



Becas universitarias en el ICFES 2019-2

Integrantes:

Daniela Olarte

Camilo Campaz

Gabriel Suarez

Johan Ricardo

David Montaña

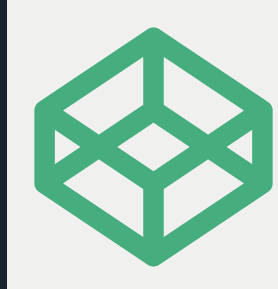


Presentación y descripción del caso

Se requiere estudiar la obtención de becas universitarias (Generación E), siendo estas:

"Generación E es el Programa que promueve el acceso, el bienestar y la permanencia de los jóvenes colombianos de escasos recursos en la Educación Superior del país." - GOV

Identificando los diferentes factores: económicos, sociales, educativos que determinaron la entrada a este programa (Generacion E - Excelencia).



Antecedentes

Un estudio realizado por el ICFES demostró que en las pruebas 2018-2:

- los hombres tuvieron mayores puntajes promedio que las mujeres en ambos pero, dichas diferencias no son considerables.
- Los colegios privados tuvieron mejor desempeño en las pruebas ICFES 2018 que los colegios públicos.
- Entre colegios del sector oficial también existe una brecha del puntaje global, ya que el puntaje de los colegios oficiales urbanos es superior al de los colegios oficiales rurales.
- Se apreció que, a medida que aumentan las características socioeconómicas de los estudiantes que componen una institución (oficial o privada), incrementa el puntaje global promedio.



Método de recolección

Base de datos oficial de ICFES, Pruebas ICFES 2019-2



Muestra

Tipo de muestreo: Intencional, opinático o de conveniencia

Tamaño: 11798



Población de estudio

Estudiantes caleños que presentaron la prueba ICFES 2019-2

VARIABLES

Cualitativas

Género
Nominal

Estrato Socio-Económico
Ordinal

Tipo de colegio
Nominal

Ubicación del colegio
Nominal

Estudio de los padres
Ordinal

Colegio Bilingüe
Nominal

Tiempo de Lectura
Ordinal

Tiempo en internet
Ordinal

Cuantitativa

Matemáticas
Continua - Razón

Sociales
Continua - Razón

Lectura crítica
Continua - Razón

Naturales
Continua - Razón

Inglés
Continua - Razón

Puntaje Global
Continua - Razón

Objetivo general

1

Objetivo

Identificar los diferentes factores que determinaron la obtención de becas para los estudiantes caleños en las pruebas ICFES 2019-2

Objetivos específicos

1

Exploración de datos

Explorar los datos obtenidos por cada una de nuestras variables

2

Hipótesis para la media

Determinar si los puntajes de los estudiantes no becados en el área de Lectura crítica fue mayor o igual al puntaje medio de todos los estudiantes en el año 2018

3

Hipótesis para la proporción

Determinar si el porcentaje de becados del año 2019 supero o fue igual a el umbral de becados frente a 2018 (1,5%)

4

Hipótesis para Muestras Independientes

Determinar cuál de los puntajes globales entre los estudiantes de estrato 1 frente a los estrato 3 es superior

Objetivos específicos

5

Planteamientos de Medias Dependientes

Determinar si la tendencia de que los resultados promedios en el área de matemáticas son mayores que en el área de inglés se repite en el año 2019

6

Análisis de varianza

Precisar cuál fue el estrato socio-económico que obtuvo un mayor puntaje en la prueba de inglés

7

Prueba de independencia (Chi - Squared)

Verificar si existe relación entre el género de los estudiantes que presentaron la prueba y su puntaje global

8

Regresión

Demostrar la relación que existe entre el puntaje global y el puntaje obtenido en la prueba de matemáticas

EXPLORACIÓN DE DATOS

Género

BaseDataBecados\$ESTU_GENERO		
	Frequency	Percent
F	32	44.44
M	40	55.56
Total	72	100.00



Ubicación

BaseDataBecados\$COLE_AREA_UBICACION		
	Frequency	Percent
RURAL	1	1.389
URBANO	71	98.611
Total	72	100.000



EXPLORACIÓN DE DATOS

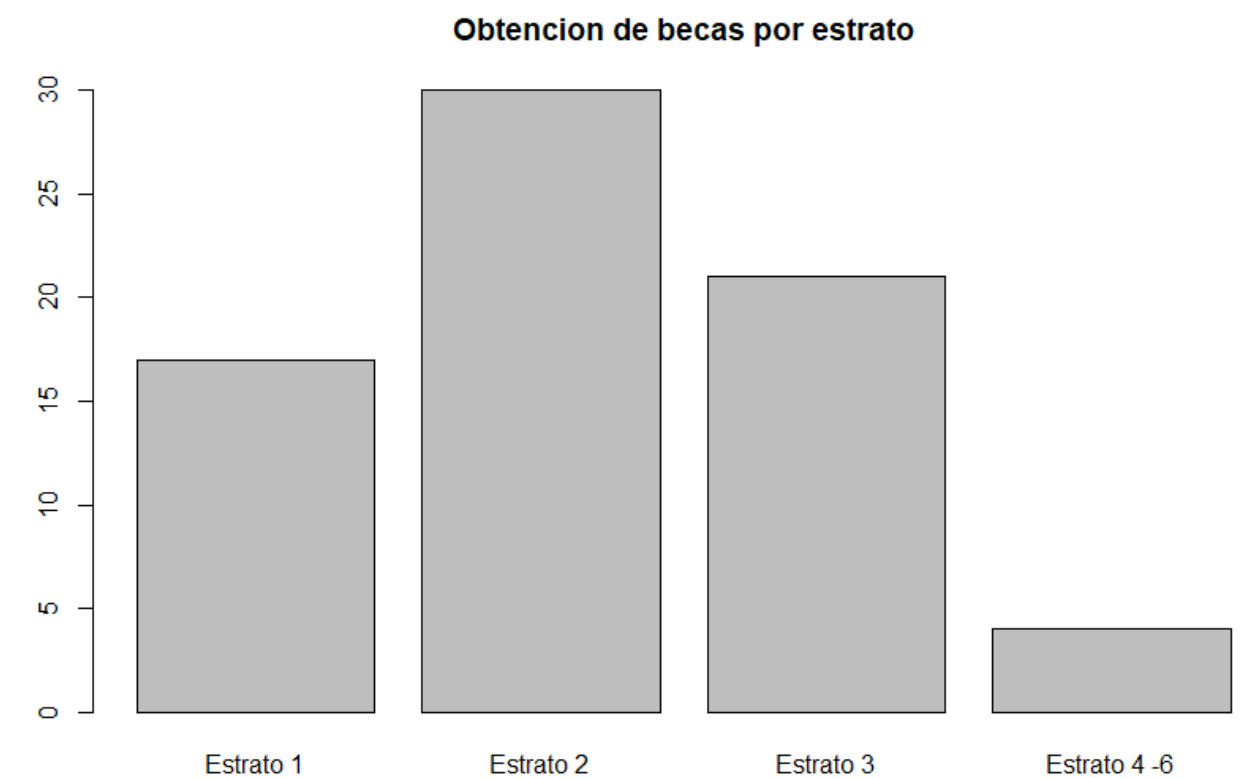
Tipo de Colegio

```
BaseDataBecados$COLE_NATURALEZA
Frequency Percent
NO OFICIAL      27    37.5
OFICIAL         45    62.5
Total          72   100.0
```



Estrato Becados

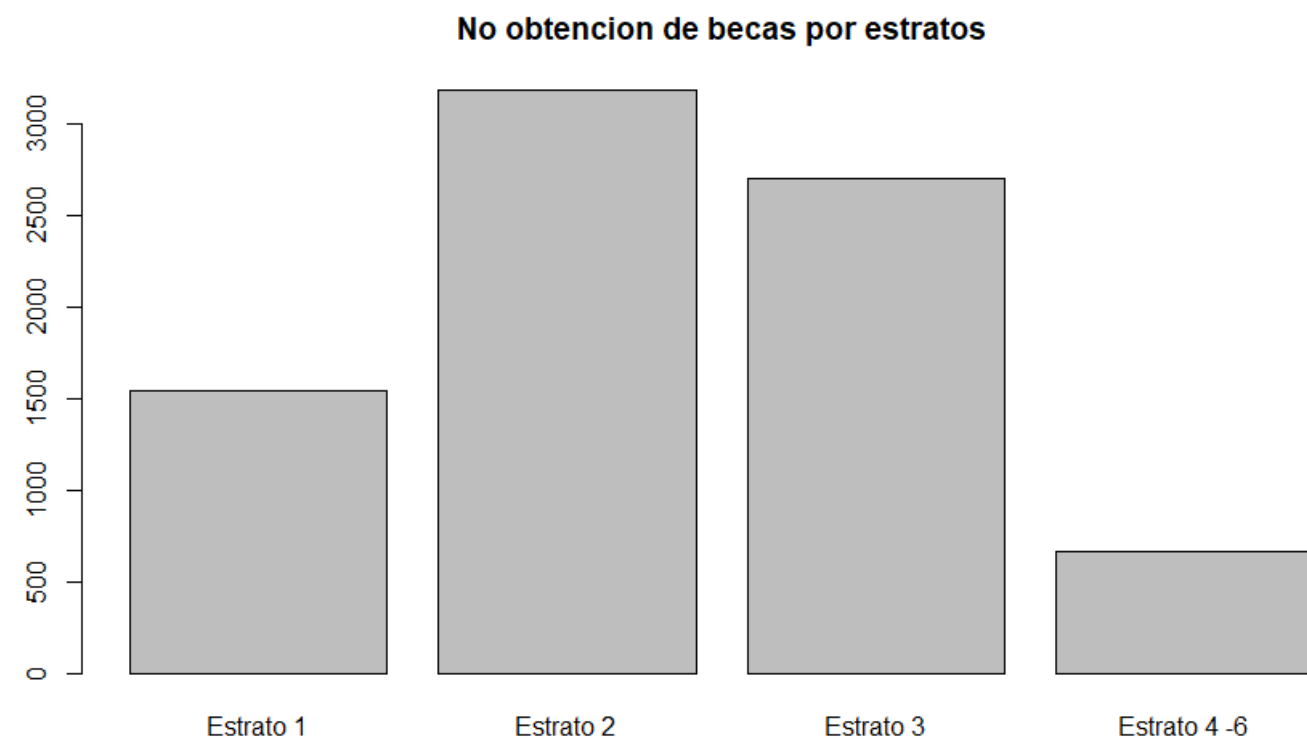
```
BaseDataBecados$FAMI_ESTRATOVIVIENDA
Frequency Percent
Estrato 1      17    23.611
Estrato 2      30    41.667
Estrato 3      21    29.167
Estrato 4 -6    4     5.556
Total          72   100.000
```



