

# Biostatistics 625 Final Project Report

*Group 2 members: Mukai Wang, Kangping Yang, Litian Zhou*

*December 20, 2019*

## Abstract

Housing costs in China have witnessed turbulent fluctuation in the past few years, especially in the country's capital Beijing. In this project, we design and implement an R shiny app that offers interactive data visualization and statistical insights about the housing price trend in Beijing. The data source is from Kaggle. It contains 300k+ transaction logs between 2011 and 2017.

## Dataset Summary and Storage

The raw data contains 26 variables. After we remove columns that are not useful (e.g. ids) and columns with too many missing values, we retain following variables that we are interested in.

Column	Description
Lng	Longitude of the housing
Lat	Latitude of the housing
tradeTime	Transaction time
totalPrice	Total Price of the Housing
price	Price per Square Meter
square	Area of the housing
livingRoom	Number of Living Rooms
drawingRoom	Number of Drawing Rooms
kitchen	Number of kitchens
bathRoom	Number of Bathrooms
floor	The floor of the housing
buildingType	"Tower" / "Plate" / "Tower/Plate"
constructionTime	Year the housing is built
renovationCondition	"Simplicity" / "HardCover" / "Rough"
buildingStructure	"Mixed" / "Brick/Wood" / "Brick/Concrete" / "Steel" / "Steel/Concrete"
elevator	Whether there is elevator
fiveYearsProperty	True/False
subway	Whether there is subway nearby
district	The district the housing is in