

# Práctica 16 de abril

- Haga la lectura del dataset en el fichero “Canada.xls”. Use la Hoja del excel nombrada ‘Canada by Citizenship’. Este fichero contiene datos de migración hacia Canada desde diferentes países en los 5 continentes (periodo 1980-2013 ).

	Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	...	2004	2005	2006	2007	2008	2009	2010	2011	2012
0	Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	...	2978	3436	3009	2652	2111	1746	1758	2203	2635
1	Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	...	1450	1223	856	702	560	716	561	539	620
2	Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	...	3616	3626	4807	3623	4005	5393	4752	4325	3774
3	Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	...	0	0	1	0	0	0	0	0	0
4	Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	...	0	0	1	1	0	0	0	0	1

Haga una limpieza de los datos para facilitar la creación de las visualizaciones.

- remueva las columnas 'AREA', 'REG', 'DEV', 'Type', 'Coverage'.
- ponga nombres más adecuados a las columnas 'OdName', 'AreaName' y 'RegName'.
- asegurese de que todas las columnas (`DataFrame.columns`) corresponden a tipo 'string'.
- use el método `*set_index()*` del `dataFrame` para que los índices de las entradas sean los países.
- adicione una columna de total, en donde esta represente la suma del flujo total de migrantes hacia Canada desde los diferentes países a lo largo de los años.

	Continent	Region	DevName	1980	1981	1982	1983	1984	1985	1986	...	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Country																					
<b>Afghanistan</b>	Asia	Southern Asia	Developing regions	16	39	39	47	71	340	496	...	3436	3009	2652	2111	1746	1758	2203	2635	2004	58639
<b>Albania</b>	Europe	Southern Europe	Developed regions	1	0	0	0	0	0	1	...	1223	856	702	560	716	561	539	620	603	15699
<b>Algeria</b>	Africa	Northern Africa	Developing regions	80	67	71	69	63	44	69	...	3626	4807	3623	4005	5393	4752	4325	3774	4331	69439
<b>American Samoa</b>	Oceania	Polynesia	Developing regions	0	1	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	6
<b>Andorra</b>	Europe	Southern Europe	Developed regions	0	0	0	0	0	0	2	...	0	1	1	0	0	0	0	1	1	15

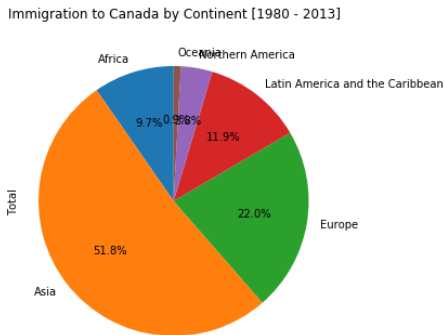
# Pie Charts

Genere un conjunto de datos que se encuentre agrupado por continente. El dataFrame de salida tendrá 6 filas y 35 columnas.  
sugerencia: aplique la función `sum()` a los datos agrupados para que el resultado del agrupamiento sea efectivamente un DataFrame (`pandas.core.frame.DataFrame`)

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	...	2005	2006	2007	2008	2009	2010	2011	2012	
Continent																				
Africa	3951	4363	3819	2671	2639	2650	3782	7494	7552	9894	...	27523	29188	28284	29890	34534	40892	35441	38083	3
Asia	31025	34314	30214	24696	27274	23850	28739	43203	47454	60256	...	159253	149054	133459	139894	141434	163845	146894	152218	15
Europe	39760	44802	42720	24638	22287	20844	24370	46698	54726	60893	...	35955	33053	33495	34692	35078	33425	26778	29177	2
Latin America and the Caribbean	13081	15215	16769	15427	13678	15171	21179	28471	21924	25060	...	24747	24676	26011	26547	26867	28818	27856	27173	2
Northern America	9378	10030	9074	7100	6661	6543	7074	7705	6469	6790	...	8394	9613	9463	10190	8995	8142	7677	7892	