EXPLORACIÓN DE DATOS

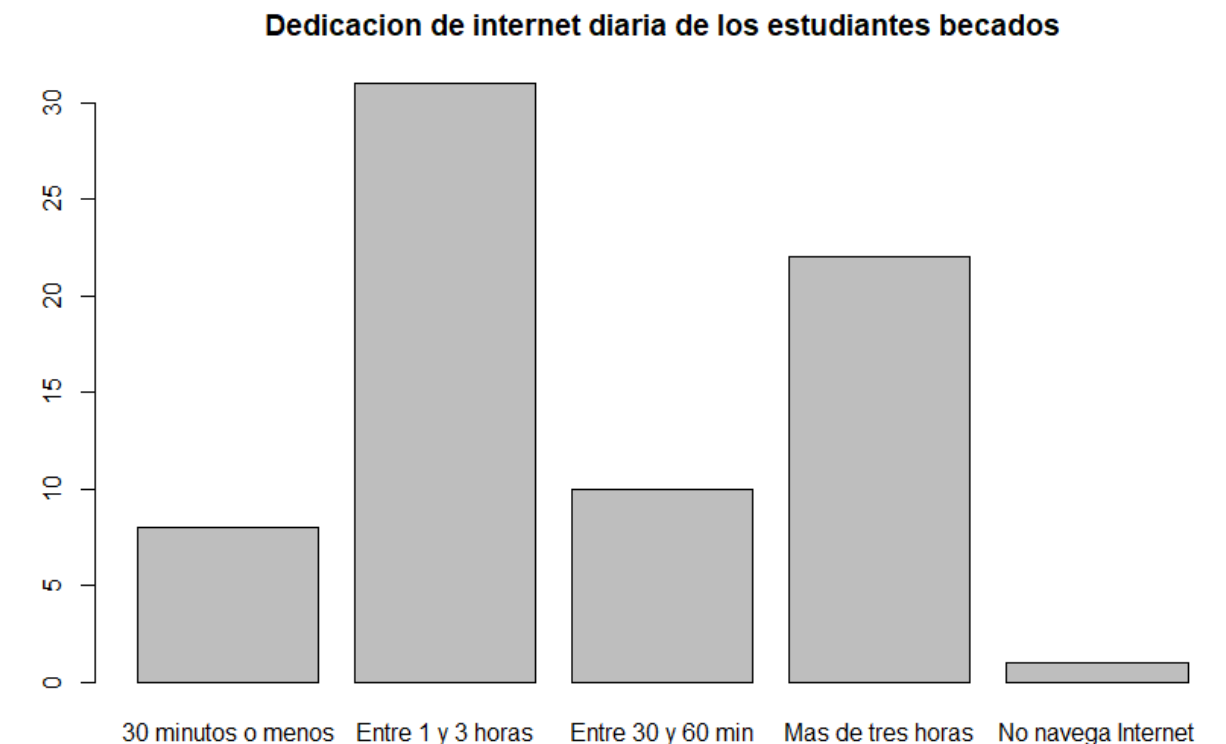
Estrato no Becados

```
BaseDataNoBecados$FAMI_ESTRATOVIVIENDA
      Frequency Percent
Estrato 1      1541   19.036
Estrato 2      3185   39.345
Estrato 3      2703   33.391
Estrato 4 -6     666    8.227
Total         8095  100.000
```



Dedicación a Internet

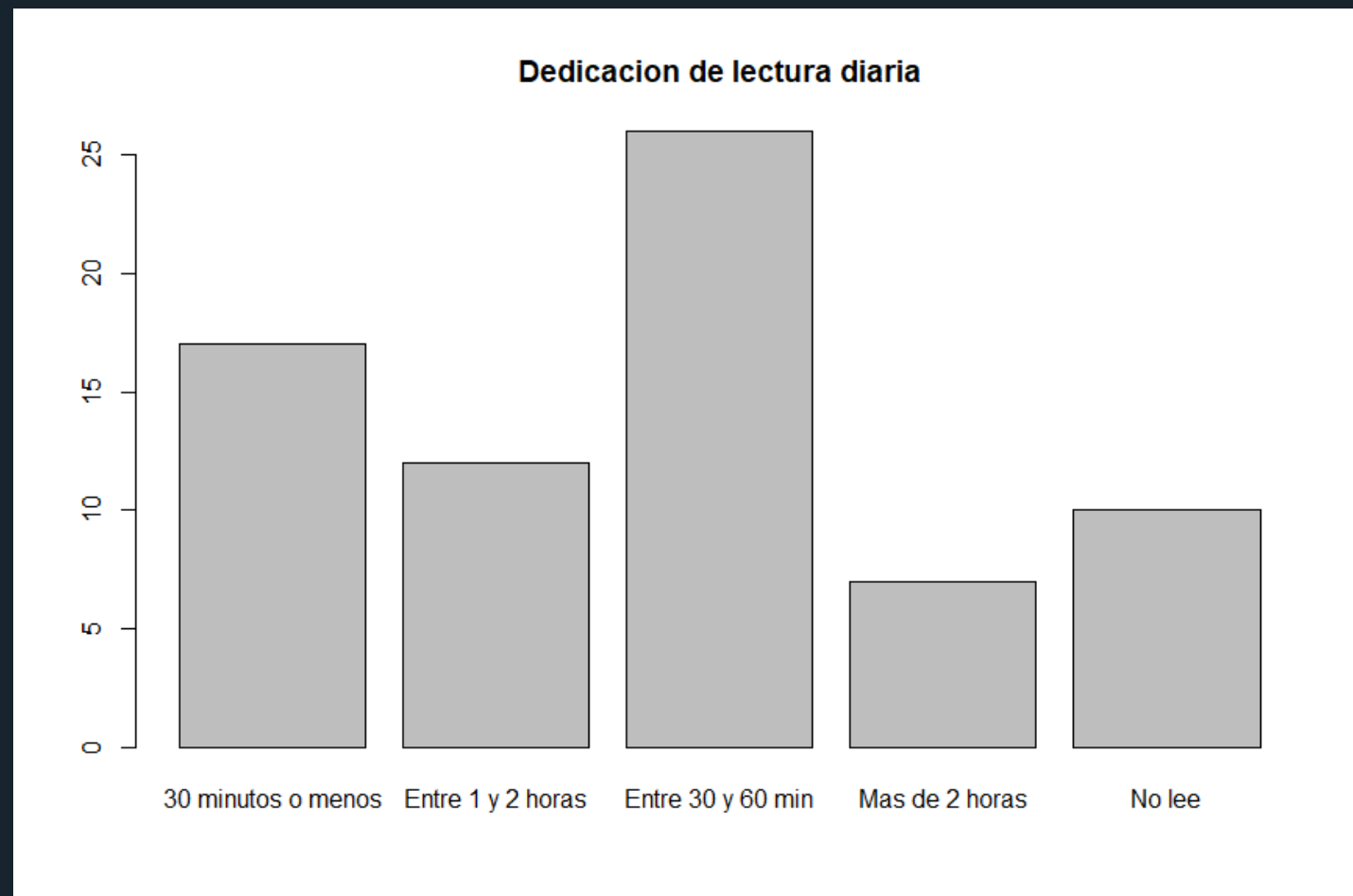
```
BaseDataBecados$ESTU_DEDICACIONINTERNET
      Frequency Percent
30 minutos o menos      8   11.111
Entre 1 y 3 horas      31  43.056
Entre 30 y 60 minutos   10  13.889
Más de 3 horas         22  30.556
No Navega Internet      1    1.389
Total                 72 100.000
```



