NewYears

2023-03-01

Setup

Load packages

Input **Load the data**

Transformation **Summary statistics on the price of the cities for the listings dataset**

```
file_names <- c( 'listings-ams.csv.gz', "listings-london.csv.gz", "listings-paris.csv.gz", "listings-ro
city_names <- c('Amsterdam', 'London', 'Paris', 'Rome')

price_summary <- data.frame(city = character(),
                            min_price = numeric(),
                            max_price = numeric(),
                            mean_price = numeric(),
                            stringsAsFactors = FALSE)

# loop through each file and calculate the summary statistics
# A filter (price above 0) has been applied to remove outliers
for (i in 1:length(file_names)) {
  list_data <- read_csv(gzfile(paste0('../../data/',file_names[i])))
  list_prices <- list_data$price[!is.na(list_data$price)]
  list_prices <- as.numeric(gsub("\\$", "", list_prices))
  list_prices <- list_prices[!is.na(list_prices)]
  list_prices <- list_prices[list_prices > 0]
  list_prices_summary <- summary(list_prices)

  # add the results to the data frame
  price_summary[i, "city"] <- city_names[i]
  price_summary[i, "min_price"] <- list_prices_summary["Min."]
  price_summary[i, "max_price"] <- list_prices_summary["Max."]
  price_summary[i, "mean_price"] <- mean(list_prices)
}
```

```
## Rows: 6809 Columns: 75
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (25): listing_url, source, name, description, neighborhood_overview, pi...
## dbl  (37): id, scrape_id, host_id, host_listings_count, host_total_listings_...
## lgl   (8): host_is_superhost, host_has_profile_pic, host_identity_verified, ...
## date  (5): last_scraped, host_since, calendar_last_scraped, first_review, la...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Warning: NAs introduced by coercion
```

```
## Rows: 71938 Columns: 75
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (24): listing_url, source, name, description, neighborhood_overview, pi...
## dbl  (37): id, scrape_id, host_id, host_listings_count, host_total_listings_...
## lgl   (9): host_is_superhost, host_has_profile_pic, host_identity_verified, ...
## date  (5): last_scraped, host_since, calendar_last_scraped, first_review, la...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Warning: NAs introduced by coercion

## Rows: 55104 Columns: 75
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (25): listing_url, source, name, description, neighborhood_overview, pi...
## dbl  (37): id, scrape_id, host_id, host_listings_count, host_total_listings_...
## lgl   (8): host_is_superhost, host_has_profile_pic, host_identity_verified, ...
## date  (5): last_scraped, host_since, calendar_last_scraped, first_review, la...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Warning: NAs introduced by coercion

## Rows: 24383 Columns: 75
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (25): listing_url, source, name, description, neighborhood_overview, pi...
## dbl  (37): id, scrape_id, host_id, host_listings_count, host_total_listings_...
## lgl   (8): host_is_superhost, host_has_profile_pic, host_identity_verified, ...
## date  (5): last_scraped, host_since, calendar_last_scraped, first_review, la...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Warning: NAs introduced by coercion
```

```
# display the price summary table
price_summary
```

```
##        city min_price max_price mean_price
## 1 Amsterdam        20       999   208.9832
## 2    London         7       999   139.4911
## 3     Paris         8       999   143.2158
## 4      Rome         9       999   122.0354
```

#dit kan weg

```r
#removing the dollar sign from the "price" column in the calendar data
city_names <- c("ams", "london", "paris", "rome")

for(city in city_names) {
  assign(paste0("cal_", city), get(paste0("cal_", city)) %>%
           mutate(price = ifelse(!is.na(price), as.numeric(gsub("\\$", "", price)), price)))
}
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `price = ifelse(!is.na(price), as.numeric(gsub("\\$", "",
##   price)), price)`.
## Caused by warning in `ifelse()`:
## ! NAs introduced by coercion
## There was 1 warning in `mutate()`.
## i In argument: `price = ifelse(!is.na(price), as.numeric(gsub("\\$", "",
##   price)), price)`.
## Caused by warning in `ifelse()`:
## ! NAs introduced by coercion
## There was 1 warning in `mutate()`.
## i In argument: `price = ifelse(!is.na(price), as.numeric(gsub("\\$", "",
##   price)), price)`.
## Caused by warning in `ifelse()`:
## ! NAs introduced by coercion
## There was 1 warning in `mutate()`.
## i In argument: `price = ifelse(!is.na(price), as.numeric(gsub("\\$", "",
##   price)), price)`.
## Caused by warning in `ifelse()`:
## ! NAs introduced by coercion
```

Summary statistics for beds per city

```
## Summary statistics for ams
## Beds
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   1.000   1.873   2.000  33.000      91
## Summary statistics for london
## Beds
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   1.000   1.768   2.000  45.000    1180
## Summary statistics for paris
## Beds
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   1.000   1.713   2.000  90.000     769
## Summary statistics for rome
## Beds
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.00    1.00    2.00    2.24    3.00   24.00     339
```

Subset of the listings dataset for all cities

Subset of the calendar dataset for all cities

Output

Number of available rooms per city between the 5th of December 2022 and the 4th of December 2023

```
##
##                Amsterdam   London    Paris     Rome
##   available       575071  8090743  5704653  4923682
##   unavailable    1910214 18165561 14405032  3968817
```

Number of rooms types per city

```
##
##                    Amsterdam London Paris  Rome
##   Entire home/apt       4910  43076 46785 16542
##   Hotel room              62    221   985   712
##   Private room          1798  28258  7048  6992
##   Shared room             39    383   286   137
```

