Sustainable Tourism: Data Analysis of the Spanish Tourism Industry (1995-2022)

Name: Kwadwo ('Ko-jo') Gyampo. Email: fkgyampo@gmail.com

August, 2024

1.0 Introduction

According to United Nations World Tourism Organisation (UNWTO), in 1950 there were 25 million international tourism arrivals, which increased to 1.3 billion in 2017. The international tourism sector is expected to grow by 3.3% annually, until 2030, representing an expected 1.8 billion tourists crossing international borders.

It is projected, as the Human Development Index (HDI) improves across the world resulting in increased incomes per capita or per household, seasonal peaks in international travel may become a mainstay of the global tourism economy. However, the growth of tourism has led to concerns of mass travel leading to overtourism at selected travel destinations in Europe causing a schism between local residents and tourists.

In Spain, several protests by local residents and activist groups were observed at favorite tourism destinations particularly in the cities of Barcelona, Malaga and Seville as well as the Balearic Islands made up of the Canary Islands, Formentera, Majorca, Menorca and Ibiza.

The general concerns of protesting groups and local residents were hinged on the cascading effects of mass tourism and overtourism resulting in an increased cost of living and a reduced quality of life for residents.

The UNWTO asserts that in 2023, "there were over 85 million tourists in Spain compared to a population of 47 million", further adding impetus to the concerns of overtourism and bringing into focus the carrying capacity of popular tourism destinations in the European country.

The goal of the data analysis of Spanish Tourism spanning the years 1995 to 2022, is to investigate the key concerns of overtourism by analysing key inbound travel metrics.

1.1 Research Objectives

The objectives of the data analysis project are as follows: a. Assess the trend of inbound travel visitors to Spain over 25 years.

b. Investigate the regions of inbound travel to Spain spanning 25 years.

c. Determine the travel expenditure and travel purpose for inbound travelers.

d. Assess the occupancy rates and availability of accommodation units for inbound visitors.

e. Establish the summary statistical metrics for inbound tourists.

f. Outline the employment rates for direct and downstream hospitality industries engaged in Spanish tourism.

1.2 Research Question

The main research question with respect to the data analysis is as follows: a. Is Overtourism a fact or a myth in the Spanish tourism industry ?

1.3 Methodology

The methodology for the analysis was carried out as follows: a. The database for UNWTO which serves as a renowned global travel repository for UN registered nations was searched for credible and verifiable data on Spain.

b. The global dataset publicly available on the unwto website (www.unwto.org) was downloaded.

c. Further, all specific sub-datasets relevant to Spain were extricated from the global repository.

d. To ensure data integrity, the sub-datasets were renamed and aligned to the file names of the global dataset.

e. The datasets were pre-processed in MS excel to remove extra spaces (rows and columns) and irrelevant symbols without altering the cell contents.

f. The cleaned and processed files were renamed as secondary files, resaved as comma separated value (CSV) files and exported to RStudio.

g. Imported CSV files into RStudio, were further wrangled and transformed with relevant codes and syntaxes prior to analyses.

h. Final data cleaning, transformation and all visualizations were carried out in RStudio in line with the research objectives.

i. Finally, a report was knitted in Rmarkdown to display all codes, visualizations and conclusion of the study.

1.4 Discussion - Exploratory Data Analysis

A preliminary investigative analysis of the dataframes, generated after cleaning, transformation and wrangling of the data files, was carried out to glean preliminary statistical trends.

The Exploratory Data Analysis (EDA) focused on statistical summaries, percentage variations, cumulative summations, means as well as peaks (maximum) and troughs (minimum) values of trends. Percentage distributions for all datasets were conducted to observe trends and glean important observations.

1.5 Conclusion

The following conclusions were drawn: a. Data analysis of the Spanish tourism industry displays a water shed moment for the Spanish tourism industry in 1999. The analysis indicates the pivotal year for tourism (1999) experienced a significant growth for accommodation units especially for the number of bed-places. The number of bed-places sequentially experienced a robust year-on-year growth from 1999 and peaked in 2022 at 2 million units.

b. Despite the tourism down-turn of 2020 possibly attributed to the Covid-19 travel shutdowns, the sharp rise in occupancy rates in 2021 to 2022 attests to the allure and attractiveness of Spanish holiday destinations to inbound tourist arrivals.

c. Data analysis of inbound tourism arrivals for the period under review, displayed a chronological year-on-year growth for groups of visitor arrivals.

d. Analysis of regional tourism arrivals to Spain showed an overwhelming patronage of Spanish tourism destinations by arrivals from Europe. It is fair to assume that travel policies including a unified European Schengen visa programme facilitates a fluid and seamless intra-European travel.

e. Tourists contributed more than 80million dollars to the Spanish economy in 2018.

f. Most tourists surveyed during the period under review indicated two main travel purposes: business and personal. A majority of tourists indicated the purpose of travel as personal.

g. It was observed that the food and beverage industry generated about 51% of total employment in tourism. This was followed by the passenger transportation and other tourism industries. In third place for tourism employment by industry sub-sector was accommodation services and hotels.

4.1 Overtourism – A fact or a myth The available data and metrics established were not sufficient to investigate the case of overtourism. To establish the existence of overtourism in the Spanish tourism industry, metrics such as described below must be established and validated by data analysis:

a. Mass tourism: This term refers to the large numbers of tourists visiting a destination, often resulting in overcrowding and strain on local resources.

b. Tourist saturation: This describes the point at which a destination becomes overwhelmed by the number of tourists, leading to negative impacts on the local environment, economy, and community.

c. Carrying capacity: This refers to the maximum number of tourists that a destination can accommodate without experiencing negative impacts on the environment, infrastructure, and local community.

d. Tourism overload: This phrase describes the situation in which a destination is unable to cope with the number of tourists, leading to congestion, overcrowding, and strain on local resources.

e. Unsustainable tourism: This term refers to tourism practices that are not environmentally, socially, or economically sustainable, often resulting in negative impacts on the local community and environment.

```
2.0 Data Analysis
Install packages for analysis of datasets
  * tidyverse
  * patchwork
  * viridis
  * RColorBrewer
  * ggmap
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
## Warning: package 'purrr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## Warning: package 'stringr' was built under R version 4.2.3
```

```
## Warning: package 'forcats' was built under R version 4.2.3
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
## Warning: package 'ggmap' was built under R version 4.2.3
```

```
## i Google's Terms of Service: <https://mapsplatform.google.com>
##   Stadia Maps' Terms of Service: <https://stadiamaps.com/terms-of-service/>
##   OpenStreetMap's Tile Usage Policy: <https://operations.osmfoundation.org/policies/tiles/>
## i Please cite ggmap if you use it! Use `citation("ggmap")` for details.
```

2.1 Load all data files

a. tourism industry (accommodation) csv file

b. inbound tourism arrivals csv file

c. inbound tourism by region csv file

d. inbound tourism expenditure csv file

e. inbound tourism purpose of travel csv file

f. tourism employment csv file

g. protests across spain csv file

3.0 Tourism Industry Accommodation Analysis

3.1 Load all data files

```r
#3.2 select relevant variables from loaded dataframe
spain.tourism.accommodation <- tourism.accommodation %>%
                                    select("Visitors.accommodation",c("X1995":"X2022"))

spain.tourism.accommodation
```

```r
#3.3 rename variables
spain.tourism.accomm.1 <- spain.tourism.accommodation %>%
                                  rename("1995" = "X1995",
                                         "1996" = "X1996",
                                         "1997" = "X1997",
                                         "1998" = "X1998",
                                         "1999" = "X1999",
                                         "2000" = "X2000",
                                         "2001" = "X2001",
                                         "2002" = "X2002",
                                         "2003" = "X2003",
                                         "2004" = "X2004",
                                         "2005" = "X2005",
                                         "2006" = "X2006",
                                         "2007" = "X2007",
                                         "2008" = "X2008",
                                         "2009" = "X2009",
                                         "2010" = "X2010",
                                         "2011" = "X2011",
```

```
                                        "2012" = "X2012",
                                        "2013" = "X2013",
                                        "2014" = "X2014",
                                        "2015" = "X2015",
                                        "2016" = "X2016",
                                        "2017" = "X2017",
                                        "2018" = "X2018",
                                        "2019" = "X2019",
                                        "2020" = "X2020",
                                        "2021" = "X2021",
                                        "2022" = "X2022")
spain.tourism.accomm.1
```

```
#3.4 Coerce years to numeric variables
spain.tourism.accomm.2 <- spain.tourism.accomm.1 %>%
                  mutate(1995 == as.numeric("1995","1996","1997","1998","1999","2000","2001",
                                            "2002","2003","2004","2005","2006","2007","2008",
                                            "2009","2010","2011","2012","2013","2014","2015",
                                            "2016","2017","2018","2019","2020","2021","2022"))

spain.tourism.accomm.2
```

```
#3.5 filter valid values from dataframe

spain.tourism.accomm.2 <- spain.tourism.accomm.2 %>%
                   filter(!is.na(Visitors.accommodation),na.rm=TRUE,
                          !is.na(1995),na.rm=TRUE,
                          !is.na(1996),na.rm=TRUE,
                          !is.na(1997),na.rm=TRUE,
                          !is.na(1998),na.rm=TRUE,
                          !is.na(1999),na.rm=TRUE,
                          !is.na(2000),na.rm=TRUE,
                          !is.na(2001),na.rm=TRUE,
                          !is.na(2002),na.rm=TRUE,
                          !is.na(2003),na.rm=TRUE,
                          !is.na(2004),na.rm=TRUE,
                          !is.na(2005),na.rm=TRUE,
                          !is.na(2006),na.rm=TRUE,
                          !is.na(2007),na.rm=TRUE,
                          !is.na(2008),na.rm=TRUE,
                          !is.na(2009),na.rm=TRUE,
                          !is.na(2010),na.rm=TRUE,
                          !is.na(2011),na.rm=TRUE,
                          !is.na(2012),na.rm=TRUE,
                          !is.na(2013),na.rm=TRUE,
                          !is.na(2014),na.rm=TRUE,
                          !is.na(2015),na.rm=TRUE,
                          !is.na(2016),na.rm=TRUE,
                          !is.na(2017),na.rm=TRUE,
                          !is.na(2018),na.rm=TRUE,
                          !is.na(2019),na.rm=TRUE,
                          !is.na(2020),na.rm=TRUE,
                          !is.na(2021),na.rm=TRUE,
                          !is.na(2022),na.rm=TRUE)
```

```
spain.tourism.accomm.2
```

```
#3.6 transform dataframe structure with pivot_longer function
spain.tourism.accomm.3 <- spain.tourism.accomm.2 %>%
                        pivot_longer(cols=`1995`:`2022`,
                           names_to = "year",
                           values_to = "total.visitors"
                          )
spain.tourism.accomm.3
```

```
#3.7 select relevant variables from transformed dataframe
spain.tourism.accomm.4 <- spain.tourism.accomm.3 %>%
                      select(Visitors.accommodation,year,total.visitors)
spain.tourism.accomm.4
```

```
#3.8 retrieve levels of Visitors.accommodation

  fct_count(spain.tourism.accomm.4$Visitors.accommodation)
```

