

Tipología y ciclo de vida de los datos: Practica 2

Autor: Borja López Gómez y Sergio Beltrán Nuez

Diciembre 2021

Contents

1. Descripción del dataset	2
Origen de datos	2
Motivación	2
Carga inicial	2
2. Integración y selección de los datos de interés	2
Análisis de variables	2
3. Limpieza de datos	5
Transformación de variables	5
3.1. Eliminación de valores nulos o no informados	5
3.2. Identificación y tratamiento de valores extremos	7
4. Análisis de los datos	11
4.1. Selección de los datos	11
4.2 Normalidad y homogeneidad de varianzas y 4.3 Contraste de hipótesis, correlación y regresión	12
Supuesto de normalidad	14
Contraste de hipótesis	15
Supuesto de homocedasticidad	15
5. Representación de los resultados	16
Representación gráfica	16
6. Resolución del problema y conclusiones	19
7. Tabla de contribuciones	20

1. Descripción del dataset

Origen de datos

Para la realización de esta practica se propone la utilización del dataset Cervical Cancer Risk Classification, un dataset que recoge diferentes marcadores relacionados con la aparición del cancer de cuello uterino. El dataset puede encontrarse tanto en el enlace de Kaggle <https://www.kaggle.com/loveall/cervical-cancer-risk-classification> como en el repositorio UCI (<https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29>) Se trata de un dataset de origen público publicado por el Hospital Universitario de Caracas, Venezuela.

Motivación

Pese a que se trata de unos de los canceres más prevenibles y el número de nuevos casos ha disminuido constantemente en los últimos años, aproximadamente 4000 mujeres en Estado Unidos y 300000 en todo el mundo son diagnosticadas cada año. La mortalidad provocada por este cancer se ha reducido de manera notable gracias a las pruebas de cribado. No obstante, numerosos estudios confirman que el nivel de pobreza y otros factores socioeconómicos están relacionados con bajas tasas de cribado. Por todo esto consideramos que se trata de un tema de actualidad que requiere de mucha colaboración ya que entender bien la información de que disponemos puede facilitar el estudio de futuros trabajos.

Carga inicial

Una vez obtenido el dataset el primer paso sería cargar los datos, entender de que tipo de información disponemos, analizar las variables y realizar un primer análisis de la información para ver la calidad de los datos y obtener las primeras conclusiones.

Para poder ilustrar el ejemplo, cargamos el fichero de datos

```
cervical_data <- read.csv('../data/kag_risk_factors_cervical_cancer.csv', stringsAsFactors = FALSE)
rows=dim(cervical_data)[1]
```

Instalamos y cargamos las librerías ggplot2 y dplyr

```
# https://cran.r-project.org/web/packages/ggplot2/index.html
if (!require('ggplot2')) install.packages('ggplot2'); library('ggplot2')
# https://cran.r-project.org/web/packages/dplyr/index.html
if (!require('dplyr')) install.packages('dplyr'); library('dplyr')
```

2. Integración y selección de los datos de interés

Análisis de variables

En primer lugar comprobamos la estructura del fichero:

```
str(cervical_data)
```

```
## 'data.frame': 858 obs. of 36 variables:
## $ Age : int 18 15 34 52 46 42 51 26 45 44 ...
## $ Number.of.sexual.partners : chr "4.0" "1.0" "1.0" "5.0" ...
## $ First.sexual.intercourse : chr "15.0" "14.0" "?" "16.0" ...
## $ Num.of.pregnancies : chr "1.0" "1.0" "1.0" "4.0" ...
## $ Smokes : chr "0.0" "0.0" "0.0" "1.0" ...
## $ Smokes..years. : chr "0.0" "0.0" "0.0" "37.0" ...
## $ Smokes..packs.year. : chr "0.0" "0.0" "0.0" "37.0" ...
## $ Hormonal.Contraceptives : chr "0.0" "0.0" "0.0" "1.0" ...
## $ Hormonal.Contraceptives..years. : chr "0.0" "0.0" "0.0" "3.0" ...
## $ IUD : chr "0.0" "0.0" "0.0" "0.0" ...
## $ IUD..years. : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs..number. : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.condylomatosis : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.cervical.condylomatosis : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.vaginal.condylomatosis : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.vulvo.perineal.condylomatosis: chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.syphilis : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.pelvic.inflammatory.disease : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.genital.herpes : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.molluscum.contagiosum : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.AIDS : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.HIV : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.Hepatitis.B : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs.HPV : chr "0.0" "0.0" "0.0" "0.0" ...
## $ STDs..Number.of.diagnosis : int 0 0 0 0 0 0 0 0 0 ...
## $ STDs..Time.since.first.diagnosis : chr "?" "?" "?" "?" ...
## $ STDs..Time.since.last.diagnosis : chr "?" "?" "?" "?" ...
## $ Dx.Cancer : int 0 0 0 1 0 0 0 0 1 0 ...
## $ Dx.CIN : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Dx.HPV : int 0 0 0 1 0 0 0 0 1 0 ...
## $ Dx : int 0 0 0 0 0 0 0 0 1 0 ...
## $ Hinselmann : int 0 0 0 0 0 0 1 0 0 0 ...
## $ Schiller : int 0 0 0 0 0 0 1 0 0 0 ...
## $ Citology : int 0 0 0 0 0 0 0 0 0 0 ...
## $ Biopsy : int 0 0 0 0 0 0 1 0 0 0 ...
```

Se puede observar como existen 858 observaciones y un total de 36 variables que corresponden a los diferentes indicadores a tener en cuenta en la identificación del cancer de cuello de útero. La descripción de las variables es la siguiente:

Age Edad del paciente

Number of sexual partners Numero de personas con las que ha mantenido relaciones

First sexual intercourse Edad a la que mantuvo su primera relación

Num of pregnancies Número de embarazos

Smokes Indicador de si es fumadora (1/0)

Smokes years Número de años siendo fumadora

Smokes pack years Valor que cuantifica el consumo de tabaco

Hormonal contraceptives Utilización de anticonceptivos hormonales (1/0)

Hormonal Contraceptives years Utilización de anticonceptivos hormonales en años

IUD Dispositivo intrauterino (1/0)

IUD years Dispositivo intrauterino en años

STDs Enfermedades de transmisión sexual (ETS)

