

Estudio del gasto y duración media de los viajes de los turistas extranjeros en distintas comunidades autónomas

Alejandro León Líndez¹, Adrian Lizzadro Pla², Marta Medina Muñiz³

¹ Máster en Ciencia de Datos; alelin@alumni.uv.es
² Máster en Ciencia de Datos; alizpla@alumni.uv.es
³ Máster en Ciencia de Datos; memuiz@alumni.uv.es
 * Correspondence: email@email.com; Tel.: +XX-000-00-0000.

Simple Summary: Resumen del trabajo

Abstract: abstract

Keywords: keyword 1; keyword 2; keyword 3 (list three to ten pertinent keywords specific to the article, yet reasonably common within the subject discipline.).

1. Introducción

2. Carga de librerías e importación del fichero

Antes de comenzar, eliminamos todas las variables guardadas.

```
rm(list=ls()) # Borrado de todas las variables
```

```
library(readr) # Librería para importación de datos
library(dplyr) # Librería para arreglo de datos
```

```
##
## Adjuntando el paquete: 'dplyr'
## The following objects are masked from 'package:stats':
##
## filter, lag
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
library(tidyr) # Librería para arreglo de datos
library(ggplot2) # Librería para gráficas
```

Importamos los datos

```
gastos<- read_delim("data/Gasto_turistas_internacionales_según_comunidad_paisresidid")

## Rows: 3072 Columns: 6
## -- Column specification -----
## Delimiter: ";"
## chr (5): País de residencia, Total Nacional y CCAA, Tipo de dato, Gastos y d...
## dbl (1): Periodo
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message
```

Citation: Estudio del gasto y duración media de los viajes de los turistas extranjeros en distintas comunidades autónomas. *Journal Not Specified* 2023, 1, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2024 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

3. Preparación de los datos

3.1. Transformación a tidydata

```
tipos_datos <- unique(gastos$`Gastos y duración media de los viajes`)
tipos_datos
```

```
## [1] "Gasto total" "Gasto medio por persona"
## [3] "Gasto medio diario por persona" "Duración media de los viajes"
```

```
gastos1 <- subset(gastos,gastos$`Gastos y duración media de los viajes`== tipos_da
colnames(gastos1)[colnames(gastos1)=="Total"]<- "Gasto_total"
gastos1<- subset(gastos1, select = -`Gastos y duración media de los viajes`)
```

```
gastos2 <- subset(gastos,gastos$`Gastos y duración media de los viajes`== tipos_da
colnames(gastos2)[colnames(gastos2)=="Total"]<- "Gasto_medio_persona"
gastos2<- subset(gastos2, select = -`Gastos y duración media de los viajes`)
```

```
gastos3 <- subset(gastos,gastos$`Gastos y duración media de los viajes`== tipos_da
colnames(gastos3)[colnames(gastos3)=="Total"]<- "Gasto_medio_diario_persona"
gastos3<- subset(gastos3, select = -`Gastos y duración media de los viajes`)
```

```
gastos4 <- subset(gastos,gastos$`Gastos y duración media de los viajes`== tipos_da
colnames(gastos4)[colnames(gastos4)=="Total"]<- "Duracion_media"
gastos4<- subset(gastos4, select = -`Gastos y duración media de los viajes`)
```

```
# Compruebo que en todas las columnas salvo la última todos las filas son iguales
any(gastos1[1:length(nrow(gastos1)-1)] !=gastos2[1:length(nrow(gastos2)-1)])
```

```
## [1] FALSE
```

```
any(gastos3[1:length(nrow(gastos1)-1)] !=gastos4[1:length(nrow(gastos2)-1)])
```

```
## [1] FALSE
```

```
# Uno los dos datasets en un único dataset con el que trabajar
```

```
gastos12 <- merge(gastos1,gastos2, by = c(colnames(gastos1[1:length(gastos1)-1])))
gastos34 <- merge(gastos3,gastos4, by = c(colnames(gastos3[1:length(gastos3)-1])))
```

```
# Compruebo que en todas las columnas salvo la última todos las filas son iguales
any(gastos12[1:length(nrow(gastos12)-2)] !=gastos34[1:length(nrow(gastos34)-2)])
```

```
## [1] FALSE
```

```
# Uno los dos datasets
```

```
datos <- merge(gastos12,gastos34, by = c(colnames(gastos12[1:4])))
```

```
unique(datos$`Tipo de dato`)
```

```
## [1] "Dato base" "Tasa de variación anual"
```

```
# Nos quedamos únicamente con los datos base, quitando las tasas de variación
datos <- subset(datos, datos$`Tipo de dato`== unique(datos$`Tipo de dato`)[1])
# Quito la columna irrelevante
datos <- subset(datos, select = -`Tipo de dato`)
```

