

*Data Journalism*  
Jornalismo Científico usando  
Bases de Dados

Jorge Meléndez

Divulgação em Astronomia – AGA421

# Data Journalism

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style

News > Media > Digital media

## Analysing data is the future for journalists, says Tim Berners-Lee

Inventor of the world wide web says reporters should be hunting for stories in datasets



**Charles Arthur**

Follow @charlesarthur Follow @guardiantech

The Guardian, Monday 22 November 2010

# *Data Journalism: começou aprox. em 2009*

- Muitas bases de dados disponíveis ao público
- Internet: acesso fácil às bases de dados
- Ferramentas simples (e gratuitas) para visualização de dados

# *Data Journalism: começou apenas em 2009*

- Possibilidade de gerar a sua própria notícia!
- Precisa conhecimentos básicos de estatística, visualização e interpretação de dados
- Os jornalistas não recebem necessariamente esse treinamento

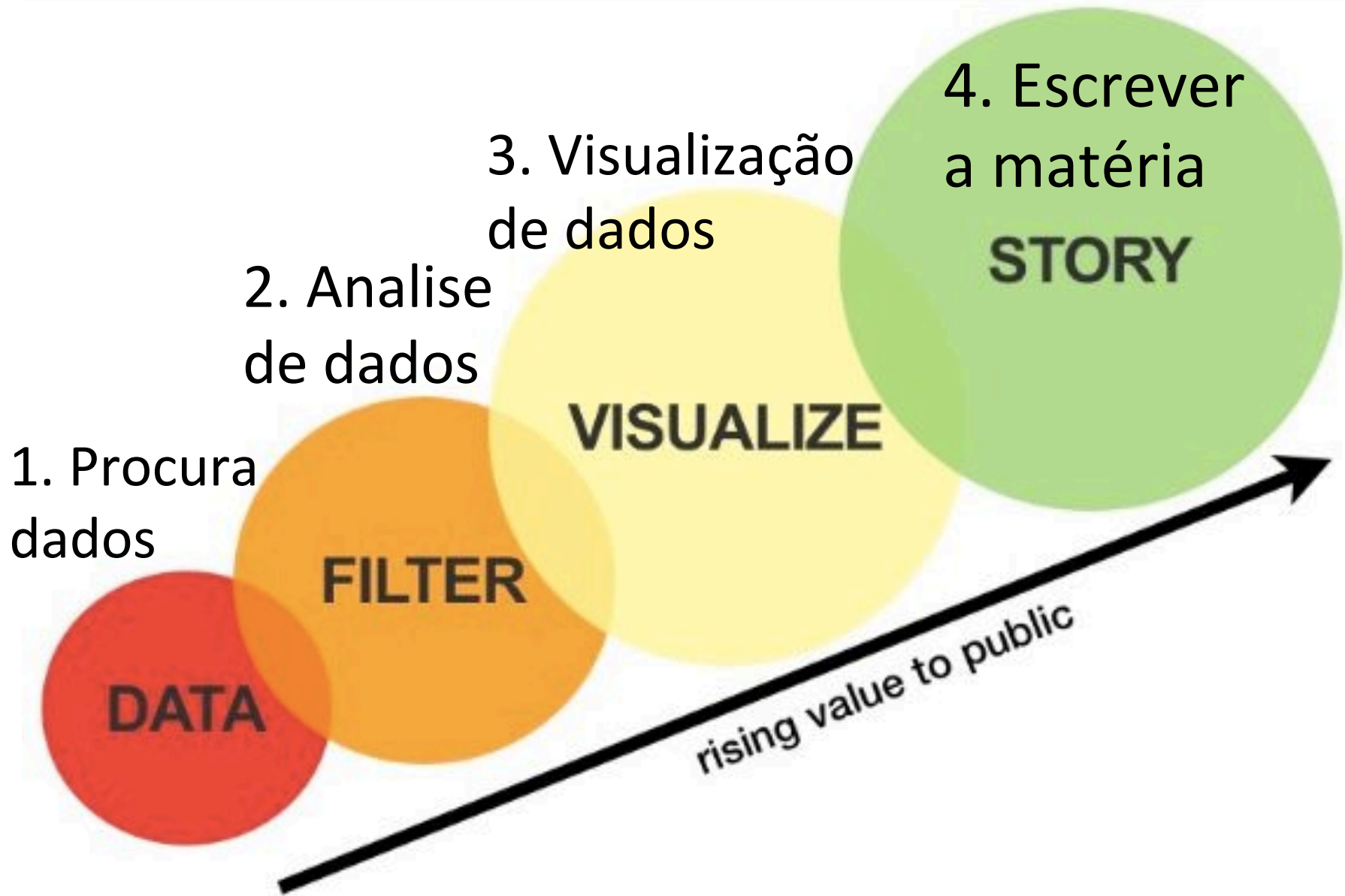
# Data Journalism: ótimo para cientistas!

Na graduação, experiência com relatórios de laboratório ou iniciação científica:

- Uso de dados
- Estatística simples
- Visualização dos dados
- Escrever relatório

	A	B	C	D	E	F	G	H	
1	Product	(All)							
2									
3	Actual Sales	Column Labels							
4	Row Labels		Jul-10	Aug-10	Sep-10	Oct-10	Nov-10	Dec-10	Grand Total
5	Alabama		4350	5142	5538	5306	5846	4916	31098
6	Alaska		4518	6186	4784	4542	6386	3110	29526
7	Arizona		6328	6556	4408	6032	6910	5198	35432
8	Idaho		6476	5492	6002	5988	7582	5280	36820
9	Illinois		4756	6356	4926	6336	4090	4956	31420
10	Indiana		5790	6534	4552	6194	5010	5032	33112
11	Ohio		3976	5792	8056	4076	5388	5566	32854
12	Oklahoma		4384	5874	7334	4514	6182	4246	32534
13	Oregon		5894	5288	6904	4910	5388	6408	34792
14	Vermont		5776	5328	5512	5248	5346	4244	31454
15	Virginia		5110	5914	3434	4686	5002	7664	31810
16	Washington		7318	5694	3672	7060	5594	4558	33896
17	Grand Total		64676	70156	65122	64892	68724	61178	394748





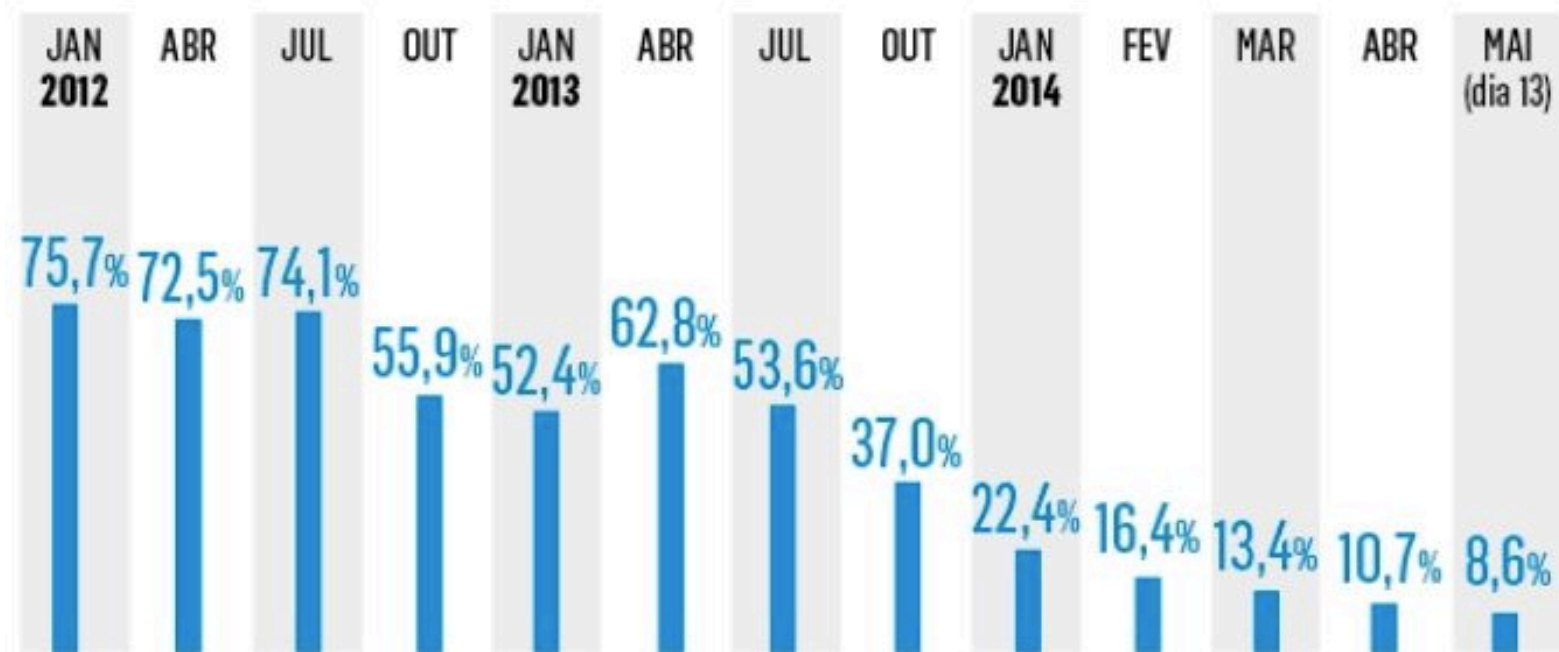
Manchete 1:

Nível do sistema Cantareira é de 86  
bilhões de litros de água

## Manchete 2:

# Nível do sistema Cantareira é de apenas 8% da sua capacidade

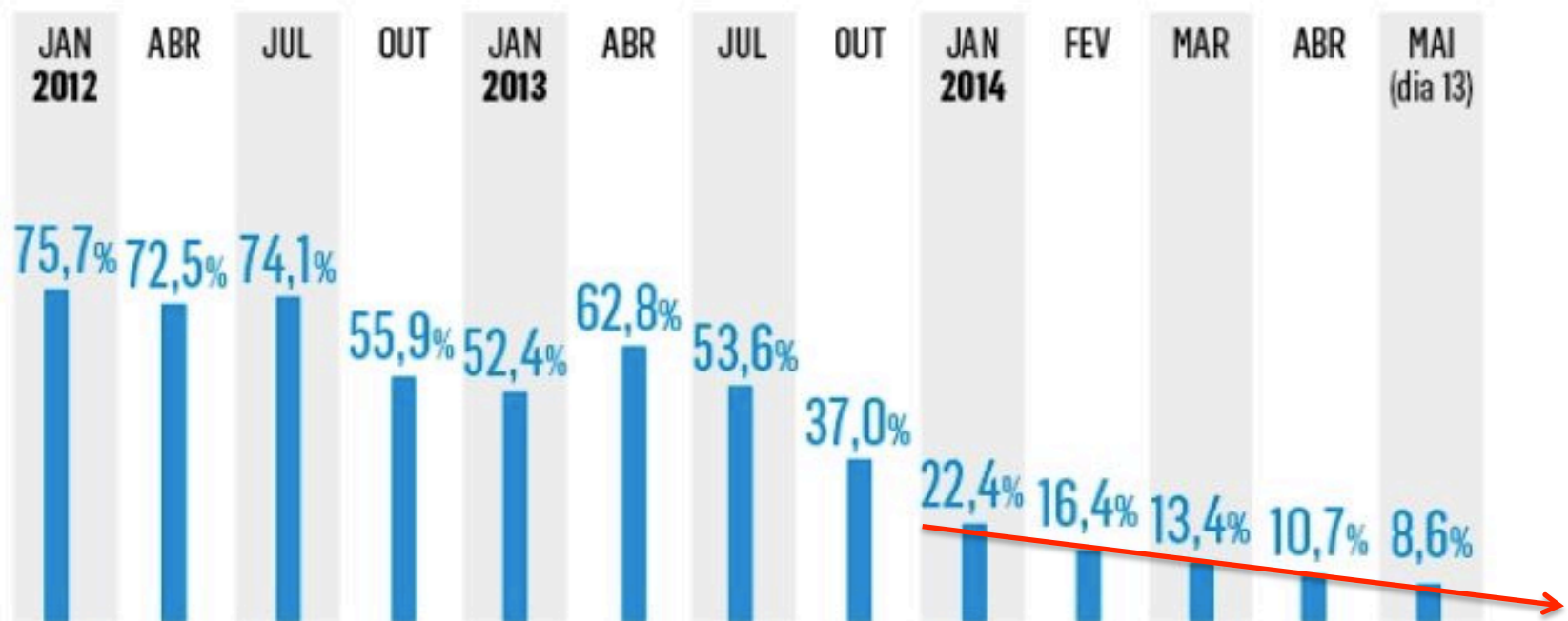
### NÍVEL DO RESERVATÓRIO





Manchete 3:

# Nível do Cantareira cai e volume útil de água pode acabar até julho

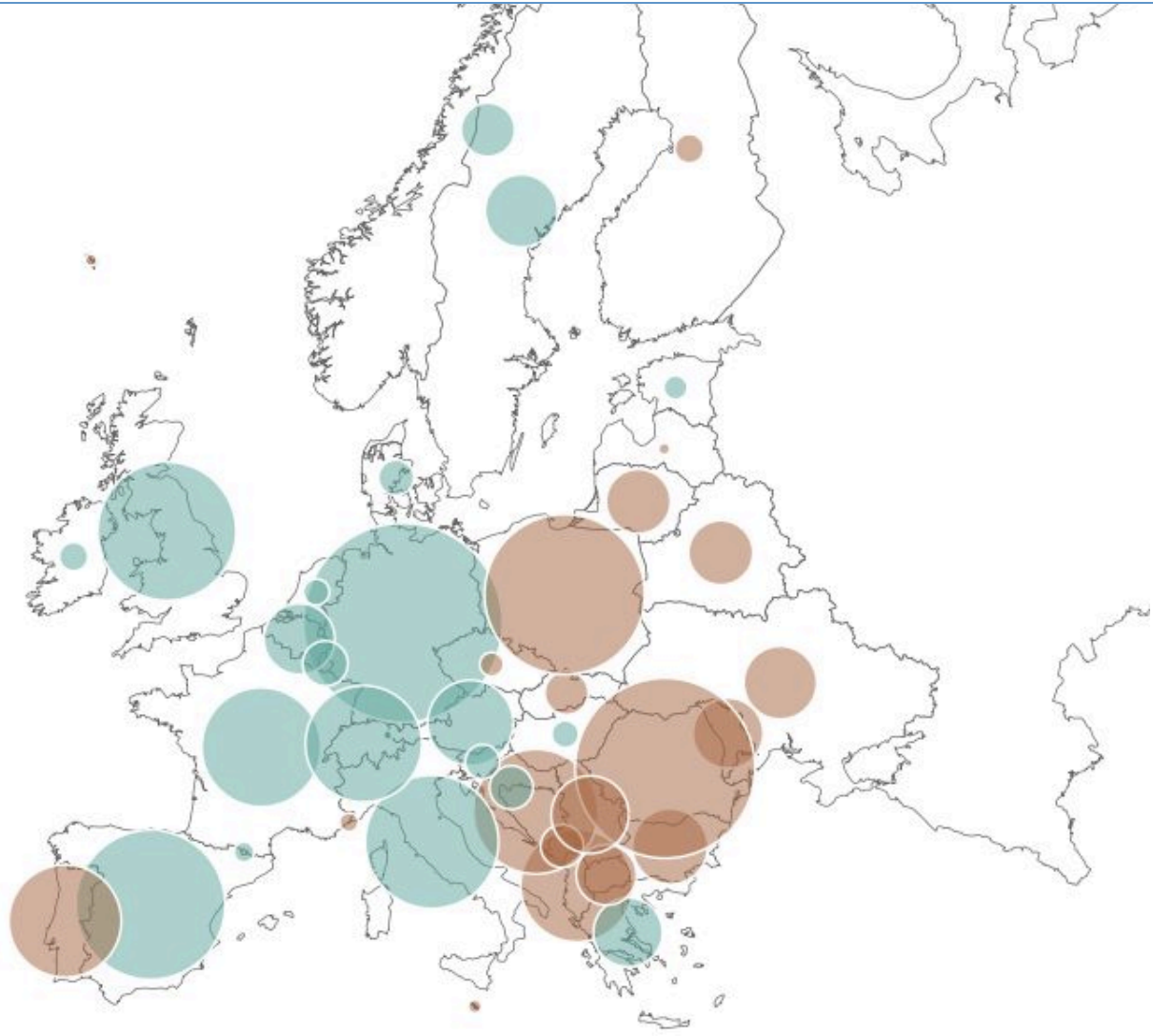


INFOGRÁFICO/ESTADÃO

<http://sao-paulo.estadao.com.br/noticias/geral,nivel-do-cantareira-cai-e-volume-util-de-agua-pode-acabar-ate-julho,1166037>

# Imigração e Emigração na Europa em 2013

● saldo migratório / fluxo migratório positivo    ● saldo migratório / fluxo migratório negativo



# Case study:

## Residentes (legais) sul-americanos na Espanha

### Extranjeros con certificado de registro o tarjeta de residencia en vigor



Esta estadística presenta datos referidos a los extranjeros residentes en España que disponen de certificado de registro o tarjeta de residencia en vigor con fecha de referencia el último día de cada trimestre. Se trata de una operación estadística incluida en el Plan Estadístico Nacional, que elabora el Observatorio Permanente de la Inmigración a partir de ficheros procedentes del Registro Central de Extranjeros, que es gestionado por la Dirección General de la Policía.

- > **Principales resultados:** Se trata de un informe con periodicidad trimestral que explora las principales características de los extranjeros residentes. Por un lado se ofrece un análisis comentado de una selección de tablas y gráficos representativa de la situación general, y por otro se incluye un anexo estadístico de tablas más completo.
- > **Resultados detallados:** Este apartado ofrece una relación exhaustiva de tablas con periodicidad trimestral en las que se describen todas las variables relativas a las autorizaciones de residencia, clasificadas por la nacionalidad y el sexo del extranjero, la comunidad autónoma y la provincia de residencia.

### Principales Resultados


Último dato publicado: 31 diciembre 2013


Fecha de publicación: 06-08-2014



[Principales resultados](#) 



[Anexo tablas \(31-12-2013\)](#) 

[Anexo tablas \(31-12-2013\)](#) 

[Anexo tablas \(30-09-2013\)](#) 



TABLA 2

EXTRANJEROS CON TARJETA O PERMISO DE RESIDENCIA EN VIGOR SEGÚN CONTINENTE Y NACIONALIDAD POR RÉGIMEN DE RESIDENCIA  
31-12-2003 / 31-12-2002

CONTINENTE / NACIONALIDAD	31-12-2003			31-12-2002			VARIACIÓN 2002 - 2003					
	TOTAL	RÉGIMEN DE RESIDENCIA		TOTAL	RÉGIMEN DE RESIDENCIA		TOTAL	RÉGIMEN DE RESIDENCIA		TOTAL	RÉGIMEN DE RESIDENCIA	
		Régimen General	Régimen Comunitario		Régimen General	Régimen Comunitario		Régimen General	Régimen Comunitario		Régimen General	Régimen Comunitario
<b>TOTAL</b>	<b>1,647,011</b>	<b>1,074,895</b>	<b>572,116</b>	<b>1,324,001</b>	<b>826,956</b>	<b>497,045</b>	<b>323,010</b>	<b>247,939</b>	<b>75,071</b>	<b>24.40</b>	<b>29.98</b>	<b>15.10</b>
<b>ESPACIO EC.EUROPEO</b>	<b>406,199</b>	-	<b>406,199</b>	<b>362,858</b>	<b>25</b>	<b>362,833</b>	<b>43,341</b>	-	<b>43,366</b>	<b>11.94</b>	-	<b>11.95</b>
Alemania	67,963	-	67,963	65,823	-	65,823	2,140	-	2,140	3.25	-	3.25
Austria	4,172	-	4,172	3,931	-	3,931	241	-	241	6.13	-	6.13
Bélgica	15,736	-	15,736	14,631	-	14,631	1,105	-	1,105	7.55	-	7.55
Dinamarca	6,568	-	6,568	6,167	-	6,167	401	-	401	6.50	-	6.50
Finlandia	5,906	-	5,906	5,672	-	5,672	234	-	234	4.13	-	4.13
Francia	49,196	-	49,196	46,986	-	46,986	2,210	-	2,210	4.70	-	4.70
Grecia	1,367	-	1,367	1,183	-	1,183	184	-	184	15.55	-	15.55
Irlanda	4,882	-	4,882	4,208	-	4,208	674	-	674	16.02	-	16.02
Italia	59,745	-	59,745	45,236	-	45,236	14,509	-	14,509	32.07	-	32.07
Noruega	8,049	-	8,049	6,717	20	6,697	1,332	-	1,352	19.83	-	20.19
Países Bajos	20,551	-	20,551	18,722	-	18,722	1,829	-	1,829	9.77	-	9.77
Portugal	45,614	-	45,614	43,309	-	43,309	2,305	-	2,305	5.32	-	5.32
Reino Unido	105,479	-	105,479	90,091	-	90,091	15,388	-	15,388	17.08	-	17.08
Suecia	10,415	-	10,415	9,652	-	9,652	763	-	763	7.91	-	7.91
Otros Esp. Económico Europeo	556	-	556	530	5	525	26	-	31	4.91	-	5.90
<b>IBEROAMÉRICA</b>	<b>514,485</b>	<b>409,982</b>	<b>104,503</b>	<b>364,569</b>	<b>282,301</b>	<b>82,268</b>	<b>149,916</b>	<b>127,681</b>	<b>22,235</b>	<b>41.12</b>	<b>45.23</b>	<b>27.03</b>
Argentina	43,347	22,353	20,994	27,937	14,606	13,331	15,410	7,747	7,663	55.16	53.04	57.48
Bolivia	7,053	6,452	601	4,995	4,529	466	2,058	1,923	135	41.20	42.46	28.97
Brasil	14,598	7,868	6,730	12,902	7,024	5,878	1,696	844	852	13.15	12.02	14.49
Colombia	107,459	90,481	16,978	71,238	59,292	11,946	36,221	31,189	5,032	50.85	52.60	42.12
Costa Rica	405	204	201	370	200	170	35	4	31	9.46	2.00	18.24
Cuba	27,323	11,611	15,712	24,226	9,640	14,586	3,097	1,971	1,126	12.78	20.45	7.72
Chile	10,869	8,316	2,553	8,257	5,936	2,321	2,612	2,380	232	31.63	40.09	10.00
Ecuador	174,289	170,652	3,637	115,301	112,802	2,499	58,988	57,850	1,138	51.16	51.28	45.54
El Salvador	1,217	871	346	1,110	786	324	107	85	22	9.64	10.81	6.79
Guatemala	631	387	244	608	370	238	23	17	6	3.78	4.59	2.52
Honduras	1,834	1,360	474	1,649	1,204	445	185	156	29	11.22	12.96	6.52
Méjico	6,714	3,192	3,522	5,894	2,910	2,984	820	282	538	13.91	9.69	18.03
Nicaragua	671	393	278	574	334	240	97	59	38	16.90	17.66	15.83
Panamá	500	230	270	421	200	221	79	30	49	18.76	15.00	22.17
Paraguay	992	597	395	758	465	293	234	132	102	30.87	28.39	34.81
Perú	57,593	51,094	6,499	39,013	33,289	5,724	18,580	17,805	775	47.63	53.49	13.54
Rep. Dominicana	36,654	24,345	12,309	32,412	21,423	10,989	4,242	2,922	1,320	13.09	13.64	12.01
Uruguay	8,852	4,423	4,429	5,995	3,232	2,763	2,857	1,191	1,666	47.66	36.85	60.30
Venezuela	13,162	4,984	8,178	10,634	3,925	6,709	2,528	1,059	1,469	23.77	26.98	21.90
Otros Iberoamérica	322	169	153	275	134	141	47	35	12	17.09	26.12	8.51

