

NAME: ANNAPOORNIMA S

225229101

Lab5. Pandas Concatenate, Merge and Join

```
In [2]: ▶ import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [3]: ▶ north_america=pd.read_csv("north_america_2000_2010.csv",index_col=0)
south_america=pd.read_csv("south_america_2000_2010.csv",index_col=0)
```

```
In [4]: ▶ north_america
```

```
Out[4]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Country											
Canada	1779.0	1771.0	1754.0	1740.0	1760.0	1747	1745.0	1741.0	1735	1701.0	1703.0
Mexico	2311.2	2285.2	2271.2	2276.5	2270.6	2281	2280.6	2261.4	2258	2250.2	2242.4
USA	1836.0	1814.0	1810.0	1800.0	1802.0	1799	1800.0	1798.0	1792	1767.0	1778.0

```
In [5]: ▶ south_america
```

```
Out[5]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Country											
Chile	2263	2242	2250	2235	2232	2157	2165	2128	2095	2074	2069.6

```
In [6]: ▶ !type north_america_2000_2010.csv
```

```
Country,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010
Canada,1779,1771,1754,1740,1760,1747,1745,1741,1735,1701,1703
Mexico,2311.2,2285.2,2271.2,2276.5,2270.6,2281,2280.6,2261.4,2258,2250.2,2242.4
USA,1836,1814,1810,1800,1802,1799,1800,1798,1792,1767,1778
```

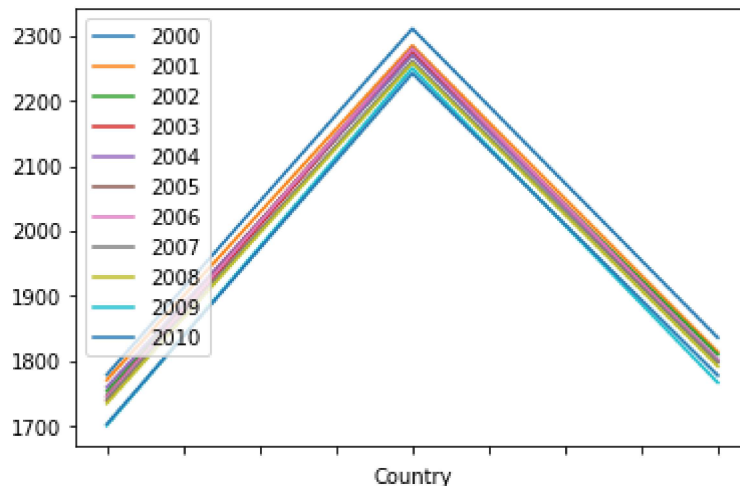
```
In [7]: ▶ !type south_america_2000_2010.csv
```

```
Country,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010
Chile,2263,2242,2250,2235,2232,2157,2165,2128,2095,2074,2069.6
```

Create line graphs for our yearly labor trends in north_america

In [9]: `north_america.plot()`

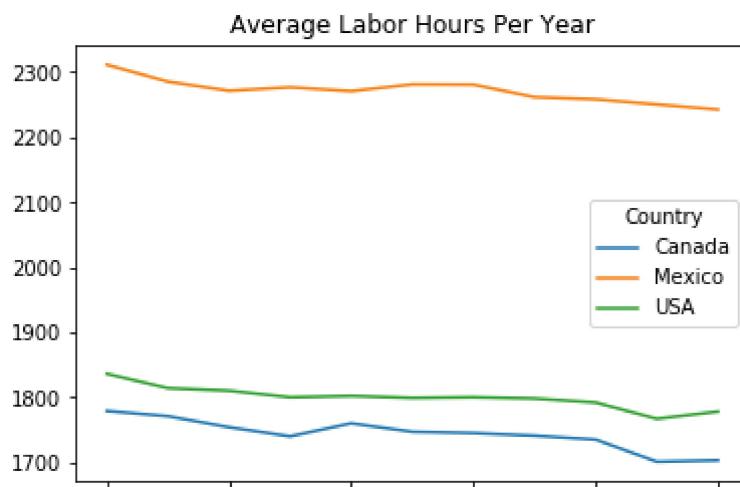
Out[9]: `<matplotlib.axes._subplots.AxesSubplot at 0x132862a5e10>`



