Bailey Williams
Problem Set 6
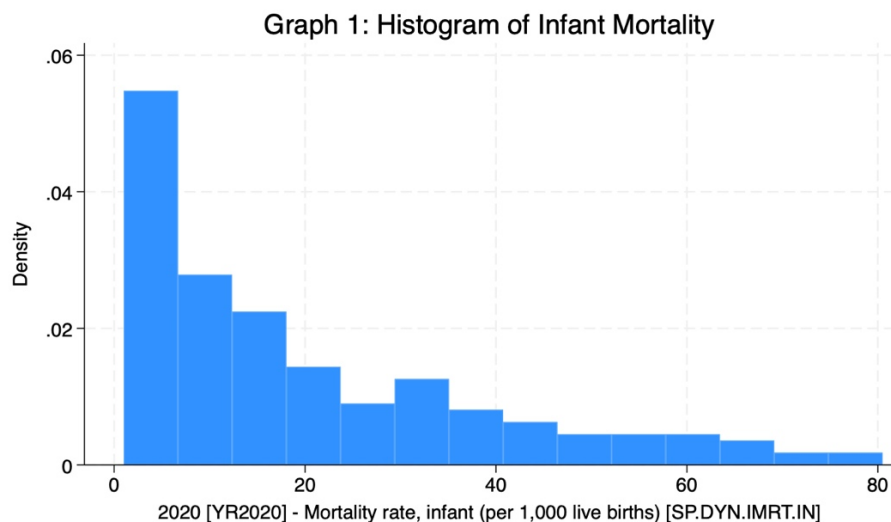Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

Problem Set 6: Descriptive Statistics

This problem set uses the World Development Index (WDI) 2020 data, particularly focusing on the infant mortality variable. Part 1 includes descriptive statistics for the infant mortality variable, discussing mean, median, and standard deviation. Part 2 analyzes infant mortality rates by region. Finally, Part 3 conducts bivariate regression analysis to determine the relationship between infant mortality and the percentage of women in parliament, both with and without controlling for political stability and region. STATA code is included in the Appendix to support these findings.

Part 1: Descriptive Statistics—Mean and Median

b) The mean of infant mortality based off the 2020 WDI data is 20.26 deaths per 1,000 live births, and the median is 13.05 deaths per live births.

c) Vanuatu is closest to the mean level of infant mortality with 20.3 deaths per live births. St. Vincent and the Grenadines and Brazil are equally representative of the median level of infant mortality, with 13 and 13.1 deaths per 1,000 live births respectfully.

d) Yes, the distribution is skewed because the mean number of infant deaths per 1,000 live births is higher than the median number. If there was no skewness the two values would be equal.

e)

Graph 1: Histogram of Infant Mortality

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

f) The state with the highest level of infant mortality is Sierra Leone, with 80.5 deaths per 1,000 live births. The state with the lowest level of infant mortality is Monaco, with 1 death per 1,000 live births. The median is the better of the two measurements for infant mortality.