# Pie Charts

Usando el método **plot** de pandas haga un gráfico tipo pastel del total de migración por cada continente, en el que incluya los porcentajes (usando el parámetro **autopct**). No olvide poner **título** a su gráfico. Ponga también el parámetro **startangle=90**



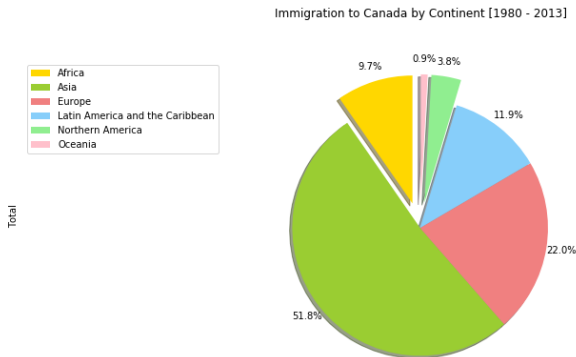
Haga una mejora del gráfico generado:

- Presente las etiquetas usando `plt.legend()`.
- Haga que los valores de los porcentajes se presenten en la parte exterior del gráfico usando el parámetro `'pctdistance'`.
- Use un esquema adecuado de colores usando el parámetro `'colors'`.
- Use el parámetro `'explode'` para hacer énfasis en los 3 continentes que representan una menor área.

Explique las mejoras que se pueden observar respecto al gráfico anterior. ¿Qué otras mejoras podría implementar? - No olvide poner título a su gráfico.

# Pie Charts

Para un listado de colores en **matplotlib** y su correspondiente valor  
>>>>>**for** name, hex **in** matplotlib.colors.cnames.items():





Una **bubble plot** es una variante del **scatter plot** que muestra 3 dimensiones de los datos (x, y, z). La posición en el plano de cada dato se representa con una burbuja mediante las variables 'x', 'y'; pero el tamaño de cada burbuja está dado por la tercera variable 'z' (conocido también como el peso). **Paso 1**

Obtener los datos para Argentina y Brasil. Convierta los años (1980:2013) en una columna (variable tipo 'int') e incorporelas en el dataframe (transpuesto).

# Burbujas

Country	Year	Afghanistan	Albania	Algeria	American Samoa	Andorra	Angola	Antigua and Barbuda	Argentina	Armenia	...	United States of America	Uruguay	Uzbekistan	Vanuatu	Venezuela (Bolivar Repu
0	1980	16	1	80	0	0	1	0	368	0	...	9378	128	0	0	1
1	1981	39	0	67	1	0	3	0	426	0	...	10030	132	0	0	1
2	1982	39	0	71	0	0	6	0	626	0	...	9074	146	0	0	1
3	1983	47	0	69	0	0	6	0	241	0	...	7100	105	0	0	1
4	1984	71	0	63	0	0	4	42	237	0	...	6661	90	0	0	1

## Paso 2

Haga normalización de cada columna al rango  $[0,1]$  usando la fórmula

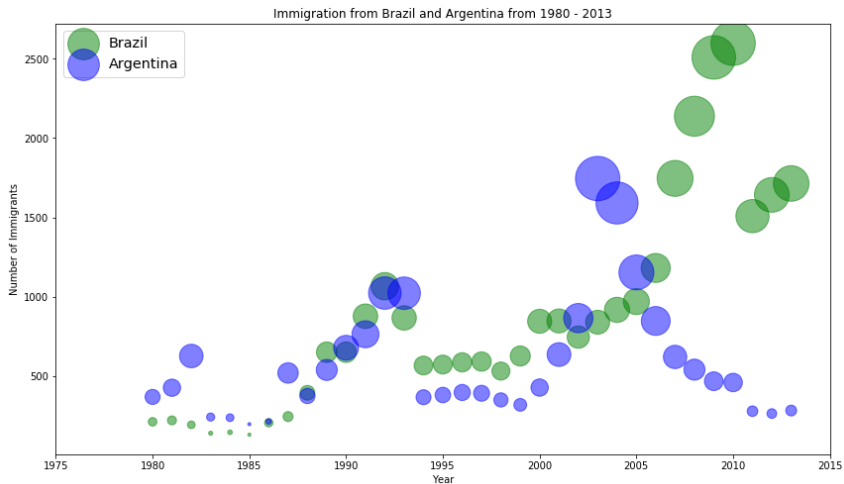
$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