Plot transposed line graph of north_america dataframe, with title "Average Labor Hours Per Year"

In [11]: `north_america.transpose().plot(title='Average Labor Hours Per Year')`

Out[11]: `<matplotlib.axes._subplots.AxesSubplot at 0x13286241278>`

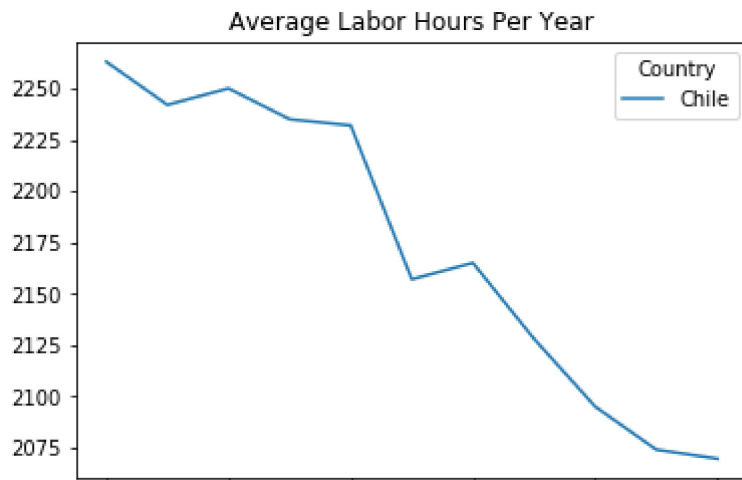


Similarly, plot transposed south_america dataframe with title "Average Labor Hours Per Year". Output chart is shown below

Concatenate America Data

Concatenate north_america and south_america dataframes and store result in a dataframe, americas

```
In [15]: ▶ south_america.transpose().plot(title='Average Labor Hours Per Year')
plt.show()
```



```
In [16]: ▶ americas=pd.concat([north_america,south_america])
americas
```

Out[16]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
--	------	------	------	------	------	------	------	------	------	------	------

Country

Canada	1779.0	1771.0	1754.0	1740.0	1760.0	1747	1745.0	1741.0	1735	1701.0	1703.0
Mexico	2311.2	2285.2	2271.2	2276.5	2270.6	2281	2280.6	2261.4	2258	2250.2	2242.4
USA	1836.0	1814.0	1810.0	1800.0	1802.0	1799	1800.0	1798.0	1792	1767.0	1778.0
Chile	2263.0	2242.0	2250.0	2235.0	2232.0	2157	2165.0	2128.0	2095	2074.0	2069.6

Load the additional files

```
In [18]: ▶ americas_dfs = [americas]
for year in range(2011, 2016):
    filename = "./americas_{}.csv".format(year)
    df = pd.read_csv(filename, index_col=0)
    americas_dfs.append(df)
```

In [19]: `americas_dfs[1]`

Out[19]: **2011**

Country	
Canada	1700.0
Chile	2047.4
Mexico	2250.2
USA	1786.0

In [20]: `americas_dfs[2]`

Out[20]: **2012**

Country	
Canada	1713.0
Chile	2024.0
Mexico	2225.8
USA	1789.0

In [21]: `americas=pd.concat(americas_dfs,axis=1)`
`americas.index.names=['country']`

Concatenate americas and americas_dfs dataframes and store result in americas

In [23]: `americas`

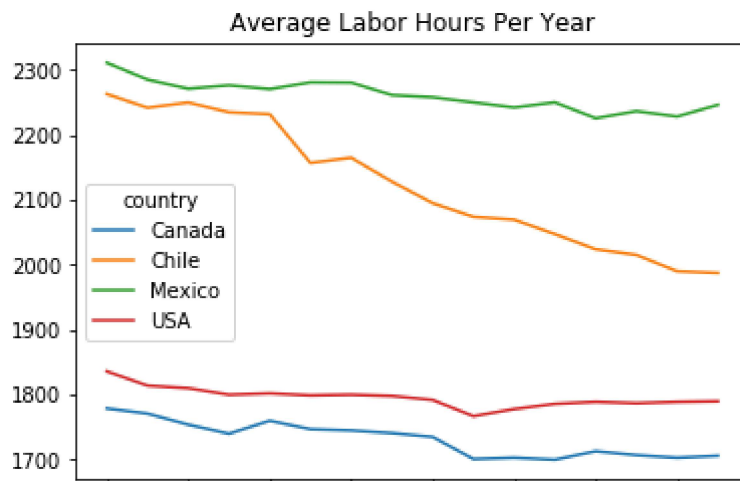
Out[23]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
--	------	------	------	------	------	------	------	------	------	------	------

