

NAME : ARUL KUMAR ARK

ROLLNO: [225229103](#)

LAB-5 (Pandas Concatenate, Merge and Join)

In [1]: *#Import necessary modules*

```
import pandas as pd
import numpy as np
```

```
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]:

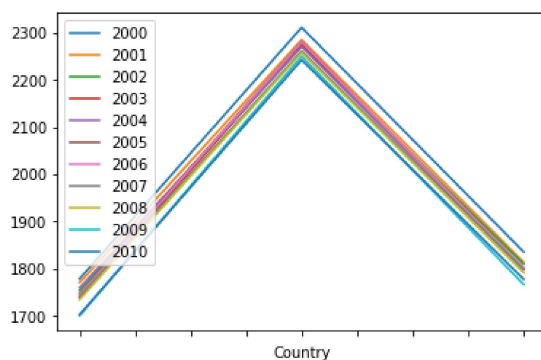
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Country											
Chile	2263	2242	2250	2235	2232	2157	2165	2128	2095	2074	2069.6

Create line graphs for our yearly labor trends in north_america

In [6]: *import matplotlib as plt*

In [8]: north_america.plot()

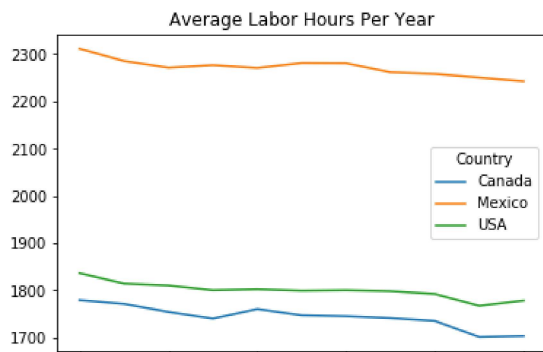
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x227fffb278>



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

```
In [9]: north_america.transpose().plot(title='Average Labor Hours Per Year')
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x22790305080>
```

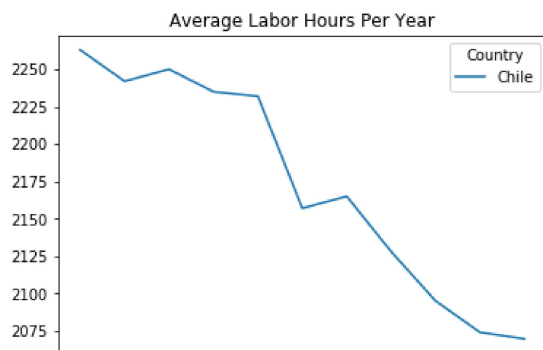


Similarly, plot transposed south_america dataframe with title "Average Labor Hours Per Year". Output chart is shown below

```
In [10]: south_america.transpose().plot(title='Average Labor Hours Per Year')
plt.show()
```

```
-----
AttributeError                                Traceback (most recent call last)
<ipython-input-10-e0491930136f> in <module>()
      1 south_america.transpose().plot(title='Average Labor Hours Per Year')
----> 2 plt.show()
```

AttributeError: module 'matplotlib' has no attribute 'show'



Concatenate America Data

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

```
In [11]: americas=pd.concat([north_america,south_america])
americas
```

```
Out[11]:
```

	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 [13]: 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 [14]: americas_dfs[1]
```

Out[14]:

	2011
Country	
Canada	1700.0
Chile	2047.4
Mexico	2250.2
USA	1786.0

```
In [15]: #americas_dfs[2]

americas_dfs[2]
```

Out[15]:

	2012
Country	
Canada	1713.0
Chile	2024.0
Mexico	2225.8
USA	1789.0

Concatenate americas and americas_dfs dataframes and store result in americas

```
In [16]: americas=pd.concat(americas_dfs,axis=1)
```

```
In [17]: americas.index.names=[ 'country' ]
```

```
In [18]: americas
```