Percentage number of rooms in Amsterdam

```
##
## Entire home/apt      Hotel room    Private room     Shared room
##     0.721104421     0.009105596     0.264062271     0.005727713
```

Percentage number of rooms in London

```
##
## Entire home/apt      Hotel room    Private room     Shared room
##     0.598793405     0.003072090     0.392810476     0.005324029
```

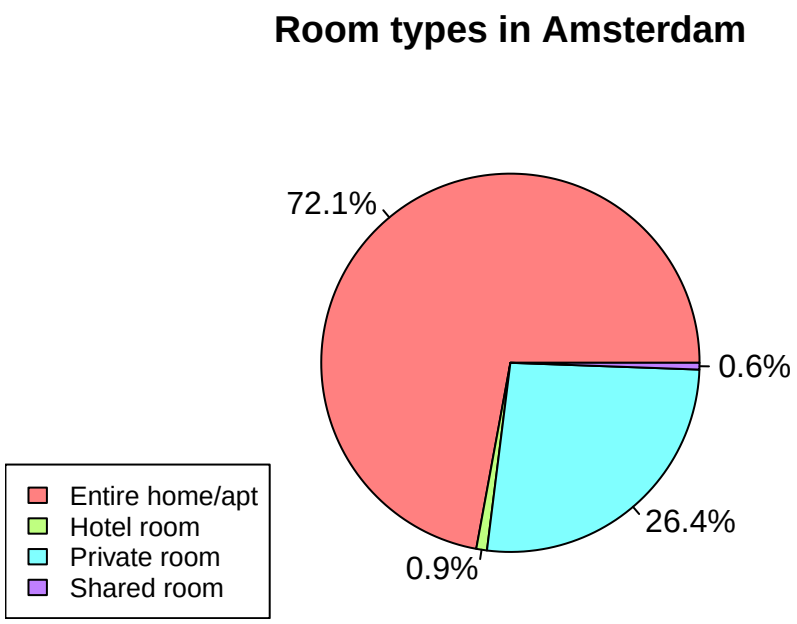
Percent number of rooms in Paris

```
##
## Entire home/apt      Hotel room    Private room     Shared room
##     0.849030923     0.017875290     0.127903600     0.005190186
```

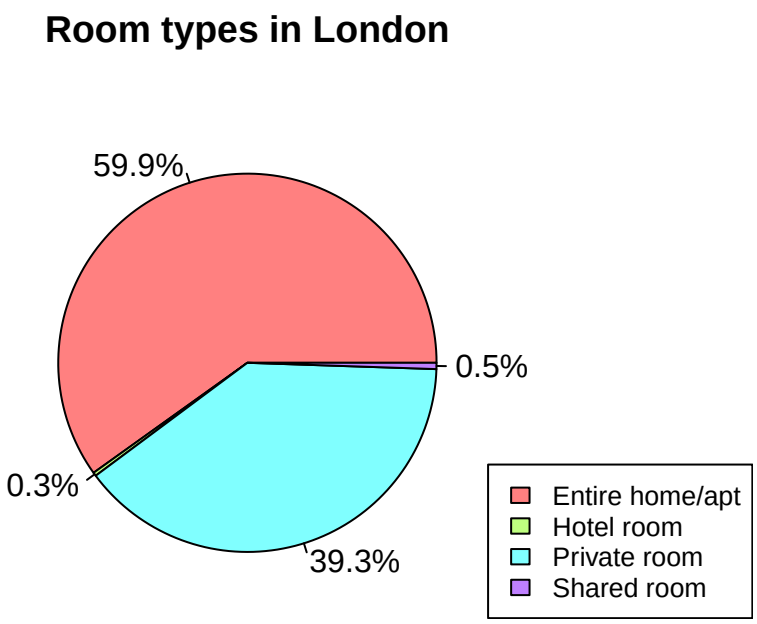
Percent number of rooms in Rome

```
##
## Entire home/apt      Hotel room    Private room     Shared room
##     0.678423492     0.029200673     0.286757167     0.005618669
```

