HW2_ggplots_instructions

Aidan O'Hara

2023-09-15

Instructions

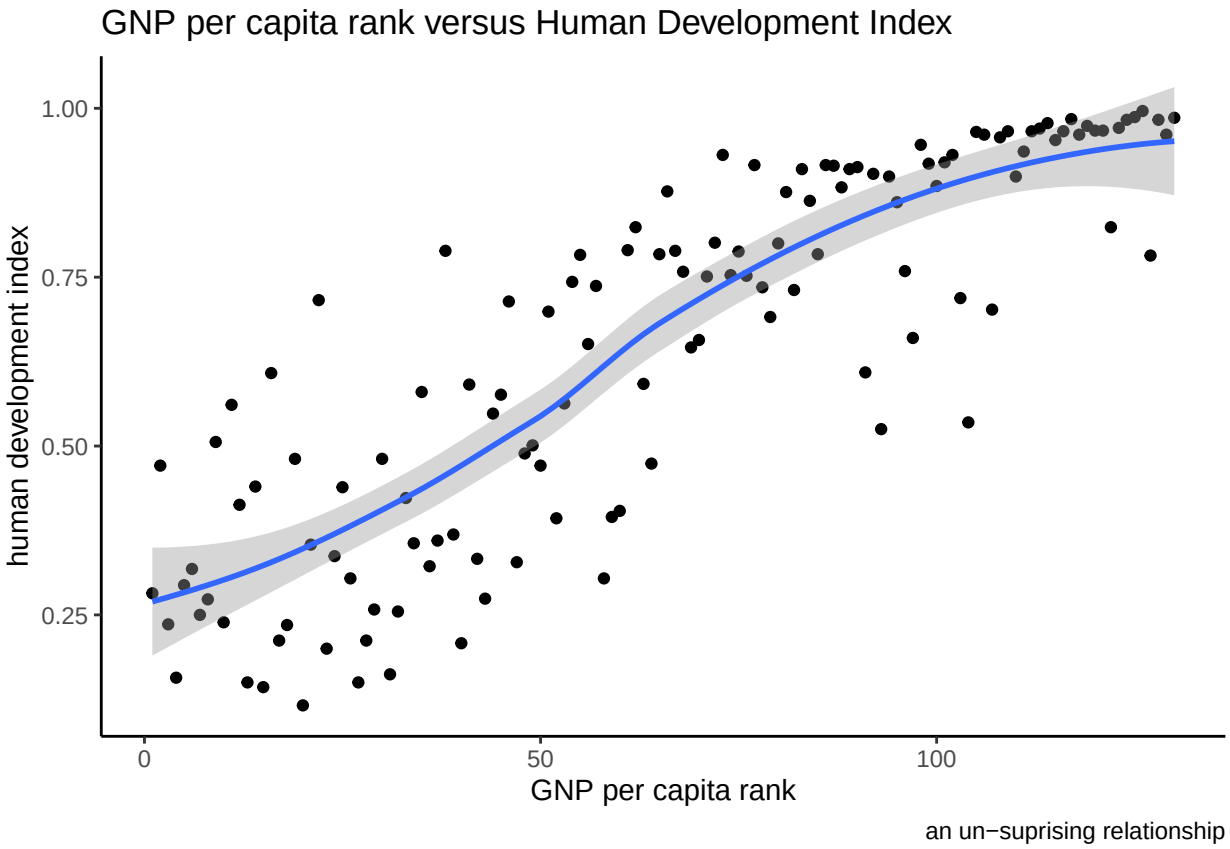
In this assignment you will practice reading in, manipulating, and plotting using three data sets, `human`, `world`, and `strikes`. You'll find the data in the included `.txt` and `.csv` files by the same name. Some data treatment has been included to make the data sets friendlier. In the future you will need to do these yourself.

Following are five example plots you are tasked with recreating. You will also find specific instructions about creating properties not included within the native data.

Gradescope will verify that your plots have exactly the labels specified. That you've applied the correct theme, and defined appropriate `aes()`. Use `geom_point()` when applicable, as opposed to `geom_jitter`. Finally, use `theme_classic()`.

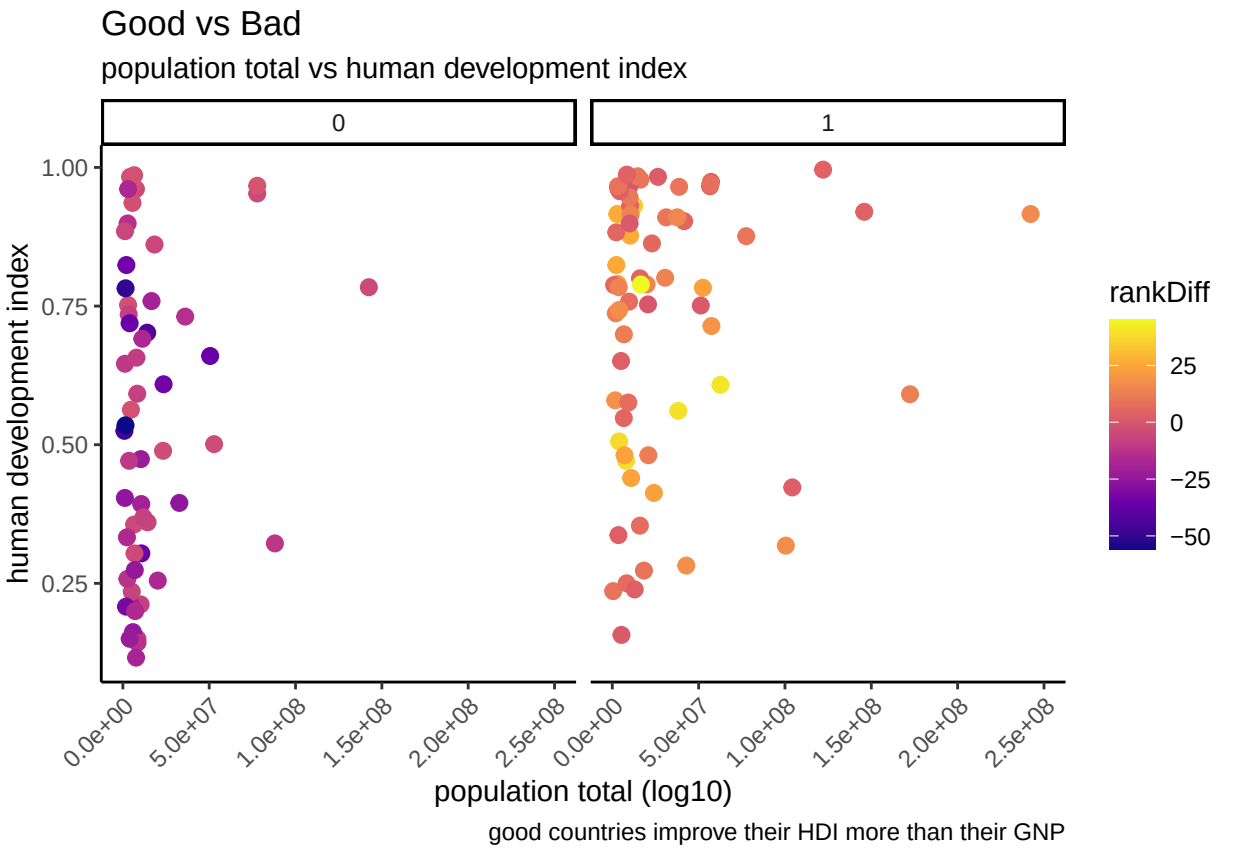
Please attempt to use only base-R dataframe operations, although `dplyr` is allowed if you would like to get a jump start on next week.

plot 1


GNP per capita rank versus Human Development Index
human development index
1.00
0.75
0.50
0.25
0
50
100
GNP per capita rank
an un−suprising relationship

plot 2

```
## Scale for x is already present.
## Adding another scale for x, which will replace the existing scale.
```



In order to proceed, you will need to add columns to the `human` set from the `world` set. Using the country names as your key. `human` should now have the following columns with values about nearly all the countries it references.

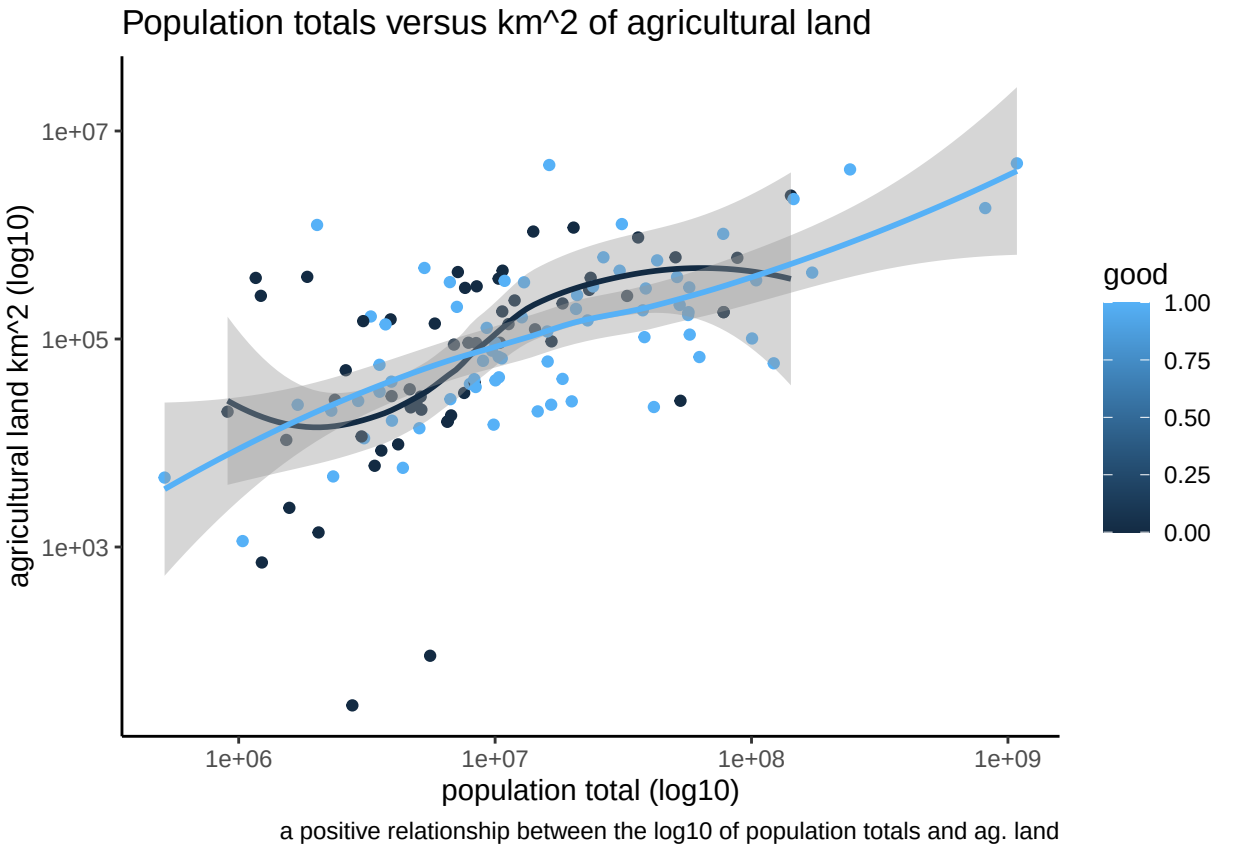
```
colnames(human)
```

```
##  [1] "rank"                      "country"
##  [3] "human_dev_index"           "gnp_percap_rank"
##  [5] "rankDiff"                  "population_total"
##  [7] "agricultural_land_km_sq"   "percent_population_female"
##  [9] "percent_land_arable"       "good"
```

Use `theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))` to turn your axis labels.

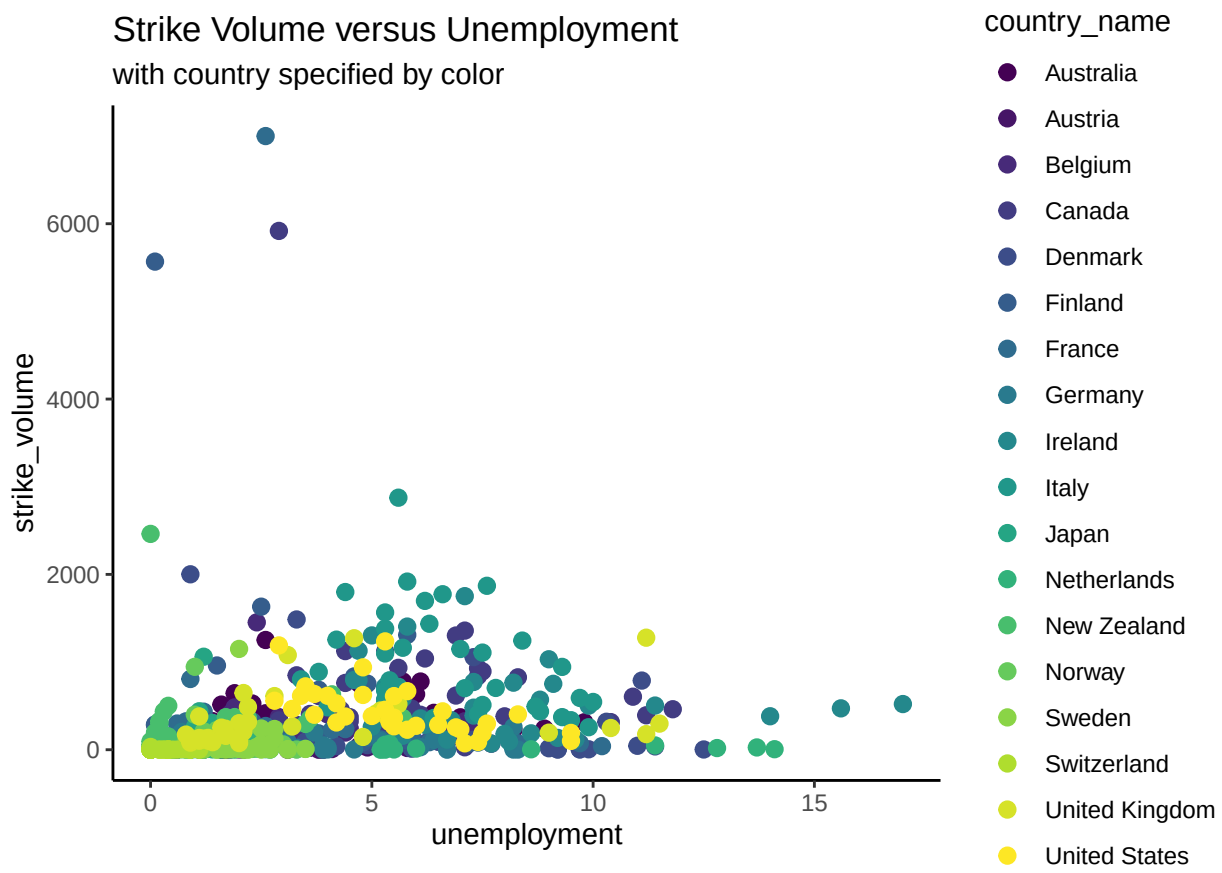
The `good` variable used here is a Boolean defined with respect to each countries `rankDiff`. True when `rankDiff >= 0`. Also be sure that you apply a `log10` transformation to total population. This will make the plot easier to read. Limit your x-values to between `0 <= x <= 2.5e08` to omit some extrema.

plot 3



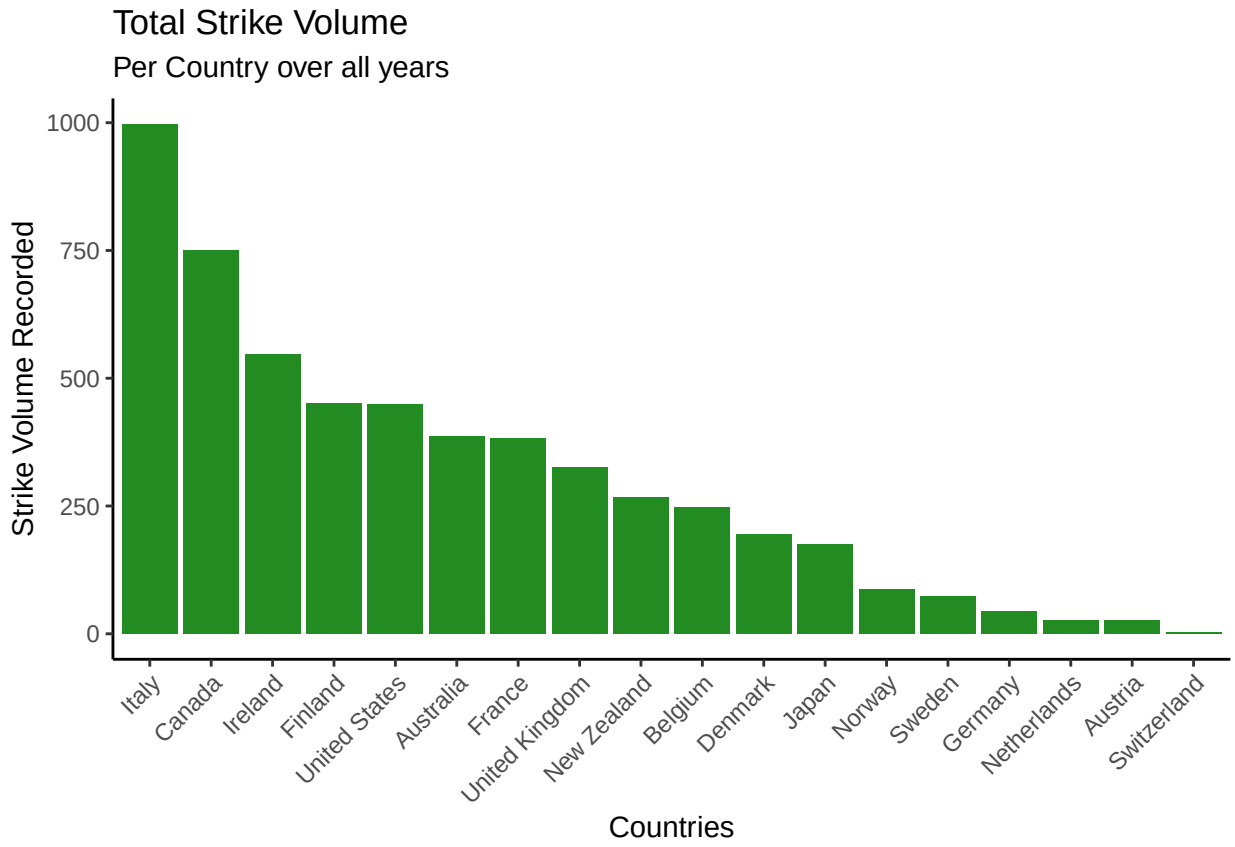
Applying a `log10` transformation to both the square kilometers of agricultural land and the total population of each country, plot the two against each other. Use `good` as your `group` and `color` argument to produce the two `geom_smooth()` lines.

plot 4



`plot4` displays `strike_volume` versus `unemployment` and colors the points according to which country they are reporting. Don't sweat the specific colors, just make sure your legend is present and your color scheme contrasts enough to provide some insight.

plot 5



To construct `plot5` you will need to create an aggregation table about `stikes`. Your aggregation table should include `country_names` and `volume`, where `volume` is the `mean(strike_volume)` calculated for each country. See if you can determine a way to order the rows of your barplot to match the example. Again rotate your axis labels so they are legible.

Appendix

human

"To measure the quality of life in a nation, the United Nations Development Program started figuring a Human Development Index. A nation's HDI is composed of life expectancy, adult literacy and Gross National Product per capita.

By combining these three elements and by pitting each nation's indicators against "the best," we come up with a worldwide HDI. Comparing the HDI rating with the traditional GNP per capita rating reveals some poor countries' remarkable progress in human development.

These countries got more bang for their development buck by giving their aid to the most needy people. The comparison also shows that some countries, including the U.S., did not translate their wealth into social benefits.

In the HDI rankings, the Arab and Moslem countries come out poorly, mainly because of low literacy among women. The formerly communist countries come out rather well because literacy is a priority and their GNP is generally low.

Latin America comes out with many plusses because their GNPs are low while they still enjoy the higher literacy and improved health-care investments of earlier years.

Africa is a mixed lot. Some oil exporters, such as Angola, Gabon, Cameroon and the Congo, did not translate their wealth into social benefits. Others–Tanzania, Madagascar, Zambia, which have poorly managed economies–were still able to improve their people's health and schooling.

Among the wealthier countries, the physical and educational benefits generally kept pace with improved economies. An exception is the U.S., where the economy flourished in the '80s but social services stagnated and declined.

The chart below lists the world's countries according to their Human Development Index–a measure of quality of life based on life expectancy, adult literacy and Gross National Product per capita. The nations are ranked from lowest quality of life to highest. The "HDI rank minus GNP rank column measures how well the nations translate the wealth they have into benefits for their citizens. A positive number in this column indicates the country makes good use of its resources to help its people. A negative number indicates it does not." humandevel

world

A collection of world demographic information from the year 1987, same year as the humandevel data set, but including more countries. Compiled by yours truly and sourced from: The World Bank

strikes

"The data consist of annual observations on the level of strike volume (days lost due to industrial disputes per 1000 wage salary earners), and their covariates in 18 OECD countries from 1951-1985. The average level and variance of strike volume varies across countries. The data distribution also features a long right tail and several large outliers. The 7 data fields include the following variables: (1) country code; (2) year; (3) strike volume; (4) unemployment; (5) inflation; (6) parliamentary representation of social democratic and labor parties; and (7) a time-invariant measure of union centralization.

These data were analyzed in the forthcoming paper by Bruce Western, "Vague Theory and Model Uncertainty in Macrosociology," which is to appear in Sociological Methodology."

strikes