An a-maizing analysis: A fun exercise to get us started back with R

By Justin Rasmussen

Introduction

Thank you DiscovR Helpers for being willing to participate as helpers this year! In preparation for that, we will do a little bit of refreshing your R skills, a little bit of looking back at the curriculum, and a little bit of talking about how to teach.

I've prepared this exercise for you all to work on before next week. This exercise is NOT an exam, or what we will learn to call a *summative assessment*. It IS what will learn to call a *formative assessment*. That means it is intended to help YOU and ME understand where you are currently with R and what you need to spend more time on before the workshop. With that in mind, feel free to use your previous work, the DiscovR curriculum, and even each other to answer questions. As you do though, please pay attention to what areas you were stuck, what questions you have, and where you want to spend more time. (You could even send questions over WhatsApp so I can look back at them).

Give this your best effort and try to send me your final product by the end of the week.

Exercise

1. Set up your project.
    a. Open RStudio. If you haven't been back here in a while look around. See if you can identify the following:
        i. Editor
        ii. Console
        iii. Environment
        iv. Files
        v. Help
    b. Buttons to create a project, an R script, an RMarkdown file.
    c. Next, create a new project.
    d. In your computer's file system, navigate to your project folder on your computer and create a folder for "data" and a folder for "reports".
    e. Download the data "kemis_drymaize.csv" from this email into the data folder.
2. Create an RMarkdown file and load the data.
    a. Back in RStudio, create a new RMarkdown file and save it to your "reports" folder.
    b. Load the tidyverse package and the lubridate package.
    c. Import the data.
    d. Look at the variables and make sure they imported correctly.
3. Clean the data.
    a. Select only the variables with values (drop the two empty variables).
    b. Filter out any rows that don't have a value for "classification".
    c. Filter to include just a few counties you're interested in.

    d. Create a new “month” and “year” variables
        i. Use the function months() to extract the month from a date
        ii. Use the function year() to extract the year.
    e. **Advanced**: Create a long dataset where “retail_kg” and “wholesale_kg” are in one column called “sale_type” and prices are in “price_kg”.
4. Make a plot
    a. Create a plot of the retail price of maize over time in the few counties you choose (with geom_point or geom_bar).
    b. Refine your plot.
        i. Filter out some outliers in the data.
        ii. Experiment with different geometries (e.g. poi
    c. **Advanced**: Use the long dataset above to compare wholesale and retail prices in different facets.
5. Create some summary tables.
    a. Create a table with mean and median prices (don’t forget rm.na = T drops missing values).
    b. Create a new table with mean and median prices by month and year.
    c. Turn your new summary table into a bar graph.
    d. **Advanced**: Create another table that looks at total monthly supply.
6. An additional challenge.
    a. With graphs and summary tables, try to argue for or against this statement: “Maize prices are basically a result of supply volume.”
7. Write down what areas were easy and what were hard. What areas would you like to focus on? What specific questions do you have?