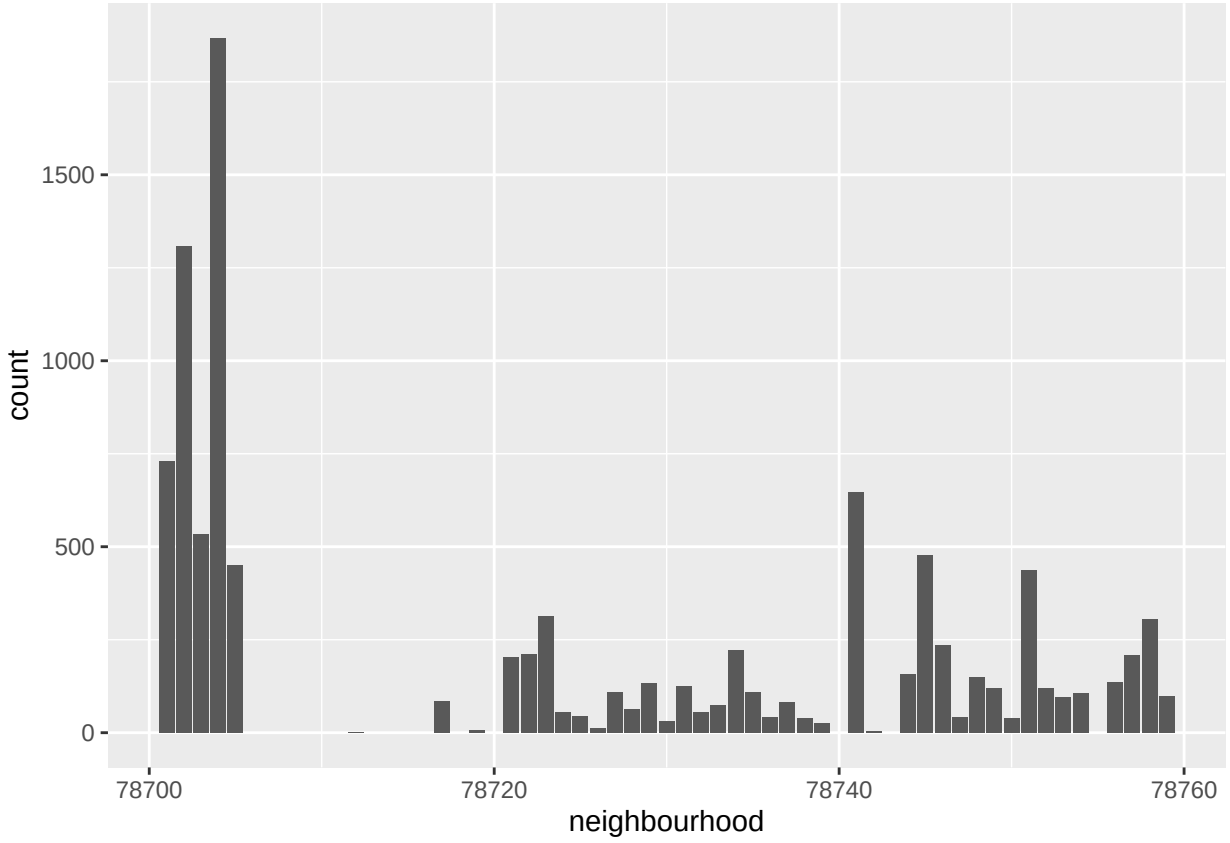
Part I

Load the necessary libraries and datasets

Variation

Perform an analysis of the variation in the "neighbourhood" column.

```
ggplot(data = listings) + geom_bar(mapping = aes(x = neighbourhood))
```




```
listings %>%
  count(neighbourhood) %>%
  arrange(desc(n))
```

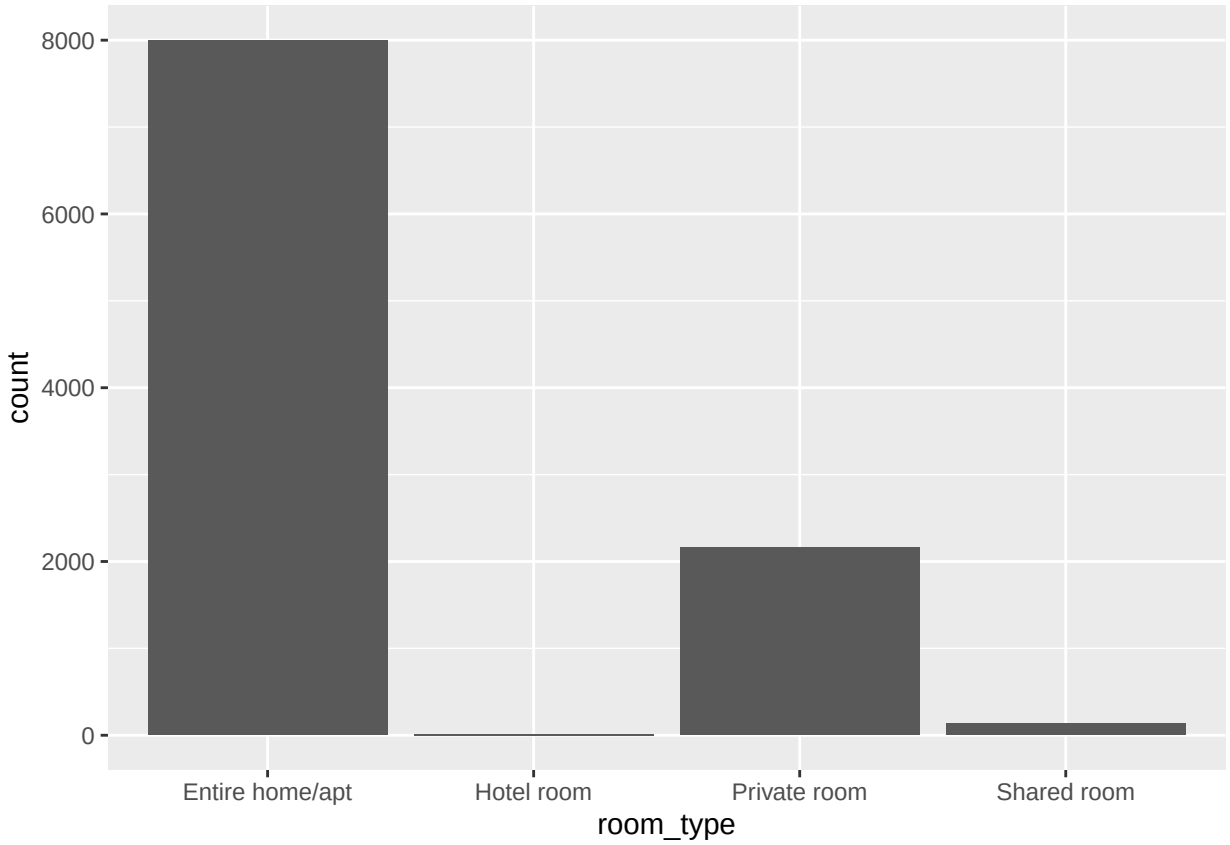
```
## # A tibble: 44 x 2
##    neighbourhood     n
##            <dbl> <int>
##  1         78704  1868
##  2         78702  1307
##  3         78701   730
##  4         78741   647
##  5         78703   534
##  6         78745   477
##  7         78705   451
```

```
##  8         78751   436
##  9         78723   313
## 10         78758   305
## # ... with 34 more rows
```

- Which values are the most common? Why?
- Which values are rare? Why? Does that match your expectations?
- Can you see any unusual patterns? What might explain them?

Perform an analysis of the variation in the "room_type" column.

```
ggplot(listings) + geom_bar(mapping = aes(x = room_type))
```



