

AMAÇ

Veri görselleştirme yöntemleri, karmaşık verileri daha anlaşılır bir biçime dönüştürmek için görsel olarak ifade edilmesini sağlayan önemli istatistiksel araçlardır. Teknolojinin hızla ilerlemesi, veri miktarının artması ve sosyal medya kullanımının yaygınlaşmasıyla da veri görselleştirme üzerine farklı çalışmalar daha da artmıştır. Bu çalışmada, tarihten günümüze farklı grafik yöntemlerinin incelenmesinin yanı sıra farklı teknolojiler kullanarak iki veri setinin interaktif olarak görselleştirilmesi üzerine bir uygulama yapılmıştır.

Bunun için Dünya Sağlık Örgütü'nden (WHO) çeşitli ülkelerde yıllara göre hava kirliliği ölçüm değerleri ve yine aynı ülkelerde yıllara göre solunum yolu hastalıklarına bağlı ölüm istatistiklerini kapsayan iki veri setini farklı açılardan birlikte görselleştirmeye çalıştık. Grafikler için Python programlama dilinin plotly kütüphanesi, harita temelli görselleştirmeler için de OpenStreetMap verileri kullanılmıştır. Front-end kısmında ise Django ve Dash kullanılmıştır.

HEDEFLER

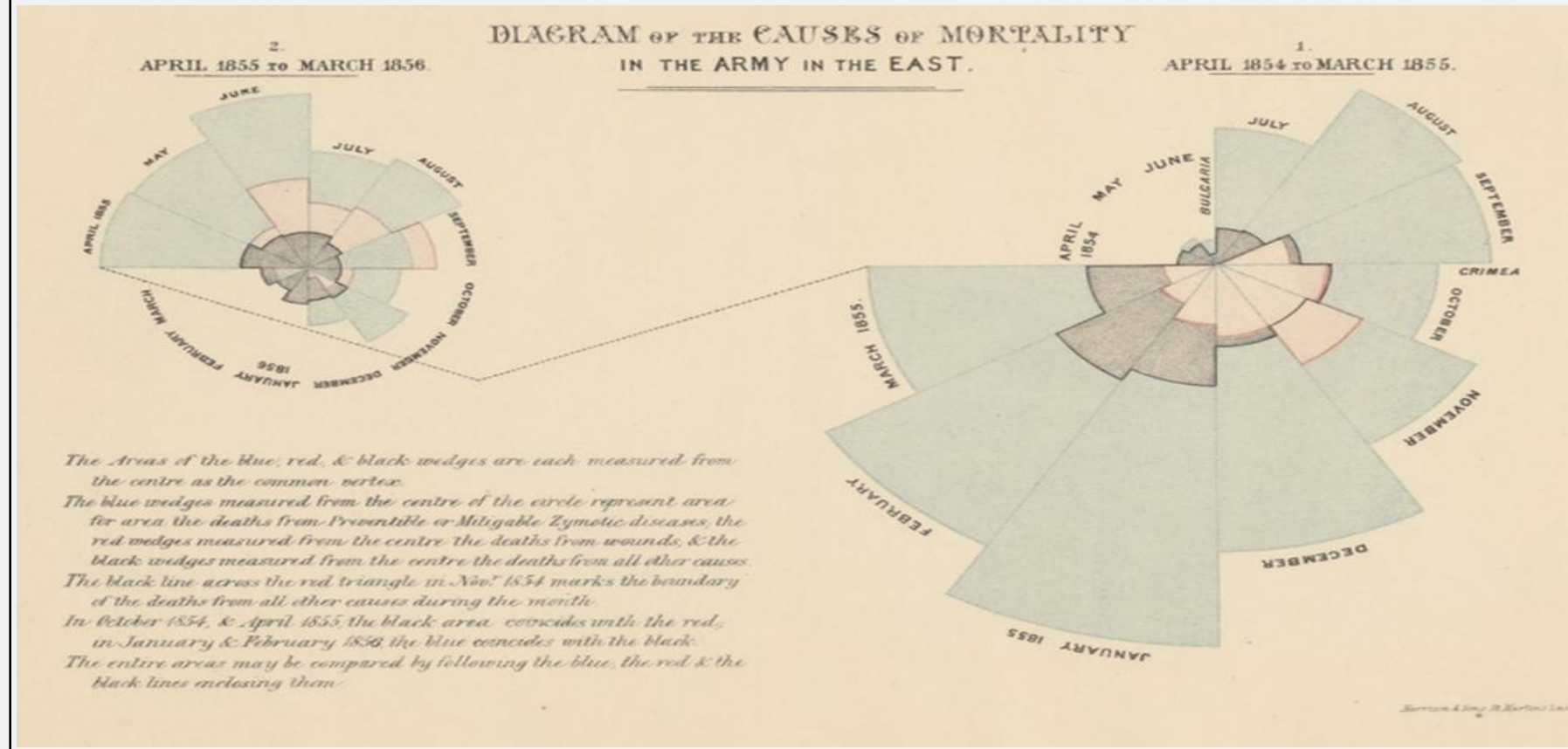
Bu çalışmada farklı grafik türleriyle ilgili detaylı bir literatür incelemesi yapılması ve çeşitli veri setlerinin farklı teknolojiler kullanılarak görselleştirilmesi hedeflenmiştir. Bu kapsamda elde edilen iki ayrı veri setinde "hava kirliliği istatistikleri" ve "solunum yolu hastalıklarından kaynaklı ölüm istatistikleri" verilerinin interaktif grafiklerle görselleştirilmesi yardımıyla veriler arasındaki olası ilişkilerin daha ileri analizlerle incelenmesine katkı sağlamak ve hava kirliliğinin ve solunum yolu hastalıklarının etkileri konularında farkındalığın artırılması hedeflenmiştir.