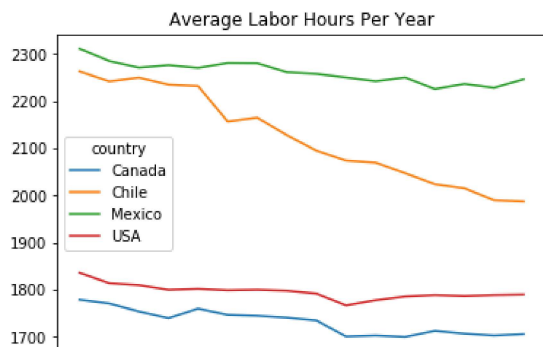
Out[18]:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
country																
Canada	1779.0	1771.0	1754.0	1740.0	1760.0	1747	1745.0	1741.0	1735	1701.0	1703.0	1700.0	1713.0	1707.0	1703.0	1706.0
Chile	2263.0	2242.0	2250.0	2235.0	2232.0	2157	2165.0	2128.0	2095	2074.0	2069.6	2047.4	2024.0	2015.3	1990.1	1987.5
Mexico	2311.2	2285.2	2271.2	2276.5	2270.6	2281	2280.6	2261.4	2258	2250.2	2242.4	2250.2	2225.8	2236.6	2228.4	2246.4
USA	1836.0	1814.0	1810.0	1800.0	1802.0	1799	1800.0	1798.0	1792	1767.0	1778.0	1786.0	1789.0	1787.0	1789.0	1790.0

Now, plot transposed americas dataframe

```
In [19]: americas.transpose().plot(title='Average Labor Hours Per Year')
```

```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x227905d9b38>
```



Appending data from other Continents

```
In [20]: asia=pd.read_csv("asia_2000_2015.csv",index_col=0)
asia
```

```
Out[20]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Country																
Israel	2017	1979	1993	1974	1942	1931	1919	1931	1929	1927	1918	1920	1910	1867	1853	1858
Japan	1821	1809	1798	1799	1787	1775	1784	1785	1771	1714	1733	1728	1745	1734	1729	1719
Korea	2512	2499	2464	2424	2392	2351	2346	2306	2246	2232	2187	2090	2163	2079	2124	2113
Russia	1982	1980	1982	1993	1993	1989	1998	1999	1997	1974	1976	1979	1982	1980	1985	1978

```
In [21]: europe=pd.read_csv("europe_2000_2015.csv",index_col=0)
europe.head()
```

```
Out[21]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Country																
Austria	1807.4	1794.6	1792.2	1783.8	1786.8	1764.0	1746.2	1736.0	1728.5	1673.0	1668.6	1675.9	1652.9	1636.7	1629.4	1624.9
Belgium	1595.0	1588.0	1583.0	1578.0	1573.0	1565.0	1572.0	1577.0	1570.0	1548.0	1546.0	1560.0	1560.0	1558.0	1560.0	1541.0
Switzerland	1673.6	1635.0	1614.0	1626.8	1656.5	1651.7	1643.2	1632.7	1623.1	1614.9	1612.4	1605.4	1590.9	1572.9	1568.3	1589.7
Czech Republic	1896.0	1818.0	1816.0	1806.0	1817.0	1817.0	1799.0	1784.0	1790.0	1779.0	1800.0	1806.0	1776.0	1763.0	1771.0	1779.0
Germany	1452.0	1441.9	1430.9	1424.8	1422.2	1411.3	1424.7	1424.4	1418.4	1372.7	1389.9	1392.8	1375.3	1361.7	1366.4	1371.0

```
In [22]: south_pacific=pd.read_csv("south_pacific_2000_2015.csv",index_col=0)
south_pacific
```

```
Out[22]:
```

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Country																
Australia	1778.7	1736.7	1731.7	1735.8	1734.5	1729.2	1720.5	1712.5	1717.2	1690	1691.5	1699.5	1678.6	1662.7	1663.6	1665
New Zealand	1836.0	1825.0	1826.0	1823.0	1830.0	1815.0	1795.0	1774.0	1761.0	1740	1755.0	1746.0	1734.0	1752.0	1762.0	1757

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

```
In [23]: world=americas.append([asia,europe,south_pacific])
```

