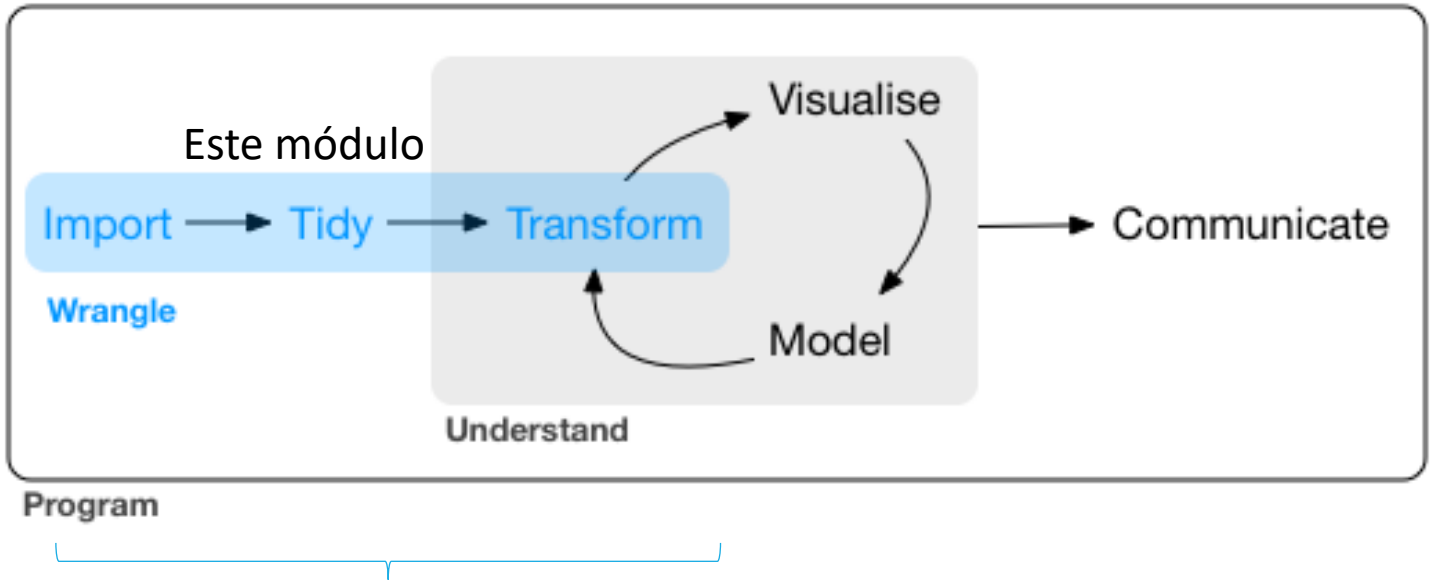
Clase 7:
Conceptos para la limpieza de bases de datos

Universidad de los Andes
Facultad de Economía

Contenido

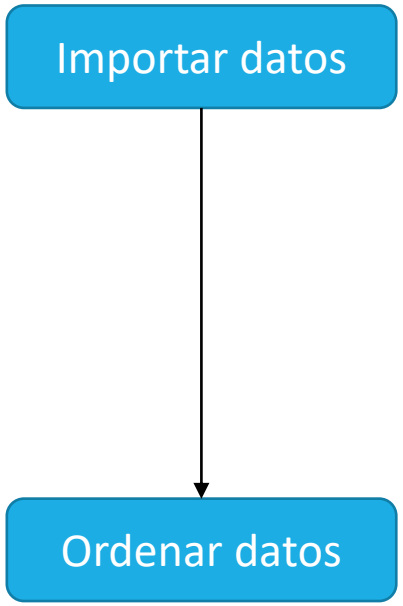
1. Introducción
2. Definiendo datos, variables, observaciones y tablas
3. La unidad de observación (*Tidy data*)
4. Variables: almacenamiento, formato y etiquetas (*parsing*)
5. Los valores faltantes (*missing values)*
6. Importando bases de datos (Ejercicio)
7. Problemas en el ordenamiento de conjuntos de datos
8. Operaciones con datos

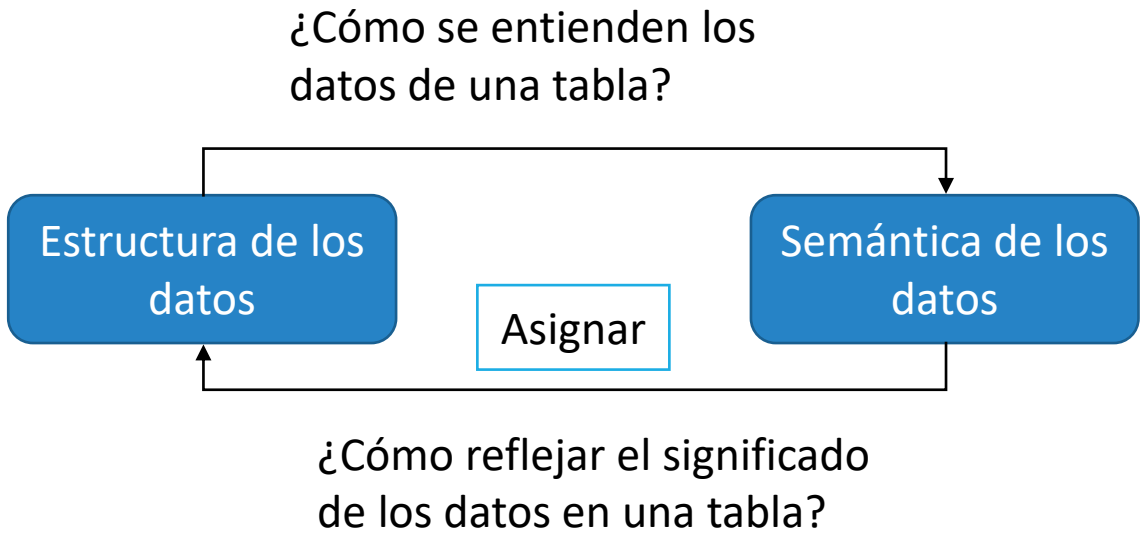
1. Introducción



80% del tiempo del análisis de datos
(Wickham, 2014)

1. Introducción

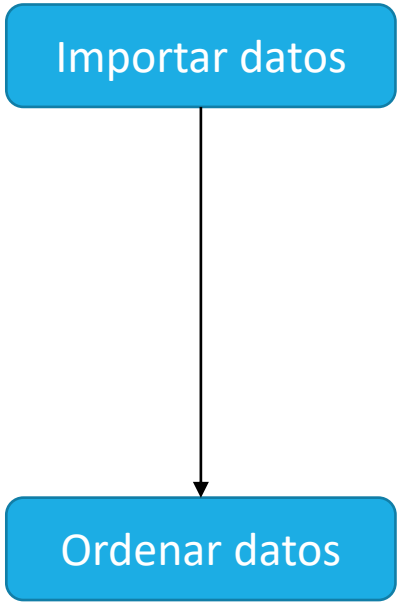




1. Introducción

	DP	DPNOM	DPMP	MPIO	THombres1985	TMujeres1985	Total1985	THombres1993	TMujeres1993	Total1993
1	05	Antioquia	05001	Medellín	692494	757683	1450177	789722	906640	1696362
2	05	Antioquia	05002	Abejorral	14066	13569	27635	13133	12732	25865
3	05	Antioquia	05004	Abriaquí	1477	1374	2851	1605	1199	2804
4	05	Antioquia	05021	Alejandría	2520	2444	4964	2468	2413	4881
5	05	Antioquia	05030	Amagá	11380	11161	22541	12266	12295	24561
6	05	Antioquia	05031	Amalfi	9565	8810	18375	10416	9910	20326
7	05	Antioquia	05034	Andes	20513	19892	40405	20957	20247	41204
8	05	Antioquia	05036	Angelópolis	3204	3003	6207	3182	2904	6086
9	05	Antioquia	05038	Angostura	6798	6856	13654	6638	6490	13128
10	05	Antioquia	05040	Anorí	6342	5850	12192	7181	6532	13713
11	05	Antioquia	05042	Santafé de…	9867	9625	19492	10828	10415	21243
12	05	Antioquia	05044	Anzá	3487	3208	6695	3509	3292	6801
13	05	Antioquia	05045	Apartadó	23835	22223	46058	32380	32245	64625