STDs number Número de enfermedades (ETS)

STDs condylomatosis ETS condilomatosis (1/0)

STDs cervical.condylomatosis ETS condilomatosis cervical (1/0)

STDs vaginal.condylomatosis ETS condilomatosis vaginal (1/0)

STDs vulvo.perineal.condylomatosis ETS condilomatosis vulvo perineal (1/0)

STDs syphilis ETS sífilis (1/0)

STDs pelvic.inflammatory.disease ETS enfermedad pélvica inflamatoria (1/0)

STDs genital.herpès ETS herpes (1/0)

STDs molluscum.contagiosum ETS molusco contagioso (1/0)

STDs AIDS ETS SIDA (1/0)

STDs HIV ETS VIH (1/0)

STDs Hepatitis.B ETS Hepatitis B (1/0)

STDs HPV ETS Virus del papiloma humano (1/0)

STDs..Number.of.diagnosis Número de ETS diagnosticadas

STDs..Time.since.first.diagnosis Tiempo desde la primera vez diagnosticada

STDs..Time.since.last.diagnosis Tiempo desde la última vez diagnosticada

Dx.Cancer Diagnostico Cancer (1/0)

Dx.CIN Diagnostico lesiones precancerosas(1/0)

Dx.HPV Diagnostico VPH (Virus del papiloma humano)

Dx Existe diagnostico

Hinselmann Prueba de Hinselmann (1/0)

Schiller Prueba de Schiller(1/0)

Citology Prueba Citología (1/0)

Biopsy Prueba Biopsia (1/0)

Selección de variables

Una vez identificadas las variables, el siguiente paso sería realizar un proceso de limpieza de los datos. Descartar los datos con mala calidad o que, a priori, pensemos no aportan mucho valor en los resultados esperados. En este caso concreto podríamos eliminar las columnas relacionadas a los diferentes tipos de ETS dejando únicamente el indicador general que informa si ha padecido alguna enfermedad de este tipo y guardar los datos en un subset adicional por si fuera necesario lanzar análisis más detallados en el futuro.

```
cervical_data_subset<-cervical_data[c(1:13,26:36)]
```

3. Limpieza de datos

Transformación de variables

Para poder obtener resultados estadísticos de los datos es necesario transformar algunos datos que viene como string a tipo entero.

```
cervical_data_subset$Number.of.sexual.partners<-as.integer(cervical_data_subset$Number.of.sexual.partners)
cervical_data_subset$First.sexual.intercourse<-as.integer(cervical_data_subset$First.sexual.intercourse)
cervical_data_subset$Smokes<-as.integer(cervical_data_subset$Smokes)
cervical_data_subset$Smokes..years.<-as.integer(cervical_data_subset$Smokes..years.)
cervical_data_subset$Num.of.pregnancies<-as.integer(cervical_data_subset$Num.of.pregnancies)
```

Otro paso dentro de la limpieza de datos podría ser eliminar los datos redundantes o duplicados, en este caso concreto, al no tener ningún campo que identifique al paciente como único aceptaremos la premisa de que diferentes pacientes pueden repetir diferentes valores y por lo tanto no eliminaremos los registros duplicados.

3.1. Eliminación de valores nulos o no informados

Siguiendo con el proceso de limpieza, el siguiente paso sería identificar los valores nulos, no informados, o informados con caracteres extraños. En el caso de este dataset, se identifica que el caracter extraño más común es el signo de interrogación (?).

Estadísticas de valores vacíos

```
colSums(is.na(cervical_data_subset))
```

```
##              Age      Number.of.sexual.partners
##              0              26
##      First.sexual.intercourse      Num.of.pregnancies
##              7              56
##              Smokes      Smokes..years.
##              13              13
##      Smokes..packs.year.      Hormonal.Contraceptives
##              0              0
##      Hormonal.Contraceptives..years.      IUD
##              0              0
##              IUD..years.      STDs
##              0              0
##      STDs..number.      STDs..Number.of.diagnosis
##              0              0
##      STDs..Time.since.first.diagnosis      STDs..Time.since.last.diagnosis
##              0              0
##              Dx.Cancer      Dx.CIN
##              0              0
##              Dx.HPV      Dx
##              0              0
##              Hinselmann      Schiller
##              0              0
##              Citology      Biopsy
##              0              0
```

Estadísticas de valores nulos

```
colSums(cervical_data_subset=="")
```

```
##           Age           Number.of.sexual.partners
##           0              NA
## First.sexual.intercourse      Num.of.pregnancies
##           NA              NA
##           Smokes           Smokes..years.
##           NA              NA
##           Smokes..packs.year.      Hormonal.Contraceptives
##           0              0
## Hormonal.Contraceptives..years.      IUD
##           0              0
##           IUD..years.           STDs
##           0              0
##           STDs..number.      STDs..Number.of.diagnosis
##           0              0
## STDs..Time.since.first.diagnosis  STDs..Time.since.last.diagnosis
##           0              0
##           Dx.Cancer           Dx.CIN
##           0              0
##           Dx.HPV            Dx
##           0              0
##           Hinselmann        Schiller
##           0              0
##           Citology          Biopsy
##           0              0
```

Estadísticas de valores informados con ‘?’

```
colSums(cervical_data_subset=="?")
```

```
##           Age           Number.of.sexual.partners
##           0              NA
## First.sexual.intercourse      Num.of.pregnancies
##           NA              NA
##           Smokes           Smokes..years.
##           NA              NA
##           Smokes..packs.year.      Hormonal.Contraceptives
##           13              108
## Hormonal.Contraceptives..years.      IUD
##           108              117
##           IUD..years.           STDs
##           117              105
##           STDs..number.      STDs..Number.of.diagnosis
##           105              0
## STDs..Time.since.first.diagnosis  STDs..Time.since.last.diagnosis
##           787              787
##           Dx.Cancer           Dx.CIN
##           0              0
##           Dx.HPV            Dx
##           0              0
##           Hinselmann        Schiller
##           0              0
##           Citology          Biopsy
##           0              0
```

Una vez identificados, se procede a aplicar el tratamiento de falta de datos que más se adapte a cada tipo de dato teniendo en cuenta que un alto número de valores sin informar puede ser un indicador de inconsistencia. Con ciertas variables podemos tomar el valor medio de los datos o el valor más repetido, en otros casos podemos tomar valores condicionales en base a la edad, por ejemplo si es fumador o no, fijando el valor NO para menores de 18 años. En otros casos es posible que nos convenga mantener los valores nulos y trabajar únicamente con el subset de datos informados como puede ser el indicador de si una persona estaba embarazada o no.

