Problem Set 1

Manuel Alcala Kovalski

2021-04-12

Question 4:

Learning about the world economy

country	code	region	income_group	gdp_1993	gdp_2018	pop_1993	pop_2018
Afghanistan	AFG	South Asia	Low income	NA	20958745169	15816603	37172386
Albania	ALB	Europe & Central Asia	Upper middle income	4424061926	14547875720	3227287	2866376
Algeria	DZA	Middle East & North Africa	Upper middle income	90578860786	203353978108	27635515	42228429
American Samoa	ASM	East Asia & Pacific	Upper middle income	NA	NA	51020	55465
Andorra	AND	Europe & Central Asia	High income	1997779424	3432140710	60971	77006
Angola	AGO	Sub-Saharan Africa	Lower middle income	23003628100	99503815532	13075049	30809762
Antigua and Barbuda	ATG	Latin America & Caribbean	High income	753373666	1450219905	65777	96286
Argentina	ARG	Latin America & Caribbean	Upper middle income	259675534968	446730617131	33970111	44494502
Armenia	ARM	Europe & Central Asia	Upper middle income	2979815090	13007610291	3363108	2951776
Aruba	ABW	Latin America & Caribbean	High income	1933189215	NA	72504	105845

There are 217 observations in the data set. Some key variables are country, code, region, income_group, gdp_1993, gdp_2018, pop_1993, pop_2018.

The type for each key variable and the extent of missingness is shown in the table below.

Data summary

Name	Piped data
Number of rows	217
Number of columns	8
_____	
Column type frequency:	
character	4
numeric	4
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
country	0	1	4	30	0	217	0
code	0	1	3	3	0	217	0
region	0	1	10	26	0	7	0
income_group	0	1	10	19	0	4	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
gdp_1993	40	0.82	222058792250	8.696209e+11	23774990	3.228369e+09	13130746305	96286128144	9.564447e+12	▇▁▁▁▁
gdp_2018	33	0.85	438901197703	1.672617e+12	41843721	1.006552e+10	35935802607	215006001953	1.784428e+13	▇▁▁▁▁
pop_1993	2	0.99	25663139	1.056904e+08	9194	5.170375e+05	4600463	14157942	1.178440e+09	▇▁▁▁▁
pop_2018	1	1.00	35033534	1.369768e+08	11508	7.728515e+05	6572040	25011584	1.392730e+09	▇▁▁▁▁

Analysis data set

Once you have done this (no need to type answers to these questions, but do answer them), create an analysis data set in which:

- You transform the population variables so that they are expressed in millions of people (for example, 158,000,000 should become 158).
- You transform the gdp variables so that they are expressed in millions of dollars
- You keep only observations that have non-missing data for both population 2018 and gdp 2018.

This will be the data set you will you for the remainder of this problem set, so assign it to an object you can use.

```
wdi <- wdi_raw %>% drop_na(ends_with("2018")) %>% pivot_longer(cols = gdp_1993:pop_2018, nam
    names_pattern = "(.+)_(.+)") %>% mutate(across(.cols = starts_with(c("pop", "gdp")), .fn
```

country	code	region	income_group	date	gdp	pop
United States	USA	North America	High income	2018	17844276	327.1674
China	CHN	East Asia & Pacific	Upper middle income	2018	10800568	1392.7300
United States	USA	North America	High income	1993	9564447	259.9190
Japan	JPN	East Asia & Pacific	High income	2018	6189748	126.5291
Japan	JPN	East Asia & Pacific	High income	1993	4880197	124.5360

Now report the mean and the number of observations for gdp 2018 for this analysis data set.

date	mean_gdp	observations
1993	234209.9	165
2018	438901.2	184

2. Totals: Please calculate and report:

a. Total World GDP (expressed in trillions of 2010 dollars) and world population in 2018 (expressed in billions of people)

Year	Total GDP	Total population
1993	38,703,380.50	5,373.78
2018	80,757,820.38	7,373.17

b. Top 5 countries in terms of GDP in 2018 and their respective GDPs and top 5 countries in terms of population in 2018 and their respective populations. You may round your values.

Richest countries in 2018

As measured by GDP

country	gdp
United States	17,844,275.61
China	10,800,568.39
Japan	6,189,748.01
Germany	3,939,226.61
France	2,924,903.22

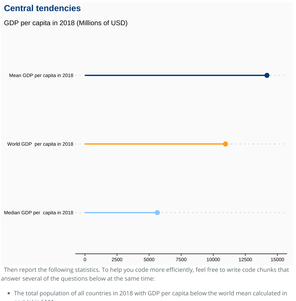
Most populated countries in 2018

In millions

Country	Population
China	1,392.73
India	1,352.62
United States	327.17
Indonesia	267.66
Pakistan	212.22

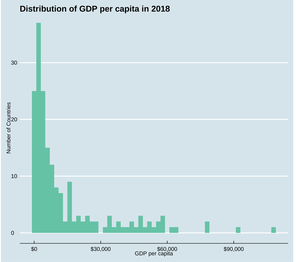
(3) **Central Tendencies:** Calculate GDP per capita (which is equal to total GDP divided over population, but beware of units) for each country in the database for 1993 and 2018.

Mean GDP per capita in 2018	Median GDP per capita in 2018	World GDP per capita in 2018
14183.25	5639.876	10952.92



Then report the following statistics. To help you code more efficiently, feel free to write code chunks that answer several of the questions below at the same time;

- The total population of all countries in 2018 with GDP per capita below the world mean calculated in part (a) is 6111
- Draw a histogram



Proportion of countries in each income category

Income	Number of countries	Proportion
Low income	26	14.1%
Lower middle income	46	25%
Upper middle income	56	30.4%
High income	56	30.4%

b. What percent of countries in Sub-Saharan Africa are low income?

46.7% percent of countries in Sub-Saharan Africa are low income

c. What share of low-income countries are in sub-Saharan Africa?

The majority of low-income countries are in sub-Saharan Africa. In particular 81% of low income countries are in sub-Saharan Africa.

Problem 5:

Dealing with missing data

The database does not have a 1993 GDP figure for 19.

Ignoring missing values, the mean GDP per capita in 1993 was 9366