	DPMP	MPIO	AÑO	THombres	TMujeres	Total
1	05001	Medellín	1985	692494	757683	1450177
2	05001	Medellín	1993	789722	906640	1696362
3	05001	Medellín	2005	951866	1094475	2046341
4	05001	Medellín	2018	1140658	1286471	2427129
5	05002	Abejorral	1985	14066	13569	27635
6	05002	Abejorral	1993	13133	12732	25865
7	05002	Abejorral	2005	11530	11412	22942
8	05002	Abejorral	2018	10534	9833	20367
9	05004	Abriaquí	1985	1477	1374	2851
10	05004	Abriaquí	1993	1605	1199	2804
11	05004	Abriaquí	2005	1392	1327	2719
12	05004	Abriaquí	2018	1452	1243	2695
13	05021	Alejandría	1985	2520	2444	4964
14	05021	Alejandría	1993	2468	2413	4881
15	05021	Alejandría	2005	2336	2388	4724



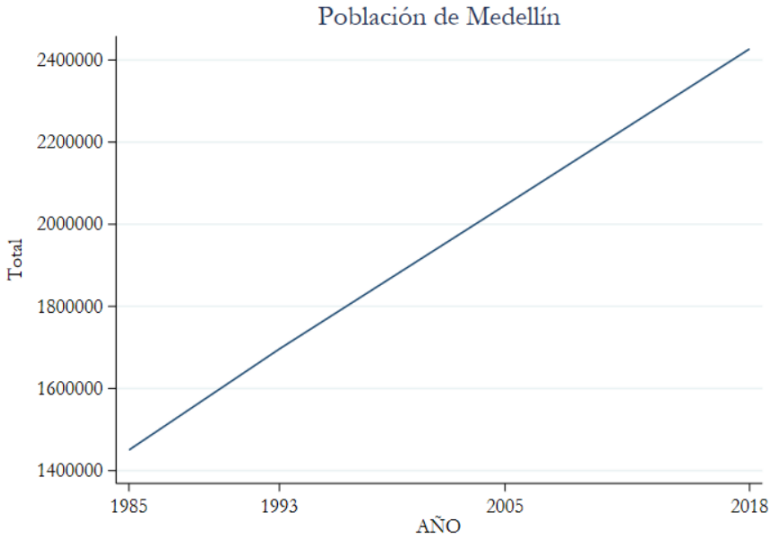
1. Introducción

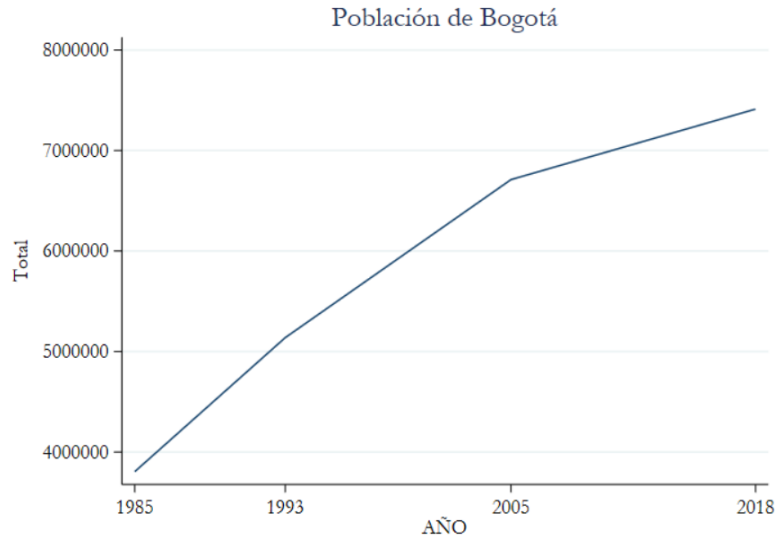
	DP	DPNOM	DPMP	MPIO	THombres1985	TMujeres1985	Total1985	THombres1993	TMujeres1993	Total1993
1	05	Antioquia	05001	Medellín	692494	757683	1450177	789722	906640	1696362
2	05	Antioquia	05002	Abejorral	14066	13569	27635	13133	12732	25865
3	05	Antioquia	05004	Abriaquí	1477	1374	2851	1605	1199	2804
4	05	Antioquia	05021	Alejandría	2520	2444	4964	2468	2413	4881
5	05	Antioquia	05030	Amagá	11380	11161	22541	12266	12295	24561
6	05	Antioquia	05031	Amalfi	9565	8810	18375	10416	9910	20326
7	05	Antioquia	05034	Andes	20513	19892	40405	20957	20247	41204
8	05	Antioquia	05036	Angelópolis	3204	3003	6207	3182	2904	6086
9	05	Antioquia	05038	Angostura	6798	6856	13654	6638	6490	13128
10	05	Antioquia	05040	Anorí	6342	5850	12192	7181	6532	13713
11	05	Antioquia	05042	Santafé de…	9867	9625	19492	10828	10415	21243
12	05	Antioquia	05044	Anzá	3487	3208	6695	3509	3292	6801
13	05	Antioquia	05045	Apartadó	23835	22223	46058	32380	32245	64625

	DPMP	MPIO	AÑO	THombres	TMujeres	Total
1	05001	Medellín	1985	692494	757683	1450177
2	05001	Medellín	1993	789722	906640	1696362
3	05001	Medellín	2005	951866	1094475	2046341
4	05001	Medellín	2018	1140658	1286471	2427129
5	05002	Abejorral	1985	14066	13569	27635
6	05002	Abejorral	1993	13133	12732	25865
7	05002	Abejorral	2005	11530	11412	22942
8	05002	Abejorral	2018	10534	9833	20367
9	05004	Abriaquí	1985	1477	1374	2851
10	05004	Abriaquí	1993	1605	1199	2804
11	05004	Abriaquí	2005	1392	1327	2719
12	05004	Abriaquí	2018	1452	1243	2695
13	05021	Alejandría	1985	2520	2444	4964
14	05021	Alejandría	1993	2468	2413	4881
15	05021	Alejandría	2005	2336	2388	4724

¿Cómo le diría al computador que cree uno de estos gráficos?

¿Con cuál tabla es más fácil crear estos gráficos?





Ordenamiento de datos que permite fácil procesamiento con lenguajes vectorizados como R o Python

Universidad de los Andes
Facultad de Economía

2. Definiciones

Datos — Colección de valores. Números o caracteres.

Variables — Todos los valores que miden el mismo atributo.