Antes de entrar a realizar tareas de transformación, es conveniente categorizar las variables y comprobar que están informadas de un modo coherente. Por ejemplo, comprobar que los campos de edad no tienen valores muy elevados o caracteres en lugar de números, eliminar los espacios de las variables de tipo string, comprobar que los indicadores se informan únicamente con valores de 0 y 1...

Transformación de las variables Number of sexual partners y first sexual intercourse por la media:

```
cervical_data_subset$Number.of.sexual.partners[is.na(cervical_data_subset$Number.of.sexual.partners)] <- median(cervical_data_subset$Number.of.sexual.partners, na.rm=T)
cervical_data_subset$First.sexual.intercourse[is.na(cervical_data_subset$First.sexual.intercourse)] <- median(cervical_data_subset$First.sexual.intercourse, na.rm=T)
```

Filtramos todos los valores con el registro IUD = '?' y les aplicamos el valor NA, se trata de una variable muy importante en nuestro estudio y no es fácilmente interpolable por lo que al ser un número elevado de muestras puede afectar al análisis de otras variables.

```
cervical_data_subset$IUD[cervical_data_subset$IUD=="?"] <- "NA"
cervical_data_subset$IUD..years.[cervical_data_subset$IUD..years=="?"] <- "0"
cervical_data_subset$IUD..years.<-as.integer(cervical_data_subset$IUD..years.)
```

3.2. Identificación y tratamiento de valores extremos

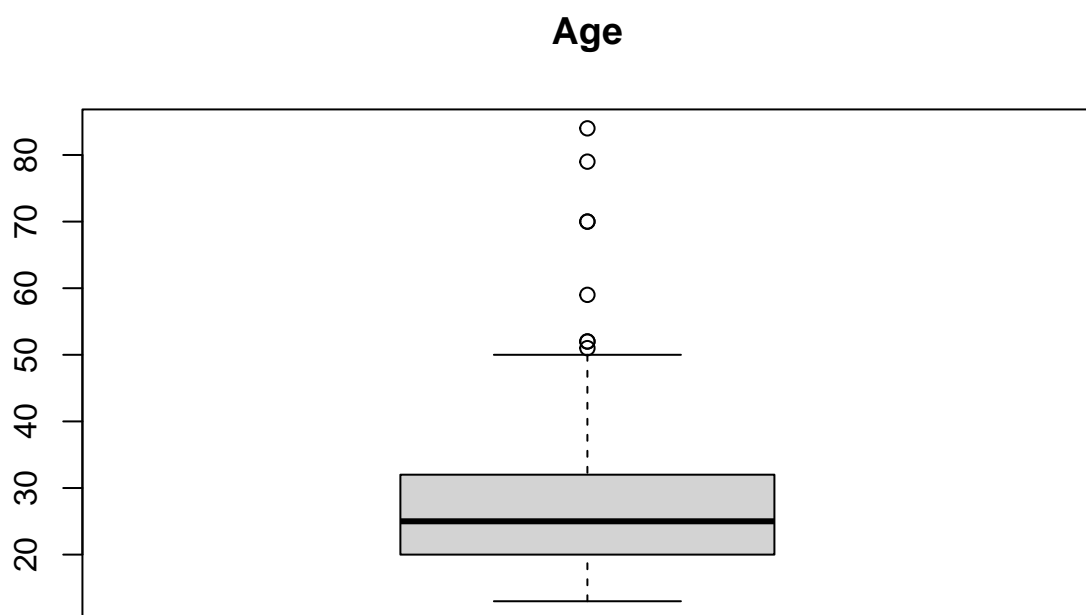
A continuación se procede a normalizar el campo edad para poder realizar representaciones gráficas de un modo más sencillo

```
summary(cervical_data_subset[, "Age"])
```

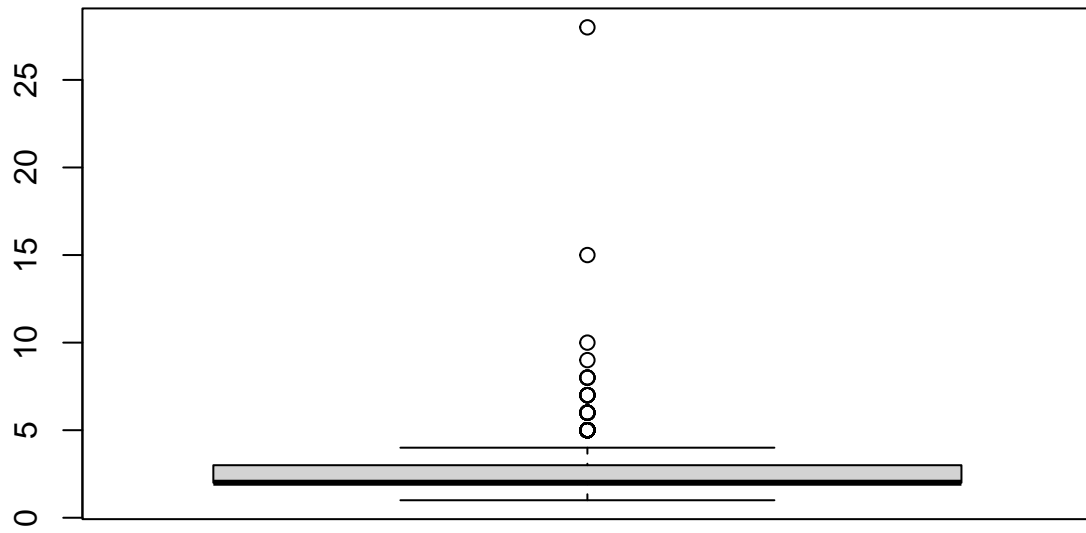
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.00   20.00   25.00   26.82   32.00   84.00
```

```
cervical_data_subset["Age_segment"] <- cut(cervical_data_subset$Age, breaks = c(0,10,20,30,40,50,60,70,80,90,100), right=F, labels=c("0-10", "10-20", "20-30", "30-40", "40-50", "50-60", "60-70", "70-80", "80-90", "90-100"))
```

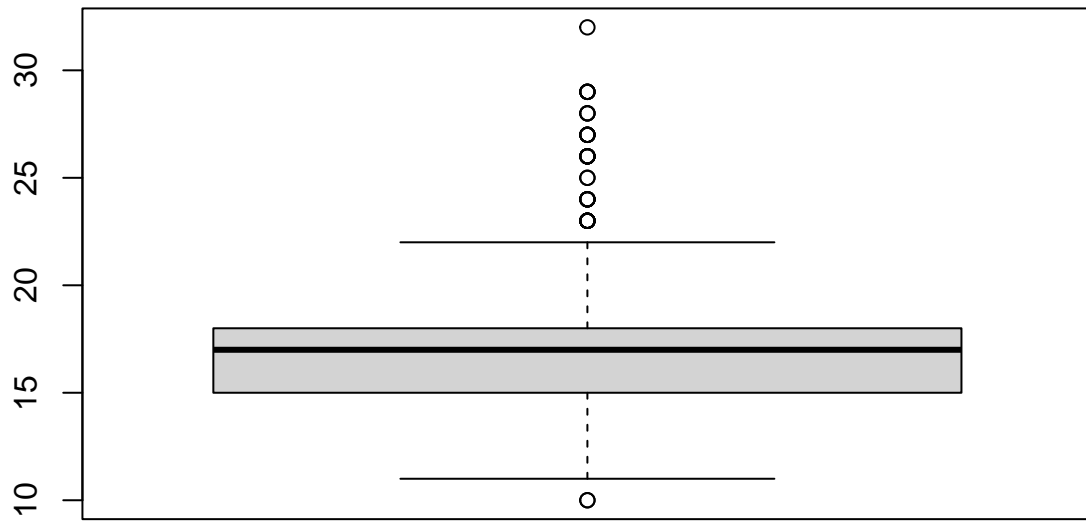
Además, para la identificación de outliers o valores extremos, utilizamos diagramas de cajas. Aparentemente no se observan valores extremos que pudieran perjudicar un posterior modelo de clasificación.



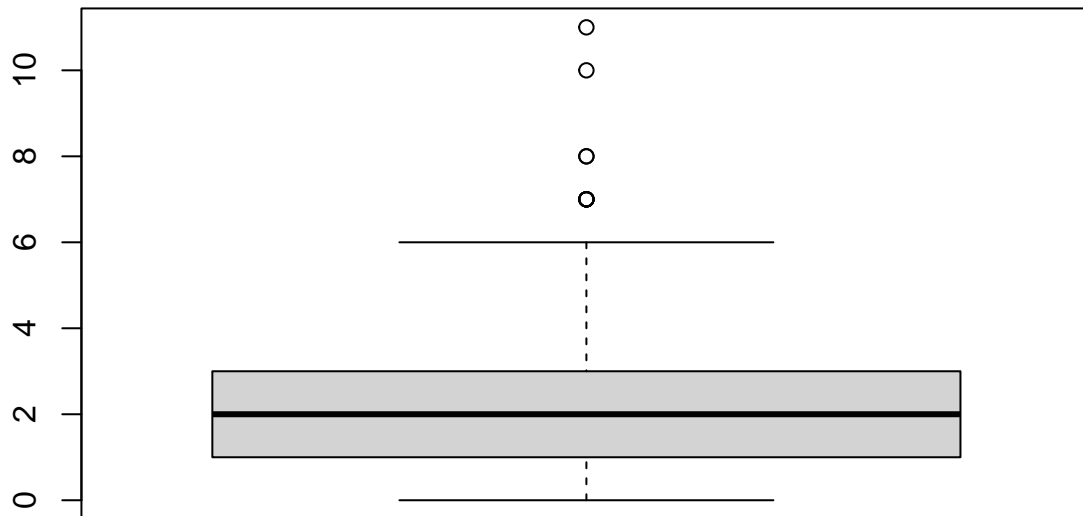
Number of sexual partners



First sexual intercourse



Number of pregnancies



4. Análisis de los datos

4.1. Selección de los datos

Precisamos un primer set de datos para continuar con los análisis escogemos las variables más significativas como edad segmentada, variables relacionadas con el tabaco, los metodos hormonales y dispositivos intrauterinos.

```
analisis_subset<-cervical_data_subset[c(1,25,5:13,17:20)]
str(analisis_subset)
```

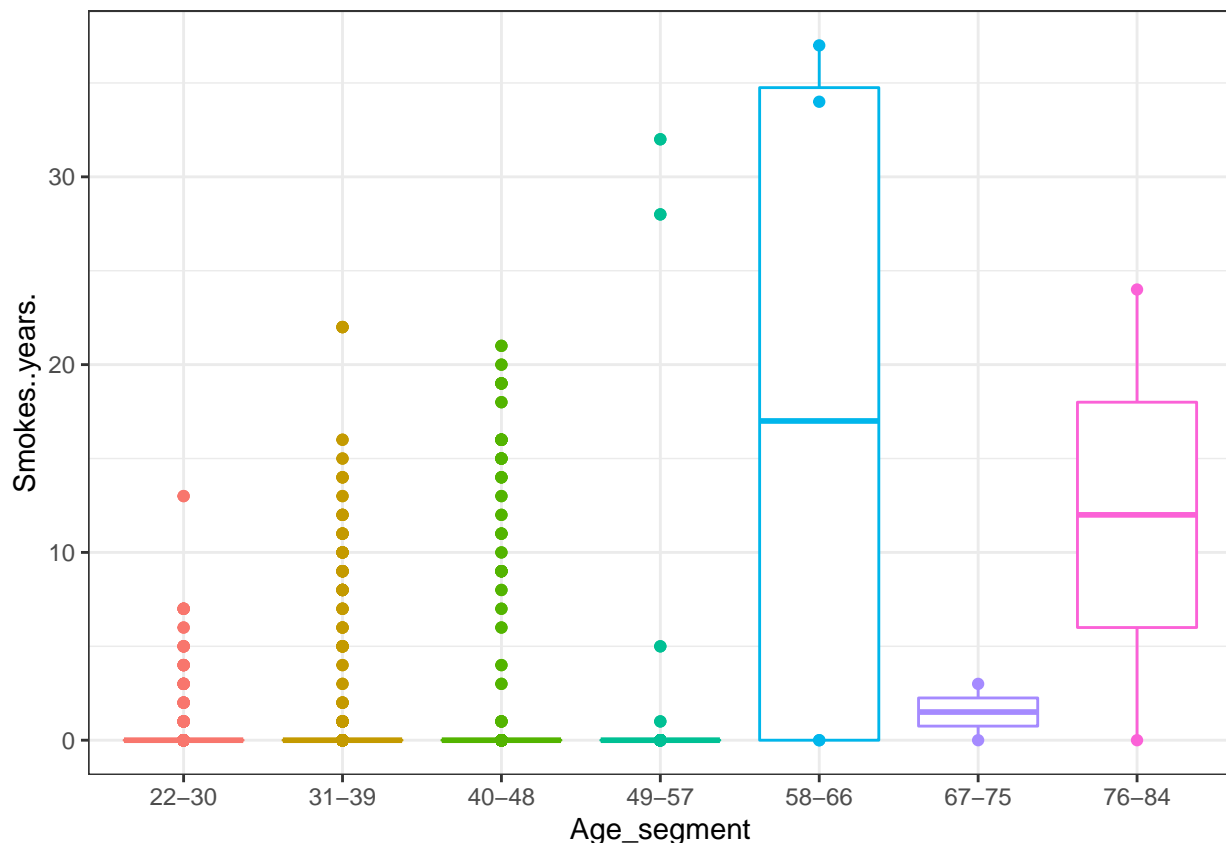
```
## 'data.frame':   858 obs. of  15 variables:
## $ Age           : int  18 15 34 52 46 42 51 26 45 44 ...
## $ Age_segment   : Factor w/ 8 levels "13-21","22-30",...: 2 2 4 6 5 5 6 3 5 5 ...
## $ Smokes        : int  0 0 0 1 0 0 1 0 0 1 ...
## $ Smokes..years. : int  0 0 0 37 0 0 34 0 0 1 ...
## $ Smokes..packs.year. : chr  "0.0" "0.0" "0.0" "37.0" ...
## $ Hormonal.Contraceptives : chr  "0.0" "0.0" "0.0" "1.0" ...
## $ Hormonal.Contraceptives..years.: chr  "0.0" "0.0" "0.0" "3.0" ...
## $ IUD           : chr  "0.0" "0.0" "0.0" "0.0" ...
## $ IUD..years.    : int  0 0 0 0 0 0 7 7 0 0 ...
## $ STDs          : chr  "0.0" "0.0" "0.0" "0.0" ...
## $ STDs..number.  : chr  "0.0" "0.0" "0.0" "0.0" ...
```

```
## $ Dx.Cancer      : int  0 0 0 1 0 0 0 0 1 0 ...
## $ Dx.CIN         : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Dx.HPV         : int  0 0 0 1 0 0 0 0 1 0 ...
## $ Dx             : int  0 0 0 0 0 0 0 0 1 0 ...
```

4.2 Normalidad y homogeneidad de varianzas y 4.3 Contraste de hipótesis, correlación y regresión

A continuación pasamos a realizar una muestra de homogeneidad entre los grupos de edad y los años consecutivos fumandos.

```
ggplot(data = analisis_subset, aes(x = Age_segment, y = Smokes..years., colour = Age_segment)) +
  geom_boxplot() +
  geom_point() +
  theme_bw() +
  theme(legend.position = "none")
```



A la vista de los resultados, los grupos con tamaños muestrales pequeños tienen mayor varianza, también es lógico pensar que a mayor edad más años consecutivos ha estado fumando la persona. Vemos como los tres grupos de 58 a 84 disponen de un tamaño muestral más grande.

```
aggregate(Smokes..years.~Age_segment, data = analisis_subset, FUN = var)
```

```
##   Age_segment Smokes..years.
## 1    22-30      2.256206
## 2    31-39      9.978527
## 3    40-48     25.672907
## 4    49-57     39.481818
```

```
## 5      58-66      421.583333
## 6      67-75       4.500000
## 7      76-84      288.000000
```

A la vista de los resultados podemos ver como la hipótesis del apartado anterior era cierta y existen grupos con mucha más cantidad de muestras y valor acumulado

A continuación utilizamos el test de Levene. Se caracteriza, además de por poder comparar 2 o más poblaciones, por permitir elegir entre diferentes estadísticos de centralidad :mediana (por defecto), media, media truncada. Esto es importante a la hora de contrastar la homocedasticidad dependiendo de si los grupos se distribuyen de forma normal o no.

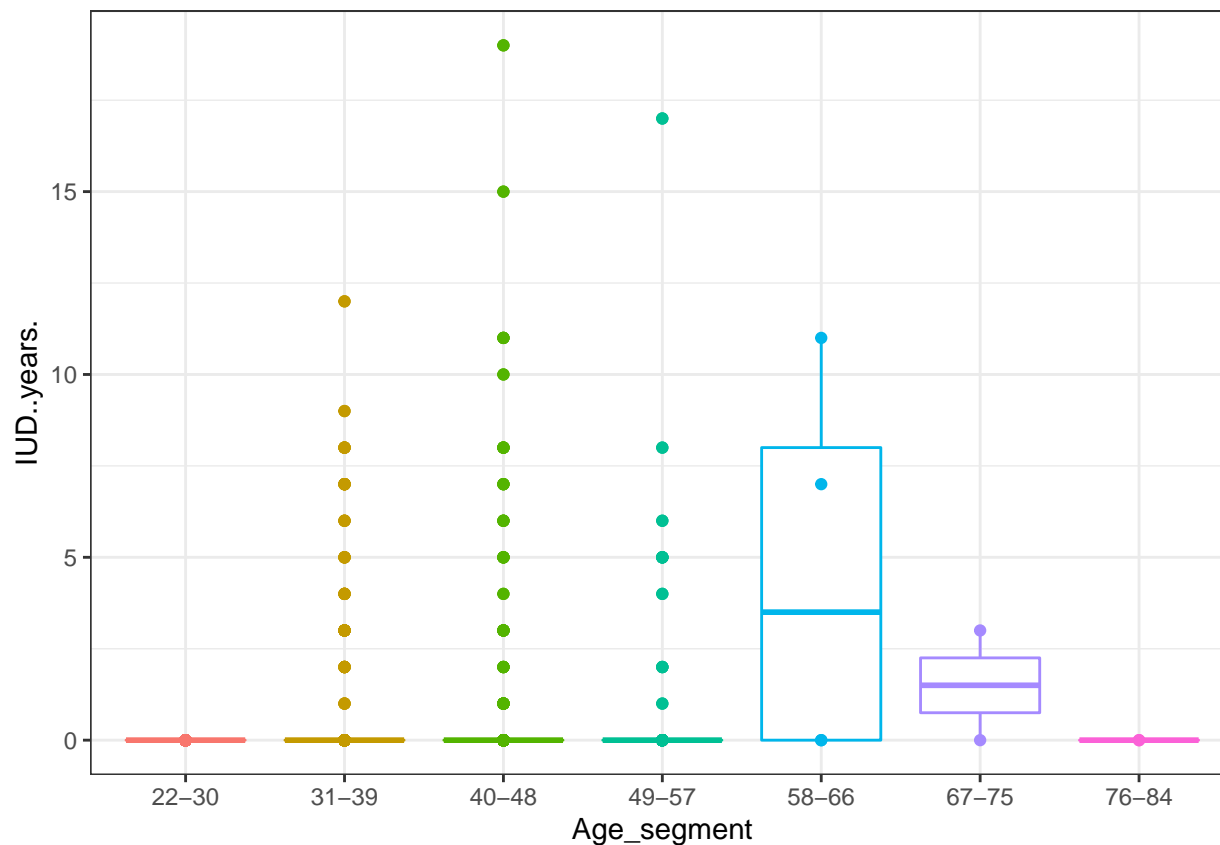
```
library(car)
data(" analisis_subset")
leveneTest(y = analisis_subset$Smokes..years., group = analisis_subset$Age_segment , center = "median")

## Levene's Test for Homogeneity of Variance (center = "median")
##      Df F value    Pr(>F)
## group  6  20.53 < 2.2e-16 ***
##      838
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El test de Levene Test encuentra diferencias significativas entre los diferentes grupos

A continuación analizamos los grupos de edad con los años de utilización de un dispositivo intra uterino.

```
ggplot(data = analisis_subset, aes(x = Age_segment, y = IUD..years., colour = Age_segment)) +
  geom_boxplot() +
  geom_point() +
  theme_bw() +
  theme(legend.position = "none")
```



En este caso vemos como las distribuciones son similares al caso anterior, aunque a priori parecen más homogéneas en el resto de grupos excepto en los extremos.

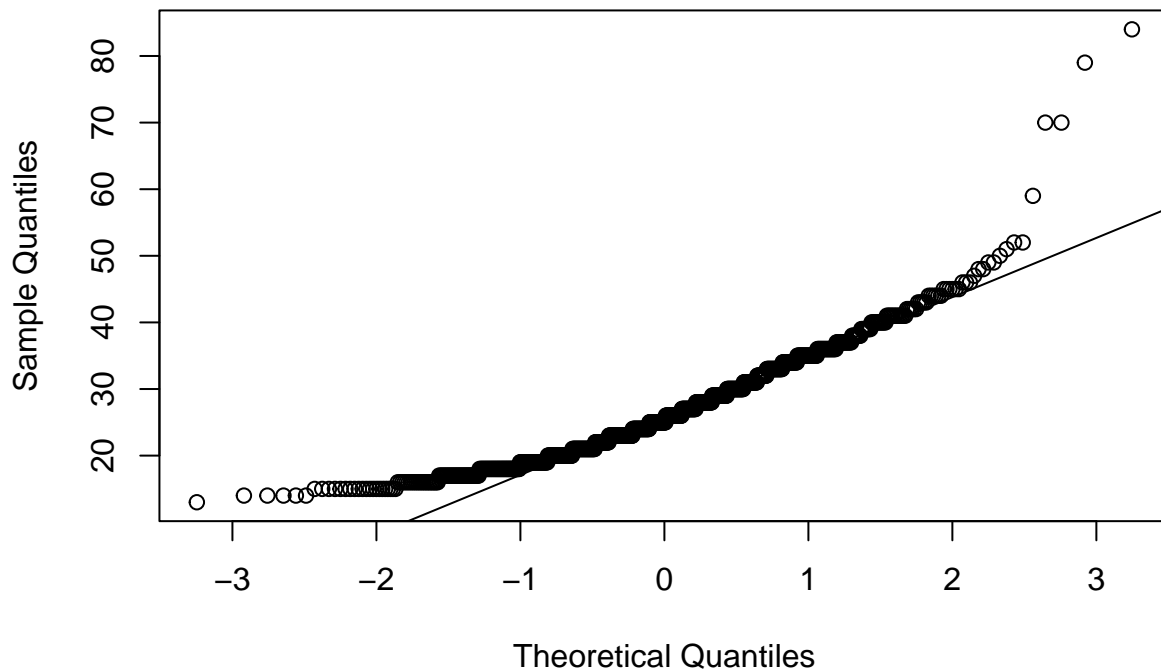
```
data(" analisis_subset")
leveneTest(y = analisis_subset$IUD..years., group = analisis_subset$Age_segment , center = "median")

## Levene's Test for Homogeneity of Variance (center = "median")
##      Df F value    Pr(>F)
## group  6  9.2597 7.566e-10 ***
##      851
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Supuesto de normalidad

```
qqnorm( analisis_subset$Age ) # la nube de puntos
qqline( analisis_subset$Age ) # la recta
```

Normal Q-Q Plot



Contraste de hipótesis

Realizamos un contraste de hipótesis para un valor de media de edad de 27 años.

```
t.test( analisis_subset$Age,  
       mu = 27,  
       alternative = "two.sided" ) # contraste bilateral
```

```
##  
## One Sample t-test  
##  
## data: analisis_subset$Age  
## t = -0.61868, df = 857, p-value = 0.5363  
## alternative hypothesis: true mean is not equal to 27  
## 95 percent confidence interval:  
## 26.25109 27.38993  
## sample estimates:  
## mean of x  
## 26.82051
```

Con un valor de p-value = 0.536 mayor que 0.05 no podemos rechazar la hipótesis nula.

Supuesto de homocedasticidad

En este caso llevamos a cabo el supuesto entre personas fumadoras o no fumadoras

```
Fumadores <- analisis_subset$Age[analisis_subset$Smokes == 1]  
NoFumadores <- analisis_subset$Age[analisis_subset$Smokes == 0]
```

```
var.test( Fumadores, NoFumadores ) # contraste de homogeneidad de varianzas
```

```
##  
## F test to compare two variances  
##  
## data: Fumadores and NoFumadores  
## F = 1.4601, num df = 122, denom df = 721, p-value = 0.003878  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 1.126914 1.942959  
## sample estimates:  
## ratio of variances  
## 1.460131
```

Con un valor de p-value = 0,003878 menor que 0,05 podemos rechazar la hipótesis nula, por lo tanto no existe homogeneidad de varianzas.

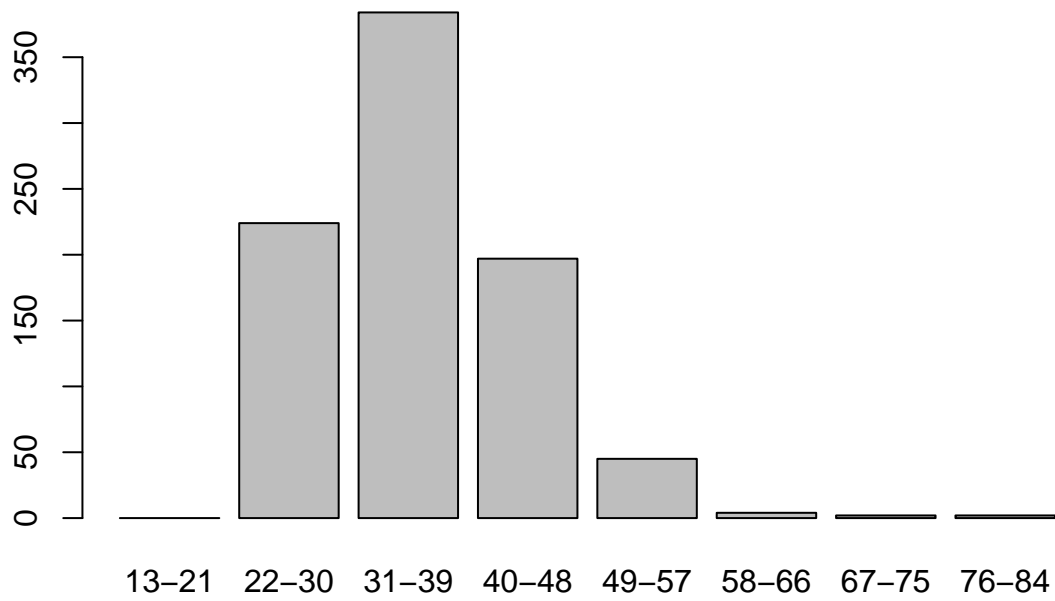
5. Representación de los resultados

Representación gráfica

Para entender los datos más en profundidad procedemos a realizar varias representaciones gráficas para obtener una visión más general de como se distribuye la información, de este modo nos será más sencillo identificar la naturaleza de los datos.

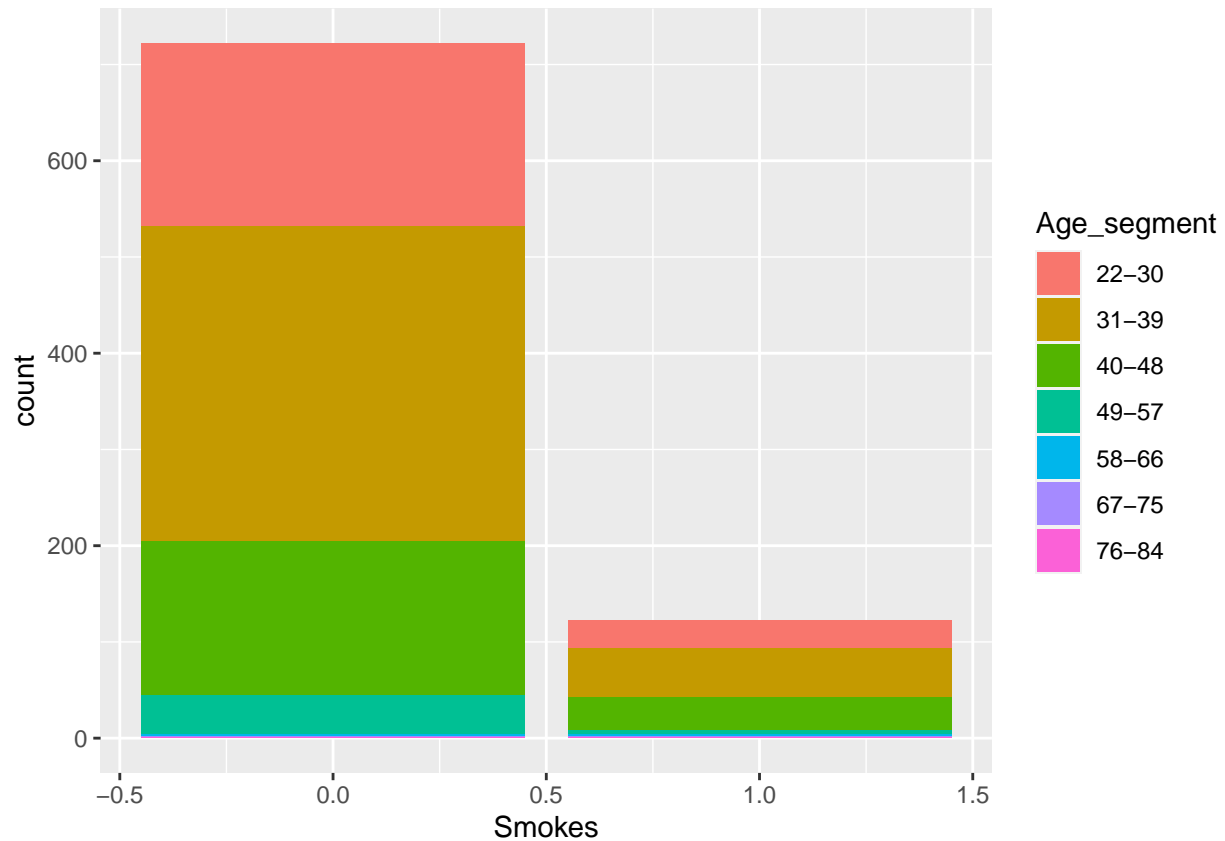
Distribución de los grupos de edad:

```
plot(cervical_data_subset$Age_segment)
```

Representación de los grupos de edad en base a si son fumadores o no:

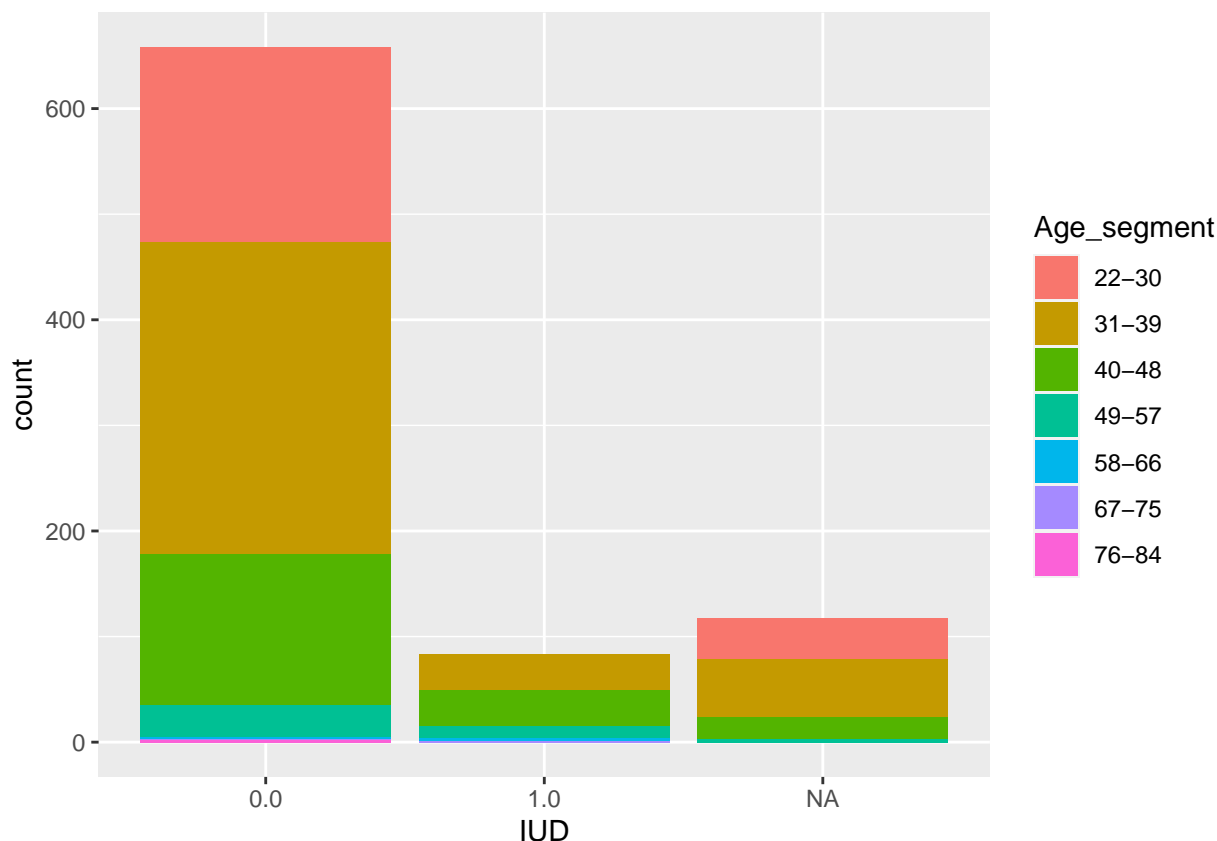
```
ggplot(data=cervical_data_subset[1:rows,], aes(x=Smokes , fill=Age_segment))+geom_bar()
```



En la grafica vemos como muchos más casos corresponden a personas no fumadoras independiente del rango de edad ya que este se distribuye de acuerdo a la distribución anterior por grupos de edad, en todos los grupos encontramos personas fumadoras excepto en el grupo 0, de 13 a 21 años. En este caso quizás podría darnos pie a pensar que el tabaco no es un factor de riesgo en este tipo de cancer pero habría que confirmarlo con análisis más avanzados.

Representación de los grupos de edad en base a si son fumadores o no:

```
ggplot(data=cervical_data_subset[1:rows,],aes(x=IUD ,fill=Age_segment))+geom_bar()
```



Por último, generamos un fichero CSV con el resultado del proceso de análisis y limpieza de datos.

```
write.csv(analysis_subset, "../data/kag_risk_factors_cervical_cancer_out.csv", row.names = TRUE)
```

6. Resolución del problema y conclusiones

Una vez analizado el dataset en profundidad se abre un amplio número de posibilidades. Por ejemplo, seríamos capaces de identificar diferentes agrupaciones en base a las variables descritas y lanzar las campañas de concienciación personalizadas a los grupos más numerosos.

Una vez establecida la clasificación sería necesario **analizar la calidad del modelo** para identificar que efectivamente los sujetos dentro de cada grupo tienen similitudes, existen diferencias entre los diferentes grupos y la distancia de cada muestra con el centro de su cluster es apropiada para definir un buen nivel de calidad. Por último, una vez establecido el modelo y teniendo en cuenta que todas las mujeres que se encuentran en el dataset han sufrido cáncer.

Si hemos sido capaces de identificar grupos con características comunes con un aceptable nivel de calidad, seremos capaces de **lanzar campañas de prevención personalizadas** lo más efectivas posible. Un posible resultado podría ser por un lado, proponer a las mujeres de entre 30 y 40 que no utilicen dispositivo intrauterino; por otro lado, proponer la realización de pruebas diagnósticas gratuitas a las mujeres de entre 20 y 30 que nunca se han realizado pruebas con anterioridad y una tercera campaña podría ir dirigida a las personas mayores de 40 años que han fumado durante más de 5 años para que dejen de fumar y concienciarles de los riesgos de seguir fumando.

7. Tabla de contribuciones

Contribuciones	Firma
Investigación previa	SBN, BLG
Redacción de las respuestas	SBN, BLG
Desarrollo del código	SBN, BLG