

Thumbtack Data Analysis & Visualization

Jong Ha Lee

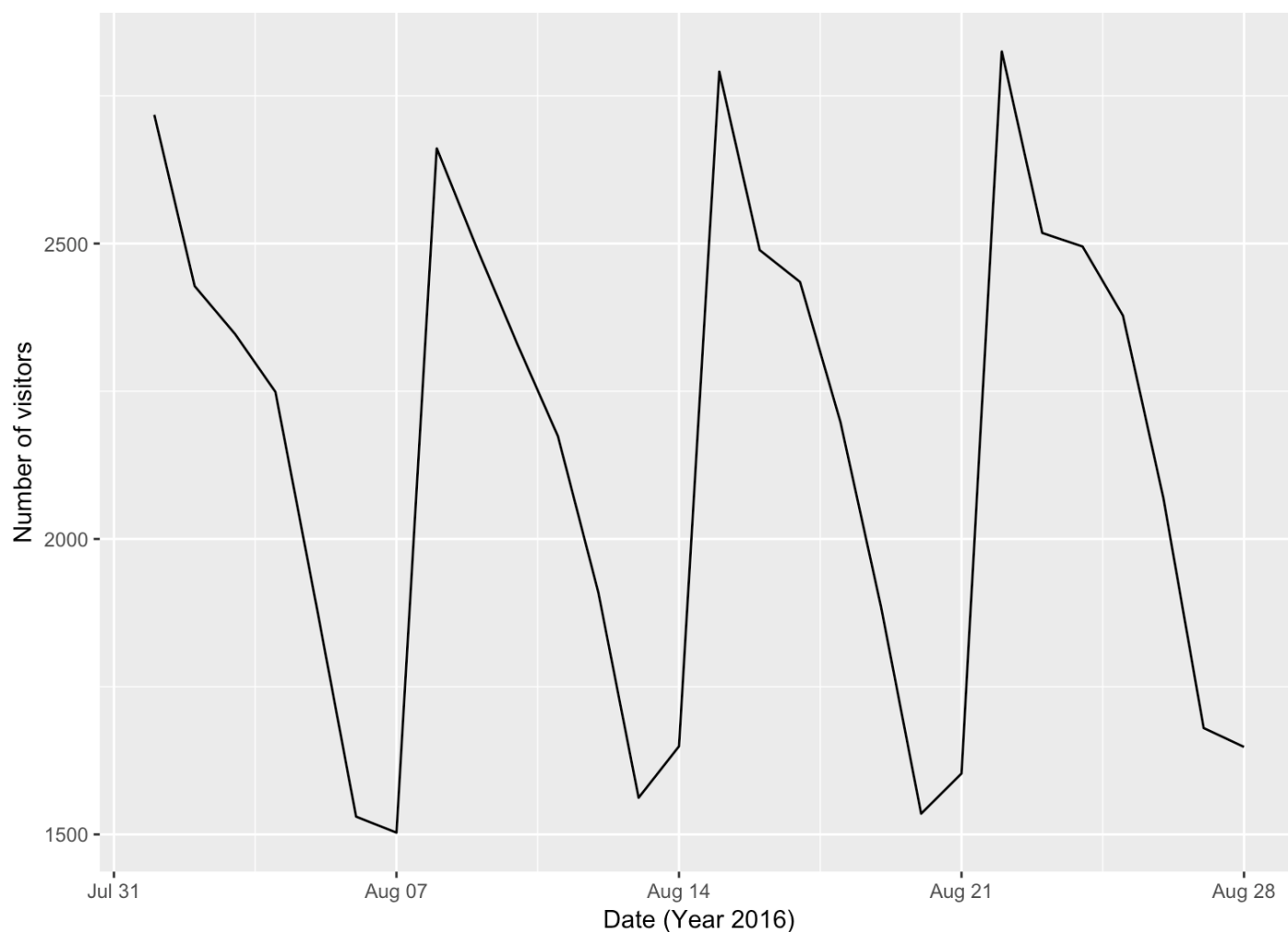
2/1/2017

1. Visitors and Conversion Rate

a) How does the number of visitors change over time?

Line Graph of Number of Visitors over time:

Number of Visitors over Time



Quick analysis shows that there is some sort of periodic activity in the number of visitors, especially on a weekly basis. Let's take August 7th for example:

```
## [1] "Sunday"
```

It looks like August 7th, 2016 is a Sunday - thus we can see that the number of visitors is highest at the beginning of the week (Monday), and diminishes as it approaches the weekend.

b) % of Visitors submitting project request

First, checking whether there are data quality issues (having values besides 0 or 1):

```
##  
##      0      1  
## 25850 34146
```

Looks okay, calculating % of Visitors submitting Project Requests:

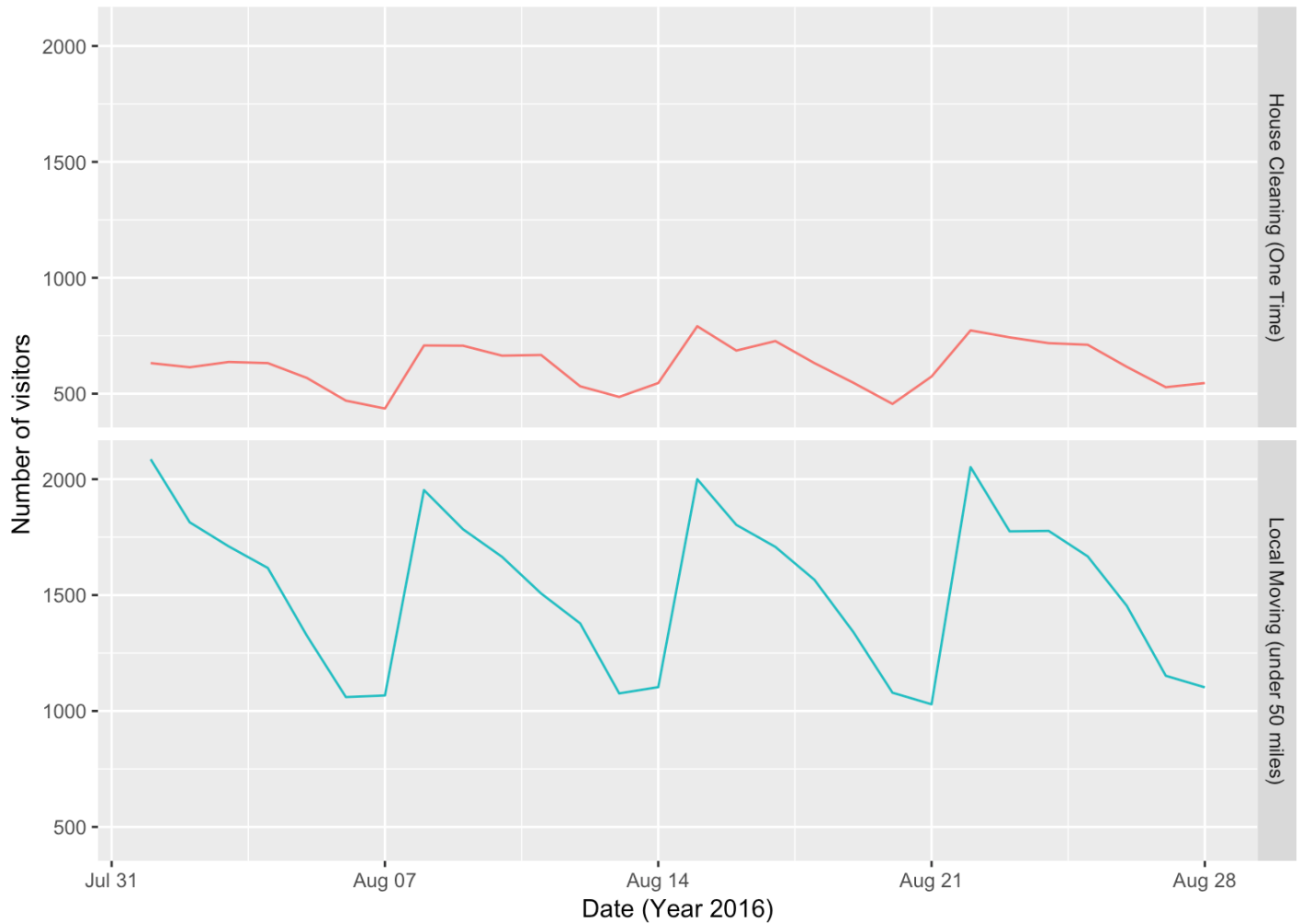
```
## [1] 56.91379
```

It looks like around 56.9 percent of visitors submit a request.

