

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Seminario de Sistemas 2
Ingeniero Marlon Orellana



Manual Técnico

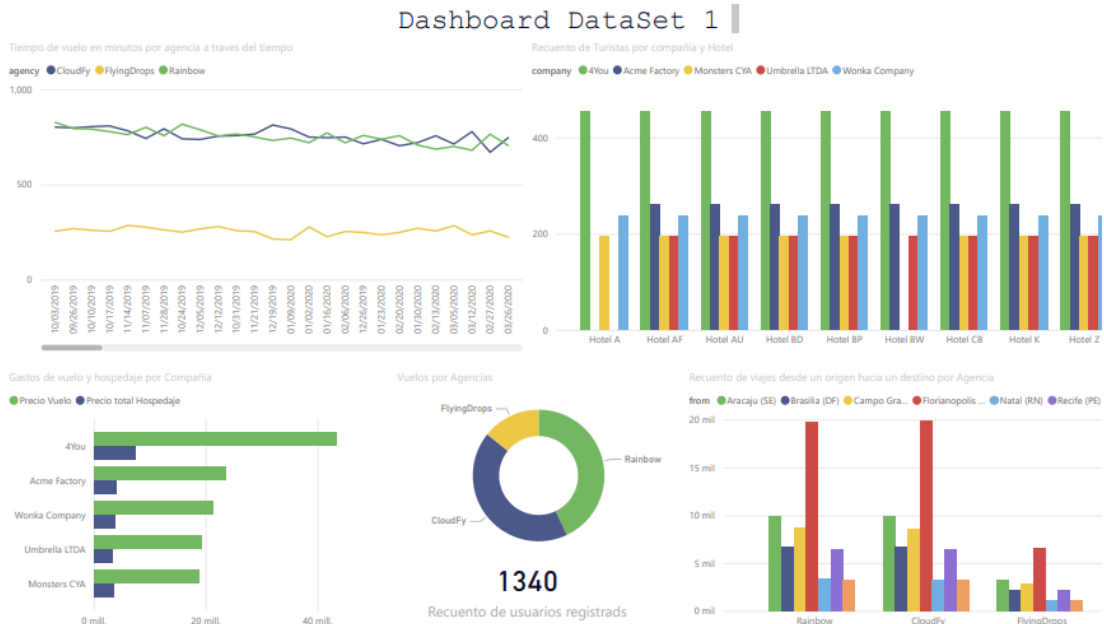
Practica 2

Nombre: Jackeline Alexandra Benitez Benitez
Carnet: 201709166

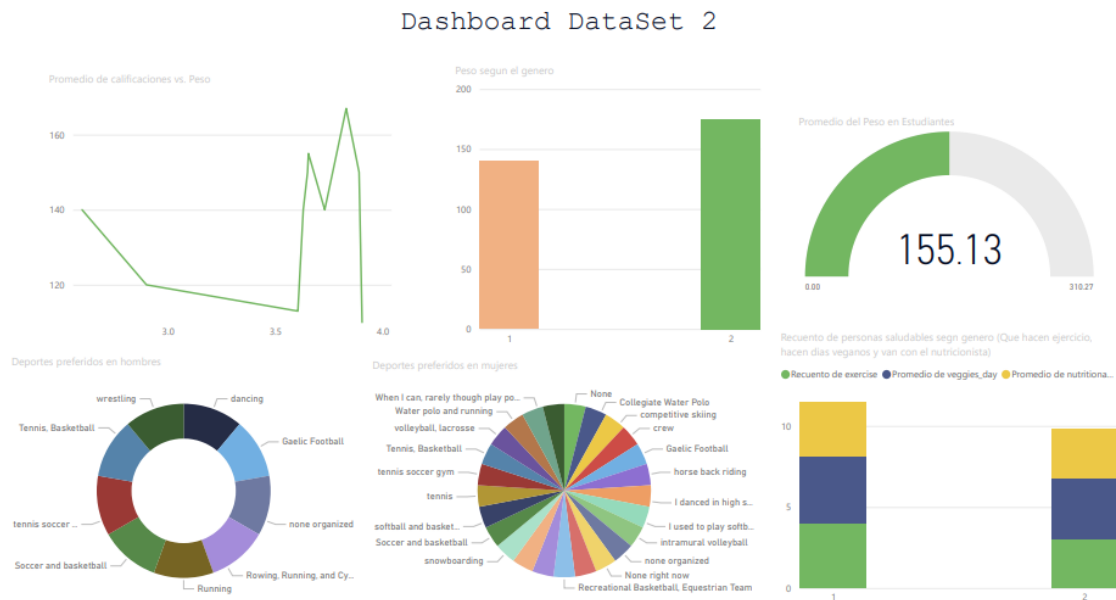
Contenido

DashBoard DataSet1	3
DashBoard DataSet2	3
DashBoard DataSet3	4
EDA de Datasets 1	6
DataSet 1 no Limpio	6
EDA Distancia de vuelos	7
EDA Tiempo de vuelos.....	8
EDA Precio de vuelos.....	9
EDA Precio total de vuelos.....	10
EDA de Edad de Usuarios:.....	11
DataSet 1 Limpio	12
EDA Distancia de vuelos	13
EDA de Datasets 2.....	19
DataSet 2 no Limpio	19
EDA de Calories per day:	21
EDA de Weight:	22
EDA de Healthy Feeling:	23
EDA de GPA:	24
EDA de Calories Scone:	25
EDA de Veggie Days:	26
EDA de Datasets 3.....	33
DataSet 3 no Limpio	33
EDA de Teaching:	36
EDA de Total Score:.....	37
EDA de Award:.....	38
EDA de National Rank:.....	39

DashBoard DataSet1



DashBoard DataSet2



Dashboard DataSet 3

DataSet 1

EDA de Datasets 1

DataSet 1 no Limpio

Resumen estadístico de Flights

```
> summary(flights)
  travelCode      userCode      from      to
Min.   : 0      Min.   : 0.0  Length:271888  Length:271888
1st Qu.: 33986    1st Qu.: 326.0  Class :character  Class :character
Median : 67972    Median : 659.0  Mode  :character  Mode  :character
Mean   : 67972    Mean   : 667.5
3rd Qu.:101957    3rd Qu.:1011.0
Max.   :135943    Max.   :1339.0

  flightType      price      time      distance
Length:271888    Min.   : 301.5  Min.   :0.440  Min.   :168.2
Class :character  1st Qu.: 672.7  1st Qu.:1.040  1st Qu.:401.7
Mode  :character  Median : 904.0  Median :1.460  Median :562.1
                  Mean   : 957.4  Mean   :1.421  Mean   :547.0
                  3rd Qu.:1222.2  3rd Qu.:1.760  3rd Qu.:676.5
                  Max.   :1754.2  Max.   :2.440  Max.   :937.8

  agency      date
Length:271888  Length:271888
Class :character  Class :character
Mode  :character  Mode  :character
```

Resumen estadístico de Hotels

```
> summary(hotels)
  travelCode      userCode      name      place
Min.   : 0      Min.   : 0      Length:40552  Length:40552
1st Qu.: 33697    1st Qu.: 323      Class :character  Class :character
Median : 67831    Median : 658      Mode  :character  Mode  :character
Mean   : 67912    Mean   : 667
3rd Qu.:102211    3rd Qu.:1013
Max.   :135942    Max.   :1339

  days      price      total      date
Min.   :1.0  Min.   : 60.39  Min.   : 60.39  Length:40552
1st Qu.:1.0  1st Qu.:165.99  1st Qu.: 247.62  Class :character
Median :2.0  Median :242.88  Median : 495.24  Mode  :character
Mean   :2.5  Mean   :214.44  Mean   : 536.23
3rd Qu.:4.0  3rd Qu.:263.41  3rd Qu.: 742.86
Max.   :4.0  Max.   :313.02  Max.   :1252.08
```

Resumen estadístico users

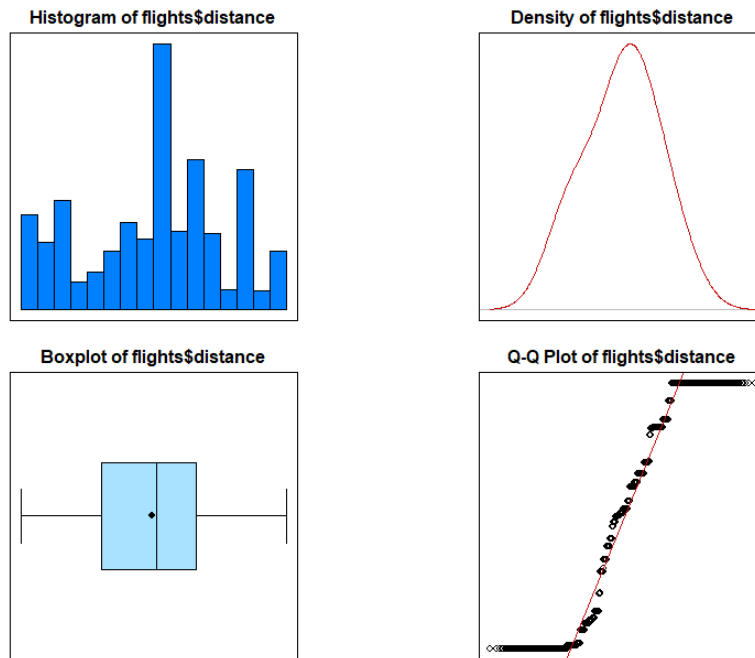
```
> summary(users)
  userCode      company      name      gender
Min.   : 0.0      Length:1340  Length:1340  Length:1340
1st Qu.: 334.8    Class :character  Class :character  Class :character
Median : 669.5    Mode  :character  Mode  :character  Mode  :character
Mean   : 669.5
3rd Qu.:1004.2
Max.   :1339.0

  age
Min.   :21.00
1st Qu.:32.00
Median :42.00
Mean   :42.74
3rd Qu.:54.00
Max.   :65.00
```

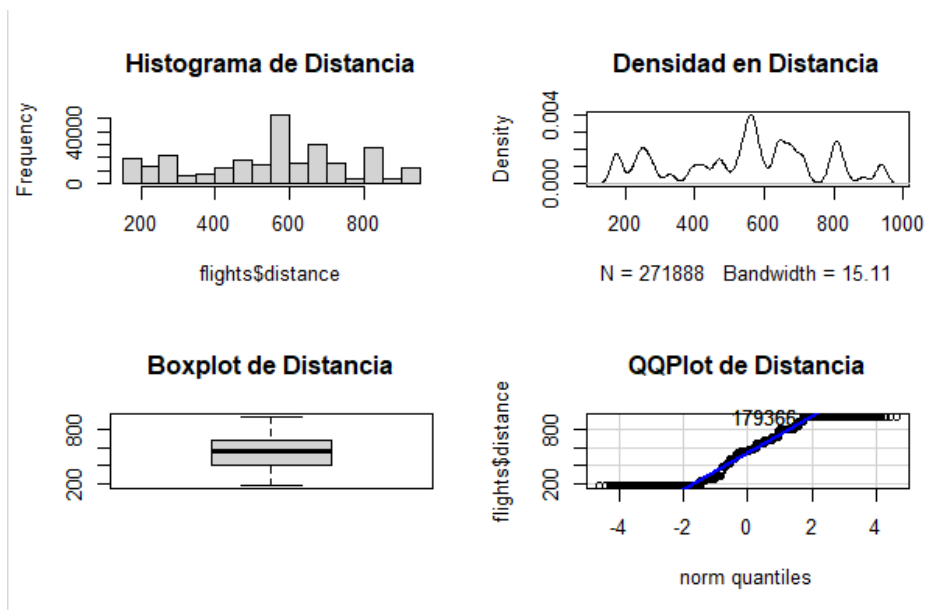
EDA Distancia de vuelos

EDA Generado Automaticamente:

EXPLORATORY DATA ANALYSIS



EDA generado manualmente



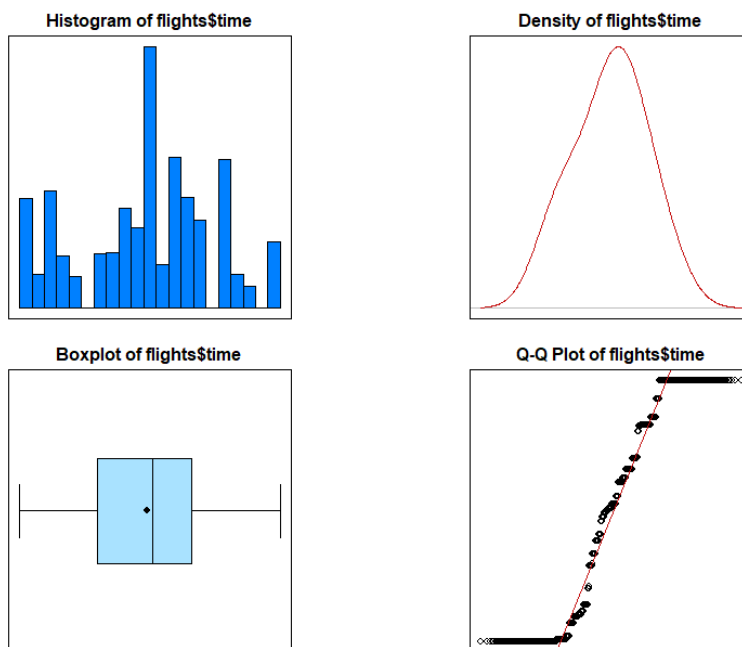
Resumen Estadístico de Distancia

```
summary(flights$distance)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
168.2  401.7   562.1   547.0  676.5   937.8
```

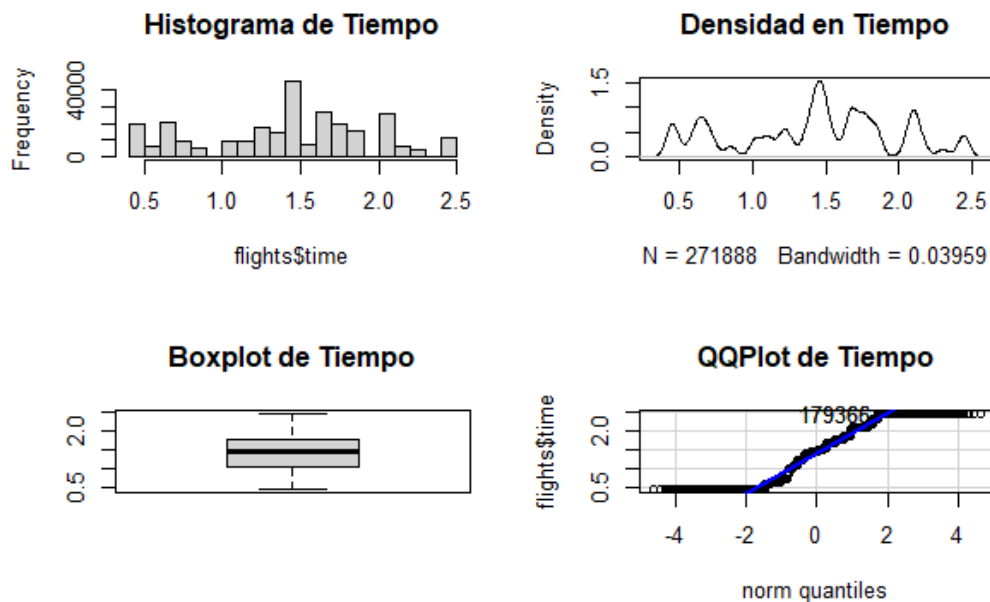
EDA Tiempo de vuelos

EDA Generado Automaticamente:

EXPLORATORY DATA ANALYSIS



EDA Generado Manualmente



Resumen estadístico de Tiempo en vuelos

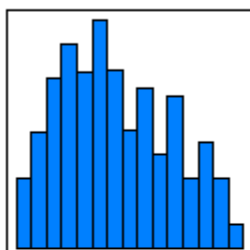
```
> summary(flights$time)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.440  1.040   1.460   1.421  1.760   2.440
> |
```


EDA Precio de vuelos

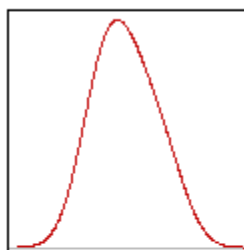
EDA Generado Automaticamente:

EXPLORATORY DATA ANALYSIS

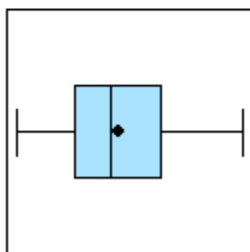
Histogram of flights\$price



Density of flights\$price



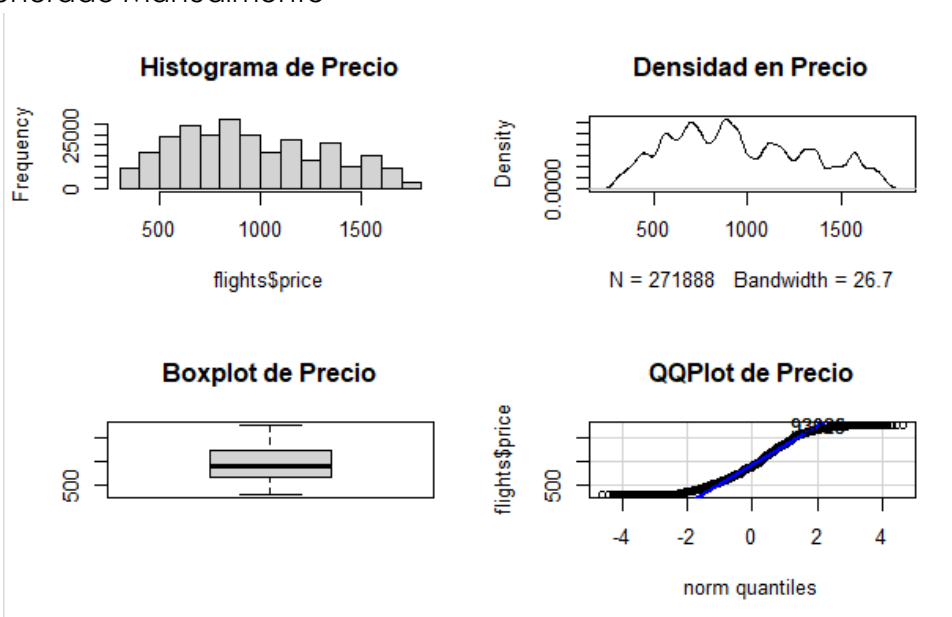
Boxplot of flights\$price



Q-Q Plot of flights\$price



EDA Generado Manualmente



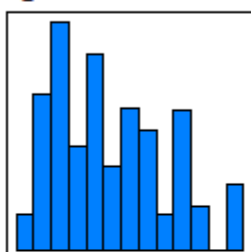
Resumen Estadístico de Precios de Vuelo

```
> summary(flights$price)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 301.5   672.7   904.0   957.4  1222.2  1754.2
```

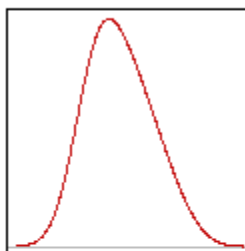
EDA Precio total de vuelos
EDA Generado Automaticamente:

EXPLORATORY DATA ANALYSIS

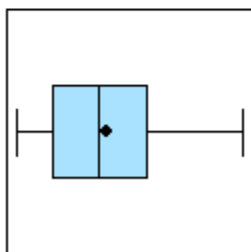
Histogram of hotels\$total



Density of hotels\$total



Boxplot of hotels\$total

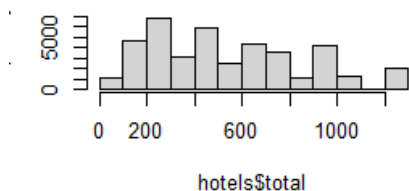


Q-Q Plot of hotels\$total

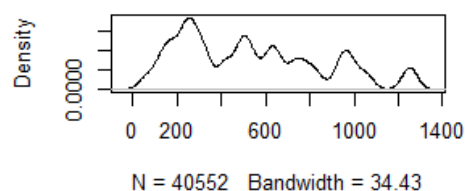


EDA generado Manualmente:

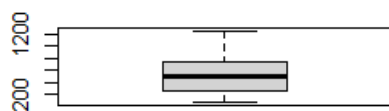
Histograma de Precio total de hotel



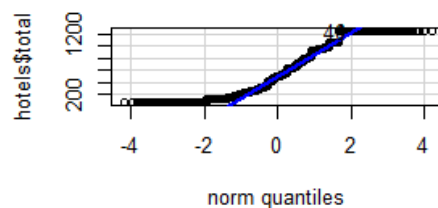
Densidad en Precio total de hotel



Boxplot de Precio total de hotel



QQPlot de Precio total de hotel



Resumen estadístico de precio total en hoteles:

```
summary(hotels$total)
```

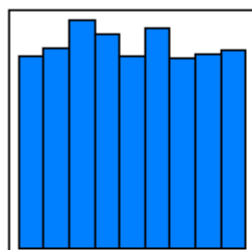
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
60.39	247.62	495.24	536.23	742.86	1252.08

EDA de Edad de Usuarios:

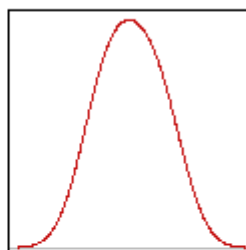
EDA Generado Automaticamente:

EXPLORATORY DATA ANALYSIS

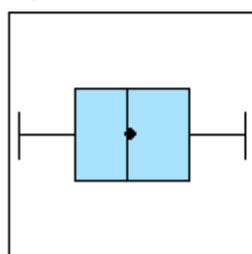
Histogram of users\$age



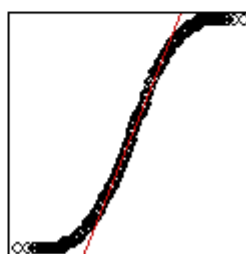
Density of users\$age



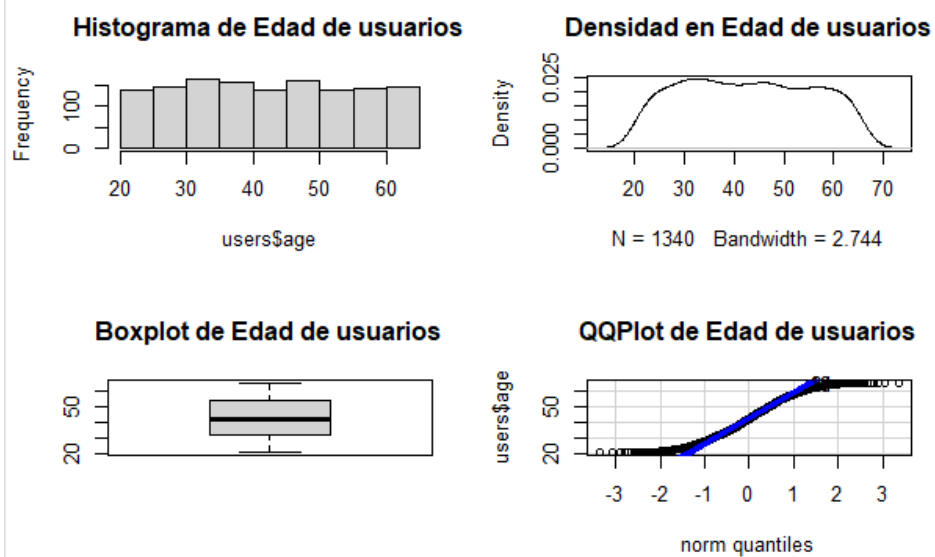
Boxplot of users\$age



Q-Q Plot of users\$age



EDA Generado Manualmente:



Resumen estadístico de la edad de los usuarios

```
> summary(users$age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 21.00  32.00   42.00   42.74  54.00   65.00
```

DataSet 1 Limpio

```
> summary(flights1)
  travelCode      userCode      from      to
Min.   :    0   Min.   :  0.0   Length:271888   Length:271888
1st Qu.: 33986   1st Qu.: 326.0   Class :character   Class :character
Median : 67972   Median : 659.0   Mode  :character   Mode  :character
Mean   : 67972   Mean   : 667.5
3rd Qu.:101957   3rd Qu.:1011.0
Max.   :135943   Max.   :1339.0
  flightType      price      time      distance
Length:271888   Min.   : 301.5   Min.   :0.440   Min.   :168.2
Class :character 1st Qu.: 672.7   1st Qu.:1.040   1st Qu.:401.7
Mode  :character Median : 904.0   Median :1.460   Median :562.1
                  Mean   : 957.4   Mean   :1.421   Mean   :547.0
                  3rd Qu.:1222.2   3rd Qu.:1.760   3rd Qu.:676.5
                  Max.   :1754.2   Max.   :2.440   Max.   :937.8
  agency      date
Length:271888   Length:271888
Class :character Class :character
Mode  :character Mode  :character

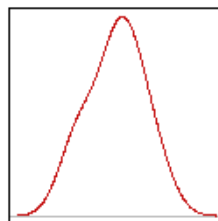
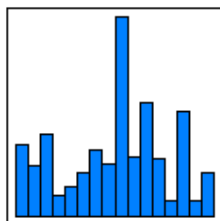
> summary(hotels1)
  travelCode      userCode      name      place
Min.   :    0   Min.   :    0   Length:40552   Length:40552
1st Qu.: 33697   1st Qu.: 323   Class :character   Class :character
Median : 67831   Median : 658   Mode  :character   Mode  :character
Mean   : 67912   Mean   : 667
3rd Qu.:102211   3rd Qu.:1013
Max.   :135942   Max.   :1339
  days      price      total      date
Min.   :1.0   Min.   : 60.39   Min.   : 60.39   Length:40552
1st Qu.:1.0   1st Qu.:165.99   1st Qu.: 247.62   Class :character
Median :2.0   Median :242.88   Median : 495.24   Mode  :character
Mean   :2.5   Mean   :214.44   Mean   : 536.23
3rd Qu.:4.0   3rd Qu.:263.41   3rd Qu.: 742.86
Max.   :4.0   Max.   :313.02   Max.   :1252.08

>
> summary(users1)
  userCode      company      name      gender
Min.   :    0.0   Length:1340   Length:1340   Length:1340
1st Qu.: 334.8   Class :character   Class :character   Class :character
Median : 669.5   Mode  :character   Mode  :character   Mode  :character
Mean   : 669.5
3rd Qu.:1004.2
Max.   :1339.0
  age
Min.   :21.00
1st Qu.:32.00
Median :42.00
Mean   :42.74
3rd Qu.:54.00
Max.   :65.00
```

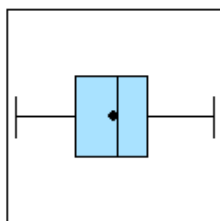
EDA Distancia de vuelos
EDA Generado Automaticamente

EXPLORATORY DATA ANALYSIS

istogram of flights1\$distanceDensity of flights1\$distance

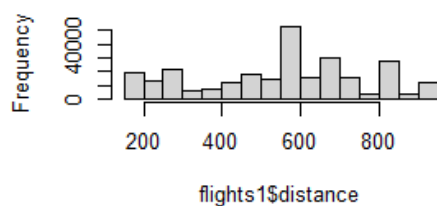


Boxplot of flights1\$distanceQ-Q Plot of flights1\$distance

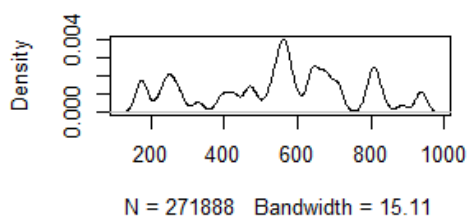


EDA Generado manualmente

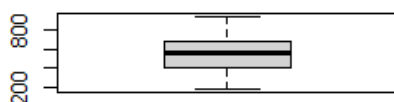
Histograma de Distancia



Densidad en Distancia



Boxplot de Distancia



Resumen Estadístico

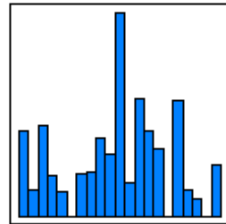
```
summary(flights1$distance)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
168.2	401.7	562.1	547.0	676.5	937.8

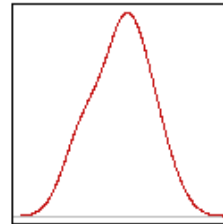
EDA tiempo de vuelos
EDA Generado Automaticamente

EXPLORATORY DATA ANALYSIS

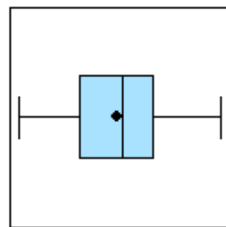
Histogram of flights\$time



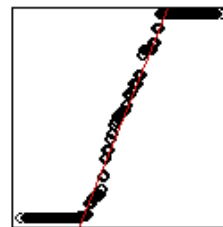
Density of flights\$time



Boxplot of flights\$time

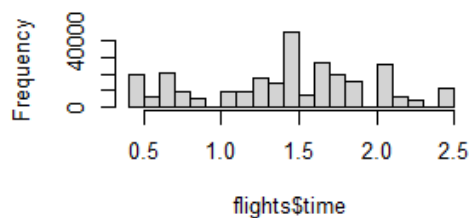


Q-Q Plot of flights\$time

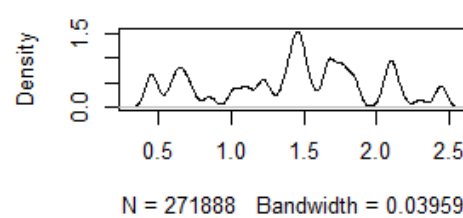


EDA generado manualmente

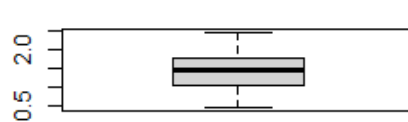
Histograma de Tiempo



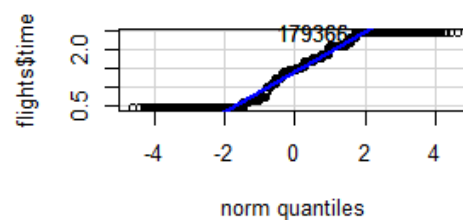
Densidad en Tiempo



Boxplot de Tiempo



QQPlot de Tiempo



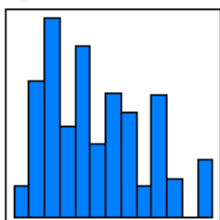
Resumen Estadístico

```
> summary(flights$time)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.440  1.040   1.460   1.421  1.760   2.440
```

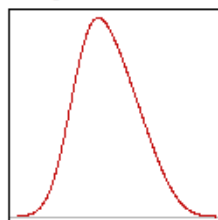
EDA precio de vuelos
EDA Generado Automaticamente

EXPLORATORY DATA ANALYSIS

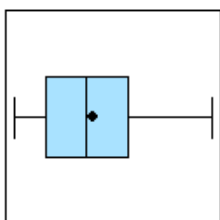
Histogram of hotels\$total



Density of hotels\$total



Boxplot of hotels\$total

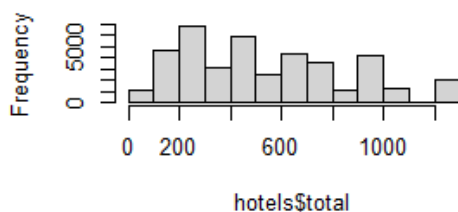


Q-Q Plot of hotels\$total

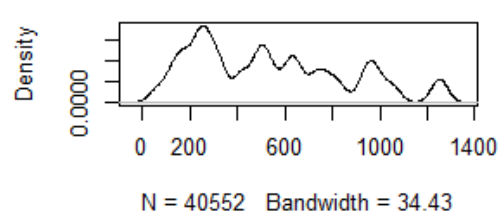


EDA generado manualmente

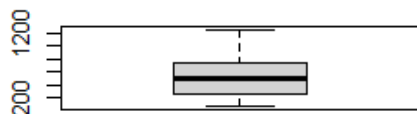
Histograma de Precio total de hotel



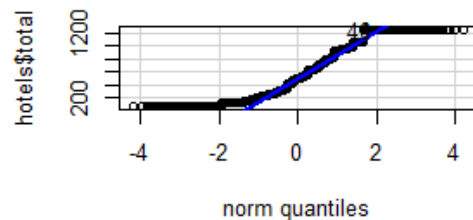
Densidad en Precio total de hotel



Boxplot de Precio total de hotel



QQPlot de Precio total de hotel



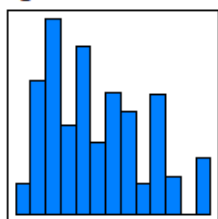
Resumen Estadístico

```
> summary(flights$price)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 301.5  672.7   904.0   957.4 1222.2  1754.2
```

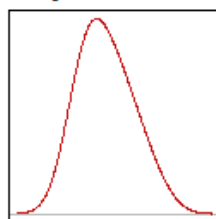
EDA precio total de hoteles
 EDA Generado Automaticamente

EXPLORATORY DATA ANALYSIS

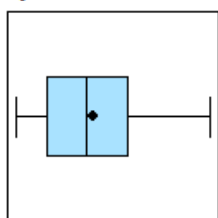
Histogram of hotels\$total



Density of hotels\$total



Boxplot of hotels\$total

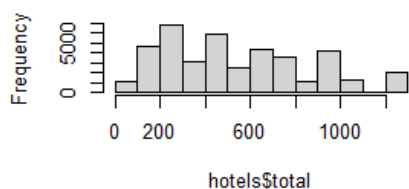


Q-Q Plot of hotels\$total

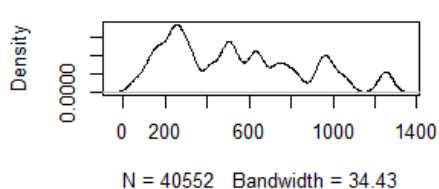


EDA Generado manualmente

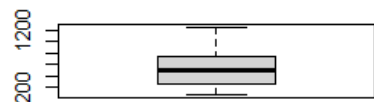
Histograma de Precio total de hotel



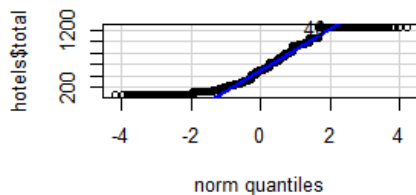
Densidad en Precio total de hotel



Boxplot de Precio total de hotel



QQPlot de Precio total de hotel



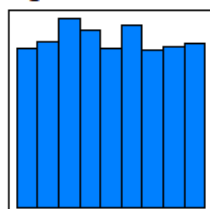
Resumen Estadístico

```
> summary(hotels$total)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 60.39 247.62  495.24  536.23  742.86 1252.08
```

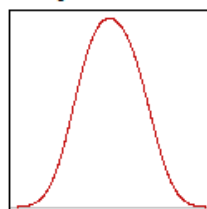

EDA edad de usuarios
EDA Generado Automaticamente

EXPLORATORY DATA ANALYSIS

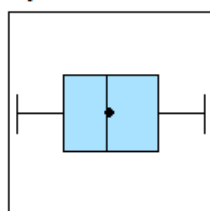
Histogram of users\$age



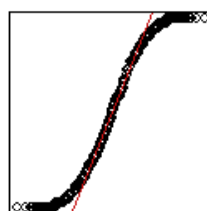
Density of users\$age



Boxplot of users\$age

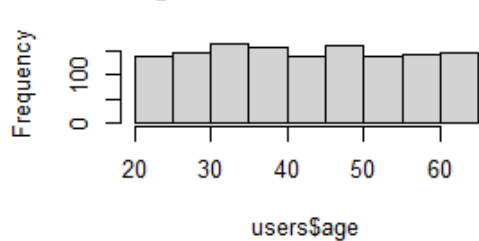


Q-Q Plot of users\$age

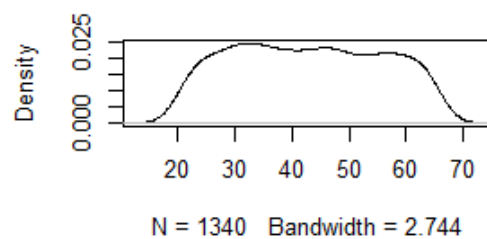


EDA Generado manualmente

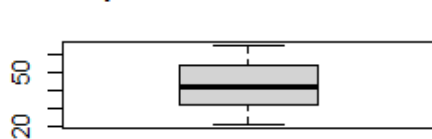
Histograma de Edad de usuarios



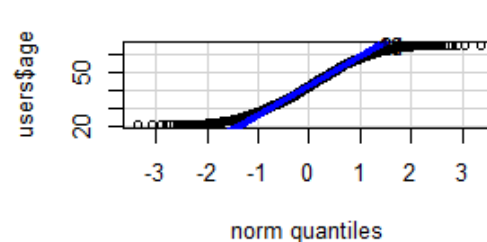
Densidad en Edad de usuarios



Boxplot de Edad de usuarios



QQPlot de Edad de usuarios



Resumen estadístico

```
> summary(users$age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  21.00  32.00  42.00  42.74  54.00  65.00
```

DataSet 2

EDA de Datasets 2

DataSet 2 no Limpio

> summary(food)

GPA	Gender	breakfast	calories_chicken	
Length:125	Min. :1.000	Min. :1.000	Min. :265.0	
Class :character	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:430.0	
Mode :character	Median :1.000	Median :1.000	Median :610.0	
	Mean :1.392	Mean :1.112	Mean :577.3	
	3rd Qu.:2.000	3rd Qu.:1.000	3rd Qu.:720.0	
	Max. :2.000	Max. :2.000	Max. :720.0	
calories_day	calories_scone	coffee	comfort_food	
Min. :2.000	Min. :315.0	Min. :1.000	Length:125	
1st Qu.:3.000	1st Qu.:420.0	1st Qu.:2.000	Class :character	
Median :3.000	Median :420.0	Median :2.000	Mode :character	
Mean :3.028	Mean :505.2	Mean :1.752		
3rd Qu.:3.000	3rd Qu.:420.0	3rd Qu.:2.000		
Max. :4.000	Max. :980.0	Max. :2.000		
NA's :19	NA's :1			
comfort_food_reasons	comfort_food_reasons_coded	cook		
Length:125	Min. :1.000	Min. :1.000		
Class :character	1st Qu.:2.000	1st Qu.:2.000		
Mode :character	Median :2.000	Median :3.000		
	Mean :2.698	Mean :2.787		
	3rd Qu.:3.000	3rd Qu.:3.000		
	Max. :9.000	Max. :5.000		
	NA's :19	NA's :3		
comfort_food_reasons_coded.1	cuisine	diet_current	diet_current_coded	
Min. :1.000	Min. :1.000	Length:125	Min. :1.00	
1st Qu.:2.000	1st Qu.:1.000	Class :character	1st Qu.:1.00	
Median :2.000	Median :1.000	Mode :character	Median :2.00	
Mean :2.688	Mean :1.389		Mean :1.76	
3rd Qu.:3.000	3rd Qu.:1.000		3rd Qu.:2.00	
Max. :9.000	Max. :6.000		Max. :4.00	
	NA's :17			
drink	eating_changes	eating_changes_coded	eating_changes_coded1	
Min. :1.000	Length:125	Min. :1.000	Min. :1.000	
1st Qu.:1.000	Class :character	1st Qu.:1.000	1st Qu.:3.000	
Median :2.000	Mode :character	Median :1.000	Median :4.000	
Mean :1.561		Mean :1.536	Mean :4.552	
3rd Qu.:2.000		3rd Qu.:2.000	3rd Qu.:5.000	
Max. :2.000		Max. :4.000	Max. :13.000	
NA's :2				
eating_out	employment	ethnic_food	exercise	father_education
Min. :1.00	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:2.00	1st Qu.:2.000	1st Qu.:3.000	1st Qu.:1.000	1st Qu.:2.000
Median :2.00	Median :2.000	Median :4.000	Median :1.000	Median :4.000
Mean :2.56	Mean :2.448	Mean :3.744	Mean :1.589	Mean :3.484
3rd Qu.:3.00	3rd Qu.:3.000	3rd Qu.:5.000	3rd Qu.:2.000	3rd Qu.:4.000
Max. :5.00	Max. :3.000	Max. :5.000	Max. :3.000	Max. :5.000
	NA's :9		NA's :13	NA's :1
father_profession	fav_cuisine	fav_cuisine_coded	fav_food	
Length:125	Length:125	Min. :0.000	Min. :1.000	
Class :character	Class :character	1st Qu.:1.000	1st Qu.:1.000	
Mode :character	Mode :character	Median :1.000	Median :1.000	
		Mean :2.424	Mean :1.715	
		3rd Qu.:4.000	3rd Qu.:3.000	
		Max. :8.000	Max. :3.000	
			NA's :2	
food_childhood	fries	fruit_day	grade_level	greek_food
Length:125	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
Class :character	1st Qu.:1.000	1st Qu.:4.000	1st Qu.:1.000	1st Qu.:3.000
Mode :character	Median :1.000	Median :5.000	Median :2.000	Median :4.000
	Mean :1.088	Mean :4.224	Mean :2.376	Mean :3.488
	3rd Qu.:1.000	3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:5.000
	Max. :2.000	Max. :5.000	Max. :4.000	Max. :5.000

healthy_feeling	healthy_meal	ideal_diet	ideal_diet_coded
Min. : 1.000	Length:125	Length:125	Min. :1.000
1st Qu.: 3.000	Class :character	Class :character	1st Qu.:2.000
Median : 5.000	Mode :character	Mode :character	Median :3.000
Mean : 5.456			Mean :3.704
3rd Qu.: 8.000			3rd Qu.:6.000
Max. :10.000			Max. :8.000

income	indian_food	italian_food	life_rewarding	marital_status
Min. :1.000	Min. :1.000	Min. :3.000	Min. : 1.000	Min. :1.0
1st Qu.:4.000	1st Qu.:2.000	1st Qu.:5.000	1st Qu.: 2.000	1st Qu.:1.0
Median :5.000	Median :3.000	Median :5.000	Median : 5.000	Median :1.0
Mean :4.532	Mean :3.152	Mean :4.728	Mean : 5.105	Mean :1.5
3rd Qu.:6.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.: 8.000	3rd Qu.:2.0
Max. :6.000	Max. :5.000	Max. :5.000	Max. :10.000	Max. :4.0
NA's :1			NA's :1	NA's :1

meals_dinner_friend	mother_education	mother_profession	nutritional_check
Length:125	Min. :1.000	Length:125	Min. :1.000
Class :character	1st Qu.:2.000	Class :character	1st Qu.:2.000
Mode :character	Median :4.000	Mode :character	Median :3.000
	Mean :3.426		Mean :3.152
	3rd Qu.:4.000		3rd Qu.:4.000
	Max. :5.000		Max. :5.000
	NA's :3		

on_off_campus	parents_cook	pay_meal_out	persian_food
Min. :1.000	Min. :1.000	Min. :2.000	Min. :1.000
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:3.000	1st Qu.:2.000
Median :1.000	Median :1.000	Median :3.000	Median :3.000
Mean :1.323	Mean :1.528	Mean :3.408	Mean :2.806
3rd Qu.:1.000	3rd Qu.:2.000	3rd Qu.:4.000	3rd Qu.:4.000
Max. :4.000	Max. :5.000	Max. :6.000	Max. :5.000
NA's :1			NA's :1

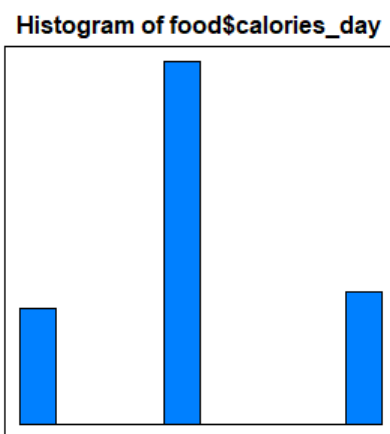
self_perception_weight	soup	sports	thai_food
Min. :1.000	Min. :1.000	Min. :1.00	Min. :1.000
1st Qu.:2.000	1st Qu.:1.000	1st Qu.:1.00	1st Qu.:2.000
Median :3.000	Median :1.000	Median :1.00	Median :3.000
Mean :3.121	Mean :1.218	Mean :1.39	Mean :3.336
3rd Qu.:4.000	3rd Qu.:1.000	3rd Qu.:2.00	3rd Qu.:5.000
Max. :6.000	Max. :2.000	Max. :2.00	Max. :5.000
NA's :1	NA's :1	NA's :2	

tortilla_calories	turkey_calories	type_sports	veggies_day
Min. : 580.0	Min. :345	Length:125	Min. :1.000
1st Qu.: 725.0	1st Qu.:500	Class :character	1st Qu.:3.000
Median : 940.0	Median :500	Mode :character	Median :4.000
Mean : 947.6	Mean :555		Mean :4.008
3rd Qu.:1165.0	3rd Qu.:690		3rd Qu.:5.000
Max. :1165.0	Max. :850		Max. :5.000
NA's :1			

vitamins	waffle_calories	weight
Min. :1.000	Min. : 575	Length:125
1st Qu.:1.000	1st Qu.: 900	Class :character
Median :2.000	Median : 900	Mode :character
Mean :1.512	Mean :1073	
3rd Qu.:2.000	3rd Qu.:1315	
Max. :2.000	Max. :1315	

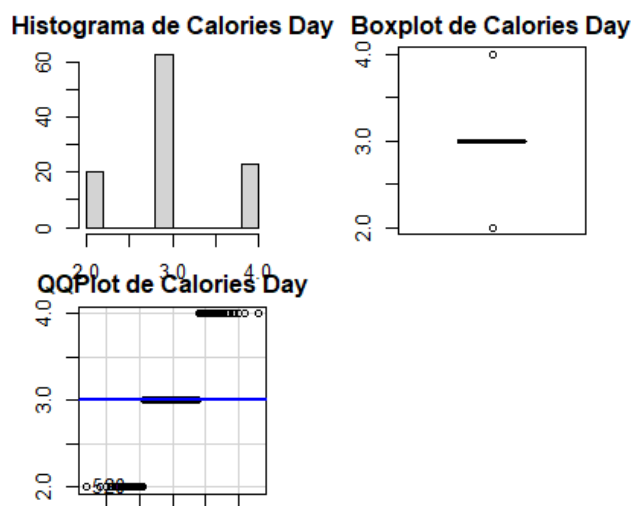
EDA de Calories per day:

EDA Generado Automaticamente:



Como se puede observar debido a la no limpieza de datos no se genera correctamente automaticamente

EDA generado Manualmente:



Se puede observar la presencia de outliers en Calories Day debido a que algunos datos aparecen como Nan

Resumen estadístico de calorías por día

```
summary(food$calories_day)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 2.000  3.000   3.000   3.028  3.000   4.000     19
```

EDA de Weight:

EDA Generado automáticamente:

```
> EDA(food$weight)
Error in Sort[V3] - Sort[V1] : non-numeric argument to binary operator
> par(mfrow=c(2,2))
```

Como se puede observar no fue posible generar el EDA automáticamente ya que este presenta datos erróneos

EDA Generado Manualmente:

```
> par(mfrow=c(2,2))
> hist(food$weight, main="Histograma de Peso")
Error in hist.default(food$weight, main = "Histograma de Peso") :
  'x' must be numeric
> plot(density(food$weight), main="Densidad en Peso")
Error in density.default(food$weight) : argument 'x' must be numeric
> boxplot(food$weight, main="Boxplot de Peso")
Error in x[floor(d)] + x[ceiling(d)] :
  non-numeric argument to binary operator
> qqPlot(food$weight,main="QQPlot de Peso")
Error in plot.window(...) : valor de 'ylim' no válido
In addition: warning message:
In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion
```

Como se puede observar no fue posible generar el EDA ni siquiera de manera manual ya que este presenta datos erróneos

Resumen estadístico:

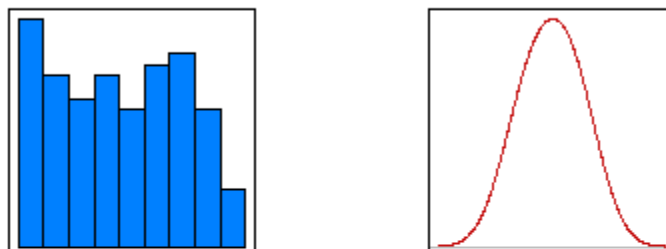
```
> summary(food$weight)
      Length      Class      Mode 
      125 character character
```

Aquí el resumen estadístico es presentado como si la variable fuera categórica por los datos erróneos ingresados

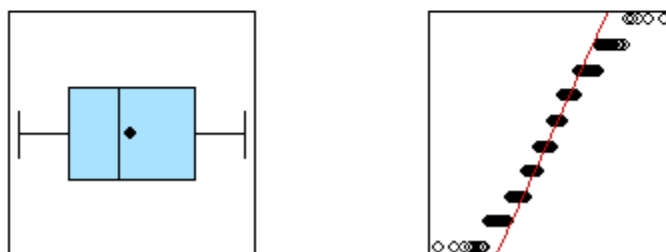
EDA de Healthy Feeling:
 EDA Generado automáticamente:

EXPLORATORY DATA ANALYSIS

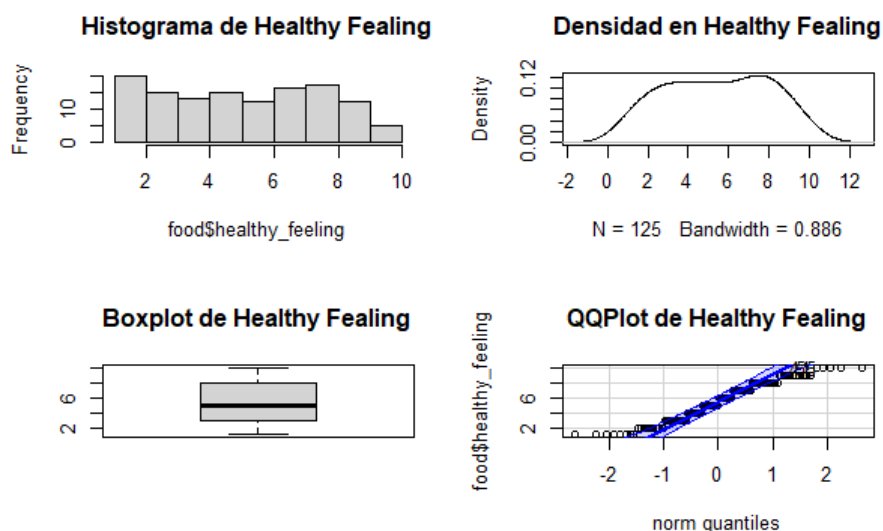
Diagrama de barras de food\$healthy_feeling vs densidad de food\$healthy_feeling



Boxplot de food\$healthy_feeling vs QQ Plot de food\$healthy_feeling



EDA generado manualmente:



Resumen estadístico de Healthy Feeling:

```
summary(food$healthy_feeling)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  3.000   5.000   5.456  8.000  10.000
```

EDA de GPA:

EDA Generado automáticamente:

```
> EDA(food$GPA)
Error in Sort[v3] - Sort[v1] : non-numeric argument to binary operator
```

Como se puede observar, el EDA no puede ser generado de manera automáticamente por los errores presentes en el DataSet

EDA generado manualmente:

```
> par(mfrow=c(2,2))
> hist(food$GPA, main="Histograma de GPA")
Error in hist.default(food$GPA, main = "Histograma de GPA") :
  'x' must be numeric
> plot(density(food$GPA), main="Densidad en GPA")
Error in density.default(food$GPA) : argument 'x' must be numeric
> boxplot(food$GPA, main="Boxplot de GPA")
Error in x[floor(d)] + x[ceiling(d)] :
  non-numeric argument to binary operator
> qqPlot(food$GPA,main="QQPlot de GPA")
Error in plot.window(...) : valor de 'ylim' no válido
In addition: warning message:
In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion
~ |
```

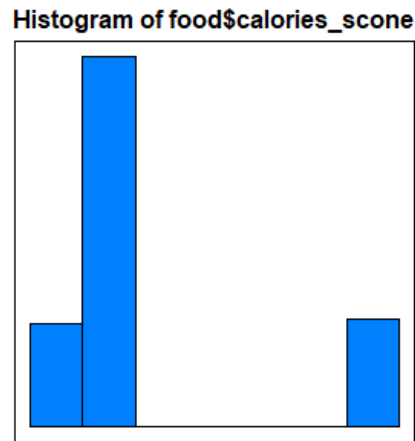
Resumen estadístico de GPA

```
> summary(food$GPA)
      Length      Class      Mode 
      125 character character
```

Aquí el resumen estadístico es presentado como si la variable fuera categórica por los datos erróneos ingresados

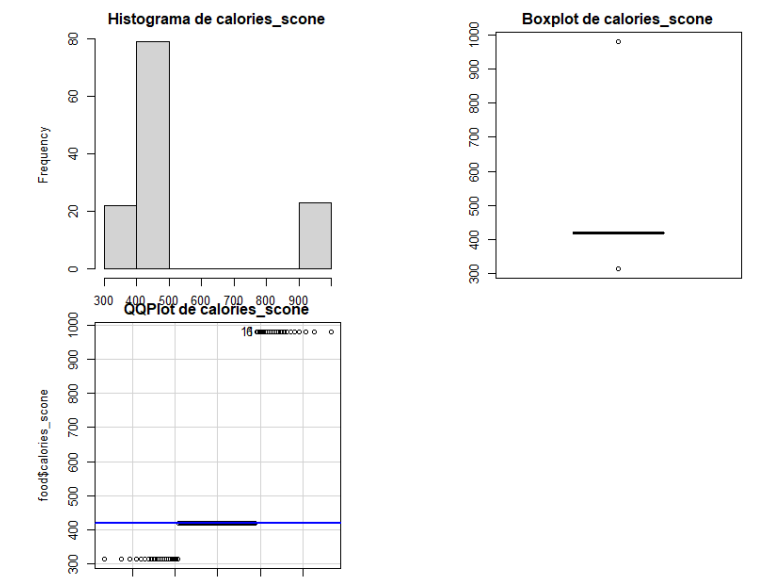
EDA de Calories Scone:

EDA Generado automáticamente:



Como se puede observar, el EDA no pudo ser generado automáticamente debido a los outliers presentes en la data

EDA Generado manualmente:



Se puede observar la presencia de outliers en Calories Day debido a que algunos datos aparecen como Nan

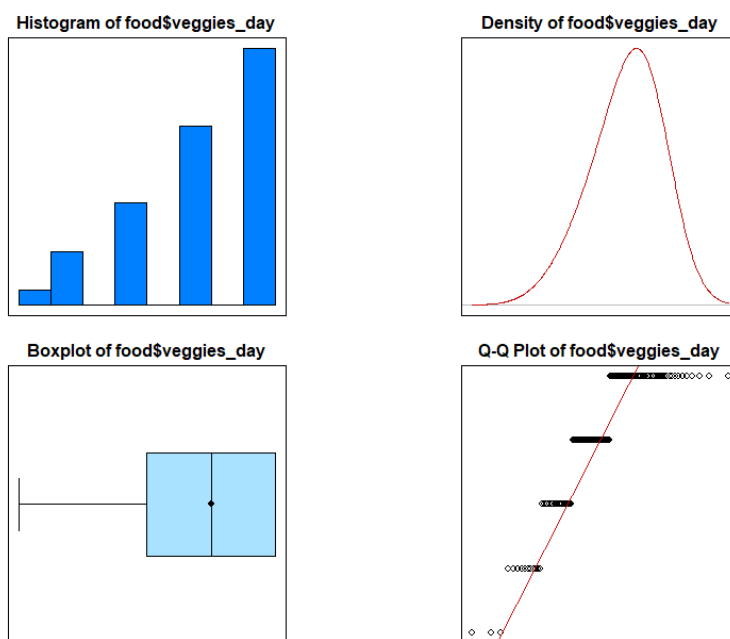
Resumen estadístico:

```
> summary(food$calories_scone)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's 
315.0  420.0   420.0  505.2  420.0   980.0     1
```

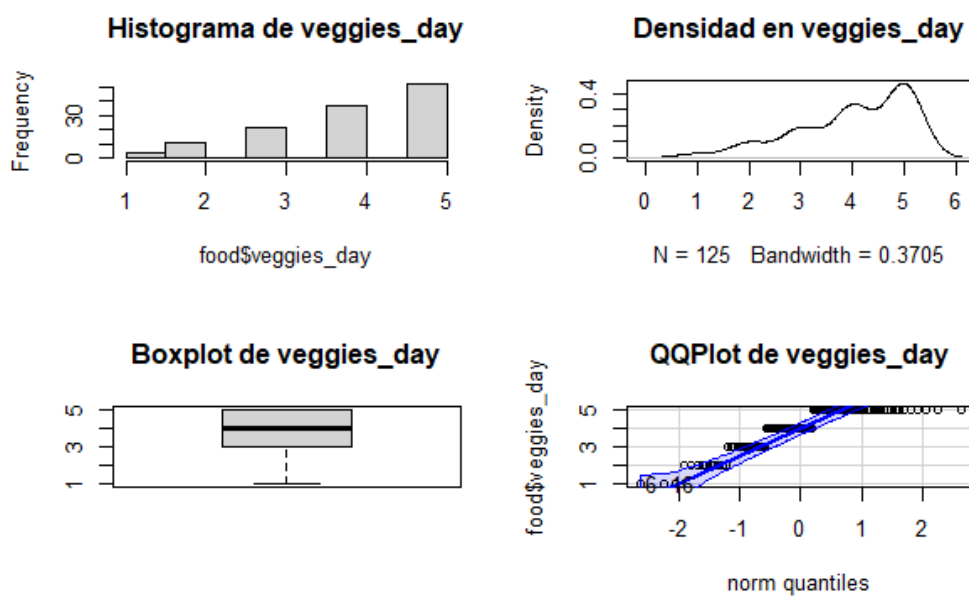
EDA de Veggie Days:

EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS



EDA Generado manualmente:



Resumen estadístico

```
> summary(food$veggies_day)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 1.000  3.000   4.000  4.008  5.000   5.000 
```

DataSet 2 Limpio

> summary(dataset2)

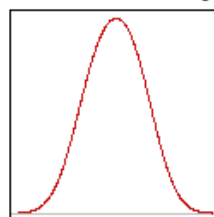
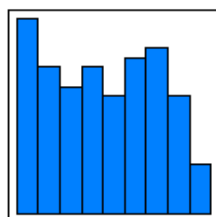
GPA	Gender	breakfast	calories_chicken	calories_day
Min. :2.600	Min. :1.00	Min. :1.000	Min. :265.0	Min. :2.00
1st Qu.:3.200	1st Qu.:1.00	1st Qu.:1.000	1st Qu.:610.0	1st Qu.:3.00
Median :3.500	Median :1.00	Median :1.000	Median :610.0	Median :3.00
Mean :3.455	Mean :1.36	Mean :1.093	Mean :597.5	Mean :3.04
3rd Qu.:3.700	3rd Qu.:2.00	3rd Qu.:1.000	3rd Qu.:665.0	3rd Qu.:3.00
Max. :4.000	Max. :2.00	Max. :2.000	Max. :720.0	Max. :4.00
calories_scone	coffee	comfort_food	cook	cuisine
Min. :315.0	Min. :1.000	Length:75	Min. :1.000	Min. :1.00
1st Qu.:420.0	1st Qu.:1.000	Class :character	1st Qu.:2.000	1st Qu.:1.00
Median :420.0	Median :2.000	Mode :character	Median :3.000	Median :1.00
Mean :504.5	Mean :1.707		Mean :2.747	Mean :1.28
3rd Qu.:420.0	3rd Qu.:2.000		3rd Qu.:3.000	3rd Qu.:1.00
Max. :980.0	Max. :2.000		Max. :5.000	Max. :6.00
diet_current_coded	drink	eating_changes_coded	eating_out	
Min. :1.000	Min. :1.000	Min. :2.00	Min. :1.000	
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:3.00	1st Qu.:2.000	
Median :2.000	Median :2.000	Median :3.00	Median :2.000	
Mean :1.667	Mean :1.547	Mean :4.24	Mean :2.427	
3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:5.00	3rd Qu.:3.000	
Max. :4.000	Max. :2.000	Max. :13.00	Max. :5.000	
employment	ethnic_food	exercise	fav_cuisine	
Min. :1.00	Min. :1.000	Length:75	Length:75	
1st Qu.:2.00	1st Qu.:3.000	Class :character	Class :character	
Median :2.00	Median :4.000	Mode :character	Mode :character	
Mean :2.36	Mean :3.667			
3rd Qu.:3.00	3rd Qu.:5.000			
Max. :3.00	Max. :5.000			
fav_cuisine_coded	fav_food	food_childhood	fries	
Min. :0.000	Min. :1.000	Length:75	Min. :1.00	
1st Qu.:1.000	1st Qu.:1.000	Class :character	1st Qu.:1.00	
Median :1.000	Median :1.000	Mode :character	Median :1.00	
Mean :2.413	Mean :1.707		Mean :1.08	
3rd Qu.:4.000	3rd Qu.:3.000		3rd Qu.:1.00	
Max. :8.000	Max. :3.000		Max. :2.00	
fruit_day	grade_level	greek_food	healthy_feeling	
Min. :2.000	Min. :1.000	Min. :1.000	Min. :1.000	
1st Qu.:4.000	1st Qu.:1.000	1st Qu.:3.000	1st Qu.:3.000	
Median :5.000	Median :2.000	Median :3.000	Median :5.000	
Mean :4.347	Mean :2.307	Mean :3.427	Mean :5.227	
3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:5.000	3rd Qu.:8.000	
Max. :5.000	Max. :4.000	Max. :5.000	Max. :10.000	
healthy_meal	ideal_diet	ideal_diet_coded	income	
Length:75	Length:75	Min. :1.000	Length:75	
Class :character	Class :character	1st Qu.:2.000	Class :character	
Mode :character	Mode :character	Median :3.000	Mode :character	
		Mean :3.547		
		3rd Qu.:5.000		
		Max. :8.000		
indian_food	italian_food	life_rewarding	marital_status	
Min. :1.00	Min. :3.000	Length:75	Min. :1.00	
1st Qu.:2.00	1st Qu.:5.000	Class :character	1st Qu.:1.00	
Median :3.00	Median :5.000	Mode :character	Median :1.00	
Mean :2.92	Mean :4.747		Mean :1.48	
3rd Qu.:4.00	3rd Qu.:5.000		3rd Qu.:2.00	
Max. :5.00	Max. :5.000		Max. :2.00	

EDA de Calories Day:

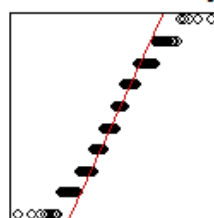
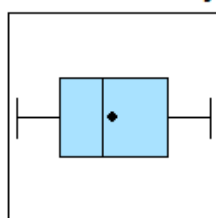
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

logram of food\$healthy_feelnsity of food\$healthy_feelin

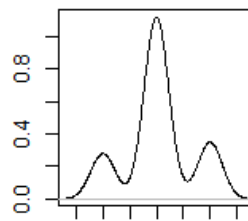
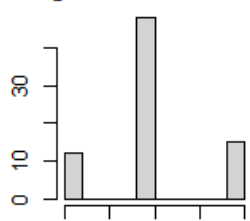


boxplot of food\$healthy_feelinQ Plot of food\$healthy_feelin

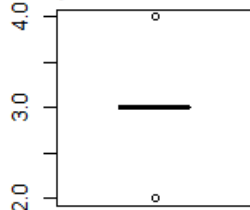


EDA generado manualmente

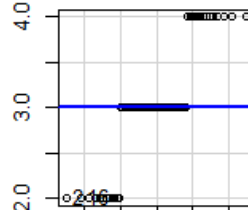
Histograma de Calories Day Densidad en Calories Day



Boxplot de Calories Day



QQPlot de Calories Day



Resumen Estadístico

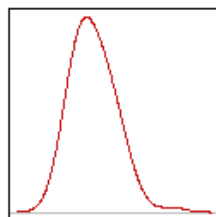
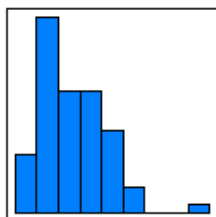
```
summary(Dataset2$calories_day)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.00   3.00   3.00   3.04   3.00   4.00
non(mfrow = c(2, 2))
```

EDA de Weight:

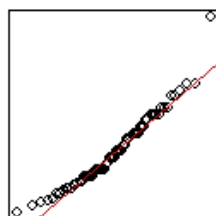
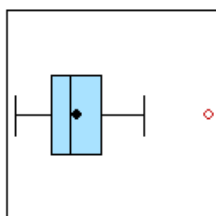
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

listogram of Dataset2\$weightDensity of Dataset2\$weight

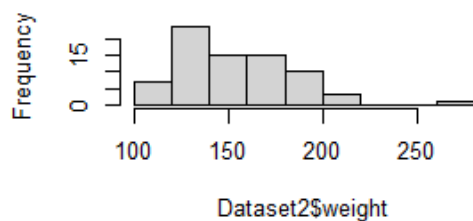


Boxplot of Dataset2\$weightQ-Q Plot of Dataset2\$weight

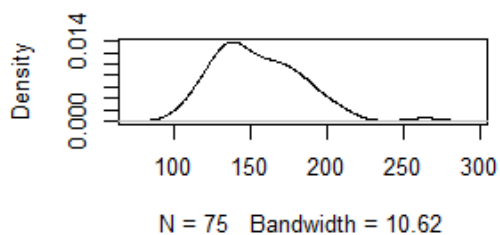


EDA generado manualmente:

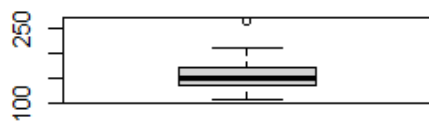
Histograma de Peso



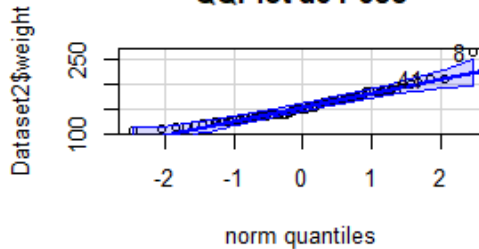
Densidad en Peso



Boxplot de Peso



QQPlot de Peso

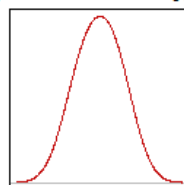
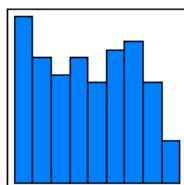


EDA de Calories Scone:

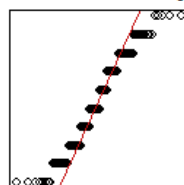
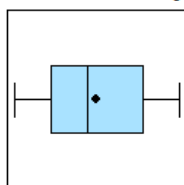
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

logram of food\$healthy_feelnsity of food\$healthy_feelir

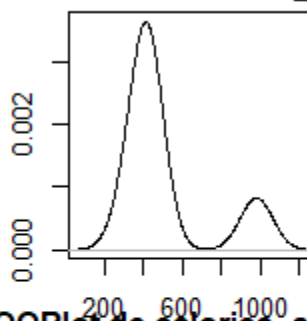
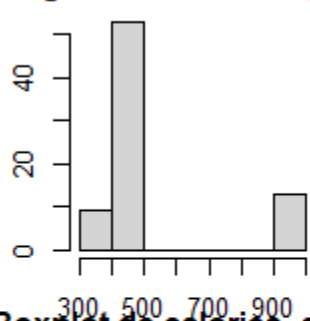


boxplot of food\$healthy_feelirQQ Plot of food\$healthy_feelir

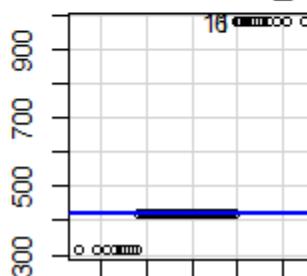
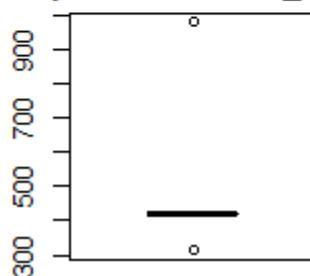


EDA generado manualmente

histograma de calories_sconeDensidad en calories_scone



Boxplot de calories_scone QQPlot de calories_scone

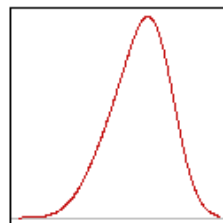
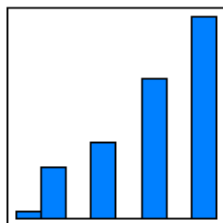


EDA de Veggie Days:

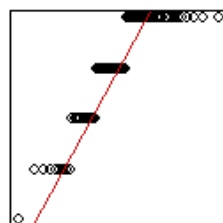
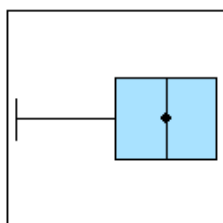
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

ogram of Dataset2\$veggies_nsity of Dataset2\$veggies_d

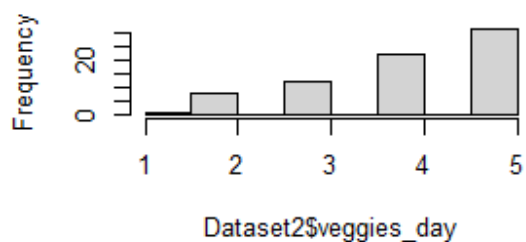


xplot of Dataset2\$veggies_d Plot of Dataset2\$veggies_c

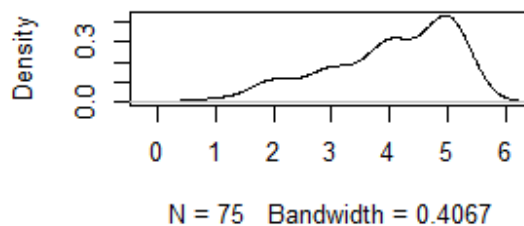


EDA generado manualmente

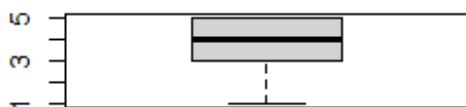
Histograma de veggies_day



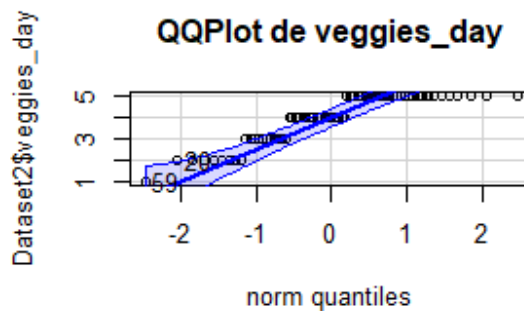
Densidad en veggies_day



Boxplot de veggies_day



QQPlot de veggies_day



DataSet 3

EDA de Datasets 3

DataSet 3 no Limpio

```
> summary(School1)
world_rank      university_name      country      teaching
Length:2603    Length:2603          Length:2603   Min.   : 9.9
Class :character Class :character    Class :character 1st Qu.:24.7
Mode  :character Mode  :character    Mode  :character Median :33.9
                                           Mean  :37.8
                                           3rd Qu.:46.4
                                           Max.   :99.7

international    research      citations      income
Length:2603      Min.   : 2.90    Min.   : 1.20   Length:2603
Class :character 1st Qu.:19.60    1st Qu.: 45.50  Class :character
Mode  :character Median :30.50     Median : 62.50  Mode  :character
                                           Mean  :35.91
                                           3rd Qu.:47.25
                                           Max.   :99.40
                                           Max.   :100.00

total_score      num_students      student_staff_ratio international_students
Length:2603      Length:2603      Min.   : 0.60   Length:2603
Class :character Class :character 1st Qu.: 11.97  Class :character
Mode  :character Mode  :character Median : 16.10   Mode  :character
                                           Mean  : 18.45
                                           3rd Qu.: 21.50
                                           Max.   :162.60
                                           NA's   :59

female_male_ratio year
Length:2603      Min.   :2011
Class :character 1st Qu.:2013
Mode  :character Median :2014
                                           Mean  :2014
                                           3rd Qu.:2016
                                           Max.   :2016

> summary(School2)
world_rank      university_name      national_rank      total_score
Length:4897     Length:4897          Length:4897        Min.   : 23.50
Class :character Class :character    Class :character   1st Qu.: 27.40
Mode  :character Mode  :character    Mode  :character   Median : 31.30
                                           Mean  : 36.38
                                           3rd Qu.: 41.80
                                           Max.   :100.00
                                           NA's   :3796

alumni          award          hici          ns
Min.   : 0.000    Min.   : 0.000    Min.   : 0.00    Min.   : 0.00
1st Qu.: 0.000    1st Qu.: 0.000    1st Qu.: 7.30    1st Qu.: 8.00
Median : 0.000    Median : 0.000    Median : 12.60   Median : 12.80
Mean   : 9.162    Mean   : 7.692    Mean   : 16.22   Mean   : 16.08
3rd Qu.: 15.600   3rd Qu.: 13.400   3rd Qu.: 21.70   3rd Qu.: 19.80
Max.   :100.000   Max.   :100.000   Max.   :100.00   Max.   :100.00
NA's   :1         NA's   :2         NA's   :2        NA's   :22

pub            pcp            year
Min.   : 7.30    Min.   : 8.30    Min.   :2005
1st Qu.: 28.90   1st Qu.: 15.60   1st Qu.:2007
Median : 36.00   Median : 19.00   Median :2009
Mean   : 38.25   Mean   : 21.24   Mean   :2010
3rd Qu.: 45.30   3rd Qu.: 24.50   3rd Qu.:2012
Max.   :100.00   Max.   :100.00   Max.   :2015
NA's   :2        NA's   :2

> summary(School3)
school_name      country
Length:818       Length:818
Class :character Class :character
Mode  :character Mode  :character
```

```
> summary(school4)
```

country_name	series_name	X1985		X1986	
Length:79055	Length:79055	Min. :	0.0	Min. :	3.74
Class :character	Class :character	1st Qu.:	1.0	1st Qu.:	4.84
Mode :character	Mode :character	Median :	5.1	Median :	5.58
		Mean :	253.1	Mean :	6.37
		3rd Qu.:	23.2	3rd Qu.:	7.26
		Max. :	745898.0	Max. :	10.19
		NA's :	27212	NA's :	79043

X1987		X1990		X1991		X1992		X1993	
Min. :	1.12	Min. :	0.0	Min. :	7.58	Min. :	5.51	Min. :	6.22
1st Qu.:	1.54	1st Qu.:	1.2	1st Qu.:	7.58	1st Qu.:	5.64	1st Qu.:	6.22
Median :	1.96	Median :	5.7	Median :	7.58	Median :	5.76	Median :	6.22
Mean :	1.99	Mean :	280.1	Mean :	7.58	Mean :	5.76	Mean :	6.22
3rd Qu.:	2.42	3rd Qu.:	24.1	3rd Qu.:	7.58	3rd Qu.:	5.88	3rd Qu.:	6.22
Max. :	2.88	Max. :	835430.0	Max. :	7.58	Max. :	6.00	Max. :	6.22
NA's :	79052	NA's :	27209	NA's :	79054	NA's :	79052	NA's :	79054

X1995		X1996		X1997		X1998		X1999	
Min. :	0.0	Min. :	7.50	Mode:logical	Min. :	0.00	Min. :	9.48	
1st Qu.:	1.5	1st Qu.:	7.65	NA's:79055	1st Qu.:	0.35	1st Qu.:	9.66	
Median :	6.2	Median :	7.80		Median :	2.39	Median :	9.88	
Mean :	306.8	Mean :	7.80		Mean :	8.81	Mean :	10.01	
3rd Qu.:	24.9	3rd Qu.:	7.95		3rd Qu.:	9.86	3rd Qu.:	10.02	
Max. :	896920.0	Max. :	8.10		Max. :	55.26	Max. :	11.14	
NA's :	27171	NA's :	79052		NA's :	79008	NA's :	79046	

X2000		X2001		X2002		X2003		X2004	
Min. :	0.0	Min. :	0.00	Min. :	0.00	Min. :	0.16	Min. :	0.00
1st Qu.:	1.7	1st Qu.:	3.09	1st Qu.:	1.56	1st Qu.:	6.91	1st Qu.:	3.55
Median :	6.8	Median :	9.07	Median :	7.46	Median :	11.25	Median :	7.30
Mean :	334.8	Mean :	16.35	Mean :	16.28	Mean :	19.53	Mean :	11.74
3rd Qu.:	25.2	3rd Qu.:	19.48	3rd Qu.:	23.97	3rd Qu.:	28.35	3rd Qu.:	10.70
Max. :	958307.0	Max. :	98.26	Max. :	91.68	Max. :	90.41	Max. :	85.40
NA's :	27085	NA's :	78830	NA's :	78871	NA's :	79000	NA's :	78918

X2005		X2006		X2007		X2008	
Min. :	0.0	Min. :	0.13	Min. :	0.00	Min. :	0.00
1st Qu.:	1.9	1st Qu.:	5.86	1st Qu.:	3.06	1st Qu.:	5.41
Median :	7.2	Median :	8.62	Median :	9.23	Median :	9.46
Mean :	363.5	Mean :	13.17	Mean :	16.07	Mean :	14.85
3rd Qu.:	26.4	3rd Qu.:	15.20	3rd Qu.:	16.02	3rd Qu.:	14.59
Max. :	1034076.0	Max. :	93.69	Max. :	99.99	Max. :	97.70
NA's :	27018	NA's :	78851	NA's :	78769	NA's :	78689

X2009		X2010		X2011		X2012	
Min. :	0.00	Min. :	0.0	Min. :	0.00	Min. :	0.00
1st Qu.:	4.46	1st Qu.:	2.0	1st Qu.:	2.74	1st Qu.:	1.50
Median :	10.17	Median :	7.5	Median :	10.80	Median :	10.60
Mean :	17.53	Mean :	386.6	Mean :	20.35	Mean :	19.48
3rd Qu.:	17.46	3rd Qu.:	27.0	3rd Qu.:	25.53	3rd Qu.:	26.93
Max. :	98.72	Max. :	1090693.0	Max. :	99.52	Max. :	98.68
NA's :	78591	NA's :	26259	NA's :	78126	NA's :	77949

X2013		X2015	
Min. :	0.00	Mode:logical	
1st Qu.:	1.05	NA's:79055	
Median :	11.88		
Mean :	23.11		
3rd Qu.:	31.54		
Max. :	100.00		
NA's :	77176		

```
> summary(school5)
```

country	institute_type	direct_expenditure_type	x1995
Length:333	Length:333	Length:333	Min. :0.100
Class :character	Class :character	Class :character	1st Qu.:1.100
Mode :character	Mode :character	Mode :character	Median :3.400
			Mean :3.106
			3rd Qu.:4.500
			Max. :6.800
			NA's :239

x2000	x2005	x2009	x2010	x2011
Min. :0.500	Min. :0.300	Min. :0.300	Min. :0.500	Min. :0.000
1st Qu.:1.200	1st Qu.:1.200	1st Qu.:1.350	1st Qu.:1.300	1st Qu.:0.700
Median :3.300	Median :3.450	Median :3.700	Median :3.600	Median :1.750
Mean :3.063	Mean :3.142	Mean :3.388	Mean :3.392	Mean :2.573
3rd Qu.:4.350	3rd Qu.:4.300	3rd Qu.:4.850	3rd Qu.:4.800	3rd Qu.:4.300
Max. :6.600	Max. :7.200	Max. :7.500	Max. :7.600	Max. :7.900
NA's :234	NA's :227	NA's :230	NA's :232	NA's :51

```
> summary(school6)
```

world_rank	institution	country	national_rank
Min. : 1.0	Length:2200	Length:2200	Min. : 1.00
1st Qu.: 175.8	Class :character	Class :character	1st Qu.: 6.00
Median : 450.5	Mode :character	Mode :character	Median : 21.00
Mean : 459.6			Mean : 40.28
3rd Qu.: 725.2			3rd Qu.: 49.00
Max. :1000.0			Max. :229.00

quality_of_education	alumni_employment	quality_of_faculty	publications
Min. : 1.0	Min. : 1.0	Min. : 1.0	Min. : 1.0
1st Qu.:175.8	1st Qu.:175.8	1st Qu.:175.8	1st Qu.: 175.8
Median :355.0	Median :450.5	Median :210.0	Median : 450.5
Mean :275.1	Mean :357.1	Mean :178.9	Mean : 459.9
3rd Qu.:367.0	3rd Qu.:478.0	3rd Qu.:218.0	3rd Qu.: 725.0
Max. :367.0	Max. :567.0	Max. :218.0	Max. :1000.0

influence	citations	broad_impact	patents	score
Min. : 1.0	Min. : 1.0	Min. : 1.0	Min. : 1.0	Min. : 43.36
1st Qu.:175.8	1st Qu.:161.0	1st Qu.: 250.5	1st Qu.:170.8	1st Qu.: 44.46
Median :450.5	Median :406.0	Median : 496.0	Median :426.0	Median : 45.10
Mean :459.8	Mean :413.4	Mean : 496.7	Mean :433.3	Mean : 47.80
3rd Qu.:725.2	3rd Qu.:645.0	3rd Qu.: 741.0	3rd Qu.:714.2	3rd Qu.: 47.55
Max. :991.0	Max. :812.0	Max. :1000.0	Max. :871.0	Max. :100.00
		NA's :200		

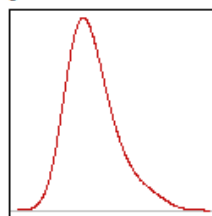
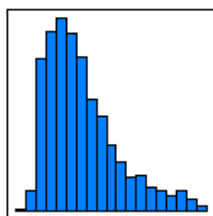
```
year
Min. :2012
1st Qu.:2014
Median :2014
Mean :2014
3rd Qu.:2015
Max. :2015
```

EDA de Teaching:

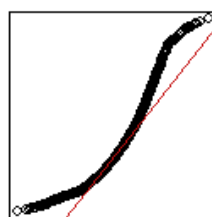
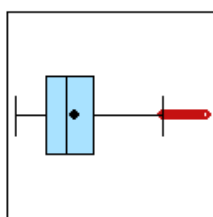
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

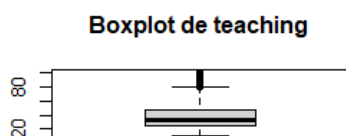
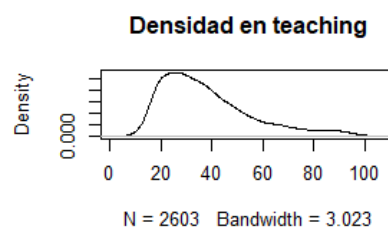
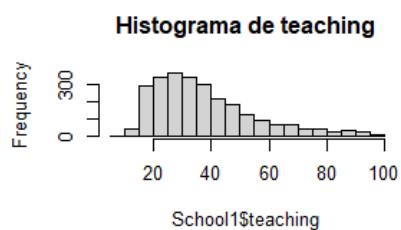
Histogram of School1\$teaching Density of School1\$teaching



Boxplot of School1\$teaching Q-Plot of School1\$teaching



EDA Generado manualmente



Como se puede visualizar, en los datos hay outliers que dificultan el correcto análisis de datos

Resumen Estadístico:

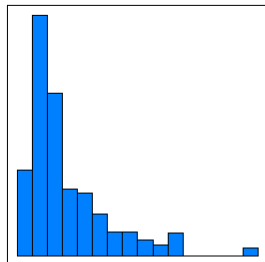
```
> summary(School1$teaching)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   9.9   24.7   33.9   37.8   46.4   99.7
```

EDA de Total Score:

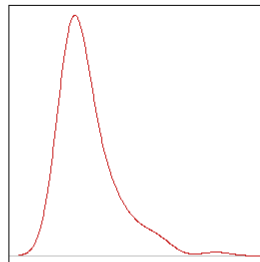
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

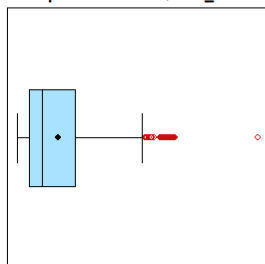
Histogram of School2\$total_score



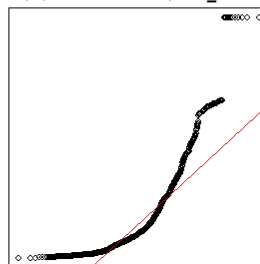
Density of School2\$total_score



Boxplot of School2\$total_score

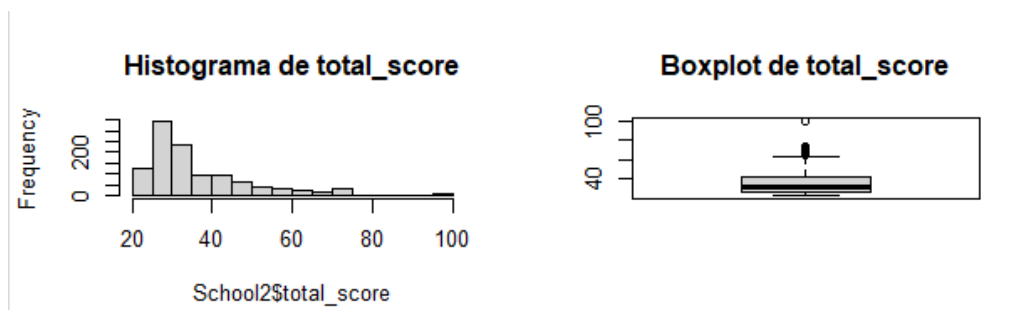


Q-Q Plot of School2\$total_score



Se puede observar la presencia de outliers en total score debido a que algunos datos aparecen como Nan

EDA Generado manualmente



Se puede observar la presencia de outliers en total score y debido a que algunos datos aparecen como Nan, por lo que tambien fue imposible generar todas las graficas

Resumen estadístico:

```
summary(School2$total_score)
```

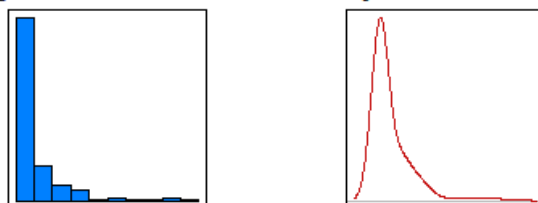
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
23.50	27.40	31.30	36.38	41.80	100.00	3796

EDA de Award:

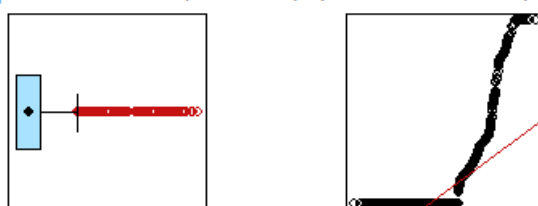
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

Histogram of School2\$award Density of School2\$award

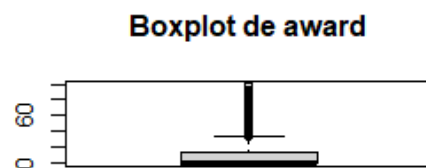
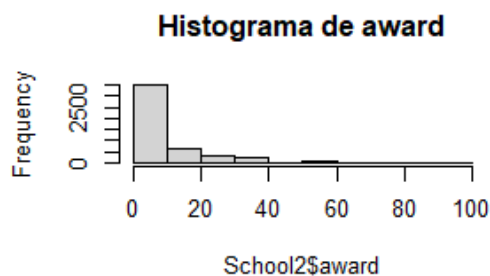


Boxplot of School2\$award Q-Q Plot of School2\$award



Se puede observar la presencia de outliers en Award debido a que algunos datos aparecen como Nan

EDA Generado manualmente:



Se puede observar la presencia de outliers en Award debido a que algunos datos aparecen como Nan, eso provoca que no se generen las demás graficas

Resumen Estadístico

```
summary(School2$award)
```

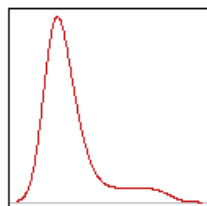
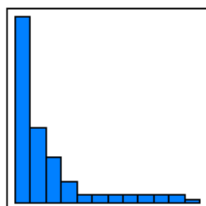
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.000	0.000	0.000	7.692	13.400	100.000	2

EDA de National Rank:

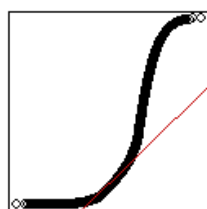
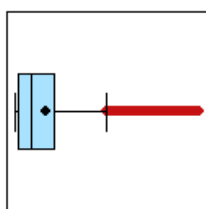
EDA Generado automaticamente:

EXPLORATORY DATA ANALYSIS

Diagram of School6\$national_rank vs Density of School6\$national_rank

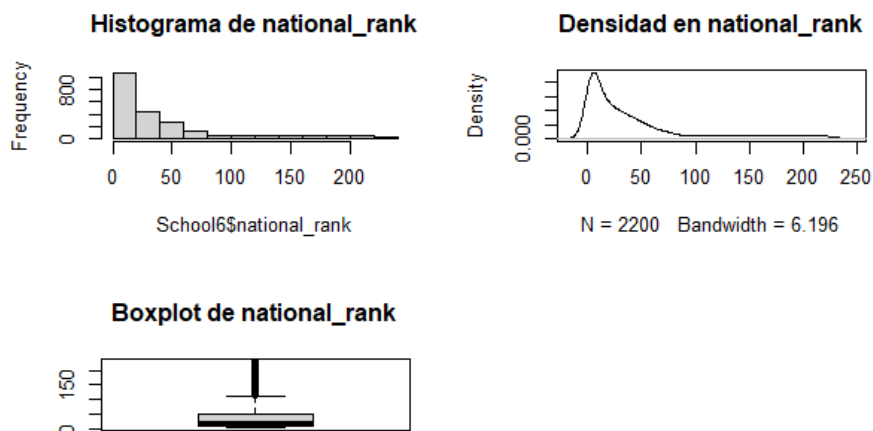


Boxplot of School6\$national_rank vs Plot of School6\$national_rank



Se puede observar la presencia de outliers en National Rank debido a que algunos datos aparecen como Nan

EDA generado Manualmente



Se puede observar la presencia de outliers en Award debido a que algunos datos aparecen como Nan, eso provoca que no se generen las demás graficas

Resumen Estadístico

```
summary(School6$national_rank)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.00   6.00   21.00   40.28  49.00  229.00
```

Conclusión

Sin duda después de la limpieza de datos se verifico como ya habían muy pocos outliers, y como todas las graficas fueron resueltas por R , cuando antes no se podían procesar por la data ilegible que traían los archivos de entrada.