# Case study:

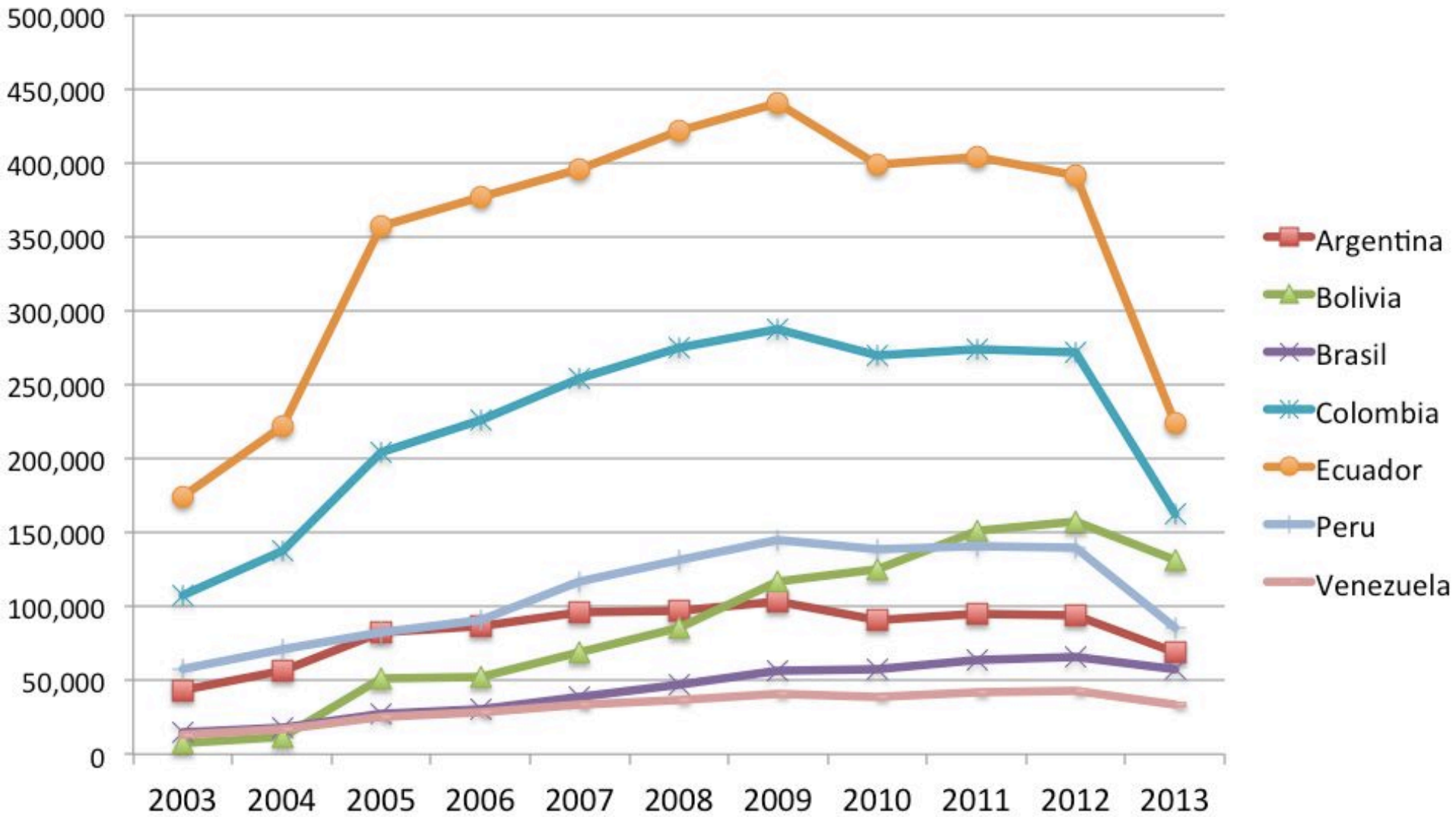
## Residentes (legais) sul-americanos na Espanha

	A	B	C	D	E	F	G	H
1	Ano	Argentina	Bolivia	Brasil	Colombia	Ecuador	Peru	Venezuela
2	2003	43,347	7,053	14,598	107,459	174,289	57,593	13,162
3	2004	56,193	11,467	17,524	137,369	221,549	71,245	16,622
4	2005	82,412	50,738	26,866	204,348	357,065	82,533	25,372
5	2006	86,921	52,587	30,242	225,504	376,233	90,906	28,188
6	2007	96,055	69,109	39,170	254,301	395,808	116,202	33,262
7	2008	97,277	85,427	47,229	274,832	421,527	130,900	36,616
8	2009	103,171	117,106	56,153	287,205	440,304	144,620	40,370
9	2010	91,056	124,695	57,459	269,687	398,724	138,211	38,775
10	2011	94,581	150,702	63,689	274,171	403,864	140,792	42,301
11	2012	93,750	157,132	65,303	271,596	391,231	139,825	43,234
12	2013	68,596	131,312	56,937	162,374	223,570	85,735	33,418



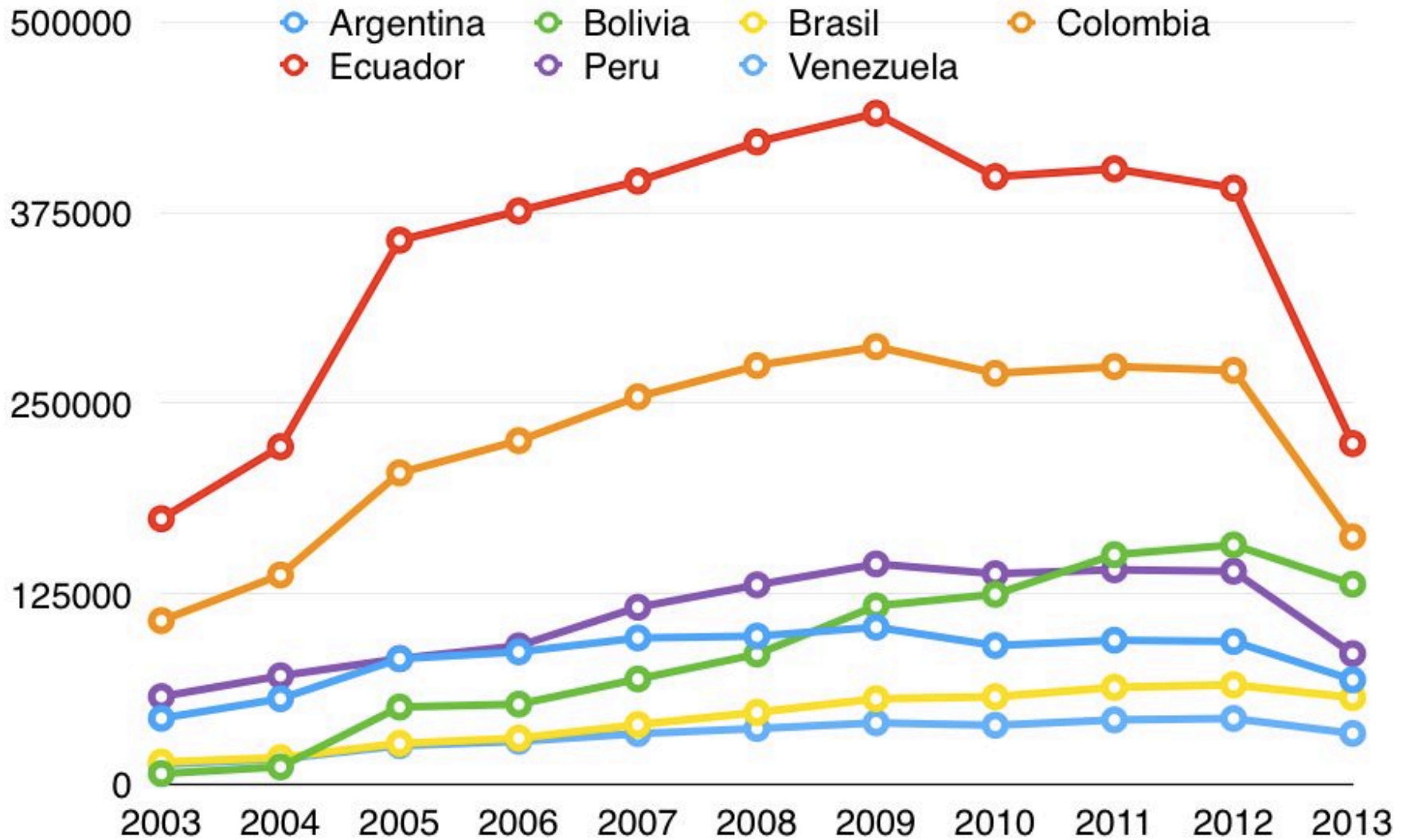
# Case study:

## Residentes (legais) sul-americanos na Espanha



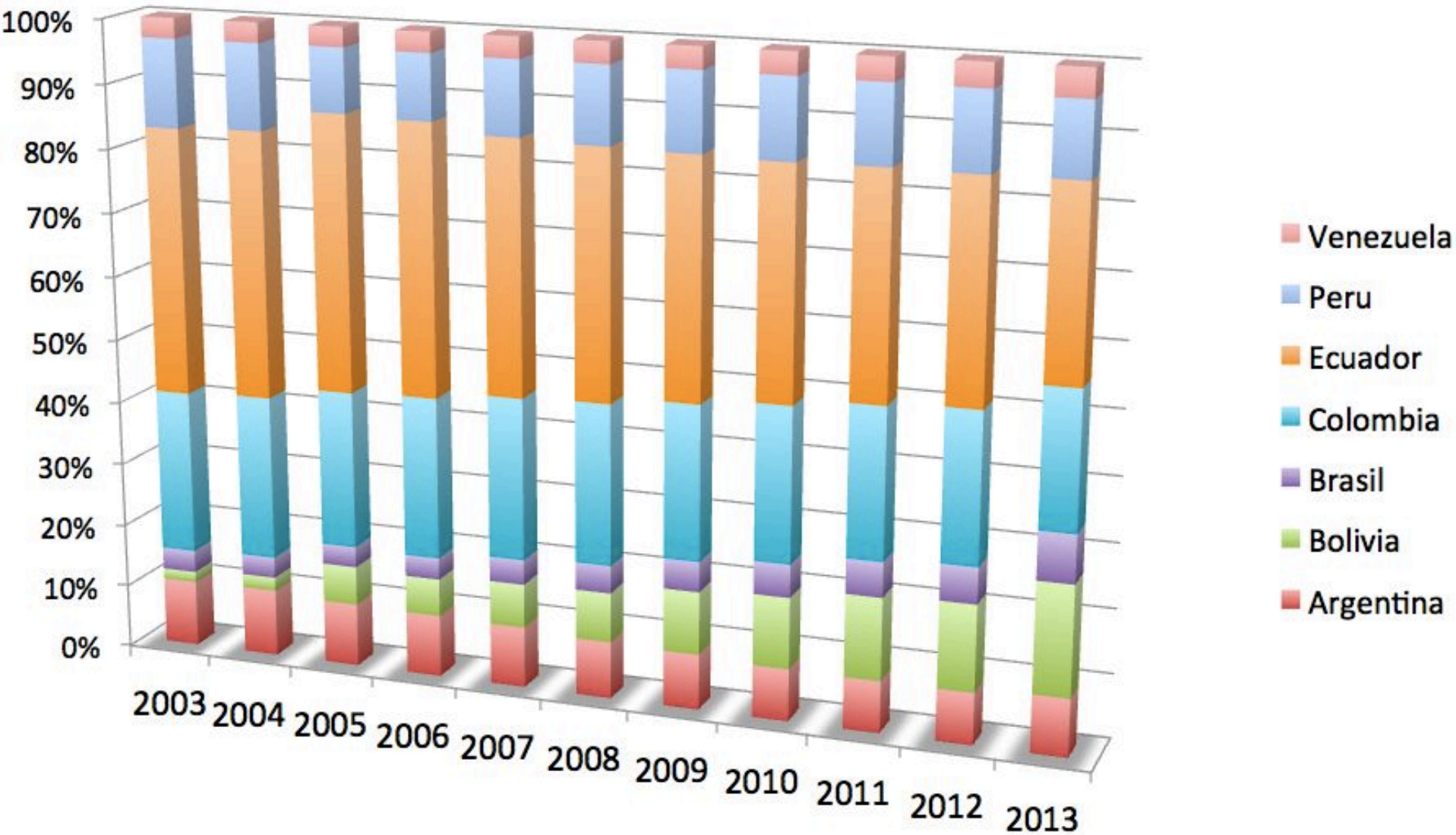
# Case study:

## Residentes (legais) sul-americanos na Espanha



# Case study:

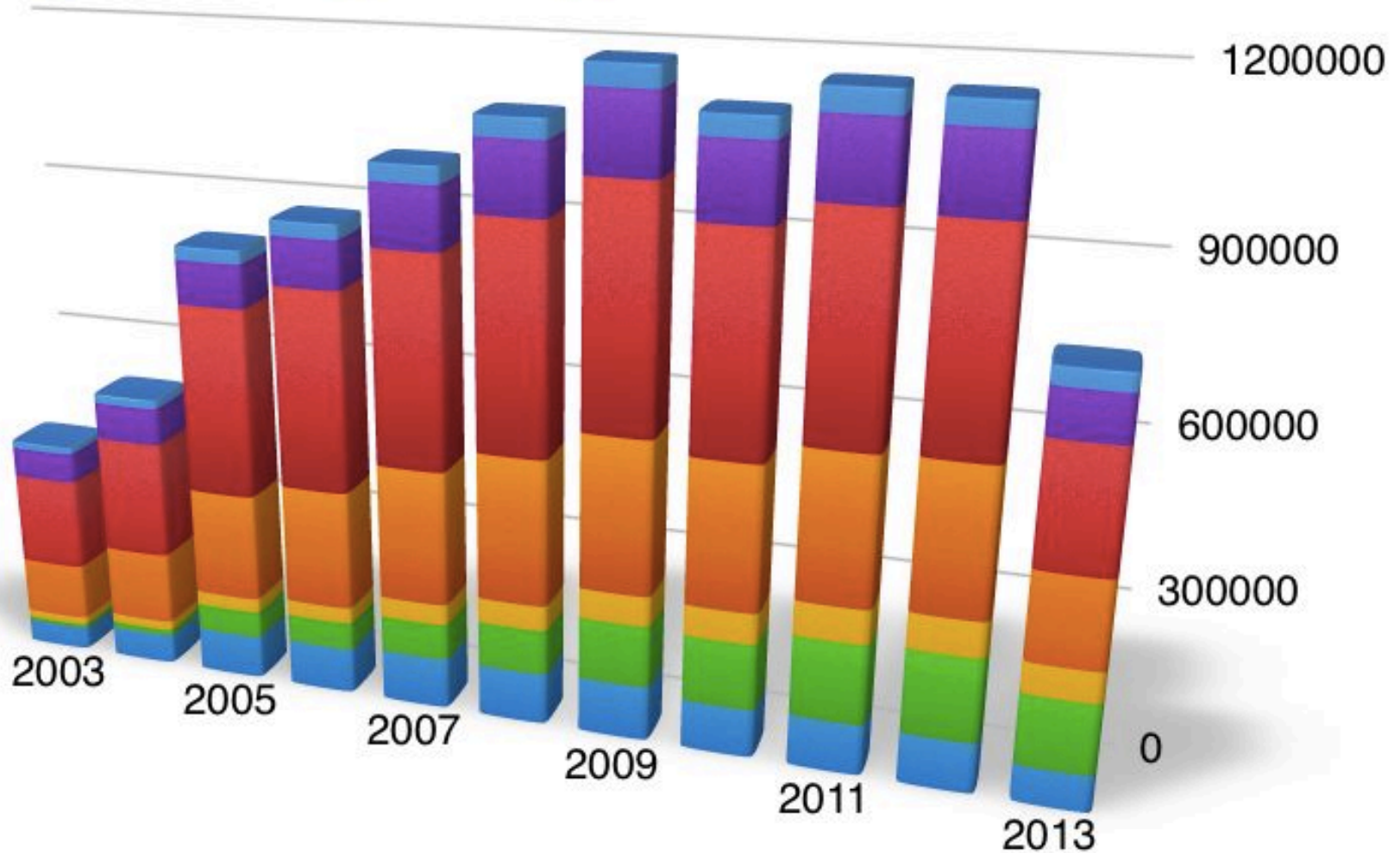
## Residentes (legais) sul-americanos na Espanha





# Case study:

## Residentes (legais) sul-americanos na Espanha



# Ferramentas online

<http://www.chartgo.com/>



Create Graphs Online

English Français Español

## Chart Settings

Chart Type

Line

Show 3D



Width

500

Height

400

Threshold Line

Chart Title

Bolivia

Sub title

X title

Ano

Y title

lentes na Espanha

Source

Title Font

Bold

Label Font

Normal

X Labels

Horizontal

Color

Blue

Y Highest

Transparent



Display Values



Legend



Color Plot



Y Lowest

Gradient Color



Round Corners



Shadow



Border



Gridlines



## Chart Data

Rules

Example

Random Data

X Axis Data

2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013

Y Axis Data

7,053  
11,467  
50,738  
52,587  
69,109  
85,427  
117,106  
124,695  
150,702  
157,132  
131,312

Group



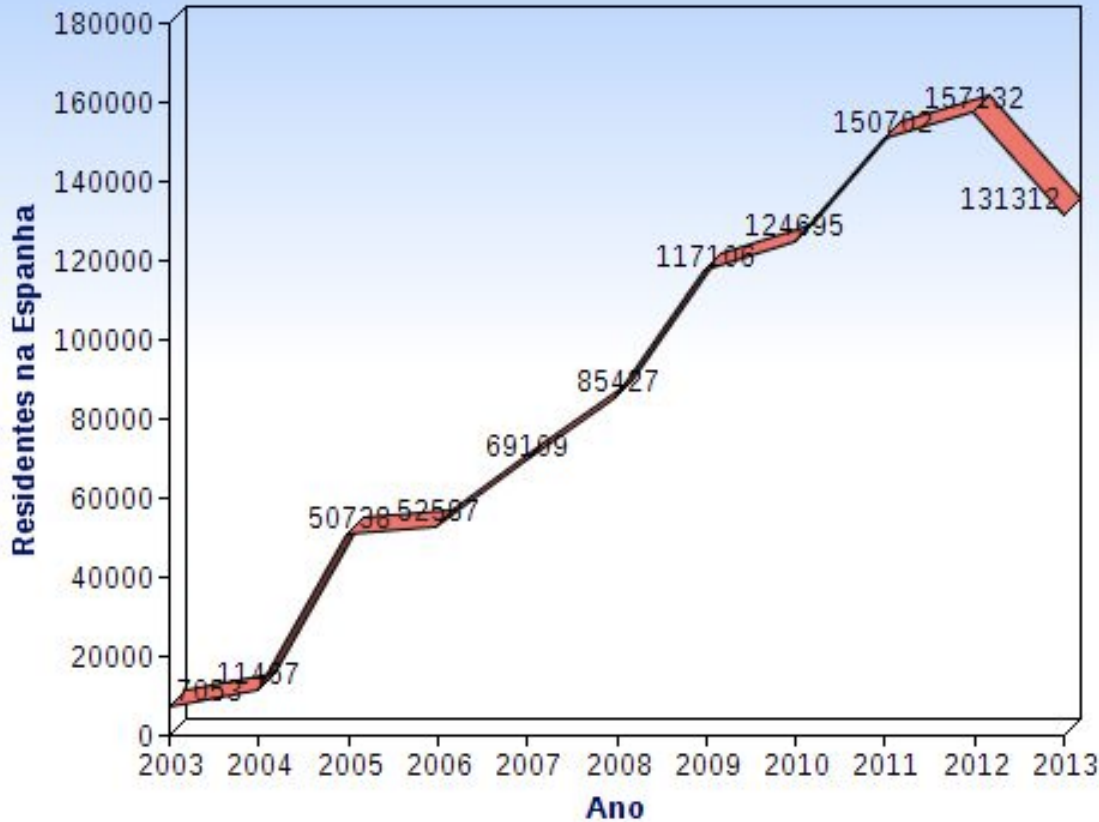
1

Create Chart

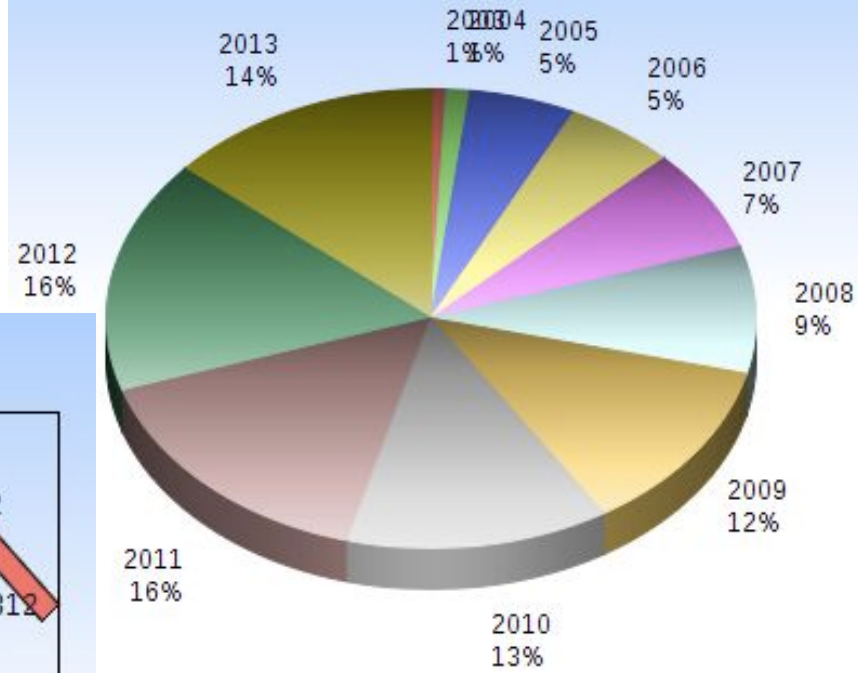
# Ferramentas online

<http://www.chartgo.com/>

### Bolivia



### Bolivia





# Ferramentas online

<http://mkweb.bcgsc.ca/tableviewer/>

Copy/paste do excel direto para um arquivo de texto:

Ano	Argentina	Bolivia	Brasil	Colombia	Ecuador	Peru	Venezuela
2003	43,347	7,053	14,598	107,459	174,289	57,593	13,162
2004	56,193	11,467	17,524	137,369	221,549	71,245	16,622
2005	82,412	50,738	26,866	204,348	357,065	82,533	25,372
2006	86,921	52,587	30,242	225,504	376,233	90,906	28,188
2007	96,055	69,109	39,170	254,301	395,808	116,202	33,262
2008	97,277	85,427	47,229	274,832	421,527	130,900	36,616
2009	103,171	117,106	56,153	287,205	440,304	144,620	40,370
2010	91,056	124,695	57,459	269,687	398,724	138,211	38,775
2011	94,581	150,702	63,689	274,171	... TABLE DATA USED FOR FIGURE		
2012	93,750	157,132	65,303	271,596			
2013	68,596	131,312	56,937	162,374			

Reconheceu corretamente  
o arquivo ascii

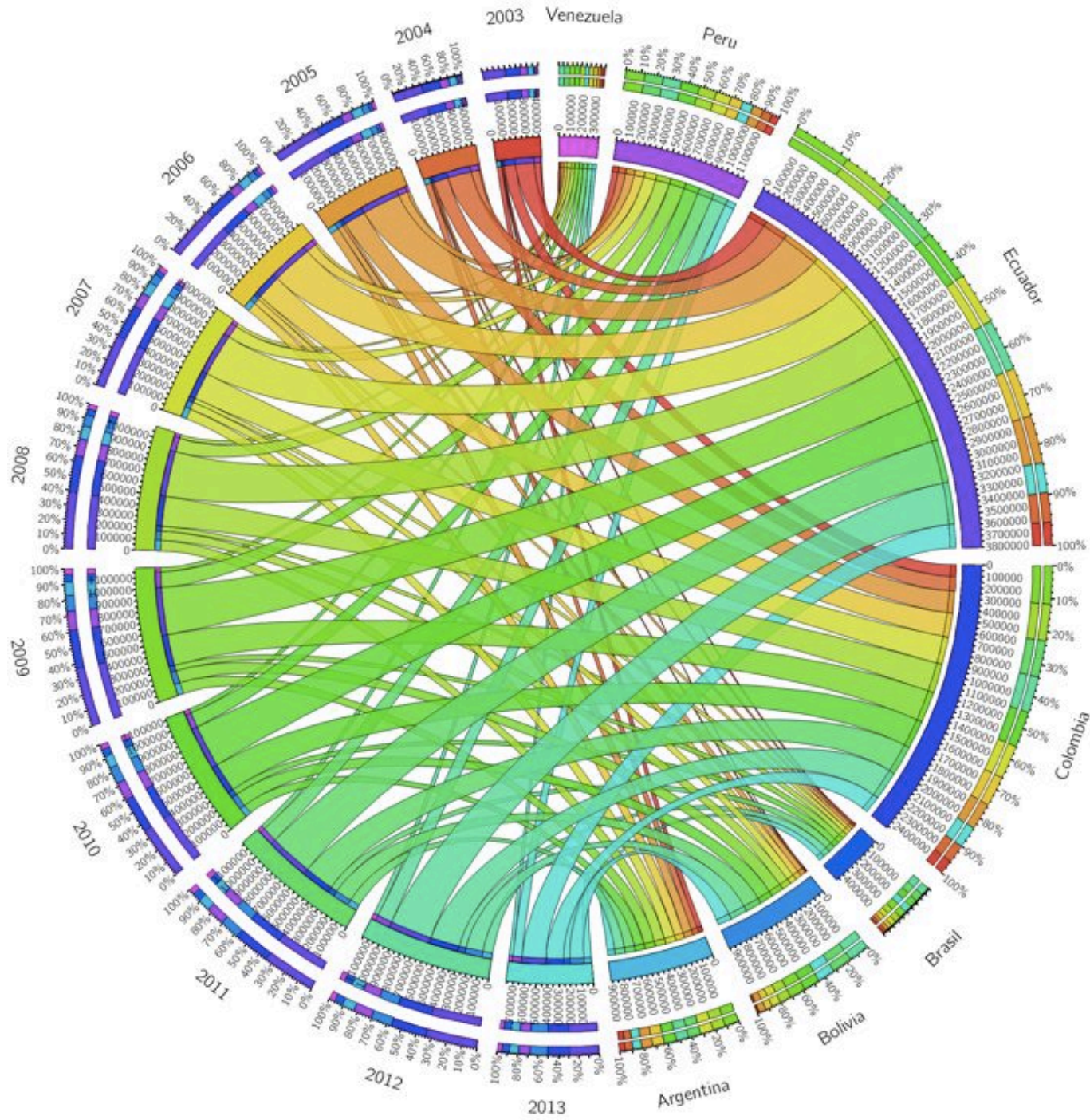


ANO	Argentina	Bolivia	Brasil	Colombia	Ecuador	PeRU	Venezuela
2003	43,347	7,053	14,598	107,459	174,289	57,593	13,162
2004	56,193	11,467	17,524	137,369	221,549	71,245	16,622
2005	82,412	50,738	26,866	204,348	357,065	82,533	25,372
2006	86,921	52,587	30,242	225,504	376,233	90,906	28,188
2007	96,055	69,109	39,170	254,301	395,808	116,202	33,262
2008	97,277	85,427	47,229	274,832	421,527	130,900	36,616
2009	103,171	117,106	56,153	287,205	440,304	144,620	40,370
2010	91,056	124,695	57,459	269,687	398,724	138,211	38,775
2011	94,581	150,702	63,689	274,171	403,864	140,792	42,301
2012	93,750	157,132	65,303	271,596	391,231	139,825	43,234
2013	68,596	131,312	56,937	162,374	223,570	85,735	33,418



CIRCOS

<http://mkweb.bcgsc.ca/tableviewer/>



# Ferramentas online



<http://filtergraph.vanderbilt.edu/tools/default/index>

You can't do this in Excel.

Create stunning interactive data visualization portals.

Try out the sample portal below, visualizing over 100,000 earthquakes worldwide.

## Earthquakes

Earthquake data from Jan 1, 2008 through December 31, 2012.

Source: <http://earthquake.usgs.gov/earthqua...>

[Click here to filter the data...](#)

Scatter Plot Histogram Table

X-axis: Longitude

Y-axis: Latitude

Color: Depth

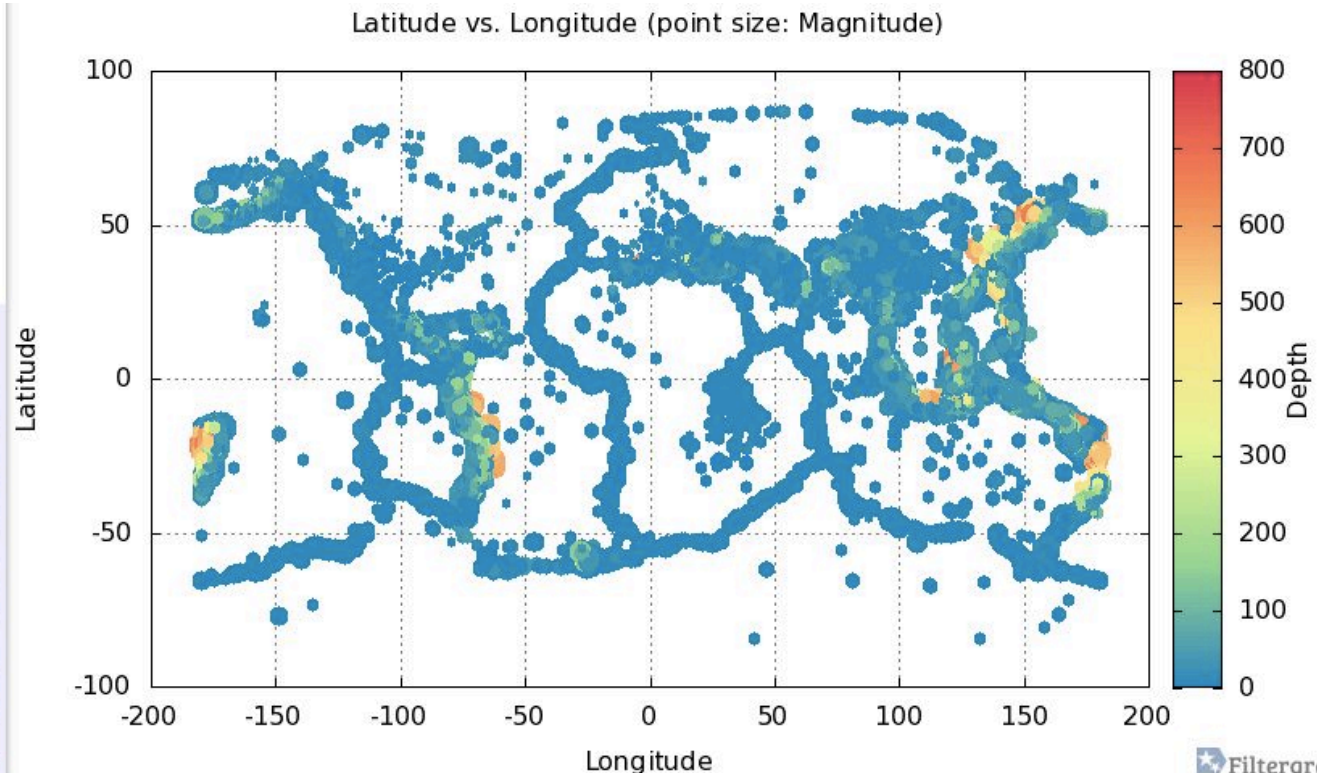
0.01 <= c <=

Size: Magnitude

3 <= s <=

Z-axis:

[more...](#)





<http://www.bizjournals.com/nashville/blog/2014/09/from-astrophysics-to-data-journalism-what-two.html?page=all>

# From astrophysics to data journalism: What two Vanderbilt grads want to do with data visualization

Sep 11, 2014, 2:01pm CDT



★ Save    📄 Order Reprints    🖨️ Print



**Eleanor Kennedy**  
Staff Reporter-  
*Nashville Business Journal*  
Email | Twitter

About 300 people in 30 countries are currently using Filtergraph, a free Web-based data visualization tool developed in a master's thesis project by Vanderbilt graduate [Dan Burger](#). Most of those users are academics, studying fields ranging from soil analysis to astrophysics.

But now Burger and [Rachel-Chloe Gibbs](#), a 2014 Vanderbilt alum, are looking to turn the tool from a free service owned by the university into a functioning business, one with potential for spinout through the [Vanderbilt Center for Technology Transfer](#)



About 300 people in 30 countries are currently using **Filtergraph**, a free Web-based data visualization tool developed by Vanderbilt graduate Dan Burger.

