

“R para Ciencia de Datos”

I. Explorar



R-Ladies La Paz

MSc. Ing. Ruth Chirinos

*Ciencia de Datos
(Data Science)...*

Casos confirmados con COVID-19 Global

Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20	2/1/20	2/2/20	2/3/20	2/4/20	2/5/20	2/6/20	2/7/20	2/8/20	2/9/20	2/10/20	2/11/20	2/12/20	2/13/20	2/14/20	2/15/20	2/16/20	2/17/20	2/18/20	
New Brunswick	Canada	46.5653	-66.4619	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Newfoundland and Labrador	Canada	53.1355	-57.6604	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Nova Scotia	Canada	44.682	-63.7443	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Ontario	Canada	51.2538	-85.3232	0	0	0	0	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
Prince Edward Island	Canada	46.5107	-63.4168	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Quebec	Canada	52.9399	-73.5491	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Saskatchewan	Canada	52.9399	-106.4509	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Central Africa	Chad	6.6111	20.9394	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	Chad	15.4542	18.7322	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	Chile	-35.6751	-71.543	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Anhui	China	31.8257	117.2264	1	9	15	39	60	70	106	152	200	237	297	340	408	480	530	591	665	733	779	830	860	889	910	934	950	962	973	98	
Beijing	China	40.1824	116.4142	14	22	36	41	68	80	91	111	114	139	168	191	212	228	253	274	297	315	326	337	342	352	366	372	375	380	381	38	
Chongqing	China	30.0572	107.874	6	9	27	57	75	110	132	147	182	211	247	300	337	366	389	411	426	428	468	486	505	518	529	537	544	551	553	55	
Fujian	China	26.0789	117.9874	1	5	10	18	35	59	80	84	101	120	144	159	179	194	205	215	224	239	250	261	267	272	279	281	285	287	290	29	
Gansu	China	37.8099	101.0583	0	2	2	4	7	14	19	24	26	29	40	51	55	57	62	62	67	79	83	83	86	87	90	90	90	90	91	9	
Guangdong	China	23.3417	113.4244	26	32	53	78	111	151	207	277	354	436	535	632	725	813	895	970	1034	1095	1131	1159	1177	1219	1241	1261	1294	1316	1322	132	
Guangxi	China	23.8298	108.7881	2	5	23	23	36	46	51	58	78	87	100	111	127	139	150	168	172	183	195	210	215	222	222	226	235	237	238	24	
Guizhou	China	26.8154	106.8748	1	3	3	4	5	7	9	12	29	38	46	58	64	71	81	89	99	109	127	133	135	140	143	144	146	144	14		
Hainan	China	19.1959	109.7453	4	5	8	19	22	33	40	43	46	52	62	64	72	80	99	106	117	124	131	138	144	157	157	159	162	162	163	16	
Hebei	China	39.549	116.1306	1	1	2	8	13	18	33	48	65	82	96	104	113	126	135	157	172	195	206	218	239	251	265	283	291	300	301	30	
Heilongjiang	China	47.862	127.7615	0	2	4	9	15	21	33	38	44	59	80	95	121	155	190	227	277	295	307	331	360	378	395	419	425	445	457	46	
Henan	China	33.882	113.614	5	5	9	32	83	128	168	206	278	352	422	493	566	675	764	851	914	981	1033	1073	1105	1135	1169	1184	1212	1231	1246	125	