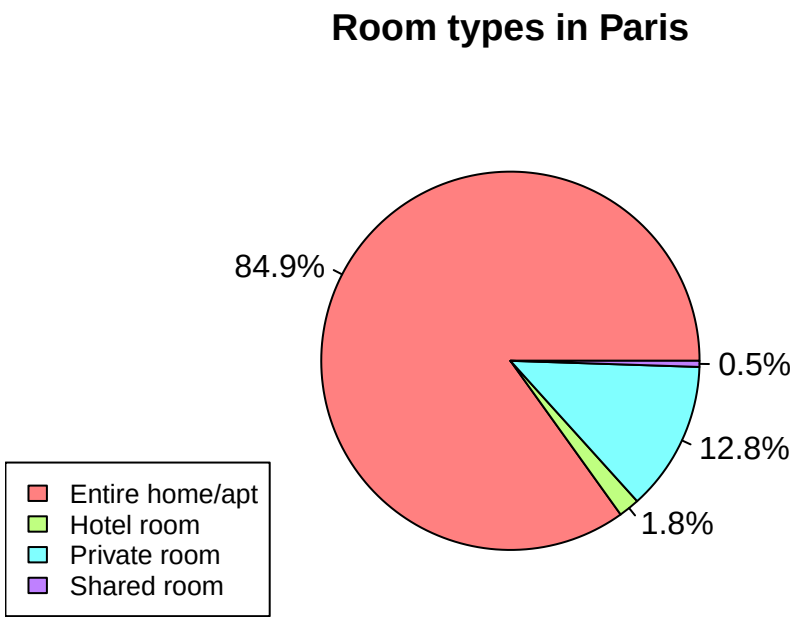
Airbnb room types offered in Amsterdam



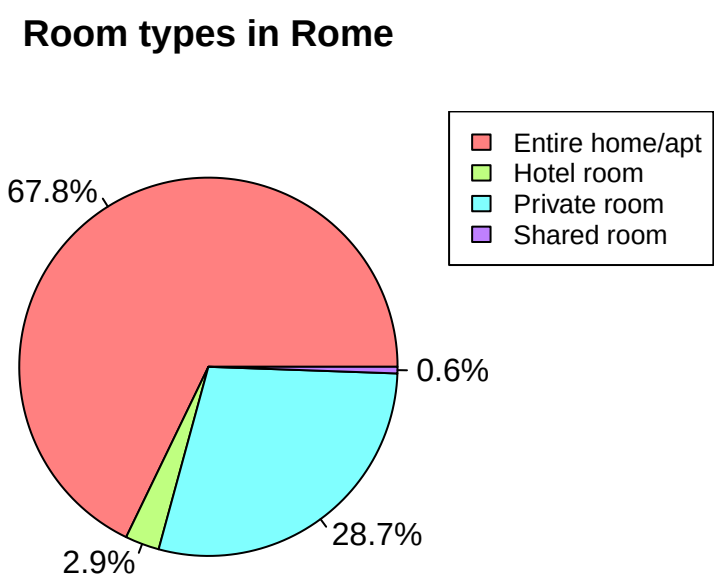
Airbnb room types offered in London