Observaciones — Todos los valores relacionados con una unidad

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
589	08	Atlántico	08849	Usiacurí	1985	2823	2641	5464
590	08	Atlántico	08849	Usiacurí	1993	3763	3582	7345
591	08	Atlántico	08849	Usiacurí	2005	5046	4743	9789
592	08	Atlántico	08849	Usiacurí	2018	6276	5974	12250
593	11	Bogotá, D.…	11001	Bogotá, D.…	1985	1916013	1888623	3804636
594	11	Bogotá, D.…	11001	Bogotá, D.…	1993	2532440	2605793	5138233
595	11	Bogotá, D.…	11001	Bogotá, D.…	2005	3236477	3474433	6710910
596	11	Bogotá, D.…	11001	Bogotá, D.…	2018	3544078	3868488	7412566
597	13	Bolívar	13001	Cartagena …	1985	231504	269406	500910
598	13	Bolívar	13001	Cartagena …	1993	306308	350528	656836
599	13	Bolívar	13001	Cartagena …	2005	402047	436710	838757
600	13	Bolívar	13001	Cartagena …	2018	468058	504987	973045
601	13	Bolívar	13006	Achí	1985	14641	13495	28136
602	13	Bolívar	13006	Achí	1993	14403	13023	27426
603	13	Bolívar	13006	Achí	2005	10220	9255	19475
604	13	Bolívar	13006	Achí	2018	12816	11536	24352

2. Definiciones

Datos

Colección de valores.
Números o caracteres.

Variables

Todos los valores que miden el mismo atributo.

Observaciones

Todos los valores relacionados con una unidad

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
589	08	Atlántico	08849	Usiacurí	1985	2823	2641	5464
590	08	Atlántico	08849	Usiacurí	1993	3763	3582	7345
591	08	Atlántico	08849	Usiacurí	2005	5046	4743	9789
592	08	Atlántico	08849	Usiacurí	2018	6276	5974	12250
593	11	Bogotá, D.…	11001	Bogotá, D.…	1985	1916013	1888623	3804636
594	11	Bogotá, D.…	11001	Bogotá, D.…	1993	2532440	2605793	5138233
595	11	Bogotá, D.…	11001	Bogotá, D.…	2005	3236477	3474433	6710910
596	11	Bogotá, D.…	11001	Bogotá, D.…	2018	3544078	3868488	7412566
597	13	Bolívar	13001	Cartagena …	1985	231504	269406	500910
598	13	Bolívar	13001	Cartagena …	1993	306308	350528	656836
599	13	Bolívar	13001	Cartagena …	2005	402047	436710	838757
600	13	Bolívar	13001	Cartagena …	2018	468058	504987	973045
601	13	Bolívar	13006	Achí	1985	14641	13495	28136
602	13	Bolívar	13006	Achí	1993	14403	13023	27426
603	13	Bolívar	13006	Achí	2005	10220	9255	19475
604	13	Bolívar	13006	Achí	2018	12816	11536	24352

2. Definiciones

Datos — Colección de valores. Números o caracteres.

Variables — Todos los valores que miden el mismo atributo.

Observaciones — Todos los valores relacionados con una unidad

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
589	08	Atlántico	08849	Usiacurí	1985	2823	2641	5464
590	08	Atlántico	08849	Usiacurí	1993	3763	3582	7345
591	08	Atlántico	08849	Usiacurí	2005	5046	4743	9789
592	08	Atlántico	08849	Usiacurí	2018	6276	5974	12250
593	11	Bogotá, D.…	11001	Bogotá, D.…	1985	1916013	1888623	3804636
594	11	Bogotá, D.…	11001	Bogotá, D.…	1993	2532440	2605793	5138233
595	11	Bogotá, D.…	11001	Bogotá, D.…	2005	3236477	3474433	6710910
596	11	Bogotá, D.…	11001	Bogotá, D.…	2018	3544078	3868488	7412566
597	13	Bolívar	13001	Cartagena …	1985	231504	269406	500910
598	13	Bolívar	13001	Cartagena …	1993	306308	350528	656836
599	13	Bolívar	13001	Cartagena …	2005	402047	436710	838757
600	13	Bolívar	13001	Cartagena …	2018	468058	504987	973045
601	13	Bolívar	13006	Achí	1985	14641	13495	28136
602	13	Bolívar	13006	Achí	1993	14403	13023	27426
603	13	Bolívar	13006	Achí	2005	10220	9255	19475
604	13	Bolívar	13006	Achí	2018	12816	11536	24352

3. La unidad de observación (Tidy data)

```
sort DP DPMP AÑO
```

Orden de la variables de menor a mayor, por jerarquía.

medición

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
589	08	Atlántico	08849	Usiacurí	1985	2823	2641	5464
590	08	Atlántico	08849	Usiacurí	1993	3763	3582	7345
591	08	Atlántico	08849	Usiacurí	2005	5046	4743	9789
592	08	Atlántico	08849	Usiacurí	2018	6276	5974	12250
593	11	Bogotá, D.…	11001	Bogotá, D.…	1985	1916013	1888623	3804636
594	11	Bogotá, D.…	11001	Bogotá, D.…	1993	2532440	2605793	5138233
595	11	Bogotá, D.…	11001	Bogotá, D.…	2005	3236477	3474433	6710910
596	11	Bogotá, D.…	11001	Bogotá, D.…	2018	3544078	3868488	7412566
597	13	Bolívar	13001	Cartagena …	1985	231504	269406	500910
598	13	Bolívar	13001	Cartagena …	1993	306308	350528	656836
599	13	Bolívar	13001	Cartagena …	2005	402047	436710	838757
600	13	Bolívar	13001	Cartagena …	2018	468058	504987	973045
601	13	Bolívar	13006	Achí	1985	14641	13495	28136
602	13	Bolívar	13006	Achí	1993	14403	13023	27426
603	13	Bolívar	13006	Achí	2005	10220	9255	19475

Identificación

Variables fijas, que describen el diseño de la medición

1. Variables organizadas por su rol en el análisis

2. Las comparaciones entre grupos deben hacerse por observaciones. e.g. Comparar municipios de los departamentos Bolivar y Meta.

3. Entre las variables puede haber una relación funcional. e.g. La suma de la población de hombres y la población de mujeres es igual a la población total.

Esta tabla reúne información municipal por año de la población total, de hombres y de mujeres

3. La unidad de observación (Tidy data)

Se registra el resultado de 2 tratamientos en 3 personas

¿Se mide a las personas?

	Nombre	Ra	Rb
1	Pedro	.	2
2	Juana	16	11
3	Oscar	3	1

¿Se mide el tratamiento?

	Trat	Pedro	Juana	Oscar
1	a	.	16	3
2	b	2	11	1

	Nombre	Trat	R
1	Juana	a	16
2	Juana	b	11
3	Oscar	a	3
4	Oscar	b	1
5	Pedro	a	.
6	Pedro	b	2

Se mide el resultado de cada tratamiento en cada persona

4. Almacenamiento y formato de variables

	make	price	rep78	length	displacement	gear_ratio	foreign	date
1	AMC Concord	4,099	3	186	121	3.58	Domestic	28dec1982
2	AMC Pacer	4,749	3	173	258	2.53	Domestic	16jun1995
3	AMC Spirit	3,799	.	168	121	3.08	Domestic	05nov2009
4	Buick Century	4,816	3	196	196	2.93	Domestic	14apr1997
5	Buick Electra	7,827	4	222	350	2.41	Domestic	02dec2006
6	Buick LeSabre	5,788	3	218	231	2.73	Domestic	08jul1963
7	Buick Opel	4,453	.	170	304	2.87	Domestic	11dec2005

Índice de observación (_n)

Variables numéricas

Storage type	Minimum	Maximum	Closest to 0 without being 0	bytes
byte	-127	100	+/-1	1
int	-32,767	32,740	+/-1	2
long	-2,147,483,647	2,147,483,620	+/-1	4
float	-1.70141173319*10^38	1.70141173319*10^38	+/-10^-38	4
double	-8.9884656743*10^307	8.9884656743*10^307	+/-10^-323	8

Precision for **float** is 3.795x10^-8.
Precision for **double** is 1.414x10^-16.

