STB Assessment

Fang Yu

2024-07-29

Question 1

```
#Setup for Question 1
setwd("D:/Projects/stb_assessment") #Set to repository directory
data = read_excel("data/mock_survey_data_3.xlsx")
```

a) Describe the approach you will take and data fields you would look into when it comes to data preparation

```


glimpse(data)
```

Preliminary Inspection

```
## Rows: 22,974
## Columns: 51
## $ case                                               <dbl> 1, 2, 3, 4, 5, 6~
## $ year                                               <dbl> 2018, 2018, 2018~
## $ month                                              <dbl> 1, 1, 1, 1, 1, 1~
## $ ‘Country of Residence‘                             <chr> "India", "India"~
## $ ‘City of Residence‘                                <chr> "Delhi", "Delhi"~
## $ ‘Purpose of Visit‘                                 <chr> "Leisure", "Leis~
## $ ‘Main Purpose of Visit‘                            <chr> "Holiday/ Rest &~
## $ Weights_QTR                                        <dbl> 660.0465, 433.21~
## $ Air_Terminal                                       <chr> "2", "2", "2", "~
## $ Sea_Terminal                                       <chr> NA, NA, NA, NA, ~
## $ Land_Terminal                                      <chr> NA, NA, NA, NA, ~
## $ ‘First Visit To Singapore‘                         <chr> "Yes", "No", "No~
## $ ‘Length of Stay‘                                   <dbl> 6, 50, 6, 7, 2, ~
## $ ‘Travel Type‘                                      <chr> "Packaged", "Non~
## $ Gender                                             <chr> "Male", "Female"~
## $ ‘Marital Status‘                                   <chr> "Married", "Sing~
## $ Occupation...17                                    <chr> "Executive (sale~
## $ Occupation...18                                    <chr> NA, NA, "InfoCom~
## $ ‘Other Designation‘                                <chr> NA, NA, "Manager~
## $ ‘Designation (free text)‘                          <chr> NA, NA, NA, NA, ~
## $ shopping_fash                                      <chr> "0", "0", "4000"~
## $ shopping_jewllery                                  <chr> "0", "0", "0", "~
## $ shopping_watches                                      <chr> "0", "0", "0", "~
## $ shopping_wellness                                     <chr> "0", "0", "0", "~
## $ shopping_food                                         <chr> "0", "0", "0", "~
## $ shopping_gifts                                        <chr> "75", "0", "0", ~
## $ shopping_consumertech                                 <chr> "0", "0", "0", "~
## $ shopping_antiques                                     <chr> "0", "0", "0", "~
## $ shopping_others                                       <chr> "0", "0", "0", "~
## $ shopping_any                                          <chr> "75", "0", "4000~
## $ totacc                                                <dbl> 880.68, 0.00, 0.~
## $ totfnb                                                <dbl> 234.79, 18.31, 9~
## $ tottran                                               <dbl> 29.35, 14.09, 24~
## $ totbiz                                                <dbl> 0, 0, 0, 0, 0, 0~
## $ totedu                                                <dbl> 0, 0, 0, 0, 0, 0~
## $ totmedi                                               <dbl> 0, 0, 0, 0, 0, 0~
## $ tototh                                                <dbl> 234.38, 40.00, 0~
## $ totshopping_rep                                       <dbl> 101.70, 26.70, 4~
## $ tot.exp                                               <dbl> 1480.90, 99.10, ~
## $ c4a_1                                                 <chr> "1", "0", "0", "~
## $ MainHotel                                             <chr> "15", NA, NA, NA~
## $ ‘Travel companion - Alone‘                            <chr> "0", "1", "1", "~
## $ ‘Travel companion - Spouse‘                           <chr> "1", "0", "0", "~
## $ ‘Travel companion - Your Child/Children‘              <chr> "0", "0", "0", "~
## $ ‘Travel companion - Parents/Parents-in-law‘           <chr> "0", "0", "0", "~
## $ ‘Travel companion - Grandparents/Grandparents-in-law‘ <chr> "0", "0", "0", "~
## $ ‘Travel companion - Siblings‘                         <chr> "0", "0", "0", "~
## $ ‘Travel companion - Other relatives‘                  <chr> "0", "0", "0", "~
## $ ‘Travel companion - Friends‘                          <chr> "0", "0", "0", "~
## $ ‘Travel companion - Business associates/Colleagues‘   <chr> "0", "0", "0", "~
## $ ‘Travel companion - Others‘                           <chr> "0", "0", "0", "~
```

```r
# Fixing column names for occupation
data <- rename(data, Occupation = `Occupation...17`, Sector = `Occupation...18`)
```

```r
# Convert to appropriate data types
shopping_columns <- grep("^shopping_", names(data), value = TRUE)
#travel_columns <- grep("^Travel", names(data), value = TRUE)

# Convert identified columns to double
data[shopping_columns] <- lapply(data[shopping_columns], as.double)
#data[travel_columns] <- lapply(data[travel_columns], as.double)
```

```r
# Checking for Missing Values
missing_values <- colSums(is.na(data))
print("Missing Values:")
```

Data Cleaning

```
## [1] "Missing Values:"
```

```
print(missing_values)
```

```
##                                                case
##                                                   2
##                                                year
##                                                   0
##                                               month
##                                                   0
##                                Country of Residence
##                                                   0
##                                   City of Residence
##                                                  12
##                                    Purpose of Visit
##                                                   0
##                               Main Purpose of Visit
##                                                   0
##                                         Weights_QTR
##                                                   0
##                                        Air_Terminal
##                                                4997
##                                        Sea_Terminal
##                                               20593
##                                       Land_Terminal
##                                               20359
##                            First Visit To Singapore
##                                                   0
##                                      Length of Stay
##                                                   0
##                                         Travel Type
##                                                   0
##                                              Gender
##                                                   0
##                                      Marital Status
##                                                   0
##                                          Occupation
##                                                   0
##                                              Sector
##                                               20184
##                                   Other Designation
##                                               20185
##                             Designation (free text)
##                                               22782
##                                       shopping_fash
##                                                  85
##                                   shopping_jewllery
##                                                   0
##                                    shopping_watches
##                                                   0
##                                   shopping_wellness
##                                                   0
##                                       shopping_food
##                                                   6
##                                      shopping_gifts
##                                                   2
##                               shopping_consumertech
##                                                   0
##                                   shopping_antiques
##                                                   0
##                                     shopping_others
##                                                   3
##                                        shopping_any
##                                                  96
##                                              totacc
##                                                   0
##                                              totfnb
##                                                   0
##                                             tottran
##                                                   0
##                                              totbiz
##                                                   0
##                                              totedu
##                                                   0
##                                             totmedi
##                                                   0
##                                              tototh
##                                                   0
##                                     totshopping_rep
##                                                   0
##                                             tot.exp
##                                                   0
##                                               c4a_1
##                                                   0
##                                           MainHotel
##                                                6690
##                            Travel companion - Alone
##                                                   0
##                           Travel companion - Spouse
##                                                   0
##              Travel companion - Your Child/Children
##                                                   0
##           Travel companion - Parents/Parents-in-law
##                                                   0
## Travel companion - Grandparents/Grandparents-in-law
##                                                   0
##                         Travel companion - Siblings
##                                                   0
##                  Travel companion - Other relatives
##                                                   0
##                          Travel companion - Friends
##                                                   0
##   Travel companion - Business associates/Colleagues
##                                                   0
##                           Travel companion - Others
##                                                   0
```

```r
# Checking for Duplicates
duplicates <- sum(duplicated(data))
```

```
print("Duplicates:")
```

```
## [1] "Duplicates:"
```

```
print(duplicates)
```

```
## [1] 0
```

```
# Checking if case is a unique identifier
data %>% count(case) %>% arrange(desc(n))
```

```
## # A tibble: 22,973 x 2
##     case     n
##    <dbl> <int>
##  1    NA     2
##  2     1     1
##  3     2     1
##  4     3     1
##  5     4     1
##  6     5     1
##  7     6     1
##  8     7     1
##  9     8     1
## 10     9     1
## # i 22,963 more rows
```

```
# Give NA cases unique identifiers
max_case <- max(data$case, na.rm = TRUE)

# Generate unique values for NA entries
data <- data %>%
  mutate(case = ifelse(is.na(case), max_case + row_number(), case))
```

```
# Fixing shopping_xxx values
data = data %>%
  mutate(shopping_fash = ifelse(is.na(shopping_fash), 0, shopping_fash),
         shopping_food = ifelse(is.na(shopping_food), 0, shopping_food),
         shopping_gifts = ifelse(is.na(shopping_gifts), 0, shopping_gifts),
         shopping_others = ifelse(is.na(shopping_others), 0, shopping_others),
         shopping_any = shopping_fash + shopping_jewllery + shopping_watches
         + shopping_wellness + shopping_gifts + shopping_consumertech
         + shopping_antiques + shopping_others)
```

```
# Check for missing values across all three terminal
missing_rows <- data %>%
  filter(is.na(Air_Terminal) & is.na(Sea_Terminal) & is.na(Land_Terminal))

# Display the rows with all three features missing
print(missing_rows)
```

```
## # A tibble: 3 x 51
##    case  year month ‘Country of Residence‘ ‘City of Residence‘
##   <dbl> <dbl> <dbl> <chr>                  <chr>
## 1   699  2018     1 Vietnam                Hanoi
## 2   811  2018     1 Thailand               Chiang Mai
## 3   824  2018     1 Italy                  Others
## # i 46 more variables: ‘Purpose of Visit‘ <chr>, ‘Main Purpose of Visit‘ <chr>,
## #   Weights_QTR <dbl>, Air_Terminal <chr>, Sea_Terminal <chr>,
## #   Land_Terminal <chr>, ‘First Visit To Singapore‘ <chr>,
## #   ‘Length of Stay‘ <dbl>, ‘Travel Type‘ <chr>, Gender <chr>,
## #   ‘Marital Status‘ <chr>, Occupation <chr>, Sector <chr>,
## #   ‘Other Designation‘ <chr>, ‘Designation (free text)‘ <chr>,
## #   shopping_fash <dbl>, shopping_jewllery <dbl>, shopping_watches <dbl>, ...
```

```
# Find most frequent terminal for tourists from Vietnam, Thailand and Italy
data %>% filter(`Country of Residence` == "Vietnam") %>% count(Land_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 4 x 2
##   Land_Terminal              n
##   <chr>                  <int>
## 1 <NA>                     583
## 2 Tuas Checkpoint           35
## 3 Woodlands Checkpoint      34
## 4 Others, please specify    19
```

```
data %>% filter(`Country of Residence` == "Vietnam") %>% count(Sea_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 5 x 2
##   Sea_Terminal            n
##   <chr>               <int>
## 1 <NA>                  655
## 2 Harbourfront Ferry      8
## 3 Tanah Merah Ferry       4
## 4 MBCCS                   3
## 5 Harbourfront Cruise     1
```

```
data %>% filter(`Country of Residence` == "Vietnam") %>% count(Air_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 4 x 2
##   Air_Terminal     n
##   <chr>        <int>
## 1 2              214
## 2 1              207
## 3 3              145
## 4 <NA>           105
```

```
data %>% filter(`Country of Residence` == "Thailand") %>% count(Land_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 4 x 2
##   Land_Terminal              n
##   <chr>                  <int>
## 1 <NA>                     726
## 2 Woodlands Checkpoint      37
## 3 Others, please specify    27
## 4 Tuas Checkpoint           27
```

```
data %>% filter(`Country of Residence` == "Thailand") %>% count(Sea_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 5 x 2
##   Sea_Terminal            n
##   <chr>               <int>
## 1 <NA>                  803
## 2 Harbourfront Ferry      5
## 3 MBCCS                   4
## 4 Tanah Merah Ferry       3
## 5 Harbourfront Cruise     2
```

```
data %>% filter(`Country of Residence` == "Thailand") %>% count(Air_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 4 x 2
##   Air_Terminal     n
##   <chr>        <int>
## 1 1              367
## 2 2              198
## 3 3              146
## 4 <NA>           106
```

```
data %>% filter(`Country of Residence` == "Italy") %>% count(Land_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 3 x 2
##   Land_Terminal            n
##   <chr>                <int>
## 1 <NA>                   124
## 2 Woodlands Checkpoint     7
## 3 Tuas Checkpoint          3
```

```
data %>% filter(`Country of Residence` == "Italy") %>% count(Sea_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 5 x 2
##   Sea_Terminal            n
##   <chr>               <int>
## 1 <NA>                  127
## 2 Tanah Merah Ferry       3
## 3 Harbourfront Ferry      2
## 4 Harbourfront Cruise     1
## 5 MBCCS                   1
```

```
data %>% filter(`Country of Residence` == "Italy") %>% count(Air_Terminal) %>% arrange(desc(n))
```

```
## # A tibble: 4 x 2
##   Air_Terminal     n
##   <chr>        <int>
## 1 1               55
## 2 3               52
## 3 <NA>            18
## 4 2                9
```

```
# Filling in Missing Values for Terminal
data <- data %>%
  mutate(
    Air_Terminal = ifelse(case == 699, 2, Air_Terminal),
    Air_Terminal = ifelse(case == 811, 3, Air_Terminal),
    Air_Terminal = ifelse(case == 824, 1, Air_Terminal)
  )
```

```
# Filling in Missing Values for Main Hotel
max_hotel <- max(data$MainHotel, na.rm = TRUE)

# Generate unique values for NA entries
data <- data %>%
  mutate(MainHotel = ifelse(is.na(MainHotel), as.character(as.double(max_hotel) + 1), MainHotel),
         MainHotel = ifelse(MainHotel %in% c(991,992,993,994,999,9996,9999),as.character(as.double(max_
```

```
# Viewing Summary Statistics
summary(data)
```