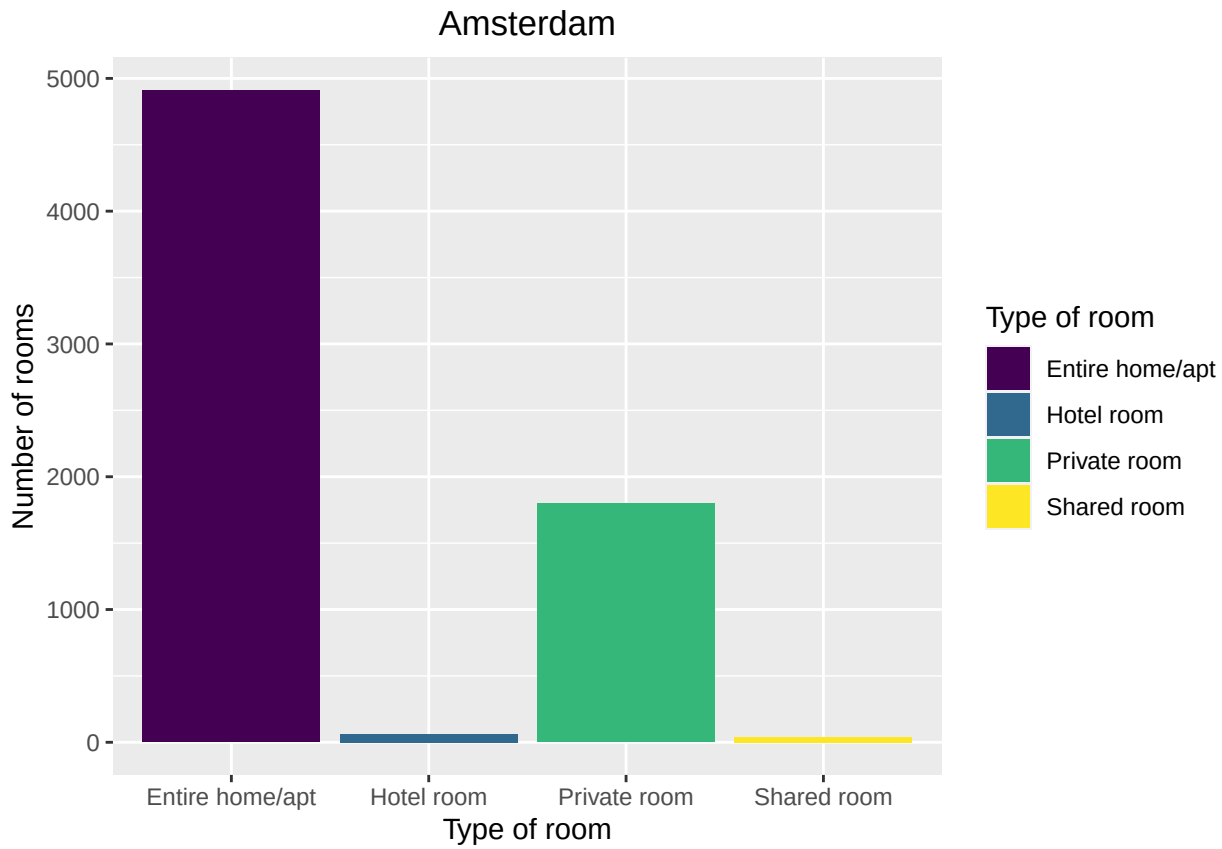
Airbnb room types offered in Paris



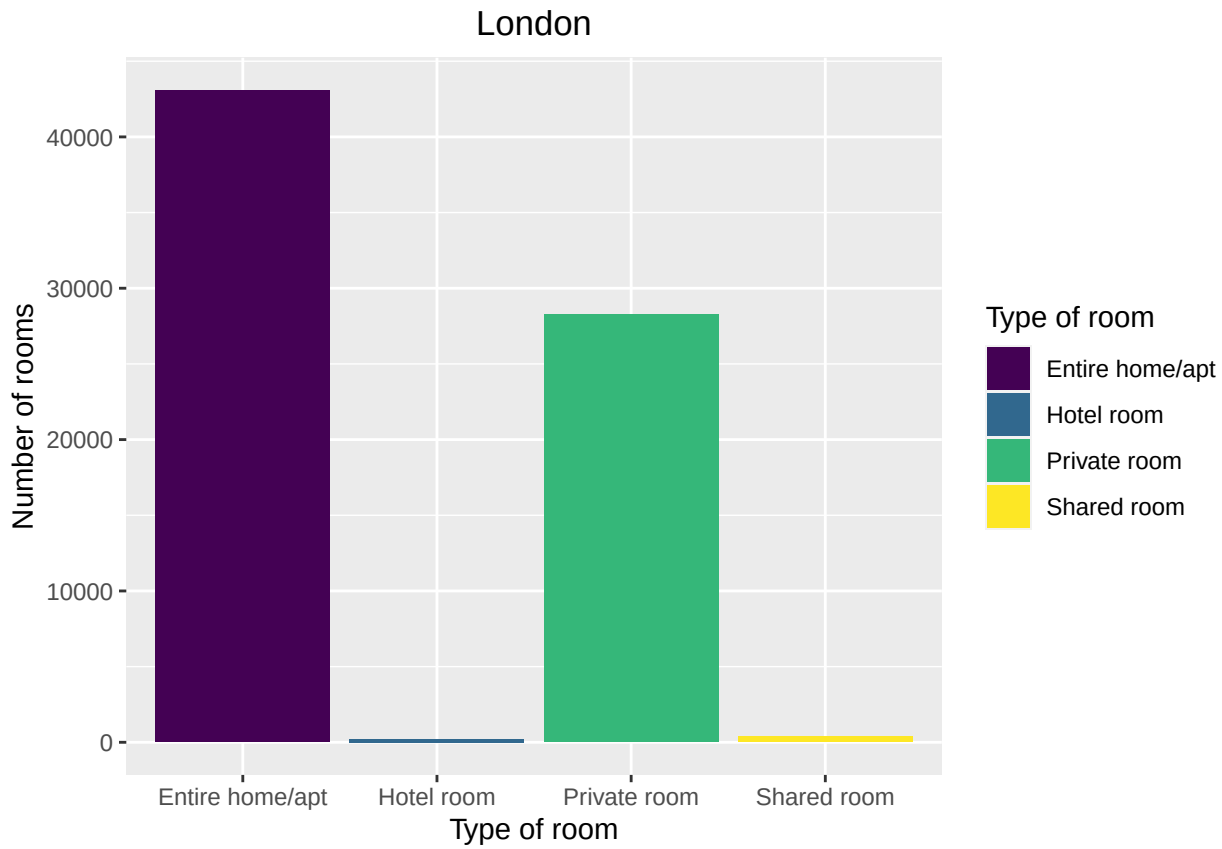
Airbnb room types offered in Rome