Eliminar variables del entorno auxiliares

33

```
rm(gastos,gastos1, gastos2, gastos3,gastos34,gastos4,gastos12)
```

3.2. Transformación de clases

34

```
lapply(datos,class)
```

```
## $'País de residencia'
## [1] "character"
##
## $'Total Nacional y CCAA'
## [1] "character"
##
## $Periodo
## [1] "numeric"
##
## $Gasto_total
## [1] "character"
##
## $Gasto_medio_persona
## [1] "character"
##
## $Gasto_medio_diario_persona
## [1] "character"
##
## $Duracion_media
## [1] "character"
```

```
# Quitar punto de miles
datos[, 4:ncol(datos)] <- lapply(datos[, 4:ncol(datos)], function(x) gsub("\\.", " ", x))
# Sustituir coma decimal por punto decimal
datos[, 4:ncol(datos)] <- lapply(datos[, 4:ncol(datos)], function(x) gsub(",", ".", x))
# Transformar a numerico
datos[, 4:ncol(datos)] <- lapply(datos[, 4:ncol(datos)], function(x) as.numeric(x))
# Comprobar la clase
lapply(datos,class)
```

```
## $'País de residencia'
## [1] "character"
##
## $'Total Nacional y CCAA'
## [1] "character"
##
## $Periodo
## [1] "numeric"
##
## $Gasto_total
## [1] "numeric"
```

```
## 66
## $Gasto_medio_persona 67
## [1] "numeric" 68
## 69
## $Gasto_medio_diario_persona 70
## [1] "numeric" 71
## 72
## $Duracion_media 73
## [1] "numeric" 74
```

```
# Comprobar si hay datos NA
any(is.na(datos))
```

```
## [1] FALSE 75
```

```
unique(datos$`País de residencia`)
```

```
## [1] "Alemania"      "Francia"      "Italia"      "Países Nórdicos" 76
## [5] "Reino Unido"   "Total" 77
```

```
# Transformacion a factor de paises de residencia
datos$`País de residencia` <- as.factor(datos$`País de residencia`)
```

Quitar filas irrelevantes (totales que se pueden calcular a partir de los datos) para tener tidy data 78
79

```
# Quitar filas de total de columna comunidades autonomas
datos <- subset(datos, datos$`Total Nacional y CCAA` != "Total")
```

```
# Transformacion a factor de nombres de comunidades
unique(datos$`Total Nacional y CCAA`)
```

```
## [1] "01 Andalucía"      "04 Balears, Illes" 80
## [3] "05 Canarias"      "09 Cataluña" 81
## [5] "10 Comunitat Valenciana" "13 Madrid, Comunidad de" 82
## [7] "Otras Comunidades Autónomas" 83
```

```
datos$`Total Nacional y CCAA` <- as.factor(datos$`Total Nacional y CCAA`)
levels(datos$`Total Nacional y CCAA`) <- c("Andalucía", "Illes Balears", "Canarias")
```

3.3. Cambios de nombres de las columnas 84

```
colnames(datos)
```

```
## [1] "País de residencia"      "Total Nacional y CCAA" 85
## [3] "Periodo"      "Gasto_total" 86
## [5] "Gasto_medio_persona"    "Gasto_medio_diario_persona" 87
## [7] "Duracion_media" 88
```

```
nombres_columnas <- c("Pais", "CCAA", "Periodo", colnames(datos)[4:7])
colnames(datos) <- nombres_columnas
colnames(datos)
```

```
## [1] "Pais"      "CCAA" 89
## [3] "Periodo"    "Gasto_total" 90
## [5] "Gasto_medio_persona"    "Gasto_medio_diario_persona" 91
## [7] "Duracion_media" 92
```

3.4. Arreglo valores de filas

93

```
# Cambiar el nombre de las filas
países <- levels(datos$País)
países[países=="Total"] <- "Otros"
levels(datos$País) <- países
```

3.4.1. Arreglar valores de Gasto_total

94

Hay que crear filas con valor en país residencia = resto de países y asignarla a total - suma del resto de países por comunidades

95

96

```
comunidades <- levels(datos$CCAA)
años <- unique(datos$Periodo)
for (i in comunidades){
  for (j in años){
    aux <- datos[datos$CCAA==i & datos$Periodo == j, "Gasto_total"]
    datos[datos$CCAA==i & datos$Periodo == j & datos$País == "Otros", "Gasto_total"] <-
  }
}
```

3.4.2. Arreglar valores de las otras 3 variables

97

Consideramos la media ponderada. Numero de países entre 194 y 206.

