# Nombres de bebés mas populares para Puerto Rico

Data del Seguro Social de EEUU

In [1]:
```
import pandas as pd
```

In [3]:
```
names1998= pd.read_csv('nb_1998.txt', names=['name', 'sex', 'births'])
```

In [4]:
```
names1998
```

Out[4]:
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 200 entries, 0 to 199
Data columns:
name      200  non-null values
sex       200  non-null values
births    200  non-null values
dtypes: int64(1), object(2)
```

In [5]:
```
names1998.groupby('sex').births.sum()
```

Out[5]:
```
sex
F      11337
M      21511
Name: births
```

In [8]:
```
names1998.groupby('sex')
```

Out[8]:
```
<pandas.core.groupby.DataFrameGroupBy at 0x3198850>
```

In [14]:
```
years = range(1998, 2012)
pieces = []
columns = ['name', 'sex', 'births']

for year in years:
    path = 'nb_%d.txt' % year
    frame = pd.read_csv(path, names=columns)
    frame['year'] = year
    pieces.append(frame)
    # Concatenate everything into a single DataFrame
    names = pd.concat(pieces, ignore_index=True)
```

In [15]:
```
names
```

Out[15]:
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2800 entries, 0 to 2799
Data columns:
name      2800  non-null values
sex       2800  non-null values
births    2800  non-null values
year      2800  non-null values
dtypes: int64(2), object(2)
```

In [46]:
```
names.groupby(['sex']).births.sum()
```

Out[46]:
```
sex
F      118227
M      219726
Name: births
```

In [40]:
```
names.groupby(['year'],sort=True).sum()
```

Out[40]:

| | births |
|---|---|
| **year** | |
| **1998** | 32848 |
| **1999** | 31895 |
| **2000** | 30931 |
| **2001** | 28830 |
| **2002** | 26441 |
| **2003** | 25141 |
| **2004** | 24288 |
| **2005** | 23516 |
| **2006** | 22148 |
| **2007** | 20761 |
| **2008** | 19288 |
| **2009** | 18484 |
| **2010** | 17188 |
| **2011** | 16194 |

In [60]:
```
names.groupby(['year','sex'],sort=True).sum()
```

Out[60]:

| | | births |
|---|---|---|
| **year** | **sex** | |
| **1998** | **F** | 11337 |
| | **M** | 21511 |
| **1999** | **F** | 10943 |
| | **M** | 20952 |
| **2000** | **F** | 10496 |
| | **M** | 20435 |
| **2001** | **F** | 10108 |
| | **M** | 18722 |
| **2002** | **F** | 9220 |
| | **M** | 17221 |
| **2003** | **F** | 8690 |
| | **M** | 16451 |
| **2004** | **F** | 8294 |
| | **M** | 15994 |
| **2005** | **F** | 8238 |
| | **M** | 15278 |
| **2006** | **F** | 7728 |
| | **M** | 14420 |
| **2007** | **F** | 7306 |
| | **M** | 13455 |
| **2008** | **F** | 6759 |
| | **M** | 12529 |
| **2009** | **F** | 6624 |
| | **M** | 11860 |
| **2010** | **F** | 6374 |
| | **M** | 10814 |
| **2011** | **F** | 6110 |
| | **M** | 10084 |

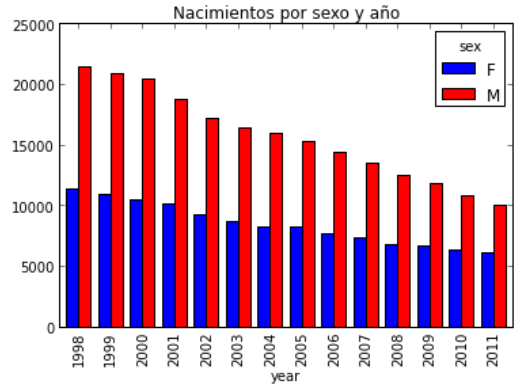
In [186]:
```
total_births = names.pivot_table('births', rows='year',cols='sex', aggfunc=sum)
total_births
```