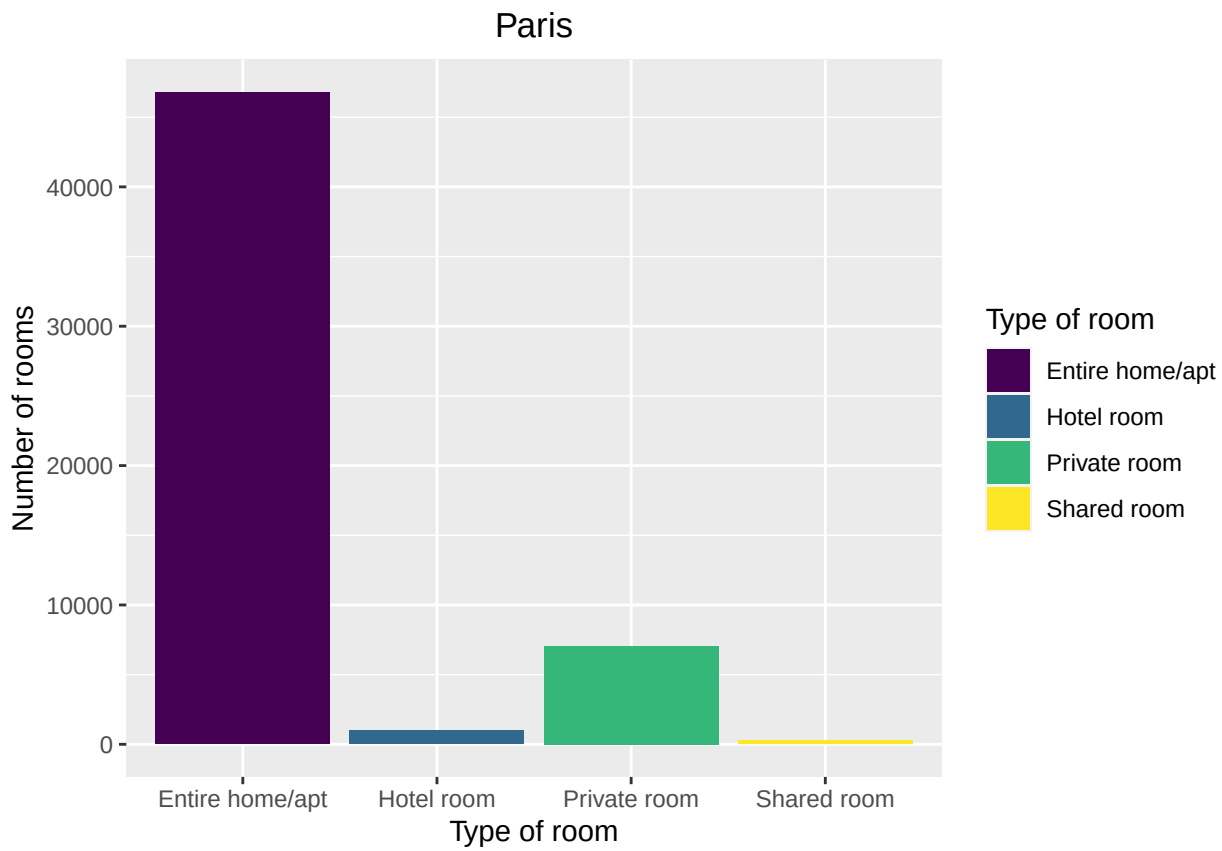
Quantity of type of rooms in Amsterdam



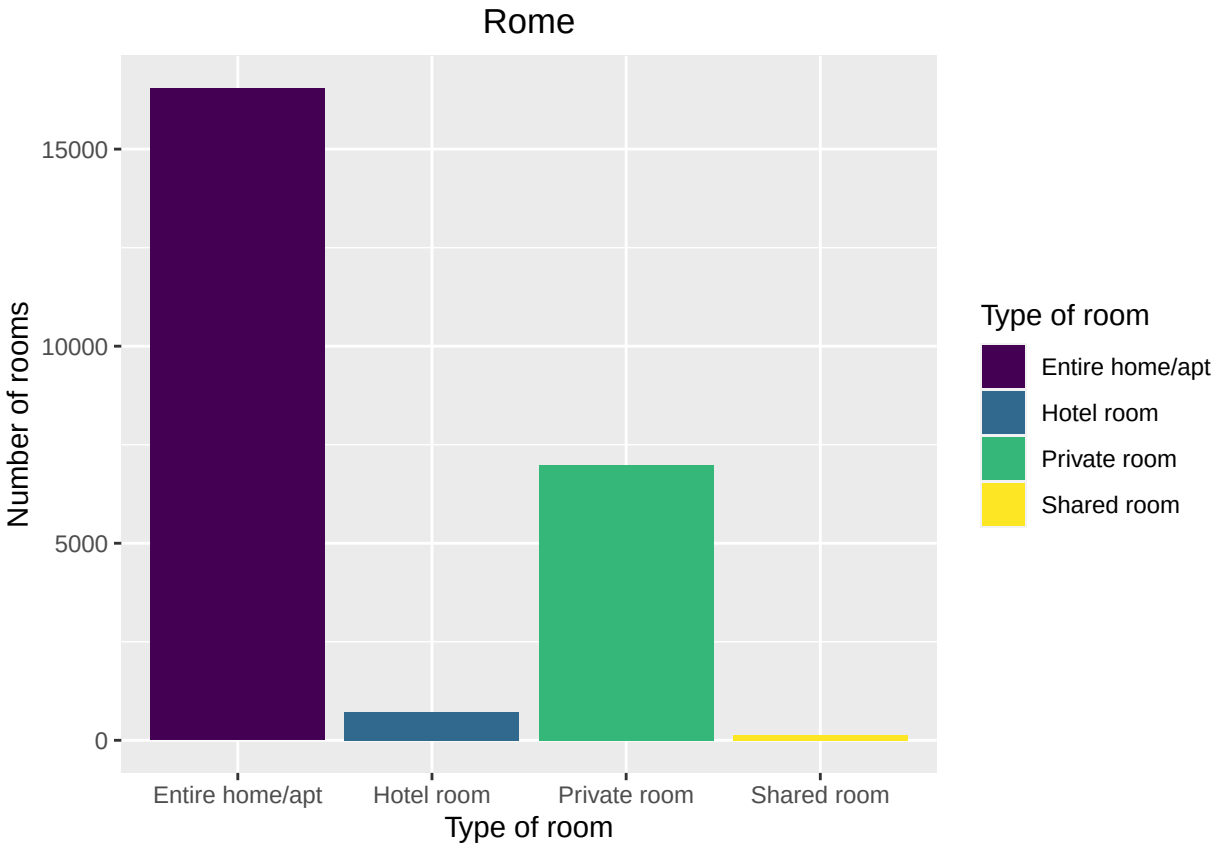
Quantity of type of rooms in London