country											
Canada	1779.0	1771.0	1754.0	1740.0	1760.0	1747	1745.0	1741.0	1735	1701.0	1703.0
Chile	2263.0	2242.0	2250.0	2235.0	2232.0	2157	2165.0	2128.0	2095	2074.0	2069.6
Mexico	2311.2	2285.2	2271.2	2276.5	2270.6	2281	2280.6	2261.4	2258	2250.2	2242.4
USA	1836.0	1814.0	1810.0	1800.0	1802.0	1799	1800.0	1798.0	1792	1767.0	1778.0

In [24]: `americas.transpose().plot(title='Average Labor Hours Per Year')`

Out[24]: `<matplotlib.axes._subplots.AxesSubplot at 0x132883b89b0>`



Appending data from other Continents

In [26]: `asia=pd.read_csv('asia_2000_2015.csv',index_col=0)`

In [27]: `asia`

Out[27]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Country													
Israel	2017	1979	1993	1974	1942	1931	1919	1931	1929	1927	1918	1920	1910
Japan	1821	1809	1798	1799	1787	1775	1784	1785	1771	1714	1733	1728	1745
Korea	2512	2499	2464	2424	2392	2351	2346	2306	2246	2232	2187	2090	2163
Russia	1982	1980	1982	1993	1993	1989	1998	1999	1997	1974	1976	1979	1982

```
In [28]: europe=pd.read_csv('europe_2000_2015.csv',index_col=0)
europe.head()
```

```
Out[28]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Country										
Austria	1807.4	1794.6	1792.2	1783.8	1786.8	1764.0	1746.2	1736.0	1728.5	1673.0
Belgium	1595.0	1588.0	1583.0	1578.0	1573.0	1565.0	1572.0	1577.0	1570.0	1548.0
Switzerland	1673.6	1635.0	1614.0	1626.8	1656.5	1651.7	1643.2	1632.7	1623.1	1614.9
Czech Republic	1896.0	1818.0	1816.0	1806.0	1817.0	1817.0	1799.0	1784.0	1790.0	1779.0
Germany	1452.0	1441.9	1430.9	1424.8	1422.2	1411.3	1424.7	1424.4	1418.4	1372.7

```
In [29]: south_pacific=pd.read_csv('south_pacific_2000_2015.csv',index_col=0)
south_pacific
```

```
Out[29]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Country											
Australia	1778.7	1736.7	1731.7	1735.8	1734.5	1729.2	1720.5	1712.5	1717.2	1690	169
New Zealand	1836.0	1825.0	1826.0	1823.0	1830.0	1815.0	1795.0	1774.0	1761.0	1740	175

Append asia, europe and south_pacific to americas dataframe and assign to new dataframe world

```
In [31]: world=americas.append([asia,europe,south_pacific])
```