98

$media_total = media_otros \times (194-5)/194 + (5/194) \times \text{sum}(media_países)$ Por comunidad autónoma y por año $media_otros = (media_total - (5/194) \times \text{sum}(media_países)) \times 194/(194-5)$

99

100

101

```
# Arreglar duracion media
for (i in comunidades){
  for (j in años){
    aux <- datos[datos$CCAA==i & datos$Periodo == j, "Duracion_media"]
    datos[datos$CCAA==i & datos$Periodo == j & datos$País == "Otros", "Duracion_media"] <-
  }
}
```

```
# Arreglar Gasto_medio_persona
for (i in comunidades){
  for (j in años){
    aux <- datos[datos$CCAA==i & datos$Periodo == j, "Gasto_medio_persona"]
    datos[datos$CCAA==i & datos$Periodo == j & datos$País == "Otros", "Gasto_medio_persona"] <-
  }
}
```

```
# Arreglar Gasto_medio_diario_persona
for (i in comunidades){
  for (j in años){
    aux <- datos[datos$CCAA==i & datos$Periodo == j, "Gasto_medio_diario_persona"]
    datos[datos$CCAA==i & datos$Periodo == j & datos$País == "Otros", "Gasto_medio_diario_persona"] <-
  }
}

rm(i,j,años,comunidades, países, aux)
```

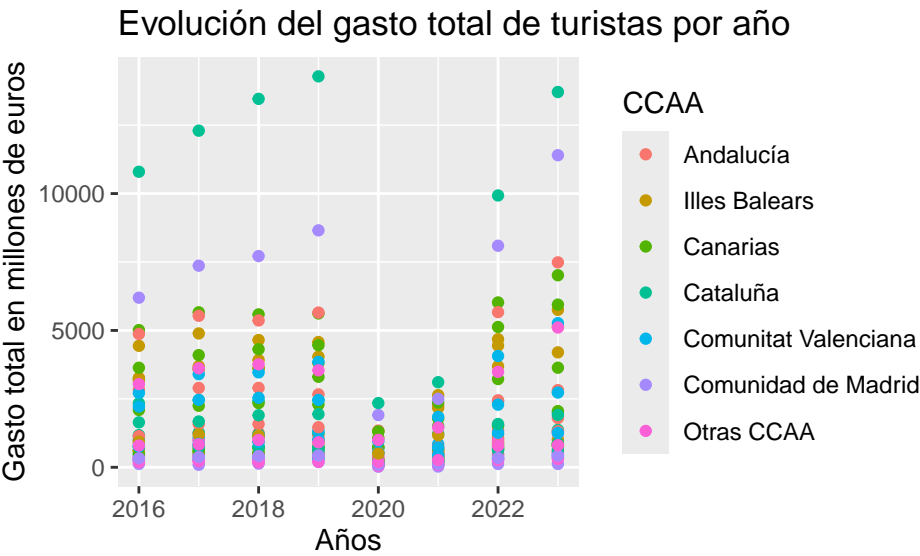
```
summary(datos)
```

##	Pais	CCAA	Periodo	
##	Alemania :56	Andalucía :48	Min. :2016	
##	Francia :56	Illes Balears :48	1st Qu.:2018	
##	Italia :56	Canarias :48	Median :2020	
##	Países Nórdicos:56	Cataluña :48	Mean :2020	
##	Reino Unido :56	Comunitat Valenciana:48	3rd Qu.:2021	
##	Otros :56	Comunidad de Madrid :48	Max. :2023	
##		Otras CCAA :48		
##	Gasto_total	Gasto_medio_persona	Gasto_medio_diario_persona	
##	Min. : 22.28	Min. : 383.0	Min. : 56.0	
##	1st Qu.: 396.06	1st Qu.: 832.8	1st Qu.:105.0	
##	Median : 849.69	Median : 980.8	Median :130.5	
##	Mean : 1775.89	Mean : 983.7	Mean :133.0	
##	3rd Qu.: 2375.74	3rd Qu.:1138.0	3rd Qu.:154.0	
##	Max. :14281.08	Max. :1612.5	Max. :277.0	
##				
##	Duracion_media			
##	Min. : 3.610			
##	1st Qu.: 5.793			
##	Median : 7.465			
##	Mean : 7.586			
##	3rd Qu.: 9.040			
##	Max. :14.270			
##				