Out[186]:

| sex | F | M |
|---|---|---|
| year | | |
| **1998** | 11337 | 21511 |
| **1999** | 10943 | 20952 |
| **2000** | 10496 | 20435 |
| **2001** | 10108 | 18722 |
| **2002** | 9220 | 17221 |
| **2003** | 8690 | 16451 |
| **2004** | 8294 | 15994 |
| **2005** | 8238 | 15278 |
| **2006** | 7728 | 14420 |
| **2007** | 7306 | 13455 |
| **2008** | 6759 | 12529 |
| **2009** | 6624 | 11860 |
| **2010** | 6374 | 10814 |
| **2011** | 6110 | 10084 |

In [291]:
```
total_births.plot(kind='bar',title=u'Nacimientos por sexo y año')
#podemos decir que hay mas variedad de nombres entre las niñas?
```

Out[291]: <matplotlib.axes.AxesSubplot at 0x23f2d7d0>




In [103]:
```
#Función para buscar los 10 mas populares
def get_top10(group):
    return group.sort_index(by='births', ascending=False)[:10]

grouped = names.groupby(['year', 'sex'])
top10 = grouped.apply(get_top10)
top10
```

Out[103]:
```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 280 entries, (1998, F, 100) to (2011, M, 2609)
Data columns:
name      280  non-null values
sex       280  non-null values
births    280  non-null values
year      280  non-null values
dtypes: int64(2), object(2)
```

In [104]:
```
boys = top10[top10.sex == 'M']
```

In [168]:
```
girls = top10[top10.sex == 'F']
```

In [180]:
```
#Todos estos nombres alguna vez llegaron a estar entre los 10 mas populares para niñas
girls_table = girls.pivot_table('births', rows='name',cols='year', aggfunc=sum,fill_value = 0, margins = True)
girls_table.sort_index(by=['All'], ascending=[False])
```

Out[180]:

| year | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| name | | | | | | | | | | | | | | | |
| All | 3754 | 3971 | 3955 | 3934 | 3488 | 3485 | 3287 | 3230 | 3027 | 2758 | 2592 | 2633 | 2469 | 2306 | 44889 |
| Alondra | 344 | 445 | 551 | 582 | 611 | 660 | 621 | 722 | 577 | 464 | 406 | 350 | 294 | 256 | 6883 |
| Paola | 724 | 761 | 764 | 718 | 641 | 492 | 414 | 367 | 321 | 285 | 253 | 0 | 0 | 0 | 5740 |
| Gabriela | 447 | 489 | 461 | 426 | 412 | 470 | 411 | 373 | 315 | 325 | 235 | 215 | 195 | 166 | 4940 |
| Andrea | 256 | 342 | 351 | 370 | 316 | 292 | 304 | 263 | 270 | 253 | 220 | 236 | 182 | 220 | 3875 |
| Adriana | 251 | 311 | 316 | 321 | 286 | 355 | 322 | 291 | 301 | 256 | 222 | 212 | 169 | 0 | 3613 |
| Valeria | 0 | 270 | 271 | 305 | 305 | 273 | 263 | 253 | 294 | 236 | 230 | 288 | 247 | 227 | 3462 |
| Genesis | 500 | 439 | 401 | 433 | 320 | 303 | 254 | 264 | 196 | 0 | 0 | 0 | 0 | 0 | 3110 |
| Amanda | 0 | 0 | 0 | 0 | 0 | 186 | 185 | 210 | 228 | 227 | 231 | 242 | 212 | 178 | 1899 |
| Mia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 217 | 223 | 211 | 323 | 386 | 358 | 1718 |
| Alanis | 0 | 0 | 0 | 0 | 0 | 262 | 282 | 267 | 308 | 257 | 0 | 198 | 0 | 0 | 1574 |
| Camila | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 232 | 326 | 283 | 298 | 249 | 1388 |
| Nicole | 392 | 370 | 333 | 255 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1350 |
| Kamila | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 258 | 286 | 320 | 292 | 1156 |
| Natalia | 0 | 0 | 251 | 310 | 213 | 192 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 966 |
| Stephanie | 273 | 272 | 256 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 801 |
| Maria | 291 | 0 | 0 | 214 | 193 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 698 |
| Ariana | 0 | 0 | 0 | 0 | 0 | 0 | 231 | 220 | 0 | 0 | 0 | 0 | 0 | 0 | 451 |
| Sofia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 166 | 166 | 332 |
| Nashaly | 276 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 276 |
| Kiara | 0 | 272 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 272 |
| Victoria | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 194 | 194 |
| Fabiola | 0 | 0 | 0 | 0 | 191 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 191 |