c) Category and Device Influences

Visitors: Categorical Influences

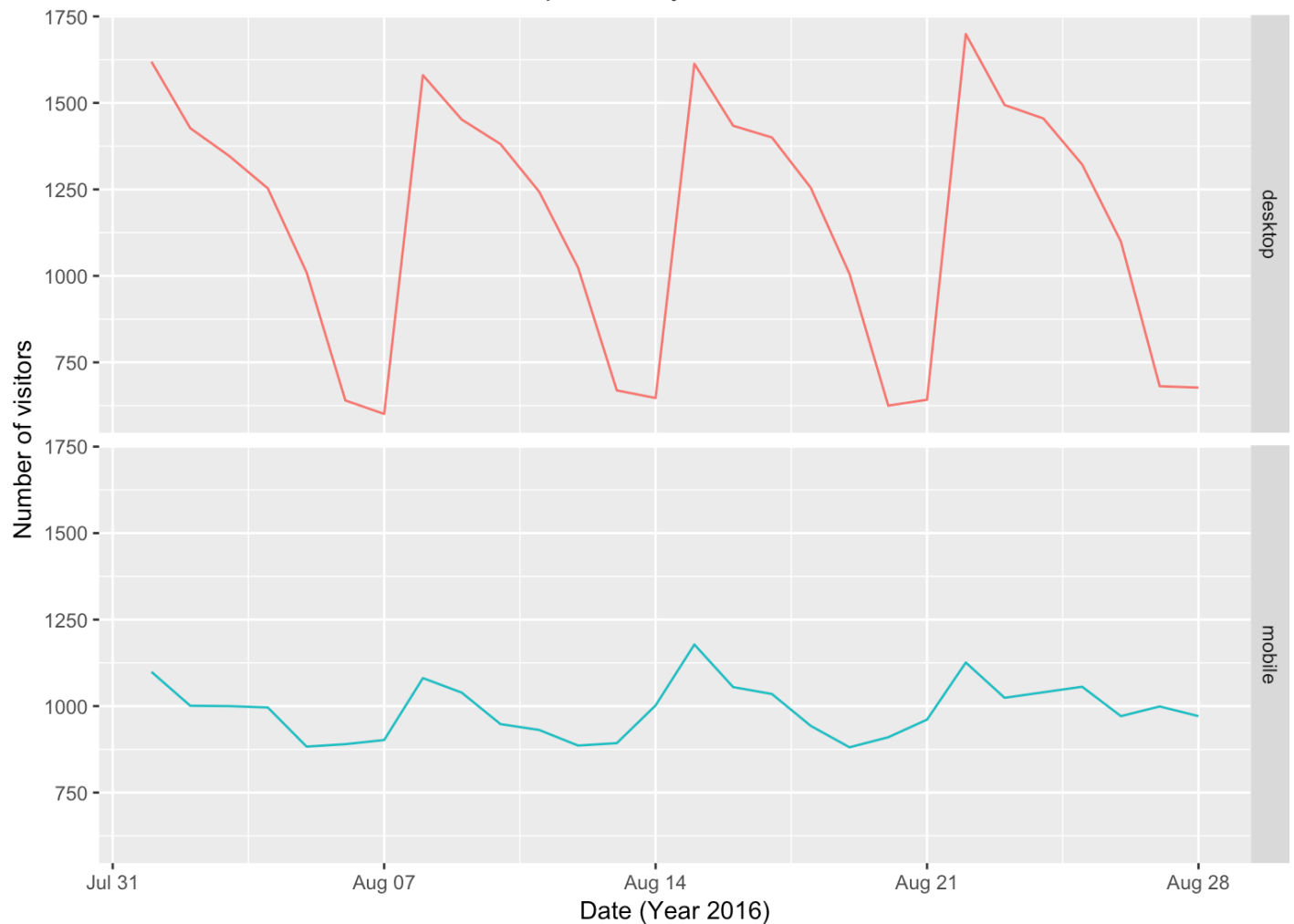
Number of Visitors Over Time, Separated by Category



Looking at the categories, first we notice that there are a lot more requests for Moving help, and thus will contribute to the overall number of visitors trend more in comparison to Cleaning help. Secondly, we can see that the overall trend we saw - decreasing number of visitors as the week go by - is prevalent in both categories. However, we also notice that the trend is more prevalent in the Moving category, with around 50% drops from the start of the week (Monday), to the end of the week (Sunday). In contrast, House Cleaning shows a slightly less degree of a drop, of around 33%.

