

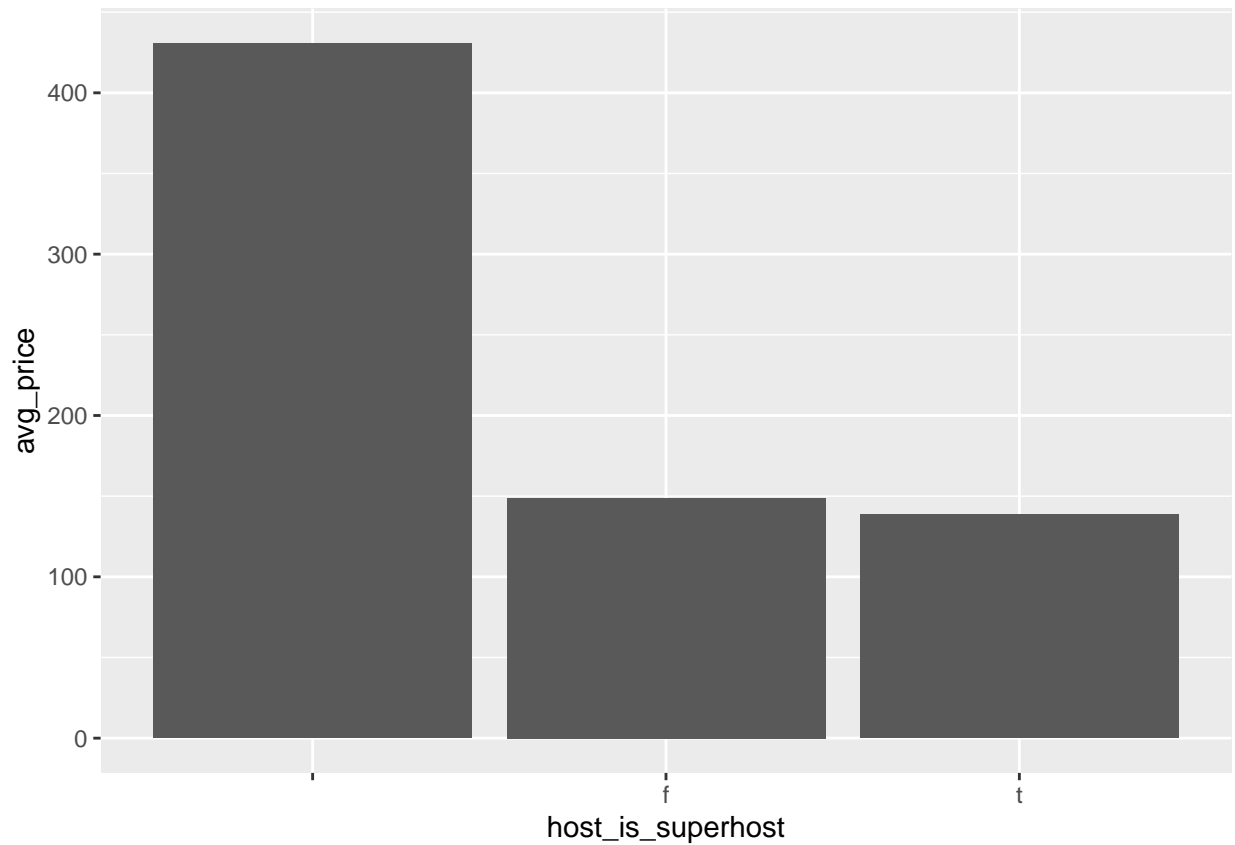
Superhost-Worth-It_pdf

Motivation

Composition

First of all, a summary of the data set that is used will be provided. This data set has been cleaned from errors, NA's and abundant data.

```
## host_is_superhost property_type room_type bedrooms
## Length:7537 Length:7537 Length:7537 Min. : 1.000
## Class :character Class :character Class :character 1st Qu.: 1.000
## Mode :character Mode :character Mode :character Median : 1.000
## Mean : 1.615
## 3rd Qu.: 2.000
## Max. :14.000
##
## beds number_of_reviews review_scores_rating reviews_per_month
## Min. : 1.0 Min. : 0.00 Min. :0.000 Min. : 0.010
## 1st Qu.: 1.0 1st Qu.: 2.00 1st Qu.:4.580 1st Qu.: 0.360
## Median : 2.0 Median : 19.00 Median :4.800 Median : 1.030
## Mean : 2.5 Mean : 56.65 Mean :4.666 Mean : 1.514
## 3rd Qu.: 3.0 3rd Qu.: 76.00 3rd Qu.:4.930 3rd Qu.: 2.160
## Max. :19.0 Max. :765.00 Max. :5.000 Max. :17.720
## NA's :1393 NA's :1393
## price_numeric
## Min. : 9.0
## 1st Qu.: 76.0
## Median : 108.0
## Mean : 153.3
## 3rd Qu.: 156.0
## Max. :9999.0
##
```



The barplot above shows that super hosts have on average lower prices.

Underneath are some summary statistics about the variables used in this study.

```
summary(df_cleaned$price_numeric)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      9.0   76.0   108.0   153.3   156.0  9999.0
```

```
summary(df_cleaned$host_is_superhost_binary)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.0000  0.0000  0.3492  1.0000  1.0000
```

Cleaning