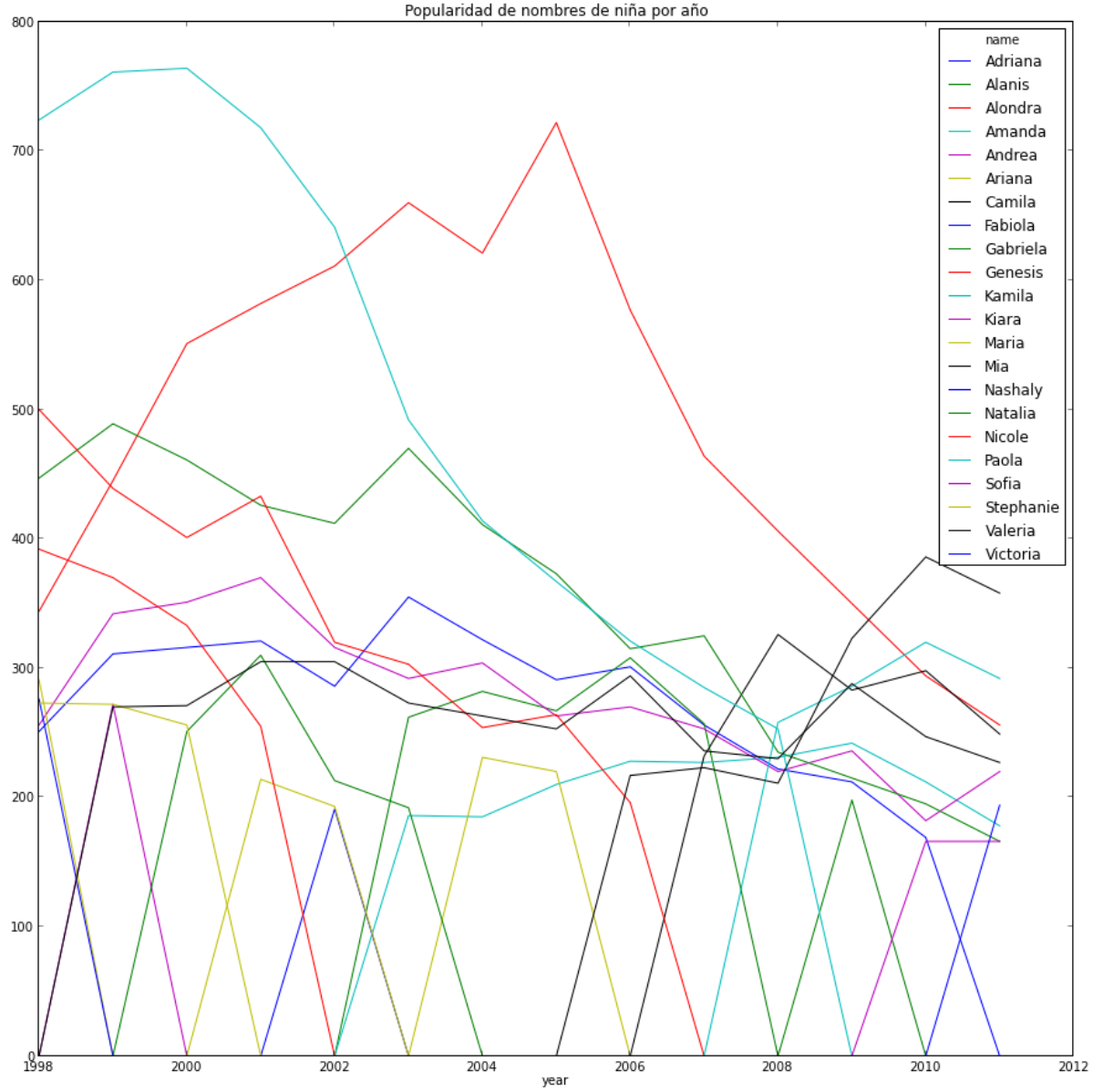
In [230]:
```
#Todos estos nombres alguna vez llegaron a estar entre los 10 mas populares para niños
boys_table = boys.pivot_table('births', rows='name',cols='year', aggfunc=sum,fill_value = 0, margins = True)
boys_table.sort_index(by=['All'], ascending=[False])
```

