.48	0.20	0.31
Bridgeport	0.54	0.38	0.08

Further causalMatch specifications

In the operationalisation of causalMatch to the experimental data in the dissertation we have rescaled the age variable to be in the range 0 to 1. Here, I provide two other additional options of working with the variable. First, I scale the age variable by dividing it by its root-mean-square defined as

$$\sqrt{\frac{1}{n}\sum_{t=1}^{n} age_t^2}$$

	τ_{ITT}^{PRED}	$\hat{\tau_{ITT}}$	SE	τ_{ITT}^{NAIVE}	NPE
Bridgeport	2.29	3.84	2.42	2.01	3.37
Raleigh	2.76	-0.90	13.38	3.19	16.77
Detroit	1.56	2.47	0.83	2.35	0.02
Minneapolis	3.46	1.99	2.18	2.47	0.23
St Paul	3.65	4.47	0.67	1.85	6.87

Finally, the table below presents the unscaled results. In this case, no transformation was applied to the age variable.

	τ_{ITT}^{PRED}	$\hat{\tau_{ITT}}$	SE	τ_{ITT}^{NAIVE}	NPE
Bridgeport	2.97	3.84	0.77	2.01	3.37
Raleigh	2.91	-0.90	14.57	3.19	16.77
Detroit	1.68	2.47	0.63	2.35	0.02
Minneapolis	3.45	1.99	2.14	2.47	0.23
St Paul	3.63	4.47	0.70	1.85	6.87