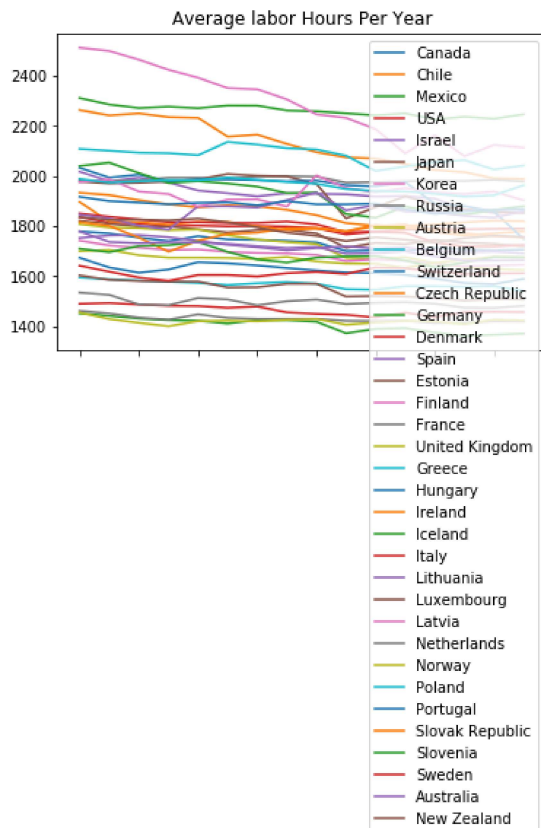
```
In [24]: world.index
```

```
Out[24]: 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 [25]: world.transpose().plot(title='Average labor Hours Per Year')
```

```
Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x227916f91d0>
```



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

```
In [26]: #Update plot() with the following features
```

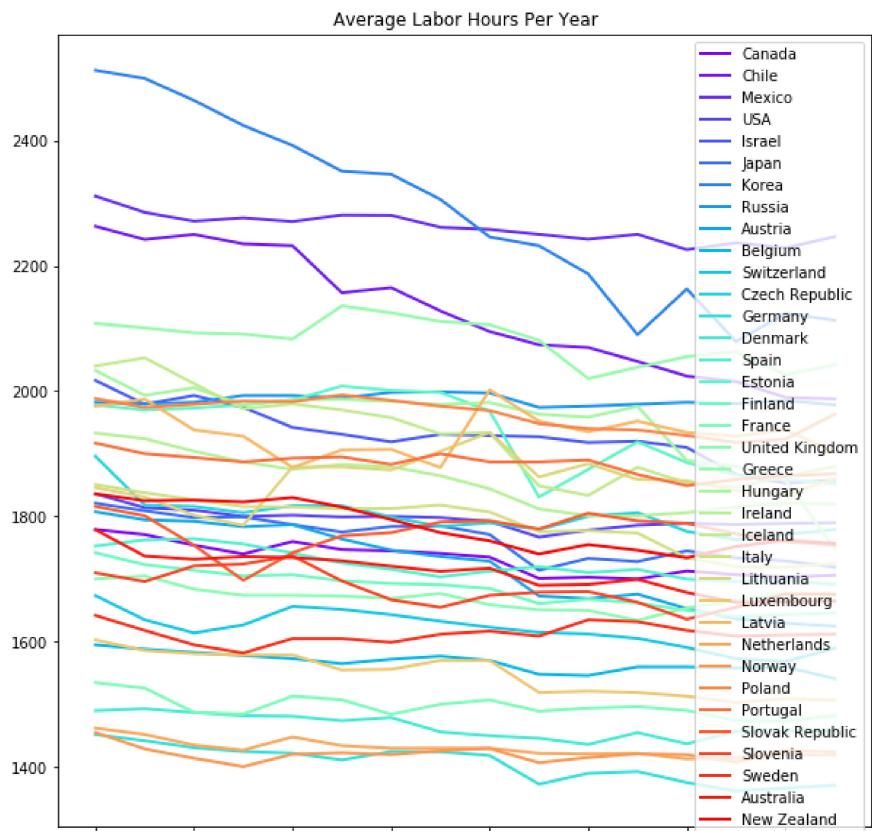
```
figsize=(10,10),  
colormap='rainbow',  
linewidth=2,  
loc='right'
```

```
In [27]: world.transpose().plot(figsize=(10,10),colormap='rainbow',linewidth=2,title='Average Labor Hours Per Year')
plt.legend(loc='right',bbox_to_anchor=(1.3,0.5))
plt.show()
```

