

Trabajo Practico R

Integrantes: Agustin Elian Fuentes, Lucas Nicolas Lenza, Felipe Sastre, Jose Ramiro Gonzalez Martin.

Ejercicio de Análisis de datos:

La consigna de este ejercicio tiene cierto grado de libertad. Deben elegir un dataset de su gusto. Hacerse tres preguntas y contestarlas con las herramientas de tidyverse. Pueden usar como referencia el ejercicio de Análisis de datos. Se sugiere realizar cierta visualización de los datos previa a la respuesta de las preguntas. Por ejemplo, si mi pregunta es cómo difieren las notas de un examen según la edad del alumno, es importante hacer gráficos y verificar, por ejemplo, que no haya datos inesperados/NA's.

En pos de realizar este ejercicio se utilizó el siguiente data set: "Used Car Price Prediction Dataset" (<https://www.kaggle.com/datasets/taeefnajib/used-car-price-prediction-dataset>). En el mismo se observa un conjunto de datos que comprende 4,009 puntos de datos, cada uno representando un listado único de vehículo, e incluye nueve características distintas que proporcionan información valiosa sobre el mundo automotriz. Separando cada auto y categorizándolos por: marca y modelo, año del modelo, cantidad de millas, tipo de combustible, tipo de motor, transmisión, colores exterior e interior, historial de accidentes, si tiene o no un título limpio y precio.

Preguntas:

- 1) Preguntas generales del dataset:
 - a) ¿Cuántos autos se vendieron?
 - b) ¿Cuántos de cada marca se vendieron?
 - c) Clasificarlos en gama económica (a los que valen menos de 25000), media (a los que valen más de 25000 y menos de 55000) y de lujo (valen más que 55000) y modificar el data set. Y ver cuántos hay de cada gama.
- 2) Cambio de características del data set:
 - a) Modificar millas a kilómetros y cambiar de dolares a pesos, cual es el auto que más kilómetros tiene y el que más vale
 - b) Agregar las clasificaciones a las columnas
- 3) Correlaciones:
 - a) Graficar la distribución de los autos con los precios
 - b) Encontrar la correlación entre precio y kilómetros realizados
 - c) Encontrar la correlación entre precio y año

Respuestas:

- 1) a) Se vendieron 4009 autos para ello usamos el siguiente código:

```
num_autos <- nrow(cardata)
```

```
cat("Cantidad total de autos:", num_autos, "\n")
```

- b) Para saber cuántos autos vendidos de cada marca se usa el siguiente código:

```

conteo_marcas <- cardata %>%
  group_by(brand) %>%
  summarise(Cantidad = n()) %>%
  arrange(desc(Cantidad))
print(head(conteo_marcas, 57), n = 57)

```

```
# A tibble: 57 x 2
```

	brand	Cantidad
	<chr>	<int>
1	Ford	386
2	BMW	375
3	Mercedes-Benz	315
4	Chevrolet	292
5	Porsche	201
6	Audi	200
7	Toyota	199
8	Lexus	163
9	Jeep	143
10	Land	130
11	Nissan	116
12	Cadillac	107
13	GMC	91
14	RAM	91
15	Dodge	90
16	Tesla	87
17	Kia	76
18	Hyundai	72
19	Acura	64
20	Mazda	64
21	Subaru	64
22	Honda	63
23	INFINITI	59
24	Volkswagen	59
25	Lincoln	52
26	Jaguar	47
27	Volvo	38
28	Maserati	34
29	Bentley	33
30	MINI	33
31	Buick	30
32	Chrysler	28
33	Lamborghini	26
34	Genesis	20
35	Mitsubishi	20
36	Alfa	19
37	Rivian	17
38	Hummer	16
39	Pontiac	15
40	Ferrari	12
41	Rolls-Royce	11
42	Aston	9
43	McLaren	6
44	Scion	6
45	FIAT	5
46	Saturn	5
47	Lotus	4
48	Lucid	3
49	Mercury	3
50	Karma	2
51	Saab	2
52	Bugatti	1
53	Maybach	1
54	Plymouth	1
55	Polestar	1
56	Suzuki	1
57	smart	1

c) Para clasificar los autos usamos:

```

cardata_eco <- cardata_limp %>%
  filter(price <= 25000)

```

```

num_autos_eco <- nrow(cardata_eco)
cat("Cantidad total de autos de gama baja:", num_autos_eco, "\n")

```

```

cardata_media <- cardata_limp %>%

```

```
filter(price > 25000 & price <= 55000)
```

```
num_autos_media <- nrow(cardata_media)
```

```
cat("Cantidad total de autos de gama media:", num_autos_media, "\n")
```

```
cardata_lujo <- cardata_limp %>%
```

```
filter(price > 55000)
```

```
num_autos_lujo <- nrow(cardata_lujo)
```

```
cat("Cantidad total de autos de gama alta:", num_autos_lujo, "\n")
```

Separando los autos:

Cantidad total de autos de gama baja: 1571

Cantidad total de autos de gama media: 1633

Cantidad total de autos de gama alta: 805

2) a) y b)

	brand	model	model_year	milage	fuel_type	engine	transmission	ext_col	int_col	accident	clean_title	price	clasificacion
1	Ford	Utility Police Interceptor Base	2013	82076.3400	E85 Flex Fuel	300.0HP 3.7L V6 Cylinder Engine Flex Fuel Capability	6-Speed A/T	Black	Black	At least 1 accident or damage reported	Yes	10300	Económico
2	Hyundai	Pelicaide SEL	2021	55911.6903	Gasoline	3.8L V6 24V GDI DOHC	8-Speed Automatic	Moonlight Cloud	Gray	At least 1 accident or damage reported	Yes	38005	Medio
3	Lexus	RX 350 RX 350	2022	36004.1545	Gasoline	3.5 Liter DOHC	Automatic	Blue	Black	None reported	N/A	54598	Medio
4	INFINITI	Q50 Hybrid Sport	2015	143070.3260	Hybrid	354.0HP 3.5L V6 Cylinder Engine Gas/Electric Hybrid	7-Speed A/T	Black	Black	None reported	Yes	15500	Económico
5	Audi	Q3 45 S line Premium Plus	2021	15827.8589	Gasoline	2.0L 14 16V GDI DOHC Turbo	8-Speed Automatic	Glacier White Metallic	Black	None reported	N/A	34999	Medio
6	Acura	ILX 2.4L	2016	219509.1480	Gasoline	2.4 Liter	F	Silver	Ebony.	None reported	N/A	14798	Económico
7	Audi	S3 2.0T Premium Plus	2017	135184.5600	Gasoline	282.0HP 2.0L 4 Cylinder Engine Gasoline Fuel	6-Speed A/T	Blue	Black	None reported	Yes	31000	Medio
8	BMW	740 iL	2001	389460.2800	Gasoline	282.0HP 4.4L 8 Cylinder Engine Gasoline Fuel	A/T	Green	Green	None reported	Yes	7300	Económico
9	Lexus	RC 350 F Sport	2021	37716.4902	Gasoline	311.0HP 3.5L V6 Cylinder Engine Gasoline Fuel	6-Speed A/T	Black	Black	None reported	Yes	41827	Medio
10	Tesla	Model X Long Range Plus	2020	54771.5600	N/A	534.0HP Electric Motor Electric Fuel System	A/T	Black	Black	None reported	Yes	69950	Lujo
11	Land	Rover Range Rover Sport 3.0 Supercharged HST	2021	44430.6587	Gasoline	V6	Automatic	Fuji White	Pimento / Ebony	None reported	N/A	73897	Lujo
12	Aston	Martin DBS Superleggera	2019	36644.6718	Gasoline	715.0HP 5.2L 12 Cylinder Engine Gasoline Fuel	8-Speed A/T	Silver	Black	None reported	Yes	184606	Medio
13	Toyota	Supra 3.0 Premium	2021	20116.7500	Gasoline	382.0HP 3.0L Straight 6 Cylinder Engine Gasoline Fuel	A/T	Yellow	Black	None reported	Yes	53500	Medio
14	Lincoln	Aviator Reserve AWD	2022	25083.5506	Gasoline	400.0HP 3.0L V6 Cylinder Engine Gasoline Fuel	Transmission w/Dual Shift Mode	Black	Brown	None reported	Yes	62000	Lujo
15	Jaguar	F-TYPE	2020	25593.3340	Gasoline	2.0 Liter Supercharged	Automatic	Silver	Black	None reported	N/A	47998	Medio
16	Land	Rover LR4 HSE	2013	128425.3320	Gasoline	375.0HP 5.0L 8 Cylinder Engine Gasoline Fuel	A/T	White	Black	None reported	Yes	29990	Medio
17	Mercedes-Benz	Metris Base	2021	2711.7379	Gasoline	2.0L 14 16V GDI DOHC Turbo	9-Speed Automatic	Gray	White	None reported	Yes	250000	Lujo
18	Dodge	Challenger SXT	2013	98288.8312	Gasoline	305.0HP 3.6L V6 Cylinder Engine Gasoline Fuel	A/T	Black	Gray	None reported	Yes	16800	Económico
19	Nissan	350Z Enthusiast	2003	119091.1600	Gasoline	287.0HP 3.5L V6 Cylinder Engine Gasoline Fuel	6-Speed M/T	Purple	—	None reported	Yes	11000	Económico
20	Jaguar	F-TYPE R	2018	56729.2350	Gasoline	550.0HP 5.0L 8 Cylinder Engine Gasoline Fuel	8-Speed A/T	Green	Black	None reported	Yes	68750	Lujo
21	Genesis	GV70 3.5T Sport	2023	8690.4360	Gasoline	375.0HP 3.5L V6 Cylinder Engine Gasoline Fuel	8-Speed A/T	Green	Beige	N/A	N/A	60000	Lujo
22	Chevrolet	S-10 LS	2000	214862.8834	E85 Flex Fuel	120.0HP 2.2L 4 Cylinder Engine Flex Fuel Capability	A/T	Blue	Black	None reported	Yes	4500	Económico
23	BMW	440 Gran Coupe 440i xDrive	2020	41826.7466	Gasoline	3.0 Liter Turbo	Automatic	White	Beige	None reported	N/A	38598	Medio
24	Ford	F-150 XLT	2023	4543.1668	Gasoline	3.5L V6 24V PDI DOHC Twin Turbo	Automatic	Iconic Silver Metallic	Black	None reported	Yes	58504	Lujo
25	Chevrolet	Suburban RST	2021	83695.6800	Gasoline	355.0HP 5.3L 8 Cylinder Engine Gasoline Fuel	10-Speed A/T	Blue	Black	None reported	Yes	61000	Lujo
26	Hyundai	Elantra N Base	2022	18748.8110	Gasoline	276.0HP 2.0L 4 Cylinder Engine Gasoline Fuel	Transmission w/Dual Shift Mode	Black	Black	None reported	Yes	32300	Medio
27	Nissan	Sentra SR	2016	155367.2929	Gasoline	1.8 Liter	Automatic	Blue	Black	None reported	N/A	13998	Económico
28	Mercedes-Benz	AMG CLA 45 Base 4MATIC	2021	28163.4500	Gasoline	382.0HP 2.0L 4 Cylinder Engine Gasoline Fuel	A/T	Gray	Black	None reported	Yes	59995	Lujo
29	BMW	650 Gran Coupe i xDrive	2015	111044.4600	Gasoline	445.0HP 4.4L 8 Cylinder Engine Gasoline Fuel	8-Speed A/T	Gray	Black	None reported	Yes	25700	Medio
30	Audi	Q5 2.0T Premium Plus	2017	131772.7592	E85 Flex Fuel	2.0L 14 16V GDI DOHC Turbo Flexible Fuel	8-Speed Automatic	Mythos Black Metallic	Black	None reported	Yes	21375	Económico
31	Mercedes-Benz	AMG C 43 Base 4MATIC	2018	46348.9920	Gasoline	362.0HP 3.0L V6 Cylinder Engine Gasoline Fuel	9-Speed A/T	Silver	Black	None reported	Yes	48000	Medio
32	Dodge	Ram 1500 Laramie Mega Cab	2006	483096.5092	Gasoline	345.0HP 5.7L 8 Cylinder Engine Gasoline Fuel	A/T	Silver	Gray	At least 1 accident or damage reported	Yes	10900	Económico