- Dados podem ser filtrados

Residentes\_sul-americanos\_em\_Espanh

✎

Set initial settings

Edit portal Edit data file

variable  
is between min max

variable  
is between min max

[How does this work?](#)

Scatter Plot Histogram Table

Sort by: Ano

Order:  Reverse

Rows: 100

Show rows: (all | none)

Ano  Argentina  Bolivia  Brasil  
 Colombia  Ecuador  Peru  
 Venezuela

Ano	Argentina	Bolivia	Brasil	Colombia	Ecuador	Peru	Venezuela
2003	43347	7053	14598	107459	174289	57593	13162
2004	56193	11467	17524	137369	221549	71245	16622
2005	82412	50738	26866	204348	357065	82533	25372
2006	86921	52587	30242	225504	376233	90906	28188
2007	96055	69109	39170	254301	395808	116202	33262
2008	97277	85427	47229	274832	421527	130900	36616
2009	103171	117106	56153	287205	440304	144620	40370
2010	91056	124695	57459	269687	398724	138211	38775
2011	94581	150702	63689	274171	403864	140792	42301
2012	93750	157132	65303	271596	391231	139825	43234
2013	68596	131312	56937	162374	223570	85735	33418

- Vantagens: muito fácil de aprender. Pode ser usada a cor para maior informação. Também opção 3D
- Desvantagem: scatter plot de apenas uma variável Y?



Exemplos de

<http://www.InformationIsBeautiful.net>

American Novels Everyone Should Read

A consensus cloud from top book lists & prizes

UNCLE TOM'S CABIN As I Lay Dying UNDERWORLD  
ANGLE OF REPOSE FAHRENHEIT 451 THE PORTRAIT OF A LADY  
On the Road A Confederacy of DUNCES FOR WHOM THE BELL TOLLS  
THE CALL OF THE WILD HOUSE KEEPING INDEPENDENCE DAY <sup>20</sup>Awakening  
Moby-Dick RABBIT AT REST BLOOD MERIDIAN Lolita  
WHITE NOISE the CATCHER in the RYE  
The Color Purple TO KILL A MOCKINGBIRD BELOVED  
CATCH-22 THE GRAPES of WRATH The GREAT GATSBY  
ADVENTURES OF HUCKLEBERRY FINN  
THE SUN ALSO RISES The House of Mirth AMERICAN PASTORAL  
The Known World INVISIBLE MAN THEIR EYES WERE WATCHING GOD  
The AMAZING ADVENTURES of HAVALIER & CLAY  
Middlesex

Exemplos de

<http://www.InformationIsBeautiful.net>

## Non-fiction Books Everyone Should Read







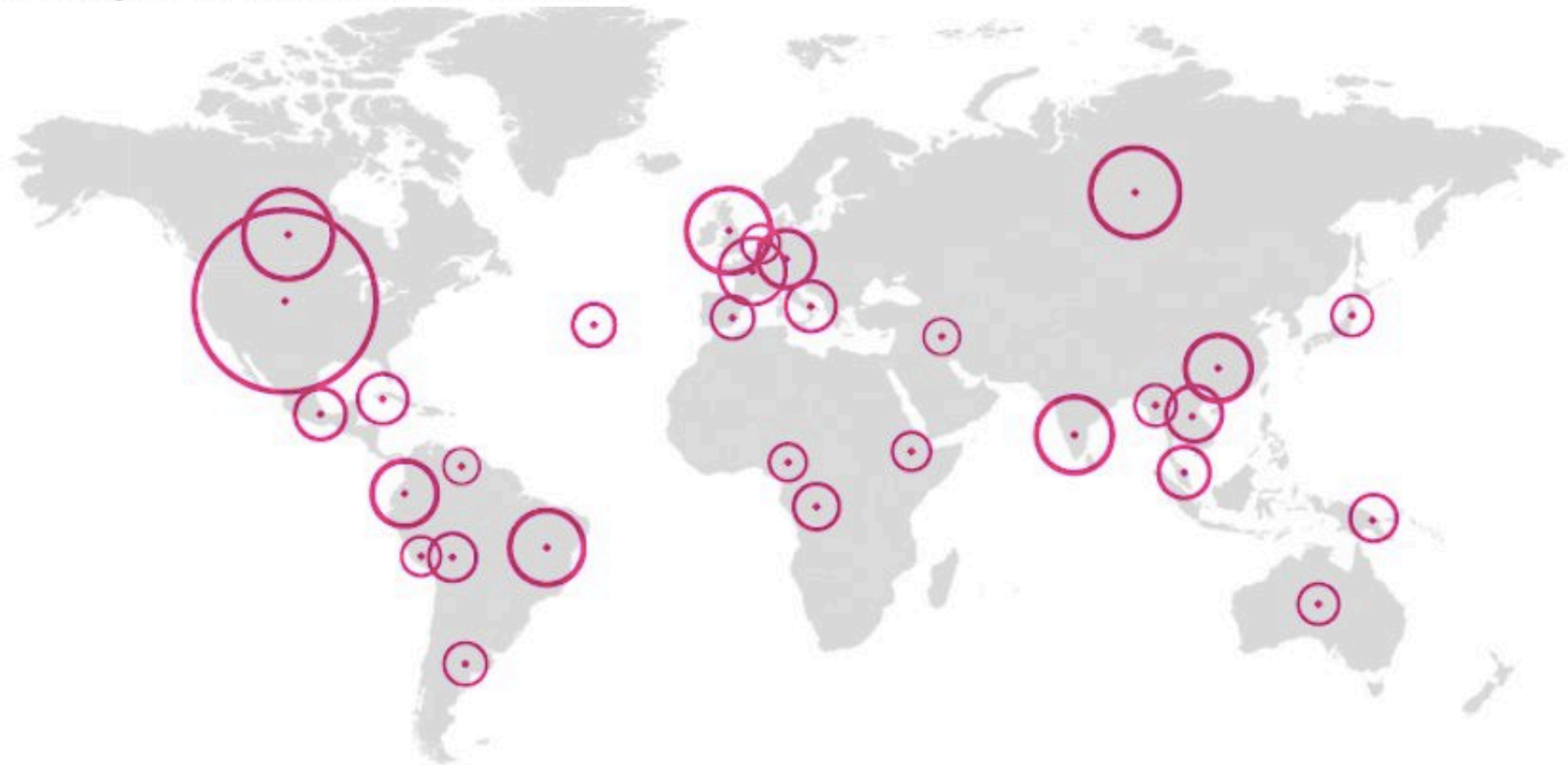
<http://www.InformationIsBeautiful.net>

# Reduce Your Odds Of Dying In A Plane Crash

A pre-flight check...

## Final Destination

Density of fatal accidents 1942-2009



<http://www.InformationIsBeautiful.net>

# Reduce Your Odds Of Dying In A Plane Crash

A pre-flight check...

## Safety Record

Fatal accidents by airline (7 or more) 1950-2008

**AIR FRANCE**

**KOREAN  
AIR**

**PIA**

Pakistan International

**AIR INDIA**

**CUBANA**

**American Airlines®**

**TURKISH  
AIRLINES**

**AEROFLOT**



**UNITED**

*Garuda Indonesia*

*Air Philippines*

CHINA AIRLINES



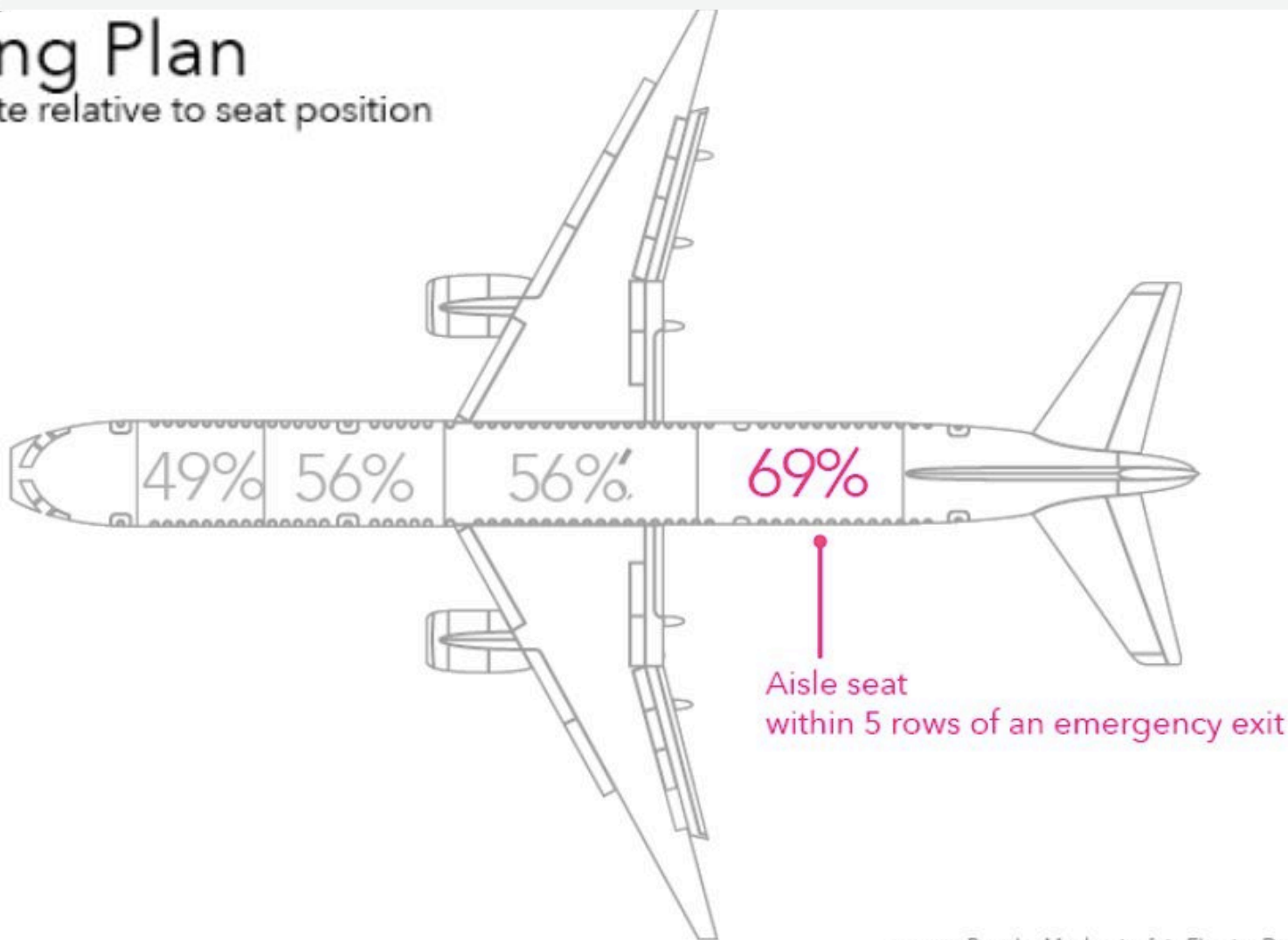
U·S AIRWAYS

# Reduce Your Odds Of Dying In A Plane Crash

A pre-flight check...

## Seating Plan

Survival rate relative to seat position





## Best in Show

The ultimate data-dog

INTELLIGENCE

SIZE

TYPE



public popularity

Inexplicably overrated



Hot dog!



