Sesion 2

Manuel Toral

January 19, 2019

Llamar los datos

Select

county_id	name	county_income
Min. : 1003	Length:467	Min. : 22545
1st Qu.: 8022	Class :character	1st Qu.: 43804
Median :24033	Mode :character	Median : 50856
Mean :25434	NA	Mean : 52527
3rd Qu.:40112	NA	3rd Qu.: 56832
Max. :56005	NA	Max. :110292

Filter

Filter + Select

raceethnicity	armed
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Vehicle
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Vehicle
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm

raceethnicity	armed
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Vehicle
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm
Hispanic/Latino	Firearm

Group by + summarise

```
## # A tibble: 47 x 2
##    state  media
##    <chr>  <dbl>
##  1 AK    77454
##  2 AL    47174.
##  3 AR    45998
##  4 AZ    48391
##  5 CA    61058.
##  6 CO    53060.
##  7 CT    61996
##  8 DC    65830
##  9 DE    59843
## 10 FL    47964.
## # ... with 37 more rows
```

Group by + arrange

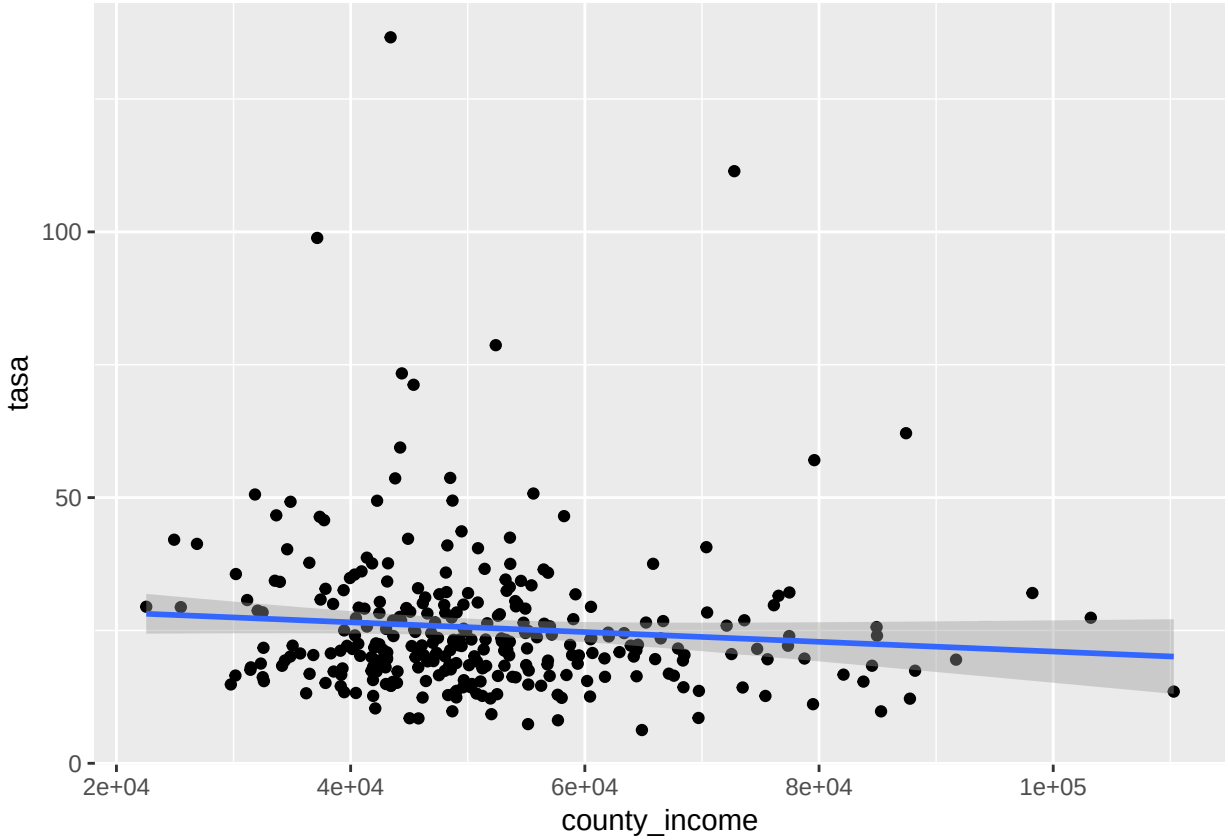
```
## # A tibble: 459 x 5
## # Groups:   state, county_id, county_income [301]
##    state county_id county_income   pop count
##    <chr>     <int>         <int> <int> <int>
##  1 AZ         4013         53596  4579     2
##  2 AZ         4025         42987  5206     2
##  3 CA         6065         56529  2886     2
##  4 NJ        34003         83794  8810     2
##  5 OK        40143         48181  3300     2
##  6 TX        48113         49481  4775     2
##  7 TX        48467         43439  6852     2
##  8 WA        53021         55177  6747     2
##  9 AK         2020         77454  2619     1
## 10 AK         2020         77454  5733     1
## # ... with 449 more rows
```