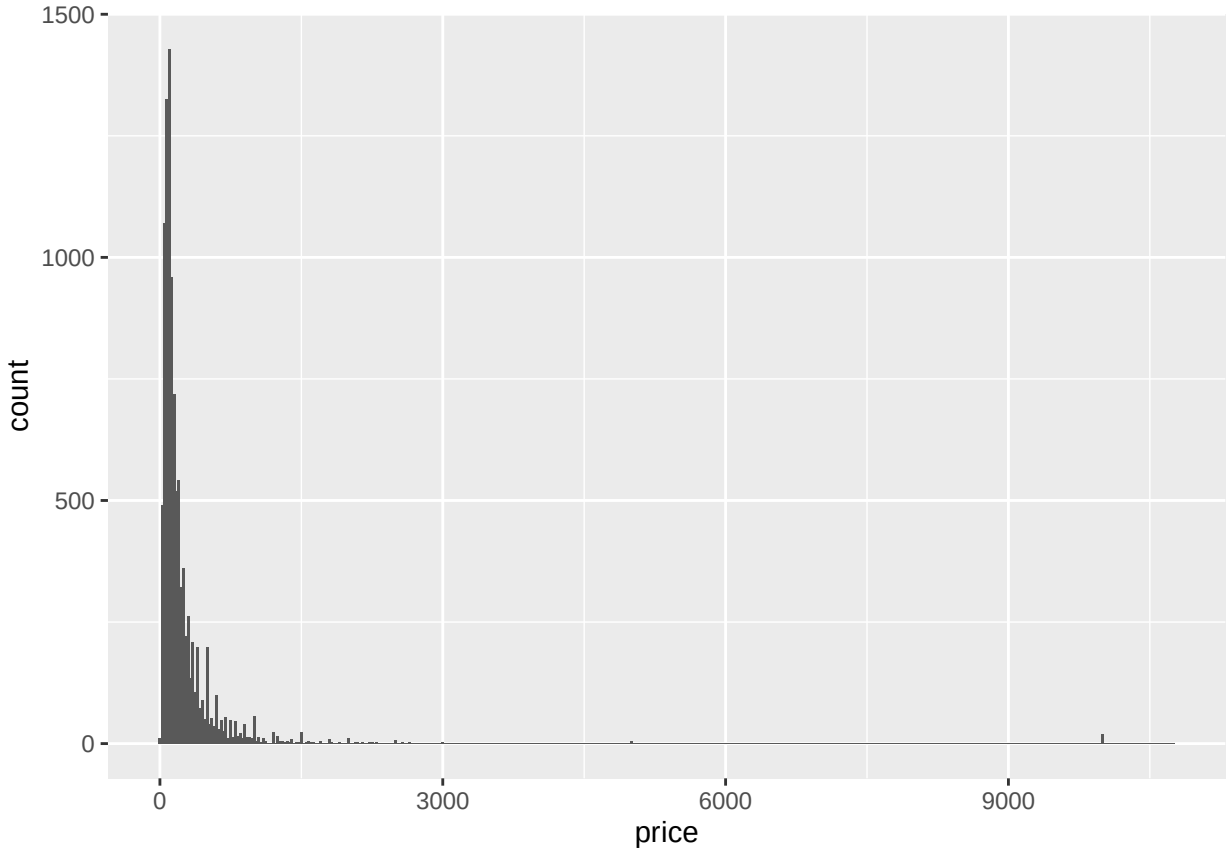

```
listings %>%
  count(room_type) %>%
  arrange(desc(n))
```

```
## # A tibble: 4 x 2
##   room_type           n
##   <chr>           <int>
## 1 Entire home/apt  7997
## 2 Private room     2161
## 3 Shared room       134
## 4 Hotel room         13
```

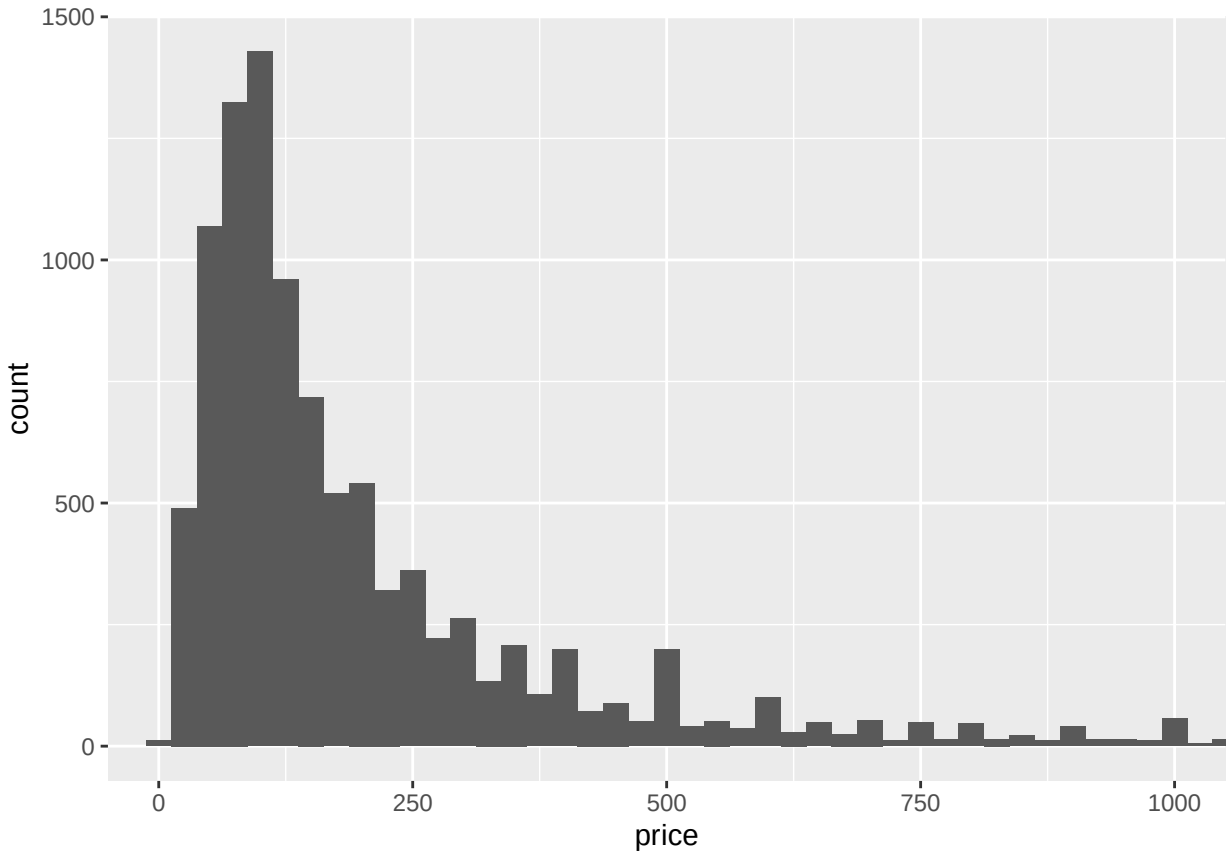
- Which values are the most common? Why?
- Which values are rare? Why? Does that match your expectations?
- Can you see any unusual patterns? What might explain them?

Perform an analysis of the variation in the "price" column. Make sure to explore different "binwidth" values in your analysis.

```
ggplot(listings) + geom_histogram(mapping = aes(x = price), binwidth = 25)
```




```
#zoomed-in
ggplot(listings) + geom_histogram(mapping = aes(x = price), binwidth = 25) +
  coord_cartesian(xlim = c(0, 1000))
```




```
listings %>%
  count(price) %>%
  arrange(desc(n))
```

```
## # A tibble: 878 x 2
##    price     n
##    <dbl> <int>
##  1   150   237
##  2   100   221
##  3   200   217
##  4    75   175
##  5   250   163
##  6    50   156
##  7    80   150
##  8   125   148
##  9   300   134
## 10    85   126
## # ... with 868 more rows
```

```
listings %>%
  count(price) %>%
  arrange(desc(price))
```

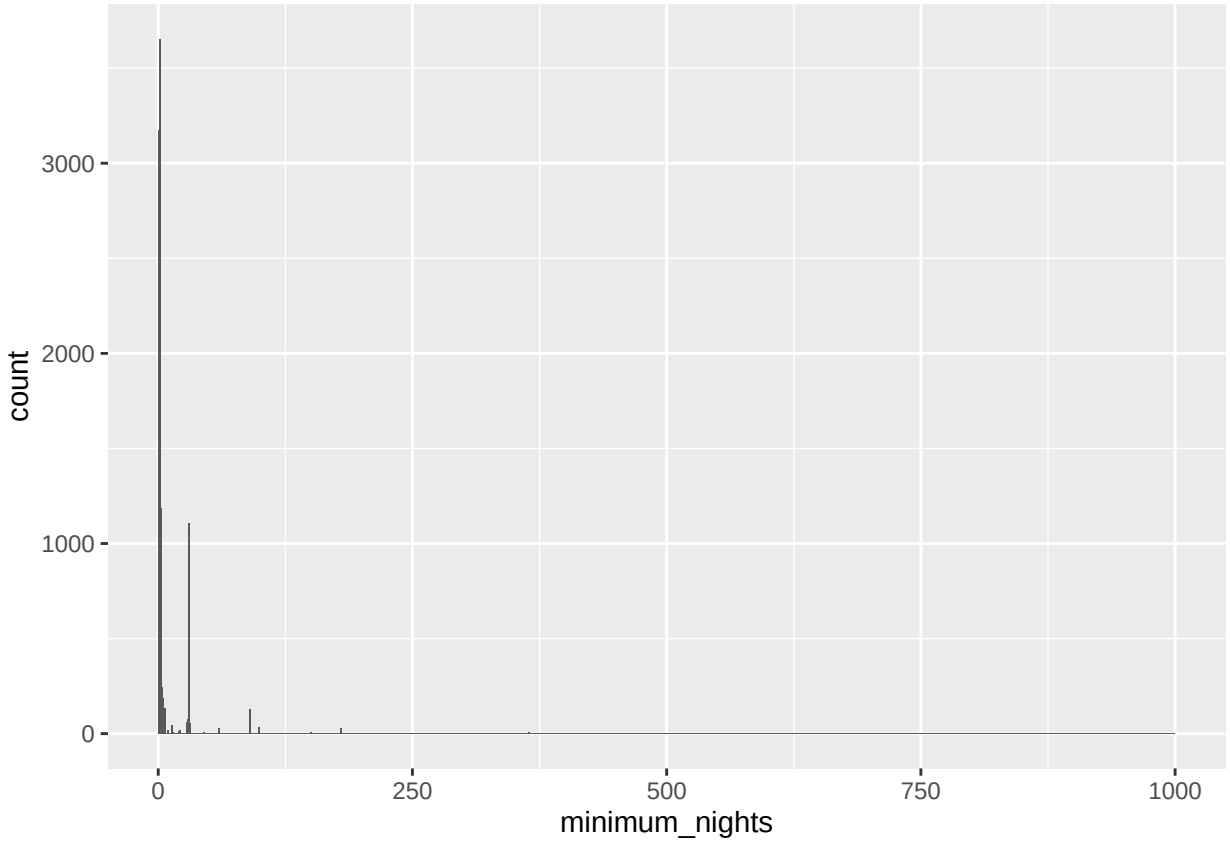
```
## # A tibble: 878 x 2
##    price     n
```

```
##    <dbl> <int>
##  1 10754     1
##  2 10000     5
##  3  9999    13
##  4  9998     1
##  5  9435     1
##  6  7229     1
##  7  5978     1
##  8  5550     1
##  9  5479     1
## 10  5413     1
## # ... with 868 more rows
```