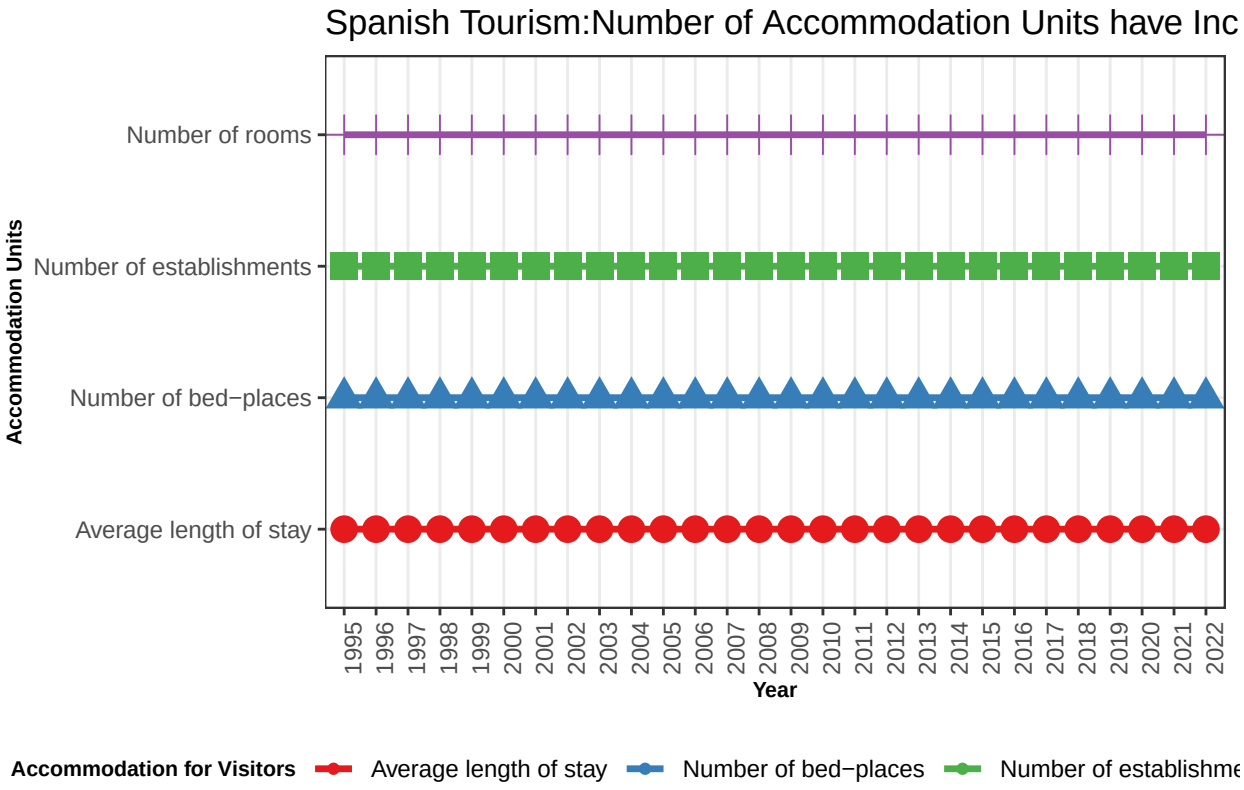
```
#3.9 plot a line graph for categorical variable visitors.accommodation

#3.9.1 group each level for visitors accommodation
      spain.tourism.accomm.4$group_var <- spain.tourism.accomm.4$Visitors.accommodation

     (spain.tourism.accomm.4.group_var <- spain.tourism.accomm.4$Visitors.accommodation)
```

```
#3.9.2 However, occupancy rates are percentages. remove these levels for homogenisation of variables

      d <- spain.tourism.accomm.4 %>%
             filter(!Visitors.accommodation %in% c("Occupancy rate / bed-places","Occupancy rate / roo
      d
```

```
#3.9.3 plot final graph - add arguments for final reporting

    d <- d %>%
        ggplot(aes(x = year, y = Visitors.accommodation, group = group_var, color = group_var,
                   linetype = group_var, size = group_var)) +
        geom_point(aes(group=group_var,size=0.5,shape=group_var))+
        geom_line(linetype="solid",size=1.2)+
        theme_bw() +
        theme(axis.text.x = element_text(angle = 90, hjust = 1),
              axis.title.x = element_text(size = 8, face = "bold"),
              axis.title.y = element_text(size = 8, face = "bold"),
              legend.title = element_text(size = 8, face = "bold"),
              legend.position = "bottom") +
        labs(title="Spanish Tourism:Number of Accommodation Units have Increased since 1998",
             x="Year",
             y="Accommodation Units",
             color = "Accommodation for Visitors",
             caption= "Source:www.unwto.org",
             legend.position="bottom")+
        scale_color_brewer(palette = "Set1") +
        scale_linetype_manual(values = c("solid", "solid", "solid","solid")) +
        #scale_size_manual(values = c(1.5, 1.5, 1.5,1.5))+
        guides(linetype = "none", size = "none",shape="none")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
d
```