Quantity of type of rooms in Paris

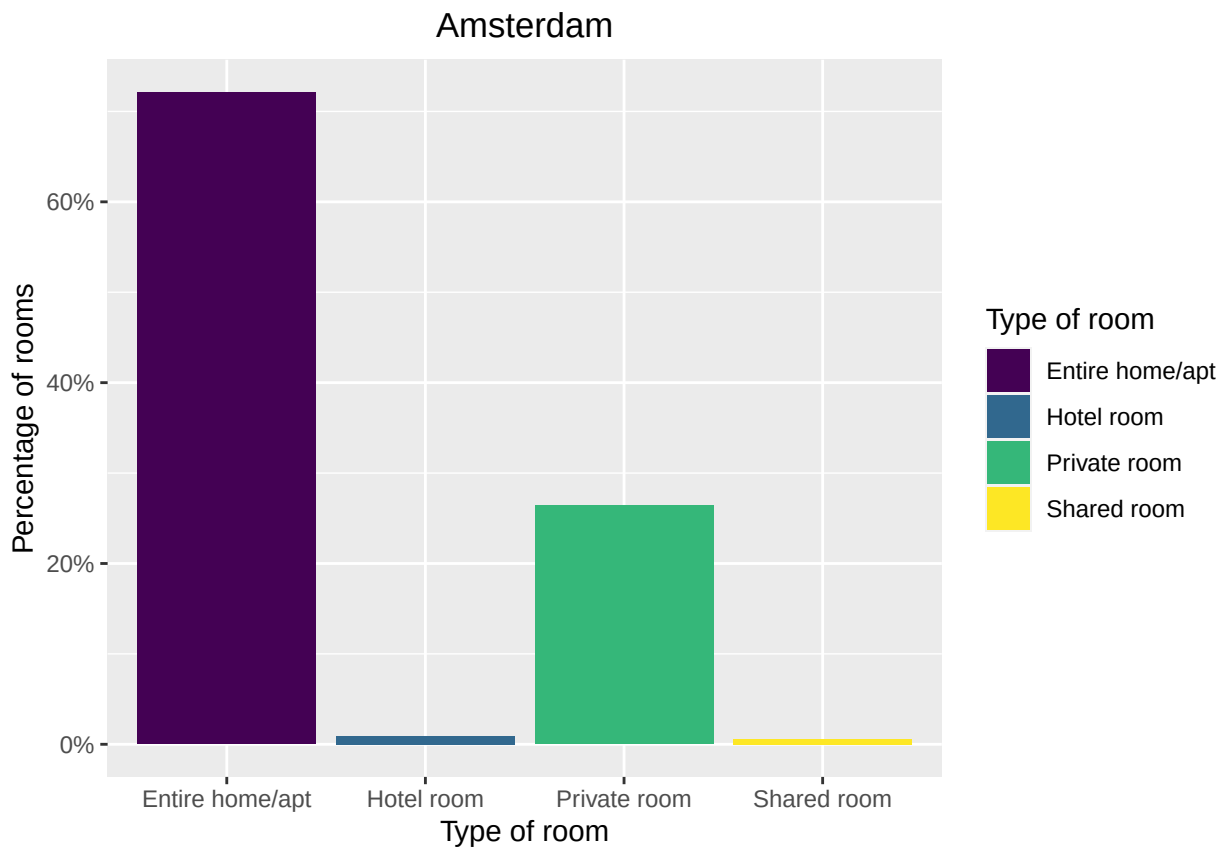


Quantity of type of rooms in Rome

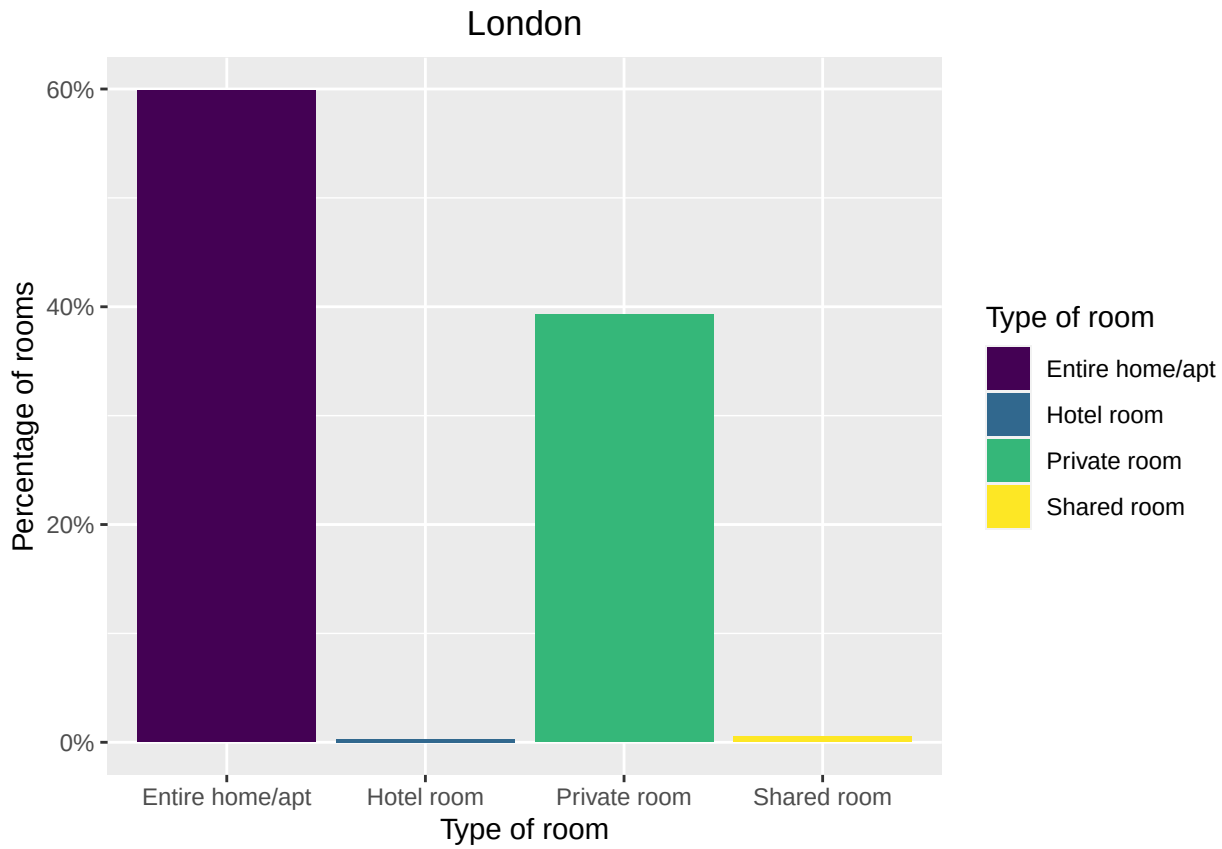