V. discretas

V. continuas

```
. describe

Contains data from https://www.stata-press.com/data/r16/aut
  obs:            74                          1978 Automobi
 vars:            13                          13 Apr 2018 1
                                              (_dta has not

-------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable labe
-------------------------------------------------------------
make            str17   %-17s                 Make and Mode
price           int     %8.0gc                Price
mpg             byte    %8.0g                 Mileage (mpg)
rep78           byte    %8.0g                 Repair Record
headroom        float   %6.1f                 Headroom (in.
trunk           byte    %8.0g                 Trunk space (
weight          int     %8.0gc                Weight (lbs.)
length          int     %8.0g                 Length (in.)
turn            byte    %8.0g                 Turn Circle (
displacement    int     %8.0g                 Displacement
gear_ratio      float   %6.2f                 Gear Ratio
foreign         byte    %8.0g       origin    Car type
date            float   %td                   Owner's date
-------------------------------------------------------------
Sorted by: foreign
```

Universidad de los Andes
Facultad de Economía

4. Almacenamiento y formato de variables

	make	price	rep78	length	displacement	gear_ratio	foreign	date
1	AMC Concord	4,099	3	186	121	3.58	Domestic	28dec1982
2	AMC Pacer	4,749	3	173	258	2.53	Domestic	16jun1995
3	AMC Spirit	3,799	.	168	121	3.08	Domestic	05nov2009
4	Buick Century	4,816	3	196	196	2.93	Domestic	14apr1997
5	Buick Electra	7,827	4	222	350	2.41	Domestic	02dec2006
6	Buick LeSabre	5,788	3	218	231	2.73	Domestic	08jul1963
7	Buick Opel	4,453	.	170	304	2.87	Domestic	11dec2005

Variable de fecha en días (numérica)

Variable discreta

```
format date %tg
```

gear_ratio	foreign	date
3.58	Domestic	8397.613
2.53	Domestic	12951
3.08	Domestic	18206.29
2.93	Domestic	13618.44
2.41	Domestic	17137.65
2.73	Domestic	1284.036

¿cuántos días han pasado desde el 1 de enero de 1960

```
. describe

Contains data from https://www.stata-press.com/data/r16/aut
  obs:            74                          1978 Automobi
 vars:            13                          13 Apr 2018 1
                                              (_dta has not

-------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable labe
-------------------------------------------------------------
make            str17   %-17s                 Make and Mode
price           int     %8.0gc                Price
mpg             byte    %8.0g                 Mileage (mpg)
rep78           byte    %8.0g                 Repair Record
headroom        float   %6.1f                 Headroom (in.
trunk           byte    %8.0g                 Trunk space (
weight          int     %8.0gc                Weight (lbs.)
length          int     %8.0g                 Length (in.)
turn            byte    %8.0g                 Turn Circle (
displacement    int     %8.0g                 Displacement
gear_ratio      float   %6.2f                 Gear Ratio
foreign         byte    %8.0g       origin    Car type
date            float   %td                   Owner's date
-------------------------------------------------------------
Sorted by: foreign
```

Universidad de los Andes
Facultad de Economía

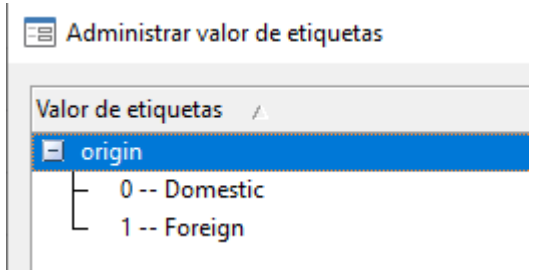
4. Almacenamiento y formato de variables

	make	price	rep78	length	displacement	gear_ratio	foreign	date
1	AMC Concord	4,099	3	186	121	3.58	Domestic	28dec1982
2	AMC Pacer	4,749	3	173	258	2.53	Domestic	16jun1995
3	AMC Spirit	3,799	.	168	121	3.08	Domestic	05nov2009
4	Buick Century	4,816	3	196	196	2.93	Domestic	14apr1997
5	Buick Electra	7,827	4	222	350	2.41	Domestic	02dec2006
6	Buick LeSabre	5,788	3	218	231	2.73	Domestic	08jul1963
7	Buick Opel	4,453	.	170	304	2.87	Domestic	11dec2005

Variable categórica

Variable numérica con etiqueta de valor

```
label define origin 0 "Domestic" 1 "Foreign"
label value foreign origin
```




```
. describe

Contains data from https://www.stata-press.com/data/r16/aut
  obs:            74                          1978 Automobi
 vars:            13                          13 Apr 2018 1
                                              (_dta has not

-------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable labe
-------------------------------------------------------------
make            str17   %-17s                 Make and Mode
price           int     %8.0gc                Price
mpg             byte    %8.0g                 Mileage (mpg)
rep78           byte    %8.0g                 Repair Record
headroom        float   %6.1f                 Headroom (in.
trunk           byte    %8.0g                 Trunk space (
weight          int     %8.0gc                Weight (lbs.)
length          int     %8.0g                 Length (in.)
turn            byte    %8.0g                 Turn Circle (
displacement    int     %8.0g                 Displacement
gear_ratio      float   %6.2f                 Gear Ratio
foreign         byte    %8.0g       origin    Car type
date            float   %td                   Owner's date
-------------------------------------------------------------
Sorted by: foreign
```

Universidad de los Andes
Facultad de Economía

4. Almacenamiento y formato de variables

	make	price	rep78	length	displacement	gear_ratio	foreign	date
1	AMC Concord	4,099	3	186	121	3.58	Domestic	28dec1982
2	AMC Pacer	4,749	3	173	258	2.53	Domestic	16jun1995
3	AMC Spirit	3,799	.	168	121	3.08	Domestic	05nov2009
4	Buick Century	4,816	3	196	196	2.93	Domestic	14apr1997
5	Buick Electra	7,827	4	222	350	2.41	Domestic	02dec2006
6	Buick LeSabre	5,788	3	218	231	2.73	Domestic	08jul1963
7	Buick Opel	4,453	.	170	304	2.87	Domestic	11dec2005

Variable con cadenas de caracteres

Variable categórica

String storage type	Maximum length	Bytes
str1	1	1
str2	2	2
...	.	.
...	.	.
...	.	.
str2045	2045	2045
strL	2000000000	2000000000

atio	foreign	date	year	month	day
3.58	Domestic	28dec1982	1982	12	28
2.53	Domestic	16jun1995	1995	6	16
3.08	Domestic	05nov2009	2009	11	5
2.93	Domestic	14apr1997	1997	4	14
2.41	Domestic	02dec2006	2006	12	2
2.73	Domestic	08jul1963	1963	7	8
2.87	Domestic	11dec2005	2005	12	11

```
. describe

Contains data from https://www.stata-press.com/data/r16/aut
  obs:            74                          1978 Automobi
 vars:            13                          13 Apr 2018 1
                                              (_dta has not

-------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable labe
-------------------------------------------------------------
make            str17   %-17s                 Make and Mode
price           int     %8.0gc                Price
mpg             byte    %8.0g                 Mileage (mpg)
rep78           byte    %8.0g                 Repair Record
headroom        float   %6.1f                 Headroom (in.
trunk           byte    %8.0g                 Trunk space (
weight          int     %8.0gc                Weight (lbs.)
length          int     %8.0g                 Length (in.)
turn            byte    %8.0g                 Turn Circle (
displacement    int     %8.0g                 Displacement
gear_ratio      float   %6.2f                 Gear Ratio
foreign         byte    %8.0g       origin    Car type
date            float   %td                   Owner's date
-------------------------------------------------------------
Sorted by: foreign
```