Mutate

```
## # A tibble: 301 x 6
##    state county_id county_income count   pop  tasa
```

```
##    <chr>     <int>         <int> <int> <int> <dbl>
##  1 FL        12111         43413     1   732 137.
##  2 HI        15003         72764     2  1795 111.
##  3 LA        22071         37146     2  2023  98.9
##  4 NY        36055         52394     1  1271  78.7
##  5 TX        48305         44375     1  1363  73.4
##  6 MI        26017         45376     1  1404  71.2
##  7 MD        24003         87430     1  1610  62.1
##  8 TX        48337         44231     1  1683  59.4
##  9 NJ        34023         79596     1  1753  57.0
## 10 LA        22033         48506     1  1862  53.7
## # ... with 291 more rows
```

Regresión



$$Y = \beta_1 X_1 + \epsilon$$

$$\text{Tasa por cada cien mil hab.}_i = \beta_1 \text{Ingreso}_1 + \epsilon$$

Modelo 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	30.2	3.16	9.558	4.628e-19
county_income	-9.175e-05	5.919e-05	-1.55	0.1222

Table 4: Fitting linear model: tasa ~ county_income

Observations	Residual Std. Error	R^2	Adjusted R^2
301	14.2	0.007971	0.004653

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: sáb, ene 19, 2019 - 13:00:50

Table 5:

	Dependent variable:
	tasa
county_income	−0.0001
	(0.0001)
Constant	30.203***
	(3.160)
Observations	301
R^2	0.008
Adjusted R^2	0.005
Residual Std. Error	14.198 (df = 299)
F Statistic	2.403 (df = 1; 299)
Note:	*p<0.1; **p<0.05; ***p<0.01

Modelo 2

Dependent variable:

tasa

(1)

(2)

county_income

-0.0001

(0.0001)

incpp

1.388***

(0.070)

Constant

30.203***

10.315***

(3.160)

(0.934)

Observations

301

301

R2

0.008

0.569

Adjusted R2

0.005

0.568

Residual Std. Error (df = 299)

14.198

9.359

F Statistic (df = 1; 299)

