Problem Set 8

Keep your rendered `.pdf` to no more than 5 pages. Only provide code in the rendered `.pdf` when it is specifically requested.

1. Revisit the data on fruit prices from the USDA: https://www.ers.usda.gov/data-products/fruit-and-vegetable-prices. Again, download the file linked below "ALL FRUITS – Average prices (CSV format)." Use the data with the `FruitType` and `Preparation` variables as created in the previous solutions.

(a) Create a table of the top 10 most expensive (in terms of price normalized for quantity as measured by `CupEquivalentPrice`) fruit types (as measured by `FruitType`) using `kable`. This table should have 10 rows and be arranged in decreasing order of price. Make sure that the table is self contained and readable. Print the code you use to do this to the rendered `.pdf`.

(b) Create a table of the top 10 least expensive (in terms of price normalized for quantity as measured by `CupEquivalentPrice`) fruit types (as measured by `FruitType`) using `kable`. This table should have 10 rows and be arranged in alphabetical order of fruit type. Make sure that the table is self contained and readable. Print the code you use to do this to the rendered `.pdf`.

(c) Create a table with 7 rows and 5 columns, where each row corresponds to a different type of preparation as measured by `Preparation` (including no information on preparation) and columns correspond to the number of fruits prepared in that way, the average retail price of fruits prepared in that way, the average yield of fruits prepared in that way, the cup equivalent size of fruits prepared in that way, and the cup equivalent price of fruits prepared in that way using `kable`. This table should rows arranged in decreasing order of the number of fruits prepared in each way. Make sure that the table is self contained and readable. Print the code you use to do this to the rendered `.pdf`.

2. Again, find the dataset summarizing some Massachusetts employment statistics called `CESReport.csv`. It was downloaded from here: https://lmi.dua.eol.mass.gov/LMI/CurrentEmploymentStatistics. Use the 170 observations and 15 variables that you obtain after carefully cleaning the data as requested in previous problem sets, removing uninformative rows and converting variables to the appropriate type. Note that you

can limit the data to just the 170 rows that contain information by specifying the `nrow` argument when you apply the `read.csv` function (although other approaches are also acceptable).

(a) Remove observations from 2025 using `filter`. Print the code you use to do this to the rendered `.pdf`.

(b) Create a new variable called `YTD` which is the sum of the month variables, omitting missing values. Print the code you use to do this to the rendered `.pdf`. Note: You may want to revisit `apply` or `rowSums`.

(c) Using `group_by` and `mutate`, create a new variable called `mYTD` which gives the average value of `YTD` across years within sector, as described by `Description`. Print the code you use to do this to the rendered `.pdf`.

(d) Using `kable`, `arrange`, and `select`, create a table that provides `Year`, `Description`, `YTD`, and `mYTD` for the years and sectors that had the three highest values of `YTD`/`mYTD`. Make sure that the table is self contained and readable. Print the code you use to do this to the rendered `.pdf`.

(e) Using `kable`, `arrange`, and `select`, create a table that provides `Year`, `Description`, `YTD`, and `mYTD` for the years and sectors that had the three lowest values of `YTD`/`mYTD`. Make sure that the table is self contained and readable. Print the code you use to do this to the rendered `.pdf`.

(f) Explain what you observe in (c) and (d) in at most two sentences.

(g) Using `kable`, `arrange`, and `summarize`, create a table with 1 column and 34 rows that is the sorted list of sectors ordered from most to least variable across years in terms of year to date employment as measured by `YTD`. Make sure that the table is self contained and readable. Print the code you use to do this to the rendered `.pdf`.