Universidad de los Andes
Facultad de Economía

4. Formato y etiquetas de variables

	make	price	rep78	length	displacement	gear_ratio	foreign	date
1	AMC Concord	4,099	3	186	121	3.58	Domestic	28dec1982
2	AMC Pacer	4,749	3	173	258	2.53	Domestic	16jun1995
3	AMC Spirit	3,799	.	168	121	3.08	Domestic	05nov2009
4	Buick Century	4,816	3	196	196	2.93	Domestic	14apr1997
5	Buick Electra	7,827	4	222	350	2.41	Domestic	02dec2006
6	Buick LeSabre	5,788	3	218	231	2.73	Domestic	08jul1963
7	Buick Opel	4,453	.	170	304	2.87	Domestic	11dec2005

Separador de miles

```
format price %8.0gc

format gear_ratio %6.2f
```

2 decimales

```
. describe make price rep78 length displacement gear_ratio foreign date

              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
make            str17   %-17s                 Make and Model
price           int     %8.0gc                Price
rep78           byte    %8.0g                 Repair Record 1978
length          int     %8.0g                 Length (in.)
displacement    int     %8.0g                 Displacement (cu. in.)
gear_ratio      float   %6.2f                 Gear Ratio
foreign         byte    %8.0g      origin     Car type
date            float   %td                   Owner's date of birth
```

```
label var fecha "Owner's date of birth"
```

Universidad de los Andes
Facultad de Economía

5. Los valores faltantes (missing values)

	make	price	rep78	length	displacement	gear_ratio	foreign	date
1	AMC Concord	4099	3	186	121	3.58	Domestic	28dec1982
2	AMC Pacer	4749	3	173	258	2.53	Domestic	16jun1995
3	AMC Spirit	3799	.	168	121	3.08	Domestic	05nov2009
4	Buick Century	4816	3	196	196	2.93	Domestic	14apr1997
5	Buick Electra	7827	4	222	350	2.41	Domestic	02dec2006
6		5788	3	218	231	2.73	Domestic	08jul1963
7	Buick Opel	4453	.	170	304	2.87	Domestic	11dec2005

```
rep78                                         Repair Record 1978

                  type:  numeric (byte)

                 range:  [1,5]                        units:  1
         unique values:  5                    missing .:  5/74

           tabulation:  Freq.  Value
                           2  1
                           8  2
                          30  3
                          18  4
                          11  5
                           5  .
```

¿Aquí podría ser cero?

Medición que debió hacerse pero no se hizo o no se puede obtener su resultado.
e.g. Problemas con el instrumento, dato ilegible

El valor faltante no necesariamente es igual a cero.

El dato existe pero no se puede acceder a él.

6. Importando bases de datos

Revisar si hay datos numéricos con cero inicial

Revisar si hay cadenas de caracteres muy largas o con tildes

Identificar el separador de las variables

```
. type poblacion.csv, lines(10)
DP,DPNOM,DPMP,MPIO,AÑO,TotalHombres,TotalMujeres,TotalGeneral
05,Antioquia,05001,Medellín,1985,692494,757683,1450177
05,Antioquia,05001,Medellín,1993,789722,906640,1696362
05,Antioquia,05002,Abejorral,1985,14066,13569,27635
05,Antioquia,05002,Abejorral,1993,13133,12732,25865
05,Antioquia,05004,Abriaquí,1985,1477,1374,2851
05,Antioquia,05004,Abriaquí,1993,1605,1199,2804
05,Antioquia,05021,Alejandría,1985,2520,2444,4964
05,Antioquia,05021,Alejandría,1993,2468,2413,4881
05,Antioquia,05030,Amagá,1985,11380,11161,22541
```

Previsualización de un archivo separado por comas (.csv) en Stata

6. Importando bases de datos

Los ceros iniciales se pierden si las variables con códigos se importan como números

Las variables que contienen códigos deben ser caracteres

Las variables con caracteres que contienen tildes deben ser legibles

Cuando las tildes no se reconocen puede haber un problema de codificación en la base. Común con datos internacionales.

	dp	dpnom	dpmp	mpio	aão	totalhom~s	totalmuj~s	totalgen~l
1	5	Antioquia	5001	MedellÃ-n	1985	692494	757683	1450177
2	5	Antioquia	5001	MedellÃ-n	1993	789722	906640	1696362
3	5	Antioquia	5002	Abejorral	1985	14066	13569	27635
4	5	Antioquia	5002	Abejorral	1993	13133	12732	25865
5	5	Antioquia	5004	AbriaquÃ-	1985	1477	1374	2851
6	5	Antioquia	5004	AbriaquÃ-	1993	1605	1199	2804

```
import delimited poblacion.csv, clear delim(",")

import delimited poblacion.csv, clear delim(",") stringcols(1 3) encoding("utf-8")
```

	dp	dpnom	dpmp	mpio	año	totalhombres	totalmujeres	totalgeneral
1	05	Antioquia	05001	Medellín	1985	692494	757683	1450177
2	05	Antioquia	05001	Medellín	1993	789722	906640	1696362
3	05	Antioquia	05002	Abejorral	1985	14066	13569	27635
4	05	Antioquia	05002	Abejorral	1993	13133	12732	25865
5	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851
6	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804

Verificar el comportamiento de las variables numéricas con histogramas o cajas de distribución.

¿Valores extremos?

6. Importando bases de datos

Clic aquí

Vaya a la dirección: http://microdatos.dane.gov.co/index.php/catalog/701/get_microdata

Descargue los datos de la Gran Encuesta Integrada de Hogares de Enero del año 2021 en un archivo separado por comas.

Universidad de los Andes
Facultad de Economía

6. Importando bases de datos

1. Abrir el archivo comprimido y colocar los datos en su carpeta de trabajo.
2. Importar con algún lenguaje el archivo separado por comas: Cabecera - Vivienda y Hogares
    1. Stata: import delimited
    2. Python: read_csv() de la librería pandas
    3. R: read_csv() de la librería readr
3. Apenas cargue los datos visualicelos en una tabla. *Recuerde, en STATA puede usar br (browse), en R puede usar View(datos), y en Python con pandas puede usar datos.head()*

6. Importando bases de datos

1. ¿Quedaron bien cargados? Para empezar. Revise algunas de las siguientes:
    1. Las columnas tienen información coherente, ¿usó el delimitador correcto?
    2. ¿Hay datos numéricos con 0 inicial?
    3. ¿Cadenas de caracteres con tildes? Revise la codificación. ¿utf-8?
    4. ¿Las columnas con info numérica son de hecho columnas numéricas?
    5. Si quiere darle elegancia, en STATA puede agregar etiquetas a valores categóricos
2. Cada vez que haga una modificación, ¡visualice! Esto no es una lista memorizable, su mapa de ruta es su base misma.

7. Problemas en el ordenamiento de conjuntos de datos

1. Varias variables almacenadas en una columna

	DP	DPNOM	DPMP_AÑO	MPIO_AÑO	TotalHombres	TotalMujeres	TotalGeneral
1	05	Antioquia	05001-1985	Medellín-1985	692494	757683	1450177
2	05	Antioquia	05001-1993	Medellín-1993	789722	906640	1696362
3	05	Antioquia	05001-2005	Medellín-2005	951866	1094475	2046341
4	05	Antioquia	05001-2018	Medellín-2018	1140658	1286471	2427129
5	05	Antioquia	05002-1985	Abejorral-1985	14066	13569	27635
6	05	Antioquia	05002-1993	Abejorral-1993	13133	12732	25865
7	05	Antioquia	05002-2005	Abejorral-2005	11530	11412	22942
8	05	Antioquia	05002-2018	Abejorral-2018	10534	9833	20367
9	05	Antioquia	05004-1985	Abriaquí-1985	1477	1374	2851
10	05	Antioquia	05004-1993	Abriaquí-1993	1605	1199	2804

```
gen AÑO1=substr(DPMP_AÑO, -4,.) // Separar año
gen AÑO2=substr(MPIO_AÑO, -4,.) // Separar año
count if AÑO1==AÑO2 // Son idénticas
drop AÑO2 // Borrar variable redundante
rename AÑO1 AÑO // Cambiar el nombre
destring AÑO, replace // Convertir a numérica
```