Visitors: Device Influences

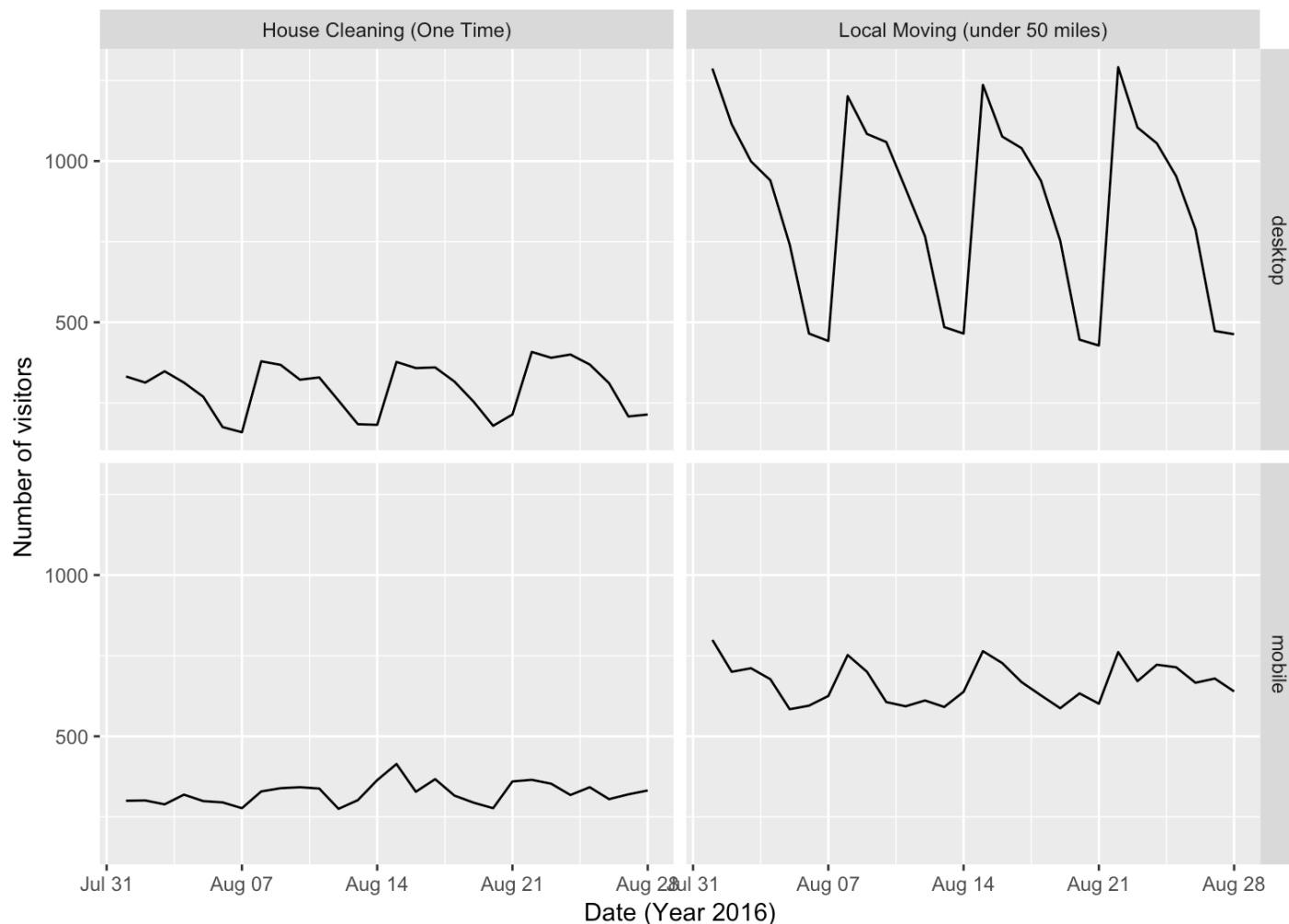
Number of Visitors Over Time, Separated by Device



Looking at the number of visitors over time, separated by device, we can see an analogous trend here. A lot more visitors come through the desktop interface as compared to the mobile application; furthermore, the degree of drop is much larger in the desktop area in comparison to the mobile area.

Visitors: Combined Influences

Number of Visitors Over Time, Separated by Device



Finally, looking at the combination of visitors and devices, we can see that desktop visitors, in need of Moving help, 1) possesses the most number of customers, and 2) clearly shows the diminishing trend we described earlier.

I hypothesize that more people tend towards the Desktop application mainly because for these help activities - especially moving - people wish to be thorough and not miss a detail in hiring a professional. These activities are a hassle for people, and they wish to avoid more hassle in the logistics of getting professional help (e.g. missing move-out date because they missed a small detail in putting in a request). Thus, since Desktop provides a more holistic view and can save details easily, I believe people, especially in need of help in these two categories, tend more towards the Desktop application.

As for the diminishing trend from weekday to weekend, I believe it is due to the fact that people want the professionals to come in during weekends to help them, because it may be a hassle to talk and let in the professionals into the house during working days. Thus, assuming the request -> quote -> agreement process takes around 4-5 days, it may seem reasonable that most people submit requests on Monday, when professionals are back to work from the weekend checking requests, and also gives ample time to walk through the process of getting professionals on board.

Percent Visitors: Category

```
##
##      House Cleaning (One Time) Local Moving (under 50 miles)
##                28.91359                      71.08641
```

This discrepancy may be due to the fact that house cleaning requests are relatively constant over seasons/months, but moving may happen more frequently during summer seasons (July - August) especially for college students. We may need to look at overall moving trends in America to see whether there is seasonality in the data.

Percent Visitors: Device

```
##
## desktop  mobile
## 53.82859 46.17141
```

Percent Visitors: Combined

```
##
##      House Cleaning (One Time) Local Moving (under 50 miles)
## desktop                13.81259                      40.01600
## mobile                 15.10101                      31.07040
```