EXPLORACIÓN DE DATOS

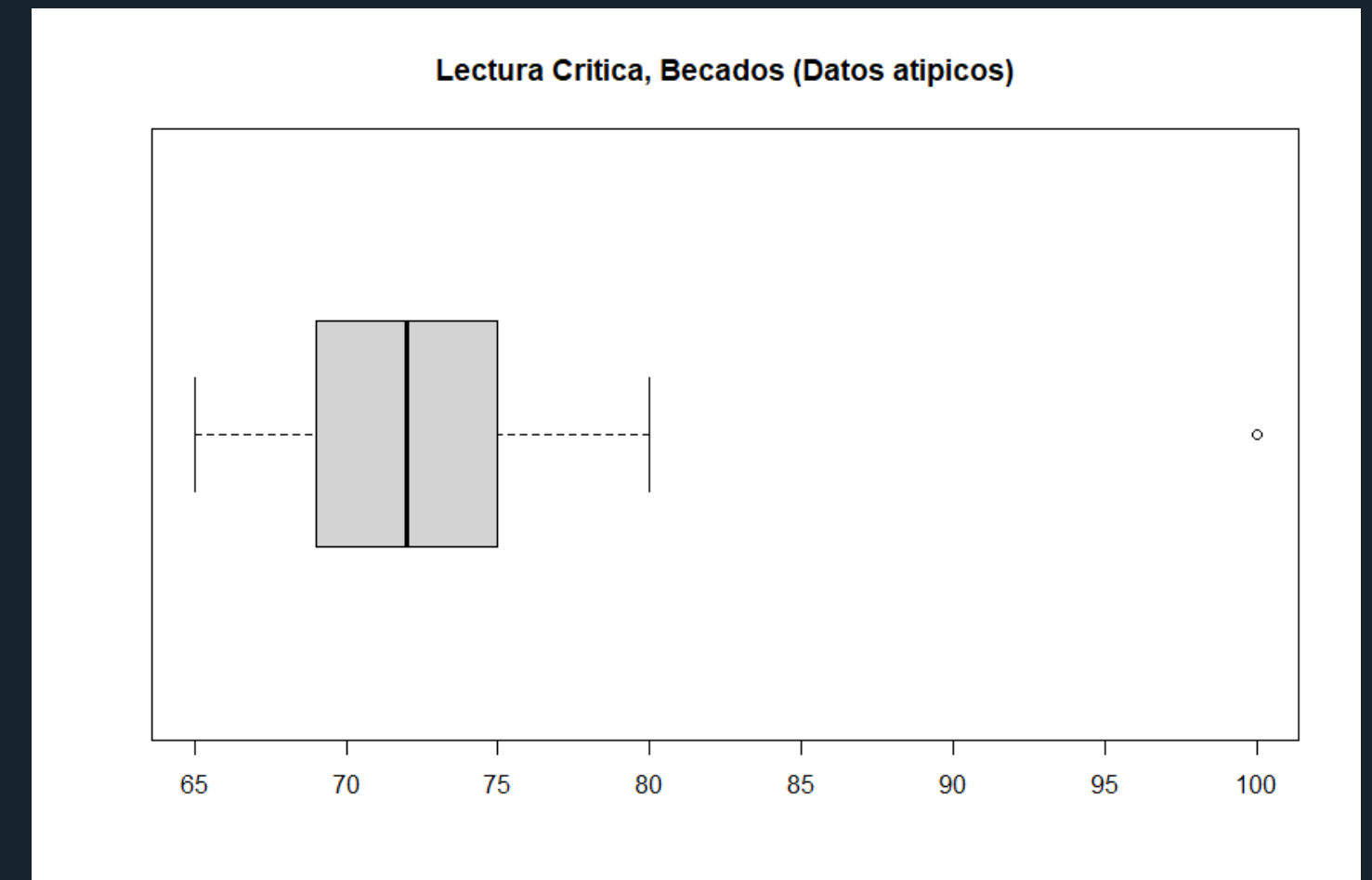
Lectura diaria

BaseDataBecados\$ESTU_DEDICACIONLECTURADIARIA		
	Frequency	Percent
30 minutos o menos	17	23.611
Entre 1 y 2 horas	12	16.667
Entre 30 y 60 minutos	26	36.111
Más de 2 horas	7	9.722
No leo por entretenimiento	10	13.889
Total	72	100.000



Lectura Crítica (Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	69.00	72.00	73.19	75.00	100.00



EXPLORACIÓN DE DATOS

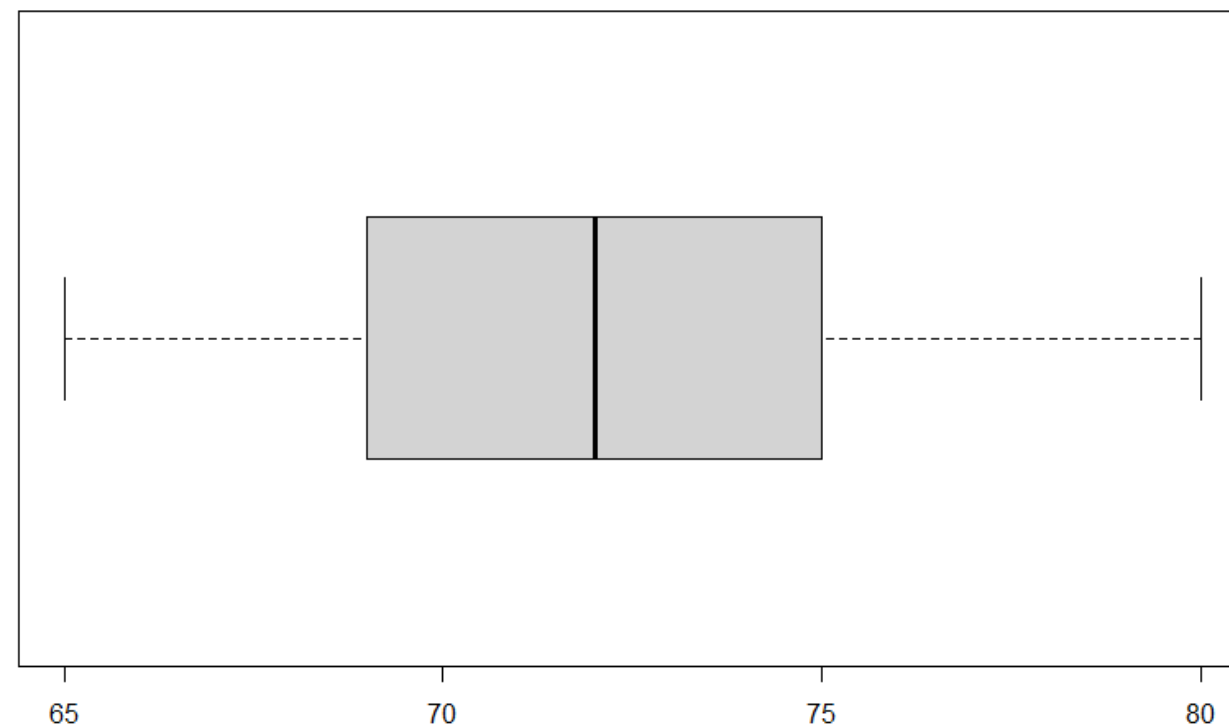
Lectura Crítica (No Atípicos)

Ciencias Naturales (Atípicos)

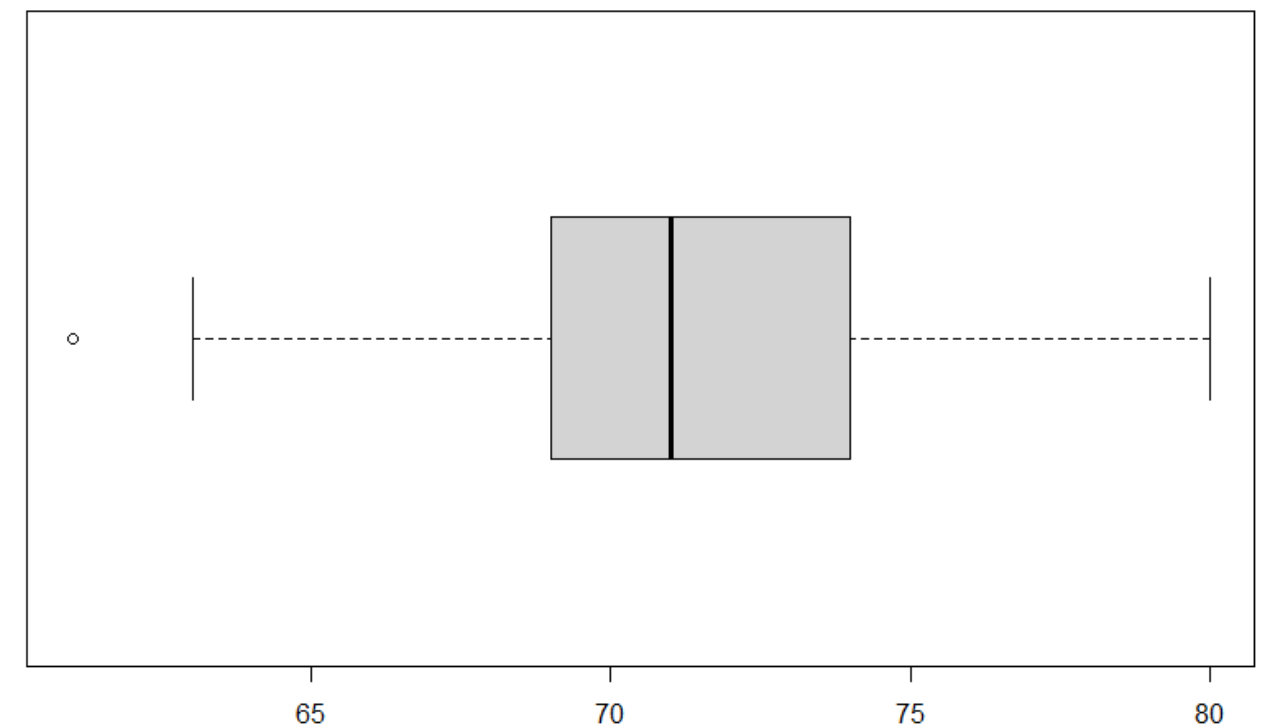
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	69.00	72.00	72.03	75.00	80.00

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
61.00	69.00	71.00	71.46	74.00	80.00

Lectura crítica, Becados (Sin Datos atípicos)



Ciencias naturales, Becados (Datos atípicos)

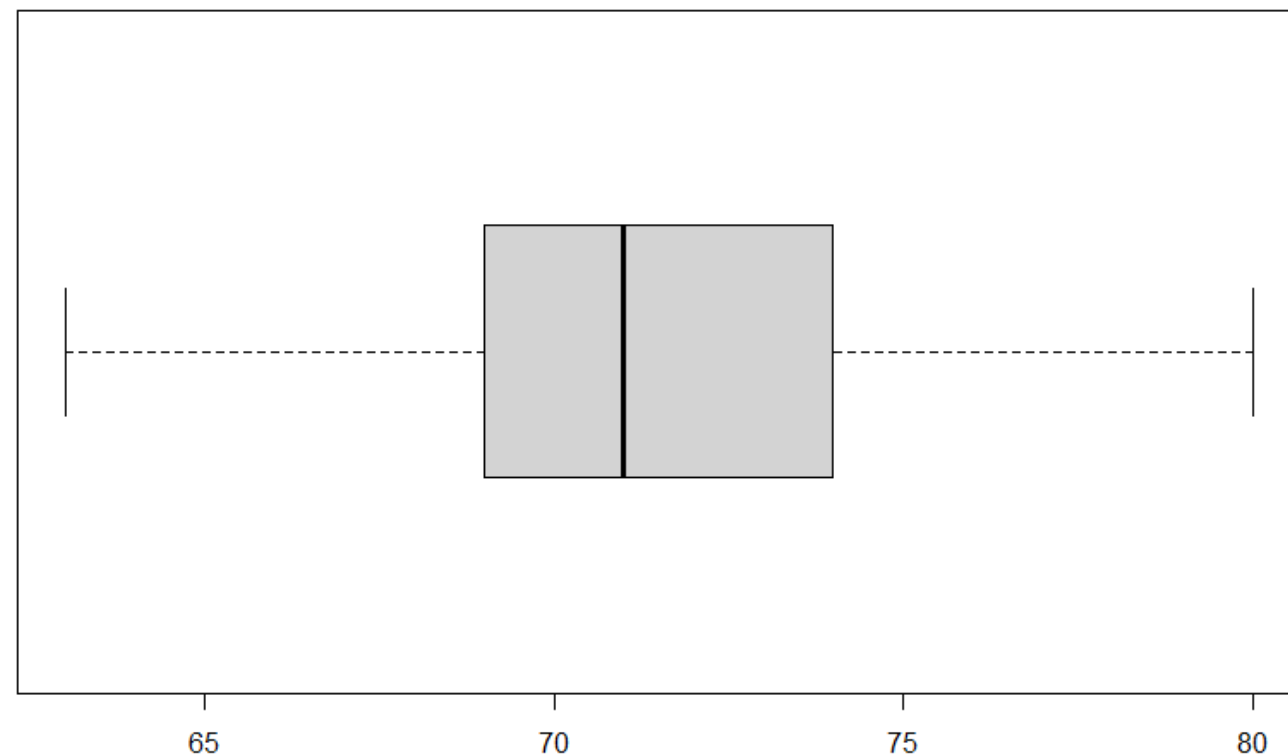


EXPLORACIÓN DE DATOS

Ciencias Naturales (No Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
63.00	69.00	71.00	71.61	74.00	80.00

