Deforestation

Data Science for Business Analytics Project

Mathieu Maréchal & Martina Kofler

14 novembre, 2021

Introduction

- Overview and Motivation
- Related Work
- Research questions

Data

Dataset index and URLs:

- AgriLandData : https://data.worldbank.org/indicator/AG.LND.AGRI.ZS
- ForestAreaData : https://data.worldbank.org/indicator/AG.LND.FRST.K2
- PopGrowth : https://data.worldbank.org/indicator/SP.POP.GROWL
- AgriVarData : https://data.worldbank.org/topic/trade
- SoybeansData : https://www.fao.org/faostat/en/#search/soy
- TradeOpenessData : https://ourworldindata.org/grapher/trade-openness
- ResourcesExtractionData : https://www.eia.gov/international/data/world
- LandAreaData : https://data.worldbank.org/indicator/AG.LND.TOTL.K2

Agricultural Land

The AgriLandData dataset shows the total area of agricultural land in square km per country/region for the years between 1960 and 2020. In this project we are just interested in the years starting from 2000.

Country	Year	AgriLand
Aruba	2000	11.1
Aruba	2001	11.1
Aruba	2002	11.1
Aruba	2003	11.1
Aruba	2004	11.1
Aruba	2005	11.1
Aruba	2006	11.1

To know how the area of agricultural land changed over time we created two new columns that contain information about the change to the year before, one in square km and one in percentage.

Country	Year	AgriLand	Change agricultural land	Change agricultural land %
Aruba	2019	NA	NA	NA
Aruba	2020	NA	NA	NA
Africa Eastern and Southern	2000	42.5	NA	NA
Africa Eastern and Southern	2001	42.6	-0.126	-0.295
Africa Eastern and Southern	2002	42.8	-0.122	-0.286
Africa Eastern and Southern	2003	43.0	-0.233	-0.545
Africa Eastern and Southern	2004	42.8	0.163	0.380
Africa Eastern and Southern	2005	43.0	-0.126	-0.293

Forest Area

The AgriLandData dataset shows the total forest area in square km per country/region for the years between 1960 and 2020. In this project we are just interested in the years starting from 2000.

Country	Year	ForestArea
Africa Western and Central	2016	20.2
Africa Western and Central	2017	20.1
Africa Western and Central	2018	20.0
Africa Western and Central	2019	19.9
Africa Western and Central	2020	19.8
Angola	2000	62.3
Angola	2001	61.9
Angola	2002	61.4

Also for this dataset we created two new columns that contain information about the change to the year before, one in square km and one in percentage.

Country	Year	ForestArea	Deforestation	Deforestation %
Africa Western and Central	2016	20.2	0.082	0.403
Africa Western and Central	2017	20.1	0.081	0.403
Africa Western and Central	2018	20.0	0.085	0.425
Africa Western and Central	2019	19.9	0.085	0.425
Africa Western and Central	2020	19.8	0.084	0.424
Angola	2000	62.3	NA	NA
Angola	2001	61.9	0.445	0.714
Angola	2002	61.4	0.445	0.719

Agricultural raw material exports, Merchandise Exports, Food exports

The dataset we took includes information about 149 indicators for 237 countries from 1960 to 2020. To get an impression about the data we show the 7 rows and 7 columns from the dataset.

Country	Indicator Name	1960	1961	1962	1963	1964
Aruba	Merchandise exports by the reporting economy, residual (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in Sub-Saharan Africa (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in South Asia (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in Middle East & North Africa (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in Latin America & the Caribbean (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise	NA	NA	NA	NA	NA

Before selecting the variables we need we changed the structure of the dataset so that it has the year and the different indicator names as column names and we also just selected data from 2000 and above.

Country	Year	Travel services (% of commercial service exports)	Transport services (% of commercial service exports)	High-technology exports (% of manufactured exports)
Aruba	2000	52.1	5.32	NA
Aruba	2001	45.3	2.86	NA
Aruba	2002	69.1	2.96	NA
Aruba	2003	53.9	2.44	NA
Aruba	2004	85.4	4.29	NA
Aruba	2005	85.0	4.39	NA
Aruba	2006	82.5	4.15	NA

Form all this variables just two are of interest for our project. Therefore we selected just these indicators and created a new dataset for each of them. The result was:

A dataset that shows the percentage of food exports from total merchandise exports in US$ per country and year

Country	Year	Food exports (% of merchandise exports)
Aruba	2000	49.96
Aruba	2001	45.08
Aruba	2002	44.09
Aruba	2003	46.87
Aruba	2004	35.86
Aruba	2005	1.18
Aruba	2006	34.38

and a dataset that shows the percentage agricultural raw material exports from total merchandise exports in US$ per country and year.

Country	Year	Agricultural raw materials exports (% of merchandise exports)
Aruba	2000	0.737
Aruba	2001	0.549
Aruba	2002	0.993
Aruba	2003	1.172
Aruba	2004	1.249
Aruba	2005	0.317
Aruba	2006	1.096

LandArea

This dataset shows how much squared kilometer every country has, this makes it possible identify how much of percent of this area is forest and how much of percent of the land is agricultural land.