g) The median is a better measure for infant mortality than the mean because the dataset is skewed. When outliers exist, the mean tends to skew higher or lower as a result. The median, however, is always the middle value of the dataset, which makes it a better measure when the data is skewed, like it is for infant mortality.
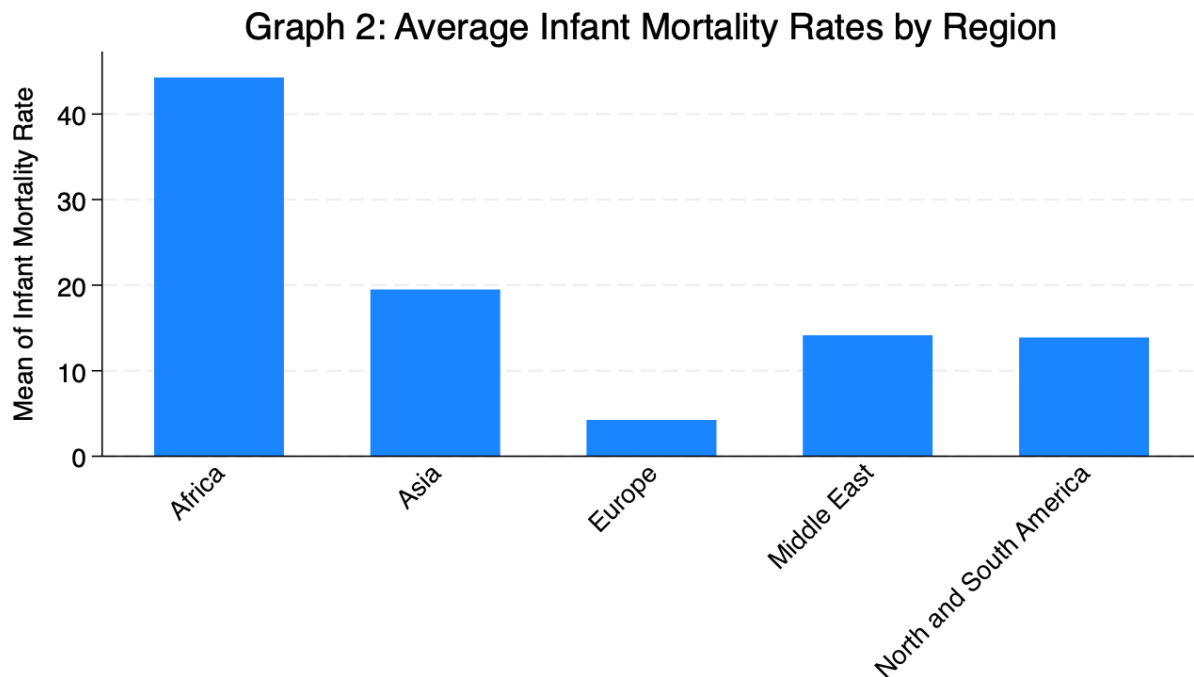
Part 2: Building Tables and Analyzing Subsamples

a)

Table 1: Infant Mortality Rates by Continent

GEO *(Separated by region of the world based on the Correlates of War (COW) scheme)*	N	Mean	SD	Median
Africa	48	44.25	17.43	42.65
Asia	43	19.47	12.89	18.70
Europe	47	4.24	3.12	3.40
Middle East	21	14.11	10.93	11.20
North and South America	37	13.86	8.60	11.80

b) Africa has the highest average infant mortality, with a mean value of 44.25 deaths per 1,000 live births.

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

c)

Graph 2: Average Infant Mortality Rates by Region

d) Europe is the region with the lowest average infant mortality rate, with a mean of 4.24 deaths per 1,000 live births.

e) Africa has the most varied infant mortality rate due to its standard deviation of 17.43 deaths per 1,000 live births, meaning infant mortality rates between countries are very spread out. Europe is the most homogenous in its infant mortality rates due to its standard deviation of 3.12 deaths per 1,000 live births, meaning infant mortality rates between countries are clustered closely around the mean.

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

Part 3: Regression Analysis

a)

Table 2: Bivariate Regression Analysis of Infant Mortality on the Function of the Percentage of Women in Parliament

Source	SS	df	MS	Number of obs = 189
Model	3836.49	1	3836.49	F(1, 187) = 11.34
Residual	63242.07	187	338.19	Prob > F = 0.00
Total	67078.56	188	356.80	R-squared = 0.06
				Adj R-squared = 0.05
				Root MSE = 18.39

Infant Mortality Rate	Coefficient	Std. err.	t	P>t	[95% conf. interval]	
Percentage of Women in Parliament	-0.37	0.11	-3.37	0.001	-0.58	-0.15
_cons	28.98	2.93	9.89	0.00	23.20	34.76

The direction of effect between infant mortality and the percentage of women in parliament is negative, meaning that as the number of women in parliament increases in a country, that country's infant mortality rate decreases. The magnitude of effect is 0.37, meaning that for every 1 percent increase of women in parliament, the number of deaths per 1,000 live births by 0.37 deaths. While there is a relationship between the variables, less than 6 percent of the variation in infant mortality rates can be explained as a function of the percentage of women in parliament. The null hypothesis of no relationship between infant mortality rates and the percentage of women in parliament is rejected due to a p-value of 0.001.

