Data Concatenation

Raw data is downloaded from GapMinder and datasets that are relatively most complete are chosen. These datasets are of the form

country	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
Aruba										
Afghanistan										
Angola										
Albania										
Andorra										
United Arab Emirates										
Argentina	151B	159B	158B	149B	164B	182B	181B	186B	195B	214B
Armenia										
American Samoa										
Antigua and Barbuda										
Australia	205B	210B	213B	226B	242B	256B	262B	279B	293B	314B
Austria	84.9B	89.6B	92B	95.8B	102B	105B	111B	115B	120B	127B
Azerbaijan										
Burundi	799M	689M	751M	782M	831M	864M	904M	1.03B	1.03B	1.01B
Belgium	107B	113B	119B	124B	132B	137B	141B	147B	153B	163B
Benin	1.65B	1.7B	1.64B	1.72B	1.84B	1.93B	2B	2.02B	2.1B	2.16B
Burkina Fa	1.17B	1.21B	1.29B	1.27B	1.3B	1.35B	1.36B	1.48B	1.52B	1.55B
Banglades	22.2B	23.6B	24.9B	24.8B	27.5B	27.9B	28.6B	28.1B	30.8B	31.1B
Bulgaria										

which are then concatenated as follows:

1. Data between 1990 – 2018 is chosen from the original sets of years included in each dataset.
2. Countries of which the number of null values is consistently across all datasets lower than 7 are chosen.
3. Existing null values in these datasets are replaced by their immediately preceding non-null values using Pandas' *fillna()* function.
4. String data (for example, '2.1M' or '-3B') is converted into their corresponding numerical values (2.100.000 or –3.000.000.000).
5. Existent data is melted into unique 'Country' - 'Year' combinations and concatenated together accordingly along the *y-axis.*

The resultant data frame contains information of 90 countries over 29 years between 1990 and 2019 and is of the following form:

	Country	Year	Democrac	Aid Receiv	Line Phon	FDI	Inequality	Income	GDP	Import	Export
0	Albania	1990	0	16300000	1.22	0	26.9	4830	4830	24	15.4
1	Albania	1991	0.764	4.63E+08	1.27	0	27	3500	3500	28.6	7.48
2	Albania	1992	0.922	5.64E+08	1.38	3.07	27	3270	3270	96.3	12.5
3	Albania	1993	0.915	4.04E+08	1.35	4.89	27	3600	3600	64.5	16
4	Albania	1994	8	2.23E+08	1.32	2.82	27	3920	3920	41.1	12
5	Albania	1995	0.923	2.26E+08	1.35	2.93	27	4470	4470	35	12.6
6	Albania	1996	0.667	2.94E+08	2.06	2.82	27	4910	4910	33.3	11.6
7	Albania	1997	6.67	2.34E+08	2.8	2.1	27.8	4400	4400	35.2	10.2
8	Albania	1998	10.7	3.79E+08	3.72	1.77	28.5	4820	4820	36.7	11.5
9	Albania	1999	10.7	7.03E+08	4.5	1.28	29.3	5480	5480	34.2	16.8
10	Albania	2000	11.1	4.91E+08	4.88	4.11	30.1	5890	5890	43.3	20.1
11	Albania	2001	13.6	4.24E+08	6.31	5.29	30.9	6440	6440	45.2	21.3
12	Albania	2002	13.1	4.53E+08	7.04	3.1	31.7	6750	6750	47.6	20.9
13	Albania	2003	12.5	4.43E+08	8.18	3.17	31.4	7150	7150	46.2	20.8
14	Albania	2004	12	3.52E+08	8.84	4.75	31	7580	7580	44.8	22.2
15	Albania	2005	12.1	3.55E+08	9.04	3.26	30.6	8040	8040	47.9	23
16	Albania	2006	11.5	3.58E+08	8.36	3.65	30.4	8570	8570	49.1	25.2
17	Albania	2007	10.8	3.08E+08	9.89	6.11	30.2	9150	9150	55	28.2
18	Albania	2008	10.1	3.39E+08	11.4	9.68	30	9910	9910	52.1	25.3

Description of the variables (taken from GapMinder website)

- *Democracy Score*: a measurement in percentage of the extent to which direct democracy is available and used, calculated from the *direct democracy indicator* of V-Dem and the *electoral indicator* from BRRD.
- *Democracy Growth*: changes in the *'Democracy Score'* between a year and the next
- *Aid Received*: net Official Development Assistance (ODA) received (in constant 2014 USD)
- *Line phone*: number of people subscribing to line phone service
- *Internet*: percentage of population using the Internet
- *FDI*: foreign direct investment, net inflows (% of GDP)
- *Inequality*: the gini coefficient of a country
- *Income*: gross domestic production per person adjusted for differences in purchasing power (in international dollars, fixed 2017 prices, PPP based on 2017 ICP)
- *GDP*: the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2010 U.S. dollars. Dollar figures for GDP are converted from domestic currencies using 2010 official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used.

Source URL: data.worldbank.org/indicator/NY.GDP.MKTP.KD

- **Export**: exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial information, business personal and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.
  Source URL: data.worldbank.org/indicator/NE.EXP.GNFS.ZS
- **Import**: Imports of goods and services represent the value of all goods and other market services received from the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial information, business personal and government services. They exclude compensation employees investment income (formerly called factor services) transfer payments.
  Source URL: data.worldbank.org/indicator/NE.IMP.GNFS.ZS