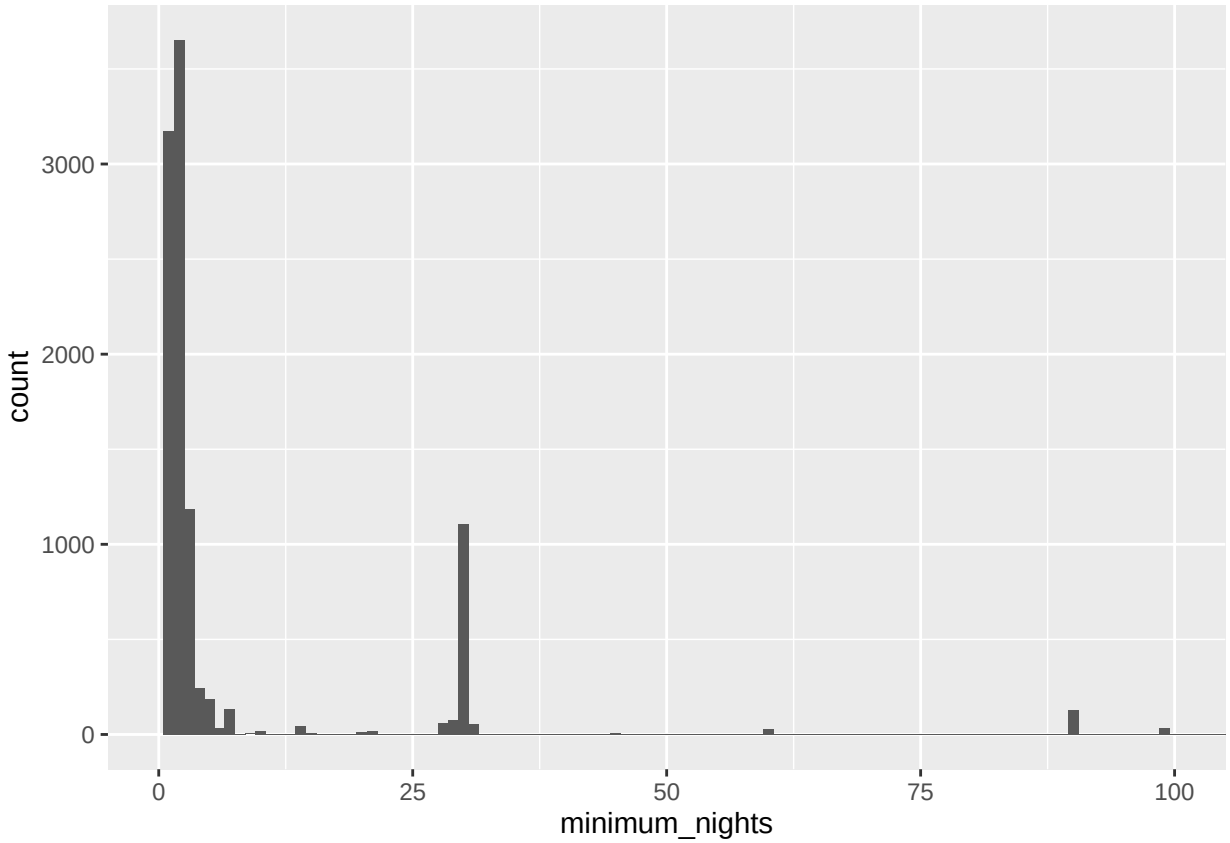
- Which values are the most common? Why?
- Which values are rare? Why? Does that match your expectations?
- Can you see any unusual patterns? What might explain them?

Perform an analysis of the variation in the "minimum_nights" column. Make sure to explore different "binwidth" values in your analysis.

```
ggplot(listings) + geom_histogram(mapping = aes(x = minimum_nights), binwidth = 1)
```



```
#zoomed-in
ggplot(listings) + geom_histogram(mapping = aes(x = minimum_nights), binwidth = 1) +
  coord_cartesian(xlim = c(0, 100))
```




```
listings %>%
  count(minimum_nights) %>%
  arrange(desc(n))
```

```
## # A tibble: 57 x 2
##    minimum_nights     n
##             <dbl> <int>
##  1              2  3654
##  2              1  3172
##  3              3  1185
##  4             30  1107
##  5              4   242
##  6              5   188
##  7              7   133
##  8             90   131
##  9             29    74
## 10             28    61
## # ... with 47 more rows
```

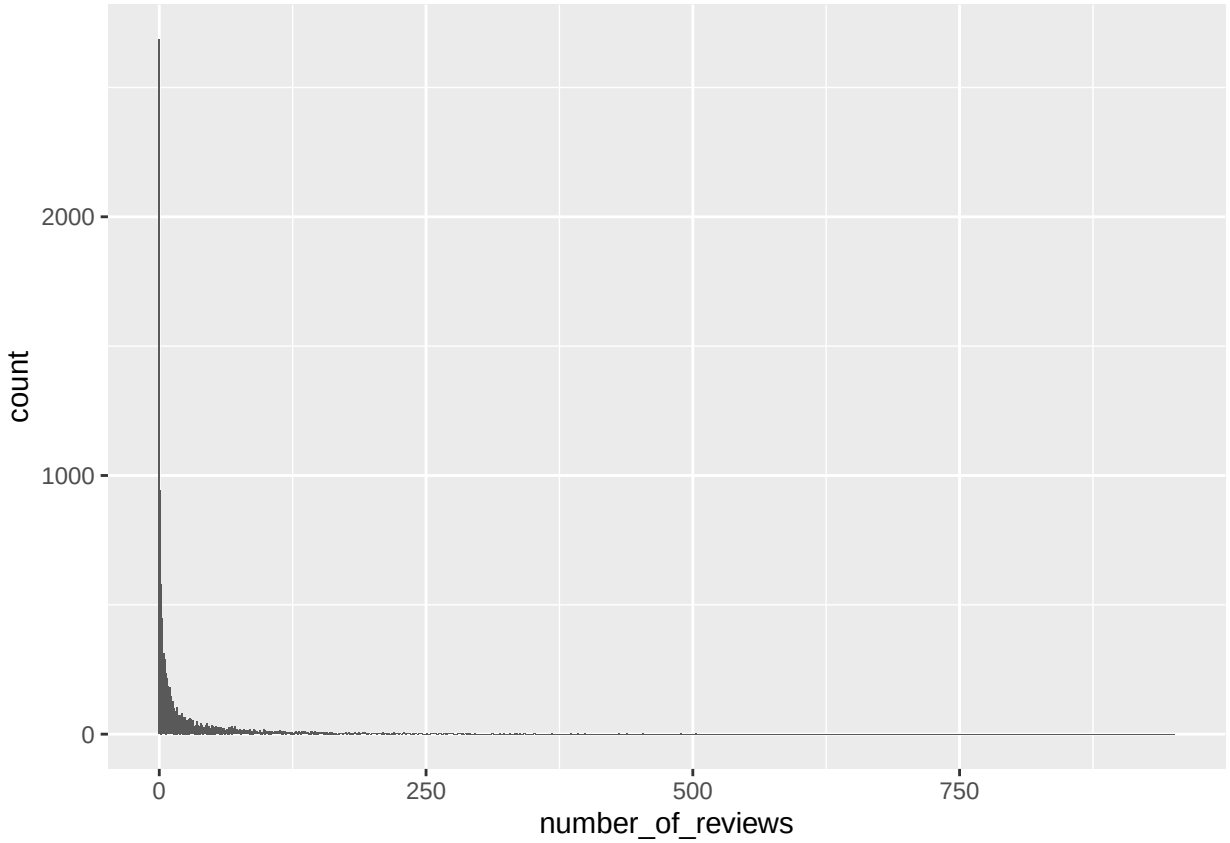
```
listings %>%
  count(minimum_nights) %>%
  arrange(desc(minimum_nights))
```

```
## # A tibble: 57 x 2
##    minimum_nights     n
##             <dbl> <int>
##  1            999     1
##  2            500     1
##  3            365     6
##  4            360     3
##  5            300     1
##  6            290     1
##  7            240     1
##  8            200     2
##  9            186     1
## 10            183     1
## # ... with 47 more rows
```

