

prueba

Rafael Chaidez Rosas

15/10/2021

Objetivo

Analizar datos de personas ## Cargar datos Se cargan los datos de la direccion <https://raw.githubusercontent.com/rpizarrog/CIIT.-Diplomado-en-Ciencia-de-los-Datos-e-IoT/main/M%C3%B3dulo%20I/SESI%C3%93N%206/scripts/generar%20personas.r>

```
source("https://raw.githubusercontent.com/rpizarrog/CIIT.-Diplomado-en-Ciencia-de-los-Datos-e-IoT/main/I
```

Mostrar los primero y ultimo registros

Utilizar las funciones head() y tail() para mostrar los primeros seis y ultimo registros.

```
head(datos.personas)
```

```
##      edad  genero      estado  feliz
## 1     21 FEMENINO BAJA CALIFORNIA NO FELIZ
## 2     18 FEMENINO      NUEVO LEÓN NO FELIZ
## 3     30 FEMENINO      TAMAULIPAS  FELIZ
## 4     23 FEMENINO      NUEVO LEÓN NO FELIZ
## 5     47 MASCULINO      TAMAULIPAS  FELIZ
## 6     38 FEMENINO      DURANGO NO FELIZ
```

```
tail(datos.personas)
```

```
##      edad  genero      estado  feliz
## 9995    28 FEMENINO BAJA CALIFORNIA  FELIZ
## 9996    58 MASCULINO BAJA CALIFORNIA NO FELIZ
## 9997    44 FEMENINO      SONORA  FELIZ
## 9998    23 FEMENINO      COAHUILA NO FELIZ
## 9999    23 FEMENINO      SONORA NO FELIZ
## 10000   39 MASCULINO      NUEVO LEÓN  FELIZ
```

Factorizar datos de personas

```
datos.personas$genero <- as.factor(datos.personas$genero)
datos.personas$estado <- as.factor(datos.personas$estado)
datos.personas$feliz <- as.factor(datos.personas$feliz)
```

Describir los datos

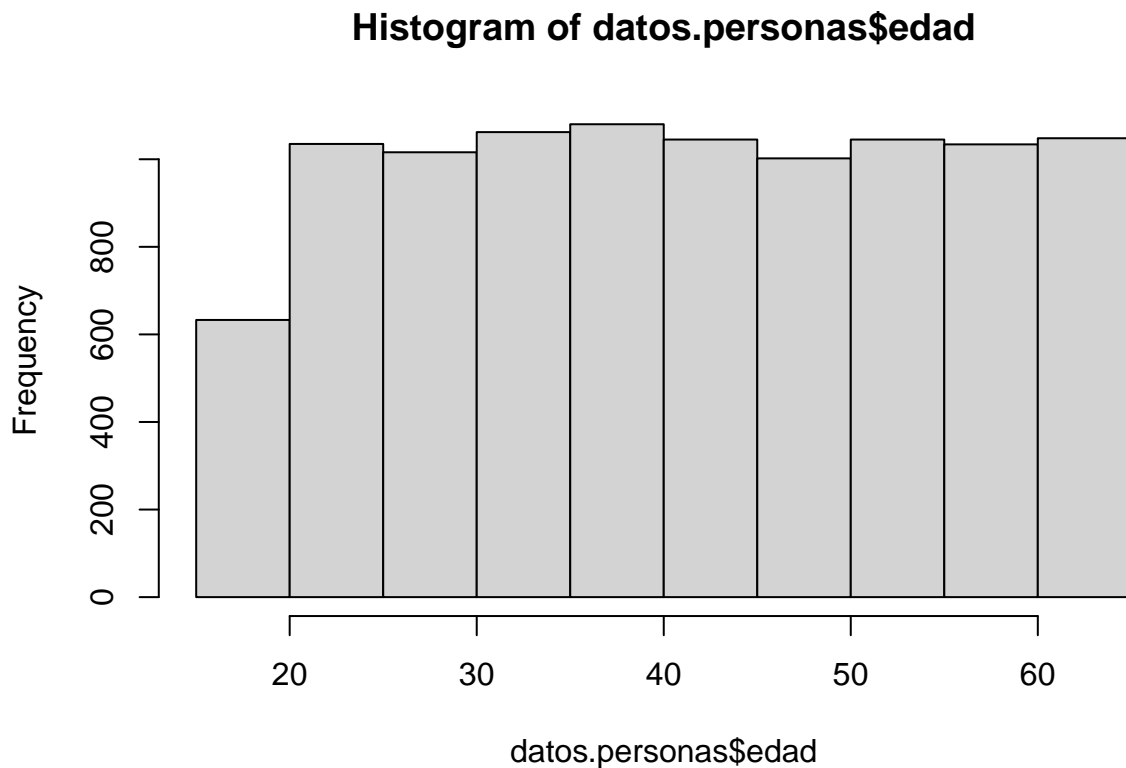
```
summary(datos.personas)
```

```
##      edad      genero      estado      feliz
```

```
## Min.      :18.0    FEMENINO :5215    DURANGO      :1278    FELIZ      :4950
## 1st Qu.   :30.0    MASCULINO:4785    NUEVO LEÓN   :1276    NO FELIZ:5050
## Median    :41.0
## Mean      :41.5
## 3rd Qu.   :53.0
## Max.      :65.0
##                                     COAHUILA     :1271
##                                     BAJA CALIFORNIA:1257
##                                     TAMAULIPAS   :1251
##                                     (Other)       :2400
```

Histograma de edades

```
hist(datos.personas$edad)
```



Frecuencia de genero

```
frecuencia.genero <- data.frame(table(datos.personas$genero))
frecuencia.genero
```

```
##      Var1 Freq
## 1 FEMENINO 5215
## 2 MASCULINO 4785
```

Grafica de barras de la variable genero

```
barplot(height = frecuencia.genero$Freq, names.arg = frecuencia.genero$Var1)
```