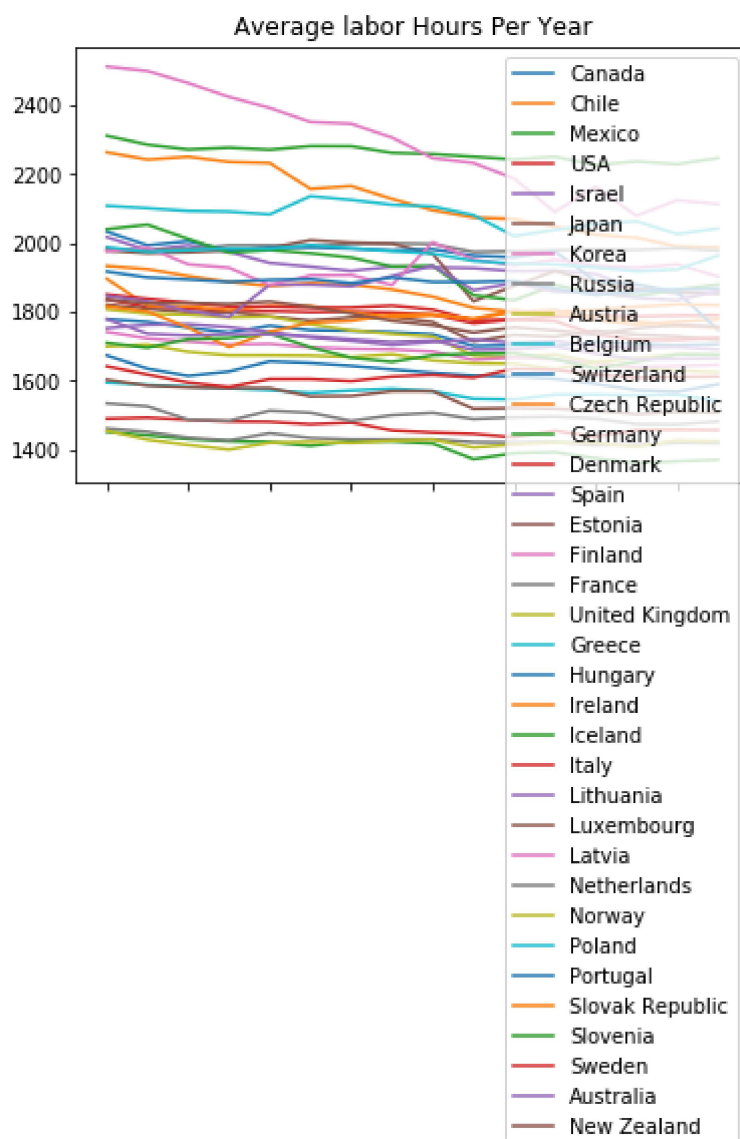
```
In [32]: world.index
```

```
Out[32]: Index(['Canada', 'Chile', 'Mexico', 'USA', 'Israel', 'Japan', 'Korea',
               'Russia', 'Austria', 'Belgium', 'Switzerland', 'Czech Republic',
               'Germany', 'Denmark', 'Spain', 'Estonia', 'Finland', 'France',
               'United Kingdom', 'Greece', 'Hungary', 'Ireland', 'Iceland', 'Italy',
               'Lithuania', 'Luxembourg', 'Latvia', 'Netherlands', 'Norway', 'Poland',
               'Portugal', 'Slovak Republic', 'Slovenia', 'Sweden', 'Australia',
               'New Zealand'],
              dtype='object')
```

Plot, transposed world dataframe

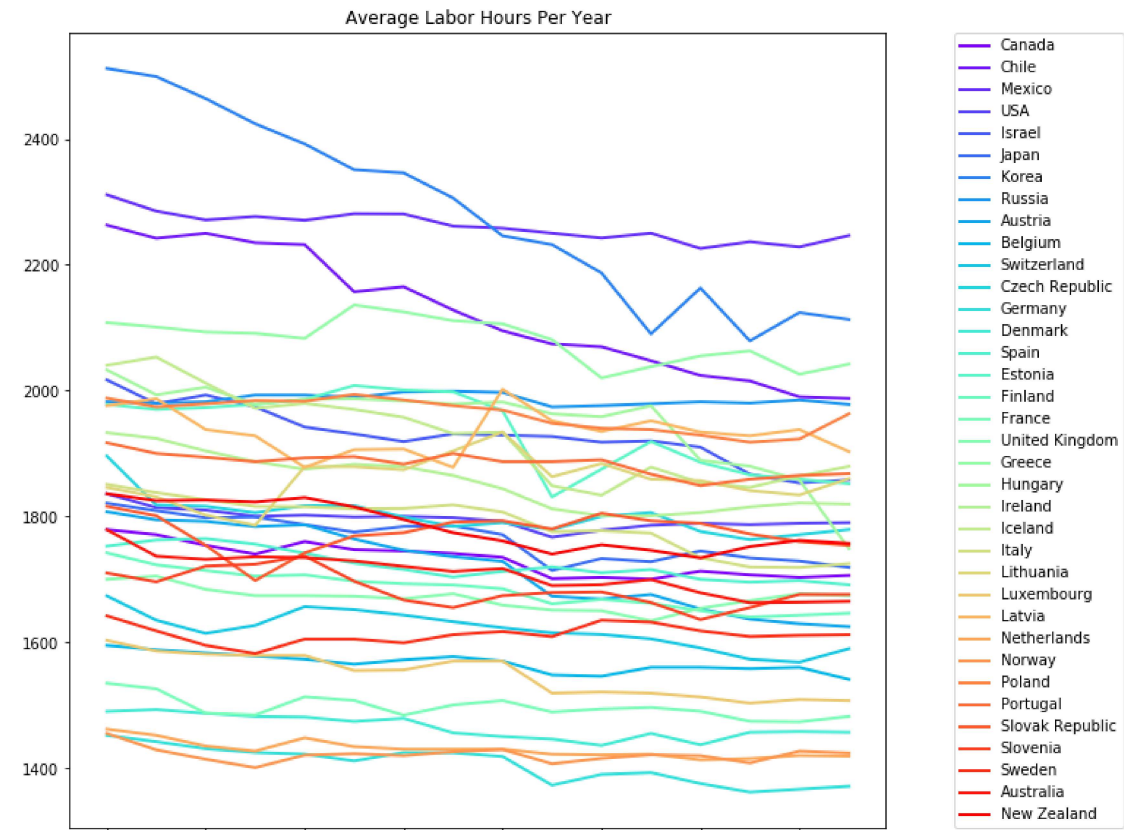
```
In [34]: world.transpose().plot(title='Average labor Hours Per Year')
```

```
Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x132884581d0>
```