LİTERATÜR

Veri türlerine göre oluşturulabilecek çok sayıda grafik bulunmaktadır. Bunlardan bazıları Sütun, Çizgi, Pasta, Dağılım, Radar, Alan, Nokta, Balon, Polar, Pareto grafikleridir. Yüzyıllar boyu insanlığın gelişmesiyle beraber kullanılan veri görselleştirme teknikleri de bilimle beraber gelişmiştir.

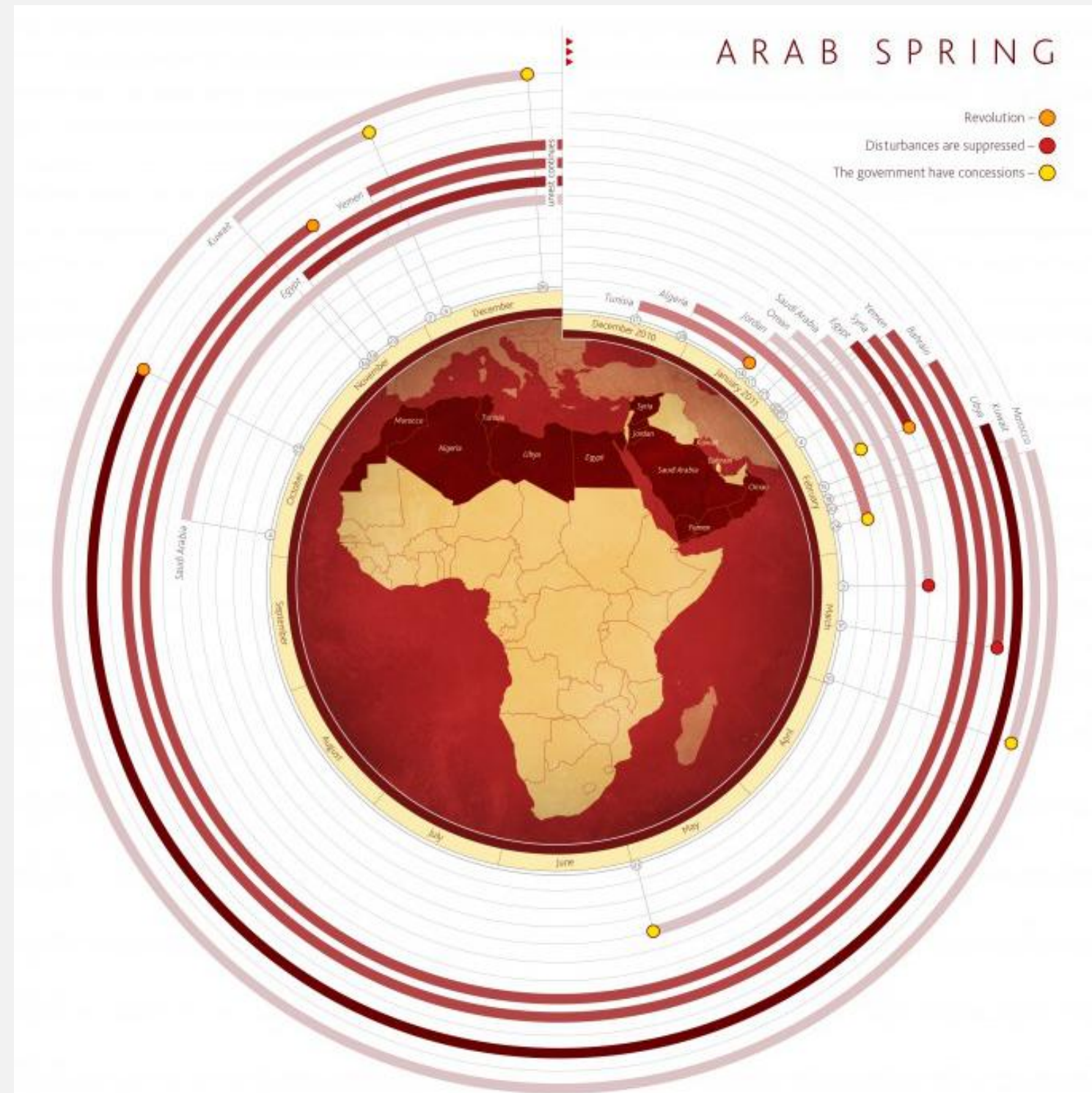
Bu tekniklerden birisi Şekil 1'de yer alan Gül grafiğidir (Rose Graph). İngiliz hemşire ve istatistikçi Florence Nightingale tarafından 19. yüzyıl ortalarında Kırım Savaşı sırasında ölen

askerlerin istatistiklerini görselleştirmek için geliştirilmiştir. Bu grafikte belirli bir zaman aralığında meydana gelen olayların zaman içindeki dağılımını göstermek için dairesel bir yapı kullanılır.



Şekil 1. Gül grafiği

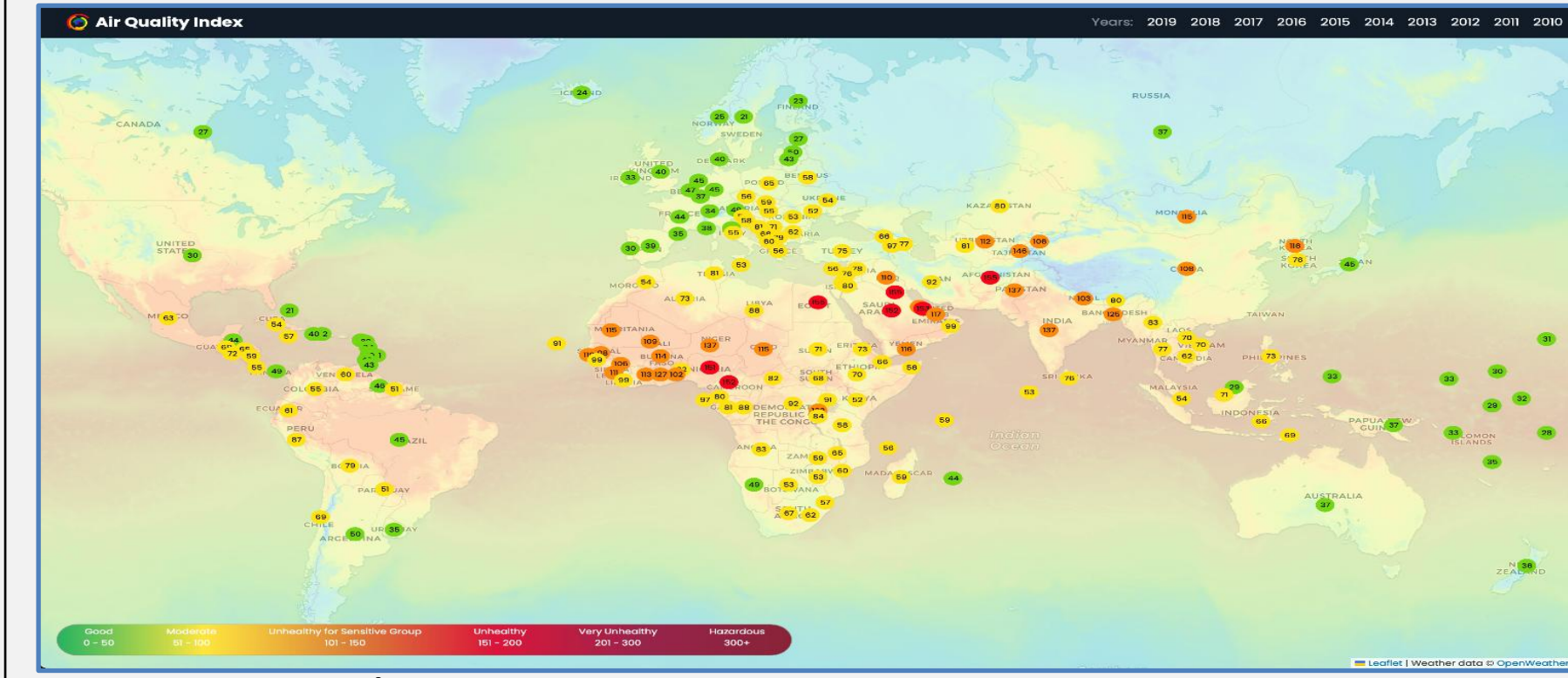
Günümüzün çarpıcı görselleştirme örneklerinden olan Kirill Khachaturov tarafından tasarlanan ve Şekil 2'de yer alan Arap Baharı veri görselleştirme çalışmasıdır. Bu grafikte süreç bir zaman çizelgesiyle ifade edilmiştir. Afrika