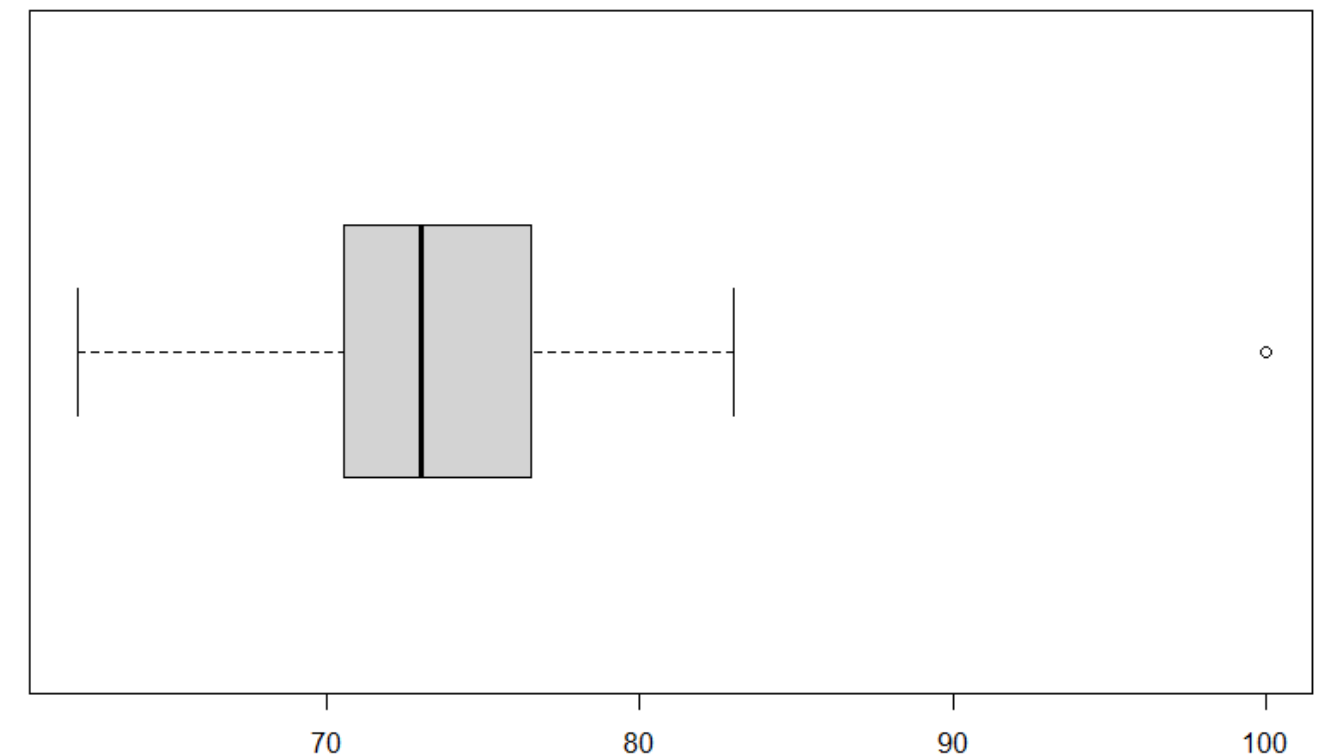
Ciencias naturales, Becados (Sin datos atipicos)



Matemáticas (Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
62.00	70.75	73.00	73.88	76.25	100.00

Matematicas, Becados (Datos atipicos)

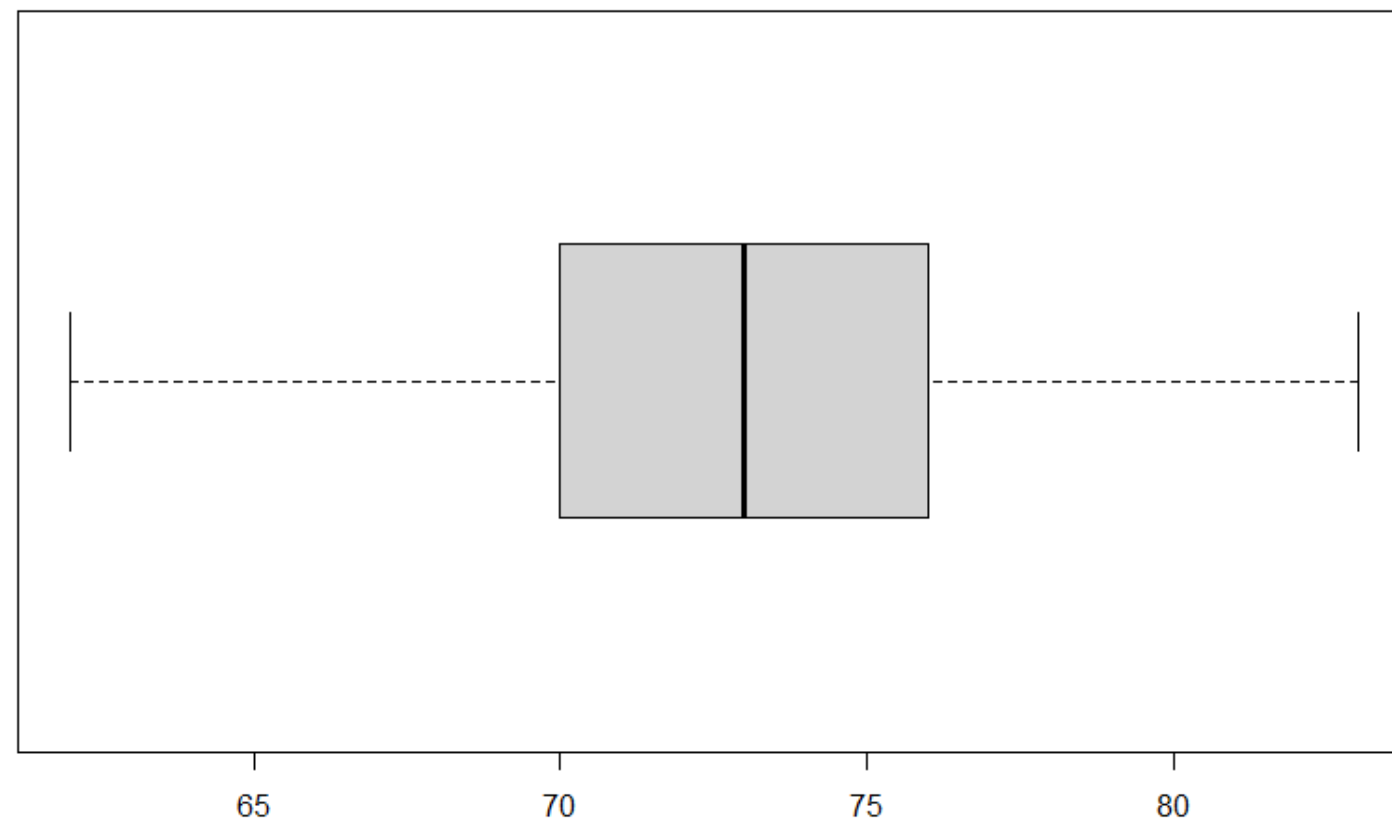


EXPLORACIÓN DE DATOS

Matemáticas (No Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
62.00	70.25	73.00	73.13	76.00	83.00