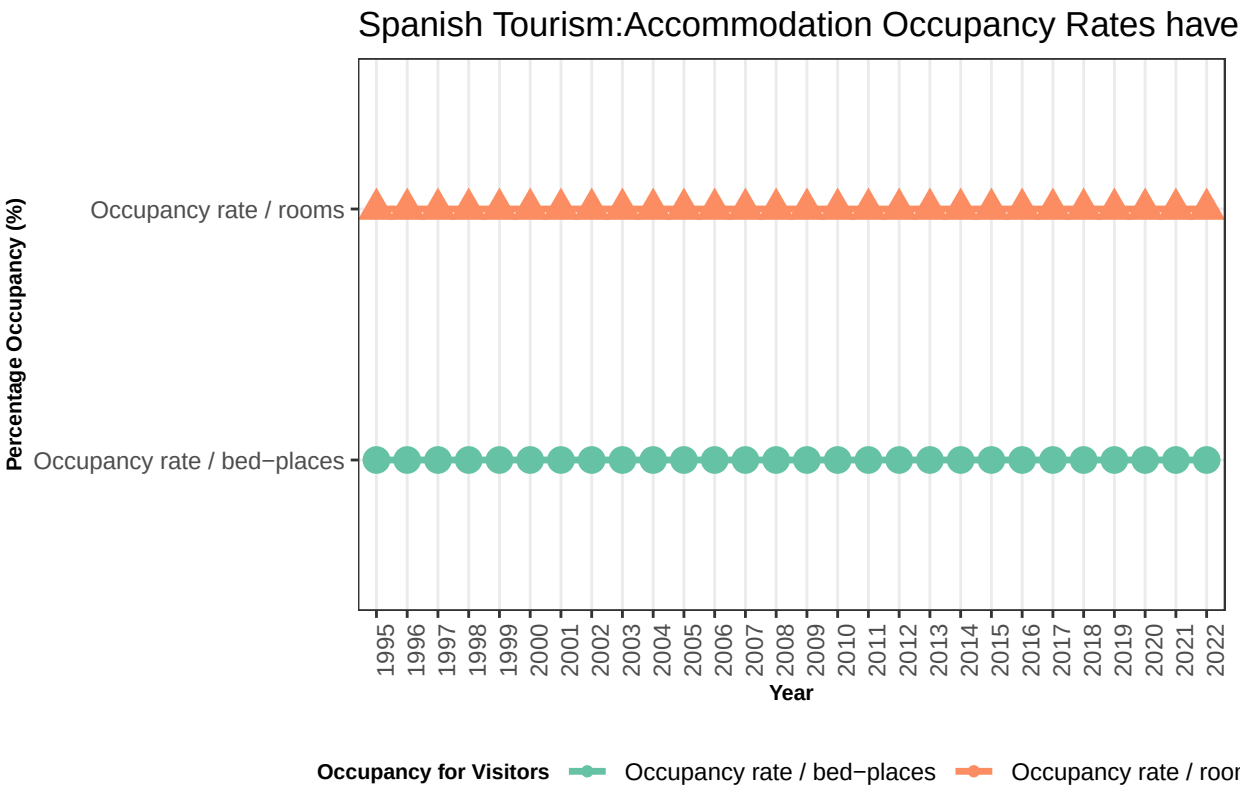


Source:www.unwto.org

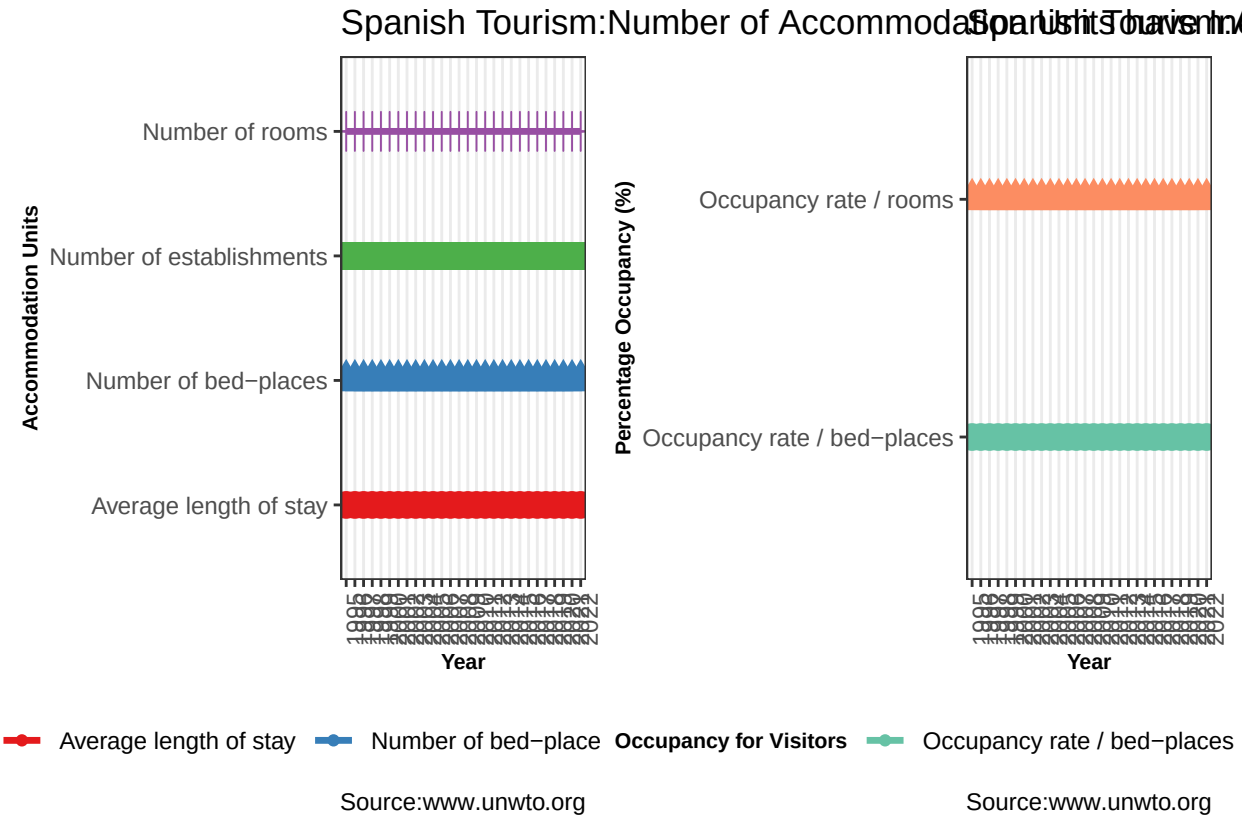
```
#3.10 plotting occupancy rates in percentages separately
      d.1 <- spain.tourism.accomm.4 %>%
                  filter(Visitors.accommodation %in% c("Occupancy rate / bed-places","Occupancy rate / r
      d.1
```

```
d.1 <- d.1 %>%
        ggplot(aes(x = year, y = Visitors.accommodation, group = group_var, color = group_var,
                   linetype = group_var, size = group_var)) +
        geom_point(aes(group=group_var,size=0.5,shape=group_var))+
        geom_line(linetype="solid",size=1.2)+
        theme_bw() +
        theme(axis.text.x = element_text(angle = 90, hjust = 1),
              axis.title.x = element_text(size = 8, face = "bold"),
              axis.title.y = element_text(size = 8, face = "bold"),
              legend.title = element_text(size = 8, face = "bold"),
              legend.position = "bottom") +
        labs(title="Spanish Tourism:Accommodation Occupancy Rates have Surged to Pre-covid Heights",
             x="Year",
             y="Percentage Occupancy (%)",
             color = "Occupancy for Visitors",
```

```
             caption= "Source:www.unwto.org") +
        scale_color_brewer(palette = "Set2") +
        scale_linetype_manual(values = c("solid", "solid")) +
        #scale_size_manual(values = c(1.5, 1.5))+
        guides(linetype = "none", size = "none",shape="none")
d.1
```




```
#3.11 Dashboard of d and d.1

  (d | d.1)+
    plot_annotation()+
    theme()
```



***

```
# *4.0 Inbound Arrivals Analysis*
```

```
# 4.1 select relevant variable columns
      inbound.spain.1 <- inbound.spain %>%
                                  select("Year",c("X1995":"X2022"))
      inbound.spain.1
```

```
# 4.2 tidy, clean and transform dataset
      inbound.spain.2 <- inbound.spain.1 %>%
                              rename("Arrivals" = "Year",
                                     "1995" = "X1995",
                                     "1996" = "X1996",
                                     "1997" = "X1997",
                                     "1998" = "X1998",
                                     "1999" = "X1999",
                                     "2000" = "X2000",
                                     "2001" = "X2001",
                                     "2002" = "X2002",
                                     "2003" = "X2003",
                                     "2004" = "X2004",
                                     "2005" = "X2005",
                                     "2006" = "X2006",
                                     "2007" = "X2007",
                                     "2008" = "X2008",
                                     "2009" = "X2009",
                                     "2010" = "X2010",
```

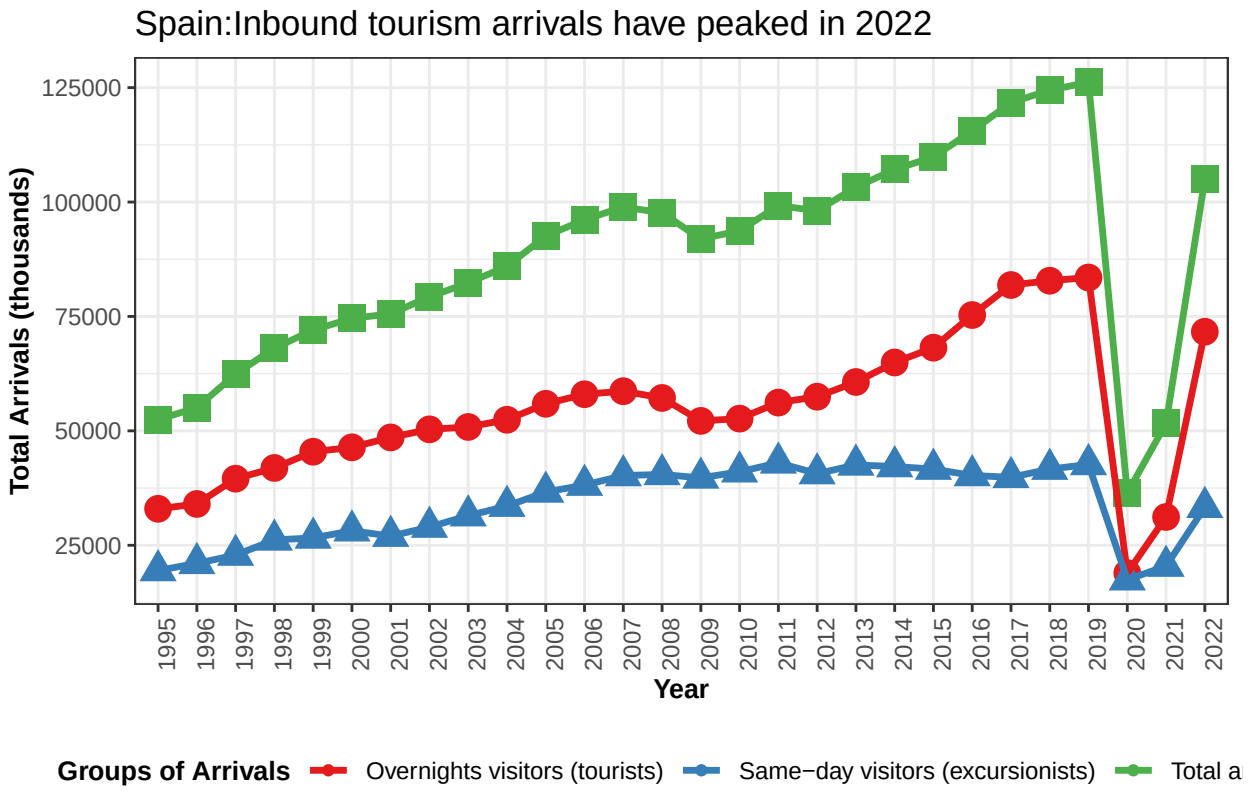
```
                                    "2011" = "X2011",
                                    "2012" = "X2012",
                                    "2013" = "X2013",
                                    "2014" = "X2014",
                                    "2015" = "X2015",
                                    "2016" = "X2016",
                                    "2017" = "X2017",
                                    "2018" = "X2018",
                                    "2019" = "X2019",
                                    "2020" = "X2020",
                                    "2021" = "X2021",
                                    "2022" = "X2022")
inbound.spain.2
```

```
# 4.3 select relevant variables and omit missing values
inbound.spain.3 <- inbound.spain.2 %>%
                        filter(!is.na("Arrivals"),na.rm=TRUE,
                               !is.na(1995),na.rm=TRUE,
                               !is.na(1996),na.rm=TRUE,
                               !is.na(1997),na.rm=TRUE,
                               !is.na(1998),na.rm=TRUE,
                               !is.na(1999),na.rm=TRUE,
                               !is.na(2000),na.rm=TRUE,
                               !is.na(2001),na.rm=TRUE,
                               !is.na(2002),na.rm=TRUE,
                               !is.na(2003),na.rm=TRUE,
                               !is.na(2004),na.rm=TRUE,
                               !is.na(2005),na.rm=TRUE,
                               !is.na(2006),na.rm=TRUE,
                               !is.na(2007),na.rm=TRUE,
                               !is.na(2008),na.rm=TRUE,
                               !is.na(2009),na.rm=TRUE,
                               !is.na(2010),na.rm=TRUE,
                               !is.na(2011),na.rm=TRUE,
                               !is.na(2012),na.rm=TRUE,
                               !is.na(2013),na.rm=TRUE,
                               !is.na(2014),na.rm=TRUE,
                               !is.na(2015),na.rm=TRUE,
                               !is.na(2016),na.rm=TRUE,
                               !is.na(2017),na.rm=TRUE,
                               !is.na(2018),na.rm=TRUE,
                               !is.na(2019),na.rm=TRUE,
                               !is.na(2020),na.rm=TRUE,
                               !is.na(2021),na.rm=TRUE,
                               !is.na(2022),na.rm=TRUE)
inbound.spain.3
```

```
# 4.4 transform structure of final dataframe with pivot_longer() function
inbound.spain.4 <- inbound.spain.3 %>%
                        pivot_longer(cols=`1995`:`2022`,
                                     names_to = "year",
                                     values_to = "total.visitors")
inbound.spain.4
```

```r
# 4.5 group levels of "Arrivals"
(inbound.spain.4$group_var <- inbound.spain.4$Arrivals)

    #or to group using mutate
    #inbound.spain.4 <- inbound.spain.4 %>%
                          #mutate(group_var = Arrivals)
    #inbound.spain.4
```

```r
# 4.6 final plot - line chart - inbound tourism group arrivals
inbound.spain.4 %>%
  ggplot(aes(x = year, y = total.visitors, group = group_var, color = group_var,
             linetype = group_var, size = group_var)) +
  geom_point(aes(group=group_var,shape=group_var,size=0.5))+
  geom_line(size=1.2) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        axis.title.x = element_text(size = 10, face = "bold"),
        axis.title.y = element_text(size = 10, face = "bold"),
        legend.title = element_text(size = 10, face = "bold"),
        legend.position = "bottom") +
  labs(title   = "Spain:Inbound tourism arrivals have peaked in 2022",
       x       = "Year",
       y       = "Total Arrivals (thousands)",
       color   = "Groups of Arrivals",
       caption = "Source:www.unwto.org.com") +
  scale_color_brewer(palette = "Set1") +
  scale_linetype_manual(values = c("solid", "solid", "solid")) +
  #scale_size_manual(values = c(1.5, 1.5, 1.5))+
  guides(linetype = "none", size = "none",shape="none")
```