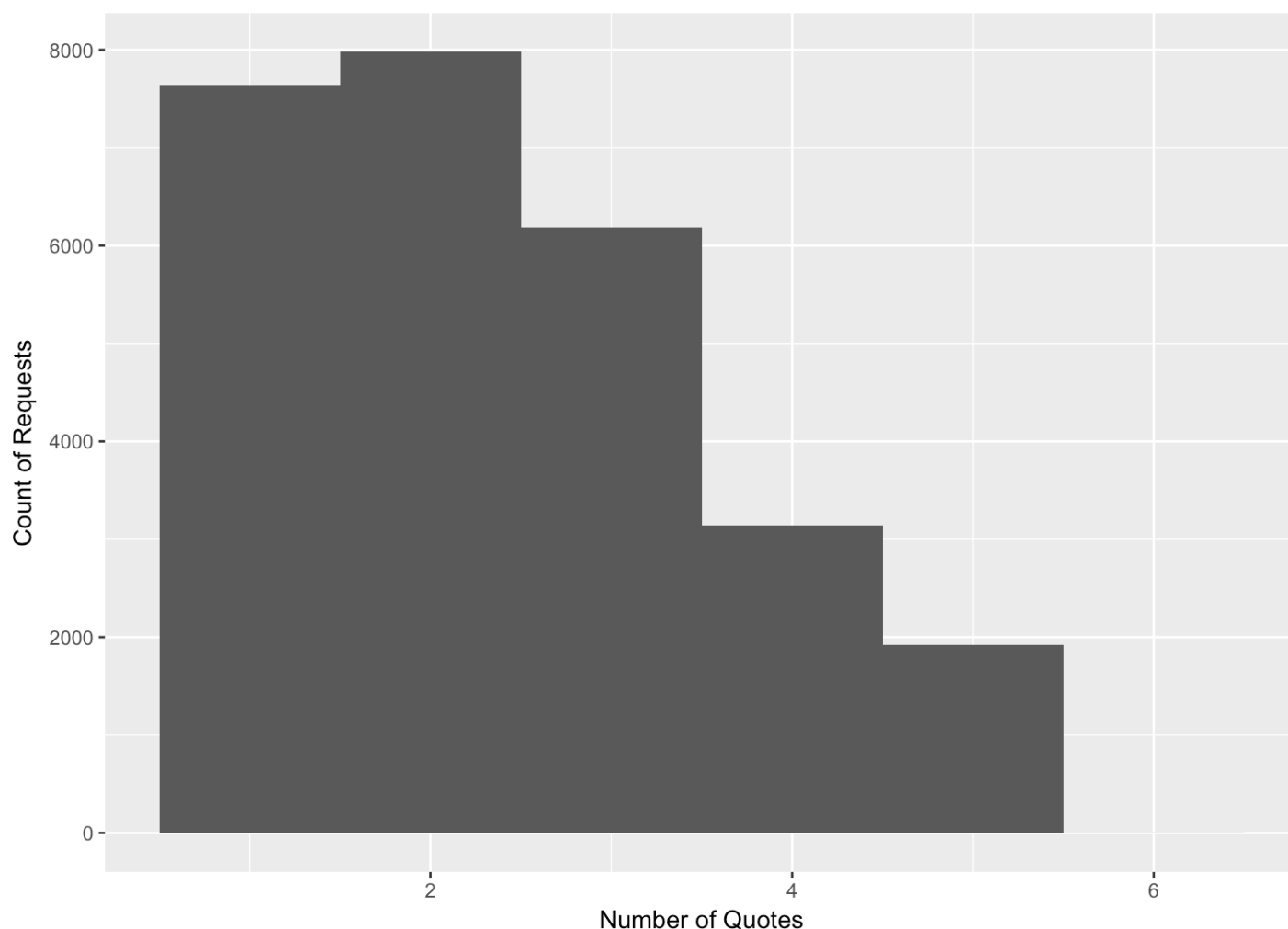
As explained before, the large discrepancy in the proportion of number of visitors in need of Local Moving Help by device may be explained by the fact that these customers want to be more thorough by using a Desktop interface. In contrast, there is little discrepancy in the House Cleaning category between the two devices.

2. Quotes Per Request

a) Distribution of Number of Quotes per Request

A Histogram of Number of Quotes Per Request:

Distribution of Number of Quotes Per Request



Five-number summary of Quotes per Request:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.000	1.000	2.000	2.395	3.000	6.000

Mean and Standard Deviation of Quotes per Request:

```
## [1] 2.3951
```

```
## [1] 1.212926
```

Based on the histogram, we can see that the distribution of number of quotes per request is skewed to the right; most of the requests only receive 1-2 quotes, with only a few receiving 5, and very rarely 6 quotes per request.

b) Factors Contributing to Distribution

I postulate that it is due to the category, how_far for Moving, and num_bedrooms/num_bathrooms for Cleaning. The further you have to move, the higher price you can charge, and thus more quotes for that request. Similar logic applies for cleaning for number of rooms to clean.

