STAT6730 Project

Exploring park access data from the Trust for Public Land

Polina Berezina, Shan Tang

12/10/2021

Introduction

Data cleaning

Delimiter: ","

Converting 3 variables from string (%) to double (0.-1), year as a factor.

```
## Rows: 713 Columns: 28
```

chr (5): city, park_pct_city_data, pct_near_park_data,
dbl (23): year, rank, med_park_size_data, med_park_size

-- Column specification -----

```
##
## i Use 'spec()' to retrieve the full column specification
## i Specify the column types or set 'show col types = FAL'
```

i Specify the column types or set 'show_col_types = FALS
spec_tbl_df [713 x 28] (S3: spec_tbl_df/tbl_df/tbl/data

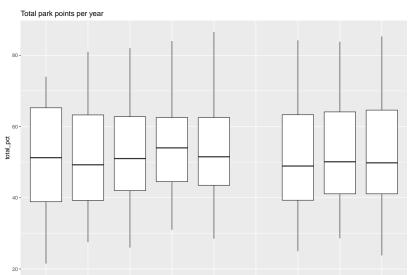
```
## $ year : num [1:713] 2020 2020 2020 ## $ rank : num [1:713] 1 2 3 4 5 6 7
```

\$ city : chr [1:713] "Minneapolis" ## \$ med_park_size_data : num [1:713] 5.7 1.4 3.2 2

Φ mad manile aims maints [1.712] OC E 1/ 10 00

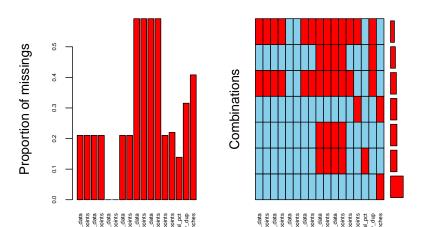
Exploratory data analysis

Not interested in the overall year trend in total park points. But observed a recently increasing trend in the Spending per resident and Percent of residents within a 10 minute walk to park



Handling missing data

- More than 30% of data is missing in 6 out of 28 variables, including restroom, splashground and park benches counts. We omit these variables from analysis.
- Assuming missing at random for others. Impute using mice package with PMM Hot Deck method











Conclusions

- 123