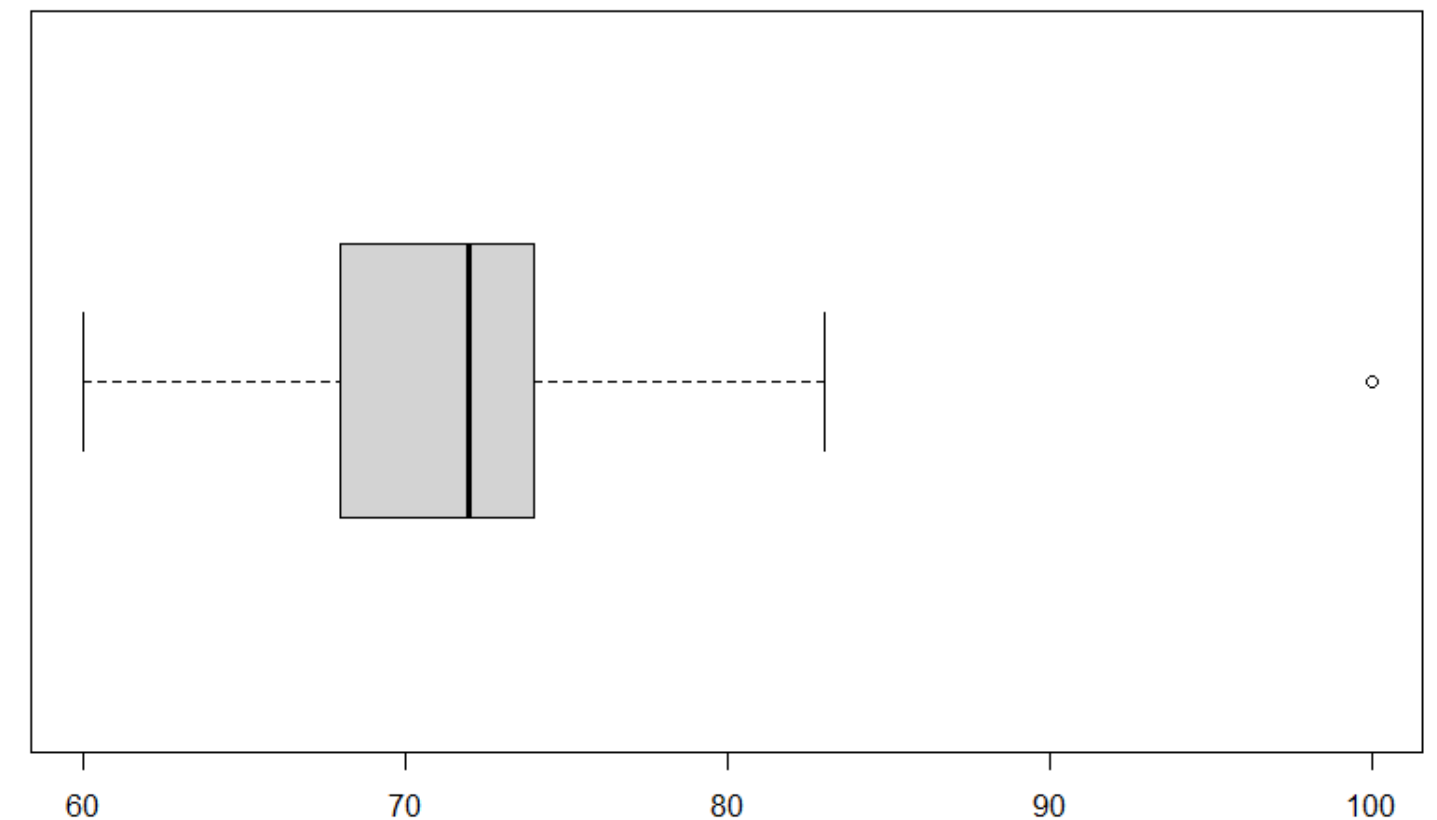
Matematicas, Becados (Sin datos atipicos)



Ciencias Sociales (Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
60.00	68.00	72.00	71.74	74.00	100.00

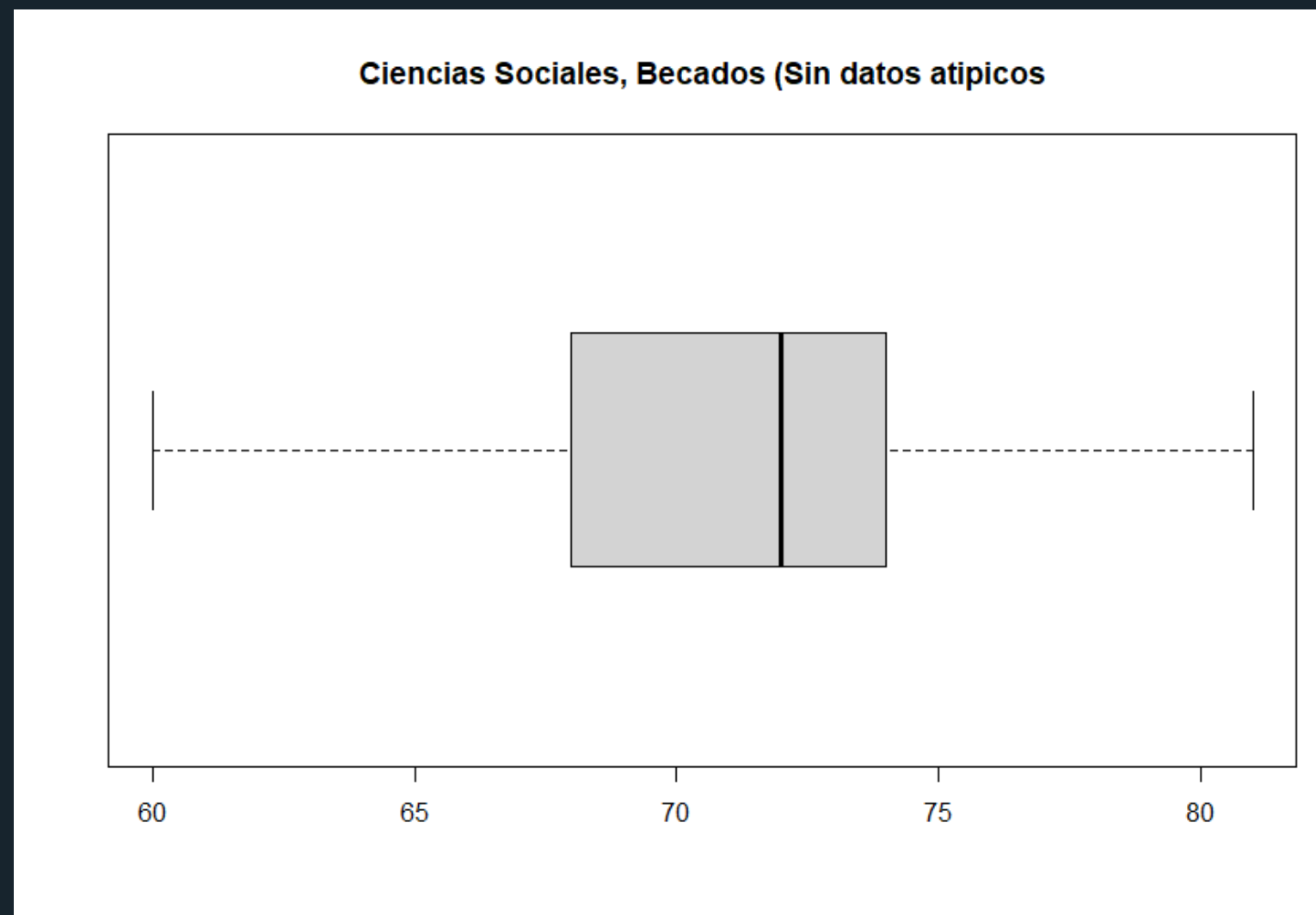
Ciencias Sociales, Becados (Datos atipicos)



EXPLORACIÓN DE DATOS

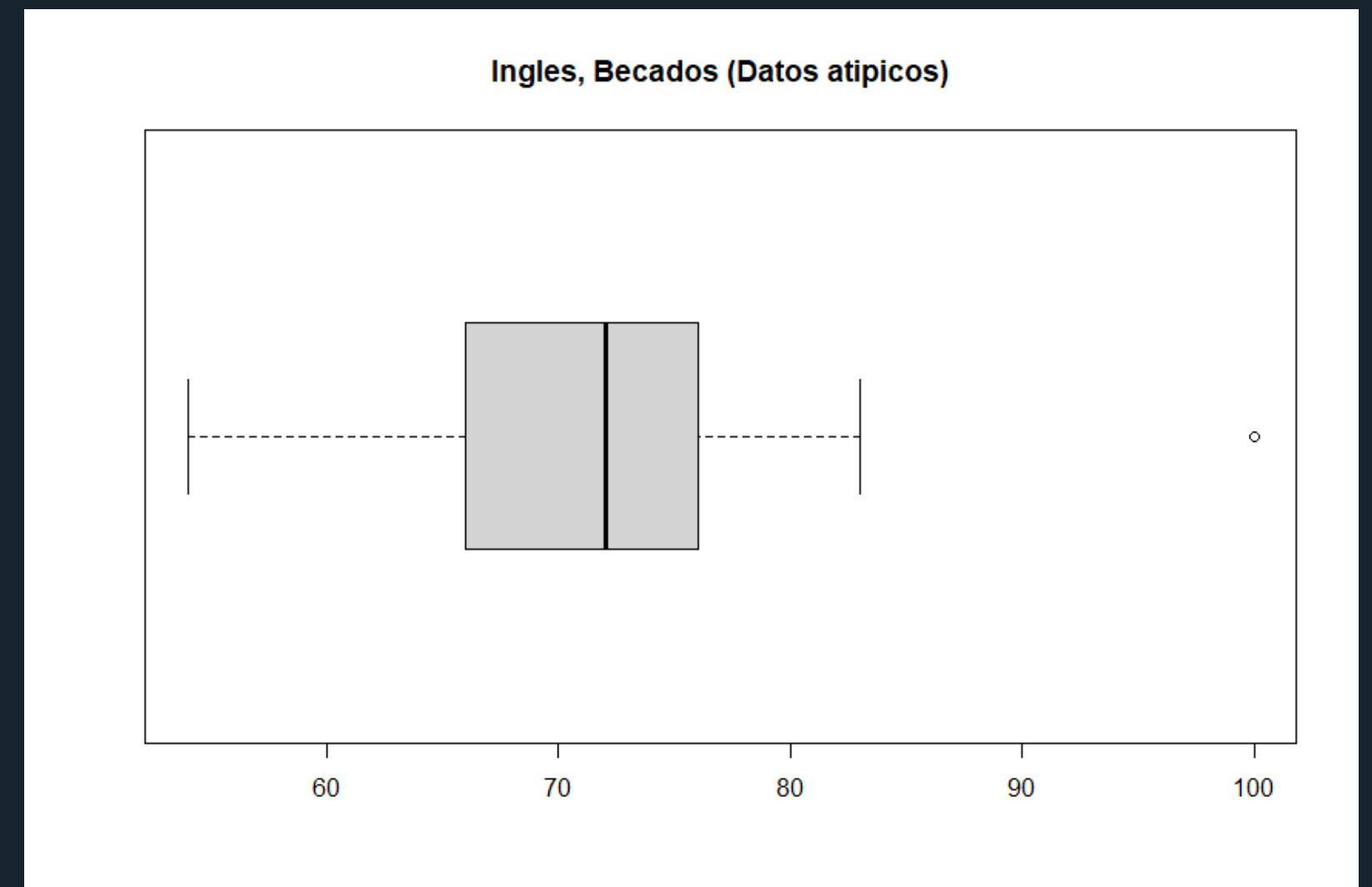
Ciencias Sociales (No Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
60.00	68.00	72.00	71.17	74.00	81.00