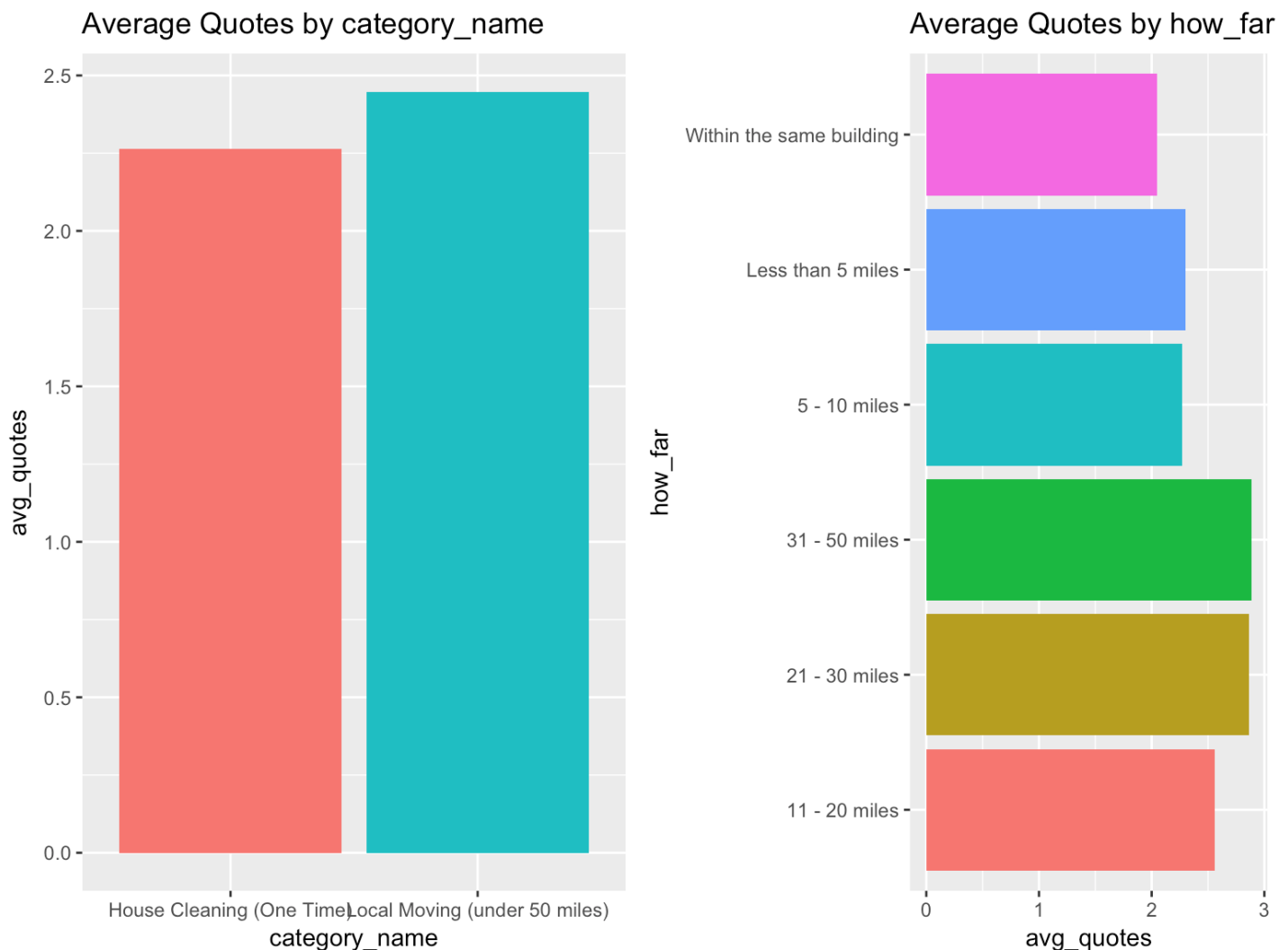
Before analysis, we note that not all the request_ids in visitors are in quotes dataset:

```
table(visitors[visitors$sent_request == 1, ]$request_id %in% grouped$request_id)
```

```
##
## FALSE TRUE
## 7287 26859
```

We will only deal with visitors data which have request_ids in the quotes dataset, meaning only requests that have received quotes after sending it in; we are making this selective process because our question is trying to understand what factors contribute to some requests getting more *quotes* than others, and with my postulation, we definitely need the visitors' request data as well.

