National Income and Infant Mortality

In this Jupyter Notebook, we'll analyze the relationship between a country's GDP per capita (a measure of average income per person) and infant mortality (in particular, the share of every 1,000 children born who do not reach their fifth birthday).

Data

Data for this analysis comes from the World Bank's *World Development Indicators* database.

```
In [ ]:  import pandas as pd
         import numpy as np

         # Download World Development Indicators
         wdi = pd.read_csv("https://media.githubusercontent.com/media/nickeubank/"
                           "MIDS_Data/master/World_Development_Indicators/wdi_small_t
```

Now let's quickly look at our data. The next command just shows us the first 5 rows of our data (we'll spend more time on these tools in later lessons). Note further than you can scroll right to see more columns!:

```
In [ ]:  wdi.head()
```

Out[ ]:

	Country Name	Adolescent fertility rate (births per 1,000 women ages 15-19)	Antiretroviral therapy coverage for PMTCT (% of pregnant women living with HIV)	Battle-related deaths (number of people)	CPIA building human resources rating (1=low to 6=high)	CPIA business regulatory environment rating (1=low to 6=high)	CPIA debt policy rating (1=low to 6=high)	ec mana a (1 (
0	Afghanistan	73.1264	NaN	17273.0	3.5	2.5	3.0	
1	Albania	20.6922	NaN	NaN	NaN	NaN	NaN	
2	Algeria	10.7052	28.0	110.0	NaN	NaN	NaN	
3	American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	
4	Andorra	NaN	NaN	NaN	NaN	NaN	NaN	

5 rows × 129 columns

```
In [ ]:  # Now we can just print out the column names:
         wdi.columns
```

```
Out[ ]: Index(['Country Name',
               'Adolescent fertility rate (births per 1,000 women ages 15-19)',
               'Antiretroviral therapy coverage for PMTCT (% of pregnant women livi
        ng with HIV)',
               'Battle-related deaths (number of people)',
               'CPIA building human resources rating (1=low to 6=high)',
               'CPIA business regulatory environment rating (1=low to 6=high)',
               'CPIA debt policy rating (1=low to 6=high)',
               'CPIA economic management cluster average (1=low to 6=high)',
               'CPIA efficiency of revenue mobilization rating (1=low to 6=high)',
               'CPIA equity of public resource use rating (1=low to 6=high)',
               ...
               'Women participating in the three decisions (own health care, major
        household purchases, and visiting family) (% of women age 15-49)',
               'Women who believe a husband is justified in beating his wife (any o
        f five reasons) (%)',
               'Women who believe a husband is justified in beating his wife when s
        he argues with him (%)',
               'Women who believe a husband is justified in beating his wife when s
        he burns the food (%)',
               'Women who believe a husband is justified in beating his wife when s
        he goes out without telling him (%)',
               'Women who believe a husband is justified in beating his wife when s
        he neglects the children (%)',
               'Women who believe a husband is justified in beating his wife when s
        he refuses sex with him (%)',
               'Women who were first married by age 15 (% of women ages 20-24)',
               'Women who were first married by age 18 (% of women ages 20-24)',
               'Women's share of population ages 15+ living with HIV (%)'],
              dtype='object', length=129)
```

Visualizing the Relationship between Log GDP Per Capita and Infant Mortality

[Now it's your turn! insert the plot from `analyze_health_and_income.py` here and make any required changes to make it work]

```python
In [ ]: wdi["Log GDP Per Capita"] = np.log(wdi["GDP per capita (constant 2010 US$)"]
        import altair as alt

        my_chart = (
            alt.Chart(wdi)
            .mark_circle(size=50)
            .encode(
                x="Log GDP Per Capita",
                y="Mortality rate, under-5 (per 1,000 live births)",
                tooltip=["Country Name"],
            )
            .interactive()
        )
        my_chart
```

Out[ ]:
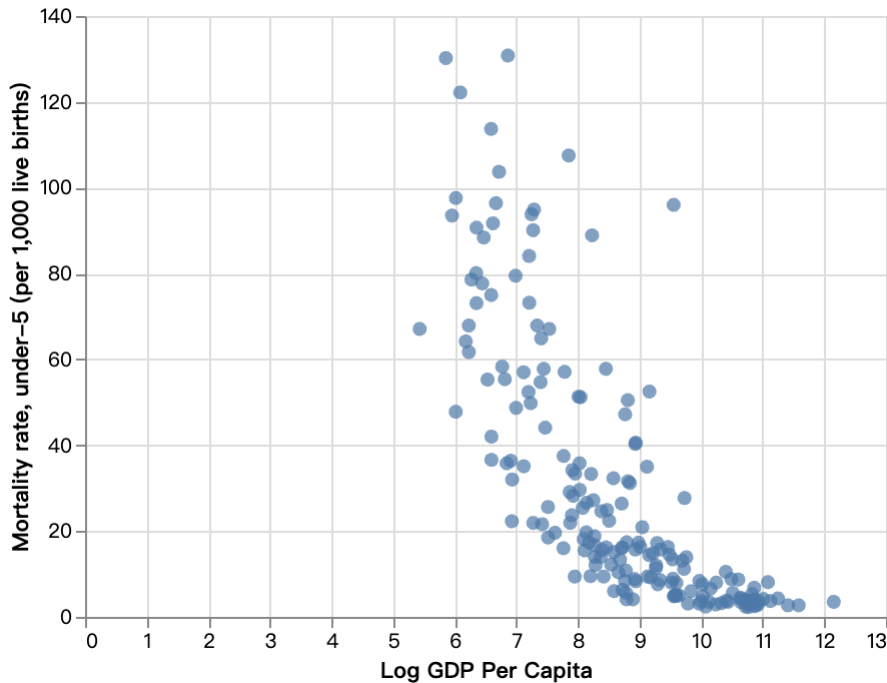
Tell Me Something Cool You've Learned

Write me a little markdown cell (with some fun formatting!) telling me something you saw in the plot you didn't expect.

There is correlation between Log GDP Per Capita and Mortality Rate

Overall, the general trend is that the higher the Log GDP Per Capita, the lower the Mortality Rate

Equatorial Guinea has a relatively high Log GDP Per Capita, but the mortality rate is still relatively high compared to the whole trend. The point might be an outlier.

Export and Send Your Notebook to Me!

When you are finished, the next step is to export this notebook.

In my experience, the best way to do this is as follows:

1. Along the top of your notebook, select "Export" (it might be in the three-dot menu).
2. Choose "HTML" and **save it next to this notebook file**. This is important because any images in the HTML have relative file paths that are set up to be in reference to the location of your notebook, so if you save it somewhere else, when you open it you may lose all your images.
3. Open that HTML in your normal web browser (Chrome, Firefox, etc), **not** in VS Code.
4. Print the page to PDF.

Why do this instead of choosing the PDF option when exporting? If you try and export a notebook directly to PDF, VS Code will actually try and use a tool to convert it to a LaTeX

document, compile that latex document, then print, and getting this setup right can be a pain, and the LaTeX conversion often causes problems. So while a little convoluted, that's my recommendation.

With that said, if you like LaTeX (and have it installed), go ahead and export directly to PDF and see if you like the result! **Just be sure to initially save the PDF next to the notebook file,** as the same issue with relative paths affects LaTeX as HTML files.

Now that you have a PDF, please upload it to Gradescope!