let us customize this plot, so that country names appear outside the chart

```
In [36]: world.transpose().plot(figsize=(10,10),colormap='rainbow',linewidth=2,
plt.legend(loc='right',bbox_to_anchor=(1.3,0.5))
plt.show()
```



Merging Historical Labor Data

```
In [38]: historical = pd.read_csv('historical.csv', index_col=0)
historical.head()
```

Out[38]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...	1990	1
Country													
Australia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1779.5	1774
Austria	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	1
Belgium	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1662.9	1625
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1789.5	1767
Switzerland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	1673

5 rows × 50 columns


```
In [39]: ▶ print("World rows & columns: ", world.shape)
          print("Historical rows & columns: ", historical.shape)
```

```
World rows & columns: (36, 16)
Historical rows & columns: (39, 50)
```

Merge historical dataframe with world dataframe and store in a new variable, world_historical

```
In [41]: ▶ world_historical = pd.merge(historical, world, left_index=True, right_index=True)
```

Print size of world_historical dataframe

```
In [43]: ▶ print(world_historical.shape)
```

```
(36, 66)
```

Print top-5 of world_historical dataframe

```
In [45]: ▶ world_historical.head()
```

```
Out[45]:
```

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Chile	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Mexico	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
USA	1960.0	1975.5	1978.0	1980.0	1970.5	1992.5	1990.0	1962.0	1936.5	1947.0	...
Israel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

```
5 rows × 66 columns
```



Joining Historical Data

```
In [47]: world_historical = historical.join(world, how='right')
world_historical.head()
```

Out[47]:

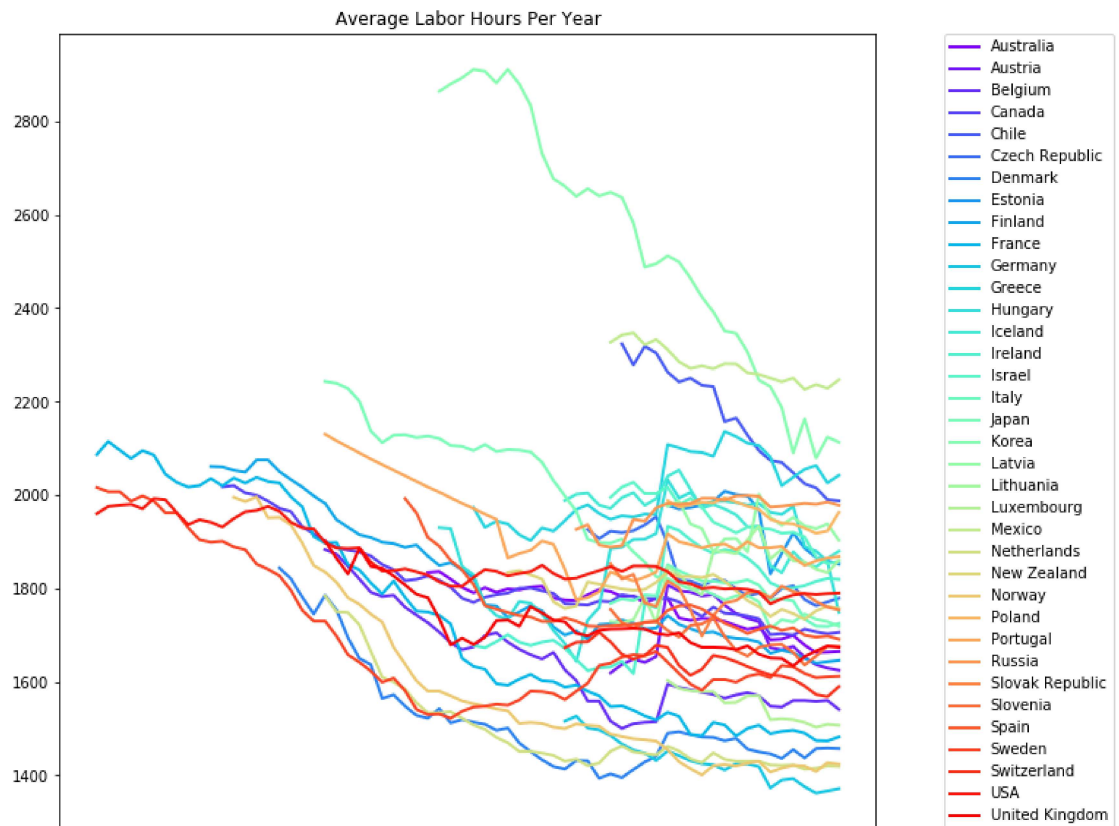
	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Chile	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Mexico	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
USA	1960.0	1975.5	1978.0	1980.0	1970.5	1992.5	1990.0	1962.0	1936.5	1947.0	...
Israel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

5 rows × 66 columns

Plot, transposed world_historical dataframe

```
In [49]: world_historical.sort_index(inplace=True)
```

```
In [50]: world_historical.transpose().plot(figsize=(10,10),colormap='rainbow',line
plt.legend(loc='right',bbox_to_anchor=(1.3,0.5))
plt.show()
```



```
In [51]: ▶ world_historical.index.name='country'
```

Which country worked longer hours per year?

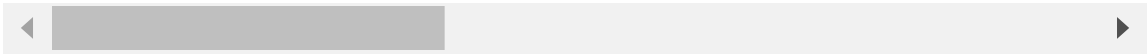
```
In [53]: ▶ world_historical.groupby('country').max()
```

Out[53]:

	1950	1951	1952	1953	1954	1955
country						
Australia	NaN	NaN	NaN	NaN	NaN	NaN
Austria	NaN	NaN	NaN	NaN	NaN	NaN
Belgium	NaN	NaN	NaN	NaN	NaN	NaN
Canada	NaN	NaN	NaN	NaN	NaN	NaN
Chile	NaN	NaN	NaN	NaN	NaN	NaN
Czech Republic	NaN	NaN	NaN	NaN	NaN	NaN
Denmark	NaN	NaN	NaN	NaN	NaN	NaN
Estonia	NaN	NaN	NaN	NaN	NaN	NaN
Finland	NaN	NaN	NaN	NaN	NaN	NaN
France	2086.380005	2114.61499	2096.035034	2078.25	2094.825012	2085.534973
Germany	NaN	NaN	NaN	NaN	NaN	NaN
Greece	NaN	NaN	NaN	NaN	NaN	NaN
Hungary	NaN	NaN	NaN	NaN	NaN	NaN
Iceland	NaN	NaN	NaN	NaN	NaN	NaN
Ireland	NaN	NaN	NaN	NaN	NaN	NaN
Israel	NaN	NaN	NaN	NaN	NaN	NaN
Italy	NaN	NaN	NaN	NaN	NaN	NaN
Japan	NaN	NaN	NaN	NaN	NaN	NaN
Korea	NaN	NaN	NaN	NaN	NaN	NaN
Latvia	NaN	NaN	NaN	NaN	NaN	NaN
Lithuania	NaN	NaN	NaN	NaN	NaN	NaN
Luxembourg	NaN	NaN	NaN	NaN	NaN	NaN
Mexico	NaN	NaN	NaN	NaN	NaN	NaN
Netherlands	NaN	NaN	NaN	NaN	NaN	NaN
New Zealand	NaN	NaN	NaN	NaN	NaN	NaN
Norway	NaN	NaN	NaN	NaN	NaN	NaN
Poland	NaN	NaN	NaN	NaN	NaN	NaN
Portugal	NaN	NaN	NaN	NaN	NaN	NaN
Russia	NaN	NaN	NaN	NaN	NaN	NaN
Slovak Republic	NaN	NaN	NaN	NaN	NaN	NaN
Slovenia	NaN	NaN	NaN	NaN	NaN	NaN
Spain	NaN	NaN	NaN	NaN	NaN	NaN
Sweden	2016.000000	2007.00000	2007.000000	1987.00	1998.000000	1987.000000