Lo estudiaremos con detalle en la clase 8

Proceso similar para arreglar las variables DPMP y MPIO

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
1	05	Antioquia	05001	Medellín	1985	692494	757683	1450177
2	05	Antioquia	05001	Medellín	1993	789722	906640	1696362
3	05	Antioquia	05001	Medellín	2005	951866	1094475	2046341
4	05	Antioquia	05001	Medellín	2018	1140658	1286471	2427129
5	05	Antioquia	05002	Abejorral	1985	14066	13569	27635
6	05	Antioquia	05002	Abejorral	1993	13133	12732	25865
7	05	Antioquia	05002	Abejorral	2005	11530	11412	22942
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367
9	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851
10	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804
11	05	Antioquia	05004	Abriaquí	2005	1392	1327	2719
12	05	Antioquia	05004	Abriaquí	2018	1452	1243	2695
13	05	Antioquia	05021	Alejandr…	1985	2520	2444	4964
14	05	Antioquia	05021	Alejandr…	1993	2468	2413	4881
15	05	Antioquia	05021	Alejandr…	2005	2336	2388	4724
16	05	Antioquia	05021	Alejandr…	2018	2320	2337	4657
17	05	Antioquia	05030	Amagá	1985	11380	11161	22541
18	05	Antioquia	05030	Amagá	1993	12266	12295	24561
19	05	Antioquia	05030	Amagá	2005	13439	13682	27121
20	05	Antioquia	05030	Amagá	2018	14862	15365	30227
21	05	Antioquia	05031	Amalfi	1985	9565	8810	18375

7. Problemas en el ordenamiento de conjuntos de datos

2. Nombres de columnas como números o valores

	DP	DPNOM	DPMP	MPIO	THom~1985	TMuj~1985	Total1985	THom~1993	TMuj~1993	Total1993
1	05	Antioquia	05001	Medellín	692494	757683	1450177	789722	906640	1696362
2	05	Antioquia	05002	Abejorral	14066	13569	27635	13133	12732	25865
3	05	Antioquia	05004	Abriaquí	1477	1374	2851	1605	1199	2804
4	05	Antioquia	05021	Alejandr…	2520	2444	4964	2468	2413	4881
5	05	Antioquia	05030	Amagá	11380	11161	22541	12266	12295	24561
6	05	Antioquia	05031	Amalfi	9565	8810	18375	10416	9910	20326
7	05	Antioquia	05034	Andes	20513	19892	40405	20957	20247	41204
8	05	Antioquia	05036	Angelópo…	3204	3003	6207	3182	2904	6086
9	05	Antioquia	05038	Angostura	6798	6856	13654	6638	6490	13128
10	05	Antioquia	05040	Anorí	6342	5850	12192	7181	6532	13713
11	05	Antioquia	05042	Santafé …	9867	9625	19492	10828	10415	21243

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
1	05	Antioquia	05001	Medellín	1985	692494	757683	1450177
2	05	Antioquia	05001	Medellín	1993	789722	906640	1696362
3	05	Antioquia	05001	Medellín	2005	951866	1094475	2046341
4	05	Antioquia	05001	Medellín	2018	1140658	1286471	2427129
5	05	Antioquia	05002	Abejorral	1985	14066	13569	27635
6	05	Antioquia	05002	Abejorral	1993	13133	12732	25865
7	05	Antioquia	05002	Abejorral	2005	11530	11412	22942
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367
9	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851
10	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804
11	05	Antioquia	05004	Abriaquí	2005	1392	1327	2719
12	05	Antioquia	05004	Abriaquí	2018	1452	1243	2695
13	05	Antioquia	05021	Alejandr…	1985	2520	2444	4964
14	05	Antioquia	05021	Alejandr…	1993	2468	2413	4881
15	05	Antioquia	05021	Alejandr…	2005	2336	2388	4724
16	05	Antioquia	05021	Alejandr…	2018	2320	2337	4657
17	05	Antioquia	05030	Amagá	1985	11380	11161	22541
18	05	Antioquia	05030	Amagá	1993	12266	12295	24561
19	05	Antioquia	05030	Amagá	2005	13439	13682	27121
20	05	Antioquia	05030	Amagá	2018	14862	15365	30227
21	05	Antioquia	05031	Amalfi	1985	9565	8810	18375

Variables de identificación

```
reshape long THombres TMujeres Total, i(DP DPNOM DPMP MPIO) j(AÑO)
```

Variable nueva con el atributo que estaba en los nombres de otras variables

Lo estudiaremos con detalle en la clase 9

7. Problemas en el ordenamiento de conjuntos de datos

3. Variables almacenadas en filas y columnas

	DP	DPNOM	DPMP	MPIO	Tipo	Total1985	Total1993	Total2005	Total2018
1	05	Antioquia	05001	Medellín	General	1450177	1696362	2046341	2427129
2	05	Antioquia	05001	Medellín	Hombres	692494	789722	951866	1140658
3	05	Antioquia	05001	Medellín	Mujeres	757683	906640	1094475	1286471
4	05	Antioquia	05002	Abejorral	General	27635	25865	22942	20367
5	05	Antioquia	05002	Abejorral	Hombres	14066	13133	11530	10534
6	05	Antioquia	05002	Abejorral	Mujeres	13569	12732	11412	9833
7	05	Antioquia	05004	Abriaquí	General	2851	2804	2719	2695
8	05	Antioquia	05004	Abriaquí	Hombres	1477	1605	1392	1452
9	05	Antioquia	05004	Abriaquí	Mujeres	1374	1199	1327	1243
10	05	Antioquia	05021	Alejandría	General	4964	4881	4724	4657
11	05	Antioquia	05021	Alejandría	Hombres	2520	2468	2336	2320
12	05	Antioquia	05021	Alejandría	Mujeres	2444	2413	2388	2337
13	05	Antioquia	05030	Amagá	General	22541	24561	27121	30227
14	05	Antioquia	05030	Amagá	Hombres	11380	12266	13439	14862
15	05	Antioquia	05030	Amagá	Mujeres	11161	12295	13682	15365

```
reshape long Total, i(DP DPNOM DPMP MPIO Tipo) j(AÑO)
reshape wide Total, i(DP DPNOM DPMP MPIO AÑO) j(Tipo) string
```

Lo estudiaremos con detalle en la clase 9

	DP	DPNOM	DPMP	MPIO	AÑO	THombres	TMujeres	Total
1	05	Antioquia	05001	Medellín	1985	692494	757683	1450177
2	05	Antioquia	05001	Medellín	1993	789722	906640	1696362
3	05	Antioquia	05001	Medellín	2005	951866	1094475	2046341
4	05	Antioquia	05001	Medellín	2018	1140658	1286471	2427129
5	05	Antioquia	05002	Abejorral	1985	14066	13569	27635
6	05	Antioquia	05002	Abejorral	1993	13133	12732	25865
7	05	Antioquia	05002	Abejorral	2005	11530	11412	22942
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367
9	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851
10	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804
11	05	Antioquia	05004	Abriaquí	2005	1392	1327	2719
12	05	Antioquia	05004	Abriaquí	2018	1452	1243	2695
13	05	Antioquia	05021	Alejandr…	1985	2520	2444	4964
14	05	Antioquia	05021	Alejandr…	1993	2468	2413	4881
15	05	Antioquia	05021	Alejandr…	2005	2336	2388	4724
16	05	Antioquia	05021	Alejandr…	2018	2320	2337	4657
17	05	Antioquia	05030	Amagá	1985	11380	11161	22541
18	05	Antioquia	05030	Amagá	1993	12266	12295	24561
19	05	Antioquia	05030	Amagá	2005	13439	13682	27121
20	05	Antioquia	05030	Amagá	2018	14862	15365	30227
21	05	Antioquia	05031	Amalfi	1985	9565	8810	18375