Hong Kong	China	22.3	114.2	0	2	2	5	8	8	8	10	10	12	13	15	15	17	21	24	25	26	29	38	49	50	53	56	56	57	60	6	
Hubei	China	30.9756	112.2707	444	444	549	761	1058	1423	1554	3554	4903	5806	7153	11177	13522	16678	19665	22112	24953	27100	29631	31728	33366	33366	33366	48206	54406	56249	58182	59989	6168
Hunan	China	27.6104	111.7088	4	9	24	43	69	100	143	221	277	332	389	463	521	593	661	711	772	803	838	879	912	946	968	988	1001	1004	1006	100	
Inner Mongolia	China	44.0935	113.9448	0	0	1	7	7	11	15	16	19	20	23	27	34	35	42	46	50	52	54	58	58	60	61	65	68	70	72	7	
Jiangsu	China	32.9711	119.455	1	5	9	18	33	47	70	99	129	168	202	236	271	308	341	373	408	439	468	492	515	543	570	593	604	617	626	62	
Jiangxi	China	27.614	115.7221	2	7	18	18	36	72	109	109	162	240	286	333	391	476	548	600	661	698	740	771	804	844	872	900	913	925	930	93	
Jilin	China	43.6661	126.1923	0	1	3	4	4	6	8	9	14	17	23	31	42	54	59	65	69	78	80	81	83	84	86	88	89	89	89	8	
Liaoning	China	41.2956	122.6085	2	3	4	17	21	27	34	39	41	48	64	70	74	81	89	94	99	105	107	108	111	116	117	119	119	121	121	12	
Macau	China	22.1667	113.55	1	2	2	5	6	7	7	7	7	8	8	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	1	
Ningxia	China	37.2692	106.1655	1	1	2	3	4	7	11	12	17	21	26	28	31	34	34	40	43	45	45	49	53	58	64	67	70	70	70	7	
Qinghai	China	35.7452	95.9956	0	0	0	1	1	6	6	6	8	8	9	11	13	15	17	18	18	18	18	18	18	18	18	18	18	18	18	1	
Shaanxi	China	35.1917	108.8701	0	3	5	15	22	35	46	56	63	87	101	116	128	142	165	173	184	195	208	213	219	225	229	230	232	236	240	24	
Shandong	China	36.3427	118.1498	2	6	15	27	46	75	95	130	158	184	206	230	259	275	307	347	386	416	444	466	487	497	509	523	532	537	541	54	
Shanghai	China	31.202	121.4491	9	16	20	33	40	53	66	96	112	135	169	182	203	219	243	257	277	286	293	299	303	311	315	318	326	328	333	33	
Shanxi	China	37.5777	112.2922	1	1	1	6	9	13	27	27	35	39	47	66	74	81	81	96	104	115	119	119	124	126	126	127	128	129	130	13	
Sichuan	China	30.6171	102.7103	5	8	15	28	44	69	90	108	142	177	207	231	254	282	301	321	344	364	386	405	417	436	451	463	470	481	495	50	
Tianjin	China	39.3054	117.323	4	4	8	10	14	23	24	27	31	32	41	48	60	67	69	79	81	88	91	95	106	112	119	120	122	124	125	12	
Tibet	China	31.6927	88.0924	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Xinjiang	China	41.1129	85.2401	0	2	2	3	4	5	10	13	14	17	18	21	24	29	32	36	39	42	45	49	55	59	63	65	70	71	75	7	
Yunnan	China	24.974	101.487	1	2	5	11	16	26	44	55	70	83	93	105	117	122	128	133	138	138	141	149	153	154	156	162	168	171	171	17	
Zhejiang	China	29.1832	120.9934	10	27	43	62	104	128	173	296	428	538	599	661	724	829	895	954	1006	1048	1075	1092	1117	1131	1145	1155	1162	1167	1171	117	
South America	Colombia	4.5709	-74.2973	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	Congo (Brazzaville)	-4.0383	21.7587	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	