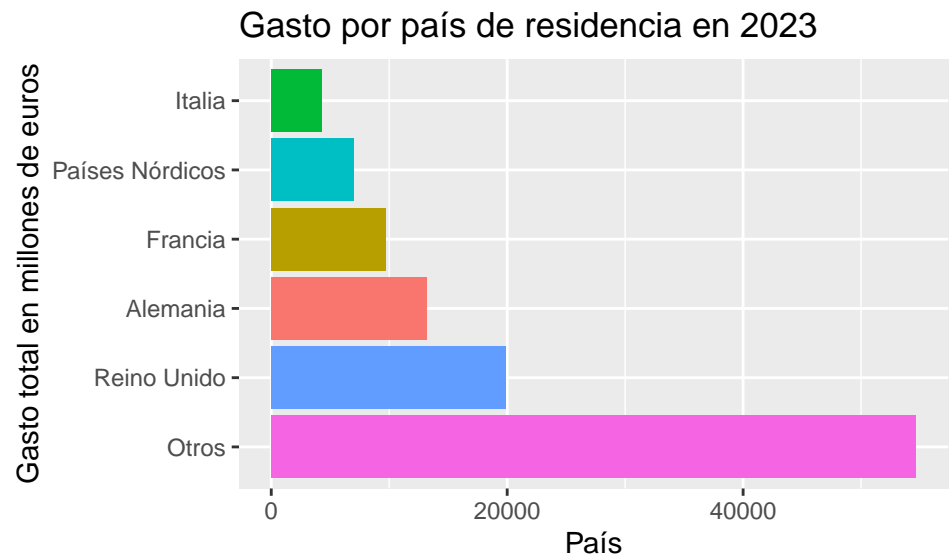
4. Prueba de ggplot

```
# Visualización de los datos e interpretación de los posibles patrones

# Visualización previa para tener una idea de ciertas variables de los datos
ggplot(datos, aes(x = Periodo, y = Gasto_total, color = CCAA, group = 1)) +
  geom_point() +
  labs(title = "Evolución del gasto total de turistas por año", x = "Años", y = "
```

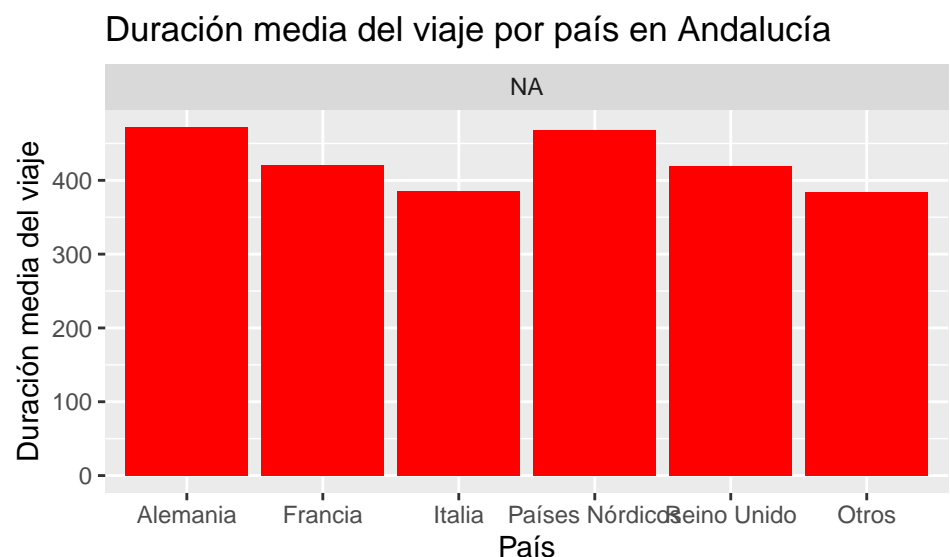


```
# Nos centramos en un año en concreto y visualizamos el gasto total por país en es
ggplot(datos[datos$Periodo == 2023, ], aes(x = reorder(Pais, -Gasto_total), y = Ga
  geom_bar(stat = "identity", aes(fill = Pais)) +
  coord_flip() +
  labs(title = "Gasto por país de residencia en 2023", x = "Gasto total en millone
  theme(legend.position = "none")
```



128

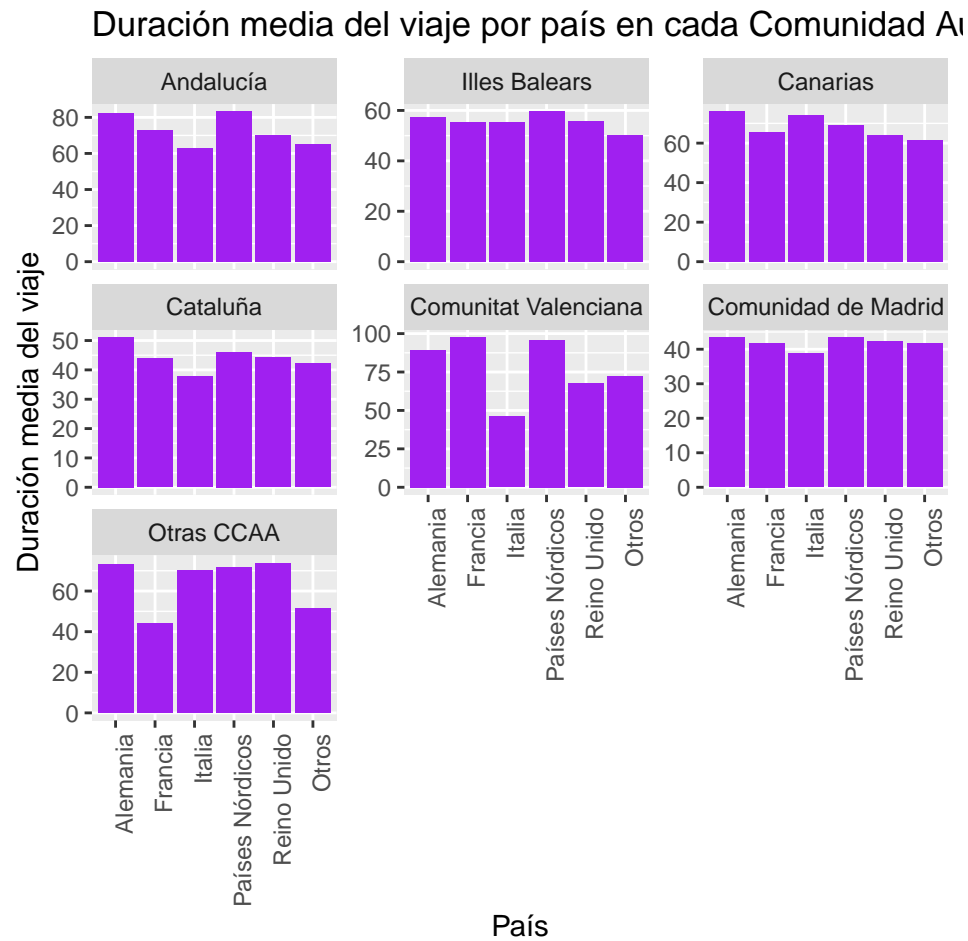
```
# Para analizar únicamente una variable por país en una sola Comunidad Autónoma.
ggplot(datos, aes(x = Pais, y = Duracion_media)) +
  geom_bar(stat = "identity", fill = "red") +
  facet_grid(~CCAA["i"], scales = "free") +
  labs(title = "Duración media del viaje por país en Andalucía", x = "País", y = "
```



129

```
# Para todas las Comunidades Autónomas por separado
ggplot(datos, aes(x = Pais, y = Duracion_media)) +
  geom_bar(stat = "identity", fill = "purple") +
  facet_wrap(~ CCAA, scales = "free_y") +
  labs(title = "Duración media del viaje por país en cada Comunidad Autónoma",
```

```
x = "País",
y = "Duración media del viaje" +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Appendix E

Appendix E.1

The appendix is an optional section that can contain details and data supplemental to the main text. For example, explanations of experimental details that would disrupt the flow of the main text, but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

Appendix F

All appendix sections must be cited in the main text. In the appendixes, Figures, Tables, etc. should be labeled starting with ‘A’, e.g., Figure A1, Figure A2, etc.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.