TypeError Traceback (most recent call last)

```
<ipython-input-27-712384a32560> in <module>()
    1 world.transpose().plot(figsize=(10,10),colormap='rainbow',linewidth=2,title='Average Labor Hours Per Year')
----> 2 plt.legend(loc='right',bbox_to_anchor=(1.3,0.5))
      3 plt.show()
```

TypeError: 'module' object is not callable



Merging Historical Labor Data

```
In [28]: historical=pd.read_csv("historical.csv",index_col=0)
historical.head()
```

Out[28]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...	1990	1991	1992	1993	1994	1995	1996
Country																		
Australia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1779.5	1774.90	1773.70	1786.50	1797.60	1793.400	1782.700
Austria	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	1619.200	1637.150
Belgium	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1662.9	1625.79	1602.72	1558.59	1558.59	1515.835	1500.295
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1789.5	1767.50	1766.00	1764.50	1773.00	1771.500	1786.500
Switzerland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	1673.10	1684.80	1685.80	1706.20	1685.500	1658.900

5 rows × 50 columns

```
In [29]: print("World rows & columns: ",world.shape)
```

World rows & columns: (36, 16)

```
In [30]: print("Historical rows & columns: ",historical.shape)
```

Historical rows & columns: (39, 50)

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

```
In [31]: world_historical = pd.merge(historical, world, left_index=True, right_index=True, how='right')
```

```
In [32]: world_historical.shape
```

Out[32]: (36, 66)

Print top-5 of world_historical dataframe

```
In [33]: world_historical.head()
```

Out[33]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...	2006	2007	2008	2009	2010	2011	2012
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1745.0	1741.0	1735.0	1701.0	1703.0	1700.0	1713.0
Chile	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2165.0	2128.0	2095.0	2074.0	2069.6	2047.4	2024.0
Mexico	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2280.6	2261.4	2258.0	2250.2	2242.4	2250.2	2225.8
USA	1960.0	1975.5	1978.0	1980.0	1970.5	1992.5	1990.0	1962.0	1936.5	1947.0	...	1800.0	1798.0	1792.0	1767.0	1778.0	1786.0	1789.0
Israel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1919.0	1931.0	1929.0	1927.0	1918.0	1920.0	1910.0

5 rows × 66 columns

Joining Historical Data

Use join method to join historical dataframe and world dataframe and store result in world_historical dataframe

```
In [34]: world_historical = historical.join(world, how='right')
```

```
In [35]: #Print head of world_historical dataframe
world_historical.head()
```

Out[35]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...	2006	2007	2008	2009	2010	2011	2012
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1745.0	1741.0	1735.0	1701.0	1703.0	1700.0	1713.0
Chile	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2165.0	2128.0	2095.0	2074.0	2069.6	2047.4	2024.0
Mexico	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	2280.6	2261.4	2258.0	2250.2	2242.4	2250.2	2225.8
USA	1960.0	1975.5	1978.0	1980.0	1970.5	1992.5	1990.0	1962.0	1936.5	1947.0	...	1800.0	1798.0	1792.0	1767.0	1778.0	1786.0	1789.0
Israel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1919.0	1931.0	1929.0	1927.0	1918.0	1920.0	1910.0

5 rows × 66 columns

Plot our world labor data

```
In [37]: world_historical.sort_index(inplace=True)
```

```
In [39]: world_historical.transpose().plot(figsize=(10,10),colormap='rainbow',linewidth=2,title='Average Labor Hours Per Year')
plt.legend(loc='right',bbox_to_anchor=(1.3,0.5))
plt.show()
```

TypeError

Traceback (most recent call last)