OUR DATA SCORE



Rightly ignored

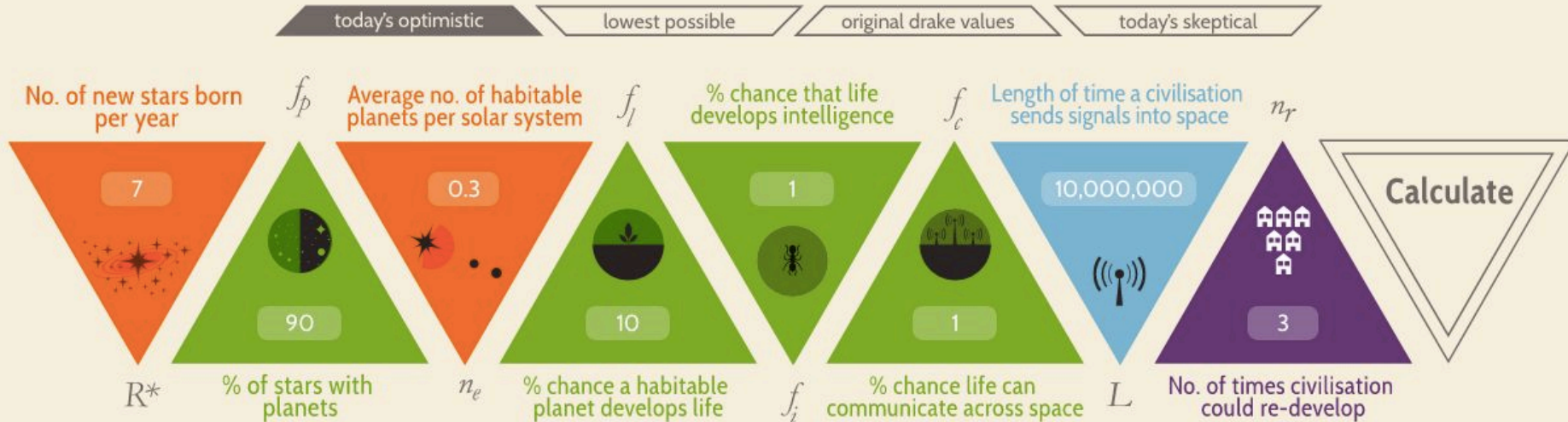
Overlooked treasures

our data score

## Are We Alone in the Universe?

Calculate the Chance of Intelligent Alien Life with the Drake Equation

In 1961, Astronomer Frank Drake came up with an equation to estimate how many detectable extraterrestrial civilizations might exist in our galaxy. Each variable is a crucial factor for the development of alien life.





# Outras ferramentas para visualização

- Online:
  - [onlinecharttool.com](http://onlinecharttool.com), [google charts](http://google.com/charts), [wordclouds.com](http://wordclouds.com), [many eyes](http://manyeyes.com) (java problem?), etc
- Desktop/laptop (linux, windows ou mac)
  - Origin, Excel, Pages
  - Python
  - IDL
  - *etc*

# Quantidades relevantes

- Media
- Mediana
- Desvio padrão
- Porcentagem
- Variação percentual
- Taxa (frequência em relação a um numero fixo)
- Analogias via proporções

Ano	Argentina
2003	43,347
2004	56,193
2005	82,412
2006	86,921
2007	96,055
2008	97,277
2009	103,171
2010	91,056
2011	94,581
2012	93,750
2013	68,596
<u>average</u>	83,033
median	91,056
r.m.s.	19,000

Varição de 2012 a 2013:

-25,154

Varição porcentual de 2012 a 2013:

-36.670

Varição de -37%  
de 2012 a 2013

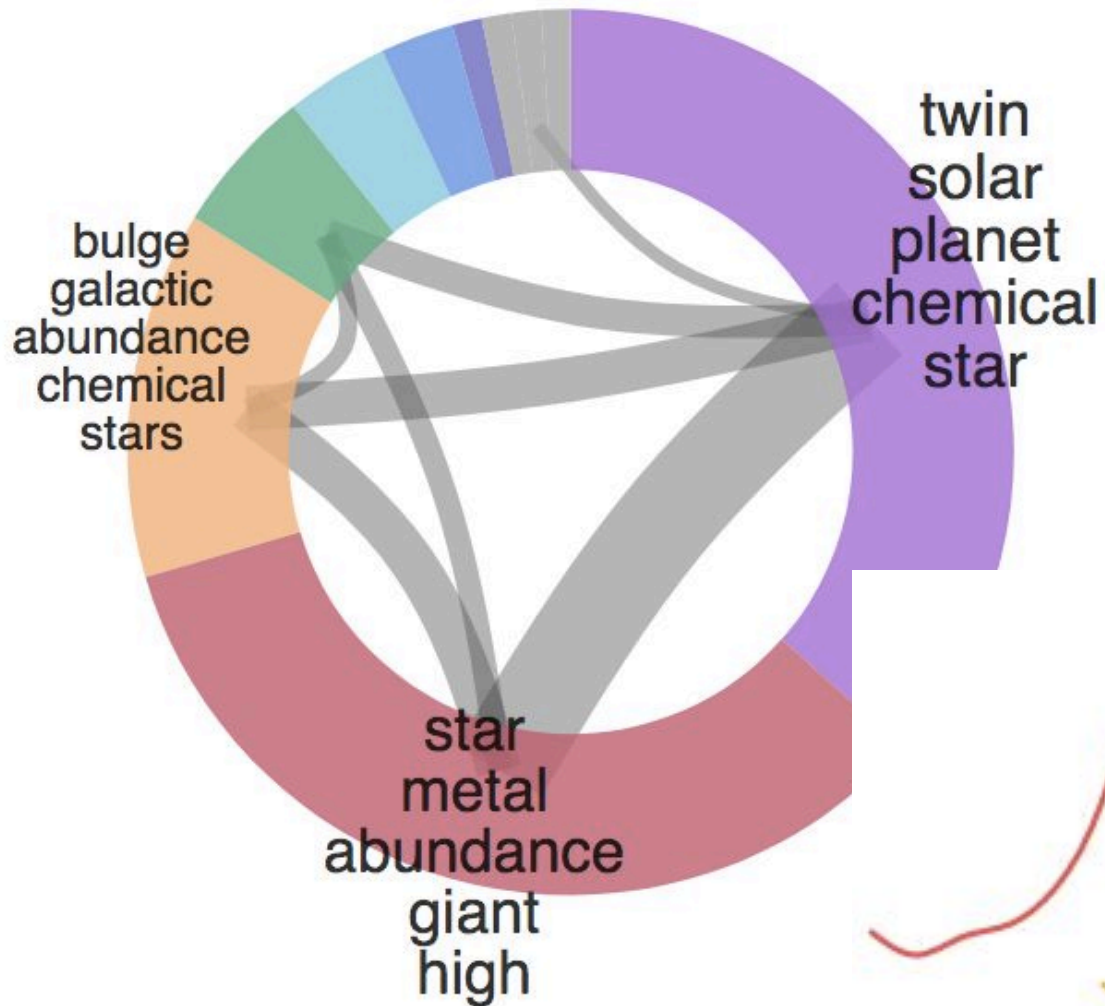
# Usar equivalências

- $\frac{1}{2}$  = metade (50%)
- 25% (um quarto)
- 1 de cada 3 (terceira parte)
- 5% (um de cada vinte)
- 10% (decima parte ou um em cada dez)
- 20% (quinta parte ou um em cada cinco)
- 30% (três em cada 10)
- 80% (quatro em cada cinco)



Pesquisa aos artigos de Melendez, J.

