ESM 201

Assignment 2

Due date: Friday March 13 by midnight

Background

In lecture and readings, we've discussed how crop yield can change over time, and how fertilizer use contributes to such trends. In this assignment, you will be working with two agricultural datasets from the United States Department of Agriculture. The first dataset (`grain_2020.csv`) contains data from the mid 1800s to present day on national production of corn, sorghum, barley, and oats:

The available columns are:

- `hectares_planted`: in million hectares
- `hectares_harvested`: in million hectares
- `production`: in million bushels (sorry about the units, the USDA won't switch to using actual units)
- `yield`: in bushels per hectare

Data from the USDA's Economic Research Service, 2019 Yearbook Tables.

year	commodity	hectares_planted	hectares_harvested	production	yield	price_per_bushel
2015	Barley	1.448776	1.2788078	218.19	170.6199	1.95
2015	Corn	35.620462	32.6783945	13601.96	416.2371	1.95
2015	Oats	1.250480	0.5179981	89.46	172.7033	1.39
2015	Sorghum	3.423644	3.1767851	596.75	187.8471	1.95
2016	Barley	1.238339	1.0400430	199.91	192.2132	1.95
2016	Corn	38.040484	35.1065105	15148.04	431.4881	1.95
2016	Oats	1.145261	0.3965923	64.63	162.9633	1.39
2016	Sorghum	2.707349	2.4928658	480.26	192.6538	1.95

The second dataset (`fertilizer_2020.csv`) contains information on fertilizer use for soybeans, corn, and wheat from 1964 to 2016. The key variable in this dataset is `kgh`, which means kilograms per hectare. These data are originally from the USDA's Economic Research Service, 2019 Fertilizer Use, but I have modified it to use kilograms per hectares instead of pounds per acre.

year	crop	fertilizer	kg_h
2015	Wheat	nitrogen	77.625
2015	Wheat	phosphate	37.125
2015	Wheat	potassium	33.750
2015	Soybeans	nitrogen	19.125
2015	Soybeans	phosphate	57.375
2015	Soybeans	potassium	93.375
2015	Corn	nitrogen	NA
2015	Corn	phosphate	NA
2015	Corn	potassium	NA
2016	Corn	nitrogen	163.125
2016	Corn	phosphate	68.625
2016	Corn	potassium	90.000

Tasks

As with Assignment 1, we recommend doing any data visualization and analyses in R and RMarkdown; however, you're welcome to use whatever language or platform you wish. **You will be turning in a pdf**.

1. Patterns in yield over time

In `grains_2020.csv`, there is a column for `yield`. As shown in *Grassini et al. 2013*, there are six possible statistical models to describe crop yield trends over time (Figure 3 in the paper shows: linear, quadratic plateau, linear piecewise, linear upper plateau, linear lower plateau, and exponential).

a. Create a finalized plot (including title and caption) of yield vs. time for each of the grains in the dataset. (5 points)

b. Which statistical model describes the trend you see for each crop? Justify your choice with the shape of the line you observe. (5 points)

c. **Extra** Choose a crop you like, and fit the statistical model you think it follows. Show your results as a regression table (`geom_smooth` is not a way to fit a model) (2.5 points)

2. Fertilizer use

a. Create a finalized plot (including title and caption) of fertilizer use through time, for each of the crops in the dataset. (5 points)

b. Describe differences in fertilizer use for each crop. What does this suggest about limiting nutrients for each crop? (5 points)

3. Case study: corn yield and fertilizer use

a. For corn, create a finalized plot (including title and caption) of yield vs fertilizer use, separated by the type of fertilizer. (5 points)

b. Explain the relationship between yield and fertilizer you see in the data. Is there a positive linear relationship between fertilizer use and yield? (Hint: recall similar plots presented in lecture) (5 points)

c. **Extra** What explains corn yield? Fit a multiple linear regression with corn yield as the outcome variable, and any relevant explanatory variables. Make sure to transform any relevant variables into units that produce meaningful coefficients. (2.5 points)