En la imagen se observa las millas pasadas a kilómetros y la nueva columna de clasificación. Para realizar esto utilizamos la código mutate:

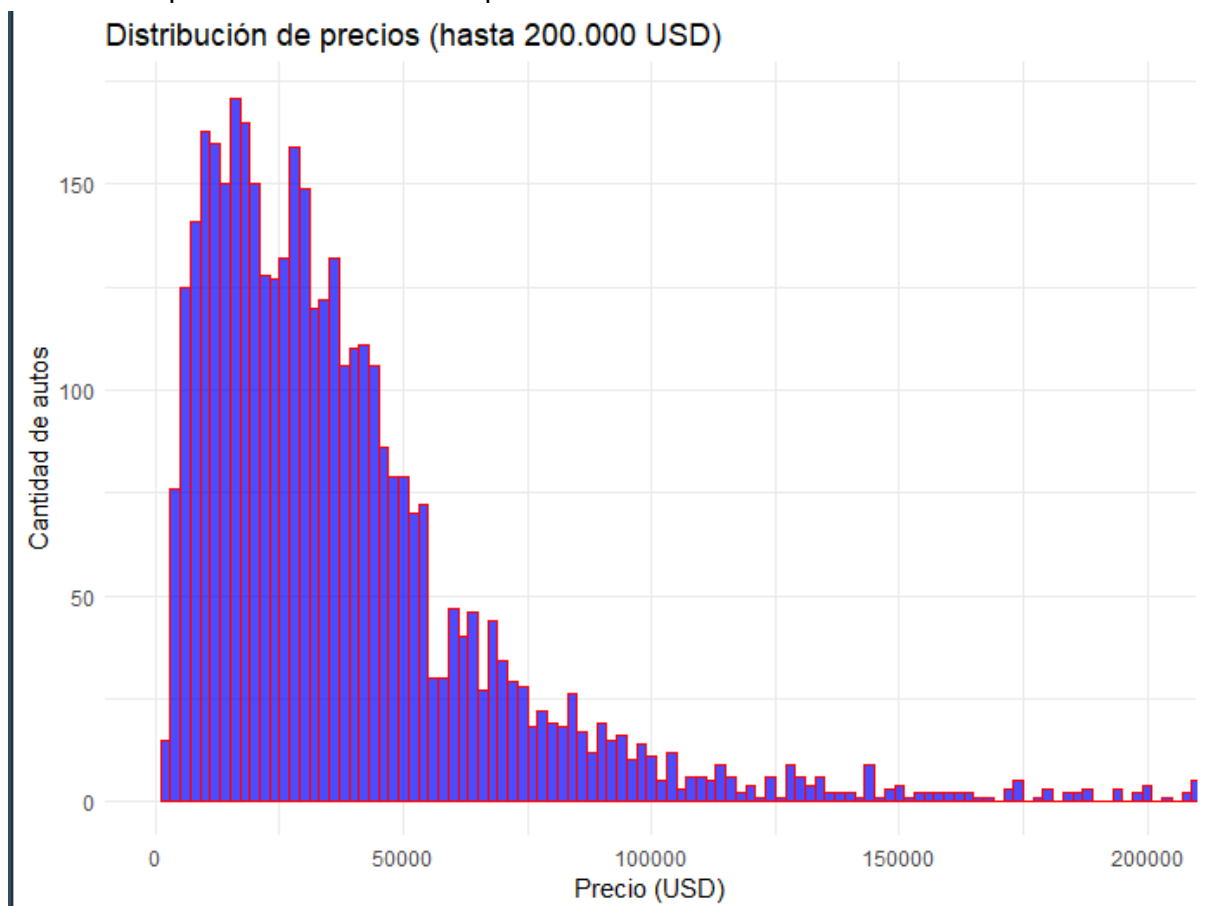
```
#A)
cardata_limp <- cardata_limp %>%
mutate(milage = milage * 1.60934) # sobrescribe la variable original
view(cardata_limp)

#B)
cardata_limp <- cardata_limp %>%
mutate(
  clasificacion = case_when(
    price <= 25000 ~ "Económico",
    price > 25000 & price <= 55000 ~ "Medio",
    price > 55000 ~ "Lujo"
  )
)
view(cardata_limp)
```

3) a) El gráfico de la distribución de los precios de los autos nos queda:



Sin embargo, en el gráfico la distribución está estirado debido a que existen ciertos autos que su precio es relativamente muy alto como por ejemplo el Bugatti haciéndolos difíciles de visualizar. Y para ellos acotamos los precios de 200000.



b) La correlación entre precios y kilómetros realizados es: -0.3055281

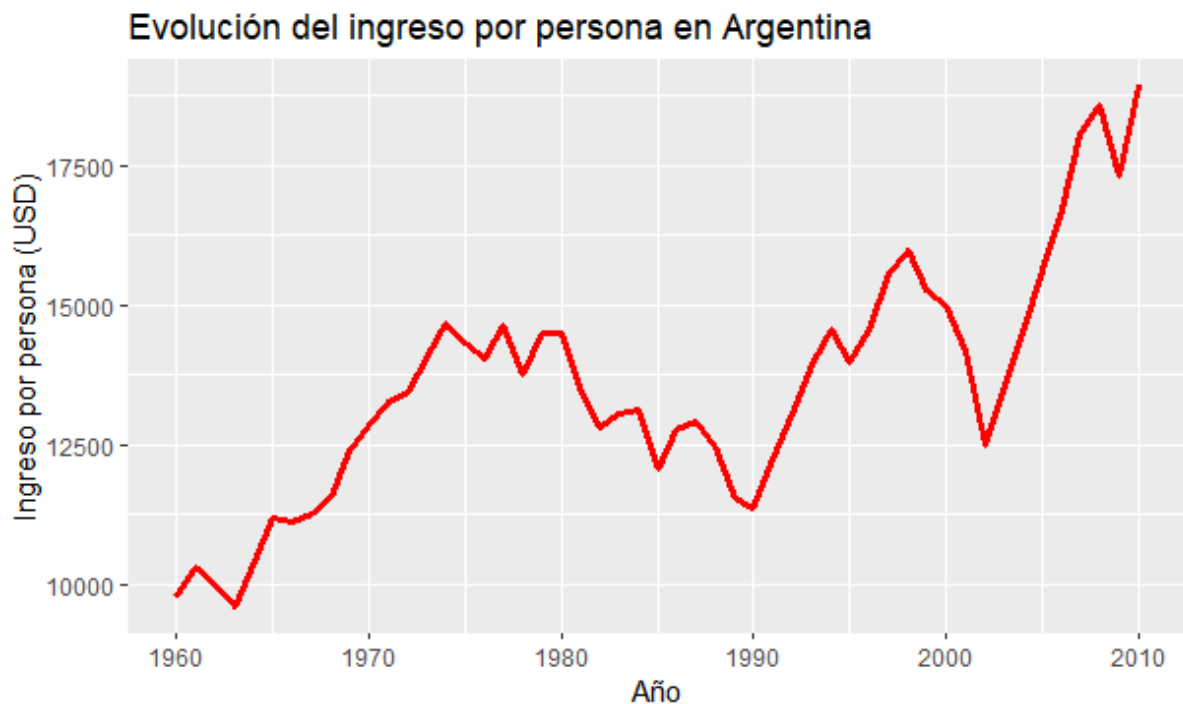
Encontramos una correlación negativa entre precio kilómetros realizados, lo cual tiene sentido

c) La correlación entre precio y año del modelo es : 0.1994962

Es decir, cuanto más nuevo es el auto los precios son mayores.

Ejercicio: Análisis econométrico con datos de gapminder

1. Grafica la evolución temporal de la variable income per person en Argentina. Comenta brevemente la tendencia observada.



Fuente: gapminder

En este ejercicio primero se filtraron los datos de argentina y se separaron del resto del data set asignado. Esto se logró gracias a los siguientes comandos:

```
argentina <- gapminder %>%  
filter(country == "Argentina")
```

A continuación, usando ggplot graficamos la evolución temporal del ingreso per cápita en la Argentina:

```
ggplot(argentina, aes(x = year, y = income_per_person)) +  
geom_line(color = "red", linewidth = 1.1) +  
labs(  
  title = "Evolución del ingreso por persona en Argentina",  
  x = "Año",  
  y = "Ingreso por persona (USD)",  
  caption = "Fuente: gapminder"  
)
```

En este mismo gráfico observamos la tendencia Argentina. Esta misma se podría llegar a separar en 2 partes; la primera de 1960 a mediados de los setenta, donde la tendencia es

claramente positiva y el crecimiento del ingreso per cápita es innegable. A partir de 1980 se observa periodos de crecimiento y decrecimiento alternados (provocados por una volatilidad e incertidumbre constantes), se destacan los descensos pronunciados durante la hiperinflación de 1989-1990 y la crisis de 2001-2002. Tras la crisis de 2001, el gráfico vuelve a su tendencia creciente antes mencionada.

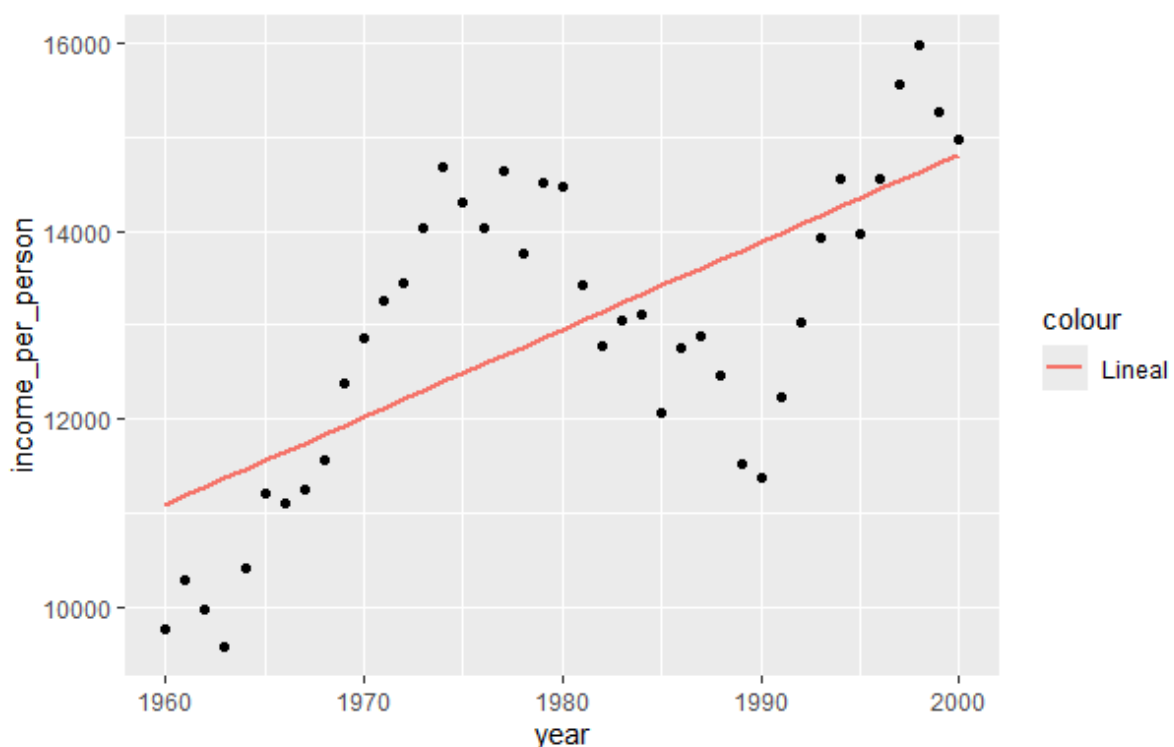
2.

Para este ejercicio, antes de realizar los diferentes modelos regresivos, primero separamos los datos para training y para testear. Para hacer ello utilizamos:

```
arg_train <- argentina %>% filter(year <= max(year) - 10)
```

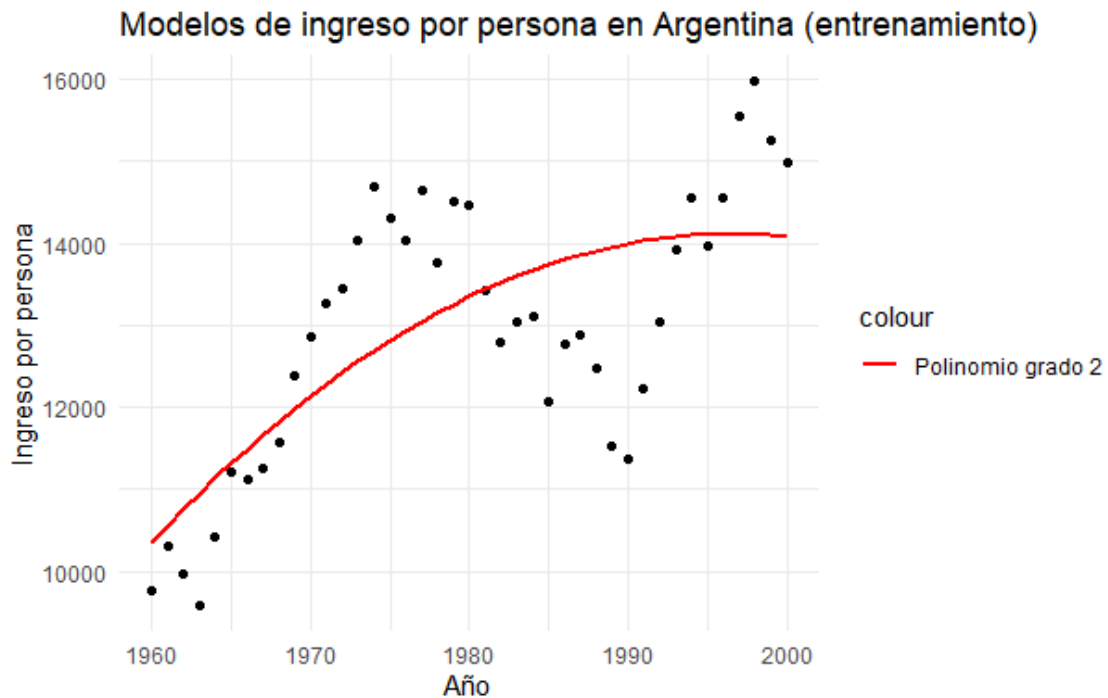
```
arg_test <- argentina %>% filter(year > max(year) - 10)
```

Primer modelo, modelo lineal:



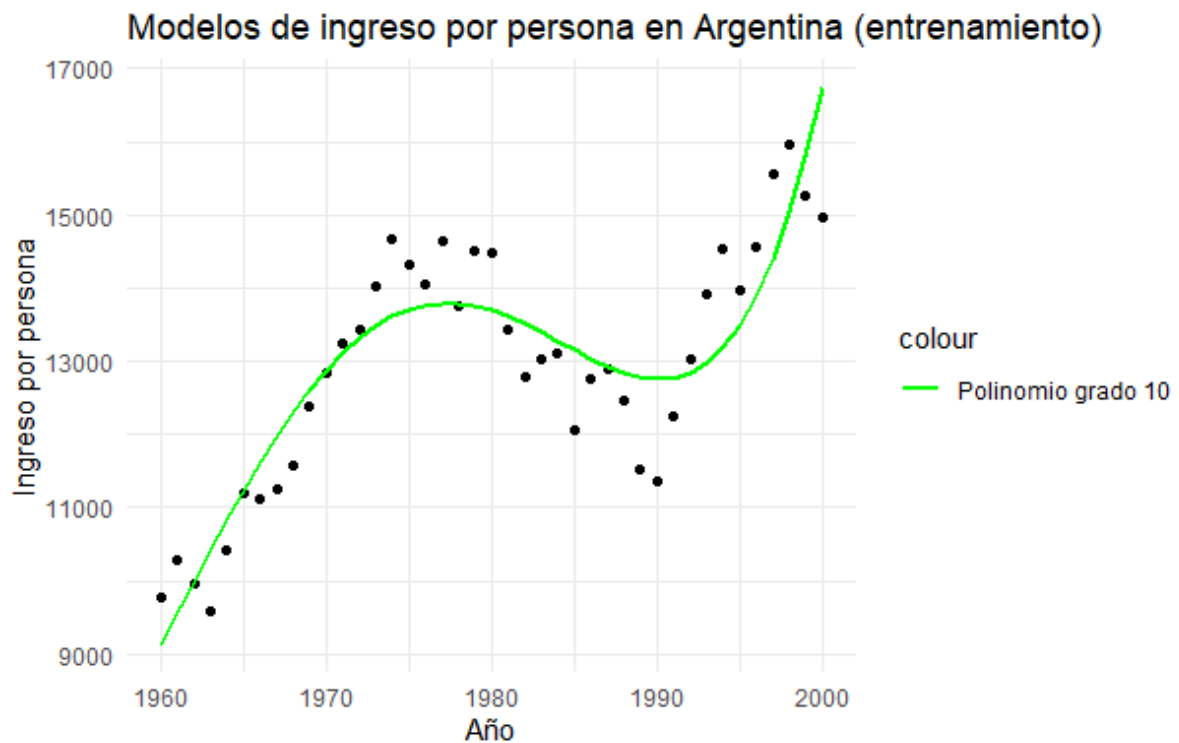
Al ser un modelo lineal, presenta constantemente una pendiente positiva, resultado de que en esos años existió un crecimiento del ingreso per cápita.

Segundo modelo, modelo polinómico de grado 2:

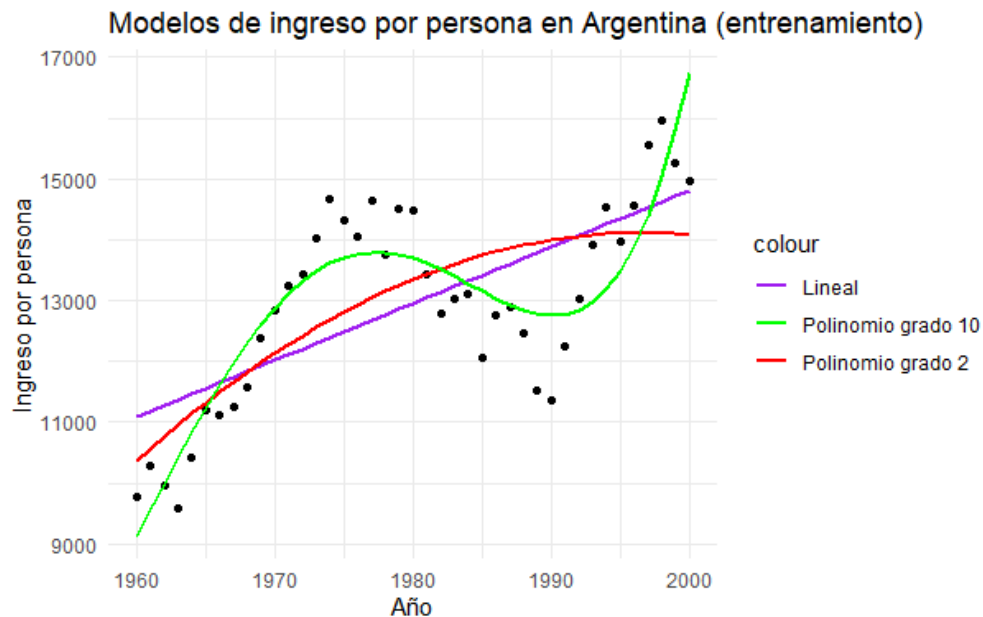


En el modelo polinómico de grado 2 observamos al igual que en el lineal una pendiente positiva; sin embargo, este modelo reconoce que no siempre creció igual y muestra una tendencia positiva pero cada vez más lenta (con 2da derivada negativa).

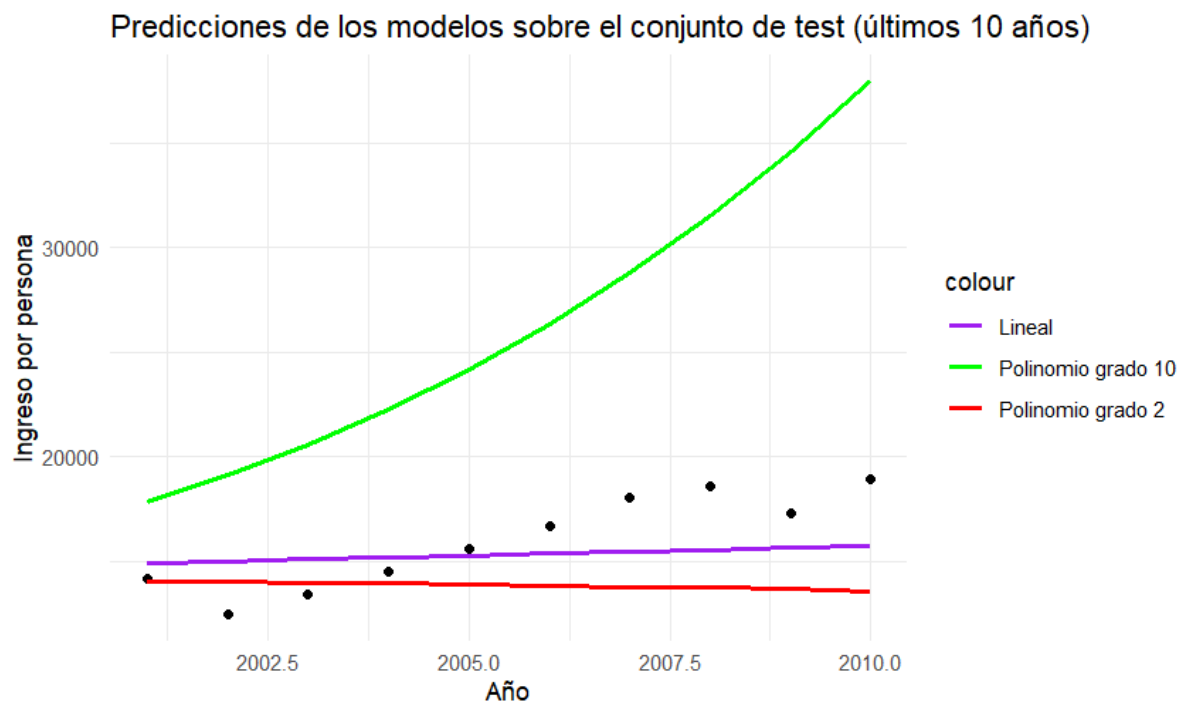
Tercer modelo, modelo polinómico de grado 10:



En el modelo polinómico de grado 10 se rompe con una predicción capaz de reconocer cuando hubo decrecimiento o crecimientos, esto se ve reflejado tanto es sus cambios de convexidades y de pendientes.



Al juntarlos se observa que los 3 modelos se ajustan al crecimiento del ingreso por persona que tuvo argentina desde 1960 hasta el año 2000. Sin embargo, es clara la diferencia en la eficiencia del training con los datos, siendo el modelo polinómico de grado 10 el que más se ajusta de manera correcta.



Sin embargo, si decidimos usar estos modelos sobre los últimos 10 años entendemos que podríamos estar ante un caso de overfitting en el polinomio de grado 10. El código que utilizamos para ver esto es muy parecido al que usamos para el testeo pero cambiando los datos a los últimos 10 años.

Comentario: Capaz esta última predicción no es muy buena referencia por tener un n chico siendo igual a 10

3. Selecciona cuatro países sudamericanos distintos de Argentina.

Con ellos, arma:

(a) Una matriz de correlaciones entre los ingresos (income per person) de los cinco países.

(b) Una matriz de correlaciones entre las variaciones porcentuales anuales (crecimiento interanual, Y/Y) de dichos ingresos.

Discute brevemente las similitudes y diferencias encontradas.

Para este ejercicio seleccionamos los siguientes países: Argentina, Paraguay, Bolivia, Guyana y Uruguay. Y a partir del siguiente código filtramos los datos de dichos países:

```
países <- c("Argentina", "Guyana", "Paraguay", "Bolivia", "Uruguay")
```

```
gap_sud <- gapminder %>%  
  filter(country %in% países) %>%  
  select(country, year, income = income_per_person)
```

Luego, construimos la matriz de correlación entre los ingresos por persona de los 5 países:

```
cor_ingresos <- gap_wide %>%  
  select(-year) %>%  
  cor(use = "pairwise.complete.obs")
```

```
print("Matriz de correlaciones entre ingresos (niveles):")
```

```
print(round(cor_ingresos, 3))
```

Matriz de varianzas y covarianzas:

	Argentina	Bolivia	Guyana	Paraguay	Uruguay
Argentina	1.000	0.924	0.869	0.673	0.829
Bolivia	0.924	1.000	0.875	0.710	0.799
Guyana	0.869	0.875	1.000	0.573	0.851
Paraguay	0.673	0.710	0.573	1.000	0.856
Uruguay	0.829	0.799	0.851	0.856	1.000

En la matriz de covarianzas observamos que todos los países tienen una correlación positiva, se podría decir que los ingresos per cápita de los países de la región se mueven más o menos de similar manera en el tiempo, creemos que es razonable porque ocupan un mismo espacio geopolítico y que las decisiones políticas de cada país tienden a afectar a los otros países también.

b) Primero calculamos variaciones porcentuales interanuales (crecimiento Y/Y) y para eso usamos el siguiente código:

```
gap_var <- gap_sud %>%  
  arrange(country, year) %>%  
  group_by(country) %>%  
  mutate(growth = (income - lag(income)) / lag(income) * 100) %>%  
  ungroup()
```

Luego calculamos la matriz de correlaciones entre variaciones:

```
cor_growth <- gap_wide_growth %>%  
  select(-year) %>%
```

`cor(use = "pairwise.complete.obs")` (sirve para que no se descarte el país completo
cuando hay datos faltantes)

Matriz de varianzas y covarianzas:

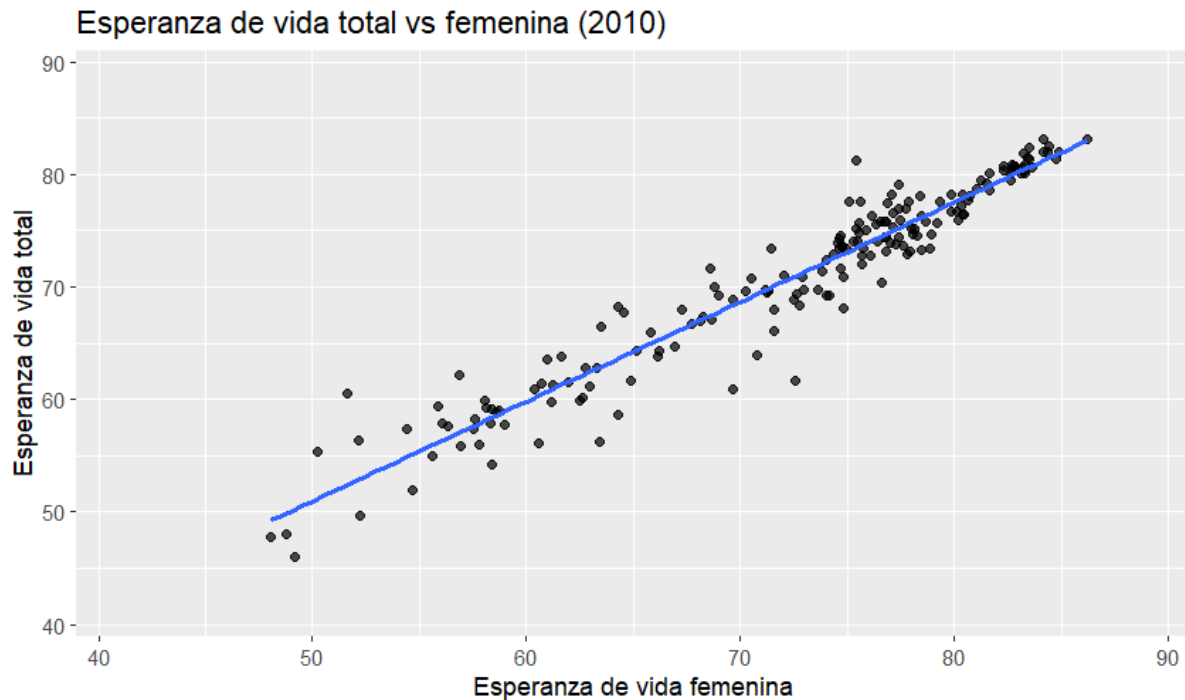
	Argentina	Bolivia	Guyana	Paraguay	Uruguay
Argentina	1.000	0.207	0.385	0.147	0.513
Bolivia	0.207	1.000	0.321	0.261	0.265
Guyana	0.385	0.321	1.000	0.094	0.460
Paraguay	0.147	0.261	0.094	1.000	0.324
Uruguay	0.513	0.265	0.460	0.324	1.000

En este caso vemos que todas las covarianzas son positivas, al igual que el caso anterior, sin embargo, presentan covarianzas más bajas, creemos que esto puede estar dado porque las variaciones interanuales de los ingresos no dependen tanto de los países externos, si no que esto sería un problema sobretodo interno de cada país. Notar el caso de Guyana y Paraguay que su correlación es relativamente muy baja, dando a entender, justamente, que la evolución del ingreso seguramente dependa mucho más de factores internos que externos.

Parte 2 : Esperanza y género

Inciso 5:

El gráfico de dispersión muestra una relación lineal positiva muy fuerte entre ambas variables. Los países con mayor esperanza de vida femenina es un buen indicador del nivel general de salud y desarrollo de cada país, ya que ambos indicadores evolucionan de manera conjunta.



Inciso 6:

Se estimó un modelo de regresión lineal simple donde la variable dependiente es la esperanza de vida total y la variable explicativa es la esperanza de vida femenina.

Resultados:

Intercepto (β_0): 5.2225

Coefficiente (β_1): 0.9037

R²: 0.8743

EL coeficiente β_1 resulta positivo y cercano a 1, lo cual implica que la esperanza de vida total correlaciona positivamente con la esperanza de vida femenina. Por otra parte, vemos un R² muy alto, lo cual implica que la variable femenina explica casi toda la variabilidad de la esperanza total. El promedio de la longevidad femenina predice casi completamente el promedio general de cada país.

Inciso 7:

Se contrastó la hipótesis de igualdad de medias entre `life_expectancy_female` y `life_expectancy`. Resultados obtenidos:

Valor del estadístico t: -3.79

P-valor: 2e-04

El resultado arroja un p-valor muy bajo (menor a 0.05), por lo que se rechaza la hipótesis nula. Esto significa que, en promedio, las mujeres viven significativamente más que el total de la población.

Inciso 8:

Estimamos una regresión múltiple de `life_expectancy` sobre `life_expectancy_female` e `income_per_person`. Los resultados obtenidos:

β_1 (vida femenina): $8.663e-01$

β_2 (ingreso): $3.202e-05$

R2: 0.8772

La esperanza de vida femenina sigue siendo la variable más influyente, pero el ingreso per cápita también contribuye positivamente: los países con mayor nivel de desarrollo económico tienden a tener mayor longevidad.

El R2 se mantiene elevado confirmando que la combinación de factores de género y nivel de vida explica prácticamente toda la variación en la esperanza de vida total.

Inciso 9:

Construimos un modelo alternativo de regresión múltiple usando las variables `child_mortality`, `life_expectancy_male` y `log(population)`. Tomamos el logaritmo de la población, ya que permite reducir asimetrías en variables muy grandes. Los resultados obtenidos:

β_1 (mortalidad infantil): -0.06549

β_2 (vida masculina): 0.71097

β_3 (población): 0.16660

R2: 0.9016

El modelo presenta un alto poder explicativo, ya que presenta un R2 alto. Por otro lado, vemos que los coeficientes son coherentes, el B1 es decir cuanto explica la mortalidad infantil a la expectativa de vida es negativa (mientras mayor sea la mortalidad infantil menor será la expectativa de vida). La esperanza de vida masculina es positiva y significativa. Y el coeficiente de la población (en log) es positivo pero es muy pequeño, lo que quiere decir que el tamaño de la población total no está tan relacionado con la expectativa de vida de la población.

Ejercicio: Simulación de ataque en el juego, TEG

Este ejercicio busca simular el proceso de ataque en el juego TEG, donde un jugador atacante enfrenta a un defensor que busca defender su país; todo esto se lleva a cabo mediante una tirada de dados.

Inciso 1:

La función `resultado_ataque` "recibe la cantidad de dados que tira el atacante y el defensor, genera tiradas aleatorias de 1 a 6 y compara los resultados ordenados de mayor a menor.

Se evalúan tantos enfrentamientos como el mínimo entre los dados lanzados de cada uno. En cada comparación de acuerdo a las reglas del TEG:

Si el dado del atacante es mayor, el defensor pierde una ficha.

En caso de empate o dado menor, pierde el atacante.

La función "resultado_ataque" devuelve un vector con las pérdidas de ambos jugadores ("perdidas_ataque", "perdidas_defensa"). Los resultados que obtuvimos de un ataque donde el atacante tiene 3 fichas y el defensor 2 (usando set.seed(1)) es:

```
resultado_ataque(3,2)
      perdidas_ataque perdidas_defensa
                2             0
```

En estos resultados, vemos que en un ataque el atacante pierde 2 fichas mientras que el defensor no pierde ninguna. Esta salida representa cuántas fichas pierde cada jugador en esa ronda puntual. No implica el resultado total de la batalla, solo la tirada.

Inciso 2:

La función "simular_batalla" simula una batalla completa, repitiendo enfrentamientos consecutivos hasta que según las reglas del TEG:

El atacante queda con una sola ficha (no puede seguir atacando).

El defensor pierde todas sus fichas (el territorio es conquistado).

En cada ronda se aplican las reglas del inciso anteriores. El atacante tira como máximo 3 dados, pero nunca más que fichas_atacante - 1. El defensor tira hasta 3 dados, sin superar ni su cantidad de fichas ni los dados del atacante.

La función simular_ataque devuelve una lista con las fichas finales de cada jugador y un indicador lógico conquistó que vale TRUE si el atacante logra vencer. El resultado que usamos de ejemplo es uno donde el atacante posee 3 fichas y el defensor 2, usando set.seed(1) obtuvimos que:

```
> simular_batalla(3,2)
$atacante
perdidas_ataque
                2

$defensor
perdidas_defensa
                0

$conquistó
perdidas_defensa
                TRUE
```

Inciso 3:

Finalmente, la función `probabilidad_ataque` ejecuta muchas simulaciones (en el ejemplo $B=1000$) de batallas con el mismo número inicial de fichas para ambos jugadores ($A=5$, $D=5$). En cada una se registra si el atacante conquista o no, y se calcula la proporción de sobre las 1000 simulaciones.

```
> probabilidad_ataque <- function(B = 1000, A = 5, D = 5) {  
+  
+   conquistas <- replicate (B , simular_batalla(A, D)$conquistó)  
+  
+   mcon <- mean(conquistas) # Media conquistas  
+  
+   list( estimación = mcon )  
+ }  
> set.seed(1)  
> probabilidad_ataque(B=1000)  
$estimación  
[1] 0.223
```

Según esta simulación vemos que el atacante gana solo en torno al 22.3% de las veces en este caso. Con 5 fichas de cada lado, el atacante tiene una probabilidad relativamente baja de conquistar, lo que coincide con la intuición de que los empates y las limitaciones de los dados benefician al defensor.