

1.Wstęp

Projekt opierał się na analizie danych importu i eksportu Kawy na całym świecie liczonych w 60kg workach. Znajdujemy współczynnik importu do eksportu na przestrzeni lat i daty w których pojawiają się ekstrema. Przystawiamy na wykresie słupkowym jak rosła suma sprzedawanych i skupowanych worków, na wykresach kołowych jaki procentowy udział w handlu miał każdy kontynent i finalnie na mapie świata jako kartogram jak zmieniał się import i eksport na przestrzeni lat.

2.Dokumentacja

Przeanalizowaliśmy dane dostępne pod adresem https://www.ico.org/new_historical.asp, do obróbki wybraliśmy te dotyczące importu i eksportu, bo uznaliśmy je za najciekawsze.

Odpowiednie spreparowanie danych obejmowało usunięcie pustych wierszy, a także wypełnienie pustych komórek zerami w miejscach gdzie było to potrzebne, żeby w późniejszej pracy nie pojawiały się w tych miejscach wartości NA.

Pierwszy napotkany problem zastał nas już na etapie wczytywania pliku, gdzie wszystkie dane domyślnie były w formacie znakowym, a do naszego zadania potrzebujemy wartości numerycznych. Żeby to zmienić najlepiej by było, żeby nazwy krajów były nazwami kolumn, ale napotkaliśmy dziwne „x” przed datami, które prawdopodobnie wynikają z formatu pliku (.csv).

Rozwiązaliśmy problem z importem zmieniając jeszcze trochę plik csv by upewnić się że komórki są w formacie liczbowym (i tak je wczyta jako character ale później nie będzie NA po konwersji). Udało się ustawić nazwy Krajów jak nazwy wierszy i zmienić wartości na numeryczne bez żadnych błędów.

Kod R i plik z notatkami przesłałem do stworzonego przez siebie repozytorium na Github.

Zacząłem obróbkę, dodałem współczynnik importu do eksportu

Pozbyliśmy się napotkanych „x” przed datami, po prostu nadpisując je datami wpisanymi ręcznie. Obliczyliśmy współczynnik importu do eksportu.

Przejrzeliśmy jakie wartości/grafy się zazwyczaj wyciąga z danych o imporcie/eksporcie i zdecydowaliśmy się na wykres słupkowy z linią oznaczającą współczynnik. Po próbach wykonania tego funkcją „barplot”, uznaliśmy że lepszą opcją będzie zgłębienie pakietu ggplot2. Znaleźliśmy na stronie „www.r-graph-gallery.com”, graf, który nam się podobał i spreparowaliśmy dane tak, by miały strukturę tych z przykładu i zadziałało to bardzo dobrze. Dodałem też funkcje która znajduje rok z najmniejszą/największą sumą importu/eksportu.

Postanowiliśmy usiąść do projektu na żywo, efektem dość długotrwałych starań są dane importu i eksportu na mapkach świata. Główną trudnością było zaadoptowanie sposobu tworzenia mapy do naszych danych. Specjalny plik z którego za pomocą funkcji plot tworzy się mapę ma wbudowane dane typu populacja itd., ale

żeby wybrać stamtąd tyle te kraje dla których mamy dane i te dane przedstawić na tej mapie trzeba było bardzo dużo kombinować.

Podczas tworzenia mapy napotkaliśmy multum problemów. W imporcie mamy dane głównie z Europy i nieporównywalnie większą wartość w Ameryce liczonej jako całość, więc po próbach przeskalowania tego jakoś tak, by było realnie widać różnice pomiędzy większością krajów, postanowiliśmy zostawić import jako mapę Europy. Co także sprawiło multum problemów, ale w końcu skopiowaliśmy część danych do osobnej zmiennej, i z niej wyjęliśmy tylko te, które my mamy w naszej tabelce i udało się przejrzystej przedstawić dane. Kolejne problemy były przy palecie kolorów, bo trzeba było ją precyzyjnie dostosować żeby było widać różnice w wartościach, były też problemy z indeksami, przez co kolorowały się złe kraje i żeby to rozwiązać (a bardziej nawet zauważyć) musieliśmy się poduczyć geografii. Później okazało się że części krajów nie ma na mapie świata, bo nie jest w 100% dokładna (nie były to duże straty, rzędu 5 rekordów), ale trzeba było to dalej uwzględniać. Z tą wiedzą szybciej już poszło stworzyć mapę eksportu, gdzie wzięliśmy cały świat bo kraje eksportujące rozkładają się generalnie wzdłuż równika. Finalny problem to animacja. Mapa świata która znaleźliśmy i wgraliśmy do naszego kodu działa bardzo dobrze, ale jest zbudowana do bazowego R, a konkretnie funkcji „plot”, więc jeśli chcielibyśmy użyć funkcji animacji z pakietu ggplot musielibyśmy zaczynać wszystko od nowa. Uznaliśmy że na potrzeby projektu wyeksportujemy mapy jako png (jest to zaledwie 60 obrazków, bo dane są z 30 lat) i złożymy w gif w zewnętrznym konwerterze.

Do utworzenia wykresu kołowego przedstawiającego procentowy udział kontynentów eksportujących kawę w latach 1990-2019 potrzebowałem odpowiednio:

- utworzyć ramki danych zawierające państwa i sumy wyeksportowanych paczek kawy w latach 1990-2019 i przypisać je do zmiennych o nazwach kontynentów (dane były wybierane z ramki danych o nazwie „eksport”, a państwa odpowiednio przypisane do kontynentów na których się znajdują)

- zsumować dane liczbowe dla każdego kontynentu

- utworzyć ramkę danych o nazwie „kontynenty”, z której później wybiorę osobno nazwy kolumn i dane liczbowe i przypisze je do zmiennych potrzebnych do utworzenia wykresu

Dodałem funkcję, która oblicza odchylenie współczynnika importu do eksportu.

3.Wnioski

Przeglądając dane i wykresy dość jasno rzuca się w oczy parę ciekawych rzeczy o rynku kawy. Współczynnik importu do eksportu ma dość małe odchylenie, na początku z roku na rok większa była raz jedna, a raz druga suma, ale od roku 2002 zawsze wyższy jest import. Import najniższy był w roku 1991, czyli prawie na początku naszych danych, za to Eksport najniższy był w 1995, wartości najwyższe zarejestrowane zostały jak można się domyślić w roku 2019. Dość uderzająco widać że Import (oprócz Stanów Zjednoczonych) skupia się w Europie, nie miało sensu

branie całego świata przy tworzeniu wykresów dla importu. W Eksporcie dominują zaś kraje które leżą blisko równika, czyli kraje Ameryki Środkowej, Afryki i z południa Azji.