Exploratory Data Analysis

```
##       case           year          month        Country of Residence
##  Min.   :    1   Min.   :2018   Min.   : 1.000   Length:22974
##  1st Qu.: 5755   1st Qu.:2018   1st Qu.: 4.000   Class :character
##  Median :11498   Median :2018   Median : 7.000   Mode  :character
##  Mean   :11498   Mean   :2018   Mean   : 6.497
##  3rd Qu.:17242   3rd Qu.:2018   3rd Qu.: 9.000
##  Max.   :24707   Max.   :2018   Max.   :12.000
##  City of Residence  Purpose of Visit   Main Purpose of Visit  Weights_QTR
##  Length:22974       Length:22974       Length:22974          Min.   :   31.0
##  Class :character   Class :character   Class :character      1st Qu.:  418.3
##  Mode  :character   Mode  :character   Mode  :character      Median :  516.8
##                                                              Mean   :  634.1
##                                                              3rd Qu.:  692.6
##                                                              Max.   :14673.0
##  Air_Terminal       Sea_Terminal       Land_Terminal
##  Length:22974       Length:22974       Length:22974
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##  First Visit To Singapore Length of Stay   Travel Type
##  Length:22974             Min.   : 0.000   Length:22974
##  Class :character         1st Qu.: 2.000   Class :character
##  Mode  :character         Median : 3.000   Mode  :character
##                           Mean   : 3.419
##                           3rd Qu.: 4.000
##                           Max.   :60.000
##     Gender          Marital Status      Occupation           Sector
##  Length:22974       Length:22974       Length:22974       Length:22974
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Other Designation  Designation (free text) shopping_fash    shopping_jewllery
##  Length:22974       Length:22974           Min.   :   0.0   Min.   :   0
##  Class :character   Class :character       1st Qu.:   0.0   1st Qu.:   0
##  Mode  :character   Mode  :character       Median :   0.0   Median :   0
##                                             Mean   : 106.7   Mean   :  13
##                                             3rd Qu.: 125.0   3rd Qu.:   0
##                                             Max.   :6000.0   Max.   :8000
##  shopping_watches   shopping_wellness shopping_food     shopping_gifts
##  Min.   :   0.000   Min.   :   0.00   Min.   :   0.00   Min.   :   0.00
##  1st Qu.:   0.000   1st Qu.:   0.00   1st Qu.:   0.00   1st Qu.:   0.00
##  Median :   0.000   Median :   0.00   Median :   0.00   Median :   0.00
##  Mean   :   9.918   Mean   :  44.89   Mean   :  22.58   Mean   :  28.84
##  3rd Qu.:   0.000   3rd Qu.:  50.00   3rd Qu.:  30.00   3rd Qu.:  50.00
##  Max.   :5244.500   Max.   :2000.00   Max.   :1000.00   Max.   :1500.00
##  shopping_consumertech shopping_antiques  shopping_others    shopping_any
##  Min.   :   0.00       Min.   :   0.000   Min.   :   0.00   Min.   :   0.0
##  1st Qu.:   0.00       1st Qu.:   0.000   1st Qu.:   0.00   1st Qu.:  20.0
##  Median :   0.00       Median :   0.000   Median :   0.00   Median : 125.0
##  Mean   :  13.69       Mean   :   3.155   Mean   :  15.69   Mean   : 235.9
##  3rd Qu.:   0.00       3rd Qu.:   0.000   3rd Qu.:   0.00   3rd Qu.: 300.0
##  Max.   :4500.00       Max.   :3500.000   Max.   :3000.00   Max.   :8200.0
##      totacc            totfnb            tottran            totbiz
##  Min.   :    0.0   Min.   :    0.00   Min.   :   0.00   Min.   :   0.000
##  1st Qu.:    0.0   1st Qu.:   38.69   1st Qu.:  20.43   1st Qu.:   0.000
##  Median :  252.3   Median :  109.02   Median :  42.45   Median :   0.000
##  Mean   :  356.1   Mean   :  155.70   Mean   :  62.11   Mean   :   5.513
##  3rd Qu.:  506.8   3rd Qu.:  209.34   3rd Qu.:  78.59   3rd Qu.:   0.000
##  Max.   :27751.5   Max.   :22999.16   Max.   :1308.09   Max.   :4000.000
##      totedu             totmedi             tototh         totshopping_rep
##  Min.   :    0.000   Min.   :   0.000   Min.   :-6295.90   Min.   :   1.62
##  1st Qu.:    0.000   1st Qu.:   0.000   1st Qu.:    0.00   1st Qu.: 111.92
##  Median :    0.000   Median :   0.000   Median :   50.00   Median : 228.54
##  Mean   :    0.954   Mean   :   2.559   Mean   :   15.45   Mean   : 331.85
##  3rd Qu.:    0.000   3rd Qu.:   0.000   3rd Qu.:  150.00   3rd Qu.: 417.48
##  Max.   :10000.000   Max.   :3000.000   Max.   :15530.00   Max.   :5148.68
##     tot.exp           c4a_1            MainHotel
##  Min.   :    0.0   Length:22974       Length:22974
##  1st Qu.:  355.9   Class :character   Class :character
##  Median :  798.1   Mode  :character   Mode  :character
##  Mean   :  923.1
##  3rd Qu.: 1280.2
##  Max.   :29206.4
##  Travel companion - Alone Travel companion - Spouse
##  Length:22974             Length:22974
##  Class :character         Class :character
##  Mode  :character         Mode  :character
##
##
##
##  Travel companion - Your Child/Children
##  Length:22974
##  Class :character
##  Mode  :character
##
##
##
##  Travel companion - Parents/Parents-in-law
##  Length:22974
##  Class :character
##  Mode  :character
##
##
##
##  Travel companion - Grandparents/Grandparents-in-law
##  Length:22974
##  Class :character
##  Mode  :character
##
##
##
##  Travel companion - Siblings Travel companion - Other relatives
##  Length:22974                Length:22974
##  Class :character            Class :character
##  Mode  :character            Mode  :character
##
##
##
##  Travel companion - Friends Travel companion - Business associates/Colleagues
##  Length:22974               Length:22974
##  Class :character           Class :character
##  Mode  :character           Mode  :character
##
##
##
##  Travel companion - Others
##  Length:22974
##  Class :character
##  Mode  :character
##
##
##
```

b) Highlight the data idiosyncrasies / issues you found in this dataset and how would you deal with it.

```r
# Setting `tototh` column values to zero is negative
data <- data %>%
  mutate(tototh = ifelse(tototh < 0, 0, tototh))

# Correcting total expenditure column
data <- data %>%
  mutate(tot.exp = totacc + totfnb + tottran + totbiz + totedu + totmedi + tototh + totshopping_rep)
```

Negative Values for Expenditure

```r
# Main Purpose of Visit does not Match Purpose of Visit
data %>% filter(`Purpose of Visit` == "Healthcare + Accompanying Pax") %>% dplyr::select("Main Purpose
```

Main Purpose of Visit does not match Purpose of Visit

```
## # A tibble: 20 x 1
##    `Main Purpose of Visit`
##    <chr>
##  1 Holiday/ Rest & Relax
##  2 General business purpose
##  3 Stopover (a planned stop of at least one night)
##  4 Outpatient consultation/ treatment (e.g. with General
##  5 Visiting friends/ relatives (who are not international
##  6 Accompanying a healthcare/ medical visitor for Outpatient
##  7 Others - Personal (e.g. weddings, funerals, etc)
##  8 In-patient (hospitalization) treatment
##  9 Corporate/ business meetings   (a. Venue of corporate/
## 10 Accompanying  a healthcare/ medical visitor for In-patient
## 11 Convention/ conference
## 12 To shop/ attend shopping events in Singapore Event name
## 13 Study mission (including government study trips)
## 14 Exhibition/ trade show
## 15 Day-surgery (with no hospitalization)
## 16 Accompanying a healthcare/ medical visitor for Day-surgery
## 17 Accompanying a business visitor
## 18 Sightseeing/ Attractions
## 19 Others, please specify
## 20 Accompanying an international student (relative or friend)
```

```r
# Obtain all distinct purpose of visits into a vector
specific_purposes = data %>% dplyr::select(`Main Purpose of Visit`) %>% distinct() %>% pull(`Main Purpo
general_purposes = data %>% dplyr::select(`Purpose of Visit`) %>% distinct() %>% pull(`Purpose of Visit

print(general_purposes)
```

```
## [1] "Leisure"                       "Business + Accompanying Pax"
## [3] "Education + Accompanying Pax"  "Others/Refused"
## [5] "Healthcare + Accompanying Pax"
```

```r
print(specific_purposes)
```

```
##  [1] "Holiday/ Rest & Relax"
##  [2] "Visiting friends/ relatives (who are not international"
##  [3] "General business purpose"
##  [4] "Visiting an international student (relative or friend)"
##  [5] "Accompanying an international student (relative or friend)"
##  [6] "To take or join a regional or international cruise"
##  [7] "Others - Personal (e.g. weddings, funerals, etc)"
##  [8] "Stopover (a planned stop of at least one night)"
##  [9] "Cultural Festivals (e.g. Lunar New Year, Deepavali, Hari"
## [10] "Accompanying a business visitor"
## [11] "Sightseeing/ Attractions"
## [12] "Outpatient consultation/ treatment (e.g. with General"
## [13] "Corporate/ business meetings   (a. Venue of corporate/"
## [14] "Convention/ conference"
## [15] "School trips"
## [16] "Accompanying a healthcare/ medical visitor for Outpatient"
## [17] "Company sponsored holiday/incentive programme (a. Venue of"
## [18] "To shop/ attend shopping events in Singapore Event name"
## [19] "Study mission (including government study trips)"
## [20] "In-patient (hospitalization) treatment"
## [21] "Gather information/ facts on the education services in"
## [22] "Music-related (e.g. concerts, festivals, etc) Event name"
## [23] "Partying at clubs/ pubs/ dance parties (e.g. Avalon &"
## [24] "Others - Work Related (e.g. visa, insurance, etc)"
## [25] "Skills development/ skills training/ vocational training"
## [26] "To visit the integrated resorts(e.g. Marina Bay Sands,"
## [27] "In-house company training (Name of company)"
## [28] "Student enrichment programmes (e.g. language courses, camps,"
## [29] "Executive training, including training workshops and"
## [30] "To experience the food/ food events in Singapore Event name"
## [31] "Others (e.g. leisure exhibitions, cinema, recreational"
## [32] "Accompanying  a healthcare/ medical visitor for In-patient"
## [33] "Personal Enrichment (cooking, batik painting class, etc)"
## [34] "Student events (e.g. sports & other competitions, music"
## [35] "Sporting (e.g. rugby matches, marathons, golf tournaments,"
## [36] "Performing Arts (e.g. musicals, theatre, dance, opera, etc)"
## [37] "Exhibition/ trade show"
## [38] "To experience different cultures"
## [39] "Using Singapore as a hub to go to other destinations:"
## [40] "Pls specify"
## [41] "Job Opportunities"
## [42] "To experience the nightlife in Singapore (e.g. all late"
## [43] "Day-surgery (with no hospitalization)"
## [44] "Accompanying a healthcare/ medical visitor for Day-surgery"
## [45] "Using Singapore as a hub to go to other destinations: Stop"
## [46] "Family Entertainment (e.g. kid’s music & theatre, ice/ magic"
## [47] "Gather information/ facts on the healthcare services in"
## [48] "Others, please specify"
```

```r
# Fix the correct Purpose of Visit based on the Main Purpose of Visit
# New Purposes
purpose_map <- list(
  Leisure = c(
    "Holiday/ Rest & Relax", "Visiting friends/ relatives (who are not international", "To take or join
    "Stopover (a planned stop of at least one night)", "Cultural Festivals (e.g. Lunar New Year, Deepava
    "Sightseeing/ Attractions", "Company sponsored holiday/incentive programme (a. Venue of", "To shop/
    "Music-related (e.g. concerts, festivals, etc) Event name", "Partying at clubs/ pubs/ dance parties
    "To visit the integrated resorts(e.g. Marina Bay Sands,", "To experience the food/ food events in Si
    "Others (e.g. leisure exhibitions, cinema, recreational", "Personal Enrichment (cooking, batik pain
    "Sporting (e.g. rugby matches, marathons, golf tournaments,", "Performing Arts (e.g. musicals, thea
    "To experience different cultures", "To experience the nightlife in Singapore (e.g. all late", "Fam
  ),
  `Business + Accompanying Pax` = c(
    "General business purpose", "Accompanying a business visitor", "Corporate/ business meetings   (a.
    "Convention/ conference", "Others - Work Related (e.g. visa, insurance, etc)", "Exhibition/ trade s
    "Job Opportunities"
  ),
  `Education + Accompanying Pax` = c(
    "Visiting an international student (relative or friend)", "Accompanying an international student (r
    "School trips", "Study mission (including government study trips)", "Gather information/ facts on t
    "Skills development/ skills training/ vocational training", "In-house company training (Name of com
    "Student enrichment programmes (e.g. language courses, camps,", "Executive training, including trai
    "Student events (e.g. sports & other competitions, music"
  ),
  `Healthcare + Accompanying Pax` = c(
    "Outpatient consultation/ treatment (e.g. with General", "In-patient (hospitalization) treatment",
    "Accompanying  a healthcare/ medical visitor for In-patient", "Day-surgery (with no hospitalization
    "Gather information/ facts on the healthcare services in"
  ),
  `Others/Refused` = c(
    "Others - Personal (e.g. weddings, funerals, etc)", "Pls specify", "Others, please specify", "Using
  )
)

# Create a reverse lookup list
reverse_map <- unlist(purpose_map)
names(reverse_map) <- rep(names(purpose_map), sapply(purpose_map, length))

# Create the new column with the general purpose
data <- data %>%
  mutate(`Purpose of Visit` = names(reverse_map[match(`Main Purpose of Visit`, reverse_map)]))
```

```r
# Check for Incosistencies for Travel Companionship
data %>%
  dplyr::select(starts_with("Travel companion")) %>%
  mutate_all(as.double) %>%
  rowSums() %>% table()
```

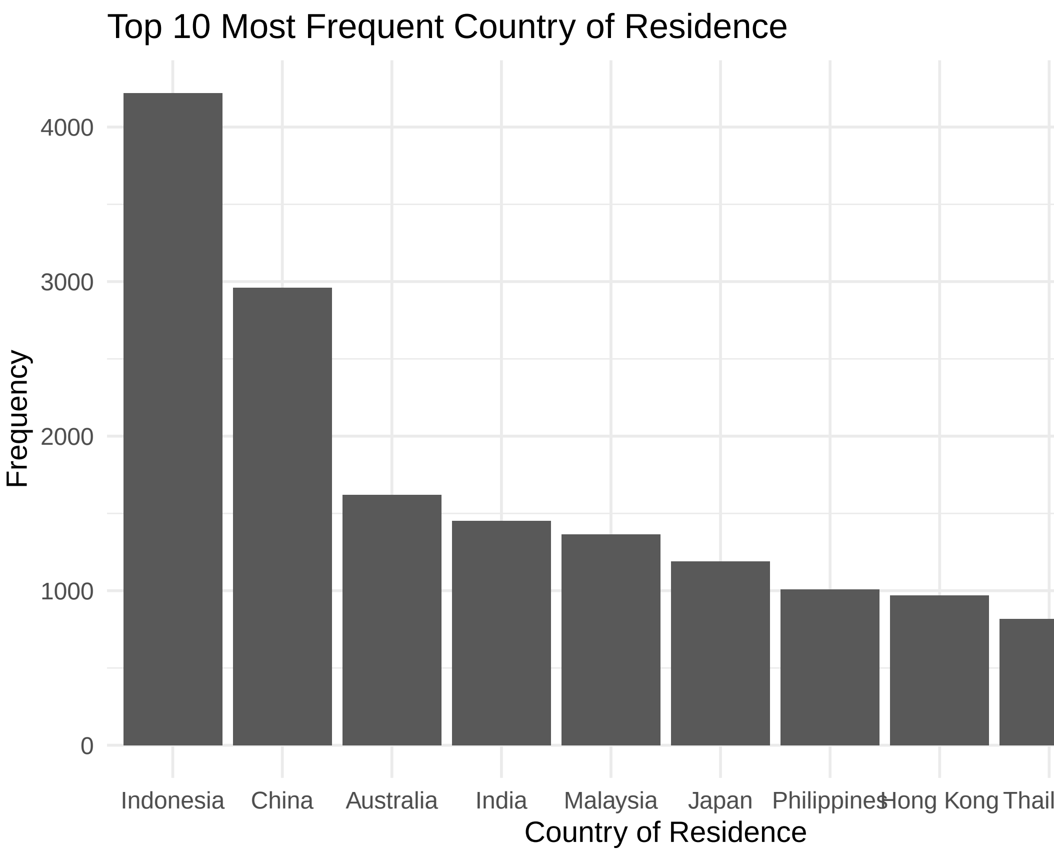
Illogical Data Entries for Travel Companions

```
## .
##     0     1     2     3     4     5
##  1326 17946  3316   345    34     7
```

```r
data %>%
  filter(`Travel companion - Alone`  == "1") %>% dplyr::select(starts_with("Travel companion")) %>%
  mutate_all(as.double) %>%
  rowSums() %>% table()
```

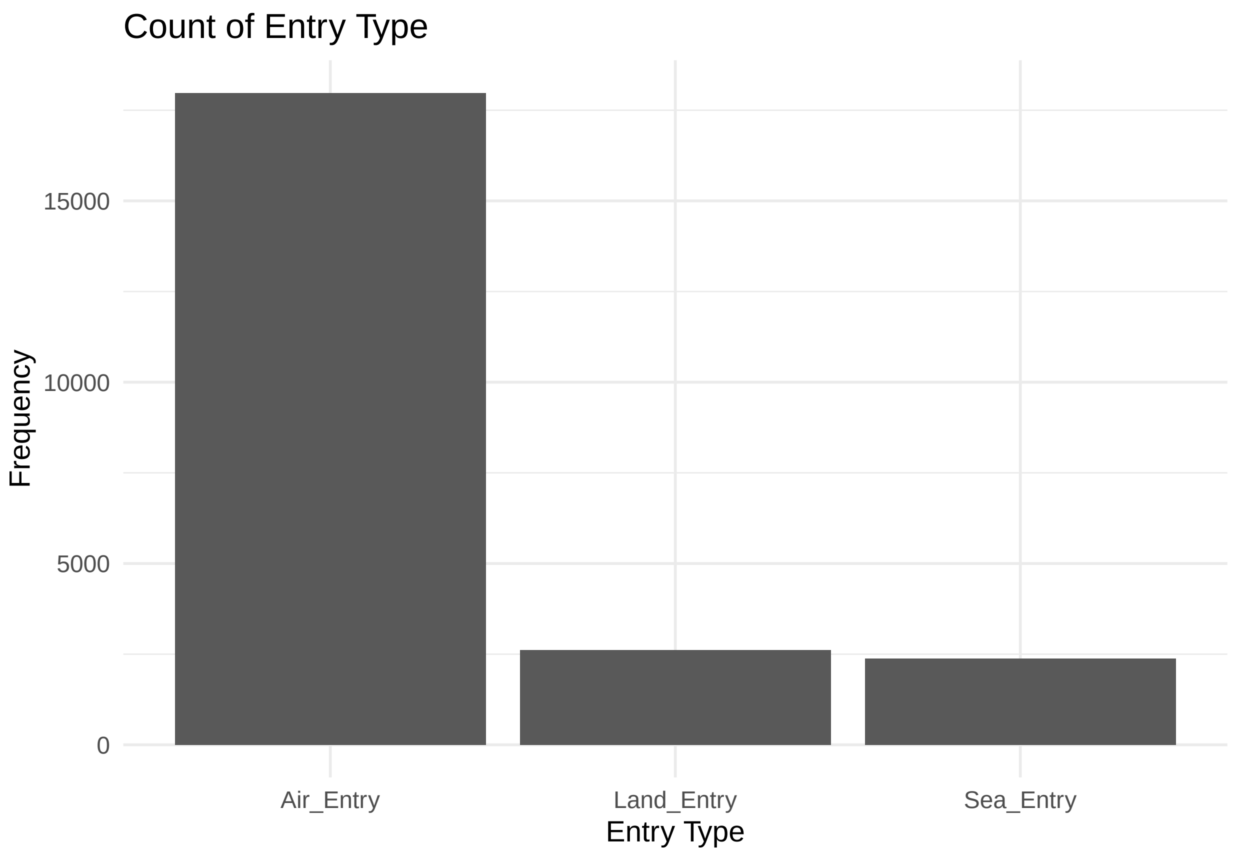
```
## .
##    1
## 5703
```

```r
# Viewing demographics of respondents
k <- 10

# Calculate the frequency of each category and select the top k
top_k_data <- data %>%
  count(`Country of Residence`) %>%
  top_n(k, n) %>%
  arrange(desc(n))

# Generate the frequency count chart for the top k categories
ggplot(top_k_data, aes(x = reorder(`Country of Residence`, -n), y = n)) +
  geom_bar(stat = "identity") +
  labs(title = paste("Top", k, "Most Frequent Country of Residence"),
       x = "Country of Residence",
       y = "Frequency") +
  theme_minimal()
```



Explaratory Data Analysis

```
data_entry <- data %>%
  summarize(
    Air_Entry = sum(!is.na(Air_Terminal)),
    Land_Entry = sum(!is.na(Land_Terminal)),
    Sea_Entry = sum(!is.na(Sea_Terminal))
  )

# Reshape the data to a long format for easier plotting
data_entry_long <- data_entry %>%
  pivot_longer(everything(), names_to = "Entry_Type", values_to = "Count")

ggplot(data_entry_long, aes(x = Entry_Type, y = Count)) +
  geom_bar(stat = "identity") +
  labs(title = "Count of Entry Type",
       x = "Entry Type",
       y = "Frequency") +
  theme_minimal()
```

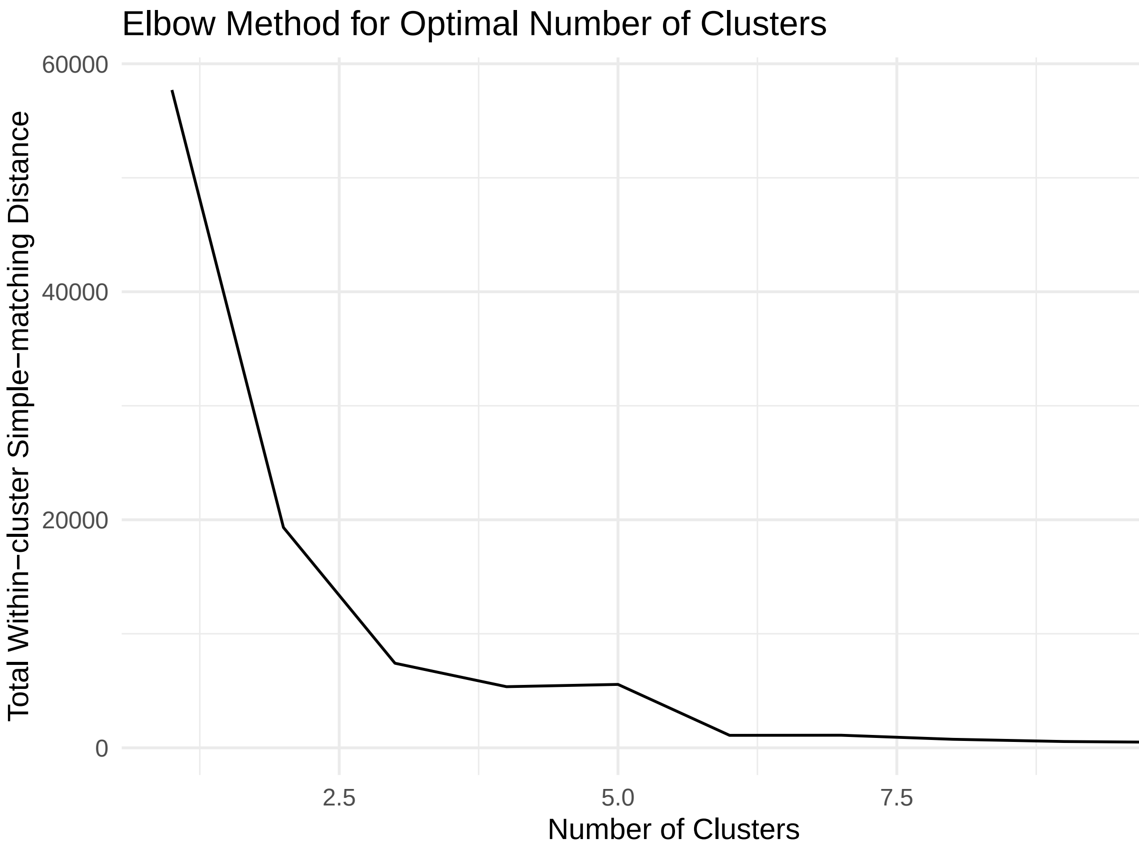


Question 2

Part 1: Clustering with Kmodes

```
# Mode of Entry
data = data %>% mutate(
  AirEntry = ifelse(!is.na(Air_Terminal),1,0),
  LandEntry = ifelse(!is.na(Land_Terminal),1,0),
  SeaEntry = ifelse(!is.na(Sea_Terminal),1,0),
  )

categorical_variables = c('Country of Residence', 'Purpose of Visit', 'AirEntry', 'LandEntry', 'SeaEntr

data_for_clustering = data[categorical_variables]
```

```
# Run Kmodes

costs <- numeric(10)

for (k in 1:10) {
  set.seed(42)
  kmodes_result <- kmodes(data_for_clustering, modes = k, iter.max = 20, weighted = FALSE)
  costs[k] <- kmodes_result$withindiff
}

# Plot the Elbow Method
qplot(1:10, costs, geom = "line") +
  labs(title = "Elbow Method for Optimal Number of Clusters", x = "Number of Clusters", y = "Total Withi
  theme_minimal()
```



Feature Engineering

```
# Run K-Modes clustering
set.seed(42)
kmodes_result <- kmodes(data_for_clustering, modes = 5, iter.max = 10, weighted = FALSE)

# Add cluster labels to the original dataset
data_for_clustering$Cluster <- kmodes_result$cluster

# Summarize the cluster profiles
cluster_profiles <- data_for_clustering %>%
  group_by(Cluster) %>%
  summarize(across(everything(), ~ names(sort(table(.), decreasing = TRUE))[1]))

print(cluster_profiles)
```

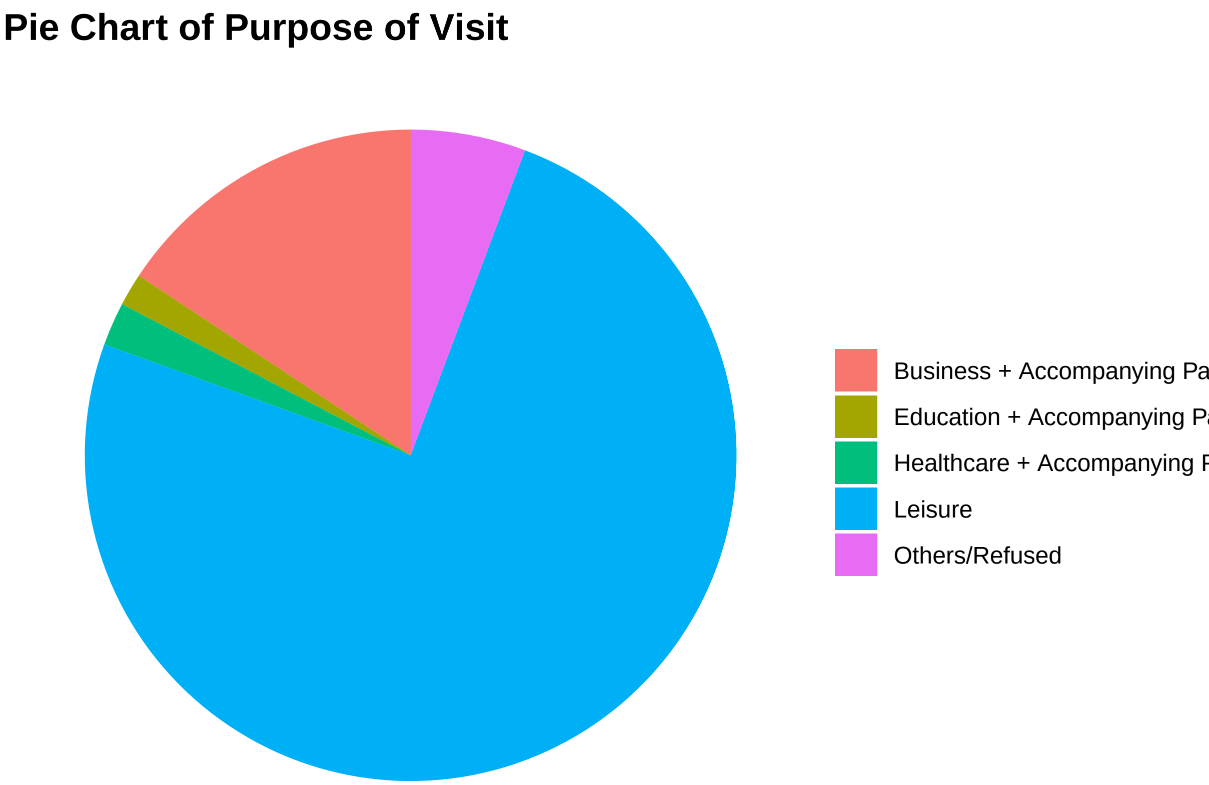
```
## # A tibble: 5 x 9
##   Cluster ‘Country of Residence‘ ‘Purpose of Visit‘  AirEntry LandEntry SeaEntry
##     <int> <chr>                  <chr>               <chr>    <chr>     <chr>
## 1       1 Indonesia              Business + Accompa~ 1        0         0
## 2       2 Indonesia              Leisure             1        0         0
## 3       3 Indonesia              Leisure             1        0         0
## 4       4 China                  Leisure             0        1         0
## 5       5 Indonesia              Leisure             1        0         0
## # i 3 more variables: Gender <chr>, ‘Marital Status‘ <chr>,
## #   ‘Travel companion - Alone‘ <chr>
```

```
# Count the number of observations in each cluster
cluster_counts <- data_for_clustering %>%
  group_by(Cluster) %>%
  summarize(Count = n())

# Combine dataframes to obtain statistics
clusters <- inner_join(cluster_profiles, cluster_counts, by = "Cluster")

print(clusters)
```

```
## # A tibble: 5 x 10
##   Cluster ‘Country of Residence‘ ‘Purpose of Visit‘  AirEntry LandEntry SeaEntry
##     <int> <chr>                  <chr>               <chr>    <chr>     <chr>
## 1       1 Indonesia              Business + Accompa~ 1        0         0
## 2       2 Indonesia              Leisure             1        0         0
## 3       3 Indonesia              Leisure             1        0         0
## 4       4 China                  Leisure             0        1         0
## 5       5 Indonesia              Leisure             1        0         0
## # i 4 more variables: Gender <chr>, ‘Marital Status‘ <chr>,
## #   ‘Travel companion - Alone‘ <chr>, Count <int>
```

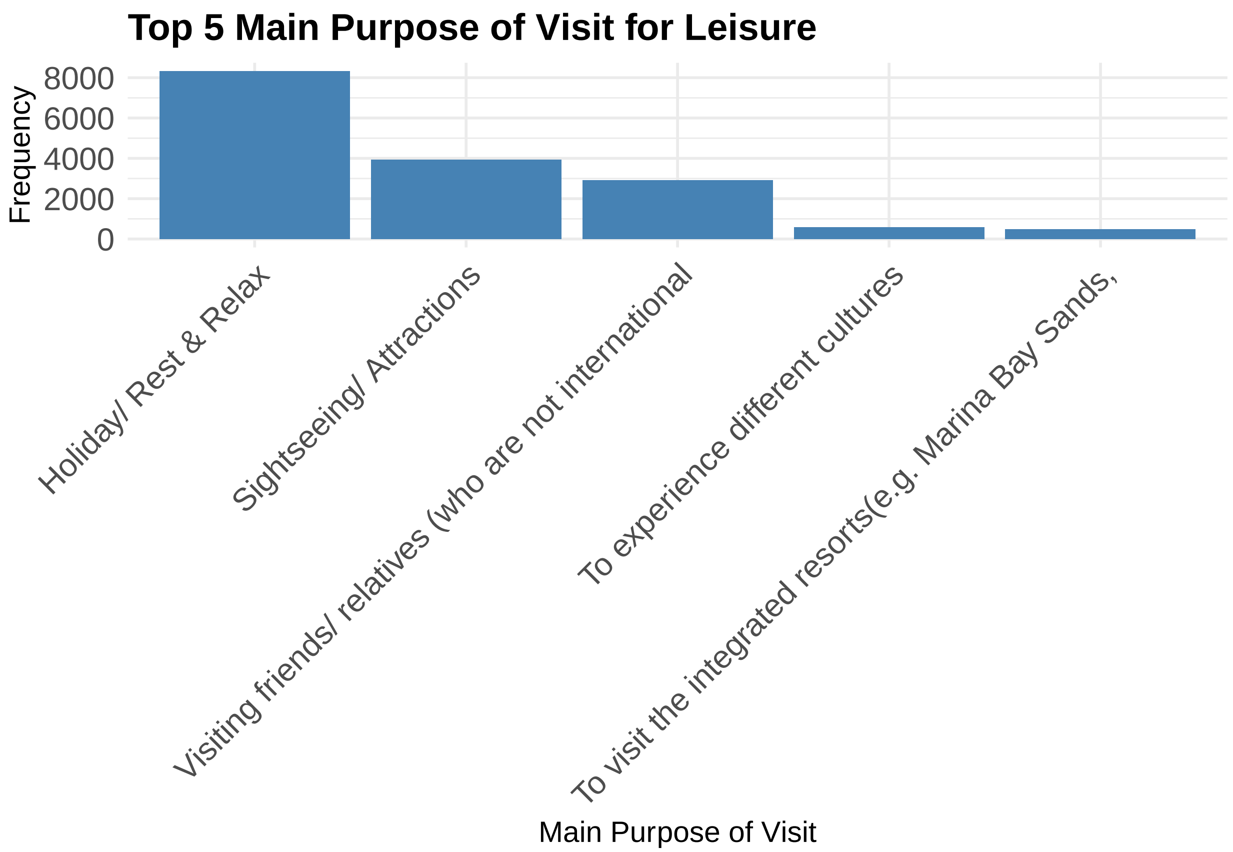
Part 2

```


### Purpose of Visit
data %>% count(`Purpose of Visit`) %>%
  ggplot(aes(x = "", y = n, fill = `Purpose of Visit`)) +
    geom_bar(stat = "identity", width = 1) +
    coord_polar("y") +
    labs(title = "Pie Chart of Purpose of Visit", x = "", y = "") +
    theme_void() +
    theme(legend.title = element_blank(), plot.title = element_text(size = 14, face = "bold"))
```

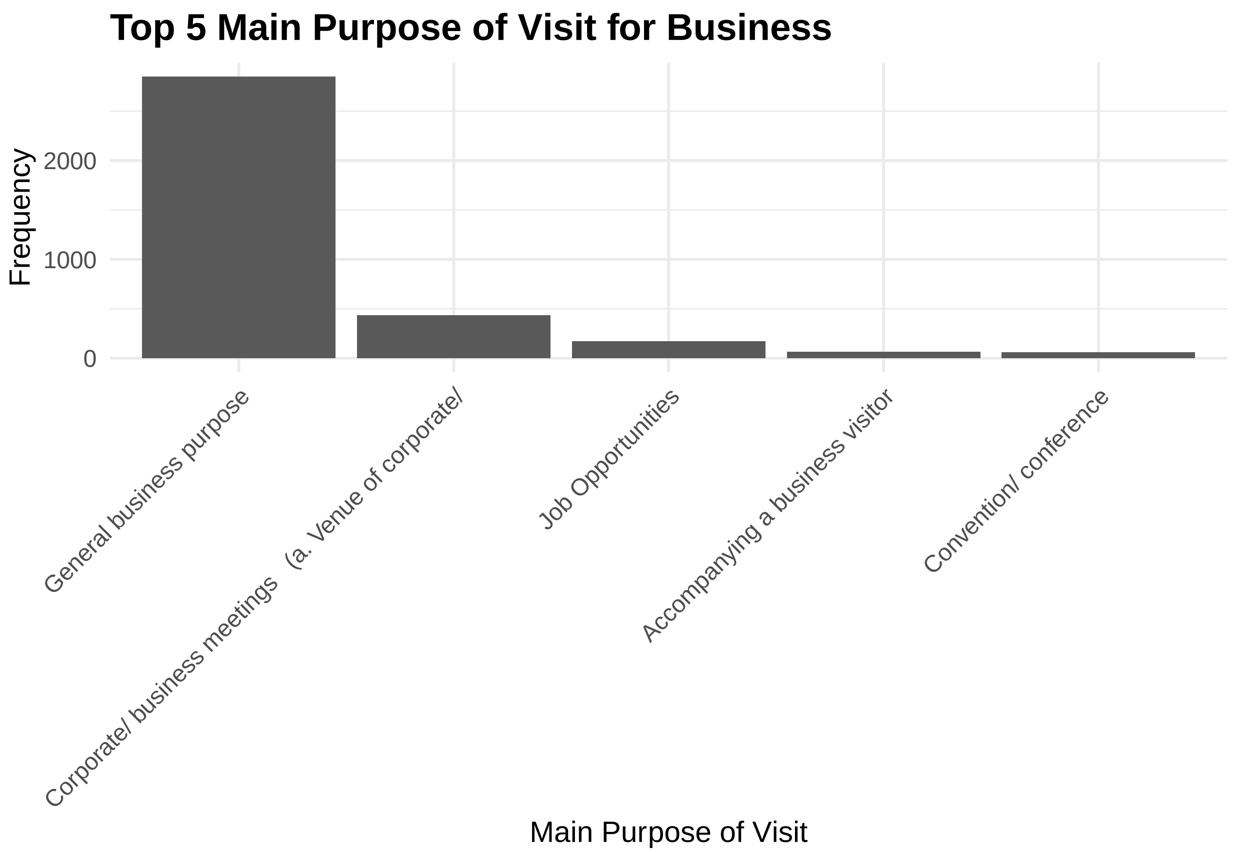


Visitor Insights

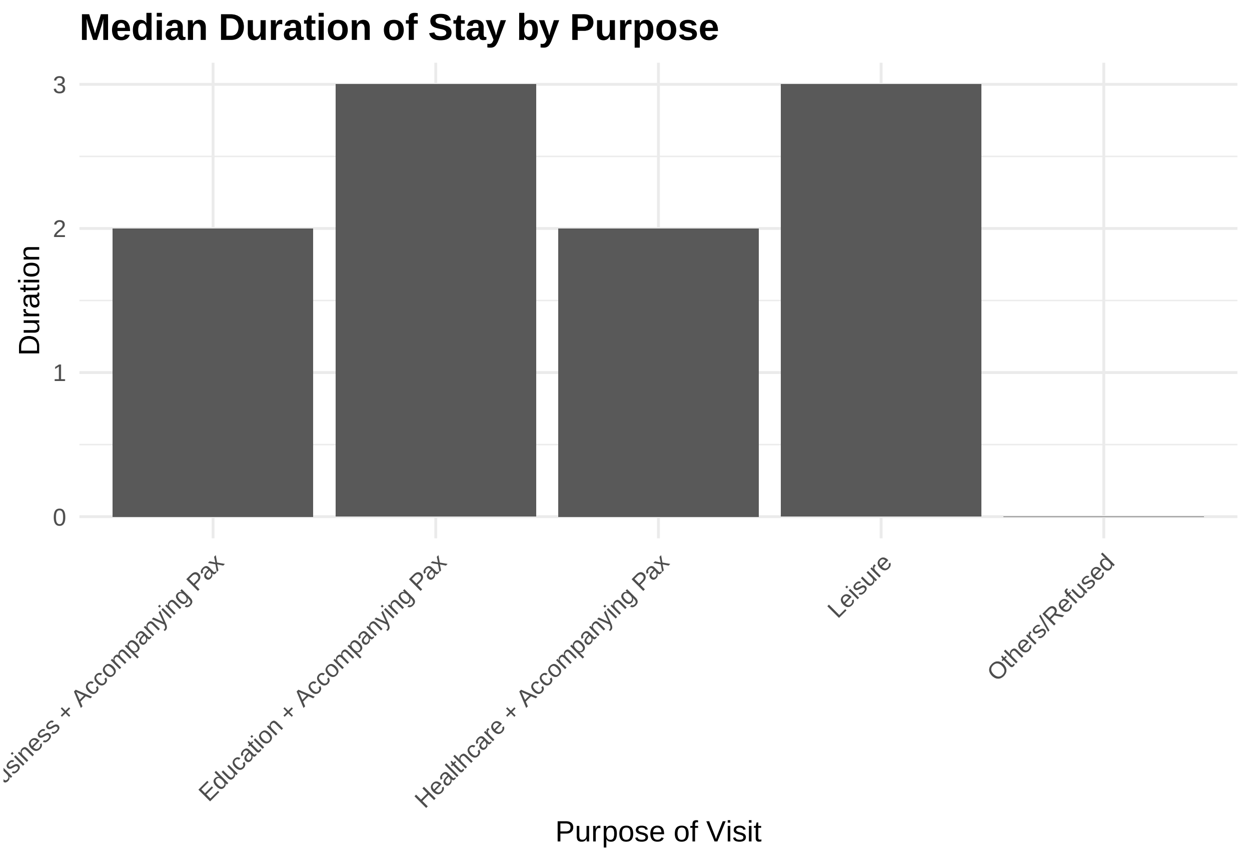
```
### Main Purpose for Leisure
k <- 5

# Calculate the frequency of each category and select the top k
top_k_data <- data %>%
  filter(`Purpose of Visit` == "Leisure") %>%
  count(`Main Purpose of Visit`) %>%
  top_n(k, n) %>%
  arrange(desc(n))

# Generate the frequency count chart for the top k categories
ggplot(top_k_data, aes(x = reorder(`Main Purpose of Visit`, -n), y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = paste("Top", k, "Main Purpose of Visit for Leisure"),
       x = "Main Purpose of Visit",
       y = "Frequency") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 12),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 14, face = "bold"))
```



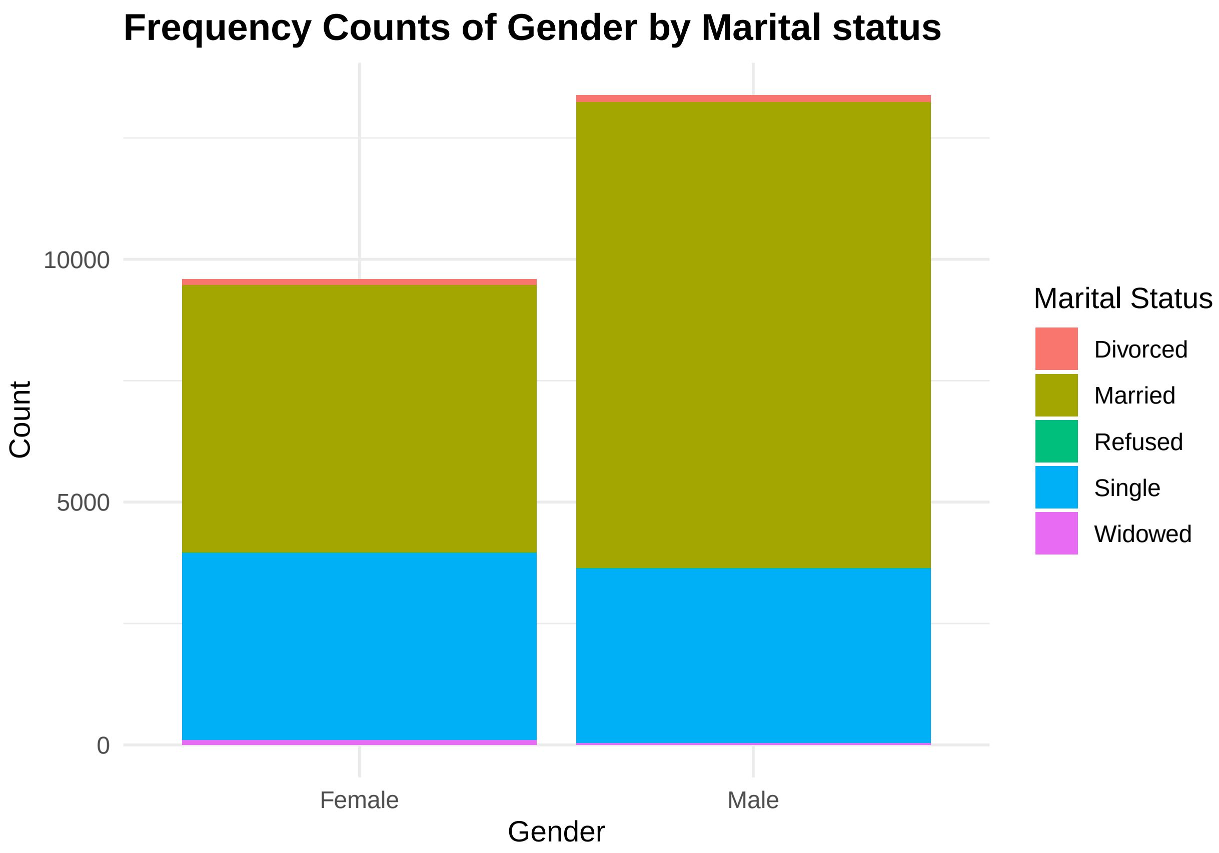
```
### Main Purpose for Business
# Calculate the frequency of each category and select the top k
top_k_data_biz <- data %>%
  filter(`Purpose of Visit` == "Business + Accompanying Pax") %>%
  count(`Main Purpose of Visit`) %>%
  top_n(k, n) %>%
  arrange(desc(n))

# Generate the frequency count chart for the top k categories
ggplot(top_k_data_biz, aes(x = reorder(`Main Purpose of Visit`, -n), y = n)) +
  geom_bar(stat = "identity") +
  labs(title = paste("Top", k, "Main Purpose of Visit for Business"),
       x = "Main Purpose of Visit",
       y = "Frequency") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(plot.title = element_text(size = 14, face = "bold"))
```



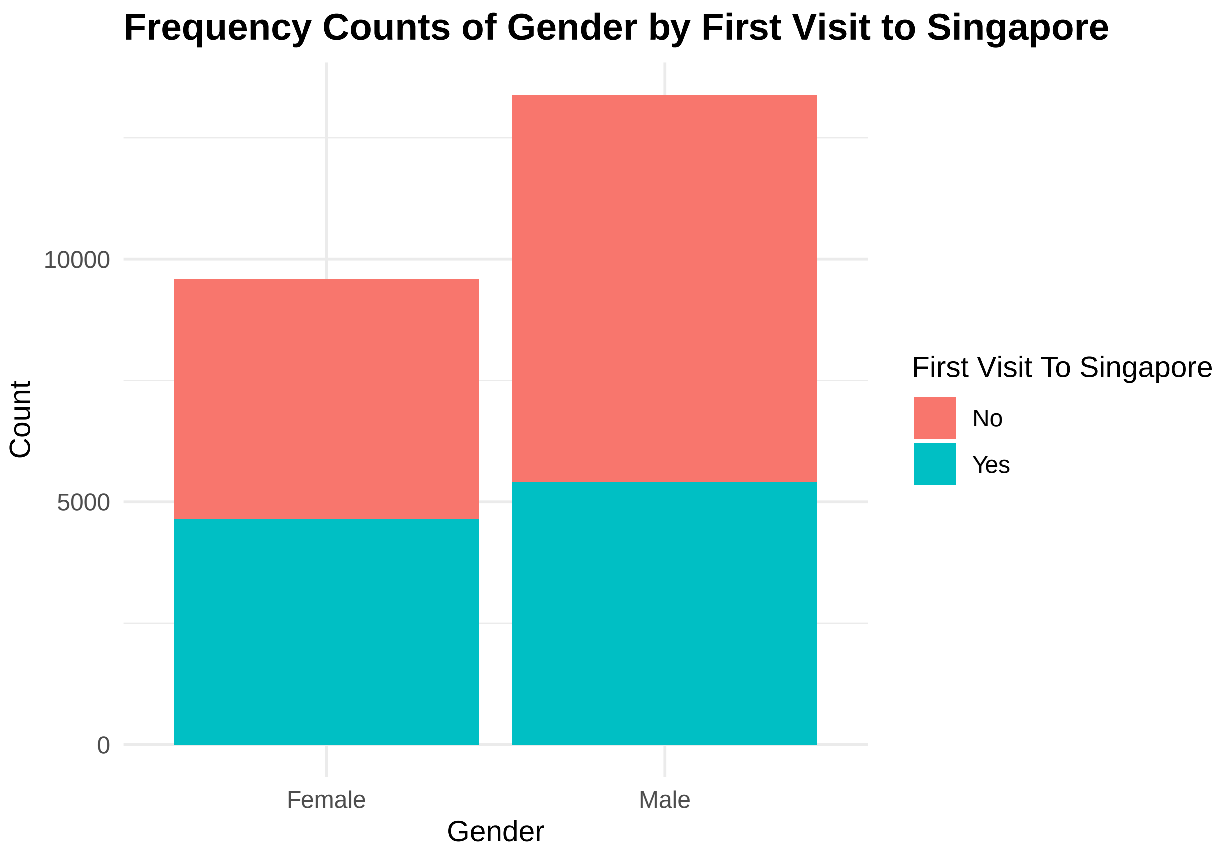
```
### Duration of Stay by Purpose
data %>% group_by(`Purpose of Visit`) %>% summarize(Duration = median(`Length of Stay`)) %>%
  ggplot(aes(x=`Purpose of Visit`, y = `Duration`)) + geom_col() + labs(title = "Median Duration of Sta
      theme_minimal() + theme(legend.title = element_blank(), plot.title = element_text(size = 14, face
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



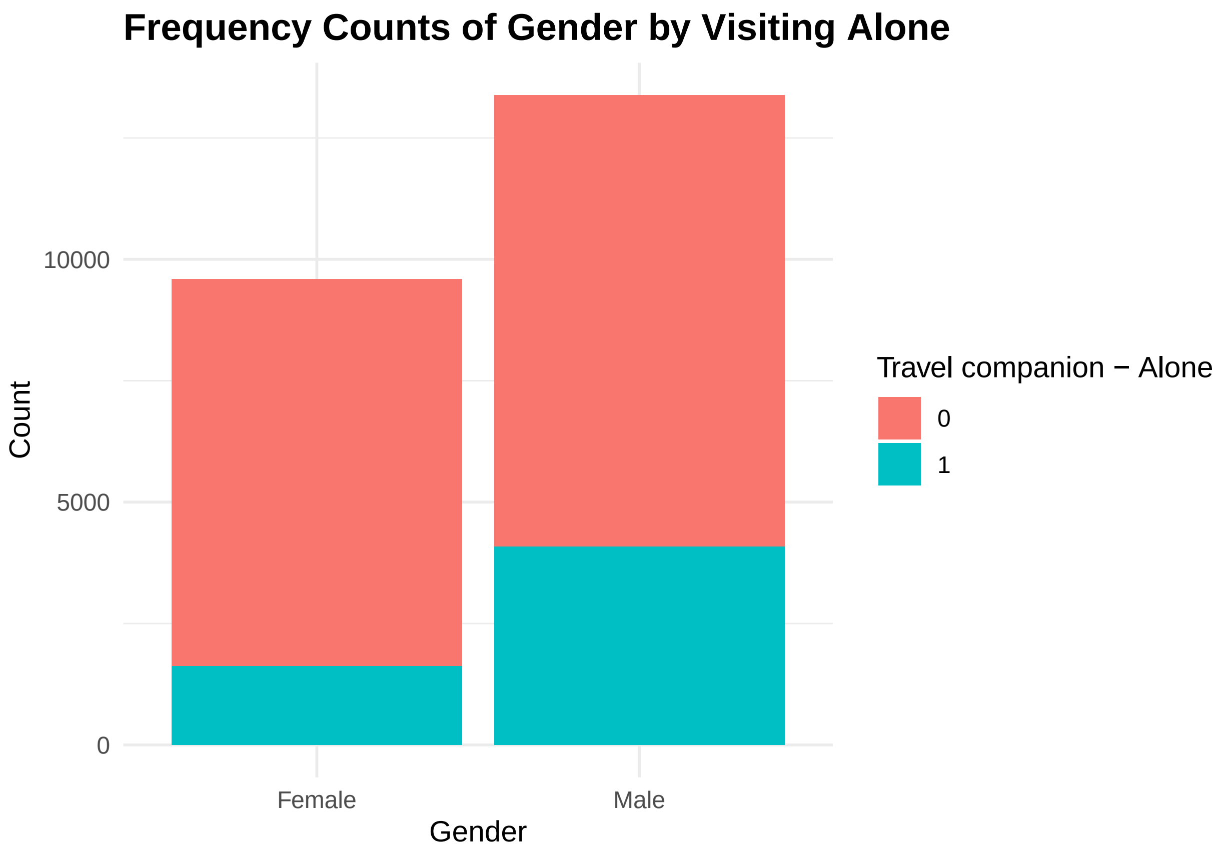
```
### Gender and Marital Status
ggplot(data, aes(x = `Gender`, fill = `Marital Status`)) +
  geom_bar() +
  labs(title = "Frequency Counts of Gender by Marital status", x = "Gender", y = "Count") + theme_minima
```



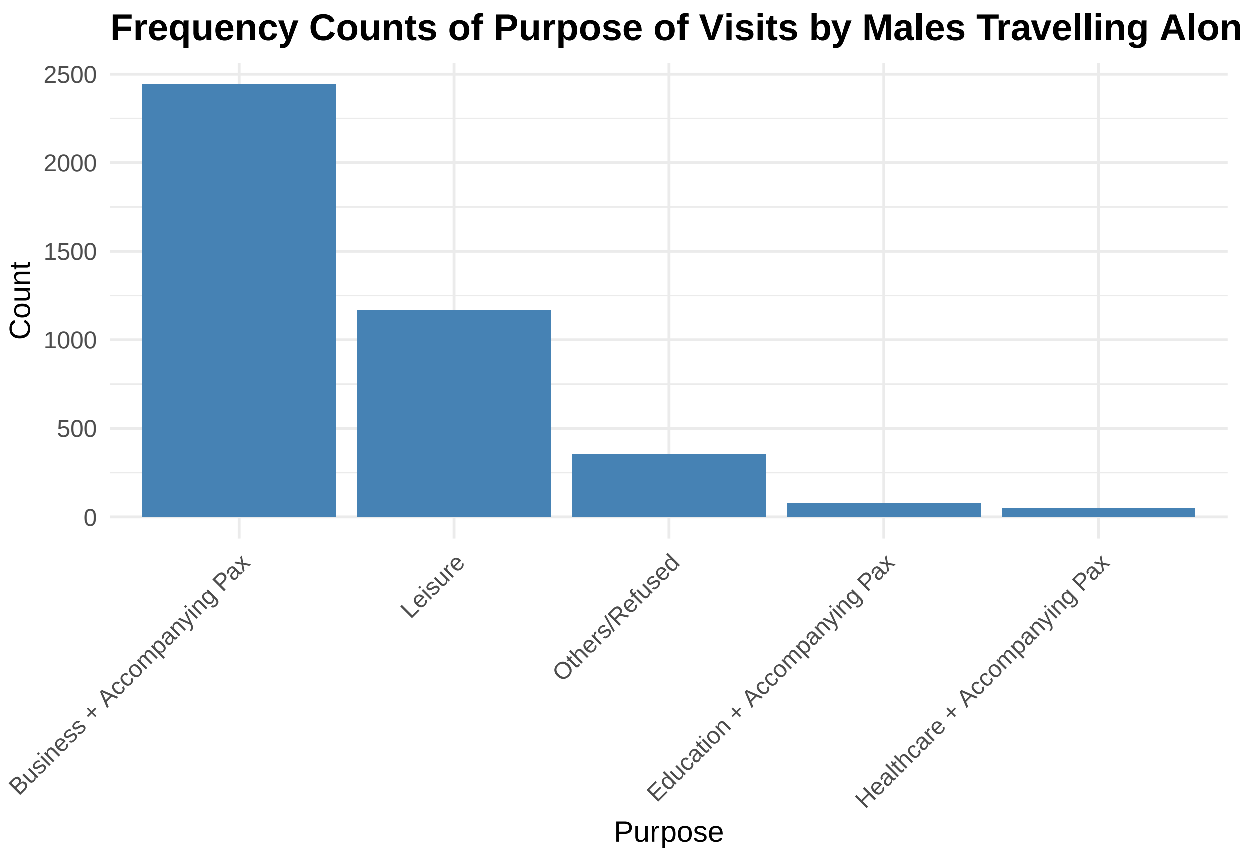
```
### Gender and First Time Visitors
ggplot(data, aes(x = `Gender`, fill = `First Visit To Singapore`)) +
  geom_bar() +
  labs(title = "Frequency Counts of Gender by First Visit to Singapore", x = "Gender", y = "Count") + th
```



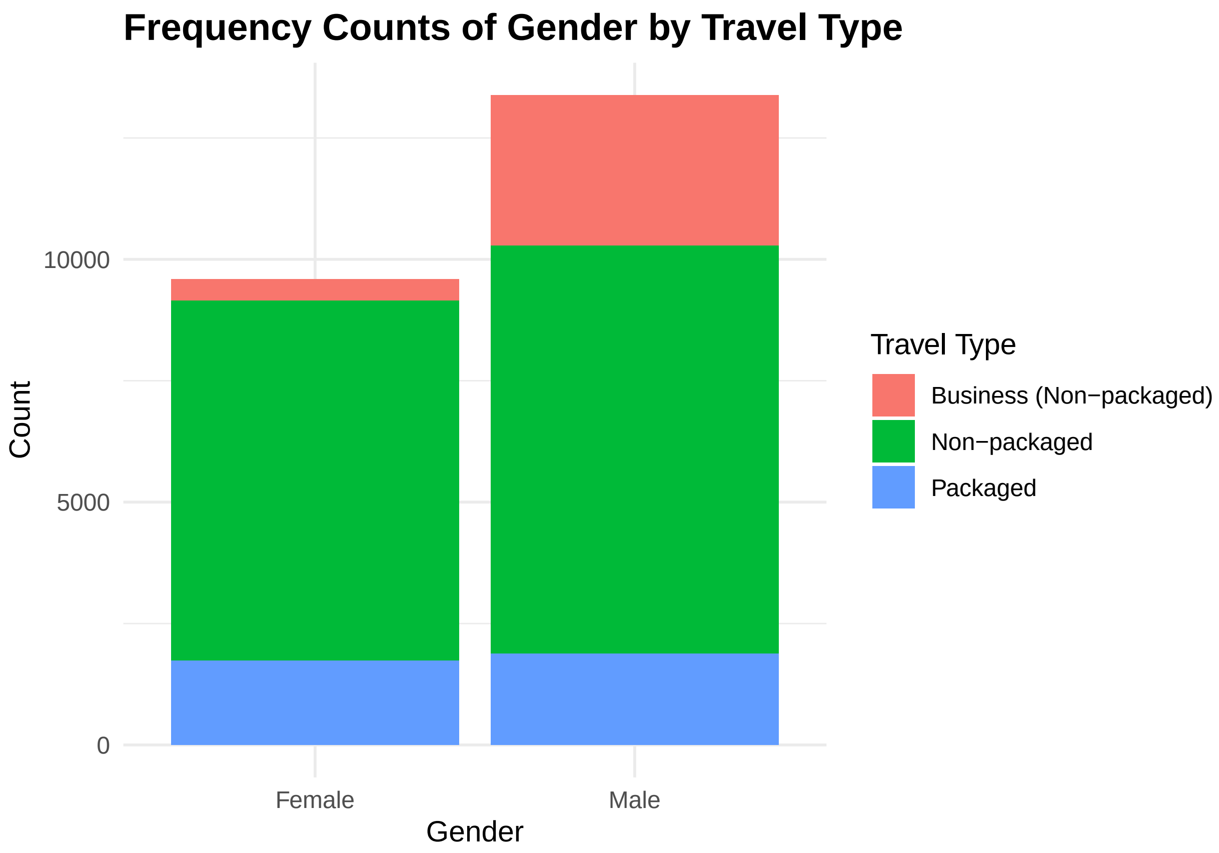
```r
### Gender and Visiting Alone
ggplot(data, aes(x = `Gender`, fill = `Travel companion - Alone`)) +
  geom_bar() +
  labs(title = "Frequency Counts of Gender by Visiting Alone", x = "Gender", y = "Count") + theme_minima
```



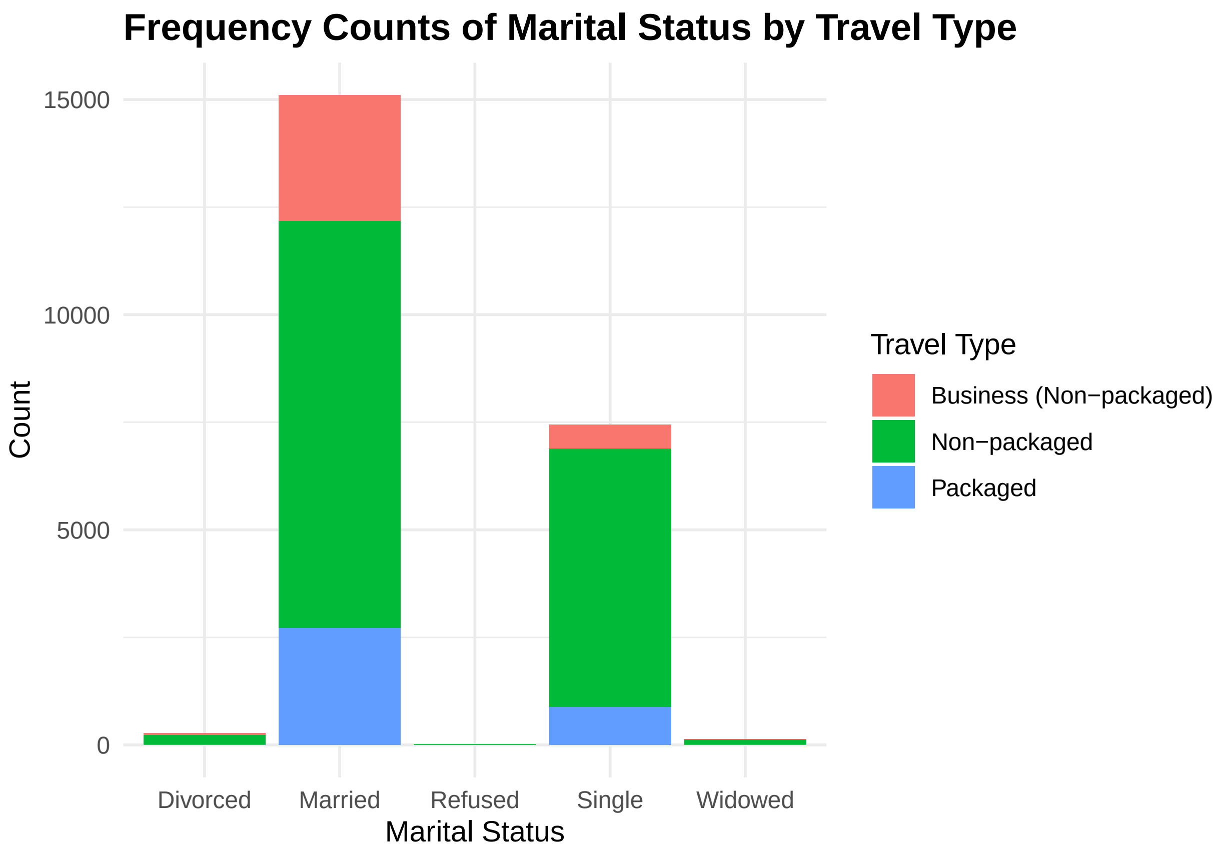
```
### Males Visiting Alone by Purpose
data %>%
  filter(`Gender` == "Male", `Travel companion - Alone` == 1) %>%
  count(`Purpose of Visit`) %>%
  ggplot(aes(x = reorder(`Purpose of Visit`, -n), y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Frequency Counts of Purpose of Visits by Males Travelling Alone", x = "Purpose", y = "C
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(size = 14, face =
```



```
### Gender and Travel Type
ggplot(data, aes(x = `Gender`, fill = `Travel Type`)) +
  geom_bar() +
  labs(title = "Frequency Counts of Gender by Travel Type", x = "Gender", y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```



```r
### Marital Status and Travel Type
ggplot(data, aes(x = `Marital Status`, fill = `Travel Type`)) +
  geom_bar() +
  labs(title = "Frequency Counts of Marital Status by Travel Type", x = "Marital Status", y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```



```
### Filtering out inconsistent entries and reducing dataset size to correct entries
corrections = data %>%
  filter(`MainHotel` != max(as.double(`MainHotel`))) %>%
  dplyr::select(starts_with("Travel companion")) %>%
  mutate_all(as.double) %>%
  rowSums()

data_reduced = data %>% filter(`MainHotel` != max(as.double(`MainHotel`))) %>%
  dplyr::select(`MainHotel`, starts_with("Travel companion")) %>%
  mutate(corrections = corrections) %>%
  filter(corrections != 0) %>% dplyr::select(-corrections)

data_reduced
```

Travel Companion and Hotel

```
## # A tibble: 14,973 x 11
##    MainHotel ‘Travel companion - Alone‘ ‘Travel companion - Spouse‘
##    <chr>     <chr>                      <chr>
##  1 15        0                          1
##  2 30        1                          0
##  3 52        0                          1
##  4 69        0                          0
##  5 51        0                          1
##  6 64        1                          0
##  7 67        0                          0
##  8 882       0                          0
##  9 80        0                          0
## 10 67        0                          0
## # i 14,963 more rows
## # i 8 more variables: ‘Travel companion - Your Child/Children‘ <chr>,
## #   ‘Travel companion - Parents/Parents-in-law‘ <chr>,
## #   ‘Travel companion - Grandparents/Grandparents-in-law‘ <chr>,
## #   ‘Travel companion - Siblings‘ <chr>,
## #   ‘Travel companion - Other relatives‘ <chr>,
## #   ‘Travel companion - Friends‘ <chr>, ...
```

```r
# Perform Chi-Square Test to determine correlation
data$MainHotel <- as.factor(data$MainHotel)

travel_companion_cols <- grep("^Travel companion", names(data), value = TRUE)

results <- list()

for (col in travel_companion_cols) {
  contingency_table <- table(data[[col]], data$MainHotel)
  chi_square_result <- chisq.test(contingency_table)
  results[[col]] <- list(
    p_value = chi_square_result$p.value,
    chi_square_statistic = chi_square_result$statistic,
    conclusion = ifelse(chi_square_result$p.value < 0.01,
                        "Significant correlation",
                        "No significant correlation")
  )
}

for (col in names(results)) {
  cat("Results for:", col, "\n")
  print(results[[col]])
  cat("\n")
}
```

```
## Results for: Travel companion - Alone
## $p_value
## [1] 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
##
## $chi_square_statistic
## X-squared
##  1959.449
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Spouse
## $p_value
## [1] 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
```

```
##
## $chi_square_statistic
## X-squared
##  1702.929
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Your Child/Children
## $p_value
## [1] 0.0000000000000000000000000000000000000000000000000000000000003381205
##
## $chi_square_statistic
## X-squared
##  1041.398
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Parents/Parents-in-law
## $p_value
## [1] 0.0001208375
##
## $chi_square_statistic
## X-squared
##  498.6644
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Grandparents/Grandparents-in-law
## $p_value
## [1] 0.5165437
##
## $chi_square_statistic
## X-squared
##  386.1803
##
## $conclusion
## [1] "No significant correlation"
##
##
## Results for: Travel companion - Siblings
## $p_value
## [1] 0.0005971083
##
## $chi_square_statistic
## X-squared
##  484.6365
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Other relatives
## $p_value
## [1] 0.000000000000000000000000000000000000000000000000000000000000004808933
##
## $chi_square_statistic
## X-squared
##  1054.839
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Friends
## $p_value
## [1] 0.00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
##
## $chi_square_statistic
## X-squared
##  1303.674
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Business associates/Colleagues
## $p_value
## [1] 0.000000000000001936283
##
## $chi_square_statistic
## X-squared
##  649.0086
##
## $conclusion
## [1] "Significant correlation"
##
##
## Results for: Travel companion - Others
## $p_value
## [1] 0.0000000000000000101254
##
## $chi_square_statistic
## X-squared
##  673.9947
##
## $conclusion
## [1] "Significant correlation"
```

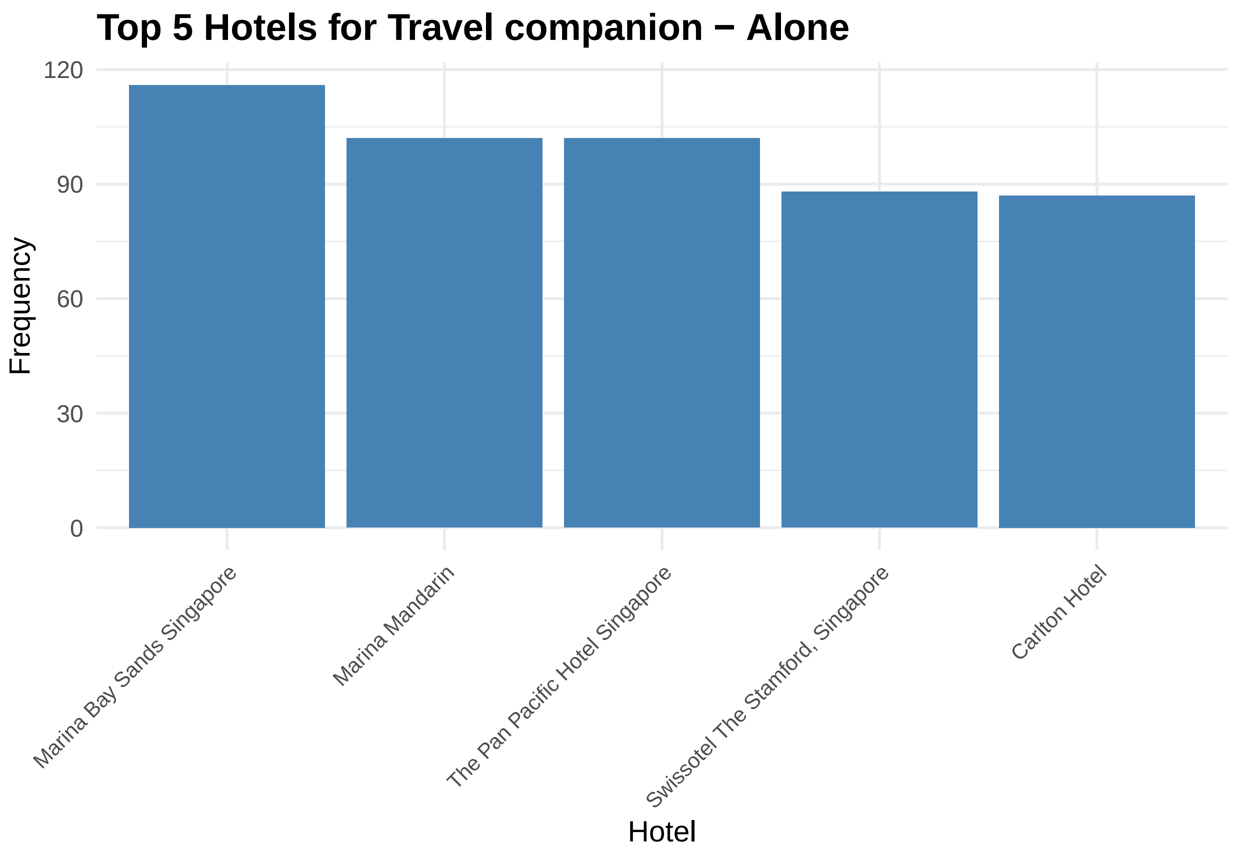
```
# load hotel dataset
```

```r
hotel_data = read_csv("data/hotel_info.csv") %>% mutate(MainHotel = as.factor(MainHotel))
data_reduced = data_reduced %>% inner_join(hotel_data, by = "MainHotel")
```

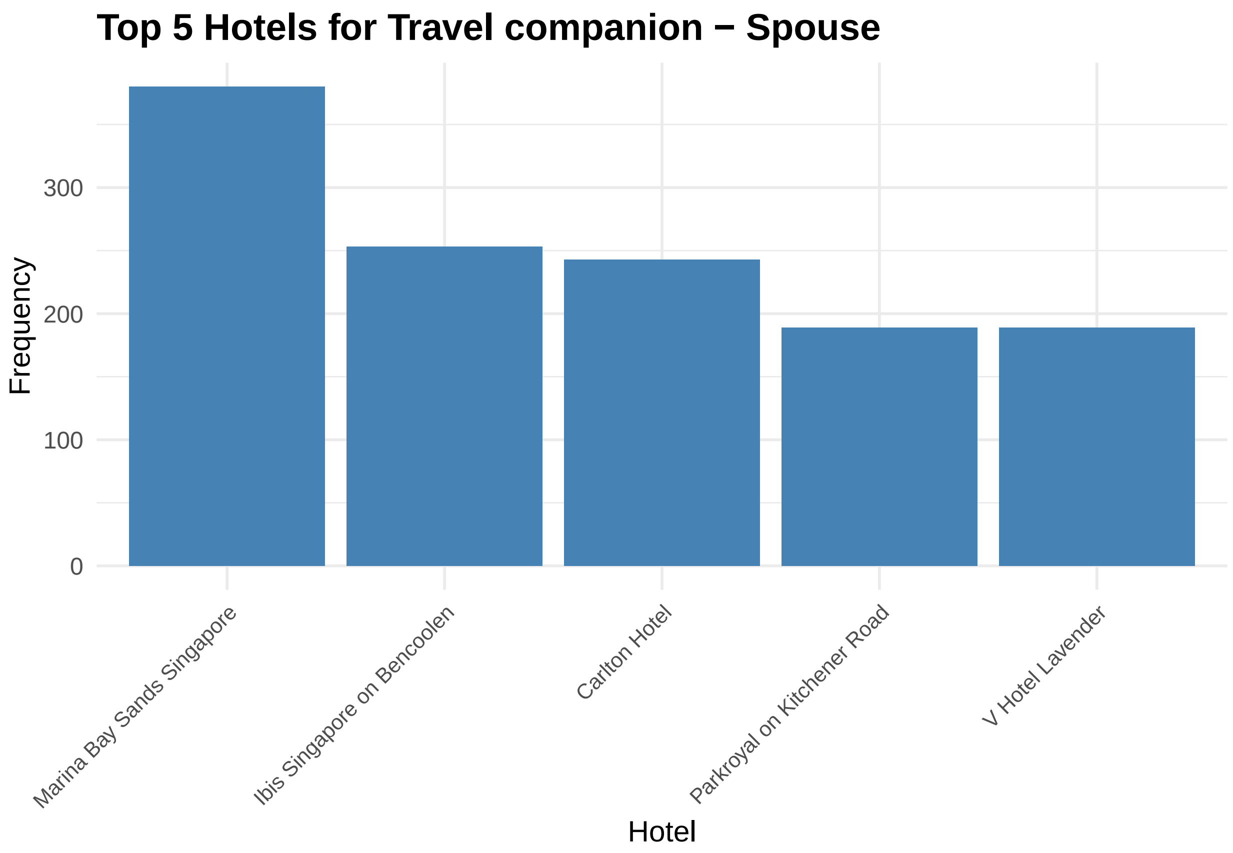
```r


# Create barplots to get top 5 hotels for each travel companion
for (col in travel_companion_cols) {

  filtered_data <- data_reduced %>% dplyr::select(col, `Hotel_Name`) %>% filter(!!sym(col) == 1)

  print(filtered_data)
  top_hotels <- filtered_data %>%
    count(`Hotel_Name`) %>%
    arrange(desc(n)) %>%
    top_n(5, n)


  dummy = ggplot(top_hotels, aes(x = reorder(`Hotel_Name`, -n), y = n)) +
    geom_bar(stat = "identity", fill = "steelblue") +
    theme_minimal() +
    labs(title = paste("Top 5 Hotels for", col),
         x = "Hotel",
         y = "Frequency") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8), plot.title = element_text(size =

  print(dummy)

}
```

Viewing Hotel Insights

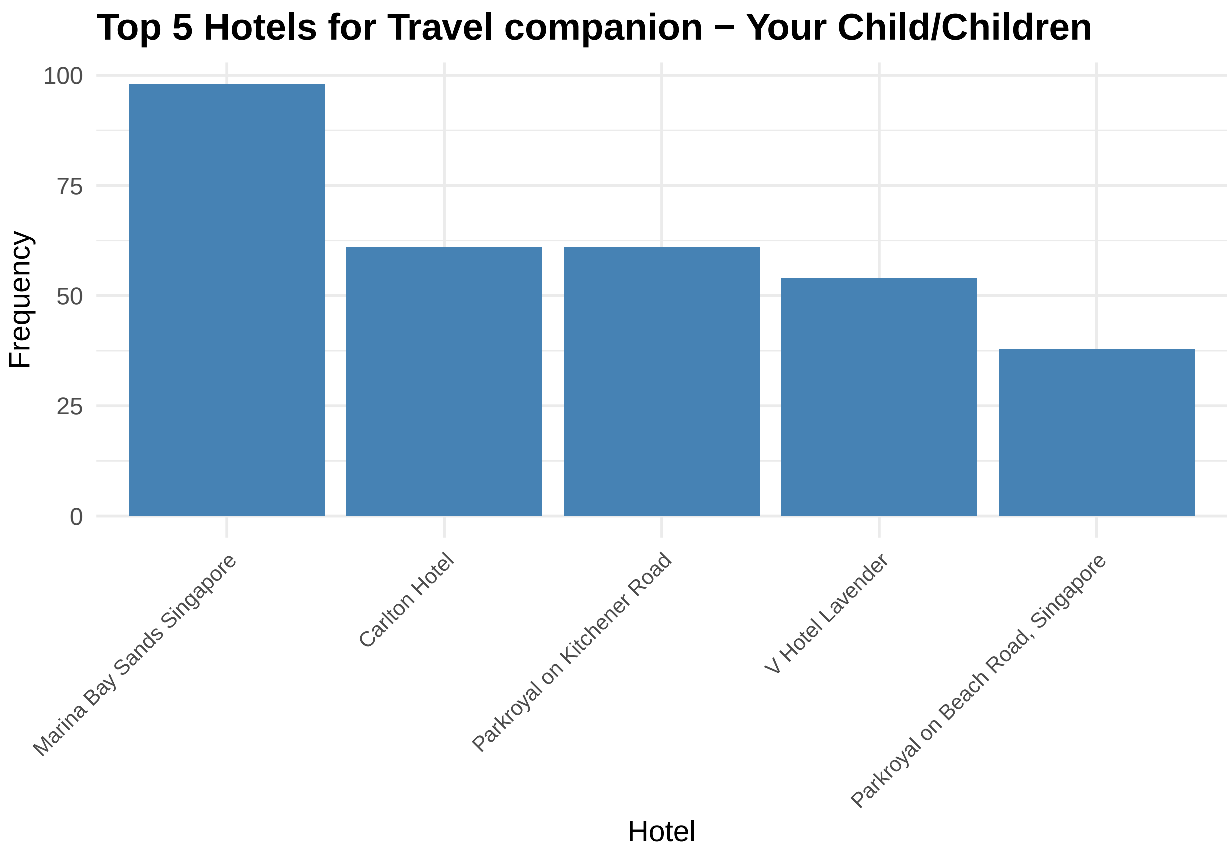
```
## # A tibble: 3,105 x 2
##    ‘Travel companion - Alone‘ Hotel_Name
##    <chr>                      <chr>
##  1 1                          Hilton Singapore
##  2 1                          Peninsula.Excelsior Hotel
##  3 1                          Parkroyal on Pickering
##  4 1                          V Hotel Lavender
##  5 1                          Carlton Hotel
##  6 1                          Concorde Hotel Singapore
##  7 1                          Hotel Grand Central
##  8 1                          Days Hotel Singapore at Zhongshan Park
##  9 1                          Amara Singapore
## 10 1                          Conrad Centennial Singapore
## # i 3,095 more rows
```



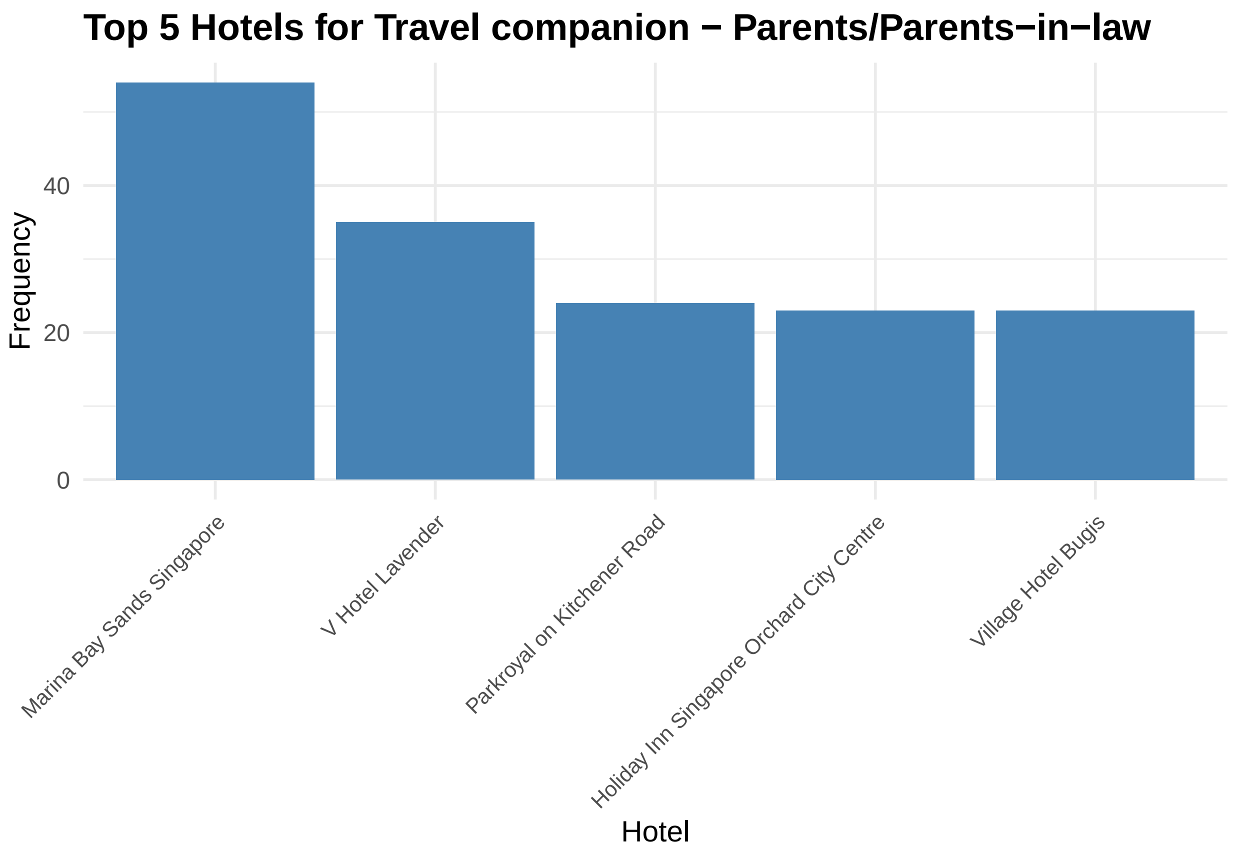
```
## # A tibble: 7,753 x 2
##    `Travel companion - Spouse` Hotel_Name
##    <chr>                       <chr>
##  1 1                           Fortuna Hotel
##  2 1                           Pan Pacific Orchard
##  3 1                           Mandarin Orchard Singapore
##  4 1                           Park Regis Singapore
##  5 1                           Holiday Inn Singapore Orchard City Centre
##  6 1                           Fragrance Hotel - Riverside
##  7 1                           Peninsula.Excelsior Hotel
##  8 1                           The Claremont
##  9 1                           Orchard Hotel Singapore
## 10 1                           Marina Mandarin
## # i 7,743 more rows
```



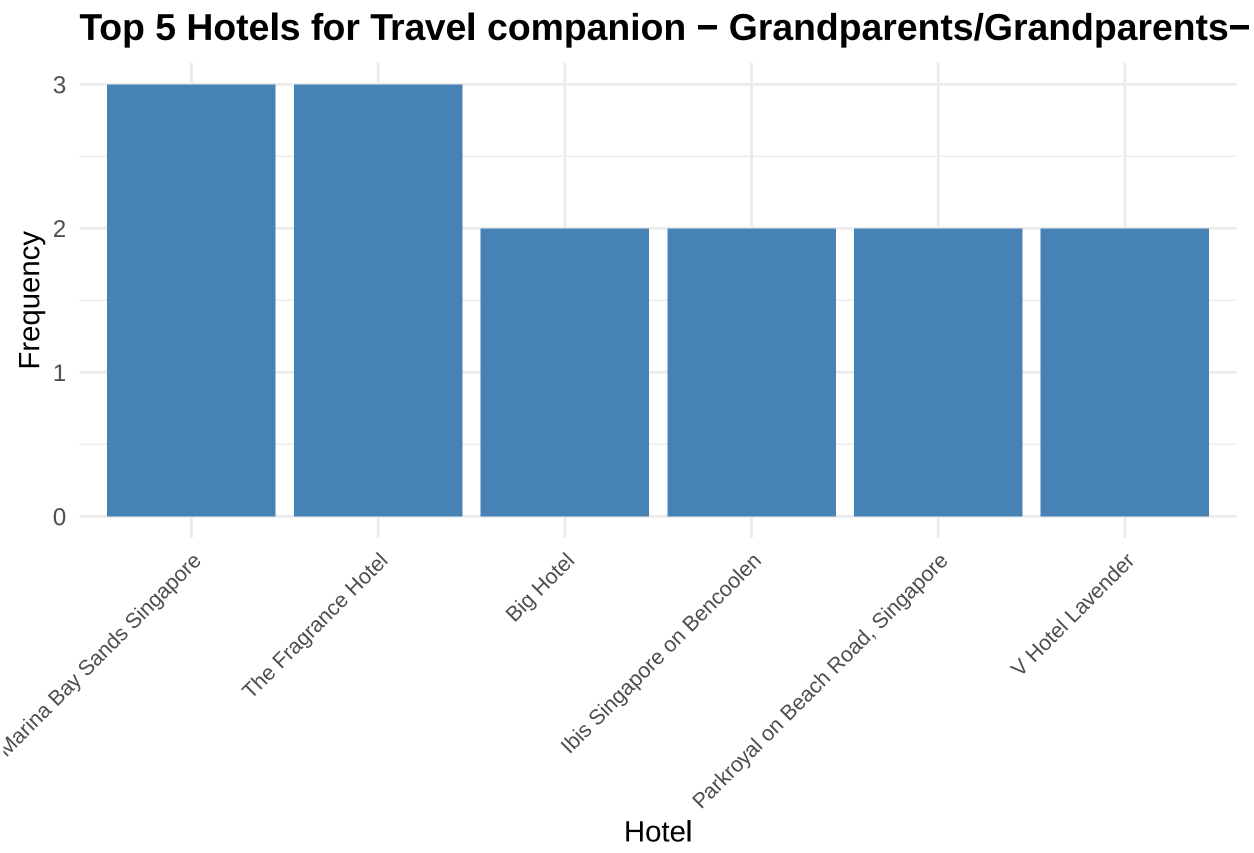
```
## # A tibble: 1,951 x 2
##    `Travel companion - Your Child/Children` Hotel_Name
##    <chr>                                    <chr>
##  1 1                                        Hotel 1929
##  2 1                                        The Claremont
##  3 1                                        Village Hotel Bugis
##  4 1                                        Shangri-La Hotel
##  5 1                                        Hotel Michael
##  6 1                                        Fragrance Hotel - Sapphire
##  7 1                                        Ramada Singapore at Zhongshan Park
##  8 1                                        Hotel Grand Pacific
##  9 1                                        InterContinental Singapore
## 10 1                                        V Hotel Lavender
## # i 1,941 more rows
```



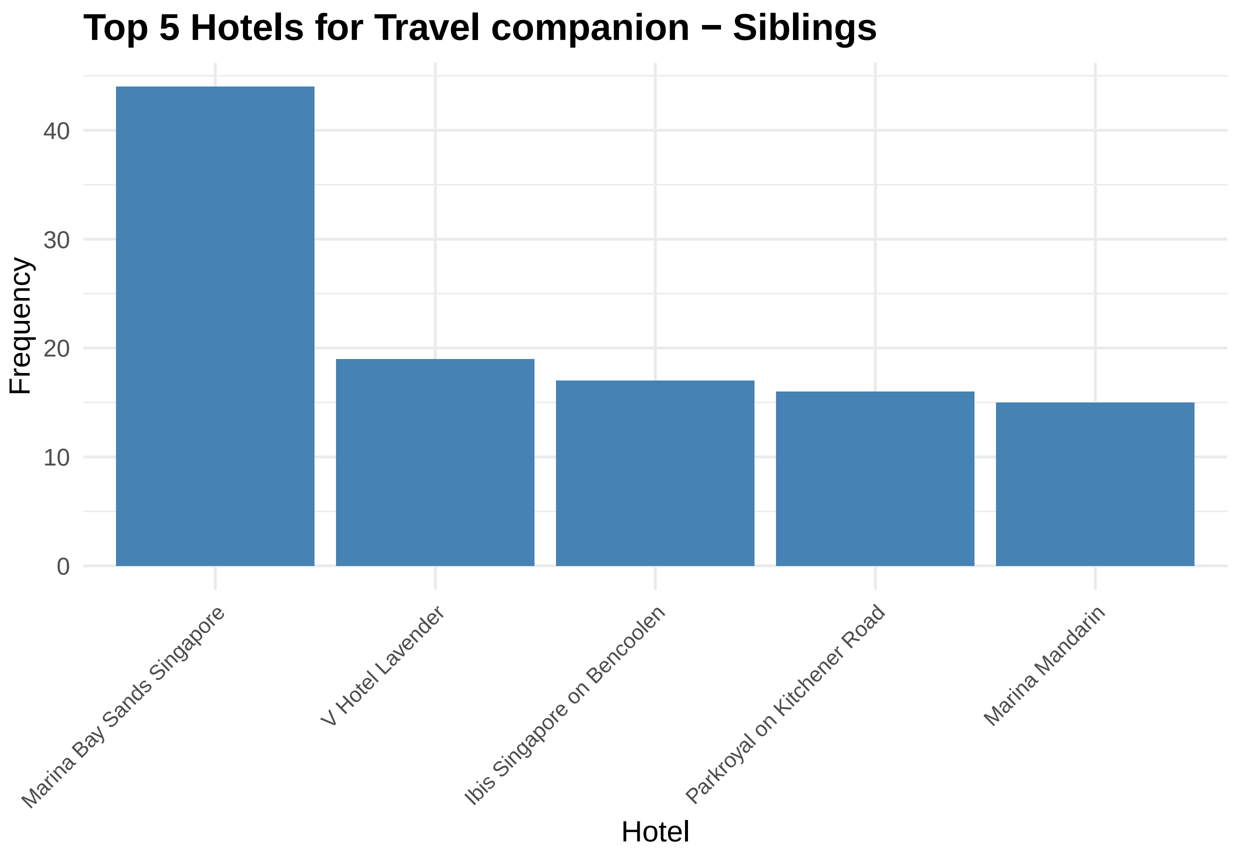
```
## # A tibble: 987 x 2
##    `Travel companion - Parents/Parents-in-law` Hotel_Name
##    <chr>                                       <chr>
##  1 1                                           Fairmont, Singapore
##  2 1                                           Fairmont, Singapore
##  3 1                                           Crowne Plaza Changi Airport
##  4 1                                           The Southbridge Hotel
##  5 1                                           Mayo Inn
##  6 1                                           Big Hotel
##  7 1                                           Hotel Chancellor @ Orchard
##  8 1                                           M Hotel Singapore
##  9 1                                           Hotel Chancellor @ Orchard
## 10 1                                           The Ritz-Carlton, Millenia Singa~
## # i 977 more rows
```



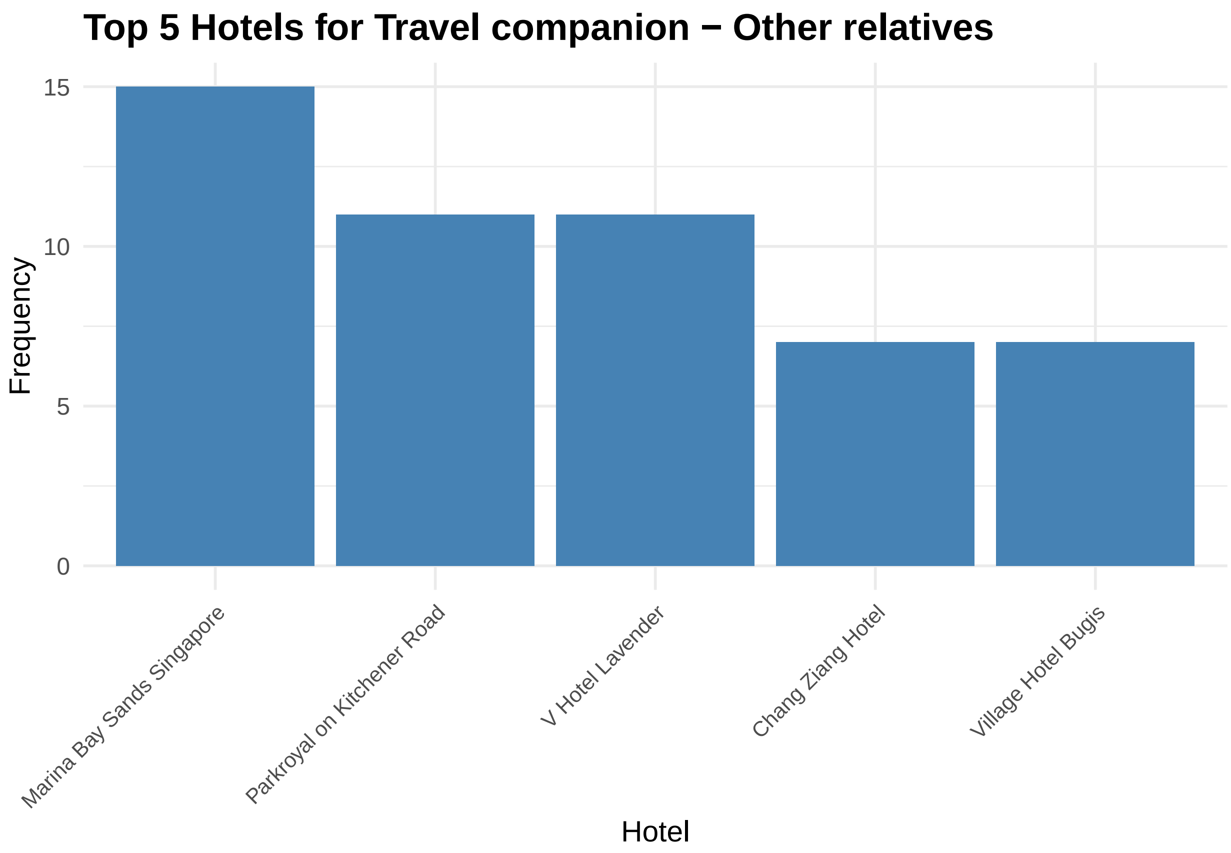
```
## # A tibble: 40 x 2
##    `Travel companion - Grandparents/Grandparents-in-law` Hotel_Name
##    <chr>                                                 <chr>
##  1 1                                                     Hotel Jen Orchardgatew~
##  2 1                                                     M Hotel Singapore
##  3 1                                                     Hotel Chancellor @ Orc~
##  4 1                                                     Grand Park Orchard
##  5 1                                                     The Scarlet Hotel
##  6 1                                                     Swissotel The Stamford~
##  7 1                                                     Amber Hotel Katong
##  8 1                                                     Grand Park City Hall
##  9 1                                                     Hotel 81 - Geylang
## 10 1                                                     Hotel 81 - Gold
## # i 30 more rows
```



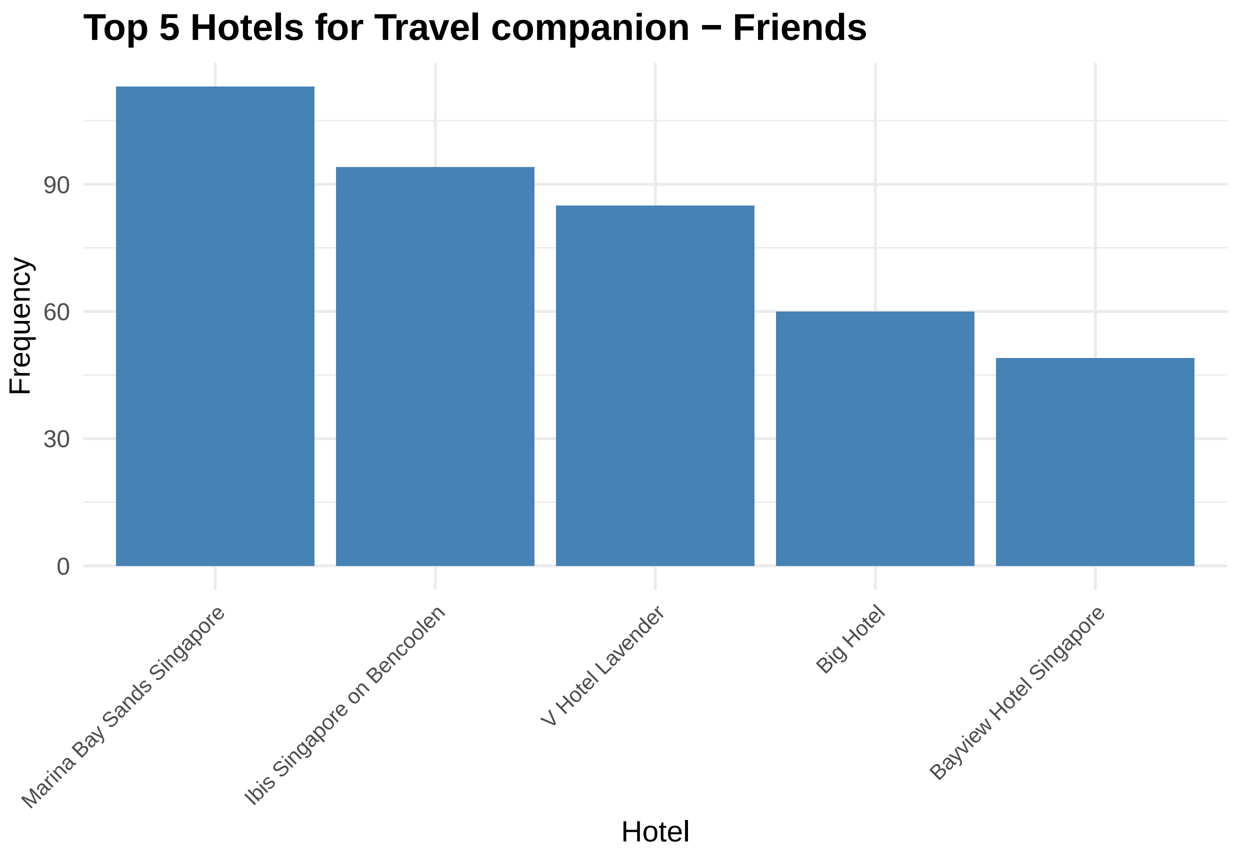
```
## # A tibble: 645 x 2
##    ‘Travel companion - Siblings‘ Hotel_Name
##    <chr>                         <chr>
##  1 1                             Fairmont, Singapore
##  2 1                             Four Seasons Hotel
##  3 1                             Crowne Plaza Changi Airport
##  4 1                             Other Hotels
##  5 1                             Hotel 81 - Classic
##  6 1                             M Hotel Singapore
##  7 1                             Hotel Chancellor @ Orchard
##  8 1                             W Singapore Sentosa Cove Hotel
##  9 1                             Marina Bay Sands Singapore
## 10 1                             Orchid Hotel
## # i 635 more rows
```



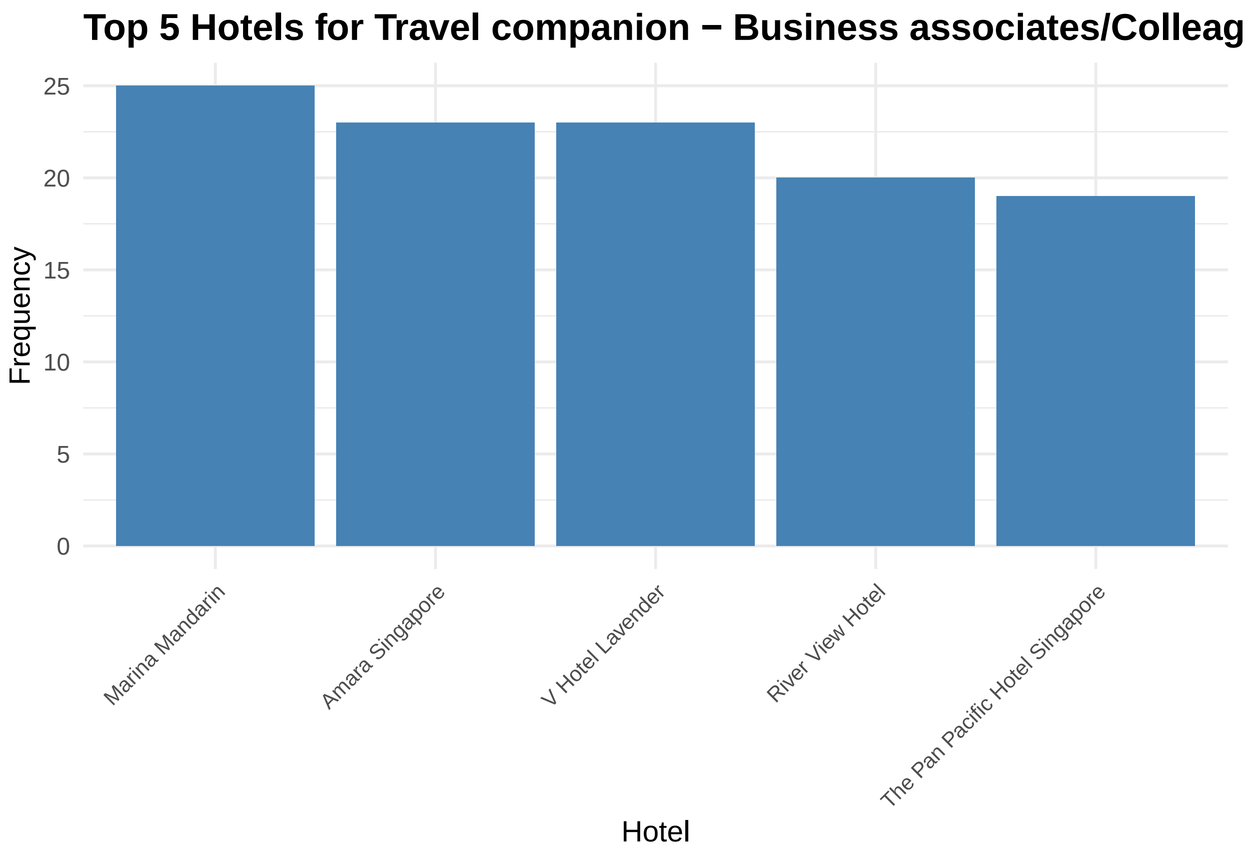
```
## # A tibble: 309 x 2
##    ‘Travel companion - Other relatives‘ Hotel_Name
##    <chr>                                <chr>
##  1 1                                    V Hotel Lavender
##  2 1                                    Hotel Chancellor @ Orchard
##  3 1                                    Marina Bay Sands Singapore
##  4 1                                    Hotel Re! @ Pearl’s Hill
##  5 1                                    Champion Hotel
##  6 1                                    Park View Hotel
##  7 1                                    Concorde Hotel Singapore
##  8 1                                    Hotel Michael
##  9 1                                    M Hotel Singapore
## 10 1                                    Hotel 81 - Hollywood
## # i 299 more rows
```



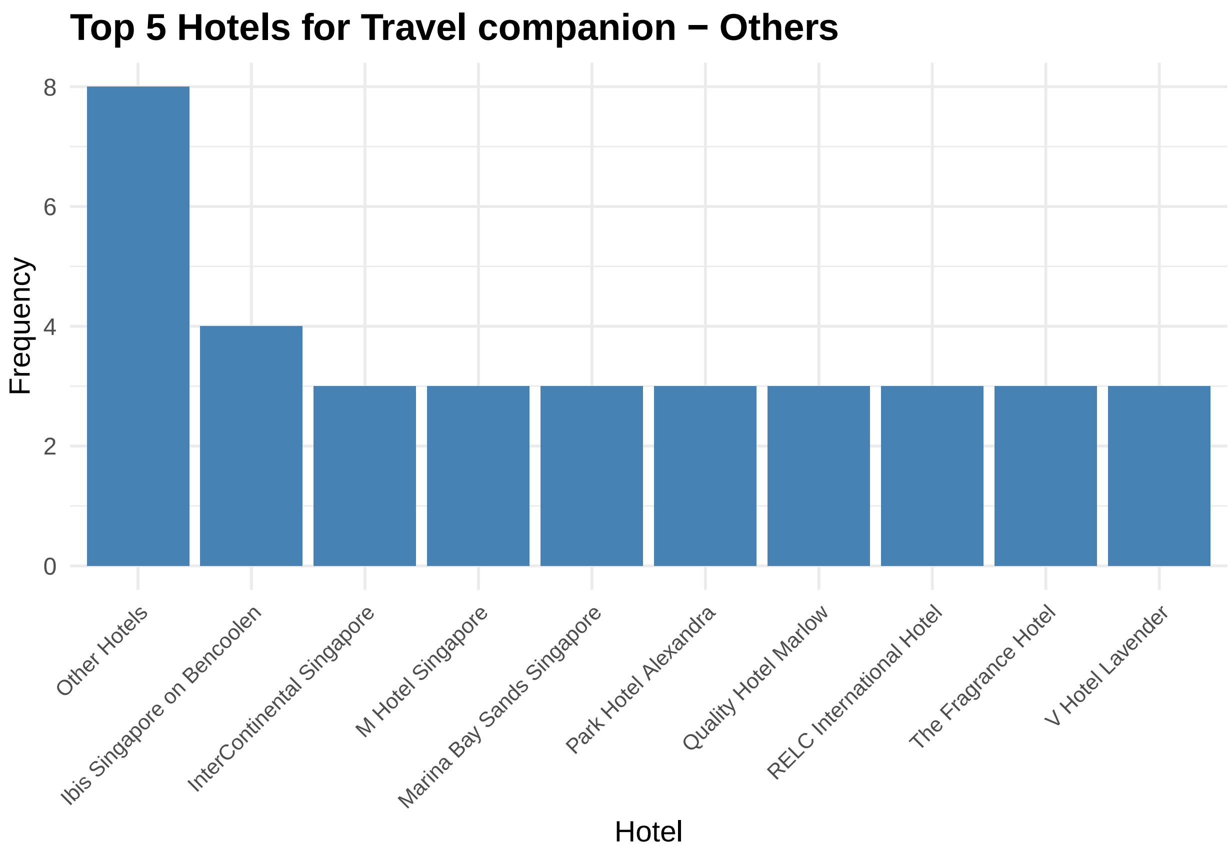
```
## # A tibble: 2,611 x 2
##    ‘Travel companion - Friends‘ Hotel_Name
##    <chr>                        <chr>
##  1 1                            5footway.inn Project (Not Specified)
##  2 1                            Swissotel Merchant Court,Singapore
##  3 1                            Swissotel Merchant Court,Singapore
##  4 1                            Fragrance Hotel - Ruby
##  5 1                            Rendezvous Hotel Singapore
##  6 1                            Marrison Hotel
##  7 1                            Hotel 81 - Classic
##  8 1                            York Hotel
##  9 1                            Robertson Quay Hotel
## 10 1                            Park Regis Singapore
## # i 2,601 more rows
```



```
## # A tibble: 756 x 2
##    `Travel companion - Business associates/Colleagues` Hotel_Name
##    <chr>                                               <chr>
##  1 1                                                   River View Hotel
##  2 1                                                   Parkroyal on Beach Road,~
##  3 1                                                   Parkroyal on Beach Road,~
##  4 1                                                   Bliss Hotel Singapore
##  5 1                                                   Hotel Clover (Not Specif~
##  6 1                                                   Orchid Hotel
##  7 1                                                   RELC International Hotel
##  8 1                                                   Carlton Hotel
##  9 1                                                   Grand Mercure Roxy Hotel
## 10 1                                                   Genting Hotel Jurong
## # i 746 more rows
```



```
## # A tibble: 95 x 2
##    ‘Travel companion - Others‘ Hotel_Name
##    <chr>                       <chr>
##  1 1                           InterContinental Singapore
##  2 1                           M Hotel Singapore
##  3 1                           Fragrance Hotel - Kovan
##  4 1                           M Hotel Singapore
##  5 1                           Marina Bay Sands Singapore
##  6 1                           Fortuna Hotel
##  7 1                           Hotel 81 (Not Specified)
##  8 1                           RELC International Hotel
##  9 1                           The Regent Singapore
## 10 1                           Other Hotels
## # i 85 more rows
```

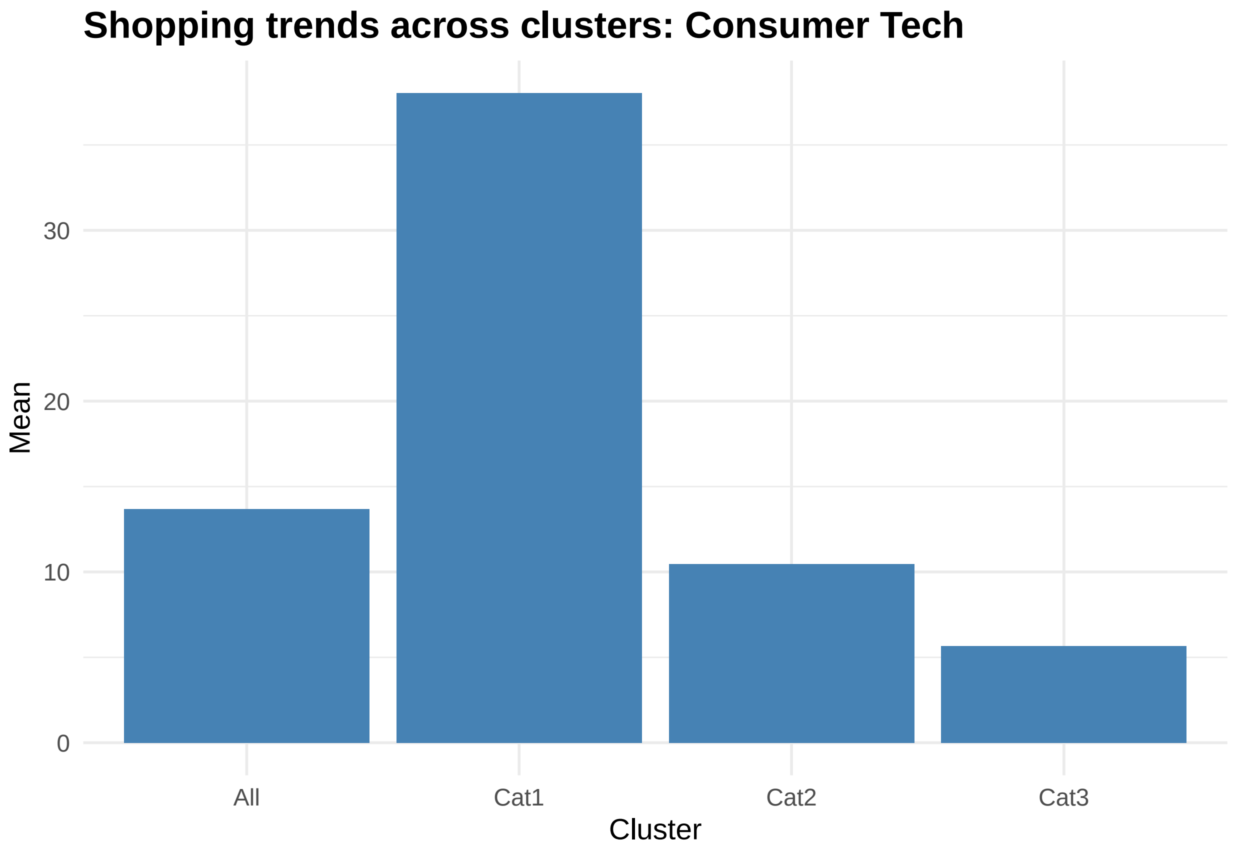


### K-Modes Analysis

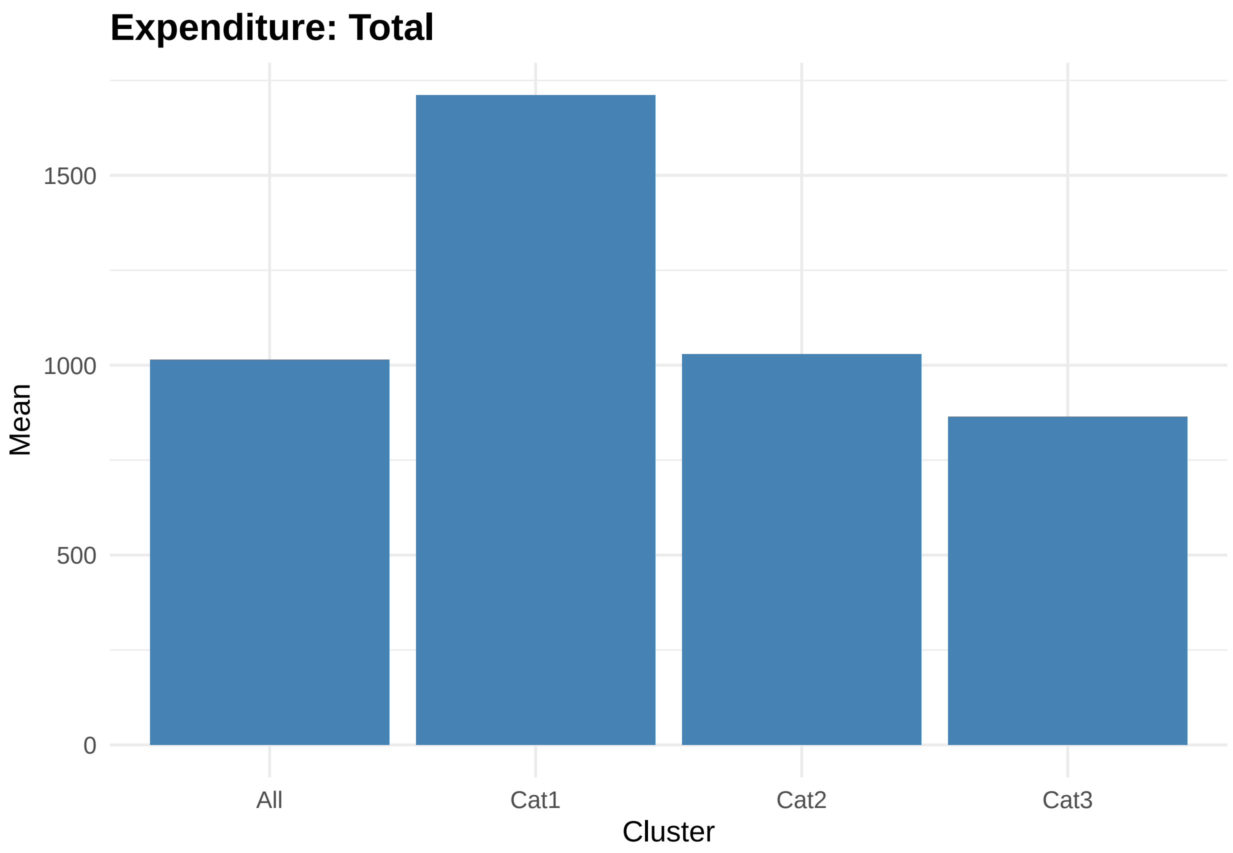
```
# All statistics
row0 = data %>% summarise(across(where(is.numeric), mean))
# Male Married Businessmen travelling alone
row1 = data %>% filter(`Purpose of Visit` == "Business + Accompanying Pax", `Gender` == "Male", `Marital
# Married Couples travelling together for Leisure
row2 = data %>% filter(`Purpose of Visit` == "Leisure", `Marital Status` == "Married", `Travel companion
# Single Females travelling with others for Leisure
row3 = data %>% filter(`Purpose of Visit` == "Leisure", `Gender` == "Female", `Marital Status` == "Sing

# Combine Statistics
cluster_statistics = bind_rows(row0,row1,row2,row3) %>% dplyr::select(-c(case,year,month,Weights_QTR)) %

cluster_statistics
```

```
## # A tibble: 4 x 24
##   Group `Length of Stay` shopping_fash shopping_jewllery shopping_watches
##   <chr>            <dbl>         <dbl>             <dbl>            <dbl>
## 1 All               3.42         107.              13.0              9.92
## 2 Cat1              2.77          54.3             16.8              6.96
## 3 Cat2              3.69         123.              17.0             12.6
## 4 Cat3              3.46         133.               6.96             7.78
## # i 19 more variables: shopping_wellness <dbl>, shopping_food <dbl>,
## #   shopping_gifts <dbl>, shopping_consumertech <dbl>, shopping_antiques <dbl>,
## #   shopping_others <dbl>, shopping_any <dbl>, totacc <dbl>, totfnb <dbl>,
## #   tottran <dbl>, totbiz <dbl>, totedu <dbl>, totmedi <dbl>, tototh <dbl>,
## #   totshopping_rep <dbl>, tot.exp <dbl>, AirEntry <dbl>, LandEntry <dbl>,
## #   SeaEntry <dbl>
```

```
# Shopping Preference Analysis
cluster_statistics_shopping = cluster_statistics %>% dplyr::select(Group, starts_with("shopping"))

cluster_statistics_shopping %>% ggplot(aes(x = Group, y = shopping_consumertech)) + geom_col(fill = "st
  labs(title = "Shopping trends across clusters: Consumer Tech", x = "Cluster", y = "Mean") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```



```
# Expenditure Analysis
cluster_statistics_expenditure = cluster_statistics %>% dplyr::select(Group, starts_with("tot"))

cluster_statistics_expenditure %>% ggplot(aes(x = Group, y = tot.exp)) + geom_col(fill = "steelblue") +
  labs(title = "Expenditure: Total", x = "Cluster", y = "Mean") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, face = "bold"))
```


Expenditure: Total
Mean
1500
1000
500
0
All
Cat1
Cat2
Cat3
Cluster