	1950	1951	1952	1953	1954	1955
country						
Switzerland	NaN	NaN	NaN	NaN	NaN	NaN
USA	1960.000000	1975.50000	1978.000000	1980.00	1970.500000	1992.500000
United Kingdom	NaN	NaN	NaN	NaN	NaN	NaN

36 rows × 66 columns



Which country worked shorter hours per year?

```
In [55]: ▶ world_historical.groupby('country').min()
```

Out[55]:

	1950	1951	1952	1953	1954	1955
country						
Australia	NaN	NaN	NaN	NaN	NaN	NaN
Austria	NaN	NaN	NaN	NaN	NaN	NaN
Belgium	NaN	NaN	NaN	NaN	NaN	NaN
Canada	NaN	NaN	NaN	NaN	NaN	NaN
Chile	NaN	NaN	NaN	NaN	NaN	NaN
Czech Republic	NaN	NaN	NaN	NaN	NaN	NaN
Denmark	NaN	NaN	NaN	NaN	NaN	NaN
Estonia	NaN	NaN	NaN	NaN	NaN	NaN
Finland	NaN	NaN	NaN	NaN	NaN	NaN
France	2086.380005	2114.61499	2096.035034	2078.25	2094.825012	2085.534973
Germany	NaN	NaN	NaN	NaN	NaN	NaN
Greece	NaN	NaN	NaN	NaN	NaN	NaN
Hungary	NaN	NaN	NaN	NaN	NaN	NaN
Iceland	NaN	NaN	NaN	NaN	NaN	NaN
Ireland	NaN	NaN	NaN	NaN	NaN	NaN
Israel	NaN	NaN	NaN	NaN	NaN	NaN
Italy	NaN	NaN	NaN	NaN	NaN	NaN
Japan	NaN	NaN	NaN	NaN	NaN	NaN
Korea	NaN	NaN	NaN	NaN	NaN	NaN
Latvia	NaN	NaN	NaN	NaN	NaN	NaN
Lithuania	NaN	NaN	NaN	NaN	NaN	NaN
Luxembourg	NaN	NaN	NaN	NaN	NaN	NaN
Mexico	NaN	NaN	NaN	NaN	NaN	NaN
Netherlands	NaN	NaN	NaN	NaN	NaN	NaN
New Zealand	NaN	NaN	NaN	NaN	NaN	NaN
Norway	NaN	NaN	NaN	NaN	NaN	NaN
Poland	NaN	NaN	NaN	NaN	NaN	NaN
Portugal	NaN	NaN	NaN	NaN	NaN	NaN
Russia	NaN	NaN	NaN	NaN	NaN	NaN
Slovak Republic	NaN	NaN	NaN	NaN	NaN	NaN
Slovenia	NaN	NaN	NaN	NaN	NaN	NaN
Spain	NaN	NaN	NaN	NaN	NaN	NaN
Sweden	2016.000000	2007.00000	2007.000000	1987.00	1998.000000	1987.000000

	1950	1951	1952	1953	1954	1955
country						
Switzerland	NaN	NaN	NaN	NaN	NaN	NaN
USA	1960.000000	1975.50000	1978.000000	1980.00	1970.500000	1992.500000
United Kingdom	NaN	NaN	NaN	NaN	NaN	NaN

36 rows × 66 columns

In []:

