Az adatvizualizációm több lépésben készült el:

1. Adatok tisztítása
2. Vizualizációs felület előkészítése
3. Adatok összevetése és vizualizációja

# Adatok tisztítása:

Az adatokat általam írt C++ és Python programocskák segítségével, valamint én tisztítottam.

Előszőr a születési hely adatait (birth) egyesítettem, ezeket a lakhely adataival vetettem össze így megkapva a születési helyen maradt és onnan elköltözött személyek arányát, valamint a vizualizáció utolsó pontjához, a születési helyek összevetéséhez is felhasználtam.

A második tisztítást a szakterületet illető adatokon (workwhat) végeztem ezeket 5 kategóriára osztottam fel:

1. Egészségügy
2. Tanügy
3. Jog
4. I.T.
5. Más

Szerintem ez volt az adattisztítás legnehezebb része, mivel a C++ korlátjai és még elég kezdő szintű Python tudásom miatt lehetetlenségnek tűnt az adatokat egyesíteni és feldolgozni. Sok gondolkozás után viszont arra jutottam, hogy a Python tudja pótolni a C++ hiányosságait. Az eredeti Json fileból egy C++ program segítségével leegyszerűsítettem és egyesítettem a munkaterületeket. Ezeket az adatokat majd egy Python kód segítségével kivettem az eredeti Json file-ból és az interneten talált egyik Json készítő applikációval egy új Json filet készítettem.

A vizualizáció utolsó alpontjánál , a születési helyek összevetésénél egy térképes vizualizációt készítettem. Ehhez az internetről szerztem be a születési helyek koordinátáit. Ezeket saját magam fordítottam le (mivel a települések románul voltak feltűntetve) és rendszereztem egy CSV file-ban.

# Vizualizációs felület előkészítése:

A vizualizációs felületnek egy hozzám közel álló felületet a HTML-t választottam. Az adatvizualizáció Javascript - ben készült a p5.js könyvtárat alkalmazva. Ezt a könyvtárat már sokszor alkalmaztam és nagyon megkönnyíti a munkát valamit vizuális programok készítésére volt kifejlesztve, így egyből ráesett a választás a vizualizációt illetően.

Az oldal háromnyelvűsége (magyar, román és angol) volt az első dolog amit elkészítettem, ezután jött maga a vizualizációs felület. Előszőr a felület felosztásán gondolkodtam mivel akartam egy rendszert ami alapján működik a vizualizáció. Ez a rendszer egy menüpontos felosztás lett amit gombnyomással lehet irányítani. Elkészült az üdvözlő és elköszönő szöveg valamint három menüpont is. A színválasztás elég sötét lett mivel sokat dolgoztam este és éjjel is és így oldottam meg, hogy ne sértse a szememet, bevallom jobb lett volna ha egy éjjszakai és nappali módot is készítek hozzá.

# Adatok összevetése és vizualizációja:

Az adatvizualizáció 5 részből áll:

1. Munka
2. Szakterület
3. Házasság
4. Költözés
5. Születési hely

A vizualizáció első 4 pontja egyszerű oszlop grafikonos vizualizációk ahol a munkanélküli és alkalmazott , a különböző szakterületű , házas és nem házas, születési helyén maradt és onnan elköltözött személyek arányát mutattam be.

Az 5. menüpont a második fordulóra készült térképes vizualizáció, melynek készítése számomra a legtöbb élvezetet szerezte. A térképet saját magam készítettem elő a mapbox API térképszerkesztőjével, majd ezt a térképet importoltam a vizualizációs felületre. A születési helyek adatait felhasználva meghatároztam az ezeket ábrázoló körök méretét, minél több ember született az adott településen annál nagyobb a kör sugara. A köröket az internetről letöltött hosszúsági és szélességi koordináták segítségével helyeztem el a térképen, ezután eltérő színeket adtam nekik, majd feltüntettem minden körnek az általa ábrázolt települését.

Az utolsó lépés a kis információs ikon elkészítése volt , hogy aki az oldalra téved tudja , hogy miről is van szó.

# Összegzés

Az adatok általi csodálatom az r/dataisbeautiful subredditen kezdődött és még régen eldöntöttem, hogy majd egyszer én is készítek egy vizualizációt, viszont nem tudtam volna ennyi adatot begyűjteni. Ez a pályázat egy álmomat valósította meg. Sokat tanúltam és mindent élvezettel csináltam. Köszönöm a lehetőséget!