Inglés (Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
54.00	66.00	72.00	71.99	76.00	100.00

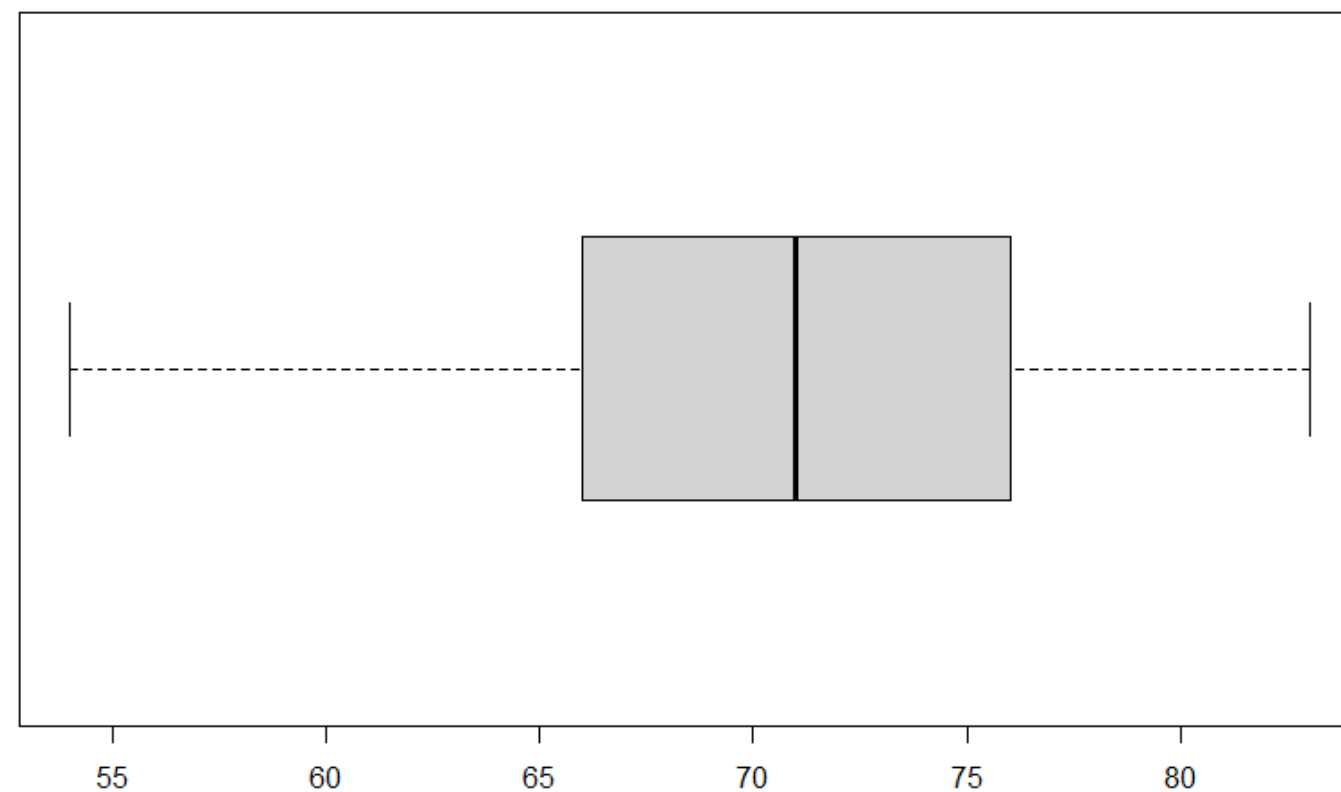


EXPLORACIÓN DE DATOS

Inglés (No Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
54.00	66.00	71.00	70.77	76.00	83.00

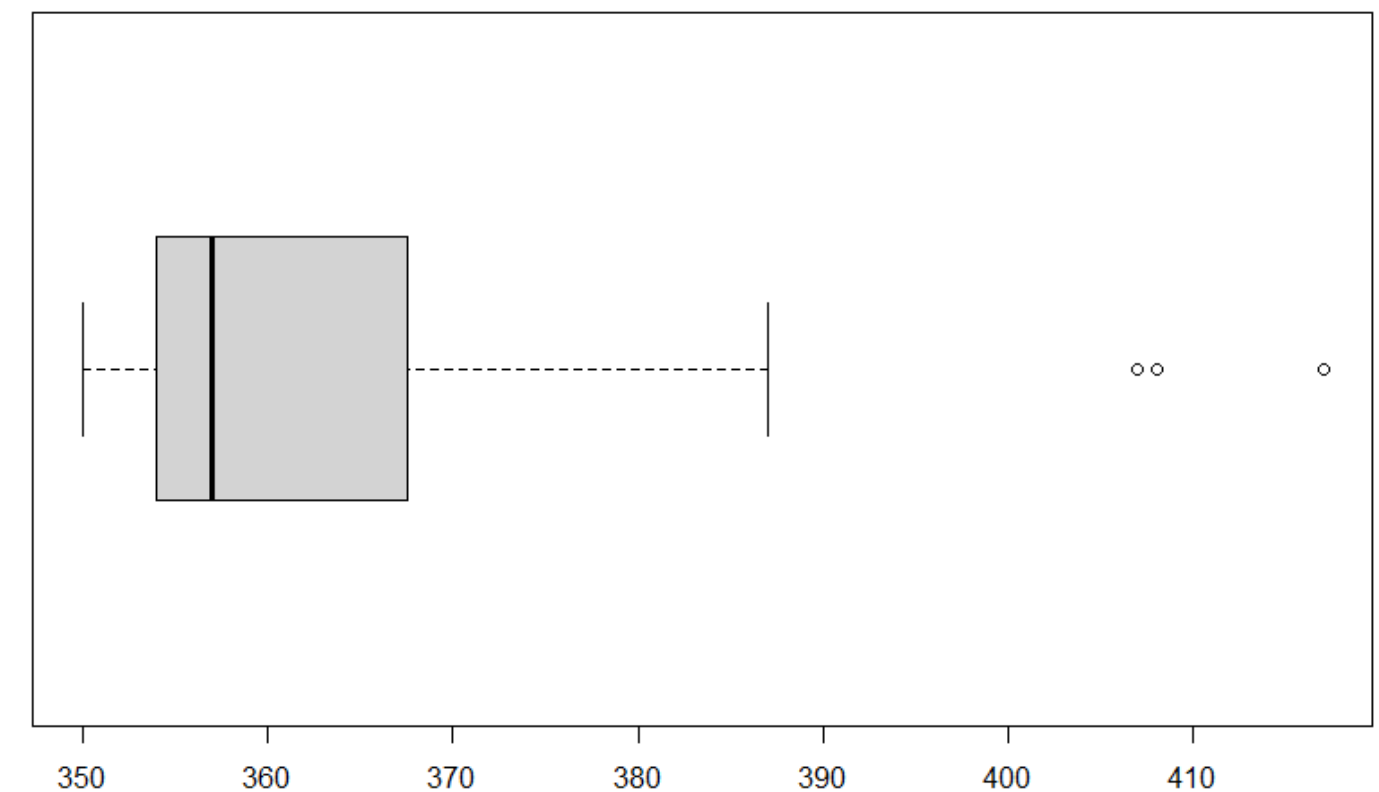
Ingles, Becados (Sin datos atipicos)



Puntaje Global (Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
350.0	354.0	357.0	362.6	367.2	417.0

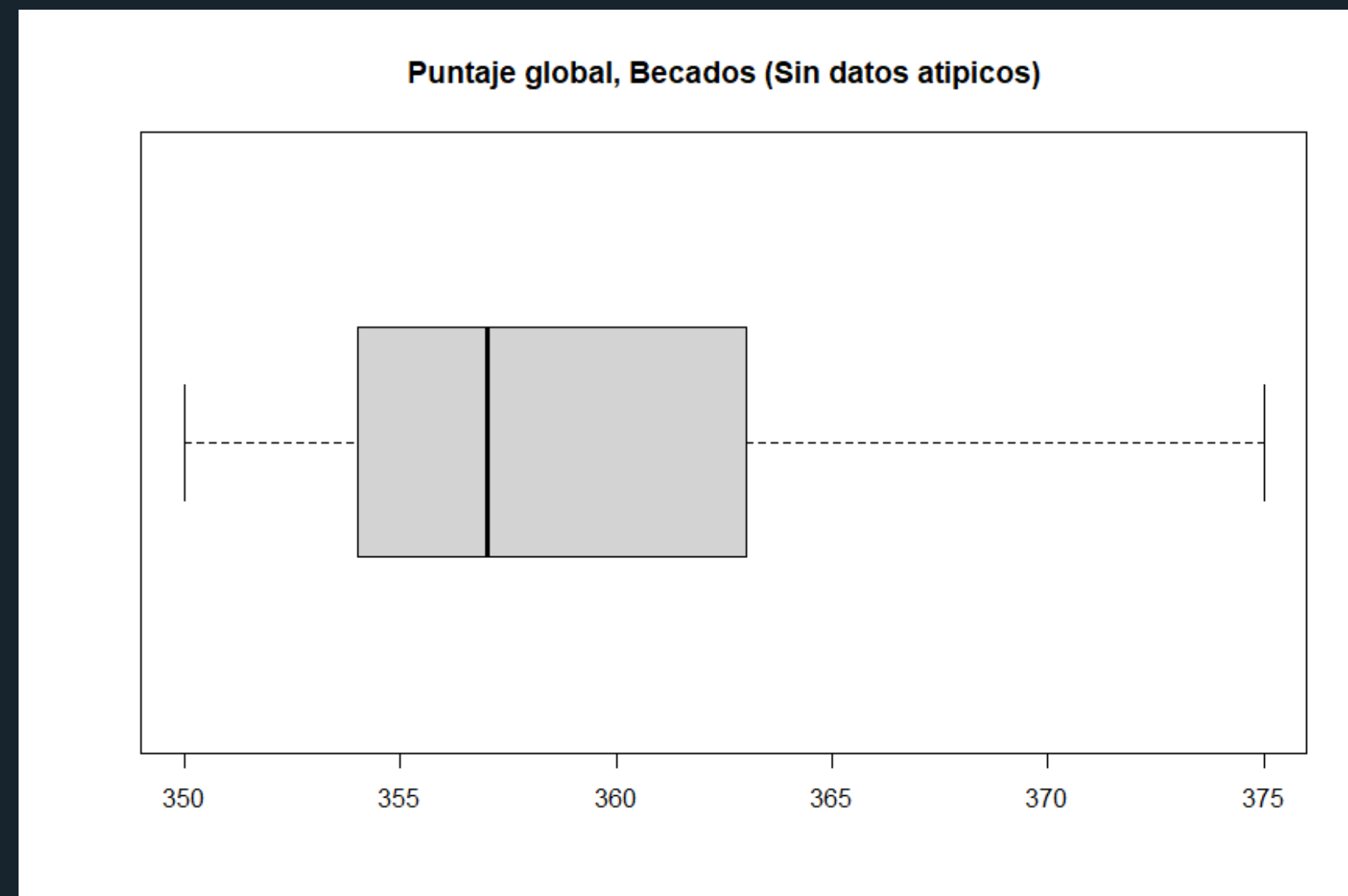
Puntaje global, Becados (Datos atipicos)