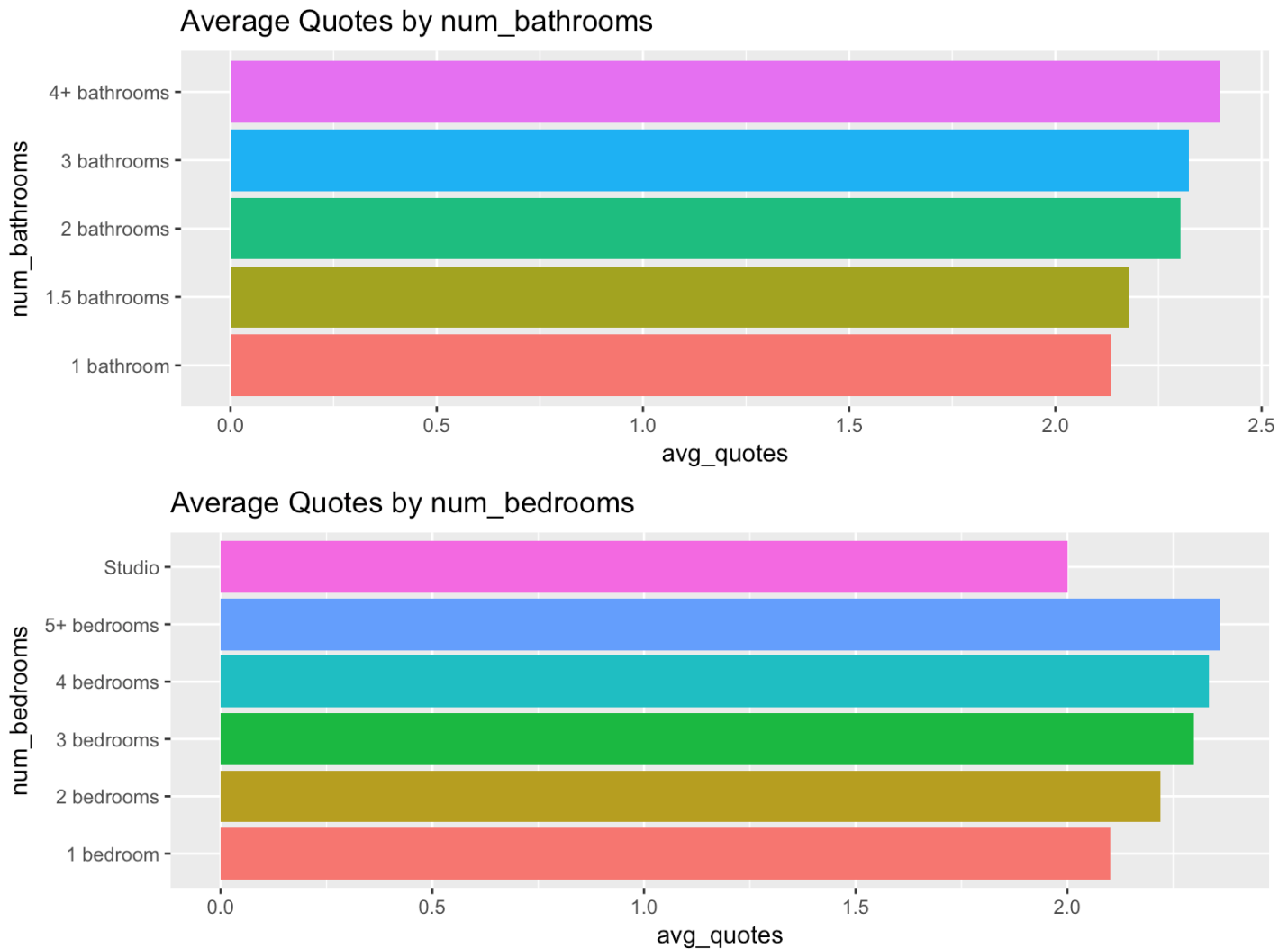
Plotting bar charts separated by category, and separated by distance for Moving requests:



We can see that Moving on average receives more quotes than house cleaning. This may be due to 1) attracting more professionals perhaps due to the higher price they can charge to help moving, or 2) simply more professionals are into Moving, regardless of reason.

Also, we can see that for the Moving Requests, the further the distance is, the more quotes a Moving request receives. This may be due to the fact that compared to the inconvenience and cost a professional burdens due to a longer distance, the revenue it receives is still more attractive than that of the shorter distance Moving requests.

Next, plotting graphs separated by number of bedrooms and number of bathrooms for Cleaning requests:



As for Cleaning Requests, we can also see a similar trend: the larger the bathroom and bedrooms are, the more quotes the Cleaning requests receive. Again similar logic applies - larger the surface area to clean, the more revenue.

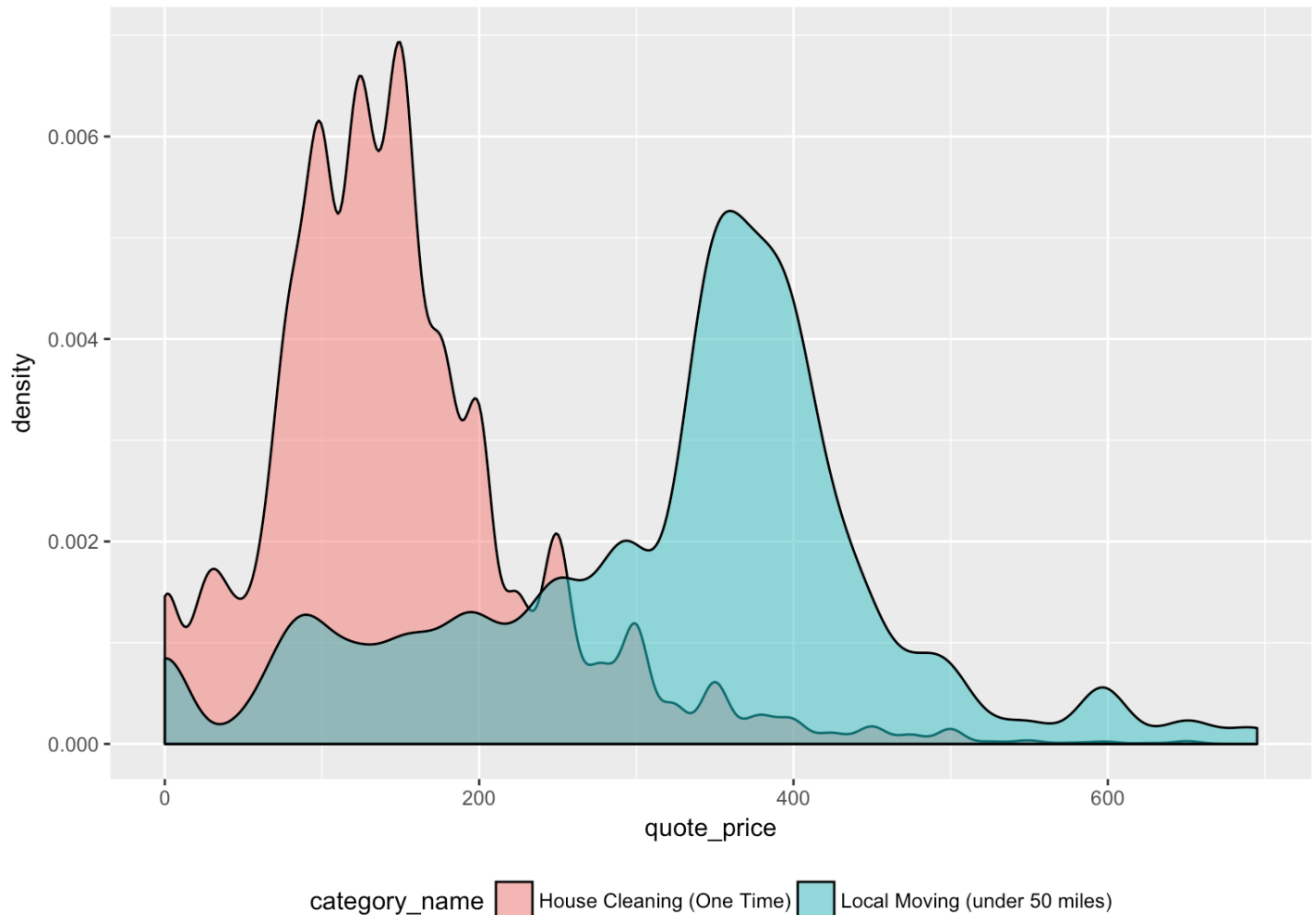
Thus, we conclude that in general, the amount of work correlates positively with the amount of quotes received in both moving and cleaning requests, due to one main reason: potentially increased revenue.

3. Job Value

a) Distribution of Prices

Note that there is missing data for the quote_prices for some quotes. We will omit these data for now, as there is no reasonable method to predict these missing quote prices.

Distributions of Quote Prices separated by Category



Though the prices are discrete variables, looking at the density plots we can see that Quote Prices for Moving requests are relatively normally distributed in contrast to the Quote Prices for Cleaning requests, which are skewed to the right. This may be due to Cleaning requests with many bedrooms and bathrooms, which may mean a higher economic class and hence the consumer's willingness to pay may be much higher than normal, creating an outlier in quote price charged.

Also to note is that the mean of quote prices for moving is larger than that of house cleaning, mainly because Moving on average costs more for a professional to conduct than cleaning.

b) Charging Prices by Thumbtack

For Cleaning Requests, I believe Thumbtack should charge pros a proportion of their quote price. We clearly saw that the distribution of quote prices pros charge for requests vary greatly, and is quite unstable. Thus, instead of charging a flat fee per quote, I believe the pricing strategy should also vary to make sure professionals cannot take advantage of Thumbtack, nor are treated with unfair prices based on the requests that come in for Cleaning. This proportion may be a 3% quoting fee.

In contrast, we can see that the quote prices for Moving requests are somewhat normally distributed, with less outliers and thus variability than that of Cleaning requests - the quote prices are relatively more stable. Thus, perhaps for Cleaning requests a flat fee per month may be a better option. This will enable professionals to quote different ranges of prices within a stable distribution we see here, and on average will be a fair price compared to the quotes received. This price per month may be calculated as:

```
#Assuming we have a month's worth of quoting data for professionals:
hired <-
  pdata_joined %>% group_by(., pro_id) %>%
  filter(., hired == 1) %>% summarise(., num_hired = n())
hired <- mean(hired$num_hired)
#So per pro, get number of times hired per month (if quote got hired, that means he got
hired) -> and get the mean of number of hired

#avg hired * avg price quoted * 0.03
move_price <- hired * mean(na.omit(pdata_joined$quote_price)) * 0.03

move_price
```

```
## [1] 26.42095
```

4. Further Explanation

a) How Elastic is the Supply of Professionals?

We saw previously that the number of requests diminishes as the weekday moves into the weekend. Do the number of professionals active (submitting quotes) follow that trend? Or rather, is it that the number of professionals is relatively fixed, but the number of quotes from pre-existing professionals simply increase? We would be interested in doing this analysis because we can flexibly alter pricing strategies based on number of professionals and the quotes they are submitting.

b) How many requests do not hire a professional - in other words, how many requests do not reach the final match between quotes/professionals and customers?

Something to notice is that some of the requests actually do not eventually hire a professional, even when many professionals are quoting for the request. We could examine which kind of requests especially show this trend, and this exploratory data analysis may lead leeways recognizing a disequilibrium in the market, suggesting better quoting strategies to professionals, and potentially asking customers for their reasons for not hiring a professional.