7. Problemas en el ordenamiento de conjuntos de datos

4. Varias unidades de observación en una misma tabla

Variables de identificación

Datos que cambian en el tiempo

Datos que NO cambian en el tiempo

	DP	DPNOM	DPMP	MPIO	AÑO	TotalHombres	TotalMujeres	TotalGeneral	MPIO_NAREA	AÑO_CREACION
1	05	Antioquia	05001	Medellín	1993	789722	906640	1696362	374.83	1965
2	05	Antioquia	05001	Medellín	2018	1140658	1286471	2427129	374.83	1965
3	05	Antioquia	05001	Medellín	1985	692494	757683	1450177	374.83	1965
4	05	Antioquia	05001	Medellín	2005	951866	1094475	2046341	374.83	1965
5	05	Antioquia	05002	Abejorral	2005	11530	11412	22942	507.13	1814
6	05	Antioquia	05002	Abejorral	1985	14066	13569	27635	507.13	1814
7	05	Antioquia	05002	Abejorral	1993	13133	12732	25865	507.13	1814
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367	507.13	1814
9	05	Antioquia	05004	Abriaquí	2005	1392	1327	2719	296.96	1912
10	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851	296.96	1912
11	05	Antioquia	05004	Abriaquí	2018	1452	1243	2695	296.96	1912
12	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804	296.96	1912
13	05	Antioquia	05021	Alejandría	2018	2320	2337	4657	128.93	1907
14	05	Antioquia	05021	Alejandría	1993	2468	2413	4881	128.93	1907
15	05	Antioquia	05021	Alejandría	1985	2520	2444	4964	128.93	1907
16	05	Antioquia	05021	Alejandría	2005	2336	2388	4724	128.93	1907

- En el análisis de datos se debe identificar si hay variables que no corresponden con la unidad de observación de la tabla.
- Es mejor crear varias tablas, cada una con variables de medición relacionadas con su propia unidad de observación.
- Estas diferentes tablas estarán relacionadas. Las relaciones pueden ser “muchos a uno” o “uno a uno”

7. Problemas en el ordenamiento de conjuntos de datos

4. Varias unidades de observación en una misma tabla

Relación muchos a uno

Variables llave

	DP	DPNOM	DPMP	MPIO	AÑO	TotalHombres	TotalMujeres	TotalGeneral
1	05	Antioquia	05001	Medellín	1993	789722	906640	1696362
2	05	Antioquia	05001	Medellín	2018	1140658	1286471	2427129
3	05	Antioquia	05001	Medellín	1985	692494	757683	1450177
4	05	Antioquia	05001	Medellín	2005	951866	1094475	2046341
5	05	Antioquia	05002	Abejorral	2005	11530	11412	22942
6	05	Antioquia	05002	Abejorral	1985	14066	13569	27635
7	05	Antioquia	05002	Abejorral	1993	13133	12732	25865
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367
9	05	Antioquia	05004	Abriaquí	2005	1392	1327	2719
10	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851
11	05	Antioquia	05004	Abriaquí	2018	1452	1243	2695
12	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804

Varias observaciones por municipio

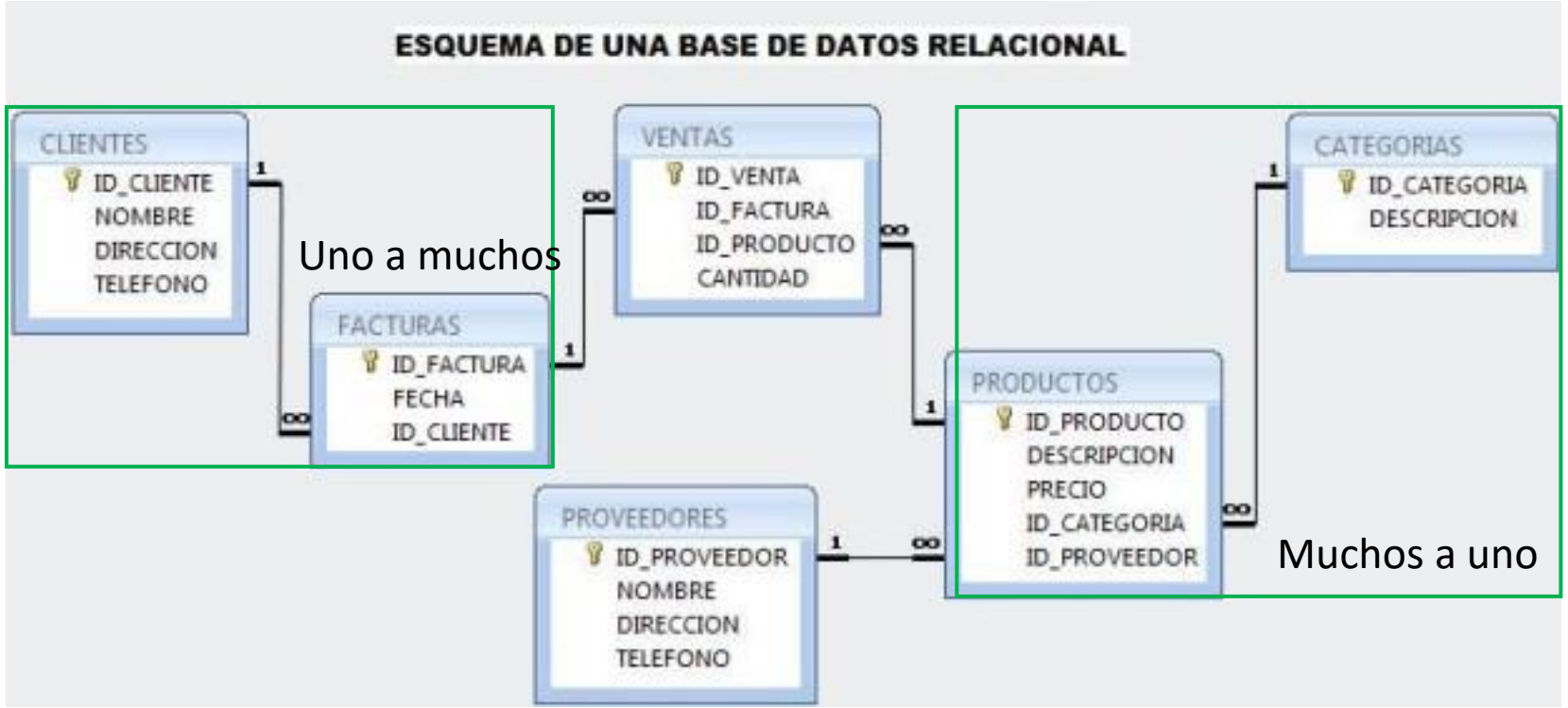
	DP	DPMP	MPIO_NAREA	AÑO_CREACION
1	05	05001	374.83	1965
2	05	05002	507.13	1814
3	05	05004	296.96	1912
4	05	05021	128.93	1907
5	05	05030	84.13	1912
6	05	05031	1209.13	1843
7	05	05034	402.50	1853
8	05	05036	81.88	1896
9	05	05038	338.67	1821
10	05	05040	1413.77	1821

Una observación por municipio

7. Problemas en el ordenamiento de conjuntos de datos

4. Varias unidades de observación en una misma tabla

Inspiración en las bases de datos relacionales



La lógica relacional lleva a evitar evaluar relaciones “muchos a muchos”