Out[230]:

| year | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| name | | | | | | | | | | | | | | | |
| All | 7969 | 7302 | 7187 | 6217 | 5614 | 5489 | 5138 | 4868 | 4661 | 4463 | 4146 | 3796 | 3500 | 3314 | 73664 |
| Luis | 1472 | 1388 | 1374 | 1152 | 1055 | 1013 | 928 | 880 | 740 | 722 | 630 | 550 | 520 | 445 | 12869 |
| Jose | 1458 | 1345 | 1253 | 1027 | 935 | 874 | 825 | 673 | 551 | 543 | 431 | 405 | 353 | 260 | 10933 |
| Angel | 829 | 774 | 802 | 726 | 722 | 730 | 739 | 647 | 633 | 575 | 546 | 486 | 505 | 404 | 9118 |
| Carlos | 843 | 749 | 732 | 741 | 611 | 564 | 518 | 544 | 493 | 390 | 397 | 357 | 284 | 253 | 7476 |
| Kevin | 847 | 752 | 762 | 614 | 497 | 477 | 406 | 314 | 245 | 0 | 0 | 0 | 0 | 0 | 4914 |
| Gabriel | 442 | 442 | 444 | 421 | 430 | 443 | 354 | 361 | 288 | 300 | 243 | 252 | 0 | 0 | 4420 |
| Diego | 0 | 0 | 0 | 0 | 340 | 402 | 409 | 478 | 655 | 545 | 514 | 389 | 316 | 299 | 4347 |
| Sebastian | 0 | 0 | 0 | 0 | 337 | 355 | 348 | 395 | 457 | 439 | 459 | 386 | 372 | 405 | 3953 |
| Juan | 508 | 435 | 447 | 0 | 343 | 294 | 324 | 293 | 0 | 0 | 0 | 0 | 0 | 0 | 2644 |
| Christian | 496 | 429 | 460 | 417 | 344 | 337 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2483 |
| Adrian | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 283 | 296 | 332 | 367 | 434 | 311 | 343 | 2366 |
| Bryan | 524 | 521 | 462 | 387 | 0 | 0 | 287 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2181 |
| Ian | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 257 | 255 | 309 | 377 | 388 | 1586 |
| Michael | 550 | 467 | 0 | 357 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1374 |
| Yadiel | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 303 | 360 | 304 | 228 | 0 | 0 | 1195 |
| Jean | 0 | 0 | 451 | 375 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 826 |
| Dylan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 247 | 243 | 490 |
| Jayden | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 274 | 274 |
| Derek | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 215 | 0 | 215 |