EXPLORACIÓN DE DATOS

Puntaje Global (No Atípicos)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
350	354	357	359	363	375



HIPÓTESIS PARA LA MEDIA

El promedio general de los estudiantes en el año 2018 en la prueba de Lectura Crítica fue de 62 puntos, ¿se podrá decir que en el año 2019 los estudiantes que no ganaron beca obtuvieron datos iguales o mayores que este?

Ho: El promedio en Lectura Crítica fue igual o mayor a 62, $\mu \geq 62$

H1: El promedio en Lectura Crítica fue menor a 62, $\mu < 62$

One Sample t-test

```
data: BaseDataNoBecados$PUNT_LECTURA_CRITICA
t = -73.789, df = 8094, p-value < 2.2e-16
alternative hypothesis: true mean is less than 62
95 percent confidence interval:
 -Inf 54.27944
sample estimates:
mean of x
 54.1034
```

Como el valor P es menor que el Valor P Crítico, Rechazamos Ho, por ende, podemos decir que el promedio de los estudiantes no superó el umbral de 62 puntos

HIPÓTESIS PARA LA PROPORCIÓN

Según un estudio realizado por el ICFES el porcentaje de becados en Colombia gracias a las pruebas ICFES por año es igual al 1,5 % de estudiantes, ¿se puede determinar si en las pruebas ICFES 2019-2 hubo mayor porcentaje de becados?

Ho: El porcentaje de becados fue mayor o igual a 1,5%, $p_i \geq 0,15$

H1: El porcentaje de becados fue menor a 1,5%, $p_i < 0,15$

```
1-sample proportions test without continuity correction

data:  nrow(PDataSucces) out of nrow(BaseData), null probability 0.15
X-squared = 1276.8, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is less than 0.15
95 percent confidence interval:
 0.00000000 0.01068752
sample estimates:
              p
0.008815967
```

Como el valor P es menor que el Valor P Crítico, Rechazamos Ho, por ende, podemos decir que en las pruebas ICFES 2019-2 la proporción de becados fue menor en la ciudad de Cali

HIPÓTESIS PARA MUESTRAS INDEPENDIENTES

Según un estudio realizado por el ICFES, hasta el 2018, la media de puntajes globales de los estudiantes de estrato 1 es mayor que la de los estudiantes de estrato 3; no obstante, ¿en las pruebas ICFES 2019-2 se podrá seguir observando esta tendencia?

Ho: Asumimos varianzas iguales, $V1 = V2$

H1: No Asumimos varianzas iguales, $V1 \neq V2$

```
F test to compare two variances

data:  E1Data$PUNT_GLOBAL and E3Data$PUNT_GLOBAL
F = 1.0132, num df = 1557, denom df = 2723, p-value = 0.7675
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9282417 1.1070798
sample estimates:
ratio of variances
      1.013182
```

Como el valor P es mayor que el Valor P Crítico, No rechazamos Ho, por ende, podemos asumir varianzas iguales, lo que nos permite seguir con nuestra prueba de hipótesis

HIPÓTESIS PARA MUESTRAS INDEPENDIENTES

Según un estudio realizado por el ICFES, hasta el 2018, la media de puntajes globales de los estudiantes de estrato 1 es mayor que la de los estudiantes de estrato 3; no obstante, ¿en las pruebas ICFES 2019-2 se podrá seguir observando esta tendencia?

Ho: Media del estrato 1 mayor o igual a media del estrato 3, $\mu_1 \geq \mu_2$

H1: Media del estrato 1 menor que media del estrato 3, $\mu_1 < \mu_2$

```
Welch Two Sample t-test

data:  E1Data$PUNT_GLOBAL and E3Data$PUNT_GLOBAL
t = -8.745, df = 3223.7, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -10.49159
sample estimates:
mean of x mean of y
 248.1714  261.0943
```

Como el valor P es menor que el Valor P Crítico, Rechazamos Ho, por ende, decimos que la media de los estratos 1 es menor a la media de los estratos 3 en el puntaje global, en las pruebas ICFES 2019-2

HIPÓTESIS PARA MUESTRAS DEPENDIENTES

Según el ICFES en el año 2018, los estudiantes becados que presentaron el ICFES obtuvieron mayor media en sus resultados en el área de matemáticas que en inglés, ¿se puede determinar si esta tendencia se vuelve a presentar en las pruebas ICFES 2019-2?

Ho: $\mu d \leq 0$

H1: $\mu d > 0$

```
Paired t-test

data:  BaseDataBecados$PUNT_MATEMATICAS and BaseDataBecados$PUNT_INGLES
t = 1.692, df = 71, p-value = 0.04752
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.02832985      Inf
sample estimates:
mean of the differences
      1.888889
```

Como el valor P es menor que el Valor P Crítico, Rechazamos Ho, por ende, los estudiantes becados obtuvieron un mayor puntaje en la prueba de matemáticas que en la de inglés, demostrándonos así una diferencia positiva, que nos ayuda a afirmar lo anterior, lo que nos dice que la tendencia se volvió a presentar

ANÁLISIS DE VARIANZA (ANOVA)

Un estudio realizado por el ifces indica qué los estudiantes , sin importar el estrato , obtienen los mismos resultados en la prueba de inglés ¿se puede determinar si esta tendencia se vuelve a presentar en las pruebas ICFES 2019-2?

Ho: todas las medias son iguales

H1: Al menos dos difieren

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FAMI_ESTRATOVIVIENDA	3	21747	7249	54.55	<2e-16 ***
Residuals	8163	1084678	133		

Rechazamos Ho entonces almenos 2 medias difieren

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = lm(PUNT_INGLES ~ FAMI_ESTRATOVIVIENDA))

```
$FAMI_ESTRATOVIVIENDA
              diff      lwr      upr    p adj
Estrato 2-Estrato 1  1.9139797  0.9996434  2.828316 0.0000005
Estrato 3-Estrato 1  4.2497205  3.3088703  5.190571 0.0000000
Estrato 4 -6-Estrato 1  4.4612094  3.0927846  5.829634 0.0000000
Estrato 3-Estrato 2   2.3357408  1.5643987  3.107083 0.0000000
Estrato 4 -6-Estrato 2  2.5472296  1.2893124  3.805147 0.0000012
Estrato 4 -6-Estrato 3  0.2114888 -1.0658303  1.488808 0.9741410
```

El pos Anova nos muestra que los estratos que poseen una media mayor son los estratos 3,4,5 y 6

PRUEBA DE INDEPENDENCIA (CHI - SQUARED)

¿Existe relación significativa entre el género y el puntaje global en las pruebas ICFES 2019-2?

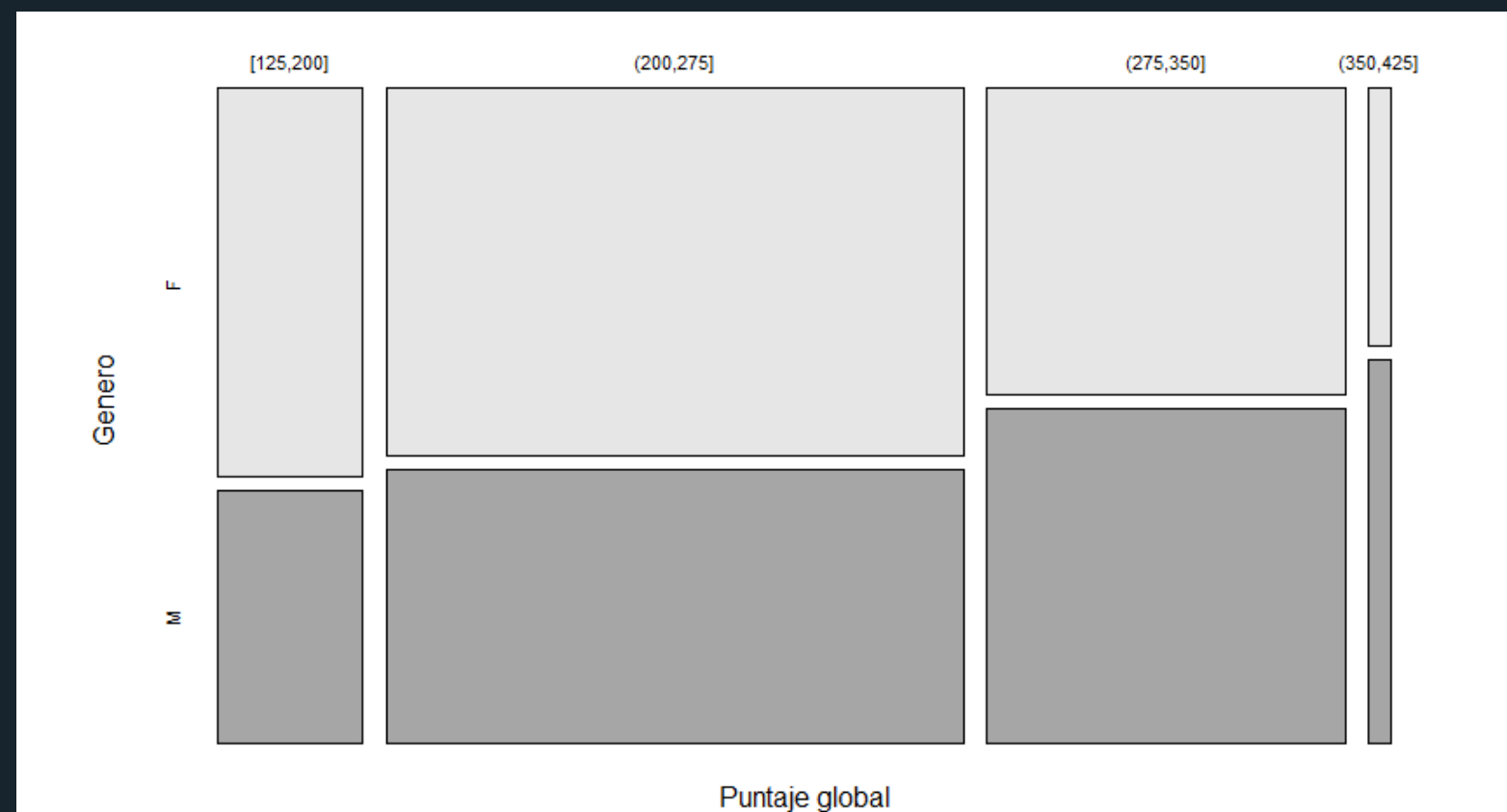
Ho: No existe relación significativa entre el género y los puntajes globales

