Bailey Williams
Problem Set 1
Due Date: July 5, 2025
Word Count: 590
Submitted: July 10, 2025

Problem Set 1: Working with World Development Indicators Data in Excel

This problem set uses the World Development Index (WDI) 2020 data, both to gain familiarity with Microsoft Excel and to practice data analysis.

1) The below table includes the median, mean, minimum, and maximum values for 217 different countries, based on the WDI 2020 data.

	Median	Mean	Minimum	Maximum
Proportion of seats held by women in national parliaments (%)	22.56	23.85	0	61.25
Mortality rate, infant (per 1,000 live births)	13.05	20.26	1	80.5
Foreign direct investment, net inflows (% of GDP)	1.75	3.85	-104.06	138.21
Control of Corruption: Estimate	-0.15	-0.03	-1.92	2.27
Political Stability and Absence of Violence/Terrorism: Estimate	0.03	-0.04	-2.71	1.91
Urban population	3914834	20302709.9	5587	866810508
Prevalence of HIV, total (% of population ages 15-49)	0.4	1.67	0.1	28.6

Bailey Williams
Problem Set 1
Due Date: July 5, 2025
Word Count: 590
Submitted: July 10, 2025

Population, total	6579900	35932626.51	11069	1411100000
Trade (% of GDP)	70.06	84.37	9.96	365.22
Military expenditure (% of GDP)	1.53	1.96	0.01	9.85

2) The missing observations for some variables, rather than entering a zero, suggest that these variables cannot be recorded by the country, or that no data exists. Many countries are included in the dataset that are territories of other countries, such as American Samoa and the Northern Mariana Islands (United States) and Bermuda and the Isle of Man (United Kingdom). In these cases, many of their records are likely included in the reports of the larger country. Other gaps in the data reported may be due to the lack of records for the specified year. For example, HIV prevalence is missing values for many countries, which may be because HIV was not measured in 2020, perhaps due to the COVID-19 pandemic, which overwhelmed healthcare systems.

3) Countries with a high infant mortality rate and high percentage of HIV cases have a lower ratio of urban to total population, such as Eswatini and Botswana. Likewise, there also appears to be a relationship between low levels of infant mortality and HIV cases with a higher ratio of urban to total population, as is the case with Switzerland and Norway. This suggests that there is a relationship between more urban areas and higher medical care.

Bailey Williams
Problem Set 1
Due Date: July 5, 2025
Word Count: 590
Submitted: July 10, 2025

4) A few notable outliers exist in the dataset, particularly with the health indicator variables. There are exceptionally high (18+%) HIV prevalence rates in Eswatini, Botswana, and South Africa, which is over 16 points above the average rate of 1.68%. In addition, infant mortality rates were exceptionally high in Chad, the Central African Republic, and Sierra Leone, with rates above 65 per 1,000 live births, more than triple the average of 20.13 per 1,000 live births. Qatar, the United Arab Emirates, and Japan all had exceptionally low infant mortality rates of less than 2 per 1,000 live births.

5) Based on the percentages of HIV included, the Sub-Saharan African region has exceptionally high rates of HIV prevalence. This is likely due to the limited healthcare infrastructure in the region, a lack of education about HIV, and because the virus likely originated in Central Africa, as suggested by the Alan Guttmacher Institute. The Institute states that the overwhelming poverty in the region makes sexual and reproductive healthcare a low priority. Education also is lacking, with less than 20% of women and less than 30% of men aged 15-19 holding more than a primary school education as of 2004 (Bankole et al., 2004, p. 2).

Bailey Williams
Problem Set 1
Due Date: July 5, 2025
Word Count: 590
Submitted: July 10, 2025

References

Bankole, A., Singh, S., Woog, V., & Wulf, D. (2004). *Risk and Protection: Youth and HIV/AIDS in*

Sub-Saharan Africa. The Allan Guttmacher Institute.

https://www.guttmacher.org/sites/default/files/pdfs/pubs/riskandprotection.pdf