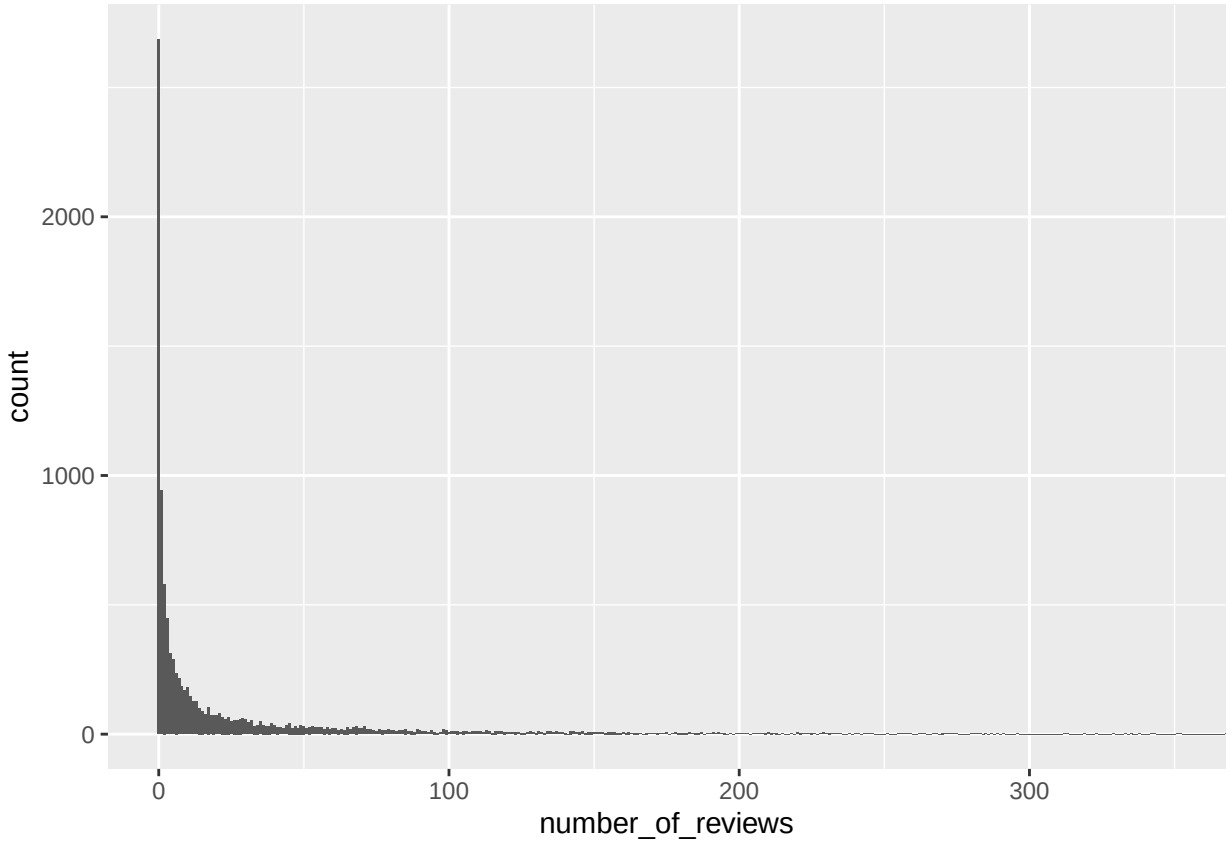
- Which values are the most common? Why?
- Which values are rare? Why? Does that match your expectations?
- Can you see any unusual patterns? What might explain them?

Perform an analysis of the variation in the "number_of_reviews" column. Make sure to explore different "binwidth" values in your analysis.

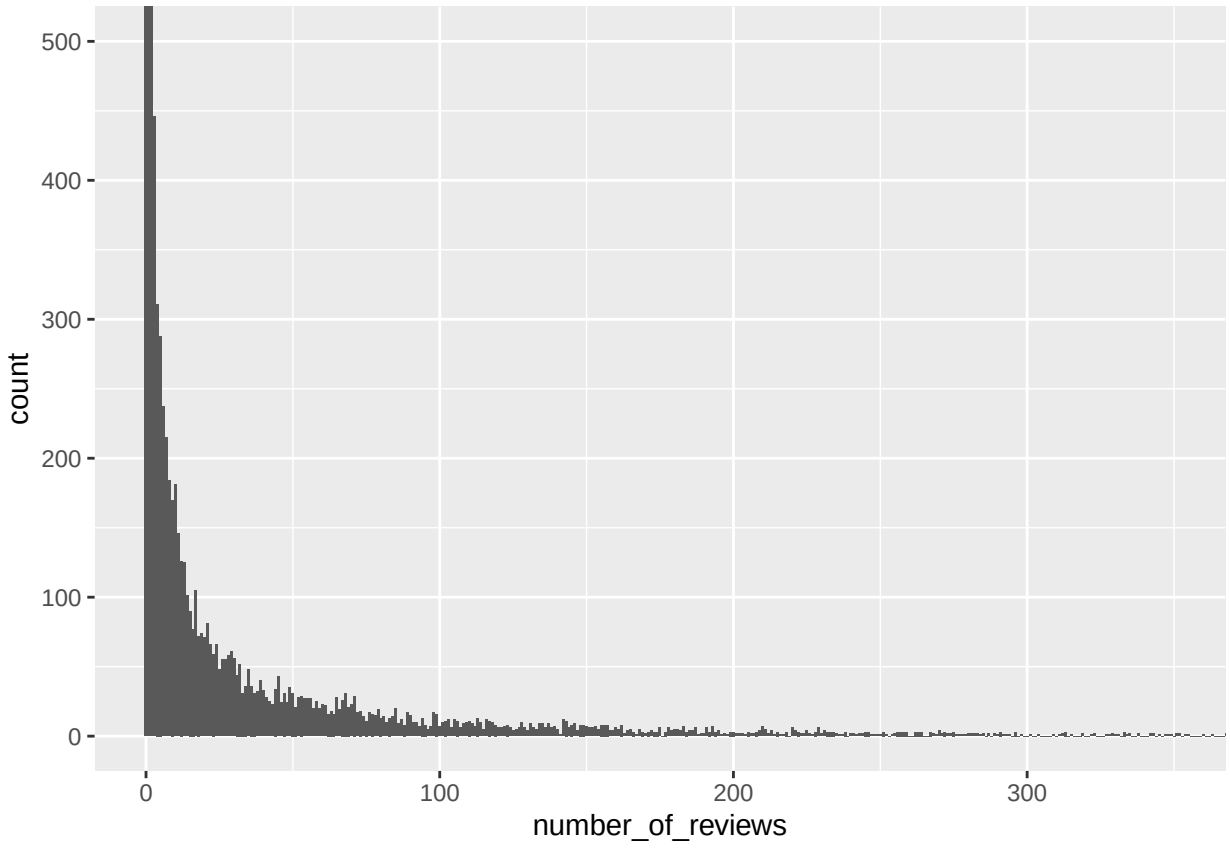
```
ggplot(listings) + geom_histogram(mapping = aes(x = number_of_reviews), binwidth = 1)
```




```
#zoomed-in x axis
ggplot(listings) + geom_histogram(mapping = aes(x = number_of_reviews), binwidth = 1) +
  coord_cartesian(xlim = c(0, 350))
```




```
#zoomed-in x and y axis
ggplot(listings) + geom_histogram(mapping = aes(x = number_of_reviews), binwidth = 1) +
  coord_cartesian(xlim = c(0, 350), ylim = c(0, 500))
```




```
listings %>%
  count(number_of_reviews) %>%
  arrange(desc(n))
```

```
## # A tibble: 383 x 2
##    number_of_reviews     n
##                <dbl> <int>
##  1                 0  2686
##  2                 1   941
##  3                 2   580
##  4                 3   446
##  5                 4   311
##  6                 5   288
##  7                 6   237
##  8                 7   215
##  9                 8   184
## 10                10   181
## # ... with 373 more rows
```

```
listings %>%
  count(number_of_reviews) %>%
  arrange(desc(number_of_reviews))
```

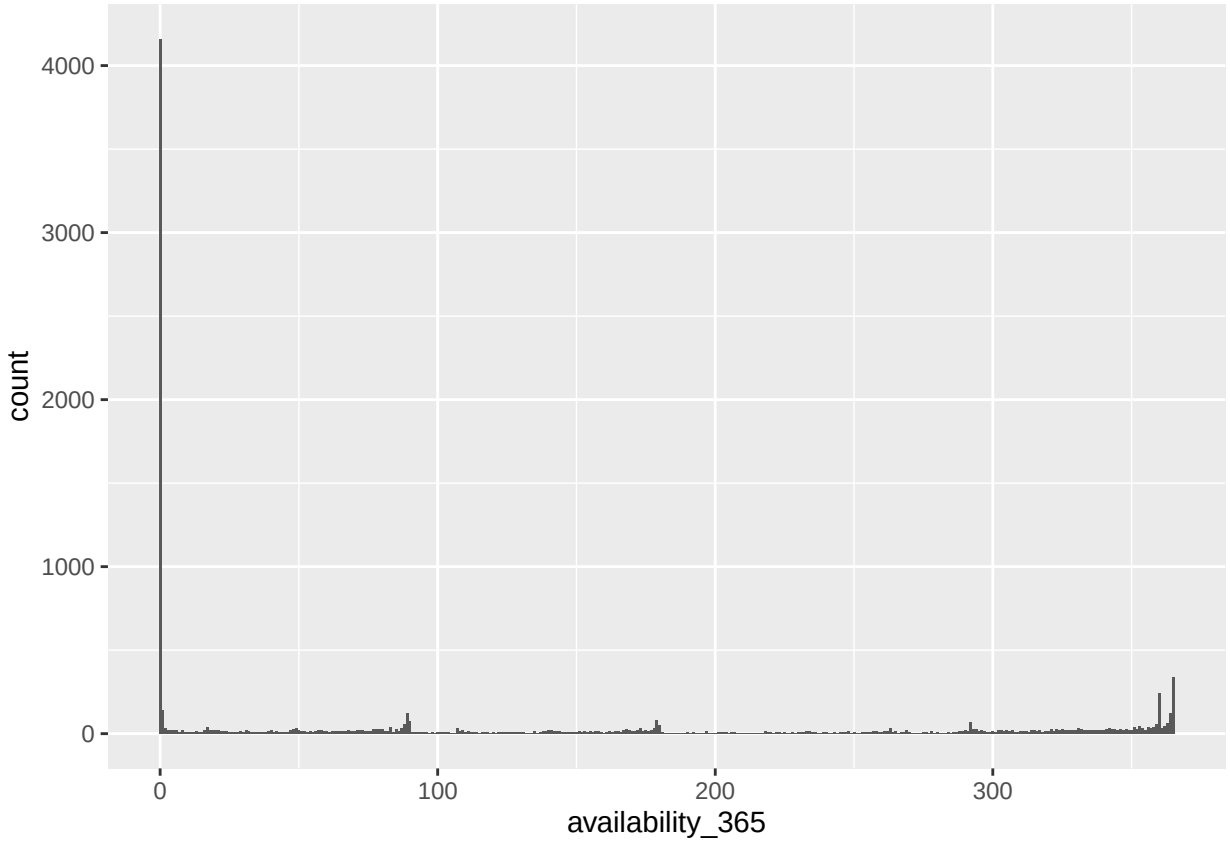
```
## # A tibble: 383 x 2
##    number_of_reviews     n
```

```
##                <dbl> <int>
##  1               951     1
##  2               836     1
##  3               825     1
##  4               746     1
##  5               745     1
##  6               720     1
##  7               690     1
##  8               689     1
##  9               677     1
## 10               651     1
## # ... with 373 more rows
```

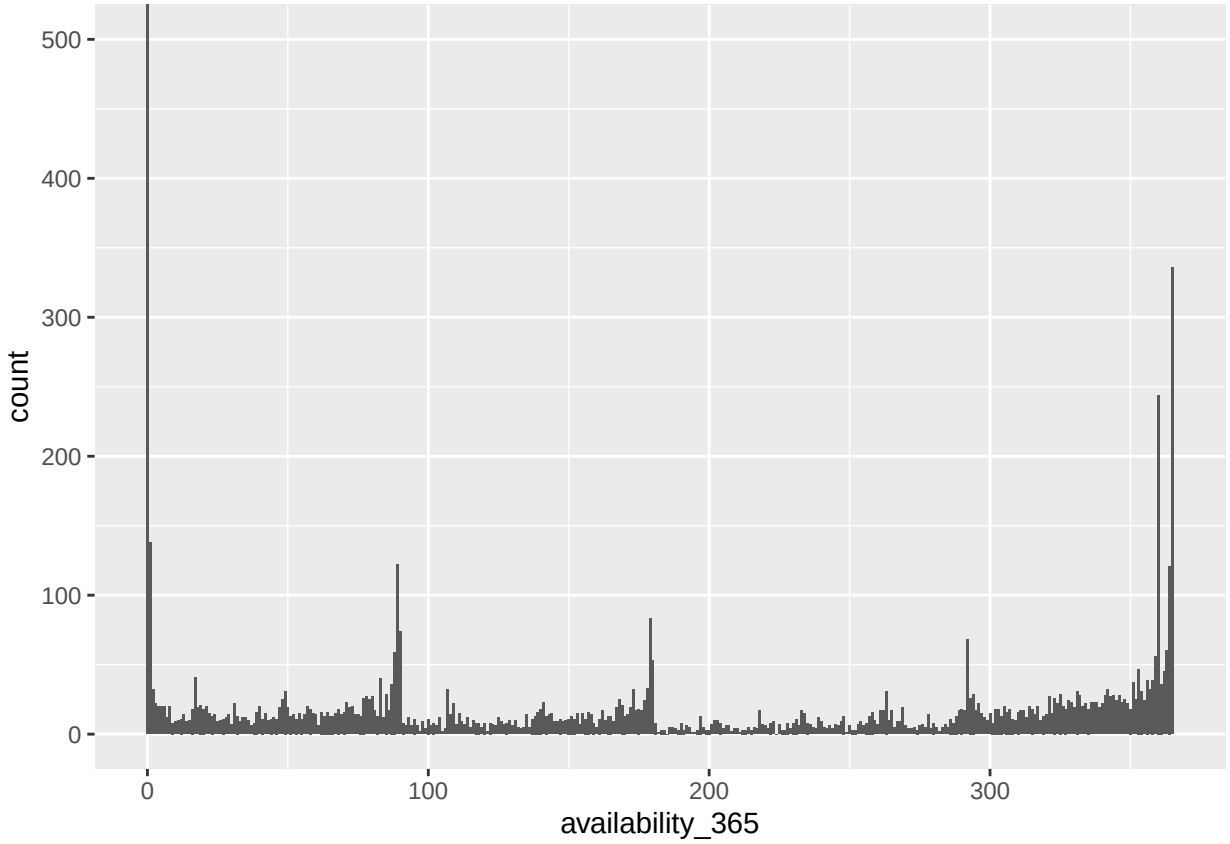
- Which values are the most common? Why?
- Which values are rare? Why? Does that match your expectations?
- Can you see any unusual patterns? What might explain them?

Perform an analysis of the variation in the "availability_365" column. Make sure to explore different "binwidth" values in your analysis.

```
ggplot(listings) + geom_histogram(mapping = aes(x = availability_365), binwidth = 1)
```



```
#zoomed-in x axis
ggplot(listings) + geom_histogram(mapping = aes(x = availability_365), binwidth = 1) +
  coord_cartesian(ylim = c(0, 500))
```




```
listings %>%
  count(availability_365) %>%
  arrange(desc(n))
```

```
## # A tibble: 364 x 2
##    availability_365     n
##               <dbl> <int>
##  1                0  4160
##  2              365   336
##  3              360   244
##  4                1   138
##  5               89   122
##  6              364   121
##  7              179    83
##  8               90    74
##  9              292    68
## 10              363    60
## # ... with 354 more rows
```

```
listings %>%
  count(availability_365) %>%
  arrange(desc(availability_365))
```

```
## # A tibble: 364 x 2
##    availability_365     n
##               <dbl> <int>
##  1              365   336
##  2              364   121
##  3              363    60
##  4              362    45
##  5              361    36
##  6              360   244
##  7              359    56
##  8              358    39
##  9              357    32
## 10              356    39
## # ... with 354 more rows
```

- Which values are the most common? Why?
- Which values are rare? Why? Does that match your expectations?
- Can you see any unusual patterns? What might explain them?

Part II

Use your dataset to answer the following questions:

- What seems to be the most common name (of a person) in the city you selected?