Source:www.unwto.org.com

```
# *5.0 Inbound Arrivals by regions Analysis*
```

```
# 5.1 select column variables

inbound.tourism.regions.1 <- inbound.tourism.regions %>%
                                  select("year",c("X1995":"X2022"))
inbound.tourism.regions.1
```

```
#5.2 rename column variables
inbound.tourism.regions.2 <- inbound.tourism.regions.1 %>%
                                  rename("Regions" = "year",
                                         "1995" = "X1995",
                                         "1996" = "X1996",
                                         "1997" = "X1997",
                                         "1998" = "X1998",
                                         "1999" = "X1999",
                                         "2000" = "X2000",
                                         "2001" = "X2001",
                                         "2002" = "X2002",
                                         "2003" = "X2003",
                                         "2004" = "X2004",
                                         "2005" = "X2005",
                                         "2006" = "X2006",
                                         "2007" = "X2007",
                                         "2008" = "X2008",
```

```
                                        "2009" = "X2009",
                                        "2010" = "X2010",
                                        "2011" = "X2011",
                                        "2012" = "X2012",
                                        "2013" = "X2013",
                                        "2014" = "X2014",
                                        "2015" = "X2015",
                                        "2016" = "X2016",
                                        "2017" = "X2017",
                                        "2018" = "X2018",
                                        "2019" = "X2019",
                                        "2020" = "X2020",
                                        "2021" = "X2021",
                                        "2022" = "X2022")
inbound.tourism.regions.2
```

```
#5.3 select relevant variables
inbound.tourism.regions.3 <- inbound.tourism.regions.2 %>%
                                    filter(!is.na(Regions),na.rm=TRUE,
                                           !is.na(1995),na.rm=TRUE,
                                           !is.na(1996),na.rm=TRUE,
                                           !is.na(1997),na.rm=TRUE,
                                           !is.na(1998),na.rm=TRUE,
                                           !is.na(1999),na.rm=TRUE,
                                           !is.na(2000),na.rm=TRUE,
                                           !is.na(2001),na.rm=TRUE,
                                           !is.na(2002),na.rm=TRUE,
                                           !is.na(2003),na.rm=TRUE,
                                           !is.na(2004),na.rm=TRUE,
                                           !is.na(2005),na.rm=TRUE,
                                           !is.na(2006),na.rm=TRUE,
                                           !is.na(2007),na.rm=TRUE,
                                           !is.na(2008),na.rm=TRUE,
                                           !is.na(2009),na.rm=TRUE,
                                           !is.na(2010),na.rm=TRUE,
                                           !is.na(2011),na.rm=TRUE,
                                           !is.na(2012),na.rm=TRUE,
                                           !is.na(2013),na.rm=TRUE,
                                           !is.na(2014),na.rm=TRUE,
                                           !is.na(2015),na.rm=TRUE,
                                           !is.na(2016),na.rm=TRUE,
                                           !is.na(2017),na.rm=TRUE,
                                           !is.na(2018),na.rm=TRUE,
                                           !is.na(2019),na.rm=TRUE,
                                           !is.na(2020),na.rm=TRUE,
                                           !is.na(2021),na.rm=TRUE,
                                           !is.na(2022),na.rm=TRUE)
inbound.tourism.regions.3
```

```
#5.4 restructure dataframe using pivot_longer function
inbound.tourism.regions.4 <- inbound.tourism.regions.3 %>%
                                     pivot_longer(cols=`1995`:`2022`,
                                                  names_to = "Years",
                                                  values_to = "Regional Arrivals")
```

```r
inbound.tourism.regions.4
```