Country	Year	Land Area in sq km
Aruba	2000	180
Aruba	2001	180
Aruba	2002	180
Aruba	2003	180
Aruba	2004	180
Aruba	2005	180
Aruba	2006	180

Extraction of Minerals Data

This data set contains the different amount of extraction of different sources of energy taken worldwide. Although it was a difficult data set to find for the whole world, it was also a hard one to tidy and here is why.

After running the four first line of code in order to deleat the unnecessary information lines and name the columns to be able to use it, we end up to this stage:

Ext	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Algeria																																								
Pro (quad Btu)	2.8	[illegible]																																						
Coa (quad Btu)	0	[illegible]																																				0	0	0
Nat gas (quad Btu)	0	[illegible]																																						
Pet and other liq- uids (quad Btu)	2.3	[illegible]																																						
Nu re- new- ables, and other (quad Btu)	0	[illegible]																																						
Nuclear (quad Btu)	0	[illegible]																																						
Ren and other (quad Btu)	0	[illegible]																																						

Now we can see our problem: The name of the country appear as a title that is in the same column as the

variables names we need. We are therefore at this stage unable to pivot our tibble. In order to solve this proble we created a new column "Country" where we take only one row every eights of the first column and the copy it for all the variable. Then we deleated the row corresponding at the country name that is now store in the column "Country" and obtain something like this:

Now we are finally able to pivot our tibble to have a tidy dataset usable for the rest of our analysis:

Country	Year	Production (quad Btu)	Coal (quad Btu)	Natural gas (quad Btu)	Petroleum and other liquids (quad Btu)	Nuclear, renew-ables, and other (quad Btu)	Nuclear (quad Btu)	Renewables and other (quad Btu)
World	1980	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
World	1981	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
World	1982	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
World	1983	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
World	1984	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
World	1985	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
World	1986	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Population Growth Data

Country	NA	Year	PopulationGrowth
Aruba	NA	2000	2.065
Aruba	NA	2001	2.205
Aruba	NA	2002	2.236
Aruba	NA	2003	2.108
Aruba	NA	2004	1.765
Aruba	NA	2005	1.292
Aruba	NA	2006	0.799

Openness to Trade Data

Country	Year	Ratio.of.exports.and.imports.to.GDP......PWT.9.1..
Afghanistan	2000	109.2
Afghanistan	2001	109.4
Afghanistan	2002	89.4
Afghanistan	2003	128.7
Afghanistan	2004	110.0
Afghanistan	2005	99.6
Afghanistan	2006	94.4

Soybeans dataset

Country	Year	Gross Production Value Soybeans
Albania	2000	292
Albania	2001	389
Albania	2002	200
Albania	2003	287
Albania	2004	272
Albania	2005	491
Albania	2006	285

Merged Data

Country	Year	Agricultural land	Change agricultural land	Change agricultural land %	Agricultural raw materials exports (% of merchandise exports)	Food exports (% of merchandise exports)	Forest Area	Deforestation	Deforestation %	Land Area in sq km	NA	Population growth	Gross Production Value Soybeans	Ratio	Production (quad Btu)	Coal (quad Btu)	Natural gas (quad Btu)	Petroleum and other liquids (quad Btu)	Nuclear, renewables, and other (quad Btu)	Nuclear (quad Btu)	GDP	Renewables and other (quad Btu)	PWT.9.1..201
Albania	2000	41.8	NA	NA	5.97	6.63	28.1	NA	NA	27400	NA	-0.637	292	63.5	0.0608	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Albania	2001	41.6	0.182	0.437	5.54	5.79	28.1	-0.047	-0.166	27400	NA	-0.938	389	66.5	0.0513	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Albania	2002	41.6	-0.036	-0.088	6.52	3.56	28.2	-0.047	-0.166	27400	NA	-0.300	200	68.5	0.0512	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Albania	2003	40.9	0.693	1.667	5.14	5.63	28.2	-0.047	-0.165	27400	NA	-0.374	287	67.0	0.0689	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Albania	2004	40.9	-0.036	-0.089	4.49	5.80	28.3	-0.047	-0.165	27400	NA	-0.418	272	67.0	0.0712	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Albania	2005	39.3	1.642	4.011	13.94	9.77	28.3	-0.047	-0.165	27400	NA	-0.512	491	70.9	0.0708	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Albania	2006	40.9	-1.569	-3.993	13.12	7.94	28.4	-0.047	-0.165	27400	NA	-0.631	285	74.3	0.0684	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

- Sources
- Description
- Wrangling/cleaning
- Spotting mistakes and missing data (could be part of EDA too)
- Listing anomalies and outliers (could be part of EDA too)

Analysis

- Answers to the research questions

- Different methods considered
- Competing approaches
- Justifications

Conclusion

- Take home message
- Limitations
- Future work?

  , 'eda.Rmd'