ve Arap Yarımadası'nın yer aldığı bir dünya haritası, sürece dahil olan ülkeleri ve zaman çizgisini vurgular. Devrimlerin, bastırılan olayların ve hükümet tavizlerinin temsil edildiği farklı renk tonlarıyla işaretlenmiş daireler ise olayların sonuçlarını gösterir.



Şekil 2. Arap baharı veri görselleştirmesi

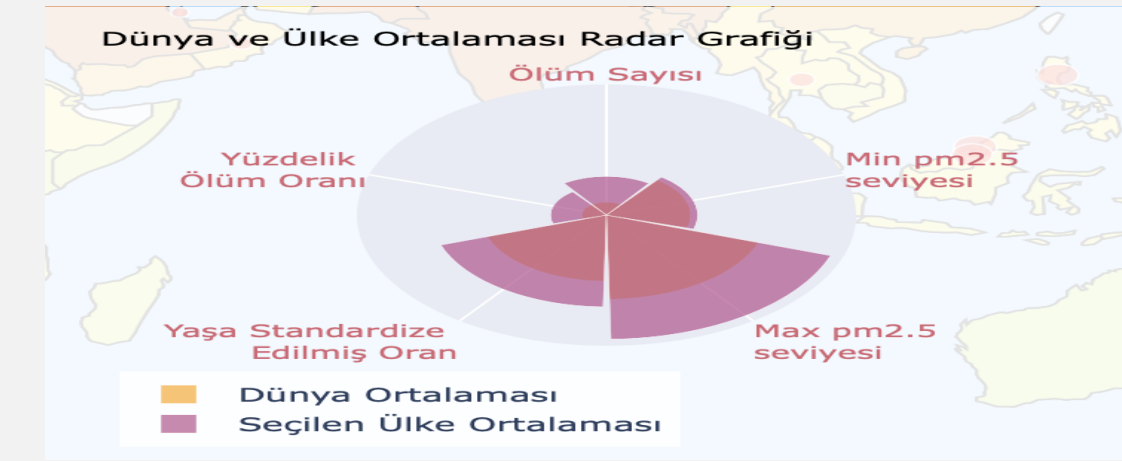
UYGULAMA

Geliştirdiğimiz uygulama üzerinde ısı haritası filtresi uygulayarak görselleştirdik ve interaktif bir arayüz oluşturduk.