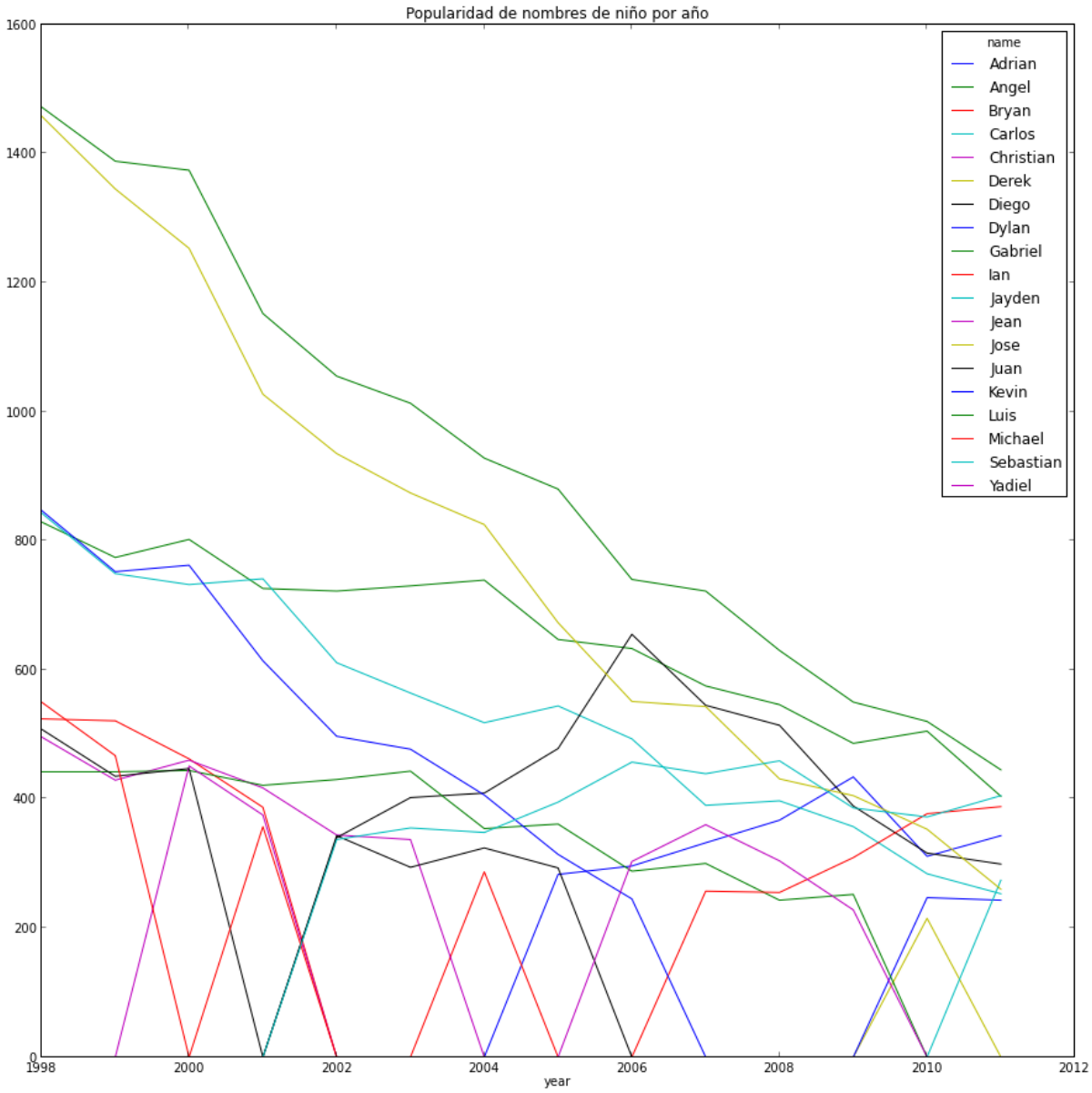
In [224]:
```
girls_chart = girls.pivot_table('births', rows='year',cols='name', aggfunc=sum, fill_value=0)
girls_chart.plot(figsize=(15, 15), grid=False, title=u'Popularidad de nombres de niña por año')
```

Out[224]: <matplotlib.axes.AxesSubplot at 0x23075f10>



In [239]:
```
boys_chart = boys.pivot_table('births', rows='year',cols='name', aggfunc=sum, fill_value=0)
boys_chart.plot(figsize=(15, 15), grid=False, title=u'Popularidad de nombres de niño por año')
```

Out[239]: <matplotlib.axes.AxesSubplot at 0x266b7c90>



In [244]:
```
sname = top10[top10.name == 'Paola']
sname_chart = sname.pivot_table('births', rows='year',cols='name', aggfunc=sum, fill_value=0)
sname_chart.plot(figsize=(15, 15), grid=False, title=u'Popularidad de nombre individual')
```

Out[244]: <matplotlib.axes.AxesSubplot at 0x27940350>