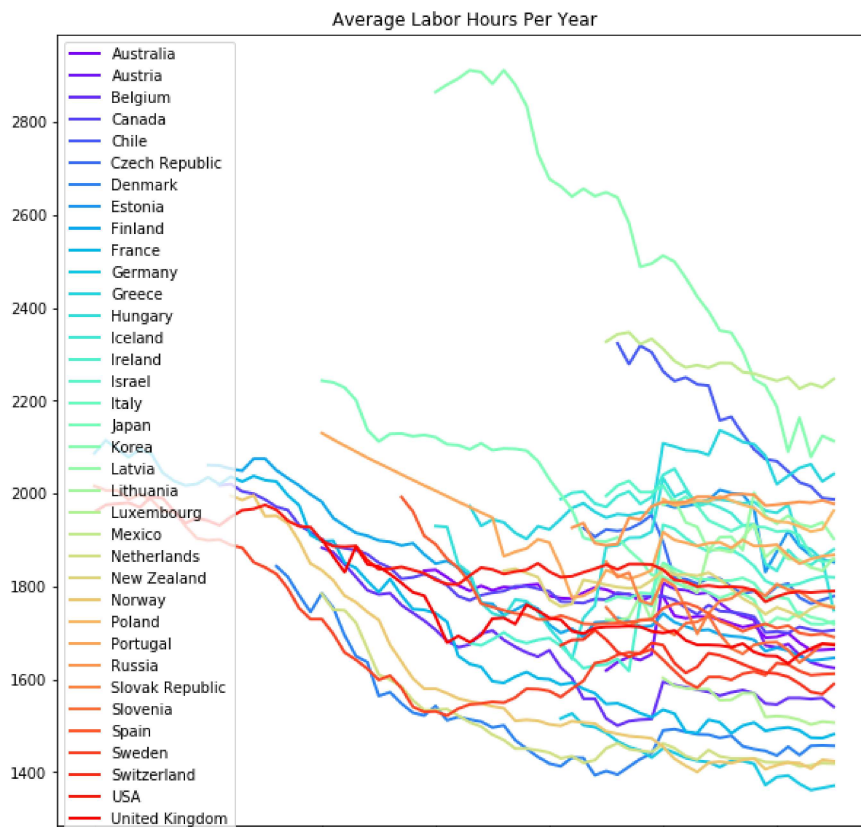
<ipython-input-39-8f322b5dfd50> in <module>()

1 world_historical.transpose().plot(figsize=(10,10),colormap='rainbow',linewidth=2,title='Average Labor Hours Per Year')

----> 2 plt.legend(loc='right',bbox_to_anchor=(1.3,0.5))

3 plt.show()

TypeError: 'module' object is not callable



Which country worked longer hours per year?

```
In [40]: world_historical.index.name='country'
```



```
In [41]: world_historical.groupby('country').max()
```

Out[41]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...
country											
Australia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Austria	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Belgium	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Chile	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Czech Republic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Denmark	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Estonia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Finland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
France	2086.380005	2114.61499	2096.035034	2078.25	2094.825012	2085.534973	2044.694946	2027.039917	2017.005005	2020.219971	...
Germany	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Greece	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Hungary	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Iceland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Ireland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Israel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Italy	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Japan	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Korea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Latvia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Lithuania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Luxembourg	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Mexico	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Netherlands	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
New Zealand	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Norway	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Poland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Portugal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Russia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Slovak Republic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Slovenia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Spain	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Sweden	2016.000000	2007.00000	2007.000000	1987.00	1998.000000	1987.000000	1962.000000	1962.000000	1930.000000	1904.000000	...
Switzerland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
USA	1960.000000	1975.50000	1978.000000	1980.00	1970.500000	1992.500000	1990.000000	1962.000000	1936.500000	1947.000000	...
United Kingdom	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

36 rows × 66 columns

Which country worked shorter hours per year?

```
In [42]: world_historical.groupby('country').min()
```

Out[42]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	...
country											
Australia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Austria	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Belgium	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Canada	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Chile	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Czech Republic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Denmark	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Estonia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Finland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
France	2086.380005	2114.61499	2096.035034	2078.25	2094.825012	2085.534973	2044.694946	2027.039917	2017.005005	2020.219971	...
Germany	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Greece	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Hungary	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Iceland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Ireland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Israel	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Italy	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Japan	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Korea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Latvia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Lithuania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Luxembourg	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Mexico	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Netherlands	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
New Zealand	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Norway	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Poland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Portugal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Russia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Slovak Republic	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Slovenia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Spain	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
Sweden	2016.000000	2007.00000	2007.000000	1987.00	1998.000000	1987.000000	1962.000000	1962.000000	1930.000000	1904.000000	...
Switzerland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
USA	1960.000000	1975.50000	1978.000000	1980.00	1970.500000	1992.500000	1990.000000	1962.000000	1936.500000	1947.000000	...
United Kingdom	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

36 rows × 66 columns