H1: Si existe relación entre el género y los puntajes globales

ESTU_GENERO	puntG [125,200]	(200,275]	(275,350]	(350,425]	Total
F	652 14.7%	2451 55.3%	1265 28.5%	66 1.5%	4434 54.3%
M	424 11.4%	1827 48.9%	1384 37.1%	98 2.6%	3733 45.7%
Total	1076	4278	2649	164	8167

Pearson's Chi-squared test

data: ESTU_GENERO and puntG
X-squared = 91.425, df = 3, p-value < 2.2e-16



Como el valor P es menor que el Valor P Crítico, Rechazamos Ho, por ende, decimos que si existe relación significativa entre el género y el puntaje obtenido

REGRESIÓN

¿Existe relación lineal entre el puntaje global y el puntaje de matemáticas?

Ho: No existe relación lineal entre el puntaje global y el puntaje en matemáticas

H1: Existe relación lineal entre el puntaje global y el puntaje en matemáticas

```
Pearson's product-moment correlation
data: Puntaje Global and Puntaje Matematicas
t = 29.513, df = 56, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9485108 0.9818042
sample estimates:
      cor
0.9693254
```

Como valor P es menor que el valor P Crítico, asumimos que existe relación, además de esto, observamos que el coeficiente de correlación es igual a 0.96, por ende decimos que existe una relación directamente proporcional

```
Residuals:
      Min       1Q   Median       3Q      Max
-2.39900 -0.39096 -0.01194  0.42355  0.98806

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -83.45796    5.30093   -15.74  <2e-16 ***
`Puntaje Global`  0.43549    0.01476    29.51  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

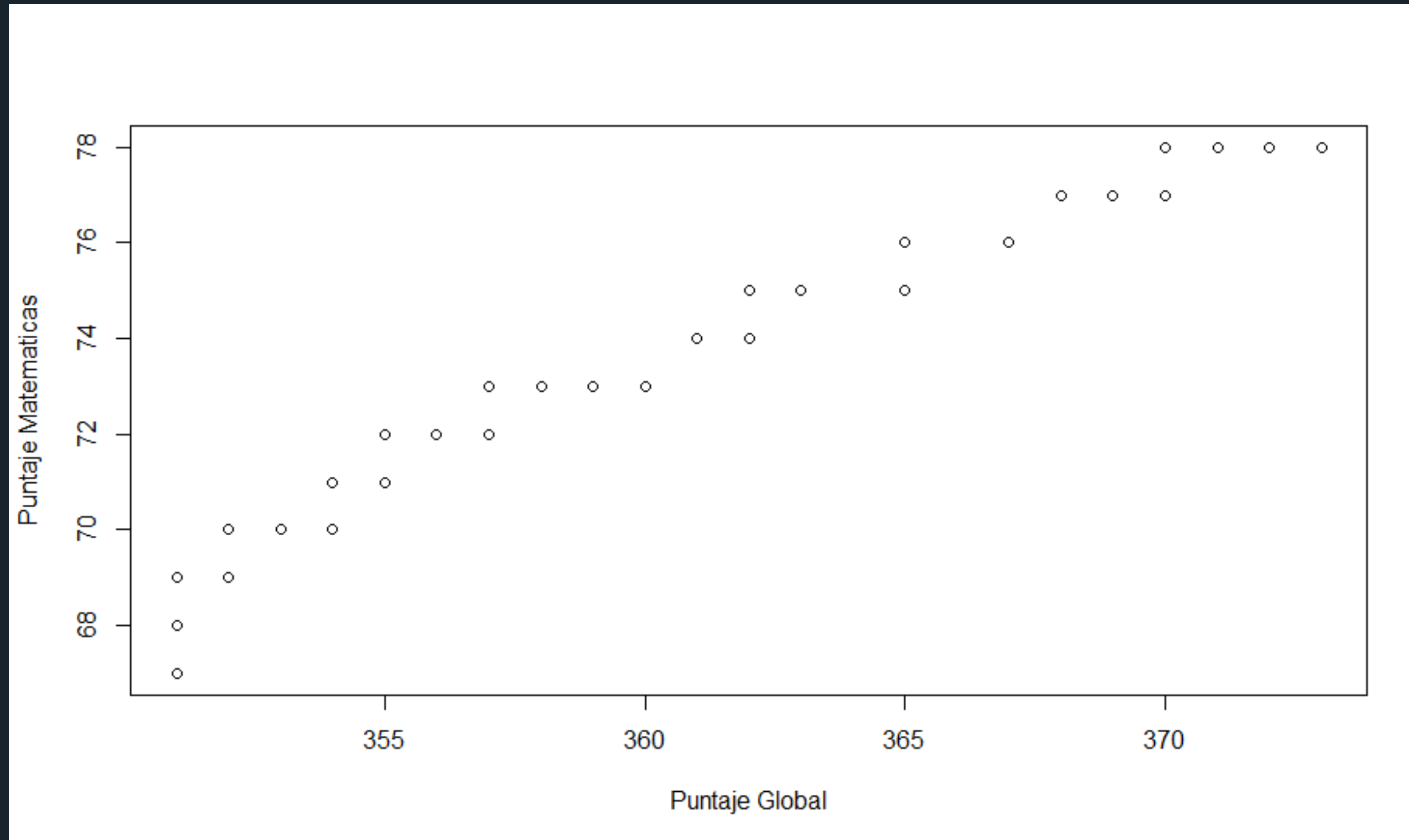
Residual standard error: 0.6998 on 56 degrees of freedom
Multiple R-squared:  0.9396,    Adjusted R-squared:  0.9385
F-statistic: 871 on 1 and 56 DF,  p-value: < 2.2e-16
```

Coeficiente de Determinación = 0.93, relación fuerte

Coeficiente Beta 1 = 0.43, por cada punto que se aumenta en el puntaje global, el puntaje en matemáticas sube en 0.43 unidades

REGRESIÓN

$$Y(X) = 0.43(X) - 83.4$$



REGRESIÓN

Durbin-Watson test

```
data: Regresion
DW = 0.51745, p-value = 2.596e-12
alternative hypothesis: true autocorrelation is not 0
```

Ho: Los residuos no están correlacionados
H1: Los residuos están correlacionados
P Valor menor que P crítico, rechazo Ho, los residuos están correlacionados

Two-sample Kolmogorov-Smirnov test

```
data: Puntaje Global and Puntaje Matematicas
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Ho: Los residuos son normales
H1: Los residuos no son normales
P Valor menor que P crítico, rechazo Ho, los residuos no son normales

studentized Breusch-Pagan test

```
data: Regresion
BP = 5.8436, df = 1, p-value = 0.01563
```

Ho: Los residuos son homocedasticos
H1: Los residuos son heterocedasticos
P Valor menor que P crítico, rechazo Ho, los residuos son heterocedastico

Los residuos no son homogéneos, lo que nos dice que presentan heterocedasticidad, lo que conlleva a una pérdida de eficiencia en el estimador mínimo cuadrático. Los estimadores siguen siendo insesgados pero ya no eficientes, es decir que los estimadores ya no serán de varianza mínima.

Conclusiones

O.E 1

Se observaron el comportamiento de todos los datos

O.E 2

Los estudiantes que presentaron las pruebas ICFES 2019-2 no obtuvieron un promedio mayor a 62 puntos en la prueba de Lectura Critica, por lo que no se supero a las pruebas del 2018

O.E 3

El porcentaje de becados en las pruebas ICFES 2019-2 no supero al porcentaje de becados del año 2018, dado a que se obtuvo un porcentaje por debajo del 1,5% de becados

O.E 4

La media de los puntajes globales para los estratos 1 es menor a la media de los estratos 3 en las pruebas ICFES 2019-2

Conclusiones

O.E 5

Los estudiantes becados obtuvieron un mayor puntaje en la prueba de matemáticas que en la de inglés, demostrándonos así una diferencia positiva, que nos dice que la tendencia se mantuvo

O.E 6

Los estratos socio-económicos que obtuvieron un mayor puntaje en la prueba de ingles son los estratos 3, 4, 5 y 6

O.E 7

Se puede decir que el genero tiene una relación significativa con los puntajes globales de las pruebas ICFES 2019-2

O.E 8

Los coeficientes de regresión nos demuestra que existe una relación lineal entre el puntaje de matemáticas e ingles, sin embargo a la hora de realizar la prueba de normalidad se observa que los datos presentan una heterocedasticidad, por lo que la regresión pierde confianza.