$$(f(x)) = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x}$$



$$a^2 + b^2 = c^2$$

$$\lim_{x \rightarrow 2} \frac{x^2 + 3x - 10}{x - 2}$$

$$\sqrt{3} =$$

$$\varphi^2 = 1 + \varphi$$

$$\frac{1}{\varphi} = \varphi - 1$$



$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, x \in \mathbb{C}$$

$$\sin A = \frac{a}{c}$$

$$\cos A = \frac{b}{c}$$

$$i = \sqrt{-1}$$

$$y = e^x$$

$$f(x) =$$



$$\sum$$



Total Confirmed

1.141.190

Confirmed Cases by
Country/Region/Sovereignty

278.942	US
124.736	Spain
119.827	Italy
92.150	Germany
83.031	France
82.543	China
55.743	Iran
42.441	United Kingdom
20.921	Turkey
20.278	Switzerland
18.431	Belgium
16.725	Netherlands
12.545	Canada
11.781	Austria
10.524	Portugal

Admin0

Admin1

Admin2

Last Updated at (M/D/YYYY)

4/4/2020 11:32:43 a. m.

181

countries/regions



Cumulative Confirmed Cases

Active Cases

Lancet Inf Dis Article: [Here](#). Mobile Version: [Here](#). Visualization: [JHU CSSE](#). Automation Support: [Esri Living Atlas team](#) and [JHU APL](#). [Contact US](#). [FAQ](#).
Data sources: WHO, CDC, ECDC, NHC, DXY, 1point3acres, Worldometers.info, BNO, state and national government health departments, and local media reports. Read more in this [blog](#).

Total Deaths

60.960

14.681 deaths	Italy
11.744 deaths	Spain
6.507 deaths	France
4.313 deaths	United Kingdom
3.452 deaths	Iran
3.207 deaths	Hubei China
1.867 deaths	New York City New York US
1.651 deaths	Netherlands

Total Recovered

233.930

76.942 recovered	China
34.219 recovered	Spain
24.575 recovered	Germany
19.758 recovered	Italy
19.736 recovered	Iran
14.138 recovered	France
9.920 recovered	US
6.325 recovered	Korea, South



Confirmed

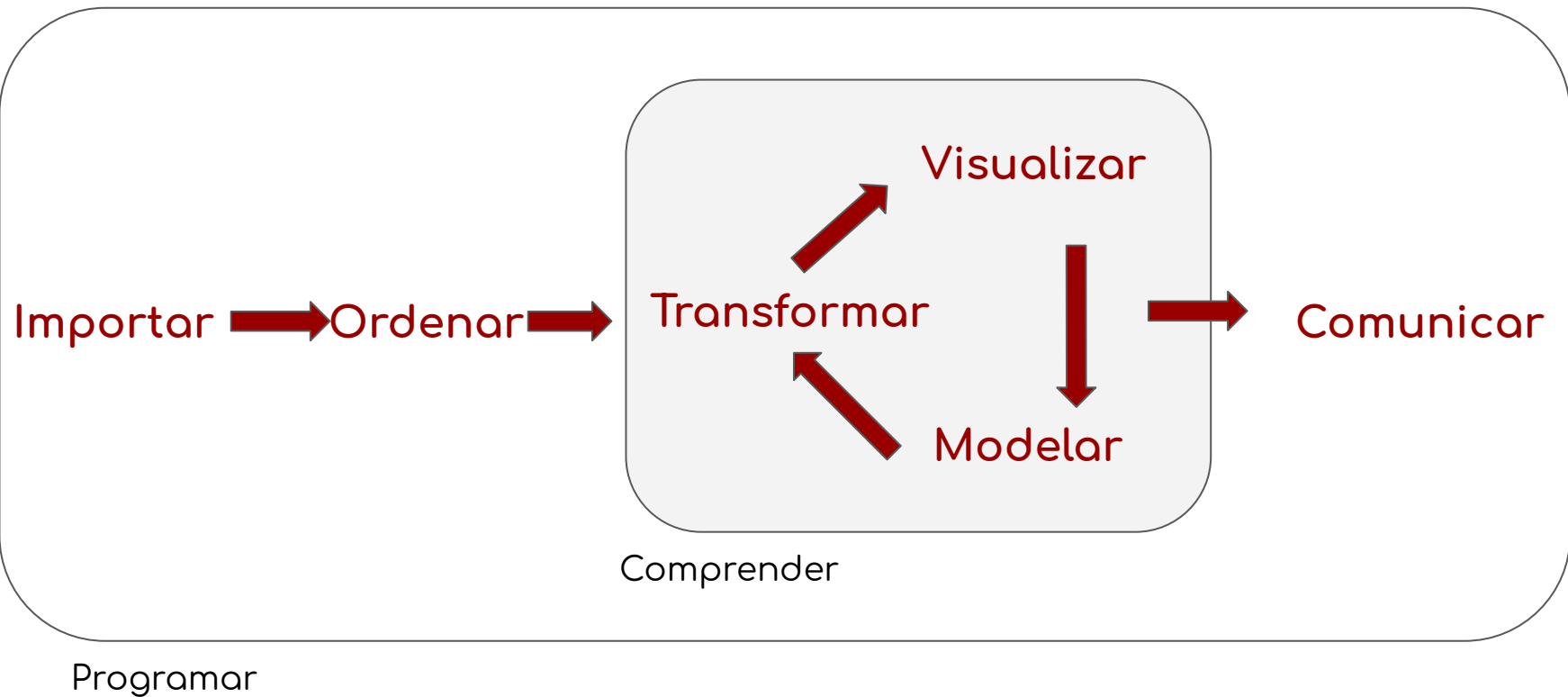
Logarithmic

Daily Increase

Lo que aprenderás...

Tools





Import



Tidy



Transform



Visualise



Model



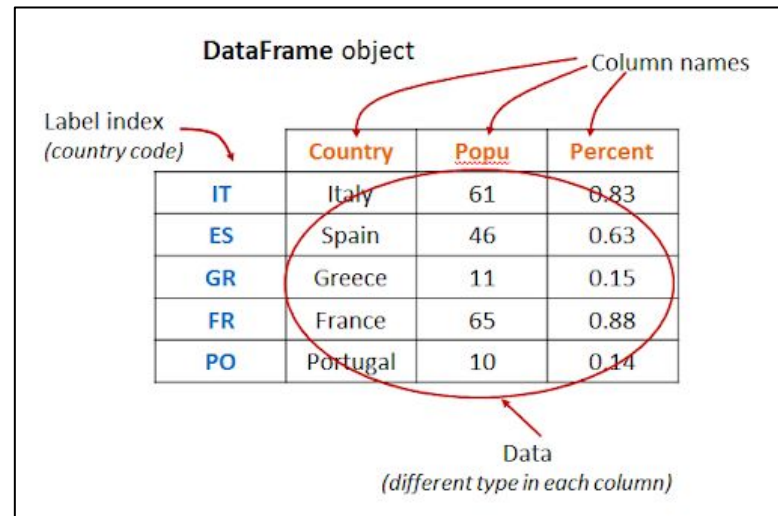
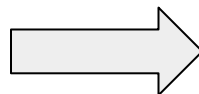
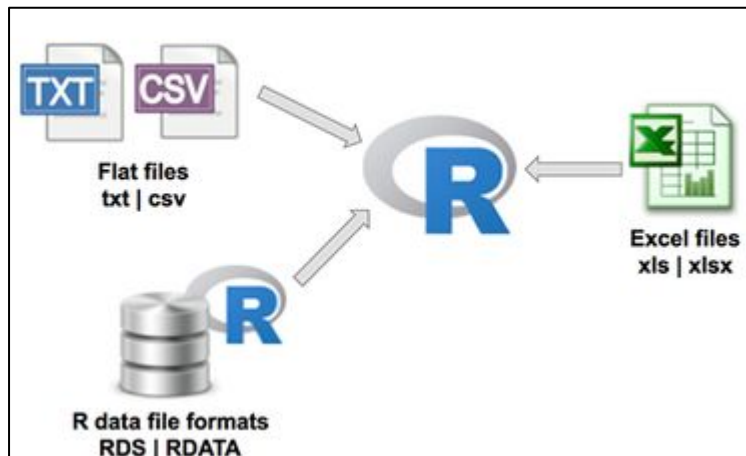
Communicate



We Can Do It!



Importar Datos



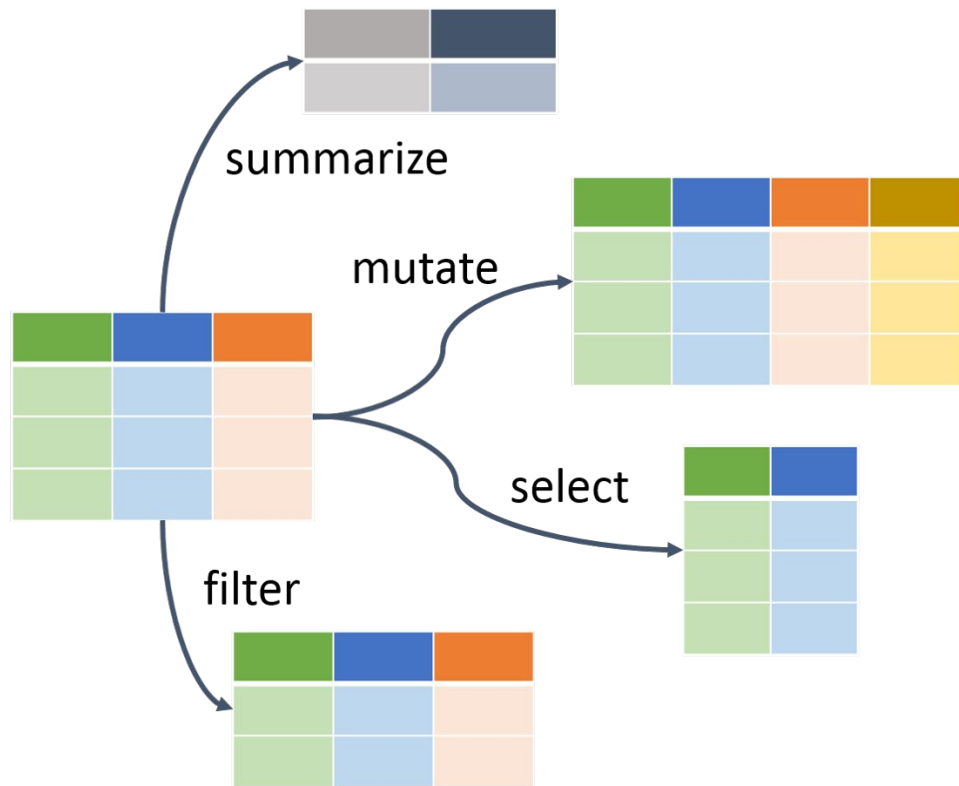
Ordenar

The diagram illustrates the structure of a data table. It features a table with four columns and four rows. Above the table, the text 'Variables (Columnas)' is centered, with four downward-pointing arrows indicating the column headers: 'Pais', 'Censo', 'Población urbana', and 'Población total'. To the left of the table, the text 'Observaciones (Filas)' is oriented vertically, with four rightward-pointing arrows indicating the row data. The table contains the following data:

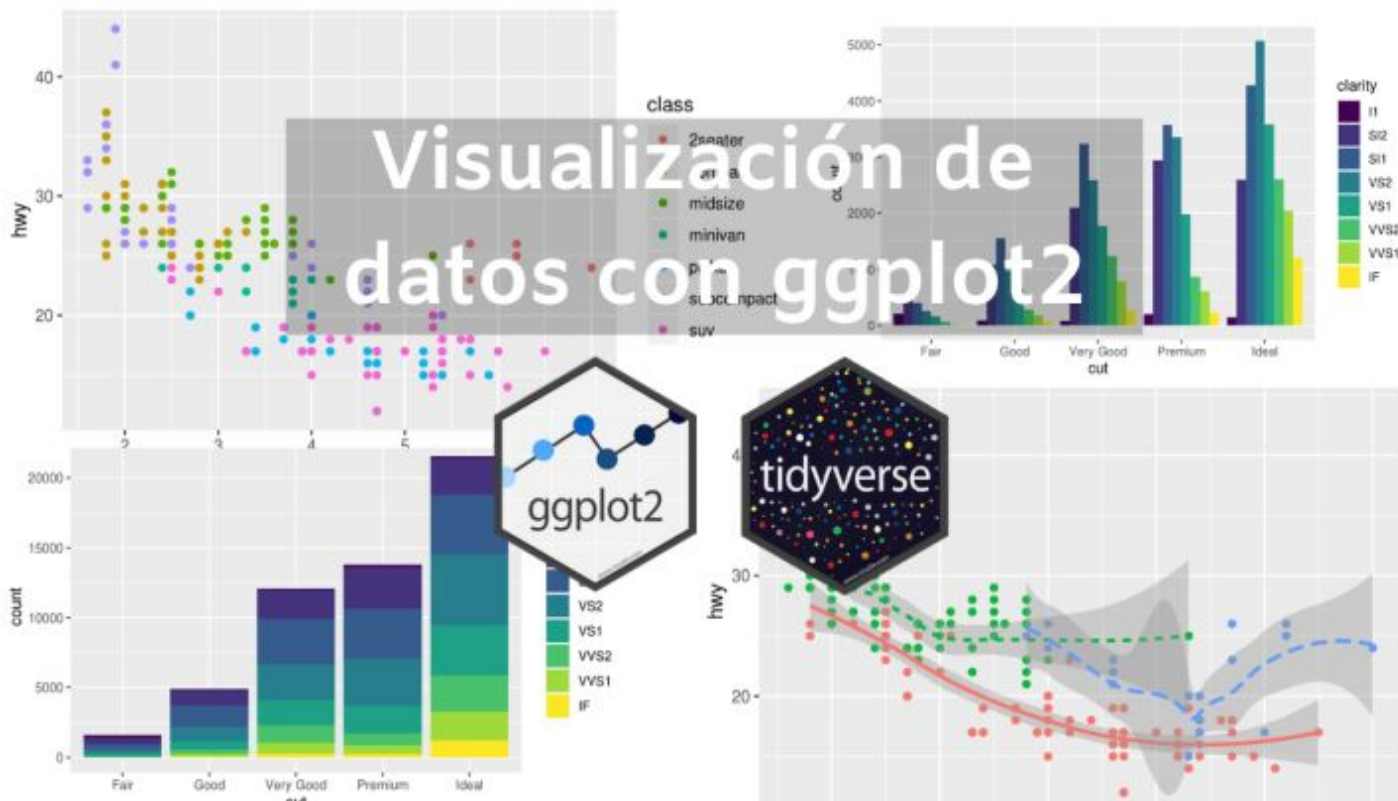
	Pais	Censo	Población urbana	Población total
→	Argentina	1980	23.198.068	27.949.480
→	Argentina	1991	28.832.127	32.615.528
→	Argentina	2001	32.380.296	36.260.130
→	Argentina	2010	36.907.728	40.117.096

The value '32.380.296' in the third row and third column is circled. An arrow points from the text 'Valores (celdas)' to this circled value.

Transformación



Visualización



Modelar

Cost function

Logistic regression:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

Neural network:

$\rightarrow h_{\Theta}(x) \in \mathbb{R}^K$ $(h_{\Theta}(x))_i = i^{th}$ output

$$\rightarrow J(\Theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(h_{\Theta}(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_{\Theta}(x^{(i)}))_k) \right]$$

$$+ \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (\Theta_{ji}^{(l)})^2$$

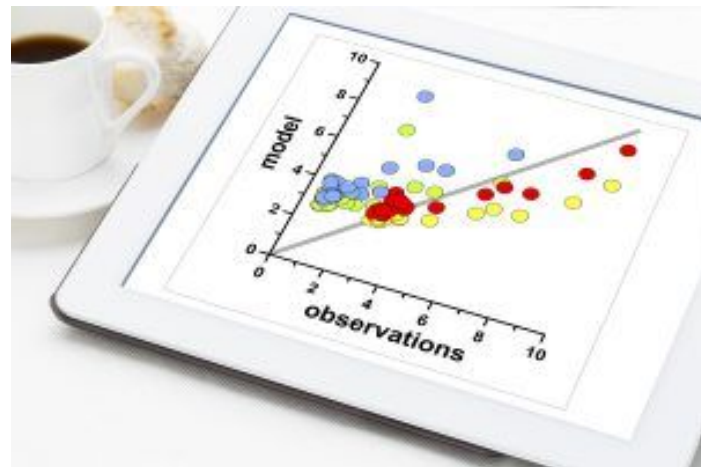
$\Theta_{ji}^{(l)} x_0 + \Theta_{ji}^{(l)} x_1 + \dots$

$\Theta_{ji}^{(l)} x_0 + \Theta_{ji}^{(l)} x_1 + \dots$

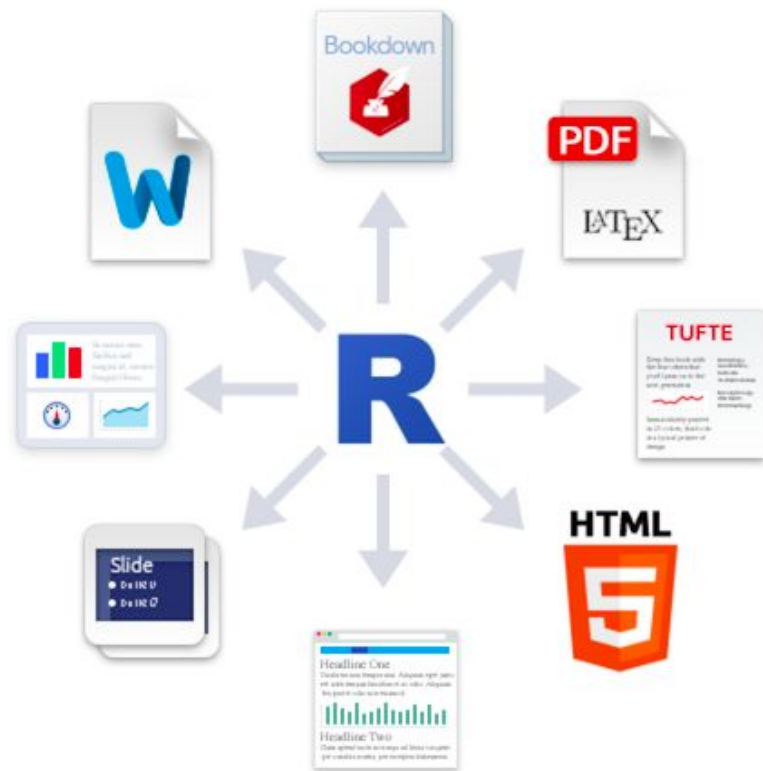
y_k

$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

Andrew Ng



Comunicar

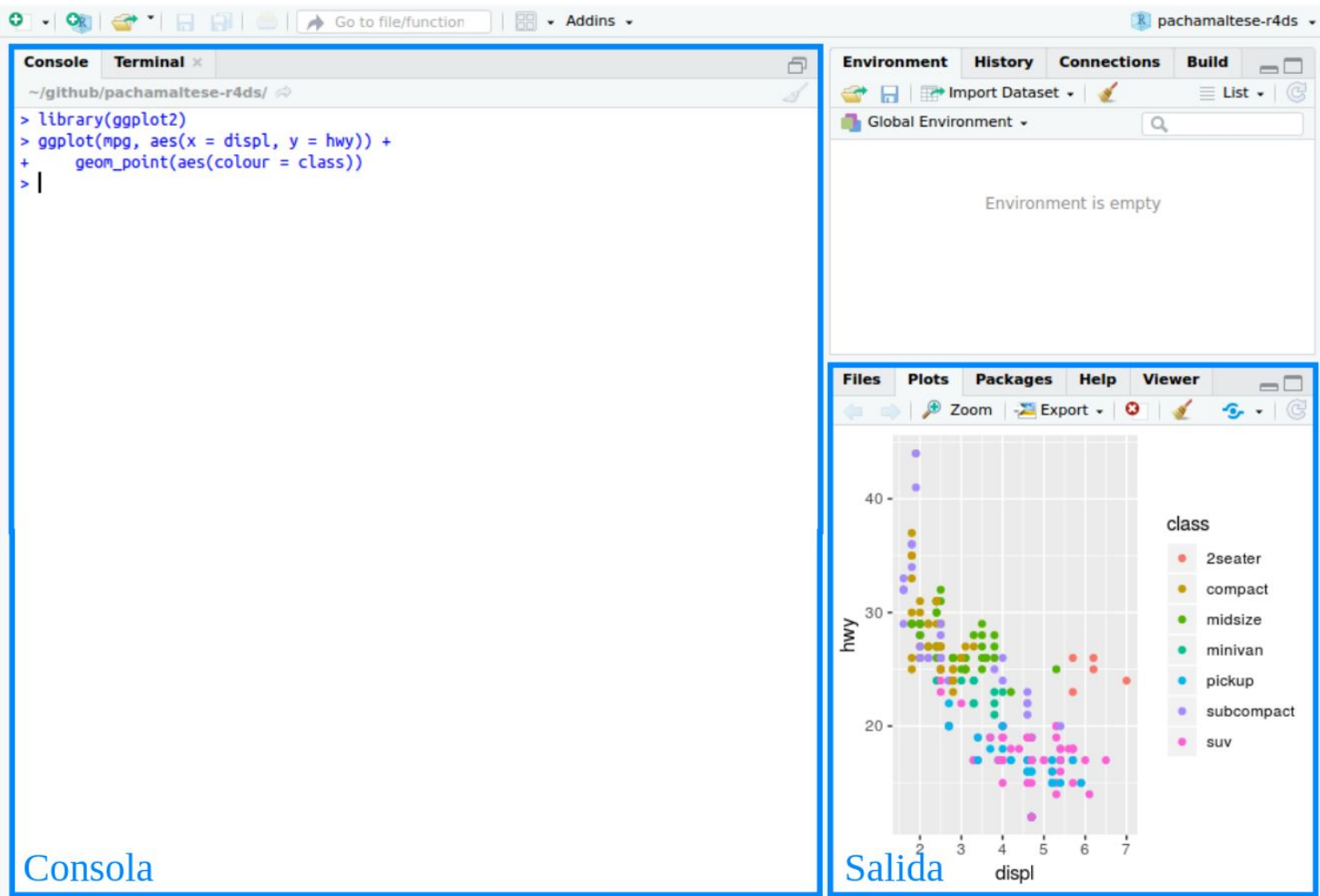


Pre Requisitos

- R-base (Última versión version)
 - Windows, Linux, (Mac) OS X:
 - <https://cloud.r-project.org/>
 - Other links for Linux:
 - <https://linuxize.com/post/how-to-install-r-on-ubuntu-18-04/>
- R-Studio (Minimo 1.0.0)
 - <http://www.rstudio.com/download>

Pre Requisitos

- Pre requisitos:
 - Leer la página Web de Introducción del libro:
<https://es.r4ds.hadley.nz/introducci%C3%B3n.html>
 - Instalar R-base
 - Instalar R-Studio
 - Instalar Tidyverse
 - Instalar paquetes de datos
 - Instalar otros paquetes
 - [Todo lo que esté en la página Web de Introducción...](#)



Laboratorio