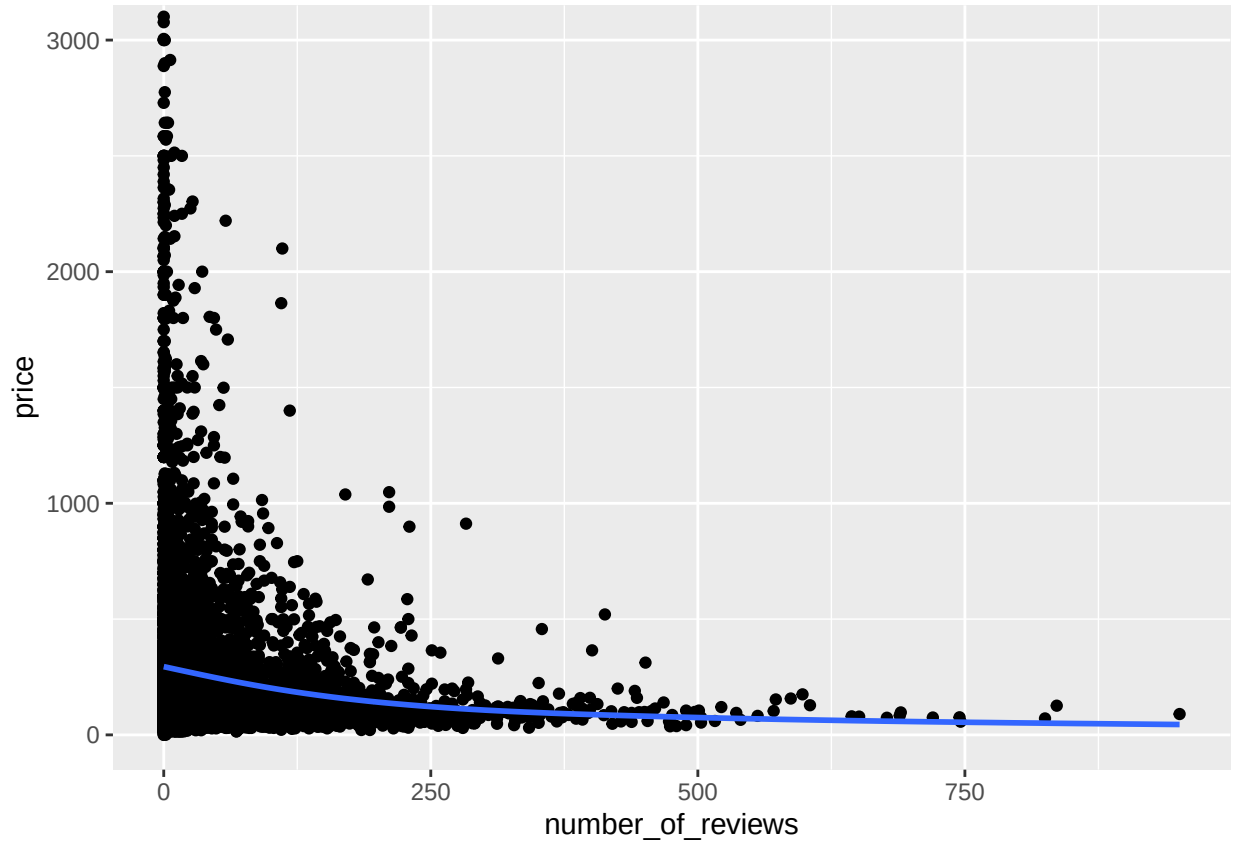
```
listings %>%
  count(host_name) %>%
  arrange(desc(n))
```

```
## # A tibble: 2,705 x 2
##    host_name                    n
##    <chr>                    <int>
##  1 Kia                        539
##  2 WanderJaunt                119
##  3 TurnKey Vacation Rentals    99
##  4 Martin                      87
##  5 Michael                     83
##  6 David                       80
##  7 Ryan                        76
##  8 Sarah                       71
##  9 Jennifer                    68
## 10 James                       60
## # ... with 2,695 more rows
```

- Do the number of reviews affect the price of the Airbnb? How? Why do you think this happens?

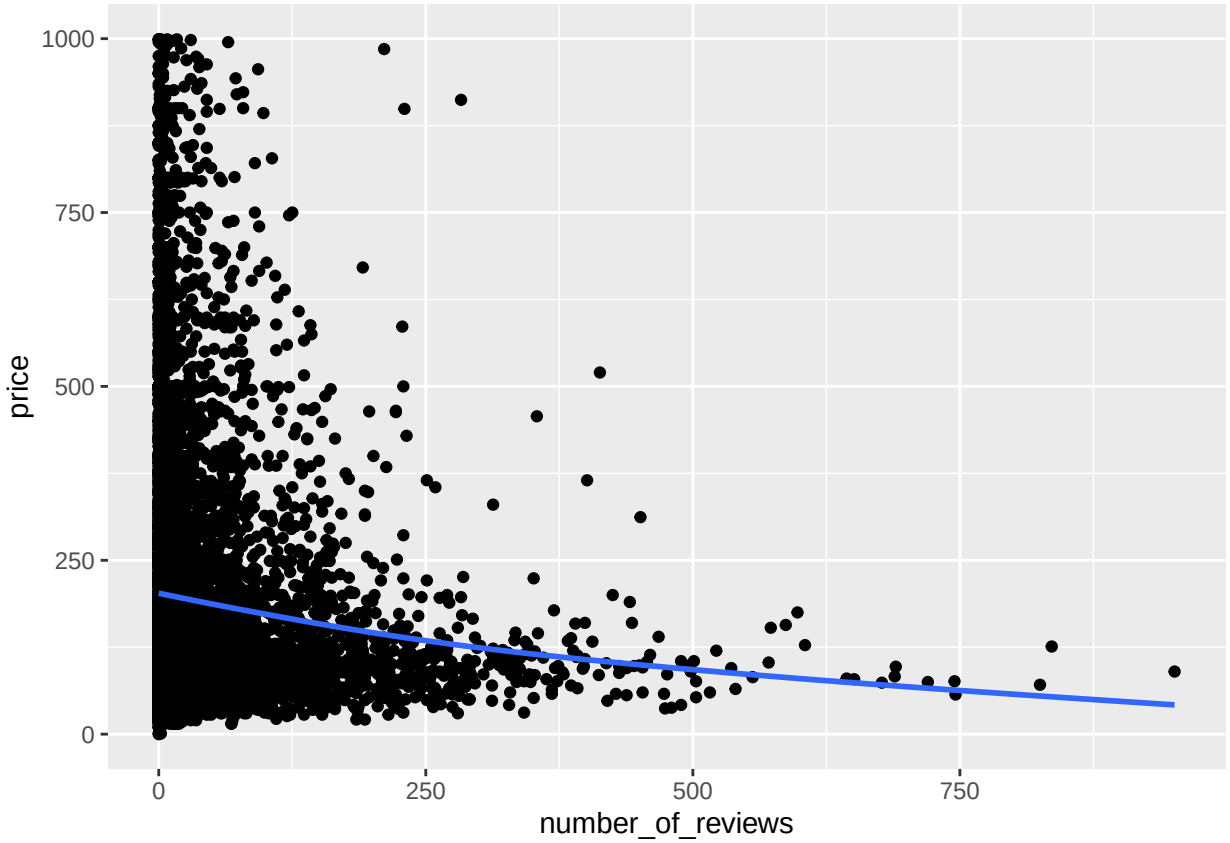
```
#accounting for all values
ggplot(data = listings, mapping = aes(x = number_of_reviews, y = price)) + geom_point() + geom_smooth(s
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



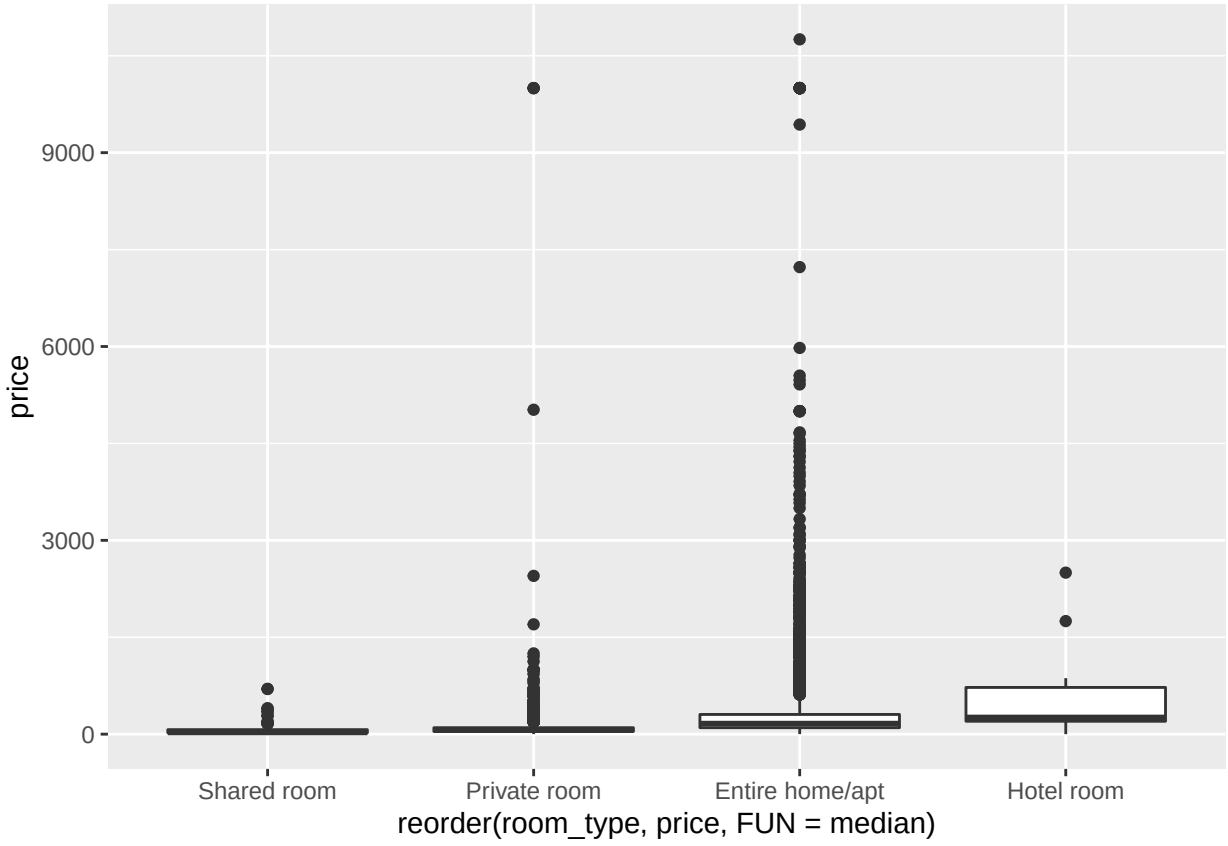