Percentage of rooms offered in Amsterdam

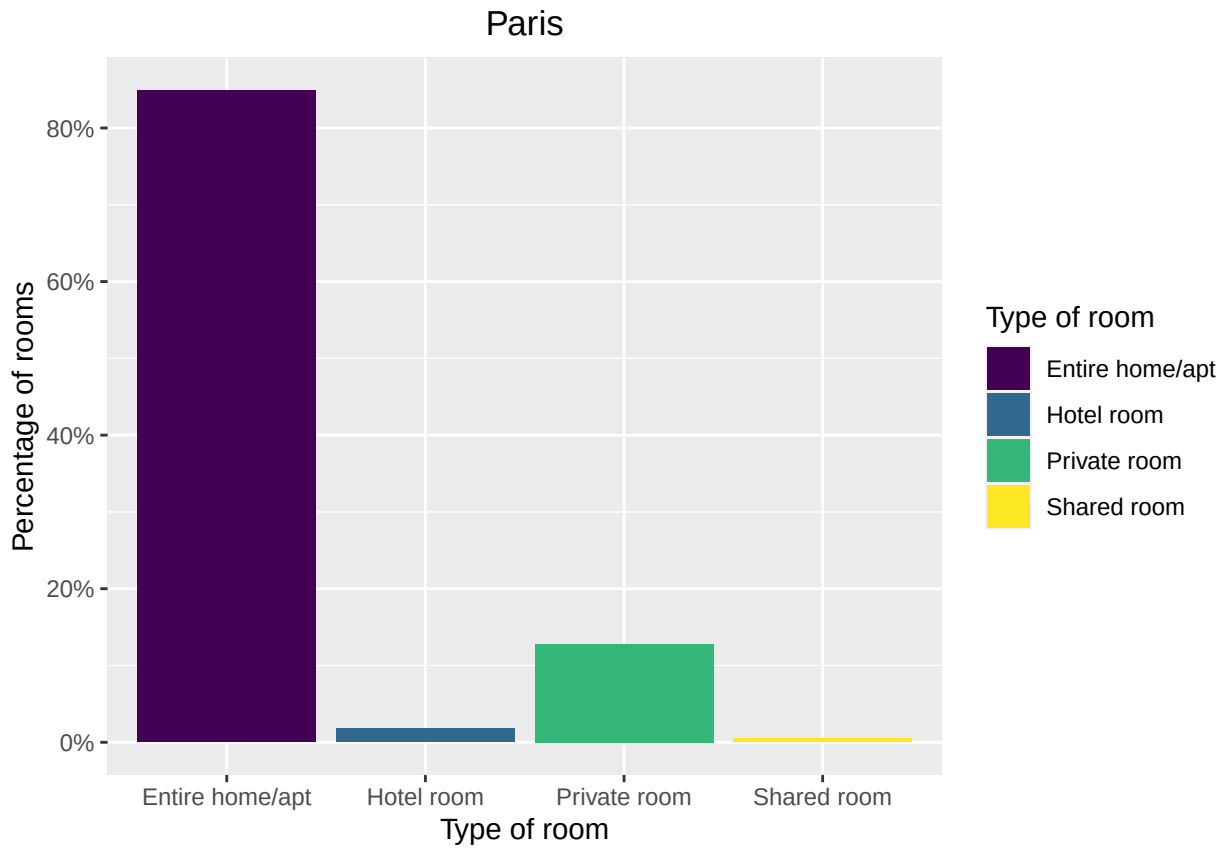
```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(count)‘ instead.
```