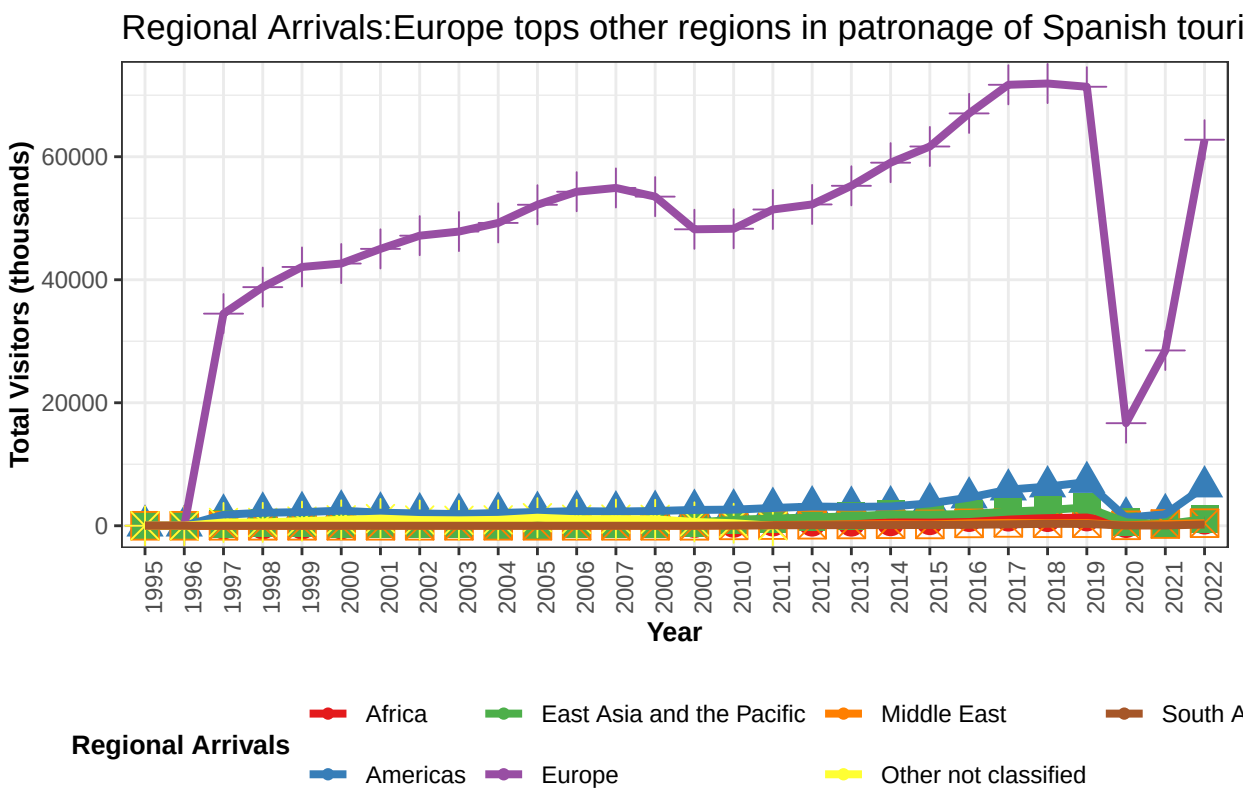
```r
#5.5 group character levels of Regions using group_var function

#5.5.1 using group_var

inbound.tourism.regions.4$group_var <- inbound.tourism.regions.4$Regions
```

```r
#5.6 plot - regional distribution of inbound tourism arrivals
inbound.tourism.regions.plota <- inbound.tourism.regions.4 %>%
                                      filter(!`Regional Arrivals` %in% c("Total")) %>%
                                      filter(!group_var %in% c("Total")) %>%
                                      ggplot(aes(x=Years,y=`Regional Arrivals`,group=group_var,color=gr
                                                 linetype = group_var, size = group_var)) +
                                      geom_point(aes(group=group_var,shape=group_var,size=0.5))+
                                      geom_line(size=1.4) +
                                      theme_bw() +
                                      theme(axis.text.x = element_text(angle = 90, hjust = 1),
                                            axis.title.x = element_text(size = 10, face = "bold"),
                                            axis.title.y = element_text(size = 10, face = "bold"),
                                            legend.title = element_text(size = 10, face = "bold"),
                                            legend.position = "bottom") +
                                      labs(title   = "Regional Arrivals:Europe tops other regions in pa
                                            x      = "Year",
                                            y      = "Total Visitors (thousands)",
                                           color   = "Regional Arrivals",
                                           caption = "Source:www.unwto.org") +
                                      scale_color_brewer(palette = "Set1") +
                                      scale_linetype_manual(values = c("solid", "solid", "solid","solid
                                      #scale_size_manual(values = c(1.5, 1.5, 1.5,1.5,1.5,1.5,1.5))+
                                      guides(linetype = "none", size = "none",shape="none")

inbound.tourism.regions.plota
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values because more
## than 6 becomes difficult to discriminate
## i you have requested 7 values. Consider specifying shapes manually if you need
##   that many have them.
```

```
## Warning: Removed 47 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 11 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



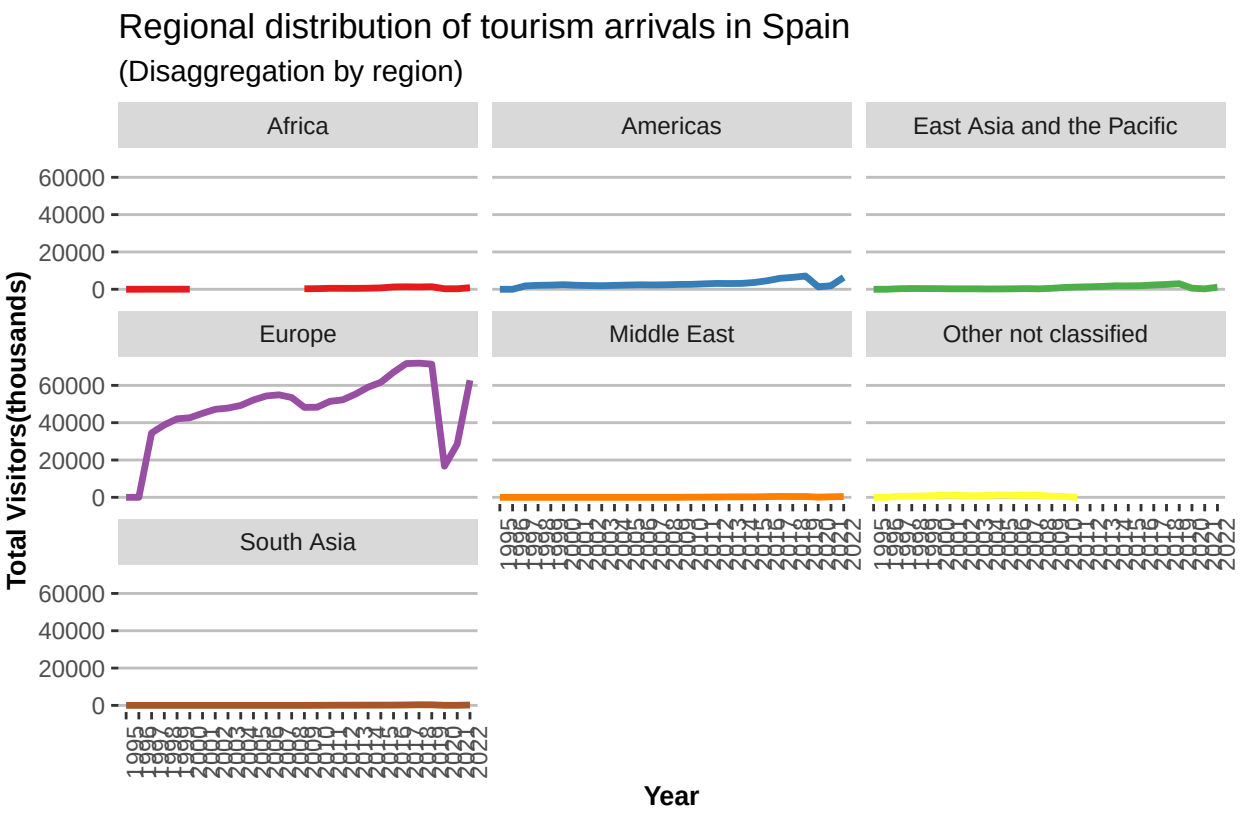
Source:www.unwto.org

```
#5.7 Regional distribution of tourism arrivals. Disaggregation by region
inbound.tourism.regions.plotb <- inbound.tourism.regions.4 %>%
                                 filter(!`Regional Arrivals` %in% c("Total")) %>%
                                 filter(!group_var %in% c("Total")) %>%
                                 ggplot(aes(x=Years,y=`Regional Arrivals`,group=group_var,color=grou
                                            linetype = group_var, size = group_var)) +
                                 geom_line(size=1.2) +
                                 facet_wrap(~group_var)+
                                 theme(panel.grid.major.x = element_blank(),
                                       panel.grid.major.y = element_line(size = 0.5, linetype = "soli
                                       panel.background = element_blank(),
                                       legend.key = element_rect(colour = "black")) +
                                 theme(axis.text.x = element_text(angle = 90, hjust = 1),
                                       axis.title.x = element_text(size = 10, face = "bold"),
                                       axis.title.y = element_text(size = 10, face = "bold"),
                                       legend.title = element_text(size = 10, face = "bold"),
                                       legend.position = "right") +
                                 labs(title = "Regional distribution of tourism arrivals in Spain",
                                   subtitle = "(Disaggregation by region)",
                                          x  = "Year",
                                          y  = "Total Visitors(thousands)",
                                      color = "Regional Arrivals",
                                      caption = "Source:www.unwto.org") +
                                 scale_color_brewer(palette = "Set1") +
                                 scale_linetype_manual(values = c("solid", "solid", "solid","solid",
                                 scale_size_manual(values = c(1.5, 1.5, 1.5,1.5,1.5,1.5,1.5))+
                                 guides(linetype = "none", size = "none",shape = "none",color="none"
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
inbound.tourism.regions.plotb
```

```
## Warning: Removed 11 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



Source:www.unwto.org

***

***

6.0 Inbound Tourism Expenditure Analysis

```
#6.1 select relevant variables
inbound.expenditure.1 <- inbound.expenditure %>%
                                 select("Year",c("X1995":"X2022"))
inbound.expenditure.1
```

```
#6.2 rename variables
inbound.expenditure.2 <- inbound.expenditure.1 %>%
                                 rename("Tourism Expenditure"="Year",
                                        "1995" = "X1995",
                                        "1996" = "X1996",
                                        "1997" = "X1997",
                                        "1998" = "X1998",
```

```
                                        "1999" = "X1999",
                                        "2000" = "X2000",
                                        "2001" = "X2001",
                                        "2002" = "X2002",
                                        "2003" = "X2003",
                                        "2004" = "X2004",
                                        "2005" = "X2005",
                                        "2006" = "X2006",
                                        "2007" = "X2007",
                                        "2008" = "X2008",
                                        "2009" = "X2009",
                                        "2010" = "X2010",
                                        "2011" = "X2011",
                                        "2012" = "X2012",
                                        "2013" = "X2013",
                                        "2014" = "X2014",
                                        "2015" = "X2015",
                                        "2016" = "X2016",
                                        "2017" = "X2017",
                                        "2018" = "X2018",
                                        "2019" = "X2019",
                                        "2020" = "X2020",
                                        "2021" = "X2021",
                                        "2022" = "X2022")
inbound.expenditure.2
```