```
#not accounting for expensive places
under_one_thousand <- listings %>%
  filter(price < 1000)

  ggplot(data = under_one_thousand, mapping = aes(x = number_of_reviews, y = price)) + geom_point() + g
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

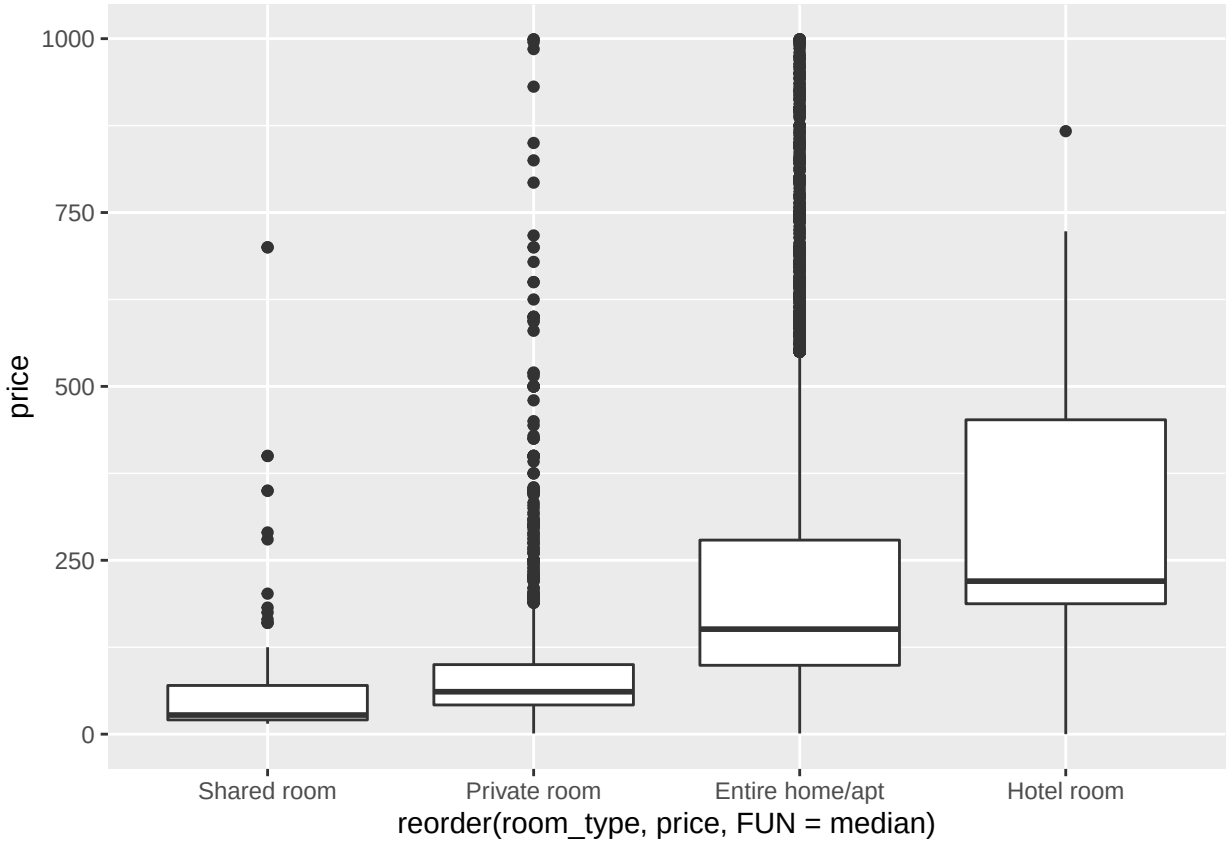


- What type of room tends to have the highest Airbnb price?

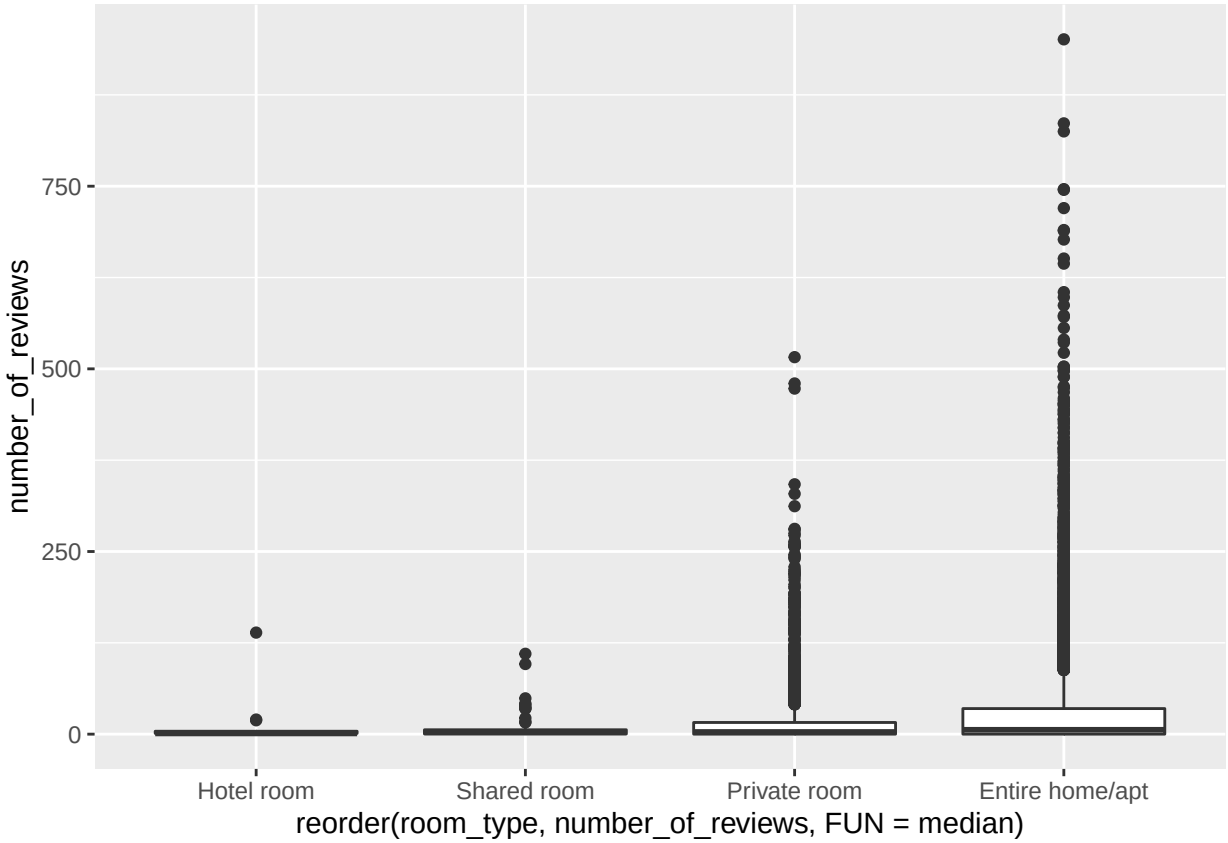
```
ggplot(data = listings) +
  geom_boxplot(mapping = aes(x = reorder(room_type, price, FUN = median), y = price))
```



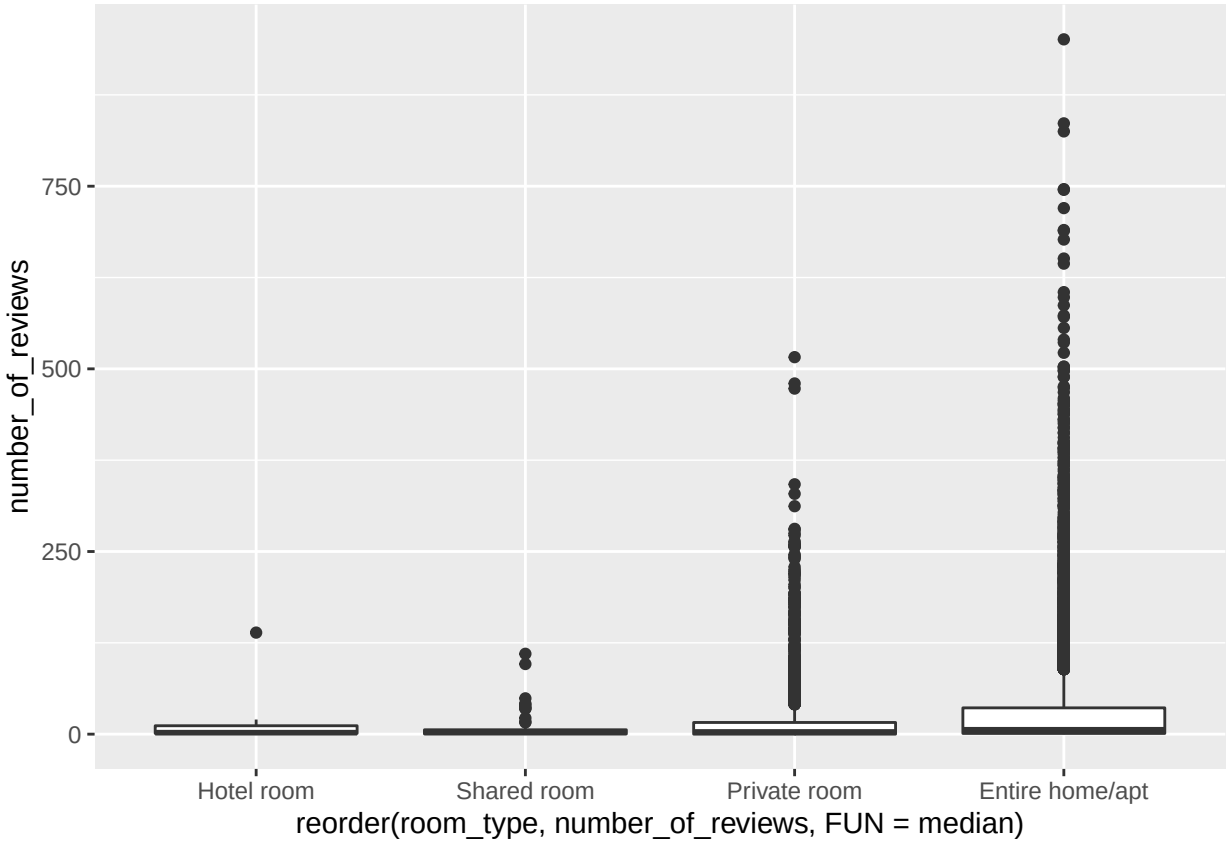

```
ggplot(data = under_one_thousand) +
  geom_boxplot(mapping = aes(x = reorder(room_type, price, FUN = median), y = price))
```




```
ggplot(data = listings) +
  geom_boxplot(mapping = aes(x = reorder(room_type, number_of_reviews, FUN = median), y = number_of_revi
```

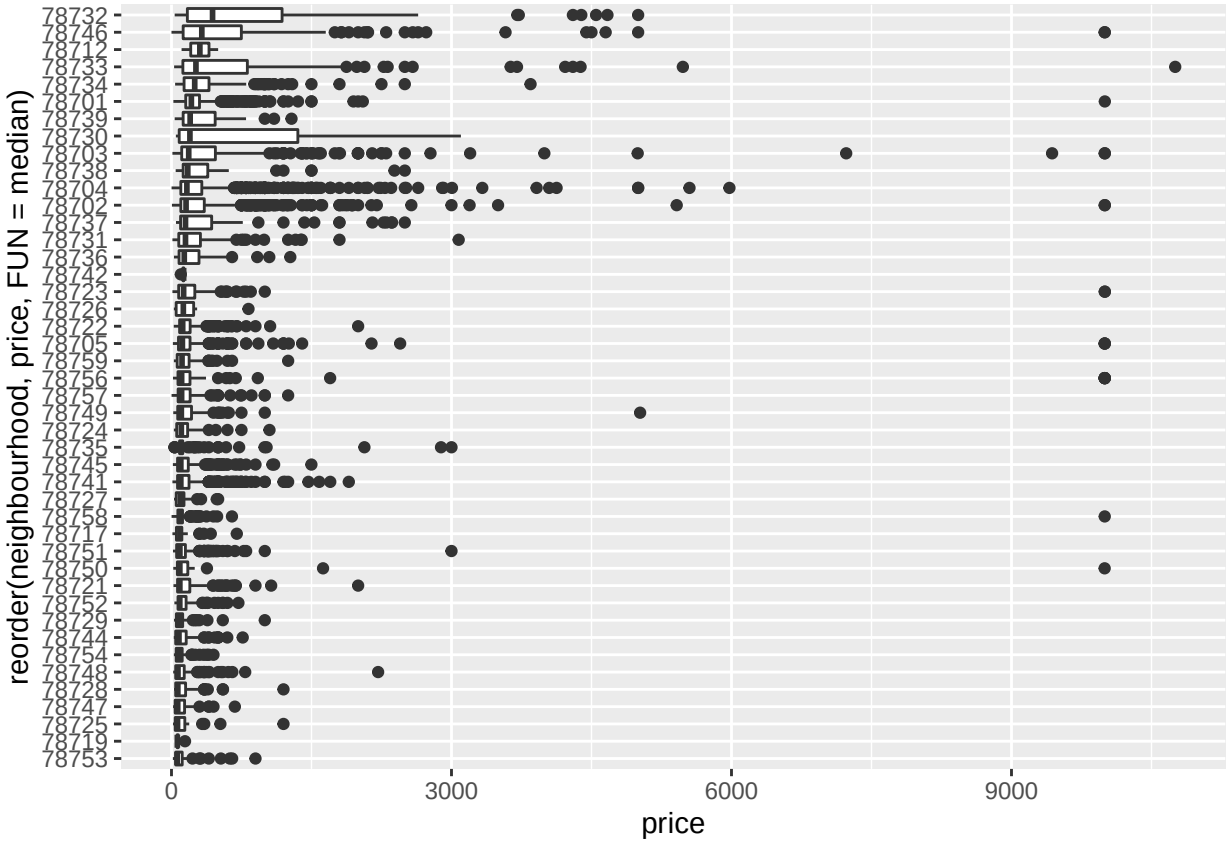



```
ggplot(data = under_one_thousand) +
  geom_boxplot(mapping = aes(x = reorder(room_type, number_of_reviews, FUN = median), y = number_of_revi
```



- What neighborhood(s) tend to have the highest Airbnb price?

```
ggplot(data = listings) +
  geom_boxplot(mapping = aes(x = reorder(neighbourhood, price, FUN = median), y = price)) + coord_flip()
```




```
by_neighborhood <- group_by(listings, neighbourhood) %>%
  summarize(avg_price = mean(price)) %>%
  arrange(desc(avg_price))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
by_neighborhood
```

```
## # A tibble: 44 x 2
##    neighbourhood avg_price
##            <dbl>     <dbl>
##  1         78732     1032.
##  2         78733      947.
##  3         78730      722.
##  4         78746      677.
##  5         78756      519.
##  6         78737      452.
##  7         78703      441.
##  8         78738      436.
##  9         78750      398.
## 10         78734      362.
## # ... with 34 more rows
```

- Suppose you could purchase a property in the city you selected, and that you could rent it to others as an Airbnb. In what neighborhood would you want to purchase your property? Why?

78732

Part III

- Visit a real estate website (such as realtor.com) and find a property that is for sale in the neighborhood you selected. Take note of the price and address of the property.
- Use your dataset to find what the average Airbnb price/night is in the neighborhood you selected.

```
by_neighborhood <- group_by(listings, neighbourhood) %>%
  summarize(avg_price = mean(price)) %>%
  arrange(desc(avg_price))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
by_neighborhood
```

```
## # A tibble: 44 x 2
##    neighbourhood avg_price
##            <dbl>     <dbl>
##  1         78732     1032.
##  2         78733      947.
##  3         78730      722.
##  4         78746      677.
##  5         78756      519.
##  6         78737      452.
##  7         78703      441.
##  8         78738      436.
##  9         78750      398.
## 10         78734      362.
## # ... with 34 more rows
```

- Use your dataset to find what the average number of available nights per year is for an Airbnb in the neighborhood you selected.

```
by_neighborhood <- group_by(listings, neighbourhood) %>%
  summarize(avg_availability = mean(availability_365)) %>%
  arrange(desc(avg_availability))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
by_neighborhood
```

```
## # A tibble: 44 x 2
##    neighbourhood avg_availability
##            <dbl>            <dbl>
##  1         78735             213.
##  2         78717             200.
##  3         78734             186.
##  4         78732             182.
```

```
##  5         78729            175.
##  6         78737            174.
##  7         78727            169.
##  8         78719            158
##  9         78752            155.
## 10         78754            154.
## # ... with 34 more rows
```

- Suppose you bought the property you selected above. If you were to rent it as an Airbnb at the average neighborhood price, for the average number of days, how long will it take you to break even?