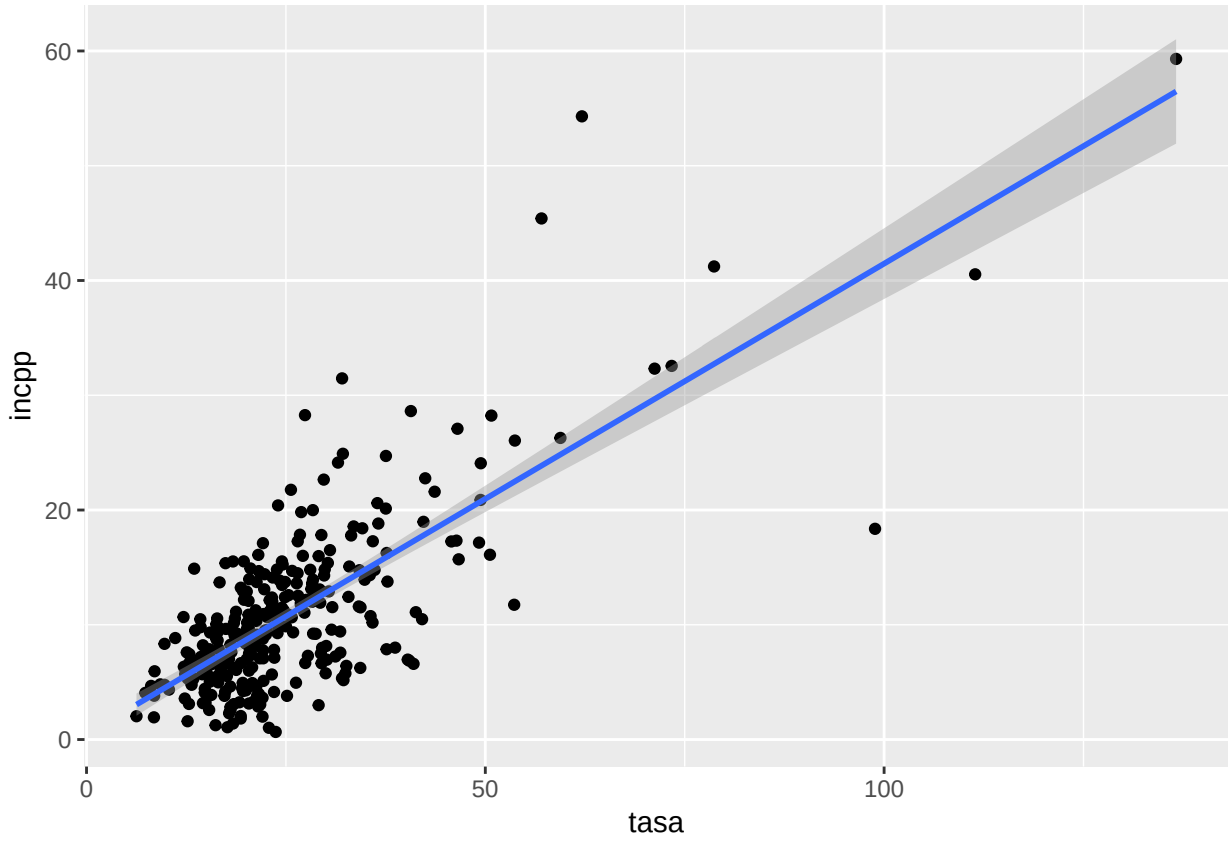
2.403

394.644***

Note:

p<0.1; ***p<0.05;*** p<0.01

Gráfica Modelo 2



Modelo 3

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

$$\text{Tasa por cada cien mil hab.}_i = \beta_1 \text{Ingreso}_i + \beta_2 \text{Porcentaje de gente blanca}_i + \epsilon$$

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: sáb, ene 19, 2019 - 13:00:54

Table 6:

	Dependent variable:		
	tasa		
	(1)	(2)	(3)
county_income	−0.0001		
	(0.0001)		
incpp		1.388***	1.388***
		(0.070)	(0.069)
pctwhite			−0.066***
			(0.020)
Constant	30.203***	10.315***	14.185***
	(3.160)	(0.934)	(1.470)
Observations	301	301	301
R^2	0.008	0.569	0.585
Adjusted R^2	0.005	0.568	0.582
Residual Std. Error	14.198 (df = 299)	9.359 (df = 299)	9.201 (df = 298)
F Statistic	2.403 (df = 1; 299)	394.644*** (df = 1; 299)	209.835*** (df = 2; 298)
Note:			*p<0.1; **p<0.05; ***p<0.01