## Paso 3

### Pintar los datos

- Incluir los dos gráficos de dispersión en una sola figura.
- En cada gráfico de **'scatter'** (**'df.plot(kind='scatter')'**) use el parámetro **'s'**. La entrada de este parámetro serán los pesos normalizados en el rango  $[0,1]$  obtenidos anteriormente. Para poder visualizar bien cada burbuja, haga **'s = norm\_values \* var\_rad + min\_rad'**, donde **'norm\_values'** son los valores normalizados, **'var\_rad'** será el valor de variación del radio de las burbujas y **'min\_rad'** será el radio mínimo para cada burbuja.

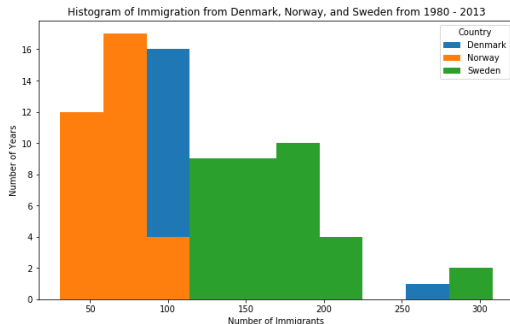
# Burbujas



- ¿Qué puede decir del anterior gráfico respecto a las tendencias de migración desde ambos países?
- Repita el proceso para los datos de migración desde los dos países que aportan mayor número de migrantes hacia Canada.

# Histogramas

- Cuál es la distribución de datos de inmigración desde Dinamarca, Noruega y Suecia para el período comprendido entre los años 1980 - 2013?
- Mejore, según su criterio, la visualización presentada por defecto.

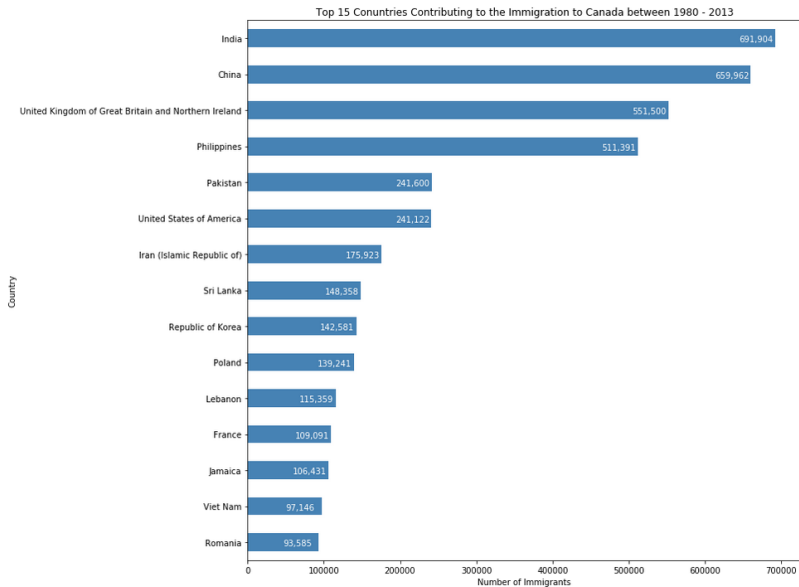


Encuentre los 15 países que más migración tuvieron hacia Canadá en el periodo de 1980-2013.

- Agregue color y legendas que permitan visualizar el continente que corresponde a cada barra horizontal.



# Barras horizontales



# Barras horizontales

Repita el proceso anterior, pero ahora hágalo para 3 períodos diferentes: 1980-1989, 1990-1999, 2000-2009; de manera que pueda visualizar la evolución de los datos.  
¿Qué conclusiones puede sacar?