Fuel Factsheet Queries

Outline

The repo contains:

- `outbound-master.R` - the script to be run: this is the only file you need to open and run.
- `\scripts\` - there are five sub-scripts in here, which are the workhorses of the master script.They
    - 1. load the required packages,
    - 2. download the data and clean it up,
    - 3. perform the calculations,
    - 4. create the excel file and
    - 5. send the email.
- `\tests\` - there are four test scripts in here, which are called from their respective scripts to test everything went OK in each one (except the email script doesn't have a test).

Flow

In order to run the script you simply source the `outbound-master.R` file.

READ THE INSTRUCTIONS AT THE TOP OF THE SCRIPT.

There are two decision points in the flow:

1. Whether or not the script is being run from a work computer or a remote computer.
2. Whether or not to proceed if sanity checks fail.

Work/homw PC: the first decision is dealt with manually in the `outbound-master.R` script by setting the logical flag `work.computer` to `TRUE/FALSE`, depending on your circumstance. Additionally, if you are not at work, you need to make sure the data folder, and email and password data are also entered correctly. **Right now they are not, because I don't have working credentials.**

Insanity checks: The second decision happens during the script execution. All other tests–if they fail–stop the script. But if the sanity checks fail e.g. the petrol price went up by more than 5p in a week, you get the option to either stop the script or continue it anyway. Because insane things happen sometimes.

Testing

During testing of this script keep in mind the following:

1. You are not using on the real googlesheet.
2. I wasn't able to test the emailing script.

ad 1. The test googlesheet: at the moment the data is being pulled from a test googlesheet, not the actual one. The reason for this is so that the test googlesheet, which is found on this link, can be manipulated to cause tests to fail and test the script.

The test googlesheet contains the 5 sheets required to perform the script, as they appeared in the googlesheet from 22.1.2020. This means one of the tests that will fail is the one to test if the data is from yesterday. You can proceed anyway.

If you don't want to play around with the test googlesheet you can switch to the real one by going into `/scripts/02-download.R` and comment out the test gs assignment in rows 15-17 while uncommenting the actual gs assignment in rows 9-12.

ad 2. The emailing: since I don't have working credentials I wasn't able to test the remote emailing script, and since I am also not on a "work PC" I wasn't able to test the Outlook one either. In principle they should work though, since they are the same as they were before, although I might have made some aestheteic changes, hopefully it will work anyway..

Either way, the script can be tested up until that point even if the email fails. If the email works, it is currently only set to send to Ivo, not everyone else.

1. How exactly does the googlesheet come into being? There are caclulations made in the sheet that I'd probably feel more comfortable making in the script. It's just that I don't like this stuff happening in the googlesheet. I'd much rather just import the raw data and manupulate it programatically in R as opposed to using googlesheet formulas.

2. For example the fuel price over time ten year maximums are not actually calculated in the googlesheet but are **fixed**. Maybe that's OK now, but what if the prices are surpassed? Who will know to manually fix that? Or what happens in 2022, when the maximum is no longer within the past ten years?

3. Some of the checking seems to be duplicating work and then comparing to itself. I'm not sure how that contributes much. Like e.g. oil max min seems to work like this: (i) the raw data is in the gs. (ii) the max and min are caclulated in the gs. (iii) the data is imported into R. (iv) the max and min. are imported into R. (v) the max and min are calculated again in R from the data imported in R. (vi) the calculated and imported max and min are compared.

Why would you do this? What kind of error is this anticipating? In the case of an error, which calculation wins out? I would just do the calculation in R.