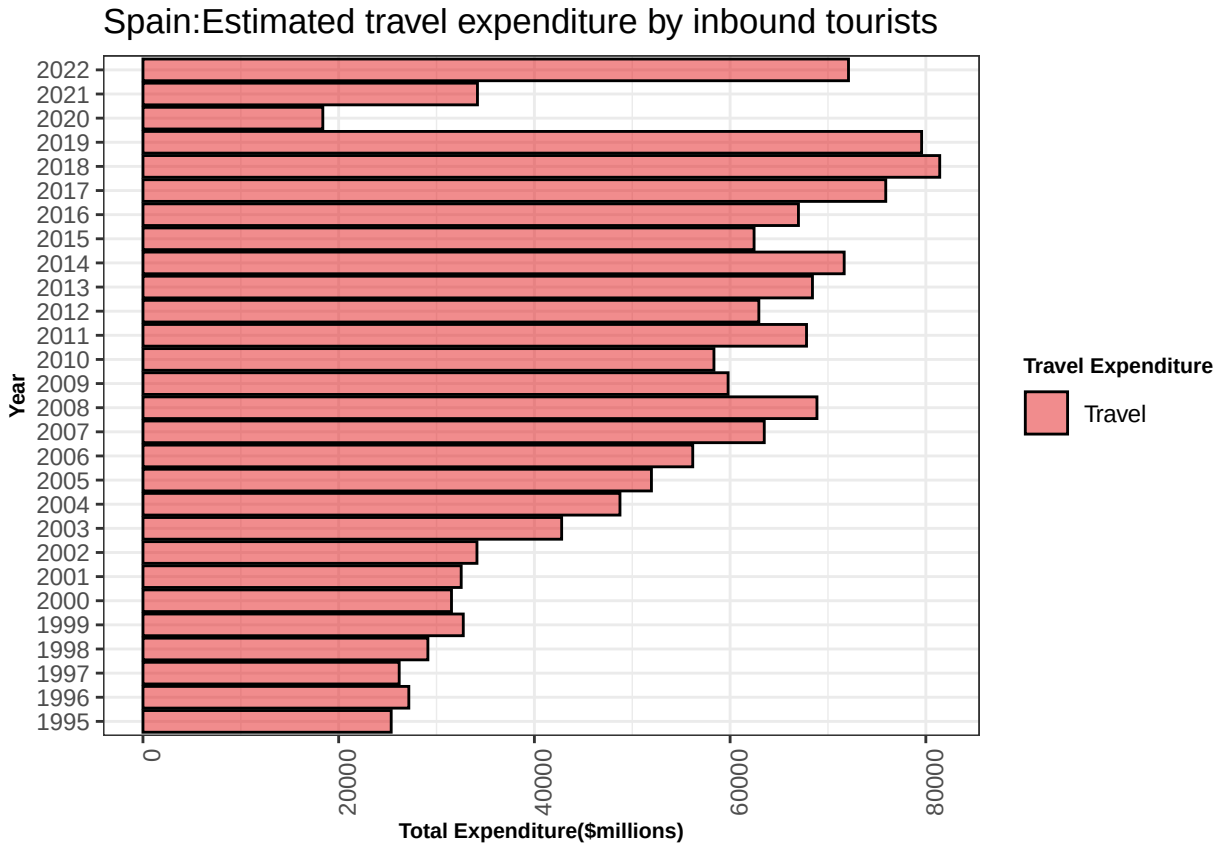
```
#6.3 transform df structure
inbound.expenditure.3 <- inbound.expenditure.2 %>%
                              pivot_longer(cols=`1995`:`2022`,
                                           names_to = "Years",
                                           values_to = "Total Expenditure")
inbound.expenditure.3
```

```
#6.4 group levels
inbound.expenditure.4 <- inbound.expenditure.3 %>%
                              mutate(group_var = `Tourism Expenditure`)
inbound.expenditure.4
```

```
#6.5 plot - Inbound expenditure in Spain
inbound.expenditure.5 <- inbound.expenditure.4 %>%
                              filter(!`Tourism Expenditure` %in% c("Tourism expenditure in the country
                              ggplot(aes(x=Years,y=`Total Expenditure`,group=group_var,color=group_var
                                         size=group_var,line = group_var)) +
                              geom_line(aes(color=group_var))+
                              theme(axis.text.x=element_text(angle=90,hjust=1))+
                              theme_bw()
inbound.expenditure.5
```

```
## Warning: Using size for a discrete variable is not advised.
```


```
#6.6 final plot - Inbound travel expenditure by tourists
inbound.expenditure.6 <- inbound.expenditure.4 %>%
                         filter(!`Tourism Expenditure` %in% c("Tourism expenditure in the country
                         ggplot(aes(x=Years, y=`Total Expenditure`, fill=group_var)) +
                         geom_bar(position="dodge", stat="identity",color="black",alpha=0.5) +
                         coord_flip()+
                         theme_bw() +
                         theme(axis.text.x  = element_text(angle = 90, hjust = 1),
                               axis.title.x = element_text(size = 8, face = "bold"),
                               axis.title.y = element_text(size = 8, face = "bold"),
                               legend.title = element_text(size = 8, face = "bold"),
                               legend.position = "right") +
                         labs(title = "Spain:Estimated travel expenditure by inbound tourists",
                              x     = "Year",
                              y     = "Total Expenditure($millions)",
                              fill  = "Travel Expenditure") +
                         scale_fill_brewer(palette = "Set1")
inbound.expenditure.6
```



7.0 Inbound Tourism Purpose of Travel Analysis

```
#7.1 rename variables
inbound.tourism.purpose.1 <- inbound.tourism.purpose %>%
                                    rename("Purpose"="X",
                                           "1995" = "X1995",
                                           "1996" = "X1996",
                                           "1997" = "X1997",
                                           "1998" = "X1998",
                                           "1999" = "X1999",
                                           "2000" = "X2000",
                                           "2001" = "X2001",
                                           "2002" = "X2002",
                                           "2003" = "X2003",
                                           "2004" = "X2004",
                                           "2005" = "X2005",
                                           "2006" = "X2006",
                                           "2007" = "X2007",
                                           "2008" = "X2008",
                                           "2009" = "X2009",
                                           "2010" = "X2010",
                                           "2011" = "X2011",
                                           "2012" = "X2012",
                                           "2013" = "X2013",
```

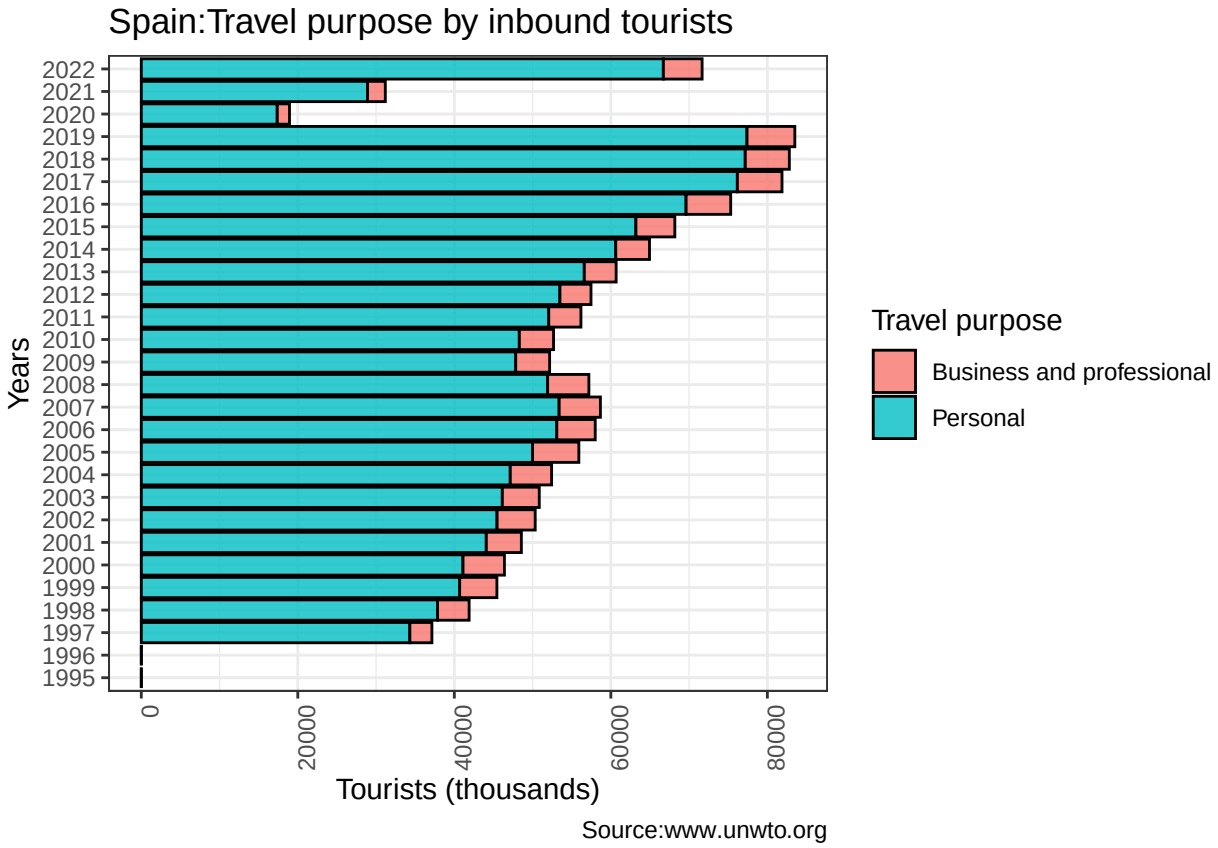
```
                                        "2014" = "X2014",
                                        "2015" = "X2015",
                                        "2016" = "X2016",
                                        "2017" = "X2017",
                                        "2018" = "X2018",
                                        "2019" = "X2019",
                                        "2020" = "X2020",
                                        "2021" = "X2021",
                                        "2022" = "X2022")
inbound.tourism.purpose.1
```

```
#7.2 select relevant variables
inbound.tourism.purpose.2 <- inbound.tourism.purpose.1 %>%
                                    select("Purpose",c("1995":"2022"))
inbound.tourism.purpose.2
```