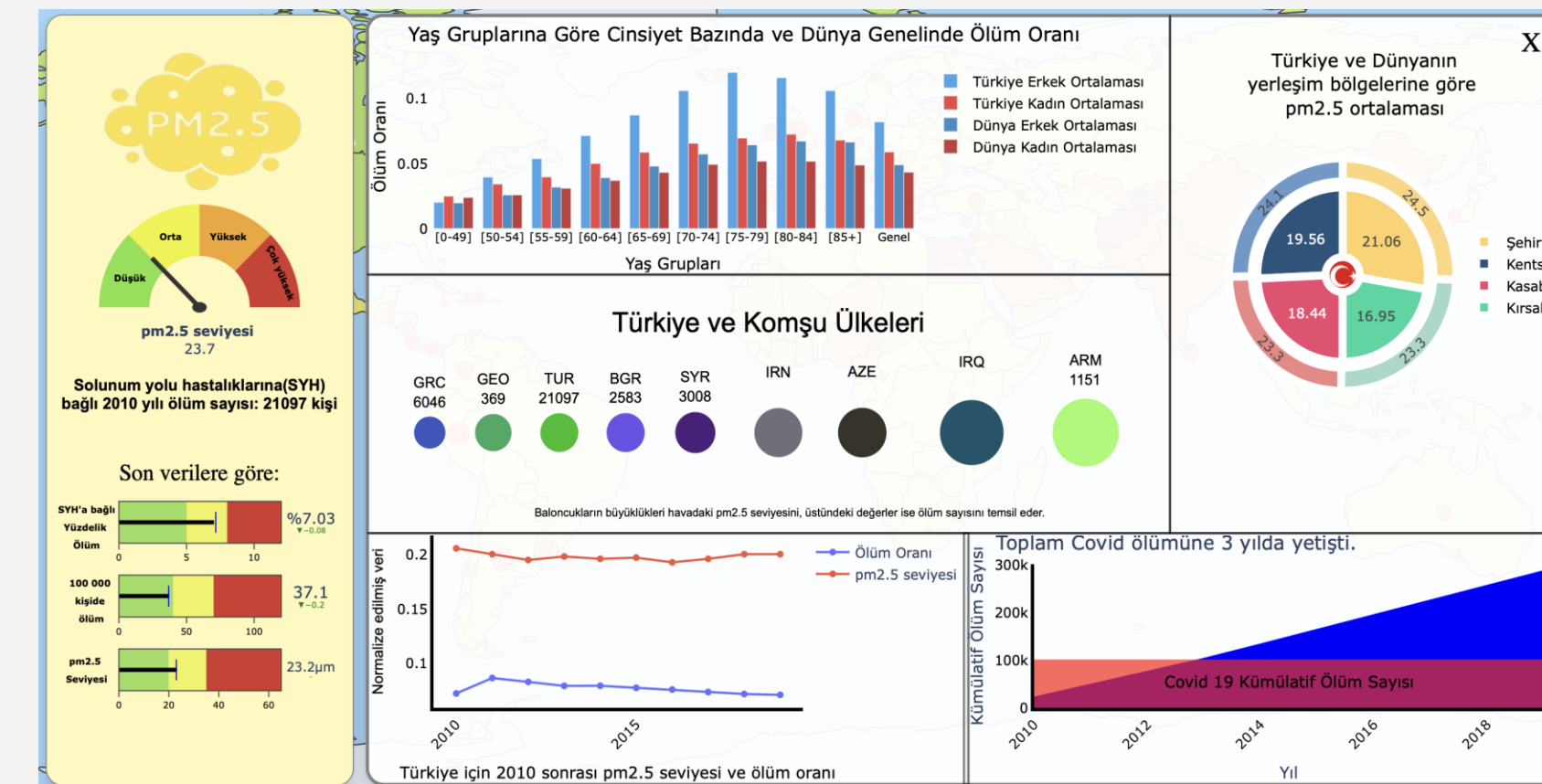


Şekil 3. İki veri setine ait çeşitli bilgi ve grafiklerin yer aldığı arayüz

Radar grafiği için Gül grafiğinden esinlenilmiş ve kursörün bulunduğu ülkeye ait çeşitli istatistikler Dünya ortalamaları ile birlikte gösterilmektedir.



Şekil 4. Radar grafiği



Şekil 5. İki veri setine ait çeşitli grafiklerin bulunduğu ekran

Her bir ülkenin sağlık ve çevre verileri yıllara göre filtrelenerek 7 farklı grafik ile görselleştirilmiştir. Bunlar, yaş ve cinsiyete göre ölüm oranlarını dünya ortalamasıyla karşılaştıran histogram, yerleşim yerlerine göre hava kirliliği seviyelerini gösteren iç içe pasta grafiği, komşu ülkelerle PM2.5 seviyeleri ve ölüm oranlarını karşılaştıran balon grafiği, yıllara göre ölüm oranı ve hava kirliliği seviyelerini gösteren dağılım çizgi grafiği, COVID-19 ile solunum yolu hastalıklarından kaybedilen insanların karşılaştırılmasını sağlayan çizgi sütun grafiği ve PM2.5 seviyelerini gösteren indikatör, son yıllardaki verilerin karşılaştırılmasını sağlayan mermi delta gösterge grafiğidir.

SONUÇ VE ÖNERİLER

Yaptığımız uygulama ile ülkelerdeki hava kirliliğinin yıllara ve alanlara göre değişimini ve solunum yolu hastalıklarını bir arada görselleştirdiğimiz bir arayüz oluşturduk. Bu arayüzü oluştururken Django ve Dash gibi iki farklı frameworkün performanslarını web platformuna veri görselleştirme bileşenleri entegre ederken inceledik. Django'nun sağladığı birçok modül ve paket, çeşitli görselleştirme kütüphanelerine erişim sağlamıştır. Ancak büyük ölçekli verilerle çalışırken performans sorunlarıyla karşılaştık. Dash kullanırken ise özelleştirme seviyesinin diğer frameworklere kıyasla düşük olduğunu gördük. Bu durum, Dash'in arayüz tasarımında sıklıkla kullanılan HTML/CSS/JS gibi teknolojileri etkin bir şekilde kullanamamasından kaynaklanmaktadır. Veri görselleştirmesi için veri setinin boyutu küçültülebilirse, veri işleme tekniklerini uygulayarak işlenmiş veri setini Django'ya entegre etmek en uygun yaklaşım olabilir. Ancak, veri seti bu işlemler için uygun değilse, Dash gibi daha performans odaklı bir çözüm tercih edilebilir. Bu durumda, Django'nun bazı özelleştirmelerinden vazgeçilerek daha akıcı bir kullanıcı deneyimi sağlanabilir. Dash, büyük veri setlerini işlemek ve hızlı bir şekilde görselleştirmek için ideal bir seçenek olabilir. Her iki durumda da, veri işleme süreci ve görselleştirme aşaması projenin gereksinimlerine ve veri setinin özelliklerine göre ayarlanmalıdır.

Ek olarak, oluşturulan grafikler ile hava kirliliği değerleri ve solunum yolu hastalıklarına bağlı ölüm oranları arasında olması beklendiği gibi pozitif bir ilişki olabileceği gözlenmiştir. Çalışmanın, teknoloji boyutu dışında da toplum sağlığı ve çevre kirliliği konularına daha çok dikkat çekilmesi ve oluşturulan bu grafiklerin de detaylı olarak incelenerek ileri analizlere ışık tutması ümit edilmektedir.

KAYNAKÇA

- C. Chen & Wolfgang Härdle & Antony Unwin (2008). *Handbook of Data Visualization*. Springer-Verlag Berlin Heidelberg, Berlin, Germany.
- WEB_1. (2023). Amazon web site. <https://aws.amazon.com/tr/what-is/data-visualization/>
- WEB_2.(2022). Scientific American web site. <https://www.scientificamerican.com/article/how-florence-nightingale-changed-data-visualization-forever/>, 01/08/2022
- PM2.5 Dataset <https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database>, 13/05/2024
- Respiratory Diseases Dataset <https://www.platform.who.int/mortality/about/about-the-who-mortality-database> 13/05/2024