1. The names of the students who will be contributing to the group project.

Haofei Xu, Linda (Xinyi) Yuan, Rebecca (YaoJia) Chen

2. A tentative title for the project.

House data research at Saratoga, New York

3. Description of the data file (what they contain including number of variables and number of records).

You do not necessarily have to list all the variables, but at least mention those of greatest importance.

Number of variables: 7

Number of records: 1057

- Price: The price of the house (US dollars)
- Living Area: Total living area (square feet)
- Lot Size: Size of lot (acres)
- Age: Age of house (years)
- Bathrooms: Number of bathrooms
- Bedrooms: Number of bedrooms
- Fireplace (categorical): Number of fireplaces
- Fireplaces (numerical): Number of fireplaces

4. Background information on the data sets, including specific citation of their source (so that I can also access it).

The data set we obtained includes a random sample of 1057 houses taken from full Saratoga Housing Data (De Veaux). The data examines the properties of houses located in Saratoga County, New York, USA.

Citation:

https://dasl.datadescription.com/datafile/housing-prices/?_sfm_methods=Multiple+Regression&_sfm_cases=4+59943&sf_paged=2

5. A brief statement of the business, science, research, or personal interest you have in the data set which you hope to explore.

This dataset explores numerous variables in predicting the housing price in New York. Since New York is a popular state for real estate investment, it seems interesting for us to explore the cost of settlement in this area. Also, just in general, we wonder what are the most important factors in predicting the price of the house in New York.

6. Evidence that the data can be loaded into R. Load the data, and print the first few values of the response variable as evidence.

```{r}
library(readr)
housing = read.csv("housing-prices.csv")
head(housing)
```

	Price <int>	Living.Area <int>	Bathrooms <dbl>	Bedrooms <int>	Fireplaces <int>	Lot.Size <dbl>	Age <int>	Fireplace <lgl>
1	142212	1982	1.0	3	0	2.00	133	FALSE
2	134865	1676	1.5	3	1	0.38	14	TRUE
3	118007	1694	2.0	3	1	0.96	15	TRUE
4	138297	1800	1.0	2	2	0.48	49	TRUE
5	129470	2088	1.0	3	1	1.84	29	TRUE
6	206512	1456	2.0	3	0	0.98	10	FALSE

6 rows