7. Problemas en el ordenamiento de conjuntos de datos

4. Varias unidades de observación en una misma tabla

Eventualmente se requiere unir tablas con diferentes unidades de observación para hacer análisis estadístico

Llamado a hacerlo con conciencia

Variables llave

	DP	DPNOM	DPMP	MPIO	AÑO	TotalHombres	TotalMujeres	TotalGeneral
1	05	Antioquia	05001	Medellín	1993	789722	906640	1696362
2	05	Antioquia	05001	Medellín	2018	1140658	1286471	2427129
3	05	Antioquia	05001	Medellín	1985	692494	757683	1450177
4	05	Antioquia	05001	Medellín	2005	951866	1094475	2046341
5	05	Antioquia	05002	Abejorral	2005	11530	11412	22942
6	05	Antioquia	05002	Abejorral	1985	14066	13569	27635
7	05	Antioquia	05002	Abejorral	1993	13133	12732	25865
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367
9	05	Antioquia	05004	Abriaquí	2005	1392	1327	2719
10	05	Antioquia	05004	Abriaquí	1985	1477	1374	2851
11	05	Antioquia	05004	Abriaquí	2018	1452	1243	2695
12	05	Antioquia	05004	Abriaquí	1993	1605	1199	2804

	DP	DPMP	MPIO_NAREA	AÑO_CREACION
1	05	05001	374.83	1965
2	05	05002	507.13	1814
3	05	05004	296.96	1912
4	05	05021	128.93	1907
5	05	05030	84.13	1912
6	05	05031	1209.13	1843
7	05	05034	402.50	1853
8	05	05036	81.88	1896
9	05	05038	338.67	1821
10	05	05040	1413.77	1821

```
merge m:1 DP DPMP using area_creacion
```

Lo estudiaremos con detalle en la clase 10

Relación muchos a uno

Universidad de los Andes
Facultad de Economía

8. Operaciones con datos

```
gen censo1985=1 if AÑO==1985
```

3. Filtrar observaciones

1. Organizar observaciones

2. Crear variables

	DP	DPNOM	DPMP	MPIO	AÑO	TotalHombres	TotalMujeres	TotalGeneral	censo1985
1	05	Antioquia	05001	Medellín	1985	692494	757683	1696362	1
2	05	Antioquia	05001	Medellín	1993	789722	906640	2427129	.
3	05	Antioquia	05001	Medellín	2005	951866	1094475	1450177	.
4	05	Antioquia	05001	Medellín	2018	1140658	1286471	2046341	.
5	05	Antioquia	05002	Abejorral	1985	14066	13569	22942	1
6	05	Antioquia	05002	Abejorral	1993	13133	12732	27635	.
7	05	Antioquia	05002	Abejorral	2005	11530	11412	25865	.
8	05	Antioquia	05002	Abejorral	2018	10534	9833	20367	.
9	05	Antioquia	05004	Abriaquí	1985	1477	1374	2719	1
10	05	Antioquia	05004	Abriaquí	1993	1605	1199	2851	.
11	05	Antioquia	05004	Abriaquí	2005	1392	1327	2695	.
12	05	Antioquia	05004	Abriaquí	2018	1452	1243	2804	.

```
gen TotalGeneral= TotalHombres+TotalMujeres
```

Universidad de los Andes
Facultad de Economía

8. Operaciones con datos

3.1. Unir variables

	DP	DPNOM	DPMP	MPIO	AÑO	TotalHombres	TotalMujeres
1	05	Antioquia	05001	Medellín	1985	692494	757683
2	05	Antioquia	05001	Medellín	1993	789722	906640
3	05	Antioquia	05001	Medellín	2005	951866	1094475
4	05	Antioquia	05001	Medellín	2018	1140658	1286471
5	05	Antioquia	05002	Abejorral	1985	14066	13569
6	05	Antioquia	05002	Abejorral	1993	13133	12732
7	05	Antioquia	05002	Abejorral	2005	11530	11412
8	05	Antioquia	05002	Abejorral	2018	10534	9833
9	05	Antioquia	05004	Abriaquí	1985	1477	1374

	DP	DPMP	MPIO_NAREA	AÑO_CREACION
1	05	05001	374.83	1965
2	05	05002	507.13	1814
3	05	05004	296.96	1912
4	05	05021	128.93	1907
5	05	05030	84.13	1912
6	05	05031	1209.13	1843
7	05	05034	402.50	1853
8	05	05036	81.88	1896
9	05	05038	338.67	1821
10	05	05040	1413.77	1821

```
merge m:1 DP DPMP using area_creacion
```

3.2. Unir observaciones

217	05	Antioquia	05318	Guarne	1985	12797	12754
218	05	Antioquia	05318	Guarne	1993	16559	16490
219	05	Antioquia	05318	Guarne	2005	21485	21934
220	05	Antioquia	05318	Guarne	2018	27308	27813
221	05	Antioquia	05321	Guatapé	1985	2151	2162
222	05	Antioquia	05321	Guatapé	1993	2675	2620
223	05	Antioquia	05321	Guatapé	2005	3352	3407
224	05	Antioquia	05321	Guatapé	2018	4097	4266

```
append guarne_guatape
```

8. Operaciones con datos

	DP	DPNOM	DPMP	MPIO	TotalHo~1985	TotalMu~1985	TotalGe~1985	TotalHo~1993	TotalMu~1993
1	05	Antioquia	05001	Medellín	692494	757683	1450177	789722	906640
2	05	Antioquia	05002	Abejorral	14066	13569	27635	13133	12732
3	05	Antioquia	05004	Abriaquí	1477	1374	2851	1605	1199
4	05	Antioquia	05021	Alejandría	2520	2444	4964	2468	2413
5	05	Antioquia	05030	Amagá	11380	11161	22541	12266	12295
6	05	Antioquia	05031	Amalfi	9565	8810	18375	10416	9910
7	05	Antioquia	05034	Andes	20513	19892	40405	20957	20247
8	05	Antioquia	05036	Angelópolis	3204	3003	6207	3182	2904
9	05	Antioquia	05038	Angostura	6798	6856	13654	6638	6490
10	05	Antioquia	05040	Anorí	6342	5850	12192	7181	6532

```
reshape long T* , i(DP DPNOM DPMP MPIO) j(AÑO)
```

4. Cambiar la estructura

	DP	DPNOM	DPMP	MPIO	AÑO	TotalHombres	TotalMujeres
1	05	Antioquia	05001	Medellín	1985	692494	757683
2	05	Antioquia	05001	Medellín	1993	789722	906640
3	05	Antioquia	05001	Medellín	2005	951866	1094475
4	05	Antioquia	05001	Medellín	2018	1140658	1286471
5	05	Antioquia	05002	Abejorral	1985	14066	13569
6	05	Antioquia	05002	Abejorral	1993	13133	12732
7	05	Antioquia	05002	Abejorral	2005	11530	11412
8	05	Antioquia	05002	Abejorral	2018	10534	9833
9	05	Antioquia	05004	Abriaquí	1985	1477	1374
10	05	Antioquia	05004	Abriaquí	1993	1605	1199
11	05	Antioquia	05004	Abriaquí	2005	1392	1327
12	05	Antioquia	05004	Abriaquí	2018	1452	1243

5. Contraer

```
collapse (sum) T*, by(DP DPNOM AÑO)
```

	DP	DPNOM	AÑO	TotalHombres	TotalMujeres	TotalGeneral
1	05	Antioquia	1985	1977354	2017630	3994984
2	05	Antioquia	1993	2242964	2364910	4607874
3	05	Antioquia	2005	2630787	2830056	5460843
4	05	Antioquia	2018	3094159	3312943	6407102
5	08	Atlántico	1985	653582	678209	1331791
6	08	Atlántico	1993	816741	864829	1681570
7	08	Atlántico	2005	1024962	1089174	2114136