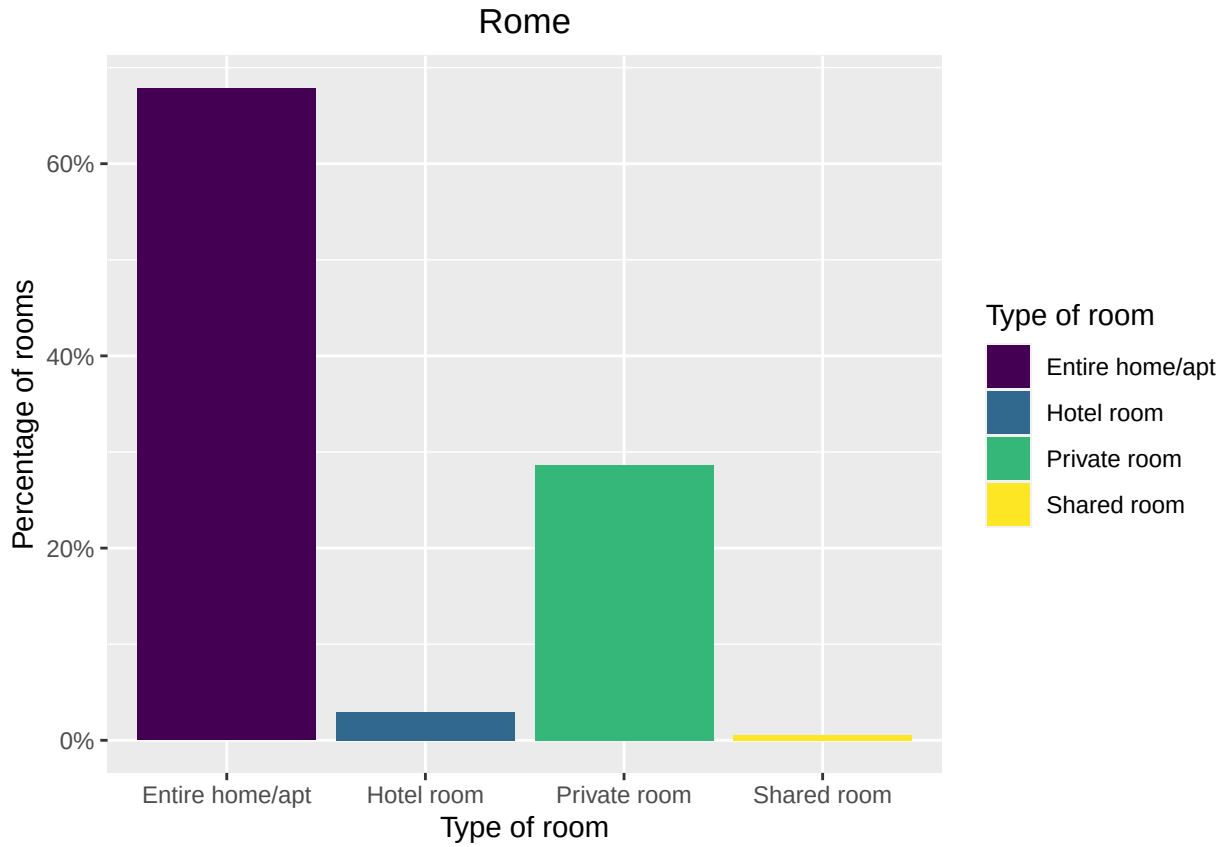
Percentage of rooms offered in London



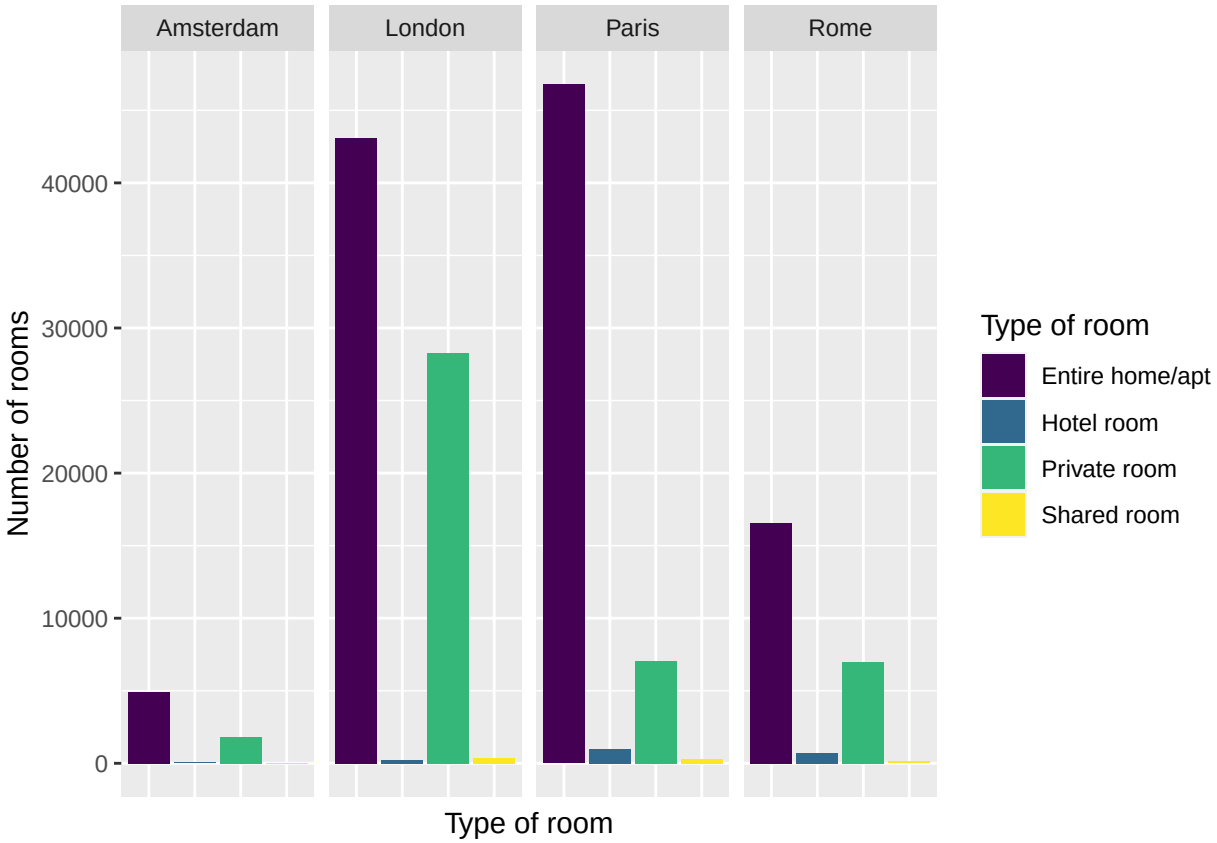
Percentage of rooms offered in Paris



##Percentage of rooms offered in Rome**



##Number and type of rooms comparison between the cities**


Amsterdam
London
Paris
Rome
40000
30000
20000
10000
0
Number of rooms
Type of room
Type of room
Entire home/apt
Hotel room
Private room
Shared room