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

b)

Table 3: Bivariate Regression Analysis of Infant Mortality on the Function of the Percentage of Women in Parliament Controlling for Region and Political Stability

Source	SS	df	MS	
Model	49740.14	6	8290.02	F(6, 182) = 87.02
Residual	17338.42	182	95.27	Prob > F = 0.00
Total	67078.56	188	356.80	R-squared = 0.74
				Adj R-squared = 0.73
				Root MSE = 9.76

Number of obs = 189

Infant Mortality Rate	Coefficient	Std. err.	t	P>t	[95% conf. interval]	
Percentage of Women in Parliament	-0.17	0.06	-2.67	0.01	-0.30	-0.04
Africa	31.19	2.29	13.62	0.00	26.67	35.71
Asia	10.92	2.24	4.86	0.00	6.49	15.35
Middle East	-3.69	3.05	-1.21	0.23	-9.70	2.32
Americas	6.91	2.22	3.11	0.00	2.53	11.30
Political Stability	-7.09	0.86	-8.24	0.00	-8.79	-5.39
_cons	12.67	2.46	5.14	0.00	7.81	17.53

The direction of effect between infant mortality rate and political stability is negative, meaning that as a state becomes more politically stable, the infant mortality rate decreases. The magnitude of effect between infant mortality rate and political stability is 7, meaning that a 1 unit increase in political stability leads to a 7 unit decrease in infant

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

mortality. 74 percent of the variation in infant mortality can be explained as a function of political stability, the region dummy variables, and percentage of women in parliament. The inclusion of these new independent variables has caused the magnitude of effect between infant mortality rate and the percentage of women in parliament to decrease to 0.17. The p-value for percentage of women in parliament also increased to 0.01, meaning that while the null hypothesis of no relationship between these variables is still rejected, less of the variation in infant mortality rate is explained by percentage of women in parliament compared to other variables.

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

Appendix

use "/Users/baileywilliams/Downloads/WDI 2020-2.dta"

fsum YR2020Mortalityrate, stats(mean p50)

histogram YR2020Mortalityrate

graph export "/Users/baileywilliams/Desktop/IM Histogram.jpg", as(jpg) name("Graph") quality(100) replace

graph save "Graph" "/Users/baileywilliams/Desktop/IM Histogram.gph", replace

sort YR2020Mortalityrate

findit kountry

kountry CountryCode, from(iso2c) geo(cow)

replace GEO = "Europe" in 4

replace GEO = "Europe" in 10

replace GEO = "Middle East" in 56

replace GEO = "Europe" in 75

replace GEO = "Middle East" in 102

replace GEO = "Africa" in 189

replace GEO = "North and South America" in 217

replace GEO = "Asia" in 207

replace GEO = "North and South America" in 37

replace GEO = "North and South America" in 79

replace GEO = "North and South America" in 197

replace GEO = "North and South America" in 198

replace GEO = "Europe" in 199

replace GEO = "Asia" in 200

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

```
replace GEO = "North and South America" in 201

replace GEO = "Asia" in 202

replace GEO = "Europe" in 203

replace GEO = "North and South America" in 204

replace GEO = "North and South America" in 206

replace GEO = "North and South America" in 208

replace GEO = "Asia" in 209

replace GEO = "North and South America" in 210

replace GEO = "Asia" in 211

replace GEO = "Asia" in 212

replace GEO = "Europe" in 213

replace GEO = "Asia" in 214

replace GEO = "Europe" in 215

replace GEO = "North and South America" in 216

save "/Users/baileywilliams/Downloads/WDI 2020-2.dta", replace

bysort GEO: fsum YR2020Mortalityrate, stats(mean p50 sd)

graph bar (mean) YR2020Mortalityrate, over(GEO, label(angle(forty_five)))

graph export "/Users/baileywilliams/Desktop/IM Bar Chart.jpg", as(jpg) name("Graph")
quality(100)

graph save "Graph" "/Users/baileywilliams/Desktop/IM Bar Chart.gph"

regress YR2020Mortalityrate YR2020Proportionofse

generate Africa = 0

replace Africa = 1 if GEO == "Africa"

generate Asia = 0
```

Bailey Williams
Problem Set 6
Due Date: July 28, 2025
Word Count: 988
Submitted: August 2, 2025

replace Asia = 1 if GEO == "Asia"

generate Americas = 0

replace Americas = 1 if GEO == "North and South America"

generate MiddleEast = 0

replace MiddleEast = 1 if GEO == "Middle East"

regress YR2020Mortalityrate YR2020Proportionofse Africa Asia MiddleEast Americas
YR2020PoliticalStabil

save "/Users/baileywilliams/Downloads/WDI 2020-2.dta", replace