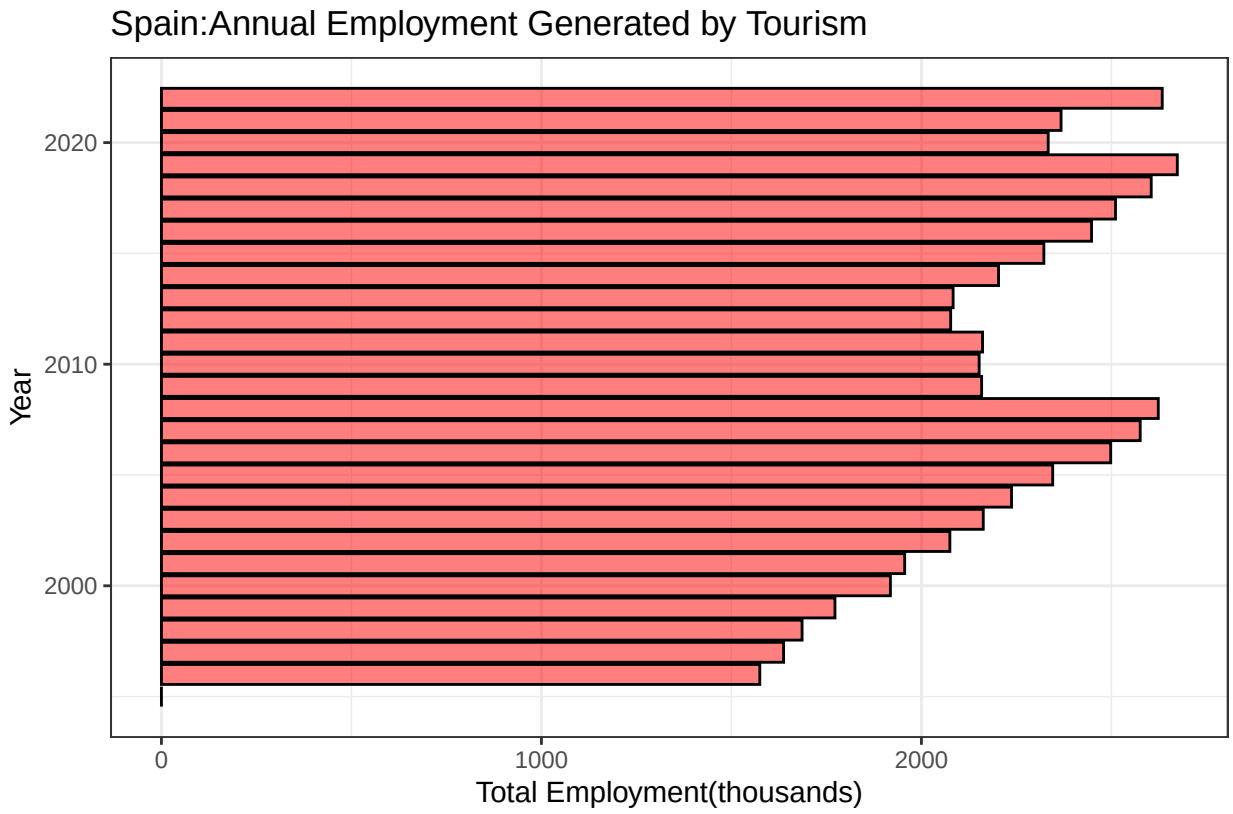
```
#7.3 restructure df with pivot_longer function
inbound.tourism.purpose.3 <- inbound.tourism.purpose.2 %>%
                                    pivot_longer(cols=`1995`:`2022`,
                                                 names_to = "Years",
                                                 values_to = "Travel purpose")
inbound.tourism.purpose.3
```

```
#7.4 group purpose
inbound.tourism.purpose.4 <- inbound.tourism.purpose.3 %>%
                                   mutate(group_var = Purpose)
inbound.tourism.purpose.4
```

```
#7.5 plot
inbound.tourism.purpose.5 <- inbound.tourism.purpose.4 %>%
                                   filter(!Purpose %in% c("Total")) %>%
                                   ggplot(aes(x=Years,y=`Travel purpose`,fill=group_var))+
                                   geom_bar(position="stack",stat="identity",alpha=0.8,color="black"
                                   coord_flip()+
                                   theme_bw()+
                                   theme(axis.text.x=element_text(angle=90,hjust=1))+
                                   labs(title = "Spain:Travel purpose by inbound tourists",
                                        x     = "Years",
                                        y     = "Tourists (thousands)",
                                        fill  = "Travel purpose",
                                        caption="Source:www.unwto.org")

inbound.tourism.purpose.5
```



8.0 Tourism Employment by Industries



Source:www.unwto.org

```
#8.2 percentage distribution by industry
# the dataframe is named 'employment.spain'
# Melt the dataframe to convert the percentage columns into a single column
employment.spain_melted <- reshape2::melt(employment.spain, id.vars = "Year",
                            variable.name = "Category",
                            value.name = "Percentage")
employment.spain_melted
```

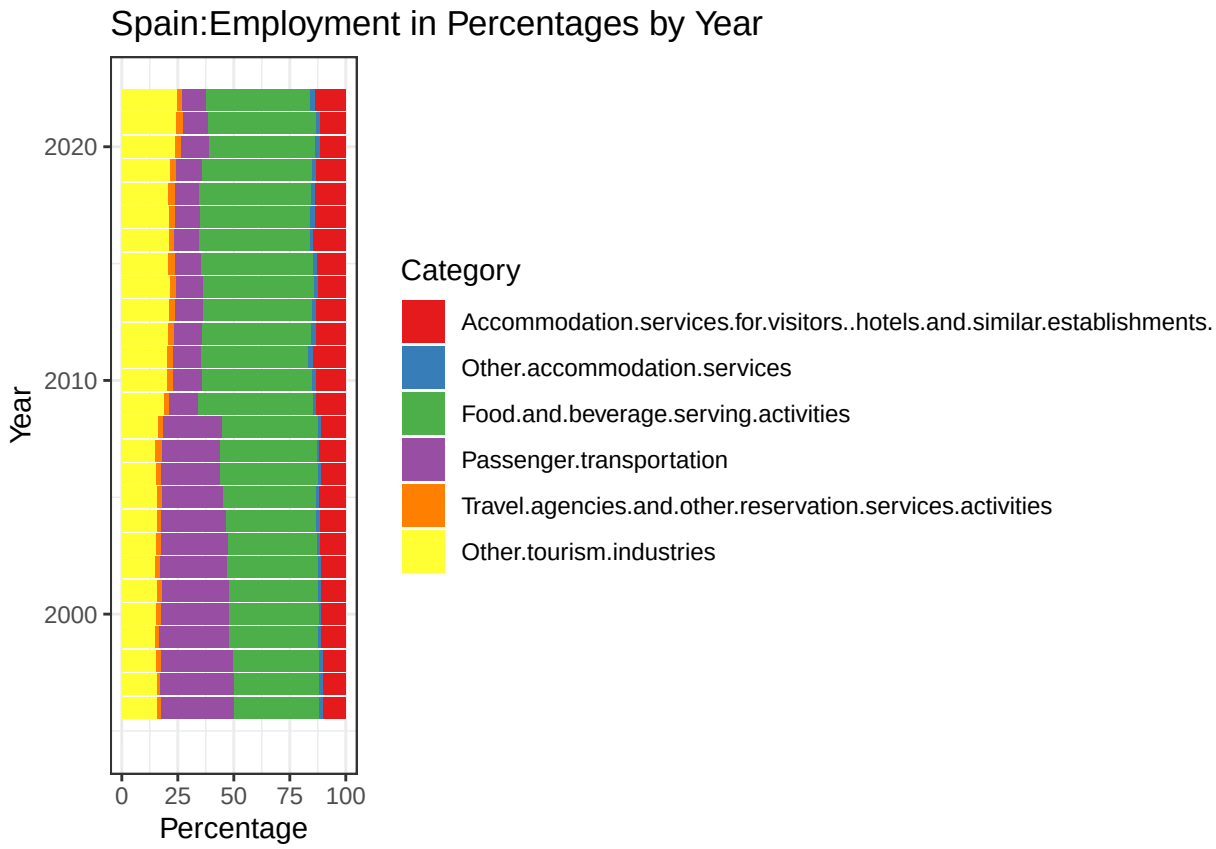
```
df.employment.spain_melted <- employment.spain_melted %>%
                                filter(!Category %in% c("Total")) %>%
                                group_by(Year,Category,Percentage) %>%
                                count(n=n()) %>%
                                arrange(desc(Percentage))
```

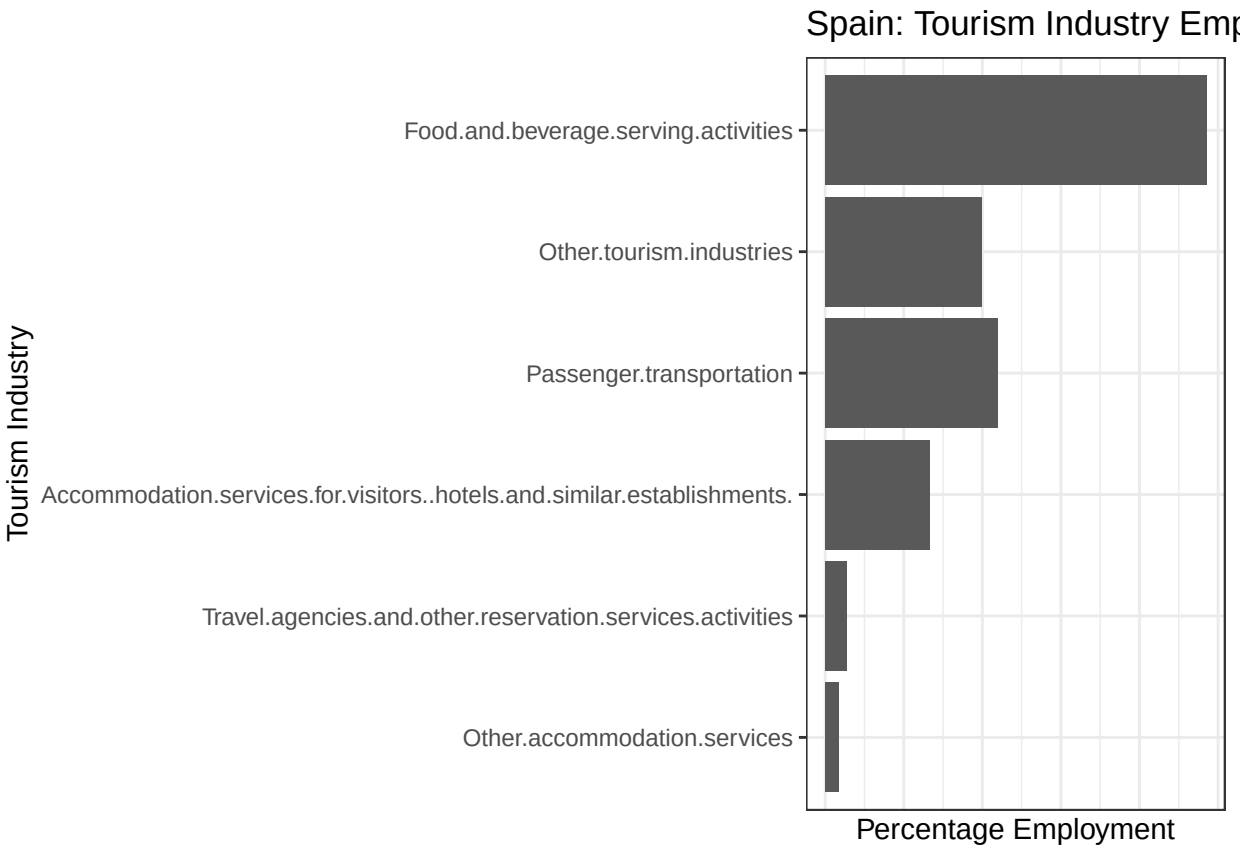
```
## Storing counts in `nn`, as `n` already present in input
## i Use `name = "new_name"` to pick a new name.
```

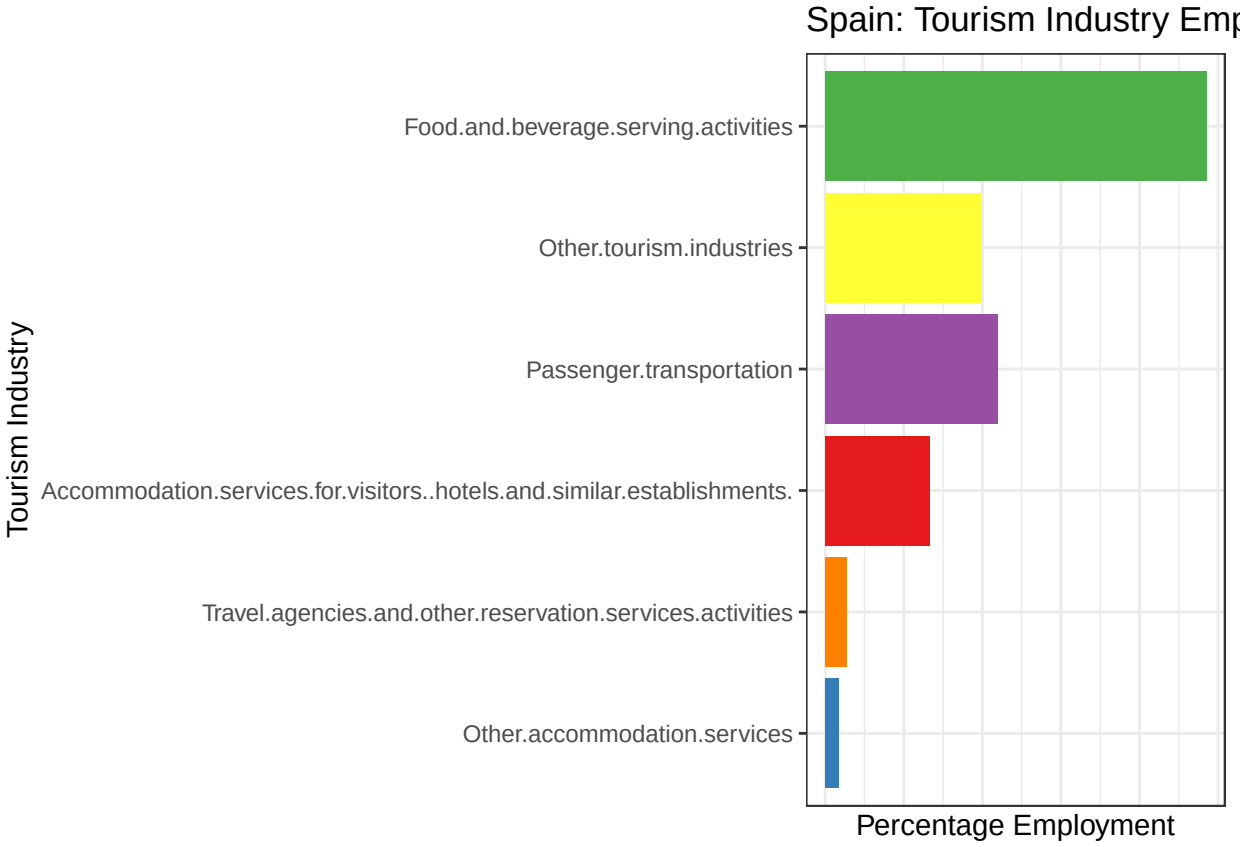
```
df.employment.spain_melted
```

```
# 8.3 Plot the bar graph using ggplot2
employment.spain_melted %>%
                        filter(!Category %in% c("Total")) %>%
                          ggplot(aes(x = Year, y = Percentage, fill = Category)) +
                            geom_bar(stat = "identity", position = "stack") +
                            coord_flip()+

                            labs(x = "Year",
                                 y = "Percentage",
```

```
     title = "Spain:Employment in Percentages by Year"
    ) +
theme(legend.position = "bottom")+
theme_bw()+
scale_fill_brewer(palette="Set1")
```

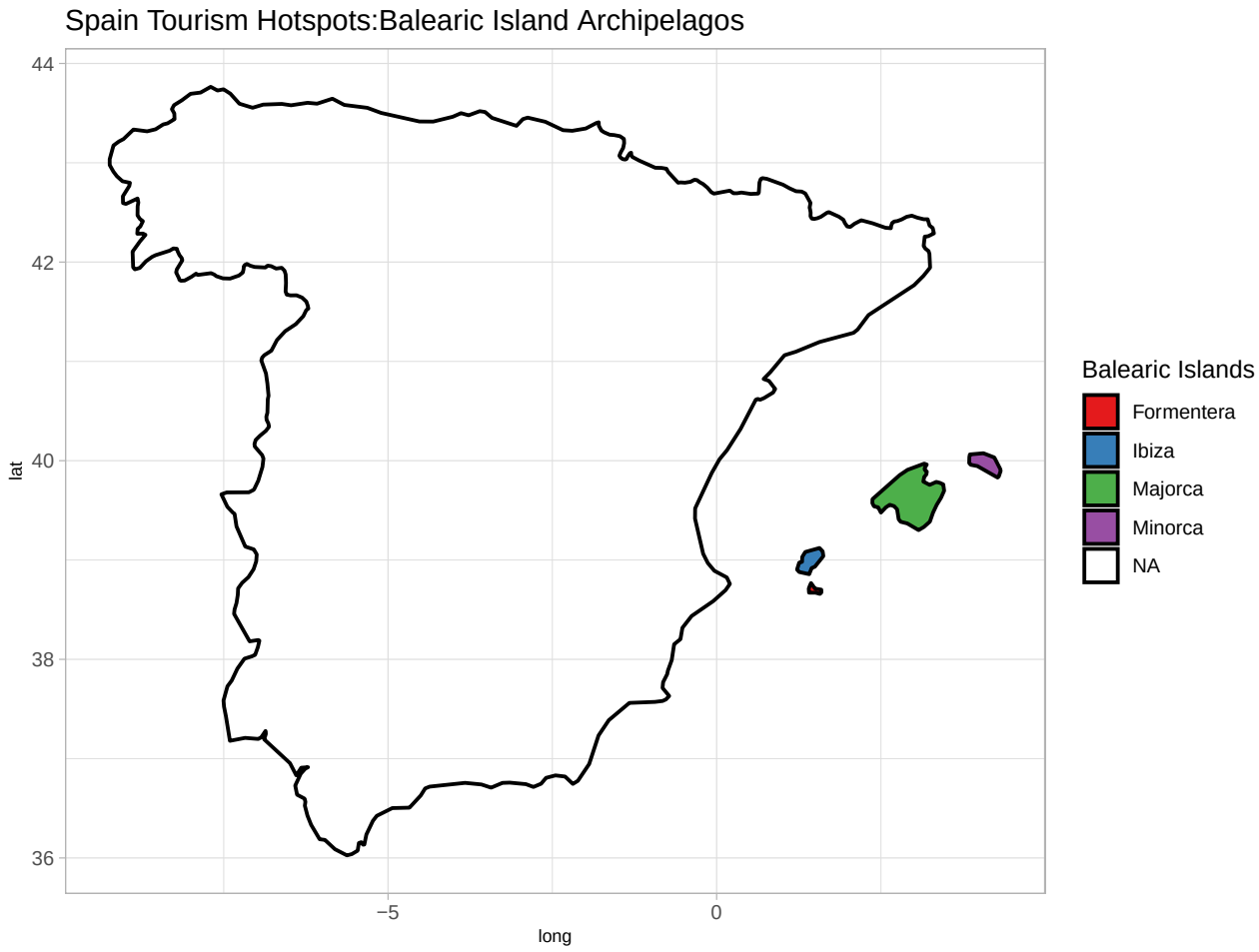



Spain: Tourism Industry Emp
Food.and.beverage.serving.activities
Other.tourism.industries
Passenger.transportation
Accommodation.services.for.visitors..hotels.and.similar.establishments.
Travel.agencies.and.other.reservation.services.activities
Other.accommodation.services
Tourism Industry
Percentage Employment


Spain: Tourism Industry Emp
Food.and.beverage.serving.activities
Other.tourism.industries
Passenger.transportation
Accommodation.services.for.visitors..hotels.and.similar.establishments.
Travel.agencies.and.other.reservation.services.activities
Other.accommodation.services
Tourism Industry
Percentage Employment

```
#9.0  plot map of region of Spain
       #install.packages("ggmap")
       #library(ggmap)

       #install.packages("mapdata")
       #library(mapdata)
```

```
Spain <-  map_data("world",region=c("Spain"),subregion=c("Formentera","Ibiza","Majorca","Minorca"),compl


Spain.map <- ggplot(Spain, aes(x=long, y=lat, group=group, fill=subregion)) +
         geom_polygon(color="black",linewidth=0.8) +
         scale_fill_brewer(palette="Set1")+
         theme_light()+
         labs(title = "Spain Tourism Hotspots:Balearic Island Archipelagos",
              fill = "Balearic Islands")+
         theme(axis.title.x = element_text(size = 8))+
         theme(axis.title.y = element_text(size = 8))
Spain.map
```



```r
#protests

anti.tourism.protests <- protests %>%
                         ggplot(aes(x=fct_reorder(City,Number.of.Protesters),y=Number.of.Protesters))+
                         geom_bar(aes(fill=City),position="dodge",stat="identity")+
                         geom_text(aes(label=Number.of.Protesters), size=4, vjust=-0.5,hjust=0.5) +
                         scale_fill_brewer(palette="Set1")+
                         theme_bw()+
                         guides(fill="none")+
                         labs(title = "Anti-tourism protests across cities in Spain",
                              x     = "Cities",
                              y     = "Number of Protesters",
                            caption = "Source:www.wikipedia.org")
anti.tourism.protests
```