Explore ▾



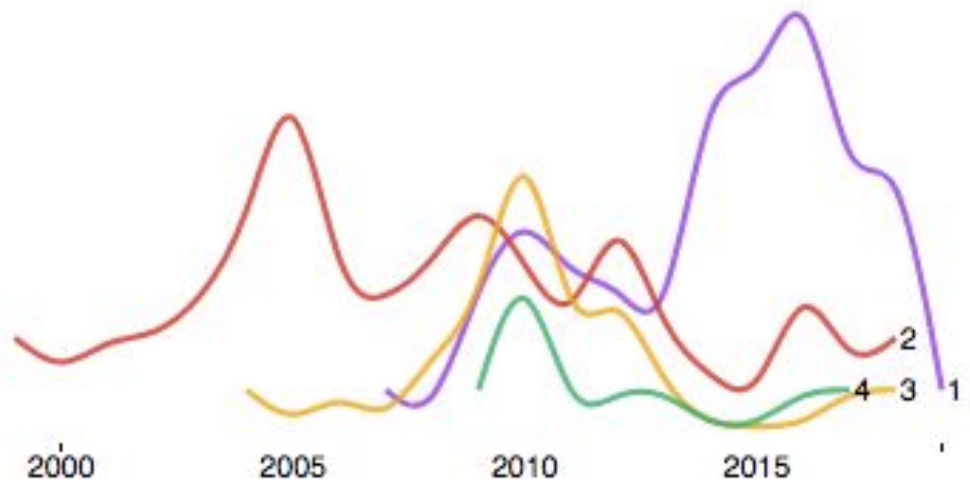
Citation Metrics

Author Network

Paper Network

Concept Cloud

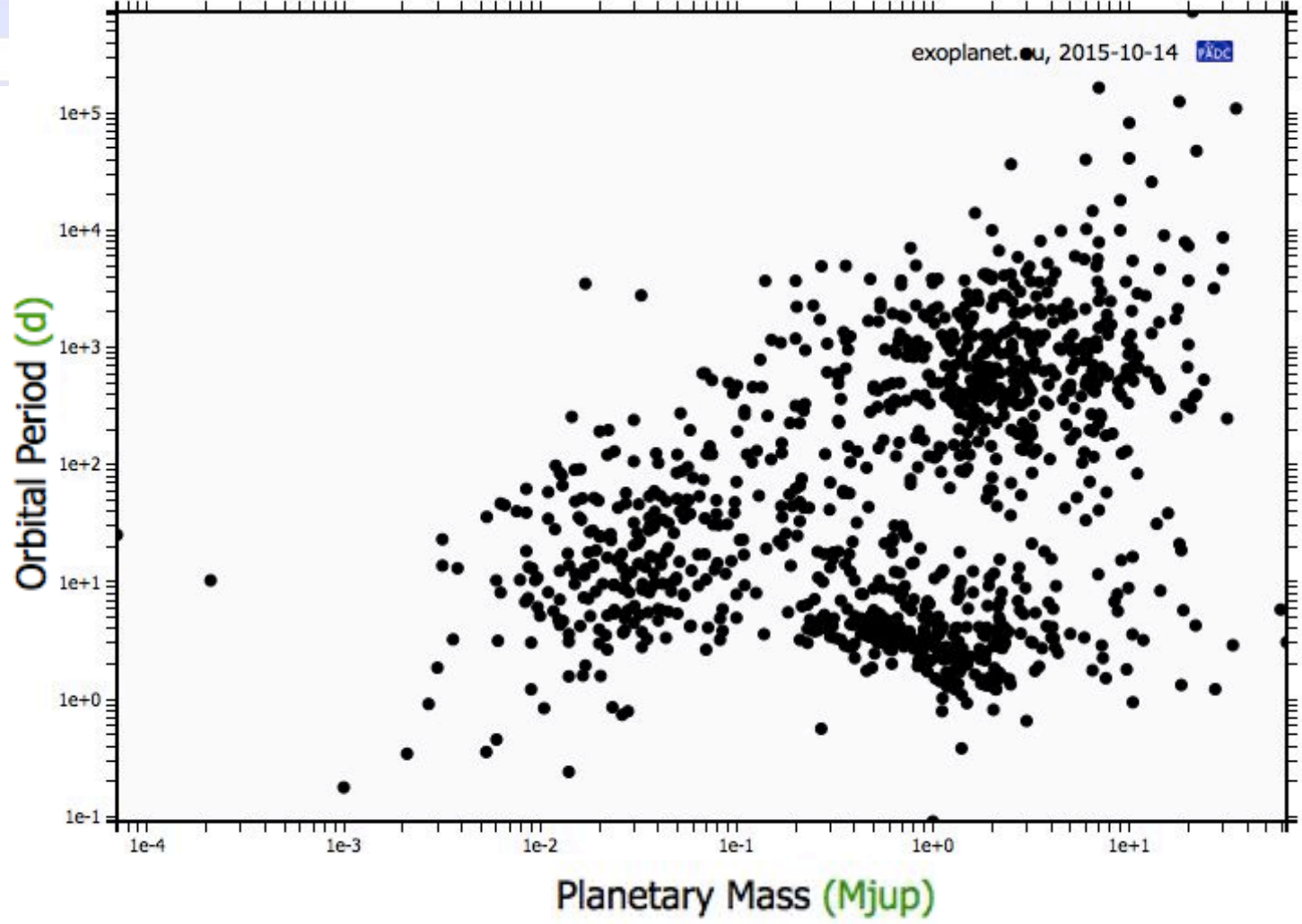
Results Graph



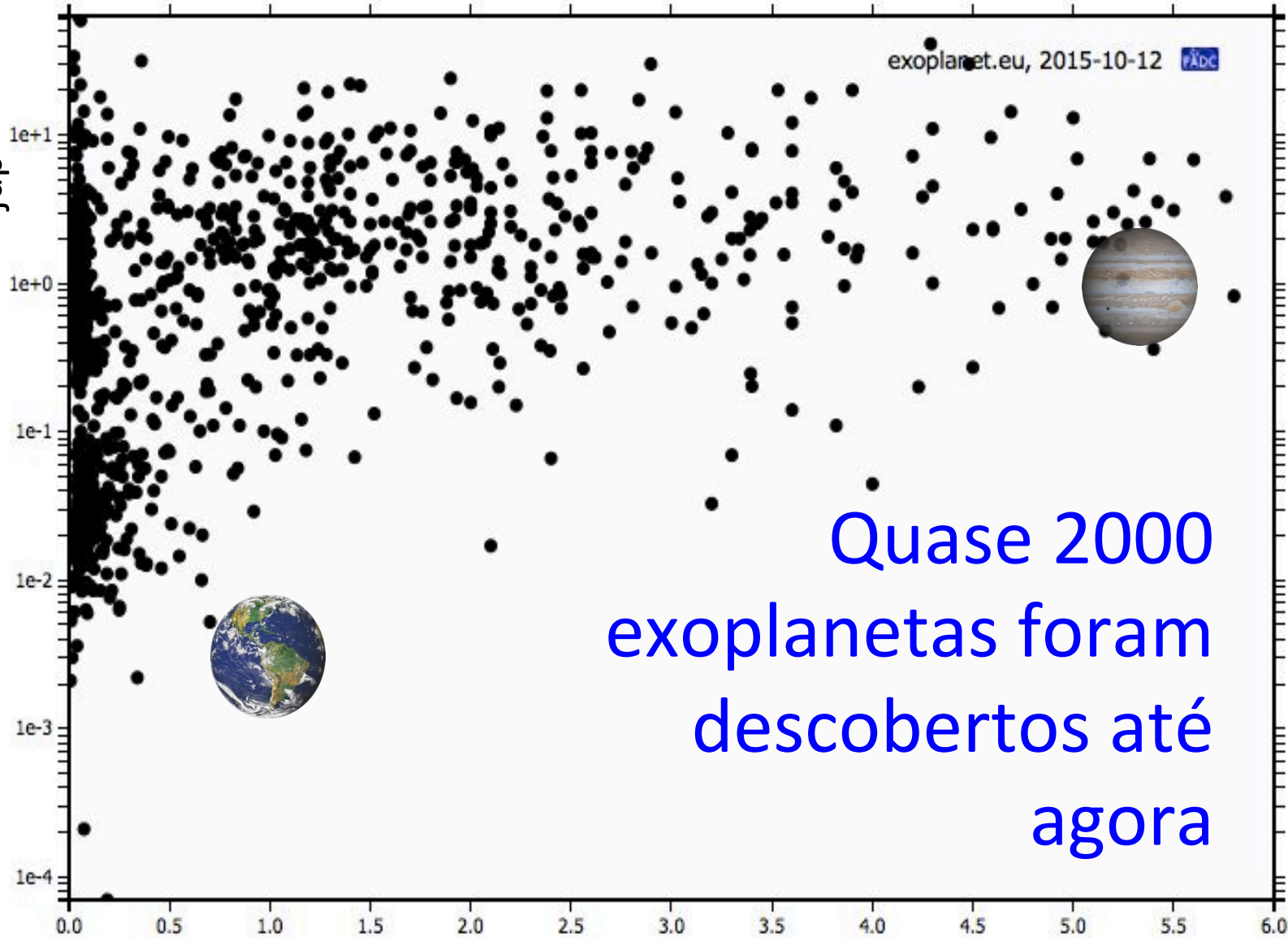
Evolução temporal da pesquisa



Planet	Mass ( $M_{Jup}$ )	Radius ( $R_{Jup}$ )	Period (day)	$a$ (AU)	$e$	$i$ (deg)	Ang. dist. (arcsec)	Discovery	Update
<a href="#">NLTT 41135</a>	33.7	1.13	2.889475	0.024	—	87.42	—	2010	2015-10-05
<a href="#">8 Umi b</a>	1.5	—	93.4	0.49	0.06	—	—	2015	2015-10-01
<a href="#">HD 11755 b</a>								015	2015-10-01
<a href="#">HD 24064 b</a>								015	2015-10-01



Massa do planeta ( $M_{\text{jup}}$ )




Quase 2000  
exoplanetas foram  
descobertos até  
agora

Distância planeta-estrela (UA)



# Sistema Planetário Solar

Mercúrio  
Vênus  
Terra  
Marte



Júpiter



# Sistema Planetário 23 Librae

23 Lib b

23 Lib  
c

# Sistema Planetário Solar

Mercúrio



Vênus



Terra



Marte



Júpiter



# Sistema Planetário Upsilon Andromedae

Ups And c


Ups And d

Ups And  
e

Ups  
And b

# Sistema Planetário Solar

Mercúrio  
Vênus  
Terra  
Marte



Júpiter



# Sistema Planetário HIP 11915

Terrestrial  
planet  
region

HIP  
11915b









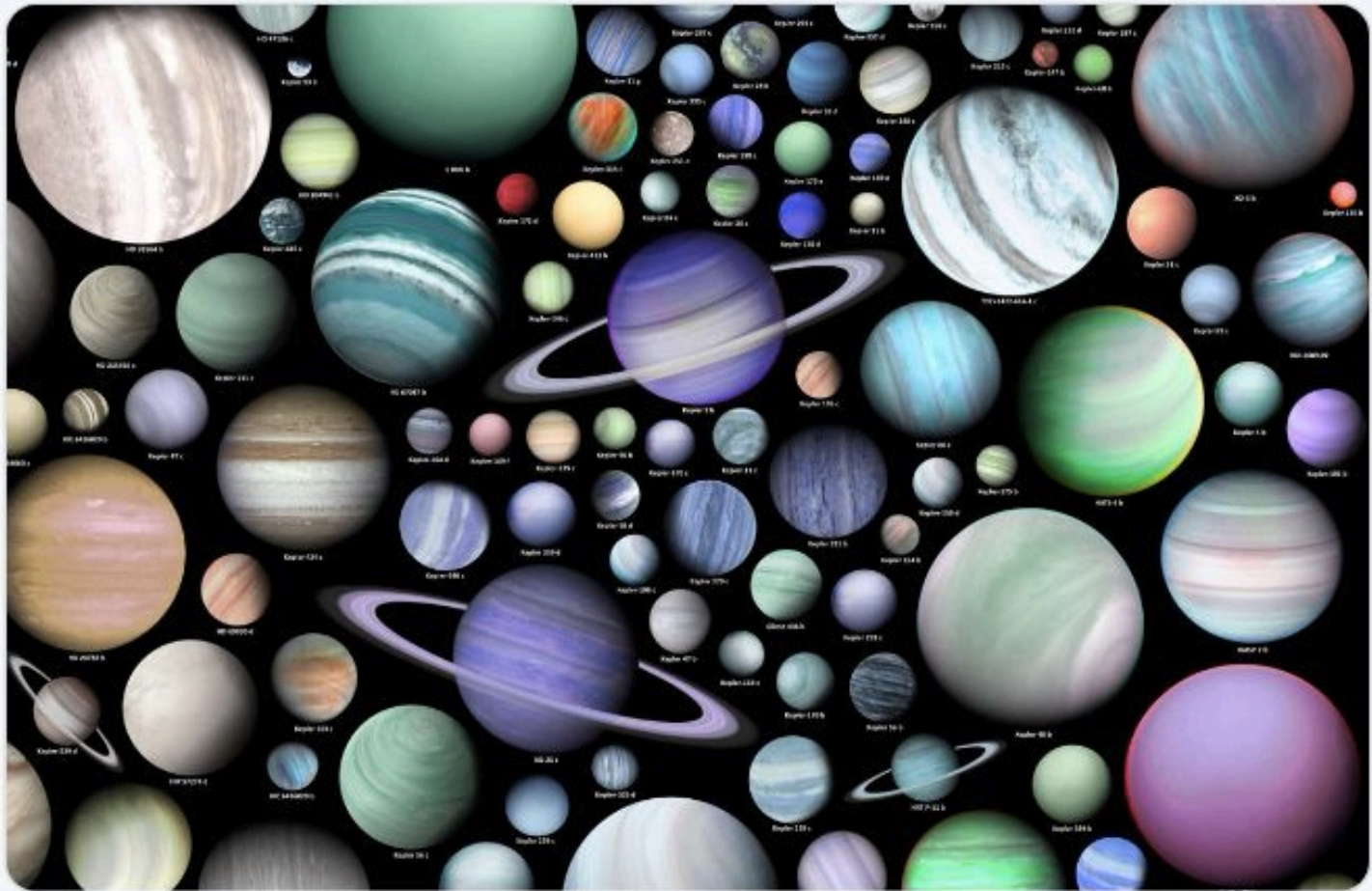


Jorge Melendez @DrJorgeMelendez · 3 de ago

#ExistePesquisaNoBR 🇧🇷

Eu procuro exoplanetas, em particular planetas como o nosso 🌍 (Terra 2.0), usando os telescópios 📡 @ESO.

Também estudo as propriedades fundamentais de estrelas (idade, temperatura, luminosidade, atividade magnética), e a evolução de nossa galáxia 🌀



20

535

2,6 mil

133.525 visualizações

# Homework#8

1. Criar [WordCloud](#) sobre algum assunto relacionado a Astronomia (p.ex., professores, alunos, artigos, divulgação, C.V., lattes [copiar para arquivo], projetos aprovados em observatórios, ficção científica relacionada a astronomia, etc)
2. Fazer uma matéria curta (mínimo 1 p., máximo 2 pp) sobre o seu wordcloud. Pode incluir 1 ou 2 gráficos. Pelo menos um dos gráficos tem que ser um WordCloud. O gráfico pode ocupar até  $\frac{1}{